Structure of an Acyl-Enzyme Intermediate during Catalysis: (Guanidinobenzoyl)trypsin[†,‡]

Walter F. Mangel, Paul T. Singer, Donna M. Cyr, Timothy C. Umland, Diana L. Toledo, Robert M. Stroud,[§] James W. Pflugrath,[∥] and Robert M. Sweet*

Biology Department, Brookhaven National Laboratory, Upton, New York 11973, S-964 Department of Biochemistry, University of California, San Francisco, California 94143-0448, and Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Received January 12, 1990; Revised Manuscript Received May 18, 1990

ABSTRACT: The crystal and molecular structure of trypsin at a transiently stable intermediate step during catalysis has been determined by X-ray diffraction methods. Bovine trypsin cleaved the substrate *p*-nitrophenyl *p*-guanidinobenzoate during crystallization under conditions in which the acyl-enzyme intermediate, (guanidinobenzoyl)trypsin, was stable. Orthorhombic crystals formed in space group $P2_12_12_1$, with a = 63.74, b = 63.54, and c = 68.93 Å. This is a crystal form of bovine trypsin for which a molecular structure has not been reported. Diffraction data were measured with a FAST (Enraf Nonius) diffractometer. The structure was refined to a crystallographic residual of R = 0.16 for data in the resolution range 7.0–2.0 Å. The refined model of (guanidinobenzoyl)trypsin provides insight into the structural basis for its slow rate of deacylation, which in solution at 25 °C and pH 7.4 exhibits a $t_{1/2}$ of 12 h. In addition to the rotation of the Ser-195 hydroxyl away from His-157, C_β of Ser-195 moves 0.7 Å toward Asp-189 at the bottom of the active site, with respect to the native structure. This allows formation of energetically favorable H bonds and an ion pair between the carboxylate of Asp-189 and the guanidino group of the substrate. This movement is dictated by the rigidity of the aromatic ring in guanidinobenzoate—model-building indicates that this should not occur when arginine, with its more flexible aliphatic backbone, forms the ester bond with Ser-195. As a consequence, highly ordered water molecules in the active site are no longer close enough to the scissile ester bond to serve as potential nucleophiles for hydrolysis. Coupled with an apparent 35% decrease in the overall temperature factor of the acyl-enzyme relative to the native structure, the tight packing and rigidity of all atoms in the active site, including solvent, prevent disordered water molecules from easily approaching the carbonyl carbon atom via diffusion.

The molecular mechanism of enzyme action is a subject of fundamental interest to researchers in a variety of fields. Structural details of catalysis not only provide insight into the physical origin of function but also serve as a rational basis for the design of improved enzymes and inhibitors by genetic engineering and chemistry. The structures of serine proteases and in particular trypsin have been extensively studied from these points of view. However, crucial details of the catalytic mechanism, especially a time-resolved description of the intermediate steps, are still missing.

Trypsin is a typical member of the serine protease family. It consists of 223 amino acid residues and an integral, octahedrally coordinated Ca^{2+} cation that serves a structural role (Bode & Schwager, 1975; Kossiakoff et al., 1977). At the active site lies a particularly reactive serine residue, Ser-195, using the residue numbering scheme for chymotrypsinogen (Hartley, 1964; Hartley & Kauffman, 1966). The residue's unusual reactivity is owed primarily to its interaction with a nearby histidine residue that is itself oriented and polarized by a buried aspartate residue (Blow et al., 1969). These three residues are conserved among all members of the family and are referred to collectively as the "catalytic triad". The net effect of this constellation of residues is that the imidazole of the histidine residue serves as a strong general base to promote nucleophilic attack by the serine oxygen atom on the electron-deficient carbonyl carbon atom of the scissile peptide bond. This results in the transfer of the acyl group from the peptide nitrogen to the serine hydroxyl and hence the formation of the acyl-enzyme intermediate.

Deacylation is similar in mechanism to acylation. Following formation of the acyl-enzyme intermediate, the newly formed amino terminus diffuses away from the active site. This allows a solvent water molecule to approach and be activated by His-57 in essentially the same manner as was the hydroxyl of Ser-195 during the acylation reaction step. This nucleophilic water molecule is positioned to hydrolyze the ester bond. Once this occurs, the second product of proteolysis is released from the enzyme, and the active enzyme is regenerated. The reversibility of this reaction is demonstrated in the case of the complex between trypsin and soybean trypsin inhibitor (STI).[1] The portion of the polypeptide that would normally be the leaving group in STI is held rigidly in place by the intimate

[†] Support for this work was provided by the OHER of the U.S. Department of Energy; NIH Grant GM24485, to R. M. Stroud; the Department of Energy's Division of University and Industry Programs, Office of Energy Research, Science and Engineering Research Semester Program, to T. C. Umland; and an Alexander Hollaender Distinguished Postdoctoral Fellowship sponsored by OHER and administered by Oak Ridge Associated Universities, to P. T. Singer.

[‡] Atomic coordinates for the refined (guanidinobenzoyl)trypsin structure have been submitted to the Brookhaven Protein Data Bank.

* To whom correspondence should be addressed.

[§] University of California.

[∥] Cold Spring Harbor Laboratory.

[1] Abbreviations: DIP-trypsin, diisopropylfluorophosphate inhibited trypsin; FMGB, fluorescein monoguanidinobenzoate; GBA, guanidinobenzoic acid; GB-trypsin, (guanidinobenzoyl)trypsin; NPGB, *p*-nitrophenyl *p'*-guanidinobenzoate; SDS, sodium dodecyl sulfate; STI, soybean trypsin inhibitor.

0006-2960/90/0429-8351$02.50/0

fit between the two proteins and does not diffuse away (Sweet et al., 1974). As a consequence, a water molecule cannot easily approach the active site to hydrolyze the acyl-enzyme. In addition to catalyzing hydrolysis of a peptide, the enzyme can activate the terminal amino group itself by this same mechanism, and the peptide bond can be reformed. When the enzyme–inhibitor complex is allowed to reach equilibrium, approximately equal quantities of the STI are found to be cleaved and uncleaved (Finkenstadt & Laskowski, 1965).

In this study, we report on the structure of trypsin at a transiently stable intermediate step during catalysis—the acyl-enzyme intermediate. When trypsin reacts with the substrate *p*-nitrophenyl *p*-guanidinobenzoate (NPGB), the rate-limiting step is not acylation but rather deacylation (Chase & Shaw, 1967). At 25 °C and pH 7.4, the half-time for deacylation is 12 h. As a consequence, this substrate is a useful active-site titrant. The slow rate of deacylation and fast data collection allowed us to crystallize the acyl-enzyme intermediate and quickly collect diffraction data. This work was performed in part for the information it would give about the details of this transient step in the catalytic pathway but also for the possibilities that might arise with this system for further dynamic studies of enzyme mechanism.

Three-dimensional structures of many trypsin-like enzymes have been determined by crystallographic methods; 38 structures have been deposited in the Brookhaven Protein Data Bank (Bernstein et al., 1977), including 7 of bovine trypsin and 12 of its zymogen. Most of these are examples of proteases derivatized to prevent autolysis during crystallization. While many of the derivatives mimic intermediates of the catalytic mechanism, most are not true intermediates themselves. The GB-trypsin employed in this study, however, is electronically identical with actual acyl-enzyme complexes formed during physiological proteolysis in that it contains a true ester bond. In this sense it is similar to the (indolylacryloyl)-α-chymotrypsin structure (Henderson, 1970). These two structures may mimic a true catalytic intermediate better than do other protease derivative structures that are available: a phosphoester (DIP) of trypsin (Chambers & Stroud, 1979) and of chymotrypsin (Sigler et al., 1966), a sulfonyl ester (tosyl) of chymotrypsin (Matthews et al., 1967), boronic esters of subtilisin (Matthews et al., 1975) and of α-lytic protease (Bone et al., 1987), and a hemiketal (APPA) of trypsin (Walter & Bode, 1983). Dixon and Matthews (1989) have recently reported a complex in which bovine γ-chymotrypsin is apparently acylated by a tetrapeptide tentatively identified as Pro-Gly-Val-Tyr. The coordinates are not yet available from the Protein Data Bank, and therefore the structure could not be compared directly to that of GB-trypsin reported here.

Experimental Procedures

The work reported here represents two structural studies separated in time by several years. For reasons not germane to the objectives of this article, Stroud and co-workers at the University of California at San Francisco obtained the new crystal form of benzamidine-inhibited trypsin and determined its structure, as described below. More recently, the acyl-enzyme intermediate, (guanidinobenzoyl)trypsin, was crystallized at Brookhaven National Laboratory. The refinement of its structure was begun by direct use of the coordinates from the structure solved by Stroud and co-workers. A difference map was generated from data obtained from crystals of native trypsin and of (guanidinobenzoyl)trypsin, both grown at Brookhaven National Laboratory.

Structure Determination of Native Trypsin in Orthorhombic Crystals. Crystals of native trypsin with benzamidine at the active site were grown following the procedures of Bode and Schwager (1975) except that the conditions for crystallization were 1.4 M $(NH_4)_2SO_4$ and pH 7.5. After 2 months, the orthorhombic crystals had grown to 0.15 × 0.15 × 0.33 mm; these were then used as seeds for a second vapor-diffusion crystallization. The unit cell dimensions were different from those reported earlier for trypsin crystallized from magnesium sulfate (Chambers & Stroud, 1979). This new form was also of space group $P2_12_12_1$ but with larger unit cell dimensions: $a = 63.7$, $b = 63.5$, $c = 68.9$ Å. In order to solve the structure in this new crystal form, data from four separate crystals were collected by an automatic diffractometer to 1.8-Å resolution. The structure was solved by rotation and translation of the benzamidine/trypsin structure that had been solved for crystals grown in magnesium sulfate (Krieger et al., 1974). The molecular replacement solution was obtained by means of the Crowther rotation function (Crowther, 1972) applied to data from 12- to 6-Å resolution by using a radius of integration of 25 Å, followed by a translation search based on the R factor using 9–5.5-Å data. From the phases of the model that resulted from the translation/rotation search, an electron density map was calculated. That benzamidine could be located provided evidence the solution was correct. At this stage the residual $R = \sum||F_o| - |F_c||/\sum|F_o|$ was 0.44.

The model was then refined by rigid-body least squares using data from ∞ to 4 Å, which gave no significant improvement. The initial residual for data out to 3-Å resolution was 0.41, which was refined by difference Fourier methods to 0.28 (Chambers & Stroud, 1977). Inclusion of data to 2.6 Å was accompanied by further difference Fourier refinement to $R = 0.23$. Inclusion of data to 2.2 Å led to $R = 0.27$ initially, which was refined to 0.23. Finally, inclusion of 1.8-Å resolution data gave an initial residual of 0.30, which was reduced by more refinement to $R = 0.19$.

The overall merging R factor, $R_1 = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, for data from four data sets was 0.061. The final R factor for 24 251 reflections between 12.0- and 1.8-Å resolution was 0.191; the average deviations in bond lengths, bond angles, and torsional angles from ideal values were 0.033 Å, 2.09°, and 5.01°, respectively.

Preparation of (Guanidinobenzoyl)trypsin Crystals. Bovine trypsin (type III) and NPGB were purchased from Sigma Chemical Co. Only lots that contained greater than 95% pure β-trypsin, as judged by SDS–polyacrylamide gel electrophoresis under reducing conditions (Laemmli, 1972), were used. The concentration of active trypsin was determined by an active-site titration with fluorescein monoguanidinobenzoate, FMGB (Melhado et al., 1982).

Crystals of GB-trypsin were grown by the hanging-drop vapor-diffusion method. Lyophilized trypsin was dissolved to a concentration of 60 mg/mL in 0.05 M Tris-HCl (pH 8.15), 0.3 M $MgSO_4$, 0.006 M $CaCl_2$, and 0.06 M benzamidine. Insoluble material was removed by centrifugation, and the active enzyme concentration was adjusted to 1.3 mM (30 mg/mL) as determined by an active-site titration of an aliquot with FMGB. The equilibration well buffer was 0.05 M Tris-HCl (pH 8.5) with 1.6–1.8 M $MgSO_4$. Prior to sealing of the wells, 0.5 μL of 0.1 M NPGB in dimethylformamide was added to the 5-μL solution of enzyme. A fine precipitate of NPGB formed throughout the droplet. The pH in the hanging droplet just prior to crystallization was measured by pH paper to be approximately 5.5. Orthorhombic crystals usually formed within 7 days at 22 °C, often with dimensions of 0.5–1.0 mm. Crystals of benzamidine-inhibited trypsin were also prepared by using this protocol except that NPGB was

not added. These crystals grew within 4 days. Isomorphous crystals could be obtained with similar solutions by substituting $(NH_4)_2SO_4$ for $MgSO_4$.

Crystals were stabilized and benzamidine and/or NPGB were removed from the mother liquor by stepwise, gradual transfer of the crystals to a synthetic mother liquor of 0.05 M citrate (pH 5.0), 2.5 M $MgSO_4$ and 0.02 M $CaCl_2$. This lower pH was used to decrease the rate of deacylation, which under these conditions exhibited a $t_{1/2}$ of 36 h.

The extent of acylation in single crystals was measured in a two-part assay. A crystal in synthetic mother liquor lacking NPGB was dissolved in a small volume of 0.001 M HCl. An aliquot was immediately titrated with FMGB to determine the concentration of active enzyme. Another aliquot was placed in a solution of 0.05 M Tris-HCl (pH 8.5) containing 0.001 M benzamidine. After 2 h at 40 °C, complete deacylation occurred, and the concentration of active sites was remeasured. For newly formed crystals, the extent of acylation was typically >95%.

Diffraction Data Collection. X-ray diffraction data were collected on a FAST (Enraf Nonius) area diffractometer at Cold Spring Harbor Laboratory. The Elliott GX-21 rotating anode generator, fitted with a graphite crystal monochromator, was operated at 40 kV and 90 mA. Frames of 0.1° crystal rotation were recorded every 100–120 s. Intensities were integrated by the computer program MADNES (Pflugrath & Messerschmidt, 1985; Messerschmidt & Pflugrath, 1987), and multiple measurements were scaled by the program suite PROTEIN (Steigemann, 1974a,b). Data were collected from two different crystals—native trypsin and acylated, (guanidinobenzoyl)trypsin. Unit cell parameters were calculated by the program MADNES during the periodic realignments of the crystal. Mean values for approximately 40 measurements of the cell parameters for the acyl-enzyme crystal were a = 63.74, b = 63.54, and c = 68.93 Å. Standard deviations of the mean values were all 0.02 Å. Data were measured with approximately 2-fold redundancy. The value of R_1 was 0.092 for GB-trypsin and 0.073 for native trypsin.

Crystallographic Refinement of GB-Trypsin. The starting point for the refinement was the model of native, benzamidine-inhibited trypsin obtained from isomorphous crystals as described above. The native model fit the acyl-enzyme data set to a crystallographic R factor of 0.27. A difference Fourier map clearly showed density in the active site consistent with acylation of Ser-195. The model was refined in cycles, each consisting of several iterations of R-factor refinement and several iterations of B-factor refinement. Following inspection of a difference Fourier map, inconsistencies between the model and the maps were corrected manually, and new solvent molecules were assigned. After four cycles of refinement, residue 195 was changed from Ser to GB-Ser. The additional atoms representing the guanidinobenzoyl moiety were manually placed into the electron density at the active site. Toward completion of refinement the fractional occupancy was refined for solvent atoms, for atoms of the guanidinobenzoyl group, and for atoms in hydrophilic side chains for which the electron density was poor. The B-factor refinement was restrained to require covalently attached atoms to have similar thermal parameters. For each of these least-squares refinements, 12934 reflections in the resolution range 7.0–2.0 Å were used. These data represented approximately 50% of all reflections possible.

The manual adjustments to the model were performed by using the molecular graphics program PS-FRODO (Jones, 1978; Pflugrath et al., 1984) on an Evans and Sutherland PS 300 graphics workstation controlled by a DEC VAX 11/730. The combined crystallographic and molecular geometry refinement of the model was performed by using the TNT program suite (Tronrud et al., 1987) on a VAX 11/780. Idealized molecular parameters for the guanidinobenzoyl group were chosen from fragments of small-molecule crystal structures. Several sets of atoms in this group were restrained to lie in common planes: the eight atoms of the benzene ring along with its para-substituted atoms and the five atoms of the guanidino group plus its attachment to the benzene ring.

To ensure that the geometry of the atoms closely associated with Ser-195 was dictated by the diffraction amplitudes rather than by restrained bond lengths and angles, we relaxed substantially the geometric restraints in this region. This was particularly important around the acyl carbon, to allow for any tetrahedral character of this atom in the catalytic intermediate. That this was appropriate was indicated by the $2F_o - F_c$ difference Fourier maps, where initial density peaks near the carbonyl oxygen atom suggested that its true position was out of the plane defined by the other three atoms comprising the carboxyl moiety. Also, the lengths of the ester bond and the carbonyl double bond were essentially unconstrained.

Results

The extent of acylation of control crystals, grown simultaneously in the same well as the diffracting crystal, was >95% at the start of diffraction data collection. After 36 h of exposure to the X-ray beam, an assay of the diffracting crystal revealed the extent of acylation to be only ~60%. The electron density associated with the guanidinobenzoyl moiety was lower than that associated with other parts of the molecule, consistent with gradual deacylation occurring during data collection. In order to account for this loss of electron density in the model, the fractional occupancies of the atoms comprising the guanidinobenzoate moiety were also fit to the crystallographic data. The occupancies of the carbon atoms ranged from 0.7 to 0.9, while the three guanidino nitrogen atoms were close to full occupancy. This result is reasonable: in the absence of guanidinobenzoate, water molecules would be expected to enter the binding cleft and to occupy positions similar to those of the guanidino nitrogen atoms, [Brookhaven Protein Data Bank (Bernstein et al., 1977) entry code 1TPO (Marquart et al., 1983)]. Visual examination of the model shows that all three guanidino nitrogen atoms form H bonds both to the protein and to additional ordered water molecules.

After 11 cycles of refinement, the atomic model fit the crystallographic data extremely well. The final crystallographic R factor was 0.16 (7–2 Å). The rms deviations from ideal bond lengths and angles were 0.018 Å and 3.6°, respectively. The rms difference of temperature factors of atoms bonded to one another was 1.4 Å^2. Treatment of the R-factor results by the method of Luzatti (1952) revealed that the likely error in atomic positions was 0.18 Å. In the final $2F_o - F_c$ Fourier map, the minimum and maximum values were 3.1 and −2.1 e/Å^3, respectively, while the standard deviation was 0.4 e/Å^3. In the $F_o - F_c$ map, the largest peaks were ±0.3 e/Å^3, with a standard deviation of 0.07 e/Å^3. The small number of peaks in this map of greater than four standard deviations in magnitude were uninterpretable in terms of movements of one or more atoms or side chains. In addition to the 223 amino acid residues of trypsin, the single calcium ion, two sulfate ions, and 95 water molecules were assigned, although the fractional occupancies of several of the water molecules were as low as 0.4. All assigned solvent molecules were plausibly bonded to either other solvent molecules or polar sites in the protein or both. The exact conformations of the side chains of several

Table I: Comparison of Structure of GB-Trypsin to Published Structures[a]

	five trypsins	2PTC	GBT
five trypsins	0.1	0.25	0.22
2PTC			0.22

[a] The coordinates of the α-carbon atoms of GB-trypsin were rotated by using the program ROTMOL (Steigemann, 1974a,b) to fit those of six other bovine trypsin models found in the Brookhaven Protein Data Bank (Bernstein et al., 1977). The root-mean-square deviation (in angstroms) between the different pairs of structures is given after rejection of outliers. Outliers were assigned according to one of two criteria: either error in fitting of the atom by $>1.5\sigma$ or misfitting of other, nearby residues. The entry "five trypsins" refers to the entries 1TPO, 1TPP, 2PTN, 3PTB, and 4PTP, all of which deviate from one another in C_α position by 0.1 Å, as shown. The structure labeled GBT is (*p*-guanidinobenzoyl)trypsin determined in this study. The Brookhaven Protein Data Bank entry codes are as follows: 1TPO, native trypsin at pH 5.0 (Marquart et al., 1983); 1TPP, trypsin inhibited by (*p*-amidinophenyl)pyruvate (Walter & Bode, 1983); 2PTC, complex of trypsin with pancreatic trypsin inhibitor (Huber et al., 1974); 2PTN, native trypsin grown in ammonium sulfate (Marquart et al., 1983); 3PTB, benzamidine-inhibited trypsin (Bode & Schwager, 1975); 4PTP, (diisopropylphosphoryl)trypsin (Chambers & Stroud, 1977).

surface-charged residues, particularly lysines, could not be determined owing to poor electron density. The occupancies of these side chains were refined, in some cases to zero.

The atoms involved in the ester link between Ser-195 and GBA were refined with only weak geometric restraints, and indeed they deviate somewhat from idealized molecular parameters. The ideal length for the C_β–O_γ bond is 1.42 Å; in the model it is longer—1.58 Å. A prototypical benzoyl ester bond is 1.38 Å, a carbonyl double bond 1.18 Å. In the final model, these distances are found to be 1.30 and 1.24 Å, respectively. Finally, the carbonyl oxygen atom is clearly out of the plane of the carboxylate moiety, lying 0.2 Å away from the plane of the other three atoms and toward the peptide nitrogen atoms of residues 193 and 195. To test the significance of these deviations we performed refinement in which the same weak constraints were applied to all main-chain peptide groups as were applied to the ester group of GB-serine. Their mean bond lengths and standard deviations were calculated as were deviations of the carbonyl oxygen atoms from the peptide plane. The C_β–O_γ bond length deviation and carbonyl oxygen atom deviation from planarity were both twice these standard deviations.

Discussion

Comparison to Other Structures. The model of guanidinobenzoyl bovine trypsin presented here was compared to others in the Brookhaven Protein Data Bank (Bernstein et al., 1977) at the level of the α-carbon atoms, Table I. In general, GB-trypsin could be aligned with each of the other structures with a root mean square (rms) deviation of ~0.2 Å, after rejection of outliers. By comparison, the estimated error in the coordinates derived from a Luzatti plot was similar: 0.18 Å. The outliers, which deviated by as much as 0.8–0.9 Å, may represent structural changes associated with the formation of the acyl-enzyme.

The GB-trypsin structure deviated markedly from the other structures in the regions around residues 37–39, 60–63, 85–93, and 115–117. The worst fit was around residues 115–117, with the notable exception of the complex with pancreatic trypsin inhibitor [Entry Code 2PTC, Huber et al. (1974)], which was quite similar in structure to GB-trypsin in this region of the molecule.

A comparison of the coordinates of GB-trypsin with those of inhibitor-free native trypsin at 1.7-Å resolution is shown in Figure 1. The coordinates of the α-carbon atoms of the

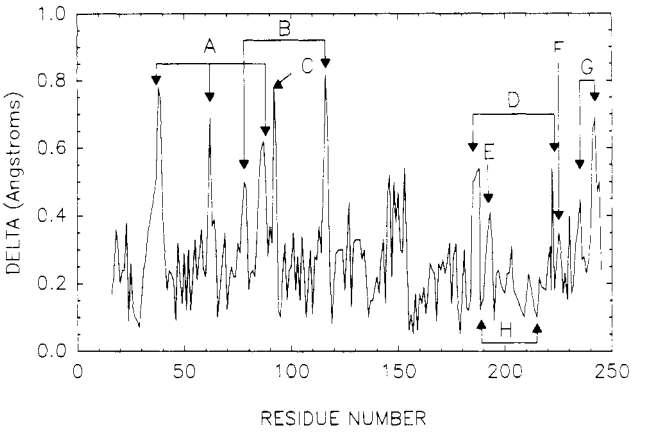

FIGURE 1: Comparison of coordinates of (guanidinobenzoyl)trypsin to native enzyme. Shown are residual distances between equivalent α-carbon atoms after fitting of all α-carbon atoms. Peaks connected by arrows represent portions of the molecule that are near one another in space and are connected by hydrogen bonds, such as neighboring strands of a β-sheet.

native enzyme were obtained from the Brookhaven Protein Data Bank [Entry Code 1TPO, Marquart et al. (1983)] and were rotated onto the refined coordinates for GB-trypsin with the computer program ROTMOL (Steigemann, 1974a). The rms deivation for all 223 α-carbon atoms was 0.31 Å; this could be lowered to 0.23 Å by elimination of 40 atoms. Most of these outliers were not isolated residues but rather comprised discrete, contiguous groups. Visual examination of the three-dimensional model revealed that most of the variable regions extended between chains that were in intimate contact, usually via hydrogen bonds. Futhermore, most of these regions were at the surface of the protein and thus subject to crystal packing forces.

The following is a description of the labeled peaks in Figure 1, which represent regions of interest in the comparison of GB-trypsin to the native enzyme (Entry Code 1TPO, Marquart et al., 1983):

(A) This region consists of three consecutive strands of an antiparallel β-barrel structure that lies at the surface of the enzyme.

(B) This region, containing residues 76–79 and 115–117, is near the Ca^{2+}-binding site and is discussed below. Gly-78 lies at the surface of the trypsin molecule.

(C) Residues 91–93 are at the surface of the protein. Because this region stands out in the comparison of GB-trypsin to each of the other trypsin structures found in Brookhaven Protein Data Bank, these differences may reflect differences in crystal packing forces.

(D) Residues 184–188, which are H bonded to residues 222 and 223, are also near the surface of the protein. They are also extremely close to the active site. Similar plots obtained by comparisons of GB-trypsin to trypsin complexes with *p*-amidinophenyl pyruvate [Entry Code 1TPP, Walter and Bode (1983)] and bovine pancreatic trypsin inhibitor [Entry Code 2PTC, Huber et al. (1974)] show only slight differences in this region. The first of these complexes involves covalent modification of the active site Ser-195; in the second, the active-site region is occupied by a small protein inhibitor, which may form a transient covalent bond to the reactive serine residue. Thus, the movement of these residues may reflect the formation of the covalent adduct of the active-site serine residue.

(E) The residues in the region 192–194 comprise the "oxyanion hole" (Robertus et al., 1972). The structure of GB-trypsin in this region differs somewhat from native [Entry

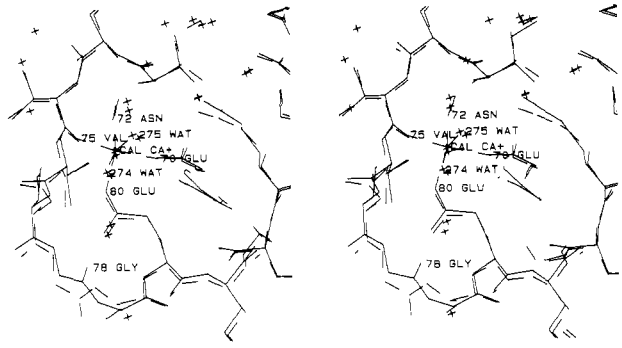

FIGURE 2: Calcium-binding site. Dashed structure is Brookhaven Protein Data Bank entry code 1TPO (Marquart et al., 1983).

Code 1TPO, Marquart et al. (1983)] and benzamidine-inhibited [Entry Code 3PTB, Bode and Schwager (1975)] enzyme but much less so from the (*p*-amidinophenyl)pyruvate complex (Entry Code 1TPP, Walter & Bode, 1983). The DIP-inhibited enzyme [Entry Code 4PTP, Chambers and Stroud (1979)] differs most markedly in this region, possibly owing to the large phosphate anion that occupies the site.

(F) These residues lie at the back of the cleft that accommodates the aromatic ring of guanidinobenzoic acid.

(G) At the carboxy terminus of trypsin are two contiguous, short α-helices, composed of residues 230–235 and 236–245. These two helices are shifted slightly with respect to the native crystal, although it cannot be determined whether this is due to acylation or to crystal packing forces.

(H) The two sides of the substrate binding pocket are labeled to point out that no movement of these residues takes place, at least not at the level of C_α coordinates. The backbone peptide bonds of these residues interact, via hydrophobic forces, with the lipophilic regions of binding-site occupants. This site can accommodate the aromatic ring of inhibitors such as benzamidine and guanidinobenzoate (and, in chymotrypsin, aromatic amino acid substrates), as well as the hydrocarbon chains of arginine and lysine. The coordinates of the three residues comprising the conserved "catalytic triad", His-57, Asp-102, and Ser-195, are also rigidly maintained among the different structures.

The Calcium-Binding Loop. Most of the differences mentioned above can be described as small shifts of short segments of domains. One notable exception is in the loop of residues that envelops the Ca^{2+} ion. The dihedral angles ψ-78 and ϕ-79 are rotated approximately 180° from those in the previously reported structures. The configuration in this region and the nearby Ca^{2+}-binding site are shown in Figure 2. The electron densities in this region of the map indicated unambiguously the interpretation we have chosen. This change allows more subtle differences in the model that extend over three contiguous (76–79) and three nearby (115–117) residues.

A consequence of this peptide flip is that the carbonyl oxygen atom of Gly-78 does not project from the surface of the molecule. Its absence exposes an almost perfectly tetrahedral binding site occupied by solvent between two adjacent, symmetry-related trypsin molecules in the crystal. The four ligands of this solvent molecule (Wat-320) are N-79 of molecule 1, an ordered water molecule (Wat-300), and O-88 and N-90 of molecule 2. If the peptide is placed in the conformation previously reported, the carbonyl oxygen atom O-78 lies within 1.2 Å of this binding site, effectively occluding it. No solvent molecule was reported at this site in any of the trypsin structures in the Brookhaven Protein Data Bank.

Further information about this peptide flip was obtained from a difference Fourier map generated between the final model of GB-trypsin and diffraction data measured from native trypsin crystals that were isomorphous to those of the acyl-enzyme. The tetrahedral solvent binding site near Gly-78 described above was clearly absent in the native enzyme, as indicated by a peak of negative density in the $F_{nat} - F_{acyl}$ difference map. While minor peaks in the difference map suggested movement of the backbone from Glu-77 to Asn-79, especially near C_α of Gly-78, they could not unambiguously be ascribed to the flipping of the 78–79 peptide bond.

With few exceptions, the environment of the octahedrally coordinated Ca^{2+} ion in GB-trypsin is similar to that found in other trypsin structures. Two opposite ligands are water molecules, while the four remaining sites are occupied by the carbonyl oxygen atoms of Val-75 and Asn-72 and the carboxylate oxygen atoms of Glu-70 and Glu-80. An additional glutamate residue, Glu-77, is involved in neutralizing the positive charge of the calcium ion, albeit indirectly—it forms a stable H bond with one of the coordinating water molecules. Each of the ligands seems to be extremely tightly bound: the two water molecules have very low temperature factors, 3.1 Å^2 for Wat-274 and ≤2.0 Å^2 for Wat-275.

In the interior, the configuration of O-78 is nicely accommodated by polar contacts between this atom and $N_{\delta 2}$-79 (3.0 Å) and $O_{\delta 1}$-80 (3.2 Å) and by an H bond to the ordered water molecule (Wat-324; 2.6 Å). Each of these ligands is in essentially the same position as in previously reported structures. However, the modest movements of the backbone atoms allow the side chains of residues 79 (Asn) and 117 (Arg) to interact much more favorably with one another than in any of the non-acyl-enzyme structures.

The Ca^{2+}-binding loop is in close proximity to the 3_{10} helix formed by residues 115–118. Already identified as an outlier in the comparison of GB-trypsin with other trypsin molecules, the helix is shifted toward the Ca^{2+}-binding loop by approximately 0.7 Å relative to its position in each of the other structures. An exception to this is in the trypsin/PTI complex [Entry Code 2PTC, Huber et al. (1974)], which is similar to GB-trypsin in this region. Because this is a crystal form of trypsin not reported before, crystal-packing forces might be responsible for this difference in structure.

The Active Site. The geometry of GB-trypsin in the vicinity of residue 195 is notable for three distinct reasons. First, the molecule as a whole fits the binding site extremely tightly. Second, the geometry of the ester bond linking the hydroxyl of Ser-195 to the benzoyl group of GBA was distorted. Last, the guanidino group was found to be significantly (~40°) out of the plane of the aromatic ring.

The aromatic ring of the guanidinobenzoate moiety is sandwiched into a narrow cleft comprised of residues 189–192 and 214–217, while the guanidino group is very tightly anchored to Asp-189, forming an ion pair and two separate H bonds. Space-filling models of the atoms in this region show that there is essentially no empty space between either face of the aromatic ring and the protein. In addition, seven highly ordered water molecules fill small gaps in the vicinity of the ester bond and the guanidino group, while forming a network of hydrogen bonds. An overall view of this region of the molecule is shown in Figure 3.

For the refinement of the guanidinobenzoate group, bond lengths and bond angles were restrained to idealized values, except where noted above. The atoms of the phenyl ring and its two para-substituent atoms were restrained to be planar, as were the five atoms comprising the guanidino group and its link to the ring. The carbonyl oxygen atom of the ester group is bent slightly out of the plane of the carboxylate toward

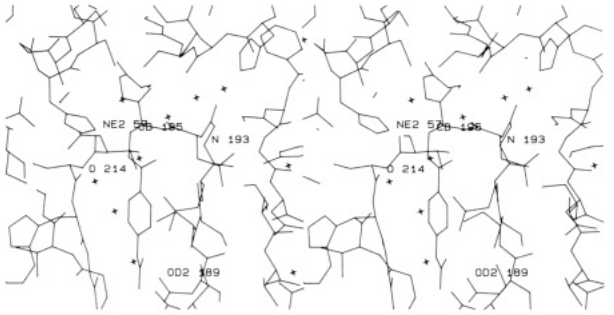

FIGURE 3: View of the trypsin active site. Guanidinobenzoate is shown in an ester linkage to the O_γ of Ser-195.

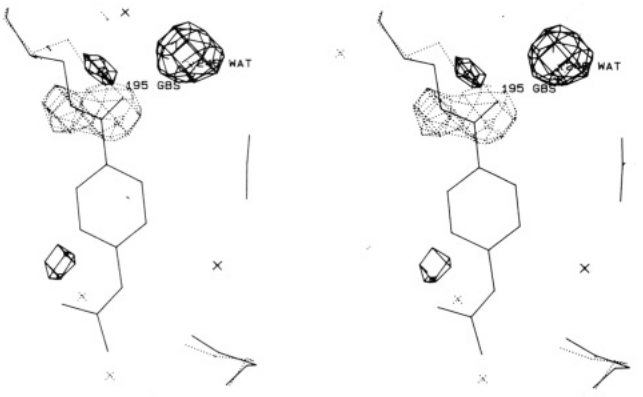

FIGURE 4: Differences between native and GB-trypsin. Continuous contour lines enclose positive density; dotted lines enclose negative density from $F_{Native} - F_{Acyl}$ Fourier synthesis. Dashed model is 1TPO (Marquart et al., 1983).

the peptide nitrogen atoms of residues 193 and 195. In this position it lies nearer to the "oxyanion hole", which is important in stabilizing the putative tetrahedral transition state intermediate in catalysis (Robertus et al., 1972). However, only weak H-bonds are formed: the O–N distances are 3.3 and 3.2 Å, respectively. An H-bond is formed between the carbonyl oxygen atom and water molecule Wat-340 (3.0 Å, C=O···H angle: 136°). Except for this small distortion, there are no apparent strains in the group or bad contacts with other groups. It fits extremely well into the active-site pocket.

The serine proteases are extremely rigid; low temperature factors are associated with most of the structures in the Brookhaven Protein Data Bank. The overall temperature factor of (guanidinobenzoyl)trypsin, mean $B = 11$ Å^2 as determined by a Wilson plot (data not shown) and also from the arithmetic average of the individual B's in the refined model, is smaller than for native trypsin, $B = 17$ Å^2 [Entry Code 1TPO, Marquart et al. (1983)]. Thus, the thermal fluctuations about the equilibrium coordinates are markedly decreased in the acyl-enzyme. A similar result was reported by Bone and co-workers (1987) for the boronic ester of α-lytic protease from *Lysobacter Enzymogenes*, which has a mean B value approximately 15% lower than that of the native enzyme. Inasmuch as the release of ligand from the active site is facilitated by thermal motion, the change in the temperature factor may be partially responsible for the exceptional stability of GB-trypsin.

Despite the overall structural similarity between native and GB-trypsin, the side-chain atoms of Ser-195 move markedly upon acylation. This movement is seen clearly as a pair of density peaks in an $F_{Native} - F_{Acyl}$ difference Fourier map generated between a data set collected from crystals of native trypsin and the GB-trypsin model coordinates, Figure 4. The native crystals were isomorphous to those of GB-trypsin. Part

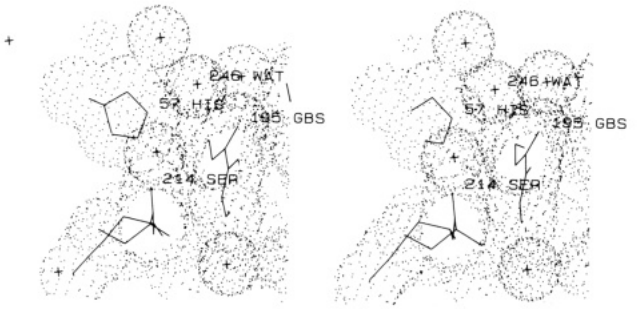

FIGURE 5: Water molecules approaching the active site of GB-trypsin. The shells shown are at one van der Waals radius.

of this movement can be explained by torsion about the C_α–C_β bond. Such a rotation has been observed for inhibitor–enzyme complexes that mimic the tetrahedral reaction intermediate, e.g., tosylchymotrypsin (Matthews et al., 1967), (diisopropylphosphoryl)trypsin (Chambers & Stroud, 1979), (*p*-amidinophenyl)pyruvate–trypsin (Walter & Bode, 1983), and the chymostatin/*Streptomyces griseus* protease A complex (Delbaere & Brayer, 1985). Unlike the tetrahedral complexes, however, we find significant movement of C_β of Ser-195 in GB-trypsin relative to its position in the native enzyme. This atom moves 0.7 Å from its position in native trypsin down into the binding cleft, as indicated by a pair of density peaks, one positive and one negative, Figure 4. Since Asp-189 is rigidly held in place by its inclusion in a β-barrel and since there is virtually no flexibility in the GBA moiety, the formation of the energetically favorable H-bonds and ion pair between the carboxylate of Asp-189 and the guanidino group of GBA apparently requires that C_β of Ser-195 be pulled significantly toward Asp-189. A similar requirement does not arise with arginine in a "natural" substrate of trypsin, because the hydrocarbon backbone of arginine is considerably more flexible than the rigid aromatic ring found in GBA. Furthermore, arginine is approximately 0.5 Å longer than GBA, even though there are equal numbers of carbon atoms between the carboxylate and guanidino groups in the two molecules. The increased length and the greater flexibility of the aliphatic backbone of arginine allows the two ends of the "natural" substrate to accommodate both ends of the binding site without distortion of the enzyme around C_β of Ser-195.

Two ordered water molecules in native trypsin are removed from the active site upon formation of the guanidinobenzoyl ester of Ser-195, Figure 4. A third water molecule, Wat-246, located near Ser-195 in the native structure moves 1.1 Å. It is displaced by the carbonyl oxygen atom O_δ of the GBA group in GB-trypsin, as indicated by a large positive peak in the $F_{Native} - F_{Acyl}$ difference Fourier map, Figure 4. In the acyl-enzyme crystal, this water molecule is held firmly in place 3.8 Å from the imidazole ring of His-57 and 3.1 Å from the carbonyl oxygen atom of the acyl group with a C=O···H angle of 103°. A view of this water molecule and others, with their van der Waals spheres represented as shells of dots, is shown in Figure 5.

In other structures of serine proteases wherein the active-site serine is reported to be covalently modified (Bone et al., 1987; Dixon & Matthews, 1989), the geometry around the carbonyl carbon atom of the ester bond is also distorted. These distorted structures seem to lie along the reaction coordinate for nucleophilic addition to the carbonyl group (Burgi et al., 1973). In these molecules, His-57 is not H bonded to the hydroxyl group of Ser-195 but rather to an ordered, active-site water molecule. This activated water molecule is held in close proximity (~2.4 Å) to the carbonyl carbon atom of the ester

bond and is thus poised to act as a nucleophile for the hydrolysis of that bond.

Two aspects of the geometry of GB-trypsin distinguish it from these other structures. First, the acylating carbon atom in GB-trypsin is not nearly as tetrahedral in nature as found in other covalently modified trypsins. Second, in GB-trypsin the only two active-site water molecules that might serve as potential nucleophiles in deacylation, Wat-246 and Wat-292, are each located >3.7 Å away from the carbonyl carbon atom. Wat-292 is closer to His-57 than Wat 246, 3.2 Å versus 3.8 Å, and therefore has the potential to interact with it more strongly. However, examination of the model makes it more likely that Wat-292 is H-bonded to the side chain of Ser-214. These two water molecules have high refined occupancies (1.0 and 0.97) and moderate crystallographic temperature factors (28 and 34 Å^2). As a consequence, the probability is low that fluctuations from the mean crystal structure will result in geometries conducive to hydrolysis. Furthermore, there is no space for disordered water molecules to diffuse into the active site to approach the carbonyl carbon atom, Figure 5. Thus, the details of the structure of the ester bond and its immediate environment—the exact positioning of water molecules and the shift of C_β of Ser-195 toward the active-site pocket—are sufficient to explain the extreme resistance of this complex to deacylation.

Acknowledgments

We gratefully acknowledge the assistance in some of the experiments of Elisa Dilg, Lisa Reynolds, and Lisa Wolf and of Ed Conti and Mike DeSousa, who were supported by the Department of Energy's Division of University and Industry programs, Office of Energy Research, as Lab Co-op Program participants and as Teacher Research Associate Program participants, respectively.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535.

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969) *Nature (London) 221*, 337.

Bode, W., & Schwager, P. (1975) *J. Mol. Biol. 98*, 693.

Bone, R., Shenve, A. B., Kettner, C. A., & Agard, D. A. (1987) *Biochemistry 26*, 7609.

Burgi, H. B., Dunitz, J. D., & Shefter, E. (1973) *J. Am. Chem. Soc. 95*, 5065.

Chambers, J. L., & Stroud, R. M. (1977) *Acta Crystallogr. B33*, 1824.

Chambers, J. L., & Stroud, R. M. (1979) *Acta Crystallogr. B33*, 1861.

Chase, T., Jr., & Shaw, E. (1967) *Biochem. Biophys. Res. Commun. 29*, 508.

Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossman, M. G., Ed.) p 713, Gordon and Breach, New York.

Delbaere, L. T. J., & Brayer, G. D. (1985) *J. Mol. Biol. 183*, 89.

Dixon, M. M., & Matthews, B. W. (1989) *Biochemistry 28*, 7033.

Finkenstadt, W. R., & Laskowski, M., Jr. (1965) *J. Biol. Chem. 240*, 962.

Hartley, B. S. (1964) *Nature (London) 201*, 1284.

Hartley, B. S., & Kauffman, D. L. (1966) *Biochem. J. 101*, 229.

Henderson, R. (1970) *J. Mol. Biol. 54*, 341.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigemann, W. (1974) *J. Mol. Biol. 89*, 73.

Jones, A. (1978) *J. Appl. Crystallogr. 11*, 268.

Kossiakoff, A. A., Chambers, J. L., Kay, L. M., & Stroud, R. M. (1977) *Biochemistry 16*, 654.

Krieger, M., Kay, L. M., & Stroud, R. M. (1974) *J. Mol. Biol. 83*, 209.

Laemmli, U. K. (1972) *Nature (London) 227*, 680.

Luzatti, V. (1952) *Acta Crystallogr. 5*, 802.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr. B39*, 480.

Matthews, B. W., Sigler, P. B., Henderson, R., & Blow, D. M. (1967) *Nature (London) 214*, 652.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1975) *J. Biol. Chem. 250*, 7120.

Melhado, L. L., Peltz, S. W., Leytus, S. P., & Mangel, W. F. (1982) *J. Am. Chem. Soc. 104*, 7299.

Messerschmidt, A., & Pflugrath, J. W. (1987) *J. Appl. Crystallogr. 20*, 306.

Pflugrath, J. W., & Messerschmidt, A. (1985) in *Crystallography in Molecular Biology*, Meeting Abstracts, Bischenberg, France.

Pflugrath, J. W., Saper, M. A., & Quiocho, F. A. (1984) in *Methods and Applications in Crystallographic Computing* (Hall, S., & Ashida, T., Eds.) pp 404–407, Clarendon Press, Oxford.

Robertus, J. D., Kraut, J., Alden, R. A., & Birktoft, J. J. (1972) *Biochemistry 11*, 4293.

Sigler, P. B., Jeffery, B. A., Matthews, B. W., & Blow, D. M. (1966) *J. Mol. Biol. 15*, 175.

Steigemann, W. (1974a) Ph.D. Dissertation, Technische Universität, Müchen.

Steigemann, W. (1974b) *J. Mol. Biol. 89*, 73.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212.

Tronrud, D. E., Ten Eyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489.

Walter, J., & Bode, W. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 949.