Supporting Information

Goulas et al. 10.1073/pnas.1012173108

SI Materials and Methods

Protein Production and Purification. Genomic DNA from *Bacteroides fragilis* strain ATCC 43858 was used as a template for PCR amplification of the coding sequence of profragilysin-3 (UniProt O86049) by using PfuTurbo DNA polymerase (Stratagene) and oligonucleotides 5′-CATTATCCATGGCATGTTCTAATGAAG-3′ (forward primer) and 5′-CTCTCTCGAGATCTAATCGCCATCT G-3′ (reverse primer; both synthesized by Sigma), which included a *Nco*I and a *Xho*I restriction site (underlined), respectively. The 1,140-bp amplicon comprised the coding sequence of the proprotein starting at Ala18; i.e., without the signal peptide as suggested by SIGNALP (1). The amplicon was cloned into a modified pET-28a expression vector (2), which attaches an N-terminal His_6-tag and a tobacco-etch virus protease cleavage site. The gene sequence encoding mature fragilysin-3 CD was amplified by PCR from the aforementioned vector with forward primer 5′-CTT CCATGGCAGTACCTTCTGAAC-3′ (encoding a *Nco*I restriction site) and the same reverse primer as for profragilysin-3. The 558-bp amplicon was cloned in a modified pETM30 vector, which attaches an N-terminal His_6-Z-tag plus a tobacco-etch virus protease cleavage site. In both constructs, the cloning strategy entailed that a tripeptide of sequence (Gly-3)—(Pro-2)—(Met-1) and (Gly-3)—(Ala-2)—(Met-1), respectively, preceded the first protein residue after removal of the tags with tobacco-etch virus protease. Profragilysin-3 Glu349Ala mutant (numbering according to the full pre-pro-protein sequence, UniProt O86049) was obtained by site-directed mutagenesis of the profragilysin-3 expression vector according to (3). The mutagenic forward and reverse primers were, respectively, 5′-GTGCTAACCATGCGGA TGATCCAAAAG-3′ and 5′-CCAATATATGACCTAGTGCGT GTGCCATAAC-3′. Reactions with restriction endonucleases (Fermentas), T4 polynucleotide kinase (Promega), and T4 DNA ligase (Invitrogen) were performed according to manufacturers' instructions. All constructs were verified by DNA sequencing. For protein overexpression, plasmids were transformed in *Escherichia coli* Origami-2 (DE3) cells (Novagen), which were grown in Luria–Bertani medium supplemented with 30 $\mu g\,mL^{-1}$ kanamycin. After initial growth of the culture at 37 °C to an $OD_{600\,nm} \approx$ 0.6, the culture was cooled to 18 °C and protein expression was induced with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside for 18–20 h. The selenomethionine variant of profragilysin-3 was obtained in the same way except that cells were in minimal medium implemented with amino acids, with selenomethionine (Sigma) replacing methionine. Pelleted cells from the cultures were washed in buffer A (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH7.4), harvested, and resuspended in buffer A, which contained EDTA-free Protease Inhibitor Cocktail tablets (Roche Applied Science) and DNAse I (Roche Applied Science). Cells were broken with a cell disrupter (Constant Cell Disruption Systems) at a pressure of 1.35 kbar and nondisrupted cells and cell debris were removed by centrifugation at 35,000 × *g* for 45 min in a Sorvall centrifuge. The supernatant was filtered (0.22 μm pore size; Millipore) and incubated at 4 °C for 1 h with nickel-nitrilotriacetic acid resin (Invitrogen) previously equilibrated with buffer A. Subsequently, the sample was applied to a batch purification open column (BioRad) and washed with five column volumes of buffer A. The His_6-tagged protein was eluted with buffer A containing 250 mM imidazole. The protein sample was dialyzed overnight at 4 °C against buffer A in the presence of 1 mM 1,4-dithio-DL-threitol (Sigma) and subsequently digested at room temperature for 24 h with His_6-tagged tobacco-etch virus protease at an enzyme:substrate ratio of 1:200 (w/w). The digested sample was again passed through the nickel-nitrilotriacetic acid resin column previously equilibrated with buffer A. The eluate was collected, concentrated by ultrafiltration, and further purified by size-exclusion chromatography with a Superdex 75 10/300 GL column (GE Healthcare) previously equilibrated with buffer B (20 mM Tris-HCl, 150 mM NaCl, pH7.4). Active fragilysin-3 was obtained by limited tryptic digestion of purified profragilysin-3 at an enzyme:substrate ratio of 1:100 (w/w) at room temperature for 3 h in buffer B. After trypsin inactivation with 1 $mg\,mL^{-1}$ Pefabloc SC (Roche), the sample was dialyzed against buffer C (20 mM Tris-HCl, 40 mM NaCl, pH7.4) overnight at 4 °C and was subjected to anion-exchange chromatography to remove the inactivated serine protease in a MonoQ HR 5/50 column (GE Healthcare) equilibrated with buffer C. Elution was carried out by using a linear gradient from 40 to 250 mM NaCl within 30 column volumes. Protein-containing fractions were pooled, concentrated by ultrafiltration, and subjected to a final size-exclusion chromatography step with a Superdex 75 10/300 GL column equilibrated with buffer B. At all stages of purification, the purity of the protein samples was assessed by 10–15% SDS-PAGE (tricine buffer) stained with Coomassie blue. All ultrafiltration steps were done with Vivaspin 500 and Vivaspin 2 filter devices of 5-kDa cut-off (Sartorius Stedim Biotech). The concentration of the proteins was determined with the BCA protein assay kit (Thermo Scientific) by using bovine serum albumin as a standard. Concentrated protein samples were stored at 4 °C.

Thermal Shift and Circular Dichroism Assays. Aliquots were prepared by mixing 7.5 μL of 300x Sypro Orange dye (Molecular Probes), 5 μL protein solution (6 $mg\,mL^{-1}$ in buffer B), and 42.5 μL of buffer B. The samples were analyzed in an iCycler iQ Real Time PCR Detection System (BioRad) by using 96-well PCR plates sealed with optical tape. Samples were heated from 20 °C to 90 °C at a rate of 1 °C min^{-1} and the change in absorbance (λ_{ex} = 490 nm; λ_{em} = 575 nm) was monitored over time. The T_m was determined for profragilysin-3 and trypsin-activated fragilysin-3 CD. Samples for far-UV circular dichroism spectroscopy were prepared by dissolving the protein at a final concentration of 0.25 $mg\,mL^{-1}$ in buffer B. Measurements were carried out in a Jasco J-810 spectrometer at 25 °C and 50 °C by using a 2-mm path length cell.

Proteolytic Activity Assays. Proteolytic activity of fragilysin-3 was routinely measured with the fluorescence-based EnzCheck assay kit containing BODIPY FL-casein (10 $\mu g\,mL^{-1}$) as a fluorescein conjugate (Invitrogen) at λ_{ex} = 485 nm and λ_{em} = 528 nm by using a microplate fluorimeter (FLx800, Biotek). Reactions were carried out in buffer D (100 mM Tris-HCl, 150 mM NaCl, pH7.4) at room temperature with a final peptidase concentration of 4.5 $\mu g\,mL^{-1}$. Proteolysis of azo-substrates (at 5 $mg\,mL^{-1}$) was assayed by incubating fragilysin-3 (at 11 $\mu g\,mL^{-1}$) with azocoll (Calbiochem), azocasein (Sigma) or azoalbumin (Sigma) in buffer D at 37 °C for up to 24 h. Reaction mixtures with the latter two substrates were quenched with an equal volume of 5% trichloroacetic acid, centrifuged at 13,000 × *g* for 5 min, and neutralized with an equal volume of 0.5 M NaOH. The enzyme activity was monitored by using a microplate spectrophotometer (PowerWave XS, Biotek) at a wavelength of 520 nm for azocoll and 440 nm for azocasein and azoalbumin. In addition, fragilysin-3 (at 9 $\mu g\,mL^{-1}$) was tested for proteolytic activity on eight fluorogenic substrates (at 10 μM) of sequence: Abz-Lys-Asp-Glu-Ser-Tyr-Arg-K(dnp) (FRET1; for definition of non-amino-acid com-

ponents, see Table S1), Abz-Thr-Val-Leu-Glu-Arg-Ser-K(dnp) (FRET2), Abz-Asp-Tyr-Val-Ala-Ser-Glu-K(dnp) (FRET3), Abz-Tyr-Gly-Lys-Arg-Val-Phe-K(dnp) (FRET4), Abz-Val-Lys-Phe-Tyr-Asp-Ile-K(dnp) (FRET5), Abz-Gly-Ile-Val-Arg-Ala-K(dnp) (FRET6) (λ_{ex} = 260 nm and λ_{em} = 420 nm) (4, 5); Mca-Pro-Leu-Gly-Leu-Dap(Dnp)-Ala-Arg-NH_2 [MMPsub; (6)]; and Mca-Arg-Pro-Lys-Pro-Val-Glu-Nva-Trp-Arg-Lys(Dnp)-NH_2 [NFF3 (7)] (λ_{ex} = 328 nm and λ_{em} = 393 nm). The reactions were performed in buffer D at room temperature and monitored for up to 24 h in a microplate fluorimeter (Infinite M200, TECAN). Finally, proteolytic activity assays against protein substrates (at 1–2 mg mL^{-1}) included human plasma fibrinogen, bovine plasma fibronectin, bovine muscle actin, and bovine milk α-casein (all from Sigma). All reactions were carried out in buffer D at 37 °C for up to 6 h and at an enzyme:substrate ratio of 1 : 100 (w/w) except for reactions with fibronectin, where the ratio was 1 : 10 (w/w). Hydrolysis was assessed by 10–15% SDS-PAGE.

Sample Identification. Samples destined to N-terminal sequencing by Edman degradation were analyzed by SDS-PAGE and electroblotted onto an Immun-Blot PVDF membrane (BioRad). Membranes were stained with Coomassie R-250 and air-dried. The bands were cut and analyzed at the Protein Chemistry Facility of the Centro de Investigaciones Biológicas in Madrid (Spain) (http://www.cib.csic.es/en/servicio.php?iddepartamento=27). Peptide cleavage sites of fluorogenic substrates were determined by MALDI-TOF fragmentation at the Laboratori de Proteòmica at the Institut de Recerca Hospital Universitari Vall d'Hebron in Barcelona (Spain) (http://www.ir.vhebron.net/easyweb_irvh/Serveis/UCTS/Proteomica/tabid/124/Default.aspx).

pH Optimum and (Autolytic) Activation in Vitro. The pH dependence of fragilysin-3 activity was assessed with the azocoll digestion assay. Reactions were carried out as described above except for the buffer, which was either 100 mM sodium acetate (pH4.0, 4.5, 5.0, and 5.5), 100 mM sodium phosphate (pH6.0, 6.5, and 7.0), or 100 mM Tris-HCl (pH7.5, 8.0, 8.5, and 9.0). The reaction mixtures were incubated at 37 °C for 6 h and monitored by using a microplate spectrophotometer (PowerWave XS, Biotek). To assess autolysis, samples of wild-type and Glu349Ala-mutant profragilysin-3 (at 25 mg mL^{-1} in buffer B) were incubated at 37 °C for up to 6 days. Autoproteolysis was assessed by 15% SDS-PAGE. To assess the importance of the N-terminus of the catalytic fragilysin moiety, limited proteolysis of profragilysin-3 was carried out with the serine proteinases α-chymotrypsin from bovine pancreas, subtilisin A from *Bacillus licheniformis*, and proteinase K from *Engyodontium album* (all Sigma) at enzyme: substrate ratios of 1 : 50, 1 : 500 and 1 : 100 (w/w), respectively, at room temperature for various times in buffer B. Activating serine proteinases were subsequently inhibited by 1 mg mL^{-1} Pefabloc and the activity of fragilysin was monitored against BODIPY FL-casein (see above).

Inhibition of Proteolytic Activity in Vitro. Fragilysin-3 (at 4.5 μg mL^{-1} in180 μL of buffer D) was incubated for 30 min with different classes of protease inhibitors (see Table S2). Subsequently, 20 μL of BODIPY FL-casein (final substrate concentration 10 μg mL^{-1}) was added to the reaction mixture, which was incubated at room temperature for 1 h. The remaining activity was measured in a microplate fluorimeter as described above.

Crystallization and Structure Analysis. Crystallization assays were performed by the sitting-drop vapor diffusion method. Reservoir solutions were prepared by a Tecan robot and 100-nL crystallization drops were dispensed on 96 × 2-well MRC plates (Innovadyne) by a Cartesian (Genomic Solutions) nanodrop robot at the High-Throughput Crystallography Platform (PAC) of the Barcelona Science Park. No crystals were obtained for trypsin-activated fragilysin-3. In contrast, crystals suitable for structure analysis were obtained for profragilysin-3 in a Bruker steady-temperature crystal farm at 4 °C from equivolumetric drops containing protein solution (at 8–15 mg mL^{-1} in 20 mM Tris-HCl, 150 mM NaCl, pH7.4) and 100 mM sodium citrate dihydrate, 20% PEG 3000, pH5.5 as reservoir solution. These conditions were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research). Prism shaped single crystals of two different types appeared within one week for wild-type, selenomethionine-derivatized, and Glu349Ala-mutant profragilysin-3. A cryocooling protocol was established consisting of successive passages through reservoir solution containing increasing concentrations of glycerol (up to 20%). Complete diffraction datasets were collected from liquid-N_2 flash-cryo-cooled crystals at 100 K (provided by an Oxford Cryosystems 700 series cryostream) on ADSC Q315R CCD detectors at beam lines ID23-1 (wild-type and selenomethionine-derivative) and ID29 (Glu349Ala-mutant) of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona." Crystals were either monoclinic (wild-type) or orthorhombic (selenomethionine-derivative and Gly349Ala-mutant), with four or two molecules per asymmetric unit, respectively. Diffraction data were integrated, scaled, merged, and reduced with programs XDS (8) and SCALA (9) within the CCP4 suite of programs (10) (see Table S3).

The structure of profragilysin-3 was solved by single-wavelength anomalous diffraction by using the selenomethionine-derivative and program SHELXD (11) (Table S3). Diffraction data of a crystal collected at the selenium absorption-peak wavelength as inferred from a XANES fluorescence scan enabled the program to identify all 20 selenium sites of the dimer present in the asymmetric unit. Subsequent phasing with SHELXE by using a higher-resolved dataset collected from the same crystal at the inflection-point as pseudonative data resolved the twofold ambiguity intrinsic to a SAD experiment due to the difference in the values of the pseudo-free correlation coefficient (11) of the two possible hands. Visual inspection of the heavy-atom sites on a Silicon Graphics Octane2 Workstation by using program TURBO-Frodo (12) allowed us to divide them into two sets of ten and to determine the noncrystallographic twofold axis that related them, with the help of program LSQKAB within the CCP4 suite (10). This symmetry operator was used in a subsequent density modification step with averaging by using program DM (13). These calculations rendered an electron density map that facilitated manual tracing of 305 residues of each protomer. This initial model was refined against the inflection-point dataset with program REFMAC5 (14), which included TLS refinement, and it was subsequently used as a searching model to solve the native structure with program PHASER (15). Four unambiguous solutions were found, which rendered a global log-likelihood gain of 11,024. Subsequently, a run with program ARP/wARP (16) was performed by starting with phases provided by the appropriately rotated and translated searching models. These calculations rendered an excellent electron density map. Thereafter, manual model building alternated with crystallographic refinement until the model was complete. The final model comprised residues 33–201 and 210–397 (molecule A), 33–200 and 211–397 (molecule B), 33–199 and 211–397 (molecule C), and 33–202 and 211–397 (molecule D). As the four molecules are equivalent, discussion focuses on molecule A unless otherwise stated. Seven residues (out of 1,418) were in disallowed regions of a Ramachandran plot and belonged to exposed surface loops of prodomain (PD) (see Table S3). In all four molecules, the linker connecting the PD with the CD was flexible and undefined by electron density for between eight and eleven residues, which could not be traced. In addition, three PD loop regions (Lys144-Glu151, loop Lα2η1; Asp160-Tyr169, Lβ9β1; and Ile183-Ile192, Lα3β11) were flexible, and traced on the basis of weak electron density maps to

preserve chain continuity. To examine such disorder/flexibility, the structure of the catalytically impaired Glu349Ala-mutant was solved by Patterson search as mentioned above and initially refined with REFMAC5 ($R_{factor} = 0.240$; free $R_{factor} = 0.259$). No significant differences were found with the wild-type structure in the critical regions, so the mutant structure was not further refined. In addition, N-terminal sequence analysis of both wild-type and mutant protein crystals revealed only intact protein starting at residue Gly-3 (see above). Accordingly, these flexible regions are intrinsic to the PD and not due to proteolysis in the crystallization drops. This is reminiscent of metallopeptidase (MP) zymogens of the matrix metalloproteinase family (MMPs), thermolysins, and astacins, which likewise evinced flexible and disordered regions within their PDs (17, 18). Such flexibility may enhance the functional properties: a PD shields the active-site but not so tightly as to prevent activation at the appropriate site and time point.

Phylogenetic Analysis. Available catalytic-domain structures of four selected human adamalysins/ADAMs and two snake-venom MPs were superimposed by using a graphic display with program TURBO-Frodo onto fragilysin-3 to ascertain the topologically equivalent common structural limits for all proteins [Arg6-Asn191 for adamalysin II; Protein Data Bank (PDB) access code 1IAG; (19)]. In addition, two fungal homologues and two potential bacterial sequences were aligned with the sequence of the minimal structure sequence of adamalysin II with program MULTALIN (20) to delimitate the corresponding sequence stretches of these latter four proteins. In the last step, all eleven sequences were aligned with MULTALIN, which computes parameters for a phylogenetic tree in rfd format. The latter was manually converted to dnd format and plotted as a circular tree with PHYLIP DRAWGRAM at http://mobyle.pasteur.fr/cgi-bin/portal.py?form=drawgram.

Miscellaneous. Figures were prepared with SETOR (21), GRASP (22), and TURBO-Frodo. Structure similarities were investigated with DALI (23). Model validation was performed with MOLPROBITY (24) and the WHATCHECK routine of WHATIF (25). The interaction surface between the prosegment and the mature enzyme moiety was calculated with CNS (26) as half of the surface area buried at the complex interface determined by using a probe radius of 1.4 Å. Close contacts were ascertained with the latter program and the PISA server at http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver (Table S4). Interface shape complementarity was computed with SC within CCP4, which rendered a value of 72%, thus indicating a good fit between the interacting surfaces. Pocket-size calculations were performed with CASTP (27). The final coordinates of wild-type profragilysin-3 have been deposited with the Protein Data Bank at www.pdb.org (access code 3P24).

1. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. *Nat Protoc* 2:953–971.
2. Marrero A, Mallorquí-Fernández G, Guevara T, García-Castellanos R, Gomis-Rüth FX (2006) Unbound and acylated structures of the MecR1 extracellular antibiotic-sensor domain provide insights into the signal-transduction system that triggers methicillin resistance. *J Mol Biol* 361:506–521.
3. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas DJ (1989) A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res* 17:6545–6551.
4. Yongzheng Y, Reymond JL (2005) Protease profiling using a fluorescent domino peptide cocktail. *Mol Biosyst* 1:57–63.
5. Cotrin SS, et al. (2004) Positional-scanning combinatorial libraries of fluorescence resonance energy transfer peptides to define substrate specificity of carboxydipeptidases: assays with human cathepsin B. *Anal Biochem* 335:244–252.
6. Knight CG, Willenbrock F, Murphy G (1992) A novel coumarin-labelled peptide for sensitive continuous assays of the matrix metalloproteinases. *FEBS Lett* 296:263–266.
7. Nagase H, Fields CG, Fields GB (1994) Design and characterization of a fluorogenic substrate selectively hydrolyzed by stromelysin 1 (matrix metalloproteinase-3). *J Biol Chem* 269:20952–20957.
8. Kabsch W (2001) *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, eds Rossmann MG, Arnold E (Kluwer Academic, The Netherlands), pp 730–734.
9. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* 62:72–82.
10. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
11. Sheldrick GM (2002) Macromolecular phasing with SHELXE. *Z Kristallogr* 217:644–650.
12. Carranza C, Inisan A-G, Mouthuy-Knoops E, Cambillau C, Roussel A (1999) *AFMB Activity Report 1996–1999* (CNRS-UPR 9039, Marseille), pp 89–90.
13. Cowtan KD, Main P (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D* 52:43–48.
14. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
15. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674.
16. Perrakis A, Sixma TK, Wilson KS, Lamzin VS (1997) wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Crystallogr D* 53:448–455.
17. Tallant C, Marrero A, Gomis-Rüth FX (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *BBA-Mol Cell Res* 1803:20–28.
18. Guevara T, et al. (2010) Proenzyme structure and activation of astacin metallopeptidase. *J Biol Chem* 285:13958–13965.
19. Gomis-Rüth FX, et al. (1994) Refined 2.0 Å X-ray crystal structure of the zinc-endopeptidase adamalysin II. Primary and tertiary structure determination, refinement, molecular structure and comparison with astacin, collagenase, and thermolysin. *J Mol Biol* 239:513–544.
20. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16:10881–10890.
21. Evans SV (1993) SETOR: Hardware lighted three-dimensional solid model representations of macromolecules. *J Mol Graphics* 11:134–138.
22. Nicholls A, Bharadwaj R, Honig B (1993) GRASP: Graphical representation and analysis of surface properties. *Biophys J* 64(2):A166–A166.
23. Holm L, Kaariainen S, Wilton C, Plewczynski D (2006) Using Dali for structural comparison of proteins. *Curr Protoc Bioinformatics* Chapter 5:Unit 5 5.
24. Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383 (Web server issue).
25. Vriend G (1990) WHAT IF: A molecular modelling and drug design program *J Mol Graph* 8:52–56.
26. Brünger AT, et al. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination *Acta Crystallogr D* 54:905–921.
27. Dundas J, et al. (2006) CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34:W116–W118 (Web Server issue).
28. Tronrud DE, Wen J, Gay L, Blankenship RE (2009) The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynth Res* 100:79–87.
29. Gomis-Rüth FX, Kress LF, Bode W (1993) First structure of a snake-venom metalloproteinase: A prototype for matrix metalloproteinases/collagenases *EMBO J* 12:4151–4157.
30. Gomis-Rüth FX, Meyer EF, Kress LF, Politi V (1998) Structures of adamalysin II with peptidic inhibitors. Implications for the design of tumor necrosis factor α convertase inhibitors. *Protein Sci* 7:283–292.
31. Gomis-Rüth FX (2003) Structural aspects of the *metzincin* clan of metalloendopeptidases. *Mol Biotech* 24:157–202.

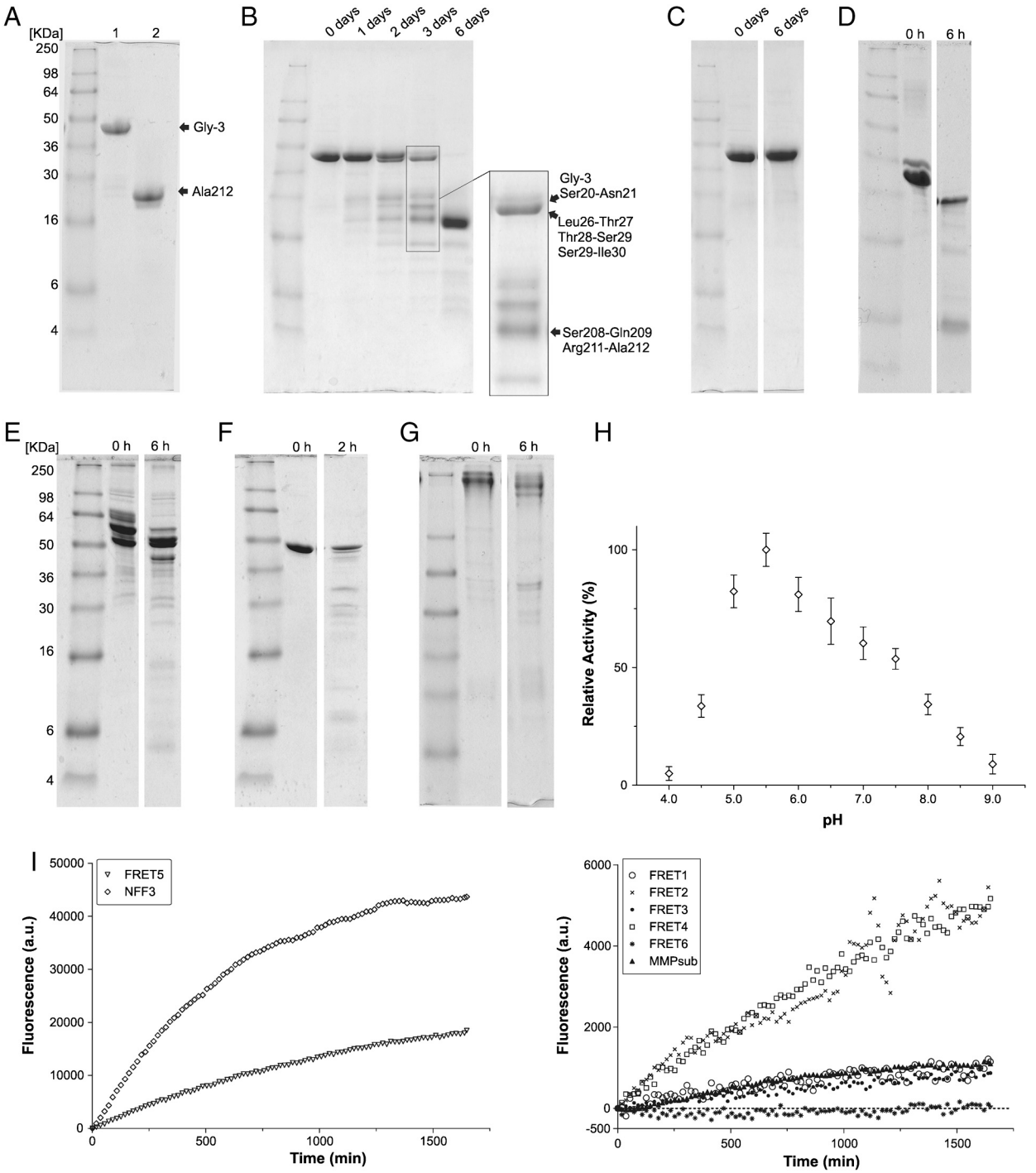

Fig. S1. (Pro)fragilysin-3 studies in vitro. (*A*) SDS-PAGE of purified profragilysin-3 (lane 1) and trypsin-activated fragilysin-3 (lane 2) with the respective N-terminal residues. (*B*) Autolytic cleavage of profragilysin-3 at 37 °C over time. The cleavage sites after 3 days are shown in the framed close-up view. (*C*) Profragilysin-3 Glu349Ala-mutant over time. (*D*)–(*G*) Proteolysis of bovine milk α-casein (*D*), human plasma fibrinogen (*E*), bovine muscle actin (*F*), and human plasma fibronectin by fragilysin-3 (*G*), respectively. (*H*) Activity dependence on pH as determined with azocoll as substrate. Values are represented as means and standard deviations of three independent experiments. (*I*) Proteolysis of fluorogenic substrates (see above for sequences). Substrates NFF3 and FRET5 were best cleaved, whereas the rest were cleaved only moderately (FRET2 and FRET4) or not at all.

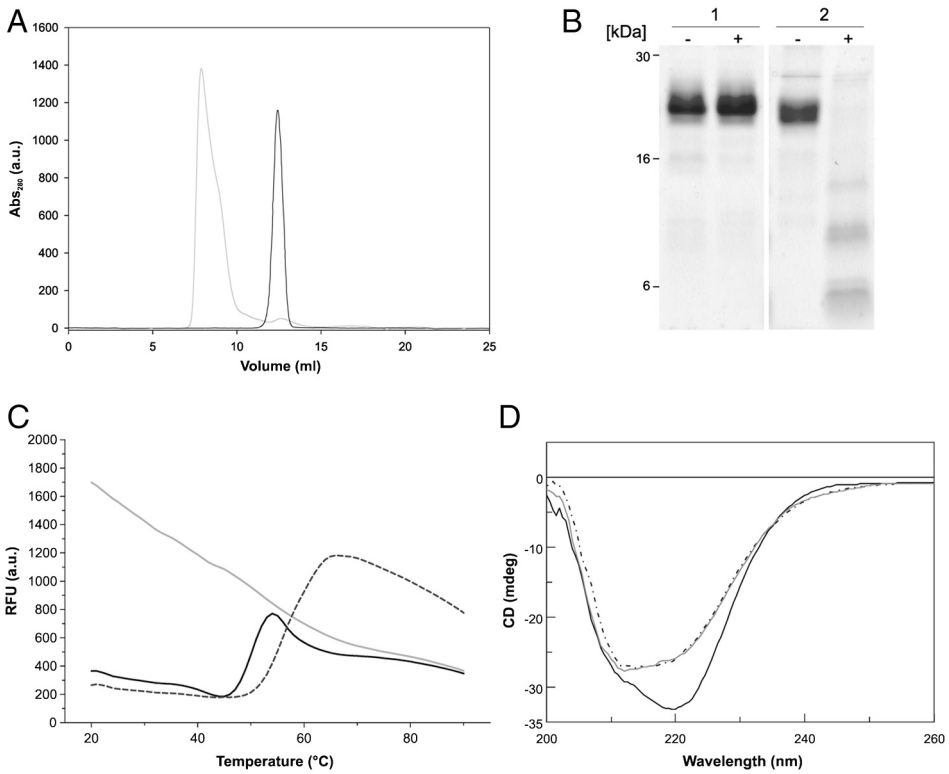

Fig. S2. Conformational analysis of (pro)fragilysin-3. (*A*) Trypsin-activated fragilysin-3 (black line) and the directly expressed variant (gray line) were loaded onto a gel-filtration column, showing different elution times. (*B*) Trypsin-activated fragilysin-3 and the directly expressed variant (1 and 2, respectively) were incubated without (–) and with (+) trypsin at 1 : 100 (w/w) ratio at room temperature for 2 h. (*C*). Thermal shift curves for profragilysin-3 (dashed line), trypsin-activated fragilysin-3 (black line) and the directly expressed variant (gray line). (*D*). Circular dichroism spectra of trypsin-activated fragilysin-3 at 25 °C (black line) and 50 °C (dashed line) and of the directly expressed variant (gray line).

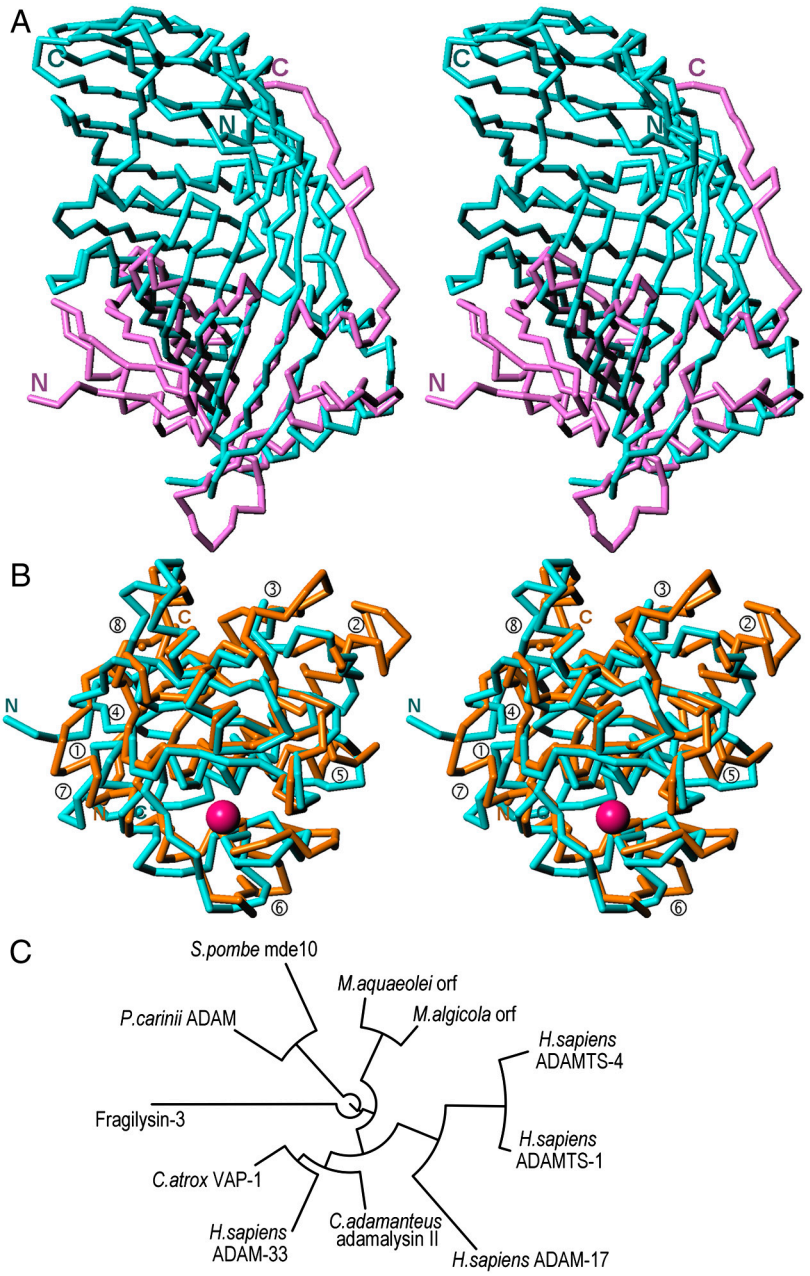

Fig. S3. Structural similarities and phylogenetic studies. (*A*) Superimposition in stereo of profragilysin-3 PD (lilac) with bacteriochlorophyl A protein from *Prosthecochloris aestuarii* 2 K [cyan; PDB 3EOJ (28)]. (*B*) Superimposition in stereo of fragilysin-3 CD (cyan) onto adamalysin II (orange) from *Crotalus adamanteus* [PDB 1IAG; (19, 29, 30)] in standard orientation. Diverging regions between both structures include: ① the N-termini are arranged distinctly due to the unique C-terminal extension after helix α7 in the bacterial enzyme; ② the backing helix is one turn longer in adamalysin II and preceded by a bulge, which is absent in the bacterial enzyme; ③ the loop connecting the outermost strand of the five-stranded β-sheet with the adamalysin helix is longer in the ophidian enzyme and includes a short helix present in all adamalysin/ADAM structures, which is absent in fragilysin-3; ④ the loop connecting the upper-rim strand with the last strand of the β-sheet is longer in fragilysin-3 and it opens to occupy the space of the C-terminal stretch after the last helix of the lower subdomain in adamalysin II; ⑤ the active-site helix starts two helical turns earlier in the ophidian enzyme; and ⑥ with seven residues separating the third zinc-binding histidine (His358) and the Met-turn methionine (Met366), fragilysin-3 better matches MMPs within this loop than adamalysin II, which has 13 residues. However, the latter finding should not be overemphasized as the length and conformation of this segment vary in adamalysins/ADAMs (31). (*C*) Circular phylogenetic tree reflecting evolutionary distances between *B. fragilis* fragilysin-3 (this work) and selected adamalysin/ADAMs, which include human ADAM-17 (PDB 1BKC), ADAM-33 (PDB 1R54), ADAMTS-1 (PDB 2JIH), and ADAMTS-4 (PDB 2RJP); snake-venom MPs adamalysin II from *Crotalus adamanteus* (PDB 1IAG) and VAP-1 from *Crotalus atrox* (PDB 2ERO); biochemically characterized fungal relatives, *Schizosaccharomyces pombe* mde10 (UniProt O13766) and *Pneumocystis carinii* ADAM (UniProt A3QZA9); and putative bacterial sequential relatives from *Marinobacter aquaeoli* (UniProt A1U5B6) and *Marinobacter algicola* (UniProt A6EZB8). No structure is available for the last four proteins.

Table S1. Fragilysin-3 cleavage sites

Autolytic cleavage I									M	–	A	–	C	–	S	–	✂	–	N	–	E	–	A	–	D
Autolytic cleavage II									A	–	D	–	S	–	L	–	✂	–	T	–	T	–	S	–	I
Autolytic cleavage III									S	–	L	–	T	–	T	–	✂	–	S	–	I	–	D	–	A
Autolytic cleavage IV									L	–	T	–	T	–	S	–	✂	–	I	–	D	–	A	–	P
Autolytic cleavage V									I	–	T	–	E	–	S	–	✂	–	Q	–	T	–	R	–	A
Autolytic cleavage VI									S	–	Q	–	T	–	R	–	✂	–	A	–	V	–	P	–	S
α-Casein									E	–	Q	–	K	–	Y	–	✂	–	I	–	Q	–	K	–	E
Fibrinogen									G	–	K	–	E	–	K	–	✂	–	V	–	T	–	S	–	G
FRET5							Abz	–	V	–	K	–	F	–	Y	–	✂	–	D	–	I	–	K (dnp)		
FRET2					Abz	–	T	–	V	–	L	–	E	–	R	–	✂	–	S	–			K (dnp)		
NFF3	Mca	–	R	–	P	–	K	–	P	–	V	–	E	–	O	–	✂	–	W	–	R	–	K (dnp)		

Mca stands for 7-methoxycoumarin-4-acetyl, abz for aminobenzoyl, O for norvaline, dnp for 2,4-dinitrophenylamino, and dap for L-diaminopropionyl.

Table S2. Inhibition of fragilysin-3 activity

Inhibitor	Concentration	Specificity	% Relative activity
None	—	—	100
Bovine lung aprotinin	0.3 μM	Serine proteases	97
PMSF	1 mM	Serine proteases	82
Benzamidine	5 mM	Serine proteases	94
Pefabloc	4 mM	Serine proteases	88
Iodoacetamide	1 mM	Cysteine proteases	92
E-64	10 μM	Cysteine proteases	94
Pepstatin A	10 μM	Aspartic proteases	92
1,10-Phenanthroline	5 mM	Metallopeptidases (MPs)	0
EDTA	5 mM	MPs	0
$ZnCl_2$	5 mM	MPs	18
CT1746	10 μM	Matrixins (MPs)	13
Phosphoramidon	10 μM	Thermolysin (MP)	88
Captopril	1 mM	Angiotensin-converting enzyme (MP)	101

Table S3. Crystallographic data

Dataset	Native	Selenomethionine (absorption-peak)*	Glu349Ala-mutant
Space group/cell constants (a, b, c, in Å; β in ° if ≠ 90)	$P2_1$/82.74, 69.14, 158.91, 91.57	$P2_12_12_1$/69.1, 83.2, 160.1	$P2_12_12_1$/69.51, 82.27, 158.71
Wavelength (Å)	1.0723	0.9793	0.9724
No. of measurements/unique reflections	749,668/164,503	119,679/31,601	474,707/72,415
Resolution range (Å) (outermost shell)[†]	45.2–1.80 (1.90–1.80)	44.9–2.50 (2.64–2.50)	42.1–1.90 (2.00–1.90)
Completeness [/Anomalous completeness] (%)	98.9 (98.0)	97.1 (97.3)/88.8 (85.2)	99.9 (99.9)
R_{merge}[‡]	0.075 (0.597)	0.073 (0.305)	0.060 (0.631)
$R_{r.i.m.} (= R_{meas})/R_{p.i.m.}$[‡]	0.084(0.714)/0.037(0.388)	0.095(0.395)/0.060(0.248)	0.065(0.687)/0.025(0.268)
Average intensity $(\langle[\langle I\rangle/\sigma(\langle I\rangle)]\rangle)$	15.9 (2.4)	15.5 (5.5)	19.4 (3.6)
B-Factor (Wilson) ($Å^2$)/Average multiplicity	24.4/4.6(3.1)	35.5/3.8(3.7)	29.4/6.6(6.3)
Resolution range used for refinement (Å)	∞–1.80		
No. of reflections used (test set)	163,722 (781)		
Crystallographic R_{factor} (free R_{factor})[‡]	0.170 (0.204)		
No. of protein atoms[§]/solvent molecules/ ligands/ions	11,395/1147/ 2 tetraethylene glycol, 6 glycerol, 2 azide/4 zinc		
Rmsd from target values			
bonds (Å)/angles (°)	0.011/1.27		
bonded B-factors (main-chain/side chain) ($Å^2$)	0.79/2.07		
Average B-factors for protein atoms ($Å^2$)	17.7		
Main-chain conformational angle analysis[¶]			
Residues in favored regions/outliers/all residues	1381/7/1418		

*Friedel-mates were treated as separate reflections.
[†]Values in parentheses refer to the outermost resolution shell.
[‡]For definitions, see table 1 in ref. 1.
[§]Including atoms and residues in alternative conformations.
[¶]According to MOLPROBITY (2).

1 Mallorquí-Fernández N, et al. (2008) A new autocatalytic activation mechanism for cysteine proteases revealed by *Prevotella intermedia* interpain A. *J Biol Chem* 283:2871–2882.
2 Davis IW, et al. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35:W375–W383 (Web server issue).

Table S4. Direct interactions between PD and CD of profragilysin-3 (molecule A)

Prodomain	Catalytic moiety	Dist.(Å)	Prodomain	Catalytic moiety	Dist.(Å)
van der Waals interactions			Hydrogen bonds		
Arg80 Cδ	Lys255 Cδ	3.92	Arg80 Nη2	Asn246 Oδ1	2.98
Thr81 Cγ2	Phe262 Cε2	3.79	Arg80 O	Tyr249 Oη	3.09
Ile83 Cγ2	Met241 Sδ	3.81	Thr81 Oγ1	Tyr249 Oη	2.92
Leu85 Cδ2	Met241 Cε	3.77	Lys82 N	Leu260 O	3.10
Leu85 Cδ2	Leu264 Cβ	3.93	Lys82 O	Phe262 N	2.83
Ile104 Cγ2	Ser238 Cβ	3.76	Gln84 Oε1	His261 Nε2	2.71
Phe109 Cε1	Pro234 Cβ	3.61	Gln84 N	Phe262 O	2.92
Phe109 Cε2	Asn235 Cα	3.59	Gln84 O	Leu264 N	2.95
Val111 Cγ2	Asn235 Cβ	3.71	Asp86 N	Leu264 O	2.82
Tyr139 Cε2	Leu340 Cδ2	3.80	Phe103 O	Gln242 Nε2	3.29
Ile175 Cδ1	Leu340 Cδ2	3.43	Tyr126 Oη	Asn235 Oδ1	2.59
Val177 Cγ1	Met341 Cα	3.90	Tyr126 Oη	Asn235 N	3.22
Ala180 Cβ	Met341 Cε	3.65	Tyr139 Oη	Asn235 Nδ2	2.95
Cys189 Sγ	Thr308 Cγ2	3.91	Tyr139 Oη	Glu236 Oε2	2.55
Cys189 Cβ	Leu314 Cδ2	3.33	Pro187 O	Tyr342 Oη	3.40
Tyr191 Cδ2	Lys369 C	3.98	Lys181 Nζ	Met341 O	2.82
Tyr195 Cδ1	Leu319 Cδ2	3.93	Cys189 N	Tyr342 Oη	3.16
Tyr195 Cδ2	His352 Cε1	3.48	Asp190 N	Asn312 O	2.85
Tyr195 Cε2	His358 Cε1	3.68	Asp190 O	Leu314 N	2.86
Ile196 Cδ1	Leu280 Cδ2	3.95	Asp190 Oδ1	Tyr370 N	2.71
Ile196 Cγ2	Trp318 Cζ3	3.48	Asp194 O	Ser317 N	3.06
Thr198 Cγ2	Trp318 Cε3	3.73	Asp194 Oδ2	His348 Nε2	3.03
Salt bridges			Asp194 Oδ2	His352 Nε2	3.44
Asp86 Oδ2	Arg226 Nη2	2.78	Asp194 Oδ1	His352 Nε2	3.21
Asp86 Oδ1	Arg226 Nη1	2.92	Asp194 Oδ1	His358 Nε2	3.00
Arg113 Nη2	Asp243 Oδ1	2.93	Ile196 N	Ser317 O	2.94
Metallorganic bonds			Ile196 O	Leu319 N	2.90
Asp194 Oδ1	Zn^{2+}	2.15	Gln200 Nε2	Asn320 Nδ2	3.49
Asp194 Oδ2	Zn^{2+}	2.44			

Table S5. Amino-acid variability in profragilysins

	Position	Profragilysin-1	Profragilysin-2	Profragilysin-3
Prodomain	32	A	T	A
	102	S	S	N
	169	D	D	Y
	170	I	L	P
	177	I	I	V
Catalytic domain	228	N	S	N
	232	I	V	V
	257	Y	F	Y
	260	F	L	L
	270	S	S	A
	275	D	N	N
	277	K	D	D
	281	E	D	D
	289	S	A	A
	312	K	N	N
	316	M	I	I
	319	F	L	L
	320	N	D	N
	331	E	K	K
	357	E	R	N
	359	T	A	A
	361	N	D	D
	362	S	P	P
	368	A	S	S
	369	T	K	K
	370	F	Y	Y
	375	S	F	F
	380	K	E	K
	383	D	Y	D
	384	I	R	I
	393	A	I	I