Acta Cryst. (1995). D**51**, 73–84

Structure of Inhibited Trypsin from *Fusarium oxysporum* at 1.55 Å

BY WOJCIECH R. RYPNIEWSKI*

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 *Hamburg, Germany*

CLAUS DAMBMANN AND CLAUS VON DER OSTEN

Novo Nordisk, Novo Alle, DK-2880 *Bagsvaerd, Denmark*

AND MIROSLAWA DAUTER AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 *Hamburg, Germany*

(*Received* 9 *March* 1994; *accepted* 5 *August* 1994)

Abstract

The structure of trypsin from the fungus *Fusarium oxysporum* has been refined at 1.55 Å resolution by restrained least-squares minimization to an *R* factor of 14.4%. The data were recorded from a single crystal on the X31 beamline at EMBL, Hamburg, using a locally developed image-plate scanner. The final model consists of 1557 protein atoms, 400 water molecules, one molecule of isopropanol and one monoisopropyl phosphoryl inhibitor group covalently bound to the catalytic Ser195. Comparison of the structure with bovine trypsin reveals significant differences in the active site and suggests a possible explanation for the difference in substrate specificity between the two enzymes. In *F. oxysporum* trypsin the specificity pocket is larger than in bovine trypsin. This explains the preference of *F. oxysporum* trypsin for the bulkier arginine over lysine and the reverse preference in bovine trypsin. The binding cavity on the C-terminal side of the substrate is more restricted in *F. oxysporum* trypsin than in mammalian and *Streptomyces griseus* trypsins, which explains the relative inactivity of *F. oxysporum* trypsin towards peptide–*p*NA substrate analogues as an unfavourable steric interaction between the side of the binding cavity and the *para*-nitroanilino group of peptide–*p*NA. The observed restriction of the binding cavity does not lead to a reduced catalytic activity compared to other trypsins.

Introduction

Trypsin (E.C. 3.4.21.4) is a serine proteinase characterized by specific cleavage of the peptide bond on the C-terminal side of lysine or arginine (for reviews see, Kraut, 1977; Steitz & Shulman, 1982). It is an important digestive enzyme and in the pancreas it is responsible for activating pancreatic enzymes including itself, by cleaving a short peptide from the amino terminus of their inactive zymogens. Historically, trypsin has been studied mostly from mammals and its ubiquity was not fully recognised until it was found in the bacterium *Streptomyces griseus* (Olafson & Smillie, 1975; Olafson, Jurásek, Carpenter & Smillie, 1975) and since then in a number of other organisms both eukaryotic and prokaryotic. More than 30 complete sequences of trypsins have been reported to date, exceeding by far the number of crystal structures. The evolutionary relationship between various trypsins has recently been established by comparing the known sequences and crystal structures (Rypniewski, Perrakis, Vorgias & Wilson, 1994). The study emphasized the need for three-dimensional information in comparison of proteins from distantly related species and in identification of features responsible for enzymatic activity and structural stability. The known crystal structures include bovine trypsin (Chambers & Stroud, 1977, 1979; PDB code 4PTP, Bernstein *et al.*, 1977; Abola, Bernstein, Bryant, Koetzle & Weng, 1987), a mutant trypsin from rat (Sprang, Standing, Fletterick & Stroud, 1987; PDB code 1TRM), trypsin from the North Atlantic salmon (Smalås & Hordvik, 1993; PDB code 1TBS) and bacterial trypsin from *S. griseus* (Read & James, 1988; PDB code 1SGT). Another bacterial trypsin from *S. erythraeus* has been determined at 2.7 Å resolution (Yamane *et al.*, 1991). The crystal structure of trypsin from *Fusarium oxysporum*, an organism distant in evolutionary terms from both bacteria and vertebrates, has previously been reported at 1.8 Å resolution (Rypniewski *et al.*, 1993). In this paper we present the refinement of *F. oxysporum* trypsin at 1.55 Å resolution and examine the structure in light of previously unpublished biochemical data.

Biochemical properties

Enzymatic activity

Proteolytic activity was measured as CPU (casein proteinase units) using Hammarsten casein as substrate. One CPU is defined as the amount of enzyme liberating

* Correspondence address: EMBL, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany.

© 1995 International Union of Crystallography
Printed in Great Britain – all rights reserved

1 mmol of primary amino groups (determined by comparison with a serine standard) per minute, at pH 9.5 and 298 K. Amino groups are determined with OPA (*ortho*-phthaldialdehyde) reagent.* Specific activity was determined by active-site titration with *p*-nitrophenyl-*p'*-guanidinobenzoate. Enzymatic activity was also measured against two commonly used trypsin substrate analogues, Gly-Pro-Arg-*p*NA and Gly-Pro-Lys-*p*NA (*p*NA = *para*-nitroaniline). One *p*NA unit is defined as the amount of enzyme liberating 1 mmol of *para*-nitroaniline per minute at pH 9.5 and 303 K. *Para*-nitroaniline is determined spectrophotometrically at 405 nm (Sarath, de la Motte & Wagner, 1989). Relative to the proteolytic activity, *F. oxysporum* trypsin has much lower activity against peptide–*p*NA substrate analogues than does bovine trypsin (Table 1). The two trypsins have comparable proteolytic activities.

Table 1. *Specific activities against casein and pNA peptides*

Enzyme	Proteolytic activity (CPU g $^{-1}$)	*p*NA activity (units mg $^{-1}$) Gly-Pro-Lys	Gly-Pro-Arg
F. oxysporum trypsin	35	2	4
Bovine trypsin	26	90	340

Specificity

Substrate specificity was examined using native human insulin and the insulin analogue MI3 in which the N terminus of the *A* chain and the C terminus of the *B* chain are connected by the tripeptide Lys-Ala-Ala. Each substrate was incubated with *F. oxysporum* trypsin and bovine trypsin and the reaction mixtures were analysed by reverse-phase high-pressure liquid chromatography and amino-acid sequencing. The analysis showed that *F. oxysporum* trypsin cleaved insulin at Arg22 in the *B* chain, while bovine trypsin cleaved it at Lys29. *F. oxysporum* trypsin cleaved MI3 at Arg22 in the *B* chain and at Lys in the connecting tripeptide, while bovine trypsin cleaved at Lys29 in the *B* chain and at Lys in the connecting tripeptide. *F. oxysporum* trypsin has a preference for arginine while bovine trypsin has a preference for lysine at position P1 in the substrate (Murao, Ikenaka, Tonomura & Hitomi, 1985).

Inhibition

Like other trypsins, and unlike most microbial proteinases, the trypsin from *F. oxysporum* is inhibited by Kunitz soybean inhibitor and Kunitz trypsin inhibitor from bovine pancreas (BPTI) (Kassell, 1970; Laskowski & Kato, 1980).

Data collection and processing

Crystals of DFP (diisopropylfluorophosphate)-inhibited trypsin were prepared as described previously (Rypniewski *et al.*, 1993). The crystals belong to space group $P2_1$ with a = 33.3, b = 67.9, c = 39.8 Å and β = 107.3°. Diffraction data were collected from a single crystal (0.3 × 0.3 × 0.5 mm) on the EMBL X31 beamline at the DORIS storage ring, DESY, Hamburg, with an imaging-plate scanner built at EMBL (Hendrix & Lentfer, unpublished work). The crystal was mounted with the a^* axis approximately parallel to the spindle axis and c^* initially perpendicular to the direction of the X-ray beam. The oscillation angle was varied to minimize the overlapping of reflections. Three data sets (Table 2), each covering 90° rotation, were collected at different exposure times to record the whole range of intensities. A total of 134 images were recorded in 14 h. The integrated intensities were measured using the program *DENZO* (Otwinowski, 1993) and scaled together by the method of Fox & Holmes (1966). The initial data set contained 76 705 raw measurements to 1.55 Å. All reflections flagged as fully recorded were used in scaling. In initial scaling the relative temperature factors for all images were close to zero, indicating that there was no significant radiation damage. Subsequently, the images were scaled without a relative temperature factor. This had no significant effect on the scaling statistics. Adjacent partially recorded reflections were summed and unmatched partials with fractions greater than 0.5 were scaled for inclusion. This gave a total of 55 012 measurements, 41 843 fully recorded and 13 169 from partially recorded reflections. The merging R factor ($R_{merge} = \Sigma|I_i-\langle I\rangle|/\Sigma\langle I\rangle$, where I_i is an individual intensity measurement, and $\langle I\rangle$ is the average intensity for this reflection with summation over all the data) was 5.1%. Fig. 1 shows the variation of R_{merge} with resolution. The reduced data were 94.7% complete (Fig. 2) and contained 23 178 unique reflections between 10 and 1.55 Å. 89.3% of all reflections were greater than three standard deviations, 69% in the highest resolution bin. The intensities were converted to structure-factor amplitudes and a correction was applied to weak or negative measurements based on the *a priori* distribution (French & Wilson, 1978). The Wilson plot (Wilson, 1942) for the data is shown in Fig. 3. The fit to the theoretical straight line is satisfactory with the expected deviations for a protein in the low-resolution range. The overall temperature factor estimated from the plot is 22.4 Å^2.

Table 2. *Summary of data collection*

	High	Medium	Low
Maximum resolution (Å)	1.55	2.0	3.3
Crystal-to-plate distance (mm)	140	200	350
Oscillation per image	1.0–1.5	1.5–2.5	3.0–5.0
No. of exposures	70	39	22
Approximate time per exposure (s)	200	30	20
Wavelength (Å)	1.009	0.990	0.990
Overall $R(I)$ (%)	5.1		
Completeness (10–1.55 Å) (%)	94.7		

* The folder AF228 describing the analytical method is available from Novo Nordisk upon request.

Refinement

Refinement of the atomic coordinates and temperature factors was carried out by stereochemically restrained least-squares minimization (Konnert, 1976; Konnert & Hendrickson, 1980) using as a starting model the structure previously determined at 1.8 Å (Rypniewski *et al.*, 1993). The geometric restraints and the final values are listed in Table 3. A factor of 0.6 was applied to calculated shifts in position and temperature factors to give a reasonable rate of convergence while allowing for the diagonal approximation in the X-ray part of the normal matrix. The relative weight of the structure-factor component to the normal matrix was initially set to 1.0 and was reduced to 0.8 towards the end of the refinement. Data between 10 and 1.55 Å were used throughout the refinement. No σ cut-off was applied to the amplitudes. Rebuilding of the model was based on $(3F_o - 2F_c)$ and $(F_o - F_c)$ maps, using an Evans and Sutherland ESV graphics station and the program *FRODO* (Jones, 1978). Solvent molecules were added to the model both manually and by an automated procedure based on searching the Fourier difference maps for peaks within hydrogen-bonding distance from protein atoms or other water molecules. Water molecules lying in density less than $1/2\sigma$ above the mean density on the $(3F_o - 2F_c)$ map were periodically deleted from the model. All solvent molecules were refined with occupancy set to 1.0. From refinement cycle 20 onwards, H-atom positions were calculated before each cycle and their contribution to

Table 3. *Weighting parameters for the least-squares refinement and the final standard deviations*

Distances (Å)	σ*	Standard deviation	No. of parameters
Bond lengths (1-2 neighbours)	0.020	0.017	1610
Bond angles (1-3 neighbours)	0.040	0.039	2185
Dihedral angles (1-4 neighbours)	0.050	0.048	578
Planar groups	0.020	0.015	276
Chiral volumes	0.200	0.171	248
Non-bonded contacts (Å)†			
Single torsion	0.500	0.179	552
Multiple torsion	0.500	0.256	756
Possible hydrogen-bonding contacts	0.500	0.255	374
Torsion angles (°)			
Peptide plane (ω)	5.0	2.9	233
Staggered	15.0	12.9	238
Orthonormal	20.0	20.4	17
Thermal factors (Å^2)			
Main-chain bond	3.0	2.13	1007
Main-chain angle	4.0	2.86	1208
Side-chain bond	4.0	3.86	603
Side-chain angle	5.0	5.32	977

* The weights correspond to $1/\sigma^2$.
† Including symmetry-related contacts.

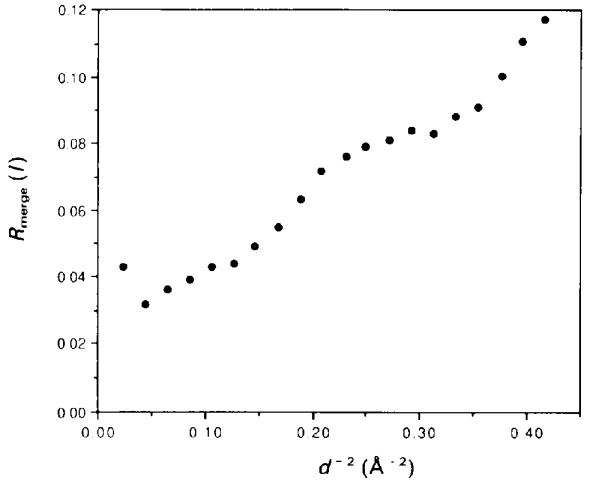

Fig. 1. $R_{merge}(I)$ as a function of $1/d^2$, where d is resolution defined in Å.

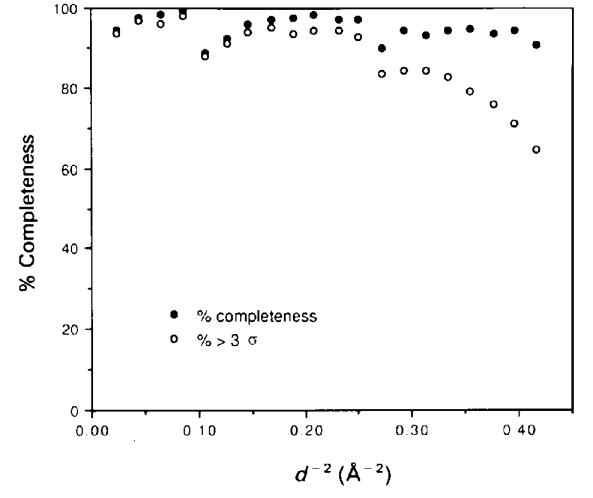

Fig. 2. Completeness of the X-ray data as a function of $1/d^2$ (d is resolution in Å) is indicated by filled circles. Open circles indicate reflections stronger than three standard deviations as a percentage of all possible reflections.

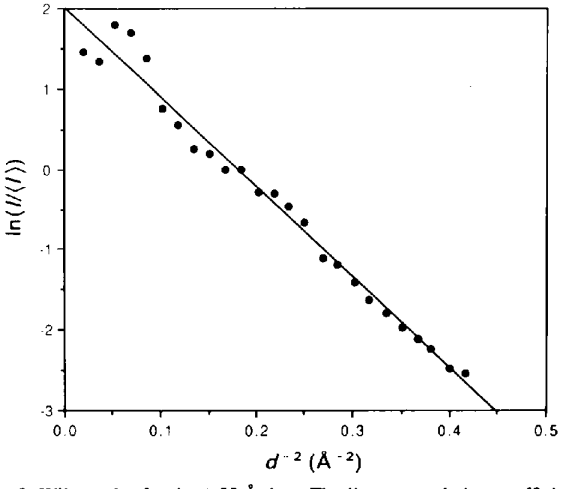

Fig. 3. Wilson plot for the 1.55 Å data. The linear correlation coefficient is 0.99. The overall temperature factor estimated from the gradient is 22.4 Å^2.

the structure factors was added to the structure factors calculated from the model. This gave a minor improvement in the R factor but significantly reduced the overall r.m.s deviation in the atomic temperature factors. The progress of the refinement is summarized in Fig. 4. Refinement was terminated when it was felt that no further significant improvement in the model could be achieved.

Accuracy of the model

The final R factor ($\Sigma||F_o| - |F_c||/\Sigma|F_o|$) is 14.4%. The estimated average error in the coordinates from the σ_A plot of Read (1986), Fig. 5, is 0.18 Å. The Luzzati (1952) plot (Fig. 6) gives a similar estimate of the average error. A weight of 0.8 for the X-ray component of the normal matrix relative to the stereochemical parameters, was sufficient to obtain good stereochemistry of the model (Table 3). The Ramachandran plot (Ramakrishnan & Ramachandran, 1965) for the refined coordinates is shown in Fig. 7. All values of the dihedral angles lie in the stereochemically allowed regions. Over 90% of those lie in the most favoured regions as defined by Morris, MacArthur, Hutchinson & Thornton (1992), based on a set of well refined protein models.*

The final ($F_o - F_c$) map was essentially flat with a standard deviation of 0.07 e Å^{-3}. The highest and the lowest features were 0.32 and −0.34 e Å^{-3}, respectively. The ($3F_o - 2F_c$) Fourier synthesis was calculated on an approximately absolute level, based on the number of electrons in the model. The average density was 0.35 e Å^{-3} with a root-mean-square deviation of 0.48 e Å^{-3}. Only one main-chain atom (Ser149O) lay in density below the 1σ level. Several side chains lie in poor density. These will be discussed below.

The overall structure

The refined model consists of 1557 protein atoms, 400 water molecules, one monoisopropyl phosphoryl group covalently bonded to the catalytic serine and one molecule

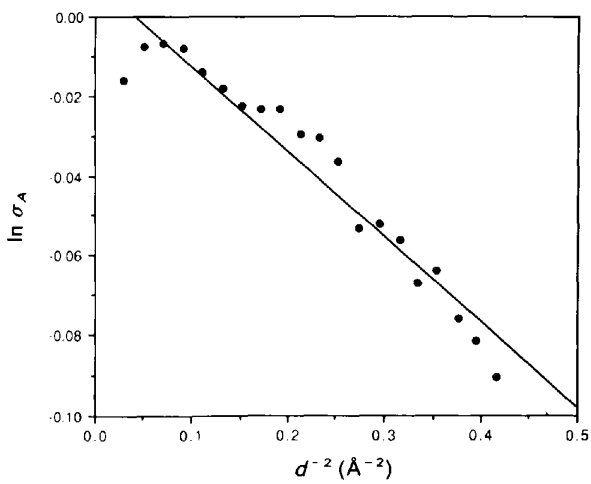

Fig. 5. The σ_A plot for the final model of *F. oxysporum* trypsin. The linear correlation coefficient is 0.96. The overall error estimated from the plot is 0.18 Å.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TRY, R1TRYSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: JN007). At the request of the authors, the atomic coordinates and structure factors will remain privileged until 1 January 1996.

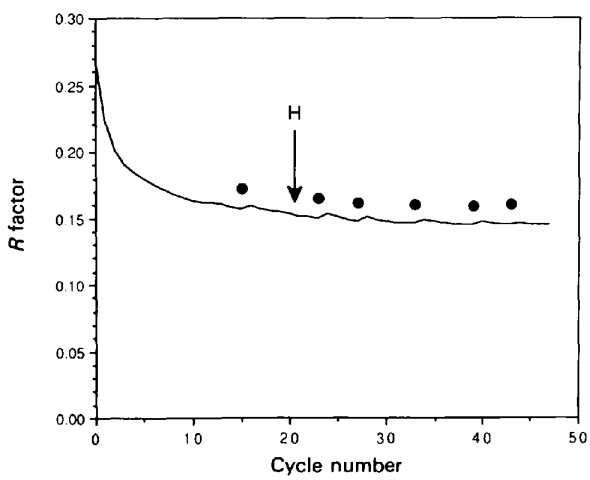

Fig. 4. Progress of the refinement: R factor ($\Sigma||F_o| - |F_c||/\Sigma|F_o|$) against the cycle number of the least-squares refinement. Dots indicate manual rebuilding of the model. The point marked H indicates where the contribution of H atoms was included in the refinement.

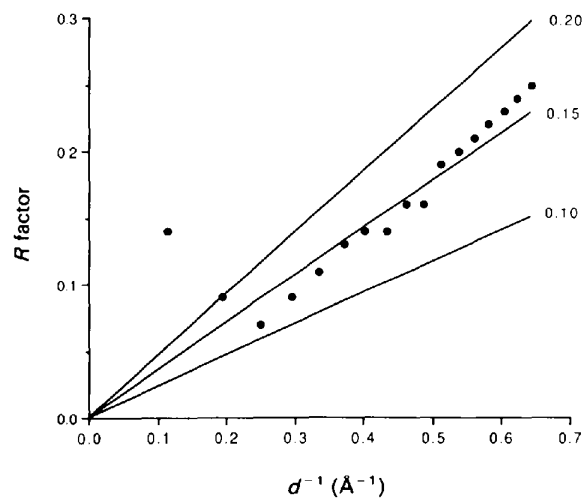

Fig. 6. Plot of the R factor as a function of inverse resolution. The straight lines show the theoretical dependence of the R factor on resolution for the mean coordinate errors shown on the right (Luzzati, 1952).

of isopropanol. The overall topology of the protein is shown schematically in Fig. 8. The folding pattern closely resembles that observed in other trypsin structures. The molecule consists of two domains. The core of each domain is a six-stranded antiparallel β-barrel with the first four strands forming a Greek key motif (Richardson, Richardson, Thomas, Silverton & Davies, 1976). Although no significant amino-acid sequence homology has been observed between the two domains in any trypsins, the topological equivalence between the two domains has been investigated by McLachlan (1979) and held as evidence of an early gene-duplication event. The three disulfide bridges are in equivalent positions to those observed in trypsin from *S. griseus* (Read & James, 1988).

Comparison with 1.8 Å *F. oxysporum trypsin structure*

The structure of *F. oxysporum* trypsin was previously refined at 1.8 Å resolution, using synchrotron data collected on film (Rypniewski *et al.*, 1993). It is interesting to compare the 1.55 Å model with the 1.8 Å model to see what structural errors still exist at 1.8 Å resolution and why they were not detected. One common error arises from building the carboxamide groups of glutamines and asparagines with the wrong polarity. It is usually difficult to determine the orientation of the carboxamide groups based on hydrogen-bonding pattern alone and a high-resolution electron-density map is required to distinguish O atoms from N atoms. In both the 1.8 and 1.55 Å structures the temperature factors were routinely examined in the course of refinement and, if necessary, the carboxamide groups were rotated to minimize the difference in the *B* factors. In addition, in the 1.55 Å structure it was often possible to see the difference between the N atoms and the O atoms by examining directly the electron-density maps. Of the 17 asparagines and glutamines in *F. oxysporum* trypsin four (Asn37, Asn100, Asn203 and Asn217) have been corrected in the 1.55 Å resolution structure. Several other side chains were refitted in the 1.55 Å structure (Gln64, Ser108, Arg122, Thr133, Ser157, Val184, Asn217, Tyr241). The second source of error arises from disorder, both static and dynamic, in residues and side chains. The electron density for Gln60 is ambiguous and it probably has two alternative conformations that overlap to a large extent. It has been fitted differently in the 1.55 Å structure to the 1.8 Å structure, but both conformations are plausible for both structures. Another ambiguity concerns serines, of which four (Ser21, 34, 61, 186) had been interpreted as double conformers in the 1.8 Å structure. It was then observed that several more residues, mostly serines, had densities indicative of multiple conformations. Of the serines above, Ser34 and Ser61 are now interpreted as having a single conformation but five additional serines have been interpreted as double conformers (Ser84, 109, 145, 190 and 236). This interpretation is supported by plausible hydrogen-bonding interactions with solvent and protein. It is difficult at this level to distinguish errors in interpretation from structural differences arising from slightly different crystallization conditions. A shift in position of up to 0.8 Å has been observed between the two models at Pro92 and adjacent residues on the molecular surface. The difference is unlikely to arise from an error in interpretation as in both models the residues fit the electron density well, although in the 1.55 Å structure the density is more clear. Similarly, in the 1.8 Å structure the side chain of Asn177 had weak but clear density indicating two alternative conformations. In the 1.55 Å structure there is only one conformation. After omitting the above residues and side chains the r.m.s. difference in position between the two models is 0.21 Å.

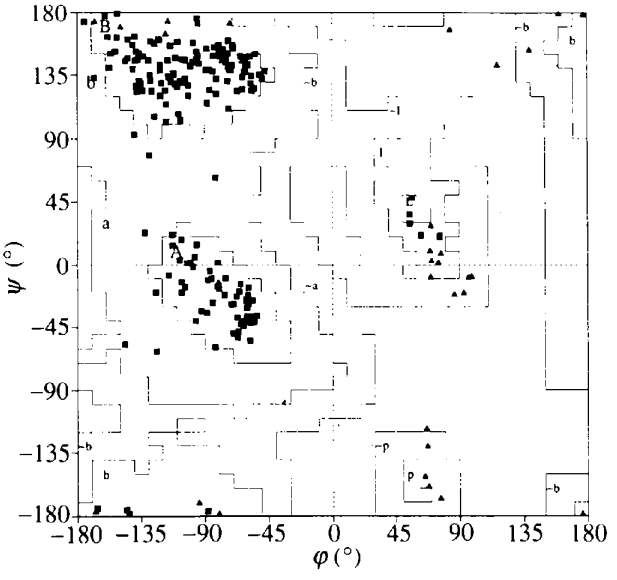

Fig. 7. Ramachandran plot of the (φ,ψ) angles for the refined model of *F. oxysporum* trypsin. Over 90% of the non-glycine residues (shown as squares) lie in the most favoured regions (*A*, *B*, *L*) defined by Morris *et al.* (1992), The remainder of non-glycine residues lie in the additional allowed regions (*a*, *b*, *l*, *p*). Glycines are shown by triangles.

Comparison with other trypsins

For comparison, the structures of *F. oxysporum*, *S. griseus* (Read & James, 1988) and bovine (Chambers & Stroud, 1977; 1979) trypsins, were initially superimposed by least-squares minimization of coordinates of 58 residues conserved in all three structures (Fig. 9). All atoms of those 58 residues were used in the initial minimization. The residues or side chains which deviated in position by more than three standard deviations were then omitted before the final superposition. The rejected residues were Gly148 and Ser150 which surround a deletion point in *S. griseus* trypsin, the side chains of Gln192 which is poorly ordered in all three structures, and Lys107 which is has a well defined electron density

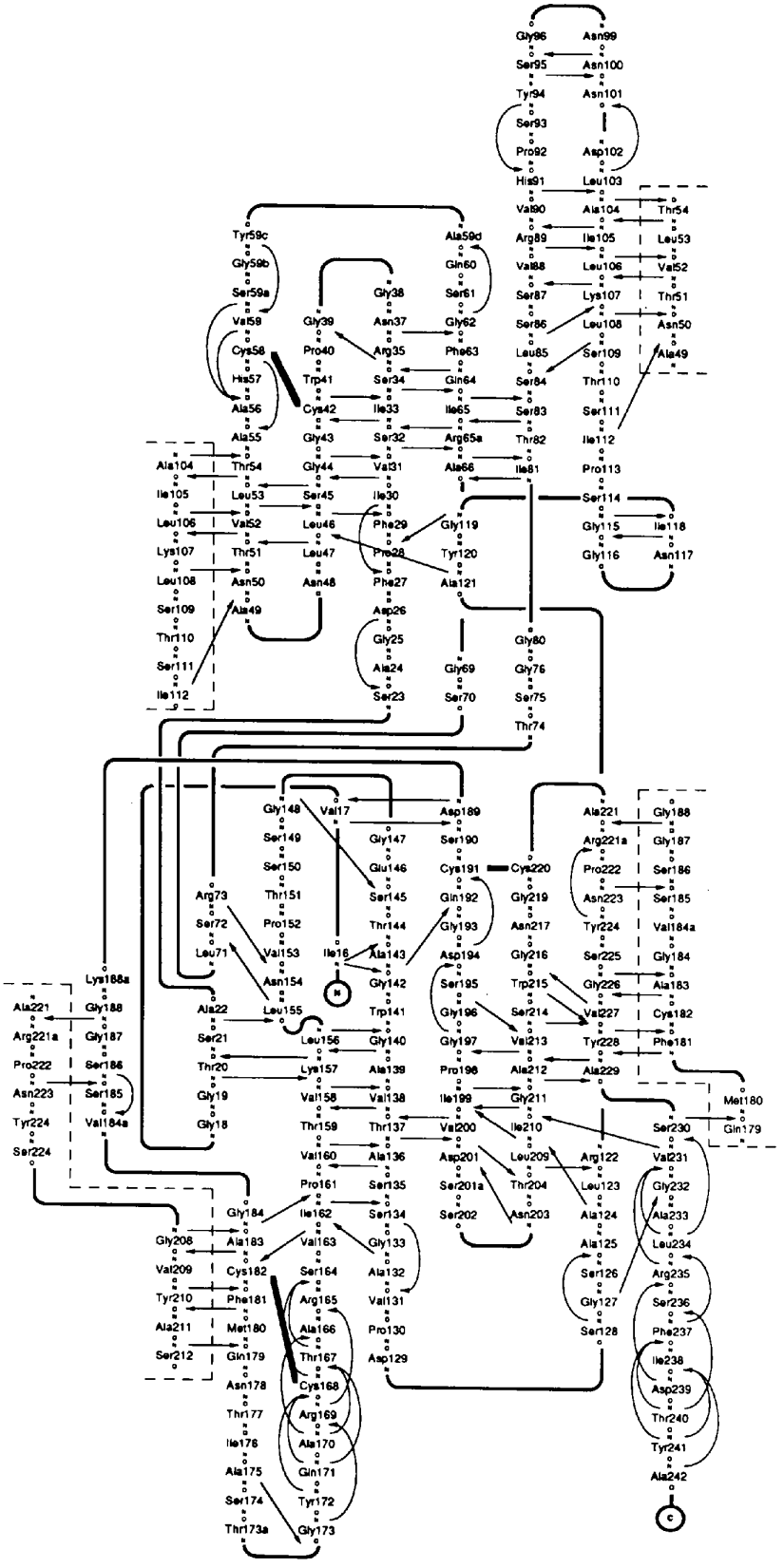

Fig. 8. Schematic representation of the hydrogen bonds between the main-chain atoms of *F. oxysporum* trypsin. Residues which appear twice due to 'unfolding' of the β-barrel in this two-dimensional representation are indicated by dashed lines. The disulfide bridges are represented by thick bars. Residues are numbered according to the convention used in other trypsins, based on bovine chymotrypsinogen (Hartley & Kauffman, 1966).

but quite different conformation in all three structures, due to amino-acid substitutions at the neighbouring position 51. After these additional rejections the final r.m.s. difference in position between *F. oxysporum* and bovine trypsins was 0.64 Å, between *F. oxysporum* and *S. griseus* trypsins 0.58 Å and between bovine and *S. griseus* trypsins 0.68 Å. Fig. 10 shows the superposition of *F. oxysporum* and bovine trypsins.

The catalytic site

Fig. 11 shows the region around the bound inhibitor. The region is expected to resemble the tetrahedral transition state of the reaction. There is density for only one of the two isopropyl groups present in DFP. The other group is most likely to have been cleaved (*c.f.*, Krieger, Kay & Stroud, 1974). Another possibility is that the second isopropyl group undergoes substantial thermal motion but even then one would expect to observe some residual density. The density for the first isopropyl group is ambiguous and probably reflects static disorder of the group. No satisfactory fit for a second conformation was achieved and the group was finally left as a single conformer.

The binding cavity in *F. oxysporum* trypsin is smaller than that in bovine (Chambers & Stroud, 1979) and *S. griseus* (Read & James, 1988) trypsins. In bovine trypsin the side chain of Phe41 points away from the active site. In *F. oxysporum* the corresponding residue is Trp41 whose side chain is rotated 130° from the plane of the phenyl ring of Phe41 in bovine trypsin. This brings the side chain of Trp41 into the binding cavity. The position of Trp41 is stabilized by the side chain of Arg35 which, in turn, is stabilized by the side chains of Tyr59C and Phe63. Tyr59C also contributes directly to the narrowing of the binding site. Both Arg35 and Tyr59C are insertions with respect to bovine trypsin, while Phe63 replaces the smaller isoleucine in the bovine enzyme. In addition to the repositioning of the side chains, the whole loop of residues 35 to 41 is shifted towards the active site. This is facilitated by the steric constraint imposed by Pro40. The shifting of the loop is accompanied by a concerted shift of β-strands 63–66 and 81–84, and loop 109–113. The steric constraint imposed by Pro113, which replaces serine in bovine trypsin, is consistent with this shift. The narrowing of the binding cavity in *F. oxysporum* trypsin in comparison with bovine trypsin explains the relative inactivity of *F. oxysporum* trypsin towards peptide–*p*NA substrate analogues. The binding of peptide–*p*NA would be obstructed by a steric collision between the bulky *para*-nitroanilino group and the side chains of Trp41 and Tyr59C. Similarly, superposition of *F. oxysporum* trypsin and bovine trypsin complexed with bovine pancreatic trypsin inhibitor (BPTI) (Marquart, Walter, Deisenhofer, Bode & Huber, 1983; PDB code 2PTC) suggests that BPTI cannot bind to *F. oxysporum* trypsin in the same way as it binds to bovine trypsin because of a severe steric clash between Trp41 and Ile18 of BPTI (Fig. 12). Although the biochemical results presented above show that BPTI does inhibit *F. oxysporum* trypsin, the interaction must involve either a rearrangement of the side chain of Trp41 and the associated residues listed above or a conformational change in BPTI.

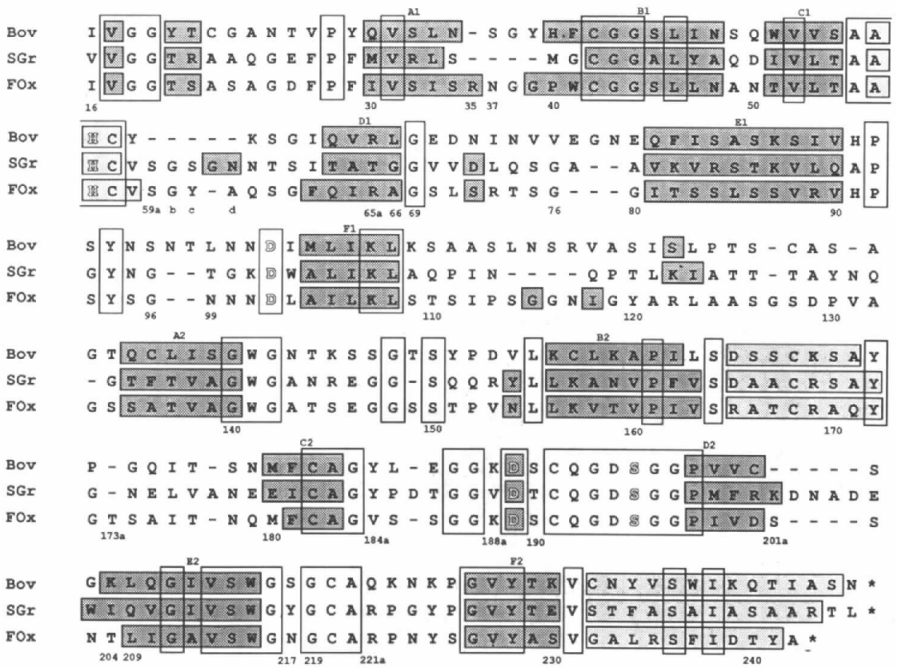

Fig. 9. Amino-acid sequence alignment of bovine (Bov), *S. griseus* (SGr) and *F. oxysporum* (FOx) trypsins, based on comparison of crystal structures. The 58 conserved residues, indicated by boxes, were used in the final superposition. Light shading is used for helices and dark shading for β-strands and isolated β-bridges, as defined by Kabsch & Sander (1983). The β-strands A1–F1 and A2–F2 form the N-terminal and the C-terminal domains, respectively (*c.f.* Fig. 8). Catalytic triad residues His57, Asp102, Ser195 and Asp189, which determines tryptic P1 specificity, are highlighted. The residues are numbered as described in Fig. 8.

Specificity pocket

In trypsins, Asp189 is the residue that determines substrate specificity for basic amino acids at the P1 position. In *F. oxysporum* trypsin residues 216–219, which form the rim of the specificity pocket, are further away from Asp189 than the equivalent residues in bovine trypsin. Also, residue 225, which is serine in *F. oxysporum* trypsin and proline in all other currently known trypsins, is shifted by about 1 Å. The overall effect is an enlargement of the P1 specificity pocket in *F. oxysporum* trypsin relative to bovine trypsin (Fig. 13). This explains the P1 substrate preference of *F. oxysporum* trypsin for arginine over lysine and the preference of the bovine enzyme for lysine over arginine. The bulkier side chain of arginine could fit more easily into the larger binding pocket of *F. oxysporum* trypsin. In *S. griseus* trypsin (Read & James, 1988) the P1 binding pocket is more similar to *F. oxysporum* trypsin than to bovine trypsin. It would be interesting to examine the P1 specificity of *S. griseus* trypsin which, based on the above criteria, should preferentially cleave on the carboxyl side of arginine. Unfortunately, no relevant data are available for *S. griseus* trypsin (James, personal communication).

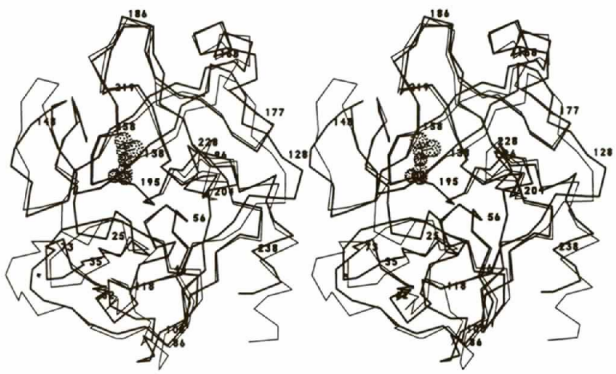

Fig. 10. Superposition of *F. oxysporum* trypsin (thick lines) and bovine trypsin (thin lines). The molecules were superimposed using the conserved residues as described in the text. Only Cα atoms are shown. Every tenth residue is labelled. The calcium ion bound to bovine trypsin is marked with a cross in the lower, left part of the figure. The monoisopropyl phosphoryl inhibitor group bound in the active site of the *F. oxysporum* trypsin is represented by the dotted region.

Other regions

The largest structural differences between *F. oxysporum* trypsin and other trypsins occur in loop regions (Fig. 10). The loop 74–80, which in bovine trypsin binds calcium, has different conformations in the various trypsins although the inter-domain interactions of nearby residues 72 and 73, with residues 153 and 155, are conserved. The loop formed by residues 95–100 has quite different conformations in all three enzymes but the side chain of the preceding Tyr94, which is in contact with His57 of the 'catalytic triad', maintains the same position. The sharp turn formed by residues 113–117 contains two glycines and a proline. In bovine trypsin the corresponding residues follow an approximately similar path despite the lack of sequence homology, while in *S. griseus* trypsin this loop is absent. Residues 120–124 maintain a similar conformation in all three structures. There is no clear sequence homology for these residues but a pattern of alternating hydrophilic and hydrophobic side chains is maintained, with the hydrophobic side chains pointing towards the bulk of the protein. Residues 125–134, which link the two structural domains in trypsin, show high variability in both sequence and structure, between the various trypsins. The loop 145–151 has weak electron density in *F. oxysporum* trypsin and makes poorer contact with the bulk of the protein than the corresponding loops in bovine and *S. griseus* trypsins. The loop 201–204 in *F. oxysporum* trypsin contains an insertion of one residue compared to bovine trypsin while in *S. griseus* trypsin the loop contains five more residues than the bovine enzyme. The insertions have only a local effect on the structure. It is interesting to compare the cluster of hydrophobic residues at positions 103, 212 and 237. The hydrophobic side chain of residue 237 anchors the C-terminal α-helix to the surface of the molecule. In *F. oxysporum* the residue is Phe, in bovine trypsin it is Trp, and in *S. griseus* trypsin it is

Fig. 11. The electron density ($3F_o - 2F_c$) for the monoisopropyl phosphoryl inhibitor group bound to Ser195. Catalytic His57 is shown on the right. Gln192, which partly closes the binding site is shown on the left. The density for the side chain is weak. The map was contoured at the 1σ level.

Ala. The size and conformation of residues 103, 212 and 91 change accordingly, illustrating the necessity for complementary mutations in order to preserve the integrity of the protein interior.

Temperature factors

The mean B value for all atoms is 23.1 Å, very close to the overall B value of 22.4 Å obtained from the Wilson plot (Fig. 3). The average temperature factor for all protein atoms is 16.3 Å^2. The mean value for the main chain is 14.5 Å^2 and for the side chains is 18.8 Å^2. Standard deviations of the mean temperature factor per residue were 5.3 and 8.4 Å^2 for the main chain and side chains, respectively. The average B value for solvent molecules is 49.6 Å^2. Fig. 14 shows a plot of the B values, averaged for the main chain and the side chains, as a function of residue number. Residues Ser149 and Ser150 have temperature factors greater than three standard deviations above the average for both main chain and side chains. These residues lie in a loop exposed to the solvent, in poor electron density. The corresponding loop in the crystal structure of trypsin from North Atlantic salmon was also reported to be disordered (Smalås & Hordvik, 1993). Similar to trypsin from salmon, *F. oxysporum* trypsin shows increasing mobility towards the end of the C-terminal helix, reflected in rising B values and poorly defined electron density. The solvent-exposed side chains of Arg122 and Tyr241 are also poorly ordered and both have B values greater than three standard deviations.

Disorder

Seven, out of a total of 38, serine residues have been modelled with two alternative conformations of the side chain. These are serines 21, 84 109, 145, 186, 190 and 236. The criteria used in identifying double conformers were both the appearance of peaks in difference electron-density maps and plausible hydrogen-bonding interactions with solvent or protein. The side chain of Arg89 has also been modelled with two conformations although the density is poor. The density for some other surface side chains suggests multiple conformations although at the present resolution it is not sufficiently clear to model satisfactorily. Although multiple conformations in

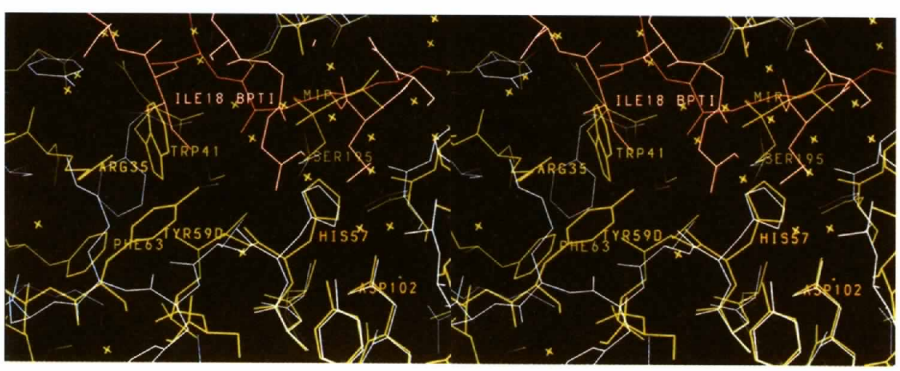

Fig. 12. Detail of the binding pocket in *F. oxysporum* trypsin (yellow). The catalytic triad residues Ser195, His57 and Asp102 are shown on the right. The monoisopropyl phosphoryl group (MIP) is bound to Ser195. The model of bovine trypsin (blue) complexed with BPTI (pink) has been superimposed as described in the text. Considerable structural differences, restricting the binding pocket of *F. oxysporum* trypsin as compared to bovine trypsin, are shown on the left. A structural rearrangement is expected to occur upon binding of BPTI to *F. oxysporum* trypsin to avoid a clash between Ile18 of BPTI and Trp41 of *F. oxysporum* trypsin.

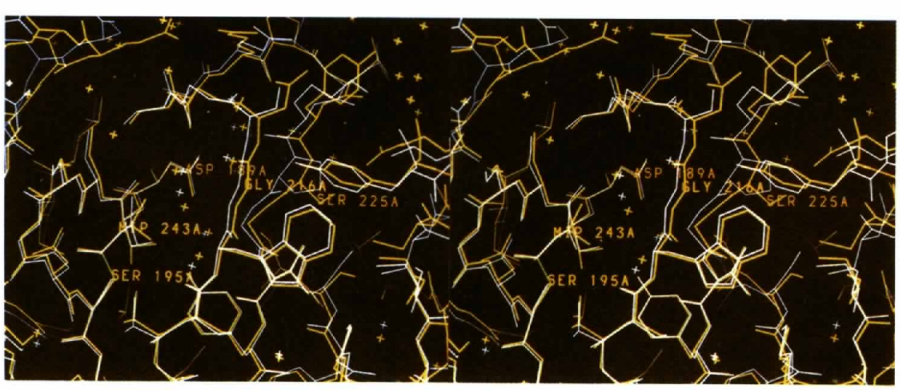

Fig. 13. Comparison of the P1 specificity pocket of *F. oxysporum* trypsin (yellow) with bovine, DFP inhibited trypsin (blue). The coordinates were aligned using Asp189, which determines tryptic specificity, and the catalytic Ser195. The entrance to the pocket is at the monoisopropyl phosphoryl group (MIP) bound to the catalytic Ser195. The largest structural differences are in the depth of the pocket, at Ser225.

protein structures are not frequently reported, they are, in the authors' experience, not uncommon in structures for which good high-resolution data are available.

The solvent

The model includes 400 water molecules. In identifying water molecules peaks greater than three standard deviations from the mean were selected from the $(F_o - F_c)$ map and tested if they lay within a hydrogen-bonding distance to protein atoms or other solvent molecules. Waters were removed from the model if, in the course of refinement, their corresponding electron density fell below $1/2\sigma$ above the mean on the $(3F_o - 2F_c)$ map, which corresponds to 0.24 e Å^{-3} on the final $(3F_o - 2F_c)$ map. This density also corresponds to the 3σ level on the final $(F_o - F_c)$ map. More waters were included or deleted during manual inspection of the difference maps. The advantage of applying the criteria of significant density and plausible hydrogen-bonding potential over simply rejecting waters above a certain B value is that it allows inclusion of partially occupied solvent sites. Since no attempt was made to refine the occupancies at this resolution, the partial occupancy is reflected in B values higher than might be expected for fully occupied sites. The distribution of the temperature factors for the 400 water molecules in the refined model (Fig. 15) gives some indication of the successive solvation shells around the protein. Fig. 16 shows the internal and cavity water molecules. One interesting feature is the absence of internal waters in the bulk of the N-terminal domain. Internal waters are found either in the C-terminal domain or at the interdomain interface. Most internal waters form clusters and channels. One extensive well defined water channel runs alongside residues 25–43 and makes contact with the highly conserved Trp141 and Leu155 (Rypniewski *et al.*, 1994). The loop formed by residues 70–73 defines the channel on the outer side of the molecule. Most of these waters are conserved, compared to bovine and *S. griseus* trypsins, but in bovine trypsin a substitution of Pro40 by histidine displaces three of the waters. Two of the same waters are displaced in *S. griseus* trypsin where Ser32 is replaced by an arginine. A large cluster of internal waters is associated with Asp189 which defines the P1 substrate specificity. The water solvates the charged but buried aspartate and forms a channel from the bottom of the specificity pocket, through the loop of residues 221–224, to the surface of the protein. Another cluster of water molecules is associated with the salt bridge between Asp129 and Arg165. The corresponding region in *S. griseus* trypsin forms a calcium-binding site. One feature of the difference maps, observed previously in the 1.8 Å structure, could not be satisfactorily modelled by water molecules and was tentatively identified as a molecule of isopropanol. The molecule is in contact with the side chains of Asn178, Arg165 and Gly116 of the neighbouring protein molecule. Isopropanol was used as solvent of the inhibitor DFP and was present in the crystallization mixture at a concentration of 5%(*v*/*v*). It was later found to facilitate crystallization of uninhibited trypsin.

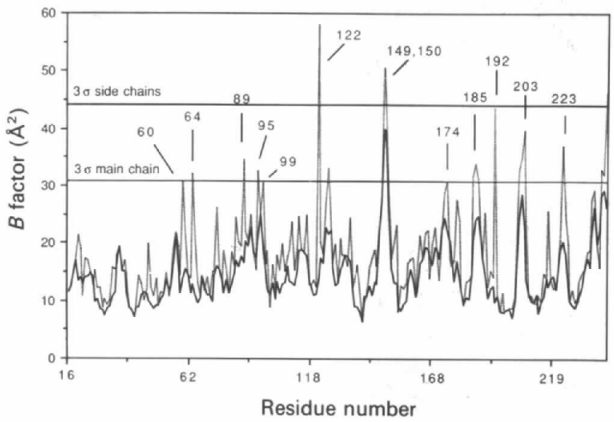

Fig. 14. Variation of the isotropic B values (Å^2) averaged for the main-chain (solid line) and the side-chain atoms (dashed line) for each residue. Horizontal lines mark the 3σ level for the main chain and the side chains. Some peaks have been labelled with corresponding residue numbers. Residues with B values greater than three standard deviations above the mean are discussed in the text.

Concluding remarks

The model of *F. oxysporum* trypsin presented, refined against high-quality synchrotron data at 1.55 Å resolution, is now comparable in accuracy to the other well refined trypsins. Comparison of trypsins from distantly related species, combined with biochemical studies, reveals the structural basis for differences in enzyme activity and P1 specificity.

Despite a large amount of structural and biochemical data on trypsins the overall picture is fragmentary.

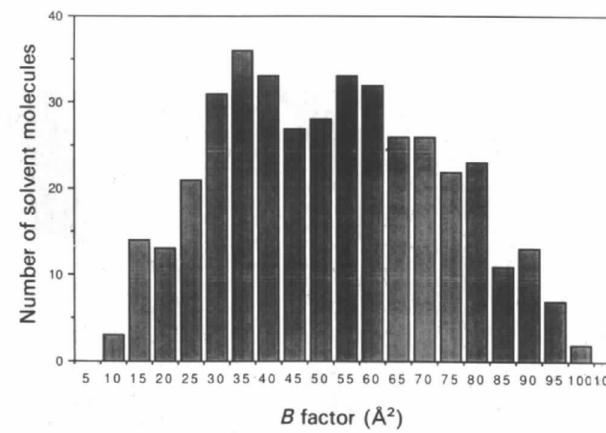

Fig. 15. Histogram of the B values (Å^2) of the solvent molecules.

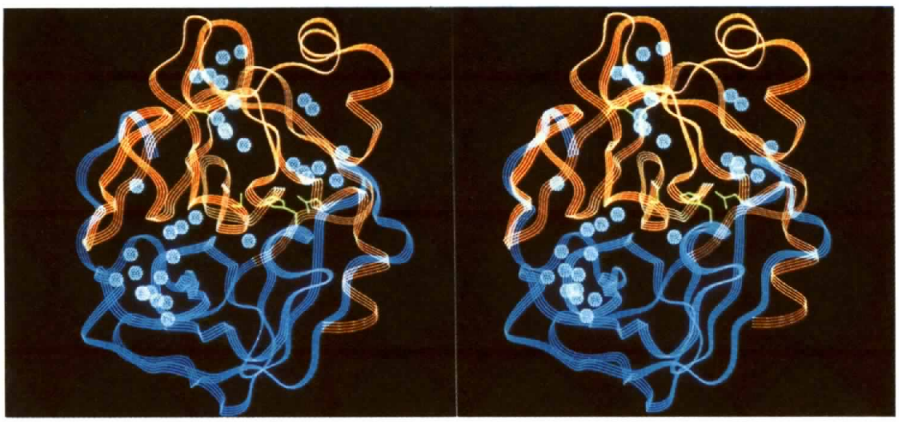

Fig. 16. Internal waters in *F. oxysporum* trypsin. The N-terminal domain is coloured blue and C-terminal domain orange. The catalytic triad residues (His57, Asp102 and Ser195) and Asp189 which determines tryptic P1 specificity are also illustrated. The water channel that runs along residues 25–43 is visible to the left. The cluster of waters associated with the P1 specificity pocket is in the upper part of the picture and the cluster associated with the salt bridge Asp129–Arg165 is on the right (see description in the text). The orientation of the molecule is similar to that shown in Fig. 10.

Many trypsins are known only from their nucleotide sequences. Several others have been studied extensively. One problem in comparing enzymatic properties of different trypsins is that substrate specificities vary. In this report the comparison of the activities of *F. oxysporum* and bovine trypsins against the same substrates reveals a difference in relative P1 specificities for Lys and Arg. The reason for the difference is suggested by comparing the three-dimensional structures and is verifiable by similar studies on other trypsins. Another difficulty in comparing different trypsins is in defining the extent of the substrate-binding site and in identifying its important features. The size of the substrate and studies of enzyme–inhibitor complexes indicate that the substrate-binding site extends beyond the immediate vicinity of the catalytic serine and the P1 specificity pocket. One approach to characterizing the significant features of the substrate-binding site involves examining the available trypsins for conserved residues (Rypniewski *et al.*, 1994). The comparison presented here of *F. oxysporum* trypsin with bovine trypsin–BPTI complex indicates that the smaller binding cavity on the C-terminal side of the binding peptide in *F. oxysporum* trypsin does not result in decreased proteolytic efficiency relative to the bovine enzyme.

A number of features in the electron-density maps remain unclear. Most of these can be explained by static disorder in the inhibitor and a number of side chains. Attempts will be made to extend the data to near atomic resolution. Preliminary results suggest that this should be possible if sufficiently large single crystals can be obtained. *F. oxysporum* trypsin shows considerable resistance to autolysis. The enzyme can be kept in solution, at room temperature, for several days without substantial degradation taking place. This made it possible to crystallize *F. oxysporum* trypsin in the absence of an inhibitor. The triclinic crystals diffract to near atomic resolution and structure solution of the native form of the enzyme is currently in progress. The inactive zymogen, containing a propeptide at the N terminus has also been purified although it has so far resisted our attempts to crystallize it.

MD was supported by a grant from Novo Nordisk. WR is partially supported by the European Science Foundation.

References

ABOLA, E. E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WENG, J. (1987). *Crystallographic Databases – Information Content, Software Systems, Scientific Applications* edited by F. H. ALLEN, G. BERGERHOFF & R. SIEVERS, pp. 107–132. Bonn/Cambridge/Chester: IUCr.

BERNSTEIN, F. C., KOETZLE, T.F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHOMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542

CHAMBERS, J. L. & STROUD, R. M. (1977). *Acta Cryst.* B**33**, 1824–1837.

CHAMBERS, J. L. & STROUD, R. M. (1979). *Acta Cryst.* B**35**, 1861–1874.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

FRENCH, S. & WILSON, K. S. (1978). *Acta Cryst.* A**34**, 517–515.

HARTLEY, B. S. & KAUFFMAN, D. L. (1966). *Biochem. J.* **101**, 229–231.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KASSELL, B. (1970). *Methods in Enzymology*, Vol. 19, *Proteolytic Enzymes*, edited by C. E. PERLMANN & L. LORAND, pp. 844–852. New York/London: Academic Press.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

KRAUT, J. (1977). *Annu. Rev. Biochem.* **46**, 331–358.

KRIEGER, M., KAY, L. M. & STROUD, R. M. (1974). *J. Mol. Biol.* **83**, 209–230.

LASKOWSKI, M. JR & KATO, I. (1980). *Annu. Rev. Biochem.* **49**, 593–626.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MCLACHLAN, A. D. (1979). *J. Mol. Biol.* **128**, 49–79.

MARQUART, M., WALTER, J., DEISENHOFER, J., BODE, W. & HUBER, R. (1983). *Acta Cryst.* B**39**, 480–490.

MORRIS, A. L., MACARTHUR, M. W., HUTCHINSON, E. G. & THORNTON, J. M. (1992). *Proteins*, **12**, 345–364.

MURAO, S., IKENAKA, T., TONOMURA, B. & HITOMI, K. (1985). *Protein Protease Inhibitor – The Case of Streptomyces Subtilisin Inhibitor*, edited by K. HIROMI, K. AKASAKA, Y. MITSUI, B. TONOMURA & S. MURAO, pp. 1–14. Amsterdam: Elsevier.

OLAFSON, R. W., JURÁSEK, L., CARPENTER, M. R. & SMILLIE, L. B. (1975). *Biochemistry*, **14**, 1168–1177.

OLAFSON, R. W. & SMILLIE, L. B. (1975). *Biochemistry*, **14**, 1161–1167.
OTWINOWSKI, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, pp. 56–62. Warrington: SERC Daresbury Laboratory.
RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
READ, R. J. & JAMES, M. N. G. (1988). *J. Mol. Biol.* **200**, 523–551.
RICHARDSON, J. S., RICHARDSON, D. C., THOMAS, K. A., SILVERTON, E. W. & DAVIES, D. R. (1976). *J. Mol. Biol.* **102**, 221–235.
RYPNIEWSKI, W. R., HASTRUP, S., BETZEL, CH., DAYTER, M., DAUTER, Z., PAPENDORF, G., BRANNER, S. & WILSON, K. S. (1993). *Protein Eng.* **6**, 341–348.
RYPNIEWSKI, W. R., PERRAKIS, A., VORGIAS, C. E. & WILSON, K. S. (1994). *Protein Eng.* **7**, 57–64.
SARATH, G., DE LA MOTTE, R. S. & WAGNER, F. W. (1989). *Proteolytic Enzymes, a Practical Approach*, edited by R. J. BEYNON & J. S. BOND, pp. 25–55. New York: IRL Press/Oxford Univ. Press.
SMALÅS, A. O. & HORDVIK, A. (1993). *Acta Cryst.* D**49**, 318–330.
SPRANG, S., STANDING, T., FLETTERICK, R. J. & STROUD, R. M. (1987). *Science*, **237**, 905–908.
STEITZ, T. A. & SHULMAN, R. G. (1982). *Annu. Rev. Biochem. Biophys.* **11**, 419–444.
WILSON, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
YAMANE, T., KOBUKE, M., TSUTSUI, H., TOIDA, T., SUZUKI, A., ASHIDA, T., KAWATA, Y. & SAKIYAMA, F. (1991). *J. Biochem.* **110**, 945–950.

Acta Cryst. (1995). D**51**, 85–92

Time-Averaging Crystallographic Refinement: Possibilities and Limitations Using α-Cyclodextrin as a Test System

By Celia A. Schiffer

Laboratory of Physical Chemistry, ETH-Zentrum, 8092 Zürich, Switzerland

Piet Gros

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

and Wilfred F. van Gunsteren

Laboratory of Physical Chemistry, ETH-Zentrum, 8092 Zürich, Switzerland

(*Received* 29 *November* 1993; *accepted* 21 *June* 1994)

Abstract

The method of time-averaging crystallographic refinement is assessed using a small molecule, α-cyclodextrin, as a test system. A total of 16 refinements are performed on simulated data. Three resolution ranges of the data are used, the memory relaxation time of the averaging is varied, and several overall temperature factors are used. The most critical factor in the reliable application of time-averaging is the resolution of the data. The ratio of data to molecular degrees of freedom should be large enough to avoid overfitting of the data by the time-averaging procedure. The use of a free *R* factor can aid in determining whether time-averaging can be reliably applied. Good ensembles of structures are obtained using data up to 1.0 or 2.0 Å resolution. Comparison of electron-density maps from time-averaging refinement and anisotropic temperature-factor refinement indicates that the former technique yields a better representation of the exact data than the latter.

Introduction

Crystallographic refinement is the technique of fitting a model structure to the experimentally observed diffraction intensities. In a perfect crystal at 0 K only one conformation of the molecular system exists. Thus, it is sufficient to represent the structure of the molecules by a single model structure. Often, however, in crystals of biological macromolecules there are regions in the asymmetric unit where the order of the crystal breaks down. Traditionally these regions are characterized by high temperature factors or disordered areas in the electron-density map.

In least-squares crystallographic refinement one structure is fitted within the constraints of chemistry to the experimental X-ray data. This method works as long as the starting configuration of the model is already close to the correct one. With the introduction of molecular dynamics (MD) (Brünger, Kuriyan & Karplus, 1987; Brünger, 1988; Fujinaga, Gros & van Gunsteren, 1989) to crystallographic refinement, the radius of convergence of the refinement increased substantially. By simulating the system at a non-zero temperature the kinetic energy of the system allows the model structure to overcome potential energy barriers and thus to search for a better fit with the experimental data. This method was still only used to search for a single model structure that fits the experimental data. This restriction prohibits an adequate analysis of regions where multiple molecular conformations exist.

The technique of time-averaging crystallographic refinement was introduced by Gros, van Gunsteren & Hol (1990). This method refines an ensemble of structures against the data rather than just a single structure. The ensemble consists of conformations of the molecule at time points along an MD trajectory in which a pseudo potential energy term restrains the molecular configurations such that the deviation of the amplitudes of the time-averaged structure factors from the observed amplitudes is minimal. Time-averaging refinement may distinguish whether a disordered part of the molecule adopts a multitude of configurations throughout the crystal, which would indicate a mobile region, or resides mainly in two or three distinct conformations, which reflects static disorder in the crystal. Better understanding of the flexibility in a molecule will aid in understanding how a particular molecule functions.

Gros *et al.* (1990) applied the time-averaging refinement technique to bovine pancreatic phospholipase A2. The resulting ensemble of protein structures showed a narrow distribution of atomic positions in secondary-structure elements and a wide variation of loop conformations. The latter were

©1995 International Union of Crystallography
Printed in Great Britain – all rights reserved