Molecular Analysis of Ulilysin, the Structural Prototype of a New Family of Metzincin Metalloproteases*

Received for publication, January 30, 2006, and in revised form, March 30, 2006. Published, JBC Papers in Press, April 20, 2006, DOI 10.1074/jbc.M600907200

Cynthia Tallant1, Raquel García-Castellanos1, Jordi Seco, Ulrich Baumann2, and F. Xavier Gomis-Ruth3

From the Departament de Biologia Estructural, Institut de Biologia Molecular de Barcelona, Consejo Superior de Investigaciones Científicas, c/Jordi Girona, 18-26, E-08034 Barcelona, Spain

The metzincin clan encompasses several families of zinc-dependent metalloproteases with proven function both in physiology and pathology. They act either as broad spectrum protein degraders or as sheddases, operating through limited proteolysis. Among the structurally uncharacterized metzincin families are the pappalysins, of which the most thoroughly studied member is human pregnancy-associated plasma protein A (PAPP-A), a heavily glycosylated 170-kDa multidomain protein specifically cleaving insulin-like growth factor (IGF)-binding proteins (IGFBPs). Proulilysin is a 38-kDa archaeal protein that shares sequence similarity with PAPP-A but encompasses only the pro-domain and the catalytic domain. It undergoes calcium-mediated autolytic activation, and the mature protein adopts a three-dimensional structure with two subdomains separated by an active site cleft containing the catalytic zinc ion. This structure is reminiscent of human members of the adamalysin/ADAMs (a disintegrin and a metalloprotease) family of metzincins. A bound dipeptide yields information on the substrate specificity of ulilysin, which specifically hydrolyzes IGFBP-2 to -6, insulin, and extracellular matrix proteins but not IGFBP-1 or IGF-II. Interestingly, ulilysin has higher proteolytic efficiency and a broader substrate specificity than human PAPP-A. The structure of ulilysin represents a prototype for the catalytic domain of pappalysins.

Insulin-like growth factors (IGF-I and -II) regulate human somatic growth and development. Their activity is also correlated with several diseases, including atherosclerosis, cardiovascular disease, diabetes, and cancer (1). Most circulating IGF molecules are sequestered in complexes with soluble IGF-binding proteins (IGFBP-1 to -6) (2). IGFBPs antagonize binding of IGFs to their receptors because of their much higher affinity, and they are thus carriers, mediators, and reservoirs of IGFs (3). IGFs are released from these complexes with IGFBP proteolytic inactivation mediated by IGFBP proteases, which include serine proteinases, cysteine proteinases, and metalloproteases (MPs). Among these, human PAPP-A specifically inactivates IGFBPs (4–6). This MP was originally identified as an antigen in human pregnancy plasma (7). It is ubiquitously expressed and plays central roles in ovarian follicular development, myogenesis, human embryo implantation, and wound healing (8). Mature PAPP-A is a glycosylated multidomain protein of 1,547 residues that specifically hydrolyzes human IGFBP-4 in an IGF-dependent manner. Only human IGFBP-5 and bovine and porcine IGFBP-2 have been identified as further substrates (6). In addition to its proteolytic domain, three Lin12-Notch repeats (LNR-1, -2, and -3), modules that regulate ligand-induced proteolytic cleavage of the Notch receptor, and five complement control protein modules (CCP1–5) have been identified (9). Human PAPP-A is the founding member of the pappalysin family of MPs, which further encompasses the paralogues PAPP-A2 (10) and has been included in the metzincin clan of MPs (5).

This clan encompasses protease families containing an extended zinc-binding consensus sequence (ZBCS), HEXXHXXGXXH/D, which comprises three zinc ligands (underlined) and a glutamate, which acts as a general base. Currently, representative three-dimensional structures of six different proteinase families bearing this motif show, despite negligible sequence similarity, topological elements in common, among them a methionine-containing 1/4–8-turn called Met turn. These are the astacins, adamalysins/ADAMs, serralysins, matrixins, snapalysins, and leishmanolysins. These two features, the ZBCS and the Met turn, gave rise to the name of the clan. They are linked by variable connecting segments that are characteristic for each constituting family (11–13). Interestingly, members of the adamalysin/ADAM (ADAM-9, -12, and -19) have been identified as IGFBP proteases (14, 15). These two latter families have previously been associated with the fate of other growth factors and derived pathologies. They mediate shedding of the ectodomains of membrane-anchored growth factors, cytokines, and receptors and thus increase their circulating forms (16).

We have identified a series of novel potential pappalysins through bioinformatic searches and studied a potential orthologue from Methanosarcina acetivorans (SwissProt protein sequence data base access code Q8TL28). This was the onlyarchaeal form found, for which we hereby propose the name “ulilysin.” We present the molecular analysis of this fully functional prokaryotic protease and discuss its mechanism of activation, as well as implications for other members of the pappalysin family.

EXPERIMENTAL PROCEDURES

Recombinant Overexpression and Purification of M. acetivorans Ulilysin—A DNA fragment encoding the full-length 342-residue protein MA3214 (38-kDa proulilysin) was amplified from total genomic DNA of M. acetivorans strain C2A and cloned into expression vectors.

The abbreviations used are: IGF, insulin-like growth factor; IGFBP, IGF-binding protein; PAPP-A, pregnancy-associated plasma protein A; MP, metalloprotease; MMP, matrix metalloprotease; NTS, N-terminal subdomain; PAPP-A, pregnancy-associated plasma protein A; ZBCS, zinc-binding consensus sequence; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine; MES, 4-morpholineethanesulfonic acid.
Cysteine-to-alanine mutants were constructed, because the wild-type protein tended to form aggregates that hampered reproducibility in crystallization. With the mutation at position 269 (Fig. 1), crystallization was reproducible. This variant was expressed in Escherichia coli BL21 StarTM cells with plasmid pET-28a, which attaches an N-terminal His6 tag and a thrombin cleavage site. A selenomethionine variant of prouliysin was obtained analogously, except that cells were grown in minimal medium containing selenomethionine and amino acids. The bacterial cell cultures were collected and centrifuged, the pellets were resuspended, and the cells were disrupted by means of a cell disruptor before centrifugation. The supernatant was subjected to protein purification through nickel-nitritotriacetic acid affinity chromatography, and the polyhistidine tag was removed with thrombin. The partially purified protein was further purified by anion exchange fast protein liquid chromatography in a calibrated Superdex 75 HR 10/30 column. N-terminal Edman degradation, enzymatic digestion, and mass spectrometry analyses were performed in collaboration with the Proteomics Unit of the Technical and Scientific Service of the Barcelona Science Park and the Laboratory of Oncology (Vall d’Hebron Hospital, Barcelona, Spain).

**Biochemical Studies in Vitro**—If not otherwise stated, the proteolytic assays were performed at 38-kDa proulilysin and 29-kDa ulilysin concentrations of 0.45 and 0.65 mg/ml, respectively, in 50 mM Tris-HCl, pH 7.5, 5 mM CaCl2 at 37 °C for 2–4 h and at (pro)protease:substrate weight ratios of 1:100 or 1:200. The proteolytic capacity of ulilysin was tested on several substrates (Table 1). For the collagen substrates the buffer was 100 mM Tris-HCl, pH 7.5, 0.2 mM NaCl. In the case of azosubstrates, absorbance was monitored at 440 nm. IGFBP-1 to -6 were tested at 1:300 and IGF-II at 1:100 at 20 °C for 4 h. Casein and gelatin zymography was performed according to the manufacturer’s instructions on precast Tris-Tricine Bio-Rad Ready Gels (with either 10% gelatin or 12% casein).

**Table 1**

| Substrates | Cleavage | Limited proteolysis |
|------------|----------|---------------------|
| Natural proteins | Human plasma α1-antitrypsin | – | Yes |
| | Bovine muscle actin | – | No |
| | Casein | + | No |
| | Bovine neck ligament elastin | + | No |
| | Bovine plasma fibronectin | + | No |
| | Human plasma fibrinogen | + | Yes |
| | Gelatin | + | No |
| | Human plasma plasmin | – | No |
| Labeled proteins | Azoalbumin | – | No |
| | Azoasepin | – | No |
| | Azoacoll | + | No |
| Collagens | Collagen type I, human placenta | – | No |
| | Collagen type I, kangaroo tail | – | No |
| | Collagen type IV, calf skin | – | No |
| | Collagen type V, human placenta | – | No |
| IGF axis proteins | IGF-II | – | – |
| | IGFBP-1 | – | Yes |
| | IGFBP-2 | + | Yes |
| | IGFBP-3 | + | Yes |
| | IGFBP-4 | + | Yes |
| | IGFBP-5 | + | Yes |
| | IGFBP-6 | + | Yes |
| | Insulin chain B | + | Yes |

The oligomerization state of proulilysin and ulilysin (at 0.5–1.8 mg/ml) in solution was assessed by size exclusion chromatography in a calibrated Superdex 75 HR 10/30 column. N-terminal Edman degradation, enzymatic digestion, and mass spectrometry analyses were performed in collaboration with the Proteomics Unit of the Technical and Scientific Service of the Barcelona Science Park and the Laboratory of Oncology (Vall d’Hebron Hospital, Barcelona, Spain).

**Crystallization of Ulilysin, Structure Solution, and Refinement**—Orthorhombic P21212 crystals with two molecules in the asymmetric unit were obtained from sitting drops consisting of full-length C269A 38-kDa proulilysin (5 mg/ml in 30 mM Tris-HCl pH 7.5, 2 mM dithio-
TABLE 2
Inhibition studies of ulilysin

| Inhibitor                        | Inhibitor concentration | Inhibitor specificity | Residual activity |
|----------------------------------|-------------------------|-----------------------|-------------------|
| Control                          | 0.3 μM                  | Serine proteases      | 100               |
| Bovine lung aprotinin            | 1 μM                    | Serine and cysteine proteases | 93                  |
| Phenylmethylsulfonyl fluoride    | 1 μM                    | Cysteine proteases    | 87                |
| Iodoacetamide                    | 1 μM                    | Cysteine proteases    | 103               |
| E-64                             | 10 μM                   | Aspartic proteases    | 95                |
| Pepstatin A                      | 1 μM                    | MMPs                  | 92                |
| Phosphoramidon                   | 10 μM                   | MMPs (thermolysin-like)| 107               |
| 1,10-Phenanthroline              | 5 μM                    | MMPs                  | 2                 |
| EDTA                             | 5 μM                    | MMPs                  | 6                 |
| ZnCl₂                            | 5 μM                    | MMPs                  | 11                |
| CT1746                           | 1 μM                    | MMPs                  | 57                |
| 409.1                            | 1 μM                    | MMPs                  | 87                |
| Galardine                        | 1 μM                    | MMPs                  | 88                |
| Batimastat                       | 1 μM                    | MMPs                  | 24                |
| RXP03-R1                        | 1 μM                    | MMPs (angiotensin converting enzyme) | 78             |
| Captopril                        | 1 μM                    | MMPs                  | 92                |
| l-Arginine                       | 5 μM                    | MMPs                  | 99/55             |
| Guanidinoethylmercaptopuscacinic acid | 1 μM/10 μM | Metallocoxbpeptidases of type B | 40                |
| Benzanamide                      | 5 μM                    | Serine proteases and metallocoxbpeptidases of type B | 56                |

Bioinformatic amino acid sequence similarities were undertaken within MEROPS data base (merops.sanger.ac.uk) and with the servers ProDom (protein.toulouse.inra.fr/promdo.html), Pfam (www.sanger.ac.uk/software/pfam), and PSI-BLAST (www.ncbi.nlm.nih.gov/blast). For the last server, the sequence of PAPP-A shown in Fig. 1, which includes the putative MP region, was employed as bait after exclusion of the two LNR domains. Structural similarity searches were performed with the DALI server (www.ebi.ac.uk/msd). Multiple sequence alignments were calculated with MULTALIN (prods.toulouse.inra.fr/multalin). Close contacts and interaction surfaces were calculated with the program CNS (25). The final coordinates of ulilysin have been deposited with the Protein Data Bank.

Bioinformatic amino acid sequence similarities were undertaken within MEROPS data base (merops.sanger.ac.uk) and with the servers ProDom (protein.toulouse.inra.fr/promdo.html), Pfam (www.sanger.ac.uk/software/pfam), and PSI-BLAST (www.ncbi.nlm.nih.gov/blast). For the last server, the sequence of PAPP-A shown in Fig. 1, which includes the putative MP region, was employed as bait after exclusion of the two LNR regions. Structural similarity searches were performed with the DALI server (www.ebi.ac.uk/msd). Multiple sequence alignments were calculated with MULTALIN (prods.toulouse.inra.fr/multalin). Close contacts and interaction surfaces were calculated with the program CNS (25). The final coordinates of ulilysin have been deposited with the Protein Data Bank.

threitol, 100 mM NaCl) and reservoir solution (18% polyethylene glycol 8000, 0.1 M MES, pH 6.5, 0.2 M CaCl₂) after several weeks at 20 °C. Mass spectrometry analysis and N-terminal sequencing of the crystallized protein revealed a molecular mass of 28,885 ± 50 Da and that the N terminus started at position Arg⁶¹, thus corresponding to activated 29-kDa ulilysin. The crystal structure was solved employing the self-rotation function derived that crystallized in the same conditions as the native protein. A three-wavelength multiple-wavelength anomalous diffraction experiment was carried out at the European Synchrotron Radiation Facility Beamline BM16 (Grenoble, France). Furthermore, a high resolution data set was collected from the same crystal. Diffraction data were collected on an MAR CCD detector and processed with MOSFLM (17); they were then scaled, reduced, and merged with SCALA (18) within the CCP4 suite (19) (Table 3). The multiple-wavelength anomalous diffraction data enabled identification of 10 of the 12 selenium sites present (6/monomer) with XPREP/SHELXD (20), and the phases were computed with SHARP (21). These phases gave rise to interpretable electron density maps. Subsequently, manual model building using TURBO-Frodo (22) alternated with crystallographic refinement with REFMAC5 (23) within CCP4, until the final model was obtained. This model contains protein residues Arg⁶¹–Ala³²². Furthermore, one zinc (Zn⁹⁹⁹) and two calcium cations (Ca⁹⁹⁸ and Ca⁹⁷⁷) were assigned to each molecule present in the asymmetric unit based on the ion coordination spheres and distance and geometry of the ligands. Comparable B-factors to the bound atoms and absence of positive or negative difference electron density after crystallographic refinement further supported this assignment, as well as the requirement of both zinc and calcium for activity. All of the protein residues were found in the most favored and additionally allowed regions of a Ramachandran plot. Each polypeptide chain displayed one disulfide bond (Cys⁵⁵⁰–Cys⁷⁷⁷). 587 solvent molecules (Hoh76W–Hoh604W), a fifth calcium cation engaged in crystal contacts (Ca1W), and six (tentatively assigned) glycerol molecules (Gol2W–Gol7W) were also identified in the structure. Finally, a dipeptide was found (Arg¹⁰¹–Val¹⁰²) in each of the two active sites.

Miscellaneous—The two molecules within the asymmetric unit (suffixes A and B) are related by an almost perfect noncrystallographic dyad, and structurally equivalent, with a root mean square deviation of 0.36 Å for all atoms of each polypeptide chain. Accordingly, discussion will focus on molecule A. The figures were prepared with TURBO-Frodo and MOLSCRIPT (24). Structural superimpositions were performed with TURBO-Frodo.

TABLE 3
Crystalllographic statistics on data collection and refinement

| Data set | Ultrafisn C269A SeMet derivative (high resolution) |
|----------|--------------------------------------------------|
| Space group | P2₁,2,2                                       |
| Cell constants (a, b, c in Å) | 49.6, 126.1, 87.4 |
| Wavelength (Å) | 0.9840                                      |
| No. of measurements | 362,539                                   |
| No. of unique reflections | 61,061                                      |
| Resolution range used for refinement (Å) | 32.1-1.70 (1.79-1.70) |
| Completeness (%) | 99.8 (99.9)                                |
| Rmerge (Å2) | 0.082 (0.438)                                 |
| Average multiplicity | 11.6 (3.4)                                 |
| B-factor (Wilson) (Å²) | 17.9                                       |
| Resolution range used for refinement (Å) | 32.1-1.70 (1.79-1.70) |
| No. of reflections used (test set) | 60,401 (610)                               |
| Crystallographic Rmerge (free Rfactor) | 0.173 (0.241)                             |
| No. of protein atoms (asymmetric unit) | 4,130                                     |
| No. of solvent molecules/ions/other molecules (asymmetric unit) | 597 (Zn²⁺), 5 (Ca²⁺)/7 (glycerol) |
| Root mean square deviation from target values | Bonds (Å) 0.012 |
| | Angles (°) 1.29 |
| | Bonded B-factors (Å²) main chain/side chain 0.91/2.34 |
| Average B-factors for protein (Å²) | 18.4                                      |

Figures in parentheses refer to the outermost resolution shell, unless otherwise indicated.

Rmerge = Σ||I(hkl)−⟨I(hkl)⟩|/ΣI(hkl), where I(hkl) is the ith intensity measurement of reflection hkl, including symmetry-related reflections, and ⟨I(hkl)⟩ is its average.

Rfactor = Σ||Fcalc−kFobs||/ΣFcalc, with Fcalc and Fobs as the observed and calculated structure factor amplitudes. The free Rmerge is the same for a test set of reflections (>500) not used during refinement.

Including atoms present with alternate occupancy.
Autolytic Activation of Ulilysin Is Mediated by Calcium—Both the full-length protein of 38 kDa and its selenomethionine variant are monomeric in solution and, once purified, stable over weeks, with no traces of degradation even at 37 °C (Fig. 2A). However, the addition of calcium causes a band of 29 kDa to appear in both cases (Fig. 2B). The full-length protein is inactive (see below and Fig. 2C), but preincubation with calcium renders activity after an initial lag phase. In turn, the 29-kDa form, which is also monomeric, is highly active. The full-length protein crystallizes in the presence of a high concentration of CaCl₂. N-terminal Edman degradation and mass spectrometry analyses of the crystalline material reveal that it contains segment Arg₆₁–Ala₃₂₂ (Fig. 1), revealing that the full-length protein was cleaved twice before crystallizing. The amino acid sequence around these two cleavage sites shows a similar motif, with an arginine in the P₁/H₁ position, the first position downstream of the scissile bond (nomenclature according to Ref. 26; see also Fig. 2E). These data indicate that the 38-kDa form is the inactivezymogen, proulilysin, which undergoes calcium-triggered autoactivation to the 29-kDa form, mature ulilysin. Autolytic activation entails removal of the first 60 residues, which would correspond to the pro-domain, and of a highly charged, 20-residue C-terminal tail that contains five glutamate and six arginine/lysine residues. Within the pro-domain, a potentially free cysteine (Cys₂₃; Fig. 1) could play the role of a “cysteine switch” or “Velcro” element, where the cysteine S atom coordinates the catalytic zinc ion in the zymogen and prevents substrates from binding to the active site. Such mechanisms for the maintenance of latency have been described in other metzincins (13, 27, 28).

Ulilysin Is a Functional Metallopeptase and a Specific and Selective IGFBP Protease—Ulilysin cleaved IGFBP-2, -3, -4, -5, and -6, as well as insulin, in an IGF-independent and mostly specific manner, but not IGFBP-1 or IGF-II (Fig. 2D). Furthermore, IGFBP-4 and -5 proteolysis occurred more efficiently in vitro than when mediated by human PAPP-A and PAPP-A2. Ulilysin produced comparable cleavage levels at much lower doses and reaction times than the human enzymes, as inferred from SDS-PAGE (data not shown). In addition, ulilysin had an optimum pH of around 6 and performed limited proteolysis of a whole series of substrates, including casein and extracellular matrix (derived) components like azocollagen, gelatin, and fibronectin (Table 1). Natural collagens of type I from kangaroo tail and of type V were also cleaved, but not collagen type I from human placenta or collagen type IV. However, these results must be taken with care as cleavage occurred at 37 °C...
but not at room temperature. Furthermore, these substrates were also cleaved by trypsin in the same conditions, thus indicating that the purchased samples contained gelatinous material. In this context, it is noteworthy that extracellular matrix components interact with IGFBP-5 and modulate its activity in vivo (1). Moreover, the rather proteolyis-resistant skeletal proteins actin and elastin were also cleaved, as was fibrinogen from the blood coagulation cascade, but not plasmin or α1-antitrypsin. N-terminal sequence analyses of the major fragments obtained after proteolysis of insulin and fibrinogen confirms the specificity for arginine in P1', (Fig. 2E), as already suggested by the autolytic cleavage points, and that the enzyme has a substrate preference with the pattern B'X-XRb(E/Q) (B, bulky hydrophobic or aromatic).

Utilelsin was only inhibited by unspecific MP inhibitors like the zinc chelators o-phenanthroline and EDTA, and by excess zinc (Table 2). Partial inhibition was only effected by batimastat and CT1746, which are broad spectrum small molecule inhibitors of MMPs. The preference for an arginine in the P1' position also led us to test i-arginine and two arginine-based molecules that inhibit serine proteases and carboxypeptidases, benzamidine, and guanidinoethylmercaptosuccinic acid. None of them caused significant inhibition. These data indicate that utilelsin is both a specific and an efficient metalloprotease that targets selected IGFBPs and extracellular matrix proteins and that selective inhibitors remain to be found.

Overall Structure of Ulilysin—Utilelsin is ellipsoidal, with an α/β topology within a polypeptide chain that runs from residue Arg61 to Ala322 (Fig. 3, a and b). The protein is partitioned into two moieties separated by an extended active site cleft running from left to right, namely an upper regular N-terminal subdomain (NTSD; Arg61–Asn235) and a rather irregular C-terminal lower domain (CTSD; Leu236–Ala322). NTSD starts on the back of the molecule and enters, through strand β1, a strongly twisted, mainly parallel five-stranded β-sheet (Figs. 3a and 4A). Following β1, helix α1 runs on the back of the molecule from the upper right to the lower left and finishes at the junction with the CTSD. Here, the polypeptide chain follows an extended loop (Lys108–Glu130) whose tip almost reaches the bottom of the molecule, resembling a cape over the back of the CTSD. This segment features two short β-1,4-helices (α2 and α3 in Figs. 3a and 4A). The peptide chain rejoins the NTSD at the second β-strand that is subdivided into two, named β2 and β3. They are separated by an insertion (Thr127–Thr136), here termed the “LNR-segment of a larger peptide and is bound to the protein mainly through hydrogen bonds with Gln305 of the C-terminal helix α5 (see below), and Tyr295 points into the NTSD hydrophobic cluster. These two interactions keep the Met turn in position under the zinc-coordinating residues, where it creates a hydrophobic pillow but does not contact the cation nor its ligands (see below). After the Met turn, the polypeptide chain turns round to reach the back surface of the protein and enters the C-terminal helix α5 (Fig. 3a). Approximately in its middle, Arg235 points into the interior of the molecule contacting both Asn229 Oβ1 and O atoms, thus contributing to structural integrity, together with the previously mentioned Asn208–Gln305 interaction. Helix α5 ends up at the molecular surface with the C-terminal residue, Ala322, solvent exposed and in the proximity of the N terminus (15.4 Å between the respective Ca atoms).

The Active Site Cleft—The active site of utilelsin traverses the molecule from left to right (Fig. 3, a and b). Its tip is framed by the antiparallel strand β2 of NTSD, which runs antiparallel to a bound peptide substrate, and by the β5β6-ribbon, which projects away from the region of the active site cleft accommodating the primed side substrate residues. The bottom of the cleft is paved by segments Tyr237–Trp240, Pro241–Gly245, the Met turn, and the subsequent four residues (Asn246–Asp249), all within the CTSD. A segment constituted by the first part of the calcium-binding double S-loop (Asp240–Arg252) protrudes from the molecular surface and further influences substrate binding, in this case mainly on the nonprimed side of the active site. The catalytic zinc cation (Zn2+), tethered by a solvent molecule and the Ne2 atoms of His229, His231, and His238 from the ZBCS, imbedded in the active site helix α4 (Figs. 3, a, b, and 4A). Binding distances range from 2.0 to 2.1 Å. The solvent molecule bound to the catalytic zinc is further anchored to the general base, Glu229 of the ZBCS. The side chain of Tyr295, immediately after the Met turn, is also close to the catalytic cation (Fig. 3b) but is swung out from its (probable) zinc-liganding position, as observed in other metzincins upon substrate binding (29, 30). In the present structure, this is due to the presence of a peptide occupying the primed side of the active site cavity, probably left behind after a proteolytic event during purification or crystallization. The electron density map unambiguously identifies the residue penetrating the deep S1′ pocket as an arginine (Arg401), followed by a possible valine fitting into S2′ (Val146). This dipeptide may represent the ordered N-terminal segment of a larger peptide and is bound to the protein mainly through two inter-main chain hydrogen bonds with the upper rim strand β7 (Arg401–Glu199 O and Arg401–O-Leu188 N). The presence of Arg401 allows us to identify the residues shaping the specificity pocket as Thr225 from α4, Leu188 from β7, Phe229 from Lys84, Met299 from the extended segment preceding α5, and the main chain from Tyr291 to Asp295 (Fig. 3b). Of particular importance for specificity is the latter aspartate at the bottom of the pocket. It strongly binds, through its Oδ1 atom, both Arg401 Nγ1 and Nγ2 atoms. The latter atom further contacts Val146 O. The nature of this cavity, mostly hydrophobic except for the pocket...
bottom, is ideally conceived to accommodate an arginine residue, thus explaining the strong preference of ulilysin for such a residue in P₁/H₁₁₀₃ (see above and Fig. 2E). The bound peptide also indicates that the S₂/H₁₁₀₃ pocket is much shallower and mainly created by the aromatic surface of Tyr₁₁₀₂, as well as by Gln₁₈₅ and Ile₁₈₇ from the 5₆ ribbon.

The Double Calcium-binding Site—Two calcium cations are present in the ulilysin CTSD, 9.1 Å apart (Figs. 3a and 4, B–D). (Pro)ulilysin is inactive in their absence and rigid in their presence. It is reversibly inhibited by EDTA (Table 2) and reactivated by dialysis against a calcium-containing buffer (data not shown). These data indicate that calcium is a switch for this protease.

The first site is centered on Ca₉₉₈ and shows eight oxygen ligands, five approximately in a plane, two apical ligands on one side of the plane, and another on the opposite side (Fig. 4, B and C). Four ligands are provided...
by the protein, and four are solvent molecules, and coordinating distances range between 2.3 and 2.6 Å. This site is reminiscent of that on Ca$^{2+}$ of thermolysin (Protein Data Bank code 8tln) (31), although in the latter case there is only one solvent ligand, and a second calcium ion is just 3.8 Å away. The second ulilysin calcium Ca$^{2+}$ has four ligands in a plane with the cation and, again, two apical ligands on one side and another at the opposite apical position (Fig. 4, B and D). Here, only one of the seven oxygen atoms comes from a solvent molecule. In this case, the site is reminiscent of the EF hands seen in calbindin 9K (32). However, in ulilysin this calcium site is not flanked by the characteristic helix-turn-helix motif found in EF hands. Taken together with the results of structural similarity searches, these findings suggest that the region of ulilysin encompassing the calcium binding sites is novel.

**Implications for the Other Members of the Pappalysin Family**—Bioinformatic sequence similarity searches suggest that pappalysins should be grouped into protease family M43 in the MEROPS data base, into the ProDom family PD332581, and into the Pfam family PF05572. Furthermore, these searches permitted the identification of two groups of sequences. The first included close relatives of human PAPP-A, with E values below 3E-20 (see Ref. 33), from several mammals, birds (chicken), fish (zebrafish and green-spotted pufferfish), amphibians (African clawed frog and pipid frog), and echinoderms (sea urchin). Besides, a second group comprised sequences with 9E-12/E value > 7E-4 from fungi (Pleurotus ostreatus PoMTP, Coccioidiodes posadasii MEP1, Ustilago maydis, Aspergillus nidulans and fumigatus, Magnaporthe grisea, Neurospora crassa, Gibberella zeae, and Metarhizium anisopliae), bacteria (Cytophaga sp. and hutchinsonii cytaphaglysin and sequences from Gloiobacter violaceus, Shewanella sp. and amazonensis), and archaea (M. acetivorans ulilysin). Among these, Mep1 (SwissProt code Q71H76) from the fungal pathogen C. posadasii, which can cause the respiratory San Joaquin Valley fever in immunocompromised humans and animals, is the only member, apart from human PAPP-A and -A2,
to have been assessed biochemically and in vivo. The 283-residue (pro)Mep1 is secreted during endospore differentiation within the host, and it digests a host-cell surface antigen, in this way preventing recognition of endospores by host phagocytes (34). Another member studied in vitro is PoMTP metalloprotease from the edible oyster mushroom, P. ostreatus (SwissProt code Q5Y972). The mRNA of this 290-residue (pro)protein is abundant at primordial and fruit body stages, thus suggesting a role in mushroom fruiting (35). These authors proposed that this protein should be grouped with a series of putative fungal orthologues in a separate metzinck family termed eucolyxins. Finally, another member is cytolygalysin, a bacterial collagenase obtained from Cytophaga sp. L43–1 (36). It is a polypeptide of 1,282 amino acid residues, putatively synthesized from cog as a zymogen. It is capable of digesting both insoluble and acid-soluble collagens and gelatin, as well as casein (37). Whereas most prokaryotic and fungal forms merely span the catalytic domain (plus a putative pro-domain), cytolygalysin encompasses a further ~950 residues C-terminal to the catalytic domain, a stretch of similar length to the one found in the vertebrate papalysins, thus potentially encompassing additional domains of distinct function (Fig. 1). Detailed inspection of selected forms reveals that papalysins present a higher similarity within their CTSDs, because most determinants of specificity are found within this subdomain. In the NTSD, the main structural features for domain integrity are the two internal hydrophobic clusters on either side of the central β-sheet. Strict residue conservation is dispensable because compensating substitutions may still maintain these clusters. Key residues for catalysis, substrate binding, and structural stability are absolutely conserved, such as Asp295, which determines the substrate specificity in P1’, and Asn268, at the beginning of the Met turn, which forms a hydrogen bond with Glu305 important for structural integrity. The same holds for Arg308, engaged in binding of the main chain at Asn235, and for the aforementioned duos: Thr302–Asp258 and Trp165–Asn172. In the calcium-binding site, most ligands are provided by solvent molecules or main chain carbonyl oxygen atoms in ulilysin (Fig. 4, B–D). Only three protein side chains participate: Asp254, Thr309, and Glu243. The first two are strictly conserved, whereas the third could be replaced by another conserved glutamate, e.g. Glu580 (following the PAPP-A numbering; see Fig. 1). This correlates well with the finding that PAPP-A activity, like ulilysin activity, depends on calcium (9).

As previously discussed, the start of the mature ulilysin sequence suggests that a pro-domain with a putative cysteine switch might be present in papalysins (Fig. 1). The disulfide bond pattern of ulilysin and, most likely, the closer bacterial and fungal relatives includes two SS bonds in the CTSD. Although this double clamp affects the region where all papalysins show the closest sequence similarity, the SS pattern reported for PAPP-A and the other more closely related vertebrate members diverges (38). This difference may be intrinsic and could therefore be compatible with the maintenance of the overall fold. Alternately, the activity of PAPP-A in vivo may be dependent on reducing agents, putatively accounting for changes in the disulfide bond pattern (8). Accordingly, reduction or formation of specific SS bonds may regulate the extracellular PAPP-A activity, and the SS bonds may show a pattern similar to those in ulilysin in certain conditions.

There is little sequence consensus between the cleavage points identified in ulilysin and those reported for PAPP-A. Further, aside from IGFBP3s, no other protein or low molecular weight substrates have been identified for the human enzyme. It has thus been proposed that steric regulation could account for substrate specificity, possibly mediated by the three LNR motifs, two of which are inserted into the catalytic protease domain (Fig. 1) (9). These LNR sequences are absent from bacterial orthologues (Fig. 1), but the region of insertion fully coincides with the LNR-like loop observed between strands β2 and β3 in ulilysin. Accordingly, this protruding 10-residue loop may be shorter, or even absent, or much longer, featuring the ~66-residue insertion of PAPP-A, than in ulilysin. In any case, it is clear that this insertion would be compatible with the overall structure of the protease moiety and that it would occur on the surface, thus in a potential disposition to carry out binding functions.

**Structural Similarities with Other MPs**—Ulilysin bears structural similarity with members of the adamalysin/ADAM family within the metzinck clan of MPs. In particular, the most closely related structures are those of ADAM-17/TACE (Protein Data Bank code 1bkc) (39) and ADAM-33 (Protein Data Bank code 1r54) (40). A total of 177 and 167 residues can be structurally aligned with an root mean square deviation over all atoms of 3.0 and 2.8 Å, respectively, despite negligible sequence similarity (12 and 16%) (Fig. 3c). These human MPs are engaged in cancer, inflammation, and modulation of the immune response and in asthma, respectively. Members of this family are IGFBP proteases. Likewise, adamalysins/ADAMs are subdivided into an NTSD and a CTSD, separated by an active site cleft harboring the catalytic zinc ion. Generally, the structural similarity depends on a series of conserved regular secondary structure elements (Fig. 3c). In detail, similarity is high within each polypeptide chain around the long ZBCS, with a superimposable catalytic base and zinc-liganding histidines. However, the structures of the CTSDs deviate strongly. Furthermore, adamalysins/ADAMs and MMPs strongly prefer bulky hydrophobic residues in P1’ of substrates, whereas ulilysin and probably other papalysins favors an arginine. Finally, the possible presence of a fifth tyrosine ligand in ulilysin that may be swung out upon substrate binding is shared with astacins and serralysins (41, 42).

**Conclusions**—We have discovered and studied a new MP that belongs to the family of the papalysins, characterized by the prototypical human PAPP-A, which was identified as a specific IGFBP protease. In contrast with PAPP-A, ulilysin displays a characteristic cleavage pattern preference and broad substrate specificity and efficiency. Recently, it has been reported that procollagen C proteinase, a multidomain MP of the astacin family of metzincins, also shows a much broader substrate profile and higher efficiency in cleaving a variety of extracellular matrix proteins than the full-length protein. This difference has been attributed to the additional domains present in the latter that may modulate and restrict the substrate specificity through their potential protein binding or steric hindering competence (43). Our results, taken together with the extensive work performed on human PAPP-A, suggest a similar scenario, where activity of a ulilysin-like protease domain could be restrained by the LNR and CCP motifs and further domains potentially contained in the additional ~1,200 residues of the vertebrate forms.

**Acknowledgments**—Protease inhibitors 409.1, RXP03-R1, and galardine were a kind gift from V. Dive (Paris, France), batimatast was from K. Maskos/ W. Bode (Munich, Germany), and CT1746 was from G. Murphy (Cambridge, UK). We are further indebted to C. Oxvig (Aarhus, Denmark) for providing IGFBP-5 and are grateful for the help provided by the EMBL and European Synchrontron Radiation Facility synchrotron local contacts, in particular by Gavin Fox, during data collection (European Synchrontron Radiation Facility, Grenoble, Grenoble). Robin Rycroft is thanked for helpful contributions to the manuscript, and Francesc Canals is thanked for help with the proteomics experiments.

**REFERENCES**

1. Bach, L. A., Heady, S. J., and Norton, R. S. (2005) Trends Endocrinol. Metab. 16, 228–234
2. Zapf, J. (1995) Eur. J. Endocrinol. 132, 645–654
3. Burger, A. M., Leyland-Jones, B., Banerjee, K., Spyropoulos, D. D., and Seth, A. K. (2005) Eur. J. Cancer 41, 1515–1527
4. Lawrence, J. B., Oxvig, C., Overgaard, M. T., Sottrup-Jensen, L., Gleich, G. J., Hays, L. G., Yates 3rd, J. R., and Conover, C. A. (1999) Proc. Natl. Acad. Sci. U.S.A. 96, 17927

JUNE 30, 2006 • VOLUME 281 • NUMBER 26

Molecular Analysis of Ulilysin Metalloprotease
Molecular Analysis of Ulilysin Metalloprotease

3149–3153
5. Boldt, H. B., Overgaard, M. T., Laursen, L. S., Weyer, K., Sotrup-Jensen, L., and Oxvig, C. (2001) Biochem. J. 358, 359–367
6. Oxvig, C., Overgaard, M. T., and Sotrup-Jensen, L. (2004) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D., and Woesner, Jr., J. F., eds) Vol. 1, 2nd Ed., pp. 754–757, Elsevier, London
7. Gall, S. A., and Halbert, S. P. (1972) Int. Arch. Allergy Appl. Immunol. 42, 503–515
8. Glerup, S., Boldt, H. B., Overgaard, M. T., Sotrup-Jensen, L., Giudice, L. C., and Oxvig, C. (2005) J. Biol. Chem. 280, 9823–9832
9. Boldt, H. B., Kjær-Sorensen, K., Overgaard, M. T., Weyer, K., Poulsen, C. B., Sotrup-Jensen, L., Conover, C. A., Giudice, L. C., and Oxvig, C. (2004) J. Biol. Chem. 279, 38525–38531
10. Overgaard, M. T., Boldt, H. B., Laursen, L. S., Sotrup-Jensen, L., Conover, C. A., and Oxvig, C. (2001) J. Biol. Chem. 276, 21849–21853
11. Bode, W., Gomis-Ruth, F. X., and Stöcker, W. (1993) FEBS Lett. 331, 134–140
12. Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B., and Bode, W. (1995) Prot. Sci. 4, 823–840
13. Gomis-Ruth, F. X. (2003) Mol. Biotech. 24, 157–202
14. Fang, P., Hwa, V., and Ronenstein, F. (2004) Novartis. Found Symp. 262, 215–234
15. Nakamura, M., Miyamoto, S., Maeda, H., Ishii, G., Hasebe, T., Chiba, T., Azaka, M., and Ochiai, A. (2005) Biochem. Biophys. Res. Commun. 333, 1011–1016
16. Blobel, C. P. (2005) Nat. Rev. Mol. Cell Biol. 6, 32–43
17. Leslie, A. G. W. (1999) Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1696–1702
18. Evans, P. (1993) in Data Collection and Processing: Proceedings of the CCP4 Study Weekend, January 29–30, 1993 (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 114–122, SERC Daresbury Laboratory, Warrington
19. Collaborative Computing Project 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
20. Sheldrick, G. M., Hauptmann, H. A., Weeke, C. M., Miller, R., and Usón, I. (2001) in International Tables for Crystallography. Arnold, E., and Rossmann, M. G., eds Vol. F, pp. 333–351, Kluwer Academic Publishers; Dordrecht, The Netherlands
21. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 2023–2030
22. Carranza, C., Insern, A.-G., Mouthy-Knoops, E., Camilliaux, C., and Roussel, A. (1999) in AFMB Activity Report 1996–1999, pp. 89–90, CNRS-UPR 9039, Marseille, France
23. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255
24. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
25. 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) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
26. Schechter, I., and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162
27. Springman, E. B., Atten, E., Birkefeld-Hansen, H., and Van Wart, H. E. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 364–368
28. Loechel, F., Overgaard, M. T., Oxvig, C., Albrechtsen, R., and Wewer, U. M. (1999) J. Biol. Chem. 274, 13427–13433
29. Grams, F., Dive, V., Iyiotaikis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stöcker, W. (1996) Nat. Struct. Biol. 3, 671–675
30. Baumann, U., Bauer, M., Letoffe, S., Depelepaire, P., and Wandersman, C. (1995) J. Mol. Biol. 248, 653–661
31. Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., and Dupoureque, D. (1972) Nature 238, 37–41
32. Szebenyi, D. M., Obendorf, S. K., and Moffat, K. (1981) Nature 294, 327–332
33. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
34. Hung, C.-Y., Seshan, K. R., Yu, J.-J., Schaller, R., Xue, J., Basrur, V., Gardner, M. J., and Cole, G. T. (2005) Infect. Immun. 73, 6689–6703
35. Joh, J.-H., Kim, B.-G., Kong, W.-S., Yoo, Y.-B., Kim, N.-K., Park, H.-R., Cho, B.-G., and Lee, C.-S. (2004) FEBS Microbiol. Lett. 239, 57–62
36. Sasagawa, Y., Ikazaki, K., Matsubara, Y., Suzuki, K., Kojima, H., and Kamiyo, Y. (1995) Biosci. Biotechnol. Biochem. 59, 2068–2073
37. Kamiyo, Y., and Sasagawa, Y. (2004) in Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D., and Woesner, Jr., J. F., eds) Vol. 1, pp. 807–808, Elsevier, London
38. Overgaard, M. T., Sorensen, E. S., Stachowiak, D., Boldt, H. B., Kristensen, L., Sotrup-Jensen, L., and Oxvig, C. (2003) J. Biol. Chem. 278, 2106–2117
39. Maskos, K., Fernández-Catalán, C., Huber, R., Bourenkov, G. P., Bartunk, H., Ellstedt, G. A., Reddy, P., Wolfson, M. F., Rauch, C. T., Castner, B. J., Davis, R., Clarke, H. R. G., Petersen, M., Fitzner, J. N., Cerretti, D. P., March, C. J., Paxton, R. J., Black, R. A., and Bode, W. (1998) Proc. Natl. Acad. Sci. U.S.A. 95, 3408–3412
40. Orth, P., Reichert, P., Wang, W., Prosi, W. W., Varosh-Tomaine, T., Hammond, G., Ingram, R. N., Xiao, L., Mirza, U. A., Zou, J., Strickland, C., Taremi, S. S., Le, H. V., and Clarke, 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) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921