Supporting Information

Dall and Brandstetter 10.1073/pnas.1300686110

SI Methods

Cloning, Expression, and Purification. Human legumain was cloned, expressed, and purified as described (1). The point mutations Glu190Lys (E190K), Asn263Gln (N263Q), and Asn272Gln (N272Q) were introduced by following a protocol based on the inverse-PCR method. Fully glycosylated prolegumain was produced using the *Leishmania tarentolae* expression system (LEXSY, Jena Bioscience) (2).

Enzymatic Activity Assays. For subsequent activity assays, 0.5 mg/mL wild-type and mutant prolegumain were activated via shifting the pH to 4.0 in a buffer containing 100 mM citric acid, 100 mM NaCl, and 5 mM DTT at 37 °C for 20 h. Progress of autocatalytic activation was monitored via SDS/PAGE and an enzymatic activity assay. Activity toward the legumain specific substrates Benzoyl-L-Asparaginyl-para-NHPhNO$_2$ (Bz-Asn-pNA; Bachem) and Z-Ala-Ala-Asn-AMC (AAN-AMC; Bachem) was assayed at 0.4 mM and 40 μM concentration, respectively. Assays were carried out at 37 °C in legumain assay buffer (100 mM citric acid at pH 5.5, 100 mM NaCl, and 2 mM DTT) at a protein concentration of 0.05–0.5 μM. Substrate turnover was measured by using an Infinite M200 Plate Reader (Tecan).

The recombinantly produced extracellular segment of human integrin $\alpha_v\beta_3$ was purchased from R&D Systems. Five hundred nanomolar integrin $\alpha_v\beta_3$ in 1× PBS was adjusted to pH 5.5 and incubated with 250 nM activated legumain at pH 4.0. After 10 min of incubation, the reaction mix was diluted 1:10 in legumain assay buffer at pH 6.0 (without DTT), and turnover of Bz-Asn-pNA was measured at 37 °C. A competition assay was carried out to see whether prolegumain binds to integrin $\alpha_v\beta_3$. Five hundred nanomolar integrin $\alpha_v\beta_3$ were preincubated with 5 μM prolegumain at pH 7.4. After 10 min, pH was adjusted to 5.5 and 250 nM activated legumain was added. After 10 min of incubation, Bz-Asn-pNA turnover was measured at pH 6.0, as described above.

To study the effect of trypsin activation, prolegumain (1 mg/mL) was incubated with trypsin (trypsin from bovine pancreas; Sigma-Aldrich) at a 1:50 molar ratio at pH 6.0 for 2 h at 20 °C. Progress of trypsin cleavage was monitored on SDS/PAGE. Enzymatic activity of trypsin-activated and pH 4.0-activated legumain was measured as turnover of AAN-AMC in legumain assay buffer at pH 5.5 (citric acid) and 6.5 (Mes).

To assay the carboxypeptidase activity of legumain, β-legumain was generated by incubation of wild-type prolegumain at pH 4.5 for 20 h in a buffer containing 100 mM citric acid, 100 mM NaCl, and 5 mM DTT. β-legumain is autocatalytically processed after Asn323 as judged by SDS/PAGE. Activated legumain was subjected to size exclusion chromatography by using an ÄKTA FPLC system equipped with a Superdex 200 10/300 GL column (GE Healthcare). Fractions were collected and analyzed by SDS/PAGE. The synthetic peptides H-AlaAlaAsn-COOH, H-AlaAlaAsn-Ala-COOH, H-AlaAlaAsn-AlaAla-COOH, H-AlaAlaAsn-AlaAlaAla-COOH, and H-AlaAlaAsn-Ala-CONH2 were synthesized by JPT (Peptide Technologies). Turnover of 10 μM AAN-AMC by pH 4.0- and pH 4.5-activated legumain in legumain assay buffer was measured in the presence of 2 mM of the respective peptide or DMSO in the control experiment at ~0.05 μM enzyme concentration and 37 °C.

Thermofluor Assay. A thermofluor assay was performed as described in more detail (3). Briefly, 1–2 mg/mL protein sample containing 50× SYPRO Orange (Invitrogen) was added in a 1:10 ratio to 22.5 μL of assay buffer composed of 60 mM Mes at pH 6.5

and 100 mM NaCl. Thermal denaturation curves were collected in a 7500 Real-Time PCR System (Applied Biosystems) from 20 to 95 °C. The fluorescence data were analyzed as described (4). Protein samples investigated were prolegumain, pH 4.0-activated legumain (asparagine-specific endopeptidase; AEP) completely lacking the C-terminal prodomain and β-legumain. β-legumain was generated from prolegumain by incubation at pH 4.5 as described before. The C-terminal portion (Asp324–Tyr433) remained bound to the protease domain as confirmed by size exclusion chromatography.

Crystallization. N263Q legumain activated at pH 4.0 was crystallized in complex with the covalent Z-Ala-Ala-AzaAsn-chloromethylketone (Z-AAN-cmk) inhibitor, which was a generous gift from U. Demuth (Probiodrug, Halle, Germany) and the Ac-Tyr-Val-Ala-Asp-chloromethylketone inhibitor (Ac-YVAD-cmk, Bachem) as described (1). The glycosylation site mutant N272Q was used to solve the crystal structure of the proenzyme.

Data Collection, Processing, and Structure Solution. X-ray diffraction data collection was performed at 100 K at beamlines ID14-4, ID29 (European Synchrotron Radiation Facility), and XRD1 (ELETTRA synchrotron light source) equipped with a Q315r ADSC CCD detector, a Pilatus 6M, and a Pilatus 2M detector, respectively. Peak datasets of ethylmercuryphosphate (EMP)-soaked Ac-YVAD-cmk and Z-AAN-cmk–inhibited legumain crystals were collected at a wavelength of 1.0059 and 1.0032 Å, respectively. Native datasets of Ac-YVAD-cmk–inhibited legumain crystallized at pH 7.5 and pH 5.0, and prolegumain were measured.

Datasets were processed by using iMOSFLM (5) and SCALA from the CCP4 program suite (6). The automatic crystal structure determination pipeline Auto-Rickshaw was used to obtain an initial model of an EMP-soaked legumain crystal by following the SAS protocol (7). Two mercury sites could be identified by using the program SHELXD (8). Initial phases were calculated by using MLPHARE (6) followed by density modification and phase extension by using the program DM (9). BUCCANEER was used to produce a partial model (10), which was further extended by manual rebuilding in COOT (11). This model was used for molecular replacement in native datasets by using PHASER (12). Iterative cycles of rebuilding in COOT followed by refinement in REFMAC (13) were performed. CIF descriptions of covalent inhibitors were created by using JLigand (14).

To solve the structure of prolegumain coordinates of Ac-YVAD-cmk–inhibited legumain were used as molecular replacement model for PHASER. Density modification using RESOLVE from the Pythonbased Hierarchical Environment for Integrated Xtallography (PHENIX) program suite (15) was carried out, followed by model building using BUCCANEER. Iterative cycles of manual rebuilding in COOT, molecular replacement using PHASER, and density modification using RESOLVE resulted in an initial model that could be further improved by phase enhancement by NCS averaging using DM and repeated cycles of rebuilding in COOT and refinement using REFMAC.

The quality of the resulting structures was assessed by using PROCHECK (16), MolProbity (17) and canonical density expansion (CDE) (18). Coordinates and structure factors were deposited at the Protein Data Bank (PDB) with ID codes 4AW9, 4AWA, 4AWB, and 4FGU. Figures illustrating structures were created by using PyMol (19).

Electrostatic surface potentials were created with APBS (20) after assigning charges at pH 7.0 by using the program Pdb2pqr (21). Surface potentials were contoured at ± 10 kT/e.

Isomorphous Difference Density. To visualize the effect of pH changes on activated legumain, the enzyme-YVAD-cmk inhibitor complex was crystallized at pH 7.5 and pH 5.0 (100 mM citric acid), respectively. Structures at both pH values were solved via molecular replacement as described in more detail above. A σ_A-weighted difference map $F_{\text{pH7.5}}$–$F_{\text{pH5.0}}$ was calculated based on the pH 7.5 structure.

Molecular Modeling of Legumain with Macromolecular Binding Partners. *In silico* models of legumain binding to the two known interaction partners cystatin C and integrin $\alpha_v\beta_3$ were created. The crystal structure of human cystatin C (3GAX) was used to predict inhibitor binding. Technically, the noncanonical legumain reactive center loop (Lys36–Asn39) was aligned to the YVAD peptide of the legumain–YVAD-cmk complex structure by using COOT. Binding of legumain to the cell surface receptor integrin $\alpha_v\beta_3$ was simulated based on 1L5G, a complex structure of the extracellular domain of human integrin $\alpha_v\beta_3$ bound to a cyclic RGD-containing peptide. To dock legumain on the receptor, the RGD motif of YVAD-cmk inhibited legumain was used to align it to the cyclic peptide. An extended state of the receptor was modeled based on ref. 22. Essentially, the C-terminal calf-1 and calf-2 domains were rotated ~120° from the bent conformation into a straightened conformation. EGF-1 and EGF-2 domains of the β_3 subunit were modeled by translocating the EGF-3 and -4 domains. Transmembrane and cytoplasmic parts were modeled at the C terminus.

1. Dall E, Brandstetter H (2012) 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* 68(Pt 1):24–31.
2. Breitling R, et al. (2002) Non-pathogenic trypanosomatid protozoa as a platform for protein research and production. *Protein Expr Purif* 25(2):209–218.
3. Ericsson UB, Hallberg BM, Detitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* 357(2):289–298.
4. Niesen F (2010) *Excel Script for the Analysis of Protein Unfolding Data Acquired by Differential Scanning Fluorimetry (DSF)* (Structural Genomics Consortium, Oxford), Version 3.0..
5. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271–281.
6. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
7. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-rickshaw: An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr* 61(Pt 4):449–457.
8. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58(Pt 10 Pt 2):1772–1779.
9. Cowtan K (1994) DM: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31:334–338.
10. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):1002–1011.
11. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
12. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
13. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
14. Lebedev AA, et al. (2012) JLigand: A graphical tool for the CCP4 template-restraint library. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):431–440.
15. Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58(Pt 11): 1948–1954.
16. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231(4):1049–1067.
17. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375–W383.
18. Ginzinger SW, Gruber M, Brandstetter H, Sippl MJ (2011) Real space refinement of crystal structures with canonical distributions of electrons. *Structure* 19(12): 1739–1743.
19. DeLano WL (2002) *PyMol* (DeLano Scientific, San Carlos, CA).
20. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041.
21. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32(Web Server issue):W665–W667.
22. Xiong JP, et al. (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294(5541):339–345.
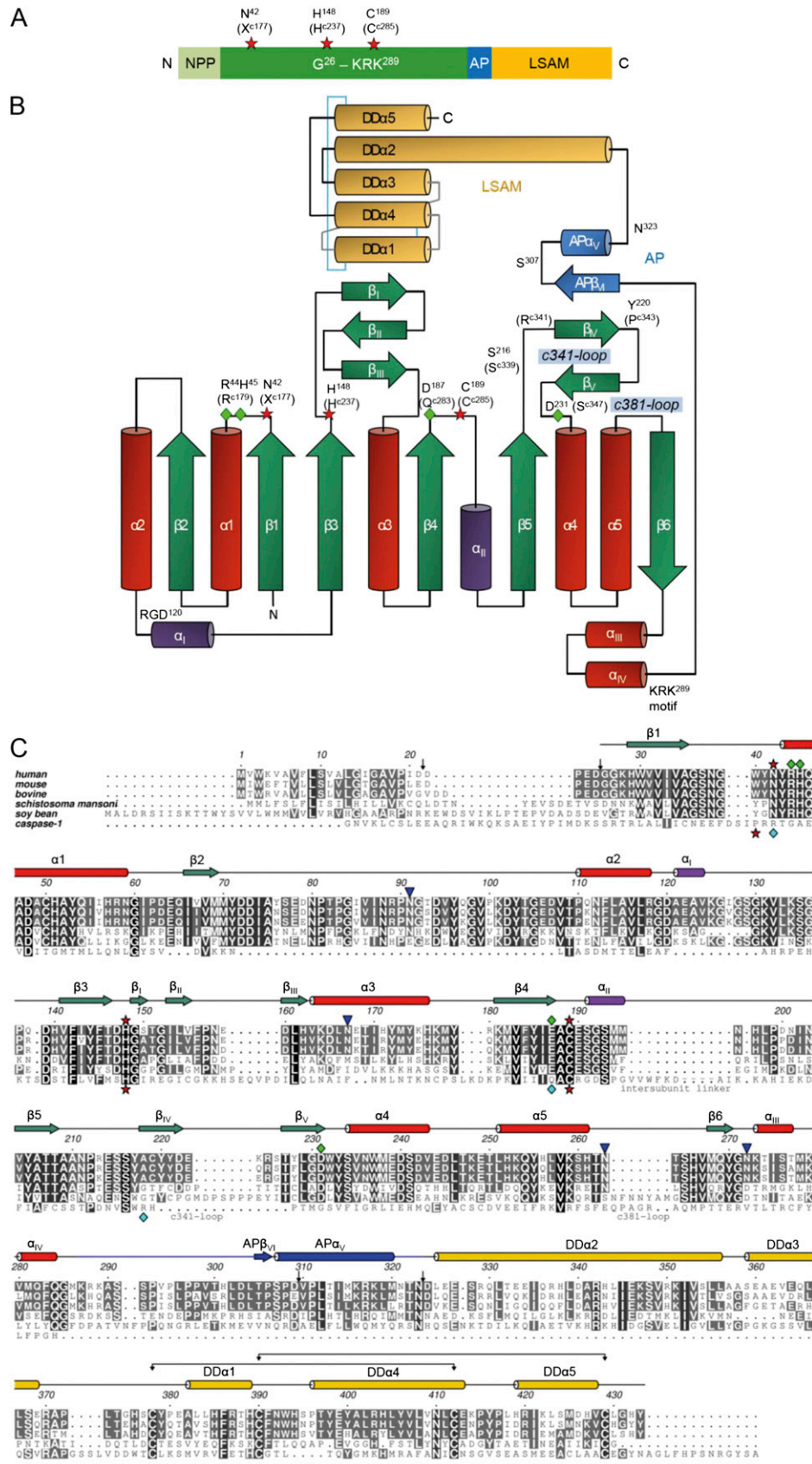
Fig. S1. Topology diagram and sequence alignment of (pro)legumain. (*A*) One-dimensional scheme of prolegumain. Active site residues are indicated by stars. AP, activation peptide (Lys287–Asn323); LSAM, legumain stabilization and activity modulation domain (Asp324–Tyr433), the catalytic domain spans Gly26-

Met286; NPP, N-terminal propeptide (Val18–Asp25). (*B*) Topology diagram modeled after ref. 1). Legumain-specific elements are colored in purple, disulfide bonds within the LSAM domain are indicated as blue lines, and linkers that deviate from standard death domain connections (DDα3→DDα1; DDα1→DDα4) are shown as gray lines. Red stars, catalytic residues; green diamonds, S1-specificity residues. (*C*) Sequence alignment of human (Q99538), mouse (NP_776526.1), bovine (NP_035305.1), schistosome (CAB71158.1), and soybean (NP_001236678.1) legumain together with human caspase-1 (P29466). Arrows, autocatalytic cleavage sites; red stars, catalytic residues; green diamonds, S1-specificity residues; blue triangles, glycosylation sites (Asn91, Asn167, Asn263, and Asn272). Secondary structure elements are based on our legumain structure. The alignment was created with ClustalW (2) and modified with Aline (3). For caspase-1 (PDB ID code 1ICE), a structure-based alignment was created by using Topmatch (4).

1. Fuentes-Prior P, Salvesen GS (2004) The protein structures that shape caspase activity, specificity, activation and inhibition. *Biochem J* 384(Pt 2):201–232.
2. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680.
3. Bond CS, Schüttelkopf AW (2009) ALINE: A WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* 65(Pt 5):510–512.
4. Sippl MJ, Wiederstein M (2008) A note on difficult structure alignment problems. *Bioinformatics* 24(3):426–427.
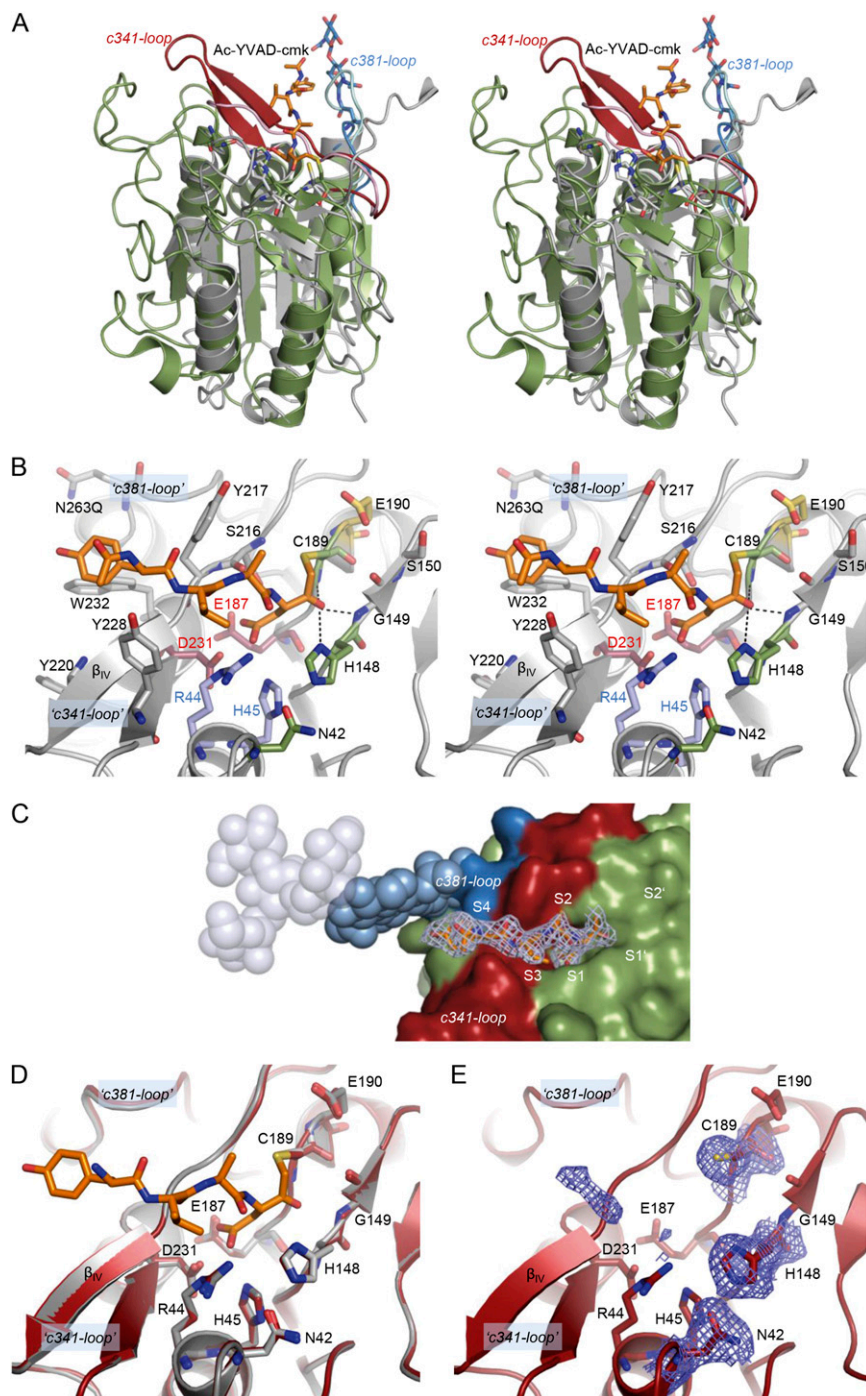
Fig. S2. Binding of the Ac-YVAD-cmk inhibitor is analogous to the Bz-AAN-cmk inhibitor. (*A*) Stereoview of a superimposition of active legumain (green) in complex with the Ac-YVAD-cmk inhibitor (orange) to caspase-9 (PDB ID code 1JXQ; gray). Active site residues are shown as sticks, and substrate specificity loops c341 (red) and c381 (blue) are shown for legumain (dark coloring) and caspase-9 (light coloring). An *N*-linked glycan on residue N263 is shown as blue sticks as observed in the prolegumain structure. (*B*) Zoom-in stereoview to the Ac-YVAD-cmk inhibitor (orange) complex. Active site residues are in green, residues forming the S1-pocket in red and blue, and residues forming the substrate binding cleft in gray. For crystallization reasons, Asn263 was mutated to Gln (N263Q) to avoid glycosylation at this site. (*C*) Zoom-in view on the substrate binding cleft of legumain in complex with Ac-YVAD-cmk (orange). The solid surface patches provided by the specificity-determining loops c341 and c381 (caspase-1 numbering) are colored in red and blue, respectively. A glycan on residue N263 was added as observed in the prolegumain structure, indicated as light blue spheres; further sugars were modeled based on a glycan structure observed in PDB ID code 4B7I (gray spheres). The electron density ($2F_{obs}-F_{calc}$) defining the inhibitor is contoured at 1σ over the mean. (*D*) Superimposition of mercury-treated, inhibitor-free (red) and Ac-YVAD-cmk–inhibited (gray) legumain. Active site residues are indicated as sticks; the Ac-YVAD-cmk inhibitor is shown in orange. (*E*) Detailed view on the active site of uninhibited legumain. Active-site residues are shown as sticks. The electron density ($2F_{obs}-F_{calc}$) defining the catalytic residues and the nonprimed substrate binding cleft is contoured at 1σ over the mean.

Fig. S3. Superimposition of procaspase-9 CARD domain on legumain's LSAM domain. The LSAM domain is colored in wheat and procaspase-9 CARD domain (PDB ID code 3YGS) in red.

Fig. S4. The LSAM domain remains bound to the catalytic domain following β-activation. (*A*) At pH 4.5, prolegumain can be activated autocatalytically to β-legumain. pH 4.5-activated legumain was subjected to size exclusion chromatography, and the peak fractions 1–5 were analyzed by SDS/PAGE. Both the catalytic domain and the LSAM domain are found in the same fractions, indicating that the LSAM domain is tightly linked to the AEP domain. M, molecular mass marker in kilodaltons. (*B*) Turnover of AAN-AMC by AEP (black bars) and β-legumain (ACP; gray bars) can be inhibited by small peptides. AEP and ACP were incubated with AAN-AMC at pH 5.5 in the presence of peptides harboring a free C terminus (AAN, AANA, AANAA, and AANAAA) and a blocked C terminus (AANA-NH2; AANAblock). Activity was normalized to maximum activity of AEP and ACP in the presence of DMSO (control). (*C*) Competition capacity of peptide substrates with varying number of primed residues. The reduction of ACP activity toward the fluorogenic substrate Z-AAN-AMC by the small peptides in *B* is represented as %-inhibition compared with the DMSO control. Black bars, AEP; gray bars, ACP. Data in *B* and *C* are represented as mean ± SD of ≥3 experiments.
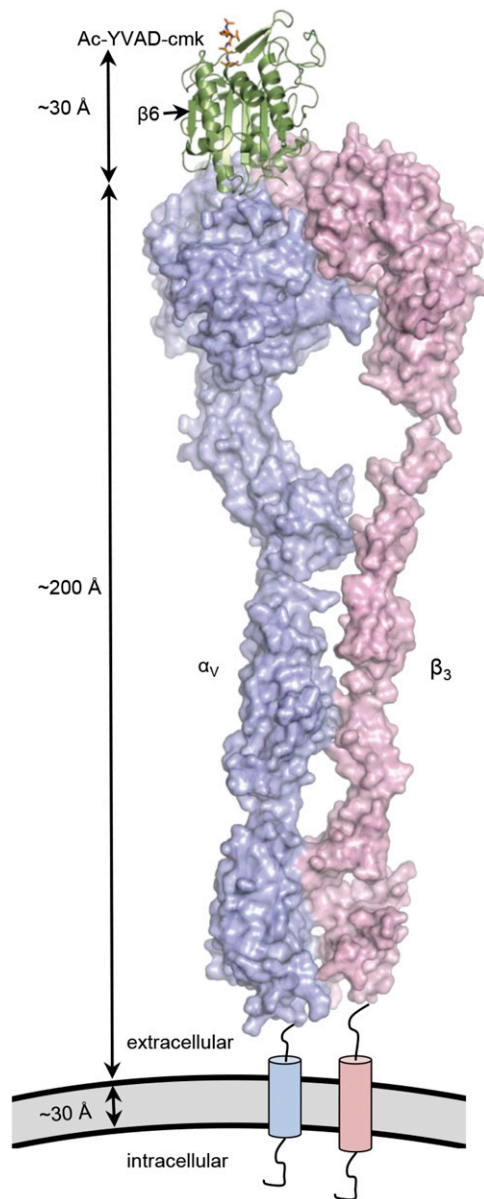
Fig. S5. Complex model of legumain with integrin $\alpha_V\beta_3$ suggesting that $\alpha_V\beta_3$ acts as an allosteric ESS (electrostatic stability switch) trigger. Ac-YVAD-cmk–inhibited legumain (green cartoon) was docked onto the extracellular segment of integrin $\alpha_V\beta_3$ by aligning its RGD motif to the cocrystallized cyclic RGD-containing peptide (PDB ID code 1L5G). The α_V subunit is shown as blue surface, and the β_3 subunit as pink surface. A straightened state of the integrin was modeled based on the reported crystal structure (1). The Ac-YVAD-cmk inhibitor is indicated as orange sticks. Transmembrane and intracellular parts were added.

1. Xiong JP, et al. (2001) Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* 294(5541):339–345.
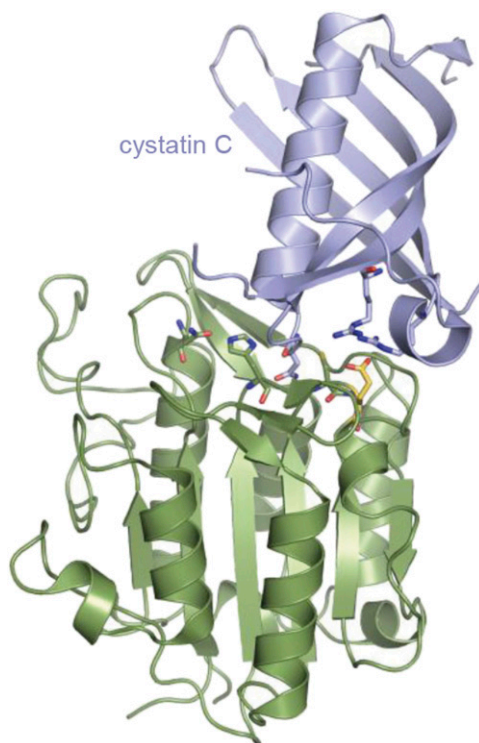
Fig. S6. Docking model suggesting that legumain recognizes cystatin C via noncanonical interactions to Glu190. Human cystatin C (PDB ID code 3GAX, blue) was superimposed onto Ac-YVAD-cmk inhibited legumain (green) by using the P4–P1 residues (Lys36$_{CysC}$–Asn39$_{CysC}$) from the cystatin with the corresponding inhibitor residues. Given the binding of Asn39$_{CysC}$ into legumain's S1 site (1), docking studies suggested that a basic cystatin C cluster binds to the primed recognition sites, allowing for attractive interactions with Glu190 (yellow sticks). Catalytic residues are indicated as green sticks.

1. Alvarez-Fernandez M, et al. (1999) Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. *J Biol Chem* 274(27):19195–19203.
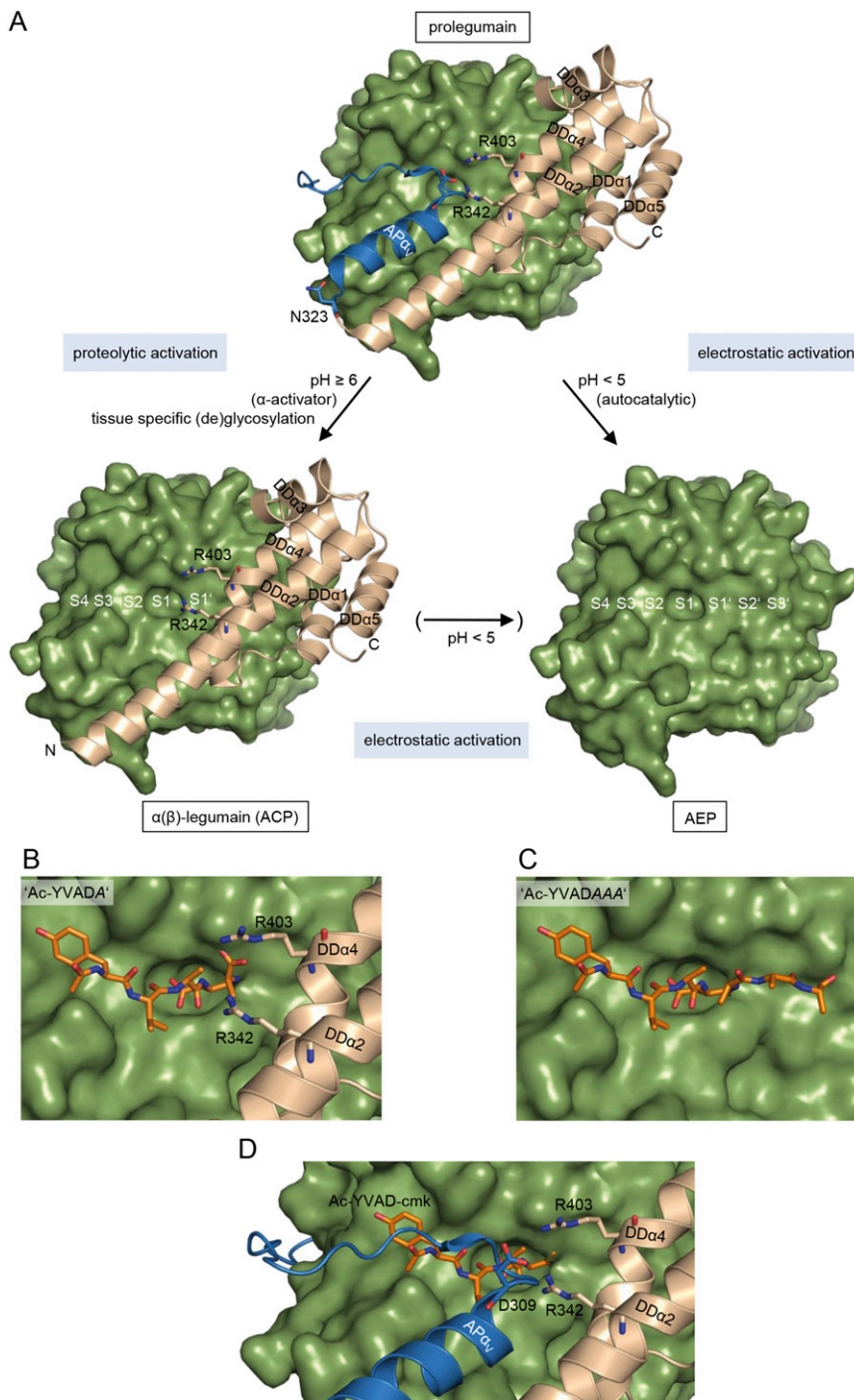
Fig. S7. Context-dependent activation of legumain. (*A*) At pH ≥ 6 and in the presence of an α-activator (e.g., PC2), prolegumain can be proteolytically converted to α(β)-legumain displaying carboxypeptidase activity (ACP). The β-cleavage can be accomplished autocatalytically either by ACP (Asp324 at P1′) or by trace amounts of AEP; the latter will require pH ∼ 6 or lower (left path). Alternatively (right path), at acidic pH < 5.0, autocatalytic activation and release of the complete prodomain will convert prolegumain to the well-known AEP activity. By lowering the pH, it should be possible to convert α(β)-legumain to AEP with concomitant release of the LSAM domain. (*B*) Surface representation of the carboxypeptidase with a pentapeptidic substrate (YVAD|A) modeled to the active site. (*C*) Asparaginylendopeptidase with heptapeptidic substrate (YVAD|AAA) modeled to the active site, spanning P4 to P3′. The substrates in *B* and *C* were modeled based on the Ac-YVAD-cmk complex structure. (*D*) Zoom-in view on the prolegumain active site. The superimposition with the Ac-YVAD-cmk inhibitor, as cocrystallized with AEP, highlights the substrate like binding of the AP and the anchoring of helix APα$_V$ to the LSAM domain via a salt bridge of Asp309 to the double Arg-motif (Arg403 and Arg342).

Table S1. X-ray data collection and refinement statistics

	EMP-AEP	Z-AAN-cmk	Ac-YVAD-cmk	Ac-YVAD-cmk (pH 5)	Prolegumain
Data collection					
Space group	$P4_2$	$P4_1$	$P4_2$	$P4_2$	$R32$
Cell dimensions					
a, b, c; Å	64.3, 64.3, 78.7	63.9, 63.9, 144.3	64.1, 64.1, 78.9	64.3, 64.3, 79.0	185.1, 185.1, 173.0
α, β, γ; °	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120
Resolution, Å	64.3–2.4 (2.6–2.4)	72.1–2.5 (2.7–2.5)	49.7–2.1 (2.2–2.1)	64.3–2.5 (2.6–2.5)	58.8–3.9 (4.1–3.9)
R_{merge}	0.09 (0.14)	0.13 (0.39)	0.12 (0.49)	0.09 (0.23)	0.17 (0.81)
$I/\sigma I$	16.4 (10.0)	8.8 (2.4)	17.4 (5.8)	16.5 (4.0)	6.9 (2.3)
Completeness, %	99.5 (99.6)	95.1 (80.2)	100.0 (100.0)	94.9 (95.2)	99.8 (99.5)
Redundancy	6.9 (6.9)	5.9 (3.1)	14.3 (14.2)	3.8 (3.7)	5.1 (5.4)
Refinement					
Resolution, Å		63.9–2.7	48.0–2.2	64.3–2.5	49.6–3.9
No. of reflections		14,732	13,893	10,128	9,976
R_{work}/R_{free}		22.9/25.5	18.0/20.7	23.7/25.5	28.6/29.9
No. of atoms					
Protein		4257	2151	2151	6548
Ligand/ion		98	57	52	84
Water		91	148	112	56
Overall B factor, Å^2		39.8	31.1	32.9	57.3
Rmsds					
Bond lengths, Å		0.006	0.005	0.007	0.007
Bond angles, °		0.96	1.03	0.94	1.16

Highest resolution shell is shown in parentheses.