Cell Research (2014) 24:344-358.

www.nature.com/cr

ORIGINAL ARTICLE

Structural analysis of asparaginyl endopeptidase reveals the activation mechanism and a reversible intermediate maturation stage

Lixia Zhao[1, 2, *], Tian Hua[1, *], Christopher Crowley[3], Heng Ru[1], Xiangmin Ni[1], Neil Shaw[1], Lianying Jiao[1], Wei Ding[1], Lu Qu[1], Li-Wei Hung[4], Wei Huang[2], Lei Liu[5], Keqiang Ye[6], Songying Ouyang[1], Genhong Cheng[3], Zhi-Jie Liu[1, 2]

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; [2]iHuman Institute, ShanghaiTech University, Shanghai 201210, China; [3]Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, CA 90095, USA; [4]Physics Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA; [5]Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology (Ministry of Education), Department of Chemistry, Tsinghua University, Beijing 100084, China; [6]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322, USA

Asparaginyl endopeptidase (AEP) is an endo/lysosomal cysteine endopeptidase with a preference for an asparagine residue at the P1 site and plays an important role in the maturation of toll-like receptors 3/7/9. AEP is known to undergo autoproteolytic maturation at acidic pH for catalytic activation. Here, we describe crystal structures of the AEP proenzyme and the mature forms of AEP. Structural comparisons between AEP and caspases revealed similarities in the composition of key residues and in the catalytic mechanism. Mutagenesis studies identified N44, R46, H150, E189, C191, S217/S218 and D233 as residues that are essential for the cleavage of the peptide substrate. During maturation, autoproteolytic cleavage of AEP's cap domain opens up access to the active site on the core domain. Unexpectedly, an intermediate autoproteolytic maturation stage was discovered at approximately pH 4.5 in which the partially activated AEP could be reversed back to its proenzyme form. This unique feature was confirmed by the crystal structure of $AEP_{pH4.5}$ (AEP was matured at pH 4.5 and crystallized at pH 8.5), in which the broken peptide bonds were religated and the structure was transformed back to its proenzyme form. Additionally, the AEP inhibitor cystatin C could be digested by the fully activated AEP, but could not be digested by activated cathepsins. Thus, we demonstrate for the first time that cystatins may regulate the activity of AEP through substrate competition for the active site.

Keywords: asparaginyl endopeptidase; autoproteolytic maturation; crystal structure; innate immunity
Cell Research (2014) **24**:344-358. doi:10.1038/cr.2014.4; published online 10 January 2014

Introduction

Mammalian asparaginyl endopeptidase (AEP) or legumain (EC 3.4.22.34) is a lysosomal cysteine endopeptidase that specifically cleaves substrates with an asparagine or aspartate residue at the P1 site [1]. It has been detected in heart, spleen, liver, brain and testis tissues, but it is particularly abundant in the kidney and placenta [2]. AEP is predominantly localized in the late endosomes and lysosomes [2] and belongs to the C13 peptidase family. AEP is unrelated to the papain-like C1 family of lysosomal-cysteine proteases, such as the cathepsins S, B, L and H, but has features that suggest an evolutionary relationship to caspases, the bacterial proteases gingipain

*These two authors contributed equally to this work.
Correspondence: Zhi-Jie Liu[a], Genhong Cheng[b], Songying Ouyang[c]
[a]Tel: + 86-10-64857988; Fax: + 86-10-64888426
E-mail: zjliu@ibp.ac.cn
[b]Tel: 310-825-8896; Fax: 310-206-5553
E-mail: gcheng@mednet.ucla.edu
[c]Tel: + 86-10-64888252; Fax: + 86-10-64888426
E-mail: ouyangsy@moon.ibp.ac.cn
Received 12 July 2013; revised 17 September 2013; accepted 5 November 2013; published online 10 January 2014

and clostripain and separase [3].

The inactive full-length pro-AEP can be activated by shifting the pH from neutral to acidic. AEP appears to be autocatalytically cleaved after asparagine or aspartate residue. AEP is also involved in the proteolytic maturation of pro-cathepsins B, H and L found in the endo/lysosomes [4]. In addition to the cathepsins, AEP also processes fibronectin [5], pro-gelatinase A [6] and α-thymosin [7].

Cystatin C is a secreted type 2 cysteine proteinase inhibitor and exerts effects on multifarious cysteine proteases including papain and cathepsins B, H, K, L and S. As the most abundant extracellular inhibitor of cysteine proteases, cystatin C is encoded by a housekeeping gene and is found ubiquitously in all organs and biological fluids at high concentration levels. Cystatin C can also inhibit AEP activity [8]. However, the inhibitory mechanisms are still elusive.

AEP is known to play a role in the processing of antigens for MHC class II presentation in the lysosomes of antigen presenting cells [2, 9-11]. More recently, AEP has been reported to regulate innate immune responses via its participation in the maturation of the toll-like receptors (TLRs) 3, 7 and 9 [12-14]. These TLRs undergo proteolytic maturation, which is required for their downstream signaling events [15]. A 72 kDa C-terminal fragment of TLR9 that is generated by proteolytic cleavage associates with the adaptor molecule MyD88 and initiates the production of pro-inflammatory cytokines and other co-immunostimulatory molecules [16].

Intriguingly, in addition to its known roles in catalyzing peptide bond cleavage, AEP is also responsible for ligation of cyclotides via transpeptidation reactions in a single processing event [17]. To date, there are at least three examples of AEP-mediated formation of cyclic peptides in phylogenetically distant plants [18], such as kalata B1 from *Oldenlandia affinis* (Rubiaceae) and cycloviolacin O11 from *Viola odorata* (Violaceae). However, little is known about the mechanism of AEP-mediated cyclic peptide formation. Additionally, the ligation properties of AEP have not yet been observed in mammalian systems.

Here, we describe crystal structures of the proenzyme and the mature forms of mouse AEP, which clearly illustrate the basis for the inability of the proenzyme to perform catalysis. We show that the maturation of AEP requires the removal of a cap that covers the active site. This process is reversible and pH-dependent. Structure-based mapping of the active site residues using site-directed mutagenesis suggests a cysteine endopeptidase-type catalytic mechanism that is similar to that observed for caspases.

Results

Characterization of the proenzyme and mature forms of AEP

Initially, human and mouse AEPs (83% sequence identity) were selected for structural and functional analyses (Figure 1A). Both proteins were expressed in insect cells, and the secreted proteins in the cell culture supernatants were collected and purified as soluble proteins (Supplementary information, Figure S1A). The subsequent structure-based experiments were primarily focused on mouse AEP; therefore, unless specified, AEP refers to mouse AEP hereafter. The theoretical molecular weight (MW) of pro-AEP without the signal peptide (residues 18-435) is ~47 kDa (Figure 1B). Mass spectrometry analysis of pro-AEP produced in insect cells and purified at pH 7.5 indicated a MW of 53.3 kDa (Supplementary information, Figure S1B). This result combined with those from endoglycosidase treatment suggested that pro-AEP was probably glycosylated (Supplementary information, Figure S1C).

The endopeptidase activity of AEP was measured using Z-AAN-NHMec as the substrate. As the pH of the incubation buffer decreased from neutral to below 5.0, the endopeptidase activity of AEP gradually increased, which peaked (full activity) at approximately pH 3.5 and then dramatically decreased as the pH further decreased (Figure 1C). The activity of AEP towards the substrate Z-AAN-NHMec was inhibited by the AAN-CMK inhibitor [19] (Supplementary information, Figure S2A). Additionally, both fully activated AEP at pH 3.5 ($AEP_{pH3.5}$) and partially activated AEP at pH 4.5 ($AEP_{pH4.5}$) were able to cleave recombinant protein GST-TEV-ubiquitin (Ub) at the asparagine (N) site, as identified by N-terminal sequencing (Supplementary information, Figure S2B and S2C).

To investigate the oligomerization and autocleavage status of AEP during the autoactivation process, several biochemical and biophysical methods were employed. The size exclusion chromatography (SEC) elution profile (Supplementary information, Figure S1D) indicated that pro-AEP forms a homodimer at neutral pH. To verify the oligomeric state of pro-AEP in solution, we performed small-angle X-ray scattering (SAXS) analysis, which confirmed that the protein exists as a homodimer in solution (Supplementary information, Figure S3). Furthermore, SDS-PAGE analysis of pro-AEPs that were incubated under different pH conditions revealed that the autoproteolytic maturation of pro-AEP was pH-dependent (Figure 1D). Interestingly, when pro-AEP was activated at pH 5.5-4.5, a ~36 kDa band corresponding to the core domain and several bands of ~13 kDa corresponding to the C-terminal domain (CTD) appeared in the SDS-PAGE gel, indicating that the peptide bonds linking the two domains were cleaved at multiple sites (Figure 1D). However, at this stage, $AEP_{pH4.5}$ exhibited only partial activity (Figure 1C) and was eluted at the same time as pro-AEP in the SEC elution profile, indicating that the cleaved CTD remained associated with the core domain (Supplementary information, Figure S1D). In contrast, $AEP_{pH3.5}$ and $AEP_{pH4.0}$ were eluted as monomers in SEC analysis (Supplementary information, Figure S1D and S1E). SDS-PAGE analysis revealed that the bands of approximately 13 kDa corresponding to the CTD disappeared when the pH was decreased from 4.5 to 4.0 (Figure 1D). This result suggests that the CTD of pro-AEP began to dissociate from the core domain and presumably was degraded when the pH was decreased below 4.5. Collectively, these results suggest that AEP activated at a pH below 4.5 existed as a monomer in solution, whereas pro-AEP existed as a homodimer, the formation of which was probably mediated by the CTD.

When the autocleavage sites were further investigated using N-terminal amino acid sequencing, two new

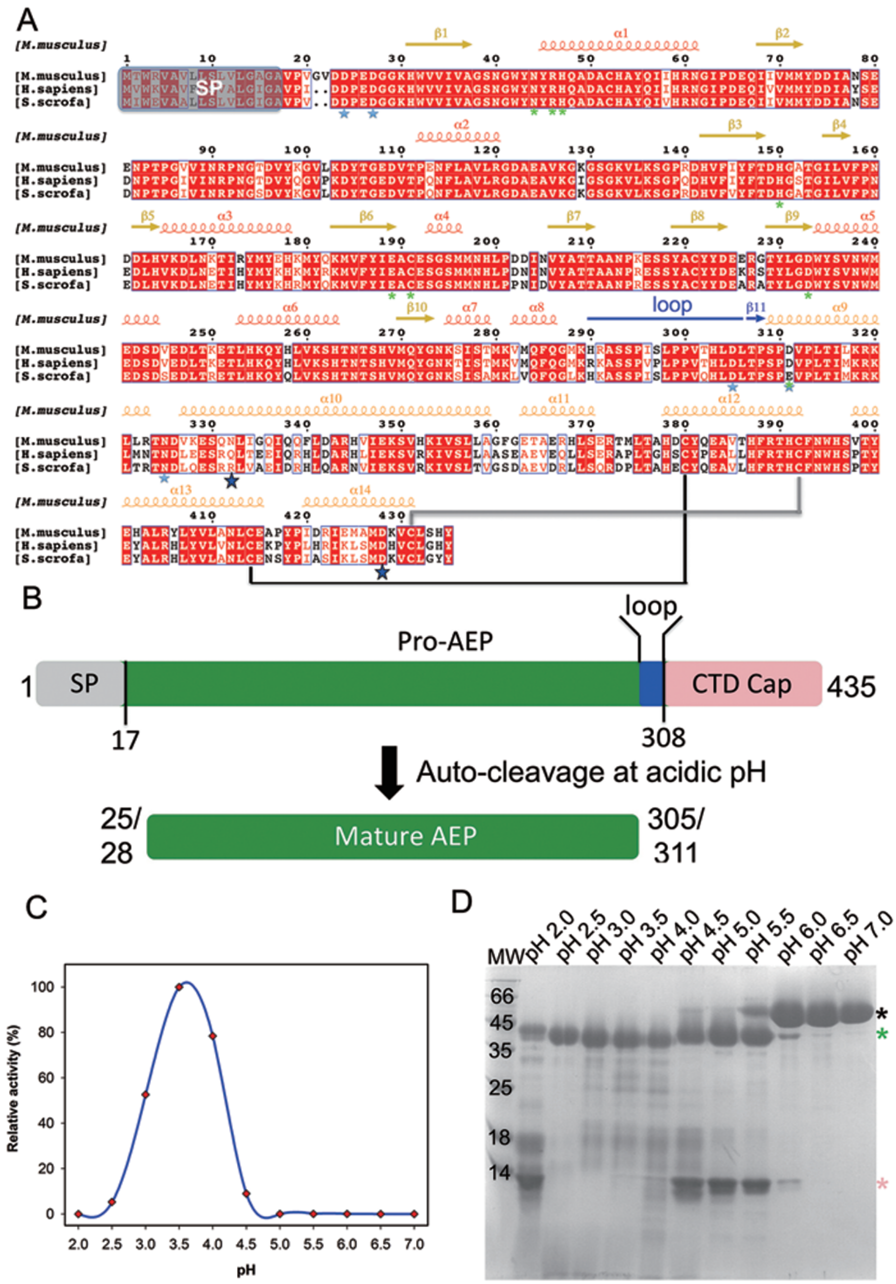

Figure 1 Characterization of the proenzyme and mature forms of AEP. **(A)** Multiple sequence alignment of AEP, which includes the secondary structure elements based on the AEP structure at pH 7.5 (refer to Figure 2B). Strictly conserved residues are boxed in white on a red background and highly conserved residues are boxed in red on a white background. The secondary structure elements of mouse AEP are shown above the sequences and every ten residues are indicated with a dot (·). Previously reported cleavage sites and two newly identified sites in this study are indicated with marine and blue five-pointed stars, respectively. Key residues for AEP activity are indicated with green asterisks. Two disulfide bonds in the cap domain are indicated with black and grey lines. SP: signal peptide. The alignment was generated using ClustalW. α-Helices are depicted as coils, and β-strands are depicted as arrows. The figure was generated using ESPript (http://espript.ibcp.fr/ESPript/ESPript/). **(B)** Domain organization of mouse AEP and its autocleavage at acidic pH. The numbers indicate residues at the domain boundaries. SP: signal peptide (residues 1-17); CTD cap: the C-terminal cap domain; and Mature AEP: the core domain. **(C)** The relative endopeptidase activity of AEPs was measured by using Z-AAN-NHMec as the substrate. AEPs were activated by incubation under different pH conditions overnight at 16 °C. **(D)** The pH-dependent autoproteolytic activation of human pro-AEP illustrated using SDS-PAGE analysis. Pro-AEP that was purified at pH 7.5 was incubated at the indicated pH (citric acid buffer), which ranged from pH 7.0 to 2.0, at 16 °C for 16 h. MW: molecular weight marker. Black, green, and wheat asterisks represent the pro-AEP, AEP core domain, and AEP CTD, respectively.

sites, N332 and D428, were identified in addition to the previously reported N325, D24 and D27 sites [20, 21] (Supplementary information, Figure S1F). However, two previously reported autocleavage sites, D305 and D311 [22], were not detected in this study.

Structure of AEP$_{pH3.5}$

To study the molecular basis of the AEP activation process, we determined and compared the crystal structures of AEPs that were processed under different pH conditions. The pro-AEP purified at pH 7.5 (AEP$_{pH7.5}$) was incubated at pH 3.5 (AEP$_{pH3.5}$) for complete activation. AEP$_{pH3.5}$ was crystallized at pH 6.7, and the crystal diffracted to a resolution of 2.80 Å (Table 1). The structure of AEP$_{pH3.5}$ (residues K30-Q289) consists of only the core domain (Figure 2A). The CTD seems to have dissociated from the core domain after autoproteolytic processing at pH 3.5. The core domain consists of a central sheet composed of six strands of which five (β1, β2, β3, β6 and β7) are parallel with a sixth short antiparallel strand (β10). The central sheet appears to be flanked by helices on both sides (Figure 2A). Dali [23] and ProFunc [24] analyses identified caspases, particularly caspases 3, 7, 8 and 9, as the closest structural matches; for example, an RMSD of 1.96 Å over 167 residues was obtained for caspase 3 (PDB ID: 4EHK) [25]. The largest difference was observed at a long loop (Y73-T111) between the β2-strand and α2-helix. Two other distinct differences were observed between the α2-helix and β3-strand (L23) and between the β7-strand and α5-helix (L75) (Supplementary information, Figure S4). These caspases use a catalytic dyad that consists of a histidine (H121 in caspase 3) and a cysteine (C163 in caspase 3) for proteolytic processing. Superposition of the caspase structures with AEP structure identified a pair of amino acids, i.e., H150 and C191, as a potential catalytic dyad (Figure 2A). This pair of residues is located on the surface of AEP and appears to be accessible (Figure 2A). It is known that a carbonyl oxygen actively participates in the catalysis of caspases [26], such as the carbonyl oxygen of G122 in caspase 3. The carbonyl oxygen atom of N44, D149, G151, E189, S218, or D233 in AEP is within hydrogen-bonding distance of the catalytic dyad (H150 and C191) and may serve as an oxyanion hole for substrate accommodation (Figure 2A). Additionally, a structural comparison between AEP$_{pH3.5}$ and the complex of caspase 3 and its inhibitor Ac-DEVD-CMK (PDB ID: 4EHK) revealed that R46, H47 and E189-E192 contribute to the formation of the S1 pocket and that two β-strands (β8 and β9) located between the β7-strand and α5-helix probably contribute to substrate recognition (Supplementary information, Figure S4). Interestingly, this comparison also revealed that the loop L47 between the α4-helix and β7-strand sterically hinders dimer formation of mature AEPs (Supplementary information, Figure S4).

Structures of AEP$_{pH7.5}$ and AEP$_{pH4.5}$

The structures of AEP$_{pH7.5}$ and AEP$_{pH4.5}$ were determined at 2.49 and 2.0 Å resolution, respectively (Table 1). AEP$_{pH7.5}$ was crystallized at pH 7.0 and adopts a fold composed of 14 α-helices and 11 β-sheets (Figure 2B). The overall structure of AEP$_{pH7.5}$ is distinctly divided into two parts: a α/β-core domain (K30-Q289) that exhibits a TIM-barrel fold, and an entirely α-helical cap domain (S309-S435) that is located directly above the conserved catalytic surface of the core domain, and is similar to the death effector domain of caspases. Two pairs of internal disulfide bonds further stabilize the cap domain (Figures 2B and 1A). The cap domain is linked to the core domain via a long loop (G290-P308) (Figure 2B and 2C). In addition, the cap domain interacts with the core domain through extensive direct or solvent-mediated hydrogen bonds, as well as salt bridges (Supplementary information, Figure S5A, Tables S1 and S2). The link-

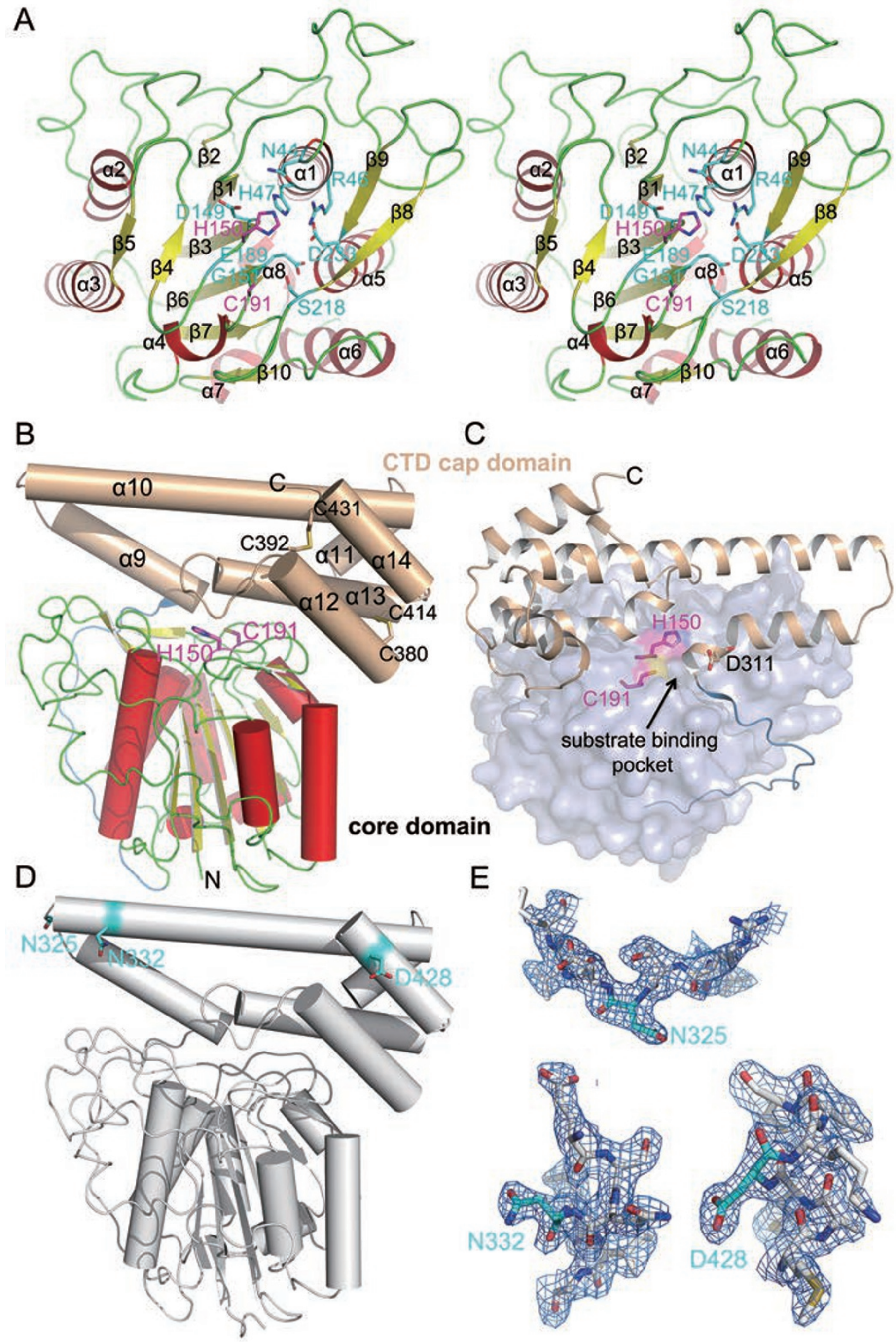

Figure 2 Overall structures of the mature AEP and pro-AEP. **(A)** Walleye stereoview of the structure of $AEP_{pH3.5}$. Key residues are shown as sticks. β-strands and α-helices are numbered. **(B)** The overall bimodular structure of the pro-AEP at pH 7.5 composed of the core domain and the entirely α-helical cap domain that is depicted as a cylindrical model (α-helices are numbered from α9 - α14, wheat). The active site, including the key residues H150 and C191, is covered by the cap domain. The long loop that connects the core domain and the cap domain is colored blue. The residues that form two pairs of internal disulfide bonds (C380 in α12 and C414 in α13, C392 in α12 and C431 in α14) in the cap domain are labeled. The N- and C-termini of each chain are labeled with the respective letters. **(C)** The indicated cavity (arrow) represents the potential substrate-binding pocket. The linkage area between the cap domain (wheat) and the long loop (blue) sterically obstructs substrate access to the active site. **(D)** The overall structure of activated $AEP_{pH4.5}$ (crystallized at pH 8.5), which is similar to the structure of $AEP_{pH7.5}$. The key cleavage sites were indicated as cyan sticks. **(E)** The 2.0 Å $2F_o$-F_c omit electron density map countered at 2.0σ for residues 30-435 is complete without breaks at either the previously reported cleavage site (N325) or the new cleavage sites of N332 and D428 identified using N-terminal amino acid sequencing in this study.

Table 1 Data collection and refinement statistics

AEP Derivatization	$AEP_{pH4.5}$ (Hg-derivatization)	$AEP_{pH7.5}$	$AEP_{pH3.5}$	AEP $D233A_{pH7.5}$
Data collection				
X-ray source	BL17U(SSRF)	BL17A(KEK)	BL17U(SSRF)	BL17U(SSRF)
Detector	ADSC Q315	ADSC 270	ADSC Q315	ADSC Q315
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12_1$	$P2_1$
Unit cell dimensions				
a, b, c (Å)	57.95, 167.72, 51.93	57.08, 168.90, 49.53	50.31, 59.58, 96.53	51.96, 163.36, 56.28
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 104.32, 90.00
Resolution range (Å)	50.00-2.00 (2.15-2.00)	50.00-2.49 (2.59-2.49)	50.00-2.80 (2.90-2.80)	50.00-2.70 (2.80-2.70)
R_{sym} (%)	0.07 (0.43)	0.08 (0.49)	0.127 (0.495)	0.08 (0.48)
Mean I/σI (I)	33.03 (4.00)	45.09 (2.45)	12.29 (2.20)	13.79 (2.38)
Completeness (%)	93.00 (65.90)	85.90 (56.20)	97.00 (95.10)	99.00 (99.80)
Redundancy	6.40 (4.10)	11.20 (5.40)	4.50 (4.50)	3.30 (3.20)
Refinement				
Resolution (Å)	32.62-2.00	19.97-2.49	47.99-2.80	43.58-2.70
No. of reflections	32142	17063	8038	23903
R_{work}/R_{free} (%)	20.44/24.28	20.22/25.19	23.27/29.40	21.67/27.52
No. of atoms				
Protein	3266	3247	2088	6190
Water	158	128	26	83
Mean B value (Å^2)	45.72	31.11	51.70	24.95
R.m.s deviations				
Bond lengths (Å)	0.007	0.003	0.006	0.003
Bond angles (°)	1.042	0.776	0.992	0.804
Ramachandran analysis				
Favored region (%)	97.04	97.26	92.64	96.19
Allowed region (%)	2.22	2.24	7.36	3.67
Outliers (%)	0.74	0.50	0	0.13

Numbers in parentheses represent the statistics for the highest resolution shell.

age area between the cap domain (wheat) and the long loop (blue) sterically obstructs substrate access to the nearby substrate-binding pocket of the core domain (Figure 2C). The negatively charged surface of the active site is neutralized by the positively charged interface of the cap domain under neutral pH conditions (Supplementary information, Figure S5B). This observation may explain the stability of pro-AEP under neutral pH conditions, whereas under acidic pH conditions, the surface of the active site would be protonated following the disassociation of the cap domain from the core domain. Although the asymmetric unit contained one molecule of pro-AEP, analysis of the symmetry mates revealed that pro-AEP was crystallized as a dimer (Supplementary information, Figure S5C). The intermonomer interactions are hydrophobic and are mediated by L321, L333, I334, I337 and L341 of the C-terminal cap domains (Supplementary information, Figure S5C), which form homotypic interactions. Thus, the packing of pro-AEP as a dimer in the crystal lattice is consistent with its oligomeric state in solution.

Superposition of the structure of the mature, active AEP with the structure of the AEP proenzyme revealed that the core domain did not undergo significant conformational changes upon activation. Superposition of the Cα atoms of the mature enzyme with those of the pro-AEP resulted in an RMSD of 0.34 Å over 248 residues (Supplementary information, Figure S5D). Because the mature, active AEP contained only the core domain, the cap domain observed in the structure of the inactive pro-AEP may sterically occlude access to the active site in the core domain.

To explore the activation process of AEP, we attempted to determine the crystal structure of $AEP_{pH4.5}$, which represents the turning point in the activation process. Unexpectedly, $AEP_{pH4.5}$ that was activated at pH 4.5 (in a buffer containing 20 mM citric acid, pH 4.5 and 150 mM NaCl) and crystallized at pH 8.5 exhibited an identical architecture as $AEP_{pH7.5}$ (Figure 2D). The electron density omit map for residues 30-435 did not contain breaks at either the previously reported cleavage sites (such as N325) or the two newly identified cleavage sites, N332 and D428 (Figure 2D and 2E). All the structural elements of the three forms of AEPs ($AEP_{pH7.5}$, $AEP_{pH4.5}$ and $AEP_{pH3.5}$) could be superposed with virtually no differences in the positions of the main-chain atoms (Supplementary information, Figure S5D).

The crystal structure of $AEP_{pH4.5}$ unveiled a striking phenomenon: the broken peptide bonds at pH 4.5 "healed" under the basic crystallization condition (pH 8.5), which resulted in the transformation of the partially activated $AEP_{pH4.5}$ back to its pro-AEP form. To eliminate the possibility of sample misplacement or contamination, a new batch of $AEP_{pH4.5}$ was prepared, and the SDS-PAGE analysis (revealed cleaved fragments), crystallization (pH 8.5), data collection and structure determination were carefully repeated. An identical crystal structure was obtained.

Proteolytic activation of AEP is reversible

To further confirm this new discovery, an aliquot of $AEP_{pH4.5}$, which exhibited a ~36 kDa band corresponding to the core domain and several additional bands (~13 kDa) in the SDS-PAGE gel (Figure 3A, lane pH 4.5), was diluted in a series of buffers of elevated pH values (5.0-7.5) and incubated overnight at 16 °C. As expected, as the pH increased, AEP gradually reversed back into its pro-AEP form (Figure 3A). To further cross-validate this reversion, the samples were then analyzed by western blot using an anti-His antibody, which recognizes the 6× His tag at the cap domain of pro-AEP (Supplementary information, Figure S5E). The result indicated that the C-terminal fragments (~13 kDa) generated during autoproteolysis were gradually re-ligated to form the proenzyme when the pH was increased towards 7.5 (Figure 3B). We then expressed AEP in HEK293 cells, and activated the purified AEP at pH 4.5. These activated AEP proteins could also be reversed to their proenzyme forms (Supplementary information, Figure S6). However, we found that AEP was not reversed to its proenzyme form if the activation pH was below 4.0 (Supplementary information, Figure S7A and Table S3).

Collectively, these results indicate that at an incubation pH of approximately 4.5, pro-AEP is partially activated and the cap domain remains associated with the core domain probably via non-covalent interactions, such as extensive salt bridges and direct or solvent-mediated hydrogen bonds, even though the peptide bond has been cleaved. However, at a pH below 4.0, the cap domain disassociates and consequently is degraded. Therefore, AEP can no longer be reversed to its proenzyme form. The core domain is fully activated at this stage.

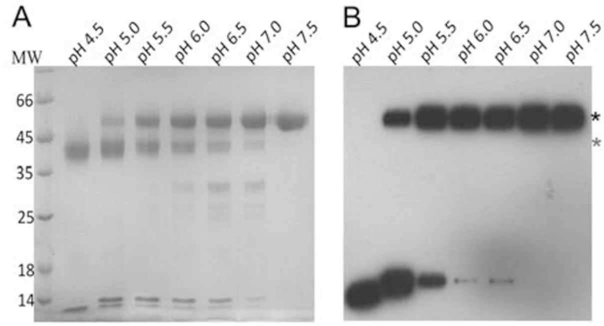

Figure 3 The proteolytic activation of AEP at a pH ≥ 4.5 is reversible. **(A)** SDS-PAGE analysis of partially activated AEP incubated at different pH values. $AEP_{pH4.5}$ was diluted in citric buffer at the indicated pH (5.0-7.5) and incubated for 16 h at 16 °C. **(B)** Western blot analysis of samples from **A** visualized using an anti-His antibody, which recognizes the C-terminal 6× His-tag of pro-AEP. Black and grey asterisks represent pro-AEP and mature AEP (the core domain), respectively.

Residues involved in the endopeptidase activity of AEP

We performed alanine-scanning mutagenesis of the conserved residues that were identified by a primary sequence alignment (Supplementary information, Figure S8). To ensure that these AEP mutants were correctly folded, 13-AEP mutants (D27A/N325A, S39A, N44A, R46A, H47A, V110A, D149A, H150A, E189A, C191A, S217A/S218A, D233A and D311A) and wild-type (WT) AEP, each prepared in a pH 7.5 buffer, were analyzed using circular dichroism (CD) spectroscopy. The CD analysis indicated that the mutants possessed similar secondary structures as WT AEP, except for H150A, which exhibited slight differences (Supplementary information, Figure S9). Additionally, the SEC profiles of all the mutants overlapped well with that of WT AEP.

Next, we evaluated the effect of the AEP mutations on its endopeptidase activity against the peptide substrate Z-AAN-NHMec. The mutants were activated at pH 3.5 overnight at 16 °C, and their relative activities were evaluated under the conditions described in the Materials and Methods. Substitution of the absolutely conserved H150 or C191 with alanine completely abolished the enzymatic activity (Figure 4A and Supplementary information, Table S3), suggesting a critical role of these amino acids in catalysis. Interestingly, substitution of N44, R46, E189, or D233 with alanine also resulted in an almost complete loss of activity. Mutants S39A, H47A, D149A and S217A/S218A exhibited a dramatic decrease in the activity of the enzyme (to less than 20% of the WT AEP activity). The V110A, P159A and C221A mutants exhibited only 21.9%, 38.8% and 49.9% of WT AEP activities, respectively. However, the S309A and E80A/E81A mutants retained approximately 80% of the WT AEP activity. The activities of the E192A, N213A and E401A mutants were barely altered in comparison with that of WT AEP (Figure 4A, 4B and Supplementary information, Table S3). The Y45A, E216A, Y222A/Y223A and S267A/H268A mutants did not exhibit any activity due to protein degradation in the 0.2 M citrate buffer at pH 3.5 or 4.0 during the activation process (Supplementary information, Figure S7G).

The residues D27, D305, D311 and N325 of AEP are known as autoproteolytic cleavage sites [20-22]. Substitution of D27 with alanine slightly decreased the enzymatic activity. This result is consistent with our structural study showing that D27 does not lie near the center of the active site (Figure 4C). It is also consistent with previous reports indicating that D27 is not essential for the AEP activity [22, 27]. In contrast, the activities of both the N325A and the D27A/N325A mutants decreased significantly (Figure 4A, 4C and Supplementary information, Table S3). Interestingly, D305A and D311A mutants exhibited 84.4% and 3.4% activity, respectively. The structure of $AEP_{pH7.5}$ indicates that D305 and D311 are located in the linker region between the core and cap domains. However, D311 sits at the turning point of a sharp turn pointing into the substrate-binding pocket and lies closer to the pocket than D305 (Figure 2C). Therefore, we speculate that D311 was more likely to be autoproteolytically cleaved to initiate AEP activity, and thus its mutation affects the activity of AEP to a greater extent than the mutation of D305.

Collectively, the structural analyses and mutagenesis studies suggest that N44, R46, H150, E189, C191, S217, S218 and D233 seem to form the active site and are essential for the cleavage of the peptide substrate. As for the autoproteolytic cleavage sites, the N-terminal cleavage site D27 is not required for the enzymatic activity. However, the C-terminal cleavage sites, D311 and N325, are essential for AEP activity.

Residues involved in the autoproteolytic maturation of AEP

To identify the residues involved in the autoproteolytic maturation of AEP, we performed an overnight incubation of pro-AEP mutants (prepared at pH 7.5) in buffers of pH 4.5 and 4.0. These samples were then examined by SDS-PAGE analyses (Supplementary information, Figure S7A-S7G). Activated, matured AEP exhibited as a band of lower molecular weight on the SDS-PAGE gel in comparison to its full-length proenzyme form. Intriguingly, the N44A, R46A, E189A and C191A mutants, which did not exhibit any activity against the peptide substrate Z-AAN-NHMec when exposed to pH 3.5 (Figure 4A and Supplementary information, Table S3), could still be autoproteolytically digested to their mature forms. However, the human AEP C189S mutant (C189 corresponds to C191 in mouse AEP) was reported to be unable to

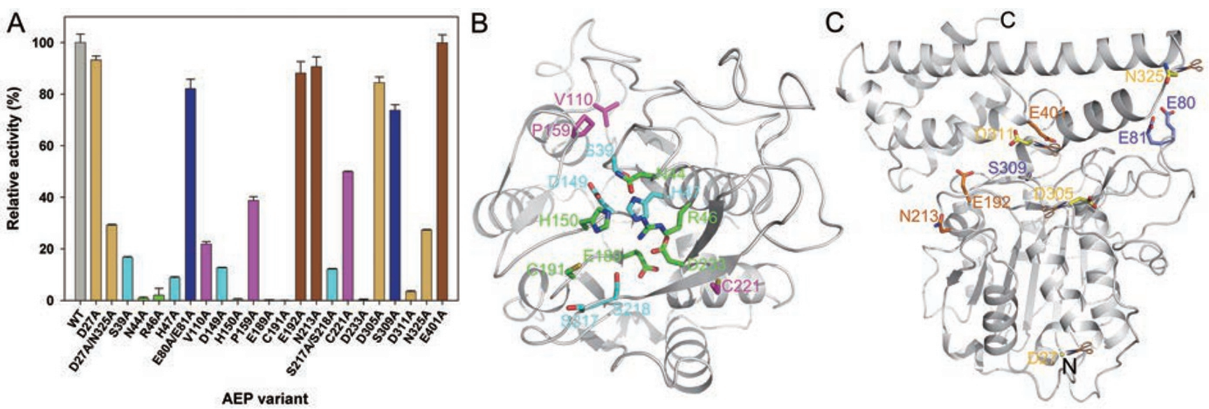

Figure 4 Mutagenesis studies of conserved AEP residues in the active site. **(A)** The relative activities of the indicated AEP variants were calculated relative to the activity of WT AEP (100%; grey). WT AEP and all the mutants were first activated at pH 3.5 overnight at 16 °C followed by the determination of the relative activities under the conditions described in the Materials and Methods. The mutants were divided into six groups (indicated by six different colors) according to their relative activities and were mapped onto the structures of the AEP core domain **(B)** and pro-AEP **(C)**.

undergo autoproteolytic maturation in an acidic buffer [21]. To further confirm the maturation ability of C191A, we repeated the maturation experiment at three different pH values (4.5, 4.0 and 3.5). C191A was not digested to its mature form when incubated at pH 4.5, but was digested to its mature form when incubated at pH 4.0 or 3.5 (Supplementary information, Figure S10A). This result suggests that C191 may not be involved in the autoproteolytic maturation, but is important for the cleavage of the AEP substrate. In contrast, H150A and D233A, which also showed little enzymatic activities, could not be autoproteolytically converted to their mature forms at pH values of 4.5 and 4.0 (Supplementary information, Figure S7C and S7E). Furthermore, the D233A mutant was not converted to its mature form even when incubated at pH 3.5 for 48 h (Supplementary information, Figure S10B). To further investigate the mechanism for the inability of the D233A mutant to undergo autoproteolysis, we determined the crystal structure of the pro-AEP D233A mutant at a resolution of 2.70 Å (Supplementary information, Figure S11). The structure of the mutant closely mirrored that of WT pro-AEP with a main-chain RMSD of 0.42 Å over 385 residues (Supplementary information, Figure S5D). In the structure of the WT pro-AEP, D233 lies in the vicinity of the catalytically critical residues. However, there is no noticeable difference between the structures of the D233A mutant and WT AEP. Perhaps D233 plays a critical role in the relay of the protons or charges that are essential for the autoproteolysis reaction.

Taken together, the above results suggest that although H150 and D233 are essential for autoproteolytic maturation, N44, R46, E189 and C191 are not required for the maturation of AEP. However, all of these residues are critical for the cleavage of the peptide substrate. Additionally, WT AEP and the N44A mutant were overexpressed in HEK293T cells, and the cell lysates were harvested for the endopeptidase activity assay using Z-AAN-NHMec as the substrate. The result confirmed that the activity of N44A mutant was dramatically decreased compared with that of WT AEP (Supplementary information, Figure S10C).

Residues involved in the re-ligation process of AEP

To identify the residues involved in the reverse process that converts partially activated AEP to the proenzyme form, we performed an overnight incubation of individual AEP mutants in buffers of pH 4.5 and 4.0. Then, the pH was adjusted to neutral pH by adding 1.0 M Tris-HCl, pH 8.1, and the samples were incubated overnight at 16 °C followed by SDS-PAGE analyses (Supplementary information, Figure S7A-S7G). The E80A/E81A, P159A, N213A, C221A and D305A mutants that were activated at pH 4.5 did not revert to the pro-AEP form when the pH was increased. The N213A mutant was one of the mutants that exhibited a level of enzymatic activity similar to that of the WT but failed to revert to the pro-AEP form. The endopeptidase activity of N213A was further confirmed by overexpression in HEK293T cells (Supplementary information, Figure S10C). Based on the pro-AEP structure, we speculated that the N213A mutation resulted in the breakage of a hydrogen bond between N213 and E216; this minor structural change may affect the re-ligation process of AEP without affecting its endopeptidase activity.

Interestingly, while WT AEP$_{pH4.0}$ was not reversed to the proenzyme form, the N44A, D27A/N325A, N325A and D311A mutants that were activated at pH 4.0 could return to their pro-AEP forms when the pH was neutralized (Supplementary information, Figure S7A-S7G). Among these residues, D27, N325 and D311 are autocleavage sites, and their mutations may somehow increase AEP resistance to proteolytic cleavage, thereby probably decreasing their maturation pH. Therefore, the corresponding mutant proteins retained the ability to be re-ligated even after activation at pH below 4.5. In addition to the mutants involving autocleavage sites, the N44A mutant also exhibited re-ligation activity when maturated at pH 4.0. To further delineate the re-ligation activity of the N44A mutant, it was initially activated at pH 3.0 or 4.5 followed by the re-ligation test in neutralized buffers. As expected, the N44A mutant activated at pH 4.5 reverted to pro-AEP when the pH was neutralized; however, it failed to return to the proenzyme form if it was activated at pH 3.0 (Supplementary information, Figure S12A). The H150A and D233A mutants, which are incapable of self-maturation, did not return to the pro-AEP form after they were maturated by WT AEP-mediated proteolysis, indicating that H150 and D233 are also important for the AEP re-ligation process (Supplementary information, Figure S12B).

A proposed model of inhibition of AEP endopeptidase activity by cystatin C

Cystatin C has been shown to inhibit mammalian AEP and other cysteine endopeptidases; however, the inhibition mechanism is not fully understood. To study the nature of the interactions between AEP and cystatin C, we co-purified and characterized the AEP/cystatin C complex. Human cystatin C (h-cystatin C) was expressed in *E. coli* and prepared as previously described [28, 29]. The h-cystatin C inhibits the endopeptidase activity of mouse AEP (m-AEP hereafter) against the peptide substrate Z-AAN-NHMec in a dose-dependent manner (Figure 5A). After verification that h-cystatin C inhibited m-AEP, we mixed pro-m-AEP with h-cystatin C at a 1:2 ratio and exposed the sample to buffers of pH 4.5 and 7.5. The two mixtures were incubated overnight at 4 °C prior to the SEC analysis. Unexpectedly, m-AEP and h-cystatin C were eluted as two separate peaks in the SEC profiles of both mixtures, indicating that m-AEP and h-cystatin C did not form complexes at either pH 4.5 or 7.5. Next, we mixed h-cystatin C and fully activated m-AEP (activated at pH 3.5) and incubated the mixture overnight at 16 °C for complex formation. Under these conditions, only m-AEP was eluted during SEC and h-cystatin C was not detected (Figure 5B). The disappearance of h-cystatin C prompted us to examine whether the m-AEP degraded h-cystatin C. SDS-PAGE analysis of the mixture of fully activated m-AEP with h-cystatin C clearly indicated that the degradation of h-cystatin C occurred in a dose-dependent manner (Figure 5C).

To determine whether the cleavage of h-cystatin C by m-AEP was due to a species discrepancy, we examined the intra- and interspecies interactions between AEP and cystatin C from human and mouse, including h-AEP vs h-cystatin C, h-AEP vs m-cystatin C, m-AEP vs m-cystatin C and m-AEP vs h-cystatin C. Interestingly, AEP-mediated degradation of cystatin C was observed in all the analyses (Figure 5D-5G). In summary, these data suggest that cystatin C functions as an AEP substrate and likely inhibits AEP activity through competition with other AEP substrates.

Pig AEP and h-cystatin C were reported to form a stable complex and were co-eluted during SEC [8]. As human and pig AEPs share high sequence identity (84.8%), we speculated that pig AEP may also degrade h-cystatin C. Indeed, as shown in Figure 5H, h-cystatin C was digested by pig AEP with a nearly identical efficiency as that of the human AEP. Cystatin C was also reported to inhibit a variety of other cysteine proteases, such as cathepsins B, H, K, L and S [30, 31]. To investigate whether cystatin C inhibits cathepsins via a similar mechanism, mouse cathepsins B, S and L were activated and their degradative activities towards cystatin C were determined. The activities of cathepsins B and S were confirmed using the fluorogenic peptide substrates Z-Leu-Arg-AMC and Mca-RPKPVENval-WRK(Dnp)-NH2, respectively (Supplementary information, Figure S13A and S13B). Therefore, the active cathepsins were used to digest h-cystatin C. Human cystatin C was not digested by cathepsin B, S, or L, but was digested by a same amount of human AEP (Supplementary information, Figure S13C-S13E).

To further confirm the specific digestion of h-cystatin C by m-AEP and to investigate the cleavage sites on h-cystatin C, N-terminal amino acid sequencing was performed to analyze the cleaved bands of h-cystatin C obtained from SDS-PAGE gel (Figure 5D). Five unique N-terminal sequences were obtained: four sequences contained AEP cleavage sites (three Asp residues and one Asn residue) and one represented the N-terminal five residues (MDIGI) of the recombinant cystatin C (Figure 5I). The identified cleavage sites on cystatin C were mapped onto the cystatin C structure (PDB ID: 1TIJ) (Figure 5I). Intriguingly, although cystatin C could inhibit the endopeptidase activity of AEP and was digested by AEP, the presence of cystatin C did not affect the religation activity of AEP (Supplementary information, Figure S12C). Thus, cystatin C may regulate the activity of AEP through competition with other substrates for the AEP active site.

Discussion

Pro-AEP is known to undergo auto-proteolytic maturation in acidic environments, and $AEP_{pH7.5}$ was fully activated when incubated at pH 3.5 (Figure 1C). Activation coincided with a shift of the protein bands observed in SDS-PAGE analysis (Figure 1D). In this study, we demonstrated that the auto-proteolytic maturation was reversible when pro-m-AEP was partially activated at a pH ≥ 4.5. This partially activated AEP was reversed back into its inactive, proenzyme form when incubated at pH 7.5 (Figure 3). The crystal structure of mouse $AEP_{pH4.5}$ (crystallized at pH 8.5) revealed a structure of the pro-AEP, and thereby independently confirmed the religation of the cleaved peptide bonds in the partially activated AEP (Figure 2D). Extensive interactions between the cap and the core domains may aid in securing and repositioning the cap during the religation process. WT AEP that was exposed to a pH below 4.0 was not reversed to its proenzyme form, which may be due to the disassociation of the cap domain from the core domain with a subsequent degradation of the cap domain at a lower pH.

Concerning the mechanism for the reversible activation and maturation of AEP, our results suggest that an intramolecular complex is formed between the partially cleaved core and cap domains at pH 4.5. The peptide bond was cleaved through a cross-proteolytic process in which the key catalytic residues H150 and D233, but not C191, were involved. According to the principle of microscopic reversibility, we propose that the peptide bond reformation may also be mediated by the identical group of catalytic residues in the intermolecular complex. The observations of the cleavage at pH 4.5 and the religation at a pH above 7.0 suggest an extraordinary protonation effect on the relative free energy between the partially activated, cleaved complex and pro-AEP. In other words, we hypothesized that protonation of the N-

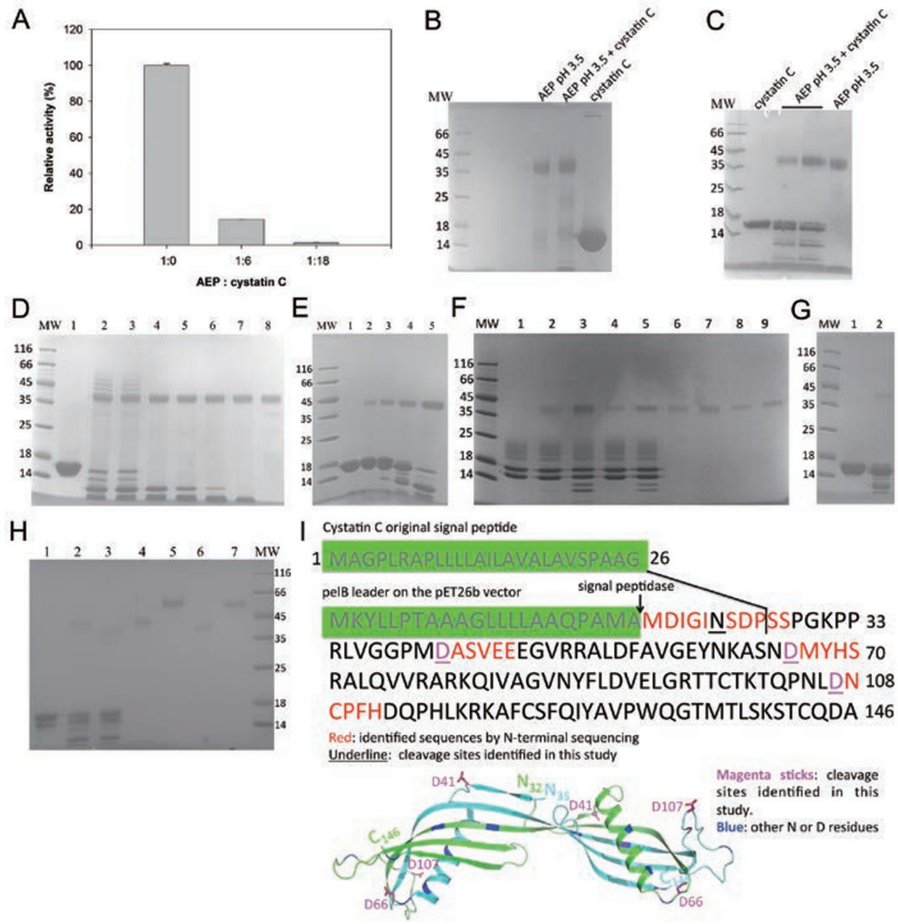

Figure 5 A proposed model for the inhibition of AEP activity by cystatin C. **(A)** Human cystatin C inhibited the activity of mouse $AEP_{pH3.5}$ in a dose-dependent manner. The enzymatic activity of AEP was measured by using the methods described in Figure 4A. **(B)** SEC combined with SDS-PAGE analysis for testing the complex formation between mature m-AEP and human cystatin C. Mature AEP and cystatin C were mixed at a 1:2 molar ratio and were incubated overnight at 16 °C for complex formation. The sample was loaded onto a Superdex 200 column. Only AEP was eluted from the SEC, whereas cystatin C was not detected. The peak fraction was analyzed by SDS-PAGE analysis. **(C)** Human cystatin C was gradually degraded by mature m-AEP. Mature AEP and human cystatin C (at a molar ratio of approximately 1:20 and 1:7) were incubated overnight at 16 °C and analyzed using SDS-PAGE analysis. **(D)** Human cystatin C was degraded by mature human AEP in a time-dependent manner at 25 °C. Lane 1: human cystatin C, 5.0 µg; loading control. Lanes 2-7: human cystatin C was degraded by mature human AEP after an incubation of 1 min, 30 min, 1 h, 2 h, 4 h and 8 h, respectively. Lane 8: mature human AEP, 2.5 µg; loading control. The unique bands of human cystatin C were further analyzed using N-terminal amino acid sequencing. **(E)** Human cystatin C was degraded by mature human AEP in a dose-dependent manner. All the incubations were performed for 5 min at 25 °C. Lane 1: human cystatin C, 5.0 µg; loading control. Lanes 2-5: human cystatin C was degraded by gradually increased amounts of mature human AEP (0.25, 0.5, 1.0 and 2.5 µg). **(F)** Mouse cystatin C, which was expressed in HEK293 cells, was cleaved by mouse and human AEPs. All the incubations were performed for 5 min at 25 °C. Lane 1: mouse cystatin C, 2.0 µg; loading control. Lanes 2 and 3: mouse cystatin C was degraded by mature mouse AEPs (0.5 and 1.0 µg, respectively). Lanes 4 and 5: mouse cystatin C was degraded by mature human AEPs (0.5 and 1.0 µg, respectively). Lanes 6 and 7: loading controls of mature mouse AEP for lanes 2 and 3. Lanes 8 and 9: loading controls of mature human AEP for lanes 4 and 5. **(G)** Human cystatin C was degraded by mature mouse AEP. Lane 1: human cystatin C, 5.0 µg; loading control. Lane 2: human cystatin C was degraded by 0.25 µg of mature mouse AEP after an incubation of 1 min at 25 °C. **(H)** Human cystatin C was degraded by activated pig AEP. Lane 1: human cystatin C, 3.3 µg; loading control. Lane 2: human cystatin C was degraded by 0.33 µg of activated pig AEP after an incubation of 5 min at 16 °C. Lane 3: human cystatin C was degraded by 0.33 µg of mature human AEP after an incubation of 5 min at 16 °C. Lanes 4 and 5: pig mature AEP and pro-AEP. Lanes 6 and 7: human mature AEP and pro-AEP. **(I)** The N-terminal amino acid sequencing results of the cleaved bands of human cystatin C obtained from **D**. Human endogenous and recombinant cystatin C amino acid sequences are shown. The cleaved bands of the recombinant human cystain C were analyzed using N-terminal amino acid sequencing. The pelB leader sequence can be cleaved by signal peptidase in *E. coli*. In the five unique sequences (red), three Asp (underlined, magenta) residues and one Asn residue (underlined, black) represent AEP cleavage sites. The three Asp residues are indicated in the structure of human cystatin C (PDB ID: 1TIJ) as magenta sticks. The N-terminal five residues (MDIGI) of recombinant cystatin C were also identified by N-terminal amino acid sequencing.

terminal amino group of the cleaved N-fragment was the driving force for the cleavage event at pH 4.5. At basic pH values, this group should be neutralized and become a strong nucleophile, shifting the equilibrium back to the peptide state. Thus, this protonation-dependent free energy shifting model can explain the reversible activation and religation of AEP. Further elucidation of the physical details of this model requires the aid of computational energy analysis.

The structural basis for AEP activation at lower pH values was previously not fully understood. In this study, we demonstrated that pro-AEP was bimodular (Figure 2): at a higher pH, the cap domain covers the core domain and sterically obstructs access to the active site, whereas at a lower pH, the cap domain is removed through auto-proteolysis. Removal of the cap domain enables substrate access to the catalytic dyad of H150 and C191 for the subsequent catalytic cleavage of the peptide substrate. Our crystallographic snapshots of the proenzyme and mature forms of mouse AEP clearly support this mechanism.

Previously, human cystatin C has been shown to inhibit both the C1 and the C13 peptidase families. Specifically, human cystatin C inhibits pig AEP with K_i values < 5 nM [8]. Human cystatin C and pig AEP were reported to form a tight complex and to be co-eluted in a single peak during SEC. Furthermore, there was no apparent degradation of cystatin C in the presence of AEP and an AEP substrate [8]. However, our initial studies on mouse AEP and its inhibition by human cystatin C indicated that AEP degraded cystatin C instead of forming a stable complex with cystatin C (Figure 5). Furthermore, AEP-mediated cystatin C cleavage was confirmed in both inter- and intraspecies assays using human and mouse proteins. Thus, cystatin C functions as an AEP substrate and likely inhibits AEP through competition with other AEP substrates. AEP from pig shares 85% sequence identity with its homolog from mouse. Our speculation that pig AEP cleaves human cystatin C was also experimentally confirmed (Figure 5H). Interestingly, previous studies indicated that mouse cathepsins could be inhibited by human cystatin C [30, 31]. However, human cystatin C was not cleaved by mouse cathepsins, indicating that cystatin C may employ a different mechanism to inhibit cathepsins.

While this manuscript was under preparation, Dall *et al*. [32] reported the crystal structures of human pro-AEP and the active AEP in complex with the substrate analogs Z-Ala-Ala-AzaAsn-chloromethyl ketone (Z-AAN-CMK) and Ac-Tyr-Val-Ala-chloromethyl ketone (Ac-YVAD-CMK, an irreversible caspase-1 inhibitor). In their study, the pH-dependent activation was proposed, and an unexpected hidden asparagine (N)-specific carboxypeptidase activity was unveiled. In this study, we demonstrated similar mouse pro-AEP and mature AEP structures (with a main chain RMSD of 1.04 Å between pro-AEPs and 0.49 Å between active AEPs). Additionally, our study demonstrates the following new findings: 1) partially activated AEP can revert to pro-AEP as the pH is shifted to neutral values if the pH of the initial activation buffer was ≥ 4.5, whereas this reversion does not occur if the pH of the initial activation buffer was ≤ 4.0; 2) cystatin C is a substrate of AEP and probably inhibits AEP through competition with other AEP substrates; and 3) two additional maturation cleavage sites (N332 and D428) were identified that have not been previously reported.

In conclusion, our structure-function studies on AEP conceptually unveiled an interesting, previously unknown intermediate stage at approximately pH 4.5 during auto-proteolytic maturation of AEP, in which the partially activated AEP can either revert to its proenzyme form if the pH is shifted from acidic (≥ 4.5) to neutral (~7.5) or proceed further into its irreversible, fully activated form if the pH is further decreased towards 3.5. Hypothetically, such an intermediate stage may function as a buffer zone for the maintenance of an appropriate AEP enzymatic activity. At the intermediate stage, the partially matured AEP may be further activated rapidly if more endopeptidase activity is required, or it may be recycled back to the proenzyme form if a sufficient endopeptidase activity has been achieved. This feature allows the tunable endopeptidase activity of AEP, which may be important for the precise control of the immune system. The physiological significance of this reversible maturation remains to be further investigated.

Materials and Methods

Cloning, protein expression and purification

Full-length mouse and human AEP genes were synthesized by Sangon Biotech Co., Ltd. (Shanghai). Mouse and human AEPs fused with 6× His tags at C-termini were subcloned into the frame of the pFastbac1 vector (Invitrogen) using the *Eco*RI and *Xho*I restriction sites. The resultant clone was confirmed using DNA sequencing. Plasmids containing the genes of interest were then transformed into DH10Bac competent cells (Invitrogen), and the recombinant bacmid DNA was isolated and verified according to the Bac-to-Bac Baculovirus Expression System instructions (Invitrogen). Sf9 insect cells were then transfected to generate recombinant baculovirus, and the titer of the baculoviral stock was amplified by infecting the Sf9 insect cells with P1 and P2 viral stocks. AEP was expressed as a secreted protein by infecting the Sf9 insect cells with recombinant baculovirus at an optimal MOI. The culture medium containing the secreted recombinant protein was harvested 72 h post infection.

For purification, the culture medium from the insect cells that contained the target protein was first centrifuged at 12 000 rpm at 4 °C for 30 min to remove the cells and cell debris, and was then loaded onto HiTrap Ni Fast Flow beads (GE Healthcare) that were previously equilibrated with 50 mM PBS (50 mM Na_2HPO_4, 10 mM KH_2PO_4, 137 mM NaCl and 2.7 mM KCl, pH 7.4) at 4 °C. The beads were sequentially washed using 50 mM PBS containing 20 mM and 50 mM imidazole and were then eluted with 50 mM PBS supplemented with 300 mM imidazole. The protein was further purified on a SEC column that was equilibrated with 20 mM HEPES, pH 7.5 and 150 mM NaCl. The pooled peak fractions ($AEP_{pH7.5}$) were concentrated to 13 mg/ml, and aliquots were flash-frozen in liquid nitrogen and stored at −80 °C until further use. Mouse and pig AEPs that were expressed in HEK293 cells to confirm functional AEP expression in insect cells were purchased from Sino Biological Inc. (Catalog #: 50051-M07H and 62001-W08H, respectively).

Mutations were designed based on mouse AEP constructs, and mutagenesis was performed using a QuikChange site-directed mutagenesis kit following the manufacturer's instructions (Stratagene). All the recombinant plasmids were sequenced to verify the sequences. AEP mutants were expressed and purified in the same manner as that for WT AEP.

Production of cystatin C and GST-TEV-Ub

For human cystatin C expression in *E. coli*, the human *cystatin C* gene without a signal peptide, C-terminal thrombin cleavage sites, and a 6× His tag cassette was subcloned into the pET26b vector using *Bam*HI and *Xho*I restriction sites. The resultant plasmid was confirmed using DNA sequencing and then transformed into *E. coli* BL21(DE3) competent cells for expression and purification according to previously described methods [28, 29]. Mouse cystatin C that was expressed in HEK293 cells was purchased from Sino Biological Inc. (Catalog #: 10439-H08H).

A GST-TEV-Ub expression plasmid was constructed by inserting the ubiquitin (Ub) gene into the pMCSG10 vector, which contained a GST tag and a TEV protease cleavage site [33].

Crystallization

The full-length mouse AEP produced by insect cells and treated with PNGase F was screened for crystallization conditions immediately after purification, as previously described [34]. Crystallization screening was performed using the sitting-drop-vapor diffusion method and commercial-screening kits from Hampton Research and Emerald BioSystems. A volume of 0.4 µl of the protein stock solution was mixed with 0.4 µl of reservoir solution using a Mosquito robot (TTP Labtech) and equilibrated against 35 µl of reservoir at 16 °C. Diffraction-quality crystals of pro-$AEP_{pH7.5}$ were obtained in a condition containing 0.2 M succinic acid, pH 7.0 and 20% PEG 3350. Crystals of native and Hg-derivatized $AEP_{pH4.5}$ were obtained in 0.2 M trimethylamine N-oxide dehydrate, 0.1 M Tris-HCl, pH 8.5 and 20% PEG 2000. Crystals of active $AEP_{pH3.5}$ were obtained in 0.01 M zinc sulfate, 0.1 M MES, pH 6.7 and 23% PEG 500. Diffraction-quality crystals of pro-AEP $D233A_{pH7.5}$ were obtained in 0.1 M ammonium acetate, 0.1 M Bis-Tris-HCl, pH 5.5 and 17% PEG 10000.

Data collection, phasing, structure solution and refinement

Crystals were harvested and flash-frozen in liquid nitrogen. The crystals were derivatized by adding a small grain of mercury potassium iodide to the crystallization drop, as previously reported [35], followed by incubation for 4 h. Diffraction data for crystals obtained at pH 7.5 were collected at beamline BL17A at the Photon Factory (KEK, Japan). All other diffraction data were collected at 100 K using an ADSC Q315 CCD detector at beamline BL17U1 at the Shanghai Synchrotron Radiation Facility (SSRF). All datasets were indexed, integrated and scaled using the HKL2000 software package [36]. The initial phases of $AEP_{pH4.5}$ were determined using the X^2DF structure determination pipeline [37, 38] and the Hg-SAD (single-wavelength anomalous dispersion) method [39], and the initial model was built using Phenix Autobuild [40]. The remaining structures were solved by molecular replacement [41] using the determined $AEP_{pH4.5}$ structure as the search model. The models were manually improved in Coot [42]. Refinement was performed by alternately using REFMAC [43] and Phenix Refine [40]. The statistics for the data collection and refinement are summarized in Table 1.

Autoactivation of AEP and enzymatic activity assay

Autoactivation of AEP was performed by diluting aliquots of the AEP proenzyme into a buffer containing 0.2 M sodium citrate/citric acid and 1 mM dithiothreitol that was adjusted to different pH values. After an overnight incubation at 16 °C, the enzymatic activity of AEP was assessed by transferring aliquots of the autoactivation reaction to the AEP assay buffer (40 mM citric acid, 121 mM Na_2HPO_4, pH 5.8, containing 1 mM dithiothreitol, 1 mM EDTA and 0.1% CHAPS), which included 20 µM Z-Ala-Ala-Asn-NHMec (Z-AAN-NHMec) (Bachem) as the substrate [21]. The rate of formation of the product was monitored in a fluorometer at 25 °C. The excitation and emission wavelengths were 360 and 460 nm, respectively.

For the enzymatic activity of AEP in the HEK293T cell lines, various AEP constructs were transfected into the cells. Autoactivation of AEP was achieved by equilibrating the cell lysates in buffers at pH 4.0. The enzymatic activity assays were performed as previously described.

Purified protein substrates of AEP, such as cystatin C and GST-TEV-Ub, were mixed with activated AEP in a reaction buffer (20 mM HEPES, pH 7.5 and 150 mM NaCl) at 16 °C or 25 °C for a suitable length of time.

For the conversion of partially activated AEP back to the proenzyme form, the pH of the samples was adjusted back to a neutral pH by adding moderate amounts of 1.0 M Tris-HCl, pH 8.1, and the samples were incubated overnight at 16 °C.

The mouse cathepsins B, S and L were purchased from Sino Biological Inc (Catalog #: 50084-M08H, 50769-M08H, and 50015-M08H, respectively) and were activated according to the manufacturer's assay procedures. The bioactivity of cathepsins B and S were measured by their abilities to cleave the fluorogenic peptide substrates Z-LR-AMC (R&D Systems, Catalog #: ES008) and Mca-RPKPVENval-WRK(Dnp)-NH2 (R&D Systems, Catalog #: ES002), respectively. Human cystatin C and the activated cathepsins were mixed in the assay buffer for 1 min, 5 min, 30 min, 1 h and 4 h at 16 °C or 25 °C and analyzed using SDS-PAGE analysis.

Small-angle X-ray scattering (SAXS)

SAXS data for $AEP_{pH4.0}$ (20 mM citric acid, pH 4.0, and 150 mM NaCl) and $AEP_{pH7.5}$ (20 mM HEPES, pH 7.5, and 150 mM

NaCl) were collected on the SIBYLS beamline at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, and the data were treated as previously described [44, 45]. Briefly, each sample was measured at three exposures (0.5, 1.0, and 6.0 sec) and three concentrations (2.5, 5.0, and 10.0 mg/ml) at 10 °C to exclude aggregation and radiation-damage effects. Both the maximum dimension (Dmax) and the MW derived from the SAXS P(r) function for $AEP_{pH7.5}$ were larger than those of $AEP_{pH4.0}$ (Dmax: 130 ± 5 Å compared with 70 ± 5 Å; MW: 97.5 ± 2.4 kDa compared with 33.1 ± 0.7 kDa, respectively). Thus, AEP exists as a homodimer in solution at pH 7.5 but as a monomer at pH 4.0. CRYSOL [46] was then used to confirm the most reasonable conformation. Ten individual GASBOR [47] calculations with P2 symmetry as a restraint were performed to construct shapes for $AEP_{pH7.5}$. The resultant solution model, within which the dimer model fits well, had an acceptable Chi value of 1.49 ± 0.07. For $AEP_{pH4.0}$, no symmetry constraint was used, and the final Chi value was 0.90 ± 0.09.

N-terminal amino acid sequencing

N-terminal amino acid sequencing was performed as previously described [48]. Briefly, the protein was first fractionated using SDS-PAGE and then electrophoretically transferred to a polyvinylidene difluoride (PVDF) membrane for 1 h (300 mA) in an ice bath using a Bio-Rad apparatus. The protein band was visualized by staining with Coomassie Brilliant Blue (0.1% Coomassie Brilliant Blue R-250 in 1.0% acetic acid and 40% methanol) and excised to air dry. The PVDF membrane containing the target band was digested and extracted. Finally, the protein was sequenced on an automated protein sequencer (ABI Procise 491) to identify the first four or five amino acids in the N-terminus.

Circular dichroism (CD) spectroscopy

CD spectra were acquired on a Chirascan CD Spectrometer (Applied Photophysics). Freshly prepared WT and mutant AEP proteins were adjusted to 0.2 mg/ml in 40 mM Tris-HCl, pH 7.5, prior to the measurements. Wavelength spectra were recorded at 20 °C using a 0.1-cm path length cuvette. Each scan was obtained by recording every 1 nm with a bandwidth of 1 nm between the wavelength ranges of 200-260 nm.

Western blot analysis

Western blot analysis was preformed according to the standard protocol and detection was performed using an HRP-conjugated mouse anti-His antibody (Sigma, USA) [29]. Immunoreactive bands were visualized using Enhanced Chemiluminescence (ECL) reagents (Santa Cruz, USA) according to the manufacturer's instructions.

Synthesis of the AEP inhibitor AAN-CMK

The AEP inhibitor AAN-CMK was synthesized as previously described [19].

Statistical analysis

The significance of differences between groups exhibiting similar variance was evaluated using Student's *t*-test.

PDB deposition

The coordinates and structure factors of $AEP_{pH4.5}$, $AEP_{pH7.5}$, $AEP_{pH3.5}$ and $AEP_{pH7.5}$D233A have been deposited in PDB under the accession codes 4NOM, 4NOK, 4NOJ and 4NOL, respectively.

Acknowledgments

The authors thank the staff at the synchrotron beamlines (17U of the SSRF, 17A of KEK, and SIBYLS of the ALS) for their assistance with the X-ray diffraction and solution X-ray scattering data collection. This work was supported by the Ministry of Science and Technology of China (2014CB910400, 2013CB911103 and 2011CB911103), the Ministry of Health of China (2013ZX10004-602) and the National Natural Science Foundation of China (31330019, 31200559, 91313301 and 31300613).

References

1 Chen JM, Dando PM, Rawlings ND, *et al.* Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J Biol Chem* 1997; **272**:8090-8098.
2 Chen JM, Dando PM, Stevens RA, Fortunato M, Barrett AJ. Cloning and expression of mouse legumain, a lysosomal endopeptidase. *Biochem J* 1998; **335**:111-117.
3 Chen JM, Rawlings ND, Stevens RA, Barrett AJ. Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS Lett* 1998; **441**:361-365.
4 Chan CB, Abe M, Hashimoto N, *et al.* Mice lacking asparaginyl endopeptidase develop disorders resembling hemophagocytic syndrome. *Proc Natl Acad Sci USA* 2009; **106**:468-473.
5 Morita Y, Araki H, Sugimoto T, *et al.* Legumain/asparaginyl endopeptidase controls extracellular matrix remodeling through the degradation of fibronectin in mouse renal proximal tubular cells. *FEBS Lett* 2007; **581**:1417-1424.
6 Chen JM, Fortunato M, Stevens RA, Barrett AJ. Activation of progelatinase A by mammalian legumain, a recently discovered cysteine proteinase. *Biol Chem* 2001; **382**:777-783.
7 Sarandeses CS, Covelo G, Diaz-Jullien C, Freire M. Prothymosin alpha is processed to thymosin alpha 1 and thymosin alpha 11 by a lysosomal asparaginyl endopeptidase. *J Biol Chem* 2003; **278**:13286-13293.
8 Alvarez-Fernandez M, Barrett AJ, Gerhartz B, *et al.* Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J Biol Chem* 1999; **274**:19195-19203.
9 Hsing LC, Rudensky AY. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol Rev* 2005; **207**:229-241.
10 Manoury B, Hewitt EW, Morrice N, *et al.* An asparaginyl endopeptidase processes a microbial antigen for class II MHC presentation. *Nature* 1998; **396**:695-699.
11 Antoniou AN, Blackwood SL, Mazzeo D, Watts C. Control of antigen presentation by a single protease cleavage site. *Immunity* 2000; **12**:391-398.
12 Ewald SE, Lee BL, Lau L, *et al.* The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. *Nature* 2008; **456**:658-662.
13 Bauer S. Toll-like receptor 9 processing: the key event in Toll-like receptor 9 activation? *Immunol Lett* 2013; **149**:85-87.
14 Ewald SE, Engel A, Lee J, *et al.* Nucleic acid recognition by Toll-like receptors is coupled to stepwise processing by

cathepsins and asparagine endopeptidase. *J Exp Med* 2011; **208**:643-651.
15 Park B, Brinkmann MM, Spooner E, *et al.* Proteolytic cleavage in an endolysosomal compartment is required for activation of Toll-like receptor 9. *Nat Immunol* 2008; **9**:1407-1414.
16 Sepulveda FE, Maschalidi S, Colisson R, *et al.* Critical role for asparagine endopeptidase in endocytic Toll-like receptor signaling in dendritic cells. *Immunity* 2009; **31**:737-748.
17 Saska I, Gillon AD, Hatsugai N, *et al.* An asparaginyl endopeptidase mediates *in vivo* protein backbone cyclization. *J Biol Chem* 2007; **282**:29721-29728.
18 Mylne JS, Chan LY, Chanson AH, *et al.* Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. *Plant Cell* 2012; **24**:2765-2778.
19 Niestroj AJ, Feussner K, Heiser U, *et al.* Inhibition of mammalian legumain by Michael acceptors and AzaAsn-halomethylketones. *Biol Chem* 2002; **383**:1205-1214.
20 Chen JM, Fortunato M, Barrett AJ. Activation of human prolegumain by cleavage at a C-terminal asparagine residue. *Biochem J* 2000; **352**:327-334.
21 Li DN, Matthews SP, Antoniou AN, Mazzeo D, Watts C. Multistep autoactivation of asparaginyl endopeptidase *in vitro* and *in vivo*. *J Biol Chem* 2003; **278**:38980-38990.
22 Dall E, Brandstetter H. Activation of legumain involves proteolytic and conformational events, resulting in a context- and substrate-dependent activity profile. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2012; **68**:24-31.
23 Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 2010; **38**:W545-W549.
24 Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 2005; **33**:W89-W93.
25 Walters J, Schipper JL, Swartz P, Mattos C, Clark AC. Allosteric modulation of caspase 3 through mutagenesis. *Biosci Rep* 2012; **32**:401-411.
26 Wilson KP, Black JA, Thomson JA, *et al.* Structure and mechanism of interleukin-1 beta converting enzyme. *Nature* 1994; **370**:270-275.
27 Kuroyanagi M, Nishimura M, Hara-Nishimura I. Activation of *Arabidopsis* vacuolar processing enzyme by self-catalytic removal of an auto-inhibitory domain of the C-terminal propeptide. *Plant Cell Physiol* 2002; **43**:143-151.
28 Dalboge H, Jensen EB, Tottrup H, *et al.* High-level expression of active human cystatin C in *Escherichia coli*. *Gene* 1989; **79**:325-332.
29 Ouyang S, Gong B, Li JZ, *et al.* Structural insights into a human anti-IFN antibody exerting therapeutic potential for systemic lupus erythematosus. *J Mol Med (Berl)* 2012; **90**:837-846.
30 Hall A, Ekiel I, Mason RW, *et al.* Structural basis for different inhibitory specificities of human cystatins C and D. *Biochemistry* 1998; **37**:4071-4079.
31 Lutgens SP, Cleutjens KB, Daemen MJ, Heeneman S. Cathepsin cysteine proteases in cardiovascular disease. *Faseb J* 2007; **21**:3029-3041.
32 Dall E, Brandstetter H. Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc Natl Acad Sci USA* 2013; **110**:10940-10945.
33 Stols L, Zhou M, Eschenfeldt WH, *et al.* New vectors for co-expression of proteins: structure of Bacillus subtilis ScoAB obtained by high-throughput protocols. *Protein Expr Purif* 2007; **53**:396-403.
34 Liang W, Ouyang S, Shaw N, *et al.* Conversion of D-ribulose 5-phosphate to D-xylulose 5-phosphate: new insights from structural and biochemical studies on human RPE. *Faseb J* 2011; **25**:497-504.
35 Mori T, Nishizawa S, Hirohashi Y, *et al.* Efficiency of G2/M-related tumor-associated antigen-targeting cancer immunotherapy depends on antigen expression in the cancer stem-like population. *Exp Mol Pathol* 2012; **92**:27-32.
36 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* 1997; **276**:307-326.
37 Liu ZJ, Lin D, Tempel W, *et al.* Parameter-space screening: a powerful tool for high-throughput crystal structure determination. *Acta Crystallogr D Biol Crystallogr* 2005; **61**:520-527.
38 Ru H, Zhao L, Ding W, *et al.* S-SAD phasing study of death receptor 6 and its solution conformation revealed by SAXS. *Acta Crystallogr D Biol Crystallogr* 2012; **68**:521-530.
39 Hendrickson WA. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 1991; **254**:51-58.
40 Adams PD, Afonine PV, Bunkoczi G, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.
41 McCoy AJ, Grosse-Kunstleve RW, Adams PD, *et al.* Phaser crystallographic software. *J Appl Crystallogr* 2007; **40**:658-674.
42 Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:486-501.
43 Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997; **53**:240-255.
44 Niu F, Shaw N, Wang YE, *et al.* Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation. *Proc Natl Acad Sci USA* 2013; **110**:9054-9059.
45 Jiao L, Ouyang S, Liang M, *et al.* Structure of severe fever with thrombocytopenia syndrome virus nucleocapsid protein in complex with suramin reveals therapeutic potential. *J Virol* 2013; **87**:6829-6839.
46 Schmidt B, Konig S, Svergun D, *et al.* Small-angle X-ray solution scattering study on the dimerization of the FKBP25mem from Legionella pneumophila. *FEBS Lett* 1995; **372**:169-172.
47 Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 2001; **80**:2946-2953.
48 Niu F, Ru H, Ding W, Ouyang S, Liu ZJ. Structural biology study of human TNF receptor associated factor 4 TRAF domain. *Protein Cell* 2013; **4**:687-694.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)