Andreas Porodko, Ana Cirnski, Drazen Petrov, Teresa Raab, Melanie Paireder, Bettina Mayer, Daniel Maresch, Lisa Nika, Martin L. Biniossek, Patrick Gallois, Oliver Schilling, Chris Oostenbrink, Marko Novinec and Lukas Mach*

The two cathepsin B-like proteases of *Arabidopsis thaliana* are closely related enzymes with discrete endopeptidase and carboxydipeptidase activities

Abstract: The genome of the model plant *Arabidopsis thaliana* encodes three paralogues of the papain-like cysteine proteinase cathepsin B (AtCathB1, AtCathB2 and AtCathB3), whose individual functions are still largely unknown. Here we show that a mutated splice site causes severe truncations of the AtCathB1 polypeptide, rendering it catalytically incompetent. By contrast, AtCathB2 and AtCathB3 are effective proteases which display comparable hydrolytic properties and share most of their substrate specificities. Site-directed mutagenesis experiments demonstrated that a single amino acid substitution (Gly$^{336}\rightarrow$Glu) is sufficient to confer AtCathB2 with the capacity to tolerate arginine in its specificity-determining S2 subsite, which is otherwise a hallmark of AtCathB3-mediated cleavages. A degradomics approach utilizing proteome-derived peptide libraries revealed that both enzymes are capable of acting as endopeptidases and exopeptidases, releasing dipeptides from the C-termini of substrates. Mutation of the carboxydipeptidase determinant His^{207} also affected the activity of AtCathB2 towards non-exopeptidase substrates, highlighting mechanistic differences between plant and human cathepsin B. This was also noted in molecular modeling studies which indicate that the occluding loop defining the dual enzymatic character of cathepsin B does not obstruct the active-site cleft of AtCathB2 to the same extent as in its mammalian orthologues.

Keywords: cysteine proteinase; degradomics; exopeptidase; plant; substrate profiling.

Introduction

Almost all biochemical data currently available on cathepsin B have been obtained studying its animal variants. As is typical for papain-like cysteine proteinases (PLCPs), the human form of the protease is characterized by a broad substrate specificity (Choe et al., 2006; Biniossek et al., 2011). However, the enzyme differs from all other mammalian PLCPs characterized to date with respect to its dual capacity of acting both as an endopeptidase and a carboxydipeptidase (Mort and Buttle, 1997; Turk et al., 2012). It has been shown that this ability to progressively remove dipeptides from the C-termini of substrates is associated with the presence of a unique structural element termed the occluding loop, which is absent from other members of the

*Corresponding author: Lukas Mach, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria,
e-mail: lukas.mach@boku.ac.at
Andreas Porodko, Teresa Raab and Melanie Paireder: Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria
Ana Cirnski and Marko Novinec: Department of Chemistry and Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Vecna pot 113, SI-1000 Ljubljana, Slovenia
Drazen Petrov and Chris Oostenbrink: Institute for Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria
Bettina Mayer and Martin L. Biniossek: Institute for Molecular Medicine and Cell Research, Faculty of Medicine, University of Freiburg, Stefan-Meier Strasse 17, D-79104 Freiburg, Germany

Daniel Maresch: Department of Chemistry, University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria
Lisa Nika: Department of Biotechnology, University of Natural Resources and Life Sciences, Muthgasse 11, A-1190 Vienna, Austria
Patrick Gallois: School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK
Oliver Schilling: Institute for Molecular Medicine and Cell Research, Faculty of Medicine, University of Freiburg, Stefan-Meier Strasse 17, D-79104 Freiburg, Germany; and BIOSS Centre for Biological Signaling Studies, University of Freiburg, Stefan-Meier Strasse 17, D-79104 Freiburg, Germany

papain superfamily (Illy et al., 1997). On the other hand, the occluding loop blocks a segment of the active-site cleft and thereby restricts the endopeptidase activity of the protease (Nägler et al., 1997). Crystallographic and mutational analyses have shed further insights into the mechanisms which allow vertebrate cathepsin B to operate as an exopeptidase. In particular, two histidine residues located in the occluding loop have been defined as critical determinants for the carboxydipeptidase activity of the human enzyme (Musil et al., 1991; Krupa et al., 2002).

Far less is known about the enzymatic properties of plant cathepsin B. A cathepsin B-like protease purified from daikon radish (*Raphanus sativus*) cotyledons was found to possess both endopeptidase and carboxydipeptidase activities when tested with various oligopeptides, although the action as an exopeptidase was less prevalent than in the case of the human enzyme (Tsuji et al., 2008). Similar observations were made for *Nicotiana benthamiana* cathepsin B (NbCathB), which displays only moderate reactivity with a synthetic substrate designed as a specific probe for the exopeptidase activity of human cathepsin B (Niemer et al., 2016). This has led to the proposal that NbCathB functions primarily as an endopeptidase, whose substrate preferences complement those of other PLCPs involved in the turnover of endogenous proteins *in planta* (Paireder et al., 2017). In this context, it is of note that the occluding loop of plant cathepsin B-like proteases is considerably shorter than in their animal relatives and contains only one histidine residue, which could account for the observed catalytic differences (Tsuji et al., 2008).

Besides its documented role in programmed cell death (McLellan et al., 2009; Ge et al., 2016; Cai et al., 2018), *Arabidopsis thaliana* cathepsin B (AtCathB) has been also studied in the context of other physiological processes (Iglesias-Fernandez et al., 2014; Lu et al., 2015). Although three cathepsin B-like loci (at1g02300, *AtCathB1*; at1g02305, *AtCathB2*; at4g01610, *AtCathB3*) have been identified in the *A. thaliana* genome, only AtCathB3 has been subjected so far to a partial characterization of its enzymatic properties (Richau et al., 2012; Lu et al., 2015; Ge et al., 2016). The biochemical features of AtCathB1 and AtCathB2 are still largely unknown. It is noteworthy that the sequence deposited for AtCathB1 displays remarkable differences to the other two enzymes (McLellan et al., 2009). In the present study, we have therefore first characterized the *AtCathB1* transcripts present in *A. thaliana* leaves. As none of them was found to encode a catalytically competent protein, all further studies were confined to AtCathB2 and AtCathB3. The latter two proteases were produced recombinantly in insect cells and then tested for their capacity to hydrolyze a range of generic cathepsin B

substrates. We also assessed the substrate specificities of AtCathB2 and AtCathB3 by mass spectrometric detection of cleavage events in proteome-derived peptide libraries. Finally, site-directed mutagenesis was used to probe the importance of selected amino acids for the P2 preferences of AtCathB2 and its carboxydipeptidase activity.

Results

Characterization of *AtCathB1* transcripts expressed in *A. thaliana* leaves

AtCathB2 (UniProt: Q93VC9) and AtCathB3 (UniProt: Q94K85) show 80% identity on the protein level (McLellan et al., 2009) and display all features characteristic for plant cathepsin B-like proteases (Gilroy et al., 2007; Tsuji et al., 2008). The sequence deposited for AtCathB1 (UniProt: F4HVZ1) is less similar due to a mutation destroying the 3′-splice site of intron 5. The latter gave rise to the prediction of a hypothetical protein carrying an insertion in the vicinity of the active-site cleft (McLellan et al., 2009), accompanied by the deletion of two adjacent residues (Asp[124], Gln[125]; preproprotein numbering) considered of key importance for the catalytic action of cathepsin B (Musil et al., 1991; Krupa et al., 2002). This suggests that the predicted AtCathB1 polypeptide would display little if any hydrolytic activity. In the light of this ambiguity, we decided to reinvestigate the splicing status of the *AtCathB1* transcripts synthesized in *A. thaliana* leaves. *AtCathB1* is constitutively expressed in this tissue, albeit at a much lower level than *AtCathB2* and *AtCathB3* (Iglesias-Fernandez et al., 2014; Cai et al., 2018). Three major splice variants were identified by reverse transcription polymerase chain reaction (RT-PCR) combined with DNA sequencing, all of them differing from the predicted transcript. The top band (884 bp) of the detected triplet is derived from incompletely spliced *AtCathB1* mRNA still harboring introns 4 and 5. Removal of intron 4 accounts for the generation of the transcript giving rise to the 778-bp PCR product. The shortest fragment (638 bp) results from the additional excision of intron 5 utilizing a different splice site than proposed previously, thereby deleting 53 bp of exon 6 (Figure 1). The presence of either intron as well as the truncation of exon 6 leads to the formation of a premature stop codon and thus a truncated protein lacking most of the catalytic domain (Table S1). Hence, none of these transcripts supports the synthesis of an AtCathB1 variant with a functional active site. These results indicate that only two catalytically active versions of cathepsin B occur
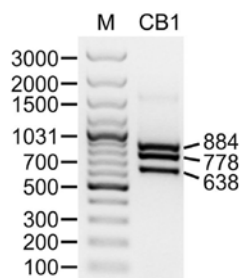
Figure 1: RT-PCR analysis of *AtCathB1* transcripts.
RT-PCR analysis was performed on RNA extracted from *A. thaliana* rosette leaves using primers encompassing exons 2–6 of the *AtCathB1* (CB1) coding region. PCR products were then separated by 1.5% agarose gel electrophoresis and visualized by ethidium bromide staining. The migration positions of DNA standards (M) are indicated, with their respective masses expressed in base pairs.

in *A. thaliana* leaves under normal growth conditions: AtCathB2 and AtCathB3.

Expression of AtCathB2 and AtCathB3 in insect cells

AtCathB2 is predicted to be synthesized as a 40-kDa pre-proenzyme of 362 amino acids, consisting of a 33-residue signal peptide, a latency-mediating proregion (72 amino acids) and the C-terminal catalytic domain (257 amino acids). The active-site residues typical for PLCPs are all present (Gln[128], Cys[134], His[289] and Asn[310]; Table S2). AtCathB2 contains four potential N-glycosylation sites, two (Asn[35], Asn[72]) present in the propeptide and the other two (Asn[154], Asn[313]) located in the catalytic domain. The theoretical molecular mass of pro-AtCathB2 is 36.5 kDa, whereas that of the catalytic domain is 28.4 kDa (not accounting for the presence of N-linked oligosaccharides). Prepro-AtCathB3 is a 359-residue protein with a predicted molecular mass of 39.4 kDa. The overall architecture of AtCathB3 is essentially the same as outlined for AtCathB2, the difference in length being accounted for by a shorter signal peptide (30 residues). Interestingly, AtCathB3 contains only two consensus motifs for N-glycosylation (Asn[69] and Asn[151]). A hallmark of cathepsin B-like proteases is the so-called occluding loop, a unique structural element affecting substrate binding to the prime-site region of the active-site cleft. As typical for plant cathepsin B variants (Tsuji et al., 2008), the occluding loop of AtCathB2 and AtCathB3 is shorter than the corresponding sequence in human cathepsin B (Illy et al., 1997) and lacks one of the two histidine residues supporting the carboxydipeptidase activity of the latter enzyme (Krupa et al., 2002; Figure S1).

It has been demonstrated recently that AtCathB3 can be functionally expressed in insect cells (Ge et al., 2016). Hence, we have chosen the baculovirus expression system for the production of recombinant AtCathB2 and AtCathB3. Upon baculovirus-mediated transduction of pro-AtCathB3, an epitope-tagged 40-kDa polypeptide could be detected in cell extracts of virus-infected *Spodoptera frugiperda* Sf21 cells. This protein was found to be secreted efficiently into the culture supernatant, which was accompanied by its transformation into a 38-kDa polypeptide (Figure 2). We have previously shown that the intracellular forms of secretory glycoproteins overexpressed in insect cells are usually trapped within the endoplasmic reticulum and thus decorated with oligomannosidic N-linked oligosaccharides. By contrast, their secreted counterparts are modified with smaller N-glycan structures (Kowalewski-Nimmerfall et al., 2014; Niemer et al., 2016). Hence, the apparent size difference between intracellular and secreted recombinant AtCathB3 is probably due to such variations in N-glycosylation.

Essentially the same results were obtained for wild-type AtCathB2 and two variants of this enzyme featuring point mutations within the occluding loop (His[207]→Ala) or the S2 subsite (Gly[336]→Glu), thus demonstrating that these substitutions are well tolerated by the protease. In analogy to human cathepsin B (Musil et al., 1991), His[207] is predicted to form a salt bridge with Asp[127]. Since this ion-pair is important for correct positioning of the occluding loop (Krupa et al., 2002), we have also aimed at mutating the latter residue. However, replacement of Asp[127] with Ala
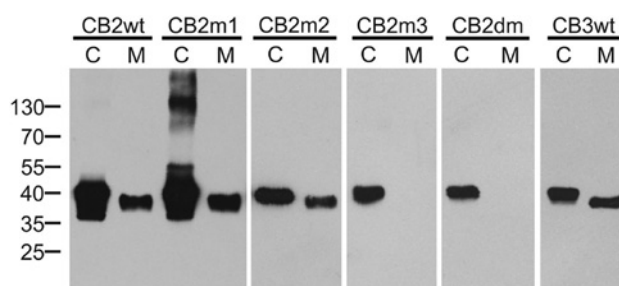

Figure 2: Expression of AtCathB variants in insect cells.
Cell extracts (C) and conditioned media (M) of Sf21 cells infected with recombinant baculoviruses encoding wild-type (wt) or mutant AtCathB2 (CB2) or AtCathB3 (CB3) were separated by SDS-PAGE, transferred to a nitrocellulose membrane and probed with anti-FLAG antibodies. No signals were obtained with extracts and media of uninfected Sf21 cells. The migration positions of prestained molecular mass standards are indicated, with their respective masses expressed in kDa. The cellular CB2m1 aggregates of 130 kDa and above were not observed in other experiments. CB2m1, AtCathB2 His[207]→Ala; CB2m2, AtCathB2 Gly[336]→Glu; CB2m3, AtCathB2 Asp[127]→Ala; CB2dm, AtCathB2 Asp[127]→Ala/His[207]→Ala.

obliterated secretion of AtCathB2 completely both in the wild-type and the His[207]→Ala background (Figure 2). This indicates that mutation of Asp[127] compromises the folding competence of AtCathB2 to a larger extent than in the case of the human enzyme (Krupa et al., 2002).

Biochemical characterization of recombinant AtCathB2 and AtCathB3

Exploiting their N-terminal leader sequences containing hexahistidine and FLAG tags, wild-type AtCathB2 and AtCathB3 could be isolated by a two-step affinity chromatography procedure in high purity and reasonable yields (0.4 mg AtCathB2 and 1.2 mg AtCathB3 per 200 ml of insect cell culture). A single band of the expected size was observed upon sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of either purified proenzyme. A more detailed characterization of the recombinant proteins was achieved by liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) (Figure 3). The molecular mass of purified wild-type AtCathB2 was determined as 42340.9 Da, which is in very good agreement with the presence of two paucimannosidic $Man_3GlcNAc_2Fuc$

N-glycans as typical for secreted glycoproteins made in insect cells (Kowalewski-Nimmerfall et al., 2014), the predicted six disulfide bridges and the loss of the two most C-terminal amino acids Ser[361] and Phe[362] (predicted mass: 42340.8 Da). A small fraction of AtCathB2 was found to exhibit a mass of 42154.8 Da, which is consistent with the additional removal of another two amino acids (Val[359], Ser[360]) from the C-terminus of the protein (predicted mass: 42154.6 Da). In the case of recombinant AtCathB3, most molecules displayed a mass of 42258.5 Da. This matches the predicted mass of the full-length proenzyme modified with two N-glycans and containing six disulfide bridges (42257.6 Da). However, some C-terminal trimming was also detected for AtCathB3, the release of the last two residues (Ser[358], Val[359]) manifesting in a minor peak of 42072.4 Da. It has been demonstrated before that the carboxydipeptidase activity of cathepsin B is capable of mediating autocatalytic C-terminal processing of the enzyme (Rowan et al., 1993).

These data demonstrated that AtCathB3 is N-glycosylated at both tractable asparagine residues (Asn[69], Asn[151]). However, the attachment sites of the two AtCathB2 N-glycans remained to be determined. Three of its four potential N-glycan acceptors (Asn[35], Asn[72], Asn[154], Asn[313]) are located in tryptic peptides which could be detected by LC-ESI-MS.
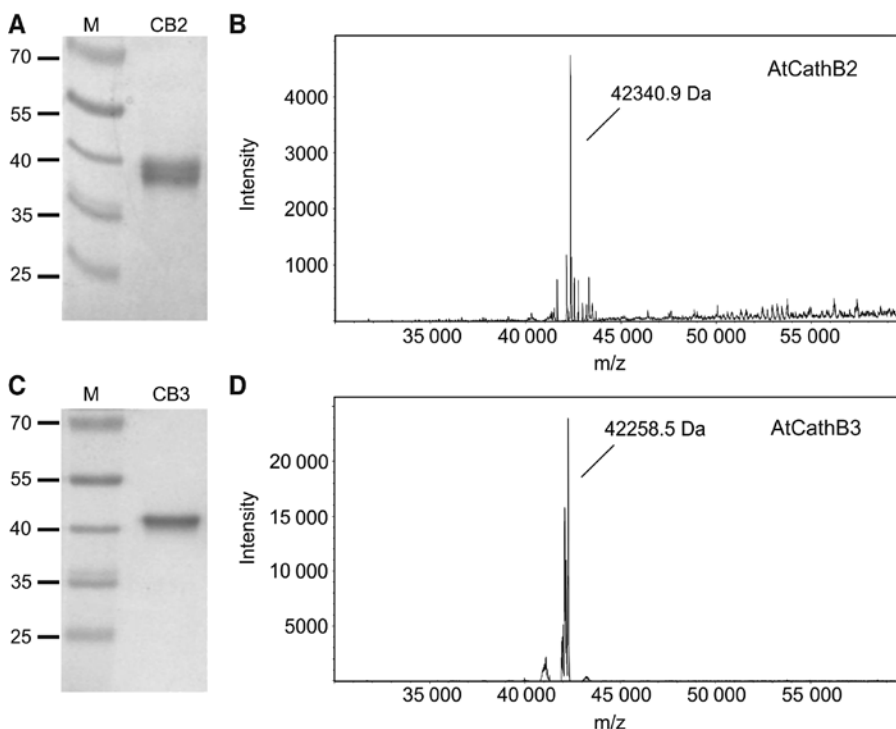

Figure 3: Characterization of recombinant AtCathB2 and AtCathB3.
Purified wild-type AtCathB2 (CB2) and AtCathB3 (CB3) were analyzed by SDS-PAGE and Coomassie Brilliant Blue staining (A, C) or LC-ESI-MS (B, D). The migration positions of prestained molecular mass standards (M) are indicated.

These analyses revealed that Asn[35] and Asn[313] are not glycosylated, while Asn[72] is modified with the classical paucimannosidic oligosaccharide $Man_3GlcNAc_2Fuc$ and minor amounts of closely related structures. Hence, the second AtCathB2 N-glycan has to be attached to Asn[154], although the glycosylation status of this asparagine residue could not be assessed directly (Figure S2).

AtCathB2 and AtCathB3 contain 14 cysteine residues within their proregions and catalytic domains, 12 of them being engaged in intramolecular disulfide bridges. Sequential labeling with two thiol-modifying reagents followed by mass spectrometric peptide mapping allowed the identification of Cys[331] (AtCathB2)/Cys[328] (AtCathB3) as the second free cysteine besides the catalytic nucleophile (AtCathB2: Cys[134]; AtCathB3: Cys[131]). Importantly, these studies also revealed that the two cysteine residues located in the occluding loop (AtCathB2: Cys[205]/Cys[210]; AtCathB3: Cys[202]/Cys[207]) form a disulfide bridge as initially observed in human cathepsin B (Musil et al., 1991).

Activity of AtCathB2 and AtCathB3 towards synthetic substrates

As was recently reported for NbCathB (Niemer et al., 2016), AtCathB2 and AtCathB3 can be activated by treatment with pepsin at acidic pH. A panel of synthetic cysteine protease substrates was then used for an initial characterization of the substrate preferences of wild-type AtCathB2 and AtCathB3, which were both found to be optimally active at pH 5.5 as reported previously (Ge et al., 2016). Either enzyme readily hydrolyzed a number of fluorogenic peptide derivatives containing phenylalanine, leucine or valine in the specificity-determining P2 position. AtCathB3 also cleaved benzyloxycarbonyl (Z)-His-Glu-Lys-4-methylcoumaryl-7-amide (MCA) and Z-Gly-Pro-Arg-MCA reasonably well. By contrast, the activity of AtCathB2 towards these two substrates was substantially weaker. A particularly striking difference between wild-type AtCathB2 and AtCathB3 was observed with respect to arginine as P2 residue. As already noted before (Ge et al., 2016), AtCathB3 displayed considerable activity against Z-Arg-Arg-MCA. By contrast, this substrate was far less sensitive to cleavage by wild-type AtCathB2 (Table S3).

A detailed kinetic analysis was undertaken for the hydrolysis of Z-Phe-Arg-MCA and Z-Arg-Arg-MCA by AtCathB2 and AtCathB3. In the case of Z-Phe-Arg-MCA, the catalytic efficiency of wild-type AtCathB2 was found to be higher than that of AtCathB3. The opposite was observed for Z-Arg-Arg-MCA, largely due to a much faster turnover of this substrate by AtCathB3. All K_m values were in the low micromolar range, with both enzymes displaying a higher affinity for Z-Phe-Arg-MCA (Table 1). This has also been observed for the AtCathB3 orthologue from *R. sativus* (Tsuji et al., 2008).

To study the catalytic properties of AtCathB2 in more depth, we generated a Gly[336]→Glu mutant with the aim to change the substrate-binding properties of its S2 subsite. It has been shown for human cathepsin B that a glutamic acid situated in this pocket of the enzyme is a key determinant

Table 1: Kinetic parameters of recombinant AtCathB variants.

Enzyme	Substrate	K_m [μm]	k_{cat} [s^{-1}]	k_{cat}/K_m [mm^{-1} s^{-1}]
AtCathB2	Z-Phe-Arg-MCA	6 ± 1	1.0 ± 0.1	174
	Z-Arg-Arg-MCA	22 ± 2	0.05 ± 0.01	2 (0.01)
	Abz-FRF(NO$_2$)A	0.47 ± 0.02	0.17 ± 0.01	362 (2.1)
	Abz-GIVRAK(Dnp)	0.9 ± 0.4	0.9 ± 0.1	957 (5.5)
His207Ala	Z-Phe-Arg-MCA	42 ± 6	0.14 ± 0.01	3.4
	Z-Arg-Arg-MCA	90 ± 27	0.005 ± 0.001	0.05 (0.01)
	Abz-FRF(NO$_2$)A	4 ± 1	0.02 ± 0.01	4.4 (1.3)
	Abz-GIVRAK(Dnp)	9 ± 6	0.015 ± 0.003	1.6 (0.5)
Gly336Glu	Z-Phe-Arg-MCA	12 ± 2	0.14 ± 0.01	12
	Z-Arg-Arg-MCA	5 ± 2	0.04 ± 0.01	8 (0.7)
	Abz-FRF(NO$_2$)A	2.0 ± 0.2	0.04 ± 0.01	18 (1.5)
	Abz-GIVRAK(Dnp)	1.1 ± 0.4	0.11 ± 0.01	104 (8.5)
AtCathB3	Z-Phe-Arg-MCA	9 ± 1	0.6 ± 0.1	65
	Z-Arg-Arg-MCA	17 ± 2	0.6 ± 0.1	34 (0.5)
	Abz-FRF(NO$_2$)A	3.7 ± 0.2	1.9 ± 0.5	518 (8.0)
	Abz-GIVRAK(Dnp)	3.1 ± 0.3	0.9 ± 0.2	290 (4.5)

Data are presented as mean ± SEM (n ≥ 3). Values in parentheses represent ratios relative to k_{cat}/K_m for Z-Phe-Arg-MCA. His207Ala, AtCathB2 His[207]→Ala; Gly336Glu, AtCathB2 Gly[336]→Glu.

for its interaction with substrates containing an arginine residue in P2 (Hasnain et al., 1993). AtCathB3 features also a glutamic acid at the equivalent position (Glu[333]), whereas this residue is replaced by glycine (Gly[336]) in AtCathB2. We reasoned that exchange of Gly[336] with glutamic acid could improve the capacity of AtCathB2 to act on Z-Arg-Arg-MCA. Although the Gly[336]→Glu substitution had a negative impact on the general catalytic properties of AtCathB2, the relative catalytic efficiency of the mutant towards Z-Arg-Arg-MCA was much higher than observed for the wild-type enzyme and comparable to that of AtCathB3. This suggests that the presence of Glu[333] (AtCathB3 numbering) influences the P2 specificity of plant cathepsin B in a similar manner as previously described for the human protease (Table 1).

Human cathepsin B has been shown to act both as an endopeptidase and an exopeptidase, removing dipeptides from the C-termini of substrates. While Z-Phe-Arg-MCA and Z-Arg-Arg-MCA are commonly used to study the endopeptidase activity of the enzyme, more sophisticated substrates are required for the measurement of its carboxydipeptidase activity. The most sensitive synthetic substrates currently available for the determination of the carboxydipeptidase activity of cathepsin B are the internally quenched fluorescent peptides 2-aminobenzoyl (Abz)-Phe-Arg-Phe(NO$_2$)-Ala [Abz-FRF(NO$_2$)A; Krupa et al., 2002] and Abz-Gly-Ile-Val-Arg-Ala-2,4-dinitrophenyl (Dnp)-Lys [Abz-GIVRAK(Dnp); Cotrin et al., 2004]. In either substrate, cathepsin B cleaves the bond following the Arg residue (Krupa et al., 2002; Cotrin et al., 2004). As shown recently for NbCathB (Niemer et al., 2016), these two substrates are also susceptible to cleavage by plant cathepsin B-like proteases. Hence, we have tested AtCathB2 and AtCathB3 for their capacity to hydrolyse Abz-FRF(NO$_2$)A and Abz-GIVRAK(Dnp). While the catalytic efficiency of both enzymes towards Abz-FRF(NO$_2$)A was lower than observed for NbCathB, AtCathB2 and AtCathB3 cleaved Abz-GIVRAK(Dnp) more avidly than their *N. benthamiana* relative. These results indicate that AtCathB2 as well as AtCathB3 possess substantial carboxydipeptidase activity, although discrete differences were noted with respect to their individual binding affinities and turnover numbers for Abz-GIVRAK(Dnp) and Abz-FRF(NO$_2$)A (Table 1).

The exopeptidase activity of human cathepsin B is strongly determined by two histidine residues present in the so-called occluding loop which projects into the active-site cleft, thereby blocking parts of the prime-site region (Musil et al., 1991). The strategic location of these residues in the S2′ subsite allows them to interact with the C-terminal carboxyl group of an exopeptidase substrate. Furthermore, one of the histidines is engaged in a salt bridge docking the occluding loop to the main body of the

enzyme, thus holding the loop in a closed position (Krupa et al., 2002). As typical for plant cathepsin B-like proteases (Tsuji et al., 2008), the occluding loop of AtCathB2 and AtCathB3 is considerably shorter than in human cathepsin B and contains only one histidine residue (Figure S1). We have therefore decided to investigate the role of this histidine (His[207]) in AtCathB2 by producing a His[207]→Ala mutant. In good agreement with data for the corresponding human cathepsin B mutants (Krupa et al., 2002), introduction of the His[207]→Ala substitution reduced the catalytic efficiency of AtCathB2 towards Abz-FRF(NO$_2$)A by two orders of magnitude. The impact of this mutation on hydrolysis of Abz-GIVRAK(Dnp) was even more pronounced. However, the k_{cat}/K_m values for cleavage of Z-Phe-Arg-MCA and Z-Arg-Arg-MCA by His[207]→Ala AtCathB2 were also 50-fold lower than in the case of the wild-type protease. Contrary to human cathepsin B (Nägler et al., 1997; Krupa et al., 2002), the effects of removing His[207] are therefore not limited to the exopeptidase activity of AtCathB2 and the accessibility of its active-site cleft (Table 1).

Different cathepsin B structures deposited in the Protein Data Bank were tested as templates for a homology-based model of AtCathB2. The best fit was obtained with the crystal structure of rat cathepsin B (Jia et al., 1995), yielding a backbone root-mean-square deviation of less than 1.3 Å. The most conspicuous differences between AtCathB2 and rat cathepsin B were noted in the region of the occluding loop, whose top segment is much shorter in AtCathB2 (Ser[206]-Gly[209]) than in the rat enzyme (10 residues). Hence, the occluding loop obstructs the active-site cleft of AtCathB2 far less than in the case of rat cathepsin B. The positions of Asp[127] and His[207] are compatible with a salt bridge formed by their side chains as observed for the equivalent residues Asp[22] and His[110] in the rat protease (Figure 4). Furthermore, the distance between His[207] and the catalytic dyad of AtCathB2 (Cys[134], His[289]) is comparable to that of His[110] in rat cathepsin B. This indicates that the main structural elements determining the carboxydipeptidase activity of human cathepsin B (Krupa et al., 2002) are well preserved in AtCathB2. Molecular dynamics (MD) studies revealed that the occluding loop of AtCathB2 is not particularly mobile, suggesting that substrates can enter the active site without substantial loop displacement (Figure S3).

Specificity profiling of AtCathB2 and AtCathB3 with proteome-derived peptide libraries

For a more comprehensive evaluation of the observed differences between the substrate preferences of wild-type
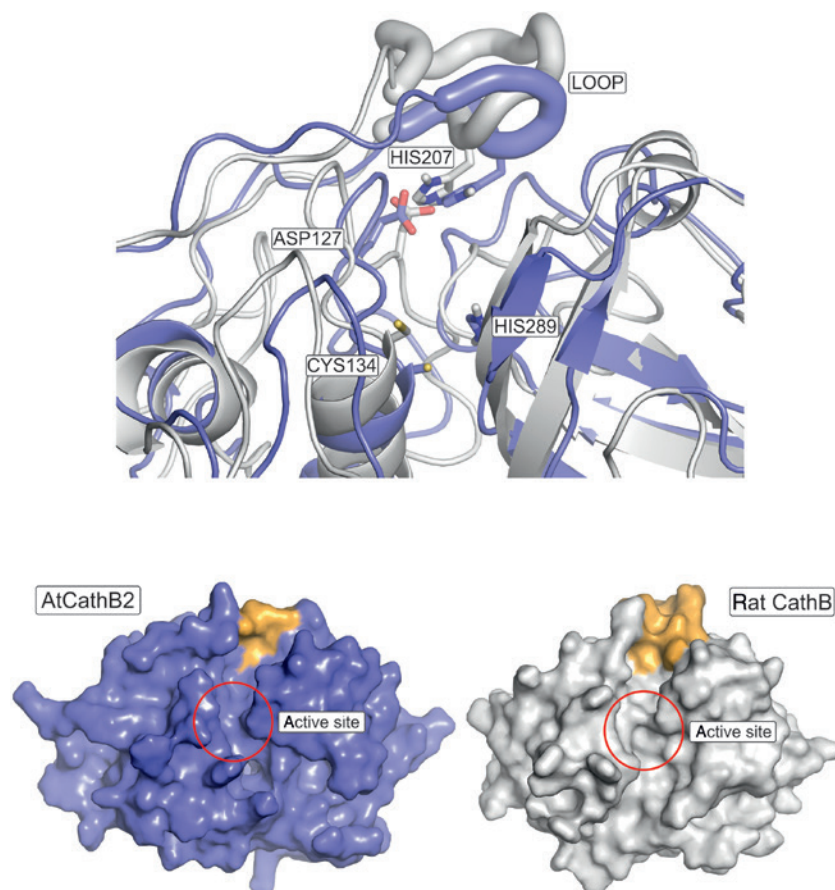
Figure 4: Molecular modeling of AtCathB2.
(Top) The occluding loops and four key amino acids are highlighted in the aligned structures of AtCathB2 (blue) and rat cathepsin B (gray). Asp[127] (AtCathB2 preproprotein numbering) forms a salt bridge with His[207] in the vicinity of the active-site residues Cys[134] and His[289]. The equivalent residues in mature rat cathepsin B are Asp[22], Cys[29], His[110] and His[199]. The latter residue is not shown since it would mask the position of AtCathB2 His[289]. (Bottom) Surface representations of AtCathB2 (blue) and rat cathepsin B (gray), with the active-site clefts displayed in perpendicular orientation. The active sites are highlighted. The occluding loops are shown in orange.

AtCathB2 and AtCathB3, we applied a modified version of the Proteomic Identification of Protease Cleavage Sites (PICS) procedure (Schilling et al., 2011; Biniossek et al., 2016). To avoid any potential interference by peptides present in the AtCathB samples, all PICS analyses were performed with substrate libraries prepared using *Escherichia coli* extracts. After digestion with a protease exhibiting a strict P1 specificity (GluC or LysC), the resulting mixtures of *E. coli* peptides were incubated with AtCathB2 or AtCathB3 at enzyme-to-library ratios of 1:10 and then subjected to reductive dimethylation with deuterated formaldehyde. After mixing with equivalent amounts of control samples labeled with natural formaldehyde, analysis by liquid chromatography coupled to mass spectrometry was performed to detect cleavage events mediated by AtCathB2 or AtCathB3 through semi-specific peptides enriched at least four-fold in the deuterated label. These data were then used to identify amino acids which are significantly enriched ($p < 0.05$) at the P3-P2' positions of AtCathB2 and AtCathB3 substrates (Figure 5).

Irrespective of the protease chosen for library generation, the PICS experiments provided firm evidence for the endo- and exopeptidase activities of wild-type AtCathB2 and AtCathB3. In the case of AtCathB2, processing of 1976 LysC- and 943 GluC-generated peptides was detected. A large number of cleavage events (LysC: 762; GluC: 523) was also observed for AtCathB3. Both enzymes exhibited pronounced endopeptidase activity (AtCathB2: 53%; AtCathB3: 23%) as defined by the liberation of prime-site fragments consisting of more than six residues. Notably, a considerable fraction of the other cleavage reactions (AtCathB2: 22%; AtCathB3: 56%) corresponded to the release of dipeptides from the C-termini of substrates, thus demonstrating that both AtCathB2 and AtCathB3 have the capacity to act as a genuine carboxydipeptidase. Substantial loss of C-terminal tetra- and hexapeptides (AtCathB2:
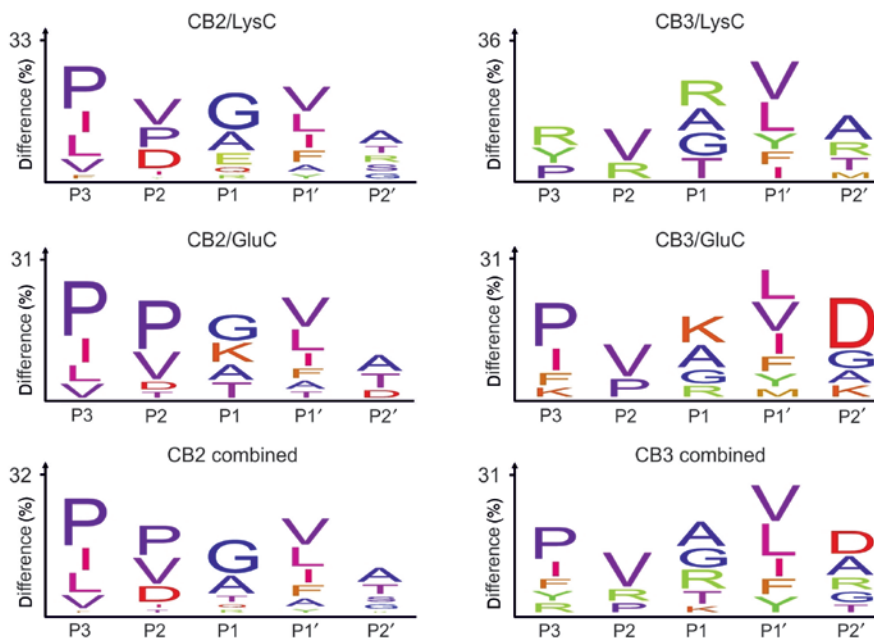
Figure 5: Specificity profiling of AtCathB2 and AtCathB3.
Escherichia coli-derived peptide libraries prepared with LysC or GluC were incubated with wild-type AtCathB2 (CB2) or AtCathB3 (CB3) at protease-to-library ratios of 1:10 for 16 h at 23°C and then subjected to PICS analysis. Semi-specific peptides enriched more than four-fold were used for reconstruction of the cleavage sites. Sequences with fewer than three prime-site residues were removed prior to presentation of the data as IceLogos. Amino acid enrichment at positions on either side of the cleaved peptide bond is expressed as percent difference to the natural abundance after normalization against the Swiss-Prot protein database for *E. coli* (strain K-12). Amino acids (abbreviated using their one-letter codes) are only shown when their enrichment was statistically significant ($p < 0.05$). Combined analyses of all observed cleavage events are also presented. P, substrate position (non-prime site); P′, substrate position (prime site).

21%; AtCathB3: 18%) was also noted, with some of these degradation events clearly attributable to progressive carboxydipeptidase action. However, neither enzyme displayed appreciable carboxymonopeptidase activity in our PICS assays.

Consistent with their similar catalytic efficiencies towards many synthetic substrates, the PICS profiles of AtCathB2 and AtCathB3 show clear signs of close resemblance. Nevertheless, these analyses revealed a few disparate properties (Figure 5). As previously observed for NbCathB (Paireder et al., 2017), the P1 preferences of AtCathB2 and AtCathB3 are far more restricted than those of other PLCPs. Only basic amino acids and small residues such as glycine, alanine or threonine were consistently found to be enriched at P1, although AtCathB2 accepts also glutamic acid and glutamine to some extent. Some striking differences were noted with respect to the P2 specificities of AtCathB2 and AtCathB3. The latter enzyme favors valine, arginine and proline at this position, which agrees well with the observation that Z-Arg-Arg-MCA and Z-Gly-Pro-Arg-MCA are suitable AtCathB3 substrates. AtCathB2 also exhibits some selectivity for proline and valine in P2. However, the tolerance of AtCathB3 for substrates with P2 arginine

residues is not shared by AtCathB2, which on the other hand, readily accepts substrates displaying aspartic acid at this position. Hence, AtCathB2 could exceed AtCathB3 in its sensitivity to JOGDA1, a cathepsin B probe based on the tripeptide Gly-Asp-Ala (Lu et al., 2015). The S3 subsites of the two enzymes interact preferentially with proline, aliphatic or aromatic residues. In addition, some enrichment of lysine and arginine was noted for AtCathB3. The latter could be related to interactions with Asp[169] and Asp[172] (AtCathB2: Gln[172] and Asn[175]), two residues predicted to contribute to the S3 pocket of cathepsin B (Jia et al., 1995). In P1′, AtCathB2 and AtCathB3 select for aliphatic and aromatic residues like their human and *N. benthamiana* orthologues (Biniossek et al., 2011; Paireder et al., 2017). Interestingly, our data indicate that AtCathB2 and AtCathB3 display subtle differences with respect to the properties of their S2′ pockets. While AtCathB2 has a preference for small uncharged amino acids at P2′, AtCathB3 also readily cleaves peptides containing aspartic acid or arginine at this position. In summary, the subsite specificity profiles of AtCathB2 and AtCathB3 indicate that these enzymes share many features besides displaying some remarkable intricacies that could relate to synergistic functions *in vivo*.

Discussion

This study aimed at a detailed characterization of the proteins encoded by the *AtCathB1*, *AtCathB2* and *AtCathB3* genes and a comparison of their enzymatic properties. While AtCathB2 and AtCathB3 were shown to represent effective proteases, AtCathB1 appears to lack hydrolytic activity. The predicted close homology between AtCathB2 and AtCathB3 is supported by their similar reactivity with several generic cathepsin B substrates. Z-Phe-Arg-MCA and Z-Leu-Arg-MCA were the best dipeptide-based substrates identified for either enzyme. The most remarkable difference between AtCathB2 and AtCathB3 relates to the hydrolysis of Z-Arg-Arg-MCA. This compound is readily cleaved by AtCathB3 (Ge et al., 2016), as also reported for its orthologue from *R. sativus* (Tsuji et al., 2008). However, Z-Arg-Arg-MCA is a poor substrate for AtCathB2. Site-directed mutagenesis studies revealed that the absence of a compensatory acidic group at the base of the S2 subsite probably accounts for the inefficiency of AtCathB2 in hydrolyzing dibasic substrates.

A degradomics approach based on proteome-derived peptide libraries (Biniossek et al., 2016) was chosen for a comprehensive assessment of the substrate specificities of AtCathB2 and AtCathB3. Recently, this technique has been used to characterize the substrate preferences of different PLCPs from *N. benthamiana*, including NbCathB (Paireder et al., 2016, 2017). In accordance with other activity-based profiling approaches (Richau et al., 2012), these studies have clearly shown that the substrate preferences of PLCPs are largely dictated by their P2 specificities. Our data indicate that this also applies to AtCathB2 and AtCathB3. However, AtCathB3 is superior to AtCathB2 in hydrolysing Z-Gly-Pro-Arg-MCA and Z-His-Glu-Lys-MCA, which is not explicable by the P2 preferences of the enzymes alone. In the case of Z-Gly-Pro-Arg-MCA, the slow action of AtCathB2 on this substrate could be due to the observed dislike of the enzyme for glycine residues in P3. Such a negative cooperativity could also account for the poor activity of AtCathB2 towards Z-His-Glu-Lys-MCA.

In proteome-derived peptide libraries, the substrate preferences of AtCathB2 and AtCathB3 are largely reminiscent of those of human cathepsin B (Biniossek et al., 2011). Nevertheless, a few notable differences exist between these enzymes. For example, AtCathB2 and AtCathB3 show some variations from human cathepsin B with respect to their P2′ selectivities. Both plant enzymes prefer small uncharged amino acids at this position, whereas the S2′ subsite of human cathepsin B accommodates bulky amino acids such as tryptophan, tyrosine and phenylalanine particularly well (Portaro et al., 2000; Krupa et al.,

2002). These specific properties could relate to the shorter occluding loop typical for all plant forms of cathepsin B as compared to their mammalian orthologues.

Although AtCathB2 shares the carboxydipeptidase activity of human cathepsin B, mutations of its evolutionarily conserved residues Asp^{127} and His^{207} affect the enzyme in a strikingly different manner. Their counterparts in the human protease, Asp^{22} and His^{110}, form a salt bridge which docks the occluding loop to the catalytic domain. Substitution of these two residues with alanine eliminates the capacity of human cathepsin B to cleave the exopeptidase substrate Abz-FRF(NO$_2$)A almost completely, while hydrolysis of the generic cysteine protease substrate Z-Phe-Arg-MCA remains essentially unchanged (Krupa et al., 2002). By contrast, replacement of His^{207} with alanine reduces the activity of AtCathB2 towards Abz-FRF(NO$_2$)A and Z-Phe-Arg-MCA to a similar extent, thus establishing that the absence of the Asp^{127}-His^{207} ion-pair compromises the active-site cleft of this enzyme beyond the S2′ subsite. Remarkably, the $Asp^{127} \rightarrow Ala$ and $Asp^{127} \rightarrow Ala/His^{207} \rightarrow Ala$ mutants could not be produced due to interference with proper protein folding, indicating that Asp^{127} is even more important than His^{207} for AtCathB2 to adopt a catalytically competent state. It is noteworthy that a second salt bridge, Arg^{116}-Asp^{224}, exists between the occluding loop and the catalytic domain of human cathepsin B (Nägler et al., 1997). The latter ion-pair is missing in plant cathepsin B, which could explain the dependence of AtCathB2 on Asp^{127} to achieve its correct conformation.

In conclusion, AtCathB2 and AtCathB3 are two closely related PLCPs which display subtle differences in their hydrolytic properties and substrate specificities. Both enzymes can act as endopeptidases and carboxydipeptidases, providing unequivocal evidence that this unique dual catalytic character is shared by animal and plant cathepsin B-like proteases.

Materials and methods

RT-PCR analysis of *AtCathB1* transcripts

Arabidopsis thaliana ecotype Columbia (Col-0) plants were grown in long-day conditions (16-h light/8-h dark photoperiod) on soil at 22°C. RNA was extracted from 35-mg samples of rosette leaves from 4-week-old plants using the SV Total RNA Isolation Kit (Promega, Madison, WI, USA) following the manufacturer's protocol. First-strand cDNA was synthesized from 1 µg RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific, Waltham, MA, USA) and oligo(dT)$_{18}$ as primer. A diagnostic fragment of the *AtCathB1* (At1g02300) coding region covering exons 2–6 was amplified with the primer combination AtCathB1_fw/AtCathB1_rev (Sigma-Aldrich,

St. Louis, MO, USA; Table S4) and GoTaq G2 DNA polymerase (Promega). PCR products were separated by 1.5% agarose gel electrophoresis, visualized by ethidium bromide staining and then analyzed by automated DNA sequencing (Microsynth, Balgach, Switzerland).

Molecular cloning and site-directed mutagenesis of *AtCathB2* and *AtCathB3* cDNAs

Plasmids containing the complete open reading frames of *AtCathB2* (At1g02305) and *AtCathB3* (At4g01610) were obtained from RIKEN BRC (Tsukuba, Japan). The regions encoding pro-AtCathB2 (amino acids 34-362, preproprotein numbering) and pro-AtCathB3 (residues 31-359) were then amplified by PCR using the primer combinations AtCathB2_fw/AtCathB2_rev and AtCathB3_fw/AtCathB3_rev (Sigma-Aldrich), respectively (Table S4). The obtained PCR products were cleaved with *Xba*I and *Kpn*I (Promega) and ligated into pBluescript II SK(+) (Agilent Technologies, La Jolla, San Diego, CA, USA) cut with the same enzymes. Point mutations were introduced by the QuikChange (Agilent Technologies) site-directed mutagenesis procedure using KapaHiFi DNA polymerase (Sigma-Aldrich) and a suitable complementary primer pair (Table S4). The fidelity of the wild-type and mutant cDNAs was verified by DNA sequencing (Microsynth).

Heterologous expression of AtCathB2 and AtCathB3 in insect cells

Wild-type and mutant pro-AtCathB sequences were isolated by digestion of the respective plasmids with *Xba*I and *Kpn*I and ligated into pVTBacHis-FLAG (Kowalewski-Nimmerfall et al., 2014) digested with the same enzymes. In the resulting pVTBacHis-FLAG-AtCathB constructs, the pro-AtCathB sequences are positioned in-frame behind a leader segment consisting of the melittin signal peptide followed by a 4-kDa linker region containing six consecutive histidine residues and the FLAG epitope. Heterologous production of wild-type and mutant pro-AtCathB in insect cells was then conducted according to previously published procedures (Schähs et al., 2008; Kowalewski-Nimmerfall et al., 2014). Briefly, recombinant baculoviruses were generated by co-transfection of *Spodoptera frugiperda* Sf9 cells with the respective pVTBacHis-FLAG construct and baculoviral DNA. The supernatant containing recombinant baculovirus was then used for infection of *S. frugiperda* Sf21 cells. After 4 days at 27°C, cells and conditioned media were harvested and subjected to immunoblotting analysis after direct lysis of the cells in SDS-PAGE sample buffer (Bencur et al., 2005).

Purification of recombinant AtCathB2 and AtCathB3 by two-step affinity chromatography

Culture supernatant of baculovirus-infected Sf21 cells (200 ml) was cleared by centrifugation and dialyzed twice against 2 l of phosphate buffer (10 mм sodium phosphate, pH 7.0, 40 mм NaCl, 0.02% NaN$_3$). After addition of 20 mм imidazole, the sample was loaded on a 5 ml column of Chelating Sepharose (GE Healthcare, Little Chalfont, UK) charged with Ni^{2+} ions, equilibrated in the same buffer. After successive washes with 40–80 mм imidazole, the recombinant protein was eluted with 250 mм imidazole in phosphate buffer.

Protein-containing eluate fractions were combined, dialyzed twice against 1 l of phosphate-buffered saline containing 0.02% (w/v) NaN$_3$ (PBSA) and then loaded on a 1 ml anti-FLAG M2 affinity column (Sigma-Aldrich). The column was washed with 10 ml PBSA prior to elution with 5 ml 100 µg/ml FLAG peptide (Sigma-Aldrich) in the same buffer. The AtCathB-containing eluate fractions were pooled, dialyzed twice against 1 l of PBSA and then concentrated by ultrafiltration. The protein content of the purified AtCathB samples was determined with the BCA Protein Assay Kit (Pierce, Rockford, IL, USA), using bovine serum albumin as standard.

Activity assays with synthetic substrates

For assays with synthetic substrates, pro-AtCathB (0.1–1.0 mg/ml) was activated by treatment with 50 µg/ml pepsin (Sigma-Aldrich) in 100 mм sodium acetate (pH 4.5), 2.5 mм DTT for 60 min at 37°C. Pepsin was then quantitatively inhibited by the addition of 10 µм pepstatin A (Sigma-Aldrich). The concentration of activated AtCathB was determined by active-site titration with E-64 (Sigma-Aldrich) as described previously (Niemer et al., 2016).

Stopped assays were performed at 37°C in 100 mм sodium acetate (pH 5.5) supplemented with 1 mм DTT and 1% dimethylsulfoxide as outlined before (Niemer et al., 2016), using 10 µм of the respective peptidyl-MCA substrate (Bachem, Bubendorf, Switzerland; PeptaNova, Sandhausen, Germany). The reactions were terminated by the addition of an equal volume of 100 mм monochloroacetic acid, 100 mм sodium acetate (pH 4.3) prior to analysis by spectrofluorimetry. Assay linearity was confirmed by variation of incubation time and amount of enzyme added.

Kinetic measurements with Z-Phe-Arg-MCA or Z-Arg-Arg-MCA as substrate were performed at 25°C in the same buffer as above, following previously published procedures (Novinec et al., 2012; Niemer et al., 2016). Briefly, the increase in fluorescence was continuously monitored for up to 10 min using an LS50B Luminescence Spectrometer (PerkinElmer, Waltham, MA, USA) with excitation and emission wavelengths set at 383 nm and 455 nm, respectively. AtCathB variants were also analyzed for hydrolysis of Abz-Phe-Arg-Phe(NO$_2$)-Ala (Krupa et al., 2002; kindly provided by John S. Mort, Shriners Hospital for Children, Montreal, Canada) or Abz-Gly-Ile-Val-Arg-Ala-Lys(Dnp) (Cotrin et al., 2004; Bachem). Hydrolysis of the internally quenched fluorescence substrates was followed fluorimetrically at 320/420 nm. Reaction rates were determined from the respective progress curves. The kinetic parameters K_m and k_{cat} were derived by non-linear regression analysis using the Henri-Michaelis-Menten equation and Graph-Pad Prism 5.0 software.

SDS-PAGE and Western blotting analysis

Samples were denatured for 5 min at 95°C under reducing conditions and then subjected to 12.5% SDS-PAGE. Separated polypeptides were then either subjected to Coomassie Brilliant Blue R-250 staining or electrophoretically transferred onto nitrocellulose membranes (GE Healthcare). After probing the membranes with monoclonal mouse anti-FLAG M2 (Sigma-Aldrich), bound immunoglobulins were visualized with peroxidase-conjugated goat anti-mouse IgG antibodies (Jackson ImmunoResearch, West Grove, PA, USA) and enhanced chemiluminescence reagents (Bio-Rad, Richmond, VA, USA).

Specificity profiling with proteome-derived peptide libraries

Comparative specificity profiling of AtCathB2 and AtCathB3 was performed using a modified PICS procedure as reported previously (Biniossek et al., 2016; Paireder et al., 2017). Briefly, proteome-derived peptide libraries were prepared by digestion of *E. coli* lysates with GluC or LysC (sequencing grade; Sigma-Aldrich). These peptide samples (50 µg) were incubated with wild-type AtCathB2 or AtCathB3 (5 µg) in 100 mM sodium acetate (pH 5.5) containing 1 mM DTT and 10 µg/ml dextran sulfate (Sigma-Aldrich) for 16 h at 23°C. Protease-treated and control samples differentially labeled with either natural ($^{12}COH_2$) or deuterated formaldehyde ($^{13}COD_2$) were then combined and analyzed by liquid chromatography-tandem mass spectrometry. Semi-specific peptides that were enriched at least four-fold in the protease-treated samples were considered as cleavage products. The corresponding prime and non-prime sequences were reconstructed bioinformatically through database searches. Sequences with a P1 or P2′ amino acid matching the specificity of the protease used for library generation were excluded prior to cleavage site analysis via iceLogo (Colaert et al., 2009).

Mass spectrometry

Wild-type AtCathB2 (1.5 mg/ml) or AtCathB3 (1.0 mg/ml) was fractionated on a Thermo ProSwift RP-4H column (250×0.20 mm) using a Dionex UltiMate 3000 HPLC system (Thermo Scientific). After application of the sample (5 µl), elution was performed at 80°C and a flow rate of 8 µl/min with a gradient of 20–95% solvent B (80% acetonitrile in 0.01% trifluoroacetic acid) in solvent A (0.05% trifluoroacetic acid) over 40 min as follows: 20–65% B (20 min), 65–95% B (20 min). Eluted polypeptides were analyzed online on a maXis 4G ETD Q-TOF mass spectrometer (Bruker, Billerica, MA, USA) equipped with an electrospray ionization source and operated in the positive ion mode (m/z range: 400–3800) as described previously (Paireder et al., 2016). The m/z values of the most prominent charge states were used to deduce the molecular masses of recombinant AtCathB2 and AtCathB3.

The N-glycosylation site occupancy of AtCathB2 was determined by liquid chromatography-mass spectrometry as described recently (Loos et al., 2015; Maresch and Altmann, 2016). Briefly, samples were separated by SDS-PAGE, stained with Coomassie Brilliant Blue and then excised from the gel. After S-alkylation with iodoacetamide and tryptic digestion, peptides were eluted with 50% acetonitrile and separated on a reversed phase column (150×0.32 mm BioBasic-18, Thermo Scientific) with a gradient of 1–80% acetonitrile followed by on-line analysis with a Q-TOF Ultima Global mass spectrometer (Waters, Milford, CT, USA).

A differential S-alkylation method was used for the detection of AtCathB cysteines not engaged in disulfide bonds. Briefly, AtCathB2 or AtCathB3 (20 µg) were treated for 1 h at 37°C with the first alkylation reagent (30 mM N-ethylmaleimide) in the presence of 6 M urea. The samples were then precipitated with 80% (v/v) acetone prior to reduction with 10 mM DTT (1 h, 56°C) and incubation with the second alkylation reagent (55 mM iodoacetamide) for 45 min at 23°C. After digestion with trypsin (AtCathB2) or trypsin/GluC (AtCathB3), peptides were analyzed by mass spectrometry as outlined above.

Molecular modeling and simulation

The structure of rat cathepsin B was taken from the Protein Data Bank (code 1CPJ; Jia et al., 1995) and used as a template to generate a model for AtCathB2 by homology modeling using SWISS-MODEL (Biasini et al., 2014) and imposing the sequence alignment shown in Figure S1. The position and flexibility of the occluding loop was then studied by MD simulations (2×5 ns) using the GROMOS11 simulation package (Schmid et al., 2012) and parameter set 54A8 of the GROMOS force field (Reif et al., 2012a), as outlined in detail earlier (Reif et al., 2012b). In both the rat and plant protein, the active-site cysteine (Cys[134], AtCathB2 preproprotein numbering) was simulated in a deprotonated form, key histidine residues (His[207] and His[289]) in a doubly protonated form, while the protonation forms of the remaining residues were chosen according to their charge states at pH 7. Final configurations of MD simulations were used for structure visualization with PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

Database accession number: Proteomic data have been deposited in ProteomeXchange under the identifier PXD008202.

References

Bencur, P., Steinkellner, H., Svoboda, B., Mucha, J., Strasser, R., Kolarich, D., Hann, S., Kollensperger, G., Glossl, J., Altmann, F., et al. (2005). *Arabidopsis thaliana* β1,2-xylosyltransferase: an unusual glycosyltransferase with the potential to act at multiple stages of the plant N-glycosylation pathway. Biochem. J. *388*, 515–525.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and

quaternary structure using evolutionary information. Nucleic Acids Res. *42*, W252–W258.

Biniossek, M.L., Nagler, D.K., Becker-Pauly, C., and Schilling, O. (2011). Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. J. Proteome Res. *10*, 5363–5373.

Biniossek, M.L., Niemer, M., Maksimchuk, K., Mayer, B., Fuchs, J., Huesgen, P.F., McCafferty, D.G., Turk, B., Fritz, G., Mayer, J., et al. (2016). Identification of protease specificity by combining proteome-derived peptide libraries and quantitative proteomics. Mol. Cell. Proteomics *15*, 2515–2524.

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

Choe, Y., Leonetti, F., Greenbaum, D.C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J.A., and Craik, C.S. (2006). Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities. J. Biol. Chem. *281*, 12824–12832.

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.

Cotrin, S.S., Puzer, L., de Souza Judice, W.A., Juliano, L., Carmona, A.K., and Juliano, M.A. (2004). Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: assays with human cathepsin B. Anal. Biochem. *335*, 244–252.

Ge, Y., Cai, Y.M., Bonneau, L., Rotari, V., Danon, A., McKenzie, E.A., McLellan, H., Mach, L., and Gallois, P. (2016). Inhibition of cathepsin B by caspase-3 inhibitors blocks programmed cell death in *Arabidopsis*. Cell Death Differ. *23*, 1493–1501.

Gilroy, E.M., Hein, I., van der Hoorn, R., Boevink, P.C., Venter, E., McLellan, H., Kaffarnik, F., Hrubikova, K., Shaw, J., Holeva, M., et al. (2007). Involvement of cathepsin B in the plant disease resistance hypersensitive response. Plant J. *52*, 1–13.

Hasnain, S., Hirama, T., Huber, C.P., Mason, P., and Mort, J.S. (1993). Characterization of cathepsin B specificity by site-directed mutagenesis. Importance of Glu245 in the S2-P2 specificity for arginine and its role in transition state stabilization. J. Biol. Chem. *268*, 235–240.

Iglesias-Fernandez, R., Wozny, D., Iriondo-de Hond, M., Onate-Sanchez, L., Carbonero, P., and Barrero-Sicilia, C. (2014). The AtCathB3 gene, encoding a cathepsin B-like protease, is expressed during germination of *Arabidopsis thaliana* and transcriptionally repressed by the basic leucine zipper protein GBF1. J. Exp. Bot. *65*, 2009–2021.

Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J.S. (1997). Role of the occluding loop in cathepsin B activity. J. Biol. Chem. *272*, 1197–1202.

Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J.S., To, R., and Huber, C.P. (1995). Crystal structures of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. Implications for structure-based inhibitor design. J. Biol. Chem. *270*, 5527–5533.

Kowalewski-Nimmerfall, E., Schahs, P., Maresch, D., Rendic, D., Kramer, H., and Mach, L. (2014). *Drosophila melanogaster* cellular repressor of E1A-stimulated genes is a lysosomal

protein essential for fly development. Biochim. Biophys. Acta *1843*, 2900–2912.

Krupa, J.C., Hasnain, S., Nagler, D.K., Menard, R., and Mort, J.S. (2002). S2′ substrate specificity and the role of His110 and His111 in the exopeptidase activity of human cathepsin B. Biochem. J. *361*, 613–619.

Loos, A., Gach, J.S., Hackl, T., Maresch, D., Henkel, T., Porodko, A., Bui-Minh, D., Sommeregger, W., Wozniak-Knopp, G., Forthal, D.N., et al. (2015). Glycan modulation and sulfoengineering of anti-HIV-1 monoclonal antibody PG9 in plants. Proc. Natl. Acad. Sci. USA *112*, 12675–12680.

Lu, H., Chandrasekar, B., Oeljeklaus, J., Misas-Villamil, J.C., Wang, Z., Shindo, T., Bogyo, M., Kaiser, M., and van der Hoorn, R.A. (2015). Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination. Plant Physiol. *168*, 1462–1475.

Maresch, D. and Altmann, F. (2016). Isotype-specific glycosylation analysis of mouse IgG by LC-MS. Proteomics *16*, 1321–1330.

McLellan, H., Gilroy, E.M., Yun, B.W., Birch, P.R., and Loake, G.J. (2009). Functional redundancy in the *Arabidopsis* cathepsin B gene family contributes to basal defence, the hypersensitive response and senescence. New Phytol. *183*, 408–418.

Mort, J.S. and Buttle, D.J. (1997). Cathepsin B. Int. J. Biochem. Cell Biol. *29*, 715–720.

Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., et al. (1991). The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J. *10*, 2321–2330.

Nägler, D.K., Storer, A.C., Portaro, F.C., Carmona, E., Juliano, L., and Menard, R. (1997). Major increase in endopeptidase activity of human cathepsin B upon removal of occluding loop contacts. Biochemistry *36*, 12608–12615.

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

Novinec, M., Pavsic, M., and Lenarcic, B. (2012). A simple and efficient protocol for the production of recombinant cathepsin V and other cysteine cathepsins in soluble form in *Escherichia coli*. Protein Expr. Purif. *82*, 1–5.

Paireder, M., Mehofer, U., Tholen, S., Porodko, A., Schahs, P., Maresch, D., Biniossek, M.L., van der Hoorn, R.A., Lenarcic, B., Novinec, M., et al. (2016). The death enzyme CP14 is a unique papain-like cysteine proteinase with a pronounced S2 subsite selectivity. Arch. Biochem. Biophys. *603*, 110–117.

Paireder, M., Tholen, S., Porodko, A., Biniossek, M.L., Mayer, B., Novinec, M., Schilling, O., and Mach, L. (2017). The papain-like cysteine proteinases NbCysP6 and NbCysP7 are highly processive enzymes with substrate specificities complementary to *Nicotiana benthamiana* cathepsin B. Biochim. Biophys. Acta *1865*, 444–452.

Portaro, F.C., Santos, A.B., Cezari, M.H., Juliano, M.A., Juliano, L., and Carmona, E. (2000). Probing the specificity of cysteine proteinases at subsites remote from the active site: analysis of P4, P3, P2′ and P3′ variations in extended substrates. Biochem. J. *347*, 123–129.

Reif, M.M., Hunenberger, P.H., and Oostenbrink, C. (2012a). New interaction parameters for charged amino acid side chains in

the GROMOS force field. J. Chem. Theory Comput. *8*, 3705–3723.

Reif, M.M., Mach, L., and Oostenbrink, C. (2012b). Molecular insight into propeptide-protein interactions in cathepsins L and O. Biochemistry *51*, 8636–8653.

Richau, K.H., Kaschani, F., Verdoes, M., Pansuriya, T.C., Niessen, S., Stuber, K., Colby, T., Overkleeft, H.S., Bogyo, M., and van der Hoorn, R.A. (2012). Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. Plant Physiol. *158*, 1583–1599.

Rowan, A.D., Feng, R., Konishi, Y., and Mort, J.S. (1993). Demonstration by electrospray mass spectrometry that the peptidyldi-peptidase activity of cathepsin B is capable of rat cathepsin B C-terminal processing. Biochem. J. *294*, 923–927.

Schähs, P., Weidinger, P., Probst, O.C., Svoboda, B., Stadlmann, J., Beug, H., Waerner, T., and Mach, L. (2008). Cellular repressor of E1A-stimulated genes is a bona fide lysosomal protein which undergoes proteolytic maturation during its biosynthesis. Exp. Cell Res. *314*, 3036–3047.

Schilling, O., Huesgen, P.F., Barre, O., auf dem Keller, U., and Overall, C.M. (2011). Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. Nat. Protoc. *6*, 111–120.

Schmid, N., Christ, C.D., Christen, M., Eichenberger, A.P., and van Gunsteren, W.F. (2012). Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation. Comput. Phys. Commun. *183*, 890–903.

Tsuji, A., Kikuchi, Y., Ogawa, K., Saika, H., Yuasa, K., and Nagahama, M. (2008). Purification and characterization of cathepsin B-like cysteine protease from cotyledons of daikon radish, *Raphanus sativus*. FEBS J. *275*, 5429–5443.

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012). Cysteine cathepsins: from structure, function and regulation to new frontiers. Biochim. Biophys. Acta *1824*, 68–88.