The death enzyme CP14 is a unique papain-like cysteine proteinase with a pronounced S2 subsite selectivity

Melanie Paireder a, Ulrich Mehofer a, Stefan Tholen b, Andreas Porodko a, Philipp Schähs a, Daniel Maresch c, Martin L. Binossek b, Renier A.L. van der Hoorn d, Brigita Lenarčič e, Marko Novince e, Oliver Schilling b,f, Lukas Mach a,e.*

a Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria
b Institute of Molecular Medicine and Cell Research, University of Freiburg, Germany
c Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria
d The Plant Chemetics Laboratory, Department of Plant Sciences, University of Oxford, United Kingdom
e BISS Centre for Biological Signaling Studies, University of Freiburg, Germany
f Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria

Abstract

The cysteine protease CP14 has been identified as a central component of a molecular module regulating programmed cell death in plant embryos. CP14 belongs to a distinct subfamily of papain-like cysteine proteinases of which no representative has been characterized thoroughly to date. However, it has been proposed that CP14 is a cathepsin H-like protease. We have now produced recombinant Nicotiana benthamiana CP14 (NbCP14) lacking the C-terminal granulin domain. As typical for papain-like cysteine proteinases, NbCP14 undergoes rapid autocatalytic activation when incubated at low pH. The mature protease is capable of hydrolysing several synthetic endopeptidase substrates, but cathepsin H-like aminopeptidase activity could not be detected. NbCP14 displays a strong preference for aliphatic over aromatic amino acids in the specificity-determining P2 position. This subsite selectivity was also observed upon digestion of proteome-derived peptide libraries. Notably, the specificity profile of NbCP14 differs from that of aleurain-like protease, the N. benthamiana orthologue of cathepsin H. We conclude that CP14 is a papain-like cysteine proteinase with unusual enzymatic properties which may prove of central importance for the execution of programmed cell death during plant development.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The tobacco-related plant species Nicotiana benthamiana is a popular model organism for the study of plant-pathogen interactions [1]. Substantial evidence has been obtained for a molecular arms race between plant-derived papain-like cysteine proteinases (PLCPs) and antagonistic effectors released by pathogens at the site of infection [2]. Intriguingly, a delicate balance between endogenous PLCPs and cysteine proteinase inhibitors also dictates proper timing of programmed cell death in tobacco embryos [3]. Hence, N. benthamiana is a suitable host for further studies on the diverse physiological functions of PLCPs. Furthermore, N. benthamiana is increasingly used as expression platform for the production of biotherapeutics [4]. In particular, N. benthamiana lines have been established which permit the production of recombinant proteins such as monoclonal antibodies (mAbs) with customized post-translational modifications and thus superior biological activities [5,6]. However, a frequent challenge encountered during the production of mAbs in plants is their proteolytic degradation [7–10]. Recent studies have highlighted that mAb fragmentation in N. benthamiana and tobacco involves serine and cysteine proteases including PLCPs [11–14].

Considerable progress is currently made in the biochemical and
structural characterization of PLCPs and their animal counterparts, the cysteine cathepsins [15,16]. Based on a comprehensive phylogenetic analysis, PLCPs have been recently divided into 9 subfamilies. The first six subfamilies comprise cathepsin L-like proteases differing in individual molecular aspects, with two of them (subfamilies 1 and 4) featuring C-terminal extensions. Subfamilies 7–9 are phylogenetically distinct and display homologies to mammalian cathepsins F, H and B, respectively [17]. The development of sophisticated activity-based probes has allowed the selective identification of several PLCPs in different plant species including N. benthamiana [18]. However, only two N. benthamiana PLCPs have been studied thoroughly on the protein level so far: aleurain-like protease (ALP), assigned to subfamily 8, and cathepsin B, a member of subfamily 9 [19–21]. Another N. benthamiana PLCP has recently received substantial attention. This enzyme promotes plant immunity in a similar way as tomato C14 (subfamily 1) and was thus initially named NbC14 [22,23]. However, it later became evident that NbC14 belongs to subfamily 4 [17] and is the orthologue of NtC14, a protease involved in the spatiotemporal control of programmed cell death [3]. Based on specificity profiling with a small number of synthetic PLCP substrates, it was concluded that the enzymatic properties of NtC14 resemble those of mammalian cathepsin H [3].

To study the catalytic features of N. benthamiana NbC14 (hereafter referred to as NbCP14) in depth, we have now produced recombinant versions of this protease in different heterologous expression systems. This enabled us to conduct a detailed characterization of its enzymatic properties using synthetic substrates as well as proteome-derived peptide libraries. Our results demonstrate that CP14 lacks cathepsin H-like aminopeptidase activity and exhibits a specificity profile distinct of ALP, the plant orthologue of cathepsin H.

2. Materials and methods

2.1. Heterologous expression of NbCP14 in insect cells

The plasmid pTP11 containing the complete open reading frame of NbCP14 [22] was used as a PCR template. Two different constructs were generated: pVTBacHis–FLAG–NbCP14 encodes a truncated version of the proenzyme lacking the C-terminal granulin domain (NbCP14 residues 30–397, preproprotein numbering), whereas both the granulin domain and the proline-rich linker region were removed in the case of pVTBacHis–FLAG–NtC14P (NbCP14 residues 30–367). The same forward primer (5′-AAATC-TAGATTACTAGTGTATTTTA-3′) was used in conjunction with different reverse primers, 5′-TCTGTACCTCAACTTGGTTTGGG-GAAAAG-3′ (NbCP14) or 5′-TCTGTACCTCAACTTGGTTTGGG-GAAAAG-3′ (NtC14P). The PCR products were cleaved with XbaI and KpnI (Fermentas, St. Leon-Rot, Germany) at the underlined sites and ligated into pVTBacHis–FLAG (Fermentas, St. Leon-Rot, Germany) at the underlined sites and ligated into the corresponding sites of pET-32/28 prior to expression in the E. coli strain Rosetta-gami B (DE3) pLysS as described previously [21,26]. The resulting cell pellet was resuspended in 20 mM Hepes (pH 7.4) supplemented with 500 mM NaCl and 20 mM imidazole. After cell lysis by sonication, debris was removed by centrifugation at 16,000g for 15 min at 4 °C. The supernatant was then passed through a column of Chelating Sepharose charged with Ni2+ ions and equilibrated in resuspension buffer. After successive washes with 40, 60 and 80 mM imidazole, recombinant pro-NbCP14 was eluted with 250 mM imidazole. Protein-containing eluate fractions were pooled, dialysed against phosphate-buffered saline (20 mM sodium phosphate, pH 7.4, 150 mM NaCl) containing 0.02% (w/v) NaN3 and then concentrated by ultrafiltration.

2.2. Production of NbCP14 in Escherichia coli

The sequence encoding NbCP14 residues 30–397 was amplified by PCR with the primer combination 5′-GCACGCTGCTTACTACTGATTCCCAATACTA-3′ (forward) and 5′-GCACGCTGCTTACTACTGATTCCCAATACTA-3′ (reverse). The PCR product was then cleaved with Nhel and Xhol at the underlined positions and ligated into the corresponding sites of pET-32/28 prior to expression in the E. coli strain Rosetta-gami B (DE3) pLysS as described previously [21,26]. The resulting cell pellet was resuspended in 20 mM Hepes (pH 7.4) supplemented with 500 mM NaCl and 20 mM imidazole. After cell lysis by sonication, debris was removed by centrifugation at 16,000g for 15 min at 4 °C. The supernatant was then passed through a column of Chelating Sepharose charged with Ni2+ ions and equilibrated in resuspension buffer. After successive washes with 40, 60 and 80 mM imidazole, recombinant pro-NbCP14 was eluted with 250 mM imidazole. Protein-containing eluate fractions were pooled, dialysed against phosphate-buffered saline (20 mM sodium phosphate, pH 7.4, 150 mM NaCl) containing 0.02% (w/v) NaN3 and then concentrated by ultrafiltration.

2.3. Activity assays

Pro-NbCP14 (0.3 mg/ml) was autocatalytically activated by incubation in 0.1 M sodium acetate (pH 5.0) containing 2.5 mM DTT for 30 min at 37 °C. Stopped assays were performed at 37 °C as outlined previously [27,28], using 10 μM of the respective peptidyl-MCA substrate (Bachem, Bubendorf, Switzerland; PeptaNova, Sandhausen, Germany) in 0.1 M sodium acetate (pH 5.0), 1% dimethylsulfoxide, 5 mM DTT. The reactions were terminated by addition of an equal volume of 0.1 M monochloracetic acid, 0.1 M sodium acetate (pH 4.3) prior to analysis by spectrofluorimetry. Active-site titration was done by preincubation with varying concentrations of the irreversible inhibitor E-64 for 60 min at 0 °C [29]. Kinetic measurements with Z-Leu-Arg-MCA or Z-Phe-Arg-MCA as substrate were performed at 25 °C following previously published procedures [26]. The assay buffer was the same as stated above. 5% dimethylsulfoxide was used to enhance the solubility of Z-Phe-Arg-MCA at high substrate concentrations. Reaction rates were determined from the respective progress curves. The kinetic parameters Km and kcat were derived by non-linear regression analysis using the Henri-Michaelis-Menten equation and GraphPad Prism 5.0 software.

2.4. Specificity profiling with proteome-derived peptide libraries

Specificity profiling of N. benthamiana aleurain-like protease (NbALP) and NbCP14 was performed using the Proteomic Identification of Protease Cleavage Sites (PICS) procedure [30–32] or an adaption thereof [33]. Briefly, proteome-derived peptide libraries were prepared by digestion of E. coli or HEK293 lysates with trypsin or GluC. These peptide samples (300 μg) were incubated with the enzyme to be tested (0.6–3 μg) in assay buffer (NbALP: 0.2 M Mes (pH 5.0), 5 mM DTT; NbCP14: 0.1 M sodium acetate (pH 5.0), 5 mM DTT) for up to 16 h at room temperature. Protease-treated and control samples were then differentially labelled by reductive dimethylation with either light (12COH2) or heavy formaldehyde (15COH2) and digested with strong-cation exchange chromatography prior to analysis by liquid chromatography-tandem mass spectrometry as described previously [34]. Spectrum to sequence assignment and relative peptide quantitation were performed as reported [35] with the following adaptations: asymmetric precursor mass error of ± 150 ppm/–0 ppm, fragment ion mass tolerance of 0.1 Da, semi-GluC or semi-tryptic specificity. Semi-specific peptides that were enriched >4-fold in the protease-
treated samples were considered as cleavage products. The corresponding prime and non-prime sequences were reconstructed bioinformatically through database searches and visualized using iceLogo [36].

2.5. Profiling with activity-based probes

Active-site labelling of NbCP14 was performed in 0.1 M sodium acetate (pH 5.5) supplemented with 5 mM DTT and varying concentrations of DCG-04 ([37]; Matthew Bogyo, Stanford University, USA), biotin-C6407 ([38]; Norbert Schaschke, University of Bielefeld, Germany) or biotin-Leu-Val-Gly-CHN2 ([39]; Gilles Lalmanach, University of Tours, France). After incubation for 30 min at 37 °C, the reactions were stopped by addition of 50 μg/ml E-64 and subsequent treatment for 5 min at 95 °C.

2.6. Digestion of mAbs

The human anti-HIV mAbs 2F5 and 2G12 (Dietmar Katinger, Polymun Scientific GmbH, Klosterneuburg, Austria) were tested for their susceptibility to NbCP14 as outlined earlier [12]. Briefly, 2F5 or 2G12 (200 μg/ml) were treated with NbCP14 (50 μg/ml) in 0.1 M sodium acetate (pH 5.5) containing 2 mM DTT at 37 °C. After incubation for up to 16 h, reactions were stopped by treatment for 5 min at 95 °C.

2.7. 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 silver staining or electrophoretically transferred onto Hybond-C nitrocellulose membranes (GE Healthcare). After probing the membranes with mouse IgG antibodies (Jackson ImmunoResearch, West Grove, USA) and the catalytic domains differ only at 11 positions (5%). Both enzymes belong to PLCP subfamily 4 [17] and are considerably larger than other PLCPs due to the additional presence of a C-terminal granulin domain (108 amino acids), which is connected to the other parts of the protease by a 30-residue proline-rich linker region. The latter segment is preceded by an N-terminal signal sequence (29 amino acids), a propeptide (116 residues) encompassing the ERFFIN motif typical for cathepsin L-like PLCPs [15,16], and the catalytic domain (222 amino acids) with the three active-site residues (Cys170, His308 and Asn328; Table S1).

Biochemical information on subfamily 4 enzymes is still sparse, but related PLCPs belonging to subfamily 1 have been characterized in more detail. Since the C-terminal granulin domain of subfamily 1 proteases is dispensable for proteolytic activity [20] but renders the enzymes prone to precipitation [40], we have first generated a truncated NbCP14 precursor (residues 30–397, preproprotein numbering) devoid of the granulin domain (pro-NbCP14) in order to improve the likelihood of proper secretion of the proenzyme into the culture medium. Although pro-NbCP14 could be expressed in insect cells, the amounts of recombinant protein accumulating in the culture supernatant were too low for further characterization. We have therefore generated a second deletion construct (residues 30–367) which additionally lacks the proline-rich region (pro-NbCP14ΔP). This protein was more efficiently secreted than pro-NbCP14 and could be purified to near homogeneity. When purified pro-NbCP14ΔP was subjected to SDS-PAGE analysis, a major 43-kDa band was observed. This apparent molecular mass is in reasonable agreement with its theoretical size (41.9 kDa). However, a 31-kDa polypeptide was also consistently present in purified pro-NbCP14ΔP samples (Fig. 1A). Incubation at low pH in the presence of a reducing agent resulted in the rapid conversion of pro-NbCP14ΔP into the faster migrating species, even in the absence of the processing enhancer dextran sulphate [21]. It is well established

2.8. Mass spectrometry

Purified pro-NbCP14 (0.5 mg/ml) was fractionated on a Thermo ProSwift RP-4H column (250 × 0.20 mm) using a Dionex UltiMate 3000 HPLC system (Thermo Scientific, Waltham, USA). 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 analysed online on a nanoES 4G ETD QTOF mass spectrometer (Bruker, Billerica, USA) equipped with an electrospray ionization source and operated in the positive ion mode (m/z range: 400–3800). The m/z values of the 8 most prominent charge states were used to deduce the molecular mass of recombinant pro-NbCP14.

2.9. Other methods

Recombinant pro-NbALP was produced and activated as described previously [21]. N-terminal sequence analysis of blotted bands was conducted at the Department of Molecular and Biomedical Sciences (Jožef Stefan Institute, Ljubljana, Slovenia) and

the Protein Micro-Analysis Facility (Medical University of Innsbruck, Austria) as outlined earlier [12]. Enzymatic deglycosylation of proteins was performed as reported [25]. Total protein content was determined with the BCA Protein Assay Kit (Pierce, Rockford, USA), using bovine serum albumin as standard.

3. Results

3.1. Heterologous expression of NbCP14

Full-length NbCP14 (505 amino acids; 56.9 kDa; GenBank: KU212214) shares 93% identity with NtCP14 (GenBank: KF113573), and the catalytic domains differ only at 11 positions (5%). Both enzymes belong to PLCP subfamily 4 [17] and are considerably larger than other PLCPs due to the additional presence of a C-terminal granulin domain (108 amino acids), which is connected to the other parts of the protease by a 30-residue proline-rich linker region. The latter segment is preceded by an N-terminal signal sequence (29 amino acids), a propeptide (116 residues) encompassing the ERFFIN motif typical for cathepsin L-like PLCPs [15,16], and the catalytic domain (222 amino acids) with the three active-site residues (Cys170, His308 and Asn328; Table S1).

Biochemical information on subfamily 4 enzymes is still sparse, but related PLCPs belonging to subfamily 1 have been characterized in more detail. Since the C-terminal granulin domain of subfamily 1 proteases is dispensable for proteolytic activity [20] but renders the enzymes prone to precipitation [40], we have first generated a truncated NbCP14 precursor (residues 30–397, preproprotein numbering) devoid of the granulin domain (pro-NbCP14) in order to improve the likelihood of proper secretion of the proenzyme into the culture medium. Although pro-NbCP14 could be expressed in insect cells, the amounts of recombinant protein accumulating in the culture supernatant were too low for further characterization. We have therefore generated a second deletion construct (residues 30–367) which additionally lacks the proline-rich region (pro-NbCP14ΔP). This protein was more efficiently secreted than pro-NbCP14 and could be purified to near homogeneity. When purified pro-NbCP14ΔP was subjected to SDS-PAGE analysis, a major 43-kDa band was observed. This apparent molecular mass is in reasonable agreement with its theoretical size (41.9 kDa). However, a 31-kDa polypeptide was also consistently present in purified pro-NbCP14ΔP samples (Fig. 1A). Incubation at low pH in the presence of a reducing agent resulted in the rapid conversion of pro-NbCP14ΔP into the faster migrating species, even in the absence of the processing enhancer dextran sulphate [21]. It is well established
that cysteine cathepsins and PLCPs can undergo autocatalytic maturation under acidic conditions [28,41]. Although pro-NbCP14AP contains an N-glycosylation sequon within its prodomain, the electrophoretic mobility of the recombinant proenzyme was not increased by treatment with peptide N-glycosidase F. This suggests that the potential N-glycosylation site of pro-NbCP14AP (Asn57) is structurally buried and thus inaccessible for N-glycosylation enzymes (Fig. 1B). Furthermore, these results indicate that the discrepancy between the size of the 31-kDa autocatalytic maturation product and its predicted molecular mass (24.4 kDa) is not due to N-glycosylation.

Opposite to the baculovirus expression system, pro-NbCP14 containing the proline-rich linker region could be produced in E. coli with good yields and purity (3 mg proenzyme per liter bacterial culture). The calculated molecular mass of the recombinant proenzyme (43220.3 Da) could be experimentally verified by liquid chromatography-electrospray ionization-mass spectrometry (43220.1 ± 1.5 Da), thus confirming proper removal of the initiator methionine and formation of the predicted 4 disulphide bridges (Fig. 2A). However, the recombinant proenzyme migrated in SDS-PAGE gels as a 51-kDa polypeptide, which gave rise to a 36-kDa band upon acid-induced autocatalytic maturation (Fig. 2B). The main N-terminal sequence of the 36-kDa protein as determined by Edman degradation was found to correspond to Ser143-Cys-Asp-Gly-CHN2 (LVG) or biotin-CA074 (CA074) followed by SDS-PAGE and western blotting using streptavidin-peroxidase for detection. The migration positions of prestained molecular mass standards are indicated.

We have then assessed the catalytic properties of mature NbCP14 using a series of synthetic substrates frequently used to monitor the activities of cysteine cathepsins and PLCPs. While the enzyme cleaved a number of endopeptidase substrates, no activity was observed with aminopeptidase substrates such as Arg-MCA (Table 1). In particular, NbCP14 was found to display high hydrolytic activity towards Z-Leu-Arg-MCA. It is of note that Z-Phe-Arg-MCA is cleaved much less efficiently by this enzyme, mainly due to a far slower turnover (Table 2). This indicates that NbCP14 favours leucine over phenylalanine in its specificity-determining S2 substrate.

We have also tested NbCP14 for its capacity to degrade native proteins. For these experiments, the two monoclonal anti-HIV antibodies 2F5 and 2G12 were chosen as substrates. We have previously observed that 2F5 is far more sensitive to PLCPs than 2G12 [12]. Treatment of 2F5 with two other N. benthamiana PLCPs, NbALP and cathepsin B, preferentially results in the formation of a discrete 40-kDa polypeptide due to cleavage within the CDR H3 loop of the antibody (Leu-Phe-Gly108; Val109-Pro-Ile) [21]. Entirely different results were obtained for NbCP14. This enzyme completely degraded both mAbs, with only transient accumulation of fragments with apparent masses of 30–40 kDa (Fig. 4). The N-terminus of the 30-kDa 2F5 polypeptide produced by NbCP14 was determined as Thr240-Cys-Pro; Table S2) as previously reported for papain [45] and human cathepsin L [12].

3.3. Subsite specificity profile of NbCP14

To assess the subsite specificity of NbCP14, we applied the conventional Proteomic Identification of Protease Cleavage Sites (PICS) procedure [30–32] as well as a recently reported adaptation thereof [33]. Preliminary experiments revealed that an enzyme-to-library ratio of 1:100 and an incubation time of 16 h result in the most informative data sets. Standard PICS analysis of tryptic E. coli and HEK293 libraries treated with NbCP14 under these conditions yielded 1971 and 1993 unique cleavage sequences, respectively (Fig. 5). The subsite specificity of cysteine cathepsins and PLCPs is largely determined by their S2 pocket [46]. Both sequence logos depict a

Fig. 2. Characterization of NbCP14 produced in E. coli. (A) Purified pro-NbCP14 was subjected to analysis by liquid chromatography-electrospray ionization-mass spectrometry. The theoretical molecular mass of the recombinant protein is 43220.3 Da. (B) Pro-NbCP14 was incubated in 0.1 M sodium acetate (pH 5.0), 2.5 mM DTT for 30 min at 37 °C (+) prior to analysis by SDS-PAGE and silver staining. An untreated sample (−) served as control. The migration positions of prestained molecular mass standards are indicated.

Fig. 3. Profiling of NbCP14 with activity-based probes. Mature NbCP14 was incubated for 30 min at 37 °C with the indicated concentrations of DCG-04, biotin-Leu-Val-Gly-CHN2 (LVG) or biotin-CA074 (CA074) followed by SDS-PAGE and western blotting using streptavidin-peroxidase for detection. The migration positions of prestained molecular mass standards are indicated.
Table 1

| Substrate               | Specific activity (mU/mg) | Relative activity (%) |
|-------------------------|---------------------------|-----------------------|
| Ala-Leu-Arg-MCA         | 771                       | 100                   |
| Boc-Val-Leu-Lys-MCA     | 158                       | 20.4                  |
| Bz-Phe-Val-Arg-MCA      | 112                       | 14.5                  |
| Boc-Val-Arg-MCA         | 48                        | 6.2                   |
| Z-Phe-Arg-MCA           | 39                        | 5.1                   |
| Z-Leu-Leu-Glu-MCA       | 26                        | 3.3                   |
| Z-Val-Val-Arg-MCA       | 25                        | 3.2                   |
| Suc-Leu-Leu-Val-Tyr-MCA | 4.4                       | 0.6                   |
| Z-Leu-Leu-Leu-MCA       | 3.8                       | 0.5                   |
| Z-His-Glu-Lys-MCA       | 0.3                       | <0.1                  |

Table 2

| Enzyme      | Substrate            | $k_v$ [mM] | $k_v$ [$s^{-1}]$ | $k_{cat}/V_{max}$ [mM$^{-1}$ s$^{-1}]$ |
|-------------|----------------------|------------|------------------|----------------------------------------|
| NbCP14      | Z-Leu-Arg-MCA        | 9.4 ± 1.5  | 1.7 ± 0.4        | 181                                    |
|             | Z-Phe-Arg-MCA        | 23 ± 2     | 0.05 ± 0.02      | 2.2                                    |
| Cat H$^a$   | Bz-Phe-Val-Arg-MCA   | 38 ± 8     | 110 ± 13         | 2900                                   |
| Papain$^b$  | Z-Phe-Arg-MCA        | 82 ± 26    | 4.3 ± 1.0        | 52.5                                   |
| Asp158Ala$^b$| Z-Phe-Arg-MCA        | 420 ± 70   | 52 ± 6           | 119                                    |
|             | Z-Phe-Arg-MCA        | 920 ± 70   | 0.46 ± 0.08      | 0.5                                    |

$^a$ Cat H, recombinant human cathepsin H lacking the mini-chain; data taken from Ref. [55].

$^b$ Asp158Ala, mutant papain (Asp$^{158} \rightarrow$ Ala); data taken from Ref. [59].

The NbCP14 subsite specificity was also assessed with an adapted PICS procedure [33]. This method avoids chemical modification of primary amines prior to exposure to a protease under investigation and thus also enables the profiling of lysine specificities. E. coli libraries generated by treatment with trypsin or GluC were incubated with NbCP14 under the same conditions as above prior to differential labelling of protease-treated and control samples by reductive dimethylation using different stable isotopes of formaldehyde. Upon analysis by liquid chromatography-mass spectrometry, NbCP14 cleavage products could be identified as semi-specific peptides enriched for the corresponding isotopic label. Two independent digests of tryptic E. coli libraries with NbCP14 yielded 149 and 207 cleavage sequences, respectively. Consistent with the data acquired by conventional PICS, the preference of NbCP14 for aliphatic amino acids in P2 was pronounced. Glycine and threonine were the most strongly enriched P1 residues (Fig. 6), which is also in good agreement with the results obtained using the conventional PICS protocol. Glycine has been previously reported to be tolerated well in P1 and P1’ of human cathepsins B, L and S [32]. Furthermore, threonine is frequently observed in the P1 position of cysteine cathepsin and PLCP specificity profiles [32,47]. P3 was again found to contribute little to the specificity of NbCP14 with solely histidine exceeding its natural abundance by more than 2-fold at this position. However, the adapted PICS procedure revealed that NbCP14 readily accepts acidic amino acids in P1’.

Many PLCPs accept lysine or arginine residues in the P1 position. Such sequences are rare in tryptic libraries due to the substrate specificity of this enzyme. Therefore, we additionally used the adapted PICS procedure to investigate the action of NbCP14 on a GluC-treated E. coli library, obtaining 213 cleavage sequences (Fig. 6). For P2, the same subsite specificity was observed as in the
amongst profiles. However, NbCP14 was found to display a preference for lysine and arginine in P1, which is in good agreement with the data available for other PLCPs [47]. An enrichment of basic amino acids was also observed at P2, whereas small and acidic amino acids dominated in P1. Taken together, these results demonstrate that the adapted and conventional PICS protocols yield comparable results for the substrate specificity of NbCP14. However, it should be pointed out that the sequence logos obtained with the modified procedure were much easier to interpret than those derived using the standard protocol.

The PLCP most closely related to cathepsin H is ALP. Like its mammalian counterpart, ALP primarily functions as aminopeptidase with moderate endopeptidolytic activity [21,48]. This is attributed to the so-called mini-chain, which partially occupies the active-site cleft and thereby restricts the substrate-binding sites of cathepsin H and related enzymes [49]. Adapted PICS assays of NbALP with tryp tic HEK293 libraries led to the identification of 252 cleavage sequences. Strikingly, a large fraction of the detected cleavage events (60%) resulted from the removal of a single residue and were thus due to the aminopeptidase activity of the enzyme. 27% corresponded to the removal of two or three amino acids. Notable differences were observed between the subsite preferences of NbALP and NbCP14. In P1, NbALP readily accepts bulky side chains such as those of phenylalanine, tyrosine and methionine. Serine and threonine are also tolerated in this position [Fig. 7]. Similar features have been reported for cathepsin H [50,51]. The S1’ pocket of NbALP prefers hydrophobic residues (valine, alanine, isoleucine) as well as tyrosine and aspartic acid. The P2 selectivity of NbALP is not as pronounced as that of NbCP14 and displays an enrichment of small neutral amino acids such as threonine, valine and alanine. Taken together, NbALP and NbCP14 clearly differ in their modes of action as well as in their substrate specificities.

4. Discussion

In tobacco, NbCP14 was identified as a key component of a bipartite module controlling the initiation of programmed cell death during the early stages of embryogenesis [3]. In the case of N. benthamiana, previous studies have revealed that NbCP14 is required for a powerful response to plant pathogens. This is reminiscent of tomato and potato C14 which also promote plant immunity [22,23]. The latter two enzymes are closely related to Arabidopsis thaliana RD21, an extensively studied protease from PLCP subfamily 1 [17,20,40]. However, NbCP14 appears to differ from RD21 in various aspects, thus providing evidence for substantial biochemical variations between subfamilies 1 and 4. For instance, our results unequivocally demonstrate that NbCP14 is capable of autocatalytic activation. By contrast, previous studies have indicated that maturation of recombinant RD21 produced in insect cells requires the addition of plant extracts [40]. Hence, it has been suggested that RD21 activation depends on the action of vacuolar processing enzymes, an unrelated class of cysteine proteases. However, we have recently obtained genetic evidence that vacuolar processing enzymes are dispensable for RD21 activation in A. thaliana [20]. Another difference between NbCP14 and RD21 relates to the functional significance of the proline-rich region connecting the catalytic and granulin domains. In the case of RD21, deletion of this spacer element resulted in accumulation of an apparently inactive precursor protein, suggesting that the proline-rich segment has to be present for proper folding of RD21-like proteases [20]. Conversely, recombinant NbCP14 lacking this linker sequence is capable of rapid autocatalytic maturation, thus highlighting remarkable differences between PLCP subfamilies 1 and 4 with respect to their folding competence and cellular activation mechanisms.

Opposite to NbALP and N. benthamiana cathepsin B [21], NbCP14 failed to cleave the anti-HIV mAb 2F5 in its CDR H3 loop. Interestingly, the cleavage site of the former enzymes in this region of the antibody (Leu-Phe-Gly Val-Pro-Ile) features a phenylalanine in the P2 position. Using synthetic peptide substrates, we observed that NbCP14 displays a pronounced preference for leucine at this location (see Table 1). This unique feature prompted us to test NbCP14 for its reaction with various activity-based probes which are frequently used to map the substrate specificities of cysteine proteinases. The prototypical PLCP probe is DCG-04, whose structure is based on the potent cysteine proteinase inhibitor E-64 [37]. DCG-04 and derivatives thereof have been successfully used to detect PLCPs in plant extracts, including the NbCP14 orthologue XBCP3 from A. thaliana [17,52]. Interestingly, all these compounds contain a leucine residue which docks into the affinity-determining S2 subsite of target PLCPs [53]. Hence, the strong labelling of NbCP14 by DCG-04 is in good agreement with the observed substrate specificity of this protease. NbCP14 also gave pronounced signals with another E-64 relative, bixin-CA074. This probe has been originally designed as a selective high-affinity label for cathepsin B [38]. The reaction of bixin-CA074 with NbCP14 indicates that the latter enzyme can tolerate the C-terminal dipeptide extension of the inhibitor in its prime-site pockets. By contrast, the binding of bixin-Leu-Val-Gly-CHN2 to NbCP14 was relatively weak. This could reflect the preference of NbCP14 for leucine over valine in the P2 position. However, it has been also noted that the bixin moiety of the probe can interfere with its efficient accommodation in the active site of PLCPs due to the absence of a linker segment [39].

For plants, very few proteomics-based studies of protease...
specification have been conducted so far. Recently, the PICS methodology has been used to determine the cleavage-site specificities of matrix metalloproteinases from A. thaliana [54]. In this study, we have now used NbCP14 to evaluate an improved PICS protocol [33] for its suitability to assess the substrate preferences of PLCPs. This new procedure circumvents a number of known PICS limitations, such as the lack of normalization and the modification of certain amino acids during library generation. The adapted method omits the dimethylation of primary amines and preserves free thiois in the peptide library allowing the recognition of unmodified N-termini, lysines and cysteines by the test protease. Importantly, the adapted and conventional PICS protocols yielded similar results with respect to the subsite preferences of NbCP14. As expected, the observed cleavage patterns are largely determined by the distinct P2 specificity of the enzyme, which clearly favours aliphatic over aromatic amino acids at this position as previously observed for human cathepsin S [32].

The strong preference of NbCP14 for leucine over phenylalanine in P2 is not only due to enhanced substrate affinity, but at large the result of a major difference in the turnover rate. Similar observations have been made previously for cathepsin H, which favours valine over phenylalanine in P2 like NbALP [55,56] (see Table 2). However, the residues forming the S2 subsite of NbCP14 (Met314, Ala280, Leu106, Ala308, Leu358) are inconspicuous when compared with papain [57]. Hence, a different molecular feature shared by NbCP14 and cathepsin H could contribute to the distinct enzymatic properties that these two enzymes have in common. In the proximity of the active-site histidine, cathepsin H contains a small four-residue loop (Lys155A-Thr-Pro-Asp155D) that is not found in other cysteine cathepsins [49]. It has been put forward that this flexible loop could cause partial obstruction of the S2 subsite [55]. Interestingly, NbCP14 also contains such an insertion (Ser301, Asn-Pro-Asp304) at the equivalent position.

Opposite to cathepsin H, NbCP14 does not display any amino-peptidase activity. Clearly, this can be attributed to the absence of the mini-chain, a segment of the propeptide which remains attached to mature cathepsin H and its plant orthologues via a disulphide bond not present in NbCP14 [21,55,58]. Another striking difference between NbCP14 and cathepsin H-like enzymes relates to the amino acid preceding the active-site histidine. For papain, it has been shown that this residue (Asp158) is involved in a hydrogen-bonding network stabilizing the thiolate-imidazolium ion pair required for catalysis [59]. While cathepsin H and NbALP feature an asparagine at this position, Asp158 is replaced by Ser307 in NbCP14. Site-directed mutagenesis experiments with papain have revealed that substitution of Asp158 with Asn has little impact on the catalytic efficiency of the protease [60]. However, the enzymatic activity of an Asp158-Ala variant was found to be severely compromised, which was mainly due to a more than 100-fold decrease in $k_{cat}$ [59] (see Table 2). Hence, Ser307 could at least in part account for the moderate processivity of NbCP14 as compared to papain and cathepsin H. Taken together, the unique molecular features of NbCP14 identified in this study may explain its unusual low-affinity interactions with endogenous cysteine proteinase inhibitors [3] and thus provide a rationale for the non-redundant role of the enzyme during execution of programmed cell death in the plant embryo.

Author contributions

MP performed experiments, analysed data and co-wrote the manuscript. UM, ST, AP, PS, DM and MB performed experiments and analysed data; RH contributed essential material; BL, MN and OS planned experiments and analysed data; LM conceived the study, planned and performed experiments, analysed data and co-wrote the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Databases

NbCP14 nucleotide sequence data are available in the DDBJ/EMBL/GenBank databases under the accession number KU212214.

Acknowledgements

The expert technical assistance of Franz Jehl, Bettina Mayer and Barbara Svoboda is gratefully acknowledged. We thank Matthew Bogoy, Dietmar Katinger, Gilles Lalmanach and Norbert Schaschke for reagents. Financial support was provided by the Austrian Science Fund (FWF): project W1224-B09; Deutsche Forschungsgemeinschaft (DFG): SCHI 871/2, SCHI 871/5, SCHI 871/6, GR 1748/6, INST 39/900-1, SPF850 - Project B8); European Research Council (Programme “Ideas” — Call identifier: ERC-2011- StG 282111-ProteaSys); and the Excellence Initiative of the German Federal and State Governments (EXC 294, BIOSS).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jab.2016.05.017.

References

[1] M.M. Goodin, D. Zaitlin, R.A. Naidu, S.A. Lommel, Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions, Mol. Plant Microbe Interact. 21 (2008) 1015–1026.
[2] A.C. Horger, R.A. van der Hoorn, The structural basis of specific protease-inhibitor interactions at the plant-pathogen interface, Curr. Opin. Struct. Biol. 23 (2013) 842–850.
[3] P. Zhao, X.M. Zhou, L.Y. Zhang, W. Wang, L.G. Ma, L.B. Yang, X.B. Peng, P.V. Bozhkov, M.X. Sun, A bipartite molecular module controls cell death activation in the basal cell lineage of plant embryos, PLoS Biol. 11 (2013) e1001655.
[4] Y.Y. Gleba, D. Tuse, A. Giritch, Plant viral vectors for delivery by Agrobacterium, Curr. Top. Microbiol. Immunol. 375 (2014) 155–192.
[5] R. Strasser, F. Altmann, H. Steinkellner, Controlled glycosylation of plant-produced recombinant proteins, Curr. Opin. Biotechnol. 30 (2014) 95–100.
[6] A. Loos, J.S. Gach, T. Hackl, D. Maresh, T. Henkel, A. Porodko, D. Bui-Minh, W. Sommeregger, G. Wozniak-Knopp, D.N. Forthal, F. Altmann, H. Steinkellner, L. Mach, Glycan modulation and sulfoengineering of anti-HIV-1 monoclonal antibody PC9 in plants, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 12675–12680.
[7] B. De Muynck, C. Navarre, M. Boutry, Production of antibodies in plants: status after twenty years, Plant Biotechnol. J. 8 (2010) 529–563.
[8] V.K. Hehle, M.J. Paul, P.M. Drake, J.K. Ma, C.J. van Dooleweerd, Antibody degradation in tobacco plants: a predominantly apoplastic process, BMC Biotechnol. 11 (2011) 128.
[9] C. Goulet, M. Khalf, F. Sainsbury, M.A. D’Aoust, D. Michaud, A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast, Plant Biotechnol. J. 10 (2012) 83–94.
[10] M. Donini, R. Lombardi, A. Aoust, F. Sainsbury, D. Michaud, Antibody production: a common picture emerging from plants, Bioengineered 6 (2015) 299–302.
[11] S. Robert, M. Khalf, M.C. Goulet, M.A. D’Aoust, F. Sainsbury, D. Michaud, Protection of recombinant mammalian antibodies from development-dependent proteinolytic degradation in leaves of Nicotiana benthamiana, Plant Cells S 1 (2012) e02633.
[12] M. Niemer, U. Mohefer, E. Benvenuto, M. Donini, J.K. Ma, Site-specific proteolytic degradation of IgG monoclonal antibodies expressed in tobacco plants, Plant Biotechnol. J. 13 (2015) 235–245.
[13] K. Dwadi, L. Chen, R. Menassa, S. Dhaubhadel, Identification, characterization
