



Mechanistic insights into CrCEP1: A dual-function cysteine protease with endo- and transpeptidase activity

Katarina P. van Midden^a, Melissa Mantz^{b,c}, Marko Fonovič^d, Martin Gazvoda^a, Jurij Svete^a, Pitter F. Huesgen^{b,c,e,f}, Renier A.L. van der Hoorn^g, Marina Klemenčič^{a,*}

^a Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

^b Central Institute for Engineering, Electronics and Analytics, ZEA-3, Forschungszentrum Jülich, Jülich, Germany

^c CECAD, Medical Faculty and University Hospital, University of Cologne, 50931 Cologne, Germany

^d Department of Biochemistry, Molecular and Structural Biology, Jozef Stefan Institute, Ljubljana, Slovenia

^e Faculty of Biology, University of Freiburg, Freiburg, Germany

^f CIBSS- Centre for Integrative Biological Signalling Studies, University of Freiburg, Freiburg, Germany

^g The Plant Chemetics Laboratory, Department of Biology, University of Oxford, OX1 3RB Oxford, UK

ARTICLE INFO

Keywords:

Proteolysis
Activity-based probes
Papain-like cysteine protease
Transpeptidation
Green algae

ABSTRACT

Proteases, essential regulators of plant stress responses, remain enigmatic in their precise functional roles. By employing activity-based probes for real-time monitoring, this study aimed to delve into protease activities in *Chlamydomonas reinhardtii* exposed to oxidative stress induced by hydrogen peroxide. However, our work revealed that the activity-based probes strongly labelled three non-proteolytic proteins—PsbO, PsbP, and PsbQ—integral components of photosystem II's oxygen-evolving complex. Subsequent biochemical assays and mass spectrometry experiments revealed the involvement of CrCEP1, a previously uncharacterized papain-like cysteine protease, as the catalyst of this labelling reaction. Further experiments with recombinant CrCEP1 and PsbO proteins replicated the reaction *in vitro*. Our data unveiled that endopeptidase CrCEP1 also has transpeptidase activity, ligating probes and peptides to the N-termini of Psb proteins, thereby expanding the repertoire of its enzymatic activities. The hitherto unknown transpeptidase activity of CrCEP1, working in conjunction with its proteolytic activity, unveils putative complex and versatile roles for proteases in cellular processes during stress responses.

1. Introduction

Proteases are a class of enzymes known for their ability to catalyse a fundamental chemical reaction: the hydrolytic cleavage of peptide bonds. This function is important not only for protein degradation, which facilitates protein turnover, but also for regulating protein structure, activity, and subcellular localization, making proteases critical enforcers of many physiological processes. In plants, proteases play a central role in regulating development, immunity, response to biotic and abiotic stresses, and the process of regulated cell death (RCD) [1–3].

Several subclasses of cysteine proteases, including metacaspases (MCAs), vacuolar processing enzymes (VPEs), and papain-like cysteine proteases (PLCPs), alongside subtilisin-like serine proteases (subtilases or saspases) and the threonine protease proteasome subunit PBA1, have all been associated with functions in developmental [4] and stress-

induced modes [5,6] of RCD in various chlorophyll-containing organisms. Despite these associations, the precise role and mode of action of the proteases involved in these processes is still the subject of ongoing research and debate [3,7,8].

Activity-based probes (ABPs) have proven to be powerful tools for detecting the activity of these enzymes in cell extracts and living cells, enabling their characterization. ABPs are small molecules designed to covalently bind to active-site residues when proteases are in their active state. Tagged proteases can then either be detected on protein gels with fluorophores or enriched by use of e. g. biotin-tags and identified by mass spectrometry [9]. Activity-based probes have previously been successfully implemented in labelling and characterising plant proteases of numerous families, including VPEs [10] and PLCPs [11]. Recently, we developed novel activity-based metacaspase targeting probes that proved useful in labelling active isoforms of metacaspases *in vitro* [12].

* Corresponding author.

E-mail address: marina.klemencic@fkkt.uni-lj.si (M. Klemenčič).

<https://doi.org/10.1016/j.ijbiomac.2024.132505>

Received 12 December 2023; Received in revised form 5 April 2024; Accepted 17 May 2024

Available online 18 May 2024

0141-8130/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

The green alga *Chlamydomonas reinhardtii* (hereafter referred to as *Chlamydomonas*) is a representative of the early-diverged Viridiplantae and therefore shares ancestral traits with higher plants. *Chlamydomonas* has been proposed as a great platform for the comparative study of metabolic and signalling pathways of higher plants because it has fewer gene duplications and thus gene redundancy, while still containing representatives of the protease families listed above. However, not a lot is known about *Chlamydomonas* proteases, with only 8 out of 352 proteases in *Chlamydomonas* being functionally characterized so far [13]. In *Chlamydomonas*, the introduction of hydrogen peroxide has been empirically demonstrated to incite oxidative stress, ultimately culminating in regulated cell death [14–16]. Notably, an increase in transcript abundance of metacaspase type I (CrMCA-I) has been observed following two sequential treatments with H_2O_2 , suggesting potential activation of metacaspases during this process in

Chlamydomonas [15].

In this study, we aimed to use metacaspase-targeting activity-based probes to monitor this process in *Chlamydomonas*. Contrary to our expectations, however, the ABPs were instead recognized as substrates by a before uncharacterized papain-like cysteine protease, which transferred a part of the probe to N-termini of the three oxygen-evolving enhancer (OEE) proteins. By studying the catalytic mechanism of this process, our results not only shed light on this non-canonical activity of cysteine proteases but also raise intriguing questions about the putative physiological significance of this process.

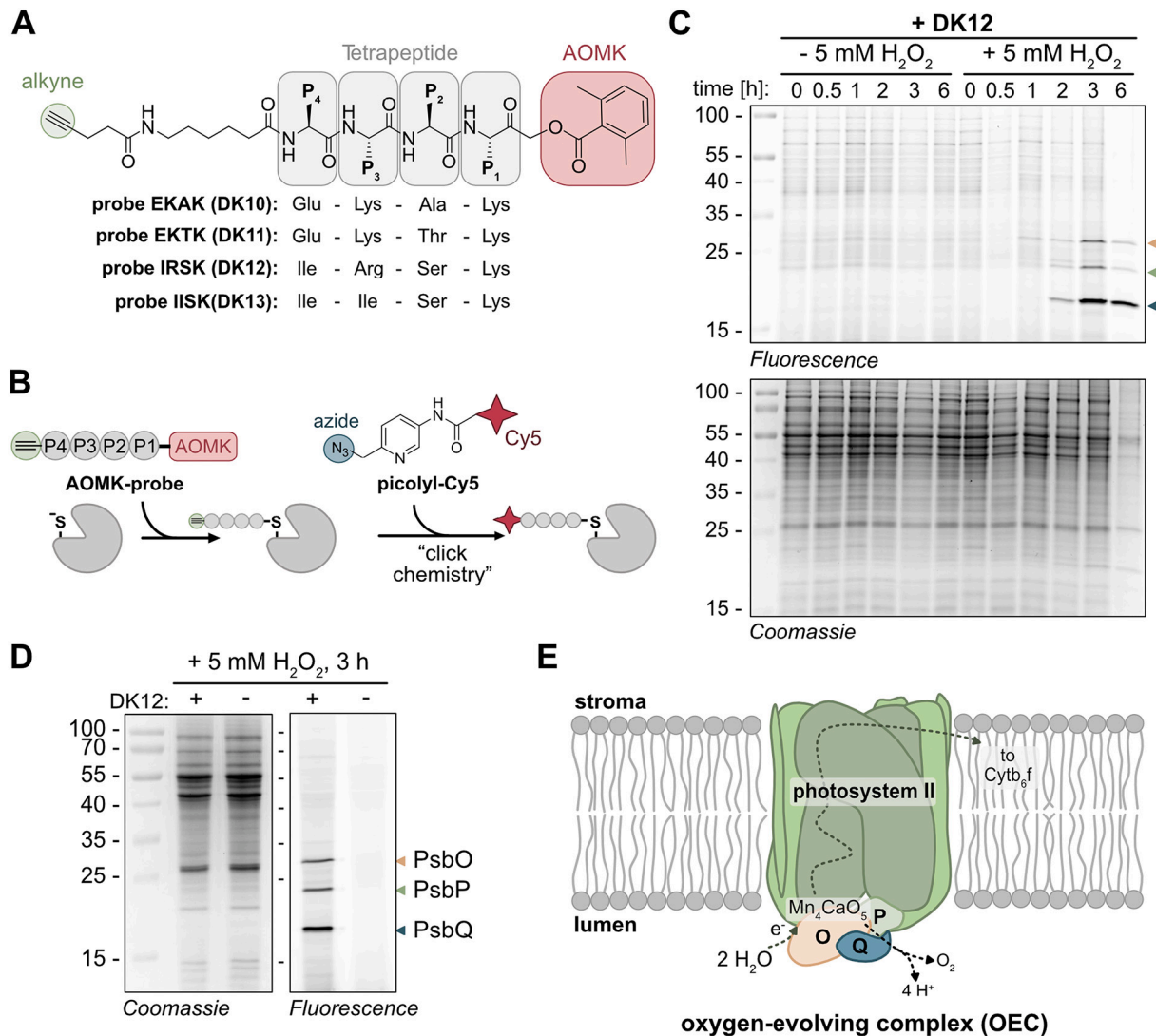


Fig. 1. AOMK-probes label proteins PsbO, PsbP and PsbQ in H_2O_2 -treated *Chlamydomonas*. (A) Structures of the four AOMK-based tetrapeptidyl probes. The probes consist of three key components: an alkyne minitag (shown in green), a tetrapeptide sequence (shown in grey), and the AOMK warhead (shown in red). The tetrapeptide sequences are shown below the structure along with the names of the probes. The design of the probes is described in [12]. (B) A schematic representation of the labelling mechanism of metacaspases. The AOMK warhead enables the covalent binding of the probes to the active site Cys. Picolyl-Cy5 or any other azide-carrying molecule can be used to detect and/or identify the labelled protease (C) Labelling of soluble proteomes of H_2O_2 -treated *Chlamydomonas* cells with DK12. The samples were separated on SDS-PAGE under reducing conditions and the gel was scanned for Cy5 fluorescence (upper panel) and stained with Coomassie (lower panel). (D) Labelled *Chlamydomonas* samples prior to MS analysis. The positions of labelled PsbO, PsbP and PsbQ of photosystem II, are denoted with sand, green, and blue arrows, respectively. (E) Schematic representation of the oxygen-evolving complex (OEC) of photosystem II. The complex includes proteins PsbO (O, sand), PsbP (P, green) and PsbQ (Q, blue) and the manganese cluster (Mn_4CaO_5) and is responsible for the splitting of water into molecular oxygen and electrons in the process of photosynthesis.

2. Results

2.1. AOMK-probes label specific proteins in peroxide-stressed *Chlamydomonas proteomes*

To monitor the activity of metacaspases in *Chlamydomonas* undergoing hydrogen peroxide-induced regulated cell death (RCD), we used our recently validated activity-based probes, which were shown to label all three metacaspase types *in vitro* [12]. The four probes contain tetrapeptides with a Lys residue at the P1 position and various residues at positions P2-P4, to direct specificity towards metacaspases (Fig. 1A). All probes contain an acyloxymethylketone (AOMK) warhead, located C-terminally to the tetrapeptide, that enables the formation of a stable covalent bond between the probe and the catalytic Cys residue in the active site of metacaspases, which irreversibly inhibits their proteolytic activity. The designed AOMK probes also contain an alkyne minitag (\equiv) that can subsequently be coupled to an azide-picolyl-Cy5 fluorophore by click chemistry, enabling the detection of tagged proteins, separated on SDS-PAGE gels, by scanning for their fluorescence (Fig. 1B).

Among the four tested probes, the probe DK12, containing tetrapeptide IRSK, showed the highest binding affinity towards metacaspases [12]. Therefore, this probe was used to investigate metacaspase activities during oxidative stress induced by H_2O_2 treatment in *Chlamydomonas*. Since studies have shown that RCD in this organism occurs at H_2O_2 concentrations of 5–10 mM [14,16], cells were treated with 5 mM H_2O_2 and collected at chosen time points after treatment. Soluble proteomes were prepared and labelled with 2 μ M DK12 to detect metacaspase activity.

Three distinct bands of approximately 28, 22, and 18 kDa were detected on reducing SDS-PAGE gels 2 h after treatment with H_2O_2 , while no labelled proteins were observed in non-treated cells (Fig. 1C). This was rather unexpected as the annotated genome of *Chlamydomonas* contains only two metacaspases (*CrMCA-I* (XP_001696956.1) and *CrMCA-II* (XP_001691826.1)), having predicted molecular weights of active isoforms of 35 kDa (*CrMCA-I* lacking the prodomain) and 25 kDa (corresponding to the processed catalytic p20 domain).

To identify the labelled proteins, we performed large scale affinity capture using azide-tagged agarose beads followed by tandem MS. The MS analysis revealed that the three bands did not correspond to metacaspases, but instead represented the three oxygen-evolving enhancer proteins (OEE1–3), also known as PsbO, PsbP, and PsbQ (Fig. 1D, Supplemental Table 1, Supplemental Fig. S2). These proteins are extrinsic components of photosystem II (PSII), localized in the lumen of thylakoid membranes inside the chloroplast and form the oxygen-evolving complex (OEC) in conjunction with the inorganic Mn_4O_5Ca cluster (Fig. 1E). The OEC plays a vital role in water splitting, generating molecular oxygen (O_2) and electrons that proceed through PSII during the process of photosynthesis [17–19]. The identification of these proteins as the labelled bands was unexpected since they are not known to possess any proteolytic activity. Moreover, two of the proteins, PsbQ and PsbP, do not contain any cysteines in their mature forms that could react with the AOMK group of the probe. To confirm that PsbO is one of the labelled proteins, we repeated the peroxide treatment and labelling on a *Chlamydomonas* knockout cell line lacking PsbO. The 28 kDa band corresponding to PsbO, was absent in this mutant (Supplemental Fig. S3), confirming that PsbO labelling causes the 28 kDa signal. Interestingly, we also observed that the amount of PsbO detected in proteomes of H_2O_2 -treated cells was greater than in the untreated ones. To assess this in more detail, we used the PsbO-specific antibody to detect PsbO in soluble proteomes from cells collected at various time points following H_2O_2 treatment. We observed a substantial increase in PsbO levels in the treated soluble proteomes 2–3 h post-treatment, coinciding with increased labelling of the three proteins (Supplemental fig. S4).

2.2. Labelling of rPsbO is blocked by cysteine protease inhibitors

To investigate the underlying mechanism behind the unexpected labelling of Psb proteins, we recombinantly expressed the mature form of *Chlamydomonas* PsbO in *E. coli*. PsbO accumulated in the insoluble fraction, which required purification of the recombinant PsbO (rPsbO) from inclusion bodies and subsequent refolding. The correct folding of rPsbO was confirmed by size exclusion chromatography and CD spectroscopy (Supplemental Fig. S5).

Only background signals appeared when rPsbO was incubated with DK12 alone (Fig. 2A), which could be attributed to nonspecific Cy5 binding and fluorescence. However, in the presence of the proteome of H_2O_2 -treated cells, labelling of PsbO with the probe was clearly detected (Fig. 2A). These results suggested that rPsbO did not have an innate affinity for the probe but that additional components in the cell proteome were required for labelling the rPsbO protein.

To identify the unknown component in the proteome of treated cells, we investigated labelling in more detail. First, we used the four available AOMK-probes for labelling to check for specificity. The probes differ only in the tetrapeptide sequence, nevertheless only the IRSK tetrapeptide probe caused strong labelling of the three proteins (Fig. 2B), indicating that the tetrapeptide sequence is important for labelling. Next, we examined how labelling depends on the pH. As shown in Fig. 2C, labelling occurred exclusively in the neutral to slightly alkaline pH range (pH 7–9). Third, we aimed to determine whether labelling could be blocked by the addition of protease inhibitors to assess the possible involvement of proteases. We tested EDTA, PMSF, E-64, and the recombinantly produced suicide inhibitor CrSerp, targeting metalloproteases, serine proteases, and cysteine proteases, respectively. Labelling was completely abolished only when soluble proteomes were pretreated with either E-64 or rCrSerp (Fig. 2D). Given that E-64 and rCrSerp both inhibit papain-like cysteine proteases (PLCPs), these results suggested the involvement of a PLCP in the process of labelling of the Psb proteins with the IRSK probe.

Chlamydomonas harbours 19 PLCPs of the C1A family [20]. Of these, at least six are likely to be active C1A family proteases since they possess the characteristic amino acid residues Cys, His, and Asn, which form the catalytic triad. However, the specific activities of these PLCPs in *Chlamydomonas* are unknown. To characterise the activity of PLCPs in *Chlamydomonas* cells exposed to oxidative stress, we performed activity-based protein profiling (ABPP) on proteomes of treated and untreated *Chlamydomonas* using the activity-based probe DCG-04 [21]. DCG-04, a biotinylated derivative of E-64, irreversibly binds to the active sites of PLCPs, allowing for their selective labelling. Proteomes of H_2O_2 -treated and untreated *Chlamydomonas* cells collected at different timepoints were incubated with DCG-04, blotted to a PVDF membrane and visualized using streptavidin-HRP. We could observe faint labelling of two bands in all samples that were suppressed upon pretreatment with E-64 (Supplemental Fig. S6).

To identify the DCG-04-labelled proteases, we performed large-scale ABPP of the *Chlamydomonas* proteome. For this experiment, proteomes of H_2O_2 -treated *Chlamydomonas* cells were incubated with DCG-04 and the biotinylated proteins were captured and enriched using biotin-binding streptavidin magnetic beads. MS analysis of the on-bead digest revealed two PLCPs: CrCEP1 (cysteine endopeptidase-1, Cre09.g407700, A0A2K3DFFE4) and CrCEP2 (cysteine endopeptidase-2, Cre05.g247851, A8I5R9) (Fig. 2E).

Both CrCEP1 and CrCEP2 were significantly enriched compared with the E-64 pretreated control, indicating their substantial activity in soluble proteomes of H_2O_2 -treated cells. These results indicate that one or both proteases might be involved in the labelling of the Psb proteins.

2.3. Labelling of rPsbO can be replicated *in vitro*

To gain further insight into the involvement of CrCEP1 and CrCEP2 in labelling, we obtained their synthetic genes and expressed them in

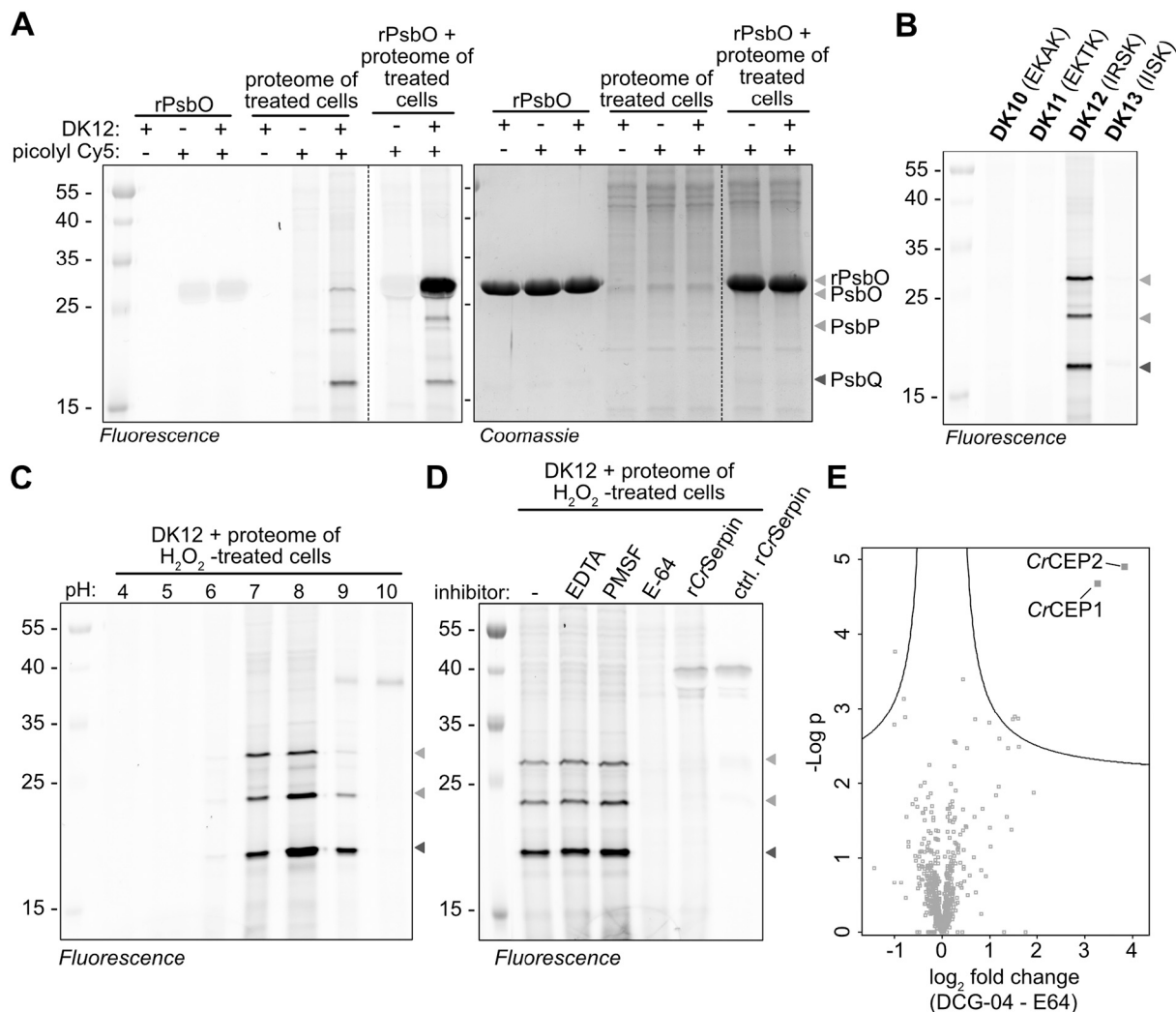


Fig. 2. Labelling is dependent on the activity of a cysteine protease. All labelling experiments were performed with 2 μ M AOMK-probes and the soluble proteome of cells collected 3 h after treatment with 5 mM H_2O_2 (proteome of treated cells). Alkyne-labelled proteins were coupled to picolyl Cy5, separated on a 15 % SDS-PAGE under reducing conditions, scanned for fluorescence, and/or stained with Coomassie. The expected positions of labelled PsbO and recombinant PsbO (rPsbO) are indicated with sand-coloured arrows, PsbP with green arrows, and PsbQ with blue arrows. (A) Labelling of rPsbO with probe DK12 in the presence of the proteome of H_2O_2 -treated *Chlamydomonas*. Dashed lines indicate cut images. (B) Labelling of the proteome of H_2O_2 -treated *Chlamydomonas* with all four AOMK-probes. Cy5 fluorescence is shown. (C) The pH profile of the labelling reaction. The proteome of H_2O_2 -treated *Chlamydomonas* was labelled in buffers with different pH values (4–10), 150 mM NaCl and 5 mM DTT. Cy5 fluorescence is shown. (D) Labelling of the proteome of H_2O_2 -treated *Chlamydomonas* in the presence of different protease inhibitors. rCrSerpin in the absence of the treated proteome was added as a negative control where background Cy5 fluorescence can be observed. (E) Volcano plot of ABPP of the proteome of H_2O_2 -treated *Chlamydomonas* using the probe DCG-04. The black line represents the threshold for statistical significance. The two proteins that showed the most significant fold change (\log_2 fold change >2 , x-axis) and high statistical significance ($-\log_{10}$ of p -values >3 , y-axis) between DCG-04-labelled and E-64-pretreated control samples are indicated in blue and represent two papain-like cysteine proteases: CrCEP1 and CrCEP2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

E. coli. These proteases contain sequences corresponding to a signal peptide, an autoinhibitory prodomain, a catalytic domain, a proline-rich region, and a granulin domain (Supplemental fig. S7). Our constructs included these sequences, lacking only the signal peptide. The autoinhibitory propeptide was included, as it has been shown to be critical for the correct folding of PLCPs [22]. Both pro-CrCEP1 and pro-CrCEP2 were expressed as soluble proteins, and subsequent isolation by metal affinity chromatography and size exclusion chromatography resulted in sufficient purity for activation (Supplemental Fig. S8).

Previous studies have shown that lowering the pH to 4.0–4.5 can induce autocatalytic maturation of PLCPs, including a granulin domain containing protease from *Nicotiana benthamiana* [23]. Therefore, purified pro-CrCEP1 and pro-CrCEP2 were incubated in acetate buffer at pH 4.0. The predicted maturation steps of granulin domain-containing proteases, including CrCEP1 and CrCEP2 are depicted in Fig. 3A

[24,25]. SDS-PAGE analysis of samples collected at different time points post-activation revealed that only CrCEP1 underwent autoactivation under these conditions (Fig. 3B).

Pro-CrCEP1, with a predicted molecular weight (MW) of 52 kDa, migrated faster than expected on the SDS-PAGE gel (apparent 47 kDa band), suggesting a possible C-terminal truncation that may have occurred during expression or purification, as compared with the higher MW of the same protein in the insoluble fraction (Supplemental Fig. S9). However, despite this truncation, pro-CrCEP1 underwent autocatalytic maturation when incubated at low pH, resulting in a progressive conversion into a mature 35-kDa form. As the predicted MW of the major and minor forms of mature CrCEP1 (with or without the proline-rich region) are 27 and 23 kDa, respectively, mature CrCEP1 showed slower electrophoretic mobility than expected. Similar observations have been made for other granulin-containing proteases, including

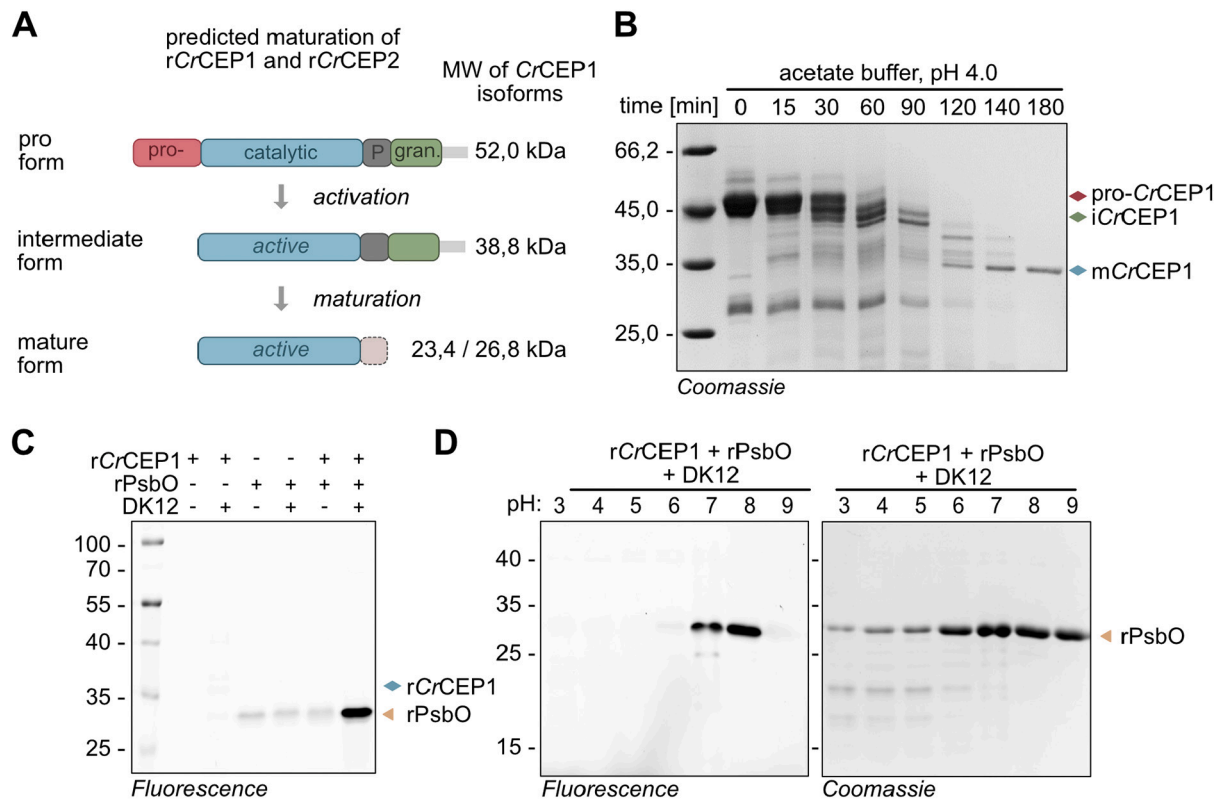


Fig. 3. Recombinant CrCEP1 enables labelling of the recombinant PsbO *in vitro* (A) Predicted maturation of rCrCEP1 and rCrCEP2. The predicted structural elements of the two proteases are depicted as follows: the autoinhibitory prodomain (pro-) in red, the catalytic domain in blue, the proline-rich region (P) in grey and the granulin domain in green. The remainder of the C-terminal sequence is depicted in grey. Predicted MW for rCrCEP1 are indicated next to the respective forms. Mature CrCEP1 can include or exclude the proline-rich region (denoted in light grey). (B) Activation of proCrCEP1 in acetate buffer. The pro, intermediate and mature forms of CrCEP1 are depicted with a red, green, and blue diamond shape, respectively. (C) *In vitro* labelling of rPsbO with active recombinant CrCEP1 (rCrCEP1). rPsbO and rCrCEP1 are denoted with a sand-coloured arrow and a blue diamond, respectively. (D) Labelling of rPsbO with rCrCEP1 at different pH points. Alkyne-labelled proteins were coupled to picolyl Cy5, separated on an SDS-PAGE under reducing conditions, scanned for fluorescence (left panel) and stained with Coomassie (right panel). Recombinant PsbO is denoted with a sand-coloured arrow.

RD21 A (Responsive to dehydration 21 A) from *Arabidopsis thaliana* [24,25] and the protease NbCysP6 from *Nicotiana benthamiana* [23]. It has been suggested that this phenomenon may be attributed to the fact that PLCPs are more resistant to denaturation by SDS [24]. Since no activation of CrCEP2 could be achieved (Supplemental Fig. S10), we used the recombinant mature CrCEP1 (rCrCEP1) for our subsequent experiments.

We first investigated if rCrCEP1 is an active protease by measuring the hydrolytic cleavage of the fluorogenic substrate Z-Phe-Arg-AMC. This substrate has previously been used for the characterization of other RD21A-like cysteine proteases including NbCysP6 from *Nicotiana benthamiana* [23]. Indeed, rCrCEP1 hydrolyzed this fluorogenic peptide efficiently, with a pH optimum at 6.0, and lower proteolytic activity at pH 8.0, the pH at which the labelling experiments were performed (Supplemental Fig. S11).

We next tested if adding rCrCEP1 would enable labelling of rPsbO. Indeed, mixing the IRSK probe with purified rPsbO and rCrCEP1 resulted in labelling of rPsbO (Fig. 3C), suggesting that rCrCEP1 enables transfer of the probe to rPsbO. Further assays show that rCrCEP1 effectively labelled rPsbO at neutral to slightly basic pH, but cleaved rPsbO at pH levels below 6.0 (Fig. 3D). These results indicated that the activity of the protease shifted from proteolysis at low pH to labelling at higher pH, indicating that the protease's proteolytic/labelling activity was pH-dependent.

2.4. rCrCEP1 labels rPsbO at its N-terminus

We further investigated the molecular details of rPsbO labelling. We

first examined if labelling requires free amino groups in rPsbO by pre-treating rPsbO with sulfo-N-hydroxysuccinimide acetate (sulfo-NHS-acetate), which reacts with primary amines, such as the α -amine at the N-terminal and the ϵ -amine of lysine side chains, thereby blocking their nucleophilic reactivity [26]. The pretreatment with sulfo-NHS-acetate before incubation with rCrCEP1 and the IRSK probe completely abolished labelling, suggesting that the IRSK probe reacts with either lysine residues or the N-termini of proteins (Fig. 4A).

There are 23 lysine residues in rPsbO, the positions of which are highlighted in Fig. 4B. To further investigate the labelling of rPsbO, we first produced and purified N-terminally truncated variants of rPsbO, lacking the first eleven ($\Delta 11$ -rPsbO) or eighteen residues ($\Delta 18$ -rPsbO), including or excluding the first lysine residue (K17) in the protein (Fig. 4B). Our hypothesis was that if labelling predominantly occurred at the ϵ -NH₂ group of lysine residues, we would observe variations in labelling efficiency between these two variants. Additionally, labelling in both would be abolished if a specific N-terminal sequence were required. Surprisingly, both truncated versions exhibited a similar labelling efficiency, failing to provide a definitive labelling site (Fig. 4C).

To address this further, we performed a competitive labelling experiment with synthetic peptides derived from the rPsbO sequence. We hypothesised that the labelling of rPsbO would decrease if peptides served as suitable substrates for this reaction, as less probe would be available to bind to the protein. We first tested a peptide consisting of the first 20 amino acid residues of rPsbO including the first lysine residue K17 (Pep20). Indeed, we observed a decreasing efficiency in the labelling of rPsbO with increasing concentrations of Pep20, confirming that labelling can also be performed with synthetic peptides (Fig. 4D).

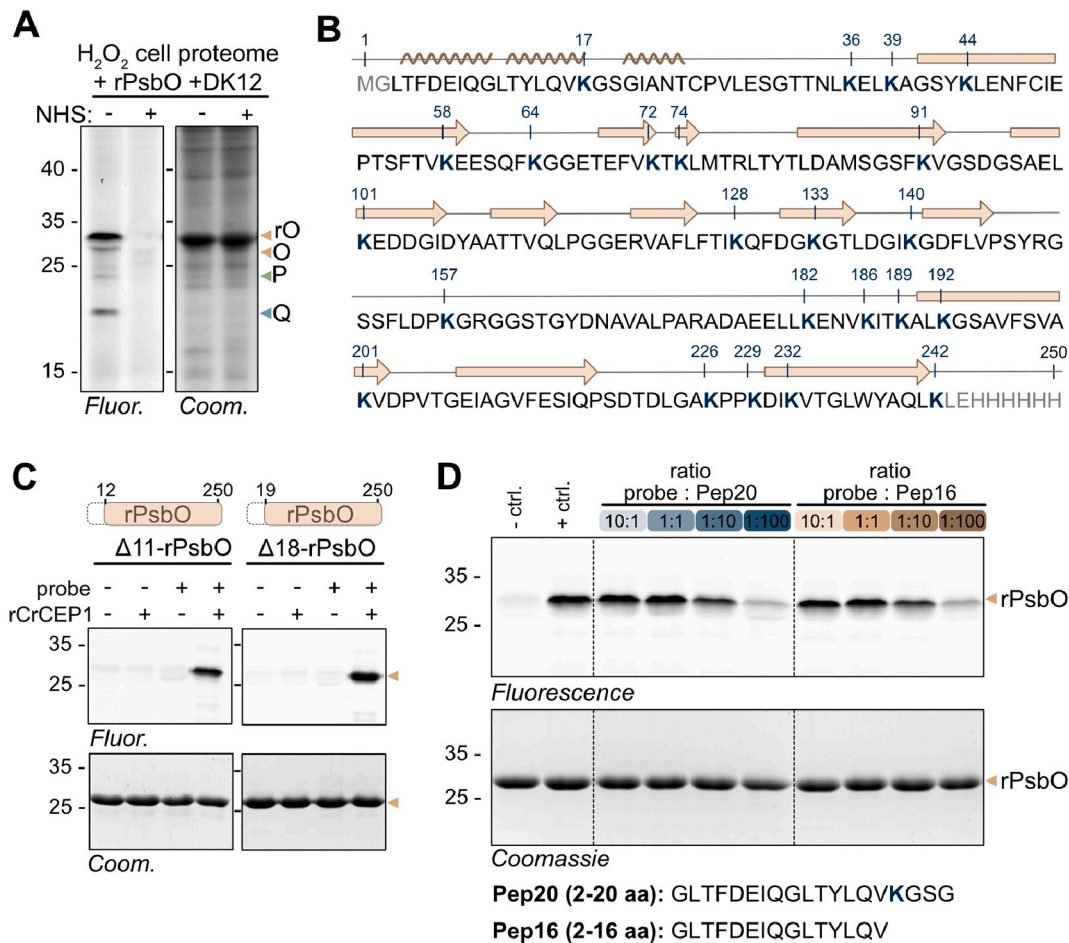


Fig. 4. Characteristics of rPsbO labelling. (A) Labelling in the presence of sulfo-NHS-acetate (NHS). Fluorescence is shown on the left and Coomassie staining on the right. (B) Annotated amino acid sequence of the mature rPsbO. Above the sequence, the corresponding secondary structures are shown according to the structure of wild-type PsbO (PDB: 6KAC, entity 14), with β -sheets and α -helices depicted as arrows and wavy lines, respectively. Lysine residues are highlighted in bold and blue, and their positions within rPsbO are indicated above the sequence. Additional N- and C-terminal sequences that enabled expression and purification of the protein, are coloured grey. (C) Labelling of truncated versions of rPsbO, lacking the first 11 ($\Delta 11$ -rPsbO) or 18 ($\Delta 18$ -rPsbO) amino acid residues. Fluorescence is shown above and Coomassie staining is shown below. Schematic representation of truncated PsbO constructs with the positions of their sequences within full-length rPsbO protein are depicted above the panels. (D) Competitive labelling of rPsbO with DK12 in presence of N-terminal rPsbO peptides. Fluorescence is shown above and Coomassie staining is shown below. The sequences of synthetic peptides are indicated below the panels (aa – amino acid residue). The expected positions of labelled PsbO (O), rPsbO (rO), PsbP (P) and PsbQ (Q) are depicted with sand-, green- and blue-coloured arrows, respectively. Dashed lines indicate cut images.

This result allowed us to further investigate whether the probe binds to ϵ -NH₂- or N-terminal primary amino groups by designing a second peptide without the lysine residue K17 (Pep16). We hypothesised that peptide labelling would be disrupted if the lysine residues were an essential part of the probe's binding mechanism, resulting in increased labelling of rPsbO. However, the truncated peptide Pep16 also showed a reduction in the efficiency of rPsbO labelling, comparable to that observed with Pep20. Since the N-terminal primary amines were the only primary amines present on both peptides, this suggests that they are essential to the probe's binding mechanism.

2.5. rCrCEP1 cleaves the tetrapeptide in the IRSK probe

While we had identified the putative binding site for the IRSK probe on the PsbO protein, the precise mechanism of rCrCEP1's involvement in mediating probe transfer remained unknown. To address this, we examined the interaction between rCrCEP1 and the IRSK probe. Since the IRSK probe was initially designed as an irreversible inhibitor of proteases, we tested whether the probe inhibits rCrCEP1 proteolytic activity. To assess this, we employed the fluorogenic Z-Phe-Arg-AMC substrate to measure the activity of rCrCEP1 in the presence of the IRSK probe at both pH 5.0 and pH 8.0. We observed no significant reduction

in proteolytic activity, indicating that the IRSK probe does not inhibit this protease (Supplemental Fig. S12).

This result prompted us to investigate whether the labelling reaction could occur without the AOMK warhead. To explore this possibility, we commercially obtained the alkyne-tagged tetrapeptide Ile-Arg-Ser-Lys (\equiv IRSK), lacking only the AOMK warhead (Fig. 5A) and conducted a labelling experiment. Indeed, labelling occurred even in the absence of the AOMK warhead (Fig. 5B). To assess how much of the peptide is necessary for labelling to still occur, two C-terminally truncated probe variants were synthesized. These truncated probes retained the N-terminal linker containing the alkyne minitag (\equiv) and included either the amino acid isoleucine (\equiv I) or the tripeptide Ile-Arg-Ser (\equiv IRS) (Fig. 5C). Our results demonstrate that while \equiv IRS still facilitated the labelling reaction, when rCrCEP1 was exposed to \equiv I, no observable labelling of rPsbO occurred (Fig. 5D), suggesting that the tetrapeptide sequence is the critical component that enables labelling.

To gain more insight into the specificity of rCrCEP1, we employed the Proteomic Identification of Protease Cleavage Sites (PICS) method [27]. To comprehensively explore the catalytic specificity of rCrCEP1, three peptide libraries were utilized, prepared with trypsin (resulting in cleavage after K/R), legumain (resulting in cleavage after N/D), or GluC (resulting in cleavage after E/D).

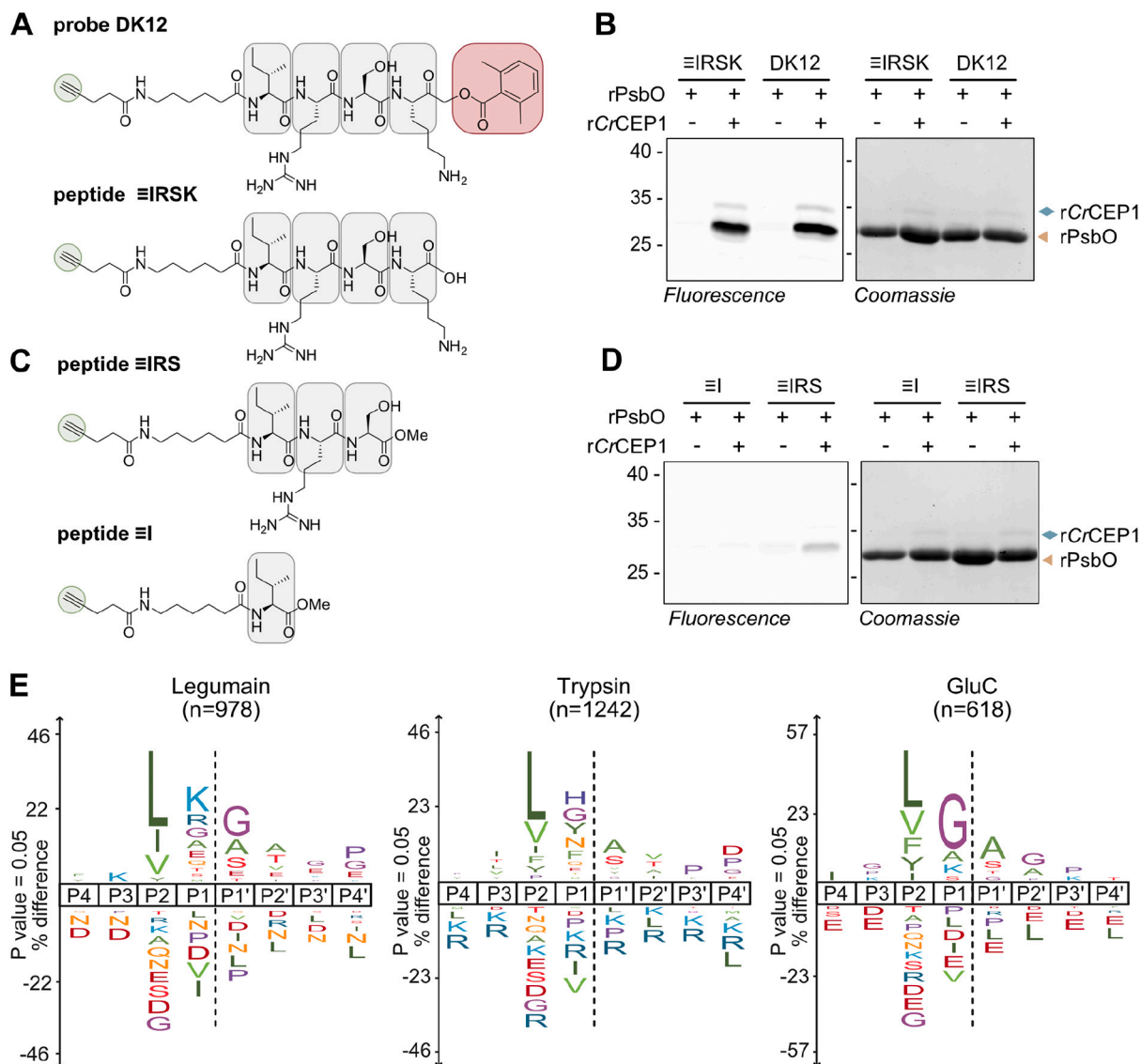


Fig. 5. rCrCEP1 interacts with the tetrapeptide sequence IRSK. (A) Chemical structures of the probes DK12 and ≡IRSK. The probes consist of a N-terminal linker containing the alkyne minitag (≡, coloured green), along with the amino acid residues Ile-Arg-Ser-Lys (IRSK, coloured grey). The AOMK group is depicted in red. (B) *In vitro* labelling of rPsbO with probes ≡IRSK and DK12 in presence of rCrCEP1. Alkyne-labelled proteins were coupled to picolyl Cy5, separated on an SDS-PAGE gel under reducing conditions, scanned for fluorescence (upper panel) and stained with Coomassie (bottom panel). Expected positions for rPsbO and rCrCEP1 are denoted by a sand-coloured arrow and a blue diamond, respectively. (C) Chemical structures of the chemically synthesized probes ≡I and ≡IRS. The probes consist of the N-terminal linker containing the alkyne minitag (≡, coloured green), along with the amino acid residues Ile (I) or Ile-Arg-Ser (IRS) that are coloured grey. (D) *In vitro* labelling of rPsbO with probes ≡IRS and ≡I in presence of rCrCEP1. The experiment was conducted as described for Fig. 5B. (E) Substrate cleavage specificity of rCrCEP1. The IceLogos visualize the amino acid frequencies surrounding the cleavage sites. These were inferred from peptides identified by semi-specific database searches after rCrCEP1 digestion of a proteome-derived peptide library created with either legumain (peptides ending with N or D), trypsin (peptides ending with E or D) or GluC (peptides ending with E or D). The number of nonredundant cleavage sites for each logo is indicated.

The PICS profiles we obtained revealed that rCrCEP1's substrate preferences primarily revolve around its specificity at the P2 position, a characteristic consistent with other members of the PLCP family, where the S2 binding pocket is recognized as the pivotal substrate binding site [28–30]. Like many other PLCPs, rCrCEP1 exhibited a preference for hydrophobic amino acids (such as Leu, Iso, Val, Phe, Tyr) at the P2 position. Additionally, rCrCEP1 displayed a preference for hydrophilic residues, particularly positively charged Lys and Arg, at the P1 position, and for small amino acid residues (Gly, Ala, Ser) at the P1' position (Fig. 5E). These findings explain the efficient cleavage of Z-Phe-Arg-AMC and suggest that the IRSK tetrapeptide could serve as a substrate for rCrCEP1 when cleaved between the arginine and serine residues.

To substantiate the hypothesis that IRSK tetrapeptide is cleaved in the middle by rCrCEP1, we incubated the tetrapeptide IRSK with

rCrCEP1 at pH 5.0 and analysed the reaction mixture using MS. Indeed, we detected both IR and SK cleavage products, confirming that the IRSK tetrapeptide serves as a suitable substrate for rCrCEP1 cleavage (Supplemental Fig. S13).

2.6. rCrCEP1 labels PsbO via transpeptidation

In view of the data presented above, we hypothesised that rCrCEP1 functions as a transpeptidase, cleaving the peptide probe and catalysing the synthesis of a peptide bond between the N-terminal part of the probe (≡IR) and the N-terminus of the Psb proteins. Considering already published data on proteases and transpeptidation (reviewed by Goettig, 2021), we propose the following reaction mechanism:

The activated thiol group of the catalytic cysteine residue initiates a

nucleophilic attack on the carbonyl group of the peptide bond between the arginine and serine residues of the probe. This process results in the creation of a thioester intermediate, wherein the peptide fragment Ile-Arg, carrying the alkyne minitag (\equiv IR), remains covalently attached to the catalytic cysteine residue of the protease. At pH values below 7.0 the dominant reaction is hydrolysis, whereas at pH values above 7.0 the intermediate reacts with the N-terminal amine leading to the formation of a new peptide bond. Through this transpeptidation process, the N-terminus of the protein is modified with the \equiv IR moiety (Fig. 6A).

To test this hypothesis experimentally, rPsbO was mixed with rCrCEP1, labelled by \equiv IRSK, and separated by SDS-PAGE. The band corresponding to the labelled rPsbO was excised and subjected to in-gel digestion. For digestion, legumain was used instead of trypsin, to prevent cleavage after the \equiv IR modification. The resulting peptides were analysed by LC-MS/MS using the MaxQuant software. A search for the N-terminal \equiv IR modification successfully identified the \equiv IR-modified N-terminal peptide of rPsbO (Supplemental Fig. S14). This confirmed our hypothesis that rCrCEP1 acts as a transpeptidase at near neutral pH. To validate these results, the same experiment was also performed with the recombinant mature PsbQ, where the same modification was observed (Supplemental Fig. S15).

3. Discussion

The data presented in this manuscript demonstrate the versatile nature of the previously uncharacterized papain-like cysteine protease rCrCEP1 from *Chlamydomonas*. Our results suggest that, in addition to its proteolytic function, rCrCEP1 also exhibits transpeptidase activity that enables it to cleave and ligate peptides to selected protein N-termini, both *in vitro* and in cellular extracts. While our exploration of the transpeptidation mechanism catalysed by rCrCEP1 provided valuable insights into this activity, several questions remained open.

First, we found it intriguing that the three Psb proteins appear to be the predominant substrates for rCrCEP1 transpeptidation in treated *Chlamydomonas* extracts. We considered the possibility that the enhanced labelling of the three proteins is a consequence of the observed increased abundance of Psb proteins in cell extracts following oxidative stress. It is well known that the oxygen evolving complex (OEC) of photosystem II is susceptible to damage under diverse abiotic stresses

[31], including light stress, high salinity, metal stress, and heat stress, and these stressors have been shown to disrupt the OEC, leading to the release of soluble PsbO, PsbP, and PsbQ (reviewed in [31–35]). We therefore hypothesise that the significant increase in labelling of Psb proteins during oxidative stress could be due to the release of OEC subunits, resulting from oxidative damage to the photosystem II.

However, while this hypothesis explains why labelling can mainly be observed in extracts of H_2O_2 -treated cells, it does not explain why no other high-abundance proteins were efficiently labelled in either the treated or untreated extracts. Transpeptidases usually require acceptor peptides to be accommodated by the S1' and S2' specificity pockets of the protease [36–38]. Considering that PICS analysis indicates that rCrCEP1 has relatively low specificity for amino acid residues located C-terminal to the cleavage bond, we would expect labelling of other proteins in the soluble proteome, as many N-termini would meet the specificity requirements of rCrCEP1. It remains plausible that either other factors influence the specificity/activity of rCrCEP1 in cell extracts or that other high abundance proteins contain modifications that prevent them from being suitable substrates for transpeptidation (including N-terminal modifications such as acetylation or the inflexibility of the N terminus). Nevertheless, the specificity of this reaction needs to be explored further in the future.

Our findings are also consistent with previous research on the granulin domain containing RD21A protease of *Arabidopsis thaliana*. It has been demonstrated that RD21A can perform transpeptidation using beta-lactone probes and peptides as donor molecules, leading to the labelling of N-termini of proteins in *Arabidopsis* extracts [39]. While RD21A labelled multiple proteins, only the OEC protein PsbP was efficiently captured on streptavidin beads and identified by MS. The correlation between our data and the findings presented by Wang et al., indicate that this characteristic is conserved among RD21A-like proteases. Notably, only RD21A was found to be responsible for labelling as it was only abolished in the knockout line of RD21A and not in knockout lines lacking other *Arabidopsis* PLCPs. Wang et al. raised a question about whether the granulin domain, as one of the defining features of RD21A, could have a role in transpeptidation [39]. However, our data suggest that rCrCEP1 can catalyse this reaction even in its mature form, i.e., without the granulin domain.

As both labelling experiments were performed on cell extracts, we

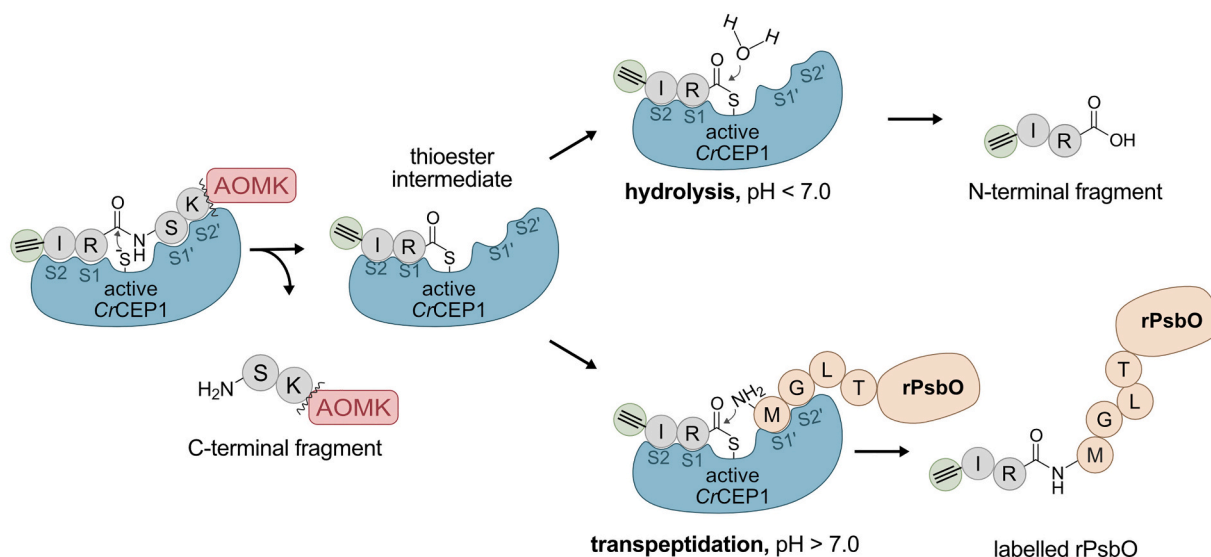


Fig. 6. Schematic representation of the proposed transpeptidation mechanism by which rCrCEP1 labels PsbO. The catalytic Cys of rCrCEP1 attacks the peptide bond within the IRSK peptide (with or without the AOMK group). A thioester intermediate with the N-terminal part of the probe forms, and the C-terminal part dissociates from the active site. At pH values below 7.0 hydrolysis is the favoured reaction, with a water molecule releasing the N-terminal fragment from the active site. At pH values above 7.0 the reaction of amino group results the formation of a novel peptide bond, labelling PsbO. The tetrapeptide is shown in grey, the alkyne group in green, the AOMK group in red, rCrCEP1 in blue, and rPsbO in sand.

cannot assume that AtRD21A or CrCEP1 execute this reaction in cells or that the OEC proteins are physiological substrates of RD21A-catalysed transpeptidation. Further investigation of the potential physiological implications of this reaction are warranted, as the characteristic imply that transpeptidation could occur *in vivo*.

The only requirements for transpeptidation are a physiological pH (between 7 and 8), available donor and acceptor peptides, and an active protease. Unlike conventional ligases, proteases can catalyse transpeptidation reactions without an additional energy source such as ATP, as the energy required for the reaction is obtained from the cleavage of a peptide bond [67]. Additionally, it has been demonstrated that macromolecular crowding, characterized by the high total concentration of background molecules, such as lipids, proteins, nucleic acids, and carbohydrates that are present in cells, shifts the equilibrium towards transpeptidation, since it is accompanied by a reduction in excluded volume [68].

For transpeptidation of the OEC proteins to occur in cells, rCrCEP1 would need to encounter them in the right environment. The OEC complex localizes to the thylakoid lumen, where pools of unassembled OEC proteins are also present [69]. Since rCrCEP1 has not yet been fully characterized, its subcellular location in *Chlamydomonas* remains unknown. Therefore, we can only speculate about its putative localization based on related plant proteases. Typically, mature RD21A-like proteases are found in acidic compartments such as the vacuole, lytic vesicles, or the apoplast [24,25,70], which primarily promote proteolytic rather than transpeptidase activity. However, exceptions exist. For instance, the protease RD21A from *Arabidopsis thaliana* was found to relocate to the nucleus and cytoplasm upon encountering a cyst nematode effector [71] and C14, the tomato ortholog of RD21, was observed to accumulate in the nuclei, cytoplasm, and even chloroplasts during drought stress [72]. Many of these compartments offer suitable pH conditions and potential contact points for transpeptidation reactions. Additionally, while the chloroplasts were long believed to be devoid of cysteine proteases, novel research suggests that PLCPs might be important regulators of photosynthetic gene expression and acclimation to high light stress [40].

The understanding that proteases can also catalyse the formation of peptide bonds is not a new concept, with the theoretical knowledge and first experimental synthesis of peptide bonds with proteases dating back to the late 1930s [41,42]. Still, transpeptidation is not universally considered as a function of proteases in cells. In the last few decades, however, an increasing number of proteases were found to catalyse peptide bond synthesis in cells. These include the animal proteasome, a threonine protease, which has been shown to use transpeptidation to produce peptides with higher immunogenicity for presentation by MHC (major histocompatibility complex) class I receptors (reviewed in [43]) and bacterial sortases, that use transpeptidation to bind proteins to the bacterial cell wall (reviewed in [44]). In addition, members of the plant asparaginyl endopeptidases (AEPs), also known as legumains or vacuolar processing enzymes (VPEs) use transpeptidation to produce cyclic polypeptides and cyclotides with antimicrobial and insecticidal properties (reviewed in [45]). Transpeptidation has therefore been proposed as a new form of post-translational modification [46].

Despite being the largest family of cysteine proteases, present in all forms of life [47], the exploration of peptide bond synthesis by PLCPs has been limited. Few PLCPs have been shown to catalyse transpeptidation *in vitro*, including cathepsin B [48] and papain [49], but *in vivo* reports are rare. Recently, cathepsin L was proposed to be responsible for the formation of chimeric MHCII autoantigen epitopes for diabetogenic CD4⁺ T cells *in vivo* [50].

It is possible that transpeptidation events in cells are more prevalent than previously believed but have been overlooked by conventional protease-monitoring techniques that focus on detecting proteolytic activity or substrate cleavage. Even advanced proteomic approaches can miss the annotation of transpeptidation products since they do not match the predicted proteome. Peptide-based probes could therefore be

a valuable tool for the detection and study of these reactions, helping to identify potential acceptor substrates of transpeptidation in the future. The dual nature of proteases highlighted in this study adds depth to our understanding of these enzymes and emphasizes the need to acknowledge and study their potential multifaceted roles in cells.

4. Materials and methods

4.1. Culture conditions, H₂O₂ treatment and soluble proteome preparation

The *Chlamydomonas reinhardtii* strain CC-5325 cw15 mt- (wild-type strain) and the mutant strain CC-4144 FUD44 mt⁺, lacking the PsbO gene, were obtained from the Chlamydomonas Resource Center, University of Minnesota, USA. Cells were grown in Tris-acetate-phosphate medium (TAP, pH 7), which was prepared as described by the Chlamydomonas Resource Center, on a gyratory shaker at 25 °C under continuous illumination with the Typ 2090118C lamp (Ritter Leuchten GmbH). For all experiments where cells were stressed by oxidative stress, $\sim 4 \times 10^6$ cells mL⁻¹ were exposed to 5 mM H₂O₂. Afterwards, the cells were incubated at 25 °C on a gyratory shaker and samples were collected at different time points (0, 0.5, 1, 2, 3 and 6 h). Growth of cell cultures after exposure to H₂O₂ was monitored turbidometrically at 740 nm using a UV-1600PC (VWR) spectrophotometer. For the preparation of soluble fractions, 25–100 mL of cell suspensions were collected at chosen time points, harvested by centrifugation (3 min, 3000 ×g) and stored at –80 °C until further processing. Samples were thawed on ice, mixed with 250–1000 µL of lysis buffer (50 mM Hepes, pH 7.5, 20 % glycerol, 0.3 % CHAPS, and 1 mM EDTA), and transferred to 1.5 mL microcentrifuge tubes. Samples were vortexed vigorously 10 × 20 s, centrifuged at 10,000 ×g for 10 min at 4 °C, and the supernatant representing the soluble proteome fraction was collected. The total protein concentration of each sample was determined by the Bradford assay [51].

4.2. Labelling of cell extracts and recombinant proteins with tetrapeptide probes

Labelling experiments were performed with various probes and peptides containing the alkyne minitag (≡) and different tetrapeptide sequences. Four probes contained the acylomethyl ketone warhead (referred to as the AOMK-probes). The synthesis details for the four AOMK-probes (DK10; DK11; DK12 and DK13) are described in our previous work [12]. Peptides ≡I and ≡IRS were synthesized as described in Section 4.16, respectively. The peptide ≡IRSK was purchased from CPC Scientific. For labelling reactions, typically 25 µL of soluble cell proteome containing about 20 µg of total protein was mixed with 25 µL of reaction buffer containing 100 mM Hepes, pH 8.0, 150 mM NaCl, 5 mM DTT (dithiothreitol). Profiling at different pHs was performed with the following reaction buffers: 100 mM acetate (pH 3.0–pH 5.5), 100 mM MES (pH 6.0–pH 6.5), 100 mM HEPES (pH 7.0–pH 8.0), 100 mM Tris (pH 8.5–pH 9.0), and 100 mM CAPS (pH 9.5–pH 11.0) supplemented with 150 mM NaCl and 5 mM DTT. For labelling of isolated proteins, 6 µg of protein was incubated in 50 µL of reaction buffer containing approximately 0.5 µg of active protease (either CrCEP1 or CrCEP2). The following proteins were used for the labelling experiments: recombinant PsbO from *C. reinhardtii* (rPsbO), recombinant PsbQ from *C. reinhardtii* (rPsbQ), and recombinant serpin-like protein from *C. reinhardtii* (rCrSerp). The expression and purification procedure for the latter is described in our previous publication [52]. Probes and peptides were typically added at a final concentration of 2 µM, and samples were incubated for 30 min to 1 h at 25 °C with gentle shaking. Proteins were then precipitated by addition of 200 µL ice-cold acetone and centrifuged at 13,000 ×g for 2 min. The acetone was removed, and 44 µL PBS containing 1 % SDS was added to the pellet. Samples were vortexed and incubated at 37 °C for 30 min. After incubation, the samples were heated

at 90 °C for 10 min and then cooled. The click reaction was performed by adding 2.5 µM picolyl-Cy5 (in dimethyl sulfoxide - DMSO), 1 mM CuSO₄, 2 mM TCEP (tris(2-carboxyethyl)phosphine), and 0.1 mM TBTA (tris((1-benzyl-4-triazolyl)methyl)amine). Samples were incubated for 1 h in the dark at 25 °C with gentle shaking. Proteins were then precipitated by addition of 200 µL ice-cold acetone and centrifuged at 13,000 ×g for 2 min. The pellets were resuspended in 25 µL 1 × SDS loading buffer containing 2 % β-mercaptoethanol and incubated at 37 °C for 30 min before loading 8 µL onto a 15 % SDS PAGE, which ran at 35 mA for approximately 1 h. Gels were scanned with ChemiDoc Imager (BioRad) using an ET700/50 emission filter and subsequently stained with Coomassie Brilliant Blue. To test different inhibitors, soluble proteomes were preincubated with DMSO (no inhibitor control) or 5 mM EDTA, (ethylenediaminetetraacetic acid), 100 µM PMSF, 10 µM E-64, and 0.1 mg/mL rCrSerp before labelling with 2 µM DK12 for 1 h. Competition assays with rPsbO peptides were performed by simultaneous addition of 2 µM DK12, 0.5–500 µM competitor peptide and 5 µM rPsbO and incubated for 1 h at 25 °C. For more information on the peptides, see the Section 4.17. Blocking of N-termini was performed by preincubation of soluble Chlamydomonas proteomes with 4 mM of freshly prepared sulfo-NHS-acetate (Thermo Scientific) for 30 min in 100 mM Hepes, pH 7.5, 150 mM NaCl, and 5 mM DTT before labelling.

4.3. Identification of DK12 labelled proteins by mass spectrometry

To identify the three labelled proteins a pull-down experiment using the Click&Go Click Chemistry Capture Kit (Click Chemistry Tools) and azide agarose resin (Jena Bioscience) was performed. Samples were prepared by mixing 3.5 mL of the treated soluble Chlamydomonas proteome (time point 3 h) with 3.5 mL of 100 mM Hepes, pH 8.0, 150 mM NaCl, 5 mM DTT. The sample was incubated with 2 µM DK12 for 30 min at 25 °C. In the control sample, DMSO was added instead of the probe in the same volume. The reaction was stopped by adding 28 mL of ice-cold acetone, and the samples were incubated at –20 °C for 2 h to ensure precipitation of the proteins. Samples were centrifuged at 15,000 ×g at 4 °C for 10 min. The pellet was additionally washed with acetone to ensure that the unbound DK12 probe was removed. The pellet was resuspended in 800 µL of Click & Go lysis buffer (200 mM Tris, 4 % CHAPS, 1 M NaCl, 8 M urea, pH 8.0). Click chemistry reaction, reduction, alkylation and washing steps were performed according to the manufacturer's instructions. Digestion of the washed agarose beads was performed with trypsin and the released peptides were desalted and analysed using Liquid Chromatography tandem-mass spectrometry (LC-MS/MS). Measurements were performed using an EASY-nanoLC II HPLC unit (Thermo Scientific) coupled to an Orbitrap LTQ Velos MS (Thermo Scientific). The peptide sample was first loaded onto a C18 trapping column (Proxeon EASY -Column™, 2 cm (length), 100 µm internal diameter, 5 µm 120 Å, C18-A1 beads) and then separated on a 10 cm long C18 PicoFrit™ AQUASIL analytical column (75 µm internal diameter, 5 µm 100 Å, C18 beads) (New Objective) using forward flushing. Peptides were eluted with a 90-min linear gradient of 5–50 % solvent B (0.1 % formic acid in acetonitrile) at a flow rate of 300 nL/min. MS spectra were acquired in the Orbitrap analyzer with a mass range of 300–2000 *m/z* and 30,000 resolution. MS/MS spectra were obtained by HCD fragmentation (normalised collision energy at 35) of the nine most intense precursor ions from the full MS scan. Dynamic exclusion was enabled with a repeat count of 2 and 120 s exclusion time. Database searches and quantification by spectral counting were performed using Mascot software version 2.6 and Scaffold software version 5.1.0. Searches were performed using the NCBI protein database, using trypsin cleavage specificity with a maximum of 2 missed cleavages. Carbamidomethylation of cysteines was set as a static modification, whereas methionine oxidation was set as a dynamic modification. Tolerances for precursor and fragment mass were set at 6 and 20 ppm, respectively. A reverse database search was performed, and the false discovery rate (FDR) was set at 1 % for peptide and protein identifications.

4.4. Molecular cloning of protein constructs

The genes encoding the mature forms of PsbO and PsbQ and the full-length forms of CrCEP1 and CrCEP2 lacking only the signal peptide, codon-optimised for expression in *Escherichia coli* (*E. coli*), were ordered from Twist Bioscience as DNA fragments containing corresponding restriction sites at the 5' and 3' ends of the genes, respectively. The genes encoding PsbO and PsbQ or shortened versions of PsbO (Δ11-rPsbO and Δ18-rPsbO) were ligated into the expression vector pET-28b(+) using the restriction sites *NcoI* and *XhoI*, and the genes encoding CrCEP1 and CrCEP2 were ligated into the expression vector pET-32/28 using the restriction sites *NheI* and *XhoI*, as described previously for cysteine cathepsins [53]. The rPsbO and PsbQ constructs contained a C-terminal hexahistidine tag, and the rCrCEP1 and rCrCEP2 constructs contained an N-terminal hexahistidine tag. The correct nucleotide sequences were verified by DNA sequencing. The amino acid sequences of all recombinant proteins are listed in the supplemental information (Supplemental Fig. S1).

4.5. Expression and purification of rPsbO and rPsbQ from the insoluble fraction

E. coli BL21(DE3) bacteria were transformed with the expression vector and grown into overnight cultures. Expression of protein was performed in 4 × 400 mL of autoinduction medium containing 50 µg/mL kanamycin (prepared as described (Studier, 2005)), which was inoculated with 4 × 400 µL of overnight bacterial culture. The cells were grown at 37 °C for 8 h. The cell pellet was then collected, each pellet was resuspended in 20 mL of resuspension buffer (1 × PBS, 5 mM EDTA), and frozen at –80 °C until further processing. The bacterial cells were thawed and sonicated for 5 × 1 min (80 % power) on ice. The lysed cells were centrifuged at 10,000 ×g for 30 min to pellet the insoluble fraction. The pellet was first resuspended in 5 mL of resuspension buffer and diluted with 40 mL of wash buffer (1 × PBS, 5 mM EDTA, 25 % Sucrose, 1 % Triton-X100). The washing step was repeated, and the washed pellet was resuspended in binding buffer (50 mM phosphate buffer, pH 7.5, 500 mM NaCl, 8 M urea, and 20 mM imidazole) and shaken overnight at 37 °C. The dissolved pellet was centrifuged at 20,000 ×g for 10 min, the supernatant filtered, and loaded onto a HisTrap™ FF column (Cytiva). After washing with binding buffer, bound proteins were eluted in 50 mM phosphate buffer, pH 7.5, 500 mM NaCl, 8 M urea, and 500 mM imidazole. Peak fractions were collected and dialyzed overnight against 20 mM Hepes, pH 7.0, 150 mM NaCl. The refolded protein solution was centrifuged to remove all insoluble debris and concentrated to approximately 5 mg/mL using an Amicon filtration unit (Millipore Corp., Temecula, California) equipped with a 10 kDa exclusion membrane. The sample was applied to a Superdex 75 size exclusion chromatography column (GE Healthcare Life Sciences, Marlborough, Massachusetts) connected to an ÄKTA FPLC system. The column was equilibrated in 20 mM HEPES pH 7.5, 500 mM NaCl, and a flow rate of 0.5 mL/min was used for protein separation. Selected fractions containing only the monomeric protein were collected and stored at –80 °C until use.

4.6. Circular dichroism spectroscopy

Circular dichroism spectra were recorded with a J-1500 Circular Dichroism Spectrometer (Jasco, MD). Circular dichroism spectra of proteins at a concentration of 1 mg/mL in phosphate buffer were recorded at 25 °C using a quartz cuvette with a 1 mm path length. Data were acquired between 250 and 200 nm with a sampling rate of 20 nm/min, a bandwidth of 1 nm, and a data spacing of 1 nm. The obtained data was converted into mean residue ellipticity (MRE, θ_{MR}) given by the following equation:

$$\theta_{MR} [deg\,cm^2\,dmol^{-1}] = \frac{\theta \times 0.1}{N_p \times P_t \times l}$$

N_p in the equation represents the number of amino acids in the sequence ($N_{\text{rPsbO}} = 250$ aa). P_i is the total monomer protein/peptide concentration and l is the optical pathway length. The theoretical CD spectra were calculated with PDB2CD [54] based on the crystal structure of Chlamydomonas PsbO and PsbQ (PDB: 6KAC, entities 14 and 16) and secondary structure content for PsbO, rPsbO, PsbQ and rPsbQ were estimated with BeStSel [55].

4.7. Immunoblotting and labelling with DCG-04

For immunoblotting, equal amounts of soluble Chlamydomonas proteomes (typically 10 µg) were mixed with 4 × SDS-PAGE loading buffer containing β-mercaptoethanol and boiled for 5 min before loading onto 12 % TGX StainFree™ FastCast™ polyacrylamide gels (Biorad) and transferring to polyvinylidene fluoride (PVDF) membranes (Biorad). The following antibodies were used: polyclonal rabbit anti-PsbO (1:1000, Agrisera), goat anti-rabbit HRP (1:10,000, Abcam), streptavidin-HRP (1:1000, GE Healthcare), anti-His6-HRP (1:1000, Roche).

DCG-04 labelling was performed with 50 µL of soluble Chlamydomonas proteomes, containing approximately 30 µg protein (prepared as described in Section 4.1). The samples were mixed with 3 µL 3 M NaOAc pH 5.5 to lower the pH of the solution and 1 M DTT was added to a final concentration of 10 mM. Control samples were preincubated with 10 µM E-64 for 30 min before labelling with 2 µM DCG-04 for 2 h at 25 °C. Proteins were then precipitated by adding 200 µL ice-cold acetone and centrifuged for 2 min at 13,000 ×g. The pellet was dissolved in 1 × SDS-PAGE loading buffer containing β-mercaptoethanol, boiled, loaded onto an SDS-PAGE gel and blotted as described above.

4.8. DCG-04 pull-down sample preparation

DCG-04 pull-down was performed on four biological replicates of soluble proteomes from H₂O₂-treated Chlamydomonas cells (time point 3 h after treatment). For each replicate, 1 mL of the soluble proteome was prepared in lysis buffer containing approximately 1 mg of protein as described in the Section 4.1. The lysate was supplemented with a final concentration of 100 mM NaOAc, pH 5.0 (to lower the pH of the solution), 150 mM NaCl, 0.1 mM PMSF and 10 mM DTT, respectively. The mixture was then separated into two parts. One was pre-incubated with 200 µM E-64 (control sample) for 1 h at 25 °C, while an equivalent amount of DMSO (DCG-04 sample) was added to the other. Both samples were then incubated with 20 µM DCG-04 for 3 h at 25 °C. The buffer was then exchanged for 50 mM Tris-HCl pH 7.5, 150 mM NaCl (TBS buffer) using NAP-5 columns (Cytiva). The Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Scientific) and 90 µL washed Streptavidin Mag Sepharose beads (Cytiva) were added to each sample and rotated overnight at 4 °C. Consecutively, beads were washed three times with 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 2 M urea according to the manufacturer's instructions and resuspended in 20 mM Hepes pH 7.5 and 150 mM NaCl. MS-grade trypsin (Serva Electrophoresis GmbH) was added at a ratio of 1:100 (enzyme:proteome) and incubated overnight at 37 °C for on-bead digestion. Eluted peptides were desalted using SDB-RP StageTips and analysed by LC-MS/MS as described in the Section 4.13.

4.9. Expression and purification of rCrCEP1 and rCrCEP2

For CrCEP1, *E. coli* strain C41 (DE3) (BioCat) was transformed with the expression vector pET-32/28 (+). Expression was performed in 8 × 400 mL of autoinduction medium containing 100 µg/mL ampicillin (prepared as described by Studier (Studier, 2005)), which was inoculated with 8 × 400 µL of an overnight bacterial culture. The cells were grown at 37 °C for 4 h and brought to 16 °C overnight. For rCrCEP2, the *E. coli* strain Rosetta Gammi 2 (DE3) pLysS (Sigma Aldrich) was used. Expression was performed in 8 × 400 mL of LB medium (lysogeny broth) containing 100 µg/mL ampicillin. The medium was inoculated with 10 mL overnight bacterial culture and grown at 37 °C until the culture

reached an optical density of 0.6. Expression of genes was then induced with the addition of 1 mM IPTG (isopropyl β- d-1-thiogalactopyranoside), and the flasks were brought to 16 °C overnight. The cells of both expressions were pelleted, and pellets obtained from 400 mL bacterial culture were resuspended in 20 mL resuspension buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 20 mM imidazole) and sonicated on ice for 5 × 1 min (80 % power). After centrifugation at 25,000 ×g for 10 min to remove insoluble debris, the supernatant was briefly sonicated again (1 × 1 min) and centrifuged at 25,000 ×g for 15 min. The supernatant was applied to a HisTrap FF column (Cytiva). After washing with resuspension buffer, bound proteins were eluted in the same buffer, containing 500 mM imidazole. Peak fractions were collected, concentrated using an Amicon filtration unit equipped with a 10-kDa exclusion membrane, and applied to a Superdex 75-size exclusion chromatography column (GE Healthcare Life Sciences) connected to an ÄKTA FPLC system (GE Healthcare/Amersham Biosciences). The column was equilibrated in 20 mM HEPES, pH 7.5, 500 mM NaCl, and a flow rate of 0.5 mL/min was used to separate the proteins. Peak proteins were collected and stored at −80 °C until further use.

4.10. Activation and kinetic assays for rCrCEP1 and rCrCEP2

50 µL aliquots of enzyme were activated by adjusting the pH to 4.0 with 3 M NaOAc, pH 3.8. DTT was added to a final concentration of 5 mM to prevent oxidation of the catalytic cysteine and the sample was incubated shaking at 25 °C and 900 rpm. The proteolytic activity was determined by monitoring the release of the fluorescent group AMC (7-amino-4-methylcoumarin) from the substrate Z-FR-AMC (Bachem) at excitation and emission wavelengths of 383 nm and 455 nm, respectively, using a FL 6500 fluorescence spectrometer (PerkinElmer) in cuvettes with a 1 cm path length. The reaction buffer consisted of 100 mM acetate pH 5.5, 150 mM NaCl, 1 mM EDTA, 5 µM Z-FR-AMC, 5 mM DTT. When peak activity was reached, pH was adjusted to 5.0 with the addition of 3 M NaOAc, pH 5.5, and the active enzyme was applied to a Superdex 75-size exclusion chromatography column (GE Healthcare Life Sciences) connected to an ÄKTA FPLC (system GE Healthcare/Amersham Biosciences), to remove aggregates and cleaved peptides. The column was equilibrated in 100 mM NaOAc, pH 5.0, 500 mM NaCl, and a flow rate of 0.5 mL/min was used to separate the proteins. Peak proteins were collected, aliquoted and stored at −80 °C until further use. During activation samples were also collected for analysis with SDS-PAGE. At each time point, approximately 5 µg of protein was collected and inhibited with 20 µM E-64 for 10 min before adding 4 × SDS-PAGE loading buffer containing β-mercaptoethanol. Samples were heated to 100 °C for 5 min, separated on a 12 % SDS-PAGE and stained with Coomassie Brilliant Blue. For the pH profile analysis, a universal buffer containing 20 mM NaOAc, 20 mM MES and 20 mM HEPES in a pH range of 3–9 was used as described by D. Brooke et al. [56]. The buffer was supplemented with 150 mM NaCl and 5 mM DTT. Reactions were performed with 10 nM enzyme and 5 µM Z-FR-AMC and the proteolytic activity was monitoring with a FL 6500 fluorescence spectrometer (PerkinElmer) as described above. The enzyme concentration is based on protein assuming purity. For the inhibition test, an aliquot of the protease was preincubated with either 2 µM DK12, 2 µM E-64 or DMSO for 30 min before measuring its fluorescence as described above. All data were analysed using GraphPad Prism 9.1.1 (GraphPad Software Inc.).

4.11. Specificity profiling of rCrCEP1

To profile rCrCEP1 sequence specificity, we used the method Proteomic Identification of protease Cleavage Sites (method PICS), as described [27]. Proteome-derived peptide libraries were prepared by digestion of *E. coli* DH5α lysates with trypsin (cleavage at K/R[X], GluC (cleavage at D/E[X]), or legumain (cleavage at D/N[X], [57]). Peptide libraries (20 µg) were incubated with CrCEP1 (0.02 µg, ratio: 1:1000) in a solution of 100 mM sodium acetate (pH 5.0), 150 mM NaCl, and 5 mM

DTT for 30 min at 25 °C. For the control sample, the protease was pre-incubated with 20 µM E-64 for one hour before being incubated with 20 µg of the peptide library. Afterwards, E-64 was added to a final concentration of 20 µM to both samples before being differentially labelled for 16 h with 30 mM formaldehyde and 30 mM NaCNBH₃. Either ¹²CH₂O (“light”) or ¹³CD₂O (“heavy”) were used for the control and protease-treated samples, respectively. To ensure a complete reaction, the same concentration of reagents was added for an additional 2 h. The reactions were quenched with 0.1 M Tris pH 7.4 at 37 °C for 1 h and pooled in a 1:1 ratio. Peptides were desalted using SDB-RP StageTips and analysed by liquid chromatography-tandem mass spectrometry (see Section 4.13 for details).

4.12. In-gel digestion of labelled rPsbO and rPsbQ

rPsbO or rPsbQ (10 µg) were mixed with 2.5 µg of active rCrCEP1 in 100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM DTT, and labelled with 2 mM \equiv IRSK for 30 min at 25 °C. The samples were separated into 4 separate lanes on a 15 % SDS-PAGE gel and stained with Coomassie Brilliant Blue for 1 h followed by destaining in 30 % ethanol and 10 % acetic acid. The four adjacent bands were combined, cut into ~1 mm³ pieces and incubated for 10 min at 25 °C in 500 µL fresh 50 mM ammonium bicarbonate (ABC). The gel pieces were then destained for 10 min at 25 °C and 900 rpm in 500 µL 50 % ACN (acetonitrile) in 25 mM ABC. Subsequently, the cysteine residues were reduced with 500 µL 10 mM DTT in 25 mM ABC for 30 min at 56 °C. The gel pieces were then cooled, and alkylation was performed with 500 µL 10 mM CAA (chloroacetamide) in 25 mM ABC in the dark for 30 min at 25 °C. The gel pieces were then washed with 500 µL 50 mM ABC and 500 µL 50 % ACN in 25 mM ABC, both for 10 min at 25 °C. Samples were partially dehydrated by adding 500 µL 100 % ACN for 5 min. ACN was removed and the gel pieces were dehydrated completely in the SpeedVac at 50 °C for 10–20 min and stored at –20 °C until digestion. For digestion, the gel pieces were covered completely with enough volume of 100 mM MES pH 5.0, 2 mM DTT, and 5 mM CaCl₂ and each was added 5 µg of activated human legumain [57]. Samples were desalted using SDB-RP StageTips and LC-MS/MS measurements were performed as described in the Section 4.13.

4.13. nanoLC-MS/MS measurements

Mass spectrometry measurements were performed on a nano-HPLC (Ultimate 3000 nano-RSLC) in a two-column setup (µPAC pillar array trap column, 1 cm length, and a µPAC pillar array analytical column of 50 cm length; PharmaFluidics) with a binary gradient from 1 to 30 % B for 90 min (DCG-04 samples) or 43 min (PICS samples) (A, 0.1 % FA; B, 0.1 % FA in acetonitrile) and a total runtime of 2 or 1 h per sample coupled to a high-resolution Q-TOF mass spectrometer (Impact II, Bruker). Data was acquired with the HyStar Software, v5.1 (Bruker Daltonics) in line-mode in a mass range from 200 to 1750 *m/z* at an acquisition rate of 5 Hz. The top 14 most intense ions were selected for fragmentation with a dynamic exclusion of previously selected precursors for the next 30 s, unless an intensity increase of factor 3 compared to the previous precursor spectrum was observed. Intensity-dependent fragmentation spectra were acquired between 5 Hz, for low intensity precursor ions (>500 cts), and 20 Hz, for high intensity (>25 k cts) spectra.

4.14. Mass spectrometry data analysis

Database searches were performed with the Maxquant software version v2.0.3.0 [58] using standard settings for Bruker Q-TOF instruments. For the DCG-04 pulldown experiment tandem mass spectra were aligned with peptide sequences from the Chlamydomonas Phytozome Proteome Database (<https://phytozome-next.jgi.doe.gov/>, genome: CC-4532 v6.1, 16,883 protein coding genes). N-terminal

acetylation and methionine oxidation of proteins were considered as variable modifications with a maximum of five variable modifications per peptide, and the “match between runs” option was enabled. The minimum score for recalibration was set to 47. Label-free quantification (LFQ) values were calculated with a minimum number of ratios set to 2. The digestion mode was set to specific trypsin. For downstream analysis of the data, the resulting MaxQuant output file “proteinGroups.txt” was loaded into Perseus software version v2.0.7.0 [59]. The potential contaminants as well as reverse hits were removed and the Phytozome annotation file for Chlamydomonas CC-4532 v6.1 was added. LFQ values were log₂(x) transformed, and protein groups were filtered to include only those with valid LFQ values in all four replicates in at least one group. For the in-gel-digest analysis of modified rPsbO and rPsbQ, tandem mass spectra were matched to peptide sequences derived from the rPsbO or rPsbQ sequence, the rCrCEP1 sequence, and the *E. coli* Uniprot proteome database (www.uniprot.org; UP000000625.83333, release 2015/11, 4305 entries). N-terminal acetylation, methionine oxidation, and the N-terminal alkynylation (\equiv IR N-term, composition C₂₃H₃₈N₆O₄) were considered as variable modifications with a maximum of five variable modifications per peptide allowed. The re-quantification feature of MaxQuant was disabled. Digestion protease was set to legumain. The MaxQuant output file “ModificationSpecificPeptides.txt” was examined to see which peptides were modified. The MS/MS spectra of the modified peptides were visualized using MaxQuant v2.0.3.0 [58]. For the PICS analysis, tandem mass spectra were matched to peptide sequences derived from the *E. coli* Uniprot proteome database (www.uniprot.org; UP000000625.83333, release 2015/11, 4305 entries). Carbamidomethylation of cysteines was set as a fixed modification and multiplicity was set to 2, considering dimethylation of lysine and peptide N-termini residues as labels (+28.0313 Da for the light channel, +34.0631 Da for the heavy channel). The re-quantification feature was disabled. Trypsin, GluC, or legumain were set as digestion proteases depending on the peptide library used, and up to two missed cleavages were allowed. Protein N-terminal acetylation and methionine oxidation were considered as variable modifications with a maximum of three variable modifications per peptide. The MaxQuant output file “peptides.txt” was analysed using the pincis.pl script (<https://pincis.sourceforge.io>) as described [27]. The PINCIS output file “Specific_Cleavage_Windows.tab file” was used for cleavage site analysis with the iceLogo software [60].

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [61] via the PRIDE [62] partner repository with the dataset identifiers PXD047220 reviewer_pxd047220@ebi.ac.uk, PXD047358 reviewer_pxd047358@ebi.ac.uk and PXD047370 reviewer_pxd047370@ebi.ac.uk.

4.15. Chemicals and general methods for the synthesis of \equiv I and \equiv IRS probes

All solvents and chemicals were used as received. Melting points were determined on a Kofler micro hot stage Leica Galen III (Leica Camera AG) and on a Mettler Toledo Melting Point System MP 30 (Mettler Toledo). The ¹H NMR and ¹³C NMR spectra were recorded in CDCl₃, DMSO-*d*₆, and CD₃OD as solvents using Me₄Si as the internal standard on a Bruker Avance III UltraShield 500 plus and on a Bruker Avance NEO 600 instruments (Bruker) at 500 and 600 MHz for ¹H and at 126 and 151 MHz for ¹³C nucleus, respectively. IR spectra were recorded on a Bruker FTIR Alpha Platinum spectrophotometer (Bruker). Mass spectra were recorded on an Agilent 6224 Accurate Mass TOF LC/MS chromatograph/spectrometer (Agilent Technologies). Microanalyses were performed by combustion analysis on a Perkin-Elmer CHNS Analyzer 2400 II (PerkinElmer). Column chromatography (CC) was performed on silica gel (Silica gel 60, particle size: 0.035–0.070 mm, Sigma-Aldrich). Catalytic hydrogenation was performed on a Parr shaker hydrogenation apparatus (Moline) at 3 bar of hydrogen in a 250 mL hydrogenation bottle. (S)-5-Amino-2-(benzyloxycarbonylamino)

pentanoic acid, *N,N'*-bis(*tert*-butoxycarbonyl)-*S*-methylisourea, *L*-serine methyl ester hydrochloride, *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC), *N*-benzyloxycarbonyl-*L*-isoleucine, pent-4-ynoic acid, 6-aminocaproic acid, 4-methylmorpholine (NMM), *N*, *N'*-carbonyldiimidazole (CDI), and trifluoroacetic acid (TFA) are commercially available (Sigma-Aldrich). *N*²-benzyloxycarbonyl-*N'*^ω,*N''*^ω-bis(*tert*-butoxycarbonyl)-*L*-arginine was prepared following the literature procedure [63].

4.16. Synthesis of \equiv I and \equiv IRS probes

The synthesis of both probes was carried out according to general solution-phase peptide synthesis procedures, following Boc- and Cbz-strategies and employing CDI and EDC as activating reagents in amidation reactions [64,65]. CDI was used as activation reagent in amidation of pent-4-ynoic acid with 6-aminocaproic acid, while EDC was used as activating reagent in all other amidation reactions. The Cbz protecting group was removed by catalytic hydrogenation, while the Boc protecting group was removed acidolytically by treatment with TFA. All synthetic steps were performed in standard glass flasks followed by standard isolation and purification procedures. 6-(Pent-4-ynamido)hexanoic acid (PAH-OH)¹ was prepared by amidation of pent-4-ynoic acid (1.055 g, 5 mmol) with 6-aminocaproic acid. Yield: 880 mg (4.17 mmol, 83 %), ¹H NMR spectral data are consistent with the literature data [66]. Methyl *N*²-benzyloxycarbonyl-*N'*^ω,*N''*^ω-bis(*tert*-butoxycarbonyl)-*L*-arginyl-*L*-serinate (Cbz-IR(Boc₂)S-OMe) was prepared in three steps from *N*²-benzyloxycarbonyl-*N'*^ω,*N''*^ω-bis(*tert*-butoxycarbonyl)-*L*-arginine (Cbz-R(Boc₂)-OH) via coupling with *L*-serine methyl ester, hydrogenolytic *N*-Cbz-deprotection, and coupling of HCl•R(Boc₂)S-OMe (501 mg, 0.98 mmol) with Cbz-*L*-isoleucine. Yield (last step): 624 mg (0.86 mmol, 88 %), HRMS (ESI): *m/z* = calculated for C₃₄H₅₅N₆O₁₁ [MH]⁺ 723.3924, found 723.3920. \equiv I probe was prepared by amidation of 6-(pent-4-ynamido)hexanoic acid (106 mg, 0.5 mmol) with *L*-isoleucine methyl ester. Yield: 142 mg (0.42 mmol, 84 %), HRMS (ESI): *m/z* = calculated for C₁₈H₃₁N₂O₄ [MH]⁺ 339.2278, found 339.2278. \equiv IRS probe was prepared as trifluoroacetate salt in three steps from Cbz-IR(Boc₂)S-OMe via hydrogenolytic *N*-Cbz-deprotection, coupling with 6-(pent-4-ynamido)hexanoic acid (PAH-OH), and acidolytic *N'*^ω,*N''*^ω-Boc-deprotection of PAH-IR(Boc₂)S-OMe (53 mg, 68 μmol). Yield (last step): 46 mg (66 μmol, 97 %), HRMS (ESI): *m/z* = calculated for C₂₇H₄₈N₇O₇ [MH]⁺ 582.3610, found 582.3599. More detailed procedures and compound characterizations can be found in the Supplemental Materials online.

4.17. Synthesis of synthetic peptides

Chemicals and solvents were purchased from commercial sources (FluoroChem, SigmaAldrich, Fluka, Novabiochem and BLDpharm) and were used as obtained unless otherwise noted. Amine trapping packets were added to the DMF. The following amino acids were used for the peptide synthesis: Fmoc-Ala-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Gly-OH, Fmoc-Ile-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH, Fmoc-Ser(tBu)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Val-OH. The following peptides were synthesized in this study: IRSK, Pep16: GLTFDEIQGLTYLQV, Pep20: GLTFDEIQGLTYLQVKGSG. All peptides were synthesized manually using a Fmoc strategy in a fritted 8-mL plastic syringe. Peptide synthesis proceeded from the C- to the N-terminus and was performed using the Grace Alltech vacuum manifold system. *Resin preparation*: peptides were synthesized using Rink amide resin (0.4–0.8 mmol/g). 300 mg of the resin was transferred to a syringe, DMF was added to half of the syringe, and the suspension was allowed to stand for 1 h. DMF was removed under *vacuum* and piperidine (20 % v/v in DMF) was added to half of the syringe and allowed to stand for 1 h. After 1 h, the piperidine was removed under *vacuum* and the resin was washed three times with DMF. *Coupling reactions*: A solution of amino

acid (0.735 mmol; 5 eq), HATU reagent (0.38 M, dissolved in DMF; 4.5 eq; 1.74 mL), and DIEA (7 eq; 179 μL) was prepared in a penicillin vial. The solution was added to the syringe and allowed to stand for 15 min with occasional mixing. After 15 min, the solution was removed under *vacuum* and the resin was washed three times with DMF. The Fmoc protection of the amino acid was then removed with piperidine DMF solution (20 % v/v in DMF, 2 × 5 min). Before the next coupling, the resin was washed five times with DMF. Amino acids were incorporated into the peptide sequence in the appropriate order following the described procedure. After completion of the synthesis, the peptide was washed three times with dichloromethane and dried for 1 h *in vacuo*. *Cleavage from the solid support*: cleavage of the synthesized peptide from the solid support was performed using a mixture of TFA, H₂O and TIS (95 %: 2.5 %: 2.5 %). The prepared mixture (5 mL) was added to the syringe containing the peptide and left for 2 h with constant shaking. It was then filtered through a frit, the filtrate was collected in 15 mL Falcon™ tube, and the TFA was evaporated in a flow of nitrogen. The crude product was triturated with cold diethyl ether (approx. 10 mL), which was added dropwise to form a white precipitate. The resulting mixture was centrifuged (7000 rpm, 5 min), after which the ether was decanted. The described procedure was repeated two more times and the obtained product was dried under *vacuum*. The crude product obtained was further purified by reverse phase chromatography. Purified peptides were analysed by HPLC and HRMS.

4.18. Analysis of IRSK peptide cleavage by mass spectrometry

200 μg of the synthetic peptide IRSK (prepared as described in Section 4.17) was incubated with 10 nM rCrCEP1 in 100 mM NaOAc, pH 5.0, 150 mM NaCl, 5 mM DTT for 1 h at 25 °C in a final volume of 200 μL, along with control samples lacking either rCrCEP1 or IRSK. Samples were applied to a 3 kDa cutoff spin column to remove rCrCEP1, and flow through was analysed by MS. Mass spectra were recorded on an Agilent 6224 Accurate Mass TOF LC/MS chromatograph/spectrometer (Agilent Technologies, Santa Clara, CA, USA).

CRedit authorship contribution statement

Katarina P. van Midden: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Melissa Mantz**: Writing – review & editing, Data curation. **Marko Fonović**: Writing – review & editing, Data curation. **Martin Gazvoda**: Writing – review & editing, Data curation. **Jurij Svete**: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **Pitter F. Huesgen**: Writing – review & editing, Supervision, Data curation. **Renier A.L. van der Hoorn**: Writing – review & editing, Writing – original draft, Supervision, Data curation, Conceptualization. **Marina Klemenčič**: Writing – original draft, Validation, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the financial support from the Slovenian Research Agency core funding programmes P1-0179 and P1-023 and research project No. J4-2550. This work was supported by an STSM grant to K.P.v.M. from the COST Action ProteoCure, CA20113, supported by COST (European Cooperation in Science and Technology). This work was in part supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Project-ID 414786233 – SFB 1403,

to P.F.H.). The authors acknowledge help from Zala Živič for help with CD spectra recordings and the help from Lara Bartol, Urban Barbič and Aljaž Renko with the preparation of synthetic peptides. Additionally, we thank Marko Novinec for the pET-32/28(+) vector.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2024.132505>.

References

- [1] R.A.L. Van Der Hoorn, Plant proteases: from phenotypes to molecular mechanisms, *Annu. Rev. Plant Biol.* 59 (1) (Jun. 2008) 191–223, <https://doi.org/10.1146/annurev.arplant.59.032607.092835>.
- [2] A.V. Balakireva, A.A. Zamyatin, Indispensable role of proteases in plant innate immunity, *Int. J. Mol. Sci.* 19, no. 2, Art. no. 2 (Feb. 2018), <https://doi.org/10.3390/ijms19020629>.
- [3] S. Stael, F. Van Breusegem, K. Gevaert, M.K. Nowack, Plant proteases and programmed cell death, *J. Exp. Bot.* 70 (7) (Apr. 2019) 1991–1995, <https://doi.org/10.1093/jxb/erz126>.
- [4] R.A. Buono, R. Hudecek, M.K. Nowack, Plant proteases during developmental programmed cell death, *J. Exp. Bot.* 70 (7) (Apr. 2019) 2097–2112, <https://doi.org/10.1093/jxb/erz072>.
- [5] A.V. Balakireva, A.A. Zamyatin, Cutting out the gaps between proteases and programmed cell death, *Front. Plant Sci.* 10 (Jun. 2019) 704, <https://doi.org/10.3389/fpls.2019.00704>.
- [6] P. Sharma, D. Gayen, Plant protease as regulator and signaling molecule for enhancing environmental stress-tolerance, *Plant Cell Rep.* 40 (11) (Nov. 2021) 2081–2095, <https://doi.org/10.1007/s00299-021-02739-9>.
- [7] D.J. Sueldo, R.A.L. van der Hoorn, Plant life needs cell death, but does plant cell death need Cys proteases? *FEBS J.* 284 (10) (May 2017) 1577–1585, <https://doi.org/10.1111/febs.14034>.
- [8] R.A.L. van der Hoorn, M. Klemenčič, Plant proteases: from molecular mechanisms to functions in development and immunity, *J. Exp. Bot.* 72 (9) (Apr. 2021) 3337–3339, <https://doi.org/10.1093/jxb/erab129>.
- [9] K. Morimoto, R.A.L. van der Hoorn, The increasing impact of activity-based protein profiling in plant science, *Plant Cell Physiol.* 57 (3) (Mar. 2016) 446–461, <https://doi.org/10.1093/pcp/pcw003>.
- [10] J.C. Misas-Villamil, et al., Activity profiling of vacuolar processing enzymes reveals a role for VPE during oomycete infection, *Plant J.* 73 (4) (2013) 689–700, <https://doi.org/10.1111/tpj.12062>.
- [11] R.A.L. van der Hoorn, M.A. Leeuwenburgh, M. Bogoy, M.H.A.J. Joosten, S.C. Peck, Activity profiling of papain-like cysteine proteases in plants, *Plant Physiol.* 135 (3) (Jul. 2004) 1170–1178, <https://doi.org/10.1104/pp.104.041467>.
- [12] V. Strancar, et al., Activity-based probes trap early active intermediates during metacaspase activation, *iScience* 25 (11) (Nov. 2022) 105247, <https://doi.org/10.1016/j.isci.2022.105247>.
- [13] Y. Zou, P.V. Bozhkov, Chlamydomonas proteases: classification, phylogeny, and molecular mechanisms, *J. Exp. Bot.* 72 (22) (Dec. 2021) 7680–7693, <https://doi.org/10.1093/jxb/erab383>.
- [14] S.L. Vavilala, K.K. Gawde, M. Sinha, J.S. D'Souza, Programmed cell death is induced by hydrogen peroxide but not by excessive ionic stress of sodium chloride in the unicellular green alga *Chlamydomonas reinhardtii*, *Eur. J. Phycol.* 50 (4) (Oct. 2015) 422–438, <https://doi.org/10.1080/09670262.2015.1070437>.
- [15] O. Murik, A. Kaplan, Paradoxically, prior acquisition of antioxidant activity enhances oxidative stress-induced cell death, *Environ. Microbiol.* 11 (9) (Sep. 2009) 2301–2309, <https://doi.org/10.1111/j.1462-2920.2009.01957.x>.
- [16] O. Murik, A. Elboher, A. Kaplan, Dehydroascorbate: a possible surveillance molecule of oxidative stress and programmed cell death in the green alga *Chlamydomonas reinhardtii*, *New Phytol.* 202 (2) (Apr. 2014) 471–484, <https://doi.org/10.1111/nph.12649>.
- [17] I. Enami, A. Okumura, R. Nagao, T. Suzuki, M. Iwai, J.-R. Shen, Structures and functions of the extrinsic proteins of photosystem II from different species, *Photosynth. Res.* 98 (1) (Oct. 2008) 349–363, <https://doi.org/10.1007/s11120-008-9343-9>.
- [18] K. Ifuku, T. Noguchi, Structural coupling of extrinsic proteins with the oxygen-evolving center in photosystem II, *Front. Plant Sci.* 7 (Feb. 2016) 84, <https://doi.org/10.3389/fpls.2016.00084>.
- [19] V.V. Terentyev, A.K. Shukshina, A.A. Ashikhmin, K.G. Tikhonov, A.V. Shitov, The main structural and functional characteristics of photosystem-II-enriched membranes isolated from wild type and cia3 mutant *Chlamydomonas reinhardtii*, *Life (Basel)* 10 (5) (May 2020) 63, <https://doi.org/10.3390/life10050063>.
- [20] N.D. Rawlings, A.J. Barrett, P.D. Thomas, X. Huang, A. Bateman, R.D. Finn, The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database, *Nucleic Acids Res.* 46 (D1) (Jan. 2018) D624–D632, <https://doi.org/10.1093/nar/gkx1134>.
- [21] D. Greenbaum, K.F. Medzihradszky, A. Burlingame, M. Bogoy, Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools, *Chem. Biol.* 7 (8) (Aug. 2000) 569–581, [https://doi.org/10.1016/S1074-5521\(00\)00014-4](https://doi.org/10.1016/S1074-5521(00)00014-4).
- [22] K. Müntener, A. Willmann, R. Zwicky, B. Svoboda, L. Mach, A. Baici, Folding competence of N-terminally truncated forms of human Procathepsin B*, *J. Biol. Chem.* 280 (12) (Mar. 2005) 11973–11980, <https://doi.org/10.1074/jbc.M413052200>.
- [23] M. Paireder, et al., The papain-like cysteine proteinases NbCysP6 and NbCysP7 are highly processive enzymes with substrate specificities complementary to Nicotiana benthamiana cathepsin B, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1865 (4) (Apr. 2017) 444–452, <https://doi.org/10.1016/j.bbapap.2017.02.007>.
- [24] K. Yamada, R. Matsushima, M. Nishimura, I. Hara-Nishimura, A slow maturation of a cysteine protease with a Granulin domain in the vacuoles of senescing Arabidopsis leaves, *Plant Physiol.* 127 (4) (Dec. 2001) 1626–1634.
- [25] C. Gu, et al., Post-translational regulation and trafficking of the Granulin-containing protease RD21 of Arabidopsis thaliana, *PLoS One* 7 (3) (Mar. 2012) e32422, <https://doi.org/10.1371/journal.pone.0032422>.
- [26] G.T. Hermanson, The Reactions of Bioconjugation, in: *Bioconjugate Techniques*, Elsevier, 2013, pp. 229–258, <https://doi.org/10.1016/B978-0-12-382239-0.00003-0>.
- [27] F. Demir, M. Kuppasamy, A. Perrar, P.F. Huesgen, Profiling sequence specificity of proteolytic activities using proteome-derived peptide libraries, *Methods Mol. Biol.* 2447 (2022) 159–174, https://doi.org/10.1007/978-1-0716-2079-3_13.
- [28] K.H. Richau, et al., Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics, *Plant Physiol.* 158 (4) (Apr. 2012) 1583–1599, <https://doi.org/10.1104/pp.112.194001>.
- [29] A.I. Petushkova, L.V. Savvateeva, A.A. Zamyatin, Structure determinants defining the specificity of papain-like cysteine proteases, *Comput. Struct. Biotechnol. J.* 20 (Nov. 2022) 6552–6569, <https://doi.org/10.1016/j.csbj.2022.11.040>.
- [30] D. Brömme, P.R. Bonneau, P. Lachance, A.C. Storer, Engineering the S2 subsite specificity of human cathepsin S to a cathepsin L- and cathepsin B-like specificity, *J. Biol. Chem.* 269 (48) (Dec. 1994) 30238–30242.
- [31] R. Gupta, The oxygen-evolving complex: a super catalyst for life on earth, in response to abiotic stresses, *Plant Signal. Behav.* 15 (12) (Dec. 2020) 1824721, <https://doi.org/10.1080/15592324.2020.1824721>.
- [32] Y.G. Song, B. Liu, L.F. Wang, M.H. Li, Y. Liu, Damage to the oxygen-evolving complex by superoxide anion, hydrogen peroxide, and hydroxyl radical in photoinhibition of photosystem II, *Photosynth. Res.* 90 (1) (Oct. 2006) 67–78, <https://doi.org/10.1007/s11120-006-9111-7>.
- [33] A. Sharma, et al., Photosynthetic response of plants under different abiotic stresses: a review, *J. Plant Growth Regul.* 39 (2) (Jun. 2020) 509–531, <https://doi.org/10.1007/s00344-019-10018-x>.
- [34] S. Fortunato, C. Lasorella, N. Dipierro, F. Vita, M.C. de Pinto, Redox signaling in plant heat stress response, *Antioxidants* 12, no. 3, Art. no. 3 (Mar. 2023), <https://doi.org/10.3390/antiox12030605>.
- [35] Q.-L. Wang, J.-H. Chen, N.-Y. He, F.-Q. Guo, Metabolic reprogramming in chloroplasts under heat stress in plants, *Int. J. Mol. Sci.* 19, no. 3, Art. no. 3 (Mar. 2018), <https://doi.org/10.3390/ijms19030849>.
- [36] P. Goettig, Reversed proteolysis—proteases as peptide ligases, *Catalysts* 11, no. 1, Art. no. 1 (Jan. 2021), <https://doi.org/10.3390/catal11010033>.
- [37] L. Frazier, A.M. Weeks, Engineered peptide ligases for cell signaling and bioconjugation, *Biochem. Soc. Trans.* 48 (3) (Jun. 2020) 1153–1165, <https://doi.org/10.1042/BST20200001>.
- [38] C.R. Berkens, A. de Jong, H. Ovaa, B. Rodenko, Transpeptidation and reverse proteolysis and their consequences for immunity, *Int. J. Biochem. Cell Biol.* 41 (1) (Jan. 2009) 66–71, <https://doi.org/10.1016/j.biocel.2008.08.036>.
- [39] Z. Wang, et al., β -Lactone probes identify a papain-like peptide ligase in Arabidopsis thaliana, *Nat. Chem. Biol.* 4 (9) (Sep. 2008) 557–563, <https://doi.org/10.1038/nchembio.104>.
- [40] S. Alomrani, K.J. Kunert, C.H. Foyer, Papain-like cysteine proteases are required for the regulation of photosynthetic gene expression and acclimation to high light stress, *J. Exp. Bot.* 72 (9) (Apr. 2021) 3441–3454, <https://doi.org/10.1093/jxb/erab101>.
- [41] M. Bergmann, H. Fraenkel-Conrat, The rôle of specificity in the enzymatic synthesis of proteins: syntheses with intracellular enzymes, *J. Biol. Chem.* 119 (2) (Jul. 1937) 707–720, [https://doi.org/10.1016/S0021-9258\(18\)74417-7](https://doi.org/10.1016/S0021-9258(18)74417-7).
- [42] M. Bergmann, H. Fraenkel-Conrat, The enzymatic synthesis of peptide bonds, *J. Biol. Chem.* 124 (1) (Jun. 1938) 1–6, [https://doi.org/10.1016/S0021-9258\(18\)74064-7](https://doi.org/10.1016/S0021-9258(18)74064-7).
- [43] N. Vigneron, V. Stroobant, V. Ferrari, J. Abi Habib, B.J. Van den Eynde, Production of spliced peptides by the proteasome, *Mol. Immunol.* 113 (Sep. 2019) 93–102, <https://doi.org/10.1016/j.molimm.2018.03.030>.
- [44] A.W. Jacobitz, M.D. Kattke, J. Wereszczynski, R.T. Clubb, Sortase Transpeptidases: structural biology and catalytic mechanism, *Adv. Protein Chem. Struct. Biol.* 109 (2017) 223–264, <https://doi.org/10.1016/bs.apcsb.2017.04.008>.
- [45] A.M. James, J. Haywood, J.S. Mylne, Macrocyclization by asparaginyl endopeptidases, *New Phytol.* 218 (3) (2018) 923–928, <https://doi.org/10.1111/nph.14511>.
- [46] I. Saska, D.J. Craik, Protease-catalysed protein splicing: a new post-translational modification? *Trends Biochem. Sci.* 33 (8) (Aug. 2008) 363–368, <https://doi.org/10.1016/j.tibs.2008.04.016>.
- [47] D. Brömme, Papain-like cysteine proteases, *Curr. Protoc. Protein Sci.* Chapter 21 (May 2001), <https://doi.org/10.1002/0471140864.ps2102s21> p. Unit 21.2.
- [48] T.R. Lambeth, Z. Dai, Y. Zhang, R.R. Julian, A two-trick pony: lysosomal protease cathepsin B possesses surprising ligase activity, *RSC Chem Biol* 2 (2) (Apr. 2021) 606–611, <https://doi.org/10.1039/d0cb00224k>.

- [49] K. Sakina, K. Kawazura, K. Morihara, Enzymatic synthesis of delta sleep-inducing peptide, *Int. J. Pept. Protein Res.* 31 (2) (Jan. 2009) 245–252, <https://doi.org/10.1111/j.1399-3011.1988.tb00030.x>.
- [50] B. Reed, et al., Lysosomal cathepsin creates chimeric epitopes for diabetogenic CD4 T cells via transpeptidation, *J. Exp. Med.* 218 (2) (Oct. 2020) e20192135, <https://doi.org/10.1084/jem.20192135>.
- [51] N.J. Kruger, The Bradford Method For Protein Quantitation, in: J.M. Walker (Ed.), *The Protein Protocols Handbook*, Springer Protocols Handbooks, Humana Press, Totowa, NJ, 2009, pp. 17–24, https://doi.org/10.1007/978-1-59745-198-7_4.
- [52] K.P. van Midden, T. Peric, M. Klemencić, Plant type I metacaspases are proteolytically active proteases despite their hydrophobic nature, *FEBS Lett.* 595 (17) (Sep. 2021) 2237–2247, <https://doi.org/10.1002/1873-3468.14165>.
- [53] M. Novinec, M. Pavšič, B. Lenarčič, 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) (Mar. 2012) 1–5, <https://doi.org/10.1016/j.pep.2011.11.002>.
- [54] L. Mavridis, R.W. Janes, PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates, *Bioinformatics* 33 (1) (Jan. 2017) 56–63, <https://doi.org/10.1093/bioinformatics/btw554>.
- [55] A. Micsonai, É. Bulyáki, J. Kardos, BeStSel: From Secondary Structure Analysis to Protein Fold Prediction by Circular Dichroism Spectroscopy, in: Y.W. Chen, C.-P. B. Yiu (Eds.), *Structural Genomics: General Applications*, Methods in Molecular Biology, Springer US, New York, NY, 2021, pp. 175–189, https://doi.org/10.1007/978-1-0716-0892-0_11.
- [56] D. Brooke, N. Movahed, B., Bothner, and Department of Chemistry and Biochemistry, Montana State University, Bozeman MT 59717, USA, 'Universal buffers for use in biochemistry and biophysical experiments', *AIMS Biophysics* 2 (3) (2015) 336–342, <https://doi.org/10.3934/biophy.2015.3.336>.
- [57] W.T. Soh, et al., ExteNDing proteome coverage with Legumain as a highly specific digestion protease, *Anal. Chem.* 92 (4) (Feb. 2020) 2961–2971, <https://doi.org/10.1021/acs.analchem.9b03604>.
- [58] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat. Protoc.* 11 (12) (Dec. 2016) 2301–2319, <https://doi.org/10.1038/nprot.2016.136>.
- [59] S. Tyanova, et al., The Perseus computational platform for comprehensive analysis of (prote)omics data, *Nat. Methods* 13 (9) (Sep. 2016) 731–740, <https://doi.org/10.1038/nmeth.3901>.
- [60] N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo, *Nat. Methods* 6 (11) (Nov. 2009) 786–787, <https://doi.org/10.1038/nmeth1109-786>.
- [61] E.W. Deutsch, et al., The ProteomeXchange consortium at 10 years: 2023 update, *Nucleic Acids Res.* 51 (D1) (Jan. 2023) D1539–D1548, <https://doi.org/10.1093/nar/gkac1040>.
- [62] Y. Perez-Riverol, et al., The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, *Nucleic Acids Res.* 50 (D1) (Jan. 2022) D543–D552, <https://doi.org/10.1093/nar/gkab1038>.
- [63] J. Izdebski, T. Gers, D. Kunce, P. Markowski, New tris-alkoxycarbonyl arginine derivatives for peptide synthesis, *J. Pept. Sci.* 11 (1) (Jan. 2005) 60–64, <https://doi.org/10.1002/psc.585>.
- [64] M. Bodanszky, *Principles of Peptide Synthesis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1993, <https://doi.org/10.1007/978-3-642-78056-1>.
- [65] M. Bodanszky, A. Bodanszky, *The Practice of Peptide Synthesis*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1994, <https://doi.org/10.1007/978-3-642-85055-4>.
- [66] N. Fischer-Durand, M. Salmain, A. Vessières, G. Jaouen, A new bioorthogonal cross-linker with alkyne and hydrazide end groups for chemoselective ligation. Application to antibody labelling, *Tetrahedron* 68 (47) (Nov. 2012) 9638–9644, <https://doi.org/10.1016/j.tet.2012.09.062>.
- [67] I. Saska, D.J. Craik, Protease-catalysed protein splicing: a new post-translational modification? *Trends Biochem. Sci.* 33 (2008) 363–368.
- [68] B.R. Somalinga, R.P. Roy, Volume Exclusion Effect as a Driving Force for Reverse Proteolysis: Implications for polypeptide assemblage in a macromolecular crowded milieu, *J. Biol. Chem.* 277 (2002) 43253–43261.
- [69] W.F. Ettinger, S.M. Theg, Physiologically active chloroplasts contain pools of unassembled extrinsic proteins of the photosynthetic oxygen-evolving enzyme complex in the thylakoid lumen, *J. Cell Biol.* 115 (1991) 321–328.
- [70] K. Clark, J.Y. Franco, S. Schwizer, Z. Pang, E. Hawara, T.W.H. Liebrand, D. Pagliaccia, L. Zeng, F.B. Gurung, P. Wang, et al., An effector from the Huanglongbing-associated pathogen targets citrus proteases, *Nat. Commun.* 9 (2018) 1718.
- [71] G.V. Pogorelko, P.S. Juvalé, W.B. Rutter, M. Hütten, T.R. Maier, T. Hewezi, J. Paulus, van der Hoorn, F.M. Grundler, S. Siddique, et al., Re-targeting of a plant defense protease by a cyst nematode effector, *Plant J.* 98 (2019) 1000–1014.
- [72] Z. Tabaeizadeh, H. Chamberland, R.-D. Chen, L.-X. Yu, G. Bellemare, J. G. Lafontaine, Identification and immunolocalization of a 65 kDa drought induced protein in cultivated tomato *Lycopersicon esculentum*, *Protoplasma* 186 (1995) 208–219.