Quantitative proteomics in plant protease substrate identification

Fatih Demir[1,*], Stefan Niedermaier[1,*], Joji Grace Villamor[1], Pitter F. Huesgen[1]

[1]Forschungszentrum Jülich, Central Institute for Engineering, Electronics and Analytics, ZEA-3Analytics, Wilhelm-Johnen-Str., 52425 Jülich, Germany

Author for correspondence:
Pitter F. Huesgen
Phone: +49 2461 61-6873
Fax: +49 2461 61-2560
Email: p.huesgen@fz-juelich.de

*These authors contributed equally to this work.

Total word count: 2707 (Introduction, Main text, and Acknowledgments)
2 Color figures
2 Text boxes
Supporting Information: 1 Supporting Table

This is the peer reviewed version of the following article: New Phytol. 2018 May;218(3):936-943. doi: 10.1111/nph.14587. Epub 2017 May 11, which has been published in final form at https://doi.org/10.1111/nph.14587.

Contents

Summary

Proteolysis is a central regulatory mechanism of protein homeostasis and protein function that affects all aspects of plant life. Higher plants encode for hundreds of proteases, but their physiological substrates and hence their molecular functions remain mostly unknown. Current quantitative mass spectrometry-based proteomics enables unbiased large-scale interrogation of the proteome and its modifications. Here we provide an overview over proteomics techniques that allow profiling of changes in protein abundance, measurement of proteome turnover rates, identification of protease cleavage sites in vivo and in vitro and determination of protease sequence specificity. We discuss how these techniques can help to reveal protease substrates and determine plant protease function, illustrated by recent studies on selected plant proteases.

Keywords: Degradomics, Plants, Proteases, Protease substrates, Proteomics, Proteolysis, Terminome

I. Introduction

Proteolysis maintains proteostasis (Nelson & Millar, 2015) and regulates signaling and other physiological processes by selective elimination of target proteins (Gibbs *et al.*, 2016) or site-specific proteolytic processing (Qiao *et al.*, 2012). The latter, also called limited proteolysis, is an essentially irreversible protein modification that generates new proteoforms with altered location, activity and/or function (Lange & Overall, 2013). Proteolytic processes occur in most compartments of the plant cell and are involved in all aspects of plant life, including growth, development and plant-environment interactions (van der Hoorn, 2008; van Wijk, 2015). Plant genomes encode for a large variety of proteases, the enzymes that catalyze peptide bond hydrolysis, including >650 protease-coding genes in rice (*Oryza sativa*) and >800 in *Arabidopsis thaliana* (van der Hoorn, 2008). Key to understanding protease function is knowledge of their physiological substrates. However, for the vast majority of plant proteases no physiological substrates have been identified to date. Consequently, the physiological role of most of these enzymes and the molecular mechanisms underlying many observed mutant phenotypes remain elusive. Since proteases directly affect protein abundance, size and sequence, sensitive modern mass spectrometry (MS)-based proteomics appears predestined to tackle this challenge. Here we review proteomics techniques suited to determine protease substrates and protease function (**Box 1**, **Table S1**), with a focus on studies of higher plant proteases published within the last two years.

II. The quest for plant protease substrates – proteomics to the rescue?

"Top-down" proteomics directly analyzes intact proteins and thus retains information on differentially modified proteoforms (Toby *et al.*, 2016). While promising for protease research, top-down proteomics is still technically challenging, places high demands on instrument performance and achieves only limited proteome coverage in complex samples. Therefore the "bottom-up" proteomics approach where proteins are enzymatically digested into peptides dominates today (**Box 2**), despite the challenge that shared non-unique peptides often impede unambiguous identification of different proteoforms (Smith *et al.*, 2013). One dimensional (1D)- and two dimensional (2D)-gel electrophoresis (GE) techniques provide information on molecular mass and isoelectric point of identified proteins, which may distinguish proteoforms (Huang *et al.*, 2015). However, these methods are labor-intensive and biased towards abundant proteins (**Table S1**). With improvements in instrumentation, sample preparation protocols and data analysis software, gel-free proteomics has become more popular and now routinely provides identification and relative quantification of

thousands of proteins and system-wide identification of post translational modifications and interaction partners (Aebersold & Mann, 2016).

III. Quantitative proteome comparison reveals candidate substrates

Quantitative proteomics is widely used in protease research to compare proteomes exposed to varying levels of protease activity in vivo. For example, a series of elegant studies on the cage-forming multi-subunit chloroplast caseinolytic protease (Clp) investigated *A. thaliana* proteomes from mutants lacking individual subunits by label-free shotgun proteomics, revealing chloroplast proteostasis phenotypes of varying severity (reviewed (Nishimura & van Wijk, 2015). Quantitative affinity proteomics identified interactors of the substrate adaptor subunit ClpS1 that recognizes and targets substrates to the Clp complex (Nishimura *et al.*, 2013). One of the interactors, ClpF, was a novel substrate adaptor component that together with ClpS1 targeted Glutamyl T-RNA Reductase, a key enzyme of tetrapyrrole biosynthesis, for Clp-mediated degradation (Nishimura *et al.*, 2015).

The same laboratory identified the *A. thaliana* plastoglobule (PG)-located metalloprotease of the M48 family (PGM48) as a positive regulator of senescence (Bhuiyan *et al.*, 2016). Label-free shotgun analysis of PG proteomes showed that PGM48, the only protease present in this compartment, significantly increased while CAROTENOID CLEAVAGE ENZYME 4 (CCD4) decreased during senescence. The accumulation CCD4 in PGM-deficient plants suggested CCD4 as a potential substrate of PGM48, which was further supported by their co-localization and evidence for direct interaction.

Two dimensional-differential fluorescence gel electrophoresis (2D-DIGE) was pivotal in establishing *Solanum lycopersicum* matrix metalloproteinases 2 and 3 (Sl2/3-MMP) as negative regulators of cell death (Zimmermann *et al.*, 2016). Sl2/3-MMP-silenced plants exhibited necrotic lesions that spread with development and were accompanied with an accumulation of the subtilisin-like serine protease P69B, which did not result from transcriptional changes. An in vitro screen for Sl2/3-MMP substrates by two-dimensional electrophoretic mobility shift assay (2D-EMSA) also identified P69B: Apoplast proteins from Sl2/3-MMP-silenced plants were separated by gel electrophoresis, lanes excised and in-gel digested with recombinant Sl2-MMP or Sl3-MMP before repeated separation in a second dimension. P69B fragments with increased electrophoretic mobility were identified by MS,

and zymography showed that they were proteolytically inactive. Thus, Sl2/3-MMPs suppressed P69B-dependent cell death in an extracellular proteolytic cascade.

In another gel-based study, 2D-DIGE comparison of mitochondrial proteomes from wild type (wt) and FTSH4-deficient *A. thaliana* found lower accumulation of the oxidative phosphorylation (OXPHOS) complexes I and IV and enzymes of the tricarboxylic acid (TCA) cycle, whereas chaperones and antioxidant enzymes accumulated (Smakowska *et al.*, 2016). Widespread protein carbonylation indicated increased oxidative stress in absence of FTSH4, likely an indirect effect arising from chronic ATP deficiency, altered phospholipid content and impairment of mitochondrial proteostasis.

IV. Dynamic metabolic stable isotope labeling to measure protein turnover in vivo

Differences in steady state protein abundance may be caused by changes in synthesis or degradation, which can be deconvoluted using metabolic stable isotope labeling in pulse or pulse-chase experiments (Nelson & Millar, 2015). In auxotrophic organisms, such dynamic metabolic labeling is mostly performed by Stable Isotope Labeling with Amino Acids (SILAC,(Ong, 2012). Plants autotrophically synthesize and interconvert all proteinogenic amino acids resulting in inefficient labeling in most growth conditions (**Table S1**). Near-complete labeling >95% has been achieved in *A. thaliana* seedlings germinated and grown in liquid culture (Lewandowska *et al.*, 2013), but this may not work in species with larger seed nitrogen stores. Labeling with ^{15}N-enriched inorganic salts in growth media is therefore the method of choice for metabolic stable isotope labeling in plants (Matthes *et al.*, 2014). In contrast to SILAC, the mass shifts introduced by ^{15}N labeling increase with the number on N atoms in the peptide. In addition, the isotopic purity of ^{15}N salts is typically only 98%. ^{15}N labeled peptides therefore exhibit variable mass shifts with wide isotope envelopes, challenging peptide identification, automatic matching of corresponding heavy/light peptides and quantification. Dynamic labeling further complicates the situation by the presence of partially labeled peptides with varying ^{14}N/^{15}N content. Nevertheless, development of dedicated data analysis pipelines enabled measurement of steady-state protein turnover rates in several species, including *Hordeum vulgare* (Nelson *et al.*, 2014), *Medicago truncatula* (Lyon *et al.*, 2016) and *A. thaliana* (Li *et al.*, 2017b).

Two publications determined changes in protein turnover rates in protease-deficient mutants by dynamic metabolic labeling. In the first, *A. thaliana* INTERMEDIATE CLEAVING PEPTIDASE55 (ICP55) was shown to remove a single N-terminal amino acid from selected mitochondrial proteins after import, a processing step that resulted in altered mitochondrial protein degradation rates in vitro and in vivo (Huang *et al.*, 2015). The second study measured protein turnover in *A. thaliana* wt and mutants lacking the mitochondrial LON1 protease (Li *et al.*, 2017a). Both lines were grown hydroponically and switched to ^{15}N-containing media after 4d to mark newly synthesized proteins with a heavy isotope label, while degradation rates were determined based on the decay of peptides with natural isotope abundance. Out of 400 observed mitochondrial proteins, 205 differed significantly in turnover rates in LON1-deficient plants. Several proteins with slow degradation rates accumulated, including TCA cycle enzymes, suggesting they are likely LON1 substrates. In contrast, several OXPHOS complex subunits showed faster degradation rates and decreased overall abundance in the absence of LON1, while assembled complexes were stable. This supported an additional function of LON1 as a chaperone in complex assembly (Li *et al.*, 2017a).

V. Terminomics - Large scale identification of protease cleavage sites

Proteolytic cleavage generates two polypeptide chains, one exposing a new protease-generated (neo-) N-terminus, the other a neo-C-terminus. Protein termini thus provide the most direct readout of protease activity and uniquely reveal the precise substrate cleavage sites defining physiologically relevant proteoforms. However, terminal peptides form only a minor fraction among the peptides released by enzymatic digest in bottom-up proteomics and are often not considered during spectra matching due to their semi-specific nature (**Box 2**). To facilitate identification of N- or C-terminal peptides from complex proteomes, the "N- and C-terminomes", techniques for selective enrichment have been developed (reviewed in (Huesgen & Overall, 2012; Plasman *et al.*, 2013).

Most widely used in plant sciences are Terminal Amine Isotope Labeling of Substrates (TAILS, (Kleifeld *et al.*, 2010), COmbined FRActional DIagonal Chromatography (COFRADIC, (Gevaert *et al.*, 2003), and its derivative Charged-based FRActional DIagonal Chromatography (ChaFRADIC, (Venne *et al.*, 2013) (**Figure 1, Table S1**). All three enrich N-terminal peptides by negative selection. First, primary amines at protein N termini and Lys side chains are modified on intact proteins, allowing simultaneous labeling with stable

isotope reagents (**Figure 1**). Subsequent enzymatic proteome digest generates new N-terminal primary amines on internal and C-terminal peptides that are used to deplete these undesired peptides either by polymer capture in the TAILS workflow or sequential chromatography with intermittent chemical primary amine derivatization in the FRADIC methods (**Figure 1**). Dedicated step-by-step protocols for plant terminome profiling are available for TAILS (Demir *et al.*, 2017) and COFRADIC (Tsiatsiani *et al.*, 2014).

N-terminome analyses have been used to study proteolytic processing after protein import into plastids of *A. thaliana* (Kohler *et al.*, 2015b; Rowland *et al.*, 2015), the diatom *Thalassiosira pseudonana* (Huesgen *et al.*, 2013) and the glaucophyte *Cyanophora paradoxa* (Kohler *et al.*, 2015a), and to investigate N-terminal sequence determinants for protein stability in *A. thaliana* (Venne *et al.*, 2015; Zhang *et al.*, 2015) revealing mostly stabilizing N-terminal amino acids in agreement with the N-end rule (Gibbs *et al.*, 2016).

Three recent publications used N termini enrichment to identify plant protease substrates. ChaFRADIC analysis of mitochondrial protein N termini in *A. thaliana* loss-of-function mutants of ICP55 and OCTAPEPTIDYL AMINOPEPTIDASE1 (OCT1) to wt identified differential processing by a single amino acid in 88 putative ICP55 substrates, and differential octapeptide processing in 7 putative OCT1 substrates (Carrie *et al.*, 2015). In agreement with their yeast homologues (Poveda-Huertes *et al.*, 2017), ICP55 and OCT1 removed destabilizing protein N-terminal residues after mitochondrial signal peptide cleavage in distinct subsets of mitochondrial proteins. This is consistent with altered turnover rates observed in ICP55-deficient plants (Huang *et al.*, 2015).

In a landmark study, COFRADIC was used to identify *A. thaliana* METACASPASE9 (MC9) substrates (Tsiatsiani *et al.*, 2013). MC9-deficient seedlings where compared to wt and plants overexpressing MC9 in the mutant background, as were isolated MC9-deficient proteomes treated with active or catalytically inactive recombinant MC9. N termini accumulating in the presence of MC9 activity were filtered using the known strict sequence preference for Arg or Lys, which classified 551 cleavage sites in 392 proteins as likely MC9-generated. Of these, 99 cleavage sites in 74 proteins were either identified only in vivo or matched to proteins identified in at least two of the N terminome screens. Cleavage assays with synthetic peptides and in vitro transcribed and translated radiolabeled proteins validated a number of these as substrates, including PHOSPHOENOLPYRUVATE CARBOXYKINASE 1 (PEPCK1),

which co-localized with MC9 in vivo. Further analysis demonstrated that MC9 contributes to control of gluconeogenesis by activating PEPCK1 in vivo.

In a similar approach, COFRADIC and 2D-DIGE quantitative proteomics were used to investigate the function of the three Deg/HtrA proteases HhoA, HhoB and HtrA in *Synechocystis* sp. PCC 6803 under normal growth conditions (Tam *et al.*, 2015). Comparisons of strains lacking individual proteases to wt were combined with analyses of mutant proteomes exposed to the corresponding recombinant enzyme in vitro, identifying both common and distinct substrates that affected major metabolic pathways.

VI. Substrate or not substrate, that is the question

Modulation of protease activity in planta affects, abundance and activity and expression of many proteins, frequently including other proteolytic enzymes and/or inhibitors as shown by the studies discussed above. Plant proteases, just as their mammalian homologues, thus do not operate in isolation, but in a complex network, termed the "protease web" (Fortelny *et al.*, 2014). Consequently, changes observed in vivo using quantitative proteome and terminome analyses often represent indirect effects. In contrast, proteins cleaved by recombinant proteases in vitro may not represent physiological substrates due to different expression patterns, modifications, subcellular localizations and reaction conditions in vivo.

Therefore, any protein affected in abundance and/or processing state in proteomics studies should only be considered as candidate substrate for the protease of interest that must be further validated as physiologically relevant substrates (**Figure 2**). Suitable experimental approaches are highly dependent on the target, but include confirmation of co-expression and co-localization in vivo (Tsiatsiani *et al.*, 2013; Bhuiyan *et al.*, 2016), tests for direct interaction in vivo and in vitro (Nishimura *et al.*, 2015; Bhuiyan *et al.*, 2016), evaluation whether complementation lines shows the expected opposite effect (Tsiatsiani *et al.*, 2013; Zimmermann *et al.*, 2016) and evaluation whether the protease is active under physiological conditions. Activity based protein profiling (ABPP) with class-specific chemical probes can monitor protease activity in vivo and in vitro and, as a chemical proteomics method (**Table S1**), reveal the proteases that participate in the active protease web of a given tissue or proteome (Morimoto & van der Hoorn, 2016).

Protease sequence specificity is a useful filter to select likely direct candidate substrates (Tsiatsiani *et al.*, 2013) or predict physiologically relevant processing sites (Schardon *et al.*, 2016). If the sequence specificity is not known, Proteomic Identification of protease Cleavage Sites (PICS) can quickly provide experimental information for both sides of the cleaved peptide bond simultaneously (Schilling & Overall, 2008). In a PICS experiment, peptide libraries are generated by proteome digests with specific proteases such as trypsin or GluC (**Figure 2**). These proteome-derived peptide libraries are incubated with the recombinant protease of interest and cleaved peptides identified after enrichment (Schilling & Overall, 2008) or using differential stable isotope labeling (Biniossek *et al.*, 2016). The full cleavage sites are inferred by matching the identified cleavage product to the proteome sequence databases. Alignment of the dozens to hundreds of cleavage sites identified in a typical PICS experiment produced a detailed sequence specificity profile that can distinguish related enzymes (Marino *et al.*, 2014).

Computational tools can assist in distinguishing indirect and direct effects. Large-scale information on gene expression (Fucile *et al.*, 2011) and subcellular localization (Hooper *et al.*, 2017) are useful predictors whether proteases and candidate substrates are likely to meet in vivo. The peptidase database MEROPS collects information on proteolytic enzymes, their substrates and inhibitors (Rawlings *et al.*, 2016) and determines protease specificity from known substrates (Rawlings, 2016). The termini-centric database TOPFIND integrates data from published terminome studies and MEROPS with information on protein domains and post-translational modifications, enabling prediction of the functional consequences of observed proteolytic cleavages (Fortelny *et al.*, 2015). The integrated TOPFINDer tool further allows identification of proteolytic pathways that link observed termini to a protease of interests, but a lack of deposited data currently limits such network predictions to the best studied proteases in mouse and man. Determination of protease function is particularly difficult if redundant enzymes are able to cut the same substrate at the same site. An elegant experimental solution of this problem was very recently presented by tissue-specific expression of a family-selective protease inhibitor (Schardon *et al.*, 2016).

VII. Concluding remarks

We have reviewed current mass spectrometry-based proteomics methods enabling proteome-wide i) quantification of changes in steady-state abundance, ii) measurement of turnover rates, iii) identification of protein termini and thus in vivo protease cleavage sites, iv)

identification of candidate protease substrates and their cleavage sites in vitro and v) rapid profiling of recombinant protease specificity. These provide fascinating insights into plant protease function, but unambiguous identification of the physiological substrates of a protease of interest remains challenging. Complementary in vitro and in vivo approaches, rational stratification of candidate substrates and thorough hypothesis testing are indispensable to discriminate direct substrates from indirect effects. Computational tools assist in this task and are expected to improve as data on more plant proteases becomes available. Major challenges are now to determine the consequences of identified substrate cleavages and to improve our currently still highly fragmented map of the protease and protease-inhibitor interactions that form the plant protease web.

Acknowledgements

We apologize to all authors whose valuable contributions could not be cited due to space limitations. This work is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (project ProPlantStress, grant agreement No 639905, to P.F.H.).

References

Aebersold R, Mann M. 2016. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**(7620): 347-355.

Bhuiyan NH, Friso G, Rowland E, Majsec K, van Wijk KJ. 2016. The Plastoglobule-Localized Metallopeptidase PGM48 Is a Positive Regulator of Senescence in *Arabidopsis thaliana*. *Plant Cell* **28**(12): 3020-3037.

Biniossek ML, Niemer M, Maksimchuk K, Mayer B, Fuchs J, Huesgen PF, McCafferty DG, Turk B, Fritz G, Mayer J, et al. 2016. Identification of Protease Specificity by Combining Proteome-Derived Peptide Libraries and Quantitative Proteomics. *Mol Cell Proteomics* **15**(7): 2515-2524.

Carrie C, Venne AS, Zahedi RP, Soll J. 2015. Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in *Arabidopsis thaliana*. *J Exp Bot* **66**(9): 2691-2708.

Demir F, Niedermaier S, Kizhakkedathu JN, Huesgen PF. 2017. Profiling of Protein N-Termini and Their Modifications in Complex Samples. *Meth Mol Biol* **1574**: DOI 10.1007/1978-1001-4939-6850-1003_1004.

Fortelny N, Cox JH, Kappelhoff R, Starr AE, Lange PF, Pavlidis P, Overall CM. 2014. Network analyses reveal pervasive functional regulation between proteases in the human protease web. *PLoS Biol* **12**(5): e1001869.

Fortelny N, Yang S, Pavlidis P, Lange PF, Overall CM. 2015. Proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of protein termini to pre- and post-translational events. *Nucleic Acids Res* **43**(D1): D290-D297.

Fucile G, Di Biase D, Nahal H, La G, Khodabandeh S, Chen Y, Easley K, Christendat D, Kelley L, Provart NJ. 2011. ePlant and the 3D data display initiative: integrative systems biology on the world wide web. *PLoS One* **6**(1): e15237.

Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, Vandekerckhove J. 2003. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* **21**(5): 566-569.

Gibbs DJ, Bailey M, Tedds HM, Holdsworth MJ. 2016. From start to finish: amino-terminal protein modifications as degradation signals in plants. *New Phytol* **211**(4): 1188-1194.

Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH. 2017. SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res* **45**(D1): D1064-1074.

Huang S, Nelson CJ, Li L, Taylor NL, Stroher E, Peteriet J, Millar AH. 2015. INTERMEDIATE CLEAVAGE PEPTIDASE55 Modifies Enzyme Amino Termini and Alters Protein Stability in *Arabidopsis* Mitochondria. *Plant Physiol* **168**(2): 415-427.

Huesgen PF, Alami M, Lange PF, Foster LJ, Schroder WP, Overall CM, Green BR. 2013. Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS One* **8**(9): e74483.

Huesgen PF, Overall CM. 2012. N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. *Physiol Plant* **145**(1): 5-17.

Kleifeld O, Doucet A, auf dem Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN, Overall CM. 2010. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* **28**(3): 281-288.

Kohler D, Dobritzsch D, Hoehenwarter W, Helm S, Steiner JM, Baginsky S. 2015a. Identification of protein N-termini in *Cyanophora paradoxa* cyanelles: transit peptide composition and sequence determinants for precursor maturation. *Front Plant Sci* **6**: 559.

Kohler D, Montandon C, Hause G, Majovsky P, Kessler F, Baginsky S, Agne B. 2015b. Characterization of chloroplast protein import without Tic56, a component of the 1-megadalton translocon at the inner envelope membrane of chloroplasts. *Plant Physiol* **167**(3): 972-990.

Lange PF, Overall CM. 2013. Protein TAILS: when termini tell tales of proteolysis and function. *Curr Opin Chem Biol* **17**(1): 73-82.

Lewandowska D, ten Have S, Hodge K, Tillemans V, Lamond AI, Brown JW. 2013. Plant SILAC: stable-isotope labelling with amino acids of *Arabidopsis* seedlings for quantitative proteomics. *PLoS One* **8**(8): e72207.

Li L, Nelson C, Fenske R, Trosch J, Pruzinska A, Millar AH, Huang S. 2017a. Changes in specific protein degradation rates in *Arabidopsis thaliana* reveal multiple roles of Lon1 in mitochondrial protein homeostasis. *Plant J* **89**: 458-471.

Li L, Nelson CJ, Trosch J, Castleden I, Huang S, Millar AH. 2017b. Protein Degradation Rate in Arabidopsis thaliana Leaf Growth and Development. *Plant Cell*: DOI 10.1105/tpc.1116.00768.

Lyon D, Castillejo MA, Mehmeti-Tershani V, Staudinger C, Kleemaier C, Wienkoop S. 2016. Drought and Recovery: Independently Regulated Processes Highlighting the Importance of Protein Turnover Dynamics and Translational Regulation in *Medicago truncatula*. *Mol Cell Proteomics* **15**(6): 1921-1937.

Marino G, Huesgen PF, Eckhard U, Overall CM, Schroder WP, Funk C. 2014. Family-wide characterization of matrix metalloproteinases from *Arabidopsis thaliana* reveals their distinct proteolytic activity and cleavage site specificity. *Biochem J* **457**(2): 335-346.

Matthes A, Kohl K, Schulze WX. 2014. SILAC and alternatives in studying cellular proteomes of plants. *Methods Mol Biol* **1188**: 65-83.

Morimoto K, van der Hoorn RA. 2016. The Increasing Impact of Activity-Based Protein Profiling in Plant Science. *Plant Cell Physiol* **57**(3): 446-461.

Nelson CJ, Alexova R, Jacoby RP, Millar AH. 2014. Proteins with high turnover rate in barley leaves estimated by proteome analysis combined with in planta isotope labeling. *Plant Physiol* **166**(1): 91-108.

Nelson CJ, Millar AH. 2015. Protein turnover in plant biology. *Nature Plants* **1**(3): 15017.

Nishimura K, Apitz J, Friso G, Kim J, Ponnala L, Grimm B, van Wijk KJ. 2015. Discovery of a Unique Clp Component, ClpF, in Chloroplasts: A Proposed Binary ClpF-ClpS1 Adaptor Complex Functions in Substrate Recognition and Delivery. *Plant Cell* **27**(10): 2677-2691.

Nishimura K, Asakura Y, Friso G, Kim J, Oh SH, Rutschow H, Ponnala L, van Wijk KJ. 2013. ClpS1 is a conserved substrate selector for the chloroplast Clp protease system in *Arabidopsis*. *Plant Cell* **25**(6): 2276-2301.

Nishimura K, van Wijk KJ. 2015. Organization, function and substrates of the essential Clp protease system in plastids. *Biochim Biophys Acta* **1847**(9): 915-930.

Ong SE. 2012. The expanding field of SILAC. *Anal Bioanal Chem* **404**(4): 967-976.

Plasman K, Van Damme P, Gevaert K. 2013. Contemporary positional proteomics strategies to study protein processing. *Curr Opin Chem Biol* **17**(1): 66-72.

Poveda-Huertes D, Mulica P, Vogtle FN. 2017. The versatility of the mitochondrial presequence processing machinery: cleavage, quality control and turnover. *Cell Tissue Res* **367**(1): 73-81.

Qiao H, Shen Z, Huang SS, Schmitz RJ, Urich MA, Briggs SP, Ecker JR. 2012. Processing and subcellular trafficking of ER-tethered EIN2 control response to ethylene gas. *Science* **338**(6105): 390-393.

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, Barrett AJ, Finn R. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **44**(D1): D343-350.

Rowland E, Kim J, Bhuiyan NH, van Wijk KJ. 2015. The *Arabidopsis* Chloroplast Stromal N-Terminome: Complexities of Amino-Terminal Protein Maturation and Stability. *Plant Physiol* **169**(3): 1881-1896.

Schardon K, Hohl M, Graff L, Pfannstiel J, Schulze WX, Stintzi A, Schaller A. 2016. Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science* **354**(6319): 1594-1597.

Schilling O, Overall CM. 2008. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **26**(6): 685-694.

Smakowska E, Skibior-Blaszczyk R, Czarna M, Kolodziejczak M, Kwasniak-Owczarek M, Parys K, Funk C, Janska H. 2016. Lack of FTSH4 Protease Affects Protein Carbonylation, Mitochondrial Morphology, and Phospholipid Content in Mitochondria of *Arabidopsis*: New Insights into a Complex Interplay. *Plant Physiol* **171**(4): 2516-2535.

Smith LM, Kelleher NL, Consortium for Top Down P. 2013. Proteoform: a single term describing protein complexity. *Nat Methods* **10**(3): 186-187.

Tam LX, Aigner H, Timmerman E, Gevaert K, Funk C. 2015. Proteomic approaches to identify substrates of the three Deg/HtrA proteases of the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochem J* **468**(3): 373-384.

Toby TK, Fornelli L, Kelleher NL. 2016. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu Rev Anal Chem* **9**(1): 499-519.

Tsiatsiani L, Stael S, Van Damme P, Van Breusegem F, Gevaert K. 2014. Preparation of *Arabidopsis thaliana* seedling proteomes for identifying metacaspase substrates by N-terminal COFRADIC. *Methods Mol Biol* **1133**: 255-261.

Tsiatsiani L, Timmerman E, De Bock PJ, Vercammen D, Stael S, van de Cotte B, Staes A, Goethals M, Beunens T, Van Damme P, et al. 2013. The *Arabidopsis* metacaspase9 degradome. *Plant Cell* **25**(8): 2831-2847.

van der Hoorn RA. 2008. Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* **59**: 191-223.

van Wijk KJ. 2015. Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. *Annu Rev Plant Biol* **66**: 75-111.

Venne AS, Solari FA, Faden F, Paretti T, Dissmeyer N, Zahedi RP. 2015. An improved workflow for quantitative N-terminal charge-based fractional diagonal chromatography (ChaFRADIC) to study proteolytic events in *Arabidopsis thaliana*. *Proteomics* **15**(14): 2458-2469.

Venne AS, Vogtle FN, Meisinger C, Sickmann A, Zahedi RP. 2013. Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. *J Proteome Res* **12**(9): 3823-3830.

Zhang H, Deery MJ, Gannon L, Powers SJ, Lilley KS, Theodoulou FL. 2015. Quantitative proteomics analysis of the Arg/N-end rule pathway of targeted degradation in *Arabidopsis* roots. *Proteomics* **15**(14): 2447-2457.

Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR, 3rd. 2013. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **113**(4): 2343-2394.

Zimmermann D, Gomez-Barrera JA, Pasule C, Brack-Frick UB, Sieferer E, Nicholson TM, Pfannstiel J, Stintzi A, Schaller A. 2016. Cell Death Control by Matrix Metalloproteinases. *Plant Physiol* **171**(2): 1456-1469.

Figure legends

Figure 1. Enrichment and identification of protease generated neo-N termini by negative selection. In a first step proteins are extracted from two proteomes exposed to differential protease activity, generating specific protease-generated neo N-termini in one condition. Next, all primary α-and ε-amines are differentially stable isotope labeled at the protein level. Samples are pooled and digested with a sequence-specific protease such as trypsin. This generates new primary a-amines at the N-termini of internal and C-terminal peptides. In the Terminal Amine Isotope Labeling of Substrates (TAILS) procedure, these are covalently captured with an aldehyde-functionalized polymer that is removed from the peptide mixture by filtration, leaving only unreactive blocked N-terminal peptides in the flow-through. Alternatively, peptides are separated into several fractions using reverse-phase (RP) or strong cation exchange (SCX) in the COmbined FRActional DIagonal Chromatography (COFRADIC) and Charged-based FRActional DIagonal Chromatography (ChaFRADIC) procedures, respectively. Peptides in each fraction are derivatized to alter the retention time of primary-amine containing peptides in a second identical chromatography step. Each fraction is subjected to repeated chromatographic separation, where only unreactive, previously blocked N-terminal peptides elute at the same retention time and are collected. N-terminal peptides are analyzed by tandem mass spectrometry (MS/MS) and identified by semi-tryptic searches. Relative quantification based on the a stable isotope label reveals N termini not affected by modulation of the protease activity with approximately equal abundance in both conditions, whereas neo-N termini appear strongly enriched or only in the condition with higher activity of the protease of interest.

Figure 2. Identification of plant protease substrates and function assisted by quantitative proteomics. Plant proteomes are exposed to different levels of proteolytic activity, either by modulation of a protease of interest in vivo or by incubation with of isolated proteomes with recombinant protease in vitro. Quantitative proteomics is used to determine differences in protein abundance, while termini enrichment techniques such as Terminal Amine Isotope Labeling of Substrates (TAILS) and Charged-based/COmbined FRActional DIagonal Chromatography (ChaFRADIC/COFRADIC) identify cleaved proteins with their precise cleavage sites. These and other approaches such as a 2D-Electrophoretic Mobility Shift Assay (2d-EMSA) identify candidate substrates that require further validation. Knowledge of

sequence specificity of a recombinant protease of interest, for example determined from proteome-derived peptide libraries by Proteomic Identification of Cleavage Sites (PICS), helps to select more likely direct candidate substrates from in vivo termini analyses. Protease-substrates relationships are further validated by proof of co-expression, co-localization, interaction, and proteolytic activity under physiological conditions. Finally, the relevance of substrate cleavage is tested in planta, e.g. using plants mutated in protease activity and/or substrate cleavage site.

Box 1: Glossary

2D-DIGE: 2- Dimensional Differential fluorescence Gel Electrophoresis, a gel-based comparative proteomics method based on protein-level labeling of primary amines with fluorescent tags

ChaFRADIC/COFRADIC: Charged-based/COmbined FRActional DIagonal Chromatography, two chromatography-based methods for protein N termini enrichment

Degradomics: High throughput OMICS techniques applied to proteolysis research

Degradome: Complete set of proteoforms produced by proteases in a given system

iTRAQ: isobaric Tags for Relative and Absolute Quantification, amine-reactive isobaric stable isotope labeling reagents for relative quantification of up to 8 samples simultaneously

Neo termini: New polypeptide chain termini produced by proteolytic processing

PICS: Proteomic Identification of protease Cleavage Sites, a method for protease sequence specificity profiling from proteome-derived peptide libraries

Proteoforms: Related, but different proteins arising from a single gene, e.g. from genetic variations, alternative transcripts or post-translational modification including proteolytic processing

Reductive dimethylation: Chemical modification of primary amines with stable formaldehyde isotopes and sodium cyanoborohydride or –deuteride, resulting in the addition of two methyl groups

Shotgun proteomics: Bottom-up analysis of in solution enzyme digested proteomes

SILAC: Stable Isotope Labeling with Amino acids in Cell culture, a method for metabolic stable isotope labeling

TAILS: Terminal Amine Isotope Labeling of Substrates, a polymer-based method for protein N termini enrichment

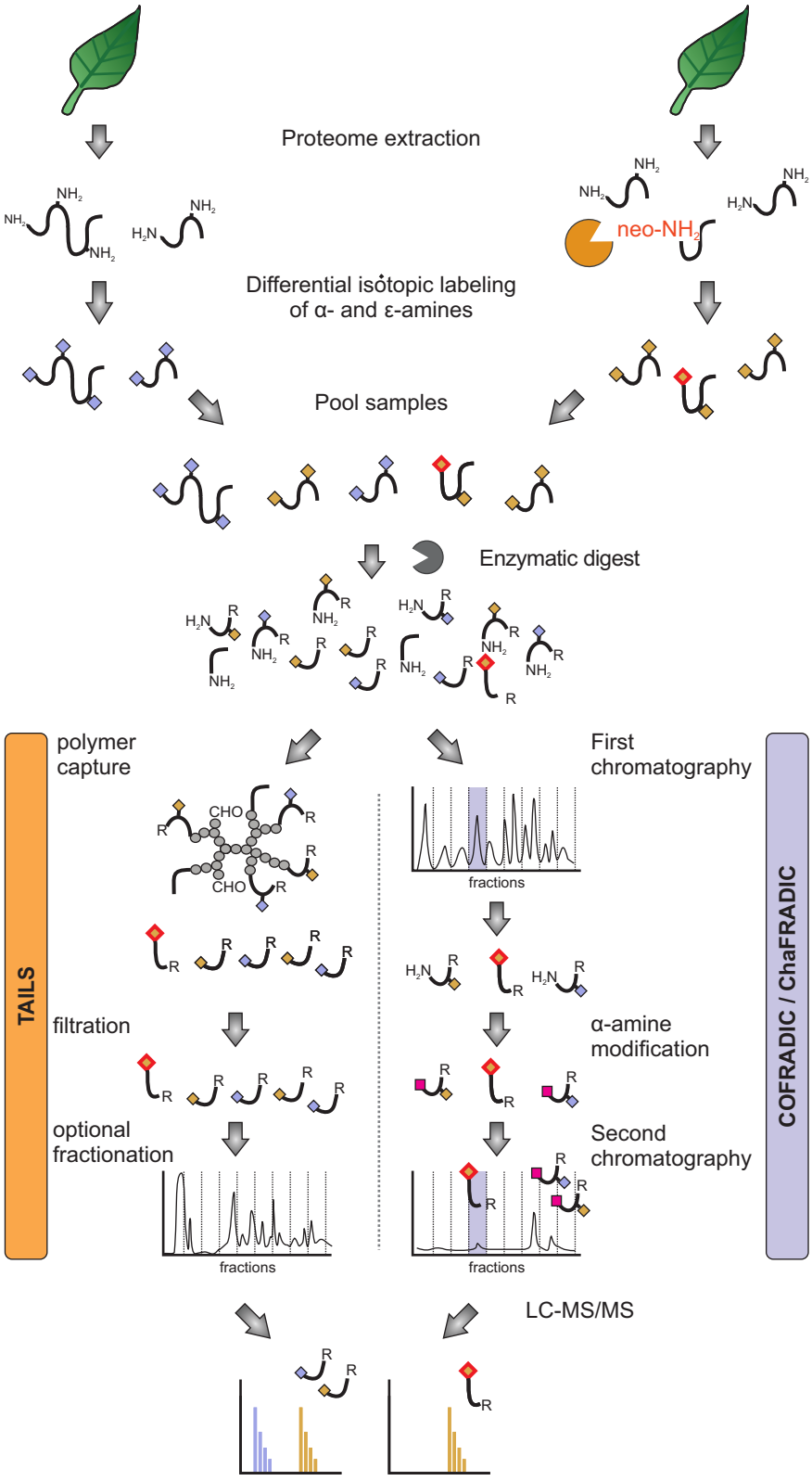
Terminome: All protein N and C termini in a sample

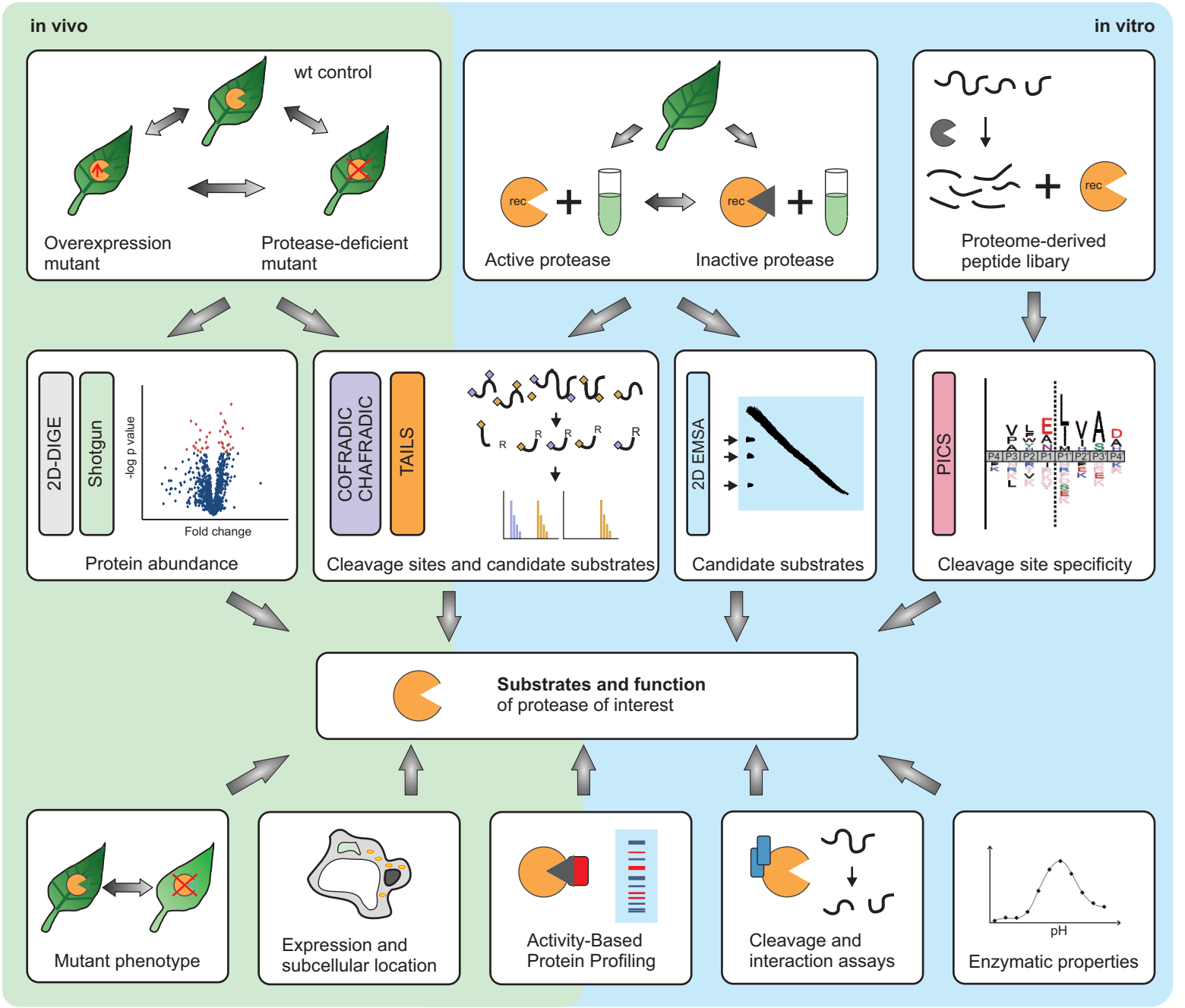
TMT: Tandem Mass Tags, amine-reactive isobaric stable isotope labeling reagents for relative quantification of up to 10 samples simultaneously

Box 2: Bottom-up proteomics – a primer

In bottom-up proteomics, proteomes are enzymatically digested with sequence-specific proteases such as trypsin to yield peptides amenable to tandem MS analysis (Aebersold & Mann, 2016). Digestion is performed either in excised gel pieces after protein-level fractionation by 1D- or 2D- gel electrophoresis, or in the "shotgun" approach directly in solution. Peptides are typically analyzed by nano-LC-MS/MS in data-dependent acquisition mode: MS1 spectra are recorded throughout the chromatography gradient and from each spectrum an instrument-dependent number of peptide precursor ions are subsequently fragmented. The resulting MS2 spectra are matched to peptide sequences computationally predicted from proteome database entries, using e.g. precursor mass and digestion protease specificity as constraints. Neo-terminal peptides generated by endogenous proteolysis are "semi-specific", i.e. the specificity of the digestion protease delimits only one side of the peptide, resulting in a vastly increased number of theoretical peptides that impede identification.

Relative quantification allows comparison of proteomes from different genotypes, stress conditions or time points (Zhang *et al.*, 2013). Label free approaches use independent MS experiments for each condition and determine peptide abundance by integration of their MS1 intensity over time or use spectra counts as proxy. Alternatively, proteins or peptides are differentially labeled using stable isotope reagents. Commonly used chemical isotope labels are ^{13}C- and/or deuterated formaldehyde isotopes for MS1 quantification and isobaric reagents such as iTRAQ or TMT that release specific reporter ions during fragmentation. Stable isotopes can also be introduced metabolically, e.g. by SILAC or growing plants on ^{14}N or ^{15}N-enriched media.


Proteome extraction
Differential isotopic labeling of α- and ε-amines
neo-NH2
Pool samples
Enzymatic digest
TAILS
polymer capture
filtration
optional fractionation
fractions
First chromatography
fractions
α-amine modification
Second chromatography
fractions
COFRADIC / ChaFRADIC
LC-MS/MS


in vivo
in vitro
wt control
Overexpression mutant
Protease-deficient mutant
rec
Active protease
Inactive protease
Proteome-derived peptide libary
2D-DIGE
Shotgun
-log p value
Fold change
Protein abundance
COFRADIC
CHAFRADIC
TAILS
Cleavage sites and candidate substrates
2D EMSA
Candidate substrates
PICS
P4 P3 P2 P1 P1 P2 P3 P4
Cleavage site specificity
Substrates and function
of protease of interest
Mutant phenotype
Expression and subcellular location
Activity-Based Protein Profiling
Cleavage and interaction assays
pH
Enzymatic properties