J. Biochem. 121, 304-308 (1997)

Structure of the Zinc Endoprotease from *Streptomyces caespitosus*[1]

Genji Kurisu,[2] Takayoshi Kinoshita, Akiko Sugimoto, Akinobu Nagara, Yasushi Kai, Nobutami Kasai, and Shigeharu Harada[3]

Department of Applied Chemistry, Faculty of Engineering, Osaka University, Suita, Osaka 565

Received for publication, September 27, 1996

A zinc endoprotease produced by *Streptomyces caespitosus* (ScNP) specifically hydrolyzes the peptide bond at the imino side of aromatic residues and is the smallest protease found to date. Although ScNP carries the zinc-binding sequence HEXXH, its primary structure of 132 amino acid residues differs from those of other known zinc metalloendoproteases. X-ray structural analysis of ScNP at 1.6 Å resolution revealed that despite a lack of sequence homology, the common topological feature of main-chain folding and a β-turn containing methionine, which is a feature of the zinc metalloendoprotease superfamily of metzincins, is conserved in ScNP. The zinc atom of ScNP is tetrahedrally ligated by the two histidines in the HEXXH sequence, an aspartate residue and a water molecule. Thus, ScNP represents a novel subfamily of metzincins with a HEXXHXXGXXD zinc-binding sequence. A plausible substrate recognition pocket to which aromatic residues bind is located near the catalytic zinc ion.

Key words: crystal structure, novel subfamily, substrate recognition pocket, zinc metalloprotease.

Zinc metalloendoproteases are widely involved in the intracellular processing and degradation of proteins. Vallee and Auld showed that these enzymes isolated from bacteria of *Bacillus* species share the consensus amino acid sequence, HEXXH, in their active sites and that the two histidine residues in the sequence ligate the catalytic zinc ion (*1, 2*). According to the sequence homology, Jiang and Bond grouped known 40 amino acid sequences of zinc endoproteases into five distinct subfamilies (astacin, serratia, matrixin, snake venom, and thermolysin) (*3*). Although the number of amino acid residues, sequence and substrate specificity are totally different among the families, all families except thermolysin have the consensus sequence HEXXHXXGXXH. The three histidine residues in the HEXXHXXGXXH sequence are zinc ligands.

In 1969, Yokote and Noguchi found a novel zinc endoprotease (ScNP) in the culture broth of *Streptomyces caespitosus* (*4, 5*). This enzyme, which belongs to the "*Streptomyces* Extracellular Neutral Proteinase Family (M7)" according to Rawlings and Barrett (*6*), consists of a single polypeptide chain of 132 amino acid residues with one disulfide bond, and it specifically cleaves the peptide bond at the imino sides of aromatic amino acid residues (*7*). Calcium ions play an essential role in the thermostability of ScNP but are not directly involved in its catalytic activity (*4*). For example, after incubating ScNP for 10 min at 60°C in the presence of calcium, it retains half the maximum catalytic activity shown at 40°C. However, ScNP from which calcium has been removed by dialysis against deionized water starts losing activity from 40°C and completely loses it at 60°C. Although ScNP does not share overall homology with the amino acid sequences of other zinc proteases, it contains the sequence HETGHVLGLPDH between residues 83 and 94. By comparison with the consensus sequence HEXXHXXGXXH, His 83 and His 87 are ScNP zinc ligands. However, it was uncertain whether the third zinc ligand is Asp 93 or His 94.

In this study we determined the crystal structure of ScNP at 1.6 Å resolution to identify the zinc-binding amino acid residues and the substrate-recognition mechanism.

METHODS

ScNP was purified and crystallized as described (*8*). Crystals of ScNP grown at 4°C from acetone were of space group $P2_12_12_1$, with the unit cell dimensions $a=55.21$ Å; $b=55.27$ Å; $c=37.60$ Å. Since acetone is volatile, the crystals were transferred to 50% MPD containing 5 mM calcium acetate (mother liquor) before X-ray studies. Five possible heavy atom derivatives [CH_3HgCl, $HgCl_2$, $Pb(CH_3COO)_2$, K_3IrCl_6, K_2PtCl_6] were obtained by soaking. The crystals soaked in 1 mM EDTA for 30 min were processed to locate the catalytically essential Zn^{2+} ion. ScNP complexed with *N*-carbobenzoxy-Gly-Tyr (*N*-CBZ-Gly-Tyr) was co-crystallized.

X-ray diffraction data were collected on a Rigaku four-circle diffractometer. Ni-filtered Cu$K\alpha$ radiation was produced by a Rigaku RU-300 rotating anode generator operated at 40 kV and 300 mA. Major heavy atom sites of the CH_3HgCl derivative were located on a difference

[1] This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas (no. 07224201) from the Ministry of Education, Science, Sports and Culture of Japan. G. Kurisu is grateful to JSPS Fellowships for Japanese Junior Scientists. S. Harada and Y. Kai are members of the TARA Sakabe project of University of Tsukuba, Japan.

[2] To whom correspondence should be addressed. Tel: +81-6-879-7410, Fax: +81-6-879-7409, e-mail: genji@ap.chem.eng.osaka-u.ac.jp

[3] Present address: Faculty of Pharmaceutical Sciences, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113.

Patterson map, and single isomorphous replacement phases were calculated. Heavy atom sites of the other derivatives were determined using difference Fourier maps. Heavy atom parameters including position, occupancy, and temperature factors were refined, and final best phase angles based on the five derivatives were calculated.

Although the electron density map at 2.5 Å resolution (MIR map) was sufficient to recognize secondary structure, it was further improved using the program SQUASH (*9*). A ($F_{NATIVE} - F_{EDTA}$, α_{MIR}) difference Fourier map revealed two prominent peaks. The catalytically essential zinc ion was assigned to the higher peak and the calcium ion to the lower. The zinc site coincided with the highest electron density in the MIR map. The model of ScNP was built using a Silicon Graphics Iris workstation with the program TURBO-FRODO (*10*).

The initial model was refined in iterative cycles consisting of crystallographic refinement with the programs X-PLOR (*11*) and PROLSQ (*12*) and manual model correction. The final model included 1,017 non-hydrogen protein atoms and 132 solvent molecules. The final R factor for the 12,926 independent reflections between 10.0 and 1.6 Å resolution with $F > 2\sigma(F)$ was 0.166. The r.m.s. bond length and bond angle deviations from ideal values were 0.017 Å and 4.3°, respectively. The coordinates were deposited in the Brookhaven Protein Data Bank. The model of the *N*-CBZ-Gly-Tyr was built in an ($F_{complex} - F_{native}$) electron density based on calculated native phase angles, and it was refined in the same way as the native form. The final R factor was 0.180 for the 8,415 independent reflections between 10.0 and 2.0 Å resolution.

A Ramachandran plot (Fig. 1) of the refined native model at 1.6 Å resolution shows that all non-glycine amino acid residues have ϕ/ψ angles in favorable regions. Examination using the program PROCHECK (*13*) showed that there were no unacceptable geometries in the model.

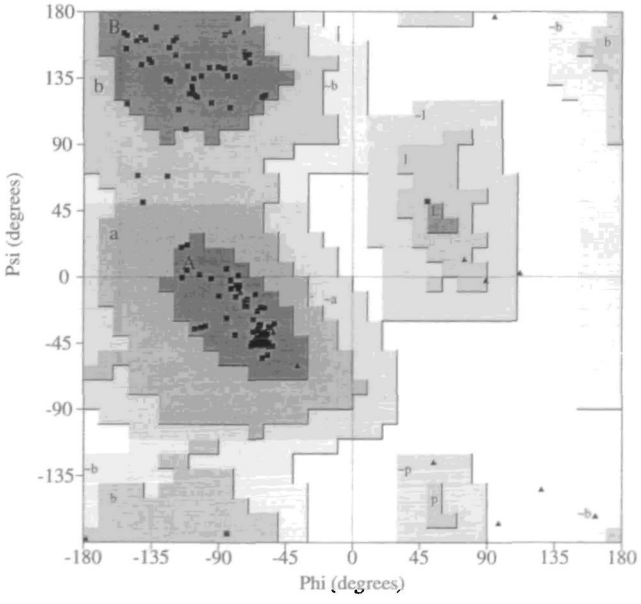

Fig. 1. **Ramachandran plot for the refined model of ScNP.** Glycines are represented by triangles. The Ramachandran plot was produced by the program PROCHECK (*13*).

RESULTS AND DISCUSSION

The refined structure of ScNP at 1.6 Å resolution shows that it consists of a highly twisted five-stranded β-sheet, four α-helices termed A, B, C, and D, one catalytically essential zinc ion and one calcium ion. The secondary structural elements connected by loops are disposed as shown in Fig. 2. The β-sheet, which contains four parallel strands (I, II, III, and V) and one antiparallel strand (IV), is flanked by helices A and C on its concave side, while its convex side is exposed to solvent (open-sandwich structure). The entire polypeptide chain of ScNP is divided into two domains. The N-domain (Thr 1–Asp 76) is organized as relatively regular secondary structural elements including the five-stranded β-sheet, helices A and B, whereas the polypeptide chain of the C-domain (Gly 90–Gly 132) folds in a more irregular manner. The long helix C, which is located immediately behind the catalytic zinc, connects the two domains.

These two domains form the zinc-containing active site cleft, which is bordered by β-strand IV, helix C, and a stretch of a loop adjacent to the C-terminus of this helix. The C helix runs roughly parallel to the cleft through the center of ScNP and includes the HEXXH sequence, of which the two histidine residues ligate the catalytic zinc through the Nε2 atoms. The helix terminates abruptly at Gly 90, and the polypeptide chain turns sharply away from it. This conformation enables Asp 93 to be near the catalytic zinc, and accordingly the Oδ1 atom of Asp 93

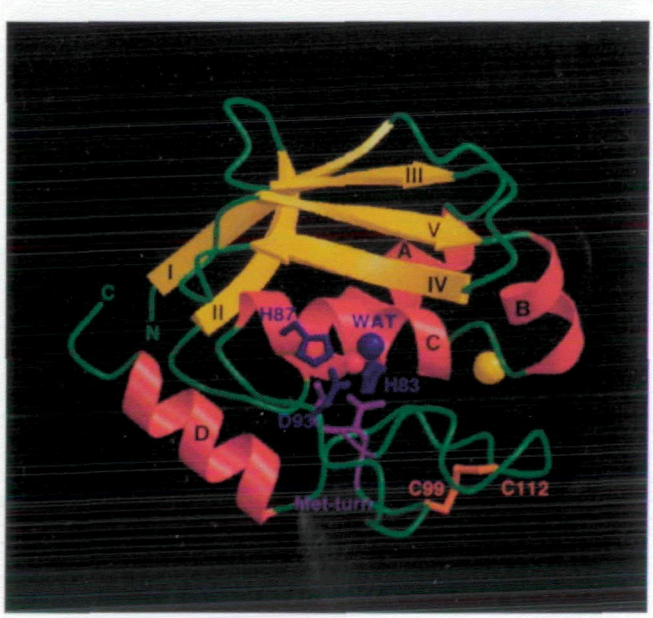

Fig. 2. **Ribbon drawing of the ScNP structure prepared using MOLSCRIPT (*22*) and RASTER3D (*23, 24*).** The amino and carboxy termini are labeled N and C, respectively. Cys 99 and Cys 112 represent the two cysteines of the disulfide bridge. The four helices (A–D) are from residues 15–26, 69–71, 77–88, and 119–128; and the five β-strands (I–V) from residues 2–7, 32–35, 42–46, 54–57, and 64–67. The zinc and calcium ions are shown as red and yellow spheres, respectively. The zinc ligands (His 83, His 87, Asp 93, and the water molecule) are shown in blue. Leu 102 and Met 103 in the Met-containing β-turn (magenta) provide a hydrophobic environment to His 83.

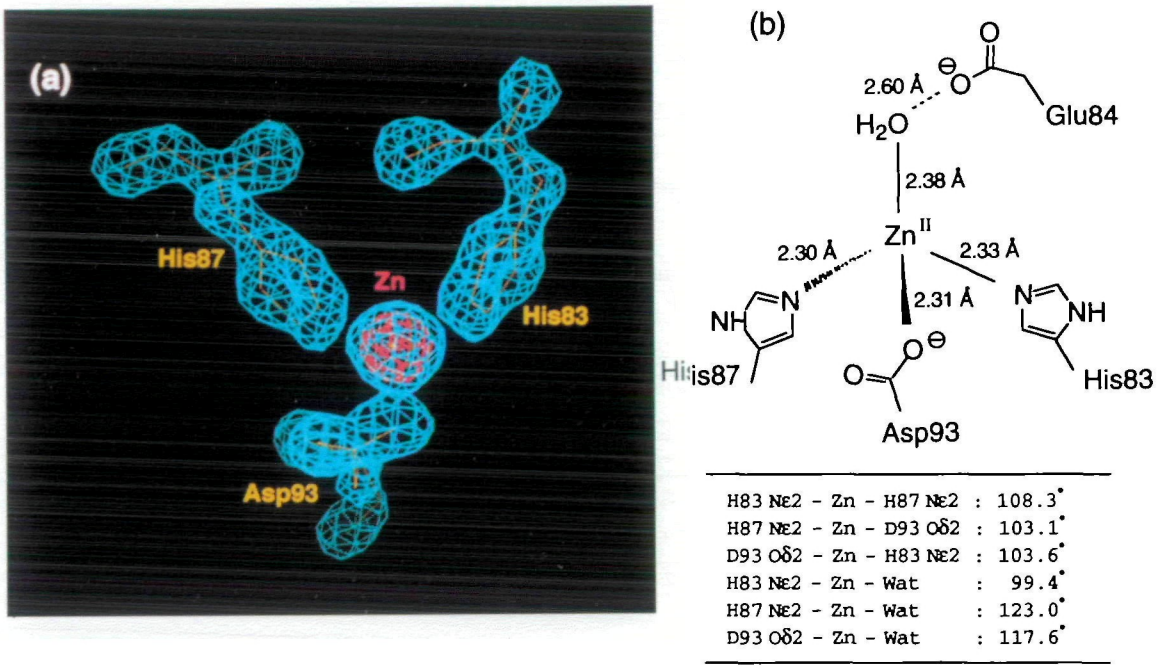

Fig. 3. **(a) The active site structure of ScNP.** The superimposed electron density was derived from a ($|F_o|-|F_c|$) map calculated with the α_c phases obtained from the refined 1.6 Å structure with the contribution from the zinc ligands omitted. **(b) Geometries of the zinc ligands.**

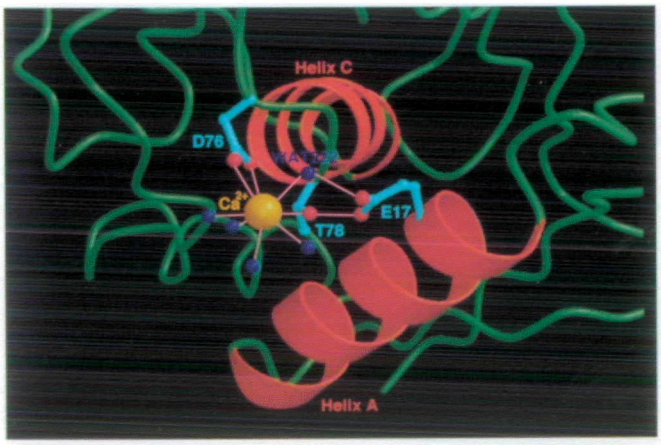

Fig. 4. **The structure and interactions (distance < 3 Å between non-hydrogen atoms) around the bound calcium site.**

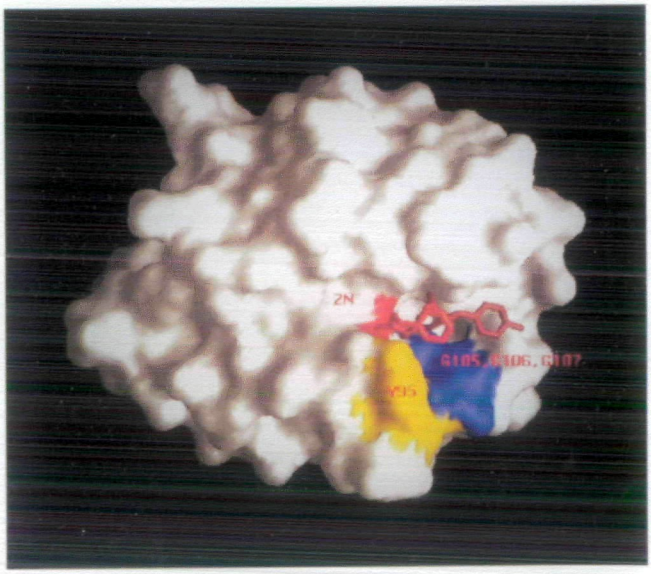

Fig. 5. **Molecular surface representation of ScNP complexed with *N*-CBZ-Gly-Tyr.** The sites of glycines 105, 106, and 107 are colored blue. The *N*-CBZ-Gly-Tyr molecule is colored orange. The zinc ion is represented as a red sphere. Tyr 95 which probably participates in catalysis is shown in yellow. This diagram was prepared using the program GRASP (*25*).

ligates the zinc ion, together with the Nε2 atoms of His 83 and His 87. These three ligands plus a water molecule complete tetrahedral ligation of the catalytic zinc (Fig. 3). By analogy with thermolysin (*14*), the bound water, the nucleophilicity of which is enhanced by a hydrogen bond to the side chain carboxyl of Glu 84, and at the same time by the ligation of the zinc, probably attacks the scissile peptide bond.

After Asp 93, the polypeptide chain forms a loop structure (loop-1) that consists of alternate and consecutive type I and type III β-turns, including His 94-Gly 97 (type I), Pro 98-Glu 101 (type III), Glu 101-Ser 104 (type I), and Gly 105-Pro 108 (type III). The third methionine-containing β-turn (Glu 101-Leu 102-Met 103-Ser 104) is equivalent to the "Met-turn" found in the "metzincins" (*15*). The side chains of Met 103 and Leu 102, which are 3.8 Å from His 83, provide a hydrophobic environment to the zinc-ligating His 83 side chain. This hydrophobic environment may influence the zinc acidity and promote the proton dissociation from Zn^{2+}-OH_2. Subsequent to loop-1, the polypeptide chain from Gly 105 to Asn 114 folds in a completely irregular manner (loop-2). The disulfide bridge between

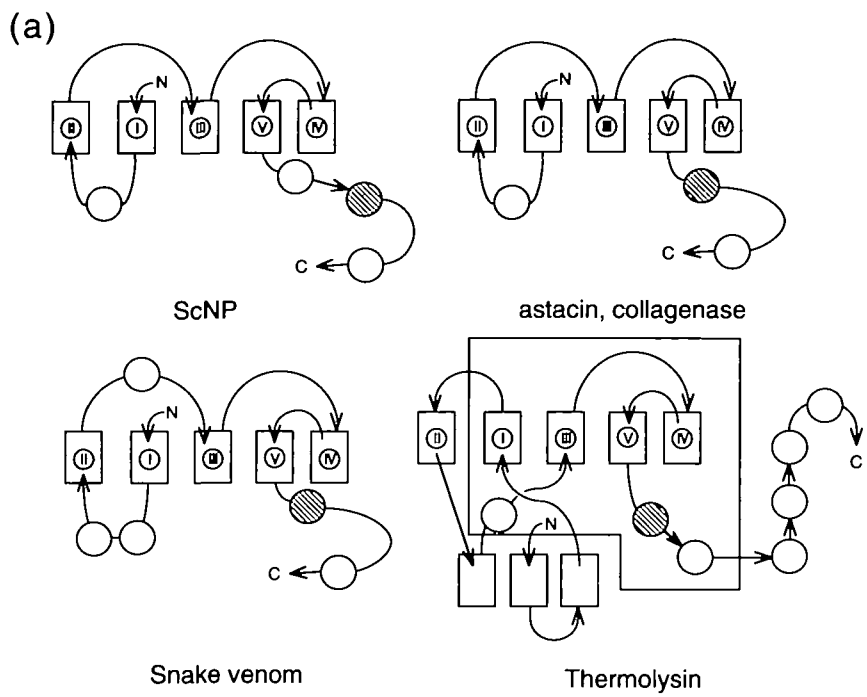

Fig. 6. **(a) Topological structure of main-chain folding of zinc metalloendoproteases.** The square and the circle represent a β-strand and an α-helix, respectively. **(b) Comparison of the amino acid sequences of zinc-binding sites of zinc metalloendoproteases.**

(b)

	75																	93
ScNP[7]	Y	D	S	T	R	V	T	A	H	E	T	G	H	V	L	G	L	P	D
Collagenase[26]	Y	N	L	H	R	V	A	A	H	E	L	G	H	S	L	G	L	S	H
Astacin[27]	C	V	Y	H	G	T	I	I	H	E	L	M	H	A	I	G	F	Y	H
Snake venom[28]	F	M	V	A	V	T	M	T	H	E	L	G	H	N	L	G	M	E	H
Serralysin[29]	D	Y	G	R	Q	T	F	T	H	E	I	G	H	A	L	G	L	S	H
Thermolysin[30]	S	G	G	I	D	V	V	A	H	E	L	T	H	A	V	T	D	Y	T

Cys 99 and Cys 112 interconnects loops-1 and -2 and probably stabilizes the more fragile three-dimensional structure of the C-domain. Helix D, which is the only secondary structure of the C-domain, is located on the C-terminus.

Figure 4 shows the structure around the calcium found in ScNP. The ion, which is located near helices A and C, 17.0 Å from the catalytic zinc, interacts with eight oxygen atoms: Oδ1 and Oδ2 of Asp 76, Oγ1 of Thr 78, and five waters. The distances between the calcium and the eight oxygen atoms are about 2.8 Å. WAT 222 and Oγ1 of Thr 78 are further connected to the carboxyl oxygen atoms of Glu 17 through hydrogen bonds. Since none of the residues in helix A other than Glu 17 interact with the other parts of the molecule, this interaction network suggests that the calcium ion contributes to the thermostability of ScNP by helping to link the A and C helices.

Inspection of the ScNP structure revealed a plausible substrate recognition pocket near the catalytic zinc ion. A pocket into which an aromatic side chain of a polypeptide may fit is formed by Gly 105, Gly 106, and Gly 107 on one side, Gln 71, Tyr 75, and Val 80 on the other and of Asp 76 at the bottom. Figure 5 shows the structure of the ScNP complexed with *N*-CBZ-Gly-Tyr. According to Yokote and Noguchi (*5*), ScNP hydrolyzed *N*-CBZ-Gly(Glu)-Xaa gradually when Xaa is Phe, but the enzyme was inactive toward Tyr. Although we supposed that the carbonyl oxygen of the Gly residue of *N*-CBZ-Gly-Tyr binds to the catalytic zinc (mode A) so that the side chain of the Tyr residue can be accommodated in the pocket, *N*-CBZ-Gly-Tyr binds to the cleft of ScNP through ligation of the carbonyl oxygen of the carbobenzoxy group (mode B). This binding mode is stabilized by the hydrogen bond between the Tyr side chain of the compound and ScNP (Tyr 75). Of course, binding mode A should occur to the same degree as B, and accordingly the compound should be hydrolyzed by ScNP. However, ScNP is bound by *N*-CBZ-Gly-Tyr again in the two modes after hydrolysis, and eventually the more stable complex with the binding mode B is accumulated. Thus the activity of ScNP toward *N*-CBZ-Gly-Tyr is lost in accordance with the accumulation of the complex B. This presumption explains why ScNP is active toward *N*-CBZ-Gly-Phe, which can not form a hydrogen bond with Tyr 75. Tyr 95 seems to play a role in catalysis similar to that of Tyr 157 in thermolysin (*15, 16*).

Figure 6a shows a topology diagram of the zinc endoproteases with known structures (*17-21*). Although the topology of the β-sheet in thermolysin differs somewhat from that of the others, their structures are basically composed of a five-stranded β-sheet and a central α-helix

with the zinc-binding HEXXH sequence as a common structural component. ScNP is more topologically similar to the catalytic domain of fibroblast collagenase and astacin than to snake venom protease. Moreover, the amino acid sequence of ScNP around the zinc ligand (Tyr 75-Asp 93) is homologous to that of collagenase with 58% identity, although the other regions are not homologous. This sequence identity is not found among ScNP and astacin, serralysin, thermolysin, except for the HEXXH sequence (Fig. 6b). Despite the homology between ScNP and collagenase, the third aspartate zinc-ligand of ScNP is replaced by a histidine in collagenase. Thus, we propose that ScNP, together with the corresponding enzymes from other species of *Streptomyces* (*6*), represents a novel subfamily with the HEXXHXXGXXD zinc-binding motif and a "Met-turn."

REFERENCES

1. Vallee, B.L. and Auld, D.S. (1990) Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **29**, 5647-5659
2. Vallee, B.L. and Auld, D.S. (1990) Active-site zinc ligands and activated H_2O of zinc enzymes. *Proc. Natl. Acad. Sci. USA* **87**, 220-224
3. Jiang, W. and Bond, J.S. (1992) Families of metalloendopeptidases and their relationships. *FEBS Lett.* **312**, 110-114
4. Yokote, Y., Kawasaki, K., Nakajima, J., and Noguchi, Y. (1969) Studies on enzymes produced by *Streptomyces caespitosus*. Part I. Production conditions and some properties of neutral protease. *Nippon Nogeikagaku Kaishi* **43**, 125-131
5. Yokote, Y. and Noguchi, Y. (1969) Studies on enzymes produced by *Streptomyces caespitosus*. Part II. Crystallization and some properties of neutral protease. *Nippon Nogeikagaku Kaishi* **43**, 132-138
6. Barrett, A.J. (1995) Evolutionary families of metallopeptidases in *Methods in Enzymology* (Rawlings, N.D. and Barrett, A.J., eds.) Vol. 248, pp. 183-228, Academic Press, New York
7. Harada, S., Kinoshita, T., Kasai, N., Tsunasawa, S., and Sakiyama, F. (1995) Complete amino acid sequence of a zinc metalloendoprotease from *Streptomyces caespitosus*. *Eur. J. Biochem.* **233**, 683-686
8. Harada, S., Kitadokoro, K., Kinoshita, T., Kai, Y., and Kasai, N. (1991) Crystallization and main-chain structure of neutral protease from *Streptomyces caespitosus*. *J. Biochem.* **110**, 46-49
9. Zhang, K.Y.J. (1993) *SQUASH*-combining constraints for macromolecular phase refinement and extension. *Acta Crystallogr.* **D49**, 213-222
10. Harold, W., Wyckoff, C.H.W., Hirs., and Serge N, Timasheff. (1985) Interactive computer graphics: FRODO in *Methods in Enzymology* (Jones, T.A., ed.) Vol. 115, pp. 157-171, Academic Press, New York
11. Brünger, A.T. (1992) *X-PLOR Version 3.1: A system for Crystallography and NMR*, Yale University Press, New Haven
12. Hendrickson, W.A. and Konnert, J.H. (1981) *Stereochemistry Restrained Crystallographic Least-Squares Refinement of Macromolecule Structures*, Pergamon Press, Oxford
13. Collaborative Computational Project (1994) Number 4. The CCP4 Suite: Programs for protein crystallography. *Acta Crystallogr.* **D50**, 760-763
14. Matthews, B.W. (1988) Structural basis of the action of thermolysin and related zinc peptidase. *Acc. Chem. Res.* **21**, 333-340
15. Bode, W., Gomis-Ruth, F.X., and Stockler, W. (1993) Astacins, serralysins, snake venom and matrix metalloendoproteinases 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
16. Hangauer, D.G., Monzingo, A.F., and Matthews, B.W. (1984) An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by *N*-carboxymethyl dipeptides. *Biochemistry* **23**, 5730-5741
17. Holmes, M.A. and Matthews, B.W. (1982) Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.* **160**, 623-639
18. 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
19. 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
20. Lovejoy, B., Cleasby, A., Hassell, A.M., Longley, K., Luther, M. A., Weigl, D., McGeehan, G., McElroy, A.B., Drewry, D., Lambert, M.H., and Jordan, S.R. (1994) Structure of the catalytic domain of fibroblast collagenase complexed with an inhibitor. *Science* **263**, 375-377
21. Borkakoti, N., Winkler, F.K., Williams, D.H., D'Arcy, A., Broadhurst, M.J., Brown, P.A., Johnson, W.H., and Murray, E.J. (1994) Structure of the catalytic domain of human fibroblast collagenase complexed with an inhibitor. *Nature Struct. Biol.* **1**, 106-110
22. Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950
23. Bacon, D.J. and Anderson, W.F. (1988) A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graphics* **6**, 219-220
24. Merritt, E.A. and Murphy, M.E.P. (1994) Raster3D Version 2.0 —A program for photorealistic molecular graphics. *Acta Cryst.* **D50**, 869-873
25. Nicholls, A., Sharp, K., and Honig, B. (1991) Protein folding and association—insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296
26. Goldberg, G.I., Wilhelm, S.M., Kronberger, A., Bauer, E.A., Grant, G.A., and Eisen, A.Z. (1986) Human fibroblast collagenase. *J. Biol. Chem.* **261**, 6600-6605
27. Titani, K., Torff, H.-J., Hormel, S., Kumar, S., Walsh, K.A., Rodl, J., Neurath, H., and Zwilling, R. (1987) Amino acid sequence of a unique protease from the crayfish *Astacus fluviatilis*. *Biochemistry* **26**, 222-226
28. Takeya, H., Arakawa, M., Miyata, T., Iwanaga, S., and Omori-Satoh, T. (1989) Primary structure of H_2-proteinase, a non-hemorrhagic metalloproteinase, isolated from the venom of the Habu snake, *Trimeresurus flavoviridis*. *J. Biochem.* **106**, 151-157
29. Braunagel, S.C. and Benedik, M.J. (1990) The metalloprotease gene of *Serratia marcescens* strain SM6. *Mol. Gen. Genet.* **222**, 446-451
30. Titani, K., Hermodson, M.A., Ericsson, L.H., Walsh, K.A., and Neurath, H. (1972) Amino-acid sequence of thermolysin. *Nature New Biol.* **238**, 35-37