

Mechanism-based traps enable protease and hydrolase substrate discovery

<https://doi.org/10.1038/s41586-022-04414-9>

Received: 12 June 2021

Accepted: 7 January 2022

Published online: 16 February 2022

Open access

 Check for updates

Shan Tang^{1✉}, Adam T. Beattie¹, Lucie Kafkova², Gianluca Petris¹, Nicolas Huguenin-Dezot¹, Marc Fiedler¹, Matthew Freeman² & Jason W. Chin^{1✉}

Hydrolase enzymes, including proteases, are encoded by 2–3% of the genes in the human genome and 14% of these enzymes are active drug targets¹. However, the activities and substrate specificities of many proteases—especially those embedded in membranes—and other hydrolases remain unknown. Here we report a strategy for creating mechanism-based, light-activated protease and hydrolase substrate traps in complex mixtures and live mammalian cells. The traps capture substrates of hydrolases, which normally use a serine or cysteine nucleophile. Replacing the catalytic nucleophile with genetically encoded 2,3-diaminopropionic acid allows the first step reaction to form an acyl-enzyme intermediate in which a substrate fragment is covalently linked to the enzyme through a stable amide bond²; this enables stringent purification and identification of substrates. We identify new substrates for proteases, including an intramembrane mammalian rhomboid protease RHBDL4 (refs. ^{3,4}). We demonstrate that RHBDL4 can shed luminal fragments of endoplasmic reticulum-resident type I transmembrane proteins to the extracellular space, as well as promoting non-canonical secretion of endogenous soluble endoplasmic reticulum-resident chaperones. We also discover that the putative serine hydrolase retinoblastoma binding protein 9 (ref. ⁵) is an aminopeptidase with a preference for removing aromatic amino acids in human cells. Our results exemplify a powerful paradigm for identifying the substrates and activities of hydrolase enzymes.

Activity-based probes have confirmed the presence and selective reactivity of the catalytic serine or cysteine nucleophile for many hydrolase proteins in cells^{6–8}. Efforts to define hydrolase specificity have captured non-covalent interactors with hydrolases and investigated the substrates that accumulate in the absence of a hydrolase or the products that accumulate in the presence of a hydrolase⁹. Current approaches to identifying protease substrates mostly aim to either co-immunoprecipitate substrates that are non-covalently bound to catalytically inactive protease variants^{10–12}, or identify the peptides resulting from the action of the protease from experiments with and without the protease^{13–15}. In the first approach, substrates may be lost in the washing steps, and bound proteins may not be substrates. The second approach typically underestimates the number of substrates, and the cleavages identified may be indirect^{16,17}. The identification of intramembrane protease substrates by current approaches is particularly challenging^{18,19}. The methodological challenges in defining hydrolase and protease specificity mean that the substrates of many proteases remain unknown or incomplete, and many hydrolases remain orphans—with unknown substrates and uncharacterized specificity.

We previously demonstrated the genetically encoded, site-specific incorporation of photocaged Dap ((2*S*)-2-amino-3-[(2-[[1-(6-nitrobenzo[*d*][1,3]dioxol-5-yl)ethyl]thio]ethoxy)carbonyl]amino]propanoic acid) (pc-Dap) into proteins expressed in *Escherichia coli*²⁰. We converted pc-Dap to Dap (2,3-diaminopropionic acid) in proteins

in vitro by illuminating purified proteins followed by incubation for up to 2 days at pH 8. By incubating purified proteases or thioesterases—in which we had replaced the catalytic cysteine or serine with Dap—with known substrates, we captured the otherwise transient thioester or ester intermediates—resulting from the first step of the reaction of these enzymes with their substrates—as their stable amide analogues. We demonstrated the utility of this approach for structural studies of acyl-enzyme intermediates².

Here we demonstrate that genetically encoded pc-Dap coupled to mass spectrometry provides a powerful approach for discovering hydrolase substrates in complex mixtures and in live mammalian cells (Extended Data Fig. 1). We first demonstrate that recombinant proteases containing Dap in place of their catalytic nucleophile can be used to covalently link the N-terminal fragment of substrates to the protease, via a stable amide bond, in cell lysates. Through stringent purification of substrates linked to high-temperature requirement protein A2 (HtrA2) protease in combination with a mass spectrometry workflow we identify more than 200 new substrates for this protease. We then demonstrate that pc-Dap can be genetically encoded into proteins in mammalian cells. Illumination effects the rapid, post-translational conversion of pc-Dap to Dap in live cells and enables the specific covalent capture of the N-terminal fragments of protease substrates in cells. We identify new substrates for the mammalian rhomboid protease RHBDL4, an intramembrane protease that resides in the endoplasmic reticulum

¹Medical Research Council Laboratory of Molecular Biology, Cambridge, UK. ²Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ✉e-mail: stang@mrc-lmb.cam.ac.uk; chin@mrc-lmb.cam.ac.uk

(ER). Our results demonstrate that RHBDL4 can shed luminal fragments of type I transmembrane proteins to the extracellular space. Upon removal of the ER retention motif, RHBDL4 promotes the secretion of a wide range of ER chaperones from cells. Finally, we develop a pipeline to directly identify the branched peptides formed between a substrate fragment and Dap and thereby reveal that the putative serine hydrolase retinoblastoma binding protein 9 (RBBP9) is an aminopeptidase with a preference for removing aromatic amino acids in mammalian cells.

Trapping substrates in complex mixtures

We demonstrated that recombinant Tobacco etch virus (TEV) protease containing Dap in place of its catalytic cysteine (TEV(C151Dap)) formed a specific conjugate with GFP with a TEV protease cleavage site appended at its C-terminus (GFP-s) when the two proteins were incubated together in human cell lysate (Extended Data Fig. 2). Although GFP-s was present in the lysate at comparable levels to many other proteins, TEV(C151Dap) formed a specific conjugate, observed by immunoblotting, with its substrate partner and not with any other proteins in the lysate. Control experiments demonstrated that conjugate formation was dependent on both the presence of Dap in the protease and the presence of the TEV cleavage site in the substrate. Replacing the catalytic cysteine of other proteases (UL36^{USP} and SCoV2-PLpro) with Dap also led to the selective capture of their known substrates (Extended Data Fig. 3). We conclude that replacing the active site nucleophile with Dap allows the specific covalent capture of substrates.

Profiling HtrA2 substrates

Next, we demonstrated the utility of Dap-containing proteases for the capture and identification of protease substrates from cell lysate. We focussed on the mature cytosolic form of HtrA2, which is released from the mitochondrial intermembrane space and relocated to the cytosol upon stress^{21,22}. Several substrates have been identified for cytosolic HtrA2, including caspase inhibitors²¹ (for example, XIAP and cIAP1/2). HtrA2 contains a PDZ domain and chymotrypsin-like protease domain; since PDZ domains bind a wide range of proteins and chymotrypsin-like domains commonly have broad specificity²³, we hypothesized that there may be undiscovered substrates of HtrA2.

We replaced the catalytic serine of HtrA2 with Dap, creating HtrA2(S306Dap)-HA-Strep (Supplementary Fig. 2). We incubated this protein with human cell lysate and enriched covalent conjugates to the protease by immunoprecipitation followed by stringent washing. The conjugates were eluted and immunoblotting against the Strep-tag on Dap-containing HtrA2 revealed numerous species with a higher apparent molecular mass (Fig. 1a). These species were not observed in the control experiments and are primarily covalent conjugates to Dap-containing HtrA2.

Proteomic analysis identified 274 proteins that were specifically captured by HtrA2(S306Dap); 237 of these proteins were not found in control experiments (Fig. 1b, Supplementary Data File 1) and 37 proteins were significantly enriched (minimum fold change (S_0) = 1, false discovery rate (FDR) < 0.01) in the HtrA2(S306Dap) elution with respect to controls (Fig. 1c, Supplementary Fig. 2). Our study identified 17 times more potential HtrA2 substrates than previous work, and this approach identified the majority of the proteins previously identified as HtrA2 substrates by N-terminal proteomics²⁴ (Supplementary Table 1).

We tested 29 of the newly identified potential substrates for proteolysis with wild-type HtrA2. Eighty-six percent of the candidates showed a decrease in the abundance of full-length protein or the generation of defined proteolytic fragments over time upon addition of the wild-type protease (Fig. 1d, Supplementary Fig. 3, Supplementary Tables 2, 3). Overall, our results demonstrate that we have developed an approach for the efficient and specific identification of protease substrates in complex mixtures.

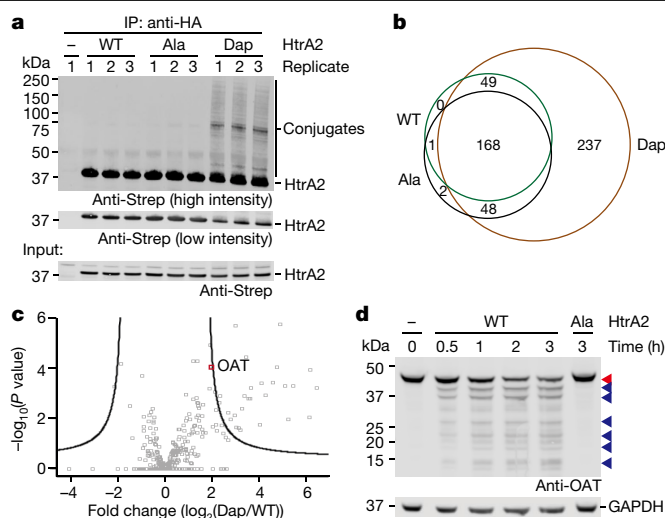


Fig. 1 | Dap-mediated HtrA2 substrate identification in mammalian cell lysate. **a**, HtrA2(S306Dap)-HA-Strep and its conjugates were enriched from cell lysate with anti-HA beads and detected with an anti-Strep antibody. Control experiments were performed with wild-type (WT) HtrA2 and the catalytically inactive S306A mutant. Input: HtrA2 variants in cell lysates before incubation. **b**, Venn diagram showing the number of proteins identified in HtrA2(S306Dap) elution compared with controls. Proteins identified in at least two of the three replicates were considered as positively identified. **c**, Volcano plot based on label-free quantification (LFQ) values for the proteins identified in HtrA2(S306Dap) and wild-type HtrA2 samples. The black line represents the cut-off curve for significance ($S_0 = 1$, FDR < 0.01). Each data point is calculated in Perseus using $n = 4$ for each HtrA2 variant. The dot representing ornithine aminotransferase (OAT) is labelled in red. **d**, Wild-type HtrA2 or HtrA2(S306A) (1 μ M) was added into Expi293 cell lysate and incubated for the indicated time at 37 °C. Red arrowhead, full-length OAT; blue arrowhead, wild-type HtrA2-dependent proteolytic fragments. GAPDH was used as a loading control. The experiment in **a** was performed in biological triplicate, and the experiment in **d** was performed in two biological replicates, both with similar results. For gel source data, see Supplementary Fig. 1.

Trapping substrates in live human cells

Next, we set out to extend our approach to capturing and identifying protease substrates in human cells (Extended Data Fig. 4a). It was unclear—on the basis of previous work²—whether pc-Dap could be converted to Dap in live cells to activate protease substrate traps. We set out to: encode pc-Dap in human cells, investigate the activation of a protease substrate trap through deprotection of pc-Dap to Dap in human cells, and demonstrate the covalent capture of protease substrates in human cells.

We demonstrated the efficient, site-specific, genetically directed incorporation of pc-Dap into proteins in human cells (Supplementary Figs. 4, 5). We produced TEV protease in which the catalytic cysteine was replaced by pc-Dap (TEV(C151pc-Dap)) in human cells (Supplementary Fig. 6). To deprotect pc-Dap we illuminated cells^{25,26} expressing TEV(C151pc-Dap) for 2 min. Mass spectrometry revealed that more than 60% of the TEV purified from cells immediately after illumination was fully deprotected to TEV(C151Dap); the remaining 40% of TEV contained the deprotection intermediate. Six hours after illumination, we detected only TEV(C151Dap) (Fig. 2a); this demonstrated that deprotection, from pc-Dap to Dap—to activate the protease substrate trap—proceeds rapidly in live cells.

To demonstrate that we can capture protease substrates in human cells, we co-expressed TEV(C151pc-Dap) and GFP-s (Fig. 2b). We illuminated HEK293T cells to generate TEV(C151Dap) and followed the formation of the TEV(Dap)-GFP conjugate in cells by immunoblotting (Fig. 2c, Supplementary Fig. 7). The conjugate was observed 30 min

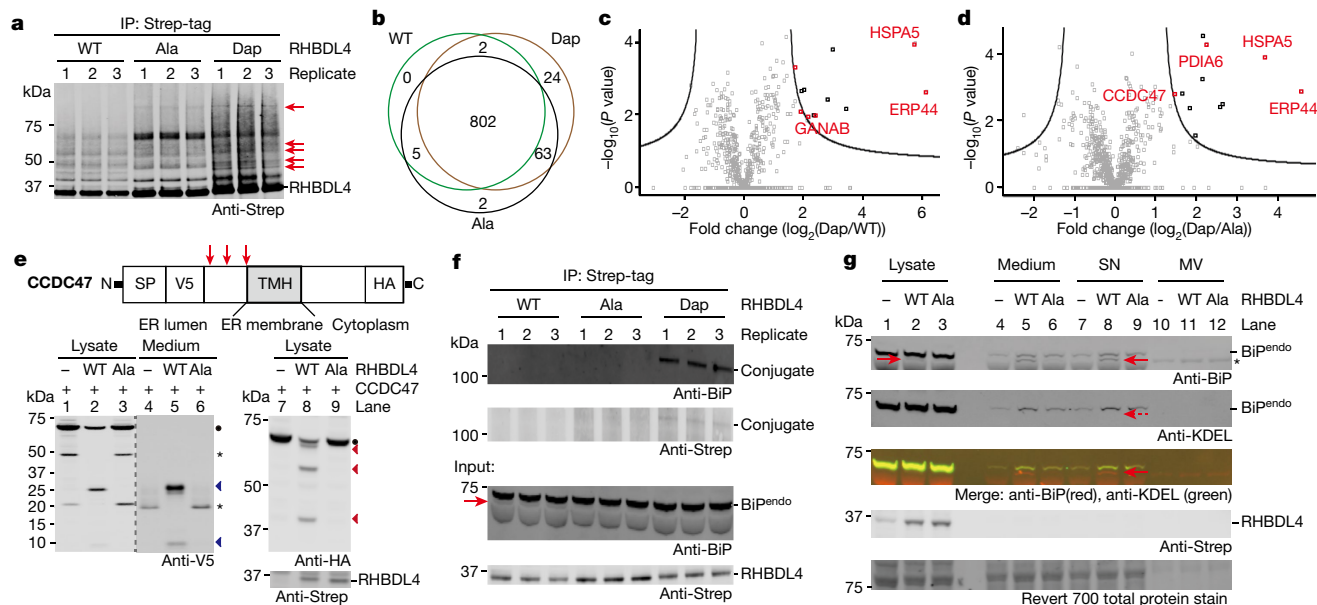


Fig. 3 | Discovery of RHBDL4 substrates. **a**, RHBDL4 variants were expressed in Expi293 cells. Immunoblotting analysis of RHBDL4 variants enriched from an equal number of cells after optical activation and substrate capture. Ala, RHBDL4(S144A); Dap, RHBDL4(S144Dap). Red arrows, RHBDL4(S144Dap)-specific higher molecular mass bands. **b**, Venn diagram showing the number of proteins identified in RHBDL4(S144Dap) elution with respect to controls. Proteins found in at least two of the three replicates were considered to be positively identified. **c, d**, Volcano plots based on the LFQ values for the identified proteins for RHBDL4(S144Dap) versus wild-type RHBDL4 (**c**; $S_0 = 1$, $FDR < 0.01$) and RHBDL4(S144Dap) versus RHBDL4(S144A) (**d**; $S_0 = 1.5$, $FDR < 0.05$). Black lines represent the cut-off curve for significance. Each point was calculated in Perseus using $n = 3$ for each RHBDL4 variant. ER-resident candidates are marked in red and select candidates are also labelled. **e**, RHBDL4 cleaves CCDC47. SP, signal peptide; V5, V5-tag; TMH, transmembrane helix; HA, HA-tag; red arrows, RHBDL4 cleavage sites; black circles, full-length CCDC47; blue triangles, N-terminal proteolytic fragments; red triangles, C-terminal

proteolytic fragments; asterisk, bands present without RHBDL4.

f, RHBDL4(S144Dap) conjugates to endogenous BiP and the cleavage of endogenous BiP by wild-type RHBDL4. Cell lysates before immunoprecipitation (input) and the conjugates were directly visualized by anti-BiP and anti-Strep antibodies after Strep-tag enrichment. Red arrow, the proteolytic fragment of endogenous BiP cleaved by wild-type RHBDL4.

g, Cleaved BiP is secreted into the medium. The medium was separated into supernatant (SN) and microvesicles (MV). Red arrow, the proteolytic fragment. Whereas full-length BiP was detected by both anti-BiP and anti-KDEL antibodies, cleaved BiP can only be detected with an anti-BiP antibody. Asterisk indicates non-specific bands. Revert 700 total protein stain was used as a loading control. Experiments in **a, f**, were performed in biological triplicate; the experiment in **e** was repeated in three biological replicates; the experiment in **g** was repeated in two biological replicates; all with similar results. For gel source data, see Supplementary Fig. 1.

Soluble proteins have not been reported as endogenous RHBDL4 substrates. However, a substantial percentage (67%) of the ER-resident proteins that we identified are soluble; the majority of these proteins (90%) are chaperones (Supplementary Table 4). We therefore focussed on further investigating the cleavage of ER soluble chaperones by RHBDL4.

We identified binding immunoglobulin protein (BiP) (encoded by *HSPA5*) as a potential substrate using our approach (Fig. 3c, d). We directly visualized the conjugates of RHBDL4(Dap) and endogenous BiP in human cells (Fig. 3f, Supplementary Fig. 12). The RHBDL4 cleavage assay confirmed that BiP can be cleaved at its interdomain linker region and C-terminus by wild-type RHBDL4. The N-terminal fragments generated by these cleavages—which do not possess the ER retention motif³⁸ (KDEL sequence from the C-terminus of BiP)—were secreted into the medium (Fig. 3g and Supplementary Fig. 13). Endogenous BiP was cleaved by exogenously expressed WT RHBDL4, and secretion of the proteolytic fragment into the medium was Brefeldin A (BFA) sensitive (Extended Data Fig. 7a), consistent with secretion by the conventional pathway. Importantly, the proteolytic fragment from endogenous BiP was detected in the extracellular medium from WT cells, but not from an RHBDL4 knockout cells (Extended Data Fig. 7b), confirming the cleavage of endogenous BiP by endogenous RHBDL4. Collectively, our data demonstrate that a fraction of the soluble BiP present in cells is a physiological substrate of endogenous RHBDL4.

RHBDL4-mediated removal of the C-terminal ER retention motif from soluble ER-resident protein candidates was also validated for protein

disulfide-isomerases (protein disulfide-isomerase A6 (gene name: *PDI6*) and ER protein 44 (gene name: *ERP44*) and calcium-binding chaperone Calreticulin (gene name: *CALR*), which were identified as potential substrates through our approach (Extended Data Fig. 8a–e). The secretion of proteolytic fragments of endogenous *PDI6* and Calreticulin was BFA sensitive (Extended Data Fig. 8f, g). The cleavage of a fraction of endogenous Calreticulin by endogenous RHBDL4 was detected in extracellular medium by an anti-Calreticulin antibody in wild-type but not in RHBDL4 knockout cells (Extended Data Fig. 8h).

Additional tests confirmed that RHBDL4 also cleaved other ER-resident soluble chaperones which were identified through our approach, including Calumenin, peptidyl-prolyl cis-trans isomerase FKBP9, Glucosidase 2 subunit α and β ; in each case the cleavage led to secretion of the resulting N-terminal fragments into extracellular media (Supplementary Fig. 14). However, RHBDL4 did not cleave Calnexin, an abundant ER-resident chaperone, which was not identified as a potential substrate by our approach (Supplementary Figs. 12 and 14); this observation is consistent with our approach specifically capturing RHBDL4 substrates.

Collectively, our results demonstrate that we can capture and identify substrates for an intramembrane protease. Moreover, we have discovered that RHBDL4 can act as a non-canonical secretase. Unlike conventional secretases—which cleave transmembrane substrates in transmembrane domains or juxtamembrane domains to release ectodomains^{16,18}—cleavage by RHBDL4 has the effect of removing the C-terminal ER-retention motif from a proportion of physiological

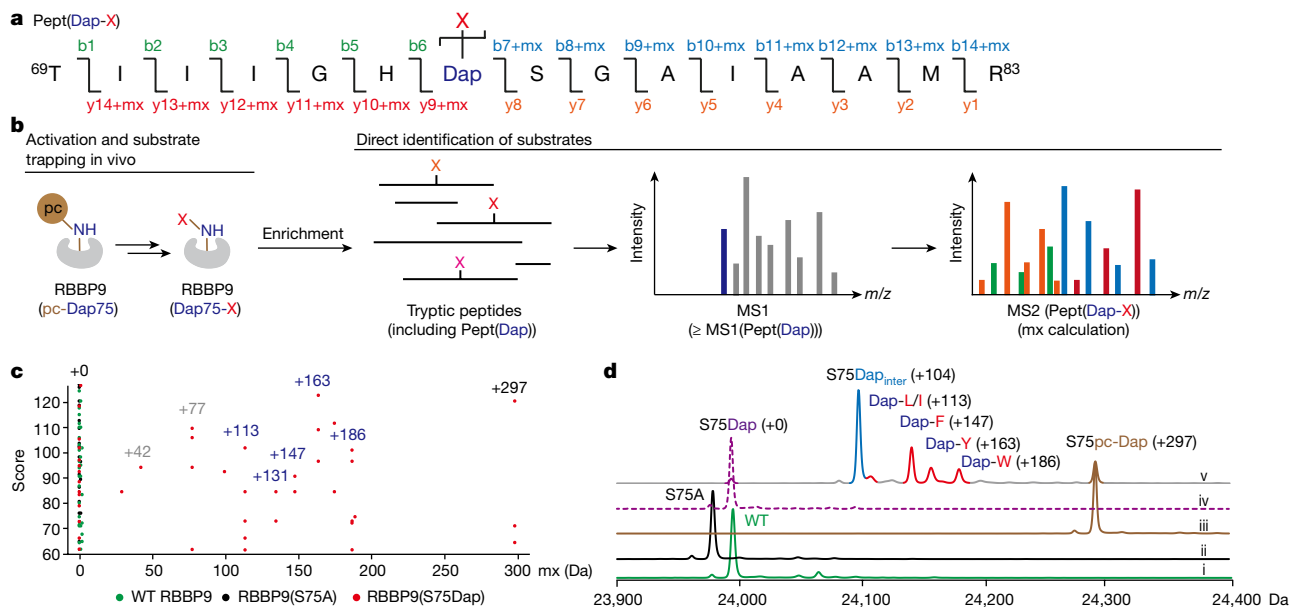


Fig. 4 | Identifying RBBP9(Dap75) conjugates. **a**, The Dap-containing tryptic peptide sequence (Pept(Dap-X)) of RBBP9 with certain b and y ion masses of Pept(Dap) modified by mx. **b**, Pipeline to identify mx in live cells. RBBP9(Dap75-X) conjugates were affinity purified and trypsinized. The resulting peptide pool was analysed by LC-MS/MS and peptides with molecular mass no less than MS(Pept(Dap)) were individually selected for mx calculation (for example, the blue peak represents Pept(Dap), mx = 0). The experimental MS2 peaks were compared to theoretical MS2 peaks for scoring as described in Methods. In the example shown, the peaks are colour coded as b, y, b+mx or y+mx ions, using the colour scheme in **a**. **c**, mx obtained from the top-scoring spectra were plotted. Each dot represents the mass shift of the observed peptide relative to the parental peptide. +113 (Leu/Ile), +131 (Met), +147 (Phe), +163 (Tyr), +186 (Trp), +297 (pc-Dap), +42 and +77 (consistent with near cognate suppression of this amber codon in mammalian cells with Gln and Tyr, respectively). **d**, The entire mass of RBBP9 variants and Dap conjugates purified from human cells. Green trace (i): wild-type RBBP9 ([Ac-(RBBP9-Met)]:

expected 23,995 Da, observed 23,994.5 Da); black trace (ii): RBBP9(S75A) ([Ac-(RBBP9(S75A)-Met)]: expected 23,979 Da, observed 23,978.5 Da); brown trace (iii): RBBP9(S75pc-Dap) before illumination ([Ac-(RBBP9(S75pc-Dap)-Met)]: expected 24,291 Da, observed 24,290.5 Da); purple dashed trace (iv): RBBP9(S75Dap) deprotected from RBBP9(S75pc-Dap) in vitro ([Ac-(RBBP9(S75Dap)-Met)]: expected 23,994 Da, observed 23,993.5 Da); multicolour trace (v): RBBP9(S75pc-Dap) purified from cells after illumination and substrate trapping for 3 h ([Ac-(RBBP9(S75Dap_{inter})-Met)]: expected 24,098 Da, observed 24,097 Da; [Ac-(RBBP9(Dap-L/I)-Met)]: expected 24,107 Da, observed 24,108 Da; [Ac-(RBBP9(Dap-F)-Met)]: expected 24,141 Da, observed 24,140.5 Da; [Ac-(RBBP9(Dap-Y)-Met)]: expected 24,157 Da, observed 24,156 Da; [Ac-(RBBP9(Dap-W)-Met)]: expected 24,180 Da, observed 24,179 Da). The experiment in **c** was performed in biological triplicate. In **d**, the entire mass acquisition of trace (v) was performed in two biological replicates with similar results, the other traces were acquired once.

substrate molecules, allowing the release of their N-terminal proteolytic fragments into the extracellular space.

Determining enzymatic function of RBBP9

Many serine hydrolases have been defined on the basis of their reactivity with activity-based probes, but their enzymatic function remains unknown³⁹. Defining the activity of these orphan hydrolases remains an outstanding challenge. We set out to address this challenge using Dap-mediated substrate trapping. We focussed on RBBP9, a tumour-associated putative serine hydrolase⁵. RBBP9 possesses a classical α/β hydrolase superfamily fold⁴⁰, and its hydrolase activity promoted tumour cell proliferation during pancreatic neoplasia⁴¹. Despite multiple efforts to characterize the cellular activity of RBBP9, its enzymatic function in cells has remained enigmatic.

Initial experiments suggested that RBBP9 is unlikely to function as an endopeptidase (Supplementary Fig. 15). We developed an approach to directly identify any entities (which we designated 'X'), conjugated to RBBP9(S75Dap), in which the catalytic serine of RBBP9 has been replaced by Dap (Fig. 4a). We first expressed RBBP9(S75Dap) in human cells and enriched any conjugates that it formed (RBBP9(Dap75-X)) in the cells, digested with trypsin, then performed LC-MS/MS on the resulting peptide pool. We then developed a computational pipeline to search for the tryptic peptides of RBBP9 containing Dap and conjugated to X (Pept(Dap-X)), based on the prediction that they should have MS2 spectra related to the corresponding non-conjugated peptide

(Pept(Dap)), in which certain fragmentation peaks are shifted by the mass of X (or the mass of a tryptic fragment of X) (mx) (Fig. 4a,b).

Our analysis revealed mass shifts of +113 Da, +131 Da, +147 Da, +163 Da and +186 Da, with respect to the mass of Pept(Dap) (Fig. 4c, Supplementary Fig. 16). These masses correspond to the amino acids leucine/isoleucine (Leu/Ile), methionine (Met), phenylalanine (Phe), tyrosine (Tyr) or tryptophan (Trp), respectively, after condensation on Dap (Supplementary Fig. 17). Additionally, we characterized the entire mass of RBBP9(Dap75-X) and further confirmed the conjugation of Leu/Ile, Phe, Tyr or Trp to RBBP9(S75Dap) (Fig. 4d).

Collectively, we demonstrated that RBBP9(S75Dap) preferentially forms conjugates with individual hydrophobic amino acids. We conclude that RBBP9 is likely to cleave amide (or ester or thioester) bonds to these hydrophobic amino acids in human cells. Amino acid sequence analysis suggested that RBBP9 belongs to DUF1234 hydrolase superfamily⁴⁰, which includes the acylpeptide hydrolase-like protein⁴² from *Arabidopsis thaliana* (AHLP; Uniprot ID: Q9FG66); this protein possesses a similar α/β hydrolase fold to RBBP9 and removes N-terminal hydrophobic residues—especially aromatic amino acids—from peptides. We therefore postulated that RBBP9 may have aminopeptidase activity in human cells.

RBBP9 is an aromatic aminopeptidase

To investigate the aminopeptidase activity of RBBP9, we performed a fluorescence-based hydrolysis assay for 19 amino acid

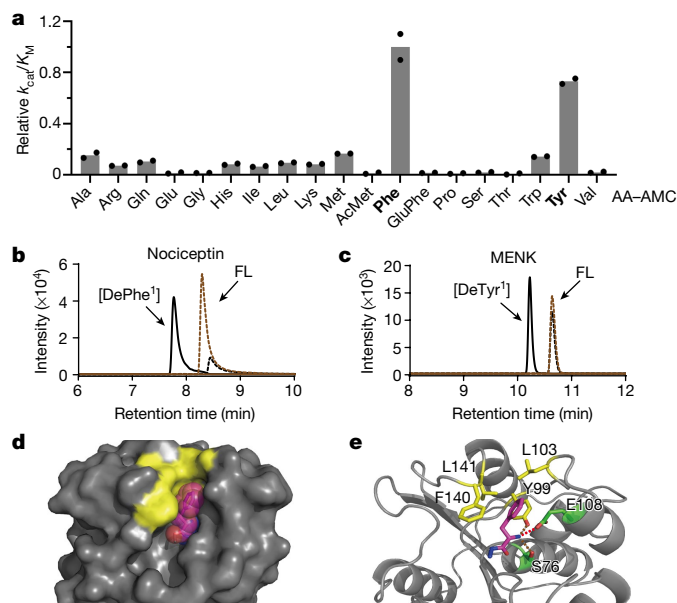


Fig. 5 | Characterization of RBBP9 aminopeptidase activity. **a**, RBBP9 is specific for aromatic amino acids. The graph shows the catalytic efficiency of RBBP9 on 19 AA-AMCs relative to its catalytic efficiency on Phe-AMC. The bar graph represents the mean of $n = 2$ independent measurements. **b**, **c**, RBBP9 cleaves aromatic amino acids from the N terminus of peptide hormones (nociceptin or MENK). The full-length (FL) peptides and the products (DePhe¹ or DeTyr¹) after incubating with wild-type RBBP9 or RBBP9(S75A) were determined by mass spectrometry. Black solid line shows detection of product after incubating with WT RBBP9; brown solid line shows detection of product after incubating with RBBP9(S75A); black dashed line shows detection of full-length peptide after incubating with wild-type RBBP9; brown dashed line shows detection of full-length peptide after incubating with RBBP9(S75A). **d**, **e**, The crystal structure of RBBP9 in complex with Phe. **d**, surface view of RBBP9 and sphere representation of Phe (purple). The surface of Tyr99, Leu103, Phe140 and Leu141 is shown in yellow. **e**, Ribbon diagram of RBBP9 with key residues shown as sticks. Side chains of Tyr99, Leu103, Phe140 and Leu141 are shown in yellow; side chains of Glu108 and Ser76 are shown in green. The hydrogen bond between the α -amino group of Phe and the side chain of Glu108 (2.7 Å) is represented by a red dashed line. The side chain of Ser76 or Tyr99 may also form hydrogen bonds with the α -amino group of Phe (dashed brown lines, bond distances of 2.8 Å or 2.9 Å, respectively).

7-amino-4-methylcoumarin (AA-AMC) compounds (Fig. 5a, Supplementary Fig. 18). RBBP9 showed a clear preference for hydrolysing aromatic residues, especially Phe and Tyr; this is generally consistent with the mass adducts identified on Dap from live cells. Moreover, RBBP9-mediated hydrolysis required a free α -amine, as acetylated Met-AMC (AcMet-AMC) and the dipeptide Glu-Phe-AMC were not hydrolysed; this explains why the aminopeptidase activity was not detected even though a subset of protease substrates (for example, Pro-pNA and Suc-Phe-pNA) were screened in previous studies⁴⁰.

The aminopeptidase activity and selectivity of RBBP9 was further validated on peptides. As expected, wild-type RBBP9 but not the catalytically inactive S75A mutant removed the first Phe (Phe¹) from neuropeptide nociceptin (Fig. 5b), the first Tyr (Tyr¹) from enkephalin (MENK) (Fig. 5c), and the first Trp (Trp¹) from fibronectin adhesion-promoting peptide, but not the first glycine (Gly¹) from a tetrapeptide (Supplementary Fig. 19, Supplementary Table 5).

To understand the preference shown by RBBP9 towards aromatic residues, we solved the crystal structure of RBBP9 in complex with Phe (Extended Data Table 1). The overall structure of this complex is almost identical (root mean squared deviation = 0.17 Å) to that of RBBP9 alone⁴⁰. Phe sits in the RBBP9 catalytic pocket (Fig. 5d). Tyr99, Leu103, Phe140 and Leu141 form a hydrophobic cage that holds the

bulky aromatic ring of Phe in position (Fig. 5d). In addition, the hydrogen bond formed between the α -amine of Phe and the side chain of Glu108 (or Ser76 or Tyr99) anchors the amino group at this position (Fig. 5e); this explains why a free α -amine is required on the N terminus of RBBP9 substrates. Overall, our results demonstrate that RBBP9 is an aminopeptidase with a preference for removing aromatic residues from the N terminus in mammalian cells. To our knowledge, this is the first reported aminopeptidase in mammals that uses a catalytic serine nucleophile to remove the N-terminal amino acid from polypeptides. Future work should aim to identify the C-terminal portion of RBBP9 substrates and understand how the hydrolase activity we have discovered relates to tumour cell proliferation.

Discussion

We have demonstrated that adding a Dap-containing protease to a complex mixture facilitates discovery of protease substrates. Extensions of this approach should facilitate substrate discovery in systems where genetic manipulation is challenging, including primary tissue samples. We have demonstrated that we can directly express and rapidly optically activate hydrolase substrate traps in live mammalian cells. As the genetic code expansion methods used to express caged hydrolase substrate traps have now been developed in several model organisms^{43,44}, including mice, future work may extend our approach to diverse physiological settings.

Hydrolases have arisen independently multiple times in evolution^{45,46}. Proteases that proceed through an acyl-enzyme intermediate naturally divide into two classes on the basis of the stereochemistry of nucleophilic attack⁴⁶. We have exemplified our approach for serine and cysteine proteases from both mechanistic classes. Proteases have also been classified into clans that have a common ancestry, as identified by structural homology⁴⁷. We have exemplified our approach for cysteine proteases from the major clan within animals (Clan PA, 66%) and viruses (Clan CA and PA, 72%), and for the major serine proteases and hydrolase clans within animals (Clans PA and SC, 70%) and viruses (Clan PA, 48%)⁴⁷ (Extended Data Fig. 9); we have also exemplified our approach for both soluble and intramembrane proteases, and provided biological insights (Supplementary Note 2). Thus, our results cover many classes of protease reaction that proceed through an acyl-enzyme intermediate and the majority of protease structural classes; this suggests that our approach will be broadly applicable.

Finally, we have demonstrated the utility of combining hydrolase substrate traps with the direct identification of Dap conjugates to define the molecular function of an orphan hydrolase. We anticipate that future work will extend the approaches we have developed to identify the activities and substrates of many other hydrolases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04414-9>.

- Klebe, G. in *Drug Design: Methodology, Concepts, and Mode-of-Action* (ed. Klebe, G.) 493–532 (Springer, 2013).
- Huguenin-Dezot, N. et al. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, 112–117 (2019).
- Freeman, M. The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu. Rev. Cell Dev. Biol.* **30**, 235–254 (2014).
- Fleig, L. et al. Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol. Cell* **47**, 558–569 (2012).
- Woitach, J. T., Zhang, M., Niu, C. H. & Thorgeirsson, S. S. A retinoblastoma-binding protein that affects cell-cycle control and confers transforming ability. *Nat. Genet.* **19**, 371–374 (1998).
- Niphakis, M. J. & Cravatt, B. F. Enzyme inhibitor discovery by activity-based protein profiling. *Annu. Rev. Biochem.* **83**, 341–377 (2014).

7. Sanman, L. E. & Bogyo, M. Activity-based profiling of proteases. *Annu. Rev. Biochem.* **83**, 249–273 (2014).
8. De Cesare, V. et al. Deubiquitinating enzyme amino acid profiling reveals a class of ubiquitin esterases. *Proc. Natl Acad. Sci. USA* **118**, e2006947118 (2021).
9. Lopez-Otin, C. & Overall, C. M. Protease degradomics: a new challenge for proteomics. *Nat. Rev. Mol. Cell Biol.* **3**, 509–519 (2002).
10. Rei Liao, J. Y. & van Wijk, K. J. Discovery of AAA+ protease substrates through trapping approaches. *Trends Biochem. Sci.* **44**, 528–545 (2019).
11. Saita, S. et al. PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. *Nat. Cell Biol.* **19**, 318–328 (2017).
12. Knopf, J. D. et al. Intramembrane protease RHBDL4 cleaves oligosaccharyltransferase subunits to target them for ER-associated degradation. *J. Cell Sci.* **133**, jcs243790 (2020).
13. Dix, M. M., Simon, G. M. & Cravatt, B. F. Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* **134**, 679–691 (2008).
14. Zhou, J. et al. Deep profiling of protease substrate specificity enabled by dual random and scanned human proteome substrate phage libraries. *Proc. Natl Acad. Sci. USA* **117**, 25464–25475 (2020).
15. Luo, S. Y., Araya, L. E. & Julien, O. Protease substrate identification using N-terminomics. *ACS Chem. Biol.* **14**, 2361–2371 (2019).
16. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's disease. *Annu. Rev. Neurosci.* **34**, 185–204 (2011).
17. Crawford, E. D. & Wells, J. A. Caspase substrates and cellular remodeling. *Annu. Rev. Biochem.* **80**, 1055–1087 (2011).
18. Wolfe, M. S. Intramembrane proteolysis. *Chem. Rev.* **109**, 1599–1612 (2009).
19. Beard, H. A., Barniol-Xicota, M., Yang, J. & Verhelst, S. H. L. Discovery of cellular roles of intramembrane proteases. *ACS Chem. Biol.* **14**, 2372–2388 (2019).
20. Tinzl, M. & Hilvert, D. Trapping transient protein species by genetic code expansion. *ChemBioChem* **22**, 92–99 (2021).
21. Suzuki, Y. et al. A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. *Mol. Cell.* **8**, 613–621 (2001).
22. Chao, J. R. et al. Hax1-mediated processing of HtrA2 by Parl allows survival of lymphocytes and neurons. *Nature* **452**, 98–102 (2008).
23. Martins, L. M. et al. Binding specificity and regulation of the serine protease and PDZ domains of HtrA2/Omi. *J. Biol. Chem.* **278**, 49417–49427 (2003).
24. Vande Walle, L. et al. Proteome-wide identification of HtrA2/Omi substrates. *J. Proteome Res.* **6**, 1006–1015 (2007).
25. Courtney, T. & Deiters, A. Recent advances in the optical control of protein function through genetic code expansion. *Curr. Opin. Chem. Biol.* **46**, 99–107 (2018).
26. Wang, J. et al. Time-resolved protein activation by proximal decaging in living systems. *Nature* **569**, 509–513 (2019).
27. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* **100**, 391–398 (2000).
28. Wolfe, M. S. & Kopan, R. Intramembrane proteolysis: theme and variations. *Science* **305**, 1119–1123 (2004).
29. Muller, S. A., Scilabra, S. D. & Lichtenthaler, S. F. Proteomic substrate identification for membrane proteases in the brain. *Front. Mol. Neurosci.* **9**, 96 (2016).
30. Dickey, S. W., Baker, R. P., Cho, S. & Urban, S. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**, 1270–1281 (2013).
31. Urban, S., Lee, J. R. & Freeman, M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**, 173–182 (2001).
32. Lemberg, M. K. & Freeman, M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res.* **17**, 1634–1646 (2007).
33. Song, W. et al. Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nat. Commun.* **6**, 8022 (2015).
34. Wan, C. et al. Exosome-related multi-pass transmembrane protein TSAP6 is a target of rhomboid protease RHBDD1-induced proteolysis. *PLoS ONE* **7**, e37452 (2012).
35. Wunderle, L. et al. Rhomboid intramembrane protease RHBDL4 triggers ER-export and non-canonical secretion of membrane-anchored TGF α . *Sci. Rep.* **6**, 27342 (2016).
36. Paschkowsky, S., Hamze, M., Oestereich, F. & Munter, L. M. Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDL4. *J. Biol. Chem.* **291**, 21903–21912 (2016).
37. Chitwood, P. J. & Hegde, R. S. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630–634 (2020).
38. Munro, S. & Pelham, H. R. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* **48**, 899–907 (1987).
39. Jessani, N. et al. A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods* **2**, 691–697 (2005).
40. Vorobiev, S. M. et al. Crystal structure of human retinoblastoma binding protein 9. *Proteins* **74**, 526–529 (2009).
41. Shields, D. J. et al. RBBP9: a tumor-associated serine hydrolase activity required for pancreatic neoplasia. *Proc. Natl Acad. Sci. USA* **107**, 2189–2194 (2010).
42. Tsuji, A., Fujisawa, Y., Mino, T. & Yuasa, K. Identification of a plant aminopeptidase with preference for aromatic amino acid residues as a novel member of the prolyl oligopeptidase family of serine proteases. *J. Biochem.* **150**, 525–534 (2011).
43. Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
44. Brown, W., Liu, J. & Deiters, A. Genetic code expansion in animals. *ACS Chem. Biol.* **13**, 2375–2386 (2018).
45. Dodson, G. & Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347–352 (1998).
46. Buller, A. R. & Townsend, C. A. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc. Natl Acad. Sci. USA* **110**, E653–E661 (2013).
47. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509 (2014).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Plasmid construction

Standard molecular biology techniques, including PCR, restriction cloning, Gibson assembly, Golden gate assembly, and quik-change mutagenesis were applied to assemble plasmids. To generate plasmids for protein expression in *E. coli*, the DNA fragment encoding human HtrA2(134–458), RBBP9 or SCoV2-PLpro was synthesized as a double stranded DNA (Integrated DNA Technologies (IDT)), and UL36^{USP} (UL36(39–285)) and UL36^{USP}(C65S) were amplified from UL36^{USP} containing plasmids⁴⁸. The encoding sequence was cloned into pNHD vector² with a C-terminal HA–Strep-tag for HtrA2 and RBBP9, and a C-terminal twin-Strep-tag for UL36^{USP} and SCoV2-PLpro. To convert catalytic serine/cysteine to alanine or amber stop codon, site directed mutagenesis was completed using quik-change primers (Agilent primer design). To generate vectors for protein expression in mammalian cells, DapRST², TEV², human RHBDL4 (ref. ⁴⁹) or RBBP9 encoding sequence was amplified and introduced into the previously reported pcDNA3.1 plasmid backbone⁵⁰ with 4 × [U6-PyltRNA_{CUA}]. A C-terminal HA–Strep-tag was introduced to TEV, and a twin-Strep-tag was introduced at the C-terminus of RHBDL4, RBBP9, UL36^{USP} or SCoV2-PLpro. The DNA fragments encoding ER-resident proteins—human CCDC47, PDIA6, CALR, GANAB, ERP44, CALU, PRKCSH, FKBP9, DNAJC3 and CANX—were amplified from HEK293T cDNA and cloned into a previously reported pcDNA3.1-based BiP expressing plasmid⁵¹, in which an ER leader peptide was followed by a V5-tag and BiP encoding sequence. An additional HA-tag was introduced at the C-terminus of CCDC47, or before the ER retention motif sequence of ERP44. The DNA fragments encoding MFN2, LEMD2, EMD, HNRNPH1 and HNRNPM were amplified from HEK293T cDNA and cloned into a pcDNA3.1 plasmid. An HA-tag was placed at the C-terminus of MFN2, LEMD2 or HNRNPM, while a GFP-tag was introduced at the C-terminus of Emerin or HNRNPH1 for better detection of the proteolytic fragments. The guide RNA (gRNA) for RHBDL4 knockout was introduced into pX330-puro plasmid with the optimized scaffold⁵² via Golden gate assembly.

Western blot

Samples were separated by SDS–PAGE (note that NuPAGE 4–12%, 10% Bis-Tris or 3–8% Tris-Acetate gels running in MES or MOPS buffer were applied to achieve the optimal separation of proteins (protein fragments) of interest) and transferred to polyvinylidene difluoride (PVDF) membrane by iBlot 2 dry blotting system (Thermo Fisher Scientific). Membrane was blocked by Odyssey blocking buffer in PBS (catalogue (cat.) no. 927-40000, Li-Cor) at room temperature for 30 min. Membrane was incubated in primary antibody solution (dilution according to manufacturer's instructions in Odyssey T20 (PBS) antibody diluent (927-75001, Li-Cor)) at 4 °C overnight. All incubations were carried out on a platform shaker. The membrane was washed three times with PBST (PBS supplemented with 0.1% Tween-20 (v/v)), and incubated with the secondary antibody solution (1: 15,000 (v/v) in PBS blocking buffer supplemented with 0.2% Tween-20 (v/v), and 0.01% SDS) at room temperature for 1 h. After washing 3 times with PBST and once with PBS, the immunoreactive proteins were visualized by the Odyssey CLx imaging system (Li-Cor) by scanning at 700 nm and/or 800 nm channels. Revert 700 Total Protein Stain (926–11015, Li-Cor) was used for total protein staining. The data were analysed by Image Studio Lite (version 5.2.5). For primary and secondary antibodies used in this study, see 'Antibodies' in Supplementary Methods.

Deprotection of pc-Dap containing proteins in buffer

To activate protease (pc-Dap), proteins were illuminated (365 nm, 4 mW cm⁻²) for 1 min in Tris buffer (5 mM DTT, pH 8.0) and incubated at 4 °C or 37 °C overnight to generate protease (Dap). MIC-LED-365 (500 mA, Prizmatix collimated modular Mic-LED light source, Supplementary Fig. 2) was used for illuminating proteins in solution. This

apparatus was also used for illuminating suspension cells (Expi293 cells) in tissue culture hood.

HtrA2 substrate trapping in cell lysate

Thirty micrograms HtrA2–HA–Strep variant (wild type, Ala or Dap) was added to 1 ml of Expi293 cell lysate (3 μg μl⁻¹) and incubated at 30 °C for 3 h. The reaction was shaken 10 s every 10 min. Fifty microlitres of anti-HA agarose slurry (A2095, Merck) was added to the reaction and mixed at 4 °C on an end-over-end rocker for 2 h. The mixture was transferred to a Bio-spin column. The resin was washed with RIPA buffer 3 times and PBST buffer 3 times using a vacuum pump, followed by centrifugation at 5,000g for 1 min to remove the residual buffer. Then, beads were incubated in 100 μl 1 × LDS loading buffer and boiled at 95 °C for 5 min. Twenty microlitres of the eluate was analysed by SDS–PAGE or western blot. Twenty microlitres of eluate was separated in a Bolt 10% Bis-Tris Plus gel for 3 min at 200 V. Gel slices containing all proteins were cut and analysed by LC–MS/MS as described in 'Electrospray ionization tandem mass spectrometry'.

Validation of HtrA2 substrates in cell lysate

Wild-type HtrA2 or HtrA2(S306A) (1 μM) was added to 1.2 ml Expi293 cell lysate. At indicated time points, 300 μl of reaction was quenched by mixing with 100 μl 4 × LDS loading buffer and boiled at 95 °C for 5 min. The samples were analysed by western blot with primary antibodies listed in Supplementary Tables 2, 3. GAPDH was used as a loading control.

Incorporation of pc-Dap in HEK293T cells

HEK293T cells were purchased from European Collection of Cell Cultures (authenticated by STR DNA profiling) and were tested negative for *Mycoplasma* contamination.

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) and Penicillin-Streptomycin (Pen/Strep, 100 IU ml⁻¹ penicillin and 100 μg ml⁻¹ streptomycin) at 37 °C in a humidified incubator supplied with 5% CO₂. Cells were passaged every 2–3 days by detaching with trypsin–EDTA solution, resuspended in DMEM with 10% FBS, and seeded into cell culture flasks.

For transfection in a 24-well plate: 0.75 μl of Lipofectamine 3000 (Thermo Fisher) was diluted in 25 μl Opti-MEM (Gibco) and vortexed briefly. DNA solution was prepared by mixing 500 ng DNA mixture (substrate:DapRST:protease, 1:1:3 or empty vector:DapRST:Protease, 1:1:3) in 25 μl Opti-MEM, followed by adding 1 μl of P3000 reagent. Then, diluted Lipofectamine was added to DNA solution (1:1 v/v). The mixture was incubated at room temperature for 10 min and the DNA–lipid complexes were added to cells. Indicated concentrations in figure legends (or 0.5 mM) of pc-Dap was added to the culture medium 30 min after transfection to achieve pc-Dap incorporation. Cells were incubated at 37 °C for 40–48 h before further analysis.

Incorporation of pc-Dap in Expi293 cells

Expi293 cells were purchased from Thermo Fisher (authenticated by STR DNA profiling) and were tested negative for *Mycoplasma* contamination.

Expi293 cells were cultured in Expi293 media (Gibco) and shaken at 125 rpm in incubator at 37 °C with 8% CO₂. Cells were passaged every 2–3 days, starting with the cell density around 0.5 × 10⁶ cells per ml. Transfection was performed at cell density around 2.5 × 10⁶ cells per ml.

Transfection of 100 ml Expi293 cells: 300 μl of polyethylenimine molecular mass 40,000 (PEI, 1 mg ml⁻¹, Polysciences) was diluted in 3.3 ml Expi293 medium. 100 μg DNA mixture (substrate:DapRST:protease, 1:1:3 or empty vector:DapRST:protease, 1:1:3) was diluted in 3.3 ml Expi293 media. Diluted DNA and PEI solution were mixed and incubated at room temperature for 15 min before adding to the cell culture. 0.5 mM (or indicated concentrations in figure legends) of pc-Dap was

added 30 min after transfection for pc-Dap incorporation. Forty to forty-eight hours after transfection, the cells were collected and photoactivated for further analysis.

Photoactivation of protease(pc-Dap) and substrate trapping in mammalian cells

Forty to forty-eight hours after transfection, cell culture medium was replaced with fresh medium, and cells were illuminated for 2 min. The apparatus for illuminating adherent mammalian cells was built in-house (Supplementary Fig. 7). LuxiGen 365 nm UV LED Emitter (LZ4-04UVOR-0000, Mouser Electronics) was used for illumination. The UV intensity at the well plate was set at 4 mW cm⁻². After illumination, cells were incubated at 37 °C for indicated period of time (Proteasome inhibitor MG132 (2 µM) was added if needed). For adherent cells, at each time point, cells in a 6-well plate were washed with PBS and lysed in 400 µl RIPA lysis buffer (89900, Thermo) supplemented with Halt Protease Inhibitor Cocktail (78429, Thermo Fisher) and the Universal Nuclease (88702, Thermo Fisher). The lysate was cleared by centrifuging at 21,000g for 5 min and the supernatant was flash frozen and stored at -80 °C. For suspension cells, at each indicated time point, 5 ml cell culture was centrifuged at 650g for 5 min and the cell pellet was flash frozen and kept at -80 °C. Then, cell pellets were lysed in 1 ml RIPA lysis buffer supplemented with protease inhibitors and the Universal Nuclease at 4 °C. Cell lysates were centrifuged at 21,000g for 5 min. The cleared lysates were used for SDS-PAGE and western blot analysis or affinity enrichment by MagStrep type3 XT beads (2-4090-002, IBA).

Trapping endogenous substrates to RHBDL4(S144Dap)

RHBDL4 variants were expressed in Expi293 cells as described in 'Incorporation of pc-Dap in Expi293 cells'. Forty hours after transfection, 50 ml cell culture was resuspended in fresh Expi293 media and illuminated (365 nm, 4 mW cm⁻²) for 2 min. Cells were incubated at 37 °C for 4 h in the presence of 2 µM MG132 before collection. After pelleting, cells were resuspended in HEPES buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM MgCl₂, 5% glycerol, 1 mM DTT) supplemented with protease inhibitors and the Universal Nuclease. The suspension was lysed by passing twice through an Avestin Emulsiflex C3 homogenizer (ATA Scientific) at 3,000–5,000 psi. The lysate was centrifuged at 1,000g for 5 min twice and the supernatant was further centrifuged at 100,000g for 1 h. The pellet was washed with Na₂CO₃ (100 mM, pH 11.3) at 4 °C for 20 min and then centrifuged at 140,000g for 1 h. The membrane fraction was dissolved in 2% SDS buffer (50 mM Tris, pH 8, 150 mM NaCl, 1 mM DTT and protease inhibitors). The solution was diluted by 10% NP40 to generate a final concentration of 0.1% SDS and 1% NP40. One-hundred microlitres MagStrep type3 XT beads were added to the solution and incubated at room temperature for 1 h. The beads were washed with RIPA and PBST three times each. Proteins attached to the beads were eluted in 1× LDS loading buffer by heating at 65 °C for 15 min. The eluates were separated in a Bolt 10% Bis-Tris Plus gel for 3 min. Gel slices containing proteins were cut and analysed by LC-MS/MS.

RHBDL4 cleavage assay

Empty vector, wild-type (WT) RHBDL4 or RHBDL4(S144A) plasmid was co-transfected with candidate substrate containing plasmid or empty vector (for endogenous substrates) into HEK293T or Expi293 cells. The amount of candidate substrate-containing plasmid was optimized for expression level. Forty to forty-eight hours after transfection, cells were collected and lysed in RIPA buffer supplemented with protease inhibitors and the Universal Nuclease. The lysate was cleared and analysed by western blot.

To analyse proteins in the extracellular medium, FBS-containing medium for HEK293T cells was replaced with hybridoma serum free medium (12045076, Thermo Fisher) 24 h before collection. Expi293 medium, which is serum-free and protein-free, can be directly collected for further analysis. The medium was collected and filtered through a

0.22 µm polyethersulfone membrane. To obtain total proteins in the medium, 1/10 volume of 100% ice-cold TCA solution (T0699, Sigma) was added at 4 °C for protein precipitation. To obtain proteins in the supernatant, the medium was centrifuged at 200,000g for 1 h to separate supernatant from microvesicles. The SN was collected and added 1/10 volume of ice-cold TCA to precipitate proteins. After centrifuging at 21,000g for 10 min, the precipitate was washed once with acetone, and dissolved in 1× LDS loading buffer. The microvesicles pellet after ultra-centrifugation was washed with PBS once and dissolved in equal volume of 1× LDS loading buffer.

Deglycosylation was performed by adding 1/10 volume of 10% NP40 and PNGase (P0704S, NEB) or DeGlycosylation mix II (P6044S, NEB) to proteins dissolved in the 1× LDS loading buffer. The reaction was incubated at 37 °C for 1 h before analysis.

Brefeldin A inhibitory assay

Twenty-four hours after transfection, Expi293 cells were split into two halves treated separately with DMSO or BFA (5 µg ml⁻¹). The BFA treatment was performed in two ways: (1) BFA was directly added into medium culture; (2) the medium culture was replaced with fresh medium supplemented with BFA. Sixteen hours after BFA treatment, the cells and extracellular medium were collected and analysed as described in 'RHBDL4 cleavage assay'.

Knockout of RHBDL4

HCT116 cells were purchased from American Type Culture Collection (authenticated by STR DNA profiling) and were tested negative for *Mycoplasma* contamination.

HCT116 cells were cultured in McCoy's 5A (modified) Media (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco) and Pen/Strep at 37 °C in humidified incubator with 5% CO₂. Cells were passaged every 2–3 days.

HCT116 cells in 6-well plates were transfected by Lipofectamine LTX (15338100, Thermo Fisher) with 2.5 µg of pX330-puro plasmid containing gRNA (5'-TCCAGTAAGTACAGAAAATG-3') and Cas9 for RHBDL4 knockout. Twenty-four hours after transfection, cells were trypsinized and plated in a 10 cm petri dish. After 24 h, the cells were treated with puromycin (1 µg ml⁻¹). The puromycin selection stopped after 48 h. Cells were trypsinized and limited dilution was performed to generate single clones, which were expanded and analysed by western blot (anti-RHBDL4) and genotyped by sequencing the genomic DNA region targeted by the gRNA.

To detect the proteolytic fragments from endogenous substrates generated by endogenous RHBDL4, 10 million wild-type or RHBDL4 knockout HCT116 cells were cultured in hybridoma serum free medium for 40 h. The medium was collected, filtered and concentrated by a 30 kDa cut-off concentrator. Proteins were precipitated by TCA and dissolved in 1× LDS loading buffer for immunoblotting analysis.

Trapping endogenous substrates to RBBP9(S75Dap)

To identify X attached to Pept(Dap), RBBP9 variants were expressed in HEK293T cells as described in 'Incorporation of pc-Dap in HEK293T cells'. pc-Dap (0.1 mM) was added to cells to produce RBBP9(S75pc-Dap). To characterize the entire masses of RBBP9 variants, RBBP9 variants were produced in 100 ml Expi293 cells as described in 'Incorporation of pc-Dap in Expi293 cells'. RBBP9(S75pc-Dap) was expressed in the presence of 0.5 mM pc-Dap. 40 h after transfection, cells were illuminated (365 nm, 4 mW cm⁻²) for 2 min, and incubated at 37 °C for 3 h. Cells were then collected and lysed in Tris buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA and Universal Nuclease) by sonication. Note that protease inhibitors were not added in lysis buffer. The lysate was cleared by centrifuging at 21,000g for 20 min. RBBP9 species in the supernatant were enriched using MagStrep type3 XT beads. Proteins attached to beads were eluted in 50 mM Biotin in Tris buffer for mass characterization.

Aminopeptidase assay of RBBP9

Fluorescence-based assay. Two micromolar RBBP9 was incubated with each AA-AMC over a range of different substrate concentrations in Tris buffer (100 mM Tris, 150 mM NaCl, pH 7.3). Fluorescence intensity (due to the release of the AMC fluorophore by hydrolysis of AA-AMC by RBBP9) was measured every 20 s over a 10-min period (MARS Data Analysis Software (version 3.20 R2)). For each substrate, the rate of fluorescence increase was converted to rate of product formation using standard curves. At substrate concentrations of greater than 10 μM , intermolecular quenching of AMC fluorescence by AA-AMC was found to be significant. Therefore, for all substrates other than Phe-AMC and Tyr-AMC, AA-AMC concentrations between 0 and 4 μM were used, and pseudo-first order kinetics were employed to calculate specificity constants. For Phe-AMC and Tyr-AMC, which showed significantly faster rates of hydrolysis when compared to the other substrates, a concentration range of 0 to 160 μM was used and converted rates were fitted to Michaelis–Menten kinetics in order to obtain specificity constants.

Peptide-based assay. Peptides (100 μM) were dissolved in Tris buffer (100 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA). Two micromolar wild-type RBBP9 or RBBP9(S75A) was added to start the hydrolysis reaction. The reaction was stopped by quenching with acetic acid and monitored by mass spectrometry. Selected ion mass (SIM) mode was applied for detection of peptide substrates and the desired products.

Protein crystallization and data collection

Human RBBP9 with a C-terminal His-tag⁴⁰ (LEHHHHHH) was expressed in BL21(DE3) cells and purified by a two-step protocol consisting of HisTrap enrichment and gel filtration (Superdex 75) chromatography. Pure fractions of RBBP9 (>98% purity determined by SDS–PAGE) were concentrated with a 10 kD MWCO Vivaspinn 20 concentrator (Sartorius) to 10 mg ml⁻¹ in buffer containing 10 mM Tris (pH 7.5), 100 mM NaCl, 5 mM DTT and 5 mM Phe. Prior to crystallization, samples were cleared by centrifugation for 15 min at 10,000g. Crystallization trials with multiple commercial crystallization kits were performed in 96-well sitting-drop vapor diffusion plates (Molecular Dimensions) at 18 °C and set up with a mosquito HTS robot (TTP Labtech). Drop ratios of 0.2 μl protein solution plus 0.2 μl reservoir solution were used for coarse and fine screening. Initial hits were obtained under multiple conditions and required no further optimization. Data was collected from crystals collected from following conditions: 30% w/v PEG 4K, 0.1 M MES sodium salt, pH 6.5.

To ensure cryo-protection, crystal-containing drops were mixed with 25% glycerol in reservoir solution prior to picking and flash frozen in liquid nitrogen. Diffraction data was collected at the Diamond Light Source (DLS, UK) on beamline I04. Datasets were auto-processed with XIA2 DIALS (version 0.7.90), scaled using Aimless and Refmac5 (version 5.8.0258) in the CCP4 suite (version 7.0.078) of programs. Structure refinement and manual model building were performed with Refmac5 and COOT (version 0.8.9.2). Colour figures were prepared with PyMol (version 2.5).

Mass characterization

Electrospray ionization mass spectrometry. Mass spectra of all protein samples were acquired on an Agilent 1200 LC-MS system equipped with a 6130 Quadrupole spectrometer. A Phenomenex Jupiter C4 column (150 \times 2 mm, 5 μm) was used to elute proteins. Buffer A (0.2% formic acid in H₂O) and buffer B (0.2% formic acid in acetonitrile) was used for RP-HPLC. Mass spectra were acquired in the positive mode and analysed by the MS Chemstation software (Rev.C.01.06[61], Agilent Technologies). The deconvolution program provided in the software was used to obtain the entire mass spectra. Theoretical molecular mass of proteins with non-canonical amino acids was calculated by correcting the calculated molecular mass of wild-type protein

(<http://www.peptidesynthetics.co.uk/tools/>) with the molecular mass of non-canonical amino acids.

Electrospray ionization tandem mass spectrometry. Proteins (including TEV-GFP conjugate, substrates trapped to HtrA2 or RHBDL4) in polyacrylamide gel slices (1–2 mm) were enzymatically digested in situ for LC–MS/MS analysis. In brief, the excised protein gel pieces were placed in a 96-well microtitre plate and destained with 50% v/v acetonitrile and 50 mM ammonium bicarbonate, followed by reduction with 10 mM DTT and alkylation with 55 mM iodoacetamide. RBBP9 eluates in solution were treated in two ways: (1) incubation at room temperature overnight in the presence of 10 mM DTT without alkylation; (2) reduction with 10 mM DTT for 30 min and alkylation with 55 mM iodoacetamide. Then, proteins were digested with trypsin/LysC (Promega) overnight at 37 °C. The resulting peptides were extracted in 2% v/v formic acid, 2% v/v acetonitrile and analyzed by nanoscale capillary LC-MS/MS, which uses an Ultimate U3000 HPLC (ThermoScientific Dionex) with a flow rate of 300 nl min⁻¹. A C18 Acclaim PepMap100 5 μm , 100 μm \times 20 mm nanoViper (ThermoScientific Dionex) was used to trap the peptides before separation on a C18 Acclaim PepMap100 3 μm , 75 μm \times 150 mm nanoViper (ThermoScientific Dionex). Peptides were eluted with a gradient of acetonitrile. The eluate was directly introduced to a modified nanoflow ESI source with a hybrid dual pressure linear ion trap mass spectrometer (Orbitrap Velos, ThermoScientific). Data-dependent analysis was carried out using a resolution of 30,000 for the full MS spectrum, followed by ten MS/MS spectra in the linear ion trap. MS spectra were collected over an *m/z* range of 100–2,000.

LC–MS/MS data analysis by Venn diagram. LC–MS/MS data were searched against an in-house protein sequence database containing Swiss-Prot and the protein constructs specific to the experiment, using the Mascot search engine program (Matrix Science, version 2.4). Database search parameters were set with a precursor tolerance of 5 p.p.m. and a fragment ion mass tolerance of 0.8 Da. Variable modifications for oxidized methionine, carbamidomethyl cysteine, pyroglutamic acid, and deamination of glutamine/asparagine were included. MS/MS data were validated using the Scaffold program (version 5, Proteome Software Inc.).

LC–MS/MS data analysis by volcano plot. For quantitative analysis, MS raw files were processed by MaxQuant software (version 1.6.3.4) and searched with the embedded Andromeda search engine against the corresponding database (Uniprot). The required FDR was set to 1% or 5% at peptide and protein levels. The maximum number of allowed missed cleavages was set to two. Protein quantification was done by LFQ with default settings. The MaxQuant ProteinGroups output file was further processed with Perseus (version 1.6.2.3)⁵³. Contaminations and reverse hits were removed by filtering. The remaining protein quantifications were log₂-transformed.

Determination of X attached to Dap in RBBP9

LC–MS/MS files (in RAW format) were first converted to mzML format⁵⁴ using ProteoWizard (version 3.0.11252)⁵⁵. Data preparation and processing were then performed using custom Python (version 3.8.1) scripts written with the pyOpenMS package (version 2.4.0)⁵⁶. In brief, collected spectra were centroided and all MS2 spectra with a precursor mass lower than that of the unconjugated Dap-containing tryptic peptide from RBBP9 (Pept(Dap)) were filtered out. For each filtered MS2 spectrum, the ten most abundant peaks in each 100 Th mass interval were extracted.

Based on the peptide sequence of Pept(Dap) and the precursor mass for each MS2 spectrum, a list of theoretical ion masses was calculated; these corresponded to the MS2 fragmentation of a substrate-conjugated Pept(Dap), (Pept(Dap-X)). This list contained the monocationic b- and y-ions, the dicationic b- and y-ions, and ions

corresponding to water or ammonium losses from the side-chains of b- or y- ions. Peaks in the MS2 spectrum were matched against this list, and a score for this matching was calculated as previously described⁵⁷. This score was ten times the negative logarithm of the approximate probability that at least k out of n masses have been matched by chance, where k is the number of matches and n is the number of masses in the list.

To extract the top-scoring spectra, the family-wise error rate for the probability values was controlled at 0.05 using the Bonferroni correction. The mass difference between Pept(Dap) and the precursor ion for each Pept(Dap-X) spectrum was calculated to determine the molecular mass of each conjugate. For each mass shift, representative top-scoring spectra were manually interrogated to verify the assignment.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The structure of RBBP9 in complex with Phe is available in the Protein Data Bank under accession code 7OEX. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PDX030381. All other datasets and materials generated or analysed in this study are available from the corresponding authors upon reasonable request. The data used to analyse serine and cysteine proteases clans were downloaded from the MEROPS database (<https://www.ebi.ac.uk/merops/>). Source data are provided with this paper.

Code availability

The code used for RBBP9 substrate identification by proteomic analysis is available at <https://doi.org/10.5281/zenodo.5768340>.

- MacDonald, C., Winistorfer, S., Pope, R. M., Wright, M. E. & Piper, R. C. Enzyme reversal to explore the function of yeast E3 ubiquitin-ligases. *Traffic* **18**, 465–484 (2017).
- Ikeda, K. N. & Freeman, M. Spatial proteomics reveal that the protein phosphatase PTP1B interacts with and may modify tyrosine phosphorylation of the rhomboid protease RHBDL4. *J. Biol. Chem.* **294**, 11486–11497 (2019).
- Schmied, W. H., Elsasser, S. J., Uttamapinant, C. & Chin, J. W. Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* **136**, 15577–15583 (2014).
- Cesaratto, F. et al. BiP/GRP78 mediates ERAD targeting of proteins produced by membrane-bound ribosomes stalled at the STOP-codon. *J. Mol. Biol.* **431**, 123–141 (2019).
- Maule, G. et al. Allele specific repair of splicing mutations in cystic fibrosis through AsCas12a genome editing. *Nat. Commun.* **10**, 3556 (2019).
- Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
- Martens, L. et al. mzML—a community standard for mass spectrometry data. *Mol. Cell Proteomics* **10**, R110.000133 (2011).

- Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
- Rost, H. L., Schmitt, U., Aebersold, R. & Malmstrom, L. pyOpenMS: a Python-based interface to the OpenMS mass-spectrometry algorithm library. *Proteomics* **14**, 74–77 (2014).
- Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
- Kattenhorn, L. M., Korb, G. A., Kessler, B. M., Spooner, E. & Ploegh, H. L. A deubiquitinating enzyme encoded by HSV-1 belongs to a family of cysteine proteases that is conserved across the family Herpesviridae. *Mol. Cell* **19**, 547–557 (2005).
- Shin, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **587**, 657–662 (2020).
- Patchett, S. et al. A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell Rep.* **36**, 109754 (2021).
- Klemm, T. et al. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **39**, e106275 (2020).
- Liu, G. et al. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nat. Microbiol.* **6**, 467–478 (2021).

Acknowledgements This work was supported by the UK Medical Research Council (MRC; MC_U105181009 and MC_UP_A024_1008) to J.W.C.; MC_U105192713 to M. Bienz (M. Fiedler's supervisor), and Wellcome Trust Investigator Awards (101035/Z/13/Z and 220887/Z/20/Z) to M. Freeman. S.T. was supported by a European Molecular Biology Organization Long-Term Fellowship (ALTF 825-2016). L.K. (no. 884511) and G.P. (no. 897663) were supported by Marie Skłodowska-Curie European Postdoctoral Fellowships. We thank M. K. Lemberg for providing the pTa construct; N. Gomez-Navarro for sharing the UL36^{USP} containing plasmids; O. R. Burrone for providing the BiP expressing plasmid; R. Hegde and S. Munro (MRC-LMB) for helpful comments on previous versions of the manuscript; O. Perisic, Y. Christova, S. Grazioli, C. Morgan and M. Spinck for helpful discussions; K. Lilley for advice on RHBDL4 proteomics analysis; the FACS facility (MRC-LMB); K. Heesom, M. Skehel, F. Begum, M. Kyte and A. Howe and the MRC-LMB Mechanical Workshop for support; and B. Cravatt (Scripps Research) for suggesting RBBP9 as a target for investigation.

Author contributions S.T. and J.W.C. wrote the paper. J.W.C. set the overall direction of research. S.T. and N.H.-D. developed the hydrolase substrate identification approaches in cell lysate (for TEV and HtrA2), and S.T. established the hydrolase substrate discovery approaches in mammalian cells (for TEV, UL36^{USP}, SCoV2-PLpro, RHBDL4 and RBBP9). N.H.-D. expressed and purified TEV and HtrA2 variants from *E. coli*. S.T. expressed and purified UL36^{USP} and SCoV2-PLpro variants from *E. coli* and performed the conjugation assays with UBL-AMC molecules. S.T. performed HtrA2 substrate validation. S.T. developed the capture of RHBDL4 model substrates with contributions from L.K. S.T. performed RHBDL4 substrate capture and identification experiments. S.T., L.K. and G.P. performed RHBDL4 substrate validation. S.T. and G.P. generated the RHBDL4 knockout cell line. M. Freeman supervised L.K. and contributed to the planning and interpretation of the RHBDL4 experiments. A.T.B. developed the computational pipeline for identifying the conjugates on RBBP9(S75Dap). A.T.B. and S.T. characterized the aminopeptidase activity of RBBP9. S.T. and M. Fiedler solved the structure of RBBP9 in complex with Phe. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Competing interests The authors declare no competing interests.

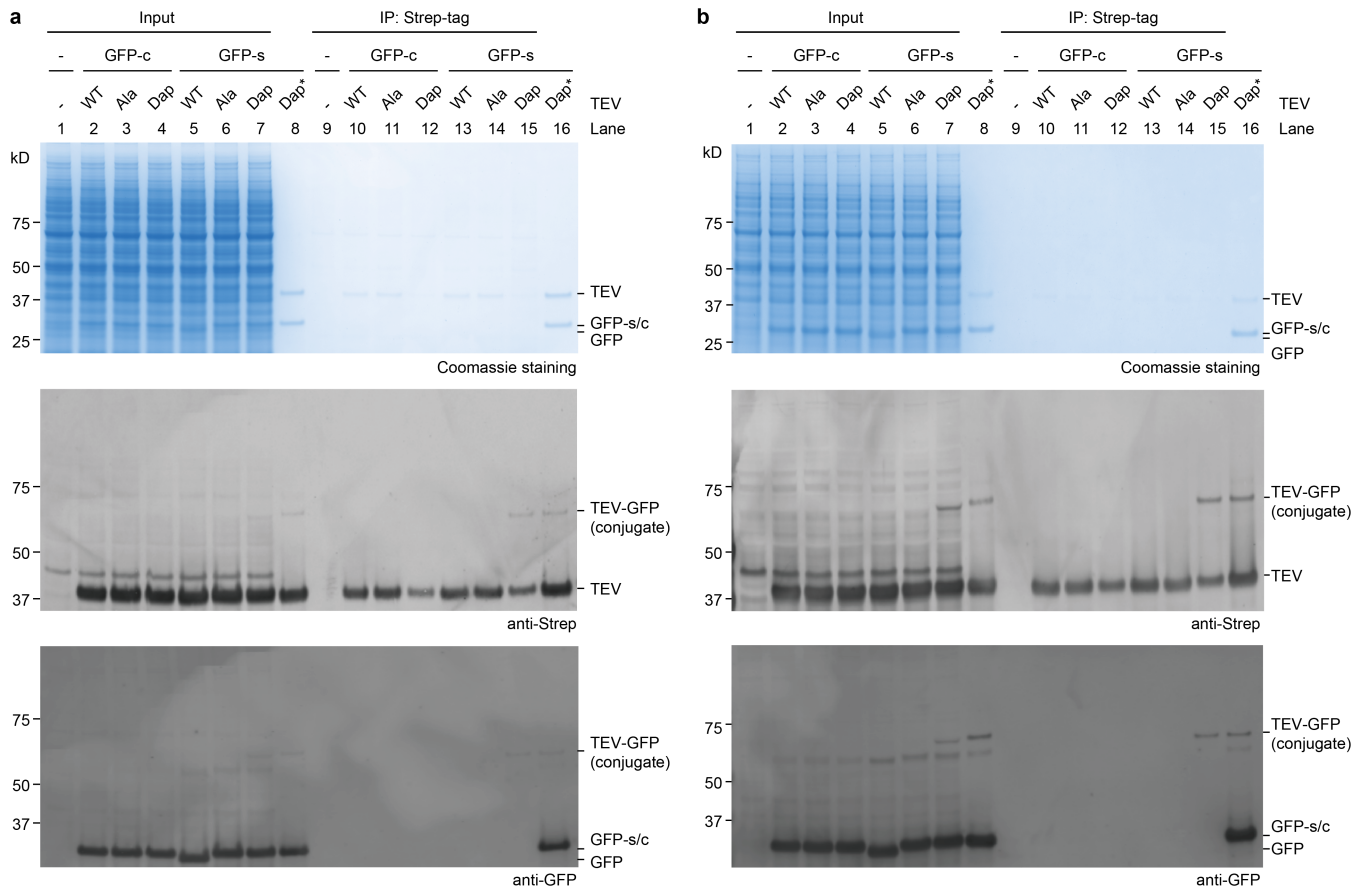
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-04414-9>.

Correspondence and requests for materials should be addressed to Shan Tang or Jason W. Chin.

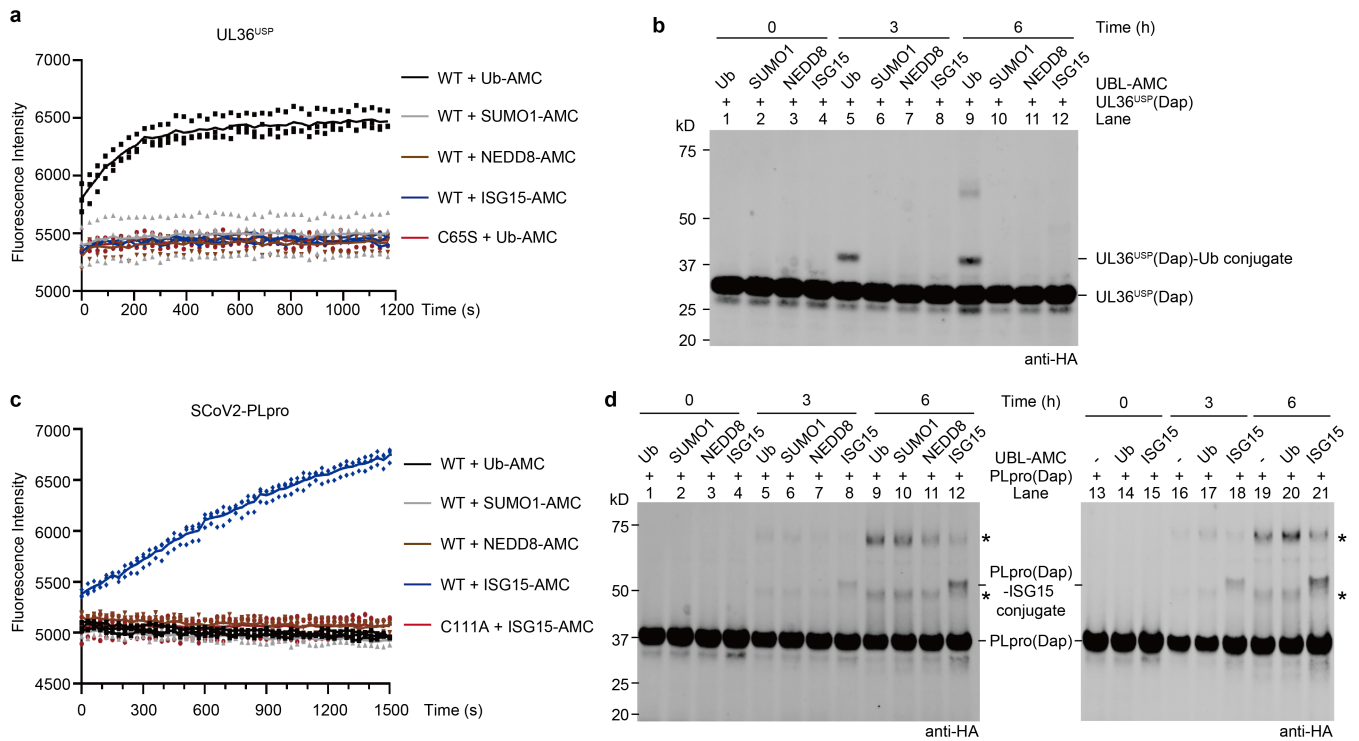
Peer review information Nature thanks Michael Bollong and the other, anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



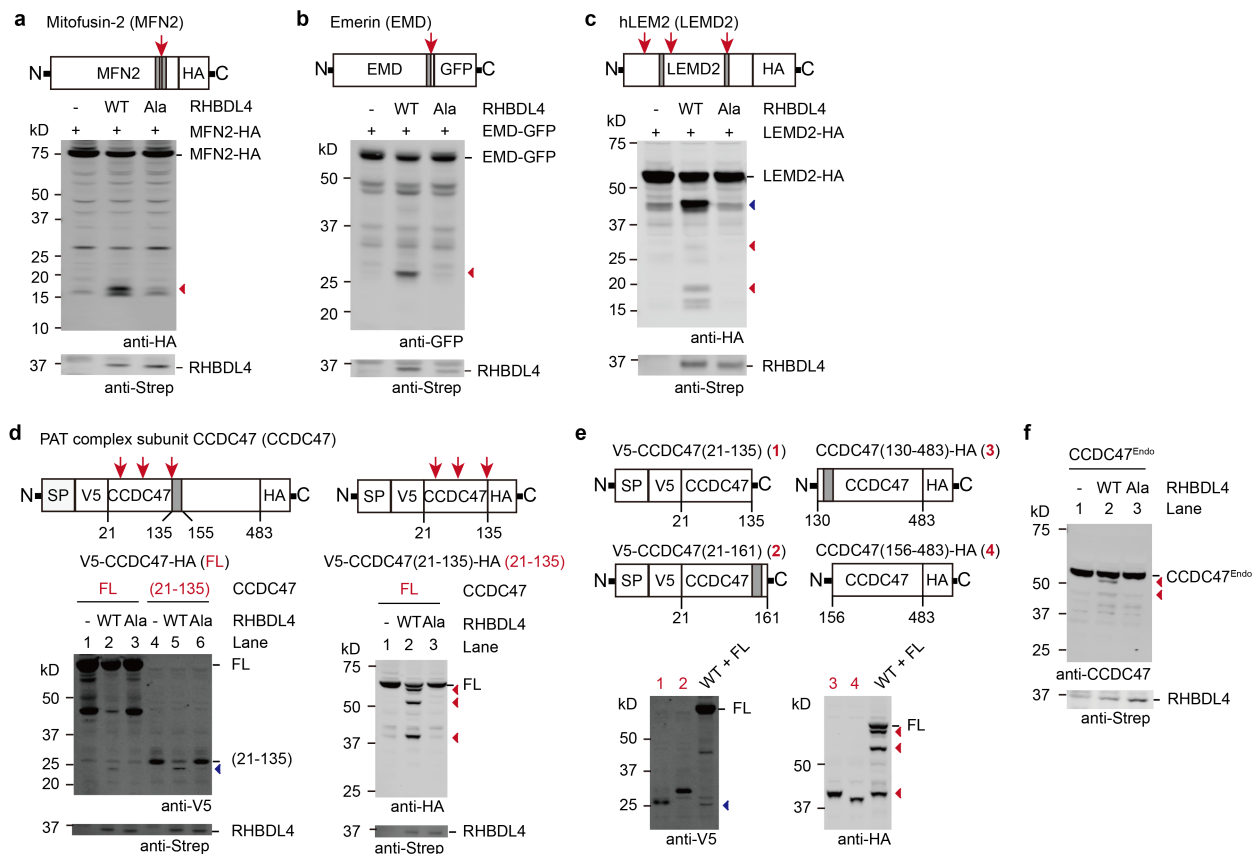
Extended Data Fig. 2 | TEV(C151Dap) specifically traps its substrate in Expi293 cell lysate. TEV variants (WT, Ala, Dap, 1 μ M) were incubated with control GFP-c (a polypeptide "Gly-Gly-Gly-Ser-Gly-Gly-Gly-His6" was attached at the C-terminus of GFP) or substrate GFP-s (a TEV cleavage sequence "Glu-Asn-Leu-Tyr-Phe-Gln-Gly-His6" was attached at the C-terminus of GFP. The cleavage between Gln and Gly is underlined) in Expi293 cell lysate at 37 °C for

3 h. The concentration of GFP-s or GFP-c was (a) 1 μ M or (b) 5 μ M. Dap⁺ reaction of TEV(C151Dap) with GFP-s in Tris buffer. The input (reaction before IP) and Strep-tag enriched TEV species were analyzed by SDS-PAGE (Coomassie staining) and WB (anti-Strep for TEV and anti-GFP). (a) and (b) were repeated in three biological replicates with similar results. For gel source data, see Supplementary Fig. 1.



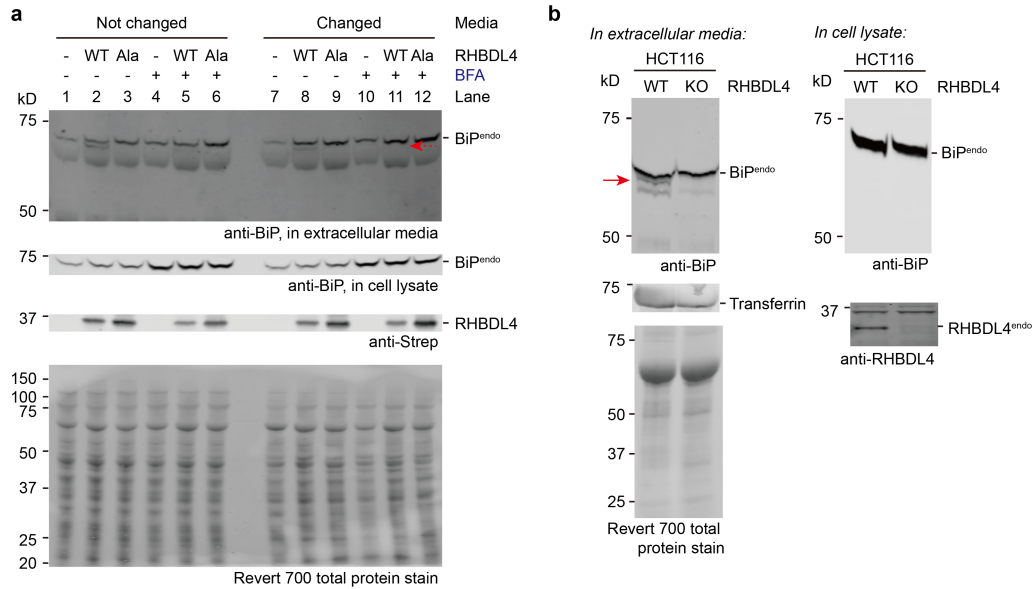
Extended Data Fig. 3 | Dap-containing proteases selectively capture known substrates of their parent enzymes. (a) The change of AMC fluorescence resulting from cleavage from the C-terminus of Ubiquitin (Ub) or ubiquitin-like molecules (SUMO, NEDD8, ISG15) was followed upon mixing with UL36^{USP}. WT UL36^{USP} from human herpesvirus 1 specifically hydrolyses Ub-AMC, but not other ubiquitin-like protein-AMC (UBL-AMC) molecules. The deubiquitination activity of UL36^{USP} is lost when Cys65 is mutated to Ser. This data is consistent with the previously reported specificity of UL36^{USP58}. (b) UL36^{USP} (C65Dap) specifically reacts with Ub-AMC to form the UL36^{USP} (Dap)-Ub conjugate. In contrast, no conjugates were observed between UL36^{USP} (Dap) and SUMO1, NEDD8 or ISG15. (c) WT SCoV2-PLpro selectively

hydrolyses ISG15-AMC in preference to other Ub/UBL-AMC molecules; this data is consistent with the previously reported specificity of SCoV2-PLpro⁵⁹. The hydrolysis of ISG15-AMC is abrogated when the catalytic Cys111 of SCoV2-PLpro is mutated to Ala. (d) SCoV2-PLpro (C111Dap) specifically reacts with ISG15-AMC, generating the SCoV2-PLpro (Dap)-ISG15 conjugate. *: UBL-AMC independent higher MW bands resulting from PLpro (C111Dap) self-reaction. (a) and (c) were generated using n = 3 independent measurements. The line represents the means of three measurements. (b) was repeated twice and (d) was performed in three biological replicates with similar results. For gel source data, see Supplementary Fig. 1.



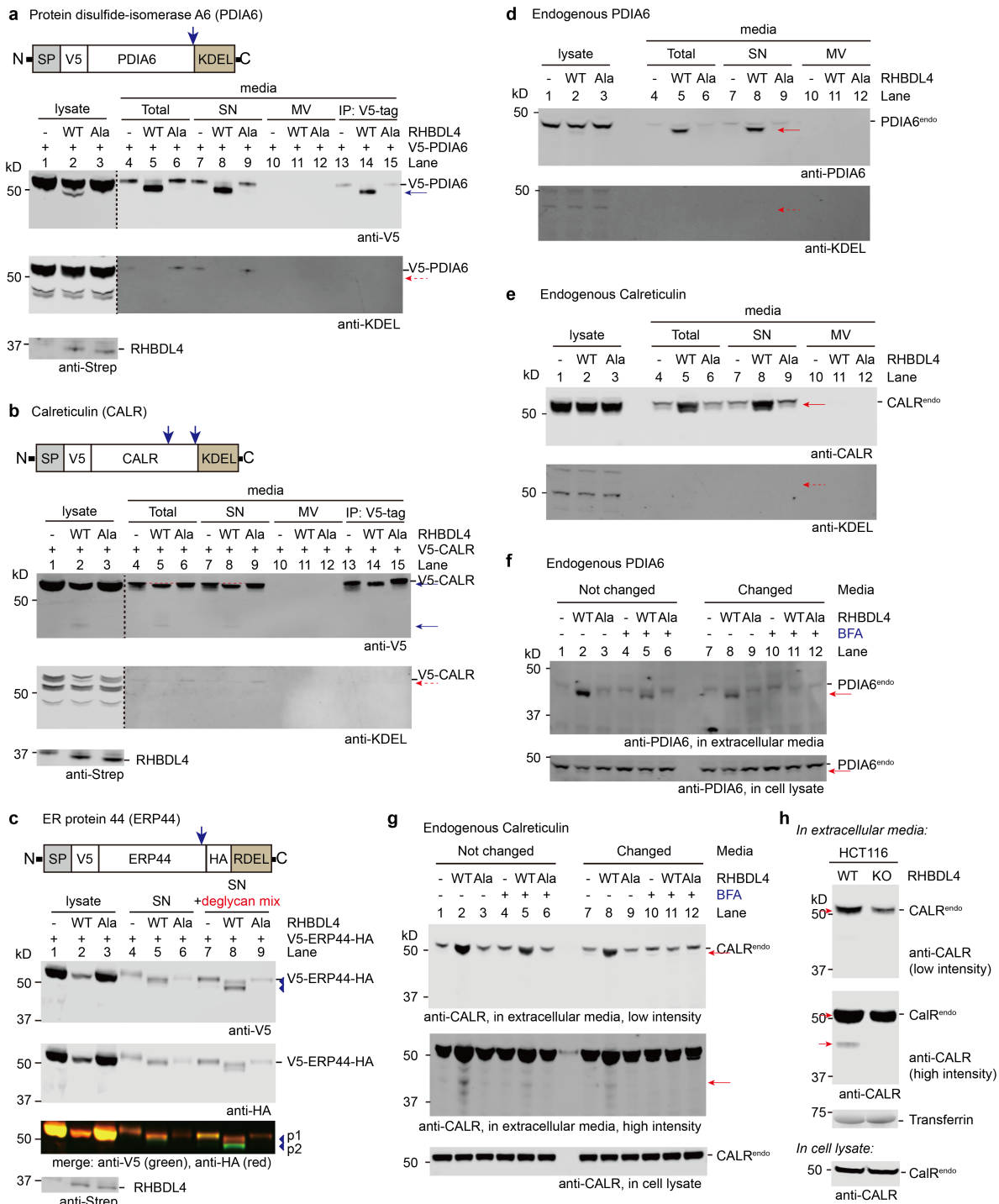
Extended Data Fig. 6 | RHBDL4 cleaves transmembrane substrate candidates. RHBDL4 cleavage assays for (a) MFN2, (b) EMD, (c) LEMD2 and (d) CCDC47. The putative cleavage sites – the approximate positions of which were estimated based on the MW of the proteolytic bands – are indicated by red arrows in schematic representations. HA: HA-tag; GFP: GFP-tag; SP: signal peptide; V5: V5-tag. The transmembrane helices are labeled in grey. (a) WT RHBDL4 cleaved MFN2, a multiple membrane-spanning protein resident in mitochondria or the ER. Red triangle: the C-terminal proteolytic fragment. (b) WT RHBDL4 cleaved EMD, a type II transmembrane protein in the nuclear inner membrane. Red triangle: the C-terminal proteolytic fragment. (c) WT RHBDL4 cleaved LEMD2, a multiple membrane-spanning protein in the nuclear inner membrane. Two minor cleavages (red triangles) would be in the perinuclear space, but the major cleavage site (blue triangle) in the domain resident in nucleus might be cut by the inverted RHBDL4 (see Supplementary Note 1). (d) RHBDL4 cleaved both full-length (FL) CCDC47 and CCDC47(21-135),

generating the same MW N-terminal proteolytic fragment (lanes 2 and 5, blue triangle); this suggested that RHBDL4-mediated cleavage is independent of the transmembrane helix of CCDC47. Blue triangle: the N-terminal fragment; Red triangles: C-terminal proteolytic fragments. (e) Comparison of the proteolytic fragments to truncated CCDC47 standards confirmed that the cleavages are in the luminal domain of CCDC47. WT + FL: FL CCDC47 cleaved by WT RHBDL4. (f) Over-expressed WT RHBDL4 cleaved endogenous CCDC47. Red triangles: proteolytic fragments detected by an anti-CCDC47 antibody. The C-terminal proteolytic fragments of MFN2, LEMD2 and CCDC47 were probed with an anti-HA antibody, while the proteolytic fragment of EMD was detected by an anti-GFP antibody. The N-terminal proteolytic fragments were probed with an anti-V5 antibody. RHBDL4 variants were detected by an anti-Strep antibody. (a-d) were repeated in three biological replicates with similar results. (e) was performed once. (f) was repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | RHBDL4 cleaves endogenous BiP in mammalian cells and facilitates the secretion of N-terminal proteolytic fragments. (a) BFA inhibitory assay was performed for endogenous BiP. The secretion of BiP proteolytic fragment into extracellular media was inhibited by Brefeldin A (BFA) compared to cells treated with DMSO (lanes 5 vs 2, 11 vs 8, anti-BiP). Especially when the medium was changed, no BiP proteolytic fragment was detected in the medium (lane 11, dashed red arrow). Treatment of BFA greatly increased the endogenous level of BiP (lanes 4–6 vs 1–3, anti-BiP in cell lysate). In comparison, the expression of endogenous BiP was slightly increased when WT RHBDL4 or RHBDL4(S144A) was expressed (lanes 2 and 3 vs 1, anti-BiP in cell lysate). BFA treatment or RHBDL4 variant expression facilitated the secretion of full-length BiP (lane 4 vs 1 or 2 vs 1), but not the proteolytic fragment

generated by RHBDL4 cleavage. RHBDL4 expression was detected by an anti-Strep antibody in cell lysate. Revert 700 total protein staining was used as loading control. (b) Endogenous BiP was cleaved by endogenously expressed RHBDL4. WT RHBDL4-dependent BiP proteolytic fragment in the media was detected by an anti-BiP antibody in WT HCT116 cells (red arrow), but not in RHBDL4 knockout HCT116 cells. Transferrin: loading control; additionally, the same samples run in different lanes were stained by Revert 700 for total protein loading control. Endogenous BiP was detected by an anti-BiP antibody and RHBDL4 was detected by an anti-RHBDL4 antibody in the cell lysate. All experiments were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.

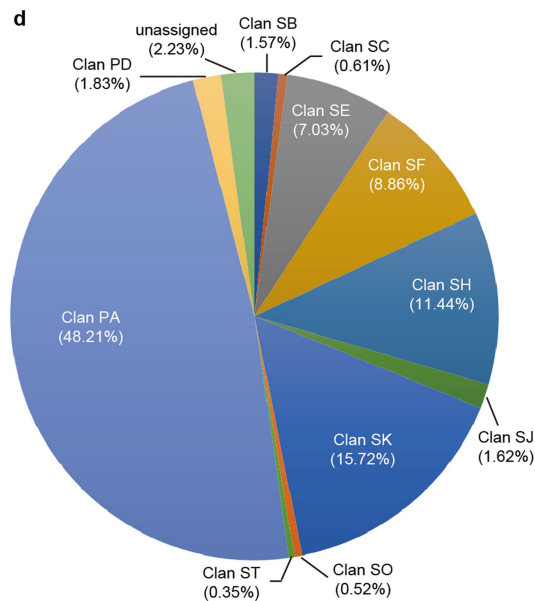
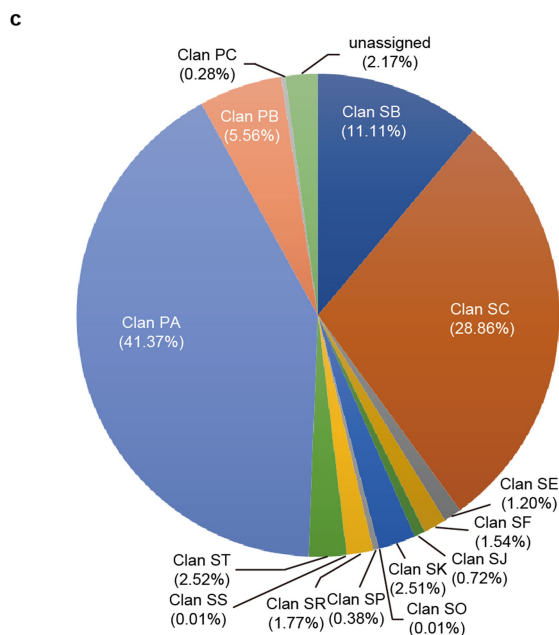
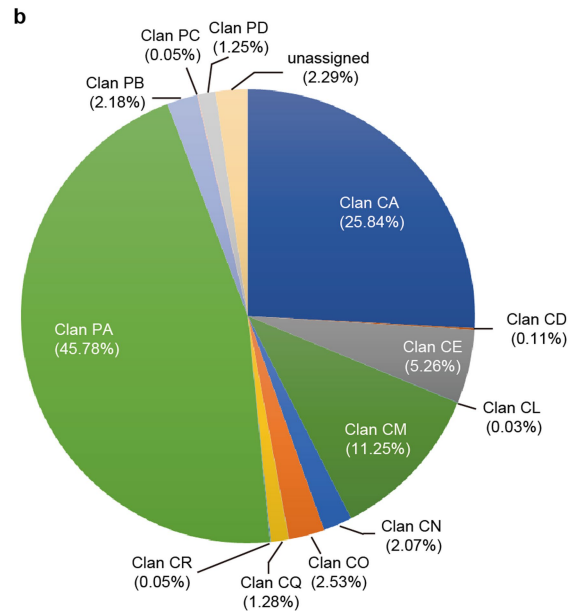
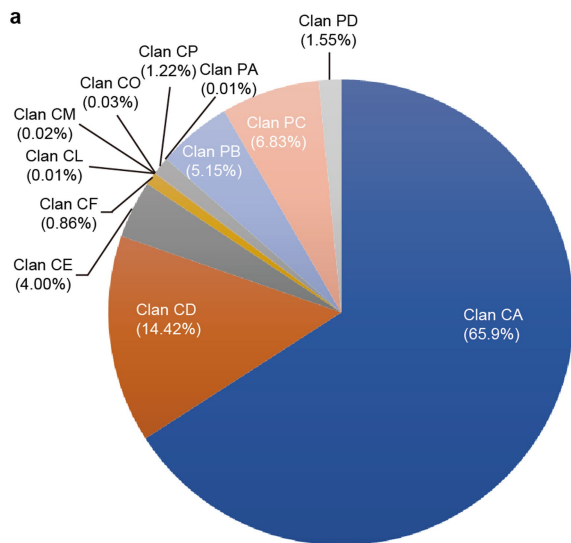


Extended Data Fig. 8 | See next page for caption.

Article

Extended Data Fig. 8 | RHBDL4 cleaves other ER-resident proteins before ER retention motifs and facilitates the secretion of N-terminal proteolytic fragments. RHBDL4 cleavage assays for V5-tagged (a) PDIA6, (b) Calreticulin and (c) ERP44. The putative cleavage sites are indicated by blue arrows in the schematic representations. (a) WT RHBDL4 cleaved PDIA6 at the C-terminus, resulting in secretion of the N-terminal fragment (blue arrow) into the medium. (b) WT RHBDL4 cleaved Calreticulin at the internal region (minor cleavage) and the C-terminus (major cleavage), resulting in secretion of the N-terminal fragments (blue arrows) into the SN. (c) WT RHBDL4 cleaved ERP44 before RDEL sequence. The secreted proteins were deglycosylation mix II sensitive. Because we could not source an antibody to specifically detect the RDEL sequence, we inserted an HA-tag four amino acids before the RDEL sequence for detection. Anti-V5 detected proteolytic fragments 1 and 2 (p1 and p2), and p1 was also detected by the anti-HA antibody, indicating that one cleavage might happen after the HA-tag. The expression of RHBDL4 was detected by an anti-Strep antibody. RHBDL4 cleavage assays for (d) endogenous PDIA6 and (e) endogenous Calreticulin. (d) WT RHBDL4 cleaved endogenous PDIA6. The proteolytic fragment of PDIA6 (red arrow) without the KDEL sequence was

secreted into the SN. (e) WT RHBDL4 cleaved endogenous Calreticulin. The proteolytic fragment of Calreticulin (red arrow) without the KDEL sequence was secreted into the SN. Expression of RHBDL4 did not increase the endogenous level of PDIA6 and Calreticulin (lanes 2 and 3 vs 1). The expression of RHBDL4 is shown in Fig. 3g. BFA inhibitory assays for (f) PDIA6 and (g) Calreticulin. The secretion of proteolytic fragments of PDIA6 and Calreticulin was inhibited compared to cells treated with DMSO (lanes 5 vs 2, 11 vs 8). Endogenous PDIA6 and Calreticulin were detected by an anti-PDIA6 and an anti-CALR antibody, respectively. The expression of RHBDL4 is shown in Extended Data Fig. 7a. (h) Endogenous RHBDL4 cleaves endogenous Calreticulin. WT RHBDL4-dependent Calreticulin proteolytic fragments in the extracellular media were detected by an anti-CALR antibody in WT HCT116 cells (red arrows), but not in RHBDL4 KO HCT116 cells. The secretion of the major proteolytic fragment (near full-length Calreticulin, low intensity) and the minor proteolytic fragment (high intensity) were both detected. Transferrin: loading control. The endogenous RHBDL4 is shown in Extended Data Fig. 7b. (a-h) were repeated in two biological replicates with similar results. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 9 | The distribution of cysteine and serine protease clans. The clan distribution of cysteine proteases from (a) animals and (b) viruses. The clan distribution of serine proteases from (c) animals and (d) viruses. We have demonstrated the Dap-mediated protease substrate capture method for multiple cysteine proteases (TEV (Clan PA), UL36^{USP} (Clan CA) and

SCoV2-PLpro (Clan CA)) and serine proteases (HtrA2 (Clan PA), RHBDL4 (Clan ST) and RBBP9 (Note that RBBP9 has not been included in the MEROPS database by September 30, 2021. RBBP9 possesses the "alpha-beta hydrolase" fold⁴⁰, suggesting that it belongs to Clan SC). The data used to generate this figure were downloaded from the MEROPS database⁴⁷.

Article

Extended Data Table 1 | Summary of crystallographic information

	RBBP9_PHE (7OEX)
Data Collection	
Space Group	P 1 21 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	36.96, 130.13, 38.99
α , β , γ (°)	90.0, 115.24, 90.0
Resolution (Å)	29.75 - 1.51
*Rmerge	0.068 (0.625)
*Mean I/ σ (I)	14.93 (2.44)
*Completeness (%)	95.9 (81.6)
*Multiplicity	7 (5.7)
Complexes in A.U.	2
Refinement	
Resolution (Å)	29.75 - 1.51
No. Reflections	49633
<i>R</i> work / <i>R</i> free (%)	16.7 / 16.4
No. Atoms	3322
Wilson B-factor (Å ²)	16.9
Anisotropy	0.668
Average B, all atoms (Å ²)	20.0
B-factors	
Protein	20.5
Ligands	21.77
Water	29.06
R.m.s. deviations	
Bond length (Å)	0.012
Bond angle (°)	1.42

* Highest resolution shell (in Å) shown in parenthesis.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Odyssey Clx imaging system (Li-Cor); MS Chemstation (Rev.C.01.06[61]); MARS Data Analysis Software (version 3.20 R2); XIA2 (version 0.7.90), DIALS (version 3.1.3), scaled using Aimless and Refmac (version 5.8.0258) in the CCP4 suite (version 7.0.078) of programs for crystal data processing
Data analysis	Image Studio Lite (version 5.2.5) for western blot analysis; Mascot Search Engine Program (version 2.4) and Scaffold Proteome Software (version 4) for protein identification (Venn Diagrams); MaxQuant Software (version 1.6.3.4) and Perseus (version 1.6.2.3) for LFQ analysis (Volcano Plots); FCS Express Flow Cytometry software (version 7) for FACS data analysis; Fiji (ImageJ2) for IF analysis; GraphPad Prism (version 8) for graphs; Refmac (version 5.8.0258) and COOT (0.8.9.2) for structure refinement and manual model building; Pymol (version 1.0r2) for protein structural visualization; ProteoWizard (version 3.0.11252) for converting LC-MS/MS RAW files to mzML format; Custom Python (version 3.8.1) scripts with the pyOpenMS package (version 2.4.0) for MS data preparation and processing. The code used for RBBP9 substrate identification by LC-MS/MS analysis is available at https://doi.org/10.5281/zenodo.5768340 .

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The structure of RBBP9 in complex with Phe is available in the Protein Data Bank under accession code 7OEX. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PDX030381. All other datasets and materials generated or analyzed in this study are available from the corresponding authors upon reasonable request. The data used to analyze serine and cysteine proteases clans were downloaded from the MEROPS database (<https://www.ebi.ac.uk/merops/>)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. These are biochemical experiments, not animal experiments, and there are therefore no individuals to sample. Sample size is not a relevant parameter, but the number of replicates is relevant. The number of replicates is indicated in the relevant figure legend.
Data exclusions	No data was excluded
Replication	Three replicates were performed for protein identification by LC-MS/MS; Two replicates were performed to obtain the entire masses. All attempts at replicates were successful. Three or two replicates (annotated in the figure legends) were performed for immunoblotting analysis. All immunoblotting analysis were reproducible. Two replicates were performed for fluorescence-based aminopeptidase activity analysis. All fluorescent measurements were reproducible. All experiments were performed in the number of replicates indicated in the figure legends. All replicates were successful.
Randomization	These are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore randomization does not make sense.
Blinding	These are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore blinding does not make sense.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies used

Anti-Strep (ab76949); Anti-Calreticulin (ab92516); Anti-PDIA6 (ab154820); Anti-KDEL (ab176333); Anti-POLR2A (ab76123); Anti-POLR2B (ab228933); Anti-DYNC1H1 (ab245554); Anti-GCN1 (ab86139); Anti-SEC16A (ab70722); Anti-RANBP2 (ab2938); Anti-FASN (ab128870); Anti-BIRC6 (ab19609); Anti-CLTC (ab172958); Anti-PLC (ab32528); Anti-POLR2C (ab182150); Anti-RBM4 (ab251923); Anti-TRIP13 (ab128171); Anti-EPRS (ab31531); Anti-DNAJA1 (ab126774); Anti-RBM14 (ab228692); Anti-RACK1 (ab129084); Anti-IRS4 (ab52622); Anti-ATP5F1A (ab176569); Anti-OAT (ab137679); Anti-TUFM (ab173300) and Anti-RBBP9 (ab157202) were from abcam. Anti-GFP (2956(D5.1)); Anti-HA (3724(C29F4)); Anti-HA (2367(6E2)); Anti-GAPDH (2118(14C10)) and Anti-Tubulin (86298(D3U1W)) were from Cell Signaling. Anti-V5 (R960-25) was from Thermo Fisher. Anti-FLAG (F3165) and Anti-CCDC47 (HPA029674) were from Merck. Anti-FLAG (AHP1074GA) was from Bio-Rad. Anti-MDN1 (A304-739A-T) and Anti-WDCP (A303-337-T) were from Bethyl Laboratories. Anti-BiP (GTX113340); Anti-MCM7 (GTX110278) and Anti-FAM120A (GTX120824) were from GeneTex. Anti-MTMR3 (H00008897-B01P) was from Novus Biologicals. Anti-RHBDL4 (20869-1-AP); Anti-Calnexin (10427-2-AP); Anti-LaminB1(12987-1-AP), Anti-Transferrin (66171-1-Ig) and Anti-SNRNP200 (23875-1-AP) were from Proteintech. Anti-DSP (MAB9080) and Anti-Ub (MAB8595) were from R&D Systems. Anti-RPS3 (A2533) was from ABclonal. Anti-GAPDH (TA802519) was from Origene. Goat anti-Mouse (925-68070; 925-32210) secondary antibodies and Goat anti-Rabbit (925-68071; 926-32211) secondary antibodies were from Li-Cor.

Validation

Validations are based on the data from the manufacturers.

Anti-Strep (ab76949) was validated by detecting recombinant proteins with strep-tag in cell lysate. <https://www.abcam.com/strep-tag-ii-antibody-ab76949.html>

Anti-Calreticulin (ab92516) was knockout validated in HAP1 cell line. <https://www.abcam.com/calreticulin-antibody-epr3924-er-marker-ab92516.html>

Anti-PDIA6 (ab154820): <https://www.abcam.com/pdia6-antibody-epr10132b-ab154820.html>

Anti-KDEL (ab176333): <https://www.abcam.com/kdel-antibody-epr12668-ab176333.html>

Anti-POLR2A (ab76123): <https://www.abcam.com/rna-polymerase-ii-rpb1-antibody-epr1509y-chip-grade-ab76123.html>

Anti-POLR2B (ab228933): <https://www.abcam.com/rpb2-antibody-c-terminal-ab228933.html>

Anti-DYNC1H1 (ab245554): <https://www.abcam.com/dync1h1-antibody-ab245554.html>

Anti-GCN1 (ab86139): <https://www.abcam.com/gcn1-antibody-ab86139.html>

Anti-SEC16A (ab70722): <https://www.abcam.com/sec16asec16-antibody-ab70722.html>

Anti-RANBP2 (ab2938): <https://www.abcam.com/ranbp2-antibody-ab2938.html>

Anti-FASN (ab128870) was knockout validated in HAP1 cell line. <https://www.abcam.com/fatty-acid-synthase-antibody-epr7466-ab128870.html?productWallTab=ShowAll>

Anti-BIRC6 (ab19609): <https://www.abcam.com/birc6apollon-antibody-ab19609.html?productWallTab=ShowAll>

Anti-CLTC (ab172958): <https://www.abcam.com/clathrin-heavy-chain-antibody-epr12235b-ab172958.html>

Anti-PLC (ab32528) was knockout validated in U2OS cell line. <https://www.abcam.com/plectin-antibody-e398p-ab32528.html?productWallTab=ShowAll>

Anti-POLR2C (ab182150): <https://www.abcam.com/rpb3-antibody-epr13294b-ab182150.html>

Anti-RBM4 (ab251923): <https://www.abcam.com/rbm4-antibody-ab251923.html>

Anti-TRIP13 (ab128171): <https://www.abcam.com/trip13pch2-antibody-ab128171.html?productWallTab=ShowAll>

Anti-EPRS (ab31531): <https://www.abcam.com/glutamyl-prolyl-trna-synthetasepars-antibody-ab31531.html?productWallTab=ShowAll>

Anti-DNAJA1 (ab126774) was knockout validated in HEK293T cell line. <https://www.abcam.com/dnaja1-antibody-epr7248-ab126774.html>

Anti-RBM14 (ab228692): <https://www.abcam.com/rbm14-antibody-n-terminal-ab228692.html>

Anti-RACK1 (ab129084): <https://www.abcam.com/rack1-antibody-epr7388-ab129084.html>

Anti-IRS4 (ab52622): <https://www.abcam.com/irs4-antibody-ep907y-ab52622.html>

Anti-ATP5F1A (ab176569): <https://www.abcam.com/atp5a-antibody-epr13030b-ab176569.html?productWallTab=ShowAll>

Anti-OAT (ab137679): <https://www.abcam.com/ornithine-aminotransferase-antibody-ab137679.html>

Anti-TUFM (ab173300): <https://www.abcam.com/tufm-antibody-epr12797b-ab173300.html>

Anti-GFP (D5.1, 2956) was validated by detecting recombinant GFP in cell lysate. <https://www.cellsignal.co.uk/products/primary-antibodies/gfp-d5-1-rabbit-mab/2956>

Anti-HA (3724(C29F4)) was validated by detecting recombinant proteins with HA-tag in cell lysate. https://www.cellsignal.co.uk/products/primary-antibodies/ha-tag-c29f4-rabbit-mab/3724?site-search-type=Products&N=4294956287&Ntt=3724s&fromPage=plp&_requestid=1510669

Anti-HA (2367(6E2)): <https://www.cellsignal.com/products/primary-antibodies/ha-tag-6e2-mouse-mab/2367>

Anti-GAPDH (2118(14C10)): <https://www.cellsignal.com/products/primary-antibodies/gapdh-14c10-rabbit-mab/2118>

Anti-GAPDH (TA802519): <https://www.origene.com/catalog/antibodies/primary-antibodies/ta802519/gapdh-mouse-monoclonal-antibody-clone-id-oti2d9>

Anti-Tubulin (86298(D3U1W)): <https://www.cellsignal.co.uk/products/primary-antibodies/b-tubulin-d3u1w-mouse-mab/86298>

Anti-V5 (R960-25) was validated by detecting V5-tagged recombinant proteins in cell lysate. <https://www.thermofisher.com/antibody/product/V5-Tag-Antibody-Monoclonal/R960-25>

Anti-FLAG (F3165) was validated by detecting FLAG-tagged recombinant proteins in cell lysate. <https://www.sigmaldrich.com/GB/en/product/sigma/f3165?context=product>

Anti-FLAG (AHP1074GA) was validated by detecting recombinant proteins with FLAG-tag in cell lysate. <https://www.bio-rad-antibodies.com/polyclonal/synthetic-peptide-dykdiddk-tag-antibody-ahp1074.html?f=purified>

Anti-CCDC47 (HPA029674): <https://www.sigmaldrich.com/GB/en/product/sigma/hpa029674>

Anti-MDN1 (A304-739A-T): <https://www.bethyl.com/product/A304-739A/Midasin+Antibody>

Anti-WDCP (A303-337-T): <https://www.bethyl.com/product/A303-337A?referrer=search#>

Anti-BiP (GTX113340): <https://www.genetex.com/Product/Detail/Grp78-antibody/GTX113340#datasheet>

Anti-MCM7 (GTX110278): <https://www.genetex.com/Product/Detail/MCM7-antibody-N2C2-Internal/GTX110278>

Anti-FAM120A (GTX120824): <https://www.genetex.com/Product/Detail/FAM120A-antibody/GTX120824>

Anti-MTMR3 (H00008897-B01P): validated by detection of MTMR3 in transfected 273T cell line. https://www.novusbio.com/products/mtmr3-antibody_h00008897-b01p?utm_source=citeab&utm_medium=referral&utm_campaign=product&utm_term=primaryantibodies

Anti-RHBDL4 (20869-1-AP) was knockout validated in HCT116 cell line in the current paper. The RHBDL4 positive bands were detected in HEK293 cell line and mouse kidney and testis tissues. <https://www.ptgcn.com/products/RHBDL4-Antibody-20869-1-AP.htm>
 Anti-Calnexin (10427-2-AP): <https://www.ptglab.com/products/CANX-Antibody-10427-2-AP.htm>
 Anti-LaminB1(12987-1-AP): <https://www.ptgcn.com/products/LMN1-Antibody-12987-1-AP.htm>
 Anti-SNRNP200 (23875-1-AP): <https://www.ptgcn.com/products/SNRNP200-Antibody-23875-1-AP.htm>
 Anti-DSP (MAB9080): https://www.rndsystems.com/cn/products/human-desmoplakin-antibody-824038_mab9080
 Anti-RPS3 (A2533): <https://abclonal.com/catalog-antibodies/RPS3RabbitpAb/A2533>
 Goat anti-Mouse (925-68070): <https://www.licor.com/bio/reagents/irdye-680rd-goat-anti-mouse-igg-secondary-antibody>
 Goat anti-Mouse (925-32210): <https://www.licor.com/bio/reagents/irdye-800cw-goat-anti-mouse-igg-secondary-antibody>
 Goat anti-Rabbit (925-68071): <https://www.licor.com/bio/reagents/irdye-680rd-goat-anti-rabbit-igg-secondary-antibody>
 Goat anti-Rabbit (926-32211): <https://www.licor.com/bio/reagents/irdye-800cw-goat-anti-rabbit-igg-secondary-antibody>
 Anti-transferrin (66171-1-ig): <https://www.ptgcn.com/products/TF-Antibody-66171-1-ig.htm>
 Anti-Ub (MAB8595): https://www.rndsystems.com/cn/products/human-ubiquitin-antibody-1002a_mab8595
 Anti-RBBP9 (ab157202): <https://www.abcam.com/bog-antibody-epr9905-ab157202.html>

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T cells were purchased from European Collection of Cell Cultures (ECACC). Expi293 cells (A14527) were purchased from Thermo Fisher. HCT116 cells were purchased from American Type Culture Collection (ATCC).
Authentication	HEK293T cell line was authenticated by ECACC, Expi293 cell line was authenticated by Thermo Fisher, and HCT116 cell line was authenticated by ATCC using STR DNA profiling analysis to rule-out intra- and inter-species contamination.
Mycoplasma contamination	All cell lines tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	None

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	48 h after transfection, cells in 24-well plates were washed with PBS, detached by trypsin/EDTA solution, and resuspended in the growth medium. Cells were pelleted and resuspended in PBS supplemented with 3% FBS.
Instrument	Becton Dickinson LSRFortessa (407 nm violet laser (V-450) for BFP excitation, 488 nm blue laser (B-525) for eGFP excitation, 561 nm yellow-green laser (YG-610) for mCherry excitation).
Software	FCS Express 7 software (De Novo software)
Cell population abundance	20,000 intact cells (gated by from scatter and side scatter) were analyzed for each sample.
Gating strategy	The front scatter and side scatter were used to identify intact cells. Non-transfected cells, cells only transfected with mCherry or eGFP plasmid and BFP positive cells were used to define the gate.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.