Acta Cryst. (1993). D**49**, 318–330

Structure Determination and Refinement of Benzamidine-Inhibited Trypsin from the North Atlantic Salmon (*Salmo salar*) at 1.82 Å Resolution

BY ARNE O. SMALÅS AND ASBJØRN HORDVIK

Protein Crystallography Group, Institute of Mathematical and Physical Sciences, University of Tromsø, 9000 Tromsø, Norway

(*Received* 29 *July* 1992; *accepted* 14 *December* 1992)

Abstract

The structure of the serine protease trypsin from the North Atlantic salmon (*Salmo salar*) has been solved by molecular replacement and refined by restrained least-squares methods to a conventional R factor of 16.4% using diffractometer data in the 6.0–1.82 Å resolution range (14443 reflections greater than 3σ). The model comprises 1793 protein atoms and 180 solvent molecules which were given unit occupancies, and the average temperature factors for protein atoms and solvent oxygen atoms are 15.2 and 36.8 Å^2, respectively. The estimated error in atomic positions is about 0.2 Å. The structure of salmon trypsin was solved and refined with only a small part of the amino-acid sequence known. However, a gene sequence of salmon trypsin has later become available. Some discrepancies between this sequence and the sequence obtained from the present X-ray crystal study indicate that the mentioned sequences may correspond to isoenzymes. The structure of salmon trypsin is similar to other trypsins of known structure.

Introduction

Trypsin, a proteolytic enzyme belonging to the large group of serine proteases, has been found in numerous higher mammals and also in fish, plants, insects and bacteria (Zwilling, Pflederer, Sonnenborn, Kraft & Stucky, 1969). Most serine proteases are extracellular enzymes which are easy to separate and purify in relatively large quantities and this is probably one of the reasons why they have been so thoroughly studied.

Three-dimensional X-ray structures have been obtained for bovine trypsin (Marquart, Walter, Deisenhofer, Bode & Huber, 1983), mutant rat trypsin (Sprang, Standing, Fletterick & Stroud, 1987), and trypsin from the bacteria *Streptomyses griseus* (Read & James, 1988). Atomic coordinates for these molecular structures, with some coordinate sets as found in different crystal forms of the same enzyme, are deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

The pancreatic family of serine proteases, comprising chymotrypsin, elastase and trypsin, collaborate with the other pancreatic proteases in the digestion of proteins by endopeptidic cleavage, carried out in the duodenum and the small intestines. The pancreatic proteases are synthesized in the pancreas as inactive proenzymes to avoid autoproteolysis. They are secreted into the duodenum where they are activated by cleavage of one specific peptide bond. This leads to the conformational change that makes the enzyme active.

At present no three-dimensional X-ray structure of a fish trypsin has been reported. However, trypsins from several cold-blooded species have been studied enzymatically, namely trypsin from shrimp (Walsh, 1970), starfish (Winter & Neurath, 1970), dogfish (Titani, Ericsson, Neurath & Walsh, 1975), sardine (Murakami & Noda, 1981), capelin (Hjelmeland & Raa, 1982), the Greenland cod (Simpson & Haard, 1984), the chum salmon (Uchida, Tsukayama & Nishide, 1984; Uchida, Anzai & Nishide, 1986), the Atlantic cod (Asgeirsson, Fox & Bjarnason, 1989), and the North Atlantic salmon (Arnesen, 1989; Torrissen, 1984).

Fish trypsins resemble mammalian trypsins in many respects, but are also different in some respects. Trypsins from most marine organisms studied are anionic and in most cases more than one form is found (Asgeirsson *et al.*, 1989). For trypsin from the North Atlantic salmon, for example, there are three forms with pI values of 5.3, 4.8 and 4.7, respectively (Arnesen, 1989). For the bovine as well as for the porcine trypsins only one form of the enzyme has been found, namely the cationic form with a pI value of 10.8 (Bier & Nord, 1951).

Bovine trypsin is stable down to pH = 3, while most of the fish trypsins studied are unstable at low pH. Salmon trypsin is reported to be stable at pH levels from 6 to 10 (Arnesen, 1989).

Most of the fish trypsins have a temperature optimum similar to the mammalian enzymes, but several trypsins from marine sources are less stable at higher temperatures than their mammalian equivalents (Hjelmeland & Raa, 1982; Asgeirsson *et al.*,

0907-4449/93/030318-13$06.00

1989; Simpson & Haard, 1984; Osnes & Mohr, 1985; Arnesen, 1989).

The k_{cat} values for fish trypsins are higher and the K_m values lower than those for the bovine enzyme, and thus the fish trypsins have a higher catalytic efficiency (k_{cat}/K_m) than trypsins from mammals. It is found that salmon trypsin at 298 K hydrolyses the substrate BAPNA (*N*-α-benzoyl-DL-arginine-*p*-nitroaniline) 20 times more efficiently (expressed by k_{cat}/K_m) than the bovine enzyme does (Arnesen, 1989). This phenomenon may be explained by the need of ectothermic species to compensate for a lower body temperature (Asgeirsson *et al.*, 1989).

The lower working temperature found for fish trypsins and the other properties unique for these fish enzymes as described above, are most likely related to the three-dimensional structure, and interests in gene modification of fish trypsins have recently been expressed (Nordisk Industrifond Protein Engineering Programme, Inaugural Symposium, September, 1989). We therefore initiated a structure study of trypsins from different fish species (Smalås, 1990; Smalås, Hordvik, Hansen, Hough, Jynge, 1990), and the results from a 1.82 Å X-ray crystallographic structure study of trypsin from the North Atlantic salmon (*Salmo salar*) as found in crystal form I (Smalås *et al.*, 1990) are presented here.

Crystallization and data collection

Salmon trypsin was purified by affinity chromatography (Hjelmeland & Raa, 1982; Smalås *et al.*, 1990) involving inhibition of the enzyme by the substrate analogue benzamidine. The molecular weight of the benzamidine complex was found to be 23400 Da from SDS–PAGE measurements (Smalås, 1990).

Crystals were grown at 301 K by the method of hanging-drop vapour diffusion, using a protein concentration of approximately 15 mg ml^{-1} in 0.8 *M* acetate buffer (pH 5.0), 20% saturated $(NH_4)_2SO_4$. Two, possibly three, crystal forms of salmon trypsin were obtained in the same hanging-drop experiment. One form was thin needles, too small to be characterized by conventional X-ray methods. The other two forms were both orthorhombic prisms, $P2_12_12$ (Smalås *et al.*, 1990).

Intensity data for salmon trypsin, type I, were collected at room temperature, 293 K, on an Enraf–Nonius CAD-4 diffractometer. Three crystals were used, and data were collected only for the unique part of the reciprocal lattice. The radiation source was a fine-focus copper tube ($\lambda = 1.5418$ Å), equipped with a graphite monochromator, and run at 45 kV and 26 mA. The detector, a scintillation counter, had a helium-filled tube at the front to minimize intensity loss over the crystal-to-detector distance of 368 mm. Intensities were measured using a 0.6° ω scan. A summary of the results and conditions for the diffractometer data collection is given in Table 1. It should be noted that the scan rate was lowered for crystals 2 and 3 relative to crystal 1 in order to maintain counting statistics. From five to eight control reflections, covering a wide resolution range, were measured for each of the three crystals every 6 h of exposure time to allow for a correction of radiation decay as function of exposure time and diffraction angle to be made. Maximum decay of the control reflections was taken as percentage of start intensity. An empirical absorption-correction curve was outlined by measuring the variation in intensity of the relatively strong reflection, 060, as the crystal was rotated about the diffraction vector (North, Phillips & Matthews, 1968). Maximum absorption corrections were taken as the ratio between the maximum and minimum intensity of the absorption curve. For each crystal the lattice constants were based on the θ values for 25 reflections measured in the θ range 17–25°. The values were derived by least-squares methods.

Table 1. *Summary of results and conditions for the diffractometer data collection*

	Crystal 1	Crystal 2	Crystal 3
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Lattice constants (Å)			
a	61.949 (13)	62.04 (3)	61.929 (23)
b	84.330 (22)	84.45 (1)	84.30 (4)
c	39.105 (11)	38.994 (22)	39.103 (15)
Crystal size (mm)	0.2 × 0.5 × 0.7	0.15 × 0.4 × 0.5	0.15 × 0.4 × 0.5
Total No. of measured reflections	15525	7974	4615
Scan rate (min $^{-1}$)	1.33	1.00	0.66
θ range ( )	0–23	17 24	23.5 25
Maximum decay of the control reflections (%)	37	34	14
Maximum absorption correction	1.38	1.22	1.49

The crystals were rather stable in the X-ray beam. More than 15500 reflections were collected from the first crystal, with a maximum intensity loss of 37% for the control reflections. A loss in scattering power of this magnitude is within the range found to be adequate for corrections (Hendrickson, 1976).

A summary of statistics for the merged data sets is given in Table 2. If reflections with $I > 2\sigma(I)$ are regarded as observed, 89% of the reflections from crystal 1, 83% of those from crystal 2, and 71% of those from crystal 3 pass this threshold. Furthermore, with this criterion 86% of the reflections in the merged data set may be considered as observed.

A division of the complete data set into resolution ranges with approximately equal numbers of reflections in each range is given in Table 3. It shows that the fraction of observed data decreases smoothly from 0.98 for the 2000 reflections in the lowest resolution range to 0.68 for the 1200 reflections in the resolution range 1.89–1.82 Å.

Table 2. *Summary of statistics from the merging of the three diffractometer data sets for salmon trypsin type* I

Data set	Resolution range (Å)	No. of unique reflections	No. of overlapping reflections*	% observed†	R_{merge}‡
1	∞–1.97	14941		89	
2	2.60–1.90	7779		83	
Merged 1 + 2	∞–1.90	16191	3182	88	0.035
3	1.93–1.82	4480		71	
Merged (1 + 2) + 3	∞–1.82	18298	1224	86	0.069

* The scaling of the data sets is based on reflections $>4\sigma(I)$.
† Criterion for observed reflection is $I > 2\sigma(I)$.
‡ $R_{merge} = \Sigma|F_1 - F_2|/\Sigma|F_1 + F_2|$, where F_1 and F_2 are the structure-factor amplitudes as found for the same reflection in each of the two data sets.

Table 3. *Fraction of observed reflections in different resolution ranges of the complete data set for salmon trypsin type* I *crystals*

Resolution range (Å)	46	3.93	3.11	2.72	2.47	2.29	2.16	2.05	1.96	1.89	1.82
Fraction observed*		0.98	0.97	0.94	0.91	0.89	0.86	0.81	0.77	0.74	0.68
No. of reflections		2047	1933	1920	1871	1877	1872	1885	1861	1823	1209

* Criterion for observed reflection is $I > 2\sigma(I)$.

A Wilson plot for the complete data set from 18.0–1.82 Å is shown in Fig. 1. The slope of the least-squares curve gives an overall temperature factor of 14.2 Å^2 which compares well with that found by refinement.

Structure solution by molecular replacement

Preliminary results from the molecular-replacement searches have been reported elsewhere (Smalås *et al.*, 1990). A more detailed account is given here.

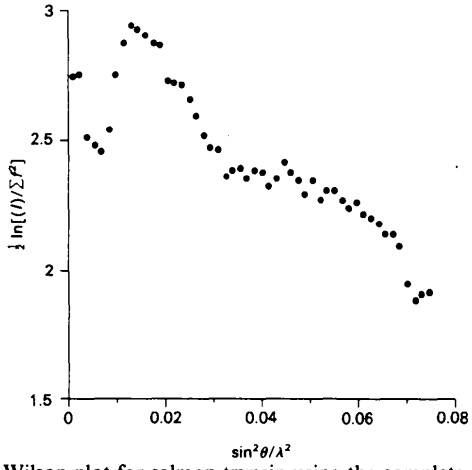

Fig. 1. Wilson plot for salmon trypsin using the complete data set from 18.0–1.82 Å resolution, f being the normalized carbon-scattering factor.

The three-dimensional structure of the North Atlantic salmon trypsin was assumed to be similar to that of bovine β-trypsin. The two trypsins have the same substrate specificity, and the estimated molecular weights are compatible. Salmon trypsin was therefore regarded as a good candidate for structure determination by the method of molecular replacement.

Molecular-replacement methods used

MERLOT (Fitzgerald, 1988), an integrated computer program package of molecular-replacement programs, was used to solve the structure of salmon trypsin.

The model of benzamidine-inhibited bovine β-trypsin (Marquart *et al.*, 1983) from the Protein Data Bank (Entry 3PTB, version July 1987) at Brookhaven National Laboratory (Bernstein *et al.*, 1977) was used as search model. This model is one of the most accurately defined protein structures of its size. It is refined to an R value of 0.157, on data to 1.5 Å resolution. The complete structure (the CS model), as well as a model with only β-carbons as side chains (the BS model), were used in the rotation searches. For both models the benzamidine molecule, the calcium ion, and all water molecules were excluded. The CS model contained 1629 atoms, while the BS model contained 1213 atoms.

Patterson functions for bovine β-trypsin, used in the rotation searches, were calculated in an artificial triclinic cell with orthogonal axes of lengths 130 Å. These Patterson functions were based on calculated structure factors for the CS and BS models, respectively, and an assumed overall temperature factor of 15 Å^2.

The Crowther rotation searches (Crowther, 1972) were carried out for several different resolution ranges, with maximum resolutions from 2.5 to 6.0 Å and a minimum resolution of 10.0 Å. In order to simplify the calculations, an intensity cut-off was used to reduce the number of reflections to about 10.000 or less. From the high intensity reflections for the model and from the observed data for salmon trypsin, harmonic coefficients were calculated, using a Patterson cut-off radius to reduce the effect of intermolecular vectors in the calculations (Smalås, 1990).

Fast rotation searches (Crowther, 1972) were carried out in the three Eulerian angles; in steps of 2.5° from 0 to 180° in α, in steps of 5° from 0 to 90° in β, and in steps of 5° from 0 to 360° in γ.

For the Lattman rotation searches (Lattman, Nickolds, Kretsinger & Love, 1971) and the Crowther & Blow (1967) translation searches, a continuous transform was formed from the calculated structure factors for the bovine β-trypsin model. The

Table 4. *Crowther (Cr) and Lattman (La) rotation angles for the false solutions for data between* 10.0 *and* 5.0 Å *resolution with all four solutions refined with searches in steps of* 1°

Model	α (°)	β (°)	γ (°)
BS (Cr)	176.9	68.7	327.5
CS (Cr)	149.5	66.1	331.0
BS (La)	160.2	65.9	328.0
CS (La)	160.0	67.0	329.0

unknown salmon structure data were expanded with reflections covering the entire hemisphere of data for each resolution range.

Crowther & Blow (1967) translation function searches were carried out for the CS as well as the BS model in fractional steps of 0.02 in all three axial directions. For each search the three independent Harker sections were plotted. Five different resolution ranges were tried for either model, namely 10.0–4.0, 10.0–4.5, 10.0–5.0, 10.0–5.5 and 10.0–6.0 Å, respectively.

A false solution

The first rotation and translation solution found later turned out to be incorrect. However, the consistency of this 'solution' was quite good, and it will therefore be described in some detail as a warning.

Rotation angles were found in the Crowther search as well as in the Lattman search restricted around the Crowther solution, namely with α 135–184, β 63–70 and γ 320–350°. This false solution showed up as the highest peak for the BS as well as the CS model in all maps based on data below 4.5 Å resolution. The rotations in β and γ were essentially the same for the two models, but there was a difference of about 26° between the respective α angles, *cf.* Table 4. However, it should be noted that the average α of 163.2° from the Crowther searches is close to that of 160.1° from the Lattman searches.

The rotation angles $\alpha = 160.1$, $\beta = 66.5$ and $\gamma = 328.5$° were applied to the CS and BS models, and Crowther and Blow translation searches were then carried out for different resolution ranges (Smalås, 1990). The same consistent set of Harker peaks could be found in four of the five Patterson maps for the CS model and in three of the five Patterson maps for the BS model, *cf.* Table 5(*a*). One notes that the best resolved peaks in the map appeared with data to 5.5 Å and with the complete structure as search model. This solution gave some intermolecular close contacts between the C-terminal helices of two neighbouring molecules. However, a minor move of the helix resulted in no close contacts at all.

The *R*-factor minimization of this rotation and translation parameters gave an *R* factor of 0.65 for data between 10.0 and 6.0 Å resolution. This high value is within the range of random positioning of atoms in the cell (Stout & Jensen, 1968). However, a slightly better *R* factor of 0.62 was obtained for data to 2.5 Å using programs in the *CCP* package (SERC Daresbury Laboratory, 1986). The *R* factor dropped to 0.32 after 13 cycles of refinement with the restrained least-squares program *PROLSQ* (Hendrickson, 1985).

After five rounds of model building and subsequent *PROLSQ* refinement, the *R* factor was brought down to 0.252, for data between 6.0 and 2.3 Å resolution. The same model was also refined using data between 6.0 and 2.0 Å resolution resulting in an *R* factor of 0.298. The resulting 2.3 Å electron-density map was reasonably good for about 60% of the structure, but the fact that the geometric parameters of the model tended to become worse after each refinement series strongly indicated that we were working on a false molecular-replacement solution. The isotropic *B* factors tended to become very low during this refinement. After each refinement round the *B* values for two to three hundred atoms were less than zero, and the overall *B* value was about 9.0 Å^2.

A new rotation solution

In order not to be affected by the 'solution' described above, the model of bovine β-trypsin was rotated 50° in all three Eulerian angles prior to the rotation search. However, the Crowther rotation searches did not give a significant solution in any of the various resolution ranges studied. The best resolved peak, about 10% higher than the next highest, appeared with data from 10.0 to 6.0 Å resolution and with CS as the model structure.

The Lattman rotation-function search, which usually determines rotation angles more accurately, was therefore carried out over the unique Eulerian rotational space, and a peak about 5% higher than any other in the map was found. This solution was similar to that found from the Crowther search, but the angles differed by 4–18° from those of the former.

Translation results

Crowther and Blow translation searches were carried out for several different rotations of the model structure. None of the Crowther rotation solutions gave a consistent set of translation vectors, but the new Lattman solution did. A single self-consistent set of translation vectors appeared in the three Harker sections regardless of resolution range. A search with the CS model on data from 10.0 to 4.0 Å resolution gave the best solution, and the corresponding Harker peaks were 64, 56 and 52% higher than the next highest peaks in the respective

Table 5. *Results from the Crowther and Blow translation search with the incorrect and the correct rotation solutions*

The translation coordinates are fractional in the axial directions. Relative heights of the solution Harker peaks compared to the highest 'false' peaks in the respective Harker sections are also given.

Model	Resolution range (Å)	Translation coordinates x	y	z	Relative heights of Harker peaks $U=0$	$V=\frac{1}{2}$	$W=\frac{1}{2}$
(*a*) Incorrect solution							
CS	10.0–4.0	0.05	0.74	0.56	0.88	0.80	—
	10.0–4.5	0.06	0.74	0.56	0.71	0.99	0.70
	10.0–5.0	0.06	0.74	0.56	1.08	1.08	0.86
	10.0–5.5	0.06	0.74	0.56	1.18	1.22	0.88
	10.0–6.0	0.05	0.74	0.56	1.13	1.23	0.88
BS	10.0–4.0	0.03	0.74	0.56	0.81	0.70	0.69
	10.0–4.5	0.03	0.74	0.56	0.83	0.89	0.69
	10.0–5.0	0.06	0.74	0.55	0.95	0.91	0.51
	10.0–5.5	0.06	0.74	0.56	0.91	0.89	—
	10.0–6.0	0.05	0.74	0.56	0.90	0.92	—
(*b*) Correct solution							
CS	10.0–4.0	0.12	0.04	0.37	1.56	1.52	1.64
	10.0–4.5	0.12	0.04	0.37	1.40	1.46	1.52
	10.0–5.0	0.12	0.04	0.37	1.37	1.29	1.45
	10.0–5.5	0.12	0.04	0.37	1.14	1.22	1.28
	10.0–6.0	0.12	0.04	0.37	1.07	1.13	1.22
BS	10.0–4.0	0.12	0.08	0.36	1.27		0.79
	10.0–4.5	0.12	0.07	0.35	1.24	—	0.83
	10.0–5.0	0.10	0.13	0.35	1.28		
	10.0–5.5	0.10	0.13	0.35	1.29	—	—
	10.0–6.0		—	—	—		—

Harker sections, *cf.* Table 5(*b*). There were no deviations between the two independent solutions for x, y and z, respectively. The same solution was also found in some of the searches with the BS model, but the consistencies between the independent solutions were much poorer, *cf.* Table 5(*b*).

In order to find how critical the orientation is for the translation solution, a test was carried out; the molecule was rotated in steps of 1° from the Lattman rotation solution in all three angles simultaneously and new translation searches carried out. The results are shown in Fig. 2. One sees that the solution is well resolved from the noise peaks with rotation of up to 3° away from the correct solution.

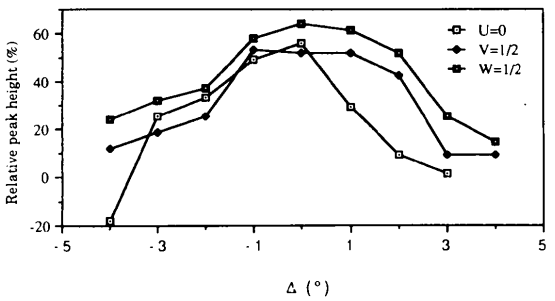

Fig. 2. Height of the peaks of the translation solution for the three Harker sections relative to the highest background peak, when the molecule is rotated $\Delta\varphi$ (°) in α, β and γ simultaneously away from the 'correct' orientation. Negative relative height means that the solution is not the highest peak in the map.

Refinement and evaluation of the molecular replacement solution

The initial R factor of the rotated and translated CS model in the cell of salmon trypsin was 0.55 for data from 10.0 to 4.0 Å resolution, decreasing to 0.525 after four cycles of R-factor minimization. The rotation parameters were $\alpha = 86.3$, $\beta = 44.75$ and $\gamma = 153.59°$ before and $\alpha = 86.12$, $\beta = 44.55$ and $\gamma = 153.39°$ after refinement, and the translation parameters changed from $x = 0.12$, $y = 0.04$ and $z = 0.37$, to $x = 0.122$, $y = 0.038$ and $z = 0.372$. This molecular-replacement solution was tested for intermolecular close contacts. Because the amino-acid sequence of the salmon trypsin was unknown, side chains were not included in the calculation of short intermolecular contacts, and hence the BS model was used in this context.

This model, oriented and positioned according to the values given above, had 14 unique contacts shorter than 3.0 Å to neighbouring molecules. Two of these short contacts were between main-chain atoms, namely between the carbonyl O atom of residue 74 and the peptide N atom of residue 174, a distance of 2.63 Å, and between the carbonyl O atom of residue 173 and the peptide N atom of residue 76, a distance of 2.82 Å. Both contacts could be almost perfect intermolecular hydrogen bonds.

The other short intermolecular contacts were mainly between atoms of the residues 123–125 of one molecule, and atoms of residues 149–152 of a neighbouring molecule. Both these regions are external loops.

Refinement and model building

Refinement procedures

Initial R-factor refinement of the CS model of bovine β-trypsin (Marquart *et al.*, 1983), oriented and positioned according to the molecular-replacement solution, gave a value of 0.45 for data from 6.0 to 2.5 Å resolution, and the corresponding value for the BS model was 0.50.

Refinement of the BS model converged with an R factor of 0.35 after eight cycles of *PROLSQ* (Hendrickson, 1985), and the CS model gave an R factor of 0.295 after three cycles. No rigid-body type refinement was carried out prior to the restraint least-squares refinements.

$2F_o - F_c$ maps with phases from the refinement of the CS model were clearer and far easier to interpret than the corresponding BS maps, and the subsequent model building was therefore based on the CS model which comprises 223 residues.

The *PROLSQ* and Fourier programs used were those of the *CCP*4 package (SERC Daresbury Laboratory, 1986) and the model building was

Table 6. *Final refinement parameters and standard deviations*

No. of protein atoms*	1582		
No. of solvent atoms	180		
No. of variable parameters	5530		
	R.m.s. deviations from ideal values		
	Number	Standard deviation	σ†
Distance restraints (Å)			
Bond distance (1–2)	1613	0.022	0.01
Angle distance (1–3)	2180	0.050	0.02
Planar distance (1–4)	597	0.088	0.03
Plane restraint (Å)	270	0.036	0.02
Chiral volume (Å^3)	242	0.304	0.12
Non-bonded contact restraints (Å)			
Single torsion contact	543	0.203	0.50
Multiple torsion contact	905	0.407	0.50
Possible hydrogen bond	167	0.227	0.50
Conformational torsion-angle restraints (°)			
Planar (ω)	210	5.5	3
Staggered	240	18.8	15
Orthonormal	23	27.5	20
Final R factor (%) in the resolution range 6.0–1.82 Å			
For 15799 reflections $>1\sigma(F)$		17.15	
For 15153 reflections $>2\sigma(F)$		16.74	
For 14443 reflections $>3\sigma(F)$		16.35	

* Including Ca and benzamidine.
† The weight on each restraint corresponds to $1/\sigma$.

carried out using the program *FRODO* (Jones, 1985) on an Evans & Sutherland PS300 colour graphical system.

The refinements were carried out with standard geometric restraints, Table 6, and from refinement round 11 the structure-factor data were weighted by the standard deviations $\sigma = 0.5\langle|F_o - F_c|\rangle$.

All the refinements from round 2 were completed with 2–3 cycles of individual B-factor refinement, during which the atomic coordinates were kept constant.

Model building

Electron-density maps for model building were calculated with $2F_o - F_c$ and $F_o - F_c$ as coefficients. The model and the electron-density maps were studied after each series of refinements, and omit maps were used for uncertain regions. They were prepared after three cycles of *PROLSQ* refinement in which the atoms of those regions were excluded.

Model adjustments were carried out where there were major differences between model and electron density and also temperature factors indicated that atoms should be moved. Since only the first 23 amino acids of the N terminal were known (Sletten, 1988), side chains were replaced according to the shape of the electron density.

Refinement after the first model building, which mainly comprised fitting of the main chain to the density and deletion of the side chains that were obviously misplaced, lowered the R factor to 0.271 for data from 6.0 to 2.0 Å, and the following two model-building and refinement rounds reduced R to 0.242, *cf.* Table 7.

Data to 1.82 Å were included from round 5, and after eight new rounds the R factor had dropped to 0.172 for the 15799 reflections with $I > 1\sigma(I)$. For the 14443 reflections with $I > 3\sigma(I)$, R dropped to 0.164. A summary of the course of refinement is given in Table 7.

Inclusion of water molecules

Water molecules were included gradually after refinemenmt round 5 when the R factor had been reduced below 0.25. The criteria used for selected water molecules were that the electron density should be well defined and that the molecules could participate in reasonably good hydrogen bonding with either protein atoms or other water molecules.

Water O atoms for which the temperature factor refined to values greater than 90 Å^2 were omitted. All solvent molecules were interpreted as water O atoms, although NH_4^+ and OH^- have approximately equivalent scattering power.

Model and electron-density maps

A complete and detailed description of the refined structure is outside the scope of this paper, but in the following we would like to discuss some features of the model and electron-density maps that are important for the evaluation of the result of the refinement.

After the last refinement round, more than 95% of the main-chain atoms were present in well defined electron density. Some residues from a possibly flexible loop, residue 146 to residue 149, and the two terminal residues of the C terminus are still not in interpretable density. The numbering system used is the one adopted from chymotrypsinogen (Hartley & Kauffman, 1966).

During the final steps of this study a gene sequence of salmon trypsin became available but it does not fully agree with the electron-density maps as described below. This indicates that the two sequences probably represent isotypes of trypsin in the North Atlantic salmon. The primary structure of the present salmon trypsin may therefore be considered as partly unknown. In the identification of residues it was of course difficult to distinguish between amidic and acidic forms of side chains. The choices made were, therefore, rather arbitrary except for residues which seemed to be conserved between the salmon and the bovine enzyme. In some cases when it was difficult to distinguish between valine and threonine side chains we allowed the environment to be decisive for the choice.

A sequence of four residues, from 146 to 149, is still not localized with interpretable density in the final map, *cf.* Fig. 3. These residues form part of an

Table 7. *Summary of the course of refinement*

Round	1	2	3	4	5	6	7	8	9	10	11	12	13
Refinement cycles	3	9	11	13	11	10	9	9	9	15	22	27	30
Resolution range (Å)	6.0–2.5	6.0–2.0	6.0–2.0	6.0–2.0	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82	6.0–1.82
Atoms refined	1638	1609	1594	1588	1594	1615	1590	1646	1636	1741	1789	1793	1762
Solvent molecules	0	0	0	0	0	5	13	49	63	125	150	160	180
Mean B factor (Å^2)	15.0	14.8	16.7	16.6	17.2	17.5	16.1	16.2	16.2	14.9	17.7	18.2	17.2
Reflections	6885	12671	12671	12671	15799	15799	15799	15799	15799	15799	15799	15799	15799
R (start of refinement)	0.452	0.401	0.388	0.381	0.394	0.357	0.330	0.318	0.320	0.325	0.285	0.271	0.280
R (end of refinement)	0.296	0.271	0.249	0.244	0.249	0.238	0.227	0.220	0.209	0.189	0.179	0.174	0.172
Standard deviation from ideal values													
Bond distances (Å) (1–2 neighbours)	0.089	0.193	0.075	0.074	0.054	0.055	0.060	0.060	0.054	0.046	0.024	0.023	0.022
Angle distances (Å) (1–3 neighbours)	0.194	0.359	0.166	0.160	0.119	0.116	0.137	0.131	0.107	0.131	0.057	0.057	0.050
Planar distances (Å) (1–4 distances)	0.202	0.333	0.204	0.189	0.143	0.155	0.161	0.153	0.126	0.138	0.093	0.090	0.088
Planar groups (Å)	0.048	0.047	0.045	0.042	0.034	0.033	0.036	0.036	0.034	0.033	0.040	0.037	0.036
Chiral volumes (Å^3)	0.423	0.410	0.446	0.459	0.343	0.337	0.405	0.418	0.397	0.345	0.345	0.336	0.304
Peptide plane (°)	8.8	8.5	9.2	6.4	10.2	9.7	9.3	9.2	6.1	4.3	6.3	6.0	5.5

external loop, and the unclear electron density may indicate disorder. As compared with bovine trypsin the two last residues are not visible in the C terminus of salmon trypsin. However, there is no clear density for a carboxyl group terminating the trypsin molecule at residue 243. This could indicate that residues 244 and 245 are present in salmon trypsin as well, although invisible in the electron-density maps due to some kind of chain disorder. Other residues for which the electron density of side chains are still not entirely satisfactory are summarized in Table 8. The somewhat unclear electron density for these side chains may indicate static disorder, but it may also in some cases indicate misinterpretation of side chains since the primary structure is still partly unknown. High-resolution data will probably clarify this ambiguity. However, all residues listed in Table 8, except Tyr 29 and Ser 192, can be classified as external and have probably greater thermal motion than internal residues.

The gene sequence which became available after the 11th refinement round (Male, 1991), had 62 of the 221 residues different from that of the present X-ray study. Most of these differences were between similar residues, and some of them were modified accordingly. In cases were residues according to the gene sequence did not fit the electron density, we chose to follow the density. The differences between the primary structure derived from the present X-ray study and the available gene sequence are shown in Fig. 4.

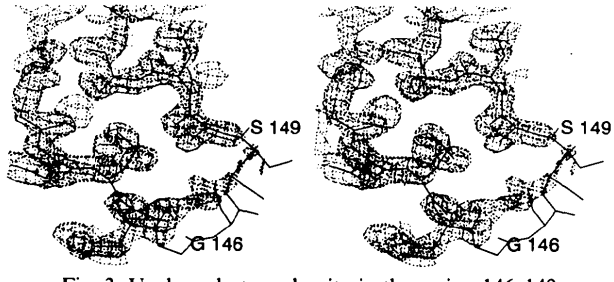

Fig. 3. Unclear electron density in the region 146–149.

Table 8. *Residues with unclear electron density in final map*

Residue	Comment
Lys 23	Poor density beyond C^γ
Asn 25	Noisy density for C^γ, $O^{\delta1}$ and $N^{\delta1}$, indicating possible static disorder about rotatable bonds
Tyr 29	Good density for the phenyl ring in omit maps, but is somewhat misshaped during refinement
Leu 33	Week density for C^α—C^β bond
Tyr 39	Noisy density indicating rotational flexibility about C^β—C^γ
His 40	Week density for C^β and C^γ
Lys 60	Good density up to N^ζ, which is not visible
Ser 61	Possible static disorder about χ^1 rotation
Val 65	Little density for C^α—C^β bond
Asn 79	Noisy density for C
Lys 87	Week and noisy density for C^ε and N^ζ
Lys 110	Week density for C^γ
Ser 125	Carbonyl group density is noisy
Thr 134	Little density for C
146–149	Noisy density, possible due to disorder
Asp 153	Little density for C^β
Phe 165	Noisy density
Lys 188B	Week density for N^ζ
Ser 192	Possible static disorder about χ^1 rotation
Ser 240	Week density for C^β

For six of the differences, namely those for positions 24, 62, 97, 186, 192 and 235, residues like glycine, alanine or serine were found in positions where the gene sequence suggests the presence of arginine, lysine or glutamine.

There are, in addition, five differences where the side chains, judging from the electron-density map, seem to be longer than those from the gene sequence. Fig. 5 shows examples of electron density for residues which are found to be alanine (27 and 31) and serine (165 and 170) in the gene sequence but could be fitted with larger side chains in the electron-density maps of the present study. This electron density was quite apparent in omit maps as well as in $2F_o - F_c$ maps. Residue 165 is interpreted as a phenylalanine, but could also be an arginine, 170 could be aspartic acid or asparagine, 27 glutamic acid or glutamine, and 31 valine or threonine. For residue 184B, the gene sequence suggests phenylalanine, but the extra electron-density bulge indicates a tyrosine instead, which is hydrogen bonded to a main-chain carbonyl O atom.

Furthermore, there are other side chains that are interpreted differently from the gene sequence, but most of these are side chains that differ in length by one or two atoms (Fig. 4).

Well defined electron density for the benzamidine molecule appeared in the difference map following refinement round 6, and the molecule was then included in the refinement. Fig. 6 shows the electron density for the benzamidine molecule and its environment in the active-site cleft at the end of refinement. Asp 189 lies in the bottom of the specificity pocket and is hydrogen bonded to benzamidine. Three hydrogen-bonded water molecules are also shown in Fig. 6.

The overall folding of the salmon trypsin molecule is shown in form of a ribbon diagram in Fig. 7. A metal binding site is found at the same sequential position as for the bovine enzyme. The metal, presumably a calcium ion as in bovine trypsin, is shown in the upper right part of the molecule, together with the six ligands. The calcium environment in salmon trypsin, shown in more detail in Fig. 8, is somewhat different from that found in bovine trypsin (Bartinuk, Summers & Bartsch, 1989). Five of the six ligands in the octahedral arrangement are the same in the two trypsin molecules. The sixth, however, which is a water molecule in bovine trypsin is a glutamic acid (Glu 77) in salmon trypsin. This is a similar arrangement to that reported for porcine pancreatic elastase (Meyer, Cole, Radhakrishnan & Epp, 1988), where the sixth ligand is Asn 77. Glu 77 of bovine trypsin also interacts with the calcium ion, but *via* a water molecule, which is shown to be the sixth ligand in the crystal structures of bovine trypsin. All the ligand-to-calcium distances in the two trypsin structures are similar.

Also the benzamidine molecule in the specificity pocket, and the catalytic triad Ser 195, His 57 and Asp 102 at the entrance of the active site cleft are indicated in Fig. 7. Six disulfide bridges are located at the same sequential positions as in the bovine enzyme.

Temperature factors

The variation of the average main-chain and side-chain atomic temperature factors along the polypep-

	I	E	E	E	E	E	I	E	E	E	E	E	E	E	I	I	I	I	E	E	E	E	E	E	I	I	I
	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	37	38	39	40	41	42	43	44
SX	Ile	Val	Gly	Gly	Tyr	Glu	Cys	Lys	Ala	Asn	Ser	Gln	Ala	Tyr	Gln	Val	Ser	Leu	Asn	Ser	Gly	Tyr	His	Phe	Cys	Gly	Gly
SG	Ile	Val	Gly	Gly	Tyr	Glu	Cys	Arg	Lys	Asn	Ser	Ala	Ser	Tyr	Gln	Ala	Ser	Leu	Gln	Ser	Gly	Tyr	His	Phe	Cys	Gly	Gly

	I	I	E	E	E	E	E	I	I	I	I	I	E	E	E	E	E	E	I	E	I	I	E	E	E
	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	69	70	71	72
SX	Ser	Leu	Val	Asn	Asn	Thr	Trp	Val	Val	Ser	Ala	Ala	His	Cys	Tyr	Lys	Ser	Gly	Ile	Gln	Val	Arg	Leu	Gly	Glu	His	Asn
SG	Ser	Leu	Ile	Ser	Ser	Thr	Trp	Val	Val	Ser	Ala	Ala	His	Cys	Tyr	Lys	Ser	Arg	Ile	Gln	Val	Arg	Leu	Gly	Glu	His	Asn

	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
SX	Ile	Ala	Val	Thr	Glu	Gly	Asn	Glu	Gln	Phe	Ile	Gly	Ser	Ser	Lys	Val	Ile	Met	His	Pro	Asn	Tyr	Ser	Gly	Gly	Asn	Leu
SG	Ile	Ala	Val	Asn	Glu	Gly	Thr	Glu	Gln	Phe	Ile	Asp	Ser	Val	Lys	Val	Ile	Met	His	Pro	Ser	Tyr	Asn	Ser	Arg	Asn	Leu

	E	E	I	I	I	I	I	E	I	E	E	E	E	E	E	E	E	E	E	E	E	I	E	E	E	E	E
	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	127
SX	Asp	Asn	Asp	Ile	Met	Leu	Ile	Lys	Leu	Ser	Lys	Pro	Ala	Thr	Leu	Asn	Ser	Tyr	Val	Gln	Pro	Val	Ala	Leu	Pro	Ser	Ser
SG	Asp	Asn	Asp	Ile	Met	Leu	Ile	Lys	Leu	Ser	Lys	Pro	Ala	Ser	Leu	Asn	Ser	Tyr	Val	Ser	Thr	Val	Ala	Leu	Pro	Ser	Ser

	E	E	E	E	E	E	E	I	I	I	I	I	I	I	E	I	E	E	E	E	E	E	E	E	E	I
	128	129	130	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155
SX	Cys	Ala	Pro	Ser	Gly	Thr	Met	Cys	Ile	Val	Ser	Gly	Trp	Gly	Asn	Leu	Ser	Gly	Ser	Ser	Ser	Ala	Asp	Ser	Asp	Thr	Leu
SG	Cys	Ala	Ser	Ser	Gly	Thr	Arg	Cys	Leu	Val	Ser	Gly	Trp	Gly	Asn	Leu	Ser	Gly	Ser	Ser	Ser	Asn	Tyr	Pro	Asp	Thr	Leu

	E	I	I	E	I	I	I	I	E	E	E	E	I	E	E	E	E	E	E	E	I	E	E	E	I	I	I
	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182
SX	Gln	Cys	Leu	Asp	Ile	Pro	Ile	Leu	Ser	Phe	Ser	Ser	Cys	Asn	Asp	Thr	Tyr	Pro	Gly	Gln	Ile	Thr	Ala	Ala	Met	Phe	Cys
SG	Arg	Cys	Leu	Asp	Leu	Pro	Ile	Leu	Ser	Ser	Ser	Ser	Cys	Asn	Ser	Ala	Tyr	Pro	Gly	Gln	Ile	Thr	Ser	Asn	Met	Phe	Cys

	I	I	E	E	E	E	E	E	I	I	E	E	E	I	E	I	I	I	I	I	I	E	E	E	I	I	I
	183	184	184 B	185	186	187	188	188 A	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	209	210	211
SX	Ala	Gly	Tyr	Met	Ala	Gly	Gly	Lys	Asp	Ser	Cys	Ser	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Cys	Asn	Gly	Gln	Leu	Gln	Gly
SG	Ala	Gly	Phe	Met	Glu	Gly	Gly	Lys	Asp	Ser	Cys	Gln	Gly	Asp	Ser	Gly	Gly	Pro	Val	Val	Cys	Asn	Gly	Gln	Leu	Gln	Gly

	I	I	I	I	I	E	E	E	E	E	E	E	E	I	I	I	I	I	E	I	E	E	E	E	I	I
	212	213	214	215	216	217	219	220	221	221 B	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238
SX	Val	Val	Ser	Trp	Gly	Tyr	Gly	Cys	Ala	Gln	Pro	Gly	Asp	Pro	Gly	Val	Tyr	Ser	Lys	Val	Cys	Ile	Phe	Ser	Gly	Trp	Ile
SG	Val	Val	Ser	Trp	Gly	Tyr	Gly	Cys	Ala	Gln	Arg	Asn	Lys	Pro	Gly	Val	Tyr	Thr	Lys	Val	Cys	Asn	Tyr	Arg	Ser	Trp	Ile

	E	E	E	E	E	E	E
	239	240	241	242	243	244	245
SX	Ser	Ser	Thr	Met	Gly		
SG	Ser	Ser	Thr	Met	Ser	Ser	Asn

Fig. 4. Primary sequence of salmon trypsin derived from the present X-ray crystallographic study (SX) and a gene sequence (SG) (Male, 1991). Differences between the two sequences are framed. Internal and external residues are indicated with I and E, respectively. The numbering system is the one adopted from chymotrypsinogen (Hartley & Kauffman, 1966).

tide chain is shown in Fig. 9. The average isotropic temperature factor is 17.2 Å^2 for all atoms, 15.2 Å^2 for protein atoms only, and 36.8 Å^2 for solvent atoms only. The average temperature factor for main-chain atoms only is 13.5 Å^2 and the corresponding value for side chains is 17.7 Å^2. The highest temperature factor for a water molecule is 84.2 Å^2, but the relatively low average temperature factor for water highlights the fact that first layer water molecules are included almost exclusively. The highest temperature factor for a protein atom, C^{ε} of Lys 87, is 138.9 Å^2. This, however, is an atom in one of the side chains that has no electron density for the outermost atoms. The average *B* factor for the benzamidine molecule is 15.3 Å^2 indicating a relatively high occupancy of the inhibitor in the active site.

The temperature factors are, in general, highest for external loop regions, reflecting a greater conformational freedom for these residues. The external loop regions in salmon trypsin comprise the residues 18–28, 33–41, 69–80, 91–101, 113–119, 122–137, 142–153, 184*B*–190, 204–207 and 221*B*–225. Residues 146 to 149, for which the temperature factors have been set to 15.0 Å^2, have not been included in the refinement. One should also note the smooth increase in average main-chain temperature factors from about residue 225 to the C terminus. Although this is a helical region, this increase could indicate some mobility of the C-terminal helix as a whole.

Reliability of the final model

The refinement statistics after each refinement round are summarized in Table 7, and the refinement

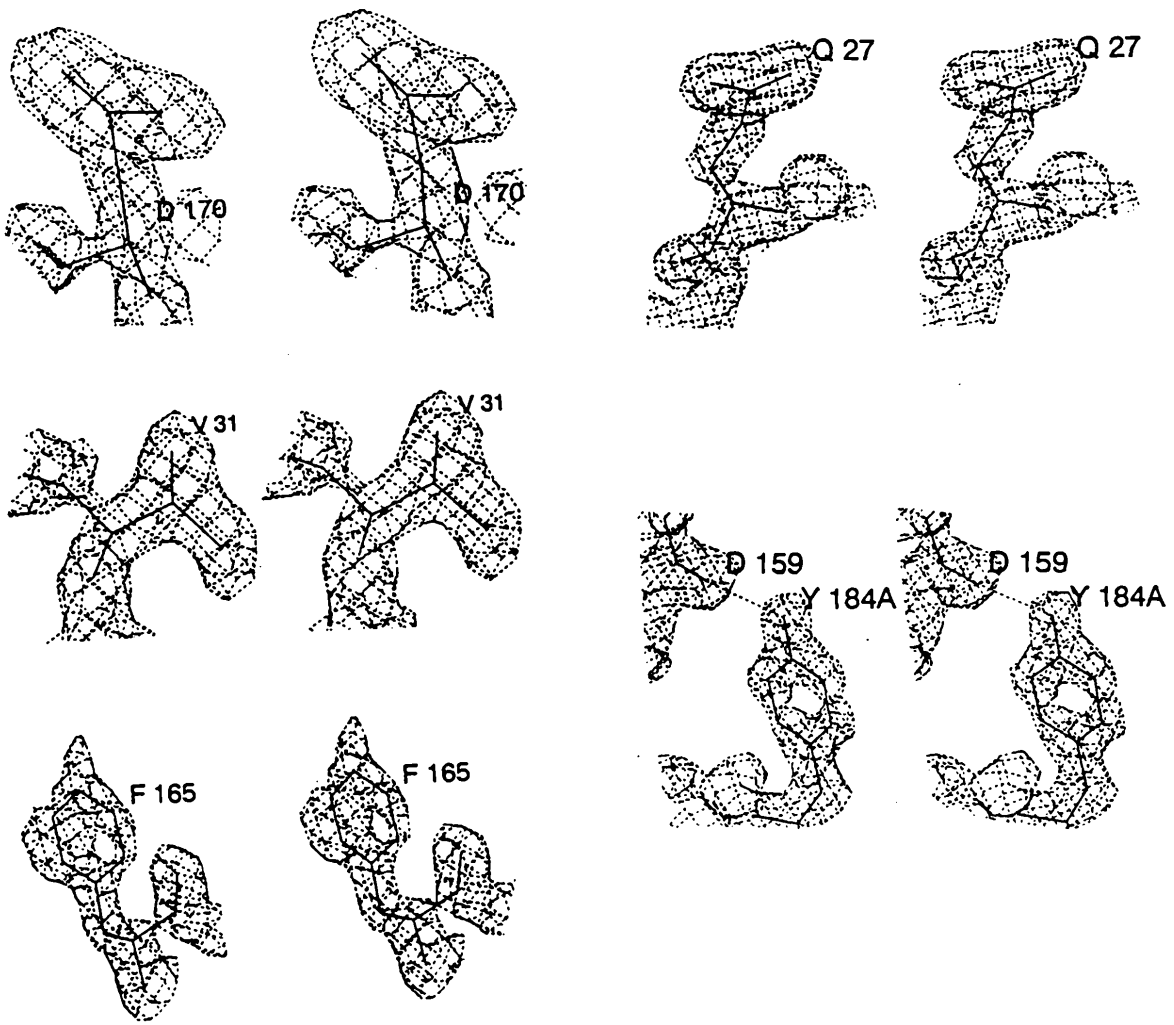

Fig. 5. Electron density for residues which are found to be alanine (27 and 31) and serine (165 and 170) from the gene sequence, but could be fitted with larger side chains. 184*B* is claimed to be phenylalanine, but is probably a tyrosine according to the electron density.

parameters and final results are given in Table 6. An important criterion for the reliability of refined protein structures is the maintenance of the correct stereochemistry after least-squares refinement. The r.m.s. deviation from ideality for the bond lengths of salmon trypsin is 0.022 Å, which is in the range found for most protein structures of related size and data resolution. Values for some other parameters (which may be taken from Table 6) are, however, not as low as found for some other comparable high-resolution structures. The fact that most of the primary structure of salmon trypsin was unknown may have effected the r.m.s. deviations from ideality to some extent because of possible misinterpretation of side chains.

Another assessment of accuracy can be made by a Luzzati plot (Luzzati, 1952), as shown in Fig. 10. The plot gives the variation in R factor as a function of $\sin\theta/\lambda$, compared to the theoretical curves for R factors *versus* $\sin\theta/\lambda$, assuming only coordinate errors to be responsible for the discrepancy between observed and calculated structure factors. This method estimates the mean positional error for atoms in salmon trypsin to be less than 0.2 Å. This value should be taken as an upper estimate of error, as the theoretical curves neglect considerations such as errors in intensity measurements and omission of most solvent molecules (Chambers & Stroud, 1979). The positional errors for atoms in well defined regions of the molecule are likely to be somewhat lower than 0.20 Å, while errors for atoms in the most unclear regions are probably larger.

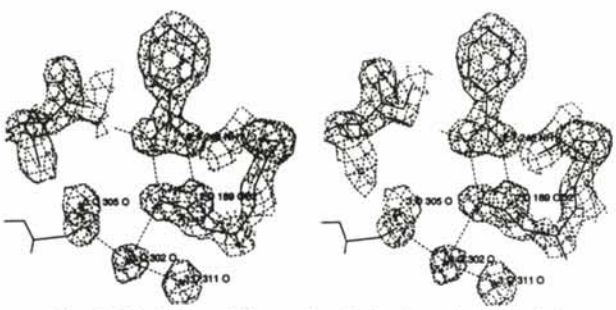

Fig. 6. The benzamidine molecule in the active-site cleft.

The distribution of the φ and ψ torsion angles, which describe the polypeptide backbone of the current salmon trypsin structure, is illustrated in the form of a Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1963) in Fig. 11. Six non-glycine residues are outside the favourable φ and ψ torsion-angle regions but are all relatively close to the boundaries of allowed regions. The six residues are Asn 25, Tyr 39, Phe 41, Asn 115, Ser 145 and Met 180. Residue 145 is in the flexible loop, which partly is invisible in electron density, Fig. 3, and may thus be erroneously positioned. Asn 25, Tyr 39, Asn 115 and Met 180 are all involved in tight turns of the polypeptide chain, which may explain why they have special conformations. However, Phe 41 (with main-chain dihedral angles of -129 and $-23°$ for φ and ψ, respectively) is the first residue of a strand possessing a somewhat irregular β-type secondary structure.

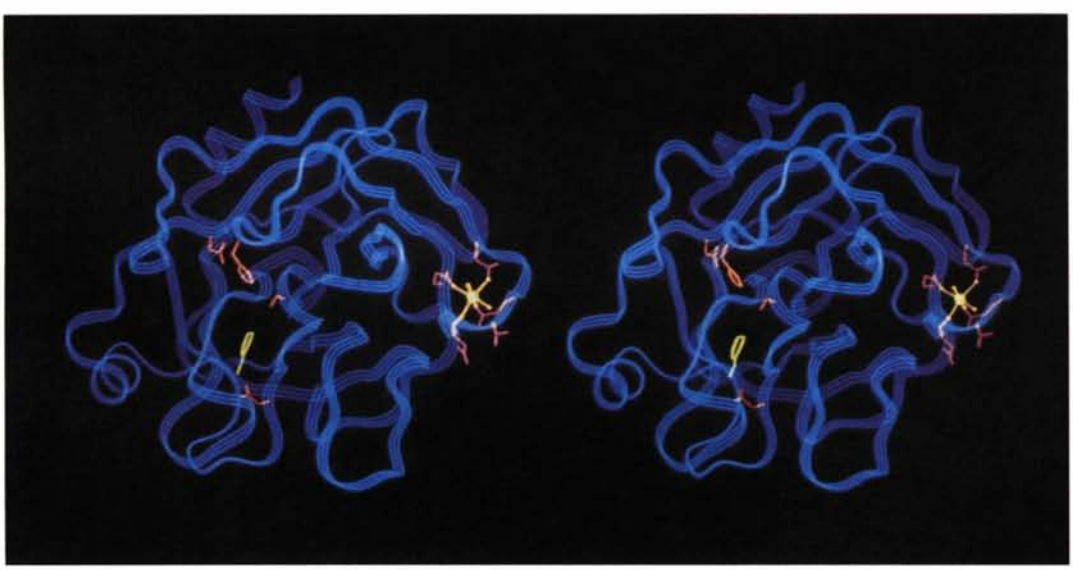

Fig. 7. The overall folding of the salmon trypsin molecule, with the C terminus at the top and the N terminus at the bottom of the picture. Benzamidine, yellow, is bonded to the aspartic acid in the active-site cleft, and the catalytic triad Ser 195, His 57 and Asp 102 sit above benzamidine at the entrance of the active-site cleft. The calcium-binding site is shown to the right. See text for further explanations.

Solvent structure

A total of 180 water molecules have so far been identified with ordered sites in the crystal structure of salmon trypsin, *cf.* Fig. 12. Of these, 21 are buried within the molecule. The distribution of ordered solvent molecules around the surface is relatively even, but with less solvent molecules identified in flexible loop regions. Such solvent molecules are probably disordered and therefore difficult to detect in the electron-density maps. Some surface areas are involved in contacts with neighbouring molecules, and thus less accessible for solvent molecules. The majority of solvent-atom peaks in the final electron-density map are well defined and spherical, but in a few cases the density is weaker and more elongated, indicating that more than one site is occupied.

The internal as well as the external water molecules tend to occur in networks or clusters. More than 50% of them appear within hydrogen-binding distances of at least one other water molecule. A predominant number of all the identified water molecules have more than one hydrogen-binding partner. Of the 21 internal water molecules seven have four hydrogen-binding partners, 11 have three, and three have two such partners.

Crystal packing and intermolecular contacts

Molecules of salmon trypsin pack tightly into the $P2_12_12$ unit cell of the type I crystals. The molecules lie in planes parallel to the crystallographic *ab* plane, Fig. 13, and consequently each molecule has 20 neighbours in the cell, six in the same plane and seven in planes above and below. It forms contacts of <4.0 Å with only eight of these, namely the six surrounding neighbours in the *ab* plane, the one above and the one below.

There are 26 contact areas (Table 9) between the reference molecule and its eight closest neighbour molecules. 204 contacts are shorter than 4.0 Å and only 30 (15 unique) of these are in the range 2.8–3.4 Å.

Sixteen of the 204 close contacts are main-chain–main-chain contacts, 90 are main-chain–side-chain contacts, and 98 are side-chain–side-chain contacts. Eleven of the fifteen unique intermolecular close contacts between 2.8 and 3.4 Å can be classified as possible hydrogen bonds.

The exposed calcium-binding loop (residues 70–80) stands for a considerable part of the intermolecular close contacts listed in Table 9. Atoms from this loop participate in the only two intermolecular main-chain–main-chain hydrogen bonds found in this structure.

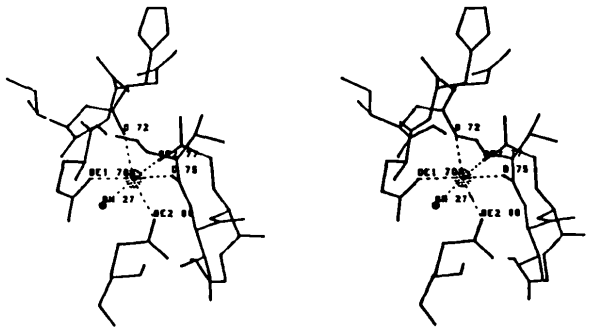

Fig. 8. The calcium-binding loop of salmon trypsin, comprising the residues 70 to 80. The six ligands are labelled and the Ca^{2+} ion is drawn as a sphere of 30% van der Waals radius.

Concluding remarks

The method of molecular replacement is widely used for solving protein structures from homologous series of molecules. As more and more 'automatic' refinement procedures have become available, such structures may now be solved and refined in a very straightforward way with a minimum of 'by-hand modelling'. This study shows, however, that great care should be taken using the method of molecular replacement. Even for structures that are expected to be very similar, erroneous rotation and translation parameters can be obtained with highly convincing resolution from the background. In this case the rotation peak of the false solution was far more resolved and clear than the one that was later found to be correct. Refinement of a totally incorrect structure can give good *R* factors and reasonable electron-density maps. As protein crystallographers have been aware for a long time, it is as important to monitor and pay attention to other parameters of the

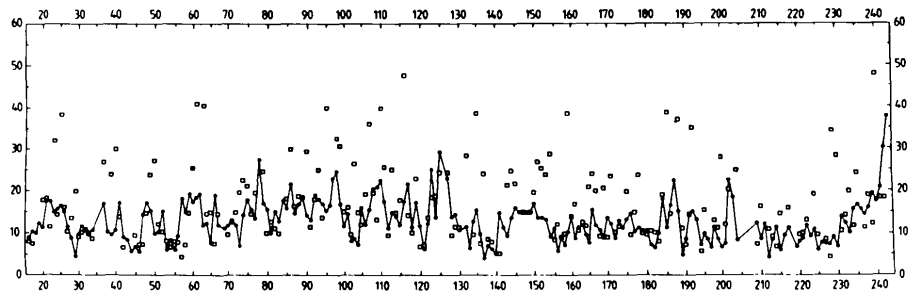

Fig. 9. Variation in isotropically refined temperature factors [B (Å^2)] along the polypeptide chain after the last refinement round, averaged over main-chain atoms and side-chain atoms. Lines are drawn between the averaged value for main-chain atoms, while the corresponding values for side-chain atoms are indicated with squares.

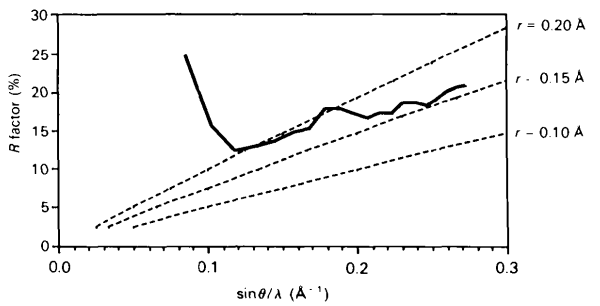

Fig. 10. Plot of R factor *versus* resolution after Luzzati (1952). All reflections in the resolution range 6.0–1.82 Å with $F_o > 3\sigma_F$ were used.

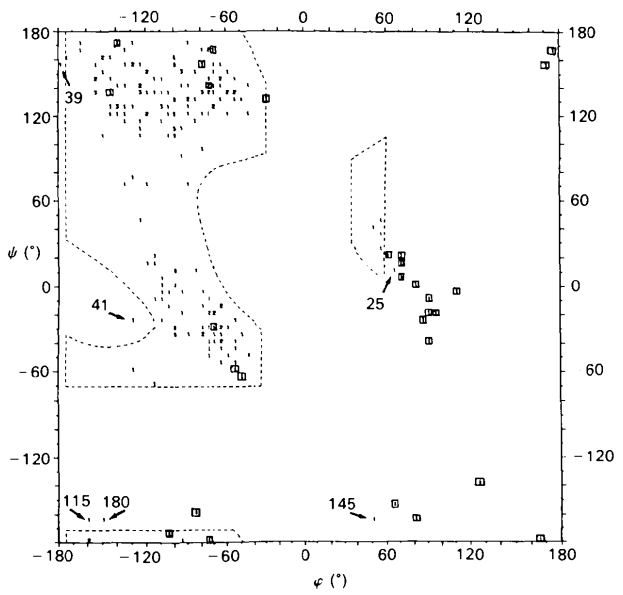

Fig. 11. Plot of φ, ψ values for salmon trypsin. 'Allowed' regions are according to Ramachandran, Ramakrishnan & Sasisekharan (1963). Glycines are framed. Non-glycine residues outside the 'allowed' regions are indicated with an arrow.

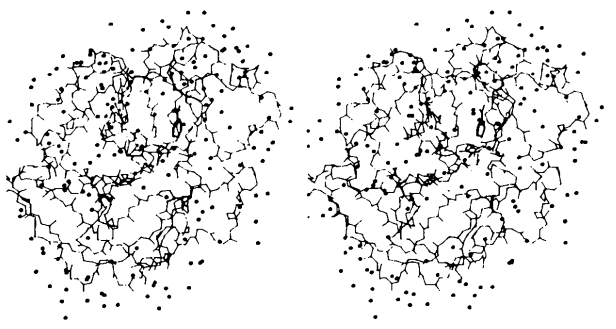

Fig. 12. Water structure found in salmon trypsin. The 180 water molecules are indicated with small filled circles. Only the main-chain atoms of trypsin are shown, together with the benzamidine molecule.

Table 9. *Unique contacts less than* 4 Å, *and in the range* 2.8–3.4 Å, *between salmon trypsin molecules in the unit cell of type* I *crystals*

The contacts between the reference molecule and molecules 1, 2 and 3 are equivalent to the contacts between the reference molecule and the molecules 4, 5 and 6. The numbering of molecules is defined in the figure legend of Fig. 12.

Residue(s) in reference molecule		Residue(s) in contact molecule	Symmetry operation	No. of contacts < 4 Å	2.8–3.4 Å
170 175	To molecule 1	66 76	x screw axis	28	7
170		34–39		19	4
78	To molecule 2	117		1	
110–113		110–113	z twofold axis	8	1
132		132		1	
130		166		4	
239	To molecule 3	145	y screw axis	1	
114–120		217–229		14	
123		143		4	1
202		39		5	
123		192		1	
48–49		221 225		15	1
93	To molecule above	188	Translation c	1	1
Total No. of unique contacts				102	15

refinement, *i.e.* geometrical parameters such as maintenance of the correct stereochemistry and folding angles (φ, ψ) of the backbone.

The refined structure of salmon trypsin* will allow a detailed comparison with the structures of other trypsins, and may thus contribute to a better under-

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TBS, R1TBSSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37074 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 1 June 1994.

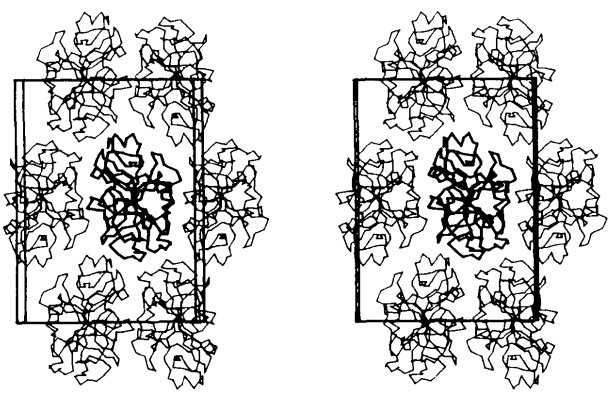

Fig. 13. Packing of salmon trypsin molecules in the unit cell of type I crystals, as seen along the c axis. The reference molecule is in the centre, molecule 1 to the right, and the others numbered clockwise relative to this, 2, 3, 4, 5, 6. Combine this with the information in Table 10.

standing of the observed differences in enzyme kinetics between trypsin molecules from some marine species and mammalian species as described in the *Introduction*. Such a comparison is now underway and will be presented in an accompanying paper (Smalås *et al.*, manuscript in preparation).

This work has been supported by the Norwegian Research Council for Science and the Humanities (NAVF), and we thank Dr Rune Male, Center for Biotechnology, University of Bergen, for giving us access to the gene sequence prior to its publication.

Note added in proof: The gene sequence of salmon trypsin (Fig. 4) has been deposited with the data library at the European Molecular Biology Laboratory (Male, 1993).

References

ARNESEN, J. A. (1989). Master degree thesis, Univ. of Tromsø, Norway.

ASGEIRSSON, B., FOX, J. W. & BJARNASON, J. B. (1989). *Eur. J. Biochem.* **180**, 85–94.

BARTINUK, H. D., SUMMERS, L. J. & BARTSCH, H. H. (1989). *J. Mol. Biol.* **210**, 813–828.

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

BIER, M. & NORD, F. F. (1951). *Arch. Biochem. Biophys.* **33**, 320–332.

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

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon & Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

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

HENDRICKSON, W. A. (1976). *J. Mol. Biol.* **106**, 889–893.

HENDRICKSON, W. (1985). *Methods Enzymol.* **115B**, 252–270.

HJELMELAND, K. & RAA, J. (1982). *Comp. Biochem. Physiol.* **71B**, 557–562.

JONES, A. W. (1985). *Methods Enzymol.* **115B**, 157–170.

LATTMAN, E. E., NICKOLDS, C. E., KRETSINGER, R. H. & LOVE, W. E. (1971). *J. Mol. Biol.* **60**, 271–277.

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

MALE, R. (1991). Personal communication.

MALE, R. (1993). EMBL Data Library, Accession No. X70074.

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

MEYER, E., COLE, G., RADHAKRISHNAN, R. & EPP, O. (1988). *Acta Cryst.* B**44**, 26–38.

MURAKAMI, K. & NODA, M. (1981). *Biochim. Biophys. Acta*, **658**, 17–26.

NORTH, A. C. T., PHILLIPS, D. C. & MATTHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.

OSNES, K. K. & MOHR, V. (1985). *Comp. Biochem. Physiol.* **82B**, 607–619.

RAMACHANDRAN, G. N., RAMAKRISHNAN, G. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.

READ, R. J. & JAMES, M. N. G. (1988). *J. Mol. Biol.* **200**, 523–551.

SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, England.

SIMPSON, B. K. & HAARD, N. F. (1984). *Can. J. Biochem. Cell Biol.* **62**, 894–900.

SLETTEN, K. (1988). Personal communication.

SMALÅS, A. O. (1990). Doctoral thesis, Univ. of Tromsø, Norway.

SMALÅS, A. O., HORDVIK, A., HANSEN, L. K., HOUGH, E. & JYNGE, K. (1990). *J. Mol. Biol.* **214**, 355–358.

SPRANG, S., STANDING, T., FLETTERICK, R. J. & STOUD, R. M. (1987). *Science*, **237**, 905–908.

STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination, A Practical Guide*, p. 246. New York: Macmillian.

TITANI, K., ERICSSON, L. H., NEURATH, H. & WALSH, K. A. (1975). *Biochemistry*, **14**, 1358–1366.

TORRISSEN, K. R. (1984). *Comp. Biochem. Physiol.* **77B**, 669–674.

UCHIDA, N., ANZAI, H. & NISHIDE, E. (1986). *Bull. Jpn. Soc. Sci. Fish.* **52**, 731–735.

UCHIDA, N., TSUKAYAMA, K. & NISHIDE, E. (1984). *Bull. Jpn. Soc. Sci. Fish.* **50**, 313–321.

WALSH, K. A. (1970). *Methods Enzymol.* **19**, 41–63.

WINTER, W. P. & NEURATH, H. (1970). *Biochemistry*, **9**, 4673–4679.

ZWILLING, R., PFLEDERER, G., SONNENBORN, H.-H., KRAFT, V. & STUCKY, I. (1969). *Comp. Biochem. Physiol.* **28**, 1275–1287.