Atomic Structure of the Trypsin−Aeruginosin 98-B Complex

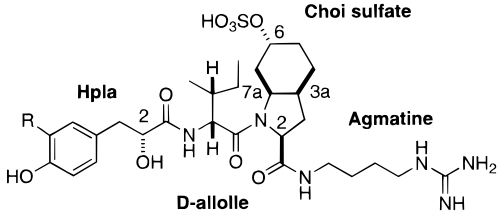
Benjamin Sandler,[†] Masahiro Murakami,[§] and Jon Clardy*[,†]

Department of Chemistry, Cornell University
Ithaca, New York 14853-1301
Laboratory of Marine Biochemistry, University of Tokyo
Bunkyo-ku, Tokyo, Japan
Received August 26, 1997

Proteolytic reactions control many biological processes, and the discovery of new proteolysis inhibitors forms an important field of chemistry. The serine protease thrombin, for example, initiates blood clot formation by cleaving fibrinogen to release fibrin.[1] The thrombin inhibitor hirudin, a 65-residue polypeptide produced by the medical leech *Hirudo medicinalis*, demonstrated antithrombotic efficacy in several clinical trials and stimulated efforts to discover lower molecular weight inhibitors.[1] Structural analyses of serine proteases complexed with hirudin and other large (30−190 residue) natural protease inhibitors have clarified the "standard mechanism" of inhibition wherein a reactive center corresponding to the scissile peptide bond is presented by an exposed loop and acts as an ideal substrate.[2] Structural analyses of much smaller (4−7 residue) naturally occurring inhibitors have also clarified their inhibitory mechanisms. Both the sponge metabolite cyclotheonamide and the blue-green algal metabolite A90720A, macrocyclic peptides with nonprotein amino acid residues, complexed with serine proteases have been analyzed by single-crystal X-ray diffraction.[3−6] Both cyclotheonamide and A90720A form tight-binding inhibitors in which the macrocyclic constraint and the nonprotein structural elements feature prominently. Cyclotheonamide inhibits through formation of a covalent bond[3, 5] while A90720A inhibits by the standard mechanism.[6] The recent discovery[7] of the serine protease inhibitor aeruginosin 98-B (**1**), an inhibitor lacking the molecular features needed to form a covalent or standard mechanism inhibitor, prompted the detailed structural analysis of its trypsin complex that forms the basis of this report.

Murakami and co-workers discovered aeruginosin 98-A (R = Cl) and 98-B (R = H) (Chart 1) from the blue-green alga *Microcystis aeruginosa* using a protease inhibition screen.[7] Aeruginosin 98-A and -B are equipotent inhibitors of thrombin, plasmin, and trypsin with IC_{50}s of 15, 10, and 0.9 μM, respectively.[7] While aeruginosin 98-B has a peptide-like structure, it contains no standard L-amino acids. The N-terminus is modified with 3-(4-hydroxyphenyl)lactic acid (Hpla), and the stereochemistry of the 2-hydroxy group was not established in the original work.[7] The next residue is D-alloisoleucine (D-alloIle), which is followed by 2-carboxy-6-hydroxyoctahydroindole sulfate (Choi sulfate). The relative configuration of all stereocenters in Choi sulfate were assigned by NMR, but the absolute configuration was not established.[7] The last residue is agmatine or decarboxylated arginine. Since the carboxy group of arginine forms the scissile peptide bond, aeruginosin 98-B lacks the structural features that figure so prominently in the well-characterized inhibitors. The atomic resolution structure of the trypsin−aeruginosin 98-B complex has established aeruginosin 98-B's complete absolute stereostructure along with the mechanism of inhibition.

Chart 1. Chemical Drawing of Aeruginosin 98-A and -B



Single crystals of the complex grow by the hanging-drop vapor-diffusion method (10 mg/mL trypsin complex in 100 mM pH 6.1 phosphate buffer against 20% PEG-4000 and 10% 2-propanol in 100 mM pH 8.5 HEPES buffer) in space group $P6_1$ with $a = b = 48.30$ and $c = 145.2$ Å. Diffraction data, which were measured on a flash frozen crystal ($0.5 \times 0.2 \times 0.2$ mm^3) at CHESS, were reduced[8] to yield 8944 reflections at 2.2 Å resolution (92% complete, merge $R = 6.3\%$). The structure was solved using molecular replacement.[9] The final model−1629 trypsin atoms, 45 ligand atoms, and 212 waters−refined to an $R = 16.4\%$ (free $R = 23.7\%$ both measured with 2.0 σ data) and rms deviations from ideality for bond lengths, bond angles, and torsional angles of 0.011 Å, 1.70°, and 25.7° respectively.[10] Bound aeruginosin 98-B was clearly defined by strong continuous electron density (Figure 1), and thermal parameters (6.97 Å^2 for ligand and 6.39 Å^2 for active site residues) suggested full occupancy.

Many of aeruginosin 98-B's interactions with trypsin employ nonpeptidic elements to form familiar serine protease interactions: a guanidinium in the primary specificity pocket, two antiparallel β-strand hydrogen bonds to Gly216, and a hydrophobic residue in the Trp215 "aryl-binding site" (Figure 2). The guanidinium of agmatine is buried deeply in trypsin's primary or S1 specificity pocket and forms two close contacts with the Asp189 side chain carboxylate (2.65 and 2.96 Å). The guanidinium interacts with a water that is also tightly bound to the side chain of Gln192 and the carbonyl oxygens of Gly216 and Gly219 (Figure 2). The antiparallel β-strand binding that characterizes serine protease recognition motifs involves the carbonyl oxygen of D-alloIle and the amide NH of Gly216 and the amide NH of D-alloIle with the carbonyl of Gly216. This series of hydrogen bonds is continued by the hydroxyl of Hpla and the amide NH of Gly219. D-alloIle participates in a van der Waals interaction with Trp215 and Leu99, and if the residue were an L-Ile or L-alloIle, the side chain would collide with Choi sulfate.

[†] Cornell University.
[§] University of Tokyo.

(1) Edmunds, J. J.; Rapundalo, S. T.; Siddiqui, M. A. *Annual Reports in Medicinal Chemistry;* Bristol, J. A., Ed. Academic Press: San Diego, 1996; Vol. 31, pp 51−60 and references therein.
(2) Bode, W.; Huber, R. *Eur. J. Biochem.* **1992,** *204,* 433−435.
(3) Lee, A. Y.; Hagihara, M.; Karmacharya, R.; Albers, M. W.; Schreiber, S. L.; Clardy, J. *J. Am. Chem. Soc.* **1993,** *115,* 12619−12620.
(4) Maryanoff, B. E.; Qiu, X.; Padmanabhan, K. P.; Tulinsky, A.; Almond, H. R.; Andrade-Grodon, P.; Greco, M. N.; Kauffman, J. A.; Nicolaou, K. C.; Liu, A.; Brungs, P. H.; Fusetani, N. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90,* 8048−8052.
(5) Ganesh, V.; Lee, A. Y.; Clardy, J.; Tulinsky, A. *Protein Sci.* **1996,** *5,* 825−835.
(6) Lee, A. Y.; Smitka, T. A.; Bonjouklian, R.; Clardy, J. *Chem. Biol.* **1994,** *1,* 113−118.
(7) Murakami, M.; Ishida, K.; Okino, T.; Okita, Y.; Matsuda, H.; Yamaguchi, K. *Tetrahedron Lett.* **1995,** *36,* 2785−2788.
(8) Diffraction data were collected at the CHESS F1 beamline using a 2k Princeton CCD detector (Tate, M. W., et al. *J. Appl. Crystallogr.* **1995**, *28*, 196−205; Thiel, D. J., et al. *Rev. Sci. Instrum.* **1996,** *67*, 1−4) and reduced using DENZO and SCALEPACK. (Z. Otwinowski and W. Minor, "The HKL Program Suite", in preparation.)
(9) Molecular replacement was carried out in AMoRe (Navaza, J. *Acta Crystallogr. A* **1994**, *50*, 157−163) using PDB coordinate file 1TPS as a search model. (Lee, A. Y.; Smitka, T. A.; Bonjouklian, R.; Clardy, J. *Chem. Biol.* **1994**, *1*, 113−118.)
(10) Model building and solvent placement were carried out in CHAIN (Sack, J. S. *J. Mol. Graph.* **1988**, *6*, 224−225) while simulated annealing and multiple cycles of rigid-body and B refinement were carried out in X-PLOR (Brunger, A. T. *X-Plor Version 3.1: A System for X-Ray Crystallography and NMR;* Yale University Press: New Haven, 1992). The geometry of the final model was evaluated in PROCHECK (Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26*, 283−291). Coordinates have been deposited at the Protein Data Bank under ID 1AQ7.

S0002-7863(97)02991-0 CCC: $15.00
Published on Web 01/03/1998

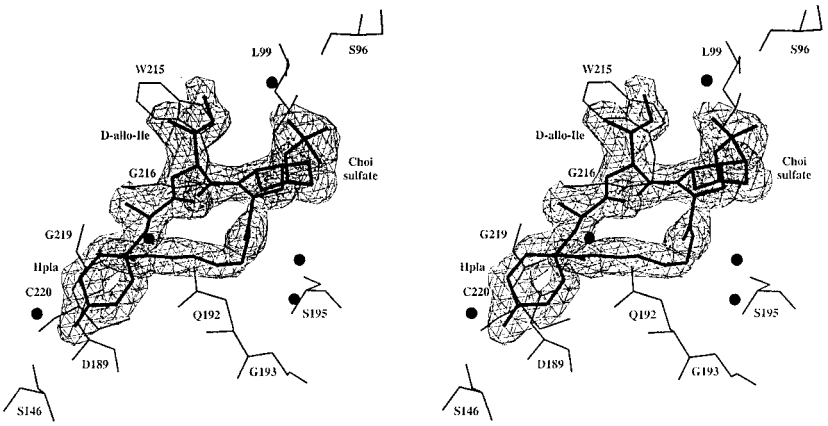


Figure 1. A stereoview of the difference ($2|F_o| - |F_c|$) electron density map of the complex. The map is contoured at 1.0 σ, and the final model of aeruginosin 98-B is shown in bold lines.

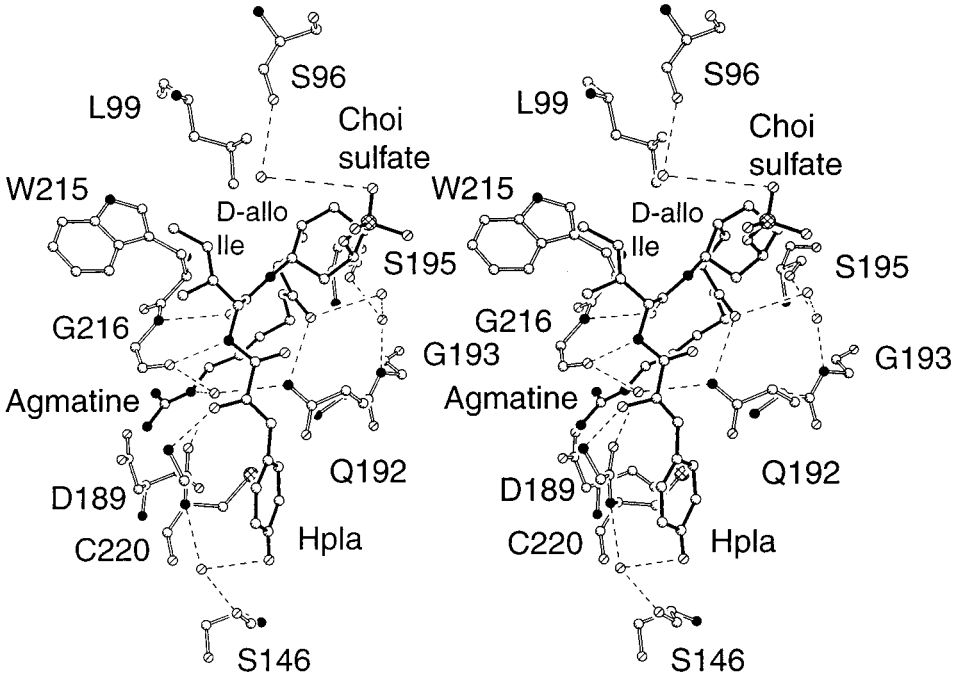


Figure 2. A stereoview of aeruginosin 98-B bound to trypsin. Aeruginosin bonds are filled, trypsin bonds are unfilled, and possible hydrogen bonds are dotted lines. Selected active site residues are indicated.

The D-alloIle has an unusual eclipsed conformation about the Cα−Cβ bond, but this conformation, which avoids contact with a neighboring trypsin, is likely a crystal packing effect. The sulfate moiety projects into solvent, but in the corresponding thrombin complex it would be in a hydrophobic pocket formed by Tyr60A and Trp60D. The sulfate group may be a primary determinant of aeruginosin 98-B's marked preference for trypsin.

Aeruginosin 98-B also interacts with trypsin in unconventional ways. At the very end of bound aeruginosin 98-B, there is a water-mediated interaction among the phenolic hydroxyl of Hpla, the amide NH of Cys220, and the carbonyl oxygen of Ser146 (Figure 2). The water binding these three residues together is seen as a tightly bound water in the structure of unliganded trypsin,[11] suggesting that aeruginosin 98-B exploits this water to extend its binding interactions.[12] The carbonyl group of Choi sulfate interacts with the side chain amide NH of Gln192, which is positioned by the tightly held water shared by the agmatine, Gln192, Gly216 and Gly219, all of which play other roles in binding. The sulfate moiety of Choi sulfate interacts with surface water molecules.

The most conspicuous absences in the inventory of aeruginosin 98-B's interactions with trypsin are Ser195, His57, and Asp102—the catalytic triad of serine peptidases. Aeruginosin 98-B binds exclusively to specificity and recognition elements in an unprecedented fashion for a natural product. The nucleophilic oxygen of Ser195 makes no close contacts with any atom of aeruginosin 98-B (Figure 2). Aeruginosin 98-B's inhibitory mode most closely resembles that of D-Phe-Pro-agmatine.[13] The current study illustrates the ability of structural studies on protein−ligand complexes to elucidate the stereochemical and conformational details of the ligand, clarifying its inhibitory mode, and suggesting new inhibitory motifs.

Acknowledgment. This work was supported by the NIH through CA24487 (J.C.) and Training Grant GM08267 in Molecular Biophysics to B.S.

Supporting Information Available: A Ramachandran plot (PROCHECK) and a list of all trypsin-1 interactions (LIGPLOT) (2 pages). See any current masthead page for ordering and Internet access instructions.

JA972991V

(11) Protein Data Bank ID 3PTN. Fehlhammer, H.; Bode, W. *J. Mol. Biol.* **1975**, *98*, 683−692. Bode, W.; Schwager. P. *J. Mol. Biol.* **1975,** *98*, 693−717.

(12) Ladbury, J. E. *Chem. Biol.* **1996,** *3*, 973−980.

(13) Wiley, M. R.; et al. *Bioorg. Med. Chem. Lett.* **1995,** *5*, 2835−2840.