Atomic Structure of the Trypsin–Cyclotheonamide A Complex: Lessons for the Design of Serine Protease Inhibitors

Angela Y. Lee,† Masahiko Hagihara,‡,§
Rakesh Karmacharya,‡ Mark W. Albers,‡
Stuart L. Schreiber,*,‡ and Jon Clardy*,†

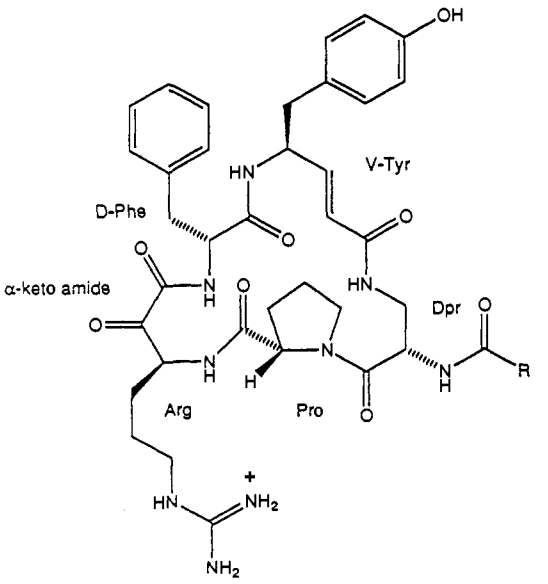
Departments of Chemistry
Cornell University, Baker Laboratory
Ithaca, New York 14853-1301
Harvard University, 12 Oxford Street
Cambridge, Massachusetts 02138

Received September 1, 1993

Studies on marine sponges of the genus *Theonella* have rewarded natural products chemists with an array of structurally novel, biologically active metabolites. Fusetani's group recently expanded *Theonella's* repertoire of natural products with the isolation of cyclotheonamides A and B—novel cyclic peptides that are potent serine protease inhibitors.[1] Following the initial isolation and characterization, a total synthesis firmly established the structures of the cyclotheonamides.[2] This bioassay-guided isolation and synthesis-assisted characterization of the cyclotheonamides completed the traditional enterprise of natural products chemistry. However, recent advances in structural chemistry and molecular biology have greatly expanded our ability to study the interactions between natural products and their biological receptors, and these studies have provided a new and fruitful approach to natural products chemistry.[3] As an example of this approach, we describe a detailed study of cyclotheonamide A bound to a serine protease, bovine β-trypsin.

Cyclotheonamides (CyA and CyB, see **1** and **2** and Figure 1) first attracted our attention because they contained two structural elements that we had studied in very different contexts: a vinylogous tyrosine (V-Tyr) and an α-keto amide. We explored vinylogous amino acids as peptidomimetics with predictable secondary structures[4] and studied α-keto amides as key structural features for complexing the immunosuppressive agents FK506 and rapamycin to FKBP.[5] We were curious to see how these structural elements, as well as the other features of cyclotheonamide A (CyA), were used in binding interactions with its receptor(s). CyA inhibits serine proteases such as human thrombin, bovine trypsin, and human two-chain tissue plasminogen activator with K_i values of 1.0, 0.2, and 40 nM, respectively.[6] Inhibition by CyA was also slow, with second-order rate constants for complex formation of 4.6×10^4, 4.8×10^4, 2.7×10^2 M^{-1} s^{-1}, respectively, for the same proteases.[6] We elected to study CyA complexed to bovine β trypsin for two reasons: this complex has the tightest binding, and trypsin is arguably the best characterized of all the serine proteases. The three-dimensional structure of the complex was revealed by a 2.0-Å resolution X-ray diffraction study.

Single crystals of the complex, grown by the hanging-drop vapor-diffusion method,[7] belong to space group $P4_32_12$ with $a = b = 71.73$ and $c = 88.82$ Å. X-ray data were collected on a



1 R = H Cyclotheonamide A (CyA)
2 R = Me Cyclotheonamide B (CyB)

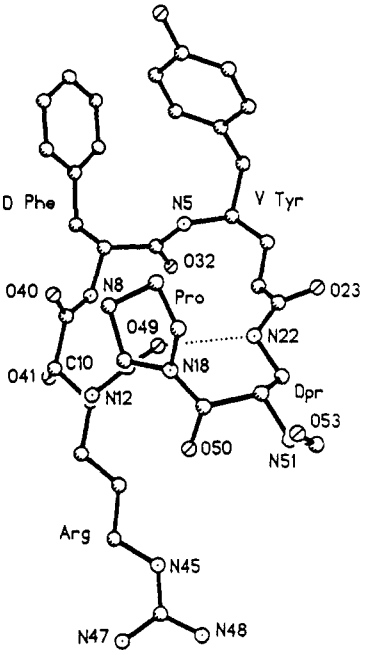


Figure 1. Two different renderings of cyclotheonamide A (CyA). On the top is a conventional structural drawing, and on the bottom is a perspective drawing of the conformation bound to bovine β-trypsin. Note that the α-keto amide is part of a hemiketal with Ser 195 O_γ (not shown) on the right.

SDMS Mark II area detector coupled to a Rigaku rotating anode X-ray generator.[8] The structure was solved using the X-PLOR[9] implementation of the molecular replacement technique with the coordinates of bovine β-trypsin as the search model.[10] Refinement used the X-PLOR slow-cooling simulated annealing and least-squares energy refinement technique.[11] The final model contains

† Cornell University.
‡ Harvard University.
§ Present address: Ube Industries, Ltd., Japan.

(1) Fusetani, N.; Matsunaga, S.; Matsumoto, H.; Tukebayashi, Y. *J. Am. Chem. Soc.* **1990**, *112*, 7053–7054.
(2) Hagihara, M.; Schreiber, S. L. *J. Am. Chem. Soc.* **1992**, *114*, 6570–6571.
(3) Rosen, M. K.; Schreiber, S. L. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 384–400.
(4) Hagihara, M.; Anthony, N. J.; Stout, T. J.; Clardy, J.; Schreiber, S. L. *J. Am. Chem. Soc.* **1992**, *114*, 6568–6570.
(5) (a) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *252*, 839–842. (b) Van Duyne, G. D.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. *J. Am. Chem. Soc.* **1991**, *113*, 7433–7434. (c) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *J. Mol. Biol.* **1993**, *229*, 105–124.
(6) Lewis, S. D.; Ng, A. S.; Baldwin, J. J.; Fusetani, N.; Naylor, A. M.; Shafer, J. A. *Thromb. Res.* **1993**, *70*, 173–190.
(7) Ten-microliter droplets containing an initial protein concentration of 15 mg/mL in 45 mM Tris, pH 7.0, buffer were equilibrated against 0.7 mL of 30% PEG 4000 and 0.2 M sodium citrate buffered with 0.1 M Tris at 4 °C. Crystals of $0.3 \times 0.4 \times 0.6$ mm^3 appeared in 1–2 months.
(8) Howard, A. D.; Nielson, C.; Zuong, N. H. In *Methods in Enzymology, Volume 114*; Wyckoff, H. W., Hirs, C. H. W., Timasheff, S. N., Eds.; Academic Press: San Diego, CA, 1985; pp 452–472. The R_{sym} was 5.0 for 16 217 unique reflections (92.2% completeness).

0002-7863/93/1515-12619$04.00/0

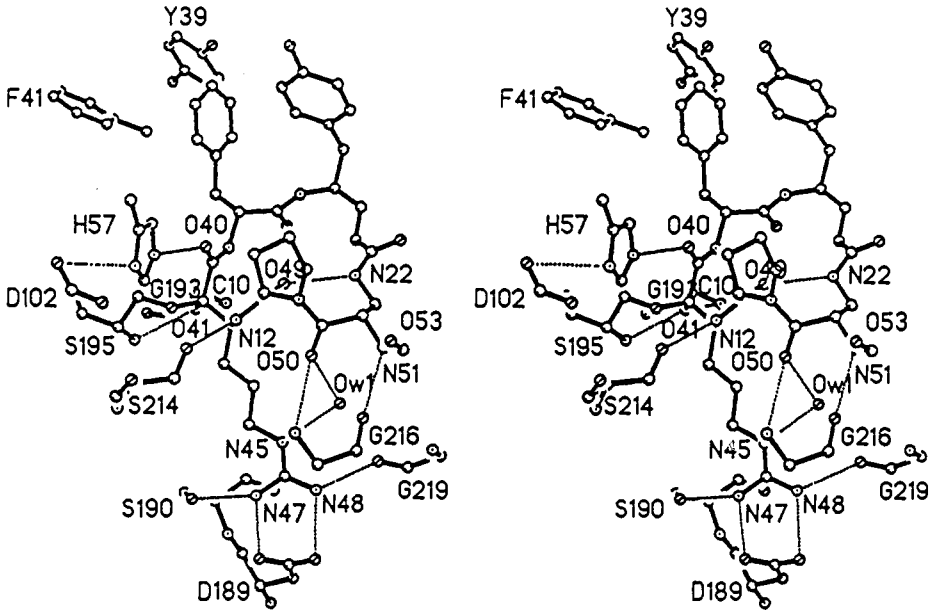


Figure 2. Stereoscopic drawing of cyclotheonamide A in the active site of bovine β-trypsin.

the protein, CyA, and 128 solvent molecules. In the final ($2|F_o| - |F_c|$) map, the electron density is well defined for all atoms except Asn 115, Ser 116, and Arg 117. The final model has an R-factor of 17.0 for the 16 217 reflections greater than 2σ in the resolution range 8.00–2.0 Å. The root mean square deviations for bond lengths and angles are 0.010 Å and 1.7°, respectively.

A stereoview of CyA bound in the active site of trypsin is shown in Figure 2. Most of CyA, the portion from Cα of diaminopropionic acid (Dpr) to the amide of D-Phe, lies along the active site groove in an extended substrate-like fashion; the rest of CyA, the upper right hand portion, is exposed to solvent. Trypsin's catalytic triad—Ser 195, His 57, and Asp 102—is on the left-hand side of Figure 2. Well-defined continuous electron density between Ser 195 O_γ and the ketone carbon of the α-keto amide strongly indicates the formation of a covalent complex through the formation of a tetrahedral hemiketal (O_γ–C10, 1.41 Å). The resulting oxygen accepts hydrogen bonds from the main-chain NH groups of Gly 193 and Ser 195 in the oxyanion binding pocket. There is also a strong hydrogen bond between the amide carbonyl of the α-keto amide and the imidazolium of His 57—the residue that deprotonates Ser 195 (O40–H57NH, 2.68 Å).

Other structural features of CyA display exquisite complementarity to the active site of trypsin. The phenyl ring of D-Phe is in a hydrophobic groove composed of Tyr 39 and Phe 41 at the top of Figure 2. The face of the phenyl ring of D-Phe is oriented toward the edges of Tyr 39, Phe 41, and the phenol ring of V-Tyr. Note that L-Phe would not fit into this binding pocket. The amide proton of Arg (N12) forms a hydrogen bond with the main-chain carbonyl of Ser 214, while its side chain extends into the deep, narrow S_1 specificity pocket at the bottom of Figure 2. The guanidinium forms hydrogen bonds to the carboxylate of Asp 189, the carbonyl of Gly 219, and the O_γ of Ser 190. Finally, there is a hydrogen bond between the carbonyl of Dpr (O50) and the NH of Gly 216 and a water-mediated (Ow1) hydrogen bond between the same carbonyl and guanidinium N45. V-Tyr is oriented away from the active site and exposed to solvent. A strong (N22–O49, 2.79 Å, see Figure 1) intramolecular hydrogen bond forms between Dpr amide N–H (N22) and the proline carbonyl (O49). The Pro exists as the *trans* rotamer in bound CyA.

This structural work provides important clues for the design of synthetic analogs that will elucidate the molecular basis of the observed slow and tight binding behavior.[12] For example, one such analog that has been synthesized and is currently under investigation is dihydro-CyB, an analog where the ketone carbonyl (C10) has been reduced to an alcohol, and results of this study will be reported separately. An independent study has provided similar insights into the complex formed between CyA and thrombin.[13] These two complementary studies should facilitate the design of new serine protease inhibitors exhibiting higher specificity. More generally, they highlight the exciting possibilities offered by studying novel natural products complexed to their macromolecular receptors.

Acknowledgment. This work was partially supported by NIH grants CA24487 (J.C.), GM44993 (S.L.S.), NIH Training Grant GM08267 in Molecular Biophysics to A.Y.L. and an HHMI Predoctoral Fellowship to M.W.A. M.H. thanks Ube Industries, Ltd. for support as a Visiting Scientist. We also thank A. Tulinsky and B. Maryanoff for sharing the results of ref 13 and J. Shafer for providing the results of ref 6 prior to publication.

(9) (a) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458. (b) Brünger, A. T. *X-PLOR* Version 3.1; Yale University: New Haven, CT, 1992. The coordinates for bovine β-trypsin were placed in a $P\bar{1}$ unit cell with dimensions $a = b = c = 100$ Å and $\alpha = \beta = \gamma = 90°$. Triclinic structure factors calculated within the 4–8-Å resolution range were used for a rotation search with a 25-Å Patterson radius. PC-refinement of the largest peaks in the rotation search gave a clear solution. The translation search was broken into two individual translation searches, corresponding to the alternative space groups $P4_32_12$ or $P4_12_12$, of this best solution using 8–4.0-Å data. A single unambiguous solution for space group $P4_32_12$ was obtained. Forty cycles of X-PLOR rigid-body refinement dropped the R-factor from 38.0 to 33.6.

(10) Protein Data Bank code 1TLD as described in Bartunik, H. D.; Summers, L. J.; Bartsch, H. H. *J. Mol. Biol.* **1989**, *210*, 813. See also: Huber, R.; Bode, W. *Acc. Chem. Res.* **1978**, *11*, 114–122. Finer-Moore, J. S.; Kossiakoff, A. A.; Harley, J. H.; Earnest, T.; Stroud, R. M. *Proteins* **1992**, *12*(3), 203–22.

(11) Model building used CHAIN. Sack, J. S. *J. Mol. Graphics* **1988**, *6*, 224–225. An initial $2|F_o| - |F_c|$ map at 2.5-Å resolution was initially used for manual adjustment of the protein model. The inhibitor, CtA, was then added to the active site cleft to account for positive peaks in an $|F_o| = |F_c|$ map. The resolution was gradually increased to 2.0 Å. In the final stages of refinement, difference electron density maps were used to locate water molecules. Restrained, individual temperature factors were refined for each residue. Average B-values were 14.63, 19.24, and 29.25 Å^2 for trypsin, CtA, and water, respectively. Coordinates have been deposited in the Protein Data Bank.

(12) Both CyA and CyB are slow-binding inhibitors of human thrombin (in 50 mM Tris, pH 7.5, 100 mM NaCl, 0.1% PEG-8000, 0.1 nM purified human thrombin, 8 μM *N-p*-tosyl-Gly-Pro-Arg-pNA (Sigma) at 37 °C and in 50 mM Tris, pH 7.5, 100 mM NaCl, 0.1 nM purified human thrombin, 100 μM H-D-Phe-Pip-Arg-pNA (Chromogenix) at 37 °C) and bovine trypsin (in 50 mM Tris, pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$, 0.1% PEG-8000, 0.4 nM purified bovine trypsin, 20 μM Sar-PR-pNA (Sigma) at 25 °C) (R.K., M.W.A., and S.L.S., unpublished results). Second-order rate constants for the slow binding of CyA to several trypsin-like serine proteases including thrombin, trypsin, plasmin, t-PA, and factor Xa have been reported by Lewis et al.[6]

(13) Maryanoff, B. E.; Qiu, X.; Padmanabhan, K. P.; Tulinsky, A.; Almond, H. R.; Andrade-Gordon, 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.