J. Biochem. 119, 844-851 (1996)

# Crystal Structure of *Serratia* Protease, a Zinc-Dependent Proteinase from *Serratia* sp. E-15, Containing a β-Sheet Coil Motif at 2.0 Å Resolution[1]

**Kensaku Hamada,*[,2] Yasuo Hata,† Yoshio Katsuya,‡ Hajime Hiramatsu,* Takaji Fujiwara,* and Yukiteru Katsube§**

**Faculty of Science, Shimane University, Matsue, Shimane 690; †Institute for Chemical Research, Kyoto University, Uji, Kyoto 611; ‡Hyogo Prefectural Institute of Industrial Technology, Kobe, Hyogo 654; and §Institute for Protein Research, Osaka University, Suita, Osaka 565*

Received for publication, October 30, 1995

**The crystal structure of *Serratia* protease from *Serratia* sp. E-15 was solved by the single isomorphous replacement method supplemented with anomalous scattering effects from both the Zn atom in the native crystal and the Sm atom in the derivative crystal, and refined at 2.0 Å resolution to a crystallographic *R*-factor of 0.194. The enzyme consists of N-terminal catalytic and C-terminal β-sandwich domains, as observed in alkaline protease from *Pseudomonas aeruginosa* IFO3080. The catalytic domain with a five-stranded antiparallel β-sheet and five α-helices shares a basically common folding topology with those of other zinc metalloendoproteases. The catalytic zinc ion at the bottom of the active site cleft is ligated by His176, His180, His186, Tyr216, and a water molecule in a distorted trigonal-bipyramidal manner. The C-terminal domain is a β-strand-rich domain containing eighteen β-strands and a short α-helix, and has seven $Ca^{2+}$ ions bound to calcium binding loops. An unusual β-sheet coil motif is observed in this domain, where the β-strands and calcium binding loops are alternately incorporated into an elliptical right-handed spiral so as to form a two-layer untwisted β-sandwich structure. The $Ca^{2+}$ ions in the C-terminal domain seem to be very important for the folding and stability of the β-sheet coil structure.**

**Key words: β-sheet coil, crystal structure, metalloprotease, *Serratia* protease, *Serratia* sp. E-15.**

*Serratia* protease, first isolated from culture medium of *Serratia* sp. E-15, is an extracellular metalloproteinase containing one catalytic zinc ion per molecule (*1, 2*). The substrate specificity of the enzyme is rather broad, compared with that of thermolysin [EC 3.4.24.4] from *Bacillus thermoproteolyticus*. *Serratia* protease is known to enhance the vascular permeability by activating the tissue Hageman factor XII and the subsequent bradykinin-generating cascade, and to degrade various human plasma proteins (*3*). The enzyme was also reported to be directly absorbed through the bowel wall and to be temporarily inhibited by a plasma $\alpha_2$-macroglobulin (*4*).

The enzyme, with a molecular weight of about 50,600, comprises a polypeptide chain of 470 residues and one catalytic zinc ion. The amino acid sequence of the protein has been deduced from the DNA sequence of its structural gene (*5*). It is known that the protein binds several calcium ions. These calcium ions are not essential for the catalytic activity of the enzyme, but are required for protecting the protein from autolysis. Some of these calcium ions seem to be strongly bound to the protein and are not removed by treatment with chelating agents such as EDTA. *Serratia* protease has the zinc binding consensus sequence HEXXHXXGXXH, where X is an arbitrary amino acid. In this sequence, the three histidines are zinc ligands, and the glutamic acid is believed to be the catalytic base. The consensus sequence has been found in several sequentially unrelated zinc-dependent enzyme families, including the astacin, snake venom proteinase, and matrixin families. These families belong to the metzincin metalloprotease superfamily (*6*), and are distinguished from the thermolysin family with the zinc binding consensus sequence HEXXH. Furthermore, an interesting feature in the amino acid sequence is that *Serratia* protease contains several repeats of the consensus sequence GGXGXD, which is believed to be a calcium binding motif, between residues 333 and 374. *Serratia* protease is the archetypical enzyme in the serralysin family of zinc-proteases, and shares significant homology in amino acid sequence with alkaline protease from another Gram-negative bacterium, *Pseudomonas aeruginosa* (*7*). For secretion of *Serratia* protease, an N-terminal signal sequence is not required, as is also the case for secretion of hemolysin A from *Escherichia coli*, which has a repeat of the same consensus sequence. This suggests that both enzymes may be secreted in a similar manner. Small-angle X-ray scattering (SAXS) studies on the structure of *Serratia* protease in solution allowed us to

[1] This work was supported in part by Grants-in-Aid for Scientific Research on Priority Areas (No. 03680049 to K.H., and No. 06235204 and No. 07224203 to Y.H.) from the Ministry of Education, Science and Culture of Japan.

[2] To whom correspondence should be addressed. Phone: +81-852-32-6480, Fax: +81-852-32-6489, E-mail: hamadak@cis.shimane-u.ac.jp

show that its molecular shape is an elongated ellipsoid of approximately 110×40×40 Å with the zinc ion located in a large cleft, and that the size of the cleft is strictly controlled to attain optimal activity (*8*).

Recently, structural studies on serralysin zinc-proteases have been carried out by X-ray diffraction methods. A comparison between the liganded and unliganded structures of *Pseudomonas aeruginosa* alkaline protease has revealed some significant conformational changes around the active site on ligand binding (*9, 10*). The structures of *Serratia* protease from *Serratia marcescens* and its complex with an inhibitor have also been studied by crystallographic methods (*11, 12*). Detailed structural studies of *Serratia* protease from *Serratia* sp. E-15, the archetypical zinc-protease of the serralysins, are essential not only for defining the structure of the serralysin zinc-protease family, but also for elucidating structure-function relationships of serralysins. In order to gain better insight into the structural characteristics for specificity and mechanism of the serralysin family, we have undertaken structural investigations of the *Serratia* protease. We published the first report on the crystal structure of *Serratia* protease from *Serratia* sp. E-15 in 1993 (*13*). In the present paper, we report the crystal structure of the *Serratia* protease at 2.0 Å resolution, and discuss the implications of the structural details for the catalytic activity and a novel $\beta$-sheet coil motif.

## EXPERIMENTAL

*Crystallization and Data Collection*—The *Serratia* protease from *Serratia* sp. E-15 was isolated and purified as described previously (*1, 2*). Crystals were prepared by dialyzing a 20 mg/ml enzyme solution containing 2% (w/v) polyethyleneglycol 8000 against 27% saturated ammonium sulfate (*14*). The crystals have orthorhombic space group $P2_12_12_1$ with cell dimensions of $a=109.18$, $b=150.89$, and $c=42.64$ Å. The density of the crystals was determined to be 1.23 g/cm$^3$ by a linear density gradient flotation method, and this was used to calculate the presence of one molecule in the asymmetric unit and the solvent content of 67%. The only useful heavy-atom derivative (Sm-derivative) was prepared by soaking the native crystals in 30% saturated ammonium sulfate solution containing 20 mM $Sm(NO_3)_3$ for 5 days.

X-ray diffraction data for the native enzyme were collected at the synchrotron radiation X-ray beamline $BL6A_2$ in the Photon Factory, National Laboratory for High Energy Physics, Tsukuba, using a screenless Weissenberg camera loaded with imaging plates (*15*). Two X-ray wavelengths ($\lambda$) of 1.000 and 1.283 Å were used for the data collection. The wavelength of 1.283 Å, which is shorter than that corresponding to the *K* absorption edge of zinc ($\lambda=1.2834$ Å), was precisely adjusted by measuring the X-ray fluorescence from the crystals of *Serratia* protease to enhance anomalous dispersion effects from zinc in the native crystal. For X-ray exposure at each wavelength, two sets of data were collected from two crystals rotated around the $a^*$ and $c^*$ axes, respectively. In all the cases, the crystals were aligned to allow simultaneous recording of as many Bijvoet pair reflections as possible. The crystals were kept at 12°C during the data collection; the oscillation range, the overlap of adjacent images and the ratio of crystal rotation to detector movement were 5.0°, 0.6°, and 1.0°/mm, respectively. Data were processed using the program WEIS (*16*). Diffraction data for the Sm-derivative were collected under two different conditions: on the Weissenberg camera with synchrotron radiation ($\lambda=1.283$ Å) at the Photon Factory and on a Rigaku R-AXIS IIC with Cu $K\alpha$ radiation ($\lambda=1.5418$ Å). Two sets of native data collected at the wavelength of 1.000 Å were scaled and merged together. Two native data sets collected at the wavelength of 1.283 Å were independently used for X-ray analysis. The data sets from Sm-derivative crystals were also independently used without merging. The data collection statistics are given in Table I.

*Structure Determination and Refinement*—The structure was solved by the conventional heavy-atom isomorphous replacement method supplemented with anomalous scattering techniques. The native data sets collected at the wavelength of 1.000 Å were merged into the single "native" data set after scaling together, while the native data sets collected at the wavelength of 1.283 Å were treated as different "Zn-derivative" data sets. The site of the zinc ion was identified from an anomalous difference Patterson map at 4.0 Å resolution using the Zn-derivative data set collected in the $c^*$-axis setting at the wavelength of 1.283 Å, and was subsequently refined against intensity differences of Bijvoet pairs in the data set using the software package PHASES (*17*). Two major sites of Sm ion were obtained from a difference Patterson map at 3.0 Å resolution and then refined using PHASES. Besides, two minor sites were obtained from a difference Fourier map. The parameters of the four Sm sites were then refined to convergence. Initial

TABLE I. **Statistics for data collection.**

| | Crystals | | | | | |
|---|---|---|---|---|---|---|
| | Native | Native | Native | $Sm(NO_3)_3$ | $Sm(NO_3)_3$ | $Sm(NO_3)_3$ |
| Resolution (Å) | 1.8 | 3.0 | 2.3 | 3.0 | 2.3 | 2.4 |
| X-ray source | SR | SR | SR | SR | SR | CuK$\alpha$ |
| Wavelength (Å) | 1.000 | 1.283 | 1.283 | 1.283 | 1.283 | 1.542 |
| No. of crystals | 2 | 1 | 1 | 1 | 1 | 1 |
| Rotation axis | $a^*$, $c^*$ | $a^*$ | $c^*$ | $a^*$ | $c^*$ | $a^*$ |
| Observations | 151,619 | 57,140 | 41,041 | 56,296 | 43,202 | 65,766 |
| Unique reflections | 54,062 | 14,412 | 14,691 | 14,328 | 15,350 | 15,350 |
| Bijvoet pairs | | 10,855 | 10,890 | 11,246 | 11,261 | 12,742 |
| Completeness (%) | 81.4 | 97.0 | 68.9 | 96.5 | 55.0 | 80.1 |
| (Å) | (1.8) | (3.0) | (3.0) | (3.0) | (3.0) | (2.4) |
| $R_{merge}$[a] (%) | 5.80 | 5.58 | 3.59 | 6.42 | 6.22 | 8.74 |

[a] $R_{merge}$ (%) $=100\ \Sigma_i|\langle I_i\rangle - I_i|/\Sigma_i I_i$; $\langle I_i\rangle$ is the average of $I_i$ over all observed symmetry equivalents.

TABLE II. **Statistics for structure refinement.**

| Parameters | Target $\sigma$ | r.m.s. deviation |
|---|---|---|
| Bond distances (Å) | | |
| Bond length | 0.020 | 0.012 |
| Angle-related distance | 0.040 | 0.038 |
| Interolanar distance | 0.050 | 0.040 |
| Planar groups (Å) | 0.020 | 0.010 |
| Chiral volume (Å) | 0.150 | 0.148 |
| Nonbonded contacts (Å) | | |
| Single torsion | 0.500 | 0.180 |
| Multiple torsion | 0.500 | 0.199 |
| Possible hydrogen bond | 0.500 | 0.185 |
| Torsion angles (°) | | |
| Planar | 3.0 | 1.8 |
| Staggered | 15.0 | 16.7 |
| Orthonormal | 20.0 | 28.9 |
| Thermal factors (Å$^2$) | | |
| Main-chain bond | 1.000 | 0.687 |
| Main-chain angle | 1.500 | 1.116 |
| Side-chain bond | 1.000 | 0.800 |
| Side-chain angle | 1.500 | 1.327 |

phases were calculated at 3.0 Å resolution by the isomorphous replacement method supplemented with anomalous dispersion effects from the zinc for each of two Zn-derivative data sets collected by $a^*$-axis setting and $c^*$-axis setting, respectively, and from the samarium for each of the Sm-derivative data sets, using the program PHASIT in the program package PHASES. The phases were used to calculate an initial electron density map at 3.0 Å resolution, and then improved considerably by four rounds of sixteen cycles of solvent-flattening procedures (*18*). In the solvent-flattening, the solvent content was set to 0.45, 0.45, 0.50, and 0.55 for the respective rounds, which were lower than the experimentally estimated value of 0.67. The sphere radius for the density summation was set to 9.0 Å at 3.0 Å resolution. The phase extension was gradually carried out from 3.0 to 2.5 Å resolution in four steps of the solvent-flattening procedures: to 2.8, 2.7, 2.6, and 2.5 Å resolutions. The resulting electron density map was interpreted with TURBO-FRODO (*19*) working on a Silicon Graphics IRIS Indigo-Elan computer. An initial molecular model was built manually and then refined with the program X-PLOR (*20*) up to an $R$-factor of 0.281 for data between 8.0 and 2.5 Å resolutions. The model was rebuilt manually based on both $2F_o-F_c$ and $F_o-F_c$ omit maps calculated with the coordinates from X-PLOR. The improved model was refined with the program GPRLSA, which was based upon PROLSQ (*21*), in the program package PHASES. In the course of refinement, the resolution was gradually improved up to 2.0 Å resolution. At this resolution, we found out that the model with the amino acid sequence FRQRL, which was assigned to the residues 334 to 338 from the DNA sequence, did not fit well into the corresponding electron density map; the sequence in this region was replaced by the sequence G$^{334}$SGNDV$^{339}$ so that the amino acid sequence was well fitted to the electron density map. The model of the molecule modified in amino acid sequence was further refined by the program GPRLSA. A difference Fourier map calculated with the refined model showed seven prominent density peaks which were significantly higher than the others. Each of the seven peaks bound to the protein *via* ligands including a vicinal Asp in the octahedral coordination geometry often observed in Ca-binding proteins whose tertiary structures are known. Since the *Serratia* protease was expected to bind several $Ca^{2+}$ ions, these peaks were assigned as $Ca^{2+}$ ions. In the difference Fourier map, 217 other significant peaks were reasonably taken as water molecules. After a total of 95 cycles of the refinement, the crystallographic $R$-factor for the current model, which consists of residues 4 to 471, one zinc ion, seven calcium ions, and 217 water molecules, was 0.184 for 38,044 reflections with $F>3\sigma(F)$ in the resolution range of 8.0 to 2.0 Å. The refinement parameters are given in Table II. The atomic coordinates of the model have been deposited in the Brookhaven protein data bank (ID code: 1SRP).

## RESULTS AND DISCUSSION

*Overall Structure of Serratia Protease*—The *Serratia* protease from *Serratia* sp. E-15 is an elongated ellipsoidal molecule with the size of approximately $90\times35\times30$ Å (Fig. 1), which is somewhat smaller than that of the hydrated molecule estimated by small-angle X-ray scattering studies (*8*). Figure 2 shows a schematic diagram of the secondary structure arrangement, whose elements were assigned on the basis of the criteria described by Kabsch and Sander (*22*). The molecule consists of two domains and an extended N-terminal tail. The N-terminal proteolytic domain having a kidney-like shape comprises residues 20 to 251 and the catalytic zinc ion. The C-terminal domain having a novel $\beta$-sheet structure comprises residues 252 to 471. In addition, the N-terminal tail of residues 4 to 19, including an $\alpha$-helix A, is in contact with the C-terminal domain. The N-terminal three residues which were invisible on the final electron density map are omitted from the model. There is a groove, between the N-terminal and C-terminal domains, which contains several water molecules. The two domains are linked by two hydrogen bonds between the main chains of Asp225 and Lys315, and between the side chains of Asp98 and Arg269. The *Serratia* protease shares 55% sequence identity with the alkaline protease from *Pseudomonas aeruginosa* IFO3080 consisting of 470 amino acid residues with the catalytic zinc ion. The overall structures of both enzymes are similar to each other (*10*).

*N-Terminal Proteolytic Domain*—The N-terminal proteolytic domain can be subdivided into two parts: the N-terminal region from residues 20 to 183 and the C-terminal region from residues 184 to 251, which are separated by the deep active site cleft. The N-terminal region consists mainly of a five-stranded mixed $\beta$-sheet, 2-1-3-5-4, and four $\alpha$-helices, B to E. The $\beta$-sheet has a parallel arrangement in strands 2, 1, 3, and 5, and an antiparallel arrangement in strand 4. Of three long $\alpha$-helices, B, C, and E, $\alpha$-helix B, the first $\alpha$-helix in the N-terminal region of this domain, is located near the surface of the C-terminal region, while $\alpha$-helices C and E run approximately parallel to $\beta$-strands 2 and 4, respectively. The $\alpha$-helix E is the "active-site $\alpha$-helix" which anchors two of the zinc ligands, His176 and His180, and the catalytic base residue, Glu177, in the zinc-binding consensus sequence motif HEXXHXXGXXH, which is characteristically found in the metzincin superfamily (*6*). The short $\alpha$-helix D following the $\beta$-sheet leads to the active-site $\alpha$-helix E. The C-terminal region of this domain consists of an irregular structure with two short $3_{10}$-helices, some turns and $\alpha$-helix F. The loop region

of Asp189 to Thr198 at the entrance to the active site cleft has a high temperature factor on average and seems to be relatively flexible, as in the equivalent region of the *P. aeruginosa* alkaline protease. The loop in the *Serratia* protease seems to protrude further from the molecular surface than that in the alkaline protease. As a consequence, the entrance to the active site in the *Serratia* protease is wider than that in the alkaline protease.

The N-terminal domain of the *Serratia* protease is topologically similar to those of an astacin (*23, 24*), a snake-venom metalloproteinase (*25*) and collagenases (*26–30*), which belong to zinc-dependent metalloprotease families different from that of the *Serratia* protease, respectively. However, the homology of amino acid sequence between these families is relatively weak. The *Serratia* protease shares less than 30% sequence identity with the others. A structural comparison shows that the N-terminal region of this domain in the *Serratia* protease is more similar to that of the astacin than that of the thermolysin. Significant structural differences between the N-terminal regions of the *Serratia* protease and the astacin are found in


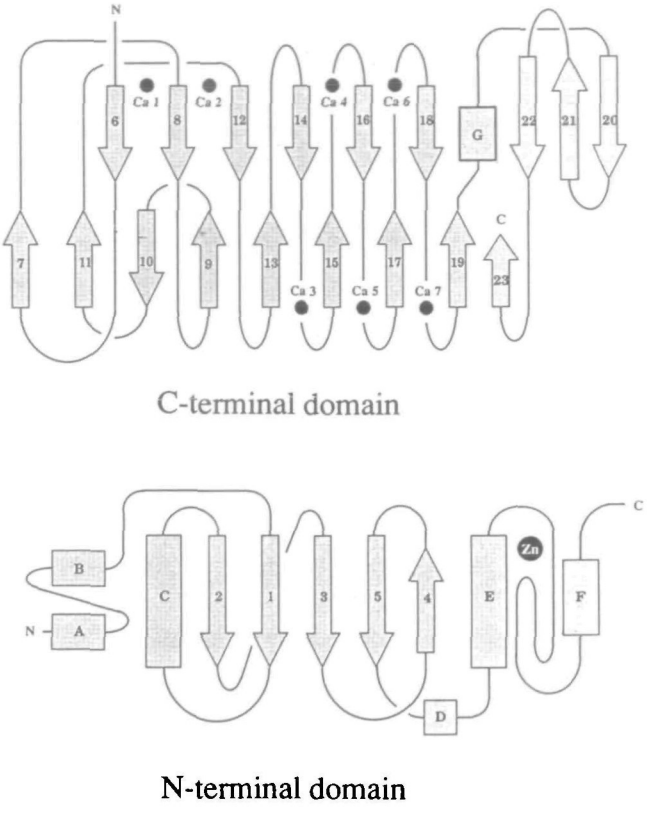


Fig. 2. **Topology diagrams of the *Serratia* protease with the catalytic zinc ion site and seven calcium binding sites: N-terminal proteolytic domain, consisting of a five-stranded $\beta$-sheet and five $\alpha$-helices with the additional N-terminal $\alpha$-helix, and C-terminal domain, consisting of eighteen $\beta$-strands and an $\alpha$-helix.** *$\alpha$-Helices are depicted as cylinders, $\beta$-strands as arrows,* the zinc ion as a large filled circle, and calcium ions as small filled circles.


Fig. 1. **The molecular structure of *Serratia* protease.** (a) A ribbon diagram showing the positions of the secondary structure elements drawn with MOLSCRIPT (*37*). $\beta$-Strands are depicted as arrows, the zinc ion is shown by a large black ball, and calcium ions by small ones. The N-terminal proteolytic domain containing the zinc ion is shown to the lower right, while the C-terminal domain adopting the $\beta$-sheet coil structure with $Ca^{2+}$ ions is shown to the upper left. (b) A stereoview of the $C_\alpha$ plot is shown in the same direction as (a). Positions of $\alpha$-carbon atoms are shown by white balls.

two loops which are positioned between strands 3 and 4, and between strands 4 and 5. These loops in the *Serratia* protease are somewhat longer than those in the astacin, and serve to narrow the entrance to the active site. An additional difference between the structures lies in the N-terminal extension of 50 residues in the *Serratia* protease, compared with the astacin. These residues form the N-terminal $\alpha$-helices A and B which are located on the molecular surface of the irregular C-terminal region of the N-terminal domain. However, the five-stranded $\beta$-sheet and two helices, C and E, in the N-terminal region of this domain in the *Serratia* protease are also topologically conserved in both the astacin (*23, 24*) and the thermolysin (*31, 32*). Thus, the folding topology of the five-stranded $\beta$-sheet and two $\alpha$-helices may be common to different families of zinc-dependent metalloendoproteases.

*Active Site*—The active site zinc ion, located at the bottom of the active site cleft, is ligated in a distorted trigonal-bipyramidal geometry by five zinc ligands. The zinc is strongly coordinated by the $N_{\epsilon 2}$ atoms of His176, His180, and His186 on the zinc-binding consensus sequence HEXXHXXGXXH and by a water molecule which makes a hydrogen bond with the side-chain carboxylate of Glu177 (Fig. 3). The $N_{\epsilon 2}$ atoms of His176 and His186 and the water molecule are arranged at the vertices of a distorted triangular plane with the central zinc ion displaced by 0.37 Å from this plane towards an axial ligand, His180. The fifth zinc ligand is the hydroxyl oxygen atom of Tyr216, in the C-terminal region of the N-terminal domain, which occupies the other axial coordination position. The distance between the zinc and hydroxyl oxygen is 3.09 Å, which is significantly longer than expected. Consequently, the coordinate bond becomes weak. The hydroxyl oxygen atom of Tyr216 is quite close to the ligand water molecule, with a distance of 2.90 Å. The zinc ligation system in the *Serratia* protease is similar to those of the *P. aeruginosa* alkaline protease and the astacin, but is quite different from those of the thermolysin, the elastase from *P. aeruginosa* (*33*) and the neutral protease from *Bacillus cereus* (*34*), which belong to the thermolysin family. In the thermolysin family, the active-site zinc ion is coordinated in a distorted tetrahedral geometry by two histidine residues from the consensus sequence HEXXH on the active-site $\alpha$-helix, a glutamyl residue that is positioned 20 residues downstream from the second histidine residue, and a water molecule. Two carboxyl oxygen atoms of the glutamyl residue in the sequence motif HEXXH make hydrogen bonds with the ligand water molecule, and the glutamyl residue has been thought to function as the active base in the enzymatic reaction (*35, 36*). This is indeed the case with Glu177 in the zinc-binding consensus sequence of the *Serratia* protease.

*C-Terminal $\beta$-Sheet Coil Domain*—The C-terminal noncatalytic domain consists of 220 residues which form eighteen $\beta$-strands and one $\alpha$-helix (Figs. 2 and 4). This domain is subdivided into two regions: the amino-proximal region of residues 252 to 404 which consists of fourteen $\beta$-strands, 6 to 19, and the carboxy-terminal region of the remaining 67 residues which contains an $\alpha$-helix G and a three-stranded antiparallel $\beta$-sheet, 20-21-22, and an additional short $\beta$-strand 23.

In the amino-proximal region, $\beta$-strands, each of which is formed by only three amino acid residues, are incorporated into an elliptical coil winding in a right-handed spiral where two layers of parallel $\beta$-sheets are formed in an untwisted $\beta$-sandwich manner. This new structural motif can be called a "$\beta$-sheet coil." In the $\beta$-sheet coil of the *Serratia* protease, one $\beta$-sheet layer comprises six parallel $\beta$-strands, and the other layer comprises seven parallel and one antiparallel $\beta$-strands. The adjacent $\beta$-strands in the $\beta$-sheet coil are separated by 4 to 20 residues along the sequence, with the most frequent length of six residues. The amino-proximal region of the C-terminal domain can be further divided into three small regions: an irregular $\beta$-sheet coil region of strands 6 to 11, a very regular $\beta$-sheet coil region of strands 12 to 17, and a somewhat irregular $\beta$-sheet coil region of strands 18 to 19. The structure of the $\beta$-sheet coil in the *Serratia* protease is bent between the first irregular $\beta$-sheet coil region and the regular $\beta$-sheet coil region. The inner core of the first irregular $\beta$-sheet coil region is mainly filled with aromatic and hydrophobic side-chains. In the regular $\beta$-sheet coil region, $\beta$-strands 12 to 17 alternating with loops are regularly coiled in a right-handed spiral to form an elliptical cylinder whose ellipse has dimensions of 7.2–9.0 Å between the sandwiched $\beta$-

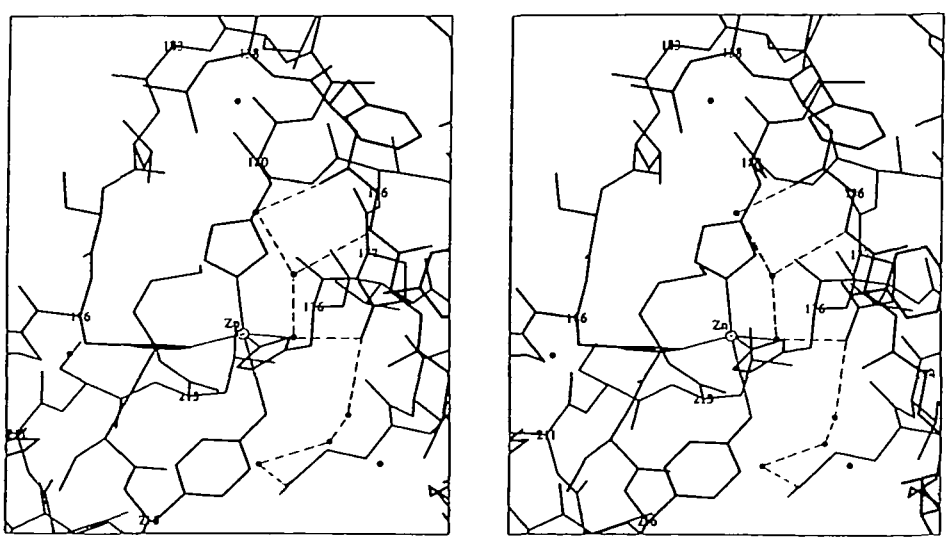

Fig. 3. **A stereoview of the zinc binding site in the *Serratia* protease.** The zinc ion is shown as a white ball, and external water molecules as black balls. The residues, His176, His180, His186, and Tyr216, and a water molecule are coordinated to the zinc ion as shown by solid lines. The residue Glu177 makes a hydrogen bond with the ligand water. Broken lines represent hydrogen bonds.

sheets and of 22.0 Å along the long axis (Fig. 5).

In the $\beta$-sheet coil region, there is a tandem repeat of the consensus amino acid sequence motif UUUGGXGXD, where U is predominantly a bulky hydrophobic residue and X is an arbitrary residue. The first three residues UUU of this motif form a short $\beta$-strand structure incorporated into a parallel $\beta$-sheet of the $\beta$-sheet coil region, where U1, U2, and U3 are preferably valine, leucine and a hydrophobic residue with a bulky side-chain, respectively (Fig. 6). The side chains of U2 residues, ideally leucines, are projected into the interior of the $\beta$-sheet coil and stacked with one another like a zipper, while those of U1 and U3 are projected to the exterior of the $\beta$-sheet coil. The last six residues GGXGXD of the consensus sequence motif form a short turn for $Ca^{2+}$ binding. X6 and X8 are positioned at the corners of the turn, and the former residue is preferably a small hydrophobic residue, glycine or alanine. The side chain of D9 projects into the inner core of the $\beta$-sheet coil to become a calcium ligand in a unique manner, as described below.

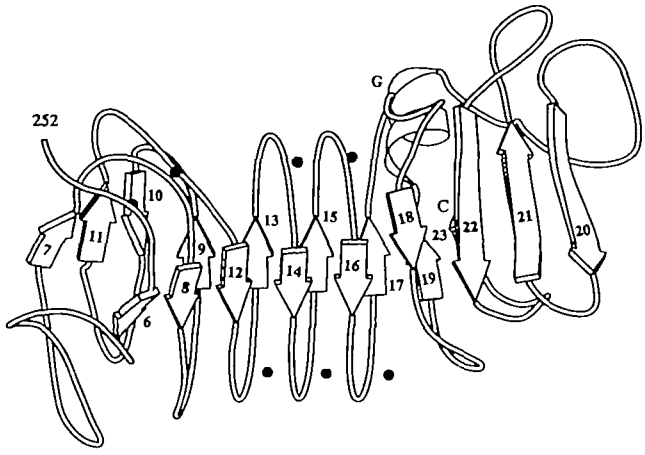


Fig. 4. **A schematic diagram of the C-terminal domain of the *Serratia* protease.** $\beta$-Strands are shown as arrows. The C-terminal domain is subdivided into two regions: an N-terminal region having a novel $\beta$-sheet coil structure motif, which consists of fourteen $\beta$-strands, 6 to 19, and a C-terminal region comprising 67 residues which consists of an $\alpha$-helix, three-stranded antiparallel $\beta$-sheet, 20-21-22, and a short $\beta$-strand 23. The $\beta$-sheet coil structure motif consists of two layers of parallel $\beta$-sheets with an antiparallel $\beta$-strand 10. Furthermore, the N-terminal region of the C-terminal domain is divided into three parts. The middle part of the N-terminal region, which is formed by the tandem repeat of the consensus sequence UUUGGXGXD, has a quite regular $\beta$-sheet coil structure including strands 13 to 17. The N-terminal part including strands 6 to 12 and the C-terminal part including strands 18 to 19 have irregular $\beta$-sheet coil structures.

The folding of the C-terminal domain in the *Serratia* protease is similar to that of the alkaline protease from *P. aeruginosa*, except for a few distinct differences. The *Serratia* protease adopts mainly an irregular structure with a short $3_{10}$-helix between strands 6 and 7, while the alkaline protease has two $\beta$-strands in this region. Besides, the $\alpha$-helix G following the $\beta$-sheet coil region in the *Serratia* protease is replaced by a $\beta$-strand, in the alkaline protease, which is somewhat shifted towards the position corresponding to $\beta$-strand 20 in the *Serratia* protease. These differences lie in the beginning of the $\beta$-sheet coil region and the C-terminal region following the $\beta$-sheet coil, which are on the surface of the molecule, and have no influence on the overall conformation of the C-terminal domain in both enzymes. Thus, the two-layer $\beta$-sheet sandwich including the $\beta$-sheet coil motif observed in the C-terminal domain of the *Serratia* protease is expected to be a unique folding conserved in the serralysin family of zinc-endoproteases.

*$Ca^{2+}$ Binding Sites*—*Serratia* protease has several bound calcium ions which are not essential for its catalytic function, but are required for protection from autolysis. At the final stage of structure refinement, we found from a difference Fourier map that at least seven calcium ions, Ca1-Ca7, were bound to the enzyme, as shown in Fig. 6. Of

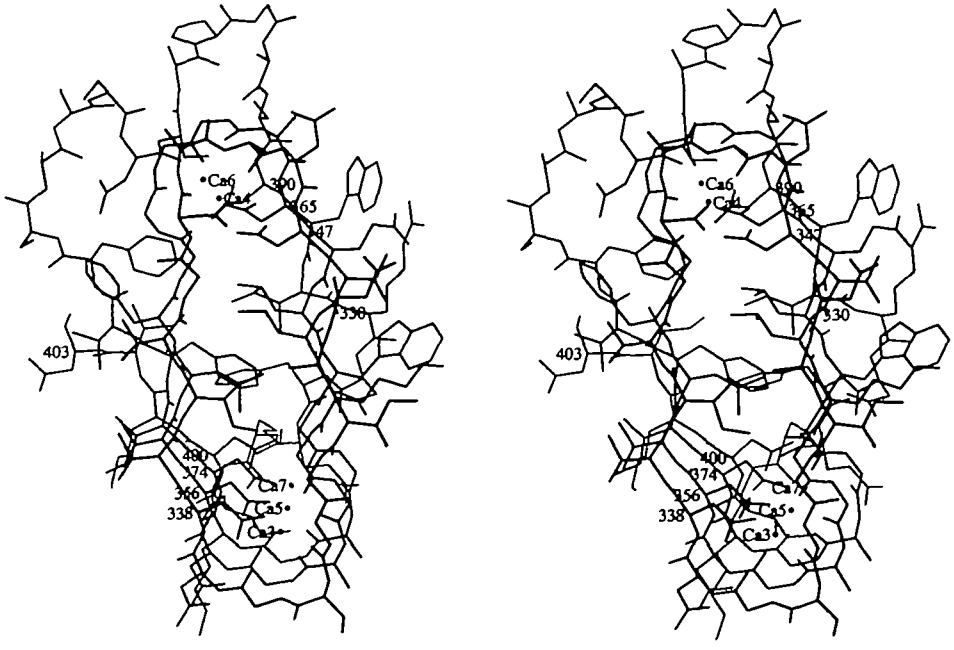


Fig. 5. **A stereoview of the regular $\beta$-sheet coil structure formed by residues 330 to 403 in the C-terminal domain.** The structure is shown in the direction viewed down along the spiral axis. Calcium ion sites are shown as small balls.

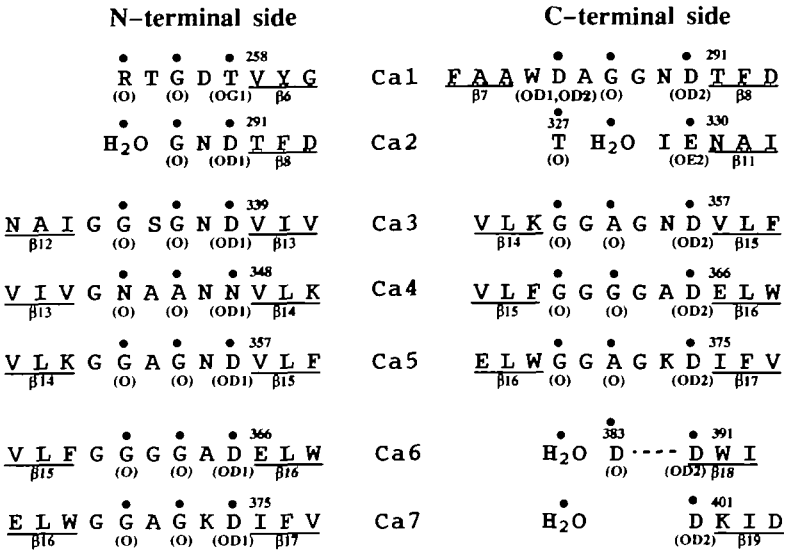


Fig. 6. **The amino acid sequence in $\beta$-sheet coil with $Ca^{2+}$ binding sites.** Amino acid residues liganded to $Ca^{2+}$ ions are indicated by black dots, and ligand atoms are shown in parentheses. $\beta$-Strand regions are indicated by underlines with sheet number.

these $Ca^{2+}$ ions, two $Ca^{2+}$ ions, Ca1 and Ca2, are located in the irregular $\beta$-sheet coil region, and the other five $Ca^{2+}$ ions, Ca3 to Ca7, are located in the regular $\beta$-sheet coil region. Four of the seven $Ca^{2+}$ ions, Ca1, Ca3, Ca4, and Ca5, are deeply buried between neighboring $Ca^{2+}$-binding loops in the $\beta$-sheet coil and shielded from the solvent. These $Ca^{2+}$ ions are bound only to atoms from the protein. Three of these four $Ca^{2+}$ ions, Ca3 to Ca5, are each coordinated by six ligands in an octahedral geometry, as is often observed in $Ca^{2+}$-binding proteins, while the other $Ca^{2+}$ ion, Ca1, is coordinated by seven ligands including two carboxyl oxygens of Asp285. These findings are in good agreement with the fact that these $Ca^{2+}$ ions can not be removed from the protein by EDTA treatment. In contrast to these $Ca^{2+}$ ions, the other three $Ca^{2+}$ ions, Ca2, Ca6, and Ca7, are easily removed from the protein by EDTA treatment, and/or substituted for $Sm^{3+}$ ions during soaking of the crystal in $Sm^{3+}$ ion-containing solution, because they are exposed to the solvent. These $Ca^{2+}$ ions have water molecules as their ligands. Ca2 is located at the same position as the $Sm^{3+}$ site 1, one of the major $Sm^{3+}$ sites in the Sm-derivative crystal. It is coordinated in an octahedral geometry by six ligands including two water molecules. Ca6 has one water molecule as a ligand and is also coordinated by six ligands in an octahedral geometry. The putative $Ca^{2+}$ ion, Ca7, is assigned to a density peak with about 40% of the electron density of a typical calcium ion site on the difference Fourier map, because this site corresponds to another $Sm^{3+}$ site in the Sm-derivative crystal and is located 2.5 to 3.0 Å away from Gly370, Gly372, and Asp374, which are included in the UUUGGXGXD motif, as also observed in other $Ca^{2+}$ binding sites. Ca7 is coordinated in an octahedral geometry by five ligands including one water molecule. This $Ca^{2+}$ ion may be liganded by two water molecules, one of which is missing in the electron density maps. The distribution of the $Ca^{2+}$ binding sites in the C-terminal domain of the *Serratia* protease is quite similar to that of alkaline protease from *P. aeruginosa*. The *Serratia* protease has seven $Ca^{2+}$ binding sites, while the alkaline protease has eight $Ca^{2+}$ binding sites. All the sites in the *Serratia* protease are in the amino-terminal region of the C-terminal domain, and are common with seven of the eight sites in the alkaline protease.

The functional role of the $\beta$-sheet coil adopted by the repeat of the $Ca^{2+}$ binding sequence UUUGGXGXD in the *Serratia* protease is not clear yet, but may be related to the translocation system for secretion of proteins, because the repeat of the consensus sequence UUUGGXGXD is also observed in other proteins secreted from Gram-negative bacteria which do not utilize any N-terminal signal peptide for that purpose (*9*).

We are grateful to Takeda Chemical Industry Ltd. for the kind gift of the *Serratia* protease. We thank Drs. Noriyoshi Sakabe, Atsushi Nakagawa and Nobuhisa Watanabe, Photon Factory, National Laboratory of High Energy Physics, Tsukuba, Japan, for their kind help in data collection using the Weissenberg camera and synchrotron radiation.

## REFERENCES

1. Miyata, K., Maejima, K., Tomoda, K., and Isono, M. (1970) *Serratia* protease. Part I. Purification and general properties of the enzyme. *Agric. Biol. Chem.* **34**, 310-318
2. Miyata, K., Tomoda, K., and Isono, M. (1971) *Serratia* protease. Part III. Characteristics of the enzyme as metalloprotease. *Agric. Biol. Chem.* **35**, 460-467
3. Kamata, R., Yamamoto, T., Matsumoto, K., and Maeda, H. (1985) A serratial protease causes vascular permeability reaction by activation of the Hageman factor-dependent pathway in guinea pigs. *Infect. Immun.* **48**, 747-753
4. Miyata, K., Hirai, S., Yashiki, T., and Tomoda, K. (1980) Intestinal absorption of *Serratia* protease. *J. Appl. Biochem.* **2**, 111-116
5. Nakahama, K., Yoshimura, K., Marumoto, R., Kikuchi, M., Lee, I.S., Hase, T., and Matsubara, H. (1986) Cloning and sequencing of *Serratia* protease gene. *Nucleic Acids Res.* **14**, 5843-5855
6. Bode, W., Gomis-Rüth, F.X., and Stöcker, 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
7. Okuda, K., Morihara, K., Atsumi, Y., Takeuchi, H., Kuwamoto, S., Kawasaki, H., Suzuki, K., and Fukushima, J. (1990) Complete nucleotide sequence of the structural gene for alkaline proteinase in *Pseudomonas aeruginosa* IFO3455. *Infect. Immun.*

**58**, 4083-4088
8. Katsuya, Y., Sato, M., Katsube, Y., Matsuura, Y., and Tomoda, K. (1992) Small-angle X-ray scattering study of metal ion-induced conformational changes in *Serratia* protease. *J. Biol. Chem.* **267**, 12668-12672
9. 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
10. Miyatake, H., Hata, Y., Fujii, T., Hamada, K., Morihara, K., and Katsube, Y. (1995) Crystal structure of the unliganded alkaline protease from *Pseudomonas aeruginosa* IFO3080 and its conformational changes on ligand binding. *J. Biochem.* **118**, 474-479
11. Baumann, U. (1994) Crystal structure of the 50 kDa metalloprotease from *Serratia marcescens*. *J. Mol. Biol.* **244**, 244-251
12. Baumann, U., Bauer, M., Letoffe, S., Delpelaire, P., and Wandersman, C. (1995) Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi*. *J. Mol. Biol.* **248**, 653-661
13. Hamada, K., Hiramatsu, H., Fujiwara, T., Katsuya, Y., Hata, Y., Matsuura, Y., and Katsube, Y. (1993) Structural analysis of *Serratia* protease. *Acta Cryst.* **A49**, suppl. 102
14. Katsuya, Y., Hamada, K., Hata, Y., Tanaka, N., Sato, M., Katsube, Y., Kakiuchi, K., and Miyata, K. (1985) Preliminary X-ray studies on *Serratia* protease. *J. Biochem.* **98**, 1139-1142
15. Sakabe, N. (1983) A focusing Weissenberg camera with multilayer-line screens for macromolecular crystallography. *J. Appl. Crystallogr.* **16**, 542-547
16. Higashi, T. (1989) The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Crystallogr.* **22**, 9-18
17. Furey, W. and Swaminathan, S. (1990) PHASES—A program package for the processing and analysis of diffraction data from macromolecules. *Am. Crystallogr. Assoc. Annu. Mtg, Program Abst.* **18**, 73
18. Wang, B.C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90-112
19. Rousell, A. and Cambillou, C. (1991) *TURBO-FRODO, Silicon Graphics Applications Directory*, p. 86, Silicon Graphics, Mountain View, CA
20. Brünger, A.T. (1990) *X-PLOR Manual. Version 2.1*, Yale University, New Haven, CT
21. Hendrickson, W.A. (1985) Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252-270
22. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637
23. Bode, W., Gomis-Rüth, F.X., Huber, R., Zwilling, R., and Stöcker, W. (1992) Structure of astacin and implications for activation of astacins and zinc-ligation of collagenase. *Nature* **358**, 164-167
24. Gomis-Rüth, F.X., Stöcker, W., Huber, R., Zwilling, R., and Bode, W. (1993) Refined 1.8 Å X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. *J. Mol. Biol.* **229**, 945-968
25. Gomis-Rüth, F.X., Kress, L.F., and Bode, W. (1993) First structure of a snake-venom metalloproteinase: A prototype for matrix metalloproteinases/collagenases. *EMBO J.* **12**, 4151-4157
26. Lovejoy, B., Cleasby, A., Hassel, A.M., Longley, K., Luther, M.A., Weigl, D., McGeehan, G., McElory, A.B., Dewry, 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
27. Borkakoti, N., Winkler, F.K., Williams, D.H., D'Arcy, A., Broadhurst, M.J., Brown, P.A., Johnson, W., and Murray, E.J. (1994) Structure of the catalytic domain of human fibroblast collagenases complexed with an inhibitor. *Nature Struct. Biol.* **1**, 106-110
28. Stamas, T., Spurlino, J.C., Smith, D.L., Walh, R.C., Ho, T.F., Qorronfleh, M.W., Banks, T.M., and Rubin, B. (1994) Structure of human neutrophil collagenase reveals large S1′ specificity pocket. *Nature Struct. Biol.* **1**, 119-123
29. Reinemer, P., Grams, F., Huber, R., Kleine, T., Schnierer, S., Pieper, M., Tschesche, H., and Bode, W. (1994) Structural implications for the role of the N terminus in the "superactivation" of collagenases. *FEBS Lett.* **338**, 227-233
30. Bode, W., Reinemer, P., Huber, R., Kleine, T., Schnierer, S., and Tschesche, H. (1994) The crystal structure of human neutrophil collagenase inhibited by a substrate analog reveals the essentials for catalysis and specificity. *EMBO J.* **13**, 1263-1269
31. Matthews, B.W., Jansonius, J.N., Colman, P.M., Schoenborn, B.P., and Dupourque, D. (1972) Three-dimensional structure of thermolysin. *Nature New Biol.* **238**, 37-41
32. Holmes, M.A. and Matthews, B.W. (1982) Structure of thermolysin at 1.6 Å resolution. *J. Mol. Biol.* **160**, 623-639
33. Thayer, M.M., Flaherty, K.M., and McKay, D.B. (1991) Three-dimensional structure of the elastase of *Pseudomonas aeruginosa*. *J. Biol. Chem.* **266**, 2864-2871
34. Pauptit, R.A., Karlsson, R., Picot, D., Jenkins, J.A., Niklaos-Reimer, A.S., and Jansonius, J.N. (1988) Crystal structure of neutral protease from *Bacillus cereus* refined at 3.0 Å resolution and comparison with the homologous, but more thermostable enzyme thermolysin. *J. Mol. Biol.* **199**, 525-537
35. Kaster, W.R. and Matthews, B.W. (1977) Crystallographic study of the binding of dipeptide inhibitors to thermolysin: Implications for mechanism of catalysis. *Biochemistry* **16**, 2506-2616
36. Holmes, M.A. and Matthews, B.W. (1981) Binding of hydroxamic acid inhibitors to crystalline thermolysin suggests a pentacoordinate zinc intermediate in catalysis. *Biochemistry* **20**, 6912-6920
37. Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950