Protein Science (1999), *8*:2537–2540. Cambridge University Press. Printed in the USA.
Copyright © 1999 The Protein Society

FOR THE RECORD

Thermolysin and mitochondrial processing peptidase: How far structure–functional convergence goes

KIRA S. MAKAROVA[1,2,3] AND NICK V. GRISHIN[1]

[1]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894
[2]Department of Pathology, F.E. Hebert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799

(RECEIVED June 25, 1999; ACCEPTED August 5, 1999)

Abstract: The structure–functional convergence between two Zn-dependent proteases, namely thermolysin and mitochondrial processing peptidase (MPP), is described. These two families of nonhomologous enzymes show not only functional convergence of several active site residues as in chymotrypsin and subtilisin, but also structural convergence of overall molecular architectures including the β-sheet arrangement and packing of the surrounding α-helices. The major functionally important structural elements are present in both enzymes with different topological connections and often in reverse main-chain orientation, but display similar packing. The structural comparison helps to rationalize sequence "inversion" of the HEXXH thermolysin consensus present as HXXEH in MPP. The described structural convergence may be due to a limited number of alternatives to build a Zn-protease that utilizes hydrogen bonding between a substrate main chain and the enzyme β-sheet for substrate binding.

Keywords: core I protein; cytochrome bc_1 complex; metalloenzymes; mitochondrial processing peptidase; molecular evolution; thermolysin

The rapidly increasing set of available protein spatial structures (Bernstein et al., 1977) guides our understanding of evolutionary rules in protein structure space. Divergence explains similarity through inheritance from the common ancestor, and convergence means that some similar properties were acquired independently in evolution. Many fascinating examples of divergent evolution have been discovered (Gardiner et al., 1997; Murzin, 1998; Holm & Sander, 1999). However, the convergent evolution aspect remains relatively unexplored and limited to the studies on arrangement of several functionally important residues in analogous proteins (Russell, 1998). The example of chymotrypsin–subtilisin proteases that utilize the same Ser-His-Asp triad still remains the classical one (Voet & Voet, 1990; Barrett & Rawlings, 1995). Because the sequence order of the triad residues and overall folds of chymotrypsin (Tsukada & Blow, 1985) (two β-barrel domains) and subtilisin (Kuhn et al., 1998) (modified Rossman fold) are different, it appears that the same catalytic triad arrangement has evolved independently in these two proteins. The question arises if it is possible to find two proteins that share not only functional and local structural similarities as chymotrypsin and subtilisin do, but are similar in overall structure and yet are related by convergence. We believe that in the course of Zn-metalloprotease sequence–structure comparison study, we stumbled upon a striking example of such structure–functional convergence.

Thermolysin, a Zn-endopeptidase from *Bacillus thermoproteolyticus*, represents zincin-fold proteases (Murzin et al., 1995) and has been thoroughly studied in terms of its spatial structure (Tronrud et al., 1987, 1992) and enzymatic mechanism (Matthews, 1988; Holland et al., 1995). At the sequence level, the thermolysin family is characterized by the signature HEXXH, which incorporates two (out of the three) Zn^{2+} ligands and the catalytic glutamate. Another large family of Zn proteases, mitochondrial processing peptidases (MPP) (Braun & Schmitz, 1995b), is represented by the recently determined structures of core I and II subunits in the mitochondrial cytochrome bc_1 complex (Xia et al., 1997; Iwata et al., 1998; Zhang et al., 1998). In contrast to thermolysin, not much is known about the MPP catalytic mechanism and specificity despite the significant effort (Braun & Schmitz, 1995b; Luciano & Geli, 1996; Tanudji et al., 1999). The sequences of catalytic β-subunit of MPP share the pattern HXXEH, which is reminiscent of that in the thermolysin Zn-binding site, but is inverted (Rawlings & Barrett, 1995). In the N-terminal domain of mammalian core I protein (homolog of MPP β-subunit), the first histidine is replaced by a tyrosine (Braun & Schmitz, 1995a). However, plant core I proteins along with MPPs contain a complete His/His/Glu Zn-binding site and display full proteolytic activity (Braun & Schmitz, 1995a).

Reprint requests to: Nick V. Grishin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894; e-mail: grishin@ncbi.nlm.nih.gov.
[3]Permanent address: Institute of Cytology and Genetics, Russian Academy of Sciences, Novosibirsk 630090, Russia.

To compare spatial structures of thermolysin (Protein Data Bank (PDB) entry 4tmn) and the N-terminal domain of the core I in a cytochrome bc_1 complex (PDB entry 1bgy, chain A), we superimposed C_α atoms of the three Zn ligands (His142, His146, and Glu166 in thermolysin with His61, Tyr57, and Glu137 in core I) and a presumably catalytic glutamate residue (Glu143 and Glu60 in thermolysin and core I, respectively). The superposition results in root-mean-square deviation (RMSD) of 1.56 Å for these four C_α atoms and leads to a striking and unexpected match of many secondary structural elements of the two proteins (Fig. 1A,B). Indeed, the two helices carrying Zn ligands (Fig. 1, *E* and *F*) and the other two helices (*B* and *G*) of thermolysin superimpose with their counterparts in the core I protein in reverse N-to-C orientation explaining the "inversion" of the family signature. Five strands of a single β-sheet (*a* to *e*) including the β-hairpin *de* are structurally equivalent in both enzymes. These superimposable elements comprise the majority of the domain structure. Thus, in addition to similarity in the arrangement of Zn-binding and catalytic residues, the two proteases have similar overall structural architecture. This architecture is based on a mixed β-sheet with a loose bundle of four α-helices packed on one side of it.

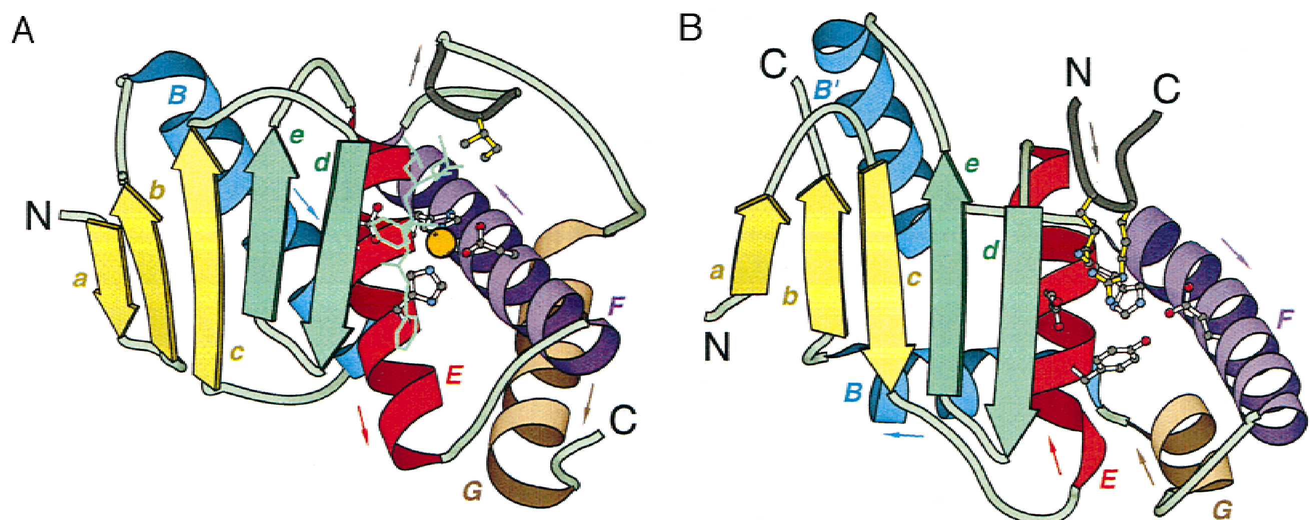
The overall structural similarity between thermolysin and the N-terminal domain of core I of bc_1 complex results in RMSD of 3.4 Å for 85 C_α atoms (Fig. 1C) if connectivity and main-chain N-to-C orientation are ignored. The resemblance between thermolysin and core I protein remained unnoticed, because it is not detected by widely used algorithms such as DALI (Holm & Sander, 1995) and VAST (Gibrat et al., 1996), which are aimed on detection of similar folds where connectivity and orientation are preserved. However, if we leave only C_α atoms in the structure of the N-terminal domain of core I protein (residues A21–A201, PDB entry 1bgy), remove loops, and change residue numbers in β-strands and α-helices to match the connectivity and orientation observed in zincins, DALI (Z-score of 2.4, RMSD 3.5 Å for 93 C_α atoms) and VAST (p-value of 0.0012) find thermolysin structure as the first hit with this set of C_α coordinates. Therefore, if the connectivity and main-chain orientation of the equivalent secondary structural elements were the same in zincins and MPPs, one would inevitably conclude that these groups of proteases shared a common ancestor. Notably, when thermolysin is used as a reference, not a single pair of structurally equivalent elements is consecutive in core I protein, and 7 out of 10 elements are present in a reversed-chain orientation (Fig. 1D). The N-terminal domain of core I protein possesses a modified ferredoxin fold (strands *b, c, d,* and *e* and helices *B′* and *E* form a ferredoxin unit), while the zincin fold of thermolysin is similar to the Rossman fold [see the more Rossman-like structure of a thermolysin homolog adamalysin (Bode et al., 1994), PDB entry 2aig]. Taking into account the differences in topology and main-chain orientation, we confidently conclude that the similarity between zincins and MPPs is convergent.

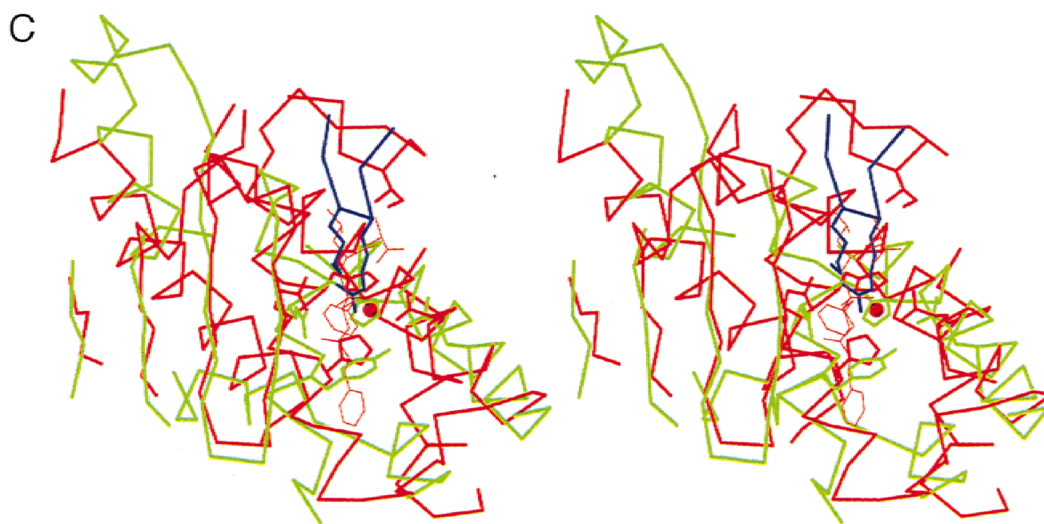
It is likely that general principles of helix-to-sheet and helix-to-helix packing contribute to the convergent similarity between zincins and MPPs, which reflects the limited number of acceptable spatial arrangements of secondary structural elements. However, functional requirements for being a Zn protease should be important as well. Indeed, the stereochemical restrains on relative spatial arrangement of Zn-binding residues, possibly catalytic glutamate, and β-hairpin *de* (Fig. 1A,B) used for substrate binding by means of main-chain hydrogen bonds limits the number of possible protein architectures. For instance, another peptide hydrolase, namely peptide deformylase (PDF), was shown to share limited structural and functional similarity with thermolysin (Meinnel et al., 1996; Chan et al., 1997; Becker et al., 1998; Dardel et al., 1998; Hao et al., 1999; Ragusa et al., 1999). PDF removes the formyl group from the N-terminus of the newly synthesized bacterial proteins. PDF sequence consensus HEXXH is indicative of a metallohydrolase. The spatial structure of PDF (Meinnel et al., 1996; Chan et al., 1997) revealed its OB-like core and demonstrated that similarly to zincins and MPP, HEXXH is a part of an α-helix with two histidines coordinating a metal. The third metal ligand (Cys90) comes from a long twisted β-hairpin spatially equivalent to the S1′-loop in zincins and MPPs, and not from an α-helix. The PDF also contains a β-hairpin structurally equivalent to the *de* β-hairpin in zincins and MPPs, but the β-hairpin with Cys90 participates in main-chain hydrogen bonding of substrate (Becker et al., 1998; Hao et al., 1999). Thus, the similarity in overall architecture and arrangement of functional groups is more pronounced between thermolysin and MPP than between PDF and either thermolysin or MPP.

Traditionally, homology is used to make functional predictions. However, the convergent similarity between thermolysin and MPP is so extensive that it allows us to formulate predictions about the substrate binding in the latter based on the former. To facilitate substrate binding, thermolysin uses the natural ability of a peptide to form main-chain hydrogen bonds with an edge of a β-sheet (Matthews, 1988). The *ed* β-hairpin carries out this function in thermolysin and probably has the same function in MPPs. The substrate chain forms hydrogen bonds with the β-strand *d* in an antiparallel way (Fig. 1A,B). The S1′ specificity pocket is large in thermolysin, which is specific toward bulky hydrophobic amino

Fig. 1. (*facing page*). Thermolysin and core I protein: an example of structure–functional convergence. Ribbon diagrams of (**A**) thermolysin (PDB entry 4tmn) and (**B**) core I protein (PDB entry 1bgy) drawn by BOBSCRIPT (Esnouf, 1997), a modified version of MOLSCRIPT (Kraulis, 1991). The structures were superimposed and then separated for clarity. The spatially equivalent structural elements are colored correspondingly in the two structures. The S1′ pocket loop in core I protein (colored dark gray) is contributed by the core II protein. N- and C-termini are labeled. Secondary structural elements are labeled consecutively in the thermolysin structure, with the same letters used for spatially equivalent elements in the core I protein. The orientations of α-helices and a loop in the S1′ pocket are shown by arrows. The Zn ligands, catalytic glutamate, and residues near the S1′ specificity pocket are shown in a ball-and-stick representation. The Zn ion is shown as an orange ball. The substrate analog inhibitor (Holden et al., 1987) (Cbz-Phep-Leu-Ala) in the thermolysin structure is shown in green lines. (**C**) The stereoview of the superposition of thermolysin (red) and core I protein (green) with the S1′ pocket loop protruding from the core II protein (blue). Side chains of Zn ligands and catalytic Glu are shown in both structures, a Zn ion and an inhibitor are shown in thermolysin as a ball and thin lines, respectively. (**D**) Structure-based sequence alignment of superimposable elements in thermolysin and core I protein. The starting and ending residue in each segment is numbered, and the segments are labeled with the same letters as in **A** and **B**. The sequences in the opposite chain orientation are shown in reverse letters. Zn ligands and a catalytic Glu residue are shown in white letters boxed with black, and the residues in the loop of S1′ pocket are in white letters boxed gray. Invariant residues are shown in bold letters, conserved hydrophobic, small, and hydrophilic residues are on yellow, green, and blue background, respectively.

acids. The pocket is capped by Ile188 and Gly189 supplied from the loop. In core I protein this loop is superimposed with the loop protruding from the core II subunit C-terminal domain, which contains Lys286 and Arg287 structurally equivalent to Ile188 and Gly189 (Fig. 1A–C). Due to the low resolution structure of bc_1 complex (Iwata et al., 1998) (3.0 Å), it is hard to deduce the exact placement of side chains, but it is clear that the S1′ pocket in MPPs should be much smaller that the one in thermolysin. The structural properties are consistent with the MPP preference for small or negatively charged residues at P1′ substrate site (Saavedra-Alanis et al., 1994; Waltner & Weiner, 1995; Tanudji et al., 1999). These hypotheses await experimental verification.





D

```
                 a               b                    B                      B
4tmn E     29-YLQD-32    40-FTYDAKY-46      64-ASYDAPAVDAH-74     75-YYAGVTYDYYKNVH-88
1bgy A     25-RLGN-21   194-RMVLAAA-200    107-PKAVELLADIV-117   191-KYHRSLYETLDARS-178
                                                  B'                     B

               c                  d      e                                 E
4tmn E     99-AIRSSVHYSQ-108 109-GYNNAFWNGSEMVYGDGDG-127 136-GIDVVAHELTHAVTDYTAG-154
1bgy A     44-GADIWVGVTC-35    84-AHLNAYSTREHTAYYIKAL-101   67-TGKFALHEVFYGAGNMKES-49

                            F                                            G
4tmn E   156-IYQNESGAINEAISDIFGTLVEFYA-180    186-WEIGED-191     268-YRALTQYLTP-277
1bgy AB  147-DRMSTDNEQLEQLIVDREKEIQSDE-123  B289-SGRKVH-B284     177-LKRVNE-SGP-169
```

We believe that thermolysin and mitochondrial peptidase give a remarkable example of convergent protein evolution. Indeed, it illustrates not only functional convergence of the active site residues as in the case of chymotrypsin and subtilisin, but, most importantly, structural convergence of overall molecular architectures including the β-sheet arrangement and packing of the surrounding α-helices. This convergence is a probable reflection of a limited number of alternatives to build a Zn protease that utilizes hydrogen bonding of a substrate chain to the enzyme β-sheet for substrate binding.

Acknowledgments: We thank Eugene V. Koonin for critical reading of the manuscript, Yuri I. Wolf for stimulating discussions, and an anonymous reviewer for helpful suggestions. Kira S. Makarova is supported by DOE OBER Grant DE-FG02-98ER62583.

References

Barrett AJ, Rawlings ND. 1995. Families and clans of serine peptidases. *Arch Biochem Biophys 318*:247–250.

Becker A, Schlichting I, Kabsch W, Groche D, Schultz S, Wagner AF. 1998. Iron center, substrate recognition and mechanism of peptide deformylase. *Nat Struct Biol 5*:1053–1058.

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem 80*:319–324.

Bode W, Kress LF, Meyer EF, Gomis-Ruth FX. 1994. The crystal structure of adamalysin II, a zinc-endopeptidase from the snake venom of the eastern diamondback rattlesnake *Crotalus adamanteus*. *Braz J Med Biol Res 27*: 2049–2068.

Braun HP, Schmitz UK. 1995a. Are the "core" proteins of the mitochondrial bc1 complex evolutionary relics of a processing protease? *Trends Biochem Sci 20*:171–175.

Braun HP, Schmitz UK. 1995b. Cytochrome-*c* reductase/processing peptidase complex from potato mitochondria. *Methods Enzymol 260*:70–82.

Chan MK, Gong W, Rajagopalan PT, Hao B, Tsai CM, Pei D. 1997. Crystal structure of the *Escherichia coli* peptide deformylase. *Biochemistry 36*:13904–13909.

Dardel F, Ragusa S, Lazennec C, Blanquet S, Meinnel T. 1998. Solution structure of nickel-peptide deformylase. *J Mol Biol 280*:501–513.

Esnouf R. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model 15*:133–138.

Gardiner EJ, Artymiuk PJ, Willett P. 1997. Clique-detection algorithms for matching three-dimensional molecular structures. *J Mol Graph Model 15*: 245–253.

Gibrat JF, Madej T, Bryant SH. 1996. Surprising similarities in structure comparison. *Curr Opin Struct Biol 6*:377–385.

Hao B, Gong W, Rajagopalan PT, Zhou Y, Pei D, Chan MK. 1999. Structural basis for the design of antibiotics targeting peptide deformylase. *Biochemistry 38*:4712–4719.

Holden HM, Tronrud DE, Monzingo AF, Weaver LH, Matthews BW. 1987. Slow- and fast-binding inhibitors of thermolysin display different modes of binding: Crystallographic analysis of extended phosphonamidate transition-state analogues. *Biochemistry 26*:8542–8553.

Holland DR, Hausrath AC, Juers D, Matthews BW. 1995. Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Sci 4*:1955–1965.

Holm L, Sander C. 1995. Dali: A network tool for protein structure comparison. *Trends Biochem Sci 20*:478–480.

Holm L, Sander C. 1999. Protein folds and families: Sequence and structure alignments. *Nucl Acids Res 27*:244–247.

Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK. 1998. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. *Science 281*:64–71.

Kraulis P. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots structures. *J Appl Crystallogr 24*:946–950.

Kuhn P, Knapp M, Soltis SM, Ganshaw G, Thoene M, Bott R. 1998. The 0.78 A structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry 37*:13446–13452.

Luciano P, Geli V. 1996. The mitochondrial processing peptidase: function and specificity. *Experientia 52*:1077–1082.

Matthews BW. 1988. Structural basis of the action of thermolysin and related zinc peptidases. *Acc Chem Res 21*:333–340.

Meinnel T, Blanquet S, Dardel F. 1996. A new subclass of the zinc metalloproteases superfamily revealed by the solution structure of peptide deformylase. *J Mol Biol 262*:375–386.

Murzin AG. 1998. How far divergent evolution goes in proteins. *Curr Opin Struct Biol 8*:380–387.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol 247*:536–540.

Ragusa S, Mouchet P, Lazennec C, Dive V, Meinnel T. 1999. Substrate recognition and selectivity of peptide deformylase. Similarities and differences with metzincins and thermolysin. *J Mol Biol 289*:1445–1457.

Rawlings ND, Barrett AJ. 1995. Evolutionary families of metallopeptidases. *Methods Enzymol 248*:183–228.

Russell RB. 1998. Detection of protein three-dimensional side-chain patterns: New examples of convergent evolution. *J Mol Biol 279*:1211–1227.

Saavedra-Alanis VM, Rysavy P, Rosenberg LE, Kalousek F. 1994. Rat liver mitochondrial processing peptidase. Both alpha- and beta-subunits are required for activity. *J Biol Chem 269*:9284–9288.

Tanudji M, Sjoling S, Glaser E, Whelan J. 1999. Signals required for the import and processing of the alternative oxidase into mitochondria. *J Biol Chem 274*:1286–1293.

Tronrud DE, Holden HM, Matthews BW. 1987. Structures of two thermolysin-inhibitor complexes that differ by a single hydrogen bond. *Science 235*: 571–574.

Tronrud DE, Roderick SL, Matthews BW. 1992. Structural basis for the action of thermolysin. *Matrix Suppl 1*:107–111.

Tsukada H, Blow DM. 1985. Structure of alpha-chymotrypsin refined at 1.68 A resolution. *J Mol Biol 184*:703–711.

Voet D, Voet JG. 1990. *Biochemistry*. New York: John Wiley & Sons, Inc.

Waltner M, Weiner H. 1995. Conversion of a nonprocessed mitochondrial precursor protein into one that is processed by the mitochondrial processing peptidase. *J Biol Chem 270*:26311–26317.

Xia D, Yu CA, Kim H, Xia JZ, Kachurin AM, Zhang L, Yu L, Deisenhofer J. 1997. Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. *Science 277*:60–66.

Zhang Z, Huang L, Shulmeister VM, Chi YI, Kim KK, Hung LW, Crofts AR, Berry EA, Kim SH. 1998. Electron transfer by domain movement in cytochrome bc1. *Nature 392*:677–684.