The Arabidopsis METACASPASE9 Degradome

Liana Tsiatsiani,1,2,3,4 Eva Timmerman,1,2,4 Pieter-Jan De Bock,1,2,4 Dominique Vercammen,1,2,4 Simon Stael,1,2,4 Brigitte van de Cotte,1,2 An Staes,1,2,4 Marc Goethals,1,2,4 Tine Beunens,1,2,4 Petra Van Damme,1,2,4 Kris Gevaert,1,2,4 and Frank Van Breusegem1,2,3

1 Department of Plant Systems Biology, VIB, 9052 Ghent, Belgium
2 Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium
3 Department of Medical Protein Research, VIB, 9000 Ghent, Belgium
4 Department of Biochemistry, Ghent University, 9000 Ghent, Belgium

Metacaspases are distant relatives of the metazoan caspases, found in plants, fungi, and protists. However, in contrast with caspases, information about the physiological substrates of metacaspases is still scarce. By means of N-terminal combined fractional diagonal chromatography, the physiological substrates of METACASPASE9 (MC9; AT5G04200) were identified in young seedlings of Arabidopsis thaliana on the proteome-wide level, providing additional insight into MC9 cleavage specificity and revealing a previously unknown preference for acidic residues at the substrate prime site position P1’. The functionalities of the identified MC9 substrates hinted at metacaspase functions other than those related to cell death. These results allowed us to resolve the substrate specificity of MC9 in more detail and indicated that the activity of phosphoenolpyruvate carboxykinase 1 (AT4G37870), a key enzyme in gluconeogenesis, is enhanced upon MC9-dependent proteolysis.

INTRODUCTION

Proteolysis is an irreversible, protease-catalyzed protein modification with diverse effects on protein function and cellular or organismal fate. Proteases mediate protein stability, turnover, and activity and assist in the correct targeting and import of proteins to different organelles. Early development, xylem formation, seed maturation, nutrient supply mobilization, responses to biotic and abiotic stresses, and programmed cell death are facilitated by proteases (Avci et al., 2008; van der Hooorn, 2008; Coll et al., 2011; Hara-Nishimura and Hatzisau, 2011; Vartapetian et al., 2011; Pesquet, 2012). Plant genomes code for numerous proteases (over 700 annotated in Arabidopsis thaliana), but their substrates remain poorly identified (Tsiatsiani et al., 2012). Nevertheless, in the near future, this knowledge gap is expected to be filled by the use of emerging technologies that allow the proteome-wide identification of protease substrates (Huesgen and Overall, 2012).

Mammalian caspases are cysteinyl proteases that, besides functioning in apoptosis, mediate inflammation, cell proliferation, and differentiation by acting on a large number of different cellular substrates (Crawford et al., 2012). Based on their structural similarity to the caspase catalytic domain, metacaspases have been identified in fungi, protozoa, and plants, whereas paracaspases have been found in metazoans and in the slime mold Dictyostelium discoideum (Aravind et al., 1999; Uren et al., 2000; Vercammen et al., 2007). Like caspases, metacaspases and paracaspases carry the hallmarks of the Cys proteases of the CD clan, including the catalytic dyad His/Cys and the hemoglobinase fold (Rawlings and Barrett, 1993; Aravind and Koonin, 2002).

In view of the crucial role of caspases in apoptosis, initial efforts assessed mainly whether metacaspases were responsible for the caspase-like activities detected in plants during cell death events (Bonneau et al., 2008). However, biochemical studies using recombinant metacaspases or protein extracts from loss-of-function paracaspase-like activities in plants, but rather act upstream of the Asp-specific caspases, cleave synthetic peptide substrates after Arg and Lys residues (Vercammen et al., 2004, 2006; Watanabe and Lam, 2005; Bozhkov et al., 2005; González et al., 2007; Moss et al., 2007; Ojha et al., 2010). Therefore, metacaspases are probably not directly responsible for the caspase-like activities in plants, but rather act upstream of the Asp-specific proteases (Meslin et al., 2011).

Besides the multifunctionality of metacaspases, which is apparent mostly in organisms encoding single metacaspase genes, such as Saccharomyces cerevisiae (yeast) and Leishmania major, redundant and antagonistic functions have been demonstrated in organisms encoding multiple metacaspases. The yeast metacaspase, termed Yeast Caspase1 (YCA1), is required for various stress-induced cell death events, including oxidative, osmotic, toxic, and virus-induced stresses and aging (Madeo et al., 2002; Herker et al., 2004; Ivanovska and Hardwick, 2005; Chahomchuen...
et al., 2009). YCA1 is also involved in cell cycle regulation (Lee et al., 2008) and clearance of insoluble protein aggregates (Lee et al., 2010). The L. major metacaspase MCA is involved in cell cycle regulation and in oxidative stress–induced cell death (Ambit et al., 2008; Zailia et al., 2011). Three of the five Trypanosoma brucei metacaspases are essential for viability of the bloodstream form of the parasite (Heims et al., 2006). Redundant functions in the regulation of senescence–associated death were also shown for two Podospora anserina metacaspases (Hamann et al., 2007) and in the resistance to endoplasmic reticulum stress for two Aspergillus fumigatus metacaspases (Richie et al., 2007).

Interestingly, antagonistic functions for homologous metacaspases were found in T. brucei and Arabidopsis in the regulation of cell death events or cell proliferation and differentiation (Coll et al., 2010; Laverrière et al., 2012).

Metacaspases are posttranslationally regulated by multiple mechanisms, and their activity depends on specific cellular conditions. The proteolytic activity of Arabidopsis MC9 is regulated by S-nitrosylation, by a Ser protease inhibitor (Serpin 1), and by autocatalytic processing (Vercammen et al., 2004, 2006; Belenghi et al., 2007). Furthermore, the activities of the Norway spruce (Picea abies) metacaspase (mcl1-Pa) and the Arabidopsis MC1, MC4, and MC5 are calcium dependent (Bozhkov et al., 2005; Watanabe and Lam, 2005, 2011a). Recently, a T. brucei metacaspase (MCA4) without detectable proteolytic activity, but necessary for parasitic proliferation and virulence, was found to be processed by another metacaspase, MCA3 (Proto et al., 2011).

Despite these results, only a handful of substrates have been identified so far (Tsatsian et al., 2011). As knowledge on the corresponding substrates of proteases with unknown function can provide insights into their potential function or role in a specific process, new tools that facilitate substrate identification should assist in expanding our understanding of plant protease functions.

Here, we assessed the degradome (Overall et al., 2004) of the Arabidopsis METACASPASE9 (MC9; AT5G04200) by N-terminal combined fractional diagonal chromatography (COFRADIC). The physiological substrates of this plant protease were identified on the proteome–wide level, and the concrete substrate cleavage sites were characterized by means of positional proteomics. Thanks to these data, the substrate specificity of MC9 could be resolved in more detail, and the activity of phosphoenolpyruvate carboxylase 1 (PEPCK1; AT4G37870), a key enzyme in gluconeogenesis, was found to be enhanced upon MC9-dependent proteolysis.

RESULTS

Expression Characteristics of the Arabidopsis MC9 Gene

The spatiotemporal expression of Arabidopsis MC9 was studied in ProMC9::GUS reporter lines by means of β-glucuronidase (GUS) histochemical analysis (for a detailed description, see Supplemental Methods 1 online). Several independent homozygous alleles were analyzed with similar results (see representative GUS staining patterns in Figures 1A to 1D). Interestingly, the MC9 promoter activity was already obvious in root caps of young 2-d-old seedlings (corresponding to the developmental stage 0.5 to 0.7, including a 3-d stratification period) (Boyce et al., 2001) (Figure 1A), remained the same in 1-week-old seedlings, and was observed also in epidermal cells (Figure 1B). In hypocotyls and cotyledons, but not in rosette leaves, discrete differentiating primary and secondary trachy-fluid elements were also stained (Figure 1C). In petals, a strong signal appeared just prior to abscission (Figure 1D). Anther connectives also displayed strong MC9 promoter activity (Figure 1D).

Figure 1. Expression Characteristics of the MC9 Gene and Molecular Characterization of the MC9 Transgenic Lines Used in This Study.

(A) to (D) GUS staining of root cap of 2-d-old seedlings (A), root cap of 1-week-old seedling (B), tracheary elements in expanding cotyledon (C), and petals prior to abscission (D).

(E) Schematic representation of the T-DNA insertion of the GK-540H0 allele. The catalytic His and Cys, the p20 and p10 MC9 protease subunits, and the linker region between the two catalytic subunits are depicted.

(F) Immunoblot analysis of Arabidopsis seedlings’ proteins with a MC9 polyclonal antibody. Full–length 35.5-kD MC9 is detected in the wild type (Col-0) and 35S::MC9 overexpressor (Vercammen et al., 2006), but not in the mc9 knockout line (GK-540H0 allele). An actin antibody directed against a common epitope of all actin isoforms (Immuno, clone C4) was used as gel loading control.

(G) Transcript levels of all Arabidopsis metacaspase genes (MC1 to MC9) in 2-d-old seedlings. Values represent the mean of three measurements ± SD relative to the MC9 gene transcript level in the wild type (Col-0).
By quantitative RT-PCR, we assessed the expression of all nine *Arabidopsis* metacaspases in 2-d-old seedlings of wild-type Columbia-0 (Col-0) and in the T-DNA insertion knockout (mc9) and MC9 overexpression (35S:MC9) (Vercammen et al., 2006) transgenic lines (for a detailed description, see Supplemental Methods 2 online) (Figures 1E to 1G). Both MC9 gain-of-function or loss-of-function plants did not display any obvious aberrant phenotypes under normal vegetative or reproductive growth conditions. MC4 expression was the highest (Figure 1G), consistent with a previous report that it is the most expressed metacaspase at various growth stages (Watanabe and Lam, 2011a). In young seedlings, corresponding to a developmentally active stage with a dynamic proteome, the MC9 expression was the second highest; therefore, this stage was chosen for analysis of the MC9 degradome (Figure 1G).

**Identification of MC9 Protein Processing Events by N-Terminal COFRADIC**

To explore the function of MC9 through identification of its physiological protein substrates, we used the N-terminal COFRADIC technique (Gevaert et al., 2003; Staes et al., 2011). In contrast with most proteomic approaches, N-terminal COFRADIC is a positional proteomics technology that uses sequential peptide chromatography to enrich for those peptides that contain the protein N termini out of the whole peptide mixture obtained by, for instance, full trypsin digestion of proteomes. Here, this so-called N-terminome consisted of all peptides that possessed the mature protein N termini and the neo-N-termini that had been generated upon proteolysis by MC9. Neo-N-termini that emerged at MC9 processing sites directly conveyed the MC9 proteolytic events and the exact site where the protein cleavage occurred (Van Damme et al., 2005). We analyzed the in vivo N-terminome of 2-d-old seedlings of wild-type and MC9 gain-of-function and loss-of-function transgenic plants (Figures 1E and 1F). First, the N-terminome of the T-DNA insertion knockout (mc9) seedlings was compared with that of wild-type seedlings and, second, to that of seedlings overexpressing MC9 (35S:MC9). Additionally, a mc9 seedling proteome to which *Arabidopsis* recombinant MC9 (rMC9) (Vercammen et al., 2004) had been added in vitro was studied versus an untreated mc9 control. To compare the N-terminomes, N-hydroxysuccinimide esters of light and heavy isotopic variants of butyric acid were used to mass tag the (neo)-N-terminal peptides (see Supplemental Figure 1 online). This differential labeling strategy was possible because MC9 introduced novel, primary α-amino groups that were uniquely tagged with either a light or heavy variant of butyric acid. After this mass tagging, equal amounts of the proteome preparations were mixed and digested with trypsin, and (neo)-N-terminal peptides were isolated with N-terminal COFRADIC (see Supplemental Figure 1 online). Mass tagging of neo-N-termini enabled us to determine the proteome origin of each labeled peptide. The respective ion signal intensities were further used to identify the MC9-processed sites.

The following MS readouts were expected: (1) Some (neo)-N-terminal peptides would be found at equal intensity in both samples. Such peptides might carry the mature protein N terminus and would not be indicative of MC9 proteolysis (Figure 2A).

Furthermore, such N-terminal peptides would imply that the particular protein (form) was not up- or downregulated in the tested proteomes. (2) Some neo-N-terminal peptides would occur in both samples, but with a significantly higher intensity in the sample with either the active MC9 or the supplemented

**Figure 2.** Illustration of the Different Categories of Isolated (Neo)-N-Terminal Peptides.

Mass tagging allowed the differential comparison of N-terminomes derived from samples in which MC9 was active (Col-0 or 35S:MC9) or added rMC9 or absent (control sample mc9). MS spectra of N-terminal peptides were, in most cases, identified as doublet ions of light (in blue) and heavy (in red) isotopes. m/z, mass-to-charge ratio. (A) Protein cleaved in both samples to the same extent (i.e., by trypsin; yellow arrow) and fragments bearing the mature protein N terminus. As the amounts of these peptides were approximately equal, no MC9-driven proteolysis could be deduced. (B) Protein cleaved in both samples, but at a higher frequency in the MC9 sample (large red arrow). The neo-N-terminus fragment was generated in both samples, but in significantly higher amounts in the MC9 sample. Proteolysis of the target in the control sample might have occurred by a redundant protease activity (blue arrow). The fragment is identified as a doublet with the highest intensity ion derived from the MC9 sample. (C) Protein solely cleaved by MC9 (red arrow) in the samples of active MC9 proteases. Thereby, a fragment with a neo-N-terminus was generated, later identified as unique in the MC9 N-terminome and, thus, as a singleton ion.
rMC9. This outcome would mean that the corresponding proteins were also cleaved in the control sample (lacking MC9 activity), although to a lesser extent, by another protease, hinting at a redundant protease activity (Figure 2B). Such events could be explained by a scenario in which the MC9 perturbation activated other proteases (i.e., metacaspases). (3) Some neo-N-terminal peptides (so-called singletons) would be uniquely found in the proteome either with the active MC9 (Col-0 or 35S:MC9 proteome) or supplemented with rMC9, implying that the corresponding proteins were cleaved by MC9 (Figure 2C).

Peptides were analyzed with an LTQ Orbitrap XL mass spectrometer. The obtained tandem mass spectrometry (MS/MS) spectra were searched against The Arabidopsis Information Resource database (release 8) with the Mascot algorithm and subsequently mapped to the TAIR10 database with DB Toolkit (Martens et al., 2005) (for a detailed description, see Supplemental Methods 3 online). In total, 75,310 MS/MS spectra were generated in the mc9/Col-0 analysis, 69,388 in the mc9/35S:MC9 analysis, and 75,042 in the in vitro analysis, identifying 3781, 2879, and 391 peptides (singletons) in the Col-0, mc9/Col-0, and 35S:MC9 proteomes, respectively. The ratios of light/heavy peptide signals for all identified peptides were calculated with the Mascot Distiller software and were individually checked for possible neo-N-terminal peptides pointing to MC9 processing. We focused on two classes of peptides: the significantly enriched neo-N-terminal peptides (Z-score ≤ −2.33 or Z-score ≥ 2.33, depending on the labeling used to calculate the light/heavy intensity ratios) and the unique peptides (singletons) in the Col-0, 35S:MC9, or rMC9-treated seedling proteomes in comparison to mc9 (Figures 2B and 2C). As the former might also appear because of differential protein levels in the input proteomes, we scanned the data for the corresponding precursor protein N termini or other neo-N-terminal peptides from these precursor proteins, which were not cleaved by MC9. Comparison of these peptide levels led us to conclude that no large proteome changes occurred between the different proteomes assayed. For an overview of all cleavage sites, their corresponding substrates, and their precursor protein N termini identified in the three analyses, see Supplemental Data Sets 1A to 1C online.

After filtering the data with the known MC9 specificity for Arg and Lys, 72% of all processed sites matched the expected MC9 specificity (2B and 44% cleavage C-terminally to Arg and Lys, respectively), without significant enrichment for a particular amino acid in the other processing events. In total, 551 cleavage events after Arg or Lys (116 in mc9/Col-0, 103 in mc9/35S:MC9, and 332 in vitro) (see Supplemental Figure 2A online) were identified in 392 proteins (97 in mc9/Col-0, 84 in mc9/35S:MC9, and 211 in vitro) (see Supplemental Figure 2B online), from which approximately one-third (160 sites) were reported as singleton peptides in either Col-0, 35S:MC9, or rMC9-treated proteomes (16 in mc9/Col-0, 11 in mc9/35S:MC9, and 133 in vitro) and 391 as significantly enriched peptides in the same proteomes (100 in mc9/Col-0, 92 in mc9/35S:MC9, and 199 in vitro) (see Supplemental Data Sets 1A to 1C online).

Positional Amino Acid Cleavage Specificity of MC9

Using iceLogo (Colaert et al., 2009), we inspected the amino acid sequences at the prime and nonprime substrate positions of the Arg and Lys P1-specific cleavage sites (Schechter and Berger, 1967) and analyzed the frequencies of specific residues after statistical correction by means of the natural occurrence of amino acids in Arabidopsis proteins. The specificity information derived from the three data sets was very similar. The large in vitro COFRADIC data set showed that the acidic residues Asp and Glu were often found at neighboring positions to the basic (Arg or Lys) cleavage site. For instance, in 46 and 47% of the Lys-cleaved and Arg-cleaved sequences, respectively, the P1′ position (C-terminally to the cleavage P1 position) was occupied by Asp or Glu and, likewise, the P2 position in 23 and 22% of the Lys-cleaved and Arg-cleaved sequences, respectively. However, only in 5 and 7% of the Lys-cleaved and Arg-cleaved sequences, respectively, both P2 and P1′ were acidic (Figure 3A; see Supplemental Figures 3 and 4 online). At the prime site, Asp or Glu was also observed in 27% of the cases at P3′. At P2′, the small amino acids Thr, Ala, and Val occurred in 38% of all cases, whereas Ala was tolerated at every position. Analogous to the in vivo COFRADIC-derived specificity (Figure 3B), Val and Glu were preferred at P4 and Lys and Arg were frequently found at P3. Ala and Thr were the sole preferred residues at P2 in the in vitro-cleaved substrates, but in vitro-cleaved proteins additionally (23% of the cleaved substrates) tolerated Asp and Glu at this position (Figures 3A and 3B; see Supplemental Figures 3 and 4 online). Regarding the prime site substrate specificity of MC9, the in vivo and in vitro data were largely similar, except for the amino acid tolerance at P2′: In the latter, small residues were present at this position, whereas aromatic residues, such as Phe and Tyr and the aliphatic Ile and Leu, were found also in the former.

In conclusion, the MC9 specificity for basic residues at the P1 site was accompanied by a strong preference for acidic amino acids at the P1′ site (see Supplemental Figure 4 online). This particular trend of acidic–basic amino acid combinations was also seen at other positions at the nonprime site (P4–P3), whereas Ala was generally tolerated at every position. A relative preference occurred for Lys over Arg at the P1 site and in vivo, often (27% of the cases) a second Arg or Lys at P3. For instance, when Lys was present at P1, 24% of these substrates contained a second Lys at P3.

Previously, in a positional scanning of a synthetic combinatorial library (PS-SCL) of 7-amino-4-methylcoumarin (AMC)–labeled synthetic tetrapeptides spanning the positions P4–P1, VRPR was found as the most preferred mc9 tetrapeptide substrate, whereas IISK, the best cleaved peptide with Lys at P1, was fivefold less efficiently cleaved than VRPR (for Val-Arg-Pro-Val) (Vercammen et al., 2006). Our current results somewhat contrast with the stronger preference for Arg over Lys at P1, but they are in agreement with the preferred presence of a dibasic amino acid motif (such as RxR) as in VRPR. When the two in vitro approaches (in vitro COFRADIC and PS-SCL) were compared (Figure 3C), Val was the most prevalent residue at P4 in both data sets of Arg-cleaved substrates. However, a major difference was observed at the P2 site: The COFRADIC data indicated that Thr and Asp were preferred and the PS-SCL data showed a preference for hydrophobic residues or Pro. Based on the in vivo and in vitro amino acid frequencies at the P4–P3′ substrate positions derived from the COFRADIC analyses (Figure 3; see Supplemental Figures 3 and 4 online), two revised
**Figure 3.** MC9 Specificity Monitored via Its Proteome-Wide Substrate Profiling.

Analysis was performed with iceLogo (Colaert et al., 2009) after multiple sequence alignment and statistical correction with the natural occurrence of amino acids in Arabidopsis proteins. Heat maps illustrating the amino acid frequencies at the P4-P3 positions in the MC9 substrates as observed in the in vivo mc9/Col-0 and mc9/3SS:MC9 COFRADIC studies (A), the in vitro COFRADIC (B), and the PS-SCL approach (C). R software was used to build the PS-SCL heat maps with the published values (Vercammen et al., 2006). Arrows mark the cleaved scissile bonds.
Experimental Validation of MC9 Protein Substrates

For the experimental validation of the MC9 substrates, we prioritized all in vivo singleton peptides together with those proteins bearing cleavage sites identified in more than one proteomic analysis. In this manner, 99 sites in 74 different proteins were selected for cleavage validation (Table 1), from which 18 sites were identical in the three different setups, 35 sites in both in vivo analyses, and finally 15 sites between the mc9/Col-0 in vivo and the in vitro analysis, and finally 15 sites between the mc9/35S:MC9 in vivo and the in vitro analysis. The remaining 12 sites were reported as in vivo singleton peptides (Table 1; see Supplemental Figure 2A online).

Two independent validation approaches were chosen. First, 18-mer amino acid peptides harboring the identified cleavage sites were synthesized for six randomly selected candidate protein substrates and used for in vitro rMC9 cleavage assays. Second, in vitro–transcribed and translated, radiolabeled (35S-Met) proteins, generated from available cDNA clones (ABRC), were incubated with rMC9 and their digestion products were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography.

After incubation with increasing concentrations of rMC9 (0, 31, 125, and 500 nM), synthetic peptides and their digestion products were separated by reverse-phase HPLC and analyzed by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (for a detailed method description, see Supplemental Methods 4 online). Peptides spanning the MC9 cleavage sites identified in trypsin inhibitor 2 (AT1G47540), general regulatory factor 5 (GRF5; AT5G16050), Gly-rich RNA binding protein (GRP7; AT2G21660), and protochlorophyllide oxidoreductase B (PORB; AT4G27440) were cleaved by rMC9 at the sites identified by COFRADIC. The degree of peptide processing increased with increasing rMC9 concentrations (see Supplemental Figure 5 online).

For proteins containing more than one identified cleavage site, such as the late-embryogenesis abundant (LEA) domain–containing protein (AT2G42560), PEPCK1 (AT4G37870), and GRP7, two synthetic peptides were tested (see Supplemental Figure 6 online). All of these peptides were cleaved by rMC9 at the sites identified by COFRADIC. However, additional alternative cleavage sites were observed as well, probably due to sequence similarities with the COFRADIC-deduced sites. For instance, the additional cleavage site R\textsuperscript{121}E in the LEA domain–containing protein precursor was identical to the COFRADIC-deduced site R\textsuperscript{124}E. As for GRP7, the peptide precursor was originally expected to be cleaved at the MK\textsuperscript{64}L site but was additionally processed atEK\textsuperscript{61}A, a similar site, because in both cases, K was adjacent to an acidic residue (Asp or Glu). Furthermore, the consensus sequence for Lys-cleaved peptides supported the preference for acidic residues either at the N-terminal or C-terminal side of the cleaved Lys.

To validate experimentally the substrate cleavages at the protein level, we employed in vitro transcription and translation coupled assays. In total, we successfully translated 16 proteins, all of which were cleaved by rMC9. Although the precursor protein levels clearly decreased or even were completely negligible after the rMC9 treatment, for some substrates, cleavage fragments were not always observed. It is possible that the substrate fragments were not detectable by autoradiography because they did not contain a radiolabeled Met or that proteolytic fragments <20 kD might have been too small for proper separation and visualization by the SDS-PAGE system used. The substrates that were processed into fragments of correctly predicted molecular weight based on the cleavage site(s) previously identified by COFRADIC and MS, were GRF5, PEPCK1, nicotinamide adenine dinucleotide (NAD) synthetase (AT1G55090), GDP-mannose 4,6-dehydratase 2 (GMD2; AT3G51160), pentatricopeptide repeat (PPR)–containing protein (AT1G02150), and citrate synthase3 (CSY3; AT2G42790) (Figure 4). Furthermore, these proteins were not cleaved by the inactive rMC9 mutant, rMC9C147AC29A (Belenghi et al., 2007), indicating that the observed fragments were indeed due to rMC9 and not to other (contaminating) proteases present in the assay mixture.

MC9 Cleave PEPC1 in Vivo, Leading to Increased PEPC1 Activity

PEPC1, which catalyzes the reversible ATP–dependent de-carboxylation of oxaloacetate to phosphoenolpyruvate, is involved in gluconeogenesis, photosynthetic CO\textsubscript{2}–concentrating mechanisms during C4 photosynthesis (Edwards et al., 1971), Crassulacean acid metabolism (Dittrich et al., 1973), and nitrogenous compound metabolism (Leegood and Walker, 2003). As PEPC is posttranslationally regulated by reversible phosphorylation (Walker and Leegood, 1995) and prone to rapid proteolytic cleavage in planta (Walker et al., 1995, 1997; Malone et al., 2007), which gives rise to fragments with sizes that correlate to the cleavage sites observed by N-terminal COFRADIC, we focused on the MC9-dependent PEPC1 cleavage. To this end, we assessed whether MC9 was responsible for in planta cleavage of PEPC1, which is highly active during germination at the 2-d-old seedling stage of the studied proteome (Rylott et al., 2003).

PEPC1 polypeptides (74 and 73 kD) were previously immunodetected in Arabidopsis cotyledons with a polyclonal antiserum raised against PEPC1 from Guinea grass (Panicum maximum) leaves (Malone et al., 2007). Similarly, in pea (Pisum sativum) cotyledons, besides the mature 72-kD form, a smaller 68-kD PEPC1 polypeptide was detected that was attributed either to a genuine isoform or to a regulatory protein processing event (Delgado-Alvarado et al., 2007). Recently, the pineapple (Ananas comosus) PEPC1 had been found to have two active polypeptide variants of 74 and 65 kD in vivo (Martin et al., 2011).
Table 1. Selection of Neo-N-Terminal Peptides Identified in at Least Two Independent Analyses.

<table>
<thead>
<tr>
<th>No. of Cleavage Sites</th>
<th>Accession</th>
<th>P4-P3'</th>
<th>Sequence Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified in all three analyses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AT1G12000</td>
<td>AVTR(10)DLT</td>
<td>DLTAVGSPENAPAKGR Phosphofructokinase family protein</td>
</tr>
<tr>
<td>2</td>
<td>AT1G42460</td>
<td>MNQK(53)DAY</td>
<td>DAYVGDEAGSKR Actin 8</td>
</tr>
<tr>
<td>3</td>
<td>AT2G21060</td>
<td>HMAR(150)ECS</td>
<td>ECS90GQGQGSGGGGGR Gly-rich protein 2B</td>
</tr>
<tr>
<td>4</td>
<td>AT2G42560</td>
<td>VGSK(454)AVD</td>
<td>AVOLTEKEAAVAADTVGYTAR LEA domain–containing protein</td>
</tr>
<tr>
<td>5</td>
<td>AT3G02480</td>
<td>MKEK(51)AQG</td>
<td>AQGAADWKKDKTMNKSHE LEA family protein</td>
</tr>
<tr>
<td>6</td>
<td>AT3G08840</td>
<td>KSKK(15)DFS</td>
<td>DFSTAILER Cell division cycle 48</td>
</tr>
<tr>
<td>7</td>
<td>AT3G17520</td>
<td>EAKK(184)EAKK</td>
<td>EAKK EAKK LEA family protein</td>
</tr>
<tr>
<td>8</td>
<td>AKDK(263)ASQ</td>
<td>ASQYSDDAAR</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>AT3G53040</td>
<td>KQK(11)ETA</td>
<td>ETATDYDKAR Putative/LEA protein</td>
</tr>
<tr>
<td>10</td>
<td>DKT(155)ETA</td>
<td>ETAETYDKAR</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>AT4G02930</td>
<td>GAKK(99)AA</td>
<td>AIAFDEIKAPEEKKR GTP binding Elongation factor Tu family protein</td>
</tr>
<tr>
<td>12</td>
<td>AT4G21020</td>
<td>EAKK(156)DYA</td>
<td>DYAYTMNNAEKAAR LEA family protein</td>
</tr>
<tr>
<td>13</td>
<td>AT1G11480</td>
<td>ARPR(404)ELV</td>
<td>ELVLER Eukaryotic translation initiation factor-related</td>
</tr>
<tr>
<td>14</td>
<td>AT1G49240</td>
<td>SIEK(241)NYE</td>
<td>NYEELPDGQVITIGAER ACTIN family protein</td>
</tr>
<tr>
<td>15</td>
<td>AT1G49970</td>
<td>FEKR(363)DYD</td>
<td>DyDGTALQAR CLP protease proteolytic subunit 1</td>
</tr>
<tr>
<td>16</td>
<td>AT1G51710</td>
<td>VATR(157)ELF</td>
<td>ELFGELDR Ubiquitin-specific protease 6</td>
</tr>
<tr>
<td>17</td>
<td>AT1G54980</td>
<td>TVAR(114)EVE</td>
<td>EVELEDPVENIGKVR Chaperonin 60β</td>
</tr>
<tr>
<td>18</td>
<td>AT1G54190</td>
<td>YTK(341)EIA</td>
<td>EIAAAPAGTAQFPLSGL Hyaluronan/mRNA binding family</td>
</tr>
<tr>
<td>Identified in both in vivo analyses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AT1G04750</td>
<td>VLDR(156)GEK</td>
<td>GEKIELLVDKTENLR Vesicle-associated membrane protein 721</td>
</tr>
<tr>
<td>2</td>
<td>AT1G07780</td>
<td>SAKK(109)DIS</td>
<td>DISKVAR Phosphoribosylanthranilate isomerase 1</td>
</tr>
<tr>
<td>3</td>
<td>AT1G09030</td>
<td>TEAK(165)EAV</td>
<td>EAAKEAVKEAAR Protein of unknown function, DUF538</td>
</tr>
<tr>
<td>4</td>
<td>AT1G11480</td>
<td>ARPR(404)ELV</td>
<td>ELVLER Eukaryotic translation initiation factor-related</td>
</tr>
<tr>
<td>5</td>
<td>AT1G29350</td>
<td>QNAR(525)ELD</td>
<td>ELDFQYSPFSAQSMQSR Kinase-related protein of unknown function</td>
</tr>
<tr>
<td>6</td>
<td>AT1G47540</td>
<td>AYDR(31)KCL</td>
<td>KCLKEYGVDVFVSCAPR Trypsin inhibitor 2</td>
</tr>
<tr>
<td>7</td>
<td>AT1G49240</td>
<td>SIEK(241)NYE</td>
<td>NYEELPDGQVITIGAER ACTIN family protein</td>
</tr>
<tr>
<td>8</td>
<td>AT1G49970</td>
<td>FEKR(363)DYD</td>
<td>DyDGTALQAR CLP protease proteolytic subunit 1</td>
</tr>
<tr>
<td>9</td>
<td>AT1G51710</td>
<td>VATR(157)ELF</td>
<td>ELFGELDR Ubiquitin-specific protease 6</td>
</tr>
<tr>
<td>10</td>
<td>AT1G54980</td>
<td>TVAR(114)EVE</td>
<td>EVELEDPVENIGKVR Chaperonin 60β</td>
</tr>
<tr>
<td>11</td>
<td>AT1G56070</td>
<td>ICLK(544)DLQ</td>
<td>DLQDFQYSPFSAQSMQSR Kinase-related protein of unknown function</td>
</tr>
<tr>
<td>12</td>
<td>AT2G01250</td>
<td>VESK(6)VVV</td>
<td>VVVVESLVKKR Ribosomal protein L30/L7 family protein</td>
</tr>
<tr>
<td>13</td>
<td>AT2G18960</td>
<td>GSYR(902)ELS</td>
<td>ELSEIQAERK H(+)ATPase 1</td>
</tr>
<tr>
<td>14</td>
<td>AT2G19730</td>
<td>SVNK(88)SIL</td>
<td>SILKEFPR Ribosomal L28e protein family</td>
</tr>
<tr>
<td>15</td>
<td>AT2G24550</td>
<td>DDRK(156)DYD</td>
<td>DDYTLAENR RAB GTPase homolog E1B</td>
</tr>
<tr>
<td>16</td>
<td>AT2G32240</td>
<td>VJKR(709)DID</td>
<td>DIDLFSSTPTKR Unknown protein</td>
</tr>
<tr>
<td>17</td>
<td>AT2G36620</td>
<td>NIPK(149)SA</td>
<td>SAAPKAAKGEGGGGRR Ribosomal protein L24</td>
</tr>
<tr>
<td>18</td>
<td>AT2G42560</td>
<td>ARAK(324)DYT</td>
<td>DYTLQKAVEKDVAEKAAR LEA domain–containing protein</td>
</tr>
<tr>
<td>19</td>
<td>AT2G53040</td>
<td>TVLK(449)EAD</td>
<td>EADQMTGQTVGIEDEEKR Putative/LEA protein</td>
</tr>
<tr>
<td>20</td>
<td>AT2G61260</td>
<td>SDRR(92)DVK</td>
<td>DVKLADLSKEKR Remorin family protein</td>
</tr>
<tr>
<td>21</td>
<td>AT3G13300</td>
<td>VGK(708)NLQ</td>
<td>NLDVSSVEEIR Transducin/WD40 repeat-like superfamily protein</td>
</tr>
<tr>
<td>22</td>
<td>AT3G04200</td>
<td>LFSK(216)DAG</td>
<td>DAGLKF Heat shock protein 81.4</td>
</tr>
<tr>
<td>23</td>
<td>AT3G04200</td>
<td>LFSK(216)DAG</td>
<td>DAGLKF Heat shock protein 81.4</td>
</tr>
<tr>
<td>24</td>
<td>AT3G24410</td>
<td>YTK(341)GKV</td>
<td>GVYSEETSGKR Protoporphyrilide oxidoreductase B</td>
</tr>
<tr>
<td>25</td>
<td>AT3G31700</td>
<td>IVK(119)GEN</td>
<td>GENDLPGLDTEKPR Ribosomal protein S6</td>
</tr>
<tr>
<td>26</td>
<td>AT3G04200</td>
<td>LFSK(216)DAG</td>
<td>DAGLKF Heat shock protein 81.4</td>
</tr>
<tr>
<td>27</td>
<td>AT3G06140</td>
<td>EQPR(9)NIS</td>
<td>NISGQSQSPR Sorting nexin 1</td>
</tr>
<tr>
<td>28</td>
<td>AT3G12140</td>
<td>GVR(15)DID</td>
<td>DIDANANLQVESLR Cystatin-1</td>
</tr>
<tr>
<td>29</td>
<td>AT3G4310</td>
<td>EYTR(150)DFS</td>
<td>DFPHEETKEVNAGASR LEA family protein</td>
</tr>
<tr>
<td>30</td>
<td>AT3G52440</td>
<td>TLR(145)EIG</td>
<td>EIGLDDISTPVNYQNKR Bacterial sec-independent translation protein</td>
</tr>
<tr>
<td>31</td>
<td>AT3G54190</td>
<td>YIK(341)GKV</td>
<td>GVYSEETSGKR Protoporphyrilide oxidoreductase B</td>
</tr>
<tr>
<td>32</td>
<td>AT3G56000</td>
<td>SSSK(604)TM</td>
<td>TMENPENSIME DELR Heat shock protein 81.4</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>No. of Cleavage Sites</th>
<th>Accession</th>
<th>P4-P3&lt;sup&gt;*&lt;/sup&gt;</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>AT5G62575</td>
<td>TREK(42)ALL</td>
<td>ALLAEDSALKR</td>
<td>Unknown protein</td>
</tr>
</tbody>
</table>

Identified in the mc9/3SS/4C9 and the in vitro analysis

1. AT1G13270 RKKR(78)ELA ELETKSKVR Met aminopeptidase 1B
2. AT1G54870 EKIK(28)NFG NFGSVPMKR Aldehyde reductase
3. AT1G72370 LGTK(33)NCA NCXYOMER 40s ribosomal protein SA
4. AT2G21660 KAMK(64)DAI DAIEMNGQDDGR Gly-rich, RNA binding protein atgrp7
5. AT2G42560 MRER(21)EJG EGKESAGGGVGR LEA domain-containing protein
6. ATEK(268)EJK EJKALEGNNMTAEQAAR
7. EKIK(270)EJK EJKALEGNNMTAEQAAR
8. VEAK(335)DJA DVAEAQR
9. AT3G17520 ESAAK(260)DKA DKASQSYYDAAAR LEA family protein
10. AT3G22640 TRSK(231)EIJ EIGGIIIR Cupin family protein
11. AT3G51160 TAPK(20)ADS ADSTTVHEPR ADSTTVHEPR
12. AT3G53040 DKT(1)3EJTA ETADYAEKAR Putative/LEA protein
13. AT4G37870 SLTR(102)ESG ESGPKVDR PEPCK1
14. AT4G39260 DLQR(25)TFS TFSQFGDVDSICR Cold, circadian rhythm, and RNA binding 1
15. AT5G10360 IVKK(119)GVS GVSDLPGDTEKPR Ribosomal protein s6e

Identified in the mc9/Col-0 and the in vitro analysis

1. AT1G47540 CAPR(49)IFP IFPTFCQNCR Trypsin inhibitor 2
2. AT1G55490 VLT(38)EJS ETSIVGDGSTDAVKKR Chaperonin 60
3. AT1G65090 TSER(162)TQK TLQDDKKSNGKASEEVDQQPKKR Unknown protein
4. AT2G1660 TDTR(23)ALE ALTEAFOYGDVDDSKINDR Gly-rich, RNA binding protein atgrp7
5. AT2G36640 QKTR(338)EST ESTQGAQGAEKTDOVVR Embryonic cell protein 63
6. AT2G40490 VTER(47)KVS KVSTASEPQLR Uroporphyrinogen decarboxylase
7. AT2G42560 OHT(607)JNV NVIVDDAPVR LEA domain-containing protein
8. AT3G02480 QQMK(49)EKA EKAQGAGGTDKGMKSH LEA family protein
9. AT3G17110 QIK(11)ALS ALSEFLMSSTLNAEASSGAGGPR Lysyl-tRNA synthetase 1
10. AT3G12960 EKIK(58)EKS EQSADGQQTIOQR Unknown protein
11. AT3G13470 VLTK(378)EMT EMTIVGDGTDQEAVKNR TCP-1/cnprotin chaperonin protein
12. AT3G25200 TEY(135)GVE GVEDLHQTGGEVKSP Unknown protein
13. AT3G5670 IKNK(134)AQD AGQDAQYKTETAGGAQYTKETAEQR LEA family protein
14. QYTK(143)JATA EATAQGAAQYTKETAEQR
15. AT3G48870 ISDR(501)FLP FLPQKAIDLEAGSR Clp ATPase
16. AT4G21020 EKAK(118)DJA DTAYNAECKAGYAEQ LEA family protein
17. ESTK(218)NAA NAQQTTEAVGPEEAEKR
18. AT4G38630 QKDR(258)DIA DGAASQASDTEVR Regulatory particle non-ATPase 10

Fraction of singleton neo-N-terminal peptides that were solely found in one of the in vivo analyses

1. AT1G21050 VTSA(474)EAV EAVELLIR Pentatricopeptide repeat-containing protein
2. AT1G50990 NSNK(710)EIG EIGWNSGDSAGL NAD synthetase
3. AT2G32990 YAK(35)EIK EIKGFGVREH Heat shock protein 60
4. AT3G45190 VHD(558)DYD DYDLAGGLNNLQFR SIT4 phosphatase-associated family protein
5. AT2G42790 KEDK(86)GLK GLKLDGYLPNTAPVR Citrate synthase 3, CSY3
6. AT2G47470 YIK(320)GSD GSYDASKEHER Thioredoxin family protein
7. AT3G21020 LMKL(49)FLP VLEDATAVKTMPKGR Heat shock protein 60
8. AT4G21020 DOAR(110)DFS DFSLEDKDR Photosystem II subunit Q-2
9. AT4G3450 GIE(320)GAV GAVLQEEHR Coatomer γ-2 subunit, putative
10. AT5G16130 LTQ(181)DJS DVVFFYVEYEA Ribosomal protein S7e family protein
11. AT5G95190 ESTR(578)DLS DLTMGAANTNEGR Polyadenylate transferase
12. AT5G4310 QSKG(468)DFD DFSDFLMDGFTKH ARF-GAP domain 5

The equivalent experiments, peptide sequence, accession number, and protein description are given. Amino acid positions of the P1<sup>*</sup> sites are indicated in parentheses in the P4-P3<sup>*</sup> column.
protease activity or degradation. Taken together, our data indicate that overexpression of MC9 increased the PEPCK1 processing in vivo.

To assess the effect of the MC9-dependent PEPCK1 cleavage, we measured the carboxylation activity of PEPCK1 in young seedling extracts of Col-0, *pepck1* (Penfield et al., 2004), and MC9 gain-of-function and loss-of-function mutants (Figure 6A). PEPCK1 activity was the highest in 35S:MC9 plant extracts (46 and 71% higher than that of Col-0 and mc9, respectively), revealing that MC9 cleavage enhanced PEPCK1 activity and that proteolysis might remove the putative PEPCK1 inhibitory phosphorylation site (Ser-62). As PEPCK1 is involved in Suc synthesis from lipids during oilseed plant germination, the *pepck1* mutants are impaired in hypocotyl elongation when grown in the dark in Suc-deprived growth media (Rylott et al., 2003; Penfield et al., 2004). Hence, under these conditions, we monitored the hypocotyl elongation of *pepck1*, mc9, Col-0, and 3SS:MC9 for a period of 10 d after germination. As expected, *pepck1* hypocotyls were the smallest and the *mc9* hypocotyl growth was significantly compromised (*P* value < 0.05) in comparison to that of Col-0. Reciprocally, 3SS:MC9 hypocotyls were the longest (17.4, 36.9, and 56.3% longer than Col-0, mc9, and *pepck1* hypocotyls, respectively) (Figures 6B and 6C). These results are indicative of a positive regulatory effect of MC9 on the PEPCK1 activity during gluconeogenesis.

**MC9 Subcellular Localization**

Because of the apparent discrepancy between the cytosolic localization of PEPCK1 and the detection of MC9 in the apoplast of 3SS:MC9 lines (Vercammen et al., 2006), we reassessed the subcellular localizations of MC9 and PEPCK1 by transiently expressing N- and C-terminal green fluorescent protein (GFP)-tagged MC9 and PEPCK1 in tobacco (*Nicotiana benthamiana*) leaves (for a detailed description, see Supplemental Methods 5 online). PEPCK1 fusion proteins were exclusively localized in the cytosol (Figure 7A) and N- and C-terminal MC9 fusion proteins in the nucleocytoplasm (Figures 7B and 7C), but not in the apoplast, possibly due to the GFP instability in the acidic apoplast environment (Ward, 1981; Patterson et al., 1997). To exclude that the nuclear GFP signal was due to diffusion of free GFP after proteolytic processing of the fusion protein, we evaluated the integrity of the MC9 fusion proteins by protein gel blot analysis (Figure 7D). An ~45-kD GFP:MC9 protein was found that corresponded with the 25-kD GFP protein fused with the large p20 subunit of MC9 (20 kD). In this case, the small p10 subunit (15 kD) was removed after autocatalytic activation of the protease (Vercammen et al., 2004), indicating that MC9 was active in the observed subcellular compartments. For the C-terminal MC9: GFP fusion, an ~60-kD protein correlated with the full length of the fusion protein, revealing that the GFP fusion at the C terminus of MC9 might have interfered with the proteolytic separation of the two subunits. However, MC9 localized in the same cellular compartments irrespective of the GFP fusion conformations.

To confirm the nucleocytoplasmic localization of MC9, we additionally analyzed *Arabidopsis* roots of the ProMC9:MC9: GFP reporter lines. Interestingly, the localization of MC9 in the nucleus and cytoplasm of epidermal cells was confirmed (Figure...
7E). Altogether, our localization study demonstrated that MC9, besides its apoplastic location, is also present in nucleus and cytoplasm.

DISCUSSION

To date, our knowledge on metacaspase substrates is limited to the At Serpin 1 inhibitor of At MC9 (Vercammen et al., 2004), the udor staphylococcal nuclease (TSN) target of Norway spruce mcII-pa (Sundström et al., 2009), the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) target of YCA1 (Silva et al., 2011), the metacaspase Tb MCA4 target of Tb MCA3 (Proto et al., 2011), and the Escherichia coli elongation factor (EF-Tu) target of Tb MCA2 (Moss et al., 2007). As EF-Tu was cleaved upon the bacterial production of the recombinant Tb MCA2, the EF-Tu cleavage had not been demonstrated to be a physiologically relevant proteolytic event (Moss et al., 2007). Interestingly, we also identified an elongation factor Tu family protein (AT4G02930) to be cleaved in vitro and in vivo by MC9 (Table 1). Furthermore, although the cleavage sites were not identical as reported previously, the Arabidopsis TSN-1 (AT5G07350), TSN-2 (AT5G61780), and GAPDH B subunit (AT1G42920), a protein sharing 37% amino acid sequence similarity to the YCA1 substrate GAPDH, were all cleaved by rMC9 in vitro at conserved Arg or Lys residues (see Supplemental Data Set 1C online). Therefore, it is tempting to hypothesize that metacaspase substrates are conserved across...
understanding of proteolysis in biological processes. Lysosomal proteolysis (Agard et al., 2012). In that study, the catalytic approach recently used to determine the kinetics of caspase-9 activity was established Arg/Lys specificity, the combined in vitro and in vivo approaches have revealed a set of 74 potential physiological MC9 substrates (Table 1) and have led to two revised cleavage signatures of MC9 for either cleavage after Arg or Lys, containing basic residues (Arg or Lys) at substrate position P3 and acidic residues (Asp or Glu) at position P1. To exclude the possibility that the MC9 substrate selection was biased by stress responses induced by the MC9 mutation in the transgenic plant lines and led to differential protein upregulation or downregulation, we examined the expression levels of the endoplasmic reticulum stress response marker genes (Liu et al., 2007) for the method description, see Supplemental Methods 2 online). This stress response, also known as the unfolded protein response, triggers the protein folding activities in the endoplasmic reticulum by the induction of marker genes. As shown in Supplemental Figure 7 online, none of these marker genes was induced in the MC9 gain-of-function and loss-of-function lines, indicating that the mc9 mutation did not provoke a stress response.

By two different in vitro approaches, we validated the MC9 cleavage of GRF5, PEPCK1, CSY3, GMD2, GRP7, PORB, trypsin inhibitor-2, NAD synthetase, a pentatricopeptide repeat-containing protein, and a LEA domain–containing protein. The functional importance of PEPCK1 cleavage by MC9 was demonstrated in planta because the PEPCK1 activity was enhanced by MC9 cleavage and it translated into improved seedling establishment. Until now, the functionality of MC9 remained elusive, mostly due to the lack of phenotypic aberrations in MC9 gain-of-function and loss-of-function mutants. Stirred by the discovery of the MC9-dependent cleavage of PEPCK1 and its role in seed germination and early seedling development, we report a hitherto unknown phenotypic aberration caused by a perturbation in MC9 expression. We believe that further confirmation of the metacaspase functionalities implied by this MC9 degradome study will greatly improve our knowledge of plant proteases.

**A Role for MC9-Dependent Proteolysis in Seed Development and Metabolic Processes**

Within the MC9 degradome, proteins of various functional categories are present, supporting the proposed multifunctionality of plant metacaspases (Tsiantisani et al., 2011). In analogy with caspases, multiple transcription and translation elongation factors, ribosomal proteins, and structural components are cleaved by MC9 (Crawford et al., 2012). Moreover, proteins involved in seed development, such as several LEA proteins, are frequently found among the cleaved proteins. LEA or LEA domain–containing proteins are small proteins encoded by multigene families in plants. They accumulate during late stages of seed development, are very hydrophilic, thermostable, and unfolded in solution, and might 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 other chaperone proteins from *Medicago truncatula*, such as a Gly-rich RNA binding protein similar to the GRP7 MC9 substrate, have been associated with seed longevity and protection against desiccation-induced protein aggregation (Chatelain et al., 2012). Possibly, MC9 cleavage of LEA proteins during water stress could induce the conformational changes necessary for LEA protein folding and their protective functions. This hypothesis is supported by the cleavage patterns of LEA proteins from which the MC9-dependent processing could not be deduced as a destructive event.

The MC9-dependent PEPCK1 proteolysis indicated its involvement in gluconeogenesis by regulating PEPCK1 activity. Although *Arabidopsis* PEPCK1 phosphorylation has not been demonstrated yet, the putative Ser-62 phosphosite that is phosphorylated in tomato (*Solanum lycopersicum*), cucumber, and Guinea grass is conserved in the *Arabidopsis* protein sequence (Leegood and Walker, 2003; Bailey et al., 2007). Hence, cleavage of PEPCK1 at the identified Arg-101 cleavage site removes the N-terminal fragment containing this putative inhibitory phosphorylated Ser residue and thus recovers the PEPCK1 activity. Despite the N-terminomics data indicating that the intensity of the PEPCK1 fragments was higher in Col-0 or 3S:S:MC9 than in *mc9* (Figure 5A), the PEPCK1 precursor was reduced strikingly only in the 3S:S:MC9 extract, without the appearance of discrete proteolytic fragments (Figure 5C), possibly because cleavage fragments at the COFRADIC-identified sites are unstable or processed further downstream, thereby losing the necessary epitopes for antibody recognition. Although high-quality MS/MS data from single peptides accurately identify protein N termini and sites of proteolytic processing, they cannot always be detected by protein gel blot analysis or other gel-based validation methods (Timmer et al., 2009).

Another MC9 substrate involved in metabolic processes is CSY3, an enzyme catalyzing the conversion of acetyl-CoA and oxaloacetate to citrate at the onset of the glyoxylate cycle in germinating seeds. Notably, the glyoxylate cycle and gluconeogenesis are consecutive steps in the catabolism of fatty acids into sugars during seed germination and seedling establishment suggesting that this pathway regulation by MC9 may happen at different levels.

**Metacaspase Maturation**

As MC9 is autocatalytically activated by proteolytic separation of its p20 and p10 subunits (Vercammen et al., 2004), it was not surprisingly detected in the in vivo degradome (Table 1). However, the autoprocessing site Arg-183 found upon bacterial overproduction was not identified. The neo-N-terminal MC9 peptide identified in both 3S:S:MC9 and Col-0 N-terminomes was generated after cleavage at Arg-214, which is the first Arg...
residue following the previously reported autoprocessing site of rMC9 (Arg-183). It is possible that only the neo-N-terminal peptide R15DAGLKC21F and not the corresponding peptide after cleavage at Arg-183 and Arg-214 has been detected by MS/MS because the latter is too long. A double processing step during zymogen maturation has also been reported for most procaspases that are processed at two Asp residues (Fuentes-Prior and Salvesen, 2004). Similarly, the Norway spruce metacaspase mcII-Pa is processed at Arg-188 and Lys-269 (Bozhkov et al., 2005). Therefore, we suggest that MC9 maturation involves processing at two sites, at Arg-183 and Arg-214, but whether cleavage at Arg-214 is necessary for proteolytic activity remains to be established. The recombinant MC4 is also processed at another site besides the primary site at Lys-225 (Watanabe and Lam, 2011b), whereas Arg-190, a site that is conserved in MC5, corresponds to one of the two processing sites in mcII-Pa (Bozhkov et al., 2005).

MC9 Specificity Properties

Thanks to the analysis of the identified cleavage sites in the in vivo and in vitro MC9 degradomes, we could refine the MC9 cleavage signature. We demonstrated that Lys is relatively preferred over Arg at the P1 site, a finding not completely in agreement with the obtained PS-SCL results (Vercammen et al., 2006). Additionally, we observed differences in the specificity derived from the in vitro COFRADIC and the PS-SCL data concerning the P2 site. One obvious explanation might be that native proteins were analyzed with COFRADIC and that the protease interaction with its substrates took place under physiological conditions, whereas with PS-SCL, small synthetic tetrapeptides were used (Turk et al., 2012). Moreover, in the calculation of the PS-SCL scores, the natural amino acid occurrence was not taken into account. Due to the limitation of PS-SCL in screening only sequences N-terminal to the cleavage bond, the role played by the C-terminal sequences in steering the MC9 specificity could not have been detected.

We demonstrated that the P1-site plays an important role in determining the enzyme specificity and is often occupied by acidic amino acids. This property also might differentiate metacaspases from caspases, with the latter generally preferring small amino acid residues, such as Ser and Ala at P1’ (Timmer and Salvesen, 2007), besides the strong metacaspase preference for basic residues at P1. Another important observation in accordance with the known tetrapeptide MC9 specificity (Vercammen et al., 2006) is the MC9 preference for basic P3 and P1 residues. In this way, KxK, KxR, RxR, and RxxK combinations as in VRPR-AMC were very frequently cleaved in native seedling proteins by MC9. Although this result could hint at substrate cooperativity (Schilling and Overall, 2008) between P3 and P1 for protease specificity, such a conclusion remains premature without supporting structural and kinetics data.

The crystal structure of the T. brucei type-I metacaspase MCA2 has recently been resolved shown to act as monomer activated in the presence of its substrate and calcium upon removal of a N-terminal domain normally blocking substrate access to the active site (McLuskey et al., 2012). As type-II metacaspases are found in plants only and, in contrast with type-I metacaspases, their processing is required for their activity (Vercammen et al., 2007), we assume there is a different activation mechanism for type-II metacaspases. Therefore, resolution of a type-II metacaspase structure is highly anticipated because it will help explain how substrates are recognized and bound.

Perspectives

The comprehensive catalog of MC9 substrates presented here will help follow-up studies on the functionality and relevance of metacaspase-dependent proteolysis. Production and characterization of gain-of-function and loss-of-function mutants, in which either the selected substrate levels are perturbed or the metacaspase cleavage sites are mutated, will provide more detailed insights into the specific roles of plant metacaspases in various biological processes. Indeed, activities and expression of metacaspases have been reported, for example, during xylem differentiation (Bollhöner et al., 2013). In addition, similar degradome studies of other Arabidopsis metacaspases, such as MC4, will unravel further the cleavage specificity of plant metacaspases and the potential redundancy or exclusivity in their substrates.

METHODS

MC9 Expression Analysis

ProMC9:GUS reporter lines were generated as described in Supplemental Methods 1 online, and plants at different stages were used for the expression analysis of MC9. In summary, whole plants or separate organs were incubated in 90% (v/v) acetone for 10 min at room temperature, washed in PBS, pH 7.4, and then incubated in 1 mg/mL 5-bromo-4-chloro-3-indolyl-β-D-glucuronide, 2 mM ferricyanide, and 0.5 mM ferrocyanide in 100 mM phosphate buffer, pH 7.4, at 37°C in the dark. The culture was stirred at 80 to 100 rpm for 2 h to overnight. The material was cleared with 85% lactate solution and observed under stereomicroscopy and conventional light microscopy. Total RNA was extracted from 2-d-old seedlings as described (Meng and Feldman, 2010) with modifications. cDNA was prepared with the iScript cDNA synthesis kit (Bio-Rad) according to the manufacturer’s instructions. Real-time quantitative PCR was performed in a final volume of 20 µL, including 5 µL of eightfold diluted cDNA, 250 nM of each primer, and 50% (v/v) SYBR green mix (Invitrogen), with an iCycler instrument (Bio-Rad), according to the manufacturer’s instructions. The housekeeping gene UBIQUITIN C (AT5G25760) was used as internal reference and for normalization of transcript levels. The gene-specific primer sequences used for the quantitative RT-PCR are listed in Supplemental Table 1 online. For a detailed description, see Supplemental Methods 2 online.

N-Terminal COFRADIC: Sample Preparation, Peptide Identification, and Quantification

Seeds of Arabidopsis thaliana, accession Col-0, the MC9 T-DNA insertion line (GK-506H04), and the MC9 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 d, seeds were transferred to 21°C with a 16-h-light/8-h-dark photoperiod with 80 to 100 μmol m−2 s−1 light intensity and left to germinate for 48 h. Young seedlings, including seed coats, were collected, frozen in liquid nitrogen, and ground into a fine powder with mortar and pestle. To achieve a total protein content of ~4 mg, 0.5 g frozen ground tissue was resuspended in 1 mL buffer containing...
1% (w/v) 3-[3-Cholamidopropyl]-dimethylammonio]-1-propanesulfonate (CHAPS), 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM ethyl-enediaminetetraacetic acid, and 10% glycerol in PBS, pH 7.5, further containing the suggested amount of protease inhibitor mixture (one tablet/10 mL buffer) according to the manufacturer’s instructions (Roche Applied Science). The sample was centrifuged at 16,000g 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 MC9 optimal activity buffer consisting of 50 mM MES, pH 5.5, 150 mM NaCl, 10% (w/v) Suc, 0.1% (w/v) CHAPS, and 10 mM dithiothreitol. Four milligrams of total seedling protein extract (48 h after germination) was incubated at 30°C with recombinant MC9 (Vercammen et al., 2004) for a total duration of 1 h; every 20 min, 10 μg of fresh protease (855 units) was preactivated in the same optimal activity buffer at room temperature for 15 min and then added to the protein extract to a total amount of 30 μg protease. Enzyme activity was 57 units μL−1, and one unit corresponded to the enzyme activity that catalyzes the hydrolysis of 1 μmol VPRLP-AMC per min at 30°C. The same procedure was used for the control sample, but without addition of rMC9. By means of the gel analysis option in the ImageJ software (Abramoff et al., 2004), the level of endogenous MC9 protein in Col-0 extracts was estimated to be 14-fold lower than in rMC9 proteins and ~10-fold lower than the total amount of recombinant MC9 (30 μg) used in the in vitro COFRADIC analysis. Protein concentrations were measured with the DC protein assay (Bio-Rad), and protein extracts were further modified for N-terminal COFRADIC analysis as described (Staes et al., 2011).

All α-amino acids, either preexisting or generated upon proteolysis, and all ω-amino (Lys residues) were blocked with N-hydroxysuccinimide esters of butyric acid–stable isotopic variants. For the analysis of the in vivo–generated samples, primary amines in the mc9 proteome were 13C4-butyrylated, whereas Col-0 or 35S:MC9 primary amines were 13C2-butyrylated. For the in vitro analysis, primary amines in the rMC9-treated sample were 13C2-butyrylated and the untreated proteome was 12C2-butyrylated. After equal amounts of the labeled proteomes had been mixed, trypptic digestion generated internal, non-N-terminal peptides that were removed by strong cation exchange at low pH (Staes et al., 2011). After reversed phase-high performance liquid chromatography peptide separations with 2,4,6-trinitrobenzene sulfonic acid modification, internal peptides were further depleted and N-terminal peptides were isolated. Finally, liquid chromatography–MS/MS analysis was used for peptide identification and quantification (for a more detailed description, see Supplemental Methods 3 online).

Given that Lys residues were all butyrylated and, therefore, uncleavable by trypsin (Gevaert et al., 2003; Van Damme et al., 2005, 2010; Staes et al., 2011), we expected to identify only tryptic peptides generated after Arg (Arg-C specificity; C-terminal to Arg). Therefore, in the search settings with one missed cleavage allowance, endoproteinase semi-Arg C/P was used that allows the identification of neo-N-terminal peptides generated after Arg-C cleavage at one terminus (using the “semi” setting) and cleavage events between Arg and Pro as well. Variable modifications were set to pyroglutamate formation of N-terminal Glu and fixed modifications included Met oxidation to its sulfoxide derivative, S-carbamidomethylation of Cys, and butyrylation (13C2 or 12C2) of the Lys side chain and peptide N termini. Identified peptides were quantified with the Mascot Distiller Toolbox (version 2.3.2.0; Matrix Science) in the precursor mode.

**Assessment of Protein Cleavage**

cDNA clones received from the ABRC were used as templates for in vitro transcription and translation (77T73 Coupled Reticulocyte Lysate System; Promega) of 35S-radiolabeled proteins, according to the manufacturer’s instructions. Translation products were prepared in 50-μL reactions, and one-tenth was used for each rMC9 digestion. The production of the recombinant MC9 protease has been described previously (Vercammen et al., 2004), and a range of concentrations (0 to 1000 nM) prepared in the MC9 activity buffer were included in the digestions. The reactions were incubated at 30°C for 30 min. The inactive MC9 mutant (rMC9C147A-C29A) was used as a negative control (Belenghi et al., 2007). One-tenth of the digestion assay products was separated with SDS-PAGE (Laemmli, 1970). These gels were dried and exposed to phospho-imaging plates and scanned by a phosphor imager (445Si; Molecular Dynamics). PageRuler prestained protein ladder was used as protein marker (Thermo Scientific), and apparent molecular weight was calculated with ImageJ software (Abramoff et al., 2004).

**Recombinant PEPCK1 Production**

The Arabidopsis PEPCK1 (AT4G37870) open reading frame was amplified by PCR from cDNA of Col-0 seeds, extended to complete attB1 and attB2 Gateway cloning sites, and cloned into the bacterial protein expression vector pDEST17 (primer sequences are listed in Supplemental Table 1 online). Bacterial (BL21-DE3-pLysS) protein production was induced in LB broth–growing cultures of successfully transformed bacterial clones with 0.2 mM isopropyl β-D-1-thiogalactopyranoside at 20°C overnight. Denatured PEPCK1 protein was isolated from protein inclusion bodies and subsequently purified on 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 the Arabidopsis PEPCK1 antibodies, 0.2 mg of purified recombinant protein was injected four times as antigen into each rabbit (Eurogentec). The rabbit antiserum was used at a 1:5000 dilution for PEPCK1 immunodetection in plant extracts.

**Protein Immunodetection**

Proteins were immunodetected in 2-d-old Arabidopsis seedling extracts (or infiltrated tobacco [Nicotiana benthamiana] leaves) prepared in 1% (w/v) CHAPS, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM ethyl-enediaminetetraacetic acid, and 10% glycerol in PBS at pH 7.5 and containing the suggested amount of protease inhibitor mixture (one tablet/10 mL buffer) according to the manufacturer’s instructions (Roche Applied Science). Extracts were centrifuged at 16,000g for 10 min at 4°C, and cleared supernatants were immediately mixed with SDS-PAGE loading buffer (Laemmli, 1970) and heated at 95°C for 5 min. Of the protein extract, 40 μg was used for the immunodetection of PEPCK1, MC9, CAT2 (AT4G35090), or GFP. For reprobing purposes, Immobilon P (Millipore) membranes were stripped with 0.1 M 2-mercaptoethanol, 2% SDS, and 62.5 mM Tris-HCl, pH 6.7. Immunoreactive proteins were visualized with a chemiluminescence kit (Western Lightning Plus-ECL; Perkin-Elmer).

**PEPCK1 Activity Assay**

Seeds were set to germinate on wet Whatman paper for 6 d at 21°C after a 3-d stratification period at 4°C in the dark. Sampling included the remaining seed coats. The carboxylation activity of PEPCK1 was measured spectrophotometrically at 340 nm as previously described (Malone et al., 2007) on UV-Star 96-well plates (Greiner Bio-One). One unit of PEPCK1 activity corresponded to the production of 1 μmol product per min at 25°C. The values represent the mean of five measurements ±95% confidence interval (Cumming et al., 2007).

**Hypocotyl Elongation Assay**

During the hypocotyl elongation experiments, seeds were sown on half-strength Murashige and Skoog salts without Suc and then stratified for 3 d at 4°C in the dark. Prior to transfer to a Lovibond incubator (24 h in the dark at 21°C), seeds were exposed for 30 min to light. Hypocotyl lengths were measured to the nearest millimeter with the ImageJ software.
Subcellular Protein Localization

N- and C-terminal GFP fusions of MC9 and PEPCK1 were generated as described in Supplemental Methods 5 online. For imaging, a confocal microscope 100M with software package LSM 510 version 3.2 (Zeiss) was used. Excitation was done with the 488-nm line of an argon laser. Emission fluorescence for GFP and for propidium iodide was captured via a 500- to 550-nm band-pass filter and a 585-nm long-pass filter, respectively.

Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: AT5G04200, AT4G37870, AT1G79350, AT1G55090, AT3G51160, AT1G02150, AT4G37870, AT1G79340, AT1G47540, AT5G16050, AT2G321660, AT4G37870, AT1G79340, AT1G79320, AT1G79310, and AT2G21660, AT4G25110, AT5G64240, AT1G55090, AT3G51160, AT1G79330, AT1G79320, AT1G79310, and AT1G16420. Furthermore, the complete set of accession numbers listed numbers: AT5G04200, AT4G37870, AT2G42560, AT1G55090, AT3G51160, AT3G51160, AT1G02150, AT2G321660, AT4G37870, AT1G79350, AT1G55090, AT3G51160, AT1G79330, AT1G79320, AT1G79310, and AT1G16420. Furthermore, the complete set of accession numbers listed in this study is included in the Supplemental Data Set 1 online.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. General Workflow of the Proteomic Methodology and Experimental Setups.

Supplemental Figure 2. Venn Diagram Generated by VENNY (Oliveros, 2007).

Supplemental Figure 3. MC9 Specificity Monitored via Iys in Vitro Proteome-Wide Substrate Profiling.

Supplemental Figure 4. iceLogo Illustrating the Amino Acid Frequencies in Substrate Positions P4–P3 (Colaert et al., 2009).

Supplemental Figure 5. Reverse-Phase HPLC UV Absorbance Chromatogram Profiles of the Trypsin Inhibitor 2, GRF5, GRP7, and iORB Oligopeptides with the MC9 Cleavage Sites.

Supplemental Figure 6. Reverse-Phase HPLC UV Absorbance Chromatogram Profiles of the PEPPK1, GR7P, and a LEA Domain-Containing Protein Oligopeptides with the MC9 Cleavage Sites.

Supplemental Figure 7. Expression of the UPR Marker Genes in the Wild-Type (Col-0), the Loss-of-Function (mc9), and the Gain-of-Function (Colaert et al., 2009).

Supplemental Table 1. Oligonucleotides Used in This Study.


Supplemental Methods 2. RNA Extraction and qRT-PCR Gene Expression Assays.

Supplemental Methods 3. Peptide Identification and Quantification.

Supplemental Methods 4. Assessment of Synthetic Peptide Cleavage.

Supplemental Methods 5. Subcellular Protein Localization.

Supplemental Data Set 1A. List of Potential Endogenous MC9 Substrates Identified in the Col-0 Proteome.

Supplemental Data Set 1B. List of Potential Endogenous MC9 Substrates Identified in the 35S:MC9 Proteome.

Supplemental Data Set 1C. List of Potential Endogenous MC9 Substrates Generated by the rMC9-Driven Proteolysis Added to the mc9 Proteome.

Supplemental Data Set 1D. Arabidopsis Proteins That Contain the Suggested MC9 Specificity Consensus Sequences.

ACKNOWLEDGMENTS

We thank Richard Leegood for useful discussion and the kind gift of the pepck1 T-DNA insertional mutant seeds, Davy Maddelein for the generation of heat maps from the PS-SCL data of Figure 3C, Pierre Hilson for fruitful comments, and Martine De Cock for help in preparing the article. This work was supported by the Research Foundation-Flanders (Grant G.0038.09N) and Ghent University (Multidisciplinary Research Partnership “Biotechnology for a Sustainable Economy” project; grant 01MRPS10W). L.T. was the recipient of a VIB International PhD Program predoctoral fellowship. D.V. was and P.V.D. is a Postdoctoral Fellow of the Research Foundation-Flanders. S.S. is indebted to the Special Research Fund of Ghent University for a postdoctoral fellowship.

AUTHOR CONTRIBUTIONS

L.T., K.G., and F.V.B. conceived the project. L.T., D.V., and P.V.D. designed the experiments. L.T., E.T., P.-J.D.B., T.B., and S.S. performed the research. L.T. and A.S. analyzed the data. B.V.d.C. and M.G. contributed to the development of analytic tools. L.T., with the help of K.G. and F.V.B., wrote the article.

Received June 20, 2013; revised July 19, 2013; accepted July 31, 2013; published August 20, 2013.

REFERENCES


