JBC Papers in Press. Published on February 9, 2017 as Manuscript M116.758110
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M116.758110

Myroilysin is a new bacterial member of the M12A family of metzincin metallopeptidases and activated by a cysteine-switch mechanism

Dongqing XU‡, Jiale ZHOU‡, Xiangdi LOU‡, Jianhua He¶, Tingting RAN‡*, Weiwu WANG‡*

From the ‡Key Laboratory of Agricultural and Environmental Microbiology, Ministry of Agriculture, College of Life Sciences, Nanjing Agricultural University, 210095 Nanjing, China, ¶ Shanghai Institute of Applied Physics, Chinese Academy of Sciences, 201204 Shanghai, China

*To whom correspondence should be addressed:
Dr. Tingting Ran, Key Laboratory of Agricultural and Environmental Microbiology, College of Life Sciences, Nanjing Agricultural University, 210095 Nanjing, China; Phone: +86-25-84395003; Fax: +86-25-84396542;
e-mail: rantt@njau.edu.cn
Prof. Dr. Weiwu Wang, Key Laboratory of Agricultural and Environmental Microbiology, Ministry of Agriculture, College of Life Sciences, Nanjing Agricultural University, 210095 Nanjing, China; Phone: +86-25-84395003; Fax: +86-25-84396542;
e-mail: wwwang@njau.edu.cn

Running title: Crystal structure of pro-myroilysin
Keywords: pro-myroilysin, crystal structure, inhibition mechanism, M12 Family protease, classification

ABSTRACT

Proteases play important roles in all living organisms and also have important industrial applications. Family M12A metalloproteases, mainly found throughout the animal kingdom, belong to the metzincin protease family and are synthesized as inactive precursors. So far, only flavastacin and myroilysin, isolated from bacteria, were reported to be M12A proteases, whilst the classification of myroilysin is still conflicting due to the lack of structural information. Here, we report the crystal structures of pro-myroilysin from bacterium *Myroides* sp. cslb8. The catalytic zinc ion of pro-myroilysin, at the bottom of a deep active site, is coordinated by three histidine residues in the conserved motif HEXXHXXGXXH; the cysteine residue in the pro-peptide coordinates the catalytic zinc ion and inhibits myroilysin activity. Structure comparisons revealed that myroilysin shares high similarity with the members of the M12A, M10A and M10B families of metalloproteases. However, a unique "cap" structure tops the active site cleft in the structure of pro-myroilysin, and this "cap" structure does not exist in the above structure-reported subfamilies. Further structure-based sequence analysis revealed that myroilysin appears to belong to the M12A family, but pro-myroilysin uses a "cysteine-switch" activation mechanism with a unique segment including the conserved cysteine residue while other reported M12A family proteases use "aspartate-switch" activation mechanism. Thus, our results suggest that myroilysin is a new bacterial member of the M12A family with an exceptional cysteine-switch activation mechanism. Our results shed new light on the classification of M12A family, and may suggest a divergent evolution of M12 family.

INTRODUCTION

Proteases play important roles during the life of all organisms, including microbes (1).

Intracellular proteases are of great importance for various metabolic processes such as protein turnover, sporulation and differentiation (2,3). Extracellular proteases usually hydrolyze proteins in the environment to provide nutrients to the bacteria (4-7), although some are thought to be involved in pathogenesis (8). As extracellular proteases are easily obtained, they are also widely applied in various industries and medical fields, including detergent, leather, food, waste treatment, diagnosis of illness and pharmaceuticals (9-13).

Metalloproteases, a group of important proteases, require a divalent metal cation (usually zinc ion) for activity. The metal ion in metalloprotease is coordinated by several amino acid residues, commonly His, Glu, Asp or Lys (14). Zinc-dependent metalloproteases contain a signature motif HEXXH (where X represents any amino acid residue) in which the two histidine residues function as the first and second zinc ligands and the glutamate residue plays an outstanding role in catalysis (15). When the third zinc ligand is a glutamate residue, zinc-dependent metalloproteases are called gluzincins, a group that consists of exopeptidases and endopeptidases (14). If the third zinc ligand is a histidine or aspartate residue in the elongated motif HEXXHXXGXXH/D, the conserved glycine residue is indispensable for forming a β-turn to bring the three zinc ligands together (16). These types of zinc-dependent metalloproteases are known as metzincins because of a conserved methionine residue in a Met-turn (such as SIMHY) that underlies the active site (17) to form a structure that acts as a "hydrophobic pillow". All metzincins are endopeptidases and are generally synthesized as inactive precursors (14). In some members of the metzincins, such as the M8, M10, M11 and M12 families, a conserved cysteine residue in the pro-peptide can interact with the catalytic zinc ion to prevent the binding of a water molecule and inactivate the protease. This inhibition mechanism is known as a "cysteine switch" (14).

The M12 family is the second largest family of metzincins and can be further divided into two subfamilies: subfamily M12A and M12B (14). Subfamily M12A are known as astacin subfamily (18). These metalloproteases from M12A subfamily contain a conserved aspartate residue (Asp21) in the pro-peptide and their activation mechanism is therefore called "aspartate-switch" (19). Proteases in the M12A family are mainly found throughout the animal kingdom and only rarely found in bacteria. Flavastacin (20-22) is the only confirmed bacterial protease in this family. Members in subfamily M12B (also called the reprolysin family) employ a cysteine-switch mechanism (14). Reprolysin, found in snake venom, is the representative member. A cysteine-switch mechanism for the pro-peptide is thought to operate in the inhibition of reprolysin. Proteases in the M12B subfamily are also widely distributed throughout the animal kingdom.

The *Myroides* sp. cslb8, isolated from silkworm excrement in our lab, secretes a 25-kDa protease. Mass spectrometry analysis and BLAST analysis of the amino acid sequence indicated that this protease shares high sequence identity with the reported myroilysin from *M. profundi* D25, which was classified as an M12A subfamily protease based on a sequence alignment (23). However, myroilysin is classified into the M10B family in the MEROPS (24) peptidase database without structural information.

The M10 family, like the M12 family, is synthesized as inactive precursor and can be divided into three subfamilies: M10A, M10B and M10C. The M10A subfamily proteases are mainly found throughout the animal kingdom and activated through a cysteine-switch mechanism.

The best-known examples are the eukaryotic matrix metallopeptidases (MMPs). The M10A proteases are also present in plants and bacteria, but the non-animal M10A proteases do not usually employ a cysteine-switch activation mechanism (14). Karilysin is a typical bacterial M10A member; it is expressed as a proenzyme with an aspartate in the pro-peptide that may act in a similar manner to proastacin (25,26). Proteases in the M10B subfamily are mainly from Proteobacteria, including the well-studied serralysin (27).

Sequence analysis revealed a cysteine (Cys26) in the predicted propeptide and a typical M-turn (SIMHY) in the amino acid sequences of the myroilysin from *Myroides* sp. cslb8. We therefore investigated the question: Should myroilysin belong to the M12 or the M10 family?

To elucidate the catalytic mechanism of myroilysin and clarify the classification of myroilysin, genes encoding myroilysin and pro-myroilysin from *Myroides* sp. cslb8 were cloned and expressed. The 1.89 Å and 1.6 Å crystal structures of pro-myroilysin were resolved. The structure comparison revealed that myroilysin shares high similarity to the members of the M12A, M10A and M10B families of metalloprotease; in particular, the cysteine residue in the pro-peptide coordinates the catalytic zinc ion and inhibits the activity of myroilysin. A unique "cap" structure tops the active site cleft in the structure of pro-myroilysin. Further structure-based sequence analyses suggest that myroilysin should be a new bacterial member of the M12A family using a cysteine-switch mechanism.

RESULTS and DISCUSSION

Isolation and characterization of strain cslb8—Strain cslb8, an orange colony with the largest clear zone on the skim milk plate, was selected for further study. The cells of cslb8 were gram-negative and rod-shaped. These cells grow well at pH 7-10, with an optimum pH for growth at 7. No acid was produced from glucose, lactose, sucrose and fructose. To determine the evolutionary relationship of cslb8 to other bacteria, the 16S rDNA was amplified and sequenced. The 16S rRNA gene of strain cslb8 is 1,473 bp (KP282832). The phylogenetic trees were constructed using the neighbor-joining method (28) with 100% bootstrap support. Phylogenetic distances were calculated using the MEGA6 software package. The 16S rRNA gene sequence of strain cslb8 shared the highest identities with the sequences of *Myroides odoratimimus* LWD09 (98%, GU570427), *M. profundi* D25 (98%, EU204978) and *M. odoratus* NBRC 14945 (95%, AB517709) (Fig. 1). These results indicated that the CSLB8 strain could be classified as *Myroides*, and thus we named it *Myroides* sp. cslb8 (CGMCC 1.15038).

Purification and characterization of protease—The protease was purified stepwise using ammonium sulfate precipitation and various chromatographic methods with monitoring protease activity. The purified enzyme appeared as a single band of approximately 25 kDa on 12% SDS-PAGE, indicating that it is the target protein. To further identify the protease, the protein sequence of the purified 25-kDa protein was determined using tandem mass spectrometry. The Mascot peptide fragment search result showed that the matching peptide fragments accounted for 16% (44/273) of the deduced amino acids of the zinc-dependent M12 metalloprotease of *M. odoratimimus* CIP101113 （EHO09775.1).

To clone the protease encoding gene, the gene encoding the protease was amplified using primers designed according to the consensus sequences of some bacterial M12 family

metalloproteases, cloned into pET24b and sequenced. The gene is 819bp long (KR611868), encoding a protein with 273 amino acids and a molecular weight of approximately 30 kDa. The deduced protein sequence analysis showed that the amino acid sequence is identical to the M12 family peptidase from *M. odoratimimus* CCUG 3837 (EKB05810.1) and CCUG 12700 (EPH13478.1), and shares 99% similarity with the peptidase from *M. odoratimimus* CIP 101113 and the myroilysin from *M. profundi* D25 (23).

Further sequence analysis showed that a signal peptide consisting of residues 1-31 and a pro-peptide consisting of residues 32-68 are present at the N-terminus, and residues 69-273 encode a mature protein (myroilysin) of approximately 25 kDa at the C-terminus.

Protein crystallization and Structure determination—As attempts to obtain diffracted crystals of native mature protein purified from the culture supernatant failed, the gene encoding the mature peptide was cloned and expressed in *E. coli* C43 (DE3). Unfortunately, no target protein was obtained, which may be ascribable to the toxic proteolytic activity of the myroilysin to host cells. Thus, the zymogen (pro-myroilysin) was expressed in *E. coli* C43 (DE3) and purified for subsequent protein crystallization. However, despite extensive crystallization trials, no crystals were obtained. We noted a relatively large number of lysine residues (14 residues), accounting for 5.79% of the total amino acid residues in pro-myroilysin. Because methylation of the ε-amino group of lysine is reported to be helpful for protein crystallization (29-31), reductive methylation was attempted to modify the pro-myroilysin surface. The methylated pro-myroilysin was further purified and crystallized. Large flake-like crystals of the pro-myroilysin were finally obtained.

As myroilysin is a zinc-dependent metalloprotease, a zinc ion should be present in the active site of pro-myroilysin. A dataset at the peak wavelength of zinc (1.2816 Å) was collected to solve the phase. The structure was determined using a single wavelength anomalous dispersion method (SAD) and refined to 1.89 Å. Another 1.6 Å dataset was collected from a different condition (0.1 M Bis-Tris pH 6.5, 40% PEG4000), and the structure was determined using the molecular replacement method with the 1.89 Å model as a template. The refined structures correspond well to crystallographic data and anticipated geometric values (Table 1).

Overall Structure—The pro-myroilysin monomer comprises 242 residues and is composed of a pro-peptide consisting of residues 1-37, an N-terminal domain consisting of residues 38-151 and residues 229-242 and a C-terminal domain consisting of residues 152-228. The overall structure indicates that pro-myroilysin is a spherical molecule (Fig. 2*A*, 2*C*). The crystal structure of pro-myroilysin from the 0.1 M HEPES-NaOH pH 7.5, 1.4 M sodium citrate condition was refined to 1.89 Å. The topology of pro-myroilysin is shown in Fig. 2*B*. The crystal belongs to the $P2_1$ space group with cell dimensions a = 72.2, b = 35.5, c = 93.8 Å. There are two almost identical molecules of pro-myroilysin with root mean square deviations (rmsd) of 0.35 Å in the asymmetric unit. In the pro-myroilysin molecule, 23% of residues form nine α-helices and 8% of residues form five β-strands. The pro-peptide in this model is fully visible (Fig. 2*A*); it includes a 3_{10}-helix (formed by residues 15-17) flanked by coils. The N-terminal domain mainly consists of six α-helices (α2-6 and α9) and a five-stranded (β1-β5) parallel β-sheet fragment. The β-sheet fragments pack against the two long α-helices (α4 and α6), and helix α9 is inserted into the cleft between helices α4 and α6. The C-terminal

domain is composed of two α-helices (α7-8) and coils. The two helices (α7-8) pack almost perpendicularly against each other and cover the active site.

The three-dimensional structure of pro-myroilysin from the 0.1 M Bis-Tris pH 6.5, 40% PEG4000 condition was resolved at 1.6 Å with R_{work} of 19.1% and R_{free} of 23.0%. The crystal belongs to the $P2_1$ space group with cell dimensions a = 51.2, b = 35.1, c = 64.3 Å. There is only one molecule of pro-myroilysin in the asymmetric unit. In the pro-myroilysin molecule, 20% of residues form seven α-helices and 7% of residues form four β-strands (Fig. 2*C*). Notably, the electron density is invisible for the fragment from residues Thr36 to Arg44. The topology of pro-myroilysin is shown in Figure 2*D*.

A comparison of the structures obtained from these two conditions revealed that they are almost identical, with an rmsd of 0.35 Å, but some differences between these two structures were still observed, including the regions from 19-23, 153-160 and 214-224 and missing residues from 36 to 44. These differences may stem from the crystal packing caused by intramolecular and intermolecular interactions.

Active site—The N-terminal domain and the C-terminal domain form a deep V-shaped cleft, and the active site is located at the bottom of the cleft (Fig. 2-3). In the long and deep active site cleft, the catalytic zinc ion at the bottom is coordinated in a triangular pyramidal geometry by three NE2 atoms of His140 (2.3 Å), His144 (2.1 Å) and His150 (2.1 Å), belonging to the signature motif HEXXHXXGXXH. The zinc cation is also coordinated by the SG atom of the side chain of the cysteine residue Cys26 in the pro-peptide, with a distance of 2.3 Å. The catalytic residue Glu141, following the first histidine zinc ligand (His140), is not involved in coordinating the zinc ion, but the OE2 atom of Glu141 forms a hydrogen bond (3.0 Å) with the SG atom of Cys26. The side chains of His140, Glu141 and His144 from the central helix α6 project into the deep active site cleft, and the α6 helix extends to Gly147, where it turns sharply to bring the third coordinator (His150) of zinc ion to the active site (Fig. 3*A*). The following Glu151, the direct neighbor of the third zinc ligand His150, is thought to be strictly conserved in the members of astacins (32).

In the other signature motif, SIMHY (residue 204-208), the oxygen atom of the Ile205 main chain carbonyl group forms a specific interaction with the ND1 atom of the first zinc ligand His140 through a hydrogen bond (2.7 Å) (Fig. 3*B*) and is involved in structure stability. However, the side chain of Tyr208 faces away from the catalytic zinc (Fig. 3), while the corresponding tyrosine is engaged in zinc and substrate binding and stabilization in mature astacin and serralysin (27,33,34).

Structure comparison of pro-myroilysin with homologues from M12A, M10A and M10B subfamilies—A structure similarity search performed with the atomic coordinates of pro-myroilysin using the DALI server yielded the M12 family proastacin (PDB:3LQ0) (19), pro-meprin (4GWM) (35), choriolysin (3VTG and 3LQB) (36), M10 family prokarilysin (PDB: 4R3V) (37), macrophage metalloelastase (PDB:1UTT) (38) and M12 family astacin (PDB: 1AST) (16), among others.

A structural comparison of pro-myroilysin with homologues from M12A, M10A and M10B subfamilies was also performed. Pro-myroilysin superimposes with the M12 family proteases pro-astacin (3LQ0) and pro-meprin (4GWM) with an rmsd of 3.2 Å (for 178 target pairs) and 2.9 Å (for 172 target pairs) and shares amino acid sequence identities of 20% and 21%, respectively. Figure 4*A* shows that the N-terminal β1-5 strands

of pro-myroilysin partially superimpose onto the corresponding β strands of pro-astacin and pro-meprin homodromously; helices α4, α6 and α9 also partially superimpose onto the corresponding helices of pro-astacin and pro-meprin homodromously. However, the β5 strand of pro-myroilysin is much shorter than the corresponding β-strand of pro-astacin and pro-meprin and instead forms a loop structure protruding outside of the protein. This may be caused by the helix α5 located between β4 and β5. There are also many structural differences between the structures; for example, the α3 helix could not superimpose onto the corresponding helix. The most remarkable structural differences are present at the C-terminus. The C-terminal domain of pro-myroilysin consists of coils and two helices (α7 and α8, purple in Fig. 2*A* and 2*C*); the two helices form the unique "cap" structure situated above the active site, while the C-terminal domains of pro-astacin and pro-meprin have four or five helices and three β strands, there is no obvious "cap"" structure above the active sites.

Moreover, pro-myroilysin superimposes onto the M10A family protease pro-karilysin (4R3V) with an rmsd of 2.7 Å (for 179 target pairs) and the M10B family protease serralysin with inhibitor (1AF0) with an rmsd of 3.5 Å (for 154 target pairs); the amino acid sequence identities were 15% and 21%, respectively. As shown in Figure 4B, the structure of pro-myroilysin also shows similar structural similarity to pro-karilysin and the catalytic domain of serralysin. The helices α4, α6 and α9 of pro-myroilysin partially superimpose onto the corresponding helices of pro-karilysin and the catalytic domain of serralysin homodromously, and β1-5 strands also superimpose well and are located on the surface of both molecules. The C-terminal domain of pro-karilysin and catalytic domain of serralysin also do not contain a "cap" structure. Overall, the superimposition of pro-myroilysin with M12A, M10A and M10B family proteases showed that the N-terminal domain of pro-myroilysin superimposes well with them, while the C-terminus is evidently structurally different, particularly the "cap" structure.

Detailed structural and sequence analyses were also performed. As mentioned above, the conserved Gly147 (equivalent to Gly99 of astacin and Gly181 of serralysin) and the M-turn SIMHY are also present, while the Tyr208 (equivalent to Tyr149 in astacin and Tyr216 in the serralysin, involved in zinc binding) points away from the zinc in pro-myroilysin (Fig. 3). In addition, the Tyr239 of pro-myroilysin also perfectly matches the Tyr194 of astacin and Tyr246 of serralysin, which are involved in locking the C-terminal helix to the molecular moiety. The 90-loop of pro-myroilysin is exactly equivalent to the loop connecting strands β2 and β3 in astacin and serralysin.

It has been reported that there are two disulfide bonds in astacin and that these two bonds are likely to be conserved among all astacins and contribute to shaping the active-site cleft (39). However, no disulfide bond is present in pro-myroilysin, as in serralysin and pro-karilysin.

Inhibition mechanism — For pro-astacin, pro-meprin and pro-karilysin, one striking difference could also be observed from pro-myroilysin. The pro-peptides of pro-astacin, pro-meprin or pro-karilysin, containing 34, 37 or 14 residues, run through the cleft, and a side-chain atom of a conserved aspartate residue (Asp21, Asp52 or Asp25) of the pro-peptide is anchored to the catalytic zinc ion to replace the zinc-binding solvent molecule (Fig. 4*C-E*). This is the "aspartate-switch" activation mechanism (19). For pro-myroilysin, the pro-peptide consisting of 37 residues runs through and occupies the active site

cleft to prevent access to peptide substrates. Instead of an aspartate residue, the side-chain sulfur atom of Cys26 of the pro-peptide coordinates the catalytic zinc ion to expel the catalytic water molecule from the activity site, allowing the inhibition of pro-myroilysin catalysis and is so-called "cysteine-switch" mechanism (Figs. 3, 4*C*-*E*). Other than this difference, the direction of binding in the pro-peptide of the active site cleft of pro-myroilysin is the same as that seen in the pro-astacin structure (Fig. 4*A*).

Interestingly, the segment including the conserved "cysteine switch" is unique in pro-myroilysin; the AKVCKDV motif has never been reported, while PRCGXPD is conserved in MMPs, PKMCGV in ADAMs, HRCIHD in leishmanolysins and CG in pappalysins (17).

PISA (40) was used to analyze the interactions between the pro-peptide and the mature peptide of pro-myroilysin. The pro-peptide main chain, consisting of 37 residues, extends along the spherical molecule surface of the mature peptide to the active site pocket. Starting at Arg9, the residues bind the surface of the mature enzyme; from Ala23, the main chain runs through the active site pocket. The pro-peptide of pro-myroilysin is firmly bound to the mature peptide through several intramolecular interactions, such as hydrogen bonds and salt bridges. Eleven hydrogen bonds are formed between the eight residues (Arg9, Asp15, Val25, Cys26, Lys27, Asp28, Asp35 and Pro36) of the pro-peptide and the eleven residues (Gly38, Ala39, Ala100, Ser102, Asn116, Glu117, His150, Tyr173, Asn175, Tyr176 and Tyr190) of the mature peptide (Fig. 5*A*); two salt bridges form between the Cys26 of the pro-peptide and Tyr170 and Tyr208 of the mature peptide (Fig. 5*B*). In addition, the cysteine residue of the pro-peptide itself (Cys26) is tightly bound to the catalytic zinc ion (Fig. 3). Therefore, the pro-peptide is firmly fixed in the active site pocket.

In the reported mature enzymes of the M10 and M12 family of proteases, the catalytic water molecule is anchored by the glutamate residue following the first histidine zinc ligand (41) and is thought to be responsible for the nucleophilic attack to the carbon atom of the carbonyl group of the substrate peptide bond (14). However, in pro-myroilysin the catalytic water molecule is absent, and instead a cysteine residue (Cys26) in the pro-peptide coordinates the zinc ion. Cys26 also forms a hydrogen bond with the catalytic Glu141 residue (Fig. 3). Thus, the presence of Cys26 expels the catalytic water molecule away from the active site and results in inhibition of protease activity.

A unique cap structure in pro-myroilysin may be involved in substrate binding—Structure-based sequence alignment of pro-myroilysin with other proteases with Z scores above 10 indicate that these proteases belong to the M10 or M12 subfamily, especially the M12A, M10A and M10B subfamilies. However, the insert between the two conserved motifs of pro-myroilysin (HEXXHXXGXXH and M–turn SIMHY) are much longer than in other homologs (Fig. 6). This fragment (residues 160-193) includes the helices α7 and α8, forming the "cap" in the pro-myroilysin structure (Fig. 2*A*, 2*C*).

Work by Chen showed that myroilysin has broad specificity and high elastinolytic activity, indicating that the active cleft of myroilysin is opened before the substrate accesses it (23). The unique "cap" structure in pro-myroilysin, which does not exist in the other structure-reported members of the M12 or M10 family, may help to stabilize the pro-peptide through hydrogen bonds (Fig. 5*A*). Tyr170 and Tyr190 are two unique amino acid residues in the "cap" structure. The oxygen atom of Tyr190 forms a hydrogen bond (2.9 Å) with the third histidine zinc ligand

(His150), and therefore the Tyr190 residue seems to stabilize the conformation of the third histidine and the pro-peptide of pro-myroilysin (Fig. 5*A*). Furthermore, the Tyr170 residue in the "cap" could also stabilize the pro-peptide through water-mediated hydrogen bonds (Fig. 5*B*).

In our pro-myroilysin structure, the "cap" domain covers the pro-peptide and forms a tunnel that holds the pro-peptide inside (Fig. 5*C*). If the tunnel exists throughout the entire catalytic cycle, the substrate must be inserted into the tunnel for cleavage. However, because the tunnel is not wide enough to directly let the substrate get into the tunnel, we hypothesized that there is a conformational change to expose the active site, with the "cap" moving away from the active site after the N-terminal pro-peptide of pro-myroilysin is proteolytically removed.

In addition, it has been reported that the Glu103 of astacin, the amino acid residue just after the third zinc-binding H in the HEXXHXXGXXH motif, is important for structural stability due to its water-mediated salt bridge to the N-terminal Ala1 after the propeptide is cleaved (42). The main chain of astacin rotates 180° around the Ψ main-chain angle of the new N-terminal residue to allow the N-terminus to bury into the mature enzyme body to maintain the structural features. The removal of the prosegment would offer sufficient space for the activation domain to enclose its substrate (39). This Glu is thought to be the family-specific residue of astacin family (M12A) proteases, while the corresponding residue is a proline or serine in the serralysin family of protease (M10B) or MMPs (M10A) (32,43). Inspection of the mature N-termini of representative astacin family members also showed that N-terminal residues are almost exclusively alanine or asparagine (39). Glu151 is the corresponding amino acid residue in the HEXXHXXGXXH motif of pro-myroilysin. However, the Glu151 is far from the Gly38, the first amino acid residue of the mature myroilysin (23) in the structure of pro-myroilysin. Instead, Glu151 can form hydrogen bonds with Thr195 and Gln196 just after the α7 and α8 helices ("cap" structure), which do not exist in astacin and keeps the Glu151 from the Gly38 in the pro-myroilysin (Fig. 7). This may imply that the unique "cap" structure in pro-myroilysin might change the structure arrangement. However, the real role of the "cap" structure in the mature form of myroilysin is unknown yet.

The structure of mature myroilysin would answer the question, but unfortunately we failed to obtain diffractable myroilysin crystals of native mature myroilysin from the culture supernatant of cslb8 and failed to obtain myroilysin expressed in *E. coli*. This may imply that myroilysin with proteolytic activity is toxic to *E. coli* cells.

Structure and sequence analysis suggest myroilysin is a bacterial M12 family protease— The M10 and M12 families of proteases are two major families of metzincins. They shares quite high sequence similarity, zymogen activation and catalytic mechanisms, including the signature conserved motifs HEXXHXXGXXH/D and Met-turn (14).

The M12 family is the second largest family in the metzincins, and proteases in this family are mainly found throughout the animal kingdom and rarely in bacteria. Though quite a few structures from the M12 family have been solved, no structural information is available for M12 family proteases from bacteria. The flavastacin from *F. meningosepticum* (20-22) is the only reported bacterial proteases in this family. Myroilysin is a newly identified metazincin from *M. profundi* D25, and amino acid sequences analysis showed that it belongs to the M12A family of proteases (23). However, myroilysin is classified into the M10B family in

the peptidase database MEROPS.

Here, we report the crystal structures of pro-myroilysin in two different crystal forms from *Myroides* sp. cslb8. A structural comparison of pro-myroilysin to some M12 and M10 family proteases, including astacin, meperin, karilysin and serralysin, is shown in Figure 4. The overall structural comparison with these M12 and M10 family proteases showed that the N-terminal domain of pro-myroilysin superimposes well, while the C-terminal domain is structurally different. The signature motif HEXXHXXGXXH and the M-turn (SIMHY) of pro-myroilysin are strictly conserved (Fig 6). The three histidine residues belonging to the HEXXHXXGXXH motif of these three proteins are structurally conserved and function as zinc ligands (Fig. 4*C-E*).

The Dali server results showed that the structure of pro-myroilysin shares some similarity with M12A, M10A and M10B family proteases, such as pro-astacin, pro-meperin and serralysin. The phylogenetic tree of the myroilysin structure-based homologs based on the Dali server (Z>10) was then constructed. Figure 8 shows that distinct clusters form for M10 and M12 family proteases; pro-myroilysin is clearly more closely related to M12A family proteases than M10B family proteases or M10A family proteases. The above results have showed a conserved Glu residue just after the third zinc-binding His in the HEXXHXXGXXH motif, which is thought to be a family-specific residue of the astacin family; non-animal M10A proteases do not usually employ a cysteine-switch activation mechanism (as in pro-karilysin). We conclude that pro-myroilysin should belong to the M12A family rather than the M10 family. However, pro-myroilysin also forms a distinct branch with structurally determined M12A family proteases in the phylogenetic tree and has a different activation mechanism (cysteine-switch mechanism) from the M12A family. In addition, due to a special "cap" structure, different N-terminus amino acid residues in the mature enzyme, a different binding character of Glu151 and its function in pro-myroilysin, we also conclude that myroilysin should be a new member of the M12A family of proteases or may even form a new M12 subfamily. In summary, our crystal structures and structural comparison with M12 and M10 family proteases contribute new insights into the classification of the M12 metalloprotease family and may imply a divergent evolution of this family.

MATERIALS AND METHODS

Biological material and culture conditions — Bacterial strain cslb8 was isolated from silkworm feces on the LB medium (1% NaCl, 1% tryptone and 0.5% yeast extract) with 2% skim milk plate. Strain cslb8 was inoculated in LB medium on rotary shaker at 180 rpm at 28 ℃ for 12 h, and then transferred into 1L LB medium at 28 ℃ for another 12 h. The culture was centrifuged at 8000 rpm, and the supernatant was then used for protein isolation and purification.

16S rDNA sequencing—The 16S rDNA gene of CSLB8 was amplified by PCR using the genomic DNA as template with the common primers, and the PCR product was then cloned and sequenced. Multiple alignment of the sequence was performed using the ClustalW program (44).

Protein purification and sequence determination — All the following purification procedures were performed at 4 ℃ if not particular indicated. Ammonium sulfate was added to the culture supernatant to reach 80 % saturation. The precipitate was collected after centrifugation at 15000 rpm for 30 min. The precipitate was dissolved in 40mM KH_2PO_4/K_2HPO_4 (pH7.6) and diluted with an equal volume of 2 M ammonium

sulfate solution. After centrifugation, the sample was then loaded onto an Octyl Sepharose 4 fast flow chromatography column (2 ml bed volume) pre-equilibrated with binding buffer (20 mM pH7.6 KH_2PO_4/K_2HPO_4 and 1 M $(NH_4)_2SO_4$). After washing with 5 bed volumes of binding buffer, the protein was eluted with a linear gradient of 1 M to 0 M ammonium sulfate. The eluted protease was then concentrated by the Amicon Ultra15 centrifugal filter unit with 10 kDa molecular weight cut-off (Merck Milipore) and further purified by the Sephadex G200 gel filtration column pre-equilibrated with 20mM Tris-HCl (pH 8.0), 300 mM NaCl. The purified protein was then eluted with the same buffer. The purified protein was then inspected on 12% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The protein concentration was measured using the PE Lamda 25 UV/VIS spectrometer at 280 nm.

After SDS-PAGE separation and gel staining, the single protein band was cut off from the gel; the protein was then sequenced by ultrafleXtreme MALDI-TOF/TOF mass spectrometer (Bruker Doltonics). Peptides sequence were identified by searching the peak list against the Mass Spectrometry Protein Sequence Database using the Mascot version 2.1 search engine (45).

Gene amplification and cloning — To amplify the gene encoding the purified protease, a pair of degenerated primers, 5'-gacgcatATGAAATTACACCACAAGATCC (upstream primer 1) and 5'-cagctcgagGTTTCTTGGATAMACTGTTGC (downstream primer 2), were designed according to the result of the mass spectrometry and the sequence alignment of the genes encoding M12 family metalloproteases from bacteria. The restriction sites (Nde I and Xho I) are underlined respectively. PCR amplification was carried out with Pfu DNA polymerase (Thermo Fermentas) using the genomic DNA of cslb8 as template. The PCR product was then cloned into the pET24b and sequenced by the Beijing Luhe Technology Co. Ltd.

Protein Expression and Purification—For expression and purification of myroilysin and pro-myroilysin, the genes encoding the two proteins were respectively amplified through PCR with the chromosomal DNA of *Myroides* sp. cslb8 as template. PCR was respectively performed with pfu DNA polymerase and the following primers: myroilysin_forward 5'-atcgcatatgGGGGCTGTTGTCAGAAGTACA AAG-3', and myroilysin_reverse 5'-cagctcgagGTTTCTTGGATAMACTGTTGC-3 '; pro-myroilysin_forward 5'-tacgcatatgAGTAGTAAGGGGCTAAAAGAA TTAAG-3', and myroilysin_reverse. The amplified genes and the pET24b vector were digested with NdeI and XhoI restriction enzymes and ligated, respectively. The recombinant plasmids (pET24b_myroilysin and pET24b_pro-myroilysin) were verified by restriction reaction and DNA sequencing, and the correct recombinant plasmids were then respectively transformed into expression strain *E. coli* C43 (DE3), respectively.

The fresh transformants were grown in LB medium containing kanamycin (30 μg/ml) at 37 ℃ until the A_{600} reached 0.7. The cells were cooled to 25 ℃, and then induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) by incubating at 25 ℃ for another 5 h. The induced cells were harvested, carefully resuspended in binding buffer (50 mM pH 7.6 KH_2PO_4/K_2HPO_4, 300 mM NaCl, 5 mM imidazole, 10% (v/v) glycerol) supplemented with 1 mM PMSF, and then sonicated in an ice bath. After removing insoluble materials and unbroken cells by centrifugation, the supernatant was applied to a 2 ml bed volume nickel-nitrilotriacetic acid (Ni-NTA, GE Life Sciences) column which was pre-equilibrated with 6 bed volumes of binding

buffer. After washing the column with the buffers containing different concentration of imidazole (5, 20 and 50 mM, six bed volumes each), the pro-myroilysin was eluted with buffer containing 100 mM imidazole.

The buffer of eluted protein was exchanged with 50 mM pH 7.6 KH_2PO_4/K_2HPO_4, 300 mM NaCl, and 10% (v/v) glycerol using a desalting column and then the target protein was concentrated to 8 mg/ml for the subsequent methylation (referring to a methylation protocol, Hampton Research). Methylation was carefully performed by dimethylamine borane complex and formaldehyde as previously reported (46). The methylated pro-myroilysin was concentrated to about 1 ml using an Amicon Ultra-10 filter (Millipore), and then loaded onto a Superdex 200 column pre-equilibrated with 20 mM Tris-Cl pH 8.5, 300 mM NaCl, and 10% (v/v) glycerol. The column was then carefully eluted with 1.2 bed volumes of the pre-equilibrated buffer. The elution pattern showed that there was only a single peak, which was further analyzed by SDS–PAGE to check the purity of pro-myroilysin. Pro-myroilysin was pooled together and concentrated to 20 mg/ml for crystallization.

Protein Crystallization — Crystallization was performed with the commercial kits from Hampton Research and Microlytic using the sitting drop vapor diffusion method at 4 ℃ and 22 ℃. The initial screen yielded flaky-like crystals at 22 ℃ from the condition: 0.1 M Bis-Tris pH 6.5, 40% PEG4000, and flaky crystals at 4 ℃ from the following three conditions: 0.1 M HEPES-NaOH pH 7.5, 1.4 M sodium citrate; 1.6 M sodium citrate; 0.1 M Tris-Cl pH 8.5, 1.25 M sodium citrate. Since flaky-like crystals at the condition (0.1 M Bis-Tris pH 6.5, 40% PEG4000) grow better and need a short time to grow to full size (about seven days), this condition was chosen for further optimization. A systematic grid screening of precipitant concentrations, pH and protein concentrations was set up to optimize the crystal. The best condition was then optimized with the commercial additive screening kit (Hampton Research).

Data Collection, Structure Determination and Refinement — The crystal was carefully looped out from the crystallization drop and quickly cooled in liquid nitrogen. Data collection was performed at 100 K. X-ray diffraction datasets were collected at beamline BL17U1 of Shanghai Synchrotron Radiation Facility (47). The crystal from the following condition (0.1 M HEPES-NaOH pH 7.5, 1.4 M sodium citrate) was diffracted to 1.89 Å. A full data set of 540 frames was collected at the peak wavelength (1.2816 Å) of Zinc ion. Each frame was exposed with a rotation range of 1.0 for 1.2 s. Diffraction data were processed with XDS (48). The structure of pro-myroilysin was solved using the SAS protocol of Auto-Rickshaw: the EMBL-Hamburg automated crystal structure determination platform (49). The input diffraction data were prepared and converted for use in Auto-Rickshaw using programs of the CCP4 suite (50). FA values were calculated using the program SHELXC (51). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 2.4 A. All of the 2 heavy atoms requested were found using the program SHELXD (52). 84.30% of the model was built using the program ARP/wARP (53,54). The crystal from condition 0.1 M Bis-Tris pH 6.5, 40% PEG4000 was diffracted to 1.6 Å. A dataset of 360 frames was collected at wavelength 0.9791 Å. The crystal structure of pro-myroilysin from this condition was solved with molecular replacement method using Phaser using the 1.89 Å crystal structure of pro-myroilysin (55). The structure model was

manually adjusted with Coot (56) and structure refinement was performed with REFMAC and Phenix (57,58). The models were validated with MolProbity (59). All figures were drawn with PyMol (60) and the TopDraw (61). Data processing and refinement statistics are summarized in Table 1.

PDB Deposition—The coordinates and the related structural factors were deposited in the Protein Data Bank (PDB) under the code number 5CZW and 5GWD.

Sequence Alignment — Structure-based amino acid sequence alignment of members of the M12 family was performed with daili server and then redrawn with ClustalX 1.81 (62) and GENEDOC (63).

ACKNOWLEDGMENTS.

We thank the staff of BL17B/BL18U1/BL19U1 beamlines at National Center for Protein Sciences Shanghai and Shanghai Synchrotron Radiation Facility, and the staff of beamline BL17U of Shanghai Synchrotron Radiation Facility Shanghai, People's Republic of China, for assistance during data collection.

FUNDING INFORMATION.

This work was supported by grants from the National Natural Science Foundation of China (31400055, 31170686 and 31100028), from the Natural Science Foundation of Jiangsu Province (BK20140690) and the Youth Science and Technology Innovation Fund from Nanjing Agricultural University (KJ2013027).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

WW, DX and TR conceived the study. DX, JZ, XL and TR conducted the experiments. All of the authors analyzed the data. All authors wrote the manuscript.

REFERENCES

1. Gupta, R., Beg, Q. K., and Lorenz, P. (2002) Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* **59**, 15-32
2. Sastry, K. J., Srivastava, O. P., Millet, J., FitzJames, P. C., and Aronson, A. I. (1983) Characterization of Bacillus subtilis mutants with a temperature-sensitive intracellular protease. *J. Bacteriol.* **153**, 511-519
3. Reysset, G., and Millet, J. (1972) Characterization of an intracellular protease in B. subtillus during sporulation. *Biochem. Biophys. Res. Commun.* **49**, 328-334
4. Okujo, N., Akiyama, T., Miyoshi, S., Shinoda, S., and Yamamoto, S. (1996) Involvement of vulnibactin and exocellular protease in utilization of transferrin- and lactoferrin-bound iron by Vibrio vulnificus. *Microbiol. Immunol.* **40**, 595-598
5. Fleming, A. B., Tangney, M., Jorgensen, P. L., Diderichsen, B., and Priest, F. G. (1995) Extracellular enzyme synthesis in a sporulation-deficient strain of Bacillus licheniformis. *Appl. Environ. Microbiol.* **61**, 3775-3780
6. Nishina, Y., Miyoshi, S., Nagase, A., and Shinoda, S. (1992) Significant role of an exocellular protease in utilization of heme by Vibrio vulnificus. *Infect. Immun.* **60**, 2128-2132
7. Frankena, J., Koningstein, G. M., Vanverseveld, H. W., and Stouthamer, A. H. (1986) Effect of Different Limitations in Chemostat Cultures on Growth and Production of Exocellular Protease by Bacillus-Licheniformis. *Appl. Microbiol. Biotechnol.* **24**, 106-112
8. Miyoshi, S., and Shinoda, S. (2000) Microbial metalloproteases and pathogenesis. *Microbes and infection / Institut Pasteur* **2**, 91-98
9. Ichida, J. M., Krizova, L., LeFevre, C. A., Keener, H. M., Elwell, D. L., and Burtt, E. H. (2001) Bacterial inoculum enhances keratin degradation and biofilm formation in poultry compost. *J. Microbiol. Methods* **47**, 199-208
10. Maeda, H. (1996) Role of microbial proteases in pathogenesis. *Microbiology and immunology* **40**, 685-699
11. Cheng, S. W., Hu, H. M., Shen, S. W., Takagi, H., Asano, M., and Tsai, Y. C. (1995) Production and characterization of keratinase of a feather-degrading Bacillus licheniformis PWD-1. *Biosci. Biotechnol. Biochem.* **59**, 2239-2243
12. Aehle, W., Sobek, H., Amory, A., Vetter, R., Wilke, D., and Schomburg, D. (1993) Rational Protein Engineering and Industrial Application - Structure Prediction by Homology and Rational Design of Protein-Variants with Improved Washing Performance - the Alkaline Protease from Bacillus-Alcalophilus. *J Biotechnol* **28**, 31-40
13. Zhu, W. S., Wojdyla, K., Donlon, K., Thomas, P. A., and Eberle, H. I. (1990) Extracellular proteases of Aspergillus flavus. Fungal keratitis, proteases, and pathogenesis. *Diagn. Microbiol. Infect. Dis.* **13**, 491-497
14. Rawlings, N. D., and Barrett, A. J. (2013) *Handbook of Proteolytic Enzymes Introduction: Metallopeptidases and Their Clans.* , Elsevier, 32 Jamestown Road, London NW1 7BY, UK
15. Tallant, C., Marrero, A., and Gomis-Ruth, F. X. (2010) Matrix metalloproteinases: Fold and function of their catalytic domains. *Bba-Mol Cell Res* **1803**, 20-28
16. Bode, W., Gomis-Ruth, F. X., Huber, R., Zwilling, R., and Stocker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* **358**, 164-167
17. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomisruth, F. X., Mckay, D. B., and Bode, W. (1995) The Metzincins - Topological and Sequential Relations between the Astacins, Adamalysins, Serralysins,

and Matrixins (Collagenases) Define a Superfamily of Zinc-Peptidases. *Protein Sci.* **4**, 823-840

18. Mac Sweeney, A., Gil-Parrado, S., Vinzenz, D., Bernardi, A., Hein, A., Bodendorf, U., Erbel, P., Logel, C., and Gerhartz, B. (2008) Structural basis for the substrate specificity of bone morphogenetic protein 1/tolloid-like metalloproteases. *J. Mol. Biol.* **384**, 228-239
19. Guevara, T., Yiallouros, I., Kappelhoff, R., Bissdorf, S., Stocker, W., and Gomis-Ruth, F. X. (2010) Proenzyme structure and activation of astacin metallopeptidase. *J. Biol. Chem.* **285**, 13958-13965
20. Tarentino, A. L., Quinones, G., Grimwood, B. G., Hauer, C. R., and Plummer, T. H., Jr. (1995) Molecular cloning and sequence analysis of flavastacin: an O-glycosylated prokaryotic zinc metalloendopeptidase. *Arch. Biochem. Biophys.* **319**, 281-285
21. Reinhold, B. B., Hauer, C. R., Plummer, T. H., and Reinhold, V. N. (1995) Detailed structural analysis of a novel, specific O-linked glycan from the prokaryote Flavobacterium meningosepticum. *J. Biol. Chem.* **270**, 13197-13203
22. Plummer, T. H., Jr., Tarentino, A. L., and Hauer, C. R. (1995) Novel, specific O-glycosylation of secreted Flavobacterium meningosepticum proteins. Asp-Ser and Asp-Thr-Thr consensus sites. *J. Biol. Chem.* **270**, 13192-13196
23. Chen, X. L., Xie, B. B., Bian, F., Zhao, G. Y., Zhao, H. L., He, H. L., Zhou, B. C., and Zhang, Y. Z. (2009) Ecological function of myroilysin, a novel bacterial M12 metalloprotease with elastinolytic activity and a synergistic role in collagen hydrolysis, in biodegradation of deep-sea high-molecular-weight organic nitrogen. *Appl. Environ. Microbiol.* **75**, 1838-1844
24. Rawlings, N. D., Barrett, A. J., and Finn, R. (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **44**, D343-D350
25. Guevara, T., Ksiazek, M., Skottrup, P. D., Cerda-Costa, N., Trillo-Muyo, S., de Diego, I., Riise, E., Potempa, J., and Gomis-Ruth, F. X. (2013) Structure of the catalytic domain of the Tannerella forsythia matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69**, 472-476
26. Cerda-Costa, N., Guevara, T., Karim, A. Y., Ksiazek, M., Nguyen, K. A., Arolas, J. L., Potempa, J., and Gomis-Ruth, F. X. (2011) The structure of the catalytic domain of Tannerella forsythia karilysin reveals it is a bacterial xenologue of animal matrix metalloproteinases. *Mol. Microbiol.* **79**, 119-132
27. Wu, D., Ran, T., Wang, W., and Xu, D. (2016) Structure of a thermostable serralysin from Serratia sp. FS14 at 1.1 A resolution. *Acta Crystallogr F Struct Biol Commun* **72**, 10-15
28. Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-425
29. Sun, L. L., Chen, Y. X., Rajendran, C., Mueller, U., Panjikar, S., Wang, M. T., Mindnich, R., Rosenthal, C., Penning, T. M., and Stoockigt, J. (2012) Crystal Structure of Perakine Reductase, Founding Member of a Novel Aldo-Keto Reductase (AKR) Subfamily That Undergoes Unique Conformational Changes during NADPH Binding. *Journal of Biological Chemistry* **287**, 11213-11221
30. Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J. S., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617-1622
31. Rayment, I. (1997) Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods in enzymology* **276**, 171-179

32. Bode, W., Gomisruth, F. X., and Stockler, W. (1993) Astacins, Serralysins, Snake-Venom and Matrix Metalloproteinases Exhibit Identical Zinc-Binding Environments (Hexxhxxgxxh and Met-Turn) and Topologies and Should Be Grouped into a Common Family, the Metzincins. *FEBS Lett.* **331**, 134-140
33. Grams, F., Dive, V., Yiotakis, A., Yiallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stocker, W. (1996) Structure of astacin with a transition-state analogue inhibitor. *Nat. Struct. Biol.* **3**, 671-675
34. Baumann, U., Wu, S., Flaherty, K. M., and McKay, D. B. (1993) Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* **12**, 3357-3364
35. Arolas, J. L., Broder, C., Jefferson, T., Guevara, T., Sterchi, E. E., Bode, W., Stocker, W., Becker-Pauly, C., and Gomis-Ruth, F. X. (2012) Structural basis for the sheddase function of human meprin beta metalloproteinase at the plasma membrane. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16131-16136
36. Okada, A., Sano, K., Nagata, K., Yasumasu, S., Ohtsuka, J., Yamamura, A., Kubota, K., Iuchi, I., and Tanokura, M. (2010) Crystal structure of zebrafish hatching enzyme 1 from the zebrafish Danio rerio. *J. Mol. Biol.* **402**, 865-878
37. Lopez-Pelegrin, M., Ksiazek, M., Karim, A. Y., Guevara, T., Arolas, J. L., Potempa, J., and Gomis-Ruth, F. X. (2015) A novel mechanism of latency in matrix metalloproteinases. *J. Biol. Chem.* **290**, 4728-4740
38. Morales, R., Perrier, S., Florent, J. M., Beltra, J., Dufour, S., De Mendez, I., Manceau, P., Tertre, A., Moreau, F., Compere, D., Dublanchet, A. C., and O'Gara, M. (2004) Crystal structures of novel non-peptidic, non-zinc chelating inhibitors bound to MMP-12. *J. Mol. Biol.* **341**, 1063-1076
39. Gomis-Ruth, F. X., Trillo-Muyo, S., and Stocker, W. (2012) Functional and structural insights into astacin metallopeptidases. *Biol. Chem.* **393**, 1027-1041
40. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797
41. Gomis-Ruth, F. X., Stocker, W., Huber, R., Zwilling, R., and Bode, W. (1993) Refined 1.8 A X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish Astacus astacus L. Structure determination, refinement, molecular structure and comparison with thermolysin. *Journal of molecular biology* **229**, 945-968
42. Yiallouros, I., Kappelhoff, R., Schilling, O., Wegmann, F., Helms, M. W., Auge, A., Brachtendorf, G., Berkhoff, E. G., Beermann, B., Hinz, H. J., Konig, S., Peter-Katalinic, J., and Stocker, W. (2002) Activation mechanism of pro-astacin: Role of the propeptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing. *J. Mol. Biol.* **324**, 237-246
43. Stocker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Ruth, F. X., McKay, D. B., and Bode, W. (1995) The metzincins--topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **4**, 823-840
44. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948
45. Liu, L., Xu, L., Yan, F., Yan, R., Song, X., and Li, X. (2009) Immunoproteomic analysis of the second-generation merozoite proteins of Eimeria tenella. *Veterinary parasitology* **164**, 173-182
46. Rosenthal, C., Mueller, U., Panjikar, S., Sun, L. L., Ruppert, M., Zhao, Y., and Stockigt, J. (2006) Expression, purification, crystallization and preliminary X-ray analysis of perakine reductase, a new

member of the aldo-keto reductase enzyme superfamily from higher plants. *Acta Crystallogr F* **62**, 1286-1289

47. Wang, Q. S., Yu, F., Huang, S., Sun, B., Zhang, K. H., Liu, K., Wang, Z. J., Xu, C. Y., Wang, S. S., Yang, L. F., Pan, Q. Y., Li, L., Zhou, H., Cui, Y., Xu, Q., Earnest, T., and He, J. H. (2015) The macromolecular crystallography beamline of SSRF. *Nucl Sci Tech* **26**, 12-17
48. Kabsch, W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132
49. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2005) Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 449-457
50. Dodson, E. J., Winn, M., and Ralph, A. (1997) Collaborative Computational Project, number 4: providing programs for protein crystallography. *Methods Enzymol.* **277**, 620-633
51. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122
52. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772-1779
53. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463
54. Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A., and Lamzin, V. S. (2004) Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat* **11**, 56-59
55. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
56. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132
57. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367
58. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221
59. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66**, 12-21
60. Bramucci, E., Paiardini, A., Bossa, F., and Pascarella, S. (2012) PyMod: sequence similarity searches, multiple sequence-structure alignments, and homology modeling within PyMOL. *BMC Bioinformatics* **13**
61. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A., and Noble, M. (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2288-2294
62. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic acids research* **25**, 4876-4882

63. Nicholas, K. B., Nicholas, H. B., Deerfield, D. W. I., Nicholas, K. B., Nicholas, H. B., and Deerfield, D. W. I. (1997) GeneDoc: Analysis and visualization of genetic variation. *Embnew News* **4**

FOOTNOTES

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 5CZW and 5GWD.

FIGURE LEGEND

Figure 1. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequences showing the phylogenetic positions of strain CSLB8, other *Myroides* species and representatives of the *Flavobacteriaceae* family. *E. coli* (X80721) was used as an out-group. Percent bootstrap values above 50 (1,000 replicates) are indicated at nodes. Scale bar = 0.05 substitutions per nucleotide position.

Figure 2. Overall structure of pro-myroilysin.

A. Overall structure of pro-myroilysin, solved at 1.89 Å. The pro-peptide is colored orange; the polypeptide chain composed of residues 160 to 193 is purple; zinc ion is grey; other parts are cyan. B. Topology of the pro-myroilysin structure solved at 1.89 Å. The N-terminal domain is formed by five helices (α3-6 and α9) and four β-strands (β1-5). The C-terminal domain is formed by two helices (α7-8) and coils. C. Overall structure of pro-myroilysin solved at 1.6 Å. The pro-peptide is orange; the polypeptide chain composed of residues 160 to 193 is purple; zinc ion is grey; other parts are cyan. D. Topology of the pro-myroilysin structure solved at 1.6 Å. The N-terminal domain is formed from four helices (α4-6 and α9) and four β-strands (β1-3, and β5). The C-terminal domain is formed from two helices (α7-8) and coils. The diagrams were drawn using TopDraw.

Figure 3. Active-site of pro-myroilysin. A. Active site of pro-myroilysin. The catalytic zinc ion is represented as a gray sphere; the zinc ligands (His140, His144 and His150), the Cys26 residue of the pro-peptide and the conserved residues Glu141, Gly147, Tyr208 are rendered in cyan sticks; the pro-peptide, helix and coils are colored orange, red and green, respectively. B. Ile205 of the M-turn forms a hydrogen bond with the first zinc ligand His140. The corresponding electron density (2Fo_Fc) is light gray and contoured at 1.0 σ.

Figure 4. Overall structure and active site comparison of pro-myroilysin with that of M12 and M10 family proteases. A. Superimposed overall structures of pro-myroilysin (purple), M12 family pro-astacin (aquamarine) and pro-meprin (light grey). B. Superimposed active site overall structures of pro-myroilysin (purple) and M10 family pro-karilysin (light grey) and catalytic domain of serralysin (aquamarine). C-E. Superimposed active site structures of pro-myroilysin (purple) with pro-astacin, pro-meprin and pro-karilysin (light gray). The zinc ion of pro-myroilysin is shown as a purple sphere; the zinc ligands of pro-myroilysin are shown as sticks with purple carbon atoms; the zinc ions of pro-astacin, pro-meprin and pro-karilysin are shown as light gray spheres; the zinc ligands of astacin, pro-meprin and pro-karilysin are shown as sticks with light gray carbon atoms.

Figure 5. The interaction of the pro-peptide and mature peptide in pro-myroilysin. A The pro-peptide forms interactions with the mature peptide in pro-myroilysin. The residues shown as sticks with green carbon atoms belong to the pro-peptide; residues shown as sticks with carbon atoms colored yellow belong to mature peptide, except for the “cap” structure; the residues shown as sticks with carbon atoms colored sky blue belong to the “cap” structure. B. Cys26 forms interactions with Tyr170 and Tyr208 though water-mediated hydrogen bonds. Water molecules are shown as red spheres; Cys26, Tyr170 and Tyr208 are shown as light gray sticks. C. Tyr190 forms interactions with the third histidine zinc ligand His150. His150 and Tyr190 residues are shown as light gray sticks; corresponding electron

densities (2Fo_Fc) are colored marine blue and contoured at 1.0 σ. D. Surface contour image showing access to the active site pocket of pro-myroilysin.

Figure 6. Structure-based sequence alignment of pro-myroilysin with members of the M12 and M10 families. Aligned sequences are representative members of the M12 and M10 families obtained from the Dali server result. The strictly conserved residues are highlighted. Conserved zinc binding motif, M-turn and unique "cap" structure are indicated. Selected protein abbreviations: HCE1 (high choriolytic enzyme 1), ZHE1 (zebrafish hatching enzyme 1), BMP1 (bone morphogenetic protein_1), FC-1 (fibroblast collagenase 1), MT1-MMP (membrane type 1 matrix_metalloproteinase), HFC (human fibroblast collagenase), ProMMP-2 (pro-matrix metalloproteinase-2).

Figure 7. The conserved Glu151 just after the third zinc-binding His in the HEXXHXXGXXH motif of pro-myroilysin interacts with Thr195 and Gln196 just after the "cap" structure. The conserved Glu151 forms hydrogen bonds with Thr195 and Gln196 and distances itself far from Gly38 (the first amino acid residue in myroilysin) in pro-myroilysin. In contrast, the Ala1 of proastacin is much closer to the corresponding residue Glu101 and forms water-mediated salt bridges to the N-terminal Ala1 after the propeptide is cleaved. The pro-peptide, the "cap" structure and the rest of pro-myroilysin (including the zinc ion) are in orange, purple and cyan, respectively; Gly38, Glu 151, Thr195 and Gln196 of pro-myroilysin are rendered in green sticks; the pro-peptide and rest of pro-astacin are light yellow and grey, respectively. Glu 101 (light grey) and Ala1 of astacin (light grey), and Ala1 in pro-astacin (dark grey), are shown as sticks; the zinc ions in pro-astacin and astacin are shown in dark grey and light grey, respectively; water molecules are shown in red.

Figure 8. Structure-based phylogenetic tree illustrating the relationship between the pro-myroilysin and its homologs. The dataset of these homologs was obtained from DALI by selecting proteins with Z score>10 when compared to pro-myroilysin. For the sake of clarity, only 46 proteins are shown in the phylogenetic tree. The scale bar indicates phylogenetic distance. PDB ID and the protein names are listed. Selected protein abbreviations: ZHE1 (zebrafish hatching enzyme 1), HCE1 (high choriolytic enzyme 1), BMP1 (bone morphogenetic protein_1), ProMMP-2 (pro-matrix metalloproteinase-2), HFC (human fibroblast collagenase), FC-1 (fibroblast collagenase 1), MT1-MMP (membrane type 1 matrix_metalloproteinase), MMP20 (enamelysin), MMP-11(stromelysin-3).

Table 1 Data collection and refinement statistics.

	Peak	Native
Crystallization condition	0.1 M HEPES-NaOH pH7.5, 1.4 M sodium citrate	0.1 M Bis-Tris pH6.5, 40 % PEG4000
Data collection		
Wavelength (Å)	1.2816	0.9791
Temperature (K)	100	100
Crystal-to-detector distance (mm)	200	200
Rotation range per image (^{o})	1	1
Total rotation range (^{o})	540	360
Space group	$P2_1$	$P2_1$
Unit-cell parameters (Å)	a=72.2, b=35.5, c=93.8 β=101.4°	a=51.2, b=35.1, c=64.3 β=110.1°
Resolution range (Å)	19.90-1.89(1.99-1.89)[a]	19.27-1.6(1.69-1.60)[a]
Observed reflections	400119(42953)	211175(30523)
Unique reflections	37660(4989)	28643(4143)
Multiplicity	10.6(8.6)	7.4(7.4)
Rmerge (%)	12.8(70.1)	10.9(81.5)
Rmeas (%)	14.3(78.1)	11.8(87.6)
Rpim (%)	4.3(26)	4.3(32)
Completeness (%)	98.6(91.0)	99.9(100)
I/σ (I)	11.9(3.7)	11.9(2.7)
Structure refinement		
Total number of atoms	4000	2056
No. of reflections used	35832	27216
R_{work} (%)	18.15	19.09
R_{free} (%)	22.38	23.00
Rmsd bonds (Å)	0.0067	0.0062
Rmsd angles (^{o})	0.855	0.857
Ramachandran plot (%)		
Favored	98.5	96.44
Allowed	1.5	3.56
Outlier	0	0

[a]**Highest resolution shell is shown in parenthesis.**

Table 2 Members of the M12 and M10 family.

Source/description	PDB ID	Z-score	Backbone rmsd (Å)	Sequence identity (%)	Resolution (Å)	target pairs
Myroides sp. cslb8, pro-myroilysin	5CZW				1.6	
	5DPU				1.89	
Astacus astacus, proastacin	3LQ0	13.6	3.2	20	1.45	178
Homo sapiens, promeprin	4GWM	13.5	2.9	21	1.85	172
Tannerella forsythia, prokarilysin	4R3V	13.0	2.7	15	2.01	179
Astacus astacus, astacin	1AST	12.3	3.0	23	1.8	200
Tannerella forsythia, karilysin	2XS4	12.0	2.4	15	1.7	167
Homo sapiens, mature meprin	4GWN	11.5.	3.1	23	3.0	158
Serratia marcescens, serralysin with inhibitor	1AF0	10.8	3.5	21	1.8	154
Serratia marcescens， serralysin	5D7W	10.7	3.7	21	1.1	156

FIGURE 1

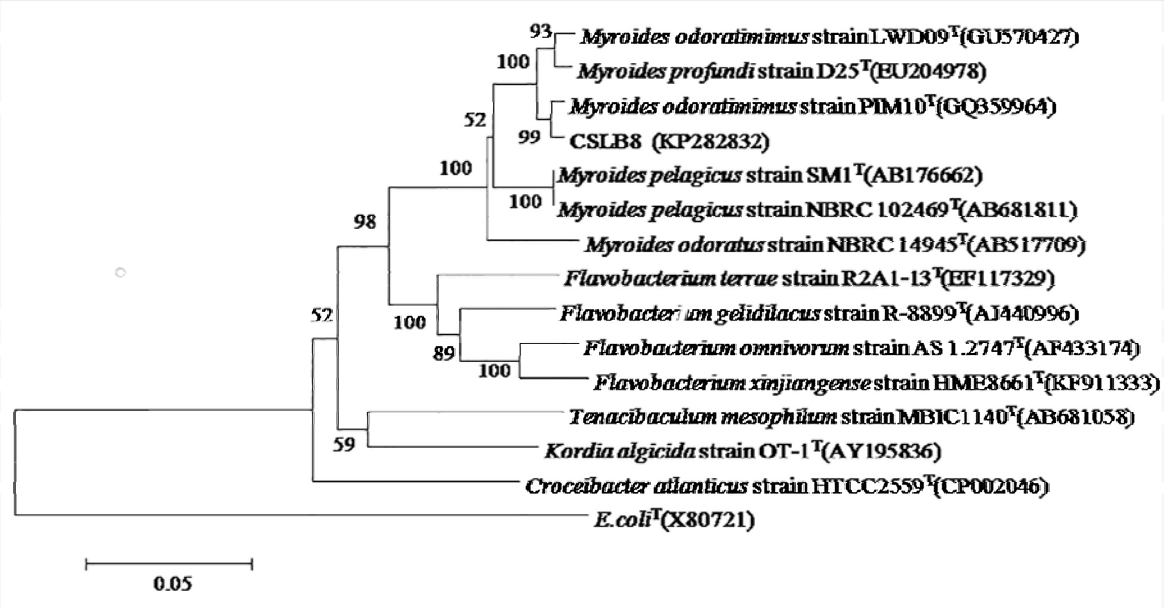


FIGURE 2

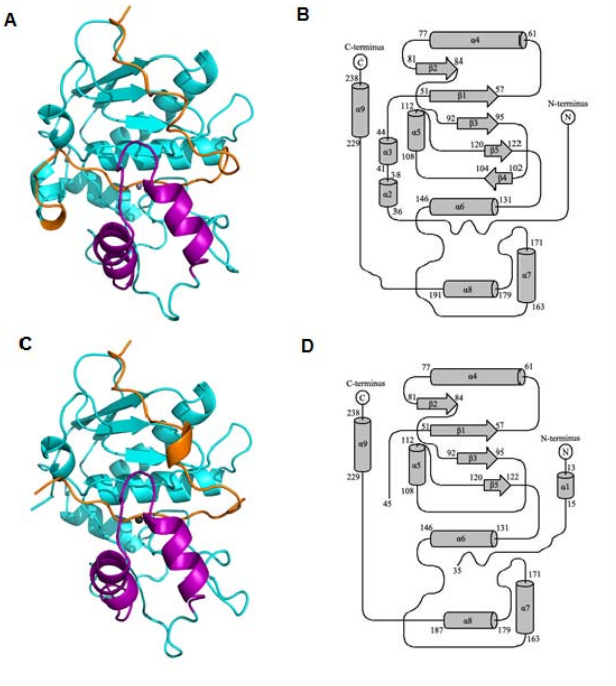


FIGURE 3

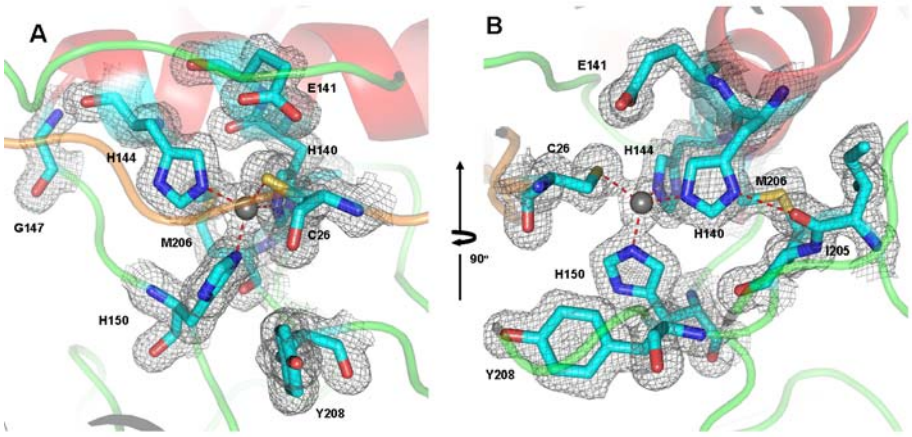


FIGURE 4

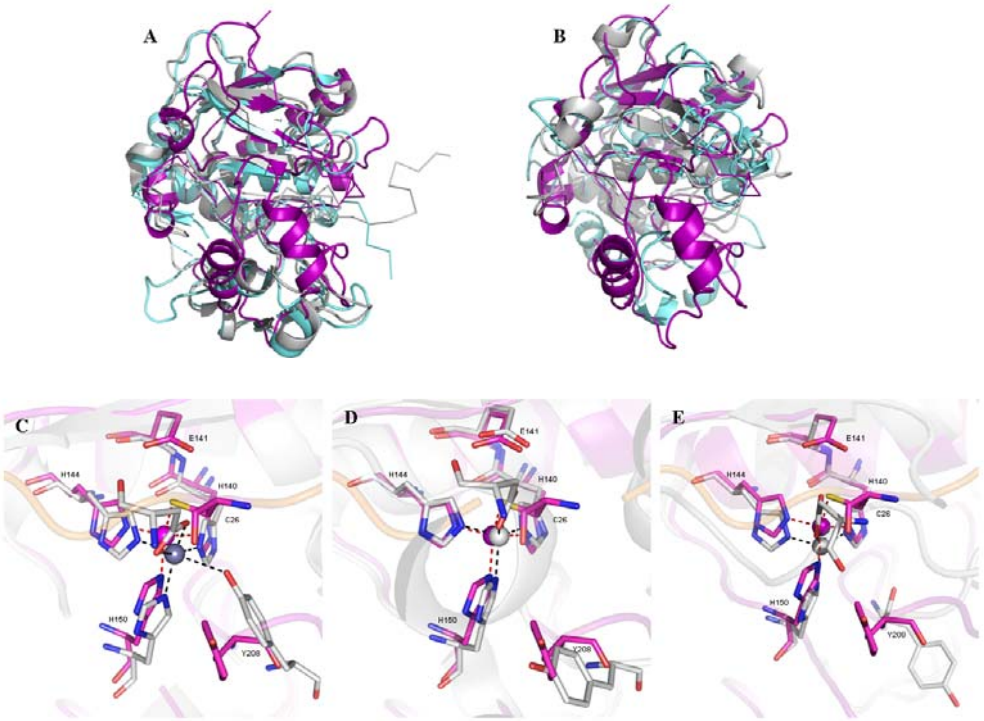


FIGURE 5

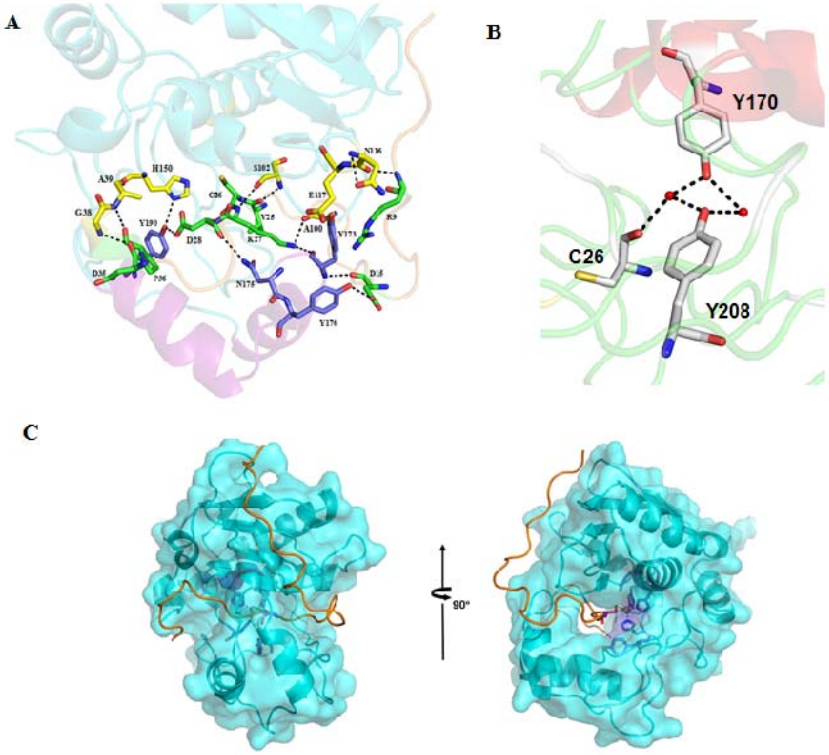


FIGURE 6

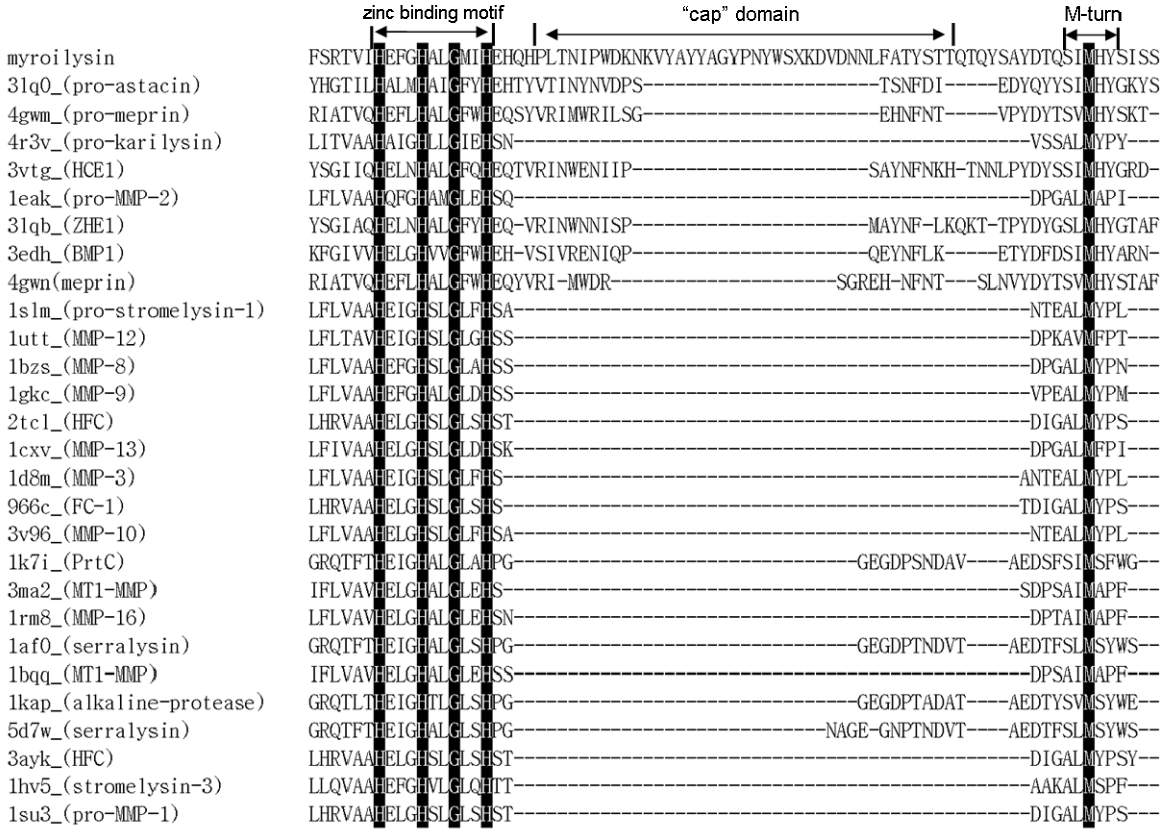


FIGURE 7

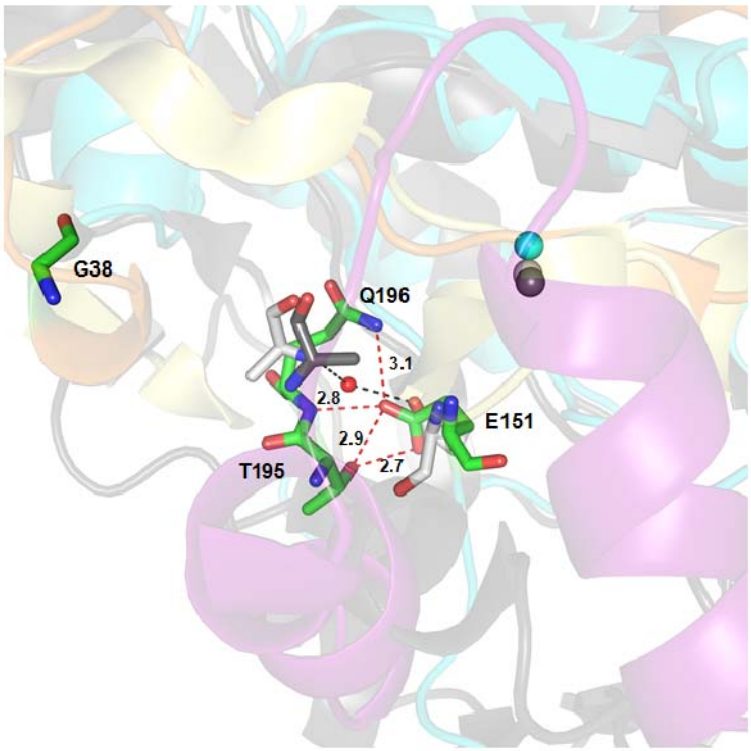
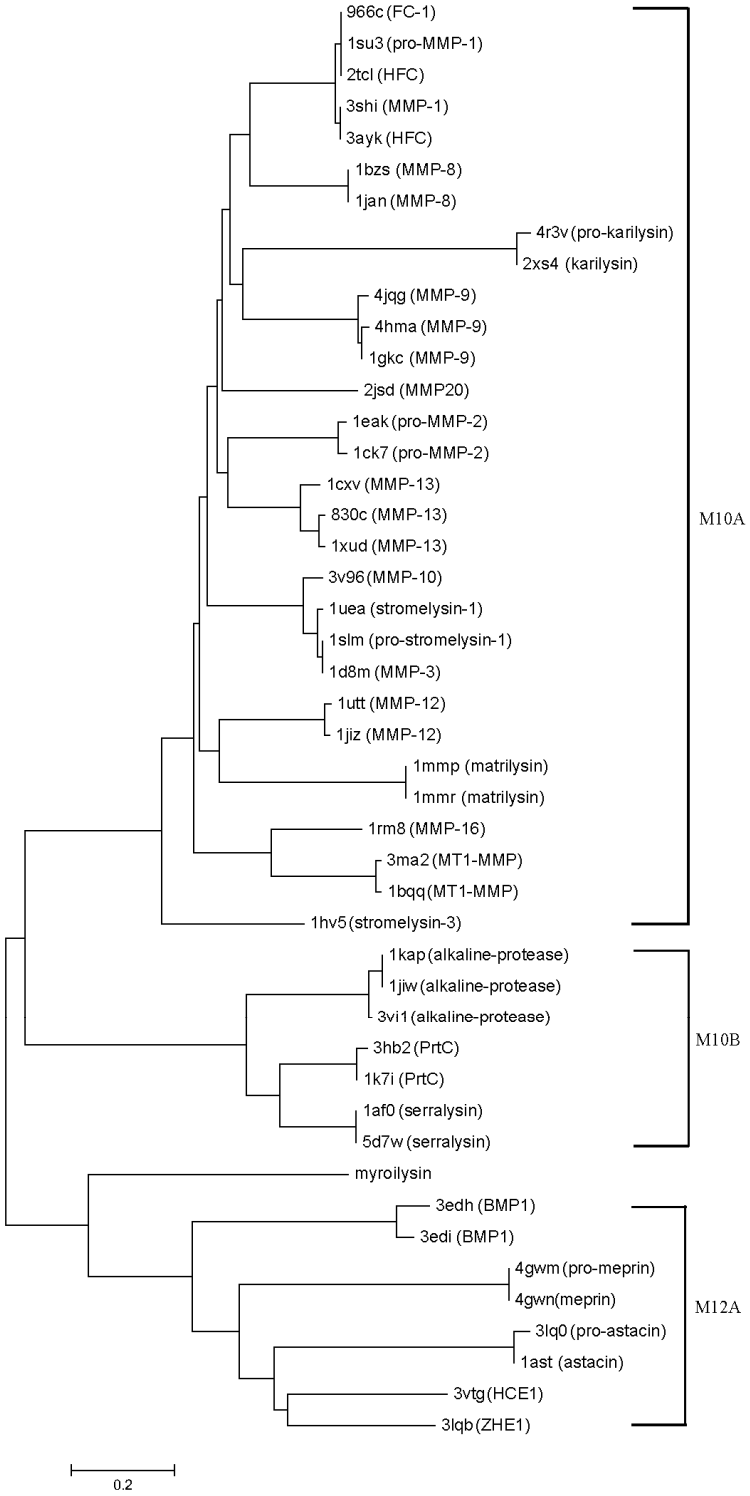


FIGURE 8



Myroilysin is a New Bacterial Member of the M12A Family of Metzincin Metallopeptidases and Activated by a Cysteine-switch Mechanism

Dongqing Xu, Jiale Zhou, Xiangdi Lou, Jianhua He, Tingting Ran and Weiwu Wang

J. Biol. Chem. published online February 9, 2017
