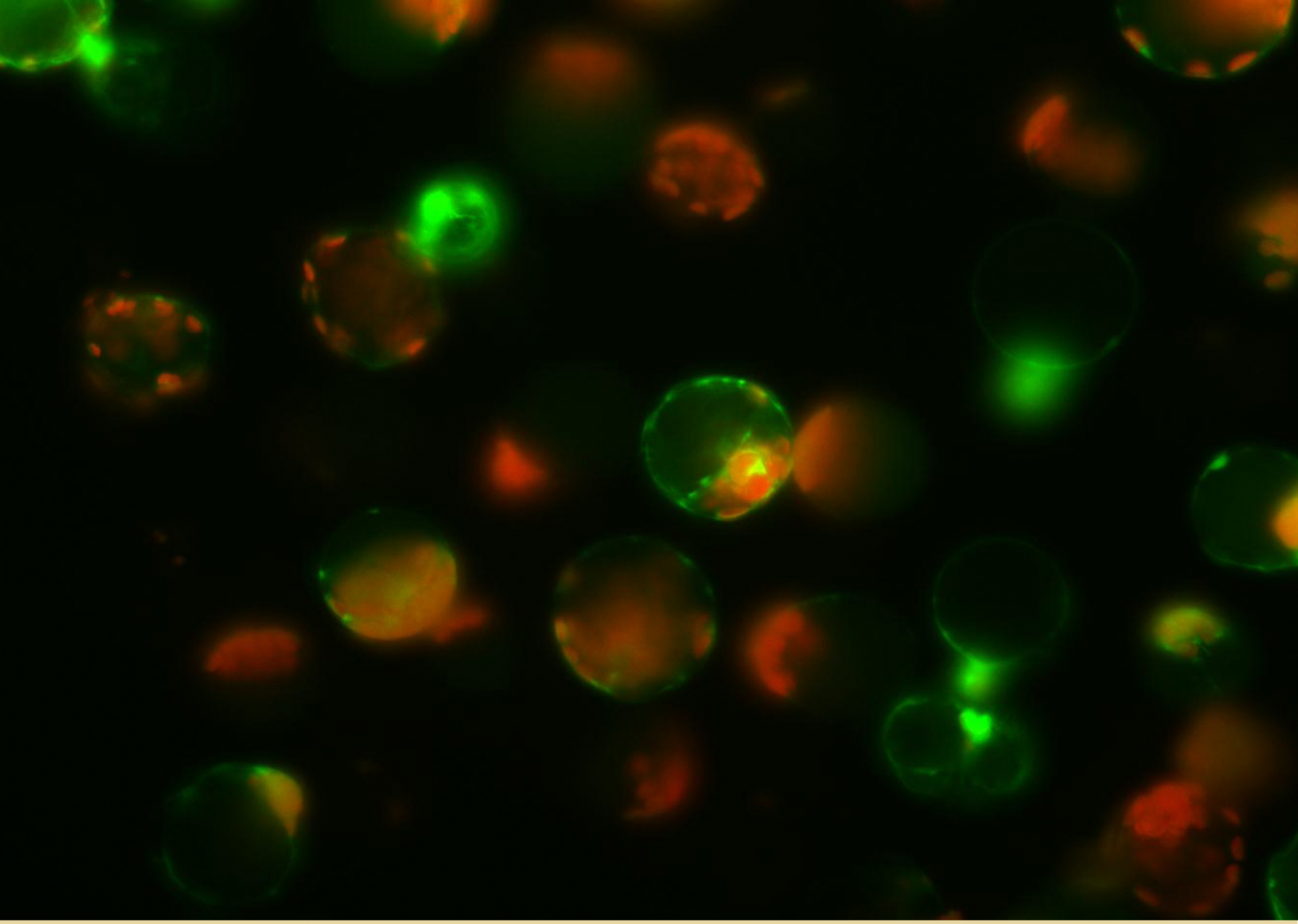


André Filipe Marques Soares

RLR1 AND RLR2, TWO NOVEL ARABIDOPSIS THALIANA ATYPICAL ASPARTIC PROTEASES INVOLVED IN PRIMARY ROOT DEVELOPMENT AND LATERAL ROOT FORMATION

Tese de Doutoramento em Biologia Experimental e Biomedicina, ramo de Biologia Molecular, Celular e do Desenvolvimento, orientada pelas Doutora Isaura Simões e Professora Doutora Alice Y. Cheung, e apresentada ao Instituto de Investigação Interdisciplinar da Universidade de Coimbra.

Julho de 2016

· U C ·

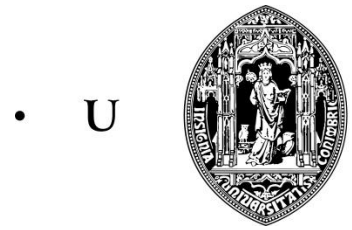
Universidade de Coimbra

André Filipe Marques Soares

RLR1 and RLR2, two novel *Arabidopsis thaliana* atypical aspartic proteases involved in primary root development and lateral root formation

2016

Thesis submitted to the Institute for Interdisciplinary Research of the University of Coimbra to apply for the degree of Doctor in Philosophy in the area of Experimental Biology and Biomedicine, specialization in Molecular, Cell and Developmental Biology



This work was conducted at the Center for Neuroscience and Cell Biology (CNC) of University of Coimbra and at Biocant - Technology Transfer Association, under the scientific supervision of Doctor Isaura Simões and at the Department of Biochemistry of University of Massachusetts, Amherst, under the scientific supervision of Doctor Alice Y. Cheung. Part of this work was also performed at the Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, under the scientific supervision of Doctor Herta Steinkellner and also at the Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, Jülich, under the schientific supervision of Doctor Pitter F. Huesgen.

André Filipe Marques Soares was a student of the Doctoral Programme in Experimental Biology and Biomedicine coordinated by the Center for Neuroscience and Cell Biology (CNC) of the University of Coimbra and a recipient of the fellowship SFRH/BD/51676/2011 from the Portuguese Foundation for Science and Technology (FCT). The execution of this work was supported by a PPP grant of the German Academic Exchange Service with funding from the Federal Ministry of Education and Research (Project-ID 57128819 to PFH) and the Fundação para a Ciência e a Tecnologia (FCT) (grant: Scientific and Technological Bilateral Agreement 2015/2016 to IS)

Agradecimentos/Acknowledgments

Esta tese e todo o percurso que culminou na sua escrita não teriam sido possíveis sem o apoio, o carinho e a amizade de várias pessoas que ainda estão ou estiveram presentes na minha vida. Sem estas pessoas eu tenho a certeza que nada disto seria possível portanto aqui fica o meu muito obrigado por todo o apoio que me deram. Os próximos parágrafos são a expressão escrita de todo o meu agradecimento mas todas as palavras serão sempre poucas para vos agradecer!

Em primeiro lugar, gostaria de agradecer ao Professor Doutor Carlos Faro e ao Professor Doutor Euclides Pires por terem proporcionado todas as condições necessárias para a realização deste trabalho quer na Unidade de Biotecnologia Molecular do Biocant, quer no Laboratório de Biotecnologia Molecular do Centro de Neurociências e Biologia Celular de Coimbra.

Se nos agradecimentos da minha tese de Mestrado me faltaram as palavras para poder agradecer devidamente à Doutora Isaura Simões, nesta tese de Doutoramento todas as palavras serão sempre poucas para lhe agradecer tudo o que fez e continua a fazer por mim. Muito obrigado por voltar a apostar em mim Professora e por me ter acompanhado ao longo destes quase 5 anos! Obrigado por ter acreditado em mim quando eu próprio duvidava. Obrigado por toda a sua preocupação e por tudo o que me ensinou ao longo de todos estes anos que trabalhámos juntos. Se sou o cientista que sou neste momento devo-o a si e a tudo o que me ensinou! Tal como escrevi na minha tese de Mestrado, espero sinceramente não a ter desiludido Professora.

To Dr. Alice Cheung and Dr. Hen-ming Wu I would like to thank the opportunity that they gave me to work in their lab at the University of Massachusetts, Amherst. Thank you for giving me all the conditions necessary to perform my work in your lab and for all the things that I learn during my stay. Especially I would like to express my deepest gratefulness to Dr. Alice for agreeing to be the co-supervisor of this thesis. Thank you for all the time and effort that you spent guiding me and discussing with me my results and ideas. Thank you very much.

To Dr. Pitter Huesgen I would like to thank the opportunity of working with you in your laboratory at Forschungszentrum Jülich. Thank you very much for all the time that you spent and continue spending analyzing my data and for all the great ideas that you gave me.

To Dr. Herta Steinkellner I also would like to thank the opportunity of working in your lab at BOKU University, Vienna. Thank you very much for allowing me in your lab and for giving

me all the conditions to express both proteins in *Nicotiana* plants. It was a great opportunity for me and also a great experience!

A todas as pessoas que fazem parte da Unidade de Biotecnologia Molecular do Biocant ou que por lá passaram enquanto eu lá trabalhei, o meu muito obrigado por tornarem o laboratório um sítio onde não só se trabalhava à séria mas onde também era possível dar umas boas gargalhadas! Apesar de ter andado a saltitar de um lado para o outro ao longo deste doutoramento, era no nosso laboratório que eu mais me sentia em casa e agora que parto tenho a certeza que vou sentir sempre a vossa falta!

Queria agradecer também à Dona Alda, à Dona Adelaide e à Filipa todo o trabalho que tiveram comigo ao longo destes anos, por todas as provetas, frascos e fraquinhos que tiveram de lavar à minha conta! Muito obrigado por todo o apoio que me deram, por todas as gargalhadas partilhadas e pela forma incansável com que não só me apoiavam a mim mas também ao resto das pessoas do Biocant. Aproveito também para agradecer a todos os meus amigos que ainda estão ou já estiveram no Biocant e que ajudaram a fazer do Biocant muito mais do que um simples local de trabalho. Muito obrigado!

To the people from the Amherst, Vienna and Jülich Laboratories, I thank you for the very good environment that I experience during my stay in your labs and for all the help that you were always willing to give me! A especial thanks to Yanjao (Amherst lab) for helping me taking my first baby steps in plant/pollen research, to Andreas and Alexandra (Vienna Lab) for all the help and advices, and finally to Stefan (Jülich lab) for all the time that you spent (and continue spending) doing my proteomics experiments and analyzing my data! Thank you very very much for everything!

A todos os meus amigos da X edição do BEB, Catarina, Tânia, Sara, Joana, Dominique, Mariline, Sofia, Patrícia, António, Marcelo e Ribas, o meu muito obrigado por serem os amigos que são. O primeiro ano do Programa Doutoral foi sem dúvida o melhor ano que passei em Coimbra e se assim foi, devo-o a todos vós! Deixo um agradecimento especial à Nina e à Tânia pela vossa amizade e por estarem sempre presentes e dispostas a dois dedos de conversa sobre tudo ou mesmo sobre nada! Aos restantes amigos de Coimbra e arredores, o meu muito obrigado por todo o apoio que me têm dado!

Deixo aqui também um agradecimento especial à Tábata e à Kaline, duas pessoas que tive a sorte de conhecer em Amherst e que tornaram a minha estadia naquele canto do mundo bastante mais alegre. Meninas, eu se calhar devia escrever este agradecimento em inglês porque eu não sei se vocês vão perceber este meu “Old Portuguese” mas o dicionário existe para essas coisas! Eheh!

To the people that I met during my stay in Vienna, thank you very much for making those 3 months that I spent there unforgettable. A special thanks to my great friend Kemo, a.k.a. “The Bosnian Guy”, and to Anja and Heidi! Those last two months that I spent in your company were truly great and I will never forget them!

Aos meus Pais, Avós e Irmã agradeço todo o incansável apoio que me deram ao longo de toda a minha vida. Eu posso andar sempre de um lado para o outro, a saltitar de país em país mas sei que terei sempre o vosso apoio e o vosso carinho! Muito obrigado por terem estado sempre presentes na minha vida, mesmo quando era eu que parecia estar mais ausente!

O meu último e mais especial agradecimento vai para a Maria. Nós só nos conhecemos a meio desta jornada mas sem ti tudo isto teria sido bem mais difícil. Muito obrigado por toda a força que me deste especialmente quando estávamos separados por todo um oceano. Muito obrigado por acreditares sempre em mim mesmo quando eu próprio duvidava! Acima de tudo, muito obrigado por fazeres parte da minha vida desta forma tão especial.

Abstract

Aspartic proteases (APs) represent the second largest class of plant proteases after serine proteases. Members of the pepsin-like family are widely distributed in plants, *e.g.* 70 APs homologues are found in *Arabidopsis thaliana* and 166 in *Oryza sativa*, the vast majority of them having atypical characteristics and properties (atypical APs). This contrasts strikingly with the considerably fewer number of APs encoded in the mammalian genomes (e.g. *Homo sapiens* has only 17 APs). The overrepresentation of APs in plants suggests potentially important and diverse roles for these proteins. Although plant APs have been much less studied than other protease classes, some functions are starting to be uncovered, with proposed roles in highly regulated processes like resistance to biotic and abiotic stresses, programmed cell death (PCD), plastid homeostasis and reproduction, which is consistent with functional specialization of plant APs and tight activity regulation.

Our major goals in this thesis were to describe the *in vivo* functions, the biochemical properties, the *in vivo* substrates and molecular pathways of two putative atypical APs from the model plant *Arabidopsis*: At2g03200 and At4g30040 gene products. At2g03200 expression is regulated by the *MS1* transcription factor whose absence leads to a severe male sterility phenotype. At4g30040 is indirectly regulated by *MYB80,* also a transcription factor whose absence leads to a severe male sterility phenotype. Because of this, our first working hypothesis was that both of these APs could be involved in pollen development. This hypothesis was further supported by the expression patterns of these proteases in reproductive tissues. Somewhat unexpected, their T-DNA knock-out (KO) mutants displayed no obvious pollen/anther phenotype. Since both proteins were also expressed in root tissues, we next sought to examine whether the absence of these genes affected root growth. Indeed, our results revealed significant reductions in primary root length and in lateral root number in both KO lines, showing that both proteins are involved in primary root development and lateral root formation. Moreover, when evaluated under nutrient limitation (N deprivation), mutant phenotypes suggest that these genes may be involved in two different regulatory mechanisms of lateral root formation. Therefore, these genes were designated Regulator of Lateral Root (RLR) 1 and 2 (At2g03200 and At4g30040, respectively). Moreover, both KO mutants showed no altered response to auxin treatment in terms of primary root growth but showed an increased response in lateral root number, connecting these enzymes to auxin signaling pathways. *RLR1* and *RLR2* overexpression mutants also showed de-regulation of lateral root formation and primary root growth further strengthening the importance of these atypical APs in both mechanisms.

To determine their biochemical properties, recombinant forms of both proteins were produced using the innovative plant-based expression platform magnICON® in *Nicotiana benthamiana* leaves, and we have successfully purified recombinant RLR1 (rRLR1). rRLR1 was shown to be glycosylated, active at acidic pHs, not completely inhibited by pepstatin A, and with a distinct specificity pattern determined by Proteomics Identification of Cleavage Sites PICS, suggesting closest resemblance with fungal APs. Redox agents also have a significant inhibitory effect on rRLR1 activity suggesting that this protein might be involved in redox sensing mechanisms. These results clearly demonstrate that RLR1 is an AP with distinct and atypical biochemical properties.

Finally, and in order to identify RLR1 and RLR2 *in vivo* substrates and molecular pathways, two proteomics techniques were used: quantitative shotgun proteomics and N-TAILS. Due to sample constrains, we were not able to obtain valid N-TAILS results on time to include in this document. However, the shotgun results allowed the identification of several de-regulated proteins in both KO lines, under normal or N-deprived growth conditions. Interestingly, our results anticipate similarities between both KO lines, with several proteins related to auxin signaling/biosynthesis and redox homeostasis being de-regulated in both *RLR1* and *RLR2* KOs. This global profiling also revealed de-regulated proteins unique for each KO line, and for each condition, which suggest that these proteases may still participate in different pathways.

Our results unveil a new role for two atypical APs in the regulation and adaptation of root development in *Arabidopsi*s, under normal growth conditions as well as under abiotic stress. To our knowledge, this is the first study linking atypical APs to root development and, hopefully, this will pave the way to a better understanding of the role of these enzymes in this important plant process and the molecular mechanism with which they mediate their biological function.

Resumo

A classe das proteases aspárticas (PAs) é a segunda mais representada em plantas e os membros da família A1 (tipo pepsina) são os mais comuns, encontrando-se em elevado número em várias espécies de plantas (por exemplo, 70 homólogos de PAs podem ser encontrados no genoma de *Arabidopsis thaliana* e 166 no genoma de *Oryza sativa*). Se considerarmos que em *Homo sapiens* só estão presentes 17 PAs, o elevado número de genes identificados em plantas sugere a sua participação e relevância funcional em diferentes processos celulares. A maioria destas PAs putativas de plantas possui características atípicas (PA atípicas) em relação à sua organização de estrutura primária e localização intracelular e, apesar de terem sido muito pouco estudadas até agora, alguns estudos reportam a sua importância em processos celulares críticos para o desenvolvimento da planta, tais como resistência a stresses bióticos e abióticos e morte celular programada durante a gametogénese, entre outros.

O presente trabalho consistiu no estudo de duas PAs atípicas putativas de *A. thaliana,* e teve como objectivos principais o estudo funcional destas duas enzimas, das suas propriedades bioquímicas, e dos mecanismos moleculares através dos quais elas exercem a sua função. O produto do gene At2g03200 foi uma das PAs estudadas, tendo sido mostrado num estudo recente que a sua expressão é regulada pelo factor de transcrição *MS1*. A segunda PA estudada foi o produto do gene At4g30040, cuja expressão é regulada pelo factor de transcrição *MYB80.* A ausência de cada um destes factores de transcrição provoca alterações severas no desenvolvimento do pólen, culminando na esterilidade da planta mutada. Suportados por estas evidências, a nossa hipótese inicial era que estas PAs pudessem estar envolvidas em processos relacionados com o desenvolvimento de pólen. De facto, esta hipótese foi reforçada pelo facto de ambas as proteínas serem expressas nos tecidos reprodutores. No entanto, linhas mutantes onde a expressão de cada PA foi abolida (mutantes KO) não apresentaram qualquer defeito ao nível do pólen/antera. Já a análise fenotípica do desenvolvimento de raiz em cada um dos mutantes KO revelou uma redução significativa no comprimento da raiz primária, bem como no número de raízes laterais em comparação com as plantas controlo. Para além disso, a análise fenotípica dos mutantes em condições de crescimento onde a disponibilidade de azoto foi limitada (stress abiótico) sugere que estas protéases controlam o aparecimento de raízes laterais através de dois mecanismos moleculares distintos. Em resultado do fenótipo observado no número de raízes laterais, as duas proteínas foram denominadas de *Regulator of Lateral Root* (RLR) 1 e 2 (At2g03200 e At4g30040, respectivamente). Os nossos resultados apontam ainda para a possibilidade destas duas PAs atípicas fazerem parte de uma das vias de sinalização da hormona auxina. Mais ainda, os mutantes de sobre-expressão de RLR1 e RLR2 mostraram também desregulação do

desenvolvimento da raiz primária e raízes laterais, reforçando a importância destas PAs para estes processos de raiz.

De modo a estudar as suas propriedades bioquímicas, as protéases RLR1 e RLR2 foram produzidas na sua forma recombinante em folhas de *Nicotiana benthamiana* usando o sistema magnICON®. A forma recombinante da RLR1 foi purificada com sucesso e demonstrámos que esta PA é glicosilada, activa a pHs acídicos, não é totalmente inibida por pepstatina A mas significativamente inibida por agentes redox, o que sugere o seu envolvimento em mecanismos de controlo redox. Mais ainda, o perfil de especificidade desta protéase revelou preferências distintas quando comparado ao de outras PAs, sendo semelhante a PA de fungos. Todos estes resultados demonstram claramente que a RLR1 é uma PA com características bioquímicas atípicas.

Por último, e de modo a identificar os substratos *in vivo* destas PAs e os seus mecanismos moleculares, recorremos a duas técnicas de proteómica: *quantitative shotgun proteomics* e *N-TAILS*. Infelizmente, não nos foi possível obter dados de *N-TAILS* a tempo de serem incluídos neste documento mas os resultados de shotgun permitiram a identificação de várias proteínas desreguladas nos dois mutantes KO crescidos em condições normais ou de stress abiótico. Apesar de revelarem proteínas desreguladas únicas para cada mutante (o que implica mecanismos moleculares distintos para cada PA), estes resultados também revelaram alterações comuns em ambos os KOs. Os resultados desta análise de proteómica antecipam a potencial relevância destas duas protéases em mecanismos de regulação de auxina e/ou de balanço redox.

Em resumo, os nossos resultados revelam um novo papel para duas PAs atípicas de *Arabidopsis* na regulação e adaptação do desenvolvimento de raiz em condições de crescimento normal ou em resposta a privação de azoto. Este é o primeiro estudo que demonstra a participação de PAs neste processo fundamental para o desenvolvimento da planta. Mais ainda, este estudo confirma a diversidade/relevância funcional das PAs atípicas de plantas, reforçando a necessidade de mais estudos integrados que permitam compreender em mais detalhe os processos biológicos regulados por este grupo de protéases.

Abbreviations

aa – Amino acid

ABA – Abscisic Acid

ACBP6 – Acyl-Coenzyme A-Binding Protein 6

ACN – Acetonitrile

Amp – Ampicillin

AOC3 – Allene Oxide Cyclase 3

AP – Aspartic protease

APCB1 – Aspartyl Protease Cleaving BAG

ARF – Auxin Response Factors

ASPG1 – Aspartic Protease in Guard Cell 1

Aux/IAA – Auxin/Indol-3-acetic acid

BAG – Bcl-2 Associated Athanogene

BR – Brassinosteroid

cds – Coding sequence

CK – Citokynin

CLV – CLAVATA genes

CND41 – 41 kDa Chloroplast Nucleoid DNA Binding Protein

CP – Cysteine Protease

DGL1 – Defective Glycosylation 1

DZ – Differentiation Zone

EDS1 – Enhanced Disease Susceptibility 1

EL – Elongation zone

ETA1 – Eternal Tapetum 1

FA – Formic acid

GFP – Green Fluorescent Protein

GO – Gene ontology

HA – Human Influenza Hemagglutinin

HR – Hypersensitive Response

IAA – Indole-3-Acetic Acid

IAM – Indole-3-Acetamide

IAN - Indole-3-Acetonitrile

IAR3 – IAA-Alanine Resistant 3

INA – Isonicotinic Acid

IT – Infection Tread

JA – Jasmonic acid

KD – Knock-down

KO – Knock-out

LAX – Like-Aux 1

LRFC – Lateral Root Founder Cells

LRP – Lateral Root Primordium

MLP – Major Latex-like Protein

MMP – matrix metalloendoprotease

MVP1 – Modified Vacuole Phenotype 1

MZ – Meristematic Zone

o.n. – Overnight

PBZ – Probenazole

PCS1 – Promotion of Cell Survival 1

PCD – Programmed Cell Death

PD – Plasmodesmata

PDGLP1 – Plasmodesmal Germin-Like Protein I

PICS – Proteomics Identification of Cleavage Sites

PIN – Pin Formed Protein

PR – Pathogenesis related genes

PSI – Plant Specific Insert

QC – Quiescent Center

RAM – Root Apical Meristem

ROS – Reactive oxygen species

RT – Room temperature

SA – Salicylic Acid

SAR – Systemic Acquired Resistance

SOD – Superoxide dismutase

SP – Signal peptide

TAILS – Terminal Amine Isotopic Labeling of Substrates

TSP – Total soluble protein

TT4 – Transparent Testa 4

WUS – WUSCHEL genes

Chapter I - Introduction

Aspartic proteases (APs, EC 3.4.23) are a class of endopeptidases that are present in a vast diversity of living organisms such as vertebrates, plants, yeast, nematodes, parasites, fungi, viruses and, more recently, they have also been identified in bacteria [1-4]. According to the MEROPS database, APs are distributed throughout fifteen families based on their amino acid sequence homology, which in turn are grouped into five different clans based on their evolutionary relationship and tertiary structure [5]. The family A1 (pepsin-like), which belongs to AA clan, is the better characterized since the majority of APs studied so far belong to this family.

When compared to their mammalian and viral counterparts, plant APs have been away from the spotlights but that condition has been changing over the last few years. The majority of plant APs characterized so far belong to the family A1 and share the general characteristics of all members of this family: they are active at acidic pHs, are specifically inhibited by pepstatin A, a natural peptide from *Streptomyces*, and they have two catalytic aspartates (organized under the sequences DTG/DSG) that are responsible for their catalytic activity [6]. Nonetheless, they have one structural feature that is unique to plant APs which is the presence of an extra domain of about 100 amino acids called the Plant Specific Insert (PSI). This PSI region is inserted in the C-terminal domain of plant AP precursors but is removed during protein maturation [6]. It is highly similar to saposins and although its function remains unclear, it has been proposed to be important in vacuolar sorting, and it was also demonstrated to induce vesicle leakage, which could make it important as a defensive weapon against pathogens [6-9].

Plant APs have been found in a variety of plants such as *Arabidopsis thaliana* [10-12], barley (*Hordeum vulgare*) [13-15], cowpea (*Vigna unguiculata*) and common bean (*Phaseolus vulgaris*) [16], buckwheat (*Fagopyrum esculentum*) [17, 18], cardoon (*Cynara cardunculus*) [19-21], grape (*Vitis vinifera*) [22], *Nepenthes* [23, 24], potato (*Solanum tuberosum*) and sweet potato (*Ipomoea batatas*) [25-27], rice (*Oryza sativa*) [28, 29], soybean (*Glycine max*) [30], tobacco (*Nicotiana tabacum*) [31], tomato (*Solanum lycopersicum*) [32] and the moss *Physcomitrella patens* [33]. Strikingly, some of these plants such as *Arabidopsis*, rice and grape seem to have a large number of APs encoded in their genomes [22, 34, 35]. Faro and Gal performed a bioinformatics analysis of the genome of *A. thaliana* searching for genes encoding APs and they found 51 sequences that potentially encode for these proteins [34]. Even more striking was the fact that the vast majority of these genes (46) have different characteristics when compared with canonical APs. Supported by these findings, Faro and Gal grouped these sequences into three classes: typical APs which have the usual characteristics of plant APs (including the presence of the PSI); nucellin-like APs which are similar to barley nucellin; and

atypical APs which, as the name implies, display atypical characteristics [34]. The same analysis was performed using the genomes of rice and grape and, as observed for *A. thaliana*, the vast majority of sequences found encodes either atypical or nucellin-like APs [22, 35].

In 2004 Simões *et al.*. provided a comprehensive review on typical plant APs [6]. In the past twelve years much has been done in the plant AP field and now the focus is on atypical and nucellin-like APs since they have been implicated in very important plant developmental processes and in stress tolerance. The present thesis will describe our efforts to characterize two atypical APs from *A. thaliana*; therefore, in this introduction we will focus on this group of APs, their characteristics and their role in plant processes. We will also focus on pollen and root development, two important development aspects where our target proteins may play a crucial role but where the participation of atypical APs is still poorly understood.

1.1 Plant Atypical and Nucellin-like Aspartic Proteases

As discussed before, a bioinformatics analysis of the *Arabidopsis* genome revealed the presence of 51 putative AP genes of which 46 encode APs with distinct characteristics when compared with the so called typical APs [34]. These 46 genes were divided in three groups: atypical APs encoded by 36 genes, nucellin-like APs encoded by 4 genes and 6 genes that could not be attributed to any of the categories [34]. When the rice and grape genomes were analyzed in a similar way the results also revealed that the majority of putative AP genes encode APs with atypical characteristics [22, 35]. Moreover, in the past few years atypical APs were described not only in *Arabidopsis* [36-43] but also in other plants such as tobacco (*Nicotiana tabacum*) [31], rice (*Oryza sativa*) [44-48], common bean (*Phaseolus vulgaris*) [49] and different species of *Nepenthes* [23, 24], playing important roles in all of these organisms. These proteins are clearly different from the well-studied typical APs and in the following sections we will analyze the differences concerning their primary structure organization, cleavage specificity and inhibition profile, as well as their distribution and intracellular localization.

- ***Primary structure organization***

Typical plant APs are synthesized as preproenzymes, with a pre domain which is a N-terminal hydrophobic signal peptide responsible for protein sorting to the ER, followed by a pro-segment of approximately 40 amino acids which is believed to be important for the inactivation of the zymogen [6]. Upon maturation, the signal peptide and the pro-segment are

removed as well as the PSI which is initially located in-between the N- and C-terminal domains of the enzyme (see Figure 1) [6]. Nonetheless, the majority of APs recently identified in different plant genomes have distinct primary structure organizations that confer them unique characteristics. Nucellin, a barley AP specifically expressed in nucellar cells of the ovary, lacks the PSI and it also lacks the pro-segment [50]. The rice orthologue of nucellin, OsAsp1, also lacks the PSI and although the authors considered that OsAsp1 has a pro-segment, it has significant deletions [51]. Faro and Gal detected the presence of 4 genes homologous to nucellin in *A. thaliana* genome. The encoded nucellin-like APs share the same features of nucellin, lacking the PSI and almost the entire pro-segment [34]. The recently described Aspartyl Protease Cleaving BAG (APCB1) is one of these nucellin-like APs encoded in the *Arabidopsis* genome. Like atypical APs, these proteins are also very different from typical APs but they were ascribed in a different group because they share low sequence similarity with atypical APs and display important differences in what concerns the amino acid residues surrounding the active site [34].

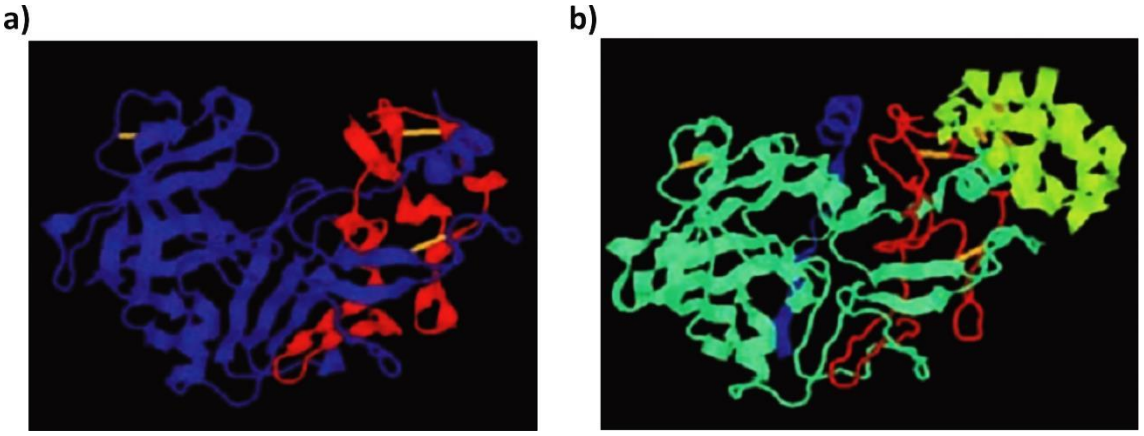


Figure 1 – Crystal structure of cardosin A and prophytepsin. a) Crystal structure of mature cardosin A from *C. cardunculus* L. (PDB code: 1B5F) [52]. In blue is depicted the heavy chain, in red the light chain and in yellow the disulfide bridges. **b)** Crystal structure of prophytepsin from *H. vulgare* L. (PDB code: 1QDM) [15]. In blue is depicted the propeptide, in light blue the heavy chain, in red the light chain, in green the PSI and in yellow the disulfide bridges. Adapted from [6].

The majority of APs identified by Faro and Gal were described as atypical APs because they display remarkable differences in their primary structures that make them very different from typical APs. These differences include the lack of PSI, an unusual high cysteine content and also the nature of the amino acid residues preceding the first catalytic triad [34]. With the exception of *Arabidopsis* Promotion of Cell Survival 1 (PCS1), Aspartic Protease in Guard Cell 1 (ASPG1), NANA [37, 38, 40], and tobacco 41kDa Chloroplast Nucleoid DNA Binding Protein (CND41) [31], all the atypical APs characterized so far have a putative signal peptide followed by a putative pro-segment with variable length when compared with typical pro-segments. In

typical APs this domain is usually removed upon protein maturation and only then the enzyme is fully active [6]. Nonetheless, recombinant CDR1 and its rice homologue were shown to be active without the removal of the pro-segment [53, 54]. In the case of PCS1 it does not appear to have a pro-segment; instead this region is substituted by a serine-rich region whose function is still unclear [37]. ASPG1 and NANA do not have a signal peptide and the latter was shown to be a chloroplastidial protein, although it also appears to lack a canonical transit peptide [38, 40]. CND41 appears to lack both pre- and pro-segments and in turn has a putative transit peptide followed by a lysine-rich region which is believed to be responsible for the ability of CND41 to bind DNA [31]. Interestingly, we have evidences that CND41 has indeed an active signal peptide that is able to target this protein to the secretory pathway when fused to Green Fluorescent Protein (GFP), suggesting that CND41 sorting may be more complex than initially anticipated (our own unpublished results). The absence of a pro-segment is a striking feature but Kato *et al.* proposed that the lysine-rich region can also interact with the active site through the positive charged residues which may keep the enzyme in the inactive form [55, 56]. However, the processed enzyme without this lysine-rich region was not purified in this work, therefore its impact in regulating enzyme activity still needs to be confirmed [56].

Another surprising feature concerning the primary structure of atypical and nucellin-like APs is their unusual high cysteine content. A mature typical AP has 6 cysteine residues usually involved in three intramolecular disulfide bridges which stabilize the overall enzyme structure (see Figure 1) [15, 52]. The PSI has another 6 cysteine residues also forming three disulfide bridges but, as discussed before, this domain is removed during protein maturation [6]. Strikingly, atypical and nucellin-like APs have as much as twelve cysteine residues and in some cases the number is even higher [24, 34, 53]. Interestingly, atypical APs have what has been denominated “the nepenthesin-type AP (NAP)-specific insertion” which is a cysteine-rich region located between the first catalytic triad and the conserved flap tyrosine. This region has seven cysteine residues and its function, as well as its presence in the mature form of the enzymes, remains unclear [24, 34, 53, 57]. Additionally, atypical APs have also a conserved cysteine residue in the extreme C terminus of the protease which is not present in typical APs [24, 34, 53].

There are no three-dimensional structures reported so far for plant atypical or nucellin-like APs. This makes the assignment of which cysteine residues are involved in disulfide-bond formation a very difficult task. Nonetheless, Athauda *et al.* proposed a model for the three-dimensional structure of nepenthesin I comprising six disulfide bridges [24]. It is well known that disulfide bridges confer stability to the overall enzyme structure and, in this case, the high number of disulfide bridges was also assumed to be the underlying reason for

the remarkable stability of this protease in a wide range of temperatures [24]. Very recently, Fejfarová *et al.* were able to produce crystals of recombinant nepenthesin I from *N. gracilis* using a low-pH crystallization screen [58]. In this work, the recombinant protein was produced in *E. coli* in the form of inclusion bodies and after refolding and purification, it was crystalized in the presence of pepstatin A, at a very low pH [58]. According to the authors, all molecular replacement methods using distant homology models proved unsuccessful so far to solve this structure, therefore it is not possible to confirm if the disulfide bond arrangement initially proposed by molecular modeling is correct and how it contributes to protein stability [58]. Cysteine residues can also be involved in intermolecular disulfide bridges which are responsible for different oligomeric states. In fact, SDS-PAGE analysis under reducing and non-reducing conditions showed that CDR1 can be present in different oligomeric states, being the dimerized form the one with higher proteolytic activity [53].

Finally another primary sequence feature that is different in atypical and nucellin-like APs is the sequence around the first Asp-Thr-Gly (DTG) motif. In typical APs the first active site motif is characterized by the presence of a hydrophobic-hydrophobic-DTG-serine-serine sequence [6]. In contrast, nucellin-like APs active site motif comprises an acidic-hydrophobic-DTG-serine-acidic sequence. For atypical APs their first active site motif is a hybrid between the typical and nucellin-like since it shares the hydrophobic-hydrophobic signature of typical APs and the serine-acidic sequence of nucellin-like APs (see Table 1) [34]. It is possible that these striking differences in the first catalytic motif of atypical and nucellin-like APs may interfere with their catalytic mechanism, most likely by changing their cleavage preferences and substrate specificity [34].

Table 1 - Active site differences between the different groups of plant APs. Adapted from [6]

AP Type	Active site motif
Typical	Hydrophobic-Hydrophobic-**Asp**-**Thr**-**Gly**-Ser-Ser
Nucellin-Like	Acidic-Hydrophobic-**Asp**-**Thr**-**Gly**-Ser-Acidic
Atypical	Hydrophobic-Hydrophobic-**Asp**-**Thr**-**Gly**-Ser-Acidic

- ***Cleavage Specificity and Inhibition***

It is well known that the catalytic activity of APs is based on a nucleophilic attack of the substrate by a water molecule that is hydrogen bonded to the two aspartate residues of the active site motifs [1]. Despite some controversy regarding the effective steps of the reaction, it is assumed that the water molecule attacks the carbonyl group of the scissile bond, with subsequent transference of a proton to one of the aspartate residues, resulting in the

formation of a non-covalent tetrahedral intermediate. The protonation of the nitrogen atom of the scissile bond results in protein cleavage [1, 2].

Typical plant APs are normally ascribed as housekeeping genes which means that they cleave a large number of different substrates [6]. As their mammalian and viral counterparts, they preferentially cleave bonds between hydrophobic residues, aliphatic or aromatic, being the Phe-Phe bond the most commonly cleaved [1, 2, 6]. In what concerns atypical and nucellin-like APs not much is known about their specificity and inhibition profiles. Bearing in mind the differences in the sequence of their first catalytic motif one would expect that their specificity requirements could be somewhat different. In fact, as we will shortly describe, there are some atypical APs that display unique specificity requirements while others appear to have a more broad specificity [24, 53].

For the majority of atypical APs described so far the specificity requirements were not studied in detail. Tobacco CND41 was shown to cleave a large number of denatured proteins, thereby suggesting a broad specificity [55]. However, we have to keep in mind that the sequence of the cleaved proteins was not determined. The optimum pH for its catalytic activity *in vitro* is 2.5 which is the normal optimum acidic pH of pepsin, the typical AP with the lower optimum pH. It is important to note that CND41 is a chloroplastidial protein and it is unlikely that under physiological conditions the pH of the chloroplast reaches this value. Interestingly, Murakami *et al.* suggested that CND41 protease activity can be important under certain stress conditions known to decrease the cytosolic pH (and thus also the chloroplastidial pH) [59]. Ge *et al.* described that PCS1 is able to partially hydrolyze casein but it is not active against the other 3 substrates tested (BSA, haemoglobin and ovalbumin) which led the authors to speculate that this enzyme has a narrow specificity [37]. Moreover, the authors found that PCS1 was mostly active at pH 5.5 - 6.5 but still displayed proteolytic activity at a wide range of pHs (2.5 - 7.5) [37]. APSG1 and NANA are *Arabidopsis* atypical APs for which no specificity requirements were analyzed. Although both proteins have different intracellular localizations and functions (both of them will be analyzed in more detail in another section) they both have an optimum pH of 6 (assessed *in vitro*) which is unusually high relative to typical APs [38, 40]. Finally, nodulin 41 is another atypical AP for which no specificity studies were made. This common bean atypical AP was shown to be most active at pH 4.5 although it has residual activity at pH values ranging from 3.5 to 7.5 (activity measured *in vitro* against casein) [49].

Despite this lack of knowledge about the specificity of atypical APs, there were some reports addressing this issue. Athauda *et al.* reported that nepenthesin I displays a broad specificity towards the oxidized insulin β-chain. The preferred bonds contained mostly aliphatic and/or aromatic residues, with leucine as prime candidate for P1 site and interestingly this

enzyme also cleaved the Leu-Cya (cysteic acid), a feature that has never been reported for APs [24]. This broad but somewhat different specificity coupled with the fact that this enzyme is very stable under very stringent conditions, recently prompt Kadek *et al.* to use recombinant nepenthesin I as a digestion tool in Hydrogen/Deuterium exchange mass spectrometry (HXMS) studies [60]. Authors showed that recombinant nepenthesin I can be successfully incorporated in HXMS workflow and that its specificity is different from the commonly used pepsin, which makes this enzyme a complementary tool to use in these experiments [60]. More recently nepenthesin II was also suggested to be a good digestion enzyme for HXMS studies [61].

In contrast to this broad specificity, CDR1 cleaved just one bond of the oxidized insulin β-chain (Leu^{15}-Tyr^{16}) making CDR1 the atypical AP characterized thus far with the narrowest specificity. Simões *et al.* tested several fluorogenic substrates in order to characterize the activity of this enzyme; however, it was necessary to design a new substrate since CDR1 was unable to cleave any of the peptide substrates commonly used to characterize APs [53]. The newly designed substrate was cleaved in a Phe-Val bond which is not that unusual for APs. However, it had a Glu residue in the P4 position which is thought to form a salt bridge with an Arg residue of the proline-rich loop and this is probably the requirement responsible for the narrow specificity of this enzyme [53]. Besides its narrow specificity, its optimum pH was described to be between 6.25 and 6.5, quite similar of that found for ASPG1 and NANA but very different from that of typical APs [53].

Concerning the inhibition profile of atypical and nucellin-like APs very little was reported. Surprisingly, these enzymes have a very broad sensitivity for pepstatin A, which is considered the canonical AP inhibitor and has been extensively used to identify/purify APs [1, 2]. Typical APs are totally inhibited by pepstatin A and while nepenthesin I is also strongly inhibited, other atypical APs show differences in sensitivity to pepstatin A [24]. CND41 was completely insensitive to this inhibitor while CDR1 and nodulin 41 conserved almost half of their activity in the presence pepstatin A and ASPG1 showed 30% of residual activity [38, 49, 53, 59]. This lack of inhibition by pepstatin A is a distinguishing feature of atypical APs and may be one of the reasons why they have remained elusive until very recently, despite the large number of putative atypical AP genes present in different plant genomes [22, 34, 35]. For some atypical APs other classes of inhibitors were also tested. CDR1 was partially inhibited by pefabloc (a serine protease inhibitor) which had not been previously observed for any plant AP. EDTA, ATP and Zn^{2+} also inhibited this enzyme as well as reduced and oxidized glutathione. This inhibition by redox agents is in agreement with the fact that CDR1 is an active homodimer, with the two monomers linked by disulfide bridges [53]. Finally, nodullin-41 was sensitive to 2-mercaptoethanol, Fe^{3+} and SDS but was completely insensitive to EDTA [49].

As we could see in this section, atypical and nucellin-like APs have very distinguishing and somewhat unexpected enzymatic properties which could be linked to their unusual localizations and functions, as will be analyzed in the following sections.

- ***Distribution and localization***

Over the last few years, an increasing number of reports show that atypical plant APs are widely distributed throughout the plant kingdom. As already mentioned, an *in silico* analysis of the *A. thaliana* genome identified a large number of genes coding for these APs [34]. However, until now, only seven *Arabidopsis* atypical APs (and one nucellin-like) were investigated and partially characterized. PCS1 and CDR1 were the first *Arabidopsis* atypical APs to be reported over ten years ago [36, 37] but just since 2011 six more were added to the list: ASP, UNDEAD, ASPG1, NANA, Apoplastic EDS1 Dependent Protein 1 (AED1) and APCB1 (nucellin-like) [38-43]. For rice (whose genome also contains a large number of genes coding for atypical APs [35]) another six atypical APs were described: CDR1 rice orthologue OsCDR1 [47], OsAP25 and OsAP37 whose expression is directly regulated by the EAT1 transcription factor [46], OsAP65 [45], the product of the S5 triallelic system [44] and OsAP77 [48]. Atypical APs were also described for other plant species: CND41 was purified from tobacco (*Nicotiana tobacum*) leaves [31], nepenthesins were isolated from the carnivorous plants *Nepenthes gracilis* and *Nepenthes distillatoria* [23, 24], nodulin 41 was isolated from common bean (*Phaseolus vulgaris* L.) root nodules [49] and a gene coding for an atypical AP was also found in buckwheat (*Fagopyrum esculentum*) seeds [18]. A summary of the known information about atypical and nucellin-like APs described so far is depicted in Table 2.

One of the special features of atypical APs that distinguish them from the typical ones is their cellular localization. Typical APs are normally intracellular proteins mostly found in plant vacuoles [6] while atypical APs can be found in a variety of intracellular compartments as well as outside the cells. *Arabidopsis* PCS1, ASPG1 and *Phaseolus* nodulin 41 were reported to be residing in the endoplasmic reticulum (ER) while UNDEAD is predicted to be a mitochondrial protein [37, 38, 41]. Surprisingly, NANA was found in the chloroplast although it appears to lack a canonical transit peptide [40]. The chloroplast is a very unusual localization for an AP but tobacco CND41 was also found in that organelle, although in this case the protease appears to have a transit peptide [31]. Concerning rice APs, OsAP65 appears to be located in the pre-vacuolar compartment, the product of the S5 triallelic system can be found in the cell wall and

Table 2 – Plant atypical and nucellin-like APs described so far.

Plant species	Protein	Cellular localization	Function	Reference
Arabidopsis thaliana	CDR1	Apoplastic fluid	Disease resistance	[36]
	AED1			[42]
	APCB1	n.d.		[43]
	ASPG1	ER	Resistance to drought stress	[38]
	NANA	Chloroplast	Chloroplast homeostasis	[40]
	PCS1	ER	PCD processes during reproduction	[37]
	UNDEAD	Mitochondria (Predicted)		[41]
	ASP (At1g66180)	Secretory pathway (Predicted)	Ascorbate responsive gene	[39]
Oryza sativa L.	OsCDR1	Apoplastic fluid (Predicted)	Disease resistance	[47]
	OsAP77			[48]
	OsAP25	n.d.	PCD processes during reproduction	[46]
	OsAP37	n.d.		
	OsAP65	Pre-vacuolar compartment		[45]
	S5 product	Cell wall	Reproductive isolation	[44]
Nepenthes spp	Nepenthesin	Extracellular	Protein degradation	[23, 24]
Nicotiana tabacum	CND41	Chloroplast	Chloroplast homeostasis	[31]
Hordeum vulgare L.	Nucellin	n.d.	PCD processes during reproduction	[50]
Phaseolus vulgaris L.	Nodulin 41	ER	n.d.	[49]
Fagopyrum esculentum	FeAPL1	n.d.	n.d.	[18]

n.d. – Not determined

OsCDR1 is located in the apoplast (intercellular fluid) which is the same localization of its *Arabidopsis* orthologue CDR1 and also of AED1 [36, 42, 44, 45, 47]. This extracellular localization was also seen for nepenthesins which are present in the pitcher fluid of the plant [23, 24]. Moreover, Faro an Gal analysis revealed that many of the *Arabidopsis* putative atypical APs are predicted to be localized along the secretory pathway, others may reside in the chloroplast and a third group may be membrane anchored [34]. All these unusual

localizations support the idea that these proteins are indeed very different from their typical counterparts and that they can play very different and important roles throughout the plant.

In addition to the atypical APs described so far, a quick search at the MEROPS peptidase database shows that other plant species such as grape vine (*Vitis vinifera*), maize (*Zea mays*), sorghum (*Sorghum bicolor*) and black cottonwood (*Populus thrichocarpa*), among others, have a high number of putative atypical APs [5]. If we think that higher organisms such as *Homo sapiens* have just a few APs encoded in their genome [5], the high number of APs encoded in plant genomes is an intriguing fact that deserves further attention by the scientific community. We are just beginning to realize the importance of these proteins as they are being linked to tightly regulated and extremely important plant processes; however, much more research is needed to fully understand the functions of these enzymes.

1.2 Biological functions of plant atypical and nucellin-like APs

Our knowledge about atypical APs has been steadily increasing over the years and we are beginning to realize the importance of these proteins in plants. All the facts that were previously discussed concerning their expanded presence in plant genomes and their unusual features strengthen the idea that plant atypical APs may not be housekeeping proteins, but may indeed have specific roles in tightly regulated processes. Unlike their typical counterparts that are normally implicated in protein turnover events (reviewed in [6]), atypical APs seam to play crucial roles in very important plant developmental processes and stress responses. They have been implicated in protein degradation (turnover), protein degradation as a source of nitrogen, chloroplast homeostasis, abiotic and biotic stress responses and also in gametophyte development and reproduction. In the following sections we will discuss what is known about the role of atypical and nucellin-like APs in these processes by analyzing the data that was published concerning the functions of the atypical and nucellin-like APs described so far.

- ***Protein degradation as a nitrogen source***

Nitrogen is the mineral nutrient that plants require the most but unfortunately its availability in terrestrial ecosystems is very limited [62]. In order to capture enough of this nutrient, plants had to adapt and carnivory is one of these adaptations [23]. Nepenthesins were isolated from the pitcher fluid of the carnivorous plant *Nepenthes gracilis* and although they were not the only APs present in that digestive fluid, they were the first atypical APs described in a carnivorous plant [23]. Their main function is protein degradation, which is very

similar to typical plant APs, although protein degradation by nepenthesins is used to provide a nitrogen source and not as a form of protein turnover [24]. Athauda *et al.* isolated two of these proteins from *Nepenthes distillatoria*, nepenthesin I and II, which were very similar in terms of biochemical properties even though they just share 66.6% of similarity in primary structure [24]. They seem to be totally adapted to their function because they are remarkably stable under a wide variety of pH values and temperatures (probably due to their high cysteine content) and are mostly active at a very acidic pH (2.5) [24]. Also they have a broad specificity which seems to be adapted for random protein degradation [24]. The function and regulation of these proteins were not further explored but the authors suggest that the stability of these enzymes allows them to maintain activity for long periods of time within the very acidic pitcher fluid [24].

- ***Chloroplast homeostasis and protein turnover***

Chloroplast is a unique organelle that is present in plant cells and plays a crucial role in primary metabolism since it is where photosynthesis takes place. It is established that this organelle derived from an endosymbiotic event that occurred approximately 3 billion years ago [63]. The acquisition of chloroplasts by eukaryotes occurred later in evolution than the acquisition of mitochondria, when one of these heterotrophic eukaryotes (already with mitochondria) engulfed a cyanobacterium-like endosymbiont to acquire photosynthesis and become autotrophic [63]. Chloroplasts have their own genome and because they have a bacterial origin, their genome is organized into DNA-protein complexes (nucleoids) very similar to bacterial genomes [63]. Because of the great importance of chloroplasts, regulation of gene expression in this organelle has been the focus of many studies throughout the years and surprisingly in one of these studies one atypical AP was described. Chloroplast Nucleoid DNA binding protein with 41 kDa (CND41) was first described by Nakano *et al.* from purified nucleoid fractions of *N. tabacum* cultured cells [31]. Since it was purified from a nucleoid fraction, the authors postulated and then confirmed that it has DNA binding activity, which is conferred by a lysine rich region present in the N-terminal of this protein [31]. The authors also showed that reduced expression of CND41 led to an increased expression of chloroplast genes (such as the *psbA* and *rbcL* gene) which suggest that CND41 may not be a merely structural protein but it can negatively regulate chloroplast gene expression [31]. In order to further explore this function, Nakano *et al.* tried to downregulate *CND41* expression by antisense suppression [64]. The transgenic plants had a dwarf phenotype with dark green and shortened leaves, consequence of an accelerated plastid development. Gibberellin-deficient mutants had a similar phenotype and since this hormone is essential for plant development, authors

assessed gibberellin levels in *CND41* knockdown plants. The levels of this hormone were indeed lower in the knockdown mutants and the addition of gibberellin partly rescued the phenotype [64]. Nonetheless, the mechanisms by which CND41 controls chloroplast development and GA biosynthesis have not been fully characterized [64].

Murakami *et al.* proved that CND41 is indeed an active atypical AP and this led to the hypothesis that CND41 may have other functions besides controlling chloroplast gene expression [59]. Since it was shown that *CND41* knockdown mutants had retarded leave senescence Kato *et al.* investigated the role of CND41 protease activity during this process [55, 64]. They reported that wild type (WT) plants had higher *CND41* transcript levels in mature leaves prior to senescence which indicates that CND41 may be a senescence-associated protein [55]. Rubisco would be one of the first targets of *in vivo* protein degradation upon senescence and indeed the authors showed that CND41 is involved in the controlled degradation of this protein upon leave senescence [55]. Nonetheless, it is noteworthy that rubisco was degraded by CND41 only in its denatured form and not in its fully active form. Under stress conditions like senescence, rubisco loses its proper folding and then becomes vulnerable to CND41 degradation and this fact ensures no rubisco degradation in healthy leaves [55]. It was also shown that post-translational processing of CND41 is implicated in the regulation of its protease activity [56]. Truncated forms of CND41 were observed although N-terminal sequencing was not performed due to limitations in purifying these forms [56]. Even so, and since DNA binding was reported to inhibit CND41 protease activity [59], the authors claim that the lysine-rich region of CND41 may be responsible for protease inactivation and that senescence conditions may trigger CND41 protease activation through N-terminal processing [56].

Another atypical AP that seems to be implicated in chloroplast homeostasis is *Arabidopsis thaliana* NANA [40]. The *NANA* mutant was isolated while the authors were searching for mutants impaired in carbohydrate metabolism. In this case, the mutation in *nana* is caused by a T-DNA insertion in the promoter region of the gene coding an atypical AP (later called NANA) that alters its pattern of expression. Like tobacco *CND41* knockdown mutants, *Arabidopsis NANA* T-DNA mutants had a dwarf phenotype and its bolting was also severely delayed [40]. Surprisingly, even without a canonical transit peptide, NANA was shown to be a chloroplastidial protein and it was implicated in chloroplast homeostasis [40]. Unlike CND41, NANA does not seem to degrade rubisco but for reasons that the authors could not fully explain, the chlorophyll content in *NANA* mutants was 50% lower than in the WT and the electron transport rate was also decreased [40]. Starch metabolism was also impaired, in particular the soluble sugars content that was lower in *NANA* mutants under low light growing

conditions and did not show the daily fluctuations observed in the WT plants. Also, the expression of some genes implicated in sugar metabolism was impaired in these mutants. Taken together these results showed that altered expression of the NANA protein lead to the down-regulation of different chloroplast-related or chloroplast-located genes which perturb chloroplast homeostasis and starch metabolism [40]. In this study, the function of NANA and the mechanisms behind its function were not completely unveiled, so more work is needed to answer these questions.

- ***Defense against biotic injury***

Defense mechanisms are very important to every living organism but they assume a special role in plants because they cannot escape when they are being attacked nor can they move to another place with more suitable growth conditions. In terms of agricultural production, plant diseases have a major economic impact and because of that plant defense mechanisms have been extensively studied throughout the years [65]. Plants lack a canonical immune system but they can recognize avirulence genes of some pathogens, triggering a defensive response that often culminates in the collapse of the infected plant cells and is known as the hypersensitive response (HR) [65]. The HR response limits the spread of the pathogen but do not confer immunity to systemic uninfected parts of the plant that remain vulnerable to the pathogen attack. This is conferred by a different immune response called systemic acquired resistance (SAR) that is induced in response to local infections. SAR triggers salicylic acid (SA) signaling and is effective in protecting systemic tissues against an attack by a broad range of pathogens which are normally sensitive to SA-mediated defense mechanisms [66]. Because SAR is induced in healthy tissues far away from the infection, this response is dependent on the generation of mobile signal(s) (elicitor), whose nature has generated much controversy over the years [66].

In an attempt to understand the complex signaling networks behind disease resistance in *Arabidopsis*, Xia *et al.* were able to isolate an atypical AP whose overexpression conferred enhanced disease resistance [36]. Constitutive Disease Resistance 1 (CDR1) was identified while studying a gain-of-function dominant mutation causing a phenotype of enhanced resistance to the bacterial pathogen *Pseudomonas syringae* and also dwarfism [36]. This protein is an extracellular atypical AP whose activation leads to systemic induction of SA-dependent defense responses, likely through the action of an endogenous mobile signal whose nature still remains unknown. Indeed, *CDR1* overexpressing mutants exhibited enhanced expression of defense genes such as pathogenesis-related genes 1 and 2 (*PR1* and *PR2)* while its knockdown led to an increase of disease susceptibility and impairment of local and systemic

induction of defense mechanisms [36]. Since protease activity is required for CDR1 function, the authors speculate that this protein might process a cell surface protein generating a peptide elicitor that would function as a SAR signal, mediating local and systemic defense responses [36]. Although this peptide elicitor remains unknown, new insights regarding CDR1 specificity have proved that this enzyme has a very narrow substrate specificity requiring a glutamate residue in the S4 pocket, which might be related with the structure of its natural substrate [53]. The mechanism behind CDR1 function is yet to be determined but the fact that it is most active in the dimerized state, which is dependent on the redox conditions, is another important evidence since it is accepted that pathogen attack or stress responses involve an oxidative burst [53].

CDR1 is not the only *Arabidopsis* atypical AP that has a major role in plant defense against pathogens. In an attempt to identify new SAR regulatory proteins Breitenbach *et al.* used the SAR-specific phenotype of *Enhanced Disease Susceptibility 1* (EDS1) mutant. The authors were able to identify 12 apoplastic *EDS1*-Dependent (AED) proteins, among them the atypical AP AED1 [42]. As we have discussed, SAR is induced in response to local infections that trigger SA signaling and a central regulator of this SA signaling is *eds1* which encodes a protein that is essential for SAR activation, among other defense-related functions [67]. Indeed, *EDS1* mutant has impaired SAR signaling while local resistance remains intact which makes this mutant an ideal candidate to identify proteins involved in SAR signaling [42]. This was the case of the atypical AP AED1 whose expression is induced both locally and systemically by infection with *P. syringae* in WT plants. In the *EDS1* mutants, *AED1* is also expressed locally but its systemic expression is diminished which means that systemic but not local *aed1* transcript accumulation is dependent upon *EDS1* and thus likely related to SAR signaling [42]. AED1 is an apoplastic protein and its knockout mutants did not have any disease-related phenotype. However, in contrast to CDR1 overexpression, AED1 conditional overexpression leads to an impaired SAR signaling although local response was not affected. This effect in the SAR response was not overcome by SA treatment (as happened for *eds1* mutants) which indicates that AED1 acts downstream of SA to suppress systemic immunity. Because conditional overexpression of AED1 repressed both SAR and SA-induced systemic resistance without affecting the growth of *P. syringae* in healthy plants, the authors propose that AED1 might be part of a homeostatic mechanism to limit SAR signaling and thus regulate resource allocation in the trade-off between defense and plant growth [42]. Also, the authors speculate that because AED1 is an atypical AP (like CDR1) it might suppress SAR by degrading one or more proteins in the apoplast of SAR-induced leaves [42]. Since nothing is known about the proteolytic activity of AED1 and since the authors did not use the active site mutant to prove

that AED1 protease activity is essential for its biological function, further studies are needed to link AED1 enzymatic activity to its biological function.

Arabidopsis Aspartyl Protease Cleaving BAG (APCB1) in another atypical AP involved in defense mechanisms, recently implicated in the resistance to the necrotrophic fungal pathogen B*otrytis cinerea* [43]**.** The authors were interested in *Arabidopsis* genes that could be controlling programmed cell death (PCD) events and in a computational screening for plant genes that structurally resembled PCD genes they found 7 genes that could belong to the Bcl-2 Associated Athanogene (BAG) protein family [68]. At this point, nothing was known about *Arabidopsis* BAG proteins but in other organisms they were found to be multifunctional proteins with proposed roles in a variety of cellular mechanisms, including apoptosis [69]. While studying these 7 genes, authors found that *BAG6* was essential for basal immunity against the fungi *Botrytis cinerea* since the T-DNA KO line of *BAG6* had an enhanced susceptibility to *B. cinerea* [68]. In a subsequent study, the molecular mechanisms underlying BAG6 function in immunity were dissected and the authors found that BAG6 needed to be processed to exert its function [43]. Using a combination of different protease inhibitors *in vivo* they found that BAG6 processing was inhibited by pepstatin A, linking BAG6 processing to an AP activity [43]. After a series of bioinformatics analysis, yeast-to-hybrid, and pull-down assays they found that APCB1 was the AP responsible for BAG6 processing and that its activity was needed to this processing (the active site mutant was not able to process BAG6) [43]. The authors found that BAG6 was cleaved by APCB1, thereby triggering autophagy and inhibiting *B. cinerea* development. However, the mechanism underlying this process is still unclear [43]. According to the bioinformatics analysis made by Faro and Gal [34], APCB1 (gene number At1g49050) is a nucellin-like AP. Despite the importance of this protease in BAG6 processing, no additional enzymatic characterization has been reported. Therefore, nothing is currently known about APCB1 biochemical properties or intracellular localization.

Arabidopsis is an important model organism that allows a relatively easy identification of important genes which then can be used to search for orthologues in plant species more important for agricultural applications. In the case of CDR1, Prasad *et al.* described a rice orthologue that was named OsCDR1 and which is also involved in disease resistance [47]. This protein shares 50% of similarity and 36% of identity with CDR1 at the amino acid level and is also located in the apoplast. As for CDR1, ectopic overexpression of OsCDR1 both in *Arabidopsis* and in rice resulted in the increased expression of defense-related genes *PR1* and *PR2,* both part of the SA pathway, culminating in an enhanced resistance against fungal and bacterial pathogens [47]. Protease activity is essential for OsCDR1 function and as described for CDR1, infiltration of intercellular fluids of transgenic OsCDR1 overexpressing *Arabidopsis*

plants in WT leaves lead to the activation of systemic defense response in these WT plants [47]. Taking all this into account, the authors concluded that OsCDR1 function is linked to the production of an elicitor molecule that is responsible for the activation of defense-related gene expression in healthy tissues far away from the infection [47]. Nothing is known about its substrate specificity so we cannot directly compare it to CDR1 [54]. It is important to note that transgenic rice lines overexpressing OsCDR1 did not show any deleterious effect in what concerns critical agronomic parameters. Because of that, the authors claim that these findings may have a profound impact on controlling diseases in economically important crops [47].

In another study of disease resistance in rice, Alam *et al.*. reported that *OsAP77* gene expression was promoted upon fungal, bacterial and viral as well as upon treatment of rice with several molecules linked to abiotic stresses such as SA, hydrogen peroxide (H_2O_2), isonicotinic acid (INA) and abscisic acid (ABA) [48]. The reason underlying the authors' interest in *OsAP77* was linked to the fact that it was previously found to be a probenazole (PBZ)-inducible gene; since PBZ in a well-known chemical inducer of plant disease resistance, it was likely that OsAP77 would also be involved in disease resistance [70]. In this following study the authors generated a GUS transgenic line using the *OsAP77* promoter and found this gene specifically expressed in rice vascular tissue, with its expression being promoted upon infection with different pathogens or upon treatment with abiotic stress related molecules [48]. In a previous bioinformatics analysis OsAP77 was ascribed as an AP [35]. Given these results, the authors concluded that OsAP77 is involved in defense mechanisms and speculated that this AP might be sent to the extracellular milieu to function as a protease for processing possible target proteins or peptides into elicitor peptides, as suggested for OsCDR1 or *Arabidopsis* CDR1 [48]**.** However, it is important to note that although being ascribed as a putative AP, a detailed analysis of its primary sequence clearly suggests that OsAP77 lacks both aspartate catalytic motifs, which likely impairs its function as an active protease. Also, the authors did not assess OsAP77 intracellular localization nor characterized OsAP77 activity. Therefore, there is no indication that OsAP77 is in fact an active AP or that it might be sent to the extracellular media to act as an AP.

- ***Defense against abiotic stresses***

Resistance to abiotic stresses like drought is also a very important aspect in terms of agricultural economics. Understanding the mechanisms underlying drought resistance will lead to the development of improved crops, more adapted to unfavorable environments, and this will have a major impact in agriculture. Because of that, a lot of effort has been devoted to understand these mechanisms. Thus far, scientists described two pathways involved in

drought stress adaptation in *Arabidopsis*: one pathway which is dependent on the elevation of ABA for the activation of downstream effectors and a second pathway that is ABA-independent [71]. Although much is already known about the adaptation to drought stress, the molecular mechanisms regulating these two pathways are far from being completely understood. In an effort to understand them better, Yao *et al.* described an *Arabidopsis* gene - *Aspartic Protease in Guard Cell 1* (*ASPG1*) - that may function in drought resistance through ABA signaling in guard cells [38]. *ASPG1* encodes an atypical AP which resides in the ER and is preferentially expressed in guard cells. Although the authors did not see any phenotype related with stomatal closure in *ASPG1* knockout mutants, its overexpressing mutants (*ASPG1*-OE) were able to recover from drought stress much easier than the WT plants [38]. Also, the authors found that ASPG1 overexpression led to a significant increase in ABA sensitivity by guard cells and that ASPG1 expression is ABA inducible. Taking all this into account, the authors speculated that increased expression of ASPG1 could confer drought resistance in *Arabidopsis* by increasing the ABA sensitivity in guard cells which would then be accompanied by a reduction in water loss by transpiration [38]. It is well known that ABA induces elevation in cytosolic calcium levels in guard cells which leads to the production of reactive oxygen species (ROS) [72]. The elevation of cytosolic ROS is responsible for stomatal closure but its excessive accumulation during drought stress can be very toxic and ultimately lead to cell death [72]. Hence, antioxidant activity is very important to prevent oxidative damage and cell death. Indeed, *ASPG1*-OE plants were capable of scavenging excessive ROS very efficiently through the activation of superoxide dismutase (SOD) and this enabled them to be more resistant to drought stress. In summary, under drought conditions ABA induces stomatal closure to avoid water loss, while inducing ASPG1 expression will activate antioxidant defenses such as SOD, preventing excessive oxidative damage and allowing plant to recover better afterwards [38]. Again, the molecular mechanisms behind ASPG1 function still remain elusive and more studies need to be done to address this issue.

- ***Gametophyte development and reproduction***

Plant reproduction is an essential process that relies on correct gametophyte development in order to occur. Despite its extreme importance, much work still needs to be done in order to fully comprehend the molecular mechanisms regulating this phenomenon.

Atypical APs were shown to play very important roles during gametophyte development and reproduction. One of these important processes is pollen germination and pollen tube growth. After pollination, pollen germinates to form the pollen tube that transports the male gametes to the female gametophytes where fertilization takes place [73].

This germination process is crucial for plant reproduction and the impairment of this process leads to plant sterility. In an attempt to study the role of rice APs, Huang *et al.* identified and characterized OsAP65, an atypical AP that plays an important role in pollen germination and pollen tube growth [45]. *OsAP65* homozygous knockout plants could not be obtained and the inability to obtain these plants was due to a severe defect in the male gametophyte. In the heterozygous plants, the total number of pollen grains and the percentage of the mature ones were equal to the WT, and the mutation did not affect sperm cell development and pollen morphology. Taking all this into account, the authors hypothesized that the problem was in pollen germination and pollen tube elongation and indeed that was the case. The authors concluded that OsAP65 was essential for pollen germination and pollen tube growth although the molecular mechanisms underlying OsAP65 function remain to be determined [45].

Reproductive isolation is crucial for maintaining species identity and rice atypical APs also seem to be important in this phenomenon. The Asian cultivated rice (*Oryza sativa* L.) is divided into two major subspecies, *indica* and *japonica* and hybrids between these two subspecies show low fertility [74]. Chen *et al.* showed that the *S5* locus was responsible for this hybrid sterility phenomenon and that there are three possible alleles at the *S5* locus: *indica* allele (*S5-i*), *japonica* allele (*S5-j*) and a neutral one (*S5-n*). Plants bearing the *S5-n* and either one of the others would always be fertile while hybrid plants *S5-j/S5-i* would be infertile and this sterility was due to female gamete abortion [44]. The *S5* locus encodes an atypical AP and while S5-j and S5-i just differ in two amino acids, the S5-n displays a large deletion at the N terminus which probably causes mislocalization of the protein, making it non-functional. This protease is expressed in the ovule and although responsible for this hybrid sterile phenomenon, it appears not to be essential for growth, development or reproduction [44]. In a subsequent study, Ji *et al.* described that S5 AP normally acts as a homodimer and the reason for the low fertility of the hybrid plants S5-j/S5-i resides in the fact that a heterodimer is formed in this case, resulting in a non-functional protein [75]. Although not much is known about the downstream signaling responsible for this hybrid incompatibility phenomenon, these studies clearly show the importance of atypical APs in this process. A deeper understanding of these mechanisms will have major impact in rice genetic improvement.

As we will see later on this chapter, PCD processes during plant reproduction are very important for many plant species. In barley, one of the crucial PCD events that occur during reproduction is the death of the nucellar cells (the cells that surround the embryo sac) which start to degenerate after fertilization [50]. Chen *et al.* described nucellin, an AP with unusual features that could be involved in nucellar cell degeneration [50]. Nucellin was found to be specifically expressed in the nucellar cells surrounding the ovary mainly after pollination, which

correlates with the timing when these cells begin to degenerate. Because of this, the authors speculate that nucellin may function as an apoptotic protease triggering nucellar cell death or, alternatively, it may function as a major protein turnover protease converting nucellar cell proteins into nutrients for embryo and endosperm development [50]. So far, its real function remains elusive.

Finally, atypical APs were shown to play very important roles during anther and pollen development mainly in controlling the PCD events that occur during those processes. Because of that, in the following sections we will focus on anther and pollen development as well as on the roles of several atypical APs implicated in these processes.

1.3 Plant reproduction and male gametophyte development

For all living organisms reproduction is a very important process since it ensures the continuation of life and the transmission of important traits to the new generations, thereby allowing evolution. Plants are no exception and despite its importance for their maintenance and evolution, much remains to be known about plant reproduction. Due to its sessile nature, plants have to take additional measures to protect the reproductive organs and to ensure successful reproduction. In fact, in most of angiosperms (like *Arabidopsis*) the reproductive tissues are very small and enclosed in the flower to be protected from the environment, and this has been linked to the evolutionary success of these plants [76]. Another feature that conferred advantage to angiosperms throughout evolution is their stringent selection of the best pollen and, for that, the correct development of the male gametophyte is of crucial importance [76]. In this work we will focus on the development of the *Arabidopsis* male gametophyte, describing anther and pollen development and highlighting key processes that are crucial for these two events, and where atypical APs have been reported.

- ***Anther development***

Anther development is crucial for normal plant reproduction because it is within this organ that the male gametophytes are formed and matured. Anthers are located in the stamen and their development starts with the appearance of the stamen primordia [77]. These stamen primordia have 3 cell layers, L1-L3; in the very first stages of anther development, L1 differentiates into the epidermis while L2 differentiates into the sporogenous cells and into the three inner somatic layers: endothecium, the middle layer, and the tapetum. L3 differentiates into the connective tissue and the vascular bundle [78, 79]. Figure 2 is a schematic representation of tobacco anther development (*A. thaliana* follows the same steps). This

process is divided in two phases, phase 1 and 2 [77, 80]. As mentioned, it begins with the appearance of the stamen primordia in the floral meristem and throughout phase 1 cell specification and tissue differentiation occur, culminating with meiosis of the microspore mother cells (microsporocytes) which gives rise to the 4 haploid microspores bound together by a callose wall (tetrad) [80]. During phase 2, pollen grains differentiate and mature while the anther enlarges and starts to degenerate, culminating in anther dehiscence and pollen grain release [80].

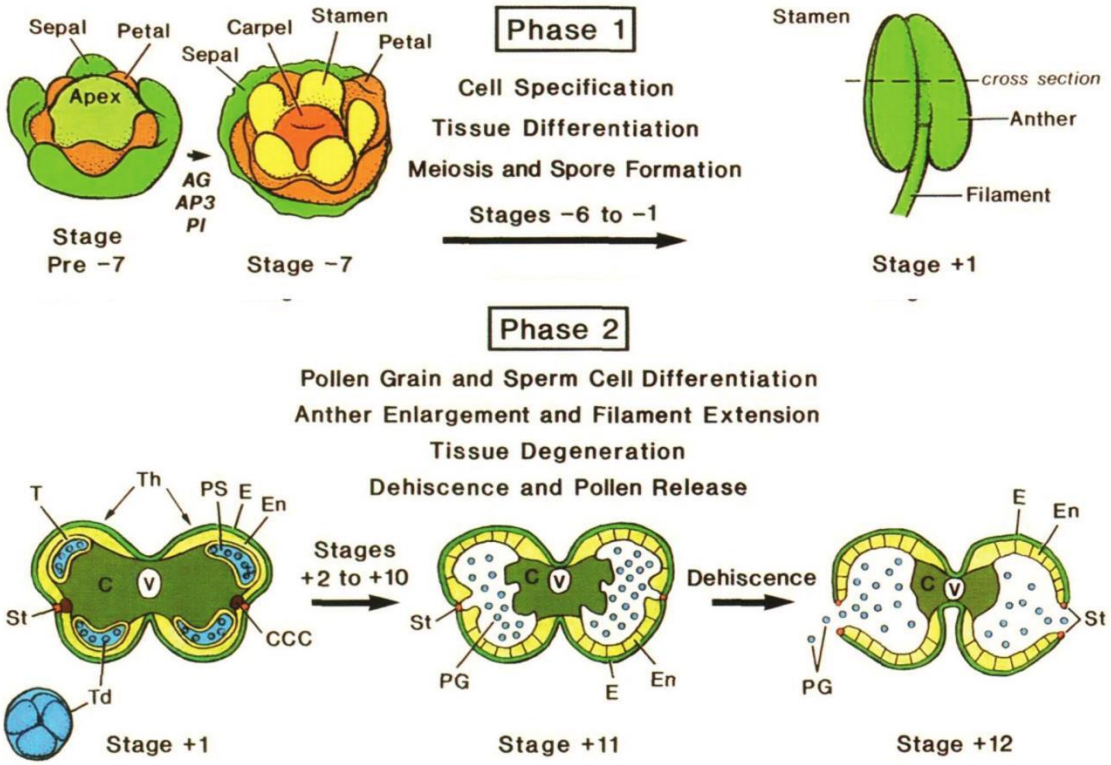


Figure 2 – Schematic representation of tobacco anther development. Anther development is divided in two phases and it starts with phase 1 when the stamen primordia appear. Throughout phase 1 cell and tissue specification occurs, giving rise to the anther; this phase culminates with meiosis which creates the tetrads (pollen primordia). In phase 2 occurs anther enlargement and filament extension as well as pollen grain and sperm cell maturation. This process culminates with anther dehiscence and releasing of the mature pollen grains. C – Connective; CCC – Circular cell cluster; E – Epidermis; En – Endothecium; PG – Pollen grain; PS – Pollen sac; St – Stomium; T – Tapetum; Td – Tetrads; Th – Theca; V – Vascular bundle. Taken from [77].

The described process is very tightly regulated and the consequences for even the smallest de-regulation can be catastrophic in terms of male fertility. This is due to the fact that anther development is tightly linked to pollen development, as the anther is not just the physical place were pollen development occurs but it also controls pollen development and maturation [80]. One of the good examples of this interaction is the role of anther tapetum cells during pollen maturation [81]. Tapetum cells form a single layer of cells that surrounds the anther locule - where pollen is formed and matured - and throughout pollen development

they have very important roles. In the first stages of pollen development the tapetum cells play an important nutritional role supporting meiosis, the formation of the haploid gametes and their nutritional demands [81]. Later on, these cells are responsible for secreting the enzymes that degrade the callose wall that surrounds the tetrade, allowing the release of the young haploid microspores and the subsequent pollen maturation. During this maturation process they also secrete important components of the pollen cell wall and at the very end of pollen maturation the tapetum cells undergo PCD. Tapetum cell death contributes to the sculpting of pollen cell wall, providing adhesive and signaling molecules [82, 83]. As we will explain shortly, the precise timing of this PCD event has a crucial role in normal pollen development which reinforces the idea that anther and pollen development are two tightly linked processes that are very important in plant reproduction.

- ***Pollen development***

Pollen development begins with the two meiotic events that occur in each microsporocyte and give origin to the tetrad, four haploid microspores tightly bound by a callose wall. In Figure 3 is depicted a schematic representation of the process of pollen development, starting with the degradation of the callose wall that binds together the 4 haploid microspores. This is accomplished by a mixture of enzymes secreted by the tapetum cells, resulting in the release of the microspores [76, 83]. After its release from the tetrad, the microspore begins the maturation process by becoming polarized due to the formation of a large vacuole, right before the first pollen mitosis takes place (PM1). Due to this polarization, PM1 gives rise to a large vegetative cell and to a much smaller generative cell, which eventually migrates into the vegetative cell [84]. This first mitosis is an important checkpoint in pollen development because it ensures cell commitment to the male gametophyte development since the reversal from gametophytic to the sporophytic development can occur before PM1 but never after [76]. Pollen maturation finishes when the anther tapetum cells undergo PCD releasing exine and other molecules that will complete pollen cell wall. In the final steps of pollen maturation, or even after maturation, a second mitotic event can take place (PMII) where the generative cell divides giving origin to the two sperm cells. This PMII occurs in some species (*A. thaliana* is one of those species) and from an evolutionary point of view this is considered to be an advanced trait [76].

As we already discussed, anther and pollen development are two processes that are tightly linked since successful pollen development depends not only on the microspore itself but also on the sporophytic tissue that releases important signals and materials [85]. At the end of pollen development, and in order to complete the maturation process and release the

fully maturated pollen grains, two PCD events must occur within the sporophytic tissue [82]. The first event is the death of tapetum cells and it occurs late in pollen development [83].

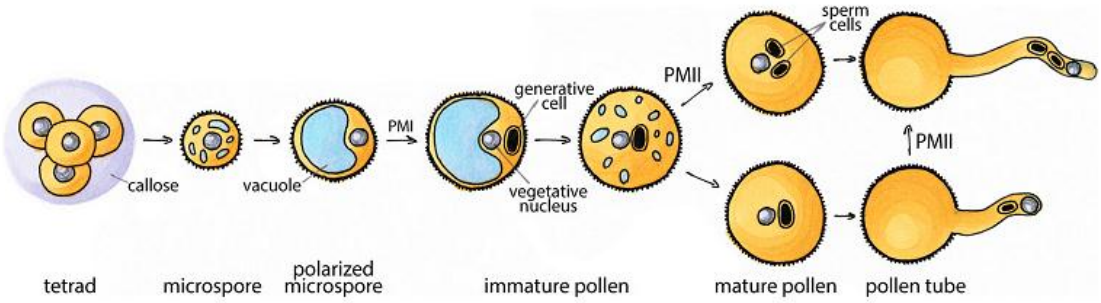


Figure 3 – Schematic representation of pollen development. In the beginning, the callose wall that binds together the 4 haploid microspores is degraded and the pollen maturation starts with the polarization of the microspore. After this polarization occurs, the first mitotic event occurs (PMI) and this gives rise to two cells: a large vegetative cell and a small generative cell which migrates into the first one. At the end of this pollen maturation process and in some plant species another mitotic even can occur (PMII) where the generative nucleus divide giving origin to two sperm cells. Taken from [76].

During most stages of pollen development, the tapetum cells remain highly active and play a crucial supporting role; however, this PCD event is needed since it releases components that will complete the extracellular sculpting of the pollen cell wall [76, 83]. Pollen development then culminates with the death of the stomium cells, resulting in the opening of the anther and the release of the fully mature pollen grains. Atypical APs were shown to have a critical role during these two developmental processes acting mainly in controlling the PCD events [37, 41]. This proves the important role of atypical APs during pollen development and because of that we will discuss what is known about plant PCD during the male gametophyte development and the role of atypical APs in this process.

- ***Role of atypical APs in PCD processes during male gametophyte development***

PCD is the controlled and organized destruction of cells and although crucial for normal gametophyte development, it is still a poorly understood plant process, with the players and pathways involved remaining largely unknown [86]. It occurs early in sexual development during the formation of male and female reproductive organs as well as in later stages of plant reproduction. Failure to properly enter the various cell death programs compromises reproductive success and can even lead to plant sterility [82, 87]. Research on this field is expected to have major impacts on plant biotechnology since it is crucial to develop efficient pollination control systems for hybrid breeding, in order to avoid self-pollination and sib-

pollination. This can be achieved by controlling PCD mechanisms during plant development and reproduction [88].

Although some of the morphological features like cytoplasm shrinkage and DNA degradation are common in both plants and animals, the pathways activated during plant PCD seem to be very different. The absence of caspase orthologs in *A. thaliana* genome is one of the most striking differences between plant PCD and animal apoptosis. Caspases are cysteine proteases that play essential roles in animal apoptosis and the activation of these enzymes is one of the hallmarks of regulated cell death in animals [89]. Since there are no caspase orthologs in *Arabidopsis*, other proteins and pathways must be involved in plant PCD and plant atypical APs were already shown to play important regulatory roles in this process [37, 41].

As we already discussed, during male gametophyte development there are two crucial events of PCD that need to take place. Failure in entering in either one of them, or even a de-regulation in the timing when they occur leads to severe male sterile phenotypes [82, 83]. Recent studies point towards the participation of atypical APs as regulators of PCD during pollen development and Promotion of Cell Survival 1 (PCS1) is one of the atypical APs involved in this process [37]. PCS1 is an *Arabidopsis* atypical AP that was shown to play an important role during gametogenesis and embryogenesis and is expressed specifically in developing gametophytes and in developing seeds [37]. According to Ge *et al.*, *PCS1* loss-of-function mutation causes male and female gametophytes degeneration as well as excessive cell death of developing embryos [37]. In contrast, *PCS1* ectopic overexpression causes a phenotype of partial or total male sterility not because of pollen degeneration but because PCS1 blocks anther dehiscence. This overexpression did not affect the female gametophyte and pollen grains were normal and viable because manual pollination was possible using the pollen grains of the overexpression mutants. Thus, the authors concluded that *PCS1* overexpression phenotype is due to the blockage of PCD processes that normally lead to the degeneration of the stomium and septum which is essential for pollen release [37]. PCS1 protease activity is essential for its function and thus, the authors proposed that PCS1 may function as an anti-cell death component by processing a peptide that function as a survival factor or by inactivating a pro-PCD component [37].

As illustrated by *PCS1* overexpression phenotype, PCD processes are very important for normal anther and pollen development and ultimately for male fertility. Transcription factors are very important in regulating many cellular processes and PCD events during anther and pollen development appear to be no exception. *Arabidopsis MYB80* is expressed in the tapetum cells and it encodes a MYB transcription factor that is essential for tapetal and pollen development [41]. Functional disruption of this gene leads to a premature tapetal

degeneration culminating in abnormal pollen grains and ultimately in male sterility [90, 91]. In an attempt to determine the genes that are directly regulated by this transcription factor, Phan *et al.* described an atypical AP, UNDEAD, that was crucial for proper tapetum PCD [41]. *UNDEAD* expression is directly regulated by *MYB80* and its knockdown leads to a premature tapetal and microspore cell death, a phenotype much similar to that of *MYB80* knockdown, indicating that this protein is one of the key regulators of tapetum PCD. Despite its importance, not much more is known about UNDEAD biochemical properties or native substrates. However, given the phenotype of the knockdown mutants, the authors concluded that MYB80 is delaying PCD by activating UNDEAD expression [41]. In this same study, a second gene encoding for a putative atypical AP (At4g30040) was also described. At4g30040 gene expression was also induced by MYB80 at almost at the same level as UNDEAD expression but since it was not directly regulated by this transcription factor, no further studies were made concerning this gene [41]. Although in this study the authors were only focused in genes whose expression was directly regulated by *MYB80* such as the *UNDEAD* gene, the presence of At4g30040 suggests that other atypical APs may be involved in these processes.

Another *Arabidopsis* transcription factor that is essential for normal tapetal development and appears to be regulating the expression of atypical APs is *MALE STERILITY 1* (*MS1*) [85]. Its knockout impairs the production of viable pollen grains which culminates in a male sterility phenotype [92]. MS1 regulates the expression of genes essential for pollen exine formation, pollen cytosolic content and tapetum development, among them At2g03200 which encodes a putative atypical AP. Although it appears that MS1 positively regulates the expression of this atypical AP gene (At2g03200 is down-regulated in the *MS1* knockout mutants), nothing more is known about this gene [85]. However, the de-regulation of this gene associated with this male sterility phenotype further reinforces the idea that other atypical APs may play crucial roles during anther and pollen development.

Tapetum development and destruction during pollen development is also very important in other plant species. In rice, transcription factors also play an important role in controlling genes essential for normal tapetum PCD processes and one of these transcription factors is *Eternal Tapetum 1* (EAT1) [46]. Niu *et al.* showed that *EAT1* knockout leads to delayed tapetal cell death which culminated in aborted pollen formation and male sterility [46]. The authors also demonstrated that EAT1 directly regulates the expression two rice atypical APs, OsAP25 and OsAP37, which are expressed in tapetum cells and promote PCD in these cells [46]. The authors showed that their overexpression in tobacco leaves lead to an increase in cell death that was strongly reversed by pepstatin A, indicating that their protease activity is essential for their function. Also these two proteins have the ability to promote cell

death in yeast cells in a way very similar to the yeast metacaspase YCA1. Since yeast YCA1 is not able to cleave caspase-specific substrates, the authors propose that OsAP25 and OsAp37 may activate downstream caspase-like protease(s) involved in PCD rather than directly acting on caspase-specific substrates (similarly to yeast YCA1) [46]. Nonetheless, more studies need to be done in order to describe the substrates of OsAP25 and OsAP37. Again, it is important to point out that our own bioinformatics analysis revealed that OsAP25 just displays one conserved catalytic motif whereas OsAP37 lacks both of them. Therefore, it is quite puzzling how both of these enzymes appear to function as active APs. This strengthens the idea that these atypical and nucellin-like APs are really different from their typical counterparts and that more studies are crucial in order to understand the functions and properties of these enzymes, and to what extent these can be considered new moonlighting enzymes.

1.4. *Arabidopsis* root development and lateral root formation

If we look just to the part of the plant that is above ground, most times is like looking at an iceberg: what we can actually see above water is just the tip of it. That is also true for plants. Although in general we cannot see the root system, it is there and is extremely important for the plant. Despite its importance, roots have traditionally been regarded as little more than a conduit for water and dissolved minerals but this vision has recently started to change. In recent years, a large volume of studies have been carried out on root development as well as on symbiotic and non-symbiotic nutrient acquisition, mostly driven by the imperative need to optimize nutrient application and uptake to improve crop productivity [93]. In this section, we will describe root development and lateral root formation and we will describe what little is known about the role of proteases in these two developmental processes.

Root development starts early in embryogenesis at the early-globular phase (see Figure 4). It is at this stage that hypophysis is formed (Figure 4, in yellow) which is the precursor cell to embryonic root formation. In the late-globular phase the hypophysis suffers an asymmetric division giving rise to an hypophysis cell and to a small organizer cell which is the precursor of the Quiescent Center (QC); by the embryo's heart stage, the embryonic Root Apical Meristem (RAM) is formed [94]. RAM has the stem cell niche that will give rise to all root tissues and it is responsible for root growth and adaptation, being present at the tip of the roots throughout all plant life-cycle [95]. Its cellular organization is very regular with initial cells (stem cells) surrounding the QC, which is a group of less mitotically active cells that is very important in the organization of cell differentiation as well as in the maintenance of the

undifferentiated stem cells [96]. Interestingly, it is the position of the stem cell regarding the QC that commits that cell to a cell lineage and so stem cells that are present shootward and in the lateral sides of the QC will give rise to the vascular, endodermal, cortex, epidermal and lateral root cap cells, whereas cells on the rootward of the QC will give rise to the columella root cap [95, 97].

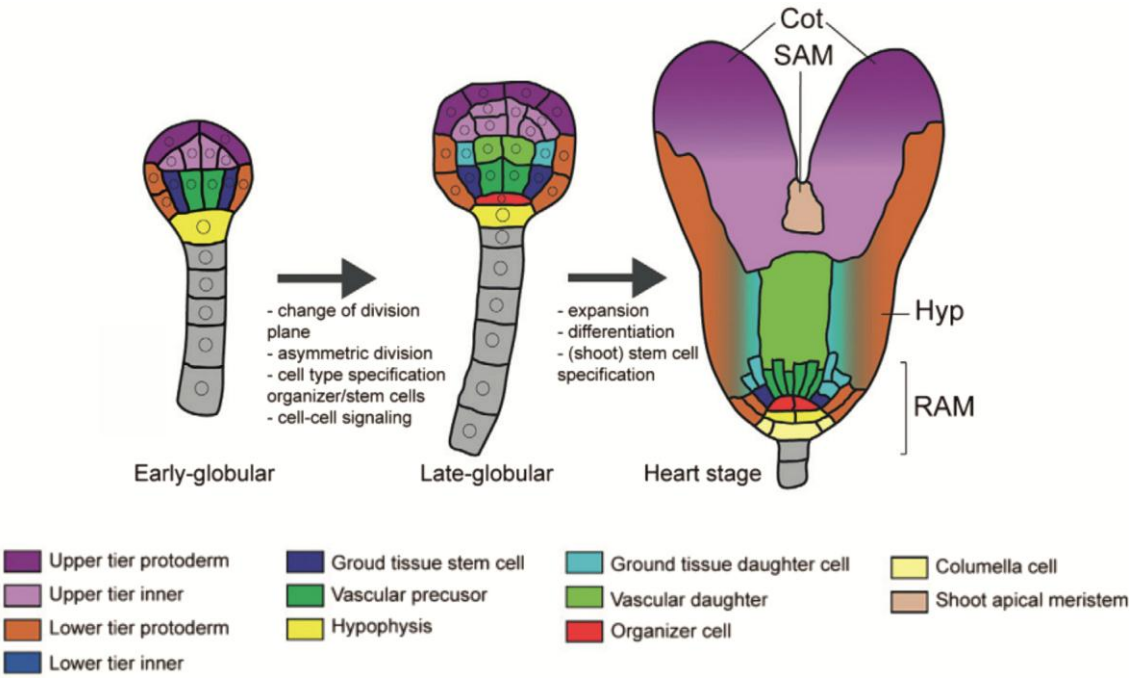


Figure 4 – Schematic representation of the late stages of *Arabidopsis* embryo development. It is in the early globular phase that root development starts with the formation of the hypophysis (in yellow) which is the embryonic root precursor cell. By late-globular phase, hypophysis suffers an asymmetric cell division which gives rise to a hypophysis cell and an organizer cell (in red). After a series of differentiation and specification processes, and by the end of embryo development at embryo's heart stage the RAM is formed and the root tissue will start to be formed. Cot – Cotyledon; SAM – Shoot Apical Meristem; Hyp – Hypophysis; RAM – Root Apical Meristem. Adapted from [94].

The initial phases of root development are followed by root elongation that is essential for normal plant development as well as for plant adaptation. As the root grows, three distinct developmental zones appear and the new cells that are formed in the RAM will progress through those zones. At the root tip the Meristematic Zone (MZ) is formed and it is here that RAM is located and cell division occurs. The newly formed cells will then pass through the Elongation Zone (EZ), which is located right above MZ, and here cells lose their ability to divide and begin to elongate to increase their length. Finally, cells exit EZ and enter the Differentiation Zone (DZ) were they finally differentiate into the different cells types forming the different tissues and structures [95]. It is also just in the DZ that lateral roots start to appear. Given the relevance of lateral root formation for normal root growth and adaptation we will also briefly describe lateral root formation.

Lateral roots do not initiate at the epidermis of the root. In fact, as shown in Figure 5**,** they originate deep in root tissue, in the pericycle cells [98]. The pericycle cell layer resides deep within the primary root, confined between the vascular bundle and the endodermis and at regular intervals some of these cells differentiate into Lateral Root Founder Cells (LRFC) [99]. As the name implies, LRFC are the initial cells that start lateral root formation and the process begins when a triplet of LRFC start to suffer asymmetric divisions giving rise to the Lateral Root Primordium (LRP) (Figure 5) [100]. As the LRP grows it starts to apply pressure in the endodermal cells which respond by becoming more and more flat until no endodermal cell content is left and its membrane fuses with LRP. The endodermis cells contain hydrophobic impregnations of the primary cell wall that fuse with each other forming the casparian strip which confers resistance to this cell layer. In order to fully cross endodermis, LRP cells then promote local degradation of this casparian strip, assume a dome shape and by continued cell division it starts to apply pressure in the cortex and epidermis cell layer [100]. Unlike the endodermis crossing, cortex and epidermis cells do not change shape, instead they are pushed apart to create an opening from where the LRP exits. This results from the expression of cell wall remodeling enzymes that are responsible for re-shaping cortex and epidermis cell wall [98]. At this stage, LRP exits the primary root tissue and root development processes start to originate the root tissue.

All these developmental processes are tightly regulated by plant hormones, such as auxin, and also by transcription factors [95, 98]. Interestingly, and despite their importance in other plant developmental processes, the role of plant proteases in root development and lateral root formation remains largely unexplored. In the following section we will describe what little is known about plant proteases and their role in root development and adaptation.

- ***Protease involvement in root processes***

In 2008 van der Hoorn published a review paper describing the roles of plant proteases reported in the literature [101]. Of the 40 proteases examined at the time only two were implicated in root processes and, somewhat unexpectedly, from 2008 to date just one more was described [102-104]. In this final section we will describe what is known about these proteases and the current ideas about their possible roles in root processes.

The first described protease implicated in root processes was the gene product of *sol1* and was shown to be involved in root meristem maintenance [102]. The authors were testing the hypothesis that root meristem maintenance was achieved via a CLV-like pathway similar to the shoot meristem maintenance. *CLAVATA* (CLV) genes (1 to 3) promote organ initiation in shoot meristem and together with *WUSCHEL* (WUS) genes, required for shoot stem cell

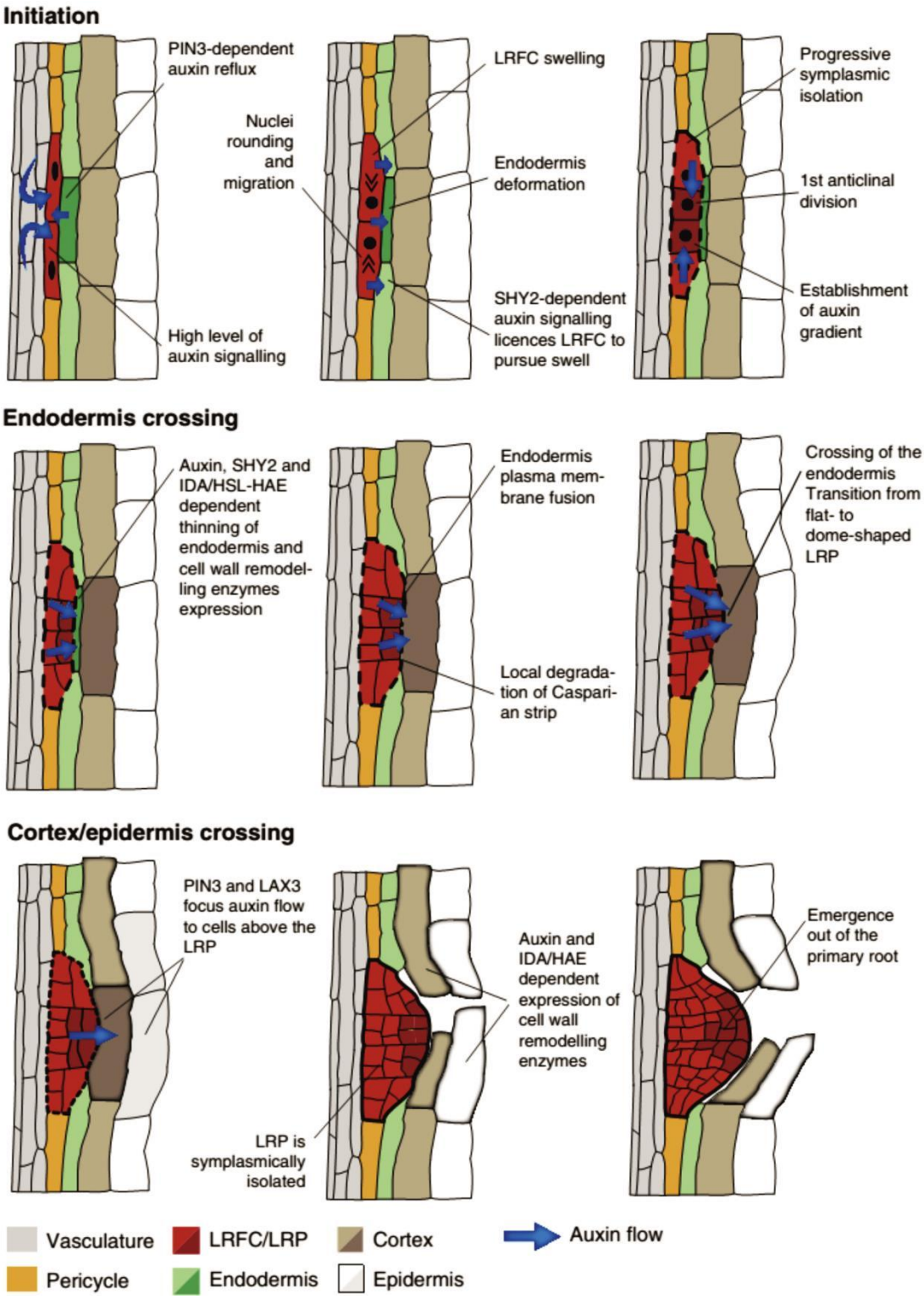


Figure 5 – Scheme of *Arabidopsis* lateral root formation (taken from [98]). In the initiation phase pericycle cells differentiate into Lateral Root Founder Cells (LRFC), the cells polarize and divide giving origin to the Lateral Root Primordium (LRP). LRP continues to grow and begin to apply pressure on the endodermis cells that become very thin. Eventually their plasma membrane fuses with LRP and local degradation of the casparian strip occurs which allow the LRP to definitely cross the endodermis and acquire a dome shape. LRP growth starts to apply pressure in the cortex and in the epidermis cell layers which will culminate in the remodeling of the cortex and epidermis cell wall which open a space in the root surface through where the new lateral root emerges. All this process is tightly regulated by the plant hormone auxin.

identity, they control shoot meristem maintenance [105]. In the CVL pathway, CVL3 is a secreted protein that activates CLV1-CLV2 receptor complex which then promotes organ initiation and it was described that overexpression of CVL3 leads to a decrease in shoot meristem size [106]. In an attempt to evaluate if a similar mechanism was also present in the root meristem, Casamitjana-Martínez *et al.* generated a transgenic *Arabidopsis* line overexpressing *CLE19* (a *CLV3* homologue) specifically in root tissue [102]. They reported that the overexpression line had shorter roots than the WT and indeed the overexpression of *CLE19* lead to a progressive loss of meristematic cells in a dose-dependent manner [102]. In a mutagenesis screen to find suppressors of this phenotype, they found that mutation in *SOL1* and *SOL2* genes completely suppress the root phenotype of *CLE19* overexpression lines, and that *SOL1* mutations maintain complete suppression throughout plant life while *SOL2* mutation effect was lost after a while. According to the authors, *SOL1* encodes a putative Zn^{2+}-carboxypeptidase that was predicted to be a transmembrane protein with a small cytoplasmic tail. Taking all these results together, authors concluded that there is indeed a CLV-like pathway in root meristem and that under normal conditions SOL1 might cleave CLE19, probably leading to cell differentiation and root growth [102].

In 2007 Combier *et al.* described another protease involved in root processes; in this case involved in the root symbiosis between the legume *Medicago truncatula* and the nitrogen-fixing bacteria *Sinorhizobium meliloti* [103]. The authors described a putative matrix metalloendoprotease (MMP) whose expression was triggered specifically by root infection with *S. meliloti* and specifically at the infection tread (IT), which is a plant cell structure that is formed in infected root hairs and through which bacterial infection takes place [103]. The protease was named MtMMPL1 (MMP-like 1) and it has the normal primary structure of a MMP: a signal peptide, a conserved pro-peptide and a sequence similar to the MMP catalytic domain. The major difference in this primary structure organization is in the MMP catalytic domain where the glutamate residue that normally appears next to the first histidine residue is changed to a glutamine. In animal MMPs this glutamate-to-glutamine change results in a dramatic decrease in MMP activity. Interestingly, authors found this mutation in other MMPs from legumes, claiming that this might be a conserved feature in some legume MMPs [103]. Although *MtMMPL1* expression appeared to be induced by *S. meliloti* infection, it does not seem to be crucial for normal infection and nodule formation since its RNAi knockdown line did not show major differences in these parameters when compared with the WT. Interestingly, the infection sites in this mutant line appear to be slightly bigger than in the WT. On the other hand, *MtMMPL1* overexpression resulted in a 60% decrease in the number of *S. meliloti*-induced nodules but those that were formed were normal. Supported by these

observations, the authors speculate that MtMMPL1 can act in two ways: i) it can play a positive role in infection by contributing to the formation of the IT cell wall or the extracellular cell matrix; or ii) it can control the number of infecting bacteria [103]. Because of the glutamate-to-glutamine exchange in MtMMPL1 active site, the authors speculate that this protein may not depend on its catalytic activity to perform its function. However, the molecular mechanisms behind MtMMPL1 function still remain unclear and additional experiments are required to describe MtMMPL1 biochemical properties, its precise subcellular localization and possible interactors [103].

The third protease involved in root processes described so far was identified in 2000 by Naito *et al.* but it was just in 2008 that its role was studied in more detail by Li and *et al.* [104, 107]. It is a cysteine protease (CP) that was named AsNODF32 and (as MtMMPL1) is involved in nitrogen-fixing symbiosis but in the legume *Astragalus sinicus*. AsNODF32 is expressed specifically in root nodules and its expression increases throughout the lifespan of the nodule, being stronger in old senescent nodules. Li and *et al.* reported that *AsNODF32* RNAi knockdown led to delayed nodule senescence that was linked to an extended period of nitrogen-fixation and to an increase of plant lifespan [104]. These results led the authors to conclude that this CP was involved in the regulation of nodule senescence and that its silencing was benefic to the plant by delaying this process [104]. Prior to this study, other authors reported the presence of CP in the nodules of different legumes and also their importance for nodule formation and maintenance [108, 109]. Nevertheless, Li *et al.* were the first researchers to study in more detail the role of CPs in nodule formation and senescence, but the molecular mechanisms behind AsNODF32 activity still remain elusive.

As illustrated by these examples, the role of proteases in root processes remains largely unclear with most of the proteases described so far linked to plant adaptations to abiotic stress conditions like drought or nitrogen starvation. Indeed, under these two conditions root proteases appear to be crucial to plant adaptation especially in plants that cannot establish symbiotic interactions like *Arabidopsis* since they are responsible for remobilizing the AA, peptides and proteins that are absorbed by the root tissues which maintains the nitrogen status of the plants [93]. Interestingly enough, no member of the AP family has been implicated to date in root development and adaptation processes, what is clearly indicative of how limited is still our understanding of this family of plant proteases.

1.5 Objectives

The large number of genes coding for proteases belonging to the A1 family of APs present in different plant species, added to the fact that all other classes of plant proteases are also expanded in number, clearly points towards a significant role of proteolysis in plant physiology and development. However, understanding proteolysis as a regulatory mechanism remains limited, as for most of these proteases their functional, structural and biochemical characterization remains very limited (not to say mostly inexistent). In the particular case of Arabidopsis APs, from the reported forty-six genes coding for atypical and nucellin-like APs, only eight of these have been studied in more detail, clearly emphasizing the current lack of knowledge about this class of proteases. Strikingly, the experimental evidences available for some of these APs implicate them in very important plant developmental processes as well as in stress responses, urging for additional and perhaps more integrative studies to fully understand the role of this family of proteases in the control of such diversified biological processes.

In line with this, we proposed a combined approach where we integrated the study of *in vivo* function, biochemical properties and the identification of native substrates for two new putative atypical APs from *A. thaliana*. The genes encoding these proteases - At2g03200 and At4g30040 - were initially described as regulated by transcription factors involved in pollen development [41, 85]. The de-regulation of these transcription factors led to severe male sterility phenotypes but the role of these two putative atypical APs in those phenotypes was not further explored. Supported by these findings, and given the critical roles of other two atypical APs – PCS1 and UNDEAD - in the regulation of different PCD events during anther development [37, 41], we initially hypothesized that these two gene products [UniProtKB accession numbers Q84M99 (At2g03200) and Q9SZV7 (At4g30040)] could also take part in pollen developmental processes. Therefore, one of the initial goals of this project was to evaluate the expression patterns of these two genes and study the *in vivo* functions of both encoded proteins, primarily focused on their roles during *A. thaliana* pollen and anther development, by making use of available gene knockout lines. Furthermore, supported by the observed expression patterns also in roots (particularly for the gene At2g03200) these studies were expanded to evaluate the potential role of the putative proteases in root growth and lateral root formation. These results are compiled in Chapter 2, and were performed in the laboratory of Professor Alice Y. Cheung at the University of Massachusetts in Amherst, USA.

Another goal of this project was the optimization and efficient production and purification of the recombinant forms of both proteases, in order to provide a detailed

biochemical characterization with respect to their enzymatic properties, inhibition profiles and specificity preferences, in the latter case, by making use of the high throughput proteomics technique called Proteomics Identification of Cleavage Sites (PICS). We started by addressing the use of *E. coli* as the heterologous system for production of these proteases in their soluble form, and next we evaluated the potential of the innovative expression system magnICON® (a very versatile and robust plant-based expression system) as an effective alternative for the production of these proteases. The magnICON® system was used in the laboratory of Professor Herta Steinkelner at BOKU University in Vienna, Austria. The results regarding the recombinant proteins are presented and discussed in Chapter 3.

Finally, to add an additional node to our integrative study, we proposed to start dissecting the *in vivo* molecular pathways in which our target atypical APs are involved. To this end, we used a combination of two high throughput proteomics approaches to analyze comparatively the proteome samples derived from the most representative phenotypes observed for each protease: i) quantitative shotgun proteomics, for the global profiling of the dynamic changes in the proteomes, not only between transgenic lines with altered protease expression and WT plants, but also under two different growth conditions; and ii) N-Terminal Amino acid Isotopic Labeling of Substrates (N-TAILS), a powerful approach to directly determine native substrates of proteases. Both proteomics approaches were conducted in collaboration with Dr. Pitter Huesgen at his laboratory at Forschungszentrum Jülich, Jülich, Germany. To our knowledge, this is the first proteomics study ever reported on plant atypical APs and the results are detailed and discussed in Chapter 4.

With the use of this integrated approach, combining data from *in vivo* and *in vitro* studies as well from proteomics analysis, we expect to pave the way for further research into dissecting the cellular functions of this important class of plant proteases.

Chapter II – *In vivo* Studies

2.1 Introduction

In Chapter I we described what is known about the role of atypical and nucellin-like APs in a variety of plant processes and it is clear that they are extremely important enzymes for the plant. They have been reported to play important roles in tightly regulated plant processes such as plastid homeostasis [31, 40], resistance to biotic and abiotic stresses [36, 38] and even PCD, mainly during anther and pollen development [37, 41]. Despite their importance, out of the 46 putative atypical and nucellin-like APs encoded in the *Arabidopsis* genome [34], just eight of them have been studied in more detail (see Table 2). Moreover, if we extend this analysis to all plant atypical and nucellin-like APs, just 19 of them were studied so far (see Table 2) which clearly demonstrates the lack of knowledge concerning this class of enzymes and the need for more studies to unveil their functions and properties.

Having all this in consideration, in this work we proposed to study in detail the function and properties of two novel *Arabidopsis* atypical APs (*At2g03200* and *At4g30040* gene products) that are regulated by two transcription factors crucial for normal pollen development [41, 85]. In 2007 Ito *et al.* reported a new *Arabidopsis* transcription factor whose absence led to a severe male sterility phenotype [85]. The authors named it *MALE STERILITY 1* (MS1) and due to its importance for proper pollen development they tried to describe the putative downstream genes regulated by this transcription factor. Among them was At2g03200, a putative AP which was positively regulated by MS1 and whose expression was significantly decreased in the *ms1* mutant plants [85]. At4g30040 appeared in another study where Phan *et al.* described the role of an atypical AP (UNDEAD) in PCD processes during pollen development [41]. In this report, the authors proved that *UNDEAD* is positively and directly regulated by the *MYB80* transcription factor and that its absence led to a very severe male sterile phenotype much similar to the phenotype observed in the absence of *MYB80* [41]. Among the genes that were regulated by *MYB80* was *At4g30040* which was also found to be positively regulated by this transcription factor, since a 9 fold increase in its expression was reported when *MYB80* transcription was promoted [41]. *At4g30040* and *UNDEAD* expression were both positively regulated by *MYB80* at almost the same level (*UNDEAD* expression was 14 fold increased upon *MYB80* induction) but unlike *UNDEAD*, *At4g30040* was not directly regulated by *MYB80* and because of that this protein was not studied in more detail [41]. In this chapter, our goal is to describe the *in vivo* function(s) of both *At2g03200* and *At4g30040* gene products. Giving the evidences that were just described, and also the role of other

atypical APs in pollen development [37, 41], our first working hypothesis was that these two proteins could be involved in PCD events during pollen development.

As we will see throughout this chapter, our results lead us to expand our studies to another plant process - root development - and the fact is that virtually nothing is known about the role of plant proteases during this process. As discussed previously, roots are essential for normal plant development and adaptation but until recently they were regarded as little more than a conduit for water and dissolved minerals [95]. Interestingly, the role of plant proteases during root development remains mostly elusive [93] and considering specifically plant atypical APs, they have never been directly associated with root development. This fact is about to change and we will present results that prove the involvement of atypical APs in root development.

During the following chapter we will present results regarding the *in vivo* functions of the putative atypical APs encoded by the genes *At2g03200* and *At4g30040*. We will demonstrate that although both proteins are indeed expressed in the reproductive tissues, the single KO lines for each gene do not have any obvious pollen related phenotype and are as fertile as the WT plants. However, both proteins are also expressed in root tissue and we will present evidences that they play an important role in primary root elongation and especially in lateral root formation. Because of this phenotype, these genes were designated *REGULATOR OF LATERAL ROOT 1* or *RLR1,* for *At2g03200*, and *REGULATOR OF LATERAL ROOT 2* or *RLR2* for *At4g30040* (this designation will be used hereafter).

2.2 Methods and Materials

- ***Plant material***

A. thaliana ecotype Col-0 seeds were used for all experiments as the WT control and were kindly provided by Dr. Alice Cheung. The KO lines used were Salk_038980.40.95.x and Salk_001270.39.20.x, for *RLR1* and *RLR2,* respectively (schematic representation in Figure 6). They are T-DNA insertion lines and were purchased from the Arabidopsis Biological Resource Center (ABRC).

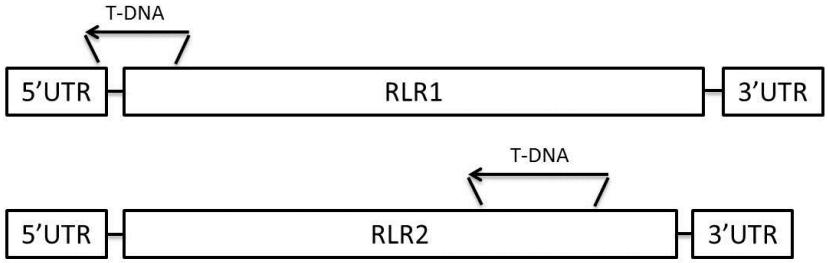


Figure 6 – Schematic representation of the T-DNA KO lines used. *RLR1* KO line (Salk_038980.40.95.x) T-DNA insertion resulted in the deletion of the first 250 bp of *RLR1* coding sequence. *RLR2* KO line (Salk_001270.39.20.x) T-DNA insertion resulted in the deletion of *RLR2* sequence from nucleotide 631 to nucleotide 1079. Both KO lines were purchased from ABRC.

- ***Plasmid DNA constructs***

  - ***Plasmid DNA constructs used for the generation of A. thaliana transgenic lines***

Several transgenic lines were generated for each atypical AP in *A. thaliana*, using different promoter sequences (caulifower mosaic virus 35S promoter or the endogenous promoters of *RLR1* and *RLR2*) as well as different fusion tags (GFP or HA). The constructs are listed in Table 3. All constructs were inserted into plasmid pAC1352 [110] using traditional cloning strategies and PCR-based methodologies. *RLR1* (At2g03200) and *RLR2* (At4g30040) endogenous promoter sequences were considered to be the 2000 bp sequence upstream the corresponding coding regions and they were amplified to include restriction sites for *Xba*I and *BamH*I at 5'- and 3'-ends, respectively, using the primers Prom1_for/Prom1_rev (*RLR1*) and Prom2_for/Prom2_rev (*RLR2*), listed in Table 4. The PCR amplified products were digested with *Xba*I and *BamH*I and cloned into pAC1352, replacing the 35S promoter flanked by the same restriction sites that was present in the original vector. The sequences encoding the full-length

proteases were PCR amplified to include restriction sites for *BamH*I and *Sal*I at 5'- and 3'-ends, respectively (primer sequences listed in Table 4), and then inserted into pAC1352 in frame with a C-terminal GFP coding sequence or HA tag. In this case, two versions of pAC1352 vector were used which had either the GFP or HA tag coding sequence following the *Sal*I restriction site. Since *RLR1* sequence contains a *BamH*I recognition sequence, a silent mutation was introduced to eliminate this restriction site, prior amplification and cloning into pAC1352. Mutagenesis was performed using the Quick Change site-directed mutagenesis kit (Stratagene) and the primers BamHI_RLR1_for and BamHI_RLR1_rev, listed in Table 4. All positive clones were selected by restriction analysis and confirmed by DNA sequencing.

Table 3 – DNA constructs used for the generation of *A. thaliana* transgenic lines.

Constructs	Description	Used
35S:: RLR1-GFP	CaMV 35S promoter::RLR1 cDNA-GFP	Yes
35S::RLR1-HA	CaMV 35S promoter::RLR1 cDNA HA tagged	No
35S:: RLR2-GFP	CaMV 35S promoter::RLR2 cDNA-GFP	Yes
35S::RLR2-HA	CaMV 35S promoter::RLR2 cDNA HA tagged	No
pRLR1::GUS	RLR1 promoter::GUS	Yes
pRLR2::GUS	RLR2 promoter::GUS	Yes

Column marked with "used" describes transgenic lines effectively used in this study.

Table 4 – Primers used to generate the PCR products used for generation of plant transgenic lines.

Name	Sequence
Prom1_for	TCT AGA TTC ATA ACA ATG GAA C
Prom1_rev	GGA TCC GGT TTG TGT GTG TGT GAG TG
Prom2_for	TCT AGA GAT TCA TTG ATT TTA CTA ACA C
Prom2_rev	GGA TCC CTT TGA AAC ACT TCA CAT ACA
BamHI_RLR1_for	CCA GTA GAT GAC TCT GGT TCC ACG GGA CTT GAC
BamHI_RLR1_rev	GTC AAG TCC CGT GGA ACC AGA GTC ATC TAC TGG
RLR1_for	GGA TCC ATG GCT TCT TCT TCT TCT TCT
RLR1_rev	GTC GAC CAA TTT TCC ACA TTC AGT GGG
RLR2_for	GGA TCC ATG GCG ATT TTC TTT ACA TCT
RLR2_rev	GTC GAC AAA AGA AAC CTC CAT TGC TTC

In grey are highlighted the restriction site sequences.
In black are highlighted the mutated nucleotides.

For the generation of the GUS transgenic lines, *RLR1* and *RLR2* promoter sequences were amplified as described before and inserted into the pBI121 vector (Clontech) upstream the GUS sequence using *BamH*I and *Sal*I restriction enzymes.

○ *Plasmid DNA constructs used for transient expression of GFP-fusion proteins in A. thaliana protoplasts*

Different chimeric constructs of RLR1 and RLR2 were designed to evaluate the subcellular localization in *A. thaliana* protoplasts. The constructs are summarized in Table 5 and consist not only of the full-length coding sequences but also several truncated constructs used to study localization signals present in RLR1 and RLR2 N-terminal sequences. For each protein three truncated N-terminal sequences were evaluated: i) the putative pre-sequence or signal peptide; ii) the putative pre-pro domain; iii) the N-terminal sequence up to the first catalytic aspartate; together with the WT full-length protein as well as its active site mutant (first catalytic aspartate substituted by an alanine). Each sequence was amplified by PCR to include the restriction sites *Pst*I and *Sal*I at 5'- and 3'-ends, respectively (primer sequences listed in Table 6) and the combination of primers used for each construct is listed in Table 5**.** The active site mutant constructs were generated by site-directed mutagenesis using the Quick Change site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. The primers used were Mut_RLR1_for/Mut_RLR1_rev for RLR1 and Mut_RLR2_for/Mut_RLR2_rev for RLR2 and primer sequences are listed in Table 6. All *RLR1* and *RLR2* cDNA sequences were inserted into a modified Bluescript pSK vector (Stratagene), immediately downstream the 35S promoter and in frame with the GFP coding sequence at the 3' end, using the restriction enzymes *Pst*I and *Sal*I restriction enzymes. All positive clones were selected by restriction analysis and confirmed by DNA sequencing.

• *Plant growth conditions*

A. thaliana ecotype Col-0 plants were used for all experiments (unless otherwise stated). Before germination, seeds were surface sterilized by washing for 5 min in 70% ethanol (twice), followed by washing for 5 min in a 70% bleach solution. After extensive washing with autoclaved water, seeds were placed in B5 medium supplemented with 1% sucrose solidified by 0.7% Agar (unless otherwise stated) and incubated in the dark at 4 ^{o}C for 2 days. Plates were then transferred to a growth chamber and seeds were allowed to germinate at 22 ^{o}C under a 16h/8h light/dark cycles and 60% humidity. For growth to maturity, seeds were either sown directly on soil, or 10-d-old tissue culture-grown seedlings were transferred to soil and maintained in a growth chamber at 22 ^{o}C under a 16h/8h light/dark cycle.

Table 5 – DNA constructs used for transient expression in protoplasts

Constructs	Description	Primer Combination
35S::RLR1$_{1-39}$-GFP	CaMV 35S promoter:: RLR1 first 39 aa cDNA sequence–GFP. This is RLR1 signal peptide construct.	RLR1_for/SP_RLR1_rev
35S::RLR1$_{1-90}$-GFP	CaMV 35S promoter:: RLR1 first 90 aa cDNA sequence–GFP. This is RLR1 prepro peptide construct.	RLR1_for/Pro_RLR1_rev
35S::RLR1$_{1-124}$-GFP	CaMV 35S promoter:: RLR1 first 124 aa cDNA sequence–GFP. This is RLR1 sequence up to the first catalytic aspartate construct.	RLR1_for/DTG_RLR1_rev
35S::RLR1-GFP	CaMV 35S promoter::RLR1 cDNA-GFP. Full-length construct.	RLR1_for/RLR1_rev
35S::RLR1$_{D125A}$-GFP	CaMV 35S promoter::RLR1 cDNA with a point mutation in aa 125 (D to A) –GFP. This is the active site mutant construct.	RLR1_for/RLR1_rev
35S::RLR2$_{1-39}$-GFP	CaMV 35S promoter:: RLR2 first 39 aa cDNA sequence–GFP. This is RLR2 signal peptide construct.	RLR2_for/SP_RLR2_rev
35S::RLR2$_{1-132}$-GFP	CaMV 35S promoter:: RLR2 first 132 aa cDNA sequence–GFP. This is RLR2 prepro peptide construct.	RLR2_for/Pro_RLR2_rev
35S::RLR2$_{1-167}$-GFP	CaMV 35S promoter:: RLR2 first 167 aa cDNA sequence–GFP. This is RLR2 sequence up to the first catalytic aspartate construct.	RLR2_for/DTG_RLR2_rev
35S::RLR2-GFP	CaMV 35S promoter::RLR2 cDNA-GFP. Full-length construct.	RLR2_for/RLR2_rev
35S::RLR2$_{D168A}$-GFP	CaMV 35S promoter::RLR2 cDNA with a point mutation in aa 168 (D to A) –GFP. This is the active site mutant construct.	RLR2_for/RLR2_rev

Table 6 – PCR primers used to generate the PCR products used for protoplast transfection

Name	Sequence
RLR1_for	CTG CAG ATG GCT TCT TCT TCT TCT TCT
RLR1_rev	GTC GAC CAA TTT TCC ACA TTC AGT GGG
SP_RLR1_rev	GTC GAC AAG GTT TTT CGG AAG AGT ACG
Pro_RLR1_rev	GTC GAC AGG CTT AGA GGC TAC AGC T
DTG_RLR1_rev	GTC GAC AAC AAT AGC AGA GTA TTT AAC C
Mut_RLR1_for	TAC TCT GCT ATT GTT GCG ACC GGA AGT GAC CTT
Mut_RLR1_rev	AAG GTC ACT TCC GGT CGC AAC AAT AGC AGA GTA
RLR2_for	CTG CAG GCG ATT TTC TTT ACA TCT
RLR2_rev	GTC GAC AAA AGA AAC CTC CAT TGC TTC
SP_RLR2_rev	GTC GAC GTG GTA GGA ATG AAC GAC A
Pro_RLR2_rev	GTC GAC AGA TCG AGA AGG ATC GAA AA
DTG_RLR2_rev	GTC GAC ATC CAC GTA TCT CAT GGA AT
Mut_RLR2_for	ATG AGA TAC GTG GAT GCG ACA GGG AGC AAA GGC
Mut_RLR2_rev	GCC TTT GCT CCC TGT CGC ATC CAC GTA TCT CAT

In grey are highlighted the restriction site sequences.
In black are highlighted the mutated nucleotides.

- ***A. thaliana transformation***

In order to generate transgenic *A. thaliana* lines, plants were transformed with *Agrobacterium tumefaciens* harboring the desired constructs (listed in Table 3). Plasmid DNA was transformed into *A. tumefaciens* strain GV3101 by direct transformation. Briefly, 10 µg of plasmid DNA were added to an aliquot of *Agrobacterium* competent cells and the mixture was incubated on ice for 30 min. The mixture was frozen in liquid nitrogen for 1 min, thawed at 37 ^{o}C for 5 min and 500 µL of LB were added. After an incubation at 30 ^{o}C for 4 h under gentle shaking, cells were spun down, spread in LB/agar plates supplemented with antibiotics (25 µg/mL rifampicin, 100 µg/mL ampicillin) and incubated at 30 ^{o}C for 72h. Positive clones were selected by colony PCR using the primers listed in Table 7. Positive clones were then grown in 200 mL of LB supplemented with rifampicin (25 µg/mL) and ampicillin (100 µg/mL) overnight at 30 ^{o}C. Cultures were pelleted at 5000*g* for 15 min and resuspended in Dipping Buffer (1 mM $MgSO_4$, 5% sucrose) supplemented with 0.02% of silwet L-77. At this point, cell suspensions were ready to be used for *A. thaliana* transformation.

A. thaliana transformation was achieved using the floral dipping methodology [111]. Briefly, flowering *A. thaliana* WT plants were dipped in the *A. tumefaciens* suspension for 2-3s

and then placed in the growth chamber in the dark for 24 h. At the end of this incubation time, plants were grown as described previously. This procedure was repeated 7 days after the first dipping, plants were allowed to grow normally and seeds were collected (T0 seeds). To select the transformed seeds, T0 seeds were surface sterilized and plated in B5 plates (as described before) supplemented with kanamycin (50 µg/mL). Seeds were allowed to germinate during 12-15 days in the growth chamber and then positive lines were transferred to soil for growth to maturity.

Table 7 – Primers used for selection of *A. tumefaciens* positive transformants by colony PCR

Name	Sequence
Agro_RLR1_for	ACA GGT GCA AGT TTA GAT GG
Agro_RLR1_rev	GAA AAT AGA CAT CCC ATT AGA A
Agro_RLR2_for	ATT TTG GGA GAT ACC ACA CC
Agro_RLR2_rev	GTT CAA ATT ACC AGG AGT CAC

- ***Root growth assays***

For root length assays, seeds were surface sterilized and plated on half-strength Murashige and Skoog medium (½ MS) (Phyto Technology Laboratories) or modified ½ MS media without ammonium nitrate (Phyto Technology Laboratories) both of them supplemented with 1% sucrose and solidified by 0.7% agar. After a 2-day incubation in the dark, at 4 ^{o}C, plates were transferred to a growth chamber and allowed to germinate vertically at 22 ^{o}C under a 16h/8h light/dark cycle and 60% humidity for 10 days. Root length was marked at 3, 5, 7 and 10 days and the number of lateral roots was determined at the end of the experiment. At this time, plates were photographed and root length was measured using the ImageJ software (http://imagej.nih.gov/ij/). 150-200 seeds from different seed batches were used in each experiment and the experiments were done in duplicate. Statistical analysis was performed by ANOVA type of analysis using the Prism software package (Graphpad Software Inc).

For the auxin experiments, seeds were surface sterilized and seeded in ½ MS plates supplemented with 1% sucrose and solidified by 0.7% agar. After a 2 day incubation in the dark, at 4 ^{o}C, plates were transferred to a growth chamber and allowed to germinate horizontally for 4 days at 22 ^{o}C under a 16h/8h light/dark cycles and 60% humidity. Plants were then transferred to new ½ MS plates supplemented with 1% sucrose and the desired concentrations of auxin. They were allowed to growth vertically for 5 days under the same

growth conditions. The number of lateral roots was determined at the end of the experiment, plates were photographed and root length was measured as described before.

- ***Protoplast isolation and transfection***

Mesophyll and root protoplasts were isolated from 3-wk-old *A. thaliana* seedlings grown in B5 medium plates supplemented with 1% sucrose as previously described. For protoplast isolation, seedlings were chopped in very small pieces and incubated overnight in a petri dish containing 25-30 mL of enzyme solution (0.5% (w/v) cellulase R10, 0.2% (w/v) macerozyme R10 in K3 medium (3.2 g/L B5 medium, 0.5 mg/L MES, 0.1 mg/L myo-inositol, 0.25 mg/L NH_4NO_3, 0.75 mg/L $CaCl_2$, 0.25 mg/L xylose, 0.4 M sucrose, adjusted to pH 5.6-5.8 with 1 M KOH)) in the dark and at room temperature. The enzyme solution was then carefully removed with a Pasteur pipet and the plant tissue incubated with K3 medium for 15 min under gently orbital agitation and in the dark. K3 medium containing the protoplast suspension was filtered into a petri dish through a piece of nylon filter, transferred to autoclaved 16 mL glass tubes and the suspension was left undisturbed for 2h in the dark, at room temperature. The upper protoplast layer was gently transferred to a new 16 mL glass tube, 5 mL of K3 medium were added and the suspension was again left undisturbed in the dark for 2h. This procedure was repeated twice and protoplast layers were pooled together in a new 16 mL glass tube and again left undisturbed in the dark for 2h. The excess of K3 solution was then removed from underneath the protoplast layer, protoplast volume was estimated and finally the protoplasts were diluted two times in Suspension solution (0.4 M mannitol, 20 mM $CaCl_2$, 5 mM MES, pH 5.7 adjusted with 1M KOH).

For protoplast transfection, aliquots of 5-10 µg of plasmid DNA (maximum volume of 10 µL) were previously prepared in 2 mL Eppendorf tubes. A volume of 100 µL of protoplast suspension was added to the DNA followed by the addition of 110 µL of PEG solution (40% PEG 4000, 0.2 M mannitol, 0.1 M $CaCl_2$). Samples were mixed gently, incubated in the dark for 20 min and then diluted with 1.5 mL of K3 medium. Protoplasts were left undisturbed overnight in the dark, at room temperature for transient transgene expression.

- ***Alexander staining***

Flower buds from WT and T-DNA mutant plants were collected at different growth stages and incubated in a 70% ethanol solution for 24 h at room temperature. After this incubation, flower buds were dissected under the dissection microscope (Olympus SZ61) and

all tissue was discarded except for the anthers. Anthers were then incubated with Alexander's staining solution [112] for 24 h and observed under the microscope.

- ***Gus histochemical staining***

Whole flower buds and 13-d-old seedlings were collected and incubated in Gus Staining Solution (0.1 M $NaPO_4$ buffer pH 7, 0.5 mM $K_3[Fe(CN)_6]$, 0.5 mM $K_4[Fe(CN)_6].3H_2O$, 10 mM Na_2EDTA) supplemented with 0.2 mg/mL of X-Gluc. Vacuum was applied for 20 min to allow tissue infiltration and the mixture was incubated overnight at 37 ^oC. The tissues were then incubated in 70% ethanol until all chlorophyll was removed (at least 24 h) and observed under the dissection microscope (Olympus SZ61).

- ***Aniline blue staining***

Young WT flower buds were emasculated and after 16 hours they were hand-pollinated either with KO pollen or WT pollen. Pollen grains were left to germinate in the pistils for 5 h and then whole pistils were collected and fixed in 80% acetone overnight. Pistils were then transferred to an aniline blue solution (0.1 mg/mL in H_2O) and incubated 24 h in the dark. At the end of this incubation, pistils were observed under the microscope (Nikon Eclipse E800 microscope).

- ***Microscopy***

GFP fluorescence in the protoplast experiments was determined by fluorescence microscopy using Nikon Eclipse E800 microscope with a Spot CCD camera. The green fluorescence was acquired using a BP 450-490 excitation filter and a BP 515-565 emission filter while the red fluorescence was acquired using a LP 546-565 excitation filter and a LP 590 emission filter. For the aniline blue experiments, pictures were taken using the same microscope but with a different set of filters: 330-380 excitation and 400-420 emission.

2.3 Results

- ***RLR1 and RLR2 are expressed in reproductive tissues as well as in root tissue***

As we stated before, our initial goal was to study *RLR1* and *RLR2 in vivo* function and our first approach was to see in which tissues they were being expressed. To do that, we used the GUS reporter system [113] where *RLR1* and *RLR2* promoter sequences were fused to the GUS reporter gene (β-glucuronidase) and created *A. thaliana* transgenic lines harboring these constructs (pRLR1::GUS or pRLR2::GUS, listed in Table 3). The GUS reporter gene is an enzyme that degrades X-Gluc (5-bromo-4chloro-3-indolyl glucuronide) and this degradation gives rise to a blue compound. Since the expression of this gene is being driven by *RLR1* or *RLR2* promoter, this enzyme will be expressed in the tissues where those genes would be expressed and when supplemented with X-Gluc, a blue color can be observed [113].

The GUS staining results for *RLR1* are depicted in Figure 7 a) and b). This gene is expressed in reproductive tissues specifically in mature pollen grains (a)) and in vegetative tissues such as roots, mainly in root tips (b)). Unfortunately we were never able to observe GUS staining in all pRLR2::GUS transgenic lines that were generated so we do not have GUS staining results for *RLR2*. This can be due to two reasons: the *RLR2* promoter can be a weak promoter under the growth conditions that we were testing or we had some problem related with the construct used to generate pRLR2::GUS transgenic line. Using the *Arabidopsis* eFP browser [114] (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) which allow us to verify the expression pattern and levels of different genes based on published microarrays data, we can clearly see that the overall *RLR2* expression levels are quite low when compared with *RLR1*. This means that the *RLR2* promoter is probably a weak promoter, not promoting enough GUS expression to allow GUS staining.

In order to circumvent this problem and to confirm the observed GUS expression patterns for *RLR1*, we extracted mRNA from whole inflorescence and root tissue and performed RT-PCR using specific primers for both *RLR1* and *RLR2*. We could have looked for *RLR1* and *RLR2* expression in other tissues as well but due to time limitations we tested only the tissues where *RLR1* expression was observed by the GUS methodology. The RT-PCR results can be seen in Figure 7 c) and d) (for whole inflorescence and root tissue, respectively) and both genes are expressed in WT plants in these two tissues, although *RLR1* expression levels in roots appear to be higher than those of *RLR2* (d)). We also performed the RT-PCR experiments in the same tissues extracted from the T-DNA KO lines that were used in this study in order to

confirm if these were indeed KO lines. As we can see in Figure 7 c) and d) (KO lanes) no mRNA was amplified from these tissues, confirming that they are true KO lines.

Taking all these results together, we can conclude that both *RLR1* and *RLR2* are expressed in roots and in the reproductive tissues of WT plants. Moreover, *RLR1* is specifically expressed in mature pollen grains and in root tips.

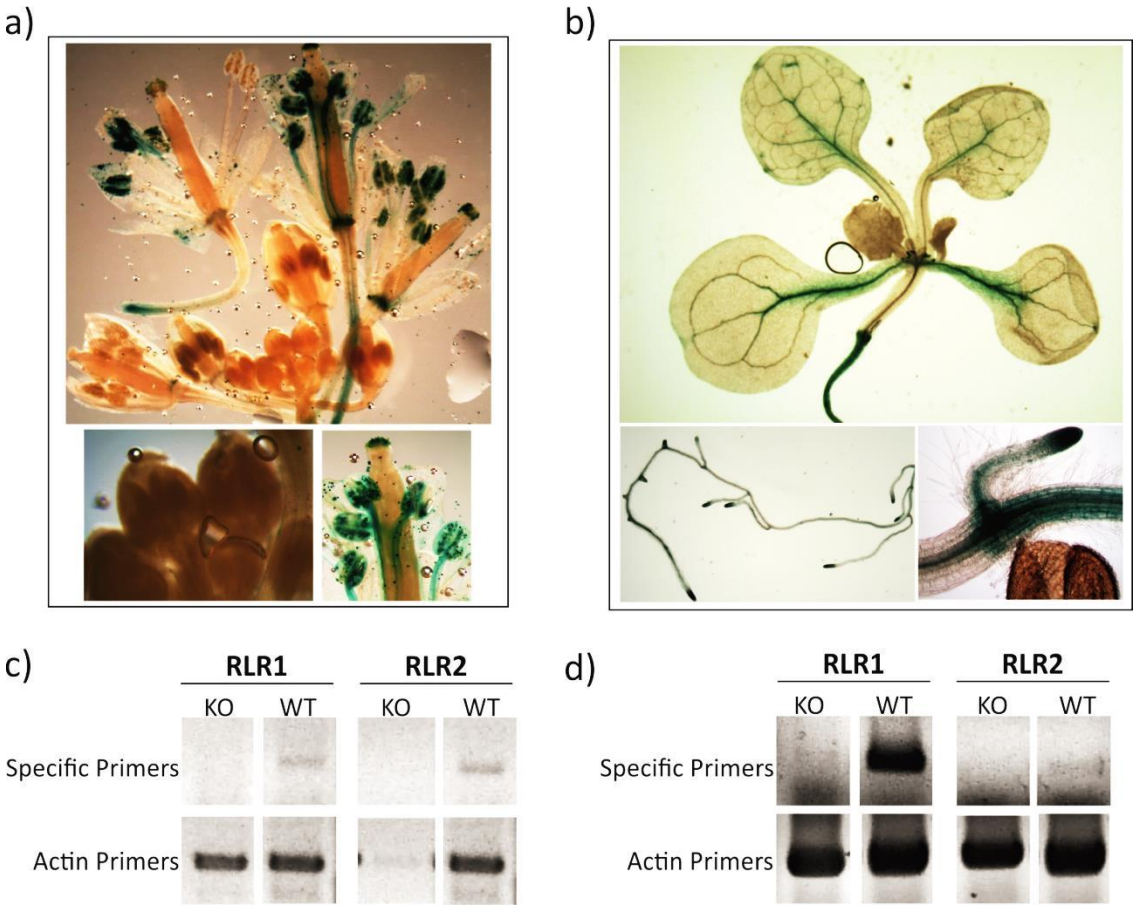


Figure 7 – *RLR1* and *RLR2* expression patterns. a) and **b)** Analysis of GUS activity in *A. thaliana* pRLR1::GUS transgenic lines in whole inflorescence buds and in 13-d-old transgenic seedlings, respectively. *RLR1* is expressed specifically in mature pollen grains and also in *A. thaliana* roots, especially in root tips. **c)** and **d)** RT-PCR analysis of Wild Type (WT) and Knockout (KO) whole inflorescence and root mRNA, respectively. Plants were grown for 10 days in normal 1/2MS media and root tissue was collected at the end of those 10 days. We used *RLR1* and *RLR2* specific primers and actin was used as a loading control. Both genes are expressed in floral buds and also in root tissue although *RLR1* expression levels are higher than those of *RLR2* in root tissue.

- ***RLR1 and RLR2 are targeted to the ER and possibly to the secretory pathway***

The GUS staining and RT-PCR assays used in the previous section provided important information on which *A. thaliana* tissues *RLR1* and *RLR2* are normally expressed but it does not allow to infer their subcellular localization. In order to do that we transfected *A. thaliana* mesophyll and root protoplasts with different *RLR1* and *RLR2* constructs fused to GFP (depicted in Figure 8) and checked for GFP expression and localization under the fluorescence microscope after several hours of expression. As we can see in Figure 8, in this experiment we used not only the full length *RLR1* and *RLR2* cDNA but also several N-terminal truncated cDNA sequences with different lengths in order to describe the minimal N-terminal sequence responsible for protein subcellular localization.

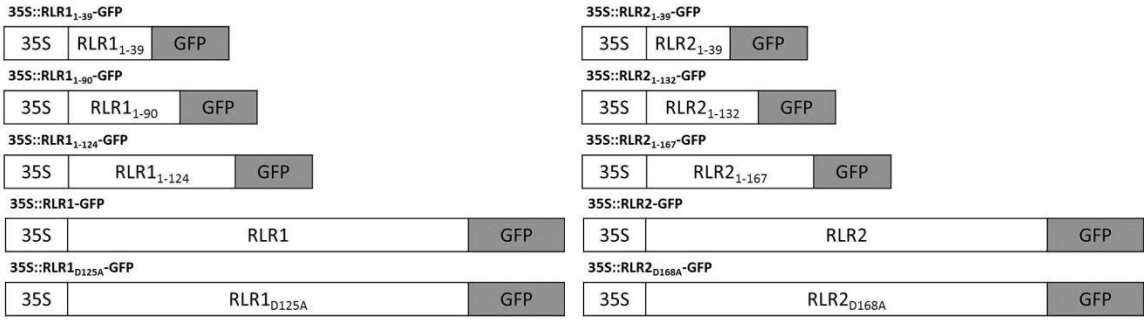


Figure 8 - DNA constructs used for protoplast transfection. All constructs had the 35S promoter driving their expression as well as the GFP gene fused to their 3' end. The first 3 constructs of each protein are different lengths of their N-terminal cDNA sequences, always starting at the first methionine (Signal Peptide, Signal Peptide + Pro Peptide and sequence until the first catalytic aspartate, respectively). $RLR1_{D125A}$ and $RLR2_{D168A}$ are the active site mutants of each protein in which the first catalytic aspartate was mutated to an alanine. Detailed description of these constructs can be found in Table 5.

The results for RLR1 are depicted in Figure 9 and indeed this protein has a functional signal peptide since construct $RLR1_{1-39}$ (signal peptide + 12aa) showed a typical ER localization pattern. The same ER localization was observed for the rest of the constructs, including the full length RLR1 and its active site mutant. In the case of RLR2 (Figure 10) the signal peptide is also active since construct $RLR2_{1-39}$ (signal peptide + 12aa) has an ER localization. Nonetheless, as we increase the length of the constructs, it is possible to observe the presence of protein aggregates suggesting that some of these longer N-terminal fusion constructs may be unstable. Interestingly, the ER staining pattern appears to be recovered for the full-length RLR2 and its active site mutant.

These results confirmed that both RLR1 and RLR2 have a functional signal peptide (SP) that targets these proteins to the secretory pathway, and that likely the SP is the only N-

terminal sequence responsible for protein targeting. The aggregated pattern observed for RLR2 truncated constructs further suggests some negative impact of the putative prosegment on the stability of this fusion protein. However, we cannot conclude that both RLR1 and RLR2 are ER proteins since they can still continue through the secretory pathway and be secreted to the extracellular space.

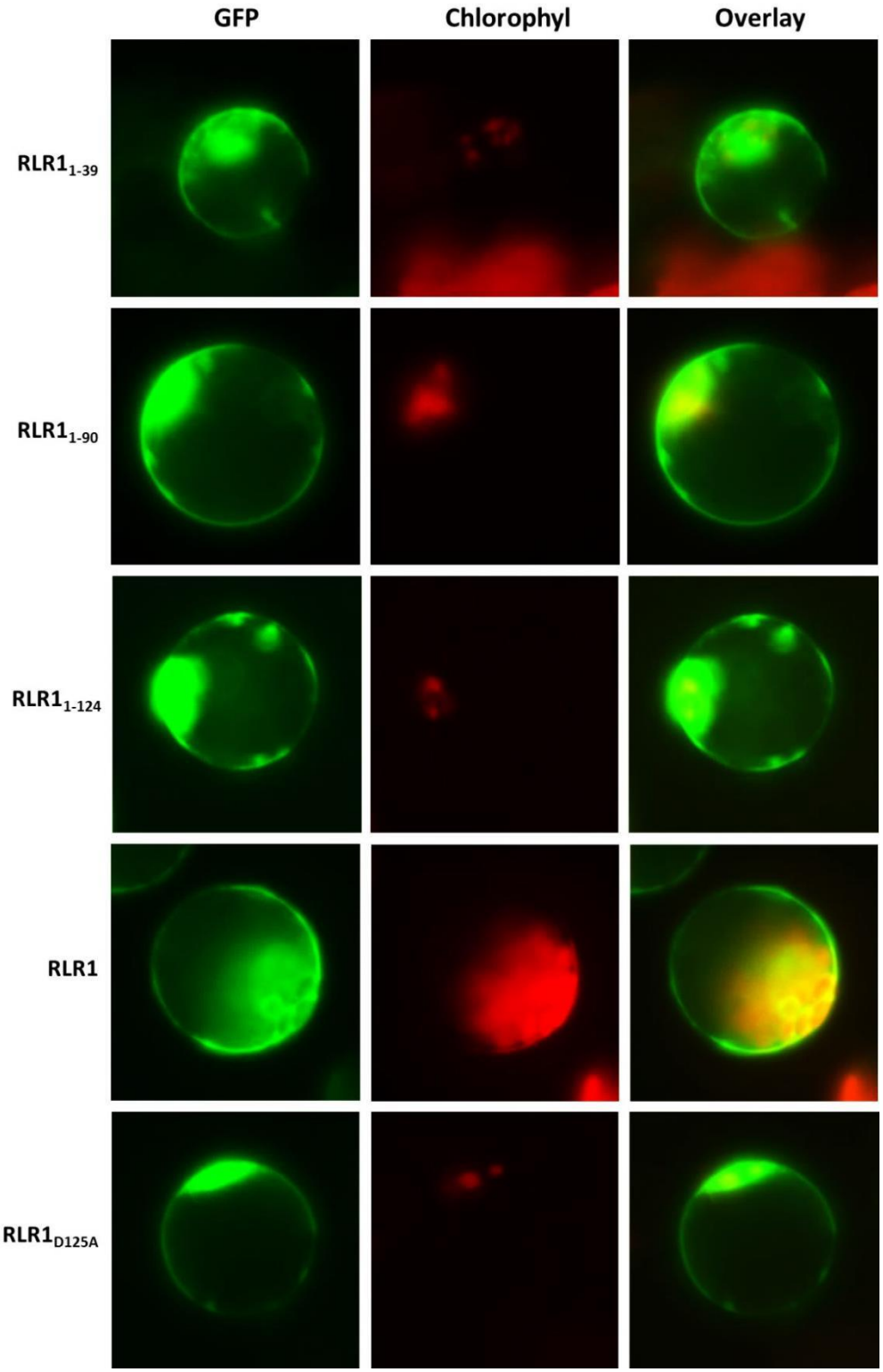


Figure 9 - RLR1 subcellular localization. Mesophyll protoplasts extracted from 3-w-old WT plants were transformed with 5-10 µg of plasmid DNA of several RLR1 constructs fused to GFP at their C-terminal. After 16h, transfected protoplasts were observed by fluorescent microscopy: green channel for GFP and red channel for chlorophyll. **RLR1$_{1-39}$)** first 39 aa (putative signal peptide) plus 12 aa; **RLR1$_{1-90}$)** first 90 aa (putative pro peptide); **RLR1$_{1-124}$)** first 124 aa **(**sequence until the first catalytic aspartate); **RLR1)** full length protein; **RLR1$_{D125A}$)** RLR1 active site mutant. The signal peptide appears to be functional and RLR1 appears to be localized in the ER although we cannot discard that it can continue through the secretory pathway and secreted to the extracellular space.

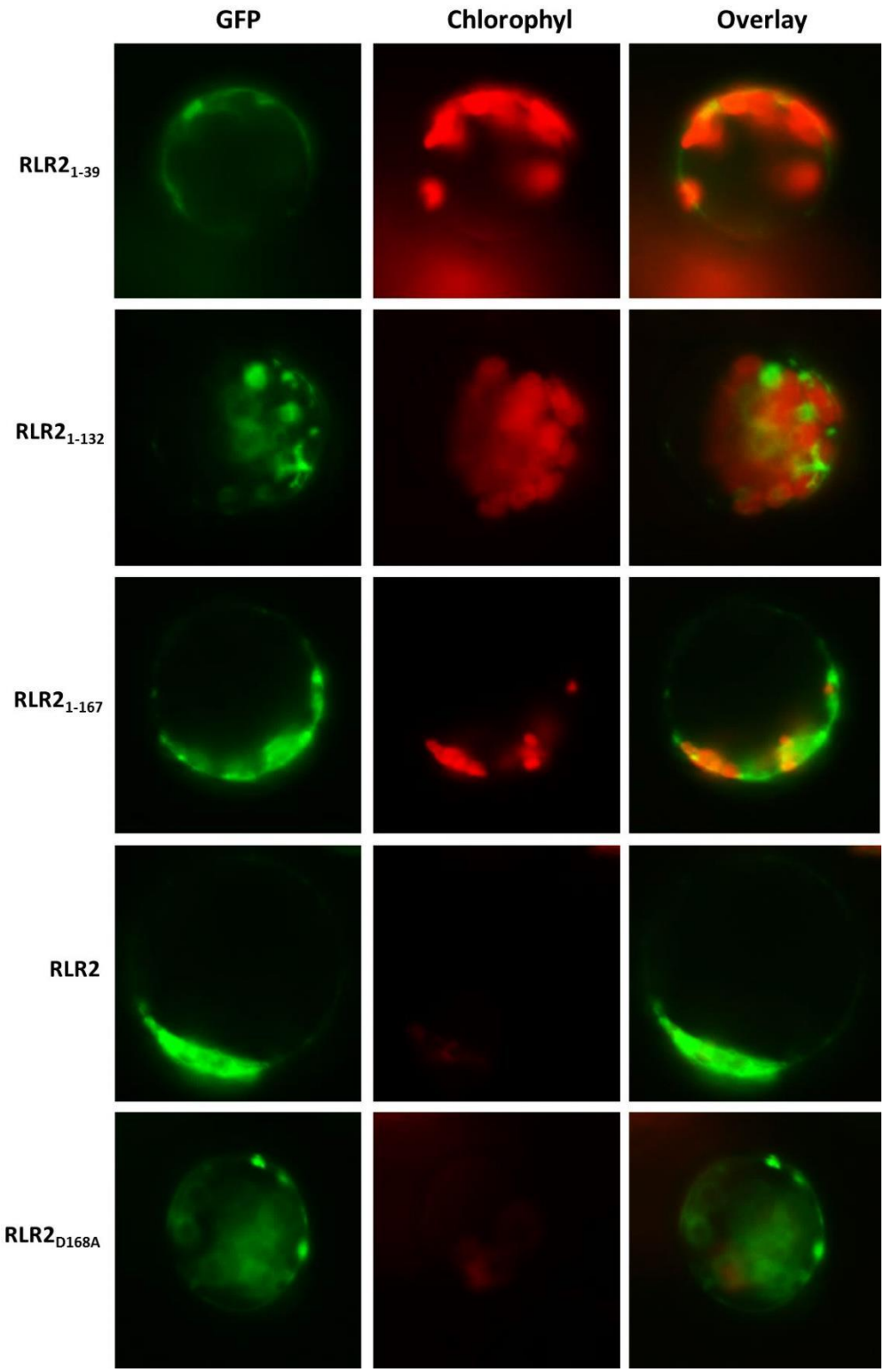


Figure 10 – RLR2 subcellular localization. Mesophyll protoplasts extracted from 3-w-old WT plants were transformed with 5-10 µg of plasmid DNA of several RLR2 constructs fused to GFP at their C-terminal. After 16 h, transfected protoplasts were observed by fluorescence microscopy: green channel for GFP and red channel for chlorophyll. **RLR2$_{1-39}$)** first 39 aa (putative signal peptide) plus 12 aa; **RLR2$_{1-132}$)** first 132 aa (putative pro peptide); **RLR2$_{1-167}$)** first 167 aa (sequence until the first catalytic aspartate); **RLR2)** full length protein; **RLR2$_{D168A}$)** RLR2 active site mutant. The signal peptide appears to be functional and RLR2 appears to be localized in the ER although we cannot discard that it continues through the secretory pathway until it is secreted to the extracellular space.

- **RLR1 *and* RLR2 *T-DNA knockout lines do not have any pollen-related phenotype***

In the first section of this chapter we confirmed that both *RLR1* and *RLR2* are being expressed in reproductive tissues and that *RLR1* is specifically expressed in mature pollen grains. As discussed, our first working hypothesis for the *in vivo* function of RLR1 and RLR2 was that they could be involved in PCD events that occur during pollen development. If this was the case, the de-regulation of these PCD events would probably lead to a male sterile phenotype very similar to the one verified in UNDEAD knockdown plants where most of the pollen grains were aborted [41]. To test this hypothesis our first approach was to study pollen viability of *RLR1* and *RLR2* KO plants and, in order to do this, we used the Alexander's staining procedure [112]. According to this methodology, we used a staining solution (Alexander solution) which is a mixture of the dyes malachite green that stains the pollen cell wall in green and acid fuchsin that stains the cytoplasm red. Since aborted pollen grains just have cell wall and no cytoplasm they will stain in green, whereas the viable ones will stain in red because they have cytoplasm which is bigger than the cell wall and makes red staining more visible [112]. Figure 11 illustrates representative results obtained for this pollen viability assay, clearly showing no difference between the WT and the KO pollen grains. These results suggest that the absence of RLR1 or RLR2 did not lead to an increase in pollen abortion.

Although we did not see any decrease in pollen viability in both *RLR1* and *RLR2* KO lines, it does not imply that the KO pollen is completely normal since pollen grains can still be viable but have a deficient germination rate. To test this hypothesis we manually pollinized WT pistils with either KO or WT pollen grains and after 5 h we collected and stained the pistils with aniline blue [115]. This dye is used to stain callose structures in plants and since these structures appear in the pollen tubes during germination, aniline blue is generally used to assess pollen germination rates [115]. The results of this *in vivo* germination rate assay are shown in Figure 12 and, again, no major differences were observed between WT and KO pollen germination. Nonetheless, the WT pollens did not germinate as well as it would be expected so we cannot withdraw any strong conclusions from these results.

At this point, there was no clear evidence for a negative impact of the lack of *RLR1* or *RLR2* in pollen development since the single KO pollen grains for both genes appear to have the same characteristics in terms of viability and germination rates as the WT pollen grains.

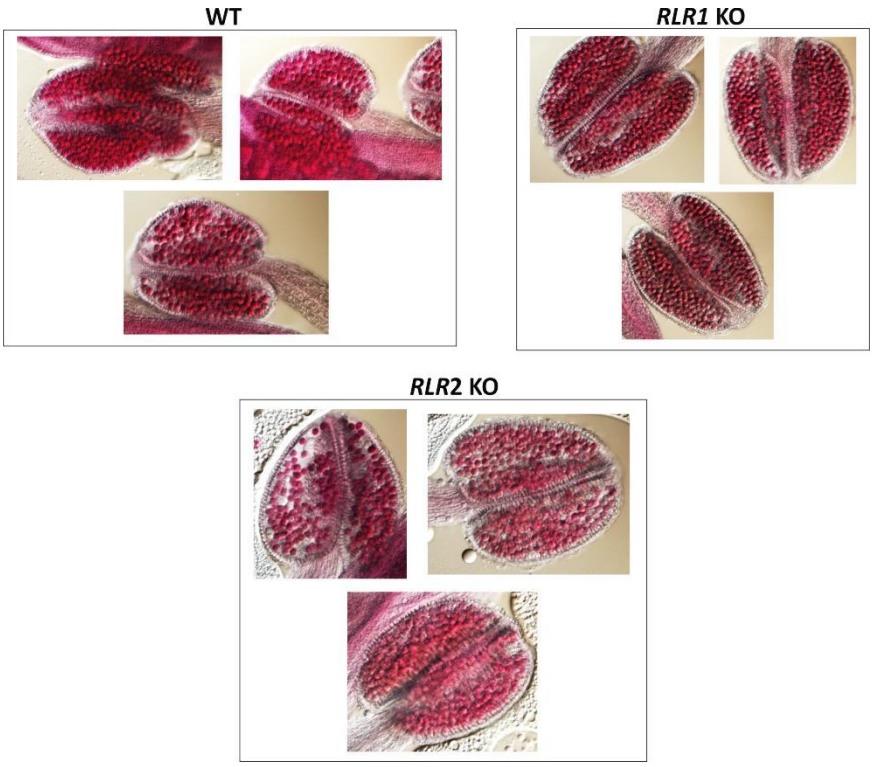


Figure 11 – Pollen viability assay. WT, *RLR1* and *RLR2* KO anthers at roughly the same growth stage were collected, de-hydrated and stained with Alexander's staining solution. Viable pollen grains stain in pink/red while aborted pollen grains stain in dark green/blue. There are no differences in pollen viability between the KO and the WT pollen grains.

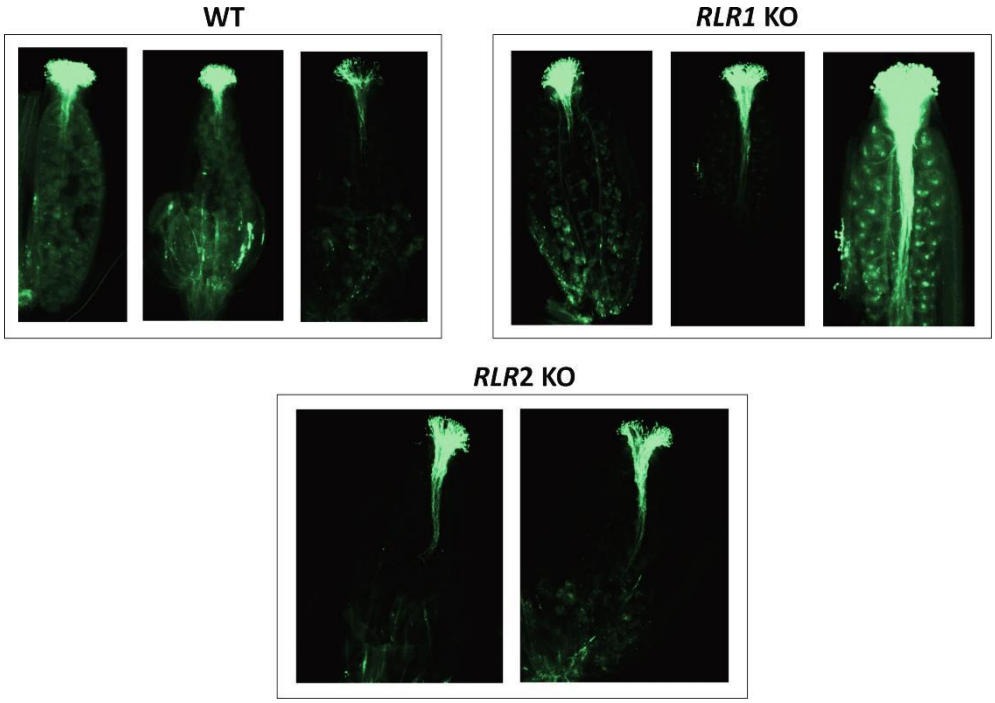


Figure 12 – *In vivo* pollen germination assay. WT pistils at roughly the same growth stage were manually pollinized with WT, *RLR1* KO or *RLR2* KO pollen grains. After a 5 h germination period, pistils were collected, stained with aniline blue and observed under the fluorescence microscope. No differences were observed between the WT and the KO pollen germination rates.

We then decided to analyze the whole flower bud morphology in order to evaluate if there would be some differences between WT and KO flowers. For each WT and KO lines we collected flower buds at different growth stages and constructed a flower development time course, comparing the KO lines with the WT. As shown in Figure 13**,** we organized the dissected flower buds from younger to older (left to right) and each column represents flower buds approximately at the same developmental stage, for comparison between the KOs and the WT. No major differences were observed between the WT and the KO lines, indicating that the absence of either *RLR1* or *RLR2* does not compromise flower development.

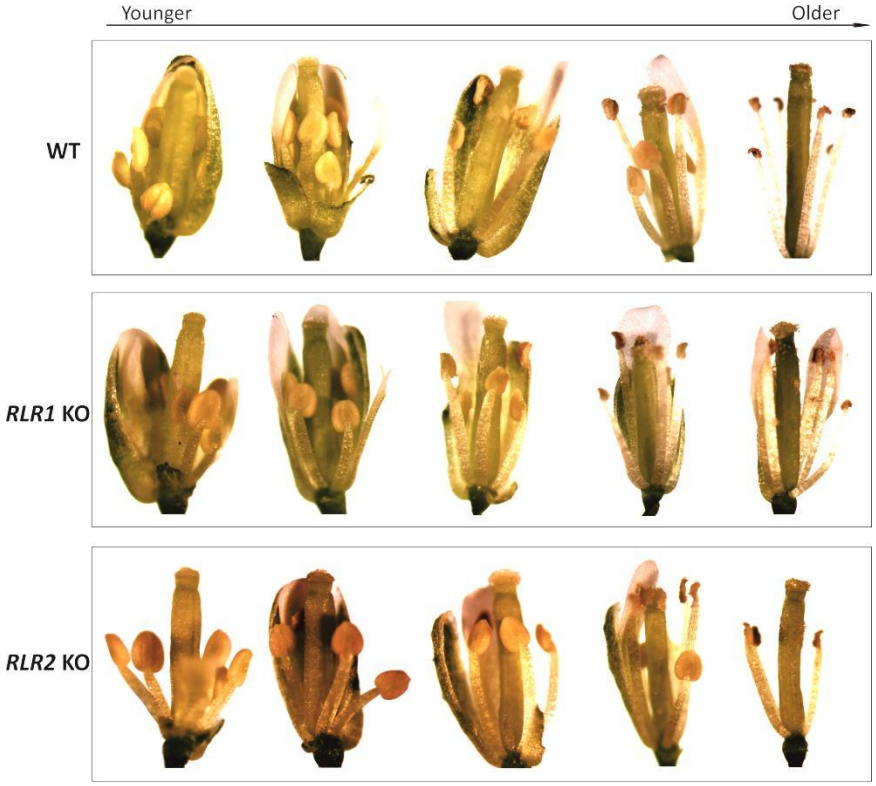


Figure 13 – Flower development time course of WT, *RLR1* and *RLR2* KO lines. Flower buds at different growth stages were collected and dissected under the dissection microscope. Images from younger to older flower buds are displayed from left to right. At each vertical column WT, *RLR1* and *RLR2* KO flower buds are approximately in the same developmental stage and so are comparable. No major differences were observed between the WT and either one of the KO lines.

Taking all these results together we can conclude that the absence of either RLR1 or RLR2 does not lead to pollen or flower abnormalities. However, this does not exclude the idea that both proteins may play a role in pollen development, just that their absence does not adequately inhibit/disable this process. Gene redundancy is an issue in *Arabidopsis* research and it may explain these results. As discussed before, there are at least 46 different atypical APs coded in the *Arabidopsis* genome [34], and some of these genes were shown to be expressed in reproductive organs [34, 37, 41] so the probability of gene redundancy events cannot be discarded for RLR1 and RLR2.

- ***RLR1 and RLR2 KO lines do not have a vegetative phenotype***

Previously we demonstrated that both *RLR1* and *RLR2* KO lines do not have any pollen or flower related phenotype. In order to evaluate if there was a vegetative phenotype, we grew both WT and KO lines under normal conditions on soil and after 6 weeks pictures were taken (Figure 14). We could not observe any difference between the WT plants and both KO lines which suggests that RLR1 and RLR2 do not seem to be essential for overall plant growth and development.

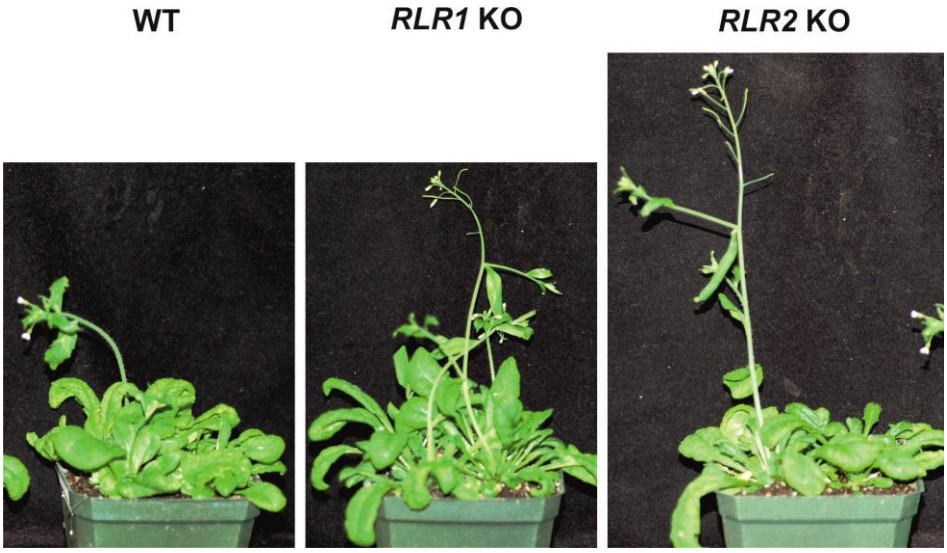


Figure 14 – *RLR1* and *RLR2* KO lines vegetative growth. Plants were grown on soil under normal conditions and pictures were taken 6 weeks after germination. No difference was observed between the WT and both KO lines.

- ***RLR1 controls primary root growth and lateral root formation under normal growth conditions.***

At the beginning of this chapter we demonstrated that both *RLR1* and *RLR2* were expressed in the floral and root tissues, being *RLR1* specifically expressed in mature pollen grains and in root tips. Because of this expression pattern we decided to expand our search for the *in vivo* functions of RLR1 and RLR2 to other developmental process: root development. To study RLR1 and RLR2 impact on primary root growth we germinated both KO and WT seeds vertically for 10 days, measuring primary root length at several time points: 3, 5, 7 and 10 days. Moreover, we aimed to analyze also the effect of RLR1 and RLR2 overexpression in root length and for that we designed several constructs, depicted in Figure 15. The goal was to generate several transgenic lines harboring *RLR1* and *RLR2* coding sequences under the regulation of CaMV 35S promoter and with different tags at their 3' terminals: GFP or HA. GFP is commonly used in these experiments as it allows the *in vivo* tracking of the protein but is also a relatively big tag which can influence the normal activity of our enzymes. Therefore, the HA-tagged transgenic lines would be an alternative strategy to avoid the possible impact of a fusion protein. Unfortunately, and due to time limitations, we were just able to generate the GFP-tagged overexpression lines which were then used in all root experiments described below.

Figure 15 – DNA constructs used for the generation of *A. thaliana* transgenic lines. RLR1 and RLR2 expression was driven by the Caulifower mosaic virus (CaMV) 35S promoter. Fused to the 3' end of *RLR1* and *RLR2* coding sequences were either the GFP coding sequence or the HA tag. Constructs were inserted in *A. tumefaciens* strain GV3101 and positive clones were used for *A. thaliana* transformation (see methods and materials). Overexpressing transgenic lines were only obtained for 35S::RLR1-GFP and 35S::RLR2-GFP.

As discussed, some atypical APs have been implicated in abiotic stress responses [38]. To start evaluating if RLR1 and RLR2 could be also implicated in this type of stress response, we decided to perform all root growth assays not only under normal conditions but also under nitrogen deprivation (an abiotic stress condition). Nitrogen is the most important inorganic nutrient for plants and root proteases have been implicated in nitrogen deprivation responses [93]. Since nitrate is the most common form of nitrogen found in normal soil [116] we decided to germinate our mutants in a media without ammonium nitrate to assess if RLR1 or RLR2 could be implicated in these type of responses. From this point forward, abiotic stress condition is always a condition of ammonium nitrate deprivation.

In this section we will present the results obtained for RLR1 and we will compare the root growth of the KO, OE and WT lines grown under normal conditions as well as under the abiotic stress condition. The results of this root length assay are depicted in Figure 16**.** Just by analysing the plate images after 10 days of growth (a) and c)) it was already obvious that both KO and OE primary roots were shorter than the WT in both conditions, although in normal conditions this was more evident. After the proper measurements at different time points (b) and d)), it was possible to confirm that KO and OE primary roots were shorter since day 5 in both growth conditions, and that this difference increased throughout the experiment. When comparing total root length at day 10 under normal and abiotic stress conditions, both KO and OE lines had shorter roots in both conditions but they could partially recover under the abiotic stress condition at almost the same rate (e)). In fact, under normal conditions *RLR1* KO and OE roots are ~26% and ~20% shorter than the WT roots, respectively; and, under abiotic stress conditions KO and OE roots are just 16% and 13% shorter, respectively (**f)**). It is also noteworthy that under this abiotic stress the WT roots also grew more when compared to the normal condition.

As described in Chapter I, lateral root formation is a very important process in root development. Since the GUS staining results showed that *RLR1* was also expressed in lateral root tips and in lateral root initiation sites, we sought to determine if RLR1 would also control lateral root formation. To this end, on the same primary root length experiments we counted the number of lateral roots at day 10 and analyzed the data in two different ways: by dividing the total number of lateral roots of each line by the number of analyzed plants or by the total root length of the analyzed plants. The results are summarized in Figure 17 and clearly demonstrate that RLR1 indeed controls lateral root formation. As expected, the results are very similar when both analysis are compared. However, taking into account the results on primary root length, it is probably more accurate to represent the results by the number of lateral root *per* cm of primary root (c) and d)). The differences between the WT and the KO or OE lines are more evident in this case. Interestingly, under normal conditions *RLR1* KO plants have 60% less lateral roots than the WT while the OE mutants have 70% less lateral roots (d)). However, under abiotic stress conditions KO mutants can recover to almost WT levels while OE recovery is more modest. It is also notheworthy that under this abiotic stress condition WT adapts by incresing the number of lateral roots (as we can see in a) and b)) but since the primary root also increases that difference is not so evident when we divide the number of lateral roots by the total root length (c) and d)).

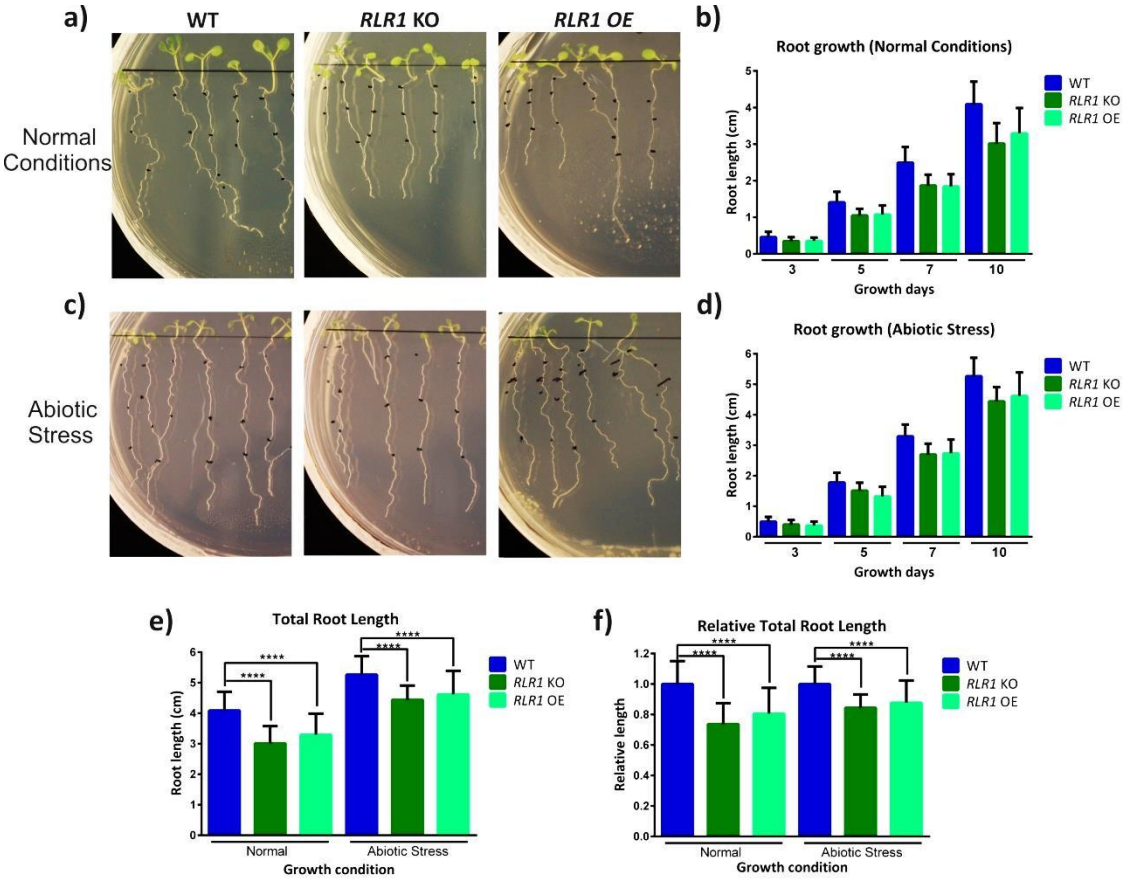


Figure 16 – Root length assays using *RLR1* KO, *RLR1* OE and WT seedlings. *RLR1* KO, *RLR1* OE and WT seedlings were grown vertically for 10 days either in normal ½ MS media (normal conditions) or in ammonium nitrate deprived media (abiotic stress). Primary root length was measured at day 3, 5, 7 and 10 (**b)** and **d)**) and at day 10 pictures were taken (**a)** and **c)**). Each experiment was repeated twice and in total 100-150 roots were measured per line. **a)** and **b)** Effect of *RLR1* absence and overexpression on primary root growth under normal conditions. **c)** and **d)** Effect of *RLR1* absence and overexpression on primary root growth under an abiotic stress condition. **e)** Total root length at the end of the experiment. **f)** Relative total root length at the end of the experiment (to each WT). Under normal and abiotic stress conditions both KO and OE lines have shorter roots than the WT and these differences are statistically significant (P values: **** <0.0001).

Taking all these results together we can conclude that RLR1 control at some extent primary root growth and, in a more pronounced way, lateral root formation. Moreover, RLR1 seems to be more important under normal growth conditions since the KO line was not affected under our abiotic stress condition. The OE results are puzzling but we have to take into consideration that we are attempting to overexpress an active protease and because of that many side effects can occur.

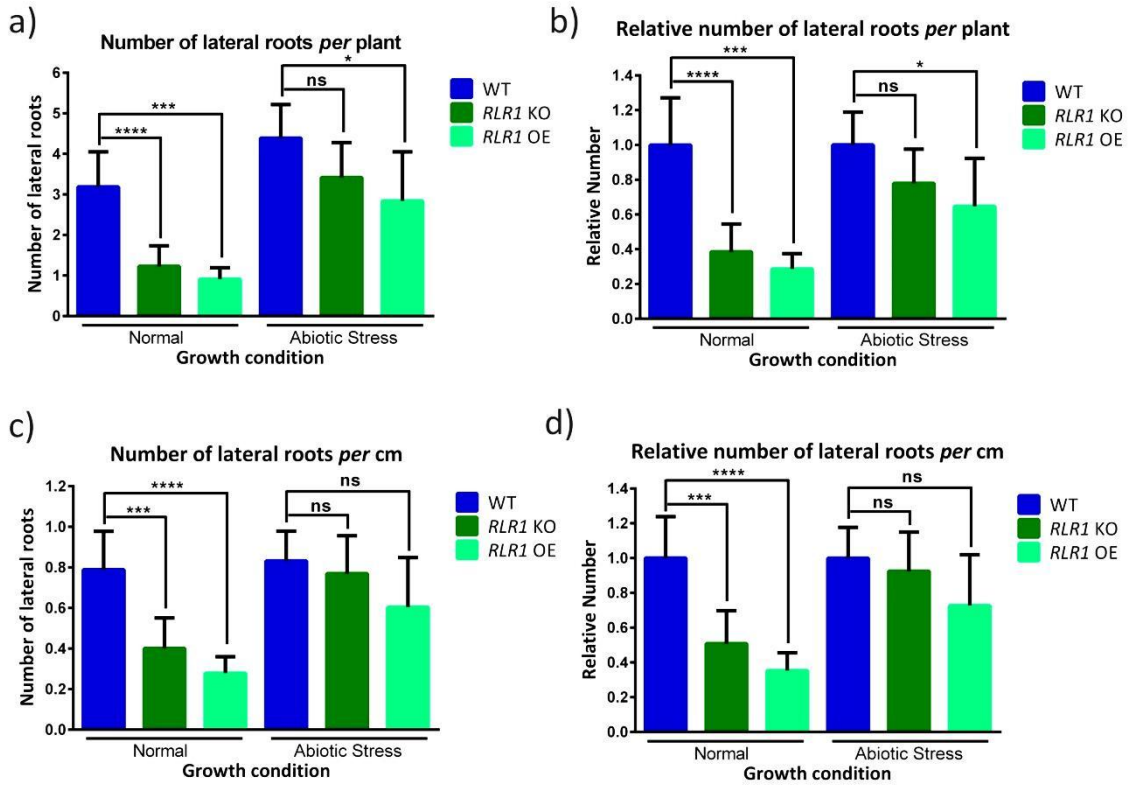


Figure 17 - Lateral root assays using *RLR1* KO, *RLR1* OE and WT seedlings. *RLR1* KO, *RLR1* OE and WT plants were grown vertically for 10 days either in normal ½ MS media (normal conditions) or in ammonium nitrate deprived media (abiotic stress) and at the 10th day lateral roots were counted under the dissection microscope. **a)** and **b)** Absolute and relative number of lateral roots *per* plant. **c)** and **d)** Absolute and relative number of lateral roots *per* cm of primary root. Under normal conditions *RLR1* KO and OE roots have 60% and 70% less lateral roots than the WT, respectively (this difference is statistically significant; P values: **** <0.0001). However, under abiotic stress condition the KO roots can recover to almost WT levels.

- ***RLR2 controls primary root growth and lateral root formation under ammonium nitrate deprived conditions***

As we described in the previous section, it was our goal to study the role of both RLR1 and RLR2 in primary root growth and lateral root formation since both proteins were expressed in root tissues. In this section we will present the results of the same primary root growth and lateral root formation experiments using *RLR2* KO and OE mutant lines. The methodology used was the same described in the previous section and the primary root growth results are depicted in Figure 18. The results under normal conditions are quite similar to those obtained for RLR1 although the RLR2 impact seems to be less pronounced under this growth condition. Indeed, *RLR2* KO and OE roots are just 18% and 10% shorter than the WT roots, respectively, when grown under normal conditions. However, under abiotic stress conditions while the OE roots can recover to almost WT levels, the KO roots are slightly shorter. In fact, when compared with the WT grown under the same condition, KO roots are 22% shorter compared to the 18 % difference observed under normal conditions.

The lateral root results are depicted in Figure 19 and again RLR2 impact is more obvious in this case. The results are expressed as the number of lateral roots *per* plant (a) and b)) or *per* cm of primary root (c) and d)). Analyzing the number of lateral root *per* cm of primary root (c) and d)), we can see that under normal conditions RLR2 KO plants have 15% fewer lateral roots than the WT although this difference is not statistically significant. However, when grown under abiotic stress conditions the KO cannot recover and, in fact, the phenotype is more severe since the KO roots have 45% fewer lateral roots than the WT. In the case of the OE mutant, under normal conditions the number of lateral roots is severely impaired but they can recover under the abiotic stress conditions.

At the end of this section we can conclude that RLR2 also regulates primary root growth but appears to be also more significant for lateral root formation. However, and unlike RLR1, RLR2 seems to be more important under the abiotic stress condition. Again, the OE results are somewhat puzzling but they tend to corroborate this conclusion since OE roots could recover the number of lateral roots under the abiotic stress condition while the KO root could not.

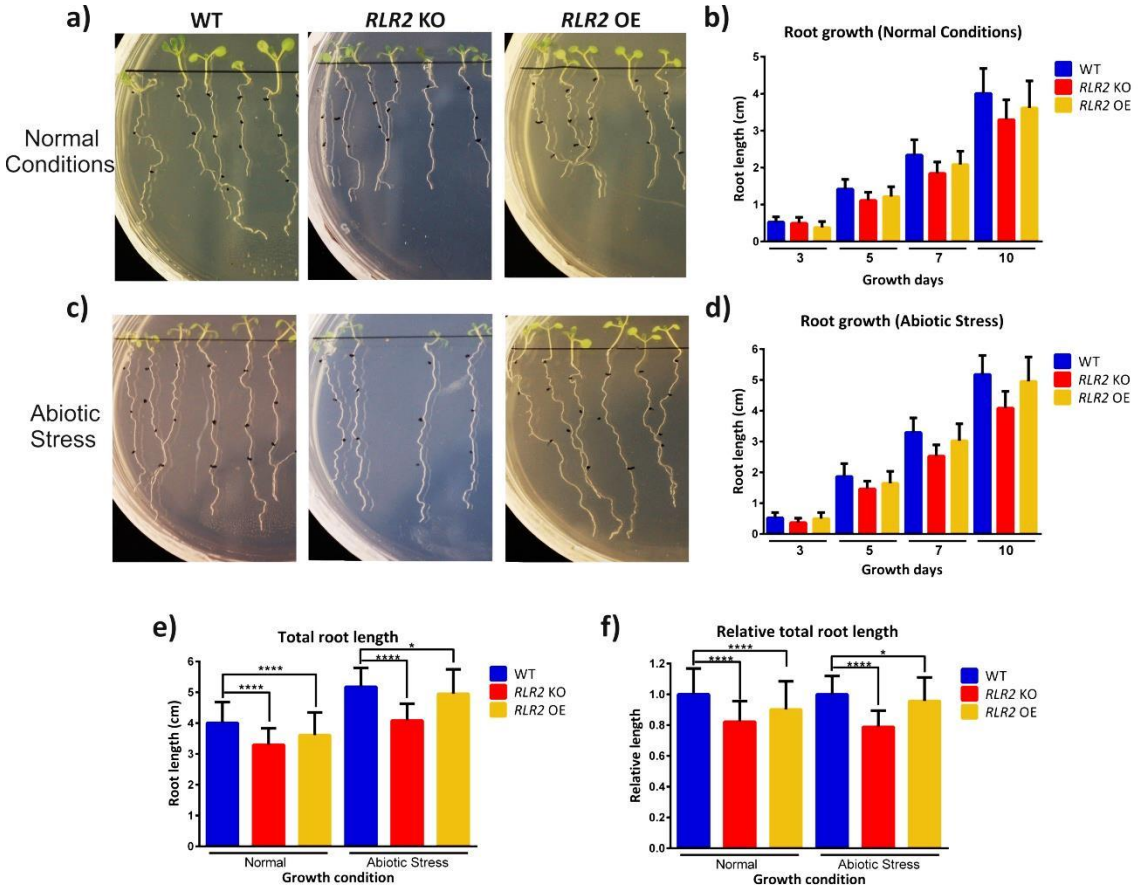


Figure 18 - Root length assays using *RLR2* KO, *RLR2* OE and WT seedlings. *RLR2* KO, *RLR2* OE and WT seedlings were grown vertically for 10 days either in normal ½ MS media (normal conditions) or in ammonium nitrate deprived media (abiotic stress). Primary root length was measured at day 3, 5, 7 and 10 (**b)** and **d)**) and at day 10 pictures were taken (**a)** and **c)**). Each experiment was repeated twice and in total 100-150 roots were measured per line. **a)** and **b)** Effect of *RLR2* absence and overexpression on primary root growth under normal conditions. **c)** and **d)** Effect of *RLR2* absence and overexpression on primary root growth under an abiotic stress condition. **e)** Total root length at the end of the experiment. **f)** Relative total root length at the end of the experiment (to each WT). Under normal and abiotic stress condition both KO and OE lines have shorter roots than the WT and these differences are statistically significant (P values: **** <0.0001).

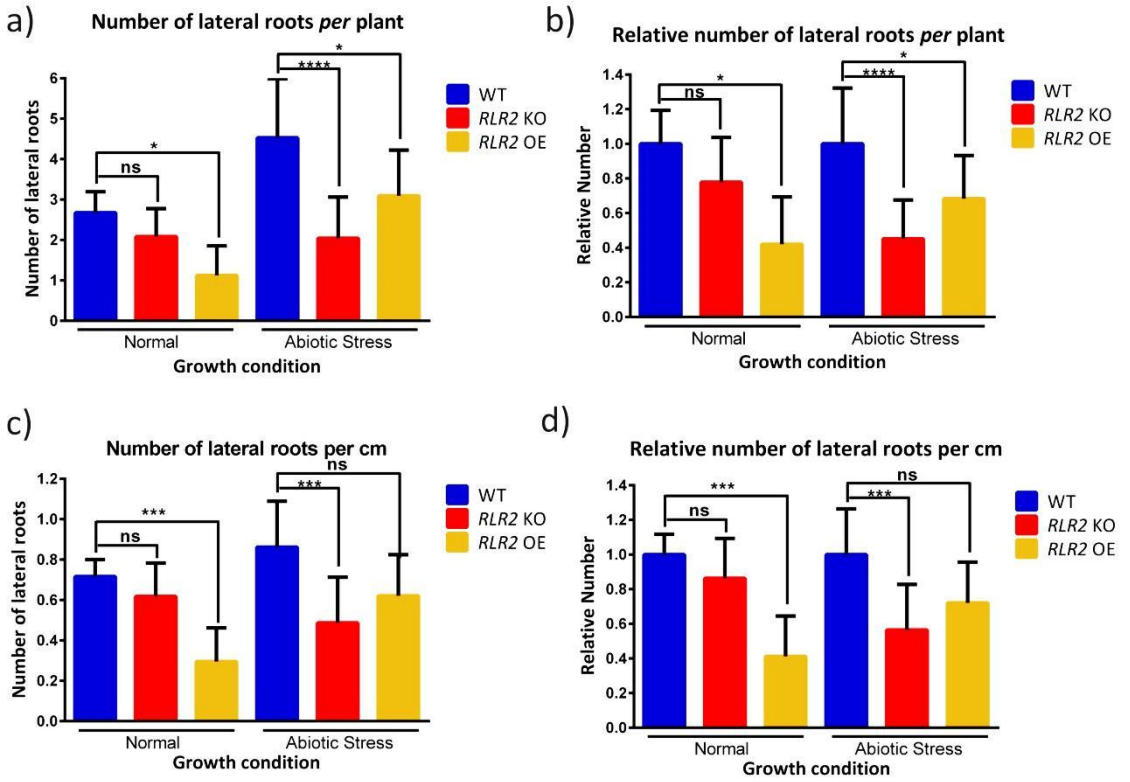


Figure 19 - Lateral root assays using *RLR2* KO, *RLR2* OE and WT seedlings. *RLR2* KO, *RLR2* OE and WT plants were grown vertically for 10 days either in normal ½ MS media (normal conditions) or in ammonium nitrate deprived media (abiotic stress) and at the 10th day lateral roots were counted under the dissection microscope. **a)** and **b)** Absolute and relative number of lateral roots *per* plant. **c)** and **d)** Absolute and relative number of lateral roots *per* cm of primary root. Under normal conditions *RLR2* KO is not significantly different than the WT while OE roots have 60% less lateral roots than the WT (this difference is statistically (P values: **** <0.0001, *** <0.001). However, under abiotic stress condition the KO roots cannot recover while the OE can.

- ***Both KO lines are more sensitive to the plant hormone auxin***

Hormones play crucial roles in a variety of plant processes and they are especially important in plant developmental processes, either during shoot or root development [117]. Amongst all plants hormones, auxin (indole-3-acetic acid, IAA) is one of the most important, if not the most important, and its crucial role during root development and lateral root formation is well documented [118, 119]. Given its importance for root development and because there is still much to be uncovered about auxin signaling pathways, we decided to evaluate if RLR1 and RLR2 role in root development and lateral root formation was connected to auxin signaling. To do that, both KO and WT seeds were germinated under normal conditions (½ MS media), transferred after 4 days to ½ MS plates supplemented with different IAA concentrations and left to grow under those conditions for 5 days. At this time point, primary root length was measured and the number of lateral roots was counted.

It is well documented that increasing auxin concentrations impair primary root growth in WT plants [119] and, as expected, our results reproduce this phenotype (Figure 20) with increasing concentrations of auxin leading to a decrease in primary root length in WT seedlings. Moreover, a similar result was observed for both KO lines and when we normalized the root length values to their respective control (each line grown without auxin) it was also clear that WT, *RLR1* and *RLR2* KO lines respond to auxin at the same extent, clearly suggesting that the auxin sensing mechanisms are not impaired in *RLR1* or *RLR2* KO lines in respect to primary root growth.

As discussed, auxin is also very important for lateral root formation and it is reported that in WT roots the number of lateral roots increases with increasing auxin concentrations [118]. In Figure 21 are depicted the results of the number of lateral roots for the auxin experiments and panels a) and c) clearly show that WT and both KO lines respond to auxin by increasing the number of lateral roots. However, when we normalized the lateral root values to each control condition we can clearly see that the response to increasing auxin concentrations is much more dramatic in *RLR1* KO line, although *RLR2* KO line also responds with an increase in the number of lateral roots when compared to the WT (b) and d)). In fact, if we consider the relative number of lateral roots *per* cm of primary root, *RLR1* KO line displayed approximately 70x more lateral roots when grown under 500 nM auxin when compared to the control condition (no auxin), whereas the WT plants just had 15x more lateral roots. *RLR2* KO line also responded more than the WT since it presented 40x more lateral roots under 500 nM auxin than its control condition. These results clearly demonstrate that the absence of both

RLR1 and RLR2 makes the plants more sensitive to auxin treatment when it comes to lateral root formation.

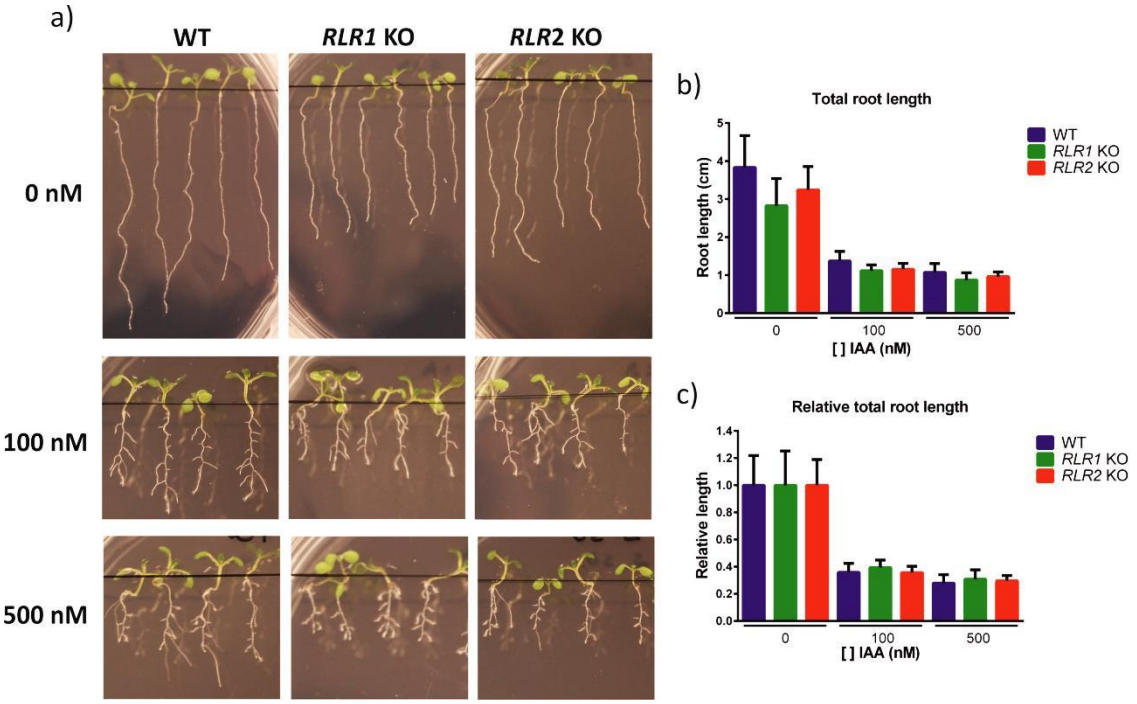


Figure 20 - Root length assays using *RLR1* KO, *RLR2* KO and WT seedlings grown under different auxin concentrations. *RLR1* KO, *RLR2* KO and WT seedlings were grown for 4 days under normal conditions (½ MS media) and then transferred to ½ MS plates supplemented with different IAA (auxin) concentrations (0, 100 and 500 nM). They were allowed to grow vertically for 5 days, pictures were taken and primary root length was measured. Each experiment was repeated with different batches of seeds and in total 20-60 roots were measured per line. **a)** Auxin effect on primary root growth after a 5 day treatment with different IAA concentrations. **b)** Total root length at the end of the experiment. **c)** Relative total root length at the end of the experiment (to each corresponding line grown without IAA). Auxin treatment affects both the KO lines and the WT primary root growth at the same extent.

Taking all these results into consideration, we conclude that the absence of either RLR1 or RLR2 does not appear to significant affect auxin response in what concerns primary root growth, suggesting that the role(s) of RLR1 and RLR2 in this case may be independent of auxin signaling. However, when it comes to lateral root formation RLR1 or RLR2 absence resulted in a significant increase in the auxin response which may anticipate a role for both proteases as negative regulators of auxin in the control of this process.

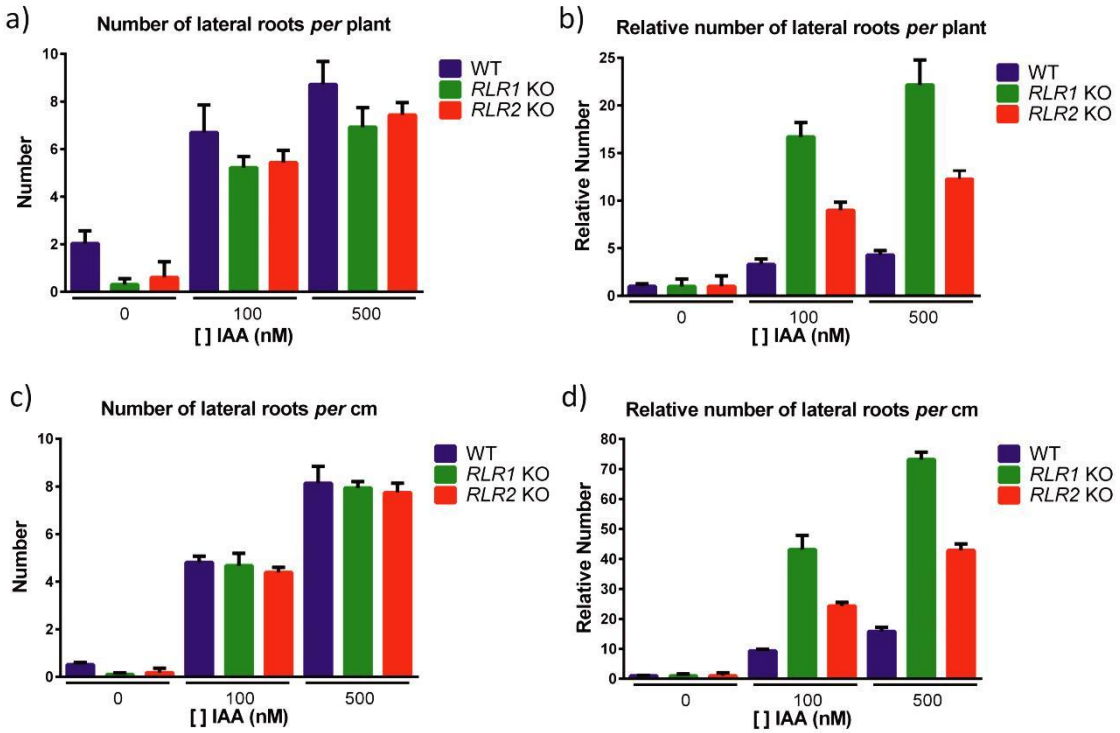


Figure 21 – Lateral root assays using *RLR1* KO, *RLR2* KO and WT seedlings grown under different auxin concentrations. *RLR1* KO and *RLR2* KO and WT seedlings were grown normally for 4 days under normal conditions (½ MS media) and then transferred to ½ MS plates supplemented with different IAA (auxin) concentrations (0, 100 and 500 nM). They were allowed to grow vertically for 5 days and the number of lateral roots was counted at the end of the experiment. Each experiment was repeated with different batches of seeds and in total 20-60 roots were measured per line. **a)** and **b)** Absolute and relative number of lateral roots *per* plant. **c)** and **d)** Absolute and relative number of lateral roots *per* cm of primary root. Both *RLR1* and *RLR2* KO lines respond more to auxin treatment than the WT line, although *RLR1* KO response is more drastic that that of *RLR2* KO lines.

2.4 Discussion

The goal of this chapter was to describe the *in vivo* funtions of two novel atypical APs that we named RLR1 and RLR2 due to their involvement in lateral root development. As we discussed before, our first working hypothesis was that both RLR1 and RLR2 would be involved in PCD events that control pollen development and this hypothesis was based on previous reports that, although not specific for these proteins, showed a de-regulation of both of these genes associated with a male sterility phenotype [41, 85]. Ito and co-workers showed that the expression of *RLR1* (Uniprot accession number Q84M99, gene ID At2g03200) is regulated by the *MALE STERILITY 1* transcription factor, which is expressed in the tapetum cells after meiosis and is very important for post meiotic pollen development. The absence of this transcription factor resulted in a severe male sterility phenotype and the observed decrease in *RLR1* levels could somehow contribute to this phenotype [85]. In the case of *RLR2* (Uniprot accession number Q9SZV7, gene ID At4g30040), this gene appeared in the same screening where the UNDEAD protein was found and, although not directly regulated by MYB80 transcription factor (like UNDEAD), a 9 fold increase in its expression was observed when MYB80 was upregulated [41]. Since UNDEAD expression is just 14 fold increased under the same conditions, we initially hypothesized that RLR2 could also play an important role during pollen development.

Our first approach was to determine in which tissues *RLR1* and *RLR2* were being expressed and the results corroborated our initial working hypothesis since both genes were expressed in reproductive tissues, being *RLR1* specifically expressed in mature pollen grains. Using T-DNA KO lines we looked for a pollen specific phenotype linked to the absence of either RLR1 or RLR2 but we failed to detect any pollen or flower phenotype. Nonetheless, these results do not necessarily mean that RLR1 and RLR2 do not play a role during pollen development; they just indicate that the absence of either one of these proteins does not disable the process. Gene redundancy is often a huge issue in plant research and it can be the reason underlying our lack of a pollen-related phenotype. As we already discussed, there are at least 46 different atypical APs encoded in the *Arabidopsis* genome [34] so redundancy can really hamper the identification of some phenotypes. Also, we cannot rule out the possibility that both proteases generate phenotypes in male or female gametophytes that are only exacerbated under specific conditions, e.g. drought or heat to which pollen is highly sensitive, that only a more thorough analysis would allow us to dissect in detail.

Since we could not confirm our first hypothesis we decided to expand the scope of our study to other plant processes. The GUS staining and the RT-PCR results showed that both

RLR1 and *RLR2* were also being expressed in root tissue and interestingly *RLR1* was being expressed specifically in root tips and in lateral root formation sites. Therefore, to further evaluate if RLR1 and RLR2 would play a role in primary root growth and/or lateral root formation, we decided to grow KO and WT seedlings vertically for 10 days and measure primary root length at different time points, as well as assess the number of lateral roots at the end of the experiment. Under normal growth conditions it was clear that KO lines had shorter primary roots, suggesting that both proteins appear to control primary root elongation to some extent. However, the more drastic phenotype was observed in lateral root formation. We have clearly demonstrated that, under normal growth conditions, the absence of *RLR1* had more impact than the absence of *RLR2* in the reduction of the number of lateral roots, anticipating a more critical role of the former protease for lateral root formation under normal conditions. Intriguingly, the OE lines had the same phenotype as the KO lines although *RLR2* OE showed also a more dramatic decrease in lateral root formation. At this point, we cannot really explain these OE phenotypes but, as previously discussed, it is important to bear in mind that the overexpression of proteases (whose expression is normally tightly regulated) may result in the deregulation of different pathways that, ultimately, result in deregulated phenotypes. In fact, if we look to some of the published evidences on the function of APs, some of the reported phenotypes resulted from the OE of the protease (and not from its absence) [36-38, 42, 47]. Here we show that the deregulation of these proteases result in both cases in an impact in root/lateral root phenotypes.

As discussed before, there is one atypical AP, ASPG1, that has been implicated in abiotic stress responses, more specifically in drought stress [38]. In our study we also wanted to start evaluating if RLR1 and RLR2 could be involved in abiotic stress responses so we germinated our KO lines under nitrogen deprivation conditions. For primary root growth, the results were quite similar to the normal growth conditions. However, when considering the number of lateral roots, the *RLR2* KO showed a significant reduction when compared to the WT whereas the *RLR1* KO line had an almost WT phenotype. These results suggest that RLR2 may play a role in the control of lateral root formation under this N-deprived condition, while the absence of RLR1 does not seem to have an impact on this process.

The role of root proteases in nitrogen uptake is still not that well understood. However, in the last few years there were a number of studies that proved the importance of root proteases in the regulation of this process [93]. Under nitrogen limitations root proteases were found to be important in protein turnover promoting protein breakdown to facilitate nitrogen uptake in form of amino acids [93]. Interestingly, Godlewski *et al.* reported that 15 plant species secrete proteases from roots and that secretion was part of the plant strategy to

increase nitrogen uptake; unfortunately the nature of these specific proteases was not identified [120]. Our own results suggest that RLR1 and RLR2 are ER proteins but we cannot exclude the hypothesis that the ER is just a step on their way through the secretory pathway to the extracellular milieu, as both proteins lack the canonical ER retention signal KDEL [121]. Thus, RLR2 secretion might occur under ammonium nitrate deprived condition as an attempt to cope with nitrogen deprivation but this hypothesis awaits future confirmation. However, that does not explain RLR2 control of lateral root formation under this stress condition and since there are no other proteases reported to be involved in lateral root control, we need to perform other studies to dissect this mechanism.

In an attempt to further characterize the molecular mechanisms underlying RLR1 and RLR2 role in primary root growth and lateral root formation, we decided to test if these proteins could be linked with the plant hormone auxin. As we already discussed, auxin is the major regulator of several plant processes and its action - sometimes in coordination with other plant hormones such as cytokinin (CK) and brassinosteroids (BR) - is responsible for primary root, lateral root and root hair development [119]. By germinating both KO and WT seedlings in normal media supplemented with several auxin (IAA) concentrations, we observed that both KO lines respond to auxin at the same rate as WT plants in what concerns primary root growth. This suggests that the lack of either RLR1 or RLR2 does not impair auxin signaling, which may indicate that the control of primary root growth by both proteases is auxin-independent. Interestingly, D'Alessandro *et al.* described very recently a KO mutant with a root phenotype similar to *RLR1* and *RLR2* KO lines which was also responsive to external auxin treatment [122]. In this case, it was a KO mutant of the p23 gene, a homolog of the animal HSP90 chaperonin and, although the mutant was responsive to auxin treatment, auxin distribution was altered in the KO, showing low auxin levels [122]. Cell-to-cell auxin transport is achieved through a combination of specific auxin influx (LAX) and efflux (PIN) carriers and this p23 KO mutant was shown to display low expression and miss-localization of several PIN transporters which explained the low auxin levels [122]. From these results, authors concluded that p23 absence could compromise the stability of HSP90 complex which would lead to alterations in the auxin transport system [122]. In our case, the root auxin levels were not measured in *RLR1* and *RLR2* KO mutants and, therefore, it is not possible to speculate if a similar mechanism is occurring in this case. Nevertheless, it would be interesting to evaluate in future assays if RLR1 and RLR2 may function on primary root growth also by controlling auxin transport/distribution.

In terms of lateral root control, our results showed that upon auxin treatment the increase in lateral root number was much more pronounced in both KO lines than in the WT,

being the *RLR1* KO line the most responsive one. This could mean that both proteins are negative regulators of auxin signaling when it regards to lateral root formation. Inside cells, Indole-3-acetic acid (IAA) is the major auxin form and one of the best characterized mechanisms of auxin signaling in root development is through ubiquitin-dependent proteolysis [119]. In this system, two types of protein families interact leading to the transcription or repression of auxin responsive genes: Auxin/indol-3-acetic acid (Aux/IAA) and Auxin Response Factors (ARF). ARF are transcription factors that either repress or promote the expression of several genes (depending on the ARF proteins) in the presence of auxin. In the absence of auxin, Aux/IAA is bound to ARF and impairs its activity but in the presence of auxin, Aux/IAA are targeted for degradation (*via* ubiquitin-proteosome pathway) and ARFs are release to perform their functions [119]. There are ARF proteins, such as ARF7 and ARF19, that specifically regulate genes involved in lateral root formation [123] and our two atypical APs might function as negative regulators in their signaling pathway, for example. As we discussed before, other hormones act with auxin during root development and CK is one of the most important ones. CK and auxin are antagonists in root development since CK promotes cell differentiation and root elongation, while auxin is essential for maintaining the undifferentiated state of cells of the stem cell niche in root apical meristem [119, 124]. Concerning lateral root formation, CK and auxin also play antagonistic roles. CK inhibits lateral root development by activating the transcription of the SHY2 gene which negatively regulates the PIN transporters while auxin facilitates the degradation of the SHY2 protein [125]. This means that the CK pathway negatively regulates the auxin pathway and thus RLR1 and RLR2 might also be positive regulators of the CK pathway or be part of the CK pathway. However, our results so far just allow us to speculate about RLR1 and RLR2 connection with auxin during lateral root formation and more studies will need to be performed to describe the RLR1 and RLR2 molecular pathways.

Although we were not able to observe a pollen/anther development phenotype for RLR1 and RLR2, our results clearly demonstrate a role of both proteases in root development. Indeed, their impact in lateral root formation seams to be the strongest, thereby justifying the proposed name which stands for Regulator of Lateral Root (RLR). To our knowledge, RLR1 and RLR2 are the first atypical APs to be implicated in these phenomena of primary root development and lateral root formation. The role of proteases during root development remains mostly elusive [93]. With this work we provide evidence for the role of two atypical APs as important regulators of lateral root growth, which we hope will open new research avenues to further explore the molecular pathways responsible for this regulation and to

extend our knowledge on the relevance of proteolysis in plant developmental processes and/or adaptation to stress conditions.

Chapter III – *In vitro* Biochemical Studies

3.1 Introduction

Until nowadays, the study of plant atypical and nucellin-like APs has been focused on their *in vivo* functional characterization and virtually nothing is known about their biochemical properties and substrate specificity. Although it is very important to know the *in vivo* functions of these proteins, the study of their structural and biochemical properties gives us valuable knowledge that ultimately helps to understand structure-function relationships. Indeed, by studying the biochemical properties we can assess if the protein is indeed an active enzyme with AP properties, we can describe its pH and inhibition profile, which will give us important clues about their possible intracellular localization and mechanism of action, and we can study its specificity profile which may also give us important information about their stringency and possible substrates. These are just a few examples of the information that we can acquire by studying the biochemical properties of these proteases but much more can be done concerning this subject.

Despite the importance of biochemical studies, until now there are just a few atypical and nucellin-like APs for which a biochemical characterization has been reported. In fact, out of the 13 atypical and nucellin-like APs characterized so far, only two were subjected to a thorough biochemical characterization: *Arabidopsis* CDR1 and *Nicotiana* CND41 [53, 59]. CDR1 biochemical characterization was performed in our laboratory and for that Simões *et al.* produced the recombinant form of this enzyme (rCDR1) using *E. coli* as the expression host and in the form of inclusion bodies [53]. After a lengthily refolding and purification procedure, the authors could obtain active enzyme that showed some remarkable properties: it was mostly active at pH 6, high above the normal optimal pH for an AP, it was greatly inhibited by redox agents and, surprisingly, not much sensitive to pepstatin A, the canonical AP inhibitor [53]. Moreover, its substrate specificity was unusually restrictive since the authors had to design a new fluorogenic substrate specifically for its biochemical characterization [53]. In the case of CND41, Murakami *et al.* isolated this enzyme form the chloroplast of *N. tabacum* cultured cells [59]. They reported that this enzyme was mostly active at pH 2.5 and that it was completely insensitive to pepstatin A [59]. CND41 biochemical characterization was not as extensive as CDR1 but we can clearly see in both cases that these two APs have some distinguishing biochemical properties which emphasize the importance of this kind of studies.

The lack of knowledge regarding the biochemical properties of these enzymes can be linked at some extent to their unusual primary structure organizations that can impair, or at least difficult their production in the recombinant form. Faro *et al.* reported that most of these

enzymes have a high number of cysteines in their amino acid sequence and that some of them have an unusually long pro-segment with unknown functions [34]. To date, the organism of choice to produce recombinant APs (plant and non-plant alike) is the bacteria *E. coli* [3, 4, 53, 126] and although it has several advantages, these bacteria cannot perform post-translational modifications such as disulfide bond formation in their cytoplasm [127]. This can be a major limitation since, as we discussed, atypical APs have a high number of cysteine residues in their amino acid sequence. Moreover, the strategy mostly used to obtain recombinant APs has been to produce them in the form of inclusion bodies and promote their *in vitro* refolding afterwards [53, 126]. This is a lengthily procedure and although it was proved to be suited for rCDR1, it may not be the most suitable strategy for other atypical APs. Thus, new approaches must be attempted in order to facilitate the production of these enzymes which will then allow their biochemical characterization in much more detail.

In the previous chapter we described RLR1 and RLR2 *in vivo* functions and one of our goals was also to study the biochemical properties of both proteases. In this chapter, we will describe the production of the recombinant forms of RLR1 and RLR2 (rRLR1 and rRLR2) as well as the biochemical characterization of rRLR1. Our first approach to produce both recombinant proteins was to use *E. coli* as the expression host but instead of producing the proteins as inclusion bodies, we tried to produce them in their soluble form in *E. coli* cytoplasm. Unfortunately, that strategy was not viable and we ended up by producing both recombinant proteins in *N. benthamiana* leaves using a powerful viral-based expression system called magnICON®. We successfully produced both proteins but we could only purify active rRLR1 which showed several interesting biochemical properties that will be described throughout this chapter.

3.2 Methods and Materials

- ***Plasmid DNA constructs***

    - ***Plasmid DNA constructs used in E. coli small scale expression screening***

Commercial vector pET23a (Novagen) was the expression vector used in all *E. coli* experiments. This vector has a T7 promoter and a T7 terminator as well as a histidine tag at the 3' end of the multiple cloning site allowing the production of histidine tagged proteins. *E. coli* codon optimized *RLR1* and *RLR2* coding sequences (GeneScript) lacking the putative signal peptide were inserted into pET23a with Fh8 tag at the 3' end giving rise to the constructs RLR1-Fh8-His and RLR2-Fh8-His (see Table 8). Both constructs were generated using Circular Polimerase Extension Cloning (CPEC) [128]. Briefly, RLR1, RLR2 and Fh8 cDNA were amplified in separate PCR reactions using the CPEC primers listed in Table 9 while pET23a was linearized with the restriction enzymes *Nde*I and *Xho*I. Two additional PCR reactions were carried out with different combinations of template DNA: RLR1 + Fh8 + linearized pET23a and RLR2 + Fh8 + linearized pET23a which gave rise to the final constructs RLR1-Fh8-His and RLR2-Fh8-His inserted into pET23a expression vector. The remaining constructs listed in Table 8 were obtained by traditional cloning techniques. PCRs were made using either RLR1-Fh8-His (for RLR1 constructs) or RLR2-Fh8-His (for RLR2 constructs) as DNA templates and the primers listed in Table 9. The PCR amplified products were digested with *Nde*I and *Xho*I restriction enzymes and cloned into pET23a, previously digested with the same enzymes. All positive clones were selected by restriction analysis and confirmed by DNA sequencing.

Table 8 - DNA constructs used for *E. coli* small scale expression screening.

Constructs	Description	Primer Combination
RLR1-Fh8-His	RLR1 without its putative pre (first 27 aa) fused to Fh8Tag and HisTag at its C-terminal.	-
RLR1-His	RLR1 without its putative pre (first 27 aa) fused to a HisTag at its C-terminal.	RLR1_for/RLR1_2_rev
ΔP_RLR1-Fh8-His	RLR1 without its putative prepro (first 90 aa) fused to Fh8 tag and HisTag at its C-terminal.	ΔP_RLR1_for/Fh8_rev
ΔP_RLR1-His	RLR1 without its putative prepro segment (first 90 aa) fused to a HisTag at its C-terminal.	ΔP_RLR1_for/RLR1_2_rev
RLR2-Fh8-His	RLR2 without its putative pre (first 26 aa) fused to Fh8 tag and HisTag at its C-terminal.	-
RLR2-His	RLR2 without its putative pre (first 26 aa) fused to a HisTag at its C-terminal.	RLR2_for/RLR1_2_rev
ΔP_RLR2-Fh8-His	RLR2 without its putative prepro (first 132 aa) fused to Fh8 tag and HisTag at its C-terminal.	ΔP_RLR2_for/Fh8_rev
ΔP_RLR2-His	RLR2 without its putative prepro (first 132 aa) fused to an HisTag at its C-terminal	ΔP_RLR2_for/RLR1_2_rev

Pre - Signal Peptide; Pro - Pro segment

Table 9 – Primers used to generate the PCR products used for *E. coli* small scale expression screening.

Name	Sequence
RLR1_CPEC_for	CTT TAA GAA GGA GAT ATA CAT ATG CGT CGT AGC CTG ATT GAT
RLR1_CPEC_rev	TGG ACC TTG AAA CAA AAC TTC TAA CAG TTT ACC ACA TTC CGT CGG
RLR2_CPEC_for	CTT TAA GAA GGA GAT ATA CAT ATG TCA CCG ACG CTG GTT CT
RLR2_CPEC_rev	TGG ACC TTG AAA CAA AAC TTC TAA AAA AGA CAC TTC CAT TGC TTC C
Fh8_CPEC_for	TTA GAA GTT TTG TTT CAA GGT CCA ATG CCT AGT GTT CAA GAG GTT
Fh8_CPEC_rev	GTG GTG GTG GTG GTG CTC GAG TGA TGA CAA AAT CGA AAC GAG T
RLR1_for	CAT ATG CGT CGT AGC CTG ATT
RLR2_for	CAT ATG TCA CCG ACG CTG GT
RLR1_2_rev	CTC GAG TGG ACC TTG AAA CAA AAC TTC TA
ΔP_RLR1_for	CAT ATG GAT GAC ACG AAC AAT ATT AAA GC
ΔP_RLR2_for	CAT ATG TAT ACC CAT CGC AAT GAA ACC T
Fh8_rev	CTC GAG TGA TGA CAA AAT CGA AA

In grey are highlighted the restriction site sequences.

○ *Plasmid DNA constructs used for protein expression in N. benthamiana*

For *N. benthamiana* expression, all constructs listed in Table 10 were inserted in magnICON® expression vector [129] using traditional cloning techniques. Two variants of this vector were used: A1TMVα and A2PVXα. The only difference between the vectors is the promoter that drives transgene expression: A1TMVα has the Tobacco Mosaic Virus Promoter while A2PVXα has the Potato virus X promoter. *RLR1* and *RLR2* cDNA sequences, previously isolated from *Arabidopsis* mRNA, were used as template for the PCRs that gave rise to the constructs listed in Table 10 and the primers used for these PCR reactions are listed in Table 11. The PCR amplifications were designed to introduce the restriction sites used for the cloning procedure and one of the two different expression tags tested: HisTag and StrepTag, both at the 3'-end. The amplified products were inserted in the magnICON® using *BsmB*I restriction enzyme for *RLR1* constructs and *Bsa*I restriction enzyme for *RLR2* constructs. All positive clones were selected by restriction analysis and confirmed by DNA sequencing.

Table 10 - DNA constructs used for protein expression in *N. benthamiana.*

Constructs	Description	Primer Combination
RLR1-Strep	RLR1 without its pre (first 27 aa) fused to the StrepTag at its C-terminal.	RLR1_for/RLR1_Strep_rev
RLR1-His	RLR1 without its pre (first 27 aa) fused to the HisTag at its C-terminal.	RLR1_for/RLR1_His_rev
ΔP_RLR1-Strep	RLR1 without its prepro (first 90 aa) fused to the StrepTag at its C-terminal.	ΔP_RLR1_for/RLR1_Strep_rev
ΔP_RLR1-His	RLR1 without its prepro (first 90 aa) fused to the HisTag at its C-terminal.	ΔP_RLR1_for/RLR1_His_rev
RLR2-Strep	RLR2 without its pre (first 26 aa) fused to the StrepTag at its C-terminal.	RLR2_for/RLR2_Strep_rev
RLR2-His	RLR2 without its pre (first 26 aa) fused to the HisTag at its C-terminal.	RLR2_for/RLR2_His_rev
ΔP_RLR2-Strep	RLR2 without its prepro (first 132 aa) fused to the StrepTag at its C-terminal.	ΔP_RLR2_for/RLR2_Strep_rev
ΔP_RLR2-His	RLR2 without its prepro (first 132 aa) fused to the HisTag at its C-terminal.	ΔP_RLR2_for/RLR2_His_rev

Pre - Signal Peptide; Pro - Pro segment

Table 11 – Primers used to generate the DNA products used for *N. benthamiana* expression.

Name	Sequence
RLR1_for	GCA CGTCTCAAGGT AGA AGA TCA TTA ATC GAC CG
ΔP_RLR1_for	GCA CGTCTCAAGGT GAT GAT ACT AAT AAC ATT AAA G
RLR1_Strep_rev	GCA CGTCTCAAAGC CTA CTT TTC GAA CTG CGG ATG GCT CCA ACC TCC CAA TTT TCC ACA TTC AGT GG
RLR1_His_rev	GCA CGTCTCAAAGC CTA ATG GTG ATG GTG ATG GTG CAA TTT TCC ACA TTC AGT GG
RLR2_for	GCA GGTCTCAAGGT TCA CCA ACT CTT GTT CTC AAT
ΔP_RLR2_for	GCA GGTCTCAAGGT TAT ACT CAC CGC AAC GAA AC
RLR2_Strep_rev	GCA GGTCTCAAAGC CTA CTT TTC GAA CTG CGG ATG GCT CCA ACC TCC AAA AGA AAC CTC CAT TGC TTC
RLR2_His_Rev	GCA GGTCTCAAAGC CTA ATG GTG ATG GTG ATG GTG AAA AGA AAC CTC CAT TGC TTC

In grey - restriction site sequence; in black - linker sequences; in green - StrepTag coding sequence; in blue – HisTag coding sequence

- ***E. coli small scale expression screening***

Several variables were tested in order to find the condition with the highest amount of soluble recombinant protein (variables listed in Table 12). We also tested the pre-expression of a sulfhydryl oxidase (Erv1p) and a disulfide isomerase (DsbC) in *E. coli* cytoplasm prior rRLR1 and rRLR2 expression. In that case, *E. coli* strains were transformed with a plasmid harboring those genes under the control of an arabinose inducible promoter, prior the transformation with the tested RLR1 and RLR2 constructs [130]. All RLR1 and RLR2 constructs (listed in Table 8) were transformed in the different *E. coli* strains tested and one colony of was used to inoculate 5 mL of liquid LB/Ampicillin (Amp) media (100 µg/mL). After an overnight growth at 37 ºC, culture OD_{600} was checked and diluted to seed 15 mL of the different tested media also supplemented with Amp (100 µg/mL) (starting OD_{600} of 0.02). Cultures were grown at 37 ºC until an OD_{600} of 0.7, moved to the tested expression temperatures and protein expression was induced with the selected IPTG concentrations. For pre-expression of ErV1P and DsbC, cultures were grown at 37 ºC until an OD_{600} of 0.4 and the expression of Erv1p and DsbC was promoted by the addition of 0.5% w/v arabinose. After 30 min, the cultures were adjusted to the selected expression temperature and rRLR1 and rRLR2 protein expression was induced by adding the different IPTG concentrations.

Table 12 – Variables tested in the *E. coli* small scale expression screening.

E. coli Strain	Expression Temperature	Expression Length	IPTG Concentration (mM)	Medium
BL21* (DE3) (Invitrogen)	37 ^{o}C	3 h	0.5	LB
C41 (Invitrogen)	20 ^{o}C	Overnight	0.1 0.05	TB

At the end of the expression 1 mL of culture was collected, cells were harvested by centrifugation at 16000*g* for 5min at room temperature (RT) and the pellet was resuspended in 200 µL of BugBuster® Master Mix (Novagen) by gentle pipetting. After a 20 min incubation at room temperature (RT) in a shaking platform, the insoluble fraction was pelleted at 16000*g* for 10 min at 4 ^{o}C. The soluble fraction (supernatant) was collected and the pellet was resuspended in 200 µL of 20 mM Tris buffer (pH 8) to produce the insoluble fraction. 50 µL of each sample were used for Western blot in order to assess the amount of protein produced. In this case, the primary antibody used was an Anti-His antibody (GenScript) diluted 1:10000 and the secondary antibody was an Anti-mouse antibody (GE Healthcare) diluted 1:10000.

- ***Recombinant protein production in N. benthamiana leaves***

Constructs listed in Table 10 were inserted in the two tested magnICON® expression vectors A1TMVα and A2PVXα. The vectors were then transformed into *Agrobacterium tumefaciens* strain GV3101 by a combined heat and cold treatment. Positive clones were selected by single colony PCR using the primers listed in Table 13 and glycerol stocks were made and stored at -80 ^{o}C.

Table 13 – Primers used for selection of *A. tumefaciens* positive transformants by colony PCR.

Name	Sequence
Agro_RLR1_for	ACA GGT GCA AGT TTA GAT GG
Agro_RLR1_rev	GAA AAT AGA CAT CCC ATT AGA A
Agro_RLR2_for	ATT TTG GGA GAT ACC ACA CC
Agro_RLR2_rev	GTT CAA ATT ACC AGG AGT CAC

For *N. benthamiana* leave infiltration, positive clones were grown overnight at 30 ^{o}C, cells were harvested by gentle centrifugation at 3200*g* for 15 min at RT and resuspended in magnicon buffer (10 mM MES pH 5.6, 10 mM $MgSO_4$) supplemented with acetosyringone (1/1000) to a final OD_{600} of 0.2. Cell suspension was used to infiltrate leaves of 5-w-old WT *N. benthamiana* grown in a growth chamber at 22 ^{o}C under a 16h/8h light/dark cycle (2-3 leaves

were infiltrated *per* plant). Infiltrated plants were left undisturbed for 3 or 5 days (depending on the experiments) in the growth chamber under the same conditions. Infiltrated leaves were collected, immediately frozen in liquid nitrogen and stored at -80 ^{o}C.

- ***Small-scale protein purification from* N. benthamiana *leaves***

For small-scale protein isolation, total soluble protein (TSP) was extracted from 1 leaf of each construct by grinding leaf material into powder. The leaf powder was resuspended in 1 mL of extraction buffer (45 mM Tris.HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 40 mM ascorbic acid) and incubated on ice for 15 min. Cellular debris were pelleted by centrifugation at 12500*g* for 10 min at 4 ^{o}C and the supernatant (TSP) was collected.

For small-scale purification of the StrepTag proteins, TSP was incubated with 80 µL of StrepTag beads (Strep-Tactin Macroprep, Iba) during 1 h 30 min at 4 ^{o}C in a roller shaker. Beads were pelleted by gentle centrifugation, placed in a spin column and washed 5x with StrepTag washing buffer (100 mM Tris.HCl pH 8.0, 150 mM NaCl, 1 mM EDTA). Protein was eluted with 50 µL of StrepTag elution buffer (100 mM Tris.HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin) and 25 µL of each sample were applied in a SDS-PAGE Coomassie stained gel while the other 25 µL were used for Western blot analysis. For Western blot the primary antibody used was an Anti-Strep antibody (Iba) diluted 1:1000 and the secondary antibody was an Anti-mouse antibody (GE Healthcare Life Sciences) diluted 1:10000.

For small-scale purification of the HisTag proteins, TSP was incubated with 80 µL of HisTrap beads (Ni-Sepharose Fast-Flow, GE Healthcare Life Sciences) during 1 h 30 min at 4 ^{o}C in a roller shaker. Beads were pelleted by gentle centrifugation, placed in a spin column and washed 5x with HisTag washing buffer (100 mM Tris.HCl pH 8.0, 500 mM NaCl, 10 mM imidazole). Protein was eluted with 50 µL of HisTag elution buffer (100 mM Tris.HCl pH 8.0, 500 mM NaCl, 500 mM imidazole) and 25 µL of each sample were applied in a SDS-PAGE Coomassie stained gel while the other 25 µL were used for Western blot analysis. In this case, the primary antibody used was an Anti-His Antibody (GenScript) diluted 1:10000 and the secondary antibody was an Anti-mouse antibody (GE Healthcare Life Sciences) diluted 1:10000.

- ***rRLR1 and rRLR2 large scale purification from N. benthamiana leaves***

For TSP extraction, 15-20 g of frozen leaf material were used *per* construct. This material was crushed into powder (in liquid nitrogen) and resuspended in extraction buffer (45 mM Tris.HCl pH 7.4, 1.5 M NaCl, 1 mM EDTA, 40 mM ascorbic acid), 2 mL of extraction buffer

per gram of frozen material. This mixture was incubated on ice for 15 min, centrifuged at 104000*g* during 20 min at 4 ºC and the supernatant (TSP) was collected.

The StrepTag recombinant proteins were purified by loading TSP extract onto a StrepTrap™ HP 1 mL column (GE Healthcare Life Sciences) previously equilibrated with StrepTag washing buffer (100 mM Tris.HCl pH 8.0, 150 mM NaCl, 1 mM EDTA). After extensive washing, the protein was eluted with 10 mL of StrepTag elution buffer (100 mM Tris.HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin). The presence of the recombinant protein and its purity state were assessed by Coomassie stained SDS-PAGE gels and by Western blot analysis using anti-Strep antibody (Iba) as the primary antibody (1:1000 dilution) and anti-mouse antibody (GE Healthcare Life Sciences) as a secondary antibody (1:10000 dilution).

To purify the HisTag recombinant proteins, TSP extract was loaded in a HisTrap HP 1 mL column (GE Healthcare Life Sciences) pre equilibrated with HisTrap washing buffer (20 mM Tris.HCl pH 7.5, 500 mM NaCl and 10 mM imidazole). After extensive washing with the same buffer, protein elution was performed over a three-step imidazole gradient (50 mM, 100 mM and 500 mM). All fractions were analyzed by Coomassie stained SDS-PAGE gels. All eluted fractions were pooled together and buffer exchanged into 20 mM Tris.HCl pH 7.5 by an overnight dialysis at 4 ºC. Dialyzed samples where further purified by anion-exchange chromatography. Samples were applied onto a MonoQ 5/50 GL column (Amhersham Biosciences), previously equilibrated with the same dialysis buffer, and protein elution was achieved by a linear gradient of NaCl (0-1 M). The presence of the recombinant protein and its purity state were assessed by SDS-PAGE gel stained with Coomassie and by Western blot analysis using anti-His antibody (GenScript) as the primary antibody (1:10000 dilution) and anti-mouse antibody (GE Healthcare Life Sciences) as a secondary antibody (1:10000 dilution).

- ***SDS-PAGE and Western blotting***

Unless stated otherwise, all samples used for either Coomassie stained SDS-PAGE or Western blotting were loaded with 6x loading buffer (0.35 M Tris.HCl/0.28% SDS buffer pH 6.8, 30% glicerol, 10% SDS, 0.6 M DTT, 0.012% bromophenol blue), denatured at 90 ºC for 8 min and separated by denaturing SDS-PAGE. The SDS-PAGE was performed in 12.5% polyacrylamide gels and the running buffer was 100 mM Tris, 100 mM bicine supplemented with 0.1% SDS. Gels were run at 150 V at RT for at least 45 min using a MiniProtean3 system (Bio-Rad) and at the end they were either stained with Coomassie Brilliant Blue R-250 or electrotransferred to a PVDF membrane for Western blotting analysis. For this analysis, gels were transferred during an overnight period using a Trans-Blot® Electrophoretic Transfer Cell

(Bio-Rad) at 40 V, 4 ^{o}C in the buffer 25 mM Tris, 192 mM glicine, 20% methanol. Membranes were then blocked during one hour in TBST buffer (20 mM Tris.HCl pH 7.6, 192 mM NaCl, 0.1% v/v Tween 20) supplemented with 5% skim milk and then incubated during 1 h with the primary antibody diluted in TBST buffer supplemented with 0.5% skim milk. Membranes were washed at least 5x (5 min each time) with TBST with 0.5% skim milk and were then incubated with the secondary antibody for 1 h, also diluted in the same TBST buffer. Membranes were washed at least 5x (5 min each time) with TBST buffer, developed with ECF^{TM} substrate (GE Healthcare Life Sciences) and scanned in a Molecular Imager FX (Bio-Rad).

For semi-denaturing SDS-PAGE, the procedure was the same except for the 6x loading buffer which did not contain DTT.

- ***rRLR1 enzymatic characterization***

rRLR1 pH and inhibition profiles were determined by fluorescence assays using the fluorogenic substrate [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP] at a final concentration of 8.18 µM. The fluorescence was measured using FluoroMax-3® spectrofluorimeter (excitation and emission wavelengths of 328nm and 393nm, respectively). To determine rRLR1 pH profile, its activity against the synthetic peptide was assessed at 25 ^{o}C in buffers ranging between pH 4 and pH 6 (50 mM sodium acetate pH 4, 4.5, 5, 5.5 and 6) supplemented with 100 mM NaCl. To assess the inhibition profile, rRLR1 was pre-incubated with the different inhibitors listed in Table 14 for 2 min at RT in 50 mM sodium acetate buffer pH 4 supplemented with 100 mM NaCl followed by the addition of the substrate and determination of rRLR1 proteolytic activity at 25 ^{o}C.

To provide additional confirmation that the activity was from rRLR1, a pull-down assay was done using His-Mag Sepharose™ Ni^{2+} magnetic beads (Ge Healthcare Life Sciences) according to the manufacturer manual with minor modifications. Briefly, purified and active rRLR1 was incubated with the magnetic beads for 2 h in a shaker, at 4 ^{o}C. After a washing step with washing buffer (20 mM sodium phosphate buffer pH 7.4, 500 mM NaCl, 10 mM imidazole), protein was eluted with the buffer 20 mM sodium phosphate buffer pH 7.4, 500 mM NaCl, 500 mM imidazole. Proteolytic activity against [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP] was checked. All fractions were also analyzed by Western blotting to evaluate the efficiency of the pull-down assay.

Table 14 – List of compounds used for the determination of rRLR1 inhibition profile.

	Compound	Concentration (mM)
Protease Inhibitors	**Pepstatin A**	0.001
	Bestatin	0.01
	Pefabloc	1
	E-64	0.01
	EDTA	10
Redox Agents	**DTT**	1
	GSH	2
	GSSG	2
	NAD	3
	NADP	2
	NADPH	2
Nucleotides	**dATP**	1
	dGTP	1
	dTTP	1
	dCTP	1
	ATP	1
	ADP	1
	GTP	1
	CTP	1
Ions	**NaCl**	20
	$CaCl_2$	2
	$MnCl_2$	1
	$MgCl_2$	2

- ***rRLR1 de-glycosylation assays***

De-glycosylation assays were done using two endoglycosidases: EndoH and PNGase F (New England Biolabs). The assays were carried-out according to the manufacturer's instructions with some minor changes and all buffers were also supplied by the manufacturer. For EndoH treatment, 25 µL of purified rRLR1 supplemented with 2.5 µL of 10x glycoprotein denaturing buffer were denatured at 100 ^{o}C for 10 min. After cooling down, the total reaction volume was brought to 50 µL by adding 5 µL of 10x glycoBuffer 3, 0.5 µL of EndoH and 17 µL of ddH_2O. The mixture was incubated at 37 ^{o}C for 1 h and then analyzed by Coomassie stained

SDS-PAGE. For PNGase F treatment, 25 µL of purified rRLR1 supplemented with 2.5 µL of 10x glycoprotein denaturating buffer were denatured at 95 ^{o}C for 5 min. After cooling down on ice, total reaction volume was brought to 50 µL by the addition of 5 µL of 10x glycobuffer 2, 5 µL of 10% NP-40, 0.5 µL of PNGase F and 12 µL of ddH_2O. The mixture was incubated for 1 h at 37 ^{o}C and analyzed by Coomasie stained SDS-PAGE.

- ***Determination of rRLR1 PICS profile***

For the determination of rRLR1 Proteomics Identification of Cleavage Sites (PICS) profile, we followed the methodology described by Schilling and co-workers [131] with some minor modifications. Peptide libraries (Tryptic and GluC) used were previously prepared using THP1 cells derived from peripheral blood of a 1-year-old male with acute monocytic leukemia [132].

  - ***Test protease assay***

One aliquot of each peptide library (200 µg) was used for each experiment and pH was adjusted to 50 mM sodium acetate buffer pH 4 supplemented with 100 mM NaCl, to a final peptide concentration of 1 mg/mL. rRLR1 was then incubated with the peptide library at a ratio of 1:20 (wt/wt) for 5 h at 25 ^{o}C followed by a 5 min incubation at 90 ^{o}C to inactivate the protease.

To purify the neo-amino termini generated by rRLR1 cleavage they had to be biotinylated first. Sample pH was adjusted to pH 7.5 by the addition of 100 mM HEPES buffer pH 7.4 and incubated with a freshly made solution of 0.5 mM sulfosuccinimidyl 2-(biotinamido)-ethyl-1,3-dithiopropionate (sulfo-NHS-SS-biotin) for 1 h at 25 ^{o}C. The biotinylated neo-amino termini were purified using high-capacity streptavidin sepharose resin (GE Healthcare Life Sciences) previously equilibrated in washing buffer (50 mM HEPES buffer pH 7.4, 150 mM NaCl). Samples were incubated with the resin at a ratio of 1 volume of sample to 1.5 volumes of resin for 16 h at 4 ^{o}C. Resin was pelleted at 200*g* for 30s and washed with 1 mL of washing buffer. The washing step was repeated 10x and resin was then transferred to a spin column with a filter of approximately 10 µm pore size and washed again 10x with washing buffer. Peptides were eluted by gently resuspension of the resin in elution buffer (50 mM HEPES buffer pH 7.5, 20 mM DTT) followed by 1 h incubation at 25 ^{o}C and 1 min centrifugation at 500*g*. The elution step was repeated twice and the two supernatants (eluted fractions) were combined. To remove all salts, they were acidified with 2% formic acid to a pH below 2.5 and loaded onto a reverse-phase C18 column (100 cc) previously equilibrated with 1 mL of 0.5% formic acid. After a centrifugation at 700*g* for 1 min, the column was washed 10x with 1 mL of

the same buffer and peptides were eluted with 1 mL of 70% acetonitrile plus 0.5% formic acid. Samples were vacuum-evaporated to near dryness (10-20 µL).

- ***LC-MS/MS data analysis***

The new peptides generated by rRLR1 cleavage were then identified by LC-MS/MS using a HALO™ C18 column (Eksigent) connected to a LC-MS/MS TripleTOF 5600 (AB SCIEX), operated by the Center for Neuroscience and Cell Biology Proteomics Unit. Purified peptides were resuspended in 0.1% formic acid and applied to the HALO™ C18 column (Eksigent). Peptides were then eluted by an acetonitrile gradient (2-40%) obtained by combining buffer A (0.1% formic acid) and buffer B (98% acetonitrile, 0.1% formic acid). MS data was acquired automatically using Analyst QS software (Applied Biosystems) for information-dependent acquisition based on a 1s MS survey scan (mass ranges listed below) followed by up to 2 MS/MS scans of 3s each. Nitrogen was used as the collision gas and the ionization tip voltage was 25,000V. LC-MS/MS raw data files (.wiff) were converted into .mgf files using ProteinPilot software (AB SCIEX) using the default parameters. The .mzxml files were further generated using the software msConvert (ProteoWizard) applying MS1 level filter [133].

The identification of prime-side sequences from LC-MS/MS data was done with the spectrum-to-sequence database search program X!Tandem [134] together with PeptideProphet [135] at a >99% confidence level. Search parameters included mass tolerances of 0.05 Da for parent ions and 0.1 Da for fragment ions and the following static modifications: carboxyamidomethylation of cysteine residues (+ 57.02 Da), dimethylation of lysines (+ 28.03 Da) and thioacylation of peptide amino termini (+ 88.00 Da). Semi-style cleavage searches were applied with no constraints for the orientation of the specific terminus. Tryptic specificity was defined to cleavage C-terminal to Lys or Arg and GluC specificity was defined to cleavage C-terminal to Glu or Asp. Up to three missed cleavages were allowed for the library-generating enzyme.

Nonprime-side sequences were inferred by the web-based integrated series of data handling scripts termed WebPICS [136]. This web service automatically retrieved non-prime cleavage sequences, rendered the list of cleavage sites nonredundant, provided heatmap-style graphical and tabular representation of subsite preferences, and screened the cleavage sites for potential subsite cooperativity. The occurrences of amino acids relative to natural abundance retrieved by WebPICS were log-transformed in order to represent the under- and over-represented amino acids for each position with the same amplitude in opposite directions. Final heatmaps were constructed using MeV software [137].

The sequence logos representations of subsite specificities encompassing the cleavage site were further obtained with IceLogo tool available at http://icelogo.googlecode.com/. Cleavage site sequences from P4 to P4' created by WebPICS were filtered to exclude redundant peptides and analyzed by the IceLogo algorithm using the proteome background (Swiss-Prot) of *Homo sapiens* as the reference set. Fold change and percentage difference graphical representations were generated with a p-value of 5%.

3.3 Results

- ***E. coli is not a suitable host for rRLR1 and rRLR2 production***

As we stated before, our first approach to produce rRLR1 and rRLR2 was using *E. coli* as the expression host since these bacteria can grow very quickly in a relatively low cost media, allowing the production of good amounts of recombinant protein in a short period of time [138]. Our goal was to produce both proteins in their soluble form in *E. coli* cytoplasm in order to circumvent the lengthy refolding procedure that is normally used when the recombinant proteins are produced in the form of inclusion bodies [53, 126]. However, one of the major limitations of *E. coli* as an expression host is its inability to perform post-translational modifications such as disulfide bond formation in the cytoplasm [127]. This is a major issue especially when trying to produce high cysteine content proteins in their soluble form in *E. coli* cytoplasm. Analyzing RLR1 and RLR2 primary structure represented in Figure 22 it is clear that both proteins have a large number of cysteine residues which may impair the amount of soluble protein that can be produced in *E. coli*. In order to circumvent this drawback, we tried the pre-expression of a sulfhydryl oxidase (Erv1p) and a disulfide isomerase (DsbC) in *E. coli* cytoplasm since, according to Nguyen *et al.*, this facilitates the formation of disulfide bonds [130]. Moreover, we also tested the addition of the Fh8 tag to the C-terminal of both recombinant proteins since this expression tag was previously reported to increase recombinant protein solubility [139].

Our goal was to obtain the maximum amount of soluble recombinant protein in *E. coli* cytoplasm so we first performed a small-scale expression screening to find the expression condition that would generate the best results. As discussed, we tested several conditions combining the different variables depicted in Table 12. Moreover, we also tested several DNA constructs (depicted in Figure 23) with different expression tags (Fh8 and HisTag) and also the removal of the putative pro-segment of each protein. In typical APs, the pro-segment is thought to be important for folding and to maintain the enzyme in its inactive form while its function is not needed [6]. For atypical and nucellin-like APs the function of the pro-segment is still unclear since there are several of these enzymes with unusually long pro-segments while others do not seem to have a pro-segment at all [34, 50]. Moreover, recombinant chlapsin, the *C. reinhardtii* typical AP, was produced without its pro-segment in an active form and was the only form of the protein that the authors were able to produce, suggesting that for some of these proteases this domain may not be that critical for protein folding [126]. Taking all this

into account, we decided to test if the removal of rRLR1 and rRLR2 putative pro-segment had any impact on the yield of recombinant protein produced.

a)

```
  1 MASSSSSSLL FPFFLILFSC LISVSSSRRS LIDRTLPKNL PRSGFRLSLR
 51 HVDSGKNLTK IQKIQRGINR GFHRLNRLGA VAVLAVASKP DDTNNIKAPT
101 HGGSGEFLME LSIGNPAVKY SAIVDTGSDL IWTQCKPCTE CFDQPTPIFD
151 PEKSSSYSKV GCSSGLCNAL PRSNCNEDKD ACEYLYTYGD YSSTRGLLAT
201 ETFTFEDENS ISGIGFGCGV ENEGDGFSQG SGLVGLGRGP LSLISQLKET
251 KFSYCLTSIE DSEASSSLFI GSLASGIVNK TGASLDGEVT KTMSLLRNPD
301 QPSFYYLELQ GITVGAKRLS VEKSTFELAE DGTGGMIIDS GTTITYLEET
351 AFKVLKEEFT SRMSLPVDDS GSTGLDLCFK LPDAAKNIAV PKMIFHFKGA
401 DLELPGENYM VADSSTGVLC LAMGSSNGMS IFGNVQQQNF NVLHDLEKET
451 VSFVPTECGK L
```

b)

```
  1 MAIFFTSPLF FLIILCFSIS VVHLSASPTL VLNLVHSYHI YSRKPPHVYH
 51 IKEASVERLE YLKAKTTGDI IAHLSPNVPI IPQAFLVNIS IGSPPITQLL
101 HMDTASDLLW IQCLPCINCY AQSLPIFDPS RSYTHRNETC RTSQYSMPSL
151 KFNANTRSCE YSMRYVDDTG SKGILAREML LFNTIYDESS SAALHDVVFG
201 CGHDNYGEPL VGTGILGLGY GEFSLVHRFG KKFSYCFGSL DDPSYPHNVL
251 VLGDDGANIL GDTTPLEIHN GFYYVTIEAI SVDGIILPID PRVFNRNHQT
301 GLGGTIIDTG NSLTSLVEEA YKPLKNRIED IFEGRFTAAD VSQDDMIKME
351 CYNGNFERDL VESGFPIVTF HFSEGAELSL DVKSLFMKLS PNVFCLAVTP
401 GNLNSIGATA QQSYNIGYDL EAMEVSF
```

Figure 22 - RLR1 and RLR2 amino acid sequence. a) RLR1 (accession number Q84M99) and **b)** RLR2 (accession number Q9SZV7) amino acid sequences. Highlighted in red is the putative signal peptide, in green the putative pro-segment, in yellow the two catalytic sequences and in dark blue the cysteine residues. RLR2 putative pro-segment is unusually long. Both proteins have a high number of cysteine residues (12 in RLR1 and 10 in RLR2).

For rRLR1 small-scale expression screening we first tested the expression of the full-length protein (excluding the signal peptide) fused to the HisTag, in order to allow protein detection by Western blotting and facilitate a subsequent purification procedure (RLR1-His construct). In this initial screening we tested a combination of several variables such as different *E. coli* strains (BL21* (DE3) and C41), expression media (LB and TB), expression temperatures (37 ^{o}C and 20 ^{o}C), expression time (3 h or overnight) and IPTG concentrations (0.1 mM and 0.5 mM). Since RLR1 has a high number of cysteine residues (see Figure 22), we also tested the pre-expression of Erv1p and DsbC [130]. In Figure 24 are depicted the results. In most conditions tested, almost all recombinant protein accumulates in the insoluble fraction. *E. coli* strain C41 (which is a strain normally used to express difficult and sometimes toxic proteins) appeared to be the best strain especially when the pre-expression of Erv1p and DsbC was promoted. Finally, low expression temperatures resulted in lower expression yields in C41 but with slightly higher levels of protein accumulated in the soluble fraction (particularly

when co-expressing Erv1p and DsbC). However, none of the tested conditions allowed the production of a significant amount of recombinant protein in the soluble form and so we decided to test other constructs and conditions.

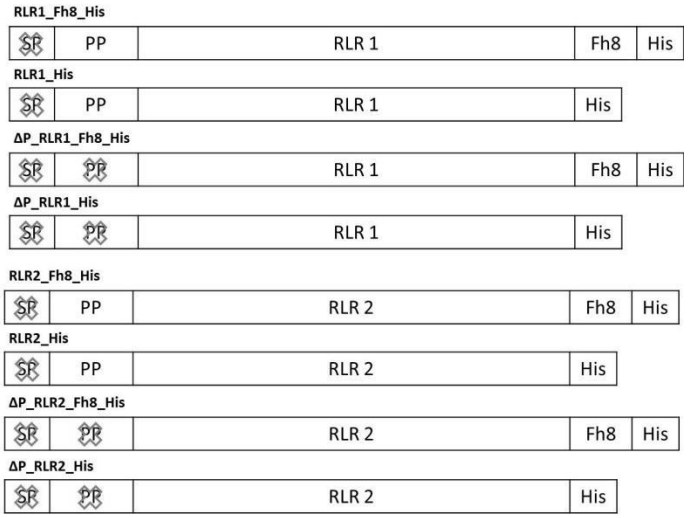


Figure 23 - Constructs used for the small-scale expression screening in *E. coli*. RLR1_Fh8_His and RLR2_Fh8_His were obtained using CPEC cloning technique while all other constructs were generated using standard cloning techniques. All constructs were inserted in the commercial vector pET23a. SP – Signal Peptide; PP – Pro Peptide/pro-segment. Sequences overlaid with the X symbol are not present in that specific construct. See also **Table 8** for a more detailed description.

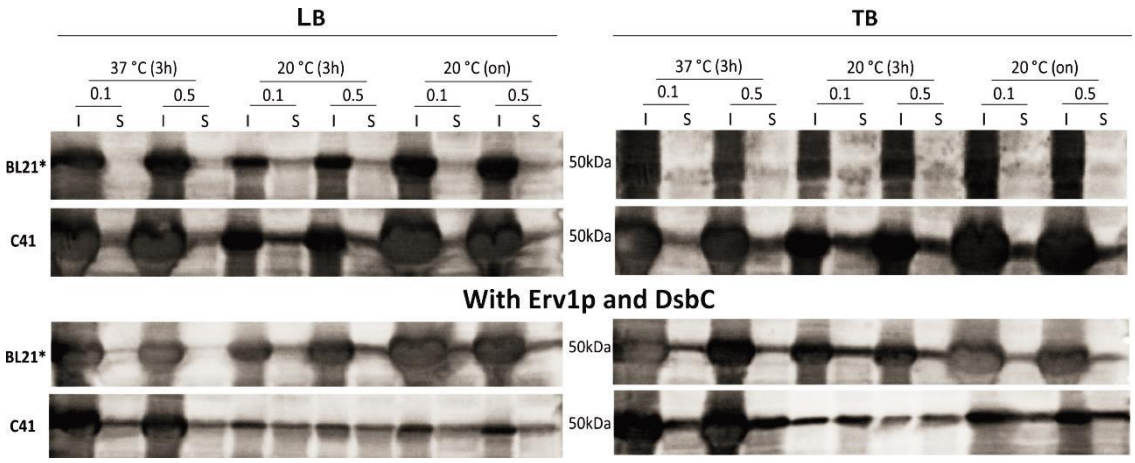


Figure 24 – Western blot analysis of the initial small-scale expression screening for production of soluble rRLR1 in *E. coli*. The only construct tested was the full-length rRLR1 fused to a HisTag at its C-terminal (RLR1-His). We tested a combination of two different *E. coli* strains (BL21* (DE3) and C41), two expression media (LB and TB), two expression temperatures (37 ^{o}C and 20 ^{o}C), two expression times (3 h and overnight) and two IPTG concentrations (0.1 and 0.5 mM). We also tested without (upper part) and with (botton part) the pre-expression of Erv1p and DsbC according to [130]. For each condition we analyzed the insoluble (I) and the soluble (S) fraction. Most of the recombinant protein appears in the insoluble fraction in all conditions.

In a second expression screening we tested different constructs (with and without RLR1 pro-segment and also the use of the Fh8 tag fused at RLR1 C-terminal) (see Figure 23). In

this case we decided to use just C41 *E. coli* strain and also the pre-expression of Erv1p and DsbC since those were the conditions in the previous screening with lower accumulation of protein in the insoluble form. We also tested just one expression temperature (20 ºC) and reduced the IPTG concentrations used (0.1 mM and 0.05 mM). The results obtained are illustrated in Figure 25. We could obtain more soluble protein in almost all of the conditions tested. Nonetheless, the amount of protein was still too low to proceed to a larger scale purification and optimization of rRLR1 purification. Moreover, the removal of the pro-segment or the inclusion of the Fh8 tag did not result in a substantial increment in the amount of protein produced.

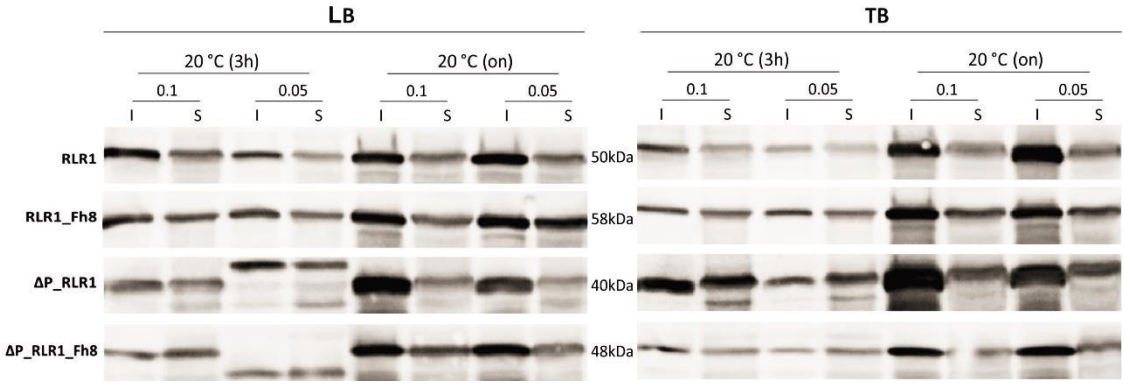


Figure 25 – Western blot analysis of the second small scale-expression screening for production of soluble rRLR1 in *E. coli*. All expression conditions were tested just with C41 *E. coli* strain and pre-expression of Erv1p and DsbC. We tested a combination of two expression media (LB and TB), one expression temperature (20 ºC), two expression lengths (3 h and overnight) and two IPTG concentrations (0.1 mM and 0.05 mM). Different constructs with the Fh8 tag (Fh8 constructs) and also with and without the pro-peptide (ΔP constructs) were also tested (see also Figure 22). The Fh8 tag was fused at the C-terminal of rRLR1 and all the constructs have a HisTag at their C-terminal. All constructs lack the signal peptide. For each condition we analyzed the insoluble (I) and the soluble (S) fraction. There was no substantial increment of rRLR1 accumulation in the soluble fraction under the conditions tested.

For rRLR2 production the initial strategy was the same. A first small-scale expression screening was performed just trying to express the full-length protein with a HisTag at its C-terminal (RLR2-His construct) and the conditions tested were also the same as those described for rRLR1 initial screening. Figure 26 shows the results of this first small-scale expression screening and, as observed for rRLR1, in almost all conditions tested rRLR2 accumulated mainly in the insoluble fraction. As before, the conditions using C41 *E. coli* strain and the pre-expression of Erv1p and DsbC under a lower expression temperature appeared to result in reduced accumulation in the insoluble form but with very low yields of soluble protein being produced. Again we decided to do a second expression screening testing different rRLR2 constructs (see Figure 22), using C41 strain and the pre-expression of Erv1p and DsbC, under the same expression conditions (different temperatures, IPTG concentrations and different

media) as those used in the initial screening. The results of this second expression screening are shown in Figure 27**.** In the large majority of the conditions tested, the protein accumulated in the insoluble fraction. As previously observed for rRLR1 production, the removal of the pro-segment and the inclusion of Fh8 tag did not have any positive impact in the amount of soluble rRLR2 produced.

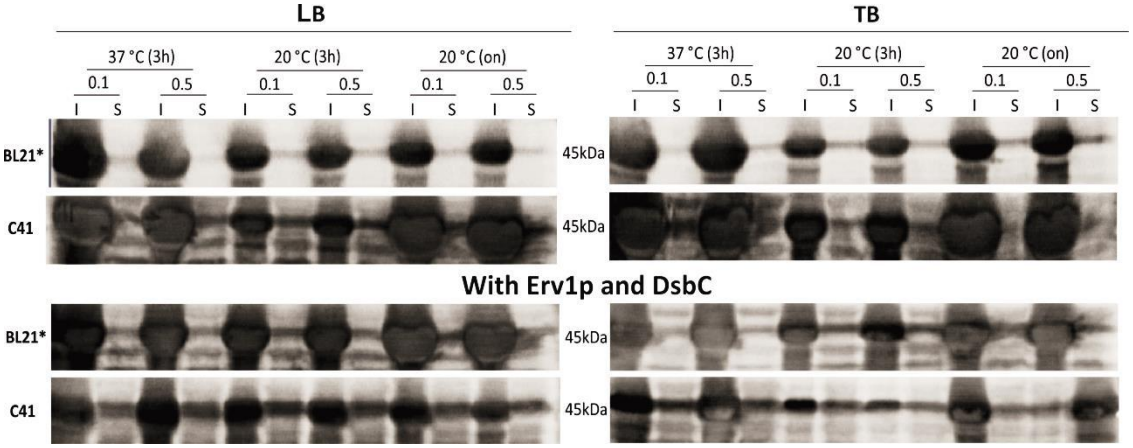


Figure 26 – Western blot analysis of the initial small-scale expression screening for production of soluble rRLR2 in *E. coli*. The only construct tested was the full-length rRLR2 fused to a HisTag at its C-terminal (RLR2-His). A combination of two different *E. coli* strains (BL21* and C41), two expression media (LB and TB), two expression temperatures (37 ºC and 20 ºC), two expression times (3 h and overnight) and two IPTG concentrations (0.1 mM and 0.5 mM) were tested. Expressions were also performed without (upper part) and with (botton part) the pre-expression of Erv1p and DsbC according to [130]. For each condition we analyzed the insoluble (I) and the soluble (S) fraction. Most of the recombinant protein appears in the insoluble fraction in all conditions.

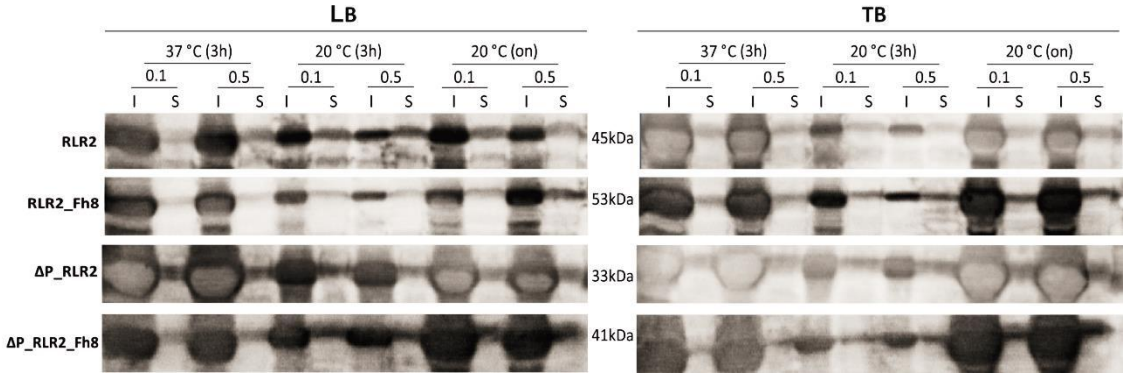


Figure 27 – Western blot analysis of the second small-scale expression screening for production of soluble rRLR2 in *E. coli*. All expression conditions were tested using C41 *E. coli* strain and pre-expression of Erv1p and DsbC. A combination of two expression media (LB and TB), two expression temperatures (37 ºC and 20 ºC), two expression times (3 h and overnight) and two IPTG concentrations (0.1 mM and 0.5 mM) were tested. Different constructs with the Fh8 tag (Fh8 constructs) and also with and without the pro-peptide (ΔP constructs) (see also Figure 22) were also tested. The Fh8 tag was fused at the C-terminal of RLR1 and all the constructs have a HisTag at their C-terminal. All constructs lack the signal peptide. For each condition we analyzed the insoluble (I) and the soluble (S) fraction. Most of the recombinant protein accumulates in the insoluble fraction in all conditions.

Taking all these results together, we can conclude that it was not possible to identify a suitable condition (from those tested) to obtain the desirable amounts of soluble rRLR1 and rRLR2 in *E. coli*. It is clear that *E. coli* is not a suitable heterologous host to produce these proteins so we decided to try a different expression system.

- ***N. benthamiana is a viable expression host to produce rRLR1 and rRLR2***

In the previous section we described the unsuccessful attempts to produce rRLR1 and rRLR2 in their soluble form in *E. coli* cytoplasm. Since RLR1 and RLR2 are plant proteins, we then decided to express both proteases in a plant based expression system. In collaboration with Dr. Herta Steinkellner at BOKU University in Vienna, Austria, we were able to transiently express rRLR1 and rRLR2 in *N. benthamiana* leaves using the magnICON® expression system [129]. This system was first developed by ICON Genetics GmbH and it consists in a highly engineered viral-based vector that promotes transient expression of a protein of interest in *N. benthamiana* leaves [129]. According to the authors, this expression system allows high yields of the protein of interest in a short period of time (usually 5 to 10 days) [129]. Although magnICON® has been used to express a variety of proteins (mainly pharmaceutical proteins) [140-143], to our knowledge, rRLR1 and rRLR2 were the first APs expressed using this system.

Our first approach was also to perform a small-scale expression screening where different constructs were tested (depicted in Figure 28). As in the *E. coli* expression screening, we decided to test the removal of the pro-segment for the reasons already discussed. Moreover, we tested two different expression tags: the commonly used HisTag, and the StrepTag that was reported to be more specific and could facilitate the downstream purification procedure [144]. All constructs were inserted in two different magnICON® vectors, A1TMVα and A2PVXα, which have different promoters driving transgene expression: A1TMVα has the Tobacco Mosaic Virus promoter while A2PVXα has the Potato Virus X promoter. In Figure 29 are depicted the results obtained for rRLR1 small-scale expression screening. In panel a) are the results for the StrepTag constructs, two of them inserted in the A2PVXα vector (RLR1-Strep and ΔP_RLR1-Strep; lanes 1 and 2, respectively) and one inserted in the A1TMVα (ΔP_RLR1-Strep; lane 3). Analyzing the TSP it is clear that we obtained a good amount of rRLR1 with each construct. Nonetheless, when a small-scale purification was performed and the purified fraction (P) analyzed, only construct number 3 (ΔP_RLR1 in A1TMVα vector) was visible in Coomassie staining. Thus, the presence of the TMVα promoter in conjugation with the absence of the pro-segment resulted in the best combination for the StrepTag constructs. Moreover, the results were better for 3 days of expression, suggesting that longer expression times may be detrimental for protein accumulation. Panel b) illustrates the results obtained for the constructs with the HisTag; in this case just one of them inserted in the A2PVXα (RLR1-His; lane 4) and two inserted in the A1TMVα vector (RLR1-His and ΔP_RLR1-His; lanes 5 and 6, respectively). We tested just 3 expression days and by analyzing the TSP extracts, all constructs

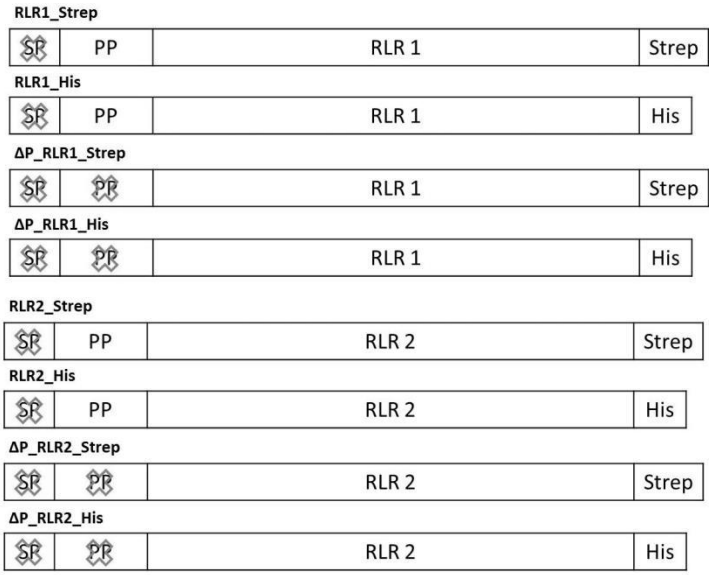


Figure 28 – Constructs used for small-scale expression screening of rRLR1 and rRLR2 in *N. benthamiana*. Native *RLR1* and *RLR2* coding sequences (cds) were inserted into Magnicon® vectors using *BsmBI* restriction enzyme for *RLR1* and *BsaI* for *RLR2* constructs. Two different expression tags were tested, StrepTag and HisTag, both fused to the 3' end of the cds. SP – Signal Peptide; PP – Pro Peptide. Sequences overlaid with the X symbol are not present in that specific construct. See also **Table 10** for a more detailed description.

displayed good yields of recombinant protein. However, when we tried the small-scale purification, construct number 5 (RLR1 in A1TMVα) gave the best results although its purification was less effective than the purification of the StrepTag constructs. Thus, the full-length RLR1 inserted in the A1TMVα was selected for rRLR1 production in larger scale (next section).

The rRLR2 small-scale expression screening results are shown in Figure 30. In this case, all tested constructs were inserted only in the A1TMVα vector since it was proved to be the best in rRLR1 expression screenings. Figure 30 a) shows the results for the two StrepTag fusion proteins (RLR1-Strep and ΔP_RLR1-Strep; lanes 1 and 2, respectively) and it was possible to obtain good amounts of recombinant protein at the end of the purification procedure with both constructs. Again, there was no difference between 3 or 5 expression days. For the HisTag fusions (b)) (RLR1-His and ΔP_RLR1-His; lanes 3 and 4, respectively) the results were quite similar although, as observed for rRLR1, the HisTag purification is not as efficient as the StrepTag purification procedure.

All together, these results clearly indicated that *N. benthamiana* may be a promising expression host to produce both rRLR1 and rRLR2 by using the magnICON® expression system. Supported by these small-scale screening assays, we then proceeded to the large scale production of these proteins using this expression system followed by the optimization of purification protocols.

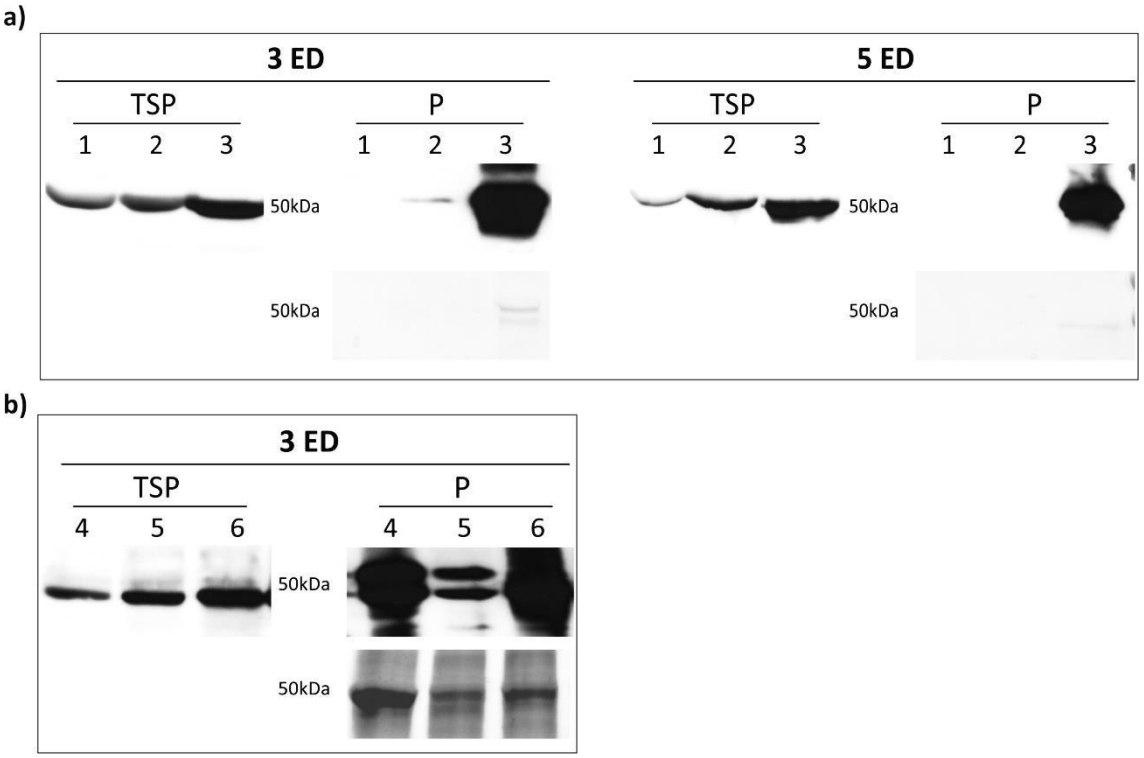


Figure 29 – Small-scale expression screening of rRLR1 in *N. benthamiana*. a) rRLR1 constructs with StrepTag (1 - RLR1 in A2PVXα, 2 - ΔP_RLR1 in A2PVXα, 3 - ΔP_RLR1 in A1TMVα). In this case we tested two different expression times (3 and 5 expression days (ED)) and we analyzed the total soluble protein (TSP) extract and the purified fraction (P) by Western blot (upper panel) and also the purified fraction by Coomassie-stained SDS PAGE (lower panel). Construct 3 gave better results after protein purification. **b)** rRLR1 constructs with HisTag (4 - RLR1 in A2PVXα, 5 - RLR1 in A1TMVα, 6 - ΔP_RLR1 in A1TMVα). In this case we only tested one expression time (3 ED) and we analyzed the total soluble protein (TSP) extract and the purified fraction (P) by Western blot (upper panel) and also the purified fraction by Coomassie-stained SDS PAGE (lower panel). All three constructs resulted in higher yields of purified protein when compared to the equivalent constructs with Strep-tag construct.

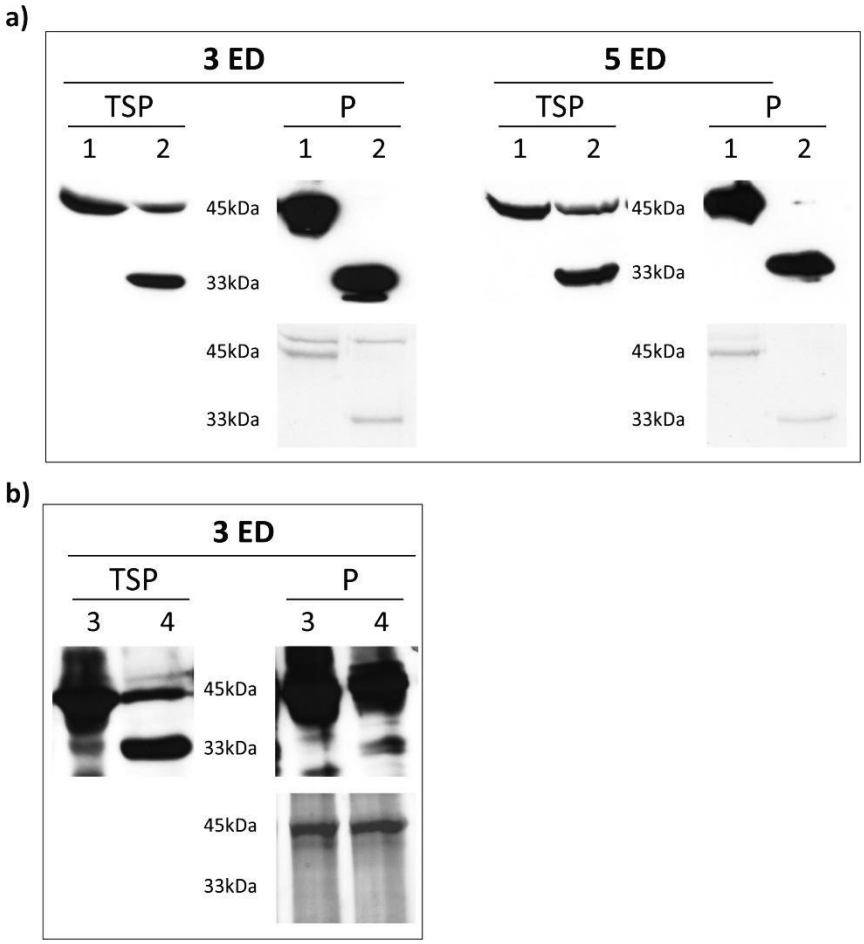


Figure 30 – Small-scale expression screening of rRLR2 in *N. benthamiana*. a) rRLR2 constructs with StrepTag, were all inserted in A1TMVα expression vector (1 - RLR2, 2 - ΔP_RLR2). Two different expression times (3 and 5 expression days (ED)) were tested and total soluble protein (TSP) extract and the purified fractions (P) were analyzed by Western blot (upper panel) and also the purified fraction by Coomassie-stained SDS PAGE (lower panel). Both fusion proteins resulted in a very good yield of purified recombinant protein after protein purification. **b)** rRLR2 constructs with HisTag were also inserted in A1TMVα expression vector (3 - RLR2, 4 - ΔP_RLR2). In this case only one expression time was tested (3 ED) and we analyzed the total soluble protein (TSP) extract and the purified fraction (P) by Western blot (upper panel) and also the purified fraction by Coomassie-stained SDS PAGE (lower panel). Again, no significant differences in protein yield were observed between the two constructs.

- ***rRLR1 was successfully expressed and purified from N. benthamiana leaves***

So far, the best results obtained concerning rRLR1 production was using *N. benthamiana* as the expression host and magnICON® vectors as the expression system. Therefore, we decided to pursue to the large scale production of this protein using these two tools. As we concluded in the previous section, the best two constructs in the small-scale screening were ΔP_RLR1-Strep and RLR1-His both using A1TMVα as the expression vector and these were selected for a large scale production trial and optimization of purification.

We started by the expression and purification of the StrepTag construct since the purification procedure was apparently more efficient, as shown in the previous section. In Figure 31 is depicted the procedure and a representative result of one purification. *A. tumefaciens* harboring the construct ΔP_RLR1_Strep (a)) was infiltrated in *N. benthamiana* leaves (b)), leaves were harvested and TSP extracted 3 days after infiltration. The TSP extract was then applied into a StrepTrap column and the eluted samples were analyzed by Coomassie-stained SDS-PAGE gel and Western blot (c)). The amount of protein obtained at the end of the procedure was very low and it was not pure enough to proceed to the biochemical studies. Moreover, every attempt that was made to further purify rRLR1 after the StrepTrap affinity chromatography resulted in the loss of the entire sample, clearly suggesting that this form of the protein is unstable.

Since the purification of the StrepTag construct was not successful we then produced and purified the proRLR1 construct fused to the HisTag (RLR1-His construct). In Figure 32 are shown the results of one representative purification procedure, being the infiltration and expression time the same as previously described. TSP extract resulting from 15-20 g of leaf material was applied into a HisTrap column and protein was eluted with 3 imidazole concentrations (50, 100 and 500 mM). Eluted fractions contained several contaminant proteins (a)) and were pooled together, dialyzed overnight to remove all the salts and applied into an anionic exchange column (MonoQ) for further purification (b)). The eluted fractions were then analyzed by Coomassie-stained SDS-PAGE gel and Western blot (b)).

At the end of this purification procedure, eluted fractions displayed either one or two bands near the expected molecular weight for the full-length rRLR1 without its signal peptide (47 kDa): one slightly above and one bellow 47 kDa (see Figure 32 b)). A bioinformatics analysis using NetNGlyc 1.0 Server algorithm [145] predicted that RLR1 is a glycosylated protein and since A1TMVα has its own signal peptide, the band above the 47 kDa could correspond to a

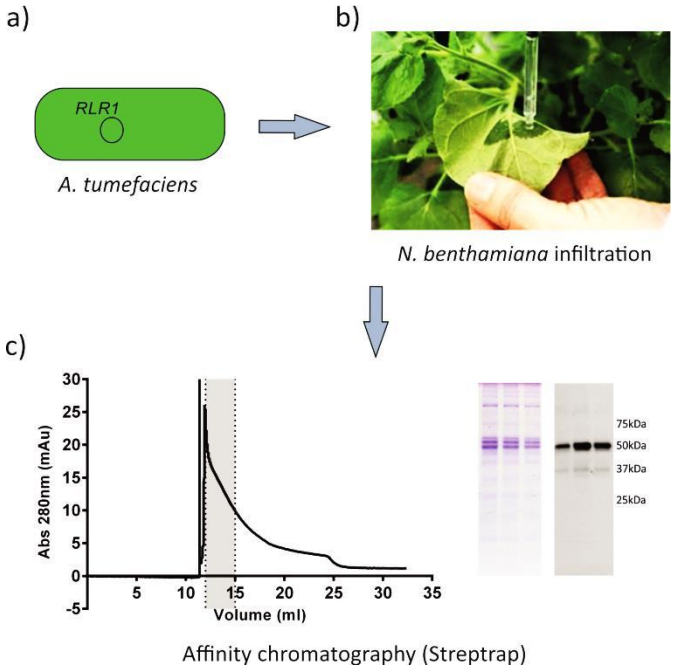


Figure 31 – Production and purification of ΔP_RLR1_Strep from *N. benthamiana* leaves. a) and **b)** *A. tumefaciens* harboring the ΔP_RLR1-Strep construct inserted in A1TMVα vector were infiltrated in 5-w-old WT *N. benthamiana* leaves; **c)** Infiltrated leaves (15-20 grams) were harvested 3 days after infiltration, total soluble protein was extracted and applied to a StrepTrap column. Eluted samples were then analyzed by Coomassie-stained SDS-PAGE and Western blot (Anti-StrepTag antibody). The amount of rRLR1 obtained was low and the eluted protein was not completely pure.

glycosylated form of rRLR1. De-glycosylation assays with EndoH and PNGaseF showed that this was indeed the case and that rRLR1 is a glycosylated protein (Figure 32 c)). EndoH can only cleave non mature N-glycosylations that occur in the Endoplasmic Reticulum (ER) but if the glycosylated protein enters the Golgi apparatus additional modifications may occur and the glycan becomes resistant to EndoH, but PNGase F can still cleave it. This is why the combination of EndoH/PNGase F treatment can be used to monitor protein trafficking throughout the secretory pathway and this is also the reason why we used both enzymes [146-148]. Both EndoH and PNGase F were able to remove rRLR1 glycosylation, clearly suggesting that rRLR1 does not have these mature glycosylations and most-likely does not enter the Golgi apparatus. The lower band can be one of two things: rRLR1 that was not glycosylated or rRLR1 without its pro-segment. The first hypothesis is linked to the fact that we were extracting TSP which likely contain protein that may still not be fully glycosylated yet. On the other hand, the removal of the pro peptide is a crucial step in the activation of typical APs [6] and this lower rRLR1 band could be glycosylated rRLR1 without its pro-peptide (RLR1 without the pro peptide is predicted to have ~40 kDa). N-terminal Edman sequencing was performed to both bands and the N-terminal sequence of the upper band (Figure 33) (sequence underlined in blue) revealed that this band corresponds to rRLR1 without its signal peptide whereas the N-terminal sequence of the lower band (sequence underlined in orange) corresponds to the

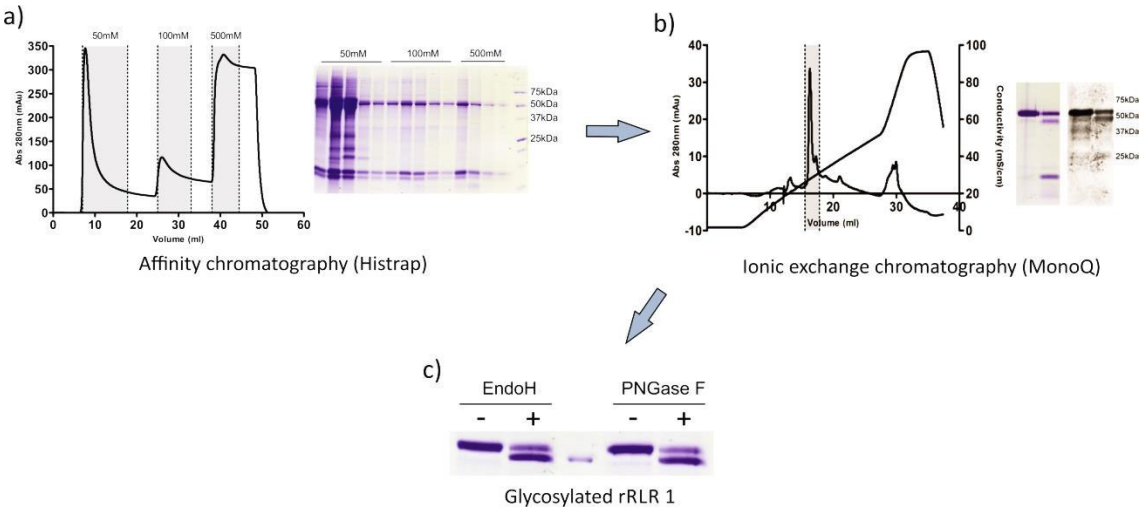


Figure 32 – Production and purification of rRLR1-His from *N. benthamiana* leaves. a) TSP extract resulting from 15-20 g of infiltrated leaves was applied into a HisTrap column and protein eluted with 3 imidazole concentrations (50, 100 and 500 mM). Eluted fractions, corresponding to the shaded areas of the chromatogram were analyzed by Coomassie-stained SDS-PAGE. **b)** All fractions were combined, dialyzed overnight and applied into an anionic exchange column (MonoQ). Fractions corresponding to the shaded area in the chromatogram were analyzed by Coomassie-stained SDS-PAGE and Western blot. **c)** The first eluted fraction was then de-glycosylated by an EndoH and PNGase F treatment, confirming that purified rRLR1 is produced as a glycosylated protein in this expression system.

processed form of RLR1 without the pro-segment. These results suggest that rRLR1 undergoes *in vivo* removal of the pro-segment in this heterologous system. It is interesting to note that both N-terminal sequences determined by Edman do not match exactly to predicted N-terminals of both rRLR1 without the signal peptide and without the putative pro-segment. If for the pro-segment this was not totally unexpected since the prediction was based on the number of amino acids typically preceding the first catalytic aspartate in mature APs and this may be slightly different for plant atypical APs, the results were more surprising for the signal peptide which was initially predicted using SignalP [149], a bioinformatics tool widely used for prediction of signal peptide cleavage sites.

At the end of this section, our results demonstrate that we were able to successfully establish a purification protocol for rRLR1. Our preliminary assays showed that rRLR1 displayed proteolytic activity and the biochemical enzymatic characterization of this recombinant protein will be described next.

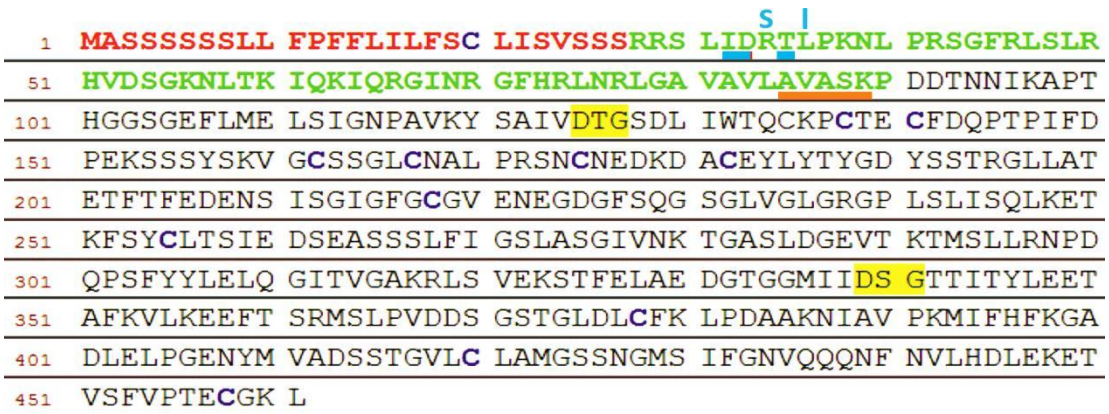


Figure 33 – rRLR1 amino acid sequence with the N-terminal Edman degradation sequencing results. Highlighted in red is the putative signal peptide, in green the putative pro-peptide, in yellow the two catalytic aspartates and in dark blue the cysteine residues. Amino acids underlined and also depicted in light blue were identified as being the N-terminal sequence of the higher molecular weight band in the SDS-PAGE analysis. Amino acids underlined in orange were identified as being the N-terminal sequence of the lower molecular weight band.

- ***rRLR1 is an active AP with atypical biochemical properties***

In the previous section we demonstrated that it was possible to purify rRLR1 from *N. benthamiana* leaves so we proceeded to the *in vitro* enzymatic characterization of this enzyme. We tested rRLR1 proteolytic activity against several fluorogenic substrates; however, the only stable activity that we reported was towards [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP]. This was an interesting finding as this substrate was specifically designed in our laboratory for the biochemical characterization of rCDR1, another *A. thaliana* atypical AP [53].

The pH and inhibition profiles of rRLR1 obtained with this substrate are illustrated in Figure 34 (a) and c), respectively). rRLR1 displayed its maximum activity at pH 4 and no activity was detected at pH 6. We also tested pH values below 4 but the activity was very unstable. For the inhibition profile, the proteolytic activity of rRLR1 was tested in the presence of a broad range of compounds (Figure 34 c) and Table 15). rRLR1 was significantly but not completely inhibited by pepstatin A (the canonical inhibitor of APs) and, interestingly, it was also strongly inhibited by redox agents such as NADP and NAPH. Other redox reagents, like NAD and oxidized gluthatione resulted also in a decrease in activity of approximately 50 %. Moreover, EDTA and pefabloc (a serine protease inhibitor) had also a partially inhibitory effect on rRLR1 activity. Nucleotides and ions did not significantly affect rRLR1 activity. Finally, to further demonstrate that the activity that was being characterized was indeed from rRLR1, a pull-down assay was performed with Ni^{2+} magnetic beads, taking advantage of the HisTag fused to rRLR1 C-terminal. The results are depicted in Figure 34 b) and analyzing the Western blot (upper part) it is possible to conclude that rRLR1 was effectively pulled down since no rRLR1 was detected in the nonbound sample or washing fractions but the signal was recovered after the elution step. Moreover, by testing the activity of the initial and the nonbound samples (lower part), activity was detected only in the initial sample confirming that it was indeed rRLR1 activity.

From these results we can conclude that rRLR1 is indeed an active AP with an optimum acidic pH, which is usually observed for the large majority of APs. However, the apparent narrow specificity of rRLR1, the fact that it was not fully inhibited by pepstatin A, together with pronounced inhibitory effect observed for several redox agents clearly anticipate some interesting and atypical properties for this protease.

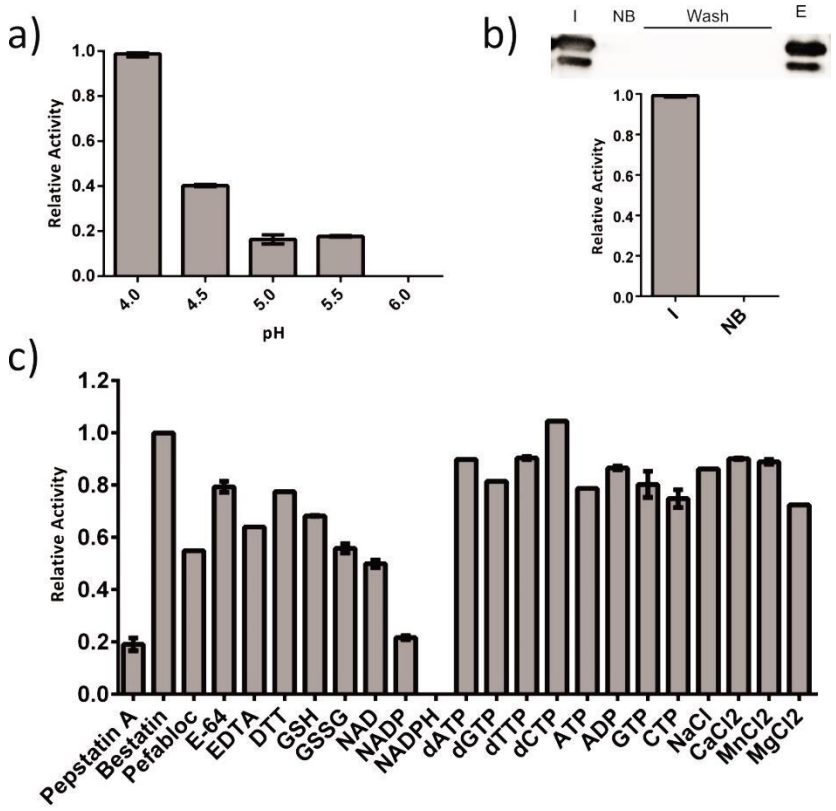


Figure 34 – rRLR1 *in vitro* biochemical characterization. All activity assays were conducted towards the fluorogenic substrate [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP] at 25˚C. **a)** Effect of pH on rRLR1 proteolytic activity. Activity studies at different pH values were performed by incubating rRLR1 with 50 mM sodium acetate buffers at pH 4.0, 4.5, 5, 5.5 and 6.0, supplemented with 100 mM NaCl. **b)** rRLR1 activity confirmation by pull-down assay using Ni^{2+} magnetic beads. Neither rRLR1 nor its activity was observed in the nonbound (NB) fraction. **c)** rRLR1 proteolytic activity inhibition profile. rRLR1 was pre-incubated with each compound for 2 min at room temperature in 50 mM sodium acetate buffer pH 4.0 with 100 mM NaCl before addition of the substrate and determination of rRLR1 proteolytic activity at 25 ^{o}C. rRLR1 is not completely inhibited by pepstatin A and is strongly inhibited by redox agents. I - Initial sample; NB - Nonbound (protein nonbound to the Ni^{2+} magnetic beads); E - Eluted protein.

Table 15 – rRLR1 inhibition profile showing the inhibitors used, their concentrations and rRLR1 activity in their presence

	Compound	Concentration (mM)	Activity (%)
Protease Inhibitors	**Pepstatin A**	0.001	19.1
	Bestatin	0.01	100
	Pefabloc	1	54.9
	E-64	0.01	79.4
	EDTA	10	64.0
Redox Agents	**DTT**	1	77.5
	GSH	2	68.3
	GSSG	2	55.8
	NAD	3	49.9
	NADP	2	21.7
	NADPH	2	0
Nucleotides	**dATP**	1	89.8
	dGTP	1	81.4
	dTTP	1	90.4
	dCTP	1	105
	ATP	1	78.7
	ADP	1	86.6
	GTP	1	80.3
	CTP	1	74.8
Ions	**NaCl**	20	86.2
	$CaCl_2$	2	90.1
	$MnCl_2$	1	88.9
	$MgCl_2$	2	72.5

- ***rRLR1 displays atypical amino acid preferences***

As discussed in the previous section, we tested rRLR1 activity against several fluorogenic substrates but the protease was found to be active only towards [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP]. This result suggested that rRLR1 could have stringent amino acid preferences and, to further evaluate this possibility, the specificity preferences of rRLR1 were profiled by a technique called Proteomics Identification of Cleavage Sites (PICS), developed by Schilling *et al.* [131, 150]. PICS is based on the proteomic identification of hundreds of newly cleaved peptide sequences after incubation of a proteome-derived peptide library with the protease of interest [131]. The sequence of the prime side (P') (Schechter and Berger nomenclature [151]) of the cleaved peptide is directly derived from its identification by mass spectrometry analysis whereas the corresponding non-prime side (P) sequence is inferred through bioinformatics analysis using the WebPICS tool [136]. PICS is a high throughput proteomics technique that has been proved to be very successful in the determination of amino acid preferences of several proteases [152-154], including non-plant APs [4, 155].

In Figure 35 is a schematic representation of the PICS workflow. The process starts with the incubation of a proteome-derived peptide library with the protease of interest (in the specific case of rRLR1 was a 5 h incubation). The new peptides generated by the protease of interest are then biotinylated and purified. Since all N-terminals were blocked during the generation of the proteome-derived library, the neo N-terminals resulting from protease activity are the only ones who react with biotin. Therefore, purification is achieved by a biotin-streptavidin pull-out, followed by salt removal using a RP-HPLC C-18 column. The prime side cleavage products are directly identified in the LC-MS/MS analysis and the non-prime side sequences are inferred bioinformatically (described in the Methods section).

For rRLR1 PICS specificity profile we used two peptide libraries derived from the digestion of human THP-1 cell protein extracts with either Trypsin or GluC. Both peptide libraries were independently incubated with purified rRLR1 and it was possible to identify 1435 C-terminal cleavage products from the trypsin library and 729 C-terminal cleavage products from the GluC library (listed in Table S1 and S2 respectively, supplementary results). These results were then analyzed using the WebPICS tool [136] where the corresponding non-prime side sequences were obtained. The complete cleavage specificity profiles are shown in Figure 36 and there is a good agreement between the results obtained from these two peptide libraries. At P1 position, rRLR1 strongly preferred leucine as well as aromatic amino acids like

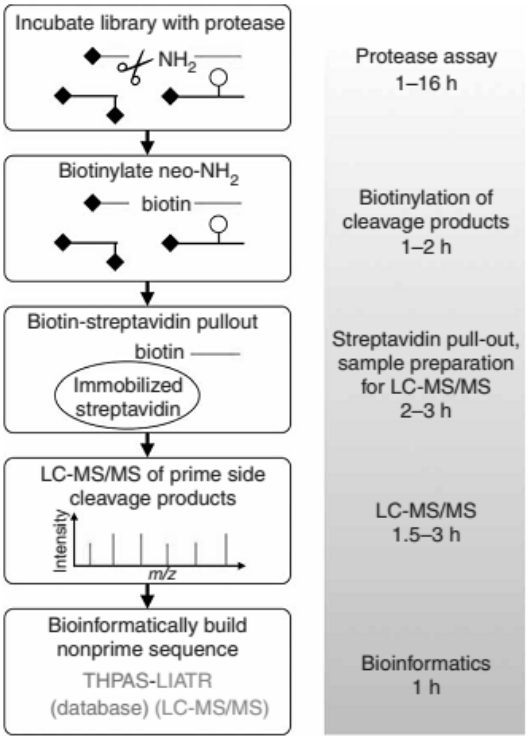


Figure 35 – PICS schematic workflow. Schematic representation of PICS workflow starting from the protease assay where the protein of interest is incubated with the peptide library. The protease assay is followed by the biotinylation of the neo N-terminals which allows their purification by a Biotin-Streptavidin pullout. After this purification, all salts must be removed from the samples and then they are ready for LC-MS/MS identification of the prime side cleavage products. The nonprime sequence is then inferred by a bioinformatics analysis. It is noteworthy that prior the protease assay the peptide libraries have to be prepared but in our case they were already prepared at the beginning of our experiment. Adapted from [131].

phenylalanine and tyrosine. Interestingly, the neutral amino acid asparagine and positively charged lysine could also be accommodated at this position. Branched aliphatic amino acids like isoleucine, valine and alanine, neutral amino acids (serine, threonine, glutamine), as well as tryptophan, glycine and the positively charged histidine were all underrepresented, whereas proline was never found at this position. At the P1' position of the scissile bond, a strong preference for the aromatic amino acids phenylalanine and tyrosine was observed although the positively charged histidine was also shown to be overrepresented. Proline was again underrepresented at this position and leucine was slightly disfavored. Interestingly, and despite the preference for hydrophobic residues like isoleucine, valine and carboxyamidomethylated cysteine (modified during library preparation) at P2, it was also striking the overrepresentation of the charged amino acids lysine and glutamate in this position. Again, proline was almost absent at P2, together with tryptophan, tyrosine, leucine and alanine. At P2', the preference for accommodating hydrophobic amino acids was maintained (valine, alanine, carboxyamidomethylated cysteine), although other aliphatic amino acids like methionine and leucine, as well as tryptophan, aspartate and lysine were

disfavored at this position. At positions more distant from the cleavage site (P3, P4, P3' and P4'), rRLR1 appears to be less stringent. At P3, rRLR1 showed a slight preference for the hydrophobic amino acids phenylalanine, tyrosine and modified cysteine but, as previously shown for P2, there were also charged amino acids overrepresented in this position (in this case, histidine and glutamate). At P3', there was a slight preference for alanine, glycine, lysine and asparagine and there were no major underrepresented amino acids. Interestingly, at P4´ position a preference for proline and asparagine was also observed.

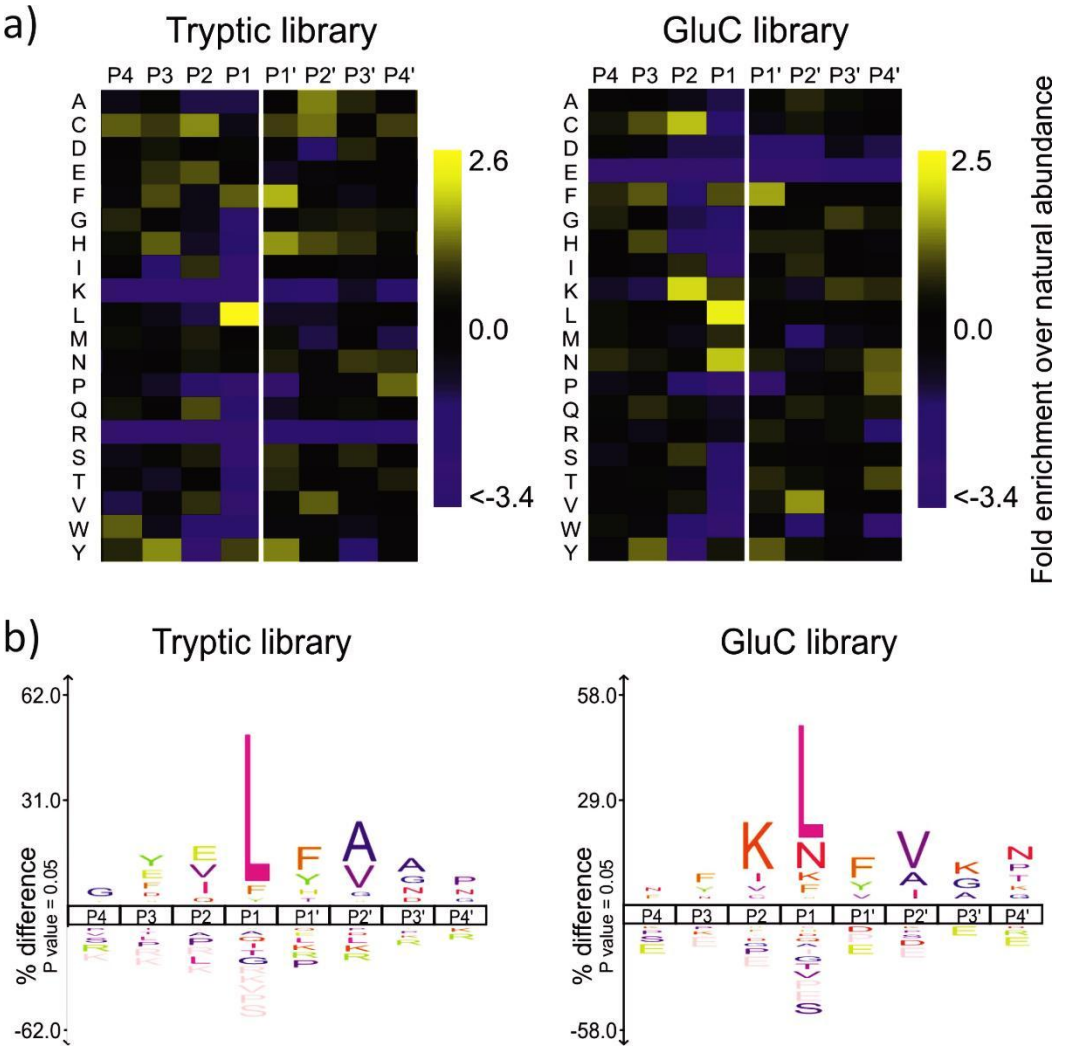


Figure 36 – rRLR1 specificity preferences. rRLR1 was incubated with either a tryptic or GluC peptide library derived from human THP1 cells, in a 1:20 ratio (enzyme:library), at pH 4.0. The average amino acid occurences in P4-P4' were calculated from one experiment for each library. **a)** Specificity profiles represented in the form of two-dimensional heatmaps of log(2) transformed values of fold-enrichment over natural abundance of amino acids. **b)** Specificity profile represented in the form of % difference IceLogos. Horizontal axis in IceLogos represents the amino acid position and vertical axis shows the over- and under-representation of amino acid occurrence compared with the Swiss-Prot *Homo sapiens* protein database. Cysteines are carboxyamidomethylated and lysines are dimethylated.

Given the unexpected preferences for asparagine and lysine at P1 as well as lysine and glutamate at P2 positions, we further evaluated subsite cooperativity between P3-P3' in an attempt to help us elucidate some of the observed preferences. Using the WebPICS tool we generated the dependency plots illustrated in Figure 37 that describe the occurrence of the different amino acids at each site when one specific amino acid is fixed at a certain position. The amino acids in positive cooperative relation with the fixed amino acid are displayed in green whereas those that stay in negative cooperativity are displayed in red. When considering the sequences with asparagine at P1, we observed a strong preference for accommodating isoleucine and leucine at P2' position whereas alanine was disfavored at this position. In fact, from the 53 unique cleavage sites containing asparagine in P1 from the GluC library, 16 had isoleucine and 23 had leucine in P2'. However, only one peptide had alanine in P2' (of a total of 50 identified cleavage sites with alanine in P2'). A similar trend was observed for the cleavage sites identified in the tryptic library (Figure 37), including the apparent mutual exclusion between asparagine in P1 and phenylalanine in P1' position (for the GluC library results, of the unique 44 cleavage sites with phenylalanine in P1' none contained asparagine in P1; for the tryptic library, of the 61 cleavage sites with phenylalanine in P1' only one had an asparagine in P1. Focusing only on the GluC library results, it was also possible to observe some positive cooperativity between lysines in positions P1 and P2 (Figure 37).

Taken together, these results suggest that rRLR1 displays unique specificity preferences beyond the somewhat expected preference for accommodating hydrophobic amino acids in S1 and S1' subsites.

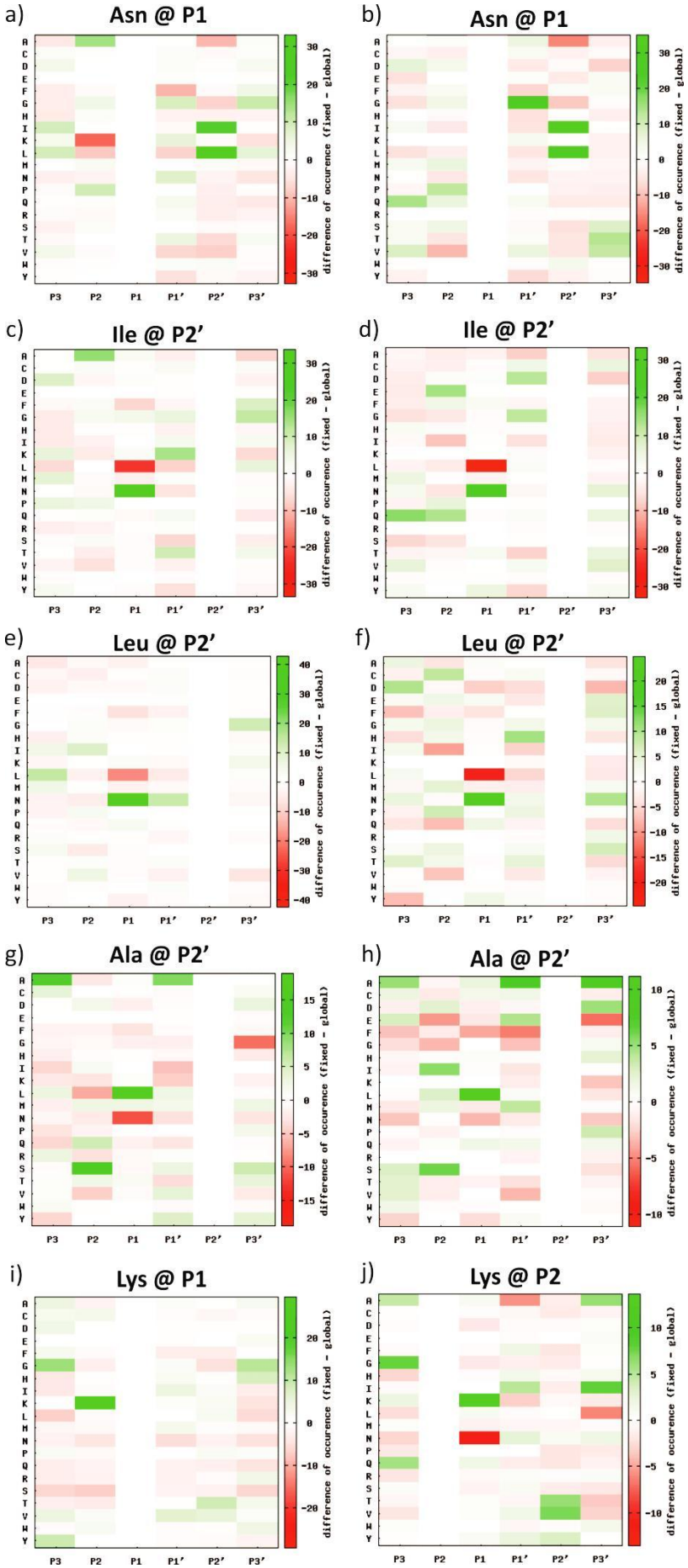

a)
Asn @ P1
b)
Asn @ P1
c)
Ile @ P2'
d)
Ile @ P2'
e)
Leu @ P2'
f)
Leu @ P2'
g)
Ala @ P2'
h)
Ala @ P2'
i)
Lys @ P1
j)
Lys @ P2
P3 P2 P1 P1' P2' P3'
difference of occurence (fixed - global)

Figure 37 – Subsite cooperativity analysis for rRLR1. Using the WebPICS tool we tested the cooperativity between amino acids by fixing one residue at a selected position to visualize the amino acid residues that have a positive (green) or negative (red) cooperative relation with each fixed residue. This analysis was performed for peptides identified when using Tryp (**a, c, e, g**) and GluC (**b, d, f, h**) peptide libraries, with the exception of panels **i)** and **j)** that were obtained with GluC library results only. Subsite cooperativity was analyzed by fixing Asn at P1 (**a** and **b**), Ile at P2' (**c** and **d**), Leu at P2' (**e** and **f**), Ala at P2' (**g** and **h**), Lys at P1 and P2 (**i** and **j**). **a)** Of the 53 unique cleavage sites that contain Asn at P1, in P2' we found 16 with Ile, 23 with Leu and only one with Ala. **b)** Of the 22 unique cleavage sites that contain Asn at P1, in P2' we found 9 with Ile, 7 with Leu and only one with Ala. **c)** Of the 34 unique cleavage sites that contain Ile at P2', 16 had Asn at P1. **d)** Of the 30 unique cleavage sites that contain Ile at P2', 9 had Asn at P1. **e)** Of the 42 unique cleavage sites that contain Leu at P2', 23 had Asn at P1. **f)** Of the 27 unique cleavage sites that contain Leu at P2', 7 had Asn at P1. **g)** Of the 50 unique cleavage sites that contain Ala at P2' in GluC library, 1 had Asn at P1. **h)** Of the 104 unique cleavage sites that contain Ala at P2', 1 had Asn at P1. **i)** Of the 39 unique cleavage sites with Lys at P1, 20 had Lys at P2 (GluC library). **j)** Of the 85 unique sites with Lys at P2, 20 had Lys at P1 (GluC library).

- ***rRLR2 large scale production and purification in N. benthamiana***

In the small-scale expression screening using *N. benthamiana* as the expression system we demonstrated that this host is also suitable for rRLR2 production and that it was possible to obtain good amounts of protein with all four constructs tested. We then decided to test a large scale production and purification of the two constructs without the putative pro-segment, ΔP_RLR2-Strep and ΔP_RLR2-His, both inserted in the A1TMVα expression vector.

We started by the expression and purification of the StrepTag construct and in Figure 38 is described the procedure and the results of one representative purification. The procedure was similar to that described for rRLR1 production. The amount of protein that we could retrieve at the end of the StrepTrap affinity chromatography was very low and the fractions contained other contaminant proteins. Moreover, and as described for rRLR1, all subsequent attempts for further purifying rRLR2 by other chromatographic approaches were unsuccessful. Also, no proteolytic activity was detected in these samples.

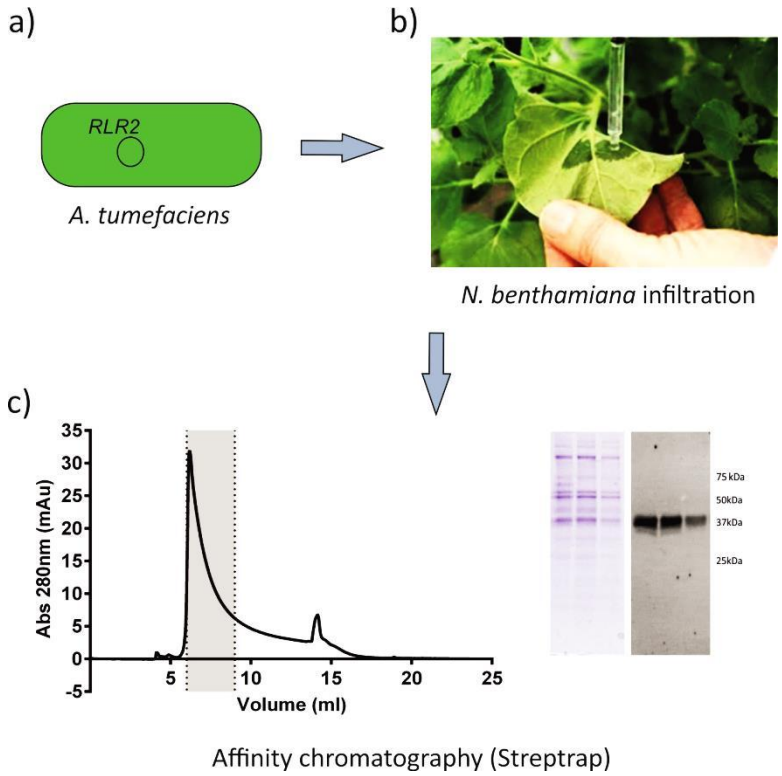


Figure 38 – Production and purification of ΔP_RLR2-Strep from *N. benthamiana* leaves. a) and **b)** *A. tumefaciens* harboring ΔP_RLR2-Strep inserted in A1TMVα expression vector were infiltrated in 5-w-old WT *N. benthamiana* leaves; **c)** Infiltrated leaves (15-20 grams) were harvested 3 days after infiltration and total soluble protein was extracted and applied to a StrepTrap column. Eluted samples were then analyzed by Coomassie-stained SDS-PAGE and Western blot. The amount of protein obtained was low and not completely pure.

Since the purification of the StrepTag fusion failed, we moved forward to the production and purification of ΔP_RLR2-His. The production procedure was the same, and Figure 39 shows the results of one representative purification procedure. TSP extract obtained from 15-20 g of leaves was applied into a HisTrap column (a)); after elution, samples were pooled and dialyzed for further purification using an anionic exchange column (MonoQ) (b)). However, most of the protein was lost in this step and this was consistent in every attempt made to purify this truncated form of rRLR2.

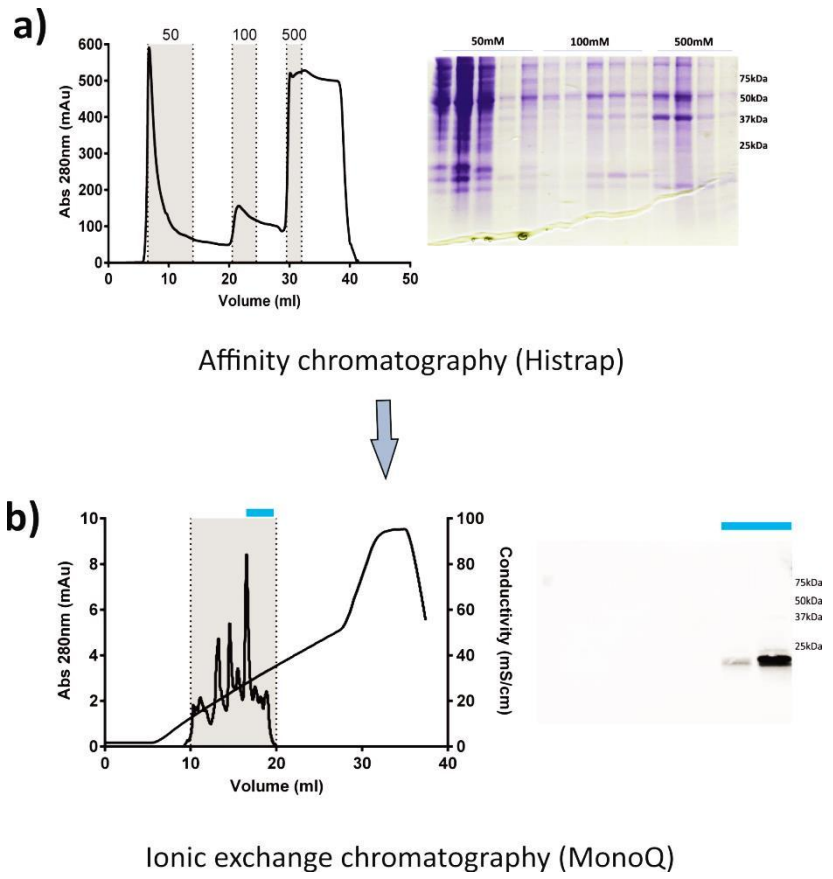


Figure 39 – Production and purification of ΔP_RLR2-His from *N. benthamiana* leaves. a) TSP extract prepared from 15-20 g of infiltrated leaves was applied into a HisTrap and protein was eluted with 3 imidazole concentrations (50, 100 and 500 mM). Eluted fractions were analyzed by Coomassie-stained SDS-PAGE. **b)** All fractions were combined, dialyzed overnight and applied into an anionic exchange column (MonoQ). Fractions corresponding to the shaded peak of the chromatogram were analyzed by Western blot.

As illustrated by the results described in this section, it was not possible to obtain purified fractions of rRLR2 expressed in *N. benthamiana* for both fusion proteins during the execution of this thesis, suggesting that additional expression and purification trials need to be performed.

3.4 Discussion

Currently, there are a large variety of organisms that can be used for recombinant protein production (expression hosts) ranging from bacteria and yeasts to mammalian cell lines or even transgenic plants or animals, each organism having advantages and disadvantages [156]. Because of their simplicity, bacteria (especially *E. coli*) were the first organisms used for the production of recombinant proteins. Bacteria can be easily handled and can grow very quickly in a relatively low cost media, allowing the production of considerable amounts of recombinant protein in a short period of time [138]. Because of these features and also because *E. coli* proved to be suitable for the production of other typical and atypical APs in our laboratory [53, 126], we first attempted to produce rRLR1 and rRLR2 using different strains of *E. coli.* We decided to test the production of the soluble form of both proteases, thereby avoiding the lengthy refolding procedure used in previous reports where the proteins were expressed in the form of inclusion bodies and then refolded prior purification [53, 126]. As demonstrated by our results, *E. coli* was not a suitable host for the production of these enzymes probably due to the fact that this organism is unable to perform post-translational modifications in its cytoplasm, such as disulfide bond formation [127]. Although we tried several strategies to overcome this limitation we could not obtain suitable amounts of rRLR1 or rRLR2 so we moved on to another expression system.

Since both RLR1 and RLR2 are plant proteins we assumed that a plant-based expression system would be the most suitable to express both proteins and, in collaboration with Dr. Herta Steinkellner, we had the opportunity to test the magnICON® expression system [129]. As discussed, the magnICON® consists in viral-based expression vectors that were engineered to produce high levels of recombinant proteins in *Nicotiana benthaminana* leaves [129]. These vectors promote a transient expression of the recombinant protein which makes magnICON® a very versatile and fast expression system. In fact, this system has been extensively used by different research groups to express a broad range of different proteins and it was very successful so far [140-142]. In our case, the small scale expression results proved that this was indeed a suitable expression system to produce both proteins so we decided to use this system for large scale production of both rRLR1 and rRLR2.

In what concerns rRLR2 production and purification, all constructs tested so far proved to be unsuccessful. Despite the very promising results obtained in the small scale expression screening with the magnICON® expression system, we were never able to obtain sufficient amounts of protein at the end of the purification procedure. Therefore, it was not possible to

pursue with the biochemical characterization or the PICS profile of rRLR2 during the course of this thesis project.

Regarding the production and purification of rRLR1, we decided to test first the StrepTag-fusion construct without the pro-segment (ΔP_RLR1-Strep). However, in parallel to what was observed for rRLR2, the protein was probably very unstable, resulting in negligible amounts of protein after the first chromatographic step. Moreover, all attempts performed to further purify ΔP_RLR1-Strep failed. This may have resulted either due to the lack of the prosegment or to some instability caused by the presence of the StrepTag at rRLR1 C-terminal which may (together or independently) impair its correct folding. When the full-length protein (without its signal peptide (RLR1-His)) was produced, the results were more consistent and, through a combination of an affinity chromatography followed by an anionic exchange chromatography, we could successfully purify rRLR1 in its glycosylated form. In fact, at the end of the purification procedure it was possible to obtain one fraction enriched mainly in the pro-protein and a second fraction with a mixture of pro-rRLR1 and its processed form with no pro-peptide, which further demonstrated that this protein was able to undergo processing *in vivo*. In typical APs the removal of the prosegment is normally an autocatalytic event [6] but in the case of rRLR1 further studies are still necessary to understand the nature of this processing. Moreover, N-terminal Edman sequencing revealed that rRLR1 prosegment was shorter than initially predicted. In fact, these differences in the predicted and actual prosegment sequences might impact our ability to produce the recombinant form of these enzymes without this region, for example. Supported by these findings, we cannot exclude that the 6 amino acids difference in our constructs without prosegment may have contributed to the instability observed with these truncated forms of rRLR1.

As discussed in the results section, we tried a number of different fluorogenic substrates to evaluate activity of purified rRLR1 but only the substrate [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP] was found to be cleaved by the protease. Interestingly, both rRLR1 purified samples showed proteolytic activity at the same extent which indicates that this enzyme might be active even without the removal of the pro-segment. As discussed, prosegment removal was found to be crucial for activation of most typical APs but CDR1 biochemical characterization also proved that this enzyme is active in the zymogen form [6, 53]. Taken together, these results suggest that proteolytic activity without irreversible prosegment removal might be a feature more common among atypical APs than initially anticipated. However, since there is a generalized lack of knowledge concerning the biochemical properties of these enzymes, more studies are needed to confirm this trend.

Using the substrate [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP] we were able to determine that rRLR1 is mostly active at pH 4 and that its activity is lost by pH 6. This is the typical pH profile of an AP [6] although it was very different from the pH profile of CDR1, that had its maximum activity at pH 6 towards the same substrate [53]. Other atypical APs such as ASPG1 and NANA also showed an optimal pH around 6 while nodulin 41 was described to be mostly active at pH 4.5, similar to rRLR1 [38, 40, 49]. We were also able to determine rRLR1 inhibition profile and some of the results were quite interesting. For instance, rRLR1 was not completely inhibited by pepstatin A, the canonical AP inhibitor, retaining 20% of activity. In fact, this is a distinguishing feature that separates typical APs from the atypical members identified so far, as the latter were generally found not to be fully inhibited by pepstatin A [6, 53](our own unpublished data). rRLR1 was also sensitive to redox agents such as NAD, NADP and NADPH and this is again an interesting feature since, with the exception of chlapsin [126], plant typical APs are not usually affected by these compounds. Interestingly, CDR1 was also shown to be susceptible to the presence of redox agents [6, 53]. This inhibitory effect might be linked to the high number of cysteines present in RLR1 (as in CDR1) primary structure. In fact, RLR1 without the signal peptide has 11 cysteines compared to the 6 cysteine residues present in mature cardosin A and phytepsin, the two plant typical APs whose structure was already determined [15, 52]. For these typical APs, all 6 cysteine residues are involved in disulfide bond formation. Since RLR1 has an odd number of cysteines, at least one of these cysteines may be unpaired and not involved in an intramolecular disulfide bridge. Since these amino acid residues are quite sensitive to the redox state of the environment, this unpaired cysteine might act as a redox sensor for this enzyme, suggesting that the *in vivo* function of RLR1 may be linked to a redox sensing mechanism. This hypothesis still needs to be experimentally confirmed but it is an exciting possibility, and is in line with the redox-dependent regulation of protease activity previously anticipated for the atypical CDR1 [53]. Other inhibitors such as pefabloc (serine protease inhibitor) or EDTA (metalloprotease inhibitor) had also some influence in rRLR1 activity. Somewhat unexpected, this was not totally surprising as the inhibitory effect of these compounds was also reported for rCDR1 at almost the same rate [53].

As discussed above, rRLR1 displayed activity against just one of the many substrates that we tested in the laboratory and this can imply more restricted specificity preferences for this protease when compared to typical housekeeping APs. Therefore, we decided to further explore the specificity profile of rRLR1 using a high throughput proteomics technique called Proteomics Identification of Cleavage Sites (PICS) [131]. This technique has been used for several other APs and it was already proved to be quite successful, although to our knowledge

this is the first report describing its use to study the specificity profile of an atypical plant AP [3, 4, 155]. The specificity profile obtained for rRLR1 confirms the common preference for accommodating hydrophobic amino acids in the S1 subsite (with a marked preference for leucine in this case), as described for other APs, such as cardosin A, pepsin, BACE1, cathepsin D, plasmepsins and shewasins D and A [155, 157-162], as well as the more stringent character of S1 in comparison with all other subsite binding pockets (particularly S1'), suggesting that the preferences at the P1 position are likely key specificity determinants. Also, in line with what has been previously described for most of these proteases [155, 157-159, 161-163], branched chain hydrophobic amino acids like isoleucine and valine, as well as proline and the neutral amino acids serine, threonine and glutamine were mostly disfavored at this position. Interestingly, rRLR1 also displayed a clear preference for accommodating asparagine and lysine in S1. This is quite striking since similar preferences have been previously reported for aspergillopepsin [163]. In fact, fungal APs like aspergillopepsin, rhizopuspepsin, endothiapepsin, penicillopepsin as well as SAP2 and SAP3 from *Candida albicans* differ from other eukaryotic APs because - besides favoring hydrophobic amino acids in positions P1 and P1' - they also cleave substrates with polar residues like lysine (or histidine) in P1, thereby having the ability to activate trypsinogen [164, 165]. Extensive characterization of this S1 subsite specificity in fungal APs has resulted in the identification of Asp77 and Ser79 (pepsin numbering), two residues in the active site flap, as critical for the accommodation of a basic amino acid at this position [164, 165]. Strikingly, by comparing the amino acid sequences of rRLR1 and fungal APs in the flap region, it was possible to confirm the presence of the same residues (Asp and Ser) in the atypical plant AP (Table 16) which may account for the observed preference for lysine in P1. However, additional site-directed mutagenesis studies will be required to further confirm the importance of these residues for defining rRLR1 S1 specificity (for both lysine and asparagine preferences).

Regarding P1' specificity, the results are consistent with what has been described for other APs which consists in the preference for accommodating hydrophobic amino acids and for excluding proline, also being much less stringent when compared with P1 position [155, 158, 159, 161, 162]. For positions distal to the scissile bond, it is interesting to note that rRLR1 shares the same preferences with shewasin D, shewasin A, pepsin, plasmepsins and BACE for P2' position [155, 158, 160, 162]; however, P2 and P3 positions appear to be much more specific when compared with results available for most eukaryotic APs. In P2 it was particularly interesting to note the preference for lysine and glutamate. Although more studies will be necessary to understand in more detail both primary and secondary specificity preferences of

rRLR1, these results clearly suggest some unique specificity requirements and unveil similarities with fungal APs that are totally unprecedented for plant APs.

Table 16 – Amino acid sequences of the active-site flap of several aspartic proteases.

Protein	Active-site flap	Reference
Aspergillopepsin I	WDISYGDGSSASGD	[166]
Penicillopepsin	WSISYGDGSSASGN	[167]
Rhizopuspepsin	WSISYGDGSSASGI	[168]
Endothiapepsin	WSISYGDGSSSSGD	[169]
Candidapepsin	FYIGYGDGSSSQGT	[170]
RLR1	YLYTYGDYSSTRGL	-
Cardosin A	GAIIYGTGSI-TGF	[171]
Phytepsin	AAIQYGTGSI-AGY	[15]
Porcine Pepsin	LSITYGTGSM-TGI	[172]

In grey are highlighted the amino acid residues that are responsible for AP activity against substrates with Lys at P1.

One of the goals of this project was to produce recombinant RLR1 and RLR2 in order to study their *in vitro* biochemical properties and amino acid preferences. The results of this chapter showed that although we were not able to achieve this goal for both proteins, we were able to succeed for rRLR1 showing that this enzyme is indeed an atypical AP with some distinct biochemical properties and with unique specific preferences. Our results strengthen the idea that these enzymes have quite unique properties, being very different from the typical APs and even from each other. They also further underscore the pattern of complexity among this class of plant proteases.

Chapter IV – Proteomics Studies

4.1 Introduction

In the introduction of the previous chapter we discussed the lack of knowledge about the biochemical properties of atypical APs and we emphasized the importance of their study in order to predict and understand the functions of these important enzymes. Nonetheless, to fully understand their biological functions it is not only necessary to study their biochemical properties and describe their role during the plant life cycle but also to dissect the molecular mechanisms underlying their functions.

At the beginning of this thesis we highlighted what is known about plant atypical and nucellin-like APs, their properties and their biological functions. These proteins play important roles in several key plant processes such as defense against biotic and abiotic stresses, gametophyte development and chloroplast homeostasis. However, the *in vivo* substrates as well as the molecular mechanisms underlying their functions remain mostly elusive, as the majority of studies reported so far consist of phenotypic analysis where just the biological function is described. In fact, out of all atypical and nucellin-like APs described so far only one biological substrate was identified - the BAG6 protein - shown to be processed by APCB1, triggering autophagy and inhibiting *B. cinerea* development [43]. For CDR1 and OsCDR1, for example, the authors hypothesize that these enzymes are responsible for the processing of a cell surface protein which generates a peptide elicitor that functions as a SAR signal [36, 47]. However, the identity of this cell surface protein and the mechanisms triggered by its processing still remain elusive. For all other atypical and nucellin-like APs described so far, just the function was described and in some cases it was just a proposed function. This clearly demonstrates the lack of knowledge that still exists concerning these enzymes and emphasizes the need of more integrative studies with new strategies to uncover the molecular mechanisms underlying atypical and nucellin-like APs biological functions.

High-throughput proteomics techniques have been widely used to describe plant proteomes and, more recently, even the *in vivo* substrates of proteases [173, 174]. Quantitative shotgun proteomics is one of the most commonly used proteomics techniques and allows a broad description of a given proteome. Figure 40 is a schematic representation of a quantitative shotgun proteomics workflow where two proteome samples are analyzed and compared. It is generally a very straightforward procedure where the proteome samples are separately digested with trypsin, labeled with different N-terminal labels, combined and analyzed by LC-MS/MS. Because they are labeled differently, both proteomes can be analyzed simultaneously and the differences between them quantified. With this relatively simple

experiment it is possible to obtain significant information on the differences between the given proteomes, thereby allowing a more thorough analysis of the molecular pathways which may be altered between samples.

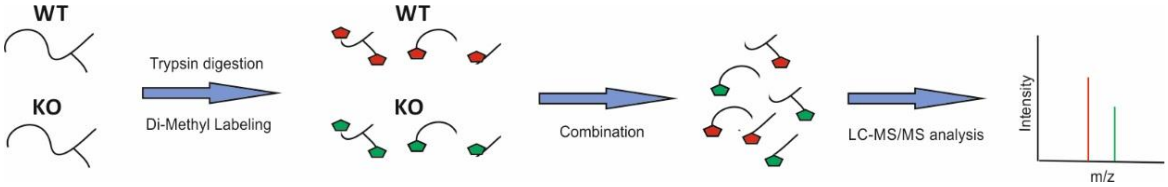


Figure 40 – Quantitative shotgun proteomics analysis workflow. From left to right: selected proteome samples of WT and KO lines are separately digested with trypsin and the N-terminal and lysine side chains are then labeled using different labels. The two samples are subsequently combined and analyzed by LC-MS/MS. The same peptide present in each sample will be separated in the spectra because it has been differentially labeled, thereby allowing peptide quantification.

As described in the previous sections, both RLR1 and RLR2 were shown to play an important role in primary root development and lateral root formation, as evidenced by the shorter primary roots and reduced number of lateral roots observed for both KO lines when compared with WT plants. Moreover, our results suggest that both proteins may be involved in two different pathways, with RLR1 apparently more important under normal growth conditions and RLR2 more important under nitrogen deprivation (see Chapter II). To try to unveil the molecular mechanisms underlying RLR1 and RLR2 functions in primary root development and lateral root formation we performed quantitative shotgun proteomics using WT and KO root proteomes. The results of these experiments will be detailed in the following sections. Both KO proteomes displayed interesting proteomic differences when compared with the WT proteome. Also, it was possible to identify some differences between RLR1 and RLR2 KO proteomes; however, it is worth to note that some proteins were identified in both KO samples, suggesting that both proteases may also participate in similar molecular pathways. With these results we expect to contribute to open new avenues of research that will help to understand the molecular mechanisms by which RLR1 and RLR2 exert their function during root development.

4.2 Methods and Materials

- ***Root proteome isolation***

Root proteins were extracted according to Zhang *et al.* with some modifications [175]. 150-200 mg of root material from 10 day-old seedlings were ground with a mortar and pestle under liquid nitrogen and resuspended in 600 µL of Extraction Buffer (EB) (3 M guanidine hydrochloride, 100 mM HEPES pH 7.5, 150 mM NaCl, 1 mM PMSF, 1X Halt Protease Inhibitor Single-Use cocktail, Thermo Scientific). After a 10 min incubation on ice, samples were centrifuged in a bench centrifuge at full speed at 4 ^{o}C for 15 min, supernatant was collected and the pellet was discarded. The supernatant contained all the extracted proteins so we continued the experiment with this fraction.

Extracted proteins were then reduced with 10 mM DTT (from a fresh 1 M DTT solution) for 1h at 50 ^{o}C and after cooling down proteins were alkylated with 20 mM Iodoacetamide (from a 0.5 M fresh solution) for 45 min in the dark and at room temperature. The remaining iodoacetamide was quenched with 10 mM DTT for 10 min at room temperature.

Proteins were then precipitated with a methanol/chloroform mixture. 4 volumes of ice cold methanol were added to each fraction followed by 2 volumes of ice cold chloroform and 3 volumes of ice cold double distillated water. The mixture was vortexed and centrifuged at full speed for 5 min at 4 ^{o}C. Proteins precipitate in the interface between the aqueous and organic phases. The upper phase was carefully removed without disturbing the interface and 1 mL of ice cold methanol was added. Samples were centrifuged in a bench centrifuge at full speed for 10 min at 4 ^{o}C and the supernatant was carefully removed and discarded. The protein pellet was washed 2-3x with 1 mL of ice cold methanol, left to dry until near dryness and resuspended in 20 to 40 µl of 100 mM NaOH. pH was adjusted to 7.5 with 1 M HEPES solution and protein was quantified using the BCA method according to the manufacturer's instructions.

- ***Quantitative shotgun proteomics***

  - ***Trypsin digestion***

Prior trypsin digestion, protein concentration of each sample was adjusted to be equal in all samples by adding small volumes of 1 M HEPES solution. pH was then adjusted to 8 with small volumes of 1 M NaOH and trypsin (Pierce™ Trypsin Protease, MS Grade, ThermoFisher Scientific) was added at a ratio of 1/100 (w/w) (trypsin/proteome) from a trypsin stock at 1

g/µL. Samples were incubated overnight at 37 ^{o}C and, after this incubation period, more trypsin was added at the same ratio and the samples were incubated for an additional 4 h at 37 ^{o}C to ensure complete trypsin digestion. Upon digestion, the proteome samples were ready to be labeled.

- ***Proteome labeling***

WT and KO root samples were labeled in separate vials according to the dimethyl labeling procedure reported by Boersema *et al.* with some modifications [176]. Two dimethyl isotopic labels were used: WT was labeled with the light label (CH_2O) while the KO samples were labeled with the medium label (CD_2O). We added 40 mM (from a fresh 1 M stock) of either CH_2O or CD_2O to each proteome sample followed immediately by the addition of 20 mM of $NaBH_3CN$ (from a 1 M stock solution). The mixture was incubated for 2 h at room temperature in a fume hood and the labeling reaction was quenched by the addition of 100 mM Tris pH 6.8 (from a 1 M stock solution). Samples were then combined (WT + *RLR1* KO and WT + *RLR2* KO), acidified with a Solution A (2% acetonitrile (ACN), 1% formic acid (FA)) to pH 3-4 and subjected to a double layer stage tip using C-18 membrane according to [177]. Briefly samples were loaded in a double layer C-18 stage tip, previously activated with 80% ACN, washed with Solution A and eluted with Solution B (80% ACN, 1% FA). Samples were vacuum-evaporated until near-dryness and resuspended in 30 µL of Solution A. At this point samples were ready for LC-MS/MS analysis.

- ***LC-MS/MS data acquisition***

Peptide identification by LC-MS/MS analysis was performed using an UltiMate3000 nano-HPLC system (Waltham, MA, USA) with Thermo PepMap 100 NanoTrap column, P/N 164564a connected to a Bruker Impact II Q-TOF mass spectrometer (Billerica, MA, USA). Samples were loaded into the C-18 column and peptides were eluted using a 5-40% gradient of organic phase. Buffer A was 0.1% formic acid whereas Buffer B was 100% acetonitrile supplemented with 0.1% formic acid.

MS data was acquired using the Bruker Control Software version 4.0.15.3248. Data-dependent acquisition was performed based on a 0.25s (4Hz) MS survey scan in m/z range 300-1750. A top 17 MS/MS method was used for fragmentation of precursor ions. Nitrogen was used as collision gas with the quadrupole-collision energy set to 5 eV and a captive spray capillary voltage of 1600 V.

- ***Data analysis***

Peptides were identified by searching the raw MS/MS data against the "TAIR10_pep_20101214_updated" database (containing 35.386 sequences) using the andromeda search engine integrated into the *MaxQuant* environment (version 1.5.3.30) [178, 179]. The false discovery rate for identified peptide sequences was set to < 1%. Searching parameters were set by using Trypsin/P specificity for *in silico* digestion with a maximum number of 2 missed cleavages. Light dimethylation (+ 28.03130 Da) and medium dimethylation (+ 32.05 Da) were set as fixed modifications on lysines and N-termini for quantification. Carbamidomethylation of cysteine (+ 56.0214 Da) was set as a fixed peptide modification. Oxidation of methionine (+ 15.9949 Da) and acetylation of protein N termini (+ 42.0105 Da) were set as variable peptide modifications. The re-quantify feature in MaxQuant software was enabled.

All the protein data resulting from this search were then analyzed using the Perseus software version 1.5.2.6. The lists of the identified proteins in each condition were loaded in Perseus together with the ratios between the medium (KO samples) and light (WT samples) for each identified protein. These ratios were converted into logarithmic values (base 2) and the potential contaminants, reverse hits and proteins only identified by site were filtered out of the final lists. Finally, to identify proteins that were significantly enriched or depleted in RLR1 or RLR2 KO proteomes in the different conditions tested, the distributions of the KO/WT ratios (medium/light labels) were divided into 10 quantiles. Proteins were considered to be significantly enriched in the KO when present in quantiles 9 and 10 (Q9-10), and significantly depleted when present in quantiles 1 and 2 (Q1-2).

4.3 Results

- **RLR1 *KO line root proteome has some interesting features***

As we stated before, our goal for this chapter was to identify the molecular pathways that were de-regulated by the absence of RLR1 or RLR2. To achieve this goal, we performed quantitative shotgun proteomics comparing KO with WT root proteomes. Moreover, and since we described a RLR-associated phenotype not only under normal growth conditions but also under N deprived conditions, we evaluated also proteomic differences between KO and WT plants under this abiotic stress.

In the introduction of this chapter we described the workflow of a general quantitative shotgun proteomics experiment (see Figure 40), which was also herein followed to compare root proteomes. Our goal was to quantify the proteome differences and the most common way to achieve this is by using isotopic labeling where protein N termini resulting from different samples are labeled differently [180]. Samples differentially labeled can then be combined, analyzed in a single MS analysis and the differences quantified by comparing the relative abundance of the same peptide in each sample [181]. There are several ways to perform isotopic labeling and several isotope labels [182-184] but in our work we decided to use the dimethyl labeling technique according to Boersema *et al.* [176]. This is a very fast and relatively cheap labeling approach that allows the combination of up to 4 different samples by using combinations of several isotopomers of formaldehyde and cyanoborohydride. This technique is based on the reaction of peptide free primary amines (free N-termini or lysine side chains) with different forms of formaldehyde that lead to the generation of a Schiff base that is rapidly reduced in the presence of cyanoborohydride [176, 184]. In our work, WT proteome was treated with a combination of regular formaldehyde (CH_2O) and cyanoborohydride ($NaBH_3CN$) which generates a mass increase of 28 Da per primary amine (light label). In the case of the KO proteomes they were treated with a combination of deuterated formaldehyde (CD_2O) and normal cyanoborohydride ($NaBH_3CN$) generating a mass increase of 32 Da per primary amine (medium label).

In this section we will present the results regarding *RLR1* KO line and in these experiments we used root proteomes resulting from two different conditions: *RLR1* KO vs WT, grown either in normal or ammonium nitrate deprived (abiotic stress) media. Three and two biological replicates were analyzed for normal growth and abiotic stress condition, respectively. In Table 17 are depicted the number of proteins that were successfully identified

and quantified in each replicate and, for all conditions, it was possible to identify and quantify a large number of proteins in each experiment.

Table 17 – Number of de-regulated proteins identified and quantified in each biological replicate of *RLR1* KO line.

Growth Condition	Replicate	Number of proteins
Normal	1	1505
	2	1752
	3	1456
Abiotic Stress	1	1145
	2	1077

To confirm whether there would be a good agreement between each biological replicate, Venn diagrams were generated to analyze the distribution of the identified and quantified proteins in each replicate. Moreover, the KO vs WT ratios of each identified protein were then converted to logarithmic values (log 2) and histograms were created to evaluate the general distribution of the identified proteins. The Venn diagrams and histograms for the samples grown under normal or abiotic stress conditions are depicted in Figure 41 and Figure 42, respectively. For both growth conditions we had a good agreement between the biological replicates since 1128 proteins were quantified in all three biological replicates grown under normal conditions, while 981 proteins were identified in the two biological replicates grown under abiotic stress conditions. Moreover, the histograms of all biological replicates show a Gaussian distribution which is a good indication of a successful shotgun experiment.

In order to find the proteins that were significantly depleted or enriched in *RLR1* KO line, histograms were divided in 10 quantiles, and we assumed as significantly depleted proteins those present in quantiles 1 and 2 (Q1 and Q2) (highlighted in green in the histograms of Figures 41 and 42), whereas the significantly enriched corresponded to those present in quantiles 9 and 10 (Q9 and Q10) (highlighted in red in the histograms of Figures 41 and 42). Moreover, only the proteins that were found to be significantly enriched or depleted in all biological replicates grown under the same growth condition were considered for further analysis. In Table S3 and Table S4 (supplementary material) are listed the significantly enriched and depleted proteins found in *RLR1* KO root proteome when plants were grown either under normal or abiotic stress conditions, respectively. A total of 37 proteins were found to be significantly de-regulated in *RLR1* KO line grown under normal condition (18 depleted and 19

enriched) while 129 proteins were significantly de-regulated in *RLR1* KO line grown under abiotic stress conditions (60 depleted and 69 enriched).

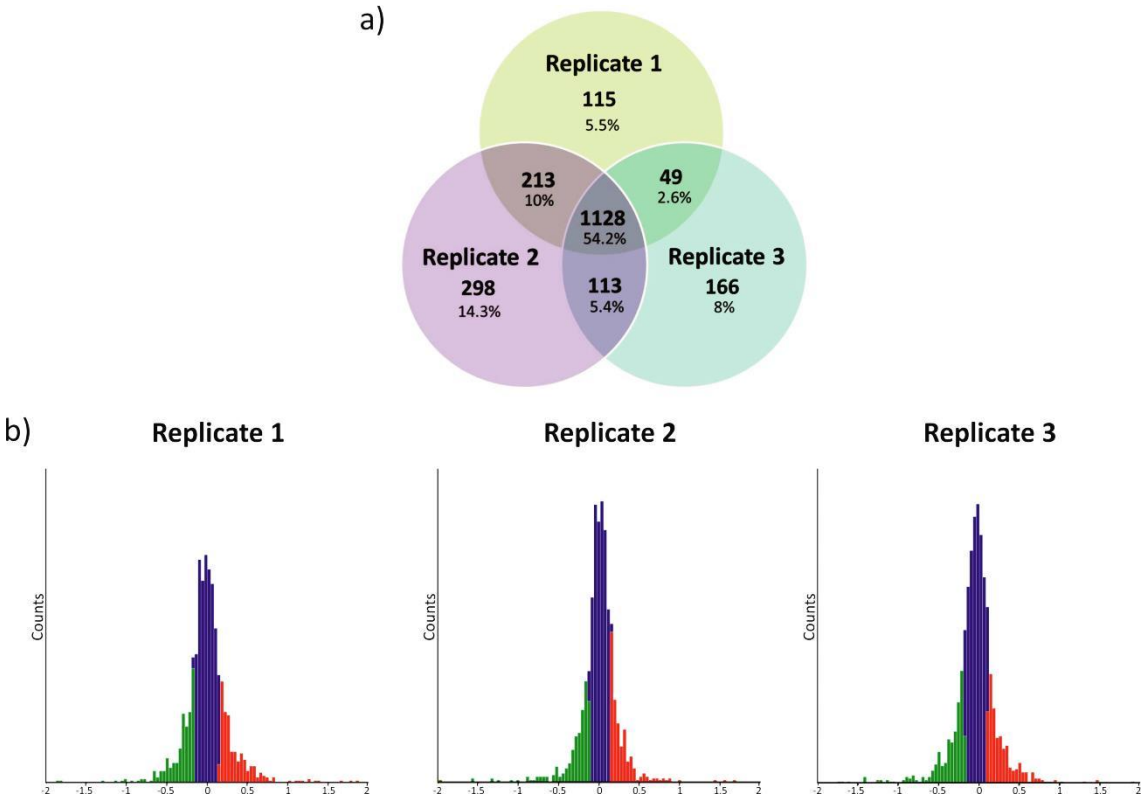


Figure 41 – Distribution of the identified and quantified proteins in *RLR1* KO and WT root proteomes grown under normal conditions. a) Venn diagrams of all the identified and quantified proteins in this experiment distributed by the biological replicates where they were identified. 1128 proteins were simultaneously identified in the three biological replicates. **b)** *RLR1* KO vs WT ratio distribution of quantified proteins in each experimental replicate. Histograms were divided in 10 quantiles and proteins present in Q1 and Q2 (depleted in *RLR1* KO) are highlighted in green whereas proteins present in Q9 and Q10 (enriched in *RLR1* KO) are highlighted in red. A Gaussian distribution was observed in all biological replicates, a good indicative of a successful shotgun experiment.

We also noticed that there were several proteins consistently present among the significantly de-regulated proteins identified in both growth conditions. To further evaluate this, a Venn diagram was generated where all the de-regulated proteins were grouped according to their de-regulation (enriched or depleted) and also according to the growth condition in which they were described. Figure 43 illustrates this Venn diagram, and it was possible to confirm that there were indeed similarities between the identified proteins since 6 proteins were found enriched in both conditions and other 5 proteins depleted in both conditions. These 11 shared proteins are listed in Table 18 and we will discuss some of them in more detail in the last section of this chapter.

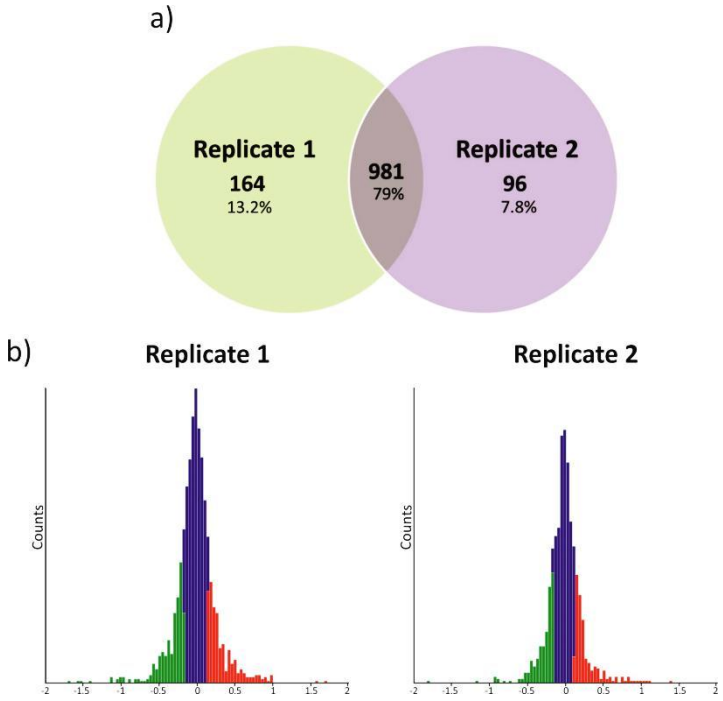


Figure 42 - Distribution of the identified and quantified proteins in *RLR1* KO and WT root proteomes grown under abiotic stress conditions. a) Venn diagrams of all the identified and quantified proteins in this experiment distributed by the biological replicates where they were identified. 981 proteins were simultaneously identified in both biological replicates. **b)** *RLR1* KO vs WT ratio distribution of quantified proteins in each experimental replicate. Histograms were divided in 10 quantiles and proteins present in Q1 and Q2 (proteins depleted in *RLR1* KO) are highlighted in green whereas proteins present in Q9 and Q10 (enriched in *RLR1* KO) are highlighted in red. A Gaussian distribution was observed in all biological replicates, a good indicative of a successful shotgun experiment.

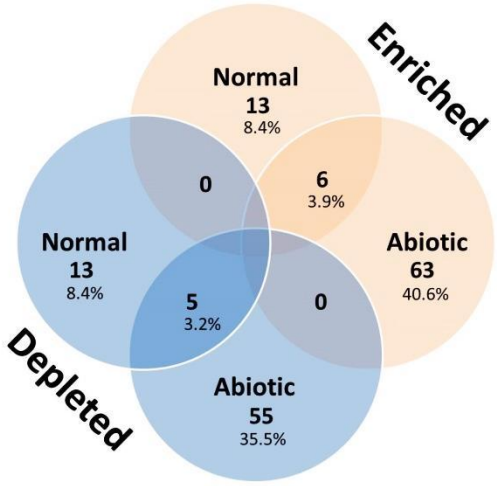


Figure 43 – Venn diagram of all enriched and depleted proteins described for *RLR1* KO line separated by growth conditions. All 166 significantly enriched (red circles) and depleted (blue circles) proteins detected in *RLR1* KO line were separated by plant growth condition.

Table 18 – De-regulated proteins found in *RLR1* KO line grown under both growth conditions.

Gene ID	Gene description	Normal conditions		Abiotic Stress	
		E	D	E	D
AT1G50480	10-Formyltettrahydrofolate Synthase				
AT2G01520	MLP-Like Protein 328				
AT4G08770	Peroxidase 37				
AT1G70850	MLP-Like Protein 34				
AT1G54010	GDSL-Like Lipase 23				
AT3G44310	Nitrilase 1				
AT2G02390	Glutathione S-Transferase 18				
AT3G44300	Nitrilase 2				
AT2G43610	Chitinase Family Protein				
AT5G05270	Chalcone Isomerase Like				
AT4G25890	60S acidic Ribossomal Protein Family				

E – Enriched; D – Depleted
Shaded rectangles represent depleted or enriched proteins under a given growth condition

In order to start exploring the molecular functions that were found de-regulated in the KO lines under both growth conditions, a gene ontology (GO) analysis was performed for the significantly de-regulated proteins identified in *RLR1* KO line. The PANTHER classification system (Protein ANalysis THrough Evolutionary Relationships) (www.pantherdb.org) was used, allowing the classification of proteins under different categories [185, 186]. The proteins were classified according to their putative molecular functions and this classification is defined by the function of a protein by itself or with directly interacting proteins at a molecular level. In Figure 44 are depicted the results of this GO analysis. The first impression is that there is a great difference between the samples grown under the two conditions. However, we have to keep in mind that we identified almost 3x more significantly de-regulated proteins in *RLR1* KO samples grown under the abiotic stress condition which may help explaining this difference. Analyzing just the de-regulated proteins identified under normal growth, 18 proteins were found to be significantly depleted in *RLR1* KO line but the GO analysis just revealed 6 different putative molecular functions grouped in 3 GO groups: Binding (comprising proteins that are involved in the selective interaction with other proteins/molecules); Catalytic activity (comprising proteins that bind to and modulate the activity of enzymes); and Structural molecule activity (comprising proteins that contribute to the structural integrity of a complex

within or outside a cell) [187]. For the enriched proteins found under the same condition, we described 19 of them and the GO analysis found 18 different molecular functions grouped into 5 GO categories: Binding, Catalytic activity and Structural molecule activity already described before; as well as Translation regulator activity (comprising proteins involved in polypeptide synthesis at the ribosome); and Transporter activity (comprising proteins involved in the movement of substances) [187]. In this case the majority of the detected functions were grouped in the "Catalytic activity" GO category.

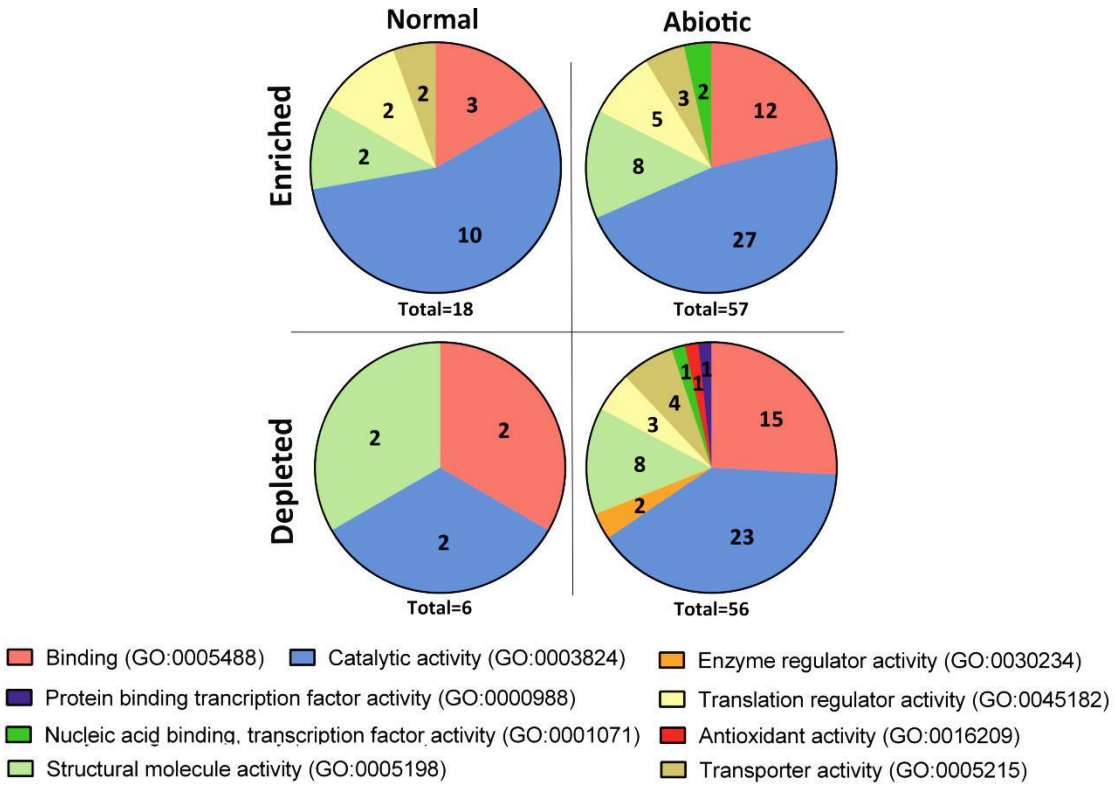


Figure 44 - Gene ontology analysis of the significantly de-regulated proteins that were identified in *RLR1* KO line grown under normal and abiotic stress conditions. Under normal growth conditions 19 and 18 proteins were found to be significantly enriched and depleted, respectively, and were used to conduct the GO analysis. Under the abiotic stress growth condition, 69 and 60 proteins were significantly enriched and depleted, respectively, and were used to perform this GO analysis. The numbers correspond to the number of predicted molecular functions for each group.

Concerning the de-regulated proteins found under abiotic stress conditions, we described 60 and 69 proteins that were significantly depleted or enriched, respectively, and the GO analysis identified 56 and 57 different putative molecular functions grouped in 9 GO categories, five of them already described in the previous analysis (normal conditions). Of the newly found 4 categories, one of them is found in both enriched and depleted proteins, "Nucleic acid binding, transcription factor activity" (comprising proteins involved in RNA and DNA transcription modulation), and the 3 remaining categories are unique for the depleted analysis: Enzyme regulator activity (comprising proteins that module enzyme activities);

Antioxidant activity (comprising proteins with an antioxidant properties); and Protein binding transcription factor activity (comprising proteins that interact and modulate transcription factors). Again, the majority of the described functions were grouped in the "Catalytic activity" GO category.

To further evaluate if there were any known or predicted interactions between RLR1 and the depleted or enriched proteins identified in this study, each group of proteins was analyzed in the STRING database (www.string-db.org), a database of known and predicted protein interactions covering 9.643.763 proteins from 2.031 organisms [188]. Figure 45 illustrates the map of protein-protein interactions among the depleted or the enriched proteins identified for *RLR1* KO line grown under normal conditions. No known or predicted interactions between RLR1 (gene ID: At2g03200) and the de-regulated proteins were found but interestingly there are already some interactions described amongst some of those deregulated proteins. Among the depleted proteins one of the predicted interaction is between two MLP-like proteins, MLP328 and MLP34, which are involved in promoting vegetative growth (MLP328) and defense responses against biotic injury (MLP34) [189, 190]. On the group of enriched proteins, there is an interaction between AT4G25890 gene product and GR-RBP2 (gene ID: AT4G13850) which are involved in RNA transcription and a cluster of interactions between the proteins TT4, CHIL, KAT5 and LACS6 which are involved in flavonoid biosynthesis (TT4) and lipid synthesis.

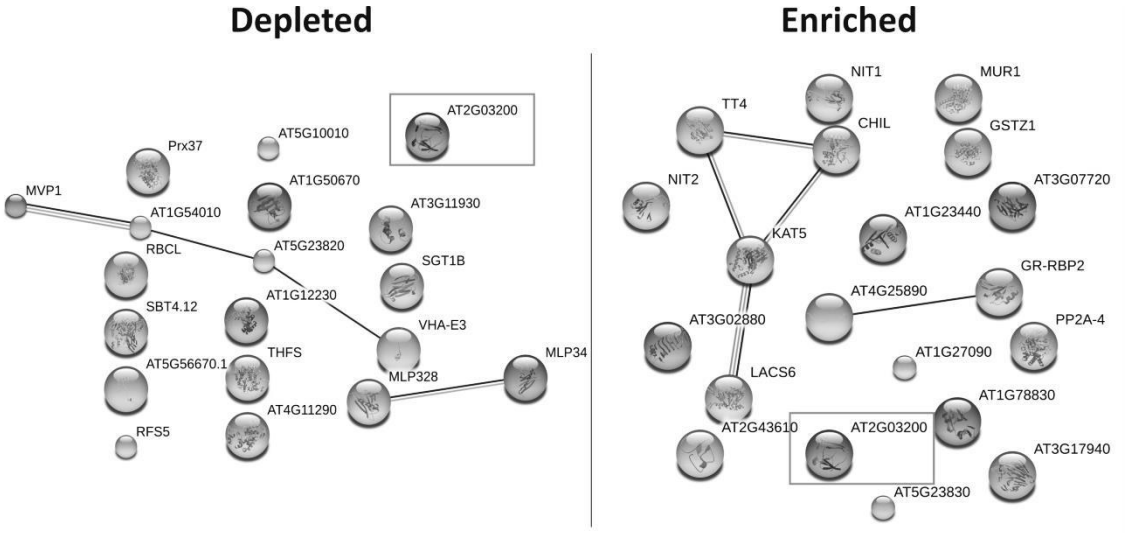


Figure 45 - Protein-protein interactions between the depleted and the enriched proteins described in *RLR1* KO line grown under normal conditions. Using the STRING database [188] we searched for interactions between RLR1 (gene ID: At2g03200) (box) and the 18 depleted or the 19 enriched proteins that were de-regulated in *RLR1* KO line grown under normal conditions. There are no described or predicted interactions between RLR1 and all these proteins but there are some known/predicted interactions between several of the de-regulated proteins.

Figure 46 depicts the known/predicted interactions between the de-regulated proteins found in *RLR1* KO line grown under abiotic stress conditions. Again, there are no interactions between RLR1 and any of the de-regulated proteins but there are several interactions between the de-regulated proteins. As highlighted in the GO analysis, it is important to keep in mind that we are analyzing 3x more proteins so it is not unexpected to detect more interactions between them. The observed interactions among de-regulated proteins will allow us to speculate about the molecular pathways that are de-regulated in *RLR1* KO line under this growth condition and we will discuss some of them in more detail in the Discussion section of this chapter.

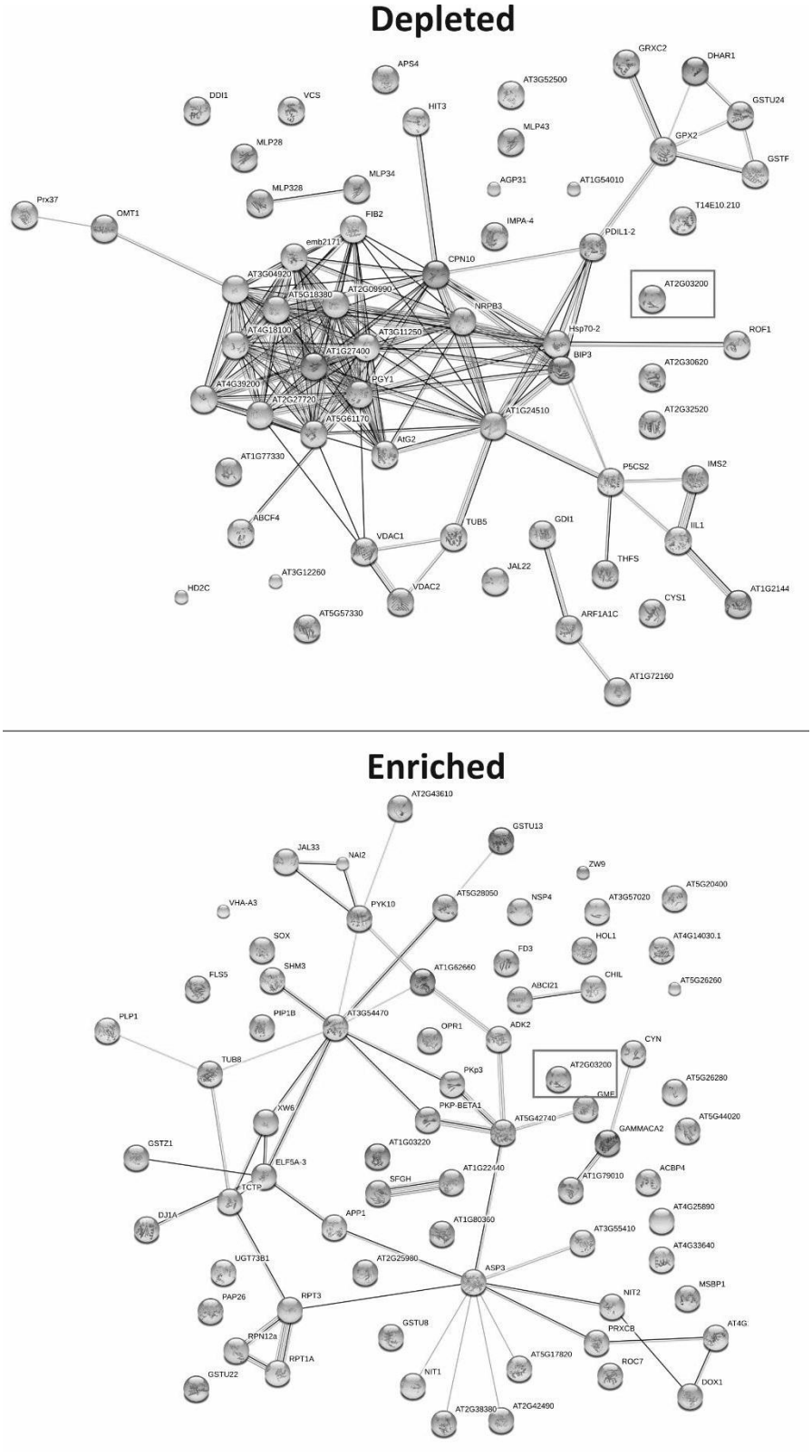


Figure 46 - Protein-protein interactions between the depleted and enriched proteins described in *RLR1* KO line grown under abiotic stress conditions. Using the STRING database [188] we looked for interactions between RLR1 (gene ID: At2g03200) (box) and the 60 depleted or the 69 enriched proteins that were de-regulated in *RLR1* KO line grown under abiotic stress conditions. There are no described or predicted interactions between RLR1 and all these proteins but there are some known/predicted interactions between several of the de-regulated proteins.

- **RLR2 *KO root protease also has some interesting features***

As stated before, we wanted to study the molecular mechanisms that were being de-regulated by the absence of both RLR1 and RLR2. In this section we will present the quantitative shotgun proteomics results concerning *RLR2* KO mutant.

We followed the same approach used for *RLR1* KO line and so we used root proteomes resulting from 2 different conditions: *RLR2* KO vs WT grown either in normal or N-deprived conditions (abiotic stress). For the normal growth conditions, three biological replicates were prepared and analyzed whereas for the abiotic stress condition just two biological replicates were analyzed. In Table 19 are described the number of proteins that were successfully identified and quantified in each replicate and as shown for *RLR1* KO line, we could identify and quantify a large number of proteins in each experiment.

Table 19 – Number of de-regulated proteins identified and quantified in each biological replicate of *RLR2* KO line.

Growth Condition	Replicate	Number of proteins
Normal	1	1975
	2	1725
	3	1623
Abiotic Stress	1	1490
	2	1017

Following the same strategy used in the previous section, the degree of similarity between the replicates was assessed through Venn diagrams. Moreover, *RLR2* KO vs WT ratios were also converted in log2 values and histograms were made to assess ratio distribution. In Figure 47 and Figure 48 are depicted the Venn diagrams and histograms concerning identified and quantified proteins from samples grown under normal and abiotic stress conditions, respectively. From the samples grown under normal conditions (Figure 47) there is a good agreement between the biological replicates with 1281 proteins identified simultaneously in the 3 biological replicates. Moreover, the histograms show the characteristic Gaussian distribution of the *RLR2* KO vs WT protein ratios, indication of a successful shotgun proteomics experiment. Concerning the results of the samples grown under abiotic stress conditions (Figure 48), they are very similar, with a good agreement between the two biological replicates (934 proteins quantified in both replicates) and a Gaussian distribution of the protein ratios.

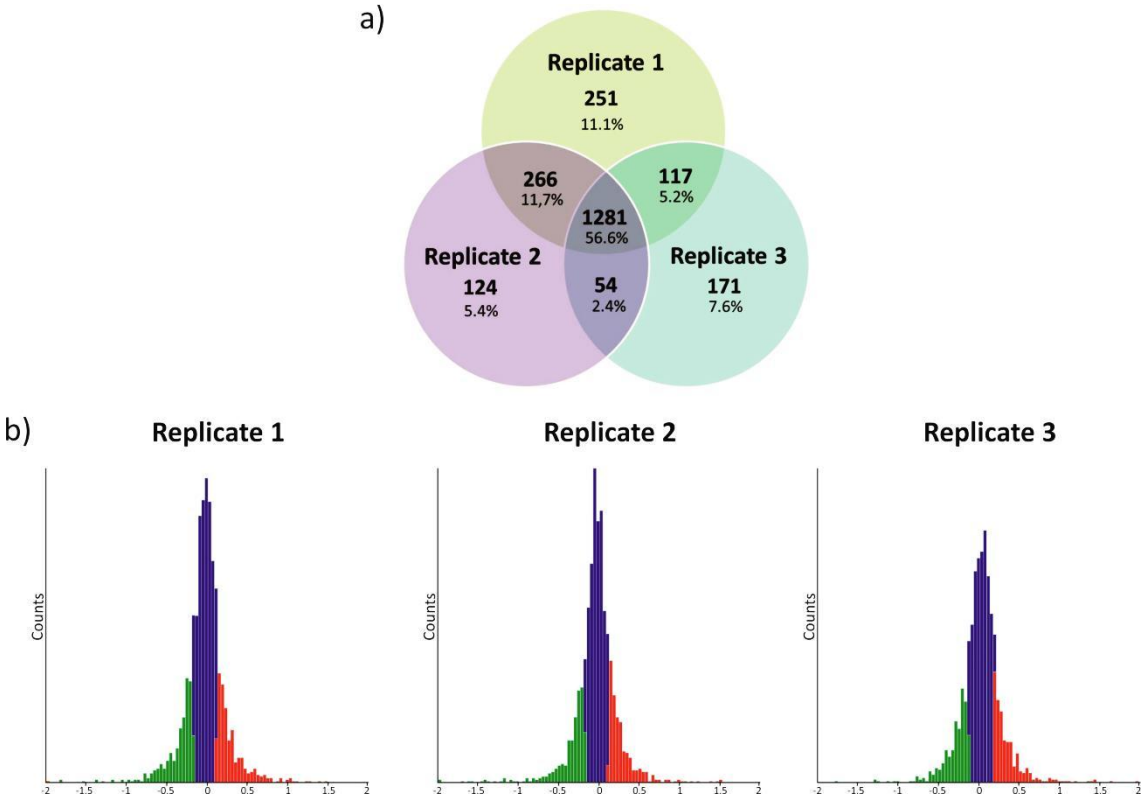


Figure 47 – Distribution of the identified and quantified proteins in *RLR2* KO and WT root proteomes grown under normal conditions. a) Venn diagrams of all the identified and quantified proteins in this experiment distributed by the biological replicates where they were identified. 1281 proteins were simultaneously identified in the three biological replicates. **b)** *RLR2* KO vs WT ratio distribution of quantified proteins in each experimental replicate. Histograms were divided in 10 quantiles and proteins present in Q1 and Q2 (depleted in *RLR2* KO) are highlighted in green while proteins present in Q9 and Q10 (enriched in *RLR2* KO) are highlighted in red. A Gaussian distribution was observed in all biological replicates, a good indicative of a successful shotgun experiment.

In the histograms of both figures are highlighted in green (depleted) and in red (enriched) what was established as significantly de-regulated proteins found in *RLR2* KO in each biological replicate. As described for RLR1, only the significantly de-regulated proteins simultaneously present in all biological replicates of a given growth condition were considered for further analysis. This resulted in the identification of 31 depleted and 14 enriched proteins in *RLR2* KO line grown under normal conditions while under abiotic stress conditions 64 depleted and 54 enriched proteins were considered. All proteins are listed in Table S5 (normal) and Table S6 (abiotic stress) that can be found in the Supplementary Material section.

As previously observed for *RLR1* KO line, there were also common proteins identified under both growth conditions. To analyze the extent of these similarities, a Venn diagram was generated where all the significantly de-regulated proteins were grouped according to their de-regulated status (depleted or enriched) and the growth condition where they were described. As shown in Figure 49, indeed there are similarities between growth conditions

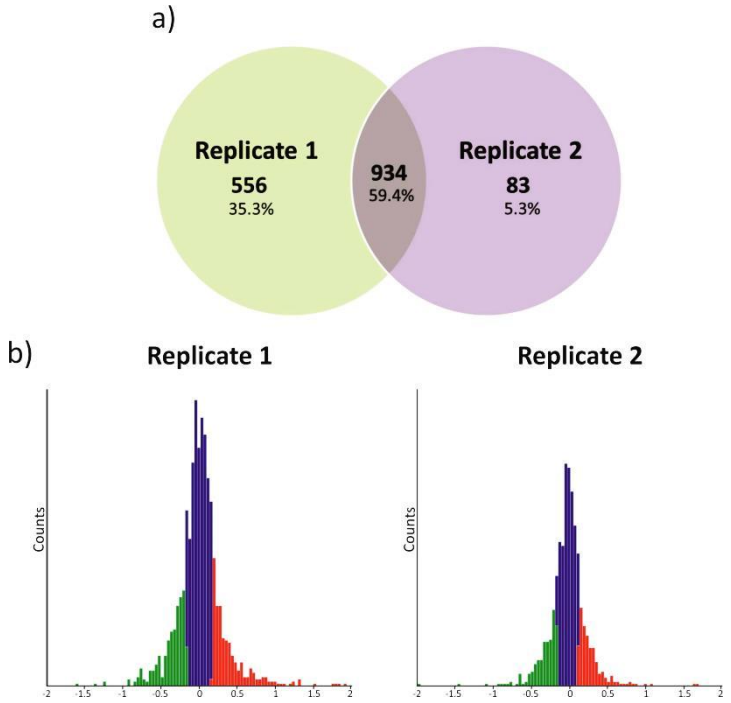


Figure 48 – Distribution of the identified and quantified proteins in *RLR2* KO and WT root proteomes grown under abiotic stress conditions. a) Venn diagrams of all the identified and quantified proteins in this experiment distributed by the biological replicates where they were identified. 934 proteins were simultaneously identified in both biological replicates. **b)** *RLR2* KO vs WT ratio distribution of quantified proteins in each experimental replicates. Histograms were divided in 10 quantiles and proteins present in Q1 and Q2 (depleted in *RLR2* KO) are highlighted in green while proteins present in Q9 and Q10 (enriched in *RLR2* KO) are highlighted in red. A Gaussian distribution was observed in all biological replicates, a good indicative of a successful shotgun experiment.

since 7 proteins were found depleted in both samples and 2 proteins enriched. Interestingly, there is one protein that was found to be enriched under normal conditions and depleted under abiotic stress conditions. All these common proteins are listed in Table 20 and we will discuss some of these proteins in more detail in the next section of this chapter.

We also performed a GO analysis using these significantly de-regulated proteins. The results are illustrated in Figure 50. Analyzing the de-regulated proteins found in *RLR2* KO line grown under normal conditions, it was possible to observe that out of the 14 enriched proteins considered, just 8 putative functions were described being grouped into 3 different GO categories. Regarding depleted proteins, out of the 31 proteins used in this analysis, 24 putative functions were described and grouped into 5 different GO categories. The GO categories that resulted from this study were the same as described for *RLR1* KO line and, again, the majority of identified functions were also grouped in the "Catalytic activity" GO family. The GO analysis for de-regulated proteins found in *RLR2* KO line grown under abiotic stress conditions gave results very similar to those obtained for *RLR1* KO line grown under similar conditions. Out of the 54 enriched proteins used in this analysis, 39 putative functions were described and grouped into the same 6 GO categories described in the RLR1 analysis.

Interestingly, out of the 64 depleted proteins considered, 68 putative functions were found which mean that some of the depleted proteins can have various functions. The putative functions were grouped into 8 different GO categories, also the same that were described in RLR1 GO analysis.

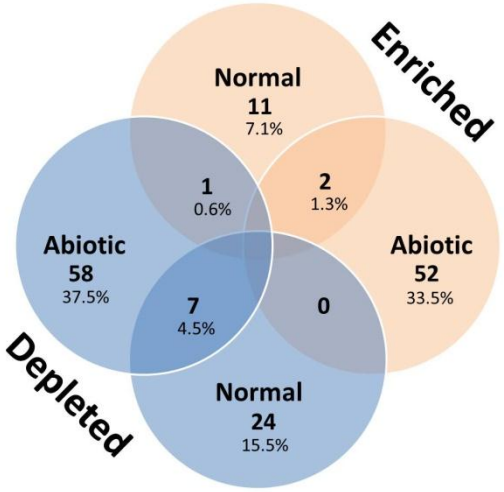


Figure 49 – Venn diagram of all enriched and depleted proteins described for *RLR2* KO line separated by growth conditions. All 163 significantly enriched (red circles) and depleted (blue circles) proteins detected in *RLR2* KO line were separated by plant growth condition.

Table 20 – De-regulated proteins found in *RLR2* KO line grown under both growth conditions.

Gene ID	Gene description	Normal conditions		Abiotic Stress	
		E	D	E	D
AT5G02490	Heat Shock Protein 70 Family Protein				
AT1G21440	Phosphoenolpyruvate Carboxylase Family Protein				
AT3G16400	Nitrile Specifier Protein 1				
AT2G36290	alpha/beta-Hydrolases Superfamily Protein				
AT2G01520	MLP-Like Protein 328				
AT4G08770	Peroxidase 37				
AT1G70850	MLP-Like Protein 34				
AT5G55480	Glycerophosphodiesterase-llike 1				
AT2G25980	Mannose-binding Lectin Superfamily Protein				
AT1G54580	Acyl Carrier Protein2				

E – Enriched; D – Depleted
Shaded rectangles represent depleted or enriched proteins under a given growth condition

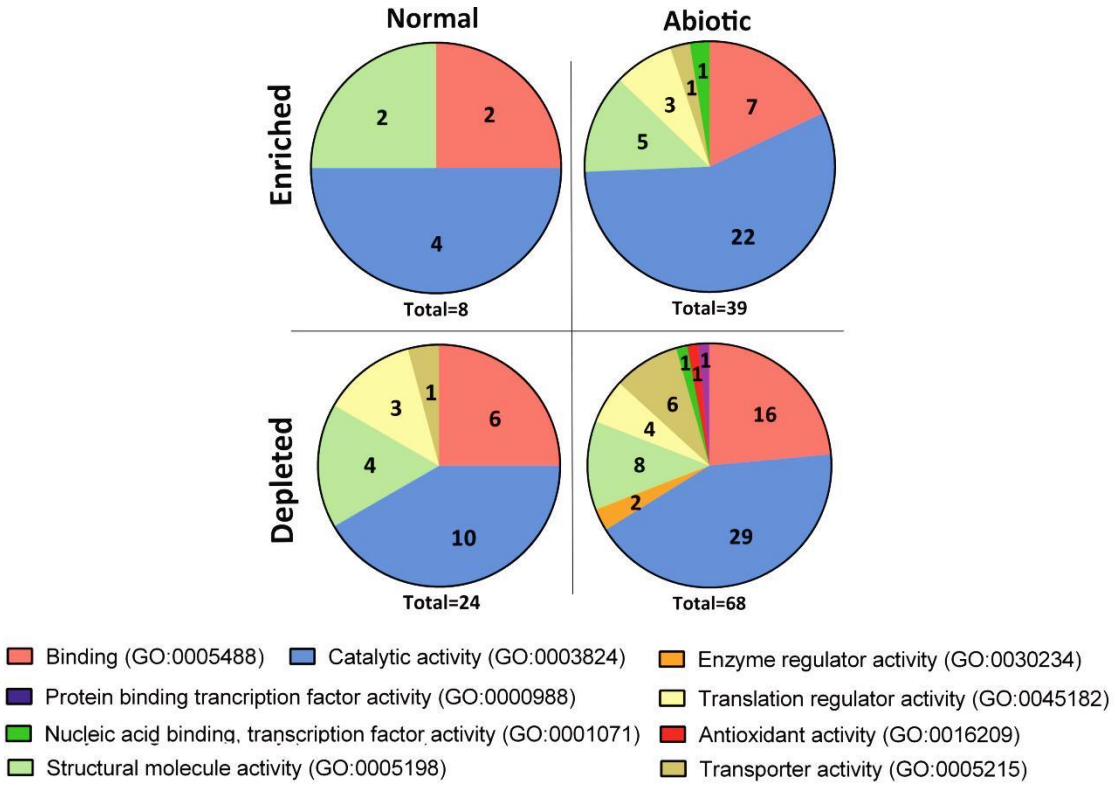


Figure 50 – Gene ontology analysis of the significantly de-regulated proteins that were identified in *RLR2* KO line grown under normal and abiotic stress conditions. Under normal growth conditions 14 and 31 proteins were found to be significantly enriched and depleted, respectively, and were used to conduct the GO analysis. Under the abiotic stress growth condition, 54 and 64 proteins were significantly enriched and depleted, respectively, and were the ones considered to perform this GO analysis. The numbers correspond to the number of predicted molecular functions for each group.

Finally, we also used the STRING database to search for interactions between RLR2 (gene ID: AT4G30040) and the de-regulated proteins found in *RLR2* KO line grown under both conditions. The results for the normal growth condition are shown in Figure 51 and, again, no interaction is predicted/described between RLR2 and any of the de-regulated proteins. However, there are several interactions between the de-regulated proteins and just by analyzing the depleted proteins we can see a string of interactions between MLP proteins that was also found in *RLR1* KO analysis (also among the depleted proteins found in normal growth conditions). Among the enriched proteins we can also see some interactions that are mainly between ribosomal subunits (AT4G25890, AT4G00810 and At3g56340 gene products) and between other 3 proteins: AT5G47890, a subunit of mitochondrial Complex I; ACP2 (gene ID: AT1G54580) which is an acyl carrier protein; and CAC1 (gene ID: AT5G16390), which interestingly is the chloroplastidial acetylcoenzyme A carboxylase 1.

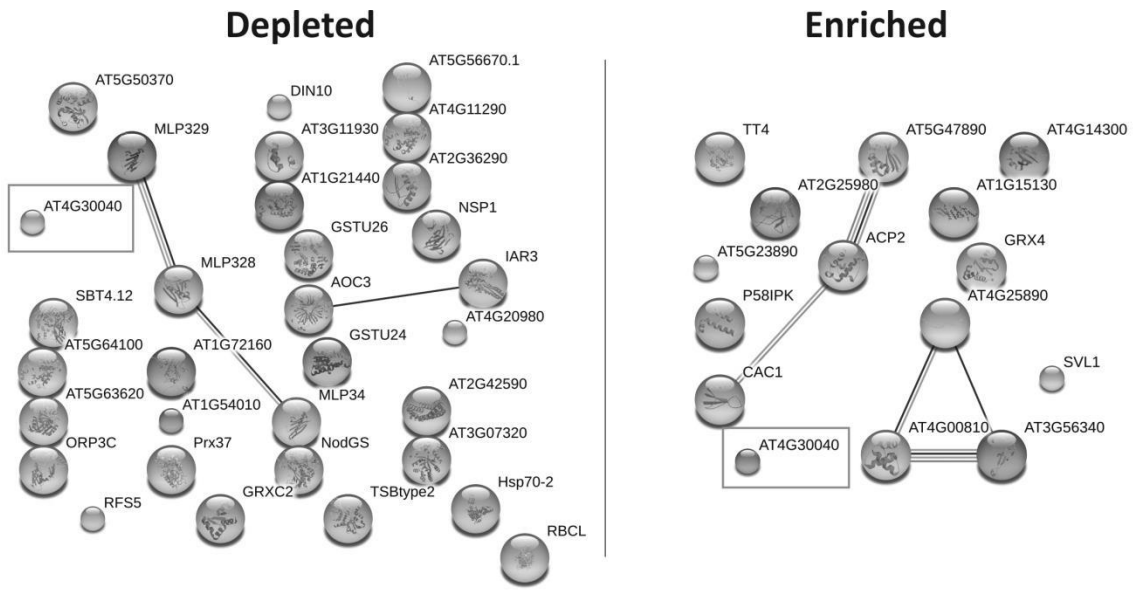


Figure 51 – Protein-protein interactions between the Depleted and the Enriched proteins described in *RLR2* KO line grown under normal conditions. Using the STRING database [188] we searched for interactions between RLR2 (gene ID: At4g30040) (box) and the 31 depleted or the 14 enriched proteins that were significantly de-regulated in *RLR2* KO line grown under normal conditions. There are no described or predicted interactions between RLR2 and all these proteins but there are some known/predicted interactions between several of the de-regulated proteins.

The STRING results concerning the significantly de-regulated proteins described under abiotic stress conditions are illustrated in Figure 52. Again, RLR2 does not interact with any of the de-regulated proteins but there are several interactions already reported/predicted between the de-regulated proteins. The potential relevance of some of these interactions to further understand the function of RLR2 will be discussed in the next section.

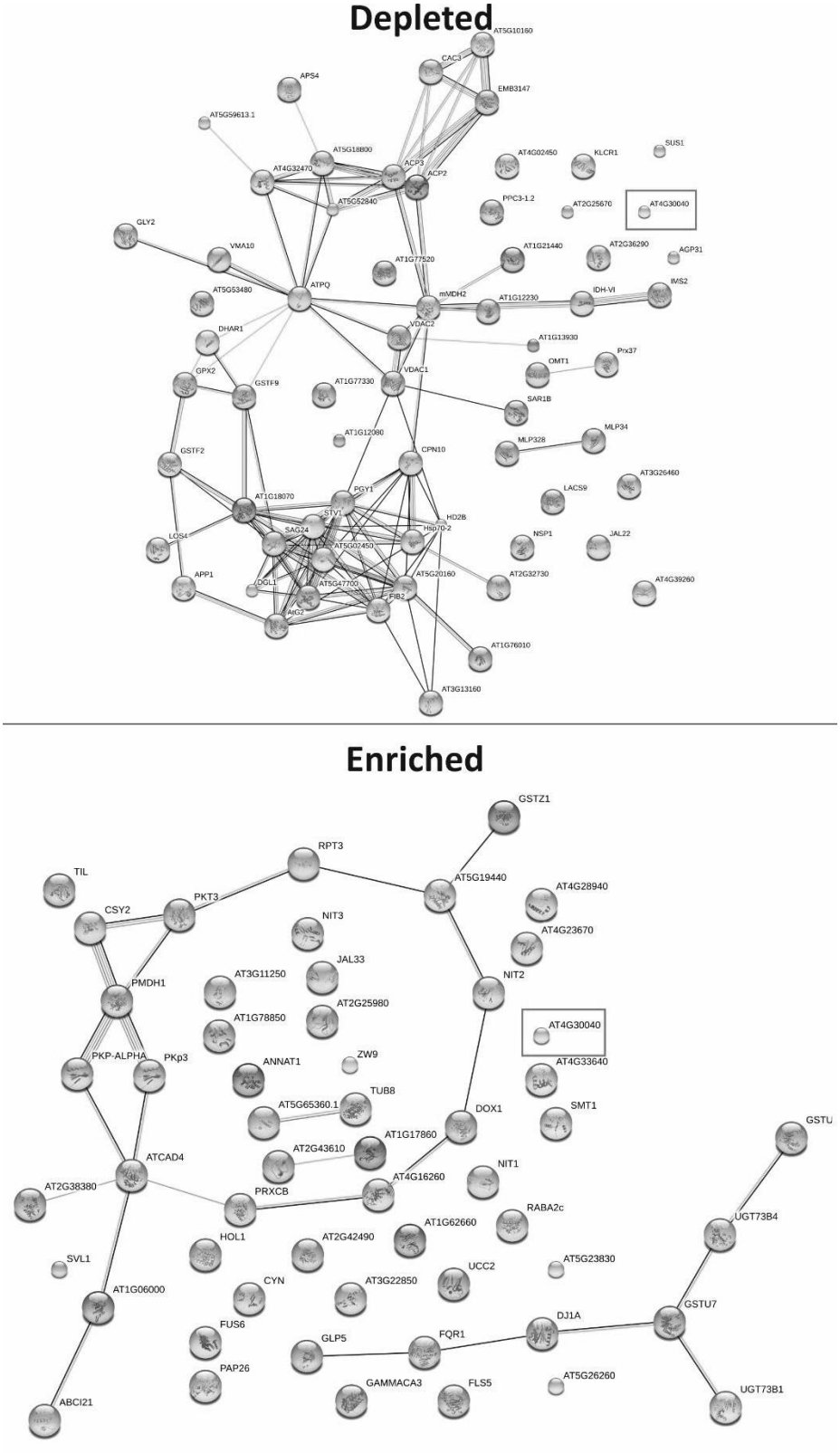


Figure 52 – Protein-protein interactions between the depleted and the enriched proteins described in *RLR2* KO line grown under abiotic stress conditions. Using the STRING database [188] we searched for interactions between RLR2 (gene ID: At4g30040) (box) and the 64 depleted or the 54 enriched proteins that were de-regulated in *RLR2* KO line grown under normal conditions. There are no described or predicted interactions between RLR2 and all these proteins but there are some known/predicted interactions between several of the de-regulated proteins.

4.4 Discussion

From the phenotypic analysis presented in Chapter 2, we have already demonstrated that RLR1 and RLR2 play an important role in primary root growth and lateral root formation. However, the molecular pathways underlying these functions remain to be clarified. Since RLR1 and RLR2 are annotated as proteases (and our results on the biochemical characterization of RLR1 confirmed that it is indeed an active AP), our initial goal was not only to describe the molecular pathways where these proteins may be involved but also the *in vivo* substrates of both proteases. In order to achieve both goals we tried two different proteomics techniques that would give us different but complementary results: quantitative shotgun proteomics, which would give us information about the molecular pathways de-regulated in the KO lines, and N-Terminal Amine Isotopic Labeling of ***S***ubstrates (N-TAILS) which would allow the identification of the *in vivo* substrates of both proteins [191-194]. Both techniques rely on N-terminal labeling of proteins/peptides but while in shotgun proteomics the labeling is done after proteome digestion with trypsin, in TAILS the labeling is prior to trypsin digestion meaning that just the native N-terminals are labeled. The labeling in the initial stages of N-TAILS makes this technique ideal to identify *in vivo* substrates of proteases and, by combining these two techniques, it would be possible to start uncovering in more detail the molecular functions of RLR1 and RLR2. Unfortunately, and due to sample limitations, it was impossible to generate TAILS data on time to be included in this thesis. However, we were able to perform shotgun proteomics analysis which provide the first global profiling of root proteome changes resulting from the absence of both APs.

As we previously discussed, both RLR1 and RLR2 are important for primary root growth, and especially for lateral root formation. Supported by the *in vivo* functional characterization results (See Chapter II), we speculated that both proteins might be part of two different molecular pathways controlling lateral root formation that are triggered under different growth conditions: RLR1 seems to be more important under normal growth conditions while RLR2 seems to be more important under ammonium nitrate deprived conditions. To validate this hypothesis, root proteome samples of *RLR1* KO, *RLR2* KO and WT plants grown under normal or N deprived conditions were prepared and analyzed by quantitative shotgun proteomics (three biological replicates for samples grown under normal conditions and two biological replicates for those grown under the abiotic stress condition).

As described in the previous section, the biological replicates were quite comparable and the majority of identified and quantified proteins were found in all replicates of the same condition. We were able to quantify the ratios of hundreds of proteins per condition; however,

and given the large amount of generated results, one of the established premises was to start by focusing our analysis on the 20 % most enriched and the 20 % most depleted proteins in the KO lines when compared with the WT plants. Moreover, and in order to be considered significantly de-regulated, a given protein would have to be present in all biological replicates of that condition and its ratio similar in all biological replicates (meaning that it would have to display the same depleted or enriched trend in all samples).

Regarding *RLR1* KO line, we were able to identify several significantly de-regulated proteins in this mutant line for each growth condition. To understand better the molecular functions that were de-regulated, a GO analysis was performed for these proteins. However, the results from this analysis were very generic, with most of the described functions grouped in the GO category "Catalytic activity". Moreover, we looked for described or predicted interactions between RLR1 and the de-regulated proteins using the STRING database but none was found. However, there were several predicted/known interactions between the de-regulated proteins. Among the proteins depleted under normal conditions we found an interesting interaction between MLP328 and MLP34. MLP stands for Major Latex-like Protein and this is a group of proteins that were first described in the opium poppy (*Papaver somniferum*) [195]. In the following years since their discovery they were also found in a variety of other plant species, including *A. thaliana* [196]. Although their functions are still unclear, they have been associated to defense responses against pathogens and wounding as well as to processes related with plant development [196-198]. In *RLR1* KO mutants grown under normal conditions, these proteins are strongly depleted which can imply that some defense mechanisms are impaired in this mutant. Interestingly, 3 MLPs are also depleted in abiotic stress conditions (MLP328, MLP43 and MLP34) which suggest that some of the de-regulated processes in *RLR1* KO line are present independent of the growth condition. Moreover, we can also find MLP328 and MLP34 depleted in *RLR2* KO line grown under both conditions. In fact, the consistent depletion of MLP-like proteins in both KO lines together with the observed root/lateral root phenotype may anticipate an additional role for (some of) these MLPs in root developmental processes. However, more studies will be necessary to confirm this hypothesis.

Among the depleted proteins described for *RLR1* KO line grown under normal conditions there is another additional string of interactions between the proteins Modified Vacuole Phenotype 1 (MVP1) (gene ID: At1g54030), At1g54010 gene product (GDSL esterase/lipase), At5g23820 gene product (MD-2-related lipid recognition domain-containing protein) and VHA-E3 (V-type proton ATPase subunit E3) which may suggest that lipid metabolism may also be impaired in *RLR1* KO line under this growth condition. Interestingly,

Marti *et al.* described that a missense mutation in MVP1 (also known as GOLD36) caused defects in ER structure and subsequently defects in protein trafficking through the secretory pathway [199]. These defects were caused by an aberrant localization of the mutated protein since it was destined to post-ER compartment but was retained in the ER [199]. Moreover, authors also described that GOLD36 KO mutant had the same phenotype what stresses the importance of this protein to proper protein trafficking through the secretory pathway [199]. In our study this protein was found to be depleted in *RLR1* KO line grown under normal conditions which may anticipate some impact in protein trafficking for this KO line under this growth condition.

Among the enriched proteins described in *RLR1* KO grown under normal conditions we can find NIT1 (At3g44310 gene product) which is the predominantly expressed nitrilase in *A. thaliana* and is part of one of IAA biosynthesis pathways, catalyzing the hydrolysis of indole-3-acetonitrile (IAN) to IAA [200]. Moreover, this enzyme can also catalyze the conversion of IAN to indole-3acetamide (IAM) [201]. Since we could find this enzyme enriched in *RLR1* KO, we can hypothesize that auxin biosynthesis is enhanced is this mutant leading to the enhanced auxin response observed in terms of lateral root development. NIT2 (gene ID: At3g44300) is another isoform of nitrilase that can be found in *A. thaliana* and interestingly it is also enriched in *RLR1* KO line strengthening the idea that auxin biosynthesis may be enhanced in this mutant. Interestingly both NIT1 and NIT2 are also enriched in both *RLR1* and *RLR2* KO lines grown under abiotic stress conditions which suggest that auxin biosynthesis may be enhanced in both growth conditions and also in both KO lines. Our phenotypic analysis supports this idea since *RLR2* KO mutant also showed an enhanced auxin response in terms of lateral root number, although less severe than *RLR1* KO response.

Another interesting enriched protein found in *RLR1* KO line grown under normal conditions is Transparent Testa 4 (TT4) (gene ID: At5g13930). This protein is a key enzyme involved in flavonol biosynthesis and its interaction with CHIL (gene ID: At5g05270) (another enriched protein) lead us to speculate that flavonol biosynthesis may be increased in *RLR1* KO line grown under normal conditions. In a recent study, Nakabayashi *et al.* demonstrated that the accumulation of flavonoids led to an enhancement of oxidative and drought tolerance [202]. Interestingly, TT4 was also found to be enriched in *RLR2* KO line grown under normal conditions showing additional similarities between the two KO lines. Supported by these results, it would be interesting to evaluate if *RLR1* and *RLR2* KO lines are also more tolerant to oxidative and drought stresses.

Concerning the de-regulated proteins described for *RLR1* KO line grown under abiotic stress conditions, we were able to obtain almost 3x more significantly de-regulated proteins in

these samples. As described, there are no predicted/known interactions between RLR1 and the de-regulated proteins but amongst them it is possible to detect a large number of described interactions. Amongst the depleted proteins most of the interactions described are between proteins that are structural subunits of ribosomes or are involved in protein translation which lead us to suggest that protein production is likely impaired under this growth condition. One interesting group of interactions can be found between DHAR1 (gene ID: At4g02520), GSTU24 (gene ID: At1g17170), GSTF2 (gene ID: At4g02520) and GPX2 (gene ID: At2g31570). All these proteins are involved in glutathione metabolism (the first 3 are glutathione transferases and the last one is a glutathione peroxidase) and they are involved in redox homeostasis and defense mechanisms against abiotic and biotic stresses [203-205]. Interestingly, glutathione peroxidases were found to influence root architecture in a redox sensing mechanism and the absence of GPX2 was described to have a deleterious effect in lateral root formation [206]. It is noteworthy that GPX2 along with DHAR1 and GSTF2 were also found to be depleted in *RLR2* KO line under the same growth condition. In what concerns the enriched proteins, the majority of them interact with each other. Among the identified interactions, it was possible to find the interactions between RPT3 (gene ID: At5g58290), RPT1A (gene ID: At1g5375) and RPN12a (gene ID: At1g64520) which are all regulatory subunits of the 26S proteasome and might imply a de-regulation in the proteasome activity.

In the case of *RLR2* KO line we were also able to identify a good number of significantly de-regulated proteins under both growth conditions, although as described for *RLR1* KO line, we identified 3x more de-regulated proteins under the abiotic stress condition. The strategy to analyze this data was the same, with a GO analysis followed by a STRING analysis to search for known/predicted interactions between RLR2 and the de-regulated proteins. The GO analysis results were also very similar, with most of the proteins being grouped in the “Catalytic activity” category. Moreover, the STRING analysis revealed that there is no known/predicted interactions between RLR2 (gene ID: At4g30040) and any of the de-regulated proteins but, again, some interesting interactions were found between the de-regulated proteins.

Among depleted proteins found in *RLR2* KO line grown under normal conditions, a string of interactions between MLP proteins was again observed, some of them also found to be de-regulated in *RLR1* KO line. The other interaction involves the proteins Allene Oxide Cyclase 3 (AOC3) (gene ID: At3g25780), an enzyme involved in jasmonic acid biosynthesis, and IAA-Alanine Resistant 3 (IAR3) (gene ID: At1g51760), an IAA-aminoacid conjugate hydrolase. Jasmonates are hormones which play very important roles in plant stress responses and in plant development and AOC3 is one of the two root AOC isoforms which establishes the enantiomeric structures of jasmonates, an essential step in jasmonic acid (JA) biosynthesis

[207]. In the case of IAR3, this enzyme is part of the auxin signaling pathway since it hydrolyses IAA-Alanine conjugates, releasing free and active IAA [208]. Interestingly, a close homolog of IAR3 from *Nicotiana attenuata* was also found to hydrolyze Jasmonoyl-Isoleucine conjugates *in vitro,* and in *Arabidopsis* IAR3 was proved to be part of JA signaling upon a wound stress response, strengthening the idea of a crosstalk between these two hormones [209, 210]. The fact that these two important enzymes appear depleted in *RLR2* KO line grown under normal conditions make us hypothesize that JA biosynthesis pathway may be compromised under this condition, as well as the crosstalk between this hormone and auxin. Concerning the enriched proteins found under these normal growth conditions, two groups of interactions were identified in the STRING analysis but no major conclusions could be withdrawn that could help elucidating the pathways affected by *RLR2* KO.

Consistent with what has been described for RLR1, 3x more de-regulated proteins were identified in *RLR2* KO line grown under abiotic stress conditions when compared with the normal growth conditions. Because of this, the STRING analysis also revealed a larger number of interactions between the de-regulated proteins. However, none of the interactions described involved RLR2. Amongst the interactions between depleted proteins, we can see Defective Glycosylation 1 (DGL1) which interacts with ribosomal proteins. DGL1 is an interesting protein because it is an *Arabidopsis* homolog of an essential oligosaccharyltransferase complex subunit, involved in protein N-glycosylation [211]. The mutant line of this gene showed a dwarf phenotype cause by reduced N-glycans transfer and concomitant underglycosylation of proteins [211]. More interestingly was the fact that recently a mutation in DGL1 rice homolog, OsDGL1, was found to cause a short root phenotype caused by shorter root cell lengths, smaller root meristem and increased cell death in root [212]. Since the mutation caused the appearance of a premature stop codon, the root phenotype was attributed to the absence of OsDGL1. In our case we are observing a depletion of this protein in *RLR2* KO line grown under abiotic conditions, which also had a short root phenotype. Our phenotype is not as severe as the one reported for OsDGL1 mutation but then we are seeing a depletion of the protein, not its absence, and this depletion might partly explain *RLR2* KO phenotype.

For the enriched proteins found in *RLR2* KO line grown under abiotic stress conditions a more reduced number of interactions were observed. An interesting group included mostly proteins involved in redox sensing mechanisms, where Plasmodesmata Germin-like Protein 1 (PDGLP1) (Gene ID: At1g09560), also known as GLP5, could also be found. Recently Ham *et al.* have shown that this protein was preferentially expressed in root tissue and an important plasmodesmata protein [213]. Interestingly, its overexpression led to the disruption of root

development causing a short root phenotype accompanied by an increase in lateral root length [213]. As we just described, this protein appears enriched in our proteomics analysis and although we did not analyze lateral root length, *RLR2* KO mutant had indeed a short root phenotype (but not as strong as the one described for PDGLP1 overexpression).

Based on the analysis of the most significantly enriched and depleted proteins identified in this study, we could not confirm our initial hypothesis that RLR1 might be more relevant as a regulator of root/lateral root growth under normal growth conditions and RLR2 under N deprived conditions. This is not totally unexpected as the most drastic effect was observed in lateral root number and, due to sample constrains, we were only able to analyze total root proteome samples, which may potentially mask these biological differences. However, our results clearly anticipate some interesting differences common to both KO lines, which may reflect the potential correlation of these newly identified APs with molecular pathways regulating not only auxin responses (illustrated by the observed enrichment of several nitrilase proteins in both KO lines) but also redox sensing mechanisms (with several proteins important for redox homeostasis depleted under N deprived conditions). All in all, these results anticipate a cross-talk between APs and each of these components during root development which, to our knowledge, has never been previously reported. Interestingly, some of the de-regulated proteins in RLR1 and RLR2 KO lines have been already reported to have an important role in primary root formation, or in other processes that can influence root development, and may be used in future studies as potential targets to start dissecting the molecular mechanisms where RLR1 and RLR2 are involved.

Chapter V – Concluding Remarks and Future Perspectives

In Chapter I we highlighted the role of plant atypical APs in several plant processes that ranged from biotic and abiotic stress responses to PCD during gametophyte development. With the integrative approach followed in this work, we provide additional evidences for the relevance of two novel atypical APs also in root development, we demonstrate that these proteases may be implicated in pathways controlling auxin responses and/or redox sensing mechanisms, and we show that one of these atypical APs display distinct biochemical properties and unique specificity preferences. In this final section we will summarize the results obtained throughout this work and also suggest what can be done not only to improve our data but also to answer some of the questions that arise from this study.

Our initial goals were to study the *in vivo* function, biochemical properties, and *in vivo* substrates of two atypical APs (that we later named RLR1 and RLR2), whose expression was reported to be regulated by two transcription factors involved in male sterility phenotypes [41, 85]. Based on this, our first working hypothesis was that both proteins would be involved in male gametophyte and pollen development probably in the control of PCD events that occur during these developmental processes. This hypothesis was also supported by the fact that there are two other atypical APs shown to be involved in this PCD control, UNDEAD and PCS1, with their de-regulation associated with severe male sterility phenotypes [37, 41]. Having this in mind, our first approach was to study the expression patterns of *RLR1* and *RLR2*. For *RLR1*, we were able to demonstrate using GUS staining that this gene is expressed specifically in mature pollen grains and also in root tips, whereas *RLR2* expression in whole flower inflorescence and also in root tissues was confirmed by RT-PCR.

To assess the intracellular localization of both proteases, we transformed mesophyll protoplasts with several DNA constructs aiming not only to describe RLR1 and RLR2 localization but also the minimal N-terminal sequence responsible for that localization. Both proteins have a putative signal peptide that theoretically should target them to the secretory pathway. Using GFP as a reporter gene we were able to confirm that both signal peptides were indeed active. Moreover, increasing the length of RLR1 N-terminal sequence fused to GFP always resulted in an ER localization pattern which suggests that the signal peptide is the only targeting sequence present in the N-terminal region of RLR1. In the case of RLR2 N-terminal sequences, an increase in length led to GFP aggregation thus impairing us to take any conclusion from those experiments. When the full-length sequences of either RLR1 or RLR2 were fused to GFP, an ER localization pattern was again observed; however, at this point, we cannot conclude if both proteases reside in the ER (as predicted by bioinformatics analysis) or if they are in transit within the secretory pathway. To further clarify these results, these experiments should be

repeated and protein expression evaluated by confocal microscopy to obtain better resolution. Also, RLR1- and RLR2-GFP fusion constructs containing an ER retention signal should be tested to clarify if the localization patterns are maintained for both proteases.

We then started assessing the *in vivo* function(s) of both RLR1 and RLR2 by making use of one *A. thaliana* T-DNA KO line for each gene. Based on our initial working hypothesis, anther/pollen phenotypes were first evaluated. However, no drastic phenotypic differences were found for both KO lines. Pollen viability was not altered and, apparently, pollen germination rates were also unchanged (although in this later case the data is not 100% clear). Moreover, whole flower morphology of both KO lines revealed no major differences when compared to WT plants. Gene redundancy is an issue in *A. thaliana* research, and may likely explain the results herein described regarding anther/pollen development. Therefore, we cannot still completely rule out a role for these proteases in any of these mechanisms and future studies with double *RLR1/RLR2* KO lines could help clarifying this point.

Because both proteins were shown to be expressed also in root tissue - being RLR1 expressed specifically in root tips - this prompted us to explore the potential relevance of these proteases in root growth. Strikingly, both KO lines showed a defect in primary root growth, with shorter primary roots, and in case of *RLR1* KO line, a significantly reduced number of lateral roots. Because another atypical AP was described to play a role in abiotic stress response [38], we decided to challenge RLR1 and RLR2 KO lines with ammonium nitrate deprivation and check for effects in primary root phenotype under this abiotic stress condition. As before, both KO lines had also shorter primary roots than the WT plants. However, when we analyzed the lateral root phenotype, *RLR2* KO showed a significant reduction in the number of lateral roots (45% less lateral roots than the WT) whereas *RLR1* KO line had almost a WT phenotype. These results lead us to speculate that both proteins may control lateral root formation through different pathways, being RLR2 more important under N-deprived conditions while RLR1 is more important under normal growth conditions. To further validate this hypothesis, it will be critical to start by evaluating if RLR2 is indeed more expressed in WT plants when they are grown under N-deprived conditions. To our knowledge, this is the first time that atypical APs are reported to have a function in root development and lateral root formation. In the future, it will be necessary to complement the KO lines with the WT gene to confirm the rescue of this root phenotype. Moreover, KO lines should also be complemented with the corresponding active site mutant gene to provide additional substantiation of the relevance of the catalytic activity of RLR1 and RLR2 for this phenotype.

Auxin is a plant hormone that play a major role in controlling primary root development and lateral root formation [119]. In order to start unveiling the molecular

mechanisms underlying RLR1 and RLR2 functions in these two processes, we tested if these enzymes were part of the auxin signaling pathway. Both KO lines were germinated in normal media and, after 4 days, seedlings were transferred to a media supplemented with different IAA concentrations (IAA is the major active auxin form found in the plant cell [119]). Our results demonstrate that KO mutants and the WT respond to IAA at the same extent in what concerns primary root growth, suggesting that RLR1 and RLR2 function(s) in this process may be auxin independent. However, as described for p23 mutants, *RLR1* and *RLR2* KO mutants may have an impaired auxin distribution, which may be caused by changes in the expression of auxin transporters [122]. In order to test this hypothesis, root auxin levels of both KO lines should be measured and the expression levels of root auxin transporters - PIN1, PIN2, PIN3, PIN4 and PIN7 - should be evaluated, since auxin fluxes can be predicted based on the distribution of these PIN transporters [214]. Interestingly, when lateral root number was evaluated the results were completely different, as the response to auxin treatment was greatly enhanced for both KO lines; with *RLR1* and *RLR2* KO lines showing approximately 70x and 40x more lateral roots (number of lateral roots *per* cm of primary root) under 500 nM auxin, respectively, when compared to the control condition (no auxin). One of two hypothesis can explain this phenotype: either RLR1 or RLR2 are negatively regulating auxin signaling or they are part of the CK pathway, as CK and auxin are antagonists in root development [119, 124]. In the first hypothesis, RLR1 and RLR2 might function as negative regulators of ARF transcription, and it would be interesting to measure the transcription levels two ARF proteins (ARF7 and ARF19) that specifically regulate genes involved in lateral root formation [123] in both KO lines, to further validate this hypothesis. On the other hand, measuring the CK levels of these KO lines as well as their CK response in terms of primary root growth and lateral root formation would provide additional information on the participation of these atypical APs in this other pathway.

RLR1 and RLR2 OE lines were also generated in this study and evaluated for a root phenotype. However, our results either in terms of primary root growth or lateral root number showed that both OE mutant lines have very similar (if not more drastic) phenotypes to those described for the KO plants. These results were somewhat unexpected, and apparently contradictory, but it is important to keep in mind that the over-expression of proteases can have several side-effects. Proteolytic activity is generally a tightly regulated process and most of these regulatory events are likely overruled under overexpression conditions. For instance, CDR1 overexpression plants were reported to have a dwarf phenotype (that was independent of its disease resistance phenotype) while CDR1 anti-sense plants appeared normal under normal growth conditions [36]. For other APs such as AED1 and ASPG1, the phenotype

reported was only for the overexpression lines while the KO lines did not exhibit any phenotype [38, 42]. These examples prove that the overexpression of these enzymes might lead to side-effects and sometimes even to unexpected phenotypes.

The biochemical characterization of RLR1 and RLR2 was also one of the major goals of this work. Our first approach was to use *E. coli* as the expression system, and to try to produce the recombinant proteins in their soluble form in *E. coli* cytoplasm. However, after several failed attempts, we were forced to conclude that *E. coli* is not the most suitable expression system to produce rRLR1 and rRLR2. As a second approach we tried the plant-based expression system magnICON® to produce the proteins in *Nicotiana benthamiana* leaves [129]. We were able to express both proteins at reasonable levels when tested in small-scale expression assays, and we have optimized a purification protocol for the pro-form of RLR1 fused to a histidine tag. Purified rRLR1 was shown to be glycosylated and to be present in two different forms: with and without the prosegment domain. Interestingly, the pro-segment is shorter than initially predicted and we cannot rule out that this small difference (6 amino acids in this case) may impair recombinant protein production or even its final activity, eventually explaining the observed instability of the truncated forms of rRLR1 lacking this segment. Strikingly, our results indicate that rRLR1 with the prosegment is also proteolytically active. This has been previously shown for recombinant CDR1 produced in *E. coli*, but for this protease no processed form was ever identified when using this heterologous system. In future studies, it will be interesting to evaluate if rRLR1 activation occurs *in vivo* prior purification or even during that procedure. Also, it will be interesting to perform activation assays *in vitro* and evaluate in detail the impact of prosegment removal on proteolytic activity.

The activity of purified rRLR1 was tested towards several fluorogenic substrates, but only one was successfully cleaved by the enzyme: [MCA-Lys]Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu[Lys-DNP]. This apparently narrow specificity is uncommon in typical APs but is also a feature of CDR1, for which a similar preference was reported (in fact, this substrate was specifically designed for CDR1) [53]. Moreover, we were able to demonstrate that rRLR1 is most active at pH 4 against this substrate, with no detectable activity at pH 6. This contrasts with the higher optimum pH reported for CDR1, resembling more the expected pH of an AP. However, the diversity observed among members of atypical APs regarding their enzymatic properties is again confirmed with the results from the inhibition profile. rRLR1 was shown not to be completely inhibited by pepstatin A, the canonical AP inhibitor. Total inhibition by pepstatin A has been the hallmark of typical AP activity but it seems that for atypical APs this is not a rule since recombinant CDR1 was also not completely inhibited by pepstatin A and others

have been shown to have different degrees of sensitivity [6, 38, 49, 53, 59]. Interestingly, rRLR1 was inhibited by redox agents such as reduced and oxidized glutathione, NADP and NADPH and this can be linked to the high number of cysteine residues present in rRLR1 primary structure. In fact, we hypothesize that rRLR1 might act as a redox sensor *in vivo* given the fact that the mature form of this enzyme has 11 cysteines, and probably one of these residues will not be forming a disulfide bridge. Since there is no crystal structure yet available for an atypical AP, it is very difficult to predict RLR1 disulfide bridges (most of these residues are not found in conserved positions), but the fact remains that its activity is highly impaired by redox agents which strengthens the redox-sensing hypothesis. In fact, this is in line with what has been previously suggested for CDR1 [53], again reinforcing the idea that atypical APs may have an important role in redox regulation.

In order to determine rRLR1 specificity profile we used a high throughput proteomics technique called PICS (Proteomics Identification of Cleavage Sites). As anticipated by our initial assays with fluorogenic substrates, rRLR1 PICS results revealed some interesting differences when compared to specificity preferences reported for other plant and non-plant APs [4, 155](our own unpublished data). Although rRLR1 appears to accommodate preferentially hydrophobic residues in S1 and S1' subsites which is very similar to other APs [155, 157-162], it is also clear a preference for Asn and Lys in S1. This ability to accommodate a Lys in S1 has been previously reported for the fungal AP aspergillopepsin [163]. In fact, several fungal APs differ from other eukaryotic counterparts in their ability to cleave substrates with polar residues such as Lys in P1, although they also favor hydrophobic amino acids at P1 and P1' [164, 165]. This fact allows fungal APs to activate trypsinogen and this special feature is conferred by just two amino acids present in the flap region, which are not present in other APs: Asp77 and Ser79 (pepsin numbering) [164, 165]. Interestingly, and as we showed in Table 16 (Chapter III), RLR1 share the same sequence in the flap region which may likely explain the observed overrepresentation of Lys in P1. To our knowledge, RLR1 is the first non-fungal AP with this feature and future site-directed mutagenesis of these flap residues will help clarifying this unique resemblance to fungal APs. Moreover, RLR1 appears also to be more stringent in S2 and S3 subsites with a unique preference for charged amino acids at S2 (Glu or Lys), two distinguishing features when compared to other PICS profiles of APs [4, 155]. Taking all of these results together we can conclude that rRLR1 has some distinctive amino acid preferences, again reinforcing the concept that atypical APs are significantly different and diverse when compared with typical plant APs.

Finally, it was also our goal to use proteomics-based strategies to start dissecting the molecular mechanisms underlying RLR1 and RLR2 *in vivo* functions. To achieve this goal we proposed the use of two state-of-the-art proteomics techniques, Quantitative Shotgun Proteomics and N-TAILS, which would allow us to unveil not only the molecular pathways de-regulated in the KO mutant lines (using shotgun proteomics), but also the *in vivo* substrates of both proteins (using N-TAILS). Unfortunately, and due to sample constrains, we were not able to obtain results from the N-TAILS analysis. However, it was possible to perform shotgun experiments comparing each KO line grown either in normal or abiotic stress conditions with WT plants grown under the same conditions. We were able to successfully identify and quantify several hundreds of proteins with a good agreement between the biological replicates. Moreover, and in order to increase the reliability of this analysis, just the 20% most enriched and depleted proteins found in all biological replicates of the same sample were considered as significant de-regulated proteins.

With these de-regulated proteins we performed a GO analysis but it generated very generic results and did not allow us to withdraw any major conclusions concerning the nature of de-regulated pathways in both KO lines. In a second approach, we evaluated the existence of predicted or known interactions between RLR1 and RLR2 and any of the de-regulated proteins found in each condition but no interaction has been previously reported. However, we found several interactions among the de-regulated proteins suggesting that auxin signaling and/or biosynthesis might be de-regulated in both KO lines, independently of the growth condition, which may help explaining the results observed upon auxin treatment. In fact, we found similarities not only between the same KO line grown under different growth conditions, but also between the two KO lines which clearly suggest that there are common pathways de-regulated in both KO lines (e.g. MLP-like proteins shown to be depleted in both KO lines under normal conditions; as well as several proteins involved in glutathione metabolism depleted in both KOs under N deprived conditions), further corroborating the phenotypic similarities between *RLR1* and *RLR2* KOs. Of course this analysis also revealed some interesting de-regulated proteins that are unique for each KO line and for each condition, which suggest that these proteases may still participate in different pathways (e.g. MVP1 shown to be depleted in *RLR1* KO line under normal conditions; AOC3 and IAR3, both depleted in *RLR2* KO line under normal conditions; and DGL1 shown to be depleted also in *RLR2* KO line but under abiotic stress conditions). Overall, these quantitative shotgun proteomics results provide the ground for new studies which will hopefully clarify the molecular functions of RLR1 and RLR2 in root development.

At the end of this work, we believe that we achieved the majority of the goals initially proposed. To our knowledge, this is the first report that link atypical APs to root development. So far, only 3 other proteases were reported to be involved in root processes [102-104]. Therefore, our evidence on RLR1 and RLR2 as novel regulators of root/lateral root growth, and the eventual crosstalk with auxin signaling and/or redox homeostasis, is a major contribution to this research area and, hopefully, it will pave the way for new experiments to uncover the relevance of proteolytic processes during root development, in particular those resulting from atypical APs. In fact, our results strengthen the idea that these enzymes have quite unique properties, further highlighting the pattern of complexity among this fascinating group of plant proteases.

References

1. Davies D.R., *The Structure and Function of the Aspartic Proteinases.* Annual Review of Biophysics and Biophysical Chemistry, 1990. **19**: p. 189-215.
2. Dunn B.M., *Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases.* Chemical Reviews, 2002. **102**(12): p. 4431-4458.
3. Simões I., Faro R., Bur D., Kay J., and Faro C., *Shewasin A, an active pepsin homolog from the bacterium Shewanella amazonensis.* FEBS Journal, 2014. **278**(17): p. 3177-3186.
4. Cruz R., Huesgen P., Riley S.P., Wlodawer A., Faro C., Overall C.M., Martinez J.J., and Simões I., *RC1339/APRc from Rickettsia conorii is a novel Aspartic Protease with properties of retropepsin-like enzymes.* PLoS Pathog, 2014. **10**(8): p. e1004324.
5. Rawlings N.D., Barrett A.J., and Finn R., *Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors.* Nucleic Acids Research, 2016. **44**(Database issue): p. D343-D350.
6. Simões I. and Faro C., *Structure and function of plant aspartic proteinases.* European Journal of Biochemistry, 2004. **271**(11): p. 2067-2075.
7. Terauchi K., Asakura T., Ueda H., Tamura T., Tamura K., Matsumoto I., Misaka T., Hara-Nishimura I., and Abe K., *Plant-specific insertions in the soybean aspartic proteinases, soyAP1 and soyAP2, perform different functions of vacuolar targeting.* Journal of Plant Physiology, 2006. **163**: p. 856-862.
8. Egas C., Lavoura N., Resende R., Brito R.M.M., Pires E., de Lima M.C.P., and Faro C., *The Saposin-like domain of the plant aspartic proteinase precursor is a potent inducer of vesicle leakage.* Journal of Biological Chemistry, 2000. **275**(49): p. 38190-38196.
9. Curto P., Lufrano D., Pinto C., Custódio V., Gomes A.C., Trejo S.A., Bakás L., Vairo-Cavalli S., Faro C., and Simões I., *Establishing the yeast Kluyveromyces lactis as an expression host for production of the Saposin-Like domain of the aspartic protease cirsin.* Applied and Environmental Microbiology, 2014. **80**: p. 86-96.
10. Chen X., Pfeil J.E., and Gal S., *The three typical aspartic proteinase genes of Arabidopsis thaliana are differentially expressed.* European Journal of Biochemistry, 2002. **269**: p. 4675-4684.
11. Mazorra-Manzano M.A. and Yada R.Y., *Expression and characterization of the recombinant aspartic proteinase A1 from Arabidopsis thaliana.* Phytochemistry, 2008. **69**: p. 2439-2448.
12. Mazorra-Manzano M.A., Tanaka T., Dee D.R., and Yada R.Y., *Structure–function characterization of the recombinant aspartic proteinase A1 from Arabidopsis thaliana.* Phytochemistry, 2010. **71**: p. 515-523.
13. Runeberg-Roos P., Törmäkangas K., and Östman A., *Primary structure of a barley-grain aspartic proteinase.* European Journal of Biochemistry, 1991. **202**: p. 1021-1027.
14. Runeberg-Roos P. and Saarma M., *Phytepsin, a barley vacuolar aspartic proteinase, is highly expressed during autolysis of developing tracheary elements and sieve cells.* The Plant Journal, 1998. **15**(1): p. 139-145.
15. Kervinen J., Tobin G.J., Costa J., Waugh D.S., Wlodawer A., and Zdanov A., *Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting.* EMBO Journal, 1999. **18**(14): p. 3947-3955.
16. Cruz de Carvalho M.H., d'Arcy-Lameta A., Roy-Macauley H., Gareil M., El Maarouf H., Pham-Thi A.-T., and Zuily-Fodil Y., *Aspartic protease in leaves of common bean (Phaseolus vulgaris L.) and cowpea (Vigna unguiculata L. Walp): enzymatic activity, gene expression and relation to drought susceptibility.* FEBS Letters, 2001. **492**: p. 242-246.
17. Timotijevic G.S., Radovic S.R., and Maksimovic V.R., *Characterization of an aspartic proteinase activity in Buckwheat (Fagopyrum esculentum Moench) seeds.* Journal of Agricultural and Food Chemistry, 2003. **51**: p. 2100-2104.

18. Milisavljevic M.D., Timotijevic G.S., Radovic S.R., Konstantinovic M.M., and Maksimovic V.R., *Two types of aspartic proteinases from buckwheat seed – Gene structure and expression analysis.* Journal of Plant Physiology, 2008. **165**: p. 983-990.
19. Cordeiro M., Xue Z.-T., Pietrzak M., Salomé Pais M., and Brodelius P., *Isolation and characterization of a cDNA from flowers of Cynara cardunculus encoding cyprosin (an aspartic proteinase) and its use to study the organ-specific expression of cyprosin.* Plant Molecular Biology, 1994. **24**(5): p. 733-741.
20. Veríssimo P., Faro C., Moir A.J.G., Lin Y., Tang J., and Pires E., *Purification, characterization and partial amino acid sequencing of two new aspartic proteinases from fresh flowers of Cynara cardunculus L.* European Journal of Biochemistry, 1996. **235**(3): p. 762-768.
21. Ramalho-Santos M., Pissarra J., Veríssimo P., Pereira S., Salema R., Pires E., and Faro C.J., *Cardosin A, an abundant aspartic proteinase, accumulates in protein storage vacuoles in the stigmatic papillae of Cynara cardunculus L.* Planta, 1997. **203**(2): p. 204-212.
22. Guo R., Xu X., Carole B., Li X., Gao M., Zheng Y., and Wang X., *Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape.* BMC Genomics, 2013. **14**: p. 554.
23. An C.-I., Fukusaki E.-i., and Kobayashi A., *Aspartic proteinases are expressed in pitchers of the carnivorous plant Nepenthes alata Blanco.* Planta, 2002. **214**: p. 661-667.
24. Athauda S.B.P., Matsumoto K., Rajapakshe S., Kuribayashi M., Kojima M., Kubomura-Yoshida N., Iwamatsu A., Shibata C., Inoue H., and Takahashi K., *Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases.* Biochem. J., 2004. **381**: p. 295-306.
25. Guevara M.G., Oliva C.R., Machinandiarena M., and Daleo G.R., *Purification and properties of an aspartic protease from potato tuber that is inhibited by a basic chitinase.* Physiologia Plantarum, 1999. **106**: p. 164-169.
26. Guevara M.G., veríssimo P., Pires E., Faro C., and daleo G.R., *Potato aspartic proteases: induction, antimicrobial activity and substrate specificity.* Journal of Plant Pathology, 2004. **86**(3): p. 233-238.
27. Chen H.-J., Huang Y.-H., Huang G.-J., Huang S.-S., Chow T.-J., and Lin Y.-H., *Sweet potato SPAP1 is a typical aspartic protease and participates in ethephon-mediated leaf senescence.* Journal of Plant Physiology, 2015. **180**: p. 1-17.
28. Asakura T., Watanabe H., Abe K., and Arai S., *Rice aspartic proteinase, Oryzasin, expressed during seed ripening and germination, has a gene organization distinct from those of animal and microbial aspartic proteinases.* European Journal of Biochemistry, 1995. **232**: p. 77-83.
29. Asakura T., Watanabe H., Abe K., and Arai S., *Oryzasin as an aspartic proteinase occurring in rice seeds: purification, characterization, and application to milk clotting.* Journal of Agricultural and Food Chemistry, 1997. **45**(4): p. 1070-1075.
30. Terauchi K., Asakura T., Nishizawa N., Matsumoto I., and Abe K., *Characterization of the genes for two soybean aspartic proteinases and analysis of their different tissue-dependent expression.* Planta, 2004. **218**(6): p. 947-957.
31. Nakano T., Murakami S., Shoji T., Yoshida S., Yamada Y., and Sato F., *A nove1 protein with DNA binding activity from Tobacco chloroplast nucleoids.* The Plant Cell, 1997. **9**: p. 1673-1682.
32. Schaller A. and Ryan C.A., *Molecular cloning of a tomato leaf cDNA encoding an aspartic protease, a systemic wound response protein.* Plant Molecular Biology, 1996. **31**: p. 1073-1077.
33. Schaaf A., Reski R., and Decker E.L., *A novel aspartic proteinase is targeted to the secretory pathway and to the vacuole in the moss Physcomitrella patens.* European Journal of Cell Biology, 2004. **83**(4): p. 145-152.

34. Faro C. and Gal S., *Aspartic proteinase content of the Arabidopsis genome.* Curr Protein Pept Sci., 2005. **6**(6): p. 493-500.
35. Chen J., Ouyang Y., Wang L., Xie W., and Zhang Q., *Aspartic proteases gene family in rice: Gene structure and expression, predicted protein features and phylogenetic relation.* Gene, 2009. **442**: p. 108-118.
36. Xia Y., Suzuki H., Borevitz J., Blount J., Guo Z., Patel K., Dixon R.A., and Lamb C., *An extracellular aspartic protease functions in Arabidopsis disease resistance signaling.* EMBO J, 2004. **23**(4): p. 980-988.
37. Ge X., Dietrich C., Matsuno M., Li G., Berg H., and Xia Y., *An Arabidopsis aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis.* EMBO Rep, 2005. **6**(3): p. 282-288.
38. Yao X., Xiong W., Ye T., and Wu Y., *Overexpression of the aspartic protease ASPG1 gene confers drought avoidance in Arabidopsis.* Journal of Experimental Botany, 2012. **63**(7): p. 2579-2593.
39. Gao Y., Nishikawa H., Badejo A.A., Shibata Y., Nakagawa T., Maruta T., Shigeoka S., Smirnoff N., and Ishikawa T., *Expression of aspartyl protease and C3HC4-type RING zinc finger genes are responsive to ascorbic acid in Arabidopsis thaliana.* Journal of Experimental Botany, 2011. **62**(10): p. 3647-3657.
40. Paparelli E., Gonzali S., Parlanti S., Novi G., Giorgi F.M., Licausi F., Kosmacz M., Feil R., Lunn J.E., Brust H., van Dongen J.T., Steup M., and Perata P., *Misexpression of a chloroplast aspartyl protease leads to severe growth defects and alters carbohydrate metabolism in Arabidopsis.* Plant Physiology, 2012. **160**: p. 1237-1250.
41. Phan H.A., Iacuone S., Li S.F., and Parish R.W., *The MYB80 Transcription Factor Is Required for Pollen Development and the Regulation of Tapetal Programmed Cell Death in Arabidopsis thaliana.* The Plant Cell Online, 2011. **23**(6): p. 2209-2224.
42. Breitenbach H.H., Wenig M., Wittek F., Jordá L., Maldonado-Alconada A.M., Sarioglu H., Colby T., Knappe C., Bichlmeier M., Pabst E., Mackey D., Parker J.E., and Vlot A.C., *Contrasting roles of apoplastic aspartyl protease AED1 and legume lectin-like protein LLP1 in Arabidopsis systemic acquired resistance.* Plant Physiology, 2014.
43. Li Y., Kabbage M., Liu W., and Dickman M.B., *Aspartyl protease mediated cleavage of AtBAG6 is necessary for autophagy and fungal resistance in plants.* The Plant Cell, 2016. **28**: p. 233-247.
44. Chen J., Ding J., Ouyang Y., Du H., Yang J., Cheng K., Zhao J., Qiu S., Zhang X., Yao J., Liu K., Wang L., Xu C., Li X., Xue Y., Xia M., Ji Q., Lu J., Xu M., and Zhang Q., *A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice.* Proceedings of the National Academy of Sciences, 2008. **105**(32): p. 11436-11441.
45. Huang J., Zhao X., Cheng K., Jiang Y., Ouyang Y., Xu C., Li X., Xiao J., and Zhang Q., *OsAP65, a rice aspartic protease, is essential for male fertility and plays a role in pollen germination and pollen tube growth.* Journal of Experimental Botany, 2013. **64**(11): p. 3351-3360.
46. Niu N., Liang W., Yang X., Jin W., Wilson Z.A., Hu J., and Zhang D., *EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice.* Nature Communications, 2013. **4**: p. 1445.
47. Prasad B.D., Creissen G., Lamb C., and Chattoo B.B., *Overexpression of rice (Oryza sativa L.) OsCDR1 leads to constitutive activation of defense responses in rice and Arabidopsis.* Mol Plant Microbe Interact, 2009. **22**(12): p. 1635-1644.
48. Alam M., Nakamura H., Ichikawa H., Miyao A., Hirochika H., Kobayashi K., Yamaoka N., and Nishiguchi M., *Response of an aspartic protease gene OsAP77 to fungal, bacterial and viral infections in rice.* Rice, 2014. **7**(1): p. 9.
49. Olivares J.E., Díaz-Camino C., Estrada-Navarrete G., Alvarado-Affantranger X., Rodríguez-Kessler M., Zamudio F.Z., Olamendi-Portugal T., Márquez Y., Servín L.E., and

Sánchez F., *Nodulin 41, a novel late nodulin of common bean with peptidase activity.* BMC Plant Biology, 2011. **11**: p. 134.

50. Chen F. and Foolad M.R., *Molecular organization of a gene in barley which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration.* Plant Molecular Biology, 1997. **35**(6): p. 821-831.
51. Bi X., Khush G.S., and Bennett J., *The Rice Nucellin Gene Ortholog OsAsp1 Encodes an Active Aspartic Protease Without a Plant-specific Insert and is Strongly Expressed in Early Embryo.* Plant and Cell Physiology, 2005. **46**(1): p. 87-98.
52. Frazão C., Bento I., Costa J., Soares C.M., Veríssimo P., Faro C., Pires E., Cooper J., and Carrondo M.A., *Crystal Structure of Cardosin A, a Glycosylated and Arg-Gly-Asp-containing Aspartic Proteinase from the Flowers of Cynara cardunculus L.* The Journal of Biological Chemistry, 1999. **274**(39): p. 27694-27701.
53. Simões I., Faro R., Bur D., and Faro C., *Characterization of Recombinant CDR1, an Arabidopsis Aspartic Proteinase Involved in Disease Resistance.* Journal of Biological Chemistry, 2007. **282**(43): p. 31358-31365.
54. Prasad B.D., Creissen G., Lamb C., and Chattoo B.B., *Heterologous expression and characterization of recombinant OsCDR1, a rice aspartic proteinase involved in disease resistance.* Protein Expression and Purification, 2010. **72**(2): p. 169-174.
55. Kato Y., Murakami S., Yamamoto Y., Chatani H., Kondo Y., Nakano T., Yokota A., and Sato F., *The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco.* Planta, 2004. **220**: p. 97-104.
56. Kato Y., Yamamoto Y., Murakami S., and Sato F., *Post-translational regulation of CND41 protease activity in senescent tobacco leaves.* Planta, 2005. **222**: p. 643-651.
57. Takahashi K., Niwa H., Yokota N., Kubota K., and Inoue H., *Widespread tissue expression of nepenthesin-like aspartic protease genes in Arabidopsis thaliana.* Plant Physiology and Biochemistry, 2008. **46**: p. 724-729.
58. Fejfarova K., Kadek A., Mrazek H., Hausner J., Tretyachenko V., Koval' T., Man P., Hasek J., and Dohnalek J., *Crystallization of nepenthesin I using a low-pH crystallization screen.* Acta Crystallographica Section F, 2016. **72**(1): p. 24-28.
59. Murakami S., Kondo Y., Nakano T., and Sato F., *Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells.* FEBS Letters, 2000. **468**(1): p. 15-18.
60. Kadek A., Mrazek H., Halada P., Rey M., Schriemer D.C., and Man P., *Aspartic Protease Nepenthesin-1 as a tool for digestion in hydrogen/deuterium exchange mass spectrometry.* Analytical Chemistry, 2014. **86**(9): p. 4287-4294.
61. Yang M., Hoeppner M., Rey M., Kadek A., Man P., and Schriemer D.C., *Recombinant nepenthesin II for hydrogen/deuterium exchange mass spectrometry.* Analytical Chemistry, 2015. **87**(13): p. 6681-6687.
62. Vitousek P. and Howarth R., *Nitrogen limitation on land and in the sea: How can it occur?* Biogeochemistry, 1991. **13**(2): p. 87-115.
63. Gould S.B., Waller R.F., and McFadden G.I., *Plastid Evolution.* Annual Review of Plant Biology, 2008. **59**(1): p. 491-517.
64. Nakano T., Nagata N., Kimura T., Sekimoto M., Kawaide H., Murakami S., Kaneko Y., Matsushima H., Kamiya Y., Sato F., and Yoshida S., *CND41, a chloroplast nucleoid protein that regulates plastid development, causes reduced gibberellin content and dwarfism in tobacco.* Physiologia Plantarum, 2003. **117**: p. 130-136.
65. Yang Y., Shah J., and Klessig D.F., *Signal perception and transduction in plant defense responses.* Genes & Development, 1997. **11**(13): p. 1621-1639.
66. Gozzo F. and Faoro F., *Systemic Acquired Resistance (50 Years after Discovery): Moving from the Lab to the Field.* Journal of Agricultural and Food Chemistry, 2013. **61**(51): p. 12473-12491.

67. Wagner S., Stuttmann J., Rietz S., Guerois R., Brunstein E., Bautor J., Niefind K., and Parker Jane E., *Structural Basis for Signaling by Exclusive EDS1 Heteromeric Complexes with SAG101 or PAD4 in Plant Innate Immunity.* Cell Host & Microbe, 2013. **14**(6): p. 619-630.
68. Doukhanina E.V., Chen S., van der Zalm E., Godzik A., Reed J., and Dickman M.B., *Identification and Functional Characterization of the BAG Protein Family in Arabidopsis thaliana.* Journal of Biological Chemistry, 2006. **281**(27): p. 18793-18801.
69. Kabbage M. and Dickman M.B., *The BAG proteins: a ubiquitous family of chaperone regulators.* Cellular and Molecular Life Sciences, 2008. **65**(9): p. 1390-1402.
70. Shimono M., Yazaki J., Nakamura K., Kishimoto N., Kikuchi S., Iwano M., Yamamoto K., Sakata K., Sasaki T., and Nishiguchi M., *cDNA microarray analysis of gene expression in rice plants treated with probenazole, a chemical inducer of disease resistance.* Journal of General Plant Pathology, 2003. **69**(1): p. 76-82.
71. Yamaguchi-Shinozaki K. and Shinozaki K., *Transcriptional regulatory networks in cellular responses and toleance to dehydration and cold stresses.* Annual Review of Plant Biology, 2006. **57**(1): p. 781-803.
72. Kwak J.M., Mori I.C., Pei Z.M., Leonhardt N., Torres M.A., Dangl J.L., Bloom R.E., Bodde S., Jones J.D.G., and Schroeder J.I., *NADPH oxidase AtrbohD and AtrbohF genes function in ROS-dependent ABA signaling in Arabidopsis.* The EMBO Journal, 2003. **22**(11): p. 2623-2633.
73. Lord E.M. and D R.S., *The mechanisms of pollination and fertilization in plants.* Annual Review of Cell and Developmental Biology, 2002. **18**: p. 81-105.
74. Liu K.D., Zhou Z.Q., Xu C.G., Zhang Q., and Saghai Maroof M.A., *An analysis of hybrid sterility in rice using a diallel cross of 21 parents involving indica, japonica and wide compatibility varieties.* Euphytica, 1996. **90**: p. 275-280.
75. Ji Q., Zhang M., Lu J., Wang H., Lin B., Liu Q., Chao Q., Zhang Y., Liu C., Gu M., and Xu M., *Molecular Basis Underlying the S5-Dependent Reproductive Isolation and Compatibility of indica/japonica Rice Hybrids.* Plant Physiology, 2012. **158**(3): p. 1319-1328.
76. Hafidh S., Fíla J., and Honys D., *Male gametophyte development and function in angiosperms: a general concept.* Plant Reproduction, 2016: p. 1-21.
77. Goldberg R.B., Beals T.P., and Sanders P.M., *Anther development: basic principles and practical applications.* The Plant Cell, 1993. **5**: p. 1217-1229.
78. Poethig R.S., *Clonal Analysis of Cell Lineage Patterns in Plant Development.* American Journal of Botany, 1987. **74**(4): p. 581-594.
79. Feng X. and Dickinson H.G., *Packaging the male germline in plants.* Trends in Genetics, 2007. **23**(10): p. 503-510.
80. Sanders M.P., Bui Q.A., Weterings K., McIntire N.K., Hsu Y.-C., Lee Y.P., Truong T.M., Beals P.T., and Goldberg B.R., *Anther developmental defects in Arabidopsis thaliana male-sterile mutants.* Sexual Plant Reproduction, 1999. **11**(6): p. 297-322.
81. Zhang D. and Yang L., *Specification of tapetum and microsporocyte cells within the anther.* Current Opinion in Plant Biology, 2014. **17**: p. 49-55.
82. Wu H.M. and Cheun A.Y., *Programmed cell death in plant reproduction.* Plant Mol Biol, 2000. **44**(3): p. 267-281.
83. Parish R.W. and Li S.F., *Death of a tapetum: A programme of developmental altruism.* Plant Science, 2010. **178**: p. 73-89.
84. Russell S.D. and Jones D.S., *The male germline of angiosperms: Repertoire of an inconspicuous but important cell lineage.* Frontiers in Plant Science, 2015. **6**.
85. Ito T., Nagata N., Yoshiba Y., Ohme-Takagi M., Ma H., and Shinozaki K., *Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development.* Plant Cell, 2007. **19**(11): p. 3549-3562.

86. Reape T.J. and McCabe P.F., *Apoptotic-like regulation of programmed cell death in plants.* Apoptosis, 2010. **15**(3): p. 249-256.
87. Huang L., Zhang C., Fang M., Xu H., and Cao M., *The Cell Death in CMS Plants.* International Journal of Bioscience, Biochemistry and Bioinformatics, 2011. **1**(4): p. 297-301.
88. Kempe K. and Gils M., *Pollination control technologies for hybrid breeding.* Molecular Breeding, 2011. **27**(4): p. 417-437.
89. Reape T.J. and McCabe P.F., *Apoptotic-like programmed cell death in plants.* New Phytol, 2008. **180**(1): p. 13-26.
90. Li S.F., Iacuone S., and Parish R.W., *Suppression and restoration of male fertility using a transcription factor.* Plant Biotechnology Journal, 2007. **5**(2): p. 297-312.
91. Zhang Z.-B., Zhu J., Gao J.-F., Wang C., Li H., Li H., Zhang H.-Q., Zhang S., Wang D.-M., Wang Q.-X., Huang H., Xia H.-J., and Yang Z.-N., *Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis.* The Plant Journal, 2007. **52**(3): p. 528-538.
92. Wilson Z.A., Morroll S.M., Dawson J., Swarup R., and Tighe P.J., *The Arabidopsis MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors.* The Plant Journal, 2001. **28**(1): p. 27-39.
93. Kohli A., Narciso J.O., Miro B., and Raorane M., *Root proteases: reinforced links between nitrogen uptake and mobilization and drought tolerance.* Physiologia Plantarum, 2012. **145**(1): p. 165-179.
94. Wendrich J.R. and Weijers D., *The Arabidopsis embryo as a miniature morphogenesis model.* New Phytologist, 2013. **199**(1): p. 14-25.
95. Petricka J.J., Winter C.M., and Benfey P.N., *Control of Arabidopsis Root Development.* Annual Review of Plant Biology, 2012. **63**(1): p. 563-590.
96. van den Berg C., Willemsen V., Hendriks G., Weisbeek P., and Scheres B., *Short-range control of cell differentiation in the Arabidopsis root meristem.* Nature, 1997. **390**(6657): p. 287-289.
97. Kidner C., Sendaresan V., Roberts K., and Dolan L., *Clonal analysis of the Arabidopsis root confirms that position, not lineage, determines cell fate.* Planta, 2000. **211**: p. 191-199.
98. Vilches-Barro A. and Maizel A., *Talking through walls: mechanisms of lateral root emergence in Arabidopsis thaliana.* Current Opinion in Plant Biology, 2015. **23**: p. 31-38.
99. Péret B., De Rybel B., Casimiro I., Benková E., Swarup R., Laplaze L., Beeckman T., and Bennett M.J., *Arabidopsis lateral root development: an emerging story.* Trends in Plant Science, 2009. **14**(7): p. 399-408.
100. Vermeer J.E.M. and Geldner N., *Lateral root initiation in Arabidopsis thaliana: a force awakens.* F1000Prime, 2015. **7**(32).
101. Hoorn R.A.L.v.d., *Plant Proteases: From Phenotypes to Molecular Mechanisms.* Annual Review of Plant Biology, 2008. **59**(1): p. 191-223.
102. Casamitjana-Martínez E., Hofhuis H.F., Xu J., Liu C.-M., Heidstra R., and Scheres B., *Root-Specific CLE19 Overexpression and the sol1/2 Suppressors Implicate a CLV-like Pathway in the Control of Arabidopsis Root Meristem Maintenance.* Current Biology, 2003. **13**(16): p. 1435-1441.
103. Combier J.-P., Vernié T., de Billy F., El Yahyaoui F., Mathis R., and Gamas P., *The MtMMPL1 Early Nodulin Is a Novel Member of the Matrix Metalloendoproteinase Family with a Role in Medicago truncatula Infection by Sinorhizobium meliloti.* Plant Physiology, 2007. **144**(2): p. 703-716.

104. Li Y., Zhou L., Li Y., Chen D., Tan X., Lei L., and Zhou J., *A nodule-specific plant cysteine proteinase, AsNODF32, is involved in nodule senescence and nitrogen fixation activity of the green manure legume Astragalus sinicus.* New Phytologist, 2008. **180**(1): p. 185-192.
105. Schoof H., Lenhard M., Haecker A., Mayer K.F.X., Jürgens G., and Laux T., *The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes.* Cell, 2000. **100**(6): p. 635-644.
106. Brand U., Fletcher J.C., Hobe M., Meyerowitz E.M., and Simon R., *Dependence of stem cell fate in Arabidopsis on a feedback loop regulated by CLV3 activity.* Science, 2000. **289**: p. 617-619.
107. Naito Y., Fujie M., Usami S., Murooka Y., and Yamada T., *The Involvement of a Cysteine Proteinase in the Nodule Development in Chinese Milk Vetch Infected with Mesorhizobium huakuii subsp. rengei.* Plant Physiology, 2000. **124**(3): p. 1087-1096.
108. Kardailsky I.V. and Brewin N.J., *Expression of Cysteine Proteases genes in Pea nodule development and senescence.* Molecular Plant-Microbe Interactions, 1996. **8**(9): p. 689-695.
109. Sheokand S., Dahiya P., Vincent J.L., and Brewin N.J., *Modified expression of cysteine protease affects seed germination, vegetative growth and nodule development in transgenic lines of Medicago truncatula.* Plant Science, 2005. **169**(5): p. 966-975.
110. Rodermel S.R., Abbott M.S., and Bogorad L., *Nuclear-organelle interactions: Nuclear antisense gene inhibits ribulose bisphosphate carboxylase enzyme levels in transformed tobacco plants.* Cell, 1988. **55**(4): p. 673-681.
111. Clough S.J. and Bent A.F., *Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.* The Plant Journal, 1998. **16**(6): p. 735-743.
112. Alexander M.P., *Differential Staining of Aborted and Nonaborted Pollen.* Stain Technology, 1969. **44**(3): p. 117-122.
113. Jefferson R.A., Kavanagh T.A., and Bevan M.W., *GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants.* The EMBO Journal, 1987. **6**(13): p. 3901-3907.
114. Winter D., Vinegar B., Nahal H., Ammar R., Wilson G.V., and Provart N.J., *An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets.* PLoS ONE, 2007. **2**(8): p. e718.
115. Martin F.W., *Staining and Observing Pollen Tubes in the Style by Means of Fluorescence.* Stain Technology, 1959. **34**(3): p. 125-128.
116. Ruffel S., Gojon A., and Lejay L., *Signal interactions in the regulation of root nitrate uptake.* Journal of Experimental Botany, 2014. **65**(19): p. 5509-5517.
117. Yruela I., *Plant development regulation: Overview and perspectives.* Journal of Plant Physiology, 2015. **182**: p. 62-78.
118. Overvoorde P., Fukaki H., and Beeckman T., *Auxin Control of Root Development.* Cold Spring Harbor Perspectives in Biology, 2010. **2**(6).
119. Saini S., Sharma I., Kaur N., and Pati P.K., *Auxin: a master regulator in plant root development.* Plant Cell Reports, 2013. **32**(6): p. 741-757.
120. Godlewski M. and Adamczyk B., *The ability of plants to secrete proteases by roots.* Plant Physiology and Biochemistry, 2007. **45**(9): p. 657-664.
121. Stornaiuolo M., Lotti L.V., Borgese N., Torrisi M.-R., Mottola G., Martire G., and Bonatti S., *KDEL and KKXX Retrieval Signals Appended to the Same Reporter Protein Determine Different Trafficking between Endoplasmic Reticulum, Intermediate Compartment, and Golgi Complex.* Molecular Biology of the Cell, 2003. **14**(3): p. 889-902.
122. D'Alessandro S., Golin S., Hardtke C.S., Lo Schiavo F., and Zottini M., *The co-chaperone p23 controls root development through the modulation of auxin distribution in the Arabidopsis root meristem.* Journal of Experimental Botany, 2015. **66**(16): p. 5113-5122.

123. Okushima Y., Fukaki H., Onoda M., Theologis A., and Tasaka M., *ARF7 and ARF19 Regulate Lateral Root Formation via Direct Activation of LBD/ASL Genes in Arabidopsis.* The Plant Cell, 2007. **19**(1): p. 118-130.
124. Dello Ioio R., Linhares F.S., Scacchi E., Casamitjana-Martinez E., Heidstra R., Costantino P., and Sabatini S., *Cytokinins Determine Arabidopsis Root-Meristem Size by Controlling Cell Differentiation.* Current Biology, 2007. **17**(8): p. 678-682.
125. Benjamins R. and Scheres B., *Auxin: The Looping Star in Plant Development.* Annual Review of Plant Biology, 2008. **59**(1): p. 443-465.
126. Almeida C.M., Pereira C., Costa D.S., Pereira S., Pissarra J., Simões I., and Faro C., *Chlapsin, a chloroplastidial aspartic proteinase from the green algae Chlamydomonas reinhardtii.* Planta, 2012. **236**(1): p. 283-296.
127. de Marco A., *Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli.* Microbial Cell Factories, 2009. **8**: p. 26-26.
128. Quan J. and Tian J., *Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways.* PLoS ONE, 2009. **4**(7): p. e6441.
129. Marillonnet S., Thoeringer C., Kandzia R., Klimyuk V., and Gleba Y., *Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants.* Nat Biotech, 2005. **23**(6): p. 718-723.
130. Nguyen V.D., Hatahet F., Salo K.E.H., Enlund E., Zhang C., and Ruddock L.W., *Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli.* Microbial Cell Factories, 2011. **10**: p. 1-1.
131. Schilling O. and Overall C.M., *Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites.* Nat Biotech, 2008. **26**(6): p. 685-694.
132. Cruz R., *New retroviral-like membrane-associated aspartic proteinases from rickettsiae: biochemical characterization and specificity profiling*, in *PhD Thesis*. 2014, University of Coimbra. p. 239.
133. Chambers M.C., Maclean B., Burke R., Amodei D., Ruderman D.L., Neumann S., Gatto L., Fischer B., Pratt B., Egertson J., Hoff K., Kessner D., Tasman N., Shulman N., Frewen B., Baker T.A., Brusniak M.-Y., Paulse C., Creasy D., Flashner L., Kani K., Moulding C., Seymour S.L., Nuwaysir L.M., Lefebvre B., Kuhlmann F., Roark J., Rainer P., Detlev S., Hemenway T., Huhmer A., Langridge J., Connolly B., Chadick T., Holly K., Eckels J., Deutsch E.W., Moritz R.L., Katz J.E., Agus D.B., MacCoss M., Tabb D.L., and Mallick P., *A Cross-platform Toolkit for Mass Spectrometry and Proteomics.* Nature biotechnology, 2012. **30**(10): p. 918-920.
134. Craig R. and Beavis R.C., *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-1467.
135. Keller A., Nesvizhskii A.I., Kolker E., and Aebersold R., *Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search.* Analytical Chemistry, 2002. **74**(20): p. 5383-5392.
136. Schilling O., Keller U., and Overall C.M., *Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites.* Biological Chemistry, 2011. **392**(11): p. 1031-1037.
137. Saeed A., Sharov V., White J., Li J., Liang W., Bhagabati N., Klapa M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Ryltsov A., Kostukovich E., Borisovky I., Liu Z., Vinsavich A., Trush V., and Quackenbush J., *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques, 2003. **34**(2): p. 374-378.
138. Rosano G.L. and Ceccarelli E.A., *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in Microbiology, 2014. **5**: p. 172.
139. Costa S.J., Almeida A., Castro A., Domingues L., and Besir H., *The novel Fh8 and H fusion partners for soluble protein expression in Escherichia coli: a comparison with the*

traditional gene fusion technology. Applied Microbiology and Biotechnology, 2012. **97**(15): p. 6779-6791.

140. Nagels B., Van Damme E.J.M., Callewaert N., Zabeau L., Tavernier J., Delanghe J.R., Boets A., Castilho A., and Weterings K., *Biologically active, magnICON®-expressed EPO-Fc from stably transformed Nicotiana benthamiana plants presenting tetra-antennary N-glycan structures.* Journal of Biotechnology, 2012. **160**(3–4): p. 242-250.
141. Klimyuk V., Pogue G., Herz S., Butler J., and Haydon H., *Production of Recombinant Antigens and Antibodies in Nicotiana benthamiana Using 'Magnifection' Technology: GMP-Compliant Facilities for Small- and Large-Scale Manufacturing*, in *Plant Viral Vectors*, K. Palmer and Y. Gleba, Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 127-154.
142. Niknejad A., Webster D., and Bhave M., *Production of bioactive wheat puroindoline proteins in Nicotiana benthamiana using a virus-based transient expression system.* Protein Expression and Purification, 2015.
143. Merlin M., Gecchele E., Arcalis E., Remelli S., Brozzetti A., Pezzotti M., and Avesani L., *Enhanced GAD65 production in plants using the MagnICON transient expression system: Optimization of upstream production and downstream processing.* Biotechnology Journal, 2016. **11**(4): p. 542-553.
144. Lichty J.J., Malecki J.L., Agnew H.D., Michelson-Horowitz D.J., and Tan S., *Comparison of affinity tags for protein purification.* Protein Expression and Purification, 2005. **41**(1): p. 98-105.
145. Gupta R. and Brunak S., *Prediction of glycosylation across the human proteome and the correlation to protein function.* Pacific Symposium on Biocomputing, 2002. **7**: p. 310-322.
146. Trimble R.B., Tarentino A.L., Evans G., and Maley F., *Purification and properties of endo-beta-N-acetylglucosaminidase L from Streptomyces plicatus.* Journal of Biological Chemistry, 1979. **254**(19): p. 9708-9713.
147. Plummer T.H., Elder J.H., Alexander S., Phelan A.W., and Tarentino A.L., *Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations.* Journal of Biological Chemistry, 1984. **259**(17): p. 10700-10704.
148. Tarentino A.L., Gomez C.M., and Plummer T.H., *Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F.* Biochemistry, 1985. **24**(17): p. 4665-4671.
149. Petersen T.N., Brunak S., von Heijne G., and Nielsen H., *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nature Methods, 2011. **8**(10): p. 785-786.
150. Schilling O., Huesgen P.F., Barre O., auf dem Keller U., and Overall C.M., *Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry.* Nat. Protocols, 2011. **6**(1): p. 111-120.
151. Schechter I. and Berger A., *On the size of the active site in proteases. I. Papain.* Biochemical and Biophysical Research Communications, 1967. **27**(2): p. 157-162.
152. Eckhard U., Huesgen P.F., Schilling O., Bellac C.L., Butler G.S., Cox J.H., Dufour A., Goebeler V., Kappelhoff R., auf dem Keller U., Klein T., Lange P.F., Marino G., Morrison C.J., Prudova A., Rodriguez D., Starr A.E., Wang Y., and Overall C.M., *Active site specificity profiling datasets of matrix metalloproteinases (MMPs) 1, 2, 3, 7, 8, 9, 12, 13 and 14.* Data in Brief, 2016. **7**: p. 299-310.
153. Zelanis A., Huesgen P.F., Oliveira A.K., Tashima A.K., Serrano S.M.T., and Overall C.M., *Snake venom serine proteinases specificity mapping by proteomic identification of cleavage sites.* Journal of Proteomics, 2015. **113**: p. 260-267.
154. Eckhard U., Huesgen P.F., Brandstetter H., and Overall C.M., *Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen.* Journal of Proteomics, 2014. **100**: p. 102-114.

155. Leal A.R., Cruz R., Bur D., Huesgen P.F., Faro R., Manadas B., Wlodawer A., Faro C., and Simões I., *Enzymatic properties, evidence for in vivo expression, and intracellular localization of shewasin D, the pepsin homolog from Shewanella denitrificans.* Scientific Reports, 2016. **6**: p. 23869.
156. Griesbeck C., Kobl I., and Heitzer M., *Chlamydomonas reinhardtii.* Molecular Biotechnology, 2006. **34**(2): p. 213-223.
157. Arnold D., Keilholz W., Schild H., Dumrese T., Stevanović S., and Rammensee H.-G., *Substrate Specificity of Cathepsins D and E Determined by N-Terminal and C-terminal Sequencing of Peptide Pools.* European Journal of Biochemistry, 1997. **249**(1): p. 171-179.
158. Beyer B.B., Johnson J.V., Chung A.Y., Li T., Madabushi A., Agbandje-McKenna M., McKenna R., Dame J.B., and Dunn B.M., *Active-Site Specificity of Digestive Aspartic Peptidases from the Four Species of Plasmodium that Infect Humans Using Chromogenic Combinatorial Peptide Libraries.* Biochemistry, 2005. **44**(6): p. 1768-1779.
159. Castanheira P., Samyn B., Sergeant K., Clemente J.C., Dunn B.M., Pires E., Van Beeumen J., and Faro C., *Activation, Proteolytic Processing, and Peptide Specificity of Recombinant Cardosin A.* Journal of Biological Chemistry, 2005. **280**(13): p. 13047-13054.
160. Dunn B.M. and Hung S.-H., *The two sides of enzyme–substrate specificity: lessons from the aspartic proteinases.* Biochimica et Biophysica Acta, 2000. **1477**(1–2): p. 231-240.
161. Koelsch G., Tang J., Loy J.A., Monod M., Jackson K., Foundling S.I., and Lin X., *Enzymic characteristics of secreted aspartic proteases of Candida albicans.* Biochimica et Biophysica Acta 2000. **1480**(1–2): p. 117-131.
162. Turner R.T., Koelsch G., Hong L., Castenheira P., Ghosh A., and Tang J., *Subsite Specificity of Memapsin 2 (β-Secretase): Implications for Inhibitor Design.* Biochemistry, 2001. **40**(34): p. 10001-10006.
163. Ahn J., Cao M.-J., Yu Y.Q., and Engen J.R., *Accessing the Reproducibility and Specificity of Pepsin and other Aspartic Proteases.* Biochimica et biophysica acta, 2013. **1834**(6): p. 1222-1229.
164. Shintani T., Kobayashi M., and Ichishima E., *Characterization of the S1 Subsite Specificity of Aspergillopepsin I by Site-Directed Mutagenesis.* Journal of Biochemistry, 1996. **120**(5): p. 974-981.
165. Shintani T., Nomura K., and Ichishima E., *Engineering of Porcine Pepsin: alteration of S1 substrate specificity of Pepsin to those of fungal Aspartic Proteinases by site-directed mutagenesis. .* Journal of Biological Chemistry, 1997. **272**(30): p. 18855-18861.
166. Shintani T. and Ichishima E., *Primary structure of aspergillopepsin I deduced from nucleotide sequence of the gene and aspartic acid-76 is an essential active site of the enzyme for trypsinogen activation.* Biochimica et Biophysica Acta, 1994. **1204**(2): p. 257-264.
167. Hsu I.N., Delbaere L.T.J., James M.N.G., and Hofmann T., *Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 [angst] and sequence homology with porcine pepsin.* Nature, 1977. **266**(5598): p. 140-145.
168. Takahashi K., *The amino acid sequence of rhizopuspepsin, an aspartic proteinase from Rhizopus chinensis.* Journal of Biological Chemistry, 1987. **262**(4): p. 1468-78.
169. Barkholt V., *Amino acid sequence of endothiapepsin.* European Journal of Biochemistry, 1987. **167**(2): p. 327-331.
170. Hube B., Turver C.J., Odds F.C., Eiffert H., Boulnois G.J., Köchel H., and Rüchel R., *Sequence of the Candida albicans gene encoding the secretory aspartate proteinase.* Journal of Medical and Veterinary Mycology, 1991. **29**(2): p. 129-132.
171. Faro C., Ramalho-Santos M., Vieira M., Mendes A., Simões I., Andrade R., Veríssimo P., Lin X.-l., Tang J., and Pires E., *Cloning and Characterization of cDNA Encoding Cardosin*

A, an RGD-containing Plant Aspartic Proteinase. Journal of Biological Chemistry, 1999. **274**(40): p. 28724-28729.

172. Sepulveda P., Marciniszyn J., Liu D., and Tang J., *Primary structure of porcine pepsin. III. Amino acid sequence of a cyanogen bromide fragment, CB2A, and the complete structure of porcine pepsin.* Journal of Biological Chemistry, 1975. **250**(13): p. 5082-8.
173. Jorrín-Novo J.V., Pascual J., Sánchez-Lucas R., Romero-Rodríguez M.C., Rodríguez-Ortega M.J., Lenz C., and Valledor L., *Fourteen years of plant proteomics reflected in Proteomics: Moving from model species and 2DE-based approaches to orphan species and gel-free platforms.* PROTEOMICS, 2015. **15**(5-6): p. 1089-1112.
174. Huesgen P.F. and Overall C.M., *N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification.* Physiologia Plantarum, 2012. **145**(1): p. 5-17.
175. Zhang H., Deery M.J., Gannon L., Powers S.J., Lilley K.S., and Theodoulou F.L., *Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots.* Proteomics, 2015. **15**(14): p. 2447-2457.
176. Boersema P.J., Raijmakers R., Lemeer S., Mohammed S., and Heck A.J.R., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat. Protocols, 2009. **4**(4): p. 484-494.
177. Rappsilber J., Mann M., and Ishihama Y., *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.* Nat. Protocols, 2007. **2**(8): p. 1896-1906.
178. Cox J. and Mann M., *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotech, 2008. **26**(12): p. 1367-1372.
179. Cox J., Neuhauser N., Michalski A., Scheltema R.A., Olsen J.V., and Mann M., *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.* Journal of Proteome Research, 2011. **10**(4): p. 1794-1805.
180. Bantscheff M., Schirle M., Sweetman G., Rick J., and Kuster B., *Quantitative mass spectrometry in proteomics: a critical review.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1017-1031.
181. Ong S.-E. and Mann M., *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. **1**(5): p. 252-262.
182. Ross P.L., Huang Y.N., Marchese J.N., Williamson B., Parker K., Hattan S., Khainovski N., Pillai S., Dey S., Daniels S., Purkayastha S., Juhasz P., Martin S., Bartlet-Jones M., He F., Jacobson A., and Pappin D.J., *Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents.* Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.
183. Harsha H.C., Molina H., and Pandey A., *Quantitative proteomics using stable isotope labeling with amino acids in cell culture.* Nat. Protocols, 2008. **3**(3): p. 505-516.
184. Hsu J.-L., Huang S.-Y., Chow N.-H., and Chen S.-H., *Stable-Isotope Dimethyl Labeling for Quantitative Proteomics.* Analytical Chemistry, 2003. **75**(24): p. 6843-6852.
185. Thomas P.D., Campbell M.J., Kejariwal A., Mi H., Karlak B., Daverman R., Diemer K., Muruganujan A., and Narechania A., *PANTHER: A Library of Protein Families and Subfamilies Indexed by Function.* Genome Research, 2003. **13**(9): p. 2129-2141.
186. Mi H., Lazareva-Ulitsky B., Loo R., Kejariwal A., Vandergriff J., Rabkin S., Guo N., Muruganujan A., Doremieux O., Campbell M.J., Kitano H., and Thomas P.D., *The PANTHER database of protein families, subfamilies, functions and pathways.* Nucleic Acids Research, 2005. **33**(suppl 1): p. D284-D288.
187. Smith A.D., *Oxford Dictionary of Biochemistry and Molecular Biology*. 2000: Oxford University Press.
188. Szklarczyk D., Franceschini A., Wyder S., Forslund K., Heller D., Huerta-Cepas J., Simonovic M., Roth A., Santos A., Tsafou K.P., Kuhn M., Bork P., Jensen L.J., and

von Mering C., *STRING v10: protein–protein interaction networks, integrated over the tree of life.* Nucleic Acids Research, 2015. **43**(Database issue): p. D447-D452.

189. Guo D., Wong W.S., Xu W.Z., Sun F.F., Qing D.J., and Li N., *Cis-cinnamic acid-enhanced 1 gene plays a role in regulation of Arabidopsis bolting.* Plant Molecular Biology, 2011. **75**(4): p. 481-495.
190. Siemens J., Keller I., Sarx J., Kunz S., Schuller A., Nagel W., Schmülling T., Parniske M., and Ludwig-Müller J., *Transcriptome Analysis of Arabidopsis Clubroots Indicate a Key Role for Cytokinins in Disease Development.* Molecular Plant-Microbe Interactions, 2006. **19**(5): p. 480-494.
191. Link A.J., Eng J., Schieltz D.M., Carmack E., Mize G.J., Morris D.R., Garvik B.M., and Yates J.R., *Direct analysis of protein complexes using mass spectrometry.* Nat Biotech, 1999. **17**(7): p. 676-682.
192. Gygi S.P., Rist B., Gerber S.A., Turecek F., Gelb M.H., and Aebersold R., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotech, 1999. **17**(10): p. 994-999.
193. Washburn M.P., Wolters D., and Yates J.R., *Large-scale analysis of the yeast proteome by multidimensional protein identification technology.* Nat Biotech, 2001. **19**(3): p. 242-247.
194. Kleifeld O., Doucet A., auf dem Keller U., Prudova A., Schilling O., Kainthan R.K., Starr A.E., Foster L.J., Kizhakkedathu J.N., and Overall C.M., *Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products.* Nat Biotech, 2010. **28**(3): p. 281-288.
195. Nessler C.L. and Burnett R.J., *Organization of the major latex protein gene family in opium poppy.* Plant Molecular Biology, 1992. **20**(4): p. 749-752.
196. Litholdo C.G., Parker B.L., Eamens A.L., Larsen M.R., Cordwell S.J., and Waterhouse P.M., *Proteomic identification of putative microRNA394 target genes in Arabidopsis thaliana identifies MLP family members critical for normal development.* Molecular & Cellular Proteomics, 2016.
197. Ruperti B., Bonghi C., Ziliotto F., Pagni S., Rasori A., Varotto S., Tonutti P., Giovannoni J.J., and Ramina A., *Characterization of a major latex protein (MLP) gene down-regulated by ethylene during peach fruitlet abscission.* Plant Science, 2002. **163**(2): p. 265-272.
198. Pozueta-Romero J., Klein M., Houlné G., Schantz M.-L., Meyer B., and Schantz R., *Characterization of a family of genes encoding a fruit-specific wound-stimulated protein of bell pepper (Capsicum annuum): identification of a new family of transposable elements.* Plant Molecular Biology, 1995. **28**(6): p. 1011-1025.
199. Marti L., Stefano G., Tamura K., Hawes C., Renna L., Held M.A., and Brandizzi F., *A missense mutation in the vacuolar protein GOLD36 causes organizational defects in the ER and aberrant protein trafficking in the plant secretory pathway.* The Plant Journal, 2010. **63**(6): p. 901-913.
200. Bartel B. and Fink G.R., *Differential regulation of an auxin-producing nitrilase gene family in Arabidopsis thaliana.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(14): p. 6649-6653.
201. Pollmann S., Müller A., Piotrowski M., and Weiler E.W., *Occurrence and formation of indole-3-acetamide in Arabidopsis thaliana.* Planta, 2002. **216**(1): p. 155-161.
202. Nakabayashi R., Yonekura-Sakakibara K., Urano K., Suzuki M., Yamada Y., Nishizawa T., Matsuda F., Kojima M., Sakakibara H., Shinozaki K., Michael A.J., Tohge T., Yamazaki M., and Saito K., *Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids.* The Plant Journal, 2014. **77**(3): p. 367-379.
203. Wang Y., Zhao W., Hao J., Xu W., Luo Y., Wu W., Yang Z., Liang Z., and Huang K., *Changes in biosynthesis and metabolism of glutathione upon ochratoxin A stress in Arabidopsis thaliana.* Plant Physiology and Biochemistry, 2014. **79**: p. 10-18.

204. Dixon David P., Sellars Jonathan D., and Edwards R., *The Arabidopsis phi class glutathione transferase AtGSTF2: binding and regulation by biologically active heterocyclic ligands.* Biochemical Journal, 2011. **438**(1): p. 63-70.
205. Lee S.C., Hwang I.S., and Hwang B.K., *Overexpression of the pepper antimicrobial protein CaAMP1 gene regulates the oxidative stress- and disease-related proteome in Arabidopsis.* Planta, 2011. **234**(6): p. 1111-1125.
206. Passaia G., Queval G., Bai J., Margis-Pinheiro M., and Foyer C.H., *The effects of redox controls mediated by glutathione peroxidases on root architecture in Arabidopsis thaliana.* Journal of Experimental Botany, 2014. **65**(5): p. 1403-1413.
207. Stenzel I., Otto M., Delker C., Kirmse N., Schmidt D., Miersch O., Hause B., and Wasternack C., *ALLENE OXIDE CYCLASE (AOC) gene family members of Arabidopsis thaliana: tissue- and organ-specific promoter activities and in vivo heteromerization(*).* Journal of Experimental Botany, 2012. **63**(17): p. 6125-6138.
208. Davies R.T., Goetz D.H., Lasswell J., Anderson M.N., and Bartel B., *IAR3 encodes an auxin conjugate hydrolase from Arabidopsis.* The Plant Cell, 1999. **11**(3): p. 365-376.
209. Woldemariam M.G., Onkokesung N., Baldwin I.T., and Galis I., *Jasmonoyl-l-isoleucine hydrolase 1 (JIH1) regulates jasmonoyl-l-isoleucine levels and attenuates plant defenses against herbivores.* The Plant Journal, 2012. **72**(5): p. 758-767.
210. Zhang T., Poudel A.N., Jewell J.B., Kitaoka N., Staswick P., Matsuura H., and Koo A.J., *Hormone crosstalk in wound stress response: wound-inducible amidohydrolases can simultaneously regulate jasmonate and auxin homeostasis in Arabidopsis thaliana.* Journal of Experimental Botany, 2016. **67**(7): p. 2107-2120.
211. Lerouxel O., Mouille G., Andème-Onzighi C., Bruyant M.-P., Séveno M., Loutelier-Bourhis C., Driouich A., Höfte H., and Lerouge P., *Mutants in DEFECTIVE GLYCOSYLATION, an Arabidopsis homolog of an oligosaccharyltransferase complex subunit, show protein underglycosylation and defects in cell differentiation and growth.* The Plant Journal, 2005. **42**(4): p. 455-468.
212. Qin C., Li Y., Gan J., Wang W., Zhang H., Liu Y., and Wu P., *OsDGL1, a Homolog of an Oligosaccharyltransferase Complex Subunit, is Involved in N-Glycosylation and Root Development in Rice.* Plant and Cell Physiology, 2013. **54**(1): p. 129-137.
213. Ham B.-K., Li G., Kang B.-H., Zeng F., and Lucas W.J., *Overexpression of Arabidopsis Plasmodesmata Germin-Like Proteins Disrupts Root Growth and Development.* The Plant Cell, 2012. **24**(9): p. 3630-3648.
214. Bailly A., Sovero V., Vincenzetti V., Santelia D., Bartnik D., Koenig B.W., Mancuso S., Martinoia E., and Geisler M., *Modulation of P-glycoproteins by Auxin Transport Inhibitors Is Mediated by Interaction with Immunophilins.* Journal of Biological Chemistry, 2008. **283**(31): p. 21817-21826.

Supplementary Material

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem. Peptides identified by LC-MS/MS spectrum-to-sequence assignment with Mascot and X!Tandem are listed with PeptideProphet probability score and one exemplary accession number of a matching Uniprot protein entry is listed. This data was further processed and rendered non-redundant for generation of cleavage specificity profiles.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
AAAAEIDEEPVSK	0.788787	7,233,453	E9PAV3
AAAIHAQVPED	0.693103	6,052,823	Q2M2U4
AAANPIESQWNPK	0.94304	7,713,746	P33991
AAANPISGHYDR	0.953512	6,803,093	Q2KIZ8
AAANPVYGR	0.672575	5,037,426	P49739
AAAVPLDPVLNP	0.0594665	4,222,252	Q9RZA4
AADESTGSIAK	0.329608	5,832,741	P04075
AADESTGSIAKR	0.895413	6,613,247	P04075
AADESTGSIAKR	0.240814	6,613,247	P04075
AAGAVAQL	0.119045	3,947,024	Q9MUT9
AAGGTAVGTGLNTR	0.999769	6,673,303	P07954
AAGGTAVGTGLNTR	0.977285	6,673,303	P07954
AAGGTAVGTGLNTR	0.965751	6,673,303	P07954
AALCSEAALQAIR	0.999997	7,313,633	Q3ZBT1
AALEPPEPK	0.0389542	5,342,759	Q2KJ93
AALEPPEPK	0.0298899	5,342,759	Q2KJ93
AALEPPEPK	0.0141065	5,342,759	Q2KJ93
AALGGNSSPSAK	0.999726	5,882,901	P42899
AALGGNSSPSAK	0.999452	5,882,901	P42899
AALGGNSSPSAK	0.777718	5,882,901	P42899
AAMAARPHSIDGR	0.999998	4,808,995	Q2HJ60
AAMAARPHSIDGR	0.994067	4,808,995	Q2HJ60
AANTFTITGHAETK	0.998637	7,893,852	P20290
AATLLANHSLR	0.995469	6,278,350	P13489
AATLLANHSLR	0.94265	6,278,350	P13489
AATLLANHSLR	0.868921	6,278,350	P13489
AAVDTSSEITTK	0.920182	6,698,267	P01252
AAYKPGMSPEECR	0.997616	8,063,520	Q3SZC2
ACGRPQATAVYK	0.0199744	7,193,527	Q3SWY3
ACPIVDSIEGVTHALR	0.716825	9,134,506	P07814
ADGAQEADLAVDA	0.0414372	6,672,826	Q53JI9
AEAMAALVLMDHLLR	0.0501482	8,714,436	Q3J713
AEEAAEIAISSPLLE	0.167456	8,158,979	O51929
AEPANAGKYICQI	0.115284	7,758,709	Q8WZ42
AEVDAAMAARPHSIDGR	0.755319	6,189,578	Q2HJ60
AFTDVHFKPGQIR	0.993311	5,446,171	P29144
AFTDVHFKPGQIR	0.965471	5,446,171	P29144
AFVNWINK	0.12774	5,542,866	P13796

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
AFVNWINK	0.0821242	5,542,866	P13796
AFVNWINK	0.0424958	5,542,866	P13796
AGFSLAEYLR	0.0462037	6,077,975	B8J8E5
AGHQTSAESWGTGR	0.999997	7,668,312	Q9SF40
AGPVAEYLK	0.998409	5,322,785	Q01518
AGPVAEYLK	0.998206	5,322,785	Q01518
AGSFGGAGGHAPGVAR	0.999997	7,288,413	P52272
AGSFGGAGGHAPGVAR	0.462241	7,288,411	P52272
AHACVTGKPISQGGIHGR	0.999997	4,912,505	P00366
AHGASPDVEDFLKR	0.0575959	5,532,718	C5CIU9
AHILSPWGAEVK	0.206031	7,123,741	P09874
AHLGGGSTSR	0.0270625	5,157,406	P04843
AHNLSMLALR	0.150572	6,153,103	Q8XD86
AISAGHGLPAK	0.730551	5,693,081	Q9QZQ8
ALFPPVEFPAPR	0.890948	7,148,710	P49327
ALPQALAQHLI	0.0170311	6,318,501	Q98C21
ALSGNEELTVK	0.0195956	6,388,266	P14625
ALYLVSGQWSSAALA	0.0406228	8,129,058	A3D4A1
ANGAVIHLALK	0.989505	6,118,526	A2VDN6
APKPGPYVK	0.996434	5,508,125	Q01518
APKPGPYVK	0.957651	5,508,125	Q01518
APPPAIVDEK	0.124491	5,768,021	Q9VW60
APPWTEDCR	0.0388275	6,102,474	Q13510
AQASHLSNLVPEP	0.536498	3,634,305	D3ZCM9
AQLQAVPAEFECIHPEK	0.999998	6,950,045	Q99829
AQLQAVPAEFECIHPEK	0.999546	6,950,045	Q99829
AQPTLLSVAFSR	0.0171901	6,893,636	Q9VQ36
AQVSIEGQI	0.620037	5,167,551	Q04JC5
ASIHSYV	0.191556	4,326,996	P66800
ASIHSYV	0.0201966	4,326,996	P66800
ASLKLPPKSEVSSDEDIQFR	0.997549	7,974,140	P40939
ASLKLPPKSEVSSDEDIQFR	0.951027	7,974,140	P40939
ASSPATCLKISDI	0.0168088	4,935,793	Q8K1Z6
ATAASSSSLEK	0.999369	5,842,819	P60712
ATAASSSSLEK	0.998457	5,842,819	P60712
ATAASSSSLEK	0.998134	5,842,819	P60712
ATAASSSSLEK	0.997588	5,842,819	P60712
ATAASSSSLEK	0.997417	5,842,819	P60712
ATAASSSSLEK	0.996283	5,842,819	P60712
ATAASSSSLEK	0.996125	5,842,819	P60712

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
ATAASSSSLEK	0.995605	5,842,819	P60712
ATAASSSSLEK	0.990124	5,842,819	P60712
ATAASSSSLEK	0.979944	5,842,819	P60712
ATAASSSSLEK	0.940005	5,842,819	P60712
ATAASSSSLEK	0.91598	5,842,819	P60712
ATAASSSSLEK	0.90458	5,842,819	P60712
ATAASSSSLEKR	0.807272	6,623,325	P53483
AVCNTPHER	0.287936	5,862,530	Q0MQA1
AVFPSIVGRPR	0.633994	6,438,557	P60712
AVKVHSFPTLK	0.99985	4,576,005	P05307
AVKVHSFPTLK	0.99955	4,576,005	P05307
AVKVHSFPTLK	0.998351	4,576,005	P05307
AVLCPPRYPK	0.973088	6,588,465	Q5E9B7
AVLCPPRYPK	0.266976	6,588,465	Q5E9B7
AVLCPPRYPK	0.198474	6,588,465	Q5E9B7
AVLCPPRYPK	0.157354	6,588,465	Q5E9B7
AVTYTEHAK	0.10114	5,682,764	Q757K0
AYEDKNDIPK	0.0421973	6,688,265	Q04760
CAGRVPEEDLKR	0.283995	5,159,220	Q2NKZ1
CAGYLEGGKDSCQGDSGGPVVCSGK	0.999998	6,730,396	P00760
CAIHAKR	0.582737	4,862,495	P68432
CALEGRDPELGLK	0.995968	7,873,895	P49411
CALEGRDPELGLK	0.677551	7,873,895	P49411
CAPPGNSDPEQLKK	0.971917	8,429,056	Q96G03
CGDLNVAHEEIDLR	0.0769347	8,648,878	A1YES6
CHNCAVEFNFGQK	0.999999	5,762,428	Q00839
CHNCAVEFNFGQK	0.999999	5,762,428	Q00839
CHNCAVEFNFGQK	0.999911	5,762,428	Q00839
CHPGHPGGVDP	0.0198777	4,064,992	Q1QWC2
CIQALPEFDGK	0.345745	6,973,284	Q58FF3
CIQALPEFDGKR	0.999504	7,753,789	Q58FF3
CIQALPEFDGKR	0.972969	7,753,789	Q58FF3
CIQALPEFDGKR	0.938045	7,753,789	Q58FF3
CLSPAVIVGLLK	0.966702	6,933,986	O43175
CMVTGGANLGR	0.99786	6,122,703	Q76N24
CMVTGGANLGR	0.937547	6,122,703	Q76N24
CNAACQTLLSLP	0.883558	7,183,227	Q8BS90
CPIEKPEEIPDD	0.0633305	7,793,444	Q9SD70
CTSLFPTIHGNDEVKR	0.992229	6,639,890	Q2KIZ8
CVISGVPGLK	0.0313103	5,733,067	Q92616

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
DAALALAAEISSK	0.991512	6,883,607	Q13011
DAALALAAEISSK	0.987908	6,883,607	Q13011
DAATQQLLSR	0.774603	5,957,955	P19483
DAATQQLLSR	0.35196	5,957,955	P19483
DADGGPLHTR	0.98815	5,637,511	Q99829
DAHLQQDPDAK	0.997616	6,773,091	P31153
DCGPESSKK	0.995687	5,762,574	P00558
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	4,435,231	P27592
DCHTAHIACK	0.999999	6,647,810	P27592
DCHTAHIACK	0.999998	4,435,231	P27592
DCHTAHIACK	0.999997	4,435,231	P27592
DCHTAHIACK	0.999971	4,435,231	P27592
DCHTAHIACK	0.999956	4,435,231	P27592
DCHTAHIACK	0.999874	6,647,810	P27592
DCHTAHIACK	0.998145	6,647,810	P27592
DCHTAHIACK	0.993869	6,647,810	P27592
DCILPPTRPTDKPLR	0.993466	6,323,342	P68103
DCILPPTRPTDKPLR	0.974515	6,323,342	P68103
DCILPPTRPTDKPLR	0.97365	6,323,342	P68103
DCILPPTRPTDKPLR	0.973339	6,323,342	P68103
DCILPPTRPTDKPLR	0.859228	6,323,342	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999999	7,206,436	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999999	7,206,436	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999999	7,206,436	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999999	7,206,436	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999755	9,605,225	P68103
DCILPPTRPTDKPLRLPLQDVYK	0.999137	7,206,436	P68103
DDGLMNKPLELLPK	0.0266459	5,763,089	A6ZZE9
DDTVSPQR	0.0866585	5,032,191	B5Y1B5
DESGPSIVHR	0.974514	5,927,721	P60712
DESGPSIVHR	0.971578	5,927,721	P60712
DFEQEMATAASSSSLEK	0.999999	9,739,217	P60712
DFEQEMQTAASSSSLEK	0.999599	6,686,241	P17126
DFEQEMQTAASSSSLEK	0.968343	6,686,241	P17126
DGQGSEHTVDK	0.999966	6,447,775	P00918

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
DGQGSEHTVDK	0.99995	6,447,775	P00918
DGQGSEHTVDK	0.999798	6,447,775	P00918
DGQGSEHTVDK	0.999764	6,447,775	P00918
DGQGSEHTVDK	0.982961	6,447,775	P00918
DGQGSEHTVDKK	0.999952	4,822,295	P00918
DGQGSEHTVDKK	0.998414	4,822,295	P00918
DGQGSEHTVDKK	0.980159	4,822,295	P00918
DGQGSEHTVDKK	0.768703	4,822,295	P00918
DGVQGPVGLPGPAGPA	0.0679717	7,388,614	Q28083
DIAALVVDNGSGMCK	0.999982	8,333,861	P60712
DIAALVVDNGSGMCK	0.999959	8,333,861	P60712
DICFTSVQK	0.0630577	6,072,834	Q3SZ62
DICFTSVQK	0.0459707	6,072,834	Q3SZ62
DICFTSVQKR	0.981448	6,853,340	Q3SZ62
DICFTSVQKR	0.685733	6,853,340	Q3SZ62
DINLAAEPK	0.277847	5,437,788	O60812
DINLAAEPK	0.226973	5,437,788	O60812
DINMAGEPKPDRPK	0.999318	5,712,885	Q9UKM9
DLPGFTNK	0.107268	5,042,472	Q19VA6
DLVDVVKNLSMF	0.0274141	5,045,878	Q01000
DLVDVVKNLSMF	0.0266777	5,045,878	Q01000
DMPSVLPEGLINGGVPL	0.182752	5,993,039	O69110
DNENIAHLACK	0.674391	4,675,482	Q60LV7
DQTEYLEER	0.785454	6,357,666	Q76LV2
DQTEYLEER	0.132821	6,357,666	Q76LV2
DQTNVSAAAQR	0.999682	6,247,857	Q00839
DQTNVSAAAQR	0.81323	6,247,857	Q00839
DSHGDPLQA	0.0157639	5,142,111	Q8LGN1
DSKPDTTAPPSSPK	0.999997	7,863,849	P27824
DSKPDTTAPPSSPK	0.996524	7,863,849	P27824
DTAVVEYLADK	0.0541983	6,703,263	Q2G6N0
DTEEHHLR	0.061353	3,754,980	Q2HJ60
DVDHQIAK	0.0653678	5,212,555	P48643
DVDHQIAK	0.0201733	5,212,555	P48643
DVLDKHLIPAANTGESK	0.997016	6,513,436	P62261
DVLDKHLIPAANTGESK	0.996329	6,513,436	P62261
DVLDKHLIPAANTGESK	0.991767	6,513,436	P62261
DVLDKHLIPAANTGESK	0.957822	6,513,436	P62261
EAAPEAPAATGAA	0.0321257	6,077,716	Q8NNJ2
EAATKGYATATD	0.0162699	6,577,979	Q3JDS2

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
EADCDILIPAASEK	0.0211416	8,243,841	P00367
EAGKDDYVK	0.850526	5,847,816	P09211
EAGKDDYVK	0.670542	5,847,816	P09211
EAHFHGCGSVNR	0.939906	4,868,717	Q9DDY9
EAIITPPPAK	0.0655951	5,768,205	Q9QZQ8
EALRVVDASIMPR	0.0164283	7,808,975	B2TCJ8
EAMNYEGSPIK	0.53225	6,778,046	P06748
EAQQAFQSVGTK	0.127733	7,053,401	P07814
EASGGGAFLVLPLGK	0.999766	7,664,133	Q92176
EASGGGAFLVLPLGK	0.999766	7,664,133	Q92176
EASGGGAFLVLPLGK	0.999766	7,664,133	Q92176
EAVQNMVSHTER	0.815755	7,448,323	Q12906
EDENFILK	0.143279	5,622,708	P62938
EDVQPHDLGK	0.993592	6,272,951	P10809
EDVQPHDLGK	0.951092	6,272,951	P10809
EFAIQPNTTGK	0.81559	6,613,267	P31976
EFAIQPNTTGK	0.239588	6,613,267	P31976
EFGDGLVAVGEPDALAK	0.230521	9,024,455	Q7UH36
EFQMPYIANIP	0.262471	7,138,229	Q1ACL5
EHGIQPDGQMPSDK	0.999999	5,522,450	P68362
EHGIQPDGQMPSDK	0.999999	5,522,450	P68362
EHGPSIVHR	0.0520683	5,602,721	Q8SWN8
EITIPVTFESR	0.243234	6,903,476	P04792
EITIPVTFESR	0.0787031	6,903,476	P04792
EITPYQQGSLK	0.0362822	6,903,476	P52279
EIWHLIGFK	0.052183	6,298,365	A3PCW8
EMPFIASMGIYVI	0.0583195	7,798,715	P52416
EMPLPPDMAAATYAK	0.045878	8,693,986	P40795
EPSANMPWFK	0.998551	6,617,992	P68103
EPSANMPWFK	0.996115	6,617,992	P68103
EPSANMPWFK	0.994079	6,617,992	P68103
EPSANMPWFK	0.992994	6,617,992	P68103
EPSANMPWFK	0.829181	6,617,992	P68103
EPVAIVF	0.0346585	4,317,226	B1KDP2
EPVVTASNGEVK	0.0172537	4,492,272	Q4KHN0
EQMPSKEDAIEHFMK	0.999696	6,553,037	P09874
ESCGVTSDNCR	0.732572	6,867,501	P13489
EVVTVETWQEGSLK	0.999539	8,609,269	P09211
EVVTVETWQEGSLK	0.997466	8,609,269	P09211
FAAEVVHPGDLK	0.999998	6,998,581	P54577

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FAAEVVHPGDLK	0.999998	6,998,581	P54577
FAAEVVHPGDLK	0.998627	6,998,581	P54577
FAAEVVHPGDLK	0.975254	6,998,581	P54577
FAASPSELHLK	0.0247613	6,583,396	O00442
FADKVPK	0.0372683	4,747,650	P62938
FADKVPK	0.0162045	4,747,650	P62938
FAGYTDNLVR	0.11232	6,222,926	P63243
FAIQPNTTGK	0.992125	5,968,054	P31976
FAIQPNTTGK	0.966871	5,968,054	P31976
FAIQPNTTGK	0.110871	5,968,054	P31976
FALPHAILR	0.950328	3,758,795	P83969
FALPHAILR	0.924378	3,758,795	P83969
FAPWCGHCK	0.895391	6,397,647	P38657
FAPWCGHCK	0.81402	6,397,647	P38657
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	8,869,536	Q01518
FAQINQGESITHALK	0.999999	8,869,538	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	8,869,538	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	5,916,383	Q01518
FAQINQGESITHALK	0.999999	8,869,538	Q01518
FAQINQGESITHALK	0.999998	5,916,383	Q01518
FAQINQGESITHALK	0.999996	5,916,382	Q01518
FAQINQGESITHALK	0.999995	5,916,383	Q01518
FAQINQGESITHALK	0.999988	5,916,383	Q01518
FAQINQGESITHALK	0.999975	5,916,383	Q01518
FAQINQGESITHALK	0.999898	5,916,383	Q01518

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FAQINQGESITHALK	0.99988	8,869,536	Q01518
FAVTVAPPGAR	0.998999	5,873,081	Q00839
FAVTVAPPGAR	0.998768	5,873,081	Q00839
FAVTVAPPGAR	0.998537	5,873,081	Q00839
FAVTVAPPGAR	0.973369	5,873,081	Q00839
FCSEYRPK	0.878524	6,017,705	P10103
FDDSQDKAVLK	0.448804	7,053,529	P00918
FDKDGDGTITTK	0.999761	7,213,478	P62147
FDKDGDGTITTK	0.999761	7,213,478	P62147
FDKDGDGTITTK	0.999523	7,213,478	P62147
FDKDGDGTITTK	0.99738	7,213,478	P62147
FDKDGDGTITTK	0.232316	7,213,478	P62147
FEALGIAQPK	0.488332	5,953,181	P62194
FEDTNLCAIHAK	0.999364	7,678,553	P68432
FEDTNLCAIHAK	0.999109	7,678,553	P68432
FEDTNLCAIHAK	0.998727	7,678,553	P68432
FEDTNLCAIHAK	0.9986	7,678,553	P68432
FEDTNLCAIHAK	0.998219	7,678,553	P68432
FEDTNLCAIHAK	0.889747	7,678,553	P68432
FEDTNLCAIHAK	0.615121	7,678,553	P68432
FEDTNLCAIHAK	0.286529	7,678,553	P68432
FEDTNLCAIHAKR	0.99204	5,642,730	P68432
FEDTNLCAIHAKR	0.856831	5,642,730	P68432
FEGFPDKQPR	0.98026	6,688,216	P35466
FEGFPDKQPR	0.973988	6,688,216	P35466
FEGFPDKQPR	0.901376	6,688,216	P35466
FEGFPDKQPR	0.189135	6,688,216	P35466
FEVTHDITK	0.741316	6,032,974	Q2I6W4
FEVTHDITK	0.219768	6,032,974	Q2I6W4
FFKPEYGRPNVSAANPR	0.848687	6,893,475	P26640
FFKPEYGRPNVSAANPR	0.752532	6,893,475	P26640
FFLVIDLEGK	0.060119	6,488,491	Q8W566
FGGGTTPDAK	0.975531	5,337,475	Q9Y4L1
FGNETWGVTK	0.723987	6,277,951	Q4VIT5
FGPDICGPGTK	0.938175	6,327,889	Q4VIT5
FGPDICGPGTK	0.0690407	6,327,889	Q4VIT5
FGSVADDQK	0.451632	5,417,450	Q3MHL3
FGTLTDCVVVVNPQTK	0.999997	9,474,761	Q13151
FGVPGAFTPGCSK	0.787776	7,208,363	Q9GLW7
FHAGGAGAADP	0.0824116	5,297,216	Q0UE25

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FHAGGAGAADP	0.014796	5,297,216	Q0UE25
FHCIDPNDSK	0.831983	6,747,866	Q3ZBF7
FHCIDPNDSK	0.468109	6,747,869	Q3ZBF7
FHHEGGVDVGDVDAK	0.999998	5,665,912	Q32PF2
FHHEGGVDVGDVDAK	0.999998	5,665,912	Q32PF2
FHHEGGVDVGDVDAK	0.998122	5,665,912	Q32PF2
FHMDPSGTFVQCDAR	0.985484	9,283,819	Q5E987
FIGNLPHEVDKSELK	0.997538	6,239,976	Q32LC7
FIGNLPHEVDKSELK	0.862075	6,239,976	Q32LC7
FINGPNNAHIQQVGDR	0.999997	6,233,005	P49951
FINGPNNAHIQQVGDR	0.989416	6,233,005	P49951
FITGLGC	0.108362	4,281,906	A7ZED8
FITGLGC	0.0330579	4,281,906	A7ZED8
FIYGHYK	0.0710089	5,222,548	P07108
FLLTEPPLNTPENR	0.995888	8,649,351	P61157
FLLTEPPLNTPENR	0.934147	8,649,351	P61157
FLLTEPPLNTPENR	0.927285	8,649,351	P61157
FLNVNSQVTTVCQALAK	0.999997	6,703,448	P50897
FLSKPVDVSK	0.0402344	6,323,547	Q9Y617
FMELYGVVQSTVNL	0.0378828	5,632,740	Q96725
FMVDNEAIYDICR	0.999998	8,673,704	P68362
FMVDNEAIYDICR	0.999997	8,673,704	P68362
FMVDNEAIYDICR	0.998312	8,753,681	P68362
FNAVGDGIVLCK	0.998954	7,048,521	P13796
FNAVGDGIVLCK	0.988914	7,048,521	P13796
FNAVGDGIVLCK	0.936602	7,048,521	P13796
FNAVGDGIVLCK	0.753155	7,048,521	P13796
FNEFPVPEQFK	0.999678	7,493,581	Q01518
FNSIATQGELVR	0.248946	7,118,561	P23381
FNSPLAHKI	0.0173491	5,718,052	Q43092
FPCVQGGTTAIPGAFGCGK	0.885572	6,809,828	P31404
FPMFVSQSALEK	0.999707	7,503,675	P07814
FPNFEQGVACLR	0.952213	7,633,502	P97275
FQIEGYNPHPTIK	0.0826207	8,304,136	P04818
FQNHDAADK	0.99522	5,812,535	Q13151
FQNISPEELK	0.990904	6,608,290	Q56JZ9
FQNISPEELK	0.979856	6,608,290	Q56JZ9
FQPSFIGME	0.580301	5,722,463	P30161
FQPSFIGME	0.299734	5,722,463	P30161
FQPSFIGME	0.065752	5,722,461	P30161

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FQTNLVPYPR	0.999978	6,618,319	P68362
FQTNLVPYPR	0.999978	6,618,319	P68362
FQTNLVPYPR	0.99953	6,618,319	P68362
FQTNLVPYPR	0.999463	6,618,319	P68362
FQTNLVPYPR	0.998858	6,618,319	P68362
FQTNLVPYPR	0.996857	6,618,319	P68362
FRDGDILGK	0.155225	5,687,923	P61603
FRDGDILGK	0.0710013	5,687,923	P61603
FRDGDILGKYVD	0.399347	7,573,716	P61603
FSELAEDKENYKK	0.0477074	5,916,308	Q76LV1
FSGDPNWFPKK	0.374618	7,338,607	P37802
FSGYVQQVK	0.982599	5,862,946	P07954
FSGYVQQVK	0.954945	5,862,946	P07954
FSITYKPVR	0.16711	6,138,341	Q9FY66
FSLPPPLQLV	0.0241038	5,998,306	P14238
FSNHGEIK	0.6023	5,242,502	P10881
FSNHGEIK	0.359155	5,242,502	P10881
FSSAACPDDPR	0.0392487	6,557,608	Q1D568
FSSPEEATK	0.0353826	5,562,526	Q98SP8
FTELAEDKENYKK	0.998788	5,963,027	Q76LV2
FTGDPAKNK	0.0470018	5,612,868	Q9SKN2
FTKILSHDVFAKN	0.0222408	5,553,009	G5EFV8
FVDPWTVQTSSAK	0.883375	7,913,847	P23381
FVEWIPNNVK	0.999925	6,813,499	Q9YHC3
FVEWIPNNVK	0.99985	6,813,499	Q9YHC3
FVEWIPNNVK	0.999839	6,813,499	Q9YHC3
FVEWIPNNVK	0.999754	6,813,499	Q9YHC3
FVEWIPNNVK	0.999743	6,813,499	Q9YHC3
FVEWIPNNVK	0.999625	6,813,499	Q9YHC3
FVEWIPNNVK	0.99882	6,813,499	Q9YHC3
FVEWIPNNVK	0.99811	6,813,501	Q9YHC3
FVEWIPNNVK	0.967485	6,813,499	Q9YHC3
FVGNLNFNK	0.100903	5,847,948	P19338
FVGNLNFNK	0.0687757	5,847,948	P19338
FVGNLNFNK	0.0642456	5,847,948	P19338
FVGSGVSGGEEGAR	0.471719	6,988,121	P38720
FVGSGVSGGEEGAR	0.0967223	6,988,121	P38720
FVITKPDVYK	0.997293	6,773,782	E9PAV3
FVITKPDVYK	0.98657	6,773,782	E9PAV3
FVPYFEAIQK	0.985056	6,793,469	P49588

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FVPYFEAIQK	0.939368	6,793,469	P49588
FYAGGQVYPGEASR	0.789504	7,953,565	Q13162
FYGPPGCGK	0.961501	5,497,412	Q3ZBT1
FYRDPEEIEKEEQAAAEK	0.999997	7,760,352	P26452
FYRDPEEIEKEEQAAAEK	0.999997	7,760,352	P26452
FYRDPEEIEKEEQAAAEK	0.999733	7,760,352	P26452
GAGLMGAGIAQVSVDK	0.999998	7,954,051	P40939
GAGLMGAGIAQVSVDK	0.999998	7,954,051	P40939
GAGLMGAGIAQVSVDK	0.999998	7,954,051	P40939
GAGLMGAGIAQVSVDK	0.916148	7,954,051	P40939
GASLSAPLTSYK	0.269659	6,558,366	Q9P2J5
GCHAYLSK	0.999879	5,262,388	Q01518
GDPGLPGEPGL	0.0218506	5,487,526	Q8NFW1
GEALQGEDGFGIDPAILDR	0.0218506	6,876,582	Q87ME0
GEMSLHGLLDHAIILAH	0.024531	4,794,947	P24800
GETYKDHENIVIAK	0.999817	5,876,346	P05307
GETYKDHENIVIAK	0.596682	5,876,346	P05307
GGFEITPPVVLR	0.633785	6,868,685	P06748
GGFEITPPVVLR	0.207926	6,868,685	P06748
GGPVGSALPASS	0.0371159	5,442,583	Q82FD8
GGQYGVQGFPTIK	0.997576	7,343,686	Q15084
GGQYGVQGFPTIK	0.991954	7,343,686	Q15084
GGSPESLLWNT	0.0450838	6,247,821	B3EWZ5
GGSSWPGYVRPLPPAAIESPAVAAPAYSR	0.914847	10,048,395	P04792
GHDVASIN	0.0170311	4,506,976	A3D039
GHFGVNESTGLSLEQVKK	0.998297	6,920,262	O46674
GHVQPQIR	0.0437901	5,117,638	Q1QVN7
GICPHMTEDNKDLLQGK	0.999622	7,006,691	P38657
GICPHMTEDNKDLLQGK	0.957607	7,006,691	P38657
GIECGGAYVK	0.97379	5,852,703	P27824
GIPAGKITAQGQGESNPVTGNTCDN	0.0483502	8,680,675	I2BAK7
GITLPVDFQGR	0.35725	6,458,294	P31943
GITLPVDFQGR	0.0820146	6,458,296	P31943
GITLPVDPEGK	0.177655	6,213,261	Q5E9J1
GITLPVDPEGK	0.02919	6,213,261	Q5E9J1
GIVPIVEPEILPDGDHDLK	0.999999	7,247,135	P04075
GIVPIVEPEILPDGDHDLK	0.999995	7,247,135	P04075
GIVPIVEPEILPDGDHDLK	0.999994	7,247,135	P04075
GIVPIVEPEILPDGDHDLK	0.999113	10,865,667	P04075
GIVPIVEPEILPDGDHDLKR	0.999999	7,767,472	P04075

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
GIVPIVEPEILPDGDHDLKR	0.999999	7,767,472	P04075
GIVPIVEPEILPDGDHDLKR	0.999999	7,767,472	P04075
GIVPIVEPEILPDGDHDLKR	0.999999	7,767,472	P04075
GIVPIVEPEILPDGDHDLKR	0.999999	7,767,472	P04075
GLEFTSSGSANTETTK	0.997852	8,733,986	P45879
GLLGCNIVLPPEAM	0.179304	7,863,851	Q9VFS8
GLMTTVHAITATQK	0.969349	7,944,156	P10096
GLMTTVHAITATQK	0.437762	7,944,156	P10096
GLTLYDLGNIMNGDIQK	0.0726229	6,663,289	A5FHI8
GMESCGIHETTFNSIMK	0.999999	6,866,316	P60712
GMESCGIHETTFNSIMK	0.999995	6,866,316	P60712
GNPLGVKA	0.0163649	4,362,391	P20721
GPATVLAIGTATPPNC	0.380073	8,143,948	P48388
GPIATADLVVAHDG	0.0217207	7,123,481	P83781
GPMGPLGPPGPKGN	0.17867	6,963,441	P08125
GPMGPLGPPGPKGNSGEPGLPGPPGPPGPP	0.0176723	6,963,478	P08125
GPPASTSRRSH	0.139518	6,207,964	Q6TCG2
GPPGPTGAPGPQGFQGNPGEPGEP	0.10685	7,630,078	Q91717
GSNMTEFHSQISK	0.987156	7,913,556	Q14204
GTHETAFLGPK	0.999997	6,373,161	Q9XSK7
GTHETAFLGPK	0.999133	6,373,161	Q9XSK7
GTHETAFLGPK	0.473453	6,373,161	Q9XSK7
GTLPAPVGSPEDPSDPPQPYR	0.0213638	7,556,882	Q9C0A1
GTNAAEFQTK	0.998297	5,917,768	Q13435
GTTLKDEGK	0.501084	5,467,841	Q2KJH4
GTTLKDEGK	0.157705	5,467,841	Q2KJH4
GTTLKDEGK	0.140472	5,467,841	Q2KJH4
GTTLKDEGK	0.0492907	5,467,841	Q2KJH4
GTVPTSIYSMSFNLSSTL	0.0486428	6,703,130	Q96U88
GVDTLVVTNAAGGLNPK	0.382662	5,813,102	P00491
GVKPNATQEELKK	0.992709	5,386,360	Q95JF4
GVKPNATQEELKK	0.971308	5,386,360	Q95JF4
GVQEPHPGER	0.422809	5,972,722	Q6XQN6
GVSHPVLK	0.0795864	4,767,681	P23004
GVTLMSTQPTV	0.0603882	6,112,965	B2VET5
GWGVMVSHR	0.546169	5,587,577	P06733
GWGVMVSHR	0.195328	5,587,577	P06733
HAAMADTFLEHMCR	0.999997	5,932,475	P00548
HAAMADTFLEHMCR	0.999997	5,932,475	P00548
HAAVGASR	0.563584	4,287,085	Q2KJE4

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
HAAVGASR	0.0420074	4,287,085	Q2KJE4
HACPGVGSCGG	0.0782743	5,737,101	B8G762
HADFAGVER	0.711111	5,452,429	P78371
HAIVVSDRDGVPVIK	0.978502	5,743,188	Q17QQ1
HALPSPL	0.217792	4,117,126	Q5W150
HAPGDYNPQVKPSSKGNYHSVSITI	0.0544608	7,108,565	A8GB11
HAQVADMK	0.998948	5,082,388	P35579
HAQVADMK	0.996673	5,082,388	P35579
HAQVADMK	0.951833	5,162,363	P35579
HAQVADMKK	0.999998	3,912,037	P35579
HAQVADMKK	0.999998	3,912,037	P35579
HAQVADMKK	0.9649	3,912,037	P35579
HCAVETAK	0.973241	5,162,363	Q9UBX3
HCAVETAK	0.295577	5,162,363	Q9UBX3
HDQTPGLPK	0.0233175	5,547,766	Q8CDU5
HENIVIAK	0.207271	5,202,841	P05307
HENIVIAK	0.0717919	5,202,841	P05307
HWGSLDGQGSEHTVDKK	0.999886	5,069,929	P00918
HWGSLDGQGSEHTVDKK	0.998652	5,069,929	P00918
HWIRDKGPVVET	0.0540859	7,769,009	A7RZW4
IAALVVDNGSGMCK	0.999296	7,758,726	P60712
IAALVVDNGSGMCK	0.97816	7,758,726	P60712
IAALVVDNGSGMCK	0.977422	7,758,726	P60712
IAELQDCEGLIVR	0.111226	8,023,946	O43175
IAGHPAFVNYSTSQK	0.991861	8,684,276	P14866
IAVPDVYSACK	0.867001	6,698,254	Q04760
IDNKVMGFGADIKSK	0.0788965	6,043,197	B8CY73
IDRFEPSSGPV	0.0299917	4,312,045	Q626H5
IEAVTGEGAITTVHADSD	0.0231213	6,252,878	A7ZQC5
IEENKYSR	0.159172	5,777,794	Q00839
IFWAPESAPLK	0.9997	6,878,601	P23528
IFWAPESAPLK	0.99955	6,878,601	P23528
IFWAPESAPLK	0.999251	6,878,601	P23528
IFWAPESAPLK	0.989714	6,878,601	P23528
IGDYNGHVGLGVK	0.994225	7,228,666	O18789
IGGLAMACHDSFLK	0.785033	8,183,886	P31949
IGGLAMACHDSFLK	0.285882	8,183,886	P31949
IGHEGPGSLLSELK	0.0917223	7,769,058	Q24K02
IGLGYTQTLKPGIK	0.9975	8,169,736	P45879
IGLGYTQTLKPGIK	0.977956	8,169,736	P45879

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
IGSLTFEAPDLEAFPCLGLAF	0.228259	7,860,463	A3DE52
IGSSHILTPTR	0.997238	6,353,348	Q28288
IGSSHILTPTR	0.199399	6,353,348	Q28288
IHAVPVEER	0.899147	5,692,899	Q1KMD3
IHAVPVEER	0.604143	5,692,899	Q1KMD3
IHAVPVEER	0.0772646	5,692,899	Q1KMD3
IHFPLVTYAPVISAEK	0.260699	6,343,463	P06603
IHFPLVTYAPVISAEK	0.212742	6,343,463	P06603
IHFPLVTYAPVISAEK	0.149549	6,343,463	P06603
IHFPLVTYAPVISAEK	0.144722	6,343,463	P06603
IINSYGTLSGEDFLR	0.071976	8,869,300	O42617
ILTLTSTDK	0.0219481	5,543,021	Q9L7Q2
INALYPEGQAPVK	0.998312	7,583,976	P37802
INALYPEGQAPVK	0.173314	7,583,976	P37802
INALYPEGQAPVKK	0.999998	8,364,608	P37802
INALYPEGQAPVKK	0.999994	8,364,608	P37802
INALYPEGQAPVKK	0.999989	8,364,608	P37802
INALYPEGQAPVKK	0.999942	5,579,763	P37802
INSQWVVSAAHCYK	0.999998	8,899,215	P00760
INSQWVVSAAHCYK	0.999998	8,899,215	P00760
INSQWVVSAAHCYK	0.999842	8,899,215	P00760
IPNGHSPLK	0.694569	5,397,896	Q54T81
IPTEGGDFNEFPVPEQFK	0.999858	7,230,067	Q01518
IPTEGGDFNEFPVPEQFK	0.999844	7,230,067	Q01518
IPTEGGDFNEFPVPEQFK	0.999763	7,230,067	Q01518
IPTEGGDFNEFPVPEQFK	0.999395	7,230,067	Q01518
IQAADTPEIQQLQADYAK	0.0341772	7,070,175	P43786
IQNVPLEDR	0.735922	5,862,926	Q00839
IQNVPLEDR	0.577031	5,862,926	Q00839
IQNVPLEDR	0.513062	5,862,926	Q00839
IQQATTVK	0.0775918	5,027,761	Q3SYR7
ISTIGVDFK	0.456297	5,482,916	P62822
ITEHGDLGNSR	0.978681	6,437,935	A4FUA8
ITEHGDLGNSR	0.915166	6,437,935	A4FUA8
ITFKEEEPVKK	0.758885	5,072,863	P17132
ITFKEEEPVKK	0.237268	5,072,863	P17132
ITGKEDAANNYAR	0.999999	7,698,672	P68362
ITGKEDAANNYAR	0.999943	7,698,672	P68362
ITKNIHLPV	0.051886	5,758,365	A6LJ47
ITSAADSEAITFQK	0.0462546	7,993,926	P17655

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
IVAAGVGEFEAGISK	0.999984	7,824,082	P68103
IVAAGVGEFEAGISK	0.999951	7,824,082	P68103
IVAAGVGEFEAGISK	0.999951	7,824,082	P68103
IVAAGVGEFEAGISK	0.999935	7,824,082	P68103
IVAAGVGEFEAGISK	0.999935	7,824,082	P68103
IVAAGVGEFEAGISK	0.99987	7,824,082	P68103
IVAAGVGEFEAGISK	0.999511	7,824,082	P68103
IVGFDFDGTPR	0.792725	6,563,056	Q9PVY6
IVLAKPPDKK	0.023161	4,276,053	O43390
IVLAKPPDKK	0.0204709	4,276,053	O43390
IVRPDNTYEVK	0.993021	7,253,742	Q4VIT5
IVRPDNTYEVK	0.99051	7,253,742	Q4VIT5
IVRPDNTYEVK	0.983456	7,253,742	Q4VIT5
IVRPDNTYEVK	0.869503	7,253,742	Q4VIT5
IVSVDETIKNPR	0.956052	7,439,005	Q2NKZ1
IVSVDETIKNPR	0.817449	7,439,005	Q2NKZ1
IYIPLPDEK	0.0949551	6,023,203	Q3ZBT1
KDECPEVR	0.112048	5,747,576	Q32PI5
KEEQTPQNK	0.275282	6,233,110	Q9BE24
KFPAIAGHPK	0.053002	6,053,445	Q75JD5
KHEDLLTMRNGSQVPLINGCDNM	0.0448677	9,254,293	Q72T26
KINYNSPLPE	0.0176676	6,458,236	Q5AD72
KIWHHTF	0.999864	3,621,849	Q92193
KIWHHTF	0.931425	3,621,849	Q92193
KIWHHTF	0.225392	3,621,849	Q92193
KNAYGLTVPL	0.0381273	5,963,259	Q7VR52
KPLRLPLQDVYK	0.992511	5,386,532	P68103
KPLRLPLQDVYK	0.908318	5,386,532	P68103
KPVAPNGSTTT	0.0385122	5,948,001	A6R7S1
KPVYFDGSVHW	0.398458	3,634,262	Q9LM74
KPVYFDGSVHW	0.0510272	3,634,262	Q9LM74
KQEVISTSSK	0.985866	6,258,369	P63243
KQEVISTSSK	0.912516	6,258,369	P63243
KQEVISTSSK	0.0959974	6,258,369	P63243
KSADMAEAL	0.0259828	5,262,436	A8LMD0
KSLLGDHMDVVAHVVYVVKD	0.0283794	7,954,168	P08941
KVLKQVHPDTGISSK	0.999999	6,036,817	P33778
KVLKQVHPDTGISSK	0.999999	4,530,131	P33778
KVLKQVHPDTGISSK	0.999999	6,036,817	P33778
KVLKQVHPDTGISSK	0.999999	4,530,131	P33778

Supplementary material

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
KVLKQVHPDTGISSK	0.999999	4,530,131	P33778
KVLKQVHPDTGISSK	0.999999	4,530,131	P33778
KVLKQVHPDTGISSK	0.999999	4,530,131	P33778
KVLKQVHPDTGISSK	0.999998	6,036,817	P33778
KVLKQVHPDTGISSK	0.999993	6,036,817	P33778
KVLKQVHPDTGISSK	0.999986	4,530,131	P33778
KVTLDWAKPK	0.233371	4,532,687	P19338
LAAGSCDFK	0.849376	5,427,439	Q1JP79
LAAGSCDFK	0.778698	5,427,439	Q1JP79
LAAGSCDFK	0.422711	5,427,439	Q1JP79
LADHANSPNKK	0.323671	4,468,995	P40939
LADLYVNDAFGTAHR	0.993375	5,842,789	Q9LD57
LADLYVNDAFGTAHR	0.955544	5,842,789	Q9LD57
LADLYVNDAFGTAHR	0.94184	5,842,789	Q9LD57
LADNVICPGAPDFLAHVR	0.999997	6,850,048	P21964
LAEVAGTGSPVIASGGIR	0.129089	8,719,591	A7IAG4
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,206	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,206	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,206	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,206	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAGPVAEYLK	0.999999	5,888,205	Q01518
LAHEIGFGSK	0.657279	5,878,001	Q5E946
LAPGGALLLLGPN	0.0322618	6,473,656	Q2RYF0
LASAVPAFQAAGVTLR	0.0273777	5,539,679	Q7U654
LASYAVQSK	0.0281018	5,417,814	Q2HJ49
LCKPEPELNAAIPSANPAK	0.999318	7,220,430	Q8WUM4
LDKPQLQGIPVLVLGNKR	0.802746	7,114,258	Q6NZW8
LDRLIEEGFDPYPP	0.0201999	8,749,138	D0LB45
LDSVQVLSADETAPL	0.040305	8,234,033	B0BR94
LEGTLLKPNMVTPGHACTQK	0.999999	7,804,022	P04075
LEGTLLKPNMVTPGHACTQK	0.999999	7,804,022	P04075
LEGTLLKPNMVTPGHACTQK	0.999999	7,804,022	P04075
LEGTLLKPNMVTPGHACTQK	0.998634	7,804,022	P04075
LGALPGAGGTQR	0.0373825	5,933,061	P40939
LGDVAPNFEANTTVGR	0.986926	8,749,174	P30041
LGERPVQHR	0.770056	3,938,734	P24452
LGETYKDHENIVIAK	0.999999	6,253,292	P05307
LGETYKDHENIVIAK	0.99925	6,253,292	P05307
LGETYKDHENIVIAK	0.999101	6,253,292	P05307
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,262	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.999999	7,296,578	P60712
LGMESCGIHETTFNSIMK	0.999999	7,296,579	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.999999	7,243,263	P60712
LGMESCGIHETTFNSIMK	0.998036	7,296,578	P60712
LGSYAVQAK	0.861365	5,267,761	P31976

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
LGSYAVQAK	0.64416	5,267,761	P31976
LIANATNPESK	0.0318536	6,373,267	Q3SZI4
LIANATNPESK	0.0304579	6,373,266	Q3SZI4
LIANATNPESK	0.0224094	6,373,267	Q3SZI4
LIDGGGEAEMLAR	0.0331149	7,183,316	Q4K6S8
LIGFNVAIGGGMGMTH	0.0335255	5,549,317	A7Z8R5
LIPEGVNGIEER	0.0216233	7,073,560	Q14194
LLTIHPALTDTDPR	0.0657972	8,259,296	B0W730
LMQILPV	0.0201999	4,592,456	O17611
LNAPLSQMASND	0.96151	6,747,974	Q7Z5Y7
LPLLQQCQVVR	0.975618	7,213,865	P13489
LPLLQQCQVVR	0.963815	7,213,865	P13489
LQNNLPAVR	0.83058	5,567,979	Q54TD3
LQVPAEYVLGK	0.139989	6,668,656	C5BQ39
LSAENADDIYDELSEILARP	0.0396354	7,746,985	Q8FMX2
LSATMPSDVLEVTKK	0.140514	8,819,701	Q3SZ54
LSDGEGPPGGR	0.156715	3,771,697	Q7L0J3
LSDGEGPPGGR	0.0842171	3,771,697	Q7L0J3
LSDGEGPPGGR	0.0735459	3,771,697	Q7L0J3
LSDGEGPPGGR	0.0515634	3,771,697	Q7L0J3
LSGAGEHLK	0.229974	5,142,659	P35579
LSGKPDDRMIEDYAD	0.0538238	6,142,747	B7UI88
LSIGSEDHKE	0.0352445	4,108,607	A5DVH6
LSQNQGGK	0.420755	4,742,346	P09211
LSQNQGGK	0.0789766	4,742,346	P09211
LSYDPPIVPPVVVK	0.0441129	5,469,775	Q57559
LTEAPLNPK	0.992725	5,497,970	P60712
LTEAPLNPK	0.990924	5,497,970	P60712
LTEAPLNPK	0.981078	5,497,970	P60712
LTEPPLNTPENR	0.995697	7,348,589	P61157
LTPVEEAPK	0.189826	5,502,890	P52566
LTPVEEAPK	0.0724963	5,502,890	P52566
LVIGGGSGGLASAR	0.0227295	6,518,456	A2TIL1
LYNLGDQYALK	0.985279	7,073,581	P04843
LYNNIVLSGGTTMFPGIADR	0.962472	7,483,661	P53498
MAASSSSLEK	0.685142	5,637,596	P0CG38
MADGTCQDAAIVGYK	0.999998	8,583,756	P07737
MADGTCQDAAIVGYK	0.942287	8,583,757	P07737
MAHAVTQLANR	0.989173	6,503,187	P78371
MAHAVTQLANR	0.988905	6,503,187	P78371

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
MAHNEPITIDD	0.336711	6,722,841	P0CS37
MAPKPGPYVK	0.999971	6,163,327	Q01518
MAPKPGPYVK	0.962468	6,163,327	Q01518
MAQITMSDMLK	0.433992	7,168,131	A1A1H9
MASIPLPDK	0.496409	5,442,801	P00491
MATAASSSSLEK	0.999999	6,498,022	P60712
MATAASSSSLEK	0.999999	6,498,022	P60712
MATAASSSSLEK	0.999999	6,498,022	P60712
MATAASSSSLEK	0.999999	6,498,022	P60712
MATAASSSSLEK	0.999999	6,498,022	P60712
MATAASSSSLEK	0.999934	6,498,022	P60712
MATAASSSSLEK	0.999881	6,577,996	P60712
MATGDLDQDGR	0.919913	6,337,583	P13796
MATGDLDQDGR	0.772428	6,337,583	P13796
MATGDLDQDGR	0.455736	6,337,583	P13796
MEAPTTAYKK	0.447968	6,423,226	P14317
MFEDPPPGVV	0.0236775	5,962,569	Q6C0B5
MFSEWHVPL	0.0250248	6,172,754	O83388
MIAPMIEKFAEQYSDAA	0.0369419	6,829,792	P22803
MILPVGAANFR	0.359498	6,388,309	P06733
MILPVGAANFR	0.256436	6,388,309	P06733
MILPVGAANFR	0.230296	6,388,309	P06733
MILPVGAANFR	0.0565973	6,388,309	P06733
MKDPSVVVR	0.0347805	5,738,043	Q14974
MLEPSANMPWFK	0.999576	7,838,615	P68103
MLEPSANMPWFK	0.996814	7,838,615	P68103
MLEPSANMPWFK	0.946193	7,838,615	P68103
MLEPSANMPWFK	0.272643	7,838,615	P68103
MNHPDLAK	0.0863587	5,212,466	Q4QXM3
MNWECAIPGKK	0.0294866	7,393,538	Q9W6H5
MRVVGHASNK	0.0341428	4,055,428	P95318
MSLPMLQVALDNQTMASAY	0.0201677	7,299,981	B7LLX8
MTFSELPYNNLR	0.0660729	7,948,606	P46892
MTSLHTK	0.898309	4,672,305	Q06323
MTSLHTK	0.600737	4,672,305	Q06323
MTTVHAITATQK	0.999999	7,093,627	P10096
MTTVHAITATQK	0.999999	7,093,627	P10096
MTTVHAITATQK	0.999936	4,732,442	P10096
MTYIPTTAAEQQQML	0.0700056	6,106,089	Q2RH48
MVDNWRPAQPLK	0.999999	5,242,673	P00918

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
MVDNWRPAQPLK	0.999999	7,858,973	P00918
MVDNWRPAQPLK	0.999999	7,858,973	P00918
MVDNWRPAQPLK	0.999999	7,858,973	P00918
MVDNWRPAQPLK	0.999987	7,858,973	P00918
MVDNWRPAQPLK	0.999977	5,295,989	P00918
MVDNWRPAQPLK	0.999969	7,858,973	P00918
MVDNWRPAQPLK	0.999963	5,242,673	P00918
MVDNWRPAQPLK	0.999847	7,938,946	P00918
MVQEHDEPILK	0.999997	7,278,547	A6H767
MVQEHDEPILK	0.999695	7,278,547	A6H767
NAAESGVQHKPSAPQGGR	0.422727	6,363,112	Q5ZKU5
NAFLLPIK	0.324219	5,163,016	P49915
NAIHNHLVK	0.0505169	3,878,782	Q6BZQ6
NAPVEFSEAR	0.035072	4,031,855	P0CD72
NASGSTSTPAPSR	0.943744	6,607,961	Q32PF2
NAVDALIDSMSLAK	0.99521	7,823,917	P13010
NCEVTNLNDYR	0.995241	7,433,087	P51122
NCLPIAAIVDEK	0.997054	7,298,706	Q3T0E7
NCLPIAAIVDEK	0.984082	7,298,706	Q3T0E7
NCLPIAAIVDEK	0.957855	7,298,706	Q3T0E7
NDEFGLEIPK	0.063058	6,393,081	P38222
NEASVLHNLK	0.676024	6,208,216	P35579
NHDGNPDR	0.262473	5,066,989	Q55242
NHDGNPDR	0.0381524	5,066,989	Q55242
NILENNSGIVDVP	0.167572	7,363,587	C3MQL0
NIQGITKPAIR	0.899054	6,638,821	Q757K0
NLALNSASAIGCHVVNIGAEDLK	0.999757	8,280,859	P13796
NLEDVQPHDLGK	0.999086	7,408,586	P10809
NLEDVQPHDLGK	0.997717	7,408,586	P10809
NLEEDKELLGTTGDSEMEELLKDEIST	0.0224361	8,003,757	B0S2B6
NLHLRLQLPL	0.0429674	6,528,792	Q0INW1
NLLAVYGAAVSLGKD	0.155905	8,039,291	A6LEU7
NLLEVQGPETVPL	0.275374	4,995,938	Q5XGW6
NMLNPPAEVTTK	0.960306	7,158,547	P13010
NPLSMNMSMPMP	0.0364557	4,798,618	Q652P9
NQLAPIVGVGMVK	0.0314121	7,213,991	P0CM47
NSFVNDIFER	0.951155	6,648,008	P33778
NSFVNDIFER	0.856188	6,648,008	P33778
NSFVNDIFER	0.553946	6,648,008	P33778
NSFVNDIFER	0.217212	6,648,008	P33778

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
NSSSASTTEPDFQK	0.988497	8,078,516	P42704
NVANPFGMILSLAMC	0.155794	8,713,926	Q2FF66
NVDLTEFQTNLVPYPR	0.999956	9,974,882	P68362
NVDLTEFQTNLVPYPR	0.999956	9,974,882	P68362
NVDLTEFQTNLVPYPR	0.999956	9,974,882	P68362
NVDLTEFQTNLVPYPR	0.277165	9,974,882	P68362
NVEFDDSQDK	0.995878	6,567,719	P00918
NVEFDDSQDK	0.980107	6,567,719	P00918
NVEFDDSQDK	0.569373	6,567,719	P00918
NVEFDDSQDKAVLK	0.997344	8,764,296	P00918
NVGAPAAGMNAAVR	0.996937	6,938,347	A1A4J1
NVIGPWIQTK	0.117738	6,363,447	Q3B7N2
NVITVGPR	0.0247323	4,722,553	P04040
NVLIPTEGGDFNEFPVPEQFK	0.999992	8,317,385	Q01518
NVLIPTEGGDFNEFPVPEQFK	0.999969	8,317,385	Q01518
NVLIPTEGGDFNEFPVPEQFK	0.999866	8,317,385	Q01518
NVSAVDKSTGK	0.172353	6,253,267	P08106
NVSAVDKSTGKENK	0.243389	5,502,908	P08106
NVSNCIQLK	0.243257	5,962,966	Q5PPY1
NVSNLVIEDTELK	0.999986	7,954,086	Q01518
NVSNLVIEDTELK	0.999983	7,954,086	Q01518
NVSNLVIEDTELK	0.999979	7,954,086	Q01518
NVSNLVIEDTELK	0.999976	7,954,086	Q01518
NVSNLVIEDTELK	0.999905	7,954,084	Q01518
NVSNLVIEDTELK	0.99986	7,954,086	Q01518
NVSNLVIEDTELK	0.999589	7,954,086	Q01518
NVTNPNSTEHMK	0.993621	7,443,347	Q14974
NWRPAQPLK	0.649446	6,133,293	P00918
NWRPAQPLK	0.491453	6,133,293	P00918
NYKPPPQK	0.999441	5,582,997	Q9TU03
NYKPPPQK	0.998696	5,582,997	Q9TU03
NYKPPPQK	0.998138	5,582,997	Q9TU03
NYKPPPQK	0.997952	5,582,997	Q9TU03
PDAVLAAEVSR	0.0252557	6,083,057	P18775
PELDHQRGSSGPGVR	0.0593899	8,403,996	Q8K0Y7
PGVIGDPQSHR	0.222048	6,258,011	Q72JU8
PLREPVVTLEGHTK	0.999997	5,646,433	Q92176
PLREPVVTLEGHTK	0.713786	5,646,433	Q92176
PNSIVEGGLFK	0.0189461	4,262,252	Q1MRT4
QAGFNTSR	0.0152275	4,847,166	Q9W2F2

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
QAGQCGNQIGAK	0.970073	6,743,111	Q9YHC3
QANCYEEVKDR	0.999202	7,643,322	P23528
QANCYEEVKDR	0.998803	7,643,322	P23528
QAVILLLMLF	0.0236448	4,222,398	P31666
QAVPLADSFR	0.577809	5,962,952	A1YES6
QAVPLADSFR	0.335954	5,962,952	A1YES6
QEESQKWGFDFELETP	0.0213314	6,959,753	Q22197
QEIVFPILSPD	0.0486794	6,733,391	Q19V68
QFLGGLGINLGGPA	0.015606	7,013,636	P9WHU4
QFVSDVEIITM	0.681718	6,853,227	Q5LII9
QGIVSWGSGCAQK	0.0552836	7,473,476	P00760
QGLQTPSCK	0.998208	5,677,679	P13489
QGLQTPSCK	0.98721	5,677,679	P13489
QGLQTPSCK	0.979721	5,677,679	P13489
QGVLTDQVAR	0.05048	5,877,981	P59579
QIVVSDYAQMDR	0.610813	7,568,449	P50991
QLIAGALQYTQ	0.0221432	6,473,291	P76562
QLMSDMAGISR	0.0321937	6,487,911	A7X715
QLNAYGGIVV	0.0249589	5,612,866	P18159
QLVWRPRPPT	0.210445	6,693,611	Q1DI97
QPQPGAALLTPGIV	0.0249589	4,839,306	Q885J7
QSVDDAAIVIK	0.433373	6,378,369	Q12906
QSVDDAAIVIK	0.395072	6,378,369	Q12906
QVLGPKPALPAGTEDTAKEDAANRK	0.99936	6,881,194	P06396
QVLPESVNVGFSAAT	0.511131	5,362,644	O24313
QVNHPHVI	0.335462	5,162,584	P07949
QVTQPTVGMNFK	0.998172	7,333,627	P0CW22
QVTQPTVGMNFK	0.981351	7,333,627	P0CW22
QYRPSFSSDMKPK	0.999998	5,722,813	P23141
QYRPSFSSDMKPK	0.999998	5,722,813	P23141
QYRPSFSSDMKPK	0.998887	5,722,813	P23141
QYRPSFSSDMKPK	0.0443073	5,722,813	P23141
RAAYSALYAPPPR	0.0224361	5,075,942	Q2P295
RAIDAIQKPDLSAETLIR	0.0175083	7,093,929	B6J505
RAQMAAEAGRPYA	0.40005	4,992,310	A5V3X4
RCAAVDGDGALR	0.0596199	6,748,086	C4IZX0
RCDSFSSFK	0.0497066	6,252,708	Q8GY91
RGQDSTGVTDRV	0.105417	4,602,176	O74941
RKPRDINEGIFAGGT	0.0164916	5,829,702	Q8Y8Q5
RLAAIYHTPK	0.0165233	6,433,581	Q2JV96

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
RLLSAATQAVPTPNQ	0.0420054	8,279,331	P20000
RLPVVQSYGMTE	0.0339368	7,423,498	Q65FT5
RLSAPEQQQQLEVPP	0.186918	9,044,542	P39001
RMSTGGPHGAA	0.0152906	3,771,675	Q753S8
RPDNFVFGQSGAGNNWAK	0.999999	6,943,269	Q9YHC3
RPDNFVFGQSGAGNNWAK	0.999999	6,943,269	Q9YHC3
RPGLVVVHAEDGTTSK	0.225732	5,946,455	P05630
RPKANGVKPSTVHVIS	0.0161749	6,120,172	Q9HAT8
RVAPEEHPVLLTEAPLNPK	0.999999	7,427,367	P60712
RVAPEEHPVLLTEAPLNPK	0.999999	7,427,367	P60712
RVAPEEHPVLLTEAPLNPK	0.999999	7,427,367	P60712
RVAPEEHPVLLTEAPLNPK	0.99999	5,573,045	P60712
RWLQMLLPMA	0.152759	6,738,431	Q60860
RYASENVNK	0.985651	5,987,903	P62822
RYASENVNK	0.828982	5,987,903	P62822
RYASENVNK	0.681048	5,987,903	P62822
SAALIQQATTVK	0.674385	6,738,711	Q3SYR7
SAALIQQATTVK	0.640303	6,738,711	Q3SYR7
SAANHDAAIFPGGFGAAK	0.993549	6,066,263	P30042
SAASCEPLASVLR	0.99968	7,248,474	P13489
SAASCEPLASVLR	0.999147	7,248,474	P13489
SAAVTALNSESNFAR	0.998472	8,133,832	O75390
SAAVTALNSESNFAR	0.950527	8,133,832	O75390
SADFPALVVK	0.0779876	5,818,126	Q59A32
SADFPALVVK	0.0481141	5,818,126	Q59A32
SAGGQLTTDQR	0.095599	6,112,803	P46940
SAGGRDEGNYLDDALVR	0.508365	6,326,234	P09525
SAGGRDEGNYLDDALVR	0.0343921	6,326,234	P09525
SAPAIGGGAHGIVFRN	0.457847	8,064,068	P15922
SAPGPITMDLTGDLEALKK	0.999999	10,510,496	Q9TU03
SAPGPITMDLTGDLEALKK	0.999999	7,010,355	Q9TU03
SAPGPITMDLTGDLEALKK	0.999999	7,010,355	Q9TU03
SAPSLAAPDGSAPSAPR	0.0199421	8,203,911	Q92529
SASFEPFSNK	0.984109	6,152,792	Q4VIT5
SASFEPFSNK	0.936067	6,152,792	Q4VIT5
SASGDKTSK	0.44407	5,127,528	Q2KJH4
SATKPGGRPR	0.0464574	5,718,088	Q24617
SCGIHETTFNSIMK	0.999999	5,809,301	P60712
SCGIHETTFNSIMK	0.999999	5,809,301	P60712
SCGIHETTFNSIMK	0.999999	5,809,301	P60712

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SCGIHETTFNSIMK	0.999999	5,809,301	P60712
SCGIHETTFNSIMK	0.999999	8,708,915	P60712
SCGIHETTFNSIMK	0.999999	8,708,915	P60712
SCGIHETTFNSIMK	0.999999	8,708,915	P60712
SCGIHETTFNSIMK	0.999982	5,862,617	P60712
SCGIHETTFNSIMK	0.999976	5,862,617	P60712
SCGIHETTFNSIMK	0.999894	5,862,617	P60712
SCGIHETTFNSIMK	0.999869	5,862,617	P60712
SCGIHETTFNSIMK	0.997177	5,862,617	P60712
SCGITHTAVVPLDLVK	0.999999	9,134,811	O61703
SCGITHTAVVPLDLVK	0.999999	9,134,811	O61703
SCGITHTAVVPLDLVK	0.999997	9,134,811	O61703
SCGITHTAVVPLDLVK	0.999938	9,134,811	O61703
SCGITHTAVVPLDLVK	0.999294	6,093,233	O61703
SDIYSFSMIIQEI	0.0200066	5,452,545	F1M5M3
SDYPPLGR	0.948323	4,967,291	P68103
SEAHSEFLK	0.997315	5,822,739	P38657
SEAHSEFLK	0.901468	5,822,739	P38657
SEAHSEFLK	0.3428	5,822,739	P38657
SEASAPIPHDGNLYPR	0.90606	9,064,231	O00560
SEATLYGNWSV	0.939593	4,388,607	P13769
SENDLNFIK	0.311381	5,982,870	Q3SYU2
SESIVNDFAYMK	0.479242	7,603,442	Q5E971
SETILLPFTR	0.0232847	4,222,252	Q16AP0
SFETTDESLR	0.969091	6,367,745	Q32P51
SFETTDESLR	0.7728	6,367,745	Q32P51
SFETTEESLR	0.237814	6,437,821	Q2HJ60
SFFPADVDECS	0.0839311	6,812,551	Q886L6
SFVNDIFER	0.0862967	6,077,794	P33778
SFVNDIFER	0.0429702	6,077,794	P33778
SGAQIDDNIPR	0.106727	6,372,961	Q16555
SGGPVVCSGK	0.961891	5,322,494	P00760
SGGTTMFNGIADR	0.0999493	7,158,058	P93584
SGGTTMYPGIADR	0.999999	7,153,082	P60712
SGGTTMYPGIADR	0.999999	7,153,082	P60712
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,153,082	P60712
SGGTTMYPGIADR	0.999999	7,153,081	P60712
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,153,081	P60712

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,073,107	P60712
SGGTTMYPGIADR	0.999999	7,153,082	P60712
SGGTTMYPGIADR	0.999996	7,153,082	P60712
SGGTTMYPGIADR	0.999996	7,073,106	P60712
SGGTTMYPGIADR	0.999986	7,153,082	P60712
SGGTTMYPGIADR	0.999979	7,153,081	P60712
SGGTTMYPGIADR	0.999927	7,153,082	P60712
SGGTTMYPGIADR	0.999918	7,153,082	P60712
SGGTTMYPGIADR	0.999914	7,153,082	P60712
SGGTTMYPGIADR	0.999576	7,153,082	P60712
SGLPTIPSPLDQEASSA	0.184032	8,794,169	Q8K0T0
SGMSDHR	0.1985	4,391,684	P31943
SGMYDHR	0.120627	4,771,840	Q5E9J1
SGQTNIHLSK	0.998925	6,008,059	P17655
SGQTNIHLSK	0.804405	6,008,059	P17655
SGSEPAYALAWIDI	0.020458	7,908,691	P56883
SGVDTVLGK	0.52013	4,962,603	O60664
SHNPEQK	0.0651684	4,782,189	P11586
SHQDLQLVK	0.587858	5,923,106	Q15418
SHSLLPALCDSK	0.999584	7,223,524	A1YES6
SHSLLPALCDSK	0.999376	7,223,524	A1YES6
SHSLLPALCDSK	0.99813	7,223,524	A1YES6
SHSLLPALCDSK	0.95346	7,223,524	A1YES6
SHSLLPALCDSK	0.382946	7,223,524	A1YES6
SHSLLPALCDSK	0.286388	7,223,521	A1YES6
SIANHQHGQK	0.505036	4,125,366	A8APY4
SIELPMDNK	0.18367	5,817,780	P17132
SIELPMDNK	0.0549534	5,817,780	P17132
SISCDHLIDKDIGSK	0.999746	6,113,026	Q99829
SISSIGVVGTAEDNR	0.999386	7,968,831	A6NHG4
SISSIGVVGTAEDNR	0.910673	7,968,831	A6NHG4
SLEEIELALTSGA	0.0224687	4,742,342	Q56ZZ7
SLEPVAVELK	0.984501	6,008,311	P00558
SLEPVAVELK	0.96734	6,008,311	P00558
SLEPVAVELK	0.940261	6,008,311	P00558
SLEPVAVELK	0.842492	6,008,311	P00558
SLFTLFGVYGDVQR	0.0212666	5,639,485	Q9UKA9

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SLGHPFGATGCR	0.848798	6,743,005	O46629
SLGHPFGATGCR	0.160358	6,743,005	O46629
SLVMEAPKGVEI	0.0248601	4,635,772	Q92629
SMTKSTVTSKELGFLT	0.0567988	6,253,265	O22795
SPSVWAAVPGK	0.689539	6,078,157	P07737
SPTIKPSPAASK	0.0877087	6,643,683	P30048
SQIIPQALGNPR	0.0259497	6,913,666	Q8H2J9
SQQVATVIQSR	0.198616	6,528,351	Q2HJG5
SQSTLLTGETK	0.900469	6,408,241	A4RN08
SQSTLLTGETK	0.842748	6,408,241	A4RN08
SQSTLLTGETK	0.0204379	6,408,241	A4RN08
SSALDHALRLLAGR	0.0636813	5,232,842	Q8KBK8
SSEALEAAR	0.058434	5,112,346	Q93VT9
SSLGAQVPDAK	0.970822	5,948,001	P04844
SSLGAQVPDAK	0.194382	5,948,001	P04844
SSMSDGAPVVK	0.0400933	5,972,806	Q6IVW0
SSNQAIQIAQHDAPVK	0.88012	6,083,088	P78406
SSTSHVPEVDPGSAELQK	0.999775	6,619,826	P49327
SSVAAGKKGYK	0.0742476	4,232,408	P07184
SSYAPVISAEK	0.0377948	6,343,156	P33623
STGGSVAGAVK	0.998977	5,252,686	P56519
SVADLAESIMK	0.392353	6,403,175	Q9BE24
SVADLAESIMK	0.114713	6,403,176	Q9BE24
SVAEITNACFEPANQMVK	0.999999	7,089,971	P68362
SVAEITNACFEPANQMVK	0.999999	7,089,971	P68362
SVAEITNACFEPANQMVK	0.999999	7,089,971	P68362
SVAEITNACFEPANQMVK	0.999999	7,089,971	P68362
SVAEITNACFEPANQMVK	0.999998	7,143,287	P68362
SVAEITNACFEPANQMVK	0.999997	7,143,287	P68362
SVAEITNACFEPANQMVK	0.999992	7,089,971	P68362
SVAEITNACFEPANQMVK	0.994714	7,143,287	P68362
SVAEITNACFEPANQMVK	0.917417	7,143,287	P68362
SVAEITNSCFEPANQMVK	0.444063	7,143,288	P09644
SVANADIINAAK	0.0196759	6,518,401	P22695
SVEDHLAWSK	0.998826	6,443,057	P30041
SVEDHLAWSK	0.997889	6,443,057	P30041
SVEDHLAWSK	0.987425	6,443,057	P30041
SVLGGMGIPGMGRK	0.0650506	7,463,776	O33013
SVNSHFMK	0.219775	5,332,466	P52505
SYLSSHIANVER	0.90539	7,323,512	P06396

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SYNINNDIMLIK	0.101299	7,773,889	P16049
SYSATEETLQEVFEK	0.999998	9,389,301	P19338
SYSATEETLQEVFEK	0.999998	9,389,301	P19338
SYSATEETLQEVFEK	0.999753	9,389,301	P19338
SYSATEETLQEVFEK	0.653259	9,389,301	P19338
SYSATEETLQEVFEK	0.0803104	9,389,301	P19338
SYVGDEAQSK	0.99883	6,002,663	P60712
SYVGDEAQSK	0.860747	6,002,663	P60712
SYVGDEAQSKR	0.991574	6,783,168	P60712
SYVGDEAQSKR	0.773365	6,783,168	P60712
TAASSSSLEK	0.997765	5,487,634	P60712
TAASSSSLEK	0.99762	5,487,634	P60712
TAASSSSLEK	0.997392	5,487,634	P60712
TAASSSSLEK	0.996327	5,487,634	P60712
TAASSSSLEK	0.996241	5,487,634	P60712
TAASSSSLEK	0.981216	5,487,634	P60712
TAASSSSLEK	0.979061	5,487,634	P60712
TAASSSSLEK	0.97862	5,487,634	P60712
TAASSSSLEK	0.978244	5,487,634	P60712
TAASSSSLEK	0.960213	5,487,634	P60712
TAASSSSLEK	0.934636	5,487,634	P60712
TAASSSSLEK	0.90776	5,487,634	P60712
TAASSSSLEK	0.822477	5,487,634	P60712
TAASSSSLEKR	0.990883	6,268,141	P53483
TAASSSSLEKR	0.712331	6,268,141	P53483
TADSQPPVFK	0.53373	6,032,974	Q13263
TAHIACK	0.327097	4,587,228	P27592
TAQVDAEEKEDVK	0.990514	8,033,877	P13620
TAQWLPDVTRPEGK	0.999418	8,574,353	Q3MHL3
TAQWLPDVTRPEGK	0.620398	8,574,353	Q3MHL3
TATGGVQSTASSK	0.997277	6,558,167	Q16658
TAVIDHHNY	0.169951	5,792,560	A6QPY0
TCLNEASVLHNLK	0.998051	8,079,026	P35579
TCLNEASVLHNLK	0.995577	8,079,026	P35579
TEALLTLSKKG	0.109039	6,528,785	Q2YAX2
TEAPLNPK	0.883888	4,932,550	P60712
TEAPLNPK	0.699176	4,932,550	P60712
TEHVAAAELGAR	0.993517	6,568,195	O43598
TEHVAAAELGAR	0.935858	6,568,195	O43598
TEHVAAAELGAR	0.911665	6,568,195	O43598

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TEKEQIVPKPEEEVAQK	0.998297	7,187,171	Q3T025
TESYISTIGVDFK	0.0194425	7,883,846	P62822
TFDPAALLPGASPK	0.996258	7,508,922	Q6JBY9
TFIQNVPLEDR	0.978004	7,103,507	Q00839
TFIQNVPLEDR	0.720786	7,103,507	Q00839
TGAGCGVLSSTLR	0.999812	6,838,265	P13489
TGAGCGVLSSTLR	0.999624	6,838,265	P13489
TGAGCGVLSSTLR	0.975764	6,838,265	P13489
TGDIPLAADVLEK	0.0771171	7,293,816	A4G1V0
TGDLEALKK	0.0657786	5,598,101	Q9TU03
TGFGGYDKVK	0.917152	6,083,077	Q99536
TGLSIGPDFQK	0.0222082	6,398,238	Q99439
TGMGNSTNKK	0.952531	5,912,865	Q28141
TGQTPVFSK	0.937097	5,407,735	Q3T087
TGSVASASATPGK	0.0156691	4,172,082	A1Z9E2
TGVHQVPTENVQVHFTER	0.999997	7,226,849	Q9HB71
TGVHQVPTENVQVHFTER	0.999997	7,226,849	Q9HB71
TGVHQVPTENVQVHFTER	0.999997	7,226,849	Q9HB71
THGEVCPANWTPDSPTIKPSPAASK	0.999997	6,988,401	P30048
THGEVCPANWTPDSPTIKPSPAASK	0.999997	6,988,401	P30048
THGIFSGPAISR	0.087689	6,658,326	Q2HJ58
THSLGGGTGSGMGTLLISK	0.999999	6,306,535	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999999	6,359,853	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999999	6,359,853	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999999	6,306,535	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999999	9,454,768	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999999	9,454,768	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999551	6,359,853	Q9YHC3
THSLGGGTGSGMGTLLISK	0.999335	6,359,853	Q9YHC3
TLADCNLLPK	0.976562	6,308,202	Q5E9B7
TLADCNLLPK	0.337014	6,308,202	Q5E9B7
TLADCNLLPK	0.317562	6,308,202	Q5E9B7
TLADCNLLPK	0.151264	6,308,202	Q5E9B7
TLADCNLLPK	0.133392	6,308,202	Q5E9B7
TLGTEITVEDQLAR	0.65772	8,174,091	P45879
TLNLEDVQPHDLGK	0.903255	8,479,246	P10809
TLNLEDVQPHDLGK	0.210584	8,479,246	P10809
TLNNDIMLIK	0.997282	6,538,411	P00760
TLNNDIMLIK	0.995296	6,538,411	P00760
TLNNDIMLIK	0.98787	6,538,411	P00760

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TLNNDIMLIK	0.986276	6,458,436	P00760
TLNNDIMLIK	0.976243	6,458,436	P00760
TLNNDIMLIK	0.974723	6,538,411	P00760
TLNNDIMLIK	0.970378	6,458,436	P00760
TLNNDIMLIK	0.950876	6,538,411	P00760
TLNNDIMLIK	0.933048	6,458,436	P00760
TLNNDIMLIK	0.898283	6,458,436	P00760
TLNNDIMLIK	0.848682	6,458,436	P00760
TMLTNETGFEISSSDATVK	0.998637	7,163,345	Q7TP98
TNLCAIHAK	0.056636	5,722,863	P68432
TNSGFAAASSGSPAGFG	0.0476933	7,873,332	O15504
TNVEIAVLVGID	0.0258835	6,658,500	A1UQA0
TPDTLTSDTYKK	0.993247	7,573,766	P17655
TPLSDGATTPADILAIK	0.0174128	6,006,535	Q5L898
TQAAPALDFSLK	0.501699	6,893,581	Q3T063
TQAAPALDFSLK	0.165991	6,893,581	Q3T063
TSDAARMAAELK	0.166303	6,983,342	A5FUG5
TSRDTHLGGEDFDNR	0.999587	6,032,605	P22623
TSRDTHLGGEDFDNR	0.988128	6,032,605	P22623
TSRDTHLGGEDFDNR	0.97985	6,032,605	P22623
TSRDTHLGGEDFDNR	0.911862	6,032,605	P22623
TTGIVFDSGDGVSYTVPIYEGY	0.257567	8,100,350	P53465
TTIAGVVYK	0.0315818	5,342,941	Q2TBP0
TTVHAITATQK	0.998742	6,438,425	P10096
TTVHAITATQK	0.991014	6,438,425	P10096
TVAEITNACFEPANQMVK	0.999465	7,136,689	P68365
TVHAITATQK	0.999641	5,933,186	P10096
TVHAITATQK	0.99749	5,933,186	P10096
TVHAITATQK	0.99194	5,933,186	P10096
TVHAITATQK	0.372511	5,933,186	P10096
TVKLPDGYEFK	0.861631	7,208,760	P09382
TVKLPDGYEFK	0.104106	7,208,760	P09382
TVKLPDGYEFK	0.047666	7,208,760	P09382
TVKLPDGYEFKFPNR	0.999998	6,523,415	P09382
TVKLPDGYEFKFPNR	0.998115	6,523,415	P09382
TVKLPDGYEFKFPNR	0.997879	6,523,415	P09382
TVLSGGTTMYPGIADR	0.999999	8,639,106	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,106	P60712
TVLSGGTTMYPGIADR	0.999999	8,719,081	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,106	P60712

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TVLSGGTTMYPGIADR	0.999999	8,719,081	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,106	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999999	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999998	8,719,082	P60712
TVLSGGTTMYPGIADR	0.999998	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999994	8,639,106	P60712
TVLSGGTTMYPGIADR	0.999993	8,639,108	P60712
TVLSGGTTMYPGIADR	0.999987	8,719,081	P60712
TVLSGGTTMYPGIADR	0.999981	8,719,082	P60712
TVLSGGTTMYPGIADR	0.999973	8,719,082	P60712
TVLSGGTTMYPGIADR	0.999781	8,719,081	P60712
TVLSGGTTMYPGIADR	0.999657	8,719,081	P60712
TVLSGGTTMYPGIADR	0.998425	8,719,082	P60712
TVLSGGTTMYPGIADR	0.997208	8,719,081	P60712
TVQQIEHLNK	0.998339	6,633,479	Q07130
TVQQIEHLNK	0.995305	6,633,479	Q07130
TVQQIEHLNK	0.706171	6,633,479	Q07130
TVSNEAQTAK	0.069121	5,827,821	A4FUA8
VADENPFAQGALK	0.777407	7,383,636	P06396
VADNAGGSHASK	0.996258	6,152,828	Q7Z5L9
VAGTPMFVVK	0.881547	5,828,116	Q3SYU2
VALNPDFKPPADYKPPATR	0.689435	7,477,295	Q15637
VALNPDFKPPADYKPPATR	0.345794	7,477,295	Q15637
VANLTDAFEGSTIR	0.0356936	7,913,827	Q7TYX4
VCSSAPGPLELDLTGDLESFKK	0.999997	8,364,186	P52565
VCSSAPGPLELDLTGDLESFKK	0.999997	8,364,186	P52565
VFGQSGAGNNWAK	0.999999	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.999999	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.999983	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.999967	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.999933	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.999749	7,263,406	Q9YHC3
VFGQSGAGNNWAK	0.99861	7,263,406	Q9YHC3
VFKDDDDVVIGK	0.285421	7,473,816	Q3MHR7
VFLPFADDK	0.910548	5,842,916	P12956

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
VFLPFADDK	0.600781	5,842,916	P12956
VFLPFADDKR	0.859009	6,623,421	P12956
VFLPFADDKR	0.602519	6,623,421	P12956
VFLPFADDKR	0.202254	6,623,421	P12956
VFLPFADDKR	0.170569	6,623,421	P12956
VGDVGQTVDDPYATFVK	0.999999	9,639,616	P23528
VGDVGQTVDDPYATFVK	0.999999	9,639,616	P23528
VGDVGQTVDDPYATFVK	0.999999	6,429,768	P23528
VGGASLKPEFVDIINAR	0.480479	9,515,116	Q1MTI4
VGLLGTEAATEPEDIK	0.0162699	8,799,453	Q2U910
VGMETCGIHETTFNSIMK	0.824595	7,243,263	P43239
VGMTPEIIQK	0.621537	6,163,251	Q08E38
VGMTPEIIQK	0.102407	6,163,251	Q08E38
VGSAKPGLQK	0.996697	5,648,261	P00918
VGSAKPGLQK	0.345132	5,648,261	P00918
VGVFQHGKVEIIANDQGNR	0.885308	7,330,445	P08106
VHVTDLSGK	0.489468	5,362,790	Q7QEH1
VHVTDLSGK	0.451747	5,362,790	Q7QEH1
VILPVPAFNVINGGSHAGNK	0.999998	7,073,825	P06733
VILPVPAFNVINGGSHAGNK	0.999808	7,073,825	P06733
VILPVPAFNVINGGSHAGNK	0.999808	7,073,825	P06733
VINGGSHAGNK	0.448042	5,852,904	P06733
VINGGSHAGNK	0.0624076	5,852,904	P06733
VISCPHLGASTK	0.215435	6,933,496	O43175
VLIPTEGGDFNEFPVPEQFK	0.999999	7,937,242	Q01518
VLIPTEGGDFNEFPVPEQFK	0.999999	7,937,242	Q01518
VNDAFGTAHR	0.992425	5,882,669	P00558
VNDAFGTAHR	0.957657	5,882,669	P00558
VPAVLLSGNHAK	0.0206518	4,412,482	Q6YQA4
VPKPPPYSK	0.0265463	3,862,195	Q5DTW7
VPLKDKTDMLHIK	0.0429674	5,759,932	Q8KAY6
VPNGITLPVDFQGR	0.206694	5,342,767	P31943
VQSGSHLAAR	0.998959	5,572,773	Q2I6W4
VQSGSHLAAR	0.989658	5,572,773	Q2I6W4
VQSGSHLAAR	0.989247	5,572,773	Q2I6W4
VSATMSGVTTCLR	0.998598	7,358,411	Q9YHC3
VSGFHPSDIEVDLLK	0.928622	8,864,506	O77523
VSGFHPSDIEVDLLK	0.894415	8,864,506	O77523
VSIDDALNH	0.0520439	3,578,308	Q74BW4
VSIDDALNH	0.0371761	3,578,308	Q74BW4

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
VSSPPPADLCHALR	0.0707276	8,043,871	P24534
VTALCDIK	0.0906753	5,182,645	Q3ZCK9
VTALNPHIGYDK	0.0688252	7,223,689	P07954
VTASQCQQPAENK	0.999991	7,888,585	Q01518
VTASQCQQPAENK	0.999883	7,888,585	Q01518
VTFDDHDSVDKIVIQK	0.999997	6,683,434	Q32P51
VTGPLSSNHG	0.119532	5,287,428	P76250
VTTSHDASVK	0.999363	5,807,846	Q2KJH4
VTTSHDASVK	0.999151	5,807,846	Q2KJH4
VTYVPVTTFK	0.999695	6,358,414	Q56JX3
VTYVPVTTFK	0.999087	6,358,414	Q56JX3
VVAANDGPMPQTR	0.0369671	7,223,396	P49411
VVDNGSGMCK	0.999985	5,917,514	P60712
VVDNGSGMCK	0.999975	5,917,514	P60712
VVDNGSGMCK	0.999408	5,917,514	P60712
VVDVAQGTQVTGR	0.575339	7,093,590	Q9UQ80
VVEDSSNMEGDNEDSK	0.0371855	9,358,697	F4JET1
VVESTGVFTTMEK	0.999827	7,723,731	P10096
VVESTGVFTTMEK	0.999827	7,723,730	P10096
VVESTGVFTTMEK	0.994547	7,723,730	P10096
VVESTGVFTTMEK	0.502125	7,723,730	P10096
VVGLCTGQIK	0.0956508	5,958,176	P06733
VVGLCTGQIK	0.0348717	5,958,176	P06733
VVMSGGTTMYPGIADR	0.0405619	8,718,991	P10989
VVNDAGRPK	0.0480897	5,362,846	P19120
VVPTLLTQM	0.317761	5,452,881	C0NX46
VVTNAAGGLNPK	0.877647	6,288,372	P00491
VYEGHVSCVK	0.998479	6,473,021	Q3T0F7
WAPESAPLK	0.421448	5,577,839	P23528
WAPESAPLK	0.312304	5,577,839	P23528
WAPESAPLK	0.164124	5,577,839	P23528
WAPESAPLK	0.0698269	5,577,839	P23528
WHDPIKEDVYR	0.994624	5,252,539	P50897
WISKEEYDESGPSIVHR	0.613604	7,166,738	Q553U6
WNNTAADDKQPYEK	0.999974	9,124,171	P10103
WNNTAADDKQPYEK	0.999923	6,086,138	P10103
WNNTAADDKQPYEKK	0.999998	6,606,560	P10103
WNNTAADDKQPYEKK	0.992244	6,606,560	P10103
WTPYAAIVGL	0.864315	3,935,355	Q111U1
YAAQAHLK	0.990918	5,092,631	Q2T9P4

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
YAAQAHLK	0.975701	5,092,631	Q2T9P4
YAAQAHLK	0.956271	5,092,631	Q2T9P4
YADPVSAQHAK	0.9994	6,518,112	P26599
YAHELPK	0.0585103	4,872,444	P15125
YAILAVLSIY	0.0209428	6,073,306	Q6FM46
YALADPSLKM	0.146785	6,208,014	P03172
YALPHAILR	0.999283	3,812,112	P60712
YALPHAILR	0.998962	3,812,112	P60712
YALPHAILR	0.998789	3,812,112	P60712
YALPHAILR	0.998608	3,812,112	P60712
YALPHAILR	0.995169	5,713,132	P60712
YALPHAILR	0.994432	5,713,132	P60712
YALPHAILR	0.993709	5,713,132	P60712
YALPHAILR	0.993496	5,713,132	P60712
YALPHAILR	0.987613	5,713,132	P60712
YALPHAILR	0.985776	5,713,132	P60712
YALPHAILR	0.979871	5,713,132	P60712
YALPHAILR	0.978782	5,713,132	P60712
YALPHAILR	0.978529	5,713,132	P60712
YALPHAILR	0.97613	5,713,132	P60712
YALPHAILR	0.964608	5,713,132	P60712
YALPHAILR	0.963029	5,713,132	P60712
YALPHAILR	0.963029	5,713,132	P60712
YALPHAILR	0.959768	5,713,132	P60712
YALPHAILR	0.955689	5,713,132	P60712
YALPHAILR	0.948574	5,713,132	P60712
YALPHAILR	0.939796	5,713,132	P60712
YALPHAILR	0.936869	5,713,131	P60712
YALPHAILR	0.839536	5,713,132	P60712
YALPHARIL	0.0221106	5,713,131	P0A527
YANPPVPIIA	0.0501482	5,717,996	O18017
YANTVLSGGTTMYPGIADR	0.999999	6,923,241	P60712
YANVIHCK	0.496183	5,607,676	Q58CP0
YAPWCGHCK	0.979707	6,477,621	P05307
YAPWCGHCK	0.375966	6,477,621	P05307
YAPWCGHCK	0.351136	6,477,621	P05307
YASPAHAVDAVK	0.996937	6,728,347	P55884
YCLPHAILR	0.435272	6,158,099	P86700
YDERGPSIVHR	0.0370854	4,728,946	P30164
YEGPPDDEAAMGIK	0.968166	8,048,496	Q3SYU2

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
YELPFIEQMVK	0.0303966	7,568,776	H8ZM73
YFLFRDGDILGK	0.338706	7,804,001	P61603
YGKIDTIEIITDR	0.999857	8,269,320	Q2HJ60
YGKIDTIEIITDR	0.999857	8,269,320	Q2HJ60
YGKIDTIEIITDR	0.999857	8,269,320	Q2HJ60
YGKIDTIEIITDR	0.999857	8,269,320	Q2HJ60
YGKIDTIEIITDR	0.999141	8,269,320	Q2HJ60
YGKIDTIEIITDR	0.284097	8,269,321	Q2HJ60
YGQDEEAVKK	0.252657	6,558,186	P09874
YHEMIESGVINLK	0.32909	8,249,075	P00829
YHGQVLCK	0.964703	5,607,677	P40939
YHGQVLCK	0.234648	5,607,677	P40939
YHIISSNLEK	0.135464	6,603,370	Q5E9G3
YKGLTVEEVTELRK	0.0461675	6,036,658	Q18CE8
YLAGMLGADPVI	0.113614	6,623,201	Q05631
YLGDAFAEWGIP	0.400729	7,138,212	Q8DJ26
YLNIYTPADLTK	0.999603	7,643,920	P23141
YLNIYTPADLTK	0.998412	7,643,920	P23141
YLNIYTPADLTK	0.998412	7,643,920	P23141
YLNIYTPADLTKK	0.998858	8,424,551	P23141
YLNIYTPADLTKK	0.0943863	8,424,551	P23141
YLNIYTPADLTKK	0.0794945	8,424,551	P23141
YLNLAHN	0.0941578	4,667,186	Q9LUL4
YLNLAHN	0.0825578	4,667,186	Q9LUL4
YPAGATTNSDSGVTDPGKITHLYEAMEK	0.0262147	10,384,865	Q0VNK6
YPLIPVGK	0.629384	5,017,885	Q2I6W4
YPLIPVGK	0.205235	5,017,885	Q2I6W4
YQALASGK	0.193312	4,772,419	Q9UBQ7
YQALASGK	0.159353	4,772,419	Q9UBQ7
YQEVEVKCVCGNTFKTGSTK	0.0176994	8,364,043	A4J9C1
YQGAGGPGPGGFGAQGPK	0.999998	5,736,035	P08107
YQGAGGPGPGGFGAQGPK	0.999998	5,736,035	P08107
YQGAGGPGPGGFGAQGPK	0.701953	5,736,035	P08107
YQPTAPSAPDLPADQGPASGYGLQR	0.0174128	8,824,148	A2QWA2
YQSNTILR	0.165657	5,417,688	P09211
YQSNTILR	0.0805321	5,417,688	P09211
YQVCHTK	0.129323	5,262,388	P61157
YREGGVESAFHK	0.999998	4,992,382	P24452
YREGGVESAFHK	0.999998	4,992,382	P24452
YREGGVESAFHK	0.89679	4,992,382	P24452

Supplementary Table S1 - rRLR1 cleavage sites identified from a tryptic peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
YREGGVESAFHK	0.519882	4,992,382	P24452
YRIPADVDPLAITSSL	0.0203289	6,069,818	P02510
YSFPVVIK	0.0351065	5,347,936	Q5ZME2
YSGQQSTYGK	0.399259	6,177,741	Q7ZX83
YSIPGDYARVEAD	0.0185617	7,723,406	E3UUE6
YSKEDATK	0.0743339	5,432,630	Q5E971
YSLEPVAVELK	0.646215	6,823,626	P00558
YSLPHAILR	0.580549	5,793,106	P26197
YSSAVTVGPHK	0.993887	6,313,161	O60496
YSVTANSK	0.781253	4,932,366	Q9BE24
YSVTANSK	0.226442	4,932,368	Q9BE24
YSVTANSK	0.104412	4,932,368	Q9BE24
YSVTANSK	0.0586286	4,932,368	Q9BE24
YTTNGKPFEGDPR	0.0222082	7,993,696	O59648
YTVLAADDPR	0.0373944	6,047,846	P52580
YVAKGEDVNR	0.719712	6,338,112	Q3T0F7
YVAKGEDVNR	0.166475	6,338,112	Q3T0F7
YVDLAPTSGTSAK	0.0446878	7,133,501	P51659
YVELQKEEAQK	0.999492	7,548,871	Q00839
YVELQKEEAQK	0.995331	7,548,871	Q00839
YVHMATADGLR	0.111623	6,613,051	Q14697
YVNGLTLGGQK	0.999794	6,333,318	P07737
YVNGLTLGGQK	0.999588	6,333,318	P07737
YVNGLTLGGQK	0.928413	6,333,318	P07737
YVVESTGVFTTMEK	0.999999	8,539,046	P10096
YVVESTGVFTTMEK	0.999999	8,539,046	P10096
YVVESTGVFTTMEK	0.99994	8,539,046	P10096
YVVESTGVFTTMEK	0.999879	8,539,046	P10096
YVVESTGVFTTMEK	0.993655	8,539,046	P10096
YYFGDFNLPR	0.996431	6,903,083	P10881
YYFGDFNLPR	0.984758	6,903,083	P10881
YYVDEHGTR	0.0412244	6,142,588	P28062

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem. Peptides identified by LC-MS/MS spectrum-to-sequence assignment with Mascot and X!Tandem are listed with PeptideProphet probability score, calculated neutral mass and one exemplary accession number of a matching Uniprot protein entry is listed. This data was further processed and rendered non-redundant for generation of cleavage specificity profiles.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SKTLAPP	0.99921	4152281	B2RU80
GLKISKKIIKKIKKKIEK	0.0988145	8219225	Q057D7
KVKLGKIKVGKPVIKE	0.132533	10106885	Q58444
ILPLITVP	0.671829	4772906	Q04734
FIQSLQINGKPIAAGGR	0.0338714	6293489	O31530
GEPCHDPSN	0.0502051	5506926	Q90839
CYCSCDSYE	0.132387	6661880	P38168
KLREPLVVKAGKPVIVKI	0.0422661	7298179	Q86VF2
QPCMNDE	0.0323597	4991566	Q5U1Y9
IGLIKVHK	0.090225	5263386	A3MXP1
LLVLHYGG	0.0459932	4802548	A7I912
SCYSYEMG	0.239155	5506723	Q7V2R3
FPPLLILPPA	0.0759728	5833383	D3ZAP3
YQVKGKKNIRATE	0.471234	4274961	P40121
AMGIDPTFGG	0.0431299	5352203	Q1AVY9
EPNPARA	0.106487	4216951	Q8N8E3
TAHMVPLH	0.0736306	4972361	P24434
HQVHNVAALPGA	0.773655	4345523	Q02251
NCLNDPTIE	0.0434624	5822392	Q03744
EAMDTTPAPVKKPT	0.113217	8234126	P15771
LRASHMAHVYD	0.99055	4632132	P54873
NMAWNEPSDSE	0.0396135	6922451	Q89A39
KDPLLASGTD	0.0572697	3781902	P22102
IGNAPAR	0.0342004	3937001	Q98E34
YAGELAHIIGCEPD	0.0433294	5449061	P60791
TLPSEPENGA	0.0457932	5517399	A8K0R7
HMACCASDTGGSPHPAASCQPSCLE	0.292165	6987630	A2A5X4
RPGLLNAGDPNYPWLAD	0.0508085	9789675	Q7TPD3
QIRQGNTKQRIDE	0.76053	5679617	Q2HJ49
KVGGLLGTLA	0.0427976	5228099	P83862
GAITSSNTRDTNAD	0.0334111	7558258	P04602
AHYNKRSTITSRE	0.846351	5602826	P02281
HATLQKSTKE	0.494881	4295641	Q13263
QNTVSDDSIKK	0.0197469	4602345	Q94819
QNTVSDDSIKK	0.0274852	4602345	Q94819
AIFDPIMGLI	0.380432	5973011	O34368
SGPWMWCDPE	0.0612641	6847385	P13691
FNKQVAKVTGTE	0.0427311	7333898	Q6F1V7

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
GGHAYIAGQCGLVPVMAEYYESSNCAIP	0.0345954	7763422	Q921I1
RKAHELKL	0.0368368	3802305	Q07657
YALCGAIRRMGE	0.189955	7428442	Q4GXP3
CINGFCVCEPD	0.0945096	7297617	P13508
QAMVDRRSAEE	0.073837	4658713	Q7U406
EFTAISIPGVFNHGNLD	0.0477291	6403051	Q3ILJ2
KAAPAPAPSSDPT	0.097189	6633241	P07530
RPGLVVVHAE	0.0731492	5828135	P05630
VEGTHELLSD	0.0452599	5942663	B7GL36
MRTPIANALA	0.155715	5732941	Q48F65
CGGGYSESIAEMC	0.0605858	5088418	Q9VR91
PVAGRFDAIRSPQGYTAIVD	0.0429305	7410406	Q5LIJ3
FLALLGHNGAGKTTLA	0.141455	5673116	Q8TQ05
GPLLIRIHPS	0.509763	3975622	P40962
CTGSNGSVNSF	0.0392163	4064904	Q54P51
APPQDNMLNSMYQLWT	0.0410714	10149288	P34498
GDPSDGFS	0.560291	4351529	A5CXU9
AFKVKDGPGGKE	0.170653	4689235	P07857
PKPKAEAQAEAQAE	0.332602	5379399	A4QEL0
NPISVSFNTPNLVYLM	0.0380257	4789852	Q9LQC1
RGTVEQDGLTDPP	0.0568645	7368381	Q5V5M5
AAYNQRKAD	0.173868	5767771	Q8WUM4
YAGKDGYNYTLSKTE	0.969739	9274351	P31949
KNVFGVEPSDAMKKAAL	0.0412041	9970287	A8FFL2
RGLMPSCSAFNE	0.109273	7288047	Q3ECH5
AAIQGLLKKHE	0.0400771	4512638	P07751
AATAGDPEALRS	0.032491	4161962	Q02835
TPSGAAVANFTVASTPRIYD	0.0686908	7093457	P0A611
QGLPLAGSQGIYQLKIGDS	0.0331481	6876945	P40354
THVQRQE	0.130335	4932298	Q3ZC50
EPSIPTQEPE	0.0403422	6077661	Q1H397
WSIDSLNKKGIL	0.0339372	5066182	B5MCY1
IPVQVTFQGGEPE	0.0376292	7448556	B1I293
EHPAAELPMA	0.0544371	5772546	B1MB85
GPAADWVRDGAPV	0.0612641	6998274	O52199
RPDSLPPIAAGE	0.0468607	6558243	Q1QI14
KEPAGSKE	0.0613998	4952526	Q9UUN3
THGVNPKAI	0.889026	3515235	Q4LAB5
YVKTRDGSDYE	0.957377	7248219	Q14697
LRFQSSHHPTD	0.0714216	4715508	P14625

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
VFPNFKGNVE	0.0732867	6338132	O74480
NLGTIAKSGTK	0.0938057	6173474	A6QBI2
MHLKPISRE	0.0671186	4095511	A1A4K8
INAPYRIE	0.0660947	5322659	Q5WEK8
LRSPHQNVCE	0.610022	6642979	O00505
FSAALAHIQCST	0.056797	6973158	A1WVL3
AGGIIGVKGAKIKE	0.292008	5049777	Q3T0D0
LFKDEVRQLGVALG	0.254371	5543080	Q7NSG1
NTPQQPKKK	0.329545	4142410	Q2HJ33
YRPGHYD	0.633695	4982058	Q96FW1
QAIAQAEVGMGE	0.112355	6542841	B1WR00
SQATLNKVFAL	0.0594339	6543552	P81921
LTSRSNAQRQE	0.337449	6893308	P08758
MSPVLAGVLQ	0.063302	5597831	Q82PI3
TPELNGAAPGCQLV	0.740876	5055709	Q9UT05
PKGMLPFSEDWFMSWE	0.9063	7123036	D5CIL6
YMVPAAFVFMADLPL	0.0559197	4479658	Q70LM5
AIRQRSLGPSLATDKS	0.99385	6059965	Q9H3N1
HVVFGKVKE	0.519625	3962267	Q6DTV9
FRPSHLTNKFE	0.432178	4979187	P38657
FSKLSDFAAALDPP	0.0539659	7978949	Q9NY46
DPVAAARDIRTTFSRMAMNDE	0.0421997	8297070	A4R5S9
APTGTGAPEGRRVPE	0.102927	5282598	P94679
TYGDSSSRFD	0.0338714	6117379	Q2FFN4
ILMIGPTGVGKTEIARRLAKLADSPFIKVE	0.0340688	11379872	B8D8E2
STGTNLLRLLE	0.0567295	4355676	A3PBC7
LVTVTNQPLLGVGLMD	0.0586894	5919758	P13574
FVSSIPSSPAGGGYNVTMIDD	0.0767318	5512488	P54602
RVGDPVT	0.0330824	4162051	P23747
SQAMAVIKE	0.0745161	5467752	P05198
ESSAGGGSDQPE	0.45635	6047224	Q6BUR3
AAGRVHAMLQNPY	0.0355835	7583636	Q01846
FLAPKDKPSGDTAAVFE	0.221267	6463291	P13010
VSMSIKCGILF	0.262909	6938453	Q9DBS4
YLRYTPQP	0.605241	5632737	O77834
YRSGQAKE	0.107414	5277531	P45478
CRGVPDAKIRIFD	0.339561	5549535	Q2TBW8
CFVTHCGWNSAVESLSCGIPVVAVPQWFD	0.0345295	8533831	Q9ZR26
YKGLGTSTPME	0.854301	6503018	P87078
MTALPAH	0.896171	4226886	C1A0X4

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
KDFPLVSGGSSL	0.0659582	6618369	O87296
ATHGDPHA	0.999999	4471824	Q2FZL2
TVRVTRAGVAVLTDSAPV	0.0365728	6340157	Q6C931
KTHTVGVILPYSD	0.0879155	3872052	P37517
PSSEFNPPMPTANTDND	0.753622	6465892	Q6BY37
KTFLLADVVVTEIPE	0.0371008	5973268	Q42695
LLLIGLRLFFGAE	0.290436	5172988	Q6D2F0
YDFTDFTLPS	0.252872	6472710	Q98TR3
KIFVGGIPHNCGETE	0.131541	8874188	Q96EP5
NNMSYDLIVCDAE	0.0441277	8243206	A4UHQ7
SQRVLSGLDITISMLDTPD	0.0490666	7170222	P08924
QGAQFVVNI	0.598478	5322661	Q0ICH2
TVLSGGTTMYPGIAD	0.925238	7938577	P53478
RVNSATQ	0.136505	4322058	D3IUT5
RYANNSNYKND	0.848227	7378228	Q0VC92
AMMDLLMDYL	0.0436619	6682706	Q04W49
SGGTTMYPGIAD	0.145096	6372576	P53478
VHMRGLPFRATE	0.501823	5012517	P31942
NTFSVVPSPKVSD	0.438464	7468714	Q17299
VVQFTVKHE	0.264717	6018157	Q4VIT5
ASRALLHPLLFE	0.135621	4855991	Q9Z8M4
KSSMKGQNFISFCRLD	0.0341346	10314996	B8BDQ4
YVTIIDAPGHRD	0.167208	7228483	P68103
AGIPYASLGLVTNM	0.227977	7558678	Q8IMU4
RNAVAVIGDGSMSAGMAFE	0.0401434	9939397	A1VMD7
FVIKPIDKKAPD	0.810063	5149704	Q2HJ49
YAKLPLPTDKIQKD	0.216044	6013427	Q8NHU6
HTNAACFACILLSHG	0.0545045	8803895	P55210
TSDTSRQPSLL	0.261397	6468114	Q4WMJ7
VKALCEMEPEGIAGAITGRAI	0.127703	7733901	Q2YAU9
FNPAMAIGGSSPAWTFMYIESLAD	0.105061	8993955	Q20848
KQYINAVRVVGRLGKNPE	0.0414695	7290808	A6VQU2
ATHGDPHALK	0.999956	3881941	Q2FZL2
FHTTGLVWSKTGPVAKE	0.432312	6680205	Q4R4I6
IFVANAGVAW	0.0927512	5682841	P87219
NVKLTKLTE	0.902586	5953469	Q3MHL4
GTGGMIIDSGTTLT	0.0417521	7063260	Q766C2
DGSTGQAIP	0.088265	4672031	Q8S9L5
KTLDPKALTGTADSMCFGPLSDI	0.0588924	6503180	Q50759
LSGARFYTYRGLGAKLE	0.0636421	6733559	B0S8Z4

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
GQKCSVIRD	0.182611	5897867	P02584
FQADELIKGYNV	0.0471278	7568740	Q9DKU0
IGSFQESMSFNENRE	0.0550433	9318881	O58931
YVVVMPKKRQALVE	0.42129	6020125	P14866
TVLSGGTTMYPGIAD	0.895264	7938577	P53478
CYKSLQACGMGIIANGELLD	0.0499371	7823598	A4WG55
TLLTSRSNAQRQE	0.976015	7963967	P08758
DIENTVTVMGG	0.102074	6202656	A6H0N4
TFSVVPSPKVSDTVVE	0.214895	9039635	P33188
FSTKGKSLE	0.0663676	5708023	Q12874
HVAVKVLQD	0.98534	5628105	Q91550
QENVLYKN	0.0680753	5622764	P39812
TVLSGGTTMYPGIAD	0.921877	7938577	P53478
HLKEDQTE	0.200537	5582557	Q76LV2
TVLSGGTTMYPGIAD	0.050562	7858602	P53478
SKKGVNLPGAAVDLPAVSE	0.923063	6660312	P14618
DYVLQNFIPGTGTM	0.085821	8303735	A1BGZ7
NYKPPPQKSLKE	0.999999	5342972	Q9TU03
KQLLGAVKGSFPSYSFE	0.0429305	6680168	Q2U9J2
EVTHAIKPVQN	0.690928	4512396	Q9VMX0
SRSCRMMANFVCEFQV	0.141455	7092965	Q71RR1
RLASLPTREEALAMLLATMKAPVQK	0.412311	7254050	Q0ABI4
SGGTTMYPGIAD	0.72458	6292601	P53478
QYIPETLMTE	0.314658	6567936	Q03CY3
RIINEPTAAALAYG	0.733773	7743982	Q04VC8
KIGYNPD	0.114275	4617208	P68103
DRAADPVSPLLHE	0.97324	7543643	Q15833
IINTFYSNKE	0.914457	6728290	Q90474
RTVSLGAGAKDE	0.999748	6603350	Q3T160
RALGQNPTNAE	0.822345	6297961	P60661
KVLQHMKAVQADQE	0.931291	5903078	Q5E9X4
YTGDYLNGTSE	0.573083	6542586	Q6GDU5
SVQPTVSLGGFE	0.976769	6548109	Q3T160
KNGRIISTGYN	0.901006	6698456	P00814
TNLVDILTGSYGN	0.356749	7278456	Q55FL5
TCVVCARPLE	0.598708	6467936	Q15942
ASHVKNLDE	0.0439281	5647715	P48734
KQAPLVH	0.942798	4547550	P04973
EILKIHARE	0.123188	4088973	D5HA94
HLIEADLRNFE	0.0705401	7228483	P18177

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
NLGTIAKSGTKAFME	0.30517	8644392	A6QBI2
TLLTNKHVVDATHGDPHALKAFPSAINQDNYPN	0.999508	9364717	Q2FZL2
KALGFDYLSLPINELPND	0.0649357	10680401	A7N6K1
CWGGGEYG	0.984287	4871626	Q6GQ48
AQAGAAGGAGGMGG	0.153115	5687266	B0R3H4
VTIMPKDIQLARRIRGE	0.691779	7047310	P84227
KVLKQVHPD	0.506344	6043472	P02281
QVAMNPTNTVFD	0.137326	7128131	P19120
KMSVQPTVSLGGFE	0.999967	7983942	Q3T160
HTVNGHNCE	0.999993	5782191	P09867
AVKNFTE	0.0503241	4627286	P04040
FAPVNVTTE	0.723314	5332499	P68103
HGLRGLRMLD	0.039481	6283237	Q3ZBI5
TSSQSYISTE	0.0667089	5957479	Q91DM0
LPSGGNRFAVNLQTGYNDSD	0.163571	7383359	Q3MHZ8
LGNAGAGGNM	0.114655	4831946	Q493W7
NGYGFINRNDTKE	0.614326	8223780	Q00436
LQQVFHTIGFR	0.536865	7173717	A0RHU0
DGTGGMIIDSGTTLT	0.400879	7638395	Q766C2
RFDSDAASPRE	0.0257605	6697911	P01889
QQPLMLPG	0.0748009	4942356	Q501R9
IINTFYSNKE	0.446144	6728290	Q90474
VTHAVVTVPA	0.995014	5412894	Q24798
FNHLSAVSE	0.219875	5462451	Q4R4I6
KTPPAANQASDPE	0.0931026	7213351	Q6W8Q3
LAVVVDDAAAGVTD	0.283976	7023400	Q5P2J9
APRAVFPSIVG	0.0456645	6013237	P53478
STLRTLPTLQE	0.0548812	6738531	P13489
KASDVHE	0.0837029	4512080	P14618
VTIMPKDIQLARRIRGE	0.24885	5288001	P84227
FVQITGKKPNFE	0.397449	7764156	Q6FI81
QFNGAVFINE	0.980766	6137794	Q6GI34
KSAADVGAPI	0.0379596	5227735	B1MIM7
KGNFTLPE	0.0774445	5112550	Q00839
LQKGIHPTIISE	0.362127	7264002	Q2T9X2
YVTIIDAPGHRD	0.727158	7228483	P68103
FFKNWRDPDQTD	0.992942	8428750	P06396
SGSNPYTTVTPQIINSKWE	0.459044	7466965	A5D7D1
AANALLLKVN	0.0750765	5718339	Q12VE5
TVLSGGTTMYPGIAD	0.0723955	7858602	P53478

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
YTGDYLNGTSE	0.616179	6542586	Q6GDU5
YGPPGTGKTLIARAVANE	0.962198	6443401	Q3ZBT1
VLGPKPALPAG	0.992744	5683310	Q3SX14
AEQITKYSGEGD	0.999935	7073140	Q2FZL2
HVAVKVLQD	0.981886	5628105	Q91550
HKLLPGMF	0.0328852	5377776	Q849Q9
MATAPGERL	0.130042	5252416	Q5NCC9
THLNLSGNKIKD	0.995267	4952695	P51122
ILFPRKPSAPKKGDSSAEE	0.150615	5580543	P26373
YPGIADRMQKE	0.0844091	7123391	P53478
QMLEAFAAMSVM	0.0731492	7168028	Q089R2
WSKTGPVAKE	0.0575422	6238288	Q4R4I6
SGGTTMYPGIADRMQKE	0.984824	6586305	P53478
FAPSGDKKNPVKFE	0.965977	5793079	P13667
FKSLDKDGTGQIQVNIQE	0.96999	7220368	P13135
TAFQKKGAE	0.373758	5622946	P30040
KGEAMQYD	0.999972	5292203	Q2FZL2
TLLTNKHVVDATHGD	0.999998	5796262	Q2FZL2
HTVNGHNCE	0.999993	5782191	P09867
KYNPPDHE	0.553975	5582451	Q15059
VGLGKKAAGIDE	0.662211	6513605	P28838
NARGLTSVINQKLKDDE	0.0822239	6823615	P07195
NVVTSMDVS	0.0384224	5282230	O97237
RIPIIEDISI	0.0326224	6288498	Q09797
LGGVQIMHYNGVSHITVPD	0.0306858	7090118	O00763
VGRPIIRSTTKVGNIE	0.45352	6193524	A7MB62
VKGKFQDNFE	0.982881	6783370	Q3ZBD9
AALLNATKNTDTYNAHD	0.988731	9749573	Q2FZL3
AALKKALAAGGYD	0.944691	6968816	P08284
RGLGMTLDLILYDGTLKVG	0.0737682	7177234	A3CSP4
FVNITPAE	0.0585667	4897339	P02584
STTGGNSGSPVFNEKNE	0.999999	9209047	Q2FZL2
LQTPKIVADKD	0.471389	6863814	P07195
HTVNGHNCE	0.999999	5782191	P09867
RFDSDAASPRE	0.0564597	6697911	P01889
RIVNEPTAAA	0.94167	5652871	Q835R7
KISKPAPYWE	0.186805	6818601	Q9BGI2
VAALSVGGAASQMAG	0.0756969	6973261	B1Y234
KASDVHE	0.115158	4512080	P14618
SNIPANLKD	0.0534949	5442764	Q86AE3

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
FVRNVADDTRSE	0.710839	7488437	O75494
NFLKQNIE	0.998735	5612868	Q2FZL2
ASKAQTQKHGITQANE	0.919629	6193160	A5D785
AQAGAAGGAGGMGG	0.519538	5687266	B0R3H4
VAMKFLRASEE	0.609558	6988520	P22392
YTGDYLNGTSE	0.907865	6542586	Q6GDU5
VAIVDPHIKVD	0.444244	6613631	Q14697
RARHAQGGTWYGVDINNE	0.893116	7113292	Q99832
VGKILGPQGNTIKRLQE	0.983074	6657190	Q8UUW7
VACTKVVAPTISSPVCQE	0.992435	6880056	P54939
THSLGGGTGSGMGTLLISKIREE	0.741323	8064104	Q17299
TGPVMPVRNVYKKE	0.908473	5879845	P26373
MGQKDSYVGDE	0.067584	6727762	P53478
ISLATGHREE	0.205523	6007877	P08133
IYAGKILNDD	0.844577	6193105	Q29RK4
AATGLHPGGKDGE	0.811967	6633116	Q15397
FQYTDKHGE	0.212546	6207690	Q9BGI2
TLLTNKHVVDATHGDPHA	0.999997	5112587	Q2FZL2
AGGIIGVKGAKIKE	0.906139	5049777	Q3T0D0
TGIPPAPRGVPQIE	0.0659445	7604007	P19120
LVGFTNKGTE	0.20141	5912974	A6QLP7
LVQPTKRPE	0.104349	5923290	Q3SZC0
RYVKAVPMGGGAD	0.0797745	7188551	A1WY07
FAQINQGE	0.229934	4977188	Q4R4I6
VTASQCQQPAE	0.631824	6537739	Q01518
ANGAVIHLALKE	0.949739	6763741	A2VDN6
AYRNTAQRQE	0.999684	6628070	P09525
HFHSPNPE	0.597866	5267166	P49327
KMAVIVMHKMMGLVMSGH	0.0773533	7366895	O24600
IASGVVVGKD	0.999984	5307892	Q2FZL2
FPVFLMFFISCLAE	0.311653	9129243	P26845
FNVINSHKQE	0.0371008	6663246	Q7YXU4
AGAPSGTPAAAPAATP	0.259049	6983325	A8IG20
ILVKHLVPQA	0.999842	6173731	Q27673
QAAMFGVLDFAPAQLPAD	0.0528898	9754606	Q1GR26
SQLGMSSSSGGGGSSGLHILQGE	0.39956	7410003	Q8L8A5
TTYSCVGVFQHGKVE	0.99603	9144240	P19120
GIGTVPVGRVE	0.279333	5863108	P68103
LAKAIMSIPAVT	0.143458	6658775	C3MZF2
TAEVADILQI	0.083801	5807971	C3PMA0

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TSATVGPKAPSGGKKATQASQEY	0.999526	8127529	P16104
AALLNATKNTD	0.584504	6243188	Q2FZL3
ISLKQAPLVH	0.999579	6113550	P04973
LTGKDVNFEFPE	0.119828	7563582	A6H769
KVLKQVHPD	0.837039	4032339	P02281
GYGFINRNDTKE	0.958365	7653565	Q00436
VCCSAKNLRDIDE	0.994165	8483788	P50990
ALLSLAKGDRSE	0.91621	6883663	P46193
RIVNEPTAAALAYGLDK	0.999608	6400084	Q835R7
YANTVLSGASTMFPGIAD	0.709382	9599319	P18499
QVFFFGTHE	0.0516807	6002633	Q9XSK7
AANTFTITGHAE	0.999052	6607983	P20290
AQSVSQNPNDPHLGHALAVVGNAKINDQE	0.580581	7856362	Q2FZL3
SGGTTMYPGIAD	0.91346	6292601	P53478
YLKSEPIPE	0.052001	5963021	P38657
LVEPISCHAWNKD	0.999052	8428949	Q58CQ2
TVSKVDDFLANE	0.51221	7273478	P52209
NYKPPPQKSLKE	0.999999	5342972	Q9TU03
LGGVQIMHYNGVSHITVPD	0.0292148	7090118	O00763
YKVPDGKPE	0.979764	5888023	Q9H2U2
AEGKAYECAGKG	0.806363	6928156	Q0TVB2
SLNIDPDAKVE	0.362639	6588240	Q95M18
LADQGQSWKEE	0.980397	7038167	P09211
STSLHQQISDD	0.923425	6597828	A6H712
TAACAQHCPE	0.99314	6167285	Q32LG3
LVVTDPRADHQPLTE	0.327392	8899409	P26452
LVKTPLVQE	0.107414	5718283	O46432
LGAGLTVGLCGL	0.0326224	6098149	P54642
HTINGHNCE	0.999526	5852270	P51968
SKYDPQKE	0.0946354	5697761	Q99439
GVTHTVPIYE	0.735203	6022896	P53478
RSPVGAGAVAVSGAVD	0.268895	7508776	A1A2J2
TLGRGTTITLVLKE	0.697636	8094660	P08110
SGRPLKVKEDPDGE	0.993064	5576189	P52272
SGLQKQASAAE	0.312006	6032954	Q03252
STTGGNSGSPVFNEKNE	0.999999	6142722	Q2FZL2
GPKFKMPE	0.0731492	5392774	Q09666
RGISTKPVYIPE	0.0412704	7384002	Q14258
IVDGRGKATISND	0.291556	7313721	Q99832
FHTTGLVWSKTGPVAKE	0.489133	5012672	Q4R4I6

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
AVEKPPQTE	0.243432	5577763	P23141
FQFTDKHGE	0.557466	6127715	Q06830
APRAVFPSIVG	0.0434996	6013237	P53478
MSNLDSNRDNE	0.955996	6917694	Q9TV56
TLLTNKHVVDATHGDPHALK	0.999999	5785612	Q2FZL2
FRAAVPSGASTGIYE	0.998813	8073852	Q9XSJ4
SYTATDRHPQALE	0.943313	7888569	P49327
FYSTFATDDRKE	0.990236	7983562	P30084
FRNGDTASPKE	0.975988	6693116	P07237
TLLTNKHVVDATHGDPHALK	0.999999	5785612	Q2FZL2
KVWLDPNETNE	0.352058	7308401	Q3T0W9
IPAMTIAKNAGVE	0.0908725	7158729	P10809
QQQQQNIDPND	0.245329	7082966	Q54WV4
QIGAKFWE	0.297157	5477708	Q9YHC3
FVRNLANTVTE	0.571233	6763376	O60506
LAVEIAGPHAQ	0.0643908	5973030	P0A9S4
TIGYKVPRNPE	0.951567	6953636	Q6PGB8
SQAMAVIKE	0.597406	5467752	P05198
GAGLQNAAQGLLDQLGDGAAVV	0.178789	7093565	Q7V419
IGLGFKTPRE	0.0851485	6173366	O65569
CILPPTRPTDKPLRLPLQD	0.964892	7827607	P68103
MVAYTFEGDQG	0.0475286	6532601	Q7URV2
NVAKAGLVDD	0.20322	5592816	P13620
IGGLAMACHD	0.841775	5667330	P31949
HVALRNRSNTPILVDGKDVMPE	0.038906	6450875	P06744
VGKILGPQGNTIKRLQE	0.924205	6657190	Q8UUW7
TLTGKTITLE	0.265674	5968285	P62992
IGIPPAPRGVPQIE	0.933092	7664189	Q3ZCH0
YVNKNSKHEE	0.999999	4645591	P07814
LRKGNYAE	0.907744	5337713	Q96QV6
FNHLSAVSE	0.70519	5462451	Q4R4I6
KSTLPDADRE	0.170219	6243007	A5D7D1
LTGKDVNFE	0.673062	5697763	A6H769
RIINEPTAAALA	0.430541	6643558	Q04VC8
MNVLHIDSGI	0.141752	6017811	C3MAW0
SQHQARIE	0.866736	5287484	Q0VCX2
LGGVTAGIHVTIAAPD	0.498452	7904111	A0QHB0
IQAVLLPKKTE	0.131763	6924178	Q96QV6
GILKVPAINVND	0.992832	6848816	Q3MHL4
QHAEPGNAQSVLE	0.998263	7343305	P21964

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
GAILTPNDGRCHLE	0.0736994	8208798	P32592
VCDVPTAKIISRE	0.986454	8024130	P31939
RLLLPTPTVKQE	0.995582	7559369	Q8IX12
VYVGNLGNNGNKTE	0.990929	7978803	Q3SZR8
FVGNLPHDIDE	0.13136	6723007	Q9UN86
AVVTVPAYFND	0.972068	6423027	P19120
LPGELAKHAVSE	0.950918	6838556	P02281
VVQFTVKHE	0.110596	6018157	Q4VIT5
ALPGQLKPFE	0.633239	6083259	P09211
VKGHAYSVTGAEE	0.662817	7323456	Q27971
TLLTNKHVVDATHGDPHAL	0.999999	5395297	Q2FZL2
VYVGNLGNNGNKTE	0.98265	7978803	Q3SZR8
QVTQPTVGMNFKTPRGPV	0.999962	10370348	P0CW22
GLTLGGQKCSVIRD	0.38738	8104161	P02584
RILNTAAPLLSLQ	0.0379596	7494267	B2DBF0
STSLHQQISDD	0.962149	6597828	A6H712
ATKNTDTYNAHD	0.985942	7338146	Q2FZL3
NFSHGTHE	0.766813	5086984	P14618
LADQGQSWKE	0.285748	6392954	P09211
YASGRTTGIVMDSGDGVTHTVPIYE	0.998446	9054195	P53478
NTLVVKKSDVE	0.239648	6883789	O60812
HVVDATHGD	0.999999	5197193	Q2FZL2
KAFNGSRMTYEAPE	0.755716	8588898	Q0W8A0
VLNMYKLIT	0.100656	6138300	Q96254
YRTVSGVAGPL	0.964282	6043108	P11574
AQKQKVVE	0.0410051	5373050	Q5XH03
TRKYTLPPGVDPTQVSSSLSPE	0.995332	8257528	P04792
KIGNFSTDIKD	0.656905	6913554	P13796
YMKARPFPDGLAE	0.993781	8058891	P13639
SADTLWDIQKD	0.203988	7043266	P07195
RALTGHTGLETGD	0.675012	7083331	B2VW14
KSTALAAIKPILEN	0.0730804	8069710	P43741
KSLAPFHPGITE	0.302868	7068661	Q9CYC5
LFLPKSVSDYD	0.989435	7003445	P07237
HTINGHNAE	0.999968	5407302	P51989
SKKGVNLPGAAVDLPAVSE	0.997598	6660312	P14618
IGIPPAPRGVPQIE	0.930487	7664189	Q3ZCH0
HGHLCPIDTGLIE	0.952025	7753607	P26358
GPKPALKE	0.98644	4922835	P40121
HRIIPGFMCQGGD	0.99997	7883471	Q6DTV9

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TIAKSGTSE	0.99254	5052474	Q95M18
GKVPTISINKTD	0.998214	7088922	Q4R4I6
AAATKKPAPE	0.677983	5643103	P36578
KLGDVGMAE	0.0449709	5182463	P13489
MKGNFTLPE	0.977683	5767752	Q00839
NQILRIEEE	0.202538	6163032	Q9XSJ4
LASLGKNPTDE	0.615371	6308109	P19105
FVKMLPDKD	0.908069	6188222	Q5E9F7
FITHAPPGEFNE	0.985224	7238218	A4FUA8
AVDKSTGKE	0.920106	5397763	P19120
FRNGDTASPKE	0.998025	6693116	P07237
GVVVPPAVAGARPE	0.988769	7038769	O95278
IIPASTGAAKAVGKVIPE	0.248453	9335423	P04406
KIFGVTTLD	0.0737682	5552994	Q32LG3
YHAFSGAQKAE	0.96796	6628034	Q3B7N2
VTRKDGNASGTTLLE	0.418477	8394276	P68103
FIIRGSPQQIDHAKQLIE	0.999972	7370649	Q92945
NLPAVRWVGGPE	0.53259	6918481	P48643
VTHAVVTVPAYFN	0.999401	7533766	Q24798
TGSDLLWTQCA	0.670961	6702866	Q6XBF8
IGIPPAPRGVPQIE	0.93567	7664189	Q3ZCH0
RLVNGLPSLNVD	0.0595622	6928666	P64880
TIFVQGLGE	0.759012	5262603	Q28009
LFLPKSVSD	0.142803	5612994	P07237
RMVNHFIAE	0.9093	6027839	P19120
FVRLGSLSE	0.867679	5482790	O60664
NLGTIAKSGTKA	0.963635	6528660	A6QBI2
RAAVPSGASTGIYE	0.999976	7338510	Q9XSJ4
LGITTKISRGTIE	0.969217	7529240	Q8NHW5
VVVPPAVAGARPE	0.989432	6753661	O95278
IVNTNVPRASVPD	0.962397	7353747	P14174
WVNTPKKQGGLGPMNIPLVSDPKRTIAQD	0.982231	8339520	Q06830
KNLPYKVTQDE	0.90738	7398818	P19338
LYGPPGTGKTLIARAVANE	0.997762	6820348	Q3ZBT1
MIKPFFHSLSE	0.999842	7263569	O97680
FQVRANSAGATRAVE	0.999922	5556107	P06396
SGGTTMYPGIADRMQKE	0.990676	9794447	P53478
LAKQAFDDAIAE	0.721873	7043451	P48347
IINSLYKNKE	0.745034	6833762	Q95M18
WSKTGPVAKE	0.10788	6238288	Q4R4I6

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
TLLTNKHVVDATHGD	0.999999	5796262	Q2FZL2
NLGTIAKSGTK	0.0960147	6173474	A6QBI2
LQANRDPDAGIDE	0.981145	7513332	P08758
RVVDSPCCLVT	0.913146	6973176	P55737
IQAVLLPKKTE	0.328091	6924178	Q96QV6
RCASIQKFGE	0.795501	6563130	P02769
RTVSLGAGAKDE	0.999812	6603350	Q3T160
APKPGPYVKE	0.115327	6153338	Q4R4I6
RLLSKYDPQKE	0.999844	5076097	Q99439
QSGLVFNPLADR	0.862397	7028506	Q2J419
RFQSSAVMALQE	0.999498	7358396	P84227
LGPPGAGKGTQAPKLAE	0.997834	5793188	Q1L8L9
THSLGGGTGSGMGTLLISKIRE	0.999703	7687279	Q17299
AILGGAKVAD	0.154429	5157839	P50310
TYCFNKPEDK	0.999696	7233258	P62992
FAVTVAPPGARQGQQQAGGDGKTE	0.997688	8294081	Q8VEK3
VTHAVVTVPAYFND	0.999148	8108902	Q0VCX2
KGKDTSCLAE	0.845412	6267994	Q9BT09
YAGKDGYNYTLSKTE	0.999999	6186258	P31949
RCIPALDSLTPANE	0.961257	8228896	P10809
RASYNNPHRAVE	0.996796	5012371	A3KMV2
SVQTFADKSKQE	0.995822	7563743	P79136
ASDALDKIRYE	0.422868	6988427	Q76LV1
KHTGPNSPDTAND	0.999999	7353201	P31943
HTINGHNAE	0.999968	5407302	P51989
FVGGLKGDVAE	0.987728	6043052	Q13151
RFQSAAIGALQEASE	0.999622	5559352	Q10453
FTRKYTLPPGVDPTQVSSSLSPE	0.999582	8747756	P04792
QLKNKYQGQDE	0.999052	7478666	Q5JTV8
GITLPVDFQGRSTGE	0.994241	8329013	P31943
TIFVQGLGE	0.812889	5262603	Q28009
MAQYNRHDSPE	0.954986	7182903	Q3ZBZ8
SLACQGKYTPSGQAGAAASE	0.999914	6906472	P04075
RLVNGLPSLNVD	0.0799508	6928666	P64880
SSGKFYGDEE	0.823668	6177505	Q4VIT5
TRKDGNASGTTLLE	0.86406	7898934	P68103
TIGADFLTKE	0.265771	6058050	Q3T0F5
LQRFIDPLAKEEE	0.997631	8524375	P33991
VIGPWIQTKMEE	0.902152	7738860	Q3B7N2
QCAILSPAFKVRE	0.155516	8179236	P34932

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
IVVVTAGVRQQE	0.222751	6938743	P07195
SNFKTAAE	0.191557	4922290	P07237
GGFTAEQITKYSGE	0.999999	8023693	Q2FZL2
GVVVPPAVAGARPE	0.992543	7038769	O95278
RQFMNKSLSGPGQ	0.999998	7833820	Q3T035
RILNNGHAFNVE	0.99055	7363594	P00918
LFLPKSVSDYD	0.99931	7003445	P07237
SVKLGHPDTLNQGE	0.999684	8058960	P06702
APLAKVIHD	0.895327	5402997	P04406
AVQSKYGDFNKE	0.999747	7653691	Q2HJ49
RIVNEPTAAALAYG	0.999999	7673903	Q835R7
LLKVNQIGSVTE	0.995264	7088922	Q9XSJ4
RTWNDPSVQQD	0.999767	7173095	Q0VCX2
YQAMAKDQAVE	0.998743	6853101	Q2KJH6
VAKNKDQGTYE	0.999512	6988427	P60661
TLKGGAAVDPDSGLE	0.992749	7733771	P18493
RATDFVVPGPGKVE	0.991021	7944136	Q9XSG3
NVAKAGLVDDFE	0.995558	6973372	P13620
GVRVVNCARGGIVDE	0.999901	8449142	O43175
FVNVVPTFGKKKGPNANS	0.996843	6927153	P62866
AALKKALAAGGYDVE	0.999999	8109371	P08284
FSPNEQNKHIGE	0.999989	7583487	Q2FZL2
VIITAGARQQE	0.922908	6373323	Q9PW06
RLVNGLPSLNVD	0.12314	6928666	P64880
TNKHVVDATHGD	0.999999	4705542	Q2FZL2
TLLTNKHVVD	0.999997	6283396	Q2FZL2
AVVGNAKINDQE	0.491513	6873403	Q2FZL3
KYLPDIIKD	0.997	6248492	P79126
IHPQTIIAGWRE	0.999526	7548878	Q3ZBH0
AVVGNAKINDQE	0.854989	6873403	Q2FZL3
RITPSYVAFTPE	0.989354	7348609	Q0VCX2
NLGTIAKSGTKAF	0.195162	7264001	A6QBI2
GLINFIKQQRE	0.92404	7313980	P06744
FLRAPGARTPVIVR	0.994794	8209746	P55308
SPNEQNKHIGE	0.999996	6848145	Q2FZL2
WLKKRTGPAATTLPDGAAAE	0.984774	7333929	P07237
SSDALDKIRYE	0.99965	7068401	Q90474
FSHPRDFTPVCTTE	0.999418	8913849	O17433
KVNQIGSVTE	0.982274	5958081	Q9XSJ4
QGLTRQMRLRV	0.115565	7233952	P47874

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
NLQNKQSLTMDPVVKSKE	0.999998	7443974	P34932
RALIAGGGAPE	0.303045	5502821	Q2T9X2
RAAVPSGASTGIYE	0.999978	7338510	Q9XSJ4
CILPPTRPTDKPLRLPLQD	0.998278	7827607	P68103
TVAKNKDQGTYE	0.999999	7493665	P60661
FAVTVAPPGARQGQQQAGGDGKTE	0.99694	8294081	Q8VEK3
RFQSSAVMALQEASE	0.998991	8713980	P84227
FDPANGKFSKSATSPE	0.9961	9139353	P13639
TGSDLLWTQCA	0.784943	6702866	Q6XBF8
VGVNLPQKAGGFLMKKE	0.996914	6633740	P50310
FQKIPTEAPQLE	0.179714	7588896	Q5VW32
TSRGPGTSFE	0.979622	5637455	Q5E946
STMKPVQKVLE	0.884822	7023857	Q0VCX2
TIVCNSKDGGAWGTE	0.999801	8558769	P09382
FLDPSGKVHPE	0.990393	6713292	O95881
SAASCEPLASVLRAKPDFKE	0.981584	7740608	P13489
YMKARPFPDGLAE	0.999641	8058891	P13639
RQQLQTTRQE	0.998736	6883412	Q08E38
NLGTIAKSGTKAFM	0.966312	7919204	A6QBI2
ATHGDPHALK	0.999999	5817876	Q2FZL2
FSPLNPMRVHIE	0.999999	7643762	Q3SZF3
GVRVVNCARGGIVD	0.998906	7803929	O43175
VSLAVCKAGAVE	0.964659	6603387	Q9XSJ4
IASGVVVGKD	0.999679	5307892	Q2FZL2
IVVVTAGVRQQE	0.136173	6938743	P07195
KMSVQPTVSLGGFE	0.997937	8063917	Q3T160
QGLTRQMRLRV	0.0829372	7233952	P47874
FNQGKIFK	0.997316	5633101	P26642
GVVVPPAVAGARPE	0.999591	7038769	O95278
VRGLPWSCSADE	0.999999	7328161	P31943
KNQIGDKE	0.22438	5382764	Q0VCX2
AAVAAGAVAIV	0.0543698	5007786	Q89FU2
RIVNEPTAAALAYGLDK	0.999999	6400084	Q835R7
VASNLNLKPGE	0.991751	6293292	P09382
IVNTNVPRASVPD	0.992409	7353747	P14174
VEAPTGTFIASGVVVGKD	0.999999	9319823	Q2FZL2
NVNAKPFVPNVHAAE	0.999999	8619354	Q8R050
GAMAKPDCIITCDGKNLTIKTE	0.999998	8701001	Q01469
IINSLYKNKE	0.731047	6833762	Q95M18
AALLNATKNTD	0.988421	6243188	Q2FZL3

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
SAYIKNSRPE	0.99842	6408190	Q9XSA7
FPASADRTVIDYNGE	0.999999	8718884	P07237
AWSRTGPVAKE	0.177818	6593346	Q3SYV4
RICKVLAVNQE	0.969875	7233840	Q3B7N2
QVTQPTVGMNFKTPRGPV	0.999991	6916923	P0CW22
RQAFQGDSIPVFD	0.987102	7843643	Q2TBQ8
WESKGKITYLKGEAMQYD	0.999999	7737182	Q2FZL2
FLQTPKIVADKD	0.999945	7599157	P07195
CLAPLAKVIHD	0.999842	6768571	P04406
MVTPGHACTQKFSHE	0.999999	6159410	P04075
RCILPFDKE	0.552318	6473203	Q9GZT3
AVAISLPKGVVE	0.977554	6498733	Q2KJH6
SAALIQQATTVKNKD	0.858982	8664691	Q3SYR7
TRKDGNASGTTLLE	0.958126	7898934	P68103
IVKFSPNEQNKHIGE	0.999999	6286611	Q2FZL2
GPKFLKSGDAAIVD	0.895187	7814186	P68103
APTGTFIASGVVVGKD	0.999999	8179267	Q2FZL2
AHAAIRENPVYE	0.999052	7293459	Q58DW5
RITPSYVAFTPE	0.99941	7348609	Q0VCX2
MAPKPGPYVKE	0.977045	6808540	Q4R4I6
GWGVMVSHRSGETE	0.9943	8103509	Q9XSJ4
IVLSQHLGKPPE	0.99637	7173951	Q60445
KIGGIGTVPVGRVE	0.16647	7494267	P68103
AEQITKYSGEGD	0.999996	7073140	Q2FZL2
GLINFIKQQRE	0.936575	7313980	P06744
MHGGTGFAGIDSSSPE	0.994988	8193321	Q5E9A3
ILGYINTGKQE	0.960133	6763501	P05091
YAAQAHLKLGE	0.999684	6588372	Q2T9P4
YARIGNSYFKE	0.99944	7323532	Q3ZBZ8
FQVRANSAGATRAVE	0.999999	8329125	P06396
QVKQKGADFLVTE	0.999641	8039293	P14618
FNRYPALHKPE	0.99584	4965871	P13796
ITYTDEEPVKKLLE	0.999842	9114816	Q5ZI72
FAVTVAPPGARQGQQQAGGKKKAE	0.99941	6498563	Q00839
KGVTIASGGVLPNIHPE	0.999998	6023258	O93327
YSILGTTLKDE	0.997631	6783420	Q2KJH4
RTALINSTGEE	0.991649	6398036	P02545
NLGTIAKSGTKAFME	0.998449	8564417	A6QBI2
VVLLQANRDPDAGIDE	0.999887	9069437	P08758
VASGGTAKALRD	0.751856	6313323	P31939

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
GGSDNFKHLNE	0.999842	6672959	Q3ZBF7
LLKVNQIGSVTE	0.999495	7088922	Q9XSJ4
QFRNEGIDLTHNPE	0.983974	8793996	Q8SS56
CLAPLAKVIHD	0.999901	6768571	P04406
AKNLPYKVTQDE	0.999741	7754004	P19338
VGMGQKDSYVGDE	0.99619	7508211	P53478
IGLGFKTPREAIE	0.414802	7739187	O65569
VQSGSHLAARE	0.999033	6217986	P04040
THSLGGGTGSGMGTLLISKIRE	0.999999	7633962	Q17299
KVVKQASEGPLKGILGYTE	0.999999	7304133	P04406
NKHVVDATHGD	0.999999	4368717	Q2FZL2
AALGGNSSPSAKD	0.994165	6458036	P42899
ARQIPQATASMKD	0.999999	7668818	O43175
SKGGVVGIKVD	0.922699	6018445	P04075
FNTISQGDVRLTPE	0.999842	8329013	Q9NP79
SAYIKNSNPALNDNLE	0.999998	9399489	Q5E9B7
VLGPKPALPAGTE	0.999999	6833762	P06396
GVKQLIVGVNKMD	0.999302	7729308	P68103
STTGGNSGSPVFNEKNE	0.999999	9209047	Q2FZL2
NALTSSIAIGALSVDAA	0.0782519	8314246	Q15X40
KVLKQVHPD	0.850432	6043472	P02281
VGNLNFNKSAPE	0.968106	7033429	P19338
FSPLNPVRVHIE	0.999999	7483901	P31943
YNRVAVQGDVVRE	0.999861	7968963	P07814
SKDGGAWGTEQRE	0.994928	7688411	P09382
GLRSVQTFADKSKQE	0.999999	6133209	P79136
YTGINAISKKE	0.999842	6843658	P61157
FSPLNPVRVHIE	0.99998	7483901	P31943
RTVSLGAGAKDE	0.999984	6603350	Q3T160
SLSTKLKQVEDE	0.968906	7608977	P35579
GSVIQKALGGE	0.963513	5878107	P12763
VEAPTGTFIASGVVVGKD	0.999999	9319823	Q2FZL2
AGKDPVQCSRD	0.999721	6748030	Q5E946
LVWVPSDKSGFEPASLKE	0.999999	7117029	P35579
ALAKIYIDSNNNPE	0.997789	8394115	P49951
IYAGKILNDDTALKE	0.999999	6033220	Q29RK4
KLSDLLAPISE	0.97933	6513549	Q4R4I6
AWSRTGPVAKE	0.206428	6593346	Q3SYV4
STTGGNSGSPVFNEKNE	0.999999	9209047	Q2FZL2
HRIIPGFMCQGGD	0.999999	7883471	Q6DTV9

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
KPVATMWESKGKITYLKG	0.999999	5600665	Q2FZL2
IVVLRNPLIAGK	0.999998	7049393	Q3SZF8
RIINEPTAAALAYG	0.999435	7743982	Q04VC8
FIASGVVVGKD	0.999999	6043234	Q2FZL2
LAIVKFSPNEQNKHIGEVVKPATMSN	0.999999	7566570	Q2FZL2
LQRAPQCLGKFIE	0.999999	8384368	Q00839
TVLSGGTTMYPGIADRMQKE	0.999922	7576989	P53478
FLQTPKIVADKD	0.999973	7599157	P07195
GSISCKSLKGGNIS	0.999998	7763973	P54577
TLLTNKHVVDATHGDPHALK	0.999999	5785612	Q2FZL2
TIVCNSKDGGAWGTEQRE	0.999999	7086545	P09382
FINENVRNFLKQNIE	0.999998	6653438	Q2FZL2
GVSLAVCKAGAVE	0.952394	6888495	Q9XSJ4
ACQGKYTPSGQAGAAASE	0.999999	9354091	P04075
ALSADVLGRCQVFE	0.999866	8268924	Q99832
KIGNFSTDIKD	0.999748	6913554	P13796
FIIRGSPQQIDHAKQLIE	0.999999	7370649	Q92945
FVGNLNFNKSAPE	0.999741	7768771	P19338
LAKNLPYKVTQDE	0.999999	5549640	P19338
RFQSSAVMALQE	0.999999	7278422	P84227
GILNVSAVDKSTGKE	0.999844	8314428	P19120
SPAVIVGLLKE	0.990565	6213625	O43175
TGMAFRVPTANVSVVD	0.999903	8764266	P04406
KIFVGGIPHNCGE	0.996749	7723736	Q96EP5
NVITVGPRGPLLVQD	0.999862	8334535	P04040
KMSVQPTVSLGGFE	0.999999	7983942	Q3T160
IYAGKILSDDVPIRD	0.999999	8959717	A3KMV2
TRKYTLPPGVDPTQVSSSLSPE	0.999999	8257528	P04792
NLIGCIIGRQGANINE	0.999861	9154537	Q5E9A3
FIIRGSPQQIDHAKQLIEE	0.999999	7800791	Q92945
LQNNLPAVRWVGGPE	0.999999	8694409	P48643
DLIGCIIGRQGAKINE	0.999684	9369873	Q15366
INRNTTIPTKKSQVFSTAADGQTQVE	0.999999	9935134	Q3ZCH0
NVSAVDKSTGKE	0.999689	6898480	P19120
FRAAVPSGASTGIYE	0.999999	8073852	Q9XSJ4
FLRAPGVQTPVIVR	0.999998	8209691	P48350
RTVPPAVTGITFLSGGQSE	0.999999	10030068	P04075
RAAVPSGASTGIYE	0.999999	7338510	Q9XSJ4
RIVNEPTAAALAYGLDK	0.999999	6400084	Q835R7
NVNAKPFVPNVHAAE	0.999999	8619354	Q8R050

Supplementary Table S2 - rRLR1 cleavage sites identified from a GluC peptide library using Mascot and X!Tandem.

Identified Peptides (prime sequence)	PeptideProphet Probability	Neutral Peptide Mass (Da)	Examplary protein ID
RFQSAAIGALQE	0.999874	6898430	Q10453
ATHGDPHALKAFPSAINQDNYPN	0.999999	8654081	Q2FZL2
AATLLANHSLRE	0.999451	6923563	P13489
NSASAIGCHVVNIGAEDLKE	0.999999	7340174	P13796
THSLGGGTGSGMGTLLISKIRE	0.999999	7633962	Q17299
VGGTTPGKGGQTHLGLPVFNTVKE	0.999998	8467845	P53597
LGPPGAGKGTQAPKLAE	0.999989	8684746	Q1L8L9
RVINEPTAAALAYGLDKSE	0.999999	7120333	Q3ZCH0
RVDKAAAAAAALQAKSDE	0.999999	6440071	P36578
IYAGKILNDDTALKE	0.999999	9044793	Q29RK4
RVINEPTAAALAYGLDKSE	0.999993	10675461	Q3ZCH0

Supplementary Table S3 – Significantly de-regulated proteins described in *RLR1* KO line grown under normal conditions by quantitative shotgun proteomics. In the table are depicted the gene ID number according to "The Arabidopsis Information Resource database (www.arabidopsis.org), protein description (according to the same database), and the protein ratios (KO *vs* WT) converted in log2 values of the different biological replicates.

Gene ID	Protein description	Ratio values (log2)		
		Replicate 1	Replicate 2	Replicate 3
AT5G56670	Ribosomal protein S30 family protein	-1.87684	-1.22041	-4.17163
AT2G01520	MLP-like protein 328	-0.60625	-0.5892	-0.5999
AT1G64200	vacuolar H+-ATPase subunit E	-0.28701	-0.34037	-0.52772
ATCG00490	ribulose-bisphosphate carboxylases	-0.38896	-0.33506	-0.46715
AT5G59090	subtilase 4.12	-0.19873	-0.24364	-0.36342
AT5G23820	MD-2-related lipid recognition domain-containing protein	-0.2917	-0.52514	-0.34075
AT4G11290	Peroxidase superfamily protein	-0.25427	-0.21342	-0.33515
AT3G11930	Adenine nucleotide alpha hydrolases-like superfamily protein	-0.6419	-0.61752	-0.29097
AT1G54010	GDSL-like Lipase/Acylhydrolase superfamily protein	-0.33526	-0.52027	-0.26802
AT5G10010	unknown protein	-0.22624	-0.32608	-0.26446
AT4G08770	Peroxidase superfamily protein	-0.30656	-0.42357	-0.26273
AT4G11260	phosphatase-related	-0.31292	-0.50672	-0.23464
AT1G50480	10-formyltetrahydrofolate synthetase	-0.24843	-0.30993	-0.2225
AT5G40390	Raffinose synthase family protein	-0.57844	-0.22159	-0.21143
AT1G70850	MLP-like protein 34;	-0.20284	-0.28638	-0.19879
AT1G50670	OTU-like cysteine protease family protein	-0.55595	-0.34975	-0.16379
AT1G12230	Aldolase superfamily protein;	-0.25577	-0.2353	-0.13903
AT1G54030	GDSL-like Lipase/Acylhydrolase superfamily protein	-0.39536	-0.26214	-0.12888
AT5G23830	MD-2-related lipid recognition domain-containing protein	0.219463	0.325386	0.123799
AT3G44310	nitrilase 1	0.245009	0.263395	0.131458
AT3G51160	NAD(P)-binding Rossmann-fold superfamily protein	0.288772	0.211137	0.135141
AT2G02390	glutathione S-transferase zeta 1	0.420294	0.280481	0.135798
AT1G23440	Peptidase C15, pyroglutamyl peptidase I-like	0.532866	0.140909	0.152118
AT3G02880	Leucine-rich repeat protein kinase family protein	0.215865	0.127501	0.158337
AT3G17940	Galactose mutarotase-like superfamily protein	0.267236	0.222434	0.166073
AT3G44300	nitrilase 2	0.206518	0.319155	0.174279
AT1G27090	glycine-rich protein	0.183328	0.188274	0.187514
AT5G48880	peroxisomal 3-keto-acyl-CoA thiolase 2	0.182438	0.301354	0.202888

Supplementary Table S3 (Cont.) – Significantly de-regulated proteins described in *RLR1* KO line grown under normal conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log2)		
		Replicate 1	Replicate 2	Replicate 3
AT4G25890	60S acidic ribosomal protein family	0.532267	0.336626	0.211137
AT5G13930	Chalcone and stilbene synthase family protein	0.607674	0.276675	0.25181
AT3G07720	Galactose oxidase/kelch repeat superfamily protein	0.167743	0.222557	0.269273
AT3G05970	long-chain acyl-CoA synthetase 6	0.35242	0.198243	0.286053
AT4G13850	glycine-rich RNA-binding protein 2;	0.193709	0.300065	0.29337
AT2G43610	Chitinase family protein	0.372729	0.757706	0.336398
AT3G58500	protein phosphatase 2A-3	0.245861	0.103397	0.357946
AT5G05270	Chalcone-flavanone isomerase family protein	0.389897	0.418568	0.760434
AT1G78830	Curculin-like (mannose-binding) lectin family protein	0.643579	0.365804	2.27226

Supplementary Table S4 – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics. In the table are depicted the gene ID number according to "The Arabidopsis Information Resource database (www.arabidopsis.org), protein description (according to the same database), and the protein ratios (KO *vs* WT) converted in log2 values of the different biological replicates.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT3G12260	LYR family of Fe/S cluster biogenesis protein	-1.14947	-0.20884
AT3G13235	ubiquitin family protein	-1.05547	-0.17365
AT1G09080	Heat shock protein 70 (Hsp 70) family protein	-1.00651	-1.16088
AT4G02520	glutathione S-transferase PHI 2	-0.98919	-0.42171
AT5G02490	Heat shock protein 70 (Hsp 70) family protein	-0.98582	-0.55011
AT1G14980	chaperonin 10	-0.9143	-0.20214
AT4G08770	Peroxidase superfamily protein	-0.59364	-0.19643
AT2G01520	MLP-like protein 328	-0.59075	-0.7337
AT2G47170	Ras-related small GTP-binding family protein	-0.58068	-0.20996
AT5G43780	Pseudouridine synthase/archaeosine transglycosylase-like family protein	-0.56785	-0.26154
AT2G27720	60S acidic ribosomal protein family	-0.53525	-0.47756
AT5G23020	2-isopropylmalate synthase 2	-0.51303	-0.91197
AT3G54540	general control non-repressible 4	-0.48855	-0.92169
AT3G52500	Eukaryotic aspartyl protease family protein	-0.45585	-0.18456
AT2G30620	winged-helix DNA-binding transcription factor family protein	-0.44737	-0.22661
AT5G40370	Glutaredoxin family protein	-0.44252	-0.52085
AT1G09270	importin alpha isoform 4	-0.44162	-0.36729
AT1G70830	MLP-like protein 28	-0.43717	-0.58875
AT1G19570	dehydroascorbate reductase	-0.41812	-0.37447
AT2G27530	Ribosomal protein L1p/L10e family	-0.40377	-0.37643
AT2G15430	DNA-directed RNA polymerase family protein	-0.39811	-0.42121
AT1G20010	tubulin beta-5 chain	-0.39295	-0.81406
AT1G28290	arabinogalactan protein 31	-0.39286	-0.36967
AT2G39310	jacalin-related lectin 22	-0.36989	-0.30082
AT2G09990	Ribosomal protein S5 domain 2-like superfamily protein	-0.36294	-0.17461
AT5G12140	cystatin-1	-0.35918	-0.38063
AT4G25630	fibrillarin 2	-0.35888	-0.32411
AT1G77330	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	-0.35268	-0.24741
AT3G04920	Ribosomal protein S24e family protein	-0.32511	-0.23194
AT3G54640	tryptophan synthase alpha chain	-0.32344	-0.18438
AT1G72160	Sec14p-like phosphatidylinositol transfer family protein	-0.31779	-0.22046
AT5G61170	Ribosomal protein S19e family protein	-0.31723	-0.18672
AT1G21440	Phosphoenolpyruvate carboxylase family protein	-0.30185	-0.44103
AT1G27400	Ribosomal protein L22p/L17e family protein	-0.29541	-0.18813
AT1G54010	GDSL-like Lipase/Acylhydrolase superfamily protein	-0.28974	-0.24001
AT3G51800	metallopeptidase M24 family protein	-0.28754	-0.27976

Supplementary Table S4 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT3G13300	Transducin/WD40 repeat-like superfamily protein	-0.28023	-0.21058
AT1G70850	MLP-like protein 34;MLP-like protein 34	-0.27249	-0.50083
AT2G31570	glutathione peroxidase 2	-0.27204	-0.36181
AT4G18100	Ribosomal protein L32e	-0.27023	-0.27787
AT5G03740	histone deacetylase 2C	-0.26484	-0.31554
AT2G32520	alpha/beta-Hydrolases superfamily protein	-0.25901	-0.19878
AT3G55610	delta 1-pyrroline-5-carboxylate synthase 2	-0.25205	-0.2313
AT1G17170	glutathione S-transferase TAU 24	-0.25018	-0.34879
AT1G70890	MLP-like protein 43	-0.24079	-0.61686
AT5G57330	Galactose mutarotase-like superfamily protein	-0.24055	-0.29717
AT3G04400	Ribosomal protein L14p/L23e family protein	-0.23496	-0.21862
AT1G24510	TCP-1/cpn60 chaperonin family protein	-0.22245	-0.19745
AT1G77510	PDI-like 1-2	-0.21846	-0.48892
AT3G01280	voltage dependent anion channel 1	-0.21402	-0.34467
AT3G11250	Ribosomal protein L10 family protein	-0.21366	-0.21195
AT5G54160	O-methyltransferase 1	-0.21233	-0.19995
AT5G67500	voltage dependent anion channel 2	-0.20788	-0.17239
AT4G13430	isopropyl malate isomerase large subunit 1	-0.20595	-0.26795
AT5G18380	Ribosomal protein S5 domain 2-like superfamily protein	-0.20541	-0.19944
AT4G39200	Ribosomal protein S25 family protein	-0.20271	-0.19349
AT1G50480	10-formyltetrahydrofolate synthetase	-0.19031	-0.1723
AT3G56490	HIS triad family protein 3	-0.18495	-0.26245
AT2G44100	guanosine nucleotide diphosphate dissociation inhibitor 1	-0.18405	-0.21508
AT3G25230	rotamase FKBP 1	-0.18318	-0.20548
AT1G64520	regulatory particle non-ATPase 12A	0.137241	0.281669
AT4G39080	vacuolar proton ATPase A3	0.138028	0.574538
AT4G14030	selenium-binding protein 1	0.13947	0.131853
AT3G15950	DNA topoisomerase-related	0.141302	0.124857
AT4G32520	serine hydroxymethyltransferase 3	0.143132	0.217975
AT5G58290	regulatory particle triple-A ATPase 3	0.145612	0.156008
AT4G16260	Glycosyl hydrolase superfamily protein	0.146394	0.811307
AT5G20400	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	0.148088	0.936892
AT1G53750	regulatory particle triple-A 1A	0.15004	0.202888
AT5G58710	rotamase CYP 7	0.151209	0.146394
AT1G47260	gamma carbonic anhydrase 2	0.152118	0.797096
AT1G79010	Alpha-helical ferredoxin	0.154583	0.230941
AT1G80360	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein	0.161049	0.146394
AT5G42740	Sugar isomerase (SIS) family protein	0.16543	1.02588
AT2G27510	ferredoxin 3	0.167615	0.203389

Supplementary Table S4 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT4G34138	UDP-glucosyl transferase 73B1	0.179384	0.178619
AT5G44020	HAD superfamily, subfamily IIIB acid phosphatase	0.182057	0.132774
AT2G02390	glutathione S-transferase zeta 1	0.182947	0.406537
AT5G44110	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.182947	0.112366
AT1G03220	Eukaryotic aspartyl protease family protein	0.184598	0.195474
AT3G16410	nitrile specifier protein 4	0.19131	0.389126
AT4G37070	Acyl transferase/acyl hydrolase/lysophospholipase superfamily protein	0.193204	0.171207
AT3G14990	Class I glutamine amidotransferase-like superfamily protein	0.196859	0.222557
AT3G01910	sulfite oxidase	0.196985	0.204642
AT3G54470	uridine 5-monophosphate synthase / UMP synthase (PYRE-F) (UMPS)	0.203139	0.141433
AT5G26280	TRAF-like family protein;TRAF-like family protein	0.209017	0.122739
AT3G44310	nitrilase 1	0.213503	0.207268
AT5G34850	purple acid phosphatase 26	0.214995	0.188021
AT3G57020	Calcium-dependent phosphotriesterase superfamily protein	0.218595	0.200881
AT1G32440	plastidial pyruvate kinase 3	0.219339	0.38659
AT2G45960	plasma membrane intrinsic protein 1B	0.219587	0.179638
AT3G09260	Glycosyl hydrolase superfamily protein	0.219711	0.125783
AT5G23860	tubulin beta 8	0.219711	0.163628
AT5G17820	Peroxidase superfamily protein	0.221444	0.263876
AT1G59359	Ribosomal protein S5 family protein	0.232661	0.285935
AT1G62660	Glycosyl hydrolases family 32 protein	0.247928	0.122474
AT3G55410	2-oxoglutarate dehydrogenase, E1 component	0.253747	0.132511
AT4G25890	60S acidic ribosomal protein family	0.275961	0.205643
AT1G76680	12-oxophytodienoate reductase 1	0.279887	0.251446
AT2G42490	Copper amine oxidase family protein	0.281194	0.41738
AT5G03300	adenosine kinase 2	0.296545	0.180275
AT3G49120	peroxidase CB	0.31904	0.199876
AT3G16640	translationally controlled tumor protein	0.331361	0.122871
AT2G43610	Chitinase family protein	0.332851	0.364909
AT1G27130	glutathione S-transferase tau 13	0.335598	0.666938
AT1G69410	eukaryotic elongation factor 5A-3	0.336055	0.157432
AT1G22440	Zinc-binding alcohol dehydrogenase family protein	0.344601	0.283566
AT3G01420	Peroxidase superfamily protein	0.35242	0.291957
AT3G23490	cyanase	0.356031	0.392207
AT5G05270	Chalcone-flavanone isomerase family protein	0.361207	0.128029
AT5G52920	plastidic pyruvate kinase beta subunit 1	0.411209	0.152897
AT4G33640	unknown protein	0.427606	0.397146

Supplementary Table S4 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT1G78340	glutathione S-transferase TAU 22	0.432104	0.226755
AT2G43910	HARMLESS TO OZONE LAYER 1	0.437015	0.203389
AT5G52240	membrane steroid binding protein 1	0.439038	0.114367
AT3G05420	acyl-CoA binding protein 4	0.45607	0.577248
AT3G44300	nitrilase 2	0.472696	0.499374
AT5G63600	flavonol synthase 5	0.482229	0.141564
AT5G11520	aspartate aminotransferase 3	0.517124	0.1125
AT2G38380	Peroxidase superfamily protein	0.560813	0.339023
AT5G28840	GDP-D-mannose 3,5-epimerase	0.592541	0.751892
AT2G25980	Mannose-binding lectin superfamily protein	0.598556	0.42094
AT5G28050	Cytidine/deoxycytidylate deaminase family protein	0.732834	0.167615
AT4G36760	aminopeptidase P1	0.766638	0.222681
AT1G58270	TRAF-like family protein	0.862908	0.27965
AT2G41530	S-formylglutathione hydrolase	0.900722	0.12075
AT5G26260	TRAF-like family protein	0.909658	0.518636
AT3G09270	glutathione S-transferase TAU 8	0.97563	0.239276
AT3G16450	Mannose-binding lectin superfamily protein	0.990084	0.390007

Supplementary Table S5 – Significantly de-regulated proteins described in *RLR2* KO line grown under normal conditions by quantitative shotgun proteomics. In the table are depicted the gene ID number according to "The Arabidopsis Information Resource database (www.arabidopsis.org), protein description (according to the same database), and the protein ratios (KO *vs* WT) converted in log2 values of the different biological replicates.

Gene ID	Protein description	Ratio value (log 2)		
		Replicate 1	Replicate 2	Replicate 3
AT5G02490	Heat shock protein 70 (Hsp 70) family protein	-1.36825	-1.08311	-0.35028
AT5G56670	Ribosomal protein S30 family protein	-1.31201	-0.68135	-0.58509
ATCG00490	ribulose-bisphosphate carboxylases	-0.84853	-0.20002	-0.65057
AT2G01520	MLP-like protein 328	-0.74541	-0.57167	-0.47572
AT5G20250	Raffinose synthase family protein	-0.67277	-0.22918	-0.51186
AT3G11930	Adenine nucleotide alpha hydrolases-like superfamily protein	-0.65579	-0.67703	-0.24795
AT5G38530	tryptophan synthase beta type 2	-0.54198	-0.33583	-0.15779
AT4G08770	Peroxidase superfamily protein	-0.4425	-0.32869	-0.22656
AT3G07320	O-Glycosyl hydrolases family 17 protein	-0.38853	-0.57145	-0.19489
AT5G63620	GroES-like zinc-binding alcohol dehydrogenase family protein	-0.34426	-0.27247	-0.1919
AT2G01530	MLP-like protein 329	-0.34307	-0.28097	-0.53299
AT4G11290	Peroxidase superfamily protein	-0.33971	-0.42415	-0.21317
AT1G70850	MLP-like protein 34	-0.33634	-0.25056	-0.36578
AT1G51760	peptidase M20/M25/M40 family protein	-0.3278	-0.49238	-1.04609
AT1G72160	Sec14p-like phosphatidylinositol transfer family protein	-0.32606	-0.22687	-0.18459
AT5G40390	Raffinose synthase family protein	-0.31935	-0.48392	-0.16965
AT1G21440	Phosphoenolpyruvate carboxylase family protein	-0.31595	-0.38991	-0.30118
AT1G54010	GDSL-like Lipase/Acylhydrolase superfamily protein	-0.30786	-0.46761	-0.33264
AT4G20980	Eukaryotic translation initiation factor 3 subunit 7 (eIF-3)	-0.28622	-0.23301	-0.27115
AT5G64100	Peroxidase superfamily protein	-0.27997	-0.21093	-0.16257
AT3G16400	nitrile specifier protein 1	-0.27565	-0.24626	-0.21951
AT5G59090	subtilase 4.12	-0.27477	-0.24945	-0.19441
AT1G17170	glutathione S-transferase TAU 24	-0.27085	-0.29439	-0.18961
AT1G17190	glutathione S-transferase tau 26	-0.2679	-0.18456	-0.30043
AT3G53180	glutamate-ammonia ligases	-0.24436	-0.25372	-0.31932
AT2G36290	alpha/beta-Hydrolases superfamily protein	-0.24139	-0.37021	-0.54968
AT3G25780	allene oxide cyclase 3	-0.21422	-0.26951	-0.35648
AT5G59420	OSBP(oxysterol binding protein)-related protein 3C	-0.2093	-0.2989	-0.29013

Supplementary Table S5 (Cont.) – Significantly de-regulated proteins described in *RLR2* KO line grown under normal conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio value (log 2)		
		Replicate 1	Replicate 2	Replicate 3
AT2G42590	general regulatory factor 9	-0.20668	-0.36058	-0.239
AT5G40370	Glutaredoxin family protein	-0.20043	-0.66399	-0.15643
AT5G50370	Adenylate kinase family protein	-0.19598	-0.58788	-0.77633
AT1G15130	Endosomal targeting BRO1-like domain-containing protein	0.111165	1.37061	0.4336
AT5G16390	chloroplastic acetylcoenzyme A carboxylase 1	0.128425	0.210763	1.41386
AT3G15660	glutaredoxin 4	0.135141	0.210638	0.390558
AT4G00810	60S acidic ribosomal protein family	0.15069	0.18574	0.185105
AT3G56340	Ribosomal protein S26e family protein	0.207268	0.160791	0.234256
AT5G23890	unknown protein	0.207893	0.677892	0.264476
AT5G03160	homolog of mamallian P58IPK	0.222186	0.169668	0.846554
AT5G55480	SHV3-like 1	0.225275	0.145612	0.363115
AT4G14300	RNA-binding (RRM/RBD/RNP motifs) family protein	0.245739	0.136454	0.220082
AT1G54580	acyl carrier protein 2	0.304277	0.333309	0.37128
AT5G47890	NADH-ubiquinone oxidoreductase B8 subunit, putative	0.465713	0.298189	0.260146
AT2G25980	Mannose-binding lectin superfamily protein	0.500292	0.545968	0.176323
AT4G25890	60S acidic ribosomal protein family	0.553655	0.420402	0.328722
AT5G13930	Chalcone and stilbene synthase family protein	0.66029	0.295958	0.282855

Supplementary Table S6 – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics. In the table are depicted the gene ID number according to "The Arabidopsis Information Resource database (www.arabidopsis.org), protein description (according to the same database), and the protein ratios (KO *vs* WT) converted in log2 values of the different biological replicates.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT5G18800	Cox19-like CHCH family protein	-2.15194	-2.03268
AT5G59613	unknown protein	-0.81495	-0.95526
AT2G39310	jacalin-related lectin 22	-0.78754	-0.41037
AT3G26460	Polyketide cyclase/dehydrase and lipid transport superfamily protein	-0.78446	-0.37148
AT4G02520	glutathione S-transferase PHI 2;glutathione S-transferase F3	-0.78217	-0.42302
AT5G02490	Heat shock protein 70 (Hsp 70) family protein	-0.77618	-0.4654
AT1G14980	chaperonin 10	-0.7598	-0.29384
AT3G01280	voltage dependent anion channel 1	-0.74434	-0.18415
AT2G01520	MLP-like protein 328	-0.65672	-0.4838
AT2G25670	unknown protein	-0.65592	-0.85194
AT5G43780	Pseudouridine synthase/archaeosine transglycosylase-like family protein	-0.65365	-0.18531
AT5G20160	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	-0.65152	-0.24044
AT5G47700	60S acidic ribosomal protein family	-0.5871	-0.35139
AT5G67500	voltage dependent anion channel 2	-0.56109	-0.1831
AT5G22650	histone deacetylase 2B	-0.55374	-0.42188
AT1G12080	Vacuolar calcium-binding protein-related	-0.53981	-0.30631
AT4G10840	Tetratricopeptide repeat (TPR)-like superfamily protein	-0.53086	-0.22948
AT5G52840	NADH-ubiquinone oxidoreductase-related	-0.5216	-0.41642
AT2G36290	alpha/beta-Hydrolases superfamily protein	-0.49419	-0.22427
AT1G54630	acyl carrier protein 3	-0.48044	-0.45328
AT1G28290	arabinogalactan protein 31	-0.47273	-0.20206
AT3G01390	vacuolar membrane ATPase 10	-0.47129	-0.54755
AT5G66680	dolichyl-diphosphooligosaccharide-protein glycosyltransferase 48kDa subunit family protein	-0.45872	-0.2134
AT3G16400	nitrile specifier protein 1	-0.43403	-0.2738
AT4G32470	Cytochrome bd ubiquinol oxidase, 14kDa subunit	-0.41444	-0.39771
AT3G13160	Tetratricopeptide repeat (TPR)-like superfamily protein	-0.41047	-0.45631
AT1G19570	dehydroascorbate reductase	-0.41001	-0.3964
AT3G52300	ATP synthase D chain, mitochondrial	-0.40993	-0.25877
AT1G54580	acyl carrier protein 2	-0.39694	-0.35003
AT3G09810	isocitrate dehydrogenase VI	-0.38259	-0.40419
AT1G13930	unknown protein	-0.37279	-0.32954
AT1G70850	MLP-like protein 34	-0.365	-0.39504
AT4G08770	Peroxidase superfamily protein	-0.33966	-0.40165
AT3G10850	Metallo-hydrolase/oxidoreductase superfamily protein	-0.33754	-0.21376

Supplementary Table S6 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT4G25630	fibrillarin 2	-0.33694	-0.22585
AT1G21440	Phosphoenolpyruvate carboxylase family protein	-0.33313	-0.21704
AT4G36760	aminopeptidase P1	-0.32675	-0.23245
AT5G02450	Ribosomal protein L36e family protein	-0.32632	-0.34747
AT5G10160	Thioesterase superfamily protein	-0.32464	-0.37653
AT2G30200	unknown protein	-0.29145	-0.30072
AT2G30860	glutathione S-transferase PHI 9;glutathione S-transferase PHI 9	-0.28919	-0.1805
AT3G53110	P-loop containing nucleoside triphosphate hydrolases superfamily protein	-0.28195	-0.28866
AT5G23020	2-isopropylmalate synthase 2	-0.281	-0.32579
AT5G54160	O-methyltransferase 1	-0.27839	-0.18797
AT4G39260	unknown protein	-0.27532	-0.23399
AT5G20830	sucrose synthase 1	-0.27324	-0.55572
AT3G53020	Ribosomal protein L24e family protein	-0.27282	-0.17754
AT1G18070	Translation elongation factor EF1A/initiation factor IF2gamma family protein	-0.27275	-0.17718
AT1G76010	Alba DNA/RNA-binding protein	-0.26528	-0.41802
AT3G51800	metallopeptidase M24 family protein;metallopeptidase M24 family protein;metallopeptidase M24 family protein	-0.25844	-0.24356
AT1G66580	senescence associated gene 24	-0.25693	-0.32648
AT2G20630	PP2C induced by AVRRPM1	-0.25599	-0.18228
AT1G77330	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	-0.24857	-0.30447
AT2G27530	Ribosomal protein L1p/L10e family	-0.23984	-0.17968
AT4G02450	HSP20-like chaperones superfamily protein	-0.23647	-0.18963
AT5G53480	ARM repeat superfamily protein	-0.23197	-0.32142
AT1G56330	secretion-associated RAS 1B	-0.21335	-0.32325
AT1G77590	long chain acyl-CoA synthetase 9	-0.21138	-0.28423
AT2G32730	26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit	-0.21116	-0.38248
AT1G77520	O-methyltransferase family protein	-0.20916	-0.55552
AT2G31570	glutathione peroxidase 2	-0.20202	-0.32339
AT1G12230	Aldolase superfamily protein	-0.19729	-0.178
AT3G15020	Lactate/malate dehydrogenase family protein	-0.19116	-0.67807
AT2G38040	acetyl Co-enzyme a carboxylase carboxyltransferase alpha subunit	-0.18922	-0.21826
AT2G29420	glutathione S-transferase tau 7	0.165044	0.341986
AT5G54500	flavodoxin-like quinone reductase 1	0.169925	0.315102
AT5G66510	gamma carbonic anhydrase 3	0.172616	0.23352
AT5G23830	MD-2-related lipid recognition domain-containing protein	0.180784	0.265437
AT4G28940	Phosphorylase superfamily protein	0.184471	0.15549
AT5G58290	regulatory particle triple-A ATPase 3	0.197614	0.203514

Supplementary Table S6 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT2G22780	peroxisomal NAD-malate dehydrogenase 1	0.201508	0.234011
AT1G35720	annexin 1	0.211884	0.170822
AT5G19440	NAD(P)-binding Rossmann-fold superfamily protein	0.212258	0.244644
AT2G15490	UDP-glycosyltransferase 73B4	0.235237	0.165815
AT5G23860	tubulin beta 8	0.242328	0.192573
AT3G14990	Class I glutamine amidotransferase-like superfamily protein	0.243425	0.123666
AT5G55480	SHV3-like 1	0.255682	0.219958
AT3G19450	GroES-like zinc-binding alcohol dehydrogenase family protein	0.260387	0.123004
AT5G13710	sterol methyltransferase 1	0.264116	0.217479
AT4G33640	unknown protein	0.272979	0.402722
AT3G22850	Aluminium induced protein with YGL and LRDR motifs	0.277271	0.478092
AT5G34850	purple acid phosphatase 26	0.281787	0.340049
AT2G33150	peroxisomal 3-ketoacyl-CoA thiolase 3	0.316146	0.251325
AT5G58070	temperature-induced lipocalin	0.317073	0.130799
AT4G16260	Glycosyl hydrolase superfamily protein	0.322505	0.194213
AT5G65360	Histone superfamily protein;Histone superfamily protein	0.324695	0.205393
AT3G44320	nitrilase 3	0.341075	0.141825
AT1G32440	plastidial pyruvate kinase 3	0.352646	0.344601
AT3G58750	citrate synthase 2	0.36614	0.21176
AT1G78850	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	0.367036	0.315914
AT3G01420	Peroxidase superfamily protein	0.372841	0.287236
AT4G34138	UDP-glucosyl transferase 73B1	0.384492	0.389787
AT1G62660	Glycosyl hydrolases family 32 protein	0.389787	0.147437
AT1G09560	germin-like protein 5	0.400975	0.217107
AT2G43610	Chitinase family protein	0.431035	0.513289
AT3G22960	Pyruvate kinase family protein	0.441058	0.166073
AT3G44310	nitrilase 1	0.451963	0.228726
AT2G02390	glutathione S-transferase zeta 1	0.472592	0.285698
AT2G38380	Peroxidase superfamily protein	0.476848	0.345964
AT4G23670	Polyketide cyclase/dehydrase and lipid transport superfamily protein	0.537147	0.115965
AT2G44790	uclacyanin 2	0.583423	0.318114
AT2G43910	HARMLESS TO OZONE LAYER 1	0.609566	0.215741
AT3G49120	peroxidase CB	0.657183	0.250234
AT3G23490	cyanase	0.657549	0.33674
AT1G06000	UDP-Glycosyltransferase superfamily protein	0.679964	0.10742
AT3G16450	Mannose-binding lectin superfamily protein	0.683472	0.402831
AT3G44300	nitrilase 2	0.698218	0.654711
AT2G25980	Mannose-binding lectin superfamily protein	0.702037	0.349139

Supplementary Table S6 (Cont.) – Significantly de-regulated proteins described in RLR1 KO line grown under abiotic stress conditions by quantitative shotgun proteomics.

Gene ID	Protein description	Ratio values (log 2)	
		Replicate 1	Replicate 2
AT3G09270	glutathione S-transferase TAU 8	0.768333	0.187768
AT2G42490	Copper amine oxidase family protein	0.801573	0.119821
AT5G44110	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.818278	0.192825
AT3G11250	Ribosomal protein L10 family protein	0.826274	0.37629
AT5G26260	TRAF-like family protein	0.876095	0.625458
AT5G63600	flavonol synthase 5	0.952781	0.192068
AT3G46830	RAB GTPase homolog A2C;RAB GTPase homolog A2B	1.00223	0.108892
AT1G17860	Kunitz family trypsin and protease inhibitor protein	1.17817	0.240253
AT3G61140	26S proteasome, regulatory subunit Rpn7	1.76299	0.559149
AT1G58270	TRAF-like family protein	1.90335	0.298776