nature chemical biology

β-Lactone probes identify a papain-like peptide ligase in *Arabidopsis thaliana*

Zheming Wang[1,2,6], Christian Gu[1,6], Tom Colby[3], Takayuki Shindo[1], Rengarajan Balamurugan[4,5], Herbert Waldmann[4], Markus Kaiser[2] & Renier A L van der Hoorn[1,2]

New activity-based probes are essential for expanding studies on the hundreds of serine and cysteine proteases encoded by the genome of *Arabidopsis thaliana*. To monitor protease activities in plant extracts, we generated biotinylated peptides containing a β-lactone reactive group. These probes cause strong labeling in leaf proteomes. Unexpectedly, labeling was detected at the N terminus of PsbP, nonproteolytic protein of photosystem II. Inhibitor studies and reverse genetics led to the discovery that this unusual modification is mediated by a single plant-specific, papain-like protease called RD21. In cellular extracts, RD21 accepts both β-lactone probes and peptides as donor molecules and ligates them, probably through a thioester intermediate, to unmodified N termini of acceptor proteins.

The genome of the model plant *Arabidopsis thaliana* encodes ~320 putative serine and ~140 putative cysteine proteases, including large families of ~60 subtilase-like proteases, ~60 serine carboxypeptidase–like proteins and ~30 papain-like cysteine proteases (PLCPs)[1]. Some of these proteases have key regulatory roles in defense and development, but the roles, substrates and activation mechanisms of most of these enzymes are unknown[2]. Studies of some of the presumed plant proteases revealed that they can also catalyze non-proteolytic reactions. Phytochelatin synthase, for example, acts as a glutathione transpeptidase, yielding phytochelatin, which is required for heavy-metal tolerance[3]. A number of serine carboxypeptidase–like proteins act as acyltransferases in the production of sinapoyl secondary metabolites, which protect plants against UV radiation[4].

Small-molecule activity-based probes (ABPs) are a powerful way to track protease activities in proteomes and living cells[5,6]. ABPs for serine and cysteine proteases contain a peptide scaffold that acts as a binding group, a reactive group that ensures an irreversible, covalent linkage to the active-site residue of the protease, a linker and a reporter tag for affinity purification and/or detection. ABPs for serine and cysteine proteases have been generated for PLCPs, caspases, legumains, gingipains, deubiquitinating and desumoylating enzymes, granzymes and other serine proteases[7–12].

So far, activity-based protein profiling of plant proteomes has been mainly done using DCG-04 (**1**), a biotinylated version of E-64 (**2**), which inhibits PLCPs[7]. Activity-based protein profiling with DCG-04 on *Arabidopsis* leaf extracts revealed the activities of six different PLCPs, including responsive to dessication-21 (RD21) and *Arabidopsis* aleurain-like protease (AALP)[13]. The functions and substrates of these PLCPs are unknown. We also used profiling with DCG-04 to show that a pathogen-derived protein specifically inhibits diversifying, defense-related PLCPs of tomato[14].

To expand the range of serine and cysteine proteases that can be monitored by activity-based protein profiling, we designed a new series of ABPs containing a β-lactone reactive group. This reactive group is found in covalent inhibitors of lipases, cysteine proteases and the proteasome[15–17]. The recent use of β-lactone reactive groups in ABPs resulted in probes that label various enzyme classes in bacterial proteomes, including proteases[18]. This labeling, however, requires side chains on the β-lactone that probably confer binding affinity for various enzymes. To target β-lactone probes to serine and cysteine proteases, we added a peptide backbone with a variant amino acid. Labeling with these probes in *Arabidopsis* leaf extracts revealed an unexpected labeling site, and further characterization led us to the discovery of a plant protease that can ligate β-lactone probes and peptides to N termini of acceptor proteins.

RESULTS

Labeling leaf proteomes with β-lactone probes

To design a new class of ABPs for serine and cysteine proteases, we used a threonine-based β-lactone linked to a variant amino acid, an amide linker and biotin ('IS' probes, **Fig. 1a**). The IS probe collection consists of six probes (**3**, **4**, **5**, **6**, **7** and **8**) with various amino acid residues representing hydrophilic, aromatic or aliphatic side chains. For all of the probes except IS6, nonbiotinylated derivatives (IS-n; **9**, **10**, **11**, **12** and **13**) were synthesized to serve as competitors of IS labeling.

[1]Plant Chemetics Lab, Max Planck Institute for Plant Breeding Research, Carl-von-Linne Weg 10, 50829 Cologne, Germany. [2]Chemical Genomics Centre of the Max Planck Society, Otto-Hahn-Straße 11, 44227 Dortmund, Germany. [3]Proteomics Service Centre, Max Planck Institute for Plant Breeding Research, Carl-von-Linne Weg 10, 50829 Cologne, Germany. [4]Max Planck Institute for Molecular Physiology, Otto-Hahn-Straße 11, 44227 Dortmund, Germany. [5]School of Chemistry, University of Hyderabad, Gachi Bowli, Hyderabad 500 046, India. [6]These authors contributed equally to this work. Correspondence should be addressed to R.A.L.v.d.H. (hoorn@mpiz-koeln.mpg.de).

Received 30 May; accepted 3 July; published online 27 July 2008; doi:10.1038/nchembio.104

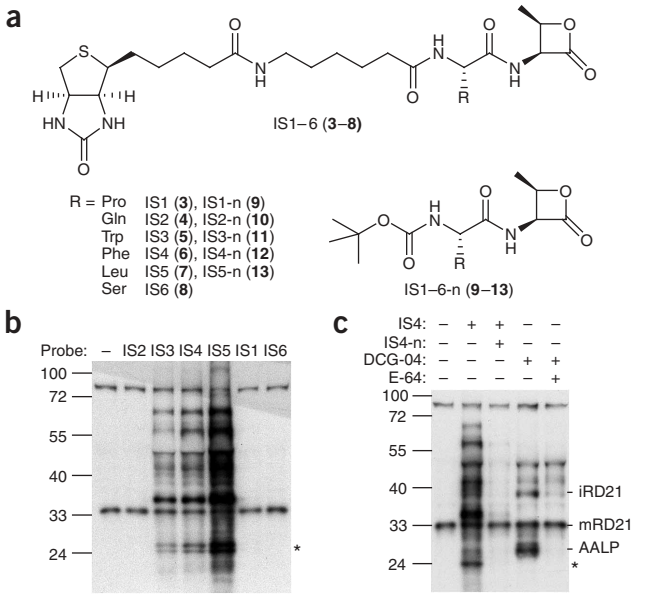

Figure 1 β-lactone probes and their labeling of leaf extracts. (**a**) Structures of β-lactone probes. Biotinylated (**3**, **4**, **5**, **6**, **7**, **8**) and nonbiotinylated (**9**, **10**, **11**, **12**, **13**) β-lactone derivatives were synthesized with various amino acid residues (R) next to the threonine-based β-lactone moiety. (**b**) Labeling with IS probes yielded similar profiles with various intensities. IS probes (2 μM) were incubated with *Arabidopsis* leaf extracts, and biotinylated proteins were detected on protein blots using streptavidin-HRP. (**c**) Labeling with IS4 and DCG-04 is specific and distinct. *Arabidopsis* leaf extracts were incubated for 2 h with 2 μM DCG-04 or IS4 in the absence or presence of nonbiotinylated 200 μM E-64 or 60 μM IS4-n, respectively. DCG-04 labeled AALP and the intermediate (i) and mature (m) isoforms of RD21. Biotinylated proteins were detected on protein blots using streptavidin-HRP. *, Hallmark band. Molecular weights (in kDa) are indicated to the left of the gels.

We first incubated *Arabidopsis* leaf extracts with the IS probes and detected biotinylated proteins on protein blots probed with streptavidin–horseradish peroxidase (HRP). IS3, IS4 and IS5 had similar labeling profiles, with strong signals at 23 and 36 kDa (**Fig. 1b**). In contrast, IS1, IS2 and IS6 did not cause any labeling compared to the no-probe control (**Fig. 1b**). Thus, IS probes with glutamine, proline and serine do not label, whereas IS probes with hydrophobic residues tryptophan, phenylalanine and especially leucine have similar labeling profiles with multiple signals.

IS4 labeling was further investigated, as it was synthesized in the largest quantity. Because IS probes potentially target serine and cysteine proteases, we compared IS4 labeling to that of DCG-04, which labels PLCPs[7]. DCG-04 labels six PLCPs in *Arabidopsis* leaf extracts, including AALP and intermediate and mature isoforms of RD21 (ref. 13). The RD21 intermediate isoform carries an additional C-terminal granulin domain of unknown function that is proteolytically removed during maturation[19]. As shown previously, the DCG-04 activity profile contains signals from the 40-kDa RD21 intermediate isoform, 30-kDa RD21 mature isoform and 25-kDa AALP, all of which were competed by adding an excess of E-64 during labeling (**Fig. 1c**). The remaining 30-kDa and 80-kDa signals are background signals, as these were also present in the no-probe control. IS4 labeling was of strong intensity compared to the DCG-04 activity profile, and its profile was different, indicating that IS4 does not label the same set of PLCPs (**Fig. 1c**). The presence of an excess of nonbiotinylated IS4-n during IS4 labeling outcompeted the biotinylation, indicating that the labeling is specific (**Fig. 1c**). IS4 labeling was also competed by IS3-n and IS5-n, but not by IS1-n or IS2-n (**Supplementary Fig. 1** online). These observations are consistent with the labeling of the biotinylated IS series, indicating that the same set of proteins is labeled with IS3, IS4 and IS5. Further characterization of IS4 labeling revealed that it occurs mainly at pH 7–9, with an optimum pH of 8 (**Supplementary Fig. 2a** online) and requires the presence of a reducing agent (**Supplementary Fig. 2b**).

IS4 labels PsbP at the N terminus

In all labeling experiments, there were consistent strong signals at 36 and 23 kDa. The 23-kDa protein was identified through large-scale affinity capture on streptavidin magnetic beads (**Supplementary Fig. 3** online) and tandem MS. The other biotinylated proteins could not be captured efficiently on streptavidin beads. MS data (explained in detail below) indisputably showed that the 23-kDa protein represents protein P of the oxygen-evolving complex of photosystem II (PsbP)[20]. The identity was confirmed by showing that purified, biotinylated proteins cross-react with PsbP-specific antiserum (**Supplementary Fig. 3**).

PsbP has a mature size of 23 kDa and is abundant in leaf extracts[20]. The labeling of PsbP by IS4 is notable, as PsbP is not a serine or cysteine protease, and no nucleophilic serine or cysteine residues have been reported for this protein. To identify the labeling site, we examined the MS spectra in detail. Comparison of the spectra of the IS4-labeled sample and the no-probe control revealed a series of peptides specific for the IS4-labeled sample (**Fig. 2a**, top). Most of these peptides matched the predicted tryptic PsbP peptides, covering 65% of the mature protein (**Fig. 2a**, bottom). The fact that these peptides were found with the predicted, unmodified masses indicates that IS4 was not attached to these regions of PsbP. Among the 'missing' peptides of PsbP was the N-terminal peptide, which should have a mass of 1,351.70 Da. Notably, the spectrum of the IS4-specific sample contained an additional mass of 1,938.99 Da (**Fig. 2a**, top inset), which fits the sum of the masses of the N-terminal peptide and the predicted mass of IS4 (**Fig. 2a**, bottom). The second additional mass at 1,954.99 Da from the IS4-specific sample is the oxidized version of this labeled N-terminal peptide.

To determine the site of IS4 labeling within the N-terminal peptide, we investigated the peptide fragmentation data of the labeled N-terminal peptide. The N-terminal peptide has the sequence AYGEAANVFGKPK. The MS/MS data contained a long series of y ions with the predicted masses, indicating that IS4 is not attached to any of the C-terminal peptide series, up to GEAANVFGKPK (**Fig. 2b**, y3–11). In contrast, masses for N-terminal peptide fragments up through AYGE were found as b ions in the MS/MS spectrum, but only if the mass of IS4 was added (**Fig. 2b**, b4–7). The presence of the b4 ion indicates that IS4 is attached to the N-terminal alanine. The MS/MS spectrum also contained the mass of IS4 itself (b3), indicating that the linkage between IS4 and alanine is likely to be an amide bond, as these bonds break during postsource fragmentation (**Fig. 2b**, b3). The b2 and b1 ions indicate that IS4 loses masses corresponding to a threonine and phenylalanine, respectively, indicating that the peptide bonds in IS4 are preserved and no other modifications occurred (**Fig. 2b**, b2 and b1). Taken together, these data indicate that IS4 is attached to the N terminus of PsbP through a peptide bond between a C-terminal threonine of the probe to the N-terminal alanine of PsbP (**Fig. 2c**).

To investigate whether other proteins are labeled by IS4 at primary amino groups, we treated the leaf proteome with

sulfo-*N*-hydroxysuccinimide acetate (sulfo-NHS-Ac) (14). Sulfo-NHS-Ac reacts with deprotonated primary amino groups of N termini above pH 7 and lysines above pH 9. Pretreatment of the leaf proteome with sulfo-NHS-Ac at pH 7.5 suppressed labeling by IS4 globally (**Supplementary Fig. 4** online), indicating that IS4 is linked to N termini of labeled proteins.

IS4 labeling requires cysteine protease RD21

The mechanism by which IS4 labels the N terminus of PsbP was puzzling, given that PsbP is not a serine or cysteine protease and that labeling does not occur at serine or cysteine residues. However, a clue to the mechanism of IS4 labeling came when protease inhibitors were studied. IS4 labeling was inhibited with cysteine protease inhibitors E-64, leupeptin (15) and antipain (16), but not with serine protease inhibitor PMSF (17; **Fig. 3a**). Because E-64 specifically targets PLCPs, these data suggest that a PLCP is required for IS4 labeling.

The *Arabidopsis* genome encodes 30 PLCPs, of which at least 10 are expressed in leaves (http://www.genevestigator.ethz.ch/) and 6 were previously identified by DCG-04 labeling in leaf extracts[13]. We reasoned that one of these leaf PLCPs could be responsible for IS4 labeling. We therefore generated PLCP-knockout lines by selecting lines carrying a T-DNA insertion in the genes encoding leaf-expressed PLCPs. IS4 labeling of extracts from these mutant plants revealed that labeling occurs in leaf extracts of all mutants except those of the *rd21-1* line (**Fig. 3b**). The absence of IS4 labeling was confirmed with an independent knockout line, *rd21-2* (data not shown). These data indicate that only RD21 is required for IS4 labeling in leaf extracts.

The absence of labeling in the *rd21* lines could be caused by the absence of PsbP and other acceptor proteins. To exclude this possibility, the *rd21-1* proteome was probed with antibody to PsbP and compared to signals from the wild-type proteome. This western blot showed that PsbP is present in both wild-type and *rd21-1* mutant plants, unaltered in size or quantity (**Supplementary Fig. 5** online). Thus, the absence of IS4 labeling is not caused by the absence of acceptor proteins in *rd21* mutants.

RD21 complements IS4 labeling *in vitro*

Having determined that RD21 is required for IS4 labeling, we tested whether RD21 could also complement IS4 labeling in *rd21* or E-64–treated proteomes. We produced *Arabidopsis* RD21 by agroinfiltration in *Nicotiana benthamiana* by expressing RD21 with the silencing inhibitor p19 (refs. 21,22). This procedure ensures high levels of recombinant proteins produced *in planta* with all required post-translational modifications. RD21 production was confirmed using RD21-specific antibodies (**Supplementary Fig. 6** online). Extracts containing recombinant RD21 were diluted to a concentration at which RD21 activity was similar to that of *Arabidopsis* leaf extract (**Fig. 3c**).

We next tested whether recombinant RD21 could complement IS4 labeling in proteomes of *rd21-1* mutant plants. Adding recombinant RD21 to proteomes of *rd21-1* mutant plants complemented IS4 labeling (**Fig. 3d**). This restoration of IS4 labeling did not occur when recombinant RD21–deficient extracts of agroinfiltrated *N. benthamiana* were added, again indicating that recombinant RD21 is required to restore IS4 labeling (data not shown). No IS4

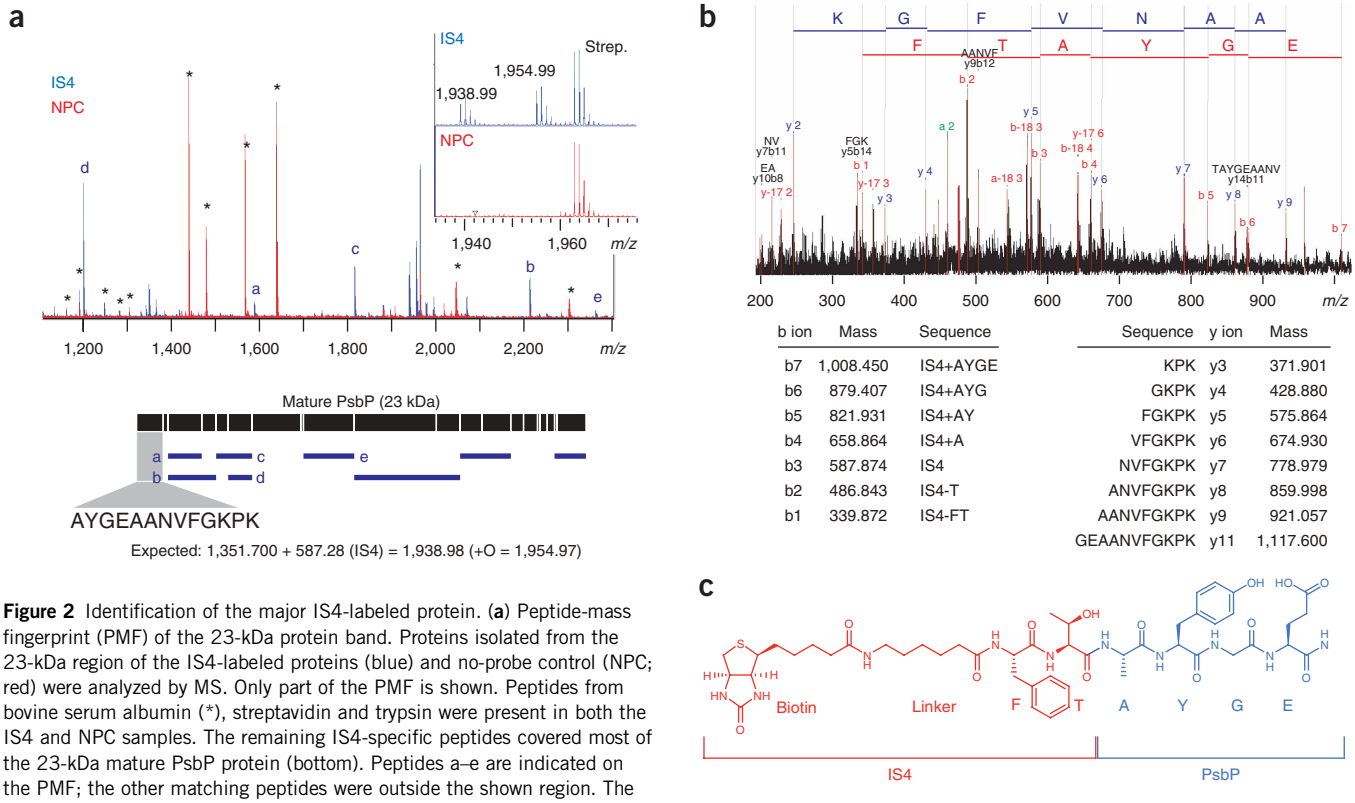

Figure 2 Identification of the major IS4-labeled protein. (**a**) Peptide-mass fingerprint (PMF) of the 23-kDa protein band. Proteins isolated from the 23-kDa region of the IS4-labeled proteins (blue) and no-probe control (NPC; red) were analyzed by MS. Only part of the PMF is shown. Peptides from bovine serum albumin (*), streptavidin and trypsin were present in both the IS4 and NPC samples. The remaining IS4-specific peptides covered most of the 23-kDa mature PsbP protein (bottom). Peptides a–e are indicated on the PMF; the other matching peptides were outside the shown region. The section of the PMF with the IS4-modified, N-terminal peptide is shown in the inset and explained by the calculation on the bottom. Both the IS4-labeled peptide and its oxidized form have predicted masses that match the masses in the PMF inset. (**b**) Fragmentation data of the IS4-labeled N-terminal peptide. The predicted y ions (bottom right) are found in the spectrum at the expected masses (top). The b ions are also found in the spectrum, with the additional mass of IS4. IS4 itself and fragments of IS4 are also found in the spectrum (b1, b2 and b3 ions). (**c**) Proposed structure of the N terminus of IS4-labeled PsbP, based on the peptide fragmentation data. IS4 is linked by a normal threonine through a peptide bond to the N-terminal alanine of PsbP.

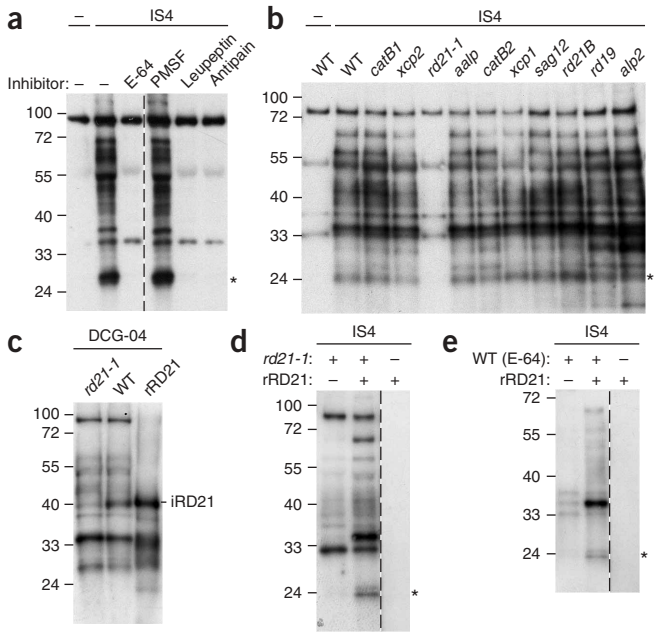

Figure 3 IS4 labeling requires cysteine protease RD21. (**a**) IS4 labeling requires cysteine, but not serine, protease activities. *Arabidopsis* leaf extracts were preincubated with 30 μM protease inhibitors and then incubated with 2 μM IS4. *, PsbP. (**b**) IS4 labeling requires the PLCP RD21. Leaf extracts of *Arabidopsis* PLCP knockout lines were incubated with IS4. WT, wild type. (**c**) Recombinant RD21 (rRD21) activity compared to that in *Arabidopsis* leaf extracts. RD21 was overexpressed by agroinfiltration in *Nicotiana benthamiana* and used as a source of rRD21. Leaf extracts from agroinfiltrated leaves and from *Arabidopsis* leaves of wild-type and *rd21* mutant plants were labeled with DCG-04 at pH 8 to reveal relative RD21 activities. To get similar levels of RD21 activities, 40× less protein was loaded from agroinfiltrated leaves compared to *Arabidopsis* leaves. iRD21, intermediate isoform. (**d**) rRD21 complements IS4 labeling in extracts of *rd21* mutant plants. Leaf extracts from wild-type and *rd21* mutant plants were labeled with IS4 in the absence or presence of rRD21-containing extracts. (**e**) Recombinant RD21 complements IS4 labeling in E-64–treated proteomes from wild-type *Arabidopsis* leaves. Leaf extracts were treated with 20 μM E-64. Excess E-64 was removed by gel filtration, and extracts were labeled with IS4 in the absence or presence of rRD21-containing extracts. In each panel, biotinylated proteins were detected on protein blots using streptavidin-HRP. Dashed lines indicate lanes that were left out from the blot. *, PsbP.

labeling was observed in recombinant RD21–containing extracts themselves, because these were diluted to adjust the recombinant RD21 concentration (**Fig. 3d**).

To investigate the mechanism by which E-64 inhibits IS4 labeling, we incubated leaf extracts of wild-type plants with E-64 to inactivate RD21 and other PLCPs. The excess of E-64 was removed by gel filtration, and the E-64–treated proteome was used for IS4 labeling. No IS4 labeling occurred on these E-64-treated proteomes (**Fig. 3e**). However, IS4 labeling was restored to normal levels by adding recombinant RD21 (**Fig. 3e**). These results indicate that E-64 does not act by occupying IS4-binding sites on the target proteins, but rather by inactivating PLCPs, presumably RD21.

We also investigated whether IS4 could label in living cells. We used both *Arabidopsis* cell cultures and detached leaves. Incubation of cell cultures with IS4 did not label specific proteins, even when the pH was increased from 5.7 to 8 (**Supplementary Fig. 7a** online). Similarly, no IS4-specific labeling was detected when detached leaves were incubated with IS4 (**Supplementary Fig. 7b**). To test whether the non-biotinylated IS4-n could enter the tissue and reach RD21, leaves were preincubated with IS4-n and E-64 *in vivo*, and then proteins were extracted and used for labeling with DCG-04 and IS4. Preincubation with E-64 blocked labeling of RD21 by DCG-04 and prevented IS4 labeling, indicating that E-64 inhibits RD21 activity *in vivo* (**Supplementary Fig. 7c**). In contrast, preincubation with IS4-n did not block labeling of RD21 by DCG-04, indicating that there is no inhibition of RD21 by IS4-n *in vivo*. Extracts from leaves pretreated with IS4-n could be labeled with IS4, consistent with the absence of inhibition of RD21 by IS4-n. Thus, IS4-n may not be reaching RD21 in living tissues because it is unstable *in vivo* or not membrane permeable. Limited *in vivo* labeling of β-lactone probes was also noted in studies on bacterial cell cultures[18].

Binding of β-lactones to RD21

The above data show that labeling of IS4 to other proteins is mediated by RD21. To investigate whether RD21 itself is labeled by IS4, we analyzed purified IS4-labeled proteins with RD21-specific antisera. Neither intermediate nor mature RD21 signals were detected among the labeled proteins, indicating that RD21 is not labeled by IS4 (**Supplementary Fig. 8** online). These data are consistent with our previous observation that the IS4-labeling pattern does not contain signals of the size of intermediate RD21 (**Fig. 1c**).

These findings suggest that IS4 binds to RD21 as an unstable intermediate that is not retained during purification of the labeled proteins. To show that IS4 binds to RD21, we labeled *Arabidopsis* leaf extracts with DCG-04 in the presence of excess nonbiotinylated IS4-n and IS2-n. IS4-n prevented labeling of intermediate RD21 by DCG-04, suggesting that IS4 occupies the substrate binding groove of RD21 (**Fig. 4a**). In contrast, IS2-n could not prevent labeling of iRD21 by DCG-04 (**Fig. 4a**), indicating that IS2 does not label (**Fig. 1b**) because it does not interact with iRD21. Notably, AALP was labeled by DCG-04, indicating that this protease does not bind IS4-n (**Fig. 4a**).

We next tested the pH dependency of IS4-n binding. RD21 reacts with DCG-04 at pH 5–10 (**Fig. 4b**), indicating that RD21 is active at a wide pH range. However, preincubation with IS4-n prevented DCG-04 labeling of RD21 at pH 6–9 (**Fig. 4b**), indicating that IS4-n binds to RD21 only in this pH range. The range of IS4-n binding coincides with that of IS4 labeling (pH 7–9; **Supplementary Fig. 2a**), except for pH 6. These data suggest that at pH 6, IS4 binds to RD21, but there is no transfer onto other proteins.

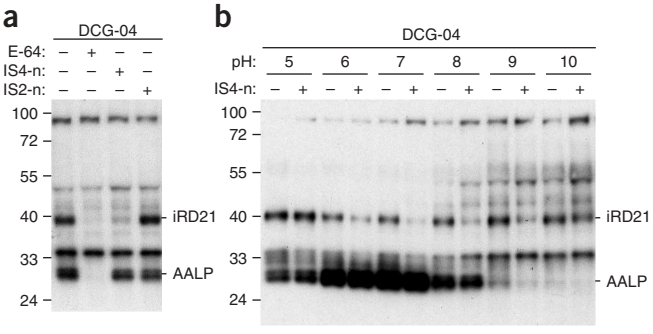

Figure 4 Binding of β-lactones to RD21. (**a**) Nonbiotinylated IS4-n prevents intermediate RD21 isoform (iRD21) labeling in leaf extracts. *Arabidopsis* leaf extracts were labeled with 2 μM DCG-04 in the presence or absence of 200 μM nonbiotinylated E-64, IS2-n or IS4-n. (**b**) IS4-n prevents iRD21 labeling at pH 6–9. *Arabidopsis* leaf extracts were incubated with 1 μM DCG-04 with or without 200 μM IS4-n at various pHs. In each panel, biotinylated proteins were detected on protein blots using streptavidin-HRP.

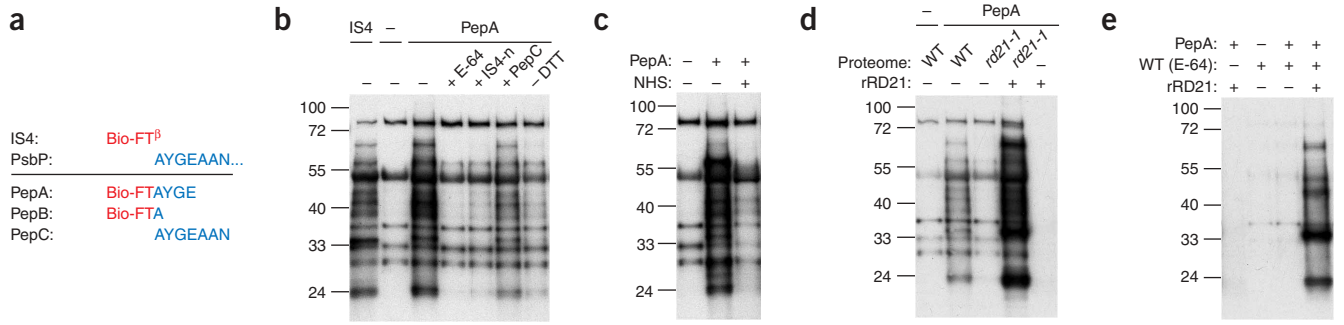

Figure 5 RD21 can ligate peptides. (**a**) Biotinylated peptides PepA, PepB and PepC were compared to sequences of IS4 (red) and PsbP (blue). (**b**) PepA labeling can be prevented by E-64, IS4-n and PepC. Leaf extracts were preincubated with or without 120 μM E-64, 120 μM IS4-n or 300 μM PepC and labeled with or without 20 μM PepA or 3 μM IS4. Only half of the IS4-labeled sample was loaded. (**c**) PepA labeling occurs at primary amino groups. Leaf extracts were preincubated with or without 600 μM sulfo-NHS-Ac (NHS) at pH 7.4 to block the N-terminal amino groups and then labeled with 30 μM PepA. (**d**) PepA labeling is RD21 dependent. Leaf extracts of wild-type and *rd21* mutant plants were labeled with PepA in the presence or absence of recombinant RD21 (rRD21). WT, wild type. (**e**) rRD21 complements PepA labeling in E-64–treated proteomes. Leaf extracts were treated with 20 μM E-64. Excess E-64 was removed by gel filtration, and extracts were labeled with 30 μM PepA in the absence or presence of rRD21. In **b**–**e**, biotinylated proteins were detected on protein blots using streptavidin-HRP.

RD21 ligates peptides

Given the above results, we hypothesized that IS4 binds to RD21 and forms a thioester intermediate that can be transferred to the N terminus of PsbP. Because a thioester bond is common to all intermediates of PLCPs with their substrates, we tested whether thioesters formed from peptides could also be ligated to other proteins by RD21. To ensure that the peptides would bind RD21, we designed them based on sequences of IS4 (Bio-FT$^{\beta}$) and the N terminus of PsbP (AYGEAAN). Three peptides were synthesized: the biotinylated peptides Bio-FTAYGE (PepA) and Bio-FTA (PepB) and the nonbiotinylated peptide AYGEAAN (PepC; **Fig. 5a**). These peptides are nonelectrophilic agents that combine key recognition elements of IS4 and the N terminus of PsbP.

Labeling of leaf extract with PepA resulted in a profile very similar to that of IS4 (**Fig. 5b**). PepB, in contrast, did not cause any biotinylation (data not shown). Analogous to labeling by IS4, PepA labeling was prevented by adding E-64 or IS4-n and required the reducing agent DTT (**Fig. 5b**). PepA labeling was reduced by adding the nonbiotinylated PepC (**Fig. 5b**), indicating that PepC competes with PepA labeling of acceptor proteins such as PsbP. Labeling by PepA was suppressed by preincubation with sulfo-NHS-Ac (**Fig. 5c**), indicating that PepA labels at primary amines, presumably N termini. Labeling with PepA was absent in *rd21-1* mutant proteomes and was complemented by recombinant RD21 (**Fig. 5d**). Finally, PepA labeling was absent from E-64–treated proteomes but was complemented by recombinant RD21 (**Fig. 5e**). Taken together, these observations indicate that RD21 ligates the N-terminal moiety of PepA to the N termini of acceptor proteins.

DISCUSSION

The data are consistent with a model in which β-lactones and peptides bind to RD21 and form a thioester bond that is transligated to abundant, unmodified N termini of acceptor proteins (**Fig. 6**). The binding of donors to RD21 is probably mediated by a phenylalanine residue at the P2 position, which is consistent with a preference for such hydrophobic P2 residues by PLCPs[23]. Transfer of the donor molecule occurs at neutral or basic pH levels, when the N terminus of acceptor molecule is deprotonated and can act as a nucleophile.

The fact that more plant PLCPs are active in leaf extracts[13], yet no labeling is observed in extracts of *rd21* knockout lines, indicates that IS4 labeling and peptide ligation are a distinctive property of RD21. RD21 carries a C-terminal granulin domain, which shows homology to animal growth factors released upon wounding[24]. Granulin-containing PLCPs are unique to the plant kingdom but highly conserved in both monocot and dicot plant species[1]. In addition,

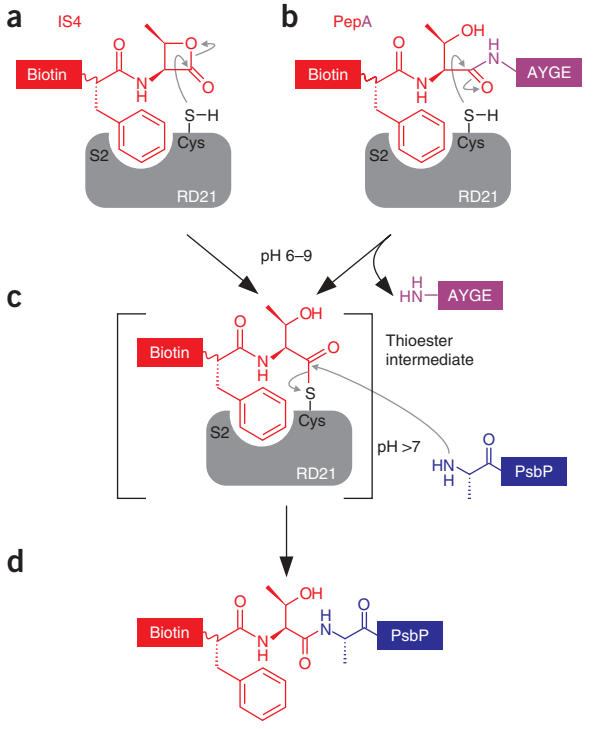

Figure 6 Model for β-lactone and peptide labeling of PsbP by RD21. (**a**–**d**) β-lactone probes (**a**) and peptides (**b**) bind to RD21. The phenylalanine residue of these acceptor molecules (IS4 and PepA, respectively) is at the P2 position, making contact with the S2 substrate binding pocket of RD21. The active-site cysteine of RD21 acts as a nucleophile, resulting in an unstable thioester intermediate (**c**). At neutral to basic pH, the N-terminal amino group of PsbP acts as a nucleophile on the thioester intermediate, resulting in labeling of the N terminus of PsbP through a peptide bond (**d**).

phylogenetic analysis of the protease domain of RD21 shows that RD21 falls in a separate clade that lacks animal counterparts (http://merops.sanger.ac.uk/). RD21 exists in two active isoforms: a 40-kDa intermediate isoform with granulin domain and a 30-kDa mature isoform without granulin domain[19]. Both isoforms are present in leaf extracts, but whether the granulin domain is required for transligation remains to be tested.

At this stage, it is not known whether PsbP is a natural acceptor for RD21 transligation, as all labeling experiments were performed on leaf extracts, and PsbP and RD21 might be compartmentalized in living cells. Post-translational modification of N termini can be an important regulatory mechanism[25], and RD21 might regulate proteins by ligating donor molecules to their N termini. One way to investigate this further is to examine which proteins are N-terminally modified by RD21 *in vivo* using labeled donor molecules. The identification of the native N-terminal modification on these natural acceptors might lead to the identification of native donor molecules and the discovery of a novel, plant-specific post-translational modification.

Although we were unable to detect labeling by IS4 in living cells, it seems likely that transligation can occur at physiological conditions, as it requires RD21, a reducing agent, neutral or basic pH, donor peptides and acceptor proteins with unmodified N termini. RD21 has been detected in vesicles and vacuoles of *Arabidopsis*[26,27], and the tomato RD21-like protease C14/TDI-65 has been detected in apoplasts, chloroplasts and nuclei[14,28]. The pH of some of these compartments would allow transligation reactions. At this stage, however, we cannot exclude that RD21 would only act as a transligase in extracts and as a protease on natural substrates *in vivo*.

So far, there are two other plant cysteine proteases described that catalyze transpeptidation reactions. Phytochelatin synthase (family C72) cleaves the tripeptide glutathione and ligates the γGlu-Cys moiety to glutathione to produce phytochelatin, which is essential for heavy-metal tolerance[29]. Vacuolar processing enzyme (family C13) is required for the production of a circular peptide (cyclotide) called kalata B1, which might have a role in insect defense[30]. RD21 is the first representative of a third cysteine protease family (C1A, the PLCPs) that can catalyze transpeptidation reactions.

Little is known of the mechanisms of transpeptidases. The data suggest that water is somehow excluded from the active site to promote transpeptidation reactions. It will be interesting to investigate whether the acceptor molecule is already bound before formation of the thioester intermediate to exclude water from the active site.

The post-translational fusion of peptides or proteins through peptide bonds using enzymes has great potential for applications in research and medicine[31]. Sortin, for example, is a bacterial transpeptidase that has been used to ligate peptides to 'sortagged' cell wall proteins on living cells[32,33]. Subtiligase, derived from a subtilase, has been used for protein semisynthesis[34,35]. Each of these enzymes has its own opportunities and limitations. The use of RD21-like PLCPs might open new avenues for controlling post-translational modifications, but further optimization and characterization are required. For example, the specificity requirements for the donor and acceptor molecules and the efficiency of transpeptidation compared to proteolysis remain to be addressed.

Although we have designed β-lactone derivatives as ABPs, the fate of these small molecules in plant extracts seems more complex than we predicted. IS4 labeling does not depend on the activity of the targeted proteins, but rather results from indirect labeling through a presumed protease. This indicates that further investigation of unexpected labeling sites can lead to intriguing molecular mechanisms. The mechanism of transligation, the selectivity for donor and acceptor substrates, and whether these reactions also occur in living cells are topics worthy of further study.

METHODS

Synthesis of β-lactone probes. The probe collection was synthesized in several steps, starting from commercially available L-Boc-threonine, which was cyclized to the corresponding β-lactone moiety by PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) and triethylamine as a base in dry dichloromethane. The nonbiotinylated probes (IS-n) were generated by cleavage of the Boc-protecting group with trifluoroacetic acid and *p*-toluolsulfonic acid, followed by coupling of the corresponding Boc-protected amino acid building blocks either by the mixed anhydride method (ethylchloroformate and triethylamine in dichloromethane) or standard peptide coupling (PyBOP, HOBt (*N*-hydroxybenzotriazole) and diisopropylamine in dichloromethane and dimethylformamide). The biotin-tagged probes (IS) were obtained by Boc cleavage of IS-n probes with trifluoroacetic acid in dichloromethane and triisopropylsilane as a scavenger and a subsequent coupling of 6-(biotinyl)-aminocaproic acid with PyBOP, HOBt and triethylamine as a base in acetonitrile. More detailed procedures and compound characterizations can be found in the **Supplementary Methods** online. Biotinylated peptides were ordered from JPT Peptide Technologies. PepA (FTAYGE) and PepB (FTA) contain a Ttds linker (4,7,10-trioxa-1,13-tridecanediamine succinimic acid) with biotin at the N terminus and a carboxyl at the C terminus. The AYGAEEN peptide has an amine at the N terminus and a carboxyl at the C terminus.

Plant materials. *Arabidopsis* plants were grown in a growth cabinet on a 12-h light regime (day, 24 °C; night, 20 °C). T-DNA insertion mutants of RD21 (At1g47128: SALK_090550 (*rd21-1*) and SALK_065256 (*rd21-2*); At5g43060: SAIL_781H05 (*rd21B*)), SAG12 (At5g45890: SALK_124030), XCP1 (At4g35350; SALK_84789), XCP2 (At4g35350: SALK_057921), RD19 (At4g39090: SALK_031088), AALP (At4g60360: SALK_075550), ALP2 (At3g45310: SALK_088620), CATB1 (At4g01610: SALK_019630) and CATB2 (At1g02300: SALK_110946) were obtained from the Salk Institute Genomic Analysis Laboratory. All are in the ecotype Columbia-0 background (Col-0). Homozygous lines were selected using T-DNA and gene-specific primers designed by SIGnAL (http://signal.salk.edu/cgi-bin/tdnaexpress/).

Sample preparation, probe labeling and protein work. Labeling was carried out as described previously[13]. All experiments were repeated once or twice, with similar results. Briefly, proteins were extracted by grinding one rosette leaf in an Eppendorf tube, mixing with 0.5 ml of water and centrifuging for 1 min at 16,000*g*. Usually, ~20 μg of protein was labeled in 0.5 ml of total volume containing 50 mM Tris buffer (pH 8), 1 mM DTT (Roche) and 2 μM probe (for example, IS4). DMSO was used in the no-probe control at a volume equal to that of to the probe. Labeling was done for 2 to 4 h at room temperature (22–25 °C) with gentle agitation. Proteins were then precipitated by adding 1 ml ice-cold acetone and centrifuging for 1 min at 16,000*g*. The acetone pellet was dissolved in 2× SDS-PAGE loading buffer containing β-mercaptoethanol. Biotinylated proteins were separated on 12% SDS gels (~4 μg protein per lane), transferred onto polyvinylidene fluoride membrane (Immobilon-P, Millipore) and detected with streptavidin-HRP (1:3,000; Ultrasensitive, Sigma) or specific antibodies (1:5,000) and HRP-conjugated secondary antibodies (1:5,000; Amersham). Profiling at various pH levels was done using 50 mM sodium acetate (pH 4–6.5) or Tris (pH 7–10) buffers. Competition or inhibition assays were done by adding probe and competitor or inhibitor molecules simultaneously or by preincubating the extracts with competitor or inhibitor molecules for 30 min. Antibodies to PsbP and bovine HRP-conjugated anti-sheep antibodies were purchased from Agrisera and Santa Cruz Biotechnology, respectively. Blocking N termini was done by preincubating leaf extracts with or without 200 μM sulfo-NHS-Ac (Pierce) for 30 min in the presence of 150 mM NaCl and 50 mM phosphate buffer (pH 7.4) and then labeling for 1 h with 2 μM IS4. Recombinant RD21 was expressed by agroinfiltration using binary vector pRH628, which was constructed by cloning the open reading frame of *RD21* into pRH385 (ref. 36).

Affinity purification of target proteins. Leaf proteins of 6-week-old plants were extracted by grinding rosette leaves in water in an ice-cold mortar and centrifuging for 10 min at 20,000*g*. Ten milliliters of supernatant at 4 mg ml^{-1} were labeled with 40 μM IS4 for 2 h at room temperature with gentle agitation. The leaf extracts were then applied to PD-10 size exclusion columns (Bio-Rad) to remove unlabeled probe. Desalted samples were incubated with 100 μl of magnetic streptavidin beads (Promega) and protease inhibitor cocktail (Complete tablet, Roche) for 1 h at room temperature with gentle agitation. Streptavidin beads were collected in an Eppendorf tube using a magnetic holder; washed twice with 0.1% SDS, twice with 6 M urea, once with 50 mM Tris (pH 8) containing 1% Triton X-100, once with 1% Triton X-100 and once with water; and boiled in 30 μl of 2× SDS-PAGE loading buffer containing β-mercaptoethanol. Affinity-purified proteins were separated on 12% SDS gel and stained with Coomassie (Imperial Protein Stain, Pierce). The specific band was excised from the Coomassie-stained gel and subjected to in-gel tryptic digestion and MS spectrum analysis.

In-gel digestion and mass spectrometry. The relevant bands were visualized on a one-dimensional gel, robotically picked as rows of 1.2-mm spots and tryptically digested in gel (PROTEINEERdp, Burker). The digests were spotted onto Bruker AnchorChip targets for subsequent MALDI-TOF analysis[37]. Peptide-mass fingerprints were taken with a Bruker Ultraflex III MALDI-TOF/TOF spectrometer, and the resulting spectra were processed in FlexAnalysis 3.0 (Bruker). Peak files were used to identify the corresponding proteins in the ProteinScape 1.3 database system (Protagen), which triggered Mascot (Matrix Science) searches. PsbP (At1g06680) was identified on the basis of peptide-mass fingerprints, but no peptide masses were matched to the predicted tryptic peptides in the first 90 amino acids of the immature protein sequence. Sample spots were recrystallized on the target by adding 1 μl of recrystallization solution (ethanol, acetone and 0.1% trifluoroacetic acid in a 6:3:1 ratio), and LIFT MS/MS spectra[38] were collected for selected precursors. The peptide-mass fingerprints were manually searched against the PsbP sequence, allowing various modifications and N-terminal modifications. This process matched the prominent mass observed at 1,938.99 Da with an IS4-modified form of a new N-terminal peptide of PsbP with the sequence AYGEAANVFGKPK, which appeared to be the N terminus of mature PsbP[20]. The fragment ions observed in the MS/MS spectrum of this precursor were then matched to the predicted daughter ions of this peptide sequence and the modification.

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We thank C. MacKintosh (University of Dundee) for providing the antibodies to RD21 and C. Koncz (Max Planck Institute for Plant Breeding Research) for providing the *rd21B* knockout. This work was supported by the International Max Planck Research School (to C.G. and T.S.), the *Arabidopsis* Functional Genomics Netwerk of the Deutsche Forschungsgemeinschaft (to T.C.), the Alexander von Humboldt Foundation (to R.B.) and the Max Planck Society (to Z.W., R.B., H.W., M.K. and R.A.L.v.d.H.).

Published online at http://www.nature.com/naturechemicalbiology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Beers, E.P., Jones, A.M. & Dickerman, A.W. The S8 serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. *Phytochemistry* **65**, 43–58 (2004).
2. Van der Hoorn, R.A.L. Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* **59**, 191–223 (2008).
3. Clemens, S. Evolution and function of phytochelatin synthases. *J. Plant Physiol.* **163**, 319–332 (2006).
4. Lehfeldt, C. *et al.* Cloning of the SNG1 gene of *Arabidopsis* reveals a role for a serine carboxypeptidase-like protein as an acyltransferase in secondary metabolism. *Plant Cell* **12**, 1295–1306 (2000).
5. Evans, M.J. & Cravatt, B.F. Mechanism-based profiling of enzyme families. *Chem. Rev.* **106**, 3279–3301 (2006).
6. Fonovic, M. & Bogyo, M. Activity-based probes for proteases: application to biomarker discovery, molecular imaging and drug screening. *Curr. Pharm. Des.* **13**, 253–261 (2007).
7. Greenbaum, D., Medzihradszky, K.F., Burlingame, A. & Bogyo, M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569–581 (2000).
8. Kato, D. *et al.* Activity-based probes that target diverse cysteine protease families. *Nat. Chem. Biol.* **1**, 33–38 (2005).
9. Hemelaar, J. *et al.* Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins. *Mol. Cell. Biol.* **24**, 84–95 (2004).
10. Mahrus, S. & Craik, C.S. Selective chemical functional probes of granzymes A and B reveal granzyme B is a major effector of natural killer cell-mediated lysis of target cells. *Chem. Biol.* **12**, 567–577 (2005).
11. Pan, Z. *et al.* Development of activity-based probes for trypsin-family serine proteases. *Bioorg. Med. Chem. Lett.* **16**, 2882–2885 (2006).
12. Liu, Y., Patricelli, M.P. & Cravatt, B.F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**, 14694–14699 (1999).
13. Van der Hoorn, R.A.L., Leeuwenburgh, M.A., Bogyo, M., Joosten, M.H.A.J. & Peck, S.C. Activity profiling of papain-like cysteine proteases in plants. *Plant Physiol.* **135**, 1170–1178 (2004).
14. Shabab, M. *et al.* Fungal effector protein AVR2 targets diversifying defence-related Cys proteases of tomato. *Plant Cell* **20**, 1169–1183 (2008).
15. Lall, M.S., Karvellas, C. & Vederas, J.C. β-Lactones as a new class of cysteine proteinase inhibitors: inhibition of hepatitis A virus 3C proteinase by *N*-Cbz-serine β-lactone. *Org. Lett.* **1**, 803–806 (1999).
16. Dick, L.R. *et al.* Mechanistic studies on the inactivation of the proteasome by lactacystin in cultured cells. *J. Biol. Chem.* **272**, 182–188 (1997).
17. Drahl, C., Cravatt, B.F. & Sorensen, E.J. Protein-reactive natural products. *Angew. Chem. Int. Edn.* **44**, 5788–5809 (2005).
18. Böttcher, T. & Sieber, S.A. β-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angew. Chem. Int. Edn.* **47**, 4600–4603 (2008).
19. Yamada, K., Matsushima, R., Nishimura, M. & Hara-Nishimura, I. A slow maturation of a cysteine protease with a granulin domain in the vacuoles of senescing *Arabidopsis* leaves. *Plant Physiol.* **127**, 1626–1634 (2001).
20. Yi, X., Hargett, S.R., Liu, H., Frankel, L.K. & Bricker, T.M. The PsbP protein is required for photosystem II complex assembly/stability and photoautotrophy in *Arabidopsis thaliana*. *J. Biol. Chem.* **282**, 24833–24841 (2007).
21. Van der Hoorn, R.A.L., Laurent, F., Roth, R. & De Wit, P.J.G.M. Agroinfiltration is a versatile tool that facilitates comparative analysis of *Avr9/Cf-9*-induced and *Avr4/Cf-4*-induced necrosis. *Mol. Plant Microbe Interact.* **13**, 439–446 (2000).
22. Voinnet, O., Rivas, S., Mestre, P. & Baulcombe, D. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**, 949–956 (2003).
23. Powers, J.C., Asgian, J.L., Ekici, O.D. & James, K.E. Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem. Rev.* **102**, 4639–4750 (2002).
24. Bateman, A. & Bennett, H.P.J. Granulins: the structure and function of an emerging family of growth factors. *J. Endocrinol.* **158**, 145–151 (1998).
25. Walling, L.L. Recycling or regulation? The role of amino-terminal modifying enzymes. *Curr. Opin. Plant Biol.* **9**, 227–233 (2006).
26. Hayashi, Y. *et al.* A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. *Plant Cell Physiol.* **42**, 894–899 (2001).
27. Carter, C. *et al.* The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. *Plant Cell* **16**, 3285–3303 (2004).
28. Tabaeizadeh, Z. *et al.* Identification and immunolocalization of a 65 kDa drought induced protein in cultivated tomato *Lycopersicon esculentum*. *Protoplasma* **186**, 208–219 (1995).
29. Cobbett, C. & Goldsbrough, P. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annu. Rev. Plant Biol.* **53**, 159–182 (2002).
30. Saska, I. *et al.* An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J. Biol. Chem.* **282**, 28721–28728 (2007).
31. Lombard, C., Saulnier, J. & Wallach, J.M. Recent trends in protease-catalyzed peptide synthesis. *Protein Pept. Lett.* **12**, 621–629 (2005).
32. Popp, M.W., Antos, J.M., Grotenberg, G.M., Spooner, E. & Ploegh, H.L. Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).
33. Tanaka, T., Yamamoto, T., Tsukiji, S. & Nagamune, T. Site-specific protein modification on living cells catalyzed by sortase. *ChemBioChem* **9**, 802–807 (2008).
34. Chang, T.K., Jackson, D.Y., Burnier, J.P. & Wells, J.A. Subtiligase: a tool for semi-synthesis of proteins. *Proc. Natl. Acad. Sci. USA* **91**, 12544–12548 (1994).
35. Tan, X.H., Zhang, X., Yang, R. & Liu, C.F. A simple method for preparing peptide C-terminal thioacids and their application in sequential chemoenzymatic ligation. *ChemBioChem* **9**, 1052–1056 (2008).
36. Van der Hoorn, R.A.L. *et al.* Structure-function analysis of Cf-9, a receptor-like protein with extracytoplasmic leucine-rich repeats. *Plant Cell* **17**, 1000–1015 (2005).
37. Gobom, J. *et al.* Alpha-cyano-4-hydroxycinnamic acid affinity sample preparation. A protocol for MALDI-MS peptide analysis in proteomics. *Anal. Chem.* **73**, 434–438 (2001).
38. Suckau, D. *et al.* A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal. Bioanal. Chem.* **376**, 952–965 (2003).