Acta Cryst. (1995). D**51**, 725–730

Structure of Anionic Salmon Trypsin in a Second Crystal Form

BY GUNNAR I. BERGLUND, ARNE O. SMALÅS* AND ASBJÖRN HORDVIK

Protein Crystallography Group, Department of Chemistry, Institute of Mathematical and Physical Sciences, University of Tromsø, N-9037 Tromsø, Norway

AND NILS P. WILLASSEN

Department of Biotechnology, Institute of Medical Biology, University of Tromsø, N-9037 Tromsø, Norway

(*Received* 9 *September* 1994; *accepted* 5 *January* 1995)

Abstract

Anionic salmon trypsin in a second crystal form (ST-IIB) has been refined at 1.83 Å resolution. The crystals are orthorhombic and belong to space group $P2_12_12$ with lattice parameters a = 77.09, b = 82.33 and c = 31.16 Å. The present structure has been compared to salmon trypsin as it appears in a previously reported crystal form (ST-IIA) with cell dimensions a = 61.95, b = 84.33 and c = 39.11 Å [Smalås & Hordvik (1993). *Acta Cryst.* D**49**, 318–330]. The presence of a sulfate group involved in several hydrogen bonds to active-site residues, and the location of an additional benzamidine site in the crystal lattice, are the most striking differences between the present and the previous structure. Superposition of main-chain atoms in the two structures give an overall r.m.s. difference of 0.26 Å, with the main differences located to areas with different molecular packing. The overall coordinate error is estimated to be between 0.20 and 0.25 Å by the method of Luzzati.

1. Introduction

The crystal structure of an anionic form of salmon trypsin was recently determined at 1.82 Å resolution (Smalås & Hordvik, 1993). Additional refinement has later been carried out after sequence data became available (Male, Lorents, Smalås & Torrissen, 1995). A comprehensive structural comparison between salmon and bovine trypsins has been reported (Smalås, Heimstad, Hordvik, Willassen & Male, 1994).

The overall tertiary structure of salmon trypsin is similar to that of other trypsins of known structure, and in particular to bovine trypsin (Bode & Schwager, 1975).

The primary structure of ST-II† shows about 65% similarity to that of bovine trypsin. It comprises 222 residues, one less than bovine trypsin. The deletion (residue 146) occurs in the region described as the autolysis loop in bovine trypsin (Marquart, Walter, Deisenhofer, Bode & Huber, 1983). Features such as disulfide bridges, calcium-binding environments and active-site regions are similar in salmon and bovine trypsins. Despite high overall agreement between the two structures, interesting differences occur (Smalås *et al.*, 1994).

The purpose of the present project is as part of a larger project to look for structure–function relationships for the cold-adaptation features of psychrophilic enzymes. The fish enzymes seem to have adopted to cold habitat by having developed a more flexible molecular structure. Molecular flexibility can hardly be estimated directly from crystal structures due to difficulties in predicting effects from intermolecular interactions. Analysis and comparison of crystal structures of the same molecule in different crystal environments may, however, give information about structural features which are caused by crystal packing effects. We report here a structure study of anionic salmon trypsin in a second crystal form.

2. Experimental

Anionic trypsin was isolated and purified from pancreatic tissue of Atlantic salmon. The purification method, which is more extensive than that used by Smalås, Hordvik, Hansen, Hough & Jynge (1990) includes affinity, hydrophobic interaction, ion-exchange and gel-filtration chromatography. Single orthorhombic crystals, slightly different from a form observed by Smalås *et al.* (1990), suitable for high-resolution X-ray analysis grew from droplets which initially contained 5 mg ml^{-1} protein equilibrated over wells containing stock solutions of 0.7 *M* ammonium sulfate, 10 m*M* $CaCl_2$ and 60 m*M* benzamidine buffered at pH 6.0 with 50 m*M* sodium citrate-phosphate.

Data were collected on an Enraf–Nonius FAST area detector using an Enraf–Nonius FR-571 rotating Cu anode generator equipped with a graphite monochromator, in 0.1° rotation frames recorded for 45 s each. The stored images were processed using the FAST version of the *MADNES* program package (Messerschmidt &

* To whom correspondence should be addressed.

† Five different gene loci coding for four different trypsins (two gene loci give rise to the same mature trypsin molecule) have been identified on the genes from Atlantic salmon (Male *et al.*, 1995). One of these, ST-II (Smalås *et al.*, 1994) has been shown to match the electron density of the first crystal form as well as the present one.

Table 1. *Summary of crystallographic data collection and refinement*

Crystal dimensions (mm)	0.9 × 0.2 × 0.1
Space group	$P2_12_12$
Cell edges (Å)	$a = 77.09$ (5) $b = 82.33$ (4) $c = 31.16$ (2)
V_m (Å^3 Da^{-1})	2.08
Resolution (Å)	1.83
No. of observations	25917
No. of unique reflections	15984
% of reflections with $I > 3\sigma_I$	82.6
Overall merging R* (%)	6.0
Overall B factor† (Å^2)	14.2
Completeness (%)	88.2
Programs used for refinement	*X-PLOR/PROLSQ*
Resolution range (Å)	8.0–1.83
Final R value	0.199
R_{free} ‡	0.272
No. of atoms	1779
No. of reflections ($I > 3\sigma_I$)	14271
Mean B values (Å^2)	
All atoms	18.6
Protein atoms	17.6
Water molecules	30.4
Protein residues	222
Ca^{2+} ion	1
SO_4^{2-} ion	1
Benzamidine molecules	2
Water molecules	125
Ramachandran plot values§ (%)	
Most favoured regions	86.6
Additional allowed regions	13.4
Mean positional error (Å) from a Luzzati plot ¶	0.20–0.25
R.m.s. bond distances (1–2)**	0.021
R.m.s. angle distances (1–3)**	0.035
R.m.s. planar distances (1–4)**	0.049
R.m.s. plane groups (Å)**	0.015

* $R_{merge} = (\sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |\langle I_h \rangle|) \times 100\%$.
† Wilson (1949).
‡ (Brünger, 1992*b*).
§ Produced with *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993).
¶ Luzzati (1952).
** R.m.s. deviations from ideal values is relative to the PROTIN.-DAT dictionary supported by the *CCP*4 package.

Pflugrath, 1987), followed by profile fitting using *PROCOR* (Kabsch, 1988). Subsequent scaling and merging were performed using the *CCP*4 (Collaborative Computational Project, Number 4, 1994) program package. Selected statistics for assessment of data quality are given in Table 1.

Given the refined coordinates of ST-IIA,* the trypsin molecules were located within the present crystals by molecular replacement, using *MERLOT* software (Fitzgerald, 1988). The calcium ion, the benzamidine inhibitor, and the water molecules were not included in the initial model.

The model was fitted to the observed data by iterative cycles of standard crystallographic refinement procedures, using *X-PLOR* (Brünger, 1992*a*) and *PROLSQ* (Hendrickson, 1985), and manual rebuilding by means of *CHAIN* (Sack, 1988) on an Indigo2 graphics terminal. Adjustments to the model comprised mainly refitting of side chains and localization of solvent molecules corresponding to positive electron-density peaks in difference Fourier maps.

Omit densities for a calcium ion and the inhibiting benzamidine molecule were localized and the atoms were built into the model (Bode & Schwager, 1975; Smalås & Hordvik, 1993). A sulfate group, initially interpreted as a water molecule, was found in the active-site region. In addition, a second benzamidine site was also localized and built into the model. Table 1 summarizes the final model with refinement statistics.*

3. Results and discussion

3.1. *The structure of ST-IIB in comparison with ST-IIA*

The overall structure of ST-IIB is similar to that of ST-IIA. The positions of the main-chain atoms of the two structures can be superimposed with an overall r.m.s. deviation of 0.26 Å. Main-chain atoms for residues building up the central β-barrels (Smalås *et al.*, 1994) of the two salmon trypsin structures can be superimposed with an r.m.s. deviation of 0.13 Å. The corresponding values with side-chain atoms included is 0.47 and 0.33 Å, respectively.

Despite good overall r.m.s. agreement, a more detailed comparison emphasizes variations along the polypeptide chains. In Fig. 1(*a*), the mean r.m.s. difference between ST-IIA and ST-IIB is shown for each residue along the polypeptide chain. For three segments (96–98, 124–127 and 148–150) located outside the central β-barrels the r.m.s. deviations exceed 0.5 Å. The highest peak in the plot (1.26 r.m.s.) occurs at residue 125, a part of an interdomain peptide stretch running from residue 109 to 133, but this residue is rather disordered in both structures and thus may not represent real differences. The regions showing the largest structural differences between ST-IIB and ST-IIA coincides well with residues having high temperature factors, *cf.* Fig. 1(*b*). Residues located mainly on the surface of the protein show a higher degree of static or dynamic disorder, *cf.* Fig. 1(*b*). In particular the residue ranges 124–127 and 145–150 are rather flexible according to their temperature factors and, therefore, able to adopt different conformations. The residues 145–150 are part of the region described as the autolysis loop (residues 142–153) in bovine trypsin (Marquart *et al.*, 1983). Comparative molecular dynamics simulations on bovine and salmon trypsins in aqueous solution show this external loop to be the most flexible part in either molecule (Heimstad, Hansen & Smalås, 1995).

* The atomic coordinates of ST-IIA are available from the Protein Data Bank (Bernstein *et al.*, 1977), reference 2TBS.

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1BIT). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: SE0164).

The r.m.s. displacement in main-chain positions for residues comprising the active-site region (His57, Asp102, Gly193, Ser195) and the specificity pocket (189–192, 214–220) is 0.13 Å after least-squares superposition. This is well below the overall r.m.s. value of 0.26 Å obtained for the entire molecule. The maximum displacement occurs at residue Gln192 where the main-chain position differs by about 0.3 Å. Possible hydrogen-bonding distances in the active-site region show close agreement within the two structures.

A possible sulfate ion was located and built into the map of ST-IIB at the active site (see Fig. 2) in a similar position as reported for bovine trypsin (Bode & Schwager, 1975; Bartunik, Summers & Bartsch, 1989). The O atoms are not fully resolved, but the central peak indicates more density than would be expected for a water molecule. Another plausible ion is phosphate, a component of the crystallization buffer. However, it is difficult to discriminate between those ions at the present resolution. All four O atoms of the possible sulfate ion

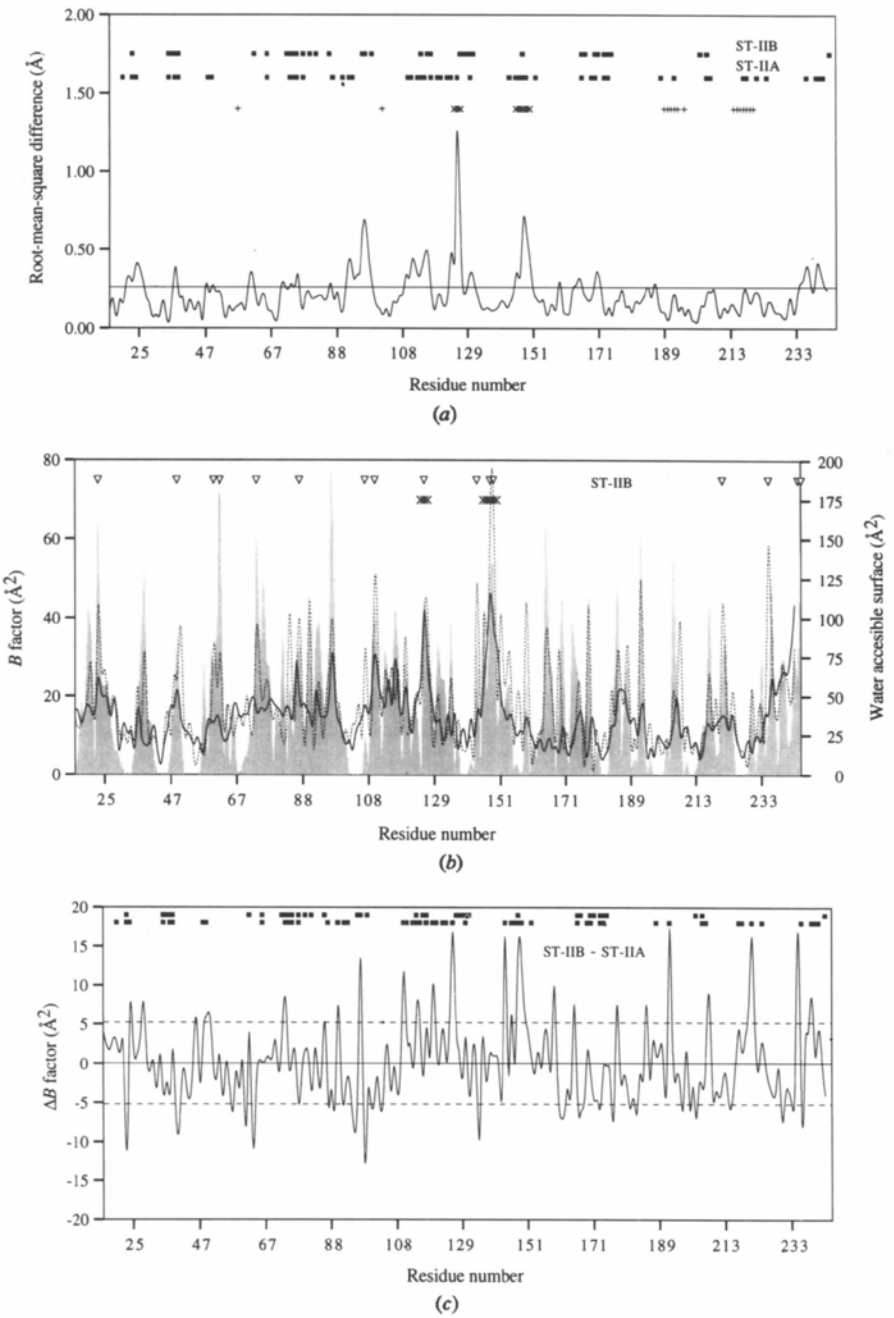


Fig. 1. Variations along the polypeptide chain. The numbering system is that adopted from chymotrypsinogen A (Hartley & Kauffman, 1966). (*a*) R.m.s. differences (Å) in main-chain atomic positions between the ST-IIB and ST-IIA structures. Average deviation of 0.26 Å is shown by a horizontal line. Closed bars denote residues involved in intermolecular contacts in either structures, respectively. The catalytic residues and residues forming the specificity pocket (in the following order: His57, Asp102, 189–192, Ser195 and 214–220) are indicated by + signs. Residues 124–127 and 145–150 (see §3.1) are marked by × signs. (*b*) Mean crystallographic temperature factors B, averaged over main-chain and side-chain atoms, respectively, are shown along with water-accessible surface area (shadowed area) averaged over all atoms per residue. Triangles indicate residues with incomplete electron density in the final $2F_o - F_c$ map contoured at 0.8σ. (*c*) Difference between mean B values (ΔB) averaged for all atoms per residue in ST-IIB and ST-IIA. Broken lines denote the r.m.s deviation of 5.2 Å^2 from the mean value of 0.03 Å^2.

seem to be involved in hydrogen bonds, *cf.* Fig. 2. The sulfate ion replaces a water molecule (Wat324) in the ST-IIA structure.

Two ordered benzamidine sites about 13 Å apart have been located in the ST-IIB structure, *cf.* Fig. 3. One benzamidine molecule (BZA246) is associated with the specificity pocket where it is hydrogen bonded to Asp189. The other (BZA247), actually a crystal artifact, is found at the surface of the protein making contacts with a neighbouring molecule. The binding of the active-site benzamidine, *cf.* Fig. 4(*a*), is almost identical to that seen in ST-IIA (Smalås *et al.*, 1994) and bovine trypsin (Bode & Schwager, 1975). BZA247 was found at the extension of the primary binding site, *cf.* Fig. 4(*b*). Multiple binding of benzamidine has been described in recombinant structures of rat trypsin (Earnest, Fauman, Craik & Stroud, 1991; Wilke, Higaki, Craik & Fletterick, 1991). These structures have in common the fact that they have all been crystallized in relatively high concentrations of benzamidine (60–100 m*M*). In contrast, the ST-IIA structure was crystallized in 1 m*M* benzamidine.

A metal-binding site, presumably occupied by a calcium ion, has been refined to agreement with the ST-IIA structure (Smalås *et al.*, 1994). The calcium-binding site, extensively described for bovine trypsin by Bode & Schwager (1975), is part of the loop running from residues 69 to 80 connecting two antiparallel β-strands in the N-terminal domain of the molecule, *cf.* Fig. 3.

3.2. *Crystal packing and intermolecular contacts*

The most pronounced deviations in atomic positions between the ST-IIB and ST-IIA structures are probably explained by different molecular packing. Fig. 1(*a*)

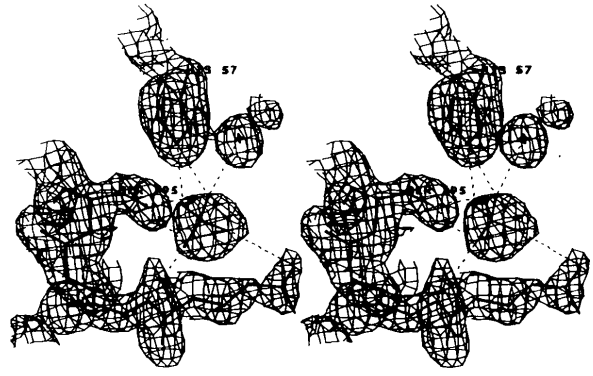


Fig. 2. Stereoview showing sulfate binding at the active site of ST-IIB. Possible hydrogen bonds to His57 $N_{\varepsilon}2$, Ser195 O_{γ}, Gly193 N (not labelled) and Gln192 $N_{\varepsilon}2$ (not labelled) are indicated with broken lines. The electron density corresponds to $2F_o - F_c$ with the sulfate ion omitted from the calculations. The map is contoured at 1σ level.

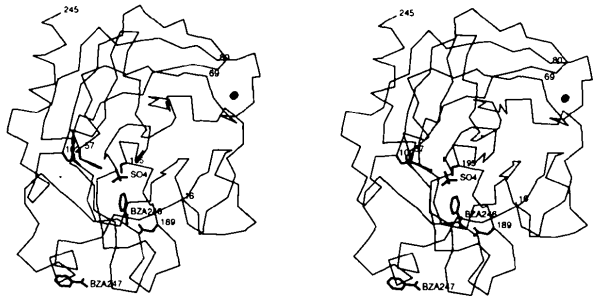
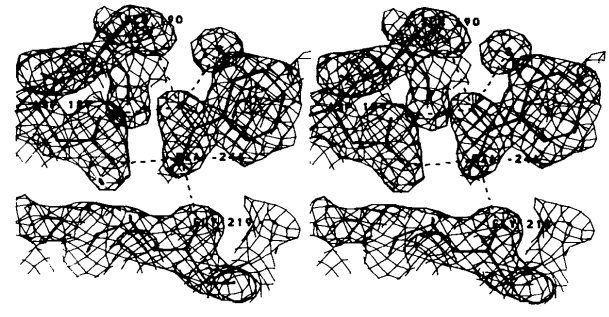


Fig. 3. Stereo drawing showing location of the sulfate ion, the two benzamidine molecules and the calcium ion in the three-dimensional structure of ST-IIB. Cα backbone is drawn with thin lines while side-chain atoms of the catalytic triad residues (His57, Asp102, Ser195) and Asp189 at the base of the specificity pocket, the sulfate ion and benzamidine molecules are drawn in thick lines. The calcium ion is indicated by a sphere. BZA246 is in the specificity pocket while BZA247 is located at the surface of the protein mediating contacts to a neighbouring molecule.



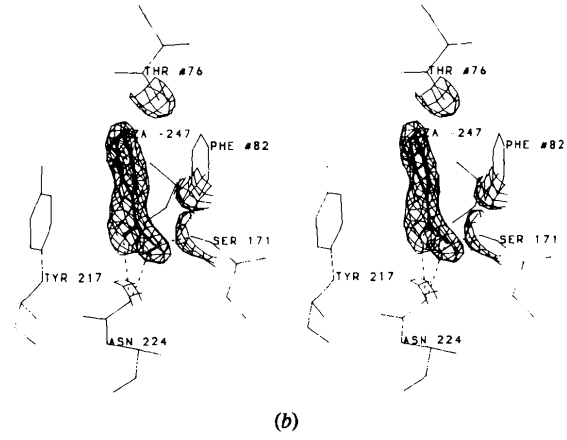


Fig. 4. Stereoview showing density for a benzamidine molecule at two sites. The electron density corresponds to $2F_o - F_c$ with the benzamidine molecule omitted from the calculations. The maps are contoured at 1.5σ level. (*a*) Benzamidine bound in the specificity pocket of ST-IIB with corresponding hydrogen-bonding pattern. (*b*) Benzamidine molecule mediating intermolecular contacts at the extension of the primary binding site. Possible hydrogen bonds to Ser171 O and Asn224 $N_{\delta}2$ are indicated with broken lines. Residues Thr#76 and Phe#82 are from a neighbouring molecule.

emphasizes that these discrepancies mainly occur at positions located close to intermolecular contact areas in one structure, but are more remote from a contact area in the other or that the same contact area is reduced. The effect of different molecular packing may also be emphasized by comparing the temperature-factor distributions for each residue along the polypeptide chains, *cf.* Fig. 1(*c*). The larger differences in temperature factors between the two structures are seen for loop regions involved in close intermolecular contacts in only one of the crystal forms.

The differences in crystal packing between ST-IIA and ST-IIB can be illustrated by focusing on the active site in the two crystal structures, *cf.* Fig. 5. In ST-IIB, the active-site region for the reference molecule pack towards the closest neighbouring molecule one cell length away in the *c* direction, *cf.* Fig.5(*a*). In ST-IIA crystals the active site in the reference molecule is closest to a symmetry-related molecule half a cell-length in the *a* direction, *cf.* Fig.5(*b*). There are no close contacts (≤ 4 Å) in the active-site environments of ST-IIB, while three residues (192, 217 and 219) of the primary binding site of ST-IIA have close contacts to a neighbouring molecule (see Fig. 1*a*). The different crystal packing in ST-IIA and ST-IIB has not affected significantly the active-site geometry.

Contact areas (side chains included) in the ST-IIA structure are more extensive than in the present structure, *cf.* Fig. 1(*a*). The total number of intermolecular contacts (≤ 4 Å) are 334 (167 unique) in ST-IIA compared to 218 (109 unique) in ST-IIB, involving 53 and 38 residues, respectively, *cf.* Fig. 1(*a*). The slightly looser packing of the ST-IIB crystals compared to the ST-IIA crystals are in accordance with their total water-accessible surface area and the *B* values for protein atoms. The water-accessible surface area of salmon trypsin in ST-IIB and ST-IIA crystals, as estimated by programs in the *CCP*4 program package, is 9335 and 9135 Å^2, respectively, and the average *B* values for protein atoms in the two structures are 17.6 and 16.9 Å^2, respectively.

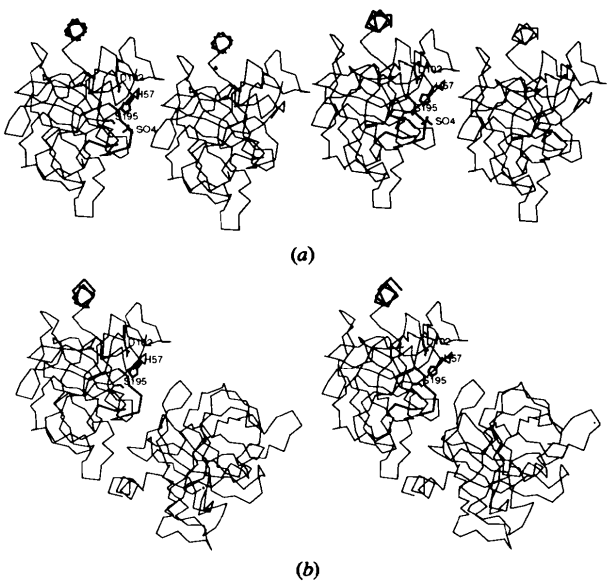


Fig. 5. Location of the active site in the crystal lattice of ST-IIB and ST-IIA. The reference molecules (left) are given the same orientation in both crystal structues to visualize the different packigns. The Cα trace of the residues forming the primary binding site are marked by heavy lines. (*a*) The closest neighbour to the active-site region of ST-IIB is actually the same molecule translated one cell length in the *c* direction. The catalytic residues and the sulfate group are shown by thick lines. (*b*) The closest neighbour to the active-site region of ST-IIA is a symmetry-related molecule translated half a cell length in *a* direction. The catalytic residues are shown by thick lines.

4. Concluding remarks

This work shows that the present anionic salmon trypsin structure corresponds to the ST-II primary sequence and, therefore, represents the previously refined structure by Smalås *et al.* (1994) in a different crystal environment. A comparison of the two salmon trypsin structures shows that the overall fold of the trypsin molecule is, to a high extent, independent of differences in crystal packing. In particular this is so for the central β-barrels building up the core of the protein, the active-site region, and the specificity pocket. The largest differences between ST-IIB and ST-IIA are found in regions which are able to adopt different conformations, and these discrepancies may probably be explained by different crystal packing. Even minor regions of high flexibility will change during the process of crystallization, and the crystal lattice accordingly.

A sulfate ion has been located at the active site in a similar position as reported for bovine trypsin (Bode & Schwager, 1975; Bartunik *et al.*, 1989). The presence of a sulfate ion in ST-IIB may indicate that the His57 $N_\varepsilon 2$ is protonated. The distance between His57 $N_\varepsilon 2$ and Ser195 O_γ is 3.3 Å, suggesting a weak hydrogen bond. The bound sulfate ion in ST-IIB replaces a water molecule in ST-IIA at an equivalent position, but does not seem to have affected the relative orientations of the catalytic residues His57 and Ser195.

This work has been supported by grants from the Norwegian Research Council.

References

BARTUNIK, 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., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BODE, W. & SCHWAGER, P. (1975). *J. Mol. Biol.* **98**, 693–717.

BRÜNGER, A. T. (1992*a*). *X-PLOR, Version* 3.1. *A System for X-ray Crystallography and NMR*. Yale Univ. Press, New Haven, CT, USA.

BRÜNGER, A. T. (1992*b*). *Nature (London)*, **355**, 472–475.

COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4. (1994). *Acta Cryst.* D**50**, 760–763.

EARNEST, T., FAUMAN, E., CRAIK, C. S. & STROUD, R. (1991). *Proteins*, **10**, 171–187.
FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.
HARTLEY, B. S. & KAUFFMAN, D. L. (1966) *Biochem. J.* **101**, 229.
HEIMSTAD, E. S., HANSEN, L. K. & SMALÅS, A. O. (1995). *Protein Eng.* In the press.
HENDRICKSON, W. (1985). *Methods. Enzymol.* **115**, 252–270.
KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
LASKOWSKI, R., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MALE, R., LORENTS, J. B., SMALÅS, A. O. & TORRISSEN, K. R. (1995). *Eur. J. Biochem.* Submitted.
MARQUART, M., WALTER, J., DEISENHOFER, J., BODE, W. & HUBER, R. (1983) *Acta Cryst.* B**39**, 480–490.
MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
SACK, J. S. (1988). *J. Mol. Graph.* **6**, 244–245.
SMALÅS, A. O., HEIMSTAD, E. S., HORDVIK, A. WILLASSEN, N. P. & MALE, R. (1994). *Proteins Struct. Funct. Genet.* **20**, 149–166.
SMALÅS, A. O. & HORDVIK, A. (1993). *Acta Cryst.* D**49**, 318–330.
SMALÅS, A. O., HORDVIK, A., HANSEN, L. K., HOUGH, E. & JYNGE, K. (1990). *J. Mol. Biol.* **214**, 355–358.
WILKE, M. E., HIGAKI, J. N., CRAIK, C. S. & FLETTERICK, R. J. (1991). *J. Mol. Biol.* **219**, 511–523.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.