Trypsin Revisited

CRYSTALLOGRAPHY AT (SUB) ATOMIC RESOLUTION AND QUANTUM CHEMISTRY REVEALING DETAILS OF CATALYSIS*

Received for publication, June 30, 2003, and in revised form, August 21, 2003
Published, JBC Papers in Press, August 22, 2003, DOI 10.1074/jbc.M306944200

Andrea Schmidt‡§, Christian Jelsch§, Peter Østergaard‖, Wojciech Rypniewski**, and Victor S. Lamzin‡

From the ∥European Molecular Biology Laboratory (EMBL) Hamburg c/o DESY, D-22607 Hamburg, Germany,
the ¶Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61–704 Poznan, Poland and Institute of Biochemistry
and Molecular Biology, University Hospital, Hamburg-Eppendorf, c/o DESY, 22603 Hamburg, Germany, the ‖Laboratoire
de Cristallographie et Modélisation des Matériaux Minéraux et Biologiques LCM3B CNRS UHP Faculté des Sciences,
54506 Vandoeuvre-les-Nancy Cedex France, and the **Novozymes Protein Design, 2cs.01 DK-2880 Bagsvaerd, Denmark

A series of crystal structures of trypsin, containing either an autoproteolytic cleaved peptide fragment or a covalently bound inhibitor, were determined at atomic and ultra-high resolution and subjected to ab initio quantum chemical calculations and multipole refinement. Quantum chemical calculations reproduced the observed active site crystal structure with severe deviations from standard stereochemistry and indicated the protonation state of the catalytic residues. Multipole refinement directly revealed the charge distribution in the active site and proved the validity of the ab initio calculations. The combined results confirmed the catalytic function of the active site residues and the two water molecules acting as the nucleophile and the proton donor. The crystal structures represent snapshots from the reaction pathway, close to a tetrahedral intermediate. The de-acylation of trypsin then occurs in true S₂,2 fashion.

Serine proteases are among the most abundant proteins. A large subgroup comprising trypsin and chymotrypsin is called trypsin-like serine proteases. Trypsin cleaves peptides at the C-terminal side of arginine or lysine. Catalysis takes place in a cleft from which a binding pocket (specificity pocket) protrudes into the interior of the enzyme. A catalytic triad (Ser195-His56-Asp99) constitutes the core of the reactive center. For the stabilization of the charges developing on the reaction intermediates during catalysis an oxyanion hole is formed by the main-chain N-H of residues 192–195 in an arrangement known as a nest (1). The specificity pocket accommodates long, basic amino acid side chains. At its bottom, an asparagine residue forms a tight salt bridge with the substrate. The difference in specificity between various serine proteases is determined by the shape and the width of the specificity pocket (2, 3).

The catalytic mechanism of trypsin isolated from various sources has been investigated in great depth (4–10). Yet several questions remain unanswered. One is the identification of the catalytically active water molecules. Also, the catalytic triad appears to function in different ways in different proteins and display different features, particularly the presence of the so-called short hydrogen bonds (11).

Trypsin from Fusarium oxysporum shows auto-protolysis activity even at crystallization conditions. A peptide fragment can always be found in the active site of the crystalline enzyme (12). This together with the fact that its crystals diffract extremely well has made it especially interesting for mechanistic studies and offered an excellent tool for the investigation of the catalytic activity at the electronic level (13, 14).

EXPERIMENTAL PROCEDURES

Crystalization—Protein samples of Fusarium oxysporum trypsin were supplied by Novozymes at a concentration of 10 mg/ml. Crystals were grown as described elsewhere (15), with the pH lowered to 5.0 instead of 6.0 using 100 mM sodium citrate buffer. A combination of streak seeding and macroseeding was applied to initiate nucleation. Crystals grew to a size of 0.3–0.5 mm within a few days.

Co-crystallization with the Inhibitor DFP—DFP (diisopropyl fluorophosphate, Mw = 184 g/mol) was purchased from Aldrich Chemicals. 5 μl of the inhibitor were dissolved in 250 μl of isopropyl alcohol. 1 μl of this solution was added to a drop containing 10 μl of protein solution and 20 μl of precipitant solution. The approximate molar ratio of inhibitor to protein was 20:1.

Co-crystallization with the Inhibitor PMSF—An aqueous stock solution of a concentration of 8 mg/ml PMSF (Mw = 174 g/mol) from Sigma was prepared, and 1 μl of this solution was added to the drops containing 10 μl of protein and 20 μl of precipitant solution. The molar ratio of inhibitor to protein was again 20:1. Nucleation was initiated as for the native crystals in both cases.

Soaking—Native trypsin crystals were soaked in a drop containing the precipitant solution at pH 4 for several minutes prior to cryo-freezing.

Cryo-freezing Conditions—In previous publications, the cryo-freezing conditions involved a gradual increase of glycerol content in the mother liquor up to 20% (v/v) glycerol (12). Now, we transferred the crystals into light paraffin oil directly from their drops. In order to avoid ice formation, all traces of the mother liquor were carefully removed from the crystal surface by moving it around in the oil drop. Crystals were then flash frozen directly in the cryo-stream at 100 K.

Data Collection and Processing—Data were collected on beamlines X11 and X13 at EMBL Hamburg, equipped with MAR CCD detectors, at wavelengths of 0.6110 and 0.5920 Å, respectively. A summary of the data statistics is given in Table I. Data were processed with denzo scalepack or HKL2000 (16), intensities were truncated to structure factors using the CCP4 suite of programs (17).
Trypsin at Atomic and Ultra-high Resolution

TABLE I
Data collection statistics

| Data set    | Beamline | Resolution | R_{sym} | (|F|/|I|) (outer shell) | Wilson B factor | Space group |
|-------------|----------|------------|---------|---------------------|----------------|-------------|
| pH 4        | X13      | 22–1.0     | 7.3     | 3.8                 | 7.1            | P1          |
| pH 5        | X11      | 30–0.83    | 3.3     | 5.5                 | 7.2            | P1          |
| DFP         | X11      | 31–1.22    | 4.5     | 5.5                 | 11.3           | P1          |
| PMSF        | X11      | 25–1.23    | 4.6     | 2.7                 | 11.6           | P2\(\tilde{}\) |
| pH 5/borax  | X11      | 22–0.80    | 3.3     | 6.2                 | 4.8            | P1          |

Refinement—Initial refinement and addition of solvent were carried out with ARP/wARP (18) and refmac (19). Subsequent steps were performed with SHELXL (20), including the addition of hydrogen atoms and anisotropic B factors, until convergence was reached. Refinement was carried out against diffraction intensities. The same refinement protocol was used for all structures described here. HeterocOMPoundS were introduced in the refinement as early as possible. Occupancies for alternate conformers, and the substrate were set according to their anisotropic atomic displacement parameters (ADP) and individually refined in a last round. Solvent molecules with ADP higher than 60 Å² were removed; those with ADP higher than 45 Å² were assigned occupancy of 0.5.

Complementary Methods—Ab initio quantum chemical calculations on the pH 4 structure were performed using GAMESS (21) with a 6–31G** basis set and a Hückel model for the construction of the initial molecular orbitals. The active site residues, the substrate and side chains involved in direct contact with the substrate were chosen as the geometrical template. Only relevant parts of the interacting compounds (i.e., atoms at an H-bonding distance) were included in the calculations. The substrate arginine was treated as alanine since only its main chain atoms interacted with the active site. Aspartate 99 was modeled as acetic acid. The system included a total of 64 atoms (285 electrons, with a net charge of –1). Atoms of chemically inert groups not involved in direct contact with the substrate were fixed at their positions. Solvent effects were neglected as the occupied active site is a closed cavity without considerable contact to the solvent space. Hydrogen atoms (at the two water molecules in the active site) were inserted into the structure according to standard stereochemistry and existing H-bonding partners. The following cases were subjected to refinement: QC1, active site in the native state; QC2, Ser195⁴¹⁸ de-protonated; QC3, W1 as hydroxyl ion; PMSF, covalently bound.

Multipolar Atoms Refinement—The 0.8 Å structure was refined using the charge density refinement program MoPro (20). The program describes the electron density of the atoms with the Hansen and Coppens model (22), which includes atomic charges, expansion/contraction coefficients, and multipoles.

The structure refined with SHELXL was used as the starting model. The charge density parameters were transferred to the structure from our data bank describing the electron density of atoms in proteins (23, 24). Protein atoms with ADPs higher than 15 Å² and multiple conformers were considered spherical and were only assigned a charge and a contraction/expansion coefficient. Water molecules were described as a neutral oxygen atom and were refined in conventional manner.

The contribution of the bulk solvent to the crystal diffraction was added to the structure factors (25, 26) in Equation 1.

\[ F_i = F_i(prot) + K_{sol} \exp(-B_{sol} s^2) F_i(sol) \]  
(Eq. 1)

These parameters were also refined in MoPro. The following weighting scheme was applied to the reflections shown in Equation 2.

\[ W_{ij} = [(|F_i|^2 + 0.02F_i^2)]^{-1/2} \]  
(Eq. 2)

The three different types of structural parameters were refined in an iterative way until convergence: scale factor with solvent parameters \( K_{sol} \) and \( B_{sol} \), atomic coordinates, ADPs. Hydrogen atoms were not refined and their coordinates were constrained according to standard stereochemistry while their isotropic B-factors were set to the \( B_{eq} \) of the parent atom. Usual distance restraints were applied on the structure according to the Eng & Huber stereochemical dictionary (target rmsd deviations of 0.02 and 0.05 Å for bond and angle distances, respectively).

RESULTS

Despite a considerable degree of disorder in the structures, all of the models refined to crystallographic \( R \) factors well below 15%. A summary of the structure refinements is given in Table II. The active site was never empty. In the native cases, an arginine molecule was bound into the specificity pocket with its main-chain atoms located in a good position for catalysis near Ser195. Arginine is the C-terminal residue of a tripeptide, which is the remaining fragment of auto-proteolysis. The substrate tripeptide shows only partial occupancy, which seems to correlate with the occupancy distribution of the multiple conformers in the structure. The arginine lacks the second oxygen atom of its carbonyl group and the C=O distance of 1.48 Å is unusually long. The arginine side chain was observed in at least two conformations, in the pH 4 structure even overlapping with a lysine. Ser195 was always in a single conformation in well defined electron density. Water molecule W2 was located close to the substrate carbonyl oxygen atom in the oxyanion hole. The substrate carbonyl carbon had contacts to W1 (1.87 Å) and Ser195 Oγ (2.05 Å) (Fig. 1), which implies covalent interaction. The geometry around the substrate carbonyl carbon was roughly tetrahedral. No electron density could be found for the N-terminal fragment of the substrate. The water molecules W1 and W2 were fully occupied as opposed to the substrate, which showed partial occupancy with the total ranging from 0.4 (pH 5) to 0.8 (pH 5/borax and pH 4).

DFP and PMSF were bound covalently to Ser195. The phosphate groups were always well visible. The respective isopropyl and phenyl moieties displayed weaker electron density and did not point into the specificity pocket but in the direction of His56. As an effect, the histidine side chain is statically disordered showing two conformations, one of which is rotated away from the inhibitors thus disrupting the catalytic triad (Fig. 1). In the DFP and PMSF structures, one oxygen atom in the sulfonyl or phosphate groups takes the position of W1, another one the position of W2.

Like the other native structures, the pH 5/borax structure also had a peptide bond into the active site. This structure had the highest quality data set and was used for multipole refinement.

Co-crystallization with DFP and PMSF changed the space group from P1 to P2₁ but left the internal protein structure unchanged and the crystal contacts mostly preserved.

Disorder and Mobility—One-fifth of the total number of residues of F. oxysporum trypsin shows at least two conformers. Their occupancies relate to the occupancy of the substrate or of the inhibitor side chain. In all the structures except native pH 5, the substrate- or inhibitor-bound state seems to correspond to the protein conformer with higher occupancy, with an average value of 0.65 in all structures. The conformers with corresponding lower occupancy (0.35) correspond to the empty active site. The rmsd for the two visible main chain conformers of disordered residues lies between 1 and 2 Å. There was no evidence for the presence of other, intermediate conformational states.

Most of the disordered residues are spatially linked to each other. The effect starts in the loops of the active site cleft and is then transferred to the adjacent regions. The patterns of disorder are different for the native and the inhibited structures.
The inhibited structures, mimicking a reaction intermediate, show higher anisotropy of the active site residues. In the native structures at pH 4 and pH 5, disorder arises in the specificity pocket, which is not the case in the DFP and PMSF structures where this site is empty. Despite the apparent movement in and around the active site, Ser195 shows only one conformation as mentioned above.

An analysis of corresponding conformers showed that there is good correlation between rmsd and ADP in the native structures indicating disorder of mostly dynamic origin. Disorder in the inhibited structure appears to be mostly static.

Quantum Chemical Calculations on the pH 4 Structure—A stable match between the crystal structure and the theoretical model could only be achieved under the assumption that the catalytic serine is de-protonated and the attacking water molecule W1 is actually a water molecule and not a hydroxyl ion. In

<table>
<thead>
<tr>
<th>Crystallographic refinement</th>
<th>pH 4</th>
<th>pH 5</th>
<th>PMSF</th>
<th>DFP</th>
<th>pH5/borax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>1.0</td>
<td>0.83</td>
<td>1.23</td>
<td>1.22</td>
<td>0.80</td>
</tr>
<tr>
<td>R (%) (all data)</td>
<td>12.8</td>
<td>9.8</td>
<td>14.1</td>
<td>14.0</td>
<td>10.7</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>86,663</td>
<td>146,839</td>
<td>44,081</td>
<td>47,457</td>
<td>165,882</td>
</tr>
<tr>
<td>Number of atoms total</td>
<td>2,139</td>
<td>2,263</td>
<td>1,935</td>
<td>1,979</td>
<td>2,162</td>
</tr>
<tr>
<td>Number of water molecules</td>
<td>330</td>
<td>489</td>
<td>269</td>
<td>265</td>
<td>495</td>
</tr>
<tr>
<td>Overall atomic displacement (all atoms)</td>
<td>11.8</td>
<td>9.2</td>
<td>16.8</td>
<td>16.0</td>
<td>9.4</td>
</tr>
<tr>
<td>Overall atomic displacement (protein)</td>
<td>9.8</td>
<td>6.6</td>
<td>15.3</td>
<td>14.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Number of disordered residues</td>
<td>40</td>
<td>33</td>
<td>16</td>
<td>27</td>
<td>14</td>
</tr>
</tbody>
</table>
all other cases, the geometry was directed toward either the products of peptide cleavage (QC3) or the covalent intermediate (see Fig. 2 for an overview of the catalytic pathway).

This implies that the crystal structure represents a state close to a tetrahedral transition state. Bonds of respective order 0.99 and 0.47 are formed between the substrate C and the Ser195 O and from the substrate C to W1, which is held in place from the other side by His56. His56 is deprotonated and acts as a Lewis base, activating the catalytic water molecule, which is indicated by a strong H-bonding contact of bond order 0.2–0.3.

Water molecule W2 in the oxyanion hole forms a very strong H bond to the carbonyl oxygen of the arginine substrate in the model matching the crystal structure. This is accompanied by a high negative charge on the W2 oxygen atom.

**Multipole Refinement**—For the ultra-high resolution structures (pH 5 and pH 5/borax), the protonation state in the active site could be visualized directly in residual electron density maps. His56 was indeed not protonated (Ne1 is non-protonated while the hydrogen atom HN1 is bound to N82). The multipole charge density parameters were transferred from the database and Fig. 3A shows the corresponding deformation electron density on the catalytic histidine. The small peak near Ne2 indicates the electron lone-pair on the nitrogen atom or a partial protonation (which at this pH would be quite probable), or a rotation of the histidine ring by 180° around the C6–C7 bond. The electrostatic potential derived from the multipoles transfer is shown for the His56–Asp59 interaction (Fig. 3B).

**DISCUSSION**

**Comparison to the Previously Refined Structures**—Remnants of the substrate have been observed in the *F. oxysporum* trypsin structures before (12). Lowering the pH to 5.0 resulted in lower occupancy and higher mobility of the substrate arginine side chain and the disappearance of two of the three water molecules arranged around the arginine carbonyl in the previous structures (12). Only water W1 remained. The structures showed no other significant differences.

**Mobility in the Crystal**—The walls of the active site cleft at the subsites S2, S3 (residues 144–150) are always in double conformation, subsites S2' and S3' (residues 93–95) only in the DFP and PMSF structures. This can be understood as an effect of the release of the N-terminal fragment of the substrate (sites S2, S3) and, for the C-terminal part of the active site cleft, adjustments in the positioning of the substrate just before the nucleophilic attack of water W1.

In the pH 4 and pH 5 structures, Cys191, Tyr225, and the stretch between amino acids 215 and 220 show main chain double conformations. Cys191 and Tyr225 are part of the wall of the specificity pocket. Cys216 forms a disulfide bridge with Cys191 and hence moves together with it. Apparently, this is related to the binding and release of the substrate arginine side chain, which also shows multiple conformers. Movement upon binding or release is supported by the fact that in the case of the covalently bound inhibitors, which did not reach into the pocket, the specificity pocket was completely rigid. An induced fit model was also supported by interpretation of the structures with inhibitors. In both cases the protein showed two different conformers in the subsites S2, S3, S2', and S3', whose occupancies relate to the occupancies of DFP (0.7) and PMSF (0.6).

Two important residues involved in this movement are Asp189 and Ser190, both located at the very bottom of the pocket, highly conserved and responsible for the preference of trypsin for basic residues (2, 3). Asp189 forms a strong ion pair with the substrate arginine side chain and its movement as well as the disorder of the other residues in the specificity pocket appears to be induced directly by the substrate. Changes in the substrate position would be transmitted to the catalytic center or vice versa. Re-positioning must occur during the formation and breaking of the covalent bond to Ser195. NMR experiments on complexes showed that neither active site nor specificity pocket is fully pre-formed in the native structures. Only binding to specificity subsites creates an environment suitable for a tetrahedral transition state (27). This seems to be the case also in *F. oxysporum* trypsin,
in contrast to the widely accepted lock-and-key model for serine protease substrate binding (28).

Increased anisotropy was observed for the residues of the catalytic triad in the inhibited structures. The shape of the thermal ellipsoids of the atoms in His $^{35}$ and Asp $^{99}$ indicated movement toward Ser $^{195}$ and the atoms in Ser $^{195}$ tend to move perpendicular to the plane formed by C $^{\beta}$, O $^{\gamma}$, and P or S (Fig. 4). In addition, some residues at sites S2 and S3 show considerable shift of one conformer toward the active site, shortening the cleft. This is not observed in the structures with the substrate. We therefore assume that a re-arrangement in the substrate binding region is required before de-acylation. In the covalent intermediate, the Ser $^{195}$ ester carbonyl can have a new orientation (9). It has to be adjusted, fixed, and a water molecule has to approach the covalent intermediate and position itself in a location suitable for catalysis. The water molecule W1 in the pH 4 and pH 5 structures is already at its original state and functions as the Henderson water is internal bonds and external H bridges of normal strength. This indicates that, after the nucleophilic attack of W1, W2 regains its original state and functions as the Henderson water is enriched in anisotropy.

**Reaction Partners and Water Molecules**—W1, attached to His $^{35}$, and W2 in the oxyanion hole were the two candidates for participation in catalysis. The role of the water molecule W1 is nucleophilic attack and steric effects might explain why the reaction did not proceed further. Embedding of protein molecules in crystals makes large movement impossible and impedes product release or binding: the active site of trypsin was in principle accessible in the crystals, but it was impossible to bind DFP to the protein by soaking; DFP could not be seen in the active site of trypsin in soaked crystals. This again strongly indicates an induced fit mechanism for substrate binding and the presence of an “open” and “closed” state or an “empty” and “loaded” conformation of the protein molecule.

**Results from the ab initio calculations** on the pH 4 structure show good agreement (low rmsd) to the crystal structure for the model in which Ser $^{195}$ is deprotonated (QC2).

<table>
<thead>
<tr>
<th>Distances (Å)</th>
<th>QC 1</th>
<th>QC 2</th>
<th>QC 3</th>
<th>Crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser $^{195}$ Oγ-Arg C</td>
<td>2.29</td>
<td>2.05</td>
<td>n/a</td>
<td>2.05</td>
</tr>
<tr>
<td>Arg C-Arg O</td>
<td>1.14</td>
<td>1.47</td>
<td>1.20</td>
<td>1.48</td>
</tr>
<tr>
<td>W1 O-Arg C</td>
<td>1.80</td>
<td>1.87</td>
<td>1.31</td>
<td>1.87</td>
</tr>
<tr>
<td>W1 O-His $^{35}$ Ne2</td>
<td>2.65</td>
<td>2.92</td>
<td>2.06</td>
<td>2.76</td>
</tr>
<tr>
<td>W2 O-Arg O</td>
<td>2.51</td>
<td>1.83</td>
<td>1.86</td>
<td>1.83</td>
</tr>
<tr>
<td>R.m.s.d. to the crystal structure</td>
<td>0.24</td>
<td>0.07</td>
<td>0.47</td>
<td></td>
</tr>
</tbody>
</table>

**Charges**

| Arg C | 0.824 | 0.939 | 0.809 |
| Arg O | -0.550 | -0.541 | -0.656 |
| Ser $^{195}$ Oγ | 0.856 | -0.668 | -0.791 |
| W2 O | -0.744 | -0.990 | -0.737 |
| W1 O | -0.700 | -0.720 | -0.637 |

**Figure 4.** Thermal ellipsoids with 35% probability for the catalytic triad of the pH 5 and PMSF structures showing the difference in anisotropy. The figure was produced in XtalView (37).
Trypsin at Atomic and Ultra-high Resolution

no longer be considered spherical. The presence of lone-pair and bonding electrons has then to be directly included into the refinement.

The electron density in the active site of the ultra-high resolution trypsin structures shows that His\(^\text{56}\) is definitely deprotonated. Because of a relatively high thermal motion on the Ser\(^\text{195}\) O\(_\text{y}\), the electron density cannot give evidence for an attached proton on this side chain. However, the serine 195 C\(_\beta\)-O\(_\text{y}\) distance is 1.22 Å (pH 5 borax structure), which is significantly smaller than the standard protonated serine C\(_\beta\)-O\(_\text{y}\) distance of 1.417 ± 0.020 Å in the Engh and Huber (35) stereochemistry dictionary. This implies the following. First, the pK\(_\text{a}\) of His\(^\text{56}\) is lower than 6.5. This is already hinted at by the presence of enzymatic activity under crystallization conditions despite a basic pH optimum and proves that His\(^\text{56}\) acts as a general base. A rotation of the imidazole ring by 180° could play a role in the substrate binding and product release steps. Second, Ser\(^\text{195}\) is de-protonated as well and is capable of covalent attachment to the substrate. And third, the assumptions made on the protonation state in the active site were correct and therefore the \textit{ab initio} calculations have produced a realistic model.

Implications for Catalysis and Conclusions—Using the atomic resolution structures and the results from the \textit{ab initio} calculations, it could be shown directly that trypsin de-acylation does occur via a tetrahedral intermediate with subsequent hydrolysis in true S\(_\text{N}_2\) fashion. The water molecule W1 is the obvious candidate to perform hydrolysis, while the preparation for this step is performed by W2 by protonating the carbonyl oxygen. Despite the requirement for W1 to carry a negative partial charge in order to perform a nucleophilic attack, complete de-protonation does not occur. Its attachment to the substrate and proton abstraction is concerted.

The lowered pK\(_\text{a}\) of His\(^\text{56}\) explains why the enzyme was still active in the crystallization solution. The fact that the His\(^\text{56}\) side chain swings out upon contact with the inhibitors implies that the catalytic triad is not rigid and tight. The inhibitors are larger than the substrate intermediate and remove the water molecules W1 and W2 from their normal positions. Therefore, no activator (W2) and no nucleophile (W1) for hydrolysis are larger than the substrate intermediate and remove the water that the catalytic triad is not rigid and tight. The inhibitors are side chain swings out upon contact with the inhibitors implies that the assignment of function to the residues involved in catalysis.

The motion of the macromolecule in the crystal structure, revealed from the anisotropic atomic displacement parameters, complements the time-average structural picture with dynamics thus opening a new prospective in the structure-functional analysis of biological processes.

Acknowledgments—We thank Richard J. Morris, Rob Meijers, and Benoit Guilot for helpful discussions.

REFERENCES
