Protein processing in *Nicotiana benthamiana* by Vacuolar Processing Enzymes and other proteases



Emma Louisa Thomas

Hertford College

Department of Plant Sciences

University of Oxford

Supervised by Professor Renier A.L. van der Hoorn

Thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2020

Abstract

Protein processing by proteases is an irreversible post-translational modification that shapes the proteomes of all organisms. The outcomes of protein processing by proteases have diverse effects on protein function. This thesis describes the investigation of the landscape of protein processing in the model plant *Nicotiana benthamiana*.

Firstly, this thesis presents a snapshot of processed proteoforms in leaves captured by two degradomic techniques: PROtein TOpography Migration Analysis Platform (PROTOMAP) and Terminal Amine Isotopic Labelling of Substrates (TAILS). In combination these techniques revealed putative processing of 58% of the 4769 identified protein groups. Several case studies of putative processing events are discussed, which illustrate the value of this dataset as a resource to support future studies of plant proteins.

Secondly, this thesis explores the roles of two major vacuolar processing enzymes (VPEs) in leaves, NbVPE1a and NbVPE1b, are explored. VPEs are described as positive regulators of programmed cell death (PCD). Using a CRISPR-Cas9 generated *vpe1a/b* double mutant in *N. benthamiana* we demonstrate that NbVPE1a and NbVPE1b are responsible for the major VPE activities in leaves. Despite this, *vpe1a/b* mutants were not compromised in hypersensitive response PCD, nor in responses to salt, nitrogen, heat or ER stresses.

Finally, to elucidate the contribution of NbVPEs to protein processing, High-Efficiency Undecanal-based N-Termini EnRichment (HUNTER) was used to identify substrates of NbVPE1a/b. The N-terminome captured by HUNTER revealed significant changes in protease activity between the *vpe1a/b* and WT leaves and a clear preference of

overexpressed NbVPE1b for asparaginase activity. Several candidate NbVPE substrates were identified, including the receptor-like protein NbFERONIA and papain-like cysteine proteases NbRD21 and NbALP. These were subject to further investigation.

Together this thesis presents an insight into general protein processing in leaves and the specific contribution of NbVPEs to the protease activities and stress responses of *N. benthamiana*.

Acknowledgements

Firstly, thank you to Renier van der Hoorn. Your guidance and support in building my scientific thinking, style, and independence have been truly invaluable on my PhD journey.

I would like to thank all the members of the van der Hoorn and Preston labs who made my project possible and pushed me through the more challenging moments. To my fellow patatas, Maria Font Farré and Laura Ossorio Carballo I thank you for the laughs, the songs and all the food. Thank you to Jiorgos Kourelis and Marcel Bach-Pages. Whilst your philosophical wisdom was lost on me, I learnt so much from you both. I'm grateful to have worked with Mariana Schuster, Friederike Grosse-Holz, Judith Paulus and Daniela Sueldo who will always be role models to me. To Tee, Flo, Kyoko, Phillippe, Pierre, Shivani, Alice, Issy, Ben and Felix, thank you for all your help and the laughs along the way. Thank you to Ursula, Sarah and Caroline for the technical support. Another thank you to Hayley Knights, who did not have the privilege of being in our lab but to whom I owe my sanity during lockdown.

I am also grateful to the teams I have worked with in Germany. Thank you to Sabrina Ninck and Farnusch Kaschani in Essen and to Andreas Perrar, Fatih Demir and Pitter Huesgen in Jülich for my education in mass spectrometry and your warm welcome.

Finally, I thank my biggest cheerleaders. Thank you to my parents Judith and Kevin Thomas, for their unwavering support and constant belief in me. And one last time, thank you to George Colgan, who was always by my side. You made me stress less and smile more.

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1].

Table of Contents

Chapter 1: Introduction 1

1.1 Protease overview 1

1.2.1 Outcomes of protease activity 2

1.2.2 Capturing the protein processing landscape 5

1.3 Vacuolar Processing Enzymes 8

1.3.1 VPE structure and activation 9

1.3.2 Endopeptidase activity 11

1.3.3 Ligase activity 13

1.3.4 Alternative activities 15

1.3.5 Inhibition of VPEs 16

1.3.6 Localisation 17

1.3.7 Seed storage protein processing 18

1.3.8 Role in programmed cell death 19

1.3.9 Biotic stress responses 25

1.3.10 Abiotic stress responses 26

1.3.11 Senescence 27

1.3.12 Interplay with other proteases 28

1.4 Thesis Structure 30

Chapter 2: The processed proteome of agroinfiltrated *Nicotiana benthamiana* 33

2.1 Abstract 33

2.2 Introduction 35

2.3 Results 37

2.3.1 Detecting protein processing by PROTOMAP 39

2.3.2 The N-terminome captured by TAILS 41

2.3.3 Overlap between TAILS and PROTOMAP 44

2.3.4 Example 1: Confirmation of known Pectin methylesterase processing 48

2.3.5 Example 2: C01 Protease maturation 49

2.3.6 Example 3: Bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase 53

2.3.7 Example 4: Release of carbohydrate binding domains 54

2.3.8 Example 5: Variation in serine-carboxypeptidase processing 57

2.3.9 Example 6: Novel processing of a GDSL lipase like protein 61

2.3.10 Example 7: Variation in Processing of Glycosyl Hydrolases 62

2.3.11 Example 8: Low abundance A1 proteases enriched in TAILS 65

2.4 Discussion 67

2.5 Supplementary figures 71

2.6 Methods 75

Chapter 3: Characterising the biological activity and roles of NbVPE1a and NbVPE1b 86

3.1 Abstract 86

3.2 Introduction 88

3.3 Results 92

3.3.1 NbVPE1a and NbVPE1b are major vegetative VPEs 93

3.3.2 Biotic stress responses in *vpe* mutants 108

3.3.3 Abiotic stress responses in the *vpe1a/b* double mutant 113

3.4 Discussion 120

3.4.1 NbVPE1a and NbVPE1b activity 120

3.4.2 Why was no phenotype observed in *vpe1a/b* mutants? 122

3.4.3 NbVPE1a and NbVPE1b are not essential for biotic stress responses 124

3.4.4 NbVPE1a and NbVPE1b are not essential for abiotic stress responses 127

3.5 Methods 130

Chapter 4: Identifying substrates of VPEs in *Nicotiana benthamiana* leaves 142

4.1 Abstract 142

4.2 Introduction 143

4.3 Results 153

4.3.1 NbFERONIA receptor is a putative NbVPE substrate 164

4.3.2 Role of NbVPEs in activation of NbRD21 167

4.3.3. Role of NbVPEs in activation of NbALP 169

4.3.4 Processing of other proteases by NbVPE1a/b 172

4.3.5 ABP19a is a putative VPE substrate 173

4.3.6 Acid Phosphatase is a putative VPE substrate 176

4.4 Discussion 178

4.5 Supplementary figures 184

4.6 Methods 188

Chapter 5: Discussion 202

5.1 Connecting substrates and proteases 202

5.2 Understanding redundancy between proteases will be crucial in investigating individual protease biological roles 205

5.3 Activity of NbVPE1a/b is interconnected with other proteases 207

5.4 NbVPE1a and NbVPE1b are not essential for PCD 209

5.5 Future perspectives in understanding protein processing in plants 211

References 213

Appendix 1: Ten Prominent Host Proteases in Plant – Pathogen Interactions 252

A1.1 Abstract 252

A1.2 Introduction 253

A1.3 Apoplastic proteases 255

A1.3.1 SBT3.3 regulates the priming of the plant immune response 255

A1.3.2 CathB is a positive regulator of HR 256

A1.3.3 CDR1 promotes the release of systemic defence signals 257

A1.3.4 Rcr3 is a coreceptor for perception of unrelated pathogens 257

A1.4 Cytonuclear proteases 258

A1.4.1 AtMC1 is a positive regulator of HR 258

A1.4.2. The proteasome is a positive regulator of HR 259

A1.5 Vacuolar proteases 260

A1.5.1 VPEs regulate vacuolar rupture during virus-induced HR 261

A1.5.2 C14/RD21 have a complex regulation 262

A1.6 Endomembrane proteases 263

A1.6.1 ER resident AtCEP1 facilitates fungal immunity 263

A1.6.2 Golgi localised S1P controls RALF23 peptide signalling 264

A1.7 Discussion 265

Additional Files

1: AFile1_Peptographs.R	*R script for PROTOMAP peptograph plotting*
2: AFile2_Nterminimaps.R	*R script for TAILS N-termini maps plotting*
3: AFile3_ACE_0277_peptides.txt	*PROTOMAP MaxQuant peptides file*
4: AFile4_ACE_0277_proteinGroups.txt	*PROTOMAP MaxQuant proteingroups file*
5: AFile5_TAILS.csv	*TAILS peptides*

6: AFile6_ACE_0282_MQ02_FAK_peptides.txt	*ISD MaxQuant peptides*
7: AFile7_ACE_0282_MQ02_FAK_proteinGroups.txt	*ISD MaxQuant proteingroups*
8: AFile8_Wtvsmut_results.txt	*WT v. vpe1a/b HUNTER results file*
9: AFile9_Actvsinact_vpe_results.txt	*NbVPE1b v. NbVPE1a^{C215S} overexpression HUNTER results file*
10. AFile10_proteinGroups.txt	*Shotgun proteingroups*

Chapter 1: Introduction

1.1 Protease overview

Proteins are the fundamental structural components and molecular machines that create cells. With the requirement to build proteins must also come the ability to degrade them, through proteases. Proteases are an ubiquitous and highly abundant class of enzymes. They represent the second largest enzyme family in the human genome with 566 members (Puente *et al.*, 2003). Arabidopsis contains a similar number at an estimated 570 proteases. *N. benthamiana* is an allotetraploid with a larger genome and contains 1243 putative proteases (Kourelis *et al.*, 2019; Jutras *et al.*, 2020). Cleavage of proteins requires a catalytic residue in an enzyme to act as, or activate, a nucleophile to attack peptide bonds between constituent amino acids of a protein. The catalytic residue is commonly used to define the type of protease into cysteine, serine, threonine, aspartic, glutamic and metalloproteases. These are classified according to the MEROPS database (Rawlings *et al.*, 2016). The composition of proteases is strikingly conserved across plants. In Arabidopsis and *N. benthamiana* serine proteases are the most numerous (at 45% and 40% respectively), followed by cysteine (25%, 25%), metalloproteases (15%, 16%), aspartic (11%, 13%) and threonine (4%, 5%) (Lallemand *et al.*, 2015; Jutras *et al.*, 2020).

Aside from the catalytic mechanism, proteases can be broadly defined in terms of their activity. Endoproteases cleave substrates within a polypeptide chain, whereas amino- or carboxy-peptidases remove residues from the N- or C-terminus respectively. Non-processive

proteases cleave a substrate once whilst processive proteases continue to remove residues after initial cleavage in an ATP-driven process.

Amino acid residues flanking the cleaved bond are crucial to the specificity of proteases, which is governed by the confirmation of the substrate binding pocket. Residues N-terminal of the cleavage site are non-primed and numbered from the cleavage site (e.g. the closest residue to cleavage is named P1), whilst residues placed C-terminally of the cleavage site are known as primed (e.g. residue immediately after cleavage site is named P1'). Thus the cleavage site nomenclature is summarised as -P2-P1-/-P1'-P2'-.

1.2.1 Outcomes of protease activity

Protease activity can be destructive and leave no stable fragments of a substrate following activity. Alternatively, limited cleavage of substrates by proteases can generate new stable proteoforms (different protein products arising from the same gene). The action of proteases has many outcomes including:

Degradation: Misfolded or damaged proteins can inhibit cellular processes, cause inappropriate interactions and sequester limited resources. The inability to clear misfolded protein aggregates is the fundamental driver of pathologies such as Alzheimer's and Parkinson's diseases (Sweeney *et al.*, 2017). Proteolysis is required to remove non-functional proteins, and in plants bulk degradation is primarily achieved through autophagy and proteasome activity.

Nutrient Remobilisation: Plants are highly efficient in retaining nitrogen as it is a common growth limiting nutrient (Tegeder and Masclaux-Daubresse, 2018). Nitrogen is

predominantly stored in amino acids of proteins. During senescence, a massive remobilisation of nutrients from older to younger tissues occurs, requiring proteolysis to release nitrogen (Diaz-Mendoza *et al.*, 2016). Likewise, grain filling consists of remobilising up to 90% of vegetative nitrogen content in wheat and rice to fuel the future generation (Hirel *et al.*, 2007; Chen *et al.*, 2019). These resources are again stored in seeds as stable proteins until the nutrients are proteolytically released to fuel seedling growth (Tegeder and Masclaux-Daubresse, 2018).

Signalling: Post-translational modification of proteins modulates activity and function. Many of these covalent modifications can be removed. In the case of ubiquitin and small ubiquitin-like modifier (SUMO), removal is dependent on proteases. Ubiquitination most commonly targets proteins for degradation. Removal of ubiquitin is catalysed by ubiquitin deconjugating proteases (UBPs) which are conserved cysteine proteases important in both stress response signalling and development processes (Zhou *et al.*, 2017). SUMOylation affects protein localisation and stability. Deconjugation is similarly dependent on cysteine proteases and is crucial for maintaining the pool of available SUMO (Müller *et al.*, 2001; Yates *et al.*, 2016). Proteases also regulate activity of intercellular signalling peptides. For instance, the signalling peptide CLE40 is released from its precursor by three redundant subtilases. Signalling is then terminated by cleavage within the CLE40 peptide by subtilases (Stührwohldt *et al.*, 2020).

Localisation: Protein localisation is determined by motifs and targeting signals in the peptide sequence. Removal of the N-terminal signal or transit peptide following protein translocation into the secretory pathway or into organelles is an essential maturation

process. Proteases can also regulate localisation in dynamic signalling events. Ethylene perception in plants is dependent on processing of ER-tethered EIN2. Upon ethylene detection, the C-terminal portion of EIN2 is released by an unknown protease and translocates to the nucleus to mediate transcriptional responses (Wang and Wen, 2015).

Activation: Proteases themselves are commonly synthesised as inactive precursors that require removal of inhibitory domains to mature into functional enzymes. Receptors can also be activated by specific protease action. Thrombin cleavage of the human PAR1 receptor generates a new N-terminus that acts as a "tethered ligand" for the receptor which then auto-stimulates the downstream signalling pathway (Hollenberg, 2003; Hansen *et al.*, 2004).

Inactivation: Degradation is an effective way to terminate protein activity. Accumulation of the mammalian tumour suppressor P53 results in apoptosis and cell cycle arrest (Chao, 2014). Therefore continual degradation of P53 is required in healthy cells. Alternatively, limited protein processing can also inactive proteins. For example, ectodomain shedding of transmembrane cell-adhesion proteins such as human L1 results in loss of intercellular connections and adhesion, thereby promoting cell migration (Mechtersheimer *et al.*, 2001; Tien *et al.*, 2017).

The diverse outcomes of protein cleavage underly the contribution of proteases to plant biology. The integral role of proteases in pathogen responses (Balakireva and Zamyatnin, 2018; Hou *et al.*, 2018), programmed cell death (Balakireva and Zamyatnin, 2019; Stael *et al.*, 2019), abiotic stress responses (Xu and Xue, 2019), signalling, development (Liu *et al.*, 2018; Buono *et al.*, 2019) and nutrient resource management (Diaz-Mendoza *et al.*, 2016;

Chen *et al.*, 2019) is well established. A further introduction to the varied contributions and roles of proteases in plant biotic stress responses is included in **Appendix 1**, which is a published review; **Thomas, E. L. and van der Hoorn, R. A. L.** (2018) '*Ten prominent host proteases in plant-pathogen interactions.*', International Journal of Molecular Sciences, 19(2). doi: 10.3390/ijms19020639. with minor textual adaptations.

This thesis examines both the landscape of protein processing in plants and delves into the contribution of one protease, vacuolar processing enzymes (VPEs), to plant cells.

1.2.2 Capturing the protein processing landscape

Proteases are a highly abundant class of enzymes and play a crucial role in modulating protein function. Despite this, our knowledge of the processing events they catalyse is sparse. On the one hand, very little is known regarding the landscape of processed protein substrates in plants. On the other hand, the vast majority of proteases have no described substrate specificity, let alone known substrates (Tsiatsiani *et al.*, 2012; Demir *et al.*, 2017; Liu and Moschou, 2018).

The field of degradomics to study proteoform accumulation has progressed greatly in the last two decades. Degradomic techniques are divided into gel-based and positional proteomics methods (Savickas *et al.*, 2020). Gel-based methods based on polyacrylamide gel electrophoresis (PAGE) infer protein processing based upon the change in molecular weight observed on cleavage. 1D-SDS-PAGE relies of denaturation of proteins with the use of a reducing agent and strong anionic detergent sodium doecyl sulfate (SDS) to produce linear protein chains associated with negatively charged SDS. Application of an electric current

to the gel causes denatured, negatively charged proteins to migrate through the gel matrix at a rate proportional to the length and charge of the protein. Alternatively, 2 dimensional PAGE additionally separates proteins by isoleletric point across a pH gradient.

The earliest approaches used 2 dimensional-PAGE for protein separation based on molecular weight and isoelectric point of proteins and subsequent identification of differential sized proteins by mass spectrometry. However this technique suffered from low reproducibility and poor separation of many protein groups (Savickas *et al.*, 2020). A more streamlined technique was developed in protein topography migration analysis platform (PROTOMAP) (Dix *et al.*, 2008). PROTOMAP separates proteins by 1D SDS-PAGE. The gel lane is then divided into separate slices spanning molecular weights and proteins are identified by mass spectrometry. Analysing the molecular weight at which the protein migrated and the region of the protein peptides are detected from provides topographical information on the putative processing event. This technique has been more widely adopted in the animal field and lead to the identification of novel processing events in apoptosis and other pathologies.

Positional proteomics is the other degradomics technique available. These methods involve selective enrichment of either the N- or C-terminus of proteins. Processing events can be inferred by detection of new termini not predicted to be generated during protein synthesis. N-terminomics dominates the application of degradomics in the published literature (Savickas *et al.*, 2020). Combined fractional directional chromatography (COFRADIC) (Gevaert *et al.*, 2003) and Terminal Amine Isotopic Labelling of Substrates (TAILS) (Kleifeld *et al.*, 2010) are the most widely applied techniques. Both COFRADIC and TAILS begin with chemical blocking of primary amines of exposed protein N-termini. Proteins are

then digested with trypsin to generate peptides. COFRADIC and TAILS then use different negative selection methods to specifically enrich for the blocked N-terminal peptides. Exclusion of peptides without a blocked N-terminus removes tryptic peptides and greatly reduces sample complexity, thereby enhancing the detection of low abundant N-terminal peptides. COFRADIC employs chemical primary amine derivatisation of new unblocked, internal peptides and sequential rounds of chromatography to separate them from the N-terminal peptides. In contrast, TAILS captures unblocked peptides with a hyper-branched polyglycerol (HPG) polymer functionalised with aldehyde groups. The retained blocked N-terminal peptides are then identified by mass spectrometry in both techniques. Detection of N-terminal peptides provides sequence specific information on putative processing events and can capture low abundant proteins or proteins resistant to gel-based separation. Adaptation of the COFRADIC technique with addition of strong cation exchange chromatography has generated the popular ChaFRADIC method boasting optimised N-terminome capture and increased versatility of the technique (Venne *et al.*, 2013). Likewise, adaptation of the TAILS technique with an optimised workflow and replacement of HPG polymer with undecanal for hydrophobic tagging step (HUNTER) has increased the power and sensitivity of N-terminome capture (Weng *et al.*, 2019). Use of different isotopes in the initial primary amine blocking step of ChaFRADIC and TAILS-based methods allows multiple samples to be compared in a single workflow.

The application of degradomic techniques in plants is still emerging (Demir *et al.*, 2017; Perrar *et al.*, 2019). Isolated studies have made significant advances in studies of stability determinants, N-terminal modifications, organellar imported protein maturation and

alternative translation initiation sites (Bienvenut *et al.*, 2012; Köhler *et al.*, 2015; Rowland *et al.*, 2015; Venne *et al.*, 2015; Zhang *et al.*, 2015; Willems *et al.*, 2017). Surprisingly though, the use of these techniques to identify protease substrates has been limited with only a handful of published studies to date (Tsiatsiani *et al.*, 2013; Carrie *et al.*, 2015; Weng *et al.*, 2019; Dall *et al.*, 2020). Enzymatic activity is intrinsic to how proteases contribute to generate a functional organism. Therefore, identifying substrates can identify new biological roles. For example, the single substrate discovery study using COFRADIC with Arabidopsis metacaspase 9 (AtMC9) identified a new substrate, PEPCK1, that resulted in discovery of a novel role for AtMC9 in primary metabolism (Tsiatsiani *et al.*, 2013). The need for these studies is clear, and degradomics will likely gain traction in upcoming years as the techniques improve further in power and reproducibility.

1.3 Vacuolar Processing Enzymes

Many names have been assigned to VPEs, including: asparaginyl endopeptidase (AEP), legumain, peptide asparagine ligase (PAL) and in more unique cases, butelase and nucellain (Hara-Nishimura *et al.*, 1991; Kembhavi *et al.*, 1993; Linnestad *et al.*, 1998; Li *et al.*, 2003; Hara-Nishimura, 2013; Nguyen *et al.*, 2014). Given the focus of this thesis on both protease substrates and general roles of VPEs in *N. benthamiana*, we will consistently use the term VPE, as the preferred term in plant biology research. VPEs from other species will be referred to by their predominant name in the literature.

VPEs are the only protease representatives of the C13 family in the CD clan of cysteine proteases, defined by the MEROPS database (Rawlings *et al.*, 2016). Furthermore, they are the only proteases reported to date to have asparaginase activity (P1 = N). Finally, they

possess both protease and ligase activities within the family. VPEs play multiple roles across the plant life cycle, from programmed cell death to nutrient mobilisation. VPEs are widely distributed in eukaryotes (Rawlings *et al.*, 2010). In contrast to most eukaryotes, plants typically harbour an expanded and diversified repertoire of VPEs in their genomes.

1.3.1 VPE structure and activation

VPEs contain the characteristic catalytic cysteine and histidine dyad of cysteine proteases that is critical for activity. From early sequence analysis it was evident that VPEs are related to caspases (C14 family), an intensively studied group of cysteine proteases involved in programmed cell death in mammals (Hiraiwa *et al.*, 1993; Bonneau *et al.*, 2008). The structural homology was confirmed first in crystal structures of mammalian VPEs, and more recently in crystal structures of three plant VPEs from Arabidopsis and *Helianthus annuus* (sunflower) (Dall and Brandstetter, 2013; Zhao *et al.*, 2014; Haywood *et al.*, 2018; Zauner *et al.*, 2018a; Dall *et al.*, 2020).

VPEs are produced as inactive enzymes that require processing to become active. Inactive VPEs form dimers at a C-terminal interface that prohibits active site access (Yang *et al.*, 2017; Zauner *et al.*, 2018a). Increasing VPE concentrations *in vitro* promotes dimerisation and *in planta* VPEs accumulate in dense ER-derived bodies, suggesting that dimerisation is a biologically relevant mechanism to suppress activity (Hara-Nishimura *et al.*, 1995, 1998; Hayashi *et al.*, 2001; Nakaune *et al.*, 2005; Zauner *et al.*, 2018a). In a cautionary note, the formation of inactive dimers is not universal to all VPE isoforms. AtVPEβ remains monomeric and activation is independent of concentration (Dall *et al.*, 2020).

VPEs undergo both N- and C-terminal processing to generate an active protease. This occurs sequentially in castor bean, with the C-terminal portion released first followed by release of N-terminal propeptide (Hiraiwa *et al.*, 1997; Kuroyanagi *et al.*, 2002). The C-terminal portion of VPEs contains the α6-helix followed by the legumain stabilisation and activity modulation (LSAM) domain. The LSAM fold bears structural homology to death domains present in caspases which similarly mediate dimerisation (Fuentes-Prior and Salvesen, 2004).

The fate of the C-terminal portion released upon processing differs between mammalian and plant VPEs. Whilst in mammals the cleaved C-terminal portion dissociates from the protein, plants retain a unique "two-chain-state" where the released polypeptide remains associated with the protease (Zauner *et al.*, 2018a). The mouse VPE can reversibly remove the C-terminal domain due to its ability to perform both proteolysis and ligation (Zhao *et al.*, 2014). At pH 4.5 the protease could autocatalytically mature, however, upon increasing pH to 8.5, the C-terminal portion was re-ligated to reconstitute the inactive form (Zhao *et al.*, 2014). The plant-specific two-chain state is maintained by either ionic or hydrophobic interactions, dependent on the VPE isoform. Arabidopsis contains four VPEs (α, β, γ, δ), and two crystal structures have been solved for AtVPEβ and AtVPEγ (Zauner *et al.*, 2018a; Dall *et al.*, 2020). AtVPEγ is dependent on ionic interactions, resulting in a highly pH sensitive two-chain state. At lower pH, the interaction is weaker and the α6-helix gatekeeper that obstructs the active site is repelled, resulting in increased substrate access (Zauner *et al.*, 2018a). At closer to neutral pH, the interaction is stronger and the LSAM stabilises the protease at the cost of reduced proteolytic activity (Haywood *et al.*, 2018; Zauner *et al.*,

2018a). In contrast, LSAM association in AtVPEβ is governed by hydrophobic interactions and is consequently less sensitive to pH fluctuations (Dall *et al.*, 2020). In the absence of an associated LSAM in mammals, integrins are required to maintain VPE stability at neutral pH (Dall and Brandstetter, 2016). VPEs of plants and animals can therefore be stable across a range of pHs and compartmentalisation and pH fluctuations may be an important mechanism for *in vivo* modulation of VPE activity (Hara-Nishimura *et al.*, 1995, 1998; Otegui *et al.*, 2006; Teper-Bamnolker *et al.*, 2019).

1.3.2 Endopeptidase activity

All studied VPEs cleave substrates exclusively at the C-terminal side of asparagine or aspartic acid residues (P1 = N/D) (Hatsugai *et al.*, 2004; Kuroyanagi *et al.*, 2005). VPEs were initially purified and identified based upon this strict processing activity. The discovery of VPEs in seeds indicated that they may be involved in maturing seed storage proteins, which were known to be processed at N, or less frequently, D residues (Mylne *et al.*, 2014). Subsequent isolation of VPEs and incubation with synthetic substrates confirmed their ability to cleave the seed storage proteins (Hara-Nishimura and Nishimura, 1987; Abe *et al.*, 1993; Hara-Nishimura *et al.*, 1993; Becker *et al.*, 1995; Hiraiwa *et al.*, 1999). Later studies used caspase-1 substrates (YVAD) to purify VPEs from plants (Hatsugai *et al.*, 2004; Rojo *et al.*, 2004). The amino acid sequences AAN, ESEN and YVAD are now frequently used in inhibitors or substrates for VPEs (James *et al.*, 2018; Yamada *et al.*, 2020).

Currently, VPEs are the only endoproteases known in humans and plants to cleave after asparagine residues (Rawlings *et al.*, 2010; Dall and Brandstetter, 2016). Aspartic acid is

highly similar in shape and size to asparagine, and at sufficiently low pH protonation of the Asp side chain causes the same charge state as Asn. The neutral charge allows Asp to be accommodated in the negatively charged substrate binding pocket of VPEs. Modifications, such as deamination or glycosylation, are not tolerated (Min *et al.*, 1992; Moss *et al.*, 2005). The strict specificity for P1 N or D of VPEs is appropriate for proteomics applications where robust cleavage is required for protein identification (Soh *et al.*, 2020). The use of VPE as an alternative to trypsin can extend the pH range of conventional proteomic digestions to acidic conditions, which may enable identification of proteins that are unstable at neutral pH.

Residues beyond P1 also contribute to substrate specificity. For instance, caspase-3 substrate DEVD is usually not cleaved by VPEs (Hatsugai *et al.*, 2004; Rojo *et al.*, 2004; Kuroyanagi *et al.*, 2005). However, although YVADase and ESENase is the consensus VPE activity from multiple studies, in rare cases other activities have been detected. For example, the papaver PrVPE1 was purified using a DEVD substrate, although upon analysis it displayed negligible DEVDase rates compared to YVADase activity (Bosch *et al.*, 2010). Furthermore, crystal structures of AtVPEβ and AtVPEγ revealed that the substrate binding pockets are suited for different P2 residues (Zauner *et al.*, 2018a; Dall *et al.*, 2020). Interestingly one VPE has been identified with greatly divergent activity. Tobacco NtTPE8 exhibited activity against FVR, a typical cathepsin H substrate (Wang *et al.*, 2018). Most VPE studies have used a restricted suite of substrates to characterise VPE activities and therefore it will be interesting to see if additional VPEs are discovered with unexpected activities.

1.3.3 Ligase activity

Alongside VPEs, the only other members of the C13 family are GPI-transamidases (Rawlings *et al.*, 2010). These enzymes catalyse ligation of glycosylphosphatidylinositol (GPI) anchors to proteins (Benghezal *et al.*, 1996). Several VPEs can also perform peptide ligation reactions between exposed protein termini to generate cyclic peptides (James *et al.*, 2018). The ability of VPEs to act as ligases has led to great interest in potential commercial applications. Ligation activity has not been universally conserved in the C13 family. Indeed, the capacity of VPEs to generate cyclic peptides has been recruited independently at multiple points in plant evolution (Gillon *et al.*, 2007; Saska *et al.*, 2007; Mylne *et al.*, 2012).

Consistent with the patchy distribution of ligation capacity, the presence of cyclic peptides in plants is similarly restricted (Weidmann and Craik, 2016). Only specific families in angiosperms, namely Violaceae and Rubiaceae contain cyclic peptides. Despite this, some VPEs from plants with no reported native cyclic peptides can ligate cyclic substrates from other species. For example, AtVPEγ and AtVPEβ can cyclise the prototypic sunflower trypsin inhibitor 1 SFTI-1 peptide (Mylne *et al.*, 2011; Zauner *et al.*, 2018b; Dall *et al.*, 2020). Cyclic peptides may be more widely distributed than currently recognised, as their prediction and discovery is by no means a trivial task. Precursors of cyclic peptides are often buried within large, complex proteins and their detection by mass spectrometry is challenging due to the large number of possible fragmentation patterns (Liu *et al.*, 2009). Furthermore, the VPEs previously described to lack ligase activity may still have the capacity to perform ligation, if it has a different substrate specificity, or requires additional modifications. For example, glycosylation is essential for the ligase activity of AtVPEγ.

However this is not universally required as HaAEPs produced in *E. coli* without glycosylation are functional ligases (Bernath-Levin *et al.*, 2015; Zauner *et al.*, 2018a).

Comparative sequence analysis and structural studies of the VPE ligases have identified features defining ligase activity. These include Ligase activity determinants LAD1, LAD2 and the Marker of Ligase Activity (MLA) region (Jackson *et al.*, 2018; Hemu *et al.*, 2019). With the exception of LAD2, these elements interact directly with the substrate in structural simulations of *Momordica cochinchinensis* (Gac) MCoAEPs (Du *et al.*, 2020).

Balancing between protease and ligase activity

The balance between the protease and ligase activities of VPEs is regulated in two ways. In plant species such as *Clitoria ternatea*, activities are divided between different VPEs. Butelase-1 exclusively ligates substrates whilst butelase-2 only performs cleavage reactions (Nguyen *et al.*, 2014; Serra *et al.*, 2016; Hemu *et al.*, 2019). Alternatively, the activities can coexist in bifunctional VPEs that are regulated by pH. Sunflower, Arabidopsis and Gac (*Momordica cochinchinensis*) contain bifunctional VPEs that preferentially ligate substrates at neutral pHs whilst cleavage dominates at acidic pH (Haywood *et al.*, 2018; Zauner *et al.*, 2018a; Du *et al.*, 2020). Mechanistically, the act of protein cleavage when coupled to ligation does not necessarily proceed via hydrolysis in the same way as VPE mediated protein cleavage. Instead, a transpeptidation reaction was preferred by sunflower HaAEP1 and *Oldenlandia affinis* OaAEP1b (Bernath-Levin *et al.*, 2015; Harris *et al.*, 2015). Both activities are relevant for cyclic substrates as plant cyclic peptides precursors typically require either N- or C-terminal processing.

Ligation and protease activities both have strong preference for P1 = N substrates. However, mutation of P1 to D dramatically reduced the efficiency of cyclisation of tested substrates to negligible levels (Bernath-Levin *et al.*, 2015). Indeed, the elevated pH permissive to ligation reactions would result in de-protonated Asp side chain that would be repulsed by the negatively charged substrate binding pocket.

1.3.4 Alternative activities

In addition to processing and ligation, some VPEs can perform other activities. The human VPE (legumain) exhibits carboxypeptidase activity dependent on pH. At closer to neutral pH, the LSAM domain is not removed from the proteolytic domain and obstructs substrate access to the active site, restricting cleavage to C-terminal residues (Dall and Brandstetter, 2013). Whilst the biological relevance for this activity is unclear, the crucial role of pH regulating VPE activity is well established and intriguing.

An unprecedented moonlighting role of VPE in transcriptional regulation has also been reported (Matarasso *et al.*, 2005). Despite the presence of a signal peptide, a *Solanum lycopersicum* VPE-GFP fusion construct was reported to have cytoplasmic localisation. Chromatin immunoprecipitation experiments indicated binding of VPE to the promoter of a gene encoding a 1-aminocyclopropane-1-caboxylic acid synthase (*SlAcs2*) (Matarasso *et al.*, 2005). The study proposed an N-terminal SUMOylation site mediates translocation to the nucleus, however VPEs undergo N-terminal and C-terminal processing which is predicted to remove the putative SUMOylation site. Furthermore, the GFP tag would be expected to be removed from a mature VPE. Indeed, no evidence was given that the full length VPE-GFP accumulates in this study. It will be interesting to observe the effect on

SlAcs2 gene expression if the experiments are repeated with LSAM-GFP fusion or GFP controls for VPE-expression.

1.3.5 Inhibition of VPEs

VPEs are not inhibited by the canonical cysteine protease inhibitor, E64 (Müntz *et al.*, 2002). Instead, synthetic peptide inhibitors based on YVAD and ESEN substrates have been crucial to the investigation of VPEs. A subgroup of cystatins (I25 inhibitors, MEROPS) are the only proteinaceous inhibitors of VPEs known in plants. Papain-like cysteine proteases (PLCPs) are inhibited by cystatins, and a C-terminal extension present in some cystatins can additionally inhibit VPEs (Benchabane *et al.*, 2010). Expression of the C-terminal extension alone from a rice cystatin specifically reduced VPE activity (AANase) compared to other protease activity (DLRase) (Christoff *et al.*, 2016). The VPE-inhibiting C-terminal extension contains a crucial SNSL motif with N as a P1 residue (Martinez *et al.*, 2007). Mutation of this sequence in strawberry and barley cystatins demonstrated it is essential for the inhibition of VPEs in barley extracts (Martinez *et al.*, 2007). The interaction between extended cystatins and VPE has been extensively studied in animal systems (Zhao *et al.*, 2014; Dall *et al.*, 2015). Interestingly, inhibition by cystatin stabilises VPE at near neutral pHs and inhibition is reversible at lower pH, thereby mimicking the effect of the LSAM domain. Cystatins C and E are known substrates of VPEs in mammals (Zhao *et al.*, 2014; Dall *et al.*, 2015). Sugarcane VPE activity is inhibited by the endogenous cystatin3 and the strong co-expression pattern of the cystatin and VPE suggest that cystatin3 is involved in *in vivo* regulation of VPE (Santos-Silva *et al.*, 2012).

Additional inhibitors of plant VPEs may yet be discovered. Mammalian VPEs are inhibited by pathogen-derived inhibitors clitocypin and macrocypin from fungi (I48 family, MEROPS) (Sabotič *et al.*, 2009; Renko *et al.*, 2010). Similar to cystatins, these inhibitors can target multiple cysteine protease families through different interfaces. I48 inhibitors appear to occur exclusively in fungi and could potentially be employed by fungal plant pathogens to manipulate VPE activity of the host (Sabotič and Kos, 2012).

1.3.6 Subcellular localisation

VPEs contain a signal peptide which leads to insertion into the ER and Golgi network following synthesis. Here VPEs are N-glycosylated (Zauner *et al.*, 2018a) and subsequently transported to the vacuole and apoplastic fluids presumably through secretory sorting pathways. The pH of the secretory pathway compartments differs in pH as vesicles mature and approach the major acidic compartments of the vacuole and apoplastic fluids (Felle, 2001; Martinière *et al.*, 2013; Shen *et al.*, 2013; Borniego *et al.*, 2020).

VPE activity has been detected in acidic compartments including vacuoles and the apoplast of plants and lysosomes and extracellularly in mammals. (**Fig. 1**) Plant VPEs were first purified from eponymous vacuolar compartments. *In vivo* vacuolar localisation has been shown in tobacco, Arabidopsis and castor bean through activity-based and immuno-based methods (Hiraiwa *et al.*, 1993; Kinoshita *et al.*, 1999; Hatsugai *et al.*, 2004; Labudda *et al.*, 2020). VPE presence and activity has also been detected in the apoplast in Solanaceous species (Goulet *et al.*, 2012; Sueldo *et al.*, 2014; Grosse-Holz *et al.*, 2018). VPEs can reach the vacuole via ER-derived bodies or by autophagasomal pathways (Hara-Nishimura *et al.*, 1995, 1998; Otegui *et al.*, 2006; Teper-Bamnolker *et al.*, 2019).

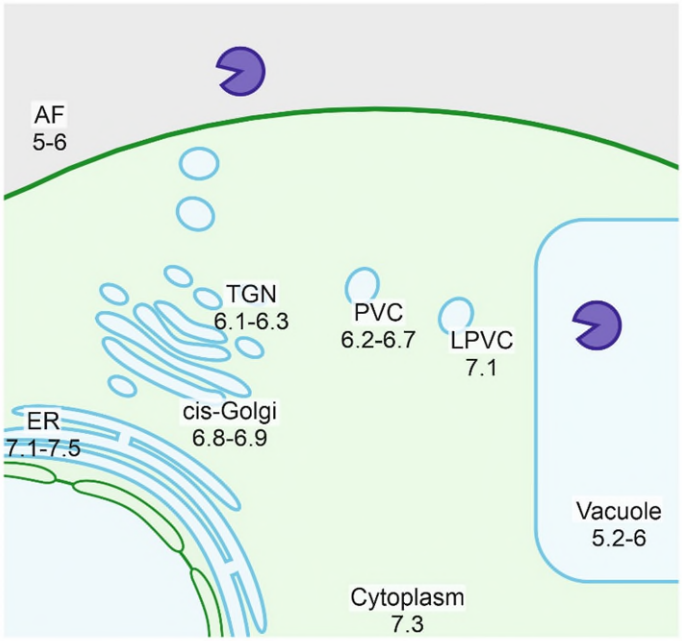


Fig. 1 Subcellular localisation of VPEs. VPE is shown in compartments with confirmed activity. The pH of each compartment is listed underneath the name (Felle, 2001; Martinière *et al.*, 2013; Shen *et al.*, 2013; Borniego *et al.*, 2020). AF apoplastic fluid, ER endoplasmic reticulum, PVC pre-vacuolar compartment, LPVC late pre-vacuolar compartment, TGN trans golgi network.

Roles of VPEs in plants

VPEs are implicated in diverse aspects of plant life across all tissues. However, the data on biological roles is fragmented between species and unifying theories of VPE roles are continually challenged.

1.3.7 Seed storage protein processing

VPEs process albumin seed storage proteins at disordered regions between domains. This processing occurs at highly conserved Asn-Gly, Asn-Gln or Asn-Pro peptide bonds and is necessary for the correct assembly of albumins into hexamers (Dickinson *et al.*, 1989; Müntz *et al.*, 2002) (Jung *et al.*, 1993). Mutation of the asparagines disrupts the formation of the

hexameric complexes. The Arabidopsis *vpe* null mutant lacking all four VPEs (α, β, γ, δ) demonstrated greatly altered seed storage protein processing. Processing in the single *vpeβ* mutant was also greatly perturbed. This was the first indicator of subfunctionalisation and the possible redundancy between VPEs. More recently, the N-terminomic technique HUNTER confirmed that the N-termini of seed storage proteins were different between WT and *vpe* null mutant, ratifying a role of VPEs (Weng *et al.*, 2019). However, despite an established role in maturing seed storage proteins, there was no effect on germination or seedling development (Gruis *et al.*, 2002, 2004). The biochemical role of native Arabidopsis VPEs therefore is uncoupled from any macroscopic phenotype. Indeed, it is known that aspartic proteases may be able to compensate for processing, albeit not at the same position (Gruis *et al.*, 2002, 2004). Serine and cysteine proteases have also been implicated in storage protein processing (Diaz-Mendoza *et al.*, 2019). It is highly plausible that redundancy exists not only between VPEs, but also with unrelated proteases.

Other targets of VPE-dependent seed protein processing release cryptic proteins from larger precursors proteins. This includes the processing of PV100, a vacuolar precursor protein that is digested to generate multiple functional mature proteins (Yamada *et al.*, 1999). Similarly, cyclic peptide substrates of VPEs such as the prototypic SFTI of sunflower are released from albumin precursors (Mylne *et al.*, 2011).

1.3.8 Role in programmed cell death

Another major biological role of VPEs in plants is in the regulation of programmed cell death (PCD). This has been reviewed in (Hatsugai *et al.*, 2015) and refreshed recently in (Vorster *et al.*, 2019; Yamada *et al.*, 2020).

VPEs in the hypersensitive response (HR)

The first evidence for a role in PCD came from the tobacco relative, *Nicotiana benthamiana.* Using a PVX-based silencing system, the two major vegetative VPEs *NbVPE1a* and *NbVPE1b* were silenced in transgenic plants carrying the *N* resistance gene. Tobacco Mosaic virus (TMV) infection induced vacuolar rupture and subsequent HR in control plants, but this was not observed in the *NbVPE1a/b* silenced plants. Although accumulation of TMV particles was higher in the *NbVPE1a/b* silenced plants, the induction of PR proteins was not perturbed (Hatsugai *et al.*, 2004). *NbVPE1a/b* silencing similarly blocked HR induced by a proteinaceous elicitor of *Xanthamonas*, harpin (Zhang *et al.*, 2010). In this case, silenced plants were not perturbed in ROS production, but were hampered in elicitor-induced stomatal closure. This indicated that VPEs act upstream of vacuolar rupture during PCD, but independent from the ROS burst and transcriptional responses. HR induced by overexpression of two calcium channels AtCNGC11 and AtCNGC12 demonstrated a weaker requirement for VPEs. Silencing of *NbVPE1a/b* delayed, but did not abolish, cell death upon overexpression of AtCNGC11 and AtCNGC12 (Urquhart *et al.*, 2007). However, HR elicited by boehmerin and nep1 was not reduced in *NbVPE1a/b* silenced plants (Zhang *et al.*, 2010). Therefore the role of VPEs in HR induction is not universal and is dependent on the HR inducer.

In Arabidopsis the role of VPEs in HR induction is similarly conflicted. VPEs are required for PCD induced by the fungal toxin FB1 in Arabidopsis leaves. The *vpe* null mutant exhibited no symptoms of PCD upon FB1 treatment. *vpeγ* had the greatest reduction of FB1 induced PCD among the individual mutants (Kuroyanagi *et al.*, 2005). Induction of

PCD by a highly similar mycotoxin, AAL, in *Nicotiana umbricata* was also reduced upon VPE silencing (Mase *et al.*, 2012). PCD in compatible biotic interactions is also VPE dependent. The mutualistic fungi *Piriformospora indica* requires PCD of root cells for successful colonisation. In line with *N. benthamiana* results, the *vpe* null mutant was compromised in vacuolar collapse and subsequent PCD during *P. indica* colonisation (Qiang *et al.*, 2012). VPEs therefore appear to have a role in PCD induced by fungal pathogens across species. However, other HR inducers in Arabidopsis do not require VPEs. *Pseudomonas syringae* DC3000 carrying AvrRpm1 is avirulent on Arabidopsis and rapidly induces a HR which was not perturbed in the *vpe* null mutant (Hatsugai *et al.*, 2009).

The variability of VPEs involvement in HR may reflect the heterogeneity of HR induction. It is not a single pathway but a broad phenomenon encompassing multiples modes of death induction (Mur *et al.*, 2008; Balint-Kurti, 2019). For example, some inducers trigger vacuolar rupture, whilst a bacterial inducer triggered vacuolar membrane fusion with the plasma membrane releasing vacuolar contents extracellularly (Hatsugai *et al.*, 2009; Pajerowska-Mukhtar and Dong, 2009). This may serve to counteract different pathogen strategies with either intra- or extra-cellular colonisation. The morphology of the HR induced by diverse pathogens needs to be studied in a single model plant to test this hypothesis.

VPEs in abiotic stress-induced PCD

PCD can proceed through ER stress to execute cell death. The PCD induced by *P. indica* discussed above proceeds in a manner highly analogous to ER stress induction by tunicamycin treatment. Tunicamycin (Tm)-induced death was reduced in Arabidopsis *vpe*

null mutant roots and delayed in the leaves, although PCD did still occur (Qiang *et al.*, 2012; Cai *et al.*, 2018). In leaves, Cathepsin B, another protease with distinct caspase-like activity, was independently required for PCD. Whilst *vpe* null lines were impaired in vacuolar rupture, cathepsin B mutants were not. However, cathepsin B mutants did still exhibit reduced death (Cai *et al.*, 2018). Crucially, this indicated that vacuolar rupture does not directly equate to cell death.

Indirect evidence suggests VPE involvement in Tm-induced PCD may be conserved in other species. VPEs are transcriptionally upregulated by known regulators of Tm-induced PCD in soybean (Mendes *et al.*, 2013). Upregulation of VPEs upon induction of PCD has been reported in response to many other stresses. Under heat stress, Arabidopsis VPEs were upregulated and the *vpe* null mutant displayed an attenuated and delayed PCD (Li *et al.*, 2012). Furthermore, transgenic expression in Arabidopsis of a radish VPE expressed in flowers, *RsVPE1,* resulted in floral bud abortion upon heat stress which was not observed in control plants (Zhang *et al.*, 2013). In *Nicotiana tabacum* cell culture, VPE transcripts and activity is upregulated upon exposure to aluminium (Kariya *et al.*, 2013). Arabidopsis VPEs are also upregulated during ozone-induced PCD (Kadono *et al.*, 2010). Finally, upregulation of VPEs were also detected prior to PCD in lethality of *N. tabacum* hybrids (Mino *et al.*, 2007).

VPEs in developmental PCD

Formation of xylem vessels requires cells to terminally differentiate and undergo PCD (Bollhöner *et al.*, 2012). Specific VPE expression has been detected in developing xylem structures. *AtVPEα* transcription was detected in developing xylem elements of roots

(Kinoshita *et al.*, 1999), whilst in a more recent paper AtVPEγ protein accumulation was detected in primary xylem development in maturing stems (Cheng *et al.*, 2019). Expression of VPEγ and VPEβ homologs in poplar has also been detected in stem xylem fiber development indicating that expression is not unique to Arabidopsis (Courtois-Moreau *et al.*, 2009). This is despite the xylem developing in *Populus* exhibiting different PCD morphologies to Arabidopsis, with the organellar degradation preceding tonoplast rupture instead of the reverse in Arabidopsis with vacuolar rupture leading cell death initiation (Courtois-Moreau *et al.*, 2009; Bollhöner *et al.*, 2012). Given the requirement of VPEs for vacuolar rupture in some PCD types, this difference in order of organellar degradation may be pertinent to the study of VPE activity in these different plants (Hatsugai *et al.*, 2004). A role of VPEs in xylem development was previously overlooked given the lack of any observable phenotype in the Arabidopsis *vpe* null mutant, such as inhibited stem development or drought-like symptoms (Gruis *et al.*, 2004; Kuroyanagi *et al.*, 2005). However, recent studies revealed a delay in developing xylem cell PCD in the *vpeγ* mutant (Cheng *et al.*, 2019). AtVPEγ contributes to cell death upstream of the cysteine protease AtCEP1, a positive regulator of xylem development PCD (Cheng *et al.*, 2019; Han *et al.*, 2019). AtCEP1 requires removal of an inhibitory propeptide for activation which could be catalysed by AtVPEγ *in vitro*. Whereas in WT plants only mature AtCEP1 accumulated, *vpeγ* plants accumulated both the pro- and mature AtCEP1 protease (Cheng *et al.*, 2019). Therefore AtVPEγ can mature AtCEP1, and does contribute to maturation *in vivo* but AtVPEγ is not the only activator of AtCEP1. Redundancy in AtCEP1 activation may explain why more drastic phenotypes are not observed in the *vpe* null mutant.

VPE-dependent activation of AtCEP1 is not restricted to xylem development. In parallel, the same group reported *in vitro* activation of AtCEP1 by AtVPEβ in the context of tapetum degradation in pollen development (Cheng *et al.*, 2020). The *vpeβ* mutant displayed a delayed vacuolar rupture prior to cell death and consequently has reduced pollen fertility. Cysteine proteases RD19 and RD21 accumulated more as the immature, inactive proteoform in the *vpeβ* mutant, but direct activation by VPE has not been shown (Cheng *et al.*, 2020). Together these reports support VPEs as indirect regulators of vacuolar integrity, lying upstream of other proteases to control this process.

VPEs have also been implicated in tapetal degradation in *N. tabacum* (Wang *et al.*, 2018). NtTPE8 was identified as a VPE expressed exclusively in the developing pollen tapetum. Silencing of *NtTPE8* reduced fertility of the resulting seeds similar to observations in Arabidopsis (Wang *et al.*, 2018). Indeed, NtTPE8 is most homologous to AtVPEβ in Arabidopsis, indicating that the role of VPEβ in pollen development may be conserved across species.

Many other species have reported VPE dependency in different aspects of reproductive developmental cell death. In Arabidopsis, AtVPEδ contributes to PCD in specific seed layers, as does HvVPE4 in barley and OsVPE3 in rice aleurone layers (Nakaune *et al.*, 2005; Radchuk *et al.*, 2018; Zhang *et al.*, 2020). Interestingly in rice aleurone death, it appears a positive feedback loop between actin depolymerisation and VPE activity may exist. Chemical inhibition of VPE blocked actin depolymerisation and subsequent vacuolar membrane fusion, and correspondingly actin depolymerisation triggered an upregulation of VPE activity (Zhang *et al.*, 2020). It is unclear how feedback between actin

depolymerisation and VPE activation before vacuolar rupture would operate but is an interesting mechanism to commit cells to undergo PCD.

In summary, extensive evidence supports VPEs as regulators of PCD but contradictory cases do exist. Despite this, literature frequently refers to VPEs as absolute regulators of PCD in blanket statements. This is a dangerous generalisation as their involvement is not consistent between cell death types and species.

1.3.9 Biotic stress responses

Independent of HR VPEs also play a role in pathogen defence. The Arabidopsis *vpeγ* mutant is more susceptible to a diverse range of pathogens, including *Botrytis cinerea*, Turnip Mosiac virus and *P. syringae* (Rojo *et al.*, 2004). In contrast, an oomycete obligate biotroph *Hyaloperonospora arabidopsidis* exhibited reduced sporulation (Misas-Villamil *et al.*, 2013) whilst *Piriformospora indica* displayed reduced colonisation on the *vpe* null mutant (Qiang *et al.*, 2012).

Despite the different outcomes reflecting pathogen lifestyles, upregulation of VPE at transcript or activity level is common to all of the interactions in the above paragraph (Rojo *et al.*, 2004; Qiang *et al.*, 2012; Misas-Villamil *et al.*, 2013). Induced transcript levels also suggest positive involvement of VPEs in pathogen responses of other plant species. Apple (*Malus domestica*) exhibited upregulation of VPEs during *Erwinia amylovora* infection (Iakimova *et al.*, 2013). In contrast, tomato plants undergoing HR cell death induced by the fungal pathogen Avr4 protein which is recognised in tomato carrying the *Cf4* gene was not associated with VPE upregulation, although the activity profile of extracellular VPEs

did change (Sueldo *et al.*, 2014). Increase of VPE transcripts correlating with PCD induction was also observed in *Phytophthora sojae* infected *N. benthamiana* and cryptogein (an elicitor of *Phytophthora cryptogea)* treatment in *Nicotiana tabacum* Xanthi cell culture (Gauthier *et al.*, 2007; Jing *et al.*, 2016). Activity labelling of barley roots infected with the nematode *Heterodera filipjevi* revealed an early increase of VPE activity (Labudda *et al.*, 2020). Clearly VPEs are present and often induced during these biotic interactions, but the role of the protease is unclear.

1.3.10 Abiotic stress responses

Reduced stomatal opening was observed in many experimental systems lacking VPEs. Silencing of *N. benthamiana NbVPE1a/b* yielded mature plants with reduced stomatal aperture (Zhang *et al.*, 2010). Similarly, the Arabidopsis *vpeγ* mutant exhibited reduced stomatal opening under native conditions and silencing of rice *OsVPE3* resulted in reduced guard cell length (Lu *et al.*, 2016). Together this suggests a VPE role in stomata development. Additionally, the Arabidopsis *vpeγ* mutant displayed reduced water loss and subsequently increased drought tolerance. Surprisingly, the difference in stomatal aperture was not observed following ABA treatment, suggesting stomatal stress responses were not AtVPEγ dependent (Albertini *et al.*, 2014). However, other plant species experience transcriptional upregulation of VPEs upon drought stress (Cilliers *et al.*, 2018; Song *et al.*, 2020). The role of VPEs in response to hydration status may therefore vary between plant species and tissues, given the different responses required between them.

In line with the potential role of VPEs in drought tolerance, further evidence supports a role of VPEs in salt stress responses of rice. Exposure to toxic salt levels can induce PCD

and coincided with a transcriptional upregulation of *OsVPE2* and *OsVPE3* downstream of calcium signalling (Kim *et al.*, 2014). Overexpression and silencing of *OsVPE3* demonstrated it to be positive driver of vacuolar rupture, PCD and salt tolerance (Lu *et al.*, 2016). Upregulation of VPE activity following toxic salt exposure has also been reported in tomato leaves, suggesting that this may be a conserved role of VPEs (Kovács, 2017). In contrast, an apple VPE that is upregulated upon drought and salt stresses conferred reduced salt tolerance when overexpressed in Arabidopsis (Song *et al.*, 2020). The salt stress administered differed vastly between these studies, and therefore it is unclear how the role of VPE in salt responses varies between species or stress response pathways.

1.3.11 Senescence

VPEs are upregulated upon senescence onset in diverse tissues and plant species. In Arabidopsis leaves transcript levels of VPEs increase with dark-induced senescence and aging (Kinoshita *et al.*, 1999; Pružinská *et al.*, 2017). Senescence of petals in *N. tabacum* and cortical root cells of Arabidopsis, and symbiotic nitrogen fixing nodule senescence in soybean and *Medicago truncatula* also demonstrated increased VPE transcript or activity levels (Oh *et al.*, 2004; Pierre *et al.*, 2014; van Wyk *et al.*, 2014). Nitrogen deprivation regimes in soybean and *Brassica napus* or carbon starvation induced senescence in *N. tabacum* cell culture also showed an upregulation of VPEs (Quain *et al.*, 2015; Poret *et al.*, 2019; Teper-Bamnolker *et al.*, 2019).

Despite the breadth of VPE upregulation reports, genetic and phenotypic evidence supporting a role of VPEs in senescence is inconsistent. Although VPEs are upregulated in Arabidopsis senescence, mutants or overexpressors are not perturbed in its onset or

progression (Pružinská *et al.*, 2017). Despite this, VPE dependent protein processing of an invertase has been reported upon senescence, indicating that this processing event is not a major contributor to senescence (Rojo *et al.*, 2004). The single *vpeα* Arabidopsis mutant exhibited elevated total protein levels under dehydration stress, suggesting a possible general role in proteolysis by VPE (Cilliers *et al.*, 2018). This result is somewhat surprising given the known redundancy of the four VPEs. In contrast, senescence of aging nodules in *Medicago truncatula* is dependent on VPE. Silencing and overexpression of VPEs in this system indicated that VPE may be a positive regulator of this process (Pierre *et al.*, 2014).

Several other stresses have also been described to involve VPE upregulation. In *Malus domesticus* VPE transcript levels increased upon drought, salt, cadmium and cold stresses (Song *et al.*, 2020). The biological roles of VPEs in these conditions are unclear, but do present interesting avenues for future research.

1.3.12 Interplay with other proteases

Whilst some activities of VPEs have been untangled, the issue of protease redundancy may have obscured their true contribution to plant life. Within the VPE family there is a clear case of redundancy. This is most evident in the numerous studies comparing the single and quadruple null mutants of Arabidopsis. Multiple VPEs are encoded in plant genomes and therefore the ability to create true null mutants in less tractable species is challenging. This, combined with the limits to silencing efficiency, makes eliminating VPE activity from plants to study their role challenging. An alternative method to knockout VPE activity is to use inhibitors. Indeed, caspase-1 inhibitor YVAD has been widely employed to this end. However, in recent years it has emerged YVAD activity is also present among serine

proteases, and in particular, the subtilases (Reichardt *et al.*, 2018). VPEs also process proteins following N. ESEN inhibitors may present more specific inhibitors to target VPEs, as no other proteases in plants are known currently to have this activity. However off target effects of ESEN based inhibitors cannot be excluded.

Alternatively, redundancy can be present with proteases that have the same substrate but not the same sequence specificity. Aspartic proteases compensate for VPE loss in seed storage processing, but do not cleave at the same sequence as VPEs (Gruis *et al.*, 2002). Furthermore, an N-terminomics screen recorded increased abundance of active PLCP proteoforms (including CP1 and Cathepsin B) in *vpe* null seedlings, suggesting cysteine proteases may also compensate for VPE loss (Weng *et al.*, 2019). Increased levels of active cathepsin B were also detected in unchallenged *vpe* null leaves compared to WT (Cai *et al.*, 2018). In this system, both VPE and Cathepsin B were shown to contribute to tunicamycin induced PCD, albeit in an independent manner.

In addition to compensating other proteases, VPEs can work in concert with other proteases. In the generation of the cyclic peptide Oak1, the rice VPE OaAEP1 relies on N -terminal pre-processing of Oak1 by the PLCP OaRD21 to generate a viable VPE substrate (Rehm *et al.*, 2019). A similar mechanism has been observed in vetch seed protein processing, whereby VPE (Proteinase B) required the preceding activity of PLCP (Proteinase A) to correctly process substrates (Shutov *et al.*, 1982).

VPEs can also impact activation of other proteases. A serine carboxypeptidase AtCPY was the first protease identified to be a substrate of VPE (Rojo *et al.*, 2003). However, this cleavage separated the catalytic triad of AtCPY, leading to questions of the biological

relevance of this processing (Paulus and Van Der Hoorn, 2019). More recently, the AtCEP1 protease was identified to be directly activated by VPE (Cheng *et al.*, 2019, 2020). In the same study, RD19A and RD19C were also shown to have reduced maturation in the *vpeβ* mutant, although it is unclear whether this is a direct or indirect effect of the absence of AtVPEβ (Cheng *et al.*, 2020). In line with this, PLCP activity was reduced in *vpeα* plants upon drought stress (Cilliers *et al.*, 2018). VPEs are not essential to all cysteine protease activation. Through activity-based probes, maturation of native papain-like cysteine proteases RD21, cathepsin B and aleurain like proteases were shown to be normal in Arabidopsis *vpe* null leaves (Lu *et al.*, 2015; Cai *et al.*, 2018). However, this does not exclude a role of VPEs in the activation of these PLCPs. Many proteases have the capacity to be activated by multiple proteases. For example, both Rcr3 and CEP1 can be matured autocatalytically in *cis,* or in *trans* by other proteases (Cheng *et al.*, 2019; Paulus *et al.*, 2020).

1.4 Thesis Structure:

This thesis contributes a significant advance in understanding protein processing in *N. benthamiana*. The aim of this thesis was to understand the contribution of proteases to shaping the proteome and stress responses. We approached this by firstly investigating the landscape of protein processing in plants. Then we focused our investigation into the contribution of one specific type of protease, VPEs, to protein processing and stress responses in *N. benthamiana*.

Firstly, in **Chapter 2** we set out to capture the landscape of protein processing in *N. benthamiana* leaves. Two different degradomic techniques, PROTOMAP and TAILS, are

employed to generate a snapshot of proteoform accumulation and to aid identification of novel processing events. In combination these techniques captured 4769 proteins and a putative 2767 processed proteins. Several processing insights are discussed that illustrate the use of this new dataset.

The remaining experimental **Chapters 3** and **4** focus on the specific contribution of NbVPEs to biotic and abiotic stress responses and protein processing in *N. benthamiana*.

In **Chapter 3** we studied the biological roles of NbVPE1a and NbVPE1b. To this end, we generated mutants lacking the two major vegetative VPEs, NbVPE1a and NbVPE1b. A drastic reduction in VPE-specific activity labelling and synthetic substrate ESENase activity was observed in the *vpe1a/b* mutant. The mutants were challenged with a number of biotic and abiotic stresses, but no differences in stress responses from WT plants were observed.

Chapter 4 describes the identification of substrates of NbVPE. The N-terminomic technique HUNTER was used to screen for differential protein processing events in the presence and absence of NbVPE activity. Several candidate substrates were identified and we have begun to validate the insights generated.

Finally, **Chapter 5** is a discussion of the major results and themes arising from the experimental chapters.

Chapter 2: The processed proteome of agroinfiltrated *Nicotiana benthamiana*

Emma Thomas[1], Fatih Demir[2], Farnusch Kaschani[3], Andreas Perrar[2], Sabrina Ninck[3], Pitter Huesgen[2], Markus Kaiser[3] and Renier A.L. van der Hoorn[1]

1. Department of Plant Sciences, University of Oxford
2. ZEA-3, Jülich Forschungszentrum, Germany
3. ZMB, University of Duisburg-Essen, Germany

2.1 Abstract

The proteins that accumulate in tissues can differ greatly from the predicted proteins, due to processing by proteases and alternative translation initiation and splicing. We applied two independent mass spectrometry techniques commonly used to study protein processing to characterise the processed proteoforms present in *N. benthamiana* leaves. The N-terminome was captured using Terminal Amine Isotopic Labelling of Substrates (TAILS) and the topography and molecular weights of proteoforms studied using Protein topography migration analysis platform (PROTOMAP). Together TAILS and PROTOMAP detected putative processing of 58% of the 4769 detected protein groups. PROTOMAP captured more than four times the number of protein groups than TAILS. The TAILS dataset had a high degree of overlap with PROTOMAP, with 68% of TAILS identified protein groups also identified in PROTOMAP and a similar overlap observed processed protein groups

detected. The combination of PROTOMAP and TAILS data generated sequence level definition of pro-domain removal of cysteine proteases and pectin methylesterases. PROTOMAP also identified partitioning of catalytic residues in serine proteases and a GDSL lipase-like protein onto separate peptides. Additionally, PROTOMAP described the separation of domains in glycosidases and a bifunctional metabolic enzyme. Here we discuss several insights generated from combining these contrasting methods. The examples discussed illustrate the power of this dataset for use in future protein studies in *N. benthamiana*.

2.2 Introduction

Nicotiana benthamiana is an important model organism in plant biology. The appeal of *N. benthamiana* stems from the ease and efficiency of transient expression in leaves by agroinfiltration (Norkunas *et al.*, 2018). Agroinfiltration utilises engineered *Agrobacterium tumefaciens* strains to express genes of interest in the plant. This system is extremely versatile for rapid production of proteins from diverse organisms. Protein interactions, localisation, activity and structures have been investigated using this system. Indeed, the method is so robust that it has been adopted in commercial settings for production of pharmaceutical products and other recombinant proteins (Sainsbury, 2020).

The rise of *N. benthamiana* in mainstream plant research has driven the curation of fundamental datasets including genome, transcriptome and proteome tools. Currently two independent genome assemblies are available, with continued efforts to improve annotations and gene models underway (Bombarely *et al.*, 2012; Naim *et al.*, 2012; Grosse-Holz *et al.*, 2018; Kourelis *et al.*, 2019). The extensive use of *N. benthamiana* for protein studies would further benefit from an understanding of the leaf proteome. The predicted proteome will not reflect the actual proteins accumulating in cells due to multiple factors. Erroneous gene model annotations can lead to incorrect protein sequence prediction. Secondly, differential pre- and post-translational processing can result in variant protein products from a single protein encoding genes. Collectively these variants are termed proteoforms. Alternative transcription start sites, alternative splicing of mRNA transcripts or alternative translation initiation sites can produce different proteoforms. Following protein synthesis, proteins can then be subject to post-translational modifications such as glycosylation and irreversibly

modified by protein processing. The vast majority of eukaryotic proteins are subject to co-translational removal of the initiating methionine and acetylation of the new N-terminus (Bienvenut *et al.*, 2012; Gibbs, 2015). Amino- or carboxy-peptidases can remove limited number of residues from the respective ends of proteins. Plants also harbour a huge diversity of endoproteases that can act at different sites within a peptide chain to drastically change the size of the protein (Van der Hoorn, 2008). This diversity is further increased by polymorphisms at sites governing transcription, splicing, translation and proteolytic processing. Consequently, the proteoforms that accumulate can be highly divergent from what is predicted.

Generation of different proteoforms is important for many properties of a protein. The nature of the exposed N-terminus has significant consequences for protein stability (Marino *et al.*, 2015). Half-life of proteins in the cytoplasm is influenced by the N-end rule pathway, where the N-terminal residue dictates the susceptibility of the protein to degradation (Dissmeyer *et al.*, 2018). Other N-terminal modifications can also modulate the stability of the protein by blocking aminopeptidase action. Acetylation promotes stability of the protein, as does pyroglutamate formation (Varland *et al.*, 2015). Larger changes in proteoforms can differ in whole domains to change the activity state, interactions with other biomolecules or localisation of the protein. Therefore, understanding the proteoforms that accumulate *in vivo* can be highly informative about the activity, regulation and stability of a protein.

Several mass spectrometry-based techniques are available to study the accumulation of proteoforms, in a field termed "degradomics" (Grozdanić *et al.*, 2019; Perrar *et al.*, 2019).

Two types of techniques dominate this field. Firstly, positional proteomics encompasses N- and C-terminomics methods that selectively enrich protein termini (Perrar *et al.*, 2019). These methods can therefore define the specific sequence at which proteoforms differ and offer insight into how these may be generated. The popular N-terminomic technique terminal amine isotopic labelling of substrates (TAILS) involves labelling of exposed N-termini with dimethyl groups followed by proteome digestion. N-terminal labelled peptides are then enriched and detected by LC-MS/MS (Kleifeld *et al.*, 2010). The second type of technique relies on gel electrophoresis for protein separation. The most widely adopted is protein topography migration analysis platform (PROTOMAP) that separates the proteome by one dimensional PAGE to determine the molecular weights that proteins accumulate at (Dix *et al.*, 2008).

N-terminomic and gel-based approaches yield highly distinct yet complementary information, but have not been used together in plants before. In this chapter, we use two independent proteomics techniques, PROTOMAP and TAILS, to generate a comprehensive snapshot of the state of proteoforms in agroinfiltrated *N. benthamiana* leaves that will support future protein work in *N. benthamiana*.

2.3 Results

To capture the proteome of agroinfiltrated *N. benthamiana* leaves, PROTOMAP and TAILS were performed (**Fig. 1**). Agroinfiltrated tissue was harvested before substantial protein expression could occur (i.e. transiently expressed proteins could not be detected by LC-MS/MS downstream of in solution digest or PROTOMAP). For PROTOMAP, protein extracts from four biological replicates were separated by 1D SDS-PAGE (**Fig. 1A**). Each

replicate gel lane was sliced into 24 slices between the highest (225 kDa) and lowest (10 kDa) MW markers (**SFig. 1**). The lowest MW marker (10 kDa) migrated with the loading dye front, suggesting smaller proteins would also be captured in the lowest slice. Proteins in each gel slice were identified by is solution digest followed by LC-MS/MS with an Orbitrap Elite (Thermo) (**Fig. 1**). TAILS was performed on protein extracts from three the biological replicates and N-terminal peptides were detected by LC-MS/MS with a Bruker QToF instrument (**Fig. 1B**).

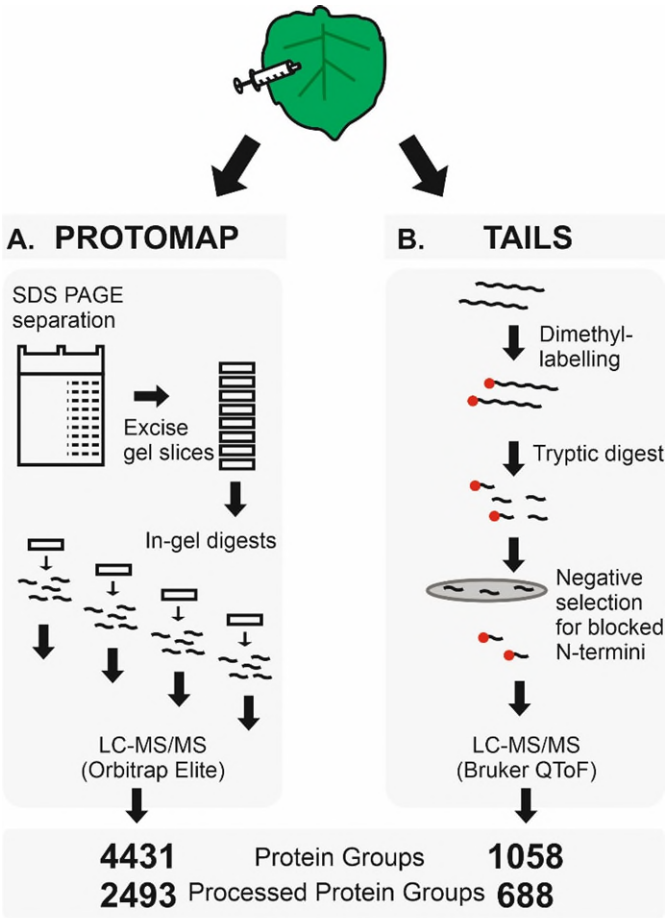


Figure 1. PROTOMAP and TAILS was performed on agroinfiltrated *N. benthamiana* leaves. Mature *N. benthamiana* leaves were agroinfiltrated and harvested at 17 hpi. **A.** PROTOMAP concept. Proteins were separated by molecular weight on denaturing SDS PAGE gel. Slices of the gel were excised and in-gel-digest was performed on each gel slice. Resulting peptides were detected by LC-MS/MS using an Orbitrap Elite instrument. **B.** TAILS concept. Proteins were extracted and primary amines at exposed N-termini were dimethylated. The mixture was digested with trypsin. All blocked N-terminal peptides were

enriched for by negative selection using a hyperbranched polyglycerol polyaldehydederivatised polymer to capture free N-termini of new trypsin generated peptides and subsequent filtration to remove the polymer and captured peptides. Enriched blocked N-terminal peptides were then detected by LC-MS/MS with a Bruker QToF instrument.

For each technique peptides detected by the mass spectrometer were mapped to protein sequences of the *N. benthamiana* proteome to identify the protein measured. Some peptides can be matched to the sequence of multiple proteins, resulting in the detection of a protein group which is comprised of multiple proteins. As the analysis of the mass spectrometry data for each technique was analysed independently, we will therefore discuss protein groups instead of proteins for comparison between techniques.

2.3.1 Detecting protein processing by PROTOMAP

PROTOMAP robustly identified 4431 protein groups across the 24 gel slices. The highest MW detected for protein groups positively correlated with the expected MW (**Fig. 2A,** Spearman's rank $r_s(4429)=0.67232$, $p < 0.0001$). However, many protein groups deviated from expected and were assigned into broad groups to discuss general migration pattern trends by visual analysis (**Fig. 2A. group1,2,3**). A large cluster of protein groups were detected in the top gel slice despite having lower predicted MW (**Fig. 2A, group2**). The observed deviation may be explained by formation of protein aggregates that failed to be denatured, although this was not observed upon Coomassie staining (**data not shown**). A second aberrant cluster is detected at higher MW than predicted (**Fig. 2A, group3**). Within this cluster we expect to observe misannotated proteins that are predicted a shorter sequence than are produced *in vivo;* identification errors from mismatching of peptides to protein groups with highly homologous domains or sequences; or proteins with extensive

post-translational modification additions, e.g. glycosylation, or phosphorylation which introduces a different charge state and thereby can influence migration on SDS-PAGE. The properties of proteins can influence the binding and coverage of SDS and therefore migration on a gel is not only dictated by protein molecular weight (Rath *et al.*, 2009; Lee *et al.*, 2019). The average of the gel slices the protein was detected in exhibited stronger correlation with expected MW (**Fig. 2B**, Spearman's rank $r_s(4429) = 0.75361$, $p < 0.0001$). The presence of processed proteins detected at lower MWs may offset the effect of the higher than expected protein group detections discussed above (**Fig. 2A, group2,3)**.

No protein groups were identified exclusively in the lowest MW gel slice 24. Indeed, the majority of protein groups (81.2%) were detected in more than one gel slice, indicating that most proteins in *N. benthamiana* exist in multiple states of post-translational modification (**Fig. 2C**).

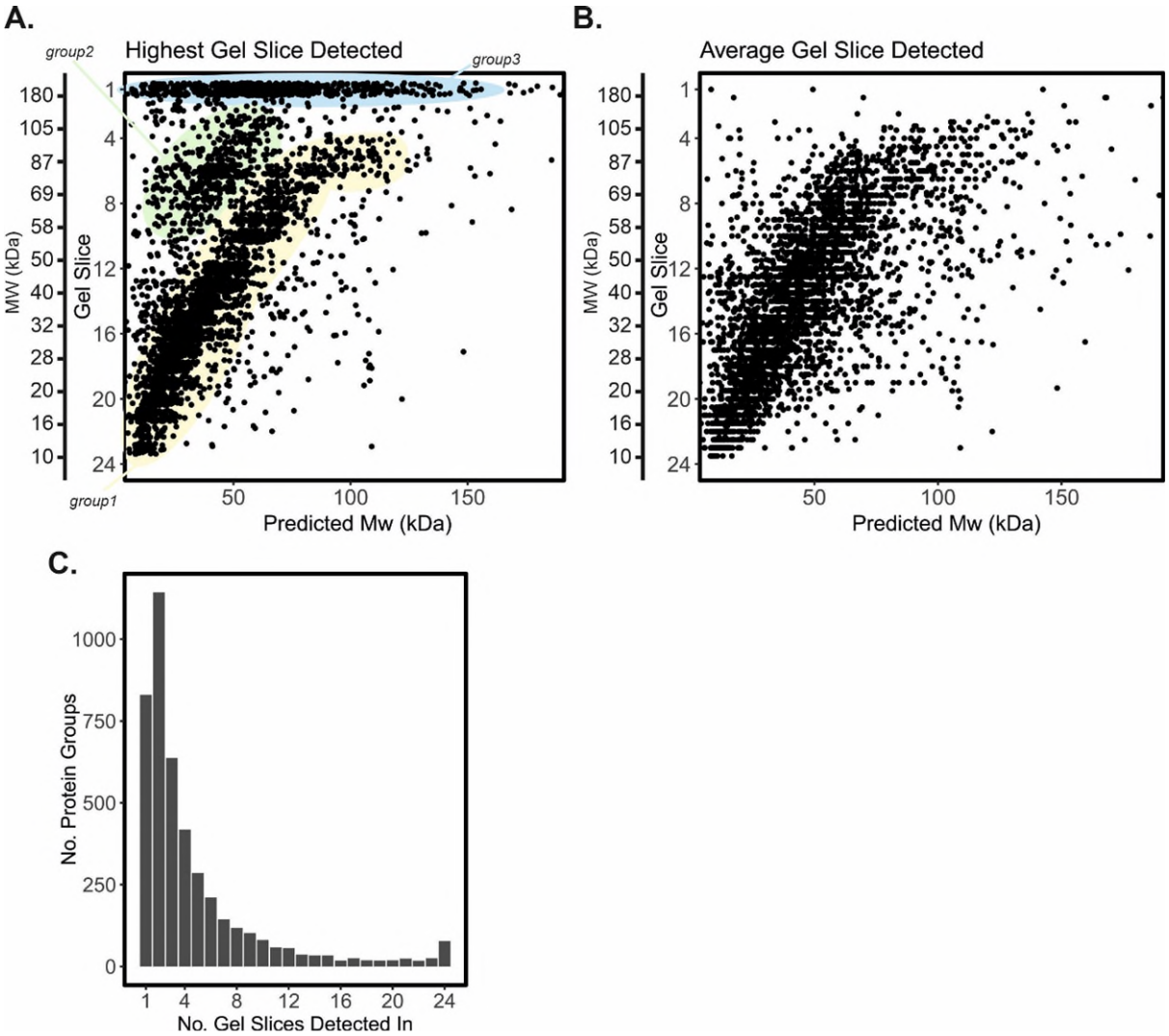


Figure 2. PROTOMAP identified migration patterns of 4431 protein groups. Proteins were separated on a denaturing gradient 4-12% SDS-PAGE gel. The gel was divided into 24 slices spanning the entire molecular weight range. Proteins in each gel slice were identified by LC-MS/MS. **A.** The highest molecular weight gel slice and **B.** the average gel slice each protein group was detected in. **C.** The number of gel slices each protein group was detected in.

2.3.2 The N-terminome captured by TAILS

The N-terminome measured by TAILS captured 1398 peptides from 1058 protein groups in at least two out of three biological replicates. 72% of the detected peptides were neo-N-termini, defined as not being the annotated start codon (peptide start position > position 2 in protein sequence) or a predicted signal peptide processing site **(Fig. 3A)**. Neo-N-

termini can be generated by alternative translation initiation, alternative splicing, or protein processing. These 1009 neo-N-termini mapped to 688 protein groups. Of the putatively processed protein groups, 513 had just a single neo-N-terminus identified but 175 had additional neo-N-termini, indicating that multiple protein processing events had occurred generating multiple stable products. Only 16 protein groups had both neo- and native N-termini detected, suggesting that cleavage events may often generate a stable and an unstable fragment which is degraded. TAILS also identified 389 native N-termini encompassing both annotated translation start sites (331) and signal peptide processing sites (58) (**Fig. 3A**). N-terminal acetylation is a common modification of newly synthesised proteins across eukaryotes and is present in >80% of proteins in *A. thaliana* and humans. (Bienvenut *et al.*, 2012; Gibbs, 2015). TAILS detected acetylation of 73.4% of the native N-termini (**Fig. 3A**).

The residues most frequently observed at P1' position were serine, alanine and threonine, which are stabilising N-terminal residues as described by the N-end rule pathway (**Fig. 3B**). This is consistent with previous N-terminome (ChaFRADIC) studies in Arabidopsis (Venne *et al.*, 2015). Correspondingly, primary destabilising residues tryptophan, histidine, tyrosine, arginine, isoleucine, leucine, and phenylalanine were observed at much lower frequencies. Proline is not detected at all at P1' or at P1, reflecting its unique structure that may be incompatible with protease active site constrictions. As the N-end rule operates within the cytoplasm, the presence of alternative P1' residues in the N-terminome may represent processing in alternative compartments, or termini buried within the mature protein structure.

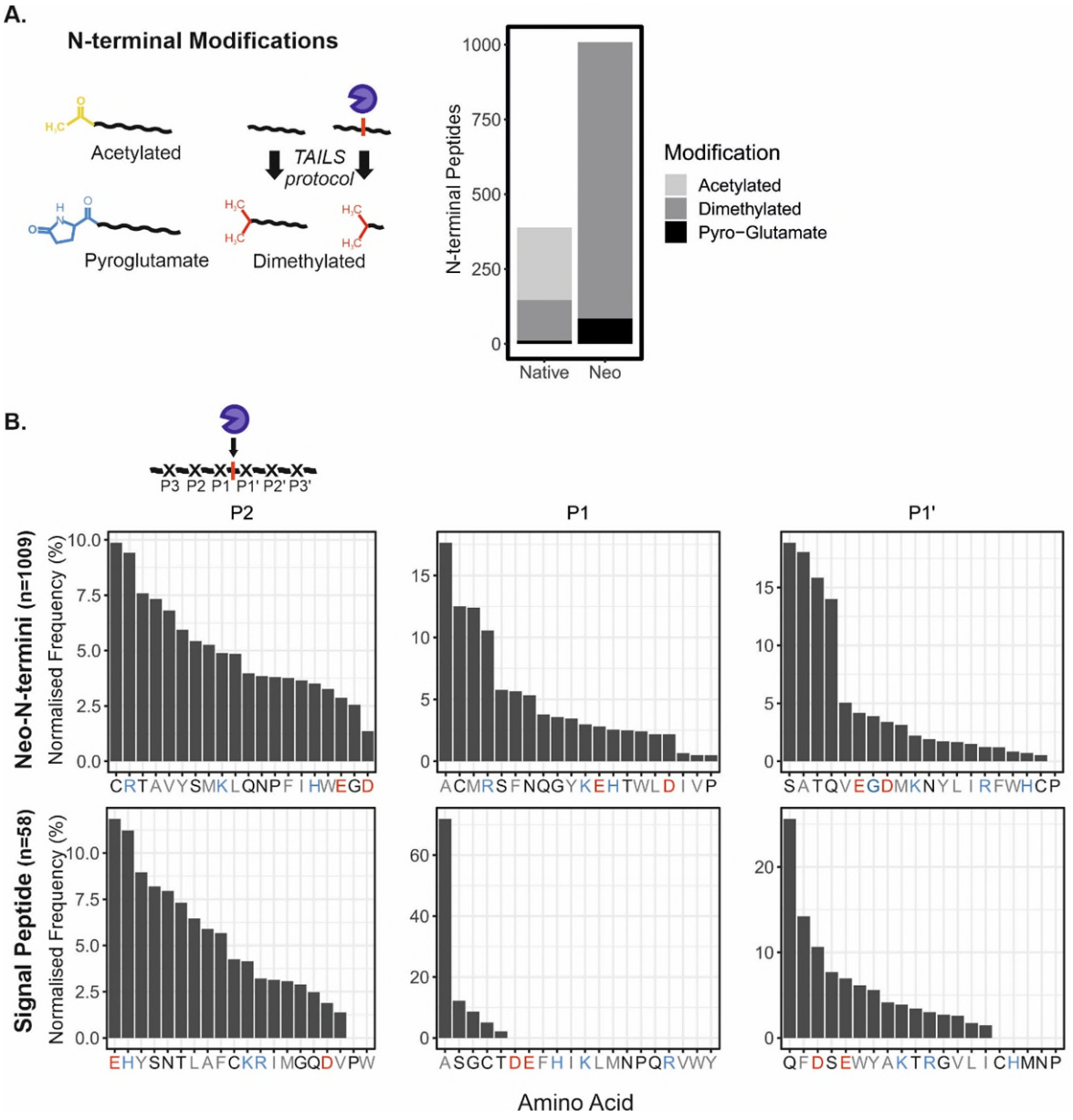


Figure 3. TAILS detected 1009 neo N-termini and 389 native N-termini. A. Peptides with blocked N-termini were detected by TAILS. 389 proteins were identified with native N-termini (mapping to the annotated protein start site or a predicted signal peptide processing event). 1009 neo N-termini were detected. N terminal modifications include acetylation (endogenous), pyroglutamate formation (endogenous or spontaneous) or dimethylation (added in the TAILS protocol). **B.** Frequency of amino acids at positions flanking the cleavage site for Neo N-termini (n = 1009) and signal peptide processing (n=58). Amino acid frequencies were normalised to the abundance of the residue in the proteome.

Detection of peptides in TAILS can report on specific site of processing, if neo-N-termini are generated by proteases. Alanine was the most common P1 amino acid, again consistent

with observations in Arabidopsis (**Fig. 3B**) (Venne *et al.*, 2015). Hydrophobic isoleucine and valine, and bulky proline were among the least common (**Fig. 3B**). The dominance of alanine at P1 is observed for signal peptide processing, as predicted by SignalP (Armenteros *et al.*, 2019). Signal peptide processing has been well studied to follow a predictable sequence, given the limited number of peptidases performing this common cleavage event (Nielsen *et al.*, 1997; Armenteros *et al.*, 2019). However, it is unclear which protease activities could cause the dominance of alanine at P1 of neo-N-termini. Indeed, the vast majority of plant proteases do not have characterised substrate specificities (Van der Hoorn, 2008).

2.3.3 Overlap between TAILS and PROTOMAP

The TAILS dataset overlapped considerably with PROTOMAP. 720 protein groups detected by TAILS (68%) were also identified in PROTOMAP, whereas only 893 (20.1%) of PROTOMAP protein groups were identified (**Fig. 4B**). Differences in data acquisition and analysis between PROTOMAP and TAILS meant composition of protein groups was not identical between datasets. To compare datasets, a protein group was classified as identified in the other experiment if one of the first three protein IDs was identified in a protein group of the other dataset. Similar levels of overlap were also observed in putatively processed protein groups (**Fig. 4A, SFig. 2**). PROTOMAP exceeded TAILS in proteome coverage, detecting more than four times the number of protein groups (**Fig. 4, SFig. 2**). The difference in proteome coverage stems from the greater resolution of protein identification in PROTOMAP due to inclusion of all tryptic peptides, as well as searches performed across a large number of samples in the gel slices. Indeed, gel separation allows

the detection of proteins not tractable in other mass-spectrometry workflows, whether due to low abundance or incompatibility with in-solution digests (ISDs). This is illustrated by 41.7% of the PROTOMAP protein groups identified exclusively by this technique.

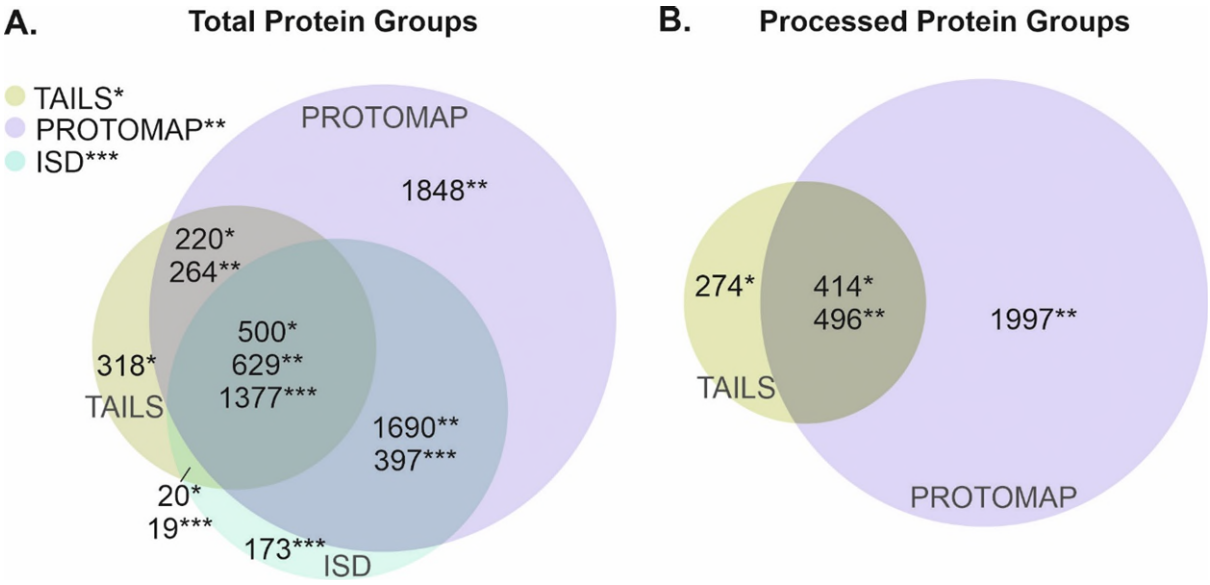


Figure 4. PROTOMAP and TAILS datasets differ in size but share high overlap. The agroinfiltrated *N. benthamiana* leaf proteome was detected by LC-MS/MS with TAILS, ISD (in solution digest) and PROTOMAP. PROTOMAP and ISD protein groups were identified in a minimum of 3 out of 4 biological replicates (independent of gel slices). TAILS protein groups were identified in a minimum of 2 out of 3 biological replicates **A.** Overlap of protein groups detected (numbers differ due to variable number of proteins assigned to each protein group in analysis of the different methods) * = TAILS, ** = PROTOMAP, *** = ISD. **B.** Overlap of processed protein groups identified by TAILS and PROTOMAP. Processed proteins were identified with neo-N termini in TAILS. PROTOMAP proteins were classified as processed if they were detected in a gel slice at least two lower than the gel slice expected to contain the full length protein.

Collectively, the two degradomic techniques identified an additional >120% more protein groups than were detected by conventional ISD alone, whereas only 8.8% of ISD protein groups were not identified in either degradome dataset (**Fig. 4B**). This highlights the power of degradomics to capture more of the protein landscape.

To investigate possible processing events at the level of protein groups, individual plots were made for all identified in TAILS and PROTOMAP. PROTOMAP data is visualised in peptographs that plot peptides detected and MS/MS counts for all peptides of a protein group, at each gel slice (**Fig. 5A**) (Dix *et al.*, 2008). Although software is available for generation of peptographs, this is not compatible with the output from MaxQuant (Dix *et al.*, 2008). Therefore, we generated a new script in R to process PROTOMAP data (**Additional File 1**), using MaxQuant peptide and protein groups files, the proteome fasta file and proteome annotation file generated by InterProScan and SignalP (Jones *et al.*, 2014; Armenteros *et al.*, 2019). The script handles filtering of protein groups and automatic generation of peptographs for each protein group. Additionally, we also generated an R script to map TAILS detected N-termini along the domain structure of the protein for each protein group detected (**Appendix 2**). Examination of the peptographs and N-termini maps revealed many interesting protein processing events. Here we describe several case studies of protein processing detected, or not, with the two experimental techniques (**Table 1**).

Protein	Description	Signal Peptide	PROTOMAP		TAILS		Known Cleavage	PROTOMAP and TAILS agreement	Figure
			Detected	Processed	Detected	Processed			
Niben101Scf03823g01005	NbPME	N	Y	Y	Y	Y	Y	Y	Fig. 5
Niben101Scf01701g00013	NbRD19	Y	Y	Y	Y	Y	Y	Y	Fig. 6
Niben101Scf09885XLOC_077597	NbRD21	Y	Y	Y	Y	Y	Y	Y	Fig. 7
Niben101Scf04989g02003	NbAE (bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1)	N	Y	Y	Y	Y	N	Y	Fig. 8
Niben101Scf04549g10022	NbBGAL1	Y	Y	Y	N	NA	Y	NA	Fig. 9
Niben101Scf04386g03005	a GH127	Y	Y	Y	N	NA	N	NA	Fig. 10
Nicotiana_denovo_assembly_Trancript_113987	NbSCPL49	N	Y	N	N	NA	N	NA	Fig.11
Niben101Scf03963g02006	NbSCPL51	Y	Y	Y	N	NA	Y	NA	Fig.11
Niben101Scf08273XLOC_072456	an S28	Y	Y	Y	N	NA	Y	NA	Fig.11
Niben101Scf01670g02001	GDSL-like lipase	Y	Y	Y	N	NA	N	NA	Fig. 12
Niben101Scf16691XLOC_086857	GH79	Y	Y	Y	N	NA	Y	NA	Fig. 13
Niben101Scf07250XLOC_068369	GH38	Y	Y	Y	N	NA	Y	NA	Fig. 13
Niben101Scf04869g03002	GH17	Y	Y	Y	N	NA	Y	NA	Fig. 13
Niben101Scf02659g02016	A01	Y	Y	N	Y	Y	N	N	Fig. 14
Niben101Scf01372g11012	A01	N	N	N	Y	Y	N	NA	Fig. 14
Niben101Scf01750g15002	A01	Y	N	N	Y	Y	N	NA	Fig. 14
Nicotiana_denovo_assembly_Transcript_5858	A01	N	N	N	Y	Y	N	NA	Fig. 14

Table 1. Summary of proteins discussed in this chapter. Y = yes, N= no, NA = not applicable.

2.3.4 Example 1: Confirmation of known Pectin methylesterase processing

Pectin methylesterases (PMEs) are secreted enzymes vital for cell wall integrity, maintenance and remodelling. PROTOMAP and TAILS identified processing of a putative pectin methylesterase (PME) in agroinfiltrated *N. benthamiana* (**Fig. 5**). A single neo-N-terminal peptide was identified mapping between the inhibitor and catalytic domains creating predicted 30 kDa N-terminal and 34 kDa C-terminal fragments (**Fig. 5B**). The neo-N-terminal peptide detected originates after a basic tetrad sequence RRLL, which is a known and conserved type-I pectinesterase processing site (**Fig. 5C**) (Wolf *et al.*, 2009; Weber *et al.*, 2013). Cleavage at this site is essential for the maturation and secretion of the closest NbPME homologs in *N. tabacum* and *A. thaliana* (Wolf *et al.*, 2009; Weber *et al.*, 2013). Tobacco NtPME-1 shares 96% amino acid similarity with NbPME and its processing described by Wolf et al. is highly similar to that observed in PROTOMAP and TAILS. The NbPME peptograph displays inhibitory and catalytic domains migrating at 18-28 kDa and 32-40 kDa respectively, confirming a cleavage event results in two stable fragments (**Fig. 5A**). Detection of the pectinesterase domain at 32-40 kDa (slice 15) corresponds with TAILS prediction. The inhibitory domain fragment accumulates at smaller MWs than expected (slices >17, >30 kDa) and therefore undergoes further processing. Whilst processing of PMEs has been characterised, the stability of the cleaved N-terminal inhibitor domain had not been reported. The complementation of TAILS and PROTOMAP data for NbPME illustrates the potential of using these techniques in combination.

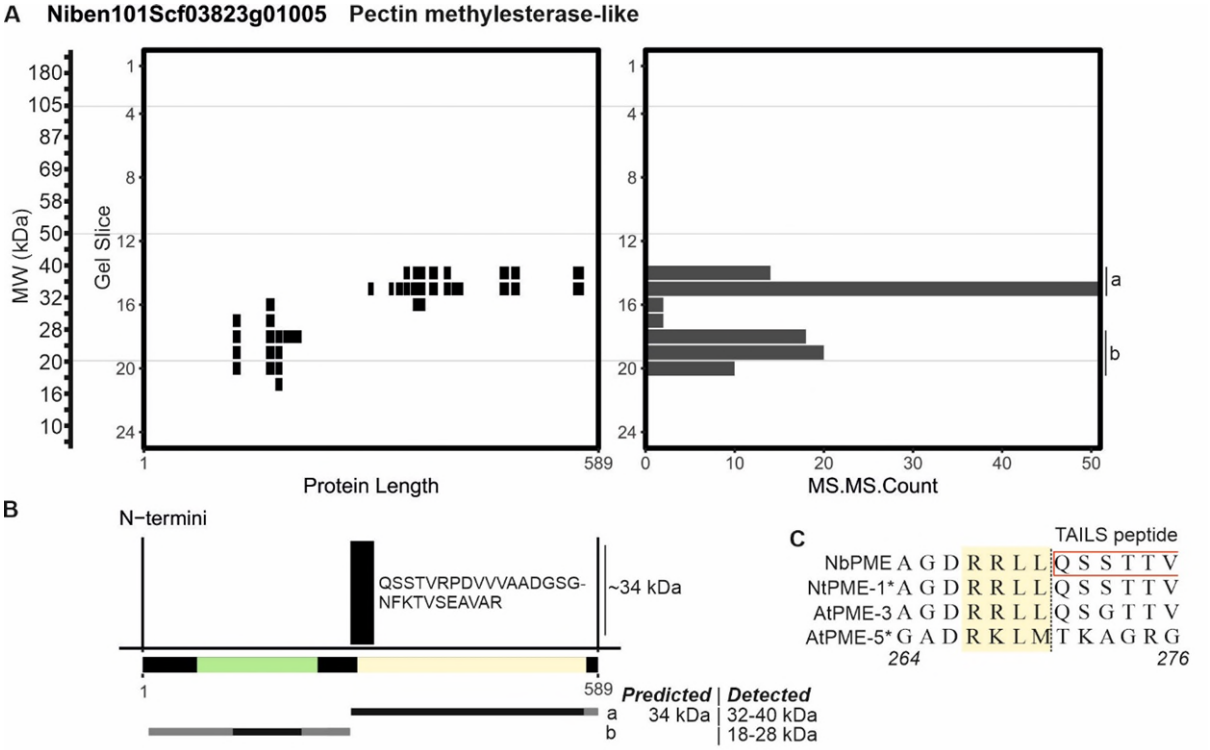


Figure 5. PROTOMAP and TAILS confirm pectin methylesterase processing. A. Peptograph of putative pectin methylesterase (NbPME) demonstrates accumulation of two cleaved products separating the catalytic and inhibitory domains**.** Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. **B.** Neo N-termini detected by TAILS at the start of pectinesterase domain. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from TAILS peptide or annotated start position) and detected MW (PROTOMAP gel slice) indicated. Yellow = PF01095, Pectinesterase. Green = PF04043, Pectin methylesterase inhibitor. **C.** Alignment of NbPME against *Nicotiana tabacum* PME1 and *Arabidopsis thaliana* PME3 closest homologs and AtPME5. * experimentally confirmed cleavage event (Wolf *et al.* 2009). Conserved basic motif preceding cleavage site highlighted in yellow. Residue numbers refer to NbPME sequence.

2.3.5 Example 2: C01 Protease maturation

Proteolytic activation of proteases is a common process across protease classes. Maturation of the papain-like cysteine proteases NbRD19 and NbRD21 was detected in both PROTOMAP and TAILS (**Fig. 6,7**). Neo-N-termini mapping close the start of the catalytic domain were detected by TAILS for both proteases (**Fig. 6B, Fig.**

7B). Cleavage at this site would result in removal of the N-terminal inhibitory I29 domain and formation of a mature, active protein. This cleavage would result in the observed 25-26 kDa and 35-35 kDa proteoforms of NbRD19 and NbRD21 respectively (**Fig. 6B, Fig. 7B**). The peptograph of NbRD19 concurs with I29 domain removal detected by TAILS at 40-45 and 30-32 kDa (slice13,16) albeit at a consistent, slightly higher MW. The increased MW may reflect N-glycosylation at the single putative N-glycosylation site (N-x-S/T) of NbRD19. AtRD19 of Arabidopsis is glycosylated, as is common for extracellular proteins (Gu *et al.*, 2012; Schoberer and Strasser, 2018).

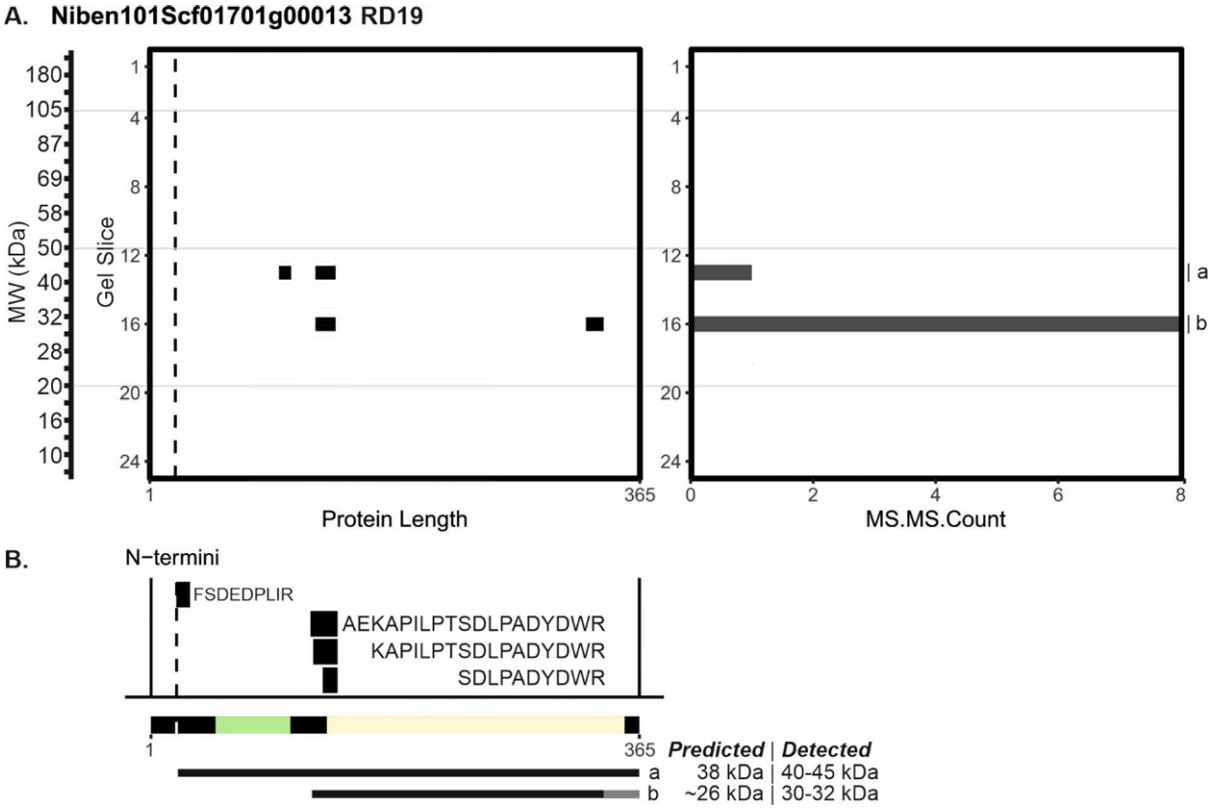


Figure 6. RD19 processing captured by PROTOMAP and TAILS. A. RD19 peptograph show accumulation at molecular weights correlating with I29 inhibitory domain removal. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four

biological replicates. **B.** N-termini detected by TAILS for RD19 map to the annotated N-terminus and the start of the Peptidase_C1 domain confirm I29 inhibitor removal. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from TAILS peptide or annotated start position) and detected MW (PROTOMAP gel slice) indicated. Yellow = PF0112, Peptidase_C1. Green = PF08246, Inhibitor_I29. Blue = PF00396, Granulin. Dashed line = predicted signal peptide.

In contrast, additional proteoforms of NbRD21 to that predicted by TAILS were detected by PROTOMAP. NbRD21 detection at 35-40 kDa (slice 14) correlates with the TAILS predicted cleavage product (**Fig. 7A**). Detection of NbRD21 at lower molecular weights (25-35 kDa, slice 15-18) indicates that further C-terminal processing may occur. Indeed, this is known for Arabidopsis AtRD21 which undergoes removal of the granulin domain. Processing of Arabidopsis and tomato RD21 homologs are well described and match the predicted NbRD21 processing described from activity-based protein labelling in *N. benthamiana,* suggesting that maturation is conserved (Gu *et al.*, 2012; Richau *et al.*, 2012; Shindo *et al.*, 2012; Jutras *et al.*, 2019).

Interestingly, both NbRD19 and NbRD21 had multiple neo-N-termini with staggered start sites detected TAILS (**Fig. 6B, Fi.g 7B**). This could be explained by two scenarios. Firstly, aminopeptidases remove a variant number of residues following protein cleavage. Alternatively, multiple bonds in a stretch of amino acids are cleaved by one, or multiple, proteases. Activation by multiple proteases is already known for some enzymes. For instance, Rcr3 can autocatalytically mature, but is more efficiently activated by subtilase P69B (Paulus *et al.*, 2020). Similarly, SH-EP of *Vigna mungo* can be activated both autocatalytically and *in trans* by VmPE-1 (Okamoto *et al.*, 1999). Of the 741 protein groups

identified with internal N-termini by TAILS, 207 exhibited N-termini with staggered start sites. Therefore, heterogeneity in protein N-termini is a fairly common occurrence in *N. benthamiana* leaves.

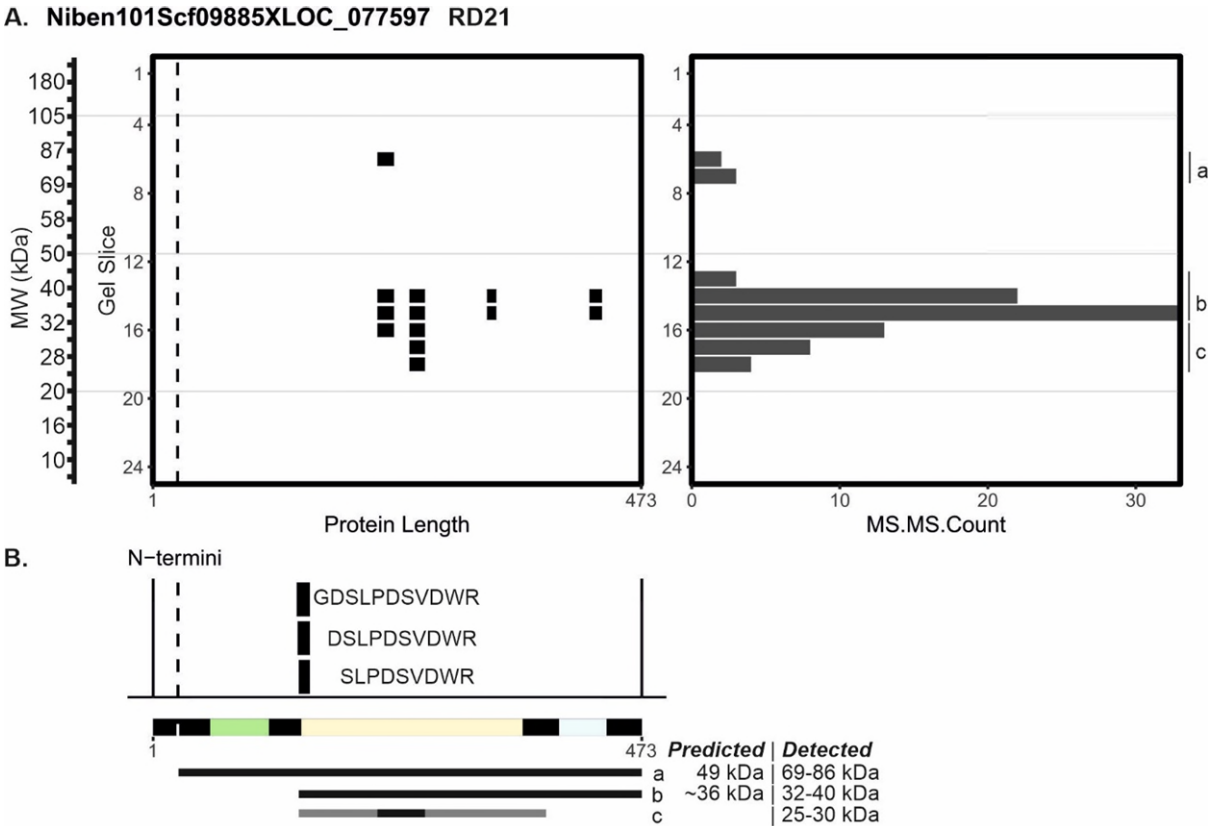


Figure 7. RD21 processing detected by PROTOMAP and TAILS. A. RD21 peptograph show accumulation at molecular weights correlating with I29 inhibitory domain removal. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. **B.** RD21 N-termini detected by TAILS map to the start of the Peptidase_C1 domain and confirm I29 inhibitor removal. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from TAILS peptide or annotated start position) and detected MW (PROTOMAP gel slice) indicated. Yellow = PF0112, Peptidase_C1. Green = PF08246, Inhibitor_I29. Blue = PF00396, Granulin. Dashed line = predicted signal peptide.

2.3.6 Example 3: Bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase

PROTOMAP and TAILS detected novel processing of a predicted bifunctional metabolic enzyme, methylthioribulose-1-phosphate dehydratase/enolase-phosphatase (NbAE), with a potential role in methionine biosynthesis pathways (**Fig. 8**). The peptograph reveals accumulation of a probable full length enzyme (55-58 kDa, slice 10) and a cleaved C-terminal enolase-phosphatase domain predominantly present at 28-30 kDa (slice 16-17) (**Fig. 8A**). This processing event matches the 28 kDa product predicted from the neo-N-terminus (**Fig. 8B**). This peptograph is representative of two highly similar protein groups detected. Separation of the catalytic domains may have regulatory consequences for the Yang cycle/methionine salvage pathways. Another bifunctional metabolic enzyme, 6-Phosphofructo-2-Kinase/Fructose-2,6-Bisphosphatase (PFKFB3), is known to be synthesised in separated isoforms in a tissue specific manner to regulate metabolic fluxes (Atsumi *et al.*, 2005). The separation of NbAE has not previously been reported in plants and represents a novel insight generated from the PROTOMAP and TAILS datasets.

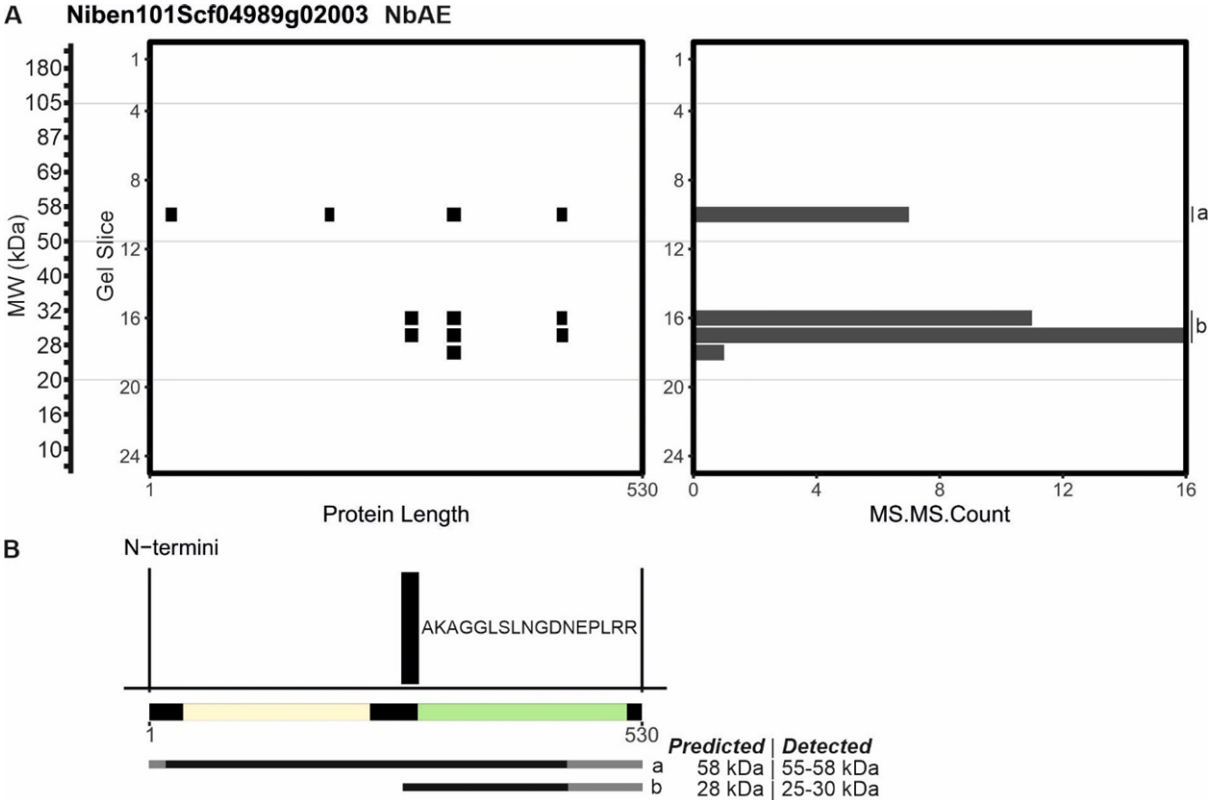


Figure 8. Novel processing of a bifunctional metabolic enzyme detected in PROTOMAP and TAILS. A. Peptograph of a putative bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1 (NbAE) indicates processing occurs to generate a stable C terminal enolase phosphatase. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. **B.** Neo N-termini detected by TAILS maps to residues immediately preceding the enolase phosphatase domain. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from TAILS peptide or annotated start position) and detected MW (PROTOMAP gel slice) indicated. Yellow = PF00596, Aldolase II (PFAM). Green = IPR0239443, Enolase-phosphatase E1 (Interpro).

2.3.7 Example 4: Release of carbohydrate binding domains

Glycosyl hydrolases commonly contain carbohydrate binding modules (CBMs) that can enhance co-localisation with substrates. PROTOMAP identified events showing separation of catalytic domains from CBMs, as illustrated by peptographs of GH35 and GH127 glycosidases (**Fig. 9, Fig.10**). The GH35, beta galactosidase (BGAL1), was detected at three MWs (**Fig. 9A**). The largest proteoform at 69-87 kDa (slice 4) represents the full

length protein without the signal peptide. The catalytic domain is detected as a 40-50 kDa proteoform (slice 12-13) lacking the C-terminal CBM. Transiently expressed BGAL1 with a N-terminal StrepII tag following the signal peptide indeed displays the ~45 kDa accumulating in apoplastic fluid (AF) (**Fig. 9B**). This data correlates well with the 45 kDa active form detected by activity-based labelling of BGAL1 with the glycosidase JJB70 probe in AF (Chandrasekar *et al.*, 2014; Buscaill *et al.*, 2019). The full length proteoform (69-87 kDa, slice 6) was not detected in the western blot, suggesting that the overexpression construct is more efficiently processed than the endogenous protein. Interestingly an additional higher ~50 kDa band was also detected in full leaf extract (FLE), which could represent the 45-50 kDa (slice 12) proteoform. Another middle to C-terminal fragment is detected in a smear of 28-40 kDa, that may arise from further post-translational modification. The CBM is probably not stable after cleavage as peptides within this region are detected exclusively in the full-length proteoform (**Fig. 9A**).

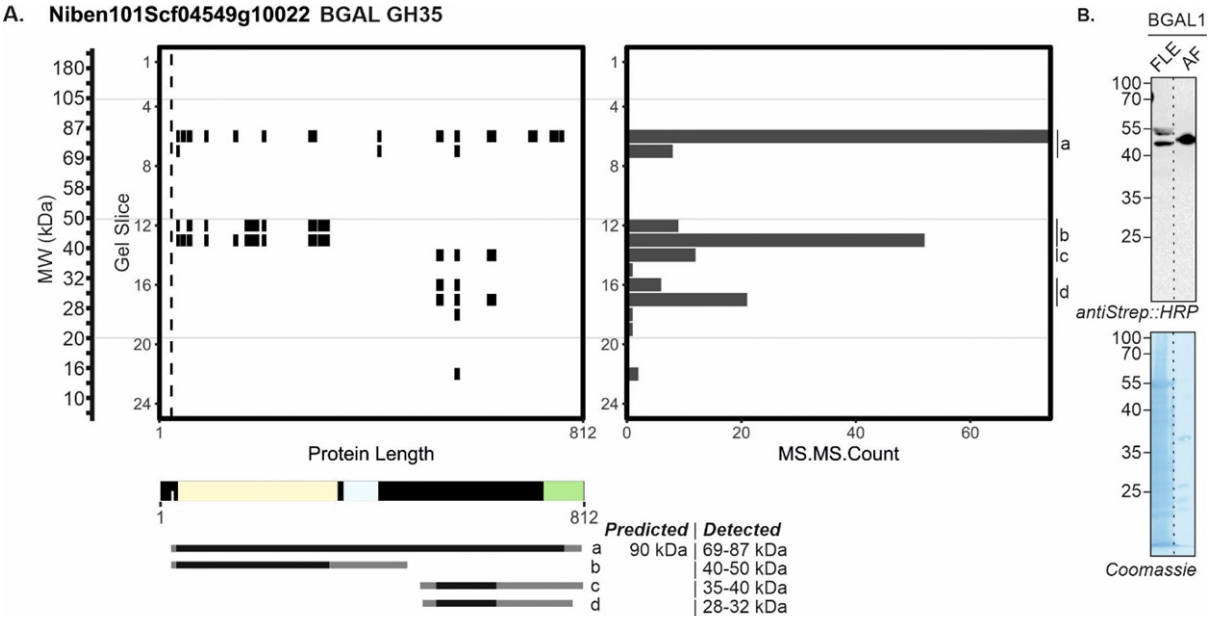


Figure 9. BGAL1 peptograph shows separation of catalytic and carbohydrate binding domains. Peptographs of the GH35 glycosidase BGAL1 exclusively detected in PROTOMAP revealed processing separating the catalytic domain from the carbohydrate binding domain. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from annotated signal peptide) and detected MW (PROTOMAP gel slice) are indicated. Yellow = PF01301, GH35. Green = PF02140, galactose binding lectin domain. Blue = PF17834, beta galactosidase beta sandwich. **B.** N-terminally Strep tagged (after a PR1 signal peptide) BGAL1 was transiently overexpressed by agroinfiltration in mature *N. benthamiana* leaves. Full leaf extract (FLE) or apoplastic fluid (AF) was harvested at 4 dpi. Proteins were detected by Western blotting.

The putative beta-l-arabinofuranosidase GH127 peptograph also demonstrates CBM separation. The longest proteoform accumulated at 78-87 kDa (slice 6), slightly lower than the expected MW following signal peptide processing, indicating additional N-terminal trimming (**Fig. 10**). A smaller 62-68 kDa (slice 8) proteoform containing the catalytic domain and a corresponding C-terminal 13–18 kDa (slice 21-22) fragment with peptides spanning the CBM were detected. Unfortunately, no neo-N-termini for these glycosidases were detected in TAILS that might have revealed the exact site of the putative processing events.

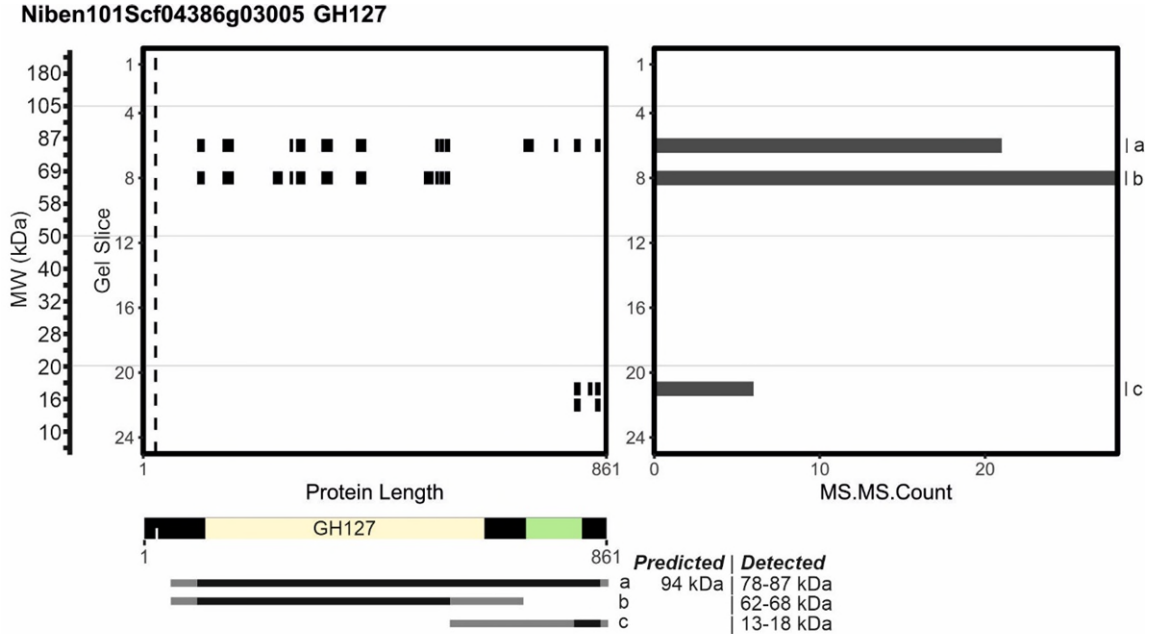


Figure 10. PROTOMAP detected processing of a GH127. An unknown GH127 exclusively detected in PROTOMAP demonstrated processing separating the catalytic domain from the carbohydrate binding domain. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for possible sequence included in this molecular weight (MW) proteoform. The predicted MW (from annotated signal peptide) and detected MW (PROTOMAP gel slice) are indicated. Yellow = PF07944, GH127 (PFAM), green = IPR036195, arabinose binding domain (Interpro).

Removal of the CBM detected by PROTOMAP may affect glycosidase activity and raises several questions. Is the CBM removed after targeting of the glycosidase to the substrate? Could the ratio of intact to cleaved CBM report on activity regulation *in planta*? Furthermore, what is the function of dissociated CBM if it remains stable in the apoplast?

2.3.8 Example 5: Variation in serine-carboxypeptidase processing

PROTOMAP also revealed how proteases containing similar domain architecture undergo divergent processing. Two serine-carboxypeptidase like proteases, NbSCPL49 (**Fig. 11A**) and NbSCPL51 (**Fig. 11B**) display strikingly different accumulation patterns. SCPLs lack

an inhibitory pro-domain that must be cleaved for activation. This is illustrated by NbSCPL49, which was detected at 45-50 kDa (slice 12) (**Fig. 11A**), with peptides spanning the protein length expected to generate a 44 kDa protein (**Fig. 11A**)**.** Labelling with the serine protease probe FP in tissue overexpressing NbSCPL49 (NbCPY-GFP) demonstrates the MW detected in PROTOMAP corresponds to an active protease (**Chapter 4, Fig. 2**)**.** This is surprising because the closest homolog in Arabidopsis, SCPL49 (AtCPY) undergoes VPE-dependent processing (Rojo *et al.*, 2003)**.**

In contrast, NbSCPL51 migrates on SDS PAGE as two stable proteoforms. S10 catalytic activity is dependent on a catalytic Ser, His, Asp triad (Van der Hoorn, 2008). Detected peptides mapping to the protein indicate a separation of the catalytic His and Asp residues in the C terminal fragment at 16-25 kDa (slice 19-21) from the Ser in a N-terminal 30-35 kDa (slice 15-16) fragment (**Fig. 11B**). Overexpression of a C-terminally Myc tagged NbSCPL51 construct corroborated the accumulation of a C-terminal ~20 kDa fragment (**Fig. 11C**). Therefore the SCPL51 protein is either inactive, does not require the canonical triad for activity, or the processed fragments remain closely associated to retain catalytic activity. A yeast SCPL (SpCPY) is active after proccessing between catalytic residues, which remain closely associated due to disulphide bonds between the two chains (Mukaiyama *et al.*, 2011). PROTOMAP indicated that the cleavage event is highly efficient as the full length protein was not detected (**Fig. 11B**). The detection of a ~60 kDa signal (larger than expected for the full length protein) upon NbSCPL51-Myc overexpression suggests high levels of expression can saturate the protease(s) responsible for NbSCPL51 processing (**Fig. 11C**).

A different class of serine carboxypeptidase, an S28 carboxypeptidase, similarly demonstrates processing that results in catalytic residue separation onto multiple peptide chains. The peptograph for the S28 shows coverage of almost the full protein length at 50-58 kDa (slice 10-11) (**Fig. 11D**). The protease is then further processed to separate the catalytic Ser from the Asp and His residues. The detection of multiple intermediate forms across a range of MWs indicates the S28 undergoes even more complex processing than NbSCPL49 and NbSCPL51.

These serine carboxypeptidase examples illustrate how homology in domain architecture and sequence does not necessarily correlate with proteoform accumulation, and that degradome techniques are necessary to understand protein accumulation *in planta*. It will be interesting to express these proteins and perform activity-based labelling using FP probe to investigate which MWs correspond to active proteases.

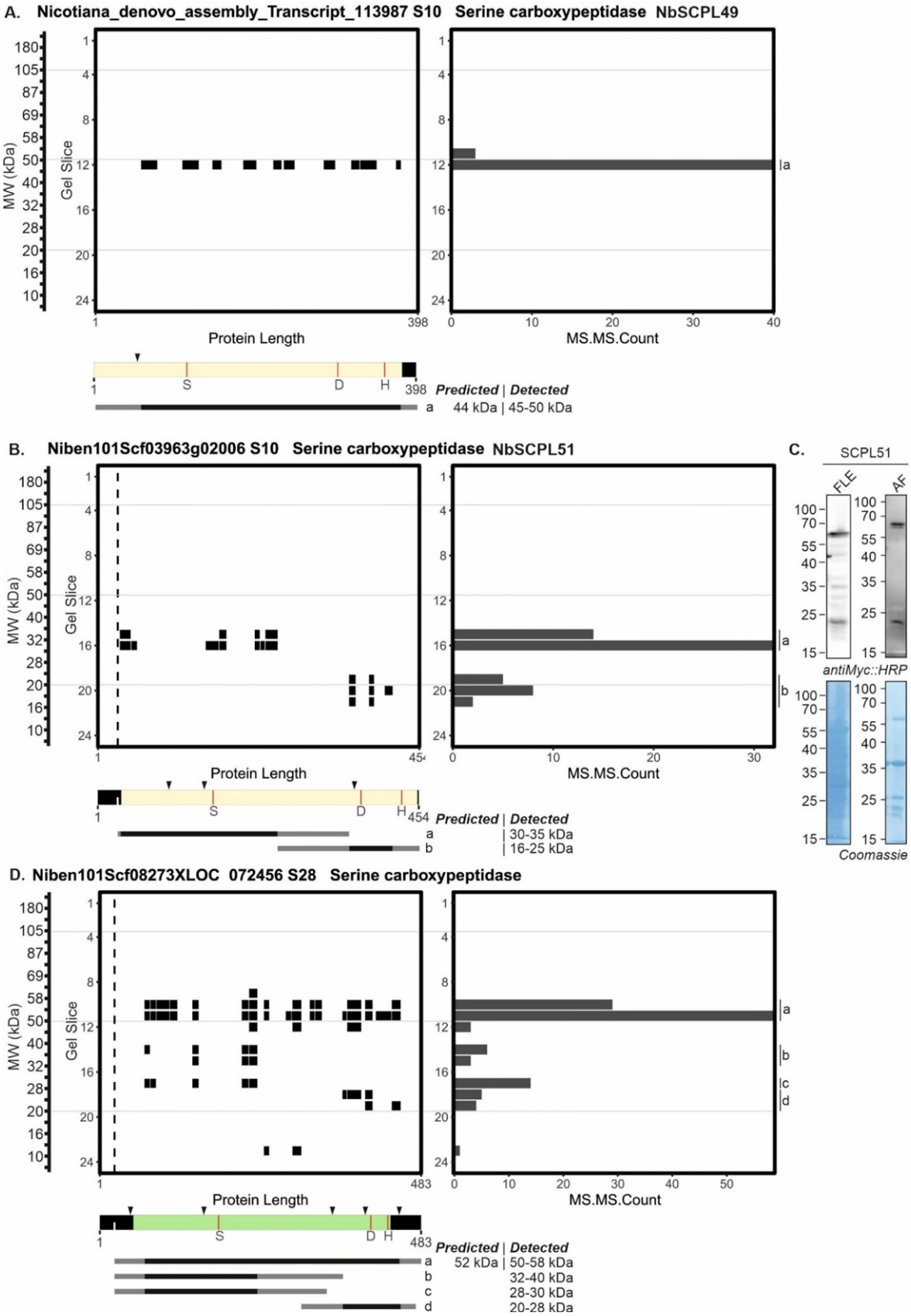


Figure 11. Serine carboxypeptidases display distinct processing patterns. Peptographs from **A** an SCPL49-like protease here named NbCPY, **B** an SCPL51-like and **D** an S28 protease exhibit different processing profiles. Left: unique peptides detected in each gel slice.

Right: total MS/MS count across the four biological replicates. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for probable sequence included in this molecular weight (MW) proteoform. The predicted MW (from annotated signal peptide or N-terminus) and detected MW (PROTOMAP gel slice) are indicated. Yellow = PF00450, Peptidase_S10. Green = PF05577, Peptidase_S28. Red lines mark catalytic residues. Black arrows mark NxS/T motifs. **C.** C-terminally Myc tagged SCPL51 was transiently overexpressed by agroinfiltration in mature *N. benthamiana* leaves. Full leaf extract (FLE) or apoplastic fluids (AF) was harvested at 4 dpi. Proteins were detected by western blotting with the antiMyc antibody.

2.3.9 Example 6: Novel processing of a GDSL-like lipase

PROTOMAP also detected processing of a GDSL-like lipase in a previously undescribed pattern. GDSL lipases catalyse the cleavage of ester bonds using a catalytic triad that is reminiscent of serine hydrolases. The GDSL-like lipase peptograph revealed accumulation as a probable full length protein at 40-45 kDa (slice 13), close to the expected MW of 40 kDa (**Fig. 12A**). An additional C-terminal fragment is detected at 28-32 kDa (slice 16-17) and an N-terminal portion at 13-18 kDa (slice 22-23) which suggests a single cleavage event separating the catalytic triad (**Fig. 12A**). Similar to the serine proteases, catalytic activity of GDSLs is dependent on a catalytic Ser, Asp and His triad (Ding *et al.*, 2019). The GDSL lipase-like protein was cloned with a C-terminal Myc tag and expressed by agroinfiltration (**Fig. 12B**). The myc tagged protein accumulated at the expected proteoforms in AF. In contrast to the peptograph, the larger proteoform was more abundant on the western blot. This may be an artefact of the overexpression that saturates the capacity to mature the GDSL-like lipase. The processing observed in agroinfiltrated *N. benthamiana* contrasts to observations for a similar GDSL lipase of *Rauvolfia verticillate* that possesses activity at a

MW corresponding to the full length protein, excluding the signal peptide (Ruppert *et al.*, 2005). No examples of similar protein processing were detected in a search of the plant literature at the time of writing.

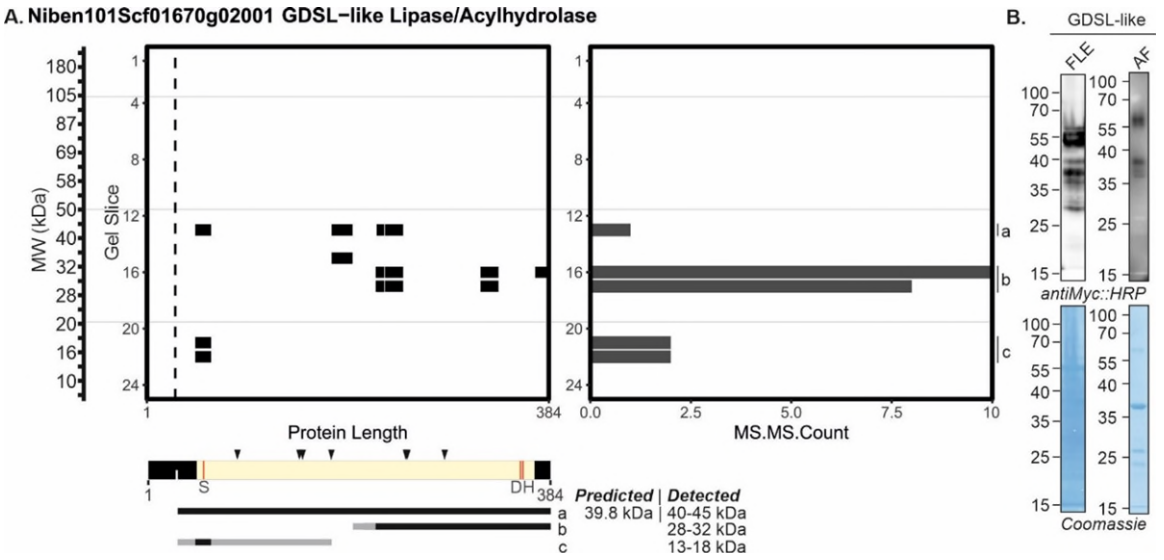


Figure 12. N-terminal processing of a GDSL-like lipase. A. The peptograph of a GDSL-like lipase indicates a cleavage of a N terminal portion of the protein. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for probable sequence included in this molecular weight (MW) proteoform. The predicted molecular weight (MW) (from annotated signal peptide) and detected MW (PROTOMAP gel slice) indicated. Yellow = PF00657, Lipase_GDSL. Red lines mark catalytic residues. Black arrows mark NxS/T motifs. **B.** C-terminally Myc tagged GDSL-like was transiently overexpressed by agroinfiltration in mature *N. benthamiana* leaves. Full leaf extract (FLE) or apoplastic fluid (AF) was harvested at 4 dpi. Proteins were detected by Western blotting.

2.3.10 Example 7: Variation in Processing of Glycosyl Hydrolases

A GH79 glycosidase was detected to undergo N-terminal processing (**Fig. 13A**). N-terminal processing of a human GH79 heparanase exposes the substrate binding cleft and is essential for activation (Wu *et al.*, 2015; Vlodavsky *et al.*, 2016). A proteoform at 55-62 kDa (slice 12-9) is probably the full length protein. N-terminal peptides are additionally detected in a 16-20 kDa (slice 20-21) proteofrom and a corresponding proteoform at 30-35 kDa (slice 16-

17) is likely to contain both the catalytic Glu residues. It will be interesting to see if this processed proteoform is active *in planta*. Previous activity labelling of another GH79 in *N. benthamiana* was detected at a MW slightly larger than the expected full length protein (Chandrasekar *et al.*, 2014), and therefore maturation of GH79s may not be conserved across the class.

The peptograph of a putative GH38 alpha-mannosidase indicates putative cleavage of a ~100 kDa protein into two stable distinct fragments of 62-78 kDa (slice 6-7) and 40-50 kDa (slice 12-13) (**Fig. 13B**). A similar profile was observed for a GH38 paralog (**SFig. 2**). GH38s are known to be processed into multiple chains in a wide range of eukaryotes (Hossain *et al.*, 2010). Processing we observed here correlates well with that of tomato alpha mannosidase, which accumulates as two separate polypeptides (Hossain *et al.*, 2009), suggesting that GH38 processing is conserved.

Finally, a GH17 hydrolase was detected to accumulate stably in at least two proteoforms. The largest proteoform detected at 30-45 kDa (slice 13-14) matched the predicted MW of 36 kDa following signal peptide processing. A 5-20 kDa portion is then removed from the C-terminus. C-terminal processing of a GH17 has also been reported in *N. tabacum* (Shinshi *et al.*, 1988) and may be a conserved event. It will be interesting to determine if processing corresponds to a change in activity state.

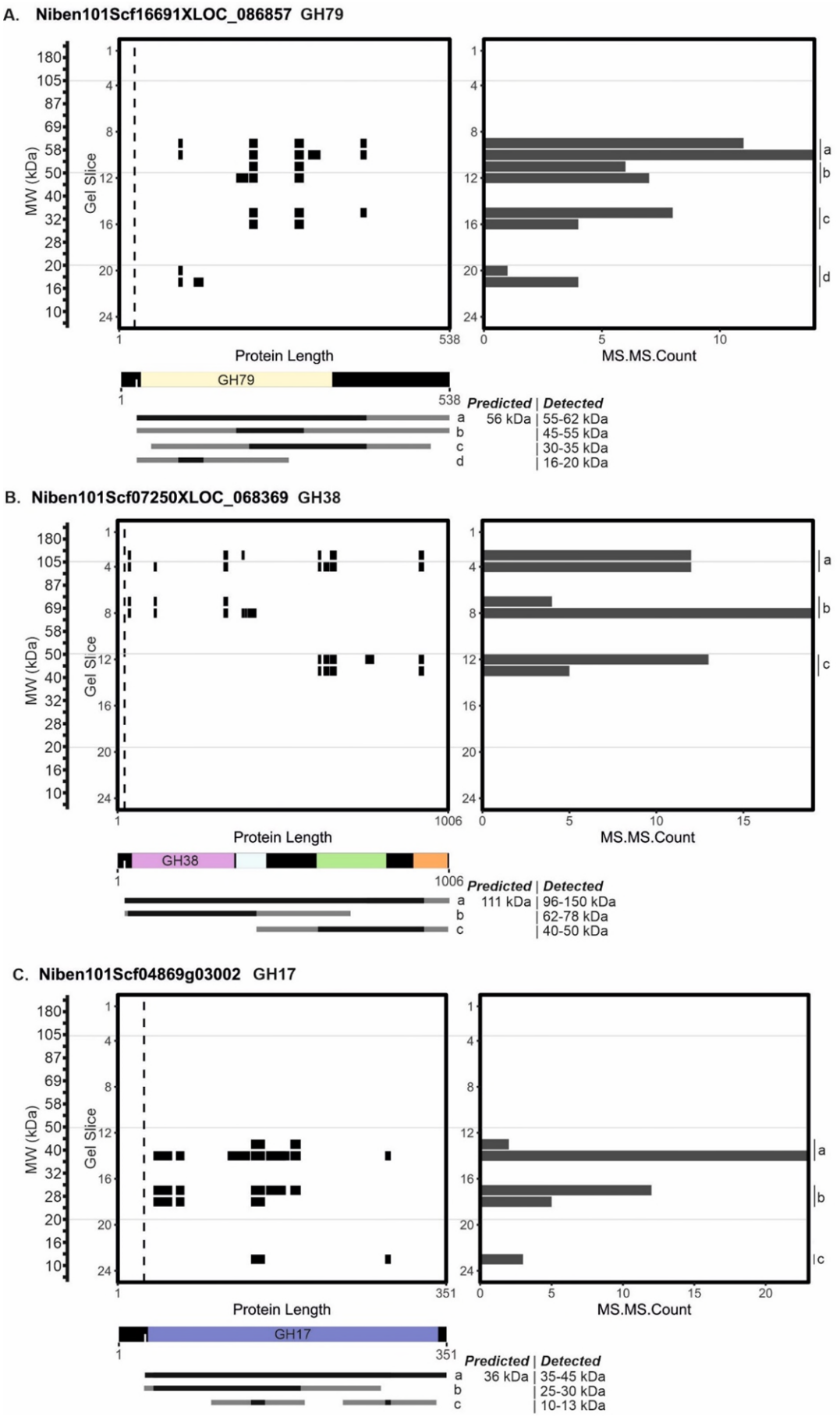


Figure 13. Different protein processing of glycosidases. Example peptographs of **A.** GH79

B. GH38 and **C.** GH17 glycosidases. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. Summary of proteoforms identified is shown below, in a black bar for confirmed detected sequence and in grey for probable sequence included in this molecular weight (MW) proteoform. The predicted Mw (from annotated signal peptide) and detected Mw (PROTOMAP gel slice) indicated. Purple = PF01074 GH38 N terminal. Green = PF07748, GH38 C terminal. Blue = PF09261, alphamannosidase middle. Orange = PF17677, GH38 C terminal beta sandwich. Yellow = PF03662, GH79 N terminal. Violet = PF00332, GH17.

2.3.11 Example 8: Low abundance A1 proteases enriched in TAILS

Whilst PROTOMAP did identify a greater number of protein groups, there were also processing events captured exclusively by TAILS. This is illustrated by some aspartic proteases, which failed to be detected by ISD and therefore are probably very low abundant proteins (**Fig. 14**). Enrichment of low abundance proteins is a commonly cited advantage of N-terminomic techniques (Schepmoes *et al.*, 2009; Rogers and Overall, 2013; Perrar *et al.*, 2019). In the case of one atypical aspartic protease, one unique peptide was detected in PROTOMAP at 45-50 kDa (slice 12) whilst two overlapping neo-N-termini were identified in TAILS (**Fig. 14A**). A further three putative A1 proteases (**Fig. 14C,D,E**) were not detected in PROTOMAP nor ISD. TAILS identified putative N-terminal processing close to the start of the PF14543 domain of atypical aspartic proteases or the PF00026 domain of the typical aspartic protease. The structure and maturation of A1 proteases is complex and divergent, with 'atypical' types that have variable structures whilst 'typical' proteases have been reported to undergo C-terminal processing to remove inhibitory regions (Faro and Gal, 2005; Soares *et al.*, 2019), C-terminal processing was not detected with TAILS, as expected if the released fragment is degraded.

Importantly, these examples are not representative of all aspartic proteases. In fact, more A1 protein groups were detected in PROTOMAP compared to TAILS (**SFig. 2**). This example does, however, illustrate that TAILS can capture a distinct fraction of the proteome.

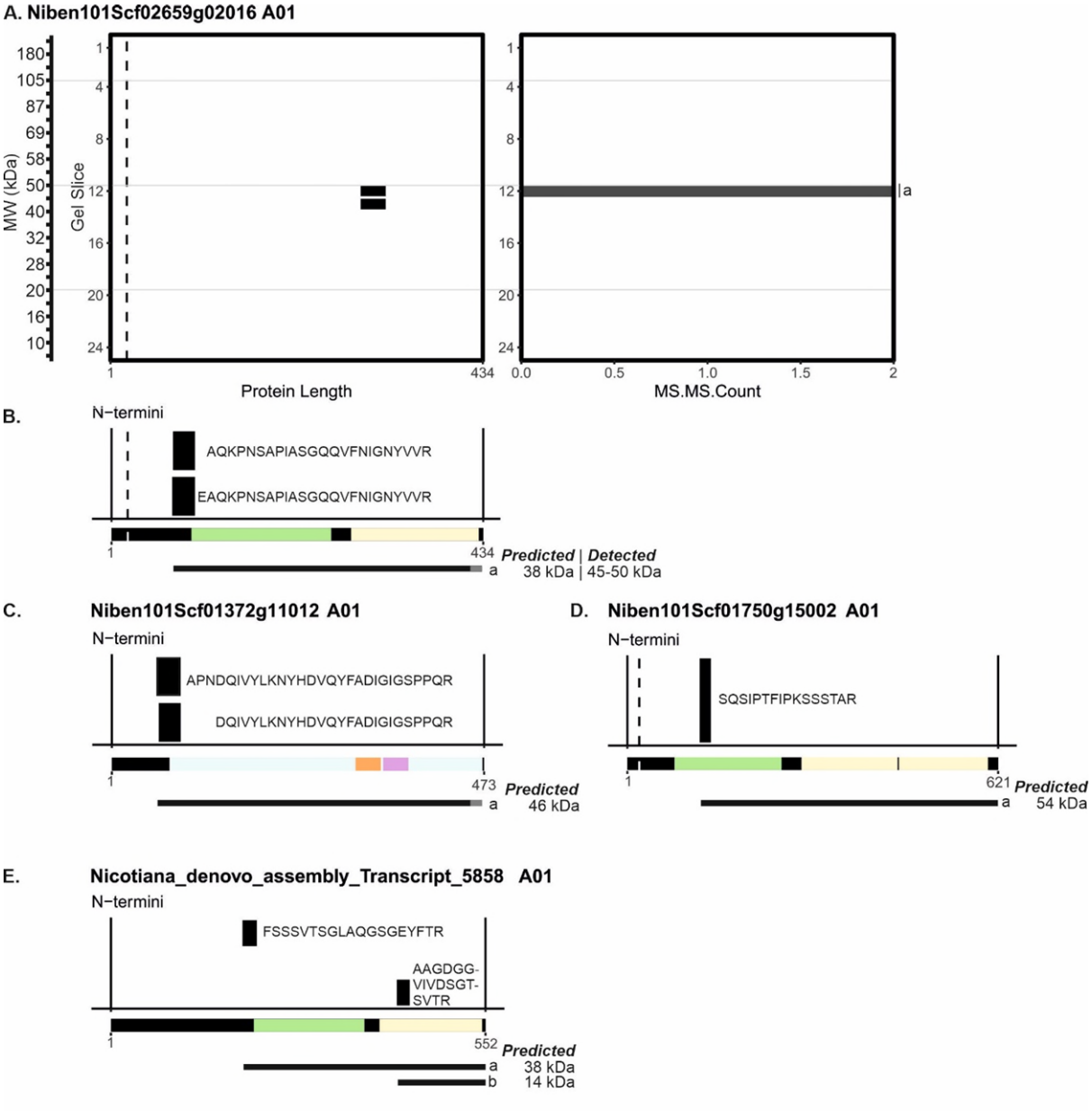


Figure 14. TAILS captures proteins not detected by ISD or PROTOMAP. A Detection of an aspartic protease is enriched in TAILS compared to PROTOMAP and not detected in ISD. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. **B** N-termini detected by TAILS. **C,D,E.** Example N-termini detected by TAILS of protein groups not identified by either PROTOMAP or ISD.

Predicted molecular weight of the proteoform stating at this N-terminus is indicated to the right. Summary of proteoforms identified is shown. The predicted MW (from TAILS peptide) is indicated. Yellow = PF14541, Xylanase inhibitor C term. Green = PF14543, Xylanase inhibitor N term. Blue = PF00026, Eukaryotic aspartyl protease. Orange = PF03489, Saposin-like 2B. Purple = PF05184, Saposin like B.

2.4 Discussion

N.benthamiana is a cornerstone of plant research due to the ease and versatility of transient protein expression by agroinfiltration. The development of tools for protein research in *N. benthamiana* is important to understand the underlying biological context. Here we present the first study of the proteoforms accumulating in agroinfiltrated *N. benthamiana* leaves with two contrasting degradome techniques, TAILS and PROTOMAP.

Altogether, a total of 2767 processed protein groups were identified from a total of 4769 identified protein groups across the two techniques. Within this, 496 protein groups were identified as processed in both TAILS and PROTOMAP, representing 60% of the TAILS proteins processed. In a previous overlap study of PROTOMAP and N-terminome capture in human cells achieved a comparable 66% of neo-N-termini also detected as PROTOMAP processed proteins (Simon *et al.*, 2009).

The use of multiple techniques in parallel is rare due to the expense of mass spectrometry and the expertise required (Perrar *et al.*, 2019). The power of using complementary approaches is evident from our data. Firstly, combining the two techniques greatly increases the coverage of the proteome. PROTOMAP identified the majority of protein groups, but TAILS allowed enrichment of additional proteins. Previous studies illustrate the strength of TAILS in enriching low abundant proteins and even to identify previously 'missing' proteins of the human proteome (Lange *et al.*, 2014; Prudova *et al.*, 2014).

Furthermore, as shown in this chapter and in previous studies, combining the precise sequence data of N-termini and the MW outputs of peptographs is a powerful approach to investigate protein processing in plants. N-terminomics offers unparalleled ability to detect sequence specific definition of processed proteoforms, but would miss events with unstable new N-termini. Furthermore, there is increased variability in quantifying single peptides and possibly higher stochasticity in detection as different methods (i.e. TAILS-based, COFRADIC, ChaFRADIC or PTAG) have previously displayed limited overlap (Chen *et al.*, 2016). PROTOMAP overcomes several of these limitations through recording the topography of the protein fragments by detecting canonical tryptic peptides, and enabling abundance to be crudely inferred through spectral counts. Here we identified a processing event in pectin methylesterases known in other species (Wolf *et al.*, 2009; Weber *et al.*, 2013). The processing of parental proteoforms into two stable fragments was reported by PROTOMAP whilst the exact processing site was identified by TAILS. Our approach similarly uncovered potential new proteoforms, such as the separation of bifunctional metabolic enzymes and added site-specific information to known cysteine protease activation events. The use of contrasting techniques increased the confidence in processing events observed in both techniques. Previous combined use of N-terminomics (Subtiligase) and PROTOMAP in human cells demonstrated that in isolation, PROTOMAP analysis may be too conservative. However in conjunction with N-terminal data from Subtiligase a number of peptographs with lower confidence could be interpreted and revealed further processing events (Johnson and Kornbluth, 2008; Simon *et al.*, 2009).

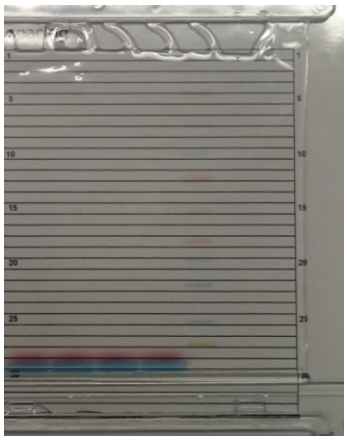
PROTOMAP also uncovered limitations of the *N. benthamiana* proteome used. In comparison to early PROTOMAP studies in human cells, the correlation between average gel slice detection and MW was much weaker (Dix *et al.*, 2014). For example, peptides detected from small proteins were often detected at higher MW than expected (**SFig. 3A**). Although some degree of gel retardation is expected for high abundance proteins such as RuBisCO, for others it likely represents incorrect genome annotation (Simon *et al.*, 2009) (**SFig 3B**). Indeed, curation of the *N. benthamiana* proteome is still in its infancy. Since this study was performed, advances in gene prediction have increased the percentage of identified MS/MS spectra from a limited 28% to 33% (Kourelis *et al.*, 2019). Clearly this is still very low and there remains a huge amount of data to be exploited from these experiments when *in silico* tools are improved. In fact, this dataset could potentially feed into future computational studies for detection of relevant alternative translation initiation sites or splicing site studies. Inaccuracy of current protein annotation may have biased towards higher detection of neo-N-termini in TAILS. However, containing 72% neo-N-termini, this dataset compared well to previous experiments in mammals and Arabidopsis. The Arabidopsis leaf N-terminome has a reported 64% neo-N-termini (Venne *et al.*, 2015). In humans, this figure is variable between 68% in erythrocytes and 77% in the most comprehensive N-terminome coverage to date (Lange *et al.*, 2014; Yeom *et al.*, 2017). Variation also exists between tissues, compartments and upon stress induction, where increased proteolysis has been observed (Auf Dem Keller *et al.*, 2013). In the absence of agroinfiltration, the proportion of N-termini detected in unchallenged leaves may in fact be lower.

The PROTOMAP and TAILS dataset is a snapshot of what is likely to be a highly dynamic landscape. Nevertheless, the resource generated here will support future agroinfiltration studies in *N. benthamiana*. Understanding the environment in which a protein is expressed is important for interpreting results. The study of native proteins in particular can be aided by the peptographs and N-termini detected. Heterologously expressed proteins may not always undergo correct maturation to form a functional protein. Fusion to tags may inhibit maturation, or the artificial overexpression saturating chaperones or proteases that are necessary for maturation. Comparison to proteoforms detected in this study can inform whether the protein is correctly processed and biologically relevant. This tool is a useful benchmark for future studies that could provide insights into how proteoform accumulation adapts in response to different stresses and conditions.

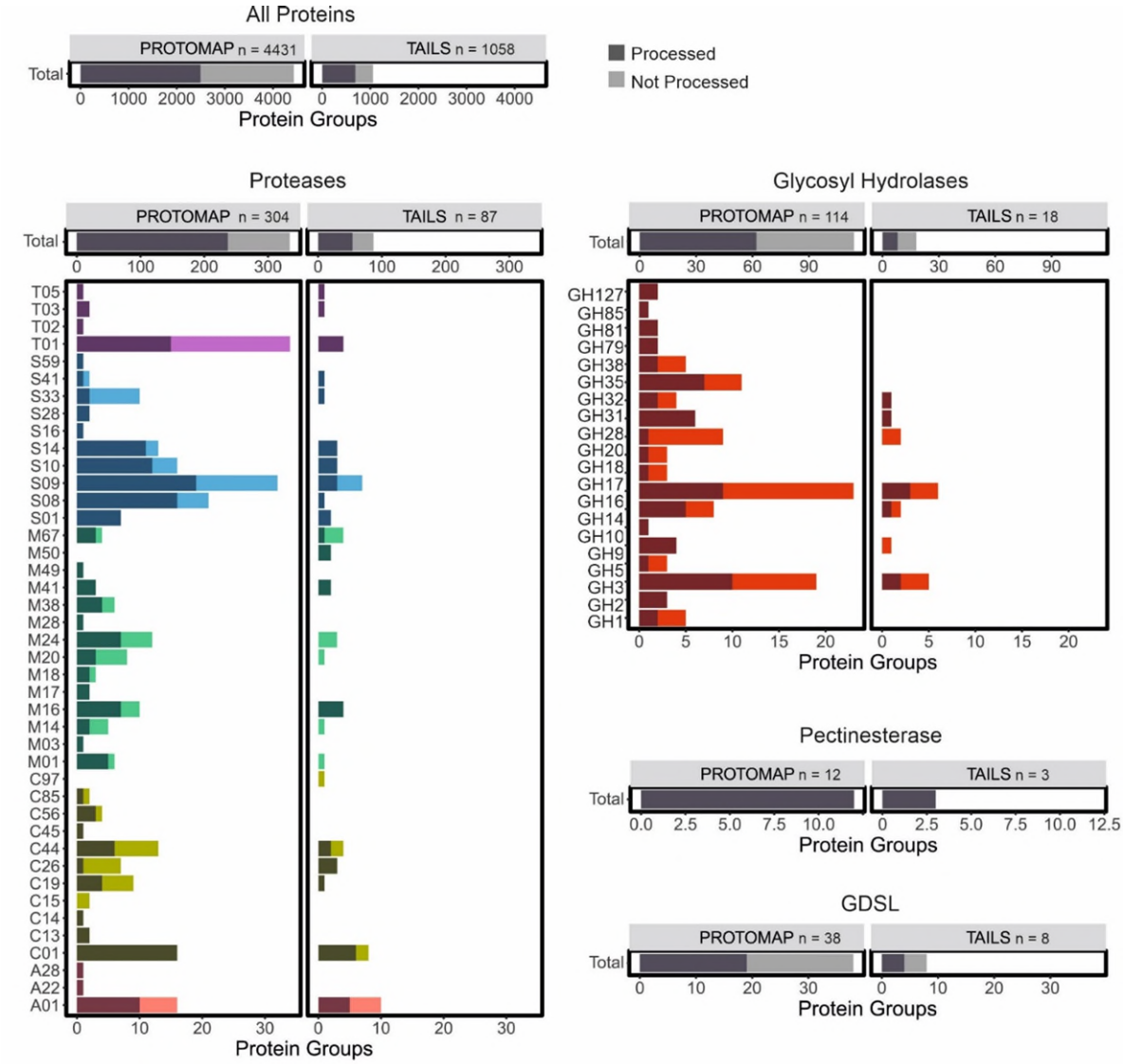
Contributions

Prof. Renier van der Hoorn conceived the research and contributed to experiment design and data interpretation. Dr Farnusch Kaschani and Prof. Markus Kaiser provided mass spectrometry expertise, equipment, materials and performed spectra searches for PROTOMAP and ISD. Dr Sabrina Ninck provided technical assistance and training in performing PROTOMAP and ISD. Prof. Pitter Huesgen provided mass spectrometry expertise, equipment, materials and data analysis for TAILS. Dr. Fatih Demir and Andreas Perrar provided technical assistance and training for performing TAILS.

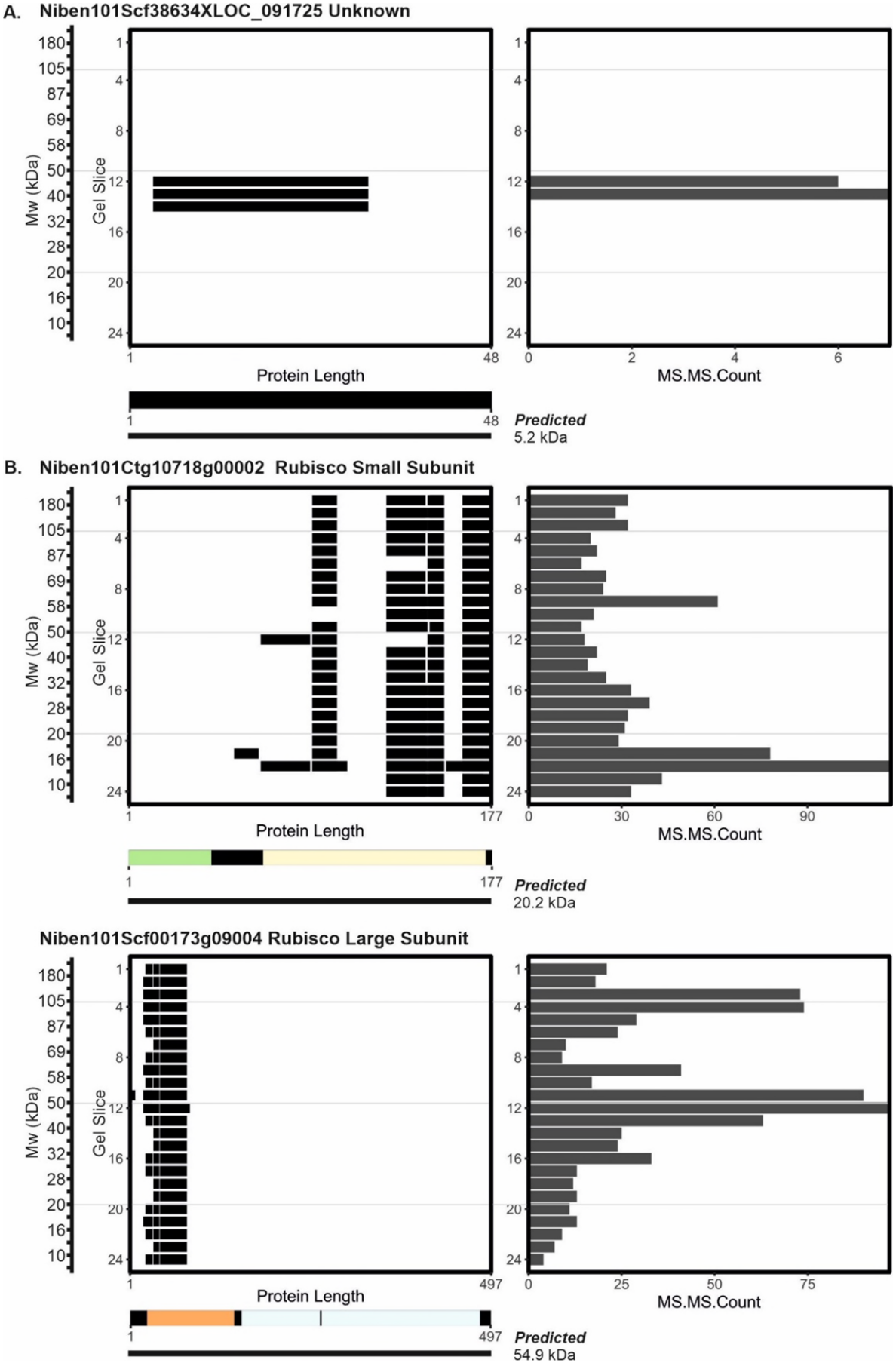
2.5 Supplementary Figures



Supplemental Figure 1. Image of PROTOMAP 4-12% gradient SDS PAGE. Gel slices were excised in 2.5 mm bands guided by gridlines in one cut and then the four biological replicate lanes separated.



Supplemental Figure 2. Protein processing is prevalent in agroinfiltrated *N. benthamiana*. Protein processing was detected by PROTOMAP or TAILS methods in the different classes of enzymes discussed in this chapter. Processed proteins were identified with neo-N termini in TAILS. PROTOMAP processed proteins were detected at a molecular weight gel slice at least two gel slices lower than the predicted molecular weight of the full length protein. Proteases and glycosidases are classified as described in the MEROPS and CAZY databases respectively.



Supplemental Figure 3. Atypical protein migration detected in PROTOMAP.

A. Peptograph of an unknown and probable misannotated protein. Left: unique peptides detected in each gel slice. Right: total MS/MS count across the four biological replicates. Predicted molecular weight from the annotated protein sequence is indicated below. **B.** Rubsico large and small subunits smear detected on peptographs. Yellow = PF00101, RuBisCO small. Green = PF12338, RbcS. Blue =PF00016, RuBisCO large. Orange =PF02788, Rubisco large.

Amino Acid	Percentage
A	6.084
R	5.575
N	4.509
D	4.746
C	2.108
E	6.049
Q	3.616
G	6.195
H	2.396
I	5.697
L	9.704
K	6.485
M	2.920
F	4.740
P	4.689
S	8.749
T	4.899
W	1.365
Y	3.000
V	6.476

Supplemental Figure 5. Amino Acid composition of the *N. benthamiana* proteome. Frequencies of amino acids in the curated FGH proteome used for this experiment were calculated in R using the PROTR package (Xiao *et al.*, 2015).

2.6 Methods

NB: This experiment was originally designed as a comparison between hypersensitive response (HR) induced and non-HR induced leaves. Coexpression of Rx and CP by agroinfiltration triggers HR in mature *N. benthamiana* leaves (Bendahmane *et al.*, 1999). However, the data analysed and discussed in this chapter includes only non-HR induced control samples, which were infiltrated with agrobacterium carrying binary vectors for either Rx or CP expression alone. Results from Rx and CP coexpression are not analysed or discussed in this chapter, but are present in the MS results files from PROTOMAP and TAILS (**AFile. 2** and **3**).

Proteomics Sample Preparation

Nicotiana benthamiana were grown under long day cycle at 16 hours of light (80 $\mu molm^{-2}s^{-1}$), 22°C and 8 hours dark at 20°C.

Agrobacterium tumefaciens C58C1 containing binary vectors with either Rx from *Solanum tuberosum* or Coat Protein from Potato Virus X (Bendahmane *et al.*, 1999) were grown overnight in Luria-Bertani medium supplemented with 100 μgml^{-1} rifampicin and kanamycin at 28°C with 220rpm overnight. Cultures were pelleted and washed with infiltration buffer (10 mM MES, 10 mM $MgCl_2$ pH 5.7). Cells were resuspended in infiltration buffer to final OD_{600} 0.5 and acetosyringone added to 100 µM. The top two fully expanded leaves of 4-5 week old plants were agroinfiltrated with either Rx or CP alone (non-HR samples). In

parallel, Rx and CP were mixed in a 1:1 ratio and co-infiltrated into leaves (HR induced samples, not discussed or analysed in this chapter). Leaves were harvested at 17 hours post infiltration (before Rx or CP protein could be detected in LC-MS/MS) and ground under liquid nitrogen. Ground frozen tissue from non-HR induced leaves agroinfiltrated with Rx or CP alone was mixed in a 1:1 ratio.

In solution digestion (ISD)

Extraction buffer (100 mM HEPES pH7.5, 5 mM DTT, MS SAFE Protease and Phosphatase Inhibitor (Sigma)) and leaf powder were mixed in a 1:1 ratio (i.e. 300 µg powder in 300 µl buffer). Extract was sonicated (Diagenode Biorupter) at full power for 10 minutes at 4°C. The samples were cleared by centrifugation at 11,000 g, 20 mins, 4°C and the retained supernatant centrifuged for a further five minutes. Protein concentration was determined using Bradford Assay. 15 µg of each sample was denatured in 8M urea. Proteins were reduced with 15 mM DTT for 30 min at RT and then alkylated with 20 mM IAA for 30 mins at 37°C in the dark. Excess IAA was quenched with final addition of DTT to final concentration of 25 mM. All solutions were prepared in 100 mM ammonium bicarbonate. Denatured protein extract was pre-digested with 500 ng Lys-C for 3 hours at 37°C, followed by digestion with 500 ng trypsin at 37°C overnight. Samples were desalted on home-made C18 stage tips as previously described (Rappsilber *et al.*, 2007). Peptides were resuspended in 0.1% (v/v) formic acid (FA) for loading onto the LC-MS/MS.

PROTOMAP

Protein Extraction and Separation: Protein was extracted as described above for PROTOMAP until protein concentration determination. Cleared lysate was mixed with lithium dodecyl sulfate loading buffer (ThermoFisher) and DTT to a final concentration of 200 mM and then boiled at 90°C for 10 min. 50 µg of sample was the loaded onto an SDS PAGE pre-cast 4-12% Bis Tris gel (ThermoFisher) and run with MOPS buffer (ThermoFisher). Each sample lane was sliced into 24 2.5 mm sections guided by a background grid (**SFig. 1**). The migration of protein ladder (Multicolor Broad Range Protein Ladder, Thermo Scientific) was used to estimate MWs of proteins in each gel slice (**SFig. 1**).

In gel-digestion: Gel slices were washed twice in HPLC grade H_2O and then 100 mM ammonium bicarbonate. Proteins were reduced by incubation with 10 mM TCEP at 62°C for 30 min and then alkylated in 55 mM iodoacetamide (IAA) for 30 min in the dark. Gel slices were then dehydrated by four washes in 50% acetonitrile (ACN) and a final wash in 100% ACN. Remaining solution was removed in a vacuum centrifuge (Eppendorf). Proteins were then digested with addition of 400 ng of trypsin in 25 mM ammonium bicarbonate and incubation overnight at 37°C. All solutions were prepared in 100 mM ammonium bicarbonate. Peptide containing solution was eluted by successive addition of 5% FA in ACN washes. Peptides were dried in a vacuum centrifuge (Eppendorf) and desalted on home-made C18 stage tips as previously described (Rappsilber *et al.*, 2007). Peptides were resuspended in 0.1% (v/v) FA for loading onto the LC-MS/MS.

PROTOMAP and ISD data acquisition

Samples were analysed with an Orbitrap Elite (Thermo) coupled to an Easy-nLC 1000 liquid chromatography (LC) system (Thermo) (Michalski *et al.*, 2012). The LC was run in a single column mode with an analytical column of a fused silica capillary (75 µm × 22 cm) with an integrated PicoFrit emitter (New Objective) packed in-house with Reprosil-Pur 120 C18-AQ 1.9 µm resin (Dr. Maisch). The analytical column was encased by a column oven (PRSO-V1; Sonation) at 45ºC and attached to a nanospray flex ion source (Thermo). All solvents were of UPLC grade (Sigma-Aldrich). Peptides were directly loaded onto the analytical column with a maximum flow rate that would not exceed the set pressure limit of 980 bar (usually around 0.6 – 1.0 µL/min). Peptides from **PROTOMAP** were separated on the running a 70 min gradient of solvent A and solvent B (start with 7% B; gradient 7% to 35% B for 60 min; gradient 35% to 80% B for 5 min and 80% B for 5 min) at a flow rate of 300 nl/min. Peptides from **ISD** were separated by using a 140 min gradient of solvent A and solvent B (start with 7% B; gradient 7% to 35% B for 120 min; gradient 35% to 80% B for 10 min and 80% B for 10 min) at a flow rate of 300 nl/min.

Xcalibur software (version 2.2 SP1.48) was used for data acquisition with the mass spectrometer in positive ion mode. Precursor ion scanning was performed in the Orbitrap analyser (FTMS; Fourier Transform Mass Spectrometry) in the scan range of m/z 300-1500 (PROTOMAP) or 1800 (ISD) and at a resolution of 60000 with the internal lock mass option turned on (lock mass was 445.120025 m/z, polysiloxane) (Olsen *et al.*, 2005). Product ion spectra were recorded in a data dependent fashion in the ion trap (ITMS) in a variable scan range and at a rapid scan rate (Wideband activation was turned on). The ionization potential (spray voltage) was set to 1.8 kV. Peptides were analysed using a repeating cycle

consisting of a full precursor ion scan (3.0×10^6 ions or 50 ms) followed by 12 (for PROTOMAP) or 15 (ISD) product ion scans (1.0×10^4 ions or 50 ms) where peptides are isolated based on their intensity in the full survey scan (threshold of 500 counts) for tandem mass spectrum (MS2) generation that permits peptide sequencing and identification. Collision induced dissociation (CID) energy was set to 35% for the generation of MS2 spectra. During MS2 data acquisition dynamic ion exclusion was set to 120 seconds with a maximum list of excluded ions consisting of 500 members and a repeat count of one. Ion injection time prediction, preview mode for the FTMS, monoisotopic precursor selection and charge state screening were enabled. Only charge states higher than 1 were considered for fragmentation.

Raw spectra were analysed with Andromeda search engine (Cox *et al.*, 2011) in Max Quant v 1.5.5.30 (Cox and Mann, 2008). Primarily default parameters were selected including LFQ and match between runs. Spectra were searched against the in-house Grosse-Holz *N. benthamiana* proteome database with appended common contaminants, Agrobacterium C58 strain (UP000000813) and Rx and CP sequences (Grosse-Holz *et al.*, 2018). Permitted modifications included static carbamidomethylation and cysteines, variable N-terminal acetylation and methionine oxidation and trypsin specific digestion. The instrument type in Andromeda searches was set to Orbitrap and the precursor mass tolerance was set to ±20 ppm (first search) and ±4.5 ppm (main search). The MS/MS match tolerance was set to ±0.5 Da. The peptide spectrum match FDR and the protein FDR were set to 0.01 (based on target-decoy approach). Minimum peptide length was 7 amino acids. For protein quantification unique and razor peptides were allowed. Modified peptides were allowed for

quantification. The minimum score for modified peptides was 40. Label-free protein quantification was switched on, and unique and razor peptides were considered for quantification with a minimum ratio count of 2. Retention times were recalibrated based on the built-in nonlinear time-rescaling algorithm. MS/MS identifications were transferred between LC-MS/MS runs with the "match between runs" option in which the maximal match time window was set to 0.7 min and the alignment time window set to 20 min. The quantification is based on the "value at maximum" of the extracted ion current. At least two quantitation events were required for a quantifiable protein.

PROTOMAP Data Processing: PROTOMAP dataset was analysed in R (**Additional File 1**). Briefly, contaminants and proteins identified only by site were removed and the remainder filtered to leave protein groups that were robustly identified in at least 3 of 4 biological replicates, independent of which gel slice each was detected in. A peptograph of unique peptides and MS/MS counts (spectral counts) detected across each gel slice were plotted for each protein group.

Terminal Amine Isotopic Labelling of Substrates (TAILS)

Protein Extraction: Protein was extracted from the same batches of ground leaf tissue as for PROTOMAP. Samples were processed at described in (Demir *et al.*, 2017), with minor modifications. Extraction buffer (100 mM HEPES pH7.5, 0.5 mM EDTA, 6 M guanidine hydrochloride, 1X HALT protease inhibitor cocktail (ThermoFisher)) was added to frozen tissue, homogenised, and filtered through miracloth. Chloroform methanol precipitation was performed and the protein pellet resuspended in extraction buffer. Protein concentration was determined by BCA assay and 1 mg of protein at 1 $mgml^{-1}$for each sample was taken

forward. Proteins were reduced in 30 mM DTT at 37°C for 30 mins and then alkylated with 15 mM IAA at room temperature for 30 mins in the dark. Excess reagent was quenched with the further addition of DTT to a final 15 mM concentration. Primary amines in the samples were dimethyllated with addition of heavy ($^{13}CD_2O$) or light ($^{12}CH_2O$) formaldehyde and sodium cyanoborohydride to 20 mM each. Samples were incubated overnight at 37°C and labelling refreshed for a further 2 hrs with repeated reagent addition to a final 40 mM. Labelling reaction was then quenched by adding Tris HCl pH6.8 to 100 mM. Proteins were isolated with chloroform methanol precipitation. Pelleted proteins were resuspended in 100 mM HEPES pH 7.5 to ~1 $mgml^{-1}$. Trypsin (Pierce, Sigma) was added in a 1:100 ratio to protein along with $CaCl_2$ to 5 mM and acetonitrile to 5%. Samples were digested overnight at 37°C. Negative selection of blocked N-termini was performed by hydrophobic tagging with addition of a pre-dialysed HPG-ALDII polymer (http://www.flintbox.com/public/project/1948/) in a 5:1 ratio to protein. Sodium cyanoborohydride was added to 20 mM and incubated overnight at 37°C. Free peptides were separated from the polymer using a 10 kDa cut off spin column (Machery Nagel). The peptide solution was fractionated at high pH over an ACN gradient through a C18 HR-X cartridge (Machery Nagel). Desalted peptides were dried in a speed vac and resuspended in 0.1% trifluoroacetic acid, 2% ACN.

Mass spectrometry data acquisition: Samples were analysed on a two-column nano-HPLC setup (Ultimate 3000 nano-RSLC system with Acclaim PepMap 100 C18, ID 75 µm, particle size 3 µm columns: a trap column of 2 cm length, ThermoFisher) with different analytical columns (Acclaim PepMap 100 C18, ID 75 µm, particle size 3 µm, 25 / 50cm length column,

coupled to a high resolution Q-TOF mass spectrometer (Impact II, Bruker) as described (Rinschen *et al.*, 2017). A binary gradient flow of two solultions (A: 0.1% FA in H_2O, B: 0.1% FA in ACN) was used for separation at 60 °C: 4-32% B for 70 minutes at 275 nL/min flow and a total runtime of 100 minutes.

Bruker HyStar Software (v3.2, Bruker Daltonics) was used for data acquisition in line-mode in a mass range from 200-1500 m/z at an acquisition rate of 4 Hz. The top 14 most intense ions were selected for fragmentation with dynamic exclusion of previously selected precursors for the next 30 sec unless intensity increased three-fold compared to the previous precursor spectrum. Intensity-dependent fragmentation spectra were acquired between 5 Hz for low intensity precursor ions (> 500 cts) and 20 Hz for high intensity (> 25k cts) spectra. Fragment spectra were averaged from t stepped parameters, with 50% of the acquisition time manner with split parameters: 61 µs transfer time, 7 eV collision energy and a collision RF of 1500 Vpp followed by 100 µs transfer time, 9 eV collision energy and a collision RF of 1800 Vpp.

Data analysis: Acquired mass spectra were matched to peptide sequences at a FDR of 0.01 using MaxQuant (Cox and Mann, 2008) v1.6.3.3 using standard Bruker QToF instrument settings. The in-house Grosse-Holz *N. benthamiana* proteome database with appended common contaminants was used (Grosse-Holz *et al.*, 2018). Searches used C-terminal cleavage by ArgC as digestion enzyme. Further search parameters included isotope labelling by light (+28.031300) or heavy (+36.075670) dimethylation of peptide N-termini or Lys residues, Cys carbamido-methylation as fixed and Met oxidation, N-terminal acetylation (+42.010565) or N-terminal pyroGlu formation from Glu (-18.010565) or Gln (-17.026549)

as variable modifications. Further statistical data analysis, filtering and annotation of the N-termini datasets was performed with the Perl script MANTI.pl 3.9.9 (https://manti.sourceforge.io). Further dataset filtering and analysis was performed using a new R script generated for this dataset. Peptides were filtered for those detected in at least two of four biological replicates and N-terminal maps plotted against protein domains annotated by Interproscan using R (**Additional File 2**).

Cloning

Open reading frames were amplified from cDNA generated from agroinfiltrated tissue using the following primers: GDSL (Niben101Scf01670g02001/NbD012971.1) with 5'-TTGGTCTCAAGGGGAAGAAGTATTAATGCTTC-3' and 5'-TTGGTCTCACACCAACTTGGCAGTTC-3':, SCPL51 (Niben101Scf08963g02006/NbD026891.1) with 5'-TTGGTCTCAAGGGGCTAGAAATCAAGATG-3' and 5'-TTGGTCTCACACCTTTCGTGATGCTG-3'. Golden gate modules used in cloning for agroinfiltration include: pJK268c (pL1V-P19-F2), pJK002 (pL0M-S-NtPR1a_PIV2), EC41414 (pL0M-T-35S-1-41414), EC51288 (pL0M-PU-35S-TMV-3-51288), EC15212 (pL0M-C2-3xMyc-15212). Expression constructs were assembled with digestion by either BsaI or BpiI and ligation with T4 DNA ligase (New England BioLabs, USA). pET194 (pL2M-p19-35S-NtPR1-His-GDSL-3xMyc) and pET196 (pL2M-p19-35S-NtPR1-SCPL51-3xMyc) constructs were validated by sequencing (Eurofins Genomics) and transformed into *Agrobacterium tumefaciens* GV3101 (a C58 based strain) by heat shock. BGAL1

(Niben101Scf04549g10022/NbD029635.1) expression construct pBK26 was used (Chandrasekar, 2018).

Agroinfiltration was performed as described for Sample Preparation above. At 4 dpi, leaf disks and apoplastic fluids were isolated. Leaf disk protein was extracted directly in SDS-gel loading buffer. Apoplastic fluid was isolated by infiltration of leaves with cold milliQ and subsequent centrifugation at 2,000 g for 15 min in the cold. Tagged proteins were detected by Western blotting with Ab1326 (Abcam) or StrepMAB-Classic (IBA Life Sciences).

Datasets

PROTOMAP MaxQuant peptides and proteingroups files are attached in **Additional File 3** and **4 respectively.** TAILS dataset is available in **Additional File 5**. ISD MaxQuant peptides and proteingroups files are attached in **Additional File 6** and **7** respectively**.**

Chapter 3: Investigation of the biological activity and roles of NbVPE1a and NbVPE1b

Emma Thomas[1], Daniel Krahn[2], Friederike Grosse-Holz[1], Chris Schofield[2] and Renier A.L. van der Hoorn[1]

1. Department of Plant Sciences, University of Oxford
2. Department of Chemistry, University of Oxford

3.1 Abstract

VPEs in plants have been characterised as positive regulators of programmed cell death in a variety of biotic and abiotic stresses. Silencing of *NbVPE1a* and *NbVPE1b* in *N. benthamiana* blocked vacuolar rupture and cell death upon biotic stress. To further understand the roles of NbVPEs, we have generated the first stable *vpe* mutant lines in *N. benthamiana* lacking the activity of the two major vegetative VPEs, NbVPE1a and NbVPE1b. NbVPE1a and NbVPE1b produced in *E. coli* could cleave the classical VPE substrates YVAD and ESEN. The *vpe1a/b* double mutant showed reduced levels of ESENase activity whereas YVADase activity was unaltered. Despite the clear biochemical phenotype, the induction of HR was not delayed in the *vpe1a/b* mutant and responses to ER, salt, heat and nitrogen stresses in the *vpe1a/b* double mutant did not differ significantly from WT. Similarly, the growth of *Pseudomonas syringae* or *Phytophthora infestans* was

not compromised in the *vpe1a/b double* mutant. NbVPE1a and NbVPE1b are therefore not essential for responses to the biotic and abiotic stresses applied in this study.

3.2 Introduction

Vacuolar processing enzymes (VPEs) are widespread cysteine proteases present exclusively in eukaryotes, spanning from mammals to plants (Rawlings *et al.*, 2016). VPEs are ubiquitous in plant genomes, and in fact most plant species contain an expanded repertoire of VPEs (Yamada *et al.*, 2019), in contrast to animals that typically contain a single VPE (legumain). Activity of VPEs is highly conserved between members. All tested to date catalyse the processing of proteins following either an asparagine or aspartic acid residue (Bernath-Levin *et al.*, 2015; Hatsugai *et al.*, 2015; Nguyen *et al.*, 2015; Dall and Brandstetter, 2016). The protease activity of VPEs is only detected at acidic pH and accordingly VPEs accumulate in vacuoles or equivalently acidic compartments (Hara-Nishimura *et al.*, 1991; Hiraiwa *et al.*, 1999; Hatsugai *et al.*, 2015; Dall and Brandstetter, 2016; Teper-Bamnolker *et al.*, 2019). As is typical of many proteases, VPEs are synthesised with inhibitory pro-peptides. Maturation of the enzyme is autocatalytic *in vitro* with removal of both N- and C-terminal pro-peptides (Hiraiwa *et al.*, 1999; Kuroyanagi *et al.*, 2002; Nguyen *et al.*, 2014; Bernath-Levin *et al.*, 2015; Haywood *et al.*, 2018). However, VPEs are distinctive in several of their properties. They are the only proteases comprising the C13 family in the MEROPS classification (Rawlings *et al.*, 2016). No other proteases are known to cleave following asparagine residues in humans or plants. Finally, some VPEs can also catalyse peptide ligation reactions following protease cleavage to form cyclic peptides (Saska *et al.*, 2007). In plants, the investigation of VPEs has primarily been split between study of the ligation capacity of select VPEs and the various biological roles of these proteases.

VPEs were initially discovered in castor bean seeds, and their identification in seeds of other species swiftly followed (Hara-Nishimura *et al.*, 1991, 1995; Becker *et al.*, 1995; Kinoshita *et al.*, 1995; Linnestad *et al.*, 1998; Okamoto *et al.*, 1999). A role of VPEs in seed storage protein processing was confirmed with the development of Arabidopsis mutants lacking either individual VPEs or all four VPEs. This demonstrated a predominant role of AtVPEβ in the seeds and a strong disruption of storage protein profile in the null mutant (Gruis *et al.*, 2002, 2004; Shimada *et al.*, 2003).

In following years, the roles of VPEs in diverse tissues has been explored. A role in pathogen defence was first revealed in the Arabidopsis *vpeγ* mutant lacking the dominant vegetative tissue VPE. The v*peγ* plants were more susceptible to *Pseudomonas syringae*, *Botrytis cinerea* and Turnip Mosaic Virus (Rojo *et al.*, 2004). Several studies have contributed more circumstantial evidence to support a role of VPEs in immunity. Upregulation of either VPE transcripts or activity was detected upon infection with *Hyaloperonospora arabidopsidis*, *Erwinia amylovora, Heterodera filipjevi* (cyst nematode) and upon *Agrobacteria tumefaciens* infiltration in various species (Iakimova *et al.*, 2013; Misas-Villamil *et al.*, 2013; Pillay *et al.*, 2016; Labudda *et al.*, 2020). Despite this, VPEs are not universally involved in pathogen responses as colonisation of *P. infestans* was not perturbed upon *NbVPE1a and NbVPE1b* silencing in *N. benthamiana* (Shibata *et al.*, 2010).

A key part of the plant immune system is induction of a programmed cell death (PCD) termed the hypersensitive response (HR). Processing after D residues is vital for PCD in animals and similar activity has been shown in plant PCD (del Pozo and Lam, 1998; Bonneau *et al.*, 2008). This led to investigation of VPEs as key regulators of plant PCD. In

N. benthamiana silencing of two VPEs, *NbVPE1a* and *NbVPE1b*, was found to abolish vacuolar rupture and cell death triggered upon recognition of Tobacco mosaic virus (TMV) by the *N* gene (Hatsugai *et al.*, 2004). Furthermore, HR induced by the elicitor harpin of *Xanthomonas oryzae* was similarly blocked by *NbVPE1a/b* silencing (Zhang *et al.*, 2010). NbVPE1a and NbVPE1b are not universally required for death however, as induction of HR by Nep1 or boehmerin elicitors was not perturbed (Zhang *et al.*, 2010).

VPEs are also required for PCD induced by the fungal toxin FB1 in Arabidopsis, and the related toxin AAL was shown to require VPEs in *NuVPE* silenced *Nicotiana umbratica* (Kuroyanagi *et al.*, 2005; Mase *et al.*, 2012). In addition, transcriptional upregulation of VPEs was correlated with HR induction in *Phytophthora sojae* infected *N. benthamiana* and cryptogein induced PCD in *Nicotiana tabacum* var. Xanthi cells (Gauthier *et al.*, 2007; Jing *et al.*, 2016). Arabidopsis VPEs were similarly required for vacuolar rupture and cell death induced by the obligate biotroph *Piriformospora indica*, which requires root cell death for successful colonisation (Qiang *et al.*, 2012). In this manner, VPE activity regulating cell death is beneficial to the pathogen reflecting the different strategies of pathogen types. Indeed, upregulation of VPE activity in *Hyaloperonospora arabidopsidis* infected Arabidopsis correlates with increased colonisation and sporulation (Misas-Villamil *et al.*, 2013). However, the HR in Arabidopsis induced by AvrRpm1 of *P. syringae* is VPE-independent (Hatsugai *et al.*, 2009). VPEs therefore do play a role in HR but are not universal regulators of cell death.

The requirement for VPEs upstream of vacuolar rupture in PCD is now recognised in a variety of abiotic, biotic and developmental PCDs. For example, in rice, salt stress induced

vacuolar rupture was promoted by OsVPE3 overexpression and reduced upon *OsVPE3* silencing (Lu *et al.*, 2016). Arabidopsis *vpe* null mutants subject to heat stress also demonstrated a delay in vacuolar rupture compared to WT plants (Li *et al.*, 2012). Furthermore, PCD induced by ER stress was restricted in the *vpe* null mutants associated with reduced vacuolar rupture (Cai *et al.*, 2018).

Another potential biological function of VPEs is in senescence. VPE is upregulated in senescent nitrogen fixing nodules from legume-rhizobia symbioses and darkened leaves (Oh *et al.*, 2004; Pierre *et al.*, 2014; van Wyk *et al.*, 2014; Pružinská *et al.*, 2017). Indeed, overexpression of a VPE in *Medicago truncatula* expedited senescence onset whilst silencing correspondingly delayed senescence (Pierre *et al.*, 2014). In summary, VPEs have been implicated in seed storage processing, pathogen defences and tolerance to various abiotic stresses.

Investigation of VPEs to date has relied on inhibition, silencing or overexpression in most species. However, the most compelling evidence has emerged from Arabidopsis, as the only species with established stable *vpe* mutant lines (Gruis *et al.*, 2004; Kuroyanagi *et al.*, 2005; Yamada *et al.*, 2019). Second to this, silencing in *N. benthamiana* has provided strong evidence regarding a role for VPEs in pathogen responses (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010). *N. benthamiana* has also proved a useful tool for the investigation of VPEs from diverse species due to the ability to transiently overexpress proteins by agroinfiltration (Conlan *et al.*, 2012; Poon *et al.*, 2018). Therefore, in this chapter we aimed to create a comparable mutant resource in *N. benthamiana* to investigate the fundamental biology of VPEs in another plant species.

3.3 Results

3.3.1 NbVPE1a and NbVPE1b are major vegetative VPEs

VPEs have traditionally been divided into three main types: β, δ and α/γ, based upon the four Arabidopsis VPEs. However this classification has been challenged as more genomes have become available (Yamada *et al.*, 2019). We generated a phylogenetic tree containing VPEs of *N. benthamiana* and other well curated species (**Fig. 1**). VPEs clustered into groups of δ VPEs, α/γ and β VPEs and an additional novel Solanum-specific expansion. Importantly most species do not contain representatives of δ VPEs, indicating these types are not essential for conserved parts of the plant life cycle. *N. benthamiana* has 11 full length VPEs encoded in the genome and a further two partial proteases predicted to be non-functional (**Fig. 1**) (Kourelis *et al.*, 2019). No δ type VPE was detected in the *N. benthamiana* genome. Four NbVPEs were positioned in the α/γ clade, including NbVPE1a and NbVPE1b which have previously been investigated for a role in leaf PCD (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010).

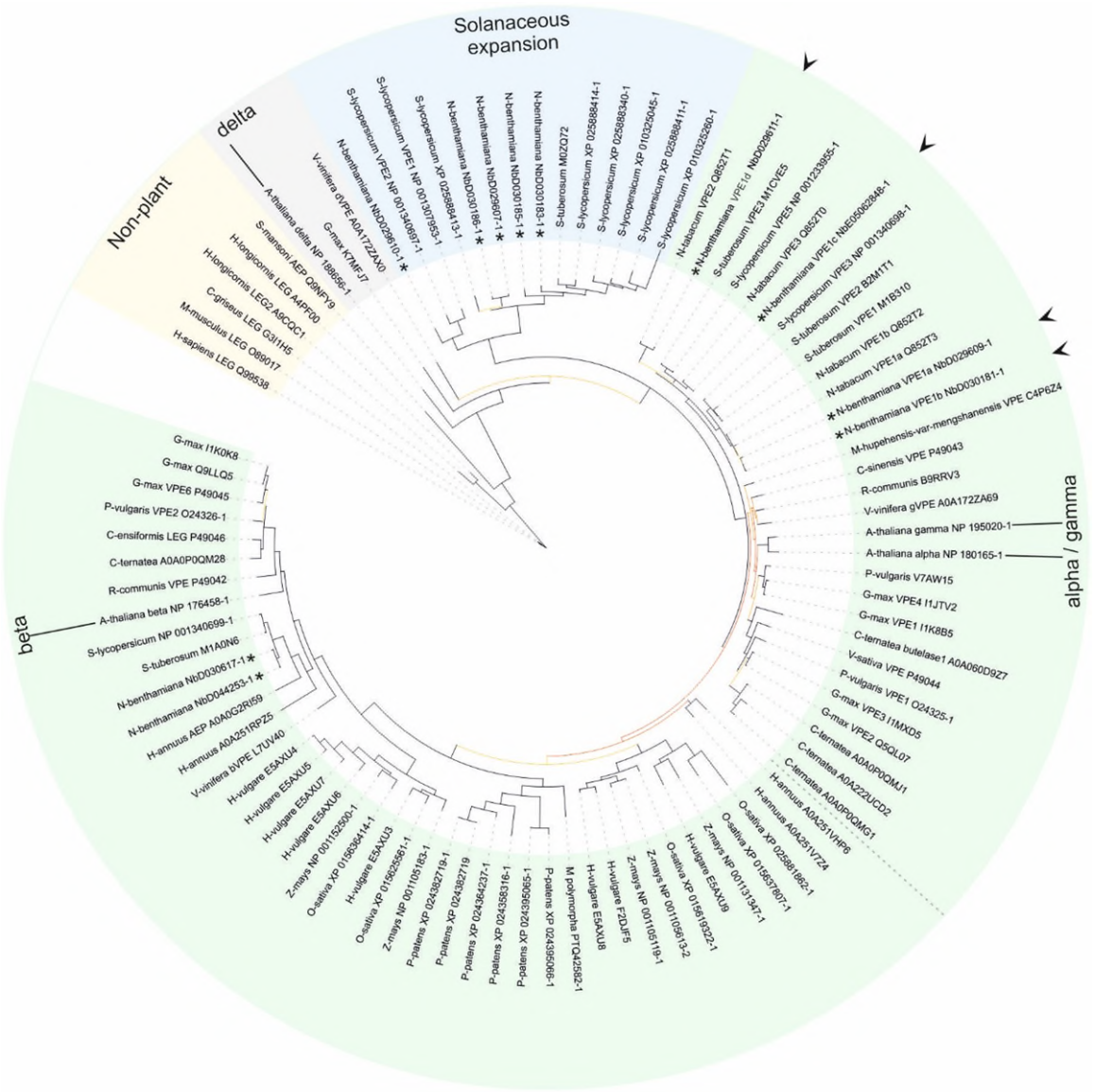


Figure 1. Unrooted tree of select plant and non-plant VPEs. A maximum likelihood tree was constructed using PhyML based on a ClustalW alignment of amino acid sequences. The Bootstrap consensus tree is shown (500 bootstrap replicates). Branches with low Bootstrap support are coloured (red < 25%, orange 25 – 40%, yellow 40-50%). Tree was rendered using iTol (Letunic and Bork, 2019). Arrows mark *NbVPE1a, NbVPE1b, NbVPE1c and NbVPE1d*. Protein IDs correspond to either the UniProt database or gene identifier. *N. benthamiana* VPEs are marked with *.

Native VPE activity in *N. benthamiana* leaves was investigated using the VPE specific activity-based probe JOPD1 (**Fig. 2A**) (Lu *et al.*, 2015). JOPD1 has previously labelled VPEs in Arabidopsis, *Solanum lycopersicum* (tomato), *Hordeum vulgare* (barley), *Brassica*

napus (rapeseed), *Aponogeton madagascariensis* (lace plant), *Helianthus annuus* (sunflower), *Clitoria ternatea* and *Crocus sativa* (saffron) (Sueldo *et al.*, 2014; Bernath-Levin *et al.*, 2015; Lu *et al.*, 2015; Husaini *et al.*, 2018; Rantong and Gunawardena, 2018; James *et al.*, 2019; Poret *et al.*, 2019; Labudda *et al.*, 2020). Seven robust fluorescent signals are detected in WT *N. benthamiana* leaf extracts (**Fig 2B**). Labelling can be prevented by preincubation with the caspase-1 inhibitor Ac-YVAD-cho, indicative of the previously characterised activity of VPEs (Hatsugai *et al.*, 2015). These signals may represent multiple processing states of individual VPE gene products or result from different gene products. Interestingly, additional signals were observed from labelled leaf extracts compared to *in vivo* labelling when the probe was infiltrated into the leaf (**Fig. 2B**). The extra *in vitro* signals could be attributed to two scenarios. Either the probe is unable to penetrate certain compartments containing VPEs, or additional processing of VPEs occurs during the *in vitro* incubation step.

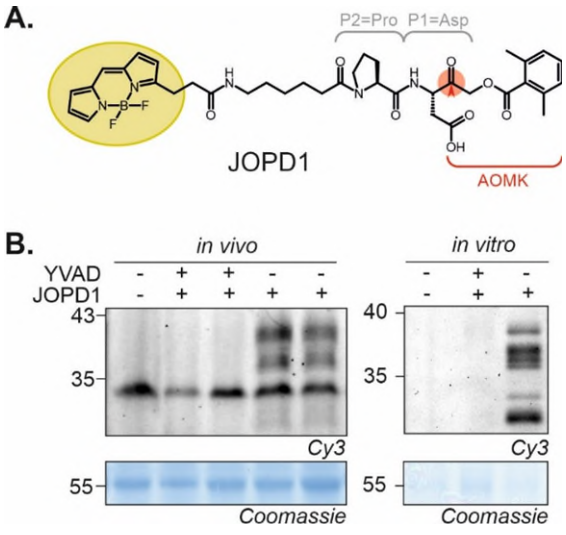


Figure 2. JOPD1 labels VPEs in *N. benthamiana in vivo* and *in vitro*. Vegetative VPEs of *N. benthamiana* were detected by activity-based protein profiling with the probe JOPD1. **A.** Structure of JOPD1 probe, reproduced from (Lu *et al.*, 2015). **B.** For *in vivo* labelling, leaves were infiltrated with either DMSO or 50 μM Ac-YVAD-cho. After 30 minutes, the

same tissue was re-infiltrated with 2 µM JOPD1 or DMSO (no probe control) and incubated for 3 hours. Tissue was then harvested and protein extracted directly in SDS gel loading buffer. For *in vitro* labelling, protein was extracted from mature leaves and preincubated with either DMSO or 50 µM Ac-YVAD-cho, followed by a 3 hours incubation with 2 µM JOPD1 or DMSO. Labelled proteins were separated by SDS PAGE and detected by fluorescence scanning.

To assess the role of VPEs in *N. benthamiana* leaves, we generated vegetative *vpe* mutants using CRISPR-Cas9. *NbVPE1a*, *NbVPE1b* and *NbVPE1c* were selected for mutagenesis as the most abundant VPEs in leaves measured in in-house transcriptomic and proteomic datasets (**Fig. 3A**) (Grosse-Holz *et al.*, 2018). Among the transformants generated, two stable lines were identified (**Fig. 3B**). Line *vpe1b* contains the *vpe1b-1* mutant allele with a single nucleotide insertion introducing a frame shift resulting in five alternative amino acids and a premature stop codon, before the catalytic cysteine (**Fig. 3C**). The second line, *vpe1a/b,* contains the mutant alleles *vpe1a-1* and *vpe1b-1* (**Fig. 3B**). Allele *vpe1a-1* carries a single nucleotide insertion that introduces a frame shift causing a subsequent variant five amino acids and a premature stop codon within the catalytic domain (**Fig. 3C**). Allele *vpe1b-2* carries a single nucleotide deletion resulting in a frame shift, translating into three alternative amino acids and a premature stop codon before the catalytic cysteine (**Fig. 3C**). The *NbVPE1c* sequence could not be amplified from gDNA or cDNA despite repeated attempts. The absence of *NbVPE1c* amplification may be a consequence of inaccurate gene prediction and annotation, as a more recent gene model predicts a different sequence for this gene (Kourelis *et al.*, 2019). Plants of both *vpe1b* or *vpe1a/b* lines were visually indistinguishable from WT in their growth from seedling to mature flowering plants (**Fig. 3B**).

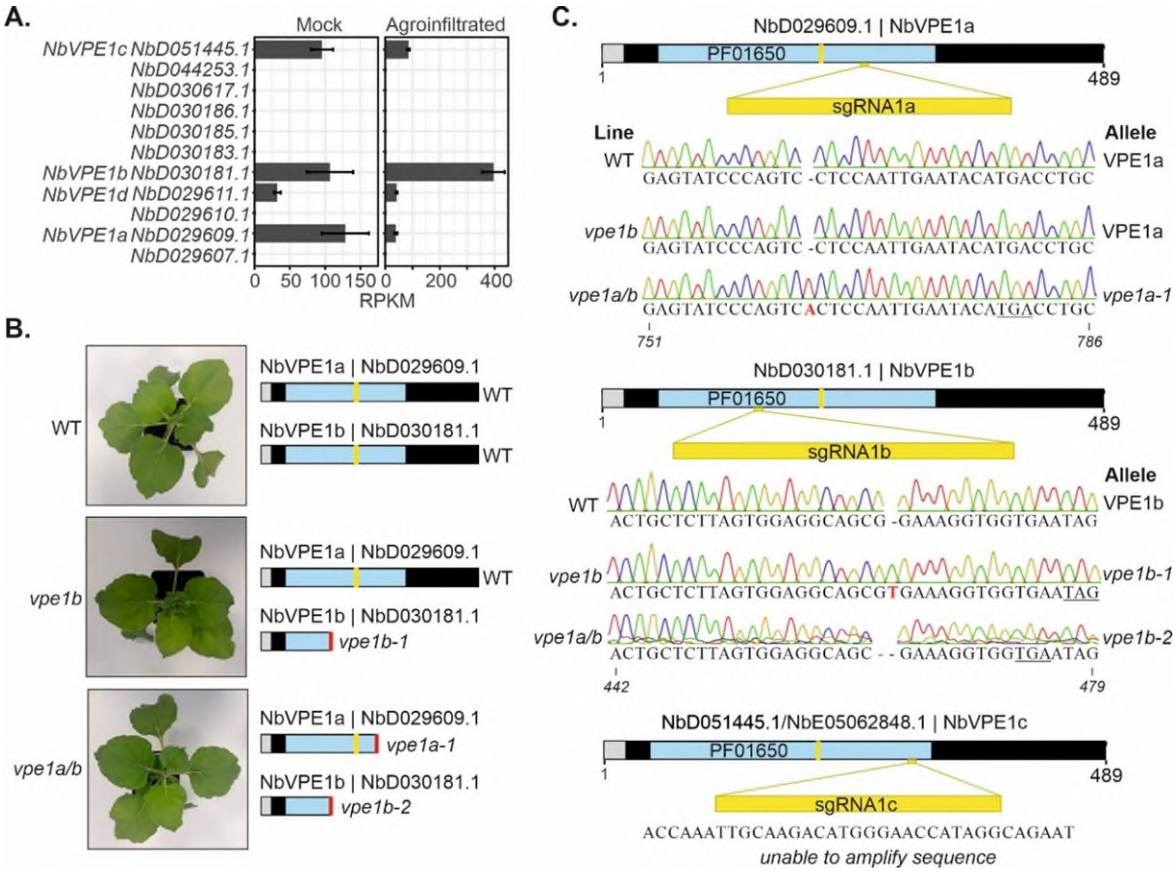


Figure 3. *N. benthamiana* mutants of γ-type VPEs were generated by CRISPR-Cas9. **A.** NbVPE1a, NbVPE1b and NbVPE1c are the most abundant transcripts in leaves. RPKM values for all full length NbVPEs from a transcriptomic dataset of mock (water) or agroinfiltrated *N. benthamiana* leaves (Grosse-Holz *et al.*, 2018)**.** Error bars represent standard deviation of n=3. **B.** *N. benthamiana* mutants were generated using CRISPR-Cas9 with three sgRNAs targeting *NbVPE1a*, *NbVPE1b* and *NbVPE1c*. Mutations confirmed by sequencing are shown. Grey = signal peptide, blue = catalytic domain, yellow = catalytic cysteine, red = premature stop codon introduced by frame shift. Images taken of 4 week old plants. **C.** Sequencing of gRNA target sites in the stable mutants *vpe1b* and *vpe1a/b*. Numbering is based on coding sequence only. Bases differing from wildtype are in red. Premature stop codons introduced by frameshifts are underlined.

Activity-based labelling of leaf extracts confirmed NbVPE activity had been reduced in the *vpe1b* and *vpe1a/b* mutants. Due to limited stock of the JOPD1 probe, we proceeded with two new synthesised probes: DK004 and DK005 (**Fig. 4A**) (unpublished, Daniel Krahn). DK004 is similar to JOPD1 with a D at P1, but contains an extended linker region. DK005 has the same structure to DK004 but contains an N at P1. We postulated DK005 would

label VPEs with higher specificity than DK004. Surprisingly, no labelling was detected in *N. benthamiana* leaf extracts with DK005 incubation, whereas DK004 generated a labelling profile very similar to that of JOPD1 (**Fig. 4B, Fig. 2B**). Importantly, *vpe1b* mutant lacked four of the seven signals observed in WT (**Fig. 4B, arrows 1,2,5,7**) and *vpe1a/b* lacked a further three signals (**Fig. 4B, arrows 3,4,6**). In labelled *vpe1b* extracts two of the remaining signals were stronger in comparison to WT (**Fig. 4B, arrows 3,4**). This increase was also observed in labelling with excess DK004 at 5 µM, suggesting that this is not an effect of reduced competition between VPEs for the probe in the mutant but an increase of an active VPE abundance. Considering signals 3, 4 and 7 are absent in the *vpe1a/b* mutant, it is likely that these arise from NbVPE1a. Importantly, labelling with excess probe reveals additional weak signals present in the *vpe1b* and *vpe1a/b* mutant, that may originate from NbVPE1c or NbVPE1d (**Fig. 4B, arrows 8,9,10**). These labelling experiments indicate that the vast majority of active VPEs are absent in the *vpe1a/b* double mutant.

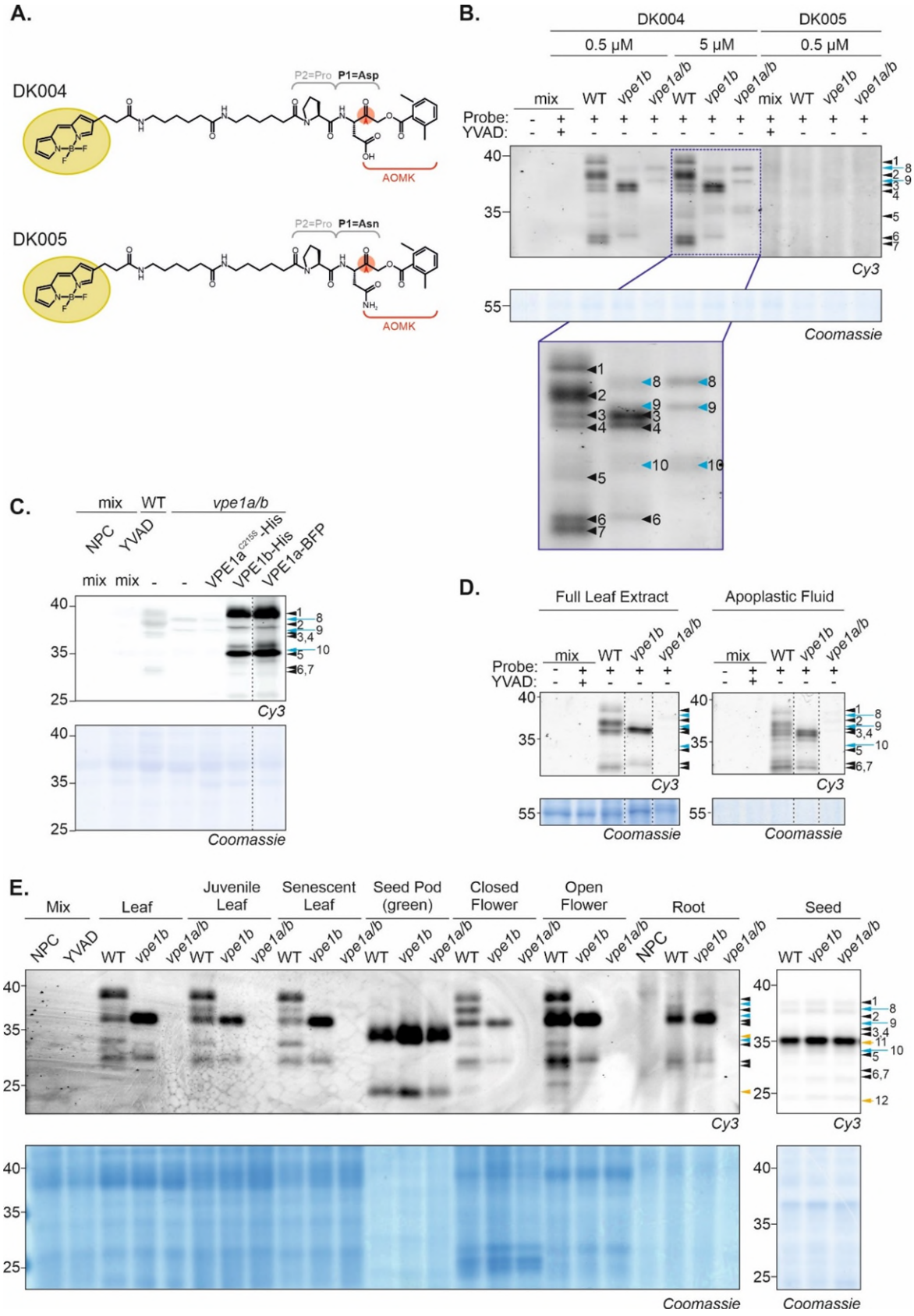


Figure 4. Mutants generated by CRISPR-Cas9 have reduced vegetative VPE activity. **A.** Probes designed to label VPEs. DK004 based on JOPD1 contains a P1 aspartic acid residue and DK005 contains an asparagine residue. BODIPY fluorophore is highlighted in

yellow. **B.** *vpe* mutants lack active VPE signals in leaves. Leaf protein extracts were preincubated with DMSO or 50 μM Ac-YVAD-cho for 30 minutes and subsequently labelled with an excess of DK004 or DK005 probes for three hours. Labelled proteins were by in-gel fluorescence scanning. Black arrows mark DK004 signals present in WT and blue arrows mark additional signals detected in DK004 profile of *vpe1a/b* mutant. **C.** Complementation of *vpe1a/b* mutant with transient expression of NbVPE1a, NbVPE1b and C215S catalytic inactive mutant by agroinfiltration. - = p19 control infiltration. Mature leaves were infiltrated with OD_{600} = 0.5 of agrobacteria and proteins extracted at 4 dpi. **D.** NbVPEs are active in apoplastic fluid and total leaf extracts. **C.** , **D.** and **E.** labelling was performed as for **B.**, but with 0.25 μM DK004. **E.** Protein was extracted from various tissues of WT and *vpe* mutant plants. Orange arrows mark the major additional DK004 signals present in developing seed pods and/or mature seeds. NPC = no probe control.

To confirm that NbVPE1a and NbVPE1b are generating these signals, we transiently overexpressed C-terminally tagged proteases in the *vpe1a/b* background and labelled leaf extracts with DK004. Surprisingly, both NbVPE1a-BFP and NbVPE1b-His + BFP expressing leaf extracts generated identical DK004 labelling profiles. *vpe1a/b* extracts overexpressing a construct with the catalytic cysteine residue substituted for serine, NbVPE1a^{C215S}-His, resulted in no additional signals in the DK004 profile compared to *vpe1a/b* background. Pre-incubation with Ac-YVAD-cho from 30 minutes prior to labelling resulted in no detectable DK004 signals, indicating that NbVPE1a and NbVPE1b can be inhibited by Ac-YVAD-cho, as expected for active VPEs (**Fig. 4C**). The activity profiles correlated with the signals lost in the *vpe1b* mutant (**Fig. 4B,C, arrows 1,2,5**), but unexpectedly not the additional signals lost in the *vpe1a/b* mutant. The artificial level of VPE overexpression may lead to different NbVPE1a processing. Previously transient expression of AtVPEs in *N. benthamiana* also did not directly reproduce labelling profiles of Arabidopsis tissue extracts (Lu *et al.*, 2015). Alternatively, these signals could be attributed to NbVPE1c/d, for which mutations could not be assessed.

The spatial distribution of NbVPE activities across compartments and tissues has not previously been studied. Reports in solanaceous species have detected active VPEs in both extracellular and intracellular locations (Goulet *et al.*, 2012; Sueldo *et al.*, 2014). The DK004 labelling profile of apoplastic fluid (AF) and total extracts isolated from the same leaves were highly similar within each genotype, (**Fig. 4D**) demonstrating that NbVPEs, including NbVPE1a and NbVPE1b, are present and active in both compartments.

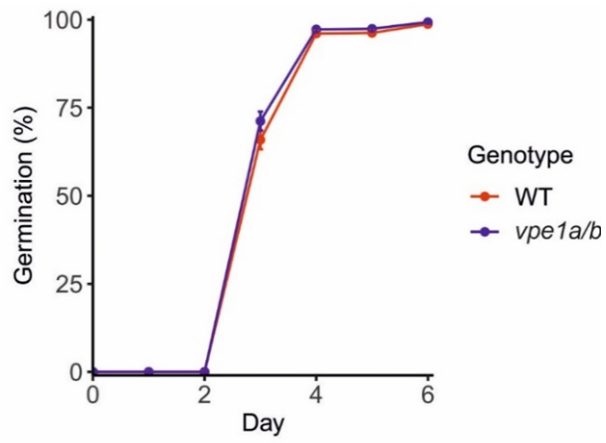


Figure 5. Germination is not compromised in the *vpe1a/b* mutant. Seeds were plated onto ½ MS 0.8% agar and germination (visible radical emergence) monitored over time. n = 4 biological replicates comprised of 6 plates of 19-20 individuals, error bars represent standard deviation.

To characterise which tissues are compromised in NbVPE activity in *vpe1b* and *vpe1a/b* mutants, extracts were made from various plant organs and DK004 labelling performed. The labelling profile was consistent across different vegetative tissues including roots, leaves and flowers for each individual genotype (**Fig. 4E**). Labelling of mature and developing seed extracts collected from dried and open, or developing seed pods respectively revealed strong labelling signals at distinct molecular weights that do not overlap with vegetative signals (**Fig. 4E, arrows 11,12**). Labelling of seed extracts was identical in WT and *vpe*

mutants at distinct molecular weight band patterns to those observed in vegetative tissues, suggesting that NbVPE1a and NbVPE1b do not accumulate to a detectable extent in these seeds. The preservation of VPE activity in seeds suggested that propagation of the mutants would not be affected, and indeed no change in seed germination rate was observed between WT and the *vpe1a/b* mutant (**Fig. 5**). Therefore, activity of NbVPE1a and NbVPE1b is primarily restricted to vegetative tissues.

NbVPE1a and NbVPE1b share 94.9% pairwise amino acid identity and probably result from the most recent genome duplication in *N. benthamiana* (**Chapter 4, Fig.1**) (Bally *et al.*, 2018). An alignment of NbVPE1a and NbVPE1b against well characterised VPEs highlights the strong conservation of the catalytic domain (**Fig. 6**) (Haywood *et al.*, 2018; Zauner *et al.*, 2018). The C-terminal legumain stabilisation and activation module (LSAM) is comparatively less well conserved between plants. Both NbVPE1a and NbVPE1b contain a putative N-glycosylation site at N146, which is absent in other plant species. NbVPE1b contains another putative N-glycosylation site at N332. Notably amongst the other variant residues of NbVPE1a/b, three were within the aligned region of confirmed VPE N-terminal processing in other species; N/S 42 D/G 46 and Y/D 48. As N and D reflect VPE substrate specificity, these could represent alternative N-terminal pro-peptide processing sites. In contrast to N terminal processing, C-terminal processing is not reported to occur at a conserved region between VPEs and may even occur at multiple positions. Of the characterised processing sites, NbVPE1a and NbVPE1b align with the C-terminal cleavage sites of sunflower HaAEP2 (351-355) and castor bean RcVPE (422-423) and could potentially be used in these proteases (**Fig. 6**) (Haywood *et al.*, 2018).

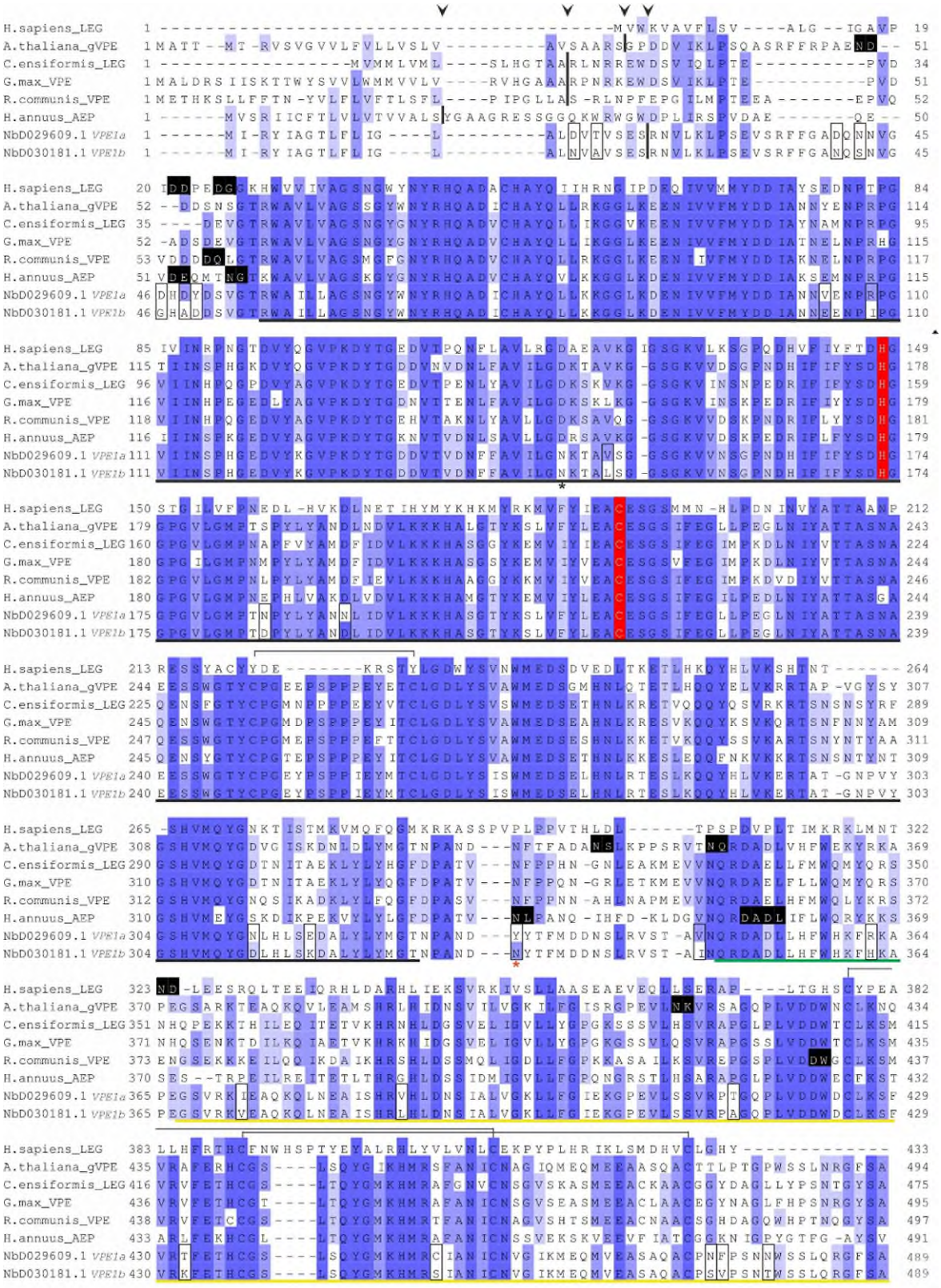


Figure 6. Alignment of well-characterised plant and human VPEs. Alignment was generated by ClustalOmega (Sievers *et al.*, 2014) and annotated in JalView (Waterhouse *et al.*, 2009). Catalytic residues are highlighted in red. Signal peptide cleavage site marked by black vertical line and downward arrow above the alignment. Black highlighted residues are flanking described autocatalytic processing sites. Cysteine bridges are marked by lines above

the sequence and based upon AtVPEγ crystal structure. Bars under the sequence mark the following domains: black = catalytic, green = activation peptide, yellow = LSAM. Residues differing between NbVPE1a and NbVPE1b are boxed. Glycosylation sites are marked with * below the sequence for NbVPE1a and NbVPE1b and present in red if not conserved (Halfon *et al.*, 1998; Hiraiwa *et al.*, 1999; Chen *et al.*, 2000; Kuroyanagi *et al.*, 2002; Li *et al.*, 2003; Hatsugai *et al.*, 2006; Haywood *et al.*, 2018; Zauner *et al.*, 2018).

To further investigate the activity of the NbVPEs, the open reading frames encoding NbVPE1a and NbVPE1b were cloned for expression in *E. coli* with N-terminal His and maltose binding protein (MBP) and C-terminal His tags (**Fig. 7A**). Activity of the recombinant proteases was confirmed by labelling with JOPD1 (**Fig. 7B**). Both proteases could autocatalytically mature at pH 5 in the presence of DTT, resulting in the release of the N-terminal His-MBP tag. The catalytic C215S mutant could not be labelled and only accumulated as a full length tagged protein product. If we assume processing of NbVPE1a and NbVPE1b occurs just before the start of the C13 peptidase domain, the molecular weights of detected JOPD1 signals demonstrate that only proteoforms processed both N- and C-terminally are labelled, suggesting that intermediates do not accumulate or cannot be labelled. Labelled NbVPE1a accumulates as a single ~37 kDa signal (**Fig. 7B**). This is slightly smaller than the three distinct JOPD1-labelled NbVPE1b signals at ~38 kDa (**Fig. 7B**). As proteins are produced without glycosylation in *E. coli*, the discrepancy is probably due to differential processing sites (**Fig. 6**). Upon prolonged incubation at 37°C, the distribution of labelled NbVPE1b shifts from the highest molecular weight band to the lowest (**Fig. 7C**). This processive cleavage may be a result of less optimal cleavage sites or dependency on initial processing events occurring first. Detection of JOPD1 labelled

NbVPE1a and NbVPE1b also demonstrates that glycosylation is not required for activity, as observed for other plant VPEs (Bernath-Levin *et al.*, 2015).

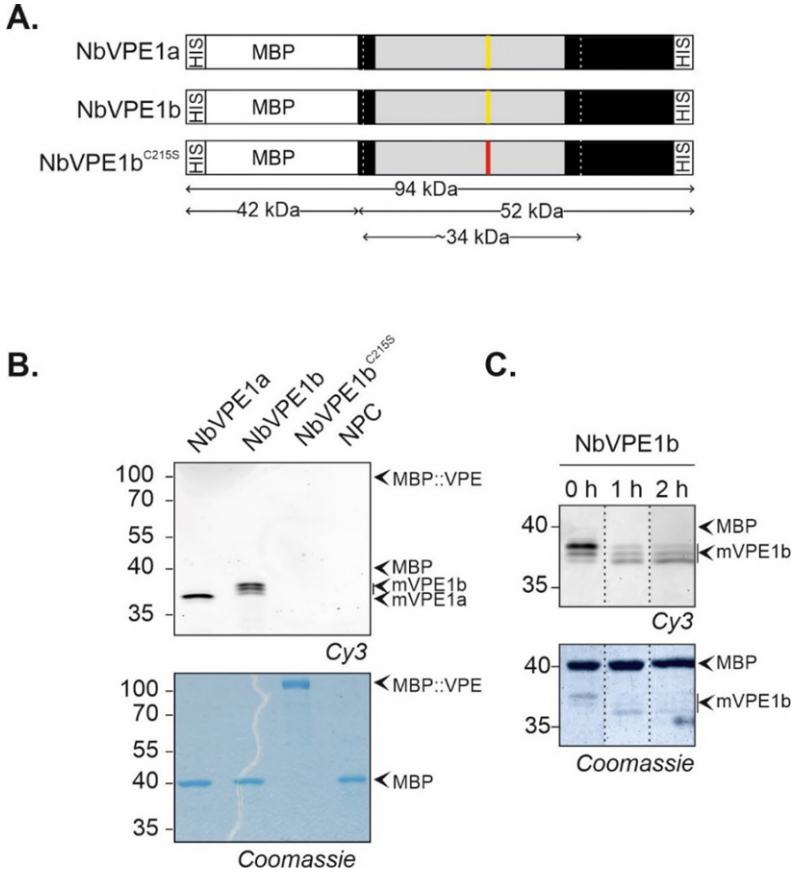


Figure 7. Active NbVPEs can be produced recombinantly in *E. coli*. A. NbVPE1a and NbVPE1b were cloned into pET28b-based expression vectors with N-terminal 6x his and maltose binding protein tags and a C terminal 6x his tag. A catalytic inactive NbVPE1b^{C215S} mutant was generated. *E. coli* T7 SHuffle cells were transformed with expression vectors and protein expression was induced by IPTG. **B.** Proteins were purified by his tag purification and labelled for 3 hours with 2 μM JOPD1. Proteins were separated by SDS PAGE and probe visualised by in gel fluorescence scanning. mVPE = matured VPE **C.** NbVPE1b autocatalytically processes at multiple sites. Purified protein was incubated for 0, 1 or 2 hours at 37°C and then active proteoforms were detected with JOPD1 labelling as performed in **B.** Putative autocatalytic processing sites are delineated with white dashed lines. NPC = no probe control.

The specificity of NbVPE activity was interrogated using synthetic peptide substrates. ESEN and YVAD sequences are established substrates for VPEs (Dall and Brandstetter,

2016). Purified recombinant His-MBP-NbVPE1a-His and His-MBP-NbVPE1b-His could both cleave ESEN and YVAD substrates, but no cleavage was detected of any other tested caspase substrates (**Fig. 8A**). In line with this, activity against Ac-ESEN-MCA was greatly reduced in leaves of the *vpe1b* mutant and was further reduced to negligible rates in the *vpe1a/b* mutant (**Fig. 8B**). Furthermore, *vpe1a/b* mutants transiently overexpressing NbVPE1a-BFP or NbVPE1b-His displayed a dramatic increase in ESENase activity which was not observed upon expression of the inactive NbVPE1a^{C215S}-His mutant (**Fig. 8C**). Therefore, NbVPE1a and NbVPE1b are the predominant contributors to ESENase activity in *N. benthamiana* leaves.

In direct contrast, activity against Ac-YVAD-AFC was not reduced in the mutants at pH 5. Similarly, transient overexpression of NbVPEs in the *vpe1a/b* mutant did not result in an increase in YVADase activity. This is surprising given that Ac-YVAD-cho can inhibit NbVPE1a and NbVPE1b (**Fig. 2B, Fig. 4C**). However, YVADase activity was found to be much greater in isolated AF compared to total leaf extracts and slightly supressed in the *vpe* mutant AF compared to WT (**Fig. 8D**). Therefore, NbVPE1a and NbVPE1b are greater contributors to YVADase activity in AF than in total extracts, but are not the only proteases involved. Altogether these results confirm NbVPE1a and NbVPE1b can process YVAD substrate but are either not responsible *in vivo* for this activity and additional proteases are known with YVADase activity. Indeed, multiple subtilases have been described to exhibit caspase-like activity in plants (Coffeen and Wolpert, 2004; Bonneau *et al.*, 2008; Vartapetian *et al.*, 2011; Reichardt *et al.*, 2018). It will be interesting to see the

contribution of VPEs to YVADase activity in leaves in the presence of the subtilase inhibitor, Epi1 (Tian *et al.*, 2004).

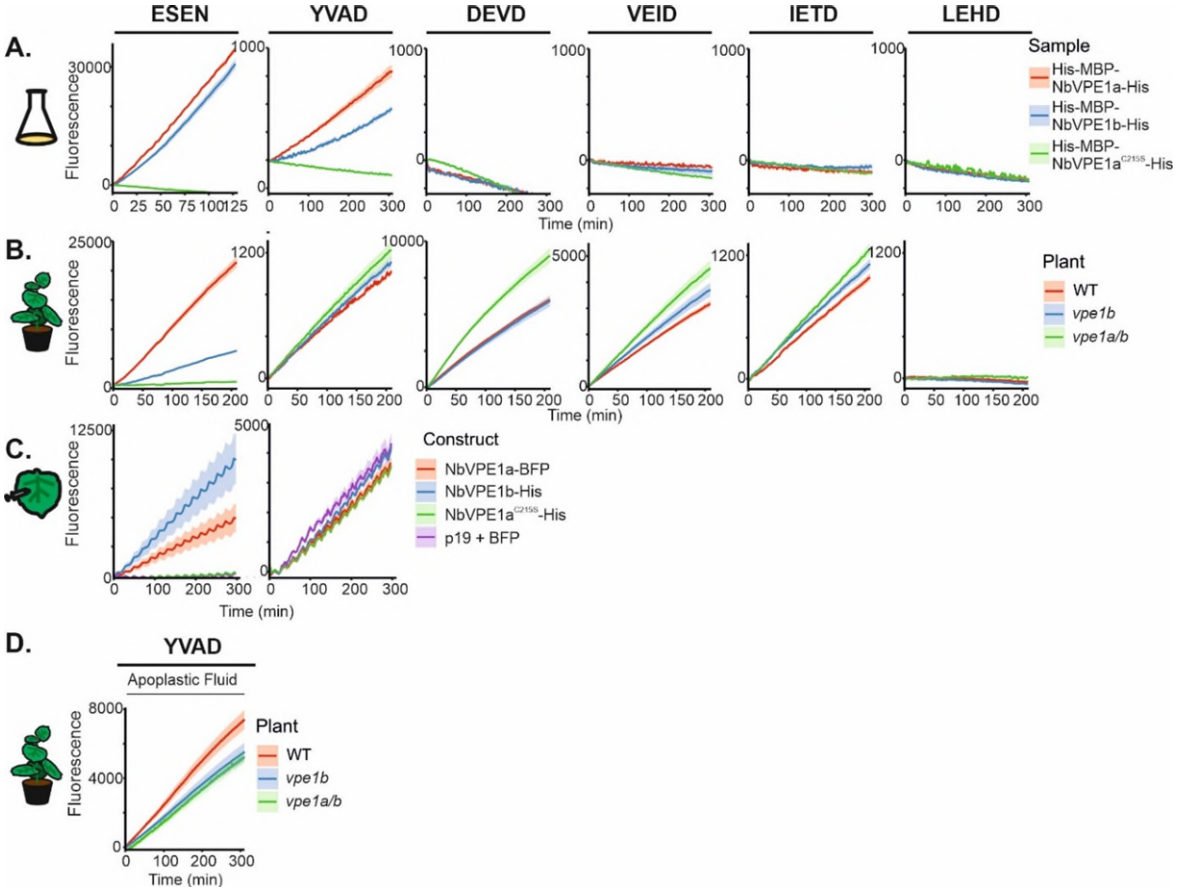


Figure 8. NbVPE1a and NbVPE1b are the major contributors to ESENase activity *in planta*. Proteolytic activities were assayed using synthetic substrates in the following samples: NbVPEs were produced and purified from *E. coli* **(A.)** or were present in leaf extracts from WT, *vpe1b* and *vpe1a/b* plants (**B.**) or NbVPEs were transiently overexpressed in the *vpe1a/b* mutant background by agroinfiltration (**C.**) or finally in apoplastic fluid isolated from mature leaves of WT, *vpe1b* or *vpe1a/b* (**D.**). Proteolytic activities were detected with 50 µM (YVAD-AFC, DEVD-AFC, VEID-AFC, IETD-AFC, LEHD-AFC) or 25 µM (ESEN-MCA) synthetic substrates. Fluorescence of fluorophore released on proteolytic cleavage was measured over time. Error bar shading represents standard deviation, n =3. **A.** and **B.** experiments performed twice, **C.** and **D.** experiments performed only once. Y axis scale differs between experiments.

Other caspase substrates tested were cleaved with similar efficiency in WT, *vpe1b* and *vpe1a/1b* mutant extracts with the exception of Ac-DEVD-AFC (caspase-3 activity), which

was processed faster in the mutant (**Fig. 8B**). Recombinant NbVPE1a and NbVPE1b were also unable to cleave the other caspase substrates, demonstrating the strict substrate specificity and importance of residues aside from P1 (**Fig. 8B**). Strikingly no activity was detected against the Ac-LEHD-AFC (caspase-9) substrate in total extracts, despite previously being reported in *N. benthamiana* leaves, albeit in a TRV silencing background (Kim *et al.*, 2003).

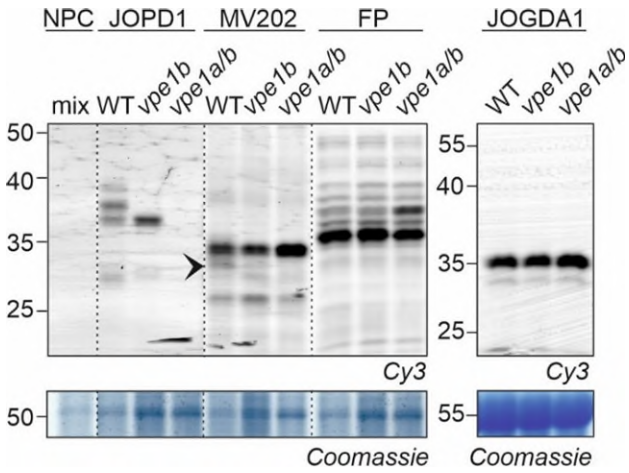


Figure 9. Cysteine proteases show differential labelling in *vpe* mutants. Activity of different protease classes were monitored in leaves of WT and *vpe* mutants. **A.** Leaf extracts were incubated with 2 µM JOPD1 (VPEs), 2 µM MV202 (cysteine proteases) 0.25 µM FP (serine hydrolases), and 2 µM JOGDA1 (cathepsin Bs). Samples were incubated for 3 hours with the exception of FP at 1 hour. Labelled proteins were detected by in gel fluorescence scanning. Samples consist of three biological replicates labelled independently and then pooled. Labelling profile is representative of at least two experiments. NPC = no probe control.

Protease activities can be highly interconnected, whereby removal of one protease often changes the regulation of another (Fortelny *et al.*, 2014). The impact of NbVPE activity on other classes of hydrolases was investigated with activity-based protein profiling (ABPP). Probes targeting serine hydrolases (FP), cysteine proteases (MV202) and cathepsin Bs

(JOGDA1) were used to reveal profiles of active proteases in WT, *vpe1b* and *vpe1a/b* mutant leaves (**Fig. 9**). For serine hydrolases and cathepsin Bs no consistent differences between genotypes were detected. In contrast, an additional cysteine protease signal around 30 kDa was present in WT, potentially representing a protease activity dependent upon VPE. It is unclear which protease(s) is contributing to this signal.

3.3.1 Biotic stress responses in *vpe* mutants

In light of the clear biochemical phenotype of the *vpe1b* and *vpe1a/b* mutants, we set out to characterise the mutants to explore the biological role of NbVPE1a and NbVPE1b. Previous studies using *NbVPE1a/b* silencing have implicated these VPEs in HR induction (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010). To investigate HR in our *vpe1b* and *vpe1a/b* mutants, we challenged leaves with a variety of HR inducers representing viral (Rx and CP), fungal (Avr9 and Cf9, Avr4 and Cf4), bacterial (AvrBS2 and BS2, AvrPto and Pto) and oomycete (Avr8 and R8, Avrvnt1 and Rpivnt1, Inf1, Avr1 and R1, Avrblb2 and Rpiblb2) pathogens. A dilution series of agrobacterium containing plasmids for transient expression of HR inducers was infiltrated in WT and *vpe* mutants. No visual phenotype in intensity or onset of HR were detected (**Fig. 10A**). Given the detection of residual VPE activity observed in ABPP (**Fig. 4B**), we hypothesised a phenotype may be too subtle for visual observation. Accordingly, the timing of HR induction was measured by ion leakage of HR induced by co-expression of Av4 and Cf4, AvrBS2 and BS2, Inf1, AvrPto and Pto, R1 and AVR1, Avrblb2 and Rpiblb2, or Rx and CP. However, induction of HR in leaves of *vpe1a/b* was not delayed for any of the inducer combinations tested (**Fig. 10B**).

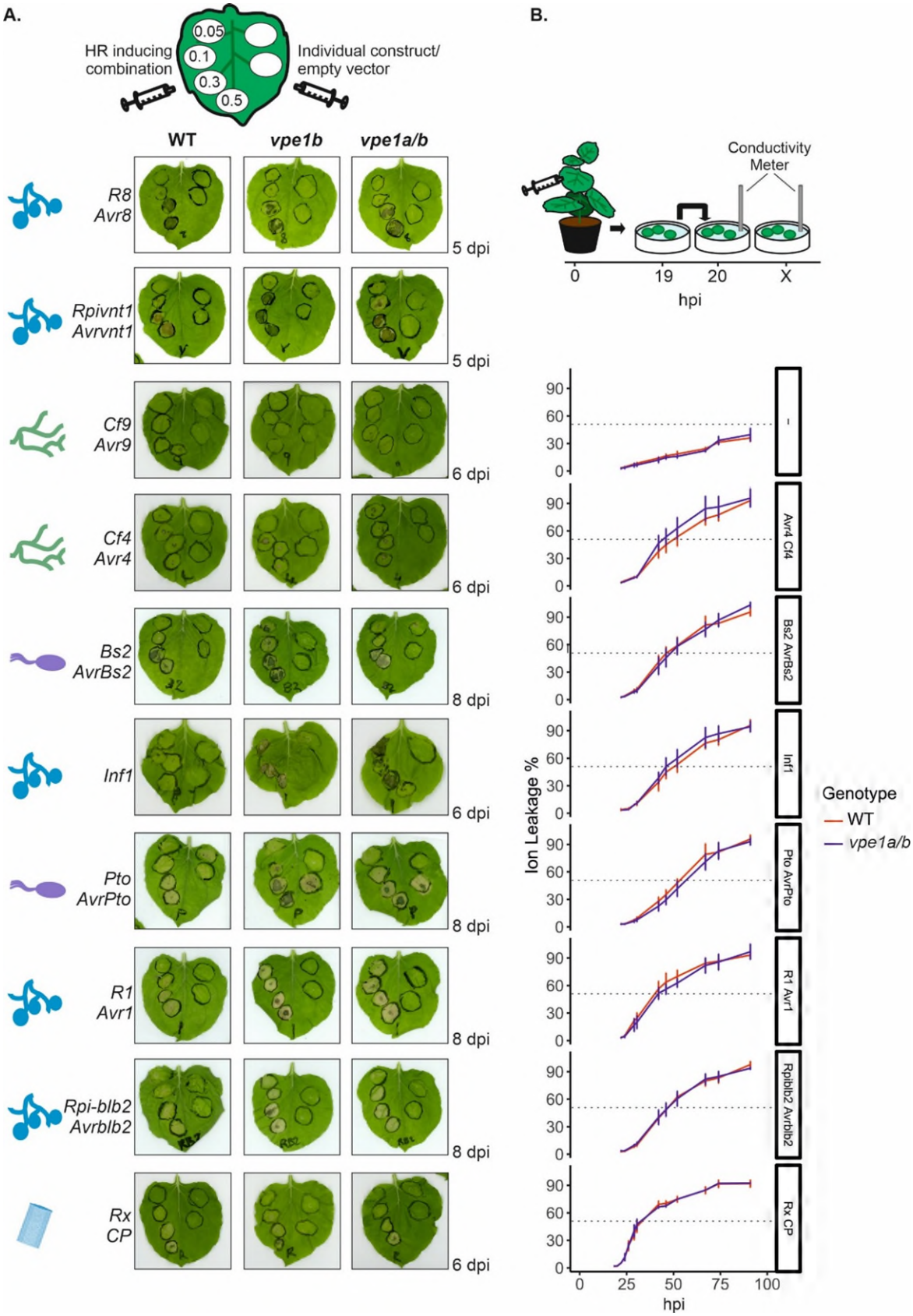


Figure 10. HR induced by agroinfiltration constructs is not blocked in the *vpe* mutants. **A.** HR was induced in WT, *vpe1b and vpe1a/1b* mutant leaves by transient expression of a dilution series of inducers by agroinfiltration. Images were taken at the indicated days post

infiltration (dpi). Images were representative of 5 biological replicates from a minimum of two experiments. **B.** *vpe1a/b* mutant were unaffected in ion leakage upon HR induction. Leaves were infiltrated with different combinations of HR inducers and leaf disks harvested at 19 hpi. Leaf disks were floated in water and a conductivity meter used to measure ion leakage over a time course of cell death. Dashed line = 50%. Error bars represent standard deviation of n = 6. Experiment repeated once with similar outcome.

Considering the resilience of HR induction in the *vpe1a/b* mutants in our agroinfiltration based system, we theorised that other proteases may be upregulated to compensate for the lack of VPE activity in leaves. We therefore monitored the presence of mature proteases upon induction of HR with ABPP. Rx and CP induction was selected given the previous reported role of NbVPEs in HR against viral pathogens (Hatsugai *et al.*, 2004). In line with the ion leakage data, HR progression was not visually different between WT and *vpe1a/b* plants (**Fig. 11A**). Activity-based labelling over a time-course upon transient Rx and CP expression revealed that no other VPEs were upregulated in the *vpe1a/b* leaves (**Fig. 11B**). Furthermore, activity profiles of serine hydrolases (FP) and cysteine proteases (TK009) did not display any consistent differences between WT and mutant in repeated experiments. The overall profile in TK009 labelling of *vpe1a/b* was subtly different to WT with a minor band at ~35 kDa that was consistently stronger in the *vpe1a/b* leaves than in WT leaves (**Fig. 11B**). It is unclear which protease may contribute to this activity. HR dependent differences were observed with the FP labelling independent of VPE. Both WT and *vpe1a/b* showed a clear reduction in three signals in the FP profile upon leaf tissue collapse at 27 and 39 hpi (**Fig. 11B**). Previously a serine-carboxypeptidase has been reported to decrease in activity in *Solanum lycopersicum* HR induced by *C. fulvum* which is therefore a possible candidate for the hydrolase causing this signal (Sueldo *et al.*, 2014).

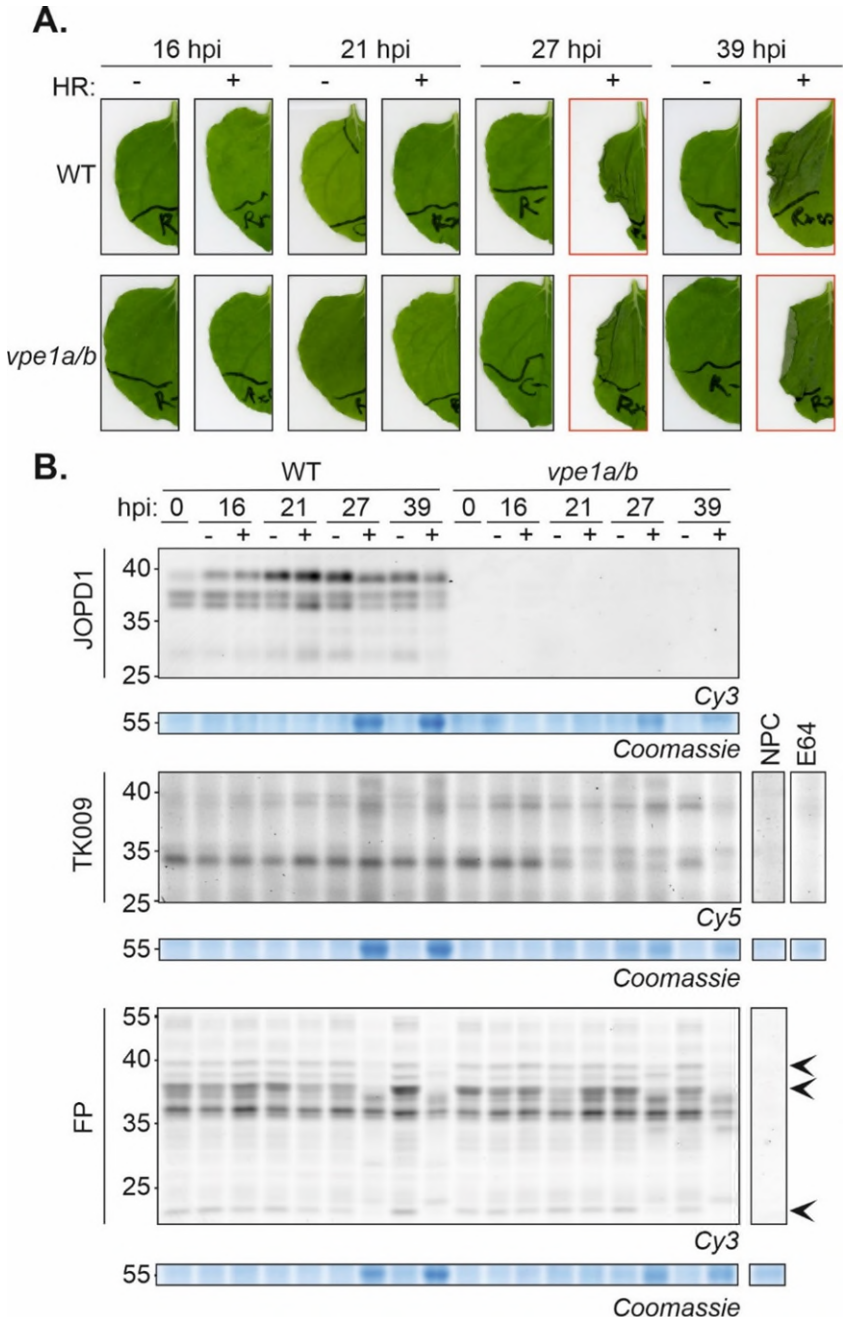


Figure 11. NbVPE1a and NbVPE1b are not necessary for HR induced by Rx and CP. **A.** Protease activities at different time points upon HR induction were investigated by addition of activity-based probes JOPD1 (VPEs), FP (serine hydrolases) and TK009 (cysteine proteases) at 2 µM, 0.25 µM and 2 µM respectively. Samples were incubated for 3 hours (JOPD1 and TK009) or 1 hour (FP) and labelled proteins detected as described previously. Samples consist of three biological replicates labelled independently and then pooled. Labelling was repeated on separate HR induction with a similar result. NPC = no probe control. **B.** Images of leaves at point of harvest. At 21 hpi no visual phenotype is observed, at 27 hpi turgor pressure has been lost (red borders). At 39 hpi tissue has lost turgor pressure and become more desiccated. Leaves were agroinfiltrated with Rx and CP (+) or with either Rx or CP alone (-) at a final concentration of OD_{600} = 0.5.

Finally, we sought to determine if NbVPE1a and NbVPE1b contribute more generally to plant immunity aside from HR. Growth of avirulent *Pseudomonas syringae* pv. tomato (Pto) DC3000 and the virulent Pst DC3000 ΔHopQ mutant which lacks the type III effector HopQ1-1 (Wei *et al.*, 2007) was monitored on WT and *vpe1a/b* mutant leaves. No difference was observed by colony count at either 2 or 3 dpi (**Fig. 12A**). In addition, lesion growth caused by *Phytophthora infestans* strain 88609 (Kamoun *et al.*, 1998) on detached leaves similarly displayed no difference between WT and *vpe1a/b* in either necrotic area or sporulating growth area (**Fig. 12B**). NbVPE1a and NbVPE1b therefore do not play an essential role in defence responses to these pathogens.

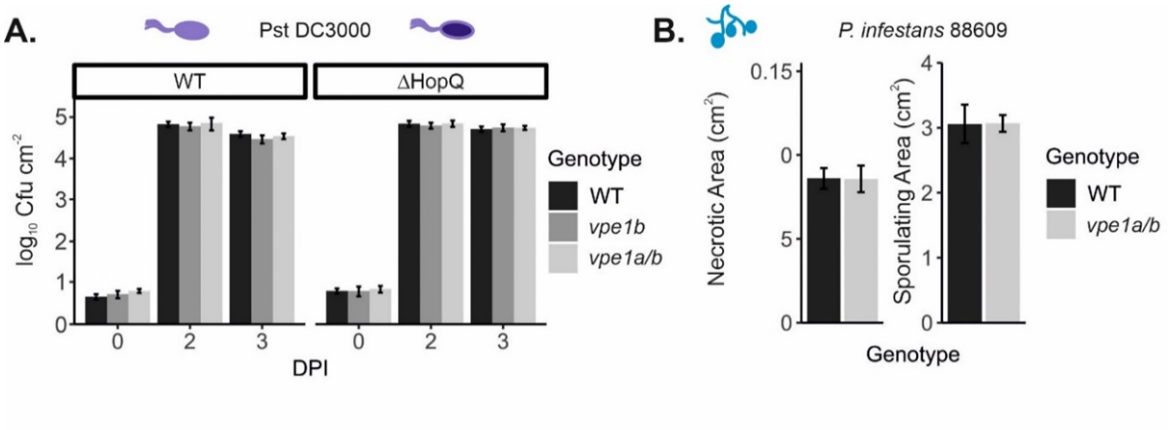


Figure 12. *Pseudomonas syringae* and *Phytophthora infestans* growth is not compromised in the *vpe* mutants. A. Mature *N. benthamiana* leaves were infiltrated with OD_{600} = 0.0002 of different *P. syringae* WT or ΔHopQ strains. Colony forming units (Cfu) were counted at 2 and 3 dpi. Error bars represent standard deviation of n = 4. Infections and colony counts were repeated once with similar results. **B.** Detached *N. benthamiana* leaves were droplet inoculated with P*. infestans* spore suspension at $5x10^4$ $sporesml^{-1}$. Images were taken and the area of necrotic tissue of total visible sporulating area were taken at 5 dpi. Infection repeated three times with similar outcomes. (Necrotic area: WT n = 12, *vpe* n = 8, sporulating area: WT n= 26 *vpe* n = 42). Error bars represent standard deviation.

3.3.2 Abiotic stress responses in the *vpe1a/b* double mutant

In the absence of a detectable defect in biotic stress response in the *vpe* mutants, several other abiotic stresses were tested for a phenotype in the *vpe1a/b* mutant; namely ER stress, senescence, salt tolerance and heat shock.

Arabidopsis VPEs are positive regulators of vacuolar rupture and PCD downstream of ER stress (Pattison and Amtmann, 2009; Qiang *et al.*, 2012; Mendes *et al.*, 2013). Tunicamycin (Tm) inhibits N-glycosylation of proteins in the ER, leading to an accumulation of incorrectly processed protein aggregates that ultimately trigger PCD. Tolerance to ER stress was tested with seedlings grown on media containing Tm for three days before returning to native recovery plates and monitoring plant recovery (**Fig. 13A**). Tm treatment inhibited root growth and biomass accumulation, but this did not differ significantly between WT and *vpe1a/b* seedlings (**Fig. 13A,B,C**). DK004 labelling showed no additional VPE signals for the *vpe1a/b* mutant at either 2 or 5 days post transfer (dpt) to recovery plates (**Fig. 13D**), so the lack of phenotype is unlikely to be due to upregulation of remaining NbVPEs. The timing of PCD induction by Tm stress was then investigated by measuring ion leakage from leaf disks floating in Tm or DMSO containing water. Consistent with seedling responses, WT and *vpe1a/b* mutant demonstrated comparable timing of cell death induction (**Fig. 13E**). Therefore, NbVPE1a and NbVPE1b are not essential for Tm-induced PCD.

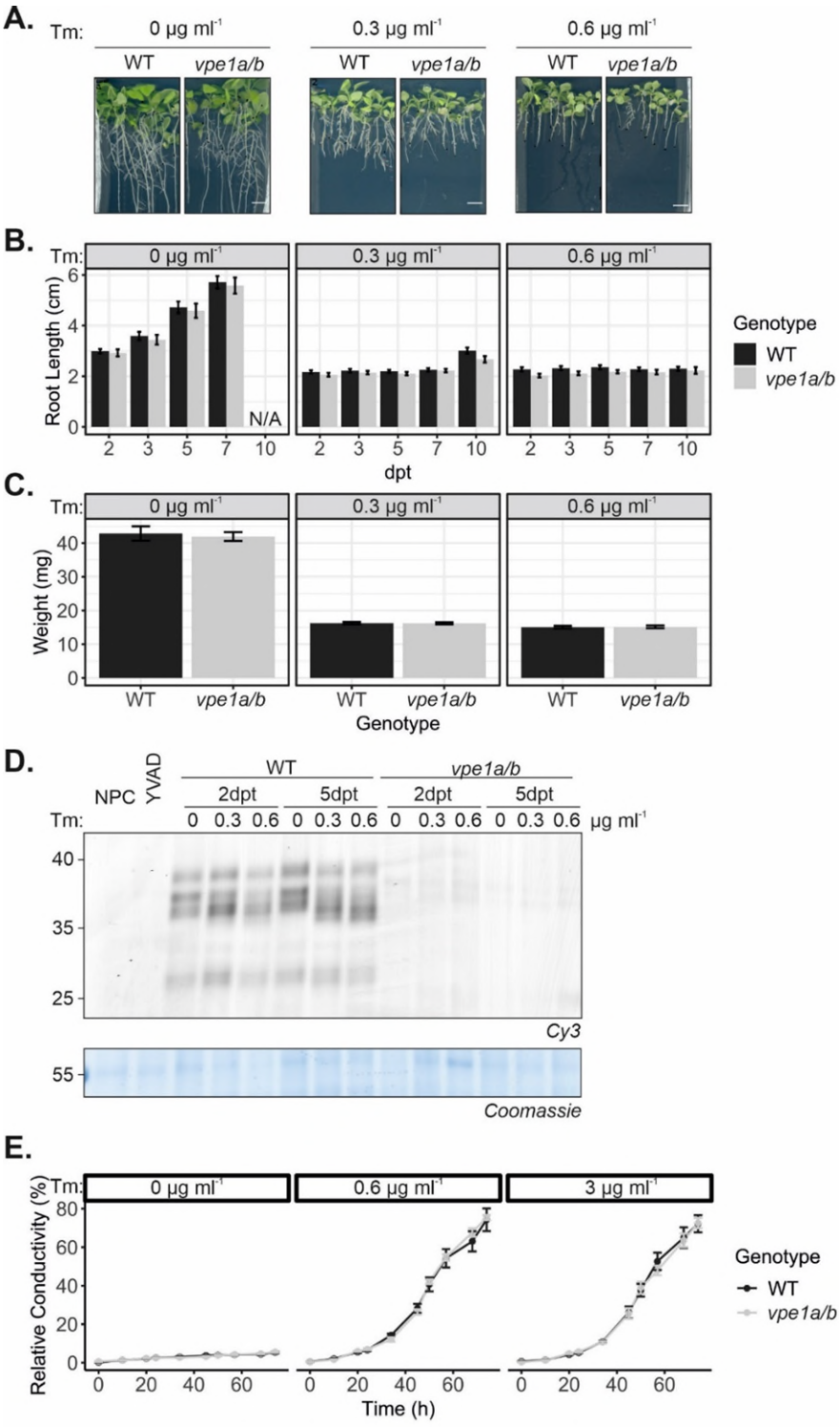


Figure 13. *vpe1a/b* mutant displays no difference in growth in the presence of ER stress inducer tunicamycin. A. Images of seedlings at 7 days post transfer (dpt) to recovery plate. Seeds were germinated on 0.4% 1/2 MS agar plates. At 10 days post germination seedlings were transferred to 0.5% agar plates supplemented with different concentrations of tunicamycin or equal volume DMSO. After 3 days of exposure, seedlings were transferred to ½ MS 0.8% agar recovery plates. Scale bar = 1 cm. **B.** Seedling primary root length. N.B. no measurement recorded for 0 µg ml^{-1} because root length exceeded plate length. n

varied for each treatment: 0 µg ml^{-1}: WT (n = 19), *vpe1a/1b* (n = 18), 0.3 µg ml^{-1}: WT (n = 46), *vpe1a/b* (n = 44), 0.6 µg ml^{-1}: WT (n = 19), *vpe1a/b* (n = 18). Error bars plot the standard deviation. **C.** Wet weight of seedling at 7 dpt to recovery plate. **A,B,C** and **E** are representative of two highly similar experiments. **D.** Activity-based labelling of VPEs. Whole seedlings were ground at 2 or 5 dpt and the protein extract labelled with 0.25 µM DK004 incubated for 3 hrs at room temperature. Proteins were separated by SDS PAGE and labelled proteins detected by in-gel fluorescence scanning. Samples are three independently labelled replicates pooled, experiment only performed once. **E.** Ion leakage measurements of leaf disks floating in milliQ supplemented with tunicamycin or equal volume DMSO. Error bars represent standard deviation of n= 6 biological replicates.

VPEs have been identified amongst upregulated genes in senescence in some species. The potential role of NbVPE1a and NbVPE1b in nitrogen metabolism was tested by growing WT and mutant seedlings on different concentrations of nitrogen. Plants were grown on commercial media with (1X) or without (0X) nitrogen sources. Additionally, half (0.5X) or double (2X) nitrogen concentration conditions were created with the addition of NH_4NO_3 in proportionate amounts (**Fig. 14**). As expected, the growth of plants was optimal in the 1X condition (**Fig. 14A,B,C**). However, no significant difference was observed in the root growth or biomass accumulation of WT and *vpe1a/b* seedlings across all conditions (**Fig. 14B,C**). DK004 labelling showed no additional signals in the *vpe1a/b* mutant upon treatment, indicating that the lack of phenotype is not due to upregulation of other NbVPEs (**Fig. 14D**).

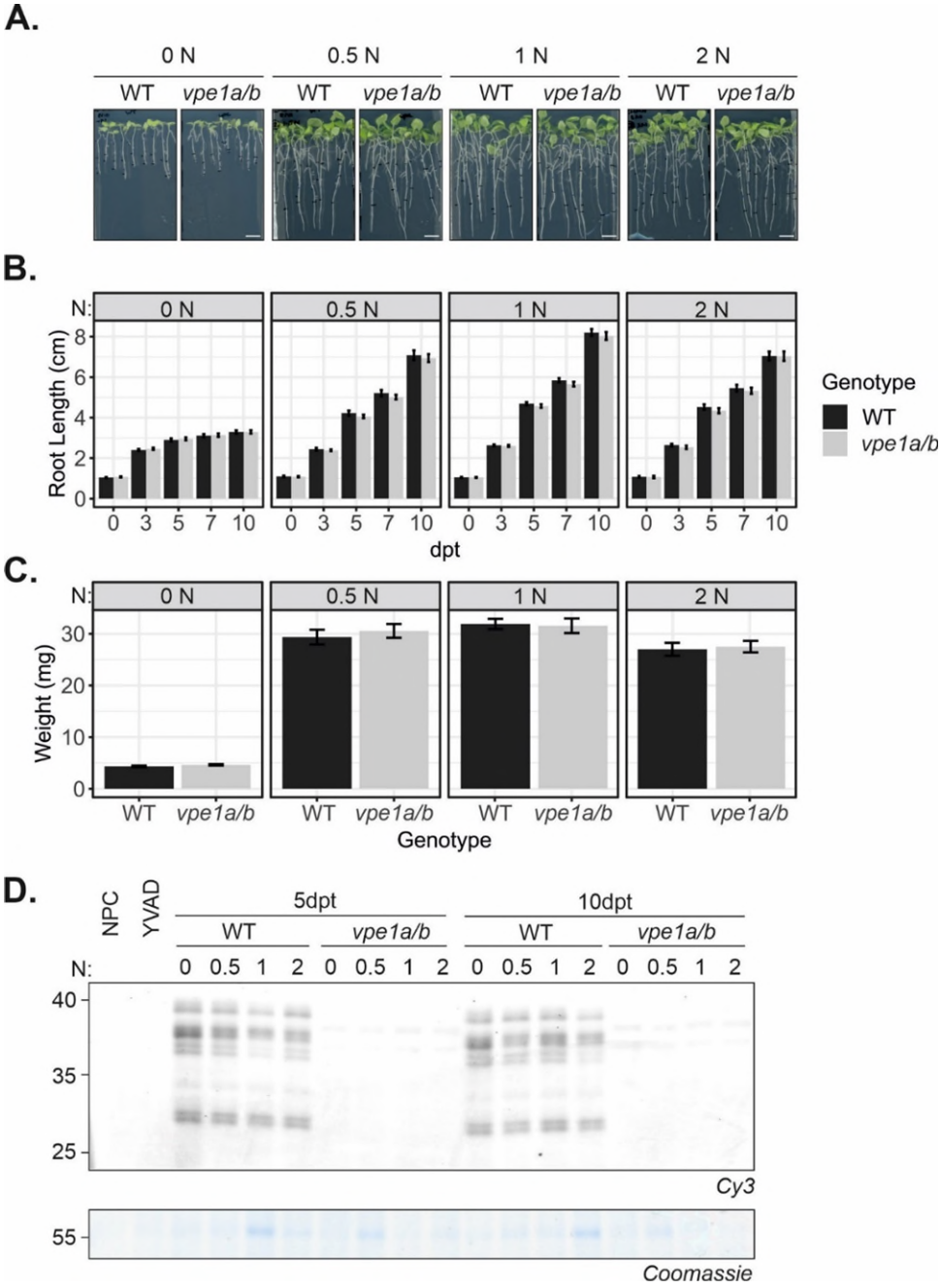


Figure 14. Seedling growth under different nitrogen conditions was not perturbed in *vpe1a/b* plants. A. Images of seedlings at 10 dpt. Seeds were germinated on 0.5% 1/2 MS agar plates. At 7 days post germination seedlings were transferred to 0.8% agar plates ½ MS with different concentrations of nitrogen (NH_4NO_3) from 0N to 2N. 0N = without nitrogen, 0.5 N = NH_4NO_3 1.2 mM, 1N = with nitrogen media (2.4 mM NH_4NO_3), 2N = with nitrogen media supplemented with NH_4NO_3 (4.8 mM NH_4NO_3). dpt = days post transfer. Scale bar = 1 cm. **B.** Length of primary root. n varied between each treatment: 0N: WT (n = 79), *vpe1a/b* (n = 79), 0.5N: WT (n = 49) *vpe1a/b* (n = 45), N: WT (n = 79), *vpe1a/b* (n = 76), 2N WT (n = 47), *vpe1a/b* (n = 46). Error bars represent standard deviation. **C.** Wet seedling weight at 10 dpt. Measurements of three biological replicates with approximately 17-20 seedlings per biological replicate. Error bars represent standard

deviation. **A,B,C** repeated experiment with highly similar outcomes. **D.** Activity based labelling of VPEs. Whole seedlings were ground and protein extract labelled with 0.25 µM DK004 as described in **Fig. 13**. Samples are three independently labelled replicates pooled, experiment only performed once.

We next tested the salt tolerance of *vpe1a/b* mutant seedlings following reports in rice, Arabidopsis and apple on a positive role of VPEs in salt stress responses (Lu *et al.*, 2016; Song *et al.*, 2020). After seven days growing on media supplemented with three different concentrations of NaCl, no difference was observed in the root growth or seedling wet weight between the genotypes (**Fig. 15A,B,C**). Again, DK004 labelling showed no upregulation of alternative VPE activities in the mutant. Therefore, NbVPE1a and NbVPE1b are not required for salt stress responses (**Fig. 15D**).

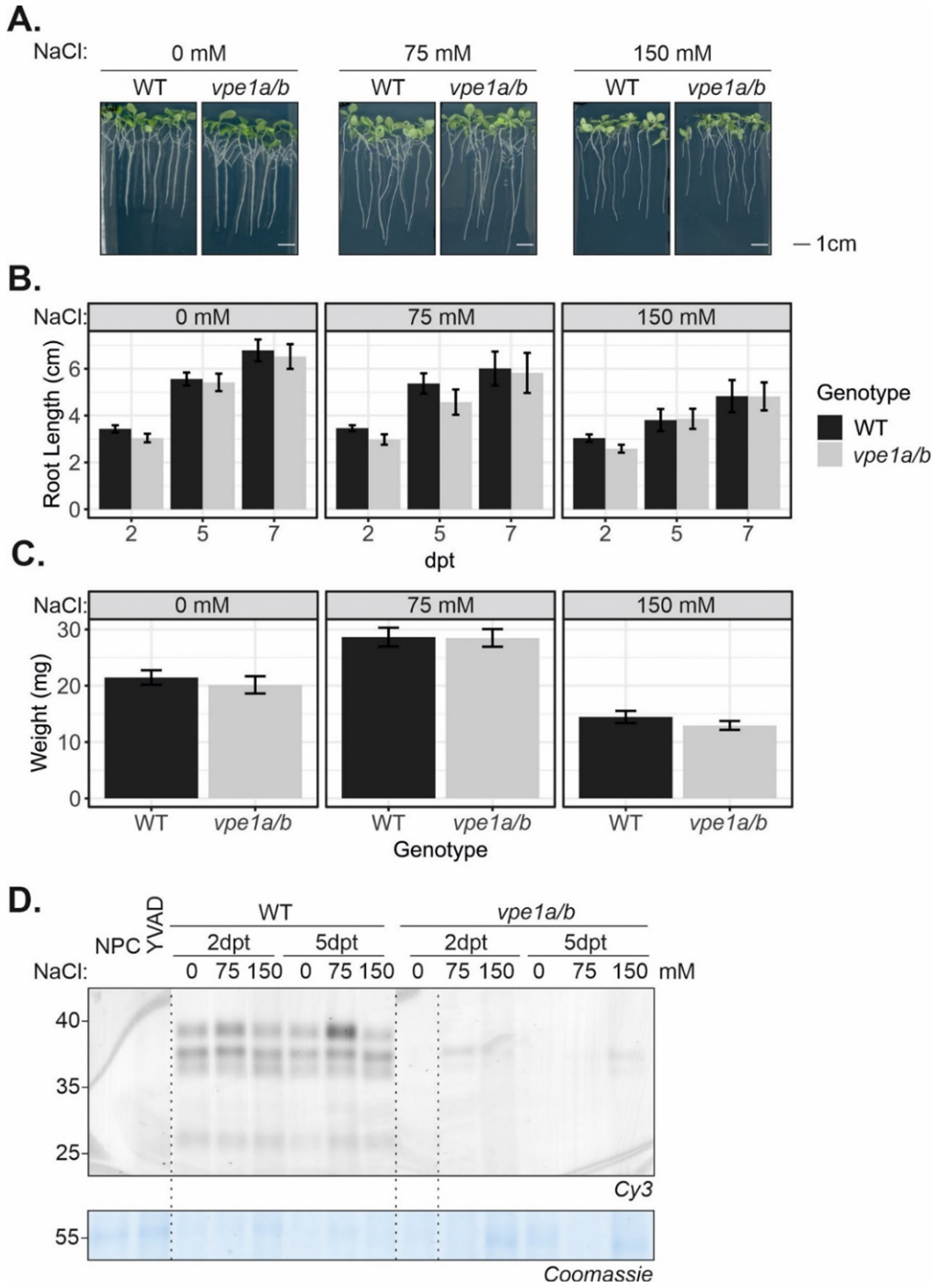


Figure 15. Seedling growth under different salt conditions was not perturbed in the *vpe1a/b* mutant. A. Images of seedlings at 7 dpt. Seeds were germinated on 0.4% 1/2 MS agar plates. At 10 days post germination seedlings were transferred to 0.8% agar plates supplemented with different concentrations of NaCl. Scale bar = 1 cm. **B.** Seedling primary root length. Minimum n per treatment: 0 mM: WT (n = 9), *vpe1a/b* (n = 10), 75 mM: WT (n = 10), *vpe1a/b* (n = 9), 150 mM: WT (n = 10), *vpe1a/b* (n = 9). Error bars represent standard deviation. **C.** Wet weight of seedlings at 7 dpt. Error bars represent standard deviation. **A,B,C** repeated experiment with highly similar outcomes. **D.** Activity based labelling of VPEs. Whole seedlings were ground and protein extracts labelled with 0.25 µM

DK004 as described in **Fig. 13**. Samples are three independently labelled replicates pooled, experiment only performed once.

The final stress applied to seedlings was heat stress, inspired by the finding that *vpeγ* mutants in Arabidopsis are compromised in vacuolar rupture and cell death upon heat shock (Li *et al.*, 2012). WT and *vpe1a/b* mutants were heat shocked for one hour at 46°C, 48°C or 52°C, before being returned to favourable conditions for recovery. Growth after heat shock was comparable in both WT and mutant with similar root length and wet weight measurements (**Fig. 16**). Therefore NbVPE1a and NbVPE1b are not essential for responses to heat stress in seedlings.

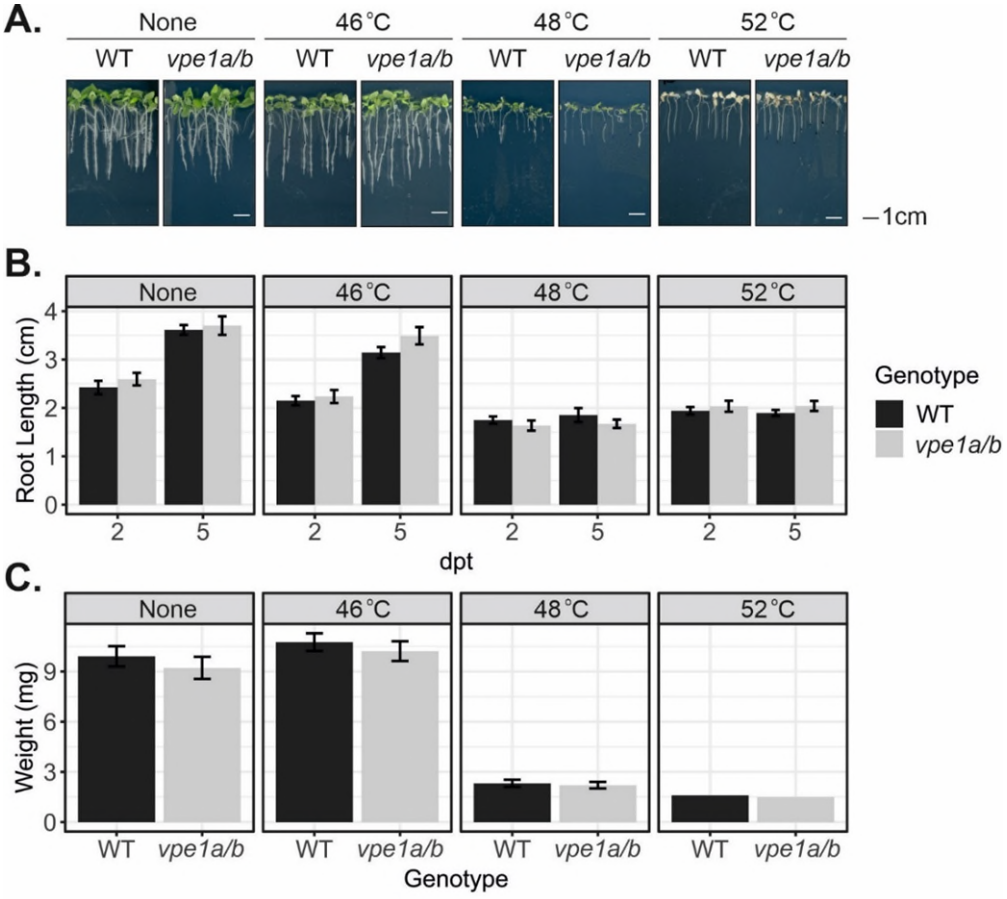


Figure 16. Response to heat shock was not perturbed in the *vpe1a/b* mutant. A. Images of seedlings at 5 dpt (days post treatment). Seeds were germinated on 0.8% 1/2 MS agar

plates. At 10 days post germination plates were wrapped in foil and heat shocked at the indicated temperature for 1 hour. Plates recovered in the dark for 2 hours before being returned to a growth chamber. Scale bar = 1 cm. **B.** Seedling primary root length n for each treatment: None: WT (n = 19), *vpe1a/b* (n = 19), 46°C: WT (n = 20), *vpe1a/b* (n = 19), 48°C: WT (n = 19), *vpe1a/b* (n = 20), 52°C: WT (n = 20), *vpe1a/b* (n = 20). Error bars represent standard deviation. **C.** Wet weight of seedlings at 5 dpt.

3.4 Discussion

In this chapter we have built a toolbox of resources to investigate NbVPE1a and NbVPE1b. This includes the generation of the first stable *N. benthamiana* CRISPR-Cas9 mutant lines lacking the two main vegetative VPEs, establishing NbVPE recombinant expression in *E. coli* and agroinfiltration constructs for NbVPEs. Using these resources we have revealed classic VPE properties of NbVPE1a and NbVPE1b activity and investigated their roles in response to biotic and abiotic stresses.

3.4.1 NbVPE1a and NbVPE1b activity

Activity-based labelling with a VPE-specific probe revealed a complex profile of VPE activity in vegetative tissues of *N. benthamiana*. The multiple bands observed in labelled extracts represent stable products of different VPE genes varying in sequence, processing and putative glycosylation sites. Additionally, multiple proteoforms of the VPE can be labelled, as illustrated by labelling of NbVPE1a/b overexpressed in leaves and produced in *E. coli*. Through the generation of the *vpe1a/b* mutant lacking functional NbVPE1a/b, we have demonstrated that NbVPE1a and NbVPE1b represent the major VPE activities in the vegetative tissues, as *vpe1a/b* extracts display negligible (but not null) ESENase activity and signals in the VPE-specific probe DK004 profile. Indeed, the ESENase activity in leaf

tissue directly correlated with DK004 labelling with WT and the *vpe* mutants. These data are consistent with the literature as VPEs are the only proteases known to cleave after N in humans and plants to date (Dall and Brandstetter, 2016). Indeed all AANase activity has been attributed to VPEs in some Arabidopsis tissues (Nakaune *et al.*, 2005).

Similar to purified VPEs of other species, NbVPE1a and NbVPE1b produced in *E. coli* could autocatalytically mature and exhibited conventional substrate specificity against ESEN and YVAD (Halfon *et al.*, 1998; Okamoto *et al.*, 1999; Li *et al.*, 2003; Kuroyanagi *et al.*, 2005; Bernath-Levin *et al.*, 2015; Hatsugai *et al.*, 2015; Dall and Brandstetter, 2016; James *et al.*, 2019). Interestingly, despite the ability of NbVPE1a/b to cleave YVAD substrate *in vitro*, they were not important contributors to YVADase activity in total leaf extracts. This contradicts previous evidence from leaves of *NbVPE1a/b* silencing where reduced ESENase activity correlated with reduced YVADase activity (Hatsugai *et al.*, 2004). However, this conflict may be reconciled when considering the discrepancies in experimental design. Hatsugai et al. performed substrate assays in the presence of a suite of inhibitors that block other protease classes, including subtilases (phytaspases and saspases) that can cleave YVAD substrates (Hatsugai *et al.*, 2004; Vartapetian *et al.*, 2011; Reichardt *et al.*, 2018). There is a widespread emphasis in the literature on the association of caspase-1 (YVADase) activity and implication of VPEs, often with no further supporting evidence. Our experiments form a cautionary note that *in vitro* and *in vivo* substrate specificity cannot be conflated and that many unrelated activities may contribute to YVADase activity *in vivo*.

NbVPE1a/b had no detectable activity against DEVD, VEID, IETD and LEHD caspase substrates. The inability of NbVPE1a/b to cleave caspase substrates other than YVAD is consistent with specificity of previously characterised VPEs. AtVPEγ is similarly unable to cleave DEVD and tobacco VPEs are insensitive to a DEVD-based inhibitor (Hatsugai *et al.*, 2004; Kuroyanagi *et al.*, 2005). In comparison to other proteases with caspase-like activity, VPEs exhibit high substrate specificity. For example, tobacco phytaspase can cleave VEID, YVAD, IETD, TATD and LEHD whilst tomato phytaspases collectively have activity against eight caspase-like substrates (VAD, YVAD, WEHD, LEHD, VEID, VNLD, IETD, STATD) (Chichkova *et al.*, 2010, 2018; Reichardt *et al.*, 2018).

In a further cautionary note, we demonstrated that the vacuole is not the only compartment with VPE activity. We detected a greater contribution of NbVPE1a/b to the extracellular YVADase activity than intracellular, despite being active in both compartments and the predicted presence of phytaspases in the apoplast. Extracellular localisation is often overlooked in the study of VPEs, but VPEs have been detected in the apoplasts of tomato and *N. benthamiana* (Goulet *et al.*, 2012; Sueldo *et al.*, 2014; Grosse-Holz *et al.*, 2018). The disparity in the relative contribution of NbVPE1a/b to YVAD activity between extracellular and vacuolar compartments, coupled with the contrasting NbVPE1a/b-dependence of intracellular ESENase activity, suggests NbVPE1a/b exhibit different activities in different compartments. Unidentified factors in these compartments may play important roles in regulating VPE activity.

3.4.2 Why was no phenotype observed in *vpe1a/b* mutants?

Our results show NbVPE1a and NbVPE1b are not essential for the stress responses tested in this chapter. This does not, however, exclude a role for VPEs in these stress responses. The *vpe1a/b* mutant is not a null knockout and still probably contains functional NbVPE1c and NbVPE1d which are also α/γ type VPEs. Therefore, it may be necessary to generate null mutants to observe a phenotype. Redundancy amongst all four VPEs has been shown in Arabidopsis. Nevertheless, some phenotypes have still been detected with single mutants (Gruis *et al.*, 2004; Rojo *et al.*, 2004; Nakaune *et al.*, 2005; Hatsugai *et al.*, 2015; Cheng *et al.*, 2019, 2020). Moreover, a dramatic reduction in ABPP labelling and the drastic suppression of ESENase activity was observed in the *vpe1a/b* mutant, although neither of these signals were abolished. Alternatively the stresses applied in this study may not be severe enough to reveal a VPE dependency, or other proteases compensate for the lack of VPEs.

The possibility of compensation by other proteases was investigated in the *vpe* mutants using ABPP in both unchallenged and HR induced plants. No upregulation of serine hydrolases or cysteine proteases was detected in the *vpe1a/b* leaves (**Fig. 9**), however compensation may not require an increase in protease abundance. Furthermore, ABPP may not be sensitive enough to capture all active proteoforms present. Subtilases are prime candidates for compensating for VPE activity as they also cleave after D (Coffeen and Wolpert, 2004; Bonneau *et al.*, 2008; Vartapetian *et al.*, 2011; Reichardt *et al.*, 2018). However, ESENase activity is unique to VPEs, and this activity is not compensated in the unchallenged leaves of the *vpe1b* and *vpe1a/b* mutants (**Fig. 8B**). Yet ESENase activity was not completely abolished in *vpe1a/b* leaves, possibly due to residual VPE activities.

Other protease classes including aspartic, metallo- and threonine proteases, in addition to atypical members of serine and cysteine proteases that escape activity-based labelling, were not investigated and could be involved. Indeed, previous investigation of seed protein processing revealed an aspartic protease that can cleave the same substrates (Gruis *et al.*, 2002, 2004; Shimada *et al.*, 2003). Unfortunately, no probe for aspartic proteases is available yet (Morimoto and van der Hoorn, 2016). Future studies in the *vpe1a/b* mutant using more sensitive activity-based pull down experiments will shed light on the effect of NbVPE1a/b on the proteolytic environment.

Using synthetic substrates we identified elevated DEVDase activity in total extract of the *vpe1a/b* mutant. Elevated DEVDase activity had also been observed in the Arabidopsis *vpe* null mutant, which was attributed to cathepsin B (CathB) (Cai *et al.*, 2018). Furthermore, processing of CathB to the active form was increased in *vpe* null seeds (Weng *et al.*, 2019). However, we found no increase in the abundance of active CathB detected by JOGDA1 labelling in *vpe1a/b* leaves (**Fig. 9**). Alternatively, a proteasome subunit PBA1 can also cleave DEVD and may be responsible (Hatsugai *et al.*, 2009; Gu *et al.*, 2010).

3.4.3 NbVPE1a and NbVPE1b are not essential for biotic stress responses

NbVPE1a and NbVPE1b do not play a role in macroscopic vegetative development, as the *N. benthamiana* CRISPR-Cas9 mutants were indistinguishable from WT plants. This is consistent with Arabidopsis *vpe* null mutants which displayed no growth defects (Gruis *et al.*, 2004; Kuroyanagi *et al.*, 2005; Mylne *et al.*, 2014) indicating that VPEs are not required for plant growth and development. However, NbVPE1a and NbVPE1b do have a reported role in regulating HR PCD (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010). Therefore, HR

induction by a suite of inducers representing all major pathogen taxonomic groups were tested in this chapter. Our results show that NbVPE1a and NbVPE1b are not essential for the induction of HR by any of the tested inducers. These data do not align with preceding evidence for the indispensable role of NbVPE1a and NbVPE1b in HR upon TMV infection in *N. benthamiana* carrying the resistance gene *N* (Hatsugai *et al.*, 2004). Crucially, the experimental system used by Hatsugai et al. differed to ours. First of all, the previous experiments were performed in a PVX infection background that was used for silencing of *NbVPEs*. Secondly, these *NbVPE1a/b* silenced plants were transgenic with constitutive expression of a foreign resistance gene, *N*. Aberrant expression of resistance genes can cause plant stress (Borrelli *et al.*, 2018). Thirdly, plants had been grown at elevated temperature for an extended period prior dropping to 23°C to induce HR synchronously across the infected tissue. Therefore, the *N. benthamiana* used in the previous study were subject to multiple long-term viral, temperature and internal autoimmune stresses that could result in atypical responses to subsequent stress. The challenge in generating *N*-transgenic plants in the *vpe1a/b* double mutant background prohibited us from repeating these experiments. Instead, plants in this chapter were unchallenged until inducers were expressed by agroinfiltration, which then introduced a background of bacterial infection. Clearly the two experimental systems are not directly comparable and induction of HR in these backgrounds may use different components.

Previous studies of NbVPE1a/b had already brought into question their role in HR regulation. Although *NbVPE1a*/b silenced plants revealed a VPE-dependency for harpin-induced HR, no difference in HR induction was observed for boehmerin or Nep1-induced

HR (Zhang *et al.*, 2010). Additionally, HR induced by infiltration of Inf1 in *N. benthamiana* was not blocked upon *NbVPE1a/b* silencing (Shibata *et al.*, 2010). Our experiments with the *vpe1a/b* double mutant further support the hypothesis that NbVPE1a and NbVPE1b are not essential to all HR types and their involvement is dependent on the HR-inducer.

Infection assays with *P. syringae* and *P. infestans* indicate that NbVPE1a/b are not essential for defence responses against these pathogens. Growth of neither virulent or avirulent *P. syringae* was altered on the *vpe1a/b* mutants, in contrast to evidence from Arabidopsis *vpeγ* mutant where more *P. syringae* growth was observed. The contribution of AtVPEγ to defence is likely to be independent of HR as no difference in HR induction was observed in the avirulent avrRpm1 strain on *vpe* null plants (Rojo *et al.*, 2004; Hatsugai *et al.*, 2009). Responses to *P. syringae* may not be conserved amongst α/γ VPEs between Arabidopsis and *N. benthamiana*, or additional NbVPE activities remaining in *vpe1a/b* leaves may be contributing to pathogen defence.

Furthermore, no difference between WT and *vpe1a/b* mutant was observed in *P. infestans* infection assays or HR induction by the *P. infestans*-derived effectors Inf1, Rpiblb2/Avrblb2 or Rpivnt1/Avrvnt1. This confirmed previous insights from *NbVPE1a*/b silenced plants that were not defective in Inf1 induced cell death or *P. infestans* colonisation (Shibata *et al.*, 2010). Additionally, VPEs generally may not be required for defence against other *Phytophthora* species. HR induction by a *P. boehmeriae* elicitor is not *NbVPE1a/b* dependent in *N. benthamiana* and no upregulation of VPEs is observed upon *P. sojae* infection in soybean (Zhang *et al.*, 2010; Jing *et al.*, 2016).

Although *P. syringae* and *P. infestans* represent two vastly different pathogen and infection strategies, the lack of NbVPE1a and NbVPE1b involvement does not disqualify any potential role of other NbVPEs such as NbVPE1c or NbVPE1d in defence, or in defence against other pathogens.

3.4.4 NbVPE1a and NbVPE1b are not essential for abiotic stress responses

VPEs have been implicated in abiotic stress responses, but their role has not been thoroughly examined in *N. benthamiana* (Hatsugai *et al.*, 2015). Using a seedling plate-based stress assays we detected no differences in WT and *vpe1a/b* responses to Tm, nitrogen, salt or heat stresses.

VPE transcription has been used in some studies as markers of ER stress (Mendes *et al.*, 2013; Kørner *et al.*, 2015). However, the *N. benthamiana vpe1a/b* mutant showed no defect in cell death induction by Tm or in seedling growth after Tm treatment. Importantly the evidence for a positive regulatory role of VPEs in Arabidopsis has come from analysing quadruple *vpe* null mutants, not single mutants. Upon application of Tm in Arabidopsis leaves and roots, reduced vacuolar rupture was observed in *vpe* null mutants compared to WT (Qiang *et al.*, 2012; Cai *et al.*, 2018). Tm exposure also induced reduced ion leakage in the Arabidopsis *vpe* null compared to WT (Cai *et al.*, 2018), contrasting to our ion leakage results in the absence of NbVPE1a/b activity. A role of NbVPEs in Tm-induced PCD may still be revealed with inhibition of residual leaf VPE activities.

Similarly, the involvement of VPEs in salt and heat stress responses reported in other species was not found for NbVPE1a and NbVPE1b in this study. VPEs were upregulated

upon heat stress in Arabidopsis with a concurrent reduction in cell death and enhanced growth in the *vpe* null mutants (Li *et al.*, 2012). Application of lethal salt stresses to tomato leaves resulted in increased VPE activity (Kovács, 2017) and salt stress in rice involves transcriptional upregulation of *OsVPE2* and *OsVPE3* (Deng *et al.*, 2011; Kim *et al.*, 2014). Silencing and overexpression of *OsVPE3* demonstrated a role of VPE in regulation of vacuolar rupture and PCD (Lu *et al.*, 2016). However, *N. benthamiana* is not a model organism typically used in heat and salt tolerance research and it may be that optimal conditions were not used in the stress assays reported here.

In contrast to ER and heat stresses, our detection of no role for NbVPE1a/b in senescence is in line with evidence from Arabidopsis research. Our results exclude an essential role of NbVPE1a and NbVPE1b in nitrogen starvation. Similarly VPEs are dispensable for senescence progression in darkened Arabidopsis leaves (Pružinská *et al.*, 2017). Despite this, VPE-dependent degradation of an invertase was detected in senescent leaves, demonstrating that VPEs are active during senescence (Rojo *et al.*, 2003). Indeed, in other species evidence suggests a role for VPEs in senescence. Silencing and overexpression of VPE show a positive contribution to nodule senescence onset (Pierre *et al.*, 2014). Transcriptional upregulation of VPEs had been associated with several senescence systems, including legume nodules, tobacco petal shedding, tobacco pollen development, nitrogen restricted oilseed rape and light deprived Arabidopsis leaves and cortical cells in root development (Kinoshita *et al.*, 1999; Oh *et al.*, 2004; Pierre *et al.*, 2014; van Wyk *et al.*, 2014; Hatsugai *et al.*, 2015; Pružinská *et al.*, 2017; Cheng *et al.*, 2020). Senescence involves degradation and nutrient mobilisation on a massive scale, and thus it is conceivable that a complex mixture of

hydrolytic enzymes could contribute to this process. Parallels can be drawn to the established role of VPEs in seed storage protein processing. Precise processing of seed storage proteins is performed by VPEs, but germination is not perturbed in Arabidopsis *vpe* null mutants due to proteolytic redundancy (Gruis *et al.*, 2004; Mylne *et al.*, 2014). It is plausible that a similar redundancy exists in both storage protein processing and senescence nitrogen mobilisation.

Summary

NbVPE1a and NbVPE1b contribute the major VPE activities in *N. benthamiana* vegetative tissues, but are not essential for the stress responses investigated in this chapter. A question remains over whether other NbVPEs, such as NbVPE1c and NbVPE1d, could compensate for the loss of NbVPE1a and NbVPE1b in the *vpe1a/b* mutant. Creation of *N. benthamiana* *vpe* null mutants are required to address whether these stress responses are truly VPE-independent.

Contributions

Dr. Daniel Krahn, with support from Prof. Chris Schofield, synthesised the new chemical probes; DK004, DK005 and TK009. Dr. Friederike Grosse-Holz designed and cloned the CRISPR-Cas9 construct used. Ursula Pyzio propagated the mutant lines. We thank Prof. Tolga Bozkurt for the gift of the NbVPE1a-BFP expression plasmid.

3.5 Methods

Alignment and phylogeny construction

Alignments were built based on amino acids sequences using ClustalOmega (Sievers *et al.*, 2014). The tree was created in MEGA X using the Maximum Likelihood method based on a Poisson correction model, with 500 bootstrap replicates to generate a consensus tree (Zuckerkandl and Pauling, 1965; Felsenstein, 1985; Kumar *et al.*, 2018).

Plant lines and growth conditions

N. benthamiana plants were grown at a constant 21°C and ~60% relative humidity under a 16 hour light (40 µmol m^{-2} s^{-1}) and 8 hour dark regime.

Generation of *N. benthamiana* mutants

CRISPR-Cas9 construct targeting *NbVPE1a, NbVPE1b* and *NbVPE1c* were designed by a previous lab member (Friederike Grosse-Holz). sgRNAs were designed using RNAscorer and off target effects minimised with use of CasOT (Xiao et al., 2014; Chari et al., 2015). The following sequences were used: *sgRNA1a:* 5'-GTCCTCCAATTGAATACATG-3', *sgRNA1b:* 5'-TCTTAGTGGAGGCAGCGGAAAGG-3' and *sgRNA1c:* *5'-*CCTATGGTTCCCATGTCTTGCAA-3'. The sgRNAs were assembled in the CRISPR-Cas9 cloning vector as described in (Buscaill et al., 2019) generating pFGH188 and the construct was checked with sanger sequencing (Eurofins Genomics). Agrobacterium tumefaciens containing pFGH188 plasmid was agroinfiltrated into N. benthamiana leaves and transformants propagated as described previously (Sparkes et al., 2006; Buscaill et al., 2019). Transformants were screened for VPE mutations in T1, T2 and T3 by ABPP with

JOPD1 probe (see ABPP methods below) and validated by amplification of target sequences using primers 21, 27, 28 and 29 (**Table 1**) and sequencing of amplicons (Eurofins Genomics and Source Bioscience).

Activity based protein profiling (ABPP)

***In vivo* labelling:** Leaves of 4 week old *N. benthamiana* plants were infiltrated with 50 μM YVAD or DMSO and incubated at room temperature for 30 minutes. The same leaf area was re-infiltrated with 2 μM JOPD1 or DMSO and incubated for 3 hours. Harvested tissue was flash frozen, ground, thawed in reducing SDS loading buffer and then separated on a 12% SDS-PAGE reducing gel.

***In vitro* labelling:** Leaf disks from 4 week old *N. benthamiana* were flash frozen, ground and suspended in extraction buffer (50 mM sodium acetate, 5 mM DTT) at pH 5 for JOPD1, DK004, DK005, FP, MV202 or TK009 labelling or at pH 6 for JOGDA1 labelling. Extracts were centrifuged at 18,000 g at 4°C and the supernatant retained for labelling. Alternatively, apoplastic fluid was extracted by complete infiltration of leaves with ice cold water and centrifugation at 1,000 g for 15 minutes with slow acceleration and deceleration in an empty syringe barrel placed within a falcon tube. Protein extracts were preincubated with protease inhibitors phenanthroline (1 mM, Calbiochem), E64 (100 μM, Calbiochem), DCI (100 μM, Calbiochem), PMSF (1 mM, Calbiochem), pepstatin A (0.03 mM, Calbiochem), Ac-YVAD-cho (50 μM, Calbiochem), DK006 (50 μM) or DMSO for 30 minutes at room temperature. Subsequently, the activity-based probes JOPD1, JOGDA1 and MV202 were added to 2 μM or DK004, DK005, FP and TK009 added to 0.25 μM. A no probe control (NPC) was made by mixing equal volumes of all tested extracts and addition of the same volume of DMSO

as volume of probe used. The sample probe mix was incubated for 3 hours at room temperature in the dark. Labelling was terminated and proteins concentrated by acetone precipitation. Samples were resuspended in reducing SDS gel loading buffer, boiled at 95°C for 5 minutes and separated using over a 12% SDS-PAGE reducing gel. Labelled proteins were detected by in-gel fluorescence scanning with Amersham Typhoon 5 Biomolecular Imager (Amersham Biosciences/GE Healthcare). Fluorophores were detected at Cy3 settings (532nm excitation, 610 band pass filter) or Cy5 setting (633nm excitation, 670 band pass filter).

VPE production in *E. coli*

VPE expression constructs were transformed by heat shock into NEB T7 SHuffle strain of *E. coli* (New England Biolabs). A starter culture of *E. coli* was grown overnight in Luria Bertani (LB) media supplemented with 100 $\mu g ml^{-1}$spectinomycin and 50 $\mu g ml^{-1}$ kanamycin at 28°C 200 rpm. A fresh culture was inoculated in a 1:100 ratio and was grown to OD_{600} = 0.5 – 0.8. Expression of VPEs was induced by addition of IPTG to a final concentration of 0.1 mM. Cultures were incubated at 16°C overnight (~20 hours) at 200 rpm. Cells were pelleted by centrifugation at 1,000 g for 10 minutes and resuspended in ice cold lysis buffer (100 mM Tris HCl pH 8, 100 mM NaCl, 0.05% Triton X-100) to an approximately 30x concentration. Extract was incubated with a pinch of lysozyme (Sigma-Aldrich) for 30 minutes and then sonicated at maximum power for five cycles of 20 s on, 60 s off over ice. DNA was sheared by passing ten times through a narrow gauge 27g needle under high pressure. Protein extracts were then cleared by centrifugation at 21,000 g at 4°C. VPEs were purified by his tag purification using a homemade column packed with Nickel NTA

agarose beads (Sigma-Aldrich). Tagged proteins were eluted in 100 mM Tris HCl pH 8, 100 mM NaCl, 300 mM imidazole. The buffer was exchanged to activation buffer (50 mM sodium acetate, pH 5, 50 mM NaCl, 5 mM DTT, 1 mM EDTA) with a 3 kDa cut off ultrafiltration spin column (Merck, Germany). Protein concentration was determined using Bradford assay against BSA standards. If not used immediately, protein aliquots were flash frozen in liquid nitrogen and stored at -80°C.

Cloning of NbVPEs

NbVPEs were selected for cloning based on biological relevance defined by; homology to AtVPE$_\gamma$, previous investigation in the literature, identification in pre-existing in-house *N. benthamiana* proteomics datasets and being amongst the 10% most abundant RNA transcripts (Grosse-Holz *et al.*, 2018). NbVPE1a (Niben101Scf04539g03005, NbD029609.1) and NbVPE1b (Niben101Scf04675g08014, NbD030181.1) were amplified without signal peptide coding sequence (as predicted by Signal P) from cDNA synthesised from *N. benthamiana* leaves (Armenteros *et al.*, 2019). Due to the use of Golden Gate cloning system, the genes were amplified in two fragments to remove an internal BsaI recognition site by introduction of a synonymous mutation in the PCR. Golden Gate compatible overhangs introduced onto the amplified fragments by PCR were digested with BpiI (New England BioLabs, USA) and ligated into a Level 0 module vector with T4 DNA ligase (New England BioLabs, USA) to generate pET003 (NbVPE1a) and pET004 (NbVPE1b). E. coli expression constructs were generated by Golden Gate assembly using BsaI (New England BioLabs, USA) and T4 DNA ligase of the following vector and modules: pJK082c (pET28-b modified), pJP001 (pL0M-PU-T7-LacO-T7 transl Strong RBS), pJP002 (pL0M-T-T7

terminator), pJK184 (pL0M-S-6xHISMBP-HRVBC), EC15259 (pL0M-S-6xHIS-15259) and pET003 or pET004. Agroinfiltration constructs were assembled with the addition of signal peptide encoding sequence annealed in two long primers to the VPE gene cloned in the original level 0 modules pET003 or pET004. These were ligated with digested modules: pJK268c (pL1V-P19-F2), EC41414 (pL0M-T-35S-1-41414), EC51288 (pL0M-PU-35S-TMV-3-51288) and EC15259 (pL0M-S-6xHIS-15259).

Correct construct assembly was confirmed by Sanger sequencing performed by Eurofins Genomics. Cloning of the agroinfiltration expression construct of NbVPE1a failed so a NbVPE1a-BFP construct gifted from Tolga Bozkurt was used for experiments (**Table 2**). Catalytic inactive mutants C215S of NbVPE1a or NbVPE1b were generated by primer directed mutagenesis and amplification of the relevant Level 0 module plasmid.

Synthetic Substrate Assays

Protein concentrations were measured using Bradford assay and extracts were normalised to the same concentration within each experiment. In experiments with inhibitors present, protein extracts were preincubated with protease inhibitors or equal volume DMSO for 30 minutes at room temperature prior to mixing with synthetic substrate mixture. Inhibitors used and concentrations are described above in ABPP methods. Substrate assays were performed using the following synthetic peptide substrates: Ac-IETD-AFC, Ac-DEVD-AFC, Ac-VEID-AFC, Ac-LEHD-AFC, Ac-YVAD-AFC (Santa Cruz Biotechnology, Inc.) and Ac-ESEN-MCA (PeptaNova, Germany). Synthetic peptides were resuspended in DMSO and diluted to 100 μM in assay buffer: 50 mM sodium acetate pH 5, 5 mM DTT, 1 mM EDTA. Synthetic substrate was then mixed in a 1:1 ratio with protein extracts in a 96

well plate clear bottom plate. Fluorescence was measured over time at room temperature in an Infinite M200 plate reader (Tecan, Mannedorf, Switzerland). Settings for AFC detection: 400 nm excitation, 505nm emission, gain = 100 and MCA detection: 360 nm excitation and 440 nm emission, gain = 90.

Agroinfiltrations

Agrobacterium were grown overnight from single colonies in LB at 28°C 200 rpm. Cultures were pelleted at 2,000g for 5 minutes and resuspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$ pH 5.7). Concentration of agrobacterium was adjusted to OD_{600} = 0.5 and acetosyringone was added to a final concentration of 100 µM. Agrobacterium were then incubated at 28°C 200 rpm for 1 hour before, if applicable, mixing in combinations at a 1:1 ratio. The two uppermost mature fully expanded leaves of 4-5 week old *N. benthamiana* plants were infiltrated with the agrobacterium suspension. See **Table 3** For details of agrobacterium used in this chapter.

Ion Leakage

Leaf disks 7 mm in diameter were harvested from mature *N. benthamiana* leaves at relevant time points and washed in sterile milliQ with gentle rotation for one hour. Leaf disks were then carefully transferred to air tight universals containing 10 ml of sterile milliQ and incubated at room temperature. For investigation of the effect of ER stress induced cell death, the 10 ml of milliQ additionally contained tunicamycin (Merck, Germany) at 0.6 μgml^{-1} or 3 μgml^{-1} or equal volume DSMO as a control. The conductivity of the solution was monitored over a time course using a Jenway 3450 pH and Conductivity Meter. Once

the time course was completed, universals were boiled at 85°C for 20 minutes and final solution conductivity measured to obtain maximum potential ion leakage measurement.

Infection Assays

Pseudomonas: *Pseudomonas syringae* WT and ΔHopQ were grown on LBA plates supplemented with rifampicin (25 μgml^{-1}). Cultures were grown at 28°C for two days and then liquid LB cultures were inoculated with single colonies. Bacteria were grown overnight at 28°C 220 rpm. The culture was pelleted and resuspended in sterile demineralised water to a final concentration of OD_{600} = 0.0002. The mature and fully expanded leaves of four week old *N. benthamiana* plants were infiltrated with the bacteria solution and incubated at the growth conditions described above. Colony counting was performed as described (Buscaill *et al.*, 2019) with minor modifications. Briefly, four leaf disks of 2.355 cm^2 were punched and surface sterilised with two washes of 15% H_2O_2 in milliQ. After a final wash with sterile milliQ, excess liquid was evaporated in a sterile flowhood. Leaf disks were then transferred to Eppendorf's with 1ml of sterile milliQ and ground in a tissue lyser (QIAGEN) at 28 shakes s^{-1} for 4 minutes. A dilution series of extracts was then plated onto LBA plates containing rifampicin (25 μgml^{-1}). Plates were incubated at 28°C for 1.5 – 2 days and then colonies were counted.

Phytophthora: *Phytophthora infestans* strain 88609 was grown as described (Caten and Jinks, 1968). *P. infestans* was culture was maintained on rye sucrose agar (RSA) at 18°C in the dark. Cultures were transferred to RSA with b-sitosterol (120 μM) for sporulation induction. To induce sporulation, ice cold sterile milliQ was added to the plates and incubated in the dark and cold for 4-6 hours. The spore containing solution was then

decanted and adjusted to 50 zoospores μl^{-1}. Mature, fully expanded leaves of 3-4 week old *N. benthamiana* plants were detached and used for the infection assay. 10 µl of zoospore suspension was droplet inoculated onto the abaxial side of leaves. The infected leaves were incubated at 18°C, ~99% humidity under 16 hours light (70 µmol $m^{-1} s^{-1}$) and 8 hour dark cycles. Leaves were imaged at 5 dpi and the visible area of infections measured in ImageJ.

Seedling Assays

Seeds were harvested at the same time from WT and *vpe1a/b* mutants for 3 separate batches of plants. Collected seeds were surface sterilised by washing for 1 minute in 50% ethanol followed by six successive sterile milliQ washes. Sterilised seeds were then plated on ½ Murashige and Skoog basal salts (MS, PhytoTech Labs, Kansas, USA) media with 1% sucrose and 0.4-0.8% phytoagar, pH 5.7. Plates were placed in a growth chamber with the same growth conditions as described for plants above.

ER stress: At 10 days post germination on 0.4% ½ MS plates, seedlings were transferred to 0.5% agar plates containing 0.3 or 6 $\mu g ml^{-1}$ of tunicamycin (Sigma-Aldrich) or equal volume of DMSO. After 3 days of incubation with tunicamycin, seedlings were transferred to ½ MS 0.8% agar recovery plates.

Nitrogen stress: At 7 days post germination on 0.5% ½ MS plates, seedlings were transferred to 0.8% agar plates ½ MS with different concentrations of nitrogen. from 0X to 2X. MS media without nitrogen (PhytoTechnology Laboratories, USA) was used at half concentration (1/2 MS) and supplemented with 0 mM NH_4NO_3 (0X) or 1.2 mM NH_4NO_3 (0.5X). MS media with nitrogen (PhytoTechnology Laboratories, USA) was used at half

concentration (1X, containing 2.4 mM NH_4NO_3) or supplemented with 4.8 mM NH_4NO_3 (2X).

Salt tolerance: At 10 days post germination on 0.4% agar plates, seedlings were transferred to 0.8% agar plates supplemented with 0, 75 and 150 mM NaCl.

Heat shock: Seeds were germinated on 0.8% 1/2 MS agar plates. At 10 days post germination plates were wrapped in foil and heat shocked in an incubator oven at the indicated temperature for 1 hour. Plates recovered in the dark for 2 hours before being returned to a growth chamber.

Length of primary root was measured using SmartRoot plugin for ImageJ (Lobet *et al.*, 2011). Wet weight measurement was performed after gentle washing to remove any excess agar on the roots.

Table 1. Primers used within this chapter.

Name	Sequence 5' - 3'	Description
5_VPE_Fw1	TTGAAGACAAAGGTCGTAATGTTTTGAAACTTCCGTCAG	Cloning *NbVPE1a* and *NbVPE1b* with BpiI restriction site and GoldenGate compatible overhangs
6_VPE_Fw2-1	TTGAAGACAAGGACTCAAAGATGAGAACATTGTAGTG	Cloning *NbVPE1a* to mutate endogenous BsaI restriction site with BpiI restrution site and GoldenGate compatible overhangs
7_VPE_Fw2-2	TTGAAGACAAGGACTCAAAGATGAGAATATTGTTGTG	Cloning *NbVPE1b* to mutate endogenous BsaI restriction site with BpiI restriction site and GoldenGate compatible overhangs
8_VPE_Rev1	TTGAAGACAAGTCCACCCTTCTTCAACAG	Cloning *NbVPE1a* and *NbVPE1b* to mutate BsaI restriction site with GoldenGate compatible overhangs
9_VPE_Rev2	TTGAAGACAACACCTGCACTAAAACCCCTCTGGA	Cloning *NbVPE1a* and *NbVPE1b* with BpiI restriction site and GoldenGate compatible overhangs
21_GVPE Fw2-2	CCAATACCAGGAGTTATCATTAATAGCCC	Genotyping *NbVPE1b*
27_GVPE Rev2-2	CCTACTTGCTGCAGTAGATATCATTTG	Genotyping *NbVPE1b*
28_GVPE Fw1	GGACTGACAAACTCATGCAAAG	Genotyping *NbVPE1a*
29_GVPE Rev1-2	CGTGTTGAAGAAGAAGCATGC	Genotyping *NbVPE1a*

117_VPE_C215S_Fw	TTGAAGACAATAGTGAGTCTGGTAGT ATATTTGAGGGACTT	*NbVPE1a* and *NbVPE1b* C215S mutagenesis primer with Bpi1 restriction site and GoldenGate compatible overhangs
118_VPE_C215S_Rev	TTGAAGACAAACTAGCTTCAAGGTAA AATACCAAG	*NbVPE1a* and *NbVPE1b* C215S mutagenesis primer with Bpi1 restriction site and GoldenGate compatible overhangs
116_VPE_Fw3	TTGAAGACAACGTAATGTTTTGAAACT TCCGTCAG	Cloning *NbVPE1a* and *NbVPE1b* with goldengate compatible overhangs for addition of signal peptide sequence

Table 2. Plasmids constructed in this chapter.

ID	Description	Selection	Gene IDs
pET003	pL0M-C1-NbVPE1a	Spec	NbD029609.1 / Niben101Scf04539g03005
pET004	pL0M-C1-NbVPE1b	Spec	NbD030181.1 / Niben101Scf04675g8014
pET012	pET28b-T7::MBP-N10-HRV3C-NbVPE1a-His	Kan	NbD029609.1 / Niben101Scf04539g03005
pET013	pET28b-T7::MBP-N10-HRV3C-NbVPE1b-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET020	pL0M-C1-NbVPE1aC215S	Spec	NbD029609.1 / Niben101Scf04539g03005
pET021	pL0M-C1-NbVPE1bC215S	Spec	NbD030181.1 / Niben101Scf04675g8014
pET022	pL0M-C1-NbVPE1a (with native signal peptide)	Spec	NbD029609.1 / Niben101Scf04539g03005
pET023	pL0M-C1-NbVPE1b (with native signal peptide)	Spec	NbD030181.1 / Niben101Scf04675g8014
pET027	pET28b-T7::MBP-N10-HRV3C-NbVPE1bC215S-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET028	pL0M-C1-NbVPE1aC215S (with native signal peptide)	Spec	NbD029609.1 / Niben101Scf04539g03005
pET029	pL0M-C1-NbVPE1bC215S (with native signal peptide)	Spec	NbD030181.1 / Niben101Scf04675g8014
pET035	pL2M-P19-2x35S::NbVPE1b-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET036	pL2M-P19-2x35S::NbVPE1aC215S-His	Kan	NbD029609.1 / Niben101Scf04539g03005

Table 3. Agrobacterium used in this chapter.

Protein Expressed	Agrobacterium Strain	Selection	Reference	Received from
Rx	C58C1	Rif, Kan	(Bendahmane *et al.*, 1999)	Baulcombe Lab
CP	C58C1	Rif, Kan	(Bendahmane *et al.*, 1999)	Baulcombe Lab
Avr4 (pRH87)	MOG101	Rif, Kan	(Van Der Hoorn *et al.*, 2000)	van der Hoorn lab
Cf4 (pRH48)	MOG101	Rif, Kan	(Van Der Hoorn *et al.*, 2000)	van der Hoorn lab

Avr9 (pMOG978)	MOG101	Rif, Kan	(Van Der Hoorn *et al.*, 2000)	van der Hoorn lab
Cf9 (pRH21)	MOG101	Rif, Kan	(Van Der Hoorn *et al.*, 2000)	van der Hoorn lab
Avr8	AGL1	Rif, Spec	(Vossen *et al.*, 2016)	Kamoun Lab
R8	AGL1	Rif, Kan	(Vossen *et al.*, 2016)	Kamoun Lab
Pto	C58	Rif, Kan	(Rathjen *et al.*, 1999; De Vries *et al.*, 2006)	Kamoun Lab
AvrPto	GV3101	Rif, Kan, Gent	(Rathjen *et al.*, 1999; De Vries *et al.*, 2006)	Kamoun Lab
Rpi-vnt1	GV3101	Rif, Kan	(Foster *et al.*, 2009; Pel, 2010)	Kamoun Lab
Avrvnt1	GV3101	Rif, Spec	(Pel, 2010)	Kamoun Lab
Bs2	GV3101	Rif, Kan, Gent	(Tai *et al.*, 1999)	Kamoun Lab
AvrBs2	GV3101	Rif, Kan, Gent	(Tai *et al.*, 1999)	Kamoun Lab
Inf1	MOG101	Rif, Kan	(Kamoun *et al.*, 2003)	van der Hoorn lab
R1	GV3101	Rif, Kan, Gent	(Ballvora *et al.*, 2002)	Kamoun Lab
Avr1	GV3101	Rif, Spec, Gent	(Du *et al.*, 2015)	Kamoun Lab
Rpi-blb2	GV3101	Rif, Spec, Gent	(Bozkurt *et al.*, 2011)	Kamoun Lab
Avrblb2	GV3101	Rif, Kan, Gent	(Oh *et al.*, 2009; Bozkurt *et al.*, 2011)	Kamoun Lab
NbVPE1a-BFP	GV3101	Rif, Spec, Gent	*Not published*	Bozkurt Lab
BFP	GV3101	Rif, Spec, Gent	*Not published*	Bozkurt Lab
NbVPE1b-6xHis	GV3101	Rif, Kan, Gent	Chapter 3	
NbVPE1a^{C215S}-6xHis	GV3101	Rif, Kan, Gent	Chapter 3	

Chapter 4: Identifying substrates of VPEs in *Nicotiana benthamiana* leaves

Emma Thomas[1], Fatih Demir[2], Andreas Perrar[2], Ben Excell[1], Pitter Huesgen[2], and Renier van der Hoorn[1]

1. Department of Plant Sciences, University of Oxford
2. ZEA-3, Jülich Forschungszentrum, Germany

4.1 Abstract

Currently there are no known substrates of VPEs in *Nicotiana benthamiana*. High-efficiency Undecanal-based N-Termini EnRichment (HUNTER) was performed on the leaf proteome of wild type and *vpe1a/b* mutant plants and upon transient overexpression of NbVPE1b and a NbVPE1a catalytic inactive mutant. Using HUNTER we identified several candidate substrates of VPEs. The N-terminome revealed significant changes in protease activity between the *vpe1a/b* and WT leaves and a clear preference of NbVPE1b for P1 N upon NbVPE1b overexpression. Several candidate substrates, including the receptor-like protein NbFERONIA and multiple cysteine proteases, were taken forward for studies to confirm NbVPE-dependent processing.

4.2 Introduction

Vacuolar processing enzymes (VPEs) are a type of cysteine protease which are widely distributed across plants and animals. As the name suggests, protease activity is restricted to acidic compartments including the vacuole and extracellular spaces. VPE autocatalytically matures through sequential removal of N- and C-terminal pro-peptides (Hiraiwa *et al.*, 1999; Zauner *et al.*, 2018).

VPEs were initially identified and characterised in seeds (Hara-Nishimura *et al.*, 1991). Subsequent investigation of proteolytic activity revealed a preference for cleavage after asparagine residues by VPEs (Müntz *et al.*, 2002). In parallel, prolific research into the processing of seed storage proteins revealed a requirement for N-specific proteolysis in their maturation (Mylne *et al.*, 2014). Storage proteins serve as a reservoir of nitrogen and sulphur to fuel the growing seedling. Many require specific proteolytic processing to generate stable structures. Mutation of N residues resulted in misprocessed protein complexes (Jung *et al.*, 1993, 1998), arousing suspicions that these could be VPE substrates. VPEs were known to colocalise with storage proteins (Hiraiwa *et al.*, 1993). Evidence then quickly emerged that VPEs could process storage proteins *in vitro*. Storage proteins from pumpkin cotyledons were shown to be digested by castor bean VPE (Hara-Nishimura *et al.*, 1991). Similarly, purified castor bean VPE could cleave the pumpkin PV100 storage protein into multiple functional products at N-Q cleavage sites (Yamada *et al.*, 1999). In subsequent years the role of VPEs in processing storage proteins was confirmed by a mutant screen in Arabidopsis (Shimada *et al.*, 2003). Mutations in single VPE genes, in particular *AtVPEβ,* perturbed the accumulation of correctly processed storage proteins, whilst a quadruple null mutant

showed complete disruption in protein processing (Gruis *et al.*, 2002, 2004; Shimada *et al.*, 2003). Despite this clear evidence for VPEs in protein processing, *vpe* null mutants did not have any macroscopic phenotype and progressed through germination and seedling growth comparable to WT (Mylne *et al.*, 2014).

Amongst the processed proteins in seeds are cyclic peptides. Cyclic peptides are highly resistant to proteolytic degradation due to their atypical conformation and are commonly associated with defence functions, including toxic peptides and potent protease inhibitors (James *et al.*, 2018). Synthesis requires excision from a precursor and ligation of exposed peptide termini. Select VPEs are able to perform both cleavage and ligation of several types of cyclic peptides (Min and Jones, 1994). The balance between processing and ligation activities differs between VPEs, from butelase-1 which cannot cleave canonical substrates, to many VPEs which cannot catalyse ligation reactions (Nguyen *et al.*, 2014).

VPE activity is not restricted to seeds and other biological roles have been a major focus in the more recent literature (Hatsugai *et al.*, 2015; Vorster *et al.*, 2019; Yamada *et al.*, 2020). VPEs have been implicated in a variety of PCD types, in addition to various roles in biotic and abiotic stress responses (Hatsugai *et al.*, 2004, 2015). Despite the intensive research into the role of VPEs, their protein processing activity has not been extensively studied (**Table 1**).

Progress was first made in Arabidopsis where two potential substrates were characterised in senescing leaves. Following previous evidence in yeast studies, a serine carboxypeptidase AtCPY was found to be subject to VPE-dependent processing in Arabidopsis senescent leaves (Kinoshita *et al.*, 1999; Rojo *et al.*, 2003). An investigation of proteins co-expressed

with VPEs upon senescence lead to identification of VPE-dependent degradation of an invertase AtFruct4 (Rojo *et al.*, 2003). However, despite the presence of VPE-dependent events, the onset of senescence is not defective in *vpe* mutants (Rojo *et al.*, 2003; Pružinská *et al.*, 2017).

More recently, a substrate and phenotype of VPE has been paired. Tapetal degradation and xylem developmental cell deaths were both delayed in single Arabidopsis *vpe* mutants (Cheng *et al.*, 2019, 2020). A similar phenotype had previously been observed in mutants lacking the cysteine protease, AtCEP1 (Han *et al.*, 2019). Both AtVPEβ and AtVPEγ were shown to activate AtCEP1 *in vitro*. Activation of AtCEP1 correlated with the presence of VPE *in vivo*, providing strong evidence for this being a true VPE substrate (**Table 1**) (Cheng *et al.*, 2019, 2020).

For a protein to be considered a true substrate of a protease, the following evidence is required: i) protease-dependent *in vivo* processing; ii) confirmed *in vivo* co-localisation of protease and substrate; iii) purified protease processing of substrate *in vitro*; and iv) mutation of the putative cleavage site to generate an "uncleavable" protein. These are challenging requirements to curate for each substrate. Evaluating the evidence for all reported VPE substrates in the literature revealed the majority of these lack direct evidence for processing (**Table 1**). Testing of VPEs with ligation activity have been the most robust studies in terms of processing evidence, however the majority of these have been tested with non-native peptides and not with endogenous substrates. Aside from this, Arabidopsis proteins dominate the list of known processing substrates.

A major drawback to this candidate-lead approach to substrate discovery is the reliance on the often-sparse literature in specific areas of plant biology. Employing high-throughput mass-spectrometry based techniques has already enabled unbiased and large-scale identification of putative substrates for other plant proteases (Demir *et al.*, 2017; Perrar *et al.*, 2019). A notable example was the use of N-terminomic mass-spectrometry based method COFRADIC to identify substrates of the Arabidopsis metacaspase AtMC9 in seedlings (Tsiatsiani *et al.*, 2013). This revealed a large number of putative candidates and resulted in the confirmation of MC9-dependent processing of PEPCK1. Other N-terminomic methods have also been successfully applied to investigation of chloroplast proteolysis and for the identification of substrates of ICP55 and OCT1 (Carrie *et al.*, 2015; Köhler *et al.*, 2015; Rowland *et al.*, 2015).

The N-terminomic method HUNTER has recently been applied to Arabidopsis quadruple *vpe* null mutant seedlings in a substantial step forward for VPE substrate identification. The technique confirmed many previous insights of seed storage protein processing by VPEs, and additionally identified multiple novel indirect or putative direct processing events (Weng *et al.*, 2019). At the time of writing the first N-terminome study of AtVPE$_\gamma$ processing in *A. thaliana* leaves was reported, although the putative substrates remain to be explored (Dall *et al.*, 2020).

Despite these advances, there are still large gaps in our knowledge of the action of VPEs. For example, *N. benthamiana* is frequently used for heterologous overexpression and VPE activity assays, yet no endogenous substrates of VPEs are known (Conlan *et al.*, 2012; Jackson *et al.*, 2018; Poon *et al.*, 2018). Multiple VPEs exist in the *N. benthamiana* genome

and activity is detected across vegetative and reproductive tissues (**Chapter 3**). Furthermore, previous studies demonstrated that vegetative VPEs NbVPE1a and NbVPE1b are essential for PCD in select pathogen defences, indicating that VPEs do perform important roles in leaves (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010). In this chapter we aim to identify VPE substrates in *N. benthamiana* leaves. Using the popular N-terminomics technique HUNTER, we have identified potential direct substrates of VPEs and initiated confirmation studies.

Evidence for processing	Species	Organ	VPE	Substrate Name	Substrate Type	Native or Non-native Substrate	*In vitro*	*In vivo*	Co-localisation	Cleavage Site	Confirmed cleavage site	Reference
Insufficient	*Abrus precatorius*	Seed	N/A	Abrin-a	Storage	Native	No	No	Predicted	NPPN|ANQS	No	(Funatsu *et al.*, 1988)
Insufficient	*Acacia confusa*	Seed	N/A	Kunitz Inhibitor	Inhibitor	Native	No	No	Predicted	CEGN|SDDE	Yes	(Wu and Lin, 1993)
Insufficient	*Actinidia chinensis*	Fruit	N/A	Actinidin protease	Protease	Native	No	No	Predicted	KYNN|QNYP	No	(Praekelt *et al.*, 1988)
Strong	*Arabidopsis thaliana*	Seed	AtVPEβ	2S Albumin 1,2,3,4,5	Storage	Native	No	Yes	Predicted	DATN|PIGP DMEN|PQGQ DASN|PMGP DIEN|PQGQ DASN|PVGP DASN|PIGP DDEN|PMGP DVSN|PQQG DIEN|PQRR	Yes	(Shimada *et al.*, 2003)
Strong	*Arabidopsis thaliana*	Seed	AtVPEβ	12S globulin 1,2,3	Storage	Native	No	Yes	Predicted	RHGN|GLEE EIAN|GLEE PGGN|GLEE	Yes	(Shimada *et al.*, 2003)
Strong	*Arabidopsis thaliana*	Leaf	AtVPEγ	AtCPY	Protease	Native	No	Yes	Yes	Unknown	No	(Rojo *et al.*, 2003)
Insufficient	*Arabidopsis thaliana*	Leaf senescent	AtVPEγ	AtFruct4	Invertase	Native	No	Yes	Yes	Unknown	No	(Rojo *et al.*, 2003)
Strong	*Arabidopsis thaliana*	Vasculature	AtVPEγ	AtCEP-1	Protease	Native	Yes	Yes	Yes	Unknown	No	(Cheng *et al.*, 2019)
Insufficient	*Arabidopsis thaliana*	Anther	AtVPEγ	RD19A	Protease	Native	No	Yes	Yes	Unknown	No	(Cheng *et al.*, 2020)
Insufficient	*Arabidopsis thaliana*	Anther	AtVPEγ	RD19C	Protease	Native	No	Yes	Yes	Unknown	No	(Cheng *et al.*, 2020)
Strong	*Arabidopsis thaliana*	Anther	AtVPEβ	AtCEP1	Protease	Native	Yes	Yes	Yes	Unknown	No	(Cheng *et al.*, 2020)

Insufficient	*Arabidopsis thaliana*	Seedling	AtVPEβ	12S albumin (cruciferins CRA1, CRB, CRC)	Storage	Native	No	Yes	Predicted	RHGN|GLEE EIAN|GLEE PQGN|GLEE	Yes	(Weng *et al.*, 2019)
Insufficient	*Arabidopsis thaliana*	Seedling	AtVPEβ	2S albumins (SSP 2,3,4)	Storage	Native	No	Yes	Predicted	PSLD|DEFD	Yes	(Weng *et al.*, 2019)
Non-native	*Arabidopsis thaliana*	Seed	AtVPEβ	SFTI	Cyclotide - inhibitor	Non-native	Yes	No	Predicted	GLDN|PRGC IEDN|GRCT	Yes	(Mylne *et al.*, 2011)
Non-native	*Arabidopsis thaliana*	Seed	AtVPEβ	SFTI	Cyclotide - inhibitor	Non-native	Yes	No	Predicted	GLDN|PRGC IEDN|GRCT	Yes	(Bernath-Levin *et al.*, 2015)
Insufficient	*Avena sativa*	Seed	N/A	12S globulin	Storage	Native	No	No	Predicted	QSFN|GLEE	No	(Shotwell *et al.*, 1988)
Insufficient	*Bertholletia excelsa*	Seed	N/A	2S Albumin	Storage	Native	No	No	Predicted	EEEN|QEEC	No	(Ampe *et al.*, 1986)
Insufficient	*Brassica napus*	Seed	N/A	2S Albumin	Storage	Native	No	No	Predicted	DMEN|PQGP DATN|SAGP	No	(Monsalve *et al.*, 1991)
Strong	*Canavalia ensiformis*	Seed	CeAEP1	Concavalin A	Cycoltide - storage	Native	Yes	No	Predicted	PDAN|VIRN	Yes	(Min and Jones, 1994)
Non-native	*Canavalia ensiformis*	Seed	CeAEP1	SFTI	Cyclotide - inhibitor	Non-native	Yes	No	Not shown	GLDN|PRGC IEDN|GRCT	Yes	(Bernath-Levin *et al.*, 2015)
Strong	*Clitoria ternatea*	Leaf	CtAEP1	Cter M precursor	Cyclotide - inhibitor	Native	No	Yes*	Predicted	CMKN|HIIA	Yes	(Poon *et al.*, 2018)
Non-native	*Clitoria ternatea*	Leaf	butelase-1	SFTI	Cyclotide - inhibitor	Non-native	Yes	No	Not shown	CFPN|HV	Yes	(Nguyen *et al.*, 2014)
Moderate	*Clitoria ternatea*	Leaf	butelase-1	kb1	Cyclotide - inhibitor	Native	Yes	No	Predicted	CTRN|HVIA	Yes	(Nguyen *et al.*, 2014)
Insufficient	*Cucurbita sp. cv Kurokawa Amakuri Nankin (Curcurbita hybrid cultivar)*	Cotyledon	N/A	Proglobulin	Storage	Native	No	No	Yes	Unknown	Yes	(Hara-Nishimura and Nishimura, 1987)

Moderate	*Cucurbita sp., cv Kurokawa Amakuri Nankin*	Cotyledon	RcVPE	Proglobulin	Storage	Native	Yes	No	Predicted	ESEN|GLEE	Yes	(Hara-Nishimura *et al.*, 1991)
Non-native	*Curcurbita maxima*	Seed	RcVPE	PV100	Storage	Non-native	Yes	No	Yes	CGVN|QRHS RGED|VD|EVER DDEN|QRDP GGRD|DEDEN|Q RDP SRVN|QVAI	Yes	(Yamada *et al.*, 1999)
Insufficient	*Curcurbita maxima*	Seed	N/A	Membrane proteins 27-32	Storage	Native	No	No	Yes	GNEN|RDKT	No	(Inoue *et al.*, 1995)
Insufficient	*Curcurbita maxima*	Seed	N/A	11S globulin	Storage	Native	No	No	Predicted	ESEN|GLEE	No	(Hayashi *et al.*, 1988)
Insufficient	*Glycine max*	Seed	N/A	Protease P34	Protease	Native	No	No	Yes	KMAN|KKMK	Yes	(Kalinski *et al.*, 1992)
Insufficient	*Glycine max*	Seed	N/A	Glycinin A5A4B3	Storage	Native	No	No	Predicted	ETRN|GVEE EQSN|RRGS	No	(Hirano *et al.*, 1985)
Insufficient	*Glycine max*	Seed	N/A	Glycinin A3B4	Storage	Native	No	No	Predicted	QTRN|GVEE	No	(Hirano *et al.*, 1984)
Insufficient	*Glycine max*	Seed	N/A	Glycinin A A2B1a	Storage	Native	No	No	Predicted	RSRN|GIDE	No	(Staswick *et al.*, 1984)
Insufficient	*Glycine max*	Seed	N/A	Glycinin A1aBx	Storage	Native	No	No	Predicted	SRRN|GIDE	No	(Negoro *et al.*, 1985)
Insufficient	*Helianthus annuus*	Seed	N/A	11S globulin HaG3	Storage	Native	No	No	Predicted	GWSN|GVEE	No	(Yonder Haar *et al.*, 1988)
Insufficient	*Helianthus annuus*	Seed	N/A	2s albumin HaG5	Storage	Native	No	No	Predicted	TESN|IDIP	Yes	(Allen *et al.*, 1987)
Strong	*Helianthus annuus*	Seed	HaAEP1	PawS1 (SFT-1)	Cyclotide - inhibitor	Native	Yes	No	Predicted	GLDN|PRGC IEDN|GRCT	Yes	(Franke *et al.*, 2017)
Strong	*Hellianthus annuus*	Seed	HaAEP1	SFTI	Cyclotide - inhibitor	Native	Yes	No	Predicted	GLDN|PRGC IEDN|GRCT	Yes	(Bernath-Levin *et al.*, 2015)
Strong	*Momordica cochinchinensis*	Seed	MCoAEP1 MCoAEP2	TIPTOP (MCoTI-II)	Cyclotide - inhibitor	Native	Yes	No	Predicted	DIN|GGVC SGSD|ALEG	Yes	(Du *et al.*, 2020)

Non-native	*Momordica cochinchinensis*	Seed	MCoAEP1 MCoAEP3	Oak1 (kB1-NAL)	Cyclotide - inhibitor	Non-native	Yes	No	Not shown	CTRN|GLPS	Yes	(Du *et al.*, 2020)
Insufficient	*Nicotiana alata*	Leaf Stigma Stylet	N/A	Proteinase inhibitor II	Inhibitor	Native	No	No	Predicted	EKKN|DRIC	Yes	(Atkinson *et al.*, 1993)
Insufficient	*Nicotiana silvestris*	Leaf	N/A	Endochitinase B	Defence	Native	No	No	Yes	Unknown	No	(Neuhaus *et al.*, 1991)
Insufficient	*Nicotiana tabaccum*	Leaf	N/A	Chitinase	Defence	Native	No	No	Predicted	SFGN|GLLV	Yes	(Sticher *et al.*, 1993)
Insufficient	*Nicotiana tabacum*	Leaf	N/A	CBP20 (PR4)	Defence	Native	No	No	Predicted	CGDN|MNVL	Yes	(Ponstein *et al.*, 1994)
Strong	*Oldenlandia affinis*	Leaf	OaAEP1b, OaAEP3	Kalata B1	Cyclotide - inhibitor	Native	No	Yes*	Predicted	CTRN|GLPS	Yes	(Poon *et al.*, 2018)
Strong	*Oldenlandia affinis*	Leaf	OaAEP1b, OaAEP3	Kalata B3/6	Cyclotide - inhibitor	Native	No	Yes*	Predicted	TRRD|GLPK CTRD|GLPS	Yes	(Poon *et al.*, 2018)
Strong	*Oldenlandia affinis*	Leaf	OaAEP1b, OaAEP3	Kalata B2 (Oak2)	Cyclotide - inhibitor	Native	No	Yes*	Predicted	CTRD|SLPM	Yes	(Poon *et al.*, 2018)
Strong	*Oldenlandia affinis*	Seed	OaAEP1b	Oak1 (kb1)	Cyclotide - inhibitor	Native	Yes	No	Predicted	CTRN|GPLS	Yes	(Harris *et al.*, 2015)
Non-native	*Oldenlandia affinis*	Seed	OaAEP1b	R1	anti-malarial peptide	Non-native	Yes	No	Not shown	CTRN|GPLS	Yes	(Harris *et al.*, 2015)
Insufficient	*Oryza sativa*	Seed	N/A	lectin	Storage	Native	No	No	Predicted	RNAN|GELC	Yes	(Wilkins and Raikhel, 1989)
Insufficient	*Phaseolus vulgaris*	Seed	N/A	alpha amylase inhibitor	Inhibitor	Native	No	No	Yes	RQAN|SAVGL	Yes	(Santino *et al.*, 1992)
Insufficient	*Pisum sativum*	Seed	N/A	legumin	Storage	Native	No	No	Predicted	QGDN|GLEE	No	(Lycett *et al.*, 1984)
Insufficient	*Pisum sativum*	Seed	N/A	legumin	Storage	Native	No	No	Predicted	ERKN|GLEE	No	(Gatehouse *et al.*, 1982)
Insufficient	*Pisum sativum*	Seed	N/A	lectin	Storage	Native	No	No	Predicted	EEEN|VTSY	No	(Gatehouse *et al.*, 1987)
Insufficient	*Pisum staivum*	Seed	N/A	vicilin	Storage	Native	No	No	Predicted	GKEN|DKEE	Yes	(Gatehouse *et al.*, 1982)

Insufficient	*Ricinus communis*	Seed	N/A	agglutinin	Storage	Native	No	No	Predicted	PNFN|ADVC	No	(Roberts, 1985)
Insufficient	*Ricinus communis*	Seed	N/A	Ricin	Defence	Native	No	No	Predicted	PNFN|ADVC	No	(Araki and Funatsu, 1987)
Non-native	*Ricinus communis*	Seed	RcAEP1	SFTI	Cyclotide - inhibitor	Non-native	Yes	No	Not shown	GLDN|PRGC IEDN|GRCT	Yes	(Bernath-Levin *et al.*, 2015)
Insufficient	*Ricinus communis*	Seed	N/A	2s albumin	Storage	Native	No	No	Predicted	TRTN|PSQQ RSDN|QERS	Yes	(Sharief and Li, 1982)
Non-native	*Sarrcharomyces cerevisiae*	N/A	AtVPE	CPY	Protease	Non-native	No	Yes**	Yes	Unknown	No	(Kinoshita *et al.*, 1999)
Insufficient	*Solanum lycopersicum*	Leaf	N/A	Proteinase inhibitor I	Inhibitor	Native	No	No	Predicted	FDSN|LMCE	Yes	(Graham *et al.*, 1985)
Moderate	*Vinga mungo*	Cotyledon	VmPE	SH-EP	Protease	Native	Yes	No	Predicted	SKVN|HHHK	Yes	(Okamoto and Minamikawa, 1999)
Non-native	*Viola canadensis*	N/A	VcAEP	GN14-SL	Cyclotide	Non-native	Yes	No	Not shown	SYRN|SL(AN)	Yes	(Hemu *et al.*, 2019)
Non-native	*Viola yedoensis*	N/A	VyPAL1-3, VyAEP1	GN14-SL	Cyclotide	Non-native	Yes	No	Not shown	SYRN|SL(AN)	Yes	(Hemu *et al.*, 2019)

Table. 1 Previously reported VPE substrates. Assessment of the strength of evidence supporting a substrate of VPEs. If the substrate is detected in the vacuole it is considered to be co-localised with VPEs. Substrate may have been shown to be in other papers in the vacuole and not recognised here. *In vivo* evidence encompasses both experiments in native systems or in model eukaryotes: **N. benthamiana* transient expression or ** expression in yeast.

4.3 Results

To investigate substrates of NbVPEs *in planta*, we generated a genetic toolbox for VPE studies. This included the CRISPSR-Cas9 generated *N. benthamiana* mutant *vpe1a/b,* which lacks NbVPE1a and NbVPE1b (**Chapter 3**). Labelling of *N. benthamiana* leaf extracts using the VPE-specific probe DK004 (**Chapter 3, Fig.4**) demonstrated negligible activity in the *vpe1a/b* mutant (**Fig. 1A**). This loss of activity in *vpe1a/b* leaves could be complemented by transient overexpression of a C-terminally His tagged NbVPE1b-His, but not with the inactive NbVPE1a^{C215S}-His mutant (**Fig. 1A**). Corresponding NbVPE1a/b WT and mutant constructs failed to assemble correctly during cloning. NbVPE1a and NbVPE1b share 94.9% amino acid identity (pairwise % positive (BLSM62) = 97.3%) (**Fig 1B**). Overlaying NbVPE1a and NbVPE1b structures generated by homology modelling on the AtVPE$_\gamma$ structure (5NIJ) (Zauner *et al.*, 2018) demonstrated variant residues occupy surface positions outside of the immediate active site area (**Fig. 1C**). Therefore NbVPE1a^{C215S} was deemed a sufficient control for NbVPE1b in future experiments.

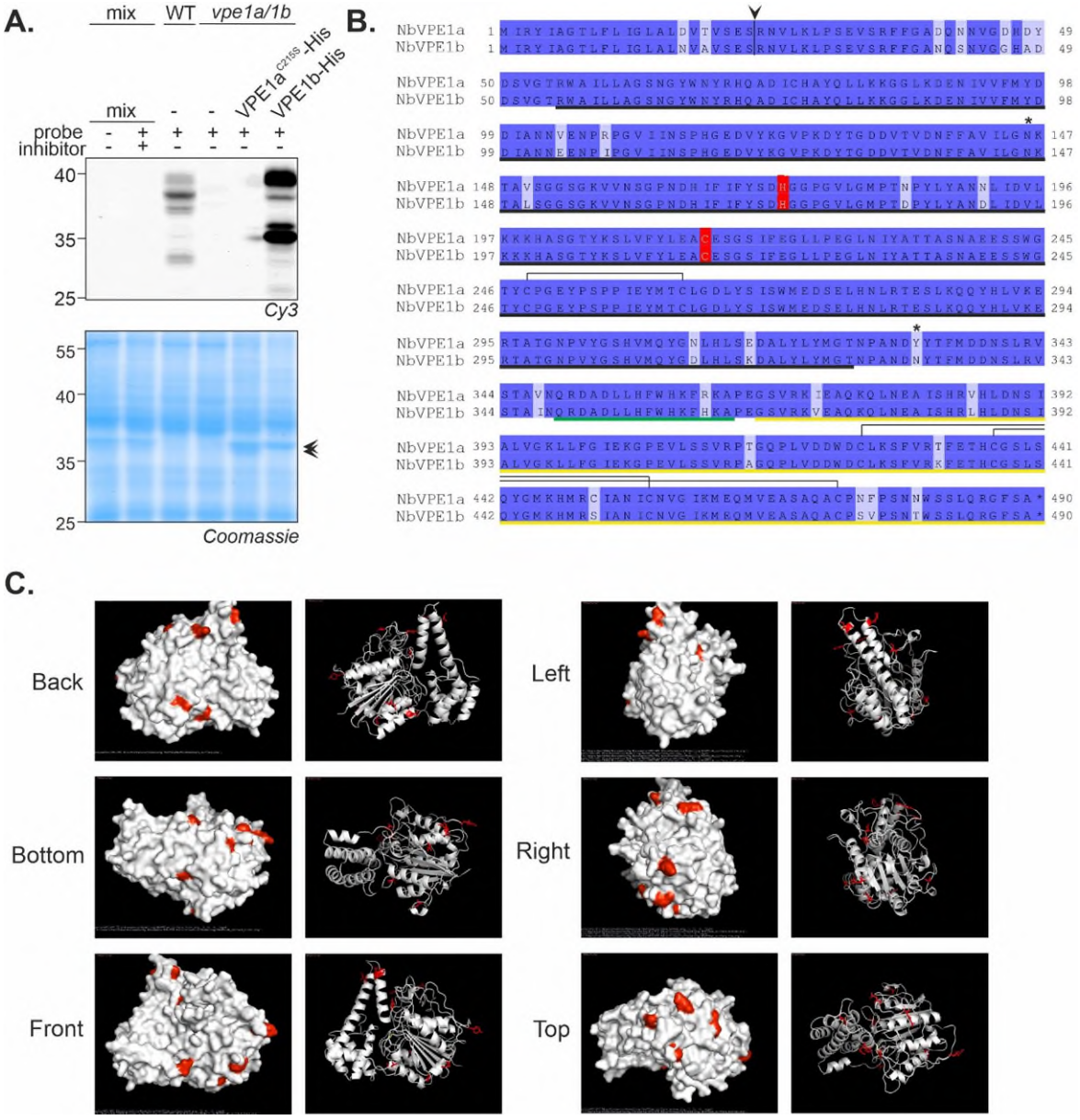


Figure 1. NbVPE1a and NbVPE1b are major vegetative VPEs. A. Activity-based labelling of NbVPEs with the DK004 probe. Protein was extracted from mature WT and *vpe* mutant plants, or from *vpe* mutants agroinfiltrated with either NbVPE1b-His or the catalytic inactive NbVPE1a^{C215S}-His mutant and harvested at 4 dpi. Extract was preincubated with either 100 µM YVAD-cho or DMSO and then 0.25 µM DK004 for 3 hours. Labelled proteins were visualised by in gel fluorescence scanning. **B.** NbVPE1a and NbVPE1b are highly similar and share 94.9% pairwise identity (97.1% BLOSUM62 identity). Amino acid alignment generated using ClustalW and visualised in JalView. Catalytic his and cys residues highlight in red. Bars underlining the sequence mark the following domains: black = catalytic, green = activation peptide, yellow = LSAM. Putative N-glycosylation sites marked with *. Signal peptide cleavage site delineated by downwards arrow. Residues coloured based on % identity. **C.** Overlay of NbVPE1a and NbVPE1b structures generated by SwissModel EXPASY homology modelling based on Arabidopsis

VPE$_\gamma$ (5NIJ) (Waterhouse *et al.*, 2018; Zauner *et al.*, 2018). Variant residues are highlighted in red. Catalytic residue C215 highlighted in yellow.

The capacity of these tools to investigate substrates of VPEs was tested using a previously described substrate of Arabidopsis VPEs in leaves, serine carboxypeptidase AtCPY (Rojo *et al.*, 2003). AtCPY and the top corresponding BLAST hit in *N. benthamiana*, NbCPY, were cloned with a fused C-terminal GFP tag for *in planta* transient overexpression by agroinfiltration. Due to its catalytic serine residue, active AtCPY can be detected using the serine hydrolase activity-based probe FP (Liu *et al.*, 1999; Kaschani *et al.*, 2009). Overexpression of AtCPY-GFP in leaves resulted in three additional fluorescent signals (arrows 1, 4 and 5) in the FP profile compared to the empty vector control (**Fig. 2A**). Previous reports showed VPE-dependent processing of AtCPY resulted in conversion of a 48 kDa protein into a 24 kDa proteoform, as detected by an AtCPY C-terminal specific antibody (Rojo *et al.*, 2003). AtCPY-GFP overexpressed in the *vpe1a/b* mutant accumulated as a ~48 kDa labelled protein, indicating that the GFP tag is removed *in planta* and the full length proteoform is active (**Fig. 2A, arrow 1**). In contrast, AtCPY-GFP overexpressed in WT plants caused a reduced ~48 kDa signal and two additional ~18 kDa fluorescent signals (**Fig. 2A, arrows 4,5**). Detected proteins must contain the labelled catalytic serine, and therefore generation of these 18 kDa proteoforms requires at least two cleavage sites to be processed. Similarly to WT, *vpe1a/b* expressing NbVPE1b-His displayed both ~48 kDa and 18 kDa signals upon AtCPY-GFP overexpression. Co-expression of AtCPY-GFP and NbVPE1a^{C215S}-His in the *vpe1a/b* mutant resulted in a prominent ~48 kDa signal and reduced 18 kDa signals (**Fig. 1C**). In contrast to AtCPY-GFP, transient

overexpression of NbCPY-GFP generated a single additional signal at ~50 kDa independent of the genetic background, at the expected molecular weight if the GFP tag is removed *in planta* (**Fig. 2A**). This indicates NbCPY does not undergo VPE-dependent processing, in contrast to AtCPY.

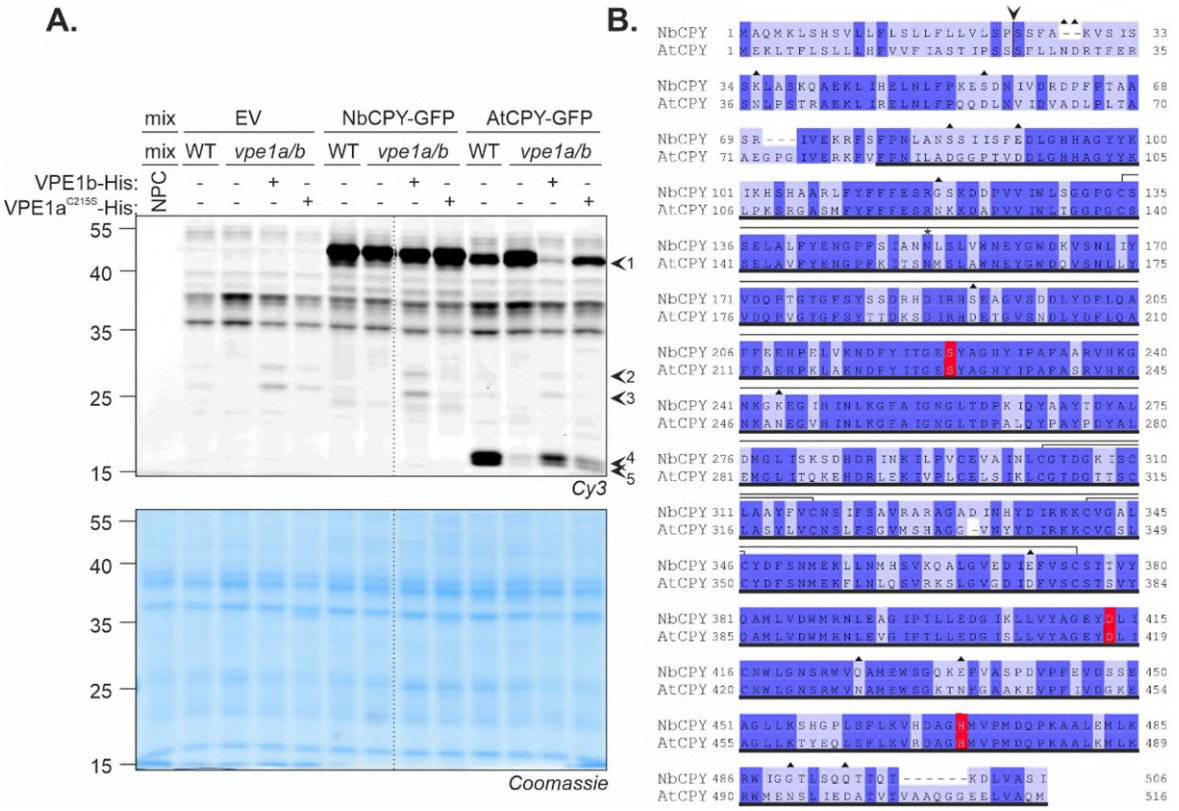


Figure 2. AtCPY processing is dependent on NbVPEs. A. Known VPE substrate AtCPY and the closest homolog in *N. benthamiana* NbCPY were transiently expressed by agroinfiltration. Samples were harvested at 4 dpi and protein extracts labelled with 0.25 μM serine hydrolase probe FP for 1 hour. Labelled proteins were visualised by in-gel fluorescence scanning. Arrows 1, 4 and 5 indicate differential CPY dependent signals. Arrows 2 and 3 mark unknown differential serine hydrolase signals. **B.** Amino acid alignment of AtCPY and NbCPY. Amino acid alignment generated using ClustalW and visualised in JalView. Catalytic his and cys residues highlight in red. Black bar underneath sequence maps catalytic domain. Putative N-glycosylation sites marked with *. Signal peptide cleavage site delineated by downwards arrow. Residues coloured based on % identity. Disulphide bridges marked by lines above the sequence. Triangle marks variant N/D residues present in AtCPY only. Annotations are made based on AtCPY uniprot entry P32826.

The sequences of AtCPY and NbCPY are relatively well conserved, but crucially AtCPY contains 14 N or D residues (representing putative VPE processing sites) that are not present in NbCPY (**Fig. 2B**). Disregarding the variant residues that are close to the N or C terminus, or are adjacent to D or Ns in the NbCPY sequence, there are seven differential VPE susceptible residues. Mutants with a combination of these residues substituted will be useful in dissecting the precise cleavage sites.

Interestingly, independent of CPY expression, signals 2 and 3 of the serine hydrolase labelling profile were consistently more prominent in the presence of VPE, suggesting there are VPE-dependent changes in serine hydrolase activation (**Fig. 2A, arrows 2,3**). It is unclear which serine hydrolases generate these signals.

Given that the *vpe1a/b* mutant and transient NbVPE expression can be used to identify VPE-dependent AtCPY cleavage, we proceeded to screen for novel NbVPE1b substrates. High- efficiency Undecanal-based N Termini EnRichment (HUNTER) was used to identify differentially abundant protein N-termini in the absence and presence of NbVPE activity (**Fig. 3**) (Weng *et al.*, 2019). Briefly, this technique involves dimethyl labelling of primary amines present at exposed protein N-termini and lysine side chains. Proteins are digested with trypsin to generate peptides and the mixture is then filtered to retain peptides with dimethylated or endogenously blocked N-termini. These N-terminal peptides are separated by liquid chromatography and detected using tandem mass-spectrometry (LC-MS/MS).

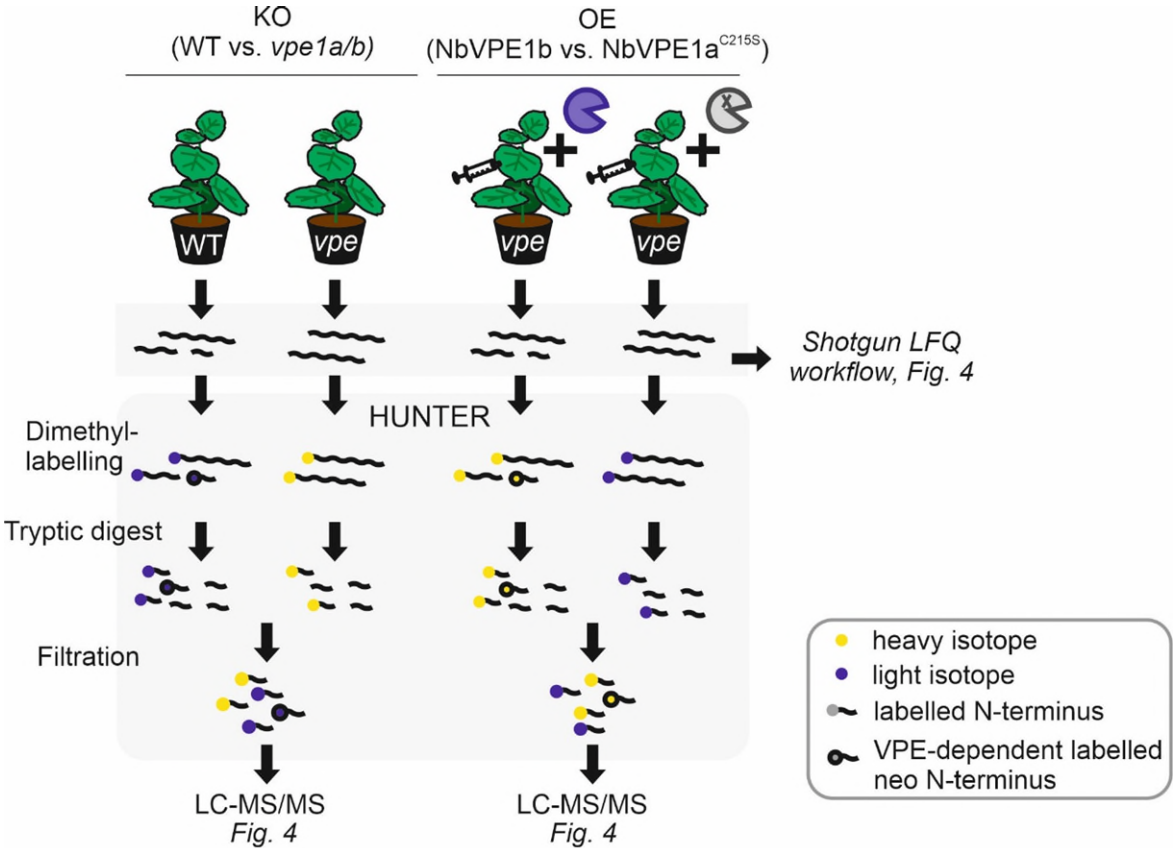


Figure 3. Experimental design for NbVPE substrate discovery by HUNTER. Differential N-terminal peptides were identified by HUNTER in two separate experiment comparisons. 1. WT and *vpe1a/b* mutant plants (KO) and 2. *vpe1a/b* mutant transiently overexpressing NbVPE1b-His or catalytic inactive NbVPE1a^{C215}-His (OE) by agroinfiltration. Leaves of 4-5 week old plants were harvested at 4 days post infiltration if agroinfiltrated. Protein extracts were subject to heavy or light isotope dimethyl labelling of primary amines (present at exposed N-termini of proteins or on lysine residue side chains). Protein extracts were then digested with trypsin to generate many peptides. Peptides with a blocked N-terminus were negatively selected for by filtration through a hydrophobic polymer. Enriched N-terminal blocked peptides were then detected by LC-MS/MS.

If a protein is a substrate of VPE, the neo-N-terminus would be more abundant in the samples with VPE activity. However, N-termini could also differ in abundance if the total protein level is different between the VPE present and absent conditions. Therefore, to control for protein abundance we first performed shotgun proteomics label free quantification (LFQ) to determine total protein levels (**Fig. 3**, **Fig. 4**). Surprisingly, despite

the drastic difference in VPE activity in the overexpression comparison, none of the 2566 robustly detected protein groups differed significantly in abundance, with the exception of the two overexpressed NbVPE proteins. Likewise, few differences in protein levels were observed in the KO comparison. Of the 2280 protein groups detected, only two proteins (CURT1c, a transmembrane protein and a dihydrolipopyl dehydrogenase) differed in abundance (**Fig. 4**). These two proteins are predicted to be present in thylakoids and mitochondria respectively, and are therefore unlikely to represent direct substrates of VPEs. The lack of large-scale difference in protein abundance indicates any N-terminal peptide abundance differences detected are more likely to be due to changes in processing activity.

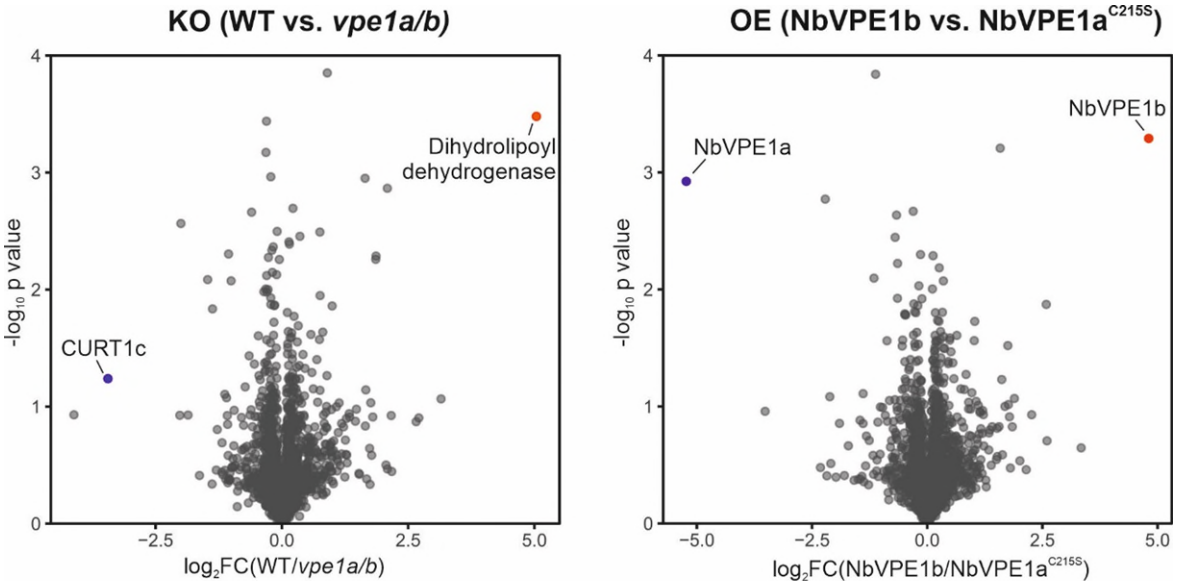


Figure 4. VPE activity does not affect protein levels. Volcano plots of protein groups identified in leaves of WT and *vpe1a/b* mutant, or *vpe1a/b* mutant transiently overexpressing NbVPE1b-His or NbVPE1a^{C215S}-His as described in **Fig. 2**. Samples were digested by trypsin and protein levels measured by label free quantification LC-MS/MS. Highlighted proteins displayed significant log_2FC with FDR = 0.01, n = 3 biological replicates.

In contrast to total protein levels, more differences were observed in N-termini detected between proteomes in the presence and absence of NbVPEs. HUNTER was performed on two comparisons: WT and *vpe1a/b* mutant leaves (KO), and NbVPE1b-His and NbVPE1a^{C215S}-His transiently overexpressed in the mutant background (OE) (**Fig. 3**). For KO comparison, 1386 N-terminal peptides were identified in at least two of three biological replicates (**Fig. 5A**). The OE comparison contained a similar 1190 peptides identified in two of three biological replicates (**Fig. 5B**). In the presence of active VPE, 90 N-termini were significantly increased in the KO comparison (**Fig. 5A**) and 37 in the OE comparison (**Fig. 5B**). 33 N-terminal peptides were significantly decreased in the KO comparison and 93 were less abundant with the OE comparison (**Fig. 5**).

Despite identifying more increasing N-termini in the KO comparison, the majority of these peptides did not represent direct substrates due the lack of a P1 N or D residue. A cleavage logo generated from the significantly increased neo-N-terminal peptides in WT v. *vpe1a/b* comparison revealed no preference for either N or D at P1, indicating that the majority of these peptides were not directly generated by VPE. However, significant differences in the cleavage logo were detected, indicating that other protease activities are dependent on VPEs (**Fig. 5C**). In contrast, the cleavage logo generated with significantly increased neo-N-termini detected in the OE comparison did reveal a strong preference for N at P1, as expected for VPE activity (**Fig. 5D**).

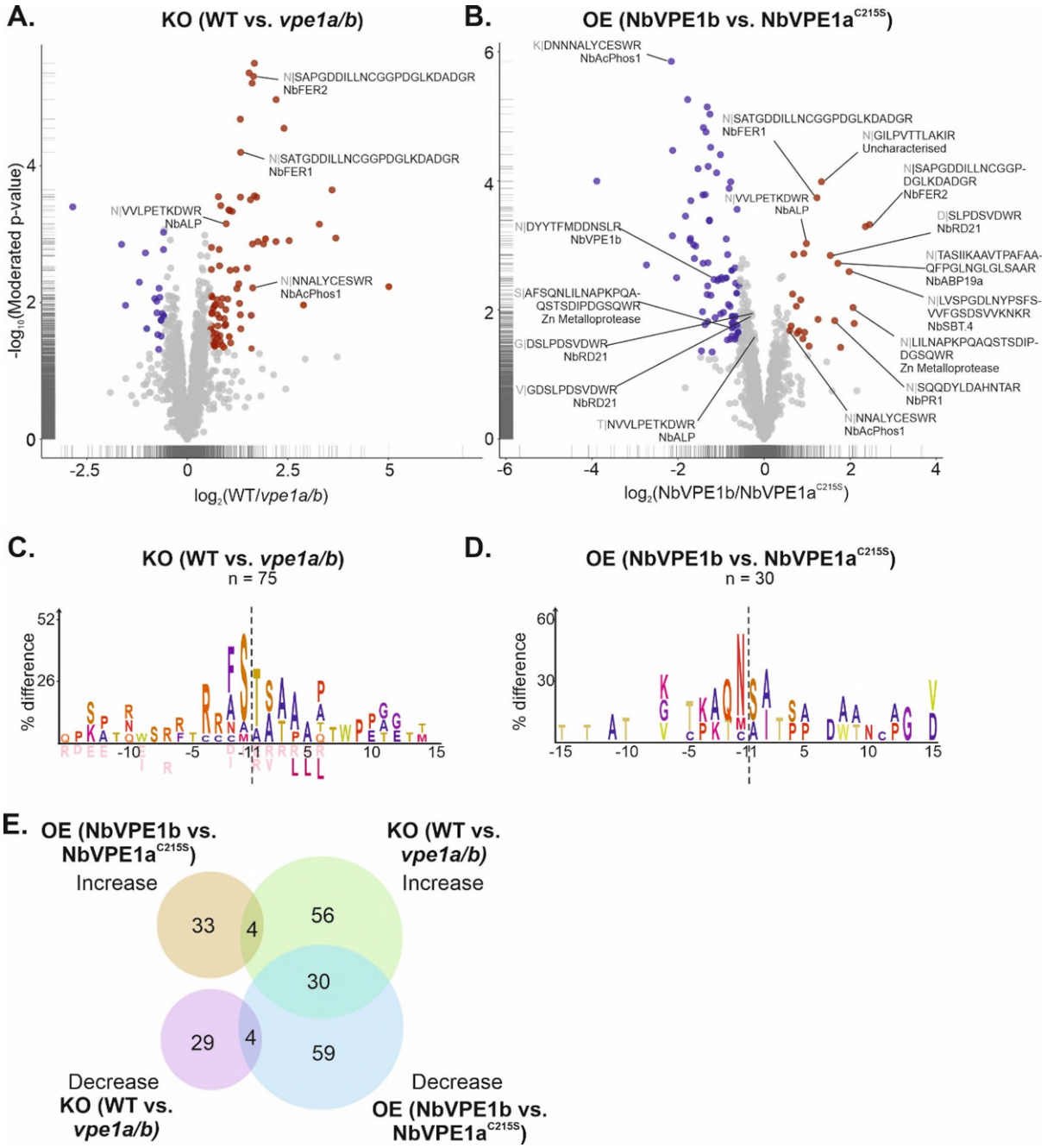


Figure 5. Identification of the N-terminome of *N. benthamiana* leaves to screen for putative NbVPE substrates. HUNTER was performed on extracts from mature WT and mutant leaves, or leaves transiently overexpressing NbVPE1b-His or NbVPE1a^{C215S}-His (see **Fig.2**). **A.** and **B.** Volcano plots of N-terminal dimethylated peptides identified by HUNTER. Peptides are coloured if $\log_2$FC $\pm >$ 0.58 with $p < 0.05$. Peptides from promising candidate substrates that were investigated further in this chapter are labelled. **C.** and **D.** Cleavage site logo for the active NbVPE positive condition generated from peptides with $\log_2$FC $\pm >$ 0.58, $p < 0.05$, start position >14. Logo was generated with iceLogo tool (Colaert *et al.*, 2009) using *N. benthamiana* Swiss-prot composition reference set, p value < 0.05. **E.** Overlap in peptides between two experiments, including peptides with $\log_2$FC $\pm >$ 0.5, p value < 0.05, peptide start position > 2.

Following on from the different cleavage logos detected in the KO and OE comparisons, the N-terminal peptides identified to significantly increase or decrease in abundance displayed very little overlap between the two experiments. Only four peptides were identified in increase and four to decrease in both experiments (**Fig. 5E**). Surprisingly the greatest overlap was observed between peptides increasing in KO that decreasing in OE comparison, where no overlap was expected to occur (**Fig. 5E**). It is unclear what underlies this unexpected phenomenon, but may potentially reflect bias in detection of different isotopic labels as heavy isotopes were used for both WT samples and NbVPE1a^{C215S}-His overexpression samples.

The combination of activity-based labelling and the HUNTER cleavage logo indicated that the OE comparison had greater differential in VPE activity and revealed more candidate substrates. Candidate substrates were selected primarily based upon the magnitude of the log_2 fold change (FC) observed in the OE comparison and the presence of a N or D at P1. Subsequently, the log_2FC and significance measured in the KO comparison was considered. Finally, the putative co-localisation with VPEs in the vacuole or apoplast was used to filter candidates to a reduced list of ten candidate NbVPE1b substrates for further investigation (**Table 2**).

To investigate if these candidate substrates are direct substrates of NbVPE1b, we cloned these with C-terminal epitope tags for transient overexpression in *N. benthamiana* and recombinant expression in *E. coli* (**Table 4**). Simultaneously, we generated uncleavable mutant constructs with the putative P1 residue substituted by alanine. However, due to time limitations not all plasmid constructs were generated or tested.

Peptide	Name	Leading MQ razor protein	NbD ID	Start Pos	P1	P1'	WT/*vpe1a/b*			NbVPE1b/NbVPE1a^{C215S}			Significant Change in Shotgun?	Signal peptide
							log_2FC	p value	adj. p value	log_2FC	p value	adj. p value		
SAPGDDILLNCGGP-DGLKDADGR	NbFER2	Niben101Scf04123_264058-268721	NbD027854.1	30	N	S	1.648	0.000	0.004	2.450	0.000	0.032	No	Yes
LILNAPKPQAQSTSD-IPDGSQWR	ATP-dependent zinc metalloprotease FTSH	Niben044Scf00026134Ctg000_22457-27974	NbD016533.1 NbD006996.1	113	N	L	0.089	0.697	0.831	2.066	0.009	0.204	No	No
LVSPGDLNYPSFSV-VFGSDSVVKNKR	Subtilase (putative NbSBT1.4)	Niben101Scf03998_52804-57134	NbD027079.1	658	N	L	ND	ND	ND	1.978	0.003	0.092	No	Yes
TASIIKAAVTPAFAA-QFPGLNGLGLSAAR	NbABP19a	Niben101Scf02749_419924-422547	NbD020273.1 NbD004644.1	60	N	T	0.425	0.283	0.565	1.713	0.002	0.078	No	Yes
SQQDYLDAHNTAR	NbPR1	Niben101Scf13926_197831-205737	NbD001203.1 NbD023903.1 NbD049561.1	33	N	S	ND	ND	ND	1.641	0.015	0.248	No	Yes
SLPDSVDWR	NbRD21	Niben101Scf04007_125083-130446 Niben101Scf09885_71970-76814	NbD027122.1 NbD045163.1 NbD045163.2	143	D	S	0.201	0.181	0.501	1.538	0.001	0.065	No	Yes
GILPVTTLAKIR	Uncharacterised	Niben101Scf02261_101899-104498	NbD017081.1	153	N	G	ND	ND	ND	1.334	0.000	0.014	ND	Yes
SATGDDILLNCGGP-DGLKDADGR	NbFER1	Niben101Scf07619_25958-30621	NbD040156.1	30	N	S	1.328	0.000	0.016	1.229	0.000	0.020	No	Yes
VVLPETKDWR	NbALP	Niben101Scf01445_48569-53417	NbD011552.1	141	N	V	0.966	0.001	0.070	0.694	0.001	0.065	No	Yes
NALYCESWR (NNALYCESWR)	Acid phosphatase 1	Niben101Scf14274_119630-135356	NbD023512.1 NbD049751.1	50/51	N	N	1.627	0.006	0.214	0.597	0.021	0.285	ND	Yes

Table 2. Putative NbVPE substrates identified by HUNTER. Leading MQ razor protein identifiers originate from an intermediate NbC proteome (Kourelis *et al.*, 2018) used for mass spectrometry searches. Corresponding NbD identifiers includes all proteins to which the peptide matches (Kourelis *et al.*, 2019). Start pos = amino acid start position of peptide. Signal peptides predicted in signalP (Armenteros *et al.*, 2019). Log_2FC values highlighted if statistically significant ($p < 0.05$). ND = protein not detected in shotgun experiment.

4.3.1 NbFERONIA receptor is a putative NbVPE substrate

The neo-N-terminus with the greatest log_2FC was that of FERONIA (NbFER2, NbD027854.1), a malectin receptor-like protein kinase in both the KO and the OE comparisons (**Fig. 5, Table 2**). This cleavage event was also detected in NbFER1 (NbD040156.1), which shares 98.9% pairwise identity with NbFER2 (**Fig. 6A**). The robust detection of this processing event increased our confidence of NbFER as a NbVPE candidate substrate, despite the atypical cleavage site, just three residues after the predicted signal peptide cleavage site. The closest FERONIA orthologs from a range of Solanaceous and other plant species indicate N or D residues are conserved at this position, even though this region is poorly conserved part of the protein. Interestingly, another secreted putative substrate NbPR1 had a neo-N-terminus at a similar sequence and position, suggesting this N-terminal might be a common VPE-dependent activity (**SFig. 1C**).

To determine if NbFER1/2 could be direct substrates of NbVPE, we generated synthetic fluorescent peptides MCA-AQNSAP-Dnp and MCA-AQASAP-Dnp, containing an N or A at the P1 position respectively. MCA (7-methoxycoumarin-4-yl) fluorescence is quenched in close proximity to Dnp (2,4-dinitrophenyl). Cleavage anywhere within the amino acid sequence results in fluorescence release. Activity against the cleavable and uncleavable peptides were assayed with leaf extracts from WT, *vpe1a/b* mutant and *vpe1a/b* mutant transiently expressing NbVPE1b-His or NbVPE1a^{C215S}-His. The rate of fluorescence increase from MCA-AQNSAP-Dnp incubation correlated with the NbVPE activity in the extract, with the greatest rate of increase observed in NbVPE1b-His overexpression extracts, followed by extracts of WT leaves (**Fig. 6B**). AQASAP-ase activity was similar between all extracts, demonstrating that

VPE-dependent difference in cleavage rate requires N in the sequence. This indicates that the observed FERONIA neo-N-terminus can be generated by endoprotease activity in a VPE-dependent manner and is therefore less likely to be generated by aminopeptidase activity against the N-terminus generated by signal peptide processing.

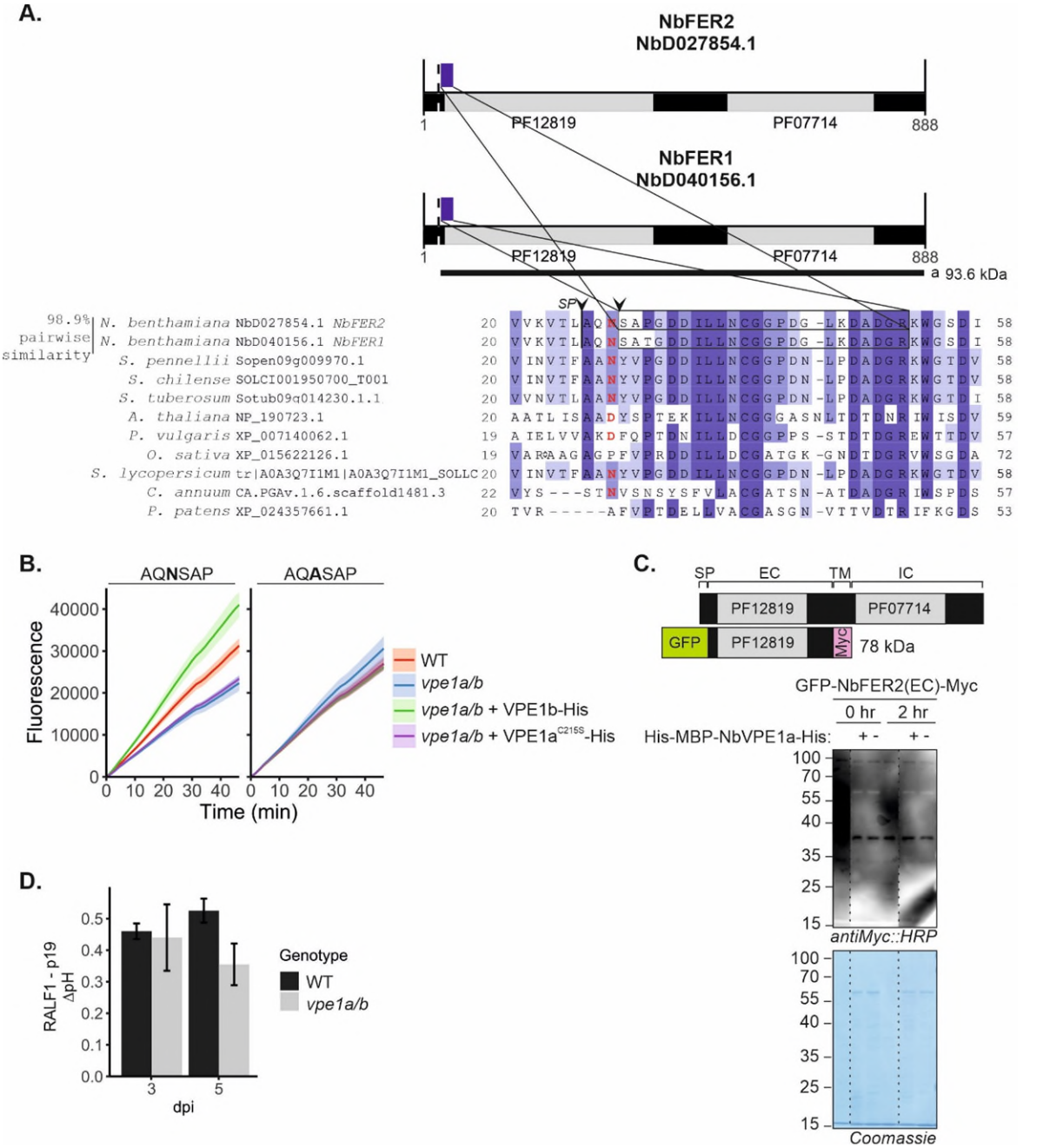


Figure 6. FERONIA is a putative substrate of NbVPE1b. A. Neo N-terminal peptides (blue) mapping to NbFER1 and NbFER2 were detected by HUNTER and significantly increased in WT and NbVPE1b-His overexpression samples compared to *vpe1a/b* and NbVPE1a^{C215S}-His overexpression. Dashed line = signal peptide processing. Domains (grey) are labelled with

PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **B.** Cleavage of synthetic AQNSAP NbFER2 peptide is NbVPE dependent. MCA-AQNSAP-dnp or MCA-AQASAP-dnp control were incubated with leaf extracts from WT, *vpe1a/b* mutant or *vpe1a/b* mutant complemented with overexpression of active NbVPE1b-His or inactive NbVPE1a^{C215S}-His. Error bars represent standard deviation of n = 6. **C.** *Preliminary data.* The extracellular portion of NbFER2 was expressed in *E. coli* and incubated with purified His-MBP-NbVPE1a-His or His-MBP-NbVPE1b^{C215S}-His for 2hrs at 37°C. Recombinant protein was detected by western blotting. SP = signal peptide, EC = extracellular, TM = transmembrane, IC = intracellular. **D.** RALF1 induced alkalinsation of the apoplast is not perturbed in the *vpe1a/b* mutant. NbRALF1 and P19, or P19 control alone, was agroinfiltrated into mature *N. benthamiana* leaves and apoplastic fluids extracted at 3 and 5 dpi. The difference in pH between control and RALF1 expressing leaves was measured. Error bars represent standard deviation of n = 4, and statistical significance tested with Students T-test.

To confirm that NbVPE1b can cleave the sequence in the context of the NbFER protein, the extracellular portion of NbFER2 was expressed in *E. coli*. The signal peptide was replaced with GFP to generate sufficient MW shift upon cleavage, and a C-terminal Myc tag was used for detection. However, the predicted 77.8 kDa GFP-NbFER2(EC)-Myc failed to accumulate and a consistent ~38 kDa band was observed independent of the co-incubation, indicating that an endogenous *E. coli* protease cleaves in the middle of the FER2 extracellular domain (**Fig. 6C**). This was a preliminary experiment and optimisation of incubation conditions and detection with an antiGFP antibody or inclusion of protease inhibitors to block *E. coli* proteases may reveal VPE-dependent processing.

FERONIA is a key receptor for the RALF family of signalling peptides in Arabidopsis (Abarca *et al.*, 2020). Perception of the AtRALF1 ligand by AtFERONIA stimulates an increase in apoplastic pH (Haruta *et al.*, 2014). To investigate if the differential NbFER1/2 processing could interfere with RALF1 induced alkalinisation, the pH was measured of apoplastic fluids (AFs) that were isolated from leaves transiently expressing NbRALF1 or an empty vector

control in WT and *vpe1a/b* mutant plants. At both 3 and 5 dpi, the AF of NbRALF1 expressing plants was more alkaline pH than that of empty vector control leaves, indicating that NbRALF1 similarly triggers AF alkalinisation. However the pH was similarly increased in WT and *vpe1a/b* mutants (**Fig. 6D**). Therefore the increased AF pH triggered by NbRALF1 is not dependent on NbVPE1a/b. If the role of FERONIA in perceiving RALF1 is conserved between Arabidopsis and *N. benthamiana,* this would indicate NbFER signal transduction is independent of NbVPE1a/b.

4.3.2 Role of NbVPEs in activation of NbRD21

HUNTER also identified a neo-N-terminus of NbRD21 that was significantly increased in samples overexpressing NbVPE1b-His (**Table 2, Fig. 5B**). Interestingly, two overlapping peptides starting a few residues upstream were found to be less abundant in the NbVPE1b-His overexpression. This suggests that multiple processing sites exist within a restricted region, and the cleavage site after D is preferred upon NbVPE1b-His overexpression. The VPE-dependent neo-N-termini mapped to the start of the NbRD21 peptidase domain, representing a mature NbRD21 proteoform with the inhibitory N-terminal domain removed (**Fig. 7A**). This VPE-dependent peptide therefore corresponds to an activation event, that would convert the 49 kDa zymogen into a 35 kDa active protease.

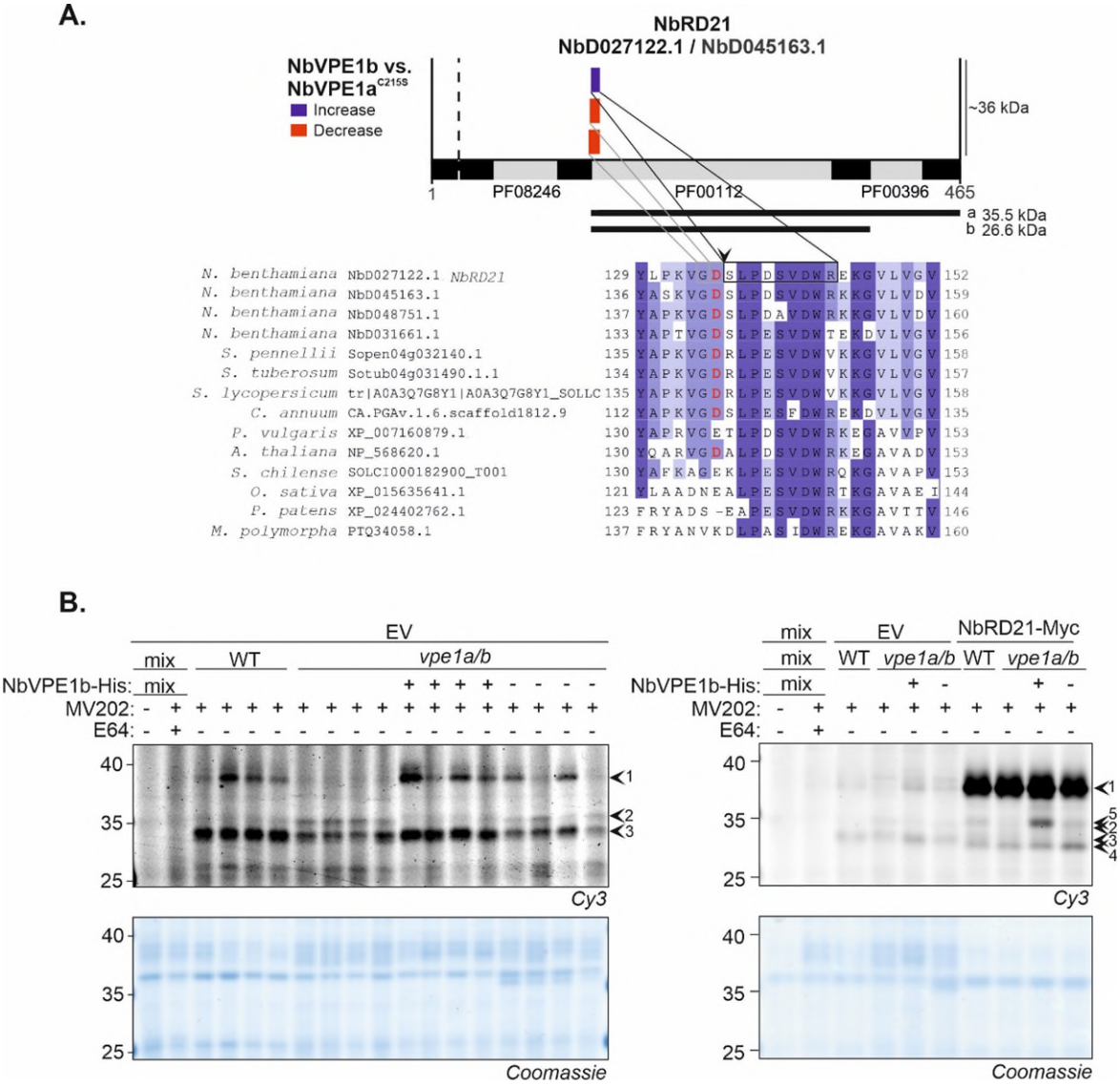


Figure 7. NbVPE1b may be involved in activation of NbRD21. A. Neo N-terminal peptides identified in HUNTER comparison of NbVPE1b-His and NbVPE1a^{C215S}-His overexpression. Dashed line = signal peptide processing. Domains (grey) are labelled with PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **B.** PLCP activity profile is dependent on VPEs. Leaves from WT, *vpe1a/b* or *vpe1a/b* leaves overexpressing NbVPE1b-His (+) or inactive NbVPE1a^{C215S}-His (-) were agroinfiltrated with the empty vector (EV) control, or NbRD21-GFP. Extracts were harvested at 4 dpi and pre-incubated with 100 µM E64 or equal volume of DMSO for 30 mins. Samples were then incubated with 2 µM of PLCP probe MV202 or an equal volume of DMSO for 3hr. All samples were mixed in equal volumes for the no probe or E64 controls. Probe was then detected using in gel fluorescence scanning.

To further investigate if VPEs play a role in activating NbRD21, the activity state of leaf papain-like cysteine proteases (PLCPs) was monitored using the activity-based probe MV202 (Richau *et al.*, 2012). In the background of a control agroinfiltration, PLCP labelling patterns in WT and *vpe1a/b* leaves differed greatly. The labelling profile of WT leaves had more prominent ~38 and ~33 kDa signals and lacked a ~35 kDa signal compared to the *vpe1a/b* mutant (**Fig. 7B, arrows 1,3,2**). Overexpression of NbVPE1b-His in the *vpe1a/b* mutant caused a comparable profile to WT, whereas overexpression of inactive NbVPE1a^{C215S}-His created profiles more similar to that of *vpe1a/b* labelling, confirming that these changes are NbVPE1a/b dependent. This upper 38 kDa band could contain NbRD21, as overexpression of NbRD21-Myc by agroinfiltration results in a very strong 38 kDa signal (**Fig. 7B, arrow 1**). This 38 kDa is larger than the expected 35.5 kDa product from the NbVPE1b cleavage. Upon NbRD21-Myc overexpression, this band did not differ between backgrounds with and without NbVPE1a/b. However a lower ~35 kDa band unique to NbRD21-Myc overexpressors was more intense in the presence of VPEs (**Fig. 7B, arrow 5**). This MW correlates with the predicted size of the NbVPE1a/b processed NbRD21, suggesting it could be the product of VPE-dependent processing.

4.3.3. Role of NbVPEs in activation of NbALP

Interestingly, inhibitory pro-domain cleavage is also observed with another PLCP aleurain-like protease, NbALP. The OE comparison identified an increased neo-N-terminal peptide of NbALP that also mapped to the start of the papain peptidase domain. An overlapping peptide spanning the putative cleavage site was less abundant upon NbVPE1b-His overexpression (**Fig. 8A**). Although the cleavage region of NbALP shares no sequence homology with

NbRD21, and the putative VPE-generated peptides initiate after D and N residues respectively, the parallels between these two examples are striking.

Considering the potential VPE-dependent activation of NbALP, the activity of endogenous NbALP was monitored in the presence and absence of NbVPE1a/b using the aleurain-specific probe FY01 (Yuan *et al.*, 2006; Lu *et al.*, 2015). Proteoforms at ~30 and ~27 kDa accumulated in both WT and *vpe1a/b* leaves, indicating that NbVPE1a/b are not essential for activating this protease (**Fig. 8B**). These signals are larger than the predicted proteoform at 23.4 kDa from the detected neo-N-terminus, which could be reconciled by glycosylation of the predicted N-glycosylation site within NbALP. The presence of multiple N-termini identified in the cleaved region indicates that multiple proteases may redundantly activate NbALP. We therefore transiently overexpressed a C-terminally GFP-tagged NbALP to saturate the capacity of the compensating proteases, so that the additive effect of NbVPE1a/b processing may be detected using the *vpe1a/b* mutant and transient NbVPE1b-His overexpression. However, activation of the overexpressed NbALP was still independent of NbVPE1a/b (**Fig. 8B**). A second weaker band at ~27 kDa was detected and the ratio of the upper to lower bands remained similar between WT, *vpe1a/b* and upon NbVPE1b-His overexpression (**Fig. 8B**). The C-terminal GFP tag was consistently removed in both WT and *vpe1a/b* background so accumulation of unprocessed NbALP could not be monitored (**SFig. 2**). The removal of the C-terminal GFP or Myc tag was also observed for NbRD21 and several other candidate substrates, suggesting that the fusion tag is cleaved in these compartments.

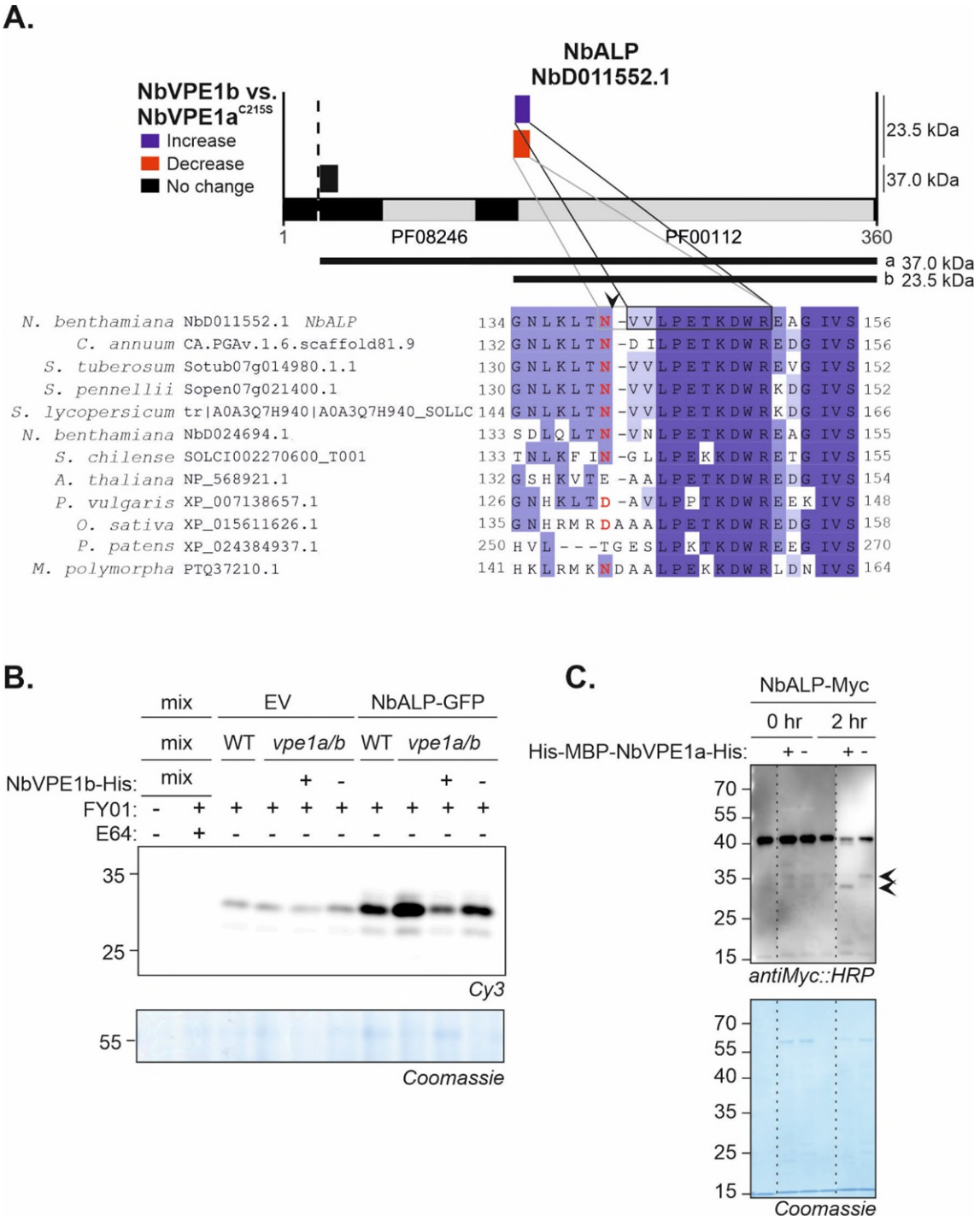


Figure 8. NbVPE1a/b may be involved in NbALP activation A. Neo N-terminal peptides identified in HUNTER comparison of NbVPE1b-His and NbVPE1a^{C215S}-His overexpression. Dashed line = signal peptide processing. Domains (grey) are labelled with PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **B.** NbALP activity profile is dependent on NbVPEs. Leaves from WT, *vpe1a/b* or *vpe1a/b* leaves overexpressing NbVPE1b-His (+) or inactive NbVPE1a^{C215S}-His (-) were agroinfiltrated with the empty vector (EV) control, or NbALP. Extracts were harvested at 4 dpi and pre-incubated with 100 µM E64 or equal volume of DMSO for 30 mins. Samples were then incubated with 2 µM of FY01 probe or an equal volume

of DMSO for 3hr. FY01 was then detected using in gel fluorescence scanning. All samples were mixed in equal volumes for the no probe or E64 controls. **C.** *Preliminary data.* NbALP produced in *E. coli* was incubated alone (blank), or with purified His-MBP-NbVPE1a-His (+) or inactive His-MBP-NbVPE1b^{C215S}-His (-). Arrows mark differential bands upon VPE incubation.

In an independent approach to investigate NbALP processing, a NbALP-Myc construct was expressed in *E. coli*. In a preliminary experiment, cleared *E. coli* extract containing NbALP-Myc was incubated with His-MBP-NbVPE1a-His or His-MBP-NbVPE1b^{C215S}-His purified from *E. coli* (**Chapter 3, Fig. 7**). Cleavage by NbVPE would be expected to generate a 27 kDa protein with complete C-terminal Myc tag. The full length 41 kDa protein decreased in abundance whereas a ~ 30 kDa proteoform appeared upon incubation with NbVPE1a (**Fig. 8C**). The ~30 kDa signal was not detected with His-MBP-NbVPE1a^{C215S}-His incubation. Levels of the full length NbALP-Myc proteoform also decreased with His-MBP-NbVPE1b^{C215S}-His incubation but this reduction was less than that observed with His-MBP-NbVPE1a-His incubation (**Fig. 8C**). This preliminary experiment needs to be repeated with optimised assay conditions to understand the role of VPE in processing of NbALP.

4.3.4 Processing of other proteases by NbVPE1a/b

PLCPs were not the only proteases present on the candidate NbVPE1b substrate list. A neo-N-terminal peptide of ATP-dependent Zinc metalloprotease was identified and maps N-terminally of the catalytic protease domain (**SFig. 1A**). Once again, an overlapping N-terminal peptide identified spanning the putative VPE cleavage site was less abundant in the active VPE samples (**Table 2**, **SFig. 1A**). Furthermore, a neo-N-terminus from subtilase NbSBT1.4 increased in active VPE samples. This putative cleavage site of NbSBT1.4 is in a

poorly conserved C-terminal region after proteolytic domains so the biological relevance of this event is unclear (**SFig. 1B**). Both proteases were cloned for transient overexpression by agroinfiltration, but they did not accumulate *in planta* (**SFig. 2**). Similarly, another candidate substrate from an uncharacterised protein was cloned and failed to accumulate when expressed by agroinfiltration (**SFig. 1D, SFig. 2**).

4.3.5 ABP19a is a putative VPE substrate

Another candidate substrate revealed was NbABP19a1 and/or NbABP19a2, highly similar germin-like proteins (GLP) which share structural homology with known VPE seed storage protein substrates (Dunwell *et al.*, 2008; Rudakova *et al.*, 2015) (**Table 1**). GLPs that are associated with plant defence form the PR16 family and have been assigned various catalytic activities including oxalate oxidase or superoxide dismutase and even cysteine protease activities (Park *et al.*, 2004; Pei *et al.*, 2019). However no publications have reported on the role of NbABP19a. Three overlapping neo-N-terminal peptides were more abundant upon NbVPE1b overexpression, although only the first originated after an N residue (**Fig. 9A**). The other two peptides could be generated by other proteases or by subsequent trimming from the NbVPE-generated N-terminus. The detected neo-N-termini suggest that the mature protein begins close to the start of the Cupin domain which is therefore likely to retain its structural fold. To our knowledge, the cleavage identified in HUNTER has not previously been reported and its biological relevance is unclear. This cleavage results in removal of a conserved box A "QDFCVAD" motif sequence in the GLP family (Dunwell *et al.*, 2008). Conserved Box B and C motifs that are key for coordination of metal ions are retained, suggesting that the putative cleavage event is unlikely to disrupt any potential catalytic activity of NbABP19a.

Transient overexpression of C-terminally myc tagged NbABP19a1-Myc and NbABP19a2-Myc *in planta* resulted in accumulation of a ~28 kDa protein, larger than the predicted ~22 kDa proteoform indicating that the putative N-glycosylation site is glycosylated (**Fig. 9B**). The pattern of NbABP19a accumulation did not differ between WT and *vpe1a/b* leaves. The smaller ~16 kDa proteoform predicted by HUNTER peptides did not strongly accumulate (**Fig. 9B**). This is surprising as we expected the cleaved product to be generated regardless of the presence of NbVPE1a/b, given the multiple peptides detected spanning the cleavage region. If NbABP19a1/2 is normally present at much lower levels, overexpression may have saturated the capacity of the proteases involved in processing this protein.

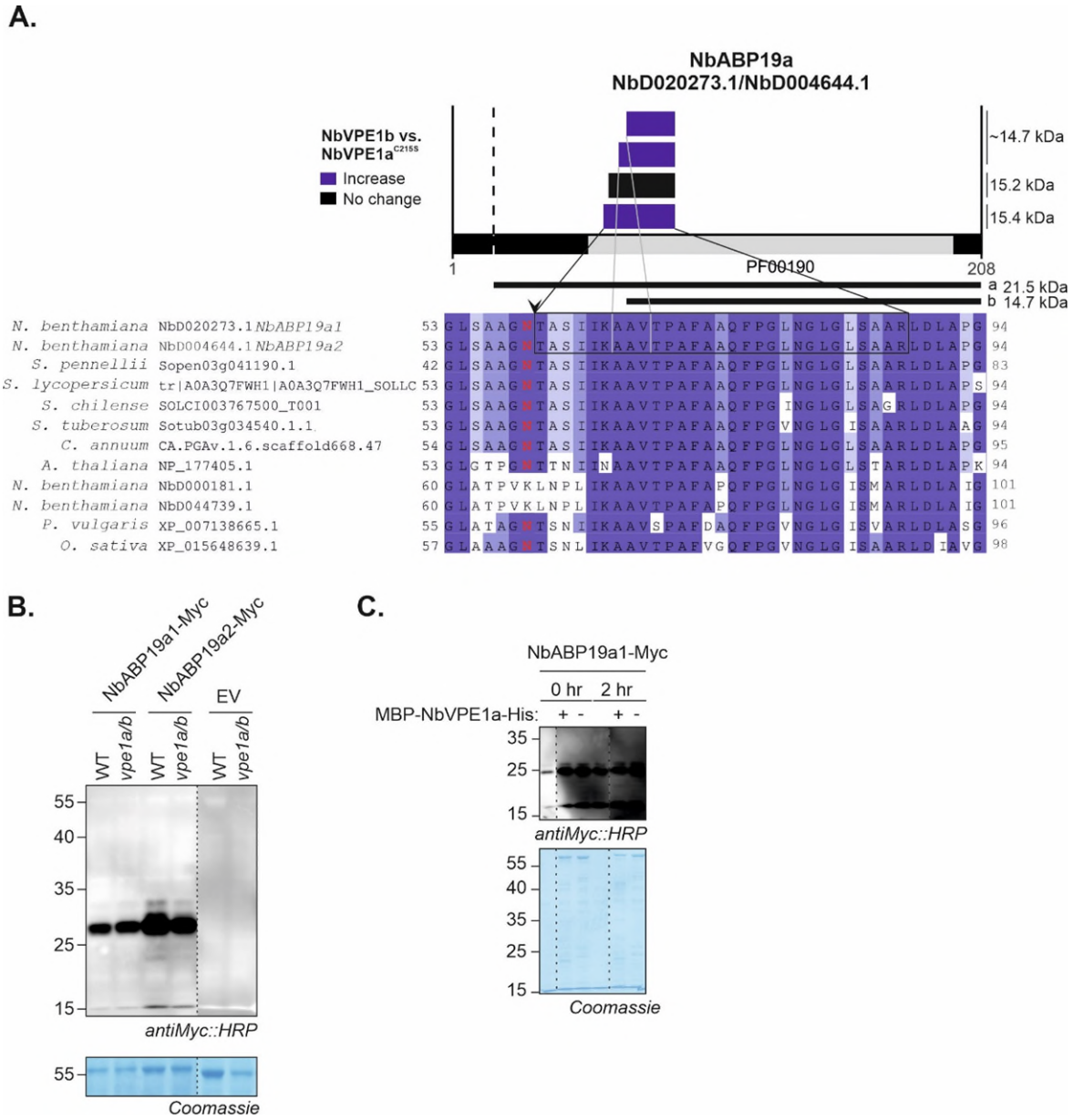


Figure 9. Putative processing of an ABP19a-like germin protein by NbVPE1b. A. Neo N-terminal peptides identified in HUNTER comparison of NbVPE1b-His and NbVPE1a^{C215S}-His overexpression. Dashed line = signal peptide processing. Domains (grey) are labelled with PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **B.** Transient overexpression of NbABP19a1-Myc and NbABP19a2-Myc by agroinfiltration in WT and *vpe* mutant mature leaves. **C.** Recombinant NbABP19a1-Myc produced in *E. coli* was incubated alone (blank), with purified His-MBP-NbVPE1a-His (+) or with inactive His-MBP-NbVPE1b^{C215S}-His (-). **B-C.** are preliminary experiments that require repeating.

In an alternative *in vitro* approach to test the dependency of NbABP19a2 processing on NbVPE1a/b, a C-terminally Myc tagged NbABP19a2-Myc was expressed in *E. coli*. Preliminary results from incubation of cleared *E. coli* extract expressing NbABP19a2-Myc with purified His-MBP-NbVPE1a-His or inactive His-MBP-NbVPE1b^{C215S}-His indicated no differential processing occurs (**Fig. 9C**). After two hours of incubation the same accumulation of ~25 kDa Myc-tagged proteoforms was detected in both incubations. Therefore, if NbVPEs are involved in processing of NbABP19a2, two hours is not sufficient to observe cleavage, or the processing of NbABP19a is inhibited when expressed in *E. coli*.

4.3.6 Acid Phosphatase is a putative VPE substrate

Acid phosphatases are vacuolar or extracellular enzymes that catalyse removal of phosphates from organic compounds. A neo-N-terminus of an acid phosphatase (NbAcPhos1a/b) was identified as significantly more abundant in both KO and OE HUNTER comparisons (**Table 2**). This peptide follows a D and N rich region that is expanded in *N. benthamiana*, but is also present in other Solanaceous species and bean (**Fig. 10A**). Transient expression of C-terminally Myc-tagged NbAcPhos1a-Myc and NbAcPhos1b-Myc resulted in accumulation of a ~36 kDa protein which is larger than the predicted 27 kDa proteoform indicating that this protein is glycosylated at the single putative glycosylation site (**Fig. 10B**). In WT plants, an additional ~28 kDa band is present compared to *vpe1a/b* mutant, however it is unclear if this is a result of a cleavage event or an artefact from higher levels of recombinant protein expression in WT. Mixing of extracts from *vpe1a/b* mutant plants overexpressing NbAcPhos1a/b with extracts containing overexpressed NbVPE1b-His or NbVPE1a^{C215S}-His did not affect protein accumulation patterns after 150 minutes (**Fig. 10C**). As with other

substrates, the presence of other N-terminal peptides from in the same region indicates that there might be redundancy that may mask the contribution of NbVPE1a/b to processing.

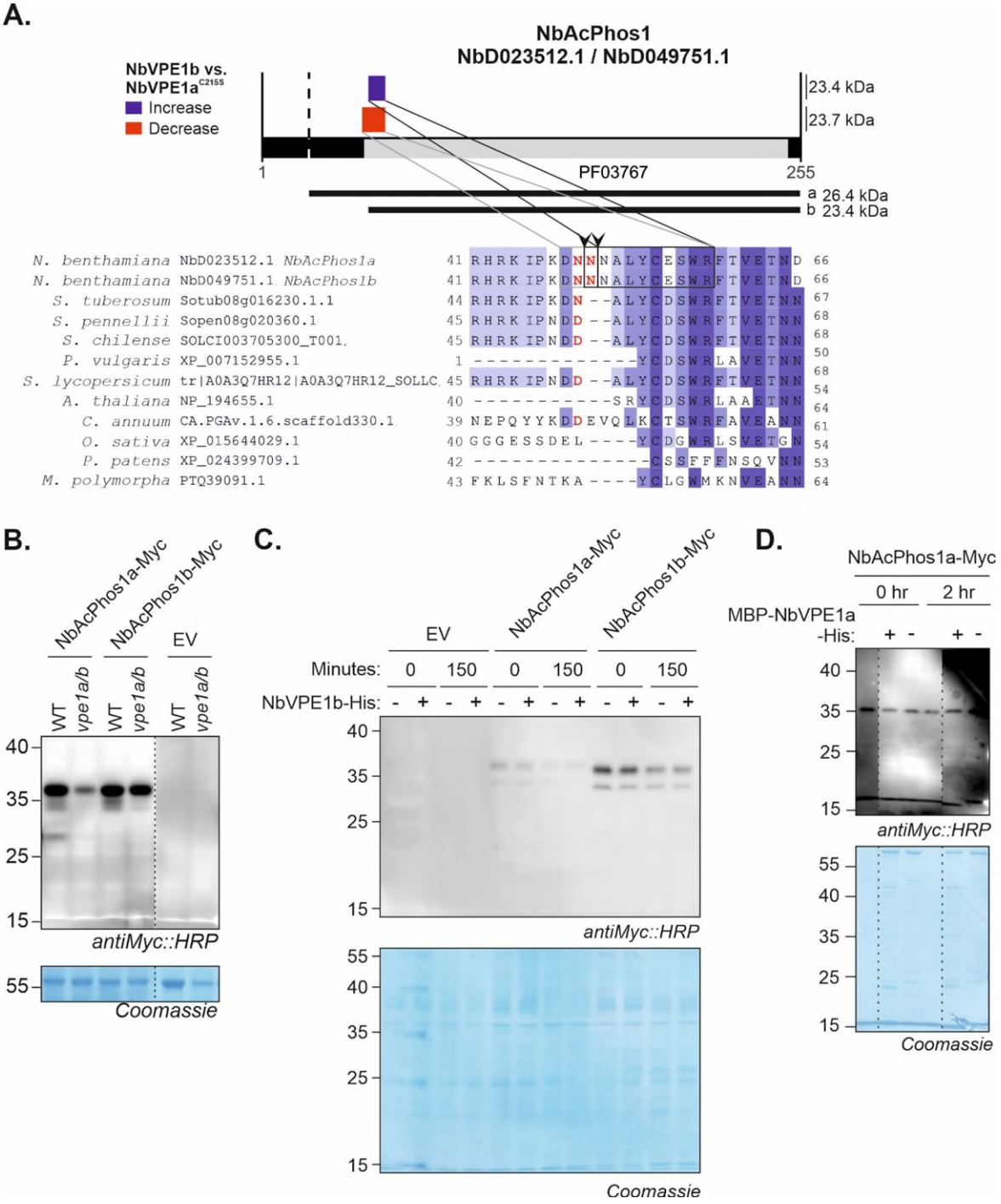


Figure 10. Putative processing of an Acid Phosphatase 1 by NbVPE1b. A. Neo N-terminal peptides identified in HUNTER screen. Dashed line = signal peptide processing. Domains (grey) are labelled with PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **B.** Transient overexpression of NbAcPhos1a-Myc and NbAcPhos1b-Myc by agroinfiltration in WT and *vpe1a/b* mutant

mature leaves. Leaves were harvested at 4 dpi and proteins were detected by western blot. **C.** NbAcPhos1a-Myc and NbAcPhos1b-Myc were co-expressed with either NbVPE1b-His (+) or NbVPE1a^{C215S}-His (-) by agroinfiltration in *vpe1a/b* mutant leaves. Protein extracts were generated at 4 dpi and proteins were visualised by western blot. **D.** Recombinant NbAcPhos1a-Myc produced in *E. coli* was incubated alone, with purified His-MBP-NbVPE1a-His (+) or with inactive His-MBP-NbVPE1b^{C215S}-His (-). **B-D.** are preliminary experiments that require repeating.

Therefore, an *in vitro* approach was employed to reduce the effect of confounding protease activities. Crude extracts from *E. coli* expressing C-terminally Myc-tagged NbAcPhos1-Myc was incubated with purified active or inactive NbVPE for 2 hours, however no processing was observed under these conditions (**Fig. 10D**).

4.4 Discussion

This chapter represents the first investigation of the contribution of NbVPE1a and NbVPE1b to shaping the proteome of *N. benthamiana*.

Comparison of *vpe1a/b* mutant and wild type plants with mass spectrometry revealed that loss of NbVPE1a and NbVPE1b activity does not have a large effect on protein accumulation *in planta*. This is in contrast to previous total protein analysis on vacuolar fractions isolated from Arabidopsis *vpeγ* mutants where multiple hydrolytic enzymes were elevated compared to WT (Rojo *et al.*, 2003). However the lack of proteomic phenotype in *N. benthamiana* mirrors the absence of a macro physiological phenotype described for both the *N. benthamiana vpe1a/b* and Arabidopsis *vpe* null mutants (Chapter 3, (Gruis *et al.*, 2004; Kuroyanagi *et al.*, 2005). However the absence of protein abundance changes does not indicate a lack of protease activity. Protease mutants of Arabidopsis lacking metacapsase 9 similarly did not have large

scale proteome changes but a number of direct substrates could still be detected using the N-terminomics technique COFRADIC (Tsiatsiani *et al.*, 2013).

NbVPE1a and NbVPE1b do contribute to the processing landscape of *N. benthamiana*. Using HUNTER, we have identified a small but significant number of processing events that are dependent on NbVPE1a/b activity. Surprisingly, we did not identify any processing events that were unique to samples with VPE activity. It would be expected that if a substrate cleavage is performed by NbVPEs, the peptide would not be identified in the absence of NbVPE1a/b. However, there are two scenarios in which this may not hold true. Firstly, the *vpe1a/b* mutant is not a complete null mutant with a further eight encoded in the genome not targeted by the CRISPR Cas9 guide RNAs. Residual signals can be detected in activity-based labelling of *vpe1a/b* (**Chapter 3, Fig 4**). The minimal activity that remains in *vpe1a/b* leaves may be sufficient to perform a fraction of the protein processing detected by HUNTER in *vpe1a/b* leaves. NbVPE1a and NbVPE1b could theoretically be compensated for by other types of VPEs, as distantly related Arabidopsis β and γ isoforms have almost identical substrate preferences and digestion of protein extracts with AtVPEβ or AtVPEγ produced a highly similar N-terminome (Misas-Villamil, 2010; Dall *et al.*, 2020).

Alternatively, VPE-generated N-termini could be produced in the absence of VPE activity by aminopeptidases trimming N-termini generated slightly upstream by other proteases. Indeed, for several proteins additional N-termini were detected a few residues upstream, that could potentially become aminopeptidase substrates. For example, some VPE-generated N-termini in seeds are substrates for aminopeptidases (Hiraiwa *et al.*, 1997; Mylne *et al.*, 2014; Weng *et al.*, 2019). Furthermore, other VPE substrates including seed storage proteins and AtCEP1

are known to also be processed by unrelated proteases, so this mechanism is highly plausible (Shimada *et al.*, 2003; Cheng *et al.*, 2019).

Redundancy in protein processing is a well reported phenomenon, and nicely exemplified by the tomato Rcr3. Rcr3 is processed by the subtilase P69B in a region containing three D residues (Paulus *et al.*, 2020). However, maturation of the protein still occurs when these residues are mutated, albeit at a reduced rate. Interestingly a distantly related subtilase NbSBT5.2 of *N. benthamiana* can also mature Rcr3, demonstrating that multiple proteases can process Rcr3 (Paulus *et al.*, 2020). The redundancy in PLCP activation may also occur for NbALP and NbRD21, as multiple peptides were identified at the cleavage site with HUNTER.

RD21 has previously been hypothesised to be activated by VPEs (Yamada 2001). However, AtRD21 processing is not perturbed in *vpe* null mutants (Gu *et al.*, 2012). Crucially though, AtRD21 can autocatalytically mature, so even if VPE does contribute to activation this would not necessarily be observed. NbRD21 has also been confirmed to autocatalytically mature when produced recombinantly in *E. coli* (Paireder *et al.*, 2017). It will be interesting to see how the maturation of catalytic inactive RD21 progresses in the absence of VPEs. Evidence suggests activation of ALP is a similarly redundant process. *In vitro* assays have shown NbALP can be activated by NbRD21 (Niemer *et al.*, 2016), however the Arabidopsis *rd21* mutant has normal AtALP maturation (Gu *et al.*, 2012). Our data suggests that inhibition of both RD21 and VPE may be required to fully perturb ALP maturation, and this could be explored in the future by silencing *NbRD21* in the *vpe1a/b* mutant background. RD21 is a highly promising candidate substrate of VPE as we have shown VPE dependent processing of active isoforms

occur. However, to confirm that NbRD21 and NbALP are direct substrates of VPEs, more evidence is required. In particular, *in vitro* cleavage assays with purified VPE using native and uncleavable substrate mutants would provide strong supporting evidence. Secondly, a difference *in vivo* needs to be observed for ALP, which may be detected using a background of protease inhibitors against other protease activities. It will also be interesting to determine *in vivo* effects of processing by testing the activity rates against synthetic substrates for PLCPs in the *vpe1a/b* mutant and upon complementation.

The HUNTER evidence for RD21 and ALP activation by VPE is in line with a recurring theme of VPE involvement in protease activation in the literature. AtCPY was among the first substrates of VPE to be confirmed (Rojo *et al.*, 2003), and has been used in this study to reveal conserved activity of *N. benthamiana* and *A. thaliana* VPEs. Another one of the few confirmed substrates is AtCEP1, a cysteine protease that is processed by both AtVPEβ and AtVPEγ. However, AtCEP1 is still matured in the *vpe* mutants and is capable of autocatalytically activating, similar to RD21. RD19A and RD19C have also been reported to be processed differently in *vpeβ* mutants (Cheng *et al.*, 2020). Clearly protease regulation is highly interconnected and is aptly described as a "web" of activities (Fortelny *et al.*, 2014).

The strongest evidence emerging from the HUNTER screen revealed a NbVPE1a/b-dependent cleavage of NbFERONIA. This detected processing occurs a few residues after the signal peptide processing site and this sequence can be cleaved directly by NbVPE1a *in vitro* (**Fig. 6B**). FERONIA is a transmembrane receptor involved in diverse immune, abiotic stress and developmental signalling processes (Li *et al.*, 2016; Stegmann *et al.*, 2017). The extracellular N-terminal region of FERONIA mediates interactions with diverse ligands, such a various

RALF peptides, pectin polymers and co-receptors such as the GPI anchored LLG1 that associates upon ligand binding (Li *et al.*, 2015; Feng *et al.*, 2018; Xiao *et al.*, 2019; Abarca *et al.*, 2020). N-terminal residues may be important in regulating protein or sugar interactions. Residues close to the N-terminus of a yeast G-protein coupled receptor regulate dimerisation and ligand binding, despite being at some distance from the binding site (Uddin *et al.*, 2017). Alternatively, N-terminal processing may somehow contribute to the stability of the FERONIA protein in the membrane. Secondary processing after signal peptide removal is a common feature in mitochondrial protein import in yeast and Arabidopsis (Carrie *et al.*, 2015). In particular, trimming by metalloprotease ICP55 of N-termini generated by the signal peptidase MPP is important for protein stability (Vögtle *et al.*, 2009, 2011). The detection of a highly similar NbVPE1a/b-dependent cleavage event of NbPR1 just after the signal peptide, suggests this activity may be more general in the apoplast. VPE cleavage could therefore have a role in stabilising proteins in the apoplast and further studies could explore whether substrate stability is dependent on NbVPE1a/b activity.

The putative substrates of NbVPE1a/b represent proteins from both the apoplast and vacuole, reflecting the known subcellular localisation of active NbVPE1a/b (Hatsugai *et al.*, 2015) (**Chapter 3**). However, protein localisation is dynamic in cells and endocytosis of apoplast proteins and transport to the vacuole is a well-documented phenomenon. In particular, in Arabidopsis FERONIA is known to be endocytosed (Yu *et al.*, 2020) and therefore processing by VPEs could occur in either the apoplastic space or upon reaching the vacuole.

The number of substrates identified here was not as high as in other N-terminomic studies. For example, HUNTER investigation of AtVPE$_\gamma$ proteome digestion identified 390 increased

cleavage events (Dall *et al.*, 2020). Crucially, however, these were from an *in vitro* digest that would include artificial protein co-localisation and conformations that are more susceptible to off-target processing. In general, a variety of reasons could contribute to lower substrate discovery. Firstly, this could be a limitation of the HUNTER method. The blocking of lysine residues in the dimethylation step restricts trypsin cleavage and results in longer peptides that are less amenable to detection by mass spectrometry. Alternatively, redundant protease activities could compensate for the lack of NbVPEs. This could be overcome by depletion of other plant proteases through the use of inhibitors. For example, Epi1 expression could be used to inhibit subtilases that can similarly process substrates after D residues. Another biological explanation is that NbVPE1a/b could have very few relevant substrates under non-challenged conditions. VPE could be a priming component that responds to relevant stresses when substrates become available. This is a less appealing hypothesis, as there are significant risks to accumulating mature, active proteases in cells. Finally, VPEs could instead be primarily involved in degradative processing. In this scenario the protein is rapidly degraded after cleavage, so no N-termini remain to be detected. It would be interesting to perform HUNTER in the presence of translation inhibitors to study the contribution of VPEs to protein turnover. Whilst the abundance of proteins did not differ much between WT and *vpe1a/b* leaves, it is possible other degradative enzymes or altered rates of transcription and translation of targets could be compensating for the lack of VPEs.

Future Work

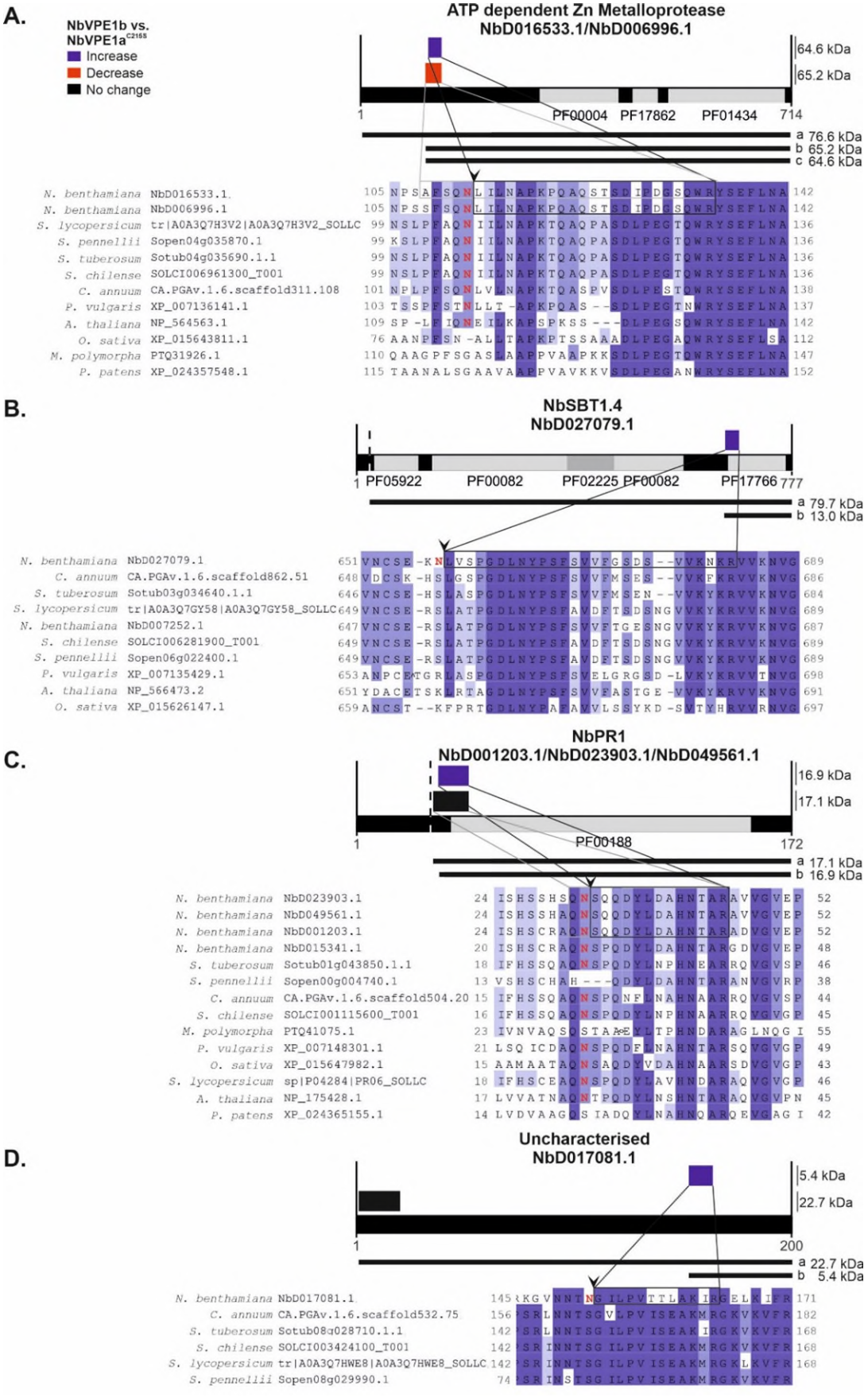
This chapter represents a great increase in knowledge of NbVPEs processing activity. A number of candidate substrates have been identified. However, these remain to be confirmed

as true direct substrates of VPE. To achieve this, we still need to establish the following supporting evidence: i) direct cleavage by NbVPE: showing that purified VPE can cleave the purified substrate *in vitro*, and that this cleavage does not occur with "uncleavable" constructs where the putative P1 residue is mutated; ii) biological relevance: establish that this processed proteoform does accumulate in WT plants and is not an artefact of VPE overexpression; and iii) *In vivo* cleavage by VPE: accumulation of processed proteoform *in planta* is dependent on VPE activity. This is supported by the HUNTER data but might be challenging to confirm given the likely redundancy in cleavage sites highlighted in this chapter, but may be achieved through a combined used of protease inhibitors that do not affect VPE and the use of mutant substrates lacking the cleavage site.

Contributions

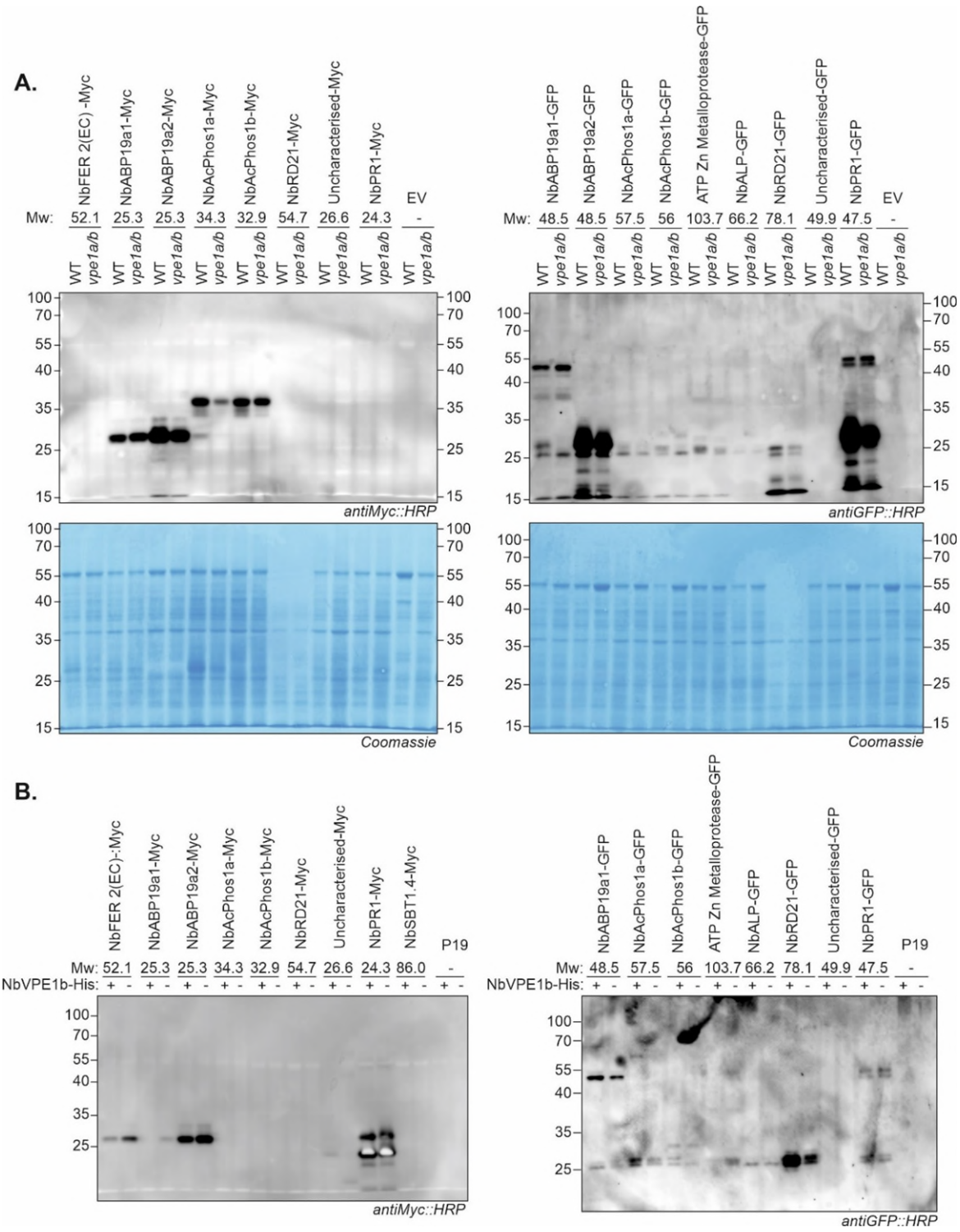
Prof. Pitter Huesgen, Dr. Fatih Demir and Andreas Perrar, provided technical assistance, materials and processing of the mass spectrometry data. Ben Excell contributed to the cloning of AtCPY and NbCPY.

4.5 Supplemental Data



Supplemental Figure 1. An ATP-dependent Zn-metalloprotease, NbSBT1.4, NbPR1 and an uncharacterised protein are putative NbVPE substrates. Neo N-terminal peptides

identified in HUNTER comparison of NbVPE1b-His and NbVPE1a^{C215S}-His overexpression. Dashed line = signal peptide processing. Domains (grey) are labelled with PFAM identifiers. ClustalW alignment of the cleaved region with the closest representatives within the orthogroup from each species. **A.** a predicted ATP-dependent Zn metalloprotease, **B.** NbSABT1.4, **C.** NbPR1 and **D.** an uncharacterised protein identified to have N-termini more abundant in NbVPE1b-His samples matching NbVPE cleavage specificity.



Supplemental Figure 2. Transient overexpression of candidate NbVPE substrates by agroinfiltration. A. Transient overexpression of C-terminally tagged candidate substrates by agroinfiltration in WT and *vpe1a/b* mutant mature leaves. Leaves were harvested at 4 dpi and proteins were detected by western blot with relevant antibodies. **B.** C-terminally myc or GFP tagged putative substrates were co-expressed with either NbVPE1b-His (+) or NbVPE1a^{C215S}-His (-) by agroinfiltration in *vpe1a/b* mutant leaves. Protein extracts were generated at 4 dpi and proteins were visualised by western blot with relevant antibodies.

4.6 Methods

Plants Growth Conditions

WT and *vpe1a/b* mutant (lacking NbVPE1a and NbVPE1b) *N. benthamiana* plants (as described in **Chapter 3**) were grown at constant temperature (21°C) and relative humidity (~60%) with a 16 hour light and 8 hour dark cycle.

Agroinfiltration

Single colonies of *Agrobacterium tumefaciens* GV3101 were used to inoculate Luria Burtani (LB) culture supplemented with rifampicin, gentamycin and kanamycin at 50 μgml^{-1} each. Cultures were incubated overnight at 28°C 200 rpm and cells collected by centrifugation at 2000g for 5 minutes. Cells were resuspended in infiltration buffer (10 mM MES, 10 mM $MgCl_2$ pH 5.7) and concentration of agrobacterium was adjusted to OD_{600} = 0.5. Acetosyringone was added to a final concentration of 100 μM. Agrobacterium were then incubated at 28°C 200 rpm for 1 hour. Where multiple constructs were infiltrated, strains were mixed at a 1:1 ratio and final OD_{600} of 0.5 was infiltrated. Fully expanded leaves of 4 week old *N. benthamiana* plants were selected for infiltration. Samples were harvested consistently at 4 days post infiltration (dpi).

Cloning

A Golden Gate based cloning system was used for all constructs assembled in this chapter. See **Table 3** for primers used for candidate amplification by PCR and **Table 4** for plasmids generated. Golden gate modules used in cloning for *E. coli* expression include: pJK082c (pET28-b modified), pPJ001 (T7 promoter), pJP002 (T7 terminator), pJK537 (pL0M-S-HIS-

sfGFP-TEV), pJK536 (pL0M-C2-TEV-sfGFP-HIS) and EC15212 (pL0M-C2-3xMyc-15212). For agroinfiltration constructs, the following modules were used: pJK268c (pL1V-P19-F2), EC41414 (pL0M-T-35S-1-41414), EC51288 (pL0M-PU-35S-TMV-3-51288), EC15095 (pL0M-C2-eGFP-15095), EC15212 (pL0M-C2-3xMyc-15212). Expression constructs were assembled with digestion by either BsaI or BpiI and ligation with T4 DNA ligase (New England BioLabs, USA). Agrobacterium GV3101 harbouring a plasmid for NbRALF1 expression was generated by Phillippe Jutras in the same Golden Gate system.

Table 3. Primers used in this study

Name	Sequence 5' - 3'	Description
NbCPY_Fw1	TTGAAGACAAAATGGCGCAAATGAAGCTTTC	Niben101Scf02803_446964-454042 / NbD020626.1
NbCPY_Rev1	TTGAAGACAACACCTATTGAAGCAACCAAGTCTTTGG	Niben101Scf02803_446964-454042 / NbD020626.1
AtCPY_F1	TTGGTCTCAAATGGAGAAACTAACTTTTCT	AT3G10410
AtCPY_R1	TTGGTCTCACACCCATCTGAGCAACC	AT3G10410
ABP19a1_Fw	TTGGTCTCAAATGTTTAAACTTCTGTTTC	Niben101Scf02749_419924-422547/ NbD020273.1
ABP19a1_Rev	TTGGTCTCACACCGTTAGTTCCTCCAAG	Niben101Scf02749_419924-422547/ NbD020273.1
ABP19a1_noSP_Fw	TTGGTCTCAAATGGCTGTCCAAGATTTC	Niben101Scf02749_419924-422547/ NbD020273.1
ABP19a2_Fw	TTGGTCTCAAATGTTTAAACTTCTTTTTCTC	Niben101Scf00493_138923-141546/ NbD004644.1
ABP19N59A_Fw	TTGGTCTCAAGCCACAGCGAGCATTATCAAAGC	Niben101Scf00493_138923-141546/ NbD004644.1
ABP19aN59A_Rev	TTGGTCTCAGGCTCCAGCTGCACTAAGCC	Niben101Scf00493_138923-141546/ NbD004644.1
ABP19a2N59A_Fw	TTGGTCTCAAGCCACAGCAAGCATAATCAAAGC	Niben101Scf00493_138923-141546/ NbD004644.1
ABP19a1_noSP_Fw	TTGGTCTCAAGGTGCTGTCCAAGATTTC	Niben101Scf02749_419924-422547/ NbD020273.1

AcPhos1_Fw	TTGGTCTCAAATGGCGTACGTTC	Niben101Scf03307_390223-401626/ NbD23512.1
AcPhos1_Rev	TTGGTCTCACACCCGCAATATAGTACATG	Niben101Scf03307_390223-401626/ NbD23512.1
AcPhos1_FwnoSP	TTGGTCTCAAATGCAGTCTGTTCTCG	Niben101Scf03307_390223-401626/ NbD23512.1
AcPhos2_Fw	TTGGTCTCAAATGGCGTACTCTC	Niben101Scf14274_119630-135356/ NbD049751.1
AcPhos2_Rev	TTGGTCTCACACCCCTTTTGGCGAC	Niben101Scf14274_119630-135356/ NbD049751.1
AcPhos2_FwnoSP	TTGGTCTCAAATGCAGTCTATTCTCGG	Niben101Scf14274_119630-135356/ NbD049751.1
AcPhosmut_Fw	TTGGTCTCAACGCCGCTTTGTACTGTGAG	N51A Niben101Scf14274_119630-135356/ NbD049751.1
AcPhosmut_Rev	TTGGTCTCAGCGTTATTGTCTTTGGGAATCTTAC	N51A Niben101Scf14274_119630-135356/ NbD049751.1
AcPhos1_FwnoSP	TTGGTCTCAAGGTCAGTCTGTTCTCG	Niben101Scf03307_390223-401626/ NbD23512.1
AcPhos2_FwnoSP	TTGGTCTCAAGGTCAGTCTATTCTCGG	Niben101Scf14274_119630-135356/ NbD049751.1
AZnMet1_Fw	TTGGTCTCAAATGGCTTTTTCAATGG	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
AZnMet1_Rev	TTGGTCTCACACCTGAAATATACAACTCGG	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
AZnMetmut_Fw	TTGGTCTCAGCTCTAATCTTGAATGCCCCAA	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
AZnMetmut_Rev	TTGGTCTCAGAGCTTGCGAAAAAGAAGATGGATT	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
AZnMet1_Fw	TTGGTCTCAAGGTGCTTTTTCAATGG	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
ALP_FwnoSP	TTGGTCTCAAATGGGACCGGCTACC	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
ALP_Fw	TTGGTCTCAAATGTCTCGTTTCTCAC	Niben101Scf01445_48569-53417/ NbD011552.1
ALP_Rev	TTGGTCTCACACCGGCAACGACAGG	Niben101Scf01445_48569-53417/ NbD011552.1
ALP_N150A_Fw	TTGGTCTCATGCCGTTGTCCTGCCGGAG	Niben101Scf01445_48569-53417/ NbD011552.1, N150A mutation
ALP_N150A_Rev	TTGGTCTCAGGCAGTGAGTTTGAGATTGCCC	Niben101Scf01445_48569-53417/ NbD011552.1, N150A mutation
ALP_FwnoSP	TTGGTCTCAAGGTGGACCGGCTACC	Niben101Scf01445_48569-53417/ NbD011552.1
RD21_NoSP	TTGGTCTCAAATGTCCGACATGTCCATCTTAAC	Niben101Scf04007_125083-130446/ NbD027122.1
RD21_Rev	TTGGTCTCACACCAGAACTGCTCTTCCGTC	Niben101Scf04007_125083-130446/ NbD027122.1
RD21_Fw	TTGGTCTCAAATGTCCGACATGTCCATCTTAAC	Niben101Scf04007_125083-130446/ NbD027122.1
RD21D134A_Fw	TTGGTCTCATCTTGGGCTTTCTCAGCAGTTG	Niben101Scf04007_125083-130446/ NbD027122.1

RD21D134A_rev	TTGGTCTCAAAGAACTCCCACATTGCCCTTG	Niben101Scf04007_125083-130446/ NbD027122.1
FER_EC_Fw	TTGGTCTCAAGGTACAGAAGGCAGTAAATT	Niben101Scf04123_264058-268721 / NbD027854.1
FER_EC_Rev1	TTGGTCTCAAAGCTGATTTATTACTTTTTGATTG	Niben101Scf04123_264058-268721 / NbD027854.1
FER_EC_Fw2	TTGGTCTCAAATGACAGAAGGCAGTAAATT	Niben101Scf04123_264058-268721 / NbD027854.1
FER_EC_BsaIremoval_Fw	TTGGTCTCATTTCTGTCCGTTCTG	Niben101Scf04007_125083-130446/ NbD027122.1
FER_EC_BsaIremoval_Rev	TTGGTCTCAGAAACCACGTAATCCTTG	Niben101Scf04007_125083-130446/ NbD027122.1
FER_ECnoSP_Fw	TTGGTCTCAAGGTCAAAATTCTGCACCTG	Niben101Scf04123_264058-268721 / NbD027854.1
FER_ECnoSP_Fw2	TTGGTCTCAAATGCAAAATTCTGCACCTG	Niben101Scf04123_264058-268721 / NbD027854.1
FER_Rev1	TTGGTCTCACACCGCGTCCTTTTGG	Niben101Scf04123_264058-268721 / NbD027854.1
FER_EC_Rev1	TTGGTCTCACACCTGATTTATTACTTTTTGATTG	Niben101Scf04123_264058-268721 / NbD027854.1
SBT1_Fw	TTGGTCTCAAATGGCACTTTCTTCATCTTC	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_FwnoSP	TTGGTCTCAAATGGTCCAATTGGACGG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_Rev	TTGGTCTCACACCCTGAGATTCTGTAGCAG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_Rev2	TTGGTCTCAAAGCCTGAGATTCTGTAGCAG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_BSAFw	TTGGTCTCAGAAACCACGCCAG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_BSARev	TTGGTCTCATTTCCGTCGTACCTG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_N657A_Fw	TTGGTCTCAGCTTTAGTTAGTCCGGGGGATTTG	Niben101Scf03998_52804-57134/NbD027079.1
SBT1_N657A_Rev	TTGGTCTCAAAGCCTTTTCGCTGCAATTCACTG	Niben101Scf03998_52804-57134/NbD027079.1
Un_Fw	TTGGTCTCAAATGGCAACAGATCAAAAATAC	Niben101Scf02261_101899-104498/ NbD017081.1
Un_Rev	TTGGTCTCACACCCTTGCAATCAAGATCTTG	Niben101Scf02261_101899-104498/ NbD017081.1
Un_N152A_Fw	TTGGTCTCAGGCAGTATTATTGACTCCTTTTCTTATG	Niben101Scf02261_101899-104498/ NbD017081.1 N152A mutation
Un_N152A_Rev	TTGGTCTCATGCCGGGATTTTGCCAGTGAC	Niben101Scf02261_101899-104498/ NbD017081.1 N152A mutation
Un_Fw no SP	TTGGTCTCAAGGTGCAACAGATCAAAAATAC	Niben101Scf02261_101899-104498/ NbD017081.1
PR11_Fw	TTGGTCTCAAATGGGATTTGTTCTCTTTTC	Niben101Scf13926_197831-205737 / NbD049561.1
PR11_FwnoSP	TTGGTCTCAAATGCAAAATTCTCAACAAGAC	Niben101Scf13926_197831-205737 / NbD049561.1
PR1_Rev	TTGGTCTCAAAGCAGCAGAAGTCCGTAC	Niben101Scf13926_197831-205737 / NbD049561.1
PR11_Fw	TTGGTCTCAAGGTGGATTTGTTCTCTTTTCAC	Niben101Scf13926_197831-205737 / NbD049561.1

PR11_FwnoSP	TTGGTCTCAAATGCAAAATTCTCAACAAGAC	Niben101Scf13926_197831-205737 / NbD049561.1
nPR11_FwnoSP	TTGGTCTCAAGGTCAAAATTCTCAACAAGAC	Niben101Scf13926_197831-205737 / NbD049561.1
PR1_Rev	TTGGTCTCACACCAGCAGAAGTCCGTAC	Niben101Scf13926_197831-205737 / NbD049561.1
PR1mut_Fw	TTGGTCTCAGCTTCTCCACAAGACTATTTGGATG	Niben101Scf13926_197831-205737 / NbD049561.1 N32A mutation
PR1mut_Rev	TTGGTCTCAAAGCTTGGGCACGACAAGAGTG	Niben101Scf13926_197831-205737 / NbD049561.1 N32A mutation
PR1mut_noSPRev1	TTGGTCTCAAAGCTTGACCTTGGAAGTATAGGTT	Niben101Scf13926_197831-205737 / NbD049561.1 N32A mutation
PR1mut_noSPRev2	TTGGTCTCAAAGCTTGACCTCCCAGATCTTC	Niben101Scf13926_197831-205737 / NbD049561.1 N32A mutation

Table 4. Plasmids generated in this study

Plasmid	Description	Selection	Gene IDs
pET003	pL0M-C1-NbVPE1a	Spec	NbD029609.1 / Niben101Scf04539g03005
pET004	pL0M-C1-NbVPE1b	Spec	NbD030181.1 / Niben101Scf04675g8014
pET012	pET28b-T7::MBP-N10-HRV3C-NbVPE1a-His	Kan	NbD029609.1 / Niben101Scf04539g03005
pET013	pET28b-T7::MBP-N10-HRV3C-NbVPE1b-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET020	pL0M-C1-NbVPE1a^{C215S}	Spec	NbD029609.1 / Niben101Scf04539g03005
pET021	pL0M-C1-NbVPE1b^{C215S}	Spec	NbD030181.1 / Niben101Scf04675g8014
pET022	pL0M-C1-NbVPE1a (with native signal peptide)	Spec	NbD029609.1 / Niben101Scf04539g03005
pET023	pL0M-C1-NbVPE1b (with native signal peptide)	Spec	NbD030181.1 / Niben101Scf04675g8014
pET027	pET28b-T7::MBP-N10-HRV3C-NbVPE1b^{C215S}-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET028	pL0M-C1-NbVPE1a^{C215S} (with native signal peptide)	Spec	NbD029609.1 / Niben101Scf04539g03005
pET029	pL0M-C1-NbVPE1b^{C215S} (with native signal peptide)	Spec	NbD030181.1 / Niben101Scf04675g8014
pET035	pL2M-P19-2x35S::NbVPE1b-His	Kan	NbD030181.1 / Niben101Scf04675g8014
pET036	pL2M-P19-2x35S::NbVPE1a^{C215S}-His	Kan	NbD029609.1 / Niben101Scf04539g03005
pET086	pL2M-P19-2x35S::ABP19a1-eGFP	Kan	Niben101Scf02749_419924-422547/ NbD020273.1
pET087	pL2M-P19-2x35S::ABP19a2-eGFP	Kan	Niben101Scf00493_138923-141546/ NbD004644.1
pET088	pL2M-P19-2x35S::AcPhos1-eGFP	Kan	Niben101Scf03307_390223-401626/ NbD23512.1

pET089	pL2M-P19-2x35S::AcPhos2-eGFP	Kan	Niben101Scf14274_119630-135356/ NbD049751.1
pET090	pL2M-P19-2x35S::ATPZnMetalloprotease1-eGFP	Kan	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
pET092	pL2M-P19-2x35S::ALP-eGFP	Kan	Niben101Scf01445_48569-53417
pET093	pL2M-P19-2x35S::RD21-eGFP	Kan	Niben101Scf04007_125083-130446/ NbD027122.1
pET094	pL2M-P19-2x35S::SBT1.4-eGFP	Kan	Niben101Scf03998_52804-57134/ NbD027079.1
pET095	pL2M-P19-2x35S::PR1-eGFP	Kan	Niben044Scf00000307Ctg000_3233-5784/ NbE03058091.1
pET099	pL2M-P19-2x35S::FERONIAextracellular-3xMyc	Kan	Niben101Scf04123_264058-268721 / NbD027854.1
pET100	pL2M-P19-2x35S::ABP19a1-3xMyc	Kan	Niben101Scf02749_419924-422547/ NbD02073.1
pET101	pL2M-P19-2x35S::ABP19a2-3xMyc	Kan	Niben101Scf00493_138923-141546/ NbD004644.1
pET102	pL2M-P19-2x35S::AcidPhos1-3xMyc	Kan	Niben101Scf03307_390223-401626/ NbD23512.1
pET103	pL2M-P19-2x35S::AcidPhos2-3xMyc	Kan	Niben101Scf14274_119630-135356/ NbD049751.1
pET107	pL2M-P19-2x35S::RD21-3xMyc	Kan	Niben101Scf04007_125083-130446/ NbD027122.1
pET108	pL2M-P19-2x35S::uncharacterised-3xMyc	Kan	Niben101Scf02261_101899-104498/ NbD017081.1
pET109	pL2M-P19-2x35S::PR1-3xMyc	Kan	Niben044Scf00000307Ctg000_3233-5784/ NbE03058091.1
pET116	pL1M-T7::ABP19a2-3xMyc	Kan	Niben101Scf00493_138923-141546/ NbD004644.1
pET117	pL1M-T7::AcPhos1-3xMyc	Kan	Niben101Scf03307_390223-401626/ NbD23512.1
pET121	pL1M-T7::ALP-3xMyc	Kan	Niben101Scf01445_48569-53417/ NbD011552.1
pET123	pL1M-T7::RD21-3xMyc	Kan	Niben101Scf04007_125083-130446/ NbD027122.1
pET125	pL1M-T7::Uncharacterised-3xMyc	Kan	Niben101Scf02261_101899-104498/ NbD017081.1
pET131	pL1M-T7::ALP-sfGFPHIS	Kan	Niben101Scf01445_48569-53417/ NbD011552.1
pET133	pL1M-T7::RD21-sfGFPHIS	Kan	Niben101Scf04007_125083-130446/ NbD027122.1
pET135	pL1M-T7::Uncharacterised-sfGFPHIS	Kan	Niben101Scf02261_101899-104498/ NbD017081.1
pET149	pL1M-T7::sfGFPHIS-FER EC-3xMyc	Kan	Niben101Scf04123_264058-268721 / NbD027854.1
pET152	pL1M-T7::sfGFPHis-PR1-3xMyc	Kan	Niben044Scf00000307Ctg000_3233-5784/ NbE03058091.1
pET154	pL1M-T7::SBT1.4-3xMyc	Kan	Niben101Scf03998_52804-57134/ NbD027079.1

pET176	uncleavable pL2M-p19-35S::AcPhosmut-eGFP	Kan	Niben101Scf03307_390223-401626/ NbD23512.1
pET178	uncleavable pL2M-p19-35S::AZnmMetmut-eGFP	Kan	Niben044Scf00026134Ctg000_22457-27974/ NbD006996.1
pET180	uncleavable pL2M-p19-35S::RD21mut-eGFP	Kan	Niben101Scf04007_125083-130446
pET182	uncleavable pL2M-p19-35S::ALPmut-eGFP	Kan	Niben101Scf01445_48569-53417/ NbD011552.1
pET184	uncleavable pL2M-p19-35S::Unmut-eGFP	Kan	Niben101Scf02261_101899-104498/ NbD017081.1
pET185	uncleavable pL2M-p19-35S::Unmut-3xMyc	Kan	Niben101Scf02261_101899-104498/ NbD017081.1
pET186	uncleavable pL2M-p19-35S::PR1mut-eGFP	Kan	Niben044Scf00000307Ctg000_3233-5784/ NbE03058091.1
pET187	uncleavable pL2M-p19-35S::PR1mut-3xMyc	Kan	Niben044Scf00000307Ctg000_3233-5784/ NbE03058091.1
pBE001	pL2M-P19-2x35S::AtCPY-GFP	Kan	AT3G10410
pBE002	pL2M-P19-2x35S::NbCPY-GFP	Kan	Niben101Scf02803_446964-454042 / NbD020626.1

Activity-based protein profiling

Leaf extracts were harvested a 4 dpi for all agroinfiltration samples. Leaf disks were ground frozen in a 85210 TissueLyser Universal Laboratory Mixer-Mill Disruptor (QIAGEN) and protein extracted in 50 mM sodium acetate pH 5, 5 mM DTT, 50 mM NaCl for labelling with DK004 or MV202, or in 50 mM Tris HCl pH 6.5, 5 mM DTT, 50 mM NaCl. Extracts were vortexed and cleared by centrifugation at 18,000 g at 4°C. The supernatant was collected and protein concentration was determined by Bradford assay. The samples were adjusted to the lowest protein concentration and 50 µl was taken for labelling. A sample mix was preincubated for 30 min with a protease inhibitor: 50 µM Ac-YVAD-cho (for DK004 labelling, or 100 µM E64 (Sigma) for FY01 and M202 labelling or equal volume DMSO control. Probes were then added to extract to 0.25 µM (DK004), 0.5 µM (FY01), 2 µM (MV202) or equal volume DMSO for the no probe control (NPC) and incubated in the dark at room temperature for 3 hours. Excess probe was removed by acetone precipitation. Protein pellets were resuspended in 2X

SDS gel loading buffer and boiled at 95°C for 5 minutes. Proteins were separated on a 12% bis tris SDS PAGE gel and labelled bands visualised by in-gel fluorescence scanning with Amersham Typhoon 5 Biomolecular Imager (Amersham Biosciences/GE Healthcare) on Cy3 detection settings (532nm excitation, 610 band pass filter).

Proteomics

Single-pot Solid-phase-enhanced Sample Preparation (SP3) Bead Preparation Hydrophilic (GE Life Sciences) and hydrophobic Sera-Mag SpeedBeads carboxylate-modified magnetic beads (GE Life Sciences) were combined in a 1:1 ratio. Beads were precipitated with a magnetic stand and solution removed by pipetting. Beads were then washed with HPLC grade water (Fisher Scientific) twice and stored at 4 °C.

HUNTER sample preparation Leaves were ground in liquid nitrogen and tissue was mixed in a 1:10 w/v ratio with extraction buffer: 100mM HEPES pH7.5, 6M guanidineHCl, 5mM EDTA, 1X HALT protease inhibitor cocktail (Thermo Fisher). Protein was purified from extract by chloroform methanol precipitation and the resulting pellet resuspended in extraction buffer without HALT protease inhibitor cocktail at 1 $mgml^{-1}$. Protein concentration was determined by BCA measurement. Samples were denatured at 56°C for 10 min and then reduced with 5 mM DTT at 56°C for 30 min. Samples were then alkylated with 15mM IAA in the dark 56°C for 30 min. The reaction was quenched by addition of DTT to a final concentration of 15 mM. Samples were then subject to dimethyl labelling by addition of formaldehyde and sodium cyanoborohydride to final concentration of 20 mM each and incubated overnight at 37°C. Heavy isotopes ($^{3}CD_2O$) were used for WT and $NbVPE1a^{C215S}$

overexpression samples, and light isotopes ($^{12}CH_2O$) used for *vpe1a/b* and NbVPE1b overexpression samples. Labelling reagents were refreshed for a further 2 hours incubation before quenching of formaldehyde by addition of Tris HCl pH 7.4 to 100 mM and mixing at 37°C for 1 hour. WT and *vpe1a/b* or NbVPE1a^{C215S} and NbVPE1b were mixed in a 1:1 ratio and chloroform methanol precipitation used to purify proteins. Proteins were resuspended in 100 mM HEPES pH7.5 to a concentration of 1 mgml^{-1}. MS-grade trypsin (Serva) was added at 1:100 ratio to protein with $CaCl_2$ and acetonitrile to final concentration of 5 mM and 5% respectively and digested overnight at 37°C. The digestion was refreshed with a further addition of trypsin at 1:200 ratio for 2 hours, and then ethanol was added to 40% final concentration. Unblocked N-termini were then filtered from the sample by hydrophobic tagging with a 50:1 ratio addition of undecanal with sodium cyanoborohydride at 20 mM incubated at 50°C for 45 min. The reaction was refreshed with further addition of undecanal to 40 mM final concentration and incubated again at 56°C for 45 min. Samples were then acidified by addition of trifluoroacetic acid (TFA) to ~1% and centrifuged. The undecanal tag was depleted using HR-X (M) spin columns (Macherey-Nagel) and subsequently fractionated with elution with acetonitrile from 10 to 50 % through a MN-HRX cartridge (Machery-Nagel). Peptide fractions were dried in a speed vacuum and resuspended in 2% acetonitirile, 0.1% TFA for LC-MS/MS detection.

Shotgun sample preparation Protein extraction was performed as described above. Following determination of protein concentration by BCA, extracts were reduced by addition of DTT to final concentration of 50 mM for 30 min at 37°C and then subsequently alkylated with IAA at 50 mM final concentration in the dark for 30 min at 37°C. Excess reagent was quenched with

addition of DTT to a final concentration of 50 mM and incubated for 20 min in the dark. Protein was mixed with SP3 magnetic beads at a ratio of 100 µg to 5 µl of prepared beads. Ethanol was added to 80% final concentration and incubated with mild rotation for 18 min. Solution was removed and the beads then washed twice with 90% ethanol. Remaining solution was evaporated and the dried beads resuspended in digestion buffer: 50 mM HEPES pH7.5, 5 mM $CaCl_2$, with addition of trypsin 1:100 (v/v). Proteins were digested for 4 hours at 42°C. Solution contained released peptides was collected and acidified with addition of TFA and peptides purified over homemade C18 stage tip columns.

Mass spectrometry data acquisition A two-column nano-HPLC setup (Ultimate 3000 nano-RSLC system with Acclaim PepMap 100 C18, ID 75 µm, particle size 3 µm columns: a trap column of 2 cm length, ThermoFisher) was used with analytical columns (Acclaim PepMap 100 C18, ID 75 µm, particle size 3 µm, 50cm length column, ThermoFisher for HUNTER and uPAC 50cm, Pharmafluidics for the label-free shotgun samples) coupled to a high resolution Q-TOF mass spectrometer (Impact II, Bruker) as described previously (Rinschen *et al.*, 2017). Peptides were separated by a binary gradient flow. For ***HUNTER*** solvent A: (H_2O + 0.1% FA) and solvent B: (ACN + 0.1% FA) were used for separation at 60 °C: 5-32.5% B for 80 minutes at 320 nL/min and a total runtime of 2 hours per sample. For ***LFQ*** a 5-30-40% B gradient was used for 120+20 minutes (140 min. separation, total of 180 minutes) at 400 nL/min flow.

Data was acquired with the Bruker HyStar Software (v3.2, Bruker Daltonics) in line-mode with a mass range from 200-1500 m/z at an acquisition rate of 4 Hz. The Top14 most intense

ions were selected for fragmentation with dynamic exclusion of previously selected precursors for the next 30 sec unless intensity increased three-fold compared to the previous precursor spectrum. Intensity-dependent fragmentation spectra were acquired between 5 Hz for low intensity precursor ions (> 500 cts) and 20 Hz for high intensity (> 25k cts) spectra. Fragment spectra were averaged from t stepped parameters, with 50% of the acquisition time manner with split parameters: 61 µs transfer time, 7 eV collision energy and a collision RF of 1500 Vpp followed by 100 µs transfer time, 9 eV collision energy and a collision RF of 1800 Vpp.

Mass spectrometry data analysis Acquired mass spectra were matched to peptide sequences at a FDR of 0.01 using MaxQuant v1.6.3.3 using standard Bruker QtoF instrument settings (Cox and Mann, 2008). Supplemental_dataset_4_NbC_AA.fasta proteome database (downloaded Jan 2019, 42851 entries (Kourelis *et al.*, 2018)) supplemented with common contaminants was used for proteome searches. Search parameters included: for ***LFQ***: tryptic peptides with fixed Cys carbamidomethylation and variable Met oxidation and for ***HUNTER***: C-terminal cleavage by ArgC with isotope labeling by light (+28.031300) or heavy (+36.075670) dimethyl labelling of peptide N termini or Lys residues, fixed Cys carbamidomethylation and variable Met oxidation, N-terminal acetylation (+42.010565) or N-terminal pyroGlu formation from Glu (-18.010565) or Gln (-17.026549). Further statistical data analysis, filtering and annotation of the N-termini datasets was performed with the Perl script MANTI.pl 3.9.9 (https://manti.sourceforge.io).

Synthetic substrate assay

Synthetic peptides MCA-AQNSAP-dnp and MCA-AQASAP-dnp were produced by GenScript (USA). Peptides were dissolved in DMSO to make a stock solution and diluted in 50 mM sodium acetate pH 5, 5 mM DTT, 50 mM NaCl. A final concentration of 50 μM was used in the assay. Total protein extract from leaves was obtained by homogenisation of leaf disks in the same buffer as peptides were diluted in. Extracts were cleared by centrifugation and supernatant protein concentration measured by Bradford assay. Extracts were normalised to the same concentration and then mixed with substrate solution in a 96 well plate in a 1:1 ratio. Fluorescence release at excitation 328 nm and emission 393 nm, gain = 100 was measured in an Infinite M200 plate reader (Tecan, Mannedorf, Switzerland).

Protein production in *E. coli*

Plasmids for *E. coli* expression were transformed into NEB T7 SHuffle strain of *E. coli* (New England Biolabs). Cultures were grown from single colonies in LB media with 100 μgml^{-1} spectinomycin and 50 μgml^{-1} kanamycin. Starter cultures were grown overnight at 30°C 200 rpm, then expression cultures inoculated in a 1:100 ratio. Once the cultures reached OD_{600} 0.6-0.8, expression was induced by addition of 0.1 mM IPTG and cultures were incubated overnight (~12 hours) at 16°C 200 rpm. Cells were pelleted by centrifugation at 2000 rpm for 5 min at 4°C and resuspended in lysis buffer (50 mM sodium acetate, pH5, 50 mM NaCl, 0.5 % Triton X100). A pinch of lysozyme was added and incubated with occasional vortexing for 30 minutes in the cold. Cell extracts were then sonicated for 5 cycles of 2min 90%, 1 min rest and finally spun down at 13000 rpm for 30 min at 4°C. Supernatant containing soluble protein was retained and the protein concentration determined by Bradford assay.

See **Chapter 3, Methods** for details of NbVPE1a and NbVPE1b^{C215S} production in and purification from *E. coli*.

Digestion assays

All VPE processing assays were performed in 50 mM sodium acetate pH 5, 5 mM DTT, 50 mM NaCl. Samples containing NbVPEs and substrate were mixed in a 1:1 ratio and incubated in lo-bind Eppendorf tubes under 400 rpm at 30°C for the indicated amount of time. Digestion mixtures were concentrated by acetone precipitation and resuspended in 2X SDS gel loading buffer. Samples were boiled at 95°C for 5 min before separation by 12% SDS PAGE. Proteins were detected by western blotting using antiMyc::HRP (Abcam, ab1326), antiGFP::HFP (Abcam, ab6663) at 1:3000 dilutions.

Datasets

Mass-spectrometry data from the HUNTER KO and OE comparison and the combined shotgun experiments are available in the **Additional Files 8**, **9** and **10** respectively.

Chapter 5: Discussion

This thesis aimed to uncover the landscape of protein processing in *N. benthamiana* leaves and elucidate the contribution of NbVPEs to protein processing and their biological roles. In **Chapter 2**, we generated a snapshot of protein processing that will serve future studies of proteins in *N. benthamiana*. In **Chapter 3** we investigated the specific roles of NbVPE1a and NbVPE1b in stress responses. Through characterisation of a new *vpe1a/b* mutant we revealed that NbVPE1a and NbVPE1b were dispensable for various types of the hypersensitive response, biotic and abiotic stress responses. In **Chapter 4** we used the materials generated in **Chapter 3** to identify novel substrates of NbVPE1a/b with the N-terminomic technique HUNTER.

5.1 Connecting substrates and proteases

The challenge in coupling protease activity to their substrates has been an overarching theme across this thesis. **Chapter 2** revealed with the degradomic techniques PROTOMAP and TAILS that the majority (58%) of the proteome is processed. For most of these processing events, the protease involved is unknown. Indeed, many of these processing events were previously unknown. Considering the prevalence of protein processing, it is somewhat surprising that more drastic phenotypes are not observed more frequently in protease mutants in the literature. However, redundancy may be expected as proteases represent one of the largest gene families in plant genomes. Conversely, the **Appendix 1** review of proteases with a role in biotic stress responses highlights how knowledge of substrates is often lacking, even

for proteases with known phenotypes. **Chapter 3** and 4 then aimed to identify substrates of a prominent plant protease, VPE.

Identification of novel substrates requires specialised mass-spectrometry degradomic techniques and subsequent experiments to confirm insights. This was the approach taken in **Chapter 4** with the application of HUNTER to identify NbVPE1a/b substrates. Whilst screens such as this are essential to capture new substrates in an unbiased manner, follow up studies are highly laborious.

In the future identification of protease-substrate modules may be expedited by also using *in silico* prediction tools (Bao *et al.*, 2019). Several *in-silico* tools for protease substrate prediction are currently available, but their use is limited. Often substrate prediction is restricted to predetermined proteases within the tools. Otherwise, knowledge of the sequence specificity of a protease is the primary factor limiting the application of these tools. More recently developed packages such as iProt-Sub and Procleave additionally integrate information regarding structural features of putative substrates to assess susceptibility to proteases (Song *et al.*, 2019; Li *et al.*, 2020). In this case, the knowledge of the characteristics of substrates is limiting the predictive power. Indeed, substrate prediction is not currently reliable, as even prediction of well-studied caspase substrates demonstrated a low level of agreement across 12 prediction tools (Bao *et al.*, 2019).

To improve prediction and identification of protease-substrate modules, two kinds of information are required; i) knowledge of biologically relevant substrates, and ii) knowledge of protease specificity determinants. Integrating the processing landscape of proteoforms that were captured by PROTOMAP and TAILS in **Chapter 2** can guide research efforts towards

cleavage events that occur *in vivo* and are therefore biologically relevant. N-terminomic datasets from TAILS and HUNTER are particularly informative because sequence level information is provided for putative processing sites which could be directly linked to protease specificity, if known. Secondly, further protease-specific studies like the characterisation of NbVPE1a/b activity presented in **Chapter 3** and 4 will build up a database of plant proteases that could be screened against substrates under investigation for a specificity match.

Integration of this data in public resources is essential for its accessibility and use in predicting protease-substrate modules. Proteomics data is routinely deposited in ProteomeXchange (Deutsch *et al.*, 2017). However, this repository is a generalised resource that is not easily used for identification of processing events. A PROTOMAP-specific database is available for these datasets but this has been underutilised and contains only a limited number of mammalian entries (Dix *et al.*, 2008). In contrast, N-termini or PTM databases offer integrated tools with search functions that are more easily navigated for the purpose of substrate identification. TopFIND is the largest centralised database of N- and C-termini integrated with the MEROPS database and equipped with PathFINDer tool that predicts pathways between N-termini and proteases (Fortelny *et al.*, 2015). Plant PTM Viewer is the most comprehensive plant-specific database available encompassing 19 post translational modifications (including N-termini from proteolysis) (Willems *et al.*, 2019). These resources have high potential to aid protein processing research, yet currently no *N. benthamiana* datasets are present in either TOPFIND or Plant PTM Viewer (October 2020). Depositing datasets such as the PROTOMAP, TAILS and HUNTER datasets generated in this thesis are vital for the success of these tools.

5.2 Understanding redundancy between proteases will be crucial in investigating individual protease biological roles

A recurring theme across the experimental results chapters was the potential for redundancy in protease activities. In **Chapter 2**, the N-terminome captured by TAILS revealed 28% of processed proteins contained 'staggered' N-termini. Staggered N-termini are multiple new N-termini present within a small region of protein sequence which could possibly be generated by different proteases. Similarly, several of the putative NbVPE substrates identified by N-terminome capture with HUNTER in **Chapter 4** also presented staggered N-termini at the region putatively cleaved by VPE. Indeed, the potential for redundancy among proteases is large as they comprise a huge gene family with more than 1200 members predicted in the *N. benthamiana* genome (Jutras *et al.*, 2020).

It is important to understand redundancy when studying the roles of individual proteases, so they can be isolated from other protease activities that would otherwise obscure phenotypes. In **Chapter 3**, we detected no phenotype of the *vpe1a/b* mutant in the responses to various biotic and abiotic stresses. Indeed the *vpe1a/b* mutant was not compromised in YVADase activity, despite NbVPE1a and NbVPE1b being able to cleave this substrate. Subtilases may compensate for the absence of NbVPE1a/b due to their shared capacity to cleave YVAD substrates. Moreover, redundancy has been demonstrated within the VPE family of Arabidopsis, and also within the subtilase family (Stührwohldt *et al.*, 2020; Yamada *et al.*, 2020). Redundancies with subtilases can be studied using Epi 1 inhibitor, whilst ESEN-cho inhibitor can be applied to investigate possible compensation from any remaining VPE activities in the *vpe1a/b* mutant.

Alternatively, the YVADase activity may not be important to the biological role of VPEs, but the other ESENase activity of NbVPE1a/b is. Indeed, upon overexpression of NbVPE1b in **Chapter 4** N-terminome capture revealed an increase in protease cleavage after N, not D. However, in **Chapter 3** we showed ESENase activity was dramatically reduced in the *vpe1a/b* mutant, indicating that no proteases are directly compensating for the reduced activity in unchallenged plants. Therefore, if a role of NbVPE1a/b mediated through asparaginase activity was compensated for by another protease, the compensating protease does not process substrates at the same peptide bond as VPE. Such indirect compensation is much more challenging to predict.

Identifying proteases that contribute to redundancy is challenging. Redundancy can be inferred between proteases that share substrate specificity, such as between the VPEs and subtilases. However, the vast majority of plant proteases have no known specificity, making prediction impossible. Furthermore, the heterogeneity in N-termini identified by TAILS in **Chapter 2** and HUNTER in **Chapter 4** illustrates that identical sequence is not essential for redundancy and that the conformation of the region may be more important. Co-expression and co-localisation of proteases may be a useful indicator of possible indirect redundancy. However, this data is also lacking for the majority of proteases. Many targeting prediction tools are available but these fail to capture the dynamics of living cells. For example, phytaspase of tomato is secreted but is subsequently re-imported by endocytosis and active within the cell (Chichkova *et al.*, 2010; Trusova *et al.*, 2019). Furthermore, predicted localisations are even less informative when studying dynamic events involving mixing of compartments such as during PCD. Alternatively, upregulation of proteases may be an indicator of redundancy. If

protease A acts redundantly with protease B, it may be expected that protease A would be upregulated in the absence of protease B through a regulatory feedback mechanism. No upregulated active proteoforms of serine hydrolases, cathepsin Bs or cysteine proteases were detected in the *vpe1a/b* mutant. However, activity may be upregulated by other mechanisms, such as a decrease in an inhibitor of protease B, that is harder to detect. Currently we lack data to predict overlap in protease activities. Further studies on different individual proteases, such as the work presented on NbVPEs in this thesis, are necessary to discern potential redundancy in activities in the future.

5.3 Activity of NbVPE1a/b is interconnected with other proteases

The investigation of NbVPEs in this thesis indicated there is a high degree of interplay between NbVPE1a/b and other protease activities. Activity-based protein profiling of *vpe1a/b* mutant revealed changes in the cysteine protease labelling profile to WT plants (**Chapter 3, Fig. 9A**). In addition, a difference in the activity-based profiles of cysteine proteases and the serine hydrolases were also observed between extracts with and without NbVPE1a/b in the background of agroinfiltration, (**Chapter 4, Fig. 2A and Fig. 7B**) demonstrating there is cross-talk between VPE activity and at least two classes of proteases. Furthermore, the HUNTER screen described in **Chapter 4** revealed multiple proteases with differentially abundant N-termini dependent upon VPE. This included two of the VPE substrate candidates, the papain-like cysteine proteases (PLCPs) NbRD21 and NbALP which may be activated by NbVPE1a/b. Other proteases including the PLCP AtCEP1 and carboxypeptidase AtCPY have previously been identified as VPE substrates (Rojo *et al.*, 2003; Cheng *et al.*, 2019, 2020). Changes in the abundance of active proteoforms of PLCPs in the absence of all, or some, VPE

activities has been also detected in western blots and by N-terminomics in Arabidopsis seeds (Weng *et al.*, 2019; Cheng *et al.*, 2020). VPEs are therefore connected directly and indirectly to other protease activities in plants.

VPEs were initially postulated to act upstream of a proteolytic cascade to induce cell death, analogous to caspases in apoptosis (Hatsugai *et al.*, 2015). However, the putative processing of NbRD21 and NbALP detected in **Chapter 4** would not represent an independent proteolytic cascade. The correct mature proteoforms of NbALP and NbRD21 were detected by ABPP in the absence of NbVPE1a/b, showing that other activators of PLCPs are present in leaves. Furthermore, AtCEP1 similarly shows redundancy in its maturation (Cheng *et al.*, 2019, 2020). Indeed, NbALP is also matured by NbRD21 (NbC14) *in vitro* (Niemer *et al.*, 2016). Even with the limited subset of three proteases, NbVPE1a/b, NbRD21 and NbALP, non-linear activity regulation is revealed, whereby a VPE may activate ALP directly and indirectly through the activation of RD21 which subsequently activates ALP. A protease network would better describe the interplay of activities proposed here. It is highly likely that once other activators of RD21 are characterised a more complex network of regulation will be revealed. Crosstalk may be further complicated by the presence of extended cystatins, that can inhibit both VPEs and PLCPs (Martinez *et al.*, 2007). Untangling this network in future studies will be interesting in understanding the contribution of cysteine proteases to protein processing.

In **Chapter 4** the application of HUNTER on WT and *vpe1a/b* mutants detected the greatest change in the P1 consensus substrate specificity not with N or D (representing VPE activity) as expected, but with other residues (**Chapter 4, Fig. 4**). Therefore, protein processing is

indirectly altered by the absence of NbVPE1a/b. It is interesting to consider if NbVPE1a and NbVPE1b might be minor protease activities that modulate major protease activities in the plant. This would be consistent with the confirmed VPE substrates AtCEP1, and putative substrates NbRD21 and NbALP, being highly expressed in their respective tissues and activated by VPE (Richau *et al.*, 2012; Han *et al.*, 2019). If the proposed PLCP candidate substrates of NbVPE1a and NbVPE1b are confirmed in follow up studies, this could reveal a general conserved role of VPEs in maturing PLCPs in plants.

5.4 NbVPE1a and NbVPE1b are not essential for PCD

An unexpected finding of *vpe1a/b* mutants in **Chapter 3** was that induction of HR was not perturbed in the mutant. This contradicted previous studies that showed NbVPE1a and NbVPE1b to be essential for HR triggered by select inducers (Hatsugai *et al.*, 2004; Zhang *et al.*, 2010). Further studies of NbVPE1a and NbVPE1b and of other plant species VPEs demonstrated that the requirement for VPE in PCD is dependent on the inducer (**Chapter 1**). Understanding the pathways and their components triggered by different inducers could provide insights into the regulation of VPE activity.

There is no single pathway to induce cell death. In animals, 17 types of regulated cell death were defined in a recent nomenclature review (Galluzzi *et al.*, 2018). Plant PCD pathways are not well characterised at a morphological level, let alone at the protein level (Kabbage *et al.*, 2017; Balakireva and Zamyatnin, 2019). However, variation in plant PCD is known. Both vacuolar rupture and fusion of vacuolar and plasma membranes has been observed in different pathosystems and may reflect the dependency on VPEs (Pajerowska-Mukhtar and Dong, 2009). Avirulent *Pseudomonas syringae* carrying AvrRpm1 induces PCD through vacuolar

and plasma membrane fusion in Arabidopsis that is not compromised in the *vpe* null mutant (Hatsugai *et al.*, 2009). In contrast, TMV/*N*-induced HR PCD that requires NbVPE1a/b proceeded through vacuolar rupture (Hatsugai *et al.*, 2004). A role of VPEs in promoting vacuolar rupture is well-established in various plant PCD systems (**Chapter 1**, (Hatsugai *et al.*, 2015)). The requirement for VPEs may therefore reflect the destiny of the vacuole in the death induction pathway.

The TMV/*N*-induced, NbVPE1a/b-dependent HR previously described by Hatsugai et al. required the resistance gene *N*, encoding a Toll/interleukin-1 receptor-like domain, nucleotide-binding leucine rich repeat protein (TIR-NB-LRR). Another class of NB-LRRs that also act upstream of HR induction are coiled-coil NB-LRRs (CC-NB-LRRs). In **Chapter 3** the *vpe1a/b* mutant was challenged with HR induction involving CC-NBS-LRRs or receptor-like-kinases recognising effectors from a variety of pathogens. We found NbVPE1a/b to be dispensable for HR induction through CC-NBS-LRRs. CC and TIR domains are distinct, with the latter having NADase activity that is necessary for immune signalling (Bayless and Nishimura, 2020). Recently, oligomerisation of CC-NBS-LRR protein Zar1 via the N-terminal CC domain into a 'resistome' was proposed to create a pore in the plasma membrane triggering death (Wang *et al.*, 2019a, 2019b). TIR-NBS-LRRs are unlikely to induce death in a similar manner as the oligomerisation is facilitated by the N-terminal domain. If a difference in HR induction mechanism is revealed between TIR-NBS-LRR and CC-NBS-LRR domains in future studies, this could present another useful predictor of VPE requirement and aid definition of cell death pathways. Likewise, the dependency on NbVPE1a/b differs between HR induced by nep1 and boehmerin (not essential), and harpin (essential) (Zhang *et al.*, 2010). The mechanism in which

death proceeds is not well-characterised for these inducers, and elucidation of the pathways they induce may reveal further factors that correlate with a requirement for NbVPE1a/b.

It is important to note that NbVPE1a and NbVPE1b being non-essential for HR does not exclude a role for NbVPEs in PCD. The *vpe1a/b* mutant is not a complete null mutant with a nominal amount of detectable activity remaining in leaves that could theoretically compensate for the loss of major activities.

No protease is universally required for plant PCD, although many proteases are essential for specific cell death types (Sueldo and van der Hoorn, 2017; Balakireva and Zamyatnin, 2019). Studies of PCD progression within a single species upon a variety of inducers are necessary to elucidate what modes of cell death induction exist in plants. The differential requirement of proteases may be a useful mechanism to inform classification of the cell death induction pathways.

5.5 Future perspectives in understanding protein processing in plants

Connecting phenotype, protease and substrate is a challenging task. This thesis has demonstrated that protein processing is a highly prevalent post-translational modification in plants and has made headway in identifying the contribution of NbVPE1a and NbVPE1b to this landscape. Release of datasets such as the *N. benthamiana* processing landscape captured in **Chapter 2** and the characterisation of NbVPE1a and NbVPE1b presented in **Chapter 3** and **4** are crucial to building our understanding of the protease network that operates within plant cells. More knowledge is required on both the processed substrates and the protease activities to predict protease-substrate modules.

References

Abarca, A., Franck, C. M. and Zipfel, C. (2020) 'Family-wide evaluation of RALF peptides in Arabidopsis thaliana', *bioRxiv*, p. 2020.06.26.174169. doi: 10.1101/2020.06.26.174169.

Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I. and Ishii, S. (1993) 'Asparaginyl endopeptidase of jack bean seeds', *The Journal of Biological Chemistry*, 268(5), p. 3525.

Albertini, A., Simeoni, F., Galbiati, M., Bauer, H., Tonelli, C. and Cominelli, E. (2014) 'Involvement of the vacuolar processing enzyme γVPE in response of Arabidopsis thaliana to water stress', *Biologia plantarum*, 58(3), pp. 531–538. doi: 10.1007/s10535-014-0417-6.

Allen, R. D., Cohen, E. A., Vonder Haar, R. A., Adams, C. A., Ma, D. P., Nessler, C. L. and Thomas, T. L. (1987) 'Sequence and expression of a gene encoding an albumin storage protein in sunflower', *MGG Molecular & General Genetics*, 210(2), pp. 211–218. doi: 10.1007/BF00325686.

Ampe, C., Van Damme, J., de Castro, L. A. B., Sampario, M. J. A. M., Van Montagu, M. and Vandekerckhove, J. (1986) 'The amino-acid sequence of the 2S sulphur-rich proteins from seeds of Brazil nut (Bertholletia excelsa H.B.K.)', *European Journal of Biochemistry*, 159(3), pp. 597–601. doi: 10.1111/j.1432-1033.1986.tb09926.x.

Araki, T. and Funatsu, G. (1987) 'The complete amino acid sequence of the B-chain of ricin E isolated from small-grain castor bean seeds. Ricin E is a gene recombination product of ricin D and ricinus communis agglutinin', *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 911(2), pp. 191–200. doi: 10.1016/0167-4838(87)90008-2.

Armenteros, J. J. A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. and Nielsen, H. (2019) 'SignalP 5.0 improves signal peptide predictions using deep neural networks', *Nature Biotechnology*, 37(4), pp. 420–423. doi: 10.1038/s41587-019-0036-z.

Atkinson, A. H., Heath, R. L., Simpson, R. J., Clarke, A. E. and Anderson, M. A. (1993) 'Proteinase inhibitors in Nicotiana alata stigmas are derived from a precursor protein which is processed into five homologous inhibitors.', *The Plant Cell*, 5(2), pp. 203–213. doi: 10.1105/tpc.5.2.203.

Atsumi, T., Nishio, T., Niwa, H., Takeuchi, J., Bando, H., Shimizu, C., Yoshioka, N., Bucala, R. and Koike, T. (2005) 'Expression of inducible 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase/PFKFB3 isoforms in adipocytes and their potential role in glycolytic regulation', *Diabetes*, 54(12), pp. 3349–3357. doi: 10.2337/diabetes.54.12.3349.

Auf Dem Keller, U., Prudova, A., Eckhard, U., Fingleton, B. and Overall, C. M. (2013) 'Systems-level analysis of proteolytic events in increased vascular permeability and complement activation in skin inflammation', *Science Signaling*, 6(258). doi: 10.1126/scisignal.2003512.

Balakireva, A. V. and Zamyatnin, A. A. (2018) 'Indispensable role of proteases in plant innate immunity', *International Journal of Molecular Sciences*, 19(2). doi: 10.3390/ijms19020629.

Balakireva, A. V. and Zamyatnin, A. A. (2019) 'Cutting out the gaps between proteases and programmed cell death', *Frontiers in Plant Science*, 10, p. 704. doi: 10.3389/fpls.2019.00704.

Balint-Kurti, P. (2019) 'The plant hypersensitive response: concepts, control and consequences', *Molecular Plant Pathology*, 0(8), pp. 1163–1178. doi: 10.1111/mpp.12821.

Bally, J., Jung, H., Mortimer, C., Naim, F., Philips, J. G., Hellens, R., Bombarely, A., Goodin, M. M. and Waterhouse, P. M. (2018) ' The Rise and Rise of Nicotiana benthamiana : A Plant for All Reasons ', *Annual Review of Phytopathology*, 56(1), pp. 405–426. doi: 10.1146/annurev-phyto-080417-050141.

Bao, Y., Marini, S., Tamura, T., Kamada, M., Maegawa, S., Hosokawa, H., Song, J. and Akutsu, T. (2019) 'Toward more accurate prediction of caspase cleavage sites: A comprehensive review of current methods, tools and features', *Briefings in Bioinformatics*, 20(5), pp. 1669–1684. doi: 10.1093/bib/bby041.

Bayless, A. M. and Nishimura, M. T. (2020) 'Reinventing the wheel with a synthetic plant inflammasome', *Proceedings of the National Academy of Sciences of the United States of America*, 117(34), pp. 20357–20359. doi: 10.1073/pnas.2013380117.

Becker, C., Shutov, A. D., Nong, V. H., Senyuk, V. I., Jung, R., Horstmann, C., Fischer, J., Nielsen, N. C. and Müntz, K. (1995) 'Purification, cDNA cloning and characterization of proteinase B, an asparagine-specific endopeptidase from germinating vetch (Vicia sativa L.) seeds.', *European Journal of Biochemistry*, 228(2), pp. 456–62.

Benchabane, M., Schlüter, U., Vorster, J., Goulet, M.-C. and Michaud, D. (2010) 'Plant cystatins', *Biochimie*, 92(11), pp. 1657–1666. doi: 10.1016/j.biochi.2010.06.006.

Bendahmane, A., Kanyuka, K. and Baulcombe, D. C. (1999) 'The Rx gene from potato controls separate virus resistance and cell death responses.', *The Plant Cell*, 11(5), pp. 781–92. doi: 10.1105/TPC.11.5.781.

Benghezal, M., Benachour, A., Rusconi, S., Aebi, M. and Conzelmann, A. (1996) 'Yeast Gpi8p is essential for GPI anchor attachment onto proteins', *EMBO Journal*, 15(23), pp. 6575–6583. doi: 10.1002/j.1460-2075.1996.tb01048.x.

Bernath-Levin, K., Nelson, C., Elliott, A. G., Jayasena, A. S., Millar, A. H., Craik, D. J. and Mylne, J. S. (2015) 'Peptide macrocyclization by a bifunctional endoprotease.', *Chemistry & Biology*, 22(5), pp. 571–582. doi: 10.1016/j.chembiol.2015.04.010.

Bienvenut, W. V., Sumpton, D., Martinez, A., Lilla, S., Espagne, C., Meinnel, T. and Giglione, C. (2012) 'Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-α-acetylation features', *Molecular and Cellular Proteomics*, 11(6), p. M111.015131. doi: 10.1074/mcp.M111.015131.

Bollhöner, B., Prestele, J. and Tuominen, H. (2012) 'Xylem cell death: emerging understanding of regulation and function', *Journal of Experimental Botany*, 63(3), pp. 1081–1094. doi: 10.1093/jxb/err438.

Bombarely, A., Rosli, H. G., Vrebalov, J., Moffett, P., Mueller, L. A. and Martin, G. B. (2012) 'A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research', *Molecular Plant-Microbe Interactions*, 25(12), pp. 1523–1530. doi: 10.1094/MPMI-06-12-0148-TA.

Bonneau, L., Ge, Y., Drury, G. E. and Gallois, P. (2008) 'What happened to plant caspases?', *Journal of Experimental Botany*, 59(3), pp. 491–499. doi: 10.1093/jxb/erm352.

Borniego, M. L., Molina, M. C., Guiamét, J. J. and Martinez, D. E. (2020) 'Physiological and Proteomic Changes in the Apoplast Accompany Leaf Senescence in Arabidopsis', *Frontiers in Plant Science*, 10, p. 1635. doi: 10.3389/fpls.2019.01635.

Borrelli, G. M., Mazzucotelli, E., Marone, D., Crosatti, C., Michelotti, V., Valè, G. and Mastrangelo, A. M. (2018) 'Regulation and evolution of NLR genes: A close interconnection for plant immunity', *International Journal of Molecular Sciences*. MDPI AG. doi: 10.3390/ijms19061662.

Bosch, M., Poulter, N. S., Perry, R. M., Wilkins, K. A. and Franklin-Tong, V. E. (2010) 'Characterization of a legumain/vacuolar processing enzyme and YVADase activity in Papaver pollen', *Plant Molecular Biology*, 74(4), pp. 381–393. doi: 10.1007/s11103-010-9681-9.

Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F. and Dong, X. (1997) 'The cpr5 mutant of Arabidopsis expresses both NPR1-dependent and NPR1-independent resistance.', *The Plant Cell*, 9(9), pp. 1573–1584. doi: 10.1105/tpc.9.9.1573.

Bozkurt, T. O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L. M., Jones, A. M. E. E., Huitema, E., van der Hoorn, R. A. L. L. and Kamoun, S. (2011) 'Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface', *Proceedings of the National Academy of Sciences*, 108(51), pp. 20832–20837. doi: 10.1073/pnas.1112708109.

Buono, R. A., Hudecek, R. and Nowack, M. K. (2019) 'Plant proteases during developmental programmed cell death', *Journal of Experimental Botany*, 70(7), pp. 2097–2112. doi: 10.1093/jxb/erz072.

Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E. L., Morimoto, K., Kaiser, M., Preston, G. M., Ichinose, Y. and Van Der Hoorn, R. A. L. (2019) 'Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides', *Science*, 364(6436). doi: 10.1126/science.aav0748.

Buscaill, P. and Rivas, S. (2014) 'Transcriptional control of plant defence responses.', *Current Opinion in Plant Biology*, 20, pp. 35–46. doi: 10.1016/J.PBI.2014.04.004.

Cai, Y.-M., Yu, J., Ge, Y., Mironov, A. and Gallois, P. (2018) 'Two proteases with caspase-3-like activity, cathepsin B and proteasome, antagonistically control ER-stress-induced programmed cell death in Arabidopsis', *New Phytologist*, 218(3), pp. 1143–1155. doi: 10.1111/nph.14676.

Carrie, C., Venne, A. S., Zahedi, R. P. and Soll, J. (2015) 'Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in Arabidopsis thaliana.', *Journal of Experimental Botany*, 66(9), pp. 2691–2708. doi: 10.1093/jxb/erv064.

Carter, W. C., Pan, S., Zouhar, J., Avila, E. L., Girke, T., Raikhel, N. V, Carter, C., Pan, S., Zouhar, J., Avila, E. L., Girke, T. and Raikhel, N. V (2004) 'The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins.', *The Plant Cell*, 16(12), pp. 3285 LP – 3303. doi: 10.1105/tpc.104.027078.

Caten, C. E. and Jinks, J. L. (1968) 'Spontaneous variability of single isolates of *Phytophthora infestans* . I. Cultural variation', *Canadian Journal of Botany*, 46(4), pp. 329–348. doi: 10.1139/b68-055.

Chandrasekar, B. (2018) *From glycosidase activity profiling to inhibitor discovery in the plant pathogen Pseudomonas syringae*. Universitäts-und Landesbibliothek Bonn.

Chandrasekar, B., Colby, T., Emon, A. E. K., Jiang, J., Ngoc Hong, T., Villamor, J. G., Harzen, A., Overkleeft, H. S. and Van Der Hoorn, R. A. L. (2014) 'Broad-range glycosidase activity profiling', *Molecular and Cellular Proteomics*, 13(10), pp. 2787–2800. doi: 10.1074/mcp.O114.041616.

Chao, C. C. K. (2014) 'Mechanisms of p53 degradation', *Clinica Chimica Acta*, 438, pp. 139–147. doi: 10.1016/j.cca.2014.08.015.

Chari, R., Mali, P., Moosburner, M. and Church, G. M. (2015) 'Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach', *Nature Methods*, 12(9), pp. 823–826. doi: 10.1038/nmeth.3473.

Chen, J. M., Fortunato, M. and Barrett, A. J. (2000) 'Activation of human prolegumain by cleavage at a C-terminal asparagine residue.', *The Biochemical Journal*, 352 Pt 2(Pt 2), pp. 327–34.

Chen, L., Shan, Y., Weng, Y., Sui, Z., Zhang, X., Liang, Z., Zhang, L. and Zhang, Y. (2016) 'Hydrophobic tagging-assisted N-termini enrichment for in-depth N-terminome analysis', *Analytical Chemistry*, 88(17), pp. 8390–8395. doi: 10.1021/acs.analchem.6b02453.

Chen, Q., Shinozaki, D., Luo, J., Pottier, M., Havé, M., Marmagne, A., Reisdorf-Cren, M., Chardon, F., Thomine, S., Yoshimoto, K. and Masclaux-Daubresse, C. (2019) 'Autophagy and nutrients management in plants', *Cells*, 8(11). doi: 10.3390/cells8111426.

Cheng, Z., Guo, X., Zhang, J., Liu, Y., Wang, B., Li, H. and Lu, H. (2020) 'βVPE is involved in tapetal degradation and pollen development by activating proprotease maturation in Arabidopsis thaliana', *Journal of Experimental Botany*, 71(6), pp. 1943–1955. doi: 10.1093/jxb/erz560.

Cheng, Z., Zhang, J., Yin, B., Liu, Y., Wang, B., Li, H. and Lu, H. (2019) 'γVPE plays an important role in programmed cell death for xylem fiber cells by activating protease CEP1 maturation in Arabidopsis thaliana.', *International Journal of Biological Macromolecules*, 137, pp. 703–711. doi: 10.1016/j.ijbiomac.2019.07.017.

Chichkova, N. V., Galiullina, R. A., Mochalova, L. V., Trusova, S. V., Sobri, Z. M., Gallois, P. and Vartapetian, A. B. (2018) 'Arabidopsis thaliana phytaspase: Identification and peculiar properties', in *Functional Plant Biology*. CSIRO, pp. 171–179. doi: 10.1071/FP16321.

Chichkova, N. V, Shaw, J., Galiullina, R. A., Drury, G. E., Tuzhikov, A. I., Kim, S. H., Kalkum, M., Hong, T. B., Gorshkova, E. N., Torrance, L., Vartapetian, A. B. and Taliansky, M. (2010) 'Phytaspase, a relocalisable cell death promoting plant protease with caspase specificity.', *The EMBO Journal*, 29(6), pp. 1149–1161. doi: 10.1038/emboj.2010.1.

Christoff, A. P., Passaia, G., Salvati, C., Alves-Ferreira, M., Margis-Pinheiro, M. and Margis, R. (2016) 'Rice bifunctional phytocystatin is a dual modulator of legumain and papain-like proteases', *Plant Molecular Biology*, 92(1–2), pp. 193–207. doi: 10.1007/s11103-016-0504-5.

Cilliers, M., van Wyk, S. G., van Heerden, P. D. R., Kunert, K. J. and Vorster, B. J. (2018) 'Identification and changes of the drought-induced cysteine protease transcriptome in soybean (Glycine max) root nodules', *Environmental and Experimental Botany*, 148, pp. 59–69. doi: 10.1016/j.envexpbot.2017.12.005.

Coffeen, W. C. and Wolpert, T. J. (2004) 'Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in Avena sativa.', *The Plant Cell*, 16(4), pp. 857–73. doi: 10.1105/tpc.017947.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J. and Gevaert, K. (2009) 'Improved visualization of protein consensus sequences by iceLogo', *Nature Methods*. Nat Methods, pp. 786–787. doi: 10.1038/nmeth1109-786.

Coll, N. S., Smidler, A., Puigvert, M., Popa, C., Valls, M. and Dangl, J. L. (2014) 'The plant metacaspase AtMC1 in pathogen-triggered programmed cell death and aging: functional linkage with autophagy.', *Cell Death and Differentiation*, 21(9), pp. 1399–1408. doi: 10.1038/cdd.2014.50.

Coll, N. S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J. L. and Epple, P. (2010) 'Arabidopsis type I metacaspases control cell death.', *Science*, 330(6009).

Conlan, B. F., Colgrave, M. L., Gillon, A. D., Guarino, R., Craik, D. J. and Anderson, M. A. (2012) 'Insights into processing and cyclization events associated with biosynthesis of the cyclic peptide kalata B1', *Journal of Biological Chemistry*, 287(33), pp. 28037–28046. doi: 10.1074/jbc.M112.347823.

Courtois-Moreau, C. L., Pesquet, E., Sjödin, A., Muñiz, L., Bollhöner, B., Kaneda, M., Samuels, L., Jansson, S. and Tuominen, H. (2009) 'A unique program for cell death in xylem fibers of *Populus* stem', *The Plant Journal*, 58(2), pp. 260–274. doi: 10.1111/j.1365-313X.2008.03777.x.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. (2011) 'Andromeda: A peptide search engine integrated into the MaxQuant environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Dahan, J., Etienne, P., Petitot, A.-S., Houot, V., P Blein, J. and Suty, L. (2001) 'Cryptogein affects expression of α3, α6 and β1 20S proteasome subunits encoding genes in tobacco.', *Journal of Experimental Botany*, 52(362), pp. 1947–1948.

Dall, E. and Brandstetter, H. (2013) 'Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), pp. 10940–5. doi: 10.1073/pnas.1300686110.

Dall, E. and Brandstetter, H. (2016) 'Structure and function of legumain in health and disease.', *Biochimie*, 122, pp. 126–50. doi: 10.1016/j.biochi.2015.09.022.

Dall, E., Fegg, J. C., Briza, P. and Brandstetter, H. (2015) 'Structure and mechanism of an aspartimide-dependent peptide ligase in human legumain', *Angewandte Chemie International Edition*, 54(10), pp. 2917–2921. doi: 10.1002/anie.201409135.

Dall, E., Zauner, F. B., Soh, W. T., Demir, F., Dahms, S. O., Cabrele, C., Huesgen, P. F. and Brandstetter, H. (2020) 'Structural and functional studies of Arabidopsis thaliana legumain beta reveal isoform specific mechanisms of activation and substrate recognition', *Journal of Biological Chemistry*, 295(37), pp. 13047–13064. doi: 10.1074/jbc.RA120.014478.

Demir, F., Niedermaier, S., Kizhakkedathu, J. N. and Huesgen, P. F. (2017a) 'Profiling of protein N-termini and their modifications in complex samples.', in Schilling, O. (ed.) *Protein Terminal Profiling: Methods and Protocols*. New York, NY: Springer New York, pp. 35–50. doi: 10.1007/978-1-4939-6850-3_4.

Demir, F., Niedermaier, S., Villamor, J. G. and Huesgen, P. F. (2017b) 'Quantitative

proteomics in plant protease substrate identification.', *New Phytologist*. doi: 10.1111/nph.14587.

Deng, M., Bian, H., Xie, Y., Kim, Y., Wang, W., Lin, E., Zeng, Z., Guo, F., Pan, J., Han, N., Wang, J., Qian, Q. and Zhu, M. (2011) 'Bcl-2 suppresses hydrogen peroxide-induced programmed cell death via OsVPE2 and OsVPE3, but not via OsVPE1 and OsVPE4, in rice', *FEBS Journal*, 278(24), pp. 4797–4810. doi: 10.1111/j.1742-4658.2011.08380.x.

Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D. S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R. L., Carver, J. J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H. and Vizcaíno, J. A. (2017) 'The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition', *Nucleic Acids Research*, 45(D1), pp. D1100–D1106. doi: 10.1093/nar/gkw936.

Diaz-Mendoza, M., Diaz, I. and Martinez, M. (2019) 'Insights on the proteases involved in barley and wheat grain germination', *International Journal of Molecular Sciences*, 20(9). doi: 10.3390/ijms20092087.

Diaz-Mendoza, M., Velasco-Arroyo, B., Santamaria, M. E., González-Melendi, P., Martinez, M. and Diaz, I. (2016) 'Plant senescence and proteolysis: Two processes with one destiny.', *Genetics and Molecular Biology*, 39(3), pp. 329–338. doi: 10.1590/1678-4685-GMB-2016-0015.

Dickinson, C. D., Hussein, E. H. and Nielsen, N. C. (1989) 'Role of posttranslational cleavage in glycinin assembly.', *The Plant Cell*, 1(4), pp. 459–469. doi: 10.1105/tpc.1.4.459.

Dickman, M. B. and Fluhr, R. (2013) 'Centrality of host cell death in plant-microbe interactions.', *Annual Review of Phytopathology*, 51, pp. 543–70. doi: 10.1146/annurev-phyto-081211-173027.

Dietrich, R. A., Delaney, T. P., Uknes, S. J., Ward, E. R., Ryals, J. A. and Dangl, J. L. (1994) 'Arabidopsis mutants simulating disease resistance response.', *Cell*, 77(4), pp. 565–77.

Ding, L.-N., Li, M., Wang, W.-J., Cao, J., Wang, Z., Zhu, K.-M., Yang, Y.-H., Li, Y.-L. and Tan, X.-L. (2019) 'Advances in plant GDSL lipases: from sequences to functional mechanisms', *Acta Physiologiae Plantarum*, 41(9), p. 151. doi: 10.1007/s11738-019-2944-4.

Dissmeyer, N., Rivas, S. and Graciet, E. (2018) 'Life and death of proteins after protease cleavage: protein degradation by the N-end rule pathway', *New Phytologist*, 218(3), pp. 929–935. doi: 10.1111/nph.14619.

Dix, M. M., Simon, G. M. and Cravatt, B. F. (2008) 'Global mapping of the topography and magnitude of proteolytic events in apoptosis', *Cell*, 134(4), pp. 679–91. doi: 10.1016/j.cell.2008.06.038.

Dix, M. M., Simon, G. M. and Cravatt, B. F. (2014) 'Global identification of caspase substrates using PROTOMAP (Protein Topography and Migration Analysis Platform)', in *Methods in Molecular Biology*, pp. 61–70. doi: 10.1007/978-1-4939-0357-3_3.

Dixon, M. S., Golstein, C., Thomas, C. M., van Der Biezen, E. A. and Jones, J. D. (2000) 'Genetic complexity of pathogen perception by plants: the example of Rcr3, a tomato gene required specifically by Cf-2.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp. 8807–14.

Dodds, P. N. and Rathjen, J. P. (2010) 'Plant immunity: Towards an integrated view of plant–pathogen interactions.', *Nature Reviews Genetics*, 11(8), pp. 539–548. doi: 10.1038/nrg2812.

Du, J., Yap, K., Chan, L. Y., Rehm, F. B. H., Looi, F. Y., Poth, A. G., Gilding, E. K., Kaas, Q., Durek, T. and Craik, D. J. (2020) 'A bifunctional asparaginyl endopeptidase efficiently catalyzes both cleavage and cyclization of cyclic trypsin inhibitors', *Nature Communications*, 11(1), pp. 1–11. doi: 10.1038/s41467-020-15418-2.

Dunwell, J. M., Gibbings, J. G., Mahmood, T. and Saqlan Naqvi, S. M. (2008) 'Germin and germin-like proteins: evolution, structure, and function', *Critical Reviews in Plant Sciences*, 27(5), pp. 342–375. doi: 10.1080/07352680802333938.

Eichmann, R. and Schäfer, P. (2012) 'The endoplasmic reticulum in plant immunity and cell death.', *Frontiers in Plant Science*, 3, p. 200. doi: 10.3389/fpls.2012.00200.

Faro, C. and Gal, S. (2005) 'Aspartic Proteinase Content of the Arabidopsis Genome', *Current Protein & Peptide Science*, 6(6), pp. 493–500. doi: 10.2174/138920305774933268.

Felle, H. H. (2001) 'pH: Signal and messenger in plant cells', *Plant Biology*, 3(6), pp. 577–591. doi: 10.1055/s-2001-19372.

Felsenstein, J. (1985) 'Confidence limits on phylogenies: an approach using the bootstrap', *Evolution*, 39(4), pp. 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x.

Feng, W., Kita, D., Peaucelle, A., Cartwright, H. N., Doan, V., Duan, Q., Liu, M. C., Maman, J., Steinhorst, L., Schmitz-Thom, I., Yvon, R., Kudla, J., Wu, H. M., Cheung, A. Y. and Dinneny, J. R. (2018) 'The FERONIA receptor kinase maintains cell-wall integrity during salt stress through Ca2+ signaling', *Current Biology*, 28(5), pp. 666-675.e5. doi: 10.1016/j.cub.2018.01.023.

Figueiredo, J., Sousa Silva, M. and Figueiredo, A. (2017) 'Subtilisin-like proteases in plant defence: the past, the present and beyond.', *Molecular Plant Pathology*. doi: 10.1111/mpp.12567.

Fortelny, N., Cox, J. H., Kappelhoff, R., Starr, A. E., Lange, P. F., Pavlidis, P. and Overall, C. M. (2014) 'Network analyses reveal pervasive functional regulation between proteases in the human protease web.', *PLoS Biology*, 12(5), p. e1001869. doi: 10.1371/journal.pbio.1001869.

Fortelny, N., Yang, S., Pavlidis, P., Lange, P. F. and Overall, C. M. (2015) 'Proteome TopFIND 3.0 with TopFINDer and PathFINDer: Database and analysis tools for the association of protein termini to pre- and post-translational events', *Nucleic Acids Research*, 43(D1), pp. D290–D297. doi: 10.1093/nar/gku1012.

Franke, B., James, A. M., Mobli, M., Colgrave, M. L., Mylne, J. S. and Rosengren, K. J. (2017) 'Two proteins for the price of one: Structural studies of the dual-destiny protein preproalbumin with sunflower trypsin inhibitor-1', *Journal of Biological Chemistry*, 292(30), pp. 12398–12411. doi: 10.1074/jbc.M117.776955.

Fu, Z. Q., Yan, S., Saleh, A., Wang, W., Ruble, J., Oka, N., Mohan, R., Spoel, S. H., Tada, Y., Zheng, N. and Dong, X. (2012) 'NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants.', *Nature*, 486(7402), p. 228. doi: 10.1038/nature11162.

Fuentes-Prior, P. and Salvesen, G. S. (2004) 'The protein structures that shape caspase activity, specificity, activation and inhibition', *Biochemical Journal*, 384(2), pp. 201–232. doi: 10.1042/BJ20041142.

Funatsu, G., Taguchi, Y., Kamenosono, M. and Yanaka, M. (1988) 'The complete amino acid sequence of the A-chain of abrin-a, a toxic protein from the seeds of Abrus precatorius',

Agricultural and Biological Chemistry, 52(4), pp. 1095–1097. doi: 10.1080/00021369.1988.10868783.

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., Andrews, D. W., Annicchiarico-Petruzzelli, M., Antonov, A. V., Arama, E., Baehrecke, E. H., Barlev, N. A., Bazan, N. G., Bernassola, F., Bertrand, M. J. M., Bianchi, K., Blagosklonny, M. V., Blomgren, K., Borner, C., Boya, P., Brenner, C., Campanella, M., Candi, E., Carmona-Gutierrez, D., Cecconi, F., Chan, F. K. M., Chandel, N. S., Cheng, E. H., Chipuk, J. E., Cidlowski, J. A., Ciechanover, A., Cohen, G. M., Conrad, M., Cubillos-Ruiz, J. R., Czabotar, P. E., D'Angiolella, V., Dawson, T. M., Dawson, V. L., De Laurenzi, V., De Maria, R., Debatin, K. M., Deberardinis, R. J., Deshmukh, M., Di Daniele, N., Di Virgilio, F., Dixit, V. M., Dixon, S. J., Duckett, C. S., Dynlacht, B. D., El-Deiry, W. S., Elrod, J. W., Fimia, G. M., Fulda, S., García-Sáez, A. J., Garg, A. D., Garrido, C., Gavathiotis, E., Golstein, P., Gottlieb, E., Green, D. R., Greene, L. A., Gronemeyer, H., Gross, A., Hajnoczky, G., Hardwick, J. M., Harris, I. S., Hengartner, M. O., Hetz, C., Ichijo, H., Jäättelä, M., Joseph, B., Jost, P. J., Juin, P. P., Kaiser, W. J., Karin, M., Kaufmann, T., Kepp, O., Kimchi, A., Kitsis, R. N., Klionsky, D. J., Knight, R. A., Kumar, S., Lee, S. W., Lemasters, J. J., Levine, B., Linkermann, A., Lipton, S. A., Lockshin, R. A., López-Otín, C., Lowe, S. W., Luedde, T., Lugli, E., MacFarlane, M., Madeo, F., Malewicz, M., Malorni, W., Manic, G., Marine, J. C., Martin, S. J., Martinou, J. C., Medema, J. P., Mehlen, P., Meier, P., Melino, S., Miao, E. A., Molkentin, J. D., Moll, U. M., Muñoz-Pinedo, C., Nagata, S., Nuñez, G., Oberst, A., Oren, M., Overholtzer, M., Pagano, M., Panaretakis, T., Pasparakis, M., Penninger, J. M., Pereira, D. M., Pervaiz, S., Peter, M. E., Piacentini, M., Pinton, P., Prehn, J. H. M., Puthalakath, H., Rabinovich, G. A., Rehm, M., Rizzuto, R., Rodrigues, C. M. P., Rubinsztein, D. C., Rudel, T., Ryan, K. M., Sayan, E., Scorrano, L., Shao, F., Shi, Y., Silke, J., Simon, H. U., Sistigu, A., Stockwell, B. R., Strasser, A., Szabadkai, G., Tait, S. W. G., Tang, D., Tavernarakis, N., Thorburn, A., Tsujimoto, Y., Turk, B., Vanden Berghe, T., Vandenabeele, P., Vander Heiden, M. G., Villunger, A., Virgin, H. W., Vousden, K. H., Vucic, D., Wagner, E. F., Walczak, H., Wallach, D., Wang, Y., Wells, J. A., Wood, W., Yuan, J., Zakeri, Z., Zhivotovsky, B., Zitvogel, L., Melino, G. and Kroemer, G. (2018) 'Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018', *Cell Death and Differentiation*, 25(3), pp. 486–541. doi: 10.1038/s41418-017-0012-4.

Gatehouse, J. A., Lycett, G. W., Croy, R. R. D. and Boulter, D. (1982) 'The post-translational proteolysis of the subunits of vicilin from pea (Pisum sativum L.)', *Biochemical Journal*, 207(3), pp. 629–632. doi: 10.1042/bj2070629.

Gatehouse, J., Bown, D., Evans, I., Gatehouse, L., Jobes, D., Preston, P. and Croy, R. (1987) 'Sequence of the seed lectin gene from pea (Pisum sativum L.).', *Nucleic Acids Research*, 15(18), p. 7642.

Gauthier, A., Lamotte, O., Reboutier, D., Bouteau, F., Pugin, A. and Wendehenne, D. (2007) 'Cryptogein-induced anion effluxes', *Plant Signaling & Behavior*, 2(2), pp. 86–95. doi: 10.4161/psb.2.2.4015.

Ge, Y., Cai, Y.-M., Bonneau, L., Rotari, V., Danon, A., McKenzie, E. A., McLellan, H., Mach, L. and Gallois, P. (2016) 'Inhibition of cathepsin B by caspase-3 inhibitors blocks programmed cell death in Arabidopsis.', *Cell Death and differentiation*, 23(9), pp. 1493–501. doi: 10.1038/cdd.2016.34.

Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. and Vandekerckhove, J. (2003) 'Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides', *Nature Biotechnology*, 21(5), pp. 566–569. doi: 10.1038/nbt810.

Ghorbani, S., Hoogewijs, K., Pečenková, T., Fernandez, A., Inzé, A., Eeckhout, D., Kawa, D., De Jaeger, G., Beeckman, T., Madder, A., Van Breusegem, F. and Hilson, P. (2016) 'The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation.', *Journal of Experimental Botany*, 67(16), pp. 4877–4887. doi: 10.1093/jxb/erw241.

Gibbs, D. J. (2015) 'Emerging functions for N-terminal protein acetylation in plants', *Trends in Plant Science*. Elsevier Ltd, pp. 599–601. doi: 10.1016/j.tplants.2015.08.008.

Gillon, A. D., Saska, I., Jennings, C. V., Guarino, R. F., Craik, D. J. and Anderson, M. A. (2007) 'Biosynthesis of circular proteins in plants', *The Plant Journal*, 53(3), pp. 505–515. doi: 10.1111/j.1365-313X.2007.03357.x.

Gilroy, E. M., Hein, I., Van Der Hoorn, R., Boevink, P. C., Venter, E., McLellan, H., Kaffarnik, F., Hrubikova, K., Shaw, J., Holeva, M., López, E. C., Borras-Hidalgo, O., Pritchard, L., Loake, G. J., Lacomme, C. and Birch, P. R. J. (2007) 'Involvement of cathepsin B in the plant disease resistance hypersensitive response.', *The Plant Journal*, 52(1), pp. 1–13. doi: 10.1111/j.1365-313X.2007.03226.x.

Goulet, C., Khalf, M., Sainsbury, F., D'Aoust, M. A. and Michaud, D. (2012) 'A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast', *Plant Biotechnology Journal*, 10(1), pp. 83–94. doi: 10.1111/j.1467-7652.2011.00643.x.

Graham, J. S., Pearce, G., Merryweather, J., Titani, K., Ericsson, L. and Ryan, C. A. (1985) 'Wound-induced proteinase inhibitors from tomato leaves. I. The cDNA-deduced primary structure of pre-inhibitor I and its post-translational processing.', *Journal of Biological Chemistry*, 260(11), pp. 6555–6560.

Grosse-Holz, F., Kelly, S., Blaskowski, S., Kaschani, F., Kaiser, M. and van der Hoorn, R. A. L. (2018) 'The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire', *Plant Biotechnology Journal*, 16(5), pp. 1068–1084. doi: 10.1111/pbi.12852.

Grosse-Holz, F. M. and Van der Hoorn, R. A. L. (2016) 'Juggling jobs: Roles and mechanisms of multifunctional protease inhibitors in plants.', *New Phytologist*, 210(3), pp. 794–807. doi: 10.1111/nph.13839.

Grozdanić, M., Vidmar, R., Vizovišek, M. and Fonović, M. (2019) 'Degradomics in Biomarker Discovery', *PROTEOMICS – Clinical Applications*, 13(6), p. 1800138. doi: 10.1002/prca.201800138.

Gruis, D., Schulze, J. and Jung, R. (2004) 'Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases.', *The Plant Cell*, 16(1), pp. 270–90. doi: 10.1105/tpc.016378.

Gruis, D., Selinger, D. A., Curran, J. M. and Jung, R. (2002) 'Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases', *Plant Cell*, 14(11), pp. 2863–2882. doi: 10.1105/tpc.005009.

Gu, C., Kolodziejek, I., Misas-Villamil, J., Shindo, T., Colby, T., Verdoes, M., Richau, K. H., Schmidt, J., Overkleeft, H. S. and van der Hoorn, R. A. L. (2010) 'Proteasome activity profiling: a simple, robust and versatile method revealing subunit-selective inhibitors and cytoplasmic, defense-induced proteasome activities.', *The Plant Journal*, 62(1), pp. 160–170. doi: 10.1111/j.1365-313X.2009.04122.x.

Gu, C., Shabab, M., Strasser, R., Wolters, P. J., Shindo, T., Niemer, M., Kaschani, F., Mach, L. and van der Hoorn, R. A. L. (2012) 'Post-translational regulation and trafficking of the granulin-containing protease RD21 of Arabidopsis thaliana.', *PLoS ONE*, 7(3), p. e32422. doi: 10.1371/journal.pone.0032422.

Halfon, S., Patel, S., Vega, F., Zurawski, S. and Zurawski, G. (1998) 'Autocatalytic activation of human legumain at aspartic acid residues', *FEBS Letters*, 438(1–2), pp. 114–118. doi: 10.1016/S0014-5793(98)01281-2.

Han, J., Li, H., Yin, B., Zhang, Y., Liu, Y., Cheng, Z., Liu, D. and Lu, H. (2019) 'The papain-like cysteine protease CEP1 is involved in programmed cell death and secondary wall thickening during xylem development in Arabidopsis', *Journal of Experimental Botany*, 70(1), pp. 205–215. doi: 10.1093/jxb/ery356.

Hansen, K. K., Saifeddine, M. and Hollenberg, M. D. (2004) 'Tethered ligand-derived peptides of proteinase-activated receptor 3 (PAR3) activate PAR1 and PAR2 in Jurkat T cells', *Immunology*, 112(2), pp. 183–190. doi: 10.1111/j.1365-2567.2004.01870.x.

Hara-Nishimura, I. (2013) 'Plant Legumain, Asparaginyl Endopeptidase, Vacuolar Processing Enzyme', *Handbook of Proteolytic Enzymes*, 2, pp. 2314–2319. doi: 10.1016/B978-0-12-382219-2.00517-2.

Hara-Nishimura, I., Inoue, K. and Nishimura, M. (1991) 'A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms', *FEBS letters*, 294(1–2), pp. 89–93. doi: 10.1016/0014-5793(91)81349-D.

Hara-Nishimura, I. and Nishimura, M. (1987) 'Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons', *Metabolism and Enzymology*, 85(2). doi: 10.1104/pp.85.2.440.

Hara-Nishimura, I., Shimada, T., Hatano, K., Takeuchi, Y. and Nishimura, M. (1998) 'Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles', *Plant Cell*, 10(5), pp. 825–836. doi: 10.1105/tpc.10.5.825.

Hara-Nishimura, I., Shimada, T., Hiraiwa, N. and Nishimura, M. (1995) 'Vacuolar Processing Enzyme responsible for maturation of seed proteins.', *Journal of Plant Physiology*, 145(5–6), pp. 632–640. doi: 10.1016/S0176-1617(11)81275-7.

Hara-Nishimura, I., Takeuchi, Y. and Nishimura, M. (1993) 'Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of Schistosoma mansoni.', *The Plant Cell*, 5(11), pp. 1651–9. doi: 10.1105/tpc.5.11.1651.

Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., Saska, I., Daly, N. L., van der Weerden, N. L., Craik, D. J. and Anderson, M. A. (2015) 'Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase', *Nature Communications*, 6, p. 10199. doi: 10.1038/ncomms10199.

Haruta, M., Sabat, G., Stecker, K., Minkoff, B. B. and Sussman, M. R. (2014) 'A peptide

hormone and its receptor protein kinase regulate plant cell expansion', *Science*, 343(6169), pp. 408–411. doi: 10.1126/science.1244454.

Hatsugai, N., Iwasaki, S., Tamura, K., Kondo, M., Fuji, K., Ogasawara, K., Nishimura, M. and Hara-Nishimura, I. (2009) 'A novel membrane fusion-mediated plant immunity against bacterial pathogens.', *Genes & Development*, 23(21), pp. 2496–2506. doi: 10.1101/gad.1825209.

Hatsugai, N., Kuroyanagi, M., Nishimura, M. and Hara-Nishimura, I. (2006) 'A cellular suicide strategy of plants: vacuole-mediated cell death', *Apoptosis*, 11, pp. 905–911. doi: 10.1007/s10495-006-6601-1.

Hatsugai, N., Kuroyanagi, M., Yamada, K., Meshi, T., Tsuda, S., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2004) 'A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death.', *Science*, 305(5685), pp. 855–8. doi: 10.1126/science.1099859.

Hatsugai, N., Yamada, K., Goto-Yamada, S. and Hara-Nishimura, I. (2015) 'Vacuolar processing enzyme in plant programmed cell death.', *Frontiers in Plant Science*, 6, p. 234. doi: 10.3389/fpls.2015.00234.

Hayashi, M., Mori, H., Nishimura, M., Akazawa, T. and Hara-Nishimura, I. (1988) 'Nucleotide sequence of cloned cDNA coding for pumpkin 11-S globulin β subunit', *European Journal of Biochemistry*, 172(3), pp. 627–632. doi: 10.1111/j.1432-1033.1988.tb13935.x.

Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N. K., Nishimura, M. and Hara-Nishimura, I. (2001) 'A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of Arabidopsis', *Plant and Cell Physiology*, 42(9), pp. 894–899. doi: 10.1093/pcp/pce144.

Haywood, J., Schmidberger, J. W., James, A. M., Nonis, S. G., Sukhoverkov, K. V, Elias, M., Bond, C. S. and Mylne, J. S. (2018) 'Structural basis of ribosomal peptide macrocyclization in plants', *eLife*, 7, p. e32955. doi: 10.7554/eLife.32955.

Hemu, X., Sahili, A. El, Hu, S., Wong, K., Chen, Y., Wong, Y. H., Zhang, X., Serra, A., Goh, B. C., Darwis, D. A., Chen, M. W., Sze, S. K., Liu, C. F., Lescar, J. and Tam, J. P. (2019) 'Structural determinants for peptide-bond formation by asparaginyl ligases', *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), pp. 11737–11746. doi: 10.1073/pnas.1818568116.

Hiraiwa, N., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1997a) 'An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants', *European Journal of Biochemistry*, 246(1), pp. 133–141. doi: 10.1111/j.1432-1033.1997.00133.x.

Hiraiwa, N., Nishimura, M. and Hara-Nishimura, I. (1997b) 'Expression and activation of the vacuolar processing enzyme in Saccharomyces cerevisiae', *The Plant Journal*, 12(4), pp. 819–829. doi: 10.1046/j.1365-313X.1997.12040819.x.

Hiraiwa, N., Nishimura, M. and Hara-Nishimura, I. (1999) 'Vacuolar processing enzyme is self-catalytically activated by sequential removal of the C-terminal and N-terminal propeptides.', *FEBS Letters*, 447(2–3), pp. 213–216. doi: 10.1016/S0014-5793(99)00286-0.

Hiraiwa, N., Takeuchi, Y., Nishimura, M. and Hara-Nishimura', I. (1993) 'A vacuolar processing enzyme in maturing and germinating seeds: its distribution and associated changes during development', *Plant and Cell Physiology*, 34(8), pp. 1197–1204. doi:

10.1093/oxfordjournals.pcp.a078541.

Hirano, H., Fukazawa, C. and Harada, K. (1984) 'The complete amino acid sequence of the A3 subunit of the glycinin seed storage protein of the soybean (Glycine max (L.) Merrill).', *Journal of Biological Chemistry*, 259(23), pp. 14371–14377.

Hirano, H., Fukazawa, C. and Harada, K. (1985) 'The primary structures of the A_4 and A_5 subunits are highly homologous to that of the A_3 subunit in the glycinin seed storage protein of soybean', *FEBS Letters*, 181(1), pp. 124–128. doi: 10.1016/0014-5793(85)81126-1.

Hirel, B., Le Gouis, J., Ney, B. and Gallais, A. (2007) 'The challenge of improving nitrogen use efficiency in crop plants: Towards a more central role for genetic variability and quantitative genetics within integrated approaches', *Journal of Experimental Botany*, 58(9), pp. 2369–2387. doi: 10.1093/jxb/erm097.

Hollenberg, M. D. (2003) 'Proteinase-activated receptors: Tethered ligands and receptor-activating peptides', *Drug Development Research*, 59(4), pp. 336–343. doi: 10.1002/ddr.10301.

Van der Hoorn, R. A. L. (2008) 'Plant proteases: From phenotypes to molecular mechanisms.', *Annual Review of Plant Biology*, 59(1), pp. 191–223. doi: 10.1146/annurev.arplant.59.032607.092835.

Hossain, M. A., Nakamura, K. and Kimura, Y. (2009) 'α-Mannosidase involved in turnover of plant complex type N-glycans in tomato (Lycopersicum esculentum) fruits', *Bioscience, Biotechnology and Biochemistry*, 73(1), pp. 140–146. doi: 10.1271/bbb.80561.

Hossain, M. A., Nakano, R., Nakamura, K., Hossain, M. T. and Kimura, Y. (2010) 'Molecular characterization of plant acidic alpha-mannosidase, a member of glycosylhydrolase family 38, involved in the turnover of N-glycans during tomato fruit ripening.', *Journal of Biochemistry*, 148(5), pp. 603–16. doi: 10.1093/jb/mvq094.

Hou, S., Jamieson, P. and He, P. (2018) 'The cloak, dagger, and shield: proteases in plant–pathogen interactions', *Biochemical Journal*, 475(15), pp. 2491–2509. doi: 10.1042/BCJ20170781.

Höwing, T., Dann, M., Hoefle, C., Hückelhoven, R. and Gietl, C. (2017) 'Involvement of Arabidopsis thaliana endoplasmic reticulum KDEL-tailed cysteine endopeptidase 1 (AtCEP1) in powdery mildew-induced and AtCPR5-controlled cell death', *PLoS ONE*. Edited by H. Lu, 12(8), p. e0183870. doi: 10.1371/journal.pone.0183870.

Höwing, T., Huesmann, C., Hoefle, C., Nagel, M.-K., Isono, E., Hückelhoven, R. and Gietl, C. (2014) 'Endoplasmic reticulum KDEL-tailed cysteine endopeptidase 1 of Arabidopsis (AtCEP1) is involved in pathogen defense.', *Frontiers in Plant Science*, 5, p. 58. doi: 10.3389/fpls.2014.00058.

Husaini, A. M., Morimoto, K., Chandrasekar, B., Kelly, S., Kaschani, F., Palmero, D., Jiang, J., Kaiser, M., Ahrazem, O., Overkleeft, H. S. and van der Hoorn, R. A. L. (2018) 'Multiplex fluorescent, activity-based protein profiling identifies active A-glycosidases and other hydrolases in plants', *Plant Physiology*, 177(1), pp. 24–37. doi: 10.1104/pp.18.00250.

Iakimova, E. T., Sobiczewski, P., Michalczuk, L., Węgrzynowicz-Lesiak, E., Mikiciński, A. and Woltering, E. J. (2013) 'Morphological and biochemical characterization of Erwinia amylovora-induced hypersensitive cell death in apple leaves', *Plant Physiology and Biochemistry*, 63, pp. 292–305. doi: 10.1016/J.PLAPHY.2012.12.006.

Ilyas, M., Hörger, A. C., Bozkurt, T. O., van den Burg, H. A., Kaschani, F., Kaiser, M.,

Belhaj, K., Smoker, M., Joosten, M. H. A. J., Kamoun, S. and van der Hoorn, R. A. L. (2015) 'Functional divergence of two secreted immune proteases of tomato.', *Current Biology*, 25(17), pp. 2300–2306. doi: 10.1016/j.cub.2015.07.030.

Inoue, K., Motozaki, A., Takeuchi, Y., Nishimura, M. and Hara-Nishimura, I. (1995) 'Molecular characterization of proteins in protein-body membrane that disappear most rapidly during transformation of protein bodies into vacuoles', *The Plant Journal*, 7(2), pp. 235–243. doi: 10.1046/j.1365-313X.1995.7020235.x.

Ito, J. and Fukuda, H. (2002) 'ZEN1 is a key enzyme in the degradation of nuclear DNA during programmed cell death of tracheary elements.', *The Plant Cell*, 14(12), pp. 3201–11. doi: 10.1105/TPC.006411.

Jackson, M. A., Gilding, E. K., Shafee, T., Harris, K. S., Kaas, Q., Poon, S., Yap, K., Jia, H., Guarino, R., Chan, L. Y., Durek, T., Anderson, M. A. and Craik, D. J. (2018) 'Molecular basis for the production of cyclic peptides by plant asparaginyl endopeptidases.', *Nature Communications*, 9(1), p. 2411. doi: 10.1038/s41467-018-04669-9.

James, A. M., Haywood, J., Leroux, J., Ignasiak, K., Elliott, A. G., Schmidberger, J. W., Fisher, M. F., Nonis, S. G., Fenske, R., Bond, C. S. and Mylne, J. S. (2019) 'The macrocyclizing protease butelase 1 remains autocatalytic and reveals the structural basis for ligase activity', *The Plant Journal*, p. tpj.14293. doi: 10.1111/tpj.14293.

James, A. M., Haywood, J. and Mylne, J. S. (2018) 'Macrocyclization by asparaginyl endopeptidases', *New Phytologist*, 218(3), pp. 923–928. doi: 10.1111/nph.14511.

Jashni, M. K., Mehrabi, R., Collemare, J., Mesarich, C. H. and de Wit, P. J. G. M. (2015) 'The battle in the apoplast: further insights into the roles of proteases and their inhibitors in plant–pathogen interactions.', *Frontiers in Plant Science*, 6, p. 584. doi: 10.3389/fpls.2015.00584.

Jing, M., Guo, B., Li, H., Yang, B., Wang, H., Kong, G., Zhao, Y., Xu, H., Wang, Y., Ye, W., Dong, S., Qiao, Y., Tyler, B. M., Ma, W. and Wang, Y. (2016) 'A Phytophthora sojae effector suppresses endoplasmic reticulum stress-mediated immunity by stabilizing plant binding immunoglobulin proteins', *Nature Communications*, 7(1), pp. 1–17. doi: 10.1038/ncomms11685.

Johnson, C. E. and Kornbluth, S. (2008) 'Caspase cleavage is not for everyone', *Cell*, 134(5), pp. 720–721. doi: 10.1016/j.cell.2008.08.019.

Jones, A. M. (2001) 'Programmed cell death in development and defense', *Plant Physiology*, 125(1), pp. 94–7. doi: 10.1104/PP.125.1.94.

Jones, J. D. G. and Dangl, J. L. (2006) 'The plant immune system.', *Nature*, 444(7117), pp. 323–329. doi: 10.1038/nature05286.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R. and Hunter, S. (2014) 'InterProScan 5: Genome-scale protein function classification', *Bioinformatics*, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031.

Jung, R., Paul Scott, M., Nam, Y.-W. W., Beaman, T. W., Bassüner, R., Saalbach, I., Müntz, K., Nielsen, N. C., Scott, M. P., Nam, Y.-W. W., Beaman, T. W., Bassüner, R., Saalbach, I., Müntz, K. and Nielsen, N. C. (1998) 'The role of proteolysis in the processing and assembly of 11S seed globulins', *The Plant Cell*, 10(3), pp. 343–357. doi: 10.1105/tpc.10.3.343.

Jung, R., Saalbach, G., Nielsen, N. C. and Muntz, K. (1993) 'Site-specific limited proteolysis of legumin chloramphenicol acetyl transferase fusions in vitro and in transgenic tobacco seeds', *Journal of Experimental Botany*, 44, pp. 343–349.

Jutras, P. V., Dodds, I. and van der Hoorn, R. AL (2020) 'Proteases of Nicotiana benthamiana: an emerging battle for molecular farming', *Current Opinion in Biotechnology*. Elsevier Ltd, pp. 60–65. doi: 10.1016/j.copbio.2019.10.006.

Jutras, P. V., Grosse-Holz, F., Kaschani, F., Kaiser, M., Michaud, D. and Hoorn, R. A. L. (2019) 'Activity-based proteomics reveals nine target proteases for the recombinant protein-stabilizing inhibitor *Sl* <scp>CYS</scp> 8 in *Nicotiana benthamiana*', *Plant Biotechnology Journal*, 17(8), p. pbi.13092. doi: 10.1111/pbi.13092.

Kabbage, M., Kessens, R., Bartholomay, L. C. and Williams, B. (2017) 'The life and death of a plant cell', *Annual Review of Plant Biology*, 68(1), pp. 375–404. doi: 10.1146/annurev-arplant-043015-111655.

Kadono, T., Tran, D., Errakhi, R., Hiramatsu, T., Meimoun, P., Briand, J., Iwaya-Inoue, M., Kawano, T. and Bouteau, F. (2010) 'Increased anion channel activity is an unavoidable event in ozone-induced programmed cell death', *PLoS ONE*, 5(10), p. e13373. doi: 10.1371/journal.pone.0013373.

Kalinski, A., Melroy, D. L., Dwivedi, R. S. and Herman, E. M. (1992) 'A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation.', *The Journal of Biological Chemistry*, 267(17), pp. 12068–76.

Kamoun, S., van West P, P. van, Vleeshouwers, V. G. A. A., de Groot KE, K. E. de and Govers, F. (1998) 'Resistance of Nicotiana benthamiana to Phytophthora infestans is mediated by the recognition of the elicitor protein INF1.', *The Plant Cell*, 10(9). doi: 10.1105/TPC.10.9.1413.

Kariya, K., Demiral, T., Sasaki, T., Tsuchiya, Y., Turkan, I., Sano, T., Hasezawa, S. and Yamamoto, Y. (2013) 'A novel mechanism of aluminium-induced cell death involving vacuolar processing enzyme and vacuolar collapse in tobacco cell line BY-2', *Journal of Inorganic Biochemistry*, 128, pp. 196–201. doi: 10.1016/J.JINORGBIO.2013.07.001.

Kaschani, F., Gu, C., Niessen, S., Hoover, H., Cravatt, B. F. and Van Der Hoorn, R. A. L. (2009) 'Diversity of serine hydrolase activities of unchallenged and Botrytis-infected Arabidopsis thaliana', *Molecular and Cellular Proteomics*, 8(5), pp. 1082–1093. doi: 10.1074/mcp.M800494-MCP200.

Kaschani, F., Shabab, M., Bozkurt, T., Shindo, T., Schornack, S., Gu, C., Ilyas, M., Win, J., Kamoun, S. and van der Hoorn, R. A. L. (2010) 'An effector-targeted protease contributes to defense against Phytophthora infestans and is under diversifying selection in natural hosts.', *Plant Physiology*, 154(4), pp. 1794–1804. doi: 10.1104/pp.110.158030.

Kembhavi, A. A., Buttle, D. J., Knight, C. G. and Barrett, A. J. (1993) 'The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays', *Archives of Biochemistry and Biophysics*, 303(2), pp. 208–213. doi: 10.1006/abbi.1993.1274.

Kim, M., Ahn, J.-W., Jin, U.-H., Choi, D., Paek, K.-H. and Pai, H.-S. (2003) 'Activation of the programmed cell death pathway by inhibition of proteasome function in plants.', *The Journal of Biological Chemistry*, 278(21), pp. 19406–15. doi: 10.1074/jbc.M210539200.

Kim, Y., Wang, M., Bai, Y., Zeng, Z., Guo, F., Han, N., Bian, H., Wang, J., Pan, J. and

Zhu, M. (2014) 'Bcl-2 suppresses activation of VPEs by inhibiting cytosolic Ca2+ level with elevated K+ efflux in NaCl-induced PCD in rice', *Plant Physiology and Biochemistry*, 80, pp. 168–175. doi: 10.1016/j.plaphy.2014.04.002.

Kinoshita, T., Nishimura, M. and Hara-Nishimura, I. (1995) 'Homologues of a vacuolar processing enzyme that are expressed in different organs in Arabidopsis thaliana.', *Plant Molecular Biology*, 29(1), pp. 81–89. doi: 10.1007/BF00019120.

Kinoshita, T., Yamada, K., Hiraiwa, N., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1999) 'Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions', *Plant Journal*, 19(1), pp. 43–53. doi: 10.1046/j.1365-313X.1999.00497.x.

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. (2010) 'Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products', *Nature Biotechnology*, 28(3), pp. 281–288. doi: 10.1038/nbt.1611.

Köhler, D., Montandon, C., Hause, G., Majovsky, P., Kessler, F., Baginsky, S. and Agne, B. (2015) 'Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts.', *Plant Physiology*, 167(3), pp. 972–90. doi: 10.1104/pp.114.255562.

Kørner, C. J., Du, X., Vollmer, M. E. and Pajerowska-Mukhtar, K. M. (2015) 'Endoplasmic reticulum stress signaling in plant lmmunity- at the crossroad of life and death', *International Journal of Molecular Sciences*, 16(11), pp. 26582–26598. doi: 10.3390/ijms161125964.

Kourelis, J., Kaschani, F., Grosse-Holz, F. M., Homma, F., Kaiser, M. and Hoorn, R. A. L. van der (2018) 'Re-annotated Nicotiana benthamiana gene models for enhanced proteomics and reverse genetics', *bioRxiv*, 9, p. 373506. doi: 10.1101/373506.

Kourelis, J., Kaschani, F., Grosse-Holz, F. M., Homma, F., Kaiser, M. and Van Der Hoorn, R. A. L. (2019) 'A homology-guided, genome-based proteome for improved proteomics in the alloploid Nicotiana benthamiana', *BMC Genomics*, 20(1), p. 722. doi: 10.1186/s12864-019-6058-6.

Kovács, J. (2017) *Comparative study of salt stress-induced physiological and molecular responses in tomato (Solanum lycopersicum L.)*. University of Szeged.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. and Battistuzzi, F. U. (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms', *Molecular Biology and Evolution*. Edited by F. U. Battistuzzi, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kurepa, J. and Smalle, J. A. (2008) 'Structure, function and regulation of plant proteasomes.', *Biochimie*, 90(2), pp. 324–335. doi: 10.1016/j.biochi.2007.07.019.

Kuroyanagi, M., Nishimura, M. and Hara-Nishimura, I. (2002) 'Activation of Arabidopsis vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide', *Plant and Cell Physiology*, 43(2), pp. 143–151. doi: 10.1093/pcp/pcf035.

Kuroyanagi, M., Yamada, K., Hatsugai, N., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2005) 'Vacuolar processing enzyme is essential for mycotoxin-induced cell death in Arabidopsis thaliana', *The Journal of Biological Chemistry*, (38), pp. 32914–20. doi: 10.1074/jbc.M504476200.

Labudda, M., Różańska, E., Prabucka, B., Muszyńska, E., Marecka, D., Kozak, M., Dababat, A. A. and Sobczak, M. (2020) 'Activity profiling of barley vacuolar processing enzymes provides new insights into the plant and cyst nematode interaction', *Molecular Plant Pathology*, 21(1), pp. 38–52. doi: 10.1111/mpp.12878.

Lallemand, J., Bouché, F., Desiron, C., Stautemas, J., Esteves, F. de L., Périlleux, C. and Tocquin, P. (2015) 'Extracellular peptidase hunting for improvement of protein production in plant cells and roots', *Frontiers in Plant Science*, 6(FEB). doi: 10.3389/fpls.2015.00037.

Lampl, N., Alkan, N., Davydov, O. and Fluhr, R. (2013) 'Set-point control of RD21 protease activity by AtSerpin1 controls cell death in Arabidopsis.', *The Plant Journal*, 74(3), pp. 498–510. doi: 10.1111/tpj.12141.

Lange, P. F., Huesgen, P. F., Nguyen, K. and Overall, C. M. (2014) 'Annotating N termini for the human proteome project: N termini and Nα-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome', *Journal of Proteome Research*, 13(4), pp. 2028–2044. doi: 10.1021/pr401191w.

Lee, C. R., Park, Y. H., Min, H., Kim, Y. R. and Seok, Y. J. (2019) 'Determination of protein phosphorylation by polyacrylamide gel electrophoresis', *Journal of Microbiology*, 57(2), pp. 93–100. doi: 10.1007/s12275-019-9021-y.

Lema Asqui, S., Vercammen, D., Serrano, I., Valls, M., Rivas, S., van Breusegem, F., Conlon, F. L., Dangl, J. L. and Coll, N. S. (2017) 'AtSERPIN1 is an inhibitor of the metacaspase AtMC1-mediated cell death and autocatalytic processing *in planta*.', *New Phytologist*. doi: 10.1111/nph.14446.

Letunic, I. and Bork, P. (2019) 'Interactive Tree Of Life (iTOL) v4: recent updates and new developments', *Nucleic Acids Research*, 47(W1), pp. W256–W259. doi: 10.1093/nar/gkz239.

Li, B., Meng, X., Shan, L. and He, P. (2016a) 'Transcriptional regulation of pattern-triggered immunity in plants.', *Cell Host & Microbe*, 19(5), pp. 641–650. doi: 10.1016/j.chom.2016.04.011.

Li, C., Wu, H.-M. and Cheung, A. Y. (2016b) 'FERONIA and her pals: Functions and mechanisms.', *Plant Physiology*, 171(4), pp. 2379–92. doi: 10.1104/pp.16.00667.

Li, C., Yeh, F.-L., Cheung, A. Y., Duan, Q., Kita, D., Liu, M.-C., Maman, J., Luu, E. J., Wu, B. W., Gates, L., Jalal, M., Kwong, A., Carpenter, H. and Wu, H.-M. (2015) 'Glycosylphosphatidylinositol-anchored proteins as chaperones and co-receptors for FERONIA receptor kinase signaling in Arabidopsis', *eLife*, 4(JUNE), pp. 1–21. doi: 10.7554/eLife.06587.

Li, D. N., Matthews, S. P., Antoniou, A. N., Mazzeo, D. and Watts, C. (2003) 'Multistep autoactivation of asparaginyl endopeptidase in vitro and in vivo.', *The Journal of Biological Chemistry*, 278(40), pp. 38980–90. doi: 10.1074/jbc.M305930200.

Li, F., Leier, A., Liu, Q., Wang, Y., Xiang, D., Akutsu, T., Webb, G. I., Smith, A. I., Marquez-Lago, T., Li, J. and Song, J. (2020) 'Procleave: Predicting protease-specific substrate cleavage sites by combining sequence and structural information', *Genomics, Proteomics and Bioinformatics*, 18(1), pp. 52–64. doi: 10.1016/j.gpb.2019.08.002.

Li, Z., Yue, H. and Xing, D. (2012) 'MAP Kinase 6-mediated activation of vacuolar processing enzyme modulates heat shock-induced programmed cell death in Arabidopsis', *New Phytologist*, 195(1), pp. 85–96. doi: 10.1111/j.1469-8137.2012.04131.x.

Linnestad, C., Doan, D. N., Brown, R. C., Lemmon, B. E., Meyer, D. J., Jung, R. and Olsen, O. A. (1998) 'Nucellain, a barley homolog of the dicot vacuolar-processing protease, is localized in nucellar cell walls.', *Plant physiology*, 118(4), pp. 1169–80. doi: 10.1104/pp.118.4.1169.

Liu, C. and Moschou, P. N. (2018) 'Cutting in the middleman: hidden substrates at the interface between proteases and plant development', *New Phytologist*, 218(3), pp. 916–922. doi: 10.1111/nph.14501.

Liu, H., Hu, M., Wang, Q., Cheng, L. and Zhang, Z. (2018) 'Role of papain-like cysteine proteases in plant development', *Frontiers in Plant Science*, 871, p. 1717. doi: 10.3389/fpls.2018.01717.

Liu, J.-X., Srivastava, R., Che, P. and Howell, S. H. (2007) 'Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling.', *The Plant Journal*, 51(5), pp. 897–909. doi: 10.1111/j.1365-313X.2007.03195.x.

Liu, W. T., Ng, J., Meluzzi, D., Bandeira, N., Gutierrez, M., Simmons, T. L., Schultz, A. W., Linington, R. G., Moore, B. S., Gerwick, W. H., Pevzner, P. A. and Dorrestein, P. C. (2009) 'Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides', *Analytical Chemistry*, 81(11), pp. 4200–4209. doi: 10.1021/ac900114t.

Liu, Y., Patricelli, M. P. and Cravatt, B. F. (1999) 'Activity-based protein profiling: The serine hydrolases', *Proceedings of the National Academy of Sciences*, 96(26), pp. 14694–14699. doi: 10.1073/pnas.96.26.14694.

Lobet, G., Pagès, L. and Draye, X. (2011) 'A novel image-analysis toolbox enabling quantitative analysis of root system architecture', *Plant Physiology*, 157(1), pp. 29–39. doi: 10.1104/pp.111.179895.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995) 'Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution.', *Science*, 268(5210), pp. 533–9.

Lozano-Torres, J. L., Wilbers, R. H. P., Gawronski, P., Boshoven, J. C., Finkers-Tomczak, A., Cordewener, J. H. G., America, A. H. P., Overmars, H. A., Van 't Klooster, J. W., Baranowski, L., Sobczak, M., Ilyas, M., van der Hoorn, R. A. L., Schots, A., de Wit, P. J. G. M., Bakker, J., Goverse, A. and Smant, G. (2012) 'Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode', *Proceedings of the National Academy of Sciences*, 109(25), pp. 10119–10124. doi: 10.1073/pnas.1202867109.

Lu, H., Chandrasekar, B., Oeljeklaus, J., Misas-Villamil, J. C., Wang, Z., Shindo, T., Bogyo, M., Kaiser, M. and van der Hoorn, R. A. L. (2015) 'Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination.', *Plant Physiology*, 168(4), pp. 1462–75. doi: 10.1104/pp.114.254466.

Lu, W., Deng, M., Guo, F., Wang, M., Zeng, Z., Han, N., Yang, Y., Zhu, M. and Bian, H. (2016) 'Suppression of OsVPE3 enhances salt tolerance by attenuating vacuole rupture during programmed cell death and affects stomata development in rice', *Rice*, 9(1), p. 65. doi: 10.1186/s12284-016-0138-x.

Lycett, G., Croy, R., Shirsat, A. and Boulter, D. (1984) 'Complete nucleotide sequence of a legumin gene from pea (Pisum sativum L.) ', *Nucleic Acids Research*, 12(11), pp. 4493–4506.

Malamy, J., Carr, J. P., Klessig, D. F. and Raskin, I. (1990) 'Salicylic acid: a likely

endogenous signal in the resistance response of tobacco to viral infection.', *Science*, 250(4983), pp. 1002–1004. doi: 10.1126/science.250.4983.1002.

Marino, G., Eckhard, U. and Overall, C. M. (2015) 'Protein termini and their modifications revealed by positional proteomics', *ACS Chemical Biology*, 10(8), pp. 1754–1764. doi: 10.1021/acschembio.5b00189.

Martinez, M., Diaz-Mendoza, M., Carrillo, L. and Diaz, I. (2007) 'Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases.', *FEBS letters*, 581(16), pp. 2914–8. doi: 10.1016/j.febslet.2007.05.042.

Martinière, A., Bassil, E., Jublanc, E., Alcon, C., Reguera, M., Sentenac, H., Blumwald, E. and Paris, N. (2013) 'In vivo intracellular pH measurements in tobacco and arabidopsis reveal an unexpected pH gradient in the endomembrane system', *Plant Cell*, 25(10), pp. 4028–4043. doi: 10.1105/tpc.113.116897.

Masachis, S., Segorbe, D., Turrà, D., Leon-Ruiz, M., Fürst, U., El Ghalid, M., Leonard, G., López-Berges, M. S., Richards, T. A., Felix, G. and Di Pietro, A. (2016) 'A fungal pathogen secretes plant alkalinizing peptides to increase infection.', *Nature Microbiology*, 1(6), p. 16043. doi: 10.1038/nmicrobiol.2016.43.

Mase, K., Mizuno, T., Ishihama, N., Fujii, T., Mori, H., Kodama, M. and Yoshioka, H. (2012) 'Ethylene signaling pathway and MAPK cascades are required for AAL toxin-induced programmed cell death.', *Molecular Plant-Microbe Interactions*, 25(8), pp. 1015–25. doi: 10.1094/MPMI-02-12-0036-R.

Matarasso, N., Schuster, S. and Avni, A. (2005) 'A novel plant cysteine protease has a dual function as a regulator of 1-aminocyclopropane-1-carboxylic acid synthase gene expression', *Plant Cell*, 17(4), pp. 1205–1216. doi: 10.1105/tpc.105.030775.

McIlwain, D. R., Berger, T. and Mak, T. W. (2013) 'Caspase functions in cell death and disease.', *Cold Spring Harbor Perspectives in Biology*, 5(4), p. a008656. doi: 10.1101/cshperspect.a008656.

McLellan, H., Gilroy, E. M., Yun, B.-W., Birch, P. R. J. and Loake, G. J. (2009) 'Functional redundancy in the Arabidopsis Cathepsin B gene family contributes to basal defence, the hypersensitive response and senescence', *The New Phytologist*, 183(2), pp. 408–18. doi: 10.1111/j.1469-8137.2009.02865.x.

Mechtersheimer, S., Gutwein, P., Agmon-Levin, N., Stoeck, A., Oleszewski, M., Riedle, S., Postina, R., Fahrenholz, F., Fogel, M., Lemmon, V. and Altevogt, P. (2001) 'Ectodomain shedding of L1 adhesion molecule promotes cell migration by autocrine binding to integrins', *Journal of Cell Biology*, 155(4), pp. 661–673. doi: 10.1083/jcb.200101099.

Mendes, G. C., Reis, P. A. B., Calil, I. P., Carvalho, H. H., Aragão, F. J. L. and Fontes, E. P. B. (2013) 'GmNAC30 and GmNAC81 integrate the endoplasmic reticulum stress- and osmotic stress-induced cell death responses through a vacuolar processing enzyme', *Proceedings of the National Academy of Sciences of the United States of America*, 110(48), pp. 19627–19632. doi: 10.1073/pnas.1311729110.

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J. J., Cox, J., Horning, S., Mann, M. and Makarov, A. (2012) 'Ultra high resolution linear ion trap orbitrap mass spectrometer (orbitrap

elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes', *Molecular and Cellular Proteomics*, 11(3). doi: 10.1074/mcp.O111.013698.

Min, W., Dunn, A. J. and Jones, D. H. (1992) 'Non-glycosylated recombinant pro-concanavalin A is active without polypeptide cleavage.', *The EMBO Journal*, 11(4), pp. 1303–7.

Min, W. and Jones, D. H. (1994) 'In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase', *Nature Structural Biology*, 1(8), pp. 502–504. doi: 10.1038/nsb0894-502.

Mino, M., Murata, N., Date, S. and Inoue, M. (2007) 'Cell death in seedlings of the interspecific hybrid of Nicotiana gossei and N. tabacum; possible role of knob-like bodies formed on tonoplast in vacuolar-collapse-mediated cell death', *Plant Cell Reports*, 26(4), pp. 407–419. doi: 10.1007/s00299-006-0261-z.

Misas-Villamil, J. C. (2010) *Activity-profiling of Vacuolar Processing Enzymes and the proteasome during plant-pathogen interactions*. University of Cologne.

Misas-Villamil, J. C., Toenges, G., Kolodziejek, I., Sadaghiani, A. M., Kaschani, F., Colby, T., Bogyo, M. and van der Hoorn, R. A. L. (2013) 'Activity profiling of vacuolar processing enzymes reveals a role for VPE during oomycete infection.', *The Plant Journal*, 73(4), pp. 689–700. doi: 10.1111/tpj.12062.

Monsalve, R. I., Villalba, M., López-Otín, C. and Rodríguez, R. (1991) 'Structural analysis of the small chain of the 2S albumin, napin nIII, from rapeseed. Chemical and spectroscopic evidence of an intramolecular bond formation', *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*, 1078(2), pp. 265–272. doi: 10.1016/0167-4838(91)90568-K.

Morimoto, K. and van der Hoorn, R. A. L. (2016) 'The increasing impact of activity-based protein profiling in plant science', *Plant and Cell Physiology*, 57(3), pp. 446–461. doi: 10.1093/pcp/pcw003.

Moss, C. X., Matthews, S. P., Lamont, D. J. and Watts, C. (2005) 'Asparagine deamidation perturbs antigen presentation on class II major histocompatibility complex molecules', *Journal of Biological Chemistry*, 280(18), pp. 18498–18503. doi: 10.1074/jbc.M501241200.

Mukaiyama, H., Iwaki, T., Idiris, A. and Takegawa, K. (2011) 'Processing and maturation of carboxypeptidase Y and alkaline phosphatase in Schizosaccharomyces pombe', *Applied Microbiology and Biotechnology*, 90(1), pp. 203–213. doi: 10.1007/s00253-010-3031-3.

Müller, S., Hoege, C., Pyrowolakis, G. and Jentsch, S. (2001) 'Sumo, ubiquitin's mysterious cousin', *Nature Reviews Molecular Cell Biology*, 2(3), pp. 202–210. doi: 10.1038/35056591.

Müntz, K., Blattner, F. R. and Shutov, A. D. (2002) 'Legumains - A family of asparagine-specific cysteine endopeptidases involved in propolypeptide processing and protein breakdown in plants', *Journal of Plant Physiology*, 159(12), pp. 1281–1293. doi: 10.1078/0176-1617-00853.

Mur, L. A. J., Kenton, P., Lloyd, A. J., Ougham, H. and Prats, E. (2008) 'The hypersensitive response; the centenary is upon us but how much do we know?', *Journal of Experimental Botany*, 59(3), pp. 501–520. doi: 10.1093/jxb/erm239.

Murata, S., Yashiroda, H. and Tanaka, K. (2009) 'Molecular mechanisms of proteasome assembly.', *Nature Reviews Molecular Cell Biology*, 10(2), pp. 104–115. doi: 10.1038/nrm2630.

Mylne, J. S., Chan, L. Y., Chanson, A. H., Daly, N. L., Schaefer, H., Bailey, T. L., Nguyencong, P., Cascales, L. and Craik, D. J. (2012) 'Cyclic peptides arising by evolutionary

parallelism via asparaginyl-endopeptidase-mediated biosynthesis', *Plant Cell*, 24(7), pp. 2765–2778. doi: 10.1105/tpc.112.099085.

Mylne, J. S., Colgrave, M. L., Daly, N. L., Chanson, A. H., Elliott, A. G., McCallum, E. J., Jones, A. and Craik, D. J. (2011) 'Albumins and their processing machinery are hijacked for cyclic peptides in sunflower', *Nature Chemical Biology*, 7(5), pp. 257–259. doi: 10.1038/nchembio.542.

Mylne, J. S., Hara-Nishimura, I. and Rosengren, K. J. (2014) 'Seed storage albumins: biosynthesis, trafficking and structures', *Functional Plant Biology*, 41(7), p. 671. doi: 10.1071/FP14035.

Naim, F., Nakasugi, K., Crowhurst, R. N., Hilario, E., Zwart, A. B., Hellens, R. P., Taylor, J. M., Waterhouse, P. M. and Wood, C. C. (2012) 'Advanced engineering of lipid metabolism in Nicotiana benthamiana using a draft genome and the V2 viral silencing-suppressor protein', *PLoS ONE*. Edited by L. Herrera-Estrella, 7(12), p. e52717. doi: 10.1371/journal.pone.0052717.

Nakaune, S., Yamada, K., Kondo, M., Kato, T., Tabata, S., Nishimura, M. and Hara-Nishimura, I. (2005) 'A vacuolar processing enzyme, deltaVPE, is involved in seed coat formation at the early stage of seed development.', *The Plant Cell*, 17(3), pp. 876–87. doi: 10.1105/tpc.104.026872.

Negoro, T., Momma, T. and Fukazawa, C. (1985) 'A cDNA clone encoding a glycinin A1a subunit precursor of soybean.', *Nucleic Acids Research*, 13(18), pp. 6719–6731.

Neuhaus, J. M., Sticher, L., Meins, F. and Boller, T. (1991) 'A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole', *Proceedings of the National Academy of Sciences of the United States of America*, 88(22), pp. 10362–10366. doi: 10.1073/pnas.88.22.10362.

Nguyen, G. K. T., Kam, A., Loo, S., Jansson, A. E., Pan, L. X. and Tam, J. P. (2015) 'Butelase 1: A versatile ligase for peptide and protein macrocyclization', *Journal of the American Chemical Society*, 137(49), pp. 15398–15401. doi: 10.1021/jacs.5b11014.

Nguyen, G. K. T. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y. and Tam, J. P. (2014) 'Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis', *Nature Chemical Biology*, 10(9), pp. 732–738. doi: 10.1038/nchembio.1586.

Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997) 'Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites', *Protein Engineering*, 10(1), pp. 1–6. doi: 10.1093/protein/10.1.1.

Niemer, M., Mehofer, U., Verdianz, M., Porodko, A., Schähs, P., Kracher, D., Lenarcic, B., Novinec, M. and Mach, L. (2016) 'Nicotiana benthamiana cathepsin B displays distinct enzymatic features which differ from its human relative and aleurain-like protease.', *Biochimie*, 122, pp. 119–125. doi: 10.1016/j.biochi.2015.06.017.

Norkunas, K., Harding, R., Dale, J. and Dugdale, B. (2018) 'Improving agroinfiltration-based transient gene expression in Nicotiana benthamiana', *Plant Methods*, 14(1), p. 71. doi: 10.1186/s13007-018-0343-2.

Oh, C. J., Lee, H., Ho, B. K. and Chung, S. A. (2004) 'Isolation and characterization of a root nodule-specific cysteine proteinase cDNA from soybean', *Journal of Plant Biology*, 47(3), pp. 216–220. doi: 10.1007/bf03030511.

Okamoto, T. and Minamikawa, T. (1999) 'Molecular cloning and characterization of Vigna mungo processing enzyme 1 (VmPE-1), an asparaginyl endopeptidase possibly involved in post-translational processing of a vacuolar cysteine endopeptidase (SH-EP).', *Plant Molecular Biology*, 39(1), pp. 63–73. doi: 10.1023/a:1006170518002.

Okamoto, T., Yuki, A., Mitsuhashi, N., Minamikawa, T. and Mimamikawa, T. (1999) 'Asparaginyl endopeptidase (VmPE-1) and autocatalytic processing synergistically activate the vacuolar cysteine proteinase (SH-EP).', *European Journal of Biochemistry*, 264(1), pp. 223–32.

Olsen, J. V., de Godoy, L. M. F., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S. and Mann, M. (2005) 'Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap', *Molecular and Cellular Proteomics*, 4(12), pp. 2010–2021. doi: 10.1074/mcp.T500030-MCP200.

Ondzighi, C. A., Christopher, D. A., Cho, E. J., Chang, S.-C. and Staehelin, L. A. (2008) 'Arabidopsis protein disulfide isomerase-5 inhibits cysteine proteases during trafficking to vacuoles before programmed cell death of the endothelium in developing seeds.', *The Plant Cell*, 20(8), pp. 2205–20. doi: 10.1105/tpc.108.058339.

Otegui, M. S., Herder, R., Schulze, J., Jung, R. and Staehelin, L. A. (2006) 'The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies', *Plant Cell*, 18(10), pp. 2567–2581. doi: 10.1105/tpc.106.040931.

Paireder, M., Tholen, S., Porodko, A., Biniossek, M. L., Mayer, B., Novinec, M., Schilling, O. and Mach, L. (2017) 'The papain-like cysteine proteinases NbCysP6 and NbCysP7 are highly processive enzymes with substrate specificities complementary to Nicotiana benthamiana cathepsin B', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1865(4), pp. 444–452. doi: 10.1016/j.bbapap.2017.02.007.

Pajerowska-Mukhtar, K. and Dong, X. (2009) 'A kiss of death--proteasome-mediated membrane fusion and programmed cell death in plant defense against bacterial infection.', *Genes & Development*, 23(21), pp. 2449–54. doi: 10.1101/gad.1861609.

Park, C. J., An, J. M., Shin, Y. C., Kim, K. J., Lee, B. J. and Paek, K. H. (2004) 'Molecular characterization of pepper germin-like protein as the novel PR-16 family of pathogenesis-related proteins isolated during the resistance response to viral and bacterial infection', *Planta*, 219(5), pp. 797–806. doi: 10.1007/s00425-004-1290-x.

Pattison, R. J. and Amtmann, A. (2009) 'N-glycan production in the endoplasmic reticulum of plants', *Trends in Plant Science*. Trends Plant Sci, pp. 92–99. doi: 10.1016/j.tplants.2008.11.008.

Paulus, J. K. and Van Der Hoorn, R. A. L. (2019) 'Do proteolytic cascades exist in plants?', *Journal of Experimental Botany*, 70(7), pp. 1997–2002. doi: 10.1093/jxb/erz016.

Paulus, J. K., Kourelis, J., Ramasubramanian, S., Homma, F., Godson, A., Hörger, A. C., Hong, T. N., Krahn, D., Ossorio Carballo, L., Wang, S., Win, J., Smoker, M., Kamoun, S., Dong, S. and van der Hoorn, R. A. L. (2020) 'Extracellular proteolytic cascade in tomato activates immune protease Rcr3', *Proceedings of the National Academy of Sciences*, 117(29), pp. 17409–17417. doi: 10.1073/pnas.1921101117.

Pei, Y., Li, X., Zhu, Y., Ge, X., Sun, Y., Liu, N., Jia, Y., Li, F. and Hou, Y. (2019) 'GhABP19, a novel germin-like protein from Gossypium hirsutum, plays an important role in the regulation of resistance to Verticillium and Fusarium wilt pathogens', *Frontiers in Plant Science*, 10,

p. 583. doi: 10.3389/fpls.2019.00583.

Perrar, A., Dissmeyer, N. and Huesgen, P. F. (2019) 'New beginnings and new ends: methods for large-scale characterization of protein termini and their use in plant biology', *Journal of Experimental Botany*, 70(7), pp. 2021–2038. doi: 10.1093/jxb/erz104.

Pierre, O., Hopkins, J., Combier, M., Baldacci, F., Engler, G., Brouquisse, R., Hérouart, D. and Boncompagni, E. (2014) 'Involvement of papain and legumain proteinase in the senescence process of *Medicago truncatula* nodules', *New Phytologist*, 202(3), pp. 849–863. doi: 10.1111/nph.12717.

Pillay, P., Kunert, K. J., van Wyk, S., Makgopa, M. E., Cullis, C. A. and Vorster, B. J. (2016) 'Agroinfiltration contributes to VP1 recombinant protein degradation', *Bioengineered*, 7(6), pp. 459–477. doi: 10.1080/21655979.2016.1208868.

Ponstein, A. S., Bres-Vloemans, S. A., Sela-Buurlage, M. B., van den Elzen, P. J. M., Melchers, L. S. and Cornelissen, B. J. C. (1994) 'A novel pathogen- and wound-inducible tobacco (Nicotiana tabacum) protein with antifungal activity', *Plant Physiology*, 104(1), pp. 109–118. doi: 10.1104/pp.104.1.109.

Poon, S., Harris, K. S., Jackson, M. A., McCorkelle, O. C., Gilding, E. K., Durek, T., van der Weerden, N. L., Craik, D. J. and Anderson, M. A. (2018) 'Co-expression of a cyclizing asparaginyl endopeptidase enables efficient production of cyclic peptides in planta', *Journal of Experimental Botany*, 69(3), pp. 633–641. doi: 10.1093/jxb/erx422.

Poret, M., Chandrasekar, B., van der Hoorn, R. A. L., Déchaumet, S., Bouchereau, A., Kim, T. H., Lee, B. R., Macquart, F., Hara-Nishimura, I. and Avice, J. C. (2019) 'A genotypic comparison reveals that the improvement in nitrogen remobilization efficiency in oilseed rape leaves is related to specific patterns of senescence-associated protease activities and phytohormones', *Frontiers in Plant Science*, 10, p. 46. doi: 10.3389/fpls.2019.00046.

del Pozo, O. and Lam, E. (1998) 'Caspases and programmed cell death in the hypersensitive response of plants to pathogens.', *Current Biology*, 8(20), pp. 1129–1132. doi: 10.1016/S0960-9822(98)70469-5.

Praekelt, U. M., McKee, R. A. and Smith, H. (1988) 'Molecular analysis of actinidin, the cysteine proteinase of Actinidia chinensis', *Plant Molecular Biology*, 10(3), pp. 193–202. doi: 10.1007/BF00027396.

Prasad, B. D., Creissen, G., Lamb, C. and Chattoo, B. B. (2009) 'Overexpression of rice (Oryza sativa L.) OsCDR1 leads to constitutive activation of defense responses in rice and Arabidopsis.', *Molecular Plant-Microbe Interactions*, 22(12), pp. 1635–1644. doi: 10.1094/MPMI-22-12-1635.

Prudova, A., Serrano, K., Eckhard, U., Fortelny, N., Devine, D. V. and Overall, C. M. (2014) 'TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage', *Blood*, 124(26), pp. e49–e60. doi: 10.1182/blood-2014-04-569640.

Pružinská, A., Shindo, T., Niessen, S., Kaschani, F., Tóth, R., Millar, A. H. and van der Hoorn, R. A. L. (2017) 'Major Cys protease activities are not essential for senescence in individually darkened Arabidopsis leaves.', *BMC Plant Biology*, 17(1), p. 4. doi: 110.1186/s12870-016-0955-5.

Puente, X. S., Sánchez, L. M., Overall, C. M. and López-Otín, C. (2003) 'Human and mouse proteases: A comparative genomic approach', *Nature Reviews Genetics*, 4(7), pp. 544–558. doi:

10.1038/nrg1111.

Qiang, X., Zechmann, B., Reitz, M. U., Kogel, K. H. and Schäfer, P. (2012) 'The mutualistic fungus Piriformospora indica colonizes Arabidopsis roots by inducing an endoplasmic reticulum stress-triggered caspase-dependent cell death', *Plant Cell*, 24(2), pp. 794–809. doi: 10.1105/tpc.111.093260.

Quain, M. D., Makgopa, M. E., Cooper, J. W., Kunert, K. J. and Foyer, C. H. (2015) 'Ectopic phytocystatin expression increases nodule numbers and influences the responses of soybean (Glycine max) to nitrogen deficiency', *Phytochemistry*, 112(1), pp. 179–187. doi: 10.1016/j.phytochem.2014.12.027.

Radchuk, V., Tran, V., Radchuk, R., Diaz-Mendoza, M., Weier, D., Fuchs, J., Riewe, D., Hensel, G., Kumlehn, J., Munz, E., Heinzel, N., Rolletschek, H., Martinez, M. and Borisjuk, L. (2018) 'Vacuolar Processing Enzyme 4 contributes to maternal control of grain size in barley by executing programmed cell death in the pericarp', *New Phytologist*, 218(3), pp. 1127–1142. doi: 10.1111/nph.14729.

Ramírez, V., López, A., Mauch-Mani, B., Gil, M. J. and Vera, P. (2013) 'An extracellular subtilase switch for immune priming in Arabidopsis.', *PLoS Pathogens*, 9(6), p. e1003445. doi: 10.1371/journal.ppat.1003445.

Rantong, G. and Gunawardena, A. H. L. A. N. (2018) 'Vacuolar processing enzymes, AmVPE1 and AmVPE2, as potential executors of ethylene regulated programmed cell death in the lace plant (*Aponogeton madagascariensis*)', *Botany*, 96(4), pp. 235–247. doi: 10.1139/cjb-2017-0184.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nature Protocols*, 2(8), pp. 1896–1906. doi: 10.1038/nprot.2007.261.

Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. and Deber, C. M. (2009) 'Detergent binding explains anomalous SDS-PAGE migration of membrane proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp. 1760–1765. doi: 10.1073/pnas.0813167106.

Rawlings, N. D., Barrett, A. J. and Bateman, A. (2010) 'MEROPS: the peptidase database', *Nucleic Acids Research*, 38(suppl_1), pp. D227–D233. doi: 10.1093/nar/gkp971.

Rawlings, N. D., Barrett, A. J. and Finn, R. (2016) 'Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors.', *Nucleic Acids Research*, 44(D1), pp. D343–D350. doi: 10.1093/nar/gkv1118.

Rehm, F. B. H., Jackson, M. A., De Geyter, E., Yap, K., Gilding, E. K., Durek, T. and Craik, D. J. (2019) 'Papain-like cysteine proteases prepare plant cyclic peptide precursors for cyclization', *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), pp. 7831–7836. doi: 10.1073/pnas.1901807116.

Reichardt, S., Repper, D., Tuzhikov, A. I., Galiullina, R. A., Planas-Marquès, M., Chichkova, N. V., Vartapetian, A. B., Stintzi, A. and Schaller, A. (2018) 'The tomato subtilase family includes several cell death-related proteinases with caspase specificity', *Scientific Reports*, 8(1). doi: 10.1038/s41598-018-28769-0.

Renko, M., Sabotič, J., Mihelič, M., Brzin, J., Kos, J. and Turk, D. (2010) 'Versatile loops in mycocypins inhibit three protease families', *Journal of Biological Chemistry*, 285(1), pp. 308–316.

doi: 10.1074/jbc.M109.043331.

Richau, K. H., Kaschani, F., Verdoes, M., Pansuriya, T. C., Niessen, S., Stüber, K., Colby, T., Overkleeft, H. S., Bogyo, M. and Van der Hoorn, R. A. L. (2012) 'Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics.', *Plant Phsiology*, 158(4). doi: 10.1104/pp.112.194001.

Rinschen, M. M., Hoppe, A. K., Grahammer, F., Kann, M., Völker, L. A., Schurek, E. M., Binz, J., Höhne, M., Demir, F., Malisic, M., Huber, T. B., Kurschat, C., Kizhakkedathu, J. N., Schermer, B., Huesgen, P. F. and Benzing, T. (2017) 'N-degradomic analysis reveals a proteolytic network processing the podocyte cytoskeleton', *Journal of the American Society of Nephrology*, 28(10), pp. 2867–2878. doi: 10.1681/ASN.2016101119.

Roberts, L. M. L. F. I. P. D. J. L. J. M. (1985) 'The primary sequence of Ricinus communis agglutinin. Comparison with ricin.', pp. 15682–15686.

Roberts, M., Tang, S., Stallmann, A., Dangl, J. L. and Bonardi, V. (2013) 'Genetic requirements for signaling from an autoactive plant NB-LRR intracellular innate immune receptor.', *PLoS Genetics*. Edited by J. M. McDowell, 9(4), p. e1003465. doi: 10.1371/journal.pgen.1003465.

Rogers, L. D. and Overall, C. M. (2013) 'Proteolytic post-translational modification of proteins: Proteomic tools and methodology', *Molecular and Cellular Proteomics*, 12(12), pp. 3532–3542. doi: 10.1074/mcp.M113.031310.

Rojo, E., Martín, R., Carter, C., Zouhar, J., Pan, S., Plotnikova, J., Jin, H., Paneque, M., Sánchez-Serrano, J. J., Baker, B., Ausubel, F. M. and Raikhel, N. V. (2004) 'VPE$_\gamma$ exhibits a caspase-like activity that contributes to defense against pathogens.', *Current Biology*, 14(21), pp. 1897–1906. doi: 10.1016/j.cub.2004.09.056.

Rojo, E., Zouhar, J., Carter, C., Kovaleva, V. and Raikhel, N. V. (2003) 'A unique mechanism for protein processing and degradation in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences*, 100(12). doi: 10.1073/pnas.1230987100.

Rooney, H. C. E., Van't Klooster, J. W., van der Hoorn, R. A. L., Joosten, M. H. A. J., Jones, J. D. G. and de Wit, P. J. G. M. (2005) 'Cladosporium Avr2 inhibits tomato Rcr3 protease required for Cf-2-dependent disease resistance.', *Science*, 308(5729), pp. 1783–1786. doi: 10.1126/science.1111404.

Rowland, E., Kim, J., Bhuiyan, N. H. and van Wijk, K. J. (2015) 'The Arabidopsis chloroplast stromal N-terminome: Complexities of amino-terminal protein maturation and stability.', *Plant Physiology*, 169(3), pp. 1881–96. doi: 10.1104/pp.15.01214.

Rudakova, A. S., Cherdivarə, A. M., Wilson, K. A. and Shutov, A. D. (2015) 'Seed storage globulins: origin and evolution of primary and higher order structures', *Biochemistry (Moscow)*, 80(10), pp. 1354–1361. doi: 10.1134/S000629791510017X.

Ruppert, M., Woll, J., Giritch, A., Genady, E., Ma, X. and Stöckigt, J. (2005) 'Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plant-virus expression system', *Planta*, 222(5), pp. 888–898. doi: 10.1007/s00425-005-0031-0.

Sabotič, J. and Kos, J. (2012) 'Microbial and fungal protease inhibitors - Current and potential applications', *Applied Microbiology and Biotechnology*, 93(4), pp. 1351–1375. doi: 10.1007/s00253-011-3834-x.

Sabotič, J., Popovič, T., Puizdar, V. and Brzin, J. (2009) 'Macrocypins, a family of cysteine protease inhibitors from the basidiomycete Macrolepiota procera', *FEBS Journal*, 276(16), pp. 4334–4345. doi: 10.1111/j.1742-4658.2009.07138.x.

Sainsbury, F. (2020) 'Innovation in plant-based transient protein expression for infectious disease prevention and preparedness', *Current Opinion in Biotechnology*, 61, pp. 110–115. doi: 10.1016/j.copbio.2019.11.002.

Salvesen, G. S., Hempel, A. and Coll, N. S. (2016) 'Protease signaling in animal and plant-regulated cell death.', *The FEBS Journal*, 283(14), pp. 2577–2598. doi: 10.1111/febs.13616.

Santino, A., Daminati, M. G., Vitale, A. and Bollini, R. (1992) 'The alpha-amylase inhibitor of bean seed: two-step proteolytic maturation in the protein storage vacuoles of the developing cotyledon', *Physiologia Plantarum*, 85(3), pp. 425–432. doi: 10.1111/j.1399-3054.1992.tb05807.x.

Santos-Silva, L. K., Soares-Costa, A., Gerald, L. T. S., Meneghin, S. P. and Henrique-Silva, F. (2012) 'Recombinant expression and biochemical characterization of sugarcane legumain', *Plant Physiology and Biochemistry*, 57, pp. 181–192. doi: 10.1016/j.plaphy.2012.05.020.

Saska, I., Gillon, A. D., Hatsugai, N., Dietzgen, R. G., Hara-Nishimura, I., Anderson, M. A. and Craik, D. J. (2007) 'An asparaginyl endopeptidase mediates in vivo protein backbone cyclization.', *The Journal of Biological Chemistry*, 282(40), pp. 29721–8. doi: 10.1074/jbc.M705185200.

Savickas, S., Kastl, P. and auf dem Keller, U. (2020) 'Combinatorial degradomics: Precision tools to unveil proteolytic processes in biological systems', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1868(6), p. 140392. doi: 10.1016/j.bbapap.2020.140392.

Schepmoes, A. A., Zhang, Q., Petritis, B. O., Qian, W. J. and Smith, R. D. (2009) 'N-terminal enrichment: Developing a protocol to detect specific proteolytic fragments', *Journal of Biomolecular Techniques*, 20(5), pp. 263–265.

Schilling, O., Huesgen, P. F., Barré, O., Auf dem Keller, U. and Overall, C. M. (2011) 'Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry.', *Nature Protocols*, 6(1), pp. 111–20. doi: 10.1038/nprot.2010.178.

Schoberer, J. and Strasser, R. (2018) 'Plant glyco-biotechnology', *Seminars in Cell & Developmental Biology*, 80, pp. 133–141. doi: 10.1016/j.semcdb.2017.07.005.

Serra, A., Hemu, X., Nguyen, G. K. T., Nguyen, N. T. K., Sze, S. K. and Tam, J. P. (2016) 'A high-throughput peptidomic strategy to decipher the molecular diversity of cyclic cysteine-rich peptides', *Scientific Reports*, 6(1), pp. 1–13. doi: 10.1038/srep23005.

Shabab, M., Shindo, T., Gu, C., Kaschani, F., Pansuriya, T., Chintha, R., Harzen, A., Colby, T., Kamoun, S. and van der Hoorn, R. A. L. (2008) 'Fungal effector protein AVR2 targets diversifying defense-related Cys proteases of tomato.', *The Plant Cell*, 20(4), pp. 1169–1183. doi: 10.1105/tpc.107.056325.

Shalini, S., Dorstyn, L., Dawar, S. and Kumar, S. (2015) 'Old, new and emerging functions of caspases.', *Cell Death & Differentiation*, 22(4), pp. 526–539. doi: 10.1038/cdd.2014.216.

Sharief, F. S. and Li, S. S. (1982) *Amino acid sequence of small and large subunits of seed storage protein from Ricinus communis.*, *Journal of Biological Chemistry*.

Shen, J., Zeng, Y., Zhuang, X., Sun, L., Yao, X., Pimpl, P. and Jiang, L. (2013) 'Organelle pH in the Arabidopsis endomembrane system', *Molecular Plant*, 6(5), pp. 1419–1437. doi: 10.1093/mp/sst079.

Shibata, Y., Kawakita, K. and Takemoto, D. (2010) 'Age-related resistance of Nicotiana benthamiana against hemibiotrophic pathogen Phytophthora infestans requires both ethylene- and salicylic acid–mediated signaling pathways', *Molecular Plant-Microbe Interactions*, 23(9), pp. 1130–1142. doi: 10.1094/MPMI-23-9-1130.

Shimada, T., Yamada, K., Kataoka, M., Nakaune, S., Koumoto, Y., Kuroyanagi, M., Tabata, S., Kato, T., Shinozaki, K., Seki, M., Kobayashi, M., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (2003) 'Vacuolar Processing Enzymes Are Essential for Proper Processing of Seed Storage Proteins in Arabidopsis thaliana', *Journal of Biological Chemistry*, 278(34), pp. 32292–32299. doi: 10.1074/jbc.M305740200.

Shindo, T., Kaschani, F., Yang, F., Kovács, J., Tian, F., Kourelis, J., Hong, T. N., Colby, T., Shabab, M., Chawla, R., Kumari, S., Ilyas, M., Hörger, A. C., Alfano, J. R. and van der Hoorn, R. A. L. (2016) 'Screen of non-annotated small secreted proteins of Pseudomonas syringae reveals a virulence factor that inhibits tomato immune proteases.', *PLoS Pathogens*. Edited by D. Desveaux, 12(9), p. e1005874. doi: 10.1371/journal.ppat.1005874.

Shindo, T., Misas-Villamil, J. C., Hörger, A. C., Song, J. and van der Hoorn, R. A. L. (2012) 'A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14.', *PLoS ONE*. Edited by C.-H. Yang, 7(1), p. e29317. doi: 10.1371/journal.pone.0029317.

Shinshi, H., Wenzler, H., Neuhaus, J.-M., Felix, G., Hofsteenge, J. and Meins, F. (1988) 'Evidence for N- and C-terminal processing of a plant defense-related enzyme: Primary structure of tobacco prepro- -1,3-glucanase', *Proceedings of the National Academy of Sciences*, 85(15), pp. 5541–5545. doi: 10.1073/pnas.85.15.5541.

Shotwell, M. A., Afonso, C., Davies, E., Chesnut, R. S. and Larkins, B. A. (1988) 'Molecular characterization of oat seed globulins', *Plant Physiology*, 87(3), pp. 698–704. doi: 10.1104/pp.87.3.698.

Shutov, A. D., Do, N. L. and Vaintraub, I. A. (1982) 'Purification and partial characterization of protease B from germinating vetch seeds.', *Biokhimiia (Moscow, Russia)*, 47(5), pp. 814–21.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. and Higgins, D. G. (2014) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7(1), pp. 539–539. doi: 10.1038/msb.2011.75.

Simon, G. M., Dix, M. M. and Cravatt, B. F. (2009) 'Comparative assessment of large-scale proteomic studies of apoptotic proteolysis', *ACS Chemical Biology*, 4(6), pp. 401–408. doi: 10.1021/cb900082q.

Soares, A., Ribeiro Carlton, S. M. and Simões, I. (2019) 'Atypical and nucellin-like aspartic proteases: Emerging players in plant developmental processes and stress responses', *Journal of Experimental Botany*. Oxford University Press, pp. 2059–2076. doi: 10.1093/jxb/erz034.

Soh, W. T., Demir, F., Dall, E., Perrar, A., Dahms, S. O., Kuppusamy, M., Brandstetter, H. and Huesgen, P. F. (2020) 'ExteNDing proteome coverage with legumain as

highly specific digestion protease', *Analytical Chemistry*, p. acs.analchem.9b03604. doi: 10.1021/acs.analchem.9b03604.

Song, J., Wang, Y., Li, F., Akutsu, T., Rawlings, N. D., Webb, G. I. and Chou, K. C. (2019) 'IProt-Sub: A comprehensive package for accurately mapping and predicting protease-specific substrates and cleavage sites', *Briefings in Bioinformatics*, 20(2), pp. 638–658. doi: 10.1093/bib/bby028.

Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., van der Hoorn, R. A. L. and Kamoun, S. (2009) 'Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3.', *Proceedings of the National Academy of Sciences*, 106(5), pp. 1654–1659. doi: 10.1073/pnas.0809201106.

Song, J., Yang, F., Xun, M., Xu, L., Tian, X., Zhang, W. and Yang, H. (2020) 'Genome-wide identification and characterization of Vacuolar Processing Enzyme gene family and diverse expression under stress in Apple (Malus × Domestic)', *Frontiers in Plant Science*, 11. doi: 10.3389/fpls.2020.00626.

Sparkes, I. A., Runions, J., Kearns, A. and Hawes, C. (2006) 'Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants', *Nature Protocols*, 1(4), pp. 2019–2025. doi: 10.1038/nprot.2006.286.

Spoel, S. H., Mou, Z., Tada, Y., Spivey, N. W., Genschik, P. and Dong, X. (2009) 'Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity.', *Cell*, 137(5), pp. 860–72. doi: 10.1016/j.cell.2009.03.038.

Srivastava, R., Liu, J.-X., Guo, H., Yin, Y. and Howell, S. H. (2009) 'Regulation and processing of a plant peptide hormone, AtRALF23, in Arabidopsis.', *The Plant Journal*, 59(6), pp. 930–939. doi: 10.1111/j.1365-313X.2009.03926.x.

Stael, S., Van Breusegem, F., Gevaert, K. and Nowack, M. K. (2019) 'Plant proteases and programmed cell death', *Journal of Experimental Botany*, 70(7), pp. 1991–1995. doi: 10.1093/jxb/erz126.

Staswick, P. E., Hermodson, M. A. and Nielsen, N. C. (1984) 'The amino acid sequence of the A2B1a subunit of glycinin.', *The Journal of Biological Chemistry*, 259(21), pp. 13424–30.

Stegmann, M., Monaghan, J., Smakowska-Luzan, E., Rovenich, H., Lehner, A., Holton, N., Belkhadir, Y. and Zipfel, C. (2017) 'The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling.', *Science*, 355(6322), pp. 287–289. doi: 10.1126/science.aal2541.

Sticher, L., Hofsteenge, J., Neuhaus, J. M., Boller, T. and Meins, F. (1993) 'Posttranslational processing of a new class of hydroxyproline-containing proteins: Prolyl hydroxylation and C-terminal cleavage of tobacco (Nicotiana tabacum) vacuolar chitinase', *Plant Physiology*, 101(4), pp. 1239–1247. doi: 10.1104/pp.101.4.1239.

Stührwohldt, N., Ehinger, A., Thellmann, K. and Schaller, A. (2020) 'Processing and formation of bioactive CLE40 peptide are controlled by post-translational proline hydroxylation', *Plant Physiology*, p. pp.00528.2020. doi: 10.1104/pp.20.00528.

Sueldo, D., Ahmed, A., Misas-Villamil, J., Colby, T., Tameling, W., Joosten, M. H. A. J. and Van der Hoorn, R. A. L. (2014) 'Dynamic hydrolase activities precede hypersensitive tissue collapse in tomato seedlings', *New Phytologist*, 203(3), pp. 913–925. doi: 10.1111/nph.12870.

Sueldo, D. J. and van der Hoorn, R. A. L. (2017) 'Plant life needs cell death, but does plant cell death need Cys proteases?', *The FEBS Journal*, 284(10), pp. 1577–1585. doi: 10.1111/febs.14034.

Suty, L., Lequeu, J., Lançon, A., Etienne, P., Petitot, A.-S. and Blein, J.-P. (2003) 'Preferential induction of 20S proteasome subunits during elicitation of plant defense reactions: towards the characterization of plant defense proteasomes.', *The International Journal of Biochemistry & Cell Biology*, 35(5), pp. 637–50.

Sweeney, P., Park, H., Baumann, M., Dunlop, J., Frydman, J., Kopito, R., McCampbell, A., Leblanc, G., Venkateswaran, A., Nurmi, A. and Hodgson, R. (2017) 'Protein misfolding in neurodegenerative diseases: Implications and strategies', *Translational Neurodegeneration*, 6(1), pp. 1–13. doi: 10.1186/s40035-017-0077-5.

Tegeder, M. and Masclaux-Daubresse, C. (2018) 'Source and sink mechanisms of nitrogen transport and use', *New Phytologist*, 217(1), pp. 35–53. doi: 10.1111/nph.14876.

Teper-Bamnolker, P., Danieli, R., Peled-Zehavi, H., Abu-Abied, M., Avin-Wittenberg, T., Sadot, E. and Eshel, D. (2019) 'Vacuolar processing enzyme translocates to the vacuole through the autophagy pathway to induce programmed cell death', *bioRxiv*, p. 831982. doi: 10.1101/831982.

Tian, M., Huitema, E., da Cunha, L., Torto-Alalibo, T. and Kamoun, S. (2004) 'A Kazal-like extracellular serine protease inhibitor from Phytophthora infestans targets the tomato pathogenesis-related protease P69B', *Journal of Biological Chemistry*, 279(25), pp. 26370–26377. doi: 10.1074/jbc.M400941200.

Tien, W. S., Chen, J. H. and Wu, K. P. (2017) 'SheddomeDB: The ectodomain shedding database for membrane-bound shed markers', *BMC Bioinformatics*, 18(3), pp. 31–48. doi: 10.1186/s12859-017-1465-7.

Trusova, S. V., Golyshev, S. A., Chichkova, N. V. and Vartapetian, A. B. (2019) 'Sometimes they come back: Endocytosis provides localization dynamics of a subtilase in cells committed to cell death', *Journal of Experimental Botany*, 70(7), pp. 2003–2007. doi: 10.1093/jxb/erz014.

Tsiatsiani, L., Gevaert, K. and Van Breusegem, F. (2012) 'Natural substrates of plant proteases: How can protease degradomics extend our knowledge?', *Physiologia Plantarum*, 145(1), pp. 28–40. doi: 10.1111/j.1399-3054.2011.01534.x.

Tsiatsiani, L., Timmerman, E., De Bock, P.-J., Vercammen, D., Stael, S., van de Cotte, B., Staes, A., Goethals, M., Beunens, T., Van Damme, P., Gevaert, K. and Van Breusegem, F. (2013) 'The Arabidopsis metacaspase9 degradome.', *The Plant Cell*, 25(8), pp. 2831–47. doi: 10.1105/tpc.113.115287.

Uddin, M. S., Naider, F. and Becker, J. M. (2017) 'Dynamic roles for the N-terminus of the yeast G protein-coupled receptor Ste2p', *Biochimica et Biophysica Acta - Biomembranes*, 1859(10), pp. 2058–2067. doi: 10.1016/j.bbamem.2017.07.014.

Urquhart, W., Gunawardena, A. H. L. A. N., Moeder, W., Ali, R., Berkowitz, G. A. and Yoshioka, K. (2007) 'The chimeric cyclic nucleotide-gated ion channel ATCNGC11/12 constitutively induces programmed cell death in a Ca2+ dependent manner.', *Plant Molecular Biology*, 65(6), pp. 747–761. doi: 10.1007/s11103-007-9239-7.

Üstün, S., Sheikh, A., Gimenez-Ibanez, S., Jones, A., Ntoukakis, V. and Börnke, F.

(2016) 'The proteasome acts as a hub for plant immunity and is targeted by Pseudomonas type III effectors.', *Plant Physiology*, 172(3), pp. 1941–1958. doi: 10.1104/pp.16.00808.

Varland, S., Osberg, C. and Arnesen, T. (2015) 'N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects', *Proteomics*, 15(14), pp. 2385–2401. doi: 10.1002/pmic.201400619.

Vartapetian, A. B., Tuzhikov, A. I., Chichkova, N. V, Taliansky, M. and Wolpert, T. J. (2011) 'A plant alternative to animal caspases: subtilisin-like proteases.', *Cell Death and Differentiation*, 18(8), pp. 1289–1297. doi: 10.1038/cdd.2011.49.

Venne, A. S., Solari, F. A., Faden, F., Paretti, T., Dissmeyer, N. and Zahedi, R. P. (2015) 'An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in Arabidopsis thaliana.', *Proteomics*, 15(14), pp. 2458–2469. doi: 10.1002/pmic.201500014.

Venne, A. S., Vögtle, F. N., Meisinger, C., Sickmann, A. and Zahedi, R. P. (2013) 'Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55', *Journal of Proteome Research*, 12(9), pp. 3823–3830. doi: 10.1021/pr400435d.

Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.-A., De Rycke, R., Brackenier, A., Inzé, D., Harris, J. L. and Van Breusegem, F. (2006) 'Serpin1 of Arabidopsis thaliana is a suicide inhibitor for metacaspase 9.', *Journal of Molecular Biology*, 364(4), pp. 625–636. doi: 10.1016/j.jmb.2006.09.010.

Vizovišek, M., Vidmar, R., Fonović, M. and Turk, B. (2016) 'Current trends and challenges in proteomic identification of protease substrates', *Biochimie*, 122, pp. 77–87. doi: 10.1016/J.BIOCHI.2015.10.017.

Vlodavsky, I., Singh, P., Boyango, I., Gutter-Kapon, L., Elkin, M., Sanderson, R. D. and Ilan, N. (2016) 'Heparanase: From basic research to therapeutic applications in cancer and inflammation', *Drug Resistance Updates*, 29, pp. 54–75. doi: 10.1016/j.drup.2016.10.001.

Vlot, A. C., Dempsey, D. A. and Klessig, D. F. (2009) 'Salicylic acid, a multifaceted hormone to combat disease.', *Annual Review of Phytopathology*, 47(1), pp. 177–206. doi: 10.1146/annurev.phyto.050908.135202.

Vögtle, F. N., Prinz, C., Kellermann, J., Lottspeich, F., Pfanner, N. and Meisinger, C. (2011) 'Mitochondrial protein turnover: Role of the precursor intermediate peptidase Oct1 in protein stabilization', *Molecular Biology of the Cell*, 22(13), pp. 2135–2143. doi: 10.1091/mbc.E11-02-0169.

Vögtle, F. N., Wortelkamp, S., Zahedi, R. P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N. and Meisinger, C. (2009) 'Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability', *Cell*, 139(2), pp. 428–439. doi: 10.1016/j.cell.2009.07.045.

Vorster, B. J., Cullis, C. A. and Kunert, K. J. (2019) 'Plant Vacuolar Processing Enzymes', *Frontiers in Plant Science*, 10, p. 479. doi: 10.3389/fpls.2019.00479.

Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H. W., Zhou, J. M. and Chai, J. (2019a) 'Reconstitution and structure of a plant NLR resistosome conferring immunity', *Science*, 364(6435). doi: 10.1126/science.aav5870.

Wang, J., Wang, J., Hu, M., Wu, S., Qi, J., Wang, G., Han, Z., Qi, Y., Gao, N., Wang, H. W., Zhou, J. M. and Chai, J. (2019b) 'Ligand-triggered allosteric ADP release primes a plant NLR complex', *Science*, 364(6435). doi: 10.1126/science.aav5868.

Wang, J. and Wen, C.-K. (2015) 'Cryptic role of the ETHYLENE INSENSITIVE2 nuclear localization signal in ethylene signaling', *Molecular Plant*, 8(8), pp. 1129–1130. doi: 10.1016/j.molp.2015.06.007.

Wang, W., Xiong, H., Lin, R., Zhao, N., Zhao, P. and Sun, M.-X. (2018) 'A VPE-like protease NtTPE8 exclusively expresses in the integumentary tapetum and involves in seed development', *Journal of Integrative Plant Biology*. doi: 10.1111/jipb.12766.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: Homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46(W1), pp. W296–W303. doi: 10.1093/nar/gky427.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009) 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

Weber, M., Deinlein, U., Fischer, S., Rogowski, M., Geimer, S., Tenhaken, R. and Clemens, S. (2013) 'A mutation in the *Arabidopsis thaliana* cell wall biosynthesis gene *pectin methylesterase 3* as well as its aberrant expression cause hypersensitivity specifically to Zn', *The Plant Journal*, 76(1), p. n/a-n/a. doi: 10.1111/tpj.12279.

Wei, C. F., Kvitko, B. H., Shimizu, R., Crabill, E., Alfano, J. R., Lin, N. C., Martin, G. B., Huang, H. C. and Collmer, A. (2007) 'A Pseudomonas syringae pv. tomato DC3000 mutant lacking the type III effector HopQ1-1 is able to cause disease in the model plant Nicotiana benthamiana', *Plant Journal*, 51(1), pp. 32–46. doi: 10.1111/j.1365-313X.2007.03126.x.

Weidmann, J. and Craik, D. J. (2016) 'Discovery, structure, function, and applications of cyclotides: circular proteins from plants', *Journal of Experimental Botany*, 67(16), pp. 4801–4812. doi: 10.1093/jxb/erw210.

Weng, S. S., Demir, F., Ergin, E. K., Dirnberger, S., Uzozie, A., Tuscher, D., Nierves, L., Tsui, J., Huesgen, P. F. and Lange, P. F. (2019) 'Sensitive determination of proteolytic proteoforms in limited microscale proteome samples', *Molecular and Cellular Proteomics*, p. mcp.TIR119.001560. doi: 10.1074/MCP.TIR119.001560.

Wilkins, T. A. and Raikhel, N. V. (1989) 'Expression of rice lectin is governed by two temporally and spatially regulated mRNAs in developing embryos.', *The Plant Cell*, 1(5), pp. 541–549. doi: 10.1105/tpc.1.5.541.

Willems, P., Horne, A., Van Parys, T., Goormachtig, S., De Smet, I., Botzki, A., Van Breusegem, F. and Gevaert, K. (2019) 'The Plant PTM Viewer, a central resource for exploring plant protein modifications', *Plant Journal*, 99(4), pp. 752–762. doi: 10.1111/tpj.14345.

Willems, P., Ndah, E., Jonckheere, V., Stael, S., Sticker, A., Martens, L., Van Breusegem, F., Gevaert, K. and Van Damme, P. (2017) 'N-terminal proteomics assisted profiling of the unexplored translation initiation landscape in Arabidopsis thaliana', *Molecular and Cellular Proteomics*, 16(6), pp. 1064–1080. doi: 10.1074/mcp.M116.066662.

Wolf, S., Rausch, T. and Greiner, S. (2009) 'The N-terminal pro region mediates retention of

unprocessed type-I PME in the Golgi apparatus', *The Plant Journal*, 58(3), pp. 361–375. doi: 10.1111/j.1365-313X.2009.03784.x.

Wu, H. C. and Lin, J. Y. (1993) 'The complete amino acid sequence of a kunitz family trypsin inhibitor from seeds of acacia confusa', *Journal of Biochemistry*, 113(2), pp. 258–263. doi: 10.1093/oxfordjournals.jbchem.a124036.

Wu, L., Viola, C. M., Brzozowski, A. M. and Davies, G. J. (2015) 'Structural characterization of human heparanase reveals insights into substrate recognition', *Nature Structural and Molecular Biology*, 22(12), pp. 1016–1022. doi: 10.1038/nsmb.3136.

van Wyk, S. G., Du Plessis, M., Cullis, C. A., Kunert, K. J. and Vorster, B. J. (2014) 'Cysteine protease and cystatin expression and activity during soybean nodule development and senescence', *BMC Plant Biology*, 14(1), p. 294. doi: 10.1186/s12870-014-0294-3.

Xia, Y., Suzuki, H., Borevitz, J., Blount, J., Guo, Z., Patel, K., Dixon, R. A. and Lamb, C. (2004) 'An extracellular aspartic protease functions in Arabidopsis disease resistance signaling.', *The EMBO Journal*, 23(4), pp. 980–8. doi: 10.1038/sj.emboj.7600086.

Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G. and Zhang, B. (2014) 'CasOT: a genome-wide Cas9/gRNA off-target searching tool', *Bioinformatics*, 30(8), pp. 1180–1182. doi: 10.1093/bioinformatics/btt764.

Xiao, Y., Stegmann, M., Han, Z., DeFalco, T. A., Parys, K., Xu, L., Belkhadir, Y., Zipfel, C. and Chai, J. (2019) 'Mechanisms of RALF peptide perception by a heterotypic receptor complex', *Nature*, 572(7768), pp. 270–274. doi: 10.1038/s41586-019-1409-7.

Xu, F. Q. and Xue, H. W. (2019) 'The ubiquitin-proteasome system in plant responses to environments', *Plant Cell and Environment*, 42(10), pp. 2931–2944. doi: 10.1111/pce.13633.

Yalpani, N., Silverman, P., Wilson, T. M., Kleier, D. A. and Raskin, I. (1991) 'Salicylic acid is a systemic signal and an inducer of pathogenesis-related proteins in virus-infected tobacco.', *The Plant Cell*, 3(8), pp. 809–18.

Yamada, K., Basak, A. K., Goto-Yamada, S., Tarnawska-Glatt, K. and Hara-Nishimura, I. (2020) 'Vacuolar processing enzymes in the plant life cycle', *New Phytologist*, 226(1), pp. 21–31. doi: 10.1111/nph.16306.

Yamada, K., Matsushima, R., Nishimura, M. and Hara-Nishimura, I. (2001) 'A slow maturation of a cysteine protease with a granulin domain in the vacuoles of senescing Arabidopsis leaves.', *Plant Physiology*, 127(4), pp. 1626–34.

Yamada, K., Shimada, T., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1999) 'Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme.', *The Journal of Biological Chemistry*, 274(4), pp. 2563–70. doi: 10.1074/jbc.274.4.2563.

Yang, R., Wong, Y. H., Nguyen, G. K. T., Tam, J. P., Lescar, J. and Wu, B. (2017) 'Engineering a catalytically efficient recombinant protein ligase', *Journal of the American Chemical Society*, 139(15), pp. 5351–5358. doi: 10.1021/jacs.6b12637.

Yates, G., Srivastava, A. K. and Sadanandom, A. (2016) 'SUMO proteases: Uncovering the roles of deSUMOylation in plants', *Journal of Experimental Botany*, 67(9), pp. 2541–2548. doi: 10.1093/jxb/erw092.

Yeom, J., Ju, S., Choi, Y. J., Paek, E. and Lee, C. (2017) 'Comprehensive analysis of human protein N-termini enables assessment of various protein forms', *Scientific Reports*, 7(1), pp. 1–13. doi: 10.1038/s41598-017-06314-9.

Yonder Haar, R. A., Allen, R. D., Cohen, E. A., Nessler, C. L. and Thomas, T. L. (1988) 'Organization of the sunflower 11S storage protein gene family', *Gene*, 74(2), pp. 433–443. doi: 10.1016/0378-1119(88)90176-X.

Yu, M., Li, R., Cui, Y., Chen, W., Li, B., Zhang, X., Bu, Y., Cao, Y., Xing, J., Jewaria, P. K., Li, X., Bhalerao, R. P., Yu, F. and Lin, J. (2020) 'The RALF1-FERONIA interaction modulates endocytosis to mediate control of root growth in Arabidopsis', *Development (Cambridge)*, 147(13). doi: 10.1242/dev.189902.

Yuan, F., Verhelst, S. H. L., Blum, G., Coussens, L. M. and Bogyo, M. (2006) 'A selective activity-based probe for the papain family cysteine protease dipeptidyl peptidase I/cathepsin C', *Journal of the American Chemical Society*, 128(17), pp. 5616–5617. doi: 10.1021/ja060835v.

Zauner, F. B., Dall, E., Regl, C., Grassi, L., Huber, C. G., Cabrele, C. and Brandstetter, H. (2018a) 'Crystal structure of plant legumain reveals a unique two-chain state with pH-dependent activity regulation', *The Plant Cell*, 30(3), pp. 686–699. doi: 10.1105/tpc.17.00963.

Zauner, F. B., Elsässer, B., Dall, E., Cabrele, C. and Brandstetter, H. (2018b) 'Structural analyses of Arabidopsis thaliana legumain γ reveal differential recognition and processing of proteolysis and ligation substrates', *Journal of Biological Chemistry*, 293(23), pp. 8934–8946. doi: 10.1074/jbc.M117.817031.

Zhang, D., Liu, D., Lv, X., Wang, Y., Xun, Z., Liu, Z., Li, F. and Lu, H. (2014) 'The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in Arabidopsis', *The Plant Cell*, 26(7), pp. 2939–2961. doi: 10.1105/tpc.114.127282.

Zhang, H., Deery, M. J., Gannon, L., Powers, S. J., Lilley, K. S. and Theodoulou, F. L. (2015) 'Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in Arabidopsis roots', *Proteomics*, 15(14), pp. 2447–57. doi: 10.1002/pmic.201400530.

Zhang, H., Dong, S., Wang, M., Wang, W., Song, W., Dou, X., Zheng, X. and Zhang, Z. (2010) 'The role of vacuolar processing enzyme (VPE) from Nicotiana benthamiana in the elicitor-triggered hypersensitive response and stomatal closure', *Journal of Experimental Botany*, 61(13), pp. 3799–3812. doi: 10.1093/jxb/erq189.

Zhang, H., Xiao, Y., Deng, X., Feng, H., Li, Z., Zhang, L. and Chen, H. (2020) 'OsVPE3 mediates GA-induced programmed cell death in rice aleurone layers via interacting with actin microfilaments', *Rice*, 13(1), p. 22. doi: 10.1186/s12284-020-00376-6.

Zhang, J., Li, Q.-F. F., Huang, W.-W. W., Xu, X.-Y. Y., Zhang, X.-L. L., Hui, M.-X. X., Zhang, M.-K. K. and Zhang, L.-G. G. (2013) 'A vacuolar processing enzyme RsVPE1 gene of radish is involved in floral bud abortion under heat stress', *International Journal of Molecular Sciences*, 14(7). doi: 10.3390/ijms140713346.

Zhao, L., Hua, T., Crowley, C., Ru, H., Ni, X., Shaw, N., Jiao, L., Ding, W., Qu, L., Hung, L. W., Huang, W., Liu, L., Ye, K., Ouyang, S., Cheng, G. and Liu, Z. J. (2014) 'Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage', *Cell Research*, 24(3), pp. 344–358. doi: 10.1038/cr.2014.4.

Zhou, H., Zhao, J., Cai, J. and Patil, S. B. (2017) 'UBIQUITIN-SPECIFIC PROTEASES

function in plant development and stress responses', *Plant Molecular Biology*. Springer Netherlands, pp. 565–576. doi: 10.1007/s11103-017-0633-5.

Zipfel, C. (2014) 'Plant pattern-recognition receptors.', *Trends in immunology*, 35(7), pp. 345–51. doi: 10.1016/j.it.2014.05.004.

Zuckerkandl, E. and Pauling, L. (1965) 'Evolutionary divergence and convergence in proteins', in *Evolving Genes and Proteins*. Elsevier, pp. 97–166. doi: 10.1016/B978-1-4832-2734-4.50017-6.

Appendix 1: Ten Prominent Host Proteases in Plant — Pathogen Interactions

Emma L. Thomas [1], Renier A. L van der Hoorn[1,]

[1]Department of Plant Sciences, University of Oxford

This chapter is a published review; **Thomas, E. L. and van der Hoorn, R. A. L**. (2018) 'Ten prominent host proteases in plant-pathogen interactions.', International Journal of Molecular Sciences, 19(2). doi: 10.3390/ijms19020639. with minor textual adaptations.

A1.1 Abstract

Proteases are integral enzymes of the plant immune system. Multiple aspects of defence are regulated by proteases, including the hypersensitive response, pathogen recognition, priming and peptide hormone release. These processes are regulated by unrelated proteases residing at different subcellular locations. In this review we discuss ten prominent plant proteases contributing to the plant immune system, highlighting the diversity of roles they perform in plant defence.

A1.2 Introduction

Numerous biological processes in plants require proteases, including the plant defence response (Van der Hoorn, 2008; Jashni *et al.*, 2015; Salvesen *et al.*, 2016; Figueiredo *et al.*, 2017). Plants can perceive pathogens by recognition of conserved pathogen associated molecular patterns (PAMPs), such as flagellin or chitin, via membrane localised pattern recognition receptors (PRRs). Alternatively secreted pathogen effectors are recognised, predominantly by nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins (Zipfel, 2014). Following pathogen recognition, signalling events transduce crucial information on pathogen attack and coordinate intracellular and tissue wide responses(Jones and Dangl, 2006; Dodds and Rathjen, 2010). The cell undergoes large scale transcriptional reprogramming and changes in salicylic acid, jasmonic acid, and ethylene hormonal signalling to control local and systemic defence responses (Malamy *et al.*, 1990; Yalpani *et al.*, 1991; Vlot *et al.*, 2009; Buscaill and Rivas, 2014; Li *et al.*, 2016). A common feature of defence is the hypersensitive response (HR), a form of programmed cell death occurring locally at the primary infection site. HR and local immune responses limit the spread of the pathogen and restrict their access to nutrients (Dickman and Fluhr, 2013).

This review focuses on the roles of host proteases in plant-pathogen interactions discussed by their subcellular localisation (**Fig. 1**). We highlight ten examples of proteases with diverse roles in defence (**Table 1**) to illustrate both the importance of proteases to multiple aspects of plant defence and the diversity of proteases in the plant immune system.

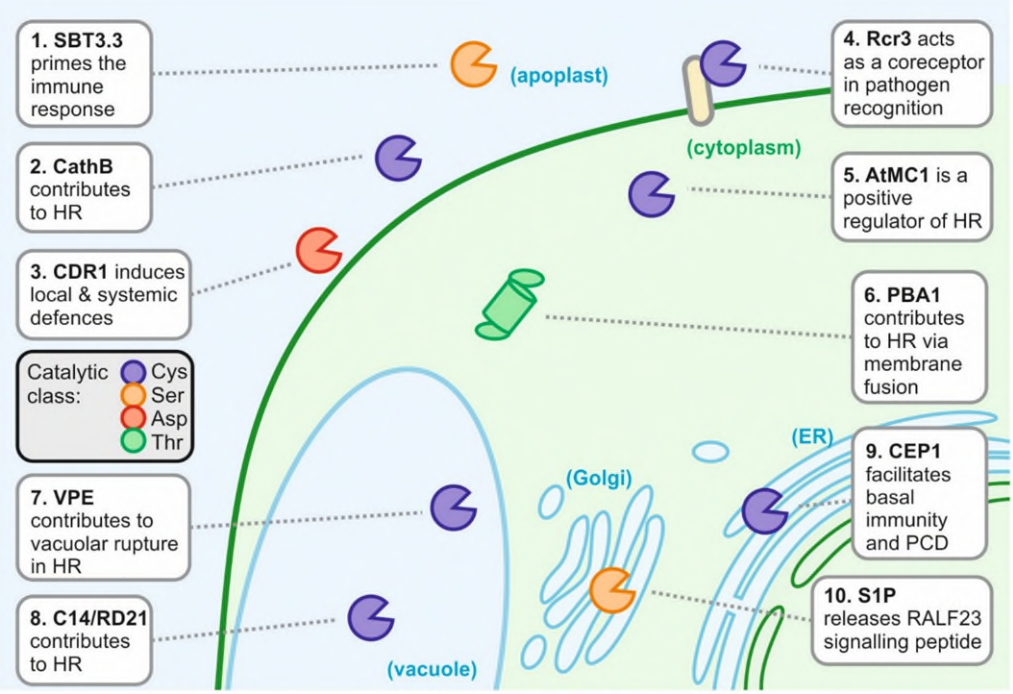


Figure 1. Ten host proteases contributing to the plant defence response. AtMC1, *Arabidopsis thaliana* Metacaspase-1; C14, Cysteine protease clone 14; CatB, Cathepsin B; CDR1, Constitutive Disease Resistance-1; CEP1, Cysteine EndoPeptidase-1; HR, Hypersensitive Response; PBA1, proteasome beta subunit-1; PCD, Programmed Cell Death; Rcr3, Required for Cladosporium Resistance-3; RD21, Responsive to Dessication-21; S1P, Site-1-Protease; SBT3.3, subtilase 3.3; VPE, Vacuolar Processing Enzyme.

Table 1. Ten prominent host proteases in plant-pathogen interactions.

	Function in defence	Subcellular localisation	MEROPS family	Organism	Known substrate in defence?
1. SBT3.3	Priming	Apoplast	S08, subtilisin-like	*A. thaliana*	No
2. CathB	HR	Apoplast (+ Vacuole)	C01, papain-like	*N. benthamiana*, *A. thaliana*	No
3. CDR1	Signalling	Apoplast	A01, pepsin-like	*A. thaliana*	No
4. Rcr3	Recognition	Apoplast	C01, papain-like	Tomato	No
5. AtMC1	HR	Cytoplasm (+ Nucleus)	C14, metacaspase	*A. thaliana*	No

6. PBA1	HR, membrane fusion	Cytoplasm	T01, proteasome	*A. thaliana*	No
7. VPE	HR, membrane fusion	Vacuole	C13, legumain-like	*N. benthamiana, A. thaliana*	No
8. C14/RD21	HR, resistance	Vacuole	C01, papain-like	Tomato, *A. thaliana*	No
9. CEP1	Basal resistance	ER-derived compartments	C01, papain-like	*A. thaliana*	No
10. S1P	Signalling, hormone release	Golgi	S08, subtilisin-like	*A. thaliana*	RALF23

MEROPS database (https://www.ebi.ac.uk/merops/). The major subcellular localisation is named, locations in brackets are other reported localisations.

A1.3 Apoplastic Proteases

Early interactions between plant and pathogen occur in the apoplast. Many bacterial, fungal and oomycete pathogens start host colonisation in the apoplastic environment, through secretion of effector molecules. Host proteins in the apoplast or plasma membrane play a role in the perception of pathogen effector molecules and in extracellular defence signalling.

A1.3.1 SBT3.3 regulates the priming of the plant immune response

Priming is the memory of a stress that enables the plant to launch an amplified and more rapid defence response upon future challenge. The Arabidopsis serine protease SBT3.3 (a subtilase member of the S8 family) regulates defence priming. Arabidopsis *sbt3.3* mutants are hypersusceptible to both the model bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 and the oomycete *Hyaloperonospora arabidopsidis* (Ramírez *et al.*, 2013). Mutant

sbt3.3 Arabidopsis plants are impaired in priming of both gene expression and signalling activity. Induction of Mitogen-associated Protein Kinase (MPK) activity is enhanced upon SBT3.3 overexpression. Furthermore, SBT3.3 overexpression increases the abundance of transcriptional activating epigenetic marks at SA regulated genes, including the promoters of WRKY transcription factors and SBT3.3, creating a positive feedback loop. Consequently these genes are induced more rapidly upon pathogen challenge (Ramírez *et al.*, 2013). Interestingly, in addition to defence priming, SBT3.3 is required for SA regulated gene activation. The mechanism of SBT3.3 controlling SA regulated gene expression and priming the immune response remains enigmatic, as do the substrates of SBT3.3.

A1.3.2 CathB is a positive regulator of HR

The papain-like cysteine protease Cathepsin B (CathB, a C1 family member) is a positive regulator of HR. Multiple *CathB* genes contribute redundantly to basal resistance in Arabidopsis (McLellan *et al.*, 2009). In *Nicotiana benthamiana*, CathB is secreted into the plant apoplast and activated (Gilroy *et al.*, 2007). Deficiency of CathB in *N. benthamiana* restricts PCD triggered by the bacterial pathogens *Erwinia amylovora* and avirulent *Pst* DC3000. Furthermore, PCD triggered by hydrogen peroxide, a prominent plant defence signal, and ER stress induced by tunicamycin requires CathB activity (Ge *et al.*, 2016; Cai *et al.*, 2018).

Whilst the involvement of CathB in HR is well established, the extent of its involvement depends on the HR inducer. CathB is required for HR triggered upon coexpression of *P. infestans* avirulence gene *AvrR3a* and potato resistance gene *R3a*, and upon *Ps* pv. *glycinea* *AvrB* expression (Gilroy *et al.*, 2007; McLellan *et al.*, 2009). However, CathB deficiency does

not perturb HR upon co-expression of the *Cladosporium fulvum* avirulence gene *Avr4* and the tomato resistance gene *Cf-4* in *N benthamiana* (Gilroy *et al.*, 2007; McLellan *et al.*, 2009). Arabidopsis CathB is also not necessary for resistance to *P. syringae* harbouring avirulence genes *AvrB* or *AvrRps4* (McLellan *et al.*, 2009). The conflicting roles indicate that CathB is important in multiple forms of HR, but is not necessarily a universal HR regulator.

A1.3.3 CDR1 promotes the release of systemic defence signals

CDR1 is an apoplastic aspartic protease of the A1 family that contributes to local and systemic defence signalling in Arabidopsis. Activation tagging of CDR1 results in enhanced resistance to multiple *P. syringae* strains, alongside constitutive *Pathogenesis Related* (*PR*) gene expression in a SA-dependent manner. *PR* gene expression is abolished in CDR1 active site mutants and upon application of the aspartic protease inhibitor pepstatin A, demonstrating that protease activity is required for the role in defence (Xia *et al.*, 2004). CDR1 generates an extracellular mobile signal capable of inducing defence responses both locally and systemically. Low molecular weight fractions of apoplastic fluids from CDR1 overexpressing plants induce defence responses in unchallenged plants, in both the infiltrated and distant leaves (Xia *et al.*, 2004). The activity of CDR1 appears to be conserved between species. Rice OsCDR1 expressed in Arabidopsis similarly generates apoplastic fluids that induce systemic defence (Prasad *et al.*, 2009). OsCDR1 overexpression in Arabidopsis also mimics the enhanced resistance to *Pst* observed on AtCDR1 overexpression (Prasad *et al.*, 2009). The nature of the signal generated by CDR1 is currently unknown. Identification of the substrates of CDR1 will lead to insights into the systemic induction of SA dependent defence responses.

A1.3.4 Rcr3 is a coreceptor for perception of unrelated pathogens

Recognition of a pathogen is the first step in mounting an immune response. The extracellular cysteine protease Rcr3 (family C1A) is crucial for the recognition of unrelated pathogens, including the fungus *C. fulvum* and nematode *Globodera rostochiensis*. Both the fungus and nematode secrete unrelated protease inhibitors (Avr2 and GrVAP1, respectively) that inhibit Rcr3 (Rooney *et al.*, 2005; Lozano-Torres *et al.*, 2012). The inhibitor-Rcr3 complex is perceived by the tomato leucine-rich repeat receptor like protein, Cf-2. Recognition triggers an oxidative burst, followed by transcriptional reprogramming and HR, culminating in disease resistance (Rooney *et al.*, 2005). This is dependent on the presence of both Cf-2 and Rcr3. Rcr3 is proposed to act as a decoy with the operative effector target, Pip1, a paralogous and more abundant immune protease (Shabab *et al.*, 2008; Ilyas *et al.*, 2015). Deficiency of Pip1 renders the plant hyper-susceptible to *P. infestans*, *C. fulvum* and *P. syringae* (Ilyas *et al.*, 2015). In addition to the role in pathogen recognition, Rcr3 contributes to resistance via alternative pathways independent of *Cf-2*. *P. infestans* produces inhibitors of Rcr3 (EPICs), but unlike *C. fulvum* and *G. rostochiensis* infection, these do not trigger HR. In the absence of Cf-2, *rcr3* mutants are hypersusceptible to *P. infestans,* but not to *C. fulvum* (Ilyas *et al.*, 2015), indicating a Rcr3 role separate from Cf-2-dependent pathogen recognition.

A1.4 Cytonuclear Proteases

The cytoplasm is an important signalling location that bridges the extracellular perception of pathogens and the intracellular responses, including changes in gene regulation, metabolite biosynthesis and induction of PCD. Cytoplasmic proteases have been implicated in HR regulation.

A1.4.1 AtMC1 is a positive regulator of HR

Two cytosolic metacaspases, AtMC1 and AtMC2 (family C14), act antagonistically in the regulation of HR in Arabidopsis. AtMC1 positively regulates HR cell death induced by *Pst* DC3000 carrying avrRPM1, although this HR does not affect pathogen growth (Coll *et al.*, 2010). Furthermore, AtMC1 is essential for the runaway cell death phenotype of defective immune components, including autoactive NLRs (key *R* genes) and *lsd1* (Coll *et al.*, 2010; Roberts *et al.*, 2013). Consistent with its pro-cell death function, AtMC1 activity is tightly controlled by two negative regulators; LSD1 and AtSERPIN1 (Coll *et al.*, 2010; Lema Asqui *et al.*, 2017). LSD1 directly interacts with AtMC1 through the N-terminal Zinc finger domain (Dietrich *et al.*, 1994), whereas suicide protease inhibitor AtSERPIN1 covalently and irreversibly inhibits AtMC1 (Lema Asqui *et al.*, 2017). Pro-death activity of AtMC1 is also suppressed by AtMC2. Overexpression of AtMC2 phenocopies the suppressed HR phenotype of *atmc1* mutants. Interestingly, whilst the role of AtMC1 in immunity requires its catalytic residues, the role of AtMC2 does not (Coll *et al.*, 2010). It is unknown how AtMC2 exerts the negative regulation of death independent of its protease activity, nor whether its catalytic activity contributes to alternative pathways in defence.

A1.4.2 The proteasome is a positive regulator of HR

The host plant proteasome is essential for protein homeostasis and is heavily implicated in plant defence (Üstün *et al.*, 2016). A notable example is that of NPR1, a transcriptional coactivator essential for SA-regulated gene expression. In non-induced cells, inappropriate transcription is restricted through degradation of NPR1 by the proteasome, whereas on SA induction degradation is required for full transcriptional activation (Spoel *et al.*, 2009; Fu *et*

al., 2012). Degradation is proposed to increase NPR1 recycling rate thereby enabling greater gene expression.

The core particle of the proteasome is comprised of multiple subunits forming heptameric rings and include three catalytic β subunits with distinct proteolytic activities (Löwe *et al.*, 1995; Kurepa and Smalle, 2008; Murata *et al.*, 2009). One catalytic subunit in particular, the threonine protease PBA1/β1 (of the T1 family), has been further investigated in the context of HR, due to its caspase-3-like activity. The presence of casapse-3 like activity is an established requirement for plant PCD and in certain forms of PCD 60% of the caspase-3 like activity can be attributed to PBA1 (del Pozo and Lam, 1998; Bonneau *et al.*, 2008; Hatsugai *et al.*, 2009; Ge *et al.*, 2016). Tobacco *PBA1* expression is induced following treatment with the fungal elicitor, cryptogein (Dahan *et al.*, 2001; Suty *et al.*, 2003). Deficiency of the PBA1 subunit compromises HR triggered by avirulent *Pst* DC3000 carrying *AvrRpm1*. This HR was observed to proceed through the fusion of tonoplast and plasma membranes (Hatsugai *et al.*, 2009). PBA1 silenced plants exhibit reduced activity of the other two catalytic subunits, PBB and PBE. Silencing of these subunits replicates the suppression of HR indicating a general role of the proteasome in HR induction (Hatsugai *et al.*, 2009). PBA1-dependent HR is distinct from the regulation of gene expression, as induction of NADPH oxidases and *PR* genes is not suppressed in PBA1 deficient plants (Hatsugai *et al.*, 2009).

A1.5 Vacuolar Proteases

The vacuole is an acidic hydrolytic storage compartment occupying the largest volume of a leaf cell. Rupture of the vacuole during HR dramatically alters the cytoplasm by acidification and the release of lytic enzymes and potential cell death mediators (Jones, 2001; Ito and

Fukuda, 2002; Rojo *et al.*, 2003). Two vacuolar proteases have been identified that contribute to HR.

A1.5.1 VPEs regulate vacuolar rupture during virus-induced HR

Vacuolar Processing Enzymes (VPEs/Asparaginyl endopeptidases/Legumains, family C13) are key regulators of tonoplast integrity in PCD. VPEs cleave after asparagine (N) but can also cleave after aspartic acid (D) and therefore have caspase-1 like activity. VPEs are essential for vacuolar rupture and HR upon infection by Tobacco Mosaic Virus (TMV) on *N. benthamiana* carrying the *N* resistance gene (Hatsugai *et al.*, 2004). Similarly, upon ER stress induced PCD, absence of VPEs prevents vacuolar rupture. VPEs mature autocatalytically and is known to activate another protease, AtCPY, in the vacuole (Rojo *et al.*, 2003; Hatsugai *et al.*, 2015). VPEs may therefore be key regulators of the PCD induction pathway.

Despite the lack of known substrates of VPEs during PCD, the requirement of VPEs for HR is well described (Hatsugai *et al.*, 2015; Kabbage *et al.*, 2017). VPEs are also required for HR triggered by mycotoxin FB1, bacterial elicitor harpin and the co-expression of calcium channels CNGC11 and CNGC12 (Kuroyanagi *et al.*, 2005; Urquhart *et al.*, 2007; Zhang *et al.*, 2010; Hatsugai *et al.*, 2015) Nonetheless, the role of VPEs is not universal. HR induced by boehmerin and Nep1 is not perturbed upon VPE silencing (Zhang *et al.*, 2010).

VPE mediated tonoplast rupture is thought to be effective against cytoplasmic pathogens like viruses that become exposed to vacuolar hydrolases and low pH (Hatsugai *et al.*, 2004). In contrast, fusion of the tonoplast and plasma membranes is dependent on PBA1 and delivers vacuolar contents to the apoplast where bacteria reside (Pajerowska-Mukhtar and Dong,

2009). Crucially however, these morphologies were observed in different plant species, *N. benthamiana* and Arabidopsis, respectively. To date it is unclear how widespread the different forms of HR PCD are.

A1.5.2 C14/RD21 have a complex regulation

C14 and RD21 are orthologous papain-like proteases from tomato and Arabidopsis, respectively (Shindo *et al.*, 2012), carrying a C-terminal granulin domain(Yamada *et al.*, 2001). Tomato C14 has been detected in the vacuole (Bozkurt *et al.*, 2011) and extracellularly (Shabab *et al.*, 2008) C14 probably plays an important role in immunity because its activity and localisation are manipulated by multiple effectors. The extracellular C14 is targeted by cystatin-like EpiC inhibitors of the oomycete pathogen *P. infestans* (Kaschani *et al.*, 2010), and the chagasin-like Cip1 inhibitor of the *Pst* DC3000 (Shindo *et al.*, 2016). In addition, RxLR effector AvrBlb2 of *P. infestans* associates with C14 and prevents its secretion into the apoplast (Bozkurt *et al.*, 2011).

Importantly, silencing or overexpression of a C14 homolog in *N. benthamiana* enhances or decreases susceptibility to *P. infestans,* respectively (Kaschani *et al.*, 2010; Bozkurt *et al.*, 2011). However, Arabidopsis *rd21* knock-out lines are not more susceptible to the oomycete *H. arabidopsidis*, even though they express genes encoding EpiC-like inhibitors (Shindo *et al.*, 2012). Nevertheless, these *rd21* lines are more susceptible to *Botrytis cinerea* when whole plants are infected (Shindo *et al.*, 2012). Remarkably, the opposite phenotype with *B. cinerea* (increased resistance) was found for the same *rd21* mutants in detached leaf assays (Lampl *et al.*, 2013). These data indicate that the role of C14/RD21 proteases depends on the pathosystem, the assay itself and on the different ways pathogens manipulate their host.

Control over RD21 activity upon release of the vacuolar content into the cytoplasm during PCD is thought to come from AtSERPIN1, a cytoplasmic serpin-like suicide inhibitor that forms a covalent complex with RD21 (Lampl *et al.*, 2013). Indeed, AtSERPIN1 overexpression causes susceptibility to *B. cinerea* (Lampl *et al.*, 2013). However, *atserpin1* mutants do not show a phenotype and AtSERPIN1 also regulates PCD via AtMC1 (Lema Asqui *et al.*, 2017). RD21 is regulated by AtSERPIN1, kunitz inhibitor WSCP (Shindo *et al.*, 2016), protein di-isomerase PDI5 (Ondzighi *et al.*, 2008), and other mechanisms (Gu *et al.*, 2012). This makes RD21 regulation a challenging and intriguing question to resolve.

A1.6 Endomembrane Proteases

The endomembrane system includes the endoplasmic reticulum and the Golgi network which are important for protein synthesis and maturation. Stress responses rely heavily on protein production to enable the cell to adapt (Kørner *et al.*, 2015). Endomembrane compartments are also involved on many viral, fungal and oomycete infections, for instance by flanking pathogen haustoria and have been implicated in PCD initiation (Eichmann and Schäfer, 2012).

A1.6.1 ER resident AtCEP1 facilitates fungal immunity

AtCEP1 is a plant-specific, papain-like cysteine endopeptidase (family C1A) that harbours a C-terminal 'KDEL' sequence that sequesters the protease within ER-derived compartments. The expression of AtCEP1 is induced upon infection with the fungal obligate biotroph *Erysiphe cruciferarum* where it contributes to basal resistance (Höwing *et al.*, 2014, 2017). Expression of GFP fusion constructs revealed AtCEP1 is enriched in endomembranes surrounding the haustorium interface during HR induction (Höwing *et al.*, 2017). However, AtCEP1 contains

a putative cleavage site that would result in the loss of the KDEL sequence, and therefore AtCEP1 activity may also be present elsewhere (Höwing *et al.*, 2014).

Arabidopsis *atcep1* mutants are hypersusceptible to *E. cruciferarum* (Höwing *et al.*, 2014, 2017). Cells penetrated by fungal haustoria characteristically undergo PCD and this is reduced in *atcep1* mutants (Höwing *et al.*, 2014, 2017). AtCEP1 is also implicated in developmental forms of PCD, specifically tapetal PCD (Zhang *et al.*, 2014). Expression of AtCEP1 is under regulation by CPR5, a major regulator of pathogenesis related (PR) gene expression (Bowling *et al.*, 1997; Höwing *et al.*, 2017). PCD on *E. cruciferarum* infection is also controlled by CPR5 (Höwing *et al.*, 2017). Conversely, however, *cpr5* mutants are resistant to *E. cruciferarum* and exhibit spontaneous cell death, in a manner epistatic to AtCEP1 (Höwing *et al.*, 2017). The deregulation of AtCEP1 in *cpr5* mutants is thought to contribute to the excessive cell death phenotype.

A1.6.2 Golgi localised S1P controls RALF23 peptide signalling

The Golgi-localised subtilase Site-1-Protease (S1P/SBT6.1, family S8), presents a rare example of a protease in immunity with not only a verified substrate, but also a known role of its identified substrate. S1P processes RAPID ALKALINISATION FACTOR 23 (RALF23) into a mature signalling peptide (Liu *et al.*, 2007; Srivastava *et al.*, 2009). RALF23 is perceived extracellularly by the transmembrane malectin-like receptor kinase FERONIA (Stegmann *et al.*, 2017). Perception of RALF23 dampens immune signalling through inhibition of PRR complex formation, thus restricting excessive defence responses that may prove costly to the plant (Stegmann *et al.*, 2017). S1P is therefore an important intracellular subtilase that negatively regulates the immune response.

Regulation of S1P could be a mechanism to rapidly control the abundance of mature RALF23 and thereby fine tune the immune response. Indeed, both S1P activity and RALF23 abundance rapidly increase upon challenge with *Pst* DC3000. RALF23 is an important substrate of S1P, as both *s1p* and *ralf23* plants exhibit comparable enhanced ROS bursts and resistance to *Pst* DC3000 (Stegmann *et al.*, 2017). Remarkably, RALF peptide mimics have also been identified in pathogenic fungi and are contributors to virulence (Masachis *et al.*, 2016), indicating that this signalling pathway may be a core component of immunity in plants.

A1.7 Discussion

Proteases have diverse roles in the plant immune system, ranging from pathogen perception (Rcr3), defence priming (SBT3.3), signalling (CDR1 and S1P) and regulation of HR (CathB, AtMC1, PBA1, VPEs, RD21 and AtCEP1). Whilst we have highlighted just ten prominent examples here, this review is not comprehensive and many more host proteases are involved in plant-pathogen interactions.

The large number of proteases involved in HR is to be expected considering the importance of proteases in animal PCD. Cysteine proteases known as caspases are essential for animal PCD in disease, acting as both regulators and executioners of cell death (Schilling *et al.*, 2011; McIlwain *et al.*, 2013; Shalini *et al.*, 2015). The absolute requirement for caspases lead to a longstanding bias in plant research that proteases with caspase-like activities are important in plant PCD. While this is true for VPEs, CathB and PBA1/proteasome, proteases without caspase-like activity are also important in HR. Furthermore, unlike caspases, plant proteases involved in HR are of unrelated families. CathB (C1A), AtMC1 (C14), RD21 (C1A), PBA1 (T1) and VPEs (C13) are all implicated in HR regulation, and represent diverse protease

classes (Hatsugai *et al.*, 2009, 2015; McLellan *et al.*, 2009; Lampl *et al.*, 2013; Coll *et al.*, 2014; Rawlings *et al.*, 2016).

Interestingly, evidence from studies on proteases involved in HR demonstrate that HR can be genetically uncoupled from restriction of pathogen growth. Of the proteases discussed, CathB and AtMC1 both contribute to HR, independent of restricting pathogen growth (McLellan *et al.*, 2009; Coll *et al.*, 2010). Furthermore, the inconsistent requirements for the CathB, VPEs, PBA1 and RD21 proteases in HR indicate that multiple pathways to HR are present. Parallel pathways to HR would be advantageous to avoid essential nodes in immune defence networks that may be targeted by effectors and increase susceptibility to a broad range of pathogens.

Strikingly there are no examples yet of proteases directly degrading pathogen proteins. There are a number of proteases linked to defence whose role in immunity is completely unknown. For example, P69B is frequently implicated in pathogen defence, but precisely what role it plays is not clear (Dixon *et al.*, 2000; Song *et al.*, 2009; Lozano-Torres *et al.*, 2012; Fortelny *et al.*, 2014; Ilyas *et al.*, 2015; Grosse-Holz and Van der Hoorn, 2016). Moreover, other proteases that already have an identified role may possess additional functions in defence.

The requirement for host proteases in plant-pathogen interactions is clear, but the mechanism in which they act is frequently not. The major factor limiting our understanding of protease roles is the general lack of known, biologically relevant, substrates. Without this knowledge it is impossible to fully understand the mechanism of a protease in immunity. Of the proteases discussed in this review, a biologically relevant substrate has only been identified for S1P (Srivastava *et al.*, 2009). Although the evidence is compelling for RALF23 being the major substrate of S1P in the context of defence, due to RALF23 depletion and overexpression

phenocopying that of S1P (Stegmann *et al.*, 2017), it is highly unlikely that proteases have exclusively one substrate. Furthermore, it is important to validate whether protease mutant phenotypes are resulting from loss of protease activities, by including catalytically dead mutants. For most immune proteases this control has not been included. This leaves open the possibility that other protein functions may be contributing to immune phenotypes.

Proteases do not act in isolation in immunity and protease mutant phenotypes may be indirect. In humans, a computational study demonstrated that proteases impact activities of other proteases in a complex web (Fortelny *et al.*, 2014). This interconnectivity is compounded by multifunctional inhibitors (Grosse-Holz and Van der Hoorn, 2016) such as AtSERPIN1, which regulates several unrelated immune proteases (Lampl *et al.*, 2013; Ghorbani *et al.*, 2016; Lema Asqui *et al.*, 2017). In addition, many proteases can possess similar activities, such as CathB and PBA1 exhibiting caspase-3 like activity, which could therefore act redundantly. Deciphering the roles of individual proteases in defence is by no means a trivial task.

It is important to note that subcellular localisations may be dynamic, especially upon stress. This is illustrated by CathB, which is not restricted to the apoplast. Mass spectrometry data from unchallenged Arabidopsis plants supports a vacuolar localisation, whilst expression of RFP fusions and apoplastic activity assays in *N. benthamiana* support CathB presence in the apoplast (Carter *et al.*, 2004; McLellan *et al.*, 2009). It may be possible that either the protease has a dual localisation, is relocated upon different stresses, or its localisation differs between species. The Golgi localised S1P has also been detected in the apoplast, where it interacts with and is inhibited by AtSERPIN1 (Ghorbani *et al.*, 2016). AtSERPIN1 has been identified in the cytoplasm, Golgi, ER and apoplast (Vercammen *et al.*, 2006; Lampl *et al.*, 2013). Thus,

caution should be exercised when assuming the location in which proteases mediate their phenotype.

Our current knowledge of the exact roles of proteases places them as key players in many facets of pathogen responses. Future efforts in this field will need to address the lack of known substrates through improving mass-spectrometry based methods (Vizovišek *et al.*, 2016; Demir *et al.*, 2017), the assignment of subcellular localisation and the role of proteases in interactions with different pathogens. Despite these limitations, proteases are now well established as important contributors to host defence. Future research addressing their regulation and substrates will undoubtedly produce greater insights into the plant immune system.

Acknowledgments:

Our research is supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1] and ERC Consolidator grant 616449 'GreenProteases', and the University of Oxford.

Conflicts of Interest

The authors declare no conflict of interest.