Crystal Structure of the Protealysin Precursor

INSIGHTS INTO PROPEPTIDE FUNCTION

Received for publication, May 29, 2009, and in revised form, November 11, 2009. Published, JBC Papers in Press, November 13, 2009, DOI 10.1074/jbc.M109.015396

Ilya V. Demidyuk†, Tania Yu. Gromova†, Konstantin M. Polyakov†, William R. Melik-Adamyan§, Inna P. Kuranova¶, and Sergey V. Kostrov‡

From the †Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq. 2, Moscow 123182, Russia, the ‡Kurchatov Center for Synchrotron Radiation and Nanotechnology, Kurchatov Sq. 1, Moscow 123182, Russia, and the ¶Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninsky Av. 59, Moscow 119333, Russia

Protealysin (PLN) belongs to the M4 family of peptidases that are commonly known as thermolysin-like proteases (TLPs). All TLPs are synthesized as precursors containing N-terminal propeptides. According to the primary structure of the N-terminal propeptides, the family is divided into two distinct groups. Representatives of the first group including thermolysin and all TLPs with known three-dimensional structures have long propeptides. A family of propeptide-containing proteases, the family is divided into two distinct groups. According to the primary structure of their propeptides, protealysin (PLN) belongs to the M4 family (1), are one of the best known groups of proteolytic enzymes. Enzymes of the second group, whose prototype is protealysin, have short (≈50 amino acids) propeptides. Here, we present the 1.8 Å crystal structure of PLN precursor (proPLN), which is the first three-dimensional structure of a TLP precursor. Whereas the structure of the catalytic domain of proPLN is similar overall to previously reported structures of mature TLPs, it has specific features, including the absence of calcium-binding sites, and different structures of the N-terminal region and substrate-binding site. PLN propeptide forms a separate domain in the precursor and likely acts as an inhibitor that blocks the substrate-binding site and fixes the “open” conformation of the active site, which is unfavorable for catalysis. Furthermore, the conserved PPL motif identified in our previous studies directly interacts with the S' subsites of the active center being a critical element of the propeptide-catalytic domain interface. Comparison of the primary structures of TLPs with short propeptides suggests that the specific features revealed in the proPLN crystal structure are typical for all protealysin-like enzymes. Thus, such proteins can be considered as a separate subfamily of TLPs.

Thermolysin-like proteases (TLPs)2, or peptidases of the M4 family (1), are one of the best known groups of proteolytic enzymes. TLPs are found in dozens of Gram-positive and negative bacterial species as well as in fungi and the archaeon Methanosarcina acetivorans (for reference see MEROPS database; 2). Many TLPs are produced by pathogenic organisms and thus can serve as potential drug targets; in addition, they have a considerable innovative potential for biotechnological application (see Ref. 3 for review). The family prototype thermolysin (MEROPS database identifier M04.001) was one of the first metallopeptidases to be sequenced (4) and one of the first to have its tertiary structure determined (5). To date, dozens of crystal structures of thermolysin with various ligands as well as crystal structures of TLPs from Bacillus cereus (6, 7), Pseudomonas aeruginosa (8), and Staphylococcus aureus (9) have been reported. Similar to many other proteolytic enzymes, TLPs are synthesized as precursors carrying N-terminal propeptides also called propeptides or prosequences. No three-dimensional structures of TLP precursors are currently available.

According to the primary structure of their propeptides, peptidases of the M4 family are divided into two distinct groups, although the mature enzymes have largely similar sequences (10). Propeptides of the first group of proteins are similar to that of thermolysin (TLN) and are ≈200 amino acids long (long propeptides). Such propeptides are found in all enzymes with determined three-dimensional structures of the mature portion. Thermolysin type propeptides are known to be crucial for enzyme functioning. They act as intramolecular chaperones (11–14), can inhibit cognate mature proteins (14–17), and influence their secretion (12, 18). Proteins of the second group resemble proteolysin from Serratia proteamaculans (MEROPS ID M04.025) (19) by the structure of their ≈50-amino acid propeptide. This group includes about one-fourth of TLPs; however, the functions of protealysin-like propeptides remain unclear.

The situation with TLP propeptides is intriguing. Their structures substantially differ at the background of highly similar catalytic domains. Moreover, the same species can produce enzymes with both propeptide types (10). This may indicate that the propeptides rather than the catalytic domains determine the biological function of M4 peptidases. If so, TLP propeptides can be considered as original evolutionary modules adapting the catalytic domains of proteases to cell requirements. However, the available structural and biochemical data are insufficient to understand how this system functions.

In this paper, we report the crystal structure of protealysin precursor, which is the first crystal structure of a precursor of thermolysin-like proteases as well as the first crystal structure of an enzyme from protealysin subfamily of M4 peptidases. The
Crystal Structure of Protealysin Precursor

data obtained give us an insight to the functioning of TLP propeptides and the underlying mechanism as well as demonstrate the difference between the catalytic domains of the two groups of the family.

EXPERIMENTAL PROCEDURES

Mutant Protein Construction, Expression, and Purification—ProPLN with a substitution of the key active site residue Glu-163 to Ala (numbering starts from the initial Met of PLN; Glu-163 corresponds to Glu-143 in mature TLN) and carrying a His6 tag at the C terminus was constructed, overproduced in Escherichia coli, and purified as described previously (20, 21). After purification, proPLN solution was dialyzed against 10 mM ammonium acetate and lyophilized.

The N-terminal amino acid sequence was determined after proPLN transfer onto a polyvinylidene difluoride membrane (Amersham Biosciences) using Edman auto-degradation on a high resolution amino acid analyzer (Applied Biosystems). The mass spectrometric analysis of proPLN was performed on a high resolution electrospray ionization orthogonal injection time-of-flight mass spectrometer (22).

Crystallization, Data Collection, and Structure Determination—The mutant proPLN was crystallized at room temperature as described previously (20) by the hanging-drop vapor diffusion method against a precipitant solution of 15% polyethylene glycol monomethyl ether 2000, 0.2M ammonium sulfate, and 0.1M sodium acetate, pH 4.6. Droplets contained 3 μl of 10 mg/ml protein solution and 1.5 μl precipitant solution with 0.06% octyl β-D-glucopyranoside in the same buffer. Crystals grew to maximum dimensions of 0.5 × 0.3 × 0.05 mm. Before data collection, the crystals were cryoprotected by rapid soaking in a solution containing mother liquor supplemented with 10% (v/v) glycerol and flash cooled with a nitrogen cryostream. X-ray diffraction data were collected on the K4.4 beamline of the Kurchatov Center for Synchrotron Radiation and Nanotechnology (Moscow, Russia) using a MAR CCD detector (23). For structure determination, a diffraction data set to 1.8 Å resolution was collected at a wavelength of 0.99 Å at 100 K. The diffraction images were processed and scaled with the XDS package (24). The proPLN crystals belonged to the space group P21212, with cell dimensions of a = 70.76, b = 78.65, and c = 59.28 Å and α, β, γ = 90°. There was one protein molecule per asymmetric unit. Structure determination was carried out with the BALBES (25) and MOLREP (26) programs using the structure of neutral protease from B. cereus (Protein Data Bank (PDB) accession code 1NPC) as the starting model. The starting model was modified and refined by using the CCP4 program suite for protein crystallography (27) and the REFMAC program (28) and manually processed with the COOT graphics program (29). An electron density map corresponding to the proPLN structure was calculated after refinement by REFMAC using coefficients (2mFo – DFo) (30). An omit map was calculated after refinement of the model lacking the propeptide residues using coefficients (mFo – DFo) (30). The final structure contained amino acid residues 6–37 and 52–341, one Zn2+ ion, and 277 water molecules yielding a crystallographic R factor of 17.1% and a free R factor of 20.2%. Seven residues (Cys-29, Glu-67, Asp-102, Met-229, Asn-260, Asp-305, and Ser-325) were modeled with alternate conformations. The details of data collection, processing, and structure refinement are summarized in Table 1. The coordinates and structure factors have

---

**TABLE 1** Data collection and structural refinement statistics

| Data collection statistics | Space group          | P2₁2₁2 (No. 18) |
|----------------------------|----------------------|-----------------|
| Cell parameters            | a = 70.76, b = 78.65, c = 59.28 Å; |
| Resolution range (Å)      | 52.6–1.8 (2.0–1.8) |
| No. unique reflections    | 28551 (6565)         |
| Completeness (%)           | 90.9 (78.6)          |
| B factor from Wilson plot (Å²) | 22.9 |
| Average I/σ(I)             | 26.92 (11.07)        |
| Multiplicity               | 2.82 (2.66)          |
| Rmax (%)                   | 3.2 (10.7)           |
| Rmerge (%)                 | 4.2 (12.5)           |

| Model refinement statistics |
|----------------------------|
| Resolution (Å)             | 52.6–1.82             |
| No. of reflections         | 27076                 |
| Rcryst (%)                 | 17.1                  |
| Rmerge (%)                 | 20.2                  |
| No. non-hydrogen atoms     | 2805                  |
| No. water molecules        | 277                   |
| B factors (Å²)             | 15.07                 |
| Average                    | 13.62                 |
| Main chain                 | 14.80                 |
| Side chains                | 23.07                 |
| Estimated coordinate error (Å) | 0.14               |
| RMS¹ deviations from ideality | Bond lengths (Å)    | 0.01             |
|                           | Bond angles           | 1.04             |
| Model geometry             | Rotamer outliers      | 0.78             |
|                           | Ramachandran outliers | 0.00             |
|                           | Ramachandran favored  | 97.2             |

a Values correspond to the highest resolution shell.

b For definition of Rcryst and Rmerge see Ref. 51.

c Rcryst = |Fobs| – |Fcalt|/|Fobs| culculated.

d Rmerge is as for Rcryst but calculated for a test set comprising 5% reflections not used in the refinement.

e Based on maximum likelihood.

f RMS, root mean square.

g Analyze was performed using MOLPROBITY (52).

---

**FIGURE 1. Overview of the protealysin precursor structure.** The catalytic domain and propeptide are shown in brown and blue, respectively. The 38–51 region of protealysin precursor is indicated by a gray dashed line. The catalytic zinc ion is represented by a green sphere.
FIGURE 2. Comparison of three-dimensional structures of protealysin (PLN, blue) and thermolysin (PDB code 8TLN; TLN, red). A, stereoview of superposition of Cα backbones of PLN and TLN (root mean square = 1.06 Å), three-dimensional structure-based sequence alignment. Secondary structures of PLN and TLN are presented over and under sequences, respectively. Numbering corresponds to proPLN (numbers in blue); for reference, numbering corresponding to mature TLN is given in parentheses (numbers in red). PLN versus TLN insertions are shown in yellow; TLN versus PLN insertions are shown in green. Residues that are Ca²⁺ ligands in TLN are marked with numbers: 1, Ca1/2; 3, Ca3; and 4, Ca4. Ligands whose side chains interact with Ca²⁺ are marked by white numbers on a black background. Z indicates zinc ion ligands; an asterisk indicates the Glu-163→Ala modification site. The residues undetectable in the electron density maps are shown in black. Amino acids identical in PLN and TLN are shaded. The PPL motif is overlined. The propeptide residues interacting with the substrate-binding site are designated as P1, P2, P1', P2', P3', P4', P5', and P6' according to Schechter and Berger (40).
been deposited in the Protein Data Bank under accession code 2VQX.

Structure and Sequence Analysis—Multiple sequence alignments were constructed by ClustalX 1.8 using the Gonnet series of protein weight matrices (31). The set of TLP sequences described previously (10) was used for alignment. The sequence logo (32) was generated using the WebLogo tool (33). Structure superimpositions were performed using DeepView (34) or PyMOL (35). All three-dimensional structure figures were generated using PyMOL.

RESULTS AND DISCUSSION

Overall Structure—Similar to all M4 peptidases, protealysin (PLN) is synthesized as a precursor with an N-terminal propeptide. The native protealysin precursor (proPLN) is unstable and autocatalytically processed in solution to mature active protein (21). The introduced Glu-163→Ala mutation (hereafter, numbering starts from the initial Met of PLN) prevented proPLN maturation, which allowed the isolation of preparative quantities of the full-length precursor. The protein was crystallized, and its three-dimensional structure was determined using x-ray crystallography, which yielded the first crystal structure of an M4 peptidase precursor.

The propeptide forms a separate domain in the proPLN molecule (Fig. 1) and interacts with the catalytic portion in the active center region. The N-terminal region of the propeptide occupies the substrate-binding site of the enzyme. It should be noted that we failed to identify some parts of the proPLN molecule in the electron density maps: 2–5, 38–51, and 342–347. At the same time, the molecular mass determined by mass spectrometry (38,103 Da) and the N-terminal sequence (PTLTAR) of the purified protein used in crystallization correspond to the entire precursor without the start methionine, with the Glu-163 to Ala substitution in the active site, and with a His6 tag at the C terminus. These findings sug-
We failed to detect calcium ions in the PLN structure, although all M4 peptidases with known three-dimensional structures bind one to four ions of this metal (5–9). Note that the significance of calcium binding for protein stabilization was demonstrated in all cases studied; accordingly, the thermostability increased with Ca\(^{2+}\) concentration (36–39). Analysis of the PLN structure in the regions corresponding to calcium binding sites in TLN and other TLPs demonstrated their notable difference in PLN.

The most surprising differences were observed in the region corresponding to the main double calcium-binding site in TLN (Ca1/2) (Fig. 3, A and B). Although most calcium ligands exist in PLN (Figs. 2B and 3B), the Glu-177\(_{\text{TLN}}\) to Lys-197 substitution (the subscript name indicates the protein whose residue is concerned except PLN whose name is omitted; TLN numbering corresponds to the mature protein) introduces an ionic interaction and additional hydrogen bonds in the PLN molecule (Fig. 3B). These interactions, as though compensating for missing coordination bonds with calcium ions, directly involve (apart from Lys-197) Asp-209 and Glu-214, analogs of the main Ca\(^{2+}\) ligands in TLN. Note that Asp-158 (an analog of another calcium ligand interacting by a side chain) is not involved in the above-mentioned interaction. The carboxyl group of Asp-158 forms hydrogen bonds with the amide group of Gly-213 and water molecule HOH-2113. The 204TADK207 and 284YAW286 insertions are other significant differences between this region in PLN and Ca1/2 in TLN. These insertions give rise to another set of interactions stabilizing the structure of this region, primarily, the Gly-285 O–Ala-205 N hydrogen bond and hydrophobic contacts of Tyr-284 and Trp-286 side chains with the aromatic rings of Phe-199 and Phe-103. Amino acid sequence comparison of the enzymes with short propeptides (data not shown) demonstrates high conservation of Lys-197. This finding suggests that the inability to bind calcium interacting by a side chain is not involved in the above-mentioned interaction. The carboxyl group of Asp-158 forms hydrogen bonds with the amide group of Gly-213 and water molecule HOH-2113. The 204TADK207 and 284YAW286 insertions are other significant differences between this region in PLN and Ca1/2 in TLN. These insertions give rise to another set of interactions stabilizing the structure of this region, primarily, the Gly-285 O–Ala-205 N hydrogen bond and hydrophobic contacts of Tyr-284 and Trp-286 side chains with the aromatic rings of Phe-199 and Phe-103. Amino acid sequence comparison of the enzymes with short propeptides (data not shown) demonstrates high conservation of Lys-197. This finding suggests that the inability to bind calcium
in the region corresponding to Ca1/2 is typical for all protealysin-like TLPs.

The organization of the N-terminal region of mature PLN differs from that of other TLPs with known crystal structures (Fig. 3C). This is largely due to a ∼30-amino acid shorter N terminus of mature PLN and other M4 enzymes with short propeptides compared with that in TLPs with long propeptides as demonstrated previously by amino acid sequence analysis of TLP precursors (10). The absence of the long sequence underlies a significantly different structure in the region corresponding to calcium-binding site Ca3 of TLN. In addition, PLN residues Asn-79 and Gly-81 correspond to the main ligands interacting with Ca2+/H11001 by the side chain (Asp-57TLN and Asp-59TLN). Considering that the short N terminus is typical of all protealysin-like enzymes, the absence of the Ca3 site is apparently their common property.

The last calcium-binding site (Ca4) is also significantly different in PLN (Fig. 3, D and E). First, the PLN molecule has Lys-224 and Ala-218 instead of Asp-200TLN and Thr-194TLN, respectively. Thus, PLN lacks the key ligands of the site, the carboxyl and hydroxyl groups directly interacting with Ca2+/H11001. In addition, PLN has a different geometry of the loop forming Ca4 in TLN (Fig. 3D), which shifts other calcium ligands: carboxyl oxygens of residues 193TLN (217), 194TLN (218), and 197TLN (221). Such loop conformation in PLN is stabilized by four hydrogen bonds (Lys-224 N–Ala-235 O, Asn-222 O–Asp-237 N, Asn-222 N–Asp-237 O, and Asn-222 N–Asp-238 Oδ1) with the residues in the structure conserved in protealysin-like proteases (residues 235–244) but missing in TLN. We designated this structure as AYDD-hairpin (Fig. 3E) and discussed it below.

This is the first demonstration that an M4 peptidase molecule can lack calcium-binding sites. Note, however, that although proPLN was purified in the presence of calcium ions, no additional calcium salts were added to the precipitant solution during crystallization. Hence, low affinity calcium-binding sites in the PLN molecule cannot be excluded. Their existence is indirectly confirmed by our previous data on decreased PLN thermostability at high Ca2+/H11001 concentrations (19), which can result from altered structure of the protein after calcium binding by such a low affinity site.

The structure of the PLN catalytic domain differs from known crystal structures of other TLPs not only by the absence of calcium ions. The most significant differences are observed in the active site. All previous three-dimensional structures of M4 peptidases belong to the enzymes with long propeptides. Active sites of these proteins are localized in the cleft between the N- and C-terminal domains (Fig. 4A). In PLN, the active site cleft is dammed on the side of the S/H11032 subsites (as designated by Schechter and Berger (40)) (Fig. 4B), which provides for extra contacts between the N- and C-terminal domains. The “dam” of the cleft is formed by the above-mentioned AYDD-hairpin (235–244) (Fig. 4C). Our analysis of a representative set of TLP sequences demonstrated that this structural element is absent from known M4 peptidases with long propeptides but present in all enzymes with short propeptides, and its sequence is highly conserved (Fig. 5). Thus, the AYDD-hairpin is a typical feature of protealysin-like enzymes like the short propeptide. Moreover, the propeptide and AYDD-hairpin directly interact, which allows us to consider the AYDD-hairpin as a structural factor providing for the complementarity between the mature portion and propeptide. (The details of the propeptide interaction with the catalytic domain and AYDD-hairpin, in particular, are described below.) In addition, the presence of the AYDD-hairpin changing the active site geometry in the S′ subsites region may be responsible for the narrowed substrate specificity of PLN; we have recently demon-
strated that, unlike TLN, PLN is highly specific toward actin (41) similar to the homologous protease ECP32/grimelysin (42, 43).

Of note, two crystal forms of TLN and \textit{P. aeruginosa} elastase (PAE) can be produced, depending on whether they bind a ligand or not (8, 44, 45). The enzyme-ligand complexes demonstrate conformations with the “closed” substrate-binding cleft, although free enzymes have more “open” conformations. These data suggested a hinge-bending motion that changes the relative position of the C- and N-terminal domains during the catalytic cycle of TLPs (46). PLN precursor, which can be considered as a mature enzyme in a complex with ligand, has a more open conformation of the substrate-binding cleft than the open form of TLN (Fig. 6) and is nearly as open as the open form of PAE (data not shown). Apparently, this results from the catalytic domain interaction with the propeptide (see below). However, it cannot be excluded that the AYDD-hairpin per se prevents approaching of the domains. In this case, mature PLN is incapable of hinge-bending motion during its catalytic cycle and always has an open conformation. Still, this situation is not very probable considering that the open conformation is catalytically unfavorable.

Thus, the PLN catalytic domain has specific properties; it contains no calcium ions and demonstrates substantial differences in the active center region. Our analysis suggests that these changes are not unique to PLN but rather are shared by all TLPs with short propeptides. This allows protealysin-like enzymes to be recognized as an individual subfamily of M4 peptidases based both on the primary structure of their precursors (10) and on the three-dimensional structure of their catalytic domains.

\textbf{Propeptide Structure and Interaction}—As mentioned above, PLN propeptide is an individual domain (Fig. 1) connected to the catalytic domain by a flexible spacer (residues 38–51). Two structural elements can be recognized in the detectable portion of the propeptide (residues 6–37): a hairpin including two \(\alpha\)-helices, \(\alpha1\) and \(\alpha2\) (residues 11–22 and 23–36, respectively), and the N-terminal segment (residues 6–10). The propeptide interacts with the catalytic domain in the active site region (Fig. 7, A and B). The \(\alpha2\) helix contacts the surface of the N-terminal domain of the mature portion; the C-terminal region of \(\alpha1\) contacts the AYDD-hairpin (which can be considered as an element of the C-terminal domain); whereas the N-terminal portion of \(\alpha1\) and the N-terminal segment go down into the socket of the active site.

Helices \(\alpha1\) and \(\alpha2\) contact each other by hydrophobic side chains of Leu-15 and Ile-19 on the \(\alpha1\) side and Arg-27 (aliphatic portion: C\(\beta\), C\(\gamma\), and C\(\delta\)), Ala-30, Leu-31, and Leu-34 on the \(\alpha2\) side. The \(\alpha1\) and \(\alpha2\) helices interact with the catalytic domain in different ways. Apparently, hydrogen bonds Gln-26 Ne2—Glu-148 Oe1, Thr-33 O\(\gamma\)1—Gln-131 N, and Cys-29 S—Glu-129 Oe1 are critical in the case of \(\alpha2\). The interactions of \(\alpha1\) are more complex and diverse. The C-terminal portion of \(\alpha1\) realizes a hydrophobic interaction between the side chain of Ile-18 and side chains of Ile-149, Phe-150, Leu-240, and Leu-241 as well as between the aliphatic portion of Arg-17 side chain and Leu-241. Significantly, residues 149 and 150 belong to the N-terminal domain of the PLN mature portion, whereas residues 240 and 241 belong to the AYDD-hairpin, which is a part of the C-terminal domain. Thus, the interaction with the propeptide fixes the N- and C-terminal domains of the proPLN mature portion in a position corresponding to the open conformation of the substrate-binding cleft.

Previously, we have identified a conserved element, the PPL motif, in the short propeptide of TLPs (10). This element is a hydrophobic cluster of seven amino acids, three of which are...
In the propeptide-catalytic domain complex, the Ser-8–Val-9 arrangement provides a hydrogen bond distance between the Ser-8 carbonyl group and the water molecule bound by the catalytic zinc (Fig. 8A), i.e. the Ser-8–Val-9 bond corresponds to the bond hydrolyzed by the active enzyme. (Previously, we have shown that cleavage of this particular bond is the first stage of processing of protealysin precursor (21).) The propeptide residues interacting with the substrate-binding site can be designated as follows in terms of the Schechter and Berger nomenclature (40): Arg-7, P2; Ser-8, P1; Val-9, P1; Ile-10, P2'; Pro-11, P3'; Pro-12, P4'; Tyr-13, P5'; and Met-14, P6' (Figs. 2B and 8A). (Note that P1’-P6’ positions correspond to amino acids in the PPL motif.) Thus, in contrast to all other TLPs with known crystal structures, the S’ region in pro-PLN is substantially extended and formally includes six instead of two subsites (47). The extra subsites are formed due to the AYDD-hairpin conserved in protealysin-like enzymes.

The structure of PLN is the first crystal structure of TLP where the entire S region of the substrate-binding site is occupied by a polypeptide ligand. Previously, peptide interaction with S subsites except S1 were considered only theoretically (47). The data obtained indicate that Ala-6 does not interact with the catalytic domain of the enzyme but is exposed to solvent (Figs. 7A and 8A). Propeptide residues 2–5, which were undetectable (see above), are likely exposed to solvent too and do not bind to the catalytic domain. This can explain high flexibility of this region. Thus, only the P1 and P2 positions of the peptide ligand nonprimed side are significant for binding to the enzyme. The interaction of the PLN propeptide with the catalytic domain at S2 and S1 is largely analogous to ligand interaction with the substrate-binding site of TLN (47). Arg-7 (P2') and Ser-8 (P1') form an antiparallel β-sheet with the 133–135 region of the catalytic domain. The side chains of residues 7 and 8 point from the active site cleft to solvent (Figs. 7A and 8A), and their nature is likely not critical for the binding. The crystal structure of proPLN is the first structure of TLP containing an unmodified amino acid at P1 linked to the residue at P1’ with a ordinary peptide bond. Thus, it is the first demonstration of an analog of a substrate-enzyme interaction immediately before the catalytic act. As mentioned above, the carbonyl oxygen of Ser-8 is within hydrogen bond distance from the catalytic water molecule (HOH-2277), which is the fourth ligand of the zinc ion (Fig. 8A).

Invariant in all protealysin-like enzymes: two neighboring Pro and a Leu two residues downstream of them. An aromatic amino acid Tyr or His immediately follows the Pro-Pro site in most cases. The obtained crystal structure data demonstrate that this motif corresponds to the N-terminal portion of α1 (residues 11–15) and to residues 9 and 10 of the propeptide N-terminal segment (Figs. 2B and 7A). Six of seven amino acids in the PPL motif directly interact with the S’ substrate-binding subsites of the active site. The last amino acid, Leu-15, residing between α1, α2, and the N-terminal domain of the mature portion is likely one of the main bridging elements that stabilizes the propeptide-catalytic domain complex (Fig. 7C). Phe-150 in the mature portion and Leu-34 in α2 are the main partners of Leu-15.

The N-terminal region of the propeptide (residues 7–14) occupies the substrate-binding site of PLN (Fig. 8A). Hence, the interaction between the catalytic domain and this propeptide region can be considered analogous to the interaction between the enzyme and substrate (and the propeptide is essentially the substrate (21)). Note the peculiar structure of the propeptide: it has a specific sequence and is anchored by the α-helical hairpin. In this context, one cannot exclude that the peptide substrate binding to the active site of PLN in general can differ from the propeptide binding.

FIGURE 8. Substrate-binding site of protealysin. A, general stereo view. Omit map contoured at the 3.0σ level in the propeptide region is shown in light blue. B, environment of the catalytic water molecule. C, S1' and S2' subsites. In B and C, electron density map contoured at the 1.5σ level is shown in light gray. The catalytic domain is colored yellow orange. The propeptide fragments are colored slate. Propeptide residues (P1, P2') are designated after Schechter and Berger (40). The zinc ion is shown as a green sphere; oxygen of HOH-2277, red sphere. The distance between atoms is given in angstroms.
The S1′ site is the major determinant of substrate specificity in TLPs (49, 50). Similar to other M4 peptides, PLN has a hydrophobic pocket in the binding region of the side chains of the amino acid at position P1′, which, however, notably differs from the S1′ pocket of TLN (Fig. 9). Phe-153, Val-159, Leu-216, and Leu-226 (corresponding to Leu-133TLN, Val-139TLN, Val-192TLN, and Leu-202TLN, which form the S1′ pocket of TLN) contribute most to the region formation in PLN (Fig. 9A). At the same time, bulkier amino acids (Phe-153 versus Leu-133TLN and Leu-216 versus Val-192TLN) make the S1′ pocket of PLN shallower compared with that of TLN. Presumably, PLN has a higher hydrolytic preference toward bonds formed by α-amino groups of small hydrophobic residues such as Ala or Val than TLN and other TLPs with a TLN-like structure of the S1′ pocket. The presence of a pronounced S2′ pocket is an important feature of the proPLN structure (Fig. 9B); this pocket results from the shift of Phe-150 (Phe-130TLN) and “elongation” of the S1′ pocket relative to TLN (Fig. 9C). Thus, the efficiency of substrate binding should depend on the amino acid at the position of the P2′ substrate more for PLN than for TLN.

As mentioned above, the S′ region in TLN and other TLPs with previously known three-dimensional structures contains only two subsites (47), whereas the catalytic domain of proPLN contains additional S′ subsites interacting with residues 11–14 of the propeptide (Fig. 10A). Because this is the most conserved region in the short propeptides, its contacts are likely crucial for their functioning. Pro-11 and Pro-12 corresponding to ligand positions P3′ and P4′ are conserved in all short propeptides without exception and, likely, provide for a precise steric complementarity between the propeptide and the catalytic domain; a tandem of prolines lowers the polypeptide chain at the N terminus of the α helix. Well-defined binding pockets in the mature portion correspond to the side chains of Tyr-13 (P5′) and Met-14 (P6′) (Fig. 10A). The S5′ pocket is a large cavity, whose entrance is formed by hydrophobic side chains of Pro-245 and Leu-265 contacting the aromatic ring of Tyr-13.

8B). The phenol group of Tyr-177 (157TLN) and the Ne2 nitrogen atom of the imidazole ring of His-264 (231TLN), which are the residues stabilizing an intermediate state of the substrate similar to the “oxygen hole” of serine propeptases (48), are oriented toward HOH-2277. The Tyr-177 OH–HOH-2277 O and His-264 Ne2–HOH-2277 O distances in the structure of proPLN are greater than the length of a strong hydrogen bond. Still, such bonds cannot be excluded in a mature molecule bond to the substrate and with a closed conformation of the substrate-binding cleft. At the same time, the Tyr-177 OH–His-264 Ne2 bond is likely formed in the proPLN molecule.

The PLN propeptide interaction with the catalytic domain at the S1′ and S2′ regions is similar to the ligand interaction with the analogous sites in TLN (47) and PAE (44). The main chain of the propeptide at the Val-9 (P1′)–Ile-10 (P2′) region forms three hydrogen bonds with the Asn-132-Ala-133 region of the catalytic domain: Val-9 N–Ala-133 O, Ile-10 N–Asn-132 Oδ1, and Ile-10 O–Asn-132 Nε2 (Fig. 8C). At the same time, the Val-9 O–Arg-227 NH1 and Val-9 O–Arg-227 NH2 distances are much longer than the hydrogen bond length, whereas hydrogen bonds between the carbonyl residue at the P1′ position and the corresponding groups of Arg-203TLN (198PAE) are formed after ligand interaction with TLN and PAE. This situation similar to the above-mentioned interaction of Tyr-177 and His-264 with HOH-2277 can result from the open conformation of the proPLN substrate-binding cleft. We believe that hydrogen bonds between Val-9 and Arg-227 are formed during ligand binding by the mature enzyme.
Deep in the pocket, there is a cluster of three water molecules (HOH-2011, HOH-2180, and HOH-2205) linked to the residues of the catalytic domain (Leu-226, Arg-227, Tyr-236, Asp-244, and His-264) and to each other with a network of hydrogen bonds. The first four residues likely contribute most to the interaction with Met-14. It should be stressed that, because the conserved residues of the AYDD-hairpin are crucial for the S5' and S6' subsites formation (Fig. 5), the observed structure is most likely not unique to the discussed molecule but rather is shared by all protealysin-like proteins.

Analysis of all interactions between the prosequence and the catalytic domain suggests that propeptide of PLN is an intramolecular inhibitor. The inhibitory effect of the prosequence most likely involves blocking the S' substrate-binding region of the enzyme by the propeptide region corresponding to the PPL motif and the fixation of the catalytically unfavorable open conformation of the active site.

CONCLUSIONS

Analysis of the proPLN crystal structure, which is the first structure of M4 peptidase precursor, suggests that protealysin-like proteases can be considered as an individual subfamily of TLPs on the basis of both precursor and catalytic domain structure. Typical features of the catalytic domain of PLN and likely other M4 peptidases with short propeptides include the absence of calcium-binding sites, different N-terminal structure, and presence of the highly conserved AYDD-hairpin modifying the structure of the substrate-binding region.

The propeptide of PLN forms a separate domain of the protein and likely acts as an intramolecular inhibitor blocking the substrate-binding region and fixing the catalytically unfavorable open conformation of the enzyme active site. The key elements of the propeptide-mature portion interface in proPLN include the AYDD-hairpin in the catalytic domain and the highly conserved PPL motif in the propeptide.
