Architecture and function of metallopeptidase catalytic domains

Núria Cerdà-Costa & F. Xavier Gomis-Ruth *

Proteolysis Lab; Department of Structural Biology; Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park; Helix Building; c/ Baldiri Reixac, 15-21; E-08028 Barcelona (Spain).

* Correspondence: e-mail: fxgr@ibmb.csic.es, phone: (+34) 934 020 186, fax: (+34) 934 034 979.

The authors state they have no competing financial interest.

Keywords: Structural biochemistry, metzincin clan, catalytic domains, active-site cleft, hydrolytic enzymes, matrix metalloproteases, astacins, ADAM, adamalysins, serralysins, metalloprotease, metalloproteinase.
ABSTRACT

The cleavage of peptide bonds by metallopeptidases (MPs) is essential for life. These ubiquitous enzymes participate in all major physiological processes, and so their deregulation leads to diseases ranging from cancer and metastasis, inflammation, and microbial infection to neurological insults and cardiovascular disorders. MPs cleave their substrates without covalent intermediate in a single-step reaction involving a solvent molecule, a general base/acid, and a mono- or dinuclear catalytic metal site. Most monometallic MPs comprise a short metal-binding motif (HEXXH), which includes two metal-binding histidines and a general base/acid glutamate, and they are grouped into the zincin tribe of MPs. The latter divides mainly into the gluzincin and metzincin clans. Metzincins consist of globular ~130-270-residue catalytic domains, which are usually preceded by N-terminal pro-segments, typically required for folding and latency maintenance. The catalytic domains are often followed by C-terminal domains for substrate recognition and other protein-protein interactions, anchoring to membranes, oligomerization, and compartmentalization. Metzincin catalytic domains consist of a structurally conserved N-terminal sub-domain spanning a five-stranded β-sheet, a backing helix, and an active-site helix. The latter contains most of the metal-binding motif, which is here characteristically extended to HEXXHXXGXX(H,D).

Downstream C-terminal sub-domains are generally shorter, differ more among metzincins, and mainly share a conserved loop—the Met-turn—and a C-terminal helix. The accumulated structural data from more than 300 deposited structures of the 12 currently characterized metzincin families reviewed here provide detailed knowledge of the molecular features of their catalytic domains, help in our understanding of their working mechanisms and form the basis for the design of novel drugs.
1. PEPTIDASES

The term “peptidase” is currently recommended for proteolytic enzymes, proteases, and proteinases in general \(^1\). Peptidases are distributed among all kingdoms of life \(^2,3\). They target peptide bonds of proteins and/or peptides and some catalyze extensive protein-processing events such as digestion or degradation of intake proteins, and development, maintenance and remodeling of tissues. Other peptidases catalyze limited, specific scission of a small number of peptide bonds, which results in the activation of deactivation of themselves and other (pro)enzymes, bioactive peptides, and DNA repressors \(^2\). In this way, peptidases participate in regulatory mechanisms, which are required in physiological processes such as blood-pressure control, hormone homeostasis, regulation of signal-transduction pathways, modulation of protein-protein and cell-cell interactions \(^4\). Such limited proteolysis through “sheddases” produces soluble forms from membrane-anchored precursors, thus decreasing protein concentration at cell surfaces and increasing it in the circulation \(^5\)-\(^7\).

Peptidases have traditionally been classified into serine, cysteine, aspartic, and metallopeptidases (MPs) based on elements key for catalysis \(^8\)-\(^10\). More recently, three categories have been added: N-terminal threonine peptidases, glutamate peptidases, and asparagine peptidases \(^11\)-\(^12\); \(^1\). In addition, peptidases are divided into endopeptidases and exopeptidases, which cut in the middle or at the ends of substrates, respectively \(^13\). The latter, in turn, can be further classified into aminopeptidases and carboxypeptidases, which cleave off N-terminal and C-terminal amino acids, respectively \(^1\). Overall, exquisite regulation of any proteolytic enzyme is essential for proper functioning and to prevent misdirected temporal and spatial proteolytic activity. Control may be exerted at the transcriptional level or via post-translational modifications. Other regulatory mechanisms include biosynthesis in the form of zymogens that require activation \(^14\), co-localization of enzyme and substrate, cofactor binding and substrate accessibility and specificity, as well as the presence of physiological protein inhibitors \(^15\); \(^16\). Failure of regulatory mechanisms can give rise to pathologies such as inflammation, tissue destruction (as in arthritis and fibrotic diseases), neurological diseases (Alzheimer’s disease, meningitis and multiple sclerosis), and cardiovascular disorders (stroke, hypertension, thrombosis, bleeding, and myocardial infarct)\(^17\)-\(^19\). Even in tumorigenesis and tumor-progression events, such as angiogenesis, tissue invasion and metastasis, peptidases play a major role \(^20\); \(^21\). A particular group of peptidases are exogenic virulence factors that lead to the depletion of host defenses and tissue destruction \(^22\)-\(^24\). This is observed during microbial infections that cause inflammation, tetanus, pneumonia, tuberculosis, AIDS, malaria, botulism, gas gangrene, bacterial meningitis, anthrax, etc. but also after poisoning. Accordingly, such diverse physiopathological implications make these proteins promising drug targets \(^25\) and their structural studies are thus essential to the understanding of their functional determinants and the design of novel therapeutic agents, now grouped under the term “degradomics” \(^26\); \(^27\); \(^4\); \(^28\); \(^29\).

Here we review the catalytic domains of MPs, in particular from the metzincin clan \(^30\)-\(^33\); \(^28\); \(^34\)-\(^37\). To date, structures are available for twelve metzincin families: astacins, adamalysins/ADAMs, serralysins, matrixins
(alias matrix metalloproteinases and MMPs), leishmanolysins, snapalysins, pappalysins, archaemetzincins, fragyllysins, cholerilysins, toxilysins, and igalysins (for the first collective mention of the first ten family names, see 20).

2. CLASSIFICATION AND CLEAVAGE MECHANISM OF METALLOPEPTIDASES

One widely accepted classification of MPs and proteolytic enzymes in general is provided by the MEROPS database, which classifies homologous sets of peptidases and protein inhibitors hierarchically into protein species, families, and clans based on sequence similarity and evolutionary distances (http://merops.sanger.ac.uk; 38). This scheme is also followed in a comprehensive study, the Handbook of Proteolytic Enzymes 39. Here, we apply a different approach based on active-site architecture and overall fold similarity (see also 40; 30; 41; 42; 28; 43; 35). According to this scheme, MPs are first divided into those that have a single catalytic metal, and those that have two. Dimetalate MPs mainly include exopeptidases and these are reviewed elsewhere 44-46, so here we focus on structurally-characterized members of the subclass of mononuclear MPs (Fig. 1). In addition, isopeptidases such as Jab1/MPN 47; 48 and AMSH-LP deubiquitinases 49, which cleave amide bonds that are not peptide bonds, and peptide deformylase, which removes N-terminal formyl groups from N-terminal methionine residues 50; 51, have also been omitted. This classification is based on recent structural data and thus supersedes previous reports by our group 28; 43; 52.

The active site of MPs accommodates a catalytic divalent metal ion, mostly zinc but also sometimes cobalt, manganese or nickel, which is anchored at the bottom of the active-site cleft of the enzyme by protein residue side chains 41; 53. Here it might be appropriate to clarify commonly-used terms which tend to be misused in the literature. The intact, active, metal-bound protein is described as the “enzyme” or the “holoenzyme”. If the enzyme has a ligand bound it can be described as an enzyme-ligand complex. The enzyme lacking a metal ion indispensable for catalysis is described as an “apoenzyme”. Metal-binding residues are mostly histidines, aspartates, and glutamates, which are included in zinc-binding motifs that are characteristic of particular MP tribes, clans and families (Fig. 1, residues in green). Tribes include αβα-exopeptidases, which comprise the funnelin clan 43; 52; LAS MPs 54-56; relatives of hydrogenase maturating factor, HybD, a proven natural nickel-dependent MP 57; 58; zincins 30; 41; 59; and inverzincins 60; 61; 11; 62. In all cases, a general/base acid, mostly a glutamate, is further found close to the catalytic metal and is required for catalysis (see section 2 and Fig. 1, residues in magenta).

In general, peptide-bond hydrolysis through monometallic MPs is an ordered single-displacement reaction that follows simple Michaelis-Menten kinetics and occurs optimally at neutral pH. A solvent molecule and a peptide substrate are bound by the active-site cleft and the catalytic metal ion to render the Michaelis complex 63; 64. In most cases, the overall active-site environment is in a competent conformation for catalysis in the mature enzyme before recruiting the peptidic substrate, which binds in extended conformation in such a way
that substrate side chains upstream (termed P₁, P₂, P₃, etc.) and downstream (P₁', P₂', P₃', etc.) of the scissile bond nestle into active-site cleft sub-sites S₁, S₂, S₃, etc. and S₁', S₂', S₃', etc. of the enzyme, respectively. In general, substrate specificity is exerted in MPs through S₁'. Cleavage can also occur within protein segments adopting compact tertiary structures, even helical conformations, in the unbound state, as interaction and binding to the peptidase can energetically compensate for substrate unwinding to yield an extended conformation suitable for catalysis. A consensus mechanism for catalysis has been derived after decades of research, mainly on thermolysin from *Bacillus thermoproteolyticus* and bovine carboxypeptidase A, which were the first two MPs to be solved for their crystal structures in the late 1960s/early 1970s, and more recently on MMPs. This mechanism entails that the solvent molecule is bound first—perhaps already part of the mature functional enzyme prior to catalysis as found in the structures of competent MPs in the absence of inhibitors or substrates—by the catalytic metal ion and the general base/acid glutamate, which polarizes the solvent and enhances its nucleophilicity (Fig. 2, I). Once the substrate to be cleaved is bound to the cleft, the scissile carbonyl oxygen is bound and polarized by the catalytic metal, so the carbonyl carbon is more susceptible to nucleophilic attack by the solvent molecule, which simultaneously transfers a proton to the general base glutamate (Fig. 2, II). The catalytic metal thus serves as a dual oxyanion hole and activates both substrates for the reaction. The attack gives rise to a gem-diolate tetrahedral reaction intermediate (Fig. 2, III), which is stabilized by the metal but also by other protein residues such as vicinal histidines, arginines and tyrosines (Fig. 1; see also section 4). Subsequently the intermediate resolves to the double-product complex through scissile bond cleavage and double proton transfer to the new α-amino terminus (Fig. 2, IV). Finally, the two products dissociate from the enzyme and leave the active-site cleft (the non-primed product possibly first, see), which is replenished with another solvent molecule and ready for a new round of catalysis.

### 3. ZINCINS SPLIT INTO GLUZINCINS AND METZINCINS

The zincins were named in 1993 by Bode and colleagues. They constitute the most extensive tribe of MPs, with >100 coding genes in humans alone. They have a characteristic short metal-binding amino-acid consensus sequence, HEXXH, first identified by McKerrow in MMPs and serralysins in 1987, which is provided by an “active-site helix” and supplies two metal-binding histidine ligands and the general base/acid glutamate (see section 2). In addition, zincins are divided by a horizontal active-site cleft into an upper N-terminal sub-domain (NTS) above the cleft and a lower C-terminal sub-domain (CTS) below the cleft when viewed in the standard orientation of MPs, i.e. with the peptide substrate running from left to right. The NTS has a helix termed “backing helix” in addition to the active-site helix and a β-sheet of at least three strands, which are mixed parallel/antiparallel. The lowermost strand horizontally frames the top of the active-site cleft, runs antiparallel to a substrate when bound to the cleft, and anchors the latter through inter-main-chain interactions mimicking a β-ribbon. The CTS varies greatly among zincins and is mainly characterized by the
presence of a residue (Fig. 1, orange residues) contained in a loop structure providing a basement for the metal-binding site. The two main zincin clans are the gluzincins and the metzincins, whose loop structures are termed, respectively, “Ser/Gly-turn” and “Met-turn” (see below and section 4). Three other zincin clans have an aspartate as third zinc ligand: (i) The S2P-zincins, which include one of the only two integral-membrane MP families structurally characterized to date (those completely imbedded in membranes and not merely type-I or type-II membrane-anchored species); their fold is different from any other zincin, although they have the active-site helix and the short consensus sequence $^{80}$; $^{52}$, and they show a proline below the metal. (ii) The aspzincins, which were named in 1999 by Ichishima and co-workers $^{81}$ in line with the terms gluzincins and metzincins (see below) due to the aspartate metal ligand $^{81}$-$^{83}$ and have an alanine below the metal. (iii) The FtsH-related enzymes, which do not have a protein residue below the metal but rather a solvent molecule $^{84}$-$^{87}$.

Gluzincins were so named because of a glutamate found as third metal-binding residue, as observed in archetypal thermolysin $^{72}$. The clan name was coined by Hooper in 1994 $^{41}$ to distinguish the latter enzyme and related structures from the metzincins, which had a histidine instead the glutamate (see below) and had been named in 1993 by Bode and colleagues based on structural similarities between astacin, adamalysin II and the serralysin Pseudomonas aeruginosa alkaline proteinase $^{30}$. Since then, a number of novel MP structures have become available, which has led us to suggest a redefinition of the common structural features of gluzincins: they are zincins, which in addition to their common structural elements contain a “glutamate helix” below the active-site helix, which runs parallel to it in a horizontal projection but is rotated backwards in a vertical projection (see Fig. 3 in $^{47}$). This helix—or a short loop segment immediately preceding it—provides the third metal-binding residue, mostly a glutamate but sometimes also a histidine as in minigluzincins $^{88}$. This residue is three positions upstream of the residue of the Ser/Gly-turn providing the basement of the metal-binding site $^{43}$ (Fig. 1, orange residues). The latter residue varies within gluzincins: it is a phenylalanine in minigluzincins $^{88}$; a threonine in clostridial neurotoxins $^{89}$; $^{90}$; a serine in thermolysins $^{72}$; $^{63}$; $^{91}$-$^{93}$; and a serine or a glycine in cowrins $^{43}$. However, the most widespread residue among gluzincin families is an alanine as found in anthrax lethal factor and related proteins such as Escherichia coli titration factor Mlc (Mtfα/Yeel)$^{94}$; $^{95}$; in M48/M56-family MPs $^{38}$; $^{88}$, which with human FACE1 and yeast Ste24p are the only other integral-membrane MP family in addition to S2P-zincins structurally analyzed to date $^{96}$; $^{97}$; in some members—others have a threonine—of the family encompassing leukotriene A4 hydrolase, tricorn interacting factor F3, and MEROPS M1-family aminopeptidases $^{98}$-$^{100}$; in bacterial collagenases such as those from Clostridium spp. $^{101}$; $^{102}$; in nephrilysins $^{103}$; $^{104}$; and in dipeptidyl peptidase III and its structural relatives, which uniquely have an additional residue inserted between the signature glutamate and the second histidine, thus resulting in an exceptional HEXXXH signature $^{105}$. This is accounted for by a wider helix turn that allows the second histidine to be suitably positioned to bind the metal ion. Overall, the general hallmark of gluzincins would thus consist of a double consensus motif, HEXXH+(E,H)XX(A,F,T,S,G)(Fig. 1), as well as the minimal structural requirements of zincins plus the presence of a glutamate helix.
In addition to the families reported in Fig. 1, other structures show similarities with gluzincins but are probably unable to cleave peptides. Three similar proteins from *Haemophilus influenzae* 106, *Termotoga maritima* 107, and *Aquifex aeolicus* 108 possess a gluzincin-like fold but have the catalytic glutamate replaced by a glycine. The structures reported have a third potential histidine metal-ligand but lack a metal at the potential metal-binding site and no function or catalytic competence has been reported for any of them. By contrast, an endonuclease function in 70S ribosome quality control and 16S rRNA maturation was reported for *Escherichia coli* protein YbeY 109. This protein shows a structure half-way between a gluzincin and a metzincin despite lacking the general base/acid glutamate—likewise replaced here by a glycine—and may thus be unable to cleave proteins 110. Its bears the extended signature of metzincins (see section 4; the glutamate is replaced by glycine though) and a methionine below the metal-binding site as metzincins do (see section 4), but it also has a glutamate helix as in gluzincins. Actually, YbeY was already identified as a close structural homolog of minigluzincins 88, placing the latter family also at the interface between gluzincins and metzincins, with a likewise bulky hydrophobic residue in the Ser/Gly-turn, a phenylalanine, and a histidine as third metal ligand instead of the usual glutamate.

Finally, inverzincins 60; 61; 11; 62, whose name was also proposed by Hooper in 1994 41, possess the sequence HXXEH, which is the inverted zincin motif. This led them to be considered a distinct MP tribe (Fig. 1). However, like zincins, they have the motif imbedded within an active-site helix, which runs here characteristically in the opposite direction to the equivalent helix in zcinics. In addition, they possess a glutamate helix below the active-site helix, in the same relative position and direction as gluzincins. They have an EXXV motif encompassing the third glutamate metal-binding residue and a valine as the Ser/Ala-turn residue, as well as a mixed β-sheet of at least three strands equivalent to zincins. Furthermore, there is likewise a backing helix in the NTS. Accordingly, although inverzincins are considered a separate tribe from the zcinics (Fig. 1), they could also be classified as a special case of gluzincins like the relatives of dipeptidyl peptidase III (see above).

4. COMMON STRUCTURAL FEATURES OF METZINCINS

Metzincins 30-33; 28; 34-37 are present in all the kingdoms of life and more than 300 structures comprising at least the catalytic domain have been deposited with the Protein Data Bank (PDB; Table 1). In the present review, twelve catalytic domain lead structures, one for each of the families depicted in Fig.1, will be discussed (Figs. 3 and 4): *Astacus astacus* crayfish astacin (hereafter AST), *Pseudomonas aeruginosa* alkaline proteinase (alias aeruginolysin; SER), *Leishmania major* leishmanolysin (LEI), *Streptomyces caespitosus* neutral protease (alias ScNP or napapalysin; SNA), human neutrophil collagenase (alias matrix metalloproteinase 8; MMP), *Crotalus adamanteus* snake venom adamalysin II (ADA), the pappalysin ulilysin from *Methanosarcina acetivorans* (PAP), the archaemetzincin AmzA from *Methanopyrus kandleri* (AMZ), the unpublished igalysin “putative MP BAC0VA_0063” from *Bacteroides ovatus* (IGA), the toxilysin EcxA from *Escherichia coli* (TOX), fragilysin-3 from
**Bacteroides fragilis** (FRA), and the cholerilysin StcE from *Escherichia coli* (CHO)\(^{75; 111; 76; 112-121}\). These prototypes represent archaea, bacteria, protozoa/protists, and eukaryotes. The structures are \(~130-270\)-residue globular moieties that share a common scaffold and active-site environment, with the common elements described for zinclins in general (see section 3), but each family has distinguishing structural elements (see section 5).

The consensus NTS of metzincins includes a five-stranded twisted \(\beta\)-sheet on the top (Figs. 3-5). All strands (\(\betaI\) to \(\betaV\)) except the fourth are parallel to each other and to any substrate that is bound in the cleft (Fig. 5a). The antiparallel strand, \(\betaIV\), forms the lower edge of this sub-domain and creates an upper rim or northern wall of the active-site crevice \(^{122}\). As discussed for zinclins under section 3, this strand binds a substrate in an antiparallel manner, mainly on its non-primed side. The loop segment connecting strands \(\betaIII\) and \(\betaIV\) (L\(\betaIII\betaIV\)), the “bulge-edge segment,” generates protruding structural elements, which mainly affect sub-sites \(S_1\)' and \(S_2\)'.

This gives rise to extensive variations in enzyme-substrate interactions on the primed side of the active-site clefts. The NTS also contains a backing helix (\(\alphaA\)) and an active-site helix (\(\alphaB\)), which are arranged on the concave side of the \(\beta\)-sheet identically in all metzincin structures (Figs. 3-5). Helix \(\alphaB\) superimposes well in all structures (Fig. 5b,c) and encompasses the first half of the zinc-binding motif, which in metzincins is expanded from HEXX to HEXXHXXGXX(H,D) and includes the first two zinc-binding histidine residues. At the end of \(\alphaB\), the polypeptide chain takes a sharp, downward turn, mediated by the glycine of the consensus sequence. The main-chain angles of this residue in the lead structures indicate that any other residue would be in a high-energy conformation and, thus, disfavored. Pappalysins provide the only notable exception, with an asparagine at this position (see section 5.7). The CSD starts right after this glycine and the chain leads to the third zinc ligand, a histidine or an aspartate, which is the last residue of the metzincin zinc-binding motif and approaches the metal from below (Fig. 5b,c). The residue immediately downstream of the latter is termed the “family-specific residue” and is characteristic for each family. The CSD contains few regular secondary-structure elements, mainly a C-terminal helix (\(\alphaC\)) at the end of the polypeptide chain. Helices \(\alphaB\) and \(\alphaC\) are connected by structures that vary in both length and conformation. However, all structures coincide at a conserved 1,4-\(\beta\)-turn containing a strictly invariant methionine at position three, the Met-turn, which is separated from the third zinc-binding histidine by “connecting segments” ranging from six to 77 amino acids. The methionine is superimposable—including the conformation of its side chain, see Fig. 5b,c—and is positioned underneath the catalytic metal, forming a hydrophobic pillow. However, no contact with the metal is observed. Mutation studies indicated a role for this methionine in the folding and stability of the catalytic domains, although the strict conservation of this residue is still a matter of debate \(^{123-129}\). The \(S_1\)' pocket of metzincins is the main determinant of substrate specificity and is shaped at the top by the bulge-edge segment and at the bottom by a “wall-forming segment” made up of residues connecting the Met-turn and the C-terminal helix \(\alphaC\) \(^{122}\). This segment also varies in structure and length, ranging from nine to 37 residues in the reference structures.

The catalytic zinc ion lies roughly at half width of the active-site cleft, at its base. It is coordinated by the \(\text{N}_{\varepsilon2}\) atoms of the three consensus histidines or two histidine \(\text{N}_{\varepsilon2}\) atoms and an aspartate \(\text{O}_{\delta1/2}\) atom (in SNA
and IGA) and the catalytic solvent molecule in an approximately tetrahedral or trigonal pyramidal manner. The solvent molecule is replaced by other ligands in enzyme+inhibitor/product complexes. Some metzincins display a further protein ligand at a slightly greater distance from the catalytic cation in the form of a tyrosine Ω atom, as seen in unbound AST \textsuperscript{75} and SER \textsuperscript{111} and suggested for unbound PAP \textsuperscript{130}. This tyrosine is found two positions downstream of the Met-turn methionine. In SNA and IGA, a tyrosine is found two positions downstream of the metal-binding aspartate, though no longer within binding distance of the catalytic metal ion. The same is observed for the tyrosine found as the family-specific residue in CHO. All these tyrosines may flip back and forth during substrate anchoring, cleavage and product release, in a motion referred to as the “tyrosine switch” in the particular cases of astacins and serralysins \textsuperscript{131-133}. In this, they may play a role in substrate and catalytic-solvent binding and stabilization of the tetrahedral intermediate or the newly created α-amino group of the downstream product \textsuperscript{75; 131-133; 28}. Other, non-conserved residues take over this role in tyrosine-lacking metzincins.

5. METZINCIN FAMILY-SPECIFIC FEATURES

5.1. ASTACINS

The first member of the astacin family and first metzincin to be structurally analyzed, in 1992, was the 200-residue digestive crayfish enzyme AST \textsuperscript{75; 134}. Astacins have disparate biological roles in digestion, development, and tissue remodeling and differentiation—by promoting cartilage and bone formation and collagen biosynthesis—exerted through protein degradation, growth-factor activation, extracellular matrix turnover, and extracellular coat degradation (hatching). Six orthologs are found in humans \textsuperscript{135}. The catalytic domain of mature AST is a packman-like ellipsoid and has two sub-domains roughly equal in size (Fig. 3). Whilst the gene of this peptidase encodes only a short pro-peptide, which shields the active-site cleft in the zymogen following an “aspartate-switch” mechanism \textsuperscript{136}, and a catalytic MP domain, other family members have additional C-terminal SHK, CUB, LC, TSP, EGF, MAM, TRAF, inserted, transmembrane and cytosolic domains. The astacin family has the longest—together with the families represented by IGA, PAP and CHO—and most sequence-conserved connecting segment among metzincins (see Table 1 in \textsuperscript{28}). The protein scaffold is cross-linked in AST by four conserved cysteine residues forming two disulfide-bridges (Fig. 3). The first two residues of the mature enzyme are buried in an internal cavity in the CSD and the N-terminal α-amino group binds the family-conserved residue, a glutamate in astacins. The unbound coordination of the catalytic zinc is trigonal-bipyramidal due to the presence of a fourth invariant, though more distal, tyrosine-switch zinc ligand and the bound (putatively) catalytic solvent molecule (see section 4 and Fig. 3). In addition to AST, the structures of mature human toloid-like protease 1, bone morphogenetic protein 1, meprin β, high choriolytic enzyme 1, and zebrafish hatching enzyme 1, as well as zymogen structures of AST and meprin β, have been reported, totaling 16 structures (see Table 1). Reviews and a monograph on the family have been published \textsuperscript{40; 137-142; 135}.
5.2. SERRALYSINS

Pathogenic γ-class proteobacterial genera such as \textit{Pseudomonas}, \textit{Erwinia}, \textit{Serratia}, \textit{Pectobacterium}, \textit{Photobacterium}, \textit{Morganella}, \textit{Preoteus}, \textit{Xenorhabdus}, \textit{Dickeya}, and \textit{Yersinia} produce ~50-kDa bacterial virulence factors, the serralysins, which are secreted as autoactivatable zymogens \cite{143,144}. Several of these bacteria infect humans and cause—mostly hospital-acquired—infected infections such as meningitis, endocarditis, pyelonephritis, plague, dermatitis, soft-tissue infections, septicemia, melioidosis, pneumonia and other respiratory and urinary tract infections. Serralysins are possibly part of the bacterial armamentarium required for such infections and they have been shown to be active against coagulation factors and defense-oriented proteins, protease inhibitors, lysozyme and transferrin. They may also trigger an anaphylactic response. The first serralysin to be biochemically and structurally characterized was \textit{SER} \cite{145,146,111}. Its mature catalytic domain spans 220-residues (Figs. 3 and 4), lacks disulfide bonds and is followed by a C-terminal β-roll domain stabilized by calcium ions. As in AST, the two SER sub-domains are similar in size, and the enzyme starts with a family-specific α-helix in the CSD that is connected with the molecular moiety by a conserved salt bridge with the C-terminal helix αC (Figs. 3 and 4). An elongated LβαA gives rise to a flap that runs across the convex surface of the β-sheet and differs among serralysins, thus distinctly conditioning substrate binding. The CSD of SER bears an extra α-helix within the wall-forming segment and a second flap shaped by residues of the connecting segment. These elements likewise modulate substrate binding. Comparison of the structures of tetrapeptide-bound SER and that of the unbound ortholog of \textit{Serratia marcescens} reveals that the unliganded zinc coordination is similar to astacin (bipyramidal-trigonal) under inclusion of a tyrosine-switch tyrosine \cite{147}. Further serralysin structures have been reported from \textit{Erwinia chrysanthemi} and \textit{Flavobacterium} sp., totaling 26 (see Table 1). Two structures correspond to complexes with cognate protein inhibitors \cite{148,149}.

5.3. LEISHMANOLYSINS

Leishmanolysins were originally identified as protozoan cell-surface proteins GP63 in trypanosomes, ciliates and sarcocysts such as those found in the genera \textit{Trypanosoma}, \textit{Leishmania}, \textit{Crithidia}, \textit{Naegleria}, \textit{Tetrahymena} and \textit{Sarcocystis}. Collectively, these organisms are associated with major human infections such as African trypanosomiasis, meningoencephalitis, Chagas’ disease, and leishmaniasis. Leishmanolysins are the major component of the promastigote surface and they cleave host surface substrates such as CD4 molecules on human T cells. They further protect protozoan promastigotes from inactivation by complement proteins, thus playing a role as virulence factors. Similar sequences to the protozoan forms have been found in mammals (here called invadolysins), insects, plants, fish, parasitic worms, sponges, mossmes, and some bacteria \cite{28,150}. The only structurally analyzed family member is \textit{LEI} \cite{114}(Table 1). It is synthesized as a 602-residue inactive precursor in the endoplasmic reticulum provided with a signal peptide and a 100-residue pro-domain, which includes a highly conserved cysteine residue potentially acting as a “cysteine switch” (see section 5.5). Maturation produces an MP of 276 residues, which includes a 63-residue insertion domain between the glycine and the
third metal-binding histidine of the metzincin consensus motif (Figs. 3 and 4), followed by a ~200-residue C-terminal domain. The MP domain is the most asymmetric among metzincins, with a 171-residue NSD and a 42-amino acid CSD, and its N-terminus is found on the back left surface. The NSD has a β-sheet that lacks strand βII (Figs. 3 and 4) and two unique ~40-residue inserted flaps preceding and following backing helix αA, which are linked to the molecular moiety through respective disulfide bonds. A slightly prominent bulge-edge segment preceding strand βIV lies on top of the shallow, medium-sized S1' pocket, which is framed by the wall-forming segment and active-site helix αB. At the end of the CSD, αC is followed by a segment in extended conformation, which runs parallel across the back surface (Fig. 3; 114).

5.4. SNAPALYSINS

Streptomyces are predominantly soil bacteria, which collectively produce over two-thirds of the clinically-relevant antibiotics in nature. Only few species are pathogenic to humans and may cause mycetoma. SNA is a secreted neutral protease from Streptomyces caespitosus, which slowly hydrolyses milk. Similar sequences to SNA have been reported from other Streptomyces species and closely-related actinomycetales only. The MP consists of a 132-residue catalytic domain preceded by an alanine-rich 100-amino acid N-terminal extension encompassing a signal-peptide and a pro-domain. SNA thus has the shortest metzincin catalytic domain reported to date and the only three reported PDB entries of the family correspond to S. caespitosus, one of them to a heterotetrameric 2+2 complex with the dimeric cognate bifunctional protein inhibitor, sermetstatin 151. The SNA structure is a flattened ellipsoid and is divided into two asymmetric sub-domains 113; 152. It displays all the characteristic metzincin features, connected by short loops, and could therefore represent a minimal metzincin. Particular distinguishing elements are a small LβIIβIII protruding from the upper sheet within the NSD, a small bulge on top of the primed side of the active-site crevice, a short helix in LβVαB, and a calcium-binding site (Figs. 3 and 4). In addition, an aspartate is the third metal-binding protein ligand and, two positions ahead in the sequence, a conserved tyrosine approaches the metal but does not bind it.

5.5. MATRIXINS

Matrixins alias MMPs are secreted or membrane-bound MPs discovered 50 years ago as responsible for tadpole tail resorption during frog metamorphosis 153. They are found in higher mammals and in animals, plants, and viruses but not in fungi and only in some bacteria, where they do not follow the Darwinian tree-based pathway but may be the result of eukaryotic-to-prokaryotic horizontal gene transfer 154; 155. MMPs exert both broad-spectrum turnover and limited proteolysis of extracellular proteins, which includes ectodomain shedding at the plasma membrane in a complementary fashion to adamalysins/ADAMs (see 5.6). Substrates include extracellular matrix proteins, peptidases and zymogens, inhibitors, clotting factors, antimicrobial peptides, chemotactic and adhesion molecules, growth factors, hormones, and cytokines, as well as their respective receptors 156. In these functions, they contribute to physiological actions ranging from tissue processing during
embryogenesis and development, morphogenesis, signaling, and angiogenesis to apoptosis and intestinal defense-protein activation. As a result of this broad spectrum of functions, they play also key roles in pathologies such as inflammation, ulcer, rheumatoid and osteoarthritis, periodontitis, heart failure, atherosclerosis, stroke and cardiovascular disease, fibrosis, emphysema, multiple sclerosis, bacterial meningitis, Alzheimer's disease, HIV-associated dementia, and cancer and metastasis 157.

Vertebrate MMPs are multimodal proteins consisting of a ~20-residue secretory signal peptide, an ~80-residue pro-peptide, a 160-to-170-residue zinc- and calcium-dependent catalytic domain, a linker region, and a fourfold-propeller hemopexin-like C-terminal domain 158. Further possible insertions include fibronectin type-II-related domains, collagen type-V-like and vitronectin-like insertion domains, cysteine-rich, proline-rich and interleukin-1 receptor-like domains, immunoglobulin-like domains, glycosyl phosphatidylinositol linkage sequences, membrane anchors and cytoplasmic tails. 23 MMP proteins encoded by 24 genes are found in humans termed MMP-1 to MMP-28 (some numbers are unassigned)159. Those MMPs encompassing a membrane anchor gave rise to the MT-MMP subfamily 157; 122. MMPs are produced as zymogens that are activated by proteolytic cleavage following a "cysteine switch" or "velcro" mechanism, in which a pro-segment shields access of substrates to the active-site cleft and a conserved cysteine binds the catalytic metal ion 160; 161. The mature catalytic domain of human neutrophil collagenase (MMP; 112) is taken here as a prototype (Figs. 3 and 4). Like its orthologs and paralogs, it has a shallow active-site cavity, which separates a larger NSD (~120 residues) from a smaller CSD (~40 amino acids), and a deep hydrophobic S₁’ pocket. No disulfide bonds are present in the structure. Characteristically, the N-terminal α-amino group of the mature MP is salt-bridged to the first of two conserved aspartates within helix aC, which also hydrogen-bonds the MMP family-specific residue, a serine or threonine. The NSD displays an S-shaped double loop connecting strands βIII and βIV, which embraces structural zinc and calcium cations. The downstream residues of this segment form a prominent bulge that protrudes into the active-site groove. In addition, LβIVβV and LβIIβIII contribute to a second calcium-binding site on top of the NSD β-sheet (Figs. 3 and 4). In the CSD, MMPs possess the shortest connecting segment within metzincins, with just six to eight residues. MMPs are the most thoroughly studied metzincin family in structural terms, with over 200 structures reported, mostly of MMP-3 and MMP-12 (Table 1). These include isolated catalytic domains, multimodal structures and even full-length structures, as well as complexes with their endogenous tissue inhibitors of metalloproteinases (TIMPS), zymogenic structures, and numerous complexes with small-molecule inhibitors. Numerous reviews and monographs have been published and only a few are cited here 162; 157; 163-166; 158; 167-170.

5.6. ADAMALYSINS/ADAMS

The family, termed either adamalysins, ADAMs or reprolysins, has also been the object of extensive studies 171-173; 7; 174-182; it can be divided into three subgroups, the snake venom MPs, the mammalian ADAMs (19 paralogs in humans), and the likewise mammalian ADAMTSs (19 paralogs in humans). A separate chapter could
be dedicated to ADAMDEC1 alias decysin, which has a unique domain structure\textsuperscript{183}, and the nine testases, which are ADAMs found only in rodents\textsuperscript{184}. The proteins from snake venom cause destruction of extracellular proteins and hemorrhage following bites, while ADAMs were originally discovered in mammalian reproductive tracts engaged in fertilisation and sperm function. They are involved in myogenesis, development and neurogenesis, differentiation of osteoblastic cells, cell-migration modulation, and muscle fusion. Due to these critical roles in physiology, they are also related to human disorders such as asthma, cardiac hypertrophy, obesity-associated adipogenesis and cachexia, rheumatoid arthritis, endotoxic shock, inflammation, and Alzheimer's disease. Together with MMPs, (section 5.5), they are major contributors to protein ectodomain shedding. Finally, the ADAMTSs disable cell adhesion by binding to integrins. They also participate in gonad formation, embryonic development and angiogenesis, and procollagen activation, as well as in inflammatory processes, cartilage (aggrecan) degradation in arthritic diseases, bleeding disorders, and glioma-tumour invasion\textsuperscript{185}.

The family members are extracellular multidomain proteins containing a zinc- and calcium-dependent MP domain. In addition, they can display a pro-domain, C-terminal disintegrin-like, cysteine-rich, C-type lectin, EGF-like, thrombospondin 1-like, and/or transmembrane domains, as well as a cytoplasmic domain. In particular, all mammalian ADAMs are type-I membrane-anchored bitopic proteins, while ADAMTSs lack the transmembrane domain and contain multiple copies of a thrombospondin 1-like repeat and a CUB domain instead, thus being soluble extracellular proteases. Zymogenicity is exerted by pro-domains and activation occurs probably following a cysteine switch-like mechanism as in MMPs (section 5.5)\textsuperscript{160,166,28}. The first structure of a catalytic domain to be solved was that of ADA\textsuperscript{76,187}. This is a compact ellipsoidal 203-residue molecule with a relatively flat active-site cleft, which separates a 148-residue NSD from a 54-residue CSD (Figs. 3 and 4). Both the N- and the C-termini are surface located; the former is linked by a hydrogen bond to the C-terminal helix αC, which also binds the family-specific residue, a glutamate (Fig. 4). ADA deviates from the metzincin consensus due to two additional helices—the one inserted between βII and βIII and running front-to-back is termed “adamylin helix”—within the NSD. In the CSD, two disulfide-bonds cross-link the irregular connecting segment and attach αC to the NSD, respectively. A calcium ion is located on the surface, opposite the active site and close to the C-terminus (Figs. 3 and 4). The Si' pocket, characterized by a pronounced bulge-edge segment, is hydrophobic and deep, reminiscent of MMPs\textsuperscript{188,189}. In addition to ADA, a number of snake venom MPs, ADAM-8, ADAM-17, ADAM-33, ADAMTS-1, -4, and -5 have been structurally analyzed to date (Table 1). These structures include one complex of ADAM-17 with the only endogenous ADAM protein inhibitor reported, TIMP-3\textsuperscript{190}.

5.7. PAPPALYSINS

This family was named after human PAPP-A alias pappalysin-1, a heavily-glycosylated 170-kDa multidomain protein specifically cleaving insulin-like growth factor binding proteins\textsuperscript{124}, and its paralog pappalysin-2. Closely related sequences are present in mammals, birds, fishes, amphibiaians, and echinoderms,
and somewhat more distant sequences belong to fungi, bacteria, and archaea. Among the latter is 38-KDa PAP from *Methanosarcina acetivorans*, which encompasses only the pro-domain and the catalytic domain. Activation of the PAP zymogen may occur through cysteine-switch as in MMPs (section 5.5) as suggested by the presence of a conserved cysteine in the pro-domain. *In vitro*, activation occurs autolytically in the presence of calcium. With 262 residues, mature PAP is the second largest metzincin-prototype catalytic domain and the only family member to be structurally analyzed to date (Table 1). As characteristic structural elements, it has a loop dividing strand βII into two sub-strands (βII and βII'; Figs. 3 and 4), and a β-ribbon inserted within LβIIIβIV that protrudes from the molecular surface and frames the active site on its primed side. The segment connecting αA with βII covers the back of the molecule from the NSD to the CSD in a cape-like fashion and includes two unique α-helices. The glycine of the metzincin zinc-binding motif is exceptionally replaced in PAP and a subset of pappalysins by an asparagine, which binds an arginine within helix αC (Fig. 4). This replacement occurs under slight variation of the main-chain angles, which do not correspond here to a high-energy conformation (see section 4) but rather to a left-handed α-helix, although overall, the chain trace is indistinguishable from other metzincins (Fig. 5b,c). The CSD evinces two disulfide bonds and a unique two-calcium site. This site is a molecular switch for activity, as the proteinase can be reversibly inhibited through calcium chelators. Finally, in the absence of an unbound structure, PAP may possess a fifth zinc-binding tyrosine ligand provided by the Met-turn that is swung out upon substrate binding.

### 5.8. ARCHAEMETZINCINS

The family name was coined in 2003 and the first members to be studied at the biochemical level were the two human orthologs, archaemetzincin-1 and 2, which were reported to hydrolyze synthetic substrates and bioactive peptides. The structures of three archaeal orthologs from *Methanopyrus kandleri*, *Archaeoglobus fulgidus*, and *Methanocorpusculum labreanum* have been reported; the latter was only deposited by a structural genomics consortium; see Table 1), which however corresponded to inactive species. The family has an extended consensus sequence immediately after the third metal-binding histidine, CX₄CXM₁ₓCXXC. This sequence encompasses the Met-turn methionine and four cysteines, which bind a second, structural zinc ion in the CTS below the Met-turn (Figs. 3 and 4) in an element termed the "Cys₄ zinc finger". The cysteines are, respectively, in the position of the family-specific residue; two residues ahead of the Met-turn methionine; and in the first and fourth position of the C-terminal helix αC. Moreover, as found in the 175-residue *M. kandleri* protein, AMZ, additional family-characteristic structural elements are an adamasin helix running front-to-back, two short 3₁₀-helices and a small three-stranded antiparallel β-sheet on the convex side of the NTS β-sheet, and an extra helix within the wall-forming segment in the CTS (Figs. 3 and 4).

### 5.9. IGALYSINS
This family was not identified in previous sequence-based studies because it does not contain the intact metzincin zinc-binding consensus sequence but has uniquely an additional glycine preceding the signature glycine and an aspartate as third zinc ligand (HEXXHXXGXXD) as found in snapalysins (see 5.4). Only the deposition—but not publication—by a structural genomics consortium of two proteins from Bacteroides fragilis and Bacteroides ovatus (see Table 1), which are thus reported here for the first time, allowed their structural analysis and ascription to the metzincins. These proteins evince significant sequence similarity (BLAST E-values <10^{-57} at www.uniprot.org) with MPs from MEROPS family M64, termed relatives of IgA-peptidase of Clostridium ramosum, which cleave prolyl bonds within an O-glycosylated, tandemly duplicated octapeptide in the hinge region of human immunoglobulin IgA1 \(193; 194\). These peptidases should not be confused with those of family M26, which groups MPs represented by IgA1-specific metallopeptidase from Streptococcus sanguinis \(195; 196\). Accordingly, the term "galysins" will be used hereafter to refer to the novel metzincin family of M64 peptidases.

Detailed inspection of B. ovatus galysin (IGA) reveals that, with 272 residues, the structure is the largest reported metzincin catalytic domain, and all the characteristic structural elements for metzincins are observed (Figs. 3 and 4). It shows an asymmetric NTS-CTS tandem (166 and 106 residues, respectively) and, within NTS, on top of the β-sheet, it has an extra sixth β-strand inserted before βI, which runs parallel to the upper-rim strand and antiparallel to the other four. In addition, an adamalysin-helix is inserted between βI and βIII and an extra helix at LaAβII (Fig. 4). The insertion of the extra glycine within the consensus sequence occurs at the end of the active-site helix and is compensated for by the extension of the helix by an additional half turn so that, overall, the chain trace hardly differs from that of the other metzincins (Fig. 5c). As to the CTS, it bears the largest connecting segment (77 residues) but the shortest wall-forming segment (nine residues) within metzincins. As in snapalysin, a tyrosine is found two positions downstream of the third zinc-binding residue—likewise an aspartate—but not within binding distance to the latter. However, the most striking feature is the presence of a second structural metal-binding site within the CTS, which coincides in space with the one found in arcahemetzelincins (chapter 5.7) despite completely different chain traces and lengths of the respective connecting segments and wall-forming segments. In IGA, this Cys4 zinc finger is bound, as in arcahemetzelincin, by the family-specific residue, which is a glutamate instead of a cysteine; by the first and fourth cysteine residues of the C-terminal helix; and by a cysteine two residues downstream of the Met-turn methionine.

5.10. TOXILYSINS

This is the most recent family, which was discovered while studying bacterial AB5 toxins \(121\). These are virulence factors of pathogenic bacteria that consist of a pentameric B subunit for host-cell invasion and an A subunit that subverts intracellular functions \(197; 198\). In contrast to previously studied related toxins, EcxB from E. coli and CfxAB from Citrobacter freundii exhibit the metzincin consensus sequence within their A subunits, and the crystal structure of the former reveals that they fulfill the structural requirements for a novel metzincin
family, the toxilysins \(^{199; 211}\), the structure of EcxA, the only one reported to date for the family (Table 1), is a compact disk of 258 residues (Figs. 3 and 4). The N-terminal segment is extended by \(~30\) residues preceding NTS strand \(\beta I\) and starts within the CTS. There it contributes with two short strands to a four-stranded antiparallel \(\beta\)-sheet. The N-terminal extension is connected with the Met-turn through a double hydrogen bond formed by the family-specific asparagine with both structural elements. Before entering strand \(\beta I\) and already within the NTS, the polypeptide chain also includes an extra helix. In addition, the NTS bears a \(\beta\)-ribbon within \(L\beta I\) and two consecutive helices within \(L\beta V\alpha B\). Within the CTS, the connecting segment provides two additional strands to the aforementioned extra \(\beta\)-sheet and the wall-forming segment contains an extra helix as found in SER and AMZ (sections 5.2 and 5.7). Finally, the C-terminal segment is also extended beyond \(\alpha C\), adopts a helical conformation in two short stretches, and is linked through a disulfide bond to the protein moiety. The final helix interacts with the pentameric B subunit to yield the \(AB_5\) complex.

### 5.11. FRAGILYSINS

This is the smallest family, with just one member occurring as three closely related isoforms—fragilysin-1, -2, and -3—with 93-96% sequence identity in enteropathogenic \(Bacteroides fragilis\) (ETBF) only. It was first reported in 1984 as a factor potentially causing acute diarrheal disease in newborn lambs \(^{200}\). The chromosomal pathogenicity islet that encodes fragilysins and is absent in non-enterotoxigenic strains contains a second gene, \(mpII\), which is countertranscribed and encodes a potential MP of similar size and moderate sequence identity (28-30%) to the three fragilysins. It has recently been shown to be an active MP with a specificity for cleavage between flanking arginines \(^{201}\). Fragilysins are the only proven virulence factors of ETBF and target E-cadherin from the intestine of infected mammalian hosts. The structural and function analysis of FRA revealed a 194-residue pro-domain—the largest among metzincins—with unique fold, which plays a major role in protein folding and inhibits the mature 188-residue MP moiety through an aspartate-switch mechanism similarly to astacins (section 5.1) \(^{117; 202}\). As distinguishing structural features, FRA has an adalysin helix within the NTS and a characteristic salt bridge that links the N- and the C-terminus at the back of the molecule (Figs. 3 and 4). Overall, the FRA catalytic domain shows remarkable structural similarity with adalysins/ADAMs despite just \(~15\%) sequence similarity (see Suppl. Fig. S3b in \(^{117}\)). Taken together with its restriction to just one protein found in just one organism, this indicates that it may be the result of horizontal gene transfer of an ADAM gene from a mammalian host and subsequent evolution in a bacterial environment, thus giving rise to a unique pro-domain, small structural changes and a different protein sequence expressed as three isoforms \(^{117; 202}\).

### 5.12. CHOLERILYSINS

In pathogenic \(Vibrio cholerae\), the transmembrane DNA-binding protein ToxR coordinates the expression of essential virulence genes encoded by the \(Vibrio\) pathogenicity island, including those encoding virulence factors such as the toxin-coregulated type-IV pilus. Among them, ToxR-activated gene A (tagA)
lipoprotein has been described. Together with a paralog, another protein from *Burholderia*, and protein StcE (secreted protease of C1-esterase inhibitor; CHO), they were proposed in 2003 to constitute the cholerilysin family, which coincides with MEROPS family M66 and the SLiMe family. CHO is an 898-residue virulence factor secreted by enterohemorrhagic *E. coli*. It acts as a mucinase hydrolyzing glycoproteins of the mucosal barrier that lines and protects the mammalian gastrointestinal tract, thus contributing to bacterial penetration during infection. CHO includes a central 258-residue catalytic domain flanked upstream by two immunoglobulin-like domains and an insertion region and downstream by two disordered regions and a C-terminal domain. This is the only structure reported for the family (Table 1): it is shaped like a banana, while comprising all the structural hallmarks of metzincins. However, in comparison with other metzincins, CHO shows an unusually open active-site cleft, which is compatible with binding of large O-glycosylated substrates. In addition, it features a helix on the convex face of the NTS β-sheet, in a similar position though different orientation to the adamalysin helix of other metzincins. Furthermore, an active-site flap with a short β-ribbon is inserted after strand βII, which replaces the bulge-edge segment in other metzincins in shaping the active-site cleft top on its primed side. Between the family-specific tyrosine, which point towards the bulk solvent in the reported structure but could belong to a tyrosine switch (section 4), and the Met-turn, two additional β-ribbons are found within the CTS; the first contains a conserved sequence, WGWD, and the second is crosslinked by a disulfide bond. Finally, the N-terminal α-amino nitrogen is linked through a salt bridge with a glutamate in the penultimate position of the C-terminal helix αC, five residues upstream of the catalytic domain C-terminus.

**CONCLUDING REMARKS**

The metzincins are ubiquitous MPs that participate in pathophysiological metabolic processes. Twelve metzincin families have been characterized at the structural level, some are restricted to certain taxa, others span several kingdoms, and even others are most likely the product of horizontal gene transfer. In addition to those structurally characterized, several more families may enlarge the clan based on the presence of the characteristic extended zinc-binding consensus sequence pattern and its small variants as found in some current family members. The structural evidence, accumulated over the past 21 years, has enabled us to confirm the hypothesis proposed in 1993 that they share a set of characteristic structural elements and a catalytic metal environment despite negligible overall sequence similarity. Accordingly, this MP clan can be envisaged as an example of divergent evolution since the onset of life—they are present in archaea—from an archetypal minimal urmetzincin, which may have spanned ~100-110 residues and may not have looked very different from the smallest current member, 132-residue snapalysin. Within this initial scaffold, functional requirements and the associated evolutionary processes would have led to the introduction of specific structural elements, which eventually resulted in the distinct current families. The structural information reviewed here may help to ascribe the function of proteins encoded by newly discovered gene sequences and in the elucidation of common
catalytic mechanisms but also of their detailed differences, which may enable the design of specific and selective inhibitors and activators to modulate their activity.

ACKNOWLEDGMENTS

Ulrich Baumann and Johann (Hans) Brandstetter are warmly thanked for critical reading of the manuscript, as is Robin Rycroft. This study was supported in part by grants from European, Spanish, and Catalan agencies (FP7-HEALTH-2010-261460 "Gums&Joints"; FP7-PEOPLE-2011-ITN-290246 "RAPID"; FP7-HEALTH-2012-306029-2 "TRIGGER"; BFU2012-32862; CSD2006-00015; Fundació “La Marató de TV3” grant 2009-100732; and Generalitat de Catalunya grant 2009SGR1036).
**FIGURE LEGENDS**

**Figure 1 — Classification of mononuclear MPs.** Within the MP class, mononuclear MPs are a subclass that is divided into tribes currently characterized at the structural level: inverzincins, zincins, relatives of hydrogenase maturing factor (HybD), LAS MPs, and \( \alpha \beta \alpha \)-exopeptidases. These tribes subdivide into clans, which in turn give rise to families. The metzincins are depicted and framed in blue. Metal-binding residues are shown in green, general base/acid residues in magenta, and residues/molecules occupying the position of the Met-turn or Ser/Gly-turn beneath the metal site are in orange. Other residues engaged in substrate binding, stabilization of the reaction intermediate, and/or catalysis are further shown in black, except for X, which stands for any residue and is here only used as a spacer within motifs. Within gluzincins, \( Z=A/F/S/G/T \). Variability within motifs was considered only if present in more than one case. * Leishmanolysins have an extra domain inserted between the glycine and the third histidine metal ligand of the metzincin extended zinc-binding signature. ** Igalsins have an extra glycine residue inserted just before the glycine of the metzincin signature. *** Dipeptidyl peptidase III has an extra residue inserted between the general base/acid glutamate and the second metal-binding histidine. This classification supersedes previous schemes \(^{28;43;52}\).  

**Figure 2 — Generally accepted catalytic mechanism of monometallic MPs.** The catalytic solvent molecule is bound first to the catalytic metal ion (white sphere) and the general base/acid in the active site in the absence of a peptidic substrate (I). Once the substrate is accommodated in the cleft and the Michaelis complex is formed (II), the polarized solvent molecule attacks the scissile carbonyl group, which leads to the tetrahedral reaction intermediate (III). The latter resolves in scissile bond breakage and double proton transfer to the newly formed \( \alpha \)-amino group to render a double-product complex (IV).

**Figure 3 — Structure of metzincin catalytic domains.** Ribbon-type Richardson plots of a representative member of each of the twelve structurally-characterized metzincin families (see sections 4 and 5, and Fig. 1) in standard orientation \(^{52}\). Depicted are astacin (AST; PDB 1AST; 200 residues), aeruginolysin (SER; PDB 1KAP; 220 residues), leishmanolysin (LEI; PDB 1LML; 213 residues), snapalysin (SNA; PDB 1C7K; 132 residues), human neutrophil collagenase (MMP; PDB 1JAN; 164 residues), adamalysin II (ADA; PDB 1IAG; 203 residues), ulilysin (PAP; PDB 2CKI; 262 residues), Methanopyrus kandleri archaemetzincin AmzA (AMZ; PDB 2X7M; 173 residues), igalysin BACOVA_0063 (IGA; PDB 3P1V; 272 residues), toxilysin EcxA (TOX; PDB 4L63; 258 residues), fragilysin-3 (FRA; PDB 3P24; 188 residues), and cholerilysin StcE (CHO; PDB 3UJZ; 258 residues). The common \( \beta \)-strands and \( \alpha \)-helices are shown in purple and cyan, respectively (see Fig. 4 for their nomenclature). Unique regular secondary structure elements of each family are in white. The N- and C-termini are labeled and the side chains of the zinc-binding histidines/aspartates (brown), general base/acid glutamates (pink), Met-turn methionines (blue), disulfide-linked cysteines (yellow), family-specific residues (red), and zinc-binding or substrate-stabilizing tyrosines (where present; in brown) are displayed as sticks. Preceding, inserted, and following (sub-)domains have been omitted for clarity. Catalytic metal ions—mostly zinc—are shown as
magenta spheres, calcium cations as red spheres, and the potassium cation of IGA is in blue. Orange arrows pinpoint the anchor points for a domain inserted in LEI after the glycine of the extended consensus sequence and the third metal-binding histidine, PAP is unique in having an asparagine instead of the glycine. Black arrows pinpoint two disordered loops in TOX.

**Figure 4 – Topology of metzincin catalytic domains.** Scheme showing the regular secondary structure elements (helices as rods, strands as arrows) of each metzincin prototype depicted for its structure in Fig. 3. The common elements are in pink (strands) and turquoise (helices) and labeled (βI-βV and αA-αC). Hydrogen bonds and metal-ligand bonds are shown as dashed lines, disulfide bonds as orange solid lines. Disordered segments or points of insertion of extra domains are characterized by blue dashes. Family-specific residues are in red, B stands for bulky, hydrophobic residues, X for any residue.

**Figure 5 – Superposition of common structural elements. (A)** Cross-eye stereo image depicting the superposition of the 12 reference structures only around the common structural elements, β-sheet strands βI (turquoise), βII (brown), βIII (orange), βIV (yellow), and βV (blue), as well as backing helices αA (pink), active-site helices αB (white), and C-terminal helices αC (red). **(B)** Superposition of the active-site helix and the downstream chain segment until the family-specific residue plus the Met-turn of AST (white), ADA (gold), SER (pink), LEI (purple), MMP (turquoise), and SNA (green) in two orthogonal views. The catalytic metal ions have been omitted for clarity. **(C)** Same as (B) but showing AMZ (blue), FRA (red), IGA (orange), CHO (tan), PAP (salmon), and TOX (sienna).
**Table 1** – Protein Data Bank entries comprising at least a metzincin catalytic domain ([www.pdb.org](http://www.pdb.org); search completed on October 14, 2013).

<table>
<thead>
<tr>
<th>Peptidase name</th>
<th>Organism</th>
<th>PDB access code</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASTACINS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astacin</td>
<td><em>Astacus astacus</em></td>
<td>1ast 1iaa 1iab 1iac 1iad 1iae 1qji 1qjj 3lq0</td>
</tr>
<tr>
<td>Tollid-like protease 1</td>
<td><em>Homo sapiens</em></td>
<td>3edi</td>
</tr>
<tr>
<td>Bone morphogenetic protein 1</td>
<td><em>H. sapiens</em></td>
<td>3edh 3edg</td>
</tr>
<tr>
<td>Meprin β</td>
<td><em>H. sapiens</em></td>
<td>4gwm 4gwn</td>
</tr>
<tr>
<td>HCE-1(high choriolyt. enzyme-1)</td>
<td><em>Oryzias latipes</em></td>
<td>3vtg</td>
</tr>
<tr>
<td>ZHE1</td>
<td><em>Danio rerio</em></td>
<td>3lqb</td>
</tr>
<tr>
<td><strong>SERRALYNS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeruginolysin</td>
<td><em>Pseudomonas aeruginosa</em></td>
<td>1kap 1jiw 1akd 3vi1</td>
</tr>
<tr>
<td>Psychrophilic alkaline protease</td>
<td><em>Pseudomonas TACII 18</em></td>
<td>1g9k 1h71 1omj 1om6 1om7 1om8 1o0t 100q</td>
</tr>
<tr>
<td>Serralysin</td>
<td><em>Serratia marcescens</em> / <em>Serratia</em> sp. E-15</td>
<td>1af0 1sat 1srp 1smp</td>
</tr>
<tr>
<td>PrtC</td>
<td><em>Erwinia chrysanthemi</em></td>
<td>1go7 1go8 1k7g 1k7i 1k7q 3hb2 3hbu 3hbv 3hda</td>
</tr>
<tr>
<td>Psychrophilic marine protease</td>
<td><em>Flavobacterium</em> sp. YS-80-122</td>
<td>3u1r</td>
</tr>
<tr>
<td><strong>ADAMALYNS/ADAMS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adamalyn II</td>
<td><em>Crotalus adamanteus</em></td>
<td>1iag 2aig 3aig 4aig</td>
</tr>
<tr>
<td>Atrolysin-C (Ht-d/c)</td>
<td><em>Crotalus atrox</em></td>
<td>1at1 1dth 1htd</td>
</tr>
<tr>
<td>Bothropasin</td>
<td><em>Bothrops jararaca</em></td>
<td>3dsl</td>
</tr>
<tr>
<td>Taiwanese habu MP</td>
<td><em>Trimeresurus mucrosquamatus</em></td>
<td>1kuf 1kug 1kui 1kuk</td>
</tr>
<tr>
<td>H₂-proteinase</td>
<td><em>Trimeresurus flavoviridis</em></td>
<td>1wni</td>
</tr>
<tr>
<td>Acutolysin-C</td>
<td><em>Agkistrodon acutus</em></td>
<td>1qua</td>
</tr>
<tr>
<td>Acutolysin-A</td>
<td><em>A. acutus</em></td>
<td>1bsw 1bud</td>
</tr>
<tr>
<td>Protein Name</td>
<td>Species</td>
<td>Matrix</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>BaP1</td>
<td><em>Bothrops asper</em></td>
<td>1nd1 2w12 2w13 2w14 2w15</td>
</tr>
<tr>
<td>Vascular apoptosis-inducing protein1 VAP1</td>
<td><em>C. atrox</em></td>
<td>2ero 2erp 2erq</td>
</tr>
<tr>
<td>Fibrin(ogen)olytic MP</td>
<td><em>A. acutus</em></td>
<td>1yp1</td>
</tr>
<tr>
<td>Catrocollastatin/VAP2B</td>
<td><em>C. atrox</em></td>
<td>2dw0 2dw1 2dw2</td>
</tr>
<tr>
<td>Russell's viper venom MP</td>
<td><em>Daboia russelli siamensis</em></td>
<td>2e3x</td>
</tr>
<tr>
<td>BmooMPα-1</td>
<td><em>Bothrops moojeni</em></td>
<td>3gbo</td>
</tr>
<tr>
<td>AaHIV</td>
<td><em>A. acutus</em></td>
<td>3hdb</td>
</tr>
<tr>
<td>Atragin</td>
<td><em>Naja atra</em></td>
<td>3k7l</td>
</tr>
<tr>
<td>K-like MP</td>
<td><em>N. atra</em></td>
<td>3k7n</td>
</tr>
<tr>
<td>TM-1 MP</td>
<td><em>T. mucrosquamatus</em></td>
<td>4j4m</td>
</tr>
<tr>
<td>TNF-α converting enzyme (ADAM-17)</td>
<td><em>H. sapiens</em></td>
<td>1bkc 3edz 3b92 3cki 2oi0 2i47 2fv9 2fv5 2ddf 2a8h 1zxc 3e8r 3ewj 3g42 3kmc 3kme 3l0t 3l0v 3le9 3lea 3lgp 3o64</td>
</tr>
<tr>
<td>ADAM-8</td>
<td><em>H. sapiens</em></td>
<td>4dd8</td>
</tr>
<tr>
<td>ADAM-33</td>
<td><em>H. sapiens</em></td>
<td>1r54 1r55</td>
</tr>
<tr>
<td>ADAMTS-1</td>
<td><em>H. sapiens</em></td>
<td>2jih 2v4b 3q2g 3q2h</td>
</tr>
<tr>
<td>ADAMTS-4</td>
<td><em>H. sapiens</em></td>
<td>2rjp 3b2z</td>
</tr>
<tr>
<td>ADAMTS-5 (alias ADAMTS-11)</td>
<td><em>H. sapiens</em></td>
<td>3b8z 2rjq 3hy7 3hy9 3hyg 3ljt</td>
</tr>
</tbody>
</table>

### MATRIXINS/MMPs

<table>
<thead>
<tr>
<th>MMP-1 Name</th>
<th>Species</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibroblast collagenase (MMP-1)</td>
<td><em>H. sapiens</em></td>
<td>966c 1ayk 1cge 1cgl 1hfc 2ayk 2tcl 3ayk 4ayk 1su3 2j0t 2clt 3shi 4auo</td>
</tr>
<tr>
<td>Gelatinase A (MMP-2)</td>
<td><em>H. sapiens</em></td>
<td>1ck7 1hov 1qib 1gxd 1eak 3ayu</td>
</tr>
<tr>
<td>Stromelysin-1 (MMP-3)</td>
<td><em>H. sapiens</em></td>
<td>1b3d 1b8y 1biw 1bm6 1bqo 1c3i 1caq 1ciz 1cqr 3usn 1uea 2d10 1c8t 2jt6 2jt5 2jnq 1d5j 1d7x 1d8f 1d8m 1g05 1g49 1g4k 1hfs 1hy7 1sh 1slm 1qia 1qic 1ums 1unm 1usn 2srt 2usn 10o9 3ohl 3oho 4dpe 4g9l 4ja1</td>
</tr>
<tr>
<td>Matrilysin (MMP-7)</td>
<td><em>H. sapiens</em></td>
<td>1mmp 1mmq 1mmr 2ddy 2y6c 2y6d</td>
</tr>
<tr>
<td>Neutrophil collagenase (MMP-8)</td>
<td><em>H. sapiens</em></td>
<td>1jan 1a85 1a86 1bzs 1i73 1i76 1jao 1jap 1jaq 1jh1 1jj9 1kbc 1mmb 1mnc 1zvx 2oy4</td>
</tr>
<tr>
<td>Protein Type</td>
<td>Species</td>
<td>Accession Numbers</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Gelatinase B (MMP-9)</td>
<td><em>H. sapiens</em></td>
<td>1gkc 1gkd 1l6j 2ow0 2ow1 2ow2 2ovz 2ovx 4h1q 4h2e 4h3x 4h82 4hma</td>
</tr>
<tr>
<td>Stromelysin-2 (MMP-10)</td>
<td><em>H. sapiens</em></td>
<td>1q3a 3v96</td>
</tr>
<tr>
<td>Stromelysin-3 (MMP-11)</td>
<td><em>Mus musculus</em></td>
<td>1hv5</td>
</tr>
<tr>
<td>Macrophage elastase (MMP-12)</td>
<td><em>H. sapiens</em></td>
<td>1jiz 1jk3 2z2d 2k2g 3ba0 2oxu 2oxz 2oxw 2hu6 1z3j 1ycm 1y93 1rmz 1os9 1os2 2p0j 1utz 1utt 1ros 2k9c 2krj 2w0d 2wo8 2wo9 2woa 3ehx 3ehy 3f15 3f16 3f17 3f18 3f19 3f1a 3lik 3lil 3lir 3ljg 3lk8 3lka 3n2u 3n2v 3nx7 3rts 3rtt 3ts4 3tsk 3uvc 4efs 4gql 4gr0 4gr3 4gr8 4guy 4h30 4h49 4h76 4h84 4i03 4ijo</td>
</tr>
<tr>
<td>Collagenase-3 (MMP-13)</td>
<td><em>H. sapiens</em></td>
<td>456c 830c 1eub 1fsl 1fm1 2pj7 2ozr 2ow9 2e2d 2d1n 1ztq 1you 1xur 1xuc 2yig 3elm 3i7g 3i7i 3kec 3kej 3kek 3kry 3l2x 3txc 3zxc 4a7b 4fuf4 4fyl4g0d</td>
</tr>
<tr>
<td></td>
<td><em>M. musculus</em></td>
<td>1cxe</td>
</tr>
<tr>
<td>MT1-MMP (MMP-14)</td>
<td><em>H. sapiens</em></td>
<td>1bqq 1buv 3ma2</td>
</tr>
<tr>
<td>MT3-MMP (MMP-16)</td>
<td><em>H. sapiens</em></td>
<td>1rm8</td>
</tr>
<tr>
<td>Enamelysin (MMP-20)</td>
<td><em>H. sapiens</em></td>
<td>2jsd</td>
</tr>
<tr>
<td>Karilysin</td>
<td><em>Tannerella forsythia</em></td>
<td>2xs3 2xs4 4in9</td>
</tr>
</tbody>
</table>

**SNAPALYSINS**

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Species</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral protease</td>
<td><em>Streptomyces caespitosus</em></td>
<td>1c7k 1kuh 4hx3</td>
</tr>
</tbody>
</table>

**LEISHMANOLYSINS**

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Species</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leishmanolysin</td>
<td><em>Leishmania major</em></td>
<td>1ml</td>
</tr>
</tbody>
</table>

**PAPPALYSINS**

<table>
<thead>
<tr>
<th>Protein Type</th>
<th>Species</th>
<th>Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ullysin</td>
<td><em>Methanosarcina acetivorans</em></td>
<td>2cki 2j83 3lum 3lun</td>
</tr>
</tbody>
</table>

**ARCHAEMETZINCINS**
<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>Access Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmznA</td>
<td><em>Methanopyrus kandleri</em></td>
<td>2x7m</td>
</tr>
<tr>
<td></td>
<td><em>Archaeoglobus fulgidus</em></td>
<td>3zv 4a3w 4axq</td>
</tr>
<tr>
<td>Putative zinc-dependent peptidase</td>
<td><em>Methanocorpusculum labreanum</em></td>
<td>3lmc</td>
</tr>
<tr>
<td>FRAGILYSINS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fragilysin-3 (profragilysin-3)</td>
<td><em>Bacteroides fragilis</em></td>
<td>3p24</td>
</tr>
<tr>
<td>CHOLERILYSINS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>StcE</td>
<td><em>Escherichia coli O157:H7</em></td>
<td>3ujz</td>
</tr>
<tr>
<td>TOXILYSINS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcxB toxin</td>
<td><em>E. coli</em></td>
<td>4l63 4l6t</td>
</tr>
<tr>
<td>IGALYSINS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative peptidase BF3526</td>
<td><em>B. fragilis</em></td>
<td>4df9</td>
</tr>
<tr>
<td>Putative metalloendopeptidase BACOVA_0063</td>
<td><em>B. ovatus</em></td>
<td>3p1v</td>
</tr>
</tbody>
</table>

The primary reference of each structure may be retrieved from the PDB with the respective access code except for coordinates deposited by structural genomics consortia but not published (denoted as “putative”).
REFERENCES


81. Fushimi N, Ee CE, Nakajima T, Ichishima E (1999) Aspzincin, a family of metalloendopeptidases with a new zinc-binding motif. Identification of new zinc-binding sites (His(128), His(132), and Asp(164)) and three catalytically crucial residues (Glu(129), Asp(143), and Tyr(106)) of deuterolysin from Aspergillus oryzae by site-directed mutagenesis. J Biol Chem 274:24195-24201.


141. Semenova SA, Rudenskaia GN (2008) [The astacin family of metalloproteinases]. Biomeditsinskaia khimiia 54:531-554. PMID: 19105396 {Medline}


Hierarchy level

Subclass

Mononuclear metallopeptidases

Tribe

Inverzincins (HXXEH)
Zincins (HEXXH)
HybD (E+?+D+H)
Las metalloproteases (HXX, D+[HXX]/[EXX])
αβ-Exopeptidases

Clan

Metzincins (HEXXHXXGGXXHD+M)
S2P-zincins (HEXXH+D+P)
FtsH-like AAA MPs (HEXXH+E+ND+H2O)
Gluzincins (HEXXH+E/HXXZ)
Asp zincins (HEXXH+DXXY+NAD)
Funnelins (HXXE+R+NR+H+Y+E)

Family

Astacins
Adamalysins/ADAMs
Archaemetzincins
Serralysins
Snapalysins
Leishmanolysins *
Matrixins
Pappalysins
Cholerilysins
Fragilysins
Igalysins **
Toxilysins

Bacterial collagenases (HEXXH+EEXXXAE+Y)
Thermolysins (HEXXH+NEXXS+S+Y+H)
Minigluzincins (HEXXH+HXXF+XXE)
Cowrins (HEXXH+EXXS/G+H+Y/R+Y)
M48/M56 IMMPs (HEXXH+EEXX+A+N+H)
Clostridial neurotoxins (HEXXH+EEXXT+RXXY)
Anthrax lethal factor (HEXXH+EEXX+A+Y)
LA4H+M1-APs (HEXXH+NEXXT/A+GXMEN+Y)
Dipeptidyl peptidase III *** (HEXXH+EE/R/KXXAE/D)
Neprilysins (HEXXH+Y+EEXXAD+R)
Fig. 2

I. Unbound enzyme

II. Michaelis complex

III. Tetrahedral intermediate

IV. Double-product complex