Proteolytic Activation of Human Cathepsin A

Nilima Kolli¹,² and Scott C. Garman¹,²

Department of Biochemistry & Molecular Biology¹ and Program in Molecular & Cellular Biology ²
University of Massachusetts, Amherst MA 01003, USA;

Running title: Structure of human cathepsin A

Address correspondence to: Scott C. Garman, PhD, Dept of Biochemistry & Molecular Biology, 710 North Pleasant Street, University of Massachusetts, Amherst MA 01003 USA. Fax 413-545-3291; E-mail: garman@biochem.umass.edu

Keywords: glycoprotein, X-ray crystallography, proteolysis

Background: The human lysosomal protease cathepsin A requires proteolytic activation.
Results: The crystal structure of mature and active cathepsin A reveals its mechanism of activation.
Conclusion: Removal of a 3.3 kDa peptide (by unidentified proteases) allows substrate access to the active site.
Significance: The results revise a two-decades old model of cathepsin A activation by proteolysis and subsequent conformational change.

ABSTRACT

Galactosialidosis is a human lysosomal storage disease caused by deficiency in the multifunctional lysosomal protease cathepsin A (also known as Protective Protein/Cathepsin A, PPCA, catA, HPP, and CTSA; E.C. 3.4.16.5). Previous structural work on the inactive precursor human cathepsin A (zymogen) led to a two-stage model for activation, where proteolysis of a 1.6 kDa excision peptide is followed by a conformational change in a blocking peptide occluding the active site. Here we present evidence for an alternate model of activation of human cathepsin A, needing only cleavage of a 3.3 kDa excision peptide to yield full enzymatic activity, with no conformational change required. We present X-ray crystallographic, mass spectrometric, amino-acid sequencing, enzymatic, and cellular data to support the cleavage-only activation model. The results clarify a long-standing question about the mechanism of cathepsin A activation and point to new avenues for the design of mechanism-based inhibitors of the enzyme.

INTRODUCTION

Galactosialidosis is a lysosomal storage disease in humans caused by the deficiency in the multifunctional lysosomal protease cathepsin A. Severity of galactosialidosis presentation varies depending on the specific mutation in the CTSA gene. Clinical manifestations include fetal hydrops, hepatosplenomegaly, skeletal dysplasia, growth retardation, and cardiac impairment. Mental retardation and neurologic deterioration are seen in some patients (1). Currently there is no treatment available for patients with galactosialidosis.

Cathepsin A forms a complex with lysosomal neuraminidase (NEU1) and β-galactosidase (GLB1), and cathepsin A is proposed to stabilize GLB1 and activate NEU1 (2-6). In addition, it exhibits enzymatic activity as a serine carboxypeptidase in the acidic pH of the lysosome (7,8) and also has deamidase and esterase activity at neutral pH (5,8,9). Cathepsin A was shown to play a role in chaperone-mediated autophagy by regulating degradation of the lysosome-associated membrane protein type 2 (LAMP2) receptor (10). Cathepsin A is expressed ubiquitously and acts on multiple peptide hormones, including endothelin-1 (11) and angiotensin I, highlighting its importance in endocrine regulation (12). Inhibitors against cathepsin A are currently in Phase I clinical trials for hypertension due to its central role in...
vasoregulation (13,14). Additionally, cathepsin A is a major hydrolase for converting peptidic prodrugs to active compounds (15). Understanding the structure and function of the catalytically active mature protease is thus important in understanding cellular regulation, and in the design and optimization of peptidic compounds with improved bioavailability via decreased proteolytic susceptibility.

Carboxypeptidases occur as serine proteases, cysteine proteases, or metalloproteases. Cathepsin A belongs to the S10 family of serine proteases, along with carboxypeptidases from yeast (CPY) (16) and wheat (CPW) (17), with which cathepsin A shares sequence identity (31% and 38% respectively) and structural similarity (18). Based upon sequence similarities, cathepsin A was originally grouped with the CPW-like carboxypeptidases (19).

Cathepsin A is synthesized as a single-chain precursorzymogen of 54 kDa (after signal peptide cleavage and addition of two N-linked glycans) that self-associates into a homodimer (Figure 1a). In 1995, the crystal structure of the precursor cathepsin A showed each monomer with two domains, a core domain (1-182 and 303-452) interrupted by a cap domain insertion (183-302) (20-22). The core domain has an \( \alpha/\beta \) hydrolase fold containing the active site, as in other serine carboxypeptidases. The cap domain is further divided into a helical subdomain (183-253) important for dimerization and a mixed three-stranded \( \beta \)-sheet maturation subdomain (254-302).

In the enzymatically inactive precursor cathepsin A structure, the catalytic triad residues (Ser150, His429, and Asp372) have a catalytically competent arrangement at the base of the substrate-binding pocket. However, in that structure, catalysis is prevented by the occlusion of the substrate-binding pocket by two segments of the maturation subdomain originally described (20) as the blocking peptide (272-277) and the excision peptide (285-298) (Figure 1).

The precursor cathepsin A structure led to a novel proposed activation mechanism, distinct from known activation mechanisms of serine-, cysteine-, metallo-, and aspartyl-proteases (20). Cathepsin A was proposed to mature by a two-step process involving cleavage of the 1.6 kDa excision peptide, followed by conformational changes in the blocking peptide of the maturation subdomain, allowing substrate access to the active site. In the proposed activation model, the 54 kDa monomer was processed into a large (32 kDa) and a small (20 kDa) subunit, covalently linked by two disulfide bonds (20,23). Nearly two decades later, the structure of mature cathepsin A bound to small-molecule inhibitors suggested a similar activation mechanism and described the removal of a 2 kDa excision peptide, although electron density was absent for a longer sequence (13). The maturation subdomain (254-302) of precursor cathepsin A has multiple surface-exposed potential protease cleavage sites, lending support to the possibility of activation by proteolytic removal of a larger segment. To date, the long-standing question of the activation mechanism of human cathepsin A remains unclear.

Here we report the crystal structure of mature and catalytically active cathepsin A, suggesting an alternate model for activation of cathepsin A. Using X-ray crystallography, mass spectrometry, amino-acid sequencing, and cellular experiments, we show activation of precursor cathepsin A by proteolytic removal of a larger 3.3 kDa peptide that includes the blocking peptide, bypassing the requirement for conformational changes. Our results suggest an activation mechanism of cathepsin A consistent with homologous carboxypeptidases such as CPW. Alanine scanning of residues in the processing region reveals alternative sites for proteolytic activation of cathepsin A. In addition, we show the importance of removal of the blocking peptide on carboxypeptidase activity of human cathepsin A, using a N-blocked dipeptide carbobenzyoxycarbonyl-phenylalanine-leucine (Z-Phe-Leu) as substrate. The structure and enzymatic studies presented here clarify the regulatory mechanisms of cathepsin A and suggest new mechanisms for the control of the cathepsin A in therapeutic uses.

**EXPERIMENTAL PROCEDURES**

**Molecular biology:** Human cathepsin A (CTSA) cDNA (National Center for Biotechnology Information Sequence ID: NM_000308.2) was purchased from Harvard Institute of Proteomics, DNA Resource Core. The complete cDNA including the native human signal sequence was cloned into the pIB/V5-HIS-TOPO-TA vector by gateway cloning methods. A His\(_6\) tag was
engineered after the C terminus of the gene, replacing the vector-encoded V5 and His6 epitopes. Mutations in the CTSA gene were generated according to the Phusion site-directed mutagenesis kit (Thermo Scientific) by PCR amplification using primers coding for the corresponding base changes.

For expression in human cells, the CTSA gene was cloned into the p3xFLAG-CMV14 vector to encode C-terminally FLAG-tagged protein. Mutations in the CTSA gene were generated as above. CTSA constructs for human-cell expression contained the nucleophile knock out S150A.

**Protein expression and purification:** We expressed and purified cathepsin A in Tn5 (Trichoplusia ni) insect cells as a secreted protein. Adherent cultures were transfected with plasmid DNA in T25 flasks. After 2 days, treatment with 100 µg/mL blasticidin for 10 days generated stable cell lines. For protein purification, cells were resuspended into 25 mL of homemade Sfx-equivalent media and then scaled up to 6 to 10 L in 3 L fernbach flasks (Corning). The cells were harvested at 3×10⁶ cells/mL and clarified by centrifugation. The spent media was concentrated and buffer exchanged into Ni²⁺ wash buffer [50 mM phosphate and 250 mM NaCl] by tangential flow filtration.

Cathepsin A was purified by affinity and ion-exchange chromatography using a FPLC (BioRad). Cathepsin A was partially purified by the Ni²⁺-sepharose FF column (GE Healthcare) with elution buffer (50 mM phosphate, 250 mM NaCl, 250 mM imidazole). The fractions containing cathepsin A were pooled, concentrated, and buffer exchanged into 50 mM sodium acetate pH 5.1 and further purified on a SOURCE15 S anion-exchange column by a gradient of 0-500 mM NaCl. The fractions containing pure cathepsin A were buffer exchanged into 20 mM Tris-HCl, 150 mM NaCl and concentrated to 1 mg/mL for storage. The purified precursor cathepsin A was observed to undergo limited proteolysis over 1-3 weeks, presumably by small amounts of co-purified endogenous proteases.

**Mammalian cell expression and immunoprecipitation (IP):** Cathepsin A plasmid DNA in the p3xFLAG-CMV14 vector was transfected into HEK293T cells using Lipofectamine (Life Technologies) according to manufacturer’s protocol. Cells were harvested 36 h post transfection and lysed in cell lysis buffer. Whole cell lysates were immunoprecipitated with anti-FLAG M2 magnetic beads (Sigma-Aldrich) for one hour at 4°C. Beads were washed in buffer (50 mM Tris, 150 mM NaCl) and acid eluted (100 mM sodium citrate, pH 3). Eluates were analyzed for proteolytic processing by western blot using anti-cathepsin A antibody (Rockland Immunochemicals) to detect precursor and large subunit cathepsin A. The small subunit of cathepsin A was detected by analyzing the whole cell lysate by western blot using anti-FLAG antibody.

**Crystallography:** Cathepsin A was purified and concentrated to 4 mg/mL for sitting drops. Crystals grew in 0.1 M sodium formate pH 7.0 (or 0.1 M ammonium tartrate), 10% (w/v) polyethylene glycol 3,350 in 7 days. Crystals were flash cooled in liquid nitrogen and X-ray data were collected at the microfocus beamline NE-CAT 24-ID-C at Argonne National Laboratory. X-ray images were processed in HKL2000 (24) and phased by molecular replacement using the cathepsin A precursor (PDB ID: 1IVY) in PHASER (25). TRUNCATE (26) reported statistics suggestive of nearly full merohedral twinning in crystal Native1, so crystal Native2 (untwinned but lower resolution) was used to independently confirm the structure. Model building and validation were done in Coot (27) and CCP4i (28) respectively. Refmac5 (29) was used for refinement. For the Native1 twinned data, amplitude-based twin refinement was used during refinement in Refmac5.

**Mass spectrometry:** Insect-cell-expressed cathepsin A glycoprotein was subjected to enzymatic deglycosylation with Endoglycosidase H and Peptide-N-Glycosidase F. However, the carbohydrate was resistant to enzymatic removal (presumably due to fucosylation (30,31), leading to poor results in electrospray ionization (ESI) experiments. The glycosylated protein was subsequently used for Matrix-assisted laser desorption/ionization (MALDI) analysis on a Bruker OmniFlex MALDI-TOF MS in linear mode to determine masses. Protein samples at different times post purification were buffer exchanged into 50 mM ammonium acetate pH 7.4 and mixed with 50 mg/ml sinapic acid in 50% acetonitrile and 0.5% trifluoroacetic acid.
N-terminal sequencing: Mature cathepsin A was concentrated and run on a reducing SDS-PAGE gel. The proteins were transferred to PVDF membrane (Immobilon) and visualized by staining with GelCode Blue (Pierce). The band corresponding to the small subunit (~20 kDa) was cut out, washed, dried, and sequenced by 10 cycles of Edman degradation (Harvard Microchemistry) with 13 pmol of Ser293 detected.

Activity assays: Carboxypeptidase activity of cathepsin A was measured using Z-Phe-Leu (Sigma) substrate in a two-step assay. Z-Phe-Leu cleavage by cathepsin A leads to a leucine primary amine, which was quantified in a modified trinitrobenzene sulfonate (TNBS) assay (32). Briefly, 40 nM of mature cathepsin A was added to assay buffer (0.1 M NaOAc and 0.15 M NaCl, pH 4.5) containing 0.01 mM to 1.0 mM Z-Phe-Leu. Every 15 seconds for 2 minutes, 10 µL of the reaction was added to 10 µL TNBS (2 mM in 0.2 M sodium borate, pH 9.8). After 30 minutes at 27°C, 80 µL of 3 mM NaS2O3 in 0.2 M KH2PO4, pH 4.2 was added. Absorbance at 420 nm was compared to controls lacking enzyme and/or substrate and calibrated against a standard curve using leucine. The pH profile of enzyme activity was determined by varying assay buffers from pH 3.5 to pH 7.0, and the resulting plot was fit with $v = \frac{v_{max}}{[H^+] / K_1 + [H^+] / K_2}$ where $K_1$ and $K_2$ are the $K_a$ values for the relevant ionizable groups (33). Kinetic parameters were determined by measuring rate of reaction using 0.01 mM to 1.0 mM Z-Phe-Leu in quadruplicate measurements. $K_M$ and $k_{cat}$ were extracted from fitting of Michaelis-Menten hyperbolae in KaleidaGraph.

To study the activation of cathepsin A, carboxypeptidase activity was measured at multiple time points during maturation. Briefly, cathepsin A purified from insect cells was concentrated and incubated at 4°C. Kinetic parameters were measured using substrate concentrations 0.01 mM to 1.0 mM of Z-Phe-Leu (as described above) at various time points (0-21 days). The processing of cathepsin A at each time point was measured by quantifying band intensities on a reducing SDS-PAGE gel using the GeneTools software (Syngene).

RESULTS
The precursor and mature structures of cathepsin A superimpose, without need for a conformational change. To understand the activation of cathepsin A by conversion of precursor to mature protease, we determined the structure of recombinant insect-cell expressed mature cathepsin A by X-ray crystallography to 2.8Å resolution (Figure 1d). The structure of mature cathepsin A superimposes with the previously reported structure of precursor cathepsin A, with a root mean square deviation (rmsd) of 0.55Å for all Ca atoms. The core domains (1-182 and 303-452) of mature and precursor cathepsin A superimpose with an rmsd of 0.28Å on 332 Ca atoms. The cap domains (183-302) of precursor and mature cathepsin A also superimpose well, with an rmsd of 0.32Å, indicating no substantial conformational changes take place after maturation. As in the precursor cathepsin A structure, each monomer of the dimeric protease has four disulfide bonds and two N-linked carbohydrates per monomer (Asn117 and Asn305). The carbohydrate at Asn117 participates in a crystal contact, making crystallization difficult.

As expected for an active mature cathepsin A, electron density for the excision peptide (285 to 297) is not observed in the structure of mature cathepsin A. However, a much larger region (259 to 297, including the putative blocking peptide 272-277) is missing from the electron density, suggesting that the excision peptide might be larger than previously hypothesized (Figure 1d). In the structure of mature cathepsin A, the absence of the blocking and excision peptides results in a completely solvent accessible substrate-binding pocket. No significant conformational changes are observed in the residues involved in catalysis or substrate binding between the precursor and mature forms (Figure 2). The precursor and mature forms of human cathepsin A can each be superimposed on those of CPY (precursor cathepsin A to CPY 0.40Å rmsd; mature cathepsin A to CPY 0.51Å rmsd for all Ca atoms) and CPW (precursor cathepsin A to CPW 0.85Å rmsd; mature cathepsin A to CPW 0.55Å rmsd for all Ca atoms).

The excision peptide is larger than 1.6 kDa. The crystal structure of mature cathepsin A suggested that the number of residues removed during maturation might be larger than the previously proposed 13 residues (285-298, 1.6 kDa). To test this possibility, we subjected...
cathepsin A to SDS-PAGE and mass spectrometry analysis at different stages in the maturation process (Figure 3). Reducing SDS-PAGE analysis of cathepsin A shows the processing of the single chain precursor polypeptide into the two-chain mature form as a function of time. MALDI mass spectrometry of cathepsin A revealed changes in mass as the protein matured. The MALDI spectrum of purified cathepsin A showed primarily a 54.4 kDa glycoprotein for the zymogen (single chain) and a 51.1 kDa glycoprotein (large plus small subunits linked by disulfide bonds) at complete maturation, with a mixture of the two during intermediate stages. The difference in the masses of the precursor and mature cathepsin A samples in the mass spectrometer reveal the size of the peptide removed during processing is 3.3 ± 0.2 kDa (Figure 3).

The precursor is cleaved at Ser293 to form the small subunit of the mature protease. Proteolytic processing of precursor cathepsin A into mature cathepsin A produces a new N-terminal residue in the small subunit. To identify the precise proteolytic location during maturation, the small subunit of mature cathepsin A was N-terminally sequenced by Edman degradation. The sequencing unambiguously identifies Ser293 as the first residue of the small subunit (Figure 4a). Because the mass spectrometry shows the excision peptide is 3.3 kDa and the first residue of the small subunit is Ser293, we estimate that the other cleavage reaction occurs after Arg262 or Lys265 (Figure 4a). Removal of the 263-292 or 266-292 peptide would result in a mass reduction of 3.6 kDa or 3.2 kDa respectively, within the error estimate of the mass spectrometry measurement on the glycoprotein (3.3 ± 0.2 kDa).

Arg292 immediately precedes Ser293, suggesting processing by a trypsin-like protease. In contrast, Arg298 (which had been previously proposed to be removed during maturation) is clearly seen in the electron density maps of mature cathepsin A. Leupeptin reduced the proteolytic processing of cathepsin A (N.K. and S.C.G., unpublished), consistent with processing by a trypsin-like protease. In an effort to suppress cleavage of the excision peptide, we expressed and purified a cathepsin A double mutant R262A/R292A, which also undergoes processing to form large and small subunits, suggesting alternative avenues for the maturation of cathepsin A. Additionally, to suppress cleavage of the predicted 1.6 kDa excision peptide, we generated a triple mutant form of cathepsin A, changing the two previously proposed cleavage sites (Arg284 and Arg298) and the catalytic nucleophile (Ser150) to alanine. The cathepsin A triple mutant S150A/R284A/R298A also undergoes cleavage into large and small subunits, comparable to the wild-type cathepsin A (Figure 4b), suggesting these sites are not mandatory for the activation of cathepsin A.

To examine maturation of cathepsin A in human cells, we expressed recombinant human cathepsin A constructs in HEK293T cells. Cathepsin A expressed in human cells mimicked that expressed in insect cells, undergoing processing into large and small subunits. We replaced Arg and Lys residues in the processing region (on either side of the blocking peptide) with Alas, and the resulting proteins were analyzed for proteolytic processing. Seven cathepsin A variants expressed in HEK293T cells were processed into large and small subunits (Figure 4c), suggesting multiple processing sites can lead to mature cathepsin A. Amino-acid sequencing results, alanine-scanning mutagenesis, and protease inhibition all point to the specificity of the maturation protease as trypsin-like.

Enzymatic activity increases with maturation of cathepsin A to the 51.1 kDa form. Cathepsin A is a serine carboxypeptidase that cleaves terminal amino acids from oligopeptides in the lysosome (7,8). The carboxypeptidase activity has been studied previously with endogenous mature cathepsin A purified from placental tissue using the synthetic dipeptide Z-Phe-Leu (34). To test the activity of the insect cell-expressed mature cathepsin A activated by removal of blocking and excision peptides by endogenous proteases, Z-Phe-Leu was used to measure $K_M$ and $k_{cat}$. The carboxypeptidase activity was measured by detecting the amount of Leu released using the amine reactive reagent TNBS. Cathepsin A activity was highest at acidic pH (Figure 5), as is common for lysosomal enzymes. Recombinant human cathepsin A showed a $K_M$ of 0.04 mM and a $k_{cat}$ of 12.1 s⁻¹, similar to previously reported values for endogenous human placental cathepsin A (0.07 mM $K_M$ and 35 s⁻¹ $k_{cat}$) against a comparable substrate (Figure 5).
We examined the activity of cathepsin A as a function of its proteolytic maturation. The maturation of full-length zymogen into cleaved mature form was followed over time and quantified by SDS-PAGE. At each maturation time point, the full kinetic profile of the carboxypeptidase activity of cathepsin A (0.01 mM to 1.0 mM substrate concentrations) was measured (Figure 5d and e). The increase in carboxypeptidase activity of cathepsin A mirrors the amount of proteolytically cleaved cathepsin A. Conversion of the intermediate form to the completely processed mature protease with large (31 kDa) and small (20 kDa) subunits results in maximal activity.

DISCUSSION
Cathepsin A undergoes processing of the inactive zymogen to generate an active mature protease. The structure of the precursor cathepsin A zymogen led to the hypothesis that enzymatic activity required cleavage of a 1.6 kDa peptide followed by conformational changes in the cap domain. The mass spectrometry, N-terminal sequencing, crystallographic, and cellular data presented here are inconsistent with the conformational change model of cathepsin A maturation. Instead, we propose an alternate model, where removal of a larger excision peptide (including the blocking peptide), allows direct access of substrate to the active site (Figure 6). Maximal cathepsin A activity is attained after removal of both the blocking and excision peptides.

Proteases are generally synthesized as inactivezymogens that undergo proteolytic activation to form the active mature enzyme, a mechanism allowing strict control of catalytic activity. Proteases use a variety of mechanisms for controlling activation. Some (e.g. lipases and caspasess) undergo significant conformational changes or dimerization to generate a viable active site. In others (e.g. chymotrypsin and trypsin), segments of the zymogen are removed either by autocatalysis or limited proteolysis, followed by conformational changes to form the functional active site. In others (e.g. wheat and metallo-carboxypeptidases, and many other cathepsins), the active site is preformed and activation requires only limited proteolysis. Our data suggest that cathepsin A is activated in a similar manner to the proteases.
multiple proteases potentially allows for regulation of cathepsin A activity across many cell types.

Our data indicate the complete removal of the blocking peptide in the mature cathepsin A generated by endogenous insect cell proteases. The enzymatic properties of mature human cathepsin A from insect cells are comparable to those of endogenous mature cathepsin A purified from placental tissue, suggesting full activation of the insect cell-expressed cathepsin A. We monitored the activity of cathepsin A during its proteolytic maturation, revealing that the activity increases as the precursor and intermediates are processed to the final mature form. A second proteolytic cleavage, which results in the removal of the blocking and excision peptides, is required for the maximal activation of cathepsin A.

The importance of cathepsin A in prodrug conversion and cardiac regulation indicates its potential as a drug target, and inhibitors of cathepsin A are currently in clinical trials. Given long-standing uncertainty in the activation mechanism of cathepsin A and its central role in human disease, the results presented here can assist in the design of mechanism-based inhibitors for use as pharmaceuticals.
REFERENCES

1. d'Azzo, A., Andria, G., Strisciuglio, P., and Galjaard, H. (2001) Galactosialidosis. in The Metabolic and Molecular Bases of Inherited Disease (Scriver, C. S., et al. ed.), 8th Ed., McGraw-Hill, New York. pp 3811-3826

2. van der Spoel, A., Bonten, E., and d'Azzo, A. (1997) Role of protective protein cathepsin A in the activation of lysosomal neuraminidase. Am J Hum Genet 61, A27

3. van der Spoel, A., Bonten, E., and d'Azzo, A. (1998) Transport of human lysosomal neuraminidase to mature lysosomes requires protective protein/cathepsin A. EMBO J 17, 1588-1597

4. Bonten, E. J., and d'Azzo, A. (2000) Lysosomal neuraminidase. Catalytic activation in insect cells is controlled by the protective protein/cathepsin A. J Biol Chem 275, 37657-37663

5. Hiraiwa, M. (1999) Cathepsin A/protective protein: an unusual lysosomal multifunctional protein. Cell Mol Life Sci 56, 894-907

6. Morreau, H., Galjart, N. J., Willemsen, R., Gillemans, N., Zhou, X. Y., and d'Azzo, A. (1992) Human lysosomal protective protein. Glycosylation, intracellular transport, and association with beta-galactosidase in the endoplasmic reticulum. J Biol Chem 267, 17949-17956

7. Galjart, N. J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E. J., and d'Azzo, A. (1991) Human lysosomal protective protein has cathepsin A-like activity distinct from its protective function. J Biol Chem 266, 14754-14762

8. Itoh, K., Takiyama, N., Kase, R., Kondoh, K., Sano, A., Oshima, A., Sakuraba, H., and Suzuki, Y. (1993) Purification and characterization of human lysosomal protective protein expressed in stably transformed Chinese hamster ovary cells. J Biol Chem 268, 1180-1186

9. Jackman, H. L., Morris, P. W., Deddish, P. A., Skidgel, R. A., and Erdos, E. G. (1992) Inactivation of endothelin I by deamidase (lysosomal protective protein). J Biol Chem 267, 2872-2875

10. Cuervo, A. M., Mann, L., Bonten, E. J., d'Azzo, A., and Dice, J. F. (2003) Cathepsin A regulates chaperone-mediated autophagy through cleavage of the lysosomal receptor. EMBO J 22, 47-59

11. Itoh, K., Kase, R., Shimmoto, M., Satake, A., Sakuraba, H., and Suzuki, Y. (2002) Protective protein as an endogenous endothelin degradation enzyme in human tissues. J Biol Chem 270, 515-518

12. Jackman, H. L., Massad, M. G., Sekosan, M., Tan, F., Brovkovych, V., Marcic, B. M., and Erdos, E. G. (2002) Angiotensin 1-9 and 1-7 release in human heart: role of cathepsin A. Hypertension 39, 976-981

13. Ruf, S., Buning, C., Schreuder, H., Horstick, G., Linz, W., Olpp, T., Pernerstorfer, J., Hiss, K., Kroll, K., Kannt, A., Kohlmann, M., Linz, D., Hubschle, T., Rutten, H., Wirth, K., Schmidt, T., and Sadowski, T. (2012) Novel beta-amino acid derivatives as inhibitors of cathepsin A. J Med Chem 55, 7636-7649

14. Ruf, S., Buning, C., Schreuder, H., Linz, W., Hubschle, T., Linz, D., Ruetten, H., Wirth, K., and Sadowski, T. (2013) Inhibition of CatA: an emerging strategy for the treatment of heart failure. Future Med Chem 5, 399-409

15. Birkus, G., Kutty, N., Frey, C. R., Shribata, R., Chou, T., Wagner, C., McDermott, M., and Cihlar, T. (2011) Role of cathepsin A and lysosomes in the intracellular activation of novel antipapillomavirus agent GS-9191. Antimicrob Agents Chemother 55, 2166-2173

16. Endrizzi, J. A., Breddam, K., and Remington, S. J. (1994) 2.8 Å structure of yeast serine carboxypeptidase. Biochemistry 33, 11106-11120

17. Bullock, T. L., Branchaud, B., and Remington, S. J. (1994) Structure of the complex of L-benzylsuccinyl with wheat serine carboxypeptidase II at 2.0 Å resolution. Biochemistry 33, 11127-11134

18. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40, D343-350
19. Olesen, K., and Breddam, K. (1995) Increase of the P1 Lys/Leu substrate preference of carboxypeptidase Y by rational design based on known primary and tertiary structures of serine carboxypeptidases. *Biochemistry* **34**, 15689-15699

20. Rudenko, G., Bonten, E., d’Azzo, A., and Hol, W. G. (1995) Three-dimensional structure of the human ‘protective protein’: structure of the precursor form suggests a complex activation mechanism. *Structure* **3**, 1249-1259

21. Rudenko, G., Bonten, E., d’Azzo, A., and Hol, W. G. (1996) Structure determination of the human protective protein: twofold averaging reveals the three-dimensional structure of a domain which was entirely absent in the initial model. *Acta Crystallogr D Biol Crystallogr* **52**, 923-936

22. Rudenko, G., Bonten, E., Hol, W. G., and d’Azzo, A. (1998) The atomic model of the human protective protein/cathepsin A suggests a structural basis for galactosialidosis. *Proc Natl Acad Sci U S A* **95**, 621-625

23. Bonten, E. J., Galjart, N. J., Willemsen, R., Usmany, M., Vlak, J. M., and d’Azzo, A. (1995) Lysosomal protective protein/cathepsin A. Role of the "linker" domain in catalytic activation. *J Biol Chem* **270**, 26441-26445

24. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology: Macromolecular Crystallography, part A* (Carter, C. W., and Sweet, R. M. eds.), Academic Press. pp 307-326

25. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *Acta Crystallogr A* **60**, 2126-2132

26. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367

27. Cornish-Bowden, A. (2004) *Fundamentals of Enzyme Kinetics*, 3rd ed., Portland Press, London

28. Pshezhetsky, A. V., Vinogradova, M. V., Elsliger, M. A., el-Zein, F., Svedas, V. K., and Potier, M. (1995) Continuous spectrophotometric assay of human lysosomal cathepsin A/protective protein in normal and galactosialidosis cells. *Anal Biochem* **230**, 303-307

29. Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whisstock, J. C., and Pike, R. N. (2012) PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One* **7**, e50300

30. Verspurten, J., Gevaert, K., Declercq, W., and Vandenabeele, P. (2009) SitePredicting the cleavage of proteinase substrates. *Trends Biochem Sci* **34**, 319-323

31. Haque, A., Banik, N. L., and Ray, S. K. (2008) New insights into the roles of endolysosomal cathepsins in the pathogenesis of Alzheimer's disease: cathepsin inhibitors as potential therapeutics. *CNS Neurol Disord Drug Targets* **7**, 270-277
38. Wittlin, S., Rosel, J., Hofmann, F., and Stover, D. R. (1999) Mechanisms and kinetics of procathepsin D activation. *Eur J Biochem* 265, 384-393

ACKNOWLEDGEMENTS
This work was funded by NIH grant R01 DK76877 to S.C.G. We thank Igor Kourinov and the staff of the Advanced Photon Source Northeastern Collaborative Access Team beamlines, which are supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. We thank Rinat Abzalimov of the UMass Amherst mass spectrometry center for the mass spectroscopy measurements.

FOOTNOTES
The abbreviations used are: PPCA, protective protein/cathepsin A; CPW, wheat carboxypeptidase; CPY, yeast carboxypeptidase; NEU1, neuraminidase; GLB1, β-galactosidase; LAMP2, lysosome-associated membrane protein type 2

COORDINATE DEPOSITION
Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 4MWS and 4MWT.

FIGURE LEGENDS

Figure 1: Structural comparison of precursor and mature cathepsin A reveals no large conformational changes. (a) Primary structure of precursor cathepsin A (green) showing the position of the N-linked glycans, blocking peptide (dark blue), excision peptide (yellow), and connecting sequence (cyan). Disulfide bonds are depicted as brackets. (b) and (c) Each monomer of cathepsin A is composed of a core hydrolase domain containing the catalytic residues (spheres) and the cap domain. In the precursor, the active site is occluded by the blocking peptide. (d) Structure of mature cathepsin A dimer (purple) superimposes on the precursor cathepsin A (green) without conformational changes. Removal of the blocking peptide, excision peptide, and connecting sequence in the mature protease renders the active site accessible for substrate binding (surface representation). In the inset, 2Fo-Fc density from a σA-weighted map contoured at 1σ shows the 253-303 disulfide flanking the region processed during maturation.

Figure 2: The active site residues in the structure of mature cathepsin A are competent for catalysis. The substrate-binding residues are matched to those of wheat carboxypeptidase complex with benzylsuccinic acid (PDB 1WHS). (a) Superposition of precursor (green) and mature cathepsin A (purple) structures show no conformational changes in the terminal carboxylicate binding site (blue), S1’ binding site (orange), the S1 binding site (cyan) of the mature protease or the residues adjoining the conserved disulfide bond (C253-C303) proximal to the cleavage sites. (b) No conformational changes are observed in the structure of mature cathepsin A for these residues on inhibitor binding (light brown, PDB 4AZ0). (c and d) Superimposition of yeast (grey) and wheat (pale green) carboxypeptidases on mature cathepsin A (purple) shows conservation of the substrate binding and catalytic residues across these homologous proteins. The disulfide bond corresponding to C253-C303 in cathepsin A is conserved among the three proteases. Wheat carboxypeptidase undergoes cleavage at the region adjoining the disulfide bond without any conformational change, similar to mature cathepsin A.

Figure 3: The excision peptide is 3.3 kDa. (a) Reducing SDS-PAGE shows full-length precursor cathepsin A (lane 2) undergoing proteolytic cleavage to form large and small subunits of the mature cathepsin A (lane 4), through a processing intermediate (lane 3). (b) MALDI spectra of cathepsin A forms (lanes 2-4 from the SDS gel) during maturation. Sample one (top panel) is predominantly 54.4 kDa precursor with a small concentration of 51.1 kDa mature cathepsin A. The middle panel shows similar
amounts of full-length and processed cathepsin A, giving rise to two peaks in the spectrum. The completely processed cathepsin A (bottom panel) shows a single peak at 51.1 kDa. Arrows point to zoomed insets of each peak. The mass measurements of the glycoprotein are accurate to 1-2 amino acids (scale bar indicates 0.1 kDa).

**Figure 4**: (a) The blocking peptide is removed during maturation. Electron density is absent for the 259-297 sequence (grey) in mature cathepsin A (purple) but appears in the precursor structure (green). N-terminal sequencing of the mature cathepsin A small subunit and MALDI analysis suggest two potential excision peptides (yellow), option 1 (266-292, 3.2 kDa) and option 2 (266-292, 3.6 kDa). The molecular weight of the excision peptide originally proposed in 1995 (1.6 kDa, yellow) is inconsistent with mass spectrometric and amino-acid sequencing of the small subunit of mature cathepsin A. (b) and (c) Cathepsin A can be activated by cleavage at multiple cleavage sites. (b) Suppressing cleavage at the sites predicted in (a) by expressing and purifying Ala variants of cathepsin A (R262A/R292A and R284A/R298A/S150A) also generates mature cathepsin A, suggesting processing at alternative sites. Wild type and variants all undergo cleavage of the precursor to form mature large and mature small subunits. Lanes 1-3 are 12% acrylamide, and lane 4 is 10%. (c) Alanine scanning of trypsin-like protease cleavage sites in human cells. Immunoprecipitation of cathepsin A variants around the blocking peptide (in the nucleophile knock out background) in HEK293T cells with anti-FLAG antibody indicates cleavage of the precursor to matured large and small subunits. IP using anti-FLAG antibody was western blotted with anti-cathepsin A antibody (top panel), and lysates were western blotted with anti-FLAG antibody (bottom panel).

**Figure 5**: Cleavage of the blocking peptide is required for maximal carboxypeptidase activity. (a) Activity of mature cathepsin A peaks at pH 5.0 (b) Michaelis-Menten plot of carboxypeptidase activity vs. substrate concentration. (c) The kinetic parameters for recombinant cathepsin A activity against Z-Phe-Leu compared to those of purified endogenous cathepsin A (34). (d) Analysis of cathepsin A on reducing SDS-PAGE shows the maturation progress of full length cathepsin A into an intermediate form upon cleavage, which is further processed to form the mature large and small subunits. (e) The maturation of cathepsin A in (c) is quantified and plotted against maturation progress by measuring the band intensities of full length (○), intermediate (♦), mature large (■) and mature small (▲) subunits. Cathepsin A activity as apparent $v_{\text{max}}$ in mM/min (grey bar) measured at each time point plotted against the maturation progress shows an increase in activity upon conversion of the intermediate to mature form.

**Figure 6**: Model of cathepsin A maturation. (a) The originally proposed model for precursor cathepsin A (green) maturation involves the cleavage of a 1.6 kDa peptide (285-298, yellow) followed by conformational changes in the blocking peptide (272-277, dark blue) and the sequence connecting the blocking and excision peptides (278-284, cyan) to allow solvent accessibility of the catalytic triad (spheres). (b) We propose an alternative model involving cleavage of a larger 3.3 kDa peptide (263-292, that includes the blocking peptide, the 1.6 kDa excision peptide and the sequence connecting the two) to form the mature active protease (purple). Alternate termini are also possible in the cleavage model.
### Table 1: Crystallographic data

|                  | Native 1          | Native 2          |
|------------------|-------------------|-------------------|
| PDB code         | 4MWS              | 4MWT              |
| Beamline         | APS 24-ID-C       | APS 24-ID-C       |
| Wavelength, Å    | 0.9792            | 0.9792            |
| Space Group      | $P3_121$          | $P3_121$          |
|                  | [52:48 twin]      |                   |
| $a$ and $c$ cell lengths, Å | 134.9, 99.9       | 134.6, 99.8       |
| Resolution, Å*   | 50-2.8 (2.85-2.8) | 50-3.85 (3.92-3.85)|
| Observations*    | 142,222 (6,536)   | 33,205 (1,694)    |
| Unique observations* | 25,989 (1,257)   | 9,724 (404)     |
| Completeness*    | 99.0 (98.0)       | 95.2 (98.0)       |
| Multiplicity*    | 5.5 (5.2)         | 3.5 (3.4)         |
| $R_{\text{sym}}$, %*  | 14.9 (88.4)      | 17.5 (52.7)      |
| $<I/\sigma_I>$,* | 15.7 (1.7)        | 5.9 (1.8)         |
| $R_{\text{work}}$/$R_{\text{free}}$, %  | 14.7 / 19.5      | 32.3 / 32.5    |
| Atoms            | 6755              | 6677              |
| Protein          | 6602              | 6596              |
| Carbohydrate     | 137               | 81                |
| Water            | 4                 | 0                 |
| Other            | 34                | 0                 |
| Average B-factor (Å²) | 58                | 81                |
| Protein          | 67                | 81                |
| Carbohydrate     | 87                | 154               |
| Water            | 40                | -                 |
| Ramachandran plot, % | 95.5              | 95.7             |
| Favored          |                    |                   |
| Allowed          | 4.2               | 3.9               |
| Outlier          | 0.4               | 0.4               |
| RMS deviations   |                   |                   |
| Bonds, Å         | 0.008             | 0.008             |
| Angles, °        | 1.16              | 1.15              |

*Highest resolution shell appears in parentheses
Kolli & Garman, Figure 2
Kolli & Garman, Figure 3
(a) Glycoprotein mass:

|                  | Calculated | Observed |
|------------------|------------|----------|
| Precursor        | 54328      | 54400    |
| Mature           |            |          |
| Excision peptides|            |          |
| Option 1         | 51127      |          |
| Option 2         | 50707      |          |
| Original prediction | 52650      |          |

(b) Reducing SDS / Coomassie

(c) IP: FLAG / WB: cathepsin A

Kolli & Garman, Figure 4
Kolli & Garman, Figure 5
(a) Conformational change model

Blocking peptide

Excision peptide

Precursor cathepsin A (PDB: 1IVY)

Cleavage

Conformational change

Proposed structure (1995)

(b) Cleavage model

Excision Peptide

(Alternate termini are possible)

Precursor cathepsin A (PDB: 1IVY)

Mature cathepsin A (PDB: 4MWS)

Kolli & Garman, Figure 6
