

The multifunctionality of *Arabidopsis* metacaspases highlighted by their degradomes Styliani Tsiatsiani

Styliani Tsiatsiani

The multifunctionality of *Arabidopsis* metacaspases highlighted by their degradomes

Ghent, October 2012



Ghent University
Faculty of Sciences
Department of Plant Biotechnology and Bioinformatics

The multifunctionality of *Arabidopsis* metacaspases highlighted by their degradomes

Styliani Tsiatsiani

Thesis submitted in partial fulfilment of the requirements
for the degree of Doctor (PhD) in Sciences, Biotechnology

Academic year: **2012-2013**

Promoter:

Prof. Dr. Frank Van Breusegem

Faculty of Sciences, Department of Plant Biotechnology
and Bioinformatics, Ghent University, Belgium

VIB Department of Plant Systems Biology
Oxidative Stress and Cell Death Group
Technologiepark 927, 9052, Belgium

Co-promoter:

Prof. Dr. Kris Gevaert

Faculty of Medicine and Health sciences, Department
of Biochemistry, Ghent University, Belgium

VIB Department of Medical Protein Research
Proteome Analysis and Bioinformatics Unit
Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

*Στην οικογένεια μου και στον Βίμ
για την αστείρευτη αγάπη και υποστήριξη.
Στον παππού μου για το όραμα.*

In the cover:

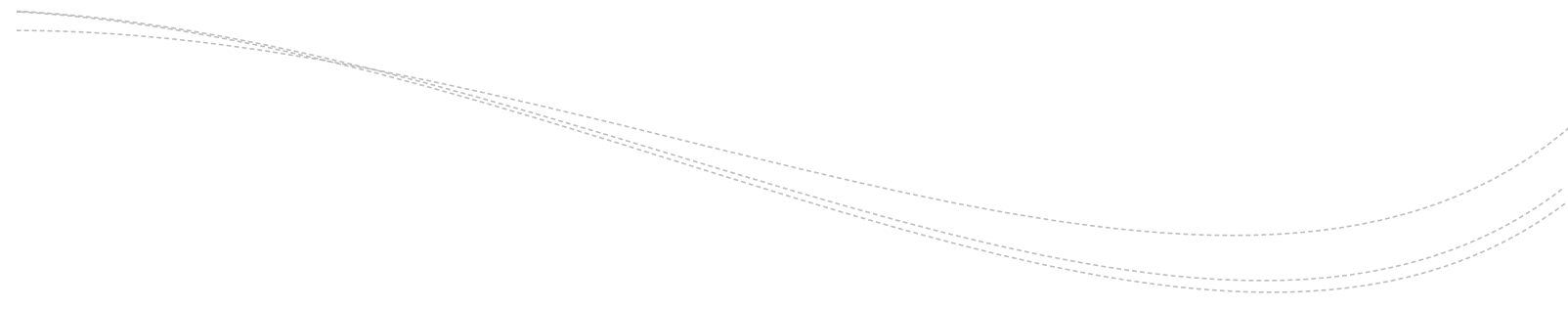
Developing lateral root primordia in close proximity
as observed in the AtMC9 knockout line (GK-540H0).

Graphic design of the thesis book was done by Sofia Tsiatsiani.

LT is indebted to VIB International PhD Program for a predoctoral fellowship.

This work was supported by Research Foundation-Flanders (grant number G.0038.09N) and Ghent University
(Multidisciplinary Research Partnership “Biotechnology for a Sustainable Economy” project).





Examination committee

Prof. Dr. Ann Depicker (Chair)

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Prof. Dr. Frank Van Breusegem (Promoter)

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Prof. Dr. Kris Gevaert (Co-promoter)

Faculty of Medicine and Health sciences, Department of Biochemistry, Ghent University, Belgium

Prof. Dr. Tom Beeckman

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Prof. Dr. Bart Devreese *

Faculty of Sciences, Department of Biochemistry and Microbiology, Laboratory for Protein Biochemistry and Biomolecular Engineering L-ProBe, Ghent University, Belgium

Prof. Dr. Sven Eyckerman *

Faculty of Medicine and Health sciences, Department of Biochemistry, Ghent University, Belgium

Prof. Dr. Pierre Hilson

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Prof. Dr. Kris Laukens *

Faculty of Sciences, Department of Mathematics and Computer Science, Intelligent Systems Lab, University of Antwerp, Belgium

Dr. Moritz Nowack

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

Prof. Dr. Hannele Tuominen *

Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Sweden

Prof. Dr. Petra Van Damme

Faculty of Medicine and Health sciences, Department of Biochemistry, Ghent University, Belgium

Dr. Dominique Vercammen

Faculty of Sciences, Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

* Members of the reading committee

Confidentiality statement concerning PhD thesis “The multifunctionality of Arabidopsis metacaspases highlighted by their degradomes” by Styliani Tsiatsiani under the promotorship of Prof. Dr. Frank Van Breusegem and Prof. Dr. Kris Gevaert.

The present document contains results that may lead to patent applications. Therefore the content of this document is to be treated as confidential and should not be disclosed in a way that may compromise possible patent applications. The document can only be passed on to or discussed with third parties after written consent of the promoters.



Table of contents

List of abbreviations		009
Scope of the research		011
Summary		013
Introduction		019
Chapter 1	Natural substrates of plant proteases: how can protease degradomics extend our knowledge?	021
Chapter 2	Metacaspases	045
Research results		067
Chapter 3	Identification of <i>Arabidopsis</i> METACASPASE 9 physiological substrates reveals its role in regulating the activity of PHOSPHOENOLPYRUVATE CARBOXYKINASE 1	069
Chapter 4	<i>Arabidopsis thaliana</i> METACASPASES are involved in cell separation during lateral root development	153
Chapter 5	<i>Arabidopsis</i> METACASPASE 1 substrate processing during <i>Pseudomonas syringae</i> - triggered cell death	201
Conclusions and Perspectives		227
Acknowledgements		237
Curriculum Vitae		241

List of abbreviations

2D	two-dimensional
2NA	β -naphthylamide
AMC	7-amino-4-methylcoumarin
Boc	tert-Butoxycarbonyl
CHCL TFA	chloromethyl trifluoro acetate
Clp	caseinolytic protease
Cmk	chloromethyl ketone
COFRADIC	combined fractional diagonal chromatography
CPY	carboxypeptidase Y
CtpA	carboxyl-terminal processing protease
CysEP	cysteine endopeptidase
DegP	degradation of periplasmic proteins
ER	endoplasmic reticulum
FB1	fumonisin B-1
fmk	fluoromethylketone
FtsH	filamentation temperature-sensitive H
GFP	green fluorescent protein
GUS	β -glucuronidase
LC	liquid chromatography
MCA	4-methylcoumarin-7-amide
MPP	mitochondria-processing peptidase
MS	mass spectrometry
NHS	N-hydroxysuccinimide
PAGE	polyacrylamide gel electrophoresis
PCD	programmed cell death
PME	pectin methylesterase
PrePs	presequence proteases
PROTOMAP	protein topography and migration analysis platform
PSII	photosystem II
RP-HPLC	reverse-phase high-performance liquid chromatography
RT-PCR	reverse transcription polymerase chain reaction
SCX	strong cation exchange
SDS	sodium dodecyl sulphate
SN	staphylococcal nuclease-like
SPP	signal-processing peptidase
SUMO	small ubiquitin-like modifier
TAILS	terminal amino isotopic labeling of substrate
TLCK	tosyl-lysyl-chloromethylketone
tmbk	trimethylbenzoyloxymethyl ketone
TPP	thylakoid-processing peptidase
TSN	tudor staphylococcal nuclease
Ub	ubiquitin
VPE	vacuolar processing enzyme

Scope of the research

Proteolysis is an irreversible, protease-catalyzed protein modification. Given that proteases transmit molecular signals by cleaving their physiological substrates and thus altering the function of the latter, identification of these substrates is essential for deciphering protease functions (Turk et al., 2012). Proteases mediate protein stability, turnover and activity and also assist in correct targeting and import of proteins into different cell organelles. In this manner, proteases are key instruments for cell and organismal fate. During this PhD project we studied a family of proteases that has attracted the interest of the plant research community since their discovery, 12 years ago (Uren et al., 2000). These metacaspases are cysteine proteases found in plants, fungi and protozoa, and whose catalytic domain is related to the metazoan caspases. Following the paradigm of caspase involvement in apoptosis and inflammation, it was shown that metacaspases are, among other processes, involved in cell death in fungi, protozoa and plants. Although the original incentive for their study was to qualify metacaspases as being responsible for the caspase-like activities observed in plants, it was unambiguously shown that metacaspases are not caspases (Enoksson and Salvesen, 2010).

To date metacaspase functions in plants concern cell death events related to biotic and abiotic responses and development. At the onset of the second decade in metacaspase research, the multifunctionality, redundancy, antagonism and regulation of these proteases has started to become apparent and until now, only a handful of substrates are identified. To understand metacaspase biological functions, detailed description of their specificity, activity and substrates is necessary. For this, our lab performed the pioneering study of one of the nine *Arabidopsis* metacaspases, metacaspase 9 (AtMC9), and determined its specificity and inhibitors (Vercammen et al., 2004 and 2006). Following this, the focus of this PhD study was on the substrate identification of selected *Arabidopsis* metacaspases and elucidation of their involvement in plant physiology.

Although proteases occupy about 2.5% of the *Arabidopsis* genome, to date only a little substrates have been identified. For this, we believe that the research performed in this PhD study will improve significantly the knowledge of the plant research community concerning dedicated protease functions and their mode of action.

Enoksson, M., and Salvesen, G.S. (2010). Metacaspases are not caspases—always doubt. *Cell Death Differ* 17, 1221.

Turk, B., Turk du, S.A., and Turk, V. (2012). Protease signalling: the cutting edge. *Embo J* 31, 1630-1643.

Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 6, 961-967.

Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.A., De Rycke, R., Brackenier, A., Inze, D., Harris, J.L., and Van Breusegem, F. (2006). Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* 364, 625-636.

Vercammen, D., van de Cotte, B., De Jaeger, G., Eeckhout, D., Casteels, P., Vandepoele, K., Vandenberghe, I., Van Beeumen, J., Inze, D., and Van Breusegem, F. (2004). Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. *J Biol Chem* 279, 45329-45336.

Summary

013

Depending on the species, protease genes correspond to 1–5% of all protein-encoding genes in a genome (Rawlings et al., 2010). Although proteases were originally considered as just protein destructors, they are currently much more seen as fine regulators of several biological processes and signalling pathways (Salvesen and Dixit, 1997; Turk et al., 2012). Proteases drive the irreversible cleavage of proteins and can thereby alter their function. In this manner, proteases are critical for controlling immune responses, cell cycle, cell death and protein turnover, amongst others. In plants, proteases are involved in xylem formation, seed maturation, nutrient supply mobilization, adaptation to biotic and abiotic stimuli and senescence (Avci et al., 2008; van der Hoorn, 2008; Pesquet et al., 2012). Because the primary function of proteases is to cleave protein substrates, determination of these substrates is essential to understand the physiological role of individual proteases.

The research performed in this PhD study aimed at the elucidation of metacaspases functions in *Arabidopsis thaliana*. For this reason, we sought to identify their protein substrates at a proteome-wide scale. Metacaspases are like caspases and paracaspases, cysteine proteases and share similar catalytic domains (Uren et al., 2000). *Arabidopsis thaliana* encodes nine metacaspase genes (The Arabidopsis Information Resource, release 10). Based on the presence or absence of an N-terminal prodomain, which contains a plant-specific LSD1-like (Lesion simulating disease 1) zinc-finger motif, *Arabidopsis* metacaspases are distinguished as type-I and type-II metacaspases, respectively.

Two metacaspases, a type-I and a type-II, were assayed in this PhD thesis for different reasons and an overview of the research results is depicted in figure 1. The degradome of AtMC9 (i.e. substrate repertoire) (Overall et al., 2004) in 2 days-old seedlings and roots was explored because AtMC9 is well biochemically characterised. Furthermore, its high gene expression in developing seeds (chapter 3) and exceptionally specific root expression (chapter 4) stirred our interest. Concerning the latter, *AtMC9* intriguing root expression is specific to sites where cell separation occurs during root development (Roberts et al., 2002). These are sites of lateral root protrusion and the root tip where cells detach from the root. The substrate degradome of AtMC1 (chapter 5), the here studied type-I metacaspase, was explored given the AtMC1 function in positively regulating pathogen-triggered hypersensitive response (HR) cell death in *Arabidopsis* (Coll et al., 2010). Finally, the specificity of AtMC9 was further resolved, while the specificity of AtMC1 could not be comprehensively assessed in our studies due to the lack of recombinant AtMC1 protease.

N-terminal COFRADIC is a technique by which peptides holding protein N-termini (N-terminome) are enriched prior to LC-MS/MS analysis (Gevaert et al., 2003; Staes et al., 2011). The N-terminome consists of all peptides, in a given proteome, which contain either mature protein N-termini or neo-N-termini that are generated upon proteolysis. Neo-N-terminal peptides are messengers of proteolytic events and indicate the exact site where protein cleavage has occurred (Van Damme et al., 2005). In this PhD study, by comparing the N-terminomes of plants that either overexpress or lack a metacaspase to wild type plants, we were able to catalogue proteins that were differentially cleaved by the studied metacaspase or possibly by another protease whose activity is metacaspase-dependent.

In chapter 3, the known specificity of AtMC9 for arginine (Arg) and lysine (Lys) (Vercammen et al., 2004) guided us for the selection of potential AtMC9 physiological substrates in germinating seeds/young seedlings. Furthermore, because recombinant AtMC9 is available and active, we were able to couple our *in vivo* degradome results to an *in vitro* substrate catalogue that we generated after the comparison of the *AtMC9* loss-of-function N-terminome to

one spiked with recombinant AtMC9. In this way, we validated some of our previously *in vivo* identified AtMC9 candidate substrates. We selected 74 potential AtMC9 protein substrates that were cleaved in at least two, out of our three, independent analyses at the exact same sites. For a random selection of candidate proteins, cleavage was confirmed using synthetic peptides and *in vitro* transcribed and translated proteins. AtMC9-mediated cleavage of phosphoenolpyruvate carboxykinase 1 (PEPCK1, AT4G37870) was confirmed *in planta* and cleavage was proven beneficiary for PEPCK1 activity. Both proteins localised in the cell cytosol and AtMC9 also localised in the nucleus.

In chapter 4, the intriguing root expression of *AtMC5* and *AtMC9* in sites where cell separation events take place during lateral root development and detachment of root cap cells from the root tip, led us to the investigation of metacaspase involvement in these processes. We identified small molecule chemical compounds that inhibit AtMC9's VRPRase activity *in vitro* and further repress the protrusion of lateral roots and the release of root cap cells. The root phenotypic affect of metacaspase inhibition by chemical means was also reproduced by genetic means. In this way, silencing of multiple metacaspases affected lateral root protrusion similarly to the chemical compounds. Although more metacaspases seem to be involved in this process, and in particular AtMC5 is also expressed in the same sites as AtMC9, we chose to analyse the degradome of AtMC9 because of its specific and high expression in endodermis cells that overlie developing lateral root primordia. In this manner, we catalogued a set of 166 potential substrates, 50 of which are co-expressed with *AtMC9*. Among them are several glycosyl hydrolases, peroxidases and components of the cell cytoskeleton, and their regulation by AtMC9 could lead to alteration of the cell wall properties but this remains to be experimentally demonstrated.

In the last chapter (chapter 5) of this PhD thesis we describe the degradome of the cell death related AtMC1, with respect to *Pseudomonas syringae* pv. tomato (*Pto*) DC3000 (*avrRpm1*) infection. Besides proteins involved in pathogen response, we identified substrates that suggest potential functions for this metacaspase in proteostasis and type-II metacaspase activation. However, further cleavage validation of the here identified potential AtMC1 substrates is necessary in order to associate AtMC1 function to the aforementioned processes.

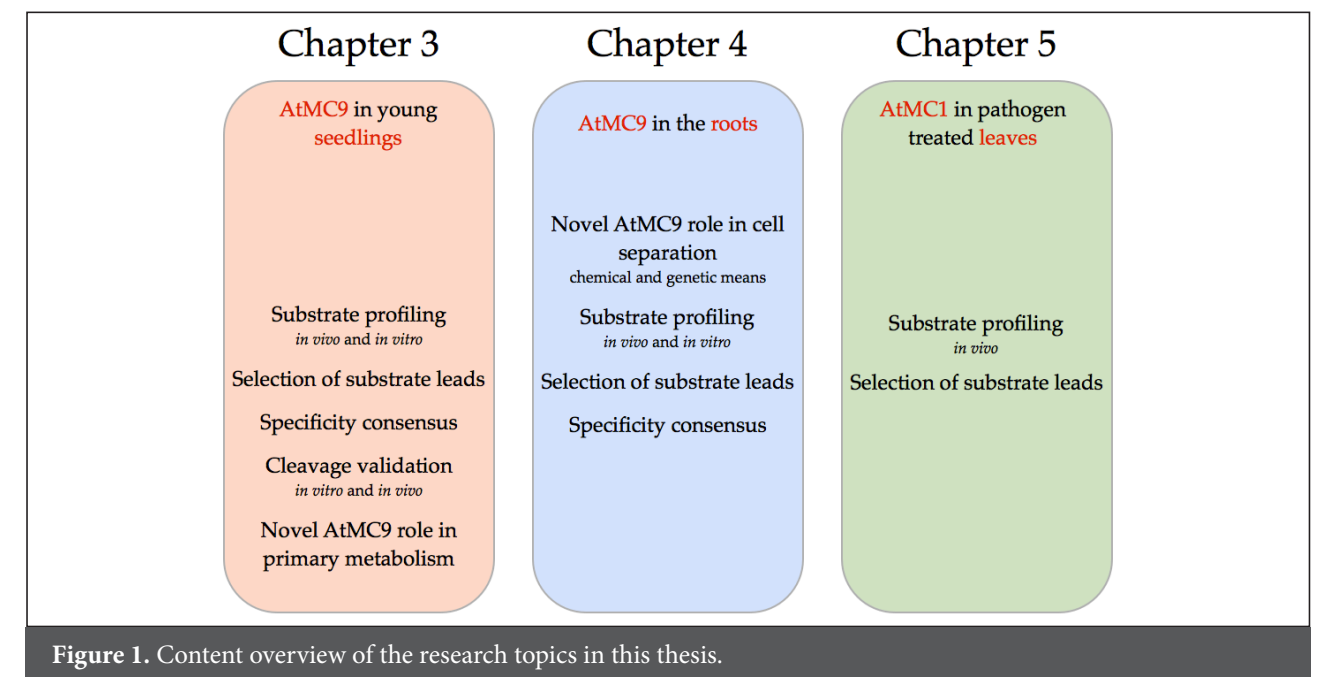


Figure 1. Content overview of the research topics in this thesis.

- Avci, U., Petzold, H.E., Ismail, I.O., Beers, E.P., and Haigler, C.H. (2008). Cysteine proteases XCP1 and XCP2 aid micro-autolysis within the intact central vacuole during xylogenesis in Arabidopsis roots. *Plant J* 56, 303-315.
- Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. *Science* 330, 1393-1397.
- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 21, 566-569.
- Overall, C.M., Tam, E.M., Kappelhoff, R., Connor, A., Ewart, T., Morrison, C.J., Puente, X., Lopez-Otin, C., and Seth, A. (2004). Protease degradomics: mass spectrometry discovery of protease substrates and the CLIP-CHIP, a dedicated DNA microarray of all human proteases and inhibitors. *Biol Chem* 385, 493-504.
- Pesquet, E. (2012). Plant proteases - from detection to function. *Physiol Plant* 145, 1-4.
- Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res* 38, D227-233.
- Roberts, J.A., Elliott, K.A., and Gonzalez-Carranza, Z.H. (2002). Abscission, dehiscence, and other cell separation processes. *Annu Rev Plant Biol* 53, 131-158.
- Salvesen, G.S., and Dixit, V.M. (1997). Caspases: intracellular signaling by proteolysis. *Cell* 91, 443-446.
- Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* 6, 1130-1141.
- Turk, B., Turk du, S.A., and Turk, V. (2012). Protease signalling: the cutting edge. *Embo J* 31, 1630-1643.
- Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 6, 961-967.
- Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005). Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat Methods* 2, 771-777.
- van der Hoorn, R.A. (2008). Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* 59, 191-223.
- Vercammen, D., van de Cotte, B., De Jaeger, G., Eeckhout, D., Casteels, P., Vandepoele, K., Vandenberghe, I., Van Beeumen, J., Inze, D., and Van Breusegem, F. (2004). Type II metacaspases Atmc4 and Atmc9 of Arabidopsis thaliana cleave substrates after arginine and lysine. *J Biol Chem* 279, 45329-45336.

Introduction

019

Chapter 1

Natural substrates of plant proteases: how can protease degradomics extend our knowledge?

Liana Tsiatsiani^{a,b,c,d}, Kris Gevaert^{c,d,*} and Frank Van Breusegem^{a,b,*}

^aDepartment of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Gent, Belgium

^bDepartment of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium

^cDepartment of Medical Protein Research, VIB, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^dDepartment of Biochemistry, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

AUTHOR CONTRIBUTIONS

LT and KG wrote the manuscript with the help of FVB.

Redrafted from:

Tsiatsiani, L., Gevaert, K., and Van Breusegem, F. (2012). Natural substrates of plant proteases: how can protease degradomics extend our knowledge? *Physiol Plant* 145, 28-40.

ABSTRACT

Despite the key role of proteolysis in various intensively studied biological processes, such as plant immunity, seed development, and abiotic stress responses, our knowledge on the identity of natural protease substrates in plants remains scarce. In the genome of the model plant *Arabidopsis thaliana*, for instance, approximately 700 genes code for proteases. However, only a few natural substrates have been identified, mainly due to the previous lack of sensitive proteomics technologies enabling the identification of low abundant proteins, together with a delay in the implementation of these technologies in the field of plant research. Here, we review the current knowledge on the identity of natural plant protease substrates and describe recently established degradomics technologies that should allow proteome-wide studies of plant proteases in the near future.

Key words: plant proteases, natural substrates, degradomics, proteomics

INTRODUCTION

Proteases hydrolyze peptide bonds between amino acids that lead to the fragmentation of their target proteins. These enzymes may act on terminal or internal amino acids of proteins and are, therefore, called exo- or endopeptidases, respectively. Proteolysis is a prominent post- and co-translational modification with diverse functional implications. It is employed either to destruct proteins (protein catabolism), to mature precursor proteins (limited and very often specific) or to remove methionine in the early translation phase by methionine aminopeptidases after new proteins are synthesized (Jackson and Hunter 1970). Proteases govern several important processes at the cell and tissue levels during growth and development of organisms and during initiation and execution of their end (van der Hoorn 2008, Kato and Sakamoto 2010, Nixon et al. 2010, Coll et al. 2011). Particularly in plants, proteases have been shown to be involved in xylem formation, seed maturation, nutrient supply mobilization, removal of signal peptides from several nucleus-encoded organellar proteins, turnover of damaged proteins, responses to environmental stimuli and virulence factors and senescence. The MEROPS database (release 9.4) (Rawlings et al. 2010) includes 695 known and putative *Arabidopsis thaliana* proteases, corresponding to approximately 2.5% of all *Arabidopsis* protein-coding genes (The Arabidopsis Information Resource, release 10). These proteases are distributed over 62 families that belong to 30 different clans of which the largest is that of the serine proteases, followed by aspartate, cysteine, metallo and threonine proteases (van der Hoorn 2008). Despite the abundance and important implication of proteases in plant biology, only a few natural substrates have been identified to date.

In contrast to the biomedical field, in which a wealth of substrate information is available on proteases, such as caspases, cathepsins, granzymes and metalloproteases (Lüthi and Martin 2007, Brix et al. 2008, Demon et al. 2009, Plasman et al. 2011), far less is known about plant proteases. Over the last years, proteolysis has gained ground in the plant research field and an increasing number of laboratories are actively engaged in the study of the mode of action, biochemical properties, putative function and natural substrates of proteases. Usually, most of these efforts have been steered by mutant phenotypes caused by gain-of-function or loss-of-function mutations. In subsequent functional studies, protease purification preceded the biochemical characterization, including assessment of the protease substrate specificity, by means of synthetic peptide-based approaches or by identification of natural substrates (Hara-Nishinura et al. 1991, Ståhl et al. 2002). Methods used comprise yeast 2-hybrid screens and two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE), 2D difference in-gel electrophoresis and diagonal gel electrophoresis combined with in-gel proteolytic digestion. Recently, so-called gel-free proteomics methods have emerged, such as combined fractional diagonal chromatography (COFRADIC; Gevaert et al. 2003, Staes et al. 2008) and terminal amine isotopic labelling of substrate (TAILS; Kleifeld et al. 2010), which overcome several pitfalls introduced by sodium dodecyl sulphate (SDS)-PAGE.

Here, we review the current knowledge on natural substrates for each family of plant proteases (Table 1 and Figure 1) and we describe the recent advances in proteomics technologies that are suitable for proteome-wide protease substrate identification. A short summary is given on the proteases involved in signal peptide or polypeptide tag removal, but for more detailed insights, the reader is referred to recent reports (Richter et al. 2005, Zybailov et al. 2008, Huang et al. 2009; Shipman and Inoue 2009). Proteases for which natural substrates have not been identified yet or proteases that act on pathogen derived substrates are outside the scope of this review and, hence, will not be discussed (van der Hoorn and Jones 2004; Chichkova et al. 2010).

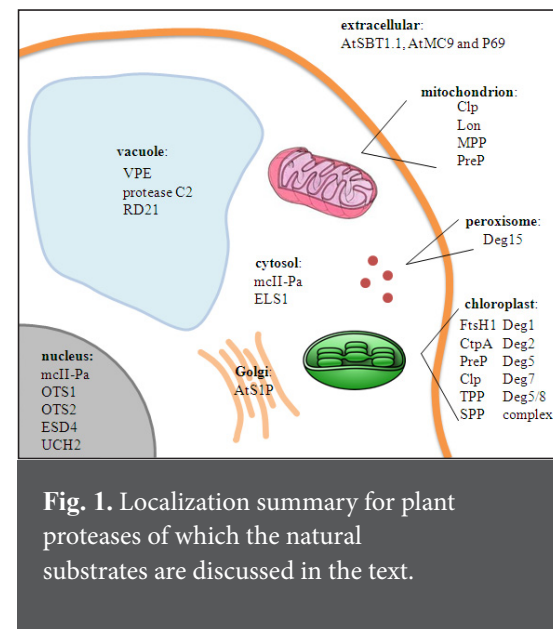
Vacuolar processing enzymes

Vacuolar processing enzymes (VPEs), also known as legumains, are cysteine-dependent proteases that share structural properties with mammalian caspases and belong to the clan CD, family C13. Although VPEs have been characterized first as asparaginyl endopeptidases (Abe et al. 1993), VPEs of castor bean (*Ricinus communis*), tobacco (*Nicotiana benthamiana*), field poppy (*Papaver rhoeas*) and vetch (*Vicia sativa*) can also cleave after aspartic acid residues (Becker et al. 1995, Hiraiwa et al. 1999, Bosch et al. 2010). VPE proteins are synthesized as inactive precursors and are self-catalytically activated under the acidic conditions of the vacuole (Kuroyanagi et al. 2002). However, no processing is required for activation of a VPE during self-incompatibility in poppy pollen (Bosch et al. 2010). Based on their sequence homology and expression patterns, VPEs are classified into seed, seed coat and vegetative types (Ariizumi et al. 2011). Seed VPEs, such as β VPE of *Arabidopsis* and VPE of castor bean, are found in the protein storage vacuoles of seeds where they are involved in the maturation of seed storage preproteins (Shimada et al. 2003). The representative seed coat type VPE of *Arabidopsis*, δ VPE, is involved in cell death during seed coat formation (Nakaune et al. 2005).

Vegetative VPEs, such as α and γ VPE from *Arabidopsis*, proteinase B from vetch and NbVPE-1a and NbVPE-1b from tobacco, reside in the lytic vacuoles of vegetative organs and their expression is rapidly upregulated during hypersensitive cell death in tobacco and various abiotic stresses, including senescence in *Arabidopsis* (Hara-Nishimura et al. 1998, Kinoshita et al. 1999, Hatsugai et al. 2004). In tobacco, VPE plays an important role in the elicitor-triggered immunity (Zhang et al. 2010), while the tomato (*Solanum lycopersicum*) SIVPE5 is involved in sugar accumulation during fruit development (Ariizumi et al. 2011).

Initial knowledge on VPE targets has been obtained for the seed-type VPEs; purified proteases from castor bean seeds converted proglubulin into mature 11S globulin (Hara-Nishimura et al. 1991). After comparative analysis of seed proteomes from several *Arabidopsis* loss-of-function VPE mutants, VPEs have been found accountable for the maturation of other seed storage proteins, such as the 2S albumins and the 12S globulins (Shimada et al. 2003). However, because seed storage proteins remain processed in the complete VPE null mutant of *Arabidopsis* (VPE α - δ), VPEs might not be the sole responsible enzymes (Gruis et al. 2004).

Ectopic expression of *Arabidopsis* γ VPE in yeast revealed that γ VPE targets yeast carboxypeptidase Y (CPY) (Kinoshita et al. 1999). By analogy to the yeast substrate, Western blot immunodetection of *Arabidopsis* AtCPY and AtFruct4 followed by comparative profiling of wild-type and γ VPE mutant proteomes, indicated that γ VPE directly degrades AtCPY and AtFruct4 and it is very likely involved in the maturation of several hydrolases such as β -glycosidases, α -mannosidases, and α -galactosidases (Rojo et al. 2003).



Deg proteases

This family of proteases was first discovered in *Escherichia coli* and designated after the null mutant phenotypes, degradation of periplasmic proteins (DegP; Strauch and Beckwith 1988), or high temperature requirement A (HtrA; Lipinska et al. 1989). Deg/HtrA serine proteases belong to the S1B subfamily of clan PA, and their catalytic domain features a His-Asp-Ser triad. *Arabidopsis* contains 16 chloroplastic ATP-independent Deg endopeptidases.

The degradation of the D1 reaction-centre protein of photosystem II (PSII) is one of the best studied proteolytic events in plant biology. The D1 protein gets readily degraded and is replaced by newly synthesized D1 proteins after damage by photoinhibition (Kyle et al. 1984). D1 has five transmembrane helices connected with loops at the luminal and stromal sides of the thylakoid membrane (Huesgen et al. 2009). The Deg1 and Deg2 proteases and the Deg5/Deg8 complex are involved in the initiation of this repair cycle by cleavage of photodamaged D1 at the stroma-exposed DE loop (Deg2) and the lumen-exposed CD loop (Deg1 and Deg5/Deg8 complex) (Haußühl et al. 2001, Kapri-Pardes et al. 2007, Sun et al. 2007).

Besides the D1 protein, Deg proteases target other components of the PSII complex in higher plants. Deg5 governs wound-stress-induced degradation of PsbF (Luciński et al. 2011) and Deg1 cleaves the luminal protein plastocyanin and the PsbO subunit *in vitro*, supporting a possible role for Deg1 as a general-purpose protease in the thylakoid lumen (Chassin et al. 2002). However, in addition to its proteolytic function, Deg1 acts also as a chaperone involved in the assembly of PSII through its direct interaction with the PSII reaction centre D2 protein (Sun et al. 2010b). Finally, Deg7 is able to cleave D1, D2, CP43 and CP47 proteins of PSII in thylakoid membrane extracts after high-light treatment *in vitro* (Sun et al. 2010a).

Deg substrates other than the aforementioned PSII components have been identified only for Deg15. This protease resides in the peroxisomes and removes the N-terminal peroxisomal transit signal 2 from several proteins, such as the glyoxysomal malate dehydrogenase, the 3-keto-acyl-CoA thiolase, the long-chain acyl-CoA synthetase 6 and the citrate synthase (Helm et al. 2007, Schuhmann et al. 2008).

FtsH proteases

Filamentation temperature-sensitive H (FtsH) proteases are plastidic or mitochondrial membrane-bound proteins. They belong to the protein family of ATPases associated with a variety of cellular activities (AAA+) and contain an ATPase and a zinc-binding domain that functions as the active site. In *Arabidopsis*, of the twelve FtsH proteases, eight are chloroplastic (FtsH1, FtsH2, FtsH5, FtsH6, FtsH7, FtsH8, FtsH9 and FtsH12), three are mitochondrial (FtsH3, FtsH4 and FtsH10) and FtsH11 is found in both organelles (Kato and Samamoto 2010, Janska et al. 2010). The plastid-localized FtsH members occur in the thylakoid or the inner envelope membranes. In particular, the thylakoid members FtsH1, FtsH2, FtsH5 and FtsH8 are well studied.

Cleavage of the D1 primary fragment has been evidenced *in vitro* only for FtsH1 in thylakoids of pea (*Pisum sativum*) (Lindahl et al. 2000). Based on the two-step model for D1 turnover (Haußühl et al. 2001), FtsH proteases act downstream of Deg proteases by performing the secondary proteolysis of Deg-generated D1 fragments. Nevertheless, a role for FtsH at an early stage of D1 degradation and not just in the removal of D1 breakdown products has been established by genetic studies of *FtsH2* and *FtsH5* mutants, in which D1 degradation is impaired (Kato and Sakamoto 2009).

Table 1. Identified natural substrates of plant proteases. Interaction data do not necessarily directly refer to proteolysis events. n/a, not available information; ↓, cleavage site; *, predicted; (hyp), hydroxyproline; #, refer to selected reference; X, any amino acid.

Family	Clan	Species	Protease	Protein substrate	Specificity	Reference
Processing data						
C1	CA	<i>R. communis</i>	CysEP	Premalate dehydrogenase P1-type extensin, tobacco	ANCR↓AKGG T(hyp)VY↓K↓S(hyp)4	Gietl et al. (1997) Helm et al. (2008)
C12	CA	<i>A. thaliana</i>	UCH2	Ubiquitin from conjugates	(Ub)G67 ↓ K(target)	Yang et al. (2004)
C13	CD	<i>R. communis</i>	VPE	Proglobulin and other vacuolar proproteins	ESEN↓GLEE among others	Hara-Nishimura et al. (1991)
		<i>A. thaliana</i>	αVPE and γVPE but mostly βVPE	Pro2s albumin and pro12s globulin	EN↓P	Shimada et al. (2003)
			VPEγ	AtCPY, AtFruct4 and several hydrolases therein	n/a	Rojo et al. (2003)
			RD21	AtSerpin1	IKLR↓GLIM	Lamp1 et al. (2010)
C14	CD	<i>A. thaliana</i>	AtMC9	AtSerpin1	IKLR↓GLIM	Vercammen et al. (2006)
		<i>P. abies</i>	mcII-Pa	Tudor staphylococcal nuclease (TSN)	ASIR↓DLPP, VLNR↓DVRI, SNSK↓AIDR, HSAR↓ESPV	Sundström et al. (2009)
C48	CE	<i>A. thaliana</i>	ESD4	De-SUMOylation of proteins	(SUMO)G ↓ K(target)	Murtas et al. (2003), Colby et al. (2006), Hermkes et al. (2011)
			ELS1			Hermkes et al. (2011)
			OTS1 and OTS2			Conti et al. (2008)
S1B	PA	<i>A. thaliana</i>	Deg1, Deg2 and Deg5/8 complex	D1 reaction centre protein of PSII	n/a	Haußühl et al. (2001), Kapri-Pardes et al. (2007), Sun et al. (2007)
			Deg1	Plastocyanin and PsbO subunit of PSII	n/a	Chassin et al. (2002)
			Deg5	PsbF subunit of PSII	n/a	Luciński et al. (2011)
			Deg7	D1, D2, CP43 and CP47 proteins of PSII	n/a	Sun et al. (2010a)
			Deg15	PTS2 removal from: 3-keto-acyl-CoA thiolase long-chain acyl-CoA synthetase 6, citrate synthase glyoxysomal malate dehydrogenase	ASAC↓LAGD* SNPT↓AGEF* AHCV↓SAQT* ANCR↓AKGG	Helm et al. (2007), Schuhmann et al. (2008)
S8A	SB	<i>A. thaliana</i>	AtSBT1.1	AtPSK4, phytosulfokine growth factor	SLVL↓HTDY	Srivastana et al. (2008)
			AtSBT6.1	AtbZIP17, transcription factor	RRIL↓RGLP	Liu et al. (2007)
				AtRALF23, peptide growth factor	RRIL↓ATTK	Srivastana et al. (2009)
		<i>S. lycopersicum</i>	P69	leucine-rich-repeat protein	ALRR↓SLSD*	Tornero et al. (1996)
		<i>G. max</i>	Protease C1	β-conglycinin	KGSE↓EEDE, NEEE↓DEDE	Qi et al. (1992)
N/A	N/A		Protease C2	β-conglycinin and β-type phaseolin	SSRK↓TISS (conglycinin) and KSLS↓KQDN (phaseolin), among others	Seo et al. (2001)
S14	SK	<i>A. thaliana</i>	Clp	Proteins therein	#	Stanne et al. (2009), Sjögren et al. (2009), Zybailov et al. (2009)
			CtpA	D1 reaction centre protein of PSII	n/a	Fabbri et al. (2005)
S16	SJ	<i>P. vulgaris</i>	Lon	ORF239	n/a	Sarria et al. (1998)
S49	SK	<i>A. thaliana</i>	SPP	Proteins therein	#	Richter et al. (2005), Zybailov et al. (2008)
S26	SF		TPP	Proteins therein	#	Shipman and Inoue (2009)
M16	ME	<i>A. thaliana</i>	PreP	F1β subunit precursor of ATP synthase	signal peptide degradation	Stahl et al. (2002)
		<i>A. thaliana</i> and <i>O. sativa</i>	MPP	Proteins therein	RX(F/Y/L)↓(S/A)(S/T)	Huang et al. (2009)
M41	MA	<i>A. thaliana</i>	FtsH1	D1 reaction centre protein of PSII	n/a	Lindahl et al. (2000)
		<i>P. sativum</i>	FtsH	Rieske Fe-S	n/a	Ostersetzer and Adam (1997)
Interaction data						
S1B	PA	<i>A. thaliana</i>	Deg1	D2 reaction centre of PSII	n/a	Sun et al. (2010b)
S8A	SB		AtSBT6.1	VGD1 pectin methyltransferase and tobacco pectin methyltransferase	n/a	Wolf et al. (2009)

Biochemical studies showed that the *in vitro*-translated Rieske Fe–S protein of the cytochrome b_6 – f complex was imported into isolated pea chloroplasts where it was gradually degraded. Degradation was inhibited by FtsH antibodies, suggesting that FtsH protease(s) is(are) responsible for Rieske Fe–S protein degradation (Ostersetzer and Adam 1997). Finally, *in vitro* evidence for the degradation of the light-harvesting complex of PSII by FtsH6 (Želisko et al. 2005) has not been confirmed *in vivo*, implying that FtsH6 is not the responsible protease for this phenomenon (Wagner et al. 2011).

Clp proteases

The caseinolytic protease (Clp) machinery consists of two components, namely a protease core and a hexameric ring-like ATP-dependent chaperone subunit that confers substrate specificity and drives the unfolding of protein substrates and their subsequent translocation to the proteolytic core. In contrast to FtsH and Deg proteases, Clps are located in the stroma. The complexity of the Clp system is well documented in bacteria (Katayama-Fujimura et al. 1997). Natural Clp substrates (Sjögren et al. 2006, Stanne et al. 2009, Zybailov et al. 2009) have been identified by means of mass spectrometry (MS)-coupled methods. However, it is still unclear whether these potential substrates are derived from genuine Clp-driven proteolysis or simply from protein upregulation caused by the Clp mutation (Olinars et al. 2011).

Lon proteases

The first ATP-dependent protease purified from *E. coli* (Swamy and Goldberg 1981) was a Lon protease that together with Clp and FtsH proteases belong to the AAA⁺ superfamily of enzymes. The catalytic site of these serine proteases consists of a Ser-Lys dyad, but without relationship to the classical catalytic Ser-His-Asp triad of serine proteases (Botos et al. 2004).

In *Arabidopsis*, four Lon proteases are found. Lon1 is mitochondrial, Lon4 has a dual localization in mitochondria and plastids (Heazlewood et al. 2004, Ostersetzer et al. 2007), Lon2 resides in the peroxisome and Lon3 is suspected to be a pseudogene (Kikuchi et al. 2004, Ostersetzer et al. 2007). So far, only one natural Lon substrate has been identified, the ORF239 from common bean (*Phaseolus vulgaris*) that is associated with cytoplasmic male sterility (Sarria et al. 1998). The *lon1* mutants of *Arabidopsis* exhibit post-germinative growth retardation, accompanied by a decreased activity of respiratory chain complexes and TCA cycle enzymes (Rigas et al. 2009). Whether they are directly cleaved by AtLON1 is still unknown.

Subtilases

Subtilases are subtilisin-like serine proteases, belonging to the clan SB. Plant subtilases occur in the pyrolysins group within the S8A subfamily that consists of a much larger multigenic family (56 members in *Arabidopsis*) than that of their mammalian homologues (nine members). In response to salt stress, the transcription factor AtbZIP17 is cleaved by subtilase AtS1P (or AtSBT6.1), as demonstrated through an *in vitro* pull-down assay with agarose bead-immobilized AtS1P. Subsequently, the N-terminal fragment of AtbZIP17 translocates to the nucleus where it is believed to initiate the salt stress response (Liu et al. 2007). In addition, AtS1P is responsible for the maturation of the rapid alkalisation growth factor AtRALF23 (Srivastana et al. 2009). Another *Arabidopsis* subtilase, AtSBT1.1, is involved in the processing of the peptide hormone AtPSK4 precursor, a

phytosulfokine growth factor, thus promoting cell proliferation during tissue culture. This effect has been observed after cleavage of a synthetic peptide harbouring the predicted cleavage site of AtPSK4. Furthermore, the AtPSK4 myc-tagged protein is not processed in the *SBT1.1* mutant (Srivastana et al. 2008).

The subtilase from soybean (*Glycine max*), protease C1, processes the α and α' subunits of the seed storage protein β -conglycinin (Qi et al. 1992). This protease performs a primary cleavage at five sites whereafter a cysteine protease, protease C2, further degrades β -conglycinin. Protease C2 has been shown to cleave also the homologous β -type phaseolin in common bean (Seo et al. 2001).

The specificity of other subtilases, such as cucumisins, hordolisins and macularisins, has been studied with synthetic peptides. Given their rather broad range specificity, a general role in protein turnover has been attributed to subtilases (Kaneda et al. 1995, Rudenskaya et al. 1995, Terp et al. 2000). However, specialized functions have been deduced after identification of the first natural substrate in tomato and the afore-mentioned *Arabidopsis* substrates. In tomato, subtilase P69 cleaves a leucine-rich repeat cell wall protein of which the expression is up-regulated during virus infection. Therefore, this particular subtilase might play a role in pathogen recognition and signalling initiated at the extracellular space (Tornerio et al. 1996).

Demethylesterification of HGA, the major component of pectin, is catalyzed by a large family of pectin-modifying enzymes, called pectin methylesterases (PMEs). AtS1P has been shown to interact with proPME1 in tobacco and its pollen-expressed *Arabidopsis* isoform VGD1 (Wolf et al. 2009). Therefore, this subtilase might be involved in the processing of PMEs in the Golgi apparatus prior to delivery and activation of PMEs at the cell wall.

KDEL-tailed cysteine endopeptidases (CysEP)

A plant-specific group of papain-type cysteine endopeptidases (CysEP) equipped with a C-terminal KDEL endoplasmic reticulum (ER) retention signal is involved in programmed cell death during endosperm degeneration in castor bean. These proteases are synthesized as prepro-enzymes and transported from the ER to the cytosol of senescing cells during seed maturation as ER-derived ribosome-studded vesicles, called ricinosomes. In *Arabidopsis*, three enzymes have been identified by homology to the castor bean CysEP that is able to cleave glyoxisomal pre-malate dehydrogenase *in vitro* (Gietl et al. 1997). The crystal structure of the castor bean KDEL-tailed CysEP has led to the investigation of its substrate specificity and has been found to degrade hydroxyproline-rich glycoproteins, such as the P1-type extensin from tobacco (Than et al. 2004, Helm et al. 2008).

Metacaspases

On the basis of sequence and structure similarity to caspases, a family of proteases known as metacaspases has been identified in fungi, protozoa and plants (Uren et al. 2000). Paracaspases, the equivalent family in metazoans and the slime mold *Dictyostelium discoideum*, had previously been discovered (Aravind et al. 1999). Like caspases, these two new families of proteases carry the hallmarks of the clan CD of cysteine proteases: the catalytic dyad His-Cys and the hemoglobinase fold (Rawlings and Barrett 1993, Aravind and Koonin 2002). In sharp contrast to the specificity of caspases, that of metacaspases is ruled by a strict preference towards arginine and lysine at position P1. The enzymes are auto-activated upon proteolytic separation of their two subunits and they are classified into two types based on the presence or absence of an

N-terminal prodomain. Interestingly, the reactive centre loop of AtSerpin1, a serine protease inhibitor, acts as bait for both AtMC9 and RD21 cysteine proteases (Vercammen et al. 2006, Lampl et al. 2010). Recently, the Tudor staphylococcal nuclease has been identified as a shared substrate of the Norway spruce (*Picea abies*) metacaspase mcII-Pa and caspase-3 (Sundström et al. 2009).

Signal-processing proteases and polypeptide tag-deconjugating enzymes

Organellar biogenesis and development depend on the proper delivery of nucleus-encoded proteins from the cytoplasm into, for instance, mitochondria and chloroplasts. During this process, signal peptides are proteolytically removed from their protein precursors. In chloroplasts, this removal is mediated by proteases, such as the signal-peptide proteases (SPPs), the thylakoid-processing peptidases (TPPs), the stromal presequence proteases (PrePs) and the carboxyl-terminal processing protease A (CtpA), whereas in mitochondria, PrePs and the mitochondria-processing peptidases (MPPs) carry out the same task. PrePs (Ståhl et al. 2002) are responsible for the *in organello* processing of the F1 β subunit precursor of ATP synthase in mitochondria and CtpA cleaves the C-terminal part of the D1 precursor protein *in vitro* (Fabbri et al. 2005). Several substrates have been identified for SPPs (Richter et al. 2005, Zybailov et al. 2008), TPPs (Shipman and Inoue 2009) and MPPs (Huang et al. 2009).

Post translational modifications alter the physicochemical properties of proteins, modulate their function, and frequently make them prone to degradation. Polypeptide tags, such as ubiquitin (Ub) and the small ubiquitin-like modifier (SUMO), conjugate to lysine residues in target proteins, cause their modification and signal their degradation by the Ub/26S proteasome (reviewed by Fu et al. 2010). Such polypeptide tags are synthesized as pre-proteins that have to be processed by proteases to expose a C-terminal glycine, which will later form a peptide bond with the ϵ -amino group of a lysine residue, located in the substrate protein. Proteases do not only cleave the proforms (protein maturation) of the above mentioned protein tags, but they also remove the mature tags from the targeted proteins. For instance, when recombinantly produced, the *Arabidopsis* Ub-deconjugating enzyme, Ub C-terminal hydrolase 2 (UCH2), is able to release Ub from conjugates *in vitro* and from the poly-Ub precursor AtUBQ10 (Yang et al. 2007). The RPN11 metalloprotease, a core subunit of the 26S proteasome de-ubiquitinates numerous proteins prior their translocation into the 20S core for degradation (Verma et al., 2002). Furthermore, SUMO peptidase activity leading to de-SUMOylation of target proteins has been attributed to the *Arabidopsis* nuclear proteins OVERLY TOLERANT TO SALT1 and 2 (OTS1 and OTS2) (Conti et al. 2008), EARLY IN SHORT DAYS4 (ESD4) and its close homologue ESD4-LIKE SUMO PROTEASE1 (ELS1) (Murtas et al. 2003, Colby et al. 2006, Hermkes et al. 2011).

Proteomics methods for identification of proteases substrates

Processing of protein substrates by proteases creates substrate fragments with reduced molecular weights and altered isoelectric points. As 2D-PAGE separates proteins according to these two physicochemical parameters (O'Farrell 1975), this proteomics technique has been used repeatedly to identify protease substrates (Sjögren et al. 2006, Stanne et al. 2009). Another gel-based technology is diagonal electrophoresis. Here, proteins are separated first by their molecular weight by means of SDS-PAGE in one gel lane. This gel lane is then incubated with a protease of interest, after which the proteins are separated a second time by SDS-PAGE, now

perpendicularly to the first separation. In such a setup, proteins processed by the added protease generate fragments with reduced molecular weights and segregate away from the diagonal that contains the unprocessed proteins, allowing their detection (Shao et al. 2007). More recently, the PRotein TOpography and Migration Analysis Platform (PROTOMAP; Dix et al. 2008) has been introduced during which proteins, including protease substrates and their fragments, are separated by SDS-PAGE. The gel lane is sliced and proteins are in-gel digested. The generated peptides are analyzed by liquid chromatography-MS/MS. The resulting MS/MS spectra are counted, linked to their parent proteins and coupled to the gel migration pattern of the proteins. Thus, protein fragments are recognized by their apparently lower molecular weight (than that of their intact precursor) and their sequence coverage by the MS/MS spectra of their peptides.

Although PROTOMAP is a clear improvement over 2D-PAGE and diagonal electrophoresis, the identification of the exact processed site is not straightforward in all these techniques. Furthermore, whereas with 2D-PAGE mainly readily soluble and abundant proteins are visualized (Gygi et al. 2000), diagonal electrophoresis has as main limitations that protein unfolding, resulting in nonphysiological substrate processing, can occur as proteins are not in their natural environment and that the access of the investigated protease to its substrates might be restricted due to the polyacrylamide maze. For these reasons and given the technological progress in the MS field and the increasing availability of whole-genome sequences, non-gel routes have been explored for the discovery of protease substrates. The most important of these technologies are discussed below.

Novel protein termini are generated through enzymatic hydrolysis by protein processing: a neo-N-terminus at the C-terminal and a neo-C-terminus at the N-terminal protein fragments. Clearly, identification of any of such peptides unmistakably points to a protein processing event and to the actual processed site as well. Until just recently, proteomics technologies focused on the enrichment of N-terminal peptides. In fact, contemporary proteomics relies on (semi-) high throughput analysis of peptides derived from whole-proteome digests by LC coupled to MS/MS (LC-MS/MS). However, given that each protein will yield only one N-terminal peptide, and several non-N-terminal, internal peptides, there is a clear need to enrich for N-terminal peptides prior to LC-MS/MS sampling because they will be outcompeted by the excess of internal peptides. Several such enrichment strategies are used nowadays and can generally be divided into two main categories: negative selection procedures that target the internal peptides and remove them from the final mixtures and positive selection procedures that target the protein N-terminal peptides (recently reviewed by Agard and Wells 2009, Impens et al. 2010a, van Domselaar et al. 2010).

A pioneering negative selection procedure is the so-called combined fractional diagonal chromatography (COFRADIC) procedure that enriches N-terminal peptides (Gevaert et al. 2003). In general, COFRADIC consists of a first separation of peptides based on differences in hydrophobicity using reverse-phase HPLC, followed by a modification step that alters the column retention of specific classes of peptides. Separation of the individual peptide fractions a second time under identical chromatographic conditions allows the segregation of the modified from the non-modified peptides and any of both peptide classes can be collected for LC-MS/MS analysis. As mentioned above, by one COFRADIC application N-terminal peptides can be overrepresented and, thus, neo-N-terminal peptides from proteome digests (Figure 2). This application was applied to an apoptotic cell model of human Jurkat T-lymphocytes and led to the direct identification of 93 processing events in 71 proteins in a differential proteomics setup (Van Damme et al. 2005). By means of this technology, diverse setups

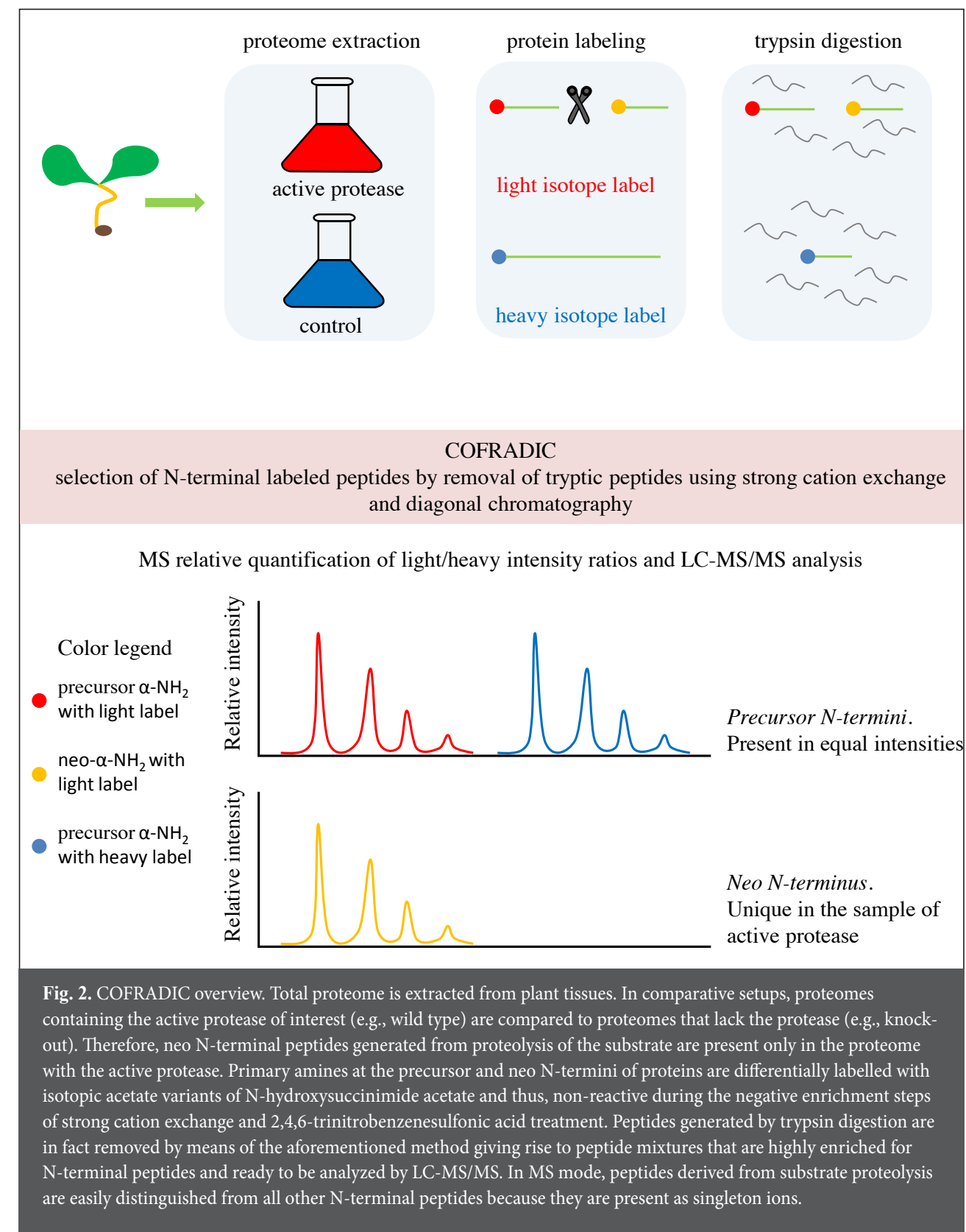
could be used, both in so-called reverse degradomics studies in which an isolated proteome is incubated with a protease of interest, as well as in forward degradomics studies in which cell or tissue samples are analyzed in which proteases are active. Interestingly, until now, emphasis was put on proteases involved in cell death models (recently reviewed in Demon et al. 2009, Impens et al. 2010b). The original COFRADIC protocol has been improved (Staes et al. 2008) and, more recently, a complementary procedure to isolate protein C-termini in addition to protein N-termini has been introduced (Van Damme et al. 2010). The latter is especially appealing because both types of protein terminal peptides have been shown to hold complementary information, thus increasing the proteome coverage.

Other negative selection procedures have been introduced based either on tagging internal peptides with affinity labels or on scavenger-carrying beads for α -amino groups. Examples include biotinylation of the α -amino group of internal peptides after trypsin digestion of proteomes with acetylated primary amino groups (McDonald et al. 2005). Various amino-scavengers immobilized on beads or polymers have been applied to trap internal peptides (Mikami and Takao 2007). Very recently, the terminal amine isotopic labeling of substrates (TAILS) technology has been reported that uses highly water-soluble dendritic polymers to isolate internal peptides (Kleifeld et al. 2010).

Positive selection techniques for protein N-terminal peptides include biotinylation, strong cation exchange chromatography (SCX) at low pH and the use of subtiligase. Specific blocking of ϵ -amines of lysines (by guanidination) with biotinylation of primary α -amino groups has been combined to enrich for protein N-terminal peptides by streptavidin chromatography (Timmer et al. 2007). This method co-enriches neo-N-terminal peptides, indicating protein processing because these peptides generally carry a free, primary α -amine, whereas most, but not all, N-termini of intact proteins from higher eukaryotes are blocked by co-translational α -N-acetylation (Arnesen et al. 2009). Strong cation exchange chromatography at low pH is known to augment the α -amino blocked peptides. This technology has been applied on a proteome-wide scale for enrichment of α -amino blocked N-terminal peptides both after trypsin digestion of a proteome (Dormeyer et al. 2007) and after endoproteinase Lys-N proteome digestion (Taouatas et al. 2009). Finally, the subtiligase approach (Mahrus et al. 2008) enzymatically tags free α -amino groups in proteins and protein fragments with a genetically engineered variant of the subtilisin protease that ligates a biotinylated peptide carrying a TEV recognition site to the N-terminus of a protein.

CONCLUSIONS

The fast progress in genomics and proteomics ensures a better understanding of proteolytic events that underlie several biological processes. Some of the unresolved issues are certainly i) the determination of proteases responsible for observed biological activities of interest, ii) the challenge to overcome functional redundancy between proteases and, thus, to characterise dedicated functions for each of them, and iii) the identification of natural protease substrates and the purpose of their cleavage. Finally, deciphering the biology of proteolysis may create opportunities for practical applications that can improve crop characteristics and performance.



REFERENCES

- Abe Y, Shirane K, Yokosawa H, Matsushita H, Mitta M, Kato I, Ishii S-I (1993) Asparaginyl endopeptidase of jack bean seeds. Purification, characterization, and high utility in protein sequence analysis. *J Biol Chem* 268: 3525-3529
- Agard NJ, Wells JA (2009) Methods for the proteomic identification of protease substrates. *Curr Opin Chem Biol* 13: 503-509
- Aravind L, Koonin LV (2002) Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* 46: 355-367
- Aravind L, Dixit VM, Koonin EV (1999) The domains of death: evolution of the apoptosis machinery. *Trends Biochem Sci* 24: 47-53
- Ariizumi T, Higuchi K, Arakaki S, Sano T, Asamizu E, Ezura H (2011) Genetic suppression analysis in novel vacuolar processing enzymes reveals their roles in controlling sugar accumulation in tomato fruits. *J Exp Bot* 62: 2773-2786
- Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F, Gevaert K (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci USA* 106: 8157-8162
- Becker C, Shutov AD, Nong VH, Senyuk VI, Jung R, Horstmann C, Fischer J, Nielsen NC, Müntz K (1995) Purification, cDNA cloning and characterization of proteinase B, an asparagine-specific endopeptidase from germinating vetch (*Vicia sativa* L.) seeds. *Eur J Biochem* 228: 456-462
- Bosch M, Poulter NS, Perry RM, Wilkins KA, Franklin-Tong VE (2010) Characterization of a legumain/vacuolar processing enzyme and YVADase activity in *Papaver* pollen. *Plant Mol Biol* 74: 381-393
- Botos I, Melnikov EE, Cherry S, Tropea JE, Khalatova AG, Rasulova F, Dauter Z, Maurizi MR, Rotanova TV, Wlodawer A, Gustchina A (2004) The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J Biol Chem* 279: 8140-8148
- Brix K, Dunkhorst A, Mayer K, Jordans S (2008) Cysteine cathepsins: cellular roadmap to different functions. *Biochimie* 90: 194-207
- Chassin Y, Kapri-Pardes E, Sinvany G, Arad T, Adam Z (2002) Expression and characterization of the thylakoid lumen protease DegP1 from *Arabidopsis*. *Plant Physiol* 130: 857-864
- Chichkova NV, Shaw J, Galiullina RA, Drury GE, Tuzhikov AI, Kim SH, Kalkum M, Hong TB, Gorshkova EN, Torrance L, Vartapetian AB, Taliansky M (2010) Phytaspase, a relocatable cell death promoting plant protease with caspase specificity. *EMBO J* 29: 1149-1161
- Colby T, Matthäi A, Boeckelmann A, Stuitable H-P (2006) SUMO-conjugating and SUMO-deconjugating enzymes from *Arabidopsis*. *Plant Physiol* 142: 318-332
- Coll NS, Epple P, Dangl JL (2011) Programmed cell death in the plant immune system. *Cell Death Differ* 18: 1247-1256
- Conti L, Price G, O'Donnell E, Schwessinger B, Dominy P, Sadanandom A (2008) Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in *Arabidopsis*. *Plant Cell* 20: 2894-2908
- Demon D, Van Damme P, Vanden Berghe T, Vandekerckhove J, Declercq W, Gevaert K, Vandenabeele P (2009)

- Caspase substrates: easily caught in deep waters? Trends Biotechnol 27: 680-688
- Dix MM, Simon GM, Cravatt BF (2008) Global mapping of the topography and magnitude of proteolytic events in apoptosis. Cell 134: 679-691
- Dormeyer W, Mohammed S, van Breukelen B, Krijgsveld J, Heck AJR (2007) Targeted analysis of protein termini. J Proteome Res 6: 4634-4645
- Fabbri BJ, Duff SMG, Remsen EE, Chen Y-CS, Anderson JC, CaJacob CA (2005) The carboxyterminal processing protease of D1 protein: expression, purification and enzymology of the recombinant and native spinach proteins. Pest Manag Sci 61: 682-690
- Fu H, Lin Y-L, Fatimababy AS (2010) Proteasomal recognition of ubiquitylated substrates. Trends Plant Sci 15: 375-386
- 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: 566-569
- Gietl C, Wimmer B, Adamec J, Kalousek F (1997) A cysteine endopeptidase isolated from castor bean endosperm microbodies processes the glyoxysomal malate dehydrogenase precursor protein. Plant Physiol 113: 863-871
- Gruis D(F), Schulze J, Jung R (2004) Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. Plant Cell 16: 270-290
- Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci USA 97: 9390-9395
- Hara-Nishimura I, Shimada T, Hatano K, Takeuchi Y, Nishimura M (1998) Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles. Plant Cell 10: 825-836
- 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 Lett 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
- Haußühl K, Andersson B, Adamska I (2001) A chloroplast DegP2 protease performs the primary cleavage of the photodamaged D1 protein in plant photosystem II. EMBO J 20: 713-722
- Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH (2004) Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. Plant Cell 16: 241-256
- Helm M, Lück C, Prestele J, Hierl G, Huesgen PF, Fröhlich T, Arnold GJ, Adamska I, Görg A, Lottspeich F, Gietl C (2007) Dual specificities of the glyoxysomal/peroxisomal processing protease Deg15 in higher plants. Proc Natl Acad Sci USA 104: 11501-11506
- Helm M, Schmid M, Hierl G, Terneus K, Tan L, Lottspeich F, Kieliszewski MJ, Gietl C (2008) KDEL-tailed cysteine endopeptidases involved in programmed cell death, intercalation of new cells, and dismantling of extensin scaffolds. Am J Bot 95: 1049-1062
- Hermkes R, Fu Y-F, Nürrenberg K, Budhiraja R, Schmelzer E, Elrouby N, Dohmen RJ, Bachmair A, Coupland G (2011) Distinct roles for Arabidopsis SUMO protease ESD4 and its closest homolog ELS1. Planta 233: 63-73
- 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 Lett 447: 213-216
- Huang S, Taylor NL, Whelan J, Millar AH (2009) Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol 150: 1272-1285
- Huesgen PF, Schuhmann H, Adamska I (2009) Deg/HtrA proteases as components of a network for photosystem II quality control in chloroplasts and cyanobacteria. Res Microbiol 160: 726-732
- Impens F, Colaert N, Helsen K, Plasman K, Van Damme P, Vandekerckhove J, Gevaert K (2010a) MS-driven protease substrate degradomics. Proteomics 10: 1284-1296
- Impens F, Vandekerckhove J, Gevaert K (2010b) Who gets cut during cell death? Curr Opin Cell Biol 22: 859-864
- Jackson R, Hunter T (1970) Role of methionine in the initiation of haemoglobin synthesis. Nature 227: 672-676
- Janska H, Piechota J, Kwasniak M (2010) ATP-dependent proteases in biogenesis and maintenance of plant mitochondria. Biochim Biophys Acta 1797: 1071-1075
- Kaneda M, Yonezawa H, Uchikoba T (1995) Improved isolation, stability and substrate specificity of cucumisin, a plant serine endopeptidase. Biotechnol Appl Biochem 22: 215-222
- Kapri-Pardes E, Naveh L, Adam Z (2007) The thylakoid lumen protease Deg1 is involved in the repair of photosystem II from photoinhibition in Arabidopsis. Plant Cell 19: 1039-1047
- Katayama-Fujimura Y, Gottesman S, Maurizi MR (1987) A multiple-component, ATP-dependent protease from Escherichia coli. J Biol Chem 262: 4477-4485
- Kato Y, Sakamoto W (2010) New insights into the types and function of proteases in plastids. Int Rev Cell Mol Biol 280: 185-218
- Kato Y, Sakamoto W (2009) Protein quality control in chloroplasts: a current model of D1 protein degradation in the photosystem II repair cycle. J Biochem 146: 463-469
- Kikuchi M, Hatano N, Yokota S, Shimosawa N, Imanaka T, Taniguchi H (2004) Proteomic analysis of rat liver peroxisome. Presence of peroxisome-specific isozyme of Lon protease. J Biol Chem 279: 421-428
- Kinoshita T, Yamada K, Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I (1999) Vacuolar processing enzyme is up-regulated in the lytic vacuoles of vegetative tissues during senescence and under various stressed conditions. Plant J 19: 43-53
- 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: 281-288
- Kuroyanagi M, Nishimura M, Hara-Nishimura I (2002) Activation of Arabidopsis vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. Plant Cell Physiol 43: 143-151
- Kyle DJ, Ohad I, Arntzen CJ (1984) Membrane protein damage and repair: selective loss of a quinone-protein function in chloroplast membranes. Proc Natl Acad Sci USA 81: 4070-4074
- LampI 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). *J Biol Chem* 285: 13550-13560
- Lindahl M, Spetea C, Hundal T, Oppenheim AB, Adam Z, Andersson B (2000) The thylakoid FtsH protease plays a role in the light-induced turnover of the photosystem II D1 protein. *Plant Cell* 12: 419-431
- Lipinska B, Fayet O, Baird L, Georgopoulos C (1989) Identification, characterization, and mapping of the *Escherichia coli* *htrA* gene, whose product is essential for bacterial growth only at elevated temperatures. *J Bacteriol* 171: 1574-1584
- Liu J-X, Srivastava R, Che P, Howell SH (2007) Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J* 51: 897-909
- Luciński R, Misztal L, Samardakiewicz S, Jackowski G (2011) Involvement of Deg5 protease in wounding-related disposal of PsbF apoprotein. *Plant Physiol Biochem* 49: 311-320
- Lüthi AU, Martin SJ (2007) The CASBAH: a searchable database of caspase substrates. *Cell Death Differ* 14: 641-650
- Mahrus S, Trinidad JC, Barkan DT, Sali A, Burlingame AL, Wells JA (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* 134: 866-876
- McDonald L, Robertson DHL, Hurst JL, Beynon RJ (2005) Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides. *Nat Methods* 2: 955-957
- Mikami T, Takao T (2007) Selective isolation of N-blocked peptides by isocyanate-coupled resin. *Anal Chem* 79: 7910-7915
- Murtas G, Reeves PH, Fu Y-F, Bancroft I, Dean C, Coupland G (2003) A nuclear protease required for flowering-time regulation in *Arabidopsis* reduces the abundance of SMALL UBIQUITIN-RELATED MODIFIER conjugates. *Plant Cell* 15: 2308-2319
- Nakaune S, Yamada K, Kondo M, Kato T, Tabata S, Nishimura M, Hara-Nishimura I (2005) A vacuolar processing enzyme, δ VPe, is involved in seed coat formation at the early stage of seed development. *Plant Cell* 17: 876-887
- Nixon PJ, Michoux F, Yu J, Boehm M, Komenda J (2010) Recent advances in understanding the assembly and repair of photosystem II. *Ann Bot* 106: 1-16
- O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250: 4007-4021
- Olinares PDB, Kim J, van Wijk KJ (2011) The Clp protease system; a central component of the chloroplast protease network. *Biochim Biophys Acta* 1807: 999-1011
- Ostersetzer O, Adam Z (1997) Light-stimulated degradation of an unassembled Rieske FeS protein by a thylakoid-bound protease: the possible role of the FtsH protease. *Plant Cell* 9: 957-965
- Ostersetzer O, Kato Y, Adam Z, Sakamoto W (2007) Multiple intracellular locations of Lon protease in *Arabidopsis*: evidence for the localization of AtLon4 to chloroplasts. *Plant Cell Physiol* 48: 881-885
- Plasman K, Van Damme P, Kaiserman D, Impens F, Demeyer K, Helsens K, Goethals M, Bird PI, Vandekerckhove J, Gevaert K (2011) Probing the efficiency of proteolytic events by positional proteomics. *Mol Cell Proteomics* 10, M110.003301
- Qi X, Wilson KA, Tan-Wilson AL (1992) Characterization of the major protease involved in the soybean β -conglycinin storage protein mobilization. *Plant Physiol* 99: 725-733
- Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. *Biochem J* 290: 205-218
- Rawlings ND, Barrett AJ, Bateman A (2010) MEROPS: the peptidase database. *Nucleic Acids Res* 38: D227-D233
- Richter S, Zhong R, Lamppa G (2005) Function of the stromal processing peptidase in the chloroplast import pathway. *Physiol Plant* 123: 362-368
- Rigas S, Daras G, Laxa M, Marathias N, Fasseas C, Sweetlove LJ, Hatzopoulos P (2009) Role of Lon1 protease in post-germinative growth and maintenance of mitochondrial function in *Arabidopsis thaliana*. *New Phytol* 181: 588-600
- Rajo E, Zouhar J, Carter C, Kovaleva V, Raikhel NV (2003) A unique mechanism for protein processing and degradation in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 100: 7389-7394
- Rudenskaya GN, Bogdanova EA, Revina LP, Golovkin BN, Stepanov VM (1995) Macluralisin - a serine proteinase from fruits of *Maclura pomifera* (Raf.) Schneid. *Planta* 196: 174-179
- Sarria R, Lyznik A, Vallejos CE, Mackenzie SA (1998) A cytoplasmic male sterility-associated mitochondrial peptide in common bean is post-translationally regulated. *Plant Cell* 10: 1217-1228
- Schuhmann H, Huesgen PF, Gietl C, Adamska I (2008) The DEG15 serine protease cleaves peroxisomal targeting signal 2-containing proteins in *Arabidopsis*. *Plant Physiol* 148: 1847-1856
- Seo S-B, Tan-Wilson A, Wilson KA (2001) Protease C2, a cysteine endopeptidase involved in the continuing mobilization of soybean β -conglycinin seed proteins. *Biochim Biophys Acta* 1545: 192-206
- Shao W, Yeretsian G, Doiron K, Hussain SN, Saleh M (2007) The caspase-1 digestome identifies the glycolysis pathway as a target during infection and septic shock. *J Biol Chem* 282: 36321-36329
- Shimada T, Yamada K, Kataoka M, Nakaune S, Koumoto Y, Kuroyanagi M, Tabata S, Kato T, Shinozaki K, Seki M, Kobayashi M, Kondo M, Nishimura M, Hara-Nishimura I (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J Biol Chem* 278: 32292-32299
- Shimada T, Yamada K, Kataoka M, Nakaune S, Koumoto Y, Kuroyanagi M, Tabata S, Kato T, Shinozaki K, Seki M, Kobayashi M, Kondo M, Nishimura M, Hara-Nishimura I (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. *J Biol Chem* 278: 32292-32299
- Shipman RL, Inoue K (2009) Suborganellar localization of plastidic type I signal peptidase 1 depends on chloroplast development. *FEBS Lett* 583: 938-942
- Sjögren LLE, Stanne TM, Zheng B, Sutinen S, Clarke AK (2006) Structural and functional insights into the chloroplast ATP-dependent Clp protease in *Arabidopsis*. *Plant Cell* 18: 2635-2649
- Srivastava R, Liu J-X, Howell SH (2008) Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in *Arabidopsis*. *Plant J* 56: 219-227
- Srivastava R, Liu J-X, Guo H, Yin Y, Howell SH (2009) Regulation and processing of a plant peptide hormone, AtRALF23, in *Arabidopsis*. *Plant J* 59: 930-939
- Staes A, Van Damme P, Helsens K, Demol H, Vandekerckhove J, Gevaert K (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* 8: 1362-1370
- Ståhl A, Moberg P, Ytterberg J, Panfilov O, Brockenhuus von Löwenhielm H, Nilsson F, Glaser E (2002)

- Isolation and identification of a novel mitochondrial metalloprotease (PreP) that degrades targeting presequences in plants. *J Biol Chem* 277: 41931-41939
- Stanne TM, Sjögren LLE, Koussevitzky S, Clarke AK (2009) Identification of new protein substrates for the chloroplast ATP-dependent Clp protease supports its constitutive role in *Arabidopsis*. *Biochem J* 417: 257-268
- Strauch KL, Beckwith J (1988) An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc Natl Acad Sci USA* 85: 1576-1580
- Sun X, Fu T, Chen N, Guo J, Ma J, Zou M, Lu C, Zhang L (2010a) The stromal chloroplast Deg7 protease participates in the repair of photosystem II after photoinhibition in *Arabidopsis*. *Plant Physiol* 152: 1263-1273
- Sun X, Ouyang M, Guo J, Ma J, Lu C, Adam Z, Zhang L (2010b) The thylakoid protease Deg1 is involved in photosystem-II assembly in *Arabidopsis thaliana*. *Plant J* 62: 240-249
- Sun X, Peng L, Guo J, Chi W, Ma J, Lu C, Zhang L (2007) Formation of DEG5 and DEG8 complexes and their involvement in the degradation of photodamaged photosystem II reaction center D1 protein in *Arabidopsis*. *Plant Cell* 19: 1347-1361
- Sundström JF, Vaculova A, Smertenko AP, Savenkov EI, Golovko A, Minina E, Tiwari BS, Rodriguez-Nieto S, Zamyatnin AA Jr, Válineva T, Saarikettu J, Frilander MJ, Suarez MF, Zavalov A, Ståhl U, Hussey PJ, Silvennoinen O, Sundberg E, Zhivotovsky B, Bozhkov PV (2009) Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat Cell Biol* 11: 1347-1354
- Swamy KHS, Goldberg AL (1981) *E. coli* contains eight soluble proteolytic activities, one being ATP dependent. *Nature* 292: 652-654
- Taouatas N, Altelar AFM, Drugan MM, Helbig AO, Mohammed S, Heck AJR (2009) Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications. *Mol Cell Proteomics* 8: 190-200
- Terp GE, Christensen IT, Jørgensen FS (2000) Structural differences of matrix metalloproteinases. Homology modeling and energy minimization of enzyme-substrate complexes. *J Biomol Struct Dyn* 17: 933-946
- Than ME, Helm M, Simpson DJ, Lottspeich F, Huber R, Gietl C (2004) The 2.0 Å crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. *J Mol Biol* 336: 1103-1116
- Timmer JC, Enoksson M, Wildfang E, Zhu W, Igarashi Y, Denault J-B, Ma Y, Dummitt B, Chang Y-H, Mast AE, Eroshkin A, Smith JW, Tao WA, Salvesen GS (2007) Profiling constitutive proteolytic events *in vivo*. *Biochem J* 407: 41-48
- Tornero P, Mayda E, Gómez MD, Cañas L, Conejero V, Vera P (1996) Characterization of LRP, a leucine-rich repeat (LRR) protein from tomato plants that is processed during pathogenesis. *Plant J* 10: 315-330
- Uren AG, O'Rourke K, Aravind L, Pisabarro MT, Seshagiri S, Koonin EV, Dixit VM (2000) Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 6: 961-967
- van der Hoorn RAL (2008) Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* 59: 191-223
- van der Hoorn RAL, Jones JDG (2004) The plant proteolytic machinery and its role in defence. *Curr Opin Plant Biol* 7: 400-407
- Van Damme P, Martens L, Van Damme J, Hugelier K, Staes A, Vandekerckhove J, Gevaert K (2005) Caspase-specific and nonspecific *in vivo* protein processing during Fas-induced apoptosis. *Nat Methods* 2: 771-777
- Van Damme P, Staes A, Bronsoms S, Helsen K, Colaert N, Timmerman E, Aviles FX, Vandekerckhove J, Gevaert K (2010) Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nat Methods* 7: 512-515
- van Domselaar R, de Poot SAH, Bovenschen N (2010) Proteomic profiling of proteases: tools for granzyme degradomics. *Expert Rev Proteomics* 7: 347-359
- 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. *J Mol Biol* 364: 625-636
- Verma R, Aravind L, Oania R, McDonald WH, Yates JR III, Koonin EV, Deshaies RJ (2002) Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* 298: 611-615
- Wagner R, Aigner H, Pružinská A, Jänkänpää HJ, Jansson S, Funk C (2011) Fitness analyses of *Arabidopsis thaliana* mutants depleted of FtsH metalloproteases and characterization of three FtsH6 deletion mutants exposed to high light stress, senescence and chilling. *New Phytol* 191: 449-458
- Wolf S, Rausch T, Greiner S (2009) The N-terminal pro region mediates retention of unprocessed type-I PME in the Golgi apparatus. *Plant J* 58: 361-375
- Yang P, Smalle J, Lee S, Yan N, Emborg TJ, Vierstra RD (2007) Ubiquitin C-terminal hydrolases 1 and 2 affect shoot architecture in *Arabidopsis*. *Plant J* 51: 441-457
- Želisko A, García-Lorenzo M, Jackowski G, Jansson S, Funk C (2005) AtFtsH6 is involved in the degradation of the light-harvesting complex II during high-light acclimation and senescence. *Proc Natl Acad Sci USA* 102: 13699-13704
- Zhang H, Dong S, Wang M, Wang W, Song W, Dou X, Zheng X, Zhang Z (2010) The role of vacuolar processing enzyme (VPE) from *Nicotiana benthamiana* in the elicitor-triggered hypersensitive response and stomatal closure. *J Exp Bot* 61: 3799-3812
- Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* 3: e1994
- Zybailov B, Friso G, Kim J, Rudella A, Ramírez Rodríguez V, Asakura Y, Sun S, van Wijk KJ (2009) Large scale comparative proteomics of a chloroplast Clp protease mutant reveals folding stress, altered protein homeostasis, and feedback regulation of metabolism. *Mol Cell Proteomics* 8: 1789-1810

Chapter 2

Metacaspases

Liana Tsiatsiani^{1,2,3,4}, Frank Van Breusegem^{1,2}, Patrick Gallois⁵, Anton Zavalov⁶, Eric Lam⁷ and Peter V. Bozhkov^{8,*}

¹Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium

²Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

³Department of Medical Protein Research, VIB, B-9000 Ghent, Belgium

⁴Department of Biochemistry, Ghent University, B-9000 Ghent, Belgium

⁵Faculty of Life Sciences, University of Manchester, Manchester, M13 9PT, UK

⁶Department of Molecular Biology, Biomedical Center, Swedish University of Agricultural Sciences, Box 590, SE-75124 Uppsala, Sweden

⁷Department of Plant Biology and Pathology, Rutgers the State University of New Jersey, 59 Dudley Road, New Brunswick, NJ 08901, USA

⁸Department of Plant Biology and Forest Genetics, Uppsala BioCentre, Swedish University of Agricultural Sciences, Box 7080, SE-75007 Uppsala, Sweden

AUTHOR CONTRIBUTIONS

LT, FVB, PG, AZ, EL and PB wrote the manuscript.

Redrafted from:

Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavalov, A., Lam, E., and Bozhkov, P.V. (2011). Metacaspases. *Cell Death Differ* 18, 1279-1288.

ABSTRACT

Metacaspases are cysteine-dependent proteases found in protozoa, fungi and plants and are distantly related to metazoan caspases. Although metacaspases share structural properties with those of caspases, they lack Asp specificity and cleave their targets after Arg or Lys residues. Studies performed over the past 10 years have demonstrated that metacaspases are multifunctional proteases essential for normal physiology of non-metazoan organisms. This article provides a comprehensive overview of the metacaspase function and molecular regulation during programmed cell death, stress and cell proliferation, as well as an analysis of the first metacaspase-mediated proteolytic pathway. To prevent further misapplication of caspase-specific molecular probes for measuring and inhibiting metacaspase activity, we provide a list of probes suitable for metacaspases.

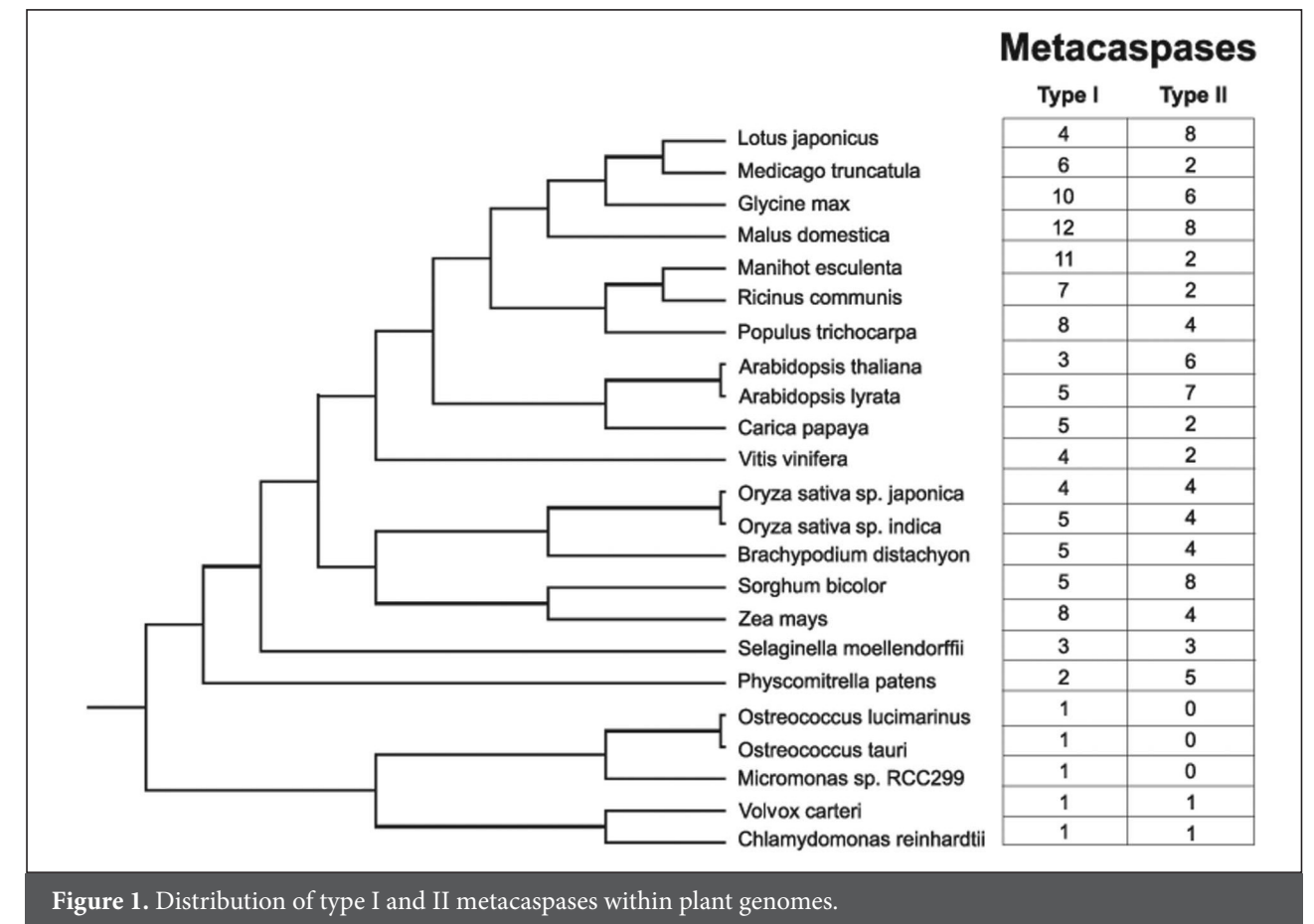
Key words: metacaspase, substrate specificity, caspase, cell death, cell proliferation, protein aggregation, *in vivo* degradome.

INTRODUCTION

More than a decade ago, an end was put to the until-then frustrating hunt for the presence of caspase orthologues in plants and other non-metazoan organisms. Based on predicted structural homologies with the catalytic domain of caspases, metacaspase sequences were identified in protozoa, fungi and plants.¹ Phylogenetic analysis indicated that eukaryotic caspases, metacaspases, and paracaspases are about equally distant from each other and together with the legumains, separases and the bacterial clostripains and gingipains, they are classified within the clan CD of Cys proteases.² A common structural feature of this clan is the presence of a caspase/hemoglobinase fold.³

Two types of metacaspases can be distinguished. Type I metacaspases have a N-terminal prodomain containing a proline-rich repeat motif and, in plant members, also a zinc finger motif. Type II metacaspases lack such a prodomain but harbor a linker region between the putative large (p20) and small (p10) subunits. In protozoa and fungi only type I metacaspases are found, whereas plant genomes encode both types with frequent prevalence of type I over type II by gene numbers (Figure 1).

Besides a few exceptions having a catalytic Ser residue instead of a Cys,⁴ metacaspases, like caspases have a His-Cys catalytic dyad in their predicted active site, with the Cys residue acting as a nucleophile for substrate peptide bond hydrolysis. For most metacaspases studied so far, enzyme maturation involves autocatalytic processing of the zymogen;⁵⁻⁸ however in some cases this step is not necessarily required for proteolytic activation.⁹⁻¹¹ The



most striking biochemical feature of all metacaspases that distinguishes them from caspases is a strict Arg and Lys substrate specificity.^{5,6,10,12,13}

In *Arabidopsis* studies, metacaspases have been described with two distinct nomenclatures. The Van Breusegem lab used the five-character name AtMC1-AtMC9 (*Arabidopsis thaliana* Metacaspase) while the Lam lab used the seven-character name AtMCP1a-AtMCP1c (*A. thaliana* Metacaspase) for the type I and AtMCP2a-AtMCP2f for the type II metacaspases. In order to avoid confusion and for the reasons of simplicity we have decided to establish a uniform nomenclature of *Arabidopsis* metacaspases namely, AtMC1-AtMC9, which we urge to be adopted by other researchers in future publications and discussions (Table 1).

The first decade of metacaspase research documented by more than 130 publications is over now. It has brought exciting information about biological functions of metacaspases and the first attempts to uncover molecular mechanisms of metacaspase action. This period was also marked by vast contradictions within the field, mainly due to different opinions about biochemical and functional relatedness between caspases and metacaspases.^{2,14-16} Here we review the essential information related to the biochemistry and function of metacaspases. In addition, we detail some of the important tools that are necessary to specifically dissect the role and activity of this highly conserved group of proteases.

Multifunctionality, functional specialization or redundancy of metacaspases

Since the metacaspase Yca1 (also termed Mca1) was found to be a positive regulator of oxidative stress- and senescence-associated cell deaths in the budding yeast (*Saccharomyces cerevisiae*),¹⁷ a great deal of reverse genetic analysis has been directed to functionally characterize metacaspases in various organisms. This analysis has revealed that cell death regulation is just one of the multifaceted abilities of metacaspases, which can also control other biological processes, either related or unrelated to cell death (Table 2).

The number of metacaspase genes in the genomes of different organisms varies considerably (Figure 1; Table 2; <http://merops.sanger.ac.uk/>). Such an unequal distribution of metacaspases between different phyla is a good paradigm to study multifunctionality and functional specialization of metacaspases.

A single metacaspase Yca1 of the budding yeast provides a striking example of a multifunctional protein, as apart from cell death activated under various settings, it is also required for cell cycle regulation and clearance of protein aggregates (Table 2).¹⁴ Deletion of the single metacaspase gene *Pca1* of the fission yeast (*Schizosaccharomyces pombe*) suppresses lipotoxic and inositol starvation-induced death.^{18,19} LmjMCA, a single metacaspase of *Leishmania major*, is involved in cell cycle regulation²⁰ and, when overexpressed, stimulates

Table 1. The two different *Arabidopsis* metacaspase nomenclatures and the unified one.

Gene AGI	Metacaspase type	Van Breusegem lab ⁵	Lam lab ¹²	Unified
At1g02170	I	AtMC1	AtMCP1b	AtMC1
At4g25110	I	AtMC2	AtMCP1c	AtMC2
At5g64240	I	AtMC3	AtMCP1a	AtMC3
At1g79340	II	AtMC4	AtMCP2d	AtMC4
At1g79330	II	AtMC5	AtMCP2b	AtMC5
At1g79320	II	AtMC6	AtMCP2c	AtMC6
At1g79310	II	AtMC7	AtMCP2a	AtMC7
At1g16420	II	AtMC8	AtMCP2e	AtMC8
At5g04200	II	AtMC9	AtMCP2f	AtMC9

Table 2. Metacaspase functions identified by reverse genetics.

Kingdom	Species	Total number of metacaspase genes	Genes with known physiological function	Method*	Function	Reference
Protozoa	<i>L. donovani</i>	2	<i>LdMc1</i> , <i>LdMc2</i>	OE	Positive regulators of oxidative stress-induced cell death	Lee et al. ¹⁰
	<i>L. major</i>	1	<i>LmjMCA</i>	KO, OE	Cell cycle regulation	Ambit et al. ²⁰
				OE	Positive regulator of oxidative stress-induced cell death	Zalila et al. ²¹
	<i>T. brucei</i>	5	<i>TbMCA2</i> , <i>TbMCA3</i> , <i>TbMCA5</i>	RNAi, KO	Required for bloodstream form	Helms et al. ²²
	<i>T. cruzi</i>	2	<i>TcMCA5</i>	OE	Positive regulator of fresh human serum-induced cell death	Kosec et al. ³⁴
Plantae	<i>A. thaliana</i>	9	<i>AtMC8</i>	KO	Required for UVC stress-induced cell death	He et al. ³⁰
			<i>AtMC1</i> , <i>AtMC2</i>	KO, OE, IM	Antagonistically control HR cell death activated by intracellular immune receptors	Coll et al. ²⁶
			<i>AtMC4</i>	KO, OE, IM	Mediates PCD activation by the fungal toxin FB1 and abiotic stress inducers	Watanabe and Lam ³²
	<i>N. benthamiana</i>	?	<i>NbMCI</i>	VIGS	Confers resistance to fungal necrotroph <i>Colletotrichum destructivum</i>	Hao et al. ⁶⁶
	<i>P. abies</i>	?	<i>MclI-Pa</i>	RNAi	Required for embryogenesis and associated PCD	Suarez et al. ²⁹
Fungi	<i>A. fumigatus</i>	2	<i>CasA</i> , <i>CasB</i>	KO	Confer cytoprotection under ER stress	Richie et al. ²⁴
	<i>A. nidulans</i>	2	<i>CasA</i> , <i>CasB</i>	KO	Antagonistically control ER stress-associated cell death	Colabardini et al. ²⁵
	<i>C. albicans</i>	1	<i>CaMCA1</i>	KO	Required for oxidative stress-induced cell death	Cao et al. ²⁸
	<i>P. anserina</i>	2	<i>PaMca1</i> , <i>PaMca2</i>	KO	Required for senescence-associated cell death	Hamann et al. ²³
	<i>S. cerevisiae</i>	1	<i>Yca1</i>	KO, OE	Required for oxidative, salt, propolis, and osmotic stress-induced cell death	Madeo et al., ²⁷ Chahomchuen et al. ⁶⁷ and de Castro et al. ⁶⁸
				KO, OE	Required for cell death associated with defects in ubiquitination, DNA replication, mRNA stability and mitochondrial metabolism	Madeo et al. ²⁷ and Walter et al. ⁶⁹
				KO	Required for virus-induced cell death	Ivanovska and Hardwick ⁷⁰
				KO	Required for chronological aging-associated death	Herker et al. ⁷¹
				KO	Mediates carnitine-dependent lifespan extension	Palermo et al. ⁷²
				KO, IM	Required for clearance of protein aggregates	Lee et al. ³⁸
				KO, IM	Cell cycle regulation	Lee et al. ³⁹
	<i>S. pombe</i>	1	<i>Pca1</i>	KO	Required for lipotoxic cell death in minimal medium and cell death induced by inositol starvation	Low et al. ¹⁸ and Guerin et al. ¹⁹

* IM, inactivation mutant; KO, knock out; OE, overexpression; RNAi, RNA interference; VIGS, virus-induced gene silencing

oxidative-stress induced cell death.²¹ Further studies using corresponding deletion and inactivation mutants are required to determine whether the single metacaspases present in *L. major* and fission yeast are multifunctional proteins comparable to Yca1.

Most organisms have two or more metacaspase genes (Table 2). There is no evidence yet to indicate significant degree of either specialization or redundancy in the physiological functions of different metacaspase genes expressed in the same organism. Partial redundancy has however been demonstrated for three *Trypanosoma brucei* metacaspases during the development of a bloodstream-specific form of the parasite,²² for two *Podospira anserina* metacaspases in the regulation of senescence-associated death²³ and for two *Aspergillus fumigatus* metacaspases in the resistance to endoplasmic reticulum (ER) stress.²⁴ Redundancy may not always be the case, as two *Aspergillus nidulans* metacaspases antagonistically control ER stress-induced cell death, one metacaspase (CasA) acting as positive regulator and another as a negative one (CasB).²⁵ Similarly, a recent study also documented the antagonistic functions of two *Arabidopsis* type I metacaspases, AtMC1 and AtMC2, in mediating hypersensitive response (HR)-associated cell death (Table 2).²⁶

Compared to genomes of protozoa and fungi with a single or a few type I metacaspase genes and no type II metacaspase genes, genomes of higher plants encode larger metacaspase families including both type I and type II members; e.g. there are nine and twelve metacaspase genes in *Arabidopsis thaliana* and *Populus trichocarpa*, respectively (Figure 1; <http://merops.sanger.ac.uk/>). Although the complete range of the roles that metacaspases may serve in plant biology remain to be understood, there is growing evidence that they are required for PCD to take place under diverse settings (Table 2; see following section).

Metacaspases in cell death

Madeo *et al.* first reported metacaspase-dependent cell death using yeast Yca1 mutant strains subjected to oxidative stress.¹⁷ Since then, a pro-cell death role for Yca1 has been demonstrated in budding yeast cells in response to viral toxins and different types of abiotic and metabolic defect-associated stresses and chronological aging (Table 2).²⁷ A similar requirement for a metacaspase has been reported in the fission yeast, albeit for a more limited number of cell death conditions, including lipotoxic stress¹⁸ and inositol starvation¹⁹ (Table 2). Further in the fungi kingdom it has been established that both PaMca1 and PaMca2 are required for senescence-associated cell death in *Podospira anserina*,²³ a single *Candida albicans* metacaspase, CaMCA1, mediates oxidative stress-induced cell death in this pathogenic yeast,²⁸ whereas one of the two metacaspases, CasA, promotes ER stress-associated cell death in *Aspergillus nidulans*²⁵ (Table 2).

A pro-cell death role of metacaspases has now been extended to protozoa and plants. In *Leishmania* species metacaspases are required for oxidative stress-induced cell death (Table 2).^{10,21} In the gymnosperm plant, Norway spruce (*Picea abies*), gene silencing of a type II metacaspase, mcII-Pa, suppressed terminal cell differentiation and PCD in the embryo-suspensor causing developmental arrest at the early stage of embryogenesis.²⁹ Unexpectedly, PCD phenotypes in metacaspase knockout lines of the model plant *Arabidopsis* have been less forthcoming, possibly due to gene redundancy. In *Arabidopsis*, all nine metacaspase genes are expressed at different levels in various parts of the plant (our unpublished results). Among these genes, a type II metacaspase gene *AtMC8* is strongly up-regulated during oxidative stress. This up-regulation requires the gene *Radical-Induced Cell Death1 (RCD1)*,³⁰ which is thought to be a mediator of oxidative stress responses in *Arabidopsis*.³¹ Consistent

with these observations, *atmc8* knockout lines exhibited reduced cell death triggered by UVC radiation or hydrogen peroxide.³⁰ Oxidative stress-induced PCD represents a dramatic example of a metacaspase-dependent process that is conserved throughout a long evolutionary distance, from protozoa to plants (Table 2). Loss of the predominant type II metacaspase in *Arabidopsis*, AtMC4, has been found recently to compromise cell death induction by the fungal toxin fumonisin B-1 (FB1), an incompatible bacterial pathogen, and chemical inducers of oxidative stress in seedlings.³² This work has further shown that AtMC4 processing from its zymogen is accelerated during activation of cell death by these agents, consistent with its deduced role in orchestrating PCD under these conditions. Lastly, a type I metacaspase AtMC1 was shown to be a positive regulator of the HR cell death in *Arabidopsis*, whereas another type I metacaspase, AtMC2, was found to counteract this pro-cell death effect of AtMC1 (see also review by Coll *et al.* in this issue).²⁶ In the case of AtMC2 as anti-death regulator, it is interesting to note that its proteolytic activity is apparently not required for this function in *Arabidopsis*. Genetic manipulation of the two metacaspases could nearly eliminate HR activated by plant intracellular immune receptors.²⁶ To conclude, trans-kingdom conservation of a role for metacaspases in PCD regulation emphasizes that these proteases are a fundamental part of the non-metazoan cell death machinery.

Although several molecular components of PCD pathways have been recently identified in non-metazoan organisms, metacaspase position(s) in these pathways remain elusive. The identification of Tudor staphylococcal nuclease (TSN) as a common substrate for both Norway spruce metacaspase mcII-Pa and human caspase-3 (see below)³³ suggests that metacaspases can execute PCD like effector caspases. Nuclear translocation of metacaspases during cell disassembly in the Norway spruce embryo-suspensors⁶ and in the dying epimastigotes of *Trypanosoma cruzi*³⁴ indirectly supports this scenario. By contrast, the antagonistic control of cell death by two type I metacaspases, such as the one shown for the *Arabidopsis* HR²⁶ and *Aspergillus nidulans* ER stress²⁵ suggests a possible role in the decision phase.

Noteworthy, in budding yeast approximately 60% of the examples of cell death reported so far are Yca1-independent.²⁷ For example, Yca1 is not required for PCD during mating,³⁵ in response to ammonia³⁶ or induced by human BAX expression.³⁷ Therefore it can be expected that further understanding of PCD mechanisms in plants, which have both type I and type II metacaspases, should facilitate a better definition of metacaspase-dependent and -independent pathways.

Metacaspases in protein aggregation and ER stress

Proteome analysis of yeast cells deleted for *Yca1* has revealed a role for this metacaspase in the clearance of insoluble protein aggregates (Table 2).³⁸ The ablation of *Yca1* was found to direct an accumulation of insoluble protein aggregates during physiological growth conditions, correlated with an increased abundance of vacuolar peptidases and stress-response chaperones. This accumulation of protein aggregates induced the autophagic pathway. Moreover, the normally cytosolic Yca1-GFP was observed to co-localize with the protein aggregation marker Hsp104-mRFP during heat stress and in aged cultures. Deletion of the poly-Q (QQXX) motif in the Yca1 prodomain demonstrated that it was largely responsible for the metacaspase localization in the protein aggregates. Interestingly, catalytic activity of Yca1 was not required for its aggregate-specific localization, representing the first indication that metacaspase could participate in the formation of protease-activation scaffold analogous to signaling platforms such as the apoptosome.^{14,38}

These data indicate that metacaspases can function to promote the dissipation of protein aggregates under normal and stress conditions in addition to their role in PCD regulation. It is possible to argue that $\Delta yca1$ yeast cells have a reduced level of PCD under stress because of an increased expression of molecular chaperones and of the activation of autophagy, a pro-survival mechanism.³⁸ This interpretation suggests a possible indirect effect of metacaspase function on PCD regulation, at least in some conditions. However, the poly-Q motif required for aggregate localization is only present in one of the nine *Arabidopsis* metacaspases (AtMC3) and is absent from the other metacaspases shown to regulate PCD, suggesting that this proposition based on observation in yeast might not hold true when investigated in other organisms.

In the human fungal pathogen *Aspergillus fumigatus*, deletion of the two metacaspase-encoding genes, *CasA* and *CasB*, leads to hypersensitivity to agents that induce ER stress (Table 2).²⁴ In contrast, no effect on growth was observed with the *DcasA/DcasB* mutant under conditions of a variety of agents that induce oxidative stress. These results indicate that in *A. fumigatus* the metacaspases may be important mediators of survival under ER stress with one possible scenario in their involvement in clearing aggregated proteins in the ER, similar to yeast Yca1. Alternatively, these proteases may be important for cleavage of specific cellular targets that play an important role in the management of ER stress. It would be important to test whether proteolytic activity of CasA and CasB is required for their role for survival under ER stress.

Metacaspases in cell proliferation

Ambit *et al.* found that the metacaspase in *Leishmania major*, LmjMCA, exhibits specific subcellular localization.²⁰ While it is relatively dispersed throughout the cell during interphase, it tends to colocalize with the kinetoplast during mitochondria segregation and it also translocates to the nucleus and mitotic spindle of the parasite during mitosis. Attempts to manipulate LmjMCA levels resulted either in a growth defect with suppression of cytokinesis frequency (overexpression) or inviability (deletion), suggesting that this protein may play an essential function(s) during cell division.²⁰ However, whether the protease activity of LmjMCA is essential for its role in cell cycle progression in this parasite remains to be clearly demonstrated.

In the budding yeast, loss of Yca1 or its replacement with a point mutation variant in its active site Cys resulted in a delay of the doubling time that correlates with a slower G1/S phase transition.³⁹ Even more dramatically, loss of Yca1 activity results in the uncoupling of cell cycle arrest at the G2/M checkpoint to nocodazole treatment, which disrupts microtubules important for cytokinesis. Interestingly, the desensitization of cell cycle progression to nocodazole treatments has also been found to result from a loss of caspase-3 activities in human hepatoma cells.⁴⁰ These studies thus underscore that metacaspase and caspase could play a common, conserved role in cell cycle progression of eukaryotes.

Measurement and inhibition of metacaspase activities

Due to the presence of the C14 caspase domain, many efforts initially aimed to demonstrate caspase-like proteolytic activity (namely, proteolysis after an Asp residue) of various metacaspases. Detection of specific proteolytic activity is usually performed by using synthetic tri- or tetra-peptides C-terminally coupled to fluorogenic moieties such as 7-amino-4-methylcoumarin (AMC), 4-methylcoumarin-7-amide (MCA) or β -naphthylamide (2NA). Initial reports using caspase substrates VEID-AMC and IETD-AMC monitored

differential caspase activity levels in lysates from wild-type, $\Delta yca1$ and overexpression yeast strains upon treatment with cell-death promoting agents. The observed activities were abrogated by the broad-range caspase inhibitor zVAD-fluoromethylketone (fmk).¹⁷ Furthermore, metacaspase involvement in the yeast cell death models was investigated after *in vivo* staining for caspase activity by flow cytometry with FITC-labeled VAD-fmk (FITC-VAD-fmk).^{17,41-44} Likewise, in the fungus *Candida albicans*, it was suggested that the antifungal compound Plagiochin E activated metacaspases based on the appearance of FITC-VAD-fmk labelled cells.⁴⁵ In the phytoplankton *Chlamydomonas reinhardtii* and other marine species, such as the diatom *Thalassiosira pseudonana* and the unicellular coccolithore *Emiliania huxleyi*, increased caspase-like activity was correlated with increased accumulation of metacaspase mRNA and protein, thereby providing indirect evidence that metacaspases had caspase-like activity.⁴⁶⁻⁴⁸ Also in plants, silencing of *mcII-Pa* in the Norway spruce embryos reduced the level of the VEIDase caspase-like activity.²⁹ Overall, these reports suggested that metacaspases might be directly responsible for cellular caspase-like activities in fungi and plants. However, a horde of biochemical studies using recombinant metacaspases or protein extracts from loss- or gain-of-function mutants have now clearly demonstrated that metacaspases are highly specific for Arg or Lys at the P1 position.^{4,6,7,9-13,22,23,30,33,49,50} This implies that conclusions, mainly based on genetic perturbation of metacaspase-encoding genes, that metacaspases have caspase-like proteolytic activities were premature.

To accurately monitor metacaspase activities in cell lysates it is therefore essential to use specific metacaspase activity assays that are rooted in substrate specificity and other biochemical characteristics of metacaspases. The determination of the *Arabidopsis* AtMC9 autoprocessing site ITSR(183) (Table 3) was the first indication for a basic amino acid cleavage preference⁵ and steered successful testing of fluorogenic tri- and tetrapeptides

Table 3. Experimentally shown metacaspase autoprocessing sites and cleavage sites in the substrate proteins.

	Metacaspase	Cleaved protein	Cleavage site sequence*	P1 position	Reference		
Autoprocessing	TbMCA2		RD A KGLHG	55	Moss et al. ⁹		
			ADV K NTAT	268	Moss et al. ⁹		
	AMca2		G K VRDLYG	15	Ojha et al. ¹¹		
			DD T RSTSS	27	Ojha et al. ¹¹		
	AtMC9		IT S RALPF	183	Vercammen et al. ⁵		
	AtMC4		AK D KSLPL	225	Watanabe and Lam ⁸		
Substrates	mcII-Pa		F E SRGIHL	188	Bozhkov et al. ⁶		
			K F VKVLVT	269	Bozhkov et al. ⁶		
	AtMC9	AtSerp		IK L RGLLM	351	Vercammen et al. ¹³	
				mcII-Pa	TS N	158	Sundström et al. ³³
					V L NRDVRI	287	Sundström et al. ³³
					S N SKAIRD	371	Sundström et al. ³³
TbMCA2	EF-Tu		PIV R HGSA	172	Moss et al. ⁹		
			S A LKALEG	177	Moss et al. ⁹		

*P1 residue is indicated in bold

with an Arg or Lys residue at the P1 position for cleavage by metacaspases. Based on the autocatalysis of mcII-Pa after KFKV(269) and FESR(188) (Table 3),⁶ an equivalent substrate analogue was employed to detect metacaspase activity in embryonic cell extracts of Norway spruce.³³ A tetrapeptide positional scanning synthetic combinatorial library screen indicated that VRPR-AMC (when Arg was fixed at P1) and IISK-AMC (when Lys was fixed at P1) were the preferred synthetic substrates for recombinant *Arabidopsis* AtMC9.¹³ Recombinant *Trypanosoma brucei* TbMCA2 was shown to autocleave at RDAK(55) and ADVK(268) (Table 3) and diverse substrates with Arg or Lys at P1 (GGR, GRR, VRPR, IKLR, IKLK) were efficiently cleaved.⁹ Similarly, *Allomyces arbuscula* AMca2 was autoprocessed at GKVR(15) and DDTR(27) (Table 3).¹¹ Furthermore, peptidyl substrates with Arg at P1 such as GRR-AMC and GGR-AMC were preferentially proteolysed by recombinant metacaspases AtMC1, AtMC4, AtMC5, AtMC8, AtMC9 and TbMCA2,^{5,8,9,12,30} by immunoprecipitated metacaspases LdMC1, LdMC2 and LmjMCA from *Leishmania* species,^{7,10} and through ectopic overexpression of AtMC1, AtMC4, AtMC5 and Yca1.^{8,12}

The substrate specificity of metacaspases towards Arg or Lys at the P1 position in peptide substrates was also consolidated through the identification of a first set of natural substrates and inhibitors. A recombinant serpin-like protein (AtSerp1) was cleaved within its reactive centre loop by recombinant AtMC9 at IKLR(351) (Table 3).¹³ During expression of TbMCA2 in *E. coli* cells, the bacterial elongation factor EF-Tu was proteolysed at TPIVR(172) and GSALK(177) (Table 3).⁹ Finally, endogenous TSN was cleaved by mcII-Pa at ASIR(158), VLNR(287), SNSK(371) and HSAR(558) during both developmental and oxidative stress-induced PCD in Norway spruce embryos (Table 3).³³

Arginal protease inhibitors leupeptin and antipain, as well as tosyl-lysyl-chloromethylketone (TLCK) have proven to be potent inhibitors of *Arabidopsis*, *Leishmania* and *Trypanosoma* metacaspases^{5,6,9,12} in addition to inhibitory substrate analogues such as FK-trimethylbenzoyloxymethyl ketone (tmbk), EGR-chloromethyl ketone (cmk) and GVR-chloromethyl trifluoro acetate (CHCL TFA).^{5,6,11}

Biochemical studies of metacaspases from *Arabidopsis* (AtMC4 and AtMC8), spruce (mcII-Pa), *Leishmania* (LdMC1 and LdMC2) and *Trypanosoma* (TbMCA2) indicate that the majority prefers neutral to slight basic (7.0-8.5) pH optima for *in vitro* activity assays.^{5,6,9,10,30} Until now, AtMC9 is the only reported metacaspase that requires acidic conditions prior to activation.⁵ Noteworthy *in vitro* activities of AtMC4, AtMC5, mcII-Pa, TbMCA2 and the *A. arbuscula* AMca2 depend on high Ca²⁺ concentrations.^{5,6,9,12} More recent biochemical studies with recombinant AtMC4 demonstrated that these high Ca²⁺ levels are required to activate the protease activity through induced cleavage at a highly conserved site AKDK(225) (Table 3) that has also been shown to be cleaved in AtMC9.⁸ This cleavage was demonstrated to be critical for *in vitro* protease activity of metacaspases^{5,8} as well as *in vivo* function of AtMC4 to complement Dyca1 mutant yeast strains in cell death activation.⁸

To conclude, caspase-specific substrates such as YVAD-AMC, DEVD-AMC, VEID-AMC and IETD-AMC (with Asp at P1) are not indicative of metacaspase catalytic activity. This implies that conclusions drawn on the involvement of metacaspases in specific processes from reports based on caspase-specific peptide substrates, should be revisited and it is advised that activity assays tailored towards the specific catalytic characteristics of metacaspases should be used instead. Since the level of VEIDase activity was reported to be decreased along with either the loss of *Yca1* in yeast¹⁷ or by suppression of *mcII-Pa* in spruce cells,²⁹ these results would indicate that cryptic caspase-like proteases may be activated downstream of the metacaspase regulation point. The

suppressing effect of pan-caspase inhibitor such as zVAD-fmk and zVEID-fmk on cell death-related phenomena in yeast and spruce cells would support the role for such caspase-like proteases in orchestrating cell death downstream of metacaspases in some of the cell death pathways.^{12,17,51} We urge the use of fluorogenic tri- or tetrapeptide substrates (eg. VRPR-AMC, IISK-AMC, GRR-AMC, GGR-AMC and FESR-AMC) with Arg or Lys residues at the P1 position to detect metacaspase activities in cellular extracts or of recombinant metacaspases. Accordingly, the use of caspase inhibitors against metacaspases is inappropriate and should be replaced by arginal protease inhibitors. It should be noted however that, in addition to metacaspases, these inhibitors can suppress other arginine-specific proteases present in the cells or test tube. Given the variability of metacaspase pH optima and possible Ca²⁺ dependence, it is advisable to screen for proteolytic activity at different levels of pH and different concentrations of Ca²⁺ as well.

***In vivo* metacaspase degradome: Tudor staphylococcal nuclease**

Unraveling the identity of natural substrates of metacaspases will give an insight into molecular mechanisms of metacaspase-dependent processes. At present, TSN is the only protein shown to be cleaved by a metacaspase *in vivo*.³³ TSN is an evolutionarily and structurally conserved protein found in all eukaryotes (except for budding yeast). It has an invariant domain architecture including a tandem of four staphylococcal nuclease-like domains (SN), followed by a Tudor domain and the fifth SN domain (Figure 2a). Owing to its complex domain structure, TSN takes part in several fundamental mechanisms of gene regulation in animal cells, including transcription,⁵² mRNA splicing⁵³ and RNA silencing.^{54,55} Decrease of *TSN* expression triggers cell death^{33,36} and impairs plant viability and stress tolerance.^{33,57}

During both developmental and oxidative stress-induced cell death in Norway spruce embryos active metacaspase mcII-Pa cleaves endogenous TSN at four sites. Three sites are situated between SN domains and the fourth site inside SN2 domain (Figure 2a).³³ All cleavage sites contain either Arg or Lys at P1 position, but there was no strict specificity beyond P1 residue (P4-P3-P2 and P1') (Table 3). In the same study caspase-3-mediated cleavage of human TSN at the canonical cleavage site DAVD/S (ref 58) situated between the Tudor and SN5 domains has been revealed in apoptotic cells.³³ A shared *in vivo* target of caspases and metacaspases suggests that plants and animals may have conserved some key proteolytic pathways to compromise cell viability, despite the large differences in morphology and biochemistry of cell death between the two kingdoms. In addition, cleavage of TSN by an effector caspase and a type II metacaspase provides a good model system for comparative analysis of the biochemical properties of caspases and metacaspases *in vivo*.¹⁴ One conclusion from such analysis is that metacaspases appear to be far more liberal than caspases in regards to P4-P3-P2 and P1' substrate residues (Figure 2a). This conclusion is consistent with earlier results of high-throughput screening of a combinatorial tetrapeptide substrate library with recombinant AtMC9.¹³ Future studies of *in vivo* metacaspase degradome might however identify more prominent metacaspase specificity towards the aforementioned residues.

The crystal structure of the Tudor-SN5 part of human TSN has been determined to high resolution (Figure 2b).⁵⁹ To complete the 3D model of the protein, each of the SN1-4 domains can be reliably reconstructed based on the available crystal structures of staphylococcal nuclease (PDB 1SNC) and human SN5 (PDB 2O4X), e.g. the SN2 domain of spruce TSN shown in Figure 2c. The analysis of the metacaspase mcII-Pa cleavage sites in

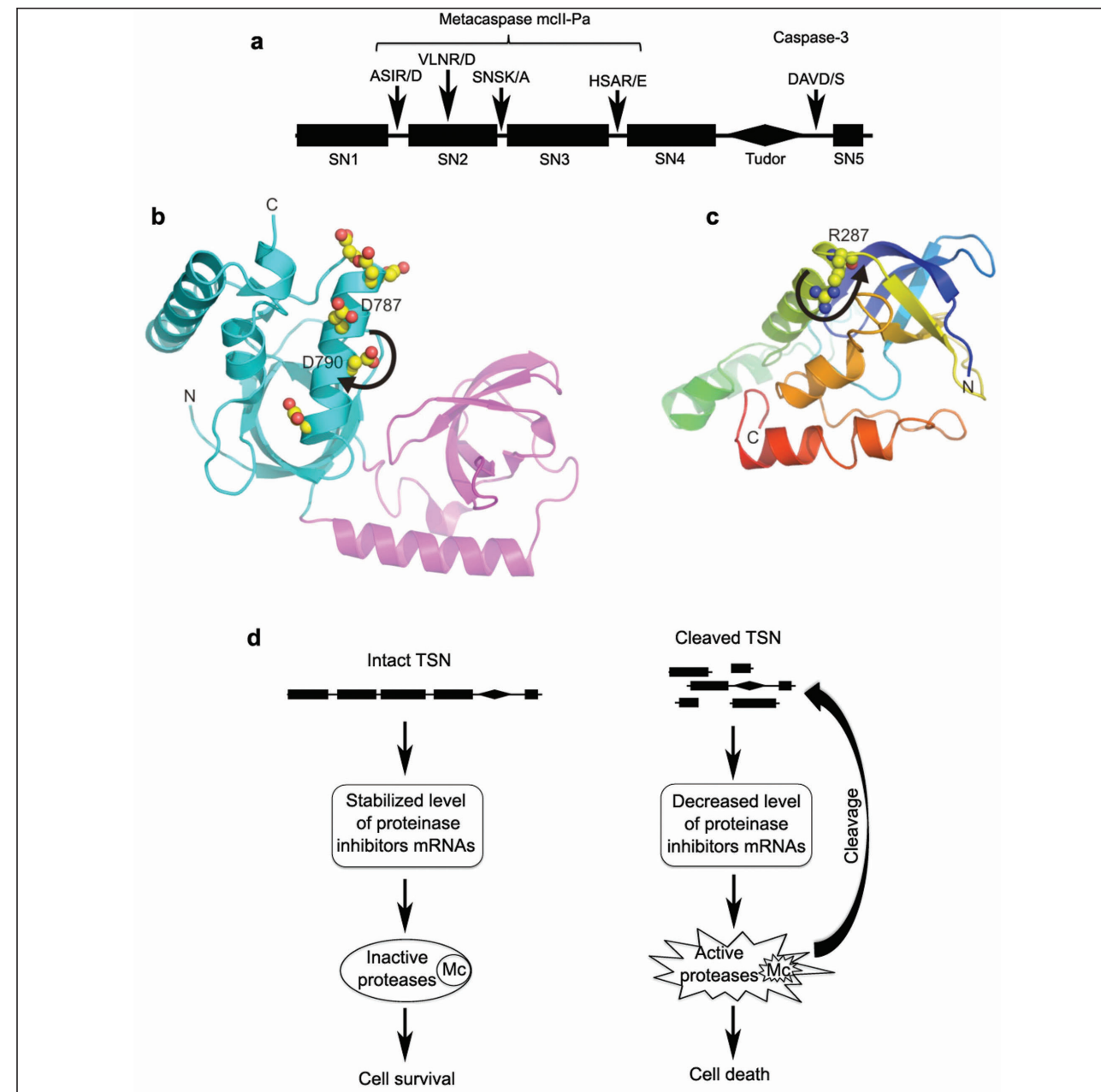


Figure 2. The mechanisms of TSN cleavage during cell death. (a) Domain structure of TSN and location of metacaspase and caspase cleavage sites. (b, c) Structural preferences for the cleavage of TSN by caspase-3 and metacaspase mclI-Pa (b) Caspase-3 cleaves helix $\alpha 2$ of human TSN. Tudor-SN5 part of human TSN is shown as cartoon with the SN5 and Tudor domains painted in cyan and pink, respectively. The caspase-3 cleavage motif (787-DAVD-790) is located in the middle of helix $\alpha 2$. Three of four residues of the motif, Asp 787 and 790 and Val 789 are accessible to the solvent. The $\alpha 2$ helix exposes six Asp residues (including two Asp residue of the DAVD motif; shown in spheres), creating a highly negatively charged patch on the surface of the helix. The abundance of Asp residues around the cleavage site might facilitate its recognition by caspase-3. (c) Metacaspase mclI-Pa cleaves a structured but solvent accessible loop ($\alpha 2$ - $\beta 8$) of the SN2 domain (cartoon painted in rainbow) of spruce TSN. The loop is cleaved after Arg 287 (shown in spheres). The structure of the SN2 domain of spruce TSN was modelled with SwissModel using staphylococcal nuclease (PDB 1SNC) and Tudor-SN5 part of human TSN (PDB 2O4X) as template structures. Arrows in b and c indicate cleavage sites. Graphics was generated using PyMol. (d) A proposed model for pro-cell death role of metacaspase-mediated cleavage of TSN. Mc, metacaspases.

this model reveals that all of them locate in loop regions: three in the loops situated between SN1-4 domains and one in the middle of a structured but solvent accessible loop ($\alpha 2$ - $\beta 8$) of the SN2 domain (Figure 2c). The preference for digesting loops is common for the majority of proteases and hence is not surprising. In contrast, the single cleavage site in human TSN for caspase-3 is located in a long helix ($\alpha 2$) of the Tudor-SN5 part (Figure 2b). Although the ability of caspase-3 to cleave α helices has been previously demonstrated,⁶⁰ this is not consistent with the structural analysis of the caspase-3-inhibitor complex.⁶¹ In this complex, the inhibitor inserts an extended conformation segment into the active site, suggesting that cleavage sites must either have this conformation or acquire it by unfolding. Interestingly, the caspase-3-cleaved helix $\alpha 2$ of human TSN contains six Asp residues (including two Asp residue of the DAVD motif), creating a highly negatively charged patch on the surface of the helix (Figure 2b). The abundance of Asp residues around the cleavage site might facilitate its recognition by caspase-3. Alternatively, the high density of an uncompensated negative charge might trigger a transient unfolding of the helix, which would facilitate not only the cleavage site recognition but also its digestion by caspase-3. Cleavage would likely to promote the unfolding of the entire Tudor-SN5 part and subsequent protein aggregation. In contrast, cleavage of spruce TSN in the loop regions (Figure 2c) is likely to generate soluble domains. Therefore mclI-Pa-mediated cleavage of spruce TSN at multiple sites may provide an alternative mechanism (to a cleavage at a single site within the $\alpha 2$ helix) to ensure efficient proteolytic inactivation of TSN.

As the list of annotated members of the caspase degradome continues to increase, it becomes evident that cleavage of many substrates is just a bystander event, which does not contribute to cell death development.⁵⁸ However, cleavage of human TSN by caspase-3 is important for the execution of apoptosis for the following reasons. The cleavage breaks up the hook-like structure formed by the Tudor and SN5 domains. An aromatic cage within the hook binds methyl groups of small nuclear ribonucleoproteins, anchoring TSN to the spliceosome and stimulating mRNA splicing.⁵⁹ Accordingly, caspase-mediated cleavage of TSN abrogates its stimulatory function in splicing.³³ As TSN is supposed to bridge transcription and splicing by interacting with the components of both processes, the cleavage by effector caspases will also uncouple these processes.

It is more difficult to explain the molecular mechanism underlying pro-cell death role of metacaspase-mediated cleavage of TSN, because the molecular role of TSN in plant biology remains poorly understood. Furthermore, plant TSN is a cytoplasmic protein,^{33,57} which does not support its involvement in transcription and splicing activation. Frei dit Frey *et al.* have recently demonstrated that TSN confers stress tolerance in *Arabidopsis* through selective stabilization of mRNAs encoding secreted proteins.⁵⁷ Notably, a significant proportion of these proteins are cysteine and serine protease inhibitors, which are known to suppress cell death in plants.^{62,63} Although the mechanism of TSN-dependent stabilization of specific mRNAs remains unknown, metacaspase cleavage of TSN molecule in at least four sites should result in deregulation of this mechanism. We propose a model, where intact TSN protein helps to protect cells from death by keeping expression of protease inhibitors at the level sufficient to suppress activation of a subset of cell-death proteases (Figure 2d, left). Contrary, metacaspase-mediated cleavage of TSN abrogates this function and promotes cell death through increased activation of cell-death proteases (including metacaspases) (Figure 2d, right). Experimental verification of this model will provide a fruitful avenue for the future research.

CONCLUSIONS

It has been approximately 10 years since metacaspases were discovered. During this time, the pendulum has shifted from naive hopes that metacaspases are responsible for cell death events and associated caspase-like activities in plants and other non-metazoans to well-grounded view that metacaspases are a distinct group of enzymes that can regulate (similar to caspases) various cell biological processes, including PCD. We expect that more mechanistic data will be obtained in the near future about metacaspases through integrated efforts of cell and structural biologists and biochemists. This will enable to resolve a number of important issues and controversies, including (but not limited to) (i) molecular conformation of active metacaspase and differences as well as similarities between type I and II metacaspases; (ii) whether activation of type I metacaspases occurs in multiprotein complexes similar to caspases-2, 8 and 9; (iii) whether the degradome of metacaspases is indeed much larger than that of caspases, considering a weak preference of metacaspases for specific residues located in P4-P2 and P1' positions in their substrates; (iv) is metacaspase-mediated proteolysis directly responsible for morphological changes that occur during PCD in nonmetazoan organisms, which differs significantly from apoptosis in animals; (v) whether metacaspase functions differ in the organisms containing a single metacaspase (e.g. budding or fission yeast) and multiple metacaspases (e.g. *Arabidopsis*) and (vi) how precisely do metacaspases regulate downstream aspartate-specific proteases.

Anticipated discovery of the natural substrates of metacaspases using advanced techniques of protease degradomics^{64,65} will assess and further consolidate metacaspase specificity in a broader biological context that will allow a more precise and rational design of synthetic substrates and inhibitors. Since metacaspases control vital cellular functions in protozoa, fungi and plants, development of approaches for artificial regulation of metacaspases can find important practical applications in the improvement of growth and disease resistance characteristics.

REFERENCES

1. Uren AG, O'Rourke K, Aravind L, Pisabarro MT, Seshagiri S, Koonin EV *et al.* Identification of paracaspases and metacaspases: Two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 2000; **6**: 961-967.
2. Vercammen D, Declercq W, Vandenabeele P, Van Breusegem F. Are metacaspases caspases? *J Cell Biol* 2007; **179**: 375-380.
3. Aravind L, Koonin EV. Classification of caspase-hemoglobinase fold: detection of new families and implications for the origin of eukaryotic separins. *Proteins* 2002; **46**: 355-367.
4. Szallies, A., Kubata, B. K., and Duszenko, M. A metacaspase of *Trypanosoma brucei* causes loss of respiration competence and clonal death in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* 2002; **517**: 144-150.
5. Vercammen D, van de Cotte B, De Jaeger G, Eeckhout D, Casteels P, Vandepoele K *et al.* Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. *J Biol Chem* 2004; **279**: 45329-45336.
6. Bozhkov PV, Suarez MF, Filonova LH, Daniel G, Zamyatnin AA Jr, Rodriguez-Nieto S *et al.* Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis. *Proc Natl Acad Sci USA* 2005; **102**: 14463-14468.
7. González IJ, Desponds C, Schaff C, Mottram JC, Fasel N. *Leishmania major* metacaspase can replace yeast metacaspase in programmed cell death and has arginine-specific cysteine peptidase activity. *Int J Parasitol* 2007; **37**: 161-172.
8. Watanabe N, Lam E. Calcium-dependent activation and autolysis of *Arabidopsis* metacaspase 2d. *J Biol Chem* 2011; in press.
9. Moss CX, Westrop GD, Juliano L, Coombs GH, Mottram JC. Metacaspase 2 of *Trypanosoma brucei* is a calcium-dependent cysteine peptidase active without processing. *FEBS Lett* 2007; **581**: 5635-5639.
10. Lee N, Gannavaram S, Selvapandiyan A, Debrabant A. Characterization of metacaspases with trypsin-like activity and their putative role in programmed cell death in the protozoan parasite *Leishmania*. *Eukaryot Cell* 2007; **6**: 1745-1757.
11. Ojha M, Cattaneo A, Hugh S, Pawlowski J, Cox JA. Structure, expression and function of *Allomyces arbuscula* CDP II (metacaspase) gene. *Gene* 2010; **457**: 25-34.
12. Watanabe N, Lam E. Two *Arabidopsis* metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. *J. Biol. Chem.* 2005; **280**: 14691-14699.
13. Vercammen D, Belenghi B, van de Cotte B, Beunens T, Gavigan JA, De Rycke R *et al.* Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* 2006; **364**: 625-636.
14. Bozhkov PV, Smertenko AP, Zhivotovsky B. Aspasing out metacaspases and caspases: Proteases of many trades. *Sci Signal* 2010; **3**: pe48.
15. Carmona-Gutierrez D, Fröhlich K-U, Kroemer G, Madeo F. Metacaspases are caspases. Doubt no more. *Cell Death Differ* 2010; **17**: 377-378.
16. Enoksson M, Salvesen GS. Metacaspases are not caspases – always doubt. *Cell Death Differ* 2010; **17**: 1221.

17. Madeo F, Herker E, Maldener C, Wissing S, Löchelt S, Herlan M *et al.* A caspase-related protease regulates apoptosis in yeast. *Mol Cell* 2002; **9**: 911-917.
18. Low CP, Shui G, Liew LP, Buttner S, Madeo F, Dawes IW *et al.* Caspase-dependent and -independent lipotoxic cell-death pathways in fission yeast. *J Cell Sci* 2008; **121**: 2671-2684.
19. Guerin R, Beauregard PB, Leroux A, Rokeach LA. Calnexin regulates apoptosis induced by inositol starvation in fission yeast. *PLoS One* 2009; **4**: e6244.
20. Ambit A, Fasel N, Coombs GH, Mottram JC. An essential role for the *Leishmania major* metacaspase in cell cycle progression. *Cell Death Differ* 2008; **15**: 113-122.
21. Zalila H, González IJ, El-Fadili AK, Delgado MB, Desponds C, Schaff C *et al.* Processing of metacaspase into a cytoplasmic catalytic domain mediating cell death in *Leishmania major*. *Mol Microbiol* 2011; **79**: 222-239.
22. Helms MJ, Ambit A, Appleton P, Tetley L, Coombs GH, Mottram JC. Bloodstream form *Trypanosoma brucei* depend upon multiple metacaspases associated with RAB11-positive endosomes. *J Cell Sci* 2006; **119**: 1105-1117.
23. Hamann A, Brust D, Osiewacz HD. Deletion of putative apoptosis factors leads to lifespan extension in the fungal ageing model *Podospora anserina*. *Mol Microbiol* 2007; **65**: 948-58.
24. Richie DL, Miley MD, Bhabhra R, Robson GD, Rhodes JC, Askew DS. The *Aspergillus fumigatus* metacaspases CasA and CasB facilitate growth under conditions of endoplasmic reticulum stress. *Mol Microbiol* 2007; **63**: 591-604.
25. Colabardini AC, De Castro PA, De Gouvêa PF, Savoldi M, Malavazi I, Goldman MH *et al.* Involvement of the *Aspergillus nidulans* protein kinase C with farnesol tolerance is related to the unfolded protein response. *Mol Microbiol* 2010; **78**: 1259-1279.
26. Coll NS, Vercaemmen D, Smidler A, Clover C, Van Breusegem F, Dangl JL *et al.* *Arabidopsis* type I metacaspases control cell death. *Science* 2010; **330**: 1393-1397.
27. Madeo F, Carmona-Gutierrez D, Ring J, Buttner S, Eisenberg T, Kroemer G. Caspase-dependent and caspase-independent cell death pathways in yeast. *Biochem Biophys Res Commun* 2009; **382**: 227-231.
28. Cao Y, Huang S, Dai B, Zhu Z, Lu H, Dong L *et al.* *Candida albicans* cells lacking CaMCA1-encoded metacaspase show resistance to oxidative stress-induced cell death and change in energy metabolism. *Fungal Genet Biol* 2009; **46**: 183-189.
29. Suarez MF, Filonova LH, Smertenko A, Savenkov EI, Clapham DH, von Arnold S *et al.* Metacaspase-dependent programmed cell death is essential for plant embryogenesis. *Curr Biol* 2004; **14**: R339-R340.
30. He R, Drury GE, Rotari VI, Gordon A, Willer M, Farzaneh T *et al.* Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H₂O₂ in *Arabidopsis*. *J Biol Chem* 2008; **283**: 774-783.
31. Jaspers P, Brosché M, Overmyer K, Kangasjärvi J. The transcription factor interacting protein RCD1 contains a novel conserved domain. *Plant Signal Behavior* 2010; **5**: 78-80
32. Watanabe N, Lam E. *Arabidopsis* metacaspase 2d is a positive mediator of cell death induced during biotic and abiotic stresses. *Plant J* 2011; in press.
33. Sundström JF, Vaculova A, Smertenko AP, Savenkov EI, Golovko A, Minina E *et al.* Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat Cell Biol* 2009; **11**: 1347-1354.
34. Kosec G, Alvarez VE, Aguero F, Sanchez D, Dolinar M, Turk B *et al.* Metacaspases of *Trypanosoma cruzi*: possible candidates for programmed cell death mediators. *Mol Biochem Parasitol* 2006; **145**: 18-28.
35. Zhang N-N, Dudgeon DD, Paliwal S, Levchenko A, Grote E, Cunningham KW. Multiple Signaling Pathways Regulate Yeast Cell Death during the Response to Mating Pheromones. *Mol Biol Cell* 2006; **17**: 3409-3422.
36. Vachova L, Palkova Z. Physiological regulation of yeast cell death in multicellular colonies is triggered by ammonia. *J Cell Biol* 2005; **169**: 711-717
37. Guscetti F, Nath N, Denko N Functional characterization of human proapoptotic molecules in yeast *S. cerevisiae*. *FASEB J* 2005; **19**: 464-466
38. Lee RE, Brunette S, Puente LG, Megeney LA. Metacaspase Yca1 is required for clearance of insoluble protein aggregates. *Proc Natl Acad Sci U S A* 2010; **107**: 13348-13353.
39. Lee RE, Puente LG, Kaern M, Megeney LA. A non-death role of the yeast metacaspase: Yca1p alters cell cycle dynamics. *PLoS One.* 2008; **3**: e2956.
40. Hsu SL, Yu CT, Yin SC, Tang MJ, Tien AC, Wu YM *et al.* Caspase 3, periodically expressed and activated at G2/M transition, is required for nocodazole-induced mitotic checkpoint. *Apoptosis* 2006; **11**: 765-771.
41. Wadskog I, Maldener C, Proksch A, Madeo F, Adler L. Yeast lacking the SRO7/SOP1-encoded tumor suppressor homologue show increased susceptibility to apoptosis-like cell death on exposure to NaCl stress. *Mol Biol Cell* 2004; **15**: 1436-1444.
42. Guaragnella N, Pereira C, Sousa MJ, Antonacci L, Passarella S, Côrte-Real M, *et al.* YCA1 participates in the acetic acid induced yeast programmed cell death also in a manner unrelated to its caspase-like activity. *FEBS Lett* 2006; **580**: 6880-6884.
43. Guaragnella N, Bobba A, Passarella S, Marra E, Giannattasio S. Yeast acetic acid-induced programmed cell death can occur without cytochrome c release which requires metacaspase YCA1. *FEBS Lett* 2010; **584**: 224-228.
44. Sripriya P, Vedantam LV, Podile AR. Involvement of mitochondria and metacaspase elevation in harpin Pss-induced cell death of *Saccharomyces cerevisiae*. *J Cell Biochem* 2009; **107**: 1150-1159.
45. Wu XZ, Chang WQ, Cheng AX, Sun LM, Lou HX. Plagiochin E, an antifungal active macrocyclic bis(bibenzyl), induced apoptosis in *Candida albicans* through a metacaspase-dependent apoptotic pathway. *Biochim Biophys Acta* 2010; **1800**: 439-447.
46. Bidle KD, Haramaty L, Barcelos E Ramos J, Falkowski P. Viral activation and recruitment of metacaspases in the unicellular cocolithophore, *Emiliania huxleyi*. *Proc Natl Acad Sci U S A* 2007; **104**: 6049-6054.
47. Bidle KD, Bender SJ. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom *Thalassiosira pseudonana*. *Eukaryot Cell* 2008; **7**: 223-236.
48. Murik O, Kaplan A. Paradoxically, prior acquisition of antioxidant activity enhances oxidative stress-induced cell death. *Environ Microbiol* 2009; **11**: 2301-2309.

49. Belenghi B, Romero-Puertas MC, Vercammen D, Brackener A, Inzé D, Delledonne M *et al.* Metacaspase activity of *Arabidopsis thaliana* is regulated by S-nitrosylation of a critical cysteine residue. *J Biol Chem* 2007; **282**: 1352-1358.
50. Helmersson A, von Arnold S, Bozhkov PV. The level of free intracellular zinc mediates programmed cell death/cell survival decisions in plant embryos. *Plant Physiol* 2008; **147**: 1158-1167.
51. Bozhkov PV, Filonova LH, Suarez MF, Helmersson A, Smertenko AP, Zhivotovsky B *et al.* VEIDase is a principal caspase-like activity involved in plant programmed cell death and essential for embryonic pattern formation. *Cell Death Differ* 2004; **11**: 175-182.
52. Yang J, Aittomäki S, Pesu M, Carter K, Saarinen J, Kalkkinen N, *et al.* Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. *EMBO J* 2002; **21**: 4950-4958.
53. Yang J, Välineva T, Hong J, Bu T, Yao Z, Jensen ON *et al.* Transcriptional co-activator protein p100 interacts with snRNP proteins and facilitates the assembly of the spliceosome. *Nucleic Acids Res* 2007; **35**: 4485-4494.
54. Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AM, Tops BB *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* 2003; **425**: 411-414.
55. Scadden ADJ. Inosine-containing dsRNA binds a stress-granule-like complex and downregulates gene expression in trans. *Mol Cell* 2007; **28**: 491-500.
56. Tong X, Drapkin R, Yalamanchill R, Mosialos G, Kieff E. The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. *Mol Cell Biol* 1995; **15**: 4735-4744.
57. Frei dit Frey N, Muller P, Jammes F, Kizis D, Leung J, *et al.* The RNA binding protein Tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in *Arabidopsis*. *Plant Cell* 2010; **22**: 1575-1591.
58. Pop C, Salvesen GS. Human caspases: activation, specificity, and regulation. *J Biol Chem* 2009; **284**: 21777-21781.
59. Shaw N, Zhao M, Cheng C, Xu H, Saarikettu J, Li Y *et al.* The multifunctional human p100 protein 'hooks' methylated ligands. *Nat Struct Mol Biol* 2007; **14**: 779-784.
60. Timmer JC, Zhu W, Pop C, Regan T, Snipas SJ, Eroshkin AM *et al.* Structural and kinetic determinants of protease substrates. *Nat Struct Mol Biol* 2009; **16**: 1101-1108.
61. Riedl SJ, Renucci M, Schwarzenbacher R, Zhou Q, Sun C, Fesik SW *et al.* Structural basis for the inhibition of caspase-3 by XIAP. *Cell* 2001; **104**: 791-800.
62. Solomon M, Belenghi B, Delledonne M, Menachem E, Levine A. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell* 1999; **11**: 431-444.
63. Coffeen WC, Wolpert TJ. Purification and characterization of serine proteases that exhibit caspase-like activity and are associated with programmed cell death in *Avena sativa*. *Plant Cell* 2004; **16**: 857-873.
64. Agard NJ, Wells JA. Method for the proteomics identification of protease substrates. *Curr Opin Chem Biol* 2009; **13**: 503-509.
65. Impens F, Colaert N, Helsens K, Plasman K, Van Damme P, Vandekerckhove J *et al.* MS-driven protease substrate degradomics. *Proteomics* 2010; **10**: 1284-1296.
66. Hao L, Goodwin PH, Hsiang T. Expression of a metacaspase gene of *Nicotiana benthamiana* after inoculation with *Colletotrichum destructivum* or *Pseudomonas syringae* pv. tomato, and the effect of silencing the gene on the host response. *Plant Cell Rep* 2007; **26**: 1879-1888.
67. Chahomchuen T, Akiyama K, Sekito T, Sugimoto N, Okabe M, Nishimoto S *et al.* Tributyltin induces Yca1p-dependent cell death of yeast *Saccharomyces cerevisiae*. *J Toxicol Sci* 2009; **34**: 541-545.
68. de Castro PA, Savoldi M, Bonatto D, Barros MH, Goldman MH, Berretta AA *et al.* Molecular characterisation of propolis-induced cell death in *Saccharomyces cerevisiae*. *Eukaryot Cell* 2011; **10**: 398-411.
69. Walter D, Matter A, Fahrenkrog B. Bre1p-mediated histone H2B ubiquitylation regulates apoptosis in *Saccharomyces cerevisiae*. *J Cell Sci* 2010; **123**: 1931-1939.
70. Ivanovska I, Hardwick JM. Viruses activate a genetically conserved cell death pathway in a unicellular organism. *J Cell Biol* 2005; **170**: 391-399.
71. Herker E, Jungwirth H, Lehmann KA, Maldener C, Fröhlich KU, Wissing S *et al.* Chronological aging leads to apoptosis in yeast. *J Cell Biol.* 2004; **164**: 501-7.
72. Palermo V, Falcone C, Calvani M, Mazzoni C. Acetyl-L-carnitine protects yeast cells from apoptosis and aging and inhibits mitochondrial fission. *Aging Cell* 2010; **9**: 570-579.

Research Results

067

Chapter 3

Identification of *Arabidopsis* METACASPASE 9 physiological substrates reveals its role in regulating the activity of PHOSPHOENOLPYRUVATE CARBOXYKINASE 1

Liana Tsiatsiani^{a,b,c,d}, Evy Timmerman^{c,d}, Pieter-Jan De Bock^{c,d}, Dominique Vercammen^e, Petra Van Damme^{c,d}, An Staes^{c,d}, Marc Goethals^{c,d}, Brigitte van de Cotte^{a,b}, Tine Beunens^{a,b}, Kris Gevaert^{c,d,*} and Frank Van Breusegem^{a,b,*}

^aDepartment of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium

^bDepartment of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

^cDepartment of Medical Protein Research, VIB, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^dDepartment of Biochemistry, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^eDevgen, Technologiepark 30, B-9052 Ghent, Belgium

Manuscript in preparation for submission

*correspondence

e-mail: frank.vanbreusegem@psb.vib-ugent.be

kris.gevaert@vib-ugent.be

AUTHOR CONTRIBUTIONS

FVB and KG conceived and supervised the project. LT conceived and designed the experiments, performed the research, analyzed the data and wrote the manuscript with the help of FVB and KG. ET and PJDB performed the MS/MS analysis. DV and PVD contributed to the experimental design. TB contributed to the data in Figure 1B. BvdC and MG contributed to the development of analytic tools. AS contributed to the computational analysis.

AIM AND CONTEXT

The *Arabidopsis thaliana* genome encodes for nine metacaspases and these are divided into two types (type I and II) based on sequence similarity and domain arrangement. Five out of six type II genes (*AtMC4-AtMC9*) originated through a segment duplication event containing the *AtMC8* gene that was followed by tandem duplications which gave rise to the *AtMC4-AtMC7* genes (Vercammen et al., 2004). Consequently, AtMC9 is the most distinct from the type II proteins with less than 40% amino acid sequence similarity to the other type II metacaspases (Watanabe and Lam, 2011). Besides its sequence divergence, we chose to study AtMC9 because, when starting this PhD project, this metacaspase was and still is the best biochemically characterized *Arabidopsis* metacaspase so far.

AtMC9 is highest expressed in mature seeds and flowers. Furthermore, transgenic plants with altered AtMC9 expression levels show differences in germination onset and seedling establishment. Our goal was to discover the functionality of AtMC9 during early vegetative stages through the identification of its protein substrates in a developmentally active stage, thus reflecting a dynamic proteome. For this purpose we used 2 days-old seedlings and we compared proteomes with different levels of AtMC9 activities, both *in vitro* and *in vivo*. Candidate substrates identified in this way were selected based on re-occurrence between independent proteome analyses and AtMC9 specificity (Arg and Lys). Further, through the identification of protein cleavage sites we were able to illustrate the AtMC9 cleavage specificity. Cleavage of some of the AtMC9 candidate substrates was confirmed *in vitro* using synthetic peptides or recombinant proteins. Finally, we showed that one of our candidate substrates, phosphoenolpyruvate carboxykinase 1 (PEPCK1), is *in vivo* cleaved by AtMC9 in plant extracts and this leads to the enhancement of PEPCK1 activity and improved seedling establishment.

Vercammen, D., van de Cotte, B., De Jaeger, G., Eeckhout, D., Casteels, P., Vandepoele, K., Vandenberghe, I., Van Beeumen, J., Inze, D., and Van Breusegem, F. (2004). Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. *J Biol Chem* 279, 45329-45336.

Watanabe, N., and Lam, E. (2011b). Calcium-dependent activation and autolysis of *Arabidopsis* metacaspase 2d. *J Biol Chem* 286, 10027-10040.

ABSTRACT

Metacaspases are distant relatives of metazoan caspases, found in plants, fungi and protists. In contrast to caspases, knowledge on physiological metacaspase substrates is scarce. We report for the first time on the proteome-wide substrate identification of a plant protease using positional proteomics and mass spectrometry. We used the N-terminal COFRADIC technology to explore the substrate repertoire of *Arabidopsis thaliana* metacaspase 9 (At5g04200, *AtMC9*) during early seedling development. This analysis provided additional insights into *AtMC9* substrate cleavage specificity and revealed metacaspase functions other than those related to programmed cell death. Stirred by the functionalities of the here identified *AtMC9* substrates, we observed a phenotype related to *AtMC9* loss-of-function. Dark-grown T-DNA insertion mutants are growth-compromised in the absence of sucrose. A phenotype indicative of *AtMC9* involvement in seedling establishment and gluconeogenesis that we afterwards verified by showing that phosphoenolpyruvate carboxykinase 1 (PEPCK1) is cleaved by *AtMC9 in vivo* and this leads to enhanced carboxylation activity of PEPCK1.

Key words: positional proteomics, *Arabidopsis thaliana*, metacaspase 9, seedling, protease degradomics, N-terminomics, substrate, phosphoenolpyruvate carboxykinase, gluconeogenesis

INTRODUCTION

Proteolysis is an irreversible, protease-catalyzed protein modification with diverse implications on protein function and cell or organismal fate. In eukaryote cells, proteases mediate protein stability, turnover and activity and also assist in correct targeting and import of proteins into different organelles. Several processes such as early development, xylem formation, seed maturation, nutrient supply mobilization, responses to biotic and abiotic stresses and programmed cell death are facilitated by diverse proteases (Avci et al., 2008; van der Hoorn et al., 2008; Coll et al., 2011; Vartapetian et al., 2011; Hara-Nishimura and Hatsugai, 2011; Pesquet et al., 2012). Given that proteases transmit molecular signals by cleaving their physiological substrates, identification of these substrates is essential for deciphering protease functions (Turk et al., 2012). Despite the numerous proteases encoded in plant genomes (2.5% of the *Arabidopsis* genes), their substrates remain poorly identified (Tsiatsiani et al., 2012). This knowledge gap is expected to be filled with the use of emerging technologies that allow for proteome-wide identification of protease substrates (Gevaert et al., 2003; Van Damme et al., 2010; Kleifeld et al., 2010; Schilling et al., 2010; Staes et al., 2011).

Plant programmed cell death (PCD), a protease-facilitated process, is classified into two major classes; vacuolar cell death and necrosis. Plant PCD cannot be compared to mammalian apoptosis (van Doorn et al., 2011), which can be mediated by caspases. The aspartate specific caspases belong to a family of evolutionarily conserved cysteinyl proteases, which, besides apoptosis, also mediate inflammation, cell proliferation and differentiation by acting on a large number of different cellular substrates (Kumar, 2007; Lamkanfi et al., 2007; Timmer and Salvesen, 2007). Based on sequence and structural similarity to the caspase catalytic domain, metacaspases were identified in fungi, protozoa and plants, while paracaspases were found in metazoans and in the slime mold *Dictiostelium discoideum*, the latter more probably being a metacaspase (Aravind et al., 1991, Uren et al., 2000; Vercammen et al., 2007). Like caspases, meta- and paracaspases carry the hallmarks of the CD clan of cysteine proteases; the catalytic dyad His/Cys and the hemoglobinase fold (Rawlings and Barrett, 1993; Aravind and Koonin, 2002).

Given the well-established role of caspases in apoptotic death of mammalian cells, a lot of research was initially focused on assessing whether metacaspases are responsible for the caspase-like activities detected in plants during PCD (Bonneau et al., 2008). Several biochemical studies using recombinant metacaspases or protein extracts from loss or gain-of-function mutants clearly demonstrated that metacaspases, in contrast to caspases, are highly specific for Arg or Lys residues (Vercammen et al., 2004; Watanabe and Lam, 2005; Bozhkov et al., 2005; Vercammen et al., 2006, Gonzalez et al., 2007; Moss et al., 2007; Ojha et al., 2010). Therefore, it is very unlikely that metacaspases are directly responsible for the caspase-like (Asp-specific) activities found in plants. Further, in the protozoan parasites *Plasmodium falciparum* and *Leishmania major* it was recently shown that metacaspases probably act upstream of such proteases (Meslin et al., 2011).

Aside from the multifunctionality of metacaspases, which is mostly apparent in organisms encoding single metacaspase genes such as *Saccharomyces cerevisiae* and *Leishmania major*, redundant and antagonistic functions have been demonstrated in organisms encoding multiple metacaspases. Yeast metacaspase YCA1 is required for various stress-induced cell death events including oxidative, osmotic, toxicity- and virus-induced stresses and aging (Madeo et al., 2002, Ivanovska and Hardwick, 2005; Herker et al., 2004; Chahomchuen et al., 2009). YCA1 is also involved in cell cycle regulation (Lee et al., 2008) and clearance of insoluble protein aggre-

gates (Lee et al., 2010). *Leishmania* LmjMCA is involved in cell cycle regulation and in oxidative stress induced cell death (Ambit et al., 2008; Zalila et al., 2011). Three of five *Trypanosoma brucei* metacaspases are required for viability of the bloodstream form of the parasite (Helms et al., 2006). Redundant functions were also shown for two *Podospira anserina* metacaspases in regulating senescence-associated death (Hamann et al., 2007), and for two *Aspergillus fumigatus* metacaspases in the resistance to endoplasmic reticulum (ER) stress (Richie et al., 2007). Interestingly, antagonistic functions for homologous metacaspases were demonstrated in *Trypanosoma brucei* and *Arabidopsis thaliana* in controlling cell death events or cell proliferation and differentiation (Coll et al., 2011; Laverrière et al., 2012).

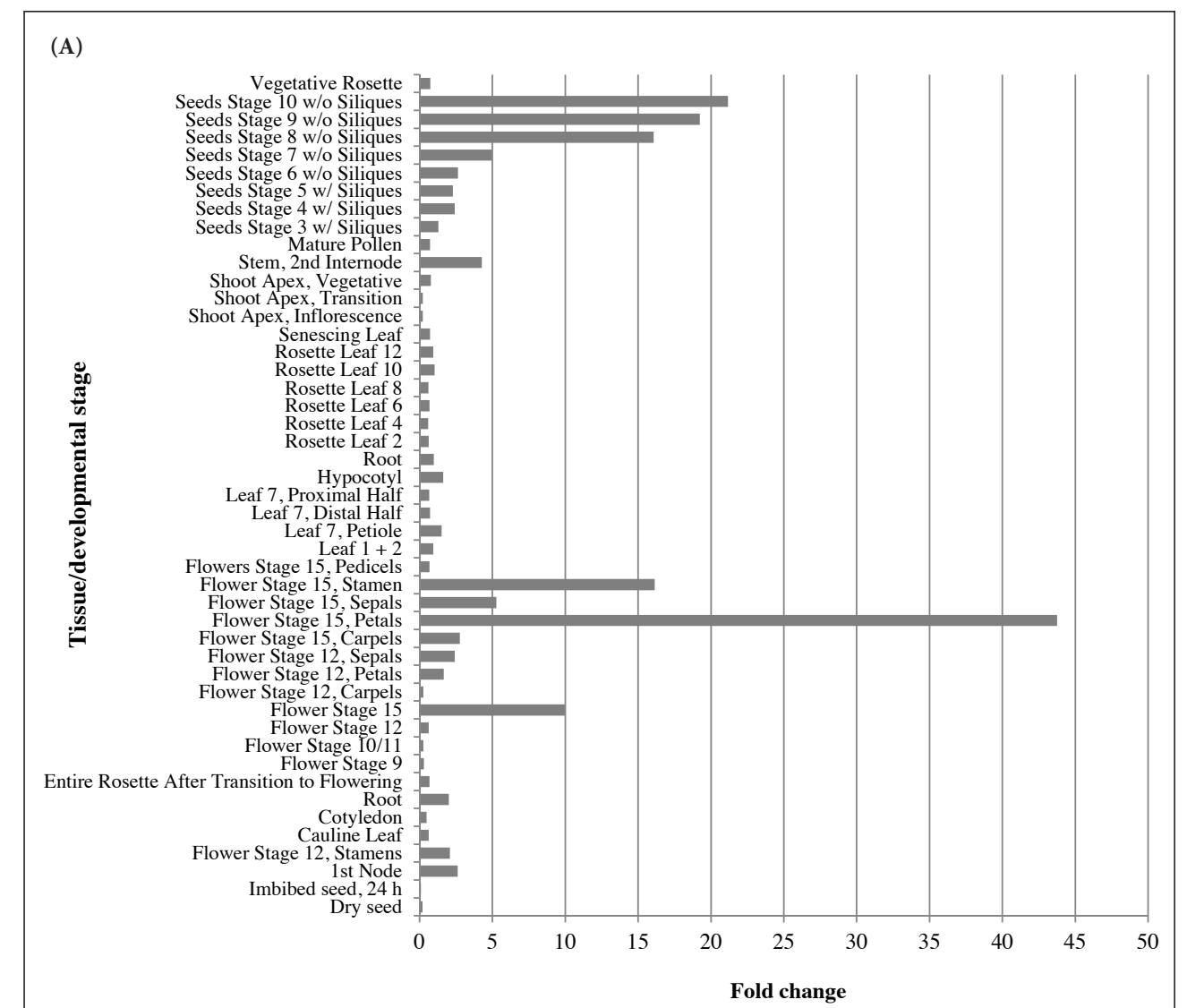
Metacaspases are known to be post-translationally regulated by multiple mechanisms and their activity is dependent on specific cellular conditions. For instance, the proteolytic activity of *Arabidopsis* AtMC9 was found to be regulated by S-nitrosylation, by a serine protease inhibitor (AtSerpin 1) and by auto-catalytic processing (Vercammen et al., 2004; Vercammen et al., 2006; Belenghi et al., 2007). Further, the activities of the Norway spruce mcII-Pa and the *Arabidopsis* AtMC1, AtMC4 and AtMC5 depend on Ca²⁺ (Bozhkov et al., 2005; Watanabe and Lam, 2005; Watanabe and Lam, 2011a). Recently, a *Trypanosoma* metacaspase (TbMCA4) with no proteolytic activity, yet necessary for parasitic proliferation and virulence was found to be processed by another metacaspase, TbMCA3 (Proto et al., 2011). At the onset of the second decade in metacaspase research, the multifunctionality, redundancy, antagonism and regulation of these proteases has become apparent although until now only a handful of substrates are identified (Tsiatsiani et al., 2011).

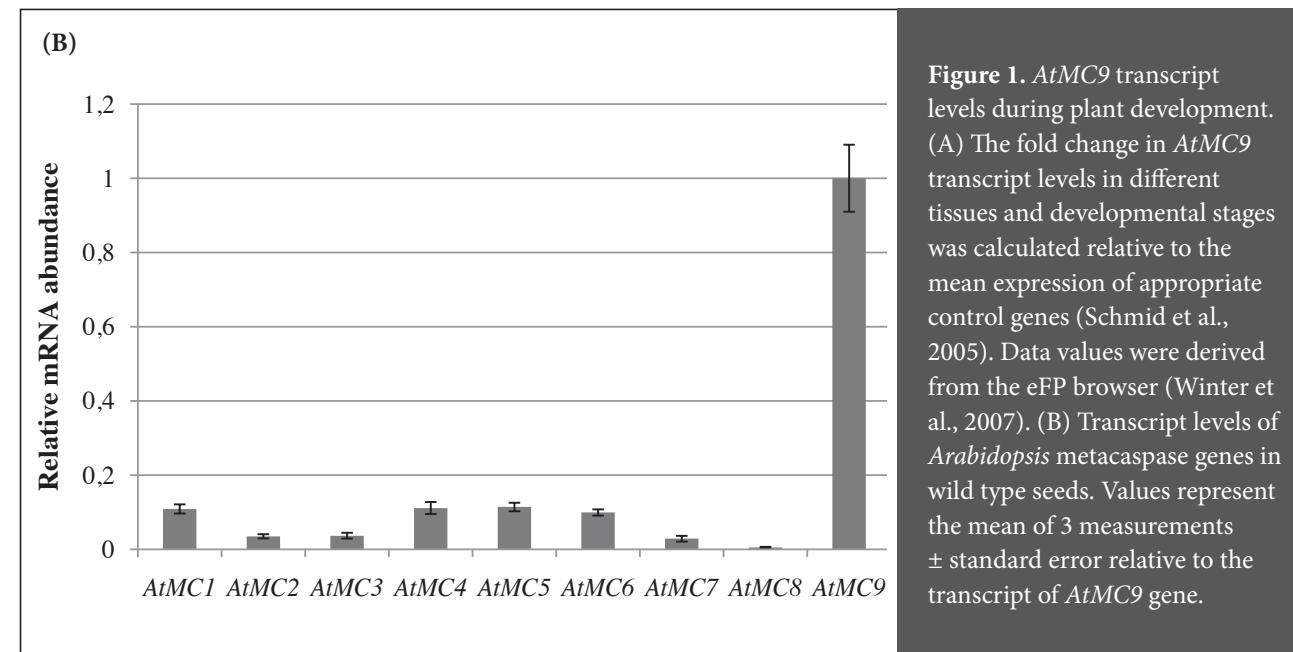
In the present study we assessed the degradome of one of the nine *Arabidopsis thaliana* metacaspases, metacaspase 9 (AtMC9, At5g04200), in young seedlings using N-terminal COFRADIC. We report for the first time on the proteome-wide identification of physiological substrates of a metacaspase and their cleavage sites using positional proteomics. These data allowed us to resolve AtMC9's substrate specificity in more detail and show that phosphoenolpyruvate carboxykinase 1 (*PEPCK1*, At4g37870), a key enzyme in gluconeogenesis, is a physiological AtMC9 substrate whose activity is enhanced upon cleavage.

RESULTS

AtMC9 is expressed in flowers and developing seeds

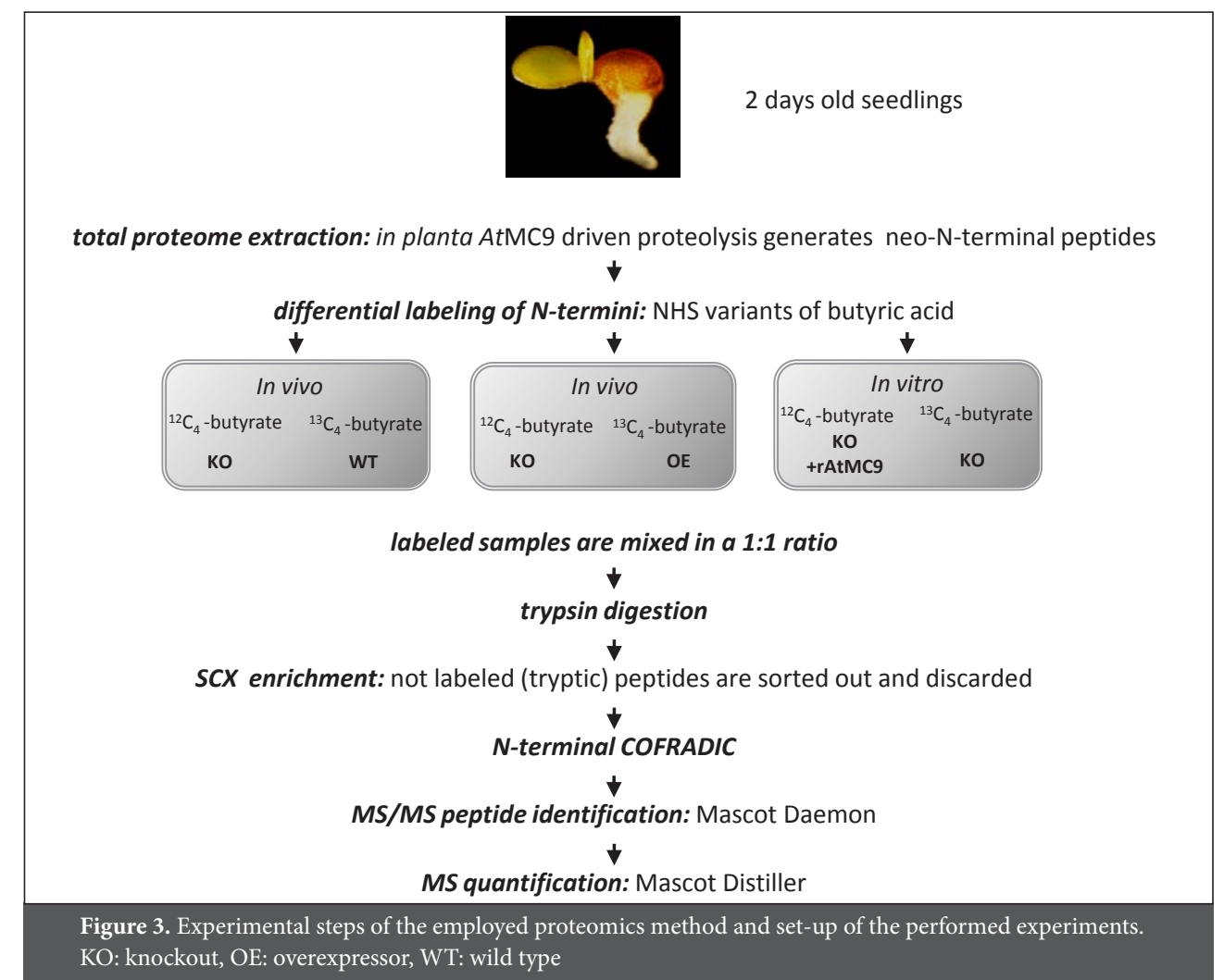
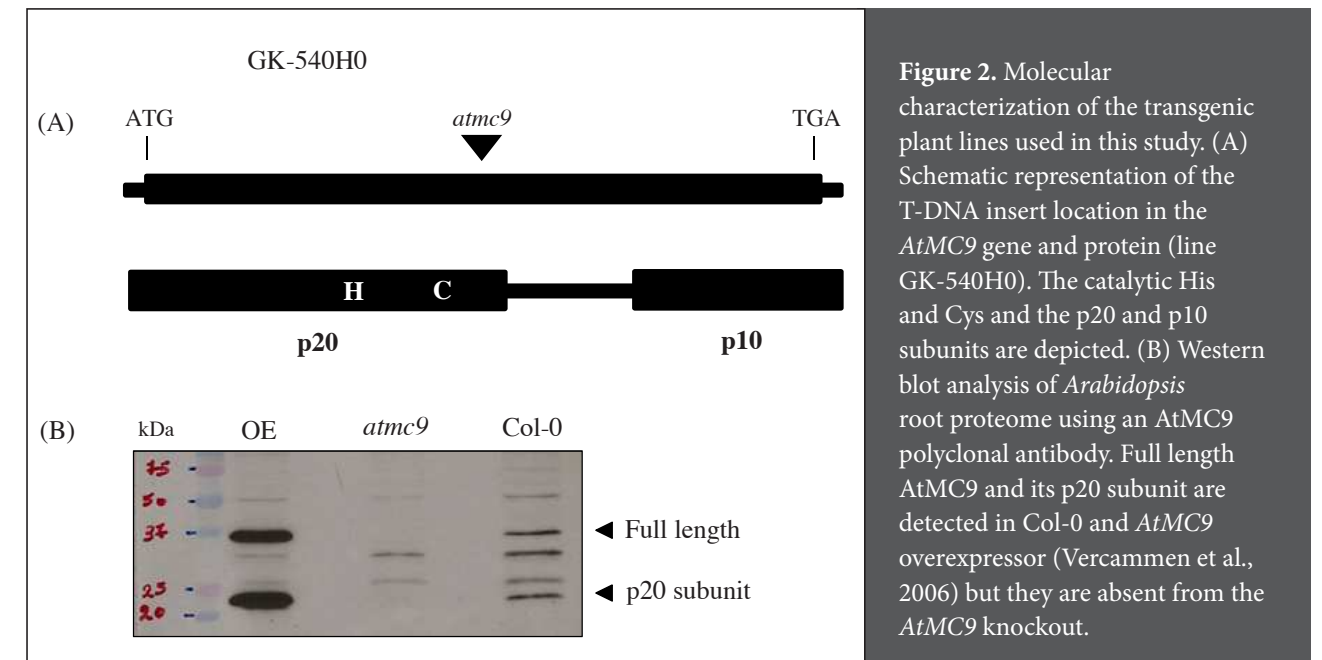
AtMC9 (At5g04200) expression characteristics were first assessed within publicly available microarray data sets using the eFP browser (Winter et al., 2007), and a clear increase in *AtMC9* transcript levels in flowers and developing seeds was found. In developing seeds, more than a 15-fold increase in transcript levels can be observed at seed stage 8 (Figure 1A) which is the stage in which the embryo is between a walking-stick shape to early curled cotyledons (Bowman, 1994). As seen in Figure 1A, *AtMC9* transcript levels further increase during later stages of seed development when cotyledons turn green (seed stages 9 and 10). *AtMC9* expression was validated by quantitative RT-PCR on dry seeds and compared to the transcript levels of the other eight *Arabidopsis* metacaspases (Figure 1B). *AtMC9* expression was found at least 6-fold higher than any other metacaspase in dry seeds. Based on the high *AtMC9* expression in seeds, we chose to analyze the proteome of germinating seeds/young seedlings, representing as well a metabolically active developmental stage.





Identification of *AtMC9* protein processing sites by N-terminal COFRADIC

To better understand the function of *AtMC9* during early vegetative stages, we set out to identify physiological *AtMC9* protein substrates. We used the N-terminal COFRADIC technique by which peptides holding protein N-termini (N-terminome) are enriched prior to LC-MS/MS analysis (Gevaert et al., 2003; Staes et al., 2011). The N-terminome consists of all peptides that contain either mature protein N-termini or neo-N-termini that are generated upon proteolysis. Neo-N-termini that sprout at sites of *AtMC9* processing directly report *AtMC9* proteolytic events and the exact site where protein cleavage has occurred (Van Damme et al., 2005). By comparative analysis of wild-type proteome and proteomes in which *AtMC9* activity was perturbed in both gain- and loss-of-function transgenics (Figure 2), we were able to identify *in vivo* *AtMC9* protein substrates. To assess *AtMC9* substrates in a metabolically active developmental stage following the seed stages in which *AtMC9* transcript levels are high, we explored the N-terminome of young seedlings harvested two days following seed germination (developmental stage 0.7; Boyes et al., 2001). First, we compared the N-terminome of *AtMC9* T-DNA insertion knockout (KO) seedlings to wild type Col-0 (WT) seedlings, while in a second *in vivo* study we compared the N-terminome of KO seedlings to that of seedlings overexpressing *AtMC9* (OE). In addition, we performed an *in vitro* study in which a KO seedling proteome was supplemented with recombinant *AtMC9* (r*AtMC9*) (Vercammen et al., 2004) and compared to an untreated KO control. To comparatively analyse the N-terminomes, N-hydroxysuccinimide (NHS) esters of isotopic variants of butyric acid were used to mass tag (neo) N-terminal peptides. This differential labeling strategy is possible since *AtMC9* introduces novel, primary alpha-amino groups which are thus differentially tagged with a light or heavy variant of butyric acid. Following this mass tagging, equal amounts of the proteome preparations were mixed, digested with trypsin and (neo) N-terminal peptides were isolated using N-terminal COFRADIC (Figure 3). Mass tagging of neo-N-termini enabled to determine the origin of each labeled peptide, and their respective ion signal intensities were further used to identify *AtMC9* processed sites.

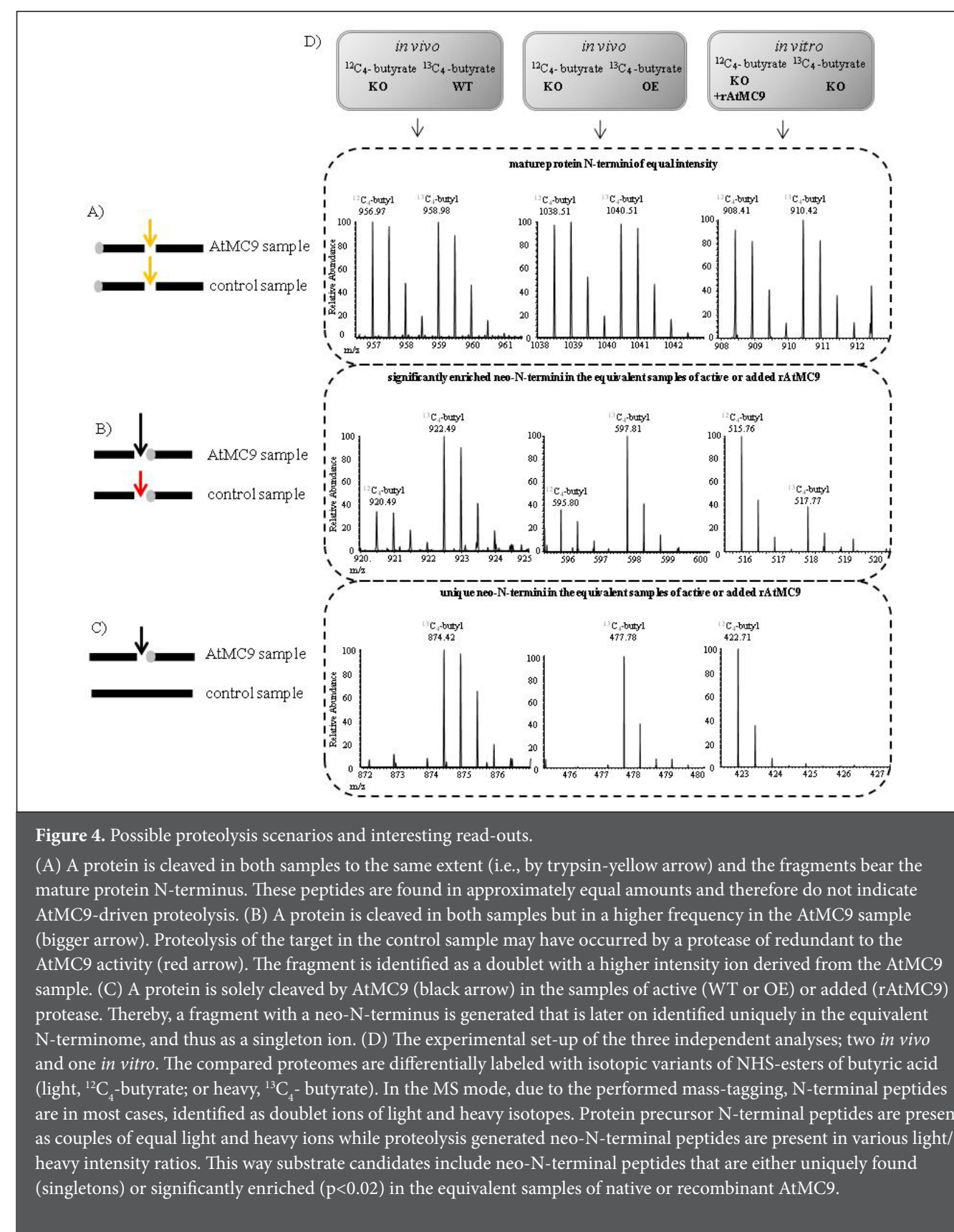


After comparison of the different N-terminomes, the following read-outs were expected: a) A (neo) N-terminal peptide is found in equal amounts in both samples. Such peptides may carry the mature protein N-terminus and are not indicative of AtMC9 proteolysis (Figure 4A). Furthermore, such N-terminal peptides indicate that the particular protein (form) was not up- or down-regulated in the tested proteomes. b) A neo-N-terminal peptide is found in both samples, but with significantly higher intensity in the sample with either active AtMC9 or supplemented rAtMC9, which indicates that the corresponding protein was also cleaved in the control sample (no AtMC9 activity), although to a lesser extent, by another protease showing redundant activity to AtMC9 (Figure 4B). Such events can be explained by the following two scenarios: AtMC9 perturbation leads to the activation of another metacaspase or these particular substrates are cut by other proteases (i.e. metacaspases) whose activity is enhanced by AtMC9 and further share the canonical AtMC9 specificity C-terminal to Arg or Lys (Vercammen et al., 2004). c) A neo-N-terminal peptide is uniquely found (singleton) in the proteome either with active AtMC9 (in a WT or OE proteome) or supplemented with rAtMC9, implying that the corresponding protein was cleaved by AtMC9 (Figure 4C).

Peptides were analyzed using a LTQ Orbitrap XL mass spectrometer and the generated MS/MS-spectra were searched against the TAIR protein database (The Arabidopsis Information Resource, release 8) using the MASCOT algorithm, thereby matching spectra and peptides to their originating proteins. In total, 75,310 fragmentation spectra were generated in the KO/WT analysis, 69,388 in the KO/OE analysis and 75,042 in the *in vitro* analysis that were subsequently used for the identification of 3,781 peptides in the KO/WT analysis, 2,879 in the KO/OE analysis and 3,050 in the *in vitro* analysis. Finally, the identified peptides were converged to 1,705 proteins in the KO/WT analysis, 1,407 proteins in the KO/OE analysis and 1,138 in the *in vitro* analysis (Table 1). Further, the MASCOT Distiller software was used to calculate the ratio of light/heavy peptide signals for all peptides identified and, for possible neo-N-terminal peptides pointing to AtMC9 processing, these ratios were individually checked. We focused on two classes of peptides; neo-N-terminal peptides that were unique (singletons) or significantly (p -value<0.02) enriched in the WT, OE or rAtMC9-treated seedlings proteomes in comparison to KO (Figure 4D). From all processed sites, 72% agreed with the expected AtMC9 specificity (28% cleavage C-terminal to Arg and 44% C-terminal to Lys), with no significant enrichment for a particular amino acid in the other processing events. In total 551 cleavage events after Arg or Lys (116 in KO/WT, 103 in KO/OE and 332 *in vitro*) in 392 proteins (97 in KO/WT, 84 in KO/OE and 211 *in vitro*) were identified. From those, approximately one third (160 sites) were reported by singleton peptides in either WT, OE or rAtMC9 treated proteomes (16 in KO/WT, 11 in KO/OE and 133 *in vitro*) and 391 were reported by significantly enriched peptides in the same proteomes (100 in KO/WT, 92 in KO/OE and 199 *in vitro*). This latter class consisted out of enriched proteolysis-reporter peptides that were 2-fold to more than 100-fold increased. An overview of all the cleavage sites and the corresponding proteins found in each of the three analyses is given in the supporting information, tables S1, S2 and S3. The number of cleaved proteins identified in the *in vitro* analysis was clearly much higher. Reasons for this could be that in such an experimental setting the substrate/protease co-localization prerequisite for cleavage is circumvented and that the added recombinant AtMC9 is present in higher levels than endogenous levels. Nevertheless, such *in vitro* data are very informative about AtMC9 positional substrate specificity (Schilling et al., 2008).

Table 1. Overview of the results of the proteome analyses performed. Numbers of generated fragmentation spectra, identified peptides and their matched proteins are listed.

Analysis	Fragmentation spectra	Peptides	Proteins
KO/WT	75,310	3,781	1,705
KO/OE	69,388	2,879	1,407
KO +/- rAtMC9	75,042	3,050	1,138



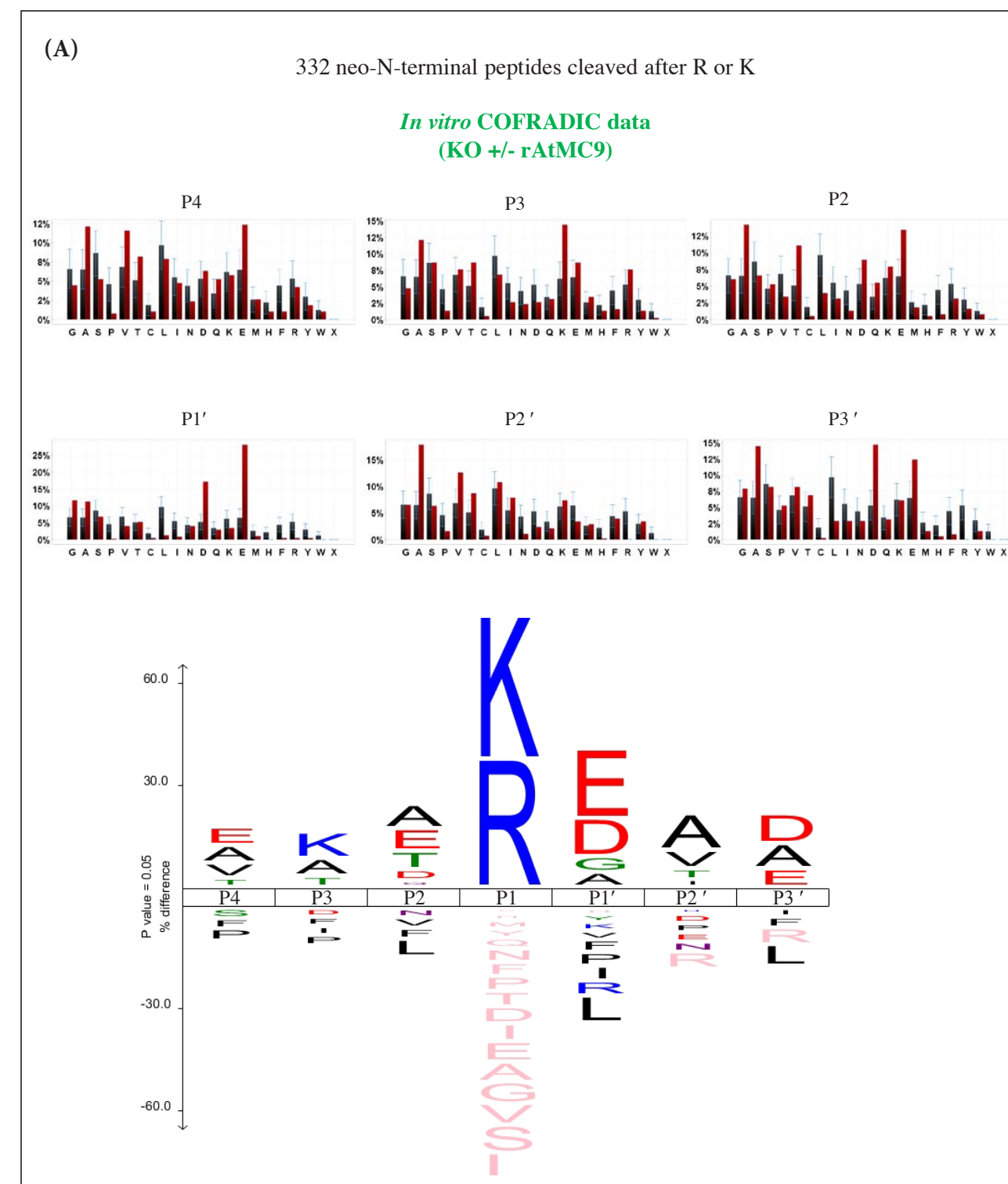
AtMC9 positional amino acid specificity for cleavage

We selected candidate AtMC9 substrates based on the AtMC9 specificity for Arg or Lys. Using Icelogo (Colaert et al., 2009) we inspected the amino acid sequences surrounding the identified cleavage sites and we analyzed the frequencies of specific residues at the prime and non-prime substrate positions (C- and N-terminally to the scissile bond) after statistical correction using the natural occurrence of amino acids in *Arabidopsis* proteins. In general, the specificity information derived from all three datasets was very similar. The larger *in vitro* COFRADIC dataset showed that the acidic residues Asp and Glu were often found at neighbouring positions to the basic (Arg or Lys) cleavage site. For instance, in 46% of the Lys-cleaved sequences and in 47% of the Arg-cleaved sequences, P1' position (C-terminal to the cleavage P1 position) was occupied by Asp or Glu. Likewise, in 23% of the Lys-cleaved and in 22% of the Arg-cleaved sequences, P2 was Asp or Glu. However, only in 5% of the Lys-cleaved and in 7% of the Arg-cleaved sequences, both P2 and P1' were acidic (Figure 5A, 5C and supporting information Figure S1). Further at the prime site, Asp or Glu were also observed in 27% of the cases at P3'. At P2', the small amino acids Thr, Ala and Val were found in 38% of all cases. Furthermore, Ala was tolerated at every position. Analogous to the *in vivo* COFRADIC derived specificity (Figure 5B), Val and Glu were preferred at P4, while Lys and Arg were frequently found at P3. Ala and Thr were the sole preferred residues at P2 in the *in vivo* cleaved substrates, while *in vitro* cleaved proteins additionally (23% of the cleaved substrates) allowed Asp and Glu (Figure 5A-B and supporting information Figure S1). Concerning the prime site specificity of AtMC9, *in vivo* and *in vitro* data largely agree apart from amino acid tolerances at P2'; in the *in vitro* dataset small residues are found at this position, whereas aromatic residues such as Phe and Tyr and the aliphatic Ile and Leu were additionally found in the *in vivo* cleaved proteins.

In conclusion, AtMC9 specificity for basic residues at P1 comes along with a strong preference for acidic amino acids at the C-terminal side of the scissile bond. This particular trend of acidic-basic amino acid combinations was also seen at other positions at the non-prime side, while Ala was generally tolerated in every position. A relative preference of Lys over Arg at P1 was found, and *in vivo* often (27% of the cases) a second Arg or Lys was found at P3. For instance, when Lys was found at P1, 24% of those substrates contained a second Lys at P3.

In a previous study using a positional-scanning synthetic combinatorial library (PS-SCL) of AMC labeled synthetic tetrapeptides spanning positions P4-P1, VRPR was found as the best cleaved tetrapeptide peptide by rAtMC9. IISK, the best cleaved peptide with Lys at P1, was five-fold less efficiently cleaved than VRPR (Vercammen et al., 2006). Our current results somehow conflict with the proposed stronger preference for Arg over Lys at P1, but they agree with the presence of a basic amino acid motif (e.g., RxR) as in VRPR. As seen in Figure 5C, where we compare the two *in vitro* approaches (*in vitro* COFRADIC and PS-SCL), Val is mostly found at P4 in both datasets. However, a major difference was observed at the P2 site; COFRADIC data indicated Thr and Asp to be preferred, while PS-SCL data indicated a preference for hydrophobic residues or Pro. One obvious explanation for these differences is that with COFRADIC, processing of native proteins is assessed and interaction of the protease with its substrates takes place under physiological conditions, whereas in PS-SCL small synthetic tetrapeptides are used (Turk et al., 2012). Moreover, for the calculation of PS-SCL scores the natural amino acid occurrence was not taken into account. As shown by our specificity data, sequences C-terminal to the cleavage bond play an important role in steering AtMC9 specificity and this could not have been detected previously due to the limitation of PS-SCL which only screened sequences N-terminally to the cleavage bond.

Based on the *in vivo* and *in vitro* observed amino acid frequencies at the P4-P3' substrate positions (Figure 5), we propose two revised consensus sequences of AtMC9 substrate specificity for either cleavage after Arg; [V]-[R]-[T/Q]-[R]-↓-[G/D/E]-[V/F/L/I]-[D/E] or Lys; [E/Q]-[K/A]-[T/E]-[K]-↓-[D/E]-[Y/T]-[D/E] for the development of biochemical assays.



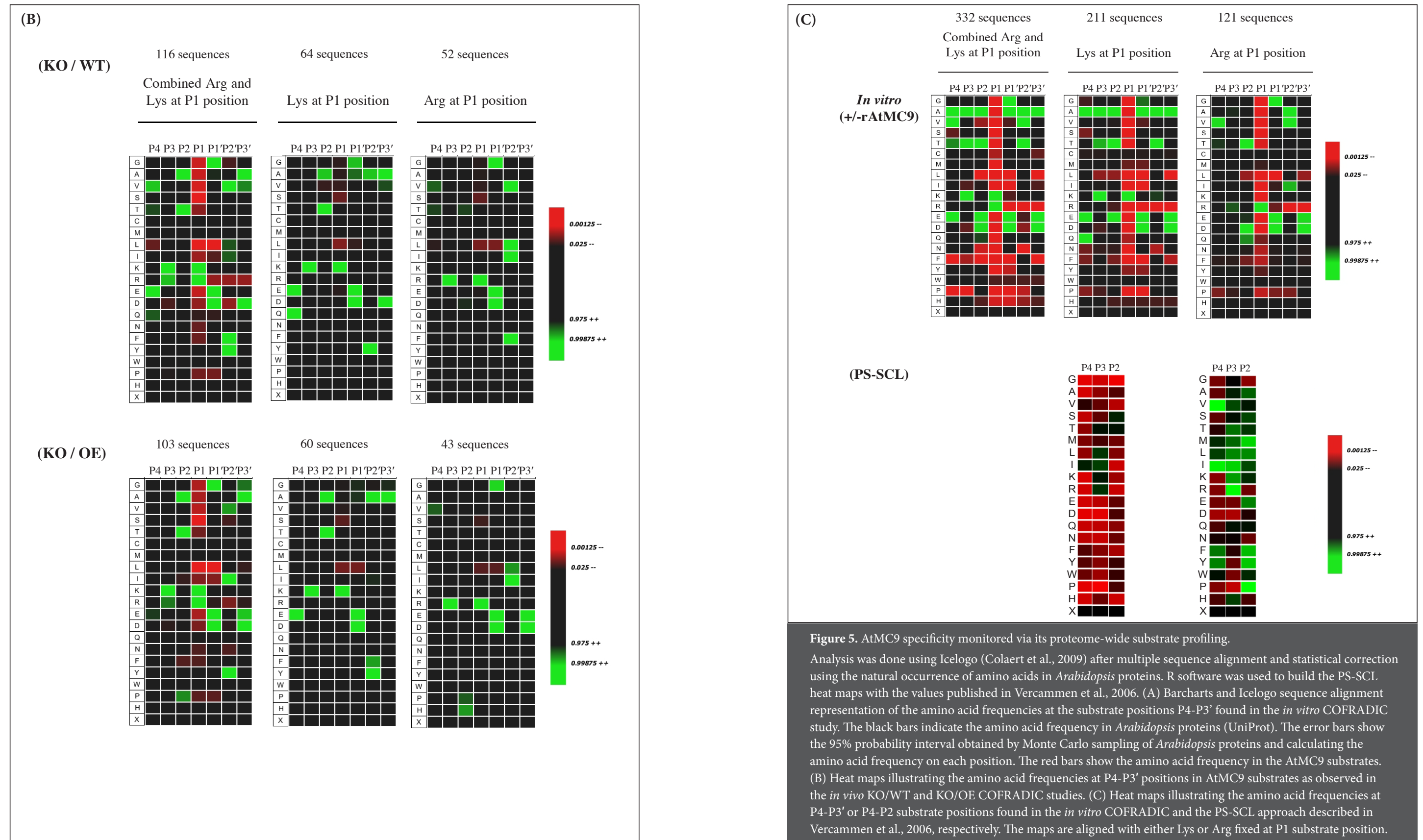


Table 2. Selection of proteolysis reporter peptides from at least two independent analyses, including all the *in vivo* singleton peptides. Amino acid positions of P1' cleavage sites are shown between brackets in the P4-P3' column.

#	Accession	P4-P3'	Sequence	Description	Shared between experiments
1	AT1G02150	VTSK(474)EAV	EAVLELLR	pentatricopeptide repeat (PPR)-containing protein	singleton KO/OE. Not Shared
2	AT1G04750	VLDR(156)GEK	GEKIELLVDKTENLR	vesicle-associated membrane protein 721	both in vivo
3	AT1G07780	SVAK(109)DIS	DISKVAR	phosphoribosylanthranilate isomerase 1	both in vivo
4	AT1G09310	TEAK(165)EAV	EAVAIKEAVAVKEAA	Protein of unknown function, DUF538	both in vivo
5	AT1G11480	ARPR(404)ELV	ELVLKER	eukaryotic translation initiation factor-related	both in vivo
6	AT1G12000	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	all three analyses
7	AT1G13270	RKMR(78)ELE	ELETKSKVR	methionine aminopeptidase 1B	KO/OE and in vitro
8	AT1G29350	QNAR(525)ELD	ELDFQYSPFSAQQMSQR	Kinase-related protein of unknown function (DUF1296)	both in vivo
9	AT1G47540	AYDR(31)KCL	KCLKEYGGDVGFSCAPR	trypsin inhibitor 2	both in vivo
	AT1G47540	CAPR(49)IFP	IFPTFCDQNCR	trypsin inhibitor 2	KO/WT and in vitro
10	AT1G49240	MNOK(53)DAY	DAYVGVDEAQSQR	actin 8	all three analyses
	AT1G49240	SIEK(241)NYE	NYELPDGQVITIGAER	actin 8	both in vivo
11	AT1G49970	FEKR(363)DYD	DYDGTLAQR	CLP protease proteolytic subunit 1	both in vivo
12	AT1G51710	VATR(157)ELF	ELFGELDR	ubiquitin-specific protease 6	both in vivo
13	AT1G54870	EKIK(288)NFG	NFGSEVPMKR	Aldehyde reductase	KO/OE and in vitro
14	AT1G55090	NSNK(710)EIG	EIGVVAANSQDPSAGL	NAD synthetase	singleton KO/OE. Not Shared
15	AT1G55490	TVAR(114)EVE	EVELEDPVENIGAKLVR	chaperonin 60 beta	both in vivo
	AT1G55490	VLTR(382)ETS	ETSTIVGDGSTDVAVKRR	chaperonin 60 beta	KO/WT and in vitro
16	AT1G56070	ICLK(544)DLQ	DLQDDFMGGAEIKSDPVVSR	Ribosomal protein S5/Elongation factor G/III/V family protein	both in vivo
17	AT1G65090	TSER(162)TLQ	TLQDDKKSNGAKSEEVQEQPEKR	unknown protein	KO/WT and in vitro
18	AT1G72370	LGTK(33)NCN	NCNYQMER	40s ribosomal protein SA	KO/OE and in vitro
19	AT2G01250	VESK(6)VVV	VVVPESVLKRR	Ribosomal protein L30/L7 family protein	both in vivo
20	AT2G18960	GSYR(902)ELS	ELSEIAEQAKR	H(+)-ATPase 1	both in vivo
21	AT2G19730	SVNK(88)SIL	SILKKEFPR	Ribosomal L28e protein family	both in vivo
22	AT2G21060	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	all three analyses
23	AT2G21660	KAMK(64)DAI	DAIEGMNGQDLDR	glycine-rich, RNA binding protein AtGRP7	KO/OE and in vitro
	AT2G21660	TDDR(23)ALE	ALETAFAYQGDVIDSKIINDR	glycine-rich, RNA binding protein AtGRP7	KO/WT and in vitro
24	AT2G26250	IVNR(18)GIE	GIEPSGPNAGSPTFSVR	3-ketoacyl-CoA synthase 10	both in vivo
25	AT2G28000	TIAR(103)AIE	AIELPNAMENAGAALIR	chaperonin-60alpha	both in vivo
	AT2G28000	PRGR(82)NVV	NVVLDEFGSPKVVNDGVTIAR	chaperonin-60alpha	both in vivo
26	AT2G32240	VKSR(709)DID	DIDLSFSSPTR	unknown protein	both in vivo
27	AT2G36620	NIPK(149)SAA	SAAPKAAKMGGGGGR	ribosomal protein L24	both in vivo
28	AT2G36640	QKTR(338)EST	ESTESGAQKAETKDSAAVR	embryonic cell protein 63	KO/WT and in vitro
29	AT2G40490	VTER(47)KVS	KVSATSEPLLLR	Uroporphyrinogen decarboxylase	KO/WT and in vitro
30	AT2G42560	VGSK(454)AVD	AVDLTKEKAAVAADTVVGYTAR	LEA domain-containing protein	all three analyses
	AT2G42560	ARAK(324)DYT	DYTLQKAVEAKDVAEKAQR	LEA domain-containing protein	both in vivo
	AT2G42560	VEAK(335)DVA	DVAEKAQR	LEA domain-containing protein	KO/OE and in vitro
	AT2G42560	EKGK(270)EAG	EAGNMTAEQAAR	LEA domain-containing protein	KO/OE and in vitro
	AT2G42560	MRER(214)EGK	EGKESAGGVGGR	LEA domain-containing protein	KO/OE and in vitro
	AT2G42560	ATEK(268)GKE	GKEAGNMTAEQAAR	LEA domain-containing protein	KO/OE and in vitro
	AT2G42560	QTTK(607)NIV	NIVIGDAPVR	LEA domain-containing protein	KO/WT and in vitro
31	AT2G42790	KEDK(86)GLK	GLKLYDPGYLNTAPVR	Citrate synthase 3, CSY3	singleton KO/WT. Not shared

32	AT2G47470	YIEK(320)GSD	GSDYASKETER	thioredoxin family protein	singleton KO/WT. Not shared
33	AT3G02480	MKEK(51)AQG	AQGAADVVKDKTGMNKSH	LEA family protein	all three analyses
	AT3G02480	QQMK(49)EKA	EKAQGAADVVKDKTGMNKSH	LEA family protein	KO/WT and in vitro
34	AT3G09840	KSKK(15)DFS	DFSTAILER	cell division cycle 48	all three analyses
35	AT3G11710	QTTK(11)ALS	ALSELAMDSSTLLNAEESAGDGAGPR	lysyl-tRNA synthetase 1	KO/WT and in vitro
36	AT3G12960	EKGK(58)EQS	EQSAASGDQTIQOR	unknown protein	KO/WT and in vitro
37	AT3G13300	VGER(708)NLD	NLDVSSVEEISR	Transducin/WD40 repeat-like superfamily protein	both in vivo
38	AT3G13470	VLTK(378)EMT	EMTIVGDGTTQEAVNKR	TCP-1/cpn60 chaperonin family protein	KO/WT and in vitro
39	AT3G15280	TEYR(135)GVE	GVEDLHQQTGGVEKSP	unknown protein	KO/WT and in vitro
40	AT3G15670	IKNK(134)AQD	AQDAAQYTKETAQGAAYTKETAEAGR	LEA family protein	KO/WT and in vitro
	AT3G15670	QYTK(143)ETA	ETAQGAAYTKETAEAGR	LEA family protein	KO/WT and in vitro
41	AT3G17520	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	all three analyses
	AT3G17520	EKAK(184)EAK	EAKEAAKR	LEA family protein	all three analyses
	AT3G17520	ESAK(260)DKA	DKASQSYDSAAR	LEA family protein	KO/OE and in vitro
42	AT3G22640	TRSK(231)EIG	EIGQGIIR	cupin family protein	KO/OE and in vitro
43	AT3G23990	YAAK(35)EIK	EIKFGVEAR	heat shock protein 60	singleton KO/OE. Not Shared
	AT3G23990	LMLK(49)GVE	GVEDLADAVKVTMGPKGR	heat shock protein 60	singleton KO/WT. Not shared
44	AT3G45190	VHDR(558)DYD	DYDLAGLANNLNQFR	SIT4 phosphatase-associated family protein	singleton KO/OE. Not Shared
45	AT3G48870	ISDR(501)FLP	FLPDKAIDLIDEAGSR	Clp ATPase	KO/WT and in vitro
46	AT3G51160	TAPK(20)ADS	ADSTVVEPR	GDP-D-mannose-4,6-dehydratase 2, GMD2	KO/OE and in vitro
47	AT3G53040	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	all three analyses
	AT3G53040	TVLK(449)EAD	EADQMTGQTFNDVGEIDDEEKVR	putative / LEA protein	both in vivo
	AT3G53040	DKTK(133)ETA	ETADYAAEKAR	putative / LEA protein	KO/OE and in vitro
	AT3G53040	DKTK(155)ETA	ETAETAEKAR	putative / LEA protein	all three analyses
48	AT3G61260	SLDR(92)DVK	DVKLADLSKEKR	Remorin family protein	both in vivo
49	AT4G02510	VSSR(607)EFS	EFSFGGKEVDQEPSGEGVTR	translocon at the outer envelope membrane of chloroplasts 159	KO/WT and in vitro
50	AT4G02930	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	all three analyses
51	AT4G05180	DQAR(110)DFS	DFSLALKDR	photosystem II subunit Q-2	singleton KO/WT. Not shared
52	AT4G15410	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	both in vivo
53	AT4G20360	ERAR(127)GIT	GITINTATVEYETENR	RAB GTPase homolog E1B	both in vivo
54	AT4G21020	EKAK(156)DYA	DYAEEDTMNAKEKAR	LEA family protein	all three analyses
	AT4G21020	EKAK(116)DTA	DTAYNAKEKAKDYAER	LEA family protein	KO/WT and in vitro
	AT4G21020	ESTK(218)NAA	NAAQTVTEAVVGPEDAEKAR	LEA family protein	KO/WT and in vitro
55	AT4G27440	YITK(337)GYV	GYVSETESGKR	protochlorophyllide oxidoreductase B	both in vivo
56	AT4G31700	IVKK(119)GEN	GENDLPLDTEKPR	ribosomal protein S6	both in vivo
57	AT4G34450	GIEK(27)GAV	GAVLQEAR	coatamer gamma-2 subunit, putative	singleton KO/WT. Not shared
58	AT4G37870	SFPK(20)GPV	GPVMPKITTGAAKR	phosphoenolpyruvate carboxylase 1	all three analyses
	AT4G37870	SLTR(102)ESG	ESGPKVVR	phosphoenolpyruvate carboxylase 1	KO/OE and in vitro
59	AT4G38630	QKDK(258)DGD	DGDTASASQETVAR	regulatory particle non-ATPase 10	KO/WT and in vitro
60	AT4G39260	DLQR(25)TFS	TFSQFGDVIDSKIINDR	cold, circadian rhythm, and RNA binding 1	KO/OE and in vitro
61	AT5G04200	LFGR(216)DAG	DAGLKFR	metacaspase 9	both in vivo
62	AT5G06140	EQPR(9)NIS	NISGSMQSPR	sorting nexin 1	both in vivo
63	AT5G10360	IVKK(119)GVS	GVSDLPLDTEKPR	Ribosomal protein S6e	KO/OE and in vitro

Protein ID	Accession	Sequence	Protein Name	Analysis	
64	AT5G12140	GGVR(15)DID	DIDANANDLQVESLAR	cystatin-1	both in vivo
65	AT5G16050	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	all three analyses
66	AT5G16130	LTGK(181)DVV	DVVFEYPVEA	Ribosomal protein S7e family protein	singleton KO/WT. Not shared
67	AT5G35910	ESTR(578)DLI	DLIMGAANTNEGR	Polynucleotidyl transferase	singleton KO/WT. Not shared
68	AT5G44310	EKTK(110)DYA	DYAEAKDKVNEGASR	LEA family protein	all three analyses
	AT5G44310	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	all three analyses
	AT5G44310	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	all three analyses
	AT5G44310	EKTK(150)DFA	DFAEETKEKVNEGASR	LEA family protein	both in vivo
69	AT5G47210	RYAK(341)EAA	EAAAPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	all three analyses
70	AT5G52440	TLER(145)EIG	EIGLDDISTPNVYNQNR	Bacterial sec-independent translocation protein mttA/Hcf106	both in vivo
71	AT5G54190	YITK(341)GYV	GYVSESEAGKR	protochlorophyllide oxidoreductase A	both in vivo
72	AT5G54310	QSGK(469)DFD	DFDFSSLMDGMFTKH	ARF-GAP domain 5	singleton KO/WT. Not shared
73	AT5G56000	SSKK(604)TME	TMEINPENSIMDELK	Heat shock protein 81.4	both in vivo
74	AT5G62575	TREK(42)ALL	ALLAEDSALKR	unknown protein	both in vivo

Selection and experimental validation of AtMC9 protein substrates

After comparative analysis of proteomes in which AtMC9 activity was abolished or enhanced, we selected neo N-terminal peptides that were unique or significantly enriched in proteomes containing active AtMC9. The reported *in vitro* and *in vivo* cleavage sites were further filtered using the known AtMC9 specificity for Arg and Lys. Potential substrates were further prioritised based on whether cleavage of the protein at that site was observed in more than one analysis. Given the high proteolysis relevance of cleavage sites that were reported by their *in vivo* singleton peptides, the corresponding proteins were anyhow included to our selection irrespectively of the number of analyses in which these were identified. In this manner, 99 sites in 74 different proteins were selected for cleavage validation (Table 2). Out of these, 53 sites (44 proteins) were identically processed in both *in vivo* analyses (Table 3), 19 sites (17 proteins) were identical between the KO/WT *in vivo* and the *in vitro* analysis, and 15 sites (12 proteins) between the KO/OE *in vivo* and the *in vitro* analysis (Figure 6). The remaining 12 sites (11 proteins) were reported by *in vivo* singleton peptides that were not found in at least two analyses. Finally, 18 sites (14 proteins) were identical in the three different setups (Figure 6 and Table 4).

We then opted to validate AtMC9 cleavage events with two independent approaches. First, peptides (18 amino acids long) harbouring the identified cleavage sites were synthesized for 6 randomly selected candidate protein substrates and used for *in vitro* rAtMC9 cleavage assays. Second, *in vitro* transcribed and translated (TnT), radio-labelled (³⁵S-Met) proteins, generated from available cDNA clones (Arabidopsis Biological Resource Centre), were incubated with rAtMC9 and their digestion products were analyzed by SDS-PAGE and autoradiography.

Following incubation with different concentrations of rAtMC9, synthetic peptides and their digestion products were separated by RP-HPLC and analyzed by MALDI-TOF-MS. Peptides spanning the AtMC9 cleavage sites identified in trypsin inhibitor 2 (AT1G47540), general regulatory factor 5 (GRF5, AT5G16050), glycine-rich RNA binding protein (GRP7, AT2G21660) and protochlorophyllide oxidoreductase B (PORB, AT4G27440) were cleaved by rAtMC9 at the sites identified by COFRADIC (Figure 7). The degree of peptide processing increased with increasing rAtMC9 concentration.

For proteins containing more than one identified cleavage site, such as late embryogenesis abundant

(LEA) domain-containing protein (AT2G42560), phosphoenolpyruvate carboxykinase 1 (PEPCK1, AT4G37870) and GRP7, two synthetic peptides were tested (Supporting information Figure S2). All these peptides were cleaved by rAtMC9 at the COFRADIC identified sites. However, alternative cleavage sites were observed, likely due to sequence similarities with the COFRADIC-deduced sites. For instance, the additional cleavage site R²¹²E in the (LEA) domain-containing protein precursor was identical to the COFRADIC-deduced site R²¹⁴E. As for GRP7, the peptide precursor was originally expected to get cleaved at the MK⁶⁴D site, while it was additionally processed at EK⁶¹A, a site similar to the first, since in both cases K is adjacent to an acidic residue (D or E). Furthermore, the consensus sequence for Lys-cleaved peptides supports the preference for acidic residues either at the N-terminal or C-terminal side of the cleaved Lys.

To assess the cleavage of substrates at the protein level, we used *in vitro* transcription and translation assays. In total, we successfully synthesized 16 proteins, and all were cleaved by rAtMC9. Although the levels of precursor proteins were clearly decreased or even completely consumed following rAtMC9 treatment, we did not always observe cleavage fragments. One reason for this is that substrate fragments not containing a radio-labeled methionine are not detectable by autoradiography. For this reason, we only focused on substrates that were processed into fragments of correct predicted molecular weight: GRF5, PEPCK1, NAD synthetase (AT1G55090), GDP-mannose 4, 6 dehydratase 2 (GMD2, AT3G51160), pentatricopeptide repeat (PPR)-containing protein (AT1G02150) and citrate synthase 3 (CSY3, AT2G42790) (Figure 8). None of these proteins were cleaved by an inactive rAtMC9 mutant, rAtMC9C147AC29A (Belenghi et al., 2007), which indicates that the observed fragments were indeed due to rAtMC9 and not to other (contaminating) proteases present in the assay mixture.

AtMC9 cleaves PEPCK1 *in vivo* leading to increased carboxylation activity

Phosphoenolpyruvate carboxykinase (PEPCK) is a cytosolic enzyme that catalyzes the reversible ATP-dependent decarboxylation of oxaloacetate to phosphoenolpyruvate. PEPCK is involved in gluconeogenesis, photosynthetic CO₂-concentrating mechanisms during C₄ photosynthesis (Edwards et al., 1971), crassulacean acid metabolism (CAM) (Dittrich et al., 1973) and nitrogenous compound metabolism (Leegood and Walker,

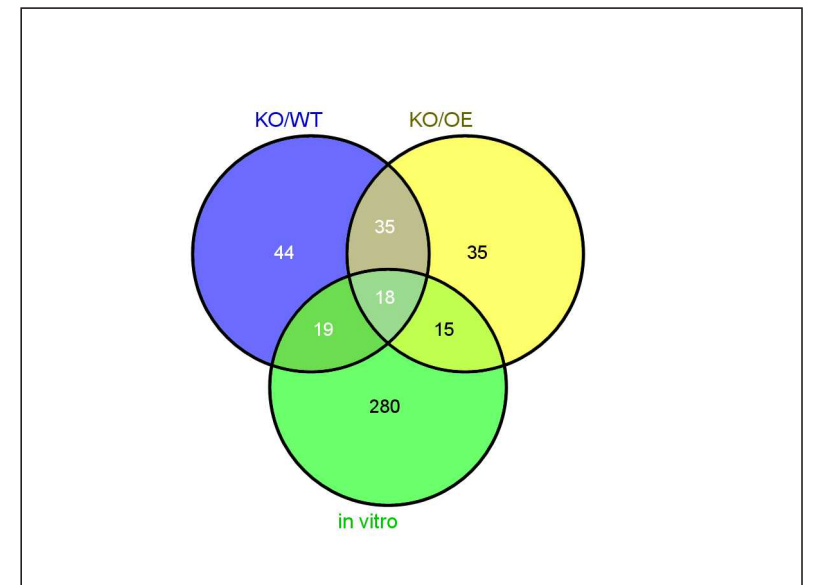
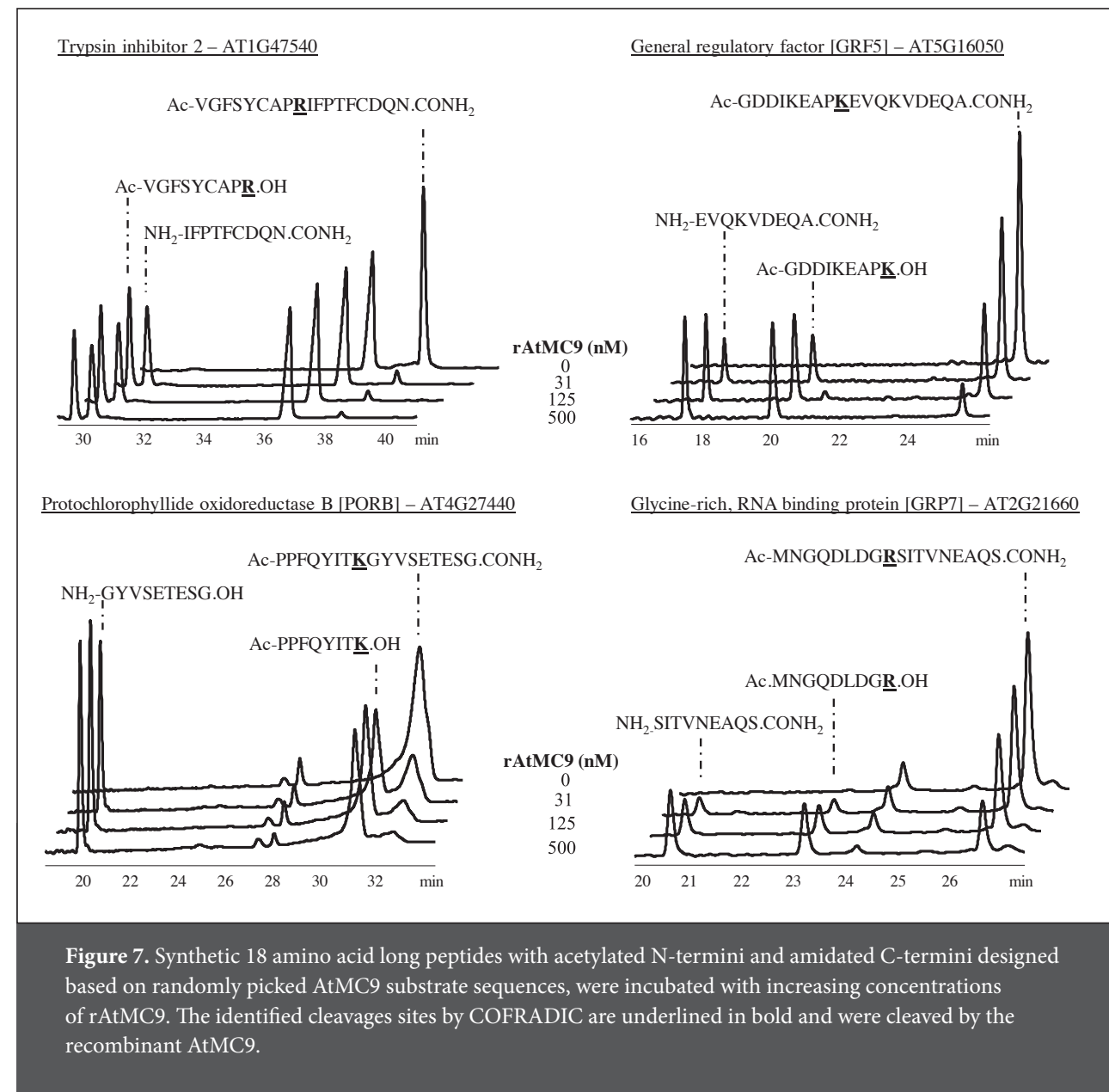
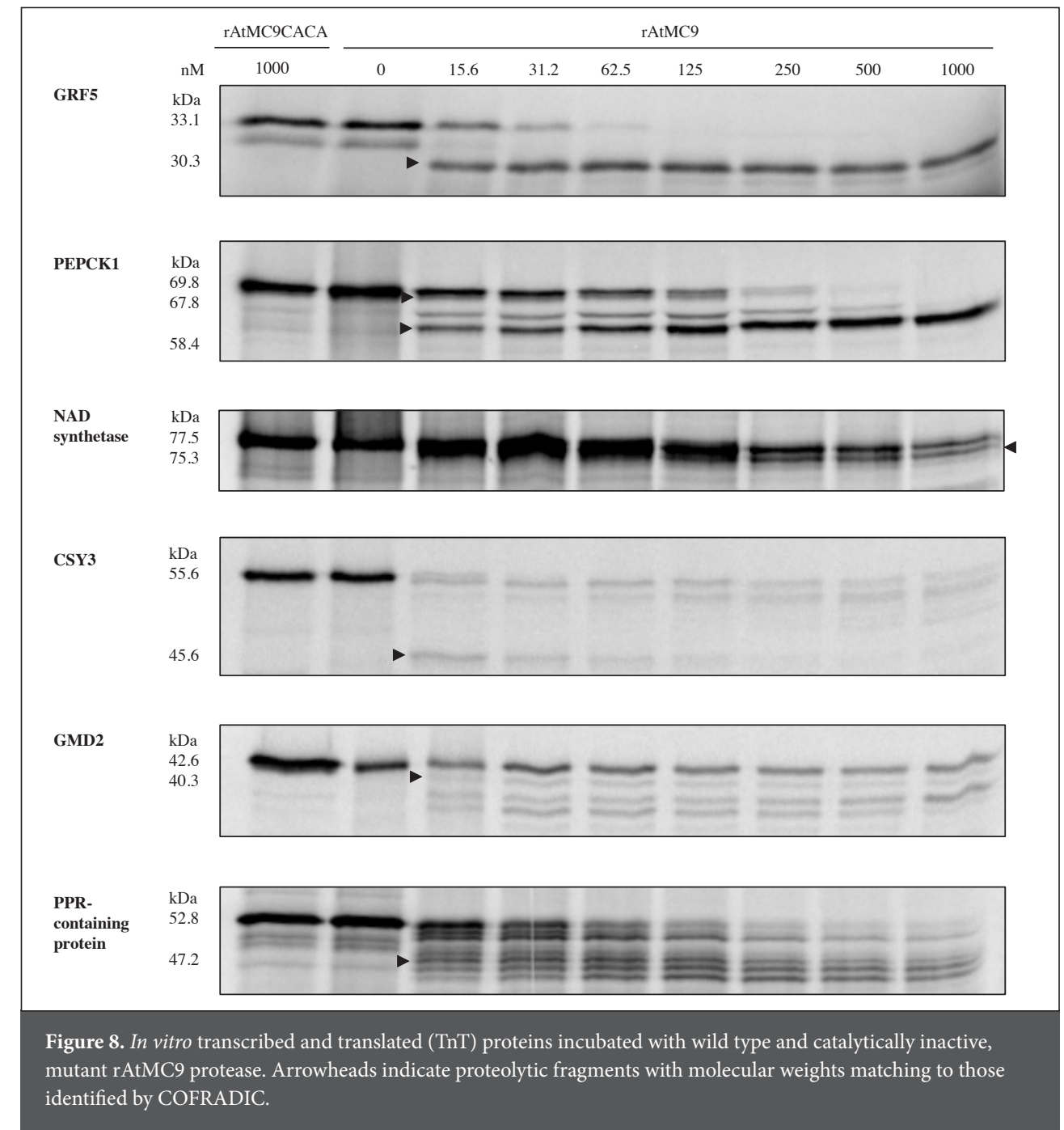


Figure 6. Venn diagram generated by VENNY (Oliveros et al., 2007). Shared cleavage sites within all three independent experiments (*in vivo* and *in vitro*).

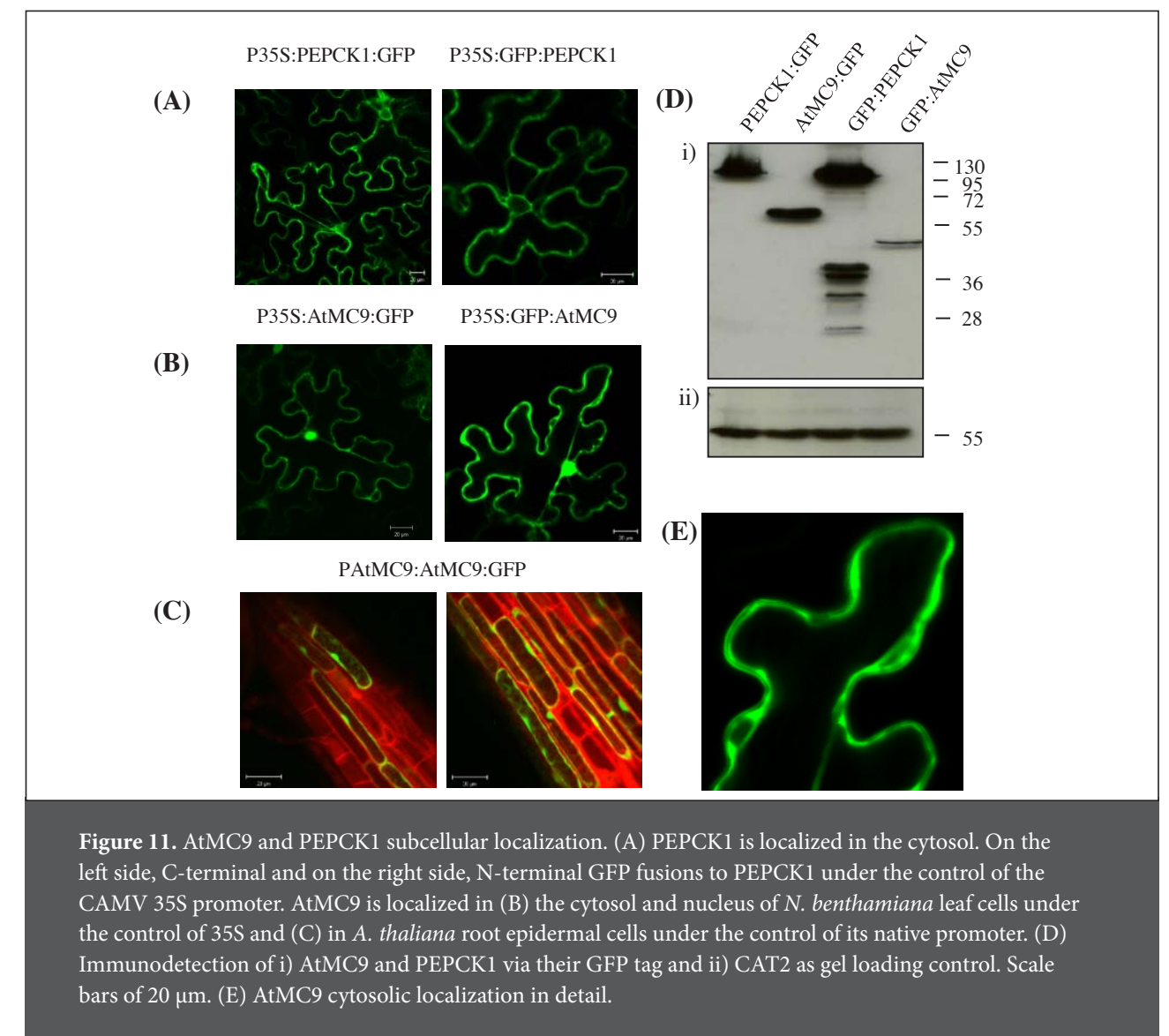
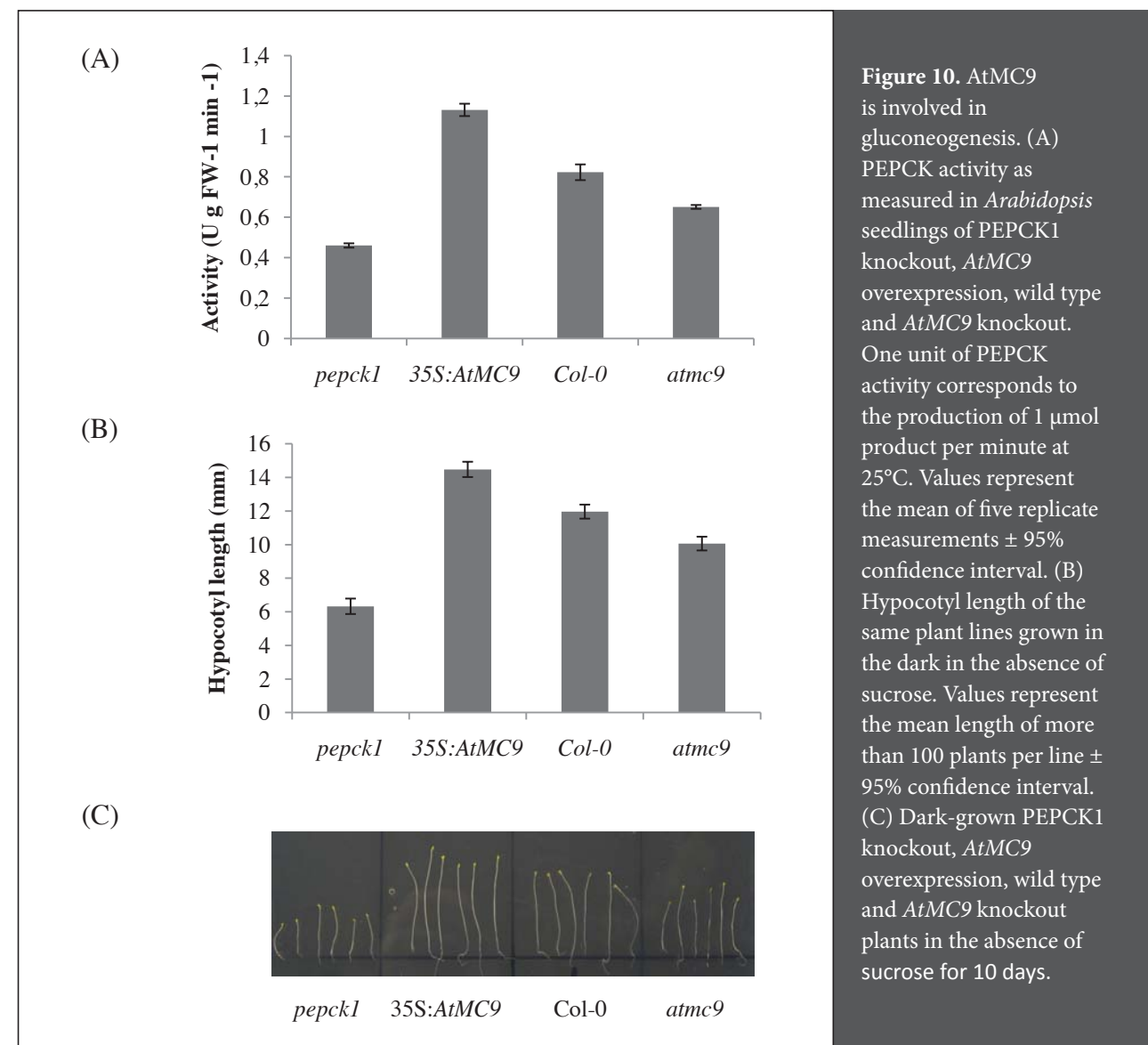


2003). It is post-translationally regulated by reversible phosphorylation (Walker and Leegood, 1995) and undergoes rapid proteolysis in various plant extracts (Walker et al., 1995 and 1997). Based on the latter, we decided to focus on this particular AtMC9 protein substrate and investigate whether AtMC9 cleaves PEPCK1 *in planta*.

Two PEPCK polypeptides of different molecular weight (74 and 73 kDa) were immunodetected in *Arabidopsis* cotyledons using a polyclonal antiserum raised against the PEPCK from Guinea grass (*Panicum maximum*) leaves (Malone et al., 2007). It was further suggested that the smaller form did not result from proteolysis of the larger, but could represent the second *Arabidopsis* PEPCK isoform, PEPCK2 (Malone et al., 2007). Similarly, in pea (*Pisum sativum*) cotyledons, besides the mature 72 kDa a smaller 68 kDa PEPCK polyptide



was detected that was attributed either to a genuine isoform or to the result of a regulatory cleavage mechanism (Delgado-Alvarado et al., 2007). Recently, it was shown that the pineapple (*Ananas comosus*) PEPCK is present *in vivo* with two polypeptides of 74 and 65 kDa that are both active (Martín et al., 2011). In cucumber (*Cucumis sativus*) an *in vivo* phosphorylation site is located in the proteolytically-removed fragment at the N-terminal region of the protein since its removal coincided with the loss of the N-terminal 32P label (Walker and Leegood,



AtMC9 subcellular localization

Given that PEPCK1 is a cytosolic enzyme and AtMC9 has been detected in the apoplast fractions of protein extracts from overexpression lines (Vercammen et al., 2006), we decided to study further the subcellular localizations of AtMC9 and PEPCK1. We transiently expressed N- and C-terminal green fluorescent protein (GFP)-tagged AtMC9 and PEPCK1 in tobacco (*Nicotiana benthamiana*) leaves under the control of the constitutive cauliflower mosaic virus (CaMV) 35S promoter. As expected, PEPCK1 was localized exclusively in the cytosol (Figure 11A). N- and C-terminal AtMC9 fusion proteins displayed nucleocytoplasmic localization (Figure 11B), but we were not able to show apoplastic AtMC9 localization. This might be explained by GFP instability in the acidic environment of the apoplast (Ward, 1981; Patterson et al., 1997). To exclude a false positive GFP signal in the nucleus due to the diffusion capacity of free GFP, we analyzed the integrity of the AtMC9 fusion proteins by Western blot analysis (Figure 11D). We detected an approximately 45 kDa GFP:AtMC9 protein that

consists of the 25 kDa GFP protein fused with the large p20 subunit of AtMC9 (20 kDa). In this case, the small p10 subunit (15 kDa) was removed after autocatalytic activation of the protease (Vercammen et al., 2004) which indicates that AtMC9 was active in the observed subcellular compartments. For the C-terminal AtMC9:GFP fusion, we detected an approximately 60 kDa protein that correlates with the full length of AtMC9:GFP fusion protein, and this shows that the GFP fusion at the C-terminus of AtMC9 may have interfered with the proteolytic separation of the two subunits, but AtMC9 activation took place under both GFP fusion conformations (data not shown) and AtMC9 localized in the same cellular compartments.

To confirm the nucleocytoplasmic localization of AtMC9 we repeated the analysis in *Arabidopsis* roots using proAtMC9:AtMC9:GFP reporter lines. Interestingly, we were able to confirm the localization of AtMC9 in the nucleus and cytoplasm of epidermal cells (Figure 11C). Altogether our localization studies revealed that AtMC9 aside from the apoplast, is also found in the cell nucleus and cytoplasm.

Table 3. List of endogenous AtMC9 substrates identified in both *in vivo* analyses. Targets are selected based on cleavage after an Arg or Lys. These peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the proteomes containing active protease (WT or OE). Moreover, they were classified as substrates in both independent *in vivo* proteomic setups (KO/WT and KO/OE). Data on the ion identification and quantification are depicted for each peptide in each experiment. Amino acid positions of P1' sites are shown between brackets in the P4-P3' column.

Fold times of higher WT or OE ion intensity in comparison to the KO ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions which is not applicable due to the absence of heavy peptide ion.

#	Accession	M/Z of best scoring peptide	Z	Intensity fold increase	P4-P3'	Sequence	Description	score (s)	threshold (t)	E value	Mass error (ppm)	# of spectra	Experiment
1	AT1G04750	990.079	2	2.64	VLDR(156)GEK	GEKIELLVDKTENLR	vesicle-associated membrane protein 721	105	41	3.98E-09	-2.022	1	KO/WT
	AT1G04750	984.059	2	2.54	VLDR(156)GEK	GEKIELLVDKTENLR	vesicle-associated membrane protein 721	77	43	3.98E-06	-2.747	2	KO/OE
2	AT1G07780	468.790	2	4.08	SVAK(109)DIS	DISKVAR	phosphoribosylanthranilate isomerase 1	55	42	5.01E-04	-0.321	2	KO/WT
	AT1G07780	468.790	2	6.09	SVAK(109)DIS	DISKVAR	phosphoribosylanthranilate isomerase 1	63	42	7.94E-05	-0.107	2	KO/OE
3	AT1G09310	861.011	2	N/A	TEAK(165)EAV	EAVAIKEAVAVKEAA	Protein of unknown function, DUF538	45	40	3.16E-03	-0.465	1	KO/WT
	AT1G09310	861.012	2	N/A	TEAK(165)EAV	EAVAIKEAVAVKEAA	Protein of unknown function, DUF538	45	40	3.16E-03	-1.919	1	KO/OE
4	AT1G11480	517.827	2	11.77	ARPR(404)ELV	ELVLKER	eukaryotic translation initiation factor-related	64	39	3.16E-05	-0.097	1	KO/WT
	AT1G11480	517.827	2	12.20	ARPR(404)ELV	ELVLKER	eukaryotic translation initiation factor-related	59	39	1.00E-04	-0.097	1	KO/OE
5	AT1G12000	865.969	2	2.68	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	94	44	1.00E-07	-1.445	3	KO/WT
	AT1G12000	861.957	2	4.23	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	46	44	6.31E-03	-2.149	2	KO/OE
6	AT1G29350	747.011	3	7.98	QNAK(525)ELD	ELDFQYSPFSAQQSMQSR	Kinase-related protein of unknown function (DUF1296)	41	39	6.31E-03	-1.072	2	KO/WT
	AT1G29350	1120.015	2	6.68	QNAK(525)ELD	ELDFQYSPFSAQQSMQSR	Kinase-related protein of unknown function (DUF1296)	59	40	1.26E-04	-2.770	1	KO/OE
7	AT1G47540	777.053	3	N/A	AYDR(31)KCL	KCLKEYGGDVGFYSYCAPR	trypsin inhibitor 2	48	43	3.16E-03	-1.890	1	KO/WT
	AT1G47540	777.053	3	N/A	AYDR(31)KCL	KCLKEYGGDVGFYSYCAPR	trypsin inhibitor 2	81	43	1.58E-06	-2.577	1	KO/OE
8	AT1G49240	743.874	2	2.72	MNQK(53)DAY	DAYVGDEAQSQR	actin 8	58	43	3.16E-04	-1.144	2	KO/WT
	AT1G49240	739.861	2	4.01	MNQK(53)DAY	DAYVGDEAQSQR	actin 8	80	43	2.00E-06	-1.015	2	KO/OE
9	AT1G49240	924.983	2	7.34	SIEK(241)NYE	NYELPDGQVITIGAER	actin 8	54	44	1.00E-03	-3.030	1	KO/WT
	AT1G49240	924.983	2	17.14	SIEK(241)NYE	NYELPDGQVITIGAER	actin 8	57	44	5.01E-04	-3.139	1	KO/OE
10	AT1G49970	556.774	2	14.57	FEKR(363)DYD	DYDGTLAQR	CLP protease proteolytic subunit 1	65	43	6.31E-05	0.360	1	KO/WT
	AT1G49970	556.774	2	31.94	FEKR(363)DYD	DYDGTLAQR	CLP protease proteolytic subunit 1	70	43	2.00E-05	-0.540	1	KO/OE
11	AT1G51710	526.776	2	4.29	VATR(157)ELF	ELFGELDR	ubiquitin-specific protease 6	54	43	7.94E-04	0.475	1	KO/WT
	AT1G51710	526.776	2	N/A	VATR(157)ELF	ELFGELDR	ubiquitin-specific protease 6	45	43	6.31E-03	-0.475	1	KO/OE
12	AT1G55490	686.717	3	9.09	TVAR(114)EVE	EVELEDPVENIGAKLVR	chaperonin 60 beta	55	43	6.31E-04	-1.215	4	KO/WT
	AT1G55490	686.717	3	12.72	TVAR(114)EVE	EVELEDPVENIGAKLVR	chaperonin 60 beta	58	43	3.16E-04	-1.215	4	KO/OE
13	AT1G56070	868.436	3	2.07	ICLK(544)DLQ	DLQDDFMGGAIEIKSDPVVSFR	Ribosomal protein S5/Elongation factor G/III/V family protein	63	44	1.26E-04	-1.114	1	KO/WT
	AT1G56070	868.771	3	2.50	ICLK(544)DLQ	DLQDDFMGGAIEIKSDPVVSFR	Ribosomal protein S5/Elongation factor G/III/V family protein	51	44	2.00E-03	-1.095	1	KO/OE
14	AT2G01250	738.484	2	8.68	VESK(6)VVV	VVVPESVLKRR	Ribosomal protein L30/L7 family protein	57	31	2.51E-05	-1.017	3	KO/WT
	AT2G01250	738.486	2	8.60	VESK(6)VVV	VVVPESVLKRR	Ribosomal protein L30/L7 family protein	54	31	5.01E-05	-3.390	2	KO/OE
15	AT2G18960	711.397	2	5.86	GSYR(902)ELS	ELSEIAEQAKR	H(+)-ATPase 1	79	44	3.16E-06	-1.478	1	KO/WT
	AT2G18960	707.383	2	3.63	GSYR(902)ELS	ELSEIAEQAKR	H(+)-ATPase 1	64	44	1.00E-04	-0.637	4	KO/OE
16	AT2G19730	670.423	2	2.96	SVNK(88)SIL	SILKKEFPR	Ribosomal L28e protein family	45	38	2.00E-03	0.224	1	KO/WT
	AT2G19730	670.423	2	4.80	SVNK(88)SIL	SILKKEFPR	Ribosomal L28e protein family	45	38	2.00E-03	-0.448	1	KO/OE
17	AT2G21060	787.327	2	8.10	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	156	34	6.31E-15	0.827	2	KO/WT
	AT2G21060	787.329	2	N/A	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	61	34	2.00E-05	-1.017	1	KO/OE
18	AT2G26250	873.949	2	3.00	IVNR(18)GIE	GIEPSGPNAGSPTFSVR	3-ketoacyl-CoA synthase 10	66	44	6.31E-05	-3.150	1	KO/WT
	AT2G26250	873.948	2	5.23	IVNR(18)GIE	GIEPSGPNAGSPTFSVR	3-ketoacyl-CoA synthase 10	53	44	1.26E-03	-3.036	1	KO/OE
19	AT2G28000	922.494	2	2.03	TIAR(103)AIE	AIELPNAMENAGAALIR	chaperonin-60alpha	75	44	7.94E-06	-2.008	2	KO/WT
	AT2G28000	922.495	2	2.83	TIAR(103)AIE	AIELPNAMENAGAALIR	chaperonin-60alpha	48	44	3.98E-03	-3.256	2	KO/OE
20	AT2G28000	793.105	3	2.95	PRGR(82)NVV	NVVLDFGSPKVVNDGVVTIAR	chaperonin-60alpha	44	43	7.94E-03	-1.389	1	KO/WT
	AT2G28000	1189.157	2	4.35	PRGR(82)NVV	NVVLDFGSPKVVNDGVVTIAR	chaperonin-60alpha	55	43	6.31E-04	-3.324	2	KO/OE
21	AT2G32240	757.410	2	2.97	VKSR(709)DID	DIDLFSSTPKR	unknown protein	52	44	1.58E-03	-1.388	1	KO/WT
	AT2G32240	757.411	2	3.58	VKSR(709)DID	DIDLFSSTPKR	unknown protein	71	44	2.00E-05	-1.454	1	KO/OE

22	AT2G36620	777.431	2	5.11	NIPK(149)SAA	SAAPKAARKMGGGGGR	ribosomal protein L24	69	44	3.16E-05	-3.091	1	KO/WT
	AT2G36620	777.430	2	4.49	NIPK(149)SAA	SAAPKAARKMGGGGGR	ribosomal protein L24	93	44	1.26E-07	-1.803	1	KO/OE
23	AT2G42560	824.466	3	5.03	VGSK(454)AVD	AVDLTKEKA AVAADTVVGYTAR	LEA domain-containing protein	83	42	7.94E-07	-1.741	2	KO/WT
	AT2G42560	824.467	3	14.94	VGSK(454)AVD	AVDLTKEKA AVAADTVVGYTAR	LEA domain-containing protein	71	42	1.26E-05	-2.672	1	KO/OE
24	AT2G42560	844.140	2	6.20	ARAK(324)DYT	DYTLQKAVEAKDVAAEKAQR	LEA domain-containing protein	57	43	3.98E-04	-2.214	2	KO/WT
	AT2G42560	844.140	3	16.05	ARAK(324)DYT	DYTLQKAVEAKDVAAEKAQR	LEA domain-containing protein	70	43	2.00E-05	-2.451	1	KO/OE
25	AT3G02480	723.721	3	N/A	MKEK(51)AQQ	AQGAADVVKDKTGMNKSH	LEA family protein	88	44	3.98E-07	-1.153	2	KO/WT
	AT3G02480	1085.079	2	3.81	MKEK(51)AQQ	AQGAADVVKDKTGMNKSH	LEA family protein	70	44	2.51E-05	-3.090	2	KO/OE
26	AT3G09840	563.302	2	4.76	KSKK(15)DFS	DFSTAILER	cell division cycle 48	77	43	3.98E-06	0	1	KO/WT
	AT3G09840	563.303	2	4.08	KSKK(15)DFS	DFSTAILER	cell division cycle 48	86	44	6.31E-07	-0.622	1	KO/OE
27	AT3G13300	711.369	2	N/A	VGER(708)NLD	NLDVSSVEEISR	Transducin/WD40 repeat-like superfamily protein	79	44	3.16E-06	-0.633	1	KO/WT
	AT3G13300	711.37	2	8.67	VGER(708)NLD	NLDVSSVEEISR	Transducin/WD40 repeat-like superfamily protein	80	44	2.51E-06	-1.619	1	KO/OE
28	AT3G17520	565.269	2	8.73	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	78	41	2.00E-06	-0.266	1	KO/WT
	AT3G17520	565.269	2	22.50	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	71	41	1.00E-05	-0.443	1	KO/OE
29	AT3G17520	562.840	2	2.90	EKAK(184)EAK	EAKEAAKR	LEA family protein	49	42	2.00E-03	-0.267	1	KO/WT
	AT3G17520	562.840	2	13.79	EKAK(184)EAK	EAKEAAKR	LEA family protein	53	42	7.94E-04	-0.356	2	KO/OE
30	AT3G53040	930.762	3	2.30	TVLK(449)EAD	EADQMTGQTFNDVGEIDDEEKVR	putative / LEA protein	54	39	3.16E-04	-2.689	1	KO/WT
	AT3G53040	930.762	3	2.87	TVLK(449)EAD	EADQMTGQTFNDVGEIDDEEKVR	putative / LEA protein	44	39	3.16E-03	-3.119	1	KO/OE
31	AT3G53040	694.849	2	9.92	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	75	43	6.31E-06	-0.360	1	KO/WT
	AT3G53040	694.850	2	11.76	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	68	44	3.98E-05	-0.865	2	KO/OE
32	AT3G53040	708.865	2	17.26	DKTK(155)ETA	ETAEYTAEKAR	putative / LEA protein	59	44	3.16E-04	-0.848	1	KO/WT
	AT3G53040	708.867	2	32.09	DKTK(155)ETA	ETAEYTAEKAR	putative / LEA protein	60	44	2.51E-04	-3.461	2	KO/OE
33	AT3G61260	849.518	2	5.21	SLDR(92)DVK	DVKLADLSKEKR	Remorin family protein	66	39	2.00E-05	-1.120	2	KO/WT
	AT3G61260	849.519	2	4.75	SLDR(92)DVK	DVKLADLSKEKR	Remorin family protein	78	39	1.26E-06	-1.532	1	KO/OE
34	AT4G15410	882.445	2	4.91	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	96	44	6.31E-08	-2.723	1	KO/WT
	AT4G15410	882.445	2	9.96	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	58	44	3.98E-04	-2.553	1	KO/OE
35	AT4G20360	942.974	2	5.55	ERAR(127)GIT	GITINTATVEYETENR	RAB GTPase homolog E1B	98	44	3.98E-08	-2.229	1	KO/WT
	AT4G20360	942.975	2	8.30	ERAR(127)GIT	GITINTATVEYETENR	RAB GTPase homolog E1B	102	44	1.58E-08	-2.760	1	KO/OE
36	AT4G21020	997.976	2	N/A	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	71	42	1.26E-05	-1.605	4	KO/WT
	AT4G21020	997.977	2	49.67	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	105	42	5.01E-09	-2.758	2	KO/OE
37	AT4G27440	680.852	2	4.30	YITK(337)GYV	GYVSETESGKR	protochlorophyllide oxidoreductase B	63	44	1.26E-04	-0.515	2	KO/WT
	AT4G27440	680.852	2	14.80	YITK(337)GYV	GYVSETESGKR	protochlorophyllide oxidoreductase B	65	44	7.94E-05	-1.030	3	KO/OE
38	AT4G29060	827.083	3	4.19	PHRR(74)ATG	ATGTDVVA AVEEQDSTPVVAEDK	elongation factor Ts family protein	53	44	1.26E-03	-1.412	1	KO/WT
	AT4G29060	827.083	3	3.43	PHRR(74)ATG	ATGTDVVA AVEEQDSTPVVAEDK	elongation factor Ts family protein	58	44	3.98E-04	-2.139	1	KO/OE
39	AT4G31700	895.465	2	3.26	IVKK(119)GEN	GENDLPGLTDTEKPR	ribosomal protein S6	66	44	6.31E-05	-2.348	2	KO/WT
	AT4G31700	891.452	2	2.54	IVKK(119)GEN	GENDLPGLTDTEKPR	ribosomal protein S6	71	44	2.00E-05	-2.976	3	KO/OE
40	AT4G37870	832.996	2	6.99	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	76	41	3.16E-06	-2.704	1	KO/WT
	AT4G37870	832.996	2	11.95	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	80	41	1.26E-06	-2.464	1	KO/OE
41	AT5G04200	473.771	2	N/A	LFGR(216)DAG	DAGLKFR	metacaspase 9	58	44	3.98E-04	0.317	2	KO/WT
	AT5G04200	477.785	2	N/A	LFGR(216)DAG	DAGLKFR	metacaspase 9	55	42	5.01E-04	0.210	1	KO/OE
42	AT5G06140	583.787	2	9.46	EQPR(9)NIS	NISGSMQSPR	sorting nexin 1	76	43	5.01E-06	-0.858	1	KO/WT
	AT5G06140	583.787	2	5.09	EQPR(9)NIS	NISGSMQSPR	sorting nexin 1	73	43	1.00E-05	-0.515	1	KO/OE
43	AT5G12140	909.458	2	4.04	GGVR(15)DID	DIDANANDLQVESLAR	cystatin-1	102	44	1.58E-08	-1.761	3	KO/WT
	AT5G12140	909.459	2	2.83	GGVR(15)DID	DIDANANDLQVESLAR	cystatin-1	54	44	1.00E-03	-2.202	1	KO/OE
44	AT5G16050	914.473	2	17.77	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	59	44	3.16E-04	-2.244	2	KO/WT
	AT5G16050	914.473	2	N/A	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	48	44	3.98E-03	-2.189	1	KO/OE
45	AT5G44310	1016.519	2	17.89	EKTK(150)DFA	DFAEETKEKVNNEGASR	LEA family protein	114	44	1.00E-09	-1.920	2	KO/WT
	AT5G44310	1016.519	2	N/A	EKTK(150)DFA	DFAEETKEKVNNEGASR	LEA family protein	128	44	3.98E-11	-1.329	2	KO/OE
46	AT5G44310	1002.504	2	23.93	EKTK(110)DYA	DYAEAKDKVNNEGASR	LEA family protein	127	44	5.01E-11	-1.897	1	KO/WT
	AT5G44310	1002.504	2	52.30	EKTK(110)DYA	DYAEAKDKVNNEGASR	LEA family protein	95	44	7.94E-08	-2.596	2	KO/OE

47	AT5G44310	1016.526	2	23.42	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	125	44	7.94E-11	-2.806	2	KO/WT
	AT5G44310	678.018	3	65.10	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	68	44	3.98E-05	-1.083	2	KO/OE
48	AT5G44310	1016.525	2	14.65	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	119	44	3.16E-10	-2.314	2	KO/WT
	AT5G44310	1016.525	2	32.12	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	71	44	2.00E-05	-2.314	2	KO/OE
49	AT5G47210	845.431	2	21.15	RYAK(341)EAA	EAAAAPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	49	44	3.16E-03	-1.954	1	KO/WT
	AT5G47210	845.431	2	44.65	RYAK(341)EAA	EAAAAPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	60	44	2.51E-04	-2.428	2	KO/OE
50	AT5G52440	1011.504	2	18.41	TLER(145)EIG	EIGLDDISTPNVYNQNR	Bacterial sec-independent translocation protein mttA/Hcf106	60	44	2.51E-04	-2.573	1	KO/WT
	AT5G52440	1011.504	2	55.35	TLER(145)EIG	EIGLDDISTPNVYNQNR	Bacterial sec-independent translocation protein mttA/Hcf106	54	44	1.00E-03	-2.573	1	KO/OE
51	AT5G54190	665.847	2	5.61	YITK(341)GYV	GYVSESEAGKR	protochlorophyllide oxidoreductase A	66	44	6.31E-05	-0.451	1	KO/WT
	AT5G54190	665.847	2	7.10	YITK(341)GYV	GYVSESEAGKR	protochlorophyllide oxidoreductase A	58	44	3.98E-04	-0.827	1	KO/OE
52	AT5G56000	949.441	2	5.11	SSKK(604)TME	TMEINPENSIMDELRL	HEAT SHOCK PROTEIN 81.4	80	41	1.26E-06	-2.636	1	KO/WT
	AT5G56000	949.441	2	7.04	SSKK(604)TME	TMEINPENSIMDELRL	HEAT SHOCK PROTEIN 81.4	69	41	1.58E-05	-2.161	1	KO/OE
53	AT5G62575	667.899	2	3.83	TREK(42)ALL	ALLAEDSALKR	unknown protein	72	41	7.94E-06	-0.825	1	KO/WT
	AT5G62575	667.900	2	8.98	TREK(42)ALL	ALLAEDSALKR	unknown protein	73	42	7.94E-06	-1.649	2	KO/OE

Table 4. List of endogenous AtMC9 substrates identified in all three analyses.

Targets are selected based on cleavage after an Arg or Lys. These peptides were identified either as unique or significantly ($p < 0.02$) up-regulated ions in the proteome of active protease (WT, OE or added recombinant). Moreover, they were classified as substrates in all three independent proteomic setups (two *in vivo* and one *in vitro*). Data on the ion identification and quantification are depicted for each peptide in each experiment. Amino acid positions of P1' sites are shown between

brackets in the P4-P3' column. Fold times of higher WT, OE or rAtMC9 treated ion intensity in comparison to the KO ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions which is not applicable due to the absence of light or heavy peptide ion.

#	Accession	M/Z of best scoring peptide	Z	Intensity fold increase	P4-P3'	Sequence	Description	score (s)	threshold (t)	E value	Mass error (ppm)	# of spectra	Experiment
1	AT1G12000	865.969	2	2.68	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	94	44	1.00E-07	-1.445	3	KO/WT
	AT1G12000	861.957	2	4.23	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	46	44	6.31E-03	-2.149	2	KO/OE
	AT1G12000	861.955	2	8.39	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	62	44	1.58E-04	-0.697	1	IN VITRO
2	AT1G49240	743.874	2	2.72	MNQK(53)DAY	DAYVGDQAQSKR	actin 8	58	43	3.16E-04	-1.144	2	KO/WT
	AT1G49240	739.861	2	4.01	MNQK(53)DAY	DAYVGDQAQSKR	actin 8	80	43	2.00E-06	-1.015	2	KO/OE
	AT1G49240	743.873	2	5.67	MNQK(53)DAY	DAYVGDQAQSKR	actin 8	74	43	7.94E-06	-0.269	4	IN VITRO
3	AT2G21060	787.327	2	8.10	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	156	34	6.31E-15	0.827	2	KO/WT
	AT2G21060	787.329	2	N/A	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	61	34	2.00E-05	-1.017	1	KO/OE
	AT2G21060	785.322	2	53.84	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	134	33	7.94E-13	-0.637	1	IN VITRO
4	AT2G42560	824.466	3	5.03	VGSK(454)AVD	AVDLTKEKA AVAADTVVGYTAR	LEA domain-containing protein	83	42	7.94E-07	-1.741	2	KO/WT
	AT2G42560	824.467	3	14.94	VGSK(454)AVD	AVDLTKEKA AVAADTVVGYTAR	LEA domain-containing protein	71	42	1.26E-05	-2.672	1	KO/OE
	AT2G42560	1230.174	2	4.07	VGSK(454)AVD	AVDLTKEKA AVAADTVVGYTAR	LEA domain-containing protein	60	43	2.00E-04	-0.529	2	IN VITRO
5	AT3G02480	723.721	3	N/A	MKEK(51)AQQ	AQGAADVVKDKTG MNKSH	LEA family protein	88	44	3.98E-07	-1.153	2	KO/WT
	AT3G02480	1085.079	2	3.81	MKEK(51)AQQ	AQGAADVVKDKTG MNKSH	LEA family protein	70	44	2.51E-05	-3.090	2	KO/OE
	AT3G02480	1077.543	2	33.01	MKEK(51)AQQ	AQGAADVVKDKTG MNKSH	LEA family protein	59	44	3.16E-04	7.226	6	IN VITRO
6	AT3G09840	563.302	2	4.76	KSKK(15)DFS	DFSTAILER	cell division cycle 48	77	43	3.98E-06	0	1	KO/WT
	AT3G09840	563.303	2	4.08	KSKK(15)DFS	DFSTAILER	cell division cycle 48	86	44	6.31E-07	-0.622	1	KO/OE
	AT3G09840	561.296	2	N/A	KSKK(15)DFS	DFSTAILER	cell division cycle 48	61	44	2.00E-04	-0.178	2	IN VITRO
7	AT3G17520	565.269	2	8.73	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	78	41	2.00E-06	-0.266	1	KO/WT
	AT3G17520	565.269	2	22.50	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	71	41	1.00E-05	-0.443	1	KO/OE
	AT3G17520	563.262	2	N/A	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	77	41	2.51E-06	0.267	3	IN VITRO
8	AT3G17520	562.840	2	2.90	EKAK(184)EAK	EAKEAAKR	LEA family protein	49	42	2.00E-03	-0.267	1	KO/WT
	AT3G17520	562.840	2	13.79	EKAK(184)EAK	EAKEAAKR	LEA family protein	53	42	7.94E-04	-0.356	2	KO/OE
	AT3G17520	556.819	2	N/A	EKAK(184)EAK	EAKEAAKR	LEA family protein	61	44	2.00E-04	0.270	6	IN VITRO

Group	Accession	Score	Length	Score	Sequence	Protein	Score	Length	Score	Score	Length	Category	
9	AT3G53040	694.849	2	9.92	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	75	43	6.31E-06	-0.360	1	KO/WT
	AT3G53040	694.850	2	11.76	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	68	44	3.98E-05	-0.865	2	KO/OE
	AT3G53040	690.836	2	23.72	QKTK(111)ETA	ETADYTADKAR	putative / LEA protein	49	43	2.51E-03	-0.217	6	IN VITRO
10	AT3G53040	708.865	2	17.26	DKTK(155)ETA	ETAEYTAEKAR	putative / LEA protein	59	44	3.16E-04	-0.848	1	KO/WT
	AT3G53040	708.867	2	32.09	DKTK(155)ETA	ETAEYTAEKAR	putative / LEA protein	60	44	2.51E-04	-3.461	2	KO/OE
	AT3G53040	704.852	2	38.41	DKTK(155)ETA	ETAEYTAEKAR	putative / LEA protein	54	43	7.94E-04	-0.426	1	IN VITRO
11	AT4G02930	1078.610	2	10.26	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	87	43	3.98E-07	-2.227	3	KO/WT
	AT4G02930	1078.609	2	10.55	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	52	43	1.26E-03	-1.949	2	KO/OE
	AT4G02930	714.056	3	6.77	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	65	44	7.94E-05	-0.280	1	IN VITRO
12	AT4G21020	997.976	2	N/A	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	71	42	1.26E-05	-1.605	4	KO/WT
	AT4G21020	997.977	2	49.67	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	105	42	5.01E-09	-2.758	2	KO/OE
	AT4G21020	661.639	3	28.31	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	43	40	5.01E-03	0.050	6	IN VITRO
13	AT4G37870	832.996	2	6.99	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	76	41	3.16E-06	-2.704	1	KO/WT
	AT4G37870	832.996	2	11.95	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	80	41	1.26E-06	-2.464	1	KO/OE
	AT4G37870	551.652	3	11.43	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	49	43	2.51E-03	0.061	3	IN VITRO
14	AT5G16050	914.473	2	17.77	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	59	44	3.16E-04	-2.244	2	KO/WT
	AT5G16050	914.473	2	N/A	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	48	44	3.98E-03	-2.189	1	KO/OE
	AT5G16050	910.458	2	N/A	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	75	44	7.94E-06	-0.605	1	IN VITRO
15	AT5G44310	1002.504	2	23.93	EKTK(110)DYA	DYAEAKDKVNEGASR	LEA family protein	127	44	5.01E-11	-1.897	1	KO/WT
	AT5G44310	1002.504	2	52.30	EKTK(110)DYA	DYAEAKDKVNEGASR	LEA family protein	95	44	7.94E-08	-2.596	2	KO/OE
	AT5G44310	996.482	2	116.70	EKTK(110)DYA	DYAEAKDKVNEGASR	LEA family protein	130	43	5.01E-04	0.100	3	IN VITRO
16	AT5G44310	1016.526	2	23.42	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	125	44	7.94E-11	-2.806	2	KO/WT
	AT5G44310	678.018	3	65.10	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	68	44	3.98E-05	-1.083	2	KO/OE
	AT5G44310	1010.504	2	N/A	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	86	44	6.31E-07	-1.040	1	IN VITRO
17	AT5G44310	1016.525	2	14.65	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	119	44	3.16E-10	-2.314	2	KO/WT
	AT5G44310	1016.525	2	32.12	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	71	44	2.00E-05	-2.314	2	KO/OE
	AT5G44310	1010.504	2	N/A	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	117	44	5.01E-10	-0.941	3	IN VITRO
18	AT5G47210	845.431	2	21.15	RYAK(341)EAA	EAAAIPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	49	44	3.16E-03	-1.954	1	KO/WT
	AT5G47210	845.431	2	44.65	RYAK(341)EAA	EAAAIPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	60	44	2.51E-04	-2.428	2	KO/OE
	AT5G47210	843.423	2	27.17	RYAK(341)EAA	EAAAIPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	80	43	2.00E-06	-0.475	1	IN VITRO

DISCUSSION

To date our knowledge on metacaspase substrates is limited to the AtSerpin 1 inhibitor of AtMC9 (Vercammen et al., 2004), the tudor staphylococcal nuclease (TSN) target of mCII-pa (Sundström et al., 2009), the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) target of YCA1 (Silva et al., 2011), the metacaspase TbmCA4 target of TbmCA3 (Proto et al., 2011) and the *E. coli* elongation factor (EF-Tu) target of TbmCA2 (Moss et al., 2007). Concerning the latter, cleavage of EF-Tu occurred upon bacterial production of TbmCA2, explaining why the authors did not claim Ef-Tu cleavage as a physiologically relevant event. Interestingly, we also identified an elongation factor of the same protein family (AT4G02930) cleaved *in vitro* and *in vivo* by AtMC9 (Table 2). Furthermore, although the cleavage sites are not conserved, *Arabidopsis* TSN-1 (AT5G07350), TSN-2 (AT5G61780) and GAPB (AT1G42920), a protein sharing 37% amino acid sequence similarity to the YCA1 substrate GAPDH, were all cleaved by rAtMC9 *in vitro* (Table S3). Based on the previous and given that metacaspase biochemical properties are conserved across kingdoms it is tempting to assume that metacaspase physiological substrates may also be conserved. Conservation of protease substrates across species has indeed been suggested as a possible parameter to distinguish physiologically relevant from bystander substrates (Crawford et al., 2011). Alternatively, selected reaction monitoring (SRM) was recently combined with positional proteomics to determine the time-course of caspase-mediated proteolysis (Agard et al., 2012). The derived catalytic efficiencies for hundreds of substrates provide a quantitative understanding of proteolysis in biological processes and this may assist in selection of candidate substrates whether we are interested in protein inactivation/degradation or limited proteolysis often leading to activation of proteins.

We report for the first time on the proteome-wide identification of substrates for a plant protease by means of positional proteomics. Using the Arg/Lys specificity of AtMC9 and the re-occurrence of these cleavage events within *in vivo* and *in vitro* analyses, our work has revealed a set of 74 potential physiological *Arabidopsis* AtMC9 substrates (Table 2). Furthermore, we validated, by *in vitro* approaches, the predicted cleavage of GRF5, PEPCK1, CSY3, GMD2, GRP7, PORB, trypsin inhibitor-2, NAD synthetase, a PPR-containing protein and a LEA domain-containing protein. The functional importance of PEPCK1 cleavage by AtMC9 was demonstrated *in planta* as PEPCK1 activity is enhanced by proteolysis, which is translated into improved seedling establishment. We believe that further confirmation of the here suggested metacaspase functionalities will greatly improve our knowledge on plant proteases.

AtMC9 during seed development

The identified AtMC9 substrates represent various functional categories, highlighting the multi-functionality of plant metacaspases (Tsiatsiani et al., 2011). As also seen with caspases, multiple transcription and translation elongation factors, ribosomal proteins and structural components were cleaved by AtMC9 (Lüthi and Martin, 2007). Moreover, proteins involved in embryonic development and growth, such as several late embryogenesis abundant (LEA) proteins were frequently found among the cleaved proteins. LEA or LEA domain-containing proteins are small proteins encoded by numerous genes in the *Arabidopsis* genome. LEA proteins accumulate during late stages of seed development, are very hydrophilic, heat stable, unfolded in solution and may gain a considerable degree of ordered structures upon water loss. Moreover, they dissociate aggregated proteins and exhibit an overall protein stabilization function that could protect cell membranes during seed dehydration and

osmotic stress (Galau et al., 1987; Chakrabortee et al., 2007; Boucher et al., 2010). Recently, certain LEA and chaperone proteins from Medicago (*Medicago truncatula*), such as a glycine-rich RNA-binding protein similar to the GRP7 AtMC9 substrate, were associated with seed longevity and protection against desiccation-induced protein aggregation (Chatelain et al., 2012). It is possible that AtMC9 cleavage of LEA proteins during water stress could induce the conformational changes necessary for LEA protein folding and their protective functions. Supporting this, from the cleavage pattern of LEA proteins we could not deduce that their processing is a destructive event.

AtMC9 during primary metabolism

An involvement of AtMC9 in primary plant metabolism is inferred by the observed cleavage of key metabolic enzymes. Based on this, we demonstrated that AtMC9 cleaves PEPCK1 *in vivo* and is thus involved in gluconeogenesis by regulating PEPCK activity. *Arabidopsis* PEPCK1 phosphorylation is not yet shown, but the putative Ser 62 phosphosite is conserved in tomato (*Lycopersicon esculentum*), cucumber and Guinea grass that undergo phosphorylation (Leegood and Walker, 2003; Bailey et al., 2007). Cleavage of PEPCK1 at Arg101 would remove the N-terminal fragment containing the putative phosphorylated Ser and thus recover PEPCK1 activity, but this remains to be identified (Figure 9A). Since our proteomics data indicated PEPCK1 fragments of higher abundance in the WT or *AtMC9*-OE than in *AtMC9*-KO (Figure 9B), we were expecting to see in the protein blot the appearance of pronounced fragments concomitant with increasing protease concentration. Instead, we found a striking reduction of the PEPCK1 precursor in the *AtMC9*-OE extract which not accompanied by the appearance of stable fragments with increased intensity (Figure 9C). We believe that this is inherent to the different nature of the two methodologies that we used for protein detection; Western blot analysis and N-terminal COFRADIC. For instance, it is possible that the generated fragments by cleavage at the COFRADIC-identified sites, are downstream re-processed and therefore they become too short to contain an epitope for antibody recognition. Thus, although the COFRADIC cleavage sites indicate abundant PEPCK1 fragments in the WT or *AtMC9*-OE proteome, these cannot be detected by Western blot analysis.

Another AtMC9 substrate involved in metabolic processes is citrate synthase-3 (CSY3), an enzyme catalyzing the conversion of acetyl-CoA and oxaloacetate (OAA) to citrate at the onset of the glyoxylate cycle in germinating seeds (otherwise the citric acid cycle). Of note is that the glyoxylate cycle and gluconeogenesis are consecutive steps in the catabolism of fatty acids into sugars during seed germination and seedling establishment. This suggests that regulation of this pathway by AtMC9 may be controlled at different levels.

Metacaspase activation

Amongst the common *in vivo* AtMC9 substrates, AtMC9 itself was identified. This confirms the validity of the used methodology and selection criteria since it is known that AtMC9 is auto-catalytically activated by proteolytic separation of its two subunits (Vercammen et al., 2004). However, the auto-processing site detected upon bacterial over-production, R¹⁸³, was not found here. The neo-N-terminal AtMC9 peptide identified in both *AtMC9*-OE and WT proteomes was generated after cleavage at R²¹⁴, which is the Arg directly following the known auto-processing site of recombinant AtMC9. The neo-N-terminal peptide spanning amino acid positions 184-214 was not identified by MS/MS here possibly because it is too long and therefore we only

identified the neo-N-terminal peptide DAGLKFR starting at amino acid position 215. All pro-caspases except pro-caspase-10 are processed at two Asp residues during zymogen maturation (Fuentes-Prior and Salvesen, 2004). Similarly, spruce metacaspase mCII-Pa is processed at R¹⁸⁸ and K²⁶⁹ (Bozhkov et al., 2005). We suggest that native separation of the two AtMC9 subunits involves processing at two sites; first at R¹⁸³ and then at R²¹⁴, however whether cleavage at the second site is needed for enzyme activity remains to be established. Supporting this, a recent study on the activity of recombinant rAtMC4 showed that additional sites may be processed besides the primary site at K²²⁵ (Watanabe and Lam, 2011b). R¹⁹⁰, a site that is conserved in AtMC5, correlates to one of the two processing sites in mCII-Pa (Bozhkov et al., 2005). Overall our results support that proteolytic separation of metacaspases *in vivo* occurs by more than one cleavage.

AtMC9 cleavage specificity

Using the sequence information spanning the cleavage sites of native proteins that were targeted by AtMC9 *in vivo* and *in vitro*, we expanded the specificity consensus of AtMC9. The number of identified proteins in the *in vitro* degradomic analysis was lower than in the *in vivo* analyses and this is due to the extraction buffer used which is optimal for rAtMC9 activity rather than for protein extraction. However, the *in vitro* set of identified cleavage sites is very informative on the protease's specificity since the enzyme was in abundance and in optimal conditions for its activity.

We showed that Lys is relatively preferred over Arg. Moreover, the P1 site plays an important role in determining the enzyme specificity and is often occupied by acidic amino acids. This is another property that may differentiate metacaspases from caspases (Timmer and Salvesen, 2007), besides the strong metacaspase preference for basic residues at P1. Another important observation in accordance with the known tetrapeptide AtMC9 specificity (Vercammen et al., 2006), is the AtMC9 preference for basic P3 and P1 residues. In this way, KxK, KxR, RxR and RxK combinations as in VRPR-AMC were very frequently cleaved in native seedling proteins by AtMC9. Although this observation could hint towards a subsite cooperativity (Schilling et al., 2008) between P3 and P1 for protease specificity, such a suggestion would be pre-mature without supporting structural and kinetic data. The crystal structure of the *T. brucei* type-I metacaspase MCA2 was recently resolved and it was shown that type-I metacaspases act as monomers that are activated in the presence of their substrates and calcium, upon removal of a N-terminal domain that normally blocks access of the substrate to the active site (McLuskey et al., 2012). As type-II metacaspases are only found in plants, and in contrast to type-I metacaspases processing is necessary for their activity (Vercammen et al., 2007), we assume that a different activation mechanism is employed. Resolution of a type-II metacaspase structure is strongly anticipated as it will additionally explain how substrates are recognised and bound.

METHODS

RT-Quantitative PCR Gene Expression Assays

Total RNA was extracted from dry seeds as described in Bentsink et al. 2006. One µg of RNA was reverse-transcribed using the iScript™ cDNA Synthesis Kit (Bio-Rad). Real-time quantitative PCR was performed in a final volume of 5 µl, including 10% (v/v) of 6-fold diluted cDNA as template for PCR, 200 nM of each primer and 50% (v/v) SYBR green mix (Invitrogen), using a LightCycler 480 II instrument (Roche Applied Science), according to the manufacturer's instructions. The housekeeping gene *UBC* (At1g14400) was used as internal reference and for normalization of transcript levels. Metacaspase gene-specific primer sequences used for quantitative PCR are listed in supporting information Table S5.

In silico analysis of AtMC9 expression levels

The fold change values in *AtMC9* transcript levels in different tissues and developmental stages were calculated relative to the mean expression of appropriate control genes (Schmid et al., 2005). Data were derived from the eFP browser (Winter et al., 2007) in the relative mode, using the gene At5g04200 as primary gene ID and the developmental map as data source.

Plant material and total proteome extraction

Seeds of *Arabidopsis thaliana* Columbia (Col-0), *AtMC9* T-DNA insertion line (GK-506H04-019739) and *AtMC9* overexpression line (Vercammen et al., 2006) were placed in round plastic plates containing wet Whatman paper for stratification at 4 °C in the dark. After 3 days, seeds were transferred to 21 °C with a 16-hour light/8-h dark photoperiod with light intensity of 80-100 µmol m⁻²s⁻¹ and left to germinate for 48 hours. Young seedlings, including seed coats, were collected, frozen in liquid nitrogen and ground into a fine powder using a mortar and a pestle. To achieve a total protein content of approximately 4 mg, 0.5 g of ground tissue was defrosted in 1 ml of buffer containing 1% (w/v) CHAPS, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA and 10% glycerol in PBS at pH 7.5, further containing the suggested amount of protease inhibitors mixture according to the manufacturer's instructions (Roche). The sample was centrifuged at 16,000 g for 10 min at 4 °C and guanidinium hydrochloride was added to the cleared supernatant to reach a final concentration of 4 M.

For the *in vitro* COFRADIC analysis, protein extracts were prepared in the *AtMC9* optimal activity buffer consisting of 50 mM MES (pH 5.5), 150 mM NaCl, 10% (w/v) sucrose, 0.1% (w/v) CHAPS and 10 mM DTT. Four mg of total seedlings protein extract was incubated at 30 °C with recombinant *AtMC9* (Vercammen et al., 2004) for a total duration of one hour; here, every 20 min 10 µg of fresh protease (855 units) was pre-activated in the same optimal activity buffer at room temperature for 15 min and then added to the protein extract (enzyme activity is 57U/µl and one unit corresponds to the enzyme activity that catalyzes the hydrolysis of 1 µmol VRPR-AMC per minute at 30 °C). This was repeated two times. The same procedure was used for the control sample, but without adding rAtMC9. Protein concentration measurements were performed using the DC protein assay of BIORAD and protein extracts were further modified for N-terminal COFRADIC analysis as described in Staes et al., 2011.

During the hypocotyl elongation experiments, seeds were sowed on half strength Murashige and Skoog (MS) salts lacking sucrose and then stratified for 3 days at 4 °C in the dark. Prior transfer to a Lovibond incuba-

tor (24 h in the dark at 21°C), seeds were exposed for 30 min to light. Hypocotyl length measurements were done using software Image J (Abramoff et al., 2004).

N-terminal COFRADIC

For the N-terminal COFRADIC analysis (Staes et al., 2011), all α -free amines, either pre-existing or generated upon proteolysis, and all ϵ -amines (lysines) were blocked using NHS-esters of butyric acid isotopic variants. Consequently, trypsin digestion only occurs after arginine.

For the analysis of the *in vivo* generated samples, primary amines in the KO proteome were $^{12}\text{C}_4$ -butyrylated, while WT or OE primary amines were $^{13}\text{C}_4$ -butyrylated. For the *in vitro* analysis, primary amines in the rAtMC9 treated sample were $^{13}\text{C}_4$ -butyrylated while the untreated proteome was $^{12}\text{C}_4$ -butyrylated. Upon mixing equal amounts of the labeled proteomes, tryptic digestion generated internal, non-N-terminal peptides which were removed by strong cation exchange at low pH. Further, following RP-HPLC peptide separations with 2,4,6-trinitrobenzene sulfonic acid (TNBS) modification in between, internal peptides were further depleted and N-terminal peptides were isolated. Finally, LC-MS/MS analysis was used for peptide identification and quantification (see below).

Peptide identification and quantification

The generated peptide mixtures were analyzed by LC-MS/MS, using an Ultimate 3000 (Dionex, Amsterdam, The Netherlands) capillary HPLC system in-line connected to an LTQ Orbitrap XL (Thermo Fisher Scientific, Bremen, Germany) mass spectrometer. Samples were first loaded on a trapping column (made in-house, 100 μm internal diameter (I.D.) x 20 mm (length) filled with 5 μm (diameter) C18 beads (Reprosil-HD, Dr. Maisch). After back-flushing from the trapping column, samples were loaded on a capillary reverse-phase column (made in-house, 75 μm I.D. x 150 mm, 5 μm C18 beads (Reprosil-HD). Here, peptides were loaded with solvent A (0.1% trifluoroacetic acid, 2% acetonitrile), and separated using a linear gradient from 2% solvent A' (0.05% formic acid) to 55% solvent B (0.05% formic acid in 80% acetonitrile) at a flow rate of 300 nl/min followed by a washing step for 15 min with 100% buffer B.

The mass spectrometer was operated in data-dependent mode, with automatic switching between MS and MS/MS acquisition for the six most abundant peaks in a given MS spectrum. Full scan MS spectra were acquired in the Orbitrap at a target value of 1E6 with a resolution of 60,000. The six most intense ions were then isolated for fragmentation in the linear ion trap, with a dynamic exclusion of 60 s. Peptides were fragmented after filling the ion trap at a target value of 1E4 ion counts. From the MS/MS data in each LC run, Mascot Generic Files were created using the Mascot Distiller software (version 2.3.2.0, Matrix Science, London, UK). While generating these peak lists, grouping of spectra was allowed with a maximum intermediate retention time of 30 s and a maximum intermediate scan count of 5 was used where possible. Grouping was done at 0.005 Da precursor tolerance. A peak list was only generated when the MS/MS spectrum contained more than 10 peaks. There was no de-isotoping and the relative signal to noise (S/N) limit was set at 2. Peak lists were then searched with the Mascot search engine (Matrix Science, Perkins et al., 1999) using the Mascot Daemon interface (version 2.2.1, Matrix Science). Spectra were searched against the TAIR database (release 8). The following search parameters were set. Mass tolerance on peptide precursor ions was set to 10 ppm (with MASCOT C13 option

set to 1) and that of fragment ions was set to 0.5 Da. Allowed peptide charges were 1+, 2+, 3+ and instrument setting were put to ESI-TRAP. Endoproteinase semi-Arg C/P was used, allowing for 1 missed cleavage. Variable modifications were set to pyro-glutamate formation of N-terminal glutamine and fixed modifications included methionine oxidation to its sulfoxide derivative, S-carbamidomethylation of cysteine and butyrylation (12C4 or 13C4) of lysine side chain and peptide N-termini. Only peptides with a score higher than the MASCOT identity threshold set at 99% confidence were withheld. False discovery rates were calculated for each experiment (0.52% for the KO/OE *in vivo* experiment, 0.49% for the KO/WT *in vivo* experiment and 0.16% for the *in vitro* experiment) as described by Käll et al., 2008. MS/MS spectra of all identified peptides are available in PRIDE database (Martens et al., 2005) (Accession number: 22147 account login username: review52029 and password: ^d4_G5Ca).

Identified peptides were quantified using Mascot Distiller Toolbox version 2.3.2.0 (Matrix Science) in the precursor mode. This software tries to fit an ideal isotopic distribution on the experimental data based on the peptide average amino acid composition. This is followed by extraction of the ion chromatogram (XIC) of both peptide components (light and heavy) from the raw data. Ratios are calculated from the area below the light and heavy isotopic envelope of the corresponding peptide (integration method 'trapezium', integration source 'survey'). To calculate this ratio value, a least squares fit to the component intensities from the different scans in the XIC peak was created. MS scans used for this ratio calculation were situated in the elution peak of the precursor determined by the Distiller software (XIC threshold 0.3, XIC smooth 1, maximum XIC width 250). To validate the calculated ratio, the standard error on the least square fit had to be below 0.16 and the correlation coefficient of the isotopic envelope should be higher than 0.97. All further data management was done in ms_lims (Helsens et al., 2010). Quantitation validation was performed by manual inspection of the MS spectrum and significantly enriched peptide ratios (outliers) were detected using robust statistics. The R software package (version 2.8.1) was used to calculate the median and Huber scale for each experimental dataset.

Assessment of synthetic peptide cleavage

Peptides (PORB; $^{338}\text{PPFQKYITKGYVSETESG}^{356}$, trypsin inhibitor 2; $^{41}\text{VGFSYCAPRIFPTFCDQN}^{58}$, GRF5; $^{245}\text{GDDIKEAPKEVQKVDEQA}^{262}$, PEPCK1; $^{11}\text{GDGGFSFPKGPVMPKIT}^{28}$ and $^{93}\text{SASLASLTRESGPKV-VRG}^{110}$, LEA-domain containing protein; $^{205}\text{AGGEEMREREGKESAGGV}^{222}$ and $^{315}\text{AAEQAAARAKDYTLQ-KAVE}^{332}$, GRP7; $^{55}\text{TFKDEKAMKDAIEGMNGQ}^{72}$ and $^{69}\text{MNGQDLGRSITVNEAQS}^{86}$) were synthesized on an Applied Biosystems 433A Peptide Synthesizer. These peptides carried an acetyl group at their N-terminus and an amide at their C-terminus. Cleavage of the peptides by recombinant AtMC9 was assessed by RP-HPLC separation and peptide fractions were analyzed by MALDI-TOF-MS. Substrate peptides were dissolved at 100 μM in AtMC9 activity buffer and the protease was used at increasing concentrations (0 to 500 nM). Reactions were done at 30°C for 30 min and quenched by adding trifluoroacetic acid to a final concentration of 1%.

Assessment of protein cleavage

cDNA clones received from the Arabidopsis Biological Resource Center (ABRC) were used as templates for *in vitro* transcription and translation (TnT[®] T7/T3 Coupled Reticulocyte Lysate System, Promega) of radio-labeled proteins. Translates were produced in 50 μl reactions and one tenth was used for each rAtMC9 digestion. The

production of recombinant AtMC9 protease is described in Vercammen et al., 2004. A range of rAtMC9 concentrations (0-1000 nM) was used, prepared in the AtMC9 activity buffer, and reactions were done at 30°C for 30 min. The inactive AtMC9 mutant (rAtMC9C147AC29A) was used as a negative control (Belenghi et al., 2007). Assay products were separated on SDS-PAGE. These gels were dried and exposed to phosphorimaging plates, which were scanned by a phosphorimager 445SI.

Recombinant PEPCK1 production and immunodetection

Arabidopsis PEPCK1 (At4g37870) open reading frame (ORF) was PCR amplified from Col-0 seed cDNA and cloned into the bacterial protein expression vector pDEST17. Bacterial (BL21-DE3-pLysE cells) protein production was induced in Luria Broth (LB) media growing cultures of successfully transformed bacterial clones using 0.2 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) overnight at 20°C. Denatured PEPCK1 protein was isolated from protein inclusion bodies and subsequently purified using His-select Nickel Affinity Gel resin (Sigma-Aldrich) in a buffer consisting of 50 mM sodium phosphate, 0.3 mM NaCl and 4 M urea. For the production of *Arabidopsis* PEPCK1 antibodies, 0.2 mg of purified recombinant protein was injected 4 times as antigen into each rabbit (Eurogentec, Seraing, Belgium) and then the rabbit antiserum was used for PEPCK1 immunodetection in plant extracts. Immunodetection of proteins was performed in 2-days old *Arabidopsis* seedling extracts (or infiltrated tobacco leaves) prepared in 1% (w/v) CHAPS, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA and 10% glycerol in PBS at pH 7.5, further containing the suggested amount of protease inhibitors mixture according to the manufacturer's instructions (Roche). Extracts were immediately mixed with Laemli SDS-PAGE loading buffer and heated at 95°C for 5 minutes. 40 μ g of protein extract was used for the immunodetection of PEPCK1, AtMC9, catalase 2 (CAT2, At4g35090) or GFP. For re-probing purposes, Immobilon P (Millipore) membranes were stripped using 0.1 M 2-mercaptoethanol, 2% SDS and 62.5 mM Tris-HCl at pH 6.7. Immunoreactive proteins were visualised using a chemiluminescence kit (Western Lightning® Plus-ECL) from Perkin Elmer.

PEPCK activity assay

Seedlings were grown on wet Whatman paper for 6 days at 21°C after a 3 day-stratification period at 4°C in the dark. Sampling included the remaining seed coats. The carboxylation activity of PEPCK was measured spectrophotometrically at 340 nm as previously described by Malone et al., 2007 using UV-Star 96-well plates from Greiner bio-one. One unit of PEPCK activity corresponds to the production of 1 μ mol product per minute at 25°C. The values represent the mean of 5 measurements \pm 95% confidence interval (Cumming et al., 2007).

Generation of marker lines and protein subcellular localization

N- and C-terminal GFP fusions to AtMC9 and PEPCK1 ORFs, under the transcriptional control of (CaMV) 35S promoter, were established using the Gateway compatible plant transformation vectors (Karimi et al., 2007) pK7WGF2 and pK7FWG2. For the development of *Arabidopsis* AtMC9:GFP reporter lines, the native AtMC9 promoter (a region of 1500 nucleotides upstream the ORF) was amplified by PCR from genomic DNA and introduced into a Gateway compatible entry vector. Using Multisite Gateway technology, AtMC9 native promoter and ORF were fused by introduction to the destination vector pGWB4 (Nakagawa et al., 2007) that was further used

to generate stable *Arabidopsis* transformants using *Agrobacterium tumefaciens* mediated plant transformation.

Tobacco leaf transient protein expression was performed as described by Sparkes et al., 2006 and tissue samples were analysed via laser scanning confocal microscopy using a Zeiss LSM5 Exiter equipped with software package ZEN 2009 Light edition (Zeiss, Jena, Germany). GFP fluorescence of tobacco leaves was imaged in a single channel setting with 488 nm excitation while *Arabidopsis* roots stained with propidium iodine were imaged in a multichannel setting with 488 and 543 nm excitation.

ACKNOWLEDGEMENTS

We thank Dr. Richard Leegood for his scientific input and the donation of *atpepck1* T-DNA insertional mutant seeds. We acknowledge the contribution of Davy Maddelein in Figure 5. LT is indebted to VIB International PhD Program for a predoctoral fellowship. This work was supported by Research Foundation-Flanders (grant number G.0038.09N) and Ghent University (Multidisciplinary Research Partnership "Biotechnology for a Sustainable Economy" project).

Supporting information

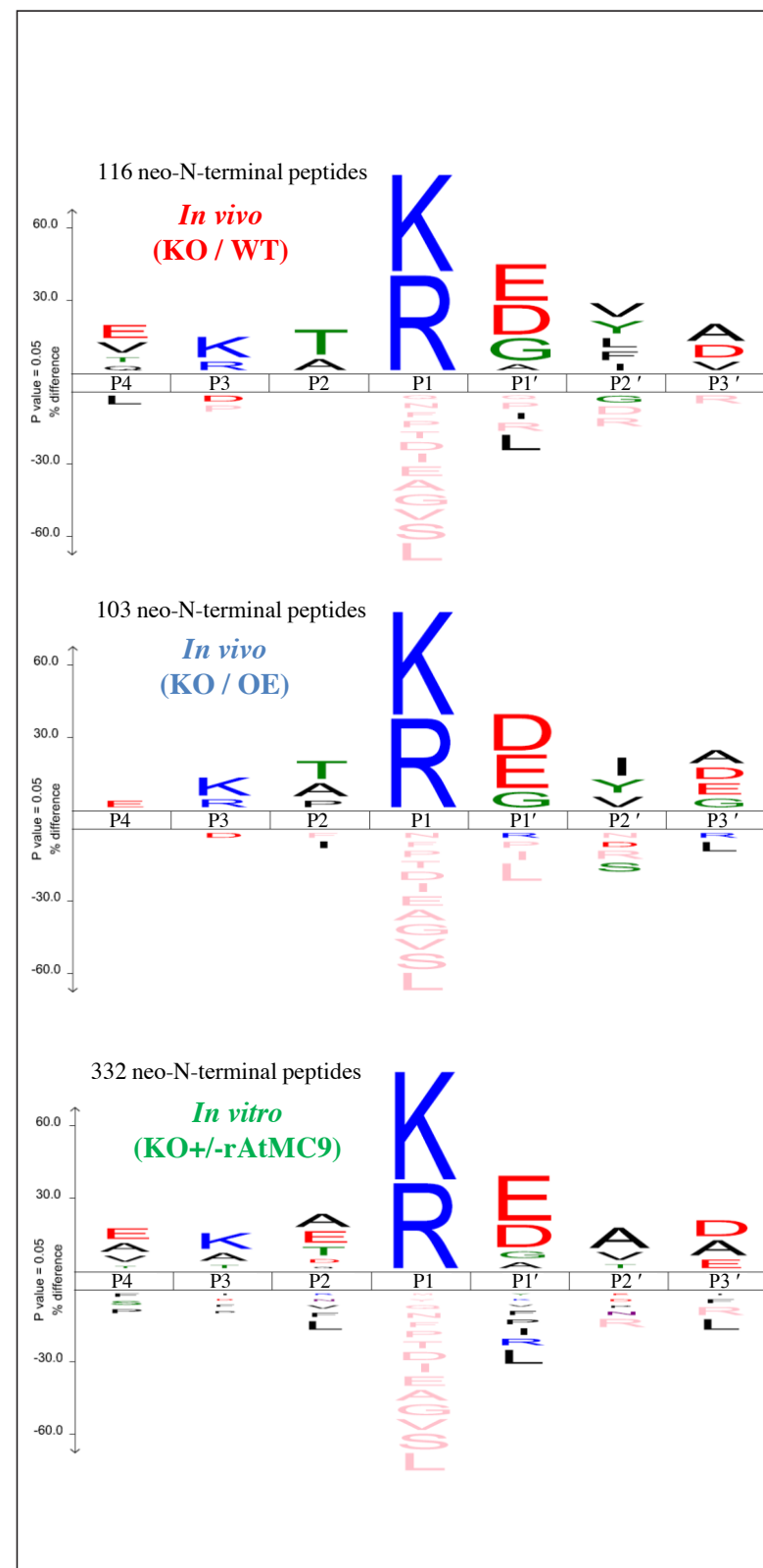


Figure S1. Icelogos illustrating the amino acid frequencies in substrate positions P4-P3' (Colaert et al., 2009).

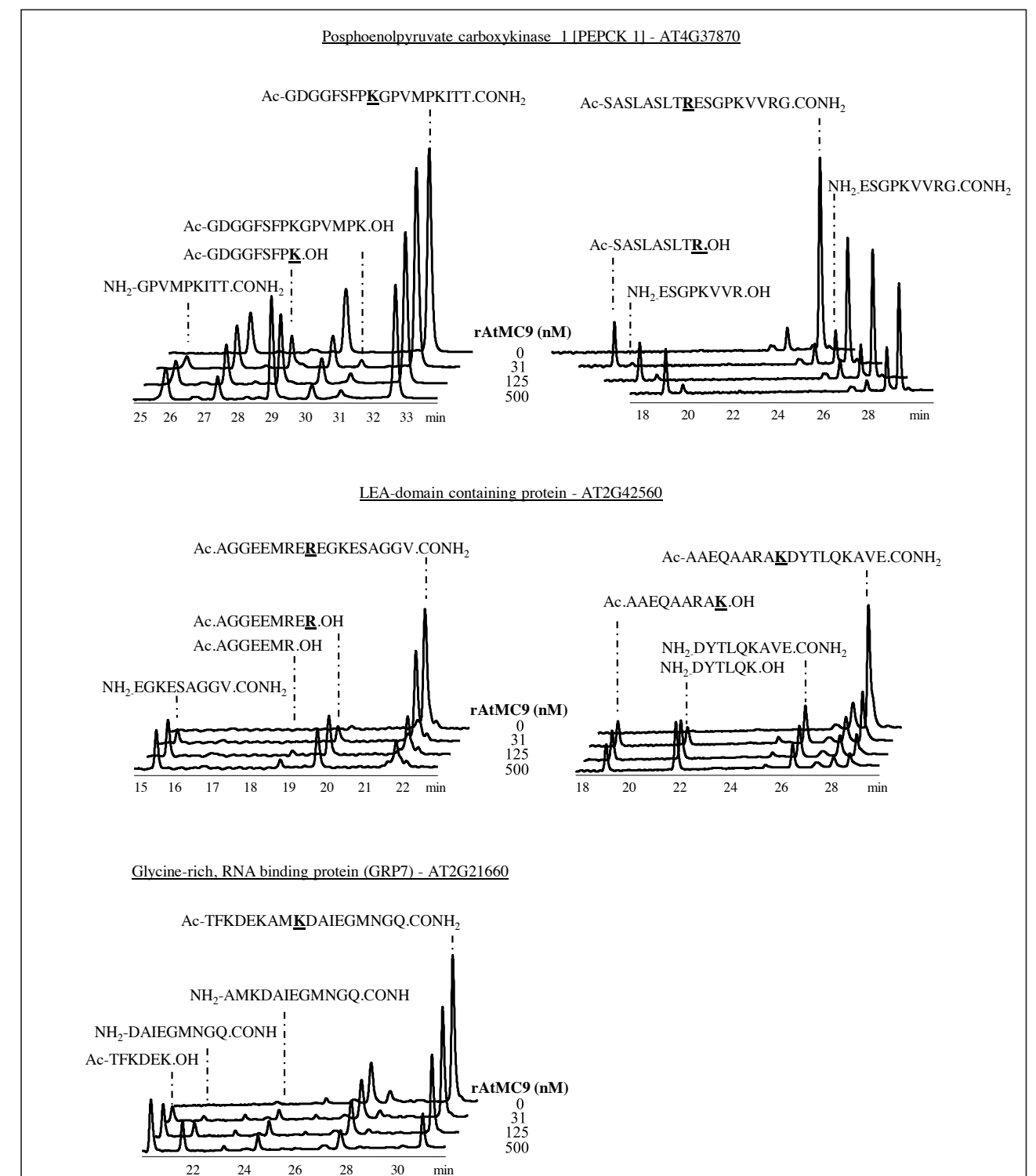


Figure S2. Synthetic 18 amino acid long peptides (50 μ M) with acetylated N-termini and amidated C-termini, designed based on the COFRADIC deduced cleavage of PEPCK1, GRP7 and a LEA-domain containing protein. Peptides were incubated with an increasing concentration of rAtMC9. The identified cleavage sites during COFRADIC are underlined in bold and were also cleaved by rAtMC9.

Table S1. List of endogenous AtMC9 substrates identified in the WT proteome.

Peptides were identified either as unique or significantly ($p < 0.02$) up-regulated ions in the proteome of active protease (WT). In the upper half of the table proteolysis reporter peptides generated after cleavage of Arg or Lys residues are presented, while at the bottom half, peptides cleaved at other residues are found. Amino acid positions of P1' sites are shown between brackets in the P4-P3' column. Fold times of higher WT ion intensity in comparison to the KO ion counterpart is

shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions which is not applicable due to the absence of heavy peptide ion.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	P4-P3'	Sequence	Description	score (s)	threshold (t)	E value	Mass error (ppm)	# of spectra
Proteolysis of Arg or Lys											
AT1G03630	686.852	2	4.79	YITK(337)GYV	GYVSEEEAGKR	protochlorophyllide oxidoreductase C	52	44	1.58E-03	-0.656	1
AT1G04750	990.079	2	2.64	VLDR(156)GEK	GEKIELLVDKTENLR	vesicle-associated membrane protein 721	105	41	3.98E-09	-2.022	1
AT1G07780	468.790	2	4.08	SVAK(109)DIS	DISKVAR	phosphoribosylanthranilate isomerase 1	55	42	5.01E-04	-0.321	2
AT1G08200	933.447	3	4.55	VSSK(329)EFY	EFYGEGYDDSDKRIIPDMTIINR	UDP-D-apiose/UDP-D-xylose synthase 2	53	42	7.94E-04	-2.288	1
AT1G09310	861.011	2	N/A	TEAK(165)EAV	EAVAIKEAVAVKEAA	Protein of unknown function, DUF538	45	40	3.16E-03	-0.465	1
AT1G11480	517.827	2	11.77	ARPR(404)ELV	ELVLKER	eukaryotic translation initiation factor-related	64	39	3.16E-05	-0.097	1
AT1G12000	865.969	2	2.68	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	94	44	1.00E-07	-1.445	3
AT1G29350	492.240	2	3.26	SRPR(86)GAN	GANNNTYNR	Kinase-related protein of unknown function (DUF1296)	46	42	3.98E-03	0.407	1
AT1G29350	747.011	3	7.98	QNAK(525)ELD	ELDFQYSPFSAQQSMQSR	Kinase-related protein of unknown function (DUF1296)	41	39	6.31E-03	-1.072	2
AT1G47540	766.347	2	5.58	CAPR(49)IFP	IFPTFCDQNCR	trypsin inhibitor 2	63	39	3.98E-05	-1.241	1
AT1G47540	777.053	3	N/A	AYDR(31)KCL	KCLKEYGGDVGFSYCAPR	trypsin inhibitor 2	48	43	3.16E-03	-1.890	1
AT1G49240	743.874	2	2.72	MNQR(53)DAY	DAYVGDEAQSKR	actin 8	58	43	3.16E-04	-1.144	2
AT1G49240	924.983	2	7.34	SIEK(241)NYE	NYELPDGQVITIGAER	actin 8	54	44	1.00E-03	-3.030	1
AT1G49970	556.774	2	14.57	FEKR(363)DYD	DYDGTLAQR	CLP protease proteolytic subunit 1	65	43	6.31E-05	0.360	1
AT1G51710	526.776	2	4.29	VATR(157)ELF	ELFGELDR	ubiquitin-specific protease 6	54	43	7.94E-04	0.475	1
AT1G55490	705.384	3	4.06	VLTK(382)ETS	ETSTIVGDGSTQDAVKKR	chaperonin 60 beta	52	44	1.58E-03	-0.237	2
AT1G55490	686.717	3	9.09	TVAR(114)EVE	EVELEDPVENIGAKLVR	chaperonin 60 beta	55	43	6.31E-04	-1.215	4
AT1G56070	868.436	3	2.07	ICLK(544)DLQ	DLQDDFMGGAEIHKSDPVVSR	Ribosomal protein S5/Elongation factor G/III/V family protein	63	44	1.26E-04	-1.114	1
AT1G59990	755.390	2	2.27	VKKR(566)AFL	AFLKAEAEPEPQ	DEA(D/H)-box RNA helicase family protein	51	44	2.00E-03	-2.452	1
AT1G64970	788.432	2	7.42	TTTR(46)GNV	GNVAVAAAATSTEALR	gamma-tocopherol methyltransferase	45	44	7.94E-03	-2.603	1
AT1G65090	998.850	3	2.45	TSER(162)TLQ	TLQDDKKSGNAKSEEVQEPEKR	unknown protein	62	44	1.58E-04	-2.773	1
AT1G65930	891.909	2	2.31	EKIK(8)VAN	VANPIVEMDGDDEMTR	cytosolic NADP+-dependent isocitrate dehydrogenase	74	40	3.98E-06	-2.357	1
AT1G79340	842.749	3	2.16	IESR(191)GFH	GFHIGGNKKDEDEAEIEITK	metacaspase 4	99	43	2.51E-08	-2.534	2
AT1G79550	785.967	2	2.52	GTLK(12)EAD	EADLKGKSVFVR	phosphoglycerate kinase	49	41	1.58E-03	-1.784	1
AT2G01250	738.484	2	8.68	VESK(6)VVV	VVVPEVSLKRR	Ribosomal protein L30/L7 family protein	57	31	2.51E-05	-1.017	3
AT2G18960	711.397	2	5.86	GSYR(902)ELS	ELSEIAEQAKR	H(+)-ATPase 1	79	44	3.16E-06	-1.478	1
AT2G19730	670.423	2	2.96	SVNK(88)SIL	SILKKEFPR	Ribosomal L28e protein family	45	38	2.00E-03	0.224	1
AT2G21060	787.327	2	8.10	HMAR(150)ECS	ECSQGGGGYSGGGGGGR	glycine-rich protein 2B	156	34	6.31E-15	0.827	2
AT2G21620	1045.106	2	3.44	YSFR(14)EVV	EVVLPVSLIPVVPEPELER	Adenine nucleotide alpha hydrolases-like superfamily protein	49	41	1.58E-03	-3.017	1
AT2G21660	829.773	3	6.55	TDDR(23)ALE	ALETAFAYGVDVIDSKIINDR	glycine-rich, RNA binding protein AtGRP7	76	44	6.31E-06	-2.574	1
AT2G23670	782.416	2	32.69	TAEK(104)GVD	GVDAAEKGLEAAER	homolog of Synechocystis YCF37	91	45	2.51E-07	-0.640	1
AT2G26250	873.949	2	3.00	IVNR(18)GIE	GIEPSGPNAGSPTFSVR	3-ketoacyl-CoA synthase 10	66	44	6.31E-05	-3.150	1
AT2G28000	922.494	2	2.03	TIAR(103)AIE	AIELPNAMENAGAALIR	chaperonin-60alpha	75	44	7.94E-06	-2.008	2
AT2G28000	793.105	3	2.95	PRGR(82)NVV	NVVLDVDFGSPKVVNDGVTIAR	chaperonin-60alpha	44	43	7.94E-03	-1.389	1
AT2G32240	757.410	2	2.97	VKSR(709)DID	DIDLFSFSPTKR	Unknown protein	52	44	1.58E-03	-1.388	1
AT2G33150	606.365	2	2.91	AAHR(60)TPL	TPLCKSKR	peroxisomal 3-ketoacyl-CoA thiolase 3	45	43	6.31E-03	-0.330	2
AT2G36620	777.431	2	5.11	NIPK(149)SAA	SAAPKAAKMGGGGGGR	ribosomal protein L24	69	44	3.16E-05	-3.091	1
AT2G36640	772.726	3	3.73	QKTR(338)EST	ESTESGAQKAETKDSAAGR	embryonic cell protein 63	71	44	2.00E-05	-1.857	1
AT2G36640	881.454	2	37.61	EAIK(378)SKL	SKLTMPSDIVEETR	embryonic cell protein 63	54	44	1.00E-03	-3.635	1
AT2G40490	731.449	2	8.05	VTER(47)KVS	KVSATSEPLLLR	Uroporphyrinogen decarboxylase	76	39	2.00E-06	-0.753	1
AT2G41560	876.957	2	2.20	NLLR(7)DFE	DFEVEAKNPSLEAR	autoinhibited Ca(2+)-ATPase, isoform 4	52	44	1.58E-03	-2.683	1
AT2G41840	848.937	2	2.47	LASK(271)ALS	ALSTSKPDPVVEDQA	Ribosomal protein S5 family protein	59	44	3.16E-04	-1.828	1

AT2G42560	824.466	3	5.03	VGSK(454)AVD	AVDLTKEKAAVAADTVVGYTAR	LEA domain-containing protein	83	42	7.94E-07	-1.741	2
AT2G42560	844.140	2	6.20	ARAK(324)DYT	DYTLQKAVEAKDVAAEKAQR	LEA domain-containing protein	57	43	3.98E-04	-2.214	2
AT2G42560	564.334	2	6.46	QTTK(607)NIV	NIVIGDAPVR	LEA domain-containing protein	87	44	5.01E-07	0.178	1
AT2G42790	963.045	2	N/A	KEDK(86)GLK	GLKLYDPGYLNTAPVR	citrate synthase 3	44	43	7.94E-03	-4.314	1
AT2G47470	695.841	2	N/A	YIEK(320)GSD	GSDYASKETER	thioredoxin family protein	46	43	5.01E-03	-2.447	1
AT3G02480	834.118	3	2.75	QQMK(49)EKA	EKAQGAADVVKDKTGMNKSH	LEA family protein	68	44	3.98E-05	-1.360	1
AT3G02480	723.721	3	N/A	MKEK(51)AQQ	AQGAADVVKDKTGMNKSH	LEA family protein	88	44	3.98E-07	-1.153	2
AT3G05420	1119.553	2	3.80	GRVR(527)EIV	EIVMDNVNPGSKVEGNSER	acyl-CoA binding protein 4	55	43	6.31E-04	-2.324	3
AT3G09840	563.302	2	4.76	KSKK(15)DFS	DFSTAILER	cell division cycle 48	77	43	3.98E-06	0	1
AT3G11710	890.419	3	2.51	QTTK(11)ALS	ALSELAMDSSTLNAEAESSAGDGAGPR	lysyl-tRNA synthetase 1	78	41	2.00E-06	-1.349	1
AT3G12960	796.890	2	N/A	EKGK(58)EQS	EQSAASGDQTIQIR	unknown protein	67	43	3.98E-05	-1.508	2
AT3G13300	711.369	2	N/A	VGER(708)NLD	NLDVSSVEEISR	Transducin/WD40 repeat-like superfamily protein	79	44	3.16E-06	-0.633	1
AT3G13470	1057.539	2	N/A	VLTK(378)EMT	EMTTIVDGTGTEAVNKR	TCP-1/cpn60 chaperonin family protein	61	44	2.00E-04	-2.698	1
AT3G15280	914.971	2	8.20	TEYR(135)GVE	GVEDLHQQTGGVEVKSP	unknown protein	60	44	2.51E-04	-2.571	1
AT3G15670	1023.501	3	2.33	IKNK(134)AQD	AQDAAQYTKETAQGAAYTKETAEAGR	LEA family protein	112	43	1.26E-09	-2.445	1
AT3G15670	1015.508	2	8.24	QYTK(143)ETA	ETAQGAAYTKETAEAGR	LEA family protein	105	44	7.94E-09	-2.169	2
AT3G16910	726.421	2	2.41	HILR(556)TKA	TKAKEMGPVPR	acyl-activating enzyme 7	57	42	3.16E-04	-1.654	1
AT3G17520	562.840	2	2.90	EKAK(184)EAK	EAKEAAKR	LEA family protein	49	42	2.00E-03	-0.267	1
AT3G17520	565.269	2	8.73	AKDK(262)ASQ	ASQSYDAAAR	LEA family protein	78	41	2.00E-06	-0.266	1
AT3G20410	1060.031	2	2.23	NTR(23)SVE	SVEVGVNTQDPPSYTPQAR	calmodulin-domain protein kinase 9	72	44	1.58E-05	-2.455	1
AT3G23990	694.384	3	N/A	LMLK(49)GVE	GVEDLADAVKVTMGPKGR	heat shock protein 60	64	44	1.00E-04	-1.394	7
AT3G48870	954.522	2	7.24	ISDR(501)FLP	FLPDKAIDLIDEAGSR	Clp ATPase	47	44	5.01E-03	-1.626	2
AT3G53040	930.762	3	2.30	TVLK(449)EAD	EADQMTGQTFNDVGEIDDEEKVR	LEA protein, putative	54	39	3.16E-04	-2.689	1
AT3G53040	694.849	2	9.92	QKTK(111)ETA	ETADYTADKAR	LEA protein, putative	75	43	6.31E-06	-0.360	1
AT3G53040	708.865	2	17.26	DKTK(155)ETA	ETAEYTAEKAR	LEA protein, putative	59	44	3.16E-04	-0.848	1
AT3G61260	849.518	2	5.21	SLDR(92)DVK	DVKLADLSKEKR	Remorin family protein	66	39	2.00E-05	-1.120	2
AT3G62870	721.945	2	2.75	MGSK(236)SQA	SQAKTKAKER	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	55	39	2.51E-04	-1.248	1
AT4G02510	768.374	3	2.55	VSSR(607)EFS	EFSFGGKEVDQEPSGEGVTR	translocon at the outer envelope membrane of chloroplasts 159	55	42	5.01E-04	-1.347	1
AT4G02930	1078.610	2	10.26	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	87	43	3.98E-07	-2.227	3
AT4G05180	606.846	2	N/A	DQAR(110)DFS	DFSLALKDR	photosystem II subunit Q-2	46	44	6.31E-03	-0.743	1
AT4G13780	574.784	2	2.02	PLFK(570)ELE	ELENDEVAR	methionine--tRNA ligase, putative	55	43	6.31E-04	0.349	2
AT4G15410	882.445	2	4.91	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	96	44	6.31E-08	-2.723	1
AT4G20360	942.974	2	5.55	ERAR(127)GIT	GITINTATVEYETENR	RAB GTPase homolog E1B	98	44	3.98E-08	-2.229	1
AT4G20360	792.396	2	5.92	SVAK(113)KYD	KYDEIDAAPEER	RAB GTPase homolog E1B	45	43	6.31E-03	-2.590	1
AT4G21020	1085.072	2	4.78	EKAK(116)DTA	DTAYNAKEKAKDYAER	LEA family protein	72	44	1.58E-05	-3.090	1
AT4G21020	778.401	3	6.24	ESTK(218)NAA	NAAQTVTEAVVGPEEDAERKAR	LEA family protein	54	44	1.00E-03	-1.372	1
AT4G21020	997.976	2	N/A	EKAK(156)DYA	DYAEDTMDNAEKAR	LEA family protein	71	42	1.26E-05	-1.605	4
AT4G21280	598.849	2	11.70	DQAR(103)DFA	DFALALKDR	photosystem II subunit QA	51	43	1.58E-03	-0.753	1
AT4G24190	891.441	3	2.31	KTLR(74)SNA	SNAEKFEFQAEVSR	Chaperone protein htpG family protein	81	43	1.58E-06	-2.358	2
AT4G26270	765.867	2	4.14	VIDK(229)SFG	SFGFDTAVEEAQR	phosphofructokinase 3	58	42	2.51E-04	-1.177	3
AT4G27320	575.856	2	4.82	GATR(223)EAI	EAIIVTKSR	Adenine nucleotide alpha hydrolases-like superfamily protein	52	40	6.31E-04	-0.522	1
AT4G27440	680.852	2	4.30	YITK(337)GYV	GYVSETESGKR	protochlorophyllide oxidoreductase B	63	44	1.26E-04	-0.515	2
AT4G29270	547.825	2	2.32	EGHR(178)SVT	SVTEKNLR	HAD superfamily, subfamily IIIB acid phosphatase	62	42	1.00E-04	-1.189	2
AT4G31700	895.465	2	3.26	IVKK(119)GEN	GENDLPGLTDTEKPR	ribosomal protein S6	66	44	6.31E-05	-2.348	2
AT4G34450	459.265	2	N/A	GIEK(27)GAV	GAVLQEAR	coatamer gamma-2 subunit, putative	53	43	1.00E-03	0.109	1
AT4G37870	832.996	2	6.99	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxylase 1	76	41	3.16E-06	-2.704	1
AT4G38630	741.350	2	11.41	QKDK(258)DGD	DGD TASASQETVAR	regulatory particle non-ATPase 10	56	42	3.98E-04	-1.283	1
AT4G39260	935.997	2	6.58	GRSR(48)GFG	GFGVTFKDEKAMR	cold, circadian rhythm, and RNA binding 1	50	44	2.51E-03	-3.102	1
AT4G39860	573.819	2	3.95	IQPR(179)SLV	SLVAAQPEAR	unknown protein	87	45	6.31E-07	0	2
AT5G04200	473.771	2	N/A	LFGR(216)DAG	DAGLKFR	metacaspase 9	58	44	3.98E-04	0.317	2

AT5G06140	583.787	2	9.46	EQPR(9)NIS	NISGSMQSPR	sorting nexin 1	76	43	5.01E-06	-0.858	1
AT5G11520	580.808	2	3.16	VVRK(99)AEQ	AEQQLINDR	aspartate aminotransferase 3	55	44	7.94E-04	0	2
AT5G12140	909.458	2	4.04	GGVR(15)DID	DIDANANDLQVESLAR	cystatin-1	102	44	1.58E-08	-1.761	3
AT5G16050	914.473	2	17.77	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	59	44	3.16E-04	-2.244	2
AT5G16130	621.31	2	N/A	LTGK(181)DVV	DVVFEYPVEA	Ribosomal protein S7e family protein	52	43	1.26E-03	-0.484	1
AT5G16840	764.391	3	9.73	ERTK(178)SVY	SVYAAAEQTVSSAGSAVMKNR	binding partner of acd11 1	111	44	2.00E-09	-1.092	1
AT5G28060	751.977	2	2.45	KKIR(121)GVK	GVKKTAKGDTK	Ribosomal protein S24e family protein	49	35	3.98E-04	-1.465	1
AT5G35910	726.354	2	N/A	ESTR(578)DLI	DLIMGAANTNEGR	Polynucleotidyl transferase	64	43	7.94E-05	-1.310	1
AT5G44120	513.321	2	2.62	NTAK(95)LSF	LSFVAKGR	RmlC-like cupins superfamily protein	57	41	2.51E-04	0.488	3
AT5G44310	1016.525	2	14.65	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	119	44	3.16E-10	-2.314	2
AT5G44310	1016.519	2	17.89	EKTK(150)DFA	DFAEETKEKVNEGASR	LEA family protein	114	44	1.00E-09	-1.920	2
AT5G44310	1168.120	2	20.22	VKEK(148)TKD	TKDFAEETKEKVNEGASR	LEA family protein	77	44	5.01E-06	-3.299	1
AT5G44310	1016.526	2	23.42	ERTK(70)DYA	DYAEQTKKNVNEGASR	LEA family protein	125	44	7.94E-11	-2.806	2
AT5G44310	1002.504	2	23.93	EKTK(110)DYA	DYAEAKDKVNEGASR	LEA family protein	127	44	5.01E-11	-1.897	1
AT5G47210	845.431	2	21.15	RYAK(341)EAA	EAAAPAIGDTAQFPSLG	Hyaluronan / mRNA binding family	49	44	3.16E-03	-1.954	1
AT5G52440	1011.504	2	18.41	TLER(145)EIG	EIGLDDISTPNVYNQNR	Bacterial sec-independent translocation protein mttA/Hef106	60	44	2.51E-04	-2.573	1
AT5G54190	665.847	2	5.61	YITK(341)GYV	GYVSESEAGKR	protochlorophyllide oxidoreductase A	66	44	6.31E-05	-0.451	1
AT5G54310	979.434	2	N/A	QSGK(469)DFD	DFDFSSSLMDGMFTKH	ARF-GAP domain 5	78	38	1.00E-06	-2.249	1
AT5G56000	949.441	2	5.11	SSKK(604)TME	TMEINPENSIMDEL	HEAT SHOCK PROTEIN 81.4	80	41	1.26E-06	-2.636	1
AT5G58070	783.926	2	3.85	TEKK(6)EME	EMEVVKGLNVER	temperature-induced lipocalin	45	44	7.94E-03	-2.171	1
AT5G61170	474.253	2	5.89	SGQR(131)DLD	DLDQVAGR	Ribosomal protein S19e family protein	48	44	3.98E-03	-0.106	1
AT5G62575	667.899	2	3.83	TREK(42)ALL	ALLAEDSALKR	unknown protein	72	41	7.94E-06	-0.825	1
AT5G62810	1012.486	2	2.61	SNIR(347)EIN	EINDMPPNPQPLSDPR	peroxin 14	56	42	3.98E-04	-3.065	1
AT5G65410	730.414	2	2.44	KRHR(196)TKF	TKFTAQKER	homeobox protein 25	51	44	2.00E-03	-1.508	1
proteolysis of other amino acids											
AT1G04270	945.993	2	3.25	LDAL(33)LDM	LDMSTDDLKLFSSR	cytosolic ribosomal protein S15	57	44	5.01E-04	-2.698	1
AT1G04820	996.450	2	2.08	AEIT(293)NSA	NSAFEPASMMAKCDPR	tubulin alpha-4 chain	46	39	2.00E-03	-2.160	1
AT1G04820	864.741	3	2.65	HEQL(287)SVA	SVAEITNSAFEPASMMAKCDPR	tubulin alpha-4 chain	70	41	1.26E-05	-3.126	2
AT1G05240	741.906	2	2.04	VSPA(22)IPQ	IPQLLDLYYR	Peroxidase superfamily protein	50	44	2.51E-03	-1.282	1
AT1G08050	961.517	2	2.00	RSQL(625)LVL	LVLESPTKPSGRGEQS	Zinc finger (C3HC4-type RING finger) family protein	56	44	6.31E-04	-2.082	1
AT1G11480	702.402	3	2.38	NLKP(378)VAQ	VAQLLEQPEVKTEKDR	eukaryotic translation initiation factor-related	48	42	2.51E-03	-1.663	1
AT1G14930	740.473	2	2.16	YTLD(64)GKP	GKPEVIKEKR	Polyketide cyclase/dehydrase and lipid transport superfamily protein	45	37	1.58E-03	-1.758	1
AT1G22530	645.406	2	2.27	AEKN(97)ALA	ALAEKELVR	PATELLIN 2	51	38	5.01E-04	-0.854	1
AT1G27400	652.389	2	3.67	FISH(118)IQV	IQVNQAAKQR	Ribosomal protein L22p/L17e family protein	47	41	2.51E-03	0.154	7
AT1G32990	959.879	2	2.29	LLKA(152)AGV	AGVEKGSKDPQQDKVGVITIDQLR	plastid ribosomal protein L11	54	41	5.01E-04	-2.503	1
AT1G49240	844.950	2	2.13	GMNQ(52)KDA	KDAYVGDQAQSKR	actin 8	97	44	5.01E-08	-1.422	2
AT1G66200	969.487	2	2.38	ILVM(92)CDA	CDAYTPAGEPIPTNKR	glutamine synthase clone F11	99	44	3.16E-08	-1.187	2
AT1G66250	595.380	2	N/A	TQVV(51)ALL	ALLKAQEIR	O-Glycosyl hydrolases family 17 protein	50	38	6.31E-04	-0.925	2
AT1G67090	1244.660	2	2.12	FETL(71)SYL	SYLPDLTDSSELAKEVDYLIR	ribulose biphosphate carboxylase small chain 1A	46	44	6.31E-03	-2.332	1
AT1G67430	892.458	2	2.53	VKYS(6)QEP	QEPDNQTKSCKAR	Ribosomal protein L22p/L17e family protein	55	44	7.94E-04	-2.580	2
AT1G79550	994.109	2	2.31	RSVG(9)TLK	TLKEADLKGKSVFVR	phosphoglycerate kinase	64	36	1.58E-05	-2.417	3
AT2G05220	885.759	3	2.78	LASL(110)GMS	GMSDTSGISAVEPQQAMAPIAAFGR	Ribosomal S17 family protein	104	42	6.31E-09	-2.223	2
AT2G14880	684.355	2	2.01	TSAT(50)ESS	ESSEPTATNKR	SWIB/MDM2 domain superfamily protein	45	44	7.94E-03	-0.220	1
AT2G15010	750.862	2	2.55	QVEA(25)QKT	QKTCPCSQSTR	Plant thionin	71	42	1.26E-05	-1.600	1
AT2G19730	498.316	2	2.18	NNCF(20)LVK	LVKQFGR	Ribosomal L28e protein family	52	40	6.31E-04	0.503	6
AT2G19730	878.055	2	3.11	PKLS(85)VNK	VNKSILKKEFPR	Ribosomal L28e protein family	66	34	6.31E-06	-1.767	9
AT2G37660	923.511	2	2.17	SVTV(69)SAA	SAAATTEPLTVLVTGAGGR	NAD(P)-binding Rossmann-fold superfamily protein	83	44	1.26E-06	-2.276	1
AT2G46280	1069.509	2	9.31	INAL(297)AFN	AFNPDGKSFSSGGEDGYVR	TGF-beta receptor interacting protein 1	105	42	5.01E-09	-3.088	2
AT3G04790	925.994	2	2.32	IRLQ(197)DLF	DLFKEFGCESKLR	Ribose 5-phosphate isomerase, type A protein	62	44	1.58E-04	-2.162	1

AT3G05420	597.295	2	N/A	RDIE(515)SEV	SEVEVSQEGR	acyl-CoA binding protein 4	48	43	3.16E-03	-0.335	1
AT3G09630	793.500	2	3.30	NVML(333)KLN	KLNPYAKTAKR	Ribosomal protein L4/L1 family	83	36	2.00E-07	-1.577	5
AT3G09680	780.496	2	2.82	LLAL(133)FKE	FKEKKEKPR	Ribosomal protein S12/S23 family protein	46	35	7.94E-04	-2.181	2
AT3G13710	498.334	2	N/A	GAVL(156)VLI	VLIHAVIR	prenylated RAB acceptor 1.F4	39	37	6.31E-03	1.862	1
AT3G14067	556.803	2	2.03	TLYD(653)ACE	ACETSKLR	Subtilase family protein	46	44	6.31E-03	0.810	1
AT3G18190	503.338	2	N/A	NVLL(298)IQK	IQKSILR	TCP-1/cpn60 chaperonin family protein	49	37	6.31E-04	-0.597	1
AT3G21055	945.98	2	2.82	KKYA(90)QVC	QVCVTMPTAKICRY	photosystem II subunit T	68	44	3.98E-05	-1.852	1
AT3G21370	664.364	2	2.31	TSFF(350)AKA	AKADQKVDSR	beta glucosidase 19	46	45	7.94E-03	0.528	1
AT3G63040	1119.547	2	2.51	MSNE(154)EVC	EVCETSCSILQEELTIR	unknown protein	76	43	5.01E-06	-2.324	1
AT3G63460	546.808	2	N/A	LHSL(399)VTE	VTEQSLVSR	transducin family protein / WD-40 repeat family protein	59	44	3.16E-04	0	1
AT4G02620	583.854	2	2.00	TNYL(45)IVD	IVDSKTTVR	vacuolar ATPase subunit F family protein	60	41	1.26E-04	-0.515	1
AT4G14880	771.048	3	2.36	SVLN(237)VDL	VDLIDEVVQVSSDESIDMAR	O-acetylserine (thiol) lyase (OAS-TL) isoform A1	68	43	3.16E-05	-2.274	1
AT4G18100	678.985	2	2.13	VPLL(7)TKK	TKKVVKKR	Ribosomal protein L32e	22	19	5.01E-03	-1.622	1
AT4G18100	766.411	2	4.34	VTLM(54)PNV	PNVGYGSDKKTR	Ribosomal protein L32e	85	44	7.94E-07	-1.568	1
AT4G18440	747.880	2	37.74	KLVT(60)STK	STKVTAMDGVSSR	L-Aspartase-like family protein	68	44	3.98E-05	-3.213	1
AT4G18970	1273.099	2	2.07	IAMG(23)DPI	DPIAPCYFIFGDSLVDSGNNR	GDSL-like Lipase/Acylhydrolase superfamily protein	73	41	6.31E-06	-1.965	2
AT4G20360	978.525	2	2.25	IGSS(110)VAK	VAKKYDEIDAAPEER	RAB GTPase homolog E1B	52	44	1.58E-03	-3.120	1
AT4G27160	722.680	3	2.25	FDFE(84)GPQ	GPQQGYQLLQCCNELR	seed storage albumin 3	49	42	2.00E-03	0.092	4
AT4G27220	524.835	2	N/A	VPLG(56)ELI	ELILEKR	NB-ARC domain-containing disease resistance protein	53	38	3.16E-04	-0.382	1
AT4G27230	679.877	2	2.12	LEYL(61)AAE	AAEVLELAGNAAR	histone H2A 2	70	44	2.51E-05	-1.178	2
AT4G28520	790.368	2	21.23	QQQP(178)WEG	WEGQQQQGQGR	cruciferin 3	92	41	7.94E-08	9.311	3
AT4G31700	711.435	3	2.28	RKFT(168)NKK	NKKGKEVSKAPKIQR	ribosomal protein S6	45	36	1.26E-03	-1.620	1
AT4G31700	606.846	2	2.47	KLFN(149)LKK	LKKEDDVR	ribosomal protein S6	48	44	3.98E-03	-0.743	1
AT4G34290	628.342	2	7.68	AAAS(54)SDP	SDPTTTTKTR	SWIB/MDM2 domain superfamily protein	58	44	3.98E-04	-1.036	1
AT5G02380	844.352	3	3.90	FVAE(60)NDA	NDACKCGSDCKCNPCTCK	metallothionein 2B	41	31	1.00E-03	-1.383	1
AT5G04200	643.812	2	N/A	AMDL(229)MDL	MDLLETMTAR	metacaspase 9	66	42	3.98E-05	-1.167	1
AT5G07350	891.488	2	N/A	IKEL(135)LQL	LQLEELAKQEGYGR	TUDOR-SN protein 1	74	44	1.00E-05	-2.190	2
AT5G10360	909.162	3	2.39	MKFN(5)VAN	VANPTTGCQKKLEIDDDQKLR	Ribosomal protein S6e	47	43	3.98E-03	-2.055	2
AT5G15200	638.379	2	2.12	YYRN(10)YGK	YGKTFKGPR	Ribosomal protein S4	50	41	1.26E-03	-0.628	1
AT5G16130	667.908	2	2.28	MLED(130)VAF	VAFPAEIVGKR	Ribosomal protein S7e family protein	62	40	6.31E-05	-2.249	1
AT5G17830	478.776	2	N/A	QQQT(46)LSG	LSGRFFR	Plasma-membrane choline transporter family protein	54	45	1.26E-03	4.500	1
AT5G19770	909.928	2	2.23	ITNA(295)VFE	VFEPASMMAKCDPR	tubulin alpha-3	72	41	7.94E-06	-2.310	1
AT5G26710	901.494	2	2.39	VNPN(665)TKK	TKKETLALGDSNMR	Glutamyl/glutaminyl-tRNA synthetase, class Ic	69	44	3.16E-05	-2.165	1
AT5G41060	463.208	2	N/A	GKAN(333)DDI	DDIEMGR	DHHC-type zinc finger family protein	68	39	1.26E-05	3.894	2
AT5G54740	1050.500	2	2.97	ENPM(92)GPQ	GPQQQSSLMCCNELR	seed storage albumin 5	81	42	1.26E-06	-1.906	4
AT5G61170	611.357	2	2.23	QTMN(113)IVD	IVDLDTKGGR	Ribosomal protein S19e family protein	61	43	1.58E-04	-0.328	4

Table S2. List of endogenous AtMC9 substrates identified in the OE proteome.

Interesting peptides were identified either as unique or significantly ($p < 0.02$) up-regulated ions in the proteome of active protease (OE). In the upper half of the table proteolysis reporter peptides generated after cleavage of Arg or Lys residues are presented, while at the bottom half, peptides cleaved at other residues are found. Amino acid positions of P1' sites are shown between brackets in the P4-P3' column. Fold times of higher OE ion intensity in comparison to the KO ion counterpart is shown

in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions which is not applicable due to the absence of heavy peptide ion.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	P4-P3'	Sequence	Description	score (s)	threshold (t)	E value	Mass error (ppm)	# of spectra
Proteolysis of Arg or Lys											
AT1G01320	581.838	2	2.56	GYSR(178)VIR	VIRDSEAPK	Tetratricopeptide repeat (TPR)-like superfamily protein	46	43	5.01E-03	-0.775	1
AT1G02150	508.812	2	N/A	VTSK(474)EAV	EAVLELLR	pentatricopeptide repeat (PPR)-containing protein	45	41	3.98E-03	0.985	1
AT1G02780	506.315	2	7.22	SIHK(145)SKA	SKAEKAR	Ribosomal protein L19e family protein	49	43	2.51E-03	0.198	1
AT1G04750	984.059	2	2.54	VLDR(156)GEK	GEKIELLVDKTENLR	vesicle-associated membrane protein 721	77	43	3.98E-06	-2.747	2
AT1G04820	1148.417	2	41.84	ALEK(431)DYE	DYEEVGAEGGDEDEDEGEEY	tubulin alpha-4 chain	52	19	5.01E-06	-1.089	4
AT1G05500	913.957	2	2.95	SLER(412)VLK	VLKNDTTDEENASSR	Calcium-dependent lipid-binding (CaLB domain) family protein	53	44	1.26E-03	-4.546	1
AT1G07780	468.790	2	6.09	SVAK(109)DIS	DISKVAR	phosphoribosylanthranilate isomerase 1	63	42	7.94E-05	-0.107	2
AT1G09310	861.012	2	N/A	TEAK(165)EAV	EAVAIKEAVAVKEAA	Protein of unknown function, DUF538	45	40	3.16E-03	-1.919	1
AT1G11480	517.827	2	12.20	ARPR(404)ELV	ELVLKER	eukaryotic translation initiation factor-related	59	39	1.00E-04	-0.097	1
AT1G12000	861.957	2	4.23	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	46	44	6.31E-03	-2.149	2
AT1G13270	656.400	2	11.92	RKMR(78)ELE	ELETKSKVR	methionine aminopeptidase 1B	44	41	5.01E-03	-0.915	1
AT1G20760	501.247	2	10.93	NSSK(928)DFG	DFGGAAFSR	Calcium-binding EF hand family protein	48	42	2.51E-03	0.500	1
AT1G27330	951.610	2	3.59	LADR(11)KIE	KIEKFDKNILKR	Ribosome associated membrane protein RAMP4	54	31	5.01E-05	-2.577	1
AT1G29350	1120.015	2	6.68	QNAR(525)ELD	ELDFQYSPFSAQQSMQSR	Kinase-related protein of unknown function (DUF1296)	59	40	1.26E-04	-2.770	1
AT1G47540	777.053	3	N/A	AYDR(31)KCL	KCLKEYGGDVGFSYCAPR	trypsin inhibitor 2	81	43	1.58E-06	-2.577	1
AT1G48920	964.997	2	2.18	IFVK(407)GFD	GFDASLSEDDIKNTLR	nucleolin like 1	76	44	6.31E-06	-2.542	2
AT1G49240	739.861	2	4.01	MNQK(53)DAY	DAYVGDEAQSQR	actin 8	80	43	2.00E-06	-1.015	2
AT1G49240	924.983	2	17.14	SIEK(241)NYE	NYELPDGQVITIGAER	actin 8	57	44	5.01E-04	-3.139	1
AT1G49970	556.774	2	31.94	FEKR(363)DYD	DYDGTLAQR	CLP protease proteolytic subunit 1	70	43	2.00E-05	-0.540	1
AT1G51710	526.776	2	N/A	VATR(157)ELF	ELFGELDR	ubiquitin-specific protease 6	45	43	6.31E-03	-0.475	1
AT1G54870	664.849	2	3.89	EKIK(288)NFG	NFGSEVPMKR	Aldehyde reductase	49	44	3.16E-03	-0.979	1
AT1G55090	765.896	2	N/A	NSNK(710)EIG	EIGVVAANSQDPSAGL	NAD synthetase	67	44	5.01E-05	-1.569	1
AT1G55490	686.717	3	12.72	TVAR(114)EVE	EVELEDPVENIGAKLVR	chaperonin 60 beta	58	43	3.16E-04	-1.215	4
AT1G56070	868.771	3	2.50	ICLK(544)DLQ	DLQDDFMGGAIEIKSDPVVSFR	Ribosomal protein S5/Elongation factor G/III/V family protein	51	44	2.00E-03	-1.095	1
AT1G72370	602.749	2	2.92	LGTK(33)NCN	NCNYQMER	40s ribosomal protein SA	45	34	7.94E-04	-0.831	1
AT1G74030	1014.510	3	2.06	VSHR(420)SGE	SGETEDNFIADLSVGLASGQIKTGAPCR	enolase 1	44	43	7.94E-03	-1.842	1
AT1G78150	919.49	2	2.42	AKLK(240)EIG	EIGGNNIFADGKVEAR	unknown protein	50	44	2.51E-03	-3.593	1
AT1G79850	573.858	2	3.82	SKTK(117)SFV	SFVALPVIAR	ribosomal protein S17	62	38	3.98E-05	-0.960	2
AT1G79920	614.788	2	32.00	AVEK(606)EFE	EFEMALQDR	Heat shock protein 70 (Hsp 70) family protein	45	41	3.98E-03	0.244	1
AT2G01250	738.486	2	8.60	VESK(6)VVV	VVVPESVLKRR	Ribosomal protein L30/L7 family protein	54	31	5.01E-05	-3.390	2
AT2G01540	1084.597	2	1.98	RDHR(26)SSD	SSDPYIVLNVADQTLKTR	Calcium-dependent lipid-binding (CaLB domain) family protein	101	44	2.00E-08	-2.769	1
AT2G18960	707.383	2	3.63	GSYR(902)ELS	ELSEIAEQAKR	H(+)-ATPase 1	64	44	1.00E-04	-0.637	4
AT2G19730	670.423	2	4.80	SVNK(88)SIL	SILKKEFPR	Ribosomal L28e protein family	45	38	2.00E-03	-0.448	1
AT2G21060	787.329	2	N/A	HMAR(150)ECS	ECSQGGGGYSGGGGGR	glycine-rich protein 2B	61	34	2.00E-05	-1.017	1
AT2G21660	790.859	2	4.44	KAMK(64)DAI	DAIEGMNGQDLDR	glycine-rich, RNA binding protein AtGRP7	41	39	6.31E-03	-3.608	2
AT2G26250	873.948	2	5.23	IVNR(18)GIE	GIEPSGPNAGSPTFSVR	3-ketoacyl-CoA synthase 10	53	44	1.26E-03	-3.036	1
AT2G28000	922.495	2	2.83	TIAR(103)AIE	AIELPNAMENAGAALIR	chaperonin-60alpha	48	44	3.98E-03	-3.256	2
AT2G28000	1189.157	2	4.35	PRGR(82)NVV	NVVLEDFGSPKVVNDGVTIAR	chaperonin-60alpha	55	43	6.31E-04	-3.324	2
AT2G32240	757.411	2	3.58	VKSR(709)DID	DIDLFSFSPTKR	unknown protein	71	44	2.00E-05	-1.454	1
AT2G33410	527.265	2	8.81	NASR(100)NFD	NFDGANVR	RNA-binding (RRM/RBD/RNP motifs) family protein	46	43	5.01E-03	-7.506	1
AT2G36620	777.430	2	4.49	NIPK(149)SAA	SAAPKAAKMGGGGGR	ribosomal protein L24	93	44	1.26E-07	-1.803	1
AT2G42560	633.312	2	2.10	EYMK(310)ETG	ETGSTAAEQAAR	LEA domain-containing protein	58	43	3.16E-04	-0.949	1

AT2G42560	568.320	2	3.39	VEAK(335)DVA	DVAAEKAQR	LEA domain-containing protein	51	44	2.00E-03	-0.176	1
AT2G42560	626.332	2	4.42	MRER(214)EGK	EGKESAGVGGRR	LEA domain-containing protein	60	44	2.51E-04	-1.039	1
AT2G42560	799.398	2	6.23	ATEK(268)GKE	GKEAGNMTAEQAAR	LEA domain-containing protein	88	43	3.16E-07	-1.628	2
AT2G42560	669.812	2	13.16	EKGK(270)EAG	EAGNMTAEQAAR	LEA domain-containing protein	45	41	3.98E-03	-1.794	1
AT2G42560	824.467	3	14.94	VGSK(454)AVD	AVDLTKEKAAVAADTVVGYTAR	LEA domain-containing protein	71	42	1.26E-05	-2.672	1
AT2G42560	844.140	3	16.05	ARAK(324)DYT	DYTLQKAVEAKDVAAEKAQR	LEA domain-containing protein	70	43	2.00E-05	-2.451	1
AT2G43160	486.234	2	2.83	GRSR(236)SVD	SVDNYGSR	ENTH/VHS family protein	49	41	1.58E-03	0	1
AT2G44650	484.793	2	2.01	LLPK(79)AAV	AAVKFER	chloroplast chaperonin 10	58	42	2.51E-04	0.310	1
AT3G02480	1085.079	2	3.81	MKEK(51)AQQ	AQGAADVVKDKTGMNKSH	LEA family protein	70	44	2.51E-05	-3.090	2
AT3G05420	775.873	2	3.94	ATTR(512)DIE	DIESEVEVSQEGR	acyl-CoA binding protein 4	66	42	3.98E-05	-1.161	1
AT3G09840	563.303	2	4.08	KSKK(15)DFS	DFSTAILER	cell division cycle 48	86	44	6.31E-07	-0.622	1
AT3G10650	764.944	2	5.45	LSSR(377)SLA	SLALPVSEPLSVR	unknown protein	56	43	5.01E-04	-2.553	1
AT3G13300	711.37	2	8.67	VGER(708)NLD	NLDVSSVEISR	Transducin/WD40 repeat-like superfamily protein	80	44	2.51E-06	-1.619	1
AT3G15280	762.873	2	3.55	VELK(123)DVQ	DVQQPVDTEYR	unknown protein	55	43	6.31E-04	-1.641	1
AT3G17520	723.858	2	2.25	ESAK(260)DKA	DKASQSYDSAAR	LEA family protein	51	43	1.58E-03	-0.761	1
AT3G17520	562.840	2	13.79	EKAK(184)EAK	EAKEAAKR	LEA family protein	53	42	7.94E-04	-0.356	2
AT3G17520	565.269	2	22.50	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	71	41	1.00E-05	-0.443	1
AT3G22640	480.289	2	9.25	TRSK(231)EIG	EIGQGIIR	cupin family protein	57	42	3.16E-04	0.417	11
AT3G23990	598.849	2	N/A	YAAK(35)EIK	EIKFGVEAR	heat shock protein 60	59	43	2.51E-04	-1.171	1
AT3G44310	1142.062	2	7.18	SSTK(6)DMS	DMSTVQNATPFNGVAPSTTVR	nitrilase 1	83	43	1.00E-06	-2.804	1
AT3G45190	899.454	2	N/A	VHDR(558)DYD	DYDLAGLANNLNQFR	SIT4 phosphatase-associated family protein	46	44	6.31E-03	-2.560	1
AT3G51160	524.279	2	2.38	TAPK(20)ADS	ADSTVVEPR	GDP-D-MANNOSE-4,6-DEHYDRATASE 2, GMD2	78	44	3.98E-06	0.287	1
AT3G53040	930.762	3	2.87	TVLK(449)EAD	EADQMTGQTFNDVGEIDDEEKVR	LEA protein, putative	44	39	3.16E-03	-3.119	1
AT3G53040	686.853	2	7.12	DKTK(133)ETA	ETADYAAEKAR	LEA protein, putative	59	44	3.16E-04	-1.385	2
AT3G53040	694.850	2	11.76	QTKK(111)ETA	ETADYTADKAR	LEA protein, putative	68	44	3.98E-05	-0.865	2
AT3G53040	708.867	2	32.09	DKTK(155)ETA	ETAETAEKAR	LEA protein, putative	60	44	2.51E-04	-3.461	2
AT3G55280	1098.241	2	3.33	KAAC(24)AVK	AVKSGQIVKKPAKKIR	ribosomal protein L23AB	33	19	3.98E-04	-2.643	3
AT3G61260	849.519	2	4.75	SLDR(92)DVK	DVKLADLSKEKR	Remorin family protein	78	39	1.26E-06	-1.532	1
AT3G62530	640.834	2	15.27	VTRR(65)GAI	GAISSAPAESK	ARM repeat superfamily protein	50	44	2.51E-03	-1.016	1
AT4G02930	1078.609	2	10.55	GKAK(99)AIA	AIAFDEIDKAPEEKRR	GTP binding Elongation factor Tu family protein	52	43	1.26E-03	-1.949	2
AT4G15410	882.445	2	9.96	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	58	44	3.98E-04	-2.553	1
AT4G20360	942.975	2	8.30	ERAR(127)GIT	GITINTATVEYETENR	RAB GTPase homolog E1B	102	44	1.58E-08	-2.760	1
AT4G21020	997.977	2	49.67	EKAK(156)DYA	DYAEDTMDNAEKAR	LEA family protein	105	42	5.01E-09	-2.758	2
AT4G27150	781.381	2	2.04	SELR(108)QEE	QEEPVCVCPTLR	seed storage albumin 2	62	43	1.26E-04	-1.153	2
AT4G27160	800.393	2	2.37	NELR(101)QEE	QEEPVCVCPTLK	seed storage albumin 3	54	43	7.94E-04	-2.127	1
AT4G27440	607.361	2	3.42	TVTK(79)SVD	SVDGKKTLR	protochlorophyllide oxidoreductase B	51	43	1.58E-03	-0.082	2
AT4G27440	680.852	2	14.80	YITK(337)GYV	GYVSETESGKR	protochlorophyllide oxidoreductase B	65	44	7.94E-05	-1.030	3
AT4G31700	891.452	2	2.54	IVKK(119)GEN	GENDLPGLTDTEKPR	ribosomal protein S6	71	44	2.00E-05	-2.976	3
AT4G37870	510.309	2	2.51	SLTR(102)ESG	ESGPKVVR	phosphoenolpyruvate carboxykinase 1	43	41	6.31E-03	0.393	1
AT4G37870	832.996	2	11.95	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	80	41	1.26E-06	-2.464	1
AT4G39260	1052.055	2	9.96	DLQR(25)TFS	TFSQFGDVIDSKIINDR	cold, circadian rhythm, and RNA binding 1	78	44	3.98E-06	-2.474	1
AT5G04200	477.785	2	N/A	LFGR(216)DAG	DAGLFR	metacaspase 9	55	42	5.01E-04	0.210	1
AT5G06140	583.787	2	5.09	EQPR(9)NIS	NISGSMQSPR	sorting nexin 1	73	43	1.00E-05	-0.515	1
AT5G07190	609.654	3	2.25	GLVR(77)GPG	GPGGLGKCSTNTFQVR	seed gene 3	51	44	2.00E-03	-1.314	11
AT5G09760	597.811	2	2.05	SEHR(127)TQS	TQSTDQALTR	Plant invertase/pectin methylesterase inhibitor superfamily	53	44	1.26E-03	-0.335	1
AT5G10360	866.973	2	2.50	IVKK(119)GVS	GVSDLPGLTDTEKPR	Ribosomal protein S6e	66	44	6.31E-05	-2.483	2
AT5G12140	909.459	2	2.83	GGVR(15)DID	DIDANANDLQVESLAR	cystatin-1	54	44	1.00E-03	-2.202	1
AT5G16050	914.473	2	N/A	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	48	44	3.98E-03	-2.189	1
AT5G44120	555.771	2	9.19	FQPR(22)FEG	FEGQGSQR	RmlC-like cupins superfamily protein	45	43	6.31E-03	0.451	30
AT5G44310	892.525	2	5.67	DVDK(274)GLE	GLEDSLKAKENR	LEA family protein	64	41	5.01E-05	-2.356	1

AT5G44310	1016.525	2	32.12	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	71	44	2.00E-05	-2.314	2
AT5G44310	1002.504	2	52.30	EKTK(110)DYA	DYAEAEAKDKVNEGASR	LEA family protein	95	44	7.94E-08	-2.596	2
AT5G44310	678.018	3	65.10	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	68	44	3.98E-05	-1.083	2
AT5G44310	1016.519	2	N/A	EKTK(150)DFA	DFAEETKEKVNNEGASR	LEA family protein	128	44	3.98E-11	-1.329	2
AT5G47210	845.431	2	44.65	RYAK(341)EAA	EAAAAPAIQDQAQFPLG	Hyaluronan / mRNA binding family	60	44	2.51E-04	-2.428	2
AT5G52440	1011.504	2	55.35	TLER(145)EIG	EIGLDDISTPNVYNQNR	Bacterial sec-independent translocation protein mttA/Hcf106	54	44	1.00E-03	-2.573	1
AT5G54190	665.847	2	7.10	YITK(341)GYV	GYVSESEAGKR	protochlorophyllide oxidoreductase A	58	44	3.98E-04	-0.827	1
AT5G56000	949.441	2	7.04	SSKK(604)TME	TMEINPENSIMDEL	HEAT SHOCK PROTEIN 81.4	69	41	1.58E-05	-2.161	1
AT5G62575	667.900	2	8.98	TREK(42)ALL	ALLAEDSALKR	unknown proteinplant structures	73	42	7.94E-06	-1.649	2
AT5G64130	1068.436	2	2.11	RYRK(93)SPC	SPCAPSEGGEDGGAAQAEAGGSGN	cAMP-regulated phosphoprotein 19-related protein	49	31	1.58E-04	-1.874	1
AT5G64400	511.262	2	2.17	VAHR(69)AVD	AVDSVMGPR	unknown protein	79	44	3.16E-06	0.784	3
proteolysis of other amino acids											
AT1G07600	1123.835	2	7.18	YNKE(28)CDN	CDNCSGSCNSCGSNCNC	metallothionein 1A	50	43	2.00E-03	-2.004	1
AT1G07670	933.832	3	2.12	NLQM(440)IAK	IAKIAAICNDANVEKSDQQFVSR	endomembrane-type CA-ATPase 4	62	44	1.58E-04	-2.073	1
AT1G11480	839.833	3	2.00	PKLN(375)JKP	LKPVAQLLEQPEVKTEKDR	eukaryotic translation initiation factor-related	41	38	5.01E-03	-3.735	1
AT1G12240	956.936	2	2.30	SVVA(504)EAE	EAEFSCERKSGGSTVR	Glycosyl hydrolases family 32 protein	75	39	2.51E-06	-2.877	2
AT1G14320	636.820	2	3.99	LRIN(101)KML	KMLSCAGADR	Ribosomal protein L16p/L10e family protein	46	43	5.01E-03	-3.067	1
AT1G15690	806.372	2	2.48	EGFS(123)TDN	TDNKPCTYDTR	Inorganic H pyrophosphatase family protein	47	40	2.00E-03	-2.918	2
AT1G17880	492.282	2	3.07	EKLM(8)KMA	KMANTVR	basic transcription factor 3	50	42	1.58E-03	-0.102	1
AT1G27400	648.376	2	3.20	FISH(118)IQV	IQVNQAAKQR	Ribosomal protein L22p/L17e family protein	59	43	2.51E-04	-1.159	3
AT1G29350	872.399	2	6.83	YAML(44)KEC	KECNMDPNETVSR	Kinase-related protein of unknown function (DUF1296)	49	40	1.26E-03	-0.459	1
AT1G47540	537.220	2	3.06	IFPT(53)FCD	FCDQNCR	trypsin inhibitor 2	40	33	2.00E-03	0	1
AT1G65910	483.225	2	N/A	RQPT(568)RGD	RGDCNRR	NAC domain containing protein 28	44	41	5.01E-03	-3.422	1
AT1G66270	846.948	2	2.36	VTTA(25)VDD	VDDPVCPTSLSLR	Glycosyl hydrolase superfamily protein	70	44	2.51E-05	-2.128	1
AT1G66280	861.953	2	2.69	STIA(25)VDD	VDDPVCPTSLSLR	Glycosyl hydrolase superfamily protein	77	44	5.01E-06	-2.671	1
AT1G67430	886.438	2	2.23	VKYS(6)QEP	QEPDNQTKSCKAR	Ribosomal protein L22p/L17e family protein	44	43	7.94E-03	-2.880	1
AT1G68560	654.392	2	2.18	PTQS(28)YKT	YKTIGKGYR	alpha-xylosidase 1	51	41	1.00E-03	-0.689	1
AT1G72370	536.218	2	2.04	IIAF(152)CDT	CDTDSMPR	40s ribosomal protein SA	45	33	6.31E-04	-0.934	2
AT2G04390	710.700	3	2.16	SDTL(116)GIS	GISAVDPQQAMAPIAFGGGR	Ribosomal S17 family protein	67	44	5.01E-05	-1.174	1
AT2G04390	917.451	3	2.60	LASL(110)GMS	GMSDTLGISAVDPQQAMAPIAFGGGR	Ribosomal S17 family protein	53	43	1.00E-03	-2.401	1
AT2G15010	750.863	2	5.46	QVEA(25)QKT	QKTCCPSQSTR	Plant thionin	67	42	3.16E-05	-1.867	1
AT2G19730	585.705	3	2.49	PKLS(85)VNK	VNKSILKKEFPR	Ribosomal L28e protein family	58	34	3.98E-05	0.399	3
AT2G19730	490.311	2	4.76	LARL(123)SAI	SAISKGLR	Ribosomal L28e protein family	56	40	2.51E-04	0.307	1
AT2G19750	550.845	2	2.04	FVTA(48)VVG	VVGFGKKR	Ribosomal protein S30 family protein	54	39	3.16E-04	0.273	2
AT2G21660	718.405	2	2.51	AQYG(33)DVI	DVIDSKIINDR	glycine-rich, RNA binding protein AtGRP7	45	43	6.31E-03	-1.812	1
AT2G22170	862.888	2	2.19	VALA(26)DDE	DDEADCYVTFLLR	Lipase/lipoxygenase, PLAT/LH2 family protein	55	39	2.51E-04	-2.030	1
AT2G27510	806.381	2	2.29	YILD(86)AAE	AAEEAGVDLPYSCR	ferredoxin 3	54	42	6.31E-04	-2.545	1
AT2G30200	523.326	2	3.55	IYVT(155)SLA	SLAAVELLR	acyl-carrier-protein S-malonyltransferase	49	41	1.58E-03	-0.383	1
AT2G33070	808.975	2	2.23	RGKW(155)IKV	IKVEQKGEQGPGR	nitrile specifier protein 2	65	41	3.98E-05	-0.681	1
AT2G38530	548.763	2	2.17	YKIS(110)AST	ASTNCNTVR	lipid transfer protein 2	67	42	3.16E-05	4.290	1
AT2G39390	784.975	2	2.45	AQVL(59)TVI	TVISQKQKSALR	Ribosomal L29 family protein	76	41	3.16E-06	-3.189	2
AT2G42000	978.370	2	4.04	VAGC(45)NDS	NDSCGCPSPCGGNSCR	Plant EC metallothionein-like protein, family 15	73	26	2.00E-07	-1.842	2
AT2G43460	677.404	2	2.29	KQIH(7)EIK	EIKDFLLTAR	Ribosomal L38e protein family	66	42	3.98E-05	-1.552	1
AT2G47610	885.014	2	2.07	GGIM(234)GSK	GSKSQAKTKAKER	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	64	42	6.31E-05	-2.149	1
AT2G47780	915.951	2	2.61	IPIE(222)KIS	KISDILEQDQCRAD	Rubber elongation factor protein (REF)	88	43	3.16E-07	-2.186	2
AT3G04790	919.975	2	2.30	IRLQ(197)DLF	DLFKEFGCESKLR	Ribose 5-phosphate isomerase, type A protein	63	44	1.26E-04	-3.917	2
AT3G07480	949.518	2	2.63	VRTS(28)ATS	ATSAPSPSLGSKKVSADR	2Fe-2S ferredoxin-like superfamily protein	62	44	1.58E-04	-2.583	1
AT3G09260	491.755	2	2.29	LISH(251)AEA	AEAVEAYR	Glycosyl hydrolase superfamily protein	51	43	1.58E-03	-0.408	1
AT3G09260	787.916	2	2.67	PANA(25)DGP	DGPVCPSPNKLRSR	Glycosyl hydrolase superfamily protein	46	44	6.31E-03	-1.843	1
AT3G09390	844.680	3	6.85	NNAE(64)NDA	NDACKCGSDCKDPCTCK	metallothionein 2A	44	29	3.16E-04	-1.620	5

AT3G09630	785.474	2	2.04	NVML(333)KLN	KLNPYAKTAKR	Ribosomal protein L4/L1 family	70	40	1.00E-05	-2.231	4
AT3G11710	610.284	2	5.15	TTLN(25)AAE	AAESSAGDGAGPR	lysyl-tRNA synthetase 1	59	41	1.58E-04	9.930	1
AT3G11830	731.879	2	2.81	QLVS(28)NIN	NINACTAVGDVVR	TCP-1/cpn60 chaperonin family protein	45	44	7.94E-03	-0.821	1
AT3G12960	682.692	3	2.00	TPLE(35)GGK	GGKIAESEPVLEFSSAQR	unknown protein	57	44	5.01E-04	-1.296	1
AT3G14415	875.504	2	3.41	YVPA(265)TIS	TISALEEVVKATQGR	Aldolase-type TIM barrel family protein	48	42	2.51E-03	-2.744	1
AT3G17520	1043.549	3	2.21	ESYE(246)TAK	TAKSKADETLESADKASQSYDSAAR	LEA family protein	76	44	6.31E-06	-2.654	1
AT3G20050	792.924	2	3.18	IDDM(534)IKL	IKLVKDESGEE	T-complex protein 1 alpha subunit	52	44	1.58E-03	-2.967	3
AT3G21370	875.950	2	2.39	PTRA(22)EEG	EEGPVCPKTETLSR	beta glucosidase 19	45	44	7.94E-03	-1.372	2
AT3G21720	552.316	2	2.08	VAAT(258)LIQ	LIQSNIDAR	isocitrate lyase	56	44	6.31E-04	-0.635	1
AT3G51610	953.954	2	2.26	FVRE(214)TVV	TVVSSSMNVGEGFGKD	unknown protein	48	42	2.51E-03	-2.309	2
AT3G53420	682.356	2	12.07	MAKD(5)VEA	VEAVPGEGFQTR	plasma membrane intrinsic protein 2A	69	44	3.16E-05	-1.248	1
AT3G63040	789.714	3	3.76	FMSN(153)EEV	EEVCETSCSILQEEPLTIR	unknown protein	66	42	3.98E-05	-1.733	14
AT4G02450	729.886	2	N/A	KAPA(231)AEE	AEETTSVKEDK	HSP20-like chaperones superfamily protein	63	44	1.26E-04	-1.784	1
AT4G04460	604.780	2	2.57	IPTQ(420)NQQ	NQQSAVDCGR	Sapoin-like aspartyl protease family protein	62	40	6.31E-05	-0.911	1
AT4G09600	1008.491	2	2.02	HVHA(27)AED	AEDSQVGEVVKIDCGGR	GAST1 protein homolog 3	80	43	2.00E-06	-1.687	3
AT4G18970	512.728	2	2.68	FRVT(293)NAG	NAGCCGVGR	GDSL-like Lipase/Acylhydrolase superfamily protein	47	38	1.26E-03	0.489	1
AT4G21020	1067.53	4	2.29	DKVN(138)EGA	EGAYKAADKAEDTKEKAKDYAEDTMDNAKEKAR	LEA family protein	46	42	3.98E-03	-3.094	1
AT4G27160	722.680	3	3.87	FDPE(84)GPQ	GPQQGYQLLQCCNELR	seed storage albumin 3	50	42	1.58E-03	-0.739	4
AT4G27220	524.835	2	N/A	VPLG(56)ELI	ELILEKR	NB-ARC domain-containing disease resistance protein	53	38	3.16E-04	-0.477	1
AT4G28520	931.434	2	N/A	QGQQ(175)GQP	GQPWEGQGQQGQGR	cruciferin 3	66	41	3.16E-05	8.276	1
AT4G34290	813.933	2	2.31	AVTS(49)AAA	AAAASSDPTTTTKTR	SWIB/MDM2 domain superfamily protein	73	44	1.26E-05	-2.399	1
AT4G39260	927.997	2	10.84	QRTE(27)SQF	SQFGDVIDSKIINDR	cold, circadian rhythm, and RNA binding 1	107	45	6.31E-09	-2.535	1
AT5G01800	534.723	2	4.50	SQVH(123)QGN	QGNCEACR	sapoin B domain-containing protein	51	35	2.51E-04	0.281	1
AT5G02380	844.352	3	7.89	FVAE(60)NDA	NDACKCGSDCKCNPCTCK	metallothionein 2B	49	31	1.58E-04	-1.344	3
AT5G08260	687.290	3	2.30	KFPA(327)GYD	GYDPCTESYAENYFNR	serine carboxypeptidase-like 35	37	33	3.98E-03	-1.166	3
AT5G15540	565.856	2	N/A	RIMA(660)IEL	IELLGTIAAR	PHD finger family protein	42	39	5.01E-03	-7.790	1
AT5G17830	478.776	2	N/A	QQQT(46)LSG	LSGRFFR	Plasma-membrane choline transporter family protein	58	45	5.01E-04	4.709	1
AT5G19550	655.325	2	2.64	QGFA(223)SGS	SGSLDTDAQSVR	aspartate aminotransferase 2	47	43	3.98E-03	-0.229	2
AT5G19770	909.929	2	2.08	ITNA(295)VFE	VFEPASMMAKCDPR	tubulin alpha-3	53	41	6.31E-04	-3.631	1
AT5G41060	463.208	2	N/A	GKAN(333)DDI	DDIEMGR	DHHC-type zinc finger family protein	62	39	5.01E-05	3.786	2
AT5G44310	742.949	2	8.06	KGLE(277)DLS	DLSKKAKENR	LEA family protein	47	41	2.51E-03	-0.607	1
AT5G54740	575.307	2	2.30	DVSN(38)PQQ	PQQGKQR	seed storage albumin 5	50	44	2.51E-03	-1.741	13
AT5G54740	1050.499	2	5.97	ENPM(92)GPQ	GPQQSSLMCCNELR	seed storage albumin 5	100	42	1.58E-08	-0.572	6
AT5G55610	683.856	2	2.10	CDAA(61)AAI	AAIDDDYLGAIR	unknown protein	52	44	1.58E-03	-1.172	1
AT5G59845	742.848	2	2.64	LSTA(26)DSS	DSSPCGGKCNVR	Gibberellin-regulated family protein	66	41	3.16E-05	-2.763	3
AT5G64620	634.300	2	2.34	MHVL(133)AAQ	AAQDYPNCR	cell wall/vacuolar inhibitor of fructosidase 2	43	42	7.94E-03	-0.316	1
AT5G67490	1017.509	2	2.22	SRFL(29)SSG	SSGTPPPQAPSPNQDLNR	unknown protein	96	44	6.31E-08	-1.525	1

Table S3. List of endogenous AtMC9 substrates raised by the rAtMC9 driven proteolysis added to the KO proteome. Interesting peptides were identified either as unique or significantly ($p < 0.02$) up-regulated ions in the treated proteome. In the upper half of the table proteolysis reporter peptides generated after cleavage of Arg or Lys residues are presented, while at the bottom half, peptides cleaved at other residues are found. Amino acid positions of P1' sites are shown between brackets in the P4-P3' column. Fold times of higher rAtMC9 treated ion intensity in comparison to the untreated ion counterpart is shown in

the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of single-ton ions which is not applicable due to the absence of light peptide ion.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	P4-P3'	Sequence	Description	score (s)	threshold (t)	E value	Mass error (ppm)	# of spectra
<i>Proteolysis of Arg or Lys</i>											
AT1G01100	941.898	2	98.60	EKKK(96)DEP	DEPAEESDGLGFGLFD	60S acidic ribosomal protein family	68	35	5.01E-06	-0.903	1
AT1G02700	580.757	2	9.18	PWPR(143)DEG	DEGEVEEQR	unknown protein	63	38	3.16E-05	-0.776	1
AT1G02780	667.851	2	N/A	AREK(154)TLS	TLSDQFEAKR	Ribosomal protein L19e family protein	56	44	6.31E-04	-0.375	1
AT1G03130	1002.858	3	13.34	PAVK(58)EAP	EAPVGFPPQLDNPSPIFAGSTGGLLR	photosystem I subunit D-2	59	44	3.16E-04	-0.965	1
AT1G03475	798.356	2	2.67	TFLR(67)DSD	DSDDVTPSSSSSVR	Coproporphyrinogen III oxidase	40	38	6.31E-03	-0.376	2
AT1G03890	642.338	2	N/A	ETAK(235)QLQ	QLQNQKDNR	RmlC-like cupins superfamily protein	55	44	7.94E-04	0.702	1
AT1G04270	738.889	2	15.26	KAKR(73)EAP	EAPQGEKPEPVR	cytosolic ribosomal protein S15	61	45	2.51E-04	-0.474	1
AT1G04510	921.457	2	N/A	NGKR(155)GID	GIDDGEGQPNACKMR	MOS4-associated complex 3A	69	43	2.51E-05	-0.435	4
AT1G04820	539.252	2	49.19	ISGK(97)EDA	EDAANNFAR	tubulin alpha-4 chain	53	41	6.31E-04	-0.093	3
AT1G04820	791.962	2	N/A	VVPK(327)DVN	DVNAAVGTIKTKR	tubulin alpha-4 chain	58	43	3.16E-04	-0.506	1
AT1G04820	601.808	2	N/A	TIGK(113)EIV	EIVDLCLDR	tubulin alpha-4 chain	49	44	3.16E-03	-0.499	1
AT1G08200	515.762	2	2.75	EPLK(252)LVD	LVDGGESQR	UDP-D-ribose/UDP-D-xylose synthase 2	58	43	3.16E-04	-0.291	1
AT1G08200	860.870	2	N/A	VSSK(329)EFY	EFYGEYDSDDKR	UDP-D-ribose/UDP-D-xylose synthase 2	81	36	3.16E-07	0.233	2
AT1G08200	784.420	2	N/A	LLRR(248)EPL	EPLKLVGGESQR	UDP-D-ribose/UDP-D-xylose synthase 2	53	44	1.26E-03	-0.511	1
AT1G09780	844.397	2	3.09	MHAK(295)ALE	ALEYEDFDKFR	Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	72	41	7.94E-06	-0.771	1
AT1G09780	700.842	2	3.59	VALR(192)ENG	ENGVDQAISGGGR	Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	105	42	5.01E-09	0.071	1
AT1G09795	573.767	2	5.62	ANMR(322)GTD	GTDAEEVAER	ATP phosphoribosyl transferase 2	42	41	7.94E-03	-0.262	1
AT1G12000	861.955	2	8.39	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	62	44	1.58E-04	-0.697	1
AT1G12800	750.384	2	N/A	MRPR(564)ENE	ENEEVEKKR	Nucleic acid-binding, OB-fold-like protein	46	44	6.31E-03	-0.801	1
AT1G13270	650.380	2	15.58	RKMR(78)ELE	ELETKSKVR	methionine aminopeptidase 1B	52	43	1.26E-03	-0.154	1
AT1G14610	717.751	3	N/A	KQAK(94)DGT	DGTNVPKKSARKSSKR	valyl-tRNA synthetase / valine-tRNA ligase (VALRS)	44	41	5.01E-03	-0.419	1
AT1G14810	668.350	2	19.61	AAEK(124)GTI	GTIVVDNSSAFR	semialdehyde dehydrogenase family protein	56	44	6.31E-04	-1.199	1
AT1G14950	804.901	2	40.53	KEKR(74)ELD	ELDDQKMAVTFR	Polyketide cyclase/dehydrase and lipid transport superfamily protein	52	43	1.26E-03	-0.684	1
AT1G15230	558.283	2	8.78	DGLK(122)DLD	DLDVEGIER	unknown protein	50	43	2.00E-03	0	1
AT1G15470	701.844	2	N/A	LETK(183)SPV	SPVTSAEVSDQGR	Transducin/WD40 repeat-like superfamily protein	84	43	7.94E-07	-0.214	1
AT1G16850	1149.135	2	7.88	EAAR(118)SAA	SAAEVVSDTAEAVKEKVKR	unknown protein	51	43	1.58E-03	-0.871	2
AT1G16850	816.884	2	10.67	KVKR(137)SVS	SVSGGVTPSEGSEEL	unknown protein	65	41	3.98E-05	-0.919	1
AT1G17210	851.382	2	N/A	VVDR(514)DGD	DGDEVNDDSAGPSKR	IAP-like protein 1	79	39	1.00E-06	-0.059	1
AT1G18070	879.465	3	4.24	AKEK(80)AAQ	AAQEKAKEEAEVVAEANKKR	Translation elongation factor EF1A	65	44	7.94E-05	-0.417	1
AT1G18070	988.524	3	11.30	AKAK(78)EKA	EKAAQEKAKEEAEVVAEANKKR	Translation elongation factor EF1A	98	44	3.98E-08	-0.540	1
AT1G18070	1020.037	2	N/A	AQEK(85)AAK	AAKEEAEVVAEANKKR	Translation elongation factor EF1A	97	44	5.01E-08	-0.981	1
AT1G18070	849.931	2	N/A	KAAC(88)EAA	EAEAEVVAEANKKR	Translation elongation factor EF1A	70	44	2.51E-05	-0.530	1
AT1G18500	734.443	2	2.73	YKLN(205)KTK	KTKAEVIEIAR	methylthioalkylmalate synthase-like 4	43	41	6.31E-03	-0.545	1
AT1G19140	589.277	2	N/A	RKPR(86)AEF	AEFQEEQAR	unknown protein	54	41	5.01E-04	0.850	1
AT1G20010	584.233	2	N/A	WDSK(299)NMM	NMMCAADPR	tubulin beta-5 chain	41	32	1.26E-03	-0.343	1
AT1G21100	652.828	2	3.10	AQLK(155)DVV	DVVLEGGDAFGR	O-methyltransferase family protein	50	43	2.00E-03	-0.767	1
AT1G21750	579.312	2	5.02	LLPR(204)GES	GESSVTGPVVR	PDI-like 1-1	69	44	3.16E-05	-1.210	1
AT1G24706	803.867	2	N/A	ISKK(129)SME	SMEQKEDTETPR	THO2	41	40	7.94E-03	0	1
AT1G26630	588.804	2	N/A	GASK(18)TYP	TYPQSAGNIR	Eukaryotic translation initiation factor 5A-1 (eIF-5A 1) protein	63	44	1.26E-04	0.170	2
AT1G27430	697.849	2	N/A	TASR(1436)AAA	AAASSYSGIEGGVR	GYF domain-containing protein	47	43	3.98E-03	0	1
AT1G29680	645.840	2	13.35	VSTR(22)LID	LIDAGSTIMQR	Protein of unknown function (DUF1264)	48	44	3.98E-03	-0.155	1

AT1G29850	985.144	3	10.98	LMAR(21)QGM	QGMGKQGNQNPQEQEQEDAKR	double-stranded DNA-binding family protein	67	42	3.16E-05	-0.424	1
AT1G29880	494.259	2	25.20	IRER(686)DSK	DSKDQVR	glycyl-tRNA synthetase / glycine--tRNA ligase	52	44	1.58E-03	0.405	1
AT1G30470	572.801	2	N/A	SSPK(754)ASG	ASGAEVTGNLR	SIT4 phosphatase-associated family protein	83	44	1.26E-06	0.087	1
AT1G30580	906.997	2	9.52	EDLK(354)ELG	ELGNPAVKAAGKYR	GTP binding protein	72	44	1.58E-05	-0.828	1
AT1G30580	822.947	2	12.51	FVGK(165)KID	KIDDVEKSMKR	GTP binding protein	64	44	1.00E-04	0.547	1
AT1G42970	696.344	2	N/A	VEKK(335)GLT	GLTAEDVNEAFR	glyceraldehyde-3-phosphate dehydrogenase B subunit	55	43	6.31E-04	-0.216	1
AT1G45000	782.441	2	N/A	RLDR(304)KIE	KIEIPLPNEQSR	AAA-type ATPase family protein	46	43	5.01E-03	-0.448	1
AT1G47540	764.340	2	97.14	CAPR(49)IFP	IFPTFCDQNCR	trypsin inhibitor 2	49	38	7.94E-04	-0.197	1
AT1G48860	598.356	2	6.18	TAEK(79)ASE	ASEIVLQPIR	RNA 3'-terminal phosphate cyclase/enolpyruvate transferase, alpha/beta	62	42	1.00E-04	-0.251	1
AT1G48860	670.393	3	N/A	QPIR(89)EIS	EISGLIKLPGSKLSNR	RNA 3'-terminal phosphate cyclase/enolpyruvate transferase, alpha/beta	65	41	3.98E-05	-0.398	1
AT1G49240	743.873	2	5.67	MNQK(53)DAY	DAYVGDQAQSKR	actin 8	74	43	7.94E-06	-0.269	4
AT1G49760	584.309	2	3.80	LATR(581)LAN	LANAAPEQQR	poly(A) binding protein 8	50	43	2.00E-03	0.600	1
AT1G49760	772.894	2	9.53	LYVK(333)NLD	NLDESVDKDLR	poly(A) binding protein 8	78	44	3.98E-06	-0.907	1
AT1G49760	861.475	2	N/A	VHHR(565)DSP	DSPTSQPVPIVALATR	poly(A) binding protein 8	55	44	7.94E-04	0.232	1
AT1G50200	1010.508	2	N/A	AEER(485)GLL	GLLVDVDGFNKAMEEAR	Alanyl-tRNA synthetase	102	44	1.58E-08	-1.040	1
AT1G52690	802.400	2	19.14	QSAK(60)DKT	DKTSQAAQTQER	LEA family protein	92	43	1.26E-07	-0.873	4
AT1G52690	645.818	2	34.28	AKDK(62)TSQ	TSQAAQTQER	LEA family protein	45	43	6.31E-03	-0.155	1
AT1G54870	643.359	2	3.64	LRGK(86)VAL	VALITGGDSGIGR	Aldehyde reductase	81	45	2.51E-06	0.156	2
AT1G54870	637.841	2	5.69	NCKR(158)VVD	VVDEVVNAFGR	Aldehyde reductase	66	44	6.31E-05	-0.236	1
AT1G54870	660.835	2	8.76	EKIK(288)NFG	NFGSEVPMKR	Aldehyde reductase	56	44	6.31E-04	-0.455	3
AT1G54870	1058.026	2	N/A	SDSK(141)EPI	EPIAIPDLGFDENCKR	Aldehyde reductase	49	43	2.51E-03	-0.757	1
AT1G55490	1051.553	2	17.30	VLTK(382)ETS	ETSTIVGDGSTQDAVKKR	chaperonin 60 beta	82	44	1.58E-06	-0.524	1
AT1G56070	488.759	2	7.91	EIK(558)SDP	SDPVVSR	Ribosomal protein S5/Elongation factor G/III/V family protein	59	44	3.16E-04	0.103	1
AT1G56340	1078.972	2	N/A	EETK(408)EAE	EAEETKEAEETDAAHDEL	calreticulin 1a	57	37	1.00E-04	-0.557	1
AT1G60080	592.338	2	N/A	ITEK(208)EAV	EAVNKEKR	3'-5'-exoribonuclease family protein	56	44	6.31E-04	-0.338	1
AT1G62380	550.788	2	16.14	VLHR(239)VVT	VVTQEGNR	ACC oxidase 2	45	44	7.94E-03	0.637	1
AT1G65090	658.377	2	N/A	EMAK(140)EST	ESTSLLEKIR	unknown protein	57	43	3.98E-04	-0.152	1
AT1G65090	998.848	3	N/A	TSER(162)TLQ	TLQDDKKSGNAKSEEVQEQPEKR	unknown protein	71	44	2.00E-05	-0.568	1
AT1G66070	705.409	3	N/A	EVPK(77)EAP	EAPKEKPLDPIAEKLR	Translation initiation factor eIF3 subunit	45	42	5.01E-03	-0.379	1
AT1G69250	846.434	2	N/A	ITQK(202)VTE	VTEPDAAPQPDGAKR	Nuclear transport factor 2 (NTF2) family protein	58	44	3.98E-04	-0.828	1
AT1G70410	703.326	2	3.38	AELK(40)ELD	ELDSSNSDAIER	beta carbonic anhydrase 4	77	41	2.51E-06	-0.214	1
AT1G70730	636.827	2	6.65	KTIK(163)EYP	EYPIAEDLPR	Phosphoglucosyltransferase/phosphomannosyltransferase family protein	63	44	1.26E-04	0.079	1
AT1G72370	600.742	2	N/A	LGTK(33)NCN	NCNYQMER	40s ribosomal protein SA	48	32	2.51E-04	-0.083	1
AT1G74320	595.812	2	N/A	GTEK(9)NVE	NVENKQYR	Protein kinase superfamily protein	46	44	6.31E-03	0.084	1
AT1G75190	954.446	2	N/A	QKAK(11)ETS	ETSSVAEASSPTESQATR	unknown protein	119	41	1.58E-10	-0.787	1
AT1G79920	700.855	2	13.34	ARQR(22)GID	GIDVVLNDESNR	Heat shock protein 70 (Hsp 70) family protein	86	44	6.31E-07	-0.286	1
AT1G79920	582.284	2	50.64	LQDR(615)VME	VMEETKDR	Heat shock protein 70 (Hsp 70) family protein	53	43	1.00E-03	0.172	1
AT2G04030	573.275	2	3.62	AACK(693)NAP	NAPESTEATR	Chaperone protein htpG family protein	59	42	2.00E-04	-0.087	1
AT2G04030	1158.530	2	N/A	VAEK(69)ETT	ETTEEGSGEKFEYQAEVSR	Chaperone protein htpG family protein	64	40	3.98E-05	-0.475	2
AT2G04030	920.802	3	N/A	YKEK(583)KfV	KFVDISKEDLELGDEDEVKDR	Chaperone protein htpG family protein	73	44	1.26E-05	-0.671	2
AT2G07698	660.844	2	23.03	EEER(277)AAE	AAELTNLFESR	ATPase, F1 complex, alpha subunit protein	58	44	3.98E-04	-0.379	1
AT2G15480	678.854	2	4.13	KAVR(438)EVI	EVIGGEKAEER	UDP-glucosyl transferase 73B5	62	44	1.58E-04	-0.148	1
AT2G16600	655.845	2	8.92	NVVR(149)DIE	DIEKVGSDSGR	rotamase CYP 3	48	44	3.98E-03	-1.069	5
AT2G20140	833.993	2	N/A	RIDR(353)KIE	KIEFPLPDIKTR	AAA-type ATPase family protein	86	42	3.98E-07	-0.360	1
AT2G21060	785.322	2	53.84	HMAR(150)ECS	ECSQGGGGYSGGGGGR	glycine-rich protein 2B	134	33	7.94E-13	-0.637	1
AT2G21660	788.849	2	3.99	KAMK(64)DAI	DAIEGMNGQDLDR	glycine-rich, RNA binding protein AtGRP7	78	38	1.00E-06	-0.444	1
AT2G21660	587.807	2	5.30	LDGR(78)SIT	SITVNEAQSR	glycine-rich, RNA binding protein AtGRP7	58	44	3.98E-04	-0.170	2
AT2G21660	476.217	2	6.69	YSSR(142)GGG	GGGGSYGGGR	glycine-rich, RNA binding protein AtGRP7	69	40	1.26E-05	0.316	1
AT2G21660	827.096	3	N/A	TDDR(23)ALE	ALETAFQYGDVIDSKIINDR	glycine-rich, RNA binding protein AtGRP7	72	44	1.58E-05	-0.847	2
AT2G22400	947.790	3	5.68	DGLK(442)ETD	ETDISVVDATPEEQAEVSDLPLER	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	52	42	1.00E-03	-0.704	1

AT2G22400	615.804	2	N/A	TEDK(567)EAN	EANSSNAGGKR	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	72	42	1.00E-05	6.100	1
AT2G23240	825.810	2	N/A	CNDR(17)CGC	CGCPSPCGGESCR	Plant EC metallothionein-like protein, family 15	41	25	2.51E-04	-0.061	1
AT2G25670	842.434	2	N/A	KSGK(282)EVD	EVDAAAKAAAEAAAAR	unknown protein	45	44	7.94E-03	-3.862	1
AT2G27530	1034.600	2	10.08	EAVR(11)EAI	EAITTIKKGSEKKR	Ribosomal protein L1p/L10e family	84	41	5.01E-07	-1.016	1
AT2G31610	747.360	2	8.60	VLTR(28)ELA	ELAEDGYSGVEVR	Ribosomal protein S3 family protein	78	42	2.51E-06	-0.737	1
AT2G32240	738.384	2	N/A	FTSR(261)DSE	DSEASSLTEKLR	unknown protein	51	44	2.00E-03	-1.153	1
AT2G32730	1009.553	2	9.04	AVEK(903)KAP	KAPEPEPAFEILVNPAPR	26S proteasome regulatory complex, non-ATPase subcomplex, Rpn2/Psmd1 subunit	60	44	2.51E-04	-0.992	1
AT2G33150	649.286	2	3.00	GVSF(212)QEQ	QEQQAAAVDSH	peroxisomal 3-ketoacyl-CoA thiolase 3	64	38	2.51E-05	0	1
AT2G33150	831.415	3	14.17	VFKK(283)DGT	DGTTTGNSSQVSDGAGAVLLMKR	peroxisomal 3-ketoacyl-CoA thiolase 3	44	43	7.94E-03	0	1
AT2G35410	998.462	3	N/A	VAEK(72)ETS	ETSADEETSQEEKTEETQNSLKR	RNA-binding (RRM/RBD/RNP motifs) family protein	58	40	1.58E-04	-0.602	1
AT2G36060	1119.963	2	4.49	RGEK(27)GIG	GIGDGTVSYGMDDGDDIYMR	MMS ZWEI homologue 3	37	32	3.16E-03	-0.492	1
AT2G36460	814.947	2	N/A	TPGK(25)GIL	GILAADESTGTIGKR	Aldolase superfamily protein	60	44	2.51E-04	-0.184	2
AT2G36640	1127.601	2	16.45	LEGK(320)ELK	ELKEEAGAKAQEASQKTR	embryonic cell protein 63	81	44	2.00E-06	-0.799	2
AT2G36640	579.81	2	N/A	AGAK(329)AQE	AQEASQKTR	embryonic cell protein 63	55	44	7.94E-04	-1.209	1
AT2G36640	835.091	3	N/A	AKGK(289)AVE	AVETKDTAKENMEKAGEVTR	embryonic cell protein 63	63	44	1.26E-04	0.280	1
AT2G36640	952.968	2	N/A	VETK(294)DTA	DTAKENMEKAGEVTR	embryonic cell protein 63	69	43	2.51E-05	-0.893	2
AT2G36640	710.343	2	N/A	DTAK(298)ENM	ENMEKAGEVTR	embryonic cell protein 63	78	43	3.16E-06	-0.423	2
AT2G36640	1152.565	2	N/A	QKTR(338)EST	ESTESGAQKAETKDSAAVR	embryonic cell protein 63	83	43	1.00E-06	-0.868	1
AT2G37220	544.765	2	3.93	NAIK(264)SLD	SLDGADLDGR	RNA-binding (RRM/RBD/RNP motifs) family protein	71	42	1.26E-05	0.184	2
AT2G37620	930.972	2	10.47	SVEK(241)NYE	NYELPDGQVITIGSER	actin 1	49	44	3.16E-03	-1.075	1
AT2G37660	1000.527	2	41.29	GGIR(258)ELL	ELLVKGDELLETETR	NAD(P)-binding Rossmann-fold superfamily protein	92	44	1.58E-07	-1.251	1
AT2G38880	1008.977	2	N/A	ARYR(110)ELE	ELEGDNKSGSGKSGDGSNR	nuclear factor Y, subunit B1	58	41	2.00E-04	-0.099	1
AT2G39460	774.811	3	2.73	KAAK(24)AVK	AVKSGQAFKKDKKIR	ribosomal protein L23AA	36	35	7.94E-03	-0.474	1
AT2G40170	791.862	2	10.05	QTRK(49)EQL	EQLGTEGYQQMGR	Stress induced protein	63	40	5.01E-05	-0.379	1
AT2G40490	727.435	2	57.42	VTER(47)KVS	KVSATSEPLLR	Uroporphyrinogen decarboxylase	72	42	1.00E-05	-0.482	1
AT2G40840	856.945	2	6.63	SSGR(928)SVP	SVPANVSGEDINKSR	disproportionating enzyme 2	52	44	1.58E-03	-0.526	1
AT2G42560	848.468	2	3.87	TLQK(330)AVE	AVEAKDVAAEKAQR	LEA domain-containing protein	85	44	7.94E-07	-0.590	3
AT2G42560	1230.174	2	4.07	VGSK(454)AVD	AVDLTKEKAAVAADTVVGYTAR	LEA domain-containing protein	60	43	2.00E-04	-0.529	2
AT2G42560	717.883	2	4.48	TKEK(462)AAV	AAVAADTVVGYTAR	LEA domain-containing protein	48	44	3.98E-03	-0.488	1
AT2G42560	696.838	2	6.21	AQQK(126)SMD	SMDSIKAAEER	LEA domain-containing protein	45	42	5.01E-03	-0.359	1
AT2G42560	562.327	2	6.66	QTTK(607)NIV	NIVIGDAPVR	LEA domain-containing protein	45	43	6.31E-03	-0.089	1
AT2G42560	571.312	3	7.09	ALQK(290)AVE	AVEAKETAAEKAQR	LEA domain-containing protein	45	44	7.94E-03	0.175	1
AT2G42560	554.749	2	8.35	EARR(565)DVG	DVGEEYGGGR	LEA domain-containing protein	88	39	1.26E-07	-0.181	1
AT2G42560	795.383	2	12.07	ATEK(268)GKE	GKEAGNMTAEQAAR	LEA domain-containing protein	90	42	1.58E-07	-0.378	1
AT2G42560	564.306	2	16.74	VEAK(335)DVA	DVAEKAQR	LEA domain-containing protein	60	44	2.51E-04	0.178	2
AT2G42560	928.478	3	45.28	VAEK(235)GQQ	GQQAKESVGEQAQKAGSATSEKAQR	LEA domain-containing protein	63	44	1.26E-04	-0.503	1
AT2G42560	667.804	2	N/A	EKGK(270)EAG	EAGNMTAEQAAR	LEA domain-containing protein	70	39	7.94E-06	-0.375	1
AT2G42560	622.318	2	N/A	MRER(214)EGK	EGKESAGGVGGR	LEA domain-containing protein	66	44	6.31E-05	-0.241	2
AT2G42560	1101.057	2	N/A	QQAQ(240)ESV	ESVGEQAQKAGSATSEKAQR	LEA domain-containing protein	68	44	3.98E-05	-1.045	4
AT2G42560	1097.505	2	N/A	LSDK(62)GSN	GSNMPVSDEGEGETKMKR	LEA domain-containing protein	95	40	3.16E-08	-0.775	6
AT2G44310	1084.525	2	16.34	STVR(13)SFV	SFVDDEEQFKKSVDER	Calcium-binding EF-hand family protein	45	43	6.31E-03	-1.246	2
AT2G45070	504.743	2	N/A	APQR(10)GSA	GSAATASMR	Sec61-beta subunit protein	56	42	3.98E-04	-0.794	1
AT2G47390	865.384	2	2.72	LFDR(516)SSE	SSEVVYSDPGSTMLR	Prolyl oligopeptidase family protein	40	38	6.31E-03	-0.752	1
AT2G47470	978.039	2	24.63	ALVK(275)ELV	ELVAASEDEKKAVALSR	thioredoxin family protein	102	44	1.58E-08	-0.665	3
AT2G47610	947.544	2	N/A	RRPK(30)QFG	QFGGALPPKKDLSR	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	47	42	3.16E-03	-0.740	1
AT2G47770	676.804	2	16.65	GDDR(15)DAA	DAATTAMAETER	TSPO(outer membrane tryptophan-rich sensory protein)-related	64	39	3.16E-05	-0.740	2
AT3G01570	847.922	2	6.87	QRTK(145)DAG	DAGQTIEDKAHDVR	Oleosin family protein	82	43	1.26E-06	-1.240	2
AT3G02480	827.429	3	2.79	QQMK(49)EKA	EKAQGAADVVKDKTGMNKSH	LEA family protein	75	44	7.94E-06	-0.484	4
AT3G02480	1174.339	4	9.06	MMDK(29)AKD	AKDAAASAQDSLQQTGQQMKEKAQGAADVVKDKTGMNKSH	LEA family protein	45	42	5.01E-03	-1.151	1
AT3G02480	1077.543	2	33.01	MKEK(51)AQG	AQGAADVVKDKTGMNKSH	LEA family protein	59	44	3.16E-04	7.226	6

AT3G02480	1107.045	4	N/A	DKAK(31)DAA	DAAASAQDSLQQTGQQMKEKAQGAADVVKDKTGMNKSH	LEA family protein	43	42	7.94E-03	-0.565	1
AT3G03780	837.441	3	2.77	LAIK(589)DEV	DEVEDLEKGGIGVIQIDEAALR	methionine synthase 2	123	44	1.26E-10	-0.498	3
AT3G03780	741.420	2	N/A	DLEK(597)GGI	GGIGVIQIDEAALR	methionine synthase 2	109	43	2.51E-09	-0.270	2
AT3G05260	652.846	2	3.03	NCKR(113)VVE	VVEEVVNSFGR	NAD(P)-binding Rossmann-fold superfamily protein	56	44	6.31E-04	-0.537	1
AT3G06610	528.788	2	6.90	AASR(54)EAD	EADLNAKR	DNA-binding enhancer protein-related	45	44	7.94E-03	0.474	1
AT3G07230	1006.007	2	9.07	DKPR(32)EQK	EQKPKASDNKPVME	wound-responsive protein-related	60	44	2.51E-04	-0.647	1
AT3G08590	572.292	2	5.56	LRAK(196)GVD	GVDAQVASGGGR	Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	53	43	1.00E-03	-1.050	2
AT3G09440	722.830	2	2.87	VDK(546)NAL	NALENYAYNMR	Heat shock protein 70 (Hsp 70) family protein	60	40	1.00E-04	-0.554	2
AT3G09440	651.348	2	N/A	QATK(164)DAG	DAGVIAGLNVMR	Heat shock protein 70 (Hsp 70) family protein	87	44	5.01E-07	0.077	1
AT3G09840	561.296	2	N/A	KSKK(15)DFS	DFSTAILER	cell division cycle 48	61	44	2.00E-04	-0.178	2
AT3G10650	1079.522	3	5.22	SVSR(495)EVL	EVLAQSEKTGDAVDGTSKTGSSKQDQMR	unknown protein	87	42	3.16E-07	-0.865	1
AT3G11400	454.711	2	N/A	GADR(200)SAV	SAVGSDMR	eukaryotic translation initiation factor 3G1	45	41	3.98E-03	0.551	1
AT3G11710	889.081	3	N/A	QTTK(11)ALS	ALSELAMDSSTLNAEASSAGDGAGPR	lysyl-tRNA synthetase 1	84	40	3.98E-07	-0.976	1
AT3G11930	1000.514	2	N/A	AVEK(18)QPE	QPETTTEAEAPSLTKR	Adenine nucleotide alpha hydrolases-like superfamily protein	70	44	2.51E-05	-0.750	1
AT3G12960	922.462	2	69.88	RIEK(56)GKE	GKEQSAASGDQTIQR	unknown protein	119	44	3.16E-10	-0.760	1
AT3G12960	794.882	2	N/A	EKGK(58)EQS	EQSAASGDQTIQR	unknown protein	74	42	6.31E-06	-0.252	1
AT3G13300	797.339	2	N/A	SRSK(72)DSN	DSNVTDDDDVSGMR	Transducin/WD40 repeat-like superfamily protein	51	35	2.51E-04	-0.063	1
AT3G13470	1053.524	2	4.27	VLT(378)EMT	EMTTIVGDGTTQEAVNKR	TCP-1/cpn60 chaperonin family protein	115	43	6.31E-10	-0.713	1
AT3G13470	635.356	2	15.13	TNAR(282)DLV	DLVGVLEDAIR	TCP-1/cpn60 chaperonin family protein	98	44	3.98E-08	-0.473	3
AT3G15280	910.955	2	31.94	TEYR(135)GVE	GVEDLHQQTGGVEKSP	unknown protein	52	44	1.58E-03	-0.440	3
AT3G15280	967.915	2	33.92	KA(94)ETG	ETGQNISDAVTGDDDDGR	unknown protein	89	36	5.01E-08	-0.362	2
AT3G15670	527.585	3	4.55	AQEK(17)TGQ	AQEKTGQAMGTMR	LEA family protein	44	42	6.31E-03	0.823	4
AT3G15670	612.638	3	7.68	GETR(15)GKA	GKAQEKTGQAMGTMR	LEA family protein	51	43	1.58E-03	0.163	4
AT3G15670	1534.745	2	14.72	IKNK(134)AQD	AQDAAQYTKETAQGAQYTKETAEAGR	LEA family protein	104	42	6.31E-09	-0.848	2
AT3G15670	674.664	3	15.86	QYTK(143)ETA	ETAQGAQYTKETAEAGR	LEA family protein	101	43	1.58E-08	0.198	4
AT3G15670	700.873	2	42.25	QAAK(103)EKT	EKTSQAGDKAR	LEA family protein	71	45	2.51E-05	-0.286	3
AT3G15790	1002.494	2	16.10	VEKK(215)TVE	TVEASDEKKNSEAETR	methyl-CPG-binding domain 11	81	43	1.58E-06	-1.747	1
AT3G16640	567.314	2	31.74	STQK(67)VVD	VVDIVDFR	translationally controlled tumor protein	76	44	6.31E-06	0.177	1
AT3G16830	956.496	3	N/A	EKPR(732)NLD	NLDSVSKSKPLELLEIVDPTQCR	TOPILESS-related 2	52	44	1.58E-03	-0.419	1
AT3G17520	719.844	2	2.76	ESAK(260)DKA	DKASQSYDSAAR	LEA family protein	88	42	2.51E-07	-0.139	2
AT3G17520	913.451	3	3.00	ETAK(249)SKA	SKADETLESKADKASQSYDSAAR	LEA family protein	73	43	1.00E-05	-0.731	4
AT3G17520	706.885	2	5.34	AKEK(222)AEK	AEKLKEEAER	LEA family protein	68	44	3.98E-05	-0.283	1
AT3G17520	655.871	2	6.09	ASQK(76)AED	AEDAKEAAKR	LEA family protein	66	43	5.01E-05	-0.076	2
AT3G17520	699.432	2	11.18	VSEK(182)AKE	AKEAKEAAKR	LEA family protein	47	40	2.00E-03	0.859	1
AT3G17520	1005.854	3	14.74	ETAK(209)EKA	EKASDMTSAAKEKAEKLKEEAER	LEA family protein	47	44	5.01E-03	-1.095	2
AT3G17520	1227.089	2	N/A	AKSK(251)ADE	ADETLKSAKDKASQSYDSAAR	LEA family protein	126	42	3.98E-11	-1.020	2
AT3G17520	1344.688	2	N/A	AKEK(211)ASD	ASDMTSAAKEKAEKLKEEAER	LEA family protein	85	44	7.94E-07	-1.302	1
AT3G17520	563.262	2	N/A	AKDK(262)ASQ	ASQSYDSAAR	LEA family protein	77	41	2.51E-06	0.267	3
AT3G17520	556.819	2	N/A	EKAK(184)EAK	EAKEAAKR	LEA family protein	61	44	2.00E-04	0.270	6
AT3G17520	870.476	2	N/A	SAAK(220)EKA	EKAELKEEAER	LEA family protein	87	44	5.01E-07	-0.748	1
AT3G20050	568.268	2	N/A	IRQR(270)EAD	EADMTKER	T-complex protein 1 alpha subunit	44	41	5.01E-03	0.793	1
AT3G22490	670.862	2	N/A	AAER(85)GVT	GVTVAQTDVPGAR	Seed maturation protein	51	44	2.00E-03	-0.224	1
AT3G22500	689.811	2	N/A	AVTR(217)EDA	EDAEGVVAEMR	Seed maturation protein	40	39	7.94E-03	0	1
AT3G22640	478.282	2	49.52	TRSK(231)EIG	EIGQGHIR	cupin family protein	48	43	3.16E-03	0	2
AT3G23580	1104.476	2	N/A	GQGR(11)DME	DMEEGESEEPPLMAQNQR	ribonucleotide reductase 2A	62	36	2.51E-05	-0.816	1
AT3G23810	556.737	2	2.75	YKVK(18)DMS	DMSQADFGR	S-adenosyl-l-homocysteine (SAH) hydrolase 2	52	37	3.16E-04	0.090	1
AT3G23810	808.449	2	11.12	MKER(199)LVG	LVGVSEETTTGVKR	S-adenosyl-l-homocysteine (SAH) hydrolase 2	69	44	3.16E-05	-0.619	1
AT3G23810	660.824	2	15.26	DVVK(276)GCA	GCAAAMKTAGAR	S-adenosyl-l-homocysteine (SAH) hydrolase 2	59	43	2.51E-04	-0.379	1
AT3G23810	950.462	2	N/A	SSGR(13)EYK	EYKVKDMSQADFGR	S-adenosyl-l-homocysteine (SAH) hydrolase 2	83	43	1.00E-06	-1.001	2
AT3G27660	959.946	2	N/A	HEAR(174)ETE	ETEFMTETHEPGKAR	oleosin 4	44	41	5.01E-03	0	1

AT3G44110	1045.397	2	N/A	QAQR(397)EAY	EAYDDDDDDDDHPGGAQR	DNAJ homologue 3	81	25	2.51E-08	-0.431	1
AT3G45190	748.386	3	8.45	GTER(777)AMD	AMDQALKEGIVGEAGPMKR	SIT4 phosphatase-associated family protein	45	44	7.94E-03	-0.669	1
AT3G48870	950.508	2	3.67	ISDR(501)FLP	FLPDKAIDLIDEAGSR	Clp ATPase	76	44	6.31E-06	-0.895	3
AT3G50820	1045.009	2	14.88	VKFK(189)EED	EEDGIDYAAVTVQLPGGER	photosystem II subunit O-2	72	43	1.26E-05	-0.623	1
AT3G51160	522.272	2	10.65	TAPK(20)ADS	ADSTVVEPR	GDP-D-MANNOSE-4,6-DEHYDRATASE 2, GMD2	54	44	1.00E-03	0.096	1
AT3G51880	591.822	2	16.02	KTTK(15)EAL	EALKPVDDR	high mobility group B1	50	43	2.00E-03	0	1
AT3G51880	810.373	2	N/A	AYNK(124)NLE	NLEEGSDESEKSR	high mobility group B1	59	40	1.26E-04	0.062	1
AT3G52380	617.289	2	22.47	EEEE(107)QTT	QTTQASGEEGR	chloroplast RNA-binding protein 33	70	41	1.26E-05	0.406	1
AT3G52930	1167.060	2	49.30	KVVK(107)GTV	GTVELAGTDGETTQGLDGLGDR	Aldolase superfamily protein	53	42	7.94E-04	-0.900	3
AT3G53040	731.715	3	10.26	AKDR(149)TAD	TADKTKETAETAEKAR	putative / LEA protein	56	44	6.31E-04	-0.228	2
AT3G53040	1075.056	2	12.77	AKDK(127)TAD	TADKTKETADYAAEKAR	putative / LEA protein	70	44	2.51E-05	-1.117	1
AT3G53040	832.433	2	13.47	TADK(131)TKE	TKETADYAAEKAR	putative / LEA protein	62	44	1.58E-04	-2.526	1
AT3G53040	781.381	2	19.86	ESTR(82)EGA	EGADIASEKAAGMR	putative / LEA protein	76	43	5.01E-06	-0.769	3
AT3G53040	989.013	2	23.09	GEYK(246)DYT	DYTAEKAKETADKAR	putative / LEA protein	76	44	6.31E-06	-1.063	3
AT3G53040	690.836	2	23.72	QTKK(111)ETA	ETADYTADKAR	putative / LEA protein	49	43	2.51E-03	-0.217	6
AT3G53040	704.852	2	38.41	DKTK(155)ETA	ETAETAEKAR	putative / LEA protein	54	43	7.94E-04	-0.426	1
AT3G53040	953.441	2	39.57	VETK(326)DTA	DTAKEKMDEAGEEAR	putative / LEA protein	77	40	2.00E-06	0	2
AT3G53040	682.838	2	45.79	DKTK(133)ETA	ETADYAAEKAR	putative / LEA protein	81	43	1.58E-06	0.220	1
AT3G53040	964.170	3	53.92	DKAK(210)EAK	EAKDKTAEKAKETAETSDKAR	putative / LEA protein	62	44	1.58E-04	-0.934	2
AT3G53040	705.841	2	119.00	EKAK(221)ETA	ETAETSDKAR	putative / LEA protein	70	42	1.58E-05	-0.497	1
AT3G53040	887.441	3	N/A	TKQK(321)AVE	AVETKDTAKEKMDEAGEEAR	putative / LEA protein	99	43	2.51E-08	-0.526	2
AT3G53040	1170.638	2	N/A	TETK(286)DAG	DAGVSKIGELKDSAVDTAKR	putative / LEA protein	115	44	7.94E-10	-1.111	1
AT3G53040	863.934	2	N/A	QGTK(67)DAV	DAVIGKSHDTAESTR	putative / LEA protein	81	43	1.58E-06	-0.579	1
AT3G53040	1430.746	2	N/A	DKAR(122)EAK	EAKDKTADKTKETADYAAEKAR	putative / LEA protein	84	44	1.00E-06	-0.804	3
AT3G53040	710.817	2	N/A	DTAK(330)EKM	EKMDEAGEEAR	putative / LEA protein	81	38	5.01E-07	-0.141	2
AT3G53040	1090.062	2	N/A	AKDK(215)TAE	TAEKAKETAETSDKAR	putative / LEA protein	55	44	7.94E-04	-1.515	2
AT3G53230	986.560	2	N/A	GTKK(16)DFS	DFSTAILEKKKAANR	ATPase, AAA-type, CDC48 protein	75	43	6.31E-06	-0.660	2
AT3G55010	842.942	2	11.25	LNYK(70)DSG	DSGVDIDAGAELVKR	phosphoribosylformylglycinamide cyclo-ligase, chloroplast	73	44	1.26E-05	-1.010	1
AT3G56130	951.505	3	2.94	LRAK(59)AAK	AAKSSTTTISDSSDASVSDGKKTVR	biotin/lipoyl attachment domain-containing protein	47	44	5.01E-03	-1.262	1
AT3G58610	710.859	2	5.57	LAYK(345)NTV	NTVECITGTISR	ketol-acid reductoisomerase	55	44	7.94E-04	-0.775	1
AT3G58840	1004.498	2	N/A	RAAK(11)GIS	GISDYDQGGVKTTELER	Tropomyosin-related	86	44	6.31E-07	-0.847	1
AT4G02510	1148.044	2	3.50	VSSR(607)EFS	EFSFGGKEVDQEPSGEGVTR	translocan at the outer envelope membrane of chloroplasts 159	67	42	3.16E-05	-1.700	1
AT4G02930	714.056	3	6.77	GKAK(99)AIA	AIAFDEIDKAPEEKKR	GTP binding Elongation factor Tu family protein	65	44	7.94E-05	-0.280	1
AT4G03200	894.971	2	N/A	ELSK(244)ALS	ALSASTGADKLSDGISR	unknown protein	73	44	1.26E-05	-0.559	1
AT4G05400	740.824	2	N/A	IQKK(122)GGA	GGAASMGVTDEPMR	unknown protein	62	37	3.16E-05	-0.608	2
AT4G08310	994.209	3	N/A	NTSK(159)SVA	SVAEQTKDEDKEVLQSDIKKALR	unknown protein	85	43	6.31E-07	-0.940	1
AT4G10320	705.422	2	N/A	DADK(406)DII	DIIEAVKAKGR	tRNA synthetase class I (I, L, M and V) family protein	51	42	1.26E-03	-0.355	1
AT4G10790	754.418	3	N/A	RQEK(390)ALA	ALALGEEPEKGPVTVQLVLR	UBX domain-containing protein	53	44	1.26E-03	-0.310	1
AT4G11420	964.497	2	N/A	IEEK(628)ELE	ELEEAQALLEETEKR	eukaryotic translation initiation factor 3A	73	44	1.26E-05	-0.882	1
AT4G11600	422.712	2	N/A	LVDK(204)DGN	DGNVVDR	glutathione peroxidase 6	59	41	1.58E-04	-0.237	1
AT4G15410	854.486	2	N/A	APSR(338)GLV	GLVVDPAAPTTSIQLR	serine/threonine protein phosphatase 2A (PUX5)	47	42	3.16E-03	-0.820	1
AT4G17520	601.307	2	2.94	RHSR(146)TAY	TAYGNEDKR	Hyaluronan / mRNA binding family	53	44	1.26E-03	0.500	2
AT4G17520	952.520	2	13.44	QTAK(37)SGK	SGKMPTKPPPSQAVR	Hyaluronan / mRNA binding family	57	44	5.01E-04	-0.788	1
AT4G17520	794.404	3	N/A	RNQR(337)DGG	DGGAAAQAPTPAIGDSAQFPPLGK	Hyaluronan / mRNA binding family	45	44	7.94E-03	-0.168	1
AT4G17720	560.267	2	14.34	ESPK(263)AAS	AASSTQEAER	RNA-binding (RRM/RBD/RNP motifs) family protein	77	42	3.16E-06	0.268	1
AT4G18440	554.769	2	2.85	TSTK(63)VTA	VTAMDGVSSR	L-Aspartase-like family protein	63	43	1.00E-04	-0.090	1
AT4G19210	984.028	2	31.84	HIPR(179)AVK	AVKGNVGEVLQDKDER	RNAse I inhibitor protein 2	84	44	1.00E-06	-2.035	1
AT4G19210	693.847	2	N/A	VIDR(213)DVE	DVENLSGGELQR	RNAse I inhibitor protein 2	63	43	1.00E-04	-0.144	1
AT4G19880	770.350	2	6.26	SMAR(35)SAV	SAVDETSDSGAFQR	Glutathione S-transferase family protein	73	40	5.01E-06	-0.195	1
AT4G21020	660.346	2	2.72	YNAK(123)EKA	EKAKDYAER	LEA family protein	62	44	1.58E-04	-0.228	2

AT4G21020	496.756	2	3.55	AKEK(125)AKD	AKDYAER	LEA family protein	53	44	1.26E-03	0.303	1
AT4G21020	718.364	3	6.55	EKAK(116)DTA	DTAYNAKEKAKDYAER	LEA family protein	58	44	3.98E-04	-0.511	4
AT4G21020	775.724	3	16.97	ESTK(218)NAA	NAAQTVTEAVVGPEEDAERKAR	LEA family protein	59	44	3.16E-04	-0.301	2
AT4G21020	661.639	3	28.31	EKAK(156)DYA	DYAEDTMDNAKEKAR	LEA family protein	43	40	5.01E-03	0.050	6
AT4G21020	1126.542	2	N/A	TKEK(154)AKD	AKDYAEDTMDNAKEKAR	LEA family protein	74	42	6.31E-06	-0.977	2
AT4G23670	836.444	2	N/A	KEKR(74)EID	EIDDENKTLTKR	Polyketide cyclase/dehydrase and lipid transport superfamily protein	71	44	2.00E-05	-0.479	1
AT4G25140	520.256	2	14.91	MSGK(34)GSD	GSDYSKSR	oleosin 1	52	43	1.26E-03	0.289	1
AT4G25140	535.732	2	21.59	GRDR(26)DQY	DQYQMSGR	oleosin 1	49	37	6.31E-04	0.281	1
AT4G25340	722.376	2	N/A	TVEK(354)QTP	QTPADSKSSQVR	FK506 BINDING PROTEIN 53	91	44	2.00E-07	-0.277	2
AT4G26970	750.852	2	18.19	LMDR(801)GVI	GVISEDFNSYGSR	aconitase 2	67	41	2.51E-05	-0.467	1
AT4G29060	661.414	2	4.43	ESIK(518)GIS	GISPALVKQLR	elongation factor Ts family protein	83	38	3.16E-07	-0.303	2
AT4G29060	1089.132	2	N/A	EEAK(747)EAV	EAVASPTTVVSAALVKQLR	elongation factor Ts family protein	76	41	3.16E-06	-1.424	1
AT4G29060	1000.993	2	N/A	AEDK(97)ETV	ETVASEKSDAPAPTSQSR	elongation factor Ts family protein	98	43	3.16E-08	-0.600	2
AT4G29060	657.375	2	N/A	LRKK(560)GLA	GLASADKKASR	elongation factor Ts family protein	75	44	7.94E-06	-0.229	4
AT4G31180	1000.477	2	2.84	QMLK(414)EAG	EAGVEVDPLGDLNTESEK	Class II aminoacyl-tRNA and biotin synthetases superfamily protein	87	42	3.16E-07	-0.500	1
AT4G32330	1000.475	3	N/A	EHVK(64)EAA	EAAEGTQVEHVDDSKCMKGEKAQR	TPX2 (targeting protein for Xklp2) protein family	56	41	3.16E-04	-0.667	1
AT4G32520	1017.490	2	16.61	ISSK(75)EIP	EIPFEDYGLGEVDPEVR	serine hydroxymethyltransferase 3	68	42	2.51E-05	-0.836	1
AT4G34490	687.367	2	N/A	RADR(294)SGA	SGAVSAVEKETR	cyclase associated protein 1	64	44	1.00E-04	-0.146	1
AT4G35090	691.683	3	4.80	KRER(420)CII	CHIEKENFKEPGER	catalase 2	57	44	5.01E-04	-0.483	1
AT4G35570	986.001	2	N/A	DFRK(51)EFN	EFNLNPDNKSVMGNVGR	high mobility group B5	101	44	2.00E-08	-0.761	1
AT4G36700	855.906	2	2.63	ERKK(469)EEE	EEEEAKREEEER	RmlC-like cupins superfamily protein	43	42	7.94E-03	-0.526	1
AT4G36760	703.791	2	N/A	EAAK(270)DME	DMEIDSDQPDR	aminopeptidase P1	68	34	3.98E-06	-0.640	2
AT4G36800	494.246	2	N/A	EKQR(13)EQA	EQAQNATR	RUB1 conjugating enzyme 1	45	42	5.01E-03	1.014	1
AT4G37870	483.303	2	2.88	VMPK(26)ITT	ITTGAAGR	phosphoenolpyruvate carboxykinase 1	49	40	1.26E-03	0.415	2
AT4G37870	506.295	2	5.48	SLTR(102)ESG	ESGPKVVR	phosphoenolpyruvate carboxykinase 1	49	43	2.51E-03	0.594	1
AT4G37870	551.652	3	11.43	SFPK(20)GPV	GPVMPKITTGAAGR	phosphoenolpyruvate carboxykinase 1	49	43	2.51E-03	0.061	3
AT4G37910	1013.026	2	2.73	SPCK(408)GVN	GVNPDEAVAMGAAIQGGILR	mitochondrial heat shock protein 70-1	76	44	6.31E-06	-0.741	1
AT4G37910	660.307	2	3.32	ITIR(547)SSG	SSGGLSDDEINR	mitochondrial heat shock protein 70-1	70	41	1.26E-05	0.076	2
AT4G37910	842.468	2	11.50	VSAK(534)DKA	DKATGKEQNITR	mitochondrial heat shock protein 70-1	67	44	5.01E-05	-0.772	1
AT4G37910	850.466	3	14.53	RMVK(562)EAE	EAEELNAQKDQEKQLIDLK	mitochondrial heat shock protein 70-1	70	43	2.00E-05	-0.706	1
AT4G37910	998.547	3	N/A	EINR(559)MVK	MVKEAELNAQKDQEKQLIDLK	mitochondrial heat shock protein 70-1	49	43	2.51E-03	-0.936	1
AT4G38630	739.342	2	N/A	QKDK(258)DGD	DGDTASASQETVAR	regulatory particle non-ATPase 10	74	41	5.01E-06	-0.203	1
AT4G38740	828.391	2	5.84	AFPK(6)VYF	VYFDMTIDGQPAGR	rotamase CYP 1	49	41	1.58E-03	-0.483	1
AT4G39260	1048.040	2	N/A	DLQR(25)TFS	TFSQFGDVIDSKIINDR	cold, circadian rhythm, and RNA binding 1	126	44	6.31E-11	-0.621	1
AT4G39260	593.825	2	N/A	LDGR(76)VIT	VITVNEAQRS	cold, circadian rhythm, and RNA binding 1	77	45	6.31E-06	-0.169	1
AT4G39680	636.844	2	10.88	DDVR(157)DIA	DIAGLDSSVVAR	SAP domain-containing protein	92	44	1.58E-07	-0.315	1
AT4G39680	542.324	2	N/A	LTTR(35)GLK	GLKEELVR	SAP domain-containing protein	56	43	5.01E-04	0	1
AT5G03860	927.479	2	5.24	SVKR(411)EDA	EDAAATEEDLLQIPR	malate synthase	73	44	1.26E-05	-0.863	1
AT5G04200	773.907	2	N/A	VKKR(11)LAV	LAVLVGCNYPNTR	metacaspase 9	49	44	3.16E-03	-1.164	1
AT5G04200	1073.246	3	N/A	FGFK(48)QDD	QDDIEVLTDPESEKVKPTGANIKAALR	metacaspase 9	71	43	1.58E-05	-0.933	1
AT5G04420	746.824	2	3.28	RRQR(490)SAS	SASDEEEDGTVQR	Galactose oxidase/kelch repeat superfamily protein	64	37	2.00E-05	-0.335	1
AT5G04430	825.429	2	23.28	VIGK(53)GGS	GGSTITEFQAKSGAR	binding to TOMV RNA 1L (long form)	57	44	5.01E-04	-0.788	1
AT5G06760	692.378	2	60.10	LTDR(116)VVE	VVEGTAVTDPGR	Late Embryogenesis Abundant 4-5	93	44	1.26E-07	-0.145	1
AT5G07350	850.926	2	7.16	EYSR(459)KVT	KVTQGDGPTTSGAADR	TUDOR-SN protein 1	79	44	3.16E-06	-0.118	2
AT5G08670	699.357	2	N/A	EGKK(74)TYD	TYDYGGKGAIGR	ATP synthase alpha/beta family protein	68	43	3.16E-05	-0.215	2
AT5G09590	889.436	2	3.16	ITIR(552)SSG	SSGGLSEDDIQKMVR	mitochondrial HSO70 2	80	43	2.00E-06	-0.619	1
AT5G10360	862.958	2	5.07	IVKK(119)GVS	GVSDLPLTDTEKPR	Ribosomal protein S6e	53	44	1.26E-03	-0.870	2
AT5G11260	546.252	2	N/A	TSGR(59)ESG	ESGATGQER	Basic-leucine zipper (bZIP) transcription factor family protein	60	40	1.00E-04	-0.275	1
AT5G12040	908.471	2	2.61	MESK(220)TLT	TLTAGETPTIIVDTDVGR	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase family protein	53	44	1.26E-03	-0.551	1
AT5G15120	741.912	2	5.63	WRRK(47)KID	KIDSPADGITAVR	Protein of unknown function (DUF1637)	53	44	1.26E-03	-0.675	1

AT5G16050	910.458	2	N/A	EAPK(254)EVQ	EVQKVDEQAQPPPSQ	general regulatory factor 5	75	44	7.94E-06	-0.605	1
AT5G19820	573.294	2	22.08	LTGK(1069)DVV	DVVTQETAGR	ARM repeat superfamily protein	50	44	2.51E-03	-0.349	1
AT5G20250	1051.537	2	N/A	SVER(252)DAT	DATVEAGDEKESPIFR	Raffinose synthase family protein	87	44	5.01E-07	-0.809	1
AT5G20890	829.942	2	3.32	LLKR(140)VID	VIDNKDNAEKFR	TCP-1/cpn60 chaperonin family protein	48	44	3.98E-03	-0.784	1
AT5G20890	867.434	2	18.76	PIDK(6)IFK	IFKDDASEEKGER	TCP-1/cpn60 chaperonin family protein	82	44	1.58E-06	-0.692	1
AT5G28500	585.343	2	19.98	ERVK(265)EEE	EEEIKIIR	unknown protein	63	44	1.26E-04	-0.257	1
AT5G35910	868.109	3	N/A	QGKR(801)EAA	EAAAGQKKGSTQESEQGQGR	Polynucleotidyl transferase, ribonuclease H fold protein with HRDC domain	61	44	2.00E-04	-0.615	1
AT5G37190	607.835	2	N/A	VSPR(203)EKL	EKLDDVVTR	COP1-interacting protein 4	53	45	1.58E-03	-0.082	1
AT5G41520	660.323	2	7.28	YGDK(142)AGA	AGAPADYQPGFR	RNA binding Plectin/S10 domain-containing protein	44	43	7.94E-03	-0.152	1
AT5G44310	785.753	3	2.88	ETAK(248)STA	STAQKVTEAVVSGEEDKAR	LEA family protein	57	44	5.01E-04	-0.467	2
AT5G44310	558.288	2	5.53	EEAK(117)DKV	DKVNEGASR	LEA family protein	59	44	3.16E-04	0.269	2
AT5G44310	773.733	3	12.45	VKEK(177)TKN	TKNYAEQTKDKVNEGASR	LEA family protein	66	44	6.31E-05	-0.216	1
AT5G44310	804.430	3	25.92	DKAR(269)DDV	DDVDKGLEDSLKAKENR	LEA family protein	58	44	3.98E-04	-0.768	2
AT5G44310	996.482	2	116.70	EKTK(110)DYA	DYAEAEKDKVNEGASR	LEA family protein	130	43	5.01E-04	0.100	3
AT5G44310	1010.504	2	N/A	ERTK(70)DYA	DYAEQTKNKVNEGASR	LEA family protein	86	44	6.31E-07	-1.040	1
AT5G44310	1010.504	2	N/A	EKTK(179)NYA	NYAEQTKDKVNEGASR	LEA family protein	117	44	5.01E-10	-0.941	3
AT5G44310	879.449	2	N/A	TAQK(253)VTE	VTEAVVSGEEDKAR	LEA family protein	70	44	2.51E-05	0.512	2
AT5G46020	556.819	2	N/A	KAAR(156)DAK	DAKKVEGR	unknown protein	46	43	5.01E-03	0	1
AT5G47210	932.552	3	6.21	KVEK(31)AAA	AAAAVQPPKAAKFPKPPPSQAVR	Hyaluronan / mRNA binding family	46	38	1.58E-03	-0.215	2
AT5G47210	843.423	2	27.17	RYAK(341)EAA	EAAAPAGDQAQFPLSG	Hyaluronan / mRNA binding family	80	43	2.00E-06	-0.475	1
AT5G47210	1121.559	2	N/A	EAKK(212)ELT	ELTAEKAKQEAEEAEAR	Hyaluronan / mRNA binding family	80	43	2.00E-06	-0.803	2
AT5G49910	861.425	2	8.57	TLTR(374)GKF	GKFEELCSDLLDR	chloroplast heat shock protein 70-2	86	43	5.01E-07	-0.755	1
AT5G49910	766.410	2	21.53	AIGK(185)QFA	QFAAEIISQVLR	chloroplast heat shock protein 70-2	47	45	6.31E-03	-0.653	2
AT5G49930	1049.019	2	40.36	GKLL(869)KMK	KMKKEYADQDEDER	zinc knuckle (CCHC-type) family protein	73	43	1.00E-05	-1.097	1
AT5G50960	661.851	2	N/A	TAPK(43)GPD	GPPDLVAIAER	nucleotide binding protein 35	55	44	7.94E-04	-0.227	1
AT5G52300	826.882	2	5.89	DVTR(205)TFA	TFAPGGEDDYLGGQR	CAP160 protein	80	41	1.26E-06	-0.545	1
AT5G52300	685.849	2	48.99	DLKK(306)ESD	ESDINKNSPAR	CAP160 protein	80	43	2.00E-06	-0.292	1
AT5G56000	488.243	2	31.28	TVTR(162)DTS	DTSGEALGR	HEAT SHOCK PROTEIN 81.4	57	43	3.98E-04	-3.694	2
AT5G56030	503.246	2	N/A	TVTR(162)DTS	DTSGETLGR	heat shock protein 81-2	44	42	6.31E-03	0.398	1
AT5G58200	596.776	2	N/A	AKKR(171)GAA	GAAEESGGTTR	Calcineurin-like metallo-phosphoesterase superfamily protein	73	40	5.01E-06	0	1
AT5G60980	612.317	2	16.98	GTTR(152)DVQ	DVQAPIEPER	Nuclear transport factor 2 (NTF2) family protein	45	44	7.94E-03	-0.082	1
AT5G61780	922.974	2	13.30	PKDK(943)QAA	QAALDALEKFQDEAR	TUDOR-SN protein 2	119	44	3.16E-10	-0.705	1
AT5G61780	1079.558	2	N/A	WEPK(941)DKQ	DKQAALDALEKFQDEAR	TUDOR-SN protein 2	51	44	2.00E-03	-2.921	1
AT5G61790	914.453	2	9.81	ESSK(508)SGD	SGDEAEKKEETAAPR	calnexin 1	47	43	3.98E-03	-0.821	1
AT5G62810	774.401	2	N/A	LERK(86)GLT	GLTKEIDEAFR	peroxin 14	45	44	7.94E-03	-0.259	1
AT5G63550	897.522	2	N/A	KETR(244)DVI	DVIADQEKAKKR	DEK domain-containing chromatin associated protein	80	42	1.58E-06	-0.223	2
AT5G63980	730.826	2	3.41	DLLR(170)AID	AIDCGTSEGGPNGR	Inositol monophosphatase family protein	48	38	1.00E-03	-0.206	1
AT5G63980	1075.512	2	N/A	VLEK(114)ELS	ELSSEPFSLVAEEDSGDLR	Inositol monophosphatase family protein	63	42	7.94E-05	-1.489	1
AT5G64250	711.902	2	31.06	VGER(124)DIP	DIPVIAAGGIVDGR	Aldolase-type TIM barrel family protein	81	43	1.58E-06	-0.774	1
AT5G66120	822.428	2	10.75	NDQR(74)SIS	SISSPTVVEVDLGDGR	3-dehydroquinate synthase, putative	73	44	1.26E-05	-0.487	1
AT5G66400	643.677	3	8.57	HEKK(172)GMM	GMMDKIKEKLPGGGR	Dehydrin family protein	58	44	3.98E-04	0.311	1
Proteolysis of Arg or Lys											
AT1G04820	942.843	2	4.97	KDYE(434)EVG	EVGAEGDDEDEDEGEEY	tubulin alpha-4 chain	45	22	5.01E-05	-0.743	1
AT1G09080	600.304	2	4.04	NKHV(73)EII	EIIANDQGNR	Heat shock protein 70 (Hsp 70) family protein	84	43	7.94E-07	0.083	5
AT1G16850	1034.581	2	5.89	RSAA(121)EVV	EVVSDTAEAVKEKVKR	unknown protein	64	43	7.94E-05	-0.919	1
AT1G47540	585.737	2	26.02	RIFP(52)TFC	TFCDQNCR	trypsin inhibitor 2	72	33	1.26E-06	0	1
AT1G50200	530.795	2	4.03	IIQL(901)DVG	DVGLDAAAVR	Alanyl-tRNA synthetase	67	44	5.01E-05	0.094	1
AT1G54870	819.381	2	2.62	AAEQ(180)YES	YESSTIEIDEPR	Aldehyde reductase	44	41	5.01E-03	-0.611	1
AT1G56340	914.411	2	N/A	KEAE(411)ETK	ETKEAETDAAHDEL	calreticulin 1a	65	39	2.51E-05	-0.657	1
AT1G76180	596.77	2	7.19	KVAT(19)EES	EESAEVTDR	Dehydrin family protein	45	39	2.51E-03	0.084	1

AT1G80230	670.403	2	4.14	RLIG(56)SAA	SAAADTAVKKR	Rubredoxin-like superfamily protein	47	42	3.16E-03	-0.075	1
AT2G05710	833.918	2	23.60	SSGE(862)DTI	DTIILAGAAYGSGSSR	aconitase 3	56	44	6.31E-04	-0.540	3
AT2G19750	592.384	2	50.79	RFVT(47)AVV	AVVGFGKKR	Ribosomal protein S30 family protein	62	37	3.16E-05	-0.169	1
AT2G21660	1083.559	2	3.63	RALE(26)TAF	TAFAYQYGDVIDSKIINDR	glycine-rich, RNA binding protein AtGRP7	81	44	2.00E-06	-1.108	1
AT3G02480	948.991	2	13.78	KAQG(54)AAD	AADVVKDKTGMMNKSH	LEA family protein	80	44	2.51E-06	-0.738	3
AT3G04050	588.846	2	3.01	CAVM(79)LDT	LDTKGPEIR	Pyruvate kinase family protein	43	42	7.94E-03	-0.680	1
AT3G05420	595.288	2	8.85	RDIE(515)SEV	SEVEVSQEGR	acyl-CoA binding protein 4	70	42	1.58E-05	-0.084	1
AT3G07720	533.762	2	N/A	WSIQ(66)EAS	EASGDAPPPR	Galactose oxidase/kelch repeat superfamily protein	48	42	2.51E-03	0.188	1
AT3G12960	730.361	2	2.84	KGKE(59)QSA	QSAASGDQTQIQR	unknown protein	108	43	3.16E-09	-0.548	1
AT3G15280	768.389	2	13.21	RGVE(138)DLH	DLHQQTGGEVKSP	unknown protein	58	44	3.98E-04	-0.326	1
AT3G15670	486.759	2	2.60	KEKT(106)SQA	SQAGDKAR	LEA family protein	59	44	3.16E-04	0	1
AT3G27660	654.826	2	3.73	GGYG(35)GYG	GYGAGSDYKSR	oleosin 4	64	43	7.94E-05	-0.229	1
AT4G18480	578.796	2	4.47	INMV(182)DLP	DLPLGATEDR	CHLI subunit of magnesium chelatase, CH-42	48	43	3.16E-03	0	1
AT4G21020	690.352	3	3.71	KNAA(221)QTV	QTVTEAVVGPEEDAER	LEA family protein	50	44	2.51E-03	-0.097	2
AT4G21020	934.441	2	5.49	KAKD(157)YAE	YAEEDTMDNAKEKAR	LEA family protein	74	41	5.01E-06	-0.161	1
AT4G27140	1036.749	4	2.84	GQQQ(90)EQQ	EQQLFQCCNELRQEEPCVPTLKQAAKAVR	seed storage albumin 1	42	40	6.31E-03	-0.459	1
AT4G28440	705.913	2	3.23	ATTG(6)TAA	TAAVATGTSTVKR	Nucleic acid-binding, OB-fold-like protein	62	43	1.26E-04	-0.567	1
AT4G34200	542.804	2	2.71	NLVN(452)ADF	ADFTAKQR	D-3-phosphoglycerate dehydrogenase	46	44	6.31E-03	-0.185	2
AT4G38740	697.325	2	4.26	PKVY(8)FDM	FDMTIDGQPAGR	rotamase CYP 1	55	41	3.98E-04	0	1
AT4G38740	930.937	2	2.80	TGGE(84)SIY	SIYGSKFEDENFER	rotamase CYP 1	50	41	1.26E-03	-0.753	1
AT4G39260	923.982	2	2.66	QRTF(27)SQF	SQFGDVIDSKIINDR	cold, circadian rhythm, and RNA binding 1	99	44	3.16E-08	-0.975	2
AT5G03860	738.373	2	4.71	YRPN(12)VAV	VAVYDSDPGVEVR	malate synthase	52	44	1.58E-03	-0.407	2
AT5G04200	602.834	2	N/A	NDVL(35)AMK	AMKETILSR	metacaspase 9	76	44	6.31E-06	-0.083	1
AT5G04200	1152.646	2	N/A	EVLTI(56)DEP	DEPESKVKPTGANIKAALR	metacaspase 9	78	43	3.16E-06	-1.216	2
AT5G04200	568.287	2	N/A	MDLM(230)DLL	DLLETMTAR	metacaspase 9	52	44	1.58E-03	0.088	1
AT5G04200	688.820	2	N/A	LYCS(314)DQN	DQNADATFLSQP	metacaspase 9	49	41	1.58E-03	-0.363	1
AT5G04200	1054.513	2	N/A	RVLN(281)ENE	ENEGAMKNKQLVMMAR	metacaspase 9	76	43	5.01E-06	-0.142	1
AT5G04200	988.514	2	N/A	NELH(28)GCI	GCINDVLAMKETILSR	metacaspase 9	118	44	3.98E-10	-0.608	1
AT5G04200	886.494	2	N/A	VSSN(169)ISP	ISPAIETTNTITSR	metacaspase 9	48	44	3.98E-03	-0.621	1
AT5G04200	454.231	2	N/A	LMDL(232)LET	LETMTAR	metacaspase 9	51	44	2.00E-03	0.221	1
AT5G04200	524.787	2	N/A	CDFN(121)LIT	LITDVDFR	metacaspase 9	66	44	6.31E-05	0.095	2
AT5G04200	567.316	2	N/A	DVLA(36)MKE	MKETILSR	metacaspase 9	56	44	6.31E-04	-0.177	1
AT5G04200	1111.535	2	N/A	QRVL(280)NEN	NENEGAMKNKQLVMMAR	metacaspase 9	72	42	1.00E-05	-0.855	2
AT5G04200	732.336	2	N/A	CLYC(313)SDQ	SDQNADATFLSQP	metacaspase 9	65	40	3.16E-05	-0.205	2
AT5G04200	987.032	2	N/A	SSVS(167)SNI	SNISPAIETTNTITSR	metacaspase 9	49	44	3.16E-03	-1.116	1
AT5G04200	1203.170	2	N/A	IEVL(55)TDE	TDEPESKVKPTGANIKAALR	metacaspase 9	63	43	1.00E-04	-1.289	1
AT5G04200	681.845	2	N/A	KRLA(13)VLV	VLVGCNYPNTR	metacaspase 9	50	44	2.51E-03	0.073	1
AT5G09590	845.920	2	5.34	TIRS(553)SGG	SGGLEDIQKMVR	mitochondrial HSO70 2	59	43	2.51E-04	-0.828	1
AT5G11520	987.513	2	2.72	FDSA(262)YQG	YQGFASGLDITDAKPIR	aspartate aminotransferase 3	58	44	3.98E-04	-0.405	1
AT5G14780	496.803	2	13.99	VINM(262)PLT	PLTEKTR	formate dehydrogenase	52	41	7.94E-04	0.303	1
AT5G16050	732.373	2	3.81	KEVQ(257)KVD	KVDEQAQPPPSQ	general regulatory factor 5	49	44	3.16E-03	-0.547	1
AT5G27220	527.816	2	5.26	DVNA(247)EKK	EKKNLGR	unknown protein	51	43	1.58E-03	0.190	2
AT5G44310	1042.548	2	4.40	KSTA(251)QKV	QKVTEAVVGSGEADKAR	LEA family protein	128	44	3.98E-11	-0.816	1
AT5G44310	939.474	2	3.34	KTKD(111)YAE	YAEAEKDKVNEGASR	LEA family protein	68	44	3.98E-05	-4.558	1
AT5G46020	598.817	2	N/A	KGAE(70)AVI	AVIEVDNPNR	unknown protein	54	44	1.00E-03	0.167	1
AT5G58290	942.931	2	N/A	KASP(18)ALM	ALMDLSTADEEDLYGR	regulatory particle triple-A ATPase 3	42	40	6.31E-03	-0.902	1
AT5G61780	787.907	2	10.60	KQAA(946)LDA	LDALFKFQDEAR	TUDOR-SN protein 2	51	44	2.00E-03	-0.572	1
AT5G67490	1015.501	2	2.59	SRFL(29)SSG	SSGTPPPQAPSPNQLNR	unknown protein	60	43	2.00E-04	-0.345	1
ATCG00480	575.774	2	3.14	PTLS(304)TEM	TEMGTLQER	ATP synthase subunit beta	44	42	6.31E-03	0.174	1

REFERENCES

- Abramoff, M.D., Magalhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International* 11, 36-42.
- Ambit, A., Fasel, N., Coombs, G.H., and Mottram, J.C. (2008). An essential role for the *Leishmania* major metacaspase in cell cycle progression. *Cell Death Differ* 15, 113-122.
- Aravind, L., Dixit, V.M., and Koonin, E.V. (1999). The domains of death: evolution of the apoptosis machinery. *Trends Biochem Sci* 24, 47-53.
- Aravind, L., and Koonin, E.V. (2002). Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* 46, 355-367.
- Avcı, U., Petzold, H.E., Ismail, I.O., Beers, E.P., and Haigler, C.H. (2008). Cysteine proteases XCP1 and XCP2 aid microautolysis within the intact central vacuole during xylogenesis in *Arabidopsis* roots. *Plant J* 56, 303-315.
- Bailey, K.J., Gray, J.E., Walker, R.P., and Leegood, R.C. (2007). Coordinate regulation of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxykinase by light and CO₂ during C₄ photosynthesis. *Plant Physiol* 144, 479-486.
- Belenghi, B., Romero-Puertas, M.C., Vercammen, D., Brackener, A., Inze, D., Delledonne, M., and Van Breusegem, F. (2007). Metacaspase activity of *Arabidopsis thaliana* is regulated by S-nitrosylation of a critical cysteine residue. *J Biol Chem* 282, 1352-1358.
- Bentsink, L., Jowett, J., Hanhart, C.J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proc Natl Acad Sci U S A* 103, 17042-17047.
- Bonneau, L., Ge, Y., Drury, G.E., and Gallois, P. (2008). What happened to plant caspases? *J Exp Bot* 59, 491-499.
- Boucher, V., Buitink, J., Lin, X., Boudet, J., Hoekstra, F.A., Hundertmark, M., Renard, D., and Leprince, O. (2010). MtPM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. *Plant Cell Environ* 33, 418-430.
- Bowman, J. (1994). *Arabidopsis: An atlas of morphology and development*. New York: Springer.
- Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., and Grolach, J. (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499-1510.
- Bozhkov, P.V., Suarez, M.F., Filonova, L.H., Daniel, G., Zamyatnin, A.A., Jr., Rodriguez-Nieto, S., Zhivotovskiy, B., and Smertenko, A. (2005). Cysteine protease mcII-Pa executes programmed cell death during plant embryogenesis. *Proc Natl Acad Sci U S A* 102, 14463-14468.
- Chahomchuen, T., Akiyama, K., Sekito, T., Sugimoto, N., Okabe, M., Nishimoto, S., Sugahara, T., and Kakinuma, Y. (2009). Tributyltin induces Yca1p-dependent cell death of yeast *Saccharomyces cerevisiae*. *J*

Table S4. Oligonucleotides used for the quantitative RT-PCR analysis of metacaspase transcripts in *Arabidopsis* seeds.

Gene	Oligo	Sequence
AtMC1 AT1G02170	Forward	tctatgcagaatgaacagac
	Reverse	catcacatccactaattgaaatgg
AtMC2 AT4G25110	Forward	ggcgctctctctgat
	Reverse	ttagcgctcaactgtggttc
AtMC3 AT5G64240	Forward	aatccagtggttacactctgtg
	Reverse	ttcaccgctttatgaagcta
AtMC4 AT1G79340	Forward	tcttgcagatgaatcga
	Reverse	agtctgaaatgattcattctgc
AtMC5 AT1G79330	Forward	cactgagctgattgatactgacg
	Reverse	agattcaacaatgcccttcg
AtMC6 AT1G79320	Forward	tcacagtgccggtctcatag
	Reverse	cttctctctgctctctctct
AtMC7 AT1G79310	Forward	agagtcaccggaaca
	Reverse	tgttccactaaatcctgaaatc
AtMC8 AT1G16420	Forward	agtggacatggcagagaa
	Reverse	caccattctctaaattgctggt
AtMC9 AT5G04200	Forward	acagatcgaccctcttcg
	Reverse	tctgttcgttctcaatagcc
UBC AT1G14400	Forward	gtataactactccatccagtc
	Reverse	gctgtacatccgagcagctt

Toxicol Sci 34, 541-545.

- Chakrabortee, S., Boschetti, C., Walton, L.J., Sarkar, S., Rubinsztein, D.C., and Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc Natl Acad Sci U S A* 104, 18073-18078.
- Chatelain, E., Hundertmark, M., Leprince, O., Gall, S.L., Satour, P., Deligny-Penninck, S., Rogniaux, H., and Buitink, J. (2012). Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ*.
- Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 6, 786-787.
- Coll, N.S., Epple, P., and Dangl, J.L. (2011). Programmed cell death in the plant immune system. *Cell Death Differ* 18, 1247-1256.
- Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). *Arabidopsis* type I metacaspases control cell death. *Science* 330, 1393-1397.
- Cumming, G., Fidler, F., and Vaux, D.L. (2007). Error bars in experimental biology. *J Cell Biol* 177, 7-11.
- Delgado-Alvarado, A., Walker, R.P., and Leegood, R.C. (2007). Phosphoenolpyruvate carboxykinase in developing pea seeds is associated with tissues involved in solute transport and is nitrogen-responsive. *Plant Cell Environ* 30, 225-235.
- Dittrich, P., Campbell, W.H., and Black, C.C. (1973). Phosphoenolpyruvate carboxykinase in plants exhibiting crassulacean Acid metabolism. *Plant Physiol* 52, 357-361.
- Edwards, G.E., Kanai, R., and Black, C.C. (1971). Phosphoenolpyruvate carboxykinase in leaves of certain plants which fix CO₂ by the C₄-dicarboxylic acid cycle of photosynthesis. *Biochem Biophys Res Commun* 45, 278-285.
- Fuentes-Prior, P., and Salvesen, G.S. (2004). The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem J* 384, 201-232.
- Galau, G.A., Bijaisoradat, N., and Hughes, D.W. (1987). Accumulation kinetics of cotton late embryogenesis-abundant mRNAs and storage protein mRNAs: coordinate regulation during embryogenesis and the role of abscisic acid. *Dev Biol* 123, 198-212.
- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 21, 566-569.
- Gonzalez, I.J., Desponds, C., Schaff, C., Mottram, J.C., and Fasel, N. (2007). *Leishmania* major metacaspase can replace yeast metacaspase in programmed cell death and has arginine-specific cysteine peptidase activity. *Int J Parasitol* 37, 161-172.
- Hamann, A., Brust, D., and Osiewacz, H.D. (2007). Deletion of putative apoptosis factors leads to lifespan extension in the fungal ageing model *Podospira anserina*. *Mol Microbiol* 65, 948-958.
- Hara-Nishimura, I., and Hatsugai, N. (2011). The role of vacuole in plant cell death. *Cell Death Differ* 18, 1298-1304.
- He, R., Drury, G.E., Rotari, V.I., Gordon, A., Willer, M., Farzaneh, T., Woltering, E.J., and Gallois, P. (2008).

- Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H₂O₂ in *Arabidopsis*. *J Biol Chem* 283, 774-783.
- Helms, M.J., Ambit, A., Appleton, P., Tetley, L., Coombs, G.H., and Mottram, J.C. (2006). Bloodstream form *Trypanosoma brucei* depend upon multiple metacaspases associated with RAB11-positive endosomes. *J Cell Sci* 119, 1105-1117.
- Helsens, K., Colaert, N., Barsnes, H., Muth, T., Flikka, K., Staes, A., Timmerman, E., Wortelkamp, S., Sickmann, A., Vandekerckhove, J., et al. (2010). ms_lim, a simple yet powerful open source laboratory information management system for MS-driven proteomics. *Proteomics* 10, 1261-1264.
- Herker, E., Jungwirth, H., Lehmann, K.A., Maldener, C., Frohlich, K.U., Wissing, S., Buttner, S., Fehr, M., Sigrist, S., and Madeo, F. (2004). Chronological aging leads to apoptosis in yeast. *J Cell Biol* 164, 501-507.
- Ivanovska, I., and Hardwick, J.M. (2005). Viruses activate a genetically conserved cell death pathway in a unicellular organism. *J Cell Biol* 170, 391-399.
- Kall, L., Storey, J.D., MacCoss, M.J., and Noble, W.S. (2008). Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J Proteome Res* 7, 29-34.
- Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. *Plant Physiol* 145, 1144-1154.
- Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R.K., Starr, A.E., Foster, L.J., Kizhakkedathu, J.N., and Overall, C.M. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* 28, 281-288.
- Kumar, S. (2007). Caspase function in programmed cell death. *Cell Death Differ* 14, 32-43.
- Lamkanfi, M., Festjens, N., Declercq, W., Vanden Berghe, T., and Vandenabeele, P. (2007). Caspases in cell survival, proliferation and differentiation. *Cell Death Differ* 14, 44-55.
- Laverriere, M., Cazzulo, J.J., and Alvarez, V.E. (2012). Antagonic activities of *Trypanosoma cruzi* metacaspases affect the balance between cell proliferation, death and differentiation. *Cell Death Differ*.
- Lee, R.E., Brunette, S., Puente, L.G., and Megeney, L.A. (2010). Metacaspase Yca1 is required for clearance of insoluble protein aggregates. *Proc Natl Acad Sci U S A* 107, 13348-13353.
- Lee, R.E., Puente, L.G., Kaern, M., and Megeney, L.A. (2008). A non-death role of the yeast metacaspase: Yca1p alters cell cycle dynamics. *PLoS One* 3, e2956.
- Leegood, R.C., and Walker, R.P. (2003). Regulation and roles of phosphoenolpyruvate carboxykinase in plants. *Arch Biochem Biophys* 414, 204-210.
- Luthi, A.U., and Martin, S.J. (2007). The CASBAH: a searchable database of caspase substrates. *Cell Death Differ* 14, 641-650.
- Madeo, F., Herker, E., Maldener, C., Wissing, S., Lachelt, S., Herlan, M., Fehr, M., Lauber, K., Sigrist, S.J., Wesselborg, S., et al. (2002). A caspase-related protease regulates apoptosis in yeast. *Molecular cell* 9, 911-917.
- Malone, S., Chen, Z.H., Bahrami, A.R., Walker, R.P., Gray, J.E., and Leegood, R.C. (2007). Phosphoenolpyruvate carboxykinase in *Arabidopsis*: changes in gene expression, protein and activity during vegetative and reproductive development. *Plant Cell Physiol* 48, 441-450.
- Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., and Apweiler, R. (2005). PRIDE: the proteomics identifications database. *Proteomics* 5, 3537-3545.
- Martin, M., Rius, S.P., and Podesta, F.E. (2011). Two phosphoenolpyruvate carboxykinases coexist in the Crassulacean Acid Metabolism plant *Ananas comosus*. Isolation and characterization of the smaller 65 kDa form. *Plant Physiol Biochem* 49, 646-653.
- McLuskey, K., Rudolf, J., Proto, W.R., Isaacs, N.W., Coombs, G.H., Moss, C.X., and Mottram, J.C. (2012). Crystal structure of a *Trypanosoma brucei* metacaspase. *Proc Natl Acad Sci U S A* 109, 7469-7474.
- Meslin, B., Beavogui, A.H., Fasel, N., and Picot, S. (2011). *Plasmodium falciparum* metacaspase PfMCA-1 triggers a z-VAD-fmk inhibitable protease to promote cell death. *PLoS One* 6, e23867.
- Moss, C.X., Westrop, G.D., Juliano, L., Coombs, G.H., and Mottram, J.C. (2007). Metacaspase 2 of *Trypanosoma brucei* is a calcium-dependent cysteine peptidase active without processing. *FEBS Lett* 581, 5635-5639.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* 104, 34-41.
- Ojha, M., Cattaneo, A., Hugh, S., Pawlowski, J., and Cox, J.A. (2010). Structure, expression and function of *Allomyces arbuscula* CDP II (metacaspase) gene. *Gene* 457, 25-34.
- Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R., and Piston, D.W. (1997). Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J* 73, 2782-2790.
- Penfield, S., Rylott, E.L., Gilday, A.D., Graham, S., Larson, T.R., and Graham, I.A. (2004). Reserve mobilization in the *Arabidopsis* endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. *Plant Cell* 16, 2705-2718.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.
- Pesquet, E. (2012). Plant proteases - from detection to function. *Physiol Plant* 145, 1-4.
- Proto, W.R., Castanys-Munoz, E., Black, A., Tetley, L., Moss, C.X., Juliano, L., Coombs, G.H., and Mottram, J.C. (2011). *Trypanosoma brucei* metacaspase 4 is a pseudopeptidase and a virulence factor. *J Biol Chem* 286, 39914-39925.
- Rawlings, N.D., and Barrett, A.J. (1993). Evolutionary families of peptidases. *Biochem J* 290 (Pt 1), 205-218.
- Richie, D.L., Miley, M.D., Bhabhra, R., Robson, G.D., Rhodes, J.C., and Askew, D.S. (2007). The *Aspergillus fumigatus* metacaspases CasA and CasB facilitate growth under conditions of endoplasmic reticulum stress. *Mol Microbiol* 63, 591-604.
- Rylott, E.L., Gilday, A.D., and Graham, I.A. (2003). The gluconeogenic enzyme phosphoenolpyruvate carboxykinase in *Arabidopsis* is essential for seedling establishment. *Plant Physiol* 131, 1834-1842.
- Schilling, O., Barre, O., Huesgen, P.F., and Overall, C.M. (2010). Proteome-wide analysis of protein carboxy termini: C terminomics. *Nature methods* 7, 508-511.
- Schilling, O., and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26, 685-694.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37, 501-506.
- Silva, A., Almeida, B., Sampaio-Marques, B., Reis, M.I., Ohlmeier, S., Rodrigues, F., Vale, A., and Ludovico,

- P. (2011). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a specific substrate of yeast metacaspase. *Biochim Biophys Acta* 1813, 2044-2049.
- Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* 1, 2019-2025.
- Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* 6, 1130-1141.
- Suarez, M.F., Filonova, L.H., Smertenko, A., Savenkov, E.I., Clapham, D.H., von Arnold, S., Zhivotovsky, B., and Bozhkov, P.V. (2004). Metacaspase-dependent programmed cell death is essential for plant embryogenesis. *Curr Biol* 14, 339-340.
- Sundstrom, J.F., Vaculova, A., Smertenko, A.P., Savenkov, E.I., Golovko, A., Minina, E., Tiwari, B.S., Rodriguez-Nieto, S., Zamyatnin, A.A., Jr., Valineva, T., et al. (2009). Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome. *Nat Cell Biol* 11, 1347-1354.
- Timmer, J.C., and Salvesen, G.S. (2007). Caspase substrates. *Cell Death Differ* 14, 66-72.
- Tsiatsiani, L., Gevaert, K., and Van Breusegem, F. (2012). Natural substrates of plant proteases: how can protease degradomics extend our knowledge? *Physiol Plant* 145, 28-40.
- Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavalov, A., Lam, E., and Bozhkov, P.V. (2011). Metacaspases. *Cell Death Differ* 18, 1279-1288.
- Turk, B., Turk du, S.A., and Turk, V. (2012). Protease signalling: the cutting edge. *Embo J* 31, 1630-1643.
- Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 6, 961-967.
- Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005). Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nature methods* 2, 771-777.
- Van Damme, P., Staes, A., Bronsoms, S., Helsens, K., Colaert, N., Timmerman, E., Aviles, F.X., Vandekerckhove, J., and Gevaert, K. (2010). Complementary positional proteomics for screening substrates of endo- and exoproteases. *Nature methods* 7, 512-515.
- van der Hoorn, R.A. (2008). Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol* 59, 191-223.
- van Doorn, W.G., Beers, E.P., Dangel, J.L., Franklin-Tong, V.E., Gallois, P., Hara-Nishimura, I., Jones, A.M., Kawai-Yamada, M., Lam, E., Mundy, J., et al. (2011). Morphological classification of plant cell deaths. *Cell Death Differ*.
- Vartapetian, A.B., Tuzhikov, A.I., Chichkova, N.V., Talianky, M., and Wolpert, T.J. (2011). A plant alternative to animal caspases: subtilisin-like proteases. *Cell Death Differ* 18, 1289-1297.
- Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.A., De Rycke, R., Brackenier, A., Inze, D., Harris, J.L., and Van Breusegem, F. (2006). Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* 364, 625-636.
- Vercammen, D., Declercq, W., Vandenabeele, P., and Van Breusegem, F. (2007). Are metacaspases caspases? *J Cell Biol* 179, 375-380.
- Vercammen, D., van de Cotte, B., De Jaeger, G., Eeckhout, D., Casteels, P., Vandepoele, K., Vandenberghe, I., Van Beeumen, J., Inze, D., and Van Breusegem, F. (2004). Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. *J Biol Chem* 279, 45329-45336.
- Walker, R.P., Chen, Z.H., Acheson, R.M., and Leegood, R.C. (2002). Effects of phosphorylation on phosphoenolpyruvate carboxykinase from the C4 plant Guinea grass. *Plant Physiol* 128, 165-172.
- Walker, R.P., and Leegood, R.C. (1995). Purification, and phosphorylation in vivo and in vitro, of phosphoenolpyruvate carboxykinase from cucumber cotyledons. *FEBS Lett* 362, 70-74.
- Walker, R.P., Acheson, R.M., Técsi, L.I., and Leegood, R.C. (1997). Phosphoenolpyruvate Carboxykinase in C4 Plants: Its Role and Regulation. *Australian Journal of Plant Physiology* 24, 459-468.
- Walker, R.P., Trevanion, S.J., and Leegood, R.C. (1995). Phosphoenolpyruvate carboxykinase from higher plants: Purification from cucumber and evidence of rapid proteolytic cleavage in extracts from a range of plant tissues. *Planta* 196, 58-63.
- Ward, W.W. (1981). Properties of the coelenterate green-fluorescent proteins. In: *Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical Applications*. DeLuca, MA and McElroy, WD, eds, Academic Press, New York, 235-242.
- Watanabe, N., and Lam, E. (2005). Two *Arabidopsis* metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. *J Biol Chem* 280, 14691-14699.
- Watanabe, N., and Lam, E. (2011a). *Arabidopsis* Metacaspase 2d Is a Positive Mediator of Cell Death Induced during Biotic and Abiotic Stresses. *Plant J*.
- Watanabe, N., and Lam, E. (2011b). Calcium-dependent activation and autolysis of *Arabidopsis* metacaspase 2d. *J Biol Chem* 286, 10027-10040.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2, e718.
- Zalila, H., Gonzalez, I.J., El-Fadili, A.K., Delgado, M.B., Desponds, C., Schaff, C., and Fasel, N. (2011). Processing of metacaspase into a cytoplasmic catalytic domain mediating cell death in *Leishmania major*. *Mol Microbiol* 79, 222-239.

Chapter 4

***Arabidopsis thaliana* METACASPASES are involved in cell separation during lateral root development**

Liana Tsiatsiani^{a,b,c,d,#}, Dominique Audenaert^{a,b,#}, Dominique Vercammen^{e,#}, Silke Jacques^{a,b,c,d}, Long Nguyen^{a,b}, Evy Timmerman^{c,d}, Tine Beunens^{a,b}, Kris Gevaert^{c,d}, Tom Beeckman^{a,b} and Frank Van Breusegem^{a,b,*}

^aDepartment of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium

^bDepartment of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

^cDepartment of Medical Protein Research, VIB, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^dDepartment of Biochemistry, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^edeVGen, Technologiepark 30, B-9052 Ghent, Belgium

equally contributed to this study

Manuscript in preparation for submission

Correspondence*

e-mail: frank.vanbreusegem@psb.vib-ugent.be

AUTHOR CONTRIBUTIONS

LT, DA and DV have contributed equally to this work. DV, FVB, TB1 and KG conceived and supervised the project. DV and TB2 performed the chemical compounds screens, tissue expression analysis and TB2 produced rAtMC5. LT, DA and LN did the phenotypic analyses. LT generated transgenic lines and performed the proteomic studies with the contribution of SJ. ET performed the MS/MS analysis. LT, DA and DV wrote the paper with the help of FVB, KG and TB1.

AIM AND CONTEXT

Metacaspases are divided into two types based on the presence (type-I) or absence (type-II) of an N-terminal prodomain and the length of an interlinker region. Among the metacaspase-containing organisms (fungi, protozoa and plants), only plants encode type-II metacaspases (Vercammen et al., 2007). Most plant genomes encode a larger number of type-I than type-II metacaspases (Tsiatsiani et al., 2011). *Arabidopsis thaliana* is an exception as it encodes three type-I (*AtMC1* to *AtMC3*) and six type-II (*AtMC4* to *AtMC9*) metacaspases. Current knowledge about redundant or antagonistic function concerns only type-I metacaspases (Helms et al., 2006, Hamann et al., 2007, Richie et al., 2007, Coll et al., 2010) whereas not much is known about it for the type-II metacaspases.

Gene expression analysis of the metacaspase protease family in roots showed that two type-II metacaspases, *AtMC5* and *AtMC9*, could be involved in lateral root protrusion, a process that requires cell separation. Cell separation takes place in most plant tissues during development and it requires appropriate cell wall modification. Despite the common expression patterns of *AtMC5* and *AtMC9* in endodermal cells, which overlie developing early-stage lateral root primordia, no phenotypic aberration could be observed in the *atmc5*, *atmc9* and the double *atmc5mc9* T-DNA lines (KO). To overcome the suggested redundancy in metacaspase functionalities, we decided to identify small-molecule chemical inhibitors able to target multiple metacaspases, and due to their small size, penetrate cell membranes. The identified metacaspase inhibitors arrested lateral root development and cell detachment from the root tip, two cell separation processes. Using type-II metacaspase RNAi lines, we could reproduce the lateral root phenotype and validate that type-II metacaspases mediate cell separation to facilitate lateral root protrusion through the endodermis. In order to explain better the mode of metacaspase action in this process, we decided to identify their potential substrates. Since *AtMC9* is the highest expressed metacaspase in the endodermis during lateral root development we chose to perform a proteome-wide analysis of *AtMC9*-induced protein processing in plant root tissues. Our analysis revealed a set of potential *AtMC9* substrates whose cleavage might affect cell wall structure, and concomitant promotion of cell separation.

Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. *Science* 330, 1393-1397.

Hamann, A., Brust, D., and Osiewacz, H.D. (2007). Deletion of putative apoptosis factors leads to lifespan extension in the fungal ageing model *Podospora anserina*. *Mol Microbiol* 65, 948-958.

Helms, M.J., Ambit, A., Appleton, P., Tetley, L., Coombs, G.H., and Mottram, J.C. (2006). Bloodstream form *Trypanosoma brucei* depend upon multiple metacaspases associated with RAB11-positive endosomes. *J Cell Sci* 119, 1105-1117.

Richie, D.L., Miley, M.D., Bhabhra, R., Robson, G.D., Rhodes, J.C., and Askew, D.S. (2007). The *Aspergillus fumigatus* metacaspases CasA and CasB facilitate growth under conditions of endoplasmic reticulum stress. *Mol Microbiol* 63, 591-604.

Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavialov, A., Lam, E., and Bozhkov, P.V. (2011). Metacaspases. *Cell death and differentiation* 18, 1279-1288.

Vercammen, D., Declercq, W., Vandenabeele, P., and Van Breusegem, F. (2007). Are metacaspases caspases? *J Cell Biol* 179, 375-380.

ABSTRACT

Cell separation is an important process that takes place in all developmental stages in the life cycle of a plant. Fine-tuned control of cell separation-controlled developmental processes such as fruit ripening and pod shattering are of high economical interest for agriculture (Roberts et al., 2002). Modification of cell walls is critical before for cell separation will take place. Metacaspases are cysteine proteases which so far have been associated with programmed cell death. Driven by the expression of *Arabidopsis thaliana* metacaspase 5 and 9 (AtMC5 and AtMC9) in root tissues where cell separation occurs, such as sites of lateral root protrusion and root cap cell release, we sought for and identified a novel metacaspase function related to cell separation. Perturbation of metacaspase activity by genetic and chemical means revealed that these proteases are important for lateral root emergence. Further, proteome-wide identification of AtMC9 substrates in roots involved, among others, cell wall proteins such as glycosyl hydrolases and peroxidases and components of the cell cytoskeleton, thus reinforcing a possible role for metacaspases in cell wall remodelling and cell shape maintenance.

Key words: *Arabidopsis*, cytoskeleton, cell separation, cell wall, DIVERSet, lateral root, metacaspase, N-terminal COFRADIC, programmed cell death, protease, proteomics, substrate

INTRODUCTION

Compared to animal cells, plant cells have a rigid cell wall at the outer side of their plasma membrane. The cell wall, surrounding plant cells, is indispensable for proper functioning of plant cells. Cell wall forms the initial barrier against pathogen infection and abiotic stresses, and in addition provides rigidity and strength to resist the positive turgor inside plant cells. These functionalities are achieved by a wide variety of polysaccharides including cellulose, hemicellulose and pectin, together with cell wall proteins such as proteases, polysaccharide hydrolytic enzymes and lipases, which together are the main constituents of the plant cell wall (Cosgrove 2005; Jamet et al., 2008). Sugar components form a fibreglass structure of cellulose microfibrils embedded in a network of hemicellulose and pectin (Carpita and Gibeau, 1993). The cell wall proteins are associated in this network, either through weak, non-covalent interactions or strong, covalent linkages (Jamet et al., 2008). A consequence of the rigid nature of the cell wall is that it imposes significant restrictions on the modification of cell shape and cell size.

Few examples exist of proteases with cell wall degrading potential. In *Giardia lamblia*, the protozoan parasite causing giardiasis, a protease related to cathepsin B is required for hatching out of the cyst wall, composed of a carbohydrate/peptide complex (Ward et al., 1997). In the unicellular green alga *Chlamydomonas reinhardtii*, cell wall degrading proteases able to cleave proline and hydroxyproline-rich substrates were identified (Jaenicke et al., 1987). In plants, group I grass pollen allergens, β -expansins, were previously reported to be related to cathepsin B and have protease activity (Grobe et al., 1999), but as it was later shown expansin-mediated cell wall loosening is not a proteolytic or lytic event (Li and Cosgrove 2001; Tabuchi et al., 2011).

Protease involvement in cell wall remodelling was shown to occur upstream of lytic enzymes that directly act on the cell wall saccharides. Using tobacco (*Nicotiana benthamiana*) transient leaf expression assays, it was shown that the subtilisin-like protease AtS1P is involved in the processing of pectin modifying enzymes, pectin methylesterases (PMEs), in the Golgi apparatus prior to delivery and activation of PMEs at the cell wall, thus regulating PME activity and consequently cell wall properties (Wolf et al., 2009). CysEP, a papain-type cysteine protease from castor bean (*Ricinus communis*), cleaved *in vitro* the tobacco (*Nicotianum tabacum*) cell wall protein P1-type extensin and, given the distinct tissue-specific expression of the *Arabidopsis* homologs *AtCEP1*, *AtCEP2* and *AtCEP3* in tissues undergoing PCD, it is believed that AtCEPs are involved in tissue collapse or remodelling (Helm et al., 2008; Hierl et al., 2012).

Cell wall remodelling is part of the morphological changes that take place during cell separation. Cell separation occurs at multiple locations in plants such as the root tip and the sites of lateral root protrusion (Roberts et al., 2002). Lateral roots originate from pericycle cells that acquire the founder cell (FC) identity and divide asymmetrically. Accumulation of auxin is one of the earliest events taking place during FC specification tightly preceding lateral root initiation (LRI) (Dubrovsky et al., 2008). FC cells proliferate to form the lateral root primordium (LRP) which then grows through the overlying tissues of endodermis and cortex until it eventually breaks through the epidermis and emerges.

Gene expression analyses of the metacaspase protease family in roots showed that *AtMC5* and *AtMC9* could be related to lateral root protrusion. In fact, *AtMC5* and *AtMC9* are commonly expressed in endodermal cells, which overlie developing early-stage lateral root primordia (LRP) and need to displace for the primordium to protrude. To date our knowledge on plant metacaspase functions is mainly on cell death events related to biotic

and abiotic responses (Bozhkov et al., 2005, He et al., 2008, Coll et al., 2010, Watanabe and Lam 2011, Wang et al., 2012,) and development (Sundström et al., 2009). Driven by the intriguing expression of these two metacaspases during later root development, we decided to explore a possible role for them in the process of cell separation. Reverse genetics approaches helped us to identify root phenotypic aberrations caused by chemical inhibition or gene perturbation of metacaspases. Finally, proteome-wide analysis of AtMC9-dependent protein degradation revealed a set of potential AtMC9 substrates whose cleavage could affect cell wall structure, and concomitant promotion of cell separation.

RESULTS

AtMC5 and AtMC9 spatiotemporal expression in Arabidopsis roots suggests a role in cell separation

The spatiotemporal expression of *Arabidopsis thaliana* metacaspases was studied in the roots using promoter GUS reporter lines. Remarkably, in the primary root, *AtMC5* and *AtMC9* were expressed in the endodermis during early lateral root primordia formation and in particular in cells overlying developing primordia (Figure 1). Both metacaspases were additionally expressed in the columella and epidermal cells at the root tip (Figure 1) showing that these two proteases are specifically expressed at locations where loosening of the cell wall is required for subsequent cell separation. Besides *AtMC6*, all metacaspase genes were expressed in the root, although with distinctive expression levels and spatial patterns (Supporting information figure S1).

To assess whether *AtMC5* and *AtMC9* are involved in cell separation, we developed and characterised transgenic lines perturbed in metacaspase function. *atmc5*, *atmc9* and the double *atmc5mc9* T-DNA lines (KO) were characterised for lateral root protrusion. The absence of *AtMC* transcripts in the T-DNA knock-out lines was validated by RT-PCR (Figure 2A and 2B). Surprisingly no significant differences were observed between wild-type plants and single or double KO lines after quantitative observation of emerged and non-emerged lateral roots (Figure 2C), indicative of possible redundancy in metacaspase function.

Identification of metacaspase inhibitors

To overcome the possible metacaspase redundancy that is suggested by the absence of a loss-of-function phenotype, we studied metacaspase functionality via a chemical biology approach. Given that *AtMC9* is so far the best biochemically described *Arabidopsis* metacaspase (Vercammen et al., 2004 and 2006), a high-throughput adaptation of an *AtMC9* fluorescence-based biochemical assay was developed by using recombinant *AtMC9* and a fluorescently-labeled VRPR tetrapeptide as an optimized fluorogenic substrate (Vercammen et al., 2006). This miniaturized *AtMC9* biochemical assay was used to screen a chemical library of 10,000 small synthetic molecules in triplicate with the aim to identify inhibitors of VRPRase activity. Ac-VRPR-fmk, a potent metacaspase inhibitor was used as a positive control. In total, 81 compounds were identified that inhibited *AtMC9* activity by at least 75% in three replicate screenings at a concentration of 10 μ M. Clustering of hit compounds by structural and physicochemical similarities revealed 30 compounds with a thioxodihydropyrimidinedione (TDP) substructure (ChemMine – Backman et al., 2011). This TDP cluster corresponds to 37% of the hit compounds. In the entire screening library, 79 compounds contain a TDP substructure (0.8%), indicating that the hit list is highly enriched in TDP-containing compounds (Figure 3A). The high proportion of compounds with a TDP substructure implies that this substructure is an important structural feature for *AtMC9* inhibition *in vitro*.

We subsequently focused on the TDP cluster and, based upon inhibitory activity and location within the cluster, we selected seven compounds for further characterization (TDP1 to TDP7, Figure 3B-C). The effect of the compounds was assessed on other available recombinant metacaspases (*AtMC4* and *AtMC5*). *AtMC5* activity was inhibited at least 50% by all compounds except TDP2, which showed a high specificity towards *AtMC9*. *AtMC4* activity was not significantly inhibited ($p < 0.05$) by any compound (Figure 4).

Chemical inhibition of metacaspases impairs cell separation

Identification of a loss-of-function phenotype caused by chemical inhibition of metacaspases was our strategy to attribute the shared *AtMC5* and *AtMC9* expression pattern to a dedicated function. For this, root growth and development of 3-days old *Arabidopsis* seedlings were monitored after a 5-day growth period in media supplemented with 50 μ M of inhibitors. Several compounds decreased the number of emerged lateral roots, with TDP3 and TDP6 having the most pronounced effect (Figure 5). Unlike TDP3, TDP6 did not have a significant effect ($p < 0.05$) on primary root length, indicating that TDP6 specifically inhibits lateral root development. Interestingly, TDP2, which showed a similar inhibitory activity towards *AtMC9* as TDP6, but did not show inhibition of *AtMC4* and *AtMC5* (Figure 4), affected lateral root development only to a minor degree, suggesting that sole inhibition of *AtMC9* is not sufficient to impair lateral root development. Detailed characterization of the TDP3 and TDP6 effect on lateral root emergence showed that the proportion of non-emerged primordia in treated seedlings was on average 20% higher compared to the untreated control, thus indicating that the compounds disrupt the emergence of lateral roots (Figure 6). The observed phenotype resembles the pertur-

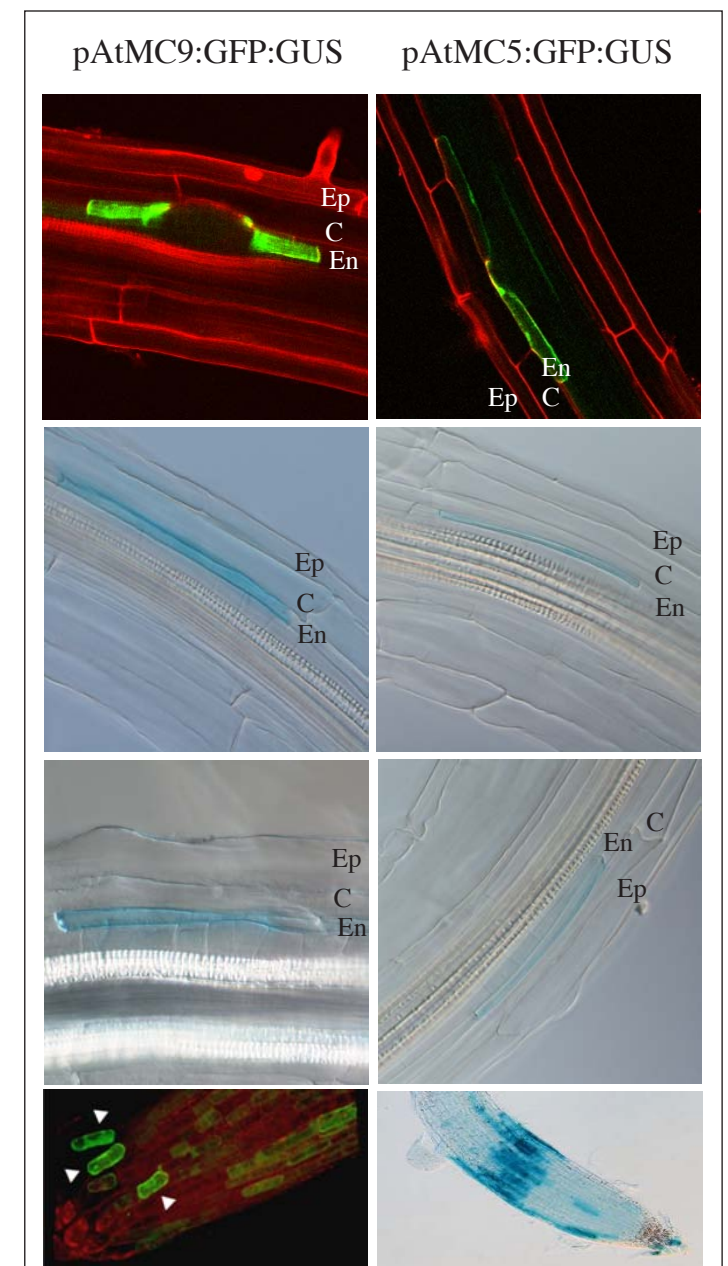
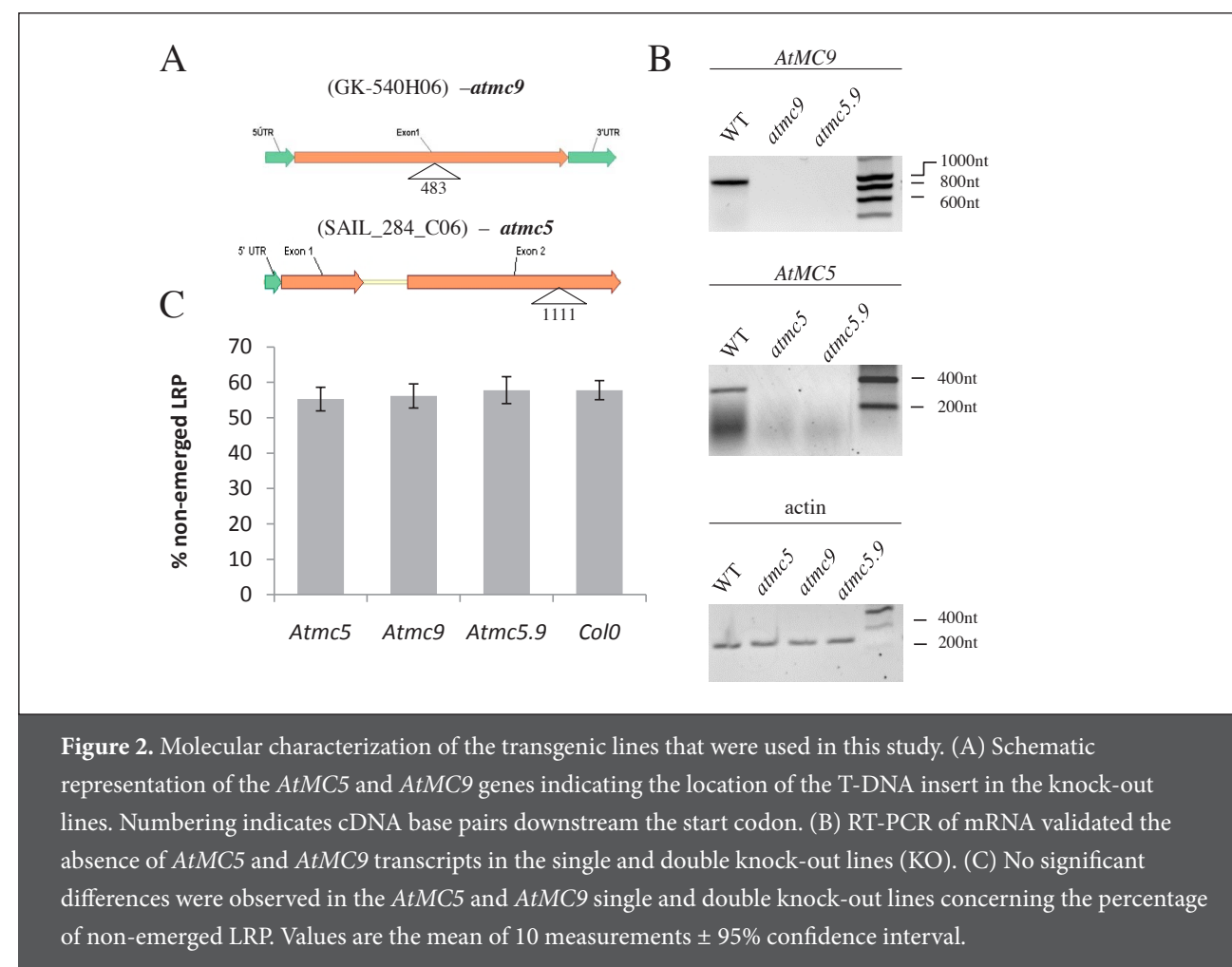


Figure 1. Expression of *AtMC5* and *AtMC9* in endodermis cells that overlie early stage lateral root primordia and in the root tip. En: endodermis, C: cortex, Ep: epidermis



bation of the auxin influx carrier, LAX3, which enhances lateral root protrusion by promoting cell separation of cortical and epidermal cells adjacent to lateral root primordia. *lax3* seedlings show a 40% increase in non-emerged lateral root primordia (Swarup et al., 2008).

In addition to the lateral root phenotype (Figure 6), plants treated with TDP3 and TDP6 also showed reduced release of root cap border cells, leading to multiple root cap cell tiers attached to the root tip. In fact, seedlings treated with the compounds had on average twice more root cap cell tiers than the untreated control (Supporting information figure S2). Taken together with the *AtMC5* and *AtMC9* expression patterns, these data suggest that chemical inhibition of *AtMC5* and *AtMC9* impairs cell separation in root tissues and results in root phenotypic aberrations including reduced detachment of the outer root tip cell layers and impaired lateral root organogenesis. Furthermore, it is suggested that type-II metacaspases might act redundantly in this process.

To validate our findings by genetic means, we developed and characterised transgenic lines perturbed in metacaspase function. Two knock-down lines (KD1 and KD2) of all type-II metacaspase genes (*AtMC4* to *AtMC9*) were characterised for LRP stages (Malamy et al., 1997). *AtMC* transcripts in the KD lines were analysed by RT-PCR and the level of down-regulation of each metacaspase gene transcript varied (Figure 7A). Interestingly, KD lines contained more non-emerged lateral root primordia than the wild-type (Figure 7B and

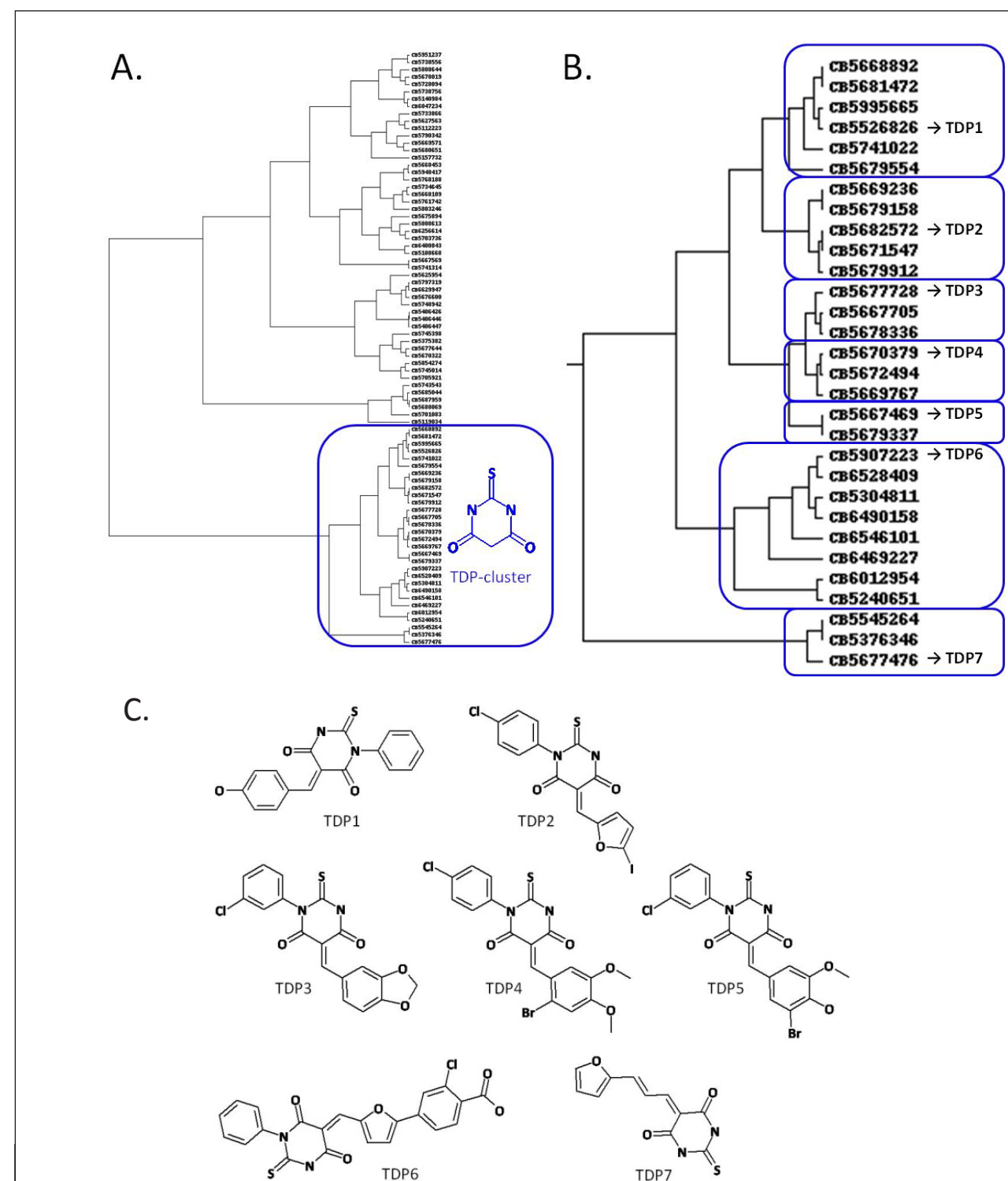


Figure 3. Structure-based clustering of chemical compounds with 75% *AtMC9* inhibitory effect (ChemMine – Backman et al., 2011). (A) 30 out of 81 *AtMC9* contain a thioxodihydropyrimidinedione (TDP) substructure. (B-C) Selection of seven representative compounds from each TDP subcluster based on their inhibitory effect against the VRPR-ase activity of *AtMC9*.

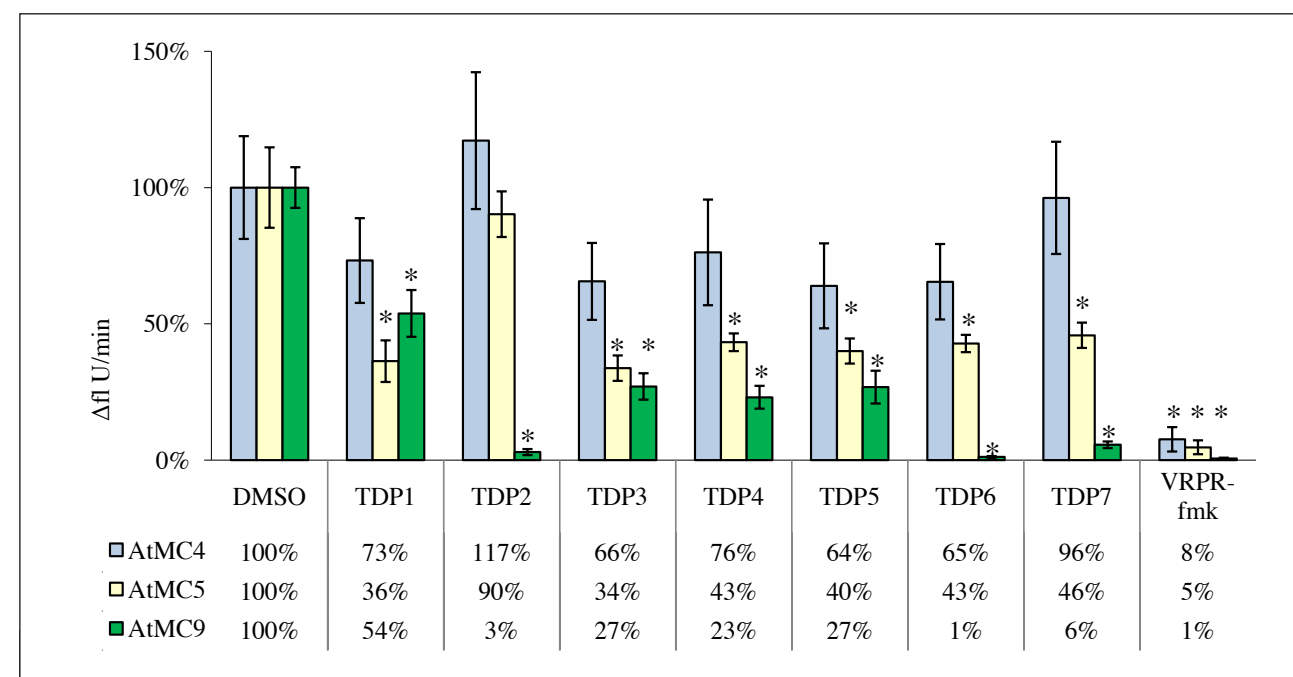


Figure 4. Residual VRPR-ase activity of AtMC9, AtMC4 and AtMC5 upon inclusion of the selected TDP-containing compounds in the assay mixture. DMSO was the solvent of the compounds while VRPR-fmk was used a positive control for metacaspase inhibition. Values are the mean of three replicate measurements \pm 95% confidence interval. Activity is measured as the rate of fluorescence release per min. *, significant difference from the DMSO control (Student's t-test, $p < 0.05$, $n = 3$).

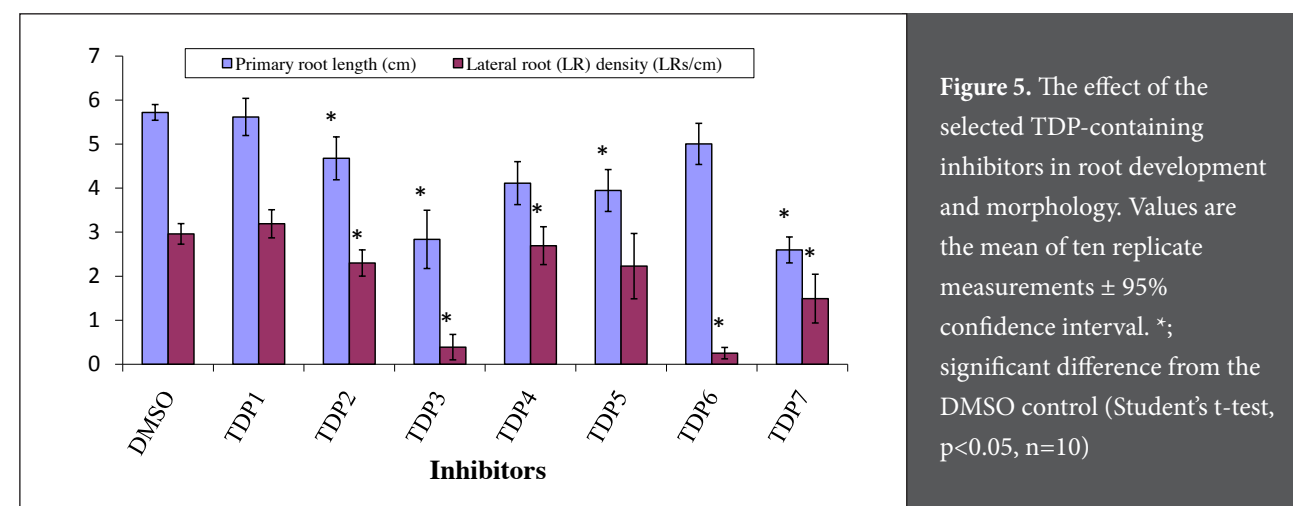


Figure 5. The effect of the selected TDP-containing inhibitors in root development and morphology. Values are the mean of ten replicate measurements \pm 95% confidence interval. *, significant difference from the DMSO control (Student's t-test, $p < 0.05$, $n = 10$).

7C). In fact, both KD lines showed a significantly increased number of lateral root primordia ($p < 0.05$) that have not yet protruded the endodermis (stages 1-3), suggesting that the endodermis is the restraining tissue for lateral root protrusion when type-II metacaspase levels are down-regulated. In accordance with the observation that *AtMC5* and *AtMC9* are expressed in the endodermis, chemical and genetic perturbation of type-II metacaspases shows that metacaspase activity is important for the early development of lateral root primordia, possibly by modulating cell-to-cell adhesion and in this manner facilitating lateral root protrusion.

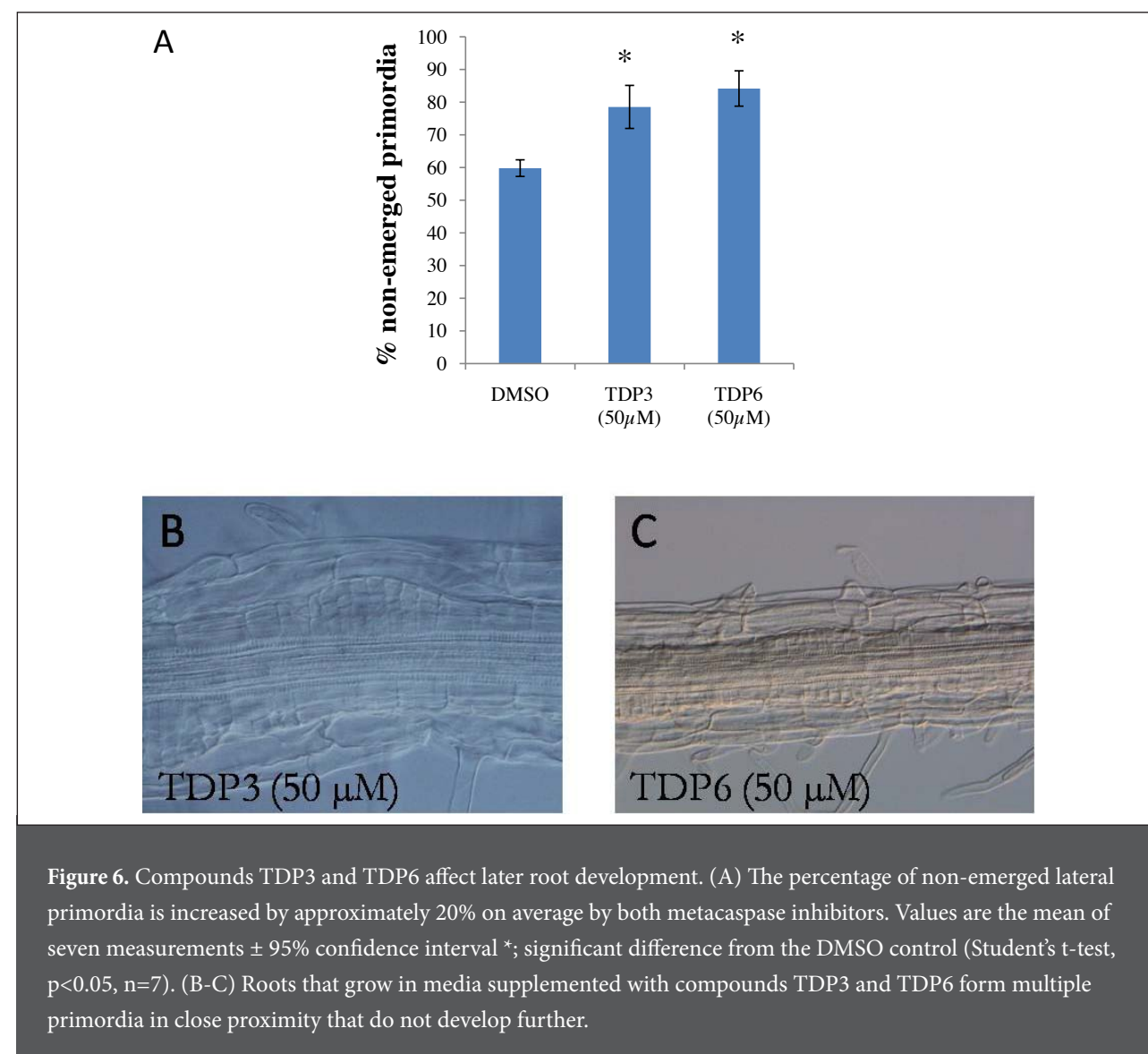


Figure 6. Compounds TDP3 and TDP6 affect later root development. (A) The percentage of non-emerged lateral primordia is increased by approximately 20% on average by both metacaspase inhibitors. Values are the mean of seven measurements \pm 95% confidence interval *, significant difference from the DMSO control (Student's t-test, $p < 0.05$, $n = 7$). (B-C) Roots that grow in media supplemented with compounds TDP3 and TDP6 form multiple primordia in close proximity that do not develop further.

AtMC9 activity is auxin-dependent and it is involved in lateral root emergence

Because *AtMC9* is the main expressed metacaspase in endodermal cells overlying lateral root primordia and lateral root emergence is dependent on auxin response in the endodermis, we assessed the effect of auxin treatment on *AtMC9* expression using the auxin analogue 1-naphthaleneacetic acid (NAA). Although auxin did not change *AtMC9* transcript levels in whole root tissues (Figure 8A), *AtMC9* activity was enhanced by NAA and in the *AtMC9* loss-of-function background NAA did not restore *AtMC9* activity (Figure 8B). It has been shown that, in the endodermis, degradation of the SHY2/IAA3 repressor in an auxin-dependent fashion drives lateral root emergence by enhancing the expression of cell wall-remodelling enzymes (Swarup et al., 2008). Interestingly, in the *shy2-2* gain-of-function mutant, that shows delayed lateral root emergence, *AtMC9* expression was notably reduced (Figure 8C). In accordance with the reduced *AtMC9* transcript level, the inhibitory effect of metacaspase inhibitors TDP3 and TDP6 on lateral root emergence was further enhanced in the *shy2-2* mutant background (Figure 8D). Taken together, these data indicate that *AtMC9* activity is triggered by auxin but given

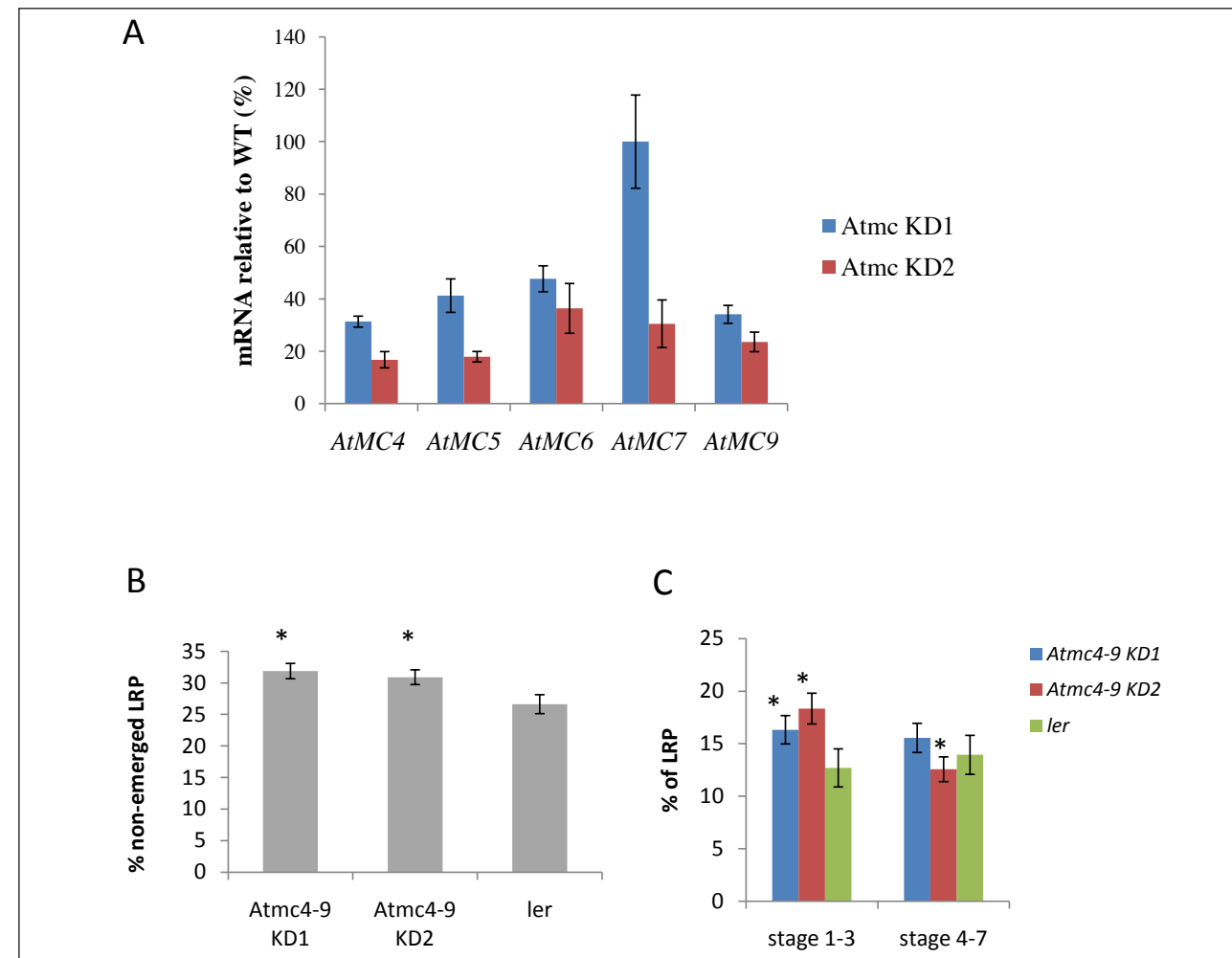


Figure 7. Characterization of the lateral root primordia (LRP) in the different metacaspase mutant lines. (A) Fold reduction of each type-II metacaspase mRNA in the two knock-down lines (KD) that were generated by means of the artificial microRNA technology (Schwab et al., 2006). Values represent the mean \pm SE of three biological replicate measurements. (B) Two independent type-II MC silencing lines show a significant increase in non-emerged LRP in comparison to the wild type roots (Ler ecotype). (C) More specifically, the morphological difference is seen at the early stages of LRP development (stages 1-3), when the primordia have not yet protruded the endodermis cell layer (Malamy et al., 2007). Values represent the mean of thirty measurements \pm 95% confidence interval. *; significant difference from the wild type Ler (Student's t-test, $p < 0.05$, $n = 10$).

that *AtMC9* expression is not dependent on auxin, it remains unclear whether *AtMC9* expression is SHY2/IAA3-dependent. However, the low *AtMC9* expression in combination to the delayed lateral root emergence rate observed in the *shy2-2* suggests that *AtMC9* is an endodermal protease involved in lateral root emergence.

Proteome-wide identification of potential *AtMC9* substrates in *Arabidopsis* roots

Upon metacaspase inhibition using chemical and genetic means, the involvement of these proteases in lateral root morphology became apparent. *AtMC9* is the predominantly expressed metacaspase in the endodermis (Figure 1) which is the restraining tissue for LRP emergence in type-II metacaspase KD lines (Figure 7C).

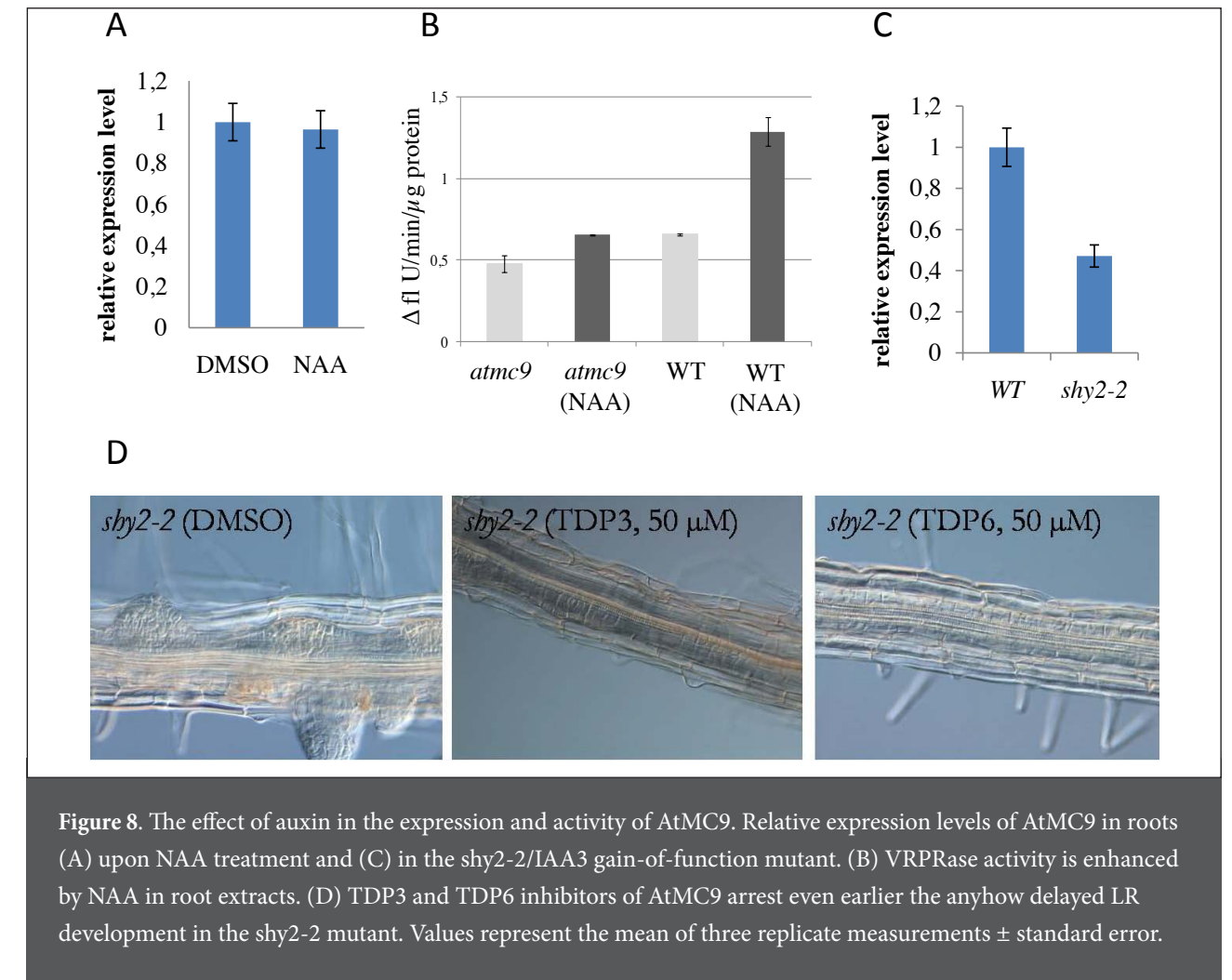


Figure 8. The effect of auxin in the expression and activity of *AtMC9*. Relative expression levels of *AtMC9* in roots (A) upon NAA treatment and (C) in the *shy2-2*/IAA3 gain-of-function mutant. (B) VRPRase activity is enhanced by NAA in root extracts. (D) TDP3 and TDP6 inhibitors of *AtMC9* arrest even earlier the anyhow delayed LR development in the *shy2-2* mutant. Values represent the mean of three replicate measurements \pm standard error.

To understand better the mechanism of *AtMC9* action in this process we opted for the identification of its substrates. A positional proteomics approach was used to screen the root proteome of *Arabidopsis* plants for proteins that are differentially cleaved in *AtMC9* loss- and gain-of-function mutants. N-terminal COFRADIC (Gevaert et al., 2003) was applied to identify N-termini that were created by proteolytic events. These neo-N-terminal peptides revealed *AtMC9*-driven proteolysis and they were used as a reference for the comparison of *AtMC9* loss- and gain-of-function N-terminal proteomes.

Two *in vivo* analyses were done in which the root N-terminal proteome of *AtMC9* loss-of-function (KO) plants was compared to wild type (Col0) or *AtMC9* gain-of-function (overexpression, OE) plants. An additional analysis was an *in vitro* set-up in which the KO root proteome treated with recombinant *AtMC9* was compared to an untreated sample. Using N-hydroxysuccinimide esters of isotopic variants of butyric acid (NHS-But) we mass-tagged neo-N-terminal peptides for further quantification (Figure 9) (Staes et al., 2011), and the peptide couples that were expected are briefly described in the following. a) A neo-N-terminal peptide is uniquely found in the sample containing active (r)*AtMC9* meaning that the corresponding protein was cleaved by *AtMC9* (Figure 10A). b) A (neo-) N-terminal peptide is uniquely found in the control sample (KO) because the corresponding protein was completely degraded in the sample of active *AtMC9* whereas in the control

sample the protein remained unprocessed. For these AtMC9 substrates we cannot draw conclusions about the cleavage site (Figure 10B).

c) A neo-N-terminal peptide is found in both samples, but with significantly higher intensity in the sample containing active AtMC9, which indicates that the corresponding protein was also cleaved in the control sample (without endogenous AtMC9 activity), although to a lesser extent, by a protease with redundant activity to AtMC9 (Figure 10C). Note that because of metacaspase redundancy in the root tissues,

such cases needed to be considered. d) (Neo-) N-terminal peptides found in approximately equal amounts in both proteomes. These do not indicate AtMC9 proteolysis (Figure 10D). Thus, the aim of the three proteomic analyses was to identify neo-N-terminal peptides that were either uniquely found in root proteomes containing active AtMC9 (singletons) or were found with significantly higher intensity ($p < 0.02$). In addition, unique N-terminal peptides found in the KO proteome of the *in vivo* analyses were considered as indirect AtMC9 targets. In total, we identified 1,141 proteins in the KO/WT analysis, 978 proteins in the KO/OE analysis and 1,315 proteins in the *in vitro* analysis. Based on our aforementioned criteria we identified in total 183 cleavage sites (77 *in vivo* and 106 *in vitro*), 87 of which were reported by singleton peptides (9 *in vivo* and 78 *in vitro*; Figure 10A scenario), 96 sites were reported by significantly enriched peptides in the equivalent proteomes of active AtMC9 (68 *in vivo* and 28 *in vitro*; Figure 10B scenario) and finally 21 peptides were reporters of potential *in vivo* AtMC9 substrate precursor proteins that were not processed in the KO samples (Figure 10C scenario). Together, all these cleavage sites were identified in 87 *in vivo* and 79 *in vitro* cleaved proteins (Supplementary information tables S1, S2 and S3) that we thus selected as candidate AtMC9 targets in the roots.

Substrate specificity of AtMC9

The canonical AtMC9 specificity for Arg and Lys (Vercammen et al., 2004) was preserved among the cleavage sites. In fact, 70% and 72% of the cleavage sites were generated after Arg or Lys in the KO/OE *in vivo* and the *in vitro* analysis, respectively (Figure 11A). However, this was not observed in the KO/WT *in vivo* study where the canonical AtMC9 specificity accounted for only 13% of the cleaved sites (Figure 11A). In the wild-type situation, the extent of AtMC9 substrate cleavage is lower than when AtMC9 is overexpressed and this could be the reason why we identified less Arg/Lys-cleaved peptides in the KO/WT than in the KO/OE analysis. Furthermore, it could be that AtMC9-generated protein fragments were further cleaved or degraded by other proteases downstream of AtMC9. Considering that AtMC9 is a genuine substrate of its own and the sole *in vivo* identified AtMC9 peptide was generated after an Asn residue (Table S2), consolidates that AtMC9 genuine fragments possibly underwent further downstream processing.

Using the amino acid sequence information spanning the P4-P3' substrate positions of only the Arg/Lys-cleaved proteins, we analysed the AtMC9 specificity in root proteins (Colaert et al., 2009). As the *in vitro* dataset describes a larger number of Arg/Lys-cleaved proteins than the *in vivo* datasets, the former is more informative

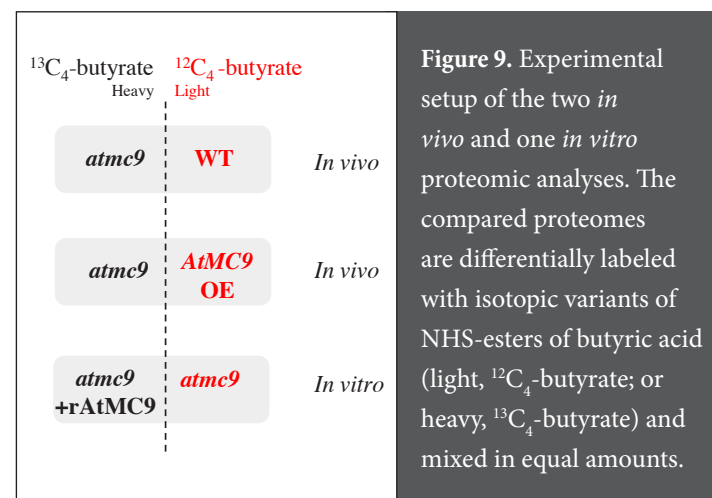


Figure 9. Experimental setup of the two *in vivo* and one *in vitro* proteomic analyses. The compared proteomes are differentially labeled with isotopic variants of NHS-esters of butyric acid (light, $^{12}\text{C}_4$ -butyrate; or heavy, $^{13}\text{C}_4$ -butyrate) and mixed in equal amounts.

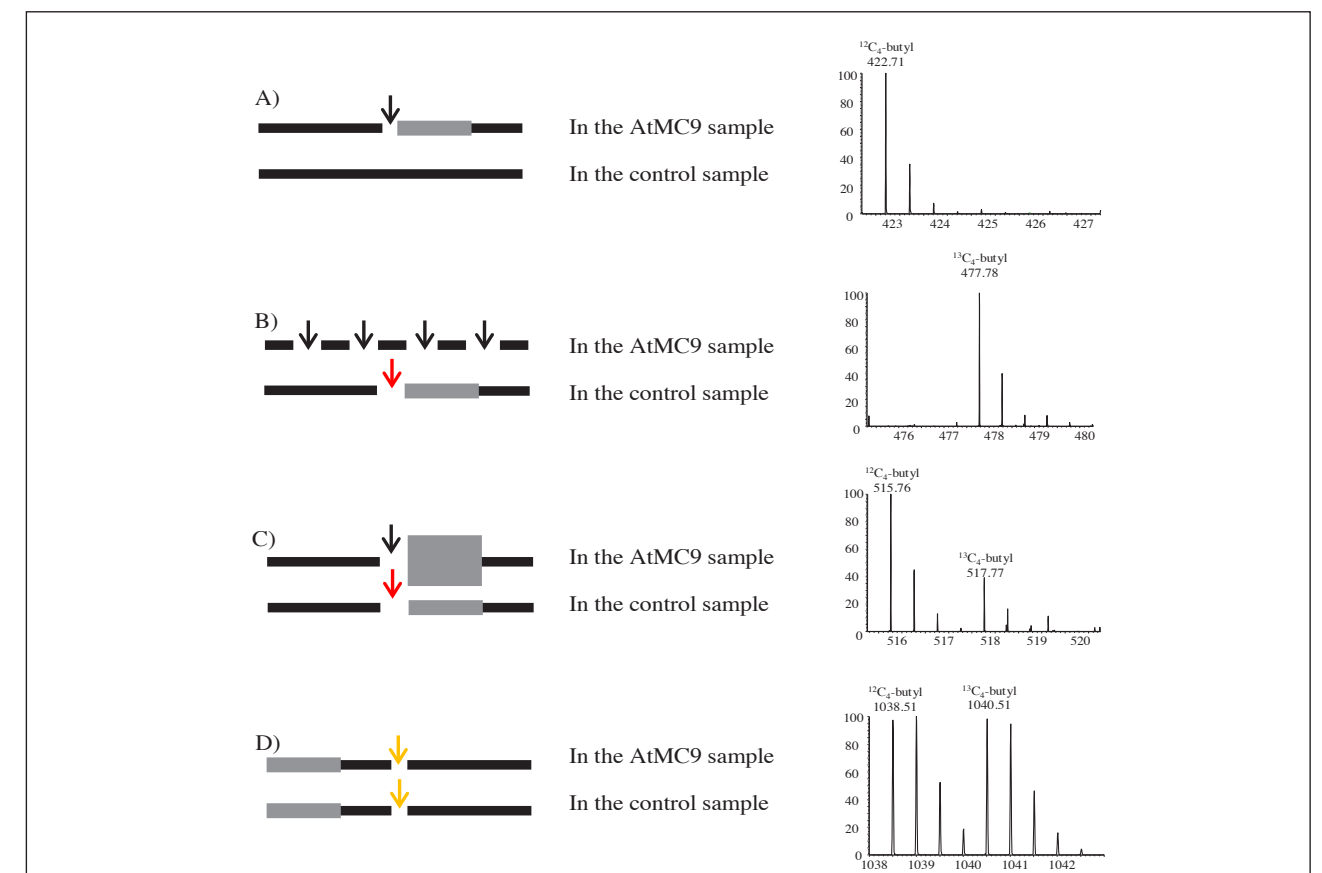
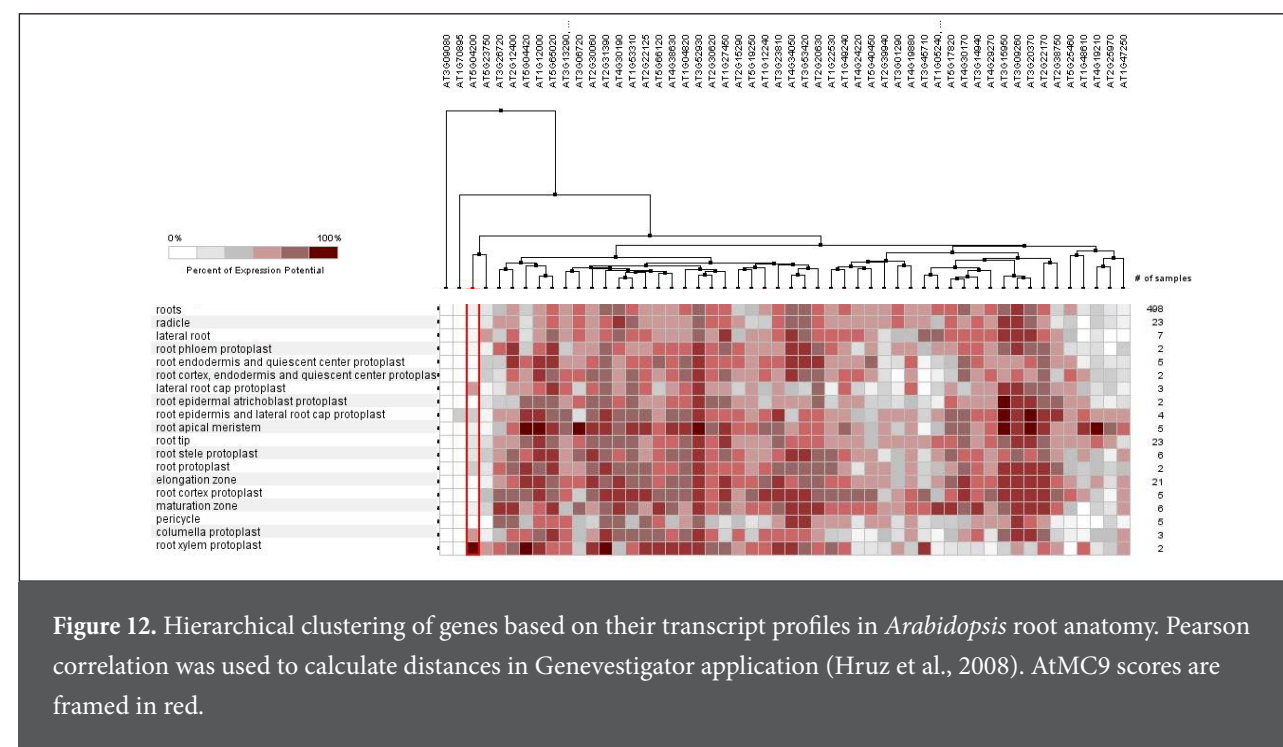
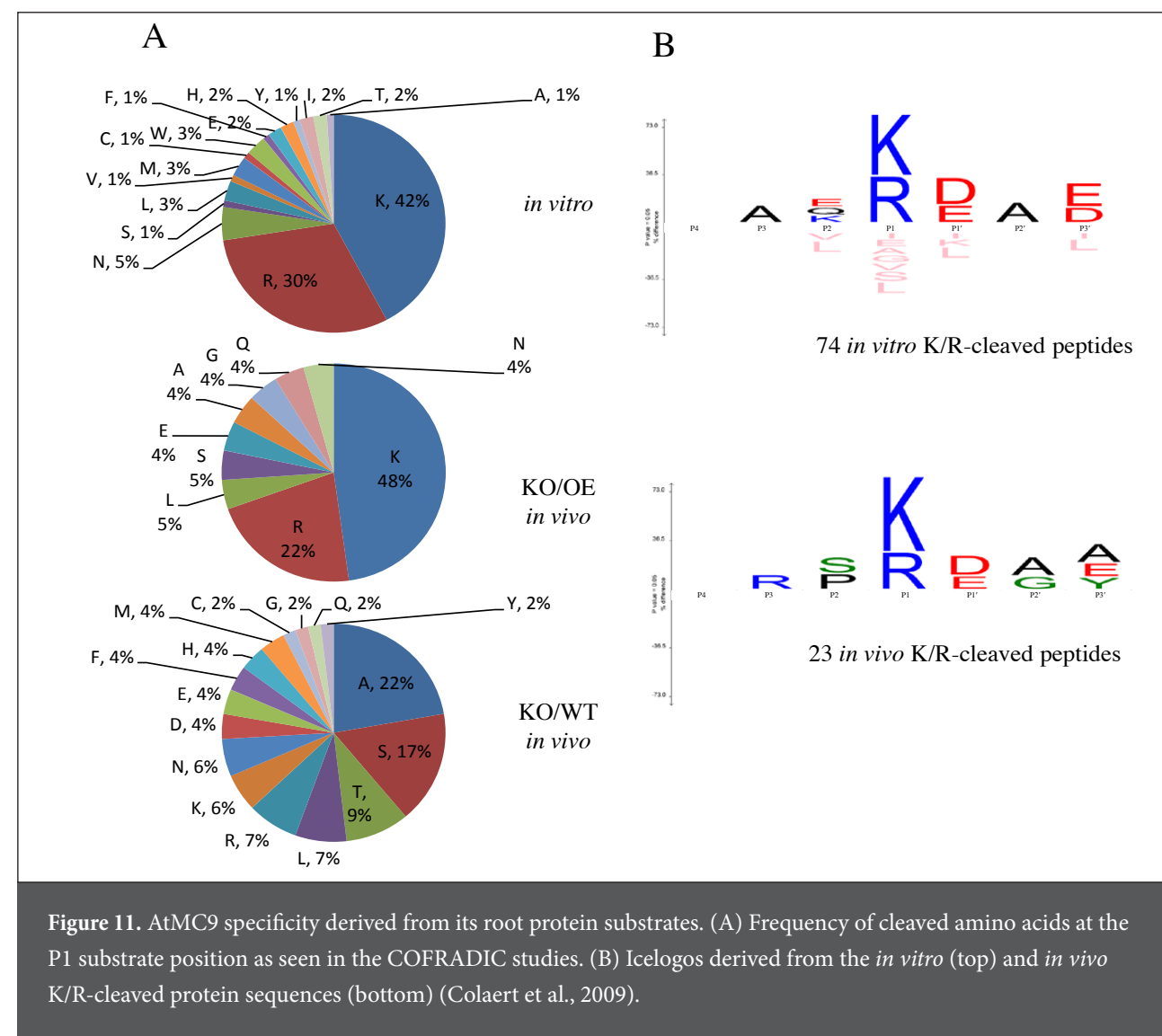


Figure 10. Possible proteolysis scenarios (left) and read-outs (right). (A) A protein is solely cleaved by AtMC9 (black arrow) in the samples of active (WT or OE) or added (rAtMC9) protease. Thereby, a fragment with a neo-N-terminus is generated that is later identified uniquely in the equivalent proteome as a singleton ion. (B) A protein is completely degraded by AtMC9 and the fragments are not identifiable. (C) A protein is cleaved in both samples but in a higher frequency in the AtMC9 sample. The neo-N-terminus fragment is generated in both samples but in significantly higher intensity in the protease sample. Proteolysis of the target in the control sample may have occurred by a protease of redundant to the AtMC9 activity (red arrow). The fragment is identified as a doublet with a higher intensity ion derived from the AtMC9 sample. (D) Fragments generated by trypsin (yellow arrows) that bear the mature N-terminus of a protein are found in approximately equal intensities in both samples.

about AtMC9 specificity. Nevertheless, recombinant and native AtMC9 specificities were very similar, particularly at the C-terminal side of the scissile bond (prime site positions P1'-P3') (Figure 11B). Furthermore, the specificity of native AtMC9 against root proteins was very similar to the one previously described towards seedling protein substrates (Supporting information Figure S3). In fact, Lys was relatively preferred over Arg at P1, and the preference for acidic residues (Asp and Glu) at the prime site positions P1' and P3' was preserved in both tissues. Moreover, at position P2 in both consensus sequences, small hydrophilic residues like Ser and Thr were found while basic Arg and Lys were preferred at P3. Overall, the specificity of native AtMC9 against either root or seedling substrates was similar. The small differences that were observed are probably inherent to the origin of the studied proteomes and to the lower number of *in vivo* root substrates.



family protein) are highly expressed in the root endodermis (Brady et al., 2007) and the proteins were both cleaved after an Arg. Taken into account the root phenotypic effects of AtMC9 inhibition and the extracellular localisation of AtMC9 (Vercammen et al., 2004), we believe that the subset of AtMC9 co-expressed substrates (Table 2) that localize in the extracellular space, such as several peroxidases and glycosyl hydrolases, are of key importance in cell wall remodelling and cell separation. In addition, considering the AtMC9 nuclear and cytosolic localization (see Chapter 3), substrates found in the nucleus and cytosol could also be targeted by AtMC9, possibly prior to AtMC9 secretion in the apoplastic space. Further tests to assess the cleavage of the here identified *in vivo* and *in vitro* AtMC9 substrates, at the peptide and protein level, would validate further our findings and confirm their physiological relevance.

Selection and *in silico* characterisation of AtMC9 root substrates

Three independent proteomic analyses led to the identification of 166 potential AtMC9 substrates. Only 4 proteins were found in at least 2 independent analyses: the cysteine protease RD21 (AT1G47128), a DNA topoisomerase-related protein (AT3G15950), a hyaluronan/mRNA binding family protein (AT4G17520) and AtMC9 (Table 1). Except for AtMC9, the identified cleavage sites in the other proteins differed between the analyses. As we have previously mentioned this is probably due to the different intensity of AtMC9 proteolysis in the tested N-terminomes and the possible downstream processing of originally AtMC9-generated fragments by other proteases. Furthermore, as usually seen in *in vivo* degradome studies, identification of primary proteolytic products can be complex due to the instability of protease fragments in the cell (Schilling et al., 2008).

Hierarchical clustering of potential AtMC9 substrates based on their gene expression patterns showed 50 genes that are co-expressed with AtMC9 throughout *Arabidopsis* anatomy (Table 2 and Figure 12). Furthermore, two of these genes (At2g20630-PP2C induced by AVRRPM1 and At4g19880-Glutathione S-transferase

Table 1. Proteins characterized as potential AtMC9 targets in two independent analyses. Amino acid position of P1' site is shown between brackets in the P4-P3' column. Fold times of ion intensity enrichment in the proteome of active AtMC9 in comparison to the KO counterpart is shown in the column 'Intensity increase fold'. N/A; not applicable refers to fold change values of singleton ions which is not applicable due to the absence of light peptide ion.

Accession	Intensity fold increase	z-score	P4-P3'	Sequence	Description	Subcellular localization SUBAMS data	Experiment
AT1G47128	3.11	2.78	FlIK(215)NGG	NGGIDTDKDYYPYKGVDTGCDQIR	Cysteine protease RD21	extracellular, plasma	KO/WT in vivo
	15.52	5.94	KYLG(115)AKM	AKMEKKGER		membrane, plastid, vacuole	KO/OE in vivo
AT3G15950	N/A	N/A	HTVK(291)DDV	DDVDDKEQDAKR	DNA topoisomerase-related	extracellular, peroxisome,	KO/WT in vivo
	5.56	4.30	ESFK(322)QLE	QLEDIADNKAEGDDESAGR		plasma membrane, plastid	in vitro
AT4G17520	N/A	N/A	SKPL(27)KVV	KVVAPVQTAQSGKMPKPPPPSQAVR	Hyaluronan / mRNA binding family	cytosol, mitochondrion,	KO/WT in vivo
	6.95	3.75	QTAK(37)SGK	SGKMPKPPPPSQAVR		peroxisome	KO/OE in vivo
AT5G04200	N/A	N/A	GCIN(32)DVL	DVLAMKETILSR	metacaspase 9	extracellular, cytosol,	in vitro
	N/A	N/A	GCIN(32)DVL	DVLAMKETILSR		nucleus	KO/OE in vivo

Table 2. List of potential AtMC9 substrate proteins that are co-expressed with *AtMC9* (Hruz et al., 2008). Amino acid position of P1' cleavage site is shown between brackets in the P4-P3' column.*; subcellular localization predicted by WoLFPSORT.

Accession	P4-P3'	Sequence	Description	Subcellular localization SUBA (MS data)
Proteolysis of Arg or Lys				
AT1G27450	AASR(58)DSE	DSEMATQVQDPR	adenine phosphoribosyl transferase 1	cytosol, extracellular, plasma membrane, plastid
AT3G20370	TLLK(254)NSY	NSYLSEVFSIGGR	TRAF-like family protein	extracellular, vacuole
AT1G49240	MNQQ(53)DAY	DAYVGDEAQSQR	actin 8	cytosol, peroxisome, plasma membrane, plastid, vacuole
AT1G12000	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	cytosol, plasma membrane
AT3G45710	VAHR(15)SSD	SSDPSEKR	Major facilitator superfamily protein	plasma membrane*
AT3G15950	ESFK(322)QLE	QLEDIADNKAEGDDESQR	DNA topoisomerase-related	extracellular, peroxisome, plasma membrane, plastid
AT3G13300	RRDR(773)DAN	DANIQDVNNDPR	Transducin/WD40 repeat-like superfamily protein	cytosol, plasma membrane
AT5G04420	RRQR(490)SAS	SASDEEEDGTVQR	Galactose oxidase/kelch repeat superfamily protein	cytosol*
AT2G25970	QVTR(172)DMD	DMDADPNCATR	KH domain-containing protein	cytosol
AT3G06720	SIER(105)SPP	SPPIEEVISAGVVPR	importin alpha isoform 1	cytosol, nucleus, plasma membrane
AT2G20630	IESR(291)GGF	GGFVSNIPGDVPR	PP2C induced by AVRRP1	cytosol*
AT5G23750	VVVK(69)EVE	EVVEEKKEGVSQR	Remorin family protein	nucleus*
AT4G19880	SMAR(35)SAV	SAVDETSDSGAFQR	Glutathione S-transferase family protein	cytosol, mitochondrion
AT1G47250	IGSR(165)SQA	SQAAKTYLER	20S proteasome alpha subunit F2	cytosol, plasma membrane
AT3G15950	HTVK(291)DDV	DDVDDKEQDAKR	DNA topoisomerase-related	extracellular, peroxisome, plasma membrane, plastid
AT5G65020	KNLK(222)EES	EESDDNDYMKLLR	annexin 2	plasma membrane, plastid
AT4G30170	TVVK(105)AKQ	AKQAVDSNPNCR	Peroxidase family protein	extracellular, plasma membrane
AT1G22530	ETKK(272)EEK	EEKATASTQVQR	PATELLIN 2	plasma membrane, plastid, vacuole
AT4G30190	LQPK(889)EAV	EAVNIFPEKGSYR	H(+)-ATPase 2	plasma membrane, vacuole
AT3G15950	AATK(434)GLE	GLEELKNESEQAENKR	DNA topoisomerase-related	extracellular, peroxisome, plasma membrane, plastid
AT4G34050	EATK(11)TSS	TSSTNGEDQKQSNLR	S-adenosyl-L-methionine-dependent methyltransferase	cytosol*
AT5G40450	HQEK(1235)NAE	NAEPVEATQNLDDAEQISR	unknown protein	plasma membrane, plastid
AT2G22125	IHPK(1001)EKE	EKEEEDDEEATENR	CELLULOSE SYNTHASE-INTERACTIVE PROTEIN 1	plasma membrane
AT1G04820	ISGK(97)EDA	EDAANNFAR	tubulin alpha-4 chain	plasma membrane, plastid, vacuole
AT3G52930	KVDK(107)GTV	GTVELAGTDGETTTQGLDGLDR	Aldolase superfamily protein	cytosol, nucleus, plasma membrane, plastid, vacuole

AT3G14940	LYAK(200)DIT	DITPDDKQELDESLQR	phosphoenolpyruvate carboxylase 3	cytosol*
AT1G53310	LYAK(200)DIT	DITPDDKQELDEALQR	phosphoenolpyruvate carboxylase 1	cytosol, plasma membrane
AT3G53420	MAK(4)DVEA	DVEAVPGEGFQTR	plasma membrane intrinsic protein 2A	plasma membrane, plastid
AT5G66120	NDQR(74)SIS	SISSPTVVEVDLGDR	3-dehydroquinate synthase, putative	plastid
AT5G40450	NISK(175)VCE	VCEEIPIKTDEVR	unknown protein	plasma membrane, plastid
AT4G38630	QKDK(258)DGD	DGDTASASQETVAR	regulatory particle non-ATPase 10	cytosol, plasma membrane
AT2G30060	RENK(14)DEE	DEEETGANEDDTGAQVAPIVR	Pleckstrin homology (PH) domain protein	nucleus*
AT2G39940	RIER(352)GAD	GADEQGMEEDEGLVSQR	RNI-like superfamily protein	cytosol*
AT3G23810	SSGR(13)EYK	EYKVKDMSQADFGR	S-adenosyl-L-homocysteine (SAH) hydrolase 2	cytosol, plasma membrane, vacuole
AT4G19210	VIDR(213)DVE	DVENLSGGELQR	RNAse I inhibitor protein 2	plasma membrane
AT4G24220	WTNR(110)ESE	ESESENCEANGSMLR	NAD(P)-binding Rossmann-fold superfamily protein	cytosol
proteolysis of other amino acids				
AT2G30620	AAVS(209)KTK	KTKAVAAKPKAKERPAKASR	winged-helix DNA-binding transcription factor protein	nucleus
AT3G09260	DIQL(102)MKN	MKNLNTDAFR	Glycosyl hydrolase superfamily protein	nucleus, peroxisome, plasma membrane, plastid, vacuole
AT5G19250	DTAD(29)EED	EEDVLLTGINSYR	Glycoprotein membrane precursor GPI-anchored	extracellular, plasma membrane
AT3G01290	DVLD(236)MVM	MVMMTQYFDTMR	SPFH/Band 7/PHB domain-containing	mitochondrion, plasma membrane, vacuole
AT2G38750	EAVA(184)SSG	SSGEEAVEKDEVVR	annexin 4	plasma membrane
AT2G12400	ERV(56)AEE	AEESGENSSLILAAKR	unknown protein	plasma membrane
AT2G31390	FANA(298)CGA	CGAITTTKGAIPALPSDAEVR	pKb-like carbohydrate kinase family protein	cytosol, plasma membrane
AT5G17820	FVRA(289)MVK	MVKMGAVDVLTR	Peroxidase superfamily protein	extracellular, plasma membrane
AT3G26720	GVTS(23)EYI	EYIEYNTKPR	Glycosyl hydrolase family 38 protein	extracellular, vacuole
AT1G05240	KLGF(307)VQI	VQILTKNGEIR	Peroxidase superfamily protein	extracellular, plasma membrane
AT3G20370	KVSA(74)SNA	SNAVKGLR	TRAF-like family protein	extracellular, vacuole
AT3G13300	PQLL(924)AMQ	AMQETMNQVMASQKEMQR	Transducin/WD40 repeat-like superfamily protein	cytosol, plasma membrane
AT1G48610	PQTD(33)ATG	ATGVSATDTASQKR	AT hook motif-containing protein	plasma membrane
AT4G29270	RGQN(202)DQG	DQGTATQYKSEQR	HAD superfamily, subfamily IIIB acid phosphatase	extracellular
AT5G17820	RVGF(29)USQ	YSQSCPQAEIVR	Peroxidase superfamily protein	extracellular, plasma membrane
AT5G25460	TAMA(20)AKS	AKSTVSFR	Protein of unknown function, DUF642	extracellular, plasma membrane
AT2G15290	TVAF(84)SYP	SYPTSPSSVPGDNEVDKAKLAQVAKR	translocon at inner membrane of chloroplasts 21	plasma membrane, plastid
AT2G22170	VALA(26)DDE	DDEADCYVTFLLR	Lipase/lipoxygenase, PLAT/LH2 family protein	extracellular, plasma membrane, plastid
AT3G09080	VFVQ(968)SIS	SISELSTASPR	Transducin/WD40 repeat-like superfamily protein	nucleus*
AT1G12240	VVAE(505)AEE	AEEFCEKSGGSTVR	Glycosyl hydrolases family 32 protein	vacuole

DISCUSSION

Driven by the expression pattern of two metacaspases during early stages of LR development, we sought to identify a functional role for these proteases in root morphology. For this we adopted two distinct reverse genetics approaches; chemical inhibition of metacaspase activity and genetic perturbation of metacaspase genes.

Chemical inhibition is a very fast and targeted way to study the effect of a protein or protein family in a certain tissue. Several chemical compound screenings have been performed successfully in mammalian setups for inhibiting proteins with therapeutic potential (French et al., 2003; Hu et al., 2004; Cheung et al., 2002; Degterev et al., 2001). Screening of a chemical compound library against the VRPRase activity of AtMC9 *in vitro* revealed a set of strong inhibitors containing the TDP substructure. Chemicals containing the same substructure have also been effective against the human Omi/HtrA2 protease, a mitochondrial serine protease that is released to the cytoplasm upon induction of apoptosis (Cilenti et al., 2003). Two of the here identified metacaspase inhibitors were effective against both AtMC5 and AtMC9 and when applied to roots, LRP development was arrested and multiple primordia were formed.

Metacaspase involvement in cell separation during lateral root development

Together with the process of root cap cell removal from the root tip, lateral root development depends on cell separation (Roberts et al., 2002). In plant roots, detachment of root cap cells and lateral root development occurs upon disruption of cell-to-cell adhesion and this requires elaborate mechanisms to first loosen-up the middle lamella, the outermost layer of cell wall that cements cells to each other, and then allow cells to separate. Plants continuously release a large number of border cells from the root cap, a process that plays an essential role in the establishment of the root-soil interface (Hawes et al., 1990). Separation of root border cells from the root cap is correlated with pectolytic enzyme activity in root cap cells, which allows hydrolysis of polygalacturonic acid, the backbone of pectin polymers, which constitute the middle lamella and cell wall (Driouich et al., 2007). Similarly, lateral root initiation occurs within internal root tissues and requires a developing lateral root primordium to break through overlying tissues before lateral root emergence can occur (Péret et al., 2009). Again, this is associated with cell separation, which is mediated through the action of cell wall remodelling enzymes and expression of these enzymes is regulated by the lateral root promoting hormone auxin.

Upon chemical inhibition or genetic perturbation of metacaspases, we noticed that LRP development was affected (Figure 5, 6 and 7B-C). Detailed staging of LRP in type-II metacaspase knock-down lines showed that growth was arrested at the endodermis (Figure 7C). The same observation was done in *shy2-2* mutant plants in which auxin responsiveness is delayed leading to the formation of numerous primordia that do not protrude the epidermis (Figure 8D). Interestingly, treatment of *shy2-2* with the compounds led to earlier LRP arrest than it is normally seen in this mutant indicating that metacaspase activity may affect auxin delivery in the root via the developing LRP by regulating its passage through the endodermis.

Candidate AtMC9 substrates potentially involved in cell wall remodelling

Cell wall remodelling leading to either loosening or thickening has been correlated with the activity of cell wall remodelling enzymes such as polygalacturonases (PG), endo- β -1,4-d-glucanases and expansins. In fact, the expression pattern of AtMC9 closely resembles that of an *Arabidopsis* PG, PGAZAT (González-Carranza et

al., 2002). Besides the abscission zones of floral organs, activity of the PGAZAT promoter could be observed, although not consistently, in root cap cells and sites of lateral root emergence, very similar to the expression pattern of AtMC9 (Figure 1). Root cap specific expression of an endo- β -1,4-d-glucanase or cellulase has been reported previously, and in pea, release of root cap cells is correlated with pectolytic activity (del Campillo et al., 2004; Hawes and Lin, 1990).

The here identified potential AtMC9 protein substrates that further co-express with *AtMC9* could be associated with cell wall remodelling that precedes cell separation (Table 2). Glycosyl hydrolases (GHs) are encoded by multiple gene families in *Arabidopsis* and hydrolyse the glycosidic bonds between two or more carbohydrates. AtMC9 substrates belong to the GH family 1 (AT3G09260), 32 (AT1G12240), 38 (AT3G26720) and 47 (AT3G21160), whose members have not yet been correlated with cell wall remodelling as for instance the GHs of family 17. However, since cell walls consist of 90% carbohydrates and only for 10% of proteins (Wojtaszek, 1997), we believe that modulation of GH by AtMC9 might impact cell wall structure.

Four different peroxidases (AT3G49120, AT4G30170, AT1G05240, and AT5G17820) were only identified in the AtMC9-KO proteome possibly as a result of their degradation in the WT proteome by AtMC9 (Table S1). Through the peroxidative cycle, secreted class III peroxidases oxidize various substrates. The radicals produced by the peroxidative cycle cause the cross-linking of phenolic compounds in the cell wall reinforcing its rigidity (reviewed in Passardi et al., 2004). This way, inactivation of peroxidases in the apoplast by AtMC9 could possibly lead to cell wall loosening. Finally, UDP-D-apiose/UDP-D-xylose synthase 2 (AT1G08200), a protein whose depletion leads to cell wall thickening and cell death (Ahn et al., 2006) underwent a 17-fold higher processing in the AtMC9-OE proteome than in the AtMC9-KO *in vivo* (Table S2).

Constitutive and associated cell cytoskeleton proteins are potential AtMC9 targets. Actin 8 (AT1G49240), tubulin alpha-4 chain (AT1G04820) and cellulose synthase interactive protein 1 (CSI1) were cleaved by rAtMC9 in the *in vitro* analysis (Table S3). The cell cytoskeleton functions in cell shape maintenance, cell division, trafficking and cell morphogenesis. During cell expansion, microtubules guide the deposition of cellulose microfibrils through their association to cellulose synthase (CESA) complexes. Furthermore, cellulose synthase interactive protein 1 (CSI1) was shown to link microtubules with cellulose synthase complexes (Li et al., 2012) and it is required for microtubule stability (Mei et al., 2012).

A cysteine protease, Responsive to Desiccation-21 (RD21), which is like AtMC9 inhibited by the extracellular AtSerp1-1, was identified as an *in vivo* substrate of AtMC9 (Table S2). RD21 is a papain-like protease that confers plant immunity to the necrotrophic pathogen *Botrytis cinerea* (Shindo et al., 2012). RD21 maturation takes place in two steps. The first step involves the removal of a N-terminal propeptide (NTPP) from the RD21 precursor to generate a RD21 intermediate form that is later processed at the C-terminus by slow removal of a C-terminal propeptide (CTTP) during the second step of protein maturation (Yamada et al., 2001). In the same study it was shown that removal of the NTPP takes place in acidic conditions (pH 5.5) and processing is performed by an endogenous *Arabidopsis* protease other than the VPEs. Interestingly, AtMC9 is most active at pH 5.5 and the identified peptide (AKMEKKGER) was 15-fold enriched in the AtMC9-OE proteome and found upstream of the protease domain at the N-terminal part of the RD21 proprotein. Therefore, AtMC9 might be responsible for processing of RD21 into its mature form. However, the abovementioned peptide is not generated after Arg or Lys, which indicates that AtMC9 protein fragments may have undergone further processing

by other proteases. Since the substrates of RD21 (besides AtSerpin 1) remain unknown we cannot deduce a function for RD21 in cell wall remodelling. However, the co-ordinated function of both proteases (RD21 and AtMC9) in the extracellular space may have an effect on cell wall components.

AtMC9 substrate conservation among tissues

We have previously reported on the AtMC9 degradome in *Arabidopsis* seedlings (see Chapter 3). Upon analysis of the potential AtMC9 substrates in roots we found proteins that were targeted in both tissues (Tables S1, S2 and S3). Moreover, some of these proteins were cleaved at the same site by AtMC9 in the root or seedling N-terminomes either *in vitro* or *in vivo* (Table S5). This observation suggests that although AtMC9 functionalities seem to be stirred by the tissue or developmental stage, certain proteins are consistently targeted and particular focus should be placed on them.

In this study we showed that metacaspases are involved in cell separation in roots during lateral root development. By exploring the substrate inventory of AtMC9 we propose that regulation cell separation in the roots could be mediated by metacaspase activity on extracellular proteins with possible implication in cell wall remodelling.

MATERIALS AND METHODS

Transgenic lines

Knock-out lines

AtMC5 T-DNA insertion line (SAIL-284-C06) and *AtMC9* T-DNA line (GK-540H06) were crossed. Homozygous F2 plants were selected by segregation analysis and T-DNA insertions were PCR-validated with gene specific and T-DNA insert specific primers (Supplementary Table S4). Subsequently, absence of *AtMC5* and *AtMC9* transcripts from the single and double KO lines was determined by RT-PCR (Figure 2B), using primers that amplify a 978 bp cDNA fragment of *AtMC9*, a 402 bp *AtMC5* fragment. The *AtMC9* KO line contains a T-DNA insert at cDNA base pair position 483 downstream the start codon, present as an inverted repeat: LB-RB|RB-LB. The *AtMC5* KO line contains a T-DNA insert at cDNA base pair position 1111 downstream the start codon (Figure 2A).

Knock-down lines

Generation of multiple type-II metacaspase silenced lines was achieved by means of the artificial microRNA (amiRNA) technology (Schwab et al., 2006). The amiRNA construct hybridises on a sequence segment commonly present in *AtMC4*, 5, 6, 7 and 9. Since *AtMC8* was not targeted, the construct was introduced into an *AtMC8* knock-out line of *Landsberg erecta* background (He et al., 2008) (Supporting information Figure S4). The amiRNA construct was engineered with 4 oligonucleotide sequences (Supplementary Table S4) and two modified A and B oligonucleotides compatible with the GatewayTM for recombination to the entry vector pDONR221. Subsequently, the construct was fused to the CAMV 35S promoter by LR reaction to the destination vector pB7WG2 (Karimi et al., 2002). p35s:amiRNA fusion construct was transformed into *Agrobacterium tumefaciens* strain C58C1RifR[pMP90] and transgenic *Arabidopsis thaliana* (L.) Ler was obtained via floral dip transformation (Clough and Bent, 1998) and subsequent Basta selection. Independent T3 homozygous lines containing a single locus were selected for further analysis.

Reporter lines

The promoter region of metacaspase genes was cloned into the binary vector pBGWFS7 (Karimi et al., 2002) giving rise to *pAtMC*:GFP:GUS fusion reporter lines. The genomic region used to study promoter activity started at 1500 bp and ended 1 bp upstream the start codon of each metacaspase gene. Forward and reverse PCR primers used for amplification of each promoter region are listed in Supporting information Table S4. Extension to complete GatewayTM attB1 and attB2 cloning sites was obtained by a second PCR reaction. Constructs were introduced into *Arabidopsis thaliana* as previously described by *A. tumefaciens* mediated floral deep transformation.

Tissue expression analysis

GUS staining of 7 days-old plants started with a 10 min incubation in 90% (v/v) acetone at RT. After washing in phosphate buffer pH 7.4, the material was incubated in 1 mg/ml 5-bromo-4-chloro-3-indolyl-D-glucuronide, 2 mM ferricyanide, and 0.5 mM ferrocyanide in 100 mM phosphate buffer, pH 7.4 at 37°C in the dark for 4 h to overnight. The material was cleared with 85% lactate solution and observed under light microscopy. Fluorescence microscopy was performed with a confocal microscope 100M and software package LSM 510 version 3.2 (Zeiss, Jena, Germany). Excitation was done using the 488 nm line of an argon laser. Emission fluorescence for GFP was captured via a 500-550 nm band-pass filter, and for propidium iodide via a 585 nm long-pass filter.

Preparation and purification of native recombinant metacaspases

The molecular cloning of metacaspase open reading frames and the preparation of recombinant AtMC4 and AtMC9 was done as described in Vercammen et al., 2004. The AtMC5 containing bacterial expression vector pDEST-17 was transformed into the *E. coli* strain BL21 (DE3) and production was induced for 3 h with 0.2 mM isopropyl 1-thio-D-galactopyranoside. Cultures were harvested, re-suspended in extraction buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 8 M urea) and sonicated. Cellular debris was removed by centrifugation, and the supernatant was mixed with TALON cobalt affinity resin (BD Biosciences) and incubated overnight at 4°C. Beads were washed eight times in the same buffer, while stepwise reducing the urea concentration from 8 to 0 M, loaded on a column, and washed twice again (50 mM HEPES, pH 7.5, 300 mM NaCl). Bound protein was eluted with extraction buffer supplemented with 150 mM imidazole.

Screening of AtMC9 inhibitors

High-throughput screen of 10,000 compounds comprising the DIVERSet™ compound library of Chembridge, was carried out in 384-well plates (Genetix X7001) and the reactions were performed in 50 µl final volume. 5 nM recombinant AtMC9, pre-activated for 15 min in 50 mM MES pH 5.5, 150 mM NaCl, 10% sucrose, 0.1% CHAPS and 1 mM DTT, was added to 10 µM of each compound in 10% DMSO. The compounds and AtMC9 were co-incubated for 15 min at room temperature and then 10 µM Ac-VRPR-AMC was added to the assay mixture. As negative control, 10% DMSO was used, while 2.3 pmol of Ac-VRPR-FMK was used a positive control (90% inhibition). The fluorescence units measured for 10% DMSO are considered as 100% activity. Time-dependent release of free amido-4-methylcoumarin (AMC) was measured on a FLUOstar OPTIMA reader (BMG Labtechnologies, Offenburg, Germany) for 45 cycles of 1 min. Activity was expressed as the rate of fluorescence release per minute in each well and the assay was performed twice. A third screen was done in small scale (96-well plates) and involved 177 selected compounds with at least 75% inhibitory effect towards AtMC9 in both previous screens. During the small-scale screen, besides AtMC9, the compounds were tested against AtMC4, and AtMC5 activities. AtMC4 and AtMC5 were used in concentrations that provide a signal similar in magnitude as AtMC9. For the AtMC4 assay, 17 nM enzyme was pre-activated at room temperature for 15 min in 50 mM HEPES (pH 7.5), 150 mM NaCl, 10% glycerol, 100 mM CaCl₂ and 10 mM DTT, while for the AtMC5 assay, 72.5 nM enzyme was pre-activated at room temperature for 15 min in 10 mM HEPES (pH 7.5), 10 mM CaCl₂ and 10 mM DTT. In each case, 10 µM compound was co-incubated with the enzymes for 15 min at room temperature and then 10 µM of the equivalent AMC substrate was added to the assay mixture.

RNA extraction, reverse-transcription PCR and real-time quantitative PCR

Total RNA was extracted from leaves as described in Bentsink et al., 2006. One µg of RNA was reverse-transcribed using the iScript™ cDNA Synthesis Kit (Bio-Rad). Real-time quantitative PCR was performed in a final volume of 5 µl including 10% (v/v) of 6-fold diluted cDNA as template for PCR, 200 nM of each primer and 50% (v/v) SYBR green mix (Invitrogen) using a LightCycler 480 II instrument (Roche Applied Science), according to the manufacturer's instructions. The genes CBP20 (At5g44200) and ARP7 (At3g60830) were used as internal references for normalization of transcript levels. Metacaspase gene-specific primer sequences used for real-time quantitative PCR and reverse-transcription PCR are listed in Supplementary Table S4.

Treatments and growth conditions

Seeds were grown on vertical plates containing half strength Murashige and Skoog (MS) salts, 1% (w/v) sucrose, 0.8% (w/v) plant tissue culture agar (LabM, Bury, UK), 0.5 mg/l nicotinic acid, 0.5 mg/l pyridoxin and 1 mg/l thiamin. The seeds were stratified at 4°C for 2 nights in the dark and then left to germinate in continuous light (80-100 µmol m⁻²s⁻¹) at 21°C. Three days after germination, seedlings were transferred to MS agar plates containing 50 µM compounds and then 5 days later their root growth and phenotypical appearance was monitored. Staining of lateral root primordia and developmental staging was performed 10 days after germination as described in Malamy et al., 2002. Auxin effect on metacaspase transcript and activity levels was studied on 7 days-old plants after they were transferred to media supplemented with 10 µM NAA (auxin analog 1-naphthaleneacetic acid) for 24 hours.

Plant material and total proteome extraction

Seeds of *Arabidopsis* Columbia (Col-0), AtMC9 T-DNA insertion line (GK-506H04-019739) and AtMC9 over-expression (Vercammen et al., 2006) were overnight gas sterilized with HCl and NaOCl, then sowed on half strength Murashige and Skoog (MS) media and stratified at 4°C in the dark. After 3 days, seeds were transferred to 21°C with a 16-hour light/8-h dark photoperiod with light intensity of 80-100 µmol m⁻²s⁻¹ and left to grow for 3 weeks. Roots were harvested, frozen in liquid nitrogen and ground into a fine powder using a mortar and a pestle. To achieve a total protein content of approximately 4 mg, 0.5 g of ground tissue was defrosted in 1 ml buffer of 1% (w/v) CHAPS, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA and 10% glycerol in PBS at pH 7.5, further containing the suggested amount of protease inhibitors mixture according to the manufacturer's instructions (Roche). The sample was centrifuged at 16,000 g for 10 min at 4°C and guanidinium hydrochloride was added to the cleared supernatant to reach a final concentration of 4 M.

For the *in vitro* COFRADIC analysis, protein extracts were prepared in the AtMC9 optimal activity buffer; 50 mM MES (pH 5.5), 150 mM NaCl, 10% (w/v) sucrose, 0.1% (w/v) CHAPS and 10 mM DTT. Four mg of total root protein extract was incubated at 30°C with recombinant AtMC9 (Vercammen et al., 2004) for a total duration of one hour; here, every 20 min 10 µg of fresh protease (855 units) was pre-activated in the same optimal activity buffer at room temperature for 15 min and then added to the protein extract (enzyme activity is 57 U/µl and one unit corresponds to the enzyme activity that catalyzes the hydrolysis of 1 µmol VRPR-AMC per min at 30°C). This was repeated twice. The same procedure was used for the control sample, but without adding rAtMC9. Reactions were terminated by adding guanidinium hydrochloride to a final concentration of 4 M. Protein concentration measurements were performed using the DC protein assay of BIORAD and protein extracts were further modified for N-terminal COFRADIC analysis as described in Staes et al., 2011.

In silico analyses

Expression data were derived from the Affymetrix GeneChip arrays and processed by Genevestigator (Hruz et al., 2008).

The spatiotemporal expression of genes in *Arabidopsis* roots (Brady et al., 2007) was assessed through the eFP browser (Winter et al., 2007).

Icelogo (Colaert et al., 2009) analyses were performed via the web application (<http://iomics.ugent.be/>

icelogoserver/logo.html) using the *Arabidopsis thaliana* Swiss-Prot proteome as reference.

Hierarchical clustering of the small-molecule metacaspase inhibitors was done via the open access software ChemMine (<http://chemmine.ucr.edu/>) based on structural and physicochemical similarities of the compounds using the Tanimoto similarity coefficient (Backman et al., 2011).

N-terminal COFRADIC, peptide identification and quantification

Here, the used methods are largely the same as described in chapter 3 of the thesis and they only differ in a few points. Briefly, in all three analyses, α and ϵ amines in the KO proteomes were $^{13}\text{C}_4$ -butyrylated, while amines of WT, OE or rAtMC9 treated-KO proteomes were $^{12}\text{C}_4$ -butyrylated. MS/MS-spectra were searched using the MASCOT algorithm against the TAIR protein database release 10. The following search parameters were set. Mass tolerance on peptide precursor ions was set to 10 ppm (with MASCOT C13 option set to 1) and that of fragment ions was set to 0.5 Da. Allowed peptide charges were 1+, 2+, 3+ and instrument setting was put on ESI-TRAP. Endoproteinase semi-Arg C/P was used, allowing for 1 missed cleavage. Two types of MASCOT searches were performed that differed in the set of variable and fixed modifications. The first type of searches included the variable modifications of pyro-glutamate formation of N-terminal glutamine and the fixed modifications of methionine oxidation to its sulfoxide derivative, S-carbamidomethylation of cysteine and butyrylation ($^{12}\text{C}_4$ or $^{13}\text{C}_4$) of lysine and peptide N-termini. The second type of searches differed in that acetylation of peptide N-termini was set as a variable modification. Only peptides with a score higher than the MASCOT identity threshold score set at 99% confidence were withheld. Finally, Mascot Distiller (version 2.4) was used for peptide quantification.

ACKNOWLEDGEMENTS

Inhibitor Ac-VRPR-FMK was kindly provided by Nicolas Fasel (University of Lausanne, Switzerland), *AtMC8* T-DNA insertion line (JII Gene Trap line GT_3_12679) by Patrick Gallois (University of Manchester, UK) and *AtMC5* T-DNA insertion line (SAIL-284-C06) by Nuria Sanchez Coll (Centre for Research in Agricultural Genomics - CRAG, Barcelona, Spain). We acknowledge the contribution of Boris Parizot in the *in silico* root spatiotemporal expression analysis of genes. LT is indebted to VIB International PhD Program for a predoctoral fellowship. This work was supported by Research Foundation-Flanders (grant number G.0038.09N) and Ghent University (Multidisciplinary Research Partnership “Biotechnology for a Sustainable Economy” project).

Supporting information

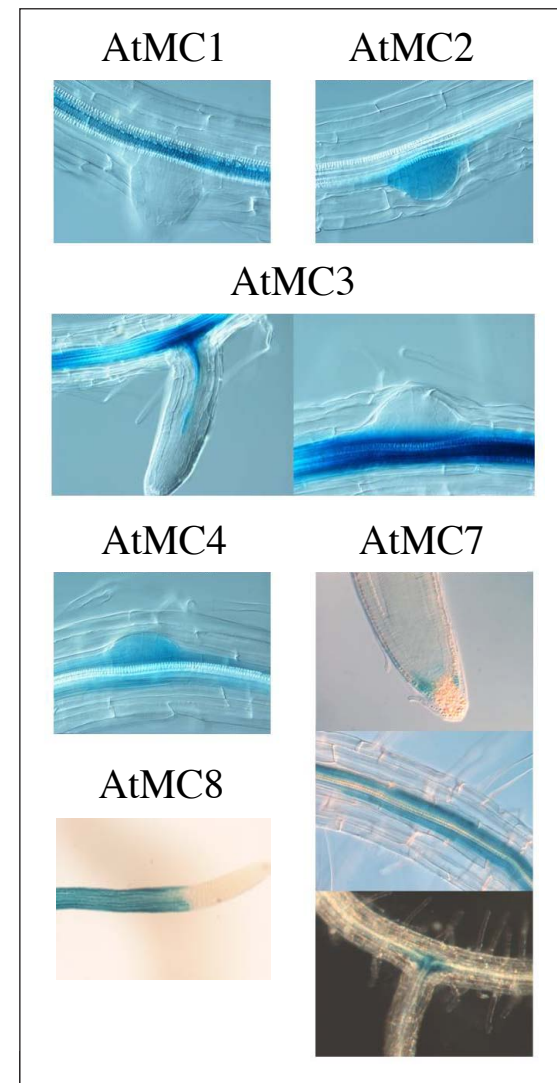


Figure S1. Expression patterns of metacaspase genes in *Arabidopsis* roots. AtMC promoter fusions to GUS and GFP (pAtMC:GFP:GUS). *AtMC1*, *AtMC3*, *AtMC4*, *AtMC7* and *AtMC9* are expressed in the vasculature tissue of the primary root, with *AtMC3* expression extending in the vasculature of lateral roots. Next to vasculature expression, *AtMC4* was also expressed in lateral root primordia, whereas *AtMC2* expression was restricted to lateral root primordia. *AtMC8* expression was restricted to the epidermis at the root elongation zone.

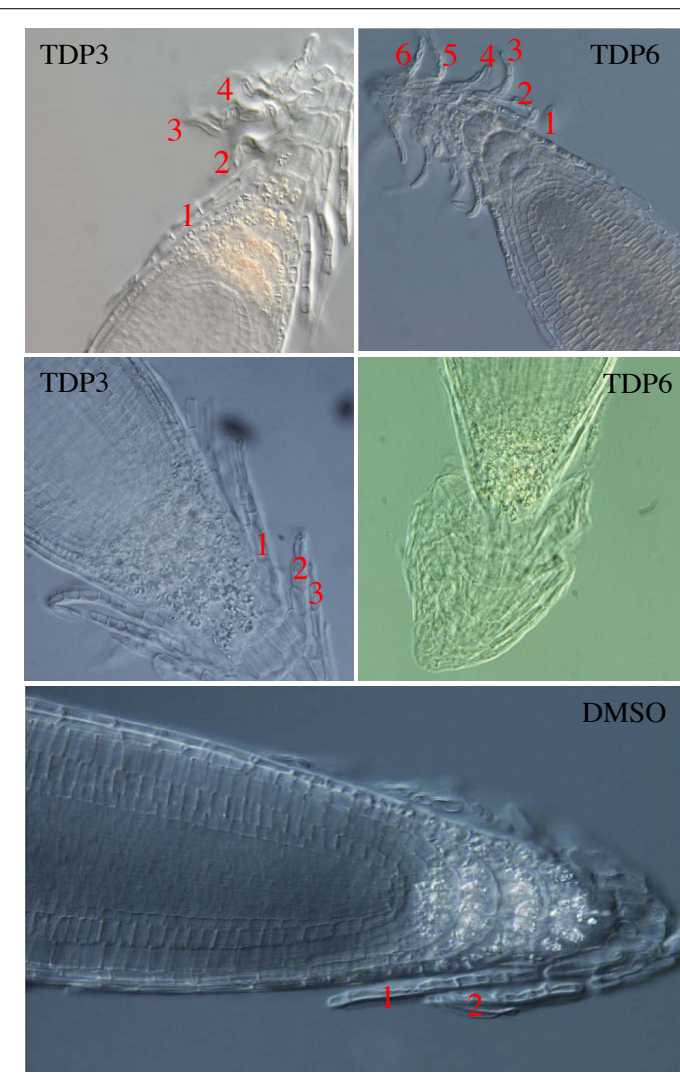


Figure S2. The effect of metacaspase inhibitors TDP3 and TDP6 on root cap. Numbers of clearly visible root cap cell tiers in untreated (DMSO control) and treated seedlings are depicted in red.

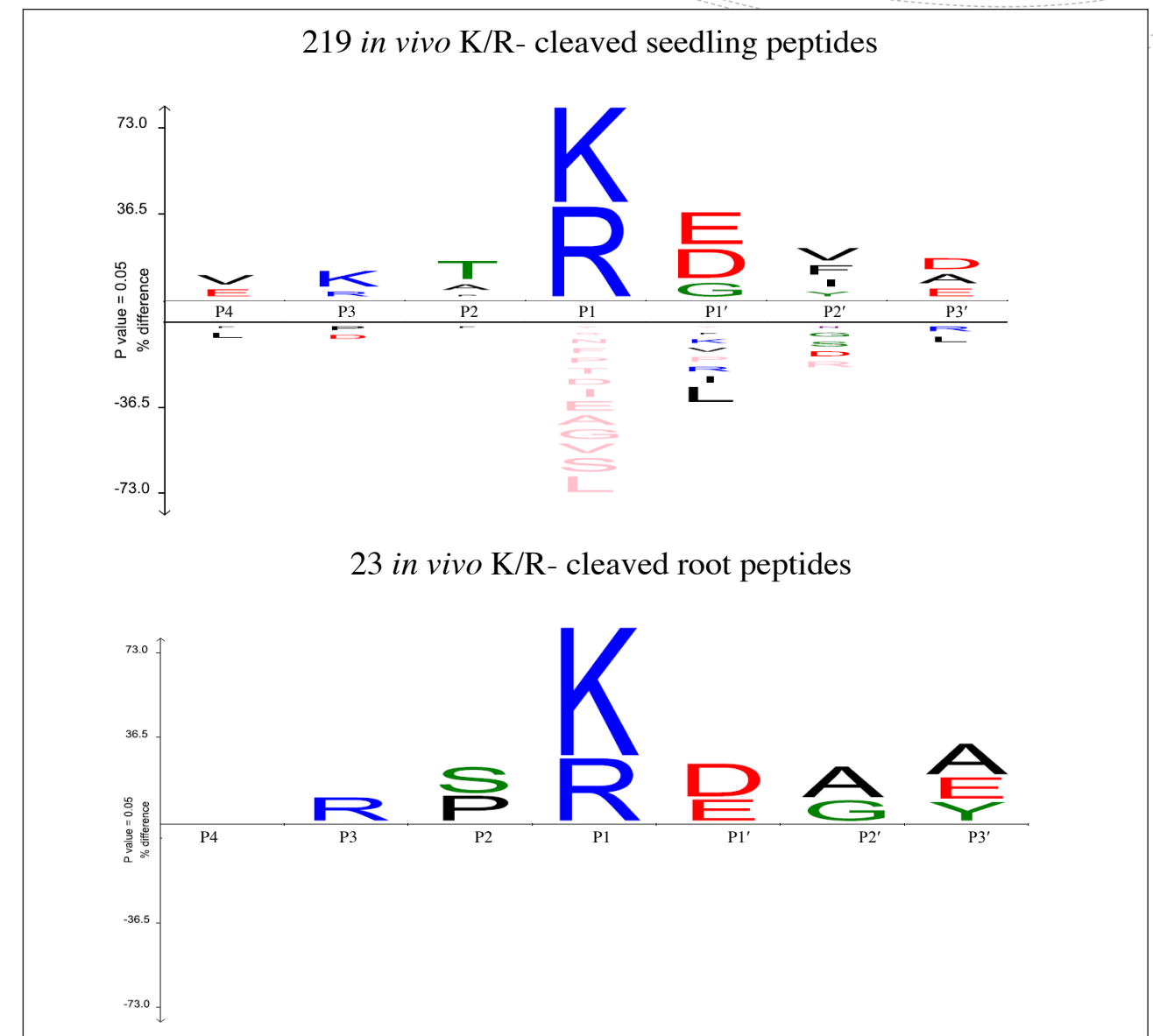


Figure S3. Native AtMC9 specificity consensus deduced from the occurrence of amino acids at the P4-P3' substrate positions after Icelogo analysis of the cleaved seedling or root peptide sequences (Colaert et al., 2009).

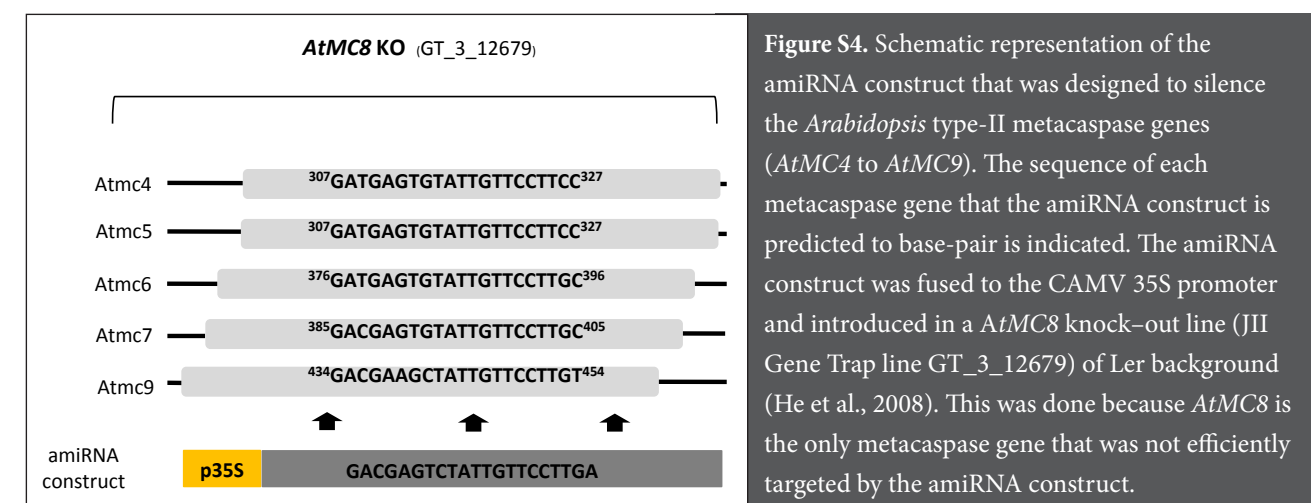


Figure S4. Schematic representation of the amiRNA construct that was designed to silence the *Arabidopsis* type-II metacaspase genes (*AtMC4* to *AtMC9*). The sequence of each metacaspase gene that the amiRNA construct is predicted to base-pair is indicated. The amiRNA construct was fused to the CAMV 35S promoter and introduced in a *AtMC8* knock-out line (JII Gene Trap line GT_3_12679) of Ler background (He et al., 2008). This was done because *AtMC8* is the only metacaspase gene that was not efficiently targeted by the amiRNA construct.

Table S1. List of endogenous AtMC9 substrates identified in the WT proteome. Peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the proteome of active protease (WT). In the upper part of the table, proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the middle part, peptides cleaved at other residues are listed. The bottom part of the table contains peptides that were uniquely found in the KO proteome and are indicative of indirect AtMC9 proteolysis. Amino acid position of the P1' site is shown

Accession	protein cleaved in seedlings?	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score	threshold	Confidence	E value	Mass error (ppm)	# of spectra	Subcellular localization SUBA MS data
Proteolysis of Arg or Lys															
AT1G12000	✓	861.956	2	4.74	3.53	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofructokinase family protein	57	45	0.01	6.31E-4	-1.162	1	cytosol, plasma membrane
AT1G47128		933.104	3	3.11	2.78	FIHK(215)NGG	NGGIDTDKDYPIKGVGDTCDQIR	Cysteine protease RD21	46	41	0.01	3.16E-3	-0.072	1	extracellular, plasma membrane, plastid, vacuole
AT1G47250		657.867	2	28.03	6.73	IGSR(165)SQA	SQAAKTYLER	20S proteasome alpha subunit F2	52	45	0.01	2.00E-3	0.533	1	cytosol, plasma membrane
AT2G07698	✓	660.843	2	2.55	2.42	EERR(277)AAE	AAELTNLFESR	ATPase, F1 complex, alpha subunit protein	53	45	0.01	1.58E-3	-0.076	1	mitochondrion, nucleus, plasma membrane, plastid, vacuole
AT3G20370		749.881	2	2.60	2.45	TLLK(254)NSY	NSYLSEVFSIGGR	TRAF-like family protein	48	44	0.01	3.98E-3	-1.002	1	extracellular, vacuole
AT3G45710		523.262	2	5.31	3.74	VAHR(15)SSD	SSDPSEKR	Major facilitator superfamily protein	61	44	0.01	2.00E-4	0.287	2	plasma membrane*
AT3G53020		589.024	3	3.04	2.73	NFPK(147)AAA	AAAASKGPKVGGGGGKR	Ribosomal protein L24e family protein	91	43	0.01	1.58E-7	-0.286	2	cytosol, nucleus, plasma membrane
Proteolysis of other amino acids															
AT1G01320	✓	439.693	2	N/A	N/A	SIST(1550)TDS	TDSGDASR	Tetratricopeptide repeat (TPR)-like superfamily protein	40	38	0.01	6.31E-3	7.522	1	nucleus*
AT1G12240		892.413	2	3.32	2.90	VVAE(505)AEE	AEEFSCSEKSGGSTVR	Glycosyl hydrolases family 32 protein	99	41	0.01	1.58E-8	-1.627	2	vacuole
AT1G19870		571.760	2	3.09	2.76	GRRT(706)SFG	SFGYDQEAR	IQ-domain 32	49	40	0.01	1.26E-3	-0.438	1	plasma membrane, plastid, vacuole
AT1G19870		580.817	2	4.23	3.33	ATPA(649)SQA	SQASSSVKAR	IQ-domain 32	78	45	0.01	5.01E-6	0.000	3	plasma membrane, plastid, vacuole
AT1G20220		757.298	2	3.48	2.98	GYVN(214)NEY	NEYNDGGMEQDR	Alba DNA/RNA-binding protein	61	29	0.01	6.31E-6	-1.190	1	cytosol, extracellular
AT1G20220		549.748	2	2.58	2.44	NGPA(175)NVE	NVEYDDGGR	Alba DNA/RNA-binding protein	54	40	0.01	3.98E-4	0.000	2	cytosol, extracellular
AT1G29250		879.465	2	4.97	3.62	EGVN(10)NMN	NMNLAVDTQKKNR	Alba DNA/RNA-binding protein	63	45	0.01	1.58E-4	-1.025	1	cytosol, nucleus, plasma membrane
AT1G48610		770.914	2	4.56	3.46	PQTD(33)ATG	ATGVSATDTASQKR	AT hook motif-containing protein	76	45	0.01	7.94E-6	-1.175	2	plasma membrane
AT1G62750		493.269	2	2.43	2.33	SVFA(82)AAE	AAEAERAKR	Translation elongation factor EFG/EF2 protein	55	45	0.01	1.00E-3	0.406	2	mitochondrion, plastid
AT1G69250	✓	838.821	3	3.44	2.96	AAPF(233)QVK	QVKSPVQAPVQPKYVGGQPR	Nuclear transport factor 2 (NTF2) family protein	70	41	0.01	1.26E-5	-1.074	1	cytosol*
AT1G75510		837.959	2	3.13	2.79	PMPG(162)MLG	MLGLVSSNSKEKR	Transcription initiation factor IIF, beta subunit	81	45	0.01	2.51E-6	-0.119	1	mitochondrion
AT1G76880		897.42	3	3.28	2.87	AVVS(384)TLD	TLDITKTDNGGDQNMTPAASASSR	Duplicated homeodomain-like superfamily protein	103	40	0.01	5.01E-9	-0.261	2	nucleus*
AT1G78580		647.300	2	4.39	3.40	SRPS(826)SDS	SDSGAKSSSGDR	trehalose-6-phosphate synthase	48	41	0.01	2.00E-3	-0.309	1	plasma membrane
AT2G02790		619.305	2	3.07	2.76	RKPT(287)GVS	GVSTTANSSSTR	IQ-domain 29	72	44	0.01	1.58E-5	-0.243	1	plasma membrane
AT2G15290		1014.220	3	2.56	2.43	TVAF(84)SYP	SYPTSPSSVPGDNEVDKAKLAQVAKR	translocon at inner membrane of chloroplasts 21	48	44	0.01	3.98E-3	-1.184	1	plasma membrane, plastid
AT2G17410		594.784	2	6.58	4.12	LEPM(602)NID	NIDNQASGSGR	ARID/BRIGHT DNA-binding domain-containing protein	77	42	0.01	3.16E-6	0.084	1	nucleus*
AT2G22170		862.887	2	2.52	2.40	VALA(26)DDE	DDEADCVYTFLLR	Lipase/lipoxygenase, PLAT/LH2 family protein	83	38	0.01	3.16E-7	1.337	5	extracellular, plasma membrane, plastid
AT2G30620		876.206	3	4.28	3.35	AAVS(209)KTK	KTKAVAAPKAKERPAKASR	winged-helix DNA-binding transcription factor family protein	36	33	0.01	5.01E-3	-0.495	1	nucleus
AT2G44730		526.298	2	2.47	2.36	RIPT(225)GVS	GVSIAQPGPR	Alcohol dehydrogenase transcription factor Myb	46	44	0.01	6.31E-3	0.095	1	nucleus*
AT2G46020		581.273	2	3.00	2.71	ATQL(1868)TVE	TVEDNAEASR	transcription regulatory protein SNF2, putative	56	42	0.01	3.98E-4	-0.345	1	plasma membrane
AT3G05730		1040.462	2	8.42	4.57	TAEA(30)NCD	NCDTYLGEVTVVYPCR	Encodes a defensin-like (DEFL) family protein.	77	38	0.01	1.26E-6	-0.385	1	extracellular*
AT3G07170		704.853	2	5.23	3.71	ARPS(120)MKS	MKSVATEITETR	Sterile alpha motif (SAM) domain-containing protein	73	44	0.01	1.26E-5	-0.497	2	nucleus*
AT3G09500		771.946	2	2.91	2.66	AQVL(59)TVS	TVSSQKQKSALR	Ribosomal L29 family protein	49	43	0.01	2.51E-3	0.195	1	cytosol, nucleus, plasma membrane
AT3G09630		785.472	2	3.40	2.94	NVML(333)KLN	KLNPYAKTAKR	Ribosomal protein L4/L1 family	62	41	0.01	7.94E-5	-0.319	1	cytosol, nucleus, plasma membrane, plastid, vacuole
AT3G10950		494.775	2	6.60	4.13	AYTM(72)NTA	NTASAVTVR	Zinc-binding ribosomal protein family protein	49	45	0.01	3.98E-3	0.000	1	cytosol, plasma membrane
AT3G11450		804.936	2	3.41	2.94	QKPD(519)SAK	SAKAFDSFLEKR	DnaJ domain *Myb-like DNA-binding domain	59	44	0.01	3.16E-4	-1.430	1	plastid*
AT3G20370		496.809	2	2.48	2.37	KVSA(74)SNA	SNAVKGLR	TRAF-like family protein	44	41	0.01	5.01E-3	0.303	1	extracellular, vacuole
AT3G21160		500.241	2	2.91	2.66	KDVQ(91)EDP	EDPVDAQR	alpha-mannosidase 2	52	42	0.01	1.00E-3	0.000	1	golgi
AT3G23390		809.472	3	5.08	3.66	HTLH(24)KVT	KVTQYKKGKDSLAAQGKR	Zinc-binding ribosomal protein family protein	47	41	0.01	2.51E-3	-0.330	1	plasma membrane
AT3G26720		726.872	2	3.66	3.07	GVTS(23)EYI	EYIEYNTIKPR	Glycosyl hydrolase family 38 protein	64	45	0.01	1.26E-4	-0.413	1	extracellular, vacuole
AT3G44100		659.311	2	3.44	2.96	ALHA(25)TSF	TSFTYCDKR	MD-2-related lipid recognition domain-containing protein	50	42	0.01	1.58E-3	-0.608	1	extracellular
AT3G57150		1134.045	2	6.55	4.12	AIIA(419)GAA	GAAAPEEIKADAENGEAGEAR	homologue of NAP57	124	43	0.01	7.94E-11	-1.721	1	nucleus, plasma membrane

between brackets in the P4-P3' column. Fold times of higher WT ion intensity in comparison to the KO ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides; Z; ion charge; Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions. *; subcellular localization predicted by WoLFPSORT.

AT4G12610		1001.982	2	2.47	2.36	AEPA(325)SAP	SAPASSSSAATGPVTEDEIR	RAP74 subunit of the general transcription factor TFIIIF	109	43	0.01	2.51E-9	-0.450	1	plastid
AT4G12770		905.891	2	3.81	3.14	GAPA(738)SQS	SQSGGFQVDVGETEER	Chaperone DnaJ-domain superfamily protein	109	37	0.01	6.31E-10	-1.658	1	cytosol, plasma membrane, plastid
AT4G16830		850.526	3	4.06	3.26	GQVS(38)SLP	SLPAKSAPKLPKPLPPAQAVR	Hyaluronan / mRNA binding family	66	34	0.01	6.31E-6	0.510	3	plastid*
AT4G17520	✓	1022.920	3	N/A	N/A	SKPL(27)KVV	KVVAVPQTAKSGKMPKPPPSQAVR	Hyaluronan / mRNA binding family	41	40	0.01	7.94E-3	0.245	1	cytosol, mitochondrion, peroxisome
AT4G26630		883.537	2	2.70	2.52	AVVA(610)AKS	AKSSPPEKITQKR	DEK domain-containing chromatin associated protein	82	40	0.01	6.31E-7	-0.963	3	cytosol
AT4G31700	✓	793.426	2	4.43	3.41	AKAN(208)SDA	SDAADYQKLLASR	ribosomal protein S6	67	45	0.01	6.31E-5	-0.571	2	cytosol, mitochondrion, nucleus, plasma membrane, plastid, vacuole
AT4G38580		896.487	2	2.92	2.66	AHPY(104)AAG	AAGVYDKKAPSGYVR	farnesylated protein 6	70	44	0.01	2.51E-5	-1.005	1	plasma membrane
AT4G39680	✓	558.288	2	2.46	2.36	AQIT(414)NSA	NSATPTTTPR	SAP domain-containing protein	52	44	0.01	1.58E-3	-0.269	1	plastid
AT5G11900		943.516	2	N/A	N/A	GAPS(79)SAQ	SAQTGGTSSKKEEVKR	Translation initiation factor SUI1 family protein	91	45	0.01	2.51E-7	-0.743	1	cytosol*
AT5G14200		538.327	2	3.18	2.82	RIRC(38)AAA	AAASPGKKR	isopropylmalate dehydrogenase 1	75	42	0.01	5.01E-6	0.091	6	plastid
AT5G16130	✓	721.405	2	2.57	2.44	AMLE(129)DVA	DVAFPAEIVGKR	Ribosomal protein S7e family protein	49	45	0.01	3.98E-3	1.874	1	cytosol, nucleus, plasma membrane
AT5G27395		702.41	3	2.50	2.38	EKRS(98)ATT	ATTQVKAPPQLQKTGAVR	Mitochondrial inner membrane translocase complex	45	41	0.01	3.98E-3	0.166	1	no data
AT5G41290		1038.932	2	6.10	3.99	QAIS(23)VSD	VSDPDDMETFCMKSSR	Receptor-like protein kinase-related family protein	52	33	0.01	1.26E-4	-0.771	1	extracellular, plasma membrane
AT5G49930	✓	736.361	2	3.76	3.12	GNTA(823)ADG	ADGNTGQEKQQR	zinc knuckle (CCHC-type) family protein	57	43	0.01	3.98E-4	-0.340	2	cytosol*
AT5G66930		852.412	2	4.01	3.23	KSHH(131)SKL	SKLVMDPGEASEER	unknown protein	67	43	0.01	3.98E-5	-0.529	1	nucleus*
Singletons in the KO proteome															
AT1G05240		738.444	2	N/A	N/A	KLGF(307)VQI	VQILTGNKEIR	Peroxidase superfamily protein	80	41	0.01	1.26E-6	-0.746	1	extracellular, plasma membrane
AT2G12400		911.992	2	N/A	N/A	ERVI(56)AEE	AEESGENSSLILAAKR	unknown protein	84	45	0.01	1.26E-6	-0.165	1	plasma membrane, plastid
AT2G31390		816.119	3	N/A	N/A	FANA(298)CGA	CGAITTTKKGAIPLPSDAEVR	pfkB-like carbohydrate kinase family protein	54	46	0.01	1.58E-3	-8.342	1	cytosol, plasma membrane
AT2G36530		744.063	3	N/A	N/A	MITTE(318)CGT	CGTEVQIVGDDLLVTNPKR	Enolase	68	48	0.01	1.00E-4	-8.299	1	cytosol, extracellular, peroxisome, plasma membrane, plastid, vacuole
AT2G38750		841.429	2	N/A	N/A	EAVA(184)SSG	SSGEEAVEKDEVVR	annexin 4	62	44	0.01	1.58E-4	-0.238	1	plasma membrane
AT3G01290		846.353	2	N/A	N/A	DVLD(236)MVM	MVMMTQYFDTMR	SPFH/Band 7/PHB domain-containing membrane	38	34	0.01	3.98E-3	-0.473	1	mitochondrion, plasma membrane, vacuole
AT3G09260		687.359	2	N/A	N/A	DIQL(102)MKN	MKNLNTDAFR	Glycosyl hydrolase superfamily protein	65	45	0.01	1.00E-4	-0.364	4	nucleus, peroxisome, plasma membrane, plastid, vacuole
AT3G15950		828.413	2	N/A	N/A	HTVK(291)DDV	DDVDDKEQDAKR	DNA topoisomerase-related	48	44	0.01	3.98E-3	-1.632	1	extracellular, peroxisome, plasma membrane, plastid
AT3G49120		764.884	2	N/A	N/A	TSFR(93)TEK	TEKDAFGNANSAR	peroxidase CB	67	44	0.01	5.01E-5	-0.065	2	extracellular, vacuole
AT4G11650		846.405	2	N/A	N/A	HRIL(143)CTA	CTADINGQCPNVLR	osmotin 34	78	43	0.01	3.16E-6	-0.177	1	extracellular
AT4G11650		710.795	2	N/A	N/A	CCTN(178)GQG	GQGSCSDTEYSR	osmotin 34	39	35	0.01	3.98E-3	0.141	1	extracellular
AT4G11650		1231.478	2	N/A	N/A	TVFQ(170)TNQ	TNQCCTNGOGSCSDTEYSR	osmotin 34	93	26	0.01	2.00E-9	-0.731	2	extracellular
AT4G11650		534.808	2	N/A	N/A	WRLD(57)VAA	VAAGTKMAR	osmotin 34	66	44	0.01	6.31E-5	-0.094	1	extracellular
AT4G29270	✓	931.489	2	N/A	N/A	RGQN(202)DQG	DQGKTATQYKSEQR	HAD superfamily, subfamily IIIB acid phosphatase	89	45	0.01	3.98E-7	-1.128	2	extracellular
AT4G30170		754.381	2	N/A	N/A	TVVK(105)AKQ	AKQAVDSNPNCR	Peroxidase family protein	62	44	0.01	1.58E-4	-0.996	2	extracellular, plasma membrane
AT4G34180		782.966	2	N/A	N/A	SLDL(115)QVL	QVLNGPALLVDVPR	Cyclase family protein	44	41	0.01	5.01E-3	0.064	1	extracellular
AT5G17820		778.924	2	N/A	N/A	FVRA(289)MVK	MVKMGAVDVLTR	Peroxidase superfamily protein	63	45	0.01	1.58E-4	-1.028	1	extracellular, plasma membrane
AT5G17820		806.887	2	N/A	N/A	RVGF(29)YSQ	YSQSCPQAEIVR	Peroxidase superfamily protein	45	43	0.01	6.31E-3	9.555	1	extracellular, plasma membrane
AT5G19250		791.910	2	N/A	N/A	DTAD(29)EED	EEDVLLTGINSYR	Glycoprotein membrane precursor GPI-anchored	52	44	0.01	1.58E-3	0.506	1	extracellular, plasma membrane
AT5G25460		522.307	2	N/A	N/A	TAMA(20)AKS	AKSTVSFR	Protein of unknown function, DUF642	74	44	0.01	1.00E-5	2.973	1	extracellular, plasma membrane
AT5G65020		896.420	2	N/A	N/A	KNLK(222)EES	EESDDNDYMKLLR	annexin 2	45	42	0.01	5.01E-3	-0.168	1	plasma membrane, plastid

Table S2. List of endogenous AtMC9 substrates identified in the OE proteome. Peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the proteome of active protease (OE). In the upper part of the table, proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the bottom half, peptides cleaved at other residues are listed. Amino acid position of the P1' site is shown between brackets in the P4-P3' column. Fold times of higher OE ion intensity in comparison to the KO ion counterpart is shown in the

column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions. *; subcellular localization predicted by WoLFPSORT.

Accession	protein cleaved in seedlings?	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score	threshold	Confidence	E value	Mass error (ppm)	# of spectra	Subcellular localization SUBAMS data
Proteolysis of Arg or Lys															
AT1G08200	✓	860.872	2	17.29	6.23	VSSK(329)EFY	EFYGEYDDSDKR	UDP-D-xylose/UDP-D-xylose synthase 2	72	37	0.01	3.16E-6	-1.570	1	cytosol, plasma membrane
AT1G22530		779.428	2	N/A	N/A	ETKK(272)EEK	EEKATASTQVQR	PATELLIN 2	50	45	0.01	3.16E-3	0.257	2	plasma membrane, plastid, vacuole
AT1G49040		931.427	2	N/A	N/A	PRPK(775)DVS	DVSVSDETQQPSEASGR	stomatal cytokinesis defective / SCD1 protein (SCD1)	72	41	0.01	7.94E-6	-2.579	1	plasma membrane
AT1G64370		649.805	2	7.63	4.00	NREK(10)DYY	DYYEVAQGQR	unknown protein	43	41	0.01	6.31E-3	-0.385	1	nucleus*
AT1G70850		892.994	2	4.33	2.46	AKER(235)IEA	IEAVDPEKNLITFR	MLP-like protein 34	55	45	0.01	1.00E-3	-0.336	1	cytosol*
AT2G20630		692.865	2	7.70	4.03	IESR(175)GGF	GGFVSNIPGDVPR	PP2C induced by AVRRPM1	53	44	0.01	1.26E-3	-0.217	1	cytosol*
AT2G33830		884.466	2	N/A	N/A	SSSK(38)TVA	TVAAVAGSPGTPPTPSGAR	Dormancy/auxin associated family protein	83	45	0.01	1.58E-6	-1.019	2	nucleus*
AT3G13300	✓	720.840	2	5.85	3.28	RRDR(773)DAN	DANIQDVNDPR	Transducin/WD40 repeat-like superfamily protein	52	42	0.01	1.00E-3	-0.417	1	cytosol, plasma membrane
AT4G15410	✓	878.431	2	5.44	3.08	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A	53	44	0.01	1.26E-3	-1.710	2	nucleus*
AT4G17520	✓	952.521	2	6.95	3.75	QTAK(37)SGK	SGKMPTKPPPSQAVR	Hyaluronan/mRNA binding family	63	44	0.01	1.26E-4	-1.734	1	cytosol, mitochondrion, peroxisome
AT4G30190		825.431	2	N/A	N/A	LQPK(889)EAV	EAVNIFPEKGSYR	H(+)-ATPase 2	48	45	0.01	5.01E-3	-0.910	1	plasma membrane, vacuole
AT4G31420		761.392	2	7.19	3.84	ITEK(314)SEN	SENTTTSKTLGSR	Zinc finger protein 622	50	45	0.01	3.16E-3	-0.197	1	nucleus*
AT5G23750		871.945	2	9.48	4.59	VVPK(69)EVE	EVEEEKKEGSVNR	Remorin family protein	48	44	0.01	3.98E-3	-0.746	1	nucleus*
AT5G46020	✓	738.383	2	5.78	3.24	LKAK(89)DLD	DLDASKTTELSR	unknown protein	48	45	0.01	5.01E-3	-0.271	1	cytosol, plasma membrane
AT5G46020	✓	727.368	2	8.20	4.20	VKKK(67)GAE	GAEAVIEVDNPNR	unknown protein	67	44	0.01	5.01E-5	-0.895	1	cytosol, plasma membrane
AT5G56000	✓	947.432	2	11.16	5.04	SSKK(604)TME	TMEINPENSIMDELRL	HEAT SHOCK PROTEIN 81.4	56	40	0.01	2.51E-4	0.423	1	mitochondrion, plasma membrane, plastid, vacuole
Proteolysis of other amino acids															
AT1G12310		890.450	2	5.38	3.05	QLKS(53)IIA	IIASENLSSPFDNR	Calcium-binding EF-hand family protein	78	44	0.01	3.98E-6	-1.124	1	cytosol, nucleus, plasma membrane
AT1G47128		686.88	2	15.52	5.94	KYLG(115)AKM	AKMEKKGER	Cysteine protease RD21	50	45	0.01	3.16E-3	-1.677	1	extracellular, plasma membrane, plastid, vacuole
AT1G51510		792.442	2	8.77	4.38	SAIA(46)GAN	GANGESAKTKKGR	RNA-binding (RRM/RBD/RNP motifs) family protein	69	45	0.01	3.98E-5	-0.569	1	nucleus
AT3G09080		609.820	2	N/A	N/A	VFVQ(968)SIS	SISELSTASPR	Transducin/WD40 repeat-like superfamily protein	48	45	0.01	5.01E-3	-817.411	2	nucleus*
AT3G13300	✓	782.346	3	5.10	2.91	PQLL(924)AMQ	AMQETMNQVMASQKEMQR	Transducin/WD40 repeat-like superfamily protein	39	37	0.01	6.31E-3	-0.299	1	cytosol, plasma membrane
AT3G52300		757.907	2	5.50	3.11	ELKE(113)AEQ	AEQKSLKESER	ATP synthase D chain, mitochondrial	60	45	0.01	3.16E-4	-1.057	1	cytosol, mitochondrion, nucleus, plasma membrane, plastid, vacuole
AT5G04200	✓	766.424	2	N/A	N/A	GCIN(32)DVL	DVLAMKETILSR	metacaspase 9	47	45	0.01	6.31E-3	-0.457	1	extracellular, cytosol, nucleus*

Table S3. List of endogenous AtMC9 substrates raised by the rAtMC9 driven proteolysis added to the KO proteome. Peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the treated proteome. In the upper half of the table proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the bottom half, peptides cleaved at other residues are listed. Amino acid position of the P1' site is shown between brackets in the P4-P3' column. Fold times of higher rAtMC9 treated ion intensity in comparison to the untreated ion counterpart

is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; # of spectra; number of identified spectra linked to the listed peptide. N/A; not applicable refers to fold change values of singleton ions. *; subcellular localization predicted by WoLFPSORT.

Accession	protein cleaved in seedlings?	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score	threshold	Confidence	E value	Mass error (ppm)	# of spectra	Subcellular localization SUBA (MS data)
Proteolysis of Arg or Lys															
AT1G04270	✓	738.889	2	6.50	4.79	KAKR(73)EAP	EAPQGEKPEPVR	cytosolic ribosomal protein S15	48	45	0.01	5.01E-3	-0.407	1	cytosol, mitochondrion, plasma membrane, plastid, vacuole
AT1G04690		493.780	2	N/A	N/A	LANR(242)SLV	SLVDDVLR	potassium channel beta subunit 1	61	45	0.01	2.51E-4	-0.609	1	plasma membrane
AT1G04820	✓	539.252	2	N/A	N/A	ISGK(242)EDA	EDAANNFAR	tubulin alpha-4 chain	46	41	0.01	3.16E-3	0.186	1	plasma membrane, plastid, vacuole
AT1G14610	✓	1158.512	2	N/A	N/A	SSKR(110)DAS	DASEENPEDFVDPETPLGER	valyl-tRNA synthetase / valine--tRNA ligase (VALRS)	90	38	0.01	6.31E-8	-1.080	1	cytosol
AT1G17210	✓	851.382	2	N/A	N/A	VVDR(514)DGD	DGDEVNDDSDAGPSKR	IAP-like protein 1	61	39	0.01	6.31E-5	-0.470	1	nucleus*
AT1G18070	✓	1020.037	2	5.35	4.17	AQEK(85)AAK	AAKEEAEDVAEANKKR	Translation elongation factor EF1A	125	45	0.01	1.00E-10	-0.981	1	cytosol, vacuole
AT1G20260		650.262	2	7.04	5.03	FFKR(224)DFE	DFEENGSMER	ATPase, V1 complex, subunit B protein	51	33	0.01	1.58E-4	-0.308	1	plastid, vacuole
AT1G20450		843.434	2	N/A	N/A	ETPK(17)VAT	VATEESSAPEIKER	Dehydrin family protein	49	45	0.01	3.98E-3	-0.831	1	cytosol, plasma membrane, plastid
AT1G22020		842.423	2	24.54	8.93	VDPR(295)TGY	TGYIDYDKLEEK	serine hydroxymethyltransferase 6	49	44	0.01	3.16E-3	-1.129	1	nucleus*
AT1G27450		789.833	2	N/A	N/A	AASR(58)DSE	DSEMATDQVQDPR	adenine phosphoribosyl transferase 1	56	35	0.01	7.94E-5	-0.380	1	cytosol, extracellular, plasma membrane, plastid
AT1G30230		1160.834	3	N/A	N/A	ADSK(99)DAA	DAAADEEDDDVDVLFGEETEEKAAEER	Glutathione S-transferase, C-terminal-like	92	33	0.01	1.26E-8	-1.150	1	cytosol, plasma membrane
AT1G49240	✓	739.860	2	4.52	3.65	MNOK(53)DAY	DAYVGDDEAQSQR	actin 8	73	43	0.01	1.00E-5	-0.135	3	cytosol, peroxisome, plasma membrane, plastid, vacuole
AT1G53310		1013.503	2	N/A	N/A	LYAK(200)DIT	DITPDDKQELDEALQR	phosphoenolpyruvate carboxylase 1	86	44	0.01	6.31E-7	-0.099	1	cytosol, plasma membrane
AT1G55860		769.833	2	N/A	N/A	AGER(232)GSS	GSSQTQAMPQDMR	ubiquitin-protein ligase 1	48	37	0.01	7.94E-4	-0.325	1	mitochondrion
AT1G56070	✓	616.285	2	N/A	N/A	WASK(67)EGP	EGPLAENMR	Ribosomal protein S5/Elongation factor G/III/V family protein	48	41	0.01	2.00E-3	-0.244	1	cytosol, extracellular, nucleus, plasma membrane, plastid, vacuole
AT1G56340	✓	687.867	2	6.95	4.99	DAEK(360)AAF	AAFDEAEKRR	calreticulin 1a	49	45	0.01	3.98E-3	-0.510	2	endoplasmic reticulum, mitochondrion, plasma membrane, plastid, vacuole
AT1G56340	✓	944.488	2	N/A	N/A	GKHK(356)DAE	DAEKAAFDEAEKRR	calreticulin 1a	75	45	0.01	1.00E-5	0.689	1	endoplasmic reticulum, mitochondrion, plasma membrane, plastid, vacuole
AT1G66410		854.902	2	9.99	6.13	KKMK(79)DTD	DTDSEELKEAFR	calmodulin 4	61	42	0.01	1.26E-4	-3.162	3	plasma membrane
AT1G75660		617.265	2	4.42	3.58	RSAR(124)DAS	DASDAAEEER	5'-3' exoribonuclease 3	69	37	0.01	6.31E-6	-0.406	1	cytosol*
AT1G76180		1277.632	2	10.21	6.20	EEIK(7)NVP	NVPEQVVPKATEESSAEVTDR	Dehydrin family protein	83	44	0.01	1.26E-6	-1.802	2	cytosol, plasma membrane, plastid, vacuole
AT1G76180		732.347	2	N/A	N/A	EVPK(16)VAT	VATEESSAEVTDR	Dehydrin family protein	88	42	0.01	2.51E-7	-0.342	8	cytosol, plasma membrane, plastid, vacuole
AT2G01250	✓	732.464	2	N/A	N/A	VESK(6)VVV	VVVPESVLKRR	Ribosomal protein L30/L7 family protein	51	37	0.01	3.98E-4	-0.273	1	cytosol, plasma membrane, plastid, vacuole
AT2G07360		648.765	2	N/A	N/A	SGSR(1122)SYE	SYESDDEEPR	SH3 domain-containing protein	37	34	0.01	5.01E-3	-0.618	1	plasma membrane
AT2G22125		938.901	2	N/A	N/A	IHPK(1001)EKE	EKEEDEEEATEENR	CELLULOSE SYNTHASE-INTERACTIVE PROTEIN 1	62	36	0.01	2.51E-5	-0.373	1	plasma membrane
AT2G22400	✓	615.807	2	N/A	N/A	TEDK(567)EAN	EANSSNAGGKR	S-adenosyl-L-methionine-dependent methyltransferase	81	44	0.01	2.00E-6	0.651	1	nucleus*
AT2G25970		676.266	2	6.60	4.83	QVTR(172)DMD	DMDADPNCATR	KH domain-containing protein	45	31	0.01	3.98E-4	-0.222	1	cytosol
AT2G30060		1208.041	2	N/A	N/A	RENR(14)DEE	DEEETGANEDDGTGAQVPIVR	Pleckstrin homology (PH) domain superfamily protein	103	39	0.01	3.98E-9	1.698	1	nucleus*
AT2G31610	✓	747.360	2	4.13	3.37	VLTR(28)ELA	ELAEDGYSQVEVR	Ribosomal protein S3 family protein	66	43	0.01	5.01E-5	-0.335	1	cytosol, mitochondrion, plasma membrane, plastid, vacuole
AT2G36060	✓	1119.963	2	N/A	N/A	RGEK(27)GIG	GIGDGTVSYGMDDGDDIYMR	MMS ZWEI homologue 3	82	32	0.01	1.00E-7	-0.492	2	cytosol, plasma membrane
AT2G39940		968.417	2	N/A	N/A	RIER(352)GAD	GADEQGMDEEGLVSQR	RNI-like superfamily protein	99	36	0.01	5.01E-9	-1.602	1	cytosol*
AT2G40510		919.956	2	4.16	3.39	AAIR(52)DVQ	DVQEASVYEGYTLPK	Ribosomal protein S26e family protein	94	44	0.01	1.00E-7	0.218	2	cytosol, plasma membrane, plastid
AT3G01390		998.488	2	N/A	N/A	EAEK(45)EIA	EIAEYKAQTEQDFQR	vacuolar membrane ATPase 10	81	43	0.01	1.58E-6	-1.404	1	cytosol, plasma membrane, plastid, vacuole
AT3G02080		472.246	2	N/A	N/A	SGQR(131)DLD	DLDQVAGR	Ribosomal protein S19e family protein	61	43	0.01	1.58E-4	0.424	1	cytosol, plasma membrane
AT3G06720		810.455	2	7.24	5.12	SIER(105)SPP	SPPIEEVISAGVVP	importin alpha isoform 1	63	44	0.01	1.26E-4	-0.988	1	cytosol, nucleus, plasma membrane
AT3G07880		466.719	2	N/A	N/A	SRTR(35)ADD	ADDDALSR	Immunoglobulin E-set superfamily protein	43	41	0.01	6.31E-3	0.537	1	nucleus*
AT3G09840	✓	564.319	2	13.44	7.05	NIEK(710)DIE	DIEKEKR	cell division cycle 48	48	44	0.01	3.98E-3	0.266	1	cytosol, plasma membrane
AT3G10650	✓	712.802	2	N/A	N/A	LRSK(173)AAD	AADSSTMNEEQR	ATNUP1, nucleoporin	48	36	0.01	6.31E-4	-0.281	1	plasma membrane
AT3G14940		1021.501	2	N/A	N/A	LYAK(200)DIT	DITPDDKQELDESLQR	phosphoenolpyruvate carboxylase 3	87	43	0.01	3.98E-7	-0.637	2	cytosol*
AT3G15790	✓	1002.493	2	N/A	N/A	VEKK(215)TVE	TVEASDEKKNSEAEATR	methyl-CPG-binding domain 11	84	44	0.01	1.00E-6	-0.549	1	nucleus*
AT3G15950		1157.557	2	5.56	4.30	ESFK(322)QLE	QLEDIADNKAEGDDESAKR	DNA topoisomerase-related	115	43	0.01	6.31E-10	-1.297	2	extracellular, peroxisome, plasma membrane, plastid
AT3G15950		1042.53	2	N/A	N/A	AATK(434)GLE	GLEELKNESEQAENKR	DNA topoisomerase-related	98	44	0.01	3.98E-8	-1.008	2	extracellular, peroxisome, plasma membrane, plastid
AT3G16640	✓	567.314	2	N/A	N/A	STQK(67)VVD	VVDIVDTFR	translationally controlled tumor protein	65	45	0.01	1.00E-4	0.353	1	cytosol, nucleus, plasma membrane, plastid, vacuole
AT3G23810	✓	950.463	2	N/A	N/A	SSGR(13)EYK	EYKVKDMSQADFGR	S-adenosyl-L-homocysteine (SAH) hydrolase 2	55	43	0.01	6.31E-4	-2.054	1	cytosol, plasma membrane, vacuole
AT3G52930	✓	1167.060	2	N/A	N/A	KVDK(107)GTV	GTVELAGTIDGETTQGLDGLGDR	Aldolase superfamily protein	130	43	0.01	2.00E-11	-1.158	5	cytosol, nucleus, plasma membrane, plastid, vacuole
AT3G53420		737.862	2	N/A	N/A	MAK(4)DVEA	DVEAVPGEFGFTR	plasma membrane intrinsic protein 2A	55	44	0.01	7.94E-4	-0.136	1	plasma membrane, plastid
AT3G54540		846.926	2	N/A	N/A	SKSK(460)GKT	GKTVEDEGPAPEAPR	general control non-repressible 4	70	44	0.01	2.51E-5	-0.591	1	cytosol*
AT3G54540		719.347	2	N/A	N/A	SKGK(462)TVD	TVDEEGPAPEAPR	general control non-repressible 4	57	43	0.01	3.98E-4	-0.418	1	cytosol*
AT3G61440		588.306	2	N/A	N/A	DLPK(43)DFP	DFPSTNAKR	cysteine synthase C1	49	45	0.01	3.98E-3	0.085	1	mitochondrion, plastid

AT4G19210	✓	693.847	2	N/A	N/A	VIDR(213)DVE	DVENLSGGELQR	RNAse I inhibitor protein 2	48	44	0.01	3.98E-3	-0.938	1	plasma membrane
AT4G19880	✓	770.350	2	12.42	6.81	SMAR(35)SAV	SAVDETSDSGAFQR	Glutathione S-transferase family protein	90	41	0.01	1.26E-7	-0.455	1	cytosol, mitochondrion
AT4G23670	✓	836.444	2	N/A	N/A	KEKR(74)EID	EIDDENKTLTKR	Polyketide cyclase/dehydrase and lipid transport protein	50	45	0.01	3.16E-3	-0.419	5	plasma membrane, plastid, vacuole
AT4G24220		899.865	2	N/A	N/A	WTNR(110)ESE	ESESENCEANGSMLR	NAD(P)-binding Rossmann-fold superfamily protein	73	32	0.01	7.94E-7	-0.779	1	cytosol
AT4G26300		558.780	2	4.77	3.82	LNSK(343)GLV	GLVEESEGAR	Arginyl-tRNA synthetase, class Ic	46	44	0.01	6.31E-3	-0.090	1	plastid
AT4G34050		966.968	2	N/A	N/A	EATK(11)ISS	TSSTNGEDQKQSQNLNR	S-adenosyl-L-methionine-dependent methyltransferase	57	43	0.01	3.98E-4	-1.191	1	cytosol*
AT4G36760	✓	703.791	2	N/A	N/A	EAAK(281)DME	DMEIDSDQDPR	aminopeptidase P1	61	34	0.01	2.00E-5	-0.854	1	cytosol
AT4G37910	✓	522.770	2	N/A	N/A	KTAR(76)VIE	VIENAEGSR	mitochondrial heat shock protein 70-1	60	44	0.01	2.51E-4	0.287	1	mitochondrion, plasma membrane
AT4G38440		920.854	2	N/A	N/A	AQPK(72)DYN	DYNDQEEEEAEER	uncharacterised protein	71	27	0.01	3.98E-7	-0.598	1	plasma membrane*
AT4G38600		766.860	2	N/A	N/A	TSGR(1144)QMN	QMNPSASGTSGAAAR	HEAT repeat *HECT-domain (ubiquitin-transferase)	77	42	0.01	3.16E-6	-0.653	1	plasma membrane
AT4G38630	✓	739.343	2	N/A	N/A	QKDK(258)DGD	DGDTASASQETVAR	regulatory particle non-ATPase 10	69	41	0.01	1.58E-5	-1.354	1	cytosol, plasma membrane
AT5G04200	✓	1169.056	2	N/A	N/A	HPFK(110)QDE	QDEAIVPCDFNLITDVFDR	metacaspase 9	86	42	0.01	3.98E-7	1.070	1	extracellular, cytosol, nucleus*
AT5G04420	✓	746.823	2	6.44	4.76	RRQR(490)SAS	SASDEEEDGTIVQR	Galactose oxidase/kelch repeat superfamily protein	57	37	0.01	1.00E-4	0.134	1	cytosol*
AT5G08450		854.854	2	N/A	N/A	ISEK(503)ESE	ESEDGCLGEGATER	uncharacterised protein	62	34	0.01	1.58E-5	-2.284	1	nucleus*
AT5G10360	✓	862.957	2	N/A	N/A	IVKK(119)GVS	GVSDLPLGLTDEKPR	Ribosomal protein S6e	53	45	0.01	1.58E-3	-0.174	2	cytosol, plasma membrane
AT5G11260	✓	546.252	2	N/A	N/A	TSGR(59)ESG	ESGSATGQER	Basic-leucine zipper (bZIP) transcription factor family protein	83	41	0.01	6.31E-7	0.183	2	nucleus*
AT5G16370		666.781	2	N/A	N/A	IADK(147)EEE	EEGGDADVADR	acyl activating enzyme 5	96	36	0.01	1.00E-8	-0.300	2	peroxisome
AT5G19990		769.824	2	9.96	6.12	GSAR(271)MES	MESGSGNGDSEVQR	regulatory particle triple-AATPase 6A	53	35	0.01	1.58E-4	-0.910	1	cytosol, plasma membrane
AT5G24710		619.746	2	12.08	6.72	RQSR(831)GDS	GDSDDIMDER	Transducin/WD40 repeat-like superfamily protein	42	32	0.01	1.00E-3	-0.323	2	plasma membrane
AT5G40450		1085.518	2	N/A	N/A	HQEK(1235)NAE	NAEPVEATQNLDDAEQISR	unknown protein	120	43	0.01	2.00E-10	-1.337	1	plasma membrane, plastid
AT5G40450		864.447	2	N/A	N/A	NISK(175)VCE	VCEEIPIKTDEVR	unknown protein	52	45	0.01	2.00E-3	0.984	1	plasma membrane, plastid
AT5G46290		537.247	2	N/A	N/A	LSQR(273)NDD	NDDPQTASR	3-ketoacyl-acyl carrier protein synthase I	49	40	0.01	1.26E-3	0.093	1	plasma membrane, plastid
AT5G47210	✓	487.725	2	N/A	N/A	KAQK(222)EAE	EAEAEAEAR	Hyaluronan / mRNA binding family	62	40	0.01	6.31E-5	0.000	3	mitochondrion, peroxisome, plasma membrane
AT5G47210	✓	748.041	3	N/A	N/A	EAKK(212)ELT	ELTAEKKAQKEAEAEAR	Hyaluronan / mRNA binding family	57	44	0.01	5.01E-4	0.134	1	mitochondrion, peroxisome, plasma membrane
AT5G57580		729.311	2	N/A	N/A	RAKR(15)NLD	NLDGNDDDQPER	Calmodulin-binding protein	39	37	0.01	6.31E-3	-1.098	1	nucleus*
AT5G66120	✓	822.429	2	N/A	N/A	NDQR(74)SIS	SISSPTVVEVDLGDGR	3-dehydroquinate synthase, putative	85	45	0.01	1.00E-6	-1.157	1	plastid
Proteolysis of other amino acids															
AT1G04820	✓	888.376	2	19.63	8.24	FVHW(408)YVG	YVGEEMEEGFSEAR	tubulin alpha-4 chain	86	35	0.01	7.94E-8	-0.733	1	plasma membrane, plastid, vacuole
AT1G09080		649.838	2	N/A	N/A	HNKH(72)VEI	VEIILANDQGNR	Heat shock protein 70 (Hsp 70) family protein	53	44	0.01	1.26E-3	0.231	1	plasma membrane
AT1G14610	✓	906.968	2	N/A	N/A	VYLE(1029)VDG	VDGAINTEAEQEKIR	valyl-tRNA synthetase / valine--tRNA ligase (VALRS)	53	45	0.01	1.58E-3	3.035	1	cytosol
AT1G35720		646.820	2	4.39	3.56	TSSN(104)QVL	QVLMVEACTR	annexin 1	57	44	0.01	5.01E-4	1.006	1	peroxisome, plasma membrane, plastid, vacuole
AT1G54270		682.323	2	N/A	N/A	IKMF(186)VLD	VLDEADEMLSR	eif4a-2	49	42	0.01	2.00E-3	1.174	1	plasma membrane, plastid, vacuole
AT1G79340	✓	723.360	2	4.25	3.45	SIQT(373)ILE	ILEETDGEISNR	metacaspase 4	52	44	0.01	1.58E-3	0.069	1	cytosol, plasma membrane
AT2G44160		544.302	2	N/A	N/A	NLGM(303)IDE	IDESKISR	methylene tetrahydrofolate reductase 2	48	45	0.01	5.01E-3	1.749	1	cytosol, plasma membrane
AT3G19760		584.245	2	N/A	N/A	LLIL(184)DES	DESDMLSR	eukaryotic initiation factor 4A-III	38	36	0.01	6.31E-3	0.429	1	nucleus
AT3G28710		780.397	2	N/A	N/A	NLMW(331)ISE	ISECVAQNQKSR	ATPase, V0/A0 complex, subunit C/D	52	45	0.01	2.00E-3	-0.706	1	plasma membrane, vacuole
AT3G62120		568.232	2	3.84	3.14	LAPW(475)CDE	CDEEEVER	Class II aaRS and biotin synthetases superfamily protein	38	34	0.01	3.98E-3	0.000	1	cytosol, plasma membrane
AT5G04200	✓	717.365	2	N/A	N/A	KKRL(12)AVL	AVLVGCNYPNTR	metacaspase 9	70	45	0.01	3.16E-5	-0.768	1	extracellular, cytosol, nucleus*
AT5G04200	✓	568.287	2	N/A	N/A	MDLM(230)DLL	DLEETMTAR	metacaspase 9	56	44	0.01	6.31E-4	-1.058	1	extracellular, cytosol, nucleus*
AT5G04200	✓	766.424	2	N/A	N/A	GCIN(32)DVL	DVLAMKETLSR	metacaspase 9	61	45	0.01	2.51E-4	-0.915	2	extracellular, cytosol, nucleus*
AT5G04200	✓	703.345	3	N/A	N/A	RVLN(281)ENE	ENEGAMKKNQVMMAR	metacaspase 9	75	43	0.01	6.31E-6	-0.237	2	extracellular, cytosol, nucleus*
AT5G04200	✓	988.513	2	N/A	N/A	NELH(28)GCI	GCINDVLAMKETLSR	metacaspase 9	80	45	0.01	3.16E-6	-0.253	2	extracellular, cytosol, nucleus*
AT5G04200	✓	526.235	2	N/A	N/A	AVLV(16)GCN	GCNYPNTR	metacaspase 9	42	40	0.01	6.31E-3	0.095	1	extracellular, cytosol, nucleus*
AT5G04200	✓	886.495	2	N/A	N/A	VSSN(169)ISP	ISPAIETTNTITSR	metacaspase 9	46	44	0.01	6.31E-3	-1.242	2	extracellular, cytosol, nucleus*
AT5G04200	✓	524.787	2	N/A	N/A	CDFN(121)LIT	LITDVFDR	metacaspase 9	60	45	0.01	3.16E-4	0.000	3	extracellular, cytosol, nucleus*
AT5G04200	✓	741.359	3	N/A	N/A	QRVL(280)NEN	NENEGAMKKNQVMMAR	metacaspase 9	47	43	0.01	3.98E-3	0.135	1	extracellular, cytosol, nucleus*
AT5G04200	✓	536.813	2	N/A	N/A	MDQ(4)QGM	QGMVKKR	metacaspase 9	47	44	0.01	5.01E-3	0.280	1	extracellular, cytosol, nucleus*
AT5G04200	✓	732.337	2	N/A	N/A	CLYC(313)SDQ	SDQNDATFLSQP	metacaspase 9	53	41	0.01	6.31E-4	-0.615	1	extracellular, cytosol, nucleus*
AT5G04200	✓	1024.145	3	N/A	N/A	GILM(249)SGC	SGCQADETSADVGVNGKAYGAFSNAIQR	metacaspase 9	101	40	0.01	7.94E-9	-1.010	4	extracellular, cytosol, nucleus*
AT5G04200	✓	987.033	2	N/A	N/A	SSVS(167)SNI	SNISPAIETTNTITSR	metacaspase 9	68	45	0.01	5.01E-5	-1.521	1	extracellular, cytosol, nucleus*
AT5G04200	✓	829.952	2	N/A	N/A	SSNI(170)SPA	SPAIETTNTITSR	metacaspase 9	47	45	0.01	6.31E-3	0.000	1	extracellular, cytosol, nucleus*
AT5G04200	✓	581.327	2	N/A	N/A	PAIE(175)TTN	TTNKITISR	metacaspase 9	50	44	0.01	2.51E-3	0.086	1	extracellular, cytosol, nucleus*
AT5G04200	✓	681.846	2	N/A	N/A	KRLA(13)VLV	VLVGCNYPNTR	metacaspase 9	47	44	0.01	5.01E-3	-0.514	1	extracellular, cytosol, nucleus*
AT5G08670	✓	695.846	2	3.82	3.13	ELGI(422)YPA	YPAVDPLDSTSR	ATP synthase alpha/beta family protein	50	44	0.01	2.51E-3	-0.432	1	mitochondrion, peroxisome, plasma membrane, plastid, vacuole
AT5G13780		696.275	2	N/A	N/A	EAKY(138)YAD	YADGEDAYDMR	Acyl-CoA N-acyltransferases (NAT) superfamily protein	40	31	0.01	1.26E-3	-0.503	1	plasma membrane
AT5G16760		857.912	2	N/A	N/A	SNLT(229)AQE	AQEDKNIEYGEDR	Inositol 1,3,4-trisphosphate 5/6-kinase family protein	53	43	0.01	1.00E-3	-1.342	1	cytosol

Table S4. Oligonucleotides used in this study.

Gene	Oligo	Sequence 5'-3'
RT-PCR		
AtMC9 AT5G04200	Forward	ATGGATCAACAAGGGATGGTCAA
	Reverse	TCAAGTTGAGAAAGGAACGTCCG
AtMC5 AT1G79330	Forward	GTGGGATATTGGGAATGATTGGGAA
	Reverse	CGGAGCATTCAAAAACATCACTGC
Actin AT3G12110	Forward	GGTTTGCTGGAGATGATGCT
	Reverse	CGATTAGCCTTTGGGTTAAGAGGCGC
T-DNA insert: GABI-LB	Reverse	CCCATTGGACGTGAATGTAGACAC
T-DNA insert: SAIL-RB	Forward	GCGCGGTGTATCATATGTTA
Real-time quantitative PCR		
AtMC4 AT1G79340	Forward	TCCTTGCACATGAATCTGA
	Reverse	AGTCTGAAATGATGTCATTCTGC
AtMC5 AT1G79330	Forward	CACTGAGCTGATTGATACTGACG
	Reverse	AGATTCAACAATGCCCTTCG
AtMC6 AT1G79320	Forward	TCACAGTGGCGGTCTCATAG
	Reverse	CTTCTCTTCGTGCTCTCTCT
AtMC7 AT1G79310	Forward	AGAGTCCACCGGAAACA
	Reverse	TGTTCCACTAAATCCCTGAAATC
AtMC9 AT5G04200	Forward	ACAGATCGGACCCCTCTCG
	Reverse	TCTTGTTCGTTGCTCAATAGCC
CBP20 At5g44200	Forward	GAAATAGGGTTCCTTTTGA
	Reverse	GCTTTGCTTCTCCTTGAAC
ARP7 At3g60830	Forward	ACTCTTCCTGATGGACAGGTG
	Reverse	CTCAACGATTCCATGCTCT
ami-RNA construct (Schwab et al., 2006)		
I	Forward	GATCAAGGAACAATAGACTCGTCTCTCTTTTGTATTCC
II	Reverse	GAGACGAGTCTATTGTTCTTGATCAAAGAGAATCAATGA
III	Forward	GAGAAGAGTCTATTGATCCTTGTTCACAGGTCGTGATATG
IV	Reverse	GAAACAAGGATCAATAGACTCTTCTTACATATATATCTCT
modified A	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCTCCCAACACACGCTCGGA
modified B	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGTCCCATGGCGATGCCTTAA
Promoter amplification		
AtMC1 AT1G02170	Forward	AAAAAGCAGGCTCCACCTTCTCTCAAATGTGAGTCAATGC
	Reverse	AGAAAGCTGGTCTATTATTCTCGGAAGGGAGGG
AtMC2 At4g25110	Forward	AAAAAGCAGGCTCCACCAGAAGTAGAGCAACATTGCT
	Reverse	AGAAAGCTGGTCTATTCTATAATAACCCAAAACCTC
AtMC3 AT5G64240	Forward	AAAAAGCAGGCTCCACCTCATATAGTATTTATAAGAGAAAA
	Reverse	AGAAAGCTGGTCTGATTGAGCTTTGTTTGGTTTTC
AtMC4 AT1G79340	Forward	AAAAAGCAGGCTCCACGATTTTCGACGATTTCAAATTTAA
	Reverse	AGAAAGCTGGTCTGAGAAAAATGATCTGATCGAT
AtMC5 AT1G79330	Forward	AAAAAGCAGGCTCCACGATGGCGGTCTCATAGATGA
	Reverse	AGAAAGCTGGTCTGAAAGATATCTCAAAACGCTCG
AtMC6 AT1G79320	Forward	AAAAAGCAGGCTCCACCGATGTCAAACAGATGAGACATC
	Reverse	AGAAAGCTGGTCTGTTGATATTTCAAACCTCTGCTT
AtMC7 AT1G79310	Forward	AAAAAGCAGGCTCCACCCACACAAGGACGCACCGATT
	Reverse	AGAAAGCTGGTCTCTTGAACCTTCAAGAGTTC
AtMC8 AT1G16420	Forward	AAAAAGCAGGCTCCACCCATCATCAGCAATTCTAGCCA
	Reverse	AGAAAGCTGGTCTGACGAATGAATGAATTATACCT
AtMC9 AT5G04200	Forward	AAAAAGCAGGCTCCACCTTTTGGATGCTTCTCGTGATAA
	Reverse	AGAAAGCTGGTCTGAGTTTCTGGATATGATTTCTT
attb1 extension	Forward	GGGGACAAGTTTGTACAAAAAAGCAGGCT
attb2 extension	Reverse	GGGGACCACTTTGTACAAGAAAGCTGGGT

Table S5. Proteins that were cleaved at the same site by AtMC9 during the N-terminome analysis of *Arabidopsis* roots and seedlings. Ion intensity increase fold of peptides derived from AtMC9 processing is shown in the column 'Intensity increase fold'. N/A; not applicable refers to fold change values of singleton ions.

#	Accession	Tissue	Experiment	Intensity fold increase	P4-P3'	Sequence	Description
1	AT1G04270	Roots	in vitro	6.50	KAKR(73)EAP	EAPQGEKPEPVR	cytosolic ribosomal protein S15
		Seedlings	in vitro	15.26			
2	AT1G04820	Roots	in vitro	N/A	ISGR(242)EDA	EDAANNFAR	tubulin alpha-4 chain
		Seedlings	in vitro	49.19			
3	AT1G08200	Seedlings	in vitro	N/A	VSSK(329)EFY	EFYGEYDSDSKR	UDP-D-xylose/UDP-D-xylose synthase 2
		Roots	KO/OE in vivo	17.29			
4	AT1G12000	Seedlings	in vitro	8.39	AVTR(10)DLT	DLTAVGSPENAPAKGR	Phosphofruktokinase family protein
		Seedlings	KO/OE	4.23			
		Seedlings	KO/WT	2.68			
		Roots	KO/WT in vivo	4.74			
5	AT1G17210	Roots	in vitro	N/A	VVDR(514)DGD	DGDEVNDDSDAGPSKR	IAP-like protein 1
		Seedlings	in vitro	N/A			
6	AT1G18070	Roots	in vitro	5.35	AQEK(85)AAK	AAKEEAEDVAEANKKR	Translation elongation factor EF1A
		Seedlings	in vitro	N/A			
7	AT1G49240	Roots	in vitro	4.52	MNQR(53)DAY	DAYVGDQAQSKR	actin 8
		Seedlings	in vitro	5.67			
		Seedlings	KO/OE	4.01			
		Seedlings	KO/WT	2.72			
8	AT2G01250	Roots	in vitro	N/A	VESK(6)VVV	VVPESVLKGR	Ribosomal protein L30/L7 family protein
		Seedlings	KO/OE	8.60			
		Seedlings	KO/WT	8.68			
9	AT2G07698	Seedlings	in vitro	23.03	EERR(27)AAE	AAELTNLFESR	ATPase, F1 complex, alpha subunit protein
		Roots	KO/WT in vivo	2.55			
10	AT2G22400	Roots	in vitro	N/A	TEDK(567)EAN	EANSSNAGGKR	S-adenosyl-L-methionine-dependent methyltransferase
		Seedlings	in vitro	N/A			
11	AT2G31610	Roots	in vitro	4.13	VLTR(28)ELA	ELAEDGYSGVEVR	Ribosomal protein S3 family protein
		Seedlings	in vitro	8.60			
12	AT2G36060	Roots	in vitro	N/A	RGEK(27)GIG	GIGDGTVSYGMDDGDDIYMR	MMS ZWEI homologue 3
		Seedlings	in vitro	4.49			
13	AT3G15790	Roots	in vitro	N/A	VEKK(215)TVE	TVEASDEKKNSEATR	methyl-CPG-binding domain 11
		Seedlings	in vitro	16.10			
14	AT3G16640	Roots	in vitro	N/A	STQK(67)VVD	VVDIVDTFR	translationally controlled tumor protein
		Seedlings	in vitro	31.74			
15	AT3G23810	Roots	in vitro	N/A	SSGR(13)EYK	EYKVKDMSQADFGR	S-adenosyl-L-homocysteine (SAH) hydrolase 2
		Seedlings	in vitro	N/A			
16	AT4G15410	Seedlings	KO/OE	9.96	LRSR(109)GGA	GGAGENKETENPSGIR	serine/threonine protein phosphatase 2A
		Roots	KO/OE in vivo	5.44			
		Seedlings	KO/WT	4.91			
17	AT4G17520	Seedlings	in vitro	13.44	QTAK(37)SGK	SGKMPTKPPPSQAVR	Hyaluronan / mRNA binding family
		Roots	KO/OE in vivo	6.95			
18	AT4G19210	Roots	in vitro	N/A	VIDR(213)DVE	DVENLSGGELQR	RNAse I inhibitor protein 2
		Seedlings	in vitro	N/A			
19	AT4G19880	Roots	in vitro	12.42	SMAR(35)SAV	SAVDETSDSGAFQR	Glutathione S-transferase family protein
		Seedlings	in vitro	6.26			
20	AT4G23670	Roots	in vitro	N/A	KEKR(74)EID	EIDDENKTLTKR	Polyketide cyclase/dehydrase and lipid transport protein
		Seedlings	in vitro	N/A			
21	AT4G36760	Roots	in vitro	N/A	EAAK(281)DME	DMEIDSQDPDR	aminopeptidase P1
		Seedlings	in vitro	N/A			
22	AT4G38630	Roots	in vitro	N/A	QKDK(258)DGD	DGDTASASQETVAR	regulatory particle non-ATPase 10
		Seedlings	in vitro	N/A			
23	AT5G04420	Roots	in vitro	6.44	RRQR(490)SAS	SASDEEEDGTVQR	Galactose oxidase/kelch repeat superfamily protein
		Seedlings	in vitro	3.28			
24	AT5G10360	Roots	in vitro	N/A	IVKK(119)GVS	GVSDLPLGLDTEKPR	Ribosomal protein S6e
		Seedlings	in vitro	5.07			
25	AT5G11260	Seedlings	KO/OE	2.50			
		Roots	in vitro	N/A	TSGR(59)ESG	ESGSATGQER	Basic-leucine zipper (bZIP) transcription factor family protein
26	AT5G47210	Roots	in vitro	N/A	EAKK(212)ELT	ELTAEKAQKEAEAEAR	Hyaluronan / mRNA binding family
		Seedlings	in vitro	N/A			
27	AT5G56000	Seedlings	KO/OE	7.04	SSKK(604)TME	TMEINPENSIMDELRL	HEAT SHOCK PROTEIN 81.4
		Roots	KO/OE in vivo	11.16			
		Seedlings	KO/WT	5.11			
28	AT5G66120	Roots	in vitro	N/A	NDQR(74)SIS	SISSPTVVEVDLGR	3-dehydroquinate synthase, putative
		Seedlings	in vitro	10.75			

REFERENCES

- Ahn, J.W., Verma, R., Kim, M., Lee, J.Y., Kim, Y.K., Bang, J.W., Reiter, W.D., and Pai, H.S. (2006). Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. *J Biol Chem* 281, 13708-13716.
- Backman, T.W., Cao, Y., and Girke, T. (2011). ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res* 39, W486-491.
- Bentsink, L., Jowett, J., Hanhart, C.J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proc Natl Acad Sci U S A* 103, 17042-17047.
- Bozhkov, P.V., Suarez, M.F., Filonova, L.H., Daniel, G., Zamyatnin, A.A., Jr., Rodriguez-Nieto, S., Zhivotovsky, B., and Smertenko, A. (2005). Cysteine protease mclII-Pa executes programmed cell death during plant embryogenesis. *Proc Natl Acad Sci U S A* 102, 14463-14468.
- Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinnyen, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318, 801-806.
- Carpita, N.C., and Gibeaut, D.M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J* 3, 1-30.
- Cheung, A., Dantzig, J.A., Hollingworth, S., Baylor, S.M., Goldman, Y.E., Mitchison, T.J., and Straight, A.F. (2002). A small-molecule inhibitor of skeletal muscle myosin II. *Nat Cell Biol* 4, 83-88.
- Cilenti, L., Lee, Y., Hess, S., Srinivasula, S., Park, K.M., Junqueira, D., Davis, H., Bonventre, J.V., Alnemri, E.S., and Zervos, A.S. (2003). Characterization of a novel and specific inhibitor for the pro-apoptotic protease Omi/HtrA2. *J Biol Chem* 278, 11489-11494.
- Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16, 735-743.
- Colaert, N., Helsen, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 6, 786-787.
- Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). *Arabidopsis* type I metacaspases control cell death. *Science* 330, 1393-1397.
- Cosgrove, D.J. (2005). Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6, 850-861.
- Degterev, A., Lugovskoy, A., Cardone, M., Mulley, B., Wagner, G., Mitchison, T., and Yuan, J. (2001). Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat Cell Biol* 3, 173-182.
- del Campillo, E., Abdel-Aziz, A., Crawford, D., and Patterson, S.E. (2004). Root cap specific expression of an endo-beta-1,4-D-glucanase (cellulase): a new marker to study root development in *Arabidopsis*. *Plant Mol Biol* 56, 309-323.
- Driouch, A., Durand, C., and Vire-Gibouin, M. (2007). Formation and separation of root border cells. *Trends Plant Sci* 12, 14-19.
- Dubrovsky, J.G., Sauer, M., Napsucialy-Mendivil, S., Ivanchenko, M.G., Friml, J., Shishkova, S., Celenza, J., and Benkova, E. (2008). Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc Natl Acad Sci U S A* 105, 8790-8794.
- French, K.J., Schrecengost, R.S., Lee, B.D., Zhuang, Y., Smith, S.N., Eberly, J.L., Yun, J.K., and Smith, C.D. (2003). Discovery and evaluation of inhibitors of human sphingosine kinase. *Cancer Res* 63, 5962-5969.
- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 21, 566-569.
- Gonzalez-Carranza, Z.H., Whitelaw, C.A., Swarup, R., and Roberts, J.A. (2002). Temporal and spatial expression of a polygalacturonase during leaf and flower abscission in oilseed rape and *Arabidopsis*. *Plant Physiol* 128, 534-543.
- Grobe, K., Becker, W.M., Schlaak, M., and Petersen, A. (1999). Grass group I allergens (beta-expansins) are novel, papain-related proteinases. *Eur J Biochem* 263, 33-40.
- Hamann, A., Brust, D., and Osiewacz, H.D. (2007). Deletion of putative apoptosis factors leads to lifespan extension in the fungal ageing model *Podospira anserina*. *Mol Microbiol* 65, 948-958.
- Hawes, M.C., and Lin, H.J. (1990). Correlation of Pectolytic Enzyme Activity with the Programmed Release of Cells from Root Caps of Pea (*Pisum sativum*). *Plant Physiol* 94, 1855-1859.
- He, R., Drury, G.E., Rotari, V.I., Gordon, A., Willer, M., Farzaneh, T., Woltering, E.J., and Gallois, P. (2008). Metacaspase-8 modulates programmed cell death induced by ultraviolet light and H₂O₂ in *Arabidopsis*. *J Biol Chem* 283, 774-783.
- Helm, M., Schmid, M., Hierl, G., Terneus, K., Tan, L., Lottspeich, F., Kieliszewski, M.J., and Gietl, C. (2008). KDEL-tailed cysteine endopeptidases involved in programmed cell death, intercalation of new cells, and dismantling of extensin scaffolds. *Am J Bot* 95, 1049-1062.
- Helms, M.J., Ambit, A., Appleton, P., Tetley, L., Coombs, G.H., and Mottram, J.C. (2006). Bloodstream form *Trypanosoma brucei* depend upon multiple metacaspases associated with RAB11-positive endosomes. *J Cell Sci* 119, 1105-1117.
- Hierl, G., Vothknecht, U., and Gietl, C. (2012). Programmed cell death in *Ricinus* and *Arabidopsis*: the function of KDEL cysteine peptidases in development. *Physiol Plant* 145, 103-113.
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008, 420747.
- Hu, Y., Helm, J.S., Chen, L., Ginsberg, C., Gross, B., Kraybill, B., Tianont, K., Fang, X., Wu, T., and Walker, S. (2004). Identification of selective inhibitors for the glycosyltransferase MurG via high-throughput screening. *Chem Biol* 11, 703-711.
- Jaenicke, L., Kuhne, W., Spessert, R., Wahle, U., and Waffenschmidt, S. (1987). Cell-wall lytic enzymes (autolysins) of *Chlamydomonas reinhardtii* are (hydroxy)proline-specific proteases. *Eur J Biochem* 170, 485-491.
- Jamet, E., Boudart, G., Borderies, G., Charmont, S., Lafitte, C., Rossignol, M., Canut, H., and Pont-Lezica, R. (2008). Isolation of plant cell wall proteins. *Methods Mol Biol* 425, 187-201.
- Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. *Plant Physiol* 145, 1144-1154.
- Li, L.C., and Cosgrove, D.J. (2001). Grass group I pollen allergens (beta-expansins) lack proteinase activity

- and do not cause wall loosening via proteolysis. *Eur J Biochem* 268, 4217-4226.
- Li, S., Lei, L., Somerville, C.R., and Gu, Y. (2012). Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. *Proc Natl Acad Sci U S A* 109, 185-190.
- Malamy, J.E., and Benfey, P.N. (1997). Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. *Development* 124, 33-44.
- Mei, Y., Gao, H.B., Yuan, M., and Xue, H.W. (2012). The *Arabidopsis* ARCP Protein, CSI1, Which Is Required for Microtubule Stability, Is Necessary for Root and Anther Development. *Plant Cell* 24, 1066-1080.
- Passardi, F., Penel, C., and Dunand, C. (2004). Performing the paradoxical: how plant peroxidases modify the cell wall. *Trends Plant Sci* 9, 534-540.
- Peret, B., De Rybel, B., Casimiro, I., Benkova, E., Swarup, R., Laplaze, L., Beeckman, T., and Bennett, M.J. (2009). *Arabidopsis* lateral root development: an emerging story. *Trends Plant Sci* 14, 399-408.
- Richie, D.L., Miley, M.D., Bhabhra, R., Robson, G.D., Rhodes, J.C., and Askew, D.S. (2007). The *Aspergillus fumigatus* metacaspases CasA and CasB facilitate growth under conditions of endoplasmic reticulum stress. *Mol Microbiol* 63, 591-604.
- Roberts, J.A., Elliott, K.A., and Gonzalez-Carranza, Z.H. (2002). Abscission, dehiscence, and other cell separation processes. *Annu Rev Plant Biol* 53, 131-158.
- Schechter, I., and Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* 27, 157-162.
- Schilling, O., and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26, 685-694.
- Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in *Arabidopsis*. *Plant Cell* 18, 1121-1133.
- Shindo, T., Misas-Villamil, J.C., Horger, A.C., Song, J., and van der Hoorn, R.A. (2012). A role in immunity for *Arabidopsis* cysteine protease RD21, the ortholog of the tomato immune protease C14. *PLoS One* 7, e29317.
- Saes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* 6, 1130-1141.
- Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Peret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., et al. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nat Cell Biol* 10, 946-954.
- Tabuchi, A., Li, L.C., and Cosgrove, D.J. (2011). Matrix solubilization and cell wall weakening by beta-expansin (group-1 allergen) from maize pollen. *Plant J* 68, 546-559.
- Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavalov, A., Lam, E., and Bozhkov, P.V. (2011). Metacaspases. *Cell death and differentiation* 18, 1279-1288.
- Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.A., De Rycke, R., Brackenier, A., Inze, D., Harris, J.L., and Van Breusegem, F. (2006). Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* 364, 625-636.
- Vercammen, D., Declercq, W., Vandenabeele, P., and Van Breusegem, F. (2007). Are metacaspases caspases? *J Cell Biol* 179, 375-380.
- Vercammen, D., van de Cotte, B., De Jaeger, G., Eeckhout, D., Casteels, P., Vandepoele, K., Vandenberghe, I., Van Beeumen, J., Inze, D., and Van Breusegem, F. (2004). Type II metacaspases Atmc4 and Atmc9 of *Arabidopsis thaliana* cleave substrates after arginine and lysine. *J Biol Chem* 279, 45329-45336.
- Wang, X., Feng, H., Tang, C., Bai, P., Wei, G., Huang, L., and Kang, Z. (2012). TaMCA4, a Novel Wheat Metacaspase Gene Functions in Programmed Cell Death Induced by the Fungal Pathogen *Puccinia striiformis* f. sp. *tritici*. *Mol Plant Microbe Interact* 25, 755-764.
- Ward, W., Alvarado, L., Rawlings, N.D., Engel, J.C., Franklin, C., and McKerrow, J.H. (1997). A primitive enzyme for a primitive cell: the protease required for excystation of *Giardia*. *Cell* 89, 437-444.
- Watanabe, N., and Lam, E. (2011). *Arabidopsis* Metacaspase 2d Is a Positive Mediator of Cell Death Induced during Biotic and Abiotic Stresses. *Plant J*.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2, e718.
- Wojtaszek, P. (1997). Oxidative burst: an early plant response to pathogen infection. *The Biochemical journal* 322 (Pt 3), 681-692.
- Wolf, S., Rausch, T., and Greiner, S. (2009). The N-terminal pro region mediates retention of unprocessed type-I PME in the Golgi apparatus. *Plant J* 58, 361-375.
- Yamada, K., Matsushima, R., Nishimura, M., and Hara-Nishimura, I. (2001). A slow maturation of a cysteine protease with a granulysin domain in the vacuoles of senescing *Arabidopsis* leaves. *Plant Physiol* 127, 1626-1634.

Chapter 5

METACASPASE 1 substrate processing during *Pseudomonas syringae*-triggered cell death

Liana Tsiatsiani^{a,b,c,d}, Nuria S. Coll^e, Pieter-Jan De Bock^{c,d}, Kris Gevaert^{c,d}, Jeffery L. Dangl^{f,g,h,i}, Petra Epple^{f,i} and Frank Van Breusegem^{a,b}

^a Department of Plant Systems Biology, VIB, Technologiepark 927, B-9052 Ghent, Belgium

^b Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

^c Department of Medical Protein Research, VIB, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^d Department of Biochemistry, Ghent University, Albert Baertsoenkaai 3, B-9000 Ghent, Belgium

^e Centre for Research in Agricultural Genomics, Autonomous University of Barcelona, Bellaterra-Cerdanyola del Valles, 08193, Barcelona, Spain

^f Department of Biology, 108 Coker Hall, University of North Carolina (UNC), CB 3280, Chapel Hill, NC 27599–3280, USA

^g Curriculum in Genetics and Molecular Biology and Department of Microbiology and Immunology, UNC, Chapel Hill, NC 27599, USA.

^h Carolina Center for Genome Sciences, UNC, Chapel Hill, NC 27599, USA.

ⁱ Howard Hughes Medical Institute, University of North Carolina, Chapel Hill, North Carolina 27599, USA

AIM AND CONTEXT

Metacaspases, the relatives of caspases, are divided into type-I and type-II based on the presence or absence of an N-terminal prodomain, respectively (Uren et al., 2000). Plant genomes contain both types of metacaspases while fungi and protozoa contain only the type-I proteases. The functionality of type-I metacaspases has been recently studied in *Arabidopsis thaliana*, where it was shown that two out of the three *Arabidopsis* type-I metacaspases antagonistically control cell death upon during the plant hypersensitive response to pathogen recognition (HR) (Coll et al., 2010). In fact, AtMC1 acts as a positive regulator of cell death whereas AtMC2 has an opposite effect. Even though the function of these two metacaspases has been reported, no experimental data exist on their substrates. To identify downstream members of the HR pathway related to the cell death-promoting AtMC1 activity, we performed *in vivo* profiling of AtMC1 substrates using N-terminal COFRADIC (Gevaert et al., 2003). By comparison of the N-terminomes derived from wild type, AtMC1 gain- and loss-of-function mutants, with and without pathogen treatment, we identified proteins that are potentially *in vivo* cleaved by AtMC1. Due to the unavailability of recombinant AtMC1 we did not confirm the cleavage of proteins *in vitro* and the protease's specificity remains enigmatic.

AUTHOR CONTRIBUTIONS

LT and NSC designed the experiments with the help of PE, JLD, KG and FVB. PJDB performed the MS/MS analysis. LT performed the experiments, and LT, KG and FVB wrote the manuscript.

ABSTRACT

Metacaspases are the structural homologs of caspases found in protozoa, fungi and plants. Genetic studies in fungi have shown that metacaspases are required for cell death, while their functions are extended to additional non-cell death roles. Recently, it was shown that in *Arabidopsis* a pair of type-I metacaspases, AtMC1 (At1g02170) and AtMC2 (At4g25110), comprise a regulatory module for cell death during the hypersensitive response to pathogen recognition. Using positional proteomics we identified candidate AtMC1 protein substrates that are involved in defence responses and innate immunity. We report for the first time on the proteome-wide specificity of a plant type-I metacaspase and we show that besides Arg and Lys, Gln was also cleaved at substrate position P1.

INTRODUCTION

Metacaspases are distinguished based on the presence (type-I) or absence (type-II) of an N-terminal prodomain that contains a plant-specific LSD1-like (Lesion simulating disease 1) zinc-finger motif (Uren et al., 2000). By homology to the caspase sequence and domain structure, type-I metacaspases were correlated to initiator caspases and type-II metacaspases to effector caspases. Type-I metacaspases are found in protozoa, fungi and plants, whereas type-II metacaspases are specific for plants (Vercammen et al., 2007). The type-I metacaspases of various fungi, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Podospora anserina*, *Candida albicans* and *Aspergillus nubilans*, are required for certain cell death events, and cell death promoting functions have been described for protozoa and plant metacaspases (reviewed in Tsatsiani et al., 2011).

Antagonistic control of cell death has been proposed for two type-I metacaspases in *Arabidopsis thaliana* (Coll et al., 2010). LSD1 is a negative regulator of cell death during the plant hypersensitive response (HR) to pathogen challenge (Dietrich et al., 1994). LSD1 contains three zinc-finger motifs that have been shown to mediate protein-protein interactions (Kaminaka et al., 2006; Coll et al., 2011) and, as mentioned above, this motif is also found in the N-terminal prodomain all three *Arabidopsis* type-I metacaspases. In yeast cells, ectopically expressed LSD1 and AtMC1 interact via their N-terminal prodomains. In this system, AtMC2 interacts only very weakly with either LSD1 or AtMC1. In addition, LSD1 co-immunoprecipitated only with AtMC1 in protein extracts from *Arabidopsis* and *Nicotiana benthamiana* leaves, whereas AtMC1 and AtMC2 interaction was not substantiated *in planta* (Coll et al., 2010). In the same study, AtMC1 was shown to positively regulate pathogen-triggered HR cell death in *Arabidopsis*, while AtMC2 had an opposite effect. These metacaspase functions depended on the catalytic activity of AtMC1 but not AtMC2 and similar to mammalian caspases, their functions are enhanced by prodomain removal.

Given the importance of AtMC1 as a positive regulator of cell death in this regulatory module, we analyzed its degradome - the substrate repertoire (Overall et al., 2004). We assessed those proteins that were differentially cleaved in the *Arabidopsis* leaf proteomes of *atmc1* knockout, wild type (Col-0) and stable transgenic plants conditionally overexpressing *AtMC1*, with and without a cell death trigger. As a cell death trigger we used infection of Col-0 plants with the avirulent bacteria *Pseudomonas syringae* pv. tomato (*Pto*) DC3000(*avrRpm1*), which causes AtMC1-dependent HR cell death upon activation of the resistance protein RPM1. In this way, we additionally aimed to specifically identify new members of the HR cell death pathway in plants downstream of AtMC1 activation.

RESULTS

Identification of potential AtMC1 substrates by N-terminal COFRADIC

Using positional proteomics we identified potential AtMC1 protein substrates by comparing the N-terminomes from extracts of *AtMC1* gain- and loss-of-function *Arabidopsis* plants. An N-terminome consists of all protein N-termini-bearing peptides and these can either be the mature protein N-termini or newly formed N-termini that are introduced in a protein by protease-mediated substrate processing. By means of N-terminal COFRADIC we identified and quantified such neo-N-terminal peptides following differential mass-tagging of their free α -amines (Gevaert et al., 2003, Staes et al., 2011). This comparative N-terminome analysis further allowed us to quantify possible differences in the extent of processing that a given site in a certain protein underwent in each proteome analyzed.

In total, we performed four N-terminome studies, with two of them involving infection with *Pto* DC3000(*avrRpm1*). In the first set of analyses we compared the N-terminomes of wild type Col-0 (WT) plants to *atmc1* knockout (KO) and then to plants conditionally expressing AtMC1 under the control of a Dexamethasone promoter (OE). The second set of analyses was performed on leaves infected with *Pto* DC3000(*avrRpm1*) and we first compared a KO N-terminome to WT and then to OE plants. For all analyses, leaves of 3-week-old plants were used and in the rest of the chapter, we will refer to the different analyses as WT/KO, OE/WT and WT/KO-P or OE/KO-P corresponding to the set-ups that were previously mentioned.

In each proteome analysis, protein (neo-)N-termini (α -amines) and lysine side-chains (ϵ -amines) were labeled with NHS esters of either $^{12}\text{C}_4$ -butyrate (light isotope, L) or $^{13}\text{C}_4$ -butyrate (heavy isotope, H) and equal amounts of material were mixed (Figure 1). Following LC-MS/MS analysis, peptide identification and quantification, the majority of peptides was found to have light/heavy ion intensity ratios of approximately 1 and these are therefore likely not reporting AtMC1 proteolysis events. Of interest are peptides that are either only found in one N-terminome (singletons) or that are significantly (here, $p < 0.02$) regulated in an N-terminome. Further, since we mass-tagged all proteome samples containing active endogenous or ectopically overexpressed *AtMC1* with light butyrate (Figure 1), we were particularly interested in peptides with L/H ratios that are significantly higher than 1.

Selection of potential *AtMC1* substrates

Our selection criteria for AtMC1 proteolysis-reporting peptides are described in the following categories: a) A neo-N-terminal peptide is only found in the sample containing active AtMC1 (WT or OE), meaning that the corresponding protein was cleaved by AtMC1 or by a protease that is activated by AtMC1 (Figure 2A). b) A (neo) N-terminal peptide is uniquely found in the control sample (KO) because for instance the corresponding protein was completely degraded in the sample containing active AtMC1. For such AtMC1 substrates we cannot draw conclusions about the cleavage site, but may derive that these proteins were indeed likely AtMC1 substrates (Figure 2B). c) A neo-N-terminal peptide is found in both samples, but in a significantly higher amount in the sample containing active AtMC1, which indicates that the corresponding protein was also cleaved in the control sample (without endogenous AtMC1 activity), although to a lesser extent, by a protease with redundant activity to AtMC1 (Figure 2C). Note that this is a case which we cannot rule out given that the *Arabidopsis* genome encodes for several metacaspases (3 type-I and 6 type-II) that might well share substrates. Alternatively these substrates could be cut by a protease whose activity is enhanced by AtMC1 in the WT or OE proteome. d) (Neo-)N-terminal peptides found in approximately

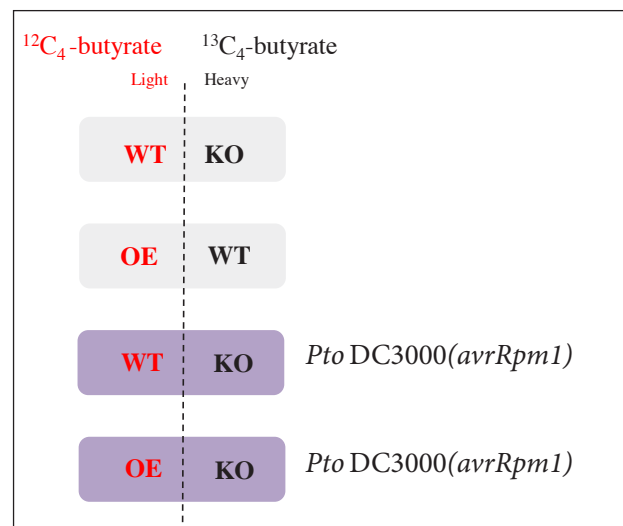


Figure 1. Experimental set-up for the proteome analyses. The compared proteomes were labeled with NHS-esters of isotopic variants of butyric acid (light, $^{12}\text{C}_4$ -butyrate or heavy, $^{13}\text{C}_4$ -butyrate) and equal amounts were mixed prior to analysis by N-terminal COFRADIC.

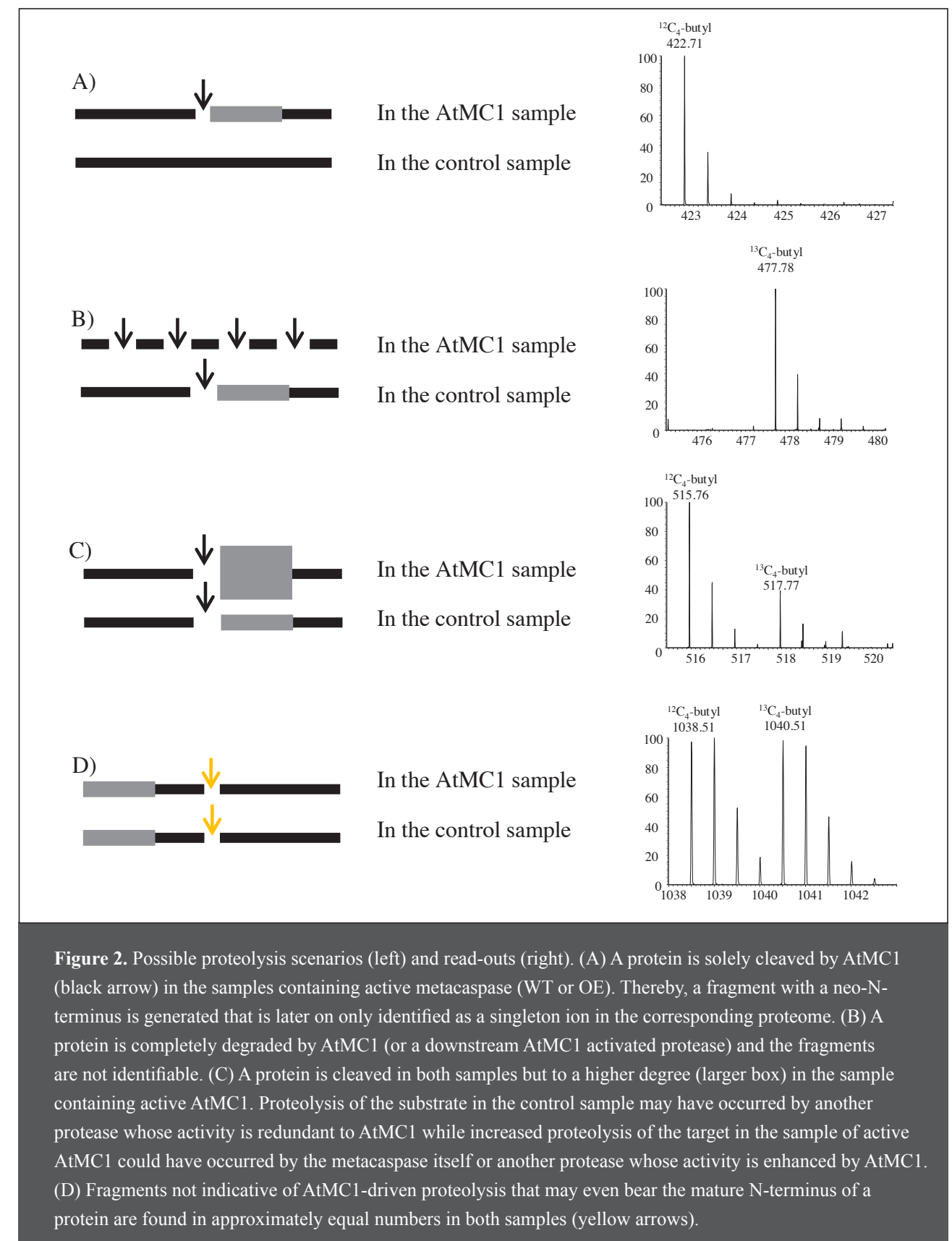


Figure 2. Possible proteolysis scenarios (left) and read-outs (right). (A) A protein is solely cleaved by AtMC1 (black arrow) in the samples containing active metacaspase (WT or OE). Thereby, a fragment with a neo-N-terminus is generated that is later on only identified as a singleton ion in the corresponding proteome. (B) A protein is completely degraded by AtMC1 (or a downstream AtMC1 activated protease) and the fragments are not identifiable. (C) A protein is cleaved in both samples but to a higher degree (larger box) in the sample containing active AtMC1. Proteolysis of the substrate in the control sample may have occurred by another protease whose activity is redundant to AtMC1 while increased proteolysis of the target in the sample of active AtMC1 could have occurred by the metacaspase itself or another protease whose activity is enhanced by AtMC1. (D) Fragments not indicative of AtMC1-driven proteolysis that may even bear the mature N-terminus of a protein are found in approximately equal numbers in both samples (yellow arrows).

equal amounts in both proteomes, and these do not indicate proteolysis related to AtMC1 activity (Figure 2D).

To identify significantly enriched neo-N-terminal peptides, we first performed a test analysis to determine the level of technical variation that was introduced in our experiments due to sample handling, efficiency of peptide labeling and the actual mass spectrometric analysis. A WT leaf proteome was divided in two parts and one part was labeled with light butyrate, while the second part was labeled with heavy butyrate and subsequently analysed by N-terminal COFRADIC similar to all other four analyses. The standard deviation calculated from this test analysis (Huber scale = 0.25) was implemented in all analyses of biological interest as technical variation (Huber, 1981).

Following individual inspection of annotated and quantified neo-N-terminal peptides, we identified 1,254 peptides in the WT/KO analysis, 938 in the WT/OE analysis, 1,195 in the WT/KO-P analysis and 1,211 in the OE/KO-P analysis (Table 1). Proteolysis-reporter peptides (singletons and AtMC1 enriched) accounted to about 1.1–2.9% of all identified peptides in each analysis (Table 1 and supporting information tables S1–S4). The number of possible substrates identified in the *Pto* DC3000(*avrRpm1*)-infected AtMC1 OE line was quite low. In fact a much broader distribution (based on L/H ratio values) was observed for this setup, which made it more difficult to identify AtMC1 proteolysis-reporter peptides. One reason for this broad distribution might be that more drastic overall proteome changes occur when infecting AtMC1 overexpressing plants. The highest number of protein processing events (35) was identified in the WT/KO-P analysis (Table 1 and S3) and 11 of these were reported as either unique processing events (singletons in WT proteome) or indirect processing events (singletons in the KO proteome).

Ten proteins were found potentially cleaved by AtMC1 in more than one independent proteome analysis, a finding that makes these proteins very likely AtMC1 substrates (Table 2). Using publicly available Affymetrix array data and Genevestigator (Hruz et al., 2008), we assessed whether all potential AtMC1 substrates identified here were transcriptionally co-expressed with AtMC1 during biotic stimuli, hormone treatment and in overall *Arabidopsis* anatomy (Table 3). In this way, among others, a type-II metacaspase (AtMC4), an Auxin efflux carrier protein (PIN3), a senescence-associated protein (SAG21) and the defence-related proteins WYRK33 and MOD1 were found clustered with *AtMC1*. The chloroplast encoded ATP synthase β -subunit (ATCG00480), a protein sharing 68% sequence similarity with its cell death regulating homolog (Chivasa et al., 2011), was more than 3-fold higher (z-score 2.64) cleaved in the WT N-terminome compared to the KO N-terminome (Table S1). Moreover, the here identified cleavage site is conserved in both proteins.

In the analyses that involved pathogen treatment, glycine-rich proteins were identified as potential AtMC1 targets and such proteins are known to be involved in signal transduction and innate immunity (Mangeon et al., 2010). Finally, AtMC1 was not identified in any of the analyses, which was somehow expected since AtMC1 peptides have not been identified so far by any LC-MS/MS method in *Arabidopsis* leaves (AtProteome and

Table 1. Overview of the results from the performed proteome analyses. The numbers of identified peptides and their matched proteins are listed, together with the types of cleavage sites identified in each analysis.

General		Cleavage sites			
Analysis	Peptides	Proteins	Singletons	AtMC1 enriched (p<0.02)	% of total peptides
WT/KO	1,254	795	-	27	2.2
WT/OE	938	623	4	19	2.5
WT/KO-P	1,195	807	11	24	2.9
OE/KO-P	1,211	795	4	9	1.1

Table 2. Proteins characterized as potential AtMC1 targets in two independent analyses. Amino acid position of P1' site is shown between brackets in the P4-P3' column. Fold times of ion intensity enrichment in the proteome of active AtMC1 is shown in the column 'Intensity increase fold'. N/A; not applicable refers to singleton ions.

	Accession	Intensity fold increase	z-score	P4-P3'	Sequence	Description	Experiment
1	AT1G12900	2.48	2.65	VTEA(64)KIK	KIKVAINGFGR	GAPA-2 (glyceraldehyde 3-phosphate dehydrogenase A subunit 2)	OE/WT
	AT1G12900	3.76	2.86	VQV(314)SKKF	KKTFAEVNAAFR		WT/KO
2	AT1G13930	3.96	2.97	LMAS(44)AKV	AKVVAEAAQAAAR	Involved in response to salt stress	WT/KO
	AT1G13930	N/A	N/A	LMAS(44)AKV	AKVVAEAAQAAAR		OE/WT
3	AT1G67740	2.93	2.29	PEAY(143)AAA	AAAEAAAASSDSR	PSBY (Photosystem II BY)	WT/KO
	AT1G67740	2.70	2.89	EAYA(144)AAE	AAAEAAAASSDSR		OE/WT
4	AT2G05520	5.75	2.05	GGNY(72)QGG	QGGGGNYQGGGGR	glycine-rich protein 3	OE/KO-P
	AT2G05520	6.54	2.20	GGRQ(107)GGG	GGGGSGGSYCR		OE/KO-P
	AT2G05520	2.05	3.45	GGRQ(107)GGG	GGGGSGGSYCR		WT/KO-P
5	AT2G05990	2.39	2.53	TRAM(77)SES	SESSENKAPSGPLIDLR	MOD1 (mosaic death 1)	OE/WT
	AT2G05990	2.99	2.33	IRVN(306)IIS	TISAGPLGSR		WT/KO
6	AT2G37220	3.36	2.60	PRSS(185)FGS	FGSSSGYGGGGSGAGSGNR	RNA-binding family protein	WT/KO
	AT2G37220	2.40	4.08	PRSS(185)FGS	FGSSSGYGGGGSGAGSGNR		WT/KO-P
7	AT2G39730	3.02	2.36	RWRG(74)LAY	LAYDTSDDQDITR	RCA (rubisco activase)	WT/KO
	AT2G39730	3.81	2.88	MMSA(192)GEL	GELESGNAGEPAKLIR		WT/KO
	AT2G39730	4.64	3.33	VNNQ(249)MVN	MVNATLMNIADNPTNVQLPGMYNKEENAR		WT/KO
	AT2G39730	1.66	2.53	NATL(255)MNI	MNIADNPTNVQLPGMYNKEENAR		WT/KO-P
	AT2G39730	6.73	4.18	IVTLV(334)DQ	VDQFPQGSIDFFGALR		WT/KO
	AT2G39730	5.27	3.62	VRKF(361)VES	VESLGVVEKIGKR		WT/KO
	AT2G39730	1.63	2.45	GNML(401)VME	VMEQENVKR		WT/KO-P
	AT2G39730	3.29	2.55	YGNM(400)LVM	LVMEQENVKR		WT/KO
	AT2G39730	8.07	4.58	AETY(417)LSQ	LSQAALGDANADAIGR		WT/KO
8	AT3G14415	2.59	4.37	KMAH(86)PDG	PDGEYATAR	Aldolase-type TIM barrel family protein	WT/KO-P
	AT3G14415	3.09	2.41	FKAL(299)ALG	ALGASGIFIGR		WT/KO
9	AT3G15356	2.77	2.16	DGSN(32)LLF	LLFLGDAELGPSSDGVSR	Legume lectin family protein	WT/KO
	AT3G15356	5.79	2.06	LQNM(248)YAG	YAGFAGSMGR		OE/KO-P
10	AT5G42530	1.78	2.84	KALG(28)AQE	AQEEAQQR	unknown protein	WT/KO-P
	AT5G42530	N/A	N/A	LGAQ(30)EEA	EEAQQR		OE/KO-P

PRIDE databases) probably due to its moderate expression level (eFP browser; Schmid et al., 2005) or possible technical limitations in identifying AtMC1 peptides by LC-MS/MS.

AtMC1 specificity

We assessed AtMC1 specificity after alignment of the cleaved amino acid sequences and statistical correction for the natural occurrence of amino acids in *Arabidopsis* proteins (Colaert et al., 2009). Singletons identified in the equivalent control samples, indicative of indirect AtMC1-driven proteolysis (Figure 2B), were excluded from the sequence alignment. Based on this analysis, we did not always observe a dominant preference for Arg (R) or Lys (K) at the P1 cleavage site. In fact, cleavage after Arg was only found in the OE/WT analysis and after Lys in the WT/KO-P analysis (Figure 3). In the latter, we could additionally observe a second basic residue at P3 (R), reminiscent of the AtMC9 signature (see chapters 3 and 4), but this dibasic motif was only found in one of the cleaved proteins, AtMC4. As for the WT/KO and OE/KO-P analyses, Gln (Q) was most frequently found at P1. Interestingly, no Q-

specific plant proteases have been reported so far (Rawlings et al, 2010). Therefore, such cleavages found in the OE/KO-P study might have resulted from pathogen-delivered proteases, although we would not expect to identify these events more frequently in the OE proteome since the KO tissue sample was similarly treated with the pathogen.

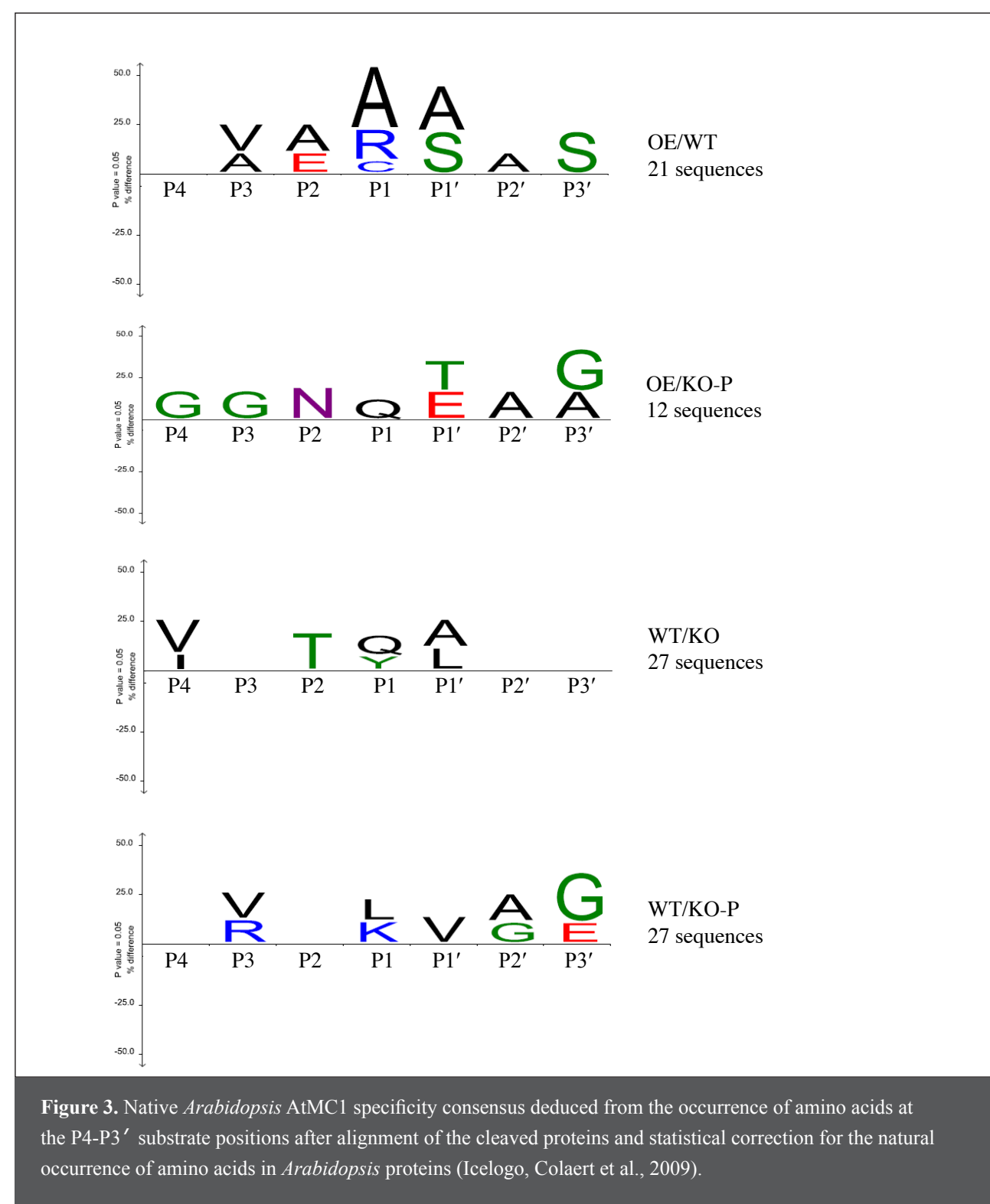


Table 3. List of potential AtMC1 substrate proteins that are co-expressed with *AtMC1* during biotic stimuli, hormone treatment and in overall *Arabidopsis* anatomy (Hruz et al., 2008). Amino acid position of P1' site is shown between brackets in the P4-P3' column. Fold times of ion intensity enrichment in the proteome of active AtMC1 is shown in the column 'Intensity increase fold'. N/A; not applicable refers to singleton ions.

Accession	Intensity fold increase	z-score	P4-P3'	Sequence	Description	Experiment
Anatomy						
AT1G75750	1.80	2.89	LVQA(24)DVE	DVENSQKNGYAKKIDCGSACVAR	GAST1 protein homolog 1	WT/KO -P
AT1G79340	2.40	4.07	LRSK(183)VEG	VEGAIESR	metacaspase 4	WT/KO -P
AT2G15620	2.61	2.79	LVAA(29)AQT	AQTTAPAESTASVDADR	NIR1 (NITRITE REDUCTASE 1)	OE/WT
AT2G30930	3.26	2.53	VSTQ(77)AKD	AKDAVDKAFSR	unknown protein	WT/KO
AT2G38280	6.28	2.16	ATPF(84)TDG	TDGGGGGGGDTGR	AMP deaminase, putative	OE/KO -P
AT2G38470	N/A	N/A	KGDN(333)ETN	ETNGGNGGGSKTVR	WRKY DNA-binding protein 33	WT/KO -P
AT3G02875	3.27	3.45	LSSA(24)GSY	GSYDSGSGLESLAR	ILR1 (IAA-LEUCINE RESISTANT 1)	OE/WT
AT4G02380	2.65	4.46	RRGY(29)AAT	AATAAQGSVSSGGR	SAG21 senescence-associated gene 21	WT/KO -P
AT4G36860	N/A	N/A	RAQL(98)EAA	EAAEEEEER	LIM domain-containing protein	OE/KO -P
Biotic						
AT1G70940	1.62	2.42	NVFG(379)GAP	GAPDNDQGGGR	PIN3, Auxin efflux carrier family protein	WT/KO -P
AT1G79340	2.40	4.07	LRSK(183)VEG	VEGAIESR	metacaspase 4	WT/KO -P
AT2G05990	2.39	2.53	TRAM(77)SES	SESSENKAPSGLPIDLR	MOD1 (MOSAIC DEATH 1)	OE/WT
AT2G05990	2.99	2.33	IRVN(306)TIS	TISAGPLGSR	MOD1 (MOSAIC DEATH 1)	WT/KO
AT2G30930	3.26	2.53	VSTQ(77)AKD	AKDAVDKAFSR	unknown protein	WT/KO
AT2G38280	6.28	2.16	ATPF(84)TDG	TDGGGGGGGDTGR	AMP deaminase, putative	OE/KO -P
AT2G38470	N/A	N/A	KGDN(333)ETN	ETNGGNGGGSKTVR	WRKY DNA-binding protein 33	WT/KO -P
AT4G02380	2.65	4.46	RRGY(29)AAT	AATAAQGSVSSGGR	senescence-associated gene 21	WT/KO -P
AT4G10480	3.21	5.18	GEAK(123)IDD	IDDMSSQLQAQAAQR	Nascent polypeptide-associated complex (NAC), alpha subunit family protein	WT/KO -P
AT4G26300	1.64	2.50	VVAM(54)AAN	AANEFTGNLKR	Arginyl-tRNA synthetase, class Ic	WT/KO -P
AT4G35630	2.42	2.57	AVRC(52)VAS	VASTTQVDGVR	PSAT phosphoserine aminotransferase	OE/WT
AT4G35630	2.31	2.44	DGVR(64)SGS	SGSVGSQER	PSAT phosphoserine aminotransferase	OE/WT
AT4G36860	N/A	N/A	RAQL(98)EAA	EAAEEEEER	LIM domain-containing protein	OE/KO -P
Hormone						
AT1G53240	N/A	N/A	RRSF(23)SSG	SSGSVPER	Lactate/malate dehydrogenase family protein	WT/KO -P
AT2G05990	2.39	2.53	TRAM(77)SES	SESSENKAPSGLPIDLR	MOD1 (MOSAIC DEATH 1)	OE/WT
AT2G05990	2.99	2.33	IRVN(306)TIS	TISAGPLGSR	MOD1 (MOSAIC DEATH 1)	WT/KO
AT2G15620	2.61	2.79	LVAA(29)AQT	AQTTAPAESTASVDADR	NIR1 (NITRITE REDUCTASE 1)	OE/WT
AT2G28000	3.07	2.39	LTTQ(565)AIV	AIVVDKPKPKAPAAAPEGLMV	CPN60A (CHAPERONIN-60ALPHA)	WT/KO
AT2G38280	6.28	2.16	ATPF(84)TDG	TDGGGGGGGDTGR	AMP deaminase, putative	OE/KO -P
AT3G53460	1.96	3.25	SQRS(24)SYG	SYGSGSGSGSGSGNR	chloroplast RNA-binding protein 29	WT/KO -P
AT4G02380	2.65	4.46	RRGY(29)AAT	AATAAQGSVSSGGR	senescence-associated gene 21	WT/KO -P
AT4G10480	3.21	5.18	GEAK(123)IDD	IDDMSSQLQAQAAQR	Nascent polypeptide-associated complex (NAC), alpha subunit family protein	WT/KO -P
AT4G24280	3.21	5.18	FEVL(291)STS	STSGDTHLGGDDFDKR	chloroplast heat shock protein 70-1	WT/KO -P
AT4G26300	1.64	2.50	VVAM(54)AAN	AANEFTGNLKR	Arginyl-tRNA synthetase, class Ic	WT/KO -P
AT4G35630	2.42	2.57	AVRC(52)VAS	VASTTQVDGVR	PSAT phosphoserine aminotransferase	OE/WT
AT4G35630	2.31	2.44	DGVR(64)SGS	SGSVGSQER	PSAT phosphoserine aminotransferase	OE/WT
ATCG00140	2.29	3.89	AVGL(20)ASI	ASIGPVGQGTAAGQAVEGIAR	ATP synthase subunit C family protein	WT/KO -P

Supporting information

Table S1. List of endogenous AtMC1 substrates identified in the WT proteome. Peptides were identified as significantly ($p < 0.02$) enriched ions in the proteome of active protease (WT). Amino acid position of the P1' site is shown between brackets in the P4-P3' column. Fold times of higher WT ion intensity in comparison to the KO ion counterpart is shown

in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score (s)	threshold (t)	Confidence	E value	Mass error (ppm)
AT1G12900	860.968	2	3.76	3.00	VQVS(314)KKF	KKTFAEVNAAFR	GAPA-2 (glyceraldehyde 3-phosphate dehydrogenase A subunit 2)	73	44	0.01	1.26E-5	-0.930
AT1G13930	698.402	2	3.96	3.12	LMAS(44)AKV	AKVVAAEAQAAAR	Involved in response to salt stress	79	43	0.01	2.51E-6	-0.789
AT1G60950	868.895	2	4.48	3.39	VYVL(78)DAA	DAEEAGIDLPYSCR	FEDA	75	40	0.01	3.16E-6	-1.267
AT1G67740	624.298	2	2.93	2.44	PEAY(143)AAA	AAAEAAAASSDSR	PSBY (Photosystem II BY)	45	42	0.01	5.01E-3	-1.284
AT2G05990	514.790	2	2.99	2.48	IRVN(306)TIS	TISAGPLGSR	MOD1 (MOSAIC DEATH 1)	64	45	0.01	1.26E-4	0.292
AT2G16600	711.332	2	7.33	4.51	MELY(27)ADT	ADTTPETAENFR	ROTAMASE CYP3	46	41	0.01	3.16E-3	-0.985
AT2G16600	527.816	2	6.27	4.16	VFGQ(140)VVE	VVEGLNVVR	ROTAMASE CYP3	48	43	0.01	3.16E-3	0.000
AT2G20260	724.861	2	3.43	2.79	ISTN(135)NYA	NYALDEVVEVK	PSAE-2 (photosystem I subunit E-2)* catalytic	65	44	0.01	7.94E-5	0.829
AT2G27710	697.753	3	5.59	3.89	IELL(43)LKE	LKEVKGKDLAELIAAGR	60S acidic ribosomal protein P2 (RPP2B)	70	39	0.01	7.94E-6	-0.718
AT2G28000	824.142	3	3.07	2.54	LTTQ(565)AIV	AIVVDKPKPKAPAAAPEGLMV	CPN60A (CHAPERONIN-60ALPHA)	51	41	0.01	1.00E-3	-2.208
AT2G30930	709.389	2	3.26	2.67	VSTQ(77)AKD	AKDAVDKAFSR	unknown protein	52	45	0.01	2.00E-3	-1.059
AT2G37220	901.889	2	3.36	2.75	PRSS(185)FGS	FGSSGSGYGGGGSGAGSGNR	29 kDa ribonucleoprotein, chloroplast	73	36	0.01	2.00E-6	-0.777
AT2G39730	890.977	2	3.81	3.03	MMSA(192)GEL	GELESGNAGEPAKLIR	RCA (RUBISCO ACTIVASE)	78	44	0.01	3.98E-6	-1.405
AT2G39730	855.895	2	3.02	2.50	RWRG(74)LAY	LAYDTSDDQDITR	RCA (RUBISCO ACTIVASE)	78	41	0.01	2.00E-6	-0.526
AT2G39730	806.919	2	8.07	4.73	AETY(417)LSQ	LSQAALGDANADAIGR	RCA (RUBISCO ACTIVASE)	80	44	0.01	2.51E-6	-1.613
AT2G39730	701.375	2	3.29	2.70	YGNM(400)LVM	LVMEQENVKR	RCA (RUBISCO ACTIVASE)	50	45	0.01	3.16E-3	-0.785
AT2G39730	1146.212	3	4.64	3.48	VNNQ(249)MVN	MVNATLMNIADNPTNVQLPGMYNKEENAR	RCA (RUBISCO ACTIVASE)	48	41	0.01	2.00E-3	-0.902
AT2G39730	933.974	2	6.73	4.32	IVTL(334)VDQ	VDQFPGQSIDFFGALR	RCA (RUBISCO ACTIVASE)	92	44	0.01	1.58E-7	-1.179
AT2G39730	762.954	2	5.27	3.76	VRKF(361)VES	VESLGVKEKIGKR	RCA (RUBISCO ACTIVASE)	52	42	0.01	1.00E-3	-0.591
AT3G14415	566.330	2	3.09	2.55	FKAL(299)ALG	ALGASGIFIGR	(S)-2-hydroxy-acid oxidase, peroxisomal, putative	73	43	0.01	1.00E-5	0.088
AT3G15356	951.995	2	2.77	2.31	DGSN(32)LLF	LLFLGDAELGPSSDGVSR	legume lectin family protein	91	45	0.01	2.51E-7	-1.209
AT3G44310	871.958	2	3.02	2.50	SVTF(334)VTK	VTKVEKAEDDSNK	NIT1 indole-3-acetonitrile nitrilase	54	45	0.01	1.26E-3	-1.435
AT3G62960	707.402	2	5.83	3.99	NDPD(49)CRE	CREIEKALVR	glutaredoxin family protein	47	44	0.01	5.01E-3	-5.379
AT4G24770	911.448	2	3.96	3.11	FVTM(199)SSV	SSVDEAETAVEKFNRR	RBP31 (31-KDA RNA BINDING PROTEIN)	79	43	0.01	2.51E-6	-1.263
AT4G32260	1060.897	3	3.66	2.94	AARA(152)EIA	EIAAALNKMKKETQVEVEEKLAEGR	ATP synthase family	64	44	0.01	1.00E-4	9.561
AT5G20720	948.503	2	3.78	3.01	VGIL(153)ETE	ETEDIKDLKPLNDR	CPN20 (CHAPERONIN 20)	49	44	0.01	3.16E-3	-1.055
ATCG00480	675.833	2	3.22	2.64	YQPT(302)LST	LSTEMGTLQER	chloroplast-encoded gene for beta subunit of ATP synthase	62	43	0.01	1.26E-4	-0.741

Table S2. List of endogenous AtMC1 substrates identified in the OE proteome. Peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the proteome of *AtMC1* overexpression. In the upper part of the table proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the middle part, peptides cleaved at other residues are listed. The bottom part of the table contains peptides that were uniquely found in the WT proteome and are indicative of indirect AtMC1 proteolysis. Amino acid position of the P1' site is shown

between brackets in the P4-P3' column. Fold times of higher OE ion intensity in comparison to the WT ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; N/A; not applicable refers to singleton ions.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score (s)	threshold (t)	Confidence	E value	Mass error (ppm)
Proteolysis of Arg or Lys												
AT2G21330	599.797	2	2.55	2.72	LASR(174)TAA	TAAYYQQGAR	fructose-bisphosphate aldolase, putative	53	43	0.01	1.00E-3	-0.418
AT3G11630	462.211	2	4.25	4.21	GIGR(219)SVD	SVDETMR	2-cys peroxiredoxin, chloroplast (BAS1)	48	40	0.01	1.58E-3	0.108
AT4G04770	451.733	2	2.26	2.37	FKLR(53)ADV	ADVGIDSR	ATABC1 (ATP BINDING CASSETTE PROTEIN 1)	50	44	0.01	2.51E-3	-0.333
AT4G35630	488.738	2	2.31	2.44	DGVR(64)SGS	SGSVGSQER	PSAT phosphoserine aminotransferase	46	43	0.01	5.01E-3	0.205
AT4G36945	643.283	2	2.39	2.53	CSSR(260)SES	SESSSLDTMSR	phospholipase C	49	38	0.01	7.94E-4	-0.389
Proteolysis of other amino acids												
AT1G09340	823.464	2	5.48	4.95	YVSA(51)SSE	SSEKILIMGGTR	CRB (CHLOROPLAST RNA BINDING)	88	44	0.01	3.98E-7	-1.277
AT1G12900	706.935	2	2.48	2.65	VTEA(64)KIK	KIKVAINGFGR	GAPA-2 (glyceraldehyde 3-phosphate dehydrogenase A subunit 2)	61	39	0.01	6.31E-5	-0.283
AT1G67740	588.778	2	2.70	2.89	EAYA(144)AAE	AAEAAAASSDSR	PSBY (Photosystem II BY)	90	42	0.01	1.58E-7	-0.255
AT1G77480	842.966	2	4.83	4.58	TSEA(35)TKD	TKDSSAQVKLQNR	Aspartyl protease-like protein	74	44	0.01	1.00E-5	-1.069
AT2G05990	970.503	2	2.39	2.53	TRAM(77)SES	SESSENKAPSGLPIDLR	MOD1 (MOSAIC DEATH 1)	72	44	0.01	1.58E-5	-1.135
AT2G15620	880.920	2	2.61	2.79	LVAA(29)AQT	AQTTAPAESTASVDADR	NIR1 (NITRITE REDUCTASE 1)	92	42	0.01	1.00E-7	-1.136
AT2G25840	836.942	2	2.94	3.14	RCCC(53)SVA	SVATDDTSPSVKRR	OVA4 (ovule abortion 4)	53	48	0.01	3.16E-3	-1.436
AT3G02875	734.850	2	3.27	3.45	LSSA(24)GSY	GSYDSGSGLESLAR	ILR1 (IAA-LEUCINE RESISTANT 1)	55	42	0.01	5.01E-4	-1.226
AT3G09080	609.820	2	N/A	N/A	VFVQ(968)SIS	SISELSTASPR	transducin family protein / WD-40 repeat family protein	56	45	0.01	7.94E-4	5.877
AT3G15520	711.913	2	2.26	2.37	PAEA(115)VLY	VLYSPDTKVPR	peptidyl-prolyl cis-trans isomerase TLP38, chloroplast	63	43	0.01	1.00E-4	1.055
AT3G21055	564.283	2	N/A	N/A	KVAM(72)AEE	AEEEEPKR	PSBTN (photosystem II subunit T)	54	43	0.01	7.94E-4	0.000
AT4G20360	904.531	2	2.97	3.16	GSAL(244)LAV	LAVETLTENPKVKR	ATRAE1B (ARABIDOPSIS RAB GTPASE HOMOLOG E1B)	85	41	0.01	3.98E-7	-0.609
AT4G35630	665.852	2	2.42	2.57	AVRC(52)VAS	VASTTQVDGVR	PSAT phosphoserine aminotransferase	68	45	0.01	5.01E-5	-0.301
AT5G08670	1211.940	3	2.58	2.76	VAEY(52)ATS	ATSSPASSAAPSSAPAKDEGKTYDYGKGAIGR	ATP binding / hydrogen ion transporting ATP synthase	76	43	0.01	5.01E-6	5.946
AT5G30510	710.859	2	2.50	2.67	TIVA(45)AVA	AVAMSSGQTKER	RPS1 (RIBOSOMAL PROTEIN S1)	80	44	0.01	2.51E-6	-0.211
AT5G35630	602.302	2	2.84	3.03	RVLA(52)LQS	LQSDNSTVNR	GS2 (GLUTAMINE SYNTHETASE 2)	81	44	0.01	2.00E-6	-0.166
Singletons in the WT proteome												
AT1G13930	702.415	2	N/A	N/A	LMAS(44)AKV	AKVVAEAAQAAAR	Involved in response to salt stress	46	41	0.01	3.16E-3	-0.713
AT5G41060	463.208	2	N/A	N/A	GKAN(333)DDI	DDIEMGR	zinc finger (DHHC type) family protein	42	39	0.01	5.01E-3	4.219

Table S3. List of endogenous AtMC1 substrates identified in the WT proteome upon *Pto* DC3000(*avrRpm1*) treatment. Peptides were identified either as unique or significantly ($p < 0.02$) enriched ions in the WT proteome. In the upper part of the table proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the middle part, peptides cleaved at other residues are listed. The bottom part of the table contains peptides that were uniquely found in the KO proteome and are indicative of indirect AtMC1 proteolysis. Amino acid position of the P1' site is

shown between brackets in the P4-P3' column. Fold times of higher WT ion intensity in comparison to the KO ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; N/A; not applicable refers to singleton ions.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence		score (s)	threshold (t)	Confidence	E value	Mass error (ppm)
Proteolysis of Arg or Lys												
AT1G07135	450.238	2	2.14	3.61	VVIK(64)KGG	KGGGGGGGGR	glycine-rich protein	64	43	0.01	7.94E-5	0.890
AT1G79340	467.756	2	2.40	4.07	LRSK(183)VEG	VEGAIESR	metacaspase 4	57	46	0.01	7.94E-4	-1.500
AT4G10480	874.420	2	3.21	5.18	GEAK(123)IDD	IDDMSSQLQAQAQR	Nascent polypeptide-associated complex (NAC) protein	45	43	0.01	6.31E-3	-2.404
AT4G21140	450.238	2	1.89	3.11	LTIK(49)KGG	KGGGNGGGR	unknown protein	52	43	0.01	1.26E-3	0.445
AT5G28840	747.937	2	1.64	2.49	GSSK(358)VVG	VVGTQAPVQLGSLR	GDP-D-mannose 3',5'-epimerase	56	43	0.01	5.01E-4	-2.343
Proteolysis of other amino acids												
AT1G70940	530.745	2	1.62	2.42	NVFG(379)GAP	GAPDNDQGGR	PIN3, Auxin efflux carrier family protein	42	41	0.01	7.94E-3	1.605
AT1G74670	1073.489	2	1.94	3.21	YHPE(32)SYG	SYGPGSLKSYQCGGQCTR	Gibberellin-regulated family protein	77	40	0.01	2.00E-6	-0.606
AT1G75750	1017.172	3	1.80	2.89	LVQA(24)DVE	DVENSQKKNGYAKKIDCGSACVAR	GAST1 protein homolog 1	46	43	0.01	5.01E-3	5.332
AT2G05520	542.726	2	2.05	3.45	GGRQ(107)GGG	GGGGSGGSYCR	glycine-rich protein 3	51	35	0.01	2.51E-4	1.938
AT2G37220	901.889	2	2.40	4.08	PRSS(185)FGS	FGSSSGYGGGGSGAGSGNR	RNA-binding (RRM/RBD/RNP motifs) family protein	65	37	0.01	1.58E-5	-1.221
AT2G38470	737.362	2	N/A	N/A	KGDN(333)ETN	ETNGGNGGSKTVR	WRKY DNA-binding protein 33	46	43	0.01	5.01E-3	8.556
AT2G39730	931.108	3	1.66	2.53	NATL(255)MNI	MNLADNPTNVQLPGMYNKEENAR	rubisco activase	86	42	0.01	3.98E-7	-2.329
AT2G39730	644.833	2	1.63	2.45	GNML(401)VME	VMEQENVKR	RCA (RUBISCO ACTIVASE)	45	44	0.01	7.94E-3	-1.398
AT3G11560	784.390	2	1.73	2.71	LVLA(74)SAE	SAEDGVAINGSPQPR	LETM1-like protein	45	44	0.01	7.94E-3	-0.830
AT3G14415	527.254	2	2.59	4.37	KMAH(86)PDG	PDGEYATAR	Aldolase-type TIM barrel family protein	56	43	0.01	5.01E-4	1.805
AT3G24255	520.335	2	N/A	N/A	YGPL(46)VEK	VEKIRVR	RNA-directed DNA polymerase (reverse transcriptase) protein	44	41	0.01	5.01E-3	0.289
AT3G53460	765.825	2	1.96	3.25	SQRS(242)SYG	SYGSGSGSGSGSGSGNR	chloroplast RNA-binding protein 29	105	36	0.01	1.26E-9	-0.392
AT4G02380	645.327	2	2.65	4.46	RRGY(29)AAT	AATAAQGSVSSGGR	SAG21 senescence-associated gene 21	94	44	0.01	1.00E-7	-1.242
AT4G09650	417.248	2	1.77	2.83	DASL(195)VAG	VAGFTIR	ATP synthase delta-subunit gene	49	45	0.01	3.98E-3	-0.961
AT4G12970	852.381	2	1.59	2.34	RRHM(58)JGS	IGSTAPTCTYNECR	stomagen	49	39	0.01	1.00E-3	0.059
AT4G24280	924.419	2	3.21	5.18	FEVL(291)STS	STSGDTHLGGDDFDKFR	chloroplast heat shock protein 70-1	75	40	0.01	3.16E-6	5.848
AT4G26300	749.400	2	1.64	2.50	VVAM(54)AAN	AANEEFTGNLKR	Arginyl-tRNA synthetase, class Ic	68	45	0.01	5.01E-5	-0.601
AT4G28080	839.377	2	1.82	2.95	SFPN(1654)STE	STESNGEANQFNGPR	Tetratricopeptide repeat (TPR)-like superfamily protein	65	39	0.01	2.51E-5	-0.537
AT4G28640	434.785	2	N/A	N/A	SRAS(71)VIA	VIAGIKR	indole-3-acetic acid inducible 11	56	42	0.01	3.98E-4	-0.231
AT5G14540	835.878	2	1.73	2.71	QGLP(459)MAS	MASAISSGGSGGSDSPR	Protein of unknown function (DUF1421)	55	40	0.01	3.16E-4	-0.599
AT5G42530	618.332	2	1.78	2.84	KALG(28)AQE	AQEEAQKQR	unknown protein	64	45	0.01	1.26E-4	1.701
ATCG00140	1019.040	2	2.29	3.89	AVGL(20)ASI	ASIGPGVQGTAAAGQAVEGIAR	ATP synthase subunit C family protein	90	45	0.01	3.16E-7	0.147
Singletons in the KO proteome												
AT1G03880	650.309	2	N/A	N/A	EIAN(270)GLE	GLEETLCTMR	cruciferin 2	51	42	0.01	1.26E-3	-0.847
AT1G53240	446.731	2	N/A	N/A	RRSF(23)SSG	SSGSVPER	Lactate/malate dehydrogenase family protein	59	44	0.01	3.16E-4	1.234
AT2G19560	503.319	2	N/A	N/A	QLKL(353)EGI	EGIAKALR	proteasome family protein	40	39	0.01	7.94E-3	-0.299
AT3G22231	865.451	3	N/A	N/A	NYFS(11)VQK	VQKPSSETSSGPTYSPPIGYPTR	pathogen and circadian controlled 1	55	45	0.01	1.00E-3	-1.041
AT4G20960	440.262	2	N/A	N/A	GQPH(122)AEV	AEVFAIR	Cytidine/deoxycytidylate deaminase family protein	60	43	0.01	2.00E-4	-4.667
AT4G28520	643.301	2	N/A	N/A	PQGN(334)GLE	GLEETICSMR	cruciferin 3	58	41	0.01	2.00E-4	-0.311
AT5G44120	605.304	2	N/A	N/A	RHGN(96)GLE	GLEETICSAR	RmlC-like cupins superfamily protein	49	44	0.01	3.16E-3	-3.144
AT5G66600	639.341	2	N/A	N/A	MGFG(5)VGG	VGGGGGRMLDLR	Protein of unknown function, DUF547	50	45	0.01	3.16E-3	6.110

Table S4. List of endogenous AtMC1 substrates identified in the OE proteome upon *Pto* DC3000(*avrRpm1*) treatment. Peptides were identified either as unique or significantly ($p < 0.05$) enriched ions in the OE proteome. In the upper part of the table proteolysis-reporter peptides generated after cleavage of Arg or Lys residues are listed, while at the middle part, peptides cleaved at other residues are listed. The bottom part of the table contains peptides that were uniquely found in the KO proteome and are indicative of indirect AtMC1 proteolysis. Amino acid position of the P1' site is shown between

brackets in the P4-P3' column. Fold times of higher OE ion intensity in comparison to the KO ion counterpart is shown in the column 'Intensity increase fold'. M/Z; mass/charge value of identified peptides, Z; ion charge, Score; Mascot ion score, threshold; Identity threshold score, confidence; confidence level for correct sequence annotation, E-value; expectation value; N/A; not applicable refers to singleton ions.

Accession	M/Z of best scoring peptide	Z	Intensity fold increase	z-score	P4-P3'	Sequence	Description	score (s)	threshold (t)	Confidence	E value	Mass error (ppm)
Proteolysis of Arg or Lys												
AT1G14980	463.277	2	6.49	2.19	NSGK(41)VIA	VIAVGPGSR	chaperonin 10	57	42	0.01	3.16E-4	-0.32
Proteolysis of other amino acids												
AT1G05070	584.734	2	7.39	2.34	KCNS(132)GME	GMETCEEAR	Protein of unknown function (DUF1068)	34	33	0.01	7.94E-3	0.60
AT1G29660	590.266	2	6.49	2.19	GFTN(295)TNT	TNTACCGIGR	GDSL-like Lipase/Acylhydrolase superfamily protein	52	40	0.01	6.31E-4	0.34
AT2G05520	617.782	2	5.75	2.05	GGNY(72)QGG	QGGGGNYQGGGGR	glycine-rich protein 3	66	40	0.01	2.51E-5	0.16
AT2G05520	542.727	2	6.54	2.20	GGRQ(107)GGG	GGGGSGGSYCR	glycine-rich protein 3	98	36	0.01	6.31E-9	0.65
AT2G19730	486.298	2	6.35	2.17	LARL(123)SAI	SAISKGLR	Ribosomal L28e protein family	45	43	0.01	6.31E-3	0.00
AT2G38280	567.244	2	6.28	2.16	ATPF(84)TDG	TDGGGGGGDTGR	AMP deaminase, putative	40	37	0.01	5.01E-3	0.97
AT3G15356	551.753	2	5.79	2.06	LQNM(248)YAG	YAGFAGSMGR	Legume lectin family protein	63	40	0.01	5.01E-5	-0.09
AT3G47070	588.299	2	9.37	2.62	KKVD(60)EKE	EKEGTTGGR	unknown protein	58	44	0.01	3.98E-4	0.00
AT4G36860	516.728	2	N/A	N/A	RAQL(98)EAA	EAAEEEEER	LIM domain-containing protein	40	38	0.01	6.31E-3	0.29
AT5G38120	434.287	2	N/A	N/A	ALGT(268)TVV	TVVILPR	AMP-dependent synthetase and ligase family protein	40	38	0.01	6.31E-3	-0.58
AT5G42530	514.772	2	N/A	N/A	LGAQ(30)EEA	EEAQKQR	unknown protein	47	44	0.01	5.01E-3	0.00
Singletons in the KO proteome												
AT3G05680.1	427.225	2	N/A	N/A	RKLP(1517)TLE	TLESSSR	embryo defective 2016	56	43	0.01	5.01E-4	3.52

REFERENCES

- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146-158.
- Chivasa, S., Tome, D.F., Hamilton, J.M., and Slabas, A.R. (2011). Proteomic analysis of extracellular ATP-regulated proteins identifies ATP synthase beta-subunit as a novel plant cell death regulator. *Mol Cell Proteomics* 10, M110 003905.
- Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 6, 786-787.
- Coll, N.S., Epple, P., and Dangl, J.L. (2011). Programmed cell death in the plant immune system. *Cell Death Differ* 18, 1247-1256.
- Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. *Science* 330, 1393-1397.
- Dietrich, R.A., Delaney, T.P., Uknes, S.J., Ward, E.R., Ryals, J.A., and Dangl, J.L. (1994). Arabidopsis mutants simulating disease resistance response. *Cell* 77, 565-577.
- Fuentes-Prior, P., and Salvesen, G.S. (2004). The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem J* 384, 201-232.
- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol* 21, 566-569.
- Grant, M.R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R.W., and Dangl, J.L. (1995). Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. *Science* 269, 843-846.
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008, 420747.
- Huber, P. (1981). *Robust Statistics*. John Wiley and Sons, Inc, New York.
- Kaminaka, H., Nake, C., Epple, P., Dittgen, J., Schutze, K., Chaban, C., Holt, B.F., 3rd, Merkle, T., Schafer, E., Harter, K., et al. (2006). bZIP10-LSD1 antagonism modulates basal defense and cell death in Arabidopsis following infection. *Embo J* 25, 4400-4411.
- Lee, R.E., Brunette, S., Puente, L.G., and Megeney, L.A. (2010). Metacaspase Yca1 is required for clearance of insoluble protein aggregates. *Proc Natl Acad Sci U S A* 107, 13348-13353.
- Mangeon, A., Junqueira, R.M., and Sabetto-Martins, G. (2010). Functional diversity of the plant glycine-rich proteins superfamily. *Plant Signal Behav* 5, 99-104.
- McLuskey, K., Rudolf, J., Proto, W.R., Isaacs, N.W., Coombs, G.H., Moss, C.X., and Mottram, J.C. (2012). Crystal structure of a Trypanosoma brucei metacaspase. *Proc Natl Acad Sci U S A* 109, 7469-7474.
- Overall, C.M., Tam, E.M., Kappelhoff, R., Connor, A., Ewart, T., Morrison, C.J., Puente, X., Lopez-Otin, C., and Seth, A. (2004). Protease degradomics: mass spectrometry discovery of protease substrates and the CLIP-CHIP, a dedicated DNA microarray of all human proteases and inhibitors. *Biol Chem* 385, 493-504.
- Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res* 38, D227-233.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. *Nat Genet* 37, 501-506.
- Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. *Nat Protoc* 6, 1130-1141.
- Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zavalov, A., Lam, E., and Bozhkov, P.V. (2011). Metacaspases. *Cell Death Differ* 18, 1279-1288.
- Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol Cell* 6, 961-967.
- Vercammen, D., Declercq, W., Vandenaabeele, P., and Van Breusegem, F. (2007). Are metacaspases caspases? *J Cell Biol* 179, 375-380.
- Watanabe, N., and Lam, E. (2005). Two Arabidopsis metacaspases AtMCP1b and AtMCP2b are arginine/lysine-specific cysteine proteases and activate apoptosis-like cell death in yeast. *J Biol Chem* 280, 14691-14699.
- Watanabe, N., and Lam, E. (2011a). Arabidopsis metacaspase 2d is a positive mediator of cell death induced during biotic and abiotic stresses. *Plant J* 66, 969-982.
- Watanabe, N., and Lam, E. (2011b). Calcium-dependent activation and autolysis of Arabidopsis metacaspase 2d. *J Biol Chem* 286, 10027-10040.
- Wong, A.H., Yan, C., and Shi, Y. (2012). Crystal structure of the yeast metacaspase Yca1. *J Biol Chem*.

Conclusions and Perspectives

227

The pioneering aspect of this study is the implementation of positional proteomics and mass spectrometry for the degradome profiling of plant proteases. Methods like COFRADIC and TAILS have been successfully used in mammalian systems where new and already known protease substrates were catalogued (Van Damme et al., 2005; Vande Walle et al., 2007; Van Damme et al., 2009; Doucet et al., 2011; de Poot et al., 2011; Starr et al., 2012 etc). Plant protease degradomics is a rather unexplored research area and positional proteomics can contribute greatly to our understanding of protease biology in plants. In this way, more rational and precise use of inhibitors can be achieved and control of the regulatory pathways governed by proteases can find practical applications in the improvement of crop performance.

The research performed in this PhD study aimed at the elucidation of metacaspase functions in *Arabidopsis thaliana*. We chose to work with this model species as it encodes both types of metacaspases (type-I and II). Metacaspase 9 (AtMC9) and metacaspase 1 (AtMC1) were our subject proteases and we sought to identify their protein substrates at a proteome-wide scale using positional proteomics (COFRADIC).

AtMC9 substrate profiling in seedlings

AtMC9 is the best biochemically described *Arabidopsis* metacaspase and it is highly expressed in developing seeds and flowers. Its involvement in plant development was studied in young seedlings, following germination, as this is an early metabolically active stage. AtMC9 substrate discovery was achieved using N-terminal COFRADIC. Three independent AtMC9 degradome analyses revealed a set of 74 potential physiological *Arabidopsis* AtMC9 substrates. Furthermore, the selected proteins were cleaved at the same Arg or Lys site in at least two of the performed *in vivo* and *in vitro* degradome analyses.

Proteins involved in post-embryonic development such as several late embryogenesis abundant (LEA) proteins, were frequently found among the cleaved proteins. Further, multiple transcription and translation elongation factors, ribosomal proteins and structural components were found cleaved by AtMC9. Involvement of AtMC9 in primary plant metabolism was inferred by the observed cleavage of key metabolic enzymes like the citrate synthase-3 (CSY3) and the phosphoenolpyruvate carboxykinase-1. Using *in vitro* transcribed and translated (TnT) proteins and synthetic 18-mer peptides, we validated AtMC9 cleavage of GRF5, PEPCK1, CSY3, GMD2, GRP7, PORB, trypsin inhibitor-2, NAD synthetase, a PPR-containing protein and a LEA domain-containing protein *in vitro*. Furthermore, we demonstrated that PEPCK1 is a biologically relevant AtMC9 substrate as the carboxylase activity of PEPCK1, in plants overexpressing AtMC9, is significantly increased ($p < 0.05$). Consequently, *atcm9* plants are growth-compromised in the absence of sucrose when grown in the dark, a phenotype reminiscent of *atpepck1* plants grown under the same conditions (Penfield et al., 2004). Moreover, during transient expression of AtMC9 in tobacco (*Nicotiana benthamiana*) leaves, it was shown that AtMC9 has a nucleocytoplasmic localisation, while PEPCK1 is exclusively found in the cytosol, where its AtMC9-mediated cleavage could occur. AtMC9 subcellular localisation was further confirmed in *Arabidopsis* roots using a reporter AtMC9 construct fused to its endogenous promoter. From the above, we concluded that PEPCK1 is *in vivo* cleaved by AtMC9 and cleavage restores PEPCK1 enzymatic activity after the latter has been phosphorylated. Although phosphorylation sites of PEPCK1 homologs from other plant species are conserved in the *Arabidopsis* protein, it yet remains unknown whether AtPEPCK1 undergoes phosphorylation. Besides this, fatty acid and metabolic profiling of *AtMC9* gain- and loss-of-function mutants appear as the next topics

for further investigation, as PEPCK1 is the key enzyme for the conversion of lipids to sucrose (Dittrich et al., 1973). Evidently, assays for *in vivo* validation of the remaining 73 potential AtMC9 substrates, might further elucidate the functional roles of this metacaspase.

A strong asset of positional proteomics, in this case N-terminal COFRADIC, is the power for exact determination of protein cleavage sites. In this manner, protease specificity can be profiled both *in vitro* and *in vivo*, using proteome extracts or proteome-derived peptide libraries (Schilling et al., 2008). As such, we confirmed the previously *in vitro* studied AtMC9 specificity for Arg or Lys and the preference for a second basic amino acid residue at the P3 substrate position (Vercammen et al., 2006). Moreover, we extended the specificity consensus sequence to the prime site (C-terminal to the cleavage bond) and concluded that AtMC9's specificity for basic residues at P1 comes along with a strong preference for acidic amino acids at the P1' position followed by aliphatic residues at P2'. Finally, Lys was relatively preferred over Arg at P1 position (Figure 1).

AtMC9 substrate profiling in roots

Caspases are, among other processes involved in regulating cell adhesion by promoting cell motility and they are related to metacaspases based on their catalytic domain (Uren et al., 2000). Using the *Arabidopsis* roots as a study system, the spatial expression of all metacaspases was explored and it was shown that *AtMC5* and *AtMC9* are expressed in tissues where cell separation is a prerequisite for root development. Genetic perturbation of *AtMC5* and/or *AtMC9* using T-DNA insertions, did not lead to root morphologic aberrations. Identification of chemical compounds that inhibit AtMC9 VRPRase activity *in vitro* led to the discovery of two thioxodihydropyrimidinedione (TDP)-containing compounds that affect root development as a consequence

of blocked cell separation. Besides AtMC9, AtMC5 was significantly inhibited by the identified compounds ($p < 0.05$). Artificial micro-RNA lines (Schwab et al., 2006) with variable levels of down-regulated type-II metacaspases (*AtMC4* to *AtMC9*) mimicked the observed root phenotypes caused by the two TDP compounds. In this manner, it became clear that metacaspases are important for cell separation. We believe that the reason why *AtMC5* and *AtMC9* single and double knockouts did not show a root phenotype is due to the plant's adaptation to the caused mutations by regulating other proteins, possibly metacaspases.

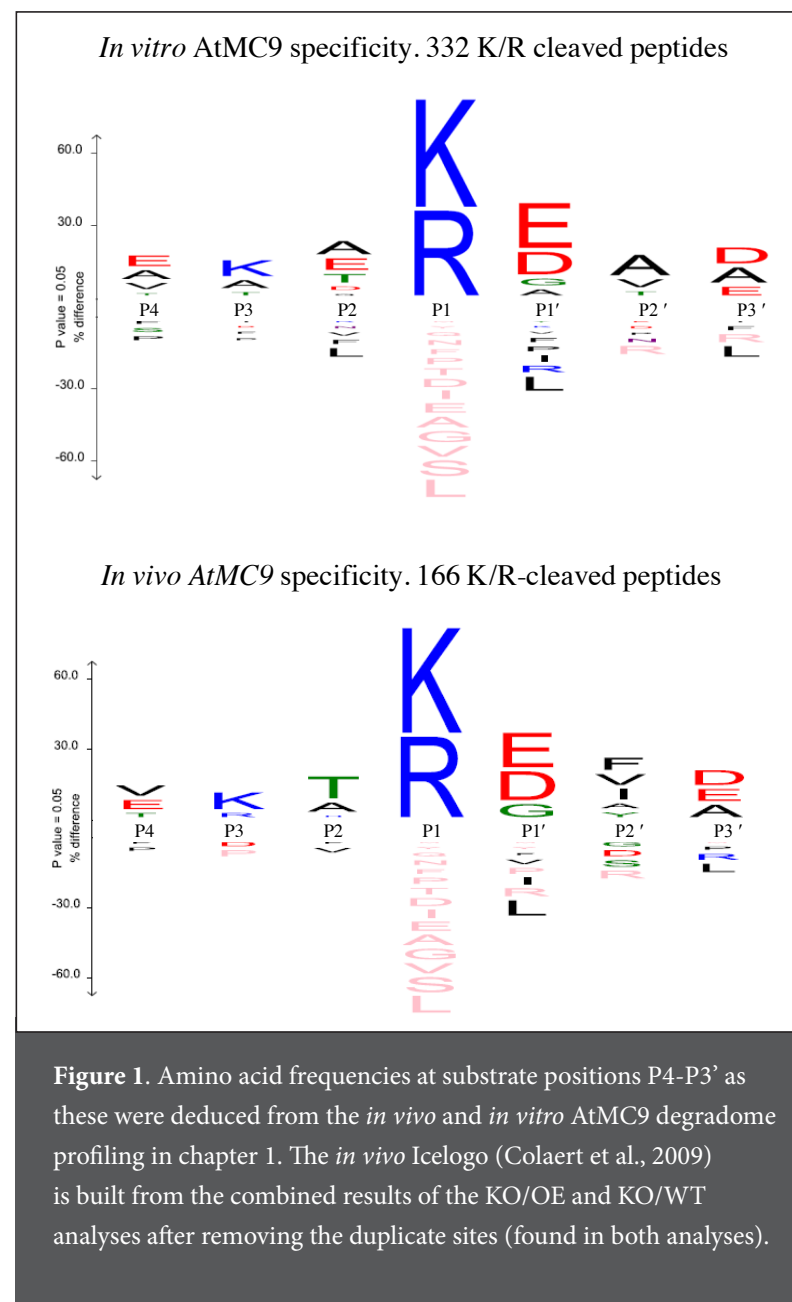
To disentangle further the involvement of metacaspases in cell separation we aimed at the identification of their protein targets. Given that *AtMC9* is the predominantly expressed metacaspase in sites of lateral root protrusion and root cap, we explored the AtMC9 degradome in plant roots. Using N-terminal COFRADIC in biologically-relevant experimental setups, 166 potential AtMC9 substrates were identified, 50 of which are co-expressed with *AtMC9*. Among them are several glycosyl hydrolases, peroxidases and components of the cell cytoskeleton, and their regulation by AtMC9 could lead to alteration of the cell wall. Evidently, *in vivo* or *in vitro* cleavage validation of selected protein substrates and downstream biochemical assays will explain in more detail the AtMC9 mode-of-action during development and in particular cell separation. Subsequently, phenotypic analysis of plants with altered levels of selected AtMC9 candidate substrates would delineate further the process of cell separation. The role of metacaspases in cell separation occurring at other locations in the plant such as abscission zones (positions of plant organ shedding) is an interesting future perspective for research.

AtMC1 substrate profiling in leaves upon pathogen-triggered cell death

Lesion simulating disease 1 (LSD1) is a negative regulator of cell death during the plant hypersensitive response (HR) to pathogen challenge (Dietrich et al., 1994) and evidence supports the LSD1 and AtMC1 interaction in yeast, *Arabidopsis* and *Nicotiana benthamiana* (Coll et al., 2011). In the same study, AtMC1 was shown to positively regulate pathogen-triggered HR cell death in *Arabidopsis*, while AtMC2 had an opposite effect.

Given the importance of AtMC1 as a positive regulator of cell death in this regulatory module, we analyzed its degradome. We assessed those proteins that were differentially cleaved in the *Arabidopsis* leaf proteomes of *atmc1* knockout, wild type (Col-0) and stable transgenic plants conditionally overexpressing AtMC1, with and without a cell death trigger; *Pseudomonas syringae* pv. tomato (*Pto*) DC3000 (*avrRpm1*). We identified 69 proteins as potential AtMC1 substrates and 20 of them were moreover transcriptionally co-regulated with AtMC1 upon biotic stimuli, hormone treatment and in overall *Arabidopsis* anatomy. Among those, a type-II metacaspase (AtMC4), an Auxin efflux carrier protein (PIN3), a senescence-associated protein (SAG21) and the defence-related proteins WYRK33 and MOD1 were present (Hruz et al., 2008).

Based on *in vitro* data, models for the activation mechanism of metacaspases have been proposed (Lam and Zhang, 2012). However, our knowledge about it *in vivo* remains poor. AtMC1, based on its long prodomain, can be associated to the initiator caspases, whereas AtMC4 that lacks this prodomain is similar to the effector caspases. Our results showed that AtMC4 is potentially cleaved *in vivo* by AtMC1 upon pathogen infection. In accordance, with the fact that effector caspases become activated upon cleavage by the initiator caspases (Pop and Salvesen, 2009), our results hint towards type-I mediated activation of type-II metacaspases and this is worth of attention and further consolidation. Supporting this, both AtMC1 and AtMC4 activities are induced



by Ca²⁺ and both proteases are involved in pathogen-triggered cell death responses (Coll et al., 2011; Watanabe and Lam., 2011a and 2011b).

The study of protease specificity *in vivo* can be elusive due to fragment instability in the cells and the action of other (downstream) proteases. For these reasons, the AtMC1 *in vivo* degradome studies did not directly point to any striking amino acid preference. However, the most prominently processed amino acid residue at P1 position, in half of our degradome analyses, was Gln. Based on the protein aggregation triggering properties of poly-Gln motifs (Alberti et al., 2009) and the involvement of another type-I metacaspase, Yca1 in yeast, in the clearance of protein aggregates (Lee et al., 2010), we envisage the investigation of AtMC1 involvement in proteostasis as very informative.

Production of active recombinant AtMC1 is vital for the validation of the findings during this study. Implementation of the active enzyme in cleavage assays of selected protein substrates and further assessment of AtMC1 specificity is necessary. Given the here suggested potential AtMC1 function in proteostasis, assessment of protein aggregation in different backgrounds of AtMC1 activity are important perspectives for the continuation of the described research findings.

In summary, the results of this PhD study learned us that positional proteomics can be used to profile the degradome and the specificity of plant proteases in whole tissue extracts. In this manner, we discovered several protein substrates of type-I and type-II metacaspases that highlight the multitude of their functions. Lastly, extrapolation of the knowledge generated here for the study of parasitic protozoa metacaspases could potentially assist in the drug discovery against human diseases caused by *Trypanosoma* and *Leishmania species* such as the sleeping sickness, the Chagas disease and Leishmaniasis.

REFERENCES

- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146-158.
- Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 6, 786-787.
- Coll, N.S., Vercammen, D., Smidler, A., Clover, C., Van Breusegem, F., Dangl, J.L., and Epple, P. (2010). Arabidopsis type I metacaspases control cell death. *Science* 330, 1393-1397.
- de Poot, S.A., Westgeest, M., Hostetter, D.R., Van Damme, P., Plasman, K., Demeyer, K., Broekhuizen, R., Gevaert, K., Craik, C.S., and Bovenschen, N. (2011). Human and mouse granzyme M display divergent and species-specific substrate specificities. *Biochem J* 437, 431-442.
- Dietrich, R.A., Delaney, T.P., Uknes, S.J., Ward, E.R., Ryals, J.A., and Dangl, J.L. (1994). Arabidopsis mutants simulating disease resistance response. *Cell* 77, 565-577.
- Dittrich, P., Campbell, W.H., and Black, C.C. (1973). Phosphoenolpyruvate carboxykinase in plants exhibiting crassulacean Acid metabolism. *Plant Physiol* 52, 357-361.
- Doucet, A., Kleifeld, O., Kizhakkedathu, J.N., and Overall, C.M. (2011). Identification of proteolytic products and natural protein N-termini by Terminal Amine Isotopic Labeling of Substrates (TAILS). *Methods Mol Biol* 753, 273-287.
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics* 2008, 420747.
- Lam, E., and Zhang, Y. (2012). Regulating the reapers: activating metacaspases for programmed cell death. *Trends Plant Sci* 17, 487-494.
- Lee, R.E., Brunette, S., Puente, L.G., and Megeney, L.A. (2010). Metacaspase Yca1 is required for clearance of insoluble protein aggregates. *Proc Natl Acad Sci U S A* 107, 13348-13353.
- Penfield, S., Rylott, E.L., Gilday, A.D., Graham, S., Larson, T.R., and Graham, I.A. (2004). Reserve mobilization in the Arabidopsis endosperm fuels hypocotyl elongation in the dark, is independent of abscisic acid, and requires PHOSPHOENOLPYRUVATE CARBOXYKINASE1. *Plant Cell* 16, 2705-2718.
- Pop, C., and Salvesen, G.S. (2009). Human caspases: activation, specificity, and regulation. *J Biol Chem* 284, 21777-21781.
- Schilling, O., and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26, 685-694.
- Schwab, R., Ossowski, S., Riester, M., Warthmann, N., and Weigel, D. (2006). Highly specific gene silencing by artificial microRNAs in Arabidopsis. *Plant Cell* 18, 1121-1133.
- Starr, A.E., Bellac, C.L., Dufour, A., Goebeler, V., and Overall, C.M. (2012). Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J Biol Chem* 287, 13382-13395.
- Uren, A.G., O'Rourke, K., Aravind, L.A., Pisabarro, M.T., Seshagiri, S., Koonin, E.V., and Dixit, V.M. (2000). Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of

which plays a key role in MALT lymphoma. *Mol Cell* 6, 961-967.

- Van Damme, P., Martens, L., Van Damme, J., Hugelier, K., Staes, A., Vandekerckhove, J., and Gevaert, K. (2005). Caspase-specific and nonspecific in vivo protein processing during Fas-induced apoptosis. *Nat Methods* 2, 771-777.
- Van Damme, P., Maurer-Stroh, S., Plasman, K., Van Durme, J., Colaert, N., Timmerman, E., De Bock, P.J., Goethals, M., Rousseau, F., Schymkowitz, J., et al. (2009). Analysis of protein processing by N-terminal proteomics reveals novel species-specific substrate determinants of granzyme B orthologs. *Mol Cell Proteomics* 8, 258-272.
- Vande Walle, L., Van Damme, P., Lamkanfi, M., Saelens, X., Vandekerckhove, J., Gevaert, K., and Vandenabeele, P. (2007). Proteome-wide Identification of HtrA2/Omi Substrates. *J Proteome Res* 6, 1006-1015.
- Vercammen, D., Belenghi, B., van de Cotte, B., Beunens, T., Gavigan, J.A., De Rycke, R., Brackenier, A., Inze, D., Harris, J.L., and Van Breusegem, F. (2006). Serpin1 of *Arabidopsis thaliana* is a suicide inhibitor for metacaspase 9. *J Mol Biol* 364, 625-636.
- Watanabe, N., and Lam, E. (2011a). *Arabidopsis* metacaspase 2d is a positive mediator of cell death induced during biotic and abiotic stresses. *Plant J* 66, 969-982.
- Watanabe, N., and Lam, E. (2011b). Calcium-dependent activation and autolysis of *Arabidopsis* metacaspase 2d. *J Biol Chem* 286, 10027-10040.

The page features a decorative graphic consisting of three wavy, horizontal lines that span across the width of the page. The top line is a solid grey, while the two lines below it are dashed grey. These lines create a subtle, flowing background element.

Acknowledgements

237

ευχαριστώ!

With the completion of five exciting years of education in PSB, Rommelaere, VIB and Belgium, there are many people to thank for it. Following the chronological series of events, I would like first to thank VIB for selecting me out of many other candidates in 2007 for a PhD fellowship. During the second phase of selection, when we had to give a talk, I was the last candidate on the schedule. I will never forget my accumulating stress as I watched all others exit the room and my joy when the results were announced.

To my PhD promotors, Prof. Dr. Frank Van Breusegem and Prof. Dr. Kris Gevaert, thank you for the tutoring, fruitful discussions, support and bright ideas in difficult moments. The interdisciplinary project that you founded is excellent and I am very happy that I could be part of it. I would also like to express my gratitude to you for reading and correcting my thesis.

I would also like to thank all the members of my PhD examination and reading committee for accepting my invitation with great enthusiasm and for their advice.

I greatly enjoyed the privilege of working in two labs and I really learned A LOT. Of course, I would have never done this without the help of my valuable colleagues.

Evy, An and PJ thank you so much for guiding me through the maze of peaks and the initially “scary” world of mass-spectrometers, HPLC machines and complicated algorithms. Also for doing the MS/MS analysis of all my greenish samples and for answering to my questions.

Dominique V. and Petra, thank you for teaching me everything I know about proteases and in general enzymology. Brigitte, Cezary and Aurine for teaching me how to produce and purify hard-to-get proteins. Again, Brigitte and Debbie for all the technical assistance.

Silke, Cezary, Michael, Pelle, Vanessa, Iker, Frank H., Annelies, Inge, Lorin, Jordi, Hans, Mathias L., Francis, Bart G., Bart R., Kim and all members of the Cell Death and Oxidative Stress lab and Proteomics lab, thank you for the nice friendly chats and atmosphere in and outside the lab. Also for helping me become better acquainted with the Belgian culture!

To my cherished partner in life, Wim, thank you αγαπάκι μου for being so patient, loving and caring with me, throughout these five years and especially the last one.

To my family;

Μαμά, μπαμπά, Χρυσάνθη, οφείλω τα πάντα σε σας και στην αγάπη σας.

Με γεμίζετε δύναμη, θάρρος και αισιοδοξία και έτσι με εσάς στο πλάι μου μπορώ να καταφέρω τα πάντα στη ζωή. Σας είμαι τόσο ευγνώμων για όλα όσα με προοικίσατε!

**Ευχαριστώ / Thank you,
Liana**

Thank you

Curriculum Vitae

241

Name: **Styliani**
 Surname: **Tsiatsiani**
 Nationality: Greek

Academic

- 2012 – pres.** Doctor in Sciences, Biotechnology
 Department of Plant Biotechnology and Bioinformatics, Department of Biochemistry,
 Ghent University, Belgium. VIB Departments of Plant Systems Biology and Medical
 Protein Research
- 2007** Master of Science, Plant Biotechnology, Functional Plant Genomics.
 Wageningen University, the Netherlands.
- 2005** Bachelor in Crop Sciences.
 School of Agricultural Technology at the Technological Education Institute (T.E.I.) of
 Crete, Greece.

Scholarships

- 2011** Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) travel grant
2010 Fonds Wetenschappelijk Onderzoek – Vlaanderen (FWO) travel grant
2007 Vlaams Instituut voor Biotechnologie (VIB) International PhD fellowship
2004 Leonardo Da Vinci European fellowship for BSc Internship

Publications

van de Mortel, J.E., Schat, H., Moerland, P.D., Ver Loren van Themaat, E., van der Ent,
 S., Blankestijn, H., Ghandilyan, A., **Tsiatsiani, S.**, and Aarts, M.G. (2008). Expression
 differences for genes involved in lignin, glutathione and sulphate metabolism in
 response to cadmium in *Arabidopsis thaliana* and the related Zn/Cd-hyperaccumulator
Thlaspi caerulescens. Plant Cell Environ 31, 301-324.

Tsiatsiani, L., Van Breusegem, F., Gallois, P., Zaviyalov, A., Lam, E., and Bozhkov, P.V.
 (2011). Metacaspases. Cell Death Differ 18, 1279-1288.

Tsiatsiani, L., Gevaert, K., and Van Breusegem, F. (2012). Natural substrates of plant
 proteases: how can protease degradomics extend our knowledge? Physiol Plant 145, 28-40.

Wrzaczek, M., Vainonen, J.P., Help, H., **Tsiatsiani, L.**, Stael, S., Gauthier, A., Kaufholdt,
 D., Bollhöner, B., Lamminmäki, A., Gevaert, K., Tuominen, H., Van Breusegem, F.,
 Helariutta and Y., Kangasjärvi, J. (2012). Metacaspase activity and receptor-ligand
 interaction control GRIM REAPER-regulated cell death in plants (**Submitted** to
 Nature Communications)

Tsiatsiani, L., Timmerman, E., De Bock, P.J., Vercammen, D., Van Damme, P., Staes, A.,
 Goethals, M., van de Cotte, B., Beunens, T., Gevaert, K., and Van Breusegem, F. (2012).
 Identification of *Arabidopsis* METACASPASE 9 physiological substrates reveals its role
 in regulating the activity of PHOSPHOENOLPYRUVATE CARBOXYKINASE1. (**In
 preparation** for submission)

Tsiatsiani, L., Audenaert, D., Vercammen, D., Jacques, S., Nguyen, L., Timmerman,
 E., Beunens, T., Gevaert, K., Beeckman, T., and Van Breusegem, F. (2012) *Arabidopsis*
 METACASPASE 9 is involved in cell separation during lateral root development by
 processing extracellular proteins. (**In preparation** for submission)

Conferences

Selected speaker:

April 2011 1st International Plant Protease Conference, Hemavan, Sweden

Poster presenter:

May 2010 Gordon Research Conference on Proteolytic Enzymes and Their Inhibitors, Lucca, Italy

March 2009 VIB seminar, Blankenberge, Belgium

Sept. 2008 VIB Science Club Technologies II, Ghent, Belgium

Research

Work Experience

November 2007 - April 2012 PhD research at Ghent University, Belgium

April - August 2007 Internship at Keygene N.V., Wageningen, The Netherlands within the sector
 of Upstream Research (UR) at the group of Gene Exploitation

March - September 2004 Internship at Plant Research International, Wageningen, The Netherlands