A novel family of soluble minimal scaffolds provides structural insight into the catalytic domains of integral-membrane metallopeptidases **

Mar López-Pelegrín[1,#], Núria Cerdà-Costa[1,#], Francisco Martínez-Jiménez[2,3], Anna Cintas-Pedrola[1], Albert Canals[4], Juan R. Peinado[1,6], Marc A. Marti-Renom[2,3], Carlos López-Otín[5], Joan L. Arolas[1,*] & F. Xavier Gomis-Rüth[1,*]

Running title: Minimal scaffold for M48 and M56 metallopeptidases

[1] Proteolysis Lab; Department of Structural Biology; Molecular Biology Institute of Barcelona, CSIC; c/ Baldiri Reixac, 15-21; 08028 Barcelona (Spain).
[2] Genome Biology Group; Centre Nacional d'Anàlisi Genòmic CNAG; c/ Baldiri Reixac, 4; 08028 Barcelona (Spain).
[3] Gene Regulation, Stem Cells and Cancer Program; Center for Genomic Regulation (CRG); c/ Dr. Aiguader, 88; 08003 Barcelona (Spain).
[4] Molecular Biology Institute of Barcelona, CSIC, and Institute for Research in Biomedicine; c/ Baldiri Reixac, 10-12; 08028 Barcelona (Spain).
[5] Departamento de Bioquímica y Biología Molecular and Instituto Universitario de Oncología; Universidad de Oviedo; 33006 Oviedo (Spain).
[6] Present address: Department of Medical Sciences; University of Castilla-La Mancha; 13071 Ciudad Real (Spain).

[#] These authors contributed equally to this work and share first authorship.
[*] Correspondence should be addressed to: xgrcri@ibmb.csic.es or jlacri@ibmb.csic.es, phone: (+34) 934 020 186/7, fax: (+34) 934 034 979.

Keywords: catalytic domain; membrane protein; metallopeptidase; zinc-binding protein

---

Background: Structural characterization of integral-membrane (IM) metallopeptidases (MPs) faces enormous technical hurdles.
Results: We have discovered a novel family of minimal MPs, minigluzincins, and determined the crystal structures of the zymogens of two family members.
Conclusion: Minigluzincins are valid models for catalytic domains of M48- and M56-family IMMPs.
Significance: They provide a high-resolution scaffold for the design of small-molecule inhibitors of IMMPs.

SUMMARY

In the search for structural models of integral-membrane metallopeptidases (MPs), we discovered three related proteins from thermophilic prokaryotes, which we grouped into a novel family called "minigluzincins". We determined the crystal structures of the zymogens of two of these (*Pyrococcus abyssi* proabylysin and *Methanocaldococcus jannaschii* projannalysin) which are soluble and, with ~100 residues, constitute the shortest structurally characterized MPs to date. Despite relevant sequence and structural similarity, the structures revealed two unique mechanisms of latency maintenance through the C-terminal segments hitherto unseen in MPs: intra-molecular, through an extended tail, in proabylysin, and crosswise inter-molecular, through a helix swap, in projannalysin. In addition, structural and sequence comparisons, as well as phylogenetic and bioinformatics analyses, revealed large similarity with MPs of the gluzincin tribe such as thermolysin, leukotriene A4 hydrolase relatives, and cowrins. Interestingly, gluzincins mostly have a glutamate as third characteristic zinc ligand while minigluzincins have a histidine. Sequence and structure similarity further allowed us to ascertain that minigluzincins are very similar to the catalytic domains of integral-membrane MPs of MEROPS database families M48 and M56, such as FACE1, HtpX, Oma1, and BlaR1/MecR1, which are provided with trans-membrane helices flanking or inserted into a minigluzincin-like catalytic domain. In a time where structural biochemistry of integral-membrane proteins in general still faces formidable challenges, the minigluzincin soluble minimal scaffold may contribute to our understanding of the working mechanisms of these membrane MPs and to the design of novel inhibitors through structure-aided rational drug design approaches.

The plasma membrane is the key information interface between the cytosol and cellular compartments or the extracellular environment, and whole genome analyses have revealed that 20-30% of proteins in humans, bacteria and fungi are polytopic integral-membrane (IM) proteins (1-4). They are of vital importance to living cells through their key roles in communication and transport, which explains why more than 60% of current pharmaceutical drug targets are membrane proteins (5,6). Among IM

proteins are members of the peptidase class of enzymes, including integral-membrane metallopeptidases (IMMPs) from families M48 and M56 (according to the MEROPS database, http://merops.sanger.ac.uk; (7)), which are α-helical for their transmembrane parts, as found in the otherwise unrelated rhodopsins, transporters, and channels (8).

M48 metallopeptidases (MPs) are subdivided into M48A, M48B, and M48C, whose prototypes are, respectively, *Saccharomyces cerevisiae* Ste24p, *Escherichia coli* HtpX, and human mitochondrial Oma1 peptidase. Ste24p was discovered as an enzyme required for maturation of the mating pheromone a-factor in yeast (9). Its mammalian ortholog, farnesylated-protein converting enzyme 1 (FACE1 *alias* Zmpste24), performs a critical cleavage that removes the hydrophobic farnesyl-modified tail of prelamin A (10). Knockout mice show growth retardation, muscular dystrophy, and premature death; similarly, disruptive mutations in humans have drastic consequences for health and lifespan (11-13). HtpX, together with the ATP-dependent protein FtsH, participates in the quality control system of bacterial membrane proteins, which is essential for growth and survival of the cell (14-17). Failure of this system under environmental stress conditions leads to the accumulation of misfolded membrane proteins. Finally, Oma1 helps in maintaining the integrity of the mitochondrial inner membrane (18,19). It plays a crucial role in the proteolytic inactivation of the dynamin-related GTPase Opa1, and its loss causes obesity and defective thermogenesis in mice (20). Family M56, in turn, includes BlaR1 from *Staphylococcus aureus* and *Bacillus licheniformis*, and MecR1 from *S. aureus*. They function as parts of the signal-transduction systems that trigger bacterial resistance to β-lactam antibiotics, a phenomenon that poses a serious threat to animal and human health and is exerted through the synthesis of a β-lactamase or a penicillin-binding protein (21-28). BlaR1 and MecR1 are composed of an N-terminal IMMP domain facing the cytosol and a C-terminal extracellular sensor domain that binds environmental β-lactams (29-32).

The strategic and widespread biological relevance of IMMPs underpins the need to acquire detailed knowledge about them at the molecular level, and atomic structural information, as provided by X-ray diffraction or nuclear magnetic resonance, makes a major contribution to this type of knowledge (33,34). However, structural biochemistry of IM proteins in general faces formidable conceptual and technical challenges due to the low concentrations at which they are naturally present in organisms, the difficulty in finding adequate recombinant overexpression systems that yield sufficient amounts of native-like protein, and their insolubility in strictly aqueous media, which requires screening for detergents and lipids that mimic the native membrane environment (35). Consequently, the only IMMP structures published to date are those of the intra-membrane site-2 protease (S2P; (36)) from the hyperthermophylic archaeon *Methanocaldococcus jannaschii* (MEROPS family M50), an ortholog of a human enzyme that releases the N-terminal transcription factor domain from membrane-bound sterol regulatory element binding proteins (36-38) and is unrelated to M48 and M56 IMMPs, and of human and yeast FACE1/Ste24p (Protein Data Bank [PDB] access codes 4AW6 and 4IL3; (39,40)), the only functional, truly integral-membrane representatives of these two latter families. In such a scenario, strategies aimed at obtaining high-resolution structural information on soluble, correctly-folded fragments of IMMP target proteins, such as globular catalytic domains (CDs) inserted into the overall transmembrane scaffold, may prove helpful in both the study of catalytic mechanisms and the design of drug-like inhibitors. The presence in the amino-acid sequence of M48 and M56 IMMPs of large segments encompassing a zinc-binding motif (HEXXH) characteristic of the gluzincin and metzincin clans of the zincin tribe of MPs (see (41-43) and Fig. 2 in (44)) between transmembrane helices (TMHs) provides a useful starting point for the identification of such CDs, which are mostly cytoplasmic (Fig. 1): in Ste24p and FACE1, these lie between the fifth and sixth of their seven TMHs; in HtpX, between the second and third of its four putative TMHs; and in BlaR1/MecR1, between the third and fourth of their four possible IMMP TMHs (26). By contrast, Oma1 is predicted to have only two transmembrane segments according to bioinformatics approaches (Fig. 1), but this topology and the localization of the active site still remain controversial (18).

In this work, we aimed to study the fold of CDs of M48 and M56 IMMPs to provide insights into the minimal scaffold required for their activity. As extensive expression trials of the putative CDs of several members of both families failed (our unpublished results), we set about searching for uncharacterized standalone hypothetic proteins with significant sequence similarity to these CDs using various bioinformatics approaches. We identified, cloned, and overexpressed three sequences from the archaeal thermophyles *Pyrococcus abyssi* and *M. jannaschii*, and the thermophilic bacterium *Aquifex aeolicus*. We further purified and crystallized the first two, determined and analyzed their crystal structures, and functionally verified their zymogenic forms. Finally, we validated their suitability as valuable structural models for the CDs of M48 and M56 IMMPs.

EXPERIMENTAL PROCEDURES

Bioinformatics sequence analyses — A search with a partial sequence of M56-family BlaR1 from *S. aureus* (residues K133-L250; UniProt [UP] sequence database access code P18357) and MecR1 (L139-L253; UP P0A0A9), putatively encompassing their TMH-flanked CDs, was carried out in 2004 using ProDom protein domain family database (release 2004.1 at http://prodes.tolouse.inra.fr), and identified ProDom family PD020519 as containing this sequence stretch. This family comprised several BlaR1/MecR1-type sequences—and others annotated as related to *E. coli* HtpX—that had been built using the PSI-BLAST program (45) with UP Q57587 (annotated as archaeal uncharacterized protein MJ0123 from the thermoautotrophican archaeon *M. jannaschii*; hereafter referred to as projannalysin) as the query. In contrast to multi-domain BlaR1/MecR1 relatives, UP Q57587 included only domain PD020519 and spanned 110

residues. Further single-domain hit sequences were the uncharacterized proteins UP Q9V1Y2 from the thermoautotrophican archaeon *P. abyssi* (105 residues; hereafter proabylysin) and UP O66407 from the bacterial thermophile *A. aeolicus* (110 residues). In the current release of ProDom (2010.1; http://prodom.prabi.fr), former family PD020519 has expanded dramatically, which has led to the three archaeal sequences being currently grouped into a separate family, PDB0S8B5, which also includes three additional sequences (http://prodom.prabi.fr/prodom/current/cgi-bin/request.pl?db_ent1=PDB0S8B5&wanted=align&SSID=363090958_11162).

The MEROPS database (release 9.7, as of 11/2012) was downloaded and used for a sequence-based search using PSI-BLAST. Starting with the three aforementioned sequences, the top 20 hits were selected (with PSI-BLAST E-values ≤ 0.005) for each search, resulting in a total of 69 unique homologous sequences. A multiple sequence alignment of all the sequences selected was obtained using the PHYLEMON2 server (46). A final consensus multiple sequence alignment was trimmed using the TRIMAl method (47) and subsequently used for searching the most likely model following the PROTEST methodology (48). BLOSUM62 was the used as input model by PHYML (49) to build a phylogenetic tree, which was then represented using the FIGTREE program (http://tree.bio.ed.ac.uk/software/figtree/). Sequence-based predictions of regular secondary structure elements were performed with JPRED3 (http://www.compbio.dundee.ac.uk/jpred; (50)).

Protein production and purification — Synthetic genes coding for the *A. aeolicus* protein (which contained a lysine-to-alanine mutation at position 2 due to the cloning strategy), proabylysin (lysine-to-valine mutation at position 2), and projannalysin (lysine-to-alanine mutation at position 2) were purchased from GeneArt (Invitrogen) and cloned into vectors pCRI7 and pCRI8 (both modified from pET-28a (Novagen) which confers resistance to kanamycin) using *Nco*I and *Xho*I restriction sites. While pCRI7 does not append any extra residue to the protein of interest, vector pCRI8 attaches an N-terminal hexahistidine-tag followed by a tobacco-etch-virus (TEV) protease recognition site. C-terminal deletion mutants lacking one to five residues of proabylysin were cloned into vector pCRI7, and a C-terminal deletion mutant lacking five residues of projannalysin was cloned into vector pCRI8. In addition, the genes coding for proabylysin and its aforementioned C-terminal deletion mutants were cloned into vector pCRI6 (derived from plasmid pETM30, which confers resistance to kanamycin), which attaches an N-terminal hexahistidine-tag followed by glutathione-S-transferase (GST) and a TEV protease recognition site, and into a variant of vector pCRI7, into which an extra N-terminal streptavidin-tag was introduced by PCR (hereafter referred to as vector pCRI7s). All constructs were verified by DNA sequencing. Proteins were produced by heterologous overexpression in *E. coli* BL21 (DE3) cells, which were grown at 37°C in Luria Bertani medium supplemented with 30µg/ml kanamycin. Cultures were induced at an OD_{600} of 0.8 with 0.2-1mM isopropyl-β-D-thiogalactopyranoside and kept growing either for 5h at 37°C or overnight at 18°C.

The purification of proabylysin (cloned into the pCRI7 vector) was performed as follows: After centrifugation at 7,000x*g* for 30min at 4°C, the cell pellet was washed twice with buffer A (50mM Tris-HCl, 500mM NaCl, pH8.0), and resuspended in the same buffer supplemented with EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics) and DNase I (Roche Diagnostics). Cells were lysed at 4ºC using a cell disrupter (Constant Systems) at a pressure of 1.35Kbar, and the cell debris was removed by centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was incubated for 1h at 85°C and the precipitated material removed by further centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was filtered (0.22µm pore size; Millipore) and dialyzed overnight at room temperature against buffer B (20mM Tris-HCl, pH8.0). The protein was subsequently purified by anion exchange chromatography using a HiLoad 16/10 Q Sepharose HP column followed by a MonoQ 4.6/100PE column (both from GE Healthcare), and finally polished with a HiLoad 16/60 Superdex 75 size-exclusion-chromatography column (GE Healthcare) previously equilibrated with buffer C (20mM Tris-HCl, 150mM NaCl, pH7.4). The purification of the proabylysin C-terminal deletion mutants was carried out in the same way except that the supernatant was incubated for 1h at 50-65°C instead of 85ºC. The purification of proabylysin and its mutants with an N-terminal GST fusion protein or a streptavidin-tag (encoded by vectors pCRI6 or pCRI7s vector, respectively, see above) was also performed in the same way except that the heat-shock step was replaced by affinity chromatography purification on a GSTrap HP (GE Healthcare) or Strep-Tactin Sepharose (IBA) column, respectively, according to the manufacturers' instructions.

The purification of projannalysin and its C-terminal deletion mutant (both cloned into the pCRI8 vector) was performed as follows: After centrifugation at 7,000x*g* for 30min at 4°C, the pellet was washed twice with buffer A, resuspended in the same buffer plus 10mM imidazole, and supplemented with EDTA-free protease inhibitor cocktail tablets and DNase I (both Roche Diagnostics). Cells were lysed using a cell disrupter as described above, and the cell debris was removed by centrifugation at 50,000x*g* for 1h at 4°C. The supernatant was filtered, incubated with nickel-nitrilotriacetic acid resin (Invitrogen) previously equilibrated with buffer A plus 10mM imidazole, and the fusion protein was eluted using buffer A plus 350mM imidazole. The sample was then dialyzed overnight at 20°C against buffer D (50mM Tris-HCl, 250mM NaCl, 1mM dithiothreitol, pH8.0) in the presence of hexahistidine-tagged TEV protease at an enzyme:substrate ratio of 1:50 (w/w). Cleavage left the dipeptide glycine-proline at the N-terminus of both proteins. Digested samples were passed several times through nickel-nitrilotriacetic acid resin previously equilibrated with buffer A plus 10mM imidazole to remove non-cleaved hexahistidine-containing molecules. Flow-throughs were collected, concentrated by ultrafiltration, and further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer C.

Protein identity and purity were assessed by mass spectrometry and 15% Tricine-SDS-PAGE stained with Coomassie blue. Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices of 5-KDa cut-off (Sartorius Stedim Biotech). Protein concentration was determined by measuring the absorbance at 280nm using a spectrophotometer (NanoDrop) and calculated absorption coefficients $E_{0.1\%}$ of 0.98 and 0.34 for proabylysin and projannalysin, respectively. N-terminal sequencing through Edman degradation, peptide mass fingerprinting of tryptic protein digests, and mass spectrometry analyses were carried out at the proteomics facilities of the Centro de Investigaciones Biológicas (http://www.cib.csic.es/en/servicio.php?iddepartamento=27) and Vall d'Hebron Institute of Oncology (http://www.vhio.net/research/proteomics/en_index.html). Melting temperatures (T_m) through differential scanning fluorimetry (thermofluor) were determined by using Sypro Orange dye (Life technologies, Invitrogen) and an iCycler iQ Real Time PCR Detection system (Bio-Rad) as published previously (51,52). Limited-proteolysis trials to remove C-terminal peptides from full-length proabylysin and projannalysin were undertaken in buffer C by overnight incubation at 37°C or 80°C with trypsin, subtilisin, thermolysin (all from Sigma), or ulilysin (produced in-house according to (53)) at different peptidase:substrate ratios (1:1, 1:5, 1:20, 1:50, 1:100, 1:200, 1:500 (w/w)), and monitored by mass spectrometry.

Proteolytic and inhibitory activity assays — Proteolytic activities were assayed at 37°C in buffer E (50mM MES, 150mM NaCl, pH 5.5), buffer F (50mM HEPES, 150mM NaCl, pH 7.5) or buffer G (50mM CHES, 150mM NaCl, pH 9.5) at a final protein concentration of 50µg/ml for proabylysin and projannalysin, unless otherwise stated.

Proteolytic activity against the fluorescein conjugates BODIPY FL casein, DQ gelatin, and DQ BSA (all from Invitrogen) was tested according to the manufacturer's instructions using a microplate fluorimeter (FLx800, BioTek or Infinite M200, Tecan). Assays with natural protein substrates (at 0.25 or 0.50mg/ml) included bovine plasma fibronectin, bovine muscle actin, human plasma fibrinogen, cold-water fish-skin gelatin, bovine milk casein, and bovine milk α-casein (all from Sigma). Reactions were carried out in buffer F at 37°C, 65°C, and 80°C overnight and at an enzyme:substrate ratio of 1/5 (w/w) for the first two substrates and 1/10 (w/w) for the others. Cleavage was assessed by 15% Tricine-SDS-PAGE stained with Coomassie blue. Proteolytic activity was further tested on eleven fluorogenic substrates of sequence: Abz-K-D-E-S-Y-R-K(dnp) (Abz, aminobenzoyl; dnp, 2,4-dinitrophenylamino), Abz-T-V-L-E-R-S-K(dnp), Abz-D-Y-V-A-S-E-K(dnp), Abz-Y-G-K-R-V-F-K(dnp), Abz-V-K-F-Y-D-I-K(dnp), Dabcyl-L-A-R-V-E-Edans (Dabcyl, *p*-dimethyl(aminophenyl)azobenzoate; Edans, 2-aminoethylamino-1-naphthalene sulfonate), Abz-G-I-V-R-A-K(dnp) (Bachem), Mca-P-L-G-L-Dap(dnp)-A-R-NH_2 (Mca, 7-methoxycoumarin-4-acetyl; Dap, L-diaminopropionyl) (Bachem), Mca-R-P-K-R-V-E-Nva-W-R-K(dnp)-NH_2 (Nva, norvaline) (Bachem), and Dnp-P-L-G-L-W-A-(*D*)R-NH_2 (Bachem) (see Ref. (54) for details on the first six substrates). Reactions were performed at enzyme:substrate molar ratios of 1/0.25, 1/0.5, 1/1.25, and 1/6.25 and monitored in a microplate fluorimeter (Infinite M200, Tecan). In addition, proteolytic assays with DQ BSA and the fluorogenic peptides were performed at enzyme:substrate molar ratios of 1/0.1 and 1/0,5, respectively, as described above at high temperatures (up to 80°C), or in the presence of 0-4M urea or 0-4M guanidinium hydrochloride.

Esterase activity against the chromogenic substrates A-ONp (ONp, *p*-nitrophenyl ester), H-ONp, and L-ONp (all from Bachem), and carboxypeptidase activity against the chromogenic substrates *N*-(3-[2-furyl]acryloyl)-F-F-OH (Bachem), *N*-(3-[2-furyl]acryloyl)-G-L-A-OH (Bachem), and *N*-(3-[2-furyl]acryloyl)-L-G-P-A-OH (Sigma) were tested using a microplate spectrophotometer (PowerWave XS, BioTek) at enzyme:substrate molar ratios of 1/25 and 1/125 for ester substrates and 1/25 for carboxypeptidase substrates. Finally, aminopeptidase activity was assayed with fluorogenic substrates F-Amc (Amc, 7-amino-4-methylcoumarin), (*H*)T-Amc, and Y-Amc (all from Bachem), and with the chromogenic *p*-nitroanilide (pNA) derivatives of a representative set of natural L-amino acids and peptides (from Bachem): A-pNA, M-pNA, L-pNA, K-pNA, V-pNA, (*H*)I-pNA, (*H*)G-pNA, N-Acetyl-F-pNA, A-A-pNA, *N*-benzyloxycarbonyl-V-G-R-pNA, *N*(α)-benzoyl-I-E-G-R-pNA, and *N*-succinyl-A-A-P-F-pNA. Reactions were monitored in a microplate fluorimeter (FLx800, BioTek) or PowerWave XS spectrophotometer (BioTek) at enzyme:substrate molar ratios of 1/125 and 1/250.

Crystallization and X-ray diffraction data collection — Crystallization assays were carried out by the sitting-drop vapor diffusion method using 96x2-well MRC plates (Innovadyne) and a Cartesian (Genomic Solutions) or a Phoenix (Art Robbins) nanodrop robot at the IBMB/IRB High-Throughput Automated Crystallography Platform. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4ºC and 20°C. Successful hits were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research). The best crystals of proabylysin appeared at 20°C from equivolumetric drops containing protein solution (2mg/ml in buffer C) and 100mM MES, 200mM sodium/potassium phosphate, 2.6M sodium chloride, pH6.5 as reservoir solution. The best crystals of projannalysin were obtained at 4°C from drops containing protein solution (20mg/ml in buffer C) and 100mM Tris-HCl, 200mM lithium sulfate monohydrate, 25% (w/v) polyethylene glycol 3,350, pH8.5 as reservoir solution at a 2:1 ratio. Crystals were cryo-protected with successive passages through reservoir solution containing increasing concentrations of glycerol—up to 10 and 20% (v/v) for projannalysin and proabylysin, respectively. A heavy-ion derivative of projannalysin was prepared by soaking native crystals for 5min in reservoir solution containing 5mM mercury(II) acetate.

Complete diffraction datasets were collected at 100K from liquid-N_2 flash-cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6M pixel and an ADSC Q315R CCD detector at beam lines ID29 (proabylysin) and ID23-1 (native and derivative projannalysin), respectively, at the European Synchrotron

Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona." Crystals of proabylysin were primitive orthorhombic, with one molecule per asymmetric unit, whereas crystals of projannalysin were trigonal with two molecules per asymmetric unit. Diffraction data were integrated, scaled, merged, and reduced with programs XDS (55), and XSCALE (56) or SCALA (57), the latter within the CCP4 suite of programs (58)(see Table 1).

Structure solution and refinement — The structure of proabylysin was solved by single-wavelength anomalous diffraction by using the diffraction data to 1.6Å resolution of a crystal collected at the zinc K-edge peak wavelength, as inferred from a previous XANES fluorescence scan, which enabled program SHELXD (59,60) to identify the zinc ion present in the asymmetric unit. Subsequent phasing with this site using program SHELXE (60,61) and higher-resolved diffraction data to 1.15Å resolution collected at a remote wavelength as a pseudo-native dataset yielded a suitable electron density map for chain tracing on a Silicon Graphics Octane2 Workstation using the TURBO-FRODO program (62) and on a Linux-operated PC using the COOT program (63,64). Model building alternated with crystallographic refinement with program BUSTER (65), which included TLS refinement, until completion of the model. The final refinement was performed with SHELXL (61) and included anisotropic refinement of atomic displacement parameters. It comprised residues M^{1}-V^{105} (superscripted residue numbering for proabylysin), one zinc ion, one glycerol molecule, and 147 solvent molecules (see Table 1).

The structure of projannalysin was solved by single-wavelength anomalous diffraction by using the diffraction data to 2.4Å resolution of a crystal collected at the mercury L_{III}-edge peak wavelength, as determined from a previous XANES fluorescence scan, which enabled program SHELXD to find six mercury sites. Visual inspection of these sites on a graphic display with TURBO-FRODO enabled the identification of two clusters of three sites, which allowed the associated dimeric non-crystallographic symmetry operator to be derived. Subsequent phasing with these sites using program SHELXE and higher-resolved native diffraction data to 2.0Å resolution, followed by density modification with two-fold averaging with the DM program (66), yielded an electron density map suitable for model building, which proceeded as for proabylysin. Crystallographic refinement was performed with BUSTER. The final model of projannalysin comprised residues G^{-2}-I^{110} (superscripted residue numbering in italics) of molecule A and thus included the N-terminal residues G^{-2} and P^{-1}, which precede M^{1} as a result of the cloning strategy and TEV cleavage (see above). In addition, residues K^{7}-I^{110} of molecule B plus seven glycerols, one Tris and two zinc cations, four sulfate and three chloride anions, and 194 solvent molecules completed the model of projannalysin (see Table 1).

Figures were prepared with the CHIMERA program (67). Interaction surfaces (taken as the total surface area buried in a complex) were calculated with CNS taking a probe radius of 1.4Å (68). Structural superpositions were performed with the SSM routine (69) within COOT. Surface complementarity was computed with SC (70) within CCP4 with default parameters. Model validation was performed with MOLPROBITY (71) and the WHATCHECK routine of WHATIF (72). The final coordinates of proabylysin and projannalysin have been deposited with the PDB (access codes 4JIU and 4JIX, respectively).

Structure-similarity analyses — The MAMMOTH program (73) was used with default parameters to search for proabylysin and projannalysin homologous structures in the PDB (74) as of 11/2012. Structural similarity searches were also carried out with the DALI program (http://ekhidna.biocenter.helsinki.fi/dali_server/start; (75)).

RESULTS AND DISCUSSION

Bioinformatics searches reveal minimal putative metallopeptidases — We performed bioinformatics searches to identify orthologs of M48 and M56 IMMPs containing only the CD by using the sequence stretch of BlaR1/MecR1 from *S. aureus* (family M56) contained between two predicted TMHs and encompassing the HEXXH zinc-binding motif (see "Experimental procedures" and Fig. 1). Three top hits were found, which were uncharacterized proteins from the thermophiles *A. aeolicus*, *P. abyssi* and *M. jannaschii*. They contained the zinc-binding motif and were closely related, aligning with 28-31% pairwise sequence identity (Fig. 2a). With a size of 105-110 residues and the absence of predicted signal peptides and transmembrane segments, they would represent a "minimal" scaffold for a soluble MP. A profile-based sequence alignment against the MEROPS database revealed that these sequences were evolutionarily related to M56-IMMPs and also close to M48-IMMPs (Fig. 2b).

Proteolytic assays suggest zymogenic forms — The three thermophilic proteins were recombinantly overexpressed in *E. coli*. The proteins from *P. abyssi* (proabylysin) and *M. jannaschii* (projannalysin) proved to be soluble and were readily purified to homogeneity. Conversely, the ortholog from *A. aeolicus* was insoluble under all expression conditions tested and no suitable refolding strategies could be found, and so was discarded. Full-length proabylysin and projannalysin were correctly folded as assessed by size-exclusion chromatography, which unambiguously revealed that they eluted as a monomer and dimer, respectively, at protein concentrations ranging from 0.1 to 2mg/ml for the former and 0.1 to 40mg/ml for the latter. Notably, proabylysin was stable for long periods only at a maximum concentration of ~0.6mg/ml at 20°C. Both proteins were inactive against a large battery of protein and peptide substrates (see "Experimental procedures"). As *P. abyssi* and *M. jannaschii* are hyperthermophilic organisms whose growing temperature and pressure can reach 95°C and 200 atmospheres, respectively, we also performed activity assays with these two proteins at high temperature (up to 80°C) and/or in the presence of chaotropic agents (urea and guanidinium hydrochloride) to mimic the physiological conditions. However, we did not observe any activity for these proteins under these conditions either (data not shown), strongly suggesting that they were zymogens, as indicated by structural studies (see following sections).

The structure of proabylysin indicates a novel intra-molecular mechanism of latency – The crystal structure of proabylysin was determined by single-wavelength anomalous diffraction of the intrinsic zinc and refined with diffraction data to 1.15Å resolution (see "Experimental procedures"). The structure shows a compact globular, almost spherical shape ~35Å in diameter (Fig. 3a). A shallow active-site cleft is carved into the front surface of the molecule when viewed in standard orientation (44) and divides it into an upper, N-terminal sub-domain (NTS; M^{1}-E^{68}; superscripted proabylysin residue numbering) and a lower, C-terminal sub-domain (CTS; T^{69}-V^{105}). The NTS starts on the right top rear of the molecule in the form of an α-helix, hereafter referred to as the "backing helix" α1 (for nomenclature and extension of regular secondary structure elements, see Fig. 2a), which runs diagonally downwards and ends in a loop connecting α1 with the first strand β1 (Lα1β1) of a front, twisted three-stranded β-sheet (β1-β3). The top strands (β1 and β3) are parallel and the lowermost (β2), which creates an "upper rim" of the active-site cleft, is antiparallel to the cleft and, thus, to any substrate bound to it. Preceding the latter strand is the "bulge-edge segment" (L^{26}-I^{30}), which contributes to shaping the top of the cleft on its primed side (for cleft and substrate sub-site nomenclature, see (44,76)). After β3, a short 3_{10}-helix (η1; the "linking helix") at the right top of the molecule links the sheet with the "active-site helix" α2, which contains the short zinc-binding consensus sequence H^{60}EXXH^{64} (see below). The helices and the sheet of the NTS contribute through their inner surfaces to an internal hydrophobic core, which traverses the entire protein moiety and glues the structure together. After α2, the polypeptide chain enters the CTS through Lα2α3, which leads to the "glutamate helix" α3. The latter is termed thus due to its topological equivalence with a similar helix in gluzincins (see below), and it lies in a horizontal plane that is roughly parallel to that of helix α2 but with the helix axis vertically rotated backwards by ~50º, so that α3 ends with V^{83} at the back molecular surface. At F^{84}, the chain turns frontwards and, after a loop that forms the bottom of the molecule (P^{85}-R^{86}), opens out into the "C-terminal helix" α4, which runs obliquely and creates the lower right part of the molecule (Fig. 3a). The last residue of α4, L^{100}, resides in a hydrophobic pocket created by the side chains of A^{45}, I^{30}, I^{57} and I^{96}, which feature the rightmost end of the internal hydrophobic core of the NTS (Fig. 3a). From Q^{101} onwards, the polypeptide chain enters the primed side of the active-site cleft in the extended conformation of—but reverse orientation to—a true peptidic substrate and reaches the zinc site with the C-terminal carboxylate of V^{105}. This C-terminal tail establishes inter-main chain hydrogen bonds with residues from the bulge-edge segment and the start of upper-rim strand β2 (N^{103} O – N^{30} N, 2.87Å; V^{105} N – A^{31} O, 2.92Å).

The catalytic zinc ion resides at the bottom of the active-site crevice, somewhat displaced from half way towards the non-primed side of the cleft (Fig. 3a). The metal is coordinated in a tetrahedral fashion by the Nε2 atoms of the two histidines of the consensus sequence, H^{60} (2.02Å apart) and H^{64} (2.00Å), one of the two C-terminal carboxylate oxygens of V^{105} (2.02Å), and the Nε2 atom of H^{72} (2.02Å apart), which is provided by Lα3α4 two positions ahead of the start of the glutamate helix α4. The other carboxylate oxygen of V^{105} is further way from the metal (2.69Å) and in a geometrically unfavorable conformation for binding (Fig. 3a). By contrast, this oxygen contacts in a distorted bidentate manner the two carboxylate oxygens of the general base/acid glutamate E^{61} (2.68Å and 3.28Å apart), thus indicating that either of the carboxylate groups must be in protonated state.

The insertion of the C-terminal tail also provides information on the distinct sub-sites on the primed side of the cleft. M^{104} nestles into what is most likely the **S_1'** sub-site of the cleft (cleft and substrate sub-sites in bold), which is usually the major determinant for specificity in MPs (44). The methionine side chain is present in double conformation and the surrounding residues give rise to a site that is wide and hydrophobic, thus suggesting that mature abylysin could easily accommodate bulky hydrophobic residues in **P_1'**, as observed in the otherwise unrelated matrix metalloproteinases, snapalysins, and ADAMs/adamalysins, which are metzincins (77-79). The **S_1'** pocket is framed at its back by residues from the start of the active-site helix (Y^{56}, I^{57}, and H^{60}), and at its right and bottom by residues from the C-terminal helix (E^{93} and I^{96}). The antepenultimate residue of the tail, N^{103}, possibly at **P_2'**, points towards the bulk solvent, suggesting that this site is not relevant for specificity as long as the extended conformation of the substrate is maintained. R^{102}, in turn, probably occupies **S_3'**, which would be shaped by the side chains of I^{30}, A^{45}, and L^{100}, and the main chain of the bulge-edge segment at P^{25}-I^{30}. The guanidinium group of R^{102} establishes two hydrogen bonds with the main-chain carbonyl oxygens of L^{26} and S^{29} (2.50Å and 2.92Å apart, respectively), and a double salt bridge with D^{43} at the end of Lβ3η1. The R^{102} Nη1 and Nη2 atoms are, respectively, 3.01Å and 3.02Å apart from the carboxylate oxygens of the aspartate. As found for M^{104} in **S_1'**, R^{102} fits neatly into the potential **S_3'** site, suggesting that mature abylysin could have a preference for arginine in **P_3'**. A look at the putative non-primed side of the cleft, in turn, suggests that V^{105} is in **S_1**. It has its side chain pointing towards the bulk solvent, thus suggesting that any residue of a potential substrate could be found in **P_1**. Upstream of **P_1**, a substrate could be accommodated in a rather shallow **S_2** or **S_3** site, putatively framed by the side chains of V^{33}, F^{35}, H^{64}, L^{65} and S^{70} (Fig. 3a).

Overall, the structure of the globular part of proabylysin spanning M^{1}-L^{100} indicates a competent MP moiety, active-site cleft, and metal-binding site—including the pivotal glutamate general base/acid required for catalysis—as found in active enzymes of this class (see also below). By contrast, the insertion of the C-terminal tail into the active-site cleft in the reverse orientation of a substrate, thus blocking structural elements essential for catalysis and access of true substrates, provides a structural explanation for the lack of hydrolytic activity of the protein as it would actually correspond to a zymogen. The latency mode has been hitherto unseen for structurally-characterized MPs, which are normally kept latent by N-terminal pro-domains or pro-peptides, as in astacin (80), meprin β (81), fragilysin (52), funnelin carboxypeptidases (82), matrix metalloproteinases (77), ADAMs/adamalysins (78), and thermolysin (83). Accordingly, the present

structure represents a novel mechanism of intra-molecular latency maintenance in MPs.

The structure of projannalysin suggests a novel inter-molecular mechanism of latency — The crystal structure of projannalysin was determined by single-wavelength anomalous diffraction of a mercury derivative and refined with diffraction data to 2.0Å resolution (see "Experimental procedures"). In one of the two molecules found in the asymmetric unit of the crystal—molecule A—the N-terminal stretch of projannalysin protrudes away from the molecular moiety and is defined from the first residue of the construct, G^{-2} (superscripted projannalysin residue numbering in italics), to N^{4}, due to an interaction with a crystallographic neighbor (Fig. 3b, left, in tan). The final electron density is somewhat weaker at E^{5} and N^{6}, but becomes fully defined again from K^{7} onwards, which is the first residue of the globular moiety. By contrast, in molecule B the polypeptide chain is defined from K^{7} onwards only (Fig. 3b, right, in turquoise).

In accordance with a 31% sequence identity (see Fig. 2a), projannalysin and proabylysin exhibit a very similar overall shape and practically overlapping chains from the beginning of helix α1 to the end of helix α4 (M^{1}-L^{100} of proabylysin and D^{9}-L^{109} of projannalysin; Z-value according to the DALI program of 16.0; *rmsd* of 1.5Å). All the regular secondary structure elements are conserved with the sole exception of proabylysin 3_{10}-helix η1, which has an extra residue in projannalysin and becomes here an α-helix (termed α1', K^{52}-N^{57}; see Figs. 2a and 3b,c). After L^{109}, the last residue of projannalysin projects away from the molecular body, pointing in the opposite direction to the proabylysin C-terminal tail and becoming engaged in binding of a neighboring molecule (see below and Fig. 3b,c).

Consistent with the dimeric behavior observed in size-exclusion chromatography, projannalysin is also a dimer in the crystal structure (Fig. 3b, right). Two molecules interact through an interaction surface of ~1,980Å^{2}, which is above the range generally described for protein-protein complexes (1,250-1,750Å^{2}; (84)), and indicates that such a dimer is likely to occur *in vivo*. The interaction surface shows a surface complementarity (Sc=0.67) that lies within the range reported for antibody/antigen interfaces (0.64-0.68; (70)). The interaction is nearly symmetric and results from 32 inter-molecular contacts (<4Å), among them nine hydrogen bonds and hydrophobic interactions between seven molecule A and six molecule B residues (Fig. 3d). Participating segments include Lβ1β2, Lβ3α1', α2, Lα2α3, and α4 of either molecule (Fig. 3d), and the catalytic metal ions are 24Å apart. Most noteworthy is the crosswise interaction between the active-site of one molecule and the C-terminal helix α4 of the other (Fig. 3b,c). The latter blocks access to the cleft and provides, thus, the structural explanation for the lack of hydrolytic activity of projannalysin. Like proabylysin, it is a zymogen but by contrast latency is exerted crosswise inter-molecularly by the swap of a C-terminal α-helix. This also entails significant differences in the zinc-binding site. In projannalysin, the metal is surrounded by a distorted octahedral hexa-coordination sphere, which is rather unusual for zinc according to MESPEUS database (http://mespeus.bch.ed.ac.uk/MESPEUS_10; see Table 2 in (85)). Protein ligands equivalent to proabylysin are the Nε2 atoms of H^{69} (2.1/2.0Å apart from the metal in molecules A/B, respectively) and H^{73} (2.1/2.1Å) of the zinc-binding sequence, and H^{81} of Lα2α3 (2.1/2.1Å). In addition, the metal is distinctly bound by a solvent molecule (2.2/2.1Å apart) and, in molecule A, a Tris molecule through its amino (2.2Å) and one of its hydroxyl groups (1.9Å) in a bidentate fashion. In molecule B, this ligand is replaced by a glycerol, two of whose hydroxyl oxygens are 2.1 and 2.2Å apart from the metal, respectively. The solvent molecule and H^{73} Nε2 are in apical positions and the other four ligands in a plane with the metal (Fig. 3c). Accordingly, and in contrast to proabylysin, the two carboxylate oxygens of the C-terminus (here, I^{110} of the neighboring molecule) are far from the zinc ion (4.1-4.2Å). Instead, they contact the carboxylate of the general base/acid glutamate, E^{70} (2.6/2.5Å in molecule A/B, respectively) *plus* the zinc-binding hydroxyl of the Tris (glycerol) molecule (both 2.6Å apart), and the zinc-bound water (2.7/2.6Å), respectively. This architecture indicates that either of the carboxylate groups of E^{70} or L^{109} must be in protonated state, as found in proabylysin (see above). The structure of the dimeric zymogenic complex further reveals that the **S_1'** pocket is partially occupied by the side chain of the last residue, I^{110}, instead of the penultimate residue as seen in proabylysin. The pocket is very similar in depth and hydrophobicity in both zymogens and framed in projannalysin by A^{39} (A^{31} in proabylysin), I^{38} (I^{30}), I^{66} (I^{58}), Y^{65} (Y^{56}), I^{105} (I^{96}), L^{109} (L^{110}), and E^{102} (E^{93}).

As for proabylysin, the globular part of projannalysin spanning D^{9}-L^{109} indicates a competent fold of catalytic moiety, active-site cleft, metal-binding site, and general base/acid glutamate, as usual for active MPs (see also below). By contrast, the insertion of the C-terminal part of helix α4 into the active-site of a neighbor contributes to another novel C-terminal latency mechanism for this type of enzymes, which is homodimeric and occurs crosswise.

Activation trials point to an intricate mechanism of peptidase activation — To provide insight into the mechanism of activation of these MPs, we attempted to subject them to limited proteolysis to remove the zymogenic segments by assaying them with a collection of standard serine- and metalloproteinases, which only yielded intact proteins or resulted in complete degradation (data not shown). We thus designed a range of constructs lacking C-terminal residues of proabylysin and projannalysin. Proabylysin constructs in which the last five residues were sequentially deleted showed dramatically decreasing expression and solubility levels when compared with the wild-type protein, thus precluding purification and/or functional assays for most of the constructs. Only those lacking the last one and two residues, respectively, could be purified. However, they showed a significantly decreased melting temperature (T_m) in a thermofluor assay (85 and 65°C, respectively; see "Experimental procedures") when compared with that of the wild type (95°C), which is indicative of structurally labile and inactive forms. Assays of fusion constructs of C-terminal truncations with N-terminal GST or streptavidin-tag, which

rendered only soluble aggregates as assessed by size-exclusion chromatography (data not shown) were also unsuccessful. Similar negative results were obtained with a projannalysin mutant lacking the last five residues (data not shown), which theoretically should avoid inhibition but not affect the overall protein moiety as inferred from the crystal structure (see above). In conclusion, we failed to obtain active abylysin and jannalysin. This strongly suggests that both proteins likely need the assistance of a specific protease for activation and/or a conformational change that is brought about in the intracellular living environment but that we were unable to reproduce *in vitro*. Nevertheless, the possibility that these proteins are inactive ancestors of metallopeptidases cannot be ruled out.

Proabylysin and projannalysin are structurally related to gluzincins — Thermolysin from *Bacillus thermoproteolyticus* was the first metalloendopeptidase structurally characterized, back in 1972 (86), and is the founding member of the gluzincins ((41,42); see also Table 2 in (44)). Superposition of both proabylysin and projannalysin onto thermolysin reveals that the polypeptide chains are very similar along almost the entire structure of the former (M^1-K^{99}) and V71-P195 of the latter (Fig. 4a,b), including the position of the N- and C-termini of the superposed parts and all regular secondary structure elements, with the exception of the linking and C-terminal helices, which are replaced by loops in the *Bacillus* enzyme. This similarity is reflected by a DALI Z-value of 3.5, an *rmsd* of 2.9Å, 83 topologically aligned residues but only 8% sequence identity. In particular, the active-site and glutamate helices overlap despite being, respectively, one and two turns shorter in the present structures (Fig. 4a; see also preceding sections). Overall, these two helices are the major characteristic features of gluzincins and provide the three protein zinc ligands—usually two histidines and a glutamate (H142, H146, and E166 in thermolysin). This shortening of the present structures entails that the third protein zinc ligand—a histidine—is found in the loop preceding the glutamate helix. Nevertheless, the position of this third ligand is practically the same and the Cα atoms of proabylysin H^{72} and thermolysin E166 are only 1.4Å apart (Fig. 4b). In addition, the three-stranded β-sheets and the backing helices fit well on top of each other in both structures, both in sequence, position and orientation, although they are linked by longer loops in thermolysin (Fig. 4a). A further feature of gluzincins is the "Ser/Gly-turn" within the glutamate helix, downstream of the metal-binding glutamate, which contains a residue—usually a serine or a glycine—whose side chain is found immediately below the metal-binding site (82). This structural element is reminiscent of the "Met-turn", which has a strictly conserved methionine and is characteristic of metzincins (51). Such a Ser/Gly-turn is also present in proabylysin and projannalysin, centered on a phenylalanine (F^{76} and F^{85}, respectively), which is also found in the *A. aeolicus* protein (Fig. 2a). In general, structurally characterized gluzincins are much larger than the minimal core found in proabylysin and projannalysin: thermolysin spans 316 residues and cowrin metallocarboxypeptidases, which also show the aforementioned structural similarity, span ~500-700 (82,86). In particular, both thermolysins and cowrins possess a downstream "tyrosine helix" below the glutamate helix (82).

Another structurally characterized gluzincin family consists of leukotriene A4 hydrolase (PDB 3B7S; (87)) and related metalloaminopeptidases such as cold-active aminopeptidase (PDB 3CIA; (88)), endoplasmic reticulum aminopeptidase 1(PDB 3QNF, 3MDJ; (89,90)), tricorn interacting factor F3 (PDB 1Z5H; (91)), M1 alanyl-aminopeptidase (PDB 3EBI; (92)), and aminopeptidase N (PDB 2HPO; (93)). Like cowrins (82), these are large multi-domain enzymes spanning ~600-950 residues that contain an inserted CD bearing a greater resemblance to proabylysin and projannalysin than that of thermolysin (DALI Z-values between 7.2 and 6.4). They even include the linking and C-terminal helices, although the latter is rotated frontwards by ~45-90 degrees. This family has a glutamate as the third zinc ligand and a threonine or alanine in the Ser/Gly-turn.

Overall, the close general similarity of proabylysin and projannalysin to distinct active gluzincin families strongly suggests that the globular parts of these two zymogens are in an overall competent conformation (see also preceding sections) and that latency is exerted through C-terminal segments by sterically blocking access to preformed CDs. Accordingly, the thermophylic proteins studied here would represent a minimal structural and functional core of the gluzincin clan, and, as such, we propose the family name "minigluzincins" for them. The fact that minigluzincins have a histidine instead of a glutamate as third zinc ligands and a phenylalanine in the Ser/Gly-turn reflects variability among gluzincins that is also observed in metzincins, which generally have a histidine as the third zinc ligand, although snapalysins and, possibly, thuringilysins, have an aspartate (94).

Minigluzincins are models for catalytic domains of M48 and M56 metallopeptidases — Despite the obvious structural similarity with the aforementioned gluzincins, automated structure-similarity searches of proabylysin and projannalysin against the PDB identified three other, more closely related structures: a putative peptidase from *Geobacter sulfurreducens* (PDB 3C37; DALI Z-score 7.4; *rmsd* 3.1Å; 98 aligned residues; 13% sequence identity); a fragment of HtpX from *Vibrio parahaemolyticus* (PDB 3CQB; DALI Z-score 6.8; *rmsd* 2.1Å; 71 aligned residues; 15% sequence identity); and human FACE1 (PDB 4AW6; (40); DALI Z-Score 4.6; *rmsd* 2.8Å; 96 aligned residues; 18% sequence identity; see Fig. 4c,d). The latter is very similar to its yeast ortholog Ste24p (PDB 4IL3; (39)). PDB entry 3C37 was deposited with the PDB by a structural genomics consortium but not published. It corresponds to a soluble 253-residue protein tagged by the depositors as an M48-family peptidase, although it is not an IMMP. It is a non-functional protein whose putative catalytic metal ion is bound in a canonical manner by three residues provided by an active-site and a glutamate helix, and also by a fourth histidine ligand coming from a second protein chain, so that the cleft is blocked. In the absence of further data and a publication, crystallization artifacts cannot be ruled out in this case. PDB entry 3CQB, in turn, is an artificial fragment between two TMHs of a *Vibrio* relative of M48-family member HtpX from *E.coli*, which was likewise

deposited by a structural genomics consortium and not published. It lacks a functional active-site cleft and metal-binding site, and could potentially correspond to the NTS of the CD of this M48 family member. Finally, PDB entries 4AW6 and 4IL3 correspond to human and yeast FACE1/Ste34p, the only two true IMMPs currently available for their 3D structures in addition to S2P MP from *M. jannaschii* (36). The latter belongs to family M50, is not a gluzincin (see Fig. 2 in (44)), and is only very distantly related to proabylysin (PDB 3B4R; DALI Z-Score 1.2; *rmsd* 3.5; 51 aligned residues; 10% sequence identity) and its structural relatives. Human and yeast FACE1/Ste24p represent the only M48-IMMPs structurally characterized to date. Further entries that are similar to minigluzincins are two closely related hypothetic proteins reported by structural genomics consortia: ybeY from *E. coli* (PDB 1XM5; (95)) and AQ-1354 from *A. aeolicus* (PDB 1OZ9; (96)), for which no function has been characterized. These are single-domain proteins of ~150 residues that contain a third metal-binding histidine and a methionine and phenylalanine, respectively, in the Ser/Gly-turn. They possess a linking helix but, despite rather large DALI Z-scores when compared with proabylysin (6.9 and 7.6, respectively), they lack the C-terminal helix of minigluzincins and have additional regular secondary structure elements inserted into the basic scaffold.

Superposition of proabylysin and projannnalysin onto the putative peptidase of *G. sulfurreducens,* human FACE1, and truncated HtpX, reveals that all the common regular secondary structure elements shared by minigluzincins with thermolysin and other gluzincins are also found in the aforementioned M48-related structures. Taken together, this ascribes M48-IMMPs to the gluzincin clan of MPs. In addition, the closer structural similarity of minigluzincins to the aforementioned three M48-like structures than to other gluzincins (see Z-scores above) is reflected by the fact that the linking helix (in PDB 3C37, 3CQB, and 4AW6) and the C-terminal helix (in PDB 3C37 and 4AW6), absent in thermolysin and cowrins, are found in both groups (see Figs. 4c,d and 5). Furthermore, in the absence of structural information on the IMMP part of M56 proteins, independent secondary-structure prediction of full-length BlaR1/MecR1, together with predictions for M48 HtpX and Oma1 and the experimental structures of the two minigluzincin zymogens and their three structural relatives, reveals that the CDs are most likely very similar in M48- and M56-IMMPs. This is further supported by the fact that minigluzincins were actually found in a search with BlaR1/MecR1 sequences and by evolutionary studies (see above), which indicate that minigluzincins are even closer to M56- than to M48-IMMPs (Fig. 2b).

CONCLUSIONS

The present work describes a novel family of minimal ~100-residue soluble MPs, the minigluzincins, which evince unique zymogenic structures that are maintained in their inactive state by intra- or inter-molecular C-terminal segments. The minigluzincins share very similar chain traces with thermolysin and other gluzincins such as leukotriene A4 relatives and cowrins, but bear even higher similarity with M48 and M56 IMMPs, thus representing a minimal scaffold for the CDs of a large cohort of integral-membrane proteins. The characterizing feature of the M48- and M56-family CDs is a compact domain containing a backing helix, a three-stranded β-sheet, a linking helix, an active-site helix providing two metal ligands, a glutamate helix providing a third metal ligand and a Ser/Gly-turn, as well as a C-terminal helix (Fig. 5). Into this minimal common scaffold, each specific family member has inserted large upstream, internal and downstream structural elements, including TMHs, to yield the specific functionality as membrane-embedded MPs. Altogether, structural knowledge of minigluzincins may be helpful in the construction of shorter soluble variants of IMMPs and thus facilitate the design of drugs to modulate their function.

ACKNOWLEDGMENTS

We are indebted to Tibisay Guevara for excellent assistance during crystallization experiments. We also thank the Automated Crystallography Platform at IBMB/IRB. We acknowledge the help provided by local contacts at the ESRF synchrotron (Grenoble, France). Funding for travelling and data collection was provided in part by ESRF.

This study was supported in part by grants from European, Spanish, and Catalan agencies (FP7-HEALTH-F3-2009-223101 "AntiPathoGN"; FP7-HEALTH-2010-261460 "Gums&Joints"; FP7-PEOPLE-2011-ITN-290246 "RAPID"; FP7-HEALTH-2012-306029-2 "TRIGGER"; BFU2010-19310; BFU2012-32862; CSD2006-00015; Fundació "La Marató de TV3" grant 2009-100732; and 2009SGR1036), as well as by a postdoctoral JAE contract from CSIC (co-funded by FSE) and two FPI Ph.D. fellowships from the Spanish Ministry for Science and Technology, currently part of the Ministry of Economy and Competitiveness.

ABBREVIATIONS

Abz, aminobenzoyl; Amc, 7-amino-4-methylcoumarin; CD, catalytic domain; CTS, C-terminal sub-domain; Dabcyl, *p*-dimethyl(aminophenyl)azobenzoate; Dap, L-diaminopropionyl; dnp, 2,4-dinitrophenylamino; Edans, 2-aminoethylamino-1-naphthalene sulfonate; FACE1, farnesylated-protein converting enzyme 1; GST, glutathione-S-transferase; IM, integral(-)membrane; IMMP, IM metallopeptidase; Mca, 7-methoxycoumarin-4-acetyl; MP, metallopeptidase; NTS, N-terminal sub-domain; Nva, norvaline; ONp, *p*-nitrophenyl ester; PDB, Protein Data Bank at www.pdb.org; pNA, *p*-nitroanilide; S2P, site-2 proteinase; TEV, tobacco-etch-virus; TMH, transmembrane helix; and UP, UniProt sequence database (www.uniprot.org) access code.

FIGURE LEGENDS

Figure 1. Transmembrane topologies of selected M48 and M56 integral-membrane metallopeptidases. The predicted transmembrane topologies of *Escherichia coli* HtpX (UP P23894), *Saccharomyces cerevisiae* Oma1 (UP P36163), and *Staphylococcus aureus* BlaR1 (UP P18357) were obtained using the TOPCONS program (99) and represented using the TOPO2 server (http://www.sacs.ucsf.edu/TOPO2/). The topology of human FACE1 (UP O75844) was derived from the experimental coordinates (PDB 4AW6). The N- and C-terminal residues, as well as the flanking transmembrane residues, are labeled. FACE1 is located in the human endoplasmic reticulum and Golgi apparatus membrane, HtpX in the bacterial cell inner membrane, Oma1 in the mitochondrial inner membrane, and BlaR1 in the bacterial cell membrane. The zinc-binding signature HEXXH residues are shown in black circles as well as the third protein ligand of FACE1, a glutamate. The segments approximately corresponding to the catalytic domains are framed where known or predictable.

Figure 2. Bioinformatics studies of the target proteins. (A) Structure-based sequence alignment of, top to bottom, *Pyrococcus abyssi* UP Q9V1Y2, *Aquifex aeolicus* UP O66407, and *Methanocaldococcus jannaschii* UP Q57587 performed using the MULTALIN program (100) and represented using the ESPRIPT server (http://espript.ibcp.fr; (100)). Residues with 100%, 75%, and 50% identity are highlighted through red, orange, and yellow background, respectively. The sequence identities to each protein (in %) with the other two proteins (top to bottom) are shown at the beginning of the second alignment block. Numbering and regular secondary structure elements correspond to *P. abyssi* UP Q9V1Y2 (top) and *M. jannaschii* UP Q57587 (bottom). **(B)** Sequence-based phylogenetic tree of the three sequences of (A) and 69 homologous sequences from MEROPS database represented by their UniProt codes. The tree shows three major groups including HtpX-/HtpX-2-like proteins of family M48 (in blue), BlaR1/MecR1- or PenR1-like proteins of family M56 (in green), an unassigned member of family M48 (in magenta), and the three sequences of (A) in red. The tree revealed another orthologous sequence from *M. jannaschii* (UP 58610) not analyzed in the present work. The scale bar represents PHYML branch length.

Figure 3. Structures of proabylysin and projannalysin. (A) Cross-eye stereo ribbon-plot of proabylysin in standard orientation (44), with coils, strands (labeled β1-β3, see also Fig. 2a) and helices (α1-α4 and η1) in pale pink, magenta, and pale purple, respectively. Selected residues are shown for their side chains as sticks with pink carbons and labeled, as is the N-terminus of the protein. The C-terminal segment running across the cleft (Q^{101}-V^{105}) is depicted as a stick model with gray carbons. **(B)** Superposition of the proabylysin (pink) and projannalysin (tan) monomers, whose termini are labeled. For both structures, the metal-binding residues, the general base/acid glutamate and the Ser/Gly-turn phenylalanine are depicted as sticks with carbons colored as the respective ribbons (left panel). Structure of the projannalysin dimer in standard orientation (i.e. as in (A)), with one monomer in tan and the other in aquamarine, and after a horizontal 65º-rotation, which depicts the view along the non-crystallographic twofold axis shown as a red ellipse (right panel). **(C)** Cross-eye stereo ribbon-plot of projannalysin in standard orientation, with coils, strands (labeled β1-β3, see also Fig. 2a) and helices (α1, α1', and α2-α4) in tan, orange, and brown, respectively. Selected

residues, as well as a Tris molecule coordinating the metal (yellow sphere) and a glycerol at the dimer interface, are shown as sticks with tan carbons and labeled in black. Three waters are shown as red spheres. The C-terminal segment (E^{102}-I^{110}) of the symmetry-related molecule in the dimer is shown as a stick model with carbons and labels in turquoise. **(D)** Close-up view in cross-eye stereo of the two projannalysin monomers displayed in Fig. 3b (right panel) superposed with the respective semi-transparent Connolly surfaces. The side chains of the residues involved in dimerization are shown as yellow and blue sticks, respectively, for molecules A and B.

Figure 4. Structural similarities of miniglutzincins. (A) Cross-eye stereo plot depicting the superposition of proabylysin and *Bacillus thermoproteolyticus* thermolysin (PDB 4TLN; (101)) as ribbons colored in pale pink, magenta, and pale lilac (proabylysin) and yellow, light green, and dark green (thermolysin) for the coils, strands, and helices of the common stretches, respectively, whose ends are highlighted by orange arrows (M^{1}-K^{99} of proabylysin and V71-P195 of thermolysin). The orientation corresponds to that of the standard orientation of MPs (44) as in Fig. 4a. The corresponding zinc ions are shown as magenta and green spheres, respectively. The remaining 70-residue N- and 121-residue C-terminal stretches of thermolysin are shown in white. The characteristic calcium ions of thermolysin have been omitted for clarity. **(B)** Close-up view of (A) in cross-eye stereo depicting the metal-binding sites of proabylysin and thermolysin. The residues of the HEXXH sequence are shown for their side chains (in pink for proabylysin [H^{60}, E^{61}, and H^{64}] and in green for thermolysin [H142, E143, and H146]). In addition, the respective downstream third metal-binding residues (H^{72} of probabylysin and E166 of thermolysin; pinpointed by a blue arrow) and the residue of the Ser/Gly-turn (F^{76} of proabylysin and S169 of thermolysin; pinpointed by a red arrow) are depicted. **(C)** Superposition of proabylysin and M48-family human FACE1 protein (PDB 4AW6) in two views resulting from a 90-degree ccw in-plane rotation of the orientation in Fig. 4a (left panel) and a subsequent vertical 180-degree rotation (right panel). The structurally equivalent parts and the catalytic metal ions are displayed in hot pink (proabylysin) and green (FACE1), the remaining parts of FACE1 are shown as a white coil. **(D)** Close-up view in cross-eye stereo of (C; left) after a 90-degree cw in-plane rotation, i.e. in standard orientation, showing only the common domains (M^{1}-Q^{101} of proabylysin and K233-G343+E415-N442 of FACE1). Coils, strands, and helices are shown, respectively, in pink, magenta, and hot pink (proabylysin) and in yellow, green, and light green (FACE1). The catalytic metal, the residues engaged in metal binding, and the general base/acid glutamate of each structure (H^{60}, E^{61}, H^{64}, and H^{72} of proabylysin and H335, E336, H339, and E415 of FACE1) are shown for their side chains with pink and green carbons, respectively. Orange arrows pinpoint the N- and C-termini of the common structure parts. Read arrows pinpoint the positions between the second and third zinc ligands, at which FACE1 has two of its TMHs inserted (segment H344-F414). These are actually a continuation of the active-site and glutamate helices. Black arrows indicate the residues flanking the disordered segment in FACE1 (E286-K321), which coincides with proabylysin linking helix η1. A blue arrow highlights the position of the Ser/Gly-turn residue (F^{76} in proabylysin and A418 in FACE1).

Figure 5. Common topology of minigluzincins and selected M48 and M56 metallopeptidases. The actual regular secondary structure elements of the experimental structures of proabylysin (UP Q9V1Y2; this work), projannalysin (UP Q57587; this work), *G. sulfurreducens* M48-like peptidase (UP Q74D82; PDB 3C37), human FACE1 (UP O75844; PDB 4AW6), and a truncated form of *V. parahaemolyticus* HtpX containing only its N-terminal sub-domain (UP Q87QN1; PDB 3CQB) are represented with dark green rods (helices) and dark purple arrows (strands) with the residues flanking each secondary structure element. In addition, the predicted regular secondary structure elements of *E. coli* HtpX C-terminal sub-domain (UP P23894), yeast Oma1 (UP P36163), and *S. aureus* BlaR1 (UP P18357) are represented with light green rods (helices) and light purple arrows (strands). The protein numbering, the HEXXH motif, the third zinc binding residue, the Ser/Gly-turn residue, and the transmembrane segments (TM) taken either from the experimental structures or the predictions are indicated by either regular or italic letters, respectively.

REFERENCES

1. Wallin, E., and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci*. 7, 1029-1038
2. Sonoda, Y., Newstead, S., Hu, N. J., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A. D., Byrne, B., Iwata, S., and Drew, D. (2011) Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* 19, 17-25
3. Almen, M. S., Nordstrom, K. J., Fredriksson, R., and Schioth, H. B. (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol*. 7, 50
4. Kim, H., Melen, K., Osterberg, M., and von Heijne, G. (2006) A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc. Natl. Acad. Sci. USA* 103, 11142-11147
5. Terstappen, G. C., and Reggiani, A. (2001) *In silico* research in drug discovery. *Trends Pharmacol. Sci*. 22, 23-26
6. Davey, J. (2004) G-protein-coupled receptors: new approaches to maximise the impact of GPCRS in drug discovery. *Expert Opin. Ther. Targets* 8, 165-170
7. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucl. Acids Res*. 40, D343-D350
8. Tusnády, G. E., Dosztányi, Z., and Simon, I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* 20, 2964-2972
9. Fujimura-Kamada, K., Nouvet, F. J., and Michaelis, S. (1997) A novel membrane-associated metalloprotease, Ste24p, is required for the first step of NH_2-terminal processing of the yeast α-factor precursor. *J. Cell Biol*. 136, 271-285
10. Freije, J. M., Blay, P., Pendas, A. M., Cadiñanos, J., Crespo, P., and López-Otín, C. (1999) Identification and chromosomal location of two human genes encoding enzymes potentially involved in proteolytic maturation of farnesylated proteins. *Genomics* 58, 270-280
11. Barrowman, J., and Michaelis, S. (2009) ZMPSTE24, an integral membrane zinc metalloprotease with a connection to progeroid disorders. *Biol. Chem*. 390, 761-773
12. Varela, I., Pereira, S., Ugalde, A. P., Navarro, C. L., Suárez, M. F., Cau, P., Cadiñanos, J., Osorio, F. G., Foray, N., Cobo, J., de Carlos, F., Levy, N., Freije, J. M., and López-Otín, C. (2008) Combined treatment with statins and aminobisphosphonates extends longevity in a mouse model of human premature aging. *Nat. Med*. 14, 767-772
13. Freije, J. M., and López-Otín, C. (2012) Reprogramming aging and progeria. *Curr. Opin. Cell Biol*. 24, 757-764
14. Akiyama, Y. (2009) Quality control of cytoplasmic membrane proteins in *Escherichia coli*. *J. Biochem*. 146, 449-454
15. Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., and Baumann, U. (2006) The molecular architecture of the metalloprotease FtsH. *Proc. Natl. Acad. Sci. USA* 103, 3066-3071
16. Sakoh, M., Ito, K., and Akiyama, Y. (2005) Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *J. Biol. Chem*. 280, 33305-33310
17. Suno, R., Niwa, H., Tsuchiya, D., Zhang, X., Yoshida, M., and Morikawa, K. (2006) Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* 22, 575-585
18. Kaser, M., Kambacheld, M., Kisters-Woike, B., and Langer, T. (2003) Oma1, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. *J. Biol. Chem*. 278, 46414-46423
19. McBride, H., and Soubannier, V. (2010) Mitochondrial function: OMA1 and OPA1, the grandmasters of mitochondrial health. *Curr. Biol*. 20, R274-R276
20. Quiros, P. M., Ramsay, A. J., Sala, D., Fernández-Vizarra, E., Rodriguez, F., Peinado, J. R., Fernández-García, M. S., Vega, J. A., Enríquez, J. A., Zorzano, A., and Lopez-Otín, C. (2012) Loss of mitochondrial protease OMA1 alters processing of the GTPase OPA1 and causes obesity and defective thermogenesis in mice. *EMBO J*. 31, 2117-2133
21. Mallorquí-Fernandez, G., Marrero, A., García-Piquè, S., García-Castellanos, R., and Gomis-Rüth, F. X. (2004) Staphylococcal methicillin resistance: fine focus on folds and functions. *FEMS Microbiol. Lett*. 235, 1-8
22. Fuda, C. C. S., Fisher, J. F., and Mobashery, S. (2005) β-Lactam resistance in *Staphylococcus aureus*: the adaptive resistance of a plastic genome. *Cell. Mol. Life Sci*. 62, 2617-2633
23. Llarrull, L. I., Testero, S. A., Fisher, J. F., and Mobashery, S. (2010) The future of the β-lactams. *Curr. Opin. Microbiol*. 13, 551-557
24. Llarrull, L. I., and Mobashery, S. (2012) Dissection of events in the resistance to β-lactam antibiotics mediated by the protein BlaR1 from *Staphylococcus aureus*. *Biochemistry* 51, 4642-4649
25. Zhu, Y., Englebert, S., Joris, B., Ghuysen, J. M., Kobayashi, T., and Lampen, J. O. (1992) Structure, function, and fate of the BlaR signal transducer involved in induction of β-lactamase in *Bacillus licheniformis*. *J. Bacteriol*. 174, 6171-6178

26. Hardt, K., Joris, B., Lepage, S., Brasseur, R., Lampen, J. O., Frere, J. M., Fink, A. L., and Ghuysen, J. M. (1997) The penicillin sensory transducer, BlaR, involved in the inducibility of β-lactamase synthesis in *Bacillus licheniformis* is embedded in the plasma membrane via a four-α-helix bundle. *Mol. Microbiol.* 23, 935-944
27. Berzigotti, S., Benlafya, K., Sepulchre, J., Amoroso, A., and Joris, B. (2012) *Bacillus licheniformis* BlaR1 L3 loop is a zinc metalloprotease activated by self-proteolysis. *PLoS ONE* 7, e36400
28. Hao, H., Dai, M., Wang, Y., Huang, L., and Yuan, Z. (2012) Key genetic elements and regulation systems in methicillin-resistant *Staphylococcus aureus*. *Fut. Microbiol.* 7, 1315-1329
29. Marrero, A., Mallorquí-Fernández, G., Guevara, T., García-Castellanos, R., and Gomis-Rüth, F. X. (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
30. Birck, C., Cha, J. Y., Cross, J., Schulze-Briese, C., Meroueh, S. O., Schlegel, H. B., Mobashery, S., and Samama, J.-P. (2004) X-ray crystal structure of the acylated β-lactam sensor domain of BlaR1 from *Staphylococcus aureus* and the mechanism of receptor activation for signal transduction. *J. Am. Chem. Soc.* 126, 13945-13947
31. Wilke, M. S., Hills, T. L., Zhang, H. Z., Chambers, H. F., and Strynadka, N. C. J. (2004) Crystal structures of the apo and penicillin-acylated forms of the BlaR1 β-lactam sensor of *Staphylococcus aureus*. *J. Biol. Chem.* 279, 47278-47287
32. Kerff, F., Charlier, P., Colombo, M.-L., Sauvage, E., Brans, A., Frère, J.-M., Joris, B., and Fonzé, E. (2003) Crystal structure of the sensor domain of the BlaR penicillin receptor from *Bacillus licheniformis*. *Biochemistry* 42, 12835-12843
33. Kuntz, I. D. (1992) Structure-based strategies for drug design and discovery. *Science* 257, 1078-1082
34. Greer, J., Erickson, J. W., Baldwin, J. J., and Varney, M. D. (1994) Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* 37, 1035-1054
35. Arinaminpathy, Y., Khurana, E., Engelman, D. M., and Gerstein, M. B. (2009) Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discov. Today* 14, 1130-1135
36. Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P. D., and Shi, Y. (2007) Structure of a site-2 protease family intramembrane metalloprotease. *Science* 318, 1608-1612
37. Wolfe, M. S. (2009) Intramembrane proteolysis. *Chem. Rev.* 109, 1599-1612
38. Wolfe, M. S. (2009) Intramembrane-cleaving proteases. *J. Biol. Chem.* 284, 13969-12973
39. Pryor, E. E., Jr., Horanyi, P. S., Clark, K. M., Fedoriw, N., Connelly, S. M., Koszelak-Rosenblum, M., Zhu, G., Malkowski, M. G., Wiener, M. C., and Dumont, M. E. (2013) Structure of the integral membrane protein CAAX protease Ste24p. *Science* 339, 1600-1604
40. Quigley, A., Dong, Y. Y., Pike, A. C., Dong, L., Shrestha, L., Berridge, G., Stansfeld, P. J., Sansom, M. S., Edwards, A. M., Bountra, C., von Delft, F., Bullock, A. N., Burgess-Brown, N. A., and Carpenter, E. P. (2013) The structural basis of ZMPSTE24-dependent laminopathies. *Science* 339, 1604-1607
41. Bode, W., Gomis-Rüth, F. X., and Stöcker, W. (1993) Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the ´metzincins´. *FEBS Lett.* 331, 134-140
42. Hooper, N. M. (1994) Families of zinc metalloproteases. *FEBS Lett.* 354, 1-6
43. McKerrow, J. H. (1987) Human fibroblast collagenase contains an amino acid sequence homologous to the zinc-binding site of *Serratia* protease. *J. Biol. Chem.* 262, 5943-5943
44. Gomis-Rüth, F. X., Botelho, T. O., and Bode, W. (2012) A standard orientation for metallopeptidases. *Biochim. Biophys. Acta* 1824, 157-163
45. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402
46. Sánchez, R., Serra, F., Tarraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutirrez, S., Huerta-Cepas, J., Gabald´ón, T., Dopazo, J., and Dopazo, H. (2011) Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucl. Acids Res.* 39, W470-W474
47. Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972-1973
48. Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164-1165
49. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696-704
50. Cole, C., Barber, J. D., and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. *Nucl. Acids Res.* 36, W197-W201
51. Tallant, C., García-Castellanos, R., Baumann, U., and Gomis-Rüth, F. X. (2010) On the relevance of the Met-turn methionine in metzincins. *J. Biol. Chem.* 285, 13951-13957

52. Goulas, T., Arolas, J. L., and Gomis-Rüth, F. X. (2010) Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc. Natl. Acad. Sci. USA* 108, 1856-1861
53. Tallant, C., García-Castellanos, R., Seco, J., Baumann, U., and Gomis-Rüth, F. X. (2006) Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J. Biol. Chem.* 281, 17920-17928
54. Botelho, T. O., Guevara, T., Marrero, A., Arêde, P., Fluxa, V. S., Reymond, J. L., Oliveira, D. C., and Gomis-Rüth, F. X. (2011) Structural and functional analyses reveal that *Staphylococcus aureus* antibiotic resistance factor HmrA is a zinc-dependent endopeptidase. *J. Biol. Chem.* 286, 25697-25709
55. Kabsch, W. (2010) XDS. *Acta Crystallogr. sect. D* 66, 125-132
56. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. sect. D* 66, 133-144
57. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. sect. D* 67, 282-292
58. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D* 67, 235-242
59. Sheldrick, G. M. (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. sect. D* 66, 479-485
60. Sheldrick, G. M. (2011) The *SHELX* approach to experimental phasing of macromolecules. *Acta Crystallogr. sect. A* 67, C13
61. Hübschle, C. B., Dittrich, B., and Sheldrick, G. M. (2012) *SHELXE* - a GUI for *SHELXL*. *Acta Crystallogr. sect. A* 68, s121
62. Carranza, C., Inisan, A.-G., Mouthuy-Knoops, E., Cambillau, C., and Roussel, A. (1999) Turbo-Frodo. in *AFMB Activity Report 1996-1999*, CNRS-UPR 9039, Marseille. pp 89-90
63. Debreczeni, J. É., and Emsley, P. (2012) Handling ligands with Coot. *Acta Crystallogr. sect. D* 68, 425-430
64. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. sect. D* 66, 486-501
65. Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. sect. D* 68, 368-380
66. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr. sect. D* 66, 470-478
67. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605-1612
68. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR System : a new software suite for macromolecular structure determination. *Acta Crystallogr. sect. D* 54, 905-921
69. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. sect. D* 60, 2256-2268
70. Lawrence, M. C., and Colman, P. M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* 234, 946-950
71. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D* 66, 12-21
72. Vriend, G. (1990) WHAT IF : a molecular modelling and drug design program. *J. Mol. Graph.* 8, 52-56
73. Ortiz, A. R., Strauss, C. E., and Olmea, O. (2002) MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Prot. Sci.* 11, 2606-2621
74. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank. *Acta crystallogr. sect. D* 58, 899-907
75. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545-W549
76. Schechter, I., and Berger, A. (1967) On the size of active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157-162
77. Tallant, C., Marrero, A., and Gomis-Rüth, F. X. (2010) Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim. Biophys. Acta - Mol. Cell Res.* 1803, 20-28
78. Takeda, S. (2009) Three-dimensional domain architecture of the ADAM family proteinases. *Semin. Cell Dev. Biol.* 20, 146-152

79. Takeda, S., Takeya, H., and Iwanaga, S. (2012) Snake venom metalloproteinases: Structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim. Biophys. Acta* 1824, 164-176
80. Guevara, T., Yiallouros, I., Kappelhoff, R., Bissdorf, S., Stöcker, W., and Gomis-Rüth, F. X. (2010) Proenzyme structure and activation of astacin metallopeptidase. *J. Biol. Chem.* 285, 13958-13965
81. Arolas, J. L., Broder, C., Jefferson, T., Guevara, T., Sterchi, E. E., Bode, W., Stöcker, W., Becker-Pauly, C., and Gomis-Rüth, F. X. (2012) Structural basis for the sheddase function of human meprin β metalloproteinase at the plasma membrane. *Proc. Natl. Acad. Sci. USA* 109, 16131-16136
82. Gomis-Rüth, F. X. (2008) Structure and mechanism of metallocarboxypeptidases. *Crit. Rev. Biochem. Mol. Biol.* 43, 319-345
83. Gao, X., Wang, J., Yu, D. Q., Bian, F., Xie, B. B., Chen, X. L., Zhou, B. C., Lai, L. H., Wang, Z. X., Wu, J. W., and Zhang, Y. Z. (2010) Structural basis for the autoprocessing of zinc metalloproteases in the thermolysin family. *Proc. Natl. Acad. Sci. USA* 107, 17569-17574
84. Janin, J., and Chothia, C. (1990) The structure of protein-protein recognition sites. *J. Biol. Chem.* 265, 16027-16030
85. Harding, M. M. (2004) The architecture of metal coordination groups in proteins. *Acta Crystallogr. sect. D* 60, 849-859
86. Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., and Dupourque, D. (1972) Three-dimensional structure of thermolysin. *Nature* 238, 37-41
87. Tholander, F., Muroya, A., Roques, B. P., Fournie-Zaluski, M. C., Thunnissen, M. M., and Haeggstrom, J. Z. (2008) Structure-based dissection of the active site chemistry of leukotriene A4 hydrolase: implications for M1 aminopeptidases and inhibitor design. *Chem. Biol.* 15, 920-929
88. Bauvois, C., Jacquamet, L., Huston, A. L., Borel, F., Feller, G., and Ferrer, J. L. (2008) Crystal structure of the cold-active aminopeptidase from *Colwellia psychrerythraea*, a close structural homologue of the human bifunctional leukotriene A4 hydrolase. *J. Biol. Chem.* 283, 23315-23325
89. Nguyen, T. T., Chang, S. C., Evnouchidou, I., York, I. A., Zikos, C., Rock, K. L., Goldberg, A. L., Stratikos, E., and Stern, L. J. (2011) Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.* 18, 604-613
90. Kochan, G., Krojer, T., Harvey, D., Fischer, R., Chen, L., Vollmar, M., von Delft, F., Kavanagh, K. L., Brown, M. A., Bowness, P., Wordsworth, P., Kessler, B. M., and Oppermann, U. (2011) Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming. *Proc. Natl. Acad. Sci. USA* 108, 7745-7750
91. Kyrieleis, O. J., Goettig, P., Kiefersauer, R., Huber, R., and Brandstetter, H. (2005) Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. *J. Mol. Biol.* 349, 787-800
92. McGowan, S., Porter, C. J., Lowther, J., Stack, C. M., Golding, S. J., Skinner-Adams, T. S., Trenholme, K. R., Teuscher, F., Donnelly, S. M., Grembecka, J., Mucha, A., Kafarski, P., Degori, R., Buckle, A. M., Gardiner, D. L., Whisstock, J. C., and Dalton, J. P. (2009) Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. *Proc. Natl. Acad. Sci. USA* 106, 2537-2542
93. Addlagatta, A., Gay, L., and Matthews, B. W. (2006) Structure of aminopeptidase N from Escherichia coli suggests a compartmentalized, gated active site. *Proc Natl Acad Sci U S A* 103, 13339-13344
94. Gomis-Rüth, F. X. (2003) Structural aspects of the *metzincin* clan of metalloendopeptidases. *Mol. Biotech.* 24, 157-202
95. Zhan, C., Fedorov, E. V., Shi, W., Ramagopal, U. A., Thirumuruhan, R., Manjasetty, B. A., Almo, S. C., Fiser, A., Chance, M. R., and Fedorov, A. A. (2005) The ybeY protein from *Escherichia coli* is a metalloprotein. *Acta Crystallogr. sect. F* 61, 959-963
96. Oganesyan, V., Busso, D., Brandsen, J., Chen, S., Jancarik, J., Kim, R., and Kim, S. H. (2003) Structure of the hypothetical protein AQ_1354 from *Aquifex aeolicus*. *Acta Crystallogr. sect. D* 59, 1219-1223
97. Weiss, M. S. (2001) Global indicators of X-ray quality. *J. Appl. Cryst.* 34, 130-135
98. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. sect. D* 62, 72-82

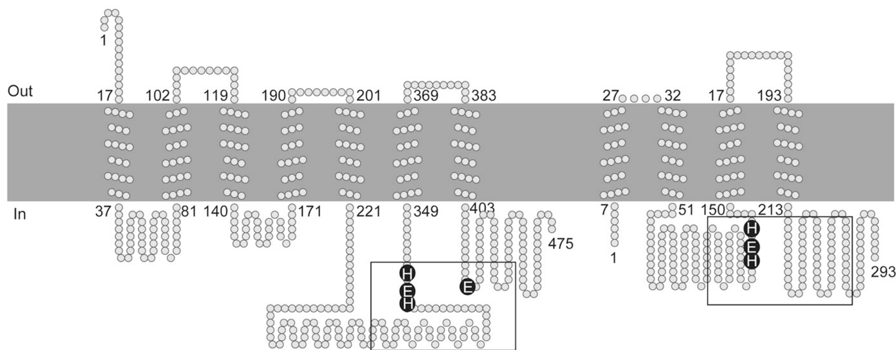
Table 1. Crystallographic data.

	Pyrococcus abyssi proabylysin (UP Q9V1Y2)		*Methanocaldococcus jannaschii* projannalysin (UP Q57587)	
Dataset	Native	Zinc [a] (absorption peak)	Native	Mercury [a] (absorption peak)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P3_121$	$P3_121$
Cell constants (a, b, c, in Å)	34.62, 44.66, 72.00	34.68, 44.90, 72.23	76.59, 76.59, 124.32	76.60, 76.60, 124.76
Wavelength (Å)	0.9540	1.2821	0.9790	1.0044
No. of measurements / unique reflections	495,316 / 40,510	99,212 / 15,317	305,164 / 29,001	179,641 / 31,910
Resolution range (Å) (outermost shell) [b]	44.7 – 1.15 (1.21 – 1.15)	44.9 – 1.60 (1.69 – 1.60)	45.4 – 2.00 (2.05– 2.00)	89.1 – 2.40 (2.46–
Completeness [/ Anom. completeness] (%)	100.0 (100.0)	98.9 (92.9) / 97.3 (86.6)	99.1 (98.6)	2.40)
R_{merge} [d]	0.072 (0.892)	0.059 (0.374)	0.042 (0.851)	99.6 (96.6)
$R_{r.i.m.}$ (= R_{meas}) [d] [/ $R_{p.i.m.}$ [d]]	0.075 (0.934) / 0.021 (0.272)	0.069 (0.453) / 0.035 (0.250)	0.044 (0.910)	0.048 (0.541)
Average intensity (<[<I> / σ(<I>)]>)	18.5 (3.4)	15.5 (4.5)	30.3 (2.8)	0.054 (0.615)
B-Factor (Wilson) ($Å^2$) / Aver. multiplicity	11.0 / 12.2 (11.3)	18.2 / 6.5 (5.2)	49.4 / 10.5 (7.9)	24.2 (3.0)
Resolution range used for refinement (Å)	∞ – 1.15		∞ – 2.00	54.4 / 5.6 (4.4)
No. of reflections used (test set)	39,668 (778)		28,924 (751)	
Crystallographic R_{factor} (free R_{factor}) [e]	0.147 (0.192)		0.183 (0.213)	
No. of protein atoms / solvent molecules / neutral ligands / ionic ligands	900 / 147 / 1 $(CH_2OH)_2CHOH$ / 1 Zn^{2+}		1,815 / 194 / 7 $(CH_2OH)_2CHOH$ / 1 $C(CH_2OH)_3NH_3^+$, 2 Zn^{2+}, 3 Cl^-, 4 SO_4^{2-}	
Rmsd from target values				
bonds (Å) / angles (°)	0.014 / 2.24		0.009 / 0.96	
Average B-factors for protein atoms ($Å^2$)	17.7		54.9	
Main-chain conformational angle analysis [c]				
Residues in favored regions / outliers / all residues	103 / 0 / 103		211 / 0 / 212	

[a] Friedel-mates were treated as separate reflections. [b] Values in parentheses refer to the outermost resolution shell. [c] According to MOLPROBITY (71).

[d] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{r.i.m.} = \Sigma_{hkl}(n_{hkl}/[n_{hkl}-1])^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{p.i.m.} = \Sigma_{hkl}(1/[n_{hkl}-1])^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the i-th intensity measurement and n_{hkl} the redundancy of reflection hkl—including symmetry-related reflections—and <I(hkl)> its average intensity. $R_{r.i.m.}$ (*alias* R_{meas}) and $R_{p.i.m.}$ are improved multiplicity-weighted indicators of the quality of the data, the redundancy-independent merging R factor and the precision-indicating merging R factor. The latter is computed after averaging over multiple measurements (for details, see (97,98)).

[e] Crystallographic $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where k is a scaling factor, and F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. This factor is calculated for the working-set reflections; free R_{factor}, same for a test-set of reflections (>500) not used during refinement.



Oma1
BlaR1
1
585
1
345
Out
63 240
7 58 102 319
In
83 220
22 38 122 300
H
E H
H
E
H

Fig.1

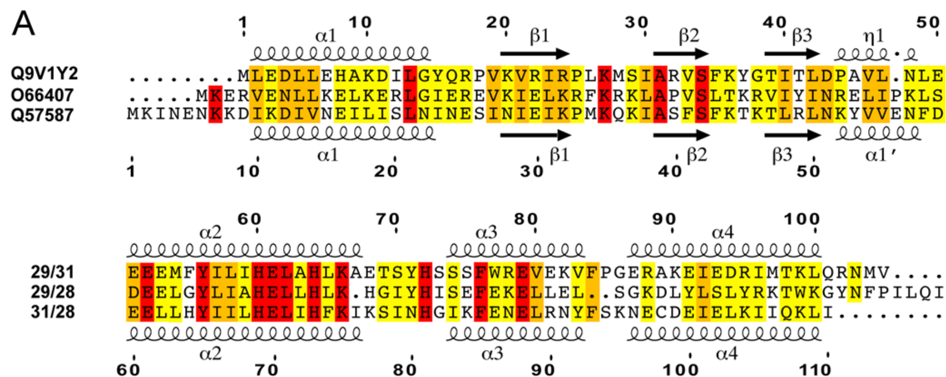
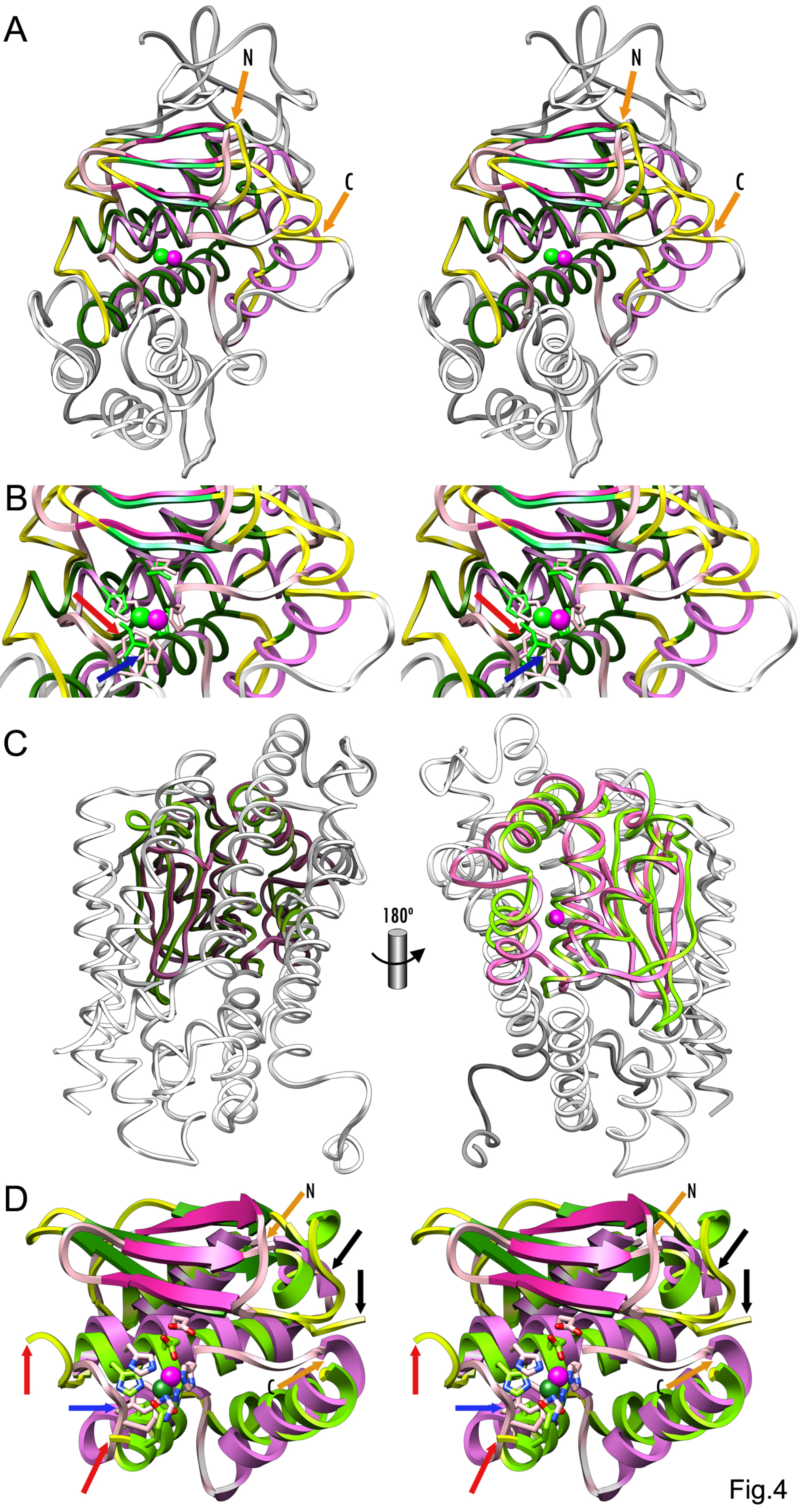


Fig.2

Fig.3


A
N
C
B
C
180°
D
N
C
Fig.4

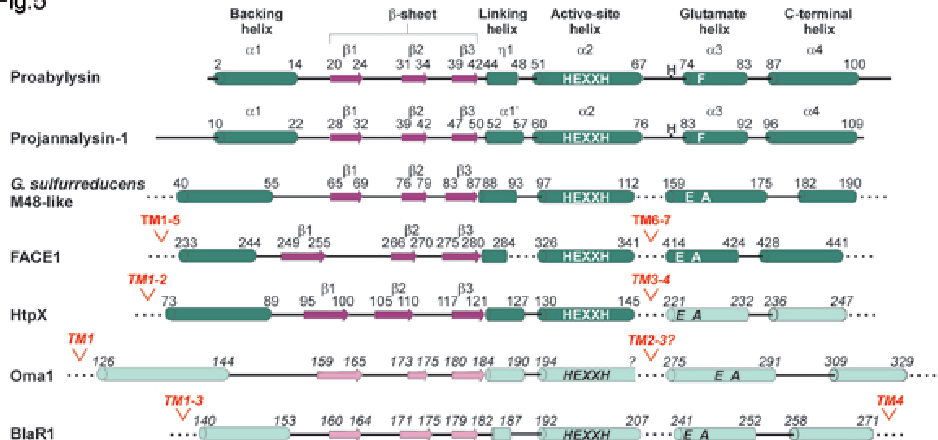


Fig.5
Backing helix
β-sheet
Linking helix
Active-site helix
Glutamate helix
C-terminal helix
Proabylysin
Projannalysin-1
G. sulfurreducens M48-like
FACE1
HtpX
Oma1
BlaR1
HEXXH
E A
TM1-5
TM6-7
TM1-2
TM3-4
TM1
TM2-3?
TM1-3
TM4