Proteases and inhibitors in the interaction between *Nicotiana benthamiana* and *Agrobacterium tumefaciens* – systematic analysis and emerging solutions for molecular farming

Friederike Maria Grosse-Holz

Somerville 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

Michelmas 2017

Proteases and inhibitors

in the interaction between *Nicotiana benthamiana* and *Agrobacterium tumefaciens* –

systematic analysis and emerging solutions for molecular farming

Friederike Maria Grosse-Holz

University of Oxford

Supervised by Professor Renier A.L. van der Hoorn

Thesis submitted for the degree of Doctor of Philosophy

Michelmas 2017

Abstract

Nicotiana benthamiana is now an established platform for molecular farming, the production of biopharmaceuticals in plants. Infiltration with *Agrobacterium tumefaciens* (agroinfiltration) is commonly used to transiently express one or multiple transgenes in *N. benthamiana* leaves. Agroinfiltrated *N. benthamiana* is a flexible and scalable recombinant protein (RP) production platform, but is impeded by low RP yields. Plant proteases can degrade RPs and thus limit RP accumulation.

To inform, design and implement strategies for enhancing RP accumulation, I present four papers about proteases and protease inhibitors in agroinfiltrated *N. benthamiana*. First, I investigated the transcriptome, extracellular proteome and active secretome to understand the plant response to agroinfiltration and investigate the expressed proteases. I show that an extracellular immune response is mounted at the expense of photosynthesis. Comprehensive annotation and monitoring uncover a large, diverse repertoire of proteases in agroinfiltrated leaves, indicating that broad-range depletion

of protease activity may be required to enhance RP accumulation. Second, I reviewed the literature on multifunctional plant protease inhibitors (PIs) and grouped them into three types of multifunctional PIs that evolved independently. Third, I screened candidate PIs and discovered that three new, unrelated PIs enhance RP accumulation. I present universal elements of the RP degradation machinery, uncovering new questions on our understanding of the protease network that degrades RPs. Fourth, I identified targets of SlCYS8, a PI that enhances RP accumulation. The target proteases of SlCYS8 are implicated in RP degradation and the high specificity of SlCYS8 can be used to study their role in other processes.

By elucidating the immune response to agroinfiltration, by uncovering the *N. benthamiana* protease repertoire and by providing new tools to deplete the activity of specific proteases, this thesis makes a relevant contribution to both basic plant research and molecular farming.

Acknowledgements

To Professor Renier van der Hoorn for inspiration, guidance and support. I am especially grateful to you for trusting me to develop bioinformatics skills on the job and for countless rounds of valuable, constructive feedback on my writing. This work could not have been achieved without your help.

To my second supervisor Dr Steven Kelly, my role model and mentor in bioinformatics. Thanks for pointing out the right direction, making sure I keep on track and signposting common pitfalls along the way, generally supporting me in learning how to do things myself.

To all members and former members of the Plant Chemetics lab. To Bala for countless tips on ABPP and for insightful discussions, scientifically and otherwise. To Mama Sueldo for babysitting me during my first MS experiment, to Judy for making sure I understand what protein structures are for and for the open access policy, to Luisa for establishing molecular farming in the lab, to Daniel for patiently explaining the organic chemistry behind the probes, to Kyoko for setting standards of flawless experimental design and to Tom for being an inspiration in the office and the most fun and unpretentious professor ever in the lab. To Philippe for inexhaustibly rational, collaborative design and completion of our SlCYS8 project. To Jiorgos for being an awesome partner in crime. To Ane for demonstrating that basically anything in life can be made possible, to Judit for good collaboration both in Maxwell's and in the lab, to Shivani for courage, open-mindedness and hospitability and to Marcel for countless little portions of encouragement. To Pierre, Maria, Tee, Emma, Tram, Amjad, Ziad, Joji, Sophie, Sonja, Nicolas, David, Izzy, Tom, Alice, Ollie, Hayley and Parvinder for the many moments that made going to the lab a fun prospect every morning. To my

students Joel, Awais, Molly and Mary for their contributions to my research and for bearing with me through my less ingenious moments. To Urszula for looking after me and my plants with endless patience and for starting every day with a smile. To Gem for looking after me and my forms with smiles and cake. To Sarah, Caroline, Sam, Sandy and Patricia for running the show behind the scenes.

To Markus Kaiser and the team in Essen, who ran all my peptides on their machines and were always happy to help and explain. Especially to Farnusch Kaschani, who supported me in designing MS experiments and racked his brain over low spectra annotation rates and alternative proteome databases during the analysis of the *N. benthamiana* extracellular proteome.

To Somerville College for making Oxford my home right from the start and supporting my conference travels, to the Studienstiftung for countless inspirational encounters and to the European Research Council for funding the GreenProteases project.

To friends both in Oxford and the rest of the world, for being inspirational, comforting, encouraging, adventurous and generally awesome. Especially to Nadyja and Ben who made Oxford special to me and to Juliette, who is always just there, wherever. To my family for being both the base I stand on and an example I aspire to. To Jan for epistemic standards, reflection and integrity, for inspiration, comfort, encouragement and adventures and for trust, bliss and hope.

My thanks to you all,

Friederike

Table of Contents

Chapter 1

General introduction

1.1 Molecular farming in agroinfiltrated *N. benthamiana* addresses global health challenges of the 21st century

Molecular farming is the production of recombinant proteins (RPs), to be used as biopharmaceuticals, in transgenic plants. After almost 30 years of development since human growth hormone was first produced in transgenic tobacco and sunflower tissues, molecular farming has reached biopharmaceutical markets (Barta *et al.*, 1986). An enzyme produced in carrot cells is licensed for treatment of Gaucher's disease, an influenza vaccine produced in *Nicotiana benthamiana* (wild tobacco) is in phase III clinical trials and a further eight plant-produced RPs are in phases I and II of clinical development (Fox, 2012; Lomonossoff & D'Aoust, 2016; ClinicalTrials.gov NCT03301051, 2017).

In the future, molecular farming could occupy niche markets where plants offer advantages over more conventional expression platforms such as bacteria, mammalian or insect cells. One such niche is the rapid production of novel biopharmaceuticals in response to emerging epidemics. Recent outbreaks of severe acute respiratory syndrome (SARS) and Ebola virus disease (EBV) illustrate how socially and economically devastating epidemics can be (Knobler *et al.*, 2004; UNDG, 2015). Indeed, it is estimated that a severe epidemic like the 19th century "Spanish Flu" would lead to the death of 51 to 81 million people globally (95 % confidence interval) if it reoccurred nowadays (Murray *et al.*, 2006). Research and production facilities for vaccines or drugs to mitigate pandemics are currently insufficient, while threats from bioterrorism increase and new diseases spread faster due to international travel (Fedson, 2003; Yamada *et al.*, 2016; WHO, 2016). Molecular farming in *N. benthamiana* excels in two key qualities needed for a successful response to epidemics: speed and scalability.

1.2 *Agrobacterium tumefaciens* is a unique plant pathogen that became a revolutionary genetic tool

Rapid genetic modification of pre-grown *N. benthamiana* is achieved by agroinfiltration, the infiltration of *Agrobacterium tumefaciens* (Agrobacterium) into leaves. Agrobacterium naturally is a soil-borne, biotrophic plant pathogen that can infect a wide range of dicot and some monocot plant species (Cleene & Ley, 1976). Infection with the wild-type strains of *Agrobacterium* (e.g. strain C58) causes crown gall disease (Braun, 1943). Upon infection, Agrobacterium transfers a single strand of DNA (the T-DNA) into the plant cell nucleus via a specialized structure called the type-IV secretion system (T4SS). This unique case of interkingdom DNA transfer involves intricate interactions of plant proteins with a specialized bacterial machinery of so-called virulence factors or *vir* proteins (reviewed by Gelvin, 2003; Pitzschke and Hirt, 2010). After successful transfer, the T-DNA can either remain separate as a subgenomic DNA molecule or integrate into the plant genome at random sites via a mechanism that is still under investigation. Plant DNA double strand breaks are likely a common site of T-DNA integration in tobacco (Salomon & Puchta, 1998; Chilton & Que, 2003) and plant polymerase θ is required for T-DNA integration in Arabidopsis (Kregten *et al.*, 2016). Genes residing on the T-DNA are expressed *in planta* and the T-DNA of *Agrobacterium* C58 encodes enzymes for the biosynthesis of plant hormones (auxin and cytokinin). Expression of these transgenes triggers formation of tumours in infected plant tissue (Thomashow *et al.*, 1986). Agrobacteria live within the plant tumour and feed on opines that are produced by a T-DNA encoded octopine synthase (Murai & Kemp, 1982).

The only difference between the wild-type, tumorigenic strains and the “disarmed” bacteria used as a genetic tool today is the binary vector system. Wild-type

Agrobacterium carries the T-DNA sequence on a plasmid together with the *vir* genes needed to transform plant cells. In disarmed strains (e.g. *Agrobacterium* GV3101), the tumorigenic T-DNA content has been removed from the *vir* plasmid and the T-DNA is localized on a second, customizable replicon, the binary plasmid (Bevan, 1984). The T-DNA can thus be modified to encode any protein of interest. Agroinfiltrated tissues are transiently transgenic, as the genetic modification occurs in individual cells and is not heritable if reproductive tissues are not infiltrated. Interestingly, T-DNAs can be converted into extrachromosomal, double-stranded circular structures (T-circles) prior to integration into plant genomes (Singer *et al.*, 2012). Extrachromosomal transgenes are expressed *in planta* (Mysore *et al.*, 1998, 2000) and a significant proportion of T-DNAs in agroinfiltrated tissues may be extrachromosomal.

During the intimate contact with plant cells that results in transfer of the T-DNA, Agrobacteria must survive in the plant extracellular space. To this end, Agrobacteria use three strategies: evasion from detection, resilience to toxic compounds and hijacking of host cellular processes. Evasion is possible because Agrobacteria carry a mutation in the conserved bacterial motor protein flagellin. The conserved plant receptor FLS2 (flagellin insensitive 2) recognizes the flagellin-derived epitope flg22 and thus triggers immune responses to most bacteria, but Agrobacteria flg22 is not recognized by FLS2 (Gómez-Gómez *et al.*, 1999).

Resilience is conferred by cytochrome P450 enzymes like Vir2H. Vir2H detoxifies a range of plant-produced benzene ring compounds (phenolics) through O-demethylation that would otherwise hamper bacterial growth (Brencic *et al.*, 2004). Similar enzymes can even break down toxic phenolics into utilizable carbon sources, creating a unique niche for Agrobacterium (Campillo *et al.*, 2014). Finally, hijacking occurs via the bacterial effector VirE2, which coats the T-DNA during translocation into

the host nucleus. VirE2 binds several homologues of the transcription factor VIP1 (VirE2 interacting protein 1). VirE2 thus hijacks VIP1 to transport the T-DNA into the plant nucleus (Citovsky *et al.*, 2004; Shi *et al.*, 2014; Wang *et al.*, 2017).

1.3 *Nicotiana benthamiana* is an ideal host for agroinfiltration-mediated molecular farming

N. benthamiana supports transient transformation by agroinfiltration much better than other species. One reason for this increased susceptibility is that *N. benthamiana* does not perceive the bacterial translation elongation factor EF-Tu because it lacks the corresponding receptor EFR (EF-Tu receptor) (Kunze *et al.*, 2004). EF-Tu perception triggers plant immunity and limits transformation in agroinfiltrated Arabidopsis (Zipfel *et al.*, 2006). Furthermore, *N. benthamiana* carries an insertion in the *RNA-dependent RNA polymerase 1* (*Rdr1*) gene (Bally *et al.*, 2015). Absence of this element of RNAi-mediated defence permits infection with a wide variety of plant viruses, which can be engineered to function as transgene vectors. The first step towards engineered viruses was the *in vitro* generation of viral RNA that could infect plants (Ahlquist *et al.*, 1984). However, plants still had to be wounded to aid viral infection, for instance by dusting the leaves with abrasives such as charcoal or carborundum (Silicon carbide) and then rubbing them with a liquid inoculum (Kalmus, 1945). Infection was simplified greatly when a viral genome was inserted into a T-DNA, facilitating viral infection via agroinfiltration, termed agroinfection (Grimsley *et al.*, 1986). Identification of non-essential sites in viral genomes then allowed the incorporation of transgenes (Dawson *et al.*, 1989). Infection with engineered viruses has fuelled both plant research and

molecular farming through the development of virus-induced gene silencing (VIGS) and replicating RP expression vectors.

VIGS is the infection of plants with an engineered virus carrying a piece of DNA derived from a gene of interest, either in sense or antisense orientation (Zhang *et al.*, 2010; Fernandez-Pozo *et al.*, 2015). As viruses spread to systemically infect the host, the gene of interest will be silenced in the entire plant (Kumagai *et al.*, 1995). VIGS has facilitated the elucidation of defence (Gilroy *et al.*, 2007; Kaschani *et al.*, 2010) and biosynthetic pathways (Sonawane *et al.*, 2016), as well as downregulation of proteases potentially involved in RP degradation (Mandal *et al.*, 2014; Duwadi *et al.*, 2015), generally turning *N. benthamiana* into a powerful tool for reverse genetics (Goodin *et al.*, 2008).

Replicating RP expression vectors based on plant viruses (replicating viral vectors) fall into two main groups: complete viruses that carry an additional transgene and deconstructed viruses, where the genes needed for replication and spreading and the transgene are delivered on separate T-DNA strands by agroinfiltration.

Complete viruses were engineered to form infectious virus particles that present an additional protein on their surface (Donson *et al.*, 1991). Such functionalized plant viruses have been used to produce for instance malaria antigens and an HIV entry inhibitor (Grill *et al.*, 1995; O'Keefe *et al.*, 2009). Although assembly of infectious virus particles *in planta* can boost RP expression, it is also a safety concern and diverts plant resources from RP production.

Deconstructed viral vectors address the concerns associated with complete viruses. In the first viral vectors designed to optimize RP expression, the tobacco mosaic virus (TMV) genome was deconstructed into several vectors, which can be delivered by

mixed agroinfiltration. A GFP-encoding transgene replaced the open reading frame of the viral coat protein, so that no viral particles would be generated (Marillonnet *et al.*, 2004; Gleba *et al.*, 2005). Viral vectors delivered by agroinfiltration were initially not very efficient, requiring large amounts of Agrobacteria to trigger viral genome replication *in planta* (Marillonnet *et al.*, 2004). Infection was limited by the primary step of viral RNA transcription in the plant nucleus and export to the cytosol, where cytoplasmic RNA replication and cell-to-cell spreading could then occur. Viral RNA stability in the plant nucleus and export to the cytosol were improved in modified Tobacco Mosaic Virus (TMV) vectors by deletion of endogenous cryptic splice sites and addition of plant introns (Marillonnet *et al.*, 2005).

Expression of multimeric RPs using replicating viral vectors was problematic at first because of viral competition. Infection of *N. benthamiana* with two versions of TMV encoding different fluorescent proteins led to mosaic patterns in leaves, because the two versions of the virus exclude each other from infected cells (Giritch *et al.*, 2006). This issue can be overcome in two ways. Either, replicating vectors derived from two non-competing viruses can be combined to produce for instance a fully assembled antibody, with the heavy chain (HC) encoding transgene delivered by TMV and the light chain (LC) by Potato Virus X (PVX) (Giritch *et al.*, 2006). The EBV treatment ZMapp, an antibody cocktail under investigation in clinical trials, was manufactured using this strategy (Qiu *et al.*, 2014; ClinicalTrials.gov NCT02363322, 2017; ClinicalTrials.gov NCT02389192, 2017). Another way to avoid viral competition is to abolish replication. The 5′ leader sequence of Cowpea Mosaic Virus (CPMV) RNA-2 has been modified by deletion of an in-frame start codon to boost translation. If an open reading frame is inserted between the modified 5′ leader sequence and 3′ untranslated region of RNA-2 and delivered by agroinfiltration, RP accumulation levels

match those achieved with replicating vectors. Replication does not occur because RNA-1 of CPMV is absent. Thus, multiple transgenes can be expressed without competition. Also, open reading frames can be larger than in replicating viral vectors (Sainsbury & Lomonossoff, 2008; Sainsbury *et al.*, 2009).

The CPMV-based “hypertranslatable”, non-replicating viral vector system has been used to assemble virus-like particles (VLPs) that require production of multiple proteins in the correct stoichiometry (Thuenemann *et al.*, 2013). VLPs can be engineered to display recombinant proteins on their surface (Peyret *et al.*, 2015), modified with dyes, affinity tags and polymers (Wen *et al.*, 2012) or loaded with metals (Sainsbury *et al.*, 2011; Meshcheriakova *et al.*, 2017). Hypertranslation in plants has facilitated the production of polio virus-like particles that do not carry a viral genome. These VLPs are a safer vaccine than the attenuated virus that is currently used and could help eradicate polio completely (Marsian *et al.*, 2017). In summary, viral genome elements boost RP expression through replication or hypertranslation of the RP-encoding mRNA. The combination of viral vectors with efficient transgene delivery by agroinfiltration has greatly advanced *N. benthamiana* as an RP expression platform (Peyret & Lomonossoff, 2015).

Interestingly, the *rdr1* mutation does not only confer unique susceptibility to viruses, but is also associated with fast growth of young *N. benthamiana* plants. Selection pressure in the arid environment of central Australia seems to have favoured *rdr1* mutants, giving rise to an ideal protein expression platform organism that can be quickly grown in large numbers and rapidly transformed (Bally *et al.*, 2015).

1.4 Proteolytic degradation hampers RP accumulation in plants

From the early days of molecular farming in transgenic tobacco, AB fragments have been detected alongside the fully assembled protein (Neve *et al.*, 1993). The first AB produced by agroinfiltration in *N. tabacum* yielded fragments (Vaquero *et al.*, 1999) and when agroinfiltrated *N. benthamiana* became common as a transient expression platform, the problem persisted (Hull *et al.*, 2005). Accumulation of fragments, likely derived from degradation, is most clear with ABs because their size makes AB fragments detectable on Western Blots. However, degradation of smaller proteins has been observed using radioactive pulse-chase experiments (Ohtani *et al.*, 1991) and proteolytic degradation is a limiting factor for human growth hormone production in tobacco cell culture, soy and maize seed (Russell *et al.*, 2005). The glycohormone erythropoietin is degraded upon expression in *N. benthamiana* until it is almost undetectable at ten days post agroinfiltration (Jez *et al.*, 2013). Post-synthesis instability of plant-produced RPs was identified more than ten years ago as a common challenge across molecular farming platform organisms and target proteins (Doran, 2006).

Since then, AB proteolysis has been studied both *in planta* and *in vitro*. Some ABs are more readily degraded than others and this correlates with the length of their respective complementarity determining region 3 (CDR3) loops (Niemer *et al.*, 2014). Cleavage sites outside the CDR3 loop are shared across different ABs and localized between the variable and constant domains of the heavy chain, as well as close to the flexible hinge region (Hehle *et al.*, 2015; Donini *et al.*, 2015). ABs are degraded in the plant extracellular space, but AB fragments are also detected in the cytosol, hinting that both intra- and extracellular enzymes might degrade RPs (Hehle *et al.*, 2011). Mammalian Ser and Cys proteases (Niemer *et al.*, 2014), as well as plant-derived

Papain-Like Cys Proteases (PLCPs) (Paireder *et al.*, 2016, 2017), can degrade ABs *in vitro*.

1.5 Lessons learned from previous attempts to limit RP degradation

Numerous approaches have been tested to prevent proteolysis of plant-produced RPs. Among them are genetic depletion of proteases by knockdown or knockout and depletion of protease activity by either treatment with chemical protease inhibitors or by co-expressing genes encoding protease inhibitors. If multiple approaches targeting the same protease family limit RP degradation, the respective protease family is implicated in the degradation process. We thus describe previous attempts to prevent proteolysis grouped by the target protease family to compile the evidence for involvement of different protease families in RP degradation.

Depletion of endogenous Cys proteases by post-transcriptional gene silencing improved accumulation of both Human Granulocyte Macrophage Colony Stimulating Factor (hGM-CSF) in rice cell cultures and human interleukin (IL10) in tobacco plants (Kim *et al.*, 2008a; Duwadi *et al.*, 2015). Overexpression of Cys protease inhibitors also increased RP accumulation in two cases. First, transgenic tobacco plants overexpressing the inhibitor Oryzacystatin-I contained higher levels of recombinant glutathione reductase produced by agroinfiltration when compared to wild-type tobacco (Pillay *et al.*, 2012). Second, co-expression of *Solanum lycopersicum* Cystatin 8 (SlCYS8) increased the accumulation of ABs in agroinfiltrated *N. benthamiana* (Robert *et al.*, 2013; Jutras *et al.*, 2016). Plant Cys proteases are thus likely responsible for degradation of different RPs across expression platforms.

Plant Ser proteases also appear to be involved in degradation because overexpression of Ser protease inhibitors can increase RP accumulation. Co-expression with Bowman-Birk serine protease inhibitor increases AB accumulation in *N. tabacum* roots (Komarnytsky *et al.*, 2006) and co-expression with a *N. alata* Ser protease inhibitor (Protease inhibitor II) increases accumulation of Human Granulocyte Macrophage Colony Stimulating Factor (hGM-CSF) in rice cell cultures (Kim *et al.*, 2008b).

Although the evidence is strongest for Cys and Ser proteases, plant aspartic and Metalloproteases are also implicated in RP degradation. The Asp protease inhibitor *Solanum lycopersicum* Cathepsin D inhibitor (SlCDI) increases accumulation of α_1-anti-chymotrypsin in transgenic potato leaves (Goulet *et al.*, 2010a). SlCDI co-expression in agroinfiltrated *N. benthamiana* also increased accumulation of ABs (Goulet *et al.*, 2012), although this seems to depend on plant growth conditions (Robert *et al.*, 2013). Plant metalloprotease depletion or inhibition has not been shown to be sufficient for increasing RP accumulation *in vivo*. However, *in vitro* degradation of both *Desmodus rotundus* Salivary Plasminogen Activator α1 (DSPAα1) in spent tobacco cell culture medium and of Bovine Serum Albumin (BSA) in tobacco leaf intercellular fluid has been impaired by addition of ethylenediaminetetraacetic acid (EDTA). EDTA chelates metal ions that are pivotal for metalloprotease activity. These results indicate that plant metalloproteases can degrade BSA and DSPAα1 *in vitro* (Schiermeyer *et al.*, 2005; Delannoy *et al.*, 2008).

In summary, proteases of all catalytic classes seem to contribute to RP degradation, with most data available on Cys and Ser proteases. Remarkably, degradation appears to be equally complex across RPs and expression platforms. This leads to low

accumulation levels of plant-produced RPs, resulting in greater efforts needed to grow and process enough plant biomass for industrial RP production.

1.6 Increasing RP accumulation is a key lever to increase the economic viability of molecular farming

For the full potential of molecular farming to be realized, plant-based RP expression must be economically attractive for companies to pursue. As an RP expression platform for complex glycoproteins, agroinfiltrated *N. benthamiana* competes directly with the leading mammalian cell culture system, Chinese Hamster Ovary (CHO) cells. Plants could outcompete CHO cells based on COGS (cost of goods sold) for ABs if accumulation levels of 1 g/kg FW (biomass fresh weight) were reached *in planta* (Nandi *et al.*, 2016). AB accumulation levels *in planta* are a key lever to decrease COGS, as both upstream (plant growth and agroinfiltration) and downstream (purification) costs decrease with increases in AB accumulation. Other main drivers of COGS are electricity, purification resin and labour cost, which are difficult to influence (Nandi *et al.*, 2016). Techno-economic analyses for other RPs yield similar conclusions (Buyel *et al.*, 2015b). However, accumulation levels of 0.1 - 0.3 g/kg FW are frequently reported for plant-produced ABs (Yusibov *et al.*, 2016), suggesting that increasing RP accumulation is a promising strategy to make molecular farming cost-competitive.

1.7 A comprehensive understanding of *N. benthamiana* proteases and responses to agroinfiltration is needed

To address the complex problem of RP degradation, a comprehensive overview of proteases in agroinfiltrated *N. benthamiana* leaves is needed. Based on which proteases are found, the activity of complete protease (sub-) families can then be depleted specifically and simultaneously. The emerging genome engineering technique based on Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) that are recognized and cut by the endonuclease Cas9 could facilitate genetic protease depletion. Cas9 can be equipped with a custom RNA sequence to direct it towards virtually any target gene, where Cas9-mediated cleavage and error-prone DNA repair then lead to mutations. CRISPR-Cas9 based techniques facilitate the generation of knockouts quickly, in non-model organisms and without the introduction of transgenes (Puchta, 2017). Furthermore, the interaction between *N. benthamiana* and *Agrobacterium* deserves more focused attention. Microarrays of *A. thaliana* suspension cells exposed to *Agrobacterium* or stalks inoculated by wounding indicate that the levels of transcripts encoding proteins associated with plant immunity increase as early as three hours upon exposure to *Agrobacterium* (Ditt *et al.*, 2006; Lee *et al.*, 2009). Different studies of the transcriptional response to *Agrobacterium* disagree regarding the extent and timing of this response (Jiang *et al.*, 2003; Ditt *et al.*, 2006; Lee *et al.*, 2009). No study was so far conducted on agroinfiltrated *N. benthamiana.* The extracellular proteome of agroinfiltrated *N. benthamiana* has been analysed using 2-dimensional gel electrophoresis. At six days post infiltration, chitinases, glycosidases and other pathogenesis-related (PR) proteins accumulated, but the timing of this response is unresolved (Goulet *et al.*, 2010b). An immune response to agroinfiltration may divert resources from RP production and could even

increase protease activity, like the tomato immune response to fungal infection (Esse *et al.*, 2008; Sueldo *et al.*, 2014). A comprehensive, time-resolved understanding of the *N. benthamiana* response to agroinfiltration and the proteases involved is needed to inform design of systematically optimized plants for molecular farming. Omics-level analyses of agroinfiltrated *N. benthamiana* are a timely endeavour, as the tools they require have recently emerged.

1.8 Transcriptomics and proteomics tools are rising to the challenge of *N. benthamiana*

N. benthamiana is an allotetraploid within the section Suavolentes of the nightshade family (Solanaceae). The *N. benthamiana* genome is now believed to have arisen from parents from the Sylvestres, Petunioides and Noctiflorae and been complicated by gene flow between the diploid parents and other members of the Suavolentes (Kelly *et al.*, 2013). The majority of the ~ 3 Gb *N. benthamiana* genome (n = 19 chromosomes) has been sequenced by two consortia. However, the sequence is still fragmented into 56,094 or 147,949 scaffolds with average lengths of 52,943 or 17,233 nucleotides (Bombarely *et al.*, 2012; Nakasugi *et al.*, 2014; BTI, 2017; QUT, 2017). A complex, incomplete genome makes both transcriptome and proteome studies challenging. In both cases, short sequences of either nucleotide reads (obtained by RNA sequencing, RNAseq) or peptides (obtained from mass spectrometry, MS) need to be mapped to longer sequences of either transcripts or proteins to estimate abundances of these transcripts or proteins, respectively. Therefore, the quality of the transcriptome, from which the proteome is then predicted, directly determines the outcome of RNAseq and MS experiments.

Recent bioinformatic developments facilitate transcriptome and proteome analyses on incompletely sequenced species and with reasonable resource input. Omics-level analysis of *N. benthamiana* is therefore possible now. This is illustrated by recently published transcriptomics studies elucidating transcription factor networks (Bond *et al.*, 2016) and defence responses to infections with fungi, viruses and oomycetes in *N. benthamiana* (Faino *et al.*, 2012; Evangelisti *et al.*, 2017; Geng *et al.*, 2017). Three specific developments that were prerequisites to the analyses of agroinfiltrated *N. benthamiana* in this thesis will be briefly described.

First, efficient data compression has improved the speed, accuracy and completeness of *de novo* transcriptome assembly (Simpson & Durbin, 2012). In general, transcriptomes can be assembled by overlapping sequence reads with each other (*de novo* assembly) or by mapping reads to a genome sequence (genome-based assembly). Simulated RNAseq datasets with known correct transcript sequences are used to compare transcriptome assemblers. Recent advances included improved data compression and the development of algorithms that can work directly with (subsets of) compressed data. StringTie, a state-of-the-art assembler, employs efficient *de novo* assembly of reads into “super reads”. These are mapped to the genome and used for transcript prediction together with a traditional mapping of the reads from algorithms like TopHat (Kim *et al.*, 2013b; Pertea *et al.*, 2015). The upstream *de novo* step of StringTie facilitates assembly of novel transcripts, making transcriptome assembly more accurate and less dependent on a correct genome sequence.

Second, new algorithms for transcript quantification and differential abundance analysis are computationally efficient and relatively straightforward to use. The transcriptome in this thesis was assembled using RNAseq reads at any location they could have originated from. Thus, reads may be re-used to ensure transcript

completeness. For accurate transcript quantification, however, each read must only be used once. A new, unambiguous alignment to the transcriptome is made by the transcript quantification algorithm Salmon. Salmon assigns reads to transcripts using short exact sequence matches (lightweight alignment). This lightweight alignment is based on a pre-processed index of the transcriptome and is thus highly parallelizable, rendering Salmon one of the fastest currently available transcript quantification algorithms (Patro *et al.*, 2015, 2017). Differentials in transcript abundance are then determined by the R package DESeq2. DESeq2 uses statistical methods tailored to the strengths and weaknesses of RNAseq data to determine differentials in transcript abundance. For instance, fold changes in transcript abundance determined from very low values are “shrunken” towards zero. This reflects the low information content of differences between very small transcript abundances (non-normality/heteroskedasticity of transcript counts) and makes the output more biologically relevant (Love *et al.*, 2014). The speed of Salmon and accessibility of DESeq2 facilitated time-resolved quantitative transcriptome analysis of *N. benthamiana* on a desktop PC and within a DPhil project.

Third, quantitative proteomics no longer requires feeding the target organism specific isotopes for heavy/light labelling or use of expensive isobaric tag labelling reagents. Traditional mass spectrometry (MS) pipelines needed peptide spectra from different samples to be obtained in the same MS run to quantify protein abundance changes between those samples. Samples were thus mixed before MS and labelling with heavy/light isotopes or isobaric tags facilitated distinction between peptides from different samples. Label-free quantification of proteins using one MS run per sample is facilitated by an algorithm termed MaxLFQ. MaxLFQ estimates protein abundances by comparing the intensities of peptide mass spectra between samples. This yields

relative abundance estimates, which are then rescaled to the summed spectral intensity for all peptides derived from a protein across samples (Cox *et al.*, 2014). MaxLFQ, which is part of the open-source MaxQuant platform (Tyanova *et al.*, 2016a), thus makes large-scale quantitative proteomics analyses of soil-grown plants affordable.

1.9 The aim of this thesis is to inform, design and implement strategies for enhancing RP accumulation in agroinfiltrated *N. benthamiana*.

Chapter 2 is aimed at elucidating **the plant response to the presence of *A. tumefaciens*** and the **protease repertoire of agroinfiltrated *N. benthamiana* leaves**. Time-resolved analyses of the transcriptome and extracellular proteome uncovered an immune response to agroinfiltration that is mounted at the expense of photosynthesis. Integrating transcriptomics, proteomics and activity-based proteomics data revealed that the response to agroinfiltration is associated with increased efficiency of extracellular protein delivery. As a basis for further work on *N. benthamiana* proteases, six large *N. benthamiana* protease families were manually curated. 975 protease and protease-homolog encoding transcripts, 196 extracellular proteases and 17 active extracellular Ser and Cys proteases were monitored in agroinfiltrated leaves, highlighting how large and diverse the protease repertoire of agroinfiltrated *N. benthamiana* is. This study was published in Plant Biotechnology Journal under the title "*The transcriptome, extracellular proteome and active secretome of agroinfiltrated N. benthamiana uncover a large, diverse protease repertoire*" (Chapter 2). This chapter demonstrates that agroinfiltrated leaves contain many different proteases, which could act redundantly to degrade RPs.

The aim of Chapter 3 and 4 is to **identify protease inhibitors that can be co-expressed with RPs to enhance RP accumulation**. To this end, a thorough analysis of the literature on multifunctional plant protease inhibitors (PIs) was conducted, resulting in a review article published in the New Phytologist under the title "*Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants*" (Chapter 3). Based on the research summarized in this review and on further mining of the literature on PIs, a systematic screen of 29 candidate PIs was designed. Candidate

PIs were co-expressed with three unrelated RPs in agroinfiltrated *N. benthamiana* and RP accumulation was monitored. The screen yielded three unrelated PIs that enhance RP accumulation upon co-expression, both separately and in combination with each other. Effects of these three PIs on leaves were characterized using ABPP and quantitative MS, to elucidate how PI co-expression prevents RP degradation. This dataset is written up for publication under the title "*Three unrelated protease inhibitors enhance accumulation of pharmaceutical RPs in N. benthamiana*" (Chapter 4).

The aim of Chapter 5 is to **identify proteases involved in RP degradation *in planta***. To this end, target proteases of a protease inhibitor that enhances RP accumulation upon co-expression were identified using ABPP and ABPP-MS. *Solanum lycopersicum* Cystatin 8 (SlCYS8) specifically inhibits four of the nine subfamilies of Papain-like Cys proteases in *N. benthamiana*. This knowledge makes SlCYS8 a powerful tool that revealed the roles of different PLCP subfamilies in RP degradation and is applicable to wider plant biology in the future. The characterisation of SlCYS8 is written up for publication under the title "Activity-based proteomics reveals target proteases of tomato cystatin SlCYS8 in agroinfiltrated *N. benthamiana*" (Chapter 5).

Results and perspectives from the four manuscripts are discussed jointly in Chapter 6.

Chapter 2

The transcriptome, extracellular proteome and active secretome of agroinfiltrated *N. benthamiana* uncover a large, diverse protease repertoire

Friederike Grosse-Holz, Steven Kelly, Svenja Blaskowski, Farnusch Kaschani, Markus Kaiser, Renier A.L. van der Hoorn. 2017. The transcriptome, extracellular proteome and active secretome of agroinfiltrated *N. benthamiana* uncover a large, diverse protease repertoire. *Plant Biotechnol. J.* **in press**.

Authors

Friederike Grosse-Holz[1], Steven Kelly[2], Svenja Blaskowski[3], Farnusch Kaschani[3], Markus Kaiser[3], Renier A.L. van der Hoorn[1]*

Affiliations

[1] Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[2] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[3] Chemische Biologie, Zentrum für Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

* Corresponding author

Running title

Proteases in agroinfiltrated *N. benthamiana*

Author contributions

R.H. conceived the research. F.G.H. designed and performed experiments and analysed data unless specified otherwise. F.G.H. and R.H interpreted the results. S.K. provided guidance for RNAseq experimental design, transcriptome assembly and differential expression analyses and performed the de novo transcriptome assembly and coding sequence prediction. S.B. performed the sample preparation for MS. F.K. and M.K. performed LC-MS/MS and peptide and protein identification steps for MS of extracellular proteomes and ABPP-MS. F.G.H wrote the manuscript with feedback from R.H. All authors read and approved the final manuscript.

2.1 Summary

Infiltration of disarmed *Agrobacterium tumefaciens* into leaves of *Nicotiana benthamiana* (agroinfiltration) facilitates quick and safe production of antibodies, vaccines, enzymes and metabolites for industrial use (Molecular Farming). However, yield and purity of proteins produced by agroinfiltration are hampered by unintended proteolysis, restricting industrial viability of the agroinfiltration platform. Proteolysis may be linked to an immune response to agroinfiltration, but understanding of the response to agroinfiltration is limited. To identify the proteases, we studied the transcriptome, extracellular proteome and active secretome of agroinfiltrated leaves over a time course, with and without the P19 silencing inhibitor. Remarkably, P19 expression had little effect on the leaf transcriptome and no effect on the extracellular proteome. 25% of the detected transcripts changed in abundance upon agroinfiltration, associated with a gradual upregulation of immunity at the expense of photosynthesis. By contrast, 70% of the extracellular proteins increase in abundance, in many cases associated with increased efficiency of extracellular delivery. We detect a dynamic reprogramming of the proteolytic machinery upon agroinfiltration by detecting transcripts encoding for 975 different proteases and protease-homologs. The extracellular proteome contains peptides derived from 196 proteases and protease homologs, and activity-based proteomics displayed 17 active extracellular Ser and Cys proteases in agroinfiltrated leaves. We discuss unique features of the *N. benthamiana* protease repertoire and highlight abundant extracellular proteases in agroinfiltrated leaves, being targets for reverse genetics. This dataset increases our understanding of the plant response to agroinfiltration and indicates ways to improve a key expression platform for both plant science and Molecular Farming.

2.2 Introduction

Agroinfiltration of *Nicotiana benthamiana* (a relative of tobacco) is widely applied to transiently express proteins, either as biopharmaceutcials, for other industrial use or to study their functions. Agroinfiltration is based on the transient genetic manipulation of leaves by infiltration with disarmed *Agrobacterium tumefaciens* (Agrobacterium) carrying gene(s) of interest on the transfer DNA (T-DNA) of binary plasmid(s) (Bevan, 1984). Agrobacterium delivers the T-DNA to the nucleus of its host plant, where genes are expressed within a few days upon agroinfiltration. Co-expression of several transgenes is simply achieved by mixing Agrobacterium cultures delivering these different transgenes before agroinfiltration. Co-expression with silencing inhibitor P19 is frequently used to boost protein overexpression by preventing the decline of the transgene transcript levels (van der Hoorn *et al.*, 2003).

The versatility and potential of agroinfiltration are illustrated by many use cases. For instance, production of biopharmaceuticals (Molecular Farming) (Stoger *et al.*, 2014) in agroinfiltrated *N. benthamiana* offers speed, scalability and low risk of contamination with human pathogens when compared to classical insect or mammalian cell culture systems. An agroinfiltration-based expression platform can now deliver ten million doses of the latest influenza vaccine within a record time of six weeks (Pillet *et al.*, 2016). Large-scale agroinfiltration has also produced many different functional monoclonal antibodies (Yusibov *et al.*, 2016), including the Ebola-neutralizing drug ZMapp (Qiu *et al.*, 2014). Transient, spatially restricted overexpression of synthetic biology building blocks can shift plant secondary metabolism towards valuable products with minor impact on fitness (Nielsen *et al.*, 2013). Along similar lines, pathogen-derived effectors that would likely have severe phenotypic effects if expressed in stable lines have been studied by agroinfiltration (Bos *et al.*, 2006;

Dagdas *et al.*, 2016; Petre *et al.*, 2016). Speed and simplicity of agroinfiltration are leveraged for high-throughput screening of fluorescently tagged proteins to study their subcellular localization (Martin *et al.*, 2009).

Although agroinfiltration is a widely-used tool, remarkably little is known about how *N. benthamiana* responds to agroinfiltration. Agrobacterium elicits immune responses, including the induction of pathogenesis-related (PR) genes and the accumulation of extracellular PR proteins (Goulet *et al.*, 2010b; Pitzschke, 2013; Zhou *et al.*, 2017). As in other plants, this immune response reduces subsequent pathogen infections (Robinette & Matthysse, 1990; Rico *et al.*, 2010; Sheikh *et al.*, 2014; Li *et al.*, 2017), and may limit transgene delivery. Transgene delivery in older, flowering *N. benthamiana* is limited due to the perception of Agrobacterium cold shock protein (Saur *et al.*, 2016). In younger plants, which are used for agroinfiltration, responses are elusive. Furthermore, the impact of silencing inhibitor P19 on the response to agroinfiltration and its timing are unresolved.

We focus on extracellular proteases, as they may limit the accumulation of recombinant proteins (RPs) passing through the secretory pathway to become glycosylated. Proteolytic degradation is a bottleneck on the way to industrial viability of agroinfiltration (Mandal *et al.*, 2016). Indeed, RP degradation can occur in the extracellular space (Hehle *et al.*, 2011) and proteolysis hampers yield and purity of biopharmaceuticals produced in *N. benthamiana* (Niemer *et al.*, 2014; Mandal *et al.*, 2014; Hehle *et al.*, 2015). Papain-like Cys proteases can degrade RPs *in vitro* (Paireder *et al.*, 2016, 2017), but the proteases degrading RP *in planta* are unidentified. Extracellular proteases commonly accumulate in leaves during immune responses. The extracellular tomato Ser protease P69 and Cys proteases Pip1 and Rcr3, for example, accumulate upon infection with viroids, oomycetes, fungi and

bacteria (Jordá *et al.*, 1999; Tian *et al.*, 2004; Kaschani *et al.*, 2010). Transcripts and proteins corresponding to proteases also accumulate in Arabidopsis infected with *Pseudomonas* (Zhao *et al.*, 2003; Xia *et al.*, 2004), and extracellular Ser and Cys protease activities increase in tomato upon fungal infection with *Cladosporium fulvum* (Esse *et al.*, 2008; Sueldo *et al.*, 2014). These examples indicate that activity and/or abundance of extracellular proteases, especially Ser and Cys proteases, may increase in *N. benthamiana* upon agroinfiltration, linking proteolytic RP degradation to plant immunity. Therefore, both comprehensive annotation of the *N. benthamiana* protease repertoire and improved understanding of the response to agroinfiltration are needed to limit undesired proteolysis. RP accumulation has been increased by depleting proteases by knockdown in rice cell cultures (Kim *et al.*, 2008a) and in *Nicotiana tabacum* (Mandal *et al.*, 2014; Duwadi *et al.*, 2015) and protease inhibitor overexpression in *N. benthamiana* (Goulet *et al.*, 2012; Sainsbury *et al.*, 2013). These studies indicate that once targets are identified, protease depletion could improve agroinfiltrated *N. benthamiana* as a protein expression platform.

Here, we investigated how RP production may be affected by the immune response to agroinfiltration, especially immune proteases. Time-resolved leaf transcriptome and extracellular proteome datasets of agroinfiltrated leaves revealed an immune response that is mounted at the expense of photosynthesis and not affected by P19. We analysed the exceptionally large *N. benthamiana* protease repertoire in the context of other plant proteases and identified active Ser and Cys proteases. Taken together, the data will advance strategies to improve transient protein expression by engineering plant immunity and depleting proteases.

2.3 Results and Discussion

To characterize agroinfiltrated *N. benthamiana* leaves, we infiltrated *N. benthamiana* leaves with wild-type *A. tumefaciens* GV3101-pMP90 (no binary vector, WT), Agrobacterium P19 (T-DNA encoding viral silencing suppressor P19 (Chapman *et al.*, 2004), P19) or buffer (mock treatment). We took samples at two, five, seven and ten days post infiltration (dpi) and quantified transcripts, extracellular proteins and extracellular protein activity using RNAseq, label-free quantification mass spectrometry (MS) and activity-based proteomics (Activity-based Protein Profiling coupled to Mass Spectrometry, ABPP-MS) (Figure 2.1).

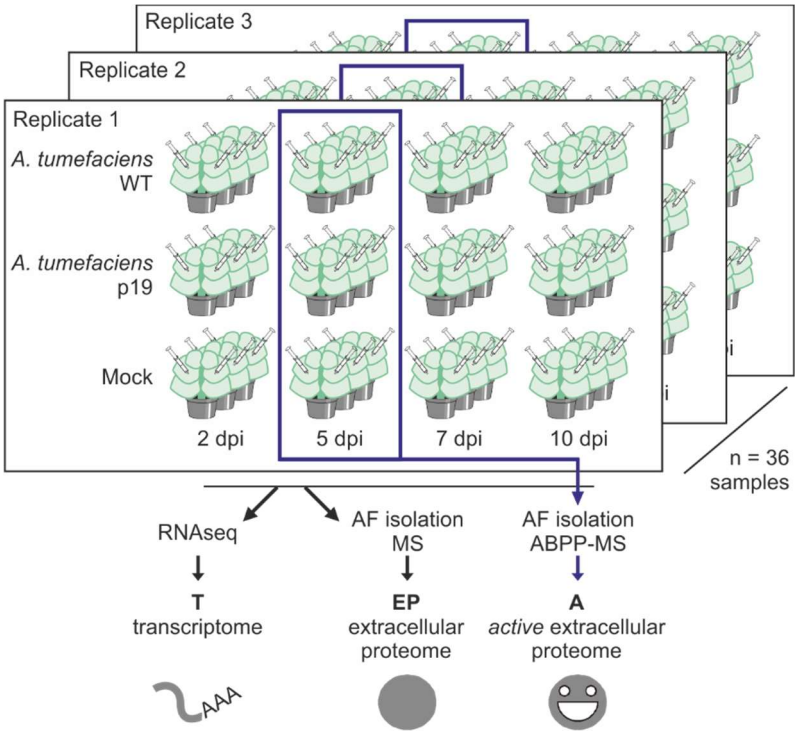


Figure 2.1: Experimental setup. Leaves of *N. benthamiana* were infiltrated with *Agrobacterium* GV3101-pMP90 without any T-DNA plasmid (WT) or carrying a plasmid for p19 expression (p19) or with buffer (mock). Abbreviations: dpi, days post infiltration; RNAseq, mRNA sequencing; AF, apoplastic fluid; MS, protein mass spectrometry; ABPP, Activity-based Protein Profiling.

During a first annotation of the transcriptome and proteome data, we observed that well-known proteases including Papain-Like Cys Proteases (PLCPs, MEROPS family C01) and subtilases (family S08) often appeared truncated or lacked conserved domains in the Niben101 proteome database (https://solgenomics.net/). To obtain a database with protease families that are adequately annotated for evaluation of transcriptomics and proteomics experiments, we compared four *N. benthamiana* proteome databases and manually curated the proteases in the best database (described in detail in Supporting File S01). Searching the extracellular proteome MS spectra with our curated proteome, we identified peptides corresponding to 30 proteins more than with the best published database, showing that the curation improved interpretation of experimental data (Supporting File 2. S01).

2.3.1 The *N. benthamiana* response to agroinfiltration

2.3.1.1 The P19 silencing suppressor has minor effects on the transcriptome and no effect on the extracellular proteome of N. benthamiana

To assess how *N. benthamiana* responds to agroinfiltration and how silencing suppression affects these responses, we sequenced mRNA from WT agroinfiltrated, P19 agroinfiltrated and mock infiltrated leaves. Euclidean distance clustering revealed that transcriptomes from agroinfiltrated samples cluster together by time point regardless of whether WT or P19 bacteria were present (Figure 2a). Surprisingly, only 0.75 % of all detected transcripts (569/75802) differed significantly in abundance at any time point between WT and P19 agroinfiltrated leaves (Supporting Table ES2.01). Among the differentials is the transcript encoding P19, which was very abundant up to 7 dpi and slightly decreased in abundance at 10 dpi, potentially because older leaves are less transcriptionally active (Supporting Figure 2.S01). Transcripts encoding components of the silencing machinery such as members of the Argonaute PFAM

family were significantly enriched among the transcripts with differential abundance between P19 and WT agroinfiltrated leaves, but most of the differential transcripts (404 out of 569) are not annotated (Supporting Figure 2.S01 and Table ES2.02). There were no significant differences between extracellular proteomes from WT and P19 agroinfiltrated leaves at any time point (Supporting Table ES2.03). We thus compare agroinfiltrated (WT and P19) to mock infiltrated leaves for further analysis.

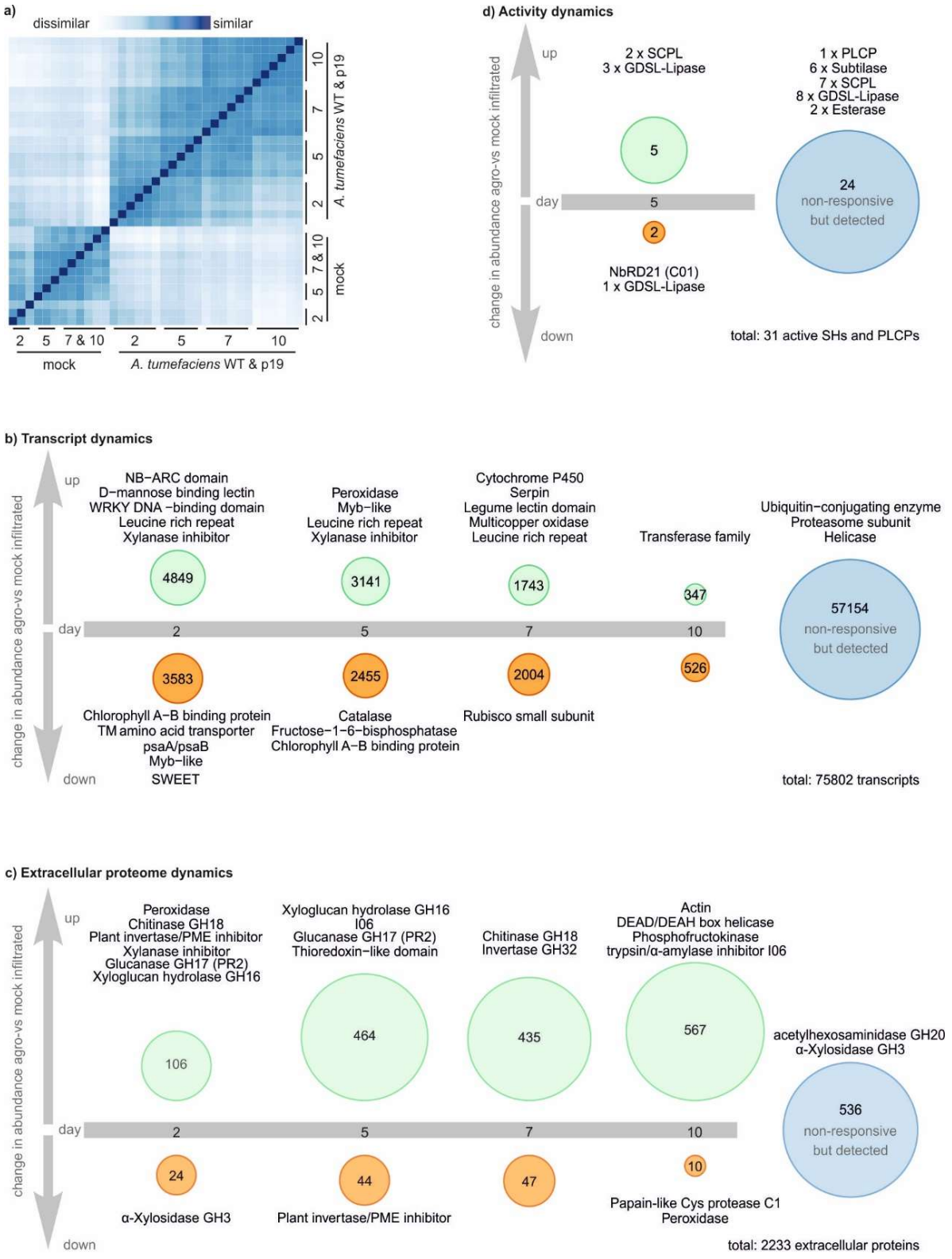

a)
dissimilar
similar
2 5 7 & 10 2 5 7 10
mock
A. tumefaciens WT & p19
A. tumefaciens WT & p19
mock
d) Activity dynamics
up
change in abundance agro-vs mock infiltrated
2 x SCPL
3 x GDSL-Lipase
1 x PLCP
6 x Subtilase
7 x SCPL
8 x GDSL-Lipase
2 x Esterase
5
day
5
24
non-responsive
but detected
2
NbRD21 (C01)
1 x GDSL-Lipase
down
total: 31 active SHs and PLCPs
b) Transcript dynamics
up
change in abundance agro-vs mock infiltrated
NB−ARC domain
D−mannose binding lectin
WRKY DNA −binding domain
Leucine rich repeat
Xylanase inhibitor
Peroxidase
Myb−like
Leucine rich repeat
Xylanase inhibitor
Cytochrome P450
Serpin
Legume lectin domain
Multicopper oxidase
Leucine rich repeat
Transferase family
Ubiquitin−conjugating enzyme
Proteasome subunit
Helicase
4849
3141
1743
347
57154
non-responsive
but detected
day
2
5
7
10
3583
2455
2004
526
Chlorophyll A−B binding protein
TM amino acid transporter
psaA/psaB
Myb−like
SWEET
Catalase
Fructose−1−6−bisphosphatase
Chlorophyll A−B binding protein
Rubisco small subunit
down
total: 75802 transcripts
c) Extracellular proteome dynamics
up
change in abundance agro-vs mock infiltrated
Peroxidase
Chitinase GH18
Plant invertase/PME inhibitor
Xylanase inhibitor
Glucanase GH17 (PR2)
Xyloglucan hydrolase GH16
Xyloglucan hydrolase GH16
I06
Glucanase GH17 (PR2)
Thioredoxin−like domain
Chitinase GH18
Invertase GH32
Actin
DEAD/DEAH box helicase
Phosphofructokinase
trypsin/α-amylase inhibitor I06
106
464
435
567
acetylhexosaminidase GH20
α-Xylosidase GH3
536
non-responsive
but detected
day
2
5
7
10
24
44
47
10
α-Xylosidase GH3
Plant invertase/PME inhibitor
Papain-like Cys protease C1
Peroxidase
down
total: 2233 extracellular proteins

Figure 2.2: The immune response to agroinfiltration entails an increase of the extracellular proteome and is not affected by p19 overexpression. a) Euclidean sample distances between the transcriptomes obtained from all 36 samples. Samples were ordered by hierarchical clustering based on the sample distances. b-c: Transcripts (b) or proteins (c) were grouped by when their abundance first changed significantly (Wald test for transcripts, Student's t-test for proteins; Benjamini-Hochberg (BH) adjusted $p < 0.05$) and more than two-fold in (WT and p19) agroinfiltrated samples compared to mock infiltrated samples. Annotations given above the circles are representatives of the PFAM families that are significantly (Hypergeometric test, BH-adjusted $p < 0.05$) overrepresented in the respective regulatory category compared to all detected transcripts (b) or proteins (c). Protein groups for which corresponding peptides were identified are counted as one protein. d) Activity of extracellular PLCPs and Ser hydrolases was assayed by ABPP-MS at 5 dpi, counting each protein group for which peptides were identified as one active protein. Proteins were grouped by whether they were enriched in the WT agroinfiltrated samples, controls or both (t-test probe sample vs no-probe control, BH-adjusted $p < 0.1$). Differences in abundance between agroinfiltrated samples and controls were not significant in any case. Only proteins annotated as SHs or PLCPs are included in the figure. Full datasets are given in Supporting Tables ES2.04 and ES2.05 (a&b), ES2.06 and ES2.07 (c) and ES2.11 (d). The R code to generate the figures is given in Supporting Files ES2.02 (a&b), ES2.03 (c) and ES2.05 (d).

2.3.1.2 Agroinfiltration induces leaf transcriptome changes associated with immune responses

Of all detected transcripts (n = 75802), 24.6 % (n = 18648) significantly changed more than two-fold in abundance at any time point and were thus considered differential in abundance. Among the differentials, the biggest category (n = 4849) is that of transcripts increasing in abundance for the first time at 2 dpi (6.4 % of the transcriptome) (Figure 2.2b, Table ES2.04, File ES2.02). In this category, transcripts encoding proteins associated with immunity are overrepresented (representatives in Figure 2.2b, complete lists in Table ES2.05). This includes transcripts encoding LRR (leucine-rich repeat) domain containing receptors such as the recently identified receptor for Agrobacterium cold-shock protein NbCSPR (Niben101Scf03240g00007) (Saur *et al.*, 2016), as well as signalling components carrying NB-ARC (nucleotide-binding adaptor shared by Apaf-1, Resistance proteins, and CED-4) and WRKY

domains. Among the categories of transcripts whose abundance first increases at 5 or 7 dpi, transcripts encoding Myb transcription factors and serpins, LRRs and xylanase inhibitors are overrepresented. Besides transcripts encoding proteins associated with immune signalling and first-line defence, we detected a 4.5-fold average decrease in abundance of 13 transcripts encoding SWEET sugar efflux transporters, which may decrease the nutrient content of the extracellular space to control bacterial growth (Chen, 2014). Differential abundance of transcripts encoding both generators and quenchers of reactive oxygen species, as well as increased accumulation of transcripts encoding cytochrome P450 enzymes, show that the plants are stressed upon agroinfiltration. Transcripts encoding members of the photosynthetic machinery and assimilatory metabolism in general are enriched among the transcripts decreasing in abundance from 2 dpi onwards, explaining the chlorotic phenotype of agroinfiltrated leaves (Pruss *et al.*, 2008). Among the transcripts detected constantly, transcripts encoding for housekeeping proteins like members of the ubiquitin-proteasome system and helicases are overrepresented. In summary, agroinfiltration is associated with an immune response mounted at the expense of photosynthesis.

2.3.1.3 Diversity and abundance of extracellular proteins increase upon agroinfiltration

We evaluated the effect of agroinfiltration on the extracellular proteome of *N. benthamiana* because the leaf extracellular space is the target site for glycoprotein accumulation in Molecular Farming, as well as the primary site of interaction with Agrobacterium and thus a promising site for improvement of the transient expression platform. Of all *N. benthamiana* proteins for which we identified extracellular peptides (n = 2233 protein groups as defined by MaxQuant (Tyanova *et al.*, 2016a)), the vast majority (n = 1697, 75.9 %) changed significantly and more than two-fold in abundance

and were thus considered differential in abundance. Among these differentials, most (n = 1572, 92.6 %) increased in abundance upon agroinfiltration (Figure 2.2c, Table ES2.06, File ES2.03). The increase of the extracellular proteome was mirrored by a corresponding increase in protein concentration in apoplastic fluid (AF) from agro- but not mock infiltrated samples (Supporting Figure 2.S02). Abundant intracellular housekeeping proteins such as actin, helicases and phosphofructokinases are overrepresented in the category of proteins that first increased in abundance at 10 dpi (n = 567), indicating that the interaction between *N. benthamiana* and Agrobacterium leads to cell content leakage at this late stage. Leakage may occur *in vivo* and during apoplastic fluid extraction.

Among proteins that first increased in abundance at 2 dpi and 5 dpi, hydrolytic enzymes and inhibitors are overrepresented (Figure 2.2c and Table ES2.07). This includes classical defence proteins such as xylanase inhibitors, chitinases (GH18) and pathogenesis-related protein 2 (PR2, a GH17 glucanase) (Cosgrove, 2016). PR2 accumulation upon agroinfiltration is consistent with an earlier study (Goulet *et al.*, 2010b). Cell-wall remodelling xyloglucan endotransglycolases/hydrolases (GH16) and versatile I06 α-amylase/Ser protease inhibitors may contribute indirectly to plant defence. Family C01 proteases (PLCPs) are overrepresented in the small category of proteins that first decreased in abundance at 10 dpi. Thus, *N. benthamiana* extracellular PLCPs do not increase as strongly and persistently in abundance upon agroinfiltration as tomato extracellular PLCPs do upon pathogen challenge (Esse *et al.*, 2008). In contrast, PLCPs may localize to intracellular compartments as shown for RD21 in Arabidopsis (Hayashi *et al.*, 2001) or may be degraded in the extracellular space. Invertase/pectin methyl esterase (PME) inhibitors are overrepresented both in the category of proteins first increasing at two and in the category of proteins first

decreasing at 5 dpi in abundance. Invertase inhibition upon agroinfiltration thus appears to be transient and indeed, invertases (GH32) are overrepresented in the category of proteins first increasing in abundance at 7 dpi. Plant Invertases cleave the transport sugar sucrose, providing vital nutrients to sink tissues (Goetz *et al.*, 2001). The chlorotic agroinfiltrated leaves may be less photosynthetically active and lose nutrients to the bacteria, turning them from a source into a sink organ. Some members of the GH32 family degrade extracellular polysaccharides from pathogens (Limoli *et al.*, 2015), suggesting that GH32 family members may promote both nutrition and defence in agroinfiltrated leaves. GH3 α-xylosidases are overrepresented in both the category of proteins first decreasing in abundance at 2 dpi and in the category of proteins with constant abundance. Some GH3 family members act in cell wall remodelling and others locally adjust auxin concentrations as auxin-amido synthetases (Zheng *et al.*, 2016; Shigeyama *et al.*, 2016). This dual role may explain why GH3 members are overrepresented in both regulatory categories. Along with hydrolases and inhibitors, peroxidases and thioredoxins are overrepresented in several regulatory categories. These modulators of ROS levels facilitate both immune signalling and cell wall remodelling by extracellular ROS (Ivanchenko *et al.*, 2013).

Besides plant proteins, we also identified peptides from bacterial proteins in the extracellular proteome of agroinfiltrated leaves. In fact, Agrobacterium proteins make up a quarter of the extracellular proteins in agroinfiltrated samples (738 bacterial vs 2233 plant proteins) and appear to mostly function in providing nutrients to the bacteria. Highly abundant bacterial proteins are ABC transporters, cytochrome P450 proteins and porins. This may include cytoplasmic bacterial proteins released into the extracellular proteome upon cell death or during the extraction of apoplastic fluid. We identified peptides corresponding to 17 different Agrobacterium proteases, including

six Ser proteases, in the extracellular space (Supporting Table ES2.08), but did not identify peptides from bacterial proteases using ABPP-MS.

Ageing of leaves irrespective of their treatment during our 10-day time course is associated with induction of defence and decrease of primary metabolism. 18.9% of detected transcripts and 6.5% of identified extracellular proteins changed significantly in abundance over time independent of the treatments. Analysis of predicted functions overrepresented among the changing transcripts and proteins suggests that while PLCPs, P450-domain containing proteins and PR proteins accumulate, components of the photosynthetic machinery, histones and cytoskeleton elements decrease in abundance. (R code in File ES2.04, data in Supporting Tables ES2.09 and ES2.10)

2.3.1.4 The repertoire of active extracellular PLCPs and Ser hydrolases is modulated, but not drastically expanded upon agroinfiltration

We next investigated extracellular hydrolase activity to identify active candidate proteases for depletion, focusing on Ser and Cys proteases because plant immune responses often result in increased abundance and activity of these protease classes and Cys proteases can degrade biopharmaceuticals *in vitro* (Paireder *et al.*, 2016, 2017). We performed Activity-based Protein Profiling (ABPP) with probes targeting Papain-like Cys Proteases (PLCPs) and Serine Hydrolases (SHs) (Greenbaum *et al.*, 2002; Kaschani *et al.*, 2009). Both probes consist of a specific inhibitor that covalently binds the active site of their respective targets, a linker and a biotin tag used for enrichment of active enzymes from extracellular proteomes prior to MS analysis. Both probes have been validated in plants using target detection, genetic target depletion and inhibition of probe binding with independent protease inhibitors, confirming that reactivity to the probe indicates availability of the active site and thus enzyme activity (Kovács & van der Hoorn, 2016). We focused on 5 dpi, when the response to

agroinfiltration is fully developed. We identified peptides corresponding to two PLCPs and 29 SHs, 17 of which are proteases, that were enriched from the extracellular proteome using ABPP-MS. (Figure 2.2d). Abundance of peptides corresponding to one Clade II and one Clade III SCPL was increased upon agroinfiltration, indicating increased activity and/or abundance. In contrast, peptides corresponding to an RD21-like PLCP were only identified in ABPP-MS samples from mock infiltrated plants, indicating depletion of enzyme activity upon agroinfiltration. Abundance of extracellular peptides from this PLCP remained constant upon agroinfiltration at 5 dpi, suggesting a post-translational regulatory mechanism. We identified peptides corresponding to six subtilases (one SBT5 and five SBT1 subtilases, including the proteins clustering with tomato P69), seven SCPLs (two Clade IB, three Clade II, four cade III) and one aleurain-like PLCP with similar abundance in both agro- and mock infiltrated samples, indicating that activity of these enzymes remains constant upon agroinfiltration. This is surprising, as in tomato, active extracellular subtilases and PLCPs drastically increase in abundance and diversity during immune responses (Esse *et al.*, 2008; Sueldo *et al.*, 2014). Besides the proteases, we identified peptides corresponding to 14 additional SHs annotated as lipases and esterases. Peptides from three GDSL-lipases (containing a GDSL sequence motif) increased in abundance upon agroinfiltration, while peptides from one GDSL-lipase decreased. Adjustment of extracellular GDSL-lipase activity may contribute to immune signalling, as lipases regulate salicylic acid as well as ethylene signalling in Arabidopsis (Falk *et al.*, 1999; Kim *et al.*, 2013a) and upon powdery mildew infection, lipase-encoding transcripts accumulate in grapevine (Szalontai *et al.*, 2012) (R code in File ES2.05, data in Table ES2.11).

2.3.1.5 The extracellular proteome and active secretome is under post-transcriptional and post-translational control

Having transcriptome, extracellular proteome and active secretome data creates a unique opportunity to detect discrepancies in abundance changes between transcripts, total extracellular proteins and active extracellular proteins. To assess how much post-transcriptional regulation shapes the extracellular proteome, we compared the fold changes of extracellular protein abundance and transcript abundance at 5 dpi, when the response to agroinfiltration is fully developed (Figure 2.3a, Supporting Table ES2.12). The extracellular protein (EP) was increased more or decreased less in abundance than its corresponding transcript (T) for 215 (9.7 %) of the 2226 extracellular proteins for which we detected the corresponding transcript (significant difference between the fold changes, EP>T). Among these proteins are two PR1 proteins, eight PR2 glucanases and two P69-like subtilases (PR7). This finding indicates that the immune response is accompanied by efficient extracellular protein delivery, as previously suggested based on transcriptional upregulation of the secretory pathway during immunity (Wang *et al.*, 2005). More efficient extracellular delivery may be accompanied by enhanced stability of the secreted proteins. In addition to classical PR proteins, two PLCPs (one XCP and one RD19-like) and three pepsin-like aspartic proteases appeared efficiently delivered to the extracellular space with EP>T, suggesting they may be candidate immune proteases. Only 45 extracellular proteins (2.0 %) increased less or decreased more in abundance than expected from their transcript level changes (EP<T).

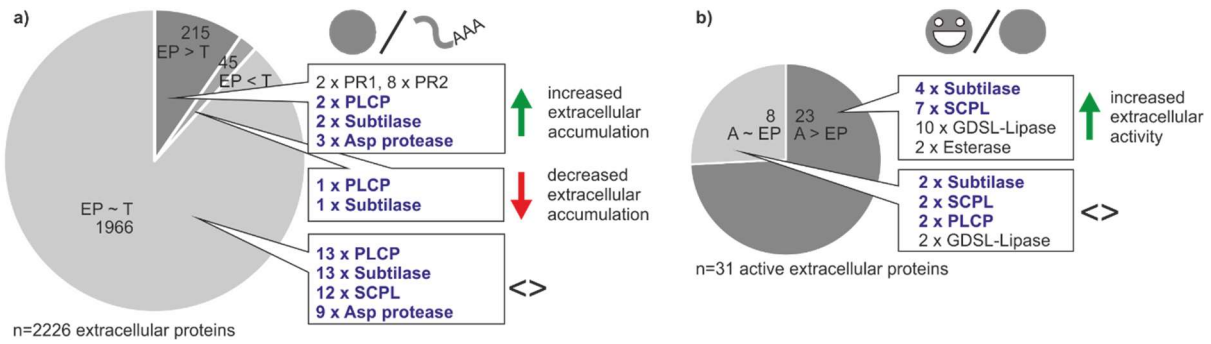


Figure 2.3: Post-transcriptional and post-translational control over the extracellular proteome. Fold changes in response to agroinfiltration at 5 dpi were compared between abundances of extracellular proteins and their transcripts (a) and between activity and abundance of extracellular proteins (b) (Student's t-test; BH adjusted $p < 0.1$). Protein groups of interest are named next to the pie charts. Full datasets are given in Supporting Tables ES2.12 and ES2.13. The R code of the analysis is given in Supporting File ES2.06. T, fold change of transcript abundance; EP, fold change of extracellular protein abundance; AEP, fold change of activity of extracellular protein.

Interestingly, for 23 of the 31 active enzymes for which we also detected the extracellular protein (74.2 %), abundance of the active protein (A) increased more or declined less than expected based on changes in total extracellular protein abundance (A>EP) (Figure 2.3b, Supporting Table ES2.13). This suggests that hydrolase activity is frequently post-translationally controlled. Among the proteins with A>EP are four subtilases, three of which contain I09 domains and are thus likely activated by cleavage upon agroinfiltration. Activation by cleavage may explain why abundance of all active subtilases remained constant while total protein abundance decreased in five cases. Seven SCPLs also remained constant or increased in active protein abundance although their total protein abundance decreased (Supporting Table ES2.13). As SCPLs lack inhibitory domains, they may undergo post-translational activation by release from an inhibitor or autoactivation triggered by pH or redox-level changes. Taken together, efficient extracellular delivery influences the increase of the

extracellular proteome upon agroinfiltration and many hydrolases for which we identified peptides by ABPP-MS appear to be activated post-translationally.

2.3.2 *N. benthamiana* deploys a large, diverse repertoire of proteases in agroinfiltrated leaves

To improve protease annotation, we analysed the *N. benthamiana* protease repertoire in the context of known plant proteases, using the MEROPS nomenclature. The MEROPS database of proteases and inhibitors defines families based on protein sequence homology that are grouped into clans based on structural homology. Protease family names consist of a letter denoting the catalytic class and a unique number (i.e. A01 for pepsin-like aspartic proteases) (Rawlings *et al.*, 2014). We identified 1245 proteases and non-catalytic protease homologs in the curated proteome of *N. benthamiana.* A smaller protease repertoire is encoded by genomes of three crop and model plants: Arabidopsis (796 proteases), tomato (901) and rice (997) (Figure 2.4a and supporting Table ES2.14). Although the *N. benthamiana* protease repertoire is much larger, the proportion of predicted proteins annotated as proteases is higher in the other plants (2.9 % in Arabidopsis, 2.6 % in tomato and 2.4 % in rice) than in *N. benthamiana* (1.6 %). The lower proportion of proteases in *N. benthamiana* may reflect the suboptimal genome annotation.

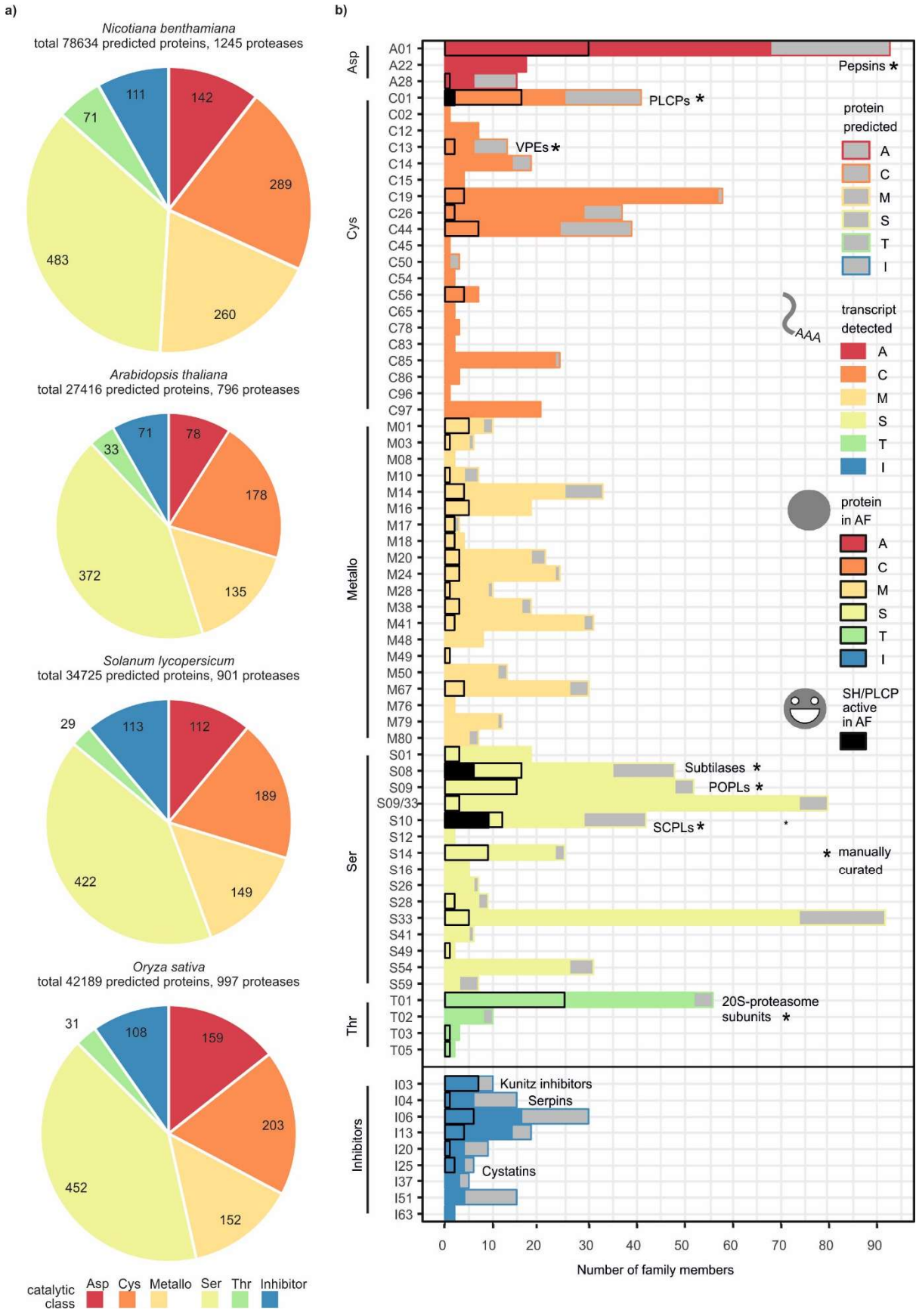

a)
Nicotiana benthamiana
total 78634 predicted proteins, 1245 proteases
142
289
260
483
71
111
Arabidopsis thaliana
total 27416 predicted proteins, 796 proteases
78
178
135
372
33
71
Solanum lycopersicum
total 34725 predicted proteins, 901 proteases
112
189
149
422
29
113
Oryza sativa
total 42189 predicted proteins, 997 proteases
159
203
152
452
31
108
catalytic class
Asp
Cys
Metallo
Ser
Thr
Inhibitor
b)
Asp
Cys
Metallo
Ser
Thr
Inhibitors
A01
A22
A28
C01
C02
C12
C13
C14
C15
C19
C26
C44
C45
C50
C54
C56
C65
C78
C83
C85
C86
C96
C97
M01
M03
M08
M10
M14
M16
M17
M18
M20
M24
M28
M38
M41
M48
M49
M50
M67
M76
M79
M80
S01
S08
S09
S09/33
S10
S12
S14
S16
S26
S28
S33
S41
S49
S54
S59
T01
T02
T03
T05
I03
I04
I06
I13
I20
I25
I37
I51
I63
Pepsins *
PLCPs *
VPEs *
Subtilases *
POPLs *
SCPLs *
20S-proteasome subunits *
Kunitz inhibitors
Serpins
Cystatins
protein predicted
A
C
M
S
T
I
AAA
transcript detected
protein in AF
SH/PLCP active in AF
* manually curated
0
10
20
30
40
50
60
70
80
90
Number of family members

Figure 2.4: *N. benthamiana* has a diverse protease and protease inhibitor repertoire. a) Number of proteases and non-catalytic protease homologs in each catalytic class or inhibitors annotated are given for each species (*N. benthamiana* curated proteome, A. thaliana TAIR10, O. sativa v7 JGI, S. lycopersicum ITAG2.4). The area of each pie chart is scaled by the total number of proteases and inhibitors. b) The protease and protease inhibitor repertoire of *N. benthamiana*. For each MEROPS family, bars give the size in the predicted proteome (grey), the number of transcripts detected in mock and/or agroinfiltrated leaves (filled), the number of proteins for which we detect corresponding extracellular peptides in agro- and/or mock infiltrated leaves (black outline) and the number of enzymes for which peptides were detected in ABPP-MS, indicating activity (black fill). For the manually curated families (marked by an asterisk), protease homologs lacking the active site were not counted. Each protein group identified in MS and ABPP-MS was counted as one family member. Note that due to the nature of the ABPP probes used, only SHs and PLCPs were monitored on the activity level. The S09 (prolyl oligopeptidase) and S33 (prolyl aminopeptidase) families share the α/β-hydrolase fold (PFAM families PF12695 and PF12697) and sequences with only these PFAM identifiers are marked S09/S33.

To characterize functional proteases and protease inhibitors in agroinfiltrated leaves, we analysed them at three levels. First, we detected transcripts for 975 proteases and 60 inhibitors. Second, we identified extracellular peptides from 196 proteases and 21 inhibitors, including proteases from every catalytic class. Third, we identified 17 active extracellular Ser and Cys proteases in agroinfiltrated leaves (Figure 2.4b). The most prominent features of the *N. benthamiana* protease repertoire are the large numbers of Cys, Metallo- and Thr proteases. Among the Cys proteases, the metacaspase family C14 is doubled in size (n = 18 members) compared to Arabidopsis (n = 9), tomato (n = 9) and rice (n = 8). We did not identify extracellular peptides corresponding to metacaspases, though 14 had detectable transcripts. The large number of metalloproteases in *N. benthamiana* (n = 260) compared to Arabidopsis (n = 135), tomato (n = 149) and rice (n = 152) is distributed among 20 families and we identified extracellular peptides corresponding to members of most metalloprotease families. 10% (n = 26) of the metalloprotease-encoding genes increased in transcript abundance upon agroinfiltration, while 13% (n = 35) decreased. In contrast, 75 % (n

= 28) of the metalloproteases for which we identified extracellular peptides increased in abundance and only one M28 protease decreased. Very few plant metalloproteases are functionally characterized, including AtSOL1, an M14 carboxypeptidase processing peptide hormones (Casamitjana-Martínez *et al.*, 2003; Tamaki *et al.*, 2013) and AtPreP1 and 2, Arabidopsis M16 proteases cleaving organellar target peptides (Bhushan *et al.*, 2005). The *N. benthamiana* M10 protease NMMP1 has been implicated in defence because silencing *NMMP1* confers susceptibility to bacterial pathogens (Kang *et al.*, 2010). The transcript corresponding to NMMP1 (Niben101Scf10336XLOC_078719) increases in abundance upon agroinfiltration, while its extracellular peptides appear constant. The large metalloprotease repertoire of *N. benthamiana* is changing upon agroinfiltration, raising the question whether these metalloproteases might regulate the immune response through protein processing. The high number of Thr proteases (n = 71) compared to Arabidopsis (n = 34), tomato (n = 29) rice (n = 31) is due to drastic expansion of the T01 family in *N. benthamiana* (n = 65, vs n = 24 in Arabidopsis, n = 20 in tomato and n = 23 in rice). T01 contains the α and β subunits of the 20S core protease of the proteasome. Phylogenetic analysis showed that *N. benthamiana* has more representatives of each subunit (Supporting Figure 2.S03). We identified extracellular peptides corresponding to 25 T01 subunits, possibly due to cell content leakage. Since transcripts of most (n = 52) T01 subunits were detected in leaves, multiple versions of the 20S proteasome may co-exist, as they do in Arabidopsis (Book *et al.*, 2010). Indeed, we recently showed that two sets of catalytic subunits are incorporated in functional 20S proteasomes in *N. benthamiana* (Misas-Villamil *et al.*, 2017).

In contrast to the protease repertoire, the protease inhibitor repertoire of *N. benthamiana* (n = 111 predicted protease inhibitors) is not much larger compared to

tomato (n = 113), rice (n = 108) and Arabidopsis (n = 71). This apparent discrepancy may reflect the multifunctionality of many protease inhibitors (Grosse-Holz & van der Hoorn, 2016) and incomplete annotation. Among the annotated protease inhibitors, the I03 (Kunitz) inhibitors probably function extracellularly, as we identified extracellular peptides corresponding to all seven Kunitz (family I03) inhibitors for which we detected transcripts. Kunitz inhibitors can inhibit both subtilases (family S08) and α-amylases, but their bifunctional structure can also target other Ser or Cys proteases, and other proteins (Renko *et al.*, 2012). *N. benthamiana* serpins (I04) appear to mostly be intracellular, as we detected six serpin-encoding transcripts, but identified corresponding extracellular peptides for only one. Serpins have been found in both the cytoplasm (Lampl *et al.*, 2013) and the extracellular space (Ghorbani *et al.*, 2016), and regulate plant defence and programmed cell death through irreversible inhibition of Ser and Cys proteases (Lampl *et al.*, 2013; Bhattacharjee *et al.*, 2017). We detected transcripts for four and identified extracellular peptides corresponding to two cystatins (family I25). Cystatins target PLCPs and VPEs, regulating storage protein accumulation, germination and defence (Benchabane *et al.*, 2008; Grosse-Holz & van der Hoorn, 2016).

Having obtained an overview of the *N. benthamiana* protease and protease inhibitor repertoire, we focused on six large protease families, which we curated manually (Supporting File 2. S01). For these six families, we performed phylogenetic analyses to resolve subfamilies and determine which *N. benthamiana* proteins are most similar to previously studied proteases (Figures 2.5 and 2.6).

2.3.2.1 The PLCP family is conserved in N. benthamiana, but PIP1- and RCR3-like PLCPs are absent from the extracellular proteome of agroinfiltrated leaves

N. benthamiana has more Papain-like Cys Proteases (PLCPs, family C01, n = 41 members) than Arabidopsis (n = 36) and tomato (n = 36), but less than rice (n = 54). PLCP subfamilies can be defined by shared sequence features (Richau *et al.*, 2012) (Figure 2.5a). For example, the NPIR vacuolar localisation signal is found in Aleurain-Like Proteases (ALPs) and the KDEL ER-retention signal in Cys Endopeptidases (CEPs). Cathepsin-B-like proteases (CTBs) have a specific prodomain (PF08127) serving as chaperone and inhibitor, like the I29 (PF08246) prodomain for other PLCPs. Most *N. benthamiana* PLCPs have a secretion signal predicted by SignalP (Dyrløv Bendtsen *et al.*, 2004). Accordingly, we identified extracellular peptides corresponding to 18 of the 25 PLCPs for which we detected transcripts. Among the extracellular PLCPs are three granulin-carrying proteases similar to the immune protease AtRD21 (Shindo *et al.*, 2012), three Cathepsin-B like proteases (CTBs) and NbCYP1 and NbCYP2, which limit susceptibility to fungal pathogens (Hao *et al.*, 2006). We identified peptides in ABPP-MS from NbRD21 and NbCYP1, indicating that these proteases are active extracellularly. Many of the PLCPs for which we detected transcripts and identified extracellular peptides contribute to *N. benthamiana* immunity. For instance, silencing *NbCathB* (Gilroy *et al.*, 2007; McLellan *et al.*, 2009) blocks the hypersensitive response (HR) and *NbC14/CP14* silencing confers susceptibility to *Phytophthora infestans* (Kaschani *et al.*, 2010; Bozkurt *et al.*, 2011). Surprisingly, we did not identify extracellular peptides corresponding to the *N. benthamiana* proteins clustering with the tomato immune proteases PIP1 (Tian *et al.*, 2004) and RCR3 (Krüger *et al.*, 2002), although we detected NbPIP1- and NbRCR3-encoding transcripts.

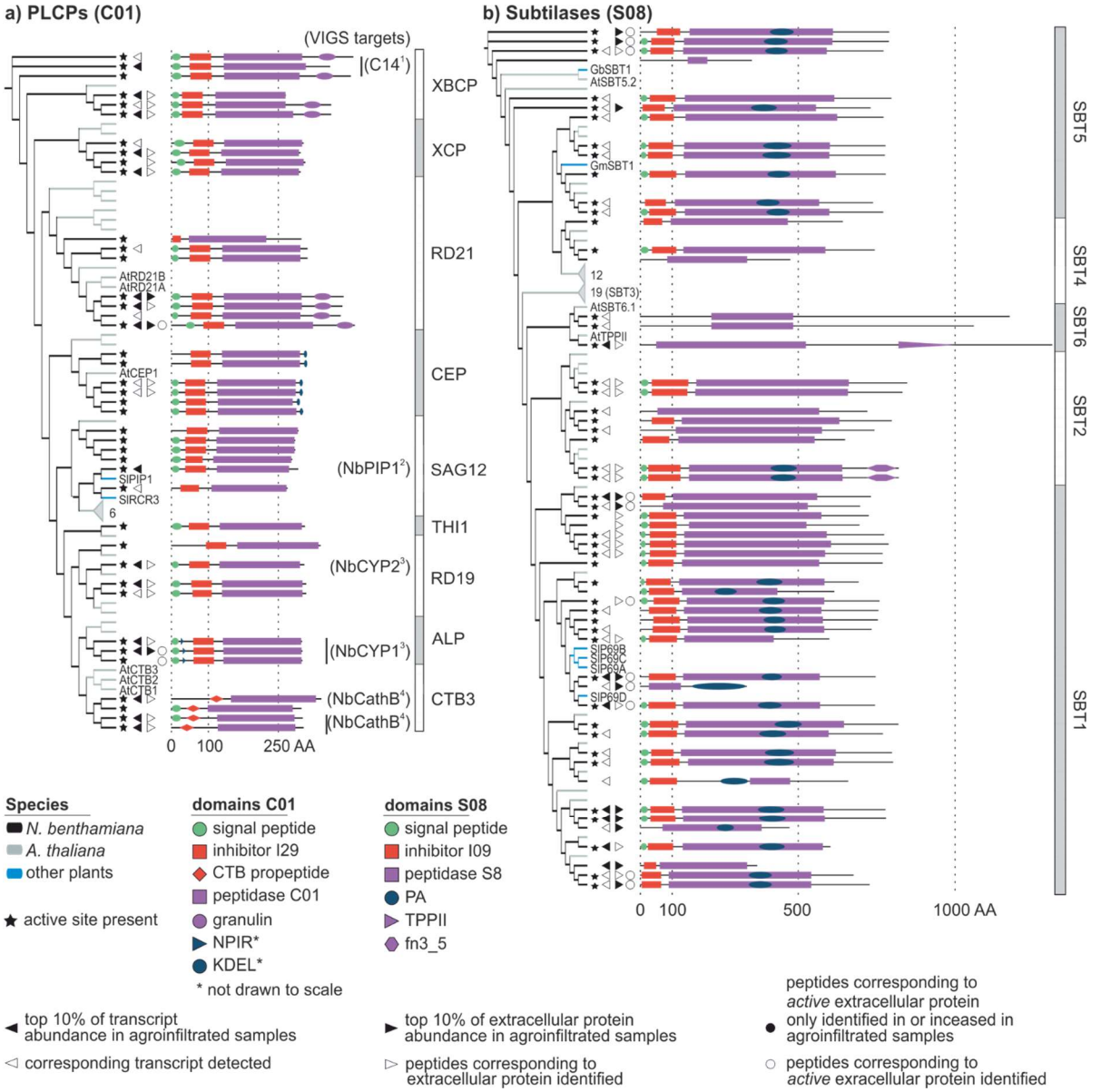


Figure 2.5: Annotation and detection of extracellular Papain-like Cys Proteases (PLCPs) and Subtilases in *N. benthamiana*. Phylogenetic trees based on the protein sequences of PLCPs (a) and subtilases (b) containing all proteases and protease homologs in the respective family in Arabidopsis (grey branches) and *N. benthamiana* (black branches), supplemented by well-studied enzymes from other plant species (blue branches). Names are given as two-letter species abbreviation followed by the name used in the literature. Grey triangles denote collapsed subtrees that contain only Arabidopsis sequences, with the number of proteins given next to the triangle. For protein abundance and activity, the respective symbols are shown next to all members of each protein group for which corresponding peptides were identified. VIGS targets were predicted based on > 90 % identical residues between the fragment used for VIGS and the respective transcript. References: 1 (Kaschani et al., 2010); 2 (Xu et al., 2012); 3 (Hao et al., 2006); 4 (Gilroy et al., 2007). Abbreviations: CTB, cathepsin B-like; TPP, tripeptidyl-peptidase; fn3_5, fibronectin-3 like domain found on streptococcal C5a peptidase.

2.3.2.2 P69-like SBT1 subtilases are abundant and active in the extracellular proteome of agroinfiltrated N. benthamiana leaves

The *N. benthamiana* subtilase family (S08, n = 56 members) is the same size as in Arabidopsis (n = 56) and smaller than in tomato (n = 90) and rice (n = 61) (Figure 2.5b). We identified extracellular peptides for 28 of the 39 subtilases whose transcript we detected. 12 of the 28 subtilases for which we identified extracellular peptides were among the top 10 % most abundant extracellular proteins and we identified peptides corresponding to 11 active subtilases using ABPP-MS (Figure 2.5b). Across the whole subtilase family, the I09 prodomain is well conserved and the PA dimerization domain (Rose *et al.*, 2010) is present in some members of each subfamily. An exception lacking SP, I09 and PA domains are the basal SBT6 subtilases (Taylor & Qiu, 2017). We detected transcripts encoding three *N. benthamiana* SBT6 subtilases and identified extracellular peptides from one. SBT6 subtilases can process peptide hormones regulating cell elongation (Ghorbani *et al.*, 2016), or degrade peptides released by the 26S proteasome (Book *et al.*, 2005). Remarkably, the SBT1 subfamily is three-fold larger in *N. benthamiana* (n = 30 members) than in Arabidopsis (n = 9), while the SBT3 subfamily is absent in *N. benthamiana*. We detected 22 SBT1 subtilase-encoding transcripts and identified corresponding extracellular peptides for 19. We also identified peptides corresponding to eight active SBT1 subtilases by ABPP-MS. The tomato P69A, B and C subtilases (Jordá *et al.*, 1999) cluster with the SBT1 subfamily, which is consistent with a recently published, updated phylogeny of the subtilase family (Taylor & Qiu, 2017). We detected transcripts for eight and identified extracellular peptides corresponding to four SBT5 subtilases. We also identified peptides corresponding to three active SBT5 subtilases by ABPP-MS. SBT5 subtilases can regulate plant immunity as receptors (Duan *et al.*, 2016), as

transcription factor binding proteins (Serrano *et al.*, 2016) or being processed to release peptide hormones (Pearce *et al.*, 2010). The updated subtilase phylogeny (Taylor & Qiu, 2017) suggests that SBT6 should be split into two subfamilies and notes that the distinction between SBT4 and SBT5 subfamilies is weakly supported. Indeed, we find two subclades of SBT6 in *N. benthamiana*, clustering with one of the Arabidopsis representatives each. SBT4 falls into a Clade containing part of SBT5 in our tree, indicating the updated phylogeny agrees with our curated *N. benthamiana* proteome.

2.3.2.3 N. benthamiana POPLs are underrepresented in the extracellular proteome of agroinfiltrated leaves

The Prolyl-oligopeptidase-like (S09, POPL) family in *N. benthamiana* (n = 180 members) is equivalent in size with the POPL families in Arabidopsis (n = 143), tomato (n = 162) and rice (n = 196) (Figure 2.6a). The S09 family is defined via the α/β-hydrolase fold (PF12695 and PF12697), but these proteins are not always proteases (Mindrebo *et al.*, 2016). Interestingly, we only identified extracellular peptides for 18 of the 122 family members for which we detect a transcript (15%), suggesting that *N. benthamiana* POPLs are primarily intracellular. S09A (prolyl-oligopeptidase-like, POPL) is the smallest subfamily and no plant POPL has been functionally characterized, but we detected 14 POPL-encoding transcripts and identified extracellular peptides from two POPLs. The S09B Dipeptidyl-Peptidase type IV (DPP-IV) subfamily contains membrane-bound exopeptidases (Tripathi & Sowdhamini, 2006), but also clusters with non-proteolytic α/β-hydrolases, including HbHNL, RsPNAE and LhMKS1 and the salicylic acid receptor NtSABP2 (Wagner *et al.*, 1996; Dogru *et al.*, 2000; Forouhar *et al.*, 2005; Auldridge *et al.*, 2012). Interestingly, we only detected a transcript for NbSABP2. We identified extracellular peptides from four DPP-

IVs. We also detected transcripts for four and identified peptides for two aminoacyl-removing peptidases (AAPs/AAREs, subfamily S09C) clustering with AtAARE. AtAARE is implicated in the cytoplasmic antioxidative system (Yamauchi *et al.*, 2003; Nakai *et al.*, 2012). We detected transcripts, but did not identify extracellular peptides for the non-proteolytic members of S09C, including the proteins clustering with the hormone receptors D14 (Yao *et al.*, 2016), KAI2 (Guo *et al.*, 2013), GID1 (Griffiths *et al.*, 2006) and the potato epoxide hydrolase StEH (Stapleton *et al.*, 1994).

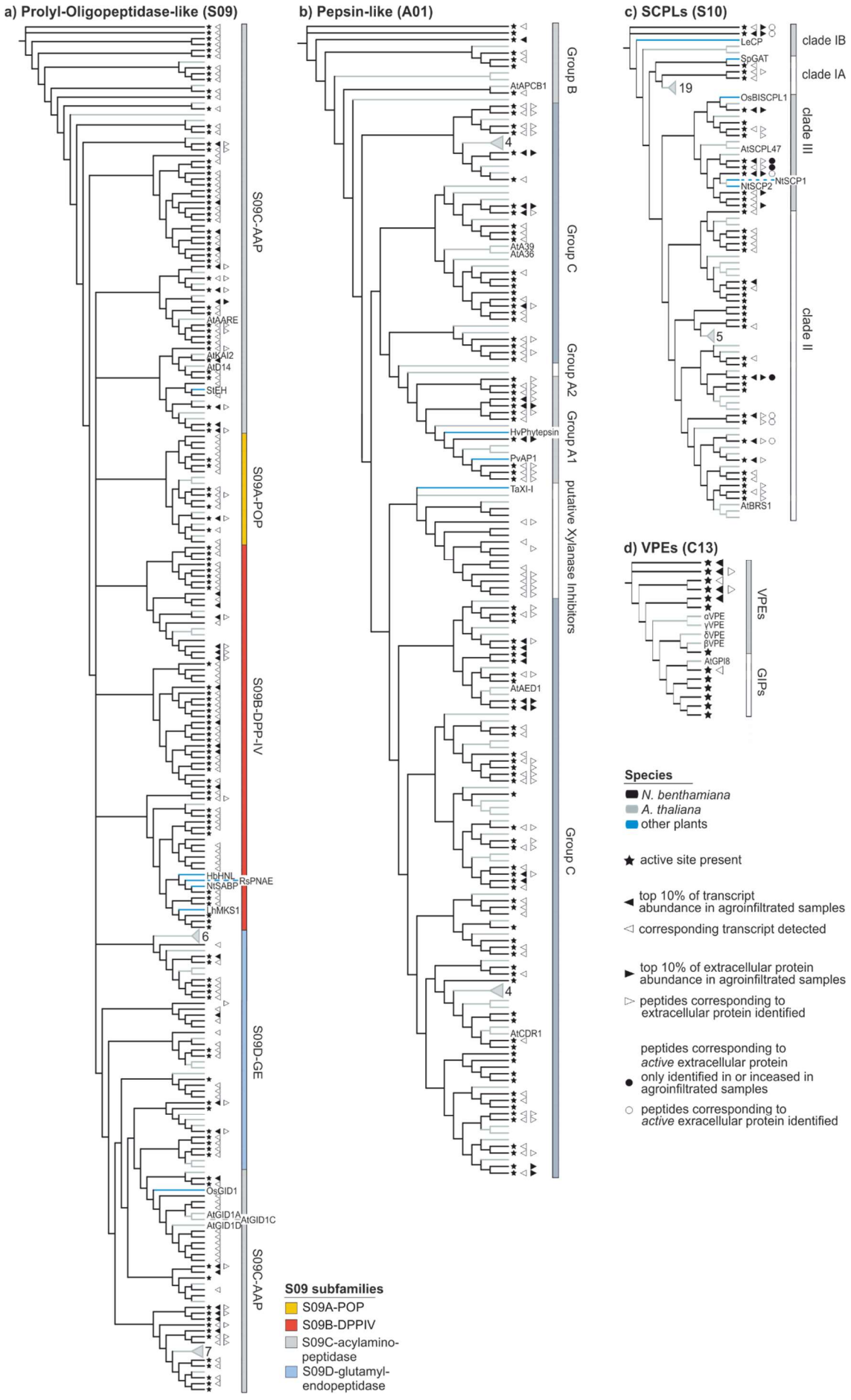

a) Prolyl-Oligopeptidase-like (S09)
S09C-AAP
S09A-POP
S09B-DPP-IV
S09D-GE
S09C-AAP
AtAARE
AtKAI2
AtD14
StEH
HbHNL
NtSABP
RsPNAE
LhMKS1
6
OsGID1
AtGID1A
AtGID1B
AtGID1C
7
b) Pepsin-like (A01)
Group B
Group C
Group A2
Group A1
putative Xylanase Inhibitors
Group C
AtAPCB1
4
AtA39
AtA36
HvPhytepsin
PvAP1
TaXI-I
AtAED1
4
AtCDR1
c) SCPLs (S10)
clade IB
clade IA
clade III
clade II
LeCP
SpGAT
19
OsBISCPL1
AtSCPL47
NtSCP2
NtSCP1
5
AtBRS1
d) VPEs (C13)
VPEs
GIPs
cVPE
vVPE
δVPE
βVPE
AtGPI8
S09 subfamilies
S09A-POP
S09B-DPPIV
S09C-acylamino-peptidase
S09D-glutamyl-endopeptidase
Species
N. benthamiana
A. thaliana
other plants
active site present
top 10% of transcript abundance in agroinfiltrated samples
corresponding transcript detected
top 10% of extracellular protein abundance in agroinfiltrated samples
peptides corresponding to extracellular protein identified
peptides corresponding to active extracellular protein only identified in or inceased in agroinfiltrated samples
peptides corresponding to active exracellular protein identified

Figure 2.6: Annotation and detection of additional protease families in *N. benthamiana*. Phylogenetic trees based on the protein sequences of POPLs (a), Pepsin-like proteases (b), SCPLs (c) and VPEs (d) containing all proteases and protease homologs of the respective family in Arabidopsis (grey branches) and *N. benthamiana* (black branches), supplemented by well-studied enzymes from other plant species (blue branches). Arabidopsis sequences that only carry the α/β-hydrolase fold PFAM identifiers are not shown in the S09 tree for readability. Names and symbols are used as described for Figure 2.5.

2.3.2.4 Pepsin-like aspartic proteases are highly abundant in agroinfiltrated leaves and pepsin-like xylanase inhibitors have expanded in N. benthamiana

Extracellular peptides from pepsin-like aspartic proteases (A01) were abundantly detected and the A01 family is expanded in *N. benthamiana* (n = 110 members) compared to Arabidopsis (n = 69) and tomato (n = 100), but is smaller than the rice A01 family (n = 130) (Figure 2.6b). We detected 76 A01 protease-encoding transcripts and identified extracellular peptides from 45 pepsin-like aspartic proteases. Eight pepsin-like aspartic proteases were among the top 10 % most abundant extracellular proteins. Pepsin-like aspartic proteases are subdivided into subfamilies A1 (typical pepsin-like), A2 (typical, but lacking the plant-specific insert), B (nucelllins) and C (atypical) (Faro & Gal, 2005). We detected transcripts and identified peptides for 10 A1 pepsin-like proteases and they cluster with two enzymes implicated in stress responses, barley phythepsin and bean AP1 (Hückelhoven *et al.*, 2001; Contour-Ansel *et al.*, 2010). Group A2 seems absent in the *N. benthamiana* predicted proteome. We detected transcripts, but did not identify extracellular peptides, for five A01 group B proteases. AtAPCB1 in group B is required for autophagy and resistance to *Botrytis* (Li *et al.*, 2016). Group C is the largest A01 subfamily in *N. benthamiana*, with 54 A01 Group C-encoding transcripts detected in leaves and corresponding extracellular peptides identified for 29. Two Arabidopsis members of group C, AtCDR1 and AtAED1, modulate plant defence responses in the extracellular space (Xia *et al.*,

2004; Breitenbach *et al.*, 2014). AtAED1 clusters with two abundant extracellular *N. benthamiana* proteins. We detected transcripts encoding several and identified extracellular peptides from one protein clustering with AtA36 and AtA39, two putative GPI-anchored aspartic proteases (Gao *et al.*, 2017). Interestingly, the A01 Clade clustering with the wheat xylanase inhibitor TaXI-I is expanded drastically with 14 members in *N. benthamiana*, compared to two in Arabidopsis (Sansen *et al.*, 2004; Brutus *et al.*, 2005). TaXIs share the fold of pepsin-like proteases, but have lost the active site and act as xylanase inhibitors. We detected transcripts for seven putative *N. benthamiana* xylanase inhibitors and identified extracellular peptides corresponding to six.

2.3.2.5 N. benthamiana has an expanded SCPL Clade III in the extracellular proteome of agroinfiltrated leaves

N. benthamiana has fewer Serine-Carboxypeptidase-like enzymes (SCPLs, S10, n = 42 members) than Arabidopsis (n = 54), tomato (n = 61) and rice (n = 59) (Figure 2.6c). We detected transcripts for 29 and identified extracellular peptides from 19 SCPLs. We also identified peptides corresponding to nine active extracellular SCPLs by ABPP-MS. SCPLs fall into four Clades (Fraser *et al.*, 2005). We detected transcripts for three and identified extracellular peptides corresponding to one member of Clade IA, which contains the *Solanum pennellii* glucose acetyltransferase (SpGAT) (Franziska, 2013). Clade IB contains the wound-inducible tomato carboxypeptidase LeCP (Moura *et al.*, 2001) and two *N. benthamiana* proteins, for which we detected transcripts, identified extracellular peptides and peptides by ABPP-MS, indicating activity. The largest S10 subfamily is Clade II, with transcripts detected for 16 members and extracellular peptides identified for eight. We also identified extracellular peptides corresponding to four active Clade II SCPLs in ABPP-MS. Interestingly,

Clade III is expanded in *N. benthamiana* (n=10) compared to Arabidopsis (n=5) and well represented in the extracellular proteome, with the encoding transcripts detected and extracellular peptides identified for eight members each. We also identified peptides corresponding to three active Clade III SCPLs in ABPP-MS. Clade III members such as NtSCP1, NtSCP2 (Bienert *et al.*, 2012) and AtSCPL47 (Charmont *et al.*, 2005) are extracellular carboxypeptidases and OsBISCPL1, a rice Clade III SCPL, enhances stress resistance when overexpressed in Arabidopsis (Liu *et al.*, 2008).

2.3.2.6 Peptides corresponding to extracellular VPEs are identified in agroinfiltrated N. benthamiana leaves

We annotated seven Vacuolar Processing Enzymes (VPEs/legumains/aspariginyl endopeptidases) and six GPI-anchor transamidases that share the VPE domain architecture (PF01650) in *N. benthamiana*. Together, they constitute family C13 (n = 13 members), which is larger than in Arabidopsis (n = 5) and rice (n = 6), but smaller than in tomato (n = 19). We detected five VPE-encoding transcripts and one NbGIP-encoding transcript. Notably, we also identified extracellular peptides corresponding to two VPEs, consistent with observations made in tomato (Sueldo *et al.*, 2014). VPEs can activate proteins in vacuoles, including proteases (Rojo *et al.*, 2003) and protease inhibitors (Heath *et al.*, 1995; Mylne *et al.*, 2011). Silencing of *NbVPEs* blocks virus-induced cell death in *N. benthamiana* (Hatsugai *et al.*, 2004) and VPEs can also act in other forms of plant cell death (Gepstein *et al.*, 2003; Nakaune *et al.*, 2005; Sueldo *et al.*, 2014; Hatsugai *et al.*, 2015).

2.4 Conclusions

Upon agroinfiltration, 25 % of the full leaf mRNA transcriptome changes in abundance, associated with an immune response mounted at the expense of photosynthesis. 70 % of all extracellular proteins increase in abundance and their predicted functions confirm that an extracellular immune response occurs. Increasing the extracellular proteome while photosynthesis is shut down appears to drive leaves into a nutrient-deprived state. Engineering *N. benthamiana* to react less strongly to Agrobacterium, or Agrobacterium to be less immunogenic in *N. benthamiana*, may enhance RP expression by re-directing limiting resources. Interestingly, expression of the silencing inhibitor P19 had minor effects on the transcriptome and no effect on the extracellular proteome.

Discrepancies between changes in transcript, extracellular protein and active extracellular protein abundances suggest that the extracellular proteome is influenced post-transcriptionally and that many extracellular enzymes are activated post-translationally. The *N. benthamiana* immune response to agroinfiltration differs from immune responses to bacterial and fungal pathogens in Arabidopsis and tomato in that there is no drastic increase in numbers or amounts of active extracellular subtilases and PLCPs (Xia *et al.*, 2004; Gilroy *et al.*, 2007; Esse *et al.*, 2008; Sueldo *et al.*, 2014). This is surprising, as *N. benthamiana* has an exceptionally large repertoire of 1245 proteases and non-catalytic protease homologs, transcripts corresponding to 975 proteases were detected in leaves and peptides corresponding to 196 proteases were identified in the extracellular space. Prominent features of the extracellular protease repertoire of agroinfiltrated leaves are an expanded Clade of SCPLs, highly abundant pepsin-like proteases and many SBT1 subtilases. Targeted depletion or inhibition of these enzymes may limit undesired proteolysis to improve

agroinfiltrated *N. benthamiana* as a protein expression platform. We have selected several proteases for genetic depletion by genome editing to investigate their role in RP degradation and how they shape the endogenous proteome.

2.5 Material and Methods

All chemicals were obtained from Sigma (Sigma-Aldrich, St. Louis, US) unless specified otherwise.

Agroinfiltration procedure: *N. benthamiana* plants were grown at 21 °C under a 16/8 h light/dark regime in a growth room. *Agrobacterium* GV3101-pMP90 (WT) and *Agrobacterium* GV3101-pMP90 carrying a P19-encoding plasmid (P19) were grown for 21h at 28 °C with agitation in LB medium (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract) containing 100 µM Rifampicin and 100 µM Gentamycin (for WT) plus 100 µM Kanamycin (for P19). The P19 plasmid was pJK050, which was a gift from Jiorgos Kourelis (Addgene plasmid # 101751), and contains a sequence encoding P19 from tomato bushy stunt virus. Bacteria were collected by centrifugation at 2000 g for 5 min at room temperature (RT), resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfone (MES), 10 mM $MgCl_2$, pH 5.7, 100 µM acetosyringone) to OD_{600} = 0.5 and left for 2 h at 28 °C with agitation to recover. The first and second fully expanded leaves of pre-flowering stage *N. benthamiana* (4-5 weeks old) were infiltrated with the bacteria suspension using a syringe without a needle.

mRNA extraction and sequencing: For each sample, two leaf discs per leaf from six leaves (three different plants) were pulverized under liquid nitrogen using a mortar and pestle. RNA was extracted from 50 mg of leaf powder using TRIZOL (Thermo Fisher Inc, Waltham, US) according to the manufacturer's instructions. DNA contamination was removed by in solution-digest with the Qiagen RNAse-free DNAse kit, followed by cleanup with the Qiagen RNeasy kit, following the manufacturer's instructions (Qiagen, Hilden, DE). RNA quality was assessed using a Bioanalyzer with the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, US) and following the manufacturer's instructions. All samples used for sequencing had a RIN (RNA integrity number, 28S

to 18S rRNA ratio) > 6.5. RNAseq library preparation and sequencing were performed by the Wellcome Trust Centre for Human Genetics, Oxford. mRNA was enriched using oligo-dT beads and sequenced over three lanes of an Illumina HiSeq device, generating on average 184 million 100 bp paired-end reads per lane.

Bioinformatics tools used for transcriptome analysis: To obtain the genome-based transcriptome (DB4), RNAseq reads were filtered to only retain those with a Phred Q Score > 30 (Ewing & Green, 1998) and aligned to the Niben101 genome (Bombarely *et al.*, 2012) using TopHat version 2.0.14 (Kim *et al.*, 2013b) with default settings. The transcriptome was assembled using StringTie (Pertea *et al.*, 2015) on these alignments, allowing for multi-mapping of reads to several transcripts. TopHat and StringTie were run via the galaxy server (Afgan *et al.*, 2016). This resulted in the genome-based transcriptome (DB4). To obtain the *de novo* assembled transcriptome (DB3), raw reads were quality-trimmed using TRIMMOMATIC-0.32 (Bolger *et al.*, 2014), BAYESHAMMER (SPADES-3.5.0) (Nikolenko *et al.*, 2013) and ALLPATHS-LG-4832 (Butler *et al.*, 2008). Ribosomal RNA was removed using SORTMERNA-1.9 (Kopylova *et al.*, 2012). The quality-trimmed reads were then normalized with a khmer size of 21 in KHMER-0.7.1 (Crusoe *et al.*, 2015). Normalized reads were then assembled and scaffolded using SGA (Simpson & Durbin, 2012), SSPACE-v.3 (Boetzer *et al.*, 2011) and CAP3 (Huang & Madan, 1999). Assembled scaffolds then underwent a final correction step using PILON-1.6 (Walker *et al.*, 2014). This resulted in the *de novo* assembled transcriptome (DB3). We manually curated protease sequences in DB4, using single transcripts from DB1-3 and 5, as described in Supporting File 2.S01. The curated transcriptome was used by the software Salmon version 0.7 (Patro *et al.*, 2015) together with the filtered reads and transcript quantification was performed in lightweight alignment mode. Thus, multi-mapping of

reads was allowed during assembly of the transcriptome in DB4, but not during quantification. The NumReads output of Salmon was used for relative expression analysis in DESeq2 (Love *et al.*, 2014, p. 2).

Bioinformatics tools for Proteome prediction All four transcriptome databases (DB1-4, see Supporting File 2.S01) were subject to coding sequence prediction using GeneMark-ST (Tang *et al.*, 2015), TransDecoder (http://transdecoder.github.io) and Prodigal (Hyatt *et al.*, 2010) using default settings for eukaryotic gene sequences. In cases where all three methods predicted an open reading frame for a transcript then the priority was given to the prediction made by GeneMark-ST unless the GeneMark-ST gene model was a substring of a longer TransDecoder gene model. Transcripts without predictions by any method were subject to an additional round of gene prediction using Prodigal settings for bacterial genes and gene predictions were compiled to create the final predicted proteome.

Apoplastic fluid (AF) extraction: Six *N. benthamiana* leaves per sample (previously infiltrated with Buffer or *A. tumefaciens* suspension) were detached and vacuum-infiltrated with ice-cold water, dried on the surface and placed in a syringe without needle and plunger that was inserted in a 50 mL falcon tube. AF was collected by centrifµgation at 2000 g, 4 °C for 25 min and stored at -80 °C until further use. Protein concentrations were determined with a Bradford assay according to (Ernst & Zor, 2010). To prove that leakage of cytosolic proteins into the extracellular proteome at later time points upon agroinfiltration is indeed caused by disease and not by our AF extraction method, we measured the activity of the intracellular enzyme malate dehydrogenase (MDH). MDH activity in our AF from mock infiltrated leaves falls within the range reported for AF that is virtually free from cytosolic contamination (Supporting Figure 2.S04) (Husted & Schjoerring, 1995; Goulet *et al.*, 2010b).

Mass spectrometry and ABPP-MS: see supplemental methods, additional File ES2.07.

Bioinformatics tools for extracellular proteome analysis: Peptide spectra were annotated using Andromeda (Cox *et al.*, 2011). Included modifications were carbamidomethylation (static) and oxidation, N-terminal acetylation and carbamylation of Lysines and N-termini (dynamic). Protein quantification was performed using MaxQuant version 1.5.5.30 (Tyanova *et al.*, 2016a), including all modifications.

Phylogenetic analyses: Sequences were aligned in Geneious (Kearse *et al.*, 2012) using a plugin for MAFFT v7.017 (Katoh & Standley, 2013). Neighbor-joining trees were constructed using the geneious tree builder with Jukes-Cantor genetic distances and bootstrapped using 1000 times resampling. Trees were edited using iTOL (Letunic & Bork, 2016). Complete versions of the trees including all sequence names are given in Supplemental File ES2.08.

Databases and protease annotation: Protease and inhibitor sequences and PFAM annotations were retrieved for Arabidopsis from TAIR10 (Berardini *et al.*, 2015) and for rice and tomato from Phytozome (Goodstein *et al.*, 2012). Protease sequences from other species to extend the family trees were retrieved from GenBank (NCBI Resource Coordinators, 2017) or UniProt (The UniProt Consortium, 2017). All Arabidopsis, rice, tomato and *N. benthamiana* proteases were annotated by mapping PFAM domains to MEROPS family annotations according to supporting Table ES2.15.

Data availability: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al.*, 2016) partner repository (https://www.ebi.ac.uk/pride/archive/) with the dataset identifier

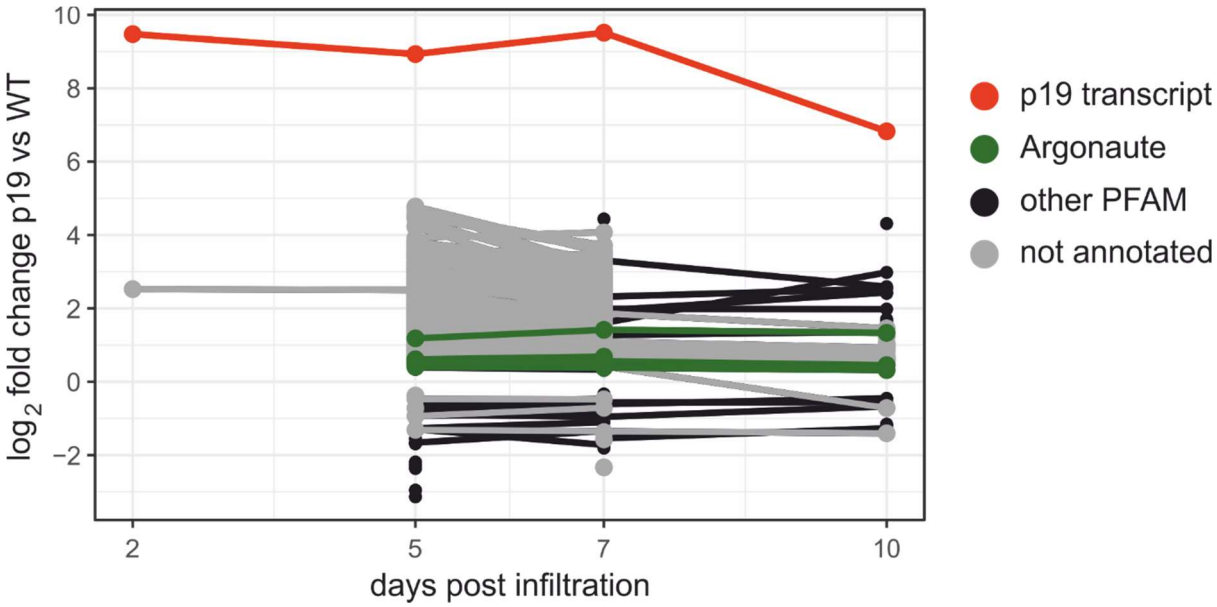
PXD006708. RNAseq data has been deposited in the NCBI Sequence Read Archive repository under identifier SRP109347.

2.6 Acknowledgments

We thank Marcel Bach, Daniela Sueldo and Philippe V. Jutras for constructive suggestions and critical reading, Jiorgos Kourelis and Jan M. Brauner for scientific discussion and Urszula Pyzio for excellent technical support. Jiorgos Kourelis also provided the P19 plasmid. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the Sequencing data. This work was financially supported by the ERC Project 'GreenProteases', University of Oxford (R.H., grant No. 616449), by Somerville College, Oxford (F.G.H. and R.H.), an ERC starting grant (M.K., grant No. 258413) and the Deutsche Forschungsgemeinschaft (M.K., grant no. INST 20876/127-1 FUGG). S.K. is a Royal Society University Research Fellow and work in S.K.'s lab received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 637765.

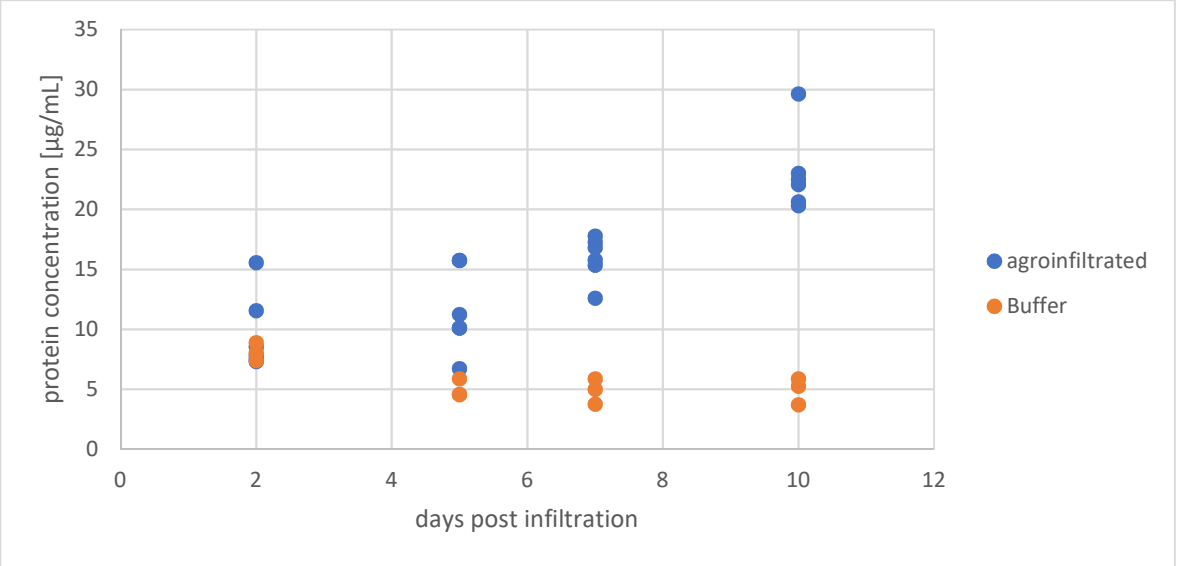
2.7 Supplementary information

Figure 2.S01: Transcripts with differential abundance between p19 and WT agroinfiltrated leaves



Testing for differential expression was performed as implemented in DESeq, using a Wald-Test and filtering for Benjamini-Hochberg adjusted p-value < 0.05. Only differentially abundant transcripts are shown.

Figure 2.S02: protein concentration in apoplastic fluid samples



Protein concentrations were determined using a microtiter-plate based Bradford assay according to (Ernst & Zor, 2010). Values are given for each biological replicate as a mean of three dilutions measured on the same plate.

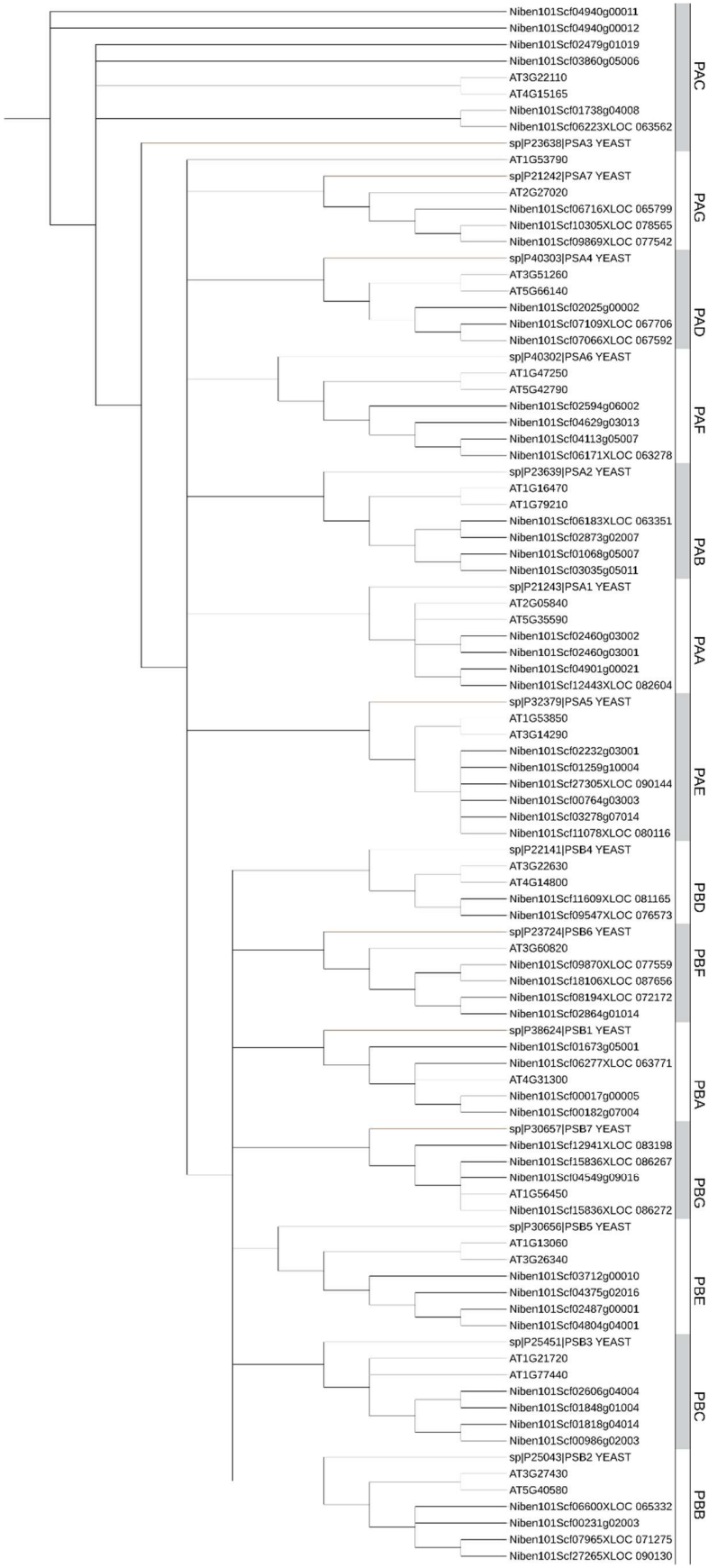
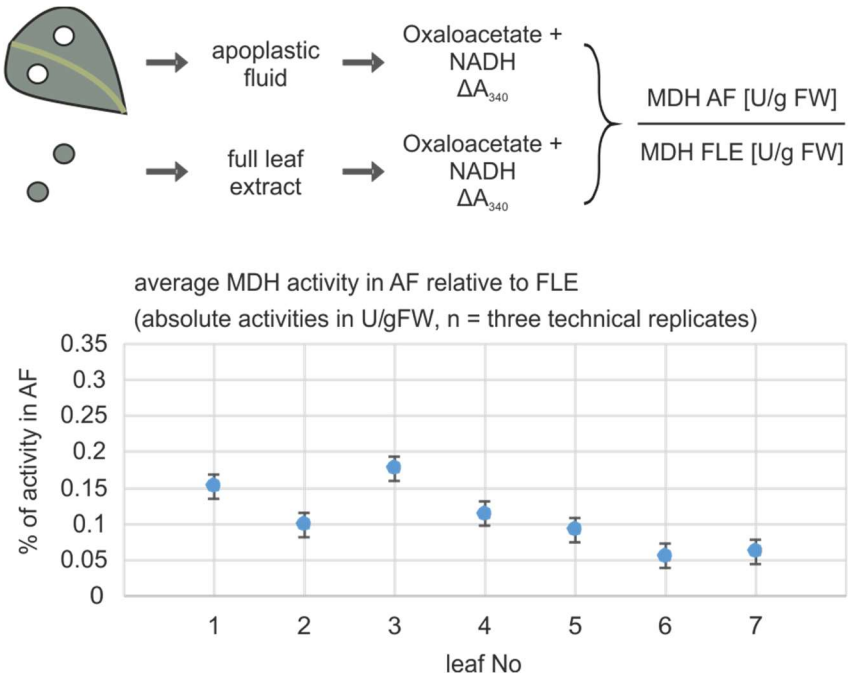


Figure 2.S03: Annotation of *N. benthamiana* proteasome subunits in MEROPS family T01
Phylogenetic tree based on the protein sequences of proteasome subunits (T01) containing all proteases and protease homologs in Arabidopsis (grey branches) and *N. benthamiana* (black branches), supplemented by yeast proteasome subunits (brown branches). Names and symbols are used as described for Figure 2.5.

Figure 2.S04: Malate dehydrogenase enzyme activity assay (MDH assay)



N. benthamiana plants were infiltrated with infiltration buffer (mock treatment) and harvested at 2 dpi. One *N. benthamiana* leaf per sample was detached, a full leaf extract (FLE) was prepared from two leaf discs and apoplastic fluid (AF) extracted from the remainder of the leaf. MDH activity in U/g fresh weight (FW) was compared between AF and FLE from the same leaf for each sample. FLE was prepared by grinding frozen tissue using metal beads and a TissueLyser (Quiagen, Hilden, DE), mixing the tissue powder with 133 µl ice-cold PBS and separating the FLE from tissue and beads by centrifugation at 4 °C, 13000 g, for 10 min. Protein concentrations of FLE and AF were determined using a Bradford assay (Ernst & Zor, 2010) and FLE was diluted to meet the protein concentration of the corresponding AF. MDH activity was measured in 50 mM Tris-HCL, pH 7.5, 0.8 mM oxaloacetate and 0.4 mM NADH, using 20 µl AF or diluted FLE for each 200 µl reaction. The reduction in absorbance at 340 nm was read every 17 seconds using a microplate reader (Tecan Group Ltd., Maennedorf, CH) and the slope used to calculate the MDH activity according to Lambert-Beers law. Mean and standard deviation of three technical replicates per sample are shown. The assay was repeated twice with similar results.

File 2.S01: A curated proteome database for apoplastic proteome analysis

During a first annotation of the transcriptome and proteome data, we observed that well-known proteases including PLCPs (MEROPS family C01) and subtilases (S08) often appeared truncated or lacked conserved domains in the Niben101 proteome database (https://solgenomics.net/). To improve the protease annotation, we evaluated the predicted proteome databases from four transcriptomes: the Niben101 transcriptome as sequenced by a consortium at the Boyce Thompson institute (https://solgenomics.net/) (DB1); the Nbv5 transcriptome as sequenced by a consortium based at the University of Sydney (Nakasugi *et al.*, 2014) (DB2); a *de novo* assembly of our own RNAseq reads (DB3) and a transcriptome generated by mapping our RNAseq reads to the Niben101 genome (Bombarely *et al.*, 2012) (DB4). We compared the predicted protein databases (DB1-4) using three different metrics.

First, we analysed protein length as a proxy for protein completeness. Protein length distributions of DB1-4 showed that our *de novo* assembly (DB3) contains a large proportion of short, probably incomplete sequences, with a median protein length of 156 amino acids (AA). This is an inherent weakness of *de novo* assemblies, which cannot draw on the genome sequence at sites with insufficient overlap between RNAseq reads. The Nbv5 transcriptome (DB2) was more carefully assembled and has a median protein length of 180 AA, but DB2 is the smallest of the four databases. DB4 has a median protein length of 172 AA and thus has a higher proportion of short sequences than DB1 (median protein length 261 AA), but DB4 is substantially larger than any of the other three databases (Figure 2.S01.1a).

Second, to assess which of the databases contains most complete plant proteins, we determined Best Hit Ratios (BHRs) between DB1-4 and the Arabidopsis TAIR10

database (Berardini *et al.*, 2015). The BHR is the number of matched positions between a *N. benthamiana* query sequence and its best blast hit in the Arabidopsis proteome divided by the length of the hit (O'Neil & Emrich, 2013). Histograms of the BHR values show that the genome-based database (DB4) has the highest number of complete best hits (BHR close to 1) (Figure 2.S01.1b).

Third, we annotated MS/MS spectra from shotgun proteomics of 36 AF samples with each of the four databases. DB4 performed best in this analysis, closely followed by the proteome predicted from the Niben101 transcriptome (DB1) (Figure 2.S01.1c).

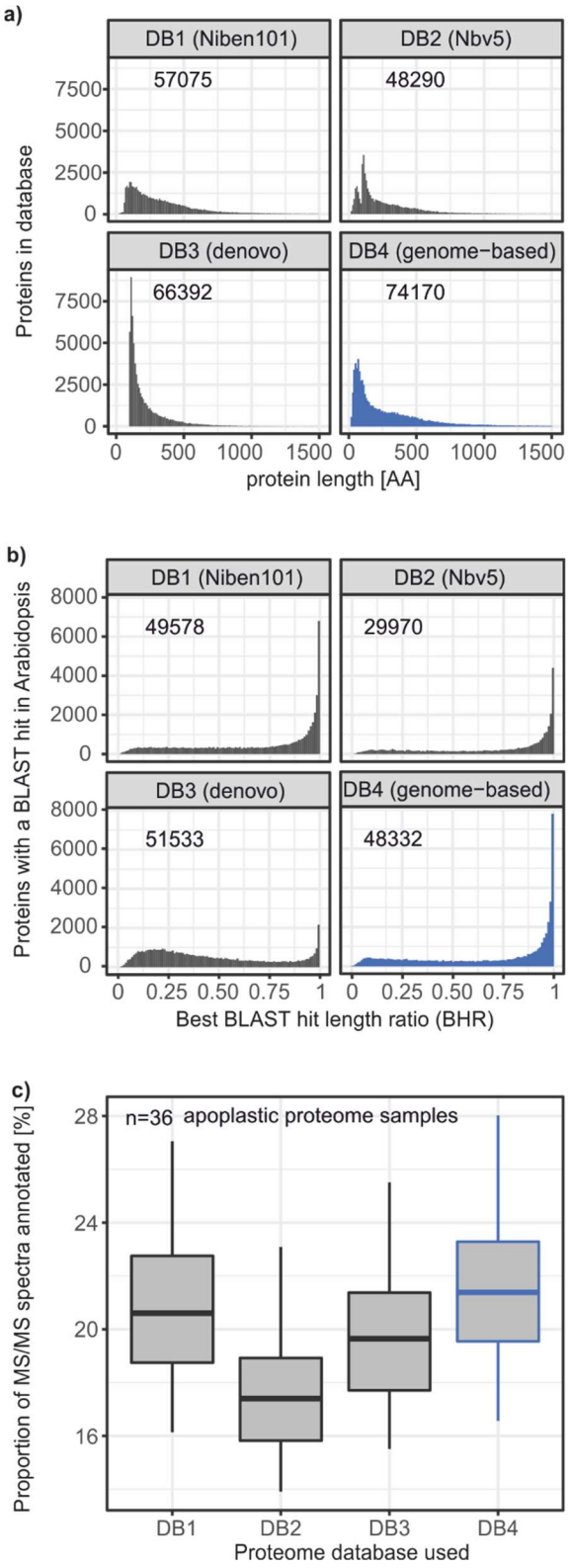


Figure 2.S01.1: Four *N. benthamiana* proteome databases Comparison of four *N. benthamiana* proteomes, predicted from the Niben101 transcriptome (DB1), the Nbv5 transcriptome (DB2), a *de novo* assembly of the RNAseq reads obtained in this study (DB3) and a genome-based assembly of the RNAseq reads obtained in this study using the Niben101 genome (DB4). a) Histograms of protein length values in the four databases and total number of sequences in each database. b) Histograms of best BLAST hit length ratio (BHR) values for the four databases and the total number of proteins with a BLAST hit in Arabidopsis in each database. c) Boxplots giving the proportion of MS spectra annotated with each database, with the upper and lower hinges corresponding to 25th and 75th percentiles, the middle line to the median and the whiskers to the lowest and highest data points within 1.5 interquartile ranges from the hinge.

We continued with our genome-based transcriptome assembly (DB4) as a starting point, because it is the largest database, has a large proportion of long sequences as well as the highest number of complete best blast hits (BHR = 1) and matched the MS data best. Protease families, however, were still poorly annotated in DB4. We therefore followed a manual curation pipeline to correct proteases in six protease families: A01, C01, C13, S08, S09 and S10. We collected the proteases from the four *N. benthamiana* proteomes and Arabidopsis TAIR10 using PFAM 30.0 (Finn *et al.*, 2016) and PFAM to MEROPS mapping (Supplementary Table

ES2.1). We generated phylogenetic trees from DB4 and TAIR10 for each protease family and used the underlying alignment to identify 197 incomplete (lacking the catalytic site or crucial motifs) and truncated sequences in DB4, corresponding to 53 % of the proteases in the six families in DB4. We replaced 64 incomplete sequences with longer versions from DB1-4. By allowing for a replacement from DB4, we ensured that duplicated sequences were collapsed into the longest available version. We retained 74 sequences that lacked the catalytic site but had no complete version in any of the four databases. We discarded 59 sequences that were truncated and had no longer version in any of the four databases (Figure 2.S01.2a). This resulted in a curated *N. benthamiana* proteome database DB5.

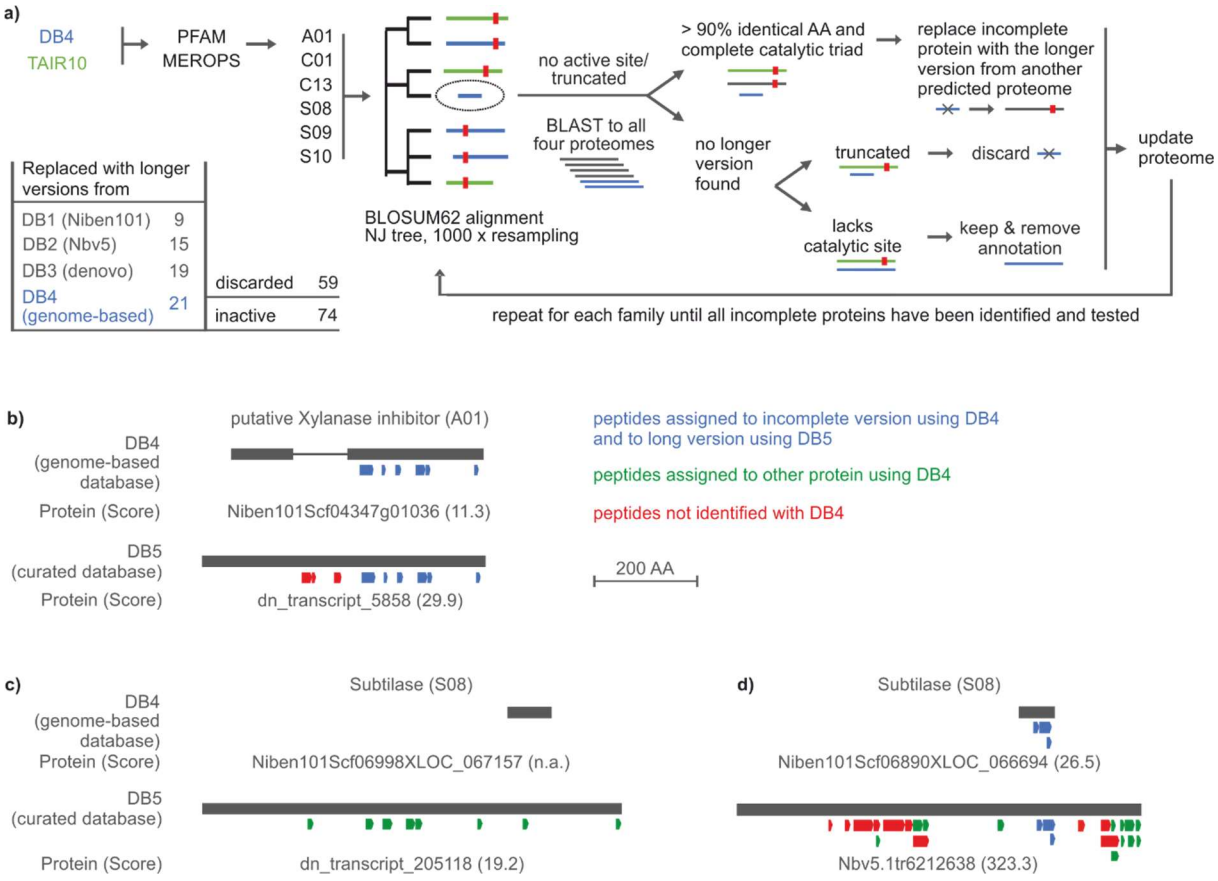


Figure 2.S01.2: Manual curation of protease families a) Flowchart detailing our manual annotation approach. b-d: Examples of proteins that could be identified with a higher confidence (higher protein score) using DB5 (curated proteome) compared to DB4 (proteome predicted from alignment of RNAseq reads to the Niben101 genome). Thick grey bars indicate protein lengths; thin grey bars indicate gaps in the alignment of the incomplete protein with its respective complete version. NJ, neighbour-joining; n.a. not applicable, as the protein was not detected.

Searching the apoplastic proteome MS spectra with DB5 allowed us to detect 30 proteins more in the apoplast than with the genome-based database DB4, showing that the considerable effort of curating is warranted. We also identified new peptides in the MS data (i.e Figures S19.2b and S19.2d) and thus increased the confidence in the protein identification, reflected in a higher protein score (Tyanova *et al.*, 2016a). Identification of new peptides with the curated database can lead to assignment of additional peptides to a protein because of the “winner takes it all”/razor principle used by the Andromeda search engine in MaxQuant (Tyanova *et al.*, 2016a) (i.e. Figures

S01.2b and S01.2d). In 151 cases, this allows us to detect proteins we were unable to identify before, either because they were not part of DB4 or because they had no peptides assigned to them (i.e. Figure 2.S01.2c).

Electronic supplementary information

Electronic supplementary information included with the thesis: Files ES2.02-08 and Tables ES2.01-15.

In Chapter 2, it became clear that *N. benthamiana* has a large, diverse repertoire of extracellular proteases from all catalytic classes. These proteases have access to recombinant proteins (RPs) that pass through the secretory pathway to become glycosylated. As outlined in Chapter 1, proteases from all catalytic classes can degrade RPs. Taken together, different proteases may redundantly degrade RPs, indicating that proteolytic activity from multiple proteases must be depleted to prevent RP degradation.

We therefore decided to use protease inhibitors (PIs) to deplete multiple protease activites. The *N. benthamiana* predicted proteome contains 111 PIs from nine families and PI repertoires in other plants are similarly large (Figure 2.2). The strongest PIs with the widest protease target ranges are promising starting points in the search for PIs that enhance RP accumulation. We thus gathered information on multifunctional PIs that can block several proteases. Interestingly, multifunctional PIs are found in many plant species, play diverse biological roles and appear to have evolved in different ways.

Chapter 3

Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants

Grosse-Holz FM, van der Hoorn RAL. **2016**. Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants. *New Phytologist* **210**: 794–807.

Authors

Friederike Grosse-Holz[1], Renier A.L. van der Hoorn[1]*

Affiliations

[1] Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

* Corresponding author

Author contributions

R.H. conceived the research. F.G.H. performed the literature searches and developed the storyline with feedback from R.H. R.H. generated the figures with feedback from F.G.H. F.G.H wrote the manuscript with feedback from R.H.

3.1 Summary

Multifunctional protease inhibitors juggle jobs by targeting different enzymes and thereby often controlling more than one biological process. Here, we discuss biological functions, mechanisms and evolution of three types of multifunctional protease inhibitors in plants. The first type are double-headed inhibitors, which feature two inhibitory sites targeting proteases with different specificities (e.g. Bowman-Birk inhibitors) or even different hydrolases (e.g. α-amylase/protease inhibitors preventing both early germination and seed predation). The second type consists of multidomain inhibitors which evolved by intragenic duplication and are released by processing (e.g. multicystatins and potato inhibitor II, implicated in tuber dormancy and defence, respectively). The third type consists of promiscuous inhibitory folds which resemble mouse traps that can inhibit different proteases cleaving the bait they offer (e.g. serpins, regulating cell death, and α-macroglobulins). Understanding how multifunctional inhibitors juggle biological jobs increases our knowledge of the connections between networks they regulate. These examples show that multifunctionality evolved independently from a remarkable diversity of molecular mechanisms that can be exploited for crop improvement and provide concepts for protein design.

3.2 Introduction

Multifunctional protease inhibitors are classic examples that contradict the old dogma of one protein – one function. This dogma accrues from the discovery that different mutations can abolish different steps in biochemical pathways, indicating that each of the mutated genes encodes one enzyme which has a particular function (Beadle & Tatum, 1941). Multifunctionality implies that a protein or a structural fold can regulate various partners, resulting in different biological outputs. About 30 years after the influential paper of Beadle and Tatum, the first report of a multifunctional protease inhibitor (Odani & Ikenaka, 1973) challenged their model. A protein from soybean was separated, yielding two parts, one of which inhibited trypsin, the other chymotrypsin (Odani & Ikenaka, 1973). Trypsin and chymotrypsin are closely related Ser proteases, but they differ in the surface loops that determine substrate specificity (Hedstrom *et al.*, 1992) and thus require distinct inhibitory sites. This first bifunctional protease inhibitor belongs to the Bowman-Birk inhibitors (BBI), classified in the MEROPS database as family I12 (see Box 3.1 for a brief description of the MEROPS system). Only a few years after the discovery that a single protein can inhibit two distinct proteases, it became clear that the Indian staple crop Ragi (*Eleusine coracana*) produces an even more peculiar multifunctional inhibitor (Shivaraj & Pattabiraman, 1981). The Ragi Bifunctional Inhibitor (RBI, family I6) can form a trimeric complex with α-amylase and trypsin, thereby simultaneously inactivating a starch-degrading enzyme and a protease (Shivaraj & Pattabiraman, 1981). Two years later, it turned out that a protein from barley also targets an α-amylase and a protease (Mundy *et al.*, 1983). However, BASI (Barley α-amylase/Subtilisin Inhibitor, family I3) is structurally unrelated to RBI. The bifunctional BBI, RBI and BASI were the first of many multifunctional plant protease inhibitors to be discovered. In these three cases,

multifunctionality is facilitated by two inhibitory sites on a single protein. More recent research has characterized other types of multifunctional protease inhibitors. Some of these contain more than one inhibitory domain (multidomain inhibitors) or use the same binding site to inhibit one of various target proteases at a time (promiscuous inhibitors) (Fig. 3.1). The frequent occurrence of multifunctionality among protease inhibitors has raised many questions about the biological roles of these proteins. Are multifunctional inhibitors regulatory links between the proteases they affect? How has multifunctionality evolved in different inhibitor structures and do they share common "roads to multifunctionality"? How can multifunctional protease inhibitors be exploited for crop improvement and protein design? These questions are addressed in this review. With multifunctional plant protease inhibitors, we bring together a range of proteins and biological processes that are not usually studied jointly. Awareness of a potential hub between proteolytic networks will help to gain a comprehensive overview. Finally, understanding the inherent power of multifunctional protease inhibitors will prove useful to design successful agricultural or biotechnological strategies.

Box 3.1: The MEROPS database of proteases and their inhibitors

(http://merops.sanger.ac.uk, (Rawlings *et al.*, 2014))

MEROPS groups proteins into families based on sequence homology and families into clans based on structural homology. Proteases or inhibitors in a clan are assumed to have evolved from a common ancestor, but are less closely related than the proteins within a family. Inhibitor families are named I1 to I93. Protease families are named with a letter indicating the catalytic type (i.e. A for Aspartic, S for Serine and C for Cysteine proteases), followed by a consecutive number. We refer to MEROPS release 9.12 throughout this article.

3.3 Three types of multifunctionality

Juggling several biological jobs is facilitated in different ways, which we group into three types of inhibitor multifunctionality (Fig. 3.1). The first type of multifunctional inhibitors are Janus-type inhibitors, double headed proteins with two inhibitory faces matching distinct targets. The second type are multidomain proteins, which resemble a necklace of pearls (inhibitor domains). Within the pearl necklace type, the I20 inhibitor domains have diversified with regard to specificity, while the macrocyclic cystine knot peptides can have both inhibitory and non-inhibitory functions. The third type of multifunctionality is represented by promiscuous inhibitors such as serpins and α-macroglobulins. These structurally different inhibitors act as mouse traps that inhibit different proteases by undergoing dramatic conformational change upon protease binding.

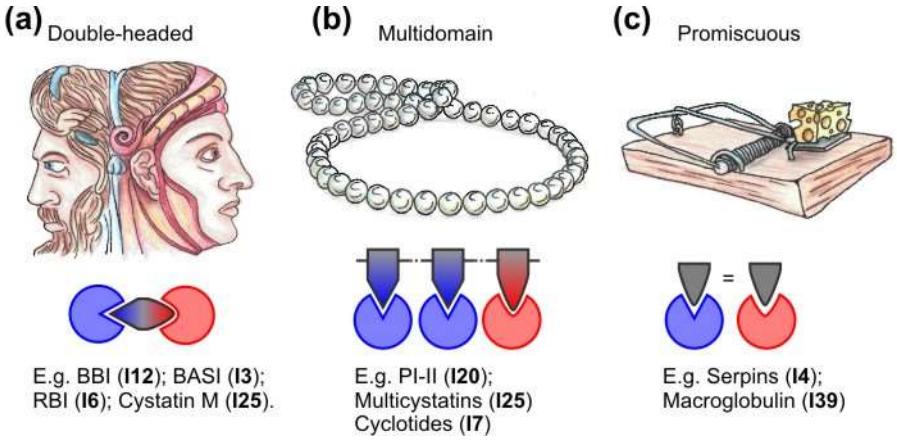
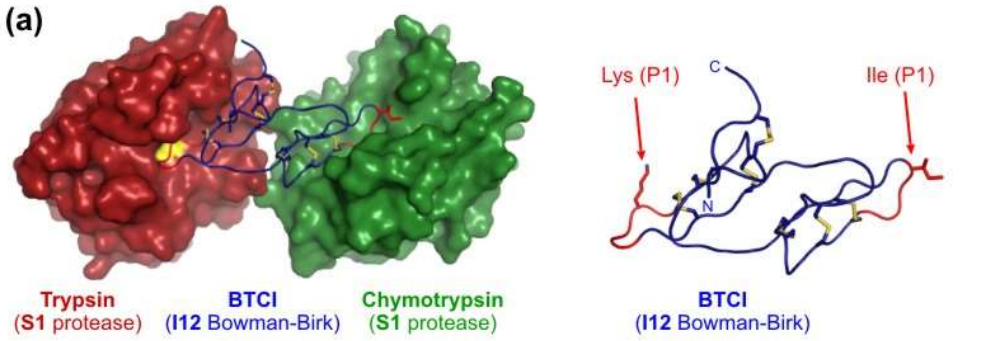
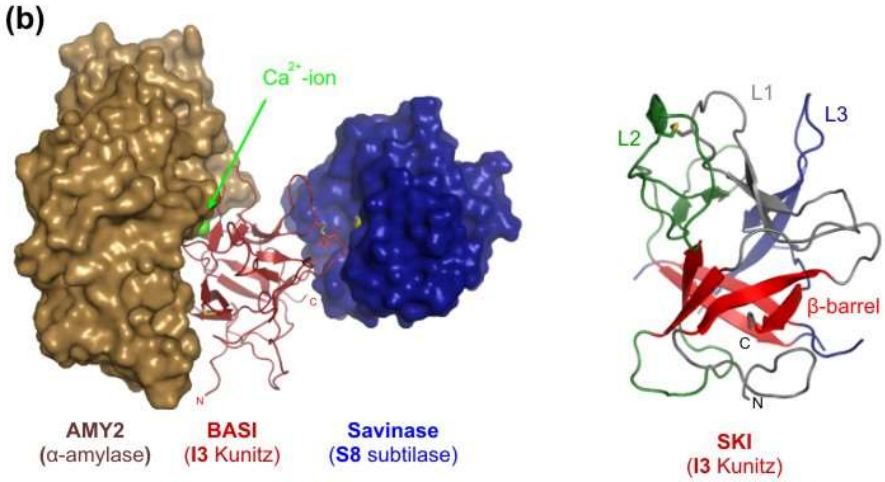


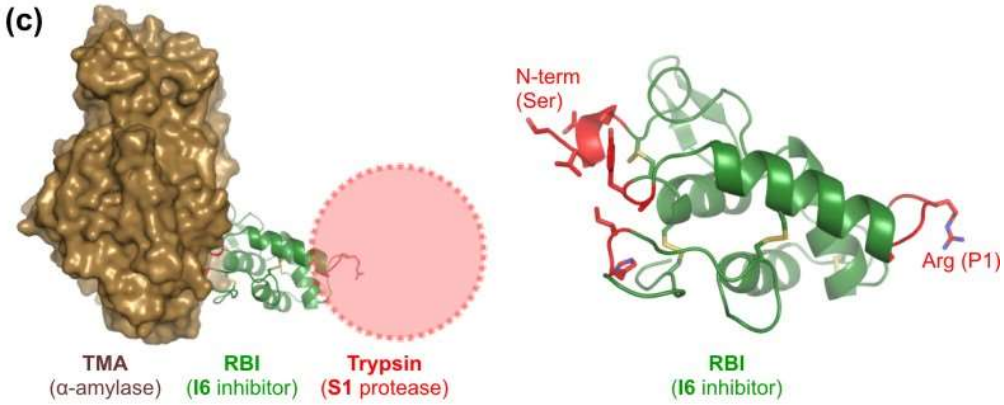
Figure 3.1: Three classes of multifunctional protease inhibitors.

We summarize the current scientific knowledge on biological role, inhibitory mechanism and evolution for eight families of multifunctional plant protease inhibitors. A comprehensive overview of all discussed families and the examples mentioned throughout this review is given in Table 3.1. The inhibitors are first discussed in groups

based on their type of multifunctionality. We conclude by discussing all eight multifunctional inhibitor families in the context of common biological roles, evolutionary history and potential applications.


(a)
Trypsin
(S1 protease)
BTCI
(I12 Bowman-Birk)
Chymotrypsin
(S1 protease)
Lys (P1)
Ile (P1)
C
N
BTCI
(I12 Bowman-Birk)


(b)
Ca2+-ion
C
N
AMY2
(α-amylase)
BASI
(I3 Kunitz)
Savinase
(S8 subtilase)
L1
L2
L3
β-barrel
C
N
SKI
(I3 Kunitz)


(c)
TMA
(α-amylase)
RBI
(I6 inhibitor)
Trypsin
(S1 protease)
N-term
(Ser)
Arg (P1)
RBI
(I6 inhibitor)

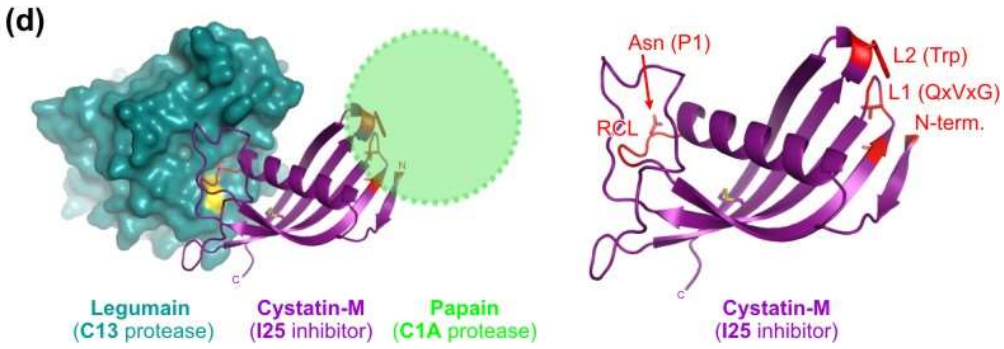

(d)
Legumain
(C13 protease)
Cystatin-M
(I25 inhibitor)
Papain
(C1A protease)
Asn (P1)
L2 (Trp)
L1 (QxVxG)
RCL
N-term.
C
Cystatin-M
(I25 inhibitor)

Figure 3.2: Four examples of double-headed inhibitors. (a) Left: Crystal structure (PDB ID 3ru4) of *Vigna unguiculata* Bowman-Birk trypsin and chymotrypsin Inhibitor (purple, BTCI, family I12) in complex with bovine trypsin (red, family S1) and bovine chymotrypsin (green, family S1). Right: BTCI inhibits trypsin and chymotrypsin using loops (red) that contain Lys and Ile residues mimicking the P1 substrate recognition sites of the respective target enzymes. (b) Left: Crystal structures of *Hordeum vulgare* (barley) α-amylase/subtilisin inhibitor (BASI, family I3, PDB ID 3bx1) in complex with barley α-amylase (AMY2, brown, PDB ID 1ava) and *Bacillus lentus* Savinase (blue, family S8, PDB ID 3bx1). A calcium ion (green) acts as molecular glue in the AMY2-BASI interaction. Right: The *Glycine max* (soybean) Kunitz inhibitor (SKI, family I3, PDB ID 1avu) illustrates the overall structure of Kunitz inhibitors: a β-barrel (red) with three extended loops (L1-L3), representing a tree trunk and branches, respectively. (c) Left: Crystal structure of Ragi Bifunctional Inhibitor (RBI, family I6, PDB ID 1tmq, green) from *Eleusine coracana* (Ragi or Indian finger millet) with *Tenebrio molitor* (yellow meal worm) α-amylase (TMA, brown, PDB ID 1tmq). The location where a trypsin-like protease could bind is indicated by a red circle. Right: Inhibition is facilitated by the N-terminal Ser residue of RBI (red) that interacts with the active site of TMA. The loop that can interact with trypsin-like proteases (family S1) contains an Arg residue (red) that mimics the P1 substrate recognition motif for trypsin. (d) Left: Crystal structure (PDB ID 4n6o) of human Cystatin M (purple, family I25) with human legumain (cyan, family C13). The location where a papain-like protease could bind is indicated by a green circle. Right: Legumains specifically cleave after Asn (P1 = Asn), therefore selectivity of the interaction is facilitated by an Asn residue in the Reactive Center Loop (RCL, red). The three regions of Cystatin M that would bind the substrate-binding groove of papain-like Cys proteases are located on the opposing side (red).

Table 3.1: Plant protease inhibitor families discussed in this review

name	family	type	inhibitory mechanism	target enzymes	implicated in defence	implicated in development	examples discussed in this review
Bowman-Birk inhibitors (BBIs)	I12	DH	Laskowski	trypsin (S1), chymotrypsin (S1)	- expression induced during plant immune responses[1] - overexpression increases resistance to a fungal pathogen[1]	ND	**BBI** (*Glycine max, Medicago truncatula, Oryza sativa, Zea mays*)
Ragi bifunctional inhibitor (RBI)	I6	DH	Laskowski	α-amylase, trypsin (S1)	- inhibits α-amylases from organisms that feed on seed, such as insects[3] and mammals[4]	ND	**RBI** (*Eleusine coracana*)
Kunitz inhibitors targeting different proteases	I3	DH	Laskowski (BASI: modified Laskowski)	α-amylase, subtilisin-like proteases (S8), PLCPs (C1), trypsin (S1), chymotrypsin (S1), cathepsin D (A1)	- inhibit insect gut proteases[7] and α-amylases[2] - reduce viability and fertility of insects [8]	- inhibit α-amylases involved in germination[5]	**BASI** (*Hordeum vulgare*) **ApKTI** (*Adenanthera pavonina*), **PSPI** (*Solanum tuberosum*), **KTI3** (*Populus trichocarpa x Populus deltoides*)
legumain-inhibiting cystatins	I25	DH	cystatin mechanism (C1), substrate mimicry (C13)	PLCPs (C1), legumains (C13)	ND	- may prevent early germination[6]	**AtCYS6** (*Arabidopsis thaliana*)
multicystatins		MD		PLCPs (C1)	- expression induced during plant immune responses[9] - inhibit insect gut proteases[10]	- regulate storage protein	**PMC** (*Solanum tuberosum*)

name	family	type	inhibitory mechanism	target enzymes	implicated in defence	implicated in development	examples discussed in this review
					- impair the growth of insects and fungal pathogens[11]	accumulation in tubers[12]	
multidomain Potato peptidase inhibitor II	I20	MD	Laskowski	trypsin (S1), chymotrypsin (S1), subtilisin (S8)	- expression increases upon wounding[13] - inhibit insect gut proteases[14] - overexpression increases resistance to insect pests[15]	ND	**PI II** (*Solanum tuberosum*), **NaProPI** (*Nicotiana alata*)
squash inhibitors	I7	MD	Laskowski	trypsin (S1)	ND	ND	**MCoTI-II** *(Momordica cochinchinensis)*
serpins	I4	P	serpin mechanism (deformation)	PLCPs (C1), metacaspases (C14), chymotrypsin (S1)	- impair insect growth and fertility[16]	ND	**AtSerpin1** (*Arabidopsis thaliana*), **BSZx** (*Hordeum vulgare*)
α-macroglobulins	I39	P	macroglobulin mechanism (caging)	endopeptidases	ND	ND	**A2M** (*Cucumis sativus*, *Fragaria vesca*, *Micromonas sp. RCC299, Populus trichocarpa*)

DH = double headed (Janus type); MD = multidomain (pearls on a string type); P = promiscuous inhibitors (mouse trap type); PLCPs = papain-like Cys proteases, ND = not determined. [1](Rakwal *et al.*, 2001; Qu *et al.*, 2003), [2](Pekkarinen & Jones, 2003), [3](Strobl *et al.*, 1998), [4](Maskos *et al.*, 1996), [5](Abdul-Hussain & Paulsen, 1989; Nielsen *et al.*, 1995; Vallée *et al.*, 1998), [6](Hwang *et al.*, 2009), [7](Da Silva *et al.*, 2014), [9](Siqueira-Júnior *et al.*, 2002; Uppalapati *et al.*, 2005; Girard *et al.*, 2007), [10](Orr *et al.*, 1994; Siqueira-Júnior *et al.*, 2002), [11](Orr *et al.*, 1994; Siqueira-Júnior *et al.*, 2002), [12](Mignery *et al.*, 1988; Pouvreau *et al.*, 2001; Weeda *et al.*, 2009), [13](Graham *et al.*, 1985; Kong & Ranganathan, 2008), [14](Tamhane *et al.*, 2005; Joshi *et al.*, 2014), [15](Johnson *et al.*, 1989; Tamhane *et al.*, 2009; Dunse *et al.*, 2010; Joshi *et al.*, 2014), [16] (Fluhr *et al.*, 2012)

3.3.1 Double-headed inhibitors: the Janus-type

With his two faces looking in opposite directions, the ancient Roman God Janus provides a good metaphor for bifunctional inhibitors. Each of the two faces stands for a binding site on which a target enzyme can be inhibited. Surprisingly many protein architectures facilitate multifunctionality by providing two sites with different inhibitory specificity, thus the Janus-type inhibitors include representatives from at least four MEROPS families that occur in plants (Table 3.1).

3.3.1.1 Bowman-Birk inhibitors (I12): losing and gaining multifunctionality

The first Janus-type inhibitor was discovered in soybean flour by Bowman in 1946 (Bowman, 1946). The inhibitor was further purified and characterized by Yehudith Birk, who showed that it inhibits both trypsin and chymotrypsin (Birk, 1961). However, it was not yet clear that this was due to two binding sites. Separation of the protein in two fragments which retained inhibitory activity towards either trypsin or chymotrypsin, respectively, elucidated that the soybean Bowman-Birk inhibitor (BBI, MEROPS family I12, see Box 3.1 for a brief description of the MEROPS system) has two independent binding sites (Odani & Ikenaka, 1973), making the soybean BBI the first known multifunctional protease inhibitor. BBI-encoding genes are also present in *Medicago*, rice and maize, but they seem absent in the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana* (Rawlings *et al.*, 2014). BBIs may have a defensive function, because *BBI* gene expression in rice is upregulated in response to wounding and the defence-related phytohormone jasmonic acid (Rakwal *et al.*, 2001; Qu *et al.*, 2003) and overexpression of an endogenous Janus-type BBI in the staple crop rice increases resistance to *Magnaporthe grisea*, a fungal pathogen causing rice blast (Qu *et al.*, 2003). BBIs fold into a core of antiparallel beta-sheets that is crosslinked by multiple disulphide bridges (seven in the case of soybean BBI). The two inhibitory sites are

located on protruding loops on opposing ends of the beta-sheet core (Fig. 3.2a) and function via the Laskowski mechanism (Box 3.2) (Laskowski & Kato, 1980; Chen *et al.*, 1992; Voss *et al.*, 1996). Monocot I12 inhibitors contain up to three BBI domains

Box 3.2: The Laskowski Mechanism of Protease Inhibition

The Laskowski mechanism of inhibition is probably the most common scenario of protease inhibition by proteinaceous inhibitors (Rawlings *et al.*, 2014). Michael Laskowski described the "standard mechanism" of protease inhibition, where the inhibitor acts as a "limited proteolysis substrate" (reviewed in Laskowski and Kato, 1980). A reactive peptide bond on this limited substrate is bound by the target protease and an acyl intermediate is formed with a high association constant. However, the rate of completion of proteolytic cleavage and dissociation is very low, resulting in an apparent equilibrium between the free enzyme and inhibitor on the one hand and the complex on the other. Both the intact and the cleaved inhibitor can bind and inhibit the protease, and cleavage is reversible.

(Qu *et al.*, 2003), thus they should in theory have up to six protease inhibitory sites. However, the monocot BBI domain has lost a disulphide bridge restraining the conformation of one of the inhibitory loops. Having lost the inhibitory activity of one of the two subdomains, monocot BBIs bind only one target protease per BBI domain. Apparently, the loss of the second inhibitory site in monocots was corrected by domain duplication, so that contemporary monocot BBIs can bind multiple proteases (Song *et al.*, 1999; Park *et al.*, 2004). Surprisingly, though BBIs can inhibit trypsin and chymotrypsin, it is still unknown what the natural targets of these seed proteins are. Identification of both endogenous (plant) and exogenous (insect/bacterial/fungal) target proteases might increase our understanding of the apparent evolutionary pressure towards multifunctional plant BBIs.

3.3.1.2 Kunitz inhibitors (I3): a very versatile fold

Ser protease inhibitors from the Kunitz (I3) family exist in most higher plants, but seem to be absent from green algae genome sequences according to the current release of the MEROPS data base (Rawlings *et al.*, 2014). The inhibitors were named after Moses Kunitz, who crystallized the first representative from soybean flour (Kunitz, 1945). The I3 family includes Janus-type inhibitors that bind different target enzymes on two reactive sites, e.g. the barley α-amylase/subtilase inhibitor BASI and its rice orthologue, both of which occur in grains (Leah & Mundy, 1989; Yamagata *et al.*, 1998). α-amylase/subtilase inhibitors are believed to regulate germination, as they inhibit endogenous α-amylases (Abdul-Hussain & Paulsen, 1989; Vallée *et al.*, 1998; Nielsen *et al.*, 2003), which mobilize storage carbohydrates during germination (Fincher, 1989). Rice α-amylase/subtilase inhibitors also block α-amylases from insects that feed on the grain starch (Bellincampi *et al.*, 2004). Using its second reactive site, BASI inhibits proteases from the fungal pathogen *Fusarium culmorum* (Pekkarinen *et al.*, 2007), also indicating defensive functions that make α-amylase/subtilase inhibitors interesting candidates for crop improvement. Another member of the Kunitz family, ApKTI (*Adenanthera pavonina* Kunitz type inhibitor), occurs in seeds of the leguminous tree *Adenanthera pavonina*. ApKTI can inhibit both trypsin (S1) and papain (C1) simultaneously (Migliolo *et al.*, 2010) and is active against gut proteases from herbivorous insects, including beetles and moths ((Da Silva *et al.*, 2014), and references therein). Artificial diets containing ApKTI reduce the viability and fertility of these insects, indicating that ApKTI acts in defence against herbivorous insects. ApKTI is a promising candidate for crop improvement, as the multifunctionality of the inhibitor may help to impede insect adaptation (Da Silva *et al.*, 2014). Potato tubers also contain a Janus-type Kunitz type inhibitor, named Potato Serine Protease

Inhibitor (PSPI). PSPI can bind simultaneously to both trypsin and chymotrypsin, two S1 Ser proteases with different substrate specificities (Valueva *et al.*, 2000; Meulenbroek *et al.*, 2012). The biological role of PSPI *in planta* is unclear. In principle, PSPI could act in defence or protect storage proteins from endogenous proteases to prevent premature sprouting. Considering that PSPI is one of the most abundant proteins in potato tubers (Meulenbroek *et al.*, 2012), it probably also serves as a storage protein itself. Recently, a new biochemical function was proposed for KTI3 (Kunitz trypsin inhibitor 3), a *Populus deltoides* (poplar) inhibitor from the Kunitz family active against trypsin and chymotrypsin (Major & Constabel, 2008). KTI3 is expressed *in planta* upon exposure to heavy metals and this protein confers heavy metal resistance when expressed in transgenic yeast. Molecular modelling suggests that KTI3 can chelate copper ions, but this remains to be demonstrated experimentally (Guerra *et al.*, 2015). Multifunctionality in Kunitz inhibitors is not always achieved via the double-headed Janus structure. For instance, a Kunitz inhibitor from *Prosopis juliflora*, a South American shrub, can inhibit either trypsin (family S1) or papain (family C1) using overlapping binding sites (Franco *et al.*, 2002).

All Kunitz (I3) inhibitors share a tree-like tertiary structure called the β-trefoil fold (Fig. 3.2b). The “tree” consists of a β-barrel (the trunk) with flexible loops protruding from each side (branches and roots, respectively) (Sweet *et al.*, 1974). The protease binding sites of β-trefoil inhibitors are located in the loops (Azarkan *et al.*, 2011) and inhibition occurs via the Laskowski mechanism (Box 3.2) (Renko *et al.*, 2012), with the notable exception of the α-amylase/subtilase inhibitors.

BASI has been crystallized in complex with savinase, a subtilase from the soil bacterium *Bacillus lentus* (Micheelsen *et al.*, 2008). Structural and mutational analysis of this complex revealed that the BASI inhibitory mechanism for subtilases differs from

the canonical Laskowski mechanism. The protease inhibitory loop of BASI is shorter than usual and pulled out of the protease active site by a disulphide bridge. Thus, BASI cannot be cleaved by the subtilase, which may result in a more stable inhibitory complex. The cysteine residues facilitating this version of the Laskowski mechanism are conserved in the rice and wheat orthologues of BASI, suggesting that they function in the same way (Micheelsen *et al.*, 2008). The inhibitory sites of BASI are known to be independent because protease inhibition still occurs when the inhibitor is saturated with α-amylase (Mundy *et al.*, 1983). Inhibition of barley α-amylase occurs via a large binding interface and involves a fully hydrated Ca^{2+} ion. This way, BASI sterically hinders the access to the active site of the α-amylase (Fig. 3.2b) (Vallée *et al.*, 1998). As affinity of BASI for barley α-amylase increases with increasing pH ((Nielsen *et al.*, 2003) and references therein), it is thought that the protonation state of the amino acid side chains involved in Ca^{2+} binding may be affected by pH changes (Nielsen *et al.*, 2003). During germination, a decrease in pH might release α-amylase from BASI (Mundy *et al.*, 1983; Vallée *et al.*, 1998; Nielsen *et al.*, 2003).

Length, orientation and amino acid composition of the loops vary between the β-trefoil inhibitors, allowing them to inhibit up to two proteases of different classes (e.g.(Dattagupta *et al.*, 1999; Meulenbroek *et al.*, 2012)). Protease families inhibited by Kunitz inhibitors include S1, S8, C1 and A1 (Azarkan *et al.*, 2011; Rawlings *et al.*, 2014). Although the overall structure is highly conserved, the inhibitory motifs of the Kunitz inhibitors are diversified across the family. The core β-barrel, the conserved element of the Kunitz fold, is a rigid structure that presumably confers stability to the Kunitz inhibitors in harsh environments. The β-barrel can fold via various routes, rendering the folding process of Kunitz proteins relatively immune to point mutations in the loops. Taken together, it appears that the stable β-trefoil fold provided a platform

for evolution of a wide range of molecular recognition mechanisms (Azarkan *et al.*, 2011). In plants, this includes the inhibition of amylases and proteases via various mechanisms. In other organisms, the Kunitz fold with its versatile loops is shared by growth factors, Interleukins and DNA binding proteins (Renko *et al.*, 2012).

3.3.1.3 RBI (I6): bifunctionality by two versions of substrate mimicry

Ragi bifunctional inhibitor (RBI) was first purified from seeds of the Indian finger millet, locally called Ragi (*Eleusine coracana*) (Shivaraj & Pattabiraman, 1981). Like BASI, RBI inhibits both insect and mammalian α-amylases as well as proteases (Maskos *et al.*, 1996; Strobl *et al.*, 1998). Astonishingly, BASI and RBI are unrelated in terms of sequence, structure and inhibitory mechanisms, thus inhibition of different hydrolase classes using one multifunctional protein must have evolved twice independently. RBI adopts an all-α-fold (Fig. 3.2c), which is unrelated to any other known protease inhibitor family (Strobl *et al.*, 1995, 1998). The trypsin binding site of RBI forms a loop that can be cleaved by bovine trypsin, but RBI remains inhibitory afterwards (Strobl *et al.*, 1995), confirming that protease inhibition follows the canonical Laskowski mechanism (Box 3.2) (Maskos *et al.*, 1996). Structural analysis of a complex between RBI and α-amylase from yellow meal worm (*Tenebrio molitor*) revealed that the amylase inhibitory mechanism differs remarkably between I3 and I6 inhibitors. RBI inserts its N-terminus into the amylase active site, almost completely filling the catalytic cleft and directly interacting with the active site residues (Strobl *et al.*, 1998). The inhibitor can be displaced by large substrates (>7 saccharide units), confirming that RBI acts in a competitive, substrate-like manner on α-amylase (Maskos *et al.*, 1996). RBI can form a ternary complex with bovine trypsin and porcine α-amylase (Maskos *et al.*, 1996), confirming that the binding sites for the protease and amylase are independent, as is the case for BASI. The extensive *in vitro* studies on the RBI

inhibitory mechanisms have, to our knowledge, not been complemented with studies on the role of this bifunctional inhibitor *in vivo*. For instance, it is unknown whether RBI inhibits endogenous α-amylases as well as pathogen enzymes and how important this is for germination and defence. The structural similarity between RBI and seed storage proteins, the albumins, provokes the speculation that this Janus-type inhibitor evolved from the small, stable albumin fold by acquiring additional inhibitory activities.

3.3.1.4 Legumain-inhibiting cystatins (I25A): More than just an extension

Plant representatives of the cystatin family (I25) are called phytocystatins (MEROPS family I25A, see Box 3.1 for a brief description of the MEROPS system) and occur in a wide range of taxonomic groups, including green algae, mosses, monocots and dicots (Martinez & Diaz, 2008). Most phytocystatins have a molecular weight of 12 – 16 kDa and inhibit Cys proteases of the family C1. Higher plants contain additional C-terminally extended cystatins with a molecular weight of ca 24 kDa (Martinez & Diaz, 2008). The C-terminal extension is a second inhibitory site, specific for C13 proteases, called aspariginyl endopeptidases (AEPs), legumains or vacuolar processing enzymes (VPEs) (Martinez *et al.*, 2007). The presence of both C1 and C13 proteases is necessary *in vitro* for complete degradation of the bean storage protein phaseolin (Zakharov *et al.*, 2004). To prevent germination, C1 and C13 proteases may thus be co-regulated by legumain-inhibiting cystatins. Indeed, overexpression of the C-terminally extended cystatin AtCYS6 in *A. thaliana* causes reduced Cys protease activity and delayed germination, whereas *atcys6* null mutant seeds exhibit higher Cys protease activity and germinate early (Hwang *et al.*, 2009). It remains unclear, however, whether this requires the C1 or the C13 protease inhibitory activity, or both. Bifunctional cystatins may be efficient regulators of germination, as they are able to prevent both the C1- and the C13-dependent steps of storage protein hydrolysis.

Bifunctional cystatins target different proteases via different mechanisms, namely steric hindrance of C1 proteases and substrate mimicry for C13 proteases. The mechanism for C1 protease inhibition by cystatins is conserved between animals and plants (Rawlings *et al.*, 2014) and is based on a tripartite wedge consisting of the N-terminus and two hairpin loops carrying the conserved QxVxG motif (Fig. 3.2d). This wedge is inserted into the active site of the C1 protease in a tight, but reversible interaction (Stubbs *et al.*, 1990). C13 protease inhibition by phytocystatins involves substrate mimicry using an exposed Asn residue following an α-helix on the C-terminal extension. Interestingly, animal cystatins can inhibit legumains (AEPs/VPEs) without an extension, as they harbour such an Asn residue next to a helix on the back of the cystatin domain itself (Alvarez-Fernandez *et al.*, 1999). Bifunctional phytocystatins may thus have evolved from an ancient domain duplication, upon which each of the two cystatin domains lost one of the inhibitory activities (Martinez *et al.*, 2007).

3.3.2 Multidomain Inhibitors: the pearl necklace

The pearl necklace-type inhibitors consist of an array of protease inhibitory domains. Their inhibitory potential is increased through cleavage, which releases the inhibitory domains like pearls from a string. Gene duplication probably gave rise to these multidomain proteins. Notably, a new folding pathway led to reorganization of the inhibitory domains of I20 inhibitors. Cyclization of the multidomain precursor or the released inhibitory peptides occurs for I20 and I7 inhibitors, respectively. The accumulation of structurally similar domains in one protein can also facilitate new regulatory mechanisms, as seen with the multicystatins (I25).

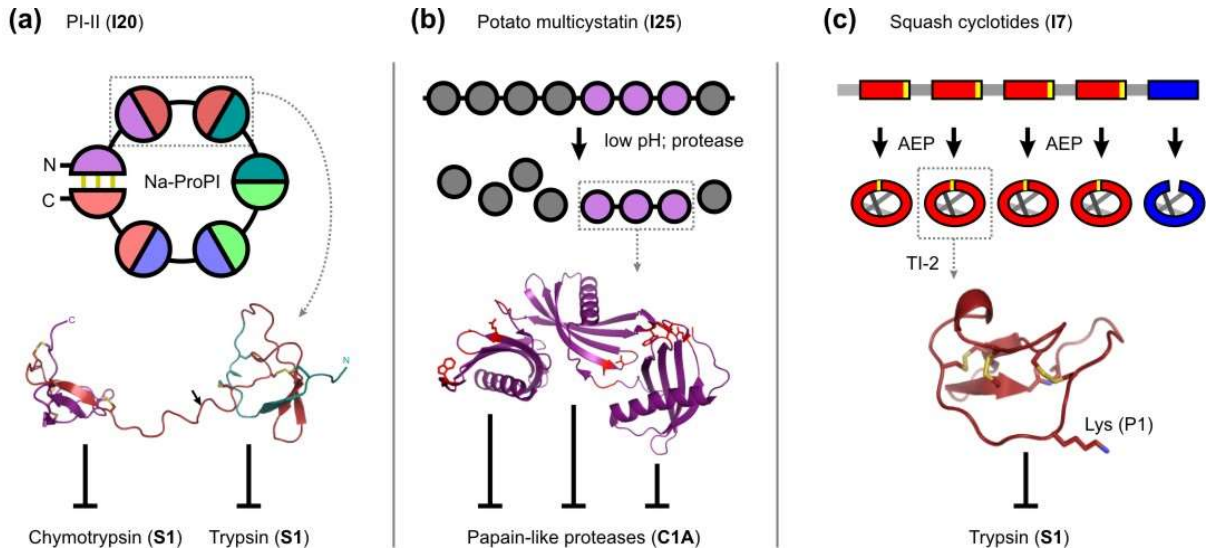


Figure 3.3: Folding and processing of multidomain inhibitors. (a) Folding and processing of the PI-II inhibitor of *Nicotiana alata* (Na-ProPI, family I20). Na-ProPI contains six sequence repeats (highlighted in different colours) that fold into six hybrid inhibitory domains (depicted as pearls). One domain consists of the N- and C-termini linked by three disulphide bridges. The inhibitory domains are released through proteolytic cleavage by Cys proteases (presumably family C13). The NMR structure of the 2nd and 3rd domain with the cleavable linker in between them has been resolved (bottom, PDB ID 1fyb). The cleavage site for release of the two domains, which target chymotrypsin and trypsin (both family S1), respectively, is indicated by a black arrow. (b) Folding and processing of potato multicystatin (PMC, family I25). The precursor consists of eight cystatin domains (depicted as pearls) that are released through proteolytic cleavage by Ser proteases at low pH. Domains 5-7 remain together and have been crystallized (PDB ID 4lzi). All released cystatin domains target papain-like Cys proteases (family C1) using a tripartite wedge inhibitory site (highlighted in red). (c) Folding and processing of squash inhibitors (family I7). The precursor consists of five sequence repeats, each but the last ending in the AEP/VPE cleavage site Asp or Asn, indicated in yellow. Four of the released peptides (red) undergo cyclization by AEP/VPE, while the last repeat (blue) remains acyclic. All five released squash inhibitors share the cystine knot motif with three disulphide bridges (grey). The crystal structure of the macrocyclic knottin MCoTI-II (PDB ID 4gux) shows the disulphide bridges linking the central loop to the periphery (yellow) and the Lys residue recognized by trypsin (family S1).

3.3.2.1 Multidomain Potato peptidase inhibitor II (I20): cyclization shuffles domains

Already in the 70s, the Japanese scientist Teruo Iwasaki determined the amino acid sequence of a fragment of potato peptidase inhibitor II (PI II) that inhibited trypsin (S1) and, to a lesser extent chymotrypsin (S1) and subtilisin (S8) (Iwasaki *et al.*, 1976). Similar proteins occur in tomato, where their production is induced by wounding

(Graham *et al.*, 1985). Today, it is clear that inhibitors of the I20 family are present in most monocots and dicots including maize, Arabidopsis, poplar and many solanaceous species. Most plants appear to only have one I20 inhibitor gene, whereas tomato and potato contain more homologues (Rawlings *et al.*, 2014). The I20 inhibitors are clearly associated with plant defence. Their expression increases upon wounding (Graham *et al.*, 1985; Kong & Ranganathan, 2008) and they are constitutively made in reproductive organs, which seem well worthy of special protection (Atkinson *et al.*, 1993). Furthermore, I20 inhibitors from pepper inhibit the gut proteases of the cotton bollworm (*Helicoverpa armigera*) *in vitro* (Tamhane *et al.*, 2005; Joshi *et al.*, 2014) and overexpression of different members of the I20 family confers increased resistance to insect pests in various species (Johnson *et al.*, 1989; Tamhane *et al.*, 2009; Dunse *et al.*, 2010; Joshi *et al.*, 2014), highlighting the potential of I20 inhibitors for crop improvement.

Structure and fold of the I20 inhibitors are prime examples for the fascinating intricacy of plant defence proteins. The *Nicotiana alata* I20 inhibitor precursor NaProPI is a well-studied example (Kong & Ranganathan, 2008). NaProPI is a 43 kDa protein produced at high levels in the female reproductive tissues of *N. alata*. The amino acid sequence of NaProPI consists of six homologous sequence repeats flanked by an N-terminal signal peptide and a C-terminal vacuolar targeting signal (Atkinson *et al.*, 1993). Both signal peptides are removed during posttranslational processing and eventually, the six-repeat precursor is cleaved into six individual protease inhibitors of 6 kDa each (Fig. 3.3a). Two of the inhibitors released from NaProPI inhibit chymotrypsin and four are specific for trypsin (Lee *et al.*, 1999). Each of the released inhibitors contains eight Cys residues forming four disulphide bonds and adopts a compact, stable fold (Nielsen *et al.*, 1995) (Fig. 3.3a). All I20 inhibitors utilize the Laskowski mechanism (Box 3.2)

(Greenblatt *et al.*, 1989), but their *in vivo* target proteases remain unknown. Five of the 6 kDa inhibitors released from NaProPI are single-chain peptides, but the sixth one consists of two chains held together by three disulphide bridges (Lee *et al.*, 1999). This peculiarity arises from the way in which NaProPI is processed into the 6 kDa mature inhibitors. Interestingly, the underlying cleavage sites are located not between, but within each of the sequence repeats of NaProPI (Heath *et al.*, 1995). After removal of the signal peptides, the NaProPI precursor adopts a cyclic conformation, linking the N-terminus to the C-terminus of the peptide chain via three disulphide bridges. This circle is then cleaved once within each sequence repeat, releasing five single-chain inhibitors as well as one two-chain protein which contains the N- and C-terminal ends of the NaProPI precursor (Lee *et al.*, 1999). NaProPI can be processed *in vitro* using C13 endopeptidases (Heath *et al.*, 1995), suggesting that asparaginyl endopeptidases (AEPs/VPEs/legumains, family C13) might release the 6 kDa inhibitors *in vivo*.

The peculiar mechanism of the NaProPI maturation prompts the question how this protein might have evolved. It is very likely that the ancestor of the I20 family consisted of a single sequence repeat folding into a single domain inhibitor. Sequence duplication events then gave rise to multi-repeat proteins that eventually adopted the cyclic fold and processing sites within each sequence repeat. This scenario is endorsed by the finding that a single ancestral NaProPI sequence repeat can fold into a functional inhibitor when it is expressed in bacteria (Scanlon *et al.*, 1999). As soon as more than one repeat is present, however, the artificial ancestral I20 inhibitor adopts a cyclic conformation and is processed within the repeats (Lee *et al.*, 1999). Notably, not all of the I20 inhibitors need to be cleaved in order to be active. The two-domain tomato inhibitor II has been crystallized in a ternary complex with two subtilisin molecules, showing that both inhibitory sites are accessible simultaneously (Barrette-

Ng *et al.*, 2003). Analysis of nonsynonymous/synonymous mutation rates within the whole I20 family revealed that that the cysteine scaffold that determines the structure of the cyclic precursors as well as of the released inhibitors is under purifying selection. This underscores the evolutionary pressure towards a cyclic precursor that may be explained by its elevated thermodynamic stability. The active residue of the released inhibitors, however, is under diversifying selection, presumably leading to specificity for different target proteases (Kong & Ranganathan, 2008).

3.3.2.2 Squash inhibitors (I7): knottins laced up

Squash inhibitors (MEROPS family I7) occur in the seeds of Cucurbitaceae and are macrocyclic knottins of less than 50 amino acids, stabilized by multiple disulphide bridges in the cystine knot motif. The squash inhibitors we discuss here are released from multidomain precursors containing up to eight repeats (Hernandez *et al.*, 2000; Mylne *et al.*, 2012). Given their abundance and strong inhibitory activity against trypsin-like proteases used by mammalian, insect and fungal seed predators, squash inhibitors might be involved in defending seeds against predation (Burman *et al.*, 2014; Mahatmanto, 2015), but a role as storage proteins in seeds has also been envisaged (Mahatmanto et al., 2015). To our knowledge, no protease has yet been identified as a natural target of squash inhibitors.

The cystine knot is a common motif that stabilizes squash inhibitors as well as a range of other, unrelated plant peptides known collectively as the cyclotides. The knot is based on a hairpin of two antiparallel β-strands containing three cysteines linked to the periphery via three disulphide bridges. A third, non-standard β-strand is present in one of the loops, which are otherwise not canonically structured (Fig. 3.3c) (Saether *et al.*, 1995; Craik *et al.*, 1999; Mylne *et al.*, 2012). Squash inhibitors are released from multidomain precursors and cyclized via transpeptidation by asparaginyl

endopeptidases, also called vacuolar processing enzymes (AEPs/VPEs, family C13) (Saska *et al.*, 2007; Gillon *et al.*, 2008; Mylne *et al.*, 2012). The last repeat of the inhibitor precursor remains acyclic, but nevertheless inhibits trypsin (Mylne et al., 2012). The squash inhibitors are members of the I7 family that shares the Laskowski mechanism of inhibition (see Box 3.2) (Otlewski & Zbyryt, 1994; Hernandez *et al.*, 2000).

Interestingly, macrocyclization of cystine knot peptides seems to have evolved several times in parallel. The whole subtropical cucurbit genus Momordica has an acyclic, single-domain cystine knot inhibitor, while multidomain precursors encoding macrocyclic knottins are only found in a subgroup of related species, indicating they may have arisen through recent gene duplication events (Mahatmanto, 2015). Additionally, cyclotides occur in the Rubiaceae, Violaceae, Fabaceae and Solanaceae. The corresponding precursors differ between species with regard to the number of cystine knot repeats (Mylne et al., 2012). All of these precursors share the recognition sites necessary for cleavage and cyclization by AEPs/VPEs (Hara-Nishimura *et al.*, 1991; Hiraiwa *et al.*, 1999) and a Gly residue needed to form a circle (Jennings *et al.*, 2001). The ability of AEPs/VPEs to cyclize peptides may have acted as an evolutionary channel, lending the selective advantage of increased stability to cystine knot peptides that start with a Gly and end in Asp/Asn (Mylne et al., 2012).

Macrocyclic knottins are an excellent template for artificial multifunctionalization. Different engineered MCoTI-II (Momordica cochinchinensis trypsin inhibitor-II) knottins specifically inhibit proteases from different catalytic classes, namely chymotrypsin (family S1, Ki in nM range), subtilisin (family S8, Ki in μM range), a viral Cys protease (family C3, Ki in in μM range) (Thongyoo *et al.*, 2009), tryptase and leukocyte elastase (both family S1) (Thongyoo et al., 2009).

3.3.2.3 Multicystatins (I25A): release at the right moment

Phytocystatins form a separate subfamily I25A within the cystatin family I25 (Margis *et al.*, 1998; Rawlings *et al.*, 2014). Phytocystatins inhibit papain-like Cys proteases (family C1) and are common among plants from green algae to Arabidopsis, tomato, rice and maize (Rawlings *et al.*, 2014). Some members of the phytocystatin family also inhibit C13 proteases using a C-terminal extension, as discussed in section 3.3.1.4. In this paragraph, we focus on the numerous cases where phytocystatins occur as multidomain proteins (Rawlings *et al.*, 2014) consisting of up to eight inhibitory cystatin domains (Walsh & Strickland, 1993). Multidomain cystatins (multicystatins) play various biological roles in both potato and tomato. The potato multicystatin (PMC) occurs in potato tubers and inhibits tuber proteases (Kumar *et al.*, 1999). PMC is produced during early stages of tuber formation, associated with a decrease in proteolytic activity in the tuber, which may facilitate storage protein accumulation. Indeed, accumulation of PMC up to 12% of total soluble tuber protein precedes the accumulation of patatin, the main storage protein in potatoes that makes up 40% of the soluble protein content (Mignery *et al.*, 1988; Pouvreau *et al.*, 2001; Weeda *et al.*, 2009). In ageing tubers, proteolytic activity increases again, coinciding with a drop in the detectable levels of PMC (Kumar *et al.*, 1999). Thus, PMC abundance appears to regulate the process from building up the protein reserves in potato tubers to their mobilisation during ageing and germination. Multicystatin expression in tomato leaves is induced by various elicitors of plant immunity (Siqueira-Júnior *et al.*, 2002; Uppalapati *et al.*, 2005; Girard *et al.*, 2007) and tomato multicystatin inhibits proteases from insect digestive tracts and impairs the growth of plant pathogenic fungi *in vitro* (Siqueira-Júnior *et al.*, 2002). Similar experiments show that potato multicystatin

inhibits the growth of corn rootworm larvae, predators of an important food crop, when added to an artificial diet (Orr *et al.*, 1994).

Multicystatins have the intriguing capacity to crystallize in native tissues. Cys protease-inhibiting crystals were first observed in potato tubers (Cohn, 1859; Rodis & Hoff, 1984), but they also occur in tomato leaves (Akers & Hoff, 1980). Multicystatin crystals localize to the cytosol in potato (Nissen *et al.*, 2009) as well as in tomato (Madureira *et al.*, 2006). The transition between crystalline and soluble states for PMC is now quite well understood. Solubilisation of PMC occurs at mildly acidic pH, exposing the inhibitory domains to target cysteine proteases (Orr *et al.*, 1994). Recent structural analyses show that low pH weakens the interdomain interactions in PMC, explaining this regulatory mechanism (Green *et al.*, 2013). As soon as it is soluble, PMC can be cleaved by serine proteases (Walsh & Strickland, 1993) to release three fragments, which collectively contain eight cystatin domains. Inhibition of Cys proteases by each of the eight domains occurs via the cystatin mechanism (section 1.4). The acidic environment of certain areas in the insect gut is thought to activate the inhibitor exactly when and where it is needed to hinder the protein catabolism of the insect pathogen, but not the host plant (Green *et al.*, 2013). Multicystatins presumably evolved from an ancestral, single-domain plant cystatin via gene duplications. Although it is known that phytocystatins inhibit papain-like cysteine proteases (family C1) (Tajima *et al.*, 2011), it is yet unknown which proteases are regulated by multicystatins (Benchabane *et al.*, 2008).

3.3.3 Promiscuous inhibitory folds: the mouse trap type

In this section, we describe two protease inhibitor families that utilize a mechanical trapping mechanism to sequester their target proteases. The nature of these mechanisms is destructive: inhibition is irreversible and dooms both the protease and the inhibitor for degradation. The active site that is recognized by the enzyme can vary in mouse trap type inhibitors without affecting functionality of the trap. Accordingly, mouse trap type inhibitors have developed multifunctionality in the sense that the same inhibitory fold can be used to target different proteases.

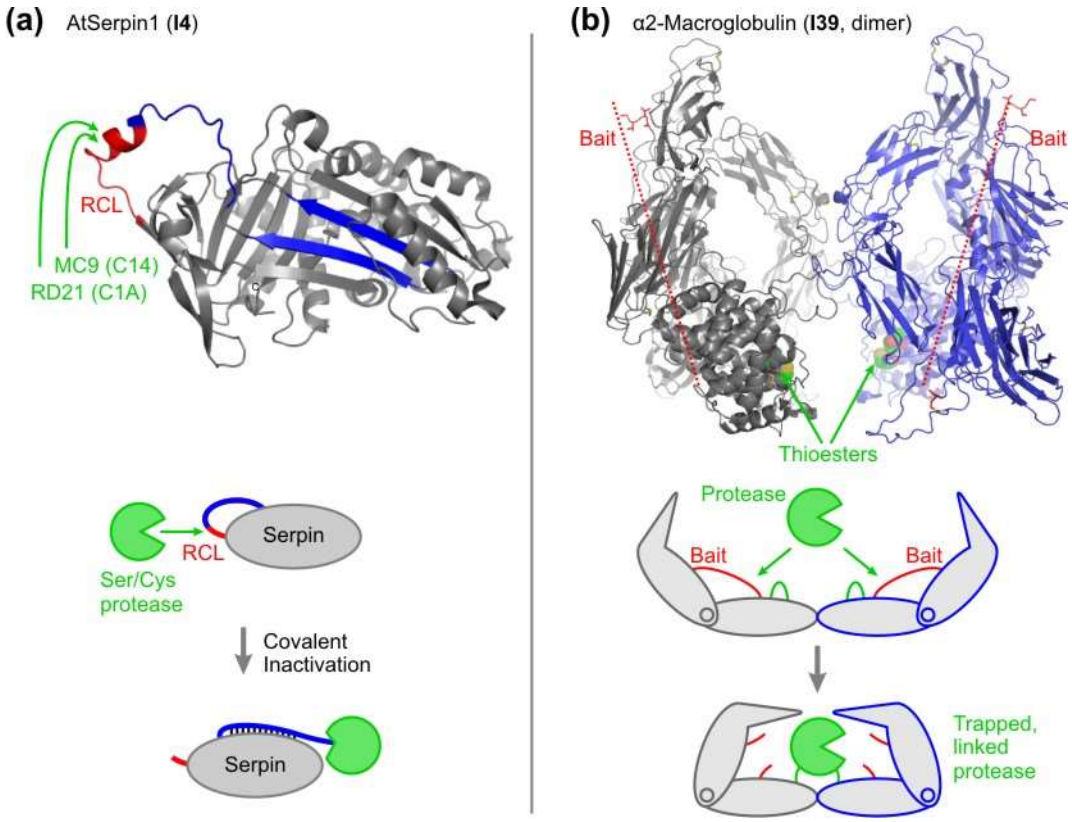


Figure 3.4: Serpin and α2-Macroglobulin, the promiscuous mouse traps. (a) Top: Crystal structure of the serpin AtSerpin1 (family I4, PDB ID 3le2) showing the reactive centre loop (RCL, red) that can be cleaved by C1 papain-like Cys protease RD21 (family C1A) or Metcaspase-9 (MC9, family C14). Cleavage causes a conformational change that inserts the loop (blue) as a new anti-parallel β-sheet in between the two β-sheets (blue arrows). Bottom: Schematic illustration of the mouse trap mechanism. Cleavage of the RCL results in a conformational change that insets the blue strand of the loop into the serpin structure, irreversibly deforming the serpin and the covalently trapped protease. (b) Top: Crystal structure of human α2-macroglobulin (family I39, PDB ID 4acq). Only two of the four subunits are shown in grey and blue, respectively. The bait peptide structure is unresolved and indicated with a red dashed line. The Cys-Glu thioester (green) traps the protease covalently. Bottom: Simplified illustration of the inhibitory mechanism of α2-macroglobulin. Any small size protease that cleaves the promiscuous bait peptide triggers a conformational change that traps the protease in a cage. The thioester (green loop) reacts with any Lys residue on the surface of the protease, thereby covalently immobilizing the protease in the cage.

3.3.3.1 Serpins (I4): springing the mouse trap on Ser and Cys proteases

Ser and Cys protease inhibitors of the MEROPS family I4, called serpins, occur in the grains of cereals of the Triticeae tribe, including barley, rye, and wheat, and in oat (Poeae tribe) (Roberts, 2003). As highly abundant grain proteins, serpins are found even in fully processed beer (Hejgaard & Kaersgaard, 1983; Hejgaard *et al.*, 1985). The biological role(s) of serpins remain somewhat unclear, although there are hints in two directions. First, artificial diets containing serpins impair growth and fertility of insect pests, indicating a potential role in defence and highlighting the potential serpins hold for crop improvement (Thomas *et al.*, 1994, 1995; Alvarez-Fernandez *et al.*, 1999; Yoo *et al.*, 2000). Second, two putative plant target proteases of the *Arabidopsis thaliana* serpin 1 (AtSerpin1) are involved in programmed cell death (PCD), suggesting that AtSerpin1 could have a pro-survival function. AtSerpin1 inhibits metacaspase AtMC9 (family C14) *in vitro* and co-localizes with AtMC9 *in vivo* in the extracellular space (Vercammen *et al.*, 2006). Electron microscopic images of *atmc9* mutant *A. thaliana* leaf cells suggest a role for AtMC9 in clearance of the cell contents after tonoplast rupture (Bollhöner *et al.*, 2013). The second putative AtSerpin1 target protease is the C1 protease RD21 (Responsive to Desiccation 21), which was identified as an *in vivo* interaction partner during pull-down of AtSerpin1 from plant extracts (Lampl *et al.*, 2010). The RD21 precursor protein accumulates in ER-derived protease storage bodies, which fuse with each other and the vacuole under stress conditions. Fusion is believed to lead to activation of the proteases which assist in recycling of cellular contents during stress-induced PCD (Hayashi *et al.*, 2001). In line with this role of RD21, PCD during plant infection with the necrotrophic fungi *Botrytis cinerea* and *Sclerotina sclerotiorum* is accelerated in the a*tserpin1* knock-out, but prevented in plants lacking *RD21*. Interestingly, Lampl *et al*. detected AtSerpin1-GFP

fusions expressed from the endogenous promoter mainly in the cytoplasm (Lampl *et al.*, 2013), in contrast to the previously reported extracellular localization (Vercammen *et al.*, 2006). In addition, increased resistance of *rd21* mutant leaves to infection by *Botrytis cinerea* is in contrast to earlier findings that *rd21* mutant plants are more susceptible than the wild type to *Botrytis cinerea* (Shindo *et al.*, 2012). Some serpins act as traps with a versatile bait, like the barley serpin BSZx. Using overlapping sites within its reactive center loop, BSZx can inhibit trypsin, chymotrypsin and, to some extent, cathepsin G (all family S1) *in vitro* (Dahl *et al.*, 1996). Evolutionarily, inhibitory serpins are conserved throughout the kingdoms of life, although they vary widely between uni- and multicellular organisms (Roberts *et al.*, 2004). *Chlamydomonas* serpins, for instance have a distinct intron-exon structure from higher plant serpins. The reactive centre loops differ remarkably between monocots and dicots, suggesting that serpins might have diverged with regard to target proteases and biological functions (Roberts & Hejgaard, 2007).

With regard to structure and mechanism, serpins are unique among plant protease inhibitors. The structure of the 43 kDa inhibitor AtSerpin1 has been resolved by crystallography and consists of three conserved β-sheets and nine conserved α-helices, as common among animal serpins (Fig. 3.4a) (Lampl *et al.*, 2010). Serpins inhibit their target proteases in a unique, suicidal manner. Ser and Cys proteases cleave the serpin reactive loop, forming an acyl-enzyme intermediate with the serpin. Cleavage triggers a profound conformational change in the metastable fold of the serpin, which deforms the active site of the protease and irreversibly binds it to the inhibitor (Fig. 3.4b). This remarkable trapping mechanism (Huntington *et al.*, 2000) allows for a structural separation of inhibitory activity from protease specificity. Thus,

members of the serpin family share the same fold, but target a range of Ser and Cys proteases (Fluhr *et al.*, 2012; Rawlings *et al.*, 2014).

3.3.3.2 α-macroglobulin (I39): the gilded cage

Functional macroglobulin genes are annotated in only a few plant species, according to the MEROPS database (Rawlings *et al.*, 2014) and a recent comparative genomics study (Santamaría *et al.*, 2014). This includes cucumber (*Cucumis sativus*), alpine strawberry (*Fragaria vesca*), the alga *Micromonas sp. RCC299* and black cottonwood (*Populus trichocarpa*). However, the structure, specificity and mechanism of action of this family are intriguing. Macroglobulins are large (ca 200 kDa) glycoproteins and their structure resembles a round cage (Fig. 3.4c) (Sottrup-Jensen, 1989; Sottrup-Jensen *et al.*, 1989). α-macroglobulins possess an exposed bait region with recognition sites for various types of endopeptidases, cleavage of which triggers a conformational change. Thus, the protease gets trapped inside the large macroglobulin protein, much like in a cage (Fig. 3.4d) (Feldman *et al.*, 1985). The caged peptidase cannot bind large targets or inhibitors any more, but remains accessible for small molecules (Sottrup-Jensen, 1989). Variation of the bait region does not affect the inhibitory mechanism, allowing for multifunctionality of the macroglobulin fold (Sottrup-Jensen *et al.*, 1989). The physiological role of macroglobulins in plants remains obscure. The versatile functions of this protease inhibitor family in animals and bacteria are reviewed elsewhere (Rehman *et al.*, 2013).

3.4 Significance of multifunctional protease inhibitors in the plant research arena

Multifunctional protease inhibitors represent hubs that regulate distinct branches of the plant physiological network, for instance defence and tuber sprouting in the case of the potato multicystatin (family I25). For one inhibitor to do several biological jobs, it must often target multiple proteases of different families or even different catalytic classes. This can be achieved through several inhibitory interfaces on the same protein, as seen with the Janus-type and the multidomain I20 inhibitors, or through one promiscuous interface, as seen in the mouse trap type inhibitors. A different way to assign multiple biological functions to one inhibitor is to produce it on different occasions in space and time. For instance, multicystatins (family I25) accumulate in tomato and potato leaves in response to wounding as well as in potato tubers when building up storage protein reserves. Through whichever route inhibitors acquire multifunctionality, the result is a protein that provides a link between the biological processes it regulates. Increased knowledge on multifunctional protease inhibitors will therefore promote a network-level understanding of plant physiology. The evolutionary history of inhibitors in their role as regulatory hubs could then reveal how the network was restructured over time.

Multifunctional protease inhibitors share stabilizing structural features, most notably a compact fold linked covalently by disulphide bridges. This structural similarity may resemble a common evolutionary road to multifunctionality. It appears that small, stable proteins, such as protease inhibitors, are well suited to acquire a (second) inhibitory function, as they are already fit to persist in harsh environments with high proteolytic activity. Additional protease inhibitory activities may await discovery in many small, stable proteins, including the known multifunctional inhibitors. A second,

even more obvious road towards multifunctionality is duplication of inhibitory domains, which can further lead to neofunctionalization of the duplicate, rearrangements in the overall protein structure or emergence of new regulatory mechanisms (sections 1.1 and 2).

On the applied side, recombinant expression of multifunctional inhibitors from the families I12, I13, I20 and I25 has been successfully used to generate pest-resistant crop plants (Orr *et al.*, 1994; Xu *et al.*, 1996; Siqueira-Júnior *et al.*, 2002; Dunse *et al.*, 2010). Furthermore, I3 and I20 inhibitors were used to limit proteolysis of recombinant human proteins produced in plants, tackling a major issue in molecular farming (Kim *et al.*, 2008b; Goulet *et al.*, 2012). Increasing our understanding how multifunctional inhibitors link physiological networks could facilitate new applications. For instance, recombinant protein degradation could be prevented very effectively by controlling regulators of proteolytic cascades in the plant. Inhibitors that interfere with endogenous as well as exogenous proteases could protect crops from pests and premature senescence. In medicine, multifunctional plant protease inhibitors of the family I12 have been used to limit undesired protease activity as potential anticancer drugs (da Costa *et al.*, 2014).

In one case, the concept of multidomain inhibitors has been taken further using artificial multidomain inhibitors consisting of five naturally occurring domains (families I25 and I31) to create a more stable and more potent inhibitor than its natural, single-domain counterparts. Overexpression of this custom-made multidomain inhibitor in potato increases resistance to the insect pest *Frankliniella occidentalis* (Outchkourov *et al.*, 2004). Some impressive examples of custom-made inhibitors based on a known multifunctional fold are found among the squash inhibitors (section 2.2). However, the relatively low affinity of the engineered squash inhibitors for their targets highlights that

a deeper understanding is needed to develop synthetic approaches. Knowing all essential features of an inhibitor structure, one could design completely novel inhibitors, customized for desired applications.

From the examples discussed throughout this review, it is clear that several roads to multifunctionality exist, starting from adaptation of target recognition sites in stable proteins or gene duplication, with the latter branching out in multiple directions. Fine-mapping these roads will facilitate the construction of custom multifunctional inhibitors while, at the same time, enhancing our knowledge about regulatory hubs in the plant physiological network.

In Chapter 3, we integrated the available information about multifunctional plant protease inhibitors (PIs) that can inhibit different proteases, often simultaneously. Multifunctional PIs can address the challenge posed by the large, diverse protease repertoire that is likely responsible for low RP accumulation in molecular farming (Chapter 2). The following chapter describes a screen carried out to find the best PIs to increase RP accumulation in agroinfiltrated *N. benthamiana*. Eleven tested PIs fall into the categories of multifunctionality discussed in Chapter 3.

Double-headed inhibitors are well represented in the screen. We included four *N. benthamiana* Kunitz inhibitors (NbK1-3 and MER41190) that may carry several inhibitory loops and the tomato Kunitz inhibitor *Solanum lycopersicum* Cathepsin D Inhibitor (SlCDI) that targets Asp and Ser proteases (Goulet *et al.*, 2010a, 2012). We also included a *N. benthamiana* cystatin carrying a legumain inhibitory site (LBNbCYS1). Specific domains of two multidomain PIs, a *Nicotiana alata* I20 inhibitor and *Solanum lycopersicum* cystatin 8 (SlCYS8), an eight-domain multicystatin, were included in the screen because these domains had previously increased RP accumulation (Kim *et al.*, 2008b; Sainsbury *et al.*, 2013; Jutras *et al.*, 2016). We also selected four mousetrap type PIs, two serpins from *N. benthamiana* (NbSRP-TMS and NbSRP-LRA), one barley serpin reported to inhibit a broad range of proteases in different tissues (Roberts *et al.*, 2003) and an *A. tumefaciens* protein annotated as macroglobulin.

Chapter 4 describes how RP accumulation in agroinfiltrated *N. benthamiana* can be increased by co-expression of PIs. This strategy emerges from the knowledge about the large, diverse *N. benthamiana* protease repertoire and about multifunctional PIs, assembled in Chapters 2 and 3, respectively.

Chapter 4

Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *N. benthamiana*

Friederike Grosse-Holz, Luisa Madeira, Muhammad Awais Zahid, Molly Songer, Mary Fresenko, Svenja Blaskowski, Sabrina Ninck, Geronimo Heilmann, Markus Kaiser, Renier A.L. van der Hoorn. (manuscript in preparation). Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in *N. benthamiana.*

Authors

Friederike Grosse-Holz[1], Luisa Madeira[1], Muhammad Awais Zahid[1], Molly Songer[1], Mary Fresenko[1], Svenja Blaskowski[2], Sabrina Ninck[2], Geronimo Heilmann[2], Markus Kaiser[2], Renier A.L. van der Hoorn[1]*

Affiliations

[1] Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[2] Chemische Biologie, Zentrum für Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

* Corresponding author

Running title

3 protease inhibitors for molecular farming

Author contributions

R.H. conceived the research. F.G.H. designed and performed experiments and analysed data unless specified otherwise. F.G.H. and R.H interpreted the results. L.M. provided plasmids and established conditions for recombinant protein expression. M.A.Z., M.S. and M.F. obtained recombinant protein accumulation data under the supervision of F.G.H. S.B. performed the sample preparation for MS using peptide samples provided by F.G.H. S.N., G.H. and M.K. performed LC-MS/MS and peptide and protein identification steps for MS. F.G.H. performed the statistical analysis of MS data, generated all figures and wrote the manuscript with feedback from R.H. All authors read and approved the final manuscript.

4.1 Summary

Agroinfiltrated *Nicotiana benthamiana* is a flexible and scalable RP production platform, but is impeded by low RP yields. Plant proteases can limit RP accumulation by degrading RPs. Co-expression of RPs with protease inhibitors (PIs) can enhance RP accumulation, but known PIs do not target all classes of *N. benthamiana* leaf proteases.

Here, we tested 29 candidate PIs for enhancing RP accumulation. We identified three new, unrelated PIs that enhanced RP accumulation. NbPR4 is a *N. benthamiana* Cys PI, NbPot1 a *N. benthamiana* Ser PI and HsTIMP is a human metalloprotease inhibitor. We compared the three new inhibitors to the known tomato Cys PI SlCYS8. Remarkably, accumulation of three unrelated RPs was enhanced by each PI, suggesting the mechanism of RP degradation may contain universal elements. Using known non-inhibitory mutant PIs, we demonstrate that inhibitory functions of HsTIMP and SlCYS8 are required to enhance RP accumulation. Different PIs additively enhance RP accumulation, but the effect of each PI is dose-dependent. Activity-based Protein Profiling (ABPP) revealed that the activity profiles of PLCPs, Ser hydrolases (SHs) or Vacuolar Processing Enzymes (VPEs) in leaves are unaffected by the new PIs. Quantitative leaf proteome data indicate that PI expression may trigger plant immunity.

Our data indicate that RPs are degraded stepwise by a redundant, dynamic network of plant proteases. NbPR4, NbPot1 and HsTIMP can be used to study new plant proteases and to enhance RP accumulation in industrial molecular farming.

4.2 Introduction

Molecular farming, the production of biopharmaceuticals in plants, offers speed, scalability and low risk of contamination with human pathogens when compared to insect or mammalian cell culture systems (Stoger et al., 2014). Plant-made biopharmaceuticals are now a reality, with the first product on the market (Fox, 2012) and more in clinical trials (Lomonossoff & D'Aoust, 2016). To produce biopharmaceuticals, *N. benthamiana* leaves can be genetically modified by infiltration with disarmed *Agrobacterium tumefaciens* (Agrobacterium) carrying gene(s) of interest on the transfer DNA (T-DNA) of binary plasmid(s) (Bevan, 1984). Agrobacterium delivers the T-DNA to the plant nucleus, where genes are transiently expressed. Co-expression of several transgenes is simply achieved by mixing Agrobacterium cultures delivering different transgenes before agroinfiltration. Co-expression with silencing inhibitor P19 is frequently used to boost protein overexpression by preventing the decline of the transgene transcript levels (Van der Hoorn et al., 2003). Agroinfiltration-mediated protein expression can now deliver ten million doses of the latest influenza vaccine within six weeks (Pillet et al., 2016). Large-scale agroinfiltration has also produced many different functional monoclonal antibodies (mABs) (Yusibov et al., 2016), including the Ebola neutralizing drug ZMapp (Qiu et al., 2014). To maximize efficacy and limit immunogenicity, biopharmaceuticals like ZMapp are produced with humanized *N*-glycans in the secretory pathway of genetically engineered *N. benthamiana* (Castilho *et al.*, 2014; Schoberer & Strasser, 2017).

A bottleneck on the road to commercialization of agroinfiltration for molecular farming are the relatively low yields of plant-produced recombinant proteins (RPs): 15 – 200 mg of monoclonal antibody (mAB) or < 1 g virus-like particles per kg fresh leaf (Yusibov

et al., 2016; Lomonossoff & D'Aoust, 2016). Both yield and purity of RPs are hampered by proteolytic degradation (Doran, 2006; Niemer *et al.*, 2014; Hehle *et al.*, 2015; Donini *et al.*, 2015), which can occur in the extracellular space (Hehle et al., 2011), affecting RPs as they pass through the secretory pathway for glycosylation. Several Papain-Like Cys Proteases (PLCPs), including two *N. benthamiana* PLCPs, can degrade RPs *in vitro* (Paireder et al., 2016, 2017), but the proteases degrading RPs in agroinfiltrated leaves remain to be identified.

The *N. benthamiana* protease repertoire is large and diverse. We recently described transcripts corresponding to 975 putative proteases of all catalytic classes present in agroinfiltrated leaves. We also detected peptides corresponding to 196 proteases in the extracellular space (Grosse-Holz *et al.*, 2017). RPs thus encounter a large, complex proteolytic network in agroinfiltrated *N. benthamiana*.

Past attempts to prevent RP degradation include supplementation of stabilizing agents, protease gene knockdown, subcellular targeting, fusion proteins and protease inhibitor co-expression (Mandal *et al.*, 2016). However, no "magic bullet" against degradation has been identified. Different strategies to limit proteolysis have been successful for each expression platform and RP.

Among the protease activity depletion approaches, protease inhibitor (PI) co-expression is promising for three reasons. First, many PIs can inhibit several enzymes, often from different families, overcoming protease redundancy (Grosse-Holz & van der Hoorn, 2016). Second, PIs can be targeted to the secretory pathway to escort RPs during secretion (Goulet *et al.*, 2012; Jutras *et al.*, 2016). Third, PI co-expression by agroinfiltration depletes protease activity only locally, thus protecting RPs with minimal effects on non-agroinfiltrated tissues.

To date, six PIs were found to enhance RP accumulation (Mandal *et al.*, 2016). Most of the PIs that enhanced *in planta* RP accumulation were expressed in stable transgenic plant cells: a Bowman-Birk serine protease inhibitor boosted mAB accumulation in *N. tabacum* roots (Komarnytsky *et al.*, 2006) and a *N. alata* Ser protease inhibitor (Protease inhibitor II) enhanced accumulation of Human Granulocyte Macrophage Colony Stimulating Factor (hGM-CSF) in rice suspension cells (Kim *et al.*, 2008b). Likewise, transgenic tobacco plants expressing Oryzacystatin I produced increased levels of recombinant glutathione reductase (Pillay *et al.*, 2012) and tomato Cathepsin D inhibitor (SlCDI) has been used to enhance human α_1-anti-chymotrypsin accumulation in transgenic potato leaves (Goulet *et al.*, 2010a) and boost mAB levels upon transient co-expression in *N. benthamiana* (Goulet *et al.*, 2012). However, effects of transient SlCDI co-expression on mAB levels seem to depend on plant growth conditions (Robert *et al.*, 2013). The tomato Cys protease inhibitor SlCYS8 enhances accumulation of mABs upon co-expression in agroinfiltrated *N. benthamiana* (Robert *et al.*, 2013; Jutras *et al.*, 2016) and can act as a stabilizing fusion partner to increase human α_1-anti-chymotrypsin levels (Sainsbury *et al.*, 2013). Importantly, the Q47P mutation abolishes protease inhibitory capacity of SlCYS8 and co-expression with SlCYS8-Q47P does not increase RP levels (Sainsbury *et al.*, 2013; Jutras *et al.*, 2016).

Here, we expand the toolbox of protease inhibitors for molecular farming, specifically in agroinfiltrated *N. benthamiana*, through a systematic screen of candidate PIs targeting various classes of *N. benthamiana* proteases. We selected three new PIs that increase levels of three unrelated RPs, separately and in combination. We also investigated suppression of protease activity and identified changes in the total proteome of leaves upon PI overexpression.

4.3 Results

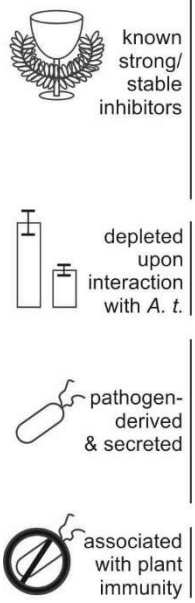
4.3.1 Selecting candidate protease inhibitors

To overcome the degradation bottleneck in molecular farming, we aimed to co-express secreted recombinant proteins (RPs) with secreted protease inhibitors (PIs). We took four approaches to select candidate PIs (Figure 4.1a). First, we mined the literature for ten strong and/or stable inhibitors targeting each class of proteases, prioritizing *N. benthamiana* proteins to simplify expression. We included UbiDomNeg (Ubiquitin-K48R) and native Ubiquitin as a control to also block proteasome-mediated degradation (Chau *et al.*, 1989). Second, we used public microarray data from the NCBI GEO database to select five putative plant PIs whose corresponding transcripts are depleted upon interaction with *A. tumefaciens* (Ditt *et al.*, 2006; Lee *et al.*, 2009; Barrett *et al.*, 2013; Lang *et al.*, 2016). Depleted putative inhibitors were chosen to shift the balance between active and inhibited proteases in agroinfiltrated leaves upon PI overexpression. In our third approach, we selected eight pathogen-derived, secreted inhibitors known to target secreted plant immune proteases. Fourth, we identified four endogenous PIs associated with plant immunity, reasoning that they might both inhibit proteases and enhance plant fitness during interaction with *A. tumefaciens*. This resulted in a collection of 29 PIs, summarized in Figure 4.1b.

The 29 candidate PIs were cloned into a golden-gate compatible a T-DNA vector containing 35S promoter and terminator and an intron to ensure *in planta* expression. All constructs also carry the signal peptide of PR1a from *N. tabacum* (NtPR1SP), replacing the endogenous signal peptide, if present (Figure 4.1c) (Ohshima *et al.*, 1987; Vancanneyt *et al.*, 1990; Engler *et al.*, 2014). To screen for PIs that enhance accumulation of a broad range of RPs, three model RPs were transiently co-expressed with each candidate PI by agroinfiltration (Figure 4.1d). As model RPs, we used an α-

Galactosidase (αGal) used to treat Fabry's disease (Garman & Garboczi, 2002), the angiogenic glycohormone erythropoietin (EPO) (Hayat *et al.*, 2008) and the HIV-neutralizing antibody VRC01 (Wu *et al.*, 2010). During the screen, leaves were sampled at three and six days post infiltration (dpi), or at four and seven dpi when P19 was co-expressed. Each RP-PI combination was tested at least once (number of separate experiments in Figure 4.1e, representative data in Supplementary File ES4.01). PIs that enhanced RP accumulation in initial experiments were tested further, focusing on three dpi without P19 as the most informative condition (black background in Figure 4.1e). For benchmark purposes, the screen included three PIs that had been used previously to increase RP accumulation upon co-expression, namely SlCYS8 (Jutras *et al.*, 2016), SlCDI (Goulet *et al.*, 2012) and the *N. benthamiana* protein most similar to *N. alata* Protease inhibitor II (NbPot2, 87 % identity) (Kim *et al.*, 2008b). However, SlCDI failed to increase RP accumulation in our system, possibly due to the differences in tested RPs (SlCDI enhanced accumulation of murine antibody C5-1) and plant growth conditions. NbPot2 was also ineffective, possibly due to differences in RPs (NbPot2 enhanced accumulation of hGM-CSF), expression systems and PI sequence differences. Remarkably, only SlCYS8 increased RP accumulation (black background in Figure 4.1e). SlCYS8 was thus used in further experiments as a positive control. Besides PIs, we tested the dominant-negative ubiquitin-K48R mutant (Chau *et al.*, 1989), but Ubi-K48R did not change RP accumulation (Figure 4.1e and Supplementary File ES4.01).

a) candidate PI selection strategies



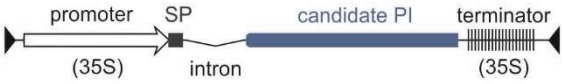
b) 29 candidate PIs

name	genbank ID	family	target proteases	reference
HaSFTI	Q4GWU5	I12	Ser	Luckett et al 1999
HsTIMP	NP_003246	I35	Metallo	Arkadash et al 2017
HvBSZx	Z15116	I4 (Serpin)	Ser/Cys	Fluhr et al 2012
LB_NbCYS1	tbd	I25 (Cystatin)	Cys	this work
NbPot2	ABA42892	I20	Ser	Kim et al 2008
NbSRP_LRA	tbd	I4 (Serpin)	Ser	this work
NbSRP_TMS	tbd	I4 (Serpin)	Ser	this work
NbHvCYS6	tbd	I25 (Cystatin)	Cys	Martinez et al 2009
SlCDI	CAC00536	I3 (Kunitz)	Ser/Asp	Goulet et al 2012
SlCYS8	AAF23128	I25 (Cystatin)	Cys	Goulet et al 2008
UbiDomNeg	tbd		T01	Chau et al 1989
Ubiquitin	tbd		T01	Chau et al 1989
MER411712	tbd	I25 (Cystatin)	Cys	this work
MER411832	tbd	I13	Ser	this work
MER411950	tbd	I3 (Kunitz)	Ser/Cys/Asp	this work
MER412033	tbd	I13	Ser	this work
NbPot1	tbd	I13	Ser	this work
AtMac	WP_010973260	I39	all	this work
AtSerPI	WP_010973636	I51	Ser	this work
CIP1	NP_793972	I42	Ser/Cys/Asp	Shindo et al 2016
EPI1	AAT00500	I1	Ser/Cys	Tian et al 2004
EPI12	XP_002904472	I1	Ser/Cys	this work
EPIC1	EEY55256	I25 (Cystatin)	Cys	Tian et al 2007
EPIC2B	EEY55258	I25 (Cystatin)	Cys	Tian et al 2007
GrVAP1	CAD60978	?	Cys	Lozano-Torres et al 2012
NbK1	tbd	I3 (Kunitz)	Ser/Cys/Asp	this work
NbK2	tbd	I3 (Kunitz)	Ser/Cys/Asp	this work
NbK3	tbd	I3 (Kunitz)	Ser/Cys/Asp	this work
NbPR4	tbd	I43	Cys	Kim et al 2015

e) screening for enhanced RP accumulation

	αGal	EPO	VRC01
HaSFTI	2	1	5
HsTIMP	10	9	21
HvBSZx	4	4	8
LB_NbCYS1	4	4	8
NbPot2	2	2	4
NbSRP_LRA	1	1	3
NbSRP_TMS	1	1	3
NbHvCYS6	0	0	1
SlCDI	2	2	5
SlCYS8	6	7	10
UbiDomNeg	1	1	3
Ubiquitin	1	1	3
MER411712	5	3	8
MER411832	1	0	4
MER411950	4	4	8
MER412033	4	4	8
NbPot1	6	7	11
AtMac	4	4	8
AtSerPI	4	4	8
CIP1	2	1	8
EPI1	2	2	4
EPI12	2	2	4
EPIC1	1	1	3
EPIC2B	2	2	4
GrVAP1	2	2	4
NbK1	1	0	4
NbK2	1	0	4
NbK3	1	0	4
NbPR4	8	8	12

c) candidate PI expression constructs



d) co-expression approach

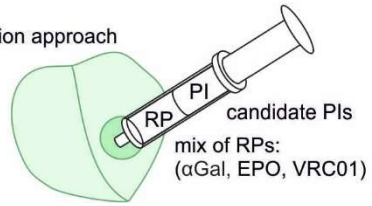


Figure 4.1: Four of 29 tested candidate protease inhibitors (PIs) enhance recombinant protein (RP) accumulation. 29 putative PIs were selected by four different strategies (a, b). PIs were cloned into an expression cassette with a 35S promoter and terminator, NtPR1 signal peptide and PIV2 intron, flanked by T-DNA borders. c) PIs were transiently co-expressed in *N. benthamiana* with three RPs [antibody VRC01, αGalactosidase (αGal), erythropoietin, (EPO)] by agroinfiltration (d). e) Number of times each PI was tested. (Western Blots are shown in Supplementary File ES4.01). Each experiment consisted of two or more biological replicates. Four PIs enhanced RP accumulation (black boxes). Sequences with genbank IDs marked "tbd" (to be determined) will be submitted to genbank upon acceptance of this manuscript.

4.3.2 NbPR4, NbPot1 and HsTIMP enhance accumulation of three model RPs upon co-expression

From the 26 previously untested PIs, we identified three PIs that enhance RP accumulation upon co-expression (black background in Figure 4.1e). Interestingly, these three new PIs are unrelated and presumably target different protease classes.

NbPR4 is the *N. benthamiana* protein most similar to CaPR4c (86.7 % identical amino acids), a novel Cys protease inhibitor associated with defence against *Xanthomonas* in pepper (Kim & Hwang, 2015). NbPot1 (*N. benthamiana* Potato inhibitor type I (I13) inhibitor) was initially selected for co-expression because a transcript corresponding to a similar Arabidopsis protein was depleted upon interaction with *A. tumefaciens.* Indeed, abundance of the *N. benthamiana* transcript corresponding to NbPot1 (Niben101Scf00750XLOC_013210) is reduced 6.7-fold at two dpi (Grosse-Holz *et al.*, 2017). HsTIMP was included in the candidate PI screen to target metalloproteases, for which we did not find plant-derived alternatives. HsTIMP is a well-studied potential anti-cancer inhibitor targeting matrix-metalloproteases (MMPs, M10) (Arkadash *et al.*, 2017). MMPs are conserved in plants (Liu *et al.*, 2001; Kang *et al.*, 2010) and both transcripts and extracellular peptides corresponding to MMPs accumulate in agroinfiltrated leaves of *N. benthamiana* (Grosse-Holz *et al.*, 2017).

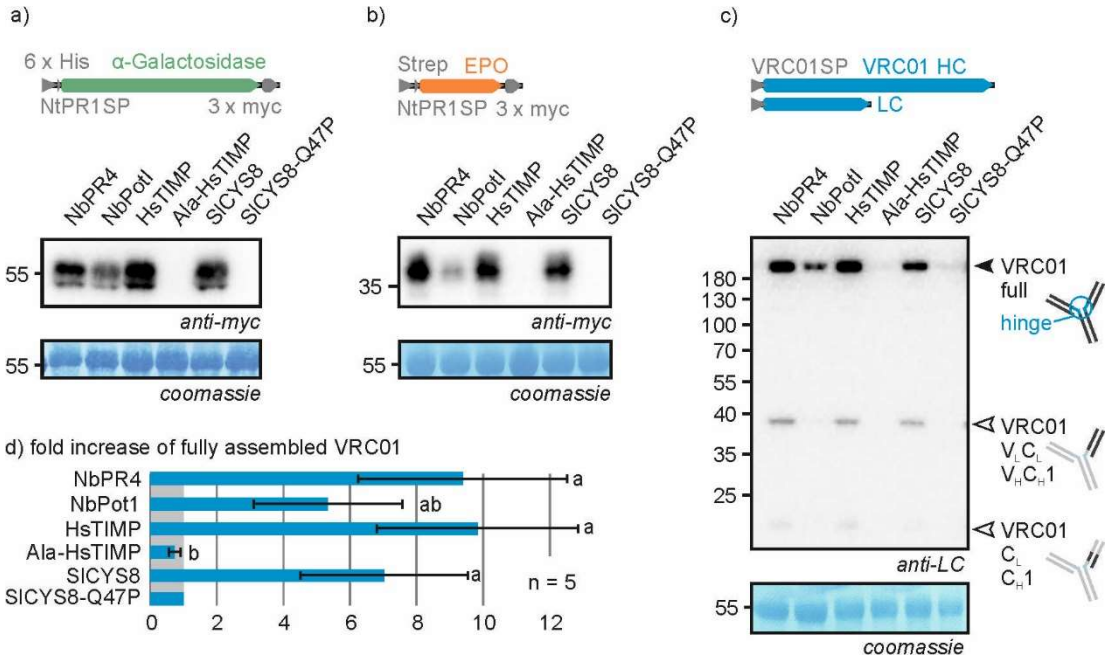


Figure 4.2: Individual co-expression with NbPR4, NbPot1 or HsTIMP enhances accumulation of αGal (a), EPO (b) and VRC01 (c). Leaves were infiltrated with 1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of αGal (a) or EPO (b) and PI or 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI (c). Full leaf extracts were harvested at 3 dpi. Proteins were subjected to reducing (a-b) or non-reducing (c) SDS-PAGE and transferred onto PVDF membranes. RP accumulation was visualized using the indicated antibodies. Closed and open triangles in (c) indicate the full length VRC01 and putative degradation products, respectively (V_L/C_L, variable/constant domain of the light chain, V_H/C_H, variable/constant domain of the heavy chain). The blots are representative of at least five biological replicates (Supplemental File ES4.02). (d) The top band in VRC01 blots was quantified using ImageJ and normalized to the SlCYS8-Q47P control (+/- stdev, n = 5, ANOVA $p<0.0001$ with post-hoc Tukey test, $p<0.05$).

Co-expression with NbPR4, NbPot1, HsTIMP or SlCYS8 enhances accumulation of αGal when compared to the SlCYS8-Q47P negative control (Figure 4.2a and Supplemental File ES4.02). Full-length αGal has a predicted molecular weight (MW) of 50.2 kDa and contains four *N*-glycosylation sites, indicating that the band at ~ 55 kDa represents the complete protein. Co-expression with HsTIMP or NbPR4 consistently enhances αGal accumulation more than co-expression with NbPot1 and SlCYS8.

Interestingly, accumulation of EPO (Figure 4.2b and Supplemental File ES4.02) is enhanced in the same way by PI co-expression. This is remarkable because the PIs are unrelated and the RPs share no similarities. EPO has a predicted MW of 23.4 kDa and contains three *N*-glycosylation sites and one *O*-glycosylation site, indicating that the band at ~ 35 kDa represents the complete protein.

The VRC01 antibody was transiently produced by co-expressing the heavy and light chains (HC and LC) and detected using an anti-κ LC antibody. Fully assembled VRC01 is detected at > 180 kDa, similar to purified VRC01, indicating that PI co-expression enhances accumulation of the fully assembled antibody, consisting of two HCs and two LCs (Figure 4.2c and Supplemental File ES4.02). Besides full-length VRC01, we also detect VRC01 fragments which likely result from cleavages close to the hinge region and after the variable domain of HC and LC. Cleavage at these positions is common for immunoglobulin G (IgG) type mABs, and similar LC fragments were detected for other IgG mABs (Niemer *et al.*, 2014; Hehle *et al.*, 2015; Donini *et al.*, 2015).

Although the levels of all RPs are boosted upon PI co-expression, the strength of this effect differs between RPs. Full-length VRC01 accumulation is 5- to 10-fold enhanced, a change that lies within the dynamic range of Western blots and could thus be quantified directly (ANOVA $p<0.0001$, post-hoc $p<0.05$) (Figure 4.2d). αGal and EPO, however, were not detectable in the control sample without overexposing the bands in PI co-expressing samples. We therefore performed Western blots on dilution series of these samples. These experiments indicate that αGal accumulation is enhanced ~14-, ~8-, ~7- and ~4-fold upon co-expression with HsTIMP, NbPR4, SlCYS8 or NbPot1, respectively (Supplemental Figure 4.S01). Accumulation of EPO is enhanced even further, namely ~27-, ~23-, ~16- and ~18-fold upon co-expression with HsTIMP,

NbPR4, SlCYS8 or NbPot1, respectively (Supplemental Figure 4.S01). For fully assembled VRC01, the values are within the range of those obtained by direct band quantification, namely ~8-, ~10-, ~10- and ~2-fold increased accumulation upon co-expression with HsTIMP, NbPR4, SlCYS8 or NbPot1, respectively (Supplemental Figure 4.S01). In summary, we discovered three PIs that enhance accumulation of three unrelated RPs upon co-expression. NbPR4 and HsTIMP are superior to NbPot1 and SlCYS8 in boosting RP accumulation.

4.3.3 The N-terminus of HsTIMP is required to boost RP accumulation

To elucidate how the three new PIs enhance RP accumulation, we mined the literature for putative inactive mutant PIs. This information is not available for NbPR4 or NbPot1, but the inhibitory function of HsTIMP can be disrupted by appending an Ala residue to the N-terminus of HsTIMP (Wingfield *et al.*, 1999). Indeed, co-expression with Ala-HsTIMP does not enhance VRC01 accumulation when compared to co-infiltration of only *A. tumefaciens*, suggesting inhibitory activity of HsTIMP is required to enhance RP accumulation (Figure 4.2, Figure 4.3a and Supplemental Figure 4.S02).

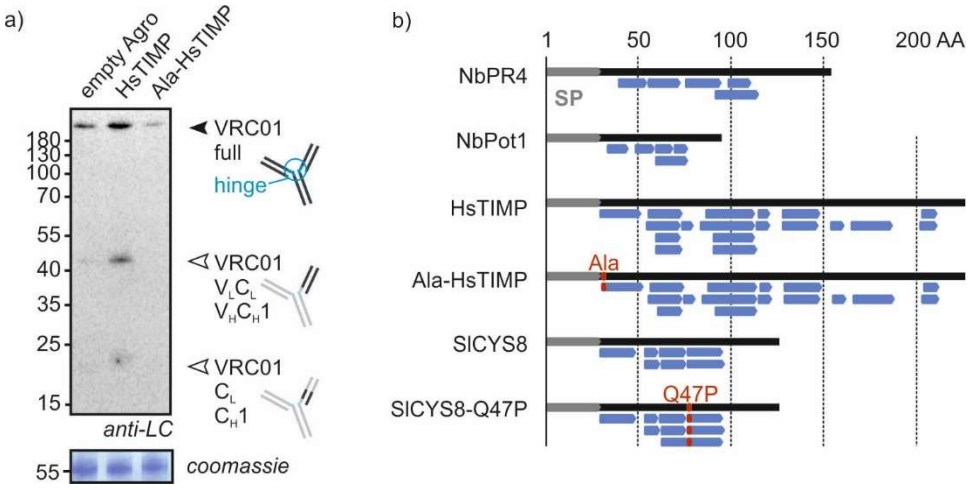


Figure 4.3: a) **Ala-HsTIMP does not enhance VRC01 accumulation**. Leaves were infiltrated with 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI or no plasmid (empty Agro). Full leaf extracts were harvested at 3 dpi. Proteins were subjected to non-reducing SDS-PAGE and transferred onto PVDF membranes. VRC01 accumulation was visualized using an anti-light chain antibody. Closed and open triangles indicate the full length VRC01 and putative degradation products, respectively (V_L/C_L, variable/constant domain of the light chain, V_H/C_H, variable/constant domain of the heavy chain). The blot shows a mixed sample from three biological replicates (Supplemental Figure 4.S02). b) **(mutant) PIs accumulate in *N. benthamiana* leaves upon transient expression.** Inhibitor-derived peptides detected by MS in extracts obtained at 4 dpi from agroinfiltrated leaves expressing (mutant) PIs in the presence of P19. Peptides (blue) are mapped to the inhibitor sequences (black) which carry the NtPR1 signal peptide (SP, grey). For HsTIMP and SlCYS8, mutations are indicated in red. Sequences are given in Supplemental Figure 4.S03.

4.3.4 (mutant) PIs accumulate in *N. benthamiana* leaves upon transient overexpression

To verify that (mutant) PIs accumulate in leaves upon transient overexpression, we performed label-free, quantitative mass spectrometry (MS) on extracts obtained at 4 dpi from agroinfiltrated leaves overexpressing the (mutant) PIs in the presence of P19. Unique peptides derived from HsTIMP and SlCYS8 were identified in the respective extracts from leaves overexpressing the (mutant) PIs. Furthermore, peptides were

identified that distinguished mutant from wild-type PIs (Figure 4.3b and Supplemental Figure 4.S03). Intensities of MS spectra corresponding to endogenous NbPot1-derived peptides were below the detection threshold in all but the NbPot1-overexpressing samples. In contrast, peptides likely derived from endogenous NbPR4 were detected in all samples, but NbPR4 abundance increased on average 69-fold in the overexpressing samples when compared to P19 controls. All overexpressed PIs were among the top 25 % of most abundant proteins in their respective samples (Supplemental File ES4.03).

4.3.5 Unrelated PIs additively enhance RP accumulation

Since NbPR4, NbPot1, HsTIMP and SlCYS8 (PI families I43, I13, I35 and I25, respectively) are predicted to target different protease classes (Cys, Ser, Metallo and Cys proteases, respectively), we tested whether they could act together to enhance RP accumulation. From an initial screen of all binary combinations between the four (Supplemental Figure 4.S04), we found that combinations without NbPot1 increased RP accumulation the most. The triple combination consisting of NbPR4, HsTIMP and SlCYS8 further enhanced accumulation of αGal (Figure 4.4a), EPO (Figure 4.4b) and VRC01 (Figure 4.4c) (replicates in Supplemental File ES4.04). Quantification of full-length VRC01 bands showed significantly higher VRC01 accumulation upon co-expression with the triple PI combination when compared to co-expression with a single PI in combination with a mutant PI (ANOVA $p<0.0001$, post-hoc $p<0.05$) (Figure 4.4d). However, RP levels were similarly increased when threefold more bacteria with PI-encoding T-DNA were infiltrated (Supplemental Figure 4.S05). The additive effects seem to be caused by increased PI levels and not by synergy between PIs, indicating the effect of each PI is dose-dependent.

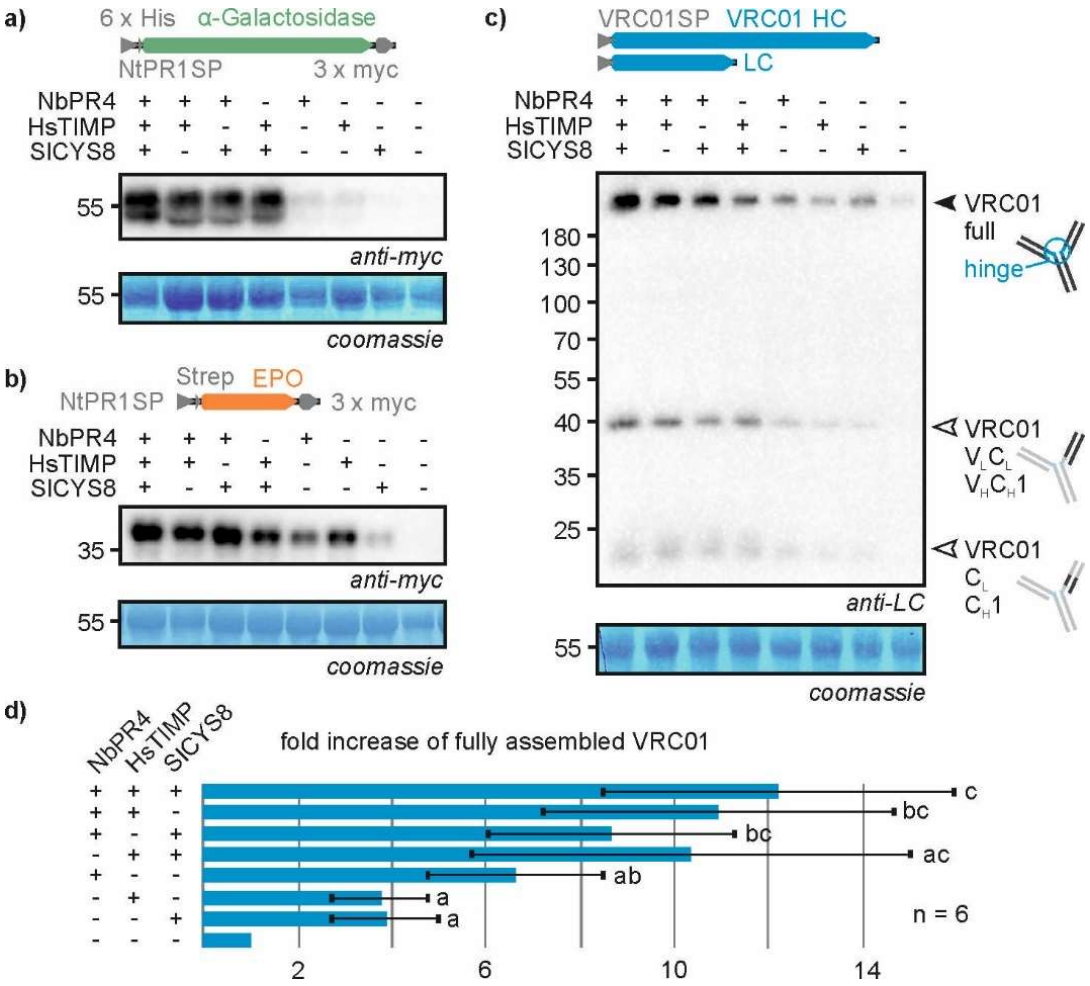


Figure 4.4: Combinations of NbPR4, HsTIMP and SlCYS8 enhance RP accumulation. Leaves were infiltrated with 1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of αGal (a) or EPO (b) and PI or 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI (c). The PI part of the mixture contained three volumes of *A. tumefaciens* strains for expression of the indicated PIs, with one part mutant PI (Ala-HsTIMP) used in lanes 2-4, two parts mutant PI in lanes 5-7 and three parts PI-mutant in lane 8 to replace the missing PIs. Full leaf extracts were harvested at 3 dpi. Proteins were subjected to reducing (a-b) or non-reducing (c) SDS-PAGE and transferred onto PVDF membranes. RP accumulation was visualized using the indicated antibodies. The blots are representative for six biological replicates each (Supplemental File ES4.04). The top band in all VRC01 blots was quantified using ImageJ and normalized to the control (PI-mutant only); average relative intensities are shown in underneath the blot (+/- stdev, n = 6, ANOVA $p<0.0001$ with post-hoc Tukey test, $p<0.05$).

4.3.6 NbPR4, NbPot1 and HsTIMP differ from SlCYS8 in their effect on leaf protease activity profiles

To understand how NbPR4, NbPot1 and HsTIMP enhance RP accumulation, we investigated how these PIs affect activity of *N. benthamiana* proteases in agroinfiltrated leaves. We thus characterized extracts harvested at 4 dpi from leaves transiently overexpressing each (mutant) PI. We assessed activity of specific protease families using Activity-based Protein Profiling (ABPP), which is based on the labelling of proteomes with fluorescent chemical probes that react covalently with the active site of enzymes in an activity-dependent manner (Morimoto & van der Hoorn, 2016). We surveyed active Papain-like Cys proteases (PLCPs, family C01), Ser Hydrolases (SHs) and legumains/Vacuolar Processing Enzymes (VPEs, family C13) using the activity-based probes MV201 (Richau *et al.*, 2012), FP-TAMRA (Kaschani *et al.*, 2009) and JOPD1 (Lu *et al.*, 2015), respectively. We have recently shown that overexpression of SlCYS8, but not the SlCYS8-Q47P mutant PI, affects PLCP activity (Jutras et al, manuscript in preparation, see Chapter 5). In contrast, overexpression of NbPR4, NbPot1, HsTIMP or Ala-HsTIMP does not affect PLCP activity when compared to the control (Figure 4.5a and Supplemental File ES4.05). This is remarkable because it was expected that NbPR4 inhibits Cys proteases, like CaPR4c in pepper (Kim & Hwang, 2015). Activity profiles of both SHs and VPEs are unaffected by PI overexpression, even though NbPot1 is annotated as a Ser protease inhibitor (PI familiy I13). To test specifically whether extracellular proteases were affected by overexpression of the PIs, we also surveyed extracellular PLCPs, SHs and VPEs using ABPP in apoplastic fluid isolated from PI overexpressing leaves. Also there, SlCYS8 overexpression affects activity of extracellular PLCPs when compared to SlCYS8-Q47P overexpression or to the control, as shown previously (Jutras et al,

manuscript in preparation, see Chapter 5). NbPot1, HsTIMP and Ala-HsTIMP do not affect extracellular PLCP activity. Also, extracellular SH and VPE activity profiles remain unchanged upon overexpression of all PIs (Supplemental Figure 4.S06).

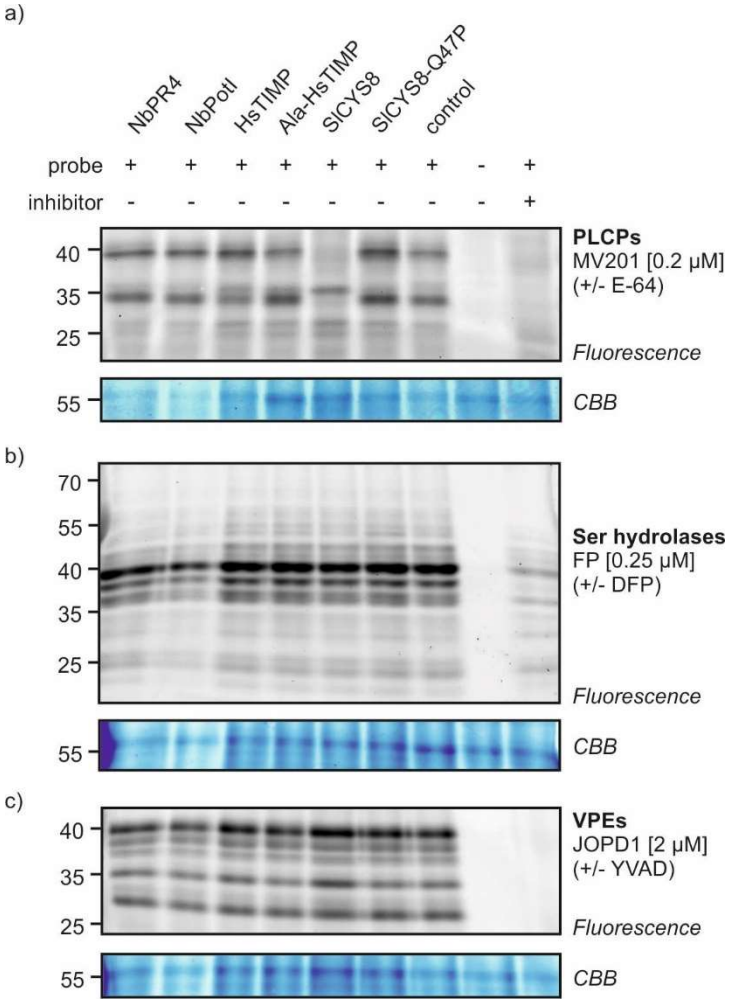


Figure 4.5: Activity-based protein profiling indicates that NbPR4, NbPot1 and HsTIMP differ from SlCYS8 in their mode of action. Activity profiles of Papain-Like Cys Proteases (PLCPs, a), Ser hydrolases (SHs, b) and Vacuolar Processing Enzymes (VPEs, c). Leaves were infiltrated with *A. tumefaciens* harbouring the indicated PI expression plasmid, mixed 1/1 (v/v) with *A. tumefaciens* harbouring the P19 expression plasmid. Leaf extracts (pH 5) were obtained at 4 dpi, adjusted to the same protein concentration and 48 µl of each sample were pre-incubated with or without 0.2 mM of inhibitor (E-64, DFP or YVAD) for 30 min and then incubated with or without the indicated probe for 4 h (MV201, JOPD1) or 1 h (FP) at room temperature. Labelled proteins were visualized by in-gel fluorescence scanning. Three additional biological replicates are shown in Supplemental File ES4.05.

4.3.7 NbPR4, NbPot1 and HsTIMP similarly affect leaf proteomes

We hypothesized that PI overexpression might affect the accumulation of endogenous proteins. For instance, inhibited proteases might be degraded, protease substrates might accumulate or PIs might be sensed by the plant and trigger a response. We thus also analysed endogenous differentials in the proteomes from leaves overexpressing the (mutant) PIs, exploring the same dataset used for Figure 4.3 (Supplemental File ES4.03).

Besides the overexpressed (mutant) PIs and P19, we quantified peptides corresponding to 3349 endogenous *N. benthamiana* and *A. tumefaciens* proteins. Of these, 3236 (97 %) remained unchanged upon overexpression of any (mutant) PI compared to the P19 control, indicating that the (mutant) PIs have rather subtle effects on the leaf proteome. However, 167 (3.4 %) proteins changed significantly (Student's t-test, $p<0.05$) and more than two-fold in abundance when comparing leaves overexpressing each (mutant) PI separately to the control (P19 only) (Supplemental File ES4.03). Euclidean distance clustering of only these endogenous differential proteins reveals two subclusters: The samples from leaves overexpressing the PIs form subcluster A and the controls form subcluster B, consisting of Ala-HsTIMP, SlCYS8 Q47P and P19 samples (Figure 4.6a). Interestingly, the samples from leaves overexpressing HsTIMP, NbPot1 or NbPR4 cluster together and separately from SlCYS8 overexpressing leaf samples within subcluster A. The effects of both SlCYS8 and HsTIMP on the endogenous proteome thus depend on protease inhibitory activity, and the effect of SlCYS8 is distinct from the effects of HsTIMP, NbPot1 and NbPR4, which have very similar effects on the leaf proteome.

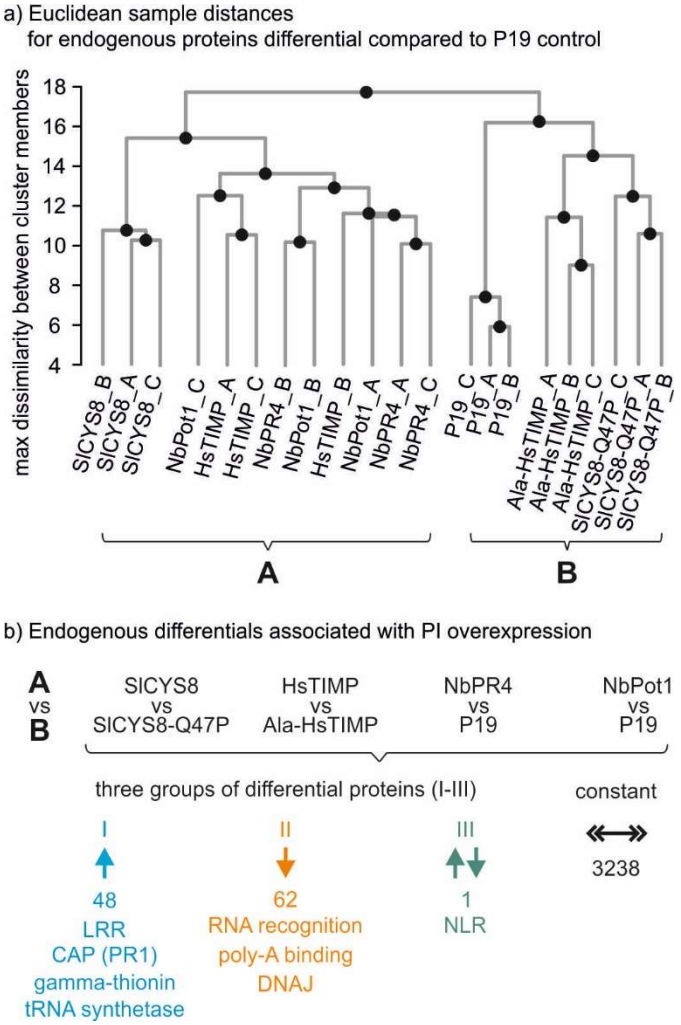


Figure 4.6: NbPR4, NbPot1 and HsTIMP similarly affect leaf proteomes. Label-free quantitative MS was performed on leaf extracts (4 dpi) from leaves overexpressing the (mutant) PIs in the presence of P19. a) Complete linkage clustering was performed using Euclidean distances between samples, based on all endogenous differential proteins. Endogenous differential proteins are all proteins that differ significantly (t-test, $p<0.05$) and >2-fold in abundance between PI overexpressing and P19 expressing control leaves, but not P19 and (mutant) PIs. b) Differentials associated with PI overexpression are endogenous proteins that differ significantly (t-test, $p<0.05$) and >2-fold when comparing NbPR4 and NbPot1 expressing leaves to the P19 control and HsTIMP and SlCYS8 expressing leaves to the respective mutant PI control. Differentials were grouped according to whether they increased for at least one PI and never decreased (group I), decreased for at least one PI and never increased (group II) or increased for some, but decreased for other PIs (group III). 3238 endogenous proteins never changed significantly and more than 2-fold in abundance (constant, black). PFAM domains named underneath groups I-II are enriched (geometric test, Benjamini-Hochberg adjusted $p<0.05$) among the proteins in the respective group, representative PFAMs are shown. The full MS dataset is in Supplemental File ES4.03.

We analysed the endogenous differentials that distinguish the effect of SlCYS8 from the effect of other PIs, but their annotations are associated only with general signalling processes and thus rather inconclusive (Supplemental File ES4.03). More targeted approaches such as activity-based proteomics or identification of direct interactors may be needed to resolve how the effect of SlCYS8 differs from the effect of the other PIs.

To elucidate what distinguishes the PI-overexpressing leaf samples in subcluster A from the controls in subcluster B, we determined which proteins changed significantly (Student's t-test, $p<0.05$) and more than two-fold in abundance when comparing NbPR4 and NbPot1 overexpressing leaves to the P19 control and SlCYS8 and HsTIMP overexpressing leaves to their respective mutant PI controls (Figure 4.6b). We divided the differential proteins into three groups: increased in abundance upon overexpression of at least one PI compared to the respective control and not decreased for another PI (n = 48 proteins, group I), decreased upon overexpression of at least one PI compared to the respective control and not increased for another PI (n = 62 proteins, group II) and increased for one/some PI(s), but decreased for other(s) (n = 1 protein, group III). Importantly, the PIs have similar effects on the leaf proteome, as only one protein was differential in opposing directions (group III). A nucleotide-binding, leucine-rich repeat receptor (NLR, Niben101Scf02118g00018) decreased in abundance upon HsTIMP overexpression when compared to the Ala-HsTIMP, but increased in abundance for all other PI/control comparisons.

We analysed the PFAM family annotation of the proteins in groups I and II as an indicator for protein function. PFAM families associated with transcription and translation are enriched among both increasing and decreasing proteins (hypergeometric test, Benjamin-Hochberg adjusted $p<0.01$) (Figure 4.6b and

Supplemental File ES4.03). This includes tRNA synthetases (2 proteins) among the proteins increasing in abundance and RNA recognition (3 proteins), poly-A binding (2 proteins) and chaperone DNAJ (1 protein) families among the proteins decreasing in abundance upon PI overexpression. Most of these proteins have several homologues within *N. benthamiana*, which might replace each other. For instance, we detect 59 proteins carrying an RNA recognition motif (PF00076) domain in leaves, five of which are differential in either direction for at least one PI vs control comparison.

Interestingly, domains associated with plant defence are enriched among the proteins increasing in abundance upon PI overexpression. This includes Leucine-Rich-Repeat (LRR) domains that are part of plant immune receptors (3 proteins), cysteine-rich secretory protein (PR 1, one protein) and gamma thionin (one protein). Gamma thionins are small, stable plant defense proteins, many of which have antifungal properties or act as protease inhibitors (Pelegrini & Franco, 2005).

We also identified peptides corresponding to 68 proteins from Agrobacterium, four of which changed in abundance (Supplemental File ES4.03). Further studies will be needed to confirm differential accumulation of the endogenous proteins we detected.

4.4 Discussion

We tested 29 candidate PIs for enhancing accumulation of three RPs upon co-expression by agroinfiltration. Of the 29 tested candidate PIs, only four enhance RP accumulation upon co-expression. The remaining 25 can be regarded negative controls for overexpression of small secreted proteins that do not affect accumulation of αGal, EPO or VRC01. Interestingly, we tested six cystatins (family I25), but only SlCYS8 enhances RP accumulation. Similarly, we tested three PIs of family I13, but only NbPot1 enhances RP accumulation. Related proteins that do not enhance RP accumulation can be used in domain swap or mutagenesis experiments to determine essential sequence and/or structural motifs. In combination with ABPP, this approach could reveal specifically which proteases or protease subfamilies need to be inhibited to enhance RP accumulation.

4.4.1 Unrelated protease inhibitors affect unrelated recombinant proteins in similar ways

The model RPs used in our screen were the lysosomal enzyme α-Galactosidase (αGal), the small glycohormone erythropoietin (EPO) and VRC01, an immunoglobulin G (IgG) type antibody. We chose completely unrelated model RPs (<23% identity between any two), reasoning they were likely to differ in sensitivity to proteases and to be degraded by different proteolytic pathways. Surprisingly, the pattern in which PIs affect RP accumulation is the same for αGal, EPO and VRC01. HsTIMP and NbPR4 enhance accumulation more than NbPot1, which has a similar effect to SlCYS8. Our data indicate that the degradation processes of unrelated RPs must have something in common that can be affected by PIs to produce the same outcome for each RP.

Astonishingly, four unrelated PIs (<13% identical amino acids between any two) enhance accumulation of αGal, EPO and VRC01, although the PIs likely target different proteases. HsTIMP presumably inhibits matrix-metalloproteases and NbPot1 belongs to the I13 family, which targets Ser proteases (Wingfield et al., 1999; Rawlings, 2016). SlCYS8 inhibits PLCPs of the XBCP3, XCP, RD21 and RD19 subfamilies (Jutras et al, manuscript in preparation, see Chapter 5) and NbPR4 likely targets Cys proteases as well, as it is 86.7% identical to the pepper Cys protease inhibitor CaPR4c (Kim & Hwang, 2015). However, activity of the PLCPs that are inhibited by SlCYS8 is not affected by NbPR4, indicating NbPR4 either targets different Cys proteases or acts in a different way.

4.4.2 Three possible proteolytic mechanisms

Different proteolytic mechanisms of RP degradation can explain how three unrelated RPs are affected similarly by three unrelated PIs.

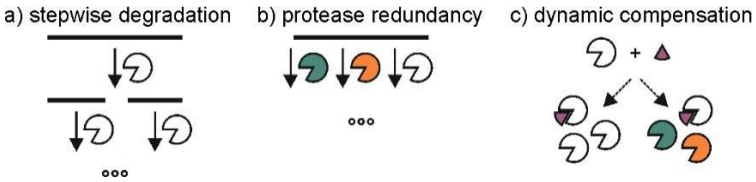


Figure 4.7: RP degradation mechanisms.

The first proteolytic mechanism is that degradation happens stepwise (Figure 4.7a). An initial cleavage exposes parts of the protein that are then degraded further. This may be the case for VRC01, where we detect fragments indicating that an initial cleavage happens in the hinge region, followed by downstream degradation of the fragments. None of the PIs cause a shift in the relative abundances of the VRC01 fragments, indicating that the initial cleavage is not prevented. In accordance with these results, N-terminal sequencing of antibody fragments revealed that SlCYS8 co-

expression does not prevent initial cleavage of the human IgG H10 (Jutras *et al.*, 2016). We do not detect fragments of αGal and EPO, indicating that fragments of αGal and EPO do either not accumulate or lack the myc tag epitope used for detection.

The second proteolytic mechanism is that different proteases act redundantly in RP degradation (Figure 4.7b). PIs can thus act on different proteases to decrease the pool of total proteolytic activity. This may be the case, as accumulation of the same RP can be enhanced by different PIs. In this scenario, one would expect different PIs to have additive effects when combined and indeed, we detected increased RP accumulation when co-expressing two or three PIs compared to one PI and a mutant PI. However, we also observed that the effect of PIs on RP levels is dose-dependent. RPs accumulate more when 1/3 of the agroinfiltration mixture delivers PI-encoding T-DNA than when only 1/9 of the mixture delivers PI-encoding T-DNA.

At the concentrations we used, we expect Agrobacteria to transform virtually all cells in the infiltrated zone (Castilho *et al.*, 2014; Buyel *et al.*, 2015a). Expression under a 35S promoter should then produce PIs in every cell. Two mechanisms can explain why we still observe a concentration-dependent effect of PIs on RP levels. First, genes on extrachromosomal T-DNA are expressed *in planta* (Mysore *et al.*, 2000; Singer *et al.*, 2012). Delivery of more T-DNAs may therefore increase PI expression further even if all cells in the infiltrated area are transformed already. Second, the transcription/translation machinery in agroinfiltrated leaves may be running at full capacity and co-expression of additional proteins may therefore divert resources away from RP and PI expression.

The third proteolytic mechanism is that protease activity is dynamic, so that inhibition can be compensated (Figure 4.7c). Plants may sense the lack of protein turnover upon

protease inhibition and upregulate proteases to ensure cellular debris and protease-inhibitor complexes are appropriately recycled. Increased protease transcription upon inhibition is seen for instance for proteasome subunits, which show increased transcript levels in the presence of the inhibitor syringolin A (Michel et al., 2006). Proteasome subunit protein abundance remains constant, indicating that transcriptional upregulation compensates for degradation of inhibited subunits (Svozil et al., 2014). We found that protease abundance does not increase in proteomes of PI overexpressing leaves, indicating there may be increased turnover of inhibited proteases. To test whether protease transcription increases upon PI overexpression, transcriptomes of PI overexpressing leaves could be analysed. Protease activity may also be compensated by post-translational activation of a non-target protease, which then compensates for the inhibited PI targets. The proteases we monitored here (PLCPs, SHs and VPEs) do not show increased activity, but the compensatory proteases could also be aspartic and metalloproteases or other protease activities that are not displayed under the tested conditions. As a broader approach, N-terminal proteomics (e.g. TAILS, COFRADIC) could reveal neo-N-termini of activated proteases, whereas gel-based proteomics (PROTOMAP) could be used to detect shifts in the size of proteases upon release of inhibitory prodomains (Dix et al., 2008; Huesgen & Overall, 2012).

4.4.3 Protease inhibitor overexpression may trigger plant immunity

The three new PIs that enhance RP accumulation have small and remarkably similar effects on the leaf proteome when compared to mutant PIs and controls. PFAM families enriched among the proteins that change in abundance upon PI overexpression are associated with increased abundance of immune signalling

components. Protease inhibition thus seems to not only prevent RP degradation, but also trigger plant immunity. Immune responses triggered by perception of Agrobacterium cold shock protein limit T-DNA delivery in flowering *N. benthamiana* (Saur *et al.*, 2016), highlighting the impact of plant immunity on the performance of the agroinfiltration expression platform. Investigating how the interaction between *N. benthamiana* and *A. tumefaciens* is altered by PI overexpression and whether this also affects T-DNA delivery or RP expression rates will be an interesting topic for further studies.

4.4.4 PIs are tools to unravel the proteolytic network degrading RPs

As RP degradation in agroinfiltrated leaves is at least partially independent of RP sequence and structure, unravelling the underlying proteolytic network will be highly valuable. PIs can be instrumental to understand degradation or processing when proteases act redundantly. For instance, the peptide hormone IDA (Inflorescence Deficient in Abscission) is processed by several subtilases that act redundantly (Schardon et al., 2016). A role for subtilases was found upon overexpression of the Kazal-like Ser protease inhibitor EPI10 from *Phytophthora infestans*, which targets subtilases and prevents abscission in Arabidopsis. Abscission can be restored in EPI10 overexpressing plants by supplying mature IDA peptide and indeed, multiple target subtilases of EPI10 can process IDA *in vitro* (Schardon et al., 2016). Analogously, target proteases of NbPot1, NbPR4 and HsTIMP are implicated in RP degradation. We recently identified nine PLCPs targeted by SlCYS8 using activity-based proteomics (see Chapter 5). For PIs that do not deplete activities that we can detect in ABPP, immunoprecipitation of PIs from leaf extracts may reveal co-purifying target proteases. We have performed preliminary experiments with N- and C-terminal

tags, but these disrupted the RP accumulation enhancing function of the PIs. NbPR4, NbPot1 and HsTIMP thus likely require intact termini for inhibitory activity. Specific antibodies raised against the plant-produced PIs or immobilisation of purified PIs may facilitate future immunoprecipitation experiments.

4.4.5 Conclusions

Our data indicate that RPs are degraded by a dynamic protease network, where different proteases act redundantly and/or compensatory. NbPR4, NbPot1 and HsTIMP can be used to unravel the *N. benthamiana* protease network. Importantly, PI co-expression substantially increases accumulation of three unrelated RPs, suggesting NbPR4, NbPot1 and HsTIMP may increase the levels of a broad variety of RPs in industrial molecular farming. We thus discovered novel tools to improve the general productivity of the agroinfiltration expression platform.

4.5 Material and Methods

All chemicals and oligonucleotides were obtained from Sigma (Sigma-Aldrich, St. Louis, US) unless specified otherwise.

Cloning of inhibitor and RP expression plasmids: The MoClo plant parts kit (Engler *et al.*, 2014) was used for cloning and all vectors are from this kit unless specified otherwise. Primer sequences and plasmid names are in Supplemental Tables 4.S01-2. The sequence encoding NbPR4 was amplified from *N. benthamiana* genomic DNA using primers #001 (including the native SP) or #005 (without the native SP) and #003 and cloned into pL0V-SC-41308 (including the native SP) or pL0V-C-41264 (without the native SP), respectively. The sequence encoding HsTIMP was codon-optimized for *in planta* expression (Supplemental Table 4.S03), synthesized as a GeneString (Thermo Fisher Inc, Waltham, US) and cloned into pL0V-C-41264 (without the native SP). The sequence encoding NbPot1 was amplified from *N. benthamiana* genomic DNA using primers #074 and #075 and cloned into pL0V-C-41264. Other inhibitors were cloned analogously; oligonucleotides designated as “s” (sense) or “as” (antisense), respectively, were annealed according to the manufacturer’s instructions and added to the golden gate reaction; where primers are designated “p1”, “p2” etc, inhibitors were cloned in multiple parts to domesticate the sequences. Each inhibitor level 0 module was then combined with pL0M-PU-35S-TMV-3-51288, pL0M-S-NtPR1a SP (X06361) PIV2 (pJK002, which was a gift from Jiorgos Kourelis, PR1 signal peptide from *Nicotiana tabaccum*), pL0M-T-35S-1-41414 and pL1VB-F-pAGM4723 in a BsaI reaction to obtain the expression plasmids pFGH008 (NbPR4), pFGH053 (NbPot1) and pFGH047 (HsTIMP). All other PIs were cloned analogously, with the exceptions of SFTI1, which retained its native signal peptide as this may be required for correct processing and folding into the rigid structure that defines this PI

(Luckett *et al.*, 1999), and NbK1-3, which were cloned into pL0V-C1-15457 (without SP, for N- and C-terminal tagging) and combined pFGH074 (generated from pL0M-C2-FLAG-15198 by adding TACCCATACGATGTTCCAGATTACGCT) to clone them with a C-terminal FLAG-HA tag. The sequences of VRC01 heavy chain (HC) and light chain (LC), codon-optimized for plant expression, were amplified from plasmids kindly provided by Julian Ma (St. George's, University of London) (Teh *et al.*, 2014) and cloned into pL0V-SC-41308 (including the native SP). Each level 0 module was then combined with pL0M-PU-35S-TMV-3-51288, pL0M-T-35S-1-41414 and pL1VB-F-pAGM4723 in a BsaI reaction to obtain the expression plasmids pLM15 (VRC01 LC) and pLM16 (VRC01 HC). The sequences encoding EPO and αGal were codon-optimized for *in planta* expression (Supplemental Table 4.S03), synthesized as GeneStrings (Thermo Fisher Inc, Waltham, US) and cloned into pL0V-C1-15457. The αGal and EPO level 0 modules were combined with pL0M-P-2x35S-45089, pLM07 (pL0M-U-TMV/NtPR1a, generated from pL0M-U-TMV-1-41402 by insertion of the NtPR1SP), pL0M-S-6xHis-15258 for αGal or pLM09 (pL0M-S-StrepII, generated by inserting TGGTCACATCCTCAATTTGAAAAG into pL0V-S-41258) for EPO, pL0M-C2-3xMyc-15212, pL0M-T-35S-1-41414 and pL1VB-F-pAGM4723 in a BsaI reaction to obtain the expression plasmids pLM25 (EPO) and pLM34 (αGal).

Plasmids were transformed into *E. coli* for amplification, purified, sequenced and transformed into *Agrobacterium* GV3101-pMP90. *Agrobacterium* GV3101-pMP90 were cultured on plates of LB medium (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract, 15 g/L agar) containing 25 µM Rifampicin, 50 µM Gentamycin and 50 µM Kanamycin to select for transformants. A single colony was picked and cultured in liquid LB medium (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract) containing 25 µM Rifampicin, 50 µM Gentamycin and 50 µM Kanamycin. Glycerol stocks were prepared

by mixing this culture 1/1 (v/v) with 50 % glycerol in water, flash-freezing in liquid nitrogen and storing at -80 °C. For each agroinfiltration experiment, a 160 µL aliquot of glycerol stock was thawed and inoculated into fresh LB.

Agroinfiltration procedure: *N. benthamiana* plants were grown at 21 °C under a 16/8 h light/dark regime in a growth room. *Agrobacterium* GV3101-pMP90 (WT), *Agrobacterium* GV3101-pMP90 carrying a P19-encoding plasmid (pJK050, which was a gift from Jiorgos Kourelis (Addgene plasmid # 101751), P19 from tomato bushy stunt virus) or *Agrobacterium* GV3101-pMP90 carrying EPO, αGal, VRC01 or inhibitor encoding plasmids were grown for 21h at 28 °C with agitation in LB containing 25 µM Rifampicin and 50 µM Gentamycin (for WT) plus 50 µM Kanamycin (for *A. tumefaciens* harbouring plasmids). Bacteria were collected by centrifugation at 2000 g for 5 min at room temperature (RT), resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfone (MES), 10 mM $MgCl_2$, pH 5.7, 100 µM acetosyringone) to OD_{600} = 0.5 (OD_{600} = 1 for αGal) and left for 2 h at 28 °C with agitation to recover. For co-expression experiments, *A. tumefaciens* suspensions were mixed in the appropriate ratios (described in Figure 4.legends). The first and second fully expanded leaves of pre-flowering *N. benthamiana* (4-5 weeks old) were infiltrated with the bacteria suspension using a syringe without a needle. For comparison, different *A. tumefaciens* suspension mixes were infiltrated into different sectors of the same leaf.

Protein extraction and detection by Western Blot: Leaf extracts were prepared at three days post infiltration (dpi), unless specified otherwise. Four leaf discs (22 mg each) from four individual plants were combined per sample, flash-frozen in liquid Nitrogen and pulverized in a TissueLyser ball mill (Qiagen, Hilden, DE). The tissue powder was mixed with 3/1 (v/fresh weight) cold Phosphate Buffered Saline (PBS, 10 mM PO_4^{3-}, 137 mM NaCl, 2.7 mM KCl, pH 7.4) and centrifuged for 10 min at 16.000

g and 4 °C. The supernatant was mixed with 4x gel loading buffer (200 mM Tris-HCl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol), heated for 5 min at 95 °C and separated on Bis-Tris gels at 100 V. Proteins were then transferred to a PVDF membrane using the TransBlot Turbo system (Biorad, Hercules, US). The membrane was blocked in 5 % milk in TBS (50 mM Tris-Cl, pH 7.6; 150 mM NaCl) for 1 h at room temperature (RT), incubated in 1/5000 anti-myc-HRP (ab1326, Abcam, Cambridge, UK) for detection of αGal and EPO or in 1/2000 anti-kappa-HRP (Sigma A7164) or 1/2000 anti-gamma-HRP (Sigma A6029) for detection of VRC01 over night at 4 °C and washed in TBST (0.005 % Tween-20) prior to detection with Clarity ECL substrate (BioRad).

Protein extraction for ABPP/mass spectrometry: Leaf extracts were prepared at four days post infiltration (dpi), to allow for maximum protein accumulation in the presence of the p19 silencing suppressor. Infiltrated tissue from three or four individual plants was combined per sample, weighed, flash-frozen in liquid Nitrogen and pulverized using pestle and mortar. The tissue powder was mixed with 3/1 (v/fresh weight) cold 500 mM NaAc, pH 5, 500 mM DTT and centrifuged for 45 min at 3500 g and 4 °C. Protein concentrations were determined using a linearized Bradford assay (Ernst & Zor, 2010).

Apoplastic fluid (AF) extraction: Six *N. benthamiana* leaves per sample were detached and vacuum-infiltrated with ice-cold water, dried on the surface and placed in a syringe without needle and plunger that was inserted in a 50 mL falcon tube. AF was collected by centrifugation at 2000 g, 4°C for 25 min and used immediately.

ABPP: Leaf extracts made in or apoplastic fluid adjusted to 500 mM NaAc, pH 5, 5 mM DTT were used for labelling. Samples of 48 µl were pre-incubated with or

without 0.2 mM of inhibitor (E-64, DFP or YVAD) for 30 min and then incubated with or without the indicated probe for 4 h (MV201, JOPD1) or 1 h (FP) at room temperature. FP was obtained from Thermo (88318), MV201 and JOPD1 were synthesized as described (Richau *et al.*, 2012; Lu *et al.*, 2015). ABPP reactions were terminated by adding 1 mL cold acetone. Samples were centrifuged for 3 min at 16000 g, the supernatant discarded and the proteins resuspended in 2 x gel loading buffer (100 mM Tris-HCl (pH 6.8), 200 mM DTT, 4% SDS, 0.02% bromophenol blue, 25% glycerol), heated for 5 min at 95 °C and separated on Bis-Tris gels at 100 V. Fluorescence scanning was performed on a Typhoon scanner (Amersham/GE Healthcare, Little Chalfont, UK), using Cy3 settings.

Mass spectrometry: see Supplemental File 4.ES06.

Bioinformatics tools for leaf proteome analysis: Peptide spectra were annotated using Andromeda (Cox *et al.*, 2011). Included modifications were carbamidomethylation (static) and oxidation, N-terminal acetylation and carbamylation of Lysines and N-termini (dynamic). Protein quantification was performed using MaxQuant version 1.5.5.30 (Tyanova *et al.*, 2016a), including all modifications. Filtering and imputation of missing values using default settings were performed in Perseus (Tyanova *et al.*, 2016b) and further data analysis carried out in R using the data.table and ggplot packages (Wickham, 2009; Dowle *et al.*, 2014; R Core Team, 2015). Western blots were quantified using imageJ (Schneider *et al.*, 2012), ANOVA and post-hoc tests were carried out in R using the Rcommander package (Fox, 2005).

Other databases and tools: We selected PIs whose corresponding transcripts were depleted upon interaction with *A. tumefaciens* using the NCBI GEO database (Barrett *et al.*, 2013), specifically the accessions GSE4116 (Ditt *et al.*, 2006), GSE14106 (Lee

et al., 2009), GSE48402 (Lang *et al.*, 2016) and GSE62751, all of which represent Arabidopsis microarrays. For each depleted transcript, we selected the most similar *N. benthamiana* transcript (Bombarely *et al.*, 2012) using BLAST (Altschul *et al.*, 1990) and identified PIs using the *N. benthamiana* PI sequences in the MEROPS database (Rawlings, 2016). Sequence analyses and plasmid design were performed in Geneious (Kearse *et al.*, 2012).

4.6 Acknowledgments

We thank Urszula Pyzio, Caroline O' Brien and Sarah Rodgers for excellent technical support, Philippe V. Jutras for scientific discussion and Jiorgos Kourelis for the P19 and NtPR1SP plasmids. This work was financially supported by the ERC Project 'GreenProteases', University of Oxford (R.H., grant No. 616449), by Somerville College, Oxford (F.G.H. and R.H.), an ERC starting grant (M.K., grant No. 258413) and the Deutsche Forschungsgemeinschaft (M.K., grant no. INST 20876/127-1 FUGG). M.A.Z. was supported by an Erasmus grant, M.S was supported by the BBSRC's Research Experience Placements scheme and M.F. was supported by the David Kirby Memorial Fund.

4.7 Supplementary information

Supplemental figures

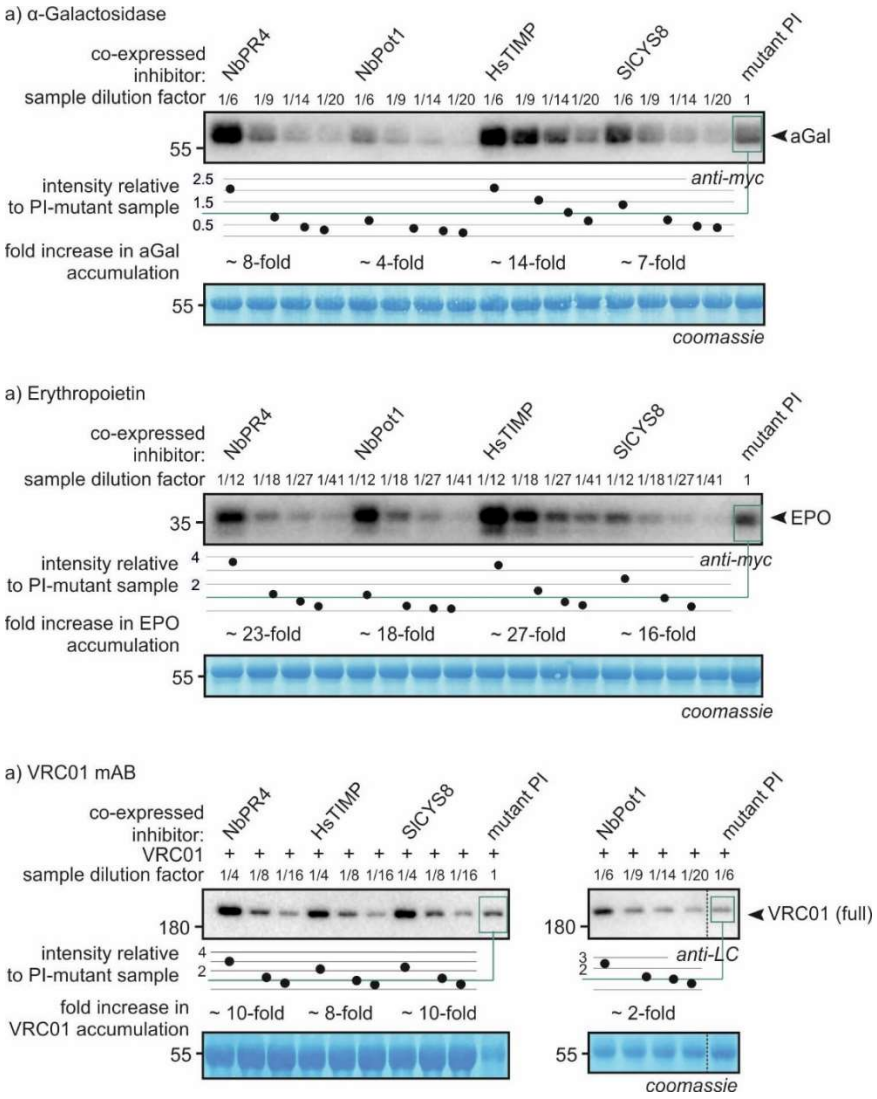


Figure 4.S01: dilution series to quantify the increase in RP accumulation upon PI co-expression. Leaves were infiltrated with 1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of αGal (a) or EPO (b) and PI or 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI (c). Full leaf extracts were harvested at 3 dpi and diluted in leaf extract from non-infiltrated leaves. Proteins were subjected to reducing (a-b) or non-reducing (c) SDS-PAGE and transferred onto PVDF membranes. αGal (a) or EPO (b) and VRC01 (c) accumulation was visualized using the indicated antibodies. Bands were quantified using ImageJ and normalized to the mutant PI (Ala-HsTIMP) co-expressing control; relative intensities are shown underneath the blots. The numbers for approximate fold increase in RP accumulation are obtained from the data point closest to the control in raw intensity.

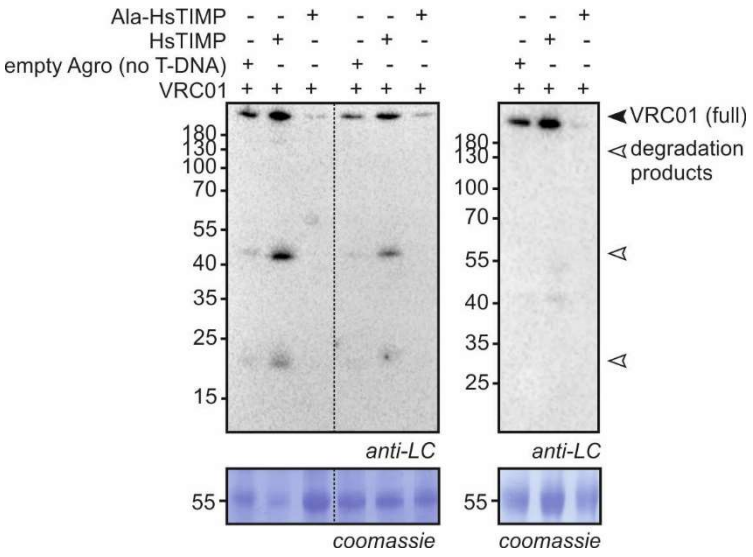


Figure 4.S02: The mutant PI Ala-HsTIMP does not enhance VRC01 accumulation. Leaves were infiltrated with 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI. Full leaf extracts were harvested at 3 dpi. Proteins were subjected to non-reducing SDS-PAGE and transferred onto PVDF membranes. VRC01 accumulation was visualized using an anti-light chain antibody. The blots show three biological replicates.

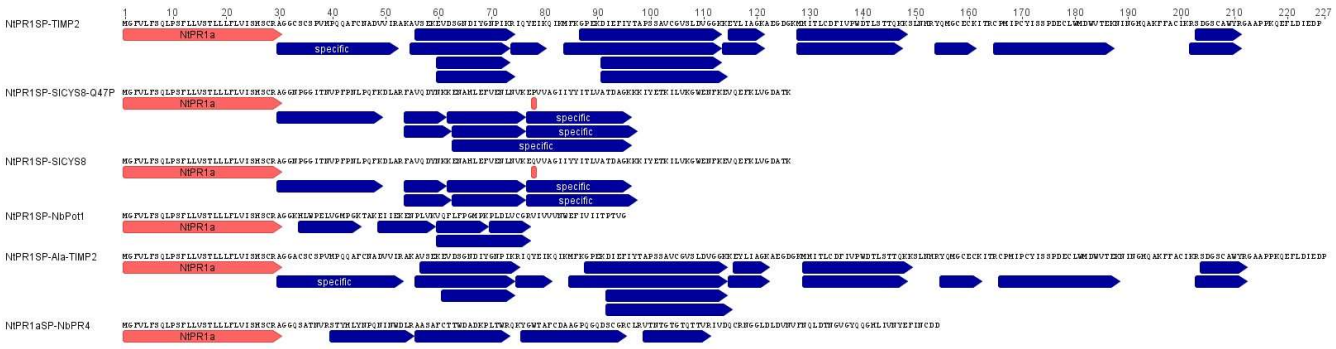


Figure 4.S03: peptides corresponding to PIs and PI mutants were identified by mass spectrometry. Inhibitor-derived peptides (blue) detected by MS in extracts (4 dpi) from agroinfiltrated leaves expressing (mutant) PIs in the presence of P19. Peptides are mapped to the inhibitor sequences which carry the NtPR1 signal peptide (red). For HsTIMP and SlCYS8, peptides that are specific to the intact or mutant inhibitor, respectively, are labelled “specific”.

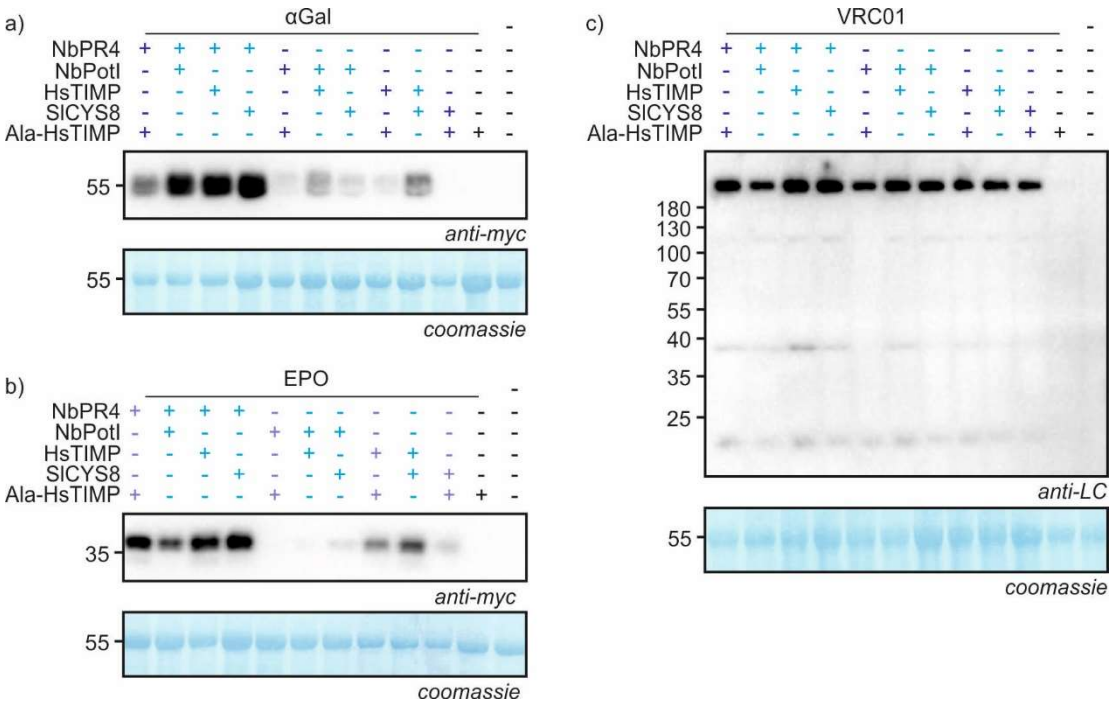


Figure 4.S04: Screen of all binary combinations between NbPR4, NbPot1, HsTIMP and SlCYS8. Leaves were infiltrated with 1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of αGal (a) or EPO (b) and PI or 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI (c). The PI part of the mixture contained two volumes of *A. tumefaciens* strains for expression of the indicated PIs, with one part Ala-HsTIMP used in lanes 1, 5, 8 and 10 (purple) and two parts Ala-HsTIMP in lane 11 to replace the missing PIs. Full leaf extracts were harvested at 3 dpi. Proteins were subjected to reducing (a-b) or non-reducing (c) SDS-PAGE and transferred onto PVDF membranes. αGal (a) or EPO (b) and VRC01 (c) accumulation was visualized using the indicated antibodies.

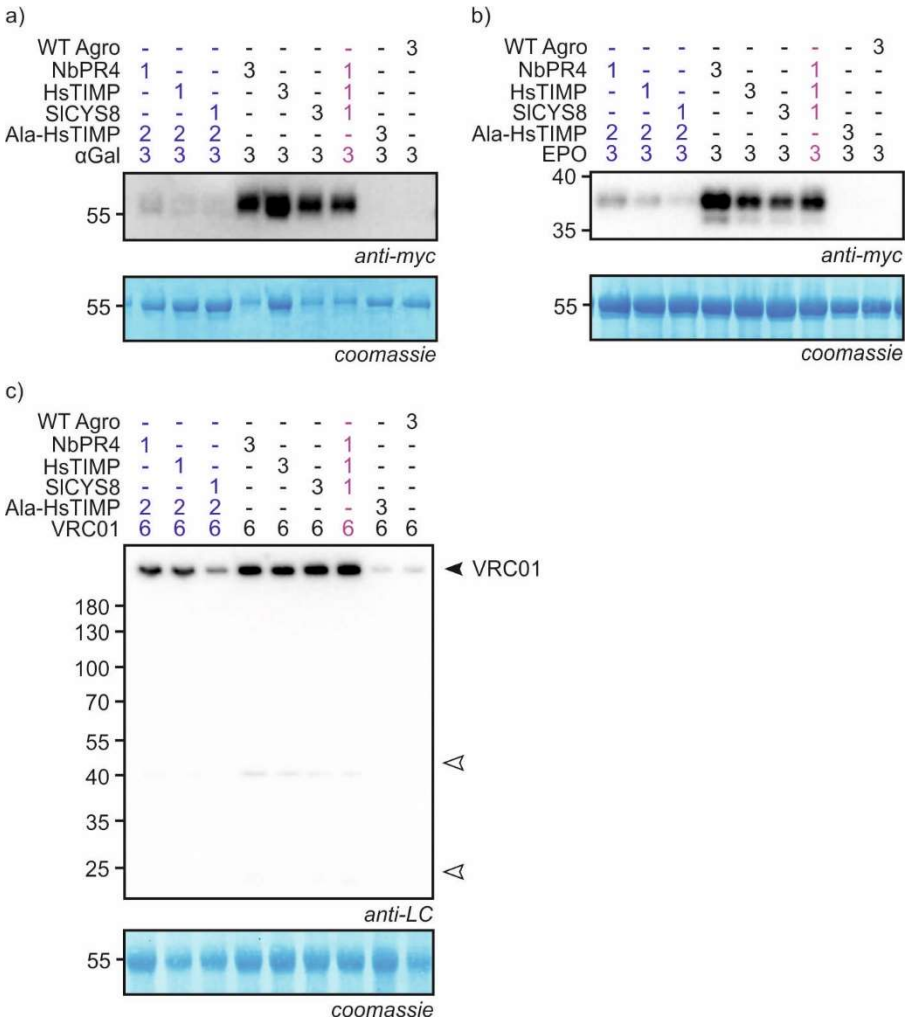


Figure 4.S05: The effect of NbPR4, HsTIMP and SlCYS8 on RP accumulation is dose dependent. Leaves were infiltrated with 1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of αGal (a) or EPO (b) and PI or 1/1/1 (v/v) mixes of *A. tumefaciens* strains carrying plasmids for expression of VRC01 heavy chain, VRC01 light chain and PI (c). The PI part of the mixture contained three volumes of *A. tumefaciens* strains for expression of the indicated PIs, with two parts Ala-HsTIMP used in lanes 1-3 (purple) and three parts Ala-HsTIMP in lane 8 to replace the missing PIs. Full leaf extracts were harvested at 3 dpi. Proteins were subjected to reducing (a-b) or non-reducing (c) SDS-PAGE and transferred onto PVDF membranes. αGal (a) or EPO (b) and VRC01 (c) accumulation was visualized using the indicated antibodies.

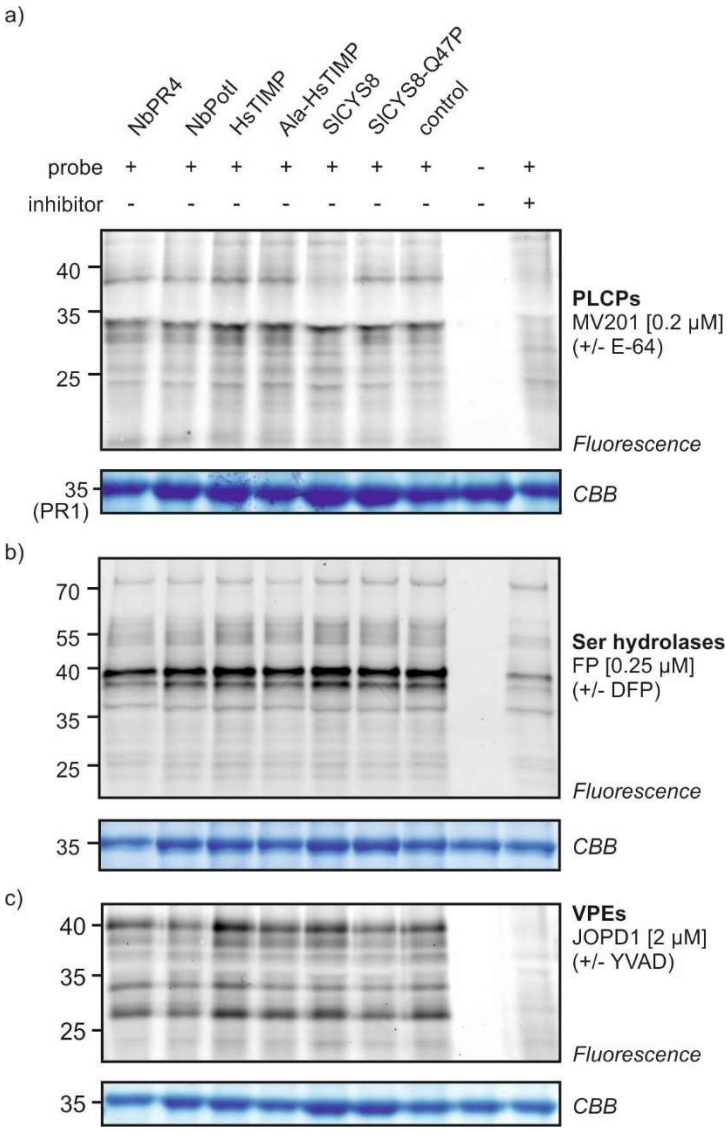


Figure 4.S06: Activity-based profiling of extracellular proteases. Activity profiles of Papain-Like Cys Proteases (PLCPs, a), Ser hydrolases (SHs, b) and Vacuolar Processing Enzymes (VPEs, c). Leaves were infiltrated with *A. tumefaciens* harbouring the indicated protease PI expression plasmid, mixed 1/1 (v/v) with *A. tumefaciens* harbouring the P19 expression plasmid. Apoplastic fluids were obtained at 4 dpi, adjusted to pH 5 (500 mM NaAc, 5 mM DTT) and 196 ul (a) or 48 µl (b, c) of each sample were pre-incubated with or without 0.2 mM of inhibitor (E-64, DFP or YVAD) for 30 min and then incubated with or without the indicated probe for 4 h (MV201, JOPD1) or 1 h (FP) at room temperature. Labelled proteins were visualized by in-gel fluorescence scanning.

Table 4.S01: Plasmids

Golden Gate modules (for cloning)	
internal lab number	**plasmid name** (C: contains CDS to be combined with NtPR1 signal peptide, SC: contains CDS to be used without extra signal peptide, C1: contains CDS to be used with C-terminal tag)
pFGH002	pL0M-SC-NbPR4_nativeSP
pFGH003	PL0M-C-NbPR4_noSP
pFGH005	pL0M-C-SlCDI
pFGH011	pL0M-C-BSZx
pFGH015	pL0M-C-NbPotII
pFGH016	pL0M-SC-Nb_ubi_native
pFGH018	pL0M-SC-Nb_ubi_K48R
pFGH019	pL0M-SC-SFTI1
pFGH021	pL0M-C-HsTIMP
pFGH022	pL0M-C-EPI1
pFGH023	pL0M-C-EPI12
pFGH025	pL0M-C-EPIC1.4
pFGH029	pL-1M-AtMac_A
pFGH030	pL-1M-AtMac_B
pFGH031	pL-1M-AtMac_C
pFGH032	pL-1M-AtMac_D
pFGH036	Pl0M-C-MER411712
pFGH038	pL0M-C-MER411950
pFGH039	pL0M-C-MER412033
pFGH040	pL0M-C-NbPot1
pFGH041	pL0M-C-SlCYS8
pFGH043	pL0M-C-AtSerPI
pFGH044	pL0M-C-AtMac
pFGH063	pL0M-C-LB_NbCYS1
pFGH065A	pL0M-C-MER411832_cDNA
pFGH067	pL0M-C-NbHvCYS6_gDNA
pFGH073	pL0M-SC-NbHvCYS6_gDNA

pFGH151	pL0M-C1-Niben101Scf04078g00002-NbK1
pFGH152	pL0M-C1-Niben101Scf06424XLOC_064533-NbK2
pFGH153	pL0M-C1-Niben101Scf06424XLOC_064534-NbK3
pLM007	pL0M-U-TMV/NtPR1a
pLM009	pL0M-S-StrepII
pLM003	pL0M-SC-nVRC01 LC
pLM004	pL0M-SC-nVRC01 HC
pLM018	pL0M-C1-hEPO
pLM028	pL0M-C1-αGal
pFGH047	pL0M-C2-flagHA

Binary vectors (for in planta expression by agroinfiltration)

internal lab number	**plasmid name** (all with 35S promoter, 35S terminator and NtPR1SP unless specified otherwise)
pFGH007	pL1MB-F-NbPR4_nativeSP
pFGH008	pL1MB-F-NbPR4
pFGH010	pL1MB-F-SlCDI
pFGH045	pL1MB-F-BSZx
pFGH046	pL1MB-F-NbPotII
pFGH047	pL1MB-F-HsTIMP
pFGH048	pL1MB-F-EPI1
pFGH049	pL1MB-F-EPI12
pFGH050	pL1MB-F-MER411712
pFGH051	pL1MB-F-MER411950
pFGH052	pL1MB-F-MER412033
pFGH053	pL1MB-F-NbPot1
pFGH054	pL1MB-F-SlCYS8
pFGH056	pL1MB-F-AtSerPI
pFGH057	pL1MB-F-AtMac
pFGH058	pL1MB-F-Nb_ubi_native

pFGH059	pL1MB-F-Nb_ubi_K48R
pFGH060	pL1MB-F-SFTI1
pFGH109	pL1MB-F-LB_NbCYS1
pFGH110	pL1MB-F-MER412218
pFGH111	pL1MB-F-MER411832
pFGH114	pL1MB-F-NbHvCYS6_nativeSP
pFGH156	pL1MB-F-NtPR1-NbK1-flagHA
pFGH157	pL1MB-F-NtPR1-NbK2-flagHA
pFGH158	pL1MB-F-NtPR1-NbK3-flagHA
pFGH203	pL1MB-F-Ala-HsTIMP
pFGH214	pL1MB-F-SlCYS8-Q47P
pLM025	pL1MB-F-Strep-EPO-myc
pLM034	pL1MB-F-His-αGal-myc
pLM015	pL1MB-F-VRC01-LC-nativeSP
pLM016	pL1MB-F-VRC01-HC-nativeSP

Table 4.S02: Primers

Name	**Sequence** (BpiI site; ***BsaI site***)
#001#NbPR4_natSP_FW	TTGAAGACAAAATGGAGAGAGTAAATAATTACTATAAG
#003#NbPR4_REV	TTGAAGACAAAAGCTTAGTCATCGCAGTTGATAAATTCATAG
#005#NbPR4_noSP_FW	TTGAAGACAAAGGTCAGAGCGCTACAAACGTGAG
#074#NbPot1_REV	TTGAAGACAAAAGCTTAACCCACTGTGGGAGTAATTATAACG
#075#NbPot1_FW	TTGAAGACAAAGGTAAGCATTTATGGCCTGAACTTGTGG
#010#SlCDI_CDS_nosP FW	TTGAAGACAAAGGTTCAAGTTTCACTTCCCAAAATCC
#011#SlCDI_CDS_REV	TTGAAGACAAAAGCTCAGACTTTCTTGAAGTAGACC
#012#SlCYS8_8d_FW	TTGAAGACAAAGGTAATCCTGGGGGCATTACCAATGTTCCA
#015#SlCYS8_8d_REV	TTGAAGACAAAAGCTCACTTAGTGGCATCACCAACAAGC
#019#TIMP2 optimized_no signalP_FW	TTGAAGACAAAGGTTGTTCTTGCTCTCCTGTTCATCC

#020#TIMP2 optimized_no signalP_REV	TTGAAGACAAAAGCCTAAGGATCCTCGATATC
#023#Nb-Pot2 - FW	TTGAAGACAAAGGTCCAAAGCCATGTCCTCGGAATTGTG
#024#Nb-Pot2 - REV	TTGAAGACAAAAGCTTATCCTTCACAAACAAAAGTTCCATCATCAC
#025#Ubi_part1 _FW	TTGAAGACAAAATGCAGATTTTTGTCAAGACTTTG
#026#UbiK48R_ part1_REV	TTGAAGACAAGACGGCCAGCAAAGATCAGCC
#027#UbiK48R_ part2_FW	TTGAAGACAACGTCAGTTGGAAGATGGTCG
#028#Ubi_part2 _REV	TTGAAGACAAAAGCTTAGAAACCACCACGTAGACGG
#029#Cip1_REV	TTGAAGACAAAAGCTCAGTTCACCGTGATTGC
#030#Cip1_FW	TTGAAGACAAAGGTCAAACGCCCAAGAACATCG
#037#EPIC1 CDS_FW	TTGAAGACAAAGGTCAAGTGGACGGCGGATACTCGAAGAAGG
#038#EPIC1 CDS_REV	TTGAAGACAAAAGCCTACTTAACTGGGGTAATCGACGTCACC
#041#A.t_macro _cl1_REV	***TTGGTCTCA***ACAAAGGAACCATTGCGTGGCGC
#042#A.t_macro _cl1_FW	***TTGGTCTCA***CCGTGGAAGAGCCACTGGAAAAGC
#043#A.t_macro _cl2_REV	***TTGGTCTCA***ACAACATCGCCTGCTTCGCGCCATCG
#044#A.t_macro _cl2_FW	***TTGGTCTCA***CTGCCATCAGCGAAAAGAGTTTCCTCG
#045#A.t_macro _cl3_REV	***TTGGTCTCA***ACAAAAGCTCATGGGGTTGCCGAACG
#046#A.t_macro _cl3_FW	***TTGGTCTCA***ACATTCGCGGTCAACGTCACG
#047#EPI10_F W	TTGAAGACAAAGGTGATGATAATTGCTCTTTCGG
#048#EPI10_RE V	TTGAAGACAAAAGCCTACAGCTTCTGCTGTTGC
#054#SlCYS8_n p1_REV	TTGAAGACAATTCTCTTTCTTATTATAATCTTGAACAGC
#055#SlCYS8_n p2_FW	TTGAAGACAAAGAATGCTCATTTGGAG
#056#AtSerPI_p 1_FW	TTGAAGACAAAGGTGACTCCACGCTGTCCGTGAAAATCG
#057#AtSerPI_p 1_REV	TTGAAGACAAGGCACGAGTCCCTCGCTTTCTGC
#058#AtSerPI_p 2_FW	TTGAAGACAATGCCGGGCGGTAAGCCGGTCG
#059#AtSerPI_p 2_REV	TTGAAGACAACCAGTGAACGGCCCGGAAAGG
#060#AtSerPI_p 3_FW	TTGAAGACAACTGGCGAGGACGTGAAAAAATCC
#061#AtSerPI_p 3_REV	TTGAAGACAAAAGCCTAGTTCGGCGTGGCGG

#064#MER411712_FW	TTGAAGACAAAGGTAGAAAAGTTGGGGGAAGAACTCC
#064B#MER411950_ REV	TTGAAGACAAAAGCTTATCCATCAGTCTTTTGGAATTTCCTAAAGAC
#065#MER411712_REV	TTGAAGACAAAAGCCTAAACTTCTGCACAATTATCCAATATTTTTCG
#065#MER411950_p1_s	AGGTTCATCTTTCACTTCCACCAATCCCATTGTCCTTCCCACCACTAC
#065B#MER411950_FW	TTGAAGACAAAGGTTCATCTTTCACTTCCACCAATCC
#066#MER411950_p1_as	ATCAGTAGTGGTGGGAAGGACAATGGGATTGGTGGAAGTGAAAGATGA
#067#MER411950_p2_REV	TTGAAGACAAAAGCTTATCCATCAGTCTTTTGGAATTTCCTAAAGAC
#068#MER411950_p2_FW	TTGAAGACAATGATGATGACAAGGGACTCCCTATCC
#070#MER412218_FW	TTGAAGACAAAGGTATGGTAACTTGCACCCCTGATACTCC
#071#MER412218_REV	TTGAAGACAAAAGCTTAGAAAGCAGATTCATGCATGATCATGC
#072#MER412218_p2_FW	TTGAAGACAACTAGTGTTCTGTCCCTATGGAGAGAATTTTACTTGC
#073#MER412218_p1_REV	TTGAAGACAACTAGCTTATACGTTGAGCCACCAAGTGAC
#074#MER412288_REV	TTGAAGACAAAAGCTTAACCCACTGTGGGAGTAATTATAACG
#075#MER412288_FW	TTGAAGACAAAGGTAAGCATTTATGGCCTGAACTTGTGG
#076#MER412033_REV	TTGAAGACAAAAGCTCAGCGAACAATAGGTACGAGAG
#077#MER412033_FW	TTGAAGACAAAGGTAGTTCTCCGTGTACGGTTCAGG
#078#NbOsCys_FW	TTGAAGACAAAGGTATGAGCAGTGATGGTGG
#079#NbOsCys_REV	TTGAAGACAAAAGCCTAAGCGTTAGCAGAAGC
#081#MER412218_np1_REV	TTGAAGACAACAGAACACTAGCTTATACGTTGAGCCACC
#083#412218_np2_s	TTGAAGACAATCTGTCCCTATGGAGAGAATTTTACTTGCCAAAATGTTGGCAGTGCCGAGGAAAATAGATATAATCGTTTGGTTCTCGCAGATTGTCTTCAA
#084#412218_np2_as	TTGAAGACAATCTGCGAGAACCAAACGATTATATCTATTTTCCTCGGCACTGCCAACATTTTGGCAAGTAAAATTCTCTCCATAGGGACAGATTGTCTTCAA
#085#412218_np3_s	TTGAAGACAACAGAGAATGCAAAGGCCTTTGTGTTCATAAAAGCGTGGTGGAATTGGAAAGGCCGAAGCATGATCATGCATGAATCTGCTTTCTAAGCTTTTGTCTTCAA
#086#412218_np3_as	TTGAAGACAAAAGCTTAGAAAGCAGATTCATGCATGATCATGCTTCGGCCTTTCCAATTCCACCACGCTTTTATGAACACAAAGGCCTTTGCATTCTCTGTTGTCTTCAA
#087#LeosNbCys1_CSC_REV	TTGAAGACAAAAGCTTAGGAGTGGTCAGGCTCCATATGGTTCAG
#088#LeosNbCys1_SC_FW	TTGAAGACAAAATGAGAGTATCTCGAAACGCCACACTGC

#088b#LBNbCys1_C_ FW	TTGAAGACAAAGGTTTAAGCGAAACCGGAGGAGGATTTTGC
#089#412206_p1_REV	TTGAAGACAAACAGGCGGGCCAAGATCGTTTACG
#090#MER412206_p1_SC_FW	TTGAAGACAAAGGTGCAAGGAATATAGAGCCCCTAGTAGTAGGGAG
#091#412206 p2_s	TTGAAGACAACTGTCGCCGCCGTGTTCTTCAATGCACAACGAGAGACAGCCGCCAGGAGGCGCTAGGCTTTTGTCTTCAA
#092#412206 p2_as	TTGAAGACAAAAGCCTAGCGCCTCCTGGCGGCTGTCTCTCGTTGTGCATTGAAGAACACGGCGGCGACAGTTGTCTTCAA
#112#411832_cDNA_C_FW	TTGAAGACAAAGGTCAATCCAGTTGCCCAGGAGTG
#113#NbHvCYS6_SC_FW	TTGAAGACAAAATGGCTCTCAAATTTAATTCC
#114#NbHvCYS6_C_FW	TTGAAGACAAAGGTACCGTTCTCTTCCATGTC
#115#NbHvCYS6_CSC_REV	TTGAAGACAAAAGCTTAAGCAAAAAGACAATTTTGTCC
#121#SFTI1_opt_SC1_FW	TTGAAGACAAAATGGCTACCACCATGGCTAAGC
#122#SFTI1_opt_SC1_K5toA_REV	TTGAAGACAACACCAGGCCTACCATCAGGGAAGCAGATAGGAGGGATAGACGCGGTACACCTACCAT
#123#SFTI1_opt_SC1_REV	TTGAAGACAACACCAGGCCTACCATCAGGGAAGC
#148#-K1p1F	TTGAAGACAAAGGTGTACCCAATCCCTCAAGG
#149#-K1p1R	TTGAAGACAAAGACCCACTTGTACTTTGTCG
#150#-K1p2F	TTGAAGACAAGTCTGAACTATTTCGTTCTACC
#151#-K1p2R	TTGAAGACAACACCGAAAGTTTTCTTGAACATAACC
#152#-K2F	TTGAAGACAAAGGTCAAGATGTTCCTGAACCGGTGC
#153#-K2R	TTGAAGACAACACCGATCTTGTGAAATGTAACTTTCAAAGG
#154#-K3p1F	TTGAAGACAAAGGTGAACCAGTTCTTGATACTAATAAAC
#155#-K3p1R	TTGAAGACAATGTCCTCATTTTCGTCAATTACTAGC
#156#-K3p2F	TTGAAGACAAGACATAAATATAAAATTTGCAGCACC
#157#-K3p2R	TTGAAGACAACACCAACCTTCTTGAACACAATCTTG
NbSRP-LRA-FW	TTGAAGACAAAGGTATGGACCTTCAAGAATCAATCAGC
NbSRP-LRA-REV	TTGAAGACAACACCGTCTACTAGAGGATTTTGCACGC
NbSRP-TMS-p1-FW	TTGAAGACAAAGGTATGGATCTCAGGGAGTCAATCTAC
NbSRP-TMS-p1-REV	TTGAAGACAATGTTTTCGTTTTAGCCCACTG
NbSRP-TMS-p2-FW	TTGAAGACAAAACAAATGATCTCATCGAAG
NbSRP-TMS-p2-REV	TTGAAGACAACACCAGCTAGAGGATTCATCACACTG

Table 4.S03: Protein coding sequences codon-optimized for *in planta* expression

HsTIMP (in pFGH047, with **native SP**, which was omitted for pFGH047)

ATGGGTGCTGCTGCTAGGACTCTTAGGCTTGCTCTTGGTCTGCTTCTTCTGGCTACT
CTTCTTAGGCCTGCTGATGCTTGTTCTTGCTCTCCTGTTCATCCTCAGCAGGCTTTCT
GCAATGCTGATGTGGTGATTAGGGCTAAGGCTGTGAGCGAGAAAGAAGTGGATAGC
GGTAACGATATCTACGGTAACCCTATCAAGAGGATCCAGTACGAGATCAAGCAGATC
AAGATGTTCAAGGGTCCTGAGAAGGATATCGAGTTTATCTACACCGCTCCTAGCTCT
GCTGTTTGCGGTGTTTCTCTTGATGTGGGTGGTAAGAAAGAGTACCTGATCGCTGGT
AAGGCTGAGGGTGATGGTAAGATGCACATTACCCTGTGCGATTTCATCGTGCCTTGG
GATACCCTTTCAACCACTCAGAAGAAGTCCCTGAACCACAGGTATCAGATGGGTTGC
GAGTGCAAGATTACCAGGTGCCCTATGATCCCTTGCTACATCTCTTCACCTGATGAGT
GCCTGTGGATGGATTGGGTTACCGAGAAGAACATCAACGGTCACCAGGCTAAGTTCT
TCGCTTGCATCAAGAGGTCCGATGGTTCTTGCGCTTGGTATAGAGGTGCTGCTCCTC
CTAAGCAAGAGTTCCTTGATATCGAGGATCCTTAG

αGal (in pLM028)

CTGGATAACGGTCTTGCTAGGACTCCTACTATGGGTTGGCTTCACTGGGAGAGATTC
ATGTGCAACCTGGATTGCCAAGAGGAACCTGATAGCTGCATCAGCGAGAAGCTGTTC
ATGGAAATGGCTGAGCTGATGGTGTCTGAGGGTTGGAAGGATGCTGGTTACGAGTA
CCTGTGCATCGATGATTGCTGGATGGCTCCTCAGAGAGATTCTGAGGGTAGACTTCA
AGCTGATCCTCAGAGGTTCCCTCACGGTATTAGGCAGCTTGCTAACTACGTGCACAG
CAAGGGTCTGAAGCTTGGTATCTACGCTGATGTGGGTAACAAGACCTGCGCTGGTTT
TCCTGGTAGCTTCGGTTACTACGATATCGATGCTCAGACCTTCGCTGATTGGGGTGT
GGATCTTCTGAAGTTCGATGGTTGCTACTGCGATAGCCTTGAGAACCTGGCTGATGG
TTACAAGCACATGTCTCTGGCTCTTAACAGGACCGGTAGATCCATCGTTTACTCTTGT
GAGTGGCCTCTGTACATGTGGCCTTTCCAGAAGCCTAACTACACCGAGATCAGGCAG
TATTGCAACCATTGGAGGAACTTCGCAGATATTGATGATAGCTGGAAGTCCATCAAGT
CTATCCTGGATTGGACCAGCTTCAATCAAGAAAGGATCGTGGATGTGGCTGGTCCTG
GTGGTTGGAATGATCCTGATATGCTGGTGATCGGTAACTTCGGTCTGAGCTGGAATC
AGCAGGTTACCCAAATGGCTCTGTGGGCTATTATGGCTGCTCCTCTGTTCATGAGCA
ACGATCTGAGGCACATTAGCCCTCAGGCTAAGGCTTTGCTGCAGGATAAGGATGTGA
TCGCTATCAACCAGGATCCTCTGGGTAAGCAGGGTTATCAGCTTAGGCAGGGTGATA
ACTTCGAGGTTTGGGAGAGGCCTTTGTCTGGTCTTGCTTGGGCTGTGGCTATGATCA
ACAGGCAAGAAATTGGTGGTCCTAGGTCCTACACCATTGCTGTGGCTTCTCTTGGTA
AGGGTGTGGCTTGTAATCCTGCTTGCTTTATCACCCAGCTGCTGCCTGTGAAGAGAA
AGCTTGGTTTTTACGAGTGGACCAGCAGGCTGAGGTCACACATTAACCCTACTGGAA
CCGTGCTTCTGCAGCTTGAGAATACCATGCAGATGAGCCTGAAGGATCTGCTT

EPO (in pLM025)

CCTAGGCTGATCTGCGATTCTAGGGTGTTGGAGAGATACCTGCTTGAGGCTAAAGAG
GCTGAGAACATTACTACCGGTTGCGCTGAGCACTGCTCTCTGAACGAGAATATTACC
GTGCCTGATACCAAGGTGAACTTCTACGCTTGGAAGAGGATGGAAGTTGGTCAGCAG
GCTGTTGAAGTTTGGCAGGGTCTTGCTCTTTTGTCTGAGGCTGTTCTTAGGGGTCAG
GCTCTGCTTGTGAATTCTTCTCAACCTTGGGAGCCTCTTCAGCTGCATGTTGATAAGG
CTGTGAGCGGTCTTAGATCTCTTACCACCCTTCTTAGGGCTCTGAGGGCTCAGAAAG
AGGCTATTTCTCCTCCTGATGCTGCTTCTGCTGCTCCTCTTAGGACTATTACCGCTGA
TACCTTTAGGAAGCTGTTCAGGGTTTACAGCAACTTCCTGAGGGGTAAGCTGAAGCT
TTACACTGGTGAGGCTTGCAGGACTGGTGATAG

Electronic supplementary information

Electronic supplementary information included with the thesis: Files ES4.01-06.

In Chapter 4, we improved the productivity of agroinfiltrated *N. benthamiana* by co-expressing recombinant proteins (RPs) with the protease inhibitors (PIs) NbPR4, NbPot1, HsTIMP and SlCYS8. Remarkably, the effects of the PIs are consistent between three unrelated RPs. This transferability makes any information on how RP accumulation is enhanced highly valuable, as it could contribute to improving production of many different RPs in *N. benthamiana*.

Interestingly, single amino acid mutations in HsTIMP and SlCYS8 abolish the RP accumulation-enhancing function. The mutant PIs are known to be non-inhibitory, indicating that protease inhibition is required for enhancing RP accumulation in the cases of HsTIMP and SlCYS8. These two PIs thus likely inhibit proteases that would usually degrade RPs. Identification of their targets is an important step towards understanding the proteolytic network that hampers RP accumulation and purity.

SlCYS8 belongs to the cystatin (I25) family of Cys protease inhibitors and activity-based protein profiling (ABPP) of Cys proteases is well established (Richau *et al.*, 2012; Lu *et al.*, 2015). Furthermore, SlCYS8 has been used previously in plant science, resulting in the availability of useful tools such as a non-inhibitory mutant and a protocol for production of recombinant SlCYS8 in *E. coli* (Goulet *et al.*, 2008; Sainsbury *et al.*, 2013). The work presented in Chapter 5 is thus a joint project with Philippe Jutras, who had previously used SlCYS8 to reduce RP degradation in plants (Jutras *et al.*, 2016). We developed the project together and Philippe carried out most of the ABPP experiments during his visit from March-May 2017. I performed additional ABPP experiments afterwards and importantly, I designed, performed and analysed the activity-based proteomics experiment in which we identified the SlCYS8 targets. Chapter 5 is based on a manuscript that we have written together.

Chapter 5

Activity-based proteomics reveals target proteases of tomato cystatin SlCYS8 in agroinfiltrated *N. benthamiana*

Philippe V. Jutras, Friederike Grosse-Holz, Farnusch Kaschani, Markus Kaiser, Dominique Michaud, Renier A.L. van der Hoorn. (manuscript in preparation). Activity-based proteomics reveals target proteases of tomato cystatin SlCYS8 in agroinfiltrated *N. benthamiana*

Authors

Philippe V. Jutras[1], Friederike Grosse-Holz[1], Farnusch Kaschani[2], Markus Kaiser[2], Dominique Michaud[3], Renier A. L. van der Hoorn[1]*

Affiliations

[1] Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

[2] Chemische Biologie, Zentrum für Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany.

[3] Centre de recherche et d'innovation sur les végétaux, Université Laval, Québec, Canada

* Corresponding author

Running title

Target proteases of SlCYS8

Author contributions

P.V.J., R.H. and D.M. conceived the research. F.G.H. and P.V.J. designed experiments. F.G.H., P.V.J. and R.H. interpreted the results. P.V.J. performed ABPP experiments for in-gel fluorescence scanning with guidance from F.G.H., F.G.H. performed ABPP-MS experiments. F.K. and M.K. performed LC-MS/MS and peptide and protein identification steps for MS. P.V.J. and F.G.H. co-wrote the manuscript with feedback from R.H. and D.M.

5.1 Summary

Co-expression with protease inhibitors is increasingly used to enhance the accumulation of recombinant proteins in plants, but the inhibited proteases have not yet been identified. We here performed activity-based protein profiling to characterize the impact of the tomato cystatin SlCYS8 on protease activities of agroinfiltrated *Nicotiana benthamiana*. SlCYS8 specifically inhibited the activity of papain-like cysteine proteases (PLCPs) in total leaf extracts and in apoplastic fluids both *in vivo* and *in vitro*. SlCYS8 did not deplete the activity of serine hydrolases and vacuolar proteases. Mutations at a positively selected amino acid site in the N-terminal region of SlCYS8 led to differential inhibitory function of the mutants. Active plant serine hydrolases and PLCPs were analysed using activity-based proteomics and we thus identified the proteases inhibited by SlCYS8. The Xylem-Bark Cys Protease (XBCP), Xylem-specific Cys protease (XCP) and Resistant-to-Desiccation 19/21-like (RD19/RD21) *N. benthamiana* PLCP subfamilies were inhibited by SlCYS8, while we found no impact of SlCYS8 on Aleurain-Like Proteases (ALP), Cathepsin-B-like proteases (CTB) or serine hydrolases. Intriguingly, the *N. benthamiana* PLCP activity profile is altered in the presence of SlCYS8 by posttranslational processing of RD21-like proteases carrying a granulin domain. We identify endogenous *N. benthamiana* proteases inhibited by SlCYS8, which are likely to be involved in the degradation of recombinant proteins.

5.2 Introduction

Although plant cells have been successfully used as alternative expression hosts to produce recombinant proteins (Daniell et al., 2015; Ma et al., 2015; Lomonossoff and D'Aoust, 2016), relatively low yields of proteins remain an important limitation for plant-based expression systems. Unintended proteolytic processing by plant proteases leads to either partial or complete degradation and is a limiting factor that affects the quantity and quality of recombinant proteins (Mandal *et al.*, 2016; Pillay *et al.*, 2016). Hundreds of genes code for proteases of diverse classes involved in various physiological processes in plants (van der Hoorn, 2008). Controlling the plant proteolytic activity is a worthwhile approach to increase both accumulation levels and quality of recombinant proteins (Benchabane *et al.*, 2008; Mandal *et al.*, 2016). Unintended proteolysis has been mitigated in plant-based expression systems by (i) targeting proteins to alternative cellular compartments (Amrani *et al.*, 2004), (ii) the addition of translational fusion partners to increase protein stability (Sainsbury *et al.*, 2013), (iii) the removal of protein domains and sequences targeted by endogenous plant proteases (Zischewski *et al.*, 2016; Hehle *et al.*, 2016) or (iv) creation of protease activity-depleted environments by cellular engineering. This includes protease gene silencing with RNAi (Mandal *et al.*, 2014; Duwadi *et al.*, 2015) and inhibition of proteases with recombinant inhibitors in stable or transient plant expression systems (Komarnytsky *et al.*, 2006; Goulet *et al.*, 2010a; Jutras *et al.*, 2016).

Plant family C1A papain-like Cys proteases (PLCPs) have been identified as a major constituent of the recombinant protein degradation machinery (Niemer *et al.*, 2014; Jutras *et al.*, 2016). PLCPs are targeted by plant cystatins, which harbour the conserved motif Gln-Xaa-Val-Xaa-Gly (QxVxG) in the central region of the polypeptide chain (Benchabane *et al.*, 2010; Martínez *et al.*, 2012). Cystatins are competitive

protease inhibitors that form non-covalent bonds to occupy the active site of target proteases (Benchabane *et al.*, 2010). Transgenic tobacco plants constitutively expressing a rice cysteine protease inhibitor (OC-I) have significantly lower protease activity, and this directly correlated with higher accumulation of a recombinant glutathione reductase (Pillay *et al.*, 2012). Cystatins have also been used as a co-expression partner for an increased accumulation of protease-susceptible proteins in the cell secretory pathway of *N. benthamiana* leaves, notably leading to higher yields of fully assembled plant-expressed antibodies (Robert *et al.*, 2013; Jutras *et al.*, 2016). However, the target proteases of plant cystatins, which are implicated in recombinant protein degradation, remain elusive.

Our objective in this study is to identify proteases that are inhibited by the *Solanum lycopersicum* (tomato) Cys protease inhibitor cystatin 8 (SlCYS8) (Girard *et al.*, 2007). In tomato, SlCYS8 expression is induced by the defence signal arachidonate and the protein forms an 88-kDa eight-unit multicystatin that accumulates in the cytosol of leaf cells upon herbivory (Girard *et al.*, 2007). Transient expression of SlCYS8 in the cell secretory pathway of agroinfiltrated *Nicotiana benthamiana* plants prevents proteolytic trimming of recombinant proteins migrating towards to the apoplast (Robert et al., 2013; Jutras et al., 2016). We here characterized the impact of SlCYS8 expression on protease activity profiles of agroinfiltrated *N. benthamiana* leaves and identified SlCYS8 target proteases by protein mass spectrometry. We thus show the practical potential of SlCYS8 to specifically inhibit plant PLCPs and pinpoint proteases likely involved in the degradation of recombinant proteins.

5.3 Results and discussion

5.3.1 SlCYS8 alters PLCP activity profiles in agroinfiltrated leaves

Because most cystatins are PLCP inhibitors (Benchabane et al., 2010), we assessed whether the expression of a secreted version of the SlCYS8 inhibitor could alter PLCP activity in agroinfiltrated leaves. Leaf total soluble proteins were extracted to provide an overview of protease activity in plants agroinfiltrated with a SlCYS8 transgene-harbouring vector, or with a mock 'empty-vector' version of the same vector. As negative control, we expressed an inactive version of the inhibitor, SlCYS8^{Q47P} (or Q47P), which bears a proline (P) instead of a glutamine (Q) in the QxVxG motif (Sainsbury et al., 2013).

Protease activity-based profiling (ABPP) uses chemical protease inhibitors fused to a reporter tag, called activity-based probes. The probes covalently bind available active sites and active proteases are thus labelled with the tag, either biotin for active protein enrichment or a fluorescent molecule for visualization of activity profiles in SDS-PAGE gels (Morimoto & van der Hoorn, 2016). We here used ABPP to characterize PLCP activity by incubating extracted proteins with the activity-based probe MV201, a fluorescent derivative of the chemical inhibitor E-64 (Richau et al., 2012). SDS-PAGE gel analysis showed a significant adjustment of PLCP activity in leaves expressing SlCYS8 compared to the empty-vector control. Fluorescence intensity of strong signals detected at ~40, ~35 and ~14 kDa (*see* Bands A, C and F, Figure 5.1A) decreased in the presence of SlCYS8, indicating that PLCP activity was depleted compared to the empty-vector control. In contrast, fluorescence of three other signals (*see* bands B, D, and E, Figure 5.1A) increased in leaves expressing SlCYS8. The rebalancing of PLCP activity was not induced by the expression of the SlCYS8^{Q47P}

inactive mutant. Pre-incubation with PLCP inhibitor E-64 suppressed MV201 labelling, indicating that the signals correspond to PLCPs. An immunoblot with an anti-SlCYS8 antibody confirmed the expression of both SlCYS8 constructs, with a lower yield of the SlCYS8^{Q47P} mutant (Figure 5.1A).

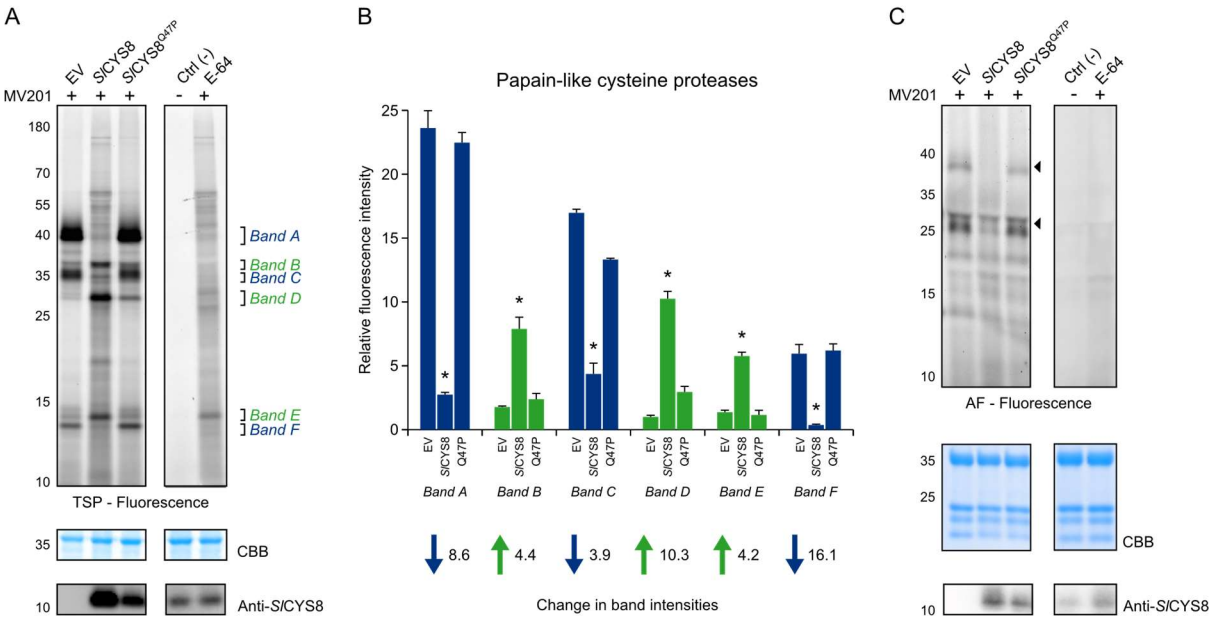


Figure 5.1. Papain-like cysteine protease activity profiles of agroinfiltrated *N. benthamiana* leaves expressing SlCYS8. (A) Activity-based profiling (ABPP) of PLCPs in total soluble protein extracts from agroinfiltrated leaves transiently expressing SlCYS8 or an inactive mutant of SlCYS8 (SlCYS8^{Q47P}), or agroinfiltrated with an empty vector construct (EV). Proteins were labelled with the MV201 fluorescent probe and revealed by in-gel fluorescence scanning. Fluorescent bands with different intensities were named band A to F. Proteins were electro-transferred onto a PVDF membrane for immunodetection with an anti-SlCYS8 antibody. Ctrl (-): mix of plant extracts with no probe, E-64: mix of plant extracts pre-incubated with the chemical cysteine protease inhibitor E-64. (B) Quantification of fluorescence intensity for bands A to F. Each bar is the mean of three biological replicate values ± SE. Bars with an asterisk are significantly different from EV (Student's t test; $P<0.05$). Arrows show fold change in fluorescence intensity of bands from plants expressing SlCYS8 compared to the empty-vector control. (C) ABPP of PLCP activity in apoplastic fluids (AF) of agroinfiltrated plants. Black arrows show PLCP bands with different fluorescence intensities in plants expressing SlCYS8. Immunoblotting confirms the presence of SlCYS8 in the apoplast.

Gel band fluorescence intensities were quantified to confirm the altered activities of PLCPs in leaves expressing SlCYS8 (Figure 5.1B). Six bands were significantly affected by the expression of the inhibitor (Student's *t* test; $P<0.05$). Fluorescence intensities of bands A, C and F decreased when SlCYS8 was expressed, with an 8.6-, 3.9- and 16.1-fold reduction, respectively. Fluorescence intensity of bands B, D and E increased in SlCYS8 expressing leaves, with a 4.4-, 10.3- and 4.2-fold increase, respectively (*see b*ands B, D and E, Figure 5.1B). This increase in fluorescence could be caused by enhanced transcription and translation of PLCPs, by post-translational activation or by conversion of active PLCPs into lower molecular weight versions. No significant effect on band fluorescence intensity was detected in protein extracts from plants expressing the SlCYS8^{Q47P} variant, confirming that the effect of SlCYS8 is associated with its inhibitory function.

To assess whether SlCYS8 inhibitory effects extend to the extracellular space of *N. benthamiana* leaves, apoplastic fluids were isolated from agroinfiltrated leaves. PLCP labelling with MV201 showed active PLCPs in the apoplast of the empty vector control (Figure 5.1C). SlCYS8 expression depleted the activity of the main extracellular PLCPs, in contrast to SlCYS8^{Q47P} (*see* black arrows, Figure 5.1C). A coomassie-stained gel of apoplastic fluids showed same protein accumulation profiles between leaf samples, suggesting that SlCYS8 expression has no effect on total protein concentration in the extracellular space (Figure 5.1C).

These data describe an alteration of PLCP activity profiles upon SlCYS8 expression in agroinfiltrated *N. benthamiana* leaves, associated with depletion of PLCP activity both inside and outside the cells.

5.3.2 SlCYS8 expression does not deplete SH and VPE activity in agroinfiltrated leaves

To confirm the specificity of SlCYS8 to PLCPs, we performed ABPP on leaf total soluble protein extracts with activity-based fluorescent probes FP-TAMRA (FP), targeting Serine Hydrolases (SHs) (Kaschani *et al.*, 2009), and JOPD1, targeting Vacuolar Processing Enzymes (VPEs) (Lu *et al.*, 2015). SH activity profiles were mainly unaffected by the expression of SlCYS8 or the SlCYS8^{Q47P} mutant, except for an additional band appearing at high molecular weight in SlCYS8-expressing leaves (*see* Band G, Figure 5.2A) and a slight increase of intensity at ~37 kDa (*see* Band H, Figure 5.2A). Quantification of fluorescence intensities revealed that intensity of bands G and H increased 5.9 and 1.1-fold (Student's *t* test; $P<0.05$), respectively, compared to control leaves (Figure 5.2B). Plant SH activity profiles have been previously characterized (Kaschani et al., 2009), and bands G and H likely correspond to subtilase (SBT, S08) and Ser Carboxypeptidase-like (SCPL, S10) activities.

VPE activity profiles were unchanged, except for intensity of one band at ~40 kDa, which increased 1.5-fold (Student's *t* test; $P<0.05$) in SlCYS8-expressing leaves when compared to the control (*see* Band I, Figure 5.2C-D). Activity profiles in protein extracts expressing the SlCYS8^{Q47P} mutant resembled the control, again confirming that the effect on *N. benthamiana* protease activity depends on SlCYS8 inhibitory function.

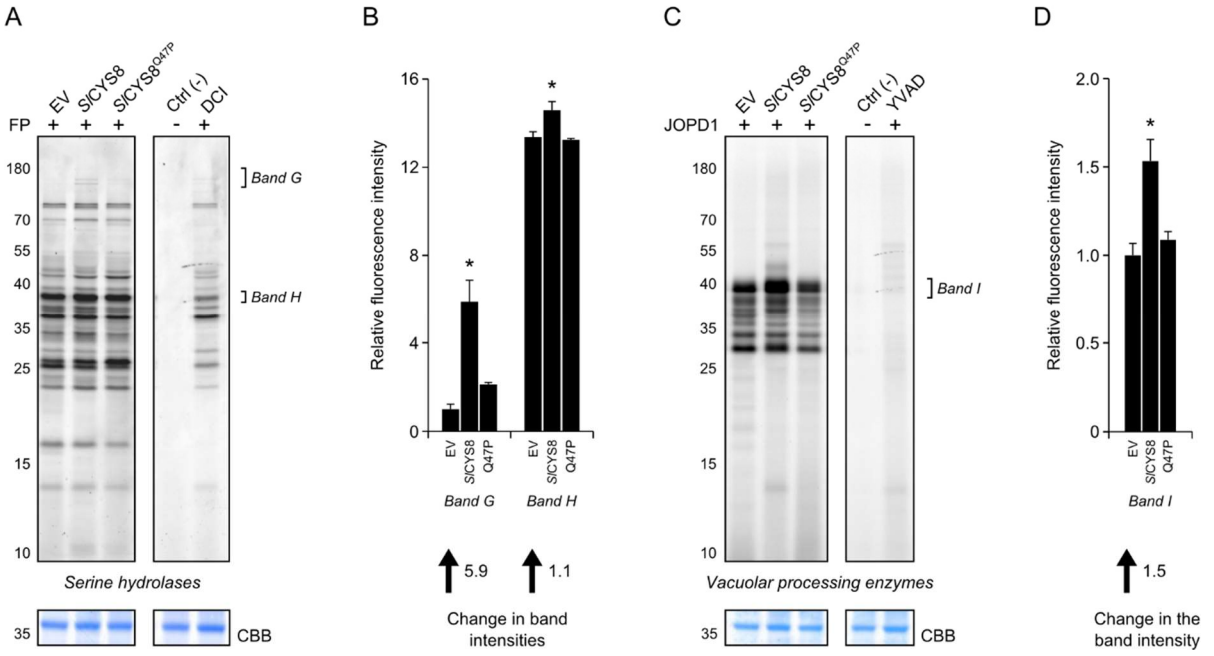


Figure 5.2. **Serine hydrolase and vacuolar processing enzyme activity profiles of agroinfiltrated *N. benthamiana* leaves expressing SlCYS8.** Activity-based profiling of (A) serine hydrolase (SH) activity and (C) vacuolar processing enzyme (VPE) activity in total soluble protein extracts from agroinfiltrated leaves transiently expressing SlCYS8 or an inactive mutant of SlCYS8 (SlCYS8Q47P), or agroinfiltrated with an empty vector construct (EV). Proteins were labelled with the FP or JOPD1 fluorescent probes to target SHs and VPEs, respectively. Bands with different fluorescent intensities were named band G to I and quantified (B-D). Each bar is the mean of three biological replicate values ± SE. Bars with an asterisk are significantly different from EV (Student's t test; $P<0.05$). Arrows show fold change in fluorescence intensity of bands from plants expressing SlCYS8 compared to the empty-vector control. Ctrl (-): mix of plant extracts with no probe, DCI: mix of plant extracts pre-incubated with the chemical serine protease inhibitor 3,4-dichloroisocoumarin, YVAD: plant extracts pre-incubated with Ac-Tyr-Val-Ala-Asp-Chloromethylketone (YVAD).

Taken together, these observations confirm the inhibitory effect of SlCYS8 on PLCPs and suggest that the activity of other plant protease families may be slightly modulated by the expression of SlCYS8. Minor increases of SH and VPE activities upon SlCYS8 expression might be a compensatory response initiated upon depletion of most PLCP activity. A compensatory response is elicited for instance by proteasome inhibition, which induces the transcription of gene groups associated with the proteasomal

degradation pathway (Michel *et al.*, 2006). Interestingly, SlCYS8 expression had no effect on the activity of apoplastic SHs and VPEs, indicating that the compensatory reaction only occurred within the cells (Supplemental Figure 5.S01).

5.3.3 *In vitro* assays confirm direct inhibition of PLCPs by SlCYS8

To determine whether SlCYS8 expression depletes PLCP activity by direct inhibition, ABPP was performed on extracts from agroinfiltrated plants upon pre-incubation with purified recombinant SlCYS8 produced in *E. coli* (Figure 5.3). PLCP labelling of leaf total soluble proteins pre-incubated with 0 to 150 nM SlCYS8 revealed that SlCYS8 addition depleted the activity of three major bands (*see* bands A, C and F, Figure 5.3) and increased intensities of three minor bands (*see* bands B, D and E, Figure 5.3), as previously observed *in vivo* (Figure 5.1A). Fluorescence intensities were quantified and, when plotted against the SlCYS8 concentration, revealed activity depletion to be SlCYS8 concentration-dependent, up to 75%, 62% and 96% decrease in signals for bands A, C and F, respectively (Figure 5.3B). Most notably, the signal of band F was rapidly suppressed by SlCYS8, reaching 85% inhibition in the presence of <50 nM SlCYS8. Remarkably, concentrations of SlCYS8 above 20 nM resulted in a 3.5- to 5-fold increase in fluorescence of bands B, D and E when compared to the no inhibitor control (Figure 5.3C). Compensation for PLCP inactivation through enhanced transcription of PLCP-encoding genes is thus unlikely to be responsible for increased PLCP signals. Alternatively, protease activity may increase or active proteases may shift in size due to *in vitro* post-translational maturation (Gu *et al.*, 2012). For instance, the Arabidopsis Cys protease Resistant-to-Desiccation 21 (RD21) is regulated at three post-translational levels, removal of the N-terminal inhibitory pro-domain, removal of

the C-terminal granulin domain and an SDS-induced process that may involve plant protease inhibitors interacting with RD21 (Yamada *et al.*, 2001; Gu *et al.*, 2012).

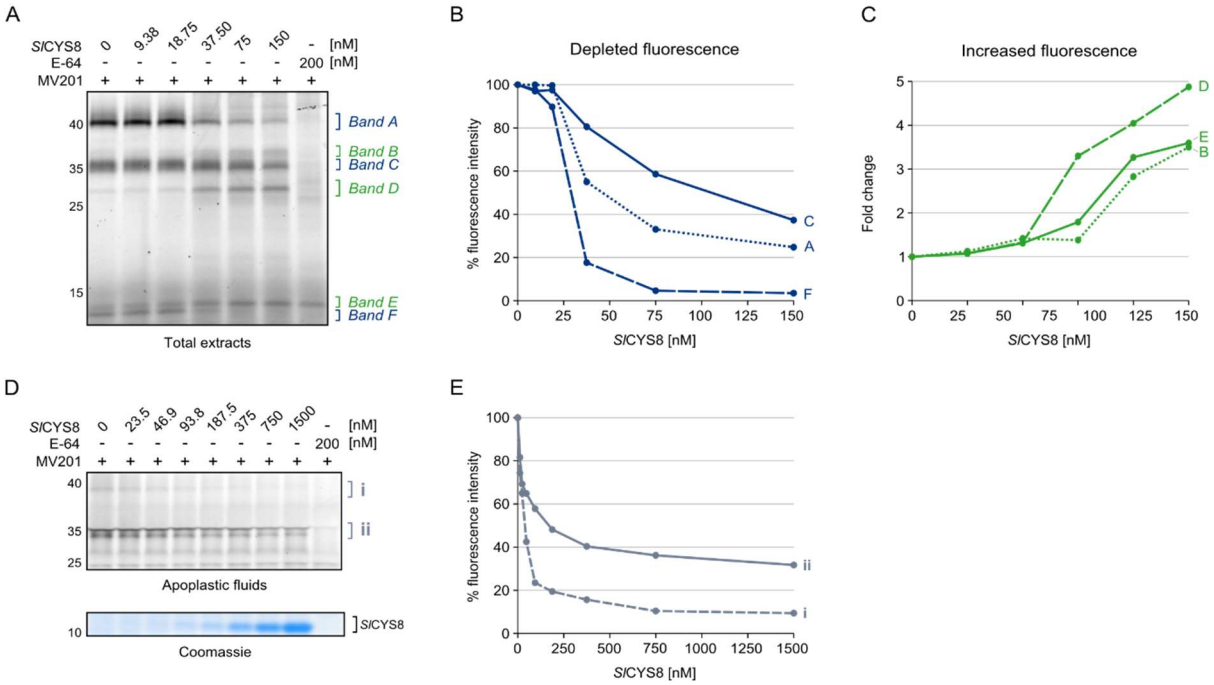


Figure 5.3. *In vitro* inhibition of papain-like cysteine proteases from agroinfiltrated *N. benthamiana* leaves. (A) Total soluble protein extracts from leaves agroinfiltrated with the empty-vector control were incubated with 0 to 150 nM SlCYS8 purified from *E. coli*, prior to PLCP activity labelling with the MV201 fluorescent probe. Bands A to F were previously described in Figure 5.1. (B) Inhibition rates of bands A, C and F upon SlCYS8 concentrations. (C) Fold change of increased fluorescence intensities of bands B, D and E. (D) Apoplastic fluids of plants agroinfiltrated with the empty-vector control were incubated with 0 to 1500 nM SlCYS8, prior to PLCP activity profiling. Two bands with different fluorescence intensities were named i and ii. A coomassie-stained gel shows the increased amount of purified SlCYS8. (E) Inhibition rates of bands i and ii in apoplastic fluids. E-64: plant extracts pre-incubated with the cysteine protease inhibitor E-64.

Inhibition assays were also performed on apoplastic fluids isolated from agroinfiltrated plants. Higher concentrations of SlCYS8 (from 0 to 1500 nM) were required to suppress labelling (Figure 5.3D-E), in accordance with previous findings that proteases are enriched in the apoplast upon agroinfiltration (Delannoy *et al.*, 2008; Goulet *et al.*, 2010b). Our data show a concentration-dependent inhibition of PLCPs

in the apoplast, with 90% and 65% inhibition of the two major bands at 500 mM SlCYS8 (Figure 5.3D-E). These results confirm the direct inhibitory effect of SlCYS8 on both intra- and extracellular PLCPs.

5.3.4 N-terminal single mutations affect SlCYS8 inhibitory function

As positively selected amino acids in the N-terminal sequence impact the inhibitory function of SlCYS8 (Kiggundu *et al.*, 2006; Vorster *et al.*, 2015; Rasoolizadeh *et al.*, 2016a), we engineered four mutants of SlCYS8 to generate cystatin variants with altered protease target ranges and/or improved inhibitory potency. Single mutations at position 2 were introduced to replace the Pro residue (Pro-2, or P2) adjacent to the conserved GG motif in the N-terminus (Goulet *et al.*, 2008). The proline residue was changed to either isoleucine (P2I), leucine (P2L), valine (P2V) or tyrosine (P2Y) (Figure 5.4A), inspired by different specificity of these SlCYS8 versions on insect digestive Cys proteases (Rasoolizadeh *et al.*, 2016a). Mutants and wild-type SlCYS8 were transiently expressed in *N. benthamiana* leaves and total soluble proteins were labelled with MV201 to compare the ability of the mutants to inhibit Cys proteases (Figure 5.4B). All mutants suppressed PLCP labelling similarly (*see* bands A to F, Figure 5.4B) and accumulated to similar levels in plant cells. Most of the bands in the PLCP activity profile showed no variation between the mutants. However, band D was less strong in leaves expressing the mutants when compared to wild-type SlCYS8 (post-ANOVA Tukey's test; $P<0.05$), except for P2V that showed non-significant difference (Figure 5.4C). The SlCYS8 P2V mutant has a stronger inhibitory potency against insect digestive proteases when compared to wild-type SlCYS8 (Rasoolizadeh *et al.*, 2016b). ABPP on total soluble protein extracts from leaves expressing the SlCYS8 mutants with activity-based fluorescent probes targeting SHs and VPEs

revealed minor impacts of SlCYS8 mutations on SH activity and no effect on VPE activity (Supplementary Figure 5.S02).

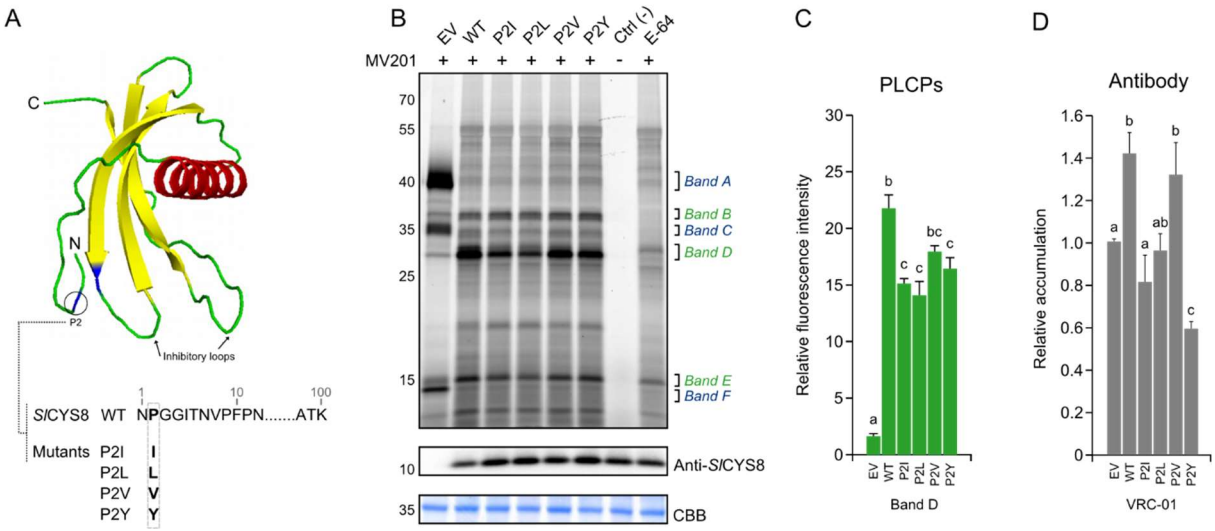


Figure 5.4. Impact of single mutations in the N-terminus of SlCYS8 on PLCP inhibition and accumulation of recombinant VRC01 antibody in *N. benthamiana*. (A) Single mutations of the proline at the position 2 in the N-terminal of SlCYS8. Structure shows the two inhibitory loops. The model was built with the NMR structure of rice cystatin I (PDB 1EQK) as a template. (B) SlCYS8 mutants were transiently expressed in *N. benthamiana* and leaf total soluble proteins were labelled with MV201 to characterize PLCP activity. Bands A to F were previously described in Figure 5.1. Protein extracts were electro-transferred onto a PVDF membrane for immunodetection with an anti-SlCYS8 antibody. (C) Quantification of band D fluorescence intensity. (D) Quantitative ELISA for VRC01 antibody co-expressed together with SlCYS8 mutants in the secretory pathway. Each bar (Panel C and D) is the mean of three biological replicates ± SE. Bars with different letters are significantly different (post-ANOVA Tukey's test; $P<0.05$). EV: extracts from plants agroinfiltrated with the empty-vector control, WT: wild-type SlCYS8, P2I: proline to isoleucine, P2L: proline to leucine, P2V: proline to valine, P2Y: proline to tyrosine, Ctrl (-): mix of plant extracts with no probe, E-64: mix of plant extracts pre-incubated with the chemical cysteine protease inhibitor E-64 before activity labelling.

The impact of the P2 N-terminal residue on SlCYS8 inhibitory potency was also assessed by characterizing antibody-stabilizing effects of the mutants (Robert et al., 2013; Jutras et al., 2016). The SlCYS8 mutants were transiently co-expressed in *N. benthamiana* leaves together with the heavy and light chains of the VRC01 antibody (Hamorsky *et al.*, 2013). The mutants differently affected the accumulation of VRC01,

as measured by a quantitative ELISA (Figure 5.4D). The highest accumulation of the VRC01 antibody was observed upon co-expression with wild-type SlCYS8 or the P2V mutant and none of the other tested mutants increased VRC01 levels more than the wild-type SlCYS8. In future, novel cystatin variants may be designed by engineering the two inhibitory loops of SlCYS8 to improve inhibition of plant proteases and stabilization of recombinant proteins in the plant secretory pathway.

5.3.5 Activity-based proteomics reveals *N. benthamiana* PLCPs inhibited by SlCYS8

To identify the proteases that are inhibited by SlCYS8 and thus potentially implicated in recombinant protein degradation, we performed activity-based proteomics (ABPP-MS) on total soluble protein extracts from plant expressing SlCYS8 (*in vivo*) and control empty-vector agroinfiltrated leaves (mock), pre-incubated with or without 280 nM of purified SlCYS8 (*in vitro*). Proteomes were incubated with a mixture of FP-biotin and DCG04 probes to label both active SHs and PLCPs, respectively. Labelled proteins were captured on avidin beads prior to label-free quantitative mass spectrometry (MS) analysis (Greenbaum, 2002; Kaschani et al., 2009). Fold change in protein abundance compared to the no-probe-control was used to identify active proteins. 48 SHs were enriched at a confidence level of 95% (P<0.05) compared to the no-probe-control, indicating they were active. Among the SHs, activity of only one pectinacetylesterase (Niben101Scf02119g00018) was increased upon expression of SlCYS8 *in vivo* (*see* black star Figure 5.5A and Supplemental File ES5.01). This protein has a very low overall abundance and was not identified in the full leaf proteome obtained in Chapter 4. 14 active PLCPs were identified (Figure 5.5A and Supplemental File ES5.01) and clustered into six subfamilies (Grosse-Holz et al.,

2017, Chapter 2) (Figure 5.5B). Nine active PLCPs were significantly less abundant in leaves expressing SlCYS8 compared to the control, suggesting their activity was depleted through inhibition by SlCYS8 *in vivo*. Inhibited PLCPs belong to the Xylem Bark Cys Protease (XBCP), Xylem-specific Cys protease (XCP) and Resistant-to-Desiccation 19/21 (RD21 and RD19) subfamilies. Among the nine PLCPs inhibited upon SlCYS8 expression, eight were also inhibited in leaf extracts treated with purified SlCYS8 *in vitro*, confirming that these eight PLCPs are directly inhibited by SlCYS8. Two XBCP3s and one RD21-like PLCP were more strongly inhibited upon SlCYS8 expression *in vivo* compared to extracts treated with SlCYS8 *in vitro* (Figure 5.5B), potentially due to an incomplete inhibition of the proteases by 280 nM SlCYS8. XCP, RD19 and RD21-like *N. benthamiana* PLCPs are interesting candidates implicated in proteolytic processing of recombinant proteins in agroinfiltrated leaves, as those protease subfamilies are strongly inhibited by SlCYS8.

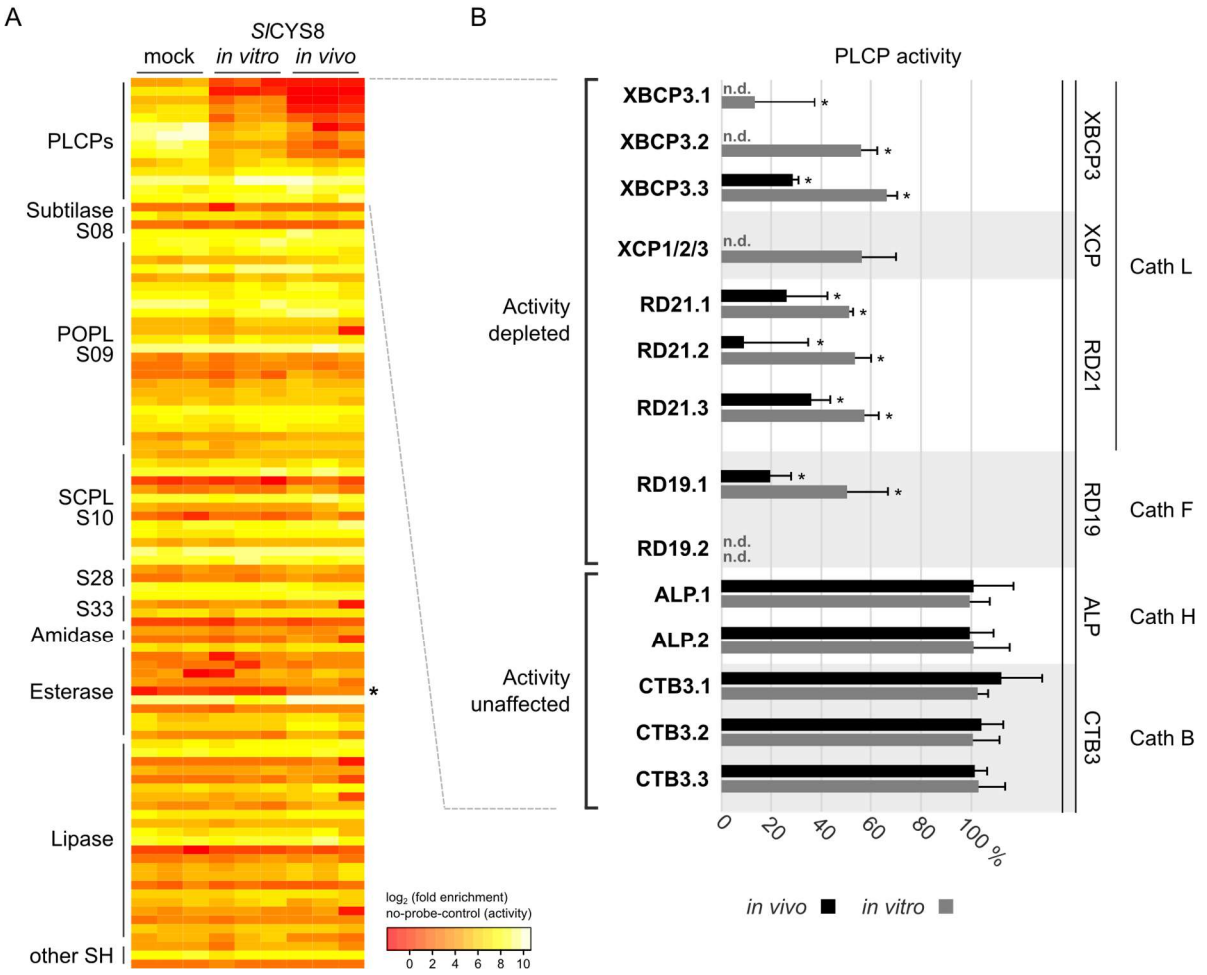


Figure 5.5. Impact of SlCYS8 expression on papain-like cysteine protease (PLCP) and serine hydrolase (SH) activity. (A) Heatmap of 101 active PLCPs and SHs identified by activity-based proteomics (ABPP-MS). ABPP-MS was performed using DCG04 and FP-biotin probes to capture active PLCPs or SHs, respectively. Proteins were enriched on avidin beads, prior to label-free quantitative MS analysis. Active proteins are enriched significantly (FDR<0.05) compared a no-probe control sample and annotated as probe targets, fold enrichment compared to the no probe control serves as a proxy for activity. The activity of only one SH was affected upon SlCYS8 expression in plants (*see* black star). Mock: extracts from leaves agroinfiltrated with the empty-vector control, SlCYS8 *in vitro*: extracts from leaves agroinfiltrated with the empty-vector control and pre-incubated with 280 nM of recombinant SlCYS8, SlCYS8 *in vivo*: extracts from agroinfiltrated leaves transiently expressing SlCYS8. (B) Activity of PLCPs, grouped by subfamily, relative to the mock control. Bars are means ± SE of three biological replicates, significant differences from the mock are indicated by a star (Student's *t* test, *P*<0.05). n.d., activity not detected. The MS dataset is in Supplemental File ES5.01.

Previous studies showed that the RD21-like NbCYSP6 and the XCP-like NbCYSP7 proteases can rapidly degrade antibodies *in vitro* (Paireder *et al.*, 2017). Furthermore, silencing the RD21-like (NtCYSP6) gene in tobacco significantly enhanced accumulation of recombinant proteins (Duwadi et al., 2015). As they are also targets of SlCYS8, RD21-like PLCPs very likely contribute to antibody degradation *in vivo*. XBCPs (C14-like PLCPs) are inhibited by SlCYS8, but likely play minor roles in recombinant protein proteolysis. Tomato XBCP proteases are specifically inhibited by cystatin-like EPIC inhibitors (Kaschani et al., 2010), but the co-expression of EPICs does not enhance the accumulation of recombinant proteins in *N. benthamiana* (Grosse-Holz et al, manuscript in preparation, Chapter 4). Interestingly, the proteolytic activity of Aleurain-Like Proteases (ALP) and Cathepsin-B-like proteases (CTB3) subfamilies was unaffected both by SlCYS8 expressed in plants (*in vivo*) or purified SlCYS8 (*in vitro*). *N. benthamiana* Cathepsin B proteases may contribute to residual antibody degradation in the presence of SlCYS8, as the NbCathB protease, a sequence 93% identical with NbCTB3.3, can degrade antibodies *in vitro* (Niemer *et al.*, 2016). In contrast, the ALP subfamily likely makes a minor contribution to proteolytic processing of recombinant proteins, as previous *in vitro* assays showed negligible effect of NbALP on antibody stability (Niemer et al., 2016). These data confirm the impact of SlCYS8 expression on host endogenous proteases and indicate that XCP, RD19 and RD21-like *N. benthamiana* PLCP subfamilies likely constitute the core of the recombinant protein degradation machinery that is blocked by SlCYS8, while Cathepsin B and other proteases may contribute to residual protein degradation in the presence of SlCYS8.

As none of the active PLCPs we identified by ABPP-MS showed increased activity upon exposure to SlCYS8 *in vivo* or *in vitro*, we reasoned that increased fluorescence intensity of bands B, D and E (*see* Figure 5.1A and Figure 5.3A) may be caused by a size shift of active proteases. We therefore performed a detailed analysis of peptides from proteases carrying granulin domains. Peptides corresponding to both the peptidase and granulin domains of three PLCPs were detected, namely RD21.1, XBCP3.2, XBCP3.3 (Figure 5.6A and Supplemental Figure 5.S03).

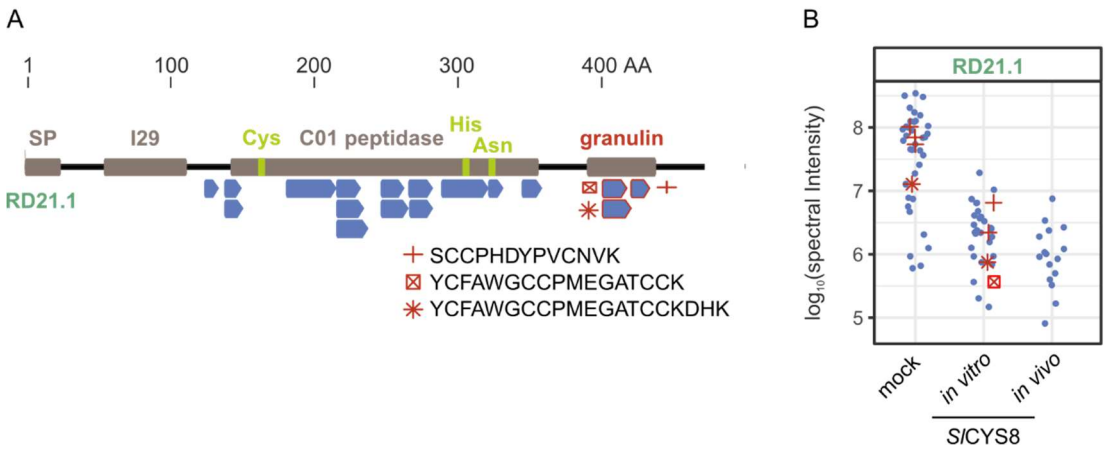


Figure 5.6. Peptides corresponding to the granulin domain of RD21.1 are depleted upon SlCYS8 expression. (A) Identified peptides (blue) from the granulin-domain containing PLCP RD21.1 (Niben101Scf09885XLOC_077597). SP: signal peptide, I29: I29 inhibitory domain, C01: peptidase, Granulin: granulin domain. The catalytic triad residues are presented in green. (B) Spectral intensity of the peptides in three biological replicates. The granulin-derived peptides are shown in red. Mock: extracts from leaves agroinfiltrated with the empty-vector control, SlCYS8 *in vitro*: extracts from leaves agroinfiltrated with the empty-vector control and pre-incubated with 280 nM of recombinant SlCYS8, SlCYS8 *in vivo*: extracts from agroinfiltrated leaves transiently expressing SlCYS8.

Spectral intensity of peptides corresponding to peptidase domains of RD21.1 decreased upon exposure to SlCYS8, indicating that the active site was less available to the probe due to inhibition by SlCYS8 (Figure 5.6B). However, peptides were still detected, indicating that RD21.1 is incompletely inhibited by SlCYS8. Interestingly, peptides corresponding to the granulin domain of RD21.1 disappeared in extracts of

leaves expressing SlCYS8 (*see* red symbols, Figure 5.6B), suggesting that the RD21.1 enzymes that remained active in leaves expressing SlCYS8 lost their granulin domains. These data highlight the impact of SlCYS8 expression on posttranslational maturation of RD21-like proteases carrying a granulin domain.

5.4 Conclusions

We here characterized the impact of SlCYS8 on protease activity in agroinfiltrated *N. benthamiana* leaves, leading to a better understanding of *in planta* protease-inhibitor interactions. We used activity-based protein profiling to study endogenous protease profiles of leaves transiently expressing SlCYS8. Our data highlight that SlCYS8 specifically inhibits PLCPs, with minor impacts on the activity of serine hydrolases and vacuolar processing enzymes. Proteomics analysis showed that PLCPs from the XBCP, XCP, RD19 and RD21-like PLCP subfamilies are inhibited by SlCYS8 and revealed an unexpected posttranslational maturation of granulin-containing RD21-like PLCPs in the presence of SlCYS8. We thus identify proteases involved in proteolytic processing of recombinant proteins. Depletion of identified proteases by genome editing and the expression of alternative Cathepsin B inhibitors could be used to reduce recombinant protein proteolysis in the secretory pathway of *N. benthamiana*. Structural/function experiments on SlCYS8 in coming years will also allow the rational design of efficient inhibitors with custom-made specificity.

5.5 Material and Methods

Plasmid constructs. Expression constructs for secreted SlCYS8 (GenBank Accession No. AF198390) and the inactive variant SlCYS8^{Q47P} were described previously (Sainsbury et al. 2013). Mutations in the N-terminal SlCYS8 sequence were introduced at position 2 by Quickchange mutagenesis (Agilent Technologies, Canada) to produce SlCYS8 variants with a proline exchanged for isoleucine (P2I), leucine (P2L), valine (P2L) or tyrosine (P2Y). Transgenes were assembled in a pCambia 2300 expression vector (CAMBIA, Australia), between a duplicated Cauliflower mosaic virus (CaMV) 35S promoter in 5' position and a nopaline synthase (NOS) terminator sequence in 3' position. An 'empty' mock vector was used as a negative control for agroinfiltration. The pCambia vectors were maintained in *Agrobacterium tumefaciens*, strain AGL1 (Lazo *et al.*, 1991). The VRC01 plasmids are described in Chapter 4 and were maintained in *Agrobacterium tumefaciens*, strain GV3101. Agrobacteria were cultured on plates of LB medium (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract, 15 g/L agar) containing 25 µM Rifampicin, 50 µM Gentamycin and 50 µM Kanamycin to select for transformants. All gene constructs were proof checked by automatic DNA sequencing.

Transient expression in leaves. *N. benthamiana* plants were grown at 21 °C under a 16/8 h light/dark regime in a growth room. *Agrobacterium* containing binary expression plasmids were grown for 21h at 28 °C with agitation in LB containing the appropriate antibiotics. Bacteria were collected by centrifugation at 2000 g for 5 min at room temperature (RT), resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfone (MES), 10 mM $MgCl_2$, pH 5.7, 100 µM acetosyringone) to OD_{600} = 0.5 and left for 2 h at 28 °C with agitation to recover. All *A. tumefaciens* suspensions were mixed in a 1:1 ratio with a suspension of *A. tumefaciens* carrying a pCambia 2300

binary vector harbouring an expression cassette for the silencing suppressor protein p19. The first and second fully expanded leaves of pre-flowering *N. benthamiana* (4-5 weeks old) were infiltrated with the bacteria suspension using a syringe without a needle. Leaf tissue was harvested 6 days post infiltration for protein extraction to allow for maximum protein accumulation in the presence of the p19 silencing suppressor. Three independent replicates including leaves of three plants each were used for each treatment to minimize variation of protein expression levels and to allow for statistical analysis of the data**.**

Protein extraction. For total soluble protein extraction, infiltrated leaf tissue was harvested as leaf discs, flash-frozen in liquid Nitrogen and pulverized using pestle and mortar. Proteins were extracted in three volumes (v/fresh weight) of cold 500 mM sodium acetate (NaAc), pH 5, 5 mM DTT and centrifuged for 20 min at 4 °C 16000 x g. For apoplastic fluid extraction, six *N. benthamiana* leaves per sample were detached and vacuum-infiltrated with ice-cold water, dried on the surface and placed in a syringe without needle and plunger that was inserted in a 50-mL falcon tube. Apoplastic fluids were collected by centrifugation at 2000 g, 4°C for 25 min and used immediately. Protein concentrations were determined using a linearised Bradford assay (Ernst and Zor, 2010).

Activity-based protein profiling. 48 µl of leaf total soluble protein extracts, or apoplastic fluids adjusted to 500 mM NaAc, pH 5, 5 mM DTT, were pre-incubated with or without 0.2 mM of inhibitor (E-64; DCI, 3,4-Dichloroisocoumarin; Caspase I) for 30 min and incubated for 4 h at room temperature with fluorescent probes targeting papain-like cysteine proteases (probe MV201) or vacuolar processing enzymes (probe JOPD1), or incubated for 1 h with an FP-TAMRA probe targeting serine hydrolase activity. FP-TAMRA was obtained from Thermo (88318), MV201 and JOPD1 were

synthesized as described (Richau et al., 2012; Lu et al., 2015). ABPP reactions were ended by adding 1 mL cold acetone. Samples were centrifuged for 3 min at 16000 and the supernatant discarded. Proteins were resuspended in 2 x gel loading buffer (100 mM Tris-HCl (pH 6.8), 200 mM DTT, 4% SDS, 0.02% bromophenol blue, 25% glycerol), heated for 5 min at 95 °C and proteins were resolved by 12% w/v SDS-PAGE in reducing conditions.

In-gel fluorescence intensity quantification. SDS-PAGE gels were scanned on a Typhoon scanner (Amersham/GE Healthcare, Little Chalfont, UK) using Cy3 settings to detect in-gel fluorescence. Image data were analyzed using the Open source software ImageJ (http://rsb.info.nih.gov/ij/) to quantify band fluorescence intensity. Background values were subtracted from each image based on the average values of images acquired from no probe controls. Bands from three biological replicates of were used to determine the fluorescence intensity.

Immunoblotting. Proteins were resolved by a 12% w/v SDS-PAGE gel and were electrotransferred onto a PVDF membrane using the TransBlot Turbo system (Biorad, Hercules, US). SlCYS8 was detected by a primary anti-SlCYS8 polyclonal IgG raised in rabbits (Agrisera, Sweden, 1/5000) and by a goat anti-rabbit horseradish peroxidase (HRP)-conjugated secondary antibody (Agrisera, Sweden, 1/10000). Non-specific binding sites were blocked with 5% w/v skimmed milk powder in PBS buffer containing 0.025% v/v Tween-20 for 1h, which also served as antibody dilution buffer. Chemiluminescent signals were revealed using the Clarity ECL Western blotting detection kit (Bio-rad, USA) and signals were captured on a Gel Doc imager (Bio-rad, USA).

Bacterial expression of recombinant SlCYS8**.** *E. coli* GST expression for purification of SlCYS8 was carried out as reported (Goulet *et al.*, 2008). Briefly, 5 ml pre-cultures of BL21 cells were incubated in LB over night at 37°C, with carbenicillin antibiotic. The pre-cultures were transferred in 250 ml LB cultures and incubated at 37°C until an OD_{600} of 0.6. Protein expression was induced by the addition of 0.5 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were incubated for another 6h at 30°C. Bacteria were collected by centrifugation at 3 500 rpm for 10 minutes and cells were lysed by 4 freeze-thaw cycles. Bacteria were resuspended in 3 ml of lysis buffer (50 mM TRIS pH8, 5% m/v sucrose, 50 mM EDTA, 5% v/v Triton X-100, 1 mM PMSF), incubated 5 minutes on ice and centrifuged at 12 000 rpm for 10 min 4°C. Supernatant was incubated with Sepharose-4B beads (GE Healthcare Life Sciences, USA) for 60 min with low agitation. Beads were washed 3 times using 5 ml of 50 mM TRIS pH 8. Factor Xa enzyme and buffer were then added. Samples were centrifuged at 8000 rpm for 5 min after 16h incubation at room temperature with agitation and supernatant was collected. Protein concentration was assayed by standard Bradford protocol and purified proteins were maintained at -80°C.

***In vitro* inhibition assays.** Total soluble protein extracts (TSP) or apoplastic fluids (AF) from plants agroinfiltrated with the control 'empty-vector' were extracted in 50 mM sodium acetate buffer pH 5 and purified SlCYS8 proteins were added at different concentration (0 to 150 nM for TSP and 0 to 1500 nM for AF) for a 30 minutes incubation at room temperature prior to PLCP ABPP labelling as previously described. The same amount of factor Xa cleavage buffer was added in the no inhibitor control. Inhibition rates were obtained by the quantification of band fluorescence intensity compared to the no inhibitor control. A coomassie-stained 12% SDS-PAGE gel showed the increased amount of purified SlCYS8 added in the samples.

Quantitative ELISA. For the quantification of VRC01 heavy chain, an enzyme-linked immunosorbent assay (ELISA) plate (Fisher Scientific, USA) was coated for 1 h at 37˚C with freshly prepared leaf protein extracts. Serial extract dilutions were prepared in PBS buffer and a relative standard curve was generated from diluted VRC01 samples. All dilutions were performed in a control extract from leaf tissue infiltrated with a mock inoculum so that any unspecific matrix effect was eliminated. The plate was washed three times in PBS buffer, blocked for 1 h at 37˚C with 5% w/v skim milk powder in PBS buffer, and washed three times again in PBS buffer. The plate was incubated with an anti-human γ chain horseradish peroxidase (HRP)-conjugated antibody (Sigma A6029, USA, 1/10 000) in blocking solution for 1h at 37ºC. Washes with PBS were repeated, and the plates were incubated with the 3,3', 5,5'- tetramethyl-benzidine (TMB) peroxidase substrate (Sigma-Aldrich, USA). The reaction was stopped by the addition of H_2SO_4, before reading the absorbance at 450 nm. Each sample was assayed in triplicate and antibody concentrations were interpolated from the linear portion of the standard curve. All measurements were made with leaf protein extracts from three independent biological replicates.

Activity-based proteomics. Agrobacteria were grown separately for each biological replicate. Three leaves from different plants were infiltrated with an agrobacterial suspension of 1/1 (v/v) SlCYS8/P19 vector harbouring bacteria at OD_{600} = 0.5, three leaves with 1/1 (v/v) $SlCYS8^{Q47P}$/P19 and six leaves from three plants with only P19 vector harbouring bacteria. Total soluble leaf protein extracts were extracted at 6 dpi, prepared as described above and diluted to 1.5 mg/mL. In the 'SlCYS8 spike' samples, 3 ng of purified SlCYS8 from *E. coli* were added per 1 ul volume. Corresponding volumes of buffer were added to the other samples and all samples were incubated at RT on a rotator for 45 min. 1 mL of each sample was then incubated with 5 μM FP-

biotin (Sigma 88317) and 5 µM DCG04-biotin (Greenbaum, 2002) for 5 h on a rotator at RT in darkness. The no-probe control contained a mix of equal volumes of each sample and DMSO instead of the probe. MS-grade water and chemicals were used from this point onwards. Samples were transferred to 15 mL falcon tubes and 4 mL ice-cold methanol, 1 mL ice-cold chloroform and 3 mL ice-cold water were added subsequently, vortexing the sample after each step. Samples were centrifuged for 30 min at 3000 x g and 4°C. The upper aqueous layer was carefully removed, 4 mL methanol added and the samples were again centrifuged for 30 min at 3000 x g and 4°C. The supernatant was discarded, the pellet dried, resuspended in 2 mL Phosphate Buffered Saline (PBS, 10 mM PO_4^{3-}, 137 mM NaCl, 2.7 mM KCl) containing 1.2% sodium dodecyl sulphate (SDS) and then diluted using 5 mL PBS. Proteins were denatured by heating at 90°C for 8 min and cooled on ice afterwards, then diluted with 3 mL of PBS and stored at -20°C until the next day. After thawing, 130 µl avidin-beads (sigma A9207, pre-washed 3 times in 1x PBS) were added to each sample. Samples were incubated on a rotator for 1h at RT, then centrifuged for 7 min at 400 x g and RT. The supernatant was discarded and the beads were washed four times in 1% SDS, then two times in water, always by adding the liquid, inverting the tubes, centrifuging for 7 min at 400 x g and discarding the supernatant. Purified proteins on the beads were reduced in 256 µL 50 mM Tris pH 8, 8M Urea, 10 mM DTT for 15 min at 65°C with agitation in darkness, then cooled to 35°C and alkylated by adding 12.5 µL of 400 mM iodoacetamide in 50 mM Tris-HCl pH8 and incubating for 30 min with agitation in darkness. 4 ul Trypsin-LysC (Promega V5071, reconstituted according to the manufacturer's instructions) were added to each sample and LysC digest was performed for 3 h at 37 °C with agitation. Samples were then diluted in 50 mM Tris pH 8 to reach a final Urea concentration of less than 1 M and Trypsin digestion was

performed for 16 h at 37 °C with agitation. Samples were centrifuged for 3 min at 1000 x g and 0.1% (v/v) trifluoracetic acid were added to the supernatant in a new tube. The peptides were purified using Sep-Pak C18 cartridges (Waters, 610 Centennial Park, Herts, UK) according to the manufacturer's instructions. Protein LoBind tubes were used throughout.

Mass Spectrometry. Supplemental File ES5.02

Bioinformatics tools for leaf proteome analysis. Peptide spectra were annotated using Andromeda (Cox et al., 2011). Included modifications were carbamidomethylation (static) and oxidation, N-terminal acetylation and carbamylation of Lysines and N-termini (dynamic). Protein quantification was performed using MaxQuant version 1.5.5.30 (Tyanova et al., 2016a), including all modifications. Filtering and imputation of missing values using default settings were performed in Perseus (Tyanova et al., 2016b) and further data analysis carried out in R using the data.table and ggplot packages (Ihaka and Gentleman, 1996; Wickham, 2016). Sequence analyses and plasmid design were performed in Geneious (Kearse et al., 2012).

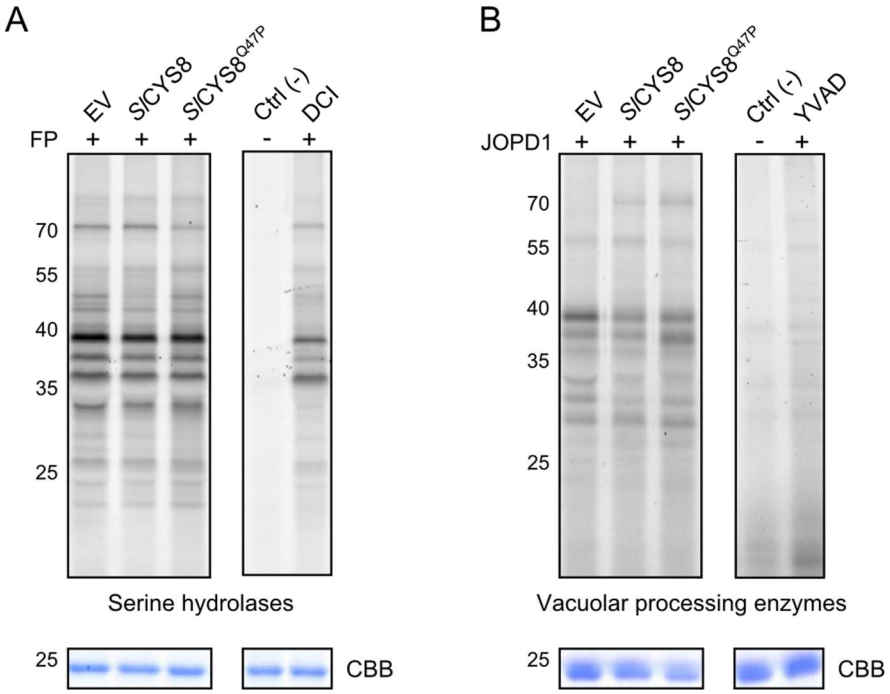
Statistical Analyses. Statistical analyses were performed using the RStudio program, v. 0.98.1103 (RStudio, Inc.). Analysis of variance (ANOVA) tests were used to compare in-gel band fluorescence intensity between treatments or ELISA values. Contrast calculations and Tukey's mean comparison tests were performed for those ANOVA giving significant *P* values at an alpha value threshold of 5%.

5.6 Acknowledgments

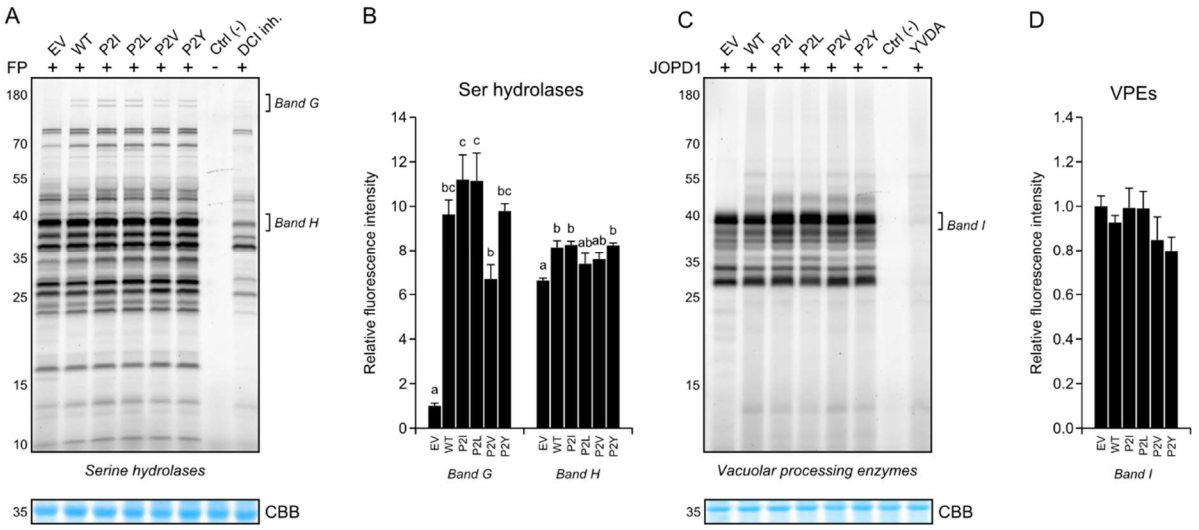
We thank Urszula Pyzio, Caroline O' Brien and Sarah Rodgers for excellent technical support. This work was financially supported by the ERC Project 'GreenProteases', University of Oxford (R.H., grant No. 616449), by Somerville College, Oxford (F.G.H. and R.H.), by an ERC starting grant (M.K., grant No. 258413), by Deutsche Forschungsgemeinschaft (M.K., grant no. INST 20876/127-1 FUGG), by a Discovery grant from the Natural Science and Engineering Research Council (NSERC) of Canada (D.M.) and by an AgroPhytoSciences NSERC–CREATE scholarship and of a BMP graduate scholarship jointly funded by Medicago inc., NSERC and Québec Government's research funding body FRQNT (P.V.J).

5.7 Supplementary information

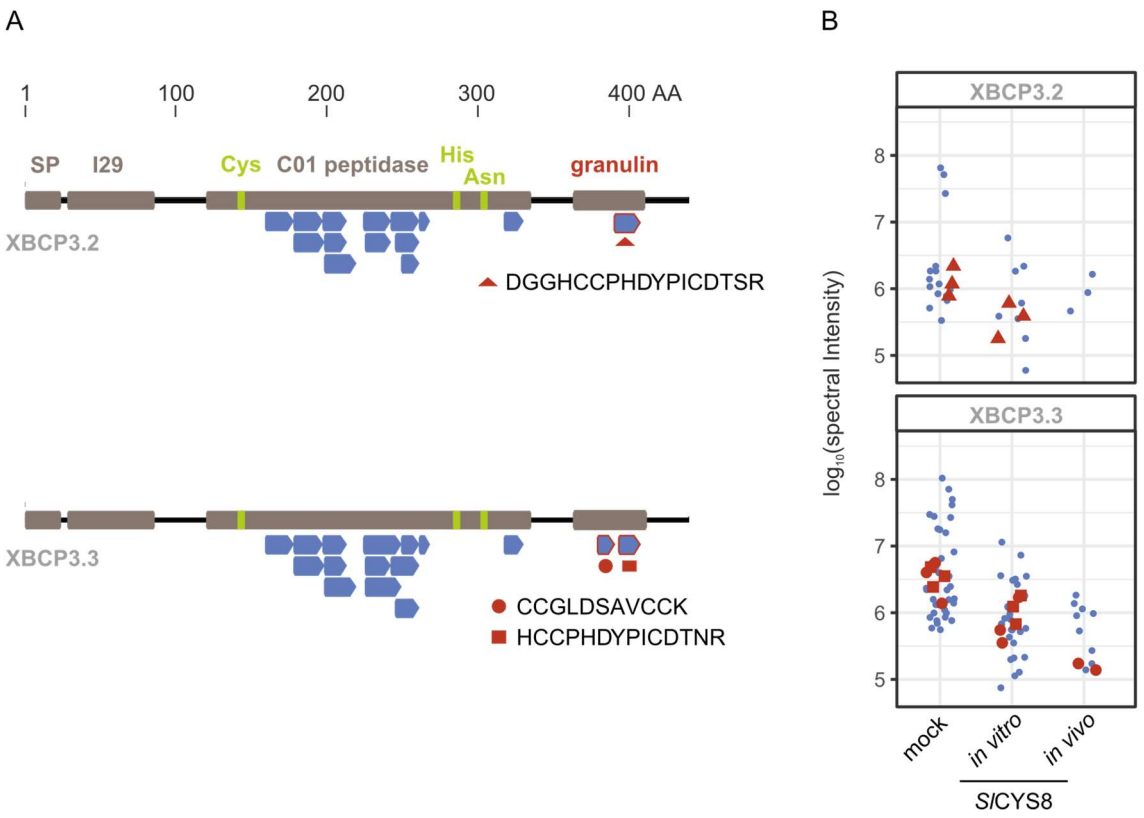
Supplemental figures



Supplemental Figure 5.S01. Complement to Figure 5.2 – Serine hydrolase and vacuolar processing enzyme activity profiles in apoplastic fluids of agroinfiltrated *N. benthamiana* leaves expressing SlCYS8**.** Apoplastic fluids from plants agroinfiltrated with the empty-vector control or expressing SlCYS8 (or the inactive mutant SlCYS8^{Q47P}) were labelled with FP or JOPD1 fluorescent probes to detect active serine hydrolases (A) or vacuolar processing enzymes (B), respectively. Ctrl (-): mix of plant extracts with no probe, DCI: mix of plant extracts pre-incubated with the chemical serine protease inhibitor 3,4-dichloroisocoumarin, YVAD: plant extracts pre-incubated with the chemical Caspase inhibitor I.



Supplemental Figure 5.S02. Complement to Figure 5.4 – Impact of SlCYS8 **single N-terminal mutations on SH and VPE activity in agroinfiltrated leaves.** (A) SlCYS8 mutants were transiently expressed in *N. benthamiana* and leaf total soluble proteins were extracted, prior to (A) serine hydrolase and (C) vacuolar processing enzyme activity labelling. Bands G to I were previously described in **Figure 5.2**. (B, D) Quantification of bands G to I fluorescence intensity. Each bar is the mean of three biological replicate values ± SE. Bars with different letters are significantly different (post-ANOVA Tukey's test; $P<0.05$). EV: extracts from plants agroinfiltrated with the empty-vector control, WT: wild-type SlCYS8, P2I: proline to isoleucine, P2L: proline to leucine, P2V: proline to valine, P2Y: proline to tyrosine, Ctrl (-): mix of plant extracts with no probe, DCI: mix of plant extracts pre-incubated with the chemical serine protease inhibitor 3,4-dichloroisocoumarin, YVAD: plant extracts pre-incubated with the chemical Caspase inhibitor I.



Supplemental Figure 5.S03. Complement to Figure 5.6 – Peptides corresponding to the granulin domain containing PLCPs XBCP3.2 and XBCP3.3. (A) Identified peptides (blue) from the granulin-domain containing PLCPs XBCP3.2 and XBCP3.3. SP: signal peptide, I29: I29 inhibitory domain, CO1: peptidase, Granulin: granulin domain. The catalytic triad residues are presented in green. (B) Spectral intensity of the peptides in three biological replicates. The granulin-derived peptides are shown in red. Mock: plants agroinfiltrated with the empty-vector control, SlCYS8 *in vitro*: agroinfiltrated plants pre-incubated with recombinant SlCYS8, SlCYS8 *in vivo*: agroinfiltrated plants expressing SlCYS8.

Electronic supplementary information

Electronic supplementary information included with the thesis: Files ES5.01-02.

Chapter 6

General discussion

6.1 Accompanying *N. benthamiana* on the way to becoming a model plant

An advantage of using model organisms compared to less well studied species for a given experiment is that for model organisms, there usually are data from other experiments using the same or similar techniques. These data can be used to estimate expected effect sizes and the number of replicates needed for significant results, thus helping to design the most informative experiments given limited resources. Once the experiment is performed, previous data can also aid the interpretation of results by providing a sense of what is expected from a well-designed, technically robust experiment. Calibrated in this way, one can then detect abnormalities pointing to either technical defects or interesting biological phenomena and potential ways to address biotechnological challenges.

Over the last years, *N. benthamiana* has become a better model for transcriptomics and proteomics because more data are available to inform experimental design and aid interpretation of results. The following sections review the contributions that this thesis has made to the body of knowledge on *N. benthamiana*. First, I summarize the proteomics and activity-based proteomics data gathered throughout this thesis on the Papain-like Cys Protease (PLCP) family. PLCPs have been studied intensively and will likely attract further interest because they can degrade RPs and are implicated in plant immunity, rendering a broad overview especially important. Second, I compare the transcriptome obtained in Chapter 2 to other transcriptomes obtained from *N. benthamiana* upon infection with different pathogens. I provide indications regarding both the number of transcripts usually detected in *N. benthamiana* leaves and the common, basal aspects of *N. benthamiana* immune responses. Third, I summarize the

proteome data obtained here and by others to provide indications regarding the size and plasticity of the *N. benthamiana* leaf proteome.

6.1.1 Defining the robust core PLCP repertoire of agroinfiltrated leaves

In Chapters 2, 4 and 5, we have monitored *N. benthamiana* PLCPs in agroinfiltrated leaves in four different proteomics datasets (Figure 6.1a). Dataset I is the 5 dpi time point from the time course of extracellular proteomes and dataset II is another extracellular proteome obtained at 5 dpi. Dataset III is an activity-based extracellular proteome obtained together with dataset II at 5 dpi (Chapter 2). Dataset IV is a full leaf proteome obtained to confirm the accumulation of overexpressed PIs. PIs were expressed in the presence of P19 using the GV3101 Agrobacteria strain and samples were taken at 4 dpi, the point of maximum accumulation of the overexpressed protein according to our experience with this strain (Chapter 4). Dataset V is an activity-based proteome of total leaf extracts obtained to identify the targets of SlCYS8. SlCYS8 was expressed in the presence of P19 using the AGL1 Agrobacteria strain (Lazo *et al.*, 1991) and samples were therefore taken at 6 dpi, which is the best harvesting time point for the AGL1 strain (Chapter 5). All five experiments should monitor different subsets of the same proteome. Extracellular proteomes represent a subset of the proteins present in whole leaf proteomes and may also contain secreted proteins that were too low in abundance in whole leaves to be identified, but are enriched in the extracellular space. Similarly, activity-based proteomics should monitor the subset of active enzymes in a proteome, but may also facilitate detection of less abundant, but active enzymes by enriching them. The five datasets reveal which PLCPs are robustly present and/or active, which are sometimes detectable and which seem absent in agroinfiltrated leaves. To define which of the 41 PLCPs in the predicted proteome

(Chapter 2) form the core PLCP repertoire in leaves, I evaluated which PLCPs were identified in at least two biological replicates and significantly enriched compared to the no-probe control (activity-based proteome datasets) or identified in all three biological replicates (total proteome datasets) using only the samples from agroinfiltrated leaves that were not expressing any transgene (datasets I-III) or P19 controls (datasets IV and V).

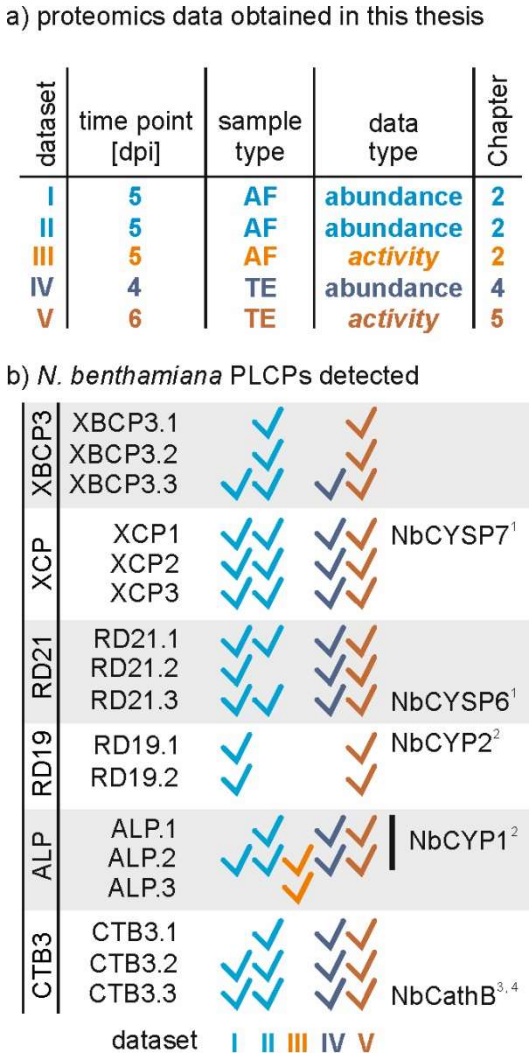


Figure 6.1: Papain-like Cys proteases (PLCPs) in agroinfiltrated *N. benthamiana* leaves detected across different MS experiments. a) overview of MS datasets. AF, apoplastic fluid (extracellular proteome); TE, total leaf extract; dpi, days post infiltration. b) *N. benthamiana* PLCP subfamilies, named according to the PLCP tree in Figure 2.5. Ticked proteins are part of a protein group (among the Majority Protein IDs) that was detected in all three replicates of agroinfiltrated leaves (proteome datasets, blue) or active in agroinfiltrated leaves (activity-based proteome datasets, orange). Previously studied enzymes are named according to: 1) Paireder et al, 2017; 2) Hao et al, 2006; 3) Gilroy et al, 2007; 4) Niemer et al, 2016. Complete Protein IDs are in Supplemental Figure S6.01, the complete comparison of MS datasets is in Supplemental File ES6.01.

16 *N. benthamiana* PLCPs were identified in at least two datasets, indicating they are robustly present in agroinfiltrated leaves (Figure 6.1b). This includes all 12 PLCPs identified in the whole leaf proteome (dataset IV). The same 12 are also identified in at least one of the extracellular proteomes (datasets I and II), in addition to four PLCPs from the RD19 (Resistant to Desiccation 19-like) and XBCP3 (Xylem bark Cys

protease 3-like) families. These four were only identified in the extracellular proteome or using activity-based proteomics, indicating they may be of low abundance, but active in leaves and enriched in the extracellular space. Surprisingly, we detected extracellular activity (dataset III) for two aleurain-like proteases (ALPs), but for none of the other 14 active PLCPs detected in full leaf extracts (dataset V). However, the active PLCPs from dataset V are mostly detected in the extracellular proteome (datasets I and II). The extracellular proteome thus either contains many inactive PLCPs or the active secretome experiment was technically suboptimal. The latter is likely, as other projects in the lab have identified >10 active PLCPs in extracellular proteomes from mock infiltrated *N. benthamiana* that were not identified in the mock samples from dataset III. Overall, extracellular proteomes likely contain additional, active PLCPs when compared to whole leaf proteomes (the PLCPs detected in datasets I, II and V but not IV). This is in accordance with higher proteolytic activity in extracellular proteomes compared to whole leaf proteomes (Delannoy *et al.*, 2008), suggesting proteases are enriched in the extracellular space.

We consistently detected NbCYSP6, NbCYSP7 and NbCathB, the *N. benthamiana* PLCPs that can degrade antibodies *in vitro* (Niemer *et al.*, 2014; Paireder *et al.*, 2016), indicating they contribute to RP degradation *in vivo.* In contrast, we never detected peptides or transcripts corresponding to members of the THI1 (THIAMINE4) subfamily. Absence of peptides and transcripts corresponding to THI1 proteins is expected, as THI1 activity is flower-specific in Arabidopsis (Richau *et al.*, 2012). We did not detect peptides corresponding to members of the CEP (Cys endopeptidase) subfamily at 4-6 dpi, although CEP-encoding transcripts were detected and extracellular peptides corresponding to CEP proteins appeared at later points in the extracellular proteome time course (Chapter 2, Figure 2.5). CEPs carry a C-terminal KDEL motif (Richau *et*

al., 2012), indicating they are retained in the ER and may not be extracted by our sample preparation method. At later time points upon agroinfiltration, they may be present in the extracellular space due to cell content leakage. In two other cases, the absence of peptides was unexpected. First, none of the members of the SAG12 (Senescence Associated Gene 12-like) family had corresponding peptides in any experiment. The SAG12 subfamily contains PIP1 and RCR3, which are major immune proteases in tomato, a relative of *N. benthamiana* (Krüger *et al.*, 2002; Tian *et al.*, 2007). We detected transcripts encoding NbPIP1 and NbRCR3 in leaves (Chapter 2, Figure 2.5), but the proteins do not seem to accumulate. Further analyses indicate that *NbPIP1* and *NbRCR3* are pseudogenes (J. Kourelis, unpublished) and the immune protease repertoires of *N. benthamiana* and tomato thus differ substantially. Second, the *N. benthamiana* predicted proteome contains two XBCPs encoded by the target genes of the hairpin RNA used to show that *NbC14* silencing enhances susceptibility of *N. benthamiana* to *P. infestans* (Kaschani *et al.*, 2010). NbC14-encoding transcripts were detected in leaves (Chapter 2, Figure 2.5). However, we did not detect peptides corresponding to these predicted NbC14 proteins in leaves, indicating that NbC14 proteins only accumulate under specific conditions. Alternatively, the other members of the XBCP3 subfamily that are detected in extracellular and total leaf proteomes may have been suppressed by *NbC14* silencing.

In summary, the core PLCP repertoire of agroinfiltrated leaves consists of ~ 16 PLCPs from the XBCP, XCP, RD19, RD21, ALP and CTB3 subfamilies, while no members of the SAG12, CEP and THI1-like subfamilies were detected in agroinfiltrated leaves.

6.1.2 Lessons learned from transcriptome studies of *N. benthamiana* immune responses

To provide an overview of typical results obtained from *N. benthamiana* leaf transcriptomes in the context of immune responses, I evaluated six recently published studies (Table 6.1). Two studies analysed leaves infected with hemibiotrophic pathogens, the fungus *Verticillium dahliae* and the oomycete *Phytophthora palmivora*, monitoring transcriptomes up to 10 dpi and 72 hours post infection (hpi), respectively (Faino *et al.*, 2012; Evangelisti *et al.*, 2017). Another two studies analysed the infection with viruses targeting Solanaceous crops or sugar beet, respectively, monitoring transcriptomes up to 10 or only at 12 dpi (Fan *et al.*, 2014; Geng *et al.*, 2017). A study investigating the effect of transcription factor overexpression by agroinfiltration at three dpi also contained empty vector and mock infiltration controls, which are comparable to our transcriptome described in Chapter 2 (Bond *et al.*, 2016).

The number of transcripts detected in each study depends on the transcriptome database used. Accordingly, studies using the relatively large transcriptome database based on the *N. benthamiana* genome (Faino *et al.*, 2012; Geng *et al.*, 2017; Grosse-Holz *et al.*, 2017) consistently detect more transcripts than those that consolidated transcripts into unigenes (Fan *et al.*, 2014; Evangelisti *et al.*, 2017) or used the smaller Nbv.5 database (Bond *et al.*, 2016). A smaller database increases the power of the study, because less statistical tests are carried out if less transcripts are potentially differential, alleviating the impact of multiple testing corrections. *N. benthamiana* leaf transcriptomes seem to be relatively stable, with < 15% of transcripts changing in abundance when a single time point is considered (Faino *et al.*, 2012; Fan *et al.*, 2014; Bond *et al.*, 2016; Geng *et al.*, 2017) and up to 25 % during time courses (Evangelisti *et al.*, 2017; Grosse-Holz *et al.*, 2017). In future, the values given in Table 6.1 can be

used to compute sample sizes needed for envisaged studies. Easily usable online tools now carry out such calculations (Ching *et al.*, 2014; Blaise *et al.*, 2016).

study setup	detected transcripts	differential transcripts	threshold FDR	threshold fold change	% of transcripts differential	timepoint(s)	reference
infection with the hemibiotrophic fungus *Verticillium dahliae*	73,041	5,596	0.05	-	7.7	4, 8, 12 and 16 dpi	(Faino et al., 2012)
infection with Beet Necrotic Yellow Vein Virus containing RNA4	27,890	2,683	0.05	-	9.6	12 dpi	(Fan et al., 2014)
infiltrated with *Agrobacterium tumefaciens*	~50,000[1]	2,220	0.001	-	4.4	3 dpi	(Bond et al., 2016)
agroinfiltration with Tobacco vein banding mosaic virus, compared to agroinfiltrated leaves	~57,000[2]	7,225	0.05	-	12.7	1, 2 and 10 dpi	(Geng et al., 2017)
infection with the hemibiotrophic oomycete *Phytophthora palmivora*, differentials at any point during time course	28,496	4,132	0.001	2	14.5	6, 18, 24, 30, 48 and 72 hpi	(Evangelisti et al., 2017)
infiltration with *Agrobacterium tumefaciens*, differentials at any point during time course	75,802	18,648	0.05	2	24.6	2, 5, 7 and 10 dpi	(Grosse-Holz et al., 2017) Chapter 2

Table 6.1: Transcriptomes obtained from *N. benthamiana* leaves. Numbers of differentials are given for infected vs mock treated leaves, unless specified otherwise. The 10 dpi TVBMV vs mock comparison is reported for (Geng *et al.*, 2017) and BN34 vs mock for (Fan *et al.*, 2014). dpi/hpi, days/hours post infection; FDR, false discovery rate. [1, 2] the authors did not specify the number of detected transcripts, sizes of the respective transcriptome databases are given as an estimate.

As the six studies used five different pathogens, they also shed light on the most basal features of the *N. benthamiana* immune response. Decreased transcript levels for the photosynthetic machinery that we detected in Chapter 2 are also detected upon both oomycete and viral infection (Fan *et al.*, 2014; Evangelisti *et al.*, 2017; Geng *et al.*, 2017). Interestingly, the degree to which photosynthesis is shut down is associated with the severity of viral infection. Beet Necrotic Yellow Vein Virus (BN3) is more virulent when it contains RNA4 (BN34), which encodes a protein localizing to the host nucleus (p31) (Rahim *et al.*, 2007; Fan *et al.*, 2014). Transcripts encoding proteins associated with photosynthesis are depleted upon infection with BN34, but not BN3. Therefore, p31 either directly or indirectly downregulates expression of the genes encoding the photosynthetic machinery (Fan *et al.*, 2014). The same pattern was observed for tobacco vein banding mosaic virus (TVBMV) and a mutant carrying a non-functional version of silencing suppressor HcPro (TVBMV-*HcPro*). Infection with TVBMV causes severe symptoms and a depletion of transcripts encoding proteins associated with photosynthesis by 10 dpi, while infection with TVBMV-*HcPro* is less severe and does not entail a shutdown of photosynthesis (Geng *et al.*, 2017). Both studies of agro- vs mock infiltrated leaves also report depletion of transcripts encoding proteins associated with photosynthesis, in accordance with the chlorotic phenotype of agroinfiltrated leaves. Furthermore, agroinfiltration resulted in increased levels of transcripts encoding proteins associated with defence, such as LRRs (leucine-rich-repeat proteins), NLRs (nucleotide-binding oligomerization domain-like receptors) and WRKY transcription factors (Bond *et al.*, 2016; Grosse-Holz *et al.*, 2017). In summary, shutdown of the photosynthetic machinery is likely to be part of the basal plant immune response, as it occurs upon infection with a fungus, an oomycete, *A. tumefaciens* and plant viruses.

6.1.3 Sizing up the *N. benthamiana* proteome

While transcriptome studies use *N. benthamiana* quite routinely now, shotgun proteomics data are still sparse. A few studies used two-dimensional gel electrophoresis and analysed gel spots by mass spectrometry (Pérez-Bueno *et al.*, 2004; Goulet *et al.*, 2010b; Villela-Dias *et al.*, 2014), but only one study published outside of this thesis analysed the whole leaf proteome (Du *et al.*, 2017). In that study, the elicitin INF-1 from *Phytophthora infestans* was overexpressed by agroinfiltration. INF-1 is recognized *in planta* and triggers cell death by 4 dpi, so the authors chose to analyse a much earlier time point to capture signalling events (8 hours post agroinfiltration) (Du *et al.*, 2017). Interestingly, proteins associated with photosynthesis are nonetheless depleted, indicating that photosynthesis is shut down during immune responses by both downregulation of gene expression and degradation of the photosynthetic machinery. The extracellular proteome data described in Chapter 2 show an immune response upon agroinfiltration that is associated with increased efficiency of extracellular delivery. Datasets I and II are replicates of each other, with the important distinction that a time course was sampled in dataset I, while dataset II shows a single time point. In Chapter 4, we characterized the changes in complete proteomes of agroinfiltrated leaves upon overexpression of four different PIs. I here show the data for NbPR4.

study setup	detected proteins	differential proteins	threshold FDR	threshold fold change	% of proteins differential	reference
overexpressing INF1 elicitin from *Phytophthora infestans* (compared to empty vector agroinfiltrated leaves)	2,964	35	0.05	1.2	1.2	(Du et al., 2017)
agroinfiltrated, differentials at any point during time course (extracellular proteome I)	2,233	1,697	0.05	2	76.0	(Grosse-Holz et al., 2017) Chapter 2
agroinfiltrated, differentials at 5 dpi (extracellular proteome II)	2,147	399	0.05	2	18.6	(Grosse-Holz et al., 2017) Chapter 2
overexpressing NbPR4 (compared to P19 agroinfiltrated leaves)	3,349	56	0.05	2	1.7	Chapter 4

Table 6.2: Proteomes obtained from *N. benthamiana* leaves. Numbers of differentials are given for infected vs mock treated leaves, unless specified otherwise.

Together, shotgun proteomics datasets of *N. benthamiana* indicate that ~ 3000 proteins are usually detected in whole leaves. If the treatment consists in overexpression of a single protein, less than 5 % of detected proteins change in abundance (Table 6.2). If the extracellular sub-proteome is analysed, changes are much more drastic, likely because protein levels depend on extracellular delivery in addition to transcription, translation and turnover. Especially during a time course, the proteome can thus expand massively (Chapter 2). For design of future studies, this implies that where overexpression of a single protein is to be studied in whole leaf proteomes, a sufficient number of replicates needs to be generated to detect the relatively subtle effects expected. Conversely, where a sub-proteome is analysed, the effect size can be expected to be larger and less replicates may be sufficient.

In context with the literature, the data from this thesis provides useful indications for experimental design and reveals commonalities in immune responses to a range of pathogens. *N. benthamiana* has thus advanced considerably on the path towards becoming a model organism in recent years, partly through our efforts. As omics level datasets for *N. benthamiana* become more abundant, true meta-analyses will be possible. Analogous to the core PLCP repertoire of agroinfiltrated leaves, a core defence machinery could then be defined in *N. benthamiana* by combining omics data on different immune responses.

6.2 The future of molecular farming

6.2.1 Resolving the degradation bottleneck with a combined effort

Given the dynamic protease network responsible for RP degradation, it is likely that combinations of RP stabilizing tools will be needed to enhance RP accumulation to the level required for commercial viability. I here discuss how we combined PIs, how the full potential of PIs may be exploited through more sophisticated ways of combining them and how PI co-expression could be interlinked with other measures to boost RP levels.

In Chapter 4, we provide three novel PIs that enhance RP accumulation upon co-expression and combine them with each other and with SlCYS8. NbPR4, NbPot1 and HsTIMP can act together to enhance RP accumulation, but each PI has a dose-dependent effect: A higher proportion of Agrobacteria delivering PI-encoding T-DNA in the agroinfiltration mixture leads to higher RP accumulation. It thus seems that PI expression levels are insufficient for each PI to reach its full potential in blocking protease activity.

This bottleneck might be overcome by co-expressing the mix with silencing suppressor P19 to ensure high transcript levels (van der Hoorn *et al.*, 2003) or by using viral genome elements that lead to hypertranslation (Sainsbury & Lomonossoff, 2008; Sainsbury *et al.*, 2009). Transgene replication *in planta*, achieved by flanking the transgene with viral genome elements to generate a replicating vector, could also help deliver sufficient amounts of transgene. However, co-expressing multiple proteins is then a challenge, as different viral vectors compete with each other for replication (Gleba *et al.*, 2005). Non-competing viral vector systems exist (Giritch *et al.*, 2006), but would have to be adapted to co-expression of multiple PIs with an AB. To

circumvent competition issues and to decrease total gene expression cost, several PIs could be combined into engineered multidomain proteins. Multidomain PIs have been generated to inhibit the gut proteases of plant parasitic nematodes and insects (Urwin *et al.*, 1998; Outchkourov *et al.*, 2004). A critical point in the design of multidomain PIs is the choice of the right peptide linker. NbPR4, NbPot1 and HsTIMP lost their RP accumulation-enhancing function when expressed with fusion tags in preliminary experiments. We thus assume that N- and C-termini of the PIs are important for protease interaction, indicating that a cleavable linker should be envisaged. Self-cleaving linkers have recently been developed from inteins, protein domains that post-translationally remove themselves from a peptide chain without any remainder (Zhang *et al.*, 2017). However, conformationally flexible glycine linkers that do not remove themselves can also enhance stability of the fusion protein (Outchkourov *et al.*, 2004), indicating some experimentation may be warranted. An additional strategy to avoid competition between different transgenes for delivery and/or expression is to generate stable plant lines expressing a PI that can then be agroinfiltrated to produce RPs (Pillay *et al.*, 2012). To minimize the PI gene expression cost in stable transgenic lines, PIs or multi-PI fusion proteins could be expressed under promoters that are known to be induced upon agroinfiltration. Such promoters may be found in the genes whose corresponding transcripts first increased in abundance at 2 dpi and were highly abundant in our transcriptome dataset (Chapter 2, see supplementary table ES2.04).

Along with PIs that can be combined with each other, the molecular farming community has different tools at hand that could jointly protect RPs from proteolysis. Genetic depletion of proteases may be part of the design for optimized molecular farming plant lines. Combined depletion of four proteases (one Asp, Ser, Cys and metalloprotease each) by gene silencing in tobacco cell suspension cultures enhanced

the accumulation of the anti-HIV antibody 2F5 fourfold (Mandal *et al.*, 2014). Multiple proteases can thus be depleted without fitness costs that negatively impact RP production. Obvious targets are the PLCPs that SlCYS8 inhibits (Chapter 5) and abundant, secreted aspartic proteases for which we do not currently have PIs. Preliminary gene silencing experiments could indicate whether depletion of these enzymes can enhance RP accumulation before protease knockout lines are generated. A more general approach is to modify the cellular environment and create suboptimal conditions for proteases. For instance, the pH of the golgi has been increased by expression of the influenza ion channel M2 in agroinfiltrated *N. benthamiana* and co-expression with M2 increased accumulation of an acid-susceptible fusion protein (Jutras *et al.*, 2015). This approach could be combined with PIs targeting the remainder of the protease repertoire that is still active at higher pH, such as some subtilases (Rawlings, 2016). Antibodies can be engineered to remove the sites most susceptible to protease cleavage (Zischewski *et al.*, 2016; Hehle *et al.*, 2016) and these stabilized antibodies could be combined with PIs to block remaining degradation.

Beyond tackling protease activity and proteolytic degradation directly, RP yields have also been improved by increasing plant growth rates and decreasing proteome complexity. Although most plant growth optimization probably happens behind the intellectual property barriers of molecular farming companies, some data on adjusting humidity, light and temperature have been published (Fujiuchi *et al.*, 2016). To simplify downstream processing, plants have been treated with the defence hormone methyl-jasmonate (MeJA) to reduce rubisco levels (Robert *et al.*, 2015). This is counterintuitive at first, as plant immune responses are known to increase protease activity (Shabab *et al.*, 2008; Esse *et al.*, 2008; Sueldo *et al.*, 2014). MeJA treatment

would thus be expected to hamper RP stability. However, MeJA treatment did not drastically increase overall proteolytic activity, in accordance with our data that show no drastic protease upregulation during the immune response upon agroinfiltration in *N. benthamiana* (Grosse-Holz *et al.*, 2017). Decreasing proteome complexity by timely depletion of rubisco can thus increase the yield of purified RPs.

In summary, a wide variety of tools to increase RP yields are now available and their potential has separately been demonstrated. In the right combination, these tools will likely resolve the degradation bottleneck and boost RP accumulation beyond the critical threshold of economic viability.

6.2.2 Molecular farming and global health

Commercialisation of molecular farming using agroinfiltrated *N. benthamiana* as a fast and flexible biopharmaceutical production platform has progressed in recent years. In the US, Kentucky BioProcessing (Kentucky BioProcessing Inc, kentuckybioprocessing.com, Owensboro KY, USA) manufactured the antibodies of the Zmapp cocktail that have been used successfully to treat Ebola in primates and are being tested in humans (Qiu *et al.*, 2014; ClinicalTrials.gov NCT02363322, 2017; ClinicalTrials.gov NCT02389192, 2017). The Canadian company Medicago (Medicago Inc, Quebec, Canada; medicago.com) is funded by the US Defense Advanced Research Projects Agency (DARPA) to develop pandemic response strategies (DARPA, 2012). In Europe, Icon Genetics (Icon Genetics GmbH, Halle/Saale, Germany; icongenetics.com) holds a family of patents covering the virus-based transient expression system magnICON® and has collaborated with university research groups to produce biopharmaceuticals for clinical studies (Tusé *et al.*, 2015).

The British company Leaf Expression Systems (Leaf Expression Systems International Ltd, Norwich, UK; leafexpressionsystems.co.uk) has won several grants to develop and produce different biopharmaceuticals together with university research groups. In summary, a robust agroinfiltration-based molecular farming industry is growing in countries of the global north.

However, a prevailing claim in the molecular farming community is that plant expression platforms will help improve global health, specifically access to vaccines and treatments in developing countries. This claim is based on three assumptions. The first assumption is that molecular farming will make biopharmaceuticals cheaper and thus charities or government programmes will be able to treat more people. The second assumption is that vaccines could be produced in edible plant parts, which can be minimally processed, stored at high temperatures for a long time and then confer immunity upon ingestion (Pascual, 2007). Such edible vaccines would simplify the workflow of charities or government programmes, which would in turn reach more people. The third assumption is that plant-based biopharmaceutical production would be easy to implement locally, fostering the independence of developing countries from international aid programmes (Rybicki *et al.*, 2013). I will briefly discuss how progress in recent years has shed light on the prospects of molecular farming to improve global health.

The potential cost advantage of molecular farming critically depends on resolving the degradation bottleneck. As outlined above, the tools presented in this thesis and elsewhere will likely resolve this bottleneck if combined in the right way. However, for decreased manufacturing costs to translate into improvements in global health, cheaper medicines must actually be an effective way to improve global health. Organisations that very cost-effectively further global health according to the charity

evaluator GiveWell (www.givewell.org) report spending ~ 1/4 of their annual budget on the medicines they use (GiveWell, 2017a,b), with the remainder covering labour, facilities and other material costs. Cheaper medicines might thus help reach more people. However, decreasing the cost of the logistics needed to administer vaccines or treatments is likely to have a bigger impact on the number of people reached by the most effective global health charities, which here serve as a proxy for the sector.

Edible vaccines could simplify the logistics of global health projects, as they do not depend on the cold chains and sterile injection by medical personnel that make vaccination campaigns difficult and expensive. Indeed, animals have been immunized against enterotoxic *E. coli* and Norwalk viruses by consuming dried maize seeds or tomato fruit expressing an antigen from the respective pathogen (subunit vaccine) (Streatfield *et al.*, 2002; Zhang *et al.*, 2006). These results are promising, but no edible vaccine has yet reached clinical trials in humans. This lag in development seems to be due to strict regulations regarding dosage control. Delivering a defined dose in minimally processed plant tissue is a challenge because RP accumulation levels in plant tissues vary too much (Arntzen, 2015). Dosing issues might be overcome if the degradation bottleneck is resolved and RP levels become more stable as a result. However, the benefits of edible vaccines would then likely be limited to some pathogens that infect humans via mucosal tissues. Protection from many pathogens requires immune responses in the blood and lymph systems throughout the body (systemic immunity). Edible vaccines elicit mucosal immune responses in the intestine and must be combined with injected “boosters” to achieve systemic immunity. Animals have been protected from the pathogens *Plasmodium falciparum* (malaria), *Vibrio cholerae* (cholera) and *Yersinia pestis* (plague) using subunit vaccines produced in transgenic chloroplasts of tobacco. However, combination with an injectable booster

was required in all cases (Arlen *et al.*, 2008; Davoodi-Semiromi *et al.*, 2010). In summary, it may be possible to make edible vaccines in controlled doses if RP levels are stabilized in crop plants. Edible vaccines could then confer mucosal immunity alone and systemic immunity in combination with injectable boosters. In many cases, cold chains and needles would thus still be needed.

Progress towards local implementation of molecular farming has been made in Brazil, South Africa and Argentina. Research groups in these regions have developed plant-produced vaccines against viral infections for both humans and livestock (Rybicki *et al.*, 2013). Multiple candidate vaccines required by the regional market are thus in pre-clinical stages of the drug development pipeline. The production capacities needed for clinical trials and commercialisation are following suit, with public-private partnerships ongoing in Brazil and envisaged in South Africa. In Brazil, the molecular farming platform provider iBioInc (New York, US) has agreed with both state authorities and the local company Fiocruz/Bio-Manguinhos (Rio de Janeiro, Brazil) to build a complete agroinfiltration-based facility and produce a yellow fever vaccine (IBIO, 2017a). A similar partnership is being established in South Africa, with exploratory negotiations under way (IBIO, 2017b). In summary, empowerment of developing nations through molecular farming seems possible. To determine whether transferring molecular farming technologies is a particularly effective way to foster local healthcare and science, however, the outcomes would need to be compared to projects that pursue the goal of empowerment by other means.

In summary, technology transfer of molecular farming is progressing and promises to empower developing nations, which may entail positive side effects on global health. However, direct global health interventions seem to be limited by the cost of logistics rather than the cost of the medicines they use and edible vaccines will not abolish the

need for cold chains and needles because they only confer mucosal immunity unless used with an injectable booster. Therefore, if one seeks to improve global health and is agnostic concerning the means of doing so, there are likely more promising opportunities than developing molecular farming. The technology nonetheless has its place in a diverse portfolio of epidemic response strategies, as outlined in Chapter 1.

.

6.3 Conclusions

This work stands at an interface between applied and basic plant science. I uncovered a large, diverse repertoire of proteases and then addressed this complexity using PI overexpression. By blocking the activity of presumably many different proteases, I succeeded in enhancing RP accumulation. Remarkably, my results indicate that a redundant protease network degrades RPs and degradation is, to some extent, a universal process across different RPs. I have uncovered some of the nodes in the RP degrading protease network by identifying the targets of SlCYS8. Identification of the targets of the other PIs will further enhance our understanding of how and where RPs are degraded.

6.4 Supplementary information

a) proteomics data obtained in this thesis

dataset	time point [dpi]	sample type	data type	Chapter
I	5	AF	abundance	2
II	5	AF	abundance	2
III	5	AF	*activity*	2
IV	4	TE	abundance	4
V	6	TE	*activity*	5

b) IDs of detected *N. benthamiana* PLCPs

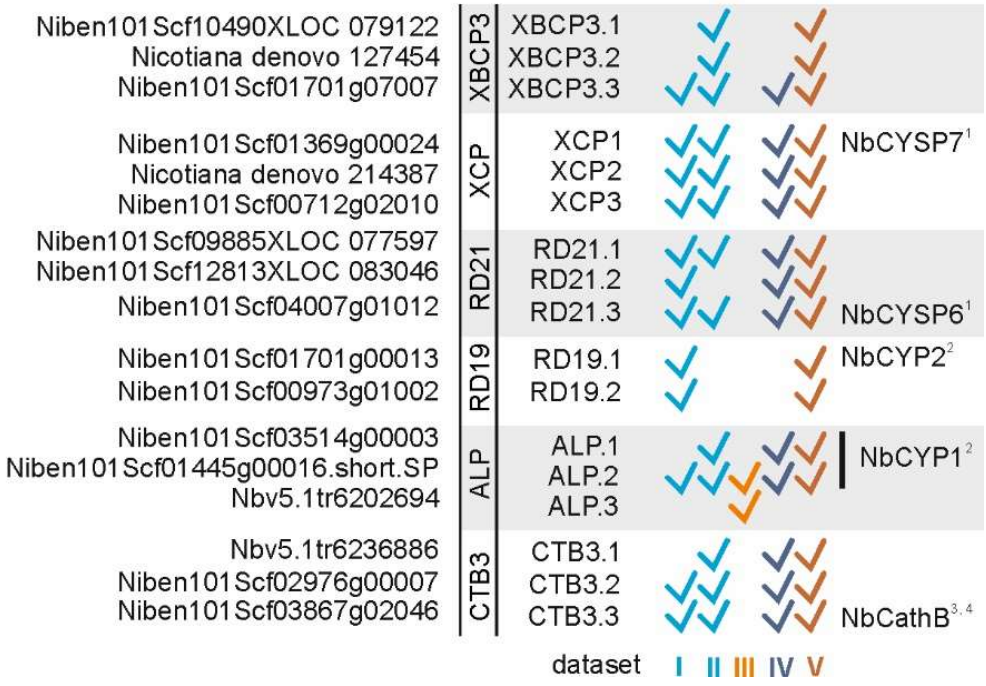


Figure 6S.01 Papain-like Cys proteases (PLCPs) in agroinfiltrated *N. benthamiana* leaves detected across different MS experiments. a) overview of MS datasets. AF, apoplastic fluid (extracellular proteome); TE, total leaf extract; dpi, days post infiltration. b) *N. benthamiana* PLCP subfamilies, named according to the PLCP tree in Figure 2.5. Ticked proteins are part of a protein group (among the Majority Protein IDs) that was detected in all three replicates of agroinfiltrated leaves (proteome datasets, blue) or active in agroinfiltrated leaves (activity-based proteome datasets, orange). Previously studied enzymes are named according to: 1) Paireder et al, 2017; 2) Hao et al, 2006; 3) Gilroy et al, 2007; 4) Niemer et al, 2016. The complete comparison of MS datasets is in Supplemental File ES6.01.

Electronic supplementary information

Electronic supplementary information included with the thesis: File ES6.01

Literature cited

Abdul-Hussain S, Paulsen GM. **1989**. Role of proteinaceous α-amylase enzyme inhibitors in preharvest sprouting of wheat grain. *Journal of Agricultural and Food Chemistry* **37**: 295–299.

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, *et al.* 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research* **44**: W3–W10.

Ahlquist P, French R, Janda M, Loesch-Fries LS. **1984**. Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proceedings of the National Academy of Sciences* **81**: 7066–7070.

Akers CP, Hoff JE. **1980**. Simultaneous formation of chymopapain inhibitor activity and cubical crystals in tomato leaves. *Canadian Journal of Botany* **58**: 1000–1003.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Alvarez-Fernandez M, Barrett AJ, Gerhartz B, Dando PM, Ni J, Abrahamson M. **1999**. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *Journal of Biological Chemistry* **274**: 19195–19203.

Amrani AE, Barakate A, Askari BM, Li X, Roberts AG, Ryan MD, Halpin C. **2004**. Coordinate expression and independent subcellular targeting of multiple proteins from a single transgene. *Plant Physiology* **135**: 16–24.

Arkadash V, Yosef G, Shirian J, Cohen I, Horev Y, Grossman M, Sagi I, Radisky ES, Shifman JM, Papo N. **2017**. Development of high affinity and high specificity inhibitors of Matrix Metalloproteinase 14 through computational design and directed evolution. *Journal of Biological Chemistry* **292**: 3481–3495.

Arlen PA, Singleton M, Adamovicz JJ, Ding Y, Davoodi-Semiromi A, Daniell H. **2008**. Effective plague vaccination via oral delivery of plant cells expressing F1-V antigens in chloroplasts. *Infection and Immunity* **76**: 3640–3650.

Arntzen C. **2015**. Plant-made pharmaceuticals: from 'Edible Vaccines' to Ebola therapeutics. *Plant Biotechnology Journal* **13**: 1013–1016.

Atkinson AH, Heath RL, Simpson RJ, Clarke AE, Anderson MA. **1993**. Proteinase inhibitors in Nicotiana alata stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *The Plant Cell* **5**: 203–213.

Auldridge ME, Guo Y, Austin MB, Ramsey J, Fridman E, Pichersky E, Noel JP. **2012**. Emergent decarboxylase activity and attenuation of α/β-hydrolase activity during the evolution of methylketone biosynthesis in tomato. *The Plant Cell* **24**: 1596–1607.

Azarkan M, Martinez-Rodriguez S, Buts L, Baeyens-Volant D, Garcia-Pino A. **2011**. The plasticity of the β-trefoil fold constitutes an evolutionary platform for protease inhibition. *Journal of Biological Chemistry* **286**: 43726–43734.

Bally J, Nakasugi K, Jia F, Jung H, Ho SYW, Wong M, Paul CM, Naim F, Wood CC, Crowhurst RN, *et al.* 2015. The extremophile Nicotiana benthamiana has traded viral defence for early vigour. *Nature Plants* **1**: 15165.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, *et al.* 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**: D991–D995.

Barrette-Ng IH, Ng KK-S, Cherney MM, Pearce G, Ghani U, Ryan CA, James MNG. **2003**. Unbound form of tomato inhibitor-II reveals interdomain flexibility and conformational variability in the reactive site loops. *Journal of Biological Chemistry* **278**: 31391–31400.

Barta A, Sommergruber K, Thompson D, Hartmuth K, Matzke MA, Matzke AJM. **1986**. The expression of a nopaline synthase — human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. *Plant Molecular Biology* **6**: 347–357.

Beadle GW, Tatum EL. **1941**. Genetic control of biochemical reactions in Neurospora. *Proceedings of the National Academy of Sciences of the United States of America* **27**: 499–506.

Bellincampi D, Camardella L, Delcour JA, Desseaux V, D'Ovidio R, Durand A, Elliot G, Gebruers K, Giovane A, Juge N, *et al.* 2004. Potential physiological role of plant glycosidase inhibitors. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1696**: 265–274.

Benchabane M, Goulet C, Rivard D, Faye L, Gomord V, Michaud D. **2008**. Preventing unintended proteolysis in plant protein biofactories. *Plant Biotechnology Journal* **6**: 633–648.

Benchabane M, Schlüter U, Vorster J, Goulet M-C, Michaud D. **2010**. Plant cystatins. *Biochimie* **92**: 1657–1666.

Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E. **2015**. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *genesis* **53**: 474–485.

Bevan M. **1984**. Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Research* **12**: 8711–8721.

Bhattacharjee L, Singh D, Gautam JK, Nandi AK. **2017**. Arabidopsis thaliana serpins AtSRP4 and AtSRP5 negatively regulate stress-induced cell death and effector-triggered immunity induced by bacterial effector AvrRpt2. *Physiologia Plantarum*: 329–339.

Bhushan S, Ståhl A, Nilsson S, Lefebvre B, Seki M, Roth C, McWilliam D, Wright SJ, Liberles DA, Shinozaki K, *et al.* 2005. Catalysis, subcellular localization, expression and evolution of the targeting peptides degrading protease, AtPreP2. *Plant & Cell Physiology* **46**: 985–996.

Bienert MD, Delannoy M, Navarre C, Boutry M. **2012**. NtSCP1 from tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. *Plant Physiology* **158**: 1220–1229.

Birk Y. **1961**. Purification and some properties of a highly active inhibitor of trypsin and α-chymotrypsin from soybeans. *Biochimica et Biophysica Acta* **54**: 378–381.

Blaise BJ, Correia G, Tin A, Young JH, Vergnaud A-C, Lewis M, Pearce JTM, Elliott P, Nicholson JK, Holmes E, *et al.* 2016. Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Analytical Chemistry* **88**: 5179–5188.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. **2011**. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bollhöner B, Zhang B, Stael S, Denancé N, Overmyer K, Goffner D, Van Breusegem F, Tuominen H. **2013**. Post mortem function of AtMC9 in xylem vessel elements. *New Phytologist* **200**: 498–510.

Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB. **2012**. A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research. *Molecular Plant-Microbe Interactions* **25**: 1523–1530.

Bond DM, Albert NW, Lee RH, Gillard GB, Brown CM, Hellens RP, Macknight RC. **2016**. Infiltration-RNAseq: transcriptome profiling of Agrobacterium-mediated infiltration of transcription factors to discover gene function and expression networks in plants. *Plant Methods* **12**: 41.

Book AJ, Gladman NP, Lee S-S, Scalf M, Smith LM, Vierstra RD. **2010**. Affinity purification of the Arabidopsis 26 S proteasome reveals a diverse array of plant proteolytic complexes. *Journal of Biological Chemistry* **285**: 25554–25569.

Book AJ, Yang P, Scalf M, Smith LM, Vierstra RD. **2005**. Tripeptidyl peptidase II. An oligomeric protease complex from Arabidopsis. *Plant Physiology* **138**: 1046–1057.

Bos JIB, Kanneganti T-D, Young C, Cakir C, Huitema E, Win J, Armstrong MR, Birch PRJ, Kamoun S. **2006**. The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana. *The Plant Journal* **48**: 165–176.

Bowman DE. **1946**. Differentiation of soybean antitryptic factors. *Proceedings of the Society for Experimental Biology and Medicine* **63**: 547–550.

Bozkurt TO, Schornack S, Win J, Shindo T, Ilyas M, Oliva R, Cano LM, Jones AME, Huitema E, van der Hoorn RAL, *et al.* 2011. Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 20832–20837.

Braun AC. **1943**. Studies on Tumor Inception in the Crown-Gall Disease. *American Journal of Botany* **30**: 674–677.

Breitenbach HH, Wenig M, Wittek F, Jorda L, Maldonado-Alconada AM, Sarioglu H, Colby T, Knappe C, Bichlmeier M, Pabst E, *et al.* 2014. Contrasting roles of apoplastic aspartyl protease AED1 and legume lectin-like protein LLP1 in Arabidopsis systemic acquired resistance. *Plant Physiology* **165**: 791–809.

Brencic A, Eberhard A, Winans SC. **2004**. Signal quenching, detoxification and mineralization of vir gene-inducing phenolics by the VirH2 protein of Agrobacterium tumefaciens. *Molecular Microbiology* **51**: 1103–1115.

Brutus A, Reca IB, Herga S, Mattei B, Puigserver A, Chaix J-C, Juge N, Bellincampi D, Giardina T. **2005**. A family 11 xylanase from the pathogen Botrytis cinerea is inhibited by plant endoxylanase inhibitors XIP-I and TAXI-I. *Biochemical and Biophysical Research Communications* **337**: 160–166.

BTI. **2017**. Sol Genomics Network, Boyce Thompson Institute. *https://solgenomics.net/organism/Nicotiana_benthamiana/genome*.

Burman R, Gunasekera S, Strömstedt AA, Göransson U. **2014**. Chemistry and biology of cyclotides: circular plant peptides outside the box. *Journal of Natural Products* **77**: 724–736.

Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. **2008**. ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Research* **18**: 810–820.

Buyel JF, Buyel JJ, Haase C, Fischer R. **2015a**. The impact of Pseudomonas syringae type III effectors on transient protein expression in tobacco. *Plant Biology* **17**: 484–492.

Buyel JF, Twyman RM, Fischer R. **2015b**. Extraction and downstream processing of plant-derived recombinant proteins. *Biotechnology Advances* **33**: 902–913.

Campillo T, Renoud S, Kerzaon I, Vial L, Baude J, Gaillard V, Bellvert F, Chamignon C, Comte G, Nesme X, *et al.* 2014. Analysis of hydroxycinnamic acid degradation in Agrobacterium fabrum reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. *Applied and Environmental Microbiology* **80**: 3341–3349.

Casamitjana-Martínez E, Hofhuis HF, Xu J, Liu C-M, Heidstra R, Scheres B. **2003**. Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of Arabidopsis root meristem maintenance. *Current Biology* **13**: 1435–1441.

Castilho A, Windwarder M, Gattinger P, Mach L, Strasser R, Altmann F, Steinkellner H. **2014**. Proteolytic and N-glycan processing of human α1-antitrypsin expressed in Nicotiana benthamiana. *Plant Physiology* **166**: 1839–1851.

Chapman EJ, Prokhnevsky AI, Gopinath K, Dolja VV, Carrington JC. **2004**. Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes & Development* **18**: 1179–1186.

Charmont S, Jamet E, Pont-Lezica R, Canut H. **2005**. Proteomic analysis of secreted proteins from Arabidopsis thaliana seedlings: improved recovery following removal of phenolic compounds. *Phytochemistry* **66**: 453–461.

Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A. **1989**. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**: 1576–1584.

Chen L-Q. **2014**. SWEET sugar transporters for phloem transport and pathogen nutrition. *New Phytologist* **201**: 1150–1155.

Chen P, Rose J, Love R, Wei CH, Wang B-C. **1992**. Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *Journal of Biological Chemistry* **267**: 1990–1994.

Chilton M-DM, Que Q. **2003**. Targeted integration of T-DNA into the tobacco genome at double-stranded breaks: new insights on the mechanism of T-DNA integration. *Plant Physiology* **133**: 956–965.

Ching T, Huang S, Garmire LX. **2014**. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA* **20**: 1684–1696.

Citovsky V, Kapelnikov A, Oliel S, Zakai N, Rojas MR, Gilbertson RL, Tzfira T, Loyter A. **2004**. Protein interactions involved in nuclear import of the Agrobacterium VirE2 protein in vivo and in vitro. *Journal of Biological Chemistry* **279**: 29528–29533.

Cleene MD, Ley JD. **1976**. The host range of crown gall. *The Botanical Review* **42**: 389–466.

ClinicalTrials.gov NCT02363322. **2017**. Putative investigational therapeutics in the treatment of patients with known ebola infection. *https://clinicaltrials.gov/ct2/show/NCT02363322*.

ClinicalTrials.gov NCT02389192. **2017**. Safety and pharmacokinetics of a single ZMappTM administration in healthy adult volunteers. *https://clinicaltrials.gov/ct2/show/NCT02389192*.

ClinicalTrials.gov NCT03301051. **2017**. Efficacy, safety, and immunogenicity of a plant-derived quadrivalent virus-like particles influenza vaccine in adults. *https://clinicaltrials.gov/ct2/show/NCT03301051*.

Cohn F. **1859**. Uber Proteinkristalle in der Kartoffel. *Jahres-Bericht der Schlesischen Gesellschaft für Vaterländische Cultur* **37**: 72–82.

Contour-Ansel D, Torres-Franklin ML, Zuily-Fodil Y, de Carvalho MHC. **2010**. An aspartic acid protease from common bean is expressed 'on call' during water stress and early recovery. *Journal of Plant Physiology* **167**: 1606–1612.

Cosgrove DJ. **2016**. Catalysts of plant cell wall loosening. *F1000Research* **5**: 1–13.

da Costa SL, Camargo R, Demasi M, Santana JM, Sá CM de, de Freitas SM. **2014**. Effects of an anticarcinogenic Bowman-Birk protease inhibitor on purified 20S proteasome and MCF-7 breast cancer cells. *PLoS ONE* **9**: e86600.

Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. **2014**. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics* **13**: 2513–2526.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. **2011**. Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research* **10**: 1794–1805.

Craik DJ, Daly NL, Bond T, Waine C. **1999**. Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *Journal of Molecular Biology* **294**: 1327–1336.

Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, *et al.* 2015. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4**.

Da Silva DS, de Oliveira CFR, Parra JRP, Marangoni S, Macedo MLR. **2014**. Short and long-term antinutritional effect of the trypsin inhibitor ApTI for biological control of sugarcane borer. *Journal of Insect Physiology* **61**: 1–7.

Dagdas YF, Belhaj K, Maqbool A, Chaparro-Garcia A, Pandey P, Petre B, Tabassum N, Cruz-Mireles N, Hughes RK, Sklenar J, *et al.* 2016. An effector of the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. *eLife* **5**: e10856.

Dahl SW, Rasmussen SK, Hejgaard J. **1996**. Heterologous expression of three plant serpins with distinct inhibitory specificities. *Journal of Biological Chemistry* **271**: 25083–25088.

DARPA. **2012**. Global biodefense program Blue Angel. *https://globalbiodefense.com/2012/07/28/darpa-program-hits-milestone-in-plant-based-vaccines-for-pandemics/*.

Dattagupta JK, Podder A, Chakrabarti C, Sen U, Mukhopadhyay D, Dutta SK, Singh M. **1999**. Refined crystal structure (2.3 Å) of a double-headed winged bean α-chymotrypsin inhibitor and location of its second reactive site. *Proteins* **35**: 321–331.

Davoodi-Semiromi A, Schreiber M, Nalapalli S, Verma D, Singh ND, Banks RK, Chakrabarti D, Daniell H. **2010**. Chloroplast-derived vaccine antigens confer dual immunity against cholera and malaria by oral or injectable delivery. *Plant Biotechnology Journal* **8**: 223–242.

Dawson WO, Lewandowski DJ, Hilf ME, Bubrick P, Raffo AJ, Shaw JJ, Grantham GL, Desjardins PR. **1989**. A tobacco mosaic virus-hybrid expresses and loses an added gene. *Virology* **172**: 285–292.

Delannoy M, Alves G, Vertommen D, Ma J, Boutry M, Navarre C. **2008**. Identification of peptidases in Nicotiana tabacum leaf intercellular fluid. *PROTEOMICS* **8**: 2285–2298.

Ditt RF, Kerr KF, de Figueiredo P, Delrow J, Comai L, Nester EW. **2006**. The Arabidopsis thaliana transcriptome in response to Agrobacterium tumefaciens. *Molecular plant-microbe interactions: MPMI* **19**: 665–681.

Dix MM, Simon GM, Cravatt BF. **2008**. Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* **134**: 679–691.

Dogru E, Warzecha H, Seibel F, Haebel S, Lottspeich F, Stöckigt J. **2000**. The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid biosynthesis in plants is an ortholog of the α/β-hydrolase super family. *European Journal of Biochemistry* **267**: 1397–1406.

Donini M, Lombardi R, Lonoce C, Carli MD, Marusic C, Morea V, Micco PD. **2015**. Antibody proteolysis: a common picture emerging from plants. *Bioengineered* **6**: 299–302.

Donson J, Kearney CM, Hilf ME, Dawson WO. **1991**. Systemic expression of a bacterial gene by a tobacco mosaic virus-based vector. *Proceedings of the National Academy of Sciences* **88**: 7204–7208.

Doran PM. **2006**. Foreign protein degradation and instability in plants and plant tissue cultures. *Trends in Biotechnology* **24**: 426–432.

Dowle M, Short T, Lianoglou S, Saporta R, Srinivasan A, Antonyan E. **2014**. data.table: Extension of data.frame. *http://CRAN.R-project.org/package=data.table*.

Du J, Guo X, Chen L, Xie C, Liu J. **2017**. Proteomic analysis of differentially expressed proteins of Nicotiana benthamiana triggered by INF1 elicitin from Phytophthora infestans. *Journal of General Plant Pathology* **83**: 66–77.

Duan X, Zhang Z, Wang J, Zuo K. **2016**. Characterization of a novel cotton subtilase gene GbSBT1 in response to extracellular stimulations and its role in Verticillium resistance. *PLOS ONE* **11**: e0153988.

Dunse KM, Stevens JA, Lay FT, Gaspar YM, Heath RL, Anderson MA. **2010**. Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 15011–15015.

Duwadi K, Chen L, Menassa R, Dhaubhadel S. **2015**. Identification, characterization and down-regulation of cysteine protease genes in tobacco for use in recombinant protein production. *PLOS ONE* **10**: e0130556.

Dyrløv Bendtsen J, Nielsen H, von Heijne G, Brunak S. **2004**. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**: 783–795.

Engler C, Youles M, Gruetzner R, Ehnert T-M, Werner S, Jones JDG, Patron NJ, Marillonnet S. **2014**. A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology* **3**: 839–843.

Ernst O, Zor T. **2010**. Linearization of the Bradford protein assay. *Journal of Visualized Experiments : JoVE* **38**.

Esse HP van, Klooster JW van't, Bolton MD, Yadeta KA, Baarlen P van, Boeren S, Vervoort J, Wit PJGM de, Thomma BPHJ. **2008**. The Cladosporium fulvum Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense. *The Plant Cell* **20**: 1948–1963.

Evangelisti E, Gogleva A, Hainaux T, Doumane M, Tulin F, Quan C, Yunusov T, Floch K, Schornack S. **2017**. Time-resolved dual transcriptomics reveal early induced Nicotiana benthamiana root genes and conserved infection-promoting Phytophthora palmivora effectors. *BMC Biology* **15**: 39.

Ewing B, Green P. **1998**. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**: 186–194.

Faino L, Jonge R de, Thomma BPHJ. **2012**. The transcriptome of Verticillium dahliae-infected Nicotiana benthamiana determined by deep RNA sequencing. *Plant Signaling & Behavior* **7**: 1065–1069.

Falk A, Feys BJ, Frost LN, Jones JDG, Daniels MJ, Parker JE. **1999**. EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases. *Proceedings of the National Academy of Sciences* **96**: 3292–3297.

Fan H, Sun H, Wang Y, Zhang Y, Wang X, Li D, Yu J, Han C. **2014**. Deep Sequencing–Based Transcriptome Profiling Reveals Comprehensive Insights into the Responses of Nicotiana benthamiana to Beet necrotic yellow vein virus Infections Containing or Lacking RNA4. *PLOS ONE* **9**: e85284.

Faro C, Gal S. **2005**. Aspartic proteinase content of the Arabidopsis genome. *Current Protein & Peptide Science* **6**: 493–500.

Fedson DS. **2003**. Pandemic influenza and the global vaccine supply. *Clinical Infectious Diseases* **36**: 1552–1561.

Feldman SR, Gonias SL, Pizzo SV. **1985**. Model of α2-macroglobulin structure and function. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 5700–5704.

Fernandez-Pozo N, Rosli HG, Martin GB, Mueller LA. **2015**. The SGN VIGS tool: user-friendly software to design virus-induced gene silencing (VIGS) constructs for functional genomics. *Molecular Plant* **8**: 486–488.

Fincher GB. **1989**. Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 305–346.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, *et al.* 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**: D279–D285.

Fluhr R, Lampl N, Roberts TH. **2012**. Serpin protease inhibitors in plant biology. *Physiologia Plantarum* **145**: 95–102.

Forouhar F, Yang Y, Kumar D, Chen Y, Fridman E, Park SW, Chiang Y, Acton TB, Montelione GT, Pichersky E, *et al.* 2005. Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 1773–1778.

Fox J. **2005**. Getting started with the R commander: a basic-statistics graphical user interface to R. *J Stat Softw* **14**: 1–42.

Fox JL. **2012**. First plant-made biologic approved. *Nature Biotechnology* **30**: 472–472.

Franco OL, Rigden DJ, Melo FR, Grossi-de-Sá MF. **2002**. Plant α-amylase inhibitors and their interaction with insect α-amylases. *European Journal of Biochemistry* **269**: 397–412.

Franziska FS. **2013**. Snap-shot of Serine Carboxypeptidase-Like acyltransferase evolution: The loss of conserved disulphide bridge is responsible for the completion of neo-functionalization. *Journal of Phylogenetics & Evolutionary Biology* **01**: 115.

Fraser CM, Rider LW, Chapple C. **2005**. An expression and bioinformatics analysis of the Arabidopsis Serine Carboxypeptidase-Like gene family. *Plant Physiology* **138**: 1136–1148.

Fujiuchi N, Matoba N, Matsuda R. **2016**. Environment control to improve recombinant protein yields in plants based on Agrobacterium-mediated transient ene expression. *Frontiers in Bioengineering and Biotechnology* **4**.

Gao H, Zhang Y, Wang W, Zhao K, Liu C, Bai L, Li R, Guo Y. **2017**. Two membrane-anchored aspartic proteases contribute to pollen and ovule development. *Plant Physiology* **173**: 219–239.

Garman SC, Garboczi DN. **2002**. Structural basis of Fabry disease. *Molecular Genetics and Metabolism* **77**: 3–11.

Gelvin SB. **2003**. Agrobacterium-Mediated Plant Transformation: the Biology behind the "Gene-Jockeying" Tool. *Microbiology and Molecular Biology Reviews* **67**: 16–37.

Geng C, Wang H-Y, Liu J, Yan Z-Y, Tian Y-P, Yuan X-F, Gao R, Li X-D. **2017**. Transcriptomic changes in Nicotiana benthamiana plants inoculated with the wild-type or an attenuated mutant of Tobacco vein banding mosaic virus. *Molecular plant pathology* **18**: 1175–1188.

Gepstein S, Sabehi G, Carp M-J, Hajouj T, Nesher MFO, Yariv I, Dor C, Bassani M. **2003**. Large-scale identification of leaf senescence-associated genes. *The Plant Journal* **36**: 629–642.

Ghorbani S, Hoogewijs K, Pečenková T, Fernandez A, Inzé A, Eeckhout D, Kawa D, De Jaeger G, Beeckman T, Madder A, *et al.* 2016. The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation. *Journal of Experimental Botany* **67**: 4877–4887.

Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA. **2008**. Biosynthesis of circular proteins in plants. *The Plant Journal* **53**: 505–515.

Gilroy EM, Hein I, van der Hoorn RAL, Boevink PC, Venter E, McLellan H, Kaffarnik F, Hrubikova K, Shaw J, Holeva M, *et al.* 2007. Involvement of cathepsin B in the plant disease resistance hypersensitive response. *The Plant Journal* **52**: 1–13.

Girard C, Rivard D, Kiggundu A, Kunert K, Gleddie SC, Cloutier C, Michaud D. **2007**. A multicomponent, elicitor-inducible cystatin complex in tomato, Solanum lycopersicum. *New Phytologist* **173**: 841–851.

Giritch A, Marillonnet S, Engler C, van Eldik G, Botterman J, Klimyuk V, Gleba Y. **2006**. Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 14701–14706.

GiveWell. **2017a**. Malaria Consortium – Seasonal Malaria Chemoprevention. *http://www.givewell.org/charities/malaria-consortium*.

GiveWell. **2017b**. Possible (formerly known as Nyaya Health). *http://www.givewell.org/international/charities/Nyaya-Health*.

Gleba Y, Klimyuk V, Marillonnet S. **2005**. Magnifection—a new platform for expressing recombinant vaccines in plants. *Vaccine* **23**: 2042–2048.

Goetz M, Godt DE, Guivarc'h A, Kahmann U, Chriqui D, Roitsch T. **2001**. Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 6522–6527.

Gómez-Gómez L, Felix G, Boller T. **1999**. A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. *The Plant Journal* **18**: 277–284.

Goodin MM, Zaitlin D, Naidu RA, Lommel SA. **2008**. Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions. *Molecular plant-microbe interactions* **21**: 1015–1026.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, *et al.* 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–D1186.

Goulet C, Benchabane M, Anguenot R, Brunelle F, Khalf M, Michaud D. **2010a**. A companion protease inhibitor for the protection of cytosol-targeted recombinant proteins in plants. *Plant Biotechnology Journal* **8**: 142–154.

Goulet M-C, Dallaire C, Vaillancourt L-P, Khalf M, Badri AM, Preradov A, Duceppe M-O, Goulet C, Cloutier C, Michaud D. **2008**. Tailoring the specificity of a plant cystatin toward herbivorous insect digestive cysteine proteases by single mutations at positively selected amino acid sites. *PLANT PHYSIOLOGY* **146**: 1010–1019.

Goulet C, Goulet C, Goulet M-C, Michaud D. **2010b**. 2-DE proteome maps for the leaf apoplast of Nicotiana benthamiana. *Proteomics* **10**: 2536–2544.

Goulet C, Khalf M, Sainsbury F, D'Aoust M-A, Michaud D. **2012**. A protease activity–depleted environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant Biotechnology Journal* **10**: 83–94.

Graham JS, Pearce G, Merryweather J, Titani K, Ericsson LH, Ryan CA. **1985**. Wound-induced proteinase inhibitors from tomato leaves. The cDNA-deduced primary structure of pre-inhibitor II. *Journal of Biological Chemistry* **260**: 6561–6564.

Green AR, Nissen MS, Kumar GNM, Knowles NR, Kang C. **2013**. Characterization of Solanum tuberosum multicystatin and the significance of core domains. *The Plant Cell* **25**: 5043–5052.

Greenbaum DC, Baruch A, Grainger M, Bozdech Z, Medzihradszky KF, Engel J, DeRisi J, Holder AA, Bogyo M. **2002**. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298**: 2002–2006.

Greenblatt HM, Ryan CA, James MNG. **1989**. Structure of the complex of Streptomyces griseus proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. *Journal of Molecular Biology* **205**: 201–228.

Griffiths J, Murase K, Rieu I, Zentella R, Zhang Z-L, Powers SJ, Gong F, Phillips AL, Hedden P, Sun T, *et al.* 2006. Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. *The Plant Cell* **18**: 3399–3414.

Grill LK, Reinl SJ, Hoffman SL, Turpen TH, Fallarme V, Charoenvit Y. **1995**. Malaria epitopes expressed on the surface of recombinant tobacco mosaic virus. *Nature Biotechnology* **13**: 53.

Grimsley N, Hohn B, Hohn T, Walden R. **1986**. "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. *Proceedings of the National Academy of Sciences* **83**: 3282–3286.

Grosse-Holz FM, van der Hoorn RAL. **2016**. Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants. *New Phytologist* **210**: 794–807.

Grosse-Holz F, Kelly S, Blaskowski S, Kaschani F, Kaiser M, van der Hoorn RAL. **2017**. The transcriptome, extracellular proteome and active secretome of agroinfiltrated N. benthamiana uncover a large, diverse protease repertoire. *Plant Biotechnology Journal* **in press**.

Gu C, Shabab M, Strasser R, Wolters PJ, Shindo T, Niemer M, Kaschani F, Mach L, van der Hoorn RAL. **2012**. Post-translational regulation and trafficking of the granulin-containing protease RD21 of Arabidopsis thaliana. *PLoS One* **7**: e32422.

Guerra FP, Reyes L, Vergara-Jaque A, Campos-Hernández C, Gutiérrez A, Pérez-Díaz J, Pérez-Díaz R, Blaudez D, Ruíz-Lara S. **2015**. Populus deltoides Kunitz trypsin inhibitor 3 confers metal tolerance and binds copper, revealing a new defensive role against heavy metal stress. *Environmental and Experimental Botany* **115**: 28–37.

Guo Y, Zheng Z, Clair JJL, Chory J, Noel JP. **2013**. Smoke-derived karrikin perception by the α/β-hydrolase KAI2 from Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 8284–8289.

Hamorsky KT, Grooms-Williams TW, Husk AS, Bennett LJ, Palmer KE, Matoba N. **2013**. Efficient single tobamoviral vector-based bioproduction of broadly neutralizing anti-HIV-1 monoclonal

antibody VRC01 in Nicotiana benthamiana plants and utility of VRC01 in combination microbicides. *Antimicrobial Agents and Chemotherapy* **57**: 2076–2086.

Hao L, Hsiang T, Goodwin PH. **2006**. Role of two cysteine proteinases in the susceptible response of Nicotiana benthamiana to Colletotrichum destructivum and the hypersensitive response to Pseudomonas syringae pv. tomato. *Plant Science* **170**: 1001–1009.

Hara-Nishimura I, Inoue K, Nishimura M. **1991**. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Letters* **294**: 89–93.

Hatsugai N, Kuroyanagi M, Yamada K, Meshi T, Tsuda S, Kondo M, Nishimura M, Hara-Nishimura I. **2004**. A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science* **305**: 855–858.

Hatsugai N, Yamada K, Goto-Yamada S, Hara-Nishimura I. **2015**. Vacuolar processing enzyme in plant programmed cell death. *Frontiers in Plant Science* **6**: 234.

Hayashi Y, Yamada K, Shimada T, Matsushima R, Nishizawa N, Nishimura M, 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**: 894–899.

Hayat A, Haria D, Salifu MO. **2008**. Erythropoietin stimulating agents in the management of anemia of chronic kidney disease. *Patient preference and adherence* **2**: 195.

Heath RL, Barton PA, Simpson RJ, Reid GE, Lim G, Anderson MA. **1995**. Characterization of the protease processing sites in a multidomain proteinase inhibitor precursor from Nicotiana alata. *European Journal of Biochemistry* **230**: 250–257.

Hedstrom L, Szilagyi L, Rutter WJ. **1992**. Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**: 1249–1253.

Hehle VK, Lombardi R, van Dolleweerd CJ, Paul MJ, Di Micco P, Morea V, Benvenuto E, Donini M, Ma JK-C. **2015**. Site-specific proteolytic degradation of IgG monoclonal antibodies expressed in tobacco plants. *Plant Biotechnology Journal* **13**: 235–245.

Hehle VK, Paul MJ, Drake PM, Ma JK, Dolleweerd CJ van. **2011**. Antibody degradation in tobacco plants: a predominantly apoplastic process. *BMC Biotechnology* **11**: 128.

Hehle VK, Paul MJ, Roberts VA, Dolleweerd CJ van, Ma JK-C. **2016**. Site-targeted mutagenesis for stabilization of recombinant monoclonal antibody expressed in tobacco (Nicotiana tabacum) plants. *The FASEB Journal* **30**: 1590–1598.

Hejgaard J, Kaersgaard P. **1983**. Purification and properties of the major antigenic beer protein of barley origin. *Journal of the Institute of Brewing* **89**: 402–410.

Hejgaard J, Rasmussen SK, Brandt A, Svendsen I. **1985**. Sequence homology between barley endosperm protein Z and protease inhibitors of the α1-antitrypsin family. *FEBS Letters* **180**: 89–94.

Hernandez J-F, Gagnon J, Chiche L, Nguyen TM, Andrieu J-P, Heitz A, Trinh Hong T, Pham TTC, Le Nguyen D. **2000**. Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. *Biochemistry* **39**: 5722–5730.

Hiraiwa N, Nishimura M, 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**: 213–216.

van der Hoorn RAL. **2008**. Plant Proteases: From Phenotypes to Molecular Mechanisms. *Annual Review of Plant Biology* **59**: 191–223.

van der Hoorn RAL, Rivas S, Wulff BBH, Jones JDG, Joosten MHAJ. **2003**. Rapid migration in gel filtration of the Cf-4 and Cf-9 resistance proteins is an intrinsic property of Cf proteins and not because of their association with high-molecular-weight proteins. *The Plant Journal* **35**: 305–315.

Huang X, Madan A. **1999**. CAP3: A DNA sequence assembly program. *Genome Research* **9**: 868–877.

Hückelhoven R, Dechert C, Trujillo M, Kogel K-H. **2001**. Differential expression of putative cell death regulator genes in near-isogenic, resistant and susceptible barley lines during interaction with the powdery mildew fungus. *Plant molecular biology* **47**: 739–748.

Huesgen PF, Overall CM. **2012**. N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. *Physiologia Plantarum* **145**: 5–17.

Hull AK, Criscuolo CJ, Mett V, Groen H, Steeman W, Westra H, Chapman G, Legutki B, Baillie L, Yusibov V. **2005**. Human-derived, plant-produced monoclonal antibody for the treatment of anthrax. *Vaccine* **23**: 2082–2086.

Huntington JA, Read RJ, Carrell RW. **2000**. Structure of a serpin–protease complex shows inhibition by deformation. *Nature* **407**: 923–926.

Husted S, Schjoerring JK. **1995**. Apoplastic pH and ammonium concentration in leaves of Brassica napus L. *Plant Physiology* **109**: 1453–1460.

Hwang JE, Hong JK, Je JH, Lee KO, Kim DY, Lee SY, Lim CO. **2009**. Regulation of seed germination and seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. *Plant Cell Reports* **28**: 1623–1632.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. **2010**. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

IBIO. **2017a**. Global Market :: iBio Inc. *https://www.ibioinc.com/about/global-market*.

IBIO. **2017b**. iBio Inc press release. *https://ir.ibioinc.com/press-releases/detail/84/ibio-cdmo-president-to-co-chair-south-african-molecular*.

Ivanchenko MG, den Os D, Monshausen GB, Dubrovsky JG, Bednářová A, Krishnan N. **2013**. Auxin increases the hydrogen peroxide concentration in tomato root tips while inhibiting root growth. *Annals of Botany* **112**: 1107–1116.

Iwasaki T, Kiyohara T, Yoshikawa M. **1976**. Amino acid sequence of an active fragment of potato proteinase inhibitor IIa. *Journal of Biochemistry* **79**: 381–391.

Jennings C, West J, Waine C, Craik D, Anderson M. **2001**. Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from Oldenlandia affinis. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10614–10619.

Jez J, Castilho A, Grass J, Vorauer-Uhl K, Sterovsky T, Altmann F, Steinkellner H. **2013**. Expression of functionally active sialylated human erythropoietin in plants. *Biotechnology Journal* **8**: 371–382.

Jiang H, Doerge R, Gelvin SB. **2003**. Transfer of T-DNA and Vir proteins to plant cells by Agrobacterium tumefaciens induces expression of host genes involved in mediating transformation and suppresses host defense gene expression. *The Plant Journal* **35**: 219–236.

Johnson R, Narvaez J, An G, Ryan C. **1989**. Expression of proteinase inhibitors I and II in transgenic tobacco plants: effects on natural defense against Manduca sexta larvae. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 9871–9875.

Jordá L, Coego A, Conejero V, Vera P. **1999**. A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. *Journal of Biological Chemistry* **274**: 2360–2365.

Joshi RS, Gupta VS, Giri AP. **2014**. Differential antibiosis against Helicoverpa armigera exerted by distinct inhibitory repeat domains of Capsicum annuum proteinase inhibitors. *Phytochemistry* **101**: 16–22.

Jutras PV, D'Aoust M-A, Couture MM-J, Vézina L-P, Goulet M-C, Michaud D, Sainsbury F. **2015**. Modulating secretory pathway pH by proton channel co-expression can increase recombinant protein stability in plants. *Biotechnology Journal* **10**: 1478–1486.

Jutras PV, Marusic C, Lonoce C, Deflers C, Goulet M-C, Benvenuto E, Michaud D, Donini M. **2016**. An accessory protease inhibitor to increase the yield and quality of a tumour-targeting mAb in Nicotiana benthamiana leaves. *PLOS ONE* **11**: e0167086.

Kalmus H. **1945**. The use of abrasives in the transmission of plant viruses. *Annals of Applied Biology* **32**: 230–234.

Kang S-R, Oh S-K, Kim J-J, Choi D-I, Baek K-H. **2010**. NMMP1, a Matrix Metalloprotease in Nicotiana benthamiana has a role in protection against bacterial infection. *The Plant Pathology Journal* **26**: 402–408.

Kaschani F, Gu C, Niessen S, Hoover H, Cravatt BF, van der Hoorn RAL. **2009**. Diversity of serine hydrolase activities of unchallenged and botrytis-infected Arabidopsis thaliana. *Molecular & Cellular Proteomics* **8**: 1082–1093.

Kaschani F, Shabab M, Bozkurt T, Shindo T, Schornack S, Gu C, Ilyas M, Win J, Kamoun S, van der Hoorn RAL. **2010**. An effector-targeted protease contributes to defense against Phytophthora infestans and is under diversifying selection in natural hosts. *Plant Physiology* **154**: 1794–1804.

Katoh K, Standley DM. **2013**. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, *et al.* 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kelly LJ, Leitch AR, Clarkson JJ, Knapp S, Chase MW. **2013**. Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (Nicotiana section Suaveolentes). *Evolution* **67**: 80–94.

Kiggundu A, Goulet M-C, Goulet C, Dubuc J-F, Rivard D, Benchabane M, Pépin G, Vyver C van der, Kunert K, Michaud D. **2006**. Modulating the proteinase inhibitory profile of a plant cystatin by single mutations at positively selected amino acid sites. *The Plant Journal* **48**: 403–413.

Kim NH, Hwang BK. **2015**. Pepper pathogenesis-related protein 4c is a plasma membrane-localized cysteine protease inhibitor that is required for plant cell death and defense signaling. *The Plant Journal* **81**: 81–94.

Kim N-S, Kim T-G, Kim O-H, Ko E-M, Jang Y-S, Jung E-S, Kwon T-H, Yang M-S. **2008a**. Improvement of recombinant hGM-CSF production by suppression of cysteine proteinase gene expression using RNA interference in a transgenic rice culture. *Plant Molecular Biology* **68**: 263–275.

Kim HG, Kwon SJ, Jang YJ, Nam MH, Chung JH, Na Y-C, Guo H, Park OK. **2013a**. GDSL LIPASE1 modulates plant immunity through feedback regulation of ethylene signaling. *Plant Physiology* **163**: 1776–1791.

Kim T-G, Lee H-J, Jang Y-S, Shin Y-J, Kwon T-H, Yang M-S. **2008b**. Co-expression of proteinase inhibitor enhances recombinant human granulocyte–macrophage colony stimulating factor production in transgenic rice cell suspension culture. *Protein Expression and Purification* **61**: 117–121.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. **2013b**. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* **14**: R36.

Knobler S, Mahmoud A, Lemon S. **2004**. *Learning from SARS: Preparing for the Next Disease Outbreak: Workshop Summary*. Washington, DC: National Academies Press.

Komarnytsky S, Borisjuk N, Yakoby N, Garvey A, Raskin I. **2006**. Cosecretion of protease inhibitor stabilizes antibodies produced by plant roots. *Plant Physiology* **141**: 1185–1193.

Kong L, Ranganathan S. **2008**. Tandem duplication, circular permutation, molecular adaptation: how Solanaceae resist pests via inhibitors. *BMC Bioinformatics* **9**: S22.

Kopylova E, Noé L, Touzet H. **2012**. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Kovács J, van der Hoorn RAL. **2016**. Twelve ways to confirm targets of activity-based probes in plants. *Bioorganic & Medicinal Chemistry* **24**: 3304–3311.

Kregten M van, Pater S de, Romeijn R, Schendel R van, Hooykaas PJJ, Tijsterman M. **2016**. T-DNA integration in plants results from polymerase-θ-mediated DNA repair. *Nature Plants* **2**: 16164.

Krüger J, Thomas CM, Golstein C, Dixon MS, Smoker M, Tang S, Mulder L, Jones JDG. **2002**. A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. *Science* **296**: 744–747.

Kumagai MH, Donson J, della-Cioppa G, Harvey D, Hanley K, Grill LK. **1995**. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences* **92**: 1679–1683.

Kumar GNM, Houtz RL, Knowles NR. **1999**. Age-induced protein modifications and increased proteolysis in potato seed-tubers. *Plant Physiology* **119**: 89–100.

Kunitz M. **1945**. Crystallization of a trypsin inhibitor from soybean. *Science* **101**: 668–669.

Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G. **2004**. The N-terminus of bacterial Elongation Factor Tu elicits Innate immunity in Arabidopsis plants. *The Plant Cell* **16**: 3496–3507.

Lampl N, Alkan N, Davydov O, Fluhr R. **2013**. Set-point control of RD21 protease activity by AtSerpin1 controls cell death in Arabidopsis. *The Plant Journal* **74**: 498–510.

Lampl N, Budai-Hadrian O, Davydov O, Joss TV, Harrop SJ, Curmi PMG, Roberts TH, Fluhr R. **2010**. Arabidopsis AtSerpin1, crystal structure and in vivo interaction with its target protease RESPONSIVE TO DESICCATION-21 (RD21). *Journal of Biological Chemistry* **285**: 13550–13560.

Lang J, Gonzalez-Mula A, Taconnat L, Clement G, Faure D. **2016**. The plant GABA signaling downregulates horizontal transfer of the Agrobacterium tumefaciens virulence plasmid. *The New Phytologist* **210**: 974–983.

Laskowski M, Kato I. **1980**. Protein inhibitors of proteinases. *Annual Review of Biochemistry* **49**: 593–626.

Lazo GR, Stein PA, Ludwig RA. **1991**. A DNA Transformation–Competent Arabidopsis Genomic Library in Agrobacterium. *Nature Biotechnology* **9**: 963–7.

Leah R, Mundy J. **1989**. The bifunctional α-amylase/subtilisin inhibitor of barley: nucleotide sequence and patterns of seed-specific expression. *Plant Molecular Biology* **12**: 673–682.

Lee C-W, Efetova M, Engelmann JC, Kramell R, Wasternack C, Ludwig-Müller J, Hedrich R, Deeken R. **2009**. Agrobacterium tumefaciens promotes tumor induction by modulating pathogen defense in Arabidopsis thaliana. *The Plant Cell* **21**: 2948–2962.

Lee MCS, Scanlon MJ, Craik DJ, Anderson MA. **1999**. A novel two-chain proteinase inhibitor generated by circularization of a multidomain precursor protein. *Nature Structural & Molecular Biology* **6**: 526–530.

Letunic I, Bork P. **2016**. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**: W242–W245.

Li W, Cao J-Y, Xu Y-P, Cai X-Z. **2017**. Artificial Agrobacterium tumefaciens strains exhibit diverse mechanisms to repress Xanthomonas oryzae pv. oryzae-induced hypersensitive response and non-host resistance in Nicotiana benthamiana. *Molecular Plant Pathology* **18**: 489–502.

Li Y, Kabbage M, Liu W, Dickman MB. **2016**. Aspartyl protease-mediated cleavage of BAG6 is necessary for autophagy and fungal resistance in plants. *The Plant Cell* **28**: 233–24.

Limoli DH, Jones CJ, Wozniak DJ. **2015**. Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiology spectrum* **3**: 223–247.

Liu Y, Dammann C, Bhattacharyya MK. **2001**. The Matrix Metalloproteinase gene GmMMP2 is activated in response to pathogenic infections in soybean. *Plant Physiology* **127**: 1788–1797.

Liu H, Wang X, Zhang H, Yang Y, Ge X, Song F. **2008**. A rice serine carboxypeptidase-like gene OsBISCPL1 is involved in regulation of defense responses against biotic and oxidative stress. *Gene* **420**: 57–65.

Lomonossoff GP, D'Aoust M-A. **2016**. Plant-produced biopharmaceuticals: A case of technical developments driving clinical deployment. *Science* **353**: 1237–1240.

Love MI, Huber W, Anders S. **2014**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Lu H, Chandrasekar B, Oeljeklaus J, Misas-Villamil JC, Wang Z, Shindo T, Bogyo M, Kaiser M, van der Hoorn RAL. **2015**. Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination. *Plant Physiology* **168**: 1462–1475.

Luckett S, Garcia RS, Barker JJ, Konarev AV, Shewry PR, Clarke AR, Brady RL. **1999**. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. *Journal of Molecular Biology* **290**: 525–533.

Madureira HC, Da Cunha M, Jacinto T. **2006**. Immunolocalization of a defense-related 87 kDa cystatin in leaf blade of tomato plants. *Environmental and Experimental Botany* **55**: 201–208.

Mahatmanto T. **2015**. Review seed biopharmaceutical cyclic peptides: From discovery to applications. *Peptide Science* **104**: 804–814.

Major IT, Constabel CP. **2008**. Functional analysis of the Kunitz Trypsin Inhibitor family in poplar reveals biochemical diversity and multiplicity in defence against herbivores. *Plant Physiology* **146**: 888–903.

Mandal MK, Ahvari H, Schillberg S, Schiermeyer A. **2016**. Tackling unwanted proteolysis in plant production hosts used for molecular farming. *Frontiers in Plant Science* **7**: 267.

Mandal MK, Fischer R, Schillberg S, Schiermeyer A. **2014**. Inhibition of protease activity by antisense RNA improves recombinant protein production in Nicotiana tabacum cv. Bright Yellow 2 (BY-2) suspension cells. *Biotechnology Journal* **9**: 1065–1073.

Margis R, Reis EM, Villeret V. **1998**. Structural and phylogenetic relationships among plant and animal cystatins. *Archives of Biochemistry and Biophysics* **359**: 24–30.

Marillonnet S, Giritch A, Gils M, Kandzia R, Klimyuk V, Gleba Y. **2004**. In planta engineering of viral RNA replicons: Efficient assembly by recombination of DNA modules delivered by Agrobacterium. *Proceedings of the National Academy of Sciences* **101**: 6852–6857.

Marillonnet S, Thoeringer C, Kandzia R, Klimyuk V, Gleba Y. **2005**. Systemic Agrobacterium tumefaciens–mediated transfection of viral replicons for efficient transient expression in plants. *Nature Biotechnology* **23**: 718–723.

Marsian J, Fox H, Bahar MW, Kotecha A, Fry EE, Stuart DI, Macadam AJ, Rowlands DJ, Lomonossoff GP. **2017**. Plant-made polio type 3 stabilized VLPs—a candidate synthetic polio vaccine. *Nature Communications* **8**: 245.

Martin K, Kopperud K, Chakrabarty R, Banerjee R, Brooks R, Goodin MM. **2009**. Transient expression in Nicotiana benthamiana fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *The Plant Journal* **59**: 150–162.

Martínez M, Cambra I, González-Melendi P, Santamaría ME, Díaz I. **2012**. C1A cysteine-proteases and their inhibitors in plants. *Physiologia Plantarum* **145**: 85–94.

Martinez M, Diaz I. **2008**. The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. *BMC Evolutionary Biology* **8**: 198.

Martinez M, Diaz-Mendoza M, Carrillo L, Diaz I. **2007**. Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. *FEBS Letters* **581**: 2914–2918.

Maskos K, Huber-Wunderlich M, Glockshuber R. **1996**. RBI, a one-domain α-amylase/trypsin inhibitor with completely independent binding sites. *FEBS Letters* **397**: 11–16.

McLellan H, Gilroy EM, Yun B-W, Birch PRJ, Loake GJ. **2009**. Functional redundancy in the Arabidopsis Cathepsin B gene family contributes to basal defence, the hypersensitive response and senescence. *New Phytologist* **183**: 408–418.

Meshcheriakova Y, Durrant A, Hesketh EL, Ranson NA, Lomonossoff GP. **2017**. Combining high-resolution cryo-electron microscopy and mutagenesis to develop cowpea mosaic virus for bionanotechnology. *Biochemical Society Transactions*: BST20160312.

Meulenbroek EM, Thomassen EAJ, Pouvreau L, Abrahams JP, Gruppen H, Pannu NS. **2012**. Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato. *Acta Crystallographica Section D* **68**: 794–799.

Micheelsen PO, Vévodová J, De Maria L, Østergaard PR, Friis EP, Wilson K, Skjøt M. **2008**. Structural and mutational analyses of the interaction between the barley α-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. *Journal of Molecular Biology* **380**: 681–690.

Michel K, Abderhalden O, Bruggmann R, Dudler R. **2006**. Transcriptional changes in powdery mildew infected wheat and Arabidopsis leaves undergoing syringolin-triggered hypersensitive cell death at infection sites. *Plant Molecular Biology* **62**: 561–578.

Migliolo L, de Oliveira AS, Santos EA, Franco OL, de Sales MP. **2010**. Structural and mechanistic insights into a novel non-competitive Kunitz trypsin inhibitor from Adenanthera pavonina L. seeds with double activity toward serine- and cysteine-proteinases. *Journal of Molecular Graphics and Modelling* **29**: 148–156.

Mignery GA, Pikaard CS, Park WD. **1988**. Molecular characterization of the patatin multigene family of potato. *Gene* **62**: 27–44.

Mindrebo JT, Nartey CM, Seto Y, Burkart MD, Noel JP. **2016**. Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Current Opinion in Structural Biology* **41**: 233–246.

Misas-Villamil JC, van der Burgh AM, Grosse-Holz F, Bach-Pages M, Kovács J, Kaschani F, Schilasky S, Emon AEK, Ruben M, Kaiser M, *et al.* 2017. Subunit-selective proteasome activity profiling uncovers uncoupled proteasome subunit activities during bacterial infections. *The Plant Journal* **90**: 418–430.

Morimoto K, van der Hoorn RAL. **2016**. The increasing impact of Activity-Based Protein Profiling in plant science. *Plant and Cell Physiology* **57**: 446–461.

Moura DS, Bergey DR, Ryan CA. **2001**. Characterization and localization of a wound-inducible type I serine-carboxypeptidase from leaves of tomato plants (Lycopersicon esculentum Mill.). *Planta* **212**: 222–230.

Mundy J, Svendsen IB, Hejgaard J. **1983**. Barley α-amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Research Communications* **48**: 81–90.

Murai N, Kemp JD. **1982**. Octopine synthase mRNA isolated from sunflower crown gall callus is homologous to the Ti plasmid of Agrobacterium tumefaciens. *Proceedings of the National Academy of Sciences* **79**: 86–90.

Murray CJ, Lopez AD, Chin B, Feehan D, Hill KH. **2006**. Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *The Lancet* **368**: 2211–2218.

Mylne JS, Chan LY, Chanson AH, Daly NL, Schaefer H, Bailey TL, Nguyencong P, Cascales L, Craik DJ. **2012**. Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase–mediated biosynthesis. *The Plant Cell* **24**: 2765–2778.

Mylne JS, Colgrave ML, Daly NL, Chanson AH, Elliott AG, McCallum EJ, Jones A, Craik DJ. **2011**. Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nature Chemical Biology* **7**: 257–259.

Mysore KS, Bassuner B, Deng X, Darbinian NS, Motchoulski A, Ream W, Gelvin SB. **1998**. Role of the Agrobacterium tumefaciens VirD2 Protein in T-DNA Transfer and Integration. *Molecular Plant-Microbe Interactions* **11**: 668–683.

Mysore KS, Nam J, Gelvin SB. **2000**. An Arabidopsis histone H2A mutant is deficient in Agrobacterium T-DNA integration. *Proceedings of the National Academy of Sciences* **97**: 948–953.

Nakai A, Yamauchi Y, Sumi S, Tanaka K. **2012**. Role of acylamino acid-releasing enzyme/oxidized protein hydrolase in sustaining homeostasis of the cytoplasmic antioxidative system. *Planta* **236**: 427–436.

Nakasugi K, Crowhurst R, Bally J, Waterhouse P. **2014**. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant Nicotiana benthamiana (O Mittapalli, Ed.). *PLoS ONE* **9**: e91776.

Nakaune S, Yamada K, Kondo M, Kato T, Tabata S, Nishimura M, Hara-Nishimura I. **2005**. A vacuolar processing enzyme, delta VPE, is involved in seed coat formation at the early stage of seed development. *Plant Cell* **17**: 876–887.

Nandi S, Kwong AT, Holtz BR, Erwin RL, Marcel S, McDonald KA. **2016**. Techno-economic analysis of a transient plant-based platform for monoclonal antibody production. *mAbs* **8**: 1456–1466.

NCBI Resource Coordinators. **2017**. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **45**: D12–D17.

Neve M de, Loose MD, Jacobs A, Houdt HV, Kaluza B, Weidle U, Montagu MV, Depicker A. **1993**. Assembly of an antibody and its derived antibody fragment in Nicotiana and Arabidopsis. *Transgenic Research* **2**: 227–237.

Nielsen KJ, Heath RL, Anderson MA, Craik DJ. **1995**. Structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of Nicotiana alata. *Biochemistry* **34**: 14304–14311.

Nielsen AZ, Ziersen B, Jensen K, Lassen LM, Olsen CE, Møller BL, Jensen PE. **2013**. Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synthetic Biology* **2**: 308–315.

Niemer M, Mehofer U, Torres Acosta JA, Verdianz M, Henkel T, Loos A, Strasser R, Maresch D, Rademacher T, Steinkellner H, *et al.* 2014. The human anti-HIV antibodies 2F5, 2G12, and PG9 differ in their susceptibility to proteolytic degradation: Down-regulation of endogenous serine and cysteine proteinase activities could improve antibody production in plant-based expression platforms. *Biotechnology Journal* **9**: 493–500.

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

Nikolenko SI, Korobeynikov AI, Alekseyev MA. **2013**. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* **14**: S7.

Nissen MS, Kumar GNM, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowles NR, Kang C. **2009**. Characterization of Solanum tuberosum multicystatin and its structural comparison with other cystatins. *The Plant Cell Online* **21**: 861–875.

Odani S, Ikenaka T. **1973**. Scission of soybean Bowman-Birk proteinase inhibitor into two small fragments having either trypsin or chymotrypsin inhibitory activity. *Journal of Biochemistry* **74**: 857–860.

Ohshima M, Matsuoka M, Yamamoto N, Tanaka Y, Kano-Murakami Y, Ozeki Y, Kato A, Harada N, Ohashi Y. **1987**. Nucleotide sequence of the PR-1 gene of Nicotiana tabacum. *FEBS letters* **225**: 243–246.

Ohtani T, Galili G, Wallace JC, Thompson GA, Larkins BA. **1991**. Normal and lysine-containing zeins are unstable in transgenic tobacco seeds. *Plant Molecular Biology* **16**: 117–128.

O'Keefe BR, Vojdani F, Buffa V, Shattock RJ, Montefiori DC, Bakke J, Mirsalis J, d'Andrea A-L, Hume SD, Bratcher B, *et al.* 2009. Scaleable manufacture of HIV-1 entry inhibitor griffithsin and validation of its safety and efficacy as a topical microbicide component. *Proceedings of the National Academy of Sciences* **106**: 6099–6104.

O'Neil ST, Emrich SJ. **2013**. Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genomics* **14**: 465.

Orr GL, Strickland JA, Walsh TA. **1994**. Inhibition of Diabrotica larval growth by a multicystatin from potato tubers. *Journal of Insect Physiology* **40**: 893–900.

Otlewski J, Zbyryt T. **1994**. Single peptide bond hydrolysis/resynthesis in squash inhibitors of serine proteinases. 1. Kinetics and thermodynamics of the interaction between squash inhibitors and bovine beta-trypsin. *Biochemistry* **33**: 200–207.

Outchkourov NS, De Kogel WJ, Wiegers GL, Abrahamson M, Jongsma MA. **2004**. Engineered multidomain cysteine protease inhibitors yield resistance against western flower thrips (Frankliniella occidentalis) in greenhouse trials. *Plant Biotechnology Journal* **2**: 449–458.

Paireder M, Mehofer U, Tholen S, Porodko A, Schähs P, Maresch D, Biniossek ML, van der Hoorn RAL, Lenarcic B, Novinec M, *et al.* 2016. The death enzyme CP14 is a unique papain-like cysteine

proteinase with a pronounced S2 subsite selectivity. *Archives of Biochemistry and Biophysics* **603**: 110–117.

Paireder M, Tholen S, Porodko A, Biniossek ML, Mayer B, Novinec M, Schilling O, 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 (BBA) - Proteins and Proteomics* **1865**: 444–452.

Park EY, Kim J-A, Kim H-W, Kim YS, Song HK. **2004**. Crystal Structure of the Bowman–Birk inhibitor from barley seeds in ternary complex with porcine trypsin. *Journal of Molecular Biology* **343**: 173–186.

Pascual DW. **2007**. Vaccines are for dinner. *Proceedings of the National Academy of Sciences* **104**: 10757–10758.

Patro R, Duggal G, Kingsford C. **2015**. Salmon: accurate, versatile and ultrafast quantification from RNA-seq data using lightweight-alignment. *bioRxiv*: 021592.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. **2017**. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**: 417–419.

Pearce G, Yamaguchi Y, Barona G, Ryan CA. **2010**. A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 14921–14925.

Pekkarinen AI, Jones BL. **2003**. Purification and identification of barley proteins that inhibit the alkaline Serine proteinases of Fusarium culmorum. *Journal of Agricultural and Food Chemistry* **51**: 1710–1717.

Pekkarinen AI, Longstaff C, Jones BL. **2007**. Kinetics of the inhibition of Fusarium Serine proteinases by barley (Hordeum vulgare L.) inhibitors. *Journal of Agricultural and Food Chemistry* **55**: 2736–2742.

Pelegrini PB, Franco OL. **2005**. Plant γ-thionins: Novel insights on the mechanism of action of a multi-functional class of defense proteins. *The International Journal of Biochemistry & Cell Biology* **37**: 2239–2253.

Pérez-Bueno ML, Rahoutei J, Sajnani C, García-Luque I, Barón M. **2004**. Proteomic analysis of the oxygen-evolving complex of photosystem II under biotec stress: Studies on Nicotiana benthamiana infected with tobamoviruses. *PROTEOMICS* **4**: 418–425.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. **2015**. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**: 290–295.

Petre B, Saunders DGO, Sklenar J, Lorrain C, Krasileva KV, Win J, Duplessis S, Kamoun S. **2016**. Heterologous expression screens in Nicotiana benthamiana identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLOS ONE* **11**: e0149035.

Peyret H, Gehin A, Thuenemann EC, Blond D, Turabi AE, Beales L, Clarke D, Gilbert RJC, Fry EE, Stuart DI, *et al.* 2015. Tandem Fusion of Hepatitis B Core Antigen Allows Assembly of Virus-Like Particles in Bacteria and Plants with Enhanced Capacity to Accommodate Foreign Proteins. *PLOS ONE* **10**: e0120751.

Peyret H, Lomonossoff GP. **2015**. When plant virology met Agrobacterium: the rise of the deconstructed clones. *Plant Biotechnology Journal* **13**: 1121–1135.

Pillay P, Kibido T, du Plessis M, van der Vyver C, Beyene G, Vorster BJ, Kunert KJ, Schlüter U. **2012**. Use of transgenic Oryzacystatin-I-expressing plants enhances recombinant protein production. *Applied Biochemistry and Biotechnology* **168**: 1608–1620.

Pillay P, Kunert KJ, Wyk SG van, Makgopa ME, Cullis CA, Vorster BJ. **2016**. Agroinfiltration contributes to VP1 recombinant protein degradation. *Bioengineered* **7**: 459–477.

Pillet S, Aubin É, Trépanier S, Bussière D, Dargis M, Poulin J-F, Yassine-Diab B, Ward BJ, Landry N. **2016**. A plant-derived quadrivalent virus like particle influenza vaccine induces cross-reactive antibody and T cell response in healthy adults. *Clinical Immunology* **168**: 72–87.

Pitzschke A. **2013**. Agrobacterium infection and plant defense—transformation success hangs by a thread. *Frontiers in Plant Science* **4**: 519.

Pitzschke A, Hirt H. **2010**. New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. *The EMBO Journal* **29**: 1021–1032.

Pouvreau L, Gruppen H, Piersma SR, van den Broek LAM, van Koningsveld GA, Voragen AGJ. **2001**. Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. *Journal of Agricultural and Food Chemistry* **49**: 2864–2874.

Pruss GJ, Nester EW, Vance V. **2008**. Infiltration with Agrobacterium tumefaciens induces host defense and development-dependent responses in the infiltrated zone. *Molecular Plant-Microbe Interactions* **21**: 1528–1538.

Puchta H. **2017**. Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Current Opinion in Plant Biology* **36**: 1–8.

Qiu X, Wong G, Audet J, Bello A, Fernando L, Alimonti JB, Fausther-Bovendo H, Wei H, Aviles J, Hiatt E, *et al.* 2014. Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature* **514**: 47–53.

Qu L-J, Chen J, Liu M, Pan N, Okamoto H, Lin Z, Li C, Li D, Wang J, Zhu G, *et al.* 2003. Molecular cloning and functional analysis of a novel type of Bowman-Birk inhibitor gene family in rice. *Plant Physiology* **133**: 560–570.

QUT. **2017**. Nicotiana benthamiana Genome & Wild Transcriptome Site, Queensland University of Technology. *http://sefapps02.qut.edu.au/benWeb/subpages/strategy.php*.

R Core Team. **2015**. R: A language and environment for statistical computing. *https://www.R-project.org/*.

Rahim MD, Andika IB, Han C, Kondo H, Tamada T. **2007**. RNA4-encoded p31 of beet necrotic yellow vein virus is involved in efficient vector transmission, symptom severity and silencing suppression in roots. *Journal of General Virology* **88**: 1611–1619.

Rakwal R, Kumar Agrawal G, Jwa N-S. **2001**. Characterization of a rice (Oryza sativa L.) Bowman–Birk proteinase inhibitor: tightly light regulated induction in response to cut, jasmonic acid, ethylene and protein phosphatase 2A inhibitors. *Gene* **263**: 189–198.

Rasoolizadeh A, Goulet M-C, Sainsbury F, Cloutier C, Michaud D. **2016a**. Single substitutions to closely related amino acids contribute to the functional diversification of an insect-inducible, positively selected plant cystatin. *FEBS Journal* **283**: 1323–1335.

Rasoolizadeh A, Munger A, Goulet M-C, Sainsbury F, Cloutier C, Michaud D. **2016b**. Functional proteomics-aided selection of protease inhibitors for herbivore insect control. *Scientific Reports* **6**: 38827.

Rawlings ND. **2016**. Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation. *Biochimie* **122**: 5–30.

Rawlings ND, Waller M, Barrett AJ, Bateman A. **2014**. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **42**: D503–D509.

Rehman AA, Ahsan H, Khan FH. **2013**. α2-macroglobulin: A physiological guardian. *Journal of Cellular Physiology* **228**: 1665–1675.

Renko M, Sabotič J, Turk D. **2012**. β-Trefoil inhibitors – from the work of Kunitz onward. *Biological Chemistry* **393**: 1043–1054.

Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stuber K, Colby T, Overkleeft HS, Bogyo M, Van der Hoorn RAL. **2012**. Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiology* **158**: 1583–1599.

Rico A, Bennett MH, Forcat S, Huang WE, Preston GM. **2010**. Agroinfiltration reduces ABA levels and suppresses Pseudomonas syringae-elicited salicylic acid production in Nicotiana tabacum. *PLoS ONE* **5**: e8977.

Robert S, Goulet M-C, D'Aoust M-A, Sainsbury F, Michaud D. **2015**. Leaf proteome rebalancing in Nicotiana benthamiana for upstream enrichment of a transiently expressed recombinant protein. *Plant Biotechnology Journal* **13**: 1169–1179.

Robert S, Khalf M, Goulet M-C, D'Aoust M-A, Sainsbury F, Michaud D. **2013**. Protection of recombinant mammalian antibodies from development-dependent proteolysis in leaves of Nicotiana benthamiana (S Park, Ed.). *PLoS ONE* **8**: e70203.

Roberts TH. **2003**. Differential gene expression for suicide-substrate serine proteinase inhibitors (serpins) in vegetative and grain tissues of barley. *Journal of Experimental Botany* **54**: 2251–2263.

Roberts TH, Hejgaard J. **2007**. Serpins in plants and green algae. *Functional & Integrative Genomics* **8**: 1–27.

Roberts TH, Hejgaard J, Saunders NFW, Cavicchioli R, Curmi PMG. **2004**. Serpins in unicellular eukarya, archaea, and bacteria: sequence analysis and evolution. *Journal of Molecular Evolution* **59**: 437–447.

Roberts TH, Marttila S, Rasmussen SK, Hejgaard J. **2003**. Differential gene expression for suicide-substrate serine proteinase inhibitors (serpins) in vegetative and grain tissues of barley. *Journal of Experimental Botany* **54**: 2251–2263.

Robinette D, Matthysse AG. **1990**. Inhibition by Agrobacterium tumefaciens and Pseudomonas savastanoi of development of the hypersensitive response elicited by Pseudomonas syringae pv. phaseolicola. *Journal of bacteriology* **172**: 5742–5749.

Rodis P, Hoff JE. **1984**. Naturally occurring protein crystals in the potato. *Plant Physiology* **74**: 907–911.

Rojo E, Zouhar J, Carter C, Kovaleva V, Raikhel NV. **2003**. A unique mechanism for protein processing and degradation in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 7389–7394.

Rose R, Schaller A, Ottmann C. **2010**. Structural features of plant subtilases. *Plant Signaling & Behavior* **5**: 180–183.

Russell DA, Spatola LA, Dian T, Paradkar VM, Dufield DR, Carroll JA, Schlittler MR. **2005**. Host limits to accurate human growth hormone production in multiple plant systems. *Biotechnology and Bioengineering* **89**: 775–782.

Rybicki E, Hitzeroth I, Meyers A, Santos M, Wigdorovitz A. **2013**. Developing country applications of molecular farming: case studies in South Africa and Argentina. *Current Pharmaceutical Design* **19**: 5612–5621.

Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG. **1995**. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry* **34**: 4147–4158.

Sainsbury F, Lomonossoff GP. **2008**. Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiology* **148**: 1212–1218.

Sainsbury F, Saunders K, Aljabali AAA, Evans DJ, Lomonossoff GP. **2011**. Peptide-controlled access to the interior surface of empty virus nanoparticles. *ChemBioChem* **12**: 2435–2440.

Sainsbury F, Thuenemann EC, Lomonossoff GP. **2009**. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* **7**: 682–693.

Sainsbury F, Varennes-Jutras P, Goulet M-C, D'Aoust M-A, Michaud D. **2013**. Tomato cystatin SlCYS8 as a stabilizing fusion partner for human serpin expression in plants. *Plant Biotechnology Journal* **11**: 1058–1068.

Salomon S, Puchta H. **1998**. Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *The EMBO Journal* **17**: 6086–6095.

Sansen S, Ranter CJD, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A. **2004**. Structural basis for inhibition of Aspergillus niger xylanase by Triticum aestivum xylanase inhibitor-I. *Journal of Biological Chemistry* **279**: 36022–36028.

Santamaría ME, Diaz-Mendoza M, Diaz I, Martinez M. **2014**. Plant protein peptidase inhibitors: an evolutionary overview based on comparative genomics. *BMC Genomics* **15**: 812–826.

Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ. **2007**. An asparaginyl endopeptidase mediates in vivo protein backbone cyclization. *Journal of Biological Chemistry* **282**: 29721–29728.

Saur IML, Kadota Y, Sklenar J, Holton NJ, Smakowska E, Belkhadir Y, Zipfel C, Rathjen JP. **2016**. NbCSPR underlies age-dependent immune responses to bacterial cold shock protein in Nicotiana benthamiana. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 3389–3394.

Scanlon MJ, Lee MC, Anderson MA, Craik DJ. **1999**. Structure of a putative ancestral protein encoded by a single sequence repeat from a multidomain proteinase inhibitor gene from Nicotiana alata. *Structure* **7**: 793–802.

Schardon K, Hohl M, Graff L, Pfannstiel J, Schulze W, Stintzi A, Schaller A. **2016**. Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science* **354**: 1594–1597.

Schiermeyer A, Schinkel H, Apel S, Fischer R, Schillberg S. **2005**. Production of Desmodus rotundus salivary plasminogen activator α1 (DSPAα1) in tobacco is hampered by proteolysis. *Biotechnology and Bioengineering* **89**: 848–858.

Schneider CA, Rasband WS, Eliceiri KW. **2012**. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Schoberer J, Strasser R. **2017**. Plant glyco-biotechnology. *Seminars in Cell & Developmental Biology*: 1084–9521.

Serrano I, Buscaill P, Audran C, Pouzet C, Jauneau A, Rivas S. **2016**. A non canonical subtilase attenuates the transcriptional activation of defence responses in Arabidopsis thaliana. *eLife* **5**: e19755.

Shabab M, Shindo T, Gu C, Kaschani F, Pansuriya T, Chintha R, Harzen A, Colby T, Kamoun S, Hoorn RAL van der. **2008**. Fungal Effector Protein AVR2 Targets Diversifying Defense-Related Cys Proteases of Tomato. *The Plant Cell* **20**: 1169–1183.

Sheikh AH, Raghuram B, Eschen-Lippold L, Scheel D, Lee J, Sinha AK. **2014**. Agroinfiltration by cytokinin-producing Agrobacterium sp. strain GV3101 primes defense responses in Nicotiana tabacum. *Molecular Plant-Microbe Interactions* **27**: 1175–1185.

Shi Y, Lee L-Y, Gelvin SB. **2014**. Is VIP1 important for Agrobacterium-mediated transformation? *The Plant Journal* **79**: 848–860.

Shigeyama T, Watanabe A, Tokuchi K, Toh S, Sakurai N, Shibuya N, Kawakami N. **2016**. α-Xylosidase plays essential roles in xyloglucan remodelling, maintenance of cell wall integrity, and seed germination in Arabidopsis thaliana. *Journal of Experimental Botany* **67**: 5615–5629.

Shindo T, Misas-Villamil JC, Hörger AC, Song J, van der Hoorn RAL. **2012**. A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14. *PLoS ONE* **7**: e29317.

Shivaraj B, Pattabiraman TN. **1981**. Natural plant enzyme inhibitors. Characterization of an unusual α-amylase/trypsin inhibitor from ragi (Eleusine coracana Geartn.). *Biochemical Journal* **193**: 29–36.

Simpson JT, Durbin R. **2012**. Efficient de novo assembly of large genomes using compressed data structures. *Genome Research* **22**: 549–556.

Singer K, Shiboleth YM, Li J, Tzfira T. **2012**. Formation of complex extrachromosomal T-DNA structures in Agrobacterium tumefaciens-infected plants. *Plant Physiology* **160**: 511–522.

Siqueira-Júnior CL, Fernandes KVS, Machado OLT, da Cunha M, Gomes VM, Moura D, Jacinto T. **2002**. 87 kDa tomato cystatin exhibits properties of a defense protein and forms protein crystals in prosystemin overexpressing transgenic plants. *Plant Physiology and Biochemistry* **40**: 247–254.

Sonawane PD, Pollier J, Panda S, Szymanski J, Massalha H, Yona M, Unger T, Malitsky S, Arendt P, Pauwels L, *et al.* 2016. Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nature Plants* **3**: 16205.

Song HK, Kim YS, Yang JK, Moon J, Lee JY, Suh SW. **1999**. Crystal structure of a 16 kda double-headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *Journal of Molecular Biology* **293**: 1133–1144.

Sottrup-Jensen L. **1989**. α-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *Journal of Biological Chemistry* **264**: 11539–11542.

Sottrup-Jensen L, Sand O, Kristensen L, Fey GH. **1989**. The α-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian alpha-macroglobulins. *Journal of Biological Chemistry* **264**: 15781–15789.

Stapleton A, Beetham JK, Pinot F, Garbarino JE, Rockhold DR, Friedman M, Hammock BD, Belknap WR. **1994**. Cloning and expression of soluble epoxide hydrolase from potato. *The Plant Journal: For Cell and Molecular Biology* **6**: 251–258.

Stoger E, Fischer R, Moloney M, Ma JK-C. **2014**. Plant molecular pharming for the treatment of chronic and infectious diseases. *Annual Review of Plant Biology* **65**: 743–768.

Streatfield SJ, Mayor JM, Barker DK, Brooks C, Lamphear BJ, Woodard SL, Beifuss KK, Vicuna DV, Massey LA, Horn ME, *et al.* 2002. Development of an edible subunit vaccine in corn against enterotoxigenic strains of escherichia coli. *In Vitro Cellular & Developmental Biology - Plant* **38**: 11–17.

Strobl S, Maskos K, Wiegand G, Huber R, Gomis-Rüth FX, Glockshuber R. **1998**. A novel strategy for inhibition of α-amylases: yellow meal worm α-amylase in complex with the Ragi bifunctional inhibitor at 2.5 Å resolution. *Structure* **6**: 911–921.

Strobl S, Muehlhahn P, Bernstein R, Wiltscheck R, Maskos K, Wunderlich M, Huber R, Glockshuber R, Holak TA. **1995**. Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin inhibitor from Ragi seeds by NMR spectroscopy. *Biochemistry* **34**: 8281–8293.

Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, Turk V. **1990**. The refined 2.4 Å x-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *The EMBO Journal* **9**: 1939–1947.

Sueldo D, Ahmed A, Misas-Villamil J, Colby T, Tameling W, Joosten MHAJ, van der Hoorn RAL. **2014**. Dynamic hydrolase activities precede hypersensitive tissue collapse in tomato seedlings. *New Phytologist* **203**: 913–925.

Svozil J, Hirsch-Hoffmann M, Dudler R, Gruissem W, Baerenfaller K. **2014**. Protein abundance changes and ubiquitylation targets identified after inhibition of the proteasome with syringolin A. *Molecular & Cellular Proteomics* **13**: 1523–1536.

Sweet RM, Wright HT, Janin J, Chothia CH, Blow DM. **1974**. Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-Å resolution. *Biochemistry* **13**: 4212–4228.

Szalontai B, Stranczinger S, Palfalvi G, Mauch-Mani B, Jakab G. **2012**. The taxon-specific paralogs of grapevine PRLIP genes are highly induced upon powdery mildew infection. *Journal of Plant Physiology* **169**: 1767–1775.

Tajima T, Yamaguchi A, Matsushima S, Satoh M, Hayasaka S, Yoshimatsu K, Shioi Y. **2011**. Biochemical and molecular characterization of senescence-related cysteine protease–cystatin complex from spinach leaf. *Physiologia Plantarum* **141**: 97–116.

Tamaki T, Betsuyaku S, Fujiwara M, Fukao Y, Fukuda H, Sawa S. **2013**. SUPPRESSOR OF LLP1 1-mediated C–terminal processing is critical for CLE19 peptide activity. *The Plant Journal* **76**: 970–981.

Tamhane VA, Chougule NP, Giri AP, Dixit AR, Sainani MN, Gupta VS. **2005**. In vivo and in vitro effect of Capsicum annum proteinase inhibitors on Helicoverpa armigera gut proteinases. *Biochimica Et Biophysica Acta* **1722**: 156–167.

Tamhane VA, Giri AP, Kumar P, Gupta VS. **2009**. Spatial and temporal expression patterns of diverse Pin-II proteinase inhibitor genes in Capsicum annuum Linn. *Gene* **442**: 88–98.

Tang S, Lomsadze A, Borodovsky M. **2015**. Identification of protein coding regions in RNA transcripts. *Nucleic Acids Research* **43**: e78–e78.

Taylor A, Qiu Y-L. **2017**. Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Molecular Plant-Microbe Interactions* **30**: 489–501.

Teh AY, Maresch D, Klein K, Ma JK. **2014**. Characterization of VRC01, a potent and broadly neutralizing anti-HIV mAb, produced in transiently and stably transformed tobacco. *Plant Biotechnology Journal* **12**: 300–311.

The UniProt Consortium. **2017**. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45**: D158–D169.

Thomas JC, Adams DG, Keppenne VD, Wasmann CC, Brown JK, Kanost MR, Bohnert HJ. **1995**. Protease inhibitors of Manduca sexta expressed in transgenic cotton. *Plant Cell Reports* **14**: 758–762.

Thomas JC, Wasmann CC, Echt C, Dunn RL, Bohnert HJ, McCoy TJ. **1994**. Introduction and expression of an insect proteinase inhibitor in alfalfa Medicago sativa L. *Plant Cell Reports* **14**: 31–36.

Thomashow MF, Hugly S, Buchholz WG, Thomashow LS. **1986**. Molecular basis for the auxin-independent phenotype of crown gall tumor tissues. *Science* **231**: 616–618.

Thongyoo P, Bonomelli C, Leatherbarrow RJ, Tate EW. **2009**. Potent inhibitors of β-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *Journal of Medicinal Chemistry* **52**: 6197–6200.

Thuenemann EC, Meyers AE, Verwey J, Rybicki EP, Lomonossoff GP. **2013**. A method for rapid production of heteromultimeric protein complexes in plants: assembly of protective bluetongue virus-like particles. *Plant Biotechnology Journal* **11**: 839–846.

Tian M, Huitema E, Cunha L da, Torto-Alalibo T, 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**: 26370–26377.

Tian M, Win J, Song J, van der Hoorn R, Knaap E van der, Kamoun S. **2007**. A Phytophthora infestans cystatin-like protein targets a novel tomato papain-like apoplastic protease. *Plant Physiology* **143**: 364–377.

Tripathi LP, Sowdhamini R. **2006**. Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics* **7**: 200.

Tusé D, Ku N, Bendandi M, Becerra C, Collins R, Langford N, Sancho SI, López-Díaz de Cerio A, Pastor F, Kandzia R, *et al.* 2015. Clinical Safety and Immunogenicity of Tumor-Targeted, Plant-Made Id-KLH Conjugate Vaccines for Follicular Lymphoma. *BioMed Research International*.

Tyanova S, Temu T, Cox J. **2016a**. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* **11**: 2301–2319.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. **2016b**. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**: 731–740.

UNDG. **2015**. Socio-Economic Impact of Ebola Virus Disease in West African Countries A call for national and regional containment, recovery and prevention. *http://www.africa.undp.org/content/dam/rba/docs/Reports/ebola-west-africa.pdf*.

Uppalapati SR, Ayoubi P, Weng H, Palmer DA, Mitchell RE, Jones W, Bender CL. **2005**. The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. *The Plant Journal* **42**: 201–217.

Urwin PE, McPherson MJ, Atkinson HJ. **1998**. Enhanced transgenic plant resistance to nematodes by dual proteinase inhibitor constructs. *Planta* **204**: 472–479.

Vallée F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R. **1998**. Barley α-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 Å resolution. *Structure* **6**: 649–659.

Valueva TA, Revina TA, Mosolov VV, Mentele R. **2000**. Primary structure of potato kunitz-type serine proteinase inhibitor. *Biological Chemistry* **381**: 1215–1221.

Vancanneyt G, Schmidt R, O'Connor-Sanchez A, Willmitzer L, Rocha-Sosa M. **1990**. Construction of an intron-containing marker gene: Splicing of the intron in transgenic plants and its use in monitoring early events in Agrobacterium-mediated plant transformation. *Molecular and General Genetics MGG* **220**: 245–250.

Vaquero C, Sack M, Chandler J, Drossard J, Schuster F, Monecke M, Schillberg S, Fischer R. **1999**. Transient expression of a tumor-specific single-chain fragment and a chimeric antibody in tobacco leaves. *Proceedings of the National Academy of Sciences* **96**: 11128–11133.

Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan J-A, De Rycke R, Brackenier A, Inzé D, Harris JL, Van Breusegem F. **2006**. Serpin1 of Arabidopsis thaliana is a suicide inhibitor for Metacaspase 9. *Journal of Molecular Biology* **364**: 625–636.

Villela-Dias C, Camillo LR, de Oliveira GAP, Sena JAL, Santiago AS, de Sousa STP, Mendes JS, Pirovani CP, Alvim FC, Costa MGC. **2014**. Nep1-like protein from Moniliophthora perniciosa induces a rapid proteome and metabolome reprogramming in cells of Nicotiana benthamiana. *Physiologia Plantarum* **150**: 1–17.

Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, *et al.* 2016. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Research* **44**: D447–D456.

Vorster J, Rasoolizadeh A, Goulet M-C, Cloutier C, Sainsbury F, Michaud D. **2015**. Positive selection of digestive Cys proteases in herbivorous Coleoptera. *Insect Biochemistry and Molecular Biology* **65**: 10–19.

Voss RH, Ermler U, Essen LO, Wenzl G, Kim YM, Flecker P. **1996**. Crystal structure of the bifunctional soybean Bowman-Birk inhibitor at 0.28-nm resolution. Structural peculiarities in a folded protein conformation. *European Journal of Biochemistry* **242**: 122–131.

Wagner UG, Hasslacher M, Griengl H, Schwab H, Kratky C. **1996**. Mechanism of cyanogenesis: the crystal structure of hydroxynitrile lyase from Hevea brasiliensis. *Structure* **4**: 811–822.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**: e112963.

Walsh TA, Strickland JA. **1993**. Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. *Plant Physiology* **103**: 1227–1234.

Wang L, Lacroix B, Guo J, Citovsky V. **2017**. The Agrobacterium VirE2 effector interacts with multiple members of the Arabidopsis VIP1 protein family. *Molecular Plant Pathology* **in press**.

Wang D, Weaver ND, Kesarwani M, Dong X. **2005**. Induction of protein secretory pathway is required for systemic acquired resistance. *Science* **308**: 1036–1040.

Weeda SM, Mohan Kumar GN, Richard Knowles N. **2009**. Developmentally linked changes in proteases and protease inhibitors suggest a role for potato multicystatin in regulating protein content of potato tubers. *Planta* **230**: 73–84.

Wen AM, Shukla S, Saxena P, Aljabali AAA, Yildiz I, Dey S, Mealy JE, Yang AC, Evans DJ, Lomonossoff GP, *et al.* 2012. Interior engineering of a viral nanoparticle and its tumor homing properties. *Biomacromolecules* **13**: 3990–4001.

WHO ERT. **2016**. After Ebola in West Africa — Unpredictable Risks, Preventable Epidemics. *New England Journal of Medicine* **375**: 587–596.

Wickham H. **2009**. Introduction. In: Use R. ggplot2. Springer, New York, NY, 1–7.

Wingfield PT, Sax JK, Stahl SJ, Kaufman J, Palmer I, Chung V, Corcoran ML, Kleiner DE, Stetler-Stevenson WG. **1999**. Biophysical and Functional Characterization of Full-length, Recombinant Human Tissue Inhibitor of Metalloproteinases-2 (TIMP-2) Produced in Escherichia coli COMPARISON OF WILD TYPE AND AMINO-TERMINAL ALANINE APPENDED VARIANT WITH IMPLICATIONS FOR THE MECHANISM OF TIMP FUNCTIONS. *Journal of Biological Chemistry* **274**: 21362–21368.

Wu X, Yang Z-Y, Li Y, Hogerkorp C-M, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, *et al.* 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* **329**: 856–861.

Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, Dixon RA, Lamb C. **2004**. An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. *The EMBO Journal* **23**: 980–988.

Xu D, Xue Q, McElroy D, Mawal Y, Hilder VA, Wu R. **1996**. Constitutive expression of a cowpea trypsin inhibitor gene, CpTi, in transgenic rice plants confers resistance to two major rice insect pests. *Molecular Breeding* **2**: 167–173.

Yamada K, Matsushima R, Nishimura M, 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**: 1626–1634.

Yamada T, Ogawa VA, Freire M, others. **2016**. Security spending must cover disease outbreaks. *https://www.nature.com/polopoly_fs/1.19836.1462381441!/menu/main/topColumns/topLeftColumn/pdf/533029a.pdf?origin=ppub*.

Yamagata H, Kunimatsu K, Kamasaka H, Kuramoto T, Iwasaki T. **1998**. Rice bifunctional α-amylase/subtilisin inhibitor: characterization, localization, and changes in developing and germinating seeds. *Bioscience, Biotechnology, and Biochemistry* **62**: 978–985.

Yamauchi Y, Ejiri Y, Toyoda Y, Tanaka K. **2003**. Identification and biochemical characterization of plant acylamino acid–releasing enzyme. *Journal of Biochemistry* **134**: 251–257.

Yao R, Ming Z, Yan L, Li S, Wang F, Ma S, Yu C, Yang M, Chen L, Chen L, *et al.* 2016. DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature* **536**: 469–473.

Yoo B-C, Aoki K, Xiang Y, Campbell LR, Hull RJ, Xoconostle-Cazares B, Monzer J, Lee J-Y, Ullman DE, Lucas WJ. **2000**. Characterization of Cucurbita maxima Phloem Serpin-1 (CmPS-1): A developmentally regulated elastase inhibitor. *Journal of Biological Chemistry* **275**: 35122–35128.

Yusibov V, Kushnir N, Streatfield SJ. **2016**. Antibody production in plants and green algae. *Annual Review of Plant Biology* **67**: 669–701.

Zakharov A, Carchilan M, Stepurina T, Rotari V, Wilson K, Vaintraub I. **2004**. A comparative study of the role of the major proteinases of germinated common bean (Phaseolus vulgaris L.) and soybean (Glycine max (L.) Merrill) seeds in the degradation of their storage proteins. *Journal of Experimental Botany* **55**: 2241–2249.

Zhang C, Bradshaw JD, Whitham SA, Hill JH. **2010**. The development of an efficient multipurpose bean pod mottle virus viral vector set for foreign gene expression and RNA silencing. *Plant Physiology* **153**: 52–65.

Zhang X, Buehner NA, Hutson AM, Estes MK, Mason HS. **2006**. Tomato is a highly effective vehicle for expression and oral immunization with Norwalk virus capsid protein. *Plant Biotechnology Journal* **4**: 419–432.

Zhang B, Rapolu M, Kumar S, Gupta M, Liang Z, Han Z, Williams P, Su WW. **2017**. Coordinated protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide. *Plant Biotechnology Journal* **15**: 718–728.

Zhao Y, Thilmony R, Bender CL, Schaller A, He SY, Howe GA. **2003**. Virulence systems of Pseudomonas syringae pv. tomato promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. *The Plant Journal* **36**: 485–499.

Zheng Z, Guo Y, Novák O, Chen W, Ljung K, Noel JP, Chory J. **2016**. Local auxin metabolism regulates environment-induced hypocotyl elongation. *Nature Plants* **2**: 16025.

Zhou Y, Cox AM, Kearney CM. **2017**. Pathogenesis-related proteins induced by agroinoculation-associated cell wall weakening can be obviated by spray-on inoculation or mannitol ex vivo culture. *Plant Biotechnology Reports* **11**: 1–9.

Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, Boller T, Felix G. **2006**. Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. *Cell* **125**: 749–760.

Zischewski J, Sack M, Fischer R. **2016**. Overcoming low yields of plant-made antibodies by a protein engineering approach. *Biotechnology Journal* **11**: 107–116.

Appendix: manuscripts from Chapters 2 and 3 as published

Plant Biotechnology Journal

Plant Biotechnology Journal (2017), pp. 1–17 doi: 10.1111/pbi.12852

The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire

Friederike Grosse-Holz[1], Steven Kelly[2], Svenja Blaskowski[3], Farnusch Kaschani[3], Markus Kaiser[3] and Renier A.L. van der Hoorn[1,*]

[1]*Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, Oxford, UK*

[2]*Department of Plant Sciences, University of Oxford, Oxford, UK*

[3]*Chemische Biologie, Zentrum für Medizinische Biotechnologie, Fakultät für Biologie, Universität Duisburg-Essen, Essen, Germany*

Received 28 June 2017;
revised 6 October 2017;
accepted 15 October 2017.
**Correspondence* (Tel +44 (0) 1865 275077; fax: +44 (0) 1865 275074; email renier.vanderhoorn@plants.ox.ac.uk)

Keywords: activity-based protein profiling, *Agrobacterium tumefaciens*, chlorosis, plant protease annotation, post-translational activation, silencing inhibitor p19.

Summary

Infiltration of disarmed *Agrobacterium tumefaciens* into leaves of *Nicotiana benthamiana* (agroinfiltration) facilitates quick and safe production of antibodies, vaccines, enzymes and metabolites for industrial use (molecular farming). However, yield and purity of proteins produced by agroinfiltration are hampered by unintended proteolysis, restricting industrial viability of the agroinfiltration platform. Proteolysis may be linked to an immune response to agroinfiltration, but understanding of the response to agroinfiltration is limited. To identify the proteases, we studied the transcriptome, extracellular proteome and active secretome of agroinfiltrated leaves over a time course, with and without the P19 silencing inhibitor. Remarkably, the P19 expression had little effect on the leaf transcriptome and no effect on the extracellular proteome. 25% of the detected transcripts changed in abundance upon agroinfiltration, associated with a gradual up-regulation of immunity at the expense of photosynthesis. By contrast, 70% of the extracellular proteins increased in abundance, in many cases associated with increased efficiency of extracellular delivery. We detect a dynamic reprogramming of the proteolytic machinery upon agroinfiltration by detecting transcripts encoding for 975 different proteases and protease homologs. The extracellular proteome contains peptides derived from 196 proteases and protease homologs, and activity-based proteomics displayed 17 active extracellular Ser and Cys proteases in agroinfiltrated leaves. We discuss unique features of the *N. benthamiana* protease repertoire and highlight abundant extracellular proteases in agroinfiltrated leaves, being targets for reverse genetics. This data set increases our understanding of the plant response to agroinfiltration and indicates ways to improve a key expression platform for both plant science and molecular farming.

Introduction

Agroinfiltration of *Nicotiana benthamiana* (a relative of tobacco) is widely applied to transiently express proteins, either as biopharmaceutcials, for other industrial use or to study their functions. Agroinfiltration is based on the transient genetic manipulation of leaves by infiltration with disarmed *Agrobacterium tumefaciens* (Agrobacterium) carrying gene(s) of interest on the transfer DNA (T-DNA) of binary plasmid(s) (Bevan, 1984). Agrobacterium delivers the T-DNA to the nucleus of its host plant, where genes are expressed within a few days upon agroinfiltration. Co-expression of several transgenes is simply achieved by mixing Agrobacterium cultures delivering these different transgenes before agroinfiltration. Co-expression with silencing inhibitor P19 is frequently used to boost protein overexpression by preventing the decline of the transgene transcript levels (Van der Hoorn *et al.*, 2003).

The versatility and potential of agroinfiltration are illustrated by many use cases. For instance, production of biopharmaceuticals (molecular farming) (Stoger *et al.*, 2014) in agroinfiltrated *N. benthamiana* offers speed, scalability and low risk of contamination with human pathogens when compared to classical insect or mammalian cell culture systems. An agroinfiltration-based expression platform can now deliver ten million doses of the latest influenza vaccine within a record time of 6 weeks (Pillet *et al.*, 2016). Large-scale agroinfiltration has also produced many different functional monoclonal antibodies (Yusibov *et al.*, 2016), including the Ebola neutralizing drug ZMapp (Qiu *et al.*, 2014). Transient, spatially restricted overexpression of synthetic biology building blocks can shift plant secondary metabolism towards valuable products with minor impact on fitness (Nielsen *et al.*, 2013). Along similar lines, pathogen-derived effectors that would likely have severe phenotypic effects if expressed in stable lines have been studied by agroinfiltration (Bos *et al.*, 2006; Dagdas *et al.*, 2016; Petre *et al.*, 2016). Speed and simplicity of agroinfiltration are leveraged for high-throughput screening of fluorescently tagged proteins to study their subcellular localization (Martin *et al.*, 2009).

Please cite this article as: Grosse-Holz, F., Kelly, S., Blaskowski, S., Kaschani, F., Kaiser, M. and van der Hoorn, R.A.L. (2017) The transcriptome, extracellular proteome and active secretome of agroinfiltrated *Nicotiana benthamiana* uncover a large, diverse protease repertoire. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.12852

Although agroinfiltration is a widely used tool, remarkably little is known about how *N. benthamiana* responds to agroinfiltration. Agrobacterium elicits immune responses, including the induction of pathogenesis-related (PR) genes and the accumulation of extracellular PR proteins (Goulet *et al.*, 2010; Pitzschke, 2013; Zhou *et al.*, 2017). As in other plants, this immune response reduces subsequent pathogen infections (Li *et al.*, 2017; Rico *et al.*, 2010; Robinette and Matthysse, 1990; Sheikh *et al.*, 2014) and may limit transgene delivery. Transgene delivery in older, flowering *N. benthamiana* is limited due to the perception of Agrobacterium cold-shock protein (Saur *et al.*, 2016). In younger plants, which are used for agroinfiltration, responses are elusive. Furthermore, the impact of silencing inhibitor P19 on the response to agroinfiltration and its timing are unresolved.

We focus on extracellular proteases, as they may limit the accumulation of recombinant proteins (RPs) passing through the secretory pathway to become glycosylated. Proteolytic degradation is a bottleneck on the way to industrial viability of agroinfiltration (Mandal *et al.*, 2016). Indeed, RP degradation can occur in the extracellular space (Hehle *et al.*, 2011) and proteolysis hampers yield and purity of biopharmaceuticals produced in *N. benthamiana* (Hehle *et al.*, 2015; Mandal *et al.*, 2014; Niemer *et al.*, 2014). Papain-like Cys proteases can degrade RPs *in vitro* (Paireder *et al.*, 2016, 2017), but the proteases degrading RP *in planta* are unidentified. Extracellular proteases commonly accumulate in leaves during immune responses. The extracellular tomato Ser protease P69 and Cys proteases Pip1 and Rcr3, for example, accumulate upon infection with viroids, oomycetes, fungi and bacteria (Jordá *et al.*, 1999; Kaschani *et al.*, 2010; Tian *et al.*, 2004). Transcripts and proteins corresponding to proteases also accumulate in Arabidopsis infected with *Pseudomonas* (Xia *et al.*, 2004; Zhao *et al.*, 2003), and extracellular Ser and Cys protease activities increase in tomato upon fungal infection with *Cladosporium fulvum* (van Esse *et al.*, 2008; Sueldo *et al.*, 2014). These examples indicate that activity and/or abundance of extracellular proteases, especially Ser and Cys proteases, may increase in *N. benthamiana* upon agroinfiltration, linking proteolytic RP degradation to plant immunity. Therefore, both comprehensive annotation of the *N. benthamiana* protease repertoire and improved understanding of the response to agroinfiltration are needed to limit undesired proteolysis. RP accumulation has been increased by depleting proteases by knockdown in rice cell cultures (Kim *et al.*, 2008) and in *Nicotiana tabacum* (Duwadi *et al.*, 2015; Mandal *et al.*, 2014) and protease inhibitor overexpression in *N. benthamiana* (Goulet *et al.*, 2012; Sainsbury *et al.*, 2013). These studies indicate that once targets are identified, protease depletion could improve agroinfiltrated *N. benthamiana* as a protein expression platform.

Here, we investigated how RP production may be affected by the immune response to agroinfiltration, especially immune proteases. Time-resolved leaf transcriptome and extracellular proteome data sets of agroinfiltrated leaves revealed an immune response that is mounted at the expense of photosynthesis and not affected by P19. We analysed the exceptionally large *N. benthamiana* protease repertoire in the context of other plant proteases and identified active Ser and Cys proteases. Taken together, the data will advance strategies to improve transient protein expression by engineering plant immunity and depleting proteases.

Results and discussion

To characterize agroinfiltrated *N. benthamiana* leaves, we infiltrated *N. benthamiana* leaves with wild-type *A. tumefaciens* GV3101-pMP90 (no binary vector, WT), Agrobacterium P19 (T-DNA encoding viral silencing suppressor P19 (Chapman *et al.*, 2004), P19) or buffer (mock treatment). We took samples at two, five, seven and 10 days postinfiltration (dpi) and quantified transcripts, extracellular proteins and extracellular protein activity using RNAseq, label-free quantification mass spectrometry (MS) and activity-based proteomics (activity-based protein profiling coupled to mass spectrometry, ABPP-MS) (Figure 1).

During a first annotation of the transcriptome and proteome data, we observed that well-known proteases including papain-like Cys proteases (PLCPs, MEROPS family C01) and subtilases (family S08) often appeared truncated or lacked conserved domains in the Niben101 proteome database (https://solgenomics.net/). To obtain a database with protease families that are adequately annotated for the evaluation of transcriptomics and proteomics experiments, we compared four *N. benthamiana* proteome databases and manually curated the proteases in the best database (described in detail in Appendix S1). Searching the extracellular proteome MS spectra with our curated proteome, we identified peptides corresponding to 30 proteins more than with the best published database, showing that the curation improved interpretation of experimental data (Appendix S1).

The *N. benthamiana* response to agroinfiltration

The P19 silencing suppressor has minor effects on the transcriptome and no effect on the extracellular proteome of N. benthamiana

To assess how *N. benthamiana* responds to agroinfiltration and how silencing suppression affects these responses, we sequenced mRNA from WT agroinfiltrated, P19 agroinfiltrated and mock-infiltrated leaves. Euclidean distance clustering revealed that transcriptomes from agroinfiltrated samples cluster together by time point regardless of whether WT or P19 bacteria were present (Figure 2a). Surprisingly, only 0.75% of all detected transcripts (569/75802) differed significantly in abundance at any time point between WT and P19 agroinfiltrated leaves (Table S1). Among the differentials is the transcript encoding P19, which was very abundant up to 7 dpi and slightly decreased in abundance at 10 dpi, potentially because older leaves are less transcriptionally active (Figure S1). Transcripts encoding components of the silencing machinery such as members of the Argonaute PFAM family were significantly enriched among the transcripts with differential abundance between P19 and WT agroinfiltrated leaves, but most of the differential transcripts (404 of 569) are not annotated (Figure S1 and Table S2). There were no significant differences between extracellular proteomes from WT and P19 agroinfiltrated leaves at any time point (Table S3). We thus compare agroinfiltrated (WT and P19) to mock-infiltrated leaves for further analysis.

Agroinfiltration induces leaf transcriptome changes associated with immune responses

Of all detected transcripts (n = 75 802), 24.6% (n = 18 648) significantly changed more than twofold in abundance at any time point and were thus considered differential in abundance. Among the differentials, the biggest category (n = 4849) is that of transcripts increasing in abundance for the first time at 2 dpi

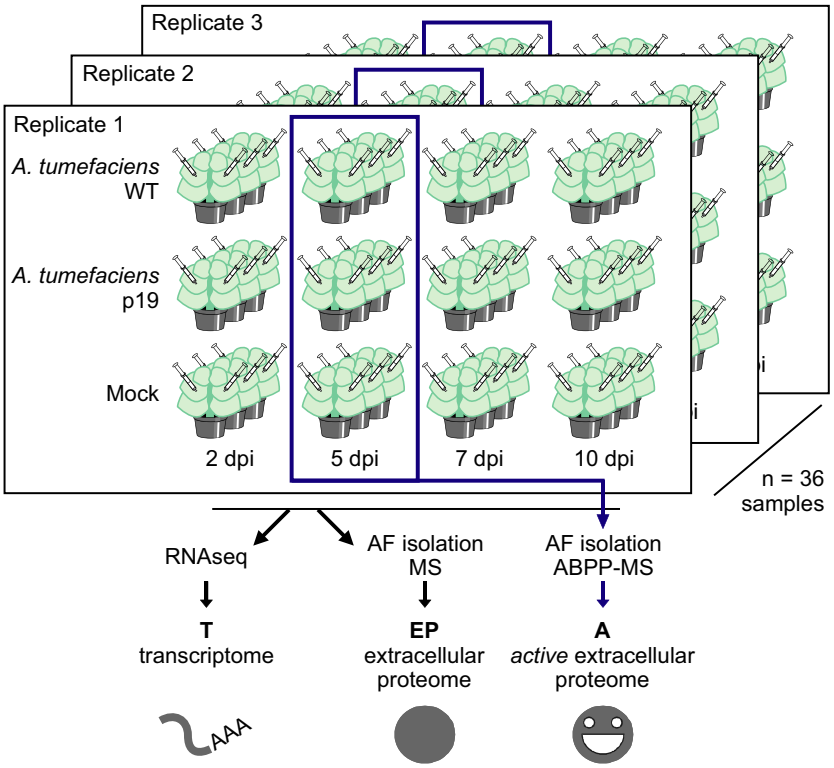


Figure 1 Experimental setup. Leaves of *Nicotiana benthamiana* were infiltrated with *Agrobacterium* GV3101-pMP90 without any T-DNA plasmid (WT) or carrying a plasmid for P19 expression (P19) or with buffer (mock). Abbreviations: dpi, days postinfiltration; RNAseq, mRNA sequencing; AF, apoplastic fluid; MS, protein mass spectrometry; ABPP, activity-based protein profiling.

(6.4% of the transcriptome) (Figure 2b, Table S4, Appendix S2). In this category, transcripts encoding proteins associated with immunity are overrepresented (representatives in Figure 2b, complete lists in Table S5). This includes transcripts encoding LRR (leucine-rich repeat) domain containing receptors such as the recently identified receptor for Agrobacterium cold-shock protein NbCSPR (Niben101Scf03240g00007) (Saur *et al.*, 2016), as well as signalling components carrying NB-ARC (nucleotide-binding adaptor shared by Apaf-1, resistance proteins and CED-4) and WRKY domains. Among the categories of transcripts whose abundance first increases at 5 or 7 dpi, transcripts encoding Myb transcription factors and serpins, LRRs and xylanase inhibitors are overrepresented. Besides transcripts encoding proteins associated with immune signalling and first-line defence, we detected a 4.5-fold average decrease in abundance of 13 transcripts encoding SWEET sugar efflux transporters, which may decrease the nutrient content of the extracellular space to control bacterial growth (Chen, 2014). Differential abundance of transcripts encoding both generators and quenchers of reactive oxygen species, as well as increased accumulation of transcripts encoding cytochrome P450 enzymes, shows that the plants are stressed upon agroinfiltration. Transcripts encoding members of the photosynthetic machinery and assimilatory metabolism in general are enriched among the transcripts decreasing in abundance from 2 dpi onwards, explaining the chlorotic phenotype of agroinfiltrated leaves (Pruss *et al.*, 2008). Among the transcripts detected constantly, transcripts encoding for housekeeping proteins like members of the ubiquitin-proteasome system and helicases are overrepresented. In summary, agroinfiltration is associated with an immune response mounted at the expense of photosynthesis.

Diversity and abundance of extracellular proteins increase upon agroinfiltration

We evaluated the effect of agroinfiltration on the extracellular proteome of *N. benthamiana* because the leaf extracellular space is the target site for glycoprotein accumulation in molecular farming, as well as the primary site of interaction with Agrobacterium and thus a promising site for improvement of the transient expression platform. Of all *N. benthamiana* proteins for which we identified extracellular peptides (n = 2233 protein groups as defined by MaxQuant (Tyanova *et al.*, 2016)), the vast majority (n = 1697, 75.9%) changed significantly and more than twofold in abundance and were thus considered differential in abundance. Among these differentials, most (n = 1572, 92.6%) increased in abundance upon agroinfiltration (Figure 2c, Table S6, Appendix S3). The increase in the extracellular proteome was mirrored by a corresponding increase in protein concentration in apoplastic fluid (AF) from agro- but not mock-infiltrated samples (Figure S2). Abundant intracellular housekeeping proteins such as actin, helicases and phosphofructokinases are overrepresented in the category of proteins that first increased in abundance at 10 dpi (n = 567), indicating that the interaction between *N. benthamiana* and Agrobacterium leads to cell content leakage at this late stage. Leakage may occur *in vivo* and during apoplastic fluid extraction.

Among proteins that first increased in abundance at 2 and 5 dpi, hydrolytic enzymes and inhibitors are overrepresented (Figure 2c and Table S7). This includes classical defence proteins such as xylanase inhibitors, chitinases (GH18) and pathogenesis-related protein 2 (PR2, a GH17 glucanase) (Cosgrove, 2016). PR2 accumulation upon agroinfiltration is consistent with an earlier study (Goulet *et al.*, 2010). Cell wall remodelling xyloglucan endotransglycolases/hydrolases (GH16) and versatile I06 α-amylase/Ser protease inhibitors may contribute indirectly to plant defence. Family C01 proteases (PLCPs) are overrepresented in the small category of proteins that first decreased in abundance at 10 dpi. Thus, *N. benthamiana* extracellular PLCPs do not increase as strongly and persistently in abundance upon agroinfiltration as tomato extracellular PLCPs do upon pathogen challenge (van Esse *et al.*, 2008). In contrast, PLCPs may localize to intracellular

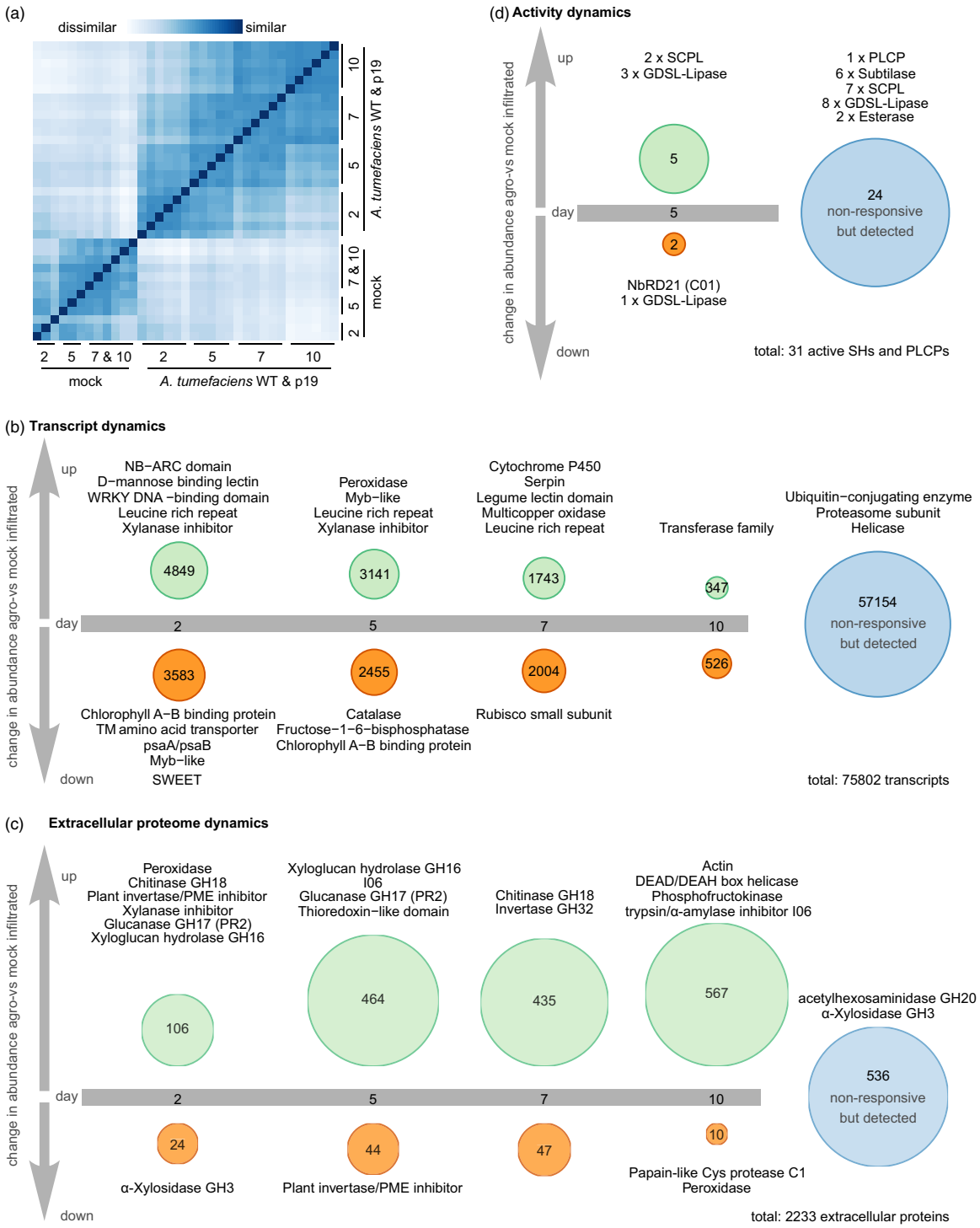


Figure 2 The immune response to agroinfiltration entails an increase in the extracellular proteome and is not affected by P19 overexpression. (a) Euclidean sample distances between the transcriptomes obtained from all 36 samples. Samples were ordered by hierarchical clustering based on the sample distances. (b, c): Transcripts (b) or proteins (c) were grouped by when their abundance first changed significantly (Wald test for transcripts, Student's *t*-test for proteins; Benjamini–Hochberg (BH) adjusted $P < 0.05$) and more than twofold in (WT and P19) agroinfiltrated samples compared to mock-infiltrated samples. Annotations given above the circles are representatives of the PFAM families that are significantly (Hypergeometric test, BH-adjusted $P < 0.05$) overrepresented in the respective regulatory category compared to all detected transcripts (b) or proteins (c). Protein groups for which corresponding peptides were identified are counted as one protein. (d) Activity of extracellular PLCPs and Ser hydrolases was assayed by ABPP-MS at 5 dpi, counting each protein group for which peptides were identified as one active protein. Proteins were grouped by whether they were enriched in the WT agroinfiltrated samples, controls or both (*t*-test probe sample vs no-probe control, BH-adjusted $P < 0.1$). Differences in abundance between agroinfiltrated samples and controls were not significant in any case. Only proteins annotated as SHs or PLCPs are included in the figure. Full data sets are given in Tables S4 and S5 (a&b), S6 and S7 (c) and S11 (d). The R code to generate the figures is given in Appendices S2 (a, b), S3 (c) and S5 (d).

compartments as shown for RD21 in Arabidopsis (Hayashi *et al.*, 2001) or may be degraded in the extracellular space. Invertase/pectin methyl esterase (PME) inhibitors are overrepresented both in the category of proteins first increasing at two and in the category of proteins first decreasing at 5 dpi in abundance. Invertase inhibition upon agroinfiltration thus appears to be transient, and indeed, invertases (GH32) are overrepresented in the category of proteins first increasing in abundance at 7 dpi. Plant invertases cleave the transport sugar sucrose, providing vital nutrients to sink tissues (Goetz *et al.*, 2001). The chlorotic agroinfiltrated leaves may be less photosynthetically active and lose nutrients to the bacteria, turning them from a source into a sink organ. Some members of the GH32 family degrade extracellular polysaccharides from pathogens (Limoli *et al.*, 2015), suggesting that GH32 family members may promote both nutrition and defence in agroinfiltrated leaves. GH3 α-xylosidases are overrepresented in both the category of proteins first decreasing in abundance at 2 dpi and in the category of proteins with constant abundance. Some GH3 family members act in cell wall remodelling and others locally adjust auxin concentrations as auxin-amido synthetases (Shigeyama *et al.*, 2016; Zheng *et al.*, 2016). This dual role may explain why GH3 members are overrepresented in both regulatory categories. Along with hydrolases and inhibitors, peroxidases and thioredoxins are overrepresented in several regulatory categories. These modulators of ROS levels facilitate both immune signalling and cell wall remodelling by extracellular ROS (Ivanchenko *et al.*, 2013).

Besides plant proteins, we also identified peptides from bacterial proteins in the extracellular proteome of agroinfiltrated leaves. In fact, Agrobacterium proteins make up a quarter of the extracellular proteins in agroinfiltrated samples (738 bacterial vs 2233 plant proteins) and appear to mostly function in providing nutrients to the bacteria. Highly abundant bacterial proteins are ABC transporters, cytochrome P450 proteins and porins. This may include cytoplasmic bacterial proteins released into the extracellular proteome upon cell death or during the extraction of apoplastic fluid. We identified peptides corresponding to 17 different Agrobacterium proteases, including six Ser proteases, in the extracellular space (Table S8), but did not identify peptides from bacterial proteases using ABPP-MS.

Ageing of leaves irrespective of their treatment during our 10-day time course is associated with induction of defence and decrease in primary metabolism. 18.9% of detected transcripts and 6.5% of identified extracellular proteins changed significantly in abundance over time independent of the treatments. Analysis of predicted functions overrepresented among the changing transcripts and proteins suggests that while PLCPs, P450-domain-containing proteins and PR proteins accumulate, components of the photosynthetic machinery, histones and cytoskeleton elements decrease in abundance (R code in Appendix S4, data in Tables S9 and S10).

The repertoire of active extracellular PLCPs and Ser hydrolases is modulated, but not drastically expanded upon agroinfiltration

We next investigated extracellular hydrolase activity to identify active candidate proteases for depletion, focusing on Ser and Cys proteases because plant immune responses often result in increased abundance and activity of these protease classes and Cys proteases can degrade biopharmaceuticals *in vitro* (Paireder *et al.*, 2016, 2017). We performed activity-based protein profiling (ABPP) with probes targeting papain-like Cys proteases (PLCPs) and serine hydrolases (SHs) (Greenbaum *et al.*, 2002; Kaschani *et al.*, 2009). Both probes consist of a specific inhibitor that covalently binds the active site of their respective targets, a linker and a biotin tag used for enrichment of active enzymes from extracellular proteomes prior to MS analysis. Both probes have been validated in plants using target detection, genetic target depletion and inhibition of probe binding with independent protease inhibitors, confirming that reactivity to the probe indicates the availability of the active site and thus enzyme activity (Kovács and van der Hoorn, 2016). We focused on 5 dpi when the response to agroinfiltration is fully developed. We identified peptides corresponding to two PLCPs and 29 SHs, 17 of which are proteases that were enriched from the extracellular proteome using ABPP-MS. (Figure 2d). The abundance of peptides corresponding to one Clade II and one Clade III SCPL was increased upon agroinfiltration, indicating increased activity and/or abundance. In contrast, peptides corresponding to an RD21-like PLCP were only identified in ABPP-MS samples from mock-infiltrated plants, indicating depletion of enzyme activity upon agroinfiltration. The abundance of extracellular peptides from this PLCP remained constant upon agroinfiltration at 5 dpi, suggesting a post-translational regulatory mechanism. We identified peptides corresponding to six subtilases (one SBT5 and five SBT1 subtilases, including the proteins clustering with tomato P69), seven SCPLs (two Clade IB, three Clade II, four cade III) and one aleurain-like PLCP with similar abundance in both agro- and mock-infiltrated samples, indicating that activity of these enzymes remains constant upon agroinfiltration. This is surprising, as in tomato, active extracellular subtilases and PLCPs drastically increase in abundance and diversity during immune responses (van Esse *et al.*, 2008; Sueldo *et al.*, 2014). Besides the proteases, we identified peptides corresponding to 14 additional SHs annotated as lipases and esterases. Peptides from three GDSL lipases (containing a GDSL sequence motif) increased in abundance upon agroinfiltration, while peptides from one GDSL-lipase decreased. Adjustment of extracellular GDSL-lipase activity may contribute to immune signalling, as lipases regulate salicylic acid as well as ethylene signalling in Arabidopsis (Falk *et al.*, 1999; Kim *et al.*, 2013b) and upon powdery mildew infection, lipase-encoding transcripts accumulate in grapevine (Szalontai *et al.*, 2012) (R code in Appendix S5, data in Table S11).

The extracellular proteome and active secretome is under post-transcriptional and post-translational control

Having transcriptome, extracellular proteome and active secretome data creates a unique opportunity to detect discrepancies in abundance changes between transcripts, total extracellular proteins and active extracellular proteins. To assess how much post-transcriptional regulation shapes the extracellular proteome, we compared the fold changes of extracellular protein abundance and transcript abundance at 5 dpi, when the response to agroinfiltration is fully developed (Figure 3a, Appendix S6, Table S12). The extracellular protein (EP) was increased more or decreased less in abundance than its corresponding transcript (T) for 215 (9.7%) of the 2226 extracellular proteins for which we detected the corresponding transcript (significant difference between the fold changes, EP > T). Among these proteins are two PR1 proteins, eight PR2 glucanases and two P69-like subtilases (PR7). This finding indicates that the immune response is accompanied by efficient extracellular protein delivery, as previously suggested based on

transcriptional up-regulation of the secretory pathway during immunity (Wang *et al.*, 2005). More efficient extracellular delivery may be accompanied by enhanced stability of the secreted proteins. In addition to classical PR proteins, two PLCPs (one XCP and one RD19-like) and three pepsin-like aspartic proteases appeared efficiently delivered to the extracellular space with EP > T, suggesting they may be candidate immune proteases. Only 45 extracellular proteins (2.0%) increased less or decreased more in abundance than expected from their transcript level changes (EP < T).

Interestingly, for 23 of the 31 active enzymes for which we also detected the extracellular protein (74.2%), abundance of the active protein (A) increased more or declined less than expected based on changes in total extracellular protein abundance (A > EP) (Figure 3b, Supporting Table S13). This suggests that hydrolase activity is frequently post-translationally controlled. Among the proteins with A > EP are four subtilases, three of which contain I09 domains and are thus likely activated by cleavage upon agroinfiltration. Activation by cleavage may explain why abundance of all active subtilases remained constant, while total protein abundance decreased in five cases. Seven SCPLs also remained constant or increased in active protein abundance although their total protein abundance decreased (Table S13). As SCPLs lack inhibitory domains, they may undergo post-translational activation by release from an inhibitor or autoactivation triggered by pH or redox-level changes. Taken together, efficient extracellular delivery influences the increase in the extracellular proteome upon agroinfiltration and many hydrolases for which we identified peptides by ABPP-MS appear to be activated post-translationally.

N. benthamiana deploys a large, diverse repertoire of proteases in agroinfiltrated leaves

To improve protease annotation, we analysed the *N. benthamiana* protease repertoire in the context of known plant proteases, using the MEROPS nomenclature. The MEROPS database of proteases and inhibitors defines families based on protein sequence homology that are grouped into clans based on structural homology. Protease family names consist of a letter denoting the catalytic class and a unique number (i.e. A01 for pepsin-like aspartic proteases) (Rawlings *et al.*, 2014). We identified 1245 proteases and noncatalytic protease homologs in the curated proteome of *N. benthamiana*. A smaller protease repertoire is encoded by genomes of three crop and model plants: Arabidopsis (796 proteases), tomato (901) and rice (997) (Figure 4a and Table S14). Although the *N. benthamiana* protease repertoire is much larger, the proportion of predicted proteins annotated as proteases is higher in the other plants (2.9% in Arabidopsis, 2.6% in tomato and 2.4% in rice) than in *N. benthamiana* (1.6%). The lower proportion of proteases in *N. benthamiana* may reflect the suboptimal genome annotation.

To characterize functional proteases and protease inhibitors in agroinfiltrated leaves, we analysed them at three levels. First, we detected transcripts for 975 proteases and 60 inhibitors. Second, we identified extracellular peptides from 196 proteases and 21 inhibitors, including proteases from every catalytic class. Third, we identified 17 active extracellular Ser and Cys proteases in agroinfiltrated leaves (Figure 4b). The most prominent features of the *N. benthamiana* protease repertoire are the large numbers of Cys, Metallo- and Thr proteases. Among the Cys proteases, the

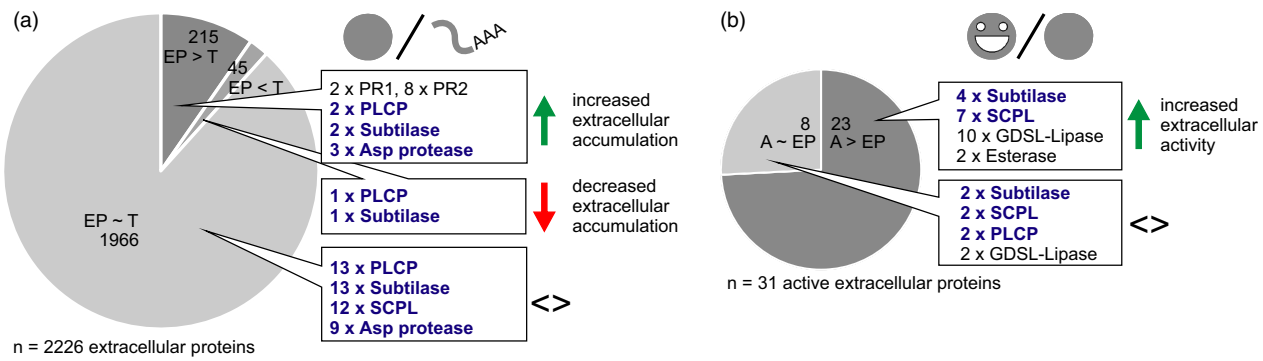
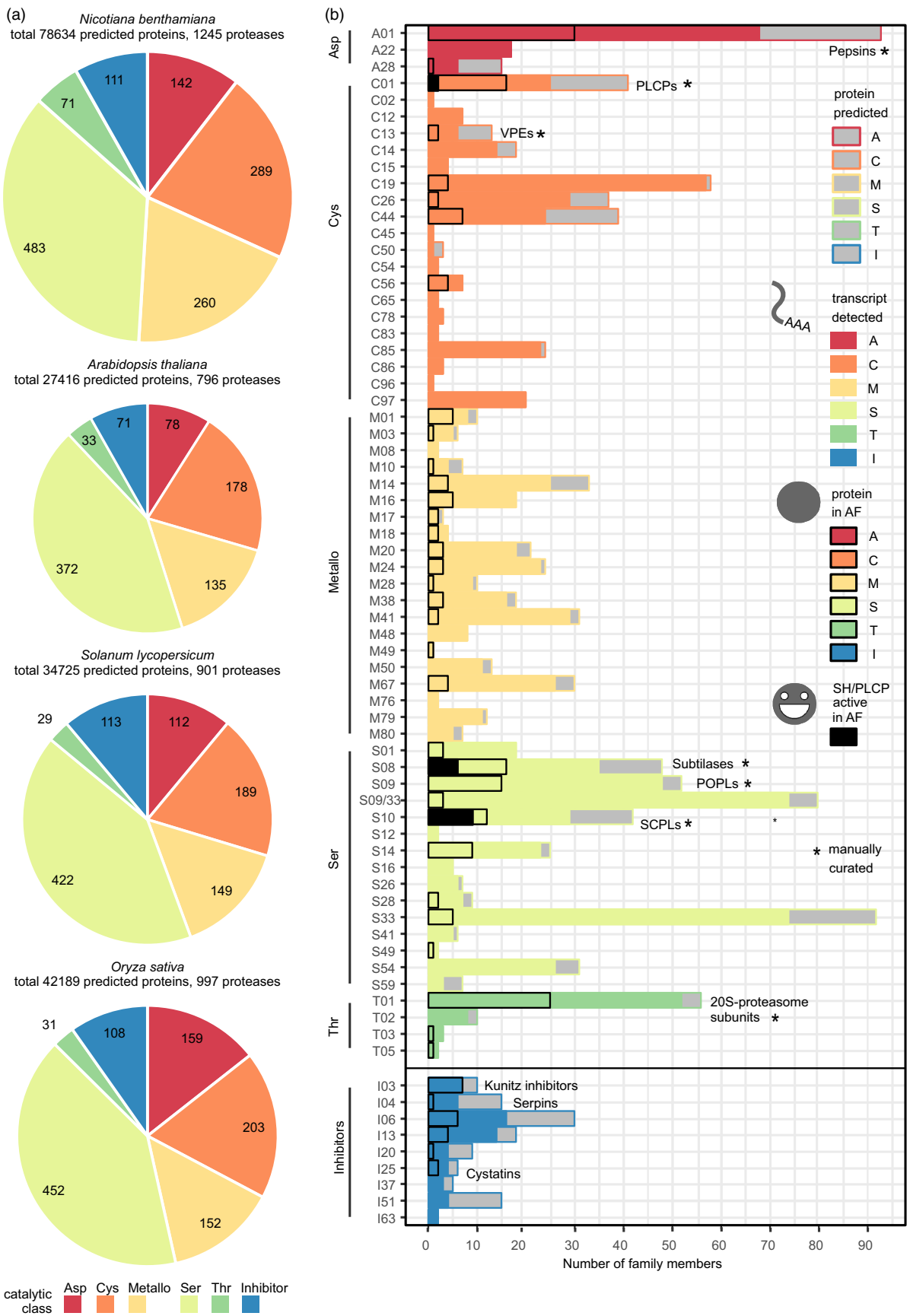


Figure 3 Post-transcriptional and post-translational control over the extracellular proteome. Fold changes in response to agroinfiltration at 5 dpi were compared between abundances of extracellular proteins and their transcripts (a) and between activity and abundance of extracellular proteins (b) (Student's *t*-test; BH-adjusted $P < 0.1$). Protein groups of interest are named next to the pie charts. Full data sets are given in Tables S12 and S13. The R code of the analysis is given in Appendix S6. T, fold change of transcript abundance; EP, fold change of extracellular protein abundance; AEP, fold change of activity of the extracellular protein.

Figure 4 *Nicotiana benthamiana* has a diverse protease and protease inhibitor repertoire. (a) Number of proteases and noncatalytic protease homologs in each catalytic class or inhibitors annotated are given for each species (*N. benthamiana* curated proteome, *Arabidopsis thaliana* TAIR10, *O. sativa* v7 JGI, *S. lycopersicum* ITAG2.4). The area of each pie chart is scaled by the total number of proteases and inhibitors. (b) The protease and protease inhibitor repertoire of *N. benthamiana*. For each MEROPS family, bars give the size in the predicted proteome (grey), the number of transcripts detected in mock and/or agroinfiltrated leaves (filled), the number of proteins for which we detect corresponding extracellular peptides in agro- and/or mock-infiltrated leaves (black outline) and the number of enzymes for which peptides were detected in ABPP-MS, indicating activity (black fill). For the manually curated families (marked by an asterisk), protease homologs lacking the active site were not counted. Each protein group identified in MS and ABPP-MS was counted as one family member. Note that due to the nature of the ABPP probes used, only SHs and PLCPs were monitored on the activity level. The S09 (prolyl oligopeptidase) and S33 (prolyl aminopeptidase) families share the α/β-hydrolase fold (PFAM families PF12695 and PF12697), and sequences with only these PFAM identifiers are marked S09/S33.



metacaspase family C14 is doubled in size (n = 18 members) compared to Arabidopsis (n = 9), tomato (n = 9) and rice (n = 8). We did not identify extracellular peptides corresponding to metacaspases, although 14 had detectable transcripts. The large number of metalloproteases in *N. benthamiana* (n = 260) compared to Arabidopsis (n = 135), tomato (n = 149) and rice (n = 152) is distributed among 20 families, and we identified extracellular peptides corresponding to members of most metalloprotease families. 10% (n = 26) of the metalloprotease-encoding genes increased in transcript abundance upon agroinfiltration, while 13% (n = 35) decreased. In contrast, 75% (n = 28) of the metalloproteases for which we identified extracellular peptides increased in abundance and only one M28 protease decreased. Very few plant metalloproteases are functionally characterized, including AtSOL1, an M14 carboxypeptidase processing peptide hormones (Casamitjana-Martínez *et al.*, 2003; Tamaki *et al.*, 2013) and AtPreP1 and 2, Arabidopsis M16 proteases cleaving organellar target peptides (Bhushan *et al.*, 2005). The *N. benthamiana* M10 protease NMMP1 has been implicated in defence because silencing *NMMP1* confers susceptibility to bacterial pathogens (Kang *et al.*, 2010). The transcript corresponding to NMMP1 (Niben101Scf10336XLOC_078719) increases in abundance upon agroinfiltration, while its extracellular peptides appear constant. The large metalloprotease repertoire of *N. benthamiana* is changing upon agroinfiltration, raising the question whether these metalloproteases might regulate the immune response through protein processing. The high number of Thr proteases (n = 71) compared to Arabidopsis (n = 34), tomato (n = 29) and rice (n = 31) is due to drastic expansion of the T01 family in *N. benthamiana* (n = 65, vs n = 24 in Arabidopsis, n = 20 in tomato and n = 23 in rice). T01 contains the α and β subunits of the 20S core protease of the proteasome. Phylogenetic analysis showed that *N. benthamiana* has more representatives of each subunit (Figure S3). We identified extracellular peptides corresponding to 25 T01 subunits, possibly due to cell content leakage. As transcripts of most (n = 52) T01 subunits were detected in leaves, multiple versions of the 20S proteasome may co-exist, as they do in Arabidopsis (Book *et al.*, 2010). Indeed, we recently showed that two sets of catalytic subunits are incorporated in functional 20S proteasomes in *N. benthamiana* (Misas-Villamil *et al.*, 2017).

In contrast to the protease repertoire, the protease inhibitor repertoire of *N. benthamiana* (n = 111 predicted protease inhibitors) is not much larger compared to tomato (n = 113), rice (n = 108) and Arabidopsis (n = 71). This apparent discrepancy may reflect the multifunctionality of many protease inhibitors (Grosse-Holz and van der Hoorn, 2016) and incomplete annotation. Among the annotated protease inhibitors, the I03 (Kunitz) inhibitors probably function extracellularly, as we identified extracellular peptides corresponding to all seven Kunitz (family I03) inhibitors for which we detected transcripts. Kunitz inhibitors can inhibit both subtilases (family S08) and α-amylases, but their bifunctional structure can also target other Ser or Cys proteases, and other proteins (Renko *et al.*, 2012). *N. benthamiana* serpins (I04) appear to mostly be intracellular, as we detected six serpin-encoding transcripts, but identified corresponding extracellular peptides for only one. Serpins have been found in both the cytoplasm (Lampl *et al.*, 2013) and the extracellular space (Ghorbani *et al.*, 2016), and regulate plant defence and programmed cell death through irreversible inhibition of Ser and Cys proteases (Bhattacharjee *et al.*, 2017; Lampl *et al.*, 2013). We detected transcripts for four and identified extracellular peptides corresponding to two cystatins (family I25). Cystatins target PLCPs and VPEs, regulating storage protein accumulation, germination and defence (Benchabane *et al.*, 2008; Grosse-Holz and van der Hoorn, 2016).

Having obtained an overview of the *N. benthamiana* protease and protease inhibitor repertoire, we focused on six large protease families, which we curated manually (Appendix S1). For these six families, we performed phylogenetic analyses to resolve subfamilies and determine which *N. benthamiana* proteins are most similar to previously studied proteases (Figures 5 and 6).

The PLCP family is conserved in N. benthamiana, but PIP1- and RCR3-like PLCPs are absent from the extracellular proteome of agroinfiltrated leaves

Nicotiana benthamiana has more papain-like Cys proteases (PLCPs, family C01, n = 41 members) than Arabidopsis (n = 36) and tomato (n = 36), but less than rice (n = 54). PLCP subfamilies can be defined by shared sequence features (Richau *et al.*, 2012) (Figure 5a). For example, the NPIR vacuolar localization signal is found in aleurain-like proteases (ALPs) and the KDEL ER-retention signal in Cys endopeptidases (CEPs). Cathepsin-B-like proteases (CTBs) have a specific prodomain (PF08127) serving as chaperone and inhibitor, like the I29 (PF08246) prodomain for other PLCPs. Most *N. benthamiana* PLCPs have a secretion signal predicted by SignalP (Dyrløv Bendtsen *et al.*, 2004). Accordingly, we identified extracellular peptides corresponding to 18 of the 25 PLCPs for which we detected transcripts. Among the extracellular PLCPs are three granulin-carrying proteases similar to the immune protease AtRD21 (Shindo *et al.*, 2012), three Cathepsin B-like proteases (CTBs) and NbCYP1 and NbCYP2, which limit susceptibility to fungal pathogens (Hao *et al.*, 2006). We identified peptides in ABPP-MS from NbRD21 and NbCYP1, indicating that these proteases are active extracellularly. Many of the PLCPs for which we detected transcripts and identified extracellular peptides contribute to *N. benthamiana* immunity. For instance, silencing *NbCathB* (Gilroy *et al.*, 2007; McLellan *et al.*, 2009) blocks the hypersensitive response (HR) and *NbC14/CP14* silencing confers susceptibility to *Phytophthora infestans* (Bozkurt *et al.*, 2011; Kaschani *et al.*, 2010). Surprisingly, we did not identify extracellular peptides corresponding to the *N. benthamiana* proteins clustering with the tomato immune proteases PIP1 (Tian *et al.*, 2004) and RCR3 (Krüger *et al.*, 2002), although we detected NbPIP1- and NbRCR3-encoding transcripts.

P69-like SBT1 subtilases are abundant and active in the extracellular proteome of agroinfiltrated N. benthamiana leaves

The *N. benthamiana* subtilase family (S08, n = 56 members) is the same size as in Arabidopsis (n = 56) and smaller than in tomato (n = 90) and rice (n = 61) (Figure 5b). We identified extracellular peptides for 28 of the 39 subtilases whose transcript we detected; 12 of the 28 subtilases for which we identified extracellular peptides were among the top 10% most abundant extracellular proteins and we identified peptides corresponding to 11 active subtilases using ABPP-MS (Figure 5b). Across the whole subtilase family, the I09 prodomain is well conserved and the PA dimerization domain (Rose *et al.*, 2010) is present in some members of each subfamily. An exception lacking SP, I09 and PA domains are the basal SBT6 subtilases (Taylor and Qiu, 2017). We detected transcripts encoding three *N. benthamiana* SBT6 subtilases and identified extracellular peptides from one. SBT6

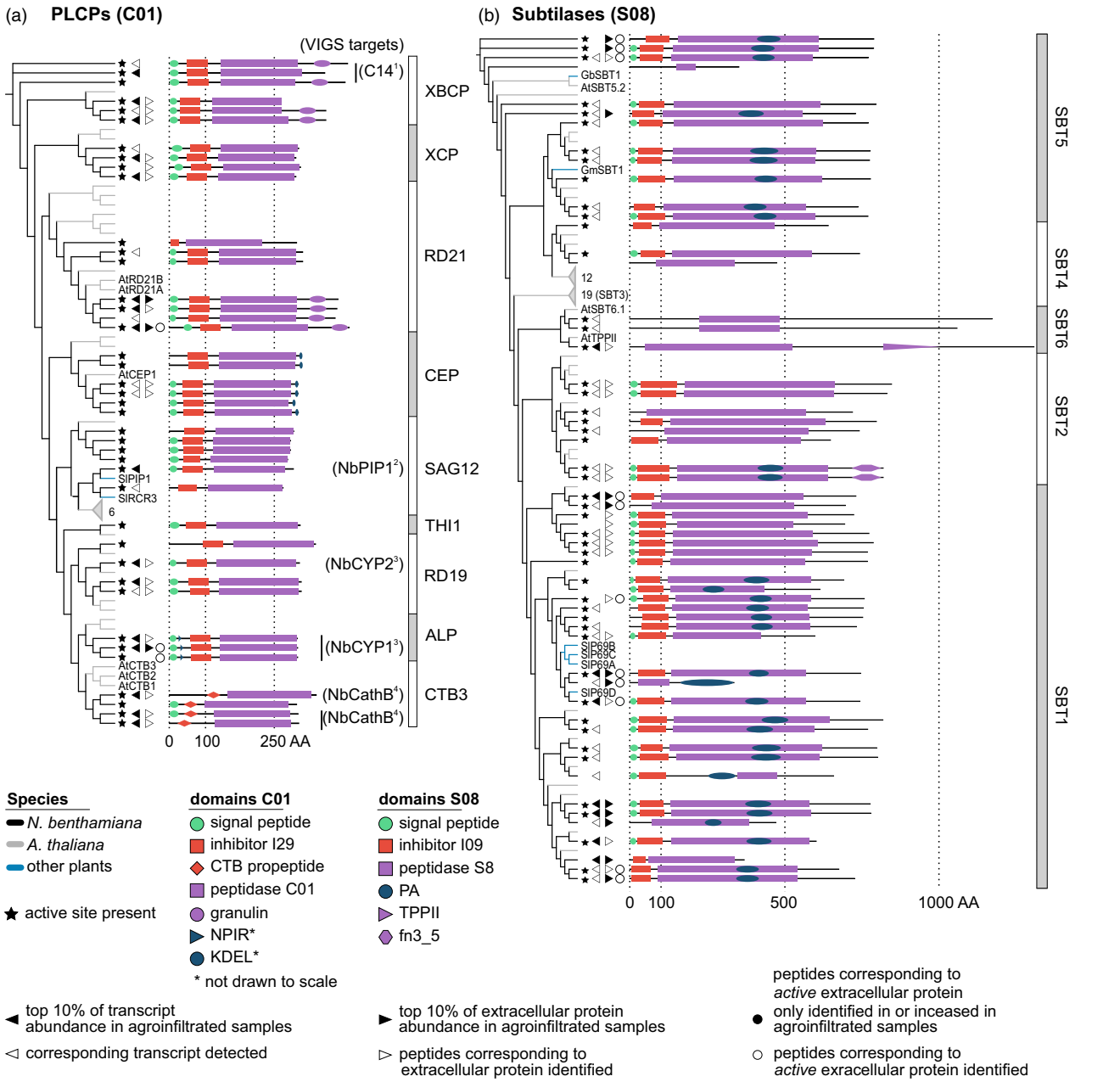


Figure 5 Annotation and detection of extracellular papain-like Cys proteases (PLCPs) and subtilases in *Nicotiana benthamiana*. Phylogenetic trees based on the protein sequences of PLCPs (a) and subtilases (b) containing all proteases and protease homologs in the respective family in Arabidopsis (grey branches) and *N. benthamiana* (black branches), supplemented by well-studied enzymes from other plant species (blue branches). Names are given as two-letter species abbreviation followed by the name used in the literature. Grey triangles denote collapsed subtrees that contain only Arabidopsis sequences, with the number of proteins given next to the triangle. For protein abundance and activity, the respective symbols are shown next to all members of each protein group for which corresponding peptides were identified. VIGS targets were predicted based on >90% identical residues between the fragment used for VIGS and the respective transcript. References: 1 (Kaschani *et al.*, 2010); 2 (Xu *et al.*, 2012); 3 (Hao *et al.*, 2006); 4 (Gilroy *et al.*, 2007). Abbreviations: CTB, cathepsin-B-like; TPP, tripeptidyl-peptidase; fn3_5, fibronectin-3 like domain found on streptococcal C5a peptidase.

subtilases can process peptide hormones regulating cell elongation (Ghorbani *et al.*, 2016), or degrade peptides released by the 26S proteasome (Book *et al.*, 2005). Remarkably, the SBT1 subfamily is threefold larger in *N. benthamiana* (n = 30 members) than in Arabidopsis (n = 9), while the SBT3 subfamily is absent in *N. benthamiana*. We detected 22 SBT1 subtilase-encoding transcripts and identified corresponding extracellular peptides for 19. We also identified peptides corresponding to eight active SBT1 subtilases by ABPP-MS. The tomato P69A, B and C subtilases (Jordá *et al.*, 1999) cluster with the SBT1 subfamily, which is consistent with a recently published, updated phylogeny of the subtilase family (Taylor and Qiu, 2017). We detected transcripts for eight and identified extracellular peptides corresponding to four SBT5 subtilases. We also identified peptides corresponding to three active SBT5 subtilases by ABPP-MS. SBT5 subtilases can regulate plant immunity as receptors (Duan *et al.*,

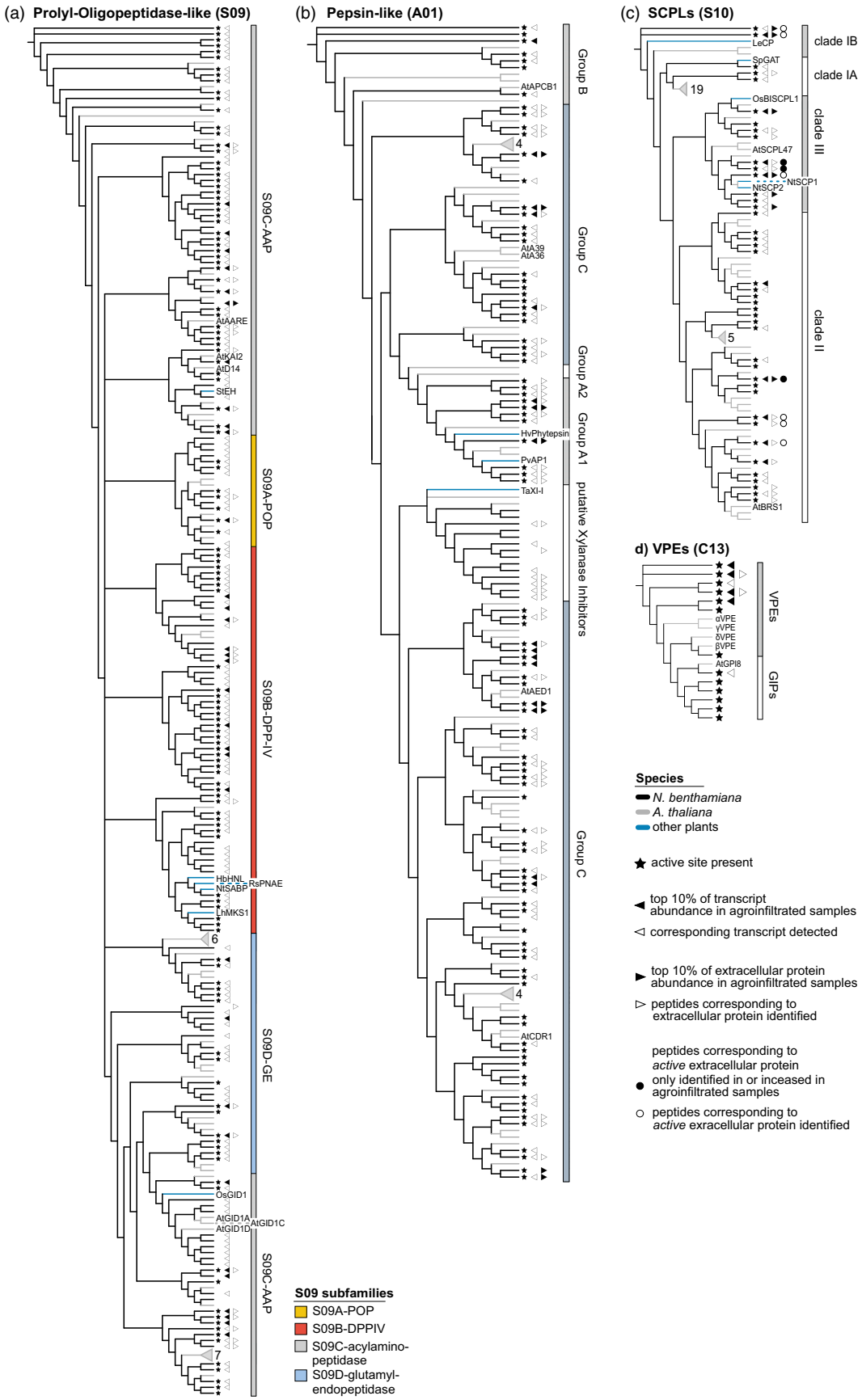


Figure 6 Annotation and detection of additional protease families in *Nicotiana benthamiana*. Phylogenetic trees based on the protein sequences of POPLs (a), pepsin-like proteases (b), SCPLs (c) and VPEs (d) containing all proteases and protease homologs of the respective family in Arabidopsis (grey branches) and *N. benthamiana* (black branches), supplemented by well-studied enzymes from other plant species (blue branches). Arabidopsis sequences that only carry the α/β-hydrolase fold PFAM identifiers are not shown in the S09 tree for readability. Names and symbols are used as described for Figure 5.

2016), as transcription factor binding proteins (Serrano *et al.*, 2016) or being processed to release peptide hormones (Pearce *et al.*, 2010). The updated subtilase phylogeny (Taylor and Qiu, 2017) suggests to split SBT6 into two subfamilies and notes that the distinction between SBT4 and SBT5 subfamilies is weakly supported. Indeed, we find two subclades of SBT6 in *N. benthamiana*, clustering with one of the Arabidopsis representatives each. SBT4 falls into a Clade containing part of SBT5 in our tree, indicating the updated phylogeny agrees with our curated *N. benthamiana* proteome.

Nicotiana benthamiana POPLs are underrepresented in the extracellular proteome of agroinfiltrated leaves

The prolyl oligopeptidase-like (S09, POPL) family in *N. benthamiana* (n = 180 members) is equivalent in size with the POPL families in Arabidopsis (n = 143), tomato (n = 162) and rice (n = 196) (Figure 6a). The S09 family is defined via the α/β-hydrolase fold (PF12695 and PF12697), but these proteins are not always proteases (Mindrebo *et al.*, 2016). Interestingly, we only identified extracellular peptides for 18 of the 122 family members for which we detect a transcript (15%), suggesting that *N. benthamiana* POPLs are primarily intracellular. S09A (prolyl oligopeptidase-like, POPL) is the smallest subfamily and no plant POPL has been functionally characterized, but we detected 14 POPL-encoding transcripts and identified extracellular peptides from two POPLs. The S09B dipeptidyl-peptidase type IV (DPP-IV) subfamily contains membrane-bound exopeptidases (Tripathi and Sowdhamini, 2006), but also clusters with nonproteolytic α/β-hydrolases, including HbHNL, RsPNAE and LhMKS1 and the salicylic acid receptor NtSABP2 (Auldridge *et al.*, 2012; Dogru *et al.*, 2000; Forouhar *et al.*, 2005; Wagner *et al.*, 1996). Interestingly, we only detected a transcript for NbSABP2. We identified extracellular peptides from four DPP-IVs. We also detected transcripts for four and identified peptides for two aminoacyl-removing peptidases (AAPs/AAREs, subfamily S09C) clustering with AtAARE. AtAARE is implicated in the cytoplasmic antioxidative system (Nakai *et al.*, 2012; Yamauchi *et al.*, 2003). We detected transcripts, but did not identify extracellular peptides for the nonproteolytic members of S09C, including the proteins clustering with the hormone receptors D14 (Yao *et al.*, 2016), KAI2 (Guo *et al.*, 2013), GID1 (Griffiths *et al.*, 2006) and the potato epoxide hydrolase StEH (Stapleton *et al.*, 1994).

Pepsin-like aspartic proteases are highly abundant in agroinfiltrated leaves and pepsin-like xylanase inhibitors have expanded in N. benthamiana

Extracellular peptides from pepsin-like aspartic proteases (A01) were abundantly detected and the A01 family is expanded in *N. benthamiana* (n = 110 members) compared to Arabidopsis (n = 69) and tomato (n = 100), but is smaller than the rice A01 family (n = 130) (Figure 6b). We detected 76 A01 protease-encoding transcripts and identified extracellular peptides from 45 pepsin-like aspartic proteases. Eight pepsin-like aspartic proteases were among the top 10% most abundant extracellular proteins. Pepsin-like aspartic proteases are subdivided into subfamilies A1 (typical pepsin-like), A2 (typical, but lacking the plant-specific insert), B (nucellins) and C (atypical) (Faro and Gal, 2005). We detected transcripts and identified peptides for 10 A1 pepsin-like proteases and they cluster with two enzymes implicated in stress responses, barley phythepsin and bean AP1 (Contour-Ansel *et al.*, 2010; Hückelhoven *et al.*, 2001). Group A2 seems absent in the *N. benthamiana* predicted proteome. We detected transcripts, but did not identify extracellular peptides, for five A01 group B proteases. AtAPCB1 in group B is required for autophagy and resistance to *Botrytis* (Li *et al.*, 2016). Group C is the largest A01 subfamily in *N. benthamiana*, with 54 A01 Group C-encoding transcripts detected in leaves and corresponding extracellular peptides identified for 29. Two Arabidopsis members of group C, AtCDR1 and AtAED1, modulate plant defence responses in the extracellular space (Breitenbach *et al.*, 2014; Xia *et al.*, 2004). AtAED1 clusters with two abundant extracellular *N. benthamiana* proteins. We detected transcripts encoding several and identified extracellular peptides from one protein clustering with AtA36 and AtA39, two putative GPI-anchored aspartic proteases (Gao *et al.*, 2017). Interestingly, the A01 Clade clustering with the wheat xylanase inhibitor TaXI-I is expanded drastically with 14 members in *N. benthamiana*, compared to two in Arabidopsis (Brutus *et al.*, 2005; Sansen *et al.*, 2004). TaXIs share the fold of pepsin-like proteases, but have lost the active site and act as xylanase inhibitors. We detected transcripts for seven putative *N. benthamiana* xylanase inhibitors and identified extracellular peptides corresponding to six.

Nicotiana benthamiana has an expanded SCPL Clade III in the extracellular proteome of agroinfiltrated leaves

Nicotiana benthamiana has less serine carboxypeptidase-like enzymes (SCPLs, S10, n = 42 members) than Arabidopsis (n = 54), tomato (n = 61) and rice (n = 59) (Figure 6c). We detected transcripts for 29 and identified extracellular peptides from 19 SCPLs. We also identified peptides corresponding to nine active extracellular SCPLs by ABPP-MS. SCPLs fall into four clades (Fraser *et al.*, 2005). We detected transcripts for three and identified extracellular peptides corresponding to one member of Clade IA, which contains the *Solanum pennellii* glucose acetyltransferase (SpGAT) (Franziska, 2013). Clade IB contains the wound-inducible tomato carboxypeptidase LeCP (Moura *et al.*, 2001) and two *N. benthamiana* proteins, for which we detected transcripts, identified extracellular peptides and peptides by ABPP-MS, indicating activity. The largest S10 subfamily is Clade II, with transcripts detected for 16 members and extracellular peptides identified for eight. We also identified extracellular peptides corresponding to four active Clade II SCPLs in ABPP-MS. Interestingly, Clade III is expanded in *N. benthamiana* (n = 10) compared to Arabidopsis (n = 5) and well represented in the extracellular proteome, with the encoding transcripts detected and extracellular peptides identified for eight members each. We also identified peptides corresponding to three active Clade III SCPLs in ABPP-MS. Clade III members such as NtSCP1, NtSCP2 (Bienert *et al.*, 2012) and AtSCPL47 (Charmont *et al.*, 2005) are extracellular carboxypeptidases, and OsBISCPL1, a rice Clade III SCPL, enhances stress resistance when overexpressed in Arabidopsis (Liu *et al.*, 2008).

Peptides corresponding to extracellular VPEs are identified in agroinfiltrated N. benthamiana leaves

We annotated seven vacuolar processing enzymes (VPEs/legumains/asparaginyl endopeptidases) and six GPI-anchor transamidases that share the VPE domain architecture (PF01650) in *N. benthamiana*. Together, they constitute family C13 (n = 13 members), which is larger than in Arabidopsis (n = 5) and rice (n = 6), but smaller than in tomato (n = 19). We detected five VPE-encoding transcripts and one NbGIP-encoding transcript. Notably, we also identified extracellular peptides corresponding to two VPEs, consistent with observations made in tomato (Sueldo *et al.*, 2014). VPEs can activate proteins in vacuoles, including proteases (Rojo *et al.*, 2003) and protease inhibitors (Heath *et al.*, 1995; Mylne *et al.*, 2011). Silencing of *NbVPEs* blocks virus-induced cell death in *N. benthamiana* (Hatsugai *et al.*, 2004) and VPEs can also act in other forms of plant cell death (Gepstein *et al.*, 2003; Hatsugai *et al.*, 2015; Nakaune *et al.*, 2005; Sueldo *et al.*, 2014).

Conclusions

Upon agroinfiltration, 25% of the full leaf mRNA transcriptome changes in abundance, associated with an immune response mounted at the expense of photosynthesis. 70% of all extracellular proteins increase in abundance and their predicted functions confirm that an extracellular immune response occurs. Increasing the extracellular proteome while photosynthesis is shut down appears to drive leaves into a nutrient-deprived state. Engineering *N. benthamiana* to react less strongly to Agrobacterium, or Agrobacterium to be less immunogenic in *N. benthamiana*, may enhance RP expression by re-directing limiting resources. Interestingly, the expression of the silencing inhibitor P19 had minor effects on the transcriptome and no effect on the extracellular proteome.

Discrepancies between changes in transcript, extracellular protein and active extracellular protein abundances suggest that the extracellular proteome is influenced post-transcriptionally and that many extracellular enzymes are activated post-translationally. The *N. benthamiana* immune response to agroinfiltration differs from immune responses to bacterial and fungal pathogens in Arabidopsis and tomato in that there is no drastic increase in numbers or amounts of active extracellular subtilases and PLCPs (van Esse *et al.*, 2008; Gilroy *et al.*, 2007; Sueldo *et al.*, 2014; Xia *et al.*, 2004). This is surprising, as *N. benthamiana* has an exceptionally large repertoire of 1245 proteases and noncatalytic protease homologs, transcripts corresponding to 975 proteases were detected in leaves and peptides corresponding to 196 proteases were identified in the extracellular space. Prominent features of the extracellular protease repertoire of agroinfiltrated leaves are an expanded clade of SCPLs, highly abundant pepsin-like proteases and many SBT1 subtilases. Targeted depletion or inhibition of these enzymes may limit undesired proteolysis to improve agroinfiltrated *N. benthamiana* as a protein expression platform. We have selected several proteases for genetic depletion by genome editing to investigate their role in RP degradation and how they shape the endogenous proteome.

Experimental procedures

All chemicals were obtained from Sigma (Sigma-Aldrich, St. Louis, MO) unless specified otherwise.

Agroinfiltration procedure

Nicotiana benthamiana plants were grown at 21 °C under a 16/8-h light/dark regime in a growth room. *Agrobacterium* GV3101-pMP90 (WT) and *Agrobacterium* GV3101-pMP90 carrying a plasmid encoding silencing inhibitor P19 of tomato bushy stunt virus, driven by a 35S promoter (pJK050, a gift from Jiorgos Kourelis), were grown for 21 h at 28C with agitation in LB medium (10 g/L NaCl, 10 g/L Tryptone, 5 g/L yeast extract) containing 100 μM rifampicin and 100 μM gentamycin (for WT) plus 100 μM kanamycin (for P19). Bacteria were collected by centrifugation at 2000 g for 5 min at room temperature (RT), resuspended in infiltration buffer (10 mM 2-(N-morpholino) ethanesulfone (MES), 10 mM $MgCl_2$, pH 5.7, 100 μM acetosyringone) to OD_{600} = 0.5 and left for 2 h at 28 °C with agitation to recover. The first and second fully expanded leaves of preflowering stage *N. benthamiana* (4–5 weeks old) were infiltrated with the bacteria suspension using a syringe without a needle.

mRNA extraction and sequencing

For each sample, two leaf discs per leaf from six leaves (three different plants) were pulverized under liquid nitrogen using a mortar and pestle. RNA was extracted from 50 mg of leaf powder using TRIZOL (Thermo Fisher Inc, Waltham, MA) according to the manufacturer's instructions. DNA contamination was removed by in-solution digest with the Qiagen RNAse-free DNAse kit, followed by cleanup with the Qiagen RNeasy kit, following the manufacturer's instructions (Qiagen, Hilden, Germany). RNA quality was assessed using a Bioanalyzer with the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA) and following the manufacturer's instructions. All samples used for sequencing had a RIN (RNA integrity number, 28S to 18S rRNA ratio) >6.5. RNAseq library preparation and sequencing were performed by the Wellcome Trust Centre for Human Genetics, Oxford. mRNA was enriched using oligo-dT beads and sequenced over three lanes of an Illumina HiSeq device, generating on average 184 million 100-bp paired-end reads per lane.

Bioinformatics tools used for transcriptome analysis

To obtain the genome-based transcriptome (DB4), RNAseq reads were filtered to only retain those with a Phred Q Score >30 (Ewing and Green, 1998) and aligned to the Niben101 genome (Bombarely *et al.*, 2012) using TopHat version 2.0.14 (Kim *et al.*, 2013a) with default settings. The transcriptome was assembled using StringTie (Pertea *et al.*, 2015) on these alignments, allowing for multimapping of reads to several transcripts. TopHat and StringTie were run via the galaxy server (Afgan *et al.*, 2016). This resulted in the genome-based transcriptome (DB4). To obtain the *de novo* assembled transcriptome (DB3), raw reads were quality-trimmed using TRIMMOMATIC-0.32 (Bolger *et al.*, 2014), BAYESHAMMER (SPADES-3.5.0) (Nikolenko *et al.*, 2013) and ALLPATHS-LG-4832 (Butler *et al.*, 2008). Ribosomal RNA was removed using SORTMERNA-1.9 (Kopylova *et al.*, 2012). The quality-trimmed reads were then normalized with a khmer size of 21 in KHMER-0.7.1 (Crusoe *et al.*, 2015). Normalized reads were then assembled and scaffolded using SGA (Simpson and Durbin, 2012), SSPACE-v.3 (Boetzer *et al.*, 2011) and CAP3 (Huang and Madan, 1999). Assembled scaffolds then underwent a final correction step using PILON-1.6 (Walker *et al.*, 2014). This resulted in the *de novo* assembled transcriptome (DB3). We manually curated protease sequences in DB4, using single

transcripts from DB1-3 and 5, as described in Appendix S1. The curated transcriptome was fed to Salmon version 0.7 (Patro *et al.*, 2017) together with the filtered reads, and transcript quantification was performed in lightweight alignment mode. Thus, multimapping of reads was allowed during assembly of the transcriptome in DB4, but not during quantification. The NumReads output of Salmon was used for relative expression analysis in DESeq2 (Love *et al.*, 2014).

Bioinformatics tools for proteome prediction

All four transcriptome databases (DB1-4, see Appendix S1) were subjected to coding sequence prediction using GeneMark-ST (Tang *et al.*, 2015), TransDecoder (http://transdecoder.github.io) and Prodigal (Hyatt *et al.*, 2010) using default settings for eukaryotic gene sequences. In cases where all three methods predicted an open reading frame for a transcript the priority was given to the prediction made by GeneMark-ST unless the GeneMark-ST gene model was a substring of a longer TransDecoder gene model. Transcripts without predictions by any method were subjected to an additional round of gene prediction using Prodigal settings for bacterial genes and gene predictions were compiled to create the final predicted proteome.

Apoplastic fluid (AF) extraction

Six *N. benthamiana* leaves per sample were detached and vacuum-infiltrated with ice-cold water, dried on the surface and placed in a syringe without needle and plunger that was inserted in a 50-mL falcon tube. AF was collected by centrifugation at 2000 g, 4 °C for 25 min and stored at -80 °C until further use. Protein concentrations were determined with a Bradford assay according to Ernst and Zor (2010). To prove that leakage of cytosolic proteins into the extracellular proteome at later time points upon agroinfiltration is indeed caused by disease and not by our AF extraction method, we measured the activity of the intracellular enzyme malate dehydrogenase (MDH). MDH activity in our AF from mock-infiltrated leaves falls within the range reported for AF that is virtually free from cytosolic contamination (Figure S4) (Goulet *et al.*, 2010; Husted and Schjoerring, 1995).

Mass spectrometry and ABPP-MS

see supplemental methods, additional Appendix S7.

Bioinformatics tools for extracellular proteome analysis

Peptide spectra were annotated using Andromeda (Cox *et al.*, 2011). Included modifications were carbamidomethylation (static) and oxidation, N-terminal acetylation and carbamylation of Lysines and N-termini (dynamic). Protein quantification was performed using MaxQuant version 1.5.5.30 (Tyanova *et al.*, 2016), including all modifications.

Phylogenetic analyses

Sequences were aligned in Geneious (Kearse *et al.*, 2012) using a plug-in for MAFFT v7.017 (Katoh and Standley, 2013). Neighbour-joining trees were constructed using the geneious tree builder with Jukes-Cantor genetic distances and bootstrapped using 1000 times resampling. Trees were edited using iTOL (Letunic and Bork, 2016). Complete versions of the trees including all sequence names are given in Appendix S8.

Databases and protease annotation

Protease and inhibitor sequences and PFAM annotations were retrieved for Arabidopsis from TAIR10 (Berardini *et al.*, 2015) and for rice and tomato from Phytozome (Goodstein *et al.*, 2012). Protease sequences from other species to extend the family trees were retrieved from GenBank (NCBI Resource Coordinators, 2017) or UniProt (The UniProt Consortium, 2017). All Arabidopsis, rice, tomato and *N. benthamiana* proteases were annotated by mapping PFAM domains to MEROPS family annotations according to Table S15.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno *et al.*, 2016) partner repository (https://www.ebi.ac.uk/pride/archive/) with the data set identifier PXD006708. RNAseq data have been deposited in the NCBI Sequence Read Archive repository under identifier. [SRP109347]

Acknowledgements

We thank Marcel Bach, Daniela Sueldo and Philippe V. Jutras for constructive suggestions and critical reading, Jiorgos Kourelis and Jan M. Brauner for scientific discussion and Urszula Pyzio for excellent technical support. Jiorgos Kourelis also provided the P19 plasmid. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the Sequencing data. This work was financially supported by the ERC Project 'GreenProteases', University of Oxford (R.H., grant No. 616449), by Somerville College, Oxford (F.G.H. and R.H.), an ERC starting grant (M.K., grant No. 258413) and the Deutsche Forschungsgemeinschaft (M.K., grant no. INST 20876/127-1 FUGG). S.K. is a Royal Society University Research Fellow and work in S.K.'s laboratory received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 637765.

The authors declare no competing interests.

Author contributions

R.H. conceived the research. F.G.H. designed and performed experiments and analysed data unless specified otherwise. F.G.H. and R.H interpreted the results. S.K. provided guidance for RNAseq experimental design, transcriptome assembly and differential expression analyses and performed the *de novo* transcriptome assembly and coding sequence prediction. S.B. performed the sample preparation for MS. F.K. and M.K. performed LC-MS/MS and peptide and protein identification steps for MS of extracellular proteomes and ABPP-MS. F.G.H wrote the manuscript with feedback from R.H. All authors read and approved the final manuscript.

References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J. *et al.* (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10.

Auldridge, M.E., Guo, Y., Austin, M.B., Ramsey, J., Fridman, E., Pichersky, E. and Noel, J.P. (2012) Emergent decarboxylase activity and attenuation of α/β-hydrolase activity during the evolution of methylketone biosynthesis in tomato. *Plant Cell*, **24**, 1596–1607.

Benchabane, M., Goulet, C., Rivard, D., Faye, L., Gomord, V. and Michaud, D. (2008) Preventing unintended proteolysis in plant protein biofactories. *Plant Biotechnol. J.* **6**, 633–648.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*, **53**, 474–485.

Bevan, M. (1984) Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Res.* **12**, 8711–8721.

Bhattacharjee, L., Singh, D., Gautam, J.K. and Nandi, A.K. (2017) *Arabidopsis thaliana* serpins AtSRP4 and AtSRP5 negatively regulate stress-induced cell death and effector-triggered immunity induced by bacterial effector AvrRpt2. *Physiol. Plant.*, **32**, 9–339.

Bhushan, S., Ståhl, A., Nilsson, S., Lefebvre, B., Seki, M., Roth, C., McWilliam, D. *et al.* (2005) Catalysis, subcellular localization, expression and evolution of the targeting peptides degrading protease, AtPreP2. *Plant Cell Physiol.* **46**, 985–996.

Bienert, M.D., Delannoy, M., Navarre, C. and Boutry, M. (2012) NtSCP1 from tobacco is an extracellular serine carboxypeptidase III that has an impact on cell elongation. *Plant Physiol.* **158**, 1220–1229.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. and Pirovano, W. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, **27**, 578–579.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

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. *Mol. Plant Microbe Interact.* **25**, 1523–1530.

Book, A.J., Yang, P., Scalf, M., Smith, L.M. and Vierstra, R.D. (2005) Tripeptidyl peptidase II. An oligomeric protease complex from Arabidopsis. *Plant Physiol.* **138**, 1046–1057.

Book, A.J., Gladman, N.P., Lee, S.-S., Scalf, M., Smith, L.M. and Vierstra, R.D. (2010) Affinity purification of the Arabidopsis 26 S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* **285**, 25554–25569.

Bos, J.I.B., Kanneganti, T.-D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R. *et al.* (2006) The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant J.* **48**, 165–176.

Bozkurt, T.O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L.M. *et al.* (2011) Phytophthora infestans effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proc. Natl Acad. Sci. USA*, **108**, 20832–20837.

Breitenbach, H.H., Wenig, M., Wittek, F., Jorda, L., Maldonado-Alconada, A.M., Sarioglu, H., Colby, T. *et al.* (2014) Contrasting roles of apoplastic aspartyl protease AED1 and legume lectin-like protein LLP1 in Arabidopsis systemic acquired resistance. *Plant Physiol.* **175**, 114.239665.

Brutus, A., Reca, I.B., Herga, S., Mattei, B., Puigserver, A., Chaix, J.-C., Juge, N. *et al.* (2005) A family 11 xylanase from the pathogen Botrytis cinerea is inhibited by plant endoxylanase inhibitors XIP-I and TAXI-I. *Biochem. Biophys. Res. Commun.* **337**, 160–166.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C. *et al.* (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**, 810–820.

Casamitjana-Martínez, E., Hofhuis, H.F., Xu, J., Liu, C.-M., Heidstra, R. and Scheres, B. (2003) Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of Arabidopsis root meristem maintenance. *Curr. Biol.* **13**, 1435–1441.

Chapman, E.J., Prokhnevsky, A.I., Gopinath, K., Dolja, V.V. and Carrington, J.C. (2004) Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes Dev.* **18**, 1179–1186.

Charmont, S., Jamet, E., Pont-Lezica, R. and Canut, H. (2005) Proteomic analysis of secreted proteins from *Arabidopsis thaliana* seedlings: improved recovery following removal of phenolic compounds. *Phytochemistry*, **66**, 453–461.

Chen, L.-Q. (2014) SWEET sugar transporters for phloem transport and pathogen nutrition. *New Phytol.* **201**, 1150–1155.

Contour-Ansel, D., Torres-Franklin, M.L., Zuily-Fodil, Y. and de Carvalho, M.H.C. (2010) An aspartic acid protease from common bean is expressed 'on call' during water stress and early recovery. *J. Plant Physiol.* **167**, 1606–1612.

Cosgrove, D.J. (2016) Catalysts of plant cell wall loosening. *F1000Research*, **5**, (F1000 Faculty Rev): 119. https://doi.org/10.12688/f1000research.7180.1

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. *J. Proteome Res.* **10**, 1794–1805.

Crusoe, M.R., Alameldin, H.F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., Charbonneau, A. *et al.* (2015) The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research*, **4**, 900.

Dagdas, Y.F., Belhaj, K., Maqbool, A., Chaparro-Garcia, A., Pandey, P., Petre, B., Tabassum, N. *et al.* (2016) An effector of the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. *eLife*, **5**, e10856.

Dogru, E., Warzecha, H., Seibel, F., Haebel, S., Lottspeich, F. and Stöckigt, J. (2000) The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid biosynthesis in plants is an ortholog of the α/β-hydrolase super family. *Eur. J. Biochem.* **267**, 1397–1406.

Duan, X., Zhang, Z., Wang, J. and Zuo, K. (2016) Characterization of a novel cotton subtilase gene GbSBT1 in response to extracellular stimulations and its role in Verticillium resistance. *PLoS ONE*, **11**, e0153988.

Duwadi, K., Chen, L., Menassa, R. and Dhaubhadel, S. (2015) Identification, characterization and down-regulation of cysteine protease genes in tobacco for use in recombinant protein production. *PLoS ONE*, **10**, e0130556.

Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G. and Brunak, S. (2004) Improved prediction of signal peptides: signalP 3.0. *J. Mol. Biol.* **340**, 783–795.

Ernst, O. and Zor, T. (2010) Linearization of the Bradford protein assay. *J. Vis. Exp. JoVE.*, **38**, http://www.jove.com/details.php?id=1918, https://doi.org/10.3791/1918.

van Esse, H.P., van't Klooster, J.W., Bolton, M.D., Yadeta, K.A., van Baarlen, P., Boeren, S., Vervoort, J. *et al.* (2008) The *Cladosporium fulvum* virulence protein Avr2 inhibits host proteases required for basal defense. *Plant Cell*, **20**, 1948–1963.

Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Falk, A., Feys, B.J., Frost, L.N., Jones, J.D.G., Daniels, M.J. and Parker, J.E. (1999) EDS1, an essential component of R gene-mediated disease resistance in Arabidopsis has homology to eukaryotic lipases. *Proc. Natl Acad. Sci. USA*, **96**, 3292–3297.

Faro, C. and Gal, S. (2005) Aspartic proteinase content of the Arabidopsis genome. *Curr. Protein Pept. Sci.* **6**, 493–500.

Forouhar, F., Yang, Y., Kumar, D., Chen, Y., Fridman, E., Park, S.W., Chiang, Y. *et al.* (2005) Structural and biochemical studies identify tobacco SABP2 as a methyl salicylate esterase and implicate it in plant innate immunity. *Proc. Natl Acad. Sci. USA*, **102**, 1773–1778.

Franziska, F.S. (2013) Snap-shot of Serine Carboxypeptidase-Like acyltransferase evolution: the loss of conserved disulphide bridge is responsible for the completion of neo-functionalization. *J. Phylogenetics Evol. Biol.* **01**, 115.

Fraser, C.M., Rider, L.W. and Chapple, C. (2005) An expression and bioinformatics analysis of the Arabidopsis serine Carboxypeptidase-like gene family. *Plant Physiol.* **138**, 1136–1148.

Gao, H., Zhang, Y., Wang, W., Zhao, K., Liu, C., Bai, L., Li, R. *et al.* (2017) Two membrane-anchored aspartic proteases contribute to pollen and ovule development. *Plant Physiol.* **173**, 219–239.

Gepstein, S., Sabehi, G., Carp, M.-J., Hajouj, T., Nesher, M.F.O., Yariv, I., Dor, C. *et al.* (2003) Large-scale identification of leaf senescence-associated genes. *Plant J.* **36**, 629–642.

Ghorbani, S., Hoogewijs, K., Pečenková, T., Fernandez, A., Inzé, A., Eeckhout, D., Kawa, D. *et al.* (2016) The SBT6.1 subtilase processes the GOLVEN1 peptide controlling cell elongation. *J. Exp. Bot.* **67**, 4877–4887.

Gilroy, E.M., Hein, I., van der Hoorn, R.A.L., Boevink, P.C., Venter, E., McLellan, H., Kaffarnik, F. *et al.* (2007) Involvement of cathepsin B in the plant disease resistance hypersensitive response. *Plant J.* **52**, 1–13.

Goetz, M., Godt, D.E., Guivarc'h, A., Kahmann, U., Chriqui, D. and Roitsch, T. (2001) Induction of male sterility in plants by metabolic engineering of the carbohydrate supply. *Proc. Natl Acad. Sci. USA*, **98**, 6522–6527.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

Goulet, C., Goulet, C., Goulet, M.-C. and Michaud, D. (2010) 2-DE proteome maps for the leaf apoplast of *Nicotiana benthamiana*. *Proteomics*, **10**, 2536–2544.

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 Biotechnol. J.* **10**, 83–94.

Greenbaum, D.C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradszky, K.F., Engel, J., DeRisi, J. *et al.* (2002) A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science*, **298**, 2002–2006.

Griffiths, J., Murase, K., Rieu, I., Zentella, R., Zhang, Z.-L., Powers, S.J., Gong, F. *et al.* (2006) Genetic characterization and functional analysis of the GID1 gibberellin receptors in Arabidopsis. *Plant Cell*, **18**, 3399–3414.

Grosse-Holz, F.M. and van der Hoorn, R.A.L. (2016) Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants. *New Phytol.* **210**, 794–807.

Guo, Y., Zheng, Z., Clair, J.J.L., Chory, J. and Noel, J.P. (2013) Smoke-derived karrikin perception by the α/β-hydrolase KAI2 from Arabidopsis. *Proc. Natl Acad. Sci. USA*, **110**, 8284–8289.

Hao, L., Hsiang, T. and Goodwin, P.H. (2006) Role of two cysteine proteinases in the susceptible response of *Nicotiana benthamiana* to *Colletotrichum destructivum* and the hypersensitive response to *Pseudomonas syringae* pv. tomato. *Plant Sci.* **170**, 1001–1009.

Hatsugai, N., Kuroyanagi, M., Yamada, K., Meshi, T., Tsuda, S., Kondo, M., Nishimura, M. *et al.* (2004) A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science*, **305**, 855–858.

Hatsugai, N., Yamada, K., Goto-Yamada, S. and Hara-Nishimura, I. (2015) Vacuolar processing enzyme in plant programmed cell death. *Front. Plant Sci.* **6**, 234.

Hayashi, Y., Yamada, K., Shimada, T., Matsushima, R., Nishizawa, N., 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 Cell Physiol.* **42**, 894–899.

Heath, R.L., Barton, P.A., Simpson, R.J., Reid, G.E., Lim, G. and Anderson, M.A. (1995) Characterization of the protease processing sites in a multidomain proteinase inhibitor precursor from *Nicotiana alata*. *Eur. J. Biochem.* **230**, 250–257.

Hehle, V.K., Paul, M.J., Drake, P.M., Ma, J.K. and van Dolleweerd, C.J. (2011) Antibody degradation in tobacco plants: a predominantly apoplastic process. *BMC Biotechnol.* **11**, 128.

Hehle, V.K., Lombardi, R., van Dolleweerd, C.J., Paul, M.J., Di Micco, P., Morea, V., Benvenuto, E. *et al.* (2015) Site-specific proteolytic degradation of IgG monoclonal antibodies expressed in tobacco plants. *Plant Biotechnol. J.* **13**, 235–245.

Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.

Hückelhoven, R., Dechert, C., Trujillo, M. and Kogel, K.-H. (2001) Differential expression of putative cell death regulator genes in near-isogenic, resistant and susceptible barley lines during interaction with the powdery mildew fungus. *Plant Mol. Biol.* **47**, 739–748.

Husted, S. and Schjoerring, J.K. (1995) Apoplastic pH and ammonium concentration in leaves of *Brassica napus* L. *Plant Physiol.* **109**, 1453–1460.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W. and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**, 119.

Ivanchenko, M.G., den Os, D., Monshausen, G.B., Dubrovsky, J.G., Bednářová, A. and Krishnan, N. (2013) Auxin increases the hydrogen peroxide concentration in tomato root tips while inhibiting root growth. *Ann. Bot.* **112**, 1107–1116.

Jordá, L., Coego, A., Conejero, V. and Vera, P. (1999) A genomic cluster containing four differentially regulated subtilisin-like processing protease genes is in tomato plants. *J. Biol. Chem.* **274**, 2360–2365.

Kang, S.-R., Oh, S.-K., Kim, J.-J., Choi, D.-I. and Baek, K.-H. (2010) NMMP1, a matrix metalloprotease in *Nicotiana benthamiana* has a role in protection against bacterial infection. *Plant Pathol. J.* **26**, 402–408.

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*. *Mol. Cell Proteomics*, **8**, 1082–1093.

Kaschani, F., Shabab, M., Bozkurt, T., Shindo, T., Schornack, S., Gu, C., Ilyas, M. *et al.* (2010) An effector-targeted protease contributes to defense against Phytophthora infestans and is under diversifying selection in natural hosts. *Plant Physiol.* **154**, 1794–1804.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S. *et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kim, N.-S., Kim, T.-G., Kim, O.-H., Ko, E.-M., Jang, Y.-S., Jung, E.-S., Kwon, T.-H. *et al.* (2008) Improvement of recombinant hGM-CSF production by suppression of cysteine proteinase gene expression using RNA interference in a transgenic rice culture. *Plant Mol. Biol.* **68**, 263–275.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013a) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.

Kim, H.G., Kwon, S.J., Jang, Y.J., Nam, M.H., Chung, J.H., Na, Y.-C., Guo, H. *et al.* (2013b) GDSL LIPASE1 modulates plant immunity through feedback regulation of ethylene signaling. *Plant Physiol.* **163**, 1776–1791.

Kopylova, E., Noé, L. and Touzet, H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

Kovács, J. and van der Hoorn, R.A.L. (2016) Twelve ways to confirm targets of activity-based probes in plants. *Bioorg. Med. Chem.* Target Identification of Small Molecule Drugs and Chemical Probes **24**, 3304–3311.

Krüger, J., Thomas, C.M., Golstein, C., Dixon, M.S., Smoker, M., Tang, S., Mulder, L. *et al.* (2002) A tomato cysteine protease required for Cf-2-dependent disease resistance and suppression of autonecrosis. *Science*, **296**, 744–747.

Lampl, N., Alkan, N., Davydov, O. and Fluhr, R. (2013) Set-point control of RD21 protease activity by AtSerpin1 controls cell death in Arabidopsis. *Plant J.* **74**, 498–510.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245.

Li, Y., Kabbage, M., Liu, W. and Dickman, M.B. (2016) Aspartyl protease-mediated cleavage of BAG6 is necessary for autophagy and fungal resistance in plants. *Plant Cell*, **28**, 233–247.

Li, W., Cao, J.-Y., Xu, Y.-P. and Cai, X.-Z. (2017) Artificial Agrobacterium tumefaciens strains exhibit diverse mechanisms to repress *Xanthomonas oryzae* pv. oryzae-induced hypersensitive response and non-host resistance in *Nicotiana benthamiana*. *Mol. Plant Pathol.* **18**, 489–502.

Limoli, D.H., Jones, C.J. and Wozniak, D.J. (2015) Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiol. Spectr.* **3**, 3.

Liu, H., Wang, X., Zhang, H., Yang, Y., Ge, X. and Song, F. (2008) A rice serine carboxypeptidase-like gene OsBISCPL1 is involved in regulation of defense responses against biotic and oxidative stress. *Gene*, **420**, 57–65.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Mandal, M.K., Fischer, R., Schillberg, S. and Schiermeyer, A. (2014) Inhibition of protease activity by antisense RNA improves recombinant protein production in *Nicotiana tabacum* cv. Bright Yellow 2 (BY-2) suspension cells. *Biotechnol. J.* **9**, 1065–1073.

Mandal, M.K., Ahvari, H., Schillberg, S. and Schiermeyer, A. (2016) Tackling unwanted proteolysis in plant production hosts used for molecular farming. *Front. Plant Sci.* **7**, 267.

Martin, K., Kopperud, K., Chakrabarty, R., Banerjee, R., Brooks, R. and Goodin, M.M. (2009) Transient expression in *Nicotiana benthamiana* fluorescent marker lines provides enhanced definition of protein localization, movement and interactions in planta. *Plant J.* **59**, 150–162.

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. *New Phytol.* **183**, 408–418.

Mindrebo, J.T., Nartey, C.M., Seto, Y., Burkart, M.D. and Noel, J.P. (2016) Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. *Curr. Opin. Struct. Biol.* Multi-protein assemblies in signaling • Catalysis and regulation, **41**, 233–246.

Misas-Villamil, J.C., van der Burgh, A.M., Grosse-Holz, F., Bach-Pages, M., Kovács, J., Kaschani, F., Schilasky, S. *et al.* (2017) Subunit-selective proteasome activity profiling uncovers uncoupled proteasome subunit activities during bacterial infections. *Plant J.* **90**, 418–430.

Moura, D.S., Bergey, D.R. and Ryan, C.A. (2001) Characterization and localization of a wound-inducible type I serine-carboxypeptidase from leaves of tomato plants (*Lycopersicon esculentum* Mill.). *Planta*, **212**, 222–230.

Mylne, J.S., Colgrave, M.L., Daly, N.L., Chanson, A.H., Elliott, A.G., McCallum, E.J., Jones, A. *et al.* (2011) Albumins and their processing machinery are hijacked for cyclic peptides in sunflower. *Nat. Chem. Biol.* **7**, 257–259.

Nakai, A., Yamauchi, Y., Sumi, S. and Tanaka, K. (2012) Role of acylamino acid-releasing enzyme/oxidized protein hydrolase in sustaining homeostasis of the cytoplasmic antioxidative system. *Planta*, **236**, 427–436.

Nakaune, S., Yamada, K., Kondo, M., Kato, T., Tabata, S., Nishimura, M. and Hara-Nishimura, I. (2005) A vacuolar processing enzyme, delta VPE, is involved in seed coat formation at the early stage of seed development. *Plant Cell*, **17**, 876–887.

NCBI Resource Coordinators. (2017) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **45**, D12–D17.

Nielsen, A.Z., Ziersen, B., Jensen, K., Lassen, L.M., Olsen, C.E., Møller, B.L. and Jensen, P.E. (2013) Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth. Biol.* **2**, 308–315.

Niemer, M., Mehofer, U., Torres Acosta, J.A., Verdianz, M., Henkel, T., Loos, A., Strasser, R. *et al.* (2014) The human anti-HIV antibodies 2F5, 2G12, and PG9 differ in their susceptibility to proteolytic degradation: down-regulation of endogenous serine and cysteine proteinase activities could improve antibody production in plant-based expression platforms. *Biotechnol. J.* **9**, 493–500.

Nikolenko, S.I., Korobeynikov, A.I. and Alekseyev, M.A. (2013) BayesHammer: bayesian clustering for error correction in single-cell sequencing. *BMC Genom.* **14**, S7.

Paireder, M., Mehofer, U., Tholen, S., Porodko, A., Schähs, P., Maresch, D., Biniossek, M.L. *et al.* (2016) The death enzyme CP14 is a unique papain-like cysteine proteinase with a pronounced S2 subsite selectivity. *Arch. Biochem. Biophys.* **603**, 110–117.

Paireder, M., Tholen, S., Porodko, A., Biniossek, M.L., Mayer, B., Novinec, M., Schilling, O. *et al.* (2017) The papain-like cysteine proteinases NbCysP6 and NbCysP7 are highly processive enzymes with substrate specificities complementary to *Nicotiana benthamiana* cathepsin B. *Biochim. Biophys. Acta BBA - Proteins Proteomics*, **1865**, 444–452.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**, 417–419.

Pearce, G., Yamaguchi, Y., Barona, G. and Ryan, C.A. (2010) A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. *Proc. Natl Acad. Sci. USA*, **107**, 14921–14925.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T. and Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295.

Petre, B., Saunders, D.G.O., Sklenar, J., Lorrain, C., Krasileva, K.V., Win, J., Duplessis, S. *et al.* (2016) Heterologous expression screens in *Nicotiana benthamiana* identify a candidate effector of the wheat yellow rust pathogen that associates with processing bodies. *PLoS ONE*, **11**, e0149035.

Pillet, S., Aubin, É., Trépanier, S., Bussière, D., Dargis, M., Poulin, J.-F., Yassine-Diab, B. *et al.* (2016) A plant-derived quadrivalent virus like particle influenza vaccine induces cross-reactive antibody and T cell response in healthy adults. *Clin. Immunol.* **168**, 72–87.

Pitzschke, A. (2013) Agrobacterium infection and plant defense—transformation success hangs by a thread. *Front. Plant Sci.* **4**, 519.

Pruss, G.J., Nester, E.W. and Vance, V. (2008) Infiltration with *Agrobacterium tumefaciens* induces host defense and development-dependent responses in the infiltrated zone. *Mol. Plant Microbe Interact.* **21**, 1528–1538.

Qiu, X., Wong, G., Audet, J., Bello, A., Fernando, L., Alimonti, J.B., Fausther-Bovendo, H. *et al.* (2014) Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. *Nature*, **514**, 47–53.

Rawlings, N.D., Waller, M., Barrett, A.J. and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509.

Renko, M., Sabotič, J. and Turk, D. (2012) β-Trefoil inhibitors – from the work of Kunitz onward. *Biol. Chem.* **393**, 1043–1054.

Richau, K.H., Kaschani, F., Verdoes, M., Pansuriya, T.C., Niessen, S., Stuber, K., Colby, T. *et al.* (2012) Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. *Plant Physiol.* **158**, 1583–1599.

Rico, A., Bennett, M.H., Forcat, S., Huang, W.E. and Preston, G.M. (2010) Agroinfiltration reduces ABA levels and suppresses *Pseudomonas syringae*-elicited salicylic acid production in *Nicotiana tabacum*. *PLoS ONE*, **5**, e8977.

Robinette, D. and Matthysse, A.G. (1990) Inhibition by *Agrobacterium tumefaciens* and *Pseudomonas savastanoi* of development of the hypersensitive response elicited by *Pseudomonas syringae* pv. phaseolicola. *J. Bacteriol.* **172**, 5742–5749.

Rojo, E., Zouhar, J., Carter, C., Kovaleva, V. and Raikhel, N.V. (2003) A unique mechanism for protein processing and degradation in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **100**, 7389–7394.

Rose, R., Schaller, A. and Ottmann, C. (2010) Structural features of plant subtilases. *Plant Signal. Behav.* **5**, 180–183.

Sainsbury, F., Varennes-Jutras, P., Goulet, M.-C., D'Aoust, M.-A. and Michaud, D. (2013) Tomato cystatin SlCYS8 as a stabilizing fusion partner for human serpin expression in plants. *Plant Biotechnol. J.* **11**, 1058–1068.

Sansen, S., Ranter, C.J.D., Gebruers, K., Brijs, K., Courtin, C.M., Delcour, J.A. and Rabijns, A. (2004) Structural basis for inhibition of *Aspergillus niger* xylanase by *Triticum aestivum* xylanase inhibitor-I. *J. Biol. Chem.* **279**, 36022–36028.

Saur, I.M.L., Kadota, Y., Sklenar, J., Holton, N.J., Smakowska, E., Belkhadir, Y., Zipfel, C. *et al.* (2016) NbCSPR underlies age-dependent immune responses to bacterial cold shock protein in *Nicotiana benthamiana*. *Proc. Natl Acad. Sci. USA*, **113**, 3389–3394.

Serrano, I., Buscaill, P., Audran, C., Pouzet, C., Jauneau, A. and Rivas, S. (2016) A non canonical subtilase attenuates the transcriptional activation of defence responses in *Arabidopsis thaliana*. *eLife*, **5**, e19755.

Sheikh, A.H., Raghuram, B., Eschen-Lippold, L., Scheel, D., Lee, J. and Sinha, A.K. (2014) Agroinfiltration by cytokinin-producing *Agrobacterium* sp. strain GV3101 primes defense responses in *Nicotiana tabacum*. *Mol. Plant Microbe Interact.* **27**, 1175–1185.

Shigeyama, T., Watanabe, A., Tokuchi, K., Toh, S., Sakurai, N., Shibuya, N. and Kawakami, N. (2016) α-Xylosidase plays essential roles in xyloglucan remodelling, maintenance of cell wall integrity, and seed germination in *Arabidopsis thaliana*. *J. Exp. Bot.* **67**, 5615–5629.

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*, **7**, e29317.

Simpson, J.T. and Durbin, R. (2012) Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* **22**, 549–556.

Stapleton, A., Beetham, J.K., Pinot, F., Garbarino, J.E., Rockhold, D.R., Friedman, M., Hammock, B.D. *et al.* (1994) Cloning and expression of soluble epoxide hydrolase from potato. *Plant J. Cell Mol. Biol.* **6**, 251–258.

Stoger, E., Fischer, R., Moloney, M. and Ma, J.K.-C. (2014) Plant molecular pharming for the treatment of chronic and infectious diseases. *Annu. Rev. Plant Biol.* **65**, 743–768.

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 Phytol.* **203**, 913–925.

Szalontai, B., Stranczinger, S., Palfalvi, G., Mauch-Mani, B. and Jakab, G. (2012) The taxon-specific paralogs of grapevine PRLIP genes are highly induced upon powdery mildew infection. *J. Plant Physiol.* **169**, 1767–1775.

Tamaki, T., Betsuyaku, S., Fujiwara, M., Fukao, Y., Fukuda, H. and Sawa, S. (2013) SUPPRESSOR OF LLP1 1-mediated C–terminal processing is critical for CLE19 peptide activity. *Plant J.* **76**, 970–981.

Tang, S., Lomsadze, A. and Borodovsky, M. (2015) Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**, e78.

Taylor, A. and Qiu, Y.-L. (2017) Evolutionary history of subtilases in land plants and their involvement in symbiotic interactions. *Mol. Plant Microbe Interact.* **30**, 489–501.

The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169.

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. *J. Biol. Chem.* **279**, 26370–26377.

Tripathi, L.P. and Sowdhamini, R. (2006) Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genom.* **7**, 200.

Tyanova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319.

Van der Hoorn, R.A.L., Rivas, S., Wulff, B.B.H., Jones, J.D.G. and Joosten, M.H.A.J. (2003) Rapid migration in gel filtration of the Cf-4 and Cf-9 resistance proteins is an intrinsic property of Cf proteins and not because of their association with high-molecular-weight proteins. *Plant J.* **35**, 305–315.

Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G. *et al.* (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456.

Wagner, U.G., Hasslacher, M., Griengl, H., Schwab, H. and Kratky, C. (1996) Mechanism of cyanogenesis: the crystal structure of hydroxynitrile lyase from Hevea brasiliensis. *Structure*, **4**, 811–822.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A. *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, **9**, e112963.

Wang, D., Weaver, N.D., Kesarwani, M. and Dong, X. (2005) Induction of protein secretory pathway is required for systemic acquired resistance. *Science*, **308**, 1036–1040.

Xia, Y., Suzuki, H., Borevitz, J., Blount, J., Guo, Z., Patel, K., Dixon, R.A. *et al.* (2004) An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. *EMBO J.* **23**, 980–988.

Xu, Q.-F., Cheng, W.-S., Li, S.-S., Li, W., Zhang, Z.-X., Xu, Y.-P., Zhou, X.-P. *et al.* (2012) Identification of genes required for Cf-dependent hypersensitive cell death by combined proteomic and RNA interfering analyses. *J. Exp. Bot.* **63**, 2421–2435.

Yamauchi, Y., Ejiri, Y., Toyoda, Y. and Tanaka, K. (2003) Identification and biochemical characterization of plant acylamino acid–releasing enzyme. *J. Biochem. (Tokyo)* **134**, 251–257.

Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., Yu, C. *et al.* (2016) DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature*, **536**, 469–473.

Yusibov, V., Kushnir, N. and Streatfield, S.J. (2016) Antibody production in plants and green algae. *Annu. Rev. Plant Biol.* **67**, 669–701.

Zhao, Y., Thilmony, R., Bender, C.L., Schaller, A., He, S.Y. and Howe, G.A. (2003) Virulence systems of *Pseudomonas syringae* pv. tomato promote bacterial speck disease in tomato by targeting the jasmonate signaling pathway. *Plant J.* **36**, 485–499.

Zheng, Z., Guo, Y., Novák, O., Chen, W., Ljung, K., Noel, J.P. and Chory, J. (2016) Local auxin metabolism regulates environment-induced hypocotyl elongation. *Nat. Plants*, **2**, 16025.

Zhou, Y., Cox, A.M. and Kearney, C.M. (2017) Pathogenesis-related proteins induced by agroinoculation-associated cell wall weakening can be obviated by spray-on inoculation or mannitol ex vivo culture. *Plant Biotechnol. Rep.* **11**, 1–9.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 Fold change of transcripts differential between P19 and WT agroinfiltrated leaves.
Figure S2 Protein concentration in apoplastic fluid over time.
Figure S3 A phylogenetic tree of proteasome subunits in MEROPS family T01
Figure S4 Malate dehydrogenase activity in apoplastic fluid.
Table S1 Differential transcript abundance data, comparing WT and p19 agroinfiltrated leaves
Table S2 PFAM Domains overrepresented among transcripts differential between WT and P19 agroinfiltrated leaves
Table S3 Differential protein abundance data, comparing WT and p19 agroinfiltrated leaves
Table S4 Differential transcript abundance data, comparing agro- and mock infiltrated leaves
Table S5 PFAM Domains overrepresented among transcripts differential between agro- and mock infiltrated leaves
Table S6 Differential protein abundance data, comparing agro- and mock infiltrated leaves
Table S7 PFAM Domains overrepresented among proteins differential between agro- and mock infiltrated leaves
Table S8 Agrobacterium proteins for which corresponding peptides were identified in the extracellular proteome
Table S9 Differential transcript abundance data over time
Table S10 Differential protein abundance data over time
Table S11 Data from ABPP-MS analyses
Table S12 Discrepancies between changes in extracellular protein and transcript levels
Table S13 Discrepancies between changes in extracellular activity and extracellular protein levels
Table S14 Protease family sizes in Arabidopsis, tomato, rice and *N. benthamiana*
Table S15 PFAM Families mapped to MEROPS families
Appendix S1 Detailing curation of the proteome database
Appendix S2 R code used for RNAseq data analysis
Appendix S3 R code used for extracellular proteome data analysis
Appendix S4 R code used for analysis of the effects of leaf ageing
Appendix S5 R code used for ABPP-MS data analysis
Appendix S6 R code used for analysis of discrepancies between changes in extracellular activity, extracellular protein abundance and transcript abundance
Appendix S7 Supplemental methods used for mass spectrometry sample preparation
Appendix S8 Full versions of the trees shown in Figures 5 and 6 with all gene names

New Phytologist

Review

Tansley review

Juggling jobs: roles and mechanisms of multifunctional protease inhibitors in plants

Author for correspondence:
Renier A. L. van der Hoorn
Tel: +44 0 1865 275077
Email: renier.vanderhoorn@plants.ox.ac.uk

Received: *27 August 2015*
Accepted: *1 December 2015*

Friederike M. Grosse-Holz and Renier A. L. van der Hoorn

Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

Contents

New Phytologist (2016)
doi: 10.1111/nph.13839

Key words: plant defence, protease inhibitors, protein structure, regulatory networks, resistance breeding, storage protein protection.

Summary

Multifunctional protease inhibitors juggle jobs by targeting different enzymes and thereby often controlling more than one biological process. Here, we discuss the biological functions, mechanisms and evolution of three types of multifunctional protease inhibitors in plants. The first type is double-headed inhibitors, which feature two inhibitory sites targeting proteases with different specificities (e.g. Bowman–Birk inhibitors) or even different hydrolases (e.g. α-amylase/protease inhibitors preventing both early germination and seed predation). The second type consists of multidomain inhibitors which evolved by intragenic duplication and are released by processing (e.g. multicystatins and potato inhibitor II, implicated in tuber dormancy and defence, respectively). The third type consists of promiscuous inhibitory folds which resemble mouse traps that can inhibit different proteases cleaving the bait they offer (e.g. serpins, regulating cell death, and α-macroglobulins). Understanding how multifunctional inhibitors juggle biological jobs increases our knowledge of the connections between the networks they regulate. These examples show that multifunctionality evolved independently from a remarkable diversity of molecular mechanisms that can be exploited for crop improvement and provide concepts for protein design.

I. Introduction

Multifunctional protease inhibitors are classic examples that contradict the old dogma of one protein – one function (Beadle & Tatum, 1941). Multifunctionality implies that a protein or a structural fold can regulate various partners, resulting in different biological outputs. Approximately 30 yr after the influential paper of Beadle and Tatum, the first report of a multifunctional protease inhibitor (Odani & Ikenaka, 1973) challenged their model. A protein from soybean (*Glycine max*) was separated into two parts, of which one inhibited trypsin and the other chymotrypsin (Odani & Ikenaka, 1973). Trypsin and chymotrypsin are closely related Ser proteases, but they differ in the surface loops that determine substrate specificity (Hedstrom *et al.*, 1992) and thus require distinct inhibitory sites. This first bifunctional protease inhibitor belongs to the Bowman–Birk inhibitors (BBIs), classified in the MEROPS database as family I12 (see Box 1 for a brief description of the MEROPS system). Only a few years after the discovery that a single protein can inhibit two distinct proteases, it became clear that the Indian staple crop ragi (*Eleusine coracana*) produces an even

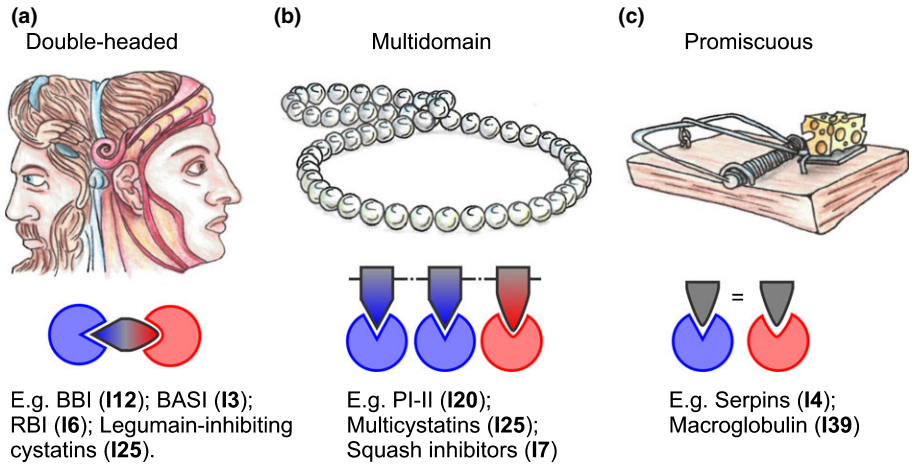


Fig. 1 Three classes of multifunctional protease inhibitors. BBI, Bowman–Birk inhibitor; BASI, barley α-amylase/subtilisin inhibitor; RBI, ragi bifunctional inhibitor; PI II, potato peptidase inhibitor II; MEROPS families are given in brackets.

more peculiar multifunctional inhibitor (Shivaraj & Pattabiraman, 1981). The ragi bifunctional inhibitor (RBI, family I6) can form a trimeric complex with α-amylase and trypsin, thereby simultaneously inactivating a starch-degrading enzyme and a protease (Shivaraj & Pattabiraman, 1981). Two years later, it turned out that a protein from barley (*Hordeum vulgare*) also targets an α-amylase and a protease (Mundy *et al.*, 1983). However, barley α-amylase/subtilisin inhibitor (BASI, family I3) is structurally unrelated to RBI. The bifunctional BBI, RBI and BASI were the first of many multifunctional plant protease inhibitors to be discovered. In these three cases, multifunctionality is facilitated by two inhibitory sites on a single protein. More recent research has characterized other types of multifunctional protease inhibitors. Some of these contain more than one inhibitory domain (multidomain inhibitors) or use the same binding site to inhibit one of various target proteases at a time (promiscuous inhibitors) (Fig. 1). The frequent occurrence of multifunctionality among protease inhibitors has raised many questions about the biological roles of these proteins. Are multifunctional inhibitors regulatory links between the proteases they affect? How has multifunctionality evolved in different inhibitor structures and do they share common 'roads to multifunctionality'? How can multifunctional protease inhibitors be exploited for crop improvement and protein design? These questions are addressed in this review. With multifunctional plant protease inhibitors, we bring together a range of proteins and biological processes that are not usually studied jointly. Awareness of a potential hub between proteolytic networks will help to gain a comprehensive overview. Finally, understanding the inherent power of multifunctional protease inhibitors will prove useful to design successful agricultural or biotechnological strategies.

Box 1 The MEROPS database of proteases and their inhibitors (http://merops.sanger.ac.uk, (Rawlings *et al.*, 2014)).

MEROPS groups proteins into families based on sequence homology and families into clans based on structural homology. Proteases or inhibitors in a clan are assumed to have evolved from a common ancestor, but are less closely related than the proteins within a family. Inhibitor families are named I1 to I93. Protease families are named with a letter indicating the catalytic type (i.e. A for aspartic, S for serine and C for cysteine proteases), followed by a consecutive number. We refer to MEROPS release 9.12 throughout this article.

II. Three types of multifunctionality

Juggling several biological jobs is facilitated in different ways, which we group into three types of inhibitor multifunctionality (Fig. 1). The first type of multifunctional inhibitors are Janus-type inhibitors, double-headed proteins with two inhibitory faces matching distinct targets. The second type are multidomain proteins, which resemble a necklace of pearls (inhibitor domains). Within the pearl necklace type, the I20 inhibitor domains have diversified with regard to specificity, while the macrocyclic cystine knot peptides can have both inhibitory and noninhibitory functions. The third type of multifunctionality is represented by promiscuous inhibitors such as serpins and α-macroglobulins. These structurally different inhibitors act as mouse traps that inhibit different proteases by undergoing dramatic conformational change upon protease binding.

We summarize the current scientific knowledge on biological role, inhibitory mechanism and evolution for eight families of multifunctional plant protease inhibitors. A comprehensive overview of all discussed families and the examples mentioned throughout this review is given in Table 1. The inhibitors are first discussed in groups based on their type of multifunctionality. We conclude by discussing all eight multifunctional inhibitor families in the context of common biological roles, evolutionary history and potential applications.

1. Double-headed inhibitors: the Janus type

With his two faces looking in opposite directions, the ancient Roman god Janus provides a good metaphor for bifunctional inhibitors. Each of the two faces stands for a binding site on which a target enzyme can be inhibited. Surprisingly many protein architectures facilitate multifunctionality by providing two sites

Table 1 Plant protease inhibitor families discussed in this review

Name	Family	Type	Inhibitory mechanism	Target enzymes	Implicated in defence	Implicated in development	Examples discussed in this review
Bowman–Birk inhibitors (BBIs)	I12	DH	Laskowski	Trypsin (S1), chymotrypsin (S1)	Expression induced during plant immune responses[1] Overexpression increases resistance to a fungal pathogen[1]	ND	BBI (*Glycine max*, *Medicago truncatula*, *Oryza sativa* and *Zea mays*)
Ragi bifunctional inhibitor (RBI)	I6	DH	Laskowski	α-amylase, trypsin (S1)	Inhibits α-amylases from organisms that feed on seed, such as insects[3] and mammals[4]	ND	RBI (*Eleusine coracana*)
Kunitz inhibitors targeting different proteases	I3	DH	Laskowski (BASI: modified Laskowski)	α-amylase, subtilisin-like proteases (S8), PLCPs (C1), trypsin (S1), chymotrypsin (S1) and cathepsin D (A1)	Inhibit insect gut proteases[7] and α-amylases[2] Reduce viability and fertility of insects[8]	Inhibit α-amylases involved in germination[5]	BASI (*Hordeum vulgare*) ApKTI (*Adenanthera pavonina*), PSPI (*Solanum tuberosum*) KTI3 (*Populus trichocarpa* ×*Populus deltoides*)
Legumain-inhibiting cystatins	I25	DH	cystatin mechanism (C1) and substrate mimicry (C13)	PLCPs (C1) and legumains (C13)	ND	May prevent early germination[6]	AtCYS6 (*Arabidopsis thaliana*)
Multicystatins	I25	MD		PLCPs (C1)	Expression induced during plant immune responses[9] Inhibit insect gut proteases[10] Impair the growth of insects and fungal pathogens[11]	Regulate storage protein accumulation in tubers[12]	PMC (*Solanum tuberosum*)
Multidomain potato peptidase inhibitor II	I20	MD	Laskowski	Trypsin (S1), chymotrypsin (S1) and subtilisin (S8)	Expression increases upon wounding[13] Inhibit insect gut proteases[14] Overexpression increases resistance to insect pests[15]	ND	PI-II (*Solanum tuberosum*) and NaProPI (*Nicotiana alata*)
Squash inhibitors	I7	MD	Laskowski	Trypsin (S1)	ND	ND	MCoTI-II (*Momordica cochinchinensis*)
Serpins	I4	P	Serpin mechanism (deformation)	PLCPs (C1), metacaspases (C14) and chymotrypsin (S1)	Impair insect growth and fertility[16]	ND	AtSerpin1 (*Arabidopsis thaliana*) and BSZx (*Hordeum vulgare*)
α-macroglobulins	I39	P	Macroglobulin mechanism (caging)	Endopeptidases	ND	ND	A2M (*Cucumis sativus*, *Fragaria vesca*, *Micromonas* sp. *RCC299* and *Populus trichocarpa*)

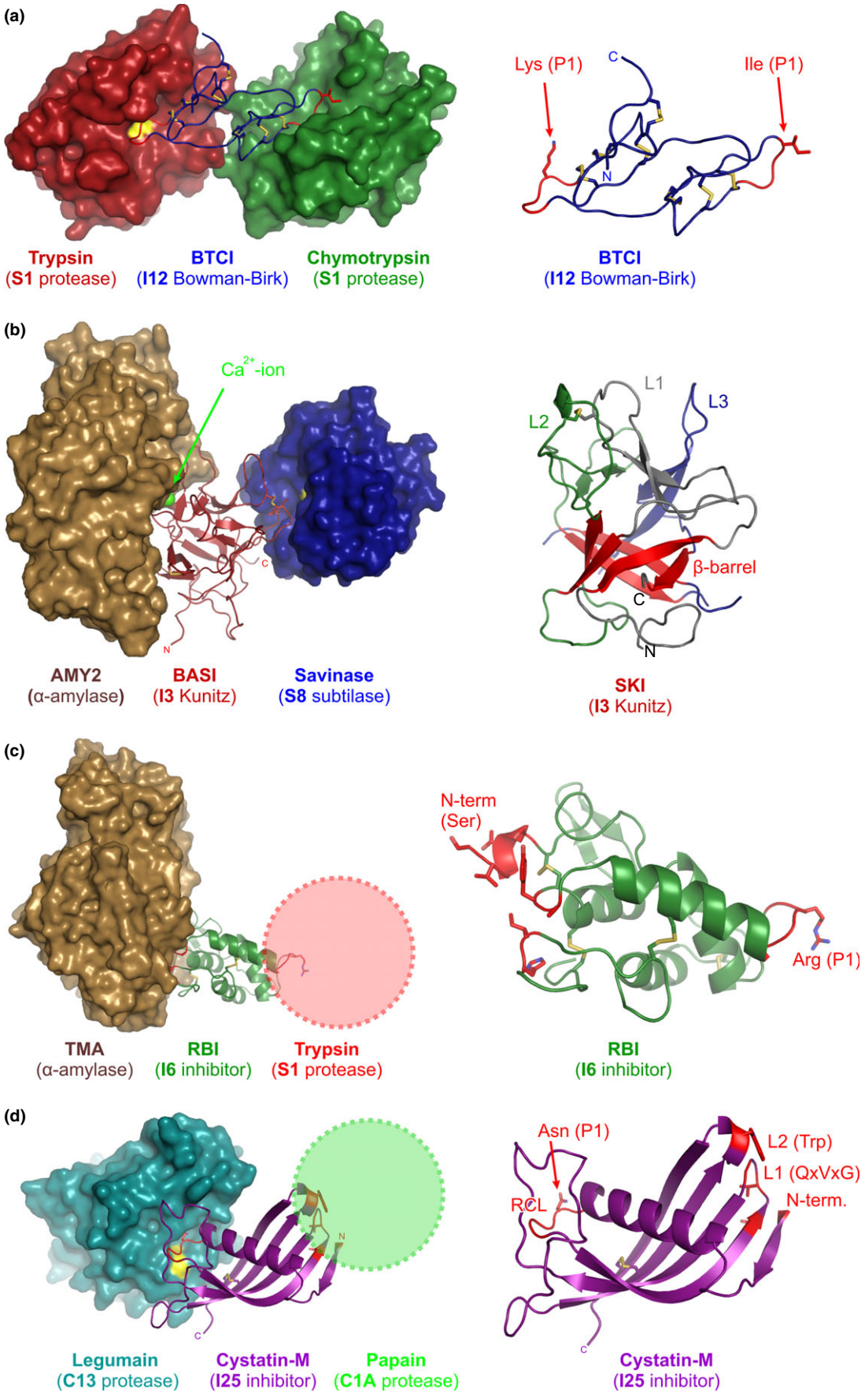
A2M, α2-macroglobulin; ApKTI, *Adenanthera pavonina* Kunitz type inhibitor; AtCYS6, *A. thaliana* cystatin 6; BASI, barley α-amylase/subtilisin inhibitor; BSZx, Barley Serpin Zx; DH, double-headed (Janus type); KTI3, Kunitz trypsin inhibitor 3; MCoTI-II, *Momordica cochinchinensis* trypsin inhibitor-II; MD, multidomain (pearls on a string type); NaProPI, *Nicotiana alata* protease inhibitor precursor; PI-II, potato peptidase inhibitor II; P, promiscuous inhibitors (mouse trap type); PLCPs, papain-like Cys proteases; PMC, potato multicystatin; PSPI, potato serine protease inhibitor; ND, not determined.[1]Rakwal *et al.* (2001); Qu *et al.* (2003). [2]Pekkarinen & Jones (2003). [3]Strobl *et al.* (1998). [4]Maskos *et al.* (1996). [5]Abdul-Hussain & Paulsen (1989); Vallée *et al.* (1998); Nielsen *et al.* (2003). [6]Hwang *et al.* (2009). [7]Da Silva *et al.* (2014). [9]Siqueira-Júnior *et al.* (2002); Uppalapati *et al.* (2005); Girard *et al.* (2007). [10]Orr *et al.* (1994); Siqueira-Júnior *et al.* (2002). [11]Orr *et al.* (1994); Siqueira-Júnior *et al.* (2002). [12]Mignery *et al.* (1988); Pouvreau *et al.* (2001); Weeda *et al.* (2009). [13]Graham *et al.* (1985); Kong & Ranganathan (2008). [14]Tamhane *et al.* (2009); Joshi *et al.* (2014). [15]Johnson *et al.* (1989); Tamhane *et al.* (2009); Dunse *et al.* (2010); Joshi *et al.* (2014). [16]Fluhr *et al.* (2012).

with different inhibitory specificity, and thus the Janus-type inhibitors include representatives from at least four MEROPS families that occur in plants (Table 1).

Bowman–Birk inhibitors (I12): losing and gaining multifunctionality The first Janus-type inhibitor was discovered in soybean flour by Bowman in 1946 (Bowman, 1946). The inhibitor was further purified and characterized by Yehudith Birk, who showed that it inhibits both trypsin and chymotrypsin (Birk, 1961). However, it was not yet clear that this was attributable to two binding sites. Separation of the protein into two fragments which retained inhibitory activity towards either trypsin or chymotrypsin, respectively, elucidated that the soybean Bowman–Birk inhibitor (BBI, MEROPS family I12; see Box 1 for a brief description of the MEROPS system) has two independent binding sites (Odani & Ikenaka, 1973), making the soybean BBI the first known multifunctional protease inhibitor. BBI encoding genes are also present in *Medicago truncatula*, rice (*Oryza sativa*) and maize (*Zea mays*), but they seem absent in the model plants *Arabidopsis thaliana* and *Nicotiana benthamiana* (Rawlings *et al.*, 2014). BBIs may have a defensive function, because *BBI* gene expression in rice is up-regulated in response to wounding and the defence-related phytohormone jasmonic acid (Rakwal *et al.*, 2001; Qu *et al.*, 2003) and overexpression of an endogenous Janus-type BBI in the staple crop rice increases resistance to *Magnaporthe grisea*, a fungal pathogen causing rice blast (Qu *et al.*, 2003). BBIs fold into a core of antiparallel beta-sheets that is crosslinked by multiple disulphide bridges (seven in the case of soybean BBI). The two inhibitory sites are located on protruding loops on opposing ends of the beta-sheet core (Fig. 2a) and function via the Laskowski mechanism (Box 2) (Chen *et al.*, 1992; Voss *et al.*, 1996). Monocot I12 inhibitors contain up to three BBI domains (Qu *et al.*, 2003), and thus they should in theory have up to six protease inhibitory sites. However, the monocot BBI domain has lost a disulphide bridge restraining the conformation of one of the inhibitory loops. Having lost the inhibitory activity of one of the two subdomains, monocot BBIs bind only one target protease per BBI domain. Apparently, the loss of the second inhibitory site in monocots was corrected by domain duplication, so that contemporary monocot BBIs can bind multiple proteases (Song *et al.*, 1999; Park *et al.*, 2004). Surprisingly, although BBIs can inhibit trypsin and chymotrypsin, it is still unknown what the natural targets of these seed proteins are. Identification of both endogenous (plant) and exogenous (insect/bacterial/fungal) target proteases might increase our understanding of the apparent evolutionary pressure towards multifunctional plant BBIs.

Kunitz inhibitors (I3): a very versatile fold Ser protease inhibitors from the Kunitz (I3) family exist in most higher plants, but seem to be absent from green alga genome sequences according to the current release of the MEROPS database (Rawlings *et al.*, 2014). The inhibitors were named after Moses Kunitz, who crystallized the first representative from soybean flour (Kunitz, 1945). The I3 family includes Janus-type inhibitors that bind different target enzymes on two reactive sites, for example the barley α-amylase/subtilase inhibitor BASI and its rice orthologue, both of which occur in grains (Leah & Mundy, 1989; Yamagata *et al.*, 1998). α-amylase/subtilase inhibitors are believed to regulate germination, as they inhibit endogenous α-amylases (Abdul-Hussain & Paulsen, 1989; Vallée *et al.*, 1998; Nielsen *et al.*, 2003), which mobilize storage carbohydrates during germination (Fincher, 1989). Rice α-amylase/subtilase inhibitors also block α-amylases from insects that feed on the grain starch (Bellincampi *et al.*, 2004). Using its second reactive site, BASI inhibits proteases from the fungal pathogen *Fusarium culmorum* (Pekkarinen *et al.*, 2007), also indicating defensive functions that make α-amylase/subtilase inhibitors interesting candidates for crop improvement. Another member of the Kunitz family, *Adenanthera pavonina* Kunitz type inhibitor (ApKTI), occurs in seeds of the leguminous tree *Adenanthera pavonina*. ApKTI can inhibit both trypsin (S1) and papain (C1) simultaneously (Migliolo *et al.*, 2010) and is active against gut proteases from herbivorous insects, including beetles and moths (Da Silva *et al.*, 2014 and references therein). Artificial diets containing ApKTI reduce the viability and fertility of these insects, indicating that ApKTI acts in defence against herbivorous insects. ApKTI is a promising candidate for crop improvement, as the multifunctionality of the inhibitor may help to impede insect adaptation (Da Silva *et al.*, 2014). Potato (*Solanum tuberosum* L.) tubers also contain a Janus-type Kunitz type inhibitor, named potato serine protease inhibitor (PSPI). PSPI can bind simultaneously to both trypsin and chymotrypsin, two S1 Ser proteases with different substrate specificities (Valueva *et al.*, 2000; Meulenbroek *et al.*, 2012). The biological role of PSPI *in planta* is unclear. In principle, PSPI could act in defence or protect storage proteins from endogenous proteases to prevent premature sprouting. Considering that PSPI is one of the most abundant proteins in

Fig. 2 Four examples of double-headed inhibitors. (a) Left panel, crystal structure (PDB ID 3ru4) of *Vigna unguiculata* Bowman–Birk trypsin and chymotrypsin inhibitor (BTCI, family I12; blue) in complex with bovine trypsin (family S1; red) and bovine chymotrypsin (family S1; green). Right panel, BTCI inhibits trypsin and chymotrypsin using loops (red) that contain Lys and Ile residues mimicking the P1 substrate recognition sites of the respective target enzymes. (b) Left panel, crystal structures of *Hordeum vulgare* (barley) α-amylase/subtilisin inhibitor (BASI, family I3; red; PDB ID 3bx1) in complex with barley α-amylase (AMY2; brown; PDB ID 1ava) and *Bacillus lentus* savinase (blue; family S8; PDB ID 3bx1). A calcium ion (green) acts as molecular glue in the AMY2–BASI interaction. Right panel, the *Glycine max* (soybean) Kunitz inhibitor (SKI, family I3; PDB ID 1avu) illustrates the overall structure of Kunitz inhibitors: a β-barrel (red) with three extended loops (L1–L3), representing a tree trunk and branches, respectively. (c) Left panel, crystal structure of ragi bifunctional inhibitor (RBI, family I6; PDB ID 1tmq; green) from *Eleusine coracana* (ragi or Indian finger millet) with *Tenebrio molitor* (yellow meal worm) α-amylase (TMA; brown; PDB ID 1tmq). The location where a trypsin-like protease could bind is indicated by a red circle. Right panel, inhibition is facilitated by the N-terminal Ser residue of RBI (red) that interacts with the active site of TMA. The loop that can interact with trypsin-like proteases (family S1) contains an Arg residue (red) that mimics the P1 substrate recognition motif for trypsin. (d) Left: crystal structure (PDB ID 4n6o) of human cystatin M (family I25; purple) with human legumain (family C13; cyan). The location where a papain-like protease could bind is indicated by a green circle. Right panel, legumains specifically cleave after Asn (P1 = Asn), and therefore selectivity of the interaction is facilitated by an Asn residue in the reactive centre loop (RCL; red). The three regions of cystatin M that would bind the substrate-binding groove of papain-like Cys proteases are located on the opposing side (red). PDB: RCSB Protein Data Bank (http://www.rcsb.org/pdb).



Box 2 The Laskowski mechanism of protease inhibition.

The Laskowski mechanism of inhibition is probably the most common scenario of protease inhibition by proteinaceous inhibitors (Rawlings *et al.*, 2014). Michael Laskowski described the 'standard mechanism' of protease inhibition, where the inhibitor acts as a 'limited proteolysis substrate' (reviewed in Laskowski & Kato, 1980). A reactive peptide bond on this limited substrate is bound by the target protease and an acyl intermediate is formed with a high association constant. However, the rate of completion of proteolytic cleavage and dissociation is very low, resulting in an apparent equilibrium between the free enzyme and inhibitor on the one hand and the complex on the other. Both the intact and the cleaved inhibitor can bind and inhibit the protease, and cleavage is reversible.

potato tubers (Meulenbroek *et al.*, 2012), it probably also serves as a storage protein itself. Recently, a new biochemical function was proposed for Kunitz trypsin inhibitor 3 (KTI3), a *Populus deltoides* (poplar) inhibitor from the Kunitz family active against trypsin and chymotrypsin (Major & Constabel, 2008). KTI3 is expressed *in planta* upon exposure to heavy metals and this protein confers heavy metal resistance when expressed in transgenic yeast. Molecular modelling suggests that KTI3 can chelate copper ions, but this remains to be demonstrated experimentally (Guerra *et al.*, 2015). Multifunctionality in Kunitz inhibitors is not always achieved via the double-headed Janus structure. For instance, a Kunitz inhibitor from *Prosopis juliflora*, a South American shrub, can inhibit either trypsin (family S1) or papain (family C1) using overlapping binding sites (Franco *et al.*, 2002).

All Kunitz (I3) inhibitors share a tree-like tertiary structure called the β-trefoil fold (Fig. 2b). The 'tree' consists of a β-barrel (the trunk) with flexible loops protruding from each side (branches and roots, respectively) (Sweet *et al.*, 1974). The protease binding sites of β-trefoil inhibitors are located in the loops (Azarkan *et al.*, 2011) and inhibition occurs via the Laskowski mechanism (Box 2) (Renko *et al.*, 2012), with the notable exception of the α-amylase/subtilase inhibitors.

BASI has been crystallized in complex with savinase, a subtilase from the soil bacterium *Bacillus lentus* (Micheelsen *et al.*, 2008). Structural and mutational analysis of this complex revealed that the BASI inhibitory mechanism for subtilases differs from the canonical Laskowski mechanism. The protease inhibitory loop of BASI is shorter than usual and pulled out of the protease active site by a disulphide bridge. Thus, BASI cannot be cleaved by the subtilase, which may result in a more stable inhibitory complex. The cysteine residues facilitating this version of the Laskowski mechanism are conserved in the rice and wheat (*Triticum aestivum*) orthologues of BASI, suggesting that they function in the same way (Micheelsen *et al.*, 2008). The inhibitory sites of BASI are known to be independent because protease inhibition still occurs when the inhibitor is saturated with α-amylase (Mundy *et al.*, 1983). Inhibition of barley α-amylase occurs via a large binding interface and involves a fully hydrated Ca^{2+} ion. In this way, BASI sterically hinders access to the active site of the α-amylase (Fig. 2b) (Vallée *et al.*, 1998). As affinity of BASI for barley α-amylase increases with increasing pH (Nielsen *et al.*, 2003 and references therein), it is thought that the protonation state of the amino acid side chains involved in Ca^{2+} binding may be affected by pH changes (Nielsen *et al.*, 2003). During germination, a pH increase might release α-amylase from BASI (Mundy *et al.*, 1983; Vallée *et al.*, 1998; Nielsen *et al.*, 2003).

Length, orientation and amino acid composition of the loops vary between the β-trefoil inhibitors, allowing them to inhibit up to two proteases of different classes (e.g. Dattagupta *et al.*, 1999; Meulenbroek *et al.*, 2012). Protease families inhibited by Kunitz inhibitors include S1, S8, C1 and A1 (Azarkan *et al.*, 2011; Rawlings *et al.*, 2014). Although the overall structure is highly conserved, the inhibitory motifs of the Kunitz inhibitors are diversified across the family. The core β-barrel, the conserved element of the Kunitz fold, is a rigid structure that presumably confers stability to the Kunitz inhibitors in harsh environments. The β-barrel can fold via various routes, rendering the folding process of Kunitz proteins relatively immune to point mutations in the loops. Taking these findings together, it appears that the stable β-trefoil fold provided a platform for evolution of a wide range of molecular recognition mechanisms (Azarkan *et al.*, 2011). In plants, this includes the inhibition of amylases and proteases via various mechanisms. In other organisms, the Kunitz fold with its versatile loops is shared by growth factors, interleukins and DNA-binding proteins (Renko *et al.*, 2012).

RBI (I6): bifunctionality via two versions of substrate mimicry Ragi bifunctional inhibitor (RBI) was first purified from seeds of the Indian finger millet, locally called ragi (*Eleusine coracana*) (Shivaraj & Pattabiraman, 1981). Like BASI, RBI inhibits both insect and mammalian α-amylases as well as proteases (Maskos *et al.*, 1996; Strobl *et al.*, 1998). Astonishingly, BASI and RBI are unrelated in terms of sequence, structure and inhibitory mechanisms, and thus inhibition of different hydrolase classes using one multifunctional protein must have evolved twice independently. RBI adopts an all-α-fold (Fig. 2c), which is unrelated to the structure of any other known protease inhibitor family (Strobl *et al.*, 1995, 1998). The trypsin-binding site of RBI forms a loop that can be cleaved by bovine trypsin, but RBI remains inhibitory afterwards (Strobl *et al.*, 1995), confirming that protease inhibition follows the canonical Laskowski mechanism (Box 2) (Maskos *et al.*, 1996). Structural analysis of a complex between RBI and α-amylase from yellow meal worm (*Tenebrio molitor*) revealed that the amylase inhibitory mechanism differs remarkably between I3 and I6 inhibitors. RBI inserts its N-terminus into the amylase active site, almost completely filling the catalytic cleft and directly interacting with the active site residues (Strobl *et al.*, 1998). The inhibitor can be displaced by large substrates (more than seven saccharide units), confirming that RBI acts in a competitive, substrate-like manner on α-amylase (Maskos *et al.*, 1996). RBI can form a ternary complex with bovine trypsin and porcine α-amylase (Maskos *et al.*, 1996), confirming that the binding sites for the protease and amylase are independent, as is the case for BASI. The extensive *in vitro* studies on the RBI inhibitory mechanisms have, to our knowledge, not been complemented by studies on the role of this bifunctional inhibitor *in vivo*. For instance, it is unknown

whether RBI inhibits endogenous α-amylases as well as pathogen enzymes and how important this is for germination and defence. The structural similarity between RBI and seed storage proteins, the albumins, provokes the speculation that this Janus-type inhibitor evolved from the small, stable albumin fold by acquiring additional inhibitory activities.

Legumain-inhibiting cystatins (I25A): more than just an extension Plant representatives of the cystatin family (I25) are called phytocystatins (MEROPS family I25A; see Box 1 for a brief description of the MEROPS system) and occur in a wide range of taxonomic groups, including green algae, mosses, monocots and dicots (Martinez & Diaz, 2008). Most phytocystatins have a molecular weight of 12–16 kDa and inhibit Cys proteases of the family C1. Higher plants contain additional C-terminally extended cystatins with a molecular weight of *c.* 24 kDa (Martinez & Diaz, 2008). The C-terminal extension is a second inhibitory site, specific for C13 proteases, called aspariginyl endopeptidases (AEPs), legumains or vacuolar processing enzymes (VPEs) (Martinez *et al.*, 2007). The presence of both C1 and C13 proteases is necessary *in vitro* for complete degradation of the bean (*Phaseolus vulgaris* L.) storage protein phaseolin (Zakharov *et al.*, 2004). To prevent germination, C1 and C13 proteases may thus be co-regulated by legumain-inhibiting cystatins. Indeed, overexpression of the C-terminally extended cystatin AtCYS6 in *A. thaliana* causes reduced Cys protease activity and delayed germination, whereas *atcys6* null mutant seeds exhibit higher Cys protease activity and germinate early (Hwang *et al.*, 2009). It remains unclear, however, whether this requires the C1 or the C13 protease inhibitory activity, or both. Bifunctional cystatins may be efficient regulators of germination, as they are able to prevent both the C1- and the C13-dependent steps of storage protein hydrolysis.

Bifunctional cystatins target different proteases via different mechanisms, namely steric hindrance of C1 proteases and substrate mimicry for C13 proteases. The mechanism for C1 protease inhibition by cystatins is conserved between animals and plants (Rawlings *et al.*, 2014) and is based on a tripartite wedge consisting of the N-terminus and two hairpin loops carrying the conserved QxVxG motif (Fig. 2d). This wedge is inserted into the active site of the C1 protease in a tight but reversible interaction (Stubbs *et al.*, 1990). C13 protease inhibition by phytocystatins involves substrate mimicry using an exposed Asn residue following an α-helix on the C-terminal extension. Interestingly, animal cystatins can inhibit legumains (AEPs/VPEs) without an extension, as they harbour such an Asn residue next to a helix on the back of the cystatin domain itself (Alvarez-Fernandez *et al.*, 1999). Bifunctional phytocystatins may thus have evolved from an ancient domain duplication, upon which each of the two cystatin domains lost one of the inhibitory activities (Martinez *et al.*, 2007).

2. Multidomain Inhibitors: the pearl necklace

The pearl necklace type inhibitors consist of an array of protease inhibitory domains. Their inhibitory potential is increased through cleavage, which releases the inhibitory domains like pearls from a string. Gene duplication probably gave rise to these multidomain proteins. Notably, a new folding pathway led to reorganization of the inhibitory domains of I20 inhibitors. Cyclization of the multidomain precursor or the released inhibitory peptides occurs for I20 and I7 inhibitors, respectively. The accumulation of structurally similar domains in one protein can also facilitate new regulatory mechanisms, as seen with the multicystatins (I25).

Multidomain potato peptidase inhibitor II (I20): cyclization shuffles domains As early as the 1970s, the Japanese scientist Teruo Iwasaki determined the amino acid sequence of a fragment of potato peptidase inhibitor II (PI-II) that inhibited trypsin (S1) and, to a lesser extent, chymotrypsin (S1) and subtilisin (S8) (Iwasaki *et al.*, 1976). Similar proteins occur in tomato (*Solanum esculentum* L.), where their production is induced by wounding (Graham *et al.*, 1985). Today, it is clear that inhibitors of the I20 family are present in most monocots and dicots including maize, *A. thaliana*, poplar and many solanaceous species. Most plants appear to have only one I20 inhibitor gene, whereas tomato and potato contain more homologues (Rawlings *et al.*, 2014). The I20 inhibitors are clearly associated with plant defence. Their expression increases upon wounding (Graham *et al.*, 1985; Kong & Ranganathan, 2008) and they are constitutively made in reproductive organs, which seem well worth special protection (Atkinson *et al.*, 1993). Furthermore, I20 inhibitors from pepper (*Capsicum annuum*) inhibit the gut proteases of the cotton bollworm (*Helicoverpa armigera*) *in vitro* (Tamhane *et al.*, 2005; Joshi *et al.*, 2014) and overexpression of different members of the I20 family confers increased resistance to insect pests in various species (Johnson *et al.*, 1989; Tamhane *et al.*, 2009; Dunse *et al.*, 2010; Joshi *et al.*, 2014), highlighting the potential of I20 inhibitors for crop improvement.

Structure and fold of the I20 inhibitors are prime examples of the fascinating intricacy of plant defence proteins. The *Nicotiana alata* protease inhibitor precursor (NaProPI) is a well-studied example (Kong & Ranganathan, 2008). NaProPI is a 43-kDa protein produced at high levels in the female reproductive tissues of *N. alata*. The amino acid sequence of NaProPI consists of six homologous sequence repeats flanked by an N-terminal signal peptide and a C-terminal vacuolar targeting signal (Atkinson *et al.*, 1993). Both signal peptides are removed during posttranslational processing and, eventually, the six-repeat precursor is cleaved into six individual protease inhibitors of 6 kDa each (Fig. 3a). Two of the inhibitors released from NaProPI inhibit chymotrypsin and four are specific for trypsin (Lee *et al.*, 1999). Each of the released inhibitors contains eight Cys residues forming four disulphide bonds and adopts a compact, stable fold (Nielsen *et al.*, 1995) (Fig. 3a). All I20 inhibitors utilize the Laskowski mechanism (Box 2) (Greenblatt *et al.*, 1989), but their *in vivo* target proteases remain unknown. Five of the 6-kDa inhibitors released from NaProPI are single-chain peptides, but the sixth one consists of two chains held together by three disulphide bridges (Lee *et al.*, 1999). This peculiarity arises from the way in which NaProPI is processed into the 6-kDa mature inhibitors. Interestingly, the underlying cleavage sites are located not between, but within each of the sequence repeats of NaProPI (Heath *et al.*, 1995). After removal of the signal peptides, the NaProPI precursor adopts a cyclic conformation, linking the N-terminus to the C-terminus of the

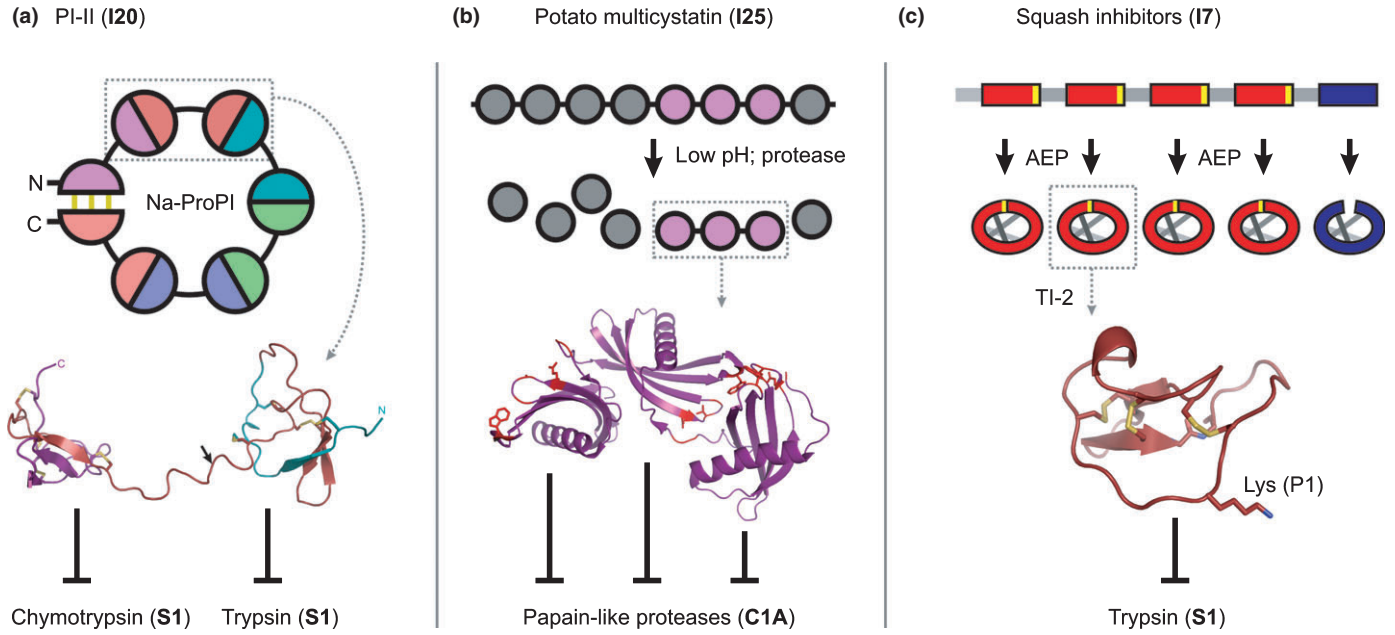


Fig. 3 Folding and processing of multidomain inhibitors. (a) Folding and processing of the PI-II (peptidase inhibitor II) inhibitor of *Nicotiana alata* (Na-ProPI, family I20). Na-ProPI contains six sequence repeats (highlighted in different colours) that fold into six hybrid inhibitory domains (depicted as pearls). One domain consists of the N- and C-termini linked by three disulphide bridges. The inhibitory domains are released through proteolytic cleavage by Cys proteases (presumably family C13). The nuclear magnetic resonance (NMR) structure of the second and third domains with the cleavable linker between them has been resolved (bottom; PDB ID 1fyb). The cleavage site for release of the two domains, which target chymotrypsin and trypsin (both family S1), respectively, is indicated by a black arrow. (b) Folding and processing of potato multicystatin (PMC, family I25). The precursor consists of eight cystatin domains (depicted as pearls) that are released through proteolytic cleavage by Ser proteases at low pH. Domains 5–7 remain together and have been crystallized (PDB ID 4lzi). All released cystatin domains target papain-like Cys proteases (family C1) using a tripartite wedge inhibitory site (highlighted in red). (c) Folding and processing of squash inhibitors (family I7). The precursor consists of five sequence repeats, each but the last ending in the asparaginyl endopeptidase (AEP)/vacuolar processing enzyme (VPE) cleavage site Asp or Asn, indicated in yellow. Four of the released peptides (red) undergo cyclization by AEP/VPE, while the last repeat (blue) remains acyclic. All five released squash inhibitors share the cystine knot motif with three disulphide bridges (grey). The crystal structure of the macrocyclic knottin MCoTI-II (*Momordica cochinchinensis* trypsin inhibitor-II; PDB ID 4gux) shows the disulphide bridges linking the central loop to the periphery (yellow) and the Lys residue recognized by trypsin (family S1).

peptide chain via three disulphide bridges. This circle is then cleaved once within each sequence repeat, releasing five single-chain inhibitors as well as one two-chain protein which contains the N- and C-terminal ends of the NaProPI precursor (Lee *et al.*, 1999). NaProPI can be processed *in vitro* using C13 endopeptidases (Heath *et al.*, 1995), suggesting that asparaginyl endopeptidases (AEPs/VPEs/legumains, family C13) might release the 6-kDa inhibitors *in vivo*.

The peculiar mechanism of NaProPI maturation prompts the question of how this protein might have evolved. It is very likely that the ancestor of the I20 family consisted of a single sequence repeat folding into a single domain inhibitor. Sequence duplication events then gave rise to multi-repeat proteins that eventually adopted the cyclic fold and processing sites within each sequence repeat. This scenario is endorsed by the finding that a single ancestral NaProPI sequence repeat can fold into a functional inhibitor when it is expressed in bacteria (Scanlon *et al.*, 1999). As soon as more than one repeat is present, however, the artificial ancestral I20 inhibitor adopts a cyclic conformation and is processed within the repeats (Lee *et al.*, 1999). Notably, not all of the I20 inhibitors need to be cleaved in order to be active. The two-domain tomato inhibitor II has been crystallized in a ternary complex with two subtilisin molecules, showing that both inhibitory sites are accessible simultaneously (Barrette-Ng *et al.*, 2003). Analysis of nonsynonymous/synonymous mutation rates within the whole I20 family revealed that that the cysteine scaffold that determines the structure of the cyclic precursors as well as of the released inhibitors is under purifying selection. This underscores the evolutionary pressure towards a cyclic precursor that may be explained by its elevated thermodynamic stability. The active residue of the released inhibitors, however, is under diversifying selection, presumably leading to specificity for different target proteases (Kong & Ranganathan, 2008).

Squash inhibitors (I7): knottins laced up Squash inhibitors (MEROPS family I7) occur in the seeds of Cucurbitaceae and are macrocyclic knottins of < 50 amino acids, stabilized by multiple disulphide bridges in the cystine knot motif. The squash inhibitors we discuss here are released from multidomain precursors containing up to eight repeats (Hernandez *et al.*, 2000; Mylne *et al.*, 2012). Given their abundance and strong inhibitory activity against trypsin-like proteases used by mammalian, insect and fungal seed predators, squash inhibitors might be involved in defending seeds against predation (Burman *et al.*, 2014; Mahatmanto, 2015), but a role as storage proteins in seeds has also been envisaged (Mahatmanto *et al.*, 2015). To our knowledge, no protease has yet been identified as a natural target of squash inhibitors.

The cystine knot is a common motif that stabilizes squash inhibitors as well as a range of other, unrelated plant peptides known collectively as the cyclotides. The knot is based on a hairpin of two antiparallel β-strands containing three cysteines linked to the periphery via three disulphide bridges. A third, nonstandard β-strand is present in one of the loops, which are otherwise not canonically structured (Fig. 3c) (Saether *et al.*, 1995; Craik *et al.*, 1999; Mylne *et al.*, 2012). Squash inhibitors are released from multidomain precursors and cyclized via transpeptidation by asparaginyl endopeptidases, also called vacuolar processing enzymes (AEPs/VPEs, family C13) (Saska *et al.*, 2007; Gillon *et al.*, 2008; Mylne *et al.*, 2012). The last repeat of the inhibitor precursor remains acyclic, but nevertheless inhibits trypsin (Mylne *et al.*, 2012). The squash inhibitors are members of the I7 family that shares the Laskowski mechanism of inhibition (Box 2) (Otlewski & Zbyryt, 1994; Hernandez *et al.*, 2000).

Interestingly, macrocyclization of cystine knot peptides seems to have evolved several times in parallel. The whole subtropical cucurbit genus *Momordica* has an acyclic, single-domain cystine knot inhibitor, while multidomain precursors encoding macrocyclic knottins are only found in a subgroup of related species, indicating they may have arisen through recent gene duplication events (Mahatmanto *et al.*, 2015). Additionally, cyclotides occur in the Rubiaceae, Violaceae, Fabaceae and Solanaceae. The corresponding precursors differ between species with regard to the number of cystine knot repeats (Mylne *et al.*, 2012). All of these precursors share the recognition sites necessary for cleavage and cyclization by AEPs/VPEs (Hara-Nishimura *et al.*, 1991; Hiraiwa *et al.*, 1999) and a Gly residue needed to form a circle (Jennings *et al.*, 2001). The ability of AEPs/VPEs to cyclize peptides may have acted as an evolutionary channel, lending the selective advantage of increased stability to cystine knot peptides that start with a Gly and end in Asp/Asn (Mylne *et al.*, 2012).

Macrocyclic knottins are an excellent template for artificial multifunctionalization. Different engineered MCoTI-II (*Momordica cochinchinensis* trypsin inhibitor-II) knottins specifically inhibit proteases from different catalytic classes, namely chymotrypsin (family S1, Ki in nM range), subtilisin (family S8, Ki in μM range), a viral Cys protease (family C3, Ki in in μM range) (Thongyoo *et al.*, 2008), tryptase and leukocyte elastase (both family S1) (Thongyoo *et al.*, 2009).

Multicystatins (I25A): release at the right moment Phytocystatins form a separate subfamily I25A within the cystatin family I25 (Margis *et al.*, 1998; Rawlings *et al.*, 2014). Phytocystatins inhibit papain-like Cys proteases (family C1) and are common among plants from green algae to *A. thaliana*, tomato, rice and maize (Rawlings *et al.*, 2014). Some members of the phytocystatin family also inhibit C13 proteases using a C-terminal extension, as discussed in subsection 1.4. In this paragraph, we focus on the numerous cases where phytocystatins occur as multidomain proteins (Rawlings *et al.*, 2014) consisting of up to eight inhibitory cystatin domains (Walsh & Strickland, 1993). Multidomain cystatins (multicystatins) play various biological roles in both potato and tomato. The potato multicystatin (PMC) occurs in potato tubers and inhibits tuber proteases (Kumar *et al.*, 1999). PMC is produced during early stages of tuber formation, which are associated with a decrease in proteolytic activity in the tuber, which may facilitate storage protein accumulation. Indeed, accumulation of PMC to up to 12% of total soluble tuber protein precedes the accumulation of patatin, the main storage protein in potatoes that makes up 40% of the soluble protein content (Mignery *et al.*, 1988; Pouvreau *et al.*, 2001; Weeda *et al.*, 2009). In ageing tubers, proteolytic activity increases again, coinciding with a drop in the detectable levels of PMC (Kumar *et al.*, 1999). Thus, PMC abundance appears to regulate the process from building up the protein reserves in potato tubers to their mobilization during ageing and germination. Multicystatin expression in tomato leaves is induced by various elicitors of plant immunity (Siqueira-Júnior *et al.*, 2002; Uppalapati *et al.*, 2005; Girard *et al.*, 2007) and tomato multicystatin inhibits proteases from insect digestive tracts and impairs the growth of plant pathogenic fungi *in vitro* (Siqueira-Júnior *et al.*, 2002). Similar experiments show that potato multicystatin inhibits the growth of corn rootworm larvae, predators of an important food crop, when added to an artificial diet (Orr *et al.*, 1994).

Multicystatins have the intriguing capacity to crystallize in native tissues. Cys protease inhibiting crystals were first observed in potato tubers (Cohn, 1859; Rodis & Hoff, 1984), but they also occur in tomato leaves (Akers & Hoff, 1980). Multicystatin crystals localize to the cytosol in potato (Nissen *et al.*, 2009) as well as in tomato (Madureira *et al.*, 2006). The transition between crystalline and soluble states for PMC is now quite well understood. Solubilization of PMC occurs at mildly acidic pH, exposing the inhibitory domains to target cysteine proteases (Orr *et al.*, 1994). Recent structural analyses show that low pH weakens the interdomain interactions in PMC, explaining this regulatory mechanism (Green *et al.*, 2013). As soon as it is soluble, PMC can be cleaved by serine proteases (Walsh & Strickland, 1993) to release three fragments, which collectively contain eight cystatin domains. Inhibition of Cys proteases by each of the eight domains occurs via the cystatin mechanism (subsection 1.4). The acidic environment of certain areas in the insect gut is thought to activate the inhibitor exactly when and where it is needed to hinder the protein catabolism of the insect pathogen, but not the host plant (Green *et al.*, 2013). Multicystatins presumably evolved from an ancestral, single-domain plant cystatin via gene duplications. Although it is known that phytocystatins inhibit papain-like cysteine proteases (family C1) (Tajima *et al.*, 2011), it is as yet unknown which proteases are regulated by multicystatins (Benchabane *et al.*, 2010).

3. Promiscuous inhibitory folds: the mouse trap type

In this section, we describe two protease inhibitor families that utilize a mechanical trapping mechanism to sequester their target proteases. The nature of these mechanisms is destructive: inhibition is irreversible and dooms both the protease and the inhibitor to degradation. The active site that is recognized by the enzyme can vary in mouse trap type inhibitors without affecting the functionality of the trap. Accordingly, mouse trap type inhibitors have developed multifunctionality in the sense that the same inhibitory fold can be used to target different proteases.

Serpins (I4): springing the mouse trap on Ser and Cys proteases Ser and Cys protease inhibitors of the MEROPS family I4, called serpins, occur in the grains of cereals of the Triticeae tribe, including barley, rye (*Secale cereale* L.), and wheat, and in oat (*Avena sativa* L., Poeae tribe) (Roberts *et al.*, 2003). As highly abundant grain proteins, serpins are found even in fully processed beer (Hejgaard & Kaersgaard, 1983; Hejgaard *et al.*, 1985). The biological role(s) of serpins remains somewhat unclear, although there are hints in two directions. First, artificial diets containing serpins impair growth and fertility of insect pests, indicating a potential role in defence and highlighting the potential serpins hold for crop improvement (Thomas *et al.*, 1994, 1995; Yoo *et al.*, 2000; Alvarez-Alfageme *et al.*, 2011). Second, two putative plant target proteases of *A. thaliana* serpin 1 (AtSerpin1) are involved in programmed cell death (PCD), suggesting that AtSerpin1 could have a prosurvival function. AtSerpin1 inhibits metacaspase AtMC9 (family C14) *in vitro* and co-localizes with AtMC9 *in vivo* in the extracellular space (Vercammen *et al.*, 2006). Electron microscopic images of *atmc9* mutant *A. thaliana* leaf cells suggest a role for AtMC9 in clearance of the cell contents after tonoplast rupture (Bollhöner *et al.*, 2013). The second putative AtSerpin1 target protease is the C1 protease Responsive to Desiccation 21 (RD21), which was identified as an *in vivo* interaction partner during pull-down of AtSerpin1 from plant extracts (Lampl *et al.*, 2010). The RD21 precursor protein accumulates in endoplasmic reticulum-derived protease storage bodies, which fuse with each other and the vacuole under stress conditions. Fusion is believed to lead to activation of the proteases which assist in recycling of cellular contents during stress-induced PCD (Hayashi *et al.*, 2001). In line with this role of RD21, PCD during plant infection with the necrotrophic fungi *Botrytis cinerea* and *Sclerotina sclerotiorum* is accelerated in the *atserpin1* knockout, but prevented in plants lacking *RD21*. Interestingly, Lampl *et al.* (2013) detected AtSerpin1-GFP fusions expressed from the endogenous promoter mainly in the cytoplasm, in contrast to the previously reported extracellular localization (Vercammen *et al.*, 2006). In addition, increased resistance of *rd21* mutant leaves to infection by *Botrytis cinerea* is in contrast to earlier findings that *rd21* mutant plants are more susceptible than the wild type to *Botrytis cinerea* (Shindo *et al.*, 2012). Some serpins act as traps with a versatile bait, like the barley serpin BSZx. Using overlapping sites within its reactive centre loop, BSZx can inhibit trypsin, chymotrypsin and, to some extent, cathepsin G (all family S1) *in vitro* (Dahl *et al.*, 1996). Evolutionarily, inhibitory serpins are conserved throughout the kingdoms of life, although they vary widely between uni- and multicellular organisms (Roberts *et al.*, 2004). *Chlamydomonas* serpins, for instance, have an intron–exon structure distinct from that of higher plant serpins. The reactive centre loops differ remarkably between monocots and dicots, suggesting that serpins might have diverged with regard to target proteases and biological functions (Roberts & Hejgaard, 2007).

With regard to structure and mechanism, serpins are unique among plant protease inhibitors. The structure of the 43-kDa inhibitor AtSerpin1 has been resolved by crystallography and consists of three conserved β-sheets and nine conserved α-helices, as is common among animal serpins (Fig. 4a) (Lampl *et al.*, 2010). Serpins inhibit their target proteases in a unique, suicidal manner. Ser and Cys proteases cleave the serpin reactive loop, forming an acyl-enzyme intermediate with the serpin. Cleavage triggers a profound conformational change in the metastable fold of the serpin, which deforms the active site of the protease and irreversibly binds it to the inhibitor (Fig. 4b). This remarkable trapping mechanism (Huntington *et al.*, 2000) allows for a structural separation of inhibitory activity from protease specificity. Thus, members of the serpin family share the same fold, but target a range of Ser and Cys proteases (Fluhr *et al.*, 2012; Rawlings *et al.*, 2014).

α-macroglobulin (I39): the gilded cage Functional macroglobulin genes are annotated in only a few plant species, according to the MEROPS database (Rawlings *et al.*, 2014) and a recent comparative genomics study (Santamaría *et al.*, 2014). This includes cucumber (*Cucumis sativus*), alpine strawberry (*Fragaria vesca*), the alga *Micromonas* sp. *RCC299* and black cottonwood (*Populus trichocarpa*). However, the structure, specificity and mechanism of action of this family are intriguing. Macroglobulins are large (*c.* 200-kDa) glycoproteins and their structure resembles a round cage (Fig. 4c) (Sottrup-Jensen, 1989). α-macroglobulins possess an exposed bait region with recognition sites for various types of endopeptidases, cleavage of which triggers a conformational change. Thus, the protease gets trapped inside the large macroglobulin protein, much like in a cage (Fig. 4d) (Feldman *et al.*, 1985). The caged peptidase cannot bind large targets or inhibitors any more, but remains accessible for small molecules (Sottrup-Jensen, 1989). Variation of the bait region does not affect the inhibitory mechanism, allowing for multifunctionality of the macroglobulin fold (Sottrup-Jensen, 1989). The physiological role of macroglobulins in plants remains obscure. The versatile functions of this protease inhibitor family in animals and bacteria are reviewed elsewhere (Budd *et al.*, 2004; Rehman *et al.*, 2013).

III. Significance of multifunctional protease inhibitors in the plant research arena

Multifunctional protease inhibitors represent hubs that regulate distinct branches of the plant physiological network, for instance defence and tuber sprouting in the case of the potato multicystatin (family I25). For one inhibitor to do several biological jobs, it must often target multiple proteases of different families or even different catalytic classes. This can be achieved through several inhibitory interfaces on the same protein, as seen with the Janus-type and the multidomain I20 inhibitors, or through one promiscuous interface, as seen in the mouse trap type inhibitors. A different way to assign multiple biological functions to one inhibitor is to produce it on different occasions in space and time. For instance, multicystatins (family I25) accumulate in tomato and potato leaves in response to wounding as well as in potato tubers when building up storage protein reserves. Through whichever route inhibitors acquire multifunctionality, the result is a protein that provides a link between the biological processes it regulates. Increased knowledge of multifunctional protease inhibitors will therefore promote a network-level understanding of plant physiology. The evolutionary history of inhibitors in their

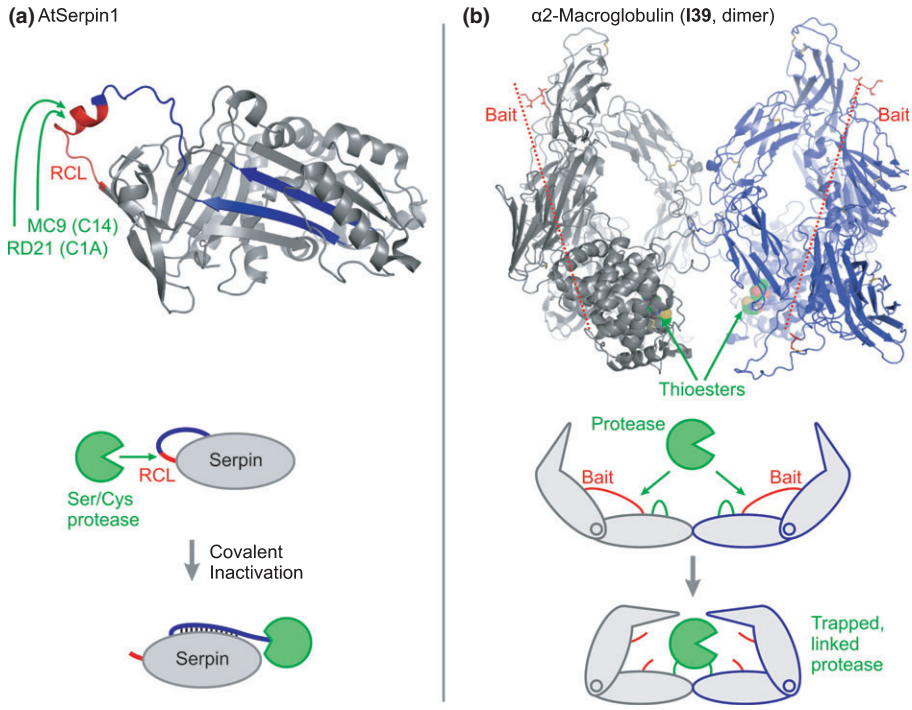


Fig. 4 Serpin and α2-macroglobulin, the promiscuous mouse traps. (a) Top pane, crystal structure of the serpin AtSerpin1 (family I4; PDB ID 3le2) showing the reactive centre loop (RCL; red) that can be cleaved by C1 papain-like Cys protease RD21 (family C1A) or metcaspase-9 (MC9, family C14). Cleavage causes a conformational change that inserts the loop (blue) as a new anti-parallel β-sheet between the two β-sheets (blue arrows). Bottom panel, schematic illustration of the mouse trap mechanism. Cleavage of the RCL results in a conformational change that insets the blue strand of the loop into the serpin structure, irreversibly deforming the serpin and the covalently trapped protease. (b) Upper panel, crystal structure of human α2-macroglobulin (family I39; PDB ID 4acq). Only two of the four subunits are shown, in grey and blue, respectively. The bait peptide structure is unresolved and indicated with a red dashed line. The Cys-Glu thioester (green) traps the protease covalently. Lower panel, simplified illustration of the inhibitory mechanism of α2-macroglobulin. Any small-size protease that cleaves the promiscuous bait peptide triggers a conformational change that traps the protease in a cage. The thioester (green loop) reacts with any Lys residue on the surface of the protease, thereby covalently immobilizing the protease in the cage.

role as regulatory hubs could then reveal how the network was restructured over time.

Multifunctional protease inhibitors share stabilizing structural features, most notably a compact fold linked covalently by disulphide bridges. This structural similarity may resemble a common evolutionary road to multifunctionality. It appears that small, stable proteins, such as protease inhibitors, are well suited to acquire a (second) inhibitory function, as they are already fit to persist in harsh environments with high proteolytic activity. Additional protease inhibitory activities may await discovery in many small, stable proteins, including the known multifunctional inhibitors. A second, even more obvious road towards multifunctionality is duplication of inhibitory domains, which can further lead to neofunctionalization of the duplicate, rearrangements in the overall protein structure or emergence of new regulatory mechanisms (subsections 1.1 and 2).

On the applied side, recombinant expression of multifunctional inhibitors from families I12, I13, I20 and I25 has been successfully used to generate pest-resistant crop plants (Orr *et al.*, 1994; Xu *et al.*, 1996; Siqueira-Júnior *et al.*, 2002; Dunse *et al.*, 2010). Furthermore, I3 and I20 inhibitors were used to limit proteolysis of recombinant human proteins produced in plants, tackling a major issue in molecular farming (Kim *et al.*, 2008; Goulet *et al.*, 2012). Increasing our understanding of how multifunctional inhibitors link physiological networks could facilitate new applications. For instance, recombinant protein degradation could be prevented very effectively by controlling regulators of proteolytic cascades in the plant. Inhibitors that interfere with endogenous as well as exogenous proteases could protect crops from pests and premature senescence. In medicine, multifunctional plant protease inhibitors of family I12 have been used to limit undesired protease activity as potential anticancer drugs (da Costa Souza *et al.*, 2014).

In one case, the concept of multidomain inhibitors has been taken further using artificial multidomain inhibitors consisting of five naturally occurring domains (families I25 and I31) to create a more stable and more potent inhibitor than its natural, single-domain counterparts. Overexpression of this custom-made multidomain inhibitor in potato increases resistance to the insect pest *Frankliniella occidentalis* (Outchkourov *et al.*, 2004). Some impressive examples of custom-made inhibitors based on a known multifunctional fold are found among the squash inhibitors (subsection 2.2). However, the relatively low affinity of the engineered squash inhibitors for their targets highlights the issue that a deeper understanding is needed to develop synthetic approaches. Knowing all essential features of an inhibitor structure, one could design completely novel inhibitors, customized for desired applications.

From the examples discussed throughout this review, it is clear that several roads to multifunctionality exist, starting from adaptation of target recognition sites in stable proteins or gene

duplication, with the latter branching out in multiple directions. Fine-mapping of these roads will facilitate the construction of custom multifunctional inhibitors while, at the same time, enhancing our knowledge about regulatory hubs in the plant physiological network.

Acknowledgements

We thank Thomas Roberts, Daniela Sueldo, Judith Paulus and Jan M. Brauner for critical reading. The authors are supported by the ERC Project 616449 'GreenProteases', University of Oxford and Somerville College, Oxford.

References

Abdul-Hussain S, Paulsen GM. 1989. Role of proteinaceous α-amylase enzyme inhibitors in preharvest sprouting of wheat grain. *Journal of Agricultural and Food Chemistry* **37**: 295–299.

Akers CP, Hoff JE. 1980. Simultaneous formation of chymopapain inhibitor activity and cubical crystals in tomato leaves. *Canadian Journal of Botany* **58**: 1000–1003.

Alvarez-Alfageme F, Maharramov J, Carrillo L, Vandenabeele S, Vercammen D, Van Breusegem F, Smagghe G. 2011. Potential use of a serpin from Arabidopsis for pest control. *PLoS ONE* **6**: e20278.

Alvarez-Fernandez M, Barrett AJ, Gerhartz B, Dando PM, Ni J, Abrahamson M. 1999. Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *Journal of Biological Chemistry* **274**: 19195–19203.

Atkinson AH, Heath RL, Simpson RJ, Clarke AE, Anderson MA. 1993. Proteinase inhibitors in *Nicotiana alata* stigmas are derived from a precursor protein which is processed into five homologous inhibitors. *Plant Cell* **5**: 203–213.

Azarkan M, Martinez-Rodriguez S, Buts L, Baeyens-Volant D, Garcia-Pino A. 2011. The plasticity of the β-trefoil fold constitutes an evolutionary platform for protease inhibition. *Journal of Biological Chemistry* **286**: 43726–43734.

Barrette-Ng IH, Ng KK-S, Cherney MM, Pearce G, Ryan CA, James MNG. 2003. Structural basis of inhibition revealed by a 1:2 complex of the two-headed tomato inhibitor-II and subtilisin Carlsberg. *Journal of Biological Chemistry* **278**: 24062–24071.

Beadle GW, Tatum EL. 1941. Genetic control of biochemical reactions in *Neurospora. Proceedings of the National Academy of Sciences, USA* **27**: 499–506.

Bellincampi D, Camardella L, Delcour JA, Desseaux V, D'Ovidio R, Durand A, Elliot G, Gebruers K, Giovane A, Juge N *et al.* 2004. Potential physiological role of plant glycosidase inhibitors. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1696**: 265–274.

Benchabane M, Schlüter U, Vorster J, Goulet M-C, Michaud D. 2010. Plant cystatins. *Biochimie* **92**: 1657–1666.

Birk Y. 1961. Purification and some properties of a highly active inhibitor of trypsin and α-chymotrypsin from soybeans. *Biochimica et Biophysica Acta* **54**: 378–381.

Bollhöner B, Zhang B, Stael S, Denancé N, Overmyer K, Goffner D, Van Breusegem F, Tuominen H. 2013. Post mortem function of AtMC9 in xylem vessel elements. *New Phytologist* **200**: 498–510.

Bowman DE. 1946. Differentiation of soybean antitryptic factors. *Proceedings of the Society for Experimental Biology and Medicine* **63**: 547–550.

Budd A, Blandin S, Levashina EA, Gibson TJ. 2004. Bacterial α2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biology* **5**: R38.

Burman R, Gunasekera S, Strömstedt AA, Göransson U. 2014. Chemistry and biology of cyclotides: circular plant peptides outside the box. *Journal of Natural Products* **77**: 724–736.

Chen P, Rose J, Love R, Wei CH, Wang B-C. 1992. Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors. *Journal of Biological Chemistry* **267**: 1990–1994.

Cohn F. 1859. Uber Proteinkristalle in der Kartoffel. *Jahres-Bericht der Schlesischen Gesellschaft für Vaterländische Cultur* **37**: 72–82.

da Costa Souza L, Camargo R, Demasi M, Santana JM, de Sá CM, de Freitas SM. 2014. Effects of an anticarcinogenic Bowman-Birk protease inhibitor on purified 20S proteasome and MCF-7 breast cancer cells. *PLoS ONE* **9**: e86600.

Craik DJ, Daly NL, Bond T, Waine C. 1999. Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *Journal of Molecular Biology* **294**: 1327–1336.

Da Silva DS, de Oliveira CFR, Parra JRP, Marangoni S, Macedo MLR. 2014. Short and long-term antinutritional effect of the trypsin inhibitor ApTI for biological control of sugarcane borer. *Journal of Insect Physiology* **61**: 1–7.

Dahl SW, Rasmussen SK, Hejgaard J. 1996. Heterologous expression of three plant serpins with distinct inhibitory specificities. *Journal of Biological Chemistry* **271**: 25083–25088.

Dattagupta JK, Podder A, Chakrabarti C, Sen U, Mukhopadhyay D, Dutta SK, Singh M. 1999. Refined crystal structure (2.3 Å) of a double-headed winged bean α-chymotrypsin inhibitor and location of its second reactive site. *Proteins* **35**: 321–331.

Dunse KM, Stevens JA, Lay FT, Gaspar YM, Heath RL, Anderson MA. 2010. Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. *Proceedings of the National Academy of Sciences, USA* **107**: 15011–15015.

Feldman SR, Gonias SL, Pizzo SV. 1985. Model of α2-macroglobulin structure and function. *Proceedings of the National Academy of Sciences, USA* **82**: 5700–5704.

Fincher GB. 1989. Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annual Review of Plant Physiology and Plant Molecular Biology* **40**: 305–346.

Fluhr R, Lampl N, Roberts TH. 2012. Serpin protease inhibitors in plant biology. *Physiologia Plantarum* **145**: 95–102.

Franco OL, Grossi de Sá MF, Sales MP, Mello LV, Oliveira AS, Rigden DJ. 2002. Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from *Prosopis juliflora. Proteins: Structure, Function, and Bioinformatics* **49**: 335–341.

Gillon AD, Saska I, Jennings CV, Guarino RF, Craik DJ, Anderson MA. 2008. Biosynthesis of circular proteins in plants. *Plant Journal* **53**: 505–515.

Girard C, Rivard D, Kiggundu A, Kunert K, Gleddie SC, Cloutier C, Michaud D. 2007. A multicomponent, elicitor-inducible cystatin complex in tomato, *Solanum lycopersicum. New Phytologist* **173**: 841–851.

Goulet C, Khalf M, Sainsbury F, D'Aoust M-A, Michaud D. 2012. A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant Biotechnology Journal* **10**: 83–94.

Graham JS, Pearce G, Merryweather J, Titani K, Ericsson LH, Ryan CA. 1985. Wound-induced proteinase inhibitors from tomato leaves. II. The cDNA-deduced primary structure of pre-inhibitor II. *Journal of Biological Chemistry* **260**: 6561–6564.

Green AR, Nissen MS, Kumar GNM, Knowles NR, Kang C. 2013. Characterization of *Solanum tuberosum* multicystatin and the significance of core domains. *Plant Cell* **25**: 5043–5052.

Greenblatt HM, Ryan CA, James MNG. 1989. Structure of the complex of *Streptomyces griseus* proteinase B and polypeptide chymotrypsin inhibitor-1 from Russet Burbank potato tubers at 2.1 Å resolution. *Journal of Molecular Biology* **205**: 201–228.

Guerra FP, Reyes L, Vergara-Jaque A, Campos-Hernández C, Gutiérrez A, Pérez-Díaz J, Pérez-Díaz R, Blaudez D, Ruíz-Lara S. 2015. *Populus deltoides* Kunitz trypsin inhibitor 3 confers metal tolerance and binds copper, revealing a new defensive role against heavy metal stress. *Environmental and Experimental Botany* **115**: 28–37.

Hara-Nishimura I, Inoue K, Nishimura M. 1991. A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. *FEBS Letters* **294**: 89–93.

Hayashi Y, Yamada K, Shimada T, Matsushima R, Nishizawa N, Nishimura M, 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**: 894–899.

Heath RL, Barton PA, Simpson RJ, Reid GE, Lim G, Anderson MA. 1995. Characterization of the protease processing sites in a multidomain proteinase inhibitor precursor from *Nicotiana alata. European Journal of Biochemistry* **230**: 250–257.

Hedstrom L, Szilagyi L, Rutter WJ. 1992. Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**: 1249–1253.

Hejgaard J, Kaersgaard P. 1983. Purification and properties of the major antigenic beer protein of barley origin. *Journal of the Institute of Brewing* **89**: 402–410.

Hejgaard J, Rasmussen SK, Brandt A, Svendsen I. 1985. Sequence homology between barley endosperm protein Z and protease inhibitors of the α1-antitrypsin family. *FEBS Letters* **180**: 89–94.

Hernandez J-F, Gagnon J, Chiche L, Nguyen TM, Andrieu J-P, Heitz A, Trinh Hong T, Pham TTC, Le Nguyen D. 2000. Squash trypsin inhibitors from *Momordica cochinchinensis* exhibit an atypical macrocyclic structure. *Biochemistry* **39**: 5722–5730.

Hiraiwa N, Nishimura M, 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**: 213–216.

Huntington JA, Read RJ, Carrell RW. 2000. Structure of a serpin–protease complex shows inhibition by deformation. *Nature* **407**: 923–926.

Hwang JE, Hong JK, Je JH, Lee KO, Kim DY, Lee SY, Lim CO. 2009. Regulation of seed germination and seedling growth by an Arabidopsis phytocystatin isoform, AtCYS6. *Plant Cell Reports* **28**: 1623–1632.

Iwasaki T, Kiyohara T, Yoshikawa M. 1976. Amino acid sequence of an active fragment of potato proteinase inhibitor IIa. *Journal of Biochemistry* **79**: 381–391.

Jennings C, West J, Waine C, Craik D, Anderson M. 2001. Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from *Oldenlandia affinis*. *Proceedings of the National Academy of Sciences, USA* **98**: 10614–10619.

Johnson R, Narvaez J, An G, Ryan C. 1989. Expression of proteinase inhibitors I and II in transgenic tobacco plants: effects on natural defense against *Manduca sexta* larvae. *Proceedings of the National Academy of Sciences, USA* **86**: 9871–9875.

Joshi RS, Gupta VS, Giri AP. 2014. Differential antibiosis against *Helicoverpa armigera* exerted by distinct inhibitory repeat domains of *Capsicum annuum* proteinase inhibitors. *Phytochemistry* **101**: 16–22.

Kim T-G, Lee H-J, Jang Y-S, Shin Y-J, Kwon T-H, Yang M-S. 2008. Co-expression of proteinase inhibitor enhances recombinant human granulocyte–macrophage colony stimulating factor production in transgenic rice cell suspension culture. *Protein Expression and Purification* **61**: 117–121.

Kong L, Ranganathan S. 2008. Tandem duplication, circular permutation, molecular adaptation: how *Solanaceae* resist pests via inhibitors. *BMC Bioinformatics* **9**: S22.

Kumar GNM, Houtz RL, Knowles NR. 1999. Age-induced protein modifications and increased proteolysis in potato seed-tubers. *Plant Physiology* **119**: 89–100.

Kunitz M. 1945. Crystallization of a trypsin inhibitor from soybean. *Science* **101**: 668–669.

Lampl N, Alkan N, Davydov O, Fluhr R. 2013. Set-point control of RD21 protease activity by AtSerpin1 controls cell death in Arabidopsis. *Plant Journal* **74**: 498–510.

Lampl N, Budai-Hadrian O, Davydov O, Joss TV, Harrop SJ, Curmi PMG, Roberts TH, Fluhr R. 2010. Arabidopsis AtSerpin1, crystal structure and *in vivo* interaction with its target protease RESPONSIVE TO DESICCATION-21 (RD21). *Journal of Biological Chemistry* **285**: 13550–13560.

Laskowski M, Kato I. 1980. Protein inhibitors of proteinases. *Annual Review of Biochemistry* **49**: 593–626.

Leah R, Mundy J. 1989. The bifunctional α-amylase/subtilisin inhibitor of barley: nucleotide sequence and patterns of seed-specific expression. *Plant Molecular Biology* **12**: 673–682.

Lee MCS, Scanlon MJ, Craik DJ, Anderson MA. 1999. A novel two-chain proteinase inhibitor generated by circularization of a multidomain precursor protein. *Nature Structural & Molecular Biology* **6**: 526–530.

Madureira HC, Da Cunha M, Jacinto T. 2006. Immunolocalization of a defense-related 87 kDa cystatin in leaf blade of tomato plants. *Environmental and Experimental Botany* **55**: 201–208.

Mahatmanto T. 2015. Seed biopharmaceutical cyclic peptides: from discovery to applications. *Peptide Science* **104**: 804–814.

Mahatmanto T, Mylne JS, Poth AG, Swedberg JE, Kaas Q, Schaefer H, Craik DJ. 2015. The evolution of *Momordica* cyclic peptides. *Molecular Biology and Evolution* **32**: 392–405.

Major IT, Constabel CP. 2008. Functional analysis of the Kunitz Trypsin Inhibitor family in poplar reveals biochemical diversity and multiplicity in defence against herbivores. *Plant Physiology* **146**: 888–903.

Margis R, Reis EM, Villeret V. 1998. Structural and phylogenetic relationships among plant and animal cystatins. *Archives of Biochemistry and Biophysics* **359**: 24–30.

Martinez M, Diaz I. 2008. The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. *BMC Evolutionary Biology* **8**: 198.

Martinez M, Diaz-Mendoza M, Carrillo L, Diaz I. 2007. Carboxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cysteine proteinases. *FEBS Letters* **581**: 2914–2918.

Maskos K, Huber-Wunderlich M, Glockshuber R. 1996. RBI, a one-domain α-amylase/trypsin inhibitor with completely independent binding sites. *FEBS Letters* **397**: 11–16.

Meulenbroek EM, Thomassen EAJ, Pouvreau L, Abrahams JP, Gruppen H, Pannu NS. 2012. Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato. *Acta Crystallographica Section D* **68**: 794–799.

Micheelsen PO, Vévodová J, De Maria L, Østergaard PR, Friis EP, Wilson K, Skjøt M. 2008. Structural and mutational analyses of the interaction between the barley α-amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. *Journal of Molecular Biology* **380**: 681–690.

Migliolo L, de Oliveira AS, Santos EA, Franco OL, de Sales MP. 2010. Structural and mechanistic insights into a novel non-competitive Kunitz trypsin inhibitor from *Adenanthera pavonina* L. seeds with double activity toward serine- and cysteine-proteinases. *Journal of Molecular Graphics and Modelling* **29**: 148–156.

Mignery GA, Pikaard CS, Park WD. 1988. Molecular characterization of the patatin multigene family of potato. *Gene* **62**: 27–44.

Mundy J, Svendsen IB, Hejgaard J. 1983. Barley α-amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Research Communications* **48**: 81–90.

Mylne JS, Chan LY, Chanson AH, Daly NL, Schaefer H, Bailey TL, Nguyencong P, Cascales L, Craik DJ. 2012. Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase–mediated biosynthesis. *Plant Cell* **24**: 2765–2778.

Nielsen KJ, Heath RL, Anderson MA, Craik DJ. 1995. Structures of a series of 6-kDa trypsin inhibitors isolated from the stigma of *Nicotiana alata*. *Biochemistry* **34**: 14304–14311.

Nielsen PK, Bønsager BC, Berland CR, Sigurskjold BW, Svensson B. 2003. Kinetics and energetics of the binding between barley α-amylase/subtilisin inhibitor and barley α-amylase 2 analyzed by surface plasmon resonance and isothermal titration calorimetry. *Biochemistry* **42**: 1478–1487.

Nissen MS, Kumar GNM, Youn B, Knowles DB, Lam KS, Ballinger WJ, Knowles NR, Kang C. 2009. Characterization of *Solanum tuberosum* multicystatin and its structural comparison with other cystatins. *Plant Cell* **21**: 861–875.

Odani S, Ikenaka T. 1973. Scission of soybean Bowman-Birk proteinase inhibitor into two small fragments having either trypsin or chymotrypsin inhibitory activity. *Journal of Biochemistry* **74**: 857–860.

Orr GL, Strickland JA, Walsh TA. 1994. Inhibition of *Diabrotica* larval growth by a multicystatin from potato tubers. *Journal of Insect Physiology* **40**: 893–900.

Otlewski J, Zbyryt T. 1994. Single peptide bond hydrolysis/resynthesis in squash inhibitors of serine proteinases. 1. Kinetics and thermodynamics of the interaction between squash inhibitors and bovine beta-trypsin. *Biochemistry* **33**: 200–207.

Outchkourov NS, De Kogel WJ, Wiegers GL, Abrahamson M, Jongsma MA. 2004. Engineered multidomain cysteine protease inhibitors yield resistance against western flower thrips (*Frankliniella occidentalis*) in greenhouse trials. *Plant Biotechnology Journal* **2**: 449–458.

Park EY, Kim J-A, Kim H-W, Kim YS, Song HK. 2004. Crystal Structure of the Bowman-Birk inhibitor from barley seeds in ternary complex with porcine trypsin. *Journal of Molecular Biology* **343**: 173–186.

Pekkarinen AI, Jones BL. 2003. Purification and identification of barley proteins that inhibit the alkaline Serine proteinases of *Fusarium culmorum*. *Journal of Agricultural and Food Chemistry* **51**: 1710–1717.

Pekkarinen AI, Longstaff C, Jones BL. 2007. Kinetics of the inhibition of *Fusarium* Serine proteinases by barley (*Hordeum vulgare* L.) inhibitors. *Journal of Agricultural and Food Chemistry* **55**: 2736–2742.

Pouvreau L, Gruppen H, Piersma SR, van den Broek LAM, van Koningsveld GA, Voragen AGJ. 2001. Relative abundance and inhibitory distribution of protease

inhibitors in potato juice from cv. Elkana. *Journal of Agricultural and Food Chemistry* **49**: 2864–2874.

Qu L-J, Chen J, Liu M, Pan N, Okamoto H, Lin Z, Li C, Li D, Wang J, Zhu G *et al.* 2003. Molecular cloning and functional analysis of a novel type of Bowman-Birk inhibitor gene family in rice. *Plant Physiology* **133**: 560–570.

Rakwal R, Kumar Agrawal G, Jwa N-S. 2001. Characterization of a rice (Oryza sativa L.) Bowman-Birk proteinase inhibitor: tightly light regulated induction in response to cut, jasmonic acid, ethylene and protein phosphatase 2A inhibitors. *Gene* **263**: 189–198.

Rawlings ND, Waller M, Barrett AJ, Bateman A. 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* **42**: D503–D509.

Rehman AA, Ahsan H, Khan FH. 2013. α2-macroglobulin: a physiological guardian. *Journal of Cellular Physiology* **228**: 1665–1675.

Renko M, Sabotič J, Turk D. 2012. β-Trefoil inhibitors – from the work of Kunitz onward. *Biological Chemistry* **393**: 1043–1054.

Roberts TH, Hejgaard J. 2007. Serpins in plants and green algae. *Functional & Integrative Genomics* **8**: 1–27.

Roberts TH, Hejgaard J, Saunders NFW, Cavicchioli R, Curmi PMG. 2004. Serpins in unicellular Eukarya, Archaea, and Bacteria: sequence analysis and evolution. *Journal of Molecular Evolution* **59**: 437–447.

Roberts TH, Marttila S, Rasmussen SK, Hejgaard J. 2003. Differential gene expression for suicide-substrate serine proteinase inhibitors (serpins) in vegetative and grain tissues of barley. *Journal of Experimental Botany* **54**: 2251–2263.

Rodis P, Hoff JE. 1984. Naturally occurring protein crystals in the potato. *Plant Physiology* **74**: 907–911.

Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG. 1995. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. *Biochemistry* **34**: 4147–4158.

Santamaría ME, Diaz-Mendoza M, Diaz I, Martinez M. 2014. Plant protein peptidase inhibitors: an evolutionary overview based on comparative genomics. *BMC Genomics* **15**: 812–826.

Saska I, Gillon AD, Hatsugai N, Dietzgen RG, Hara-Nishimura I, Anderson MA, Craik DJ. 2007. An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *Journal of Biological Chemistry* **282**: 29721–29728.

Scanlon MJ, Lee MC, Anderson MA, Craik DJ. 1999. Structure of a putative ancestral protein encoded by a single sequence repeat from a multidomain proteinase inhibitor gene from *Nicotiana alata*. *Structure* **7**: 793–802.

Shindo T, Misas-Villamil JC, Hörger AC, Song J, Van der Hoorn RAL. 2012. A role in immunity for Arabidopsis cysteine protease RD21, the ortholog of the tomato immune protease C14. *PLoS ONE* **7**: e29317.

Shivaraj N, Pattabiraman T. 1981. Natural plant enzyme inhibitors. Characterization of an unusual alpha-amylase/trypsin inhibitor from ragi (*Eleusine coracana* Geartn.). *Biochemical Journal* **193**: 29–36.

Siqueira-Júnior CL, Fernandes KVS, Machado OLT, da Cunha M, Gomes VM, Moura D, Jacinto T. 2002. 87 kDa tomato cystatin exhibits properties of a defense protein and forms protein crystals in prosystemin overexpressing transgenic plants. *Plant Physiology and Biochemistry* **40**: 247–254.

Song HK, Kim YS, Yang JK, Moon J, Lee JY, Suh SW. 1999. Crystal structure of a 16 kda double-headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *Journal of Molecular Biology* **293**: 1133–1144.

Sottrup-Jensen L. 1989. α-macroglobulins: structure, shape, and mechanism of proteinase complex formation. *Journal of Biological Chemistry* **264**: 11539–11542.

Strobl S, Maskos K, Wiegand G, Huber R, Gomis-Rüth FX, Glockshuber R. 1998. A novel strategy for inhibition of α-amylases: yellow meal worm α-amylase in complex with the Ragi bifunctional inhibitor at 2.5 Å resolution. *Structure* **6**: 911–921.

Strobl S, Muehlhahn P, Bernstein R, Wiltscheck R, Maskos K, Wunderlich M, Huber R, Glockshuber R, Holak TA. 1995. Determination of the three-dimensional structure of the bifunctional α-amylase/trypsin inhibitor from Ragi seeds by NMR spectroscopy. *Biochemistry* **34**: 8281–8293.

Stubbs MT, Laber B, Bode W, Huber R, Jerala R, Lenarcic B, Turk V. 1990. The refined 2.4 Å x-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO Journal* **9**: 1939–1947.

Sweet RM, Wright HT, Janin J, Chothia CH, Blow DM. 1974. Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6-Å resolution. *Biochemistry* **13**: 4212–4228.

Tajima T, Yamaguchi A, Matsushima S, Satoh M, Hayasaka S, Yoshimatsu K, Shioi Y. 2011. Biochemical and molecular characterization of senescence-related cysteine protease–cystatin complex from spinach leaf. *Physiologia Plantarum* **141**: 97–116.

Tamhane VA, Chougule NP, Giri AP, Dixit AR, Sainani MN, Gupta VS. 2005. *In vivo* and *in vitro* effect of *Capsicum annum* proteinase inhibitors on *Helicoverpa armigera* gut proteinases. *Biochimica et Biophysica Acta* **1722**: 156–167.

Tamhane VA, Giri AP, Kumar P, Gupta VS. 2009. Spatial and temporal expression patterns of diverse Pin-II proteinase inhibitor genes in *Capsicum annuum* Linn. *Gene* **442**: 88–98.

Thomas JC, Adams DG, Keppenne VD, Wasmann CC, Brown JK, Kanost MR, Bohnert HJ. 1995. Protease inhibitors of *Manduca sexta* expressed in transgenic cotton. *Plant Cell Reports* **14**: 758–762.

Thomas JC, Wasmann CC, Echt C, Dunn RL, Bohnert HJ, McCoy TJ. 1994. Introduction and expression of an insect proteinase inhibitor in alfalfa *Medicago sativa* L. *Plant Cell Reports* **14**: 31–36.

Thongyoo P, Bonomelli C, Leatherbarrow RJ, Tate EW. 2009. Potent inhibitors of β-tryptase and human leukocyte elastase based on the MCoTI-II scaffold. *Journal of Medicinal Chemistry* **52**: 6197–6200.

Thongyoo P, Roqué-Rosell N, Leatherbarrow RJ, Tate EW. 2008. Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Organic & Biomolecular Chemistry* **6**: 1462–1470.

Uppalapati SR, Ayoubi P, Weng H, Palmer DA, Mitchell RE, Jones W, Bender CL. 2005. The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. *The Plant Journal* **42**: 201–217.

Vallée F, Kadziola A, Bourne Y, Juy M, Rodenburg KW, Svensson B, Haser R. 1998. Barley α-amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 Å resolution. *Structure* **6**: 649–659.

Valueva TA, Revina TA, Mosolov VV, Mentele R. 2000. Primary structure of potato kunitz-type serine proteinase inhibitor. *Biological Chemistry* **381**: 1215–1221.

Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan J-A, De Rycke R, Brackenier A, Inzé D, Harris JL, Van Breusegem F. 2006. Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for Metacaspase 9. *Journal of Molecular Biology* **364**: 625–636.

Voss RH, Ermler U, Essen LO, Wenzl G, Kim YM, Flecker P. 1996. Crystal structure of the bifunctional soybean Bowman-Birk inhibitor at 0.28-nm resolution. Structural peculiarities in a folded protein conformation. *European Journal of Biochemistry* **242**: 122–131.

Walsh TA, Strickland JA. 1993. Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato releases functional cystatin domains. *Plant Physiology* **103**: 1227–1234.

Weeda SM, Mohan Kumar GN, Richard Knowles N. 2009. Developmentally linked changes in proteases and protease inhibitors suggest a role for potato multicystatin in regulating protein content of potato tubers. *Planta* **230**: 73–84.

Xu D, Xue Q, McElroy D, Mawal Y, Hilder VA, Wu R. 1996. Constitutive expression of a cowpea trypsin inhibitor gene, CpTi, in transgenic rice plants confers resistance to two major rice insect pests. *Molecular Breeding* **2**: 167–173.

Yamagata H, Kunimatsu K, Kamasaka H, Kuramoto T, Iwasaki T. 1998. Rice bifunctional α-amylase/subtilisin inhibitor: characterization, localization, and changes in developing and germinating seeds. *Bioscience, Biotechnology, and Biochemistry* **62**: 978–985.

Yoo B-C, Aoki K, Xiang Y, Campbell LR, Hull RJ, Xoconostle-Cazares B, Monzer J, Lee J-Y, Ullman DE, Lucas WJ. 2000. Characterization of *Cucurbita maxima* Phloem Serpin-1 (CmPS-1): a developmentally regulated elastase inhibitor. *Journal of Biological Chemistry* **275**: 35122–35128.

Zakharov A, Carchilan M, Stepurina T, Rotari V, Wilson K, Vaintraub I. 2004. A comparative study of the role of the major proteinases of germinated common bean (*Phaseolus vulgaris* L.) and soybean (*Glycine max* (L.) Merrill) seeds in the degradation of their storage proteins. *Journal of Experimental Botany* **55**: 2241–2249.