*Acta Cryst.* (1998). D**54**, 780–798

# Structure of a Non-psychrophilic Trypsin from a Cold-Adapted Fish Species

HANNA-KIRSTI SCHRØDER,[a] NILS P. WILLASSEN[b] AND ARNE O. SMALÅS[a]*

[a]*Protein Crystallography Group, Institute of Chemistry, Faculty of Science, University of Tromsø, N-9037, Tromsø, Norway, and* [b]*Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037, Tromsø, Norway. E-mail: arne.smalas@chem.uit.no*

(*Received 4 October 1997; accepted 2 December 1997*)

## Abstract

The crystal structure of cationic trypsin (CST) from the Atlantic salmon (*Salmo salar*) has been refined at 1.70 Å resolution. The crystals are orthorhombic, belong to space group $P2_12_12_1$, with lattice parameters $a$ = 65.91, $b$ = 83.11 and $c$ = 154.79 Å, and comprise four molecules per asymmetric unit. The structure was solved by molecular replacement with *AMoRe* and refined with *X-PLOR* to an $R$ value of 17.4% and $R_{free}$ of 21.5% for reflections $|F| > 3\sigma_F$ between 8.0 and 1.7 Å resolution. The four non-crystallographic symmetry (NCS) related molecules in the asymmetric unit display r.m.s. deviations in the range 0.31–0.74 Å for main-chain atoms, with the largest differences confined to two loops. One of these is the calcium-binding loop where the electron-density indicates a calcium ion for only one of the four molecules. In order to find structural rationalizations for the observed difference in thermostability and catalytic efficiency of CST, anionic salmon trypsin (AST) and bovine trypsin (BT), the three structures have been extensively compared. The largest deviations for the superimposed structures occur in the surface loops and particularly in the so-called 'autolysis loop'. Both the salmon enzymes possess a high methionine content, lower overall hydrophobicity and enhanced surface hydrophilicity, compared with BT. These properties have so far been correlated to cold-adaptation features, while in this work it is shown that the non-psychrophilic cationic salmon trypsin shares these features with the psychrophilic anionic salmon trypsin.

## 1. Abbreviations

CST, cationic salmon trypsin; AST, anionic salmon trypsin crystal form 1 (PDB entry 2TBS); AST-II, anionic salmon trypsin crystal form 2 (PDB entry 1BIT); BT, bovine trypsin (PDB entry 3PTB); NCS, non-crystallographic symmetry; Mol *A*–*D*, molecules *A*–*D* of cationic salmon trypsin; r.m.s., root-mean-squares; res., residue(s); $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |\langle I_h \rangle| \times 100\%$.

## 2. Introduction

The digestive proteinase trypsin is found in multiple forms in many species, both mammalian *e.g.* cattle (le Huerou *et al.*, 1990; Titani *et al.*, 1975), rat (Craik *et al.*, 1984; Fletcher *et al.*, 1987) and other vertebrates *e.g.* chicken (Wang *et al.*, 1995). In particular many marine species are known to have a rather complex isoenzyme pattern. Capelin (Hjelmeland & Raa, 1982), Antarctic krill (Osnes & Mohr, 1985), cod (Gudmundsdóttir *et al.*, 1993), anchovy (Martínes *et al.*, 1988) and crayfish (Kim *et al.*, 1994) have all been identified with multiple forms of trypsin. While the predominant isoforms in mammals are cationic, only anionic variants have been reported for the above-mentioned marine species. To our knowledge, Atlantic salmon and Chum salmon are the only fish species for which a cationic isoform has been isolated and characterized (Male *et al.*, 1995; Outzen *et al.*, 1996; Uchida *et al.*, 1986), although anionic isoforms seem to dominate in both species. Little is known about the possible functions of the multiple isoenzymes in an organism, but studies by Torrissen (1987) showed that there were variations in trypsin isoenzyme patterns between families of Atlantic salmon and one particular genotype could, for example, be associated with fish size (Torrissen, 1987, 1991). In Atlantic salmon five different clones of trypsin have been identified and sequenced (Male *et al.*, 1995), and judging from the amino-acid sequences one of these is cationic and four are anionic isoforms. One cationic and one main anionic fraction of salmon trypsin have been isolated and characterized (Outzen *et al.*, 1996), and the results from this study show that the two trypsin fractions possess notable differences in both activity and stability. The cationic form (CST) resembles the cationic bovine trypsin (BT), while the anionic form (AST) is considerably less stable at high temperature and low pH, and about 20–30 times more efficient (in terms of $k_{cat}/K_m$) in hydrolysing an amide substrate, compared to the cationic salmon and bovine enzymes. The GdnHCl-induced unfolding experiments showed that the free energy of unfolding of CST, AST and BT were 10.5, 4.7 and 14.8 kcal mol$^{-1}$ (1 kcal = 4.184 kJ), respectively. Unfolding by GdnHCl is believed to monitor the hydrophobic stability of a

Table 1. *Summary of data-collection statistics*

| | |
|---|---|
| Crystal size (mm) | 0.2 × 0.1 × 0.1 |
| Space group | $P2_12_12_1$ |
| Cell parameters (Å) | $a$ = 65.91, $b$ = 83.11, $c$ = 154.79 |
| Resolution range (Å) | 28.0–1.7 |
| No. of observed reflections | 451079 |
| No. of unique $h,k,l$ reflections | 76357 |
| Multiplicity | 5.9 |
| % reflections $I > 3\sigma_{(I)}$ | 90.5 |
| Overall $B$ factor from the Wilson plot (Å$^2$) | 19.6 |
| $I/\sigma_{(I)}$ | 3.6 |
| $R_{\rm merge}$ (%) | 12.3 |
| Completeness (%) | 81.2 |
| Last resolution shell 1.79–1.70 Å | |
| Completeness (%) | 52.3 |
| $R_{\rm merge}$ (%) | 25.8 |

protein (Monera *et al.*, 1994), indicating fewer and/or weaker stabilizing hydrophobic interactions in AST compared to the cationic enzymes. Anionic salmon trypsin was the first psychrophilic enzyme for which a three-dimensional structure was solved at atomic resolution. Coordinates for two different crystal forms of the enzyme have been deposited to the Protein Data Bank (Bernstein *et al.*, 1977; PDB entry 2TBS: Smalås & Hordvik, 1993; 1BIT: Berglund *et al.*, 1995*a*).

The cold-adaptation behaviour, characterized by a much higher physiological efficiency and a lower thermostability of the psychrophilic (cold loving) enzymes compared to their mesophilic counterparts, is not unique to the trypsin enzymes (Simpson & Hard, 1984; Genicot *et al.*, 1988; Ásgeirsson *et al.*, 1989; Genicot *et al.*, 1996). This behaviour has also been demonstrated in the case of $\alpha$-amylase (Feller *et al.*, 1992), triose phosphate isomerase (Rentier-Delrue *et al.*, 1993), alkaline phosphatase (Kobori *et al.*, 1984), lactate dehydrogenase (Vckovski *et al.*, 1990), lipases (Arpigny *et al.*, 1993) and subtilisin (Davail *et al.*, 1994). In spite of the differences in physical behaviour, our recent study of the anionic salmon trypsin revealed a high degree of similarity with bovine trypsin in both primary and tertiary structure (Smalås *et al.*, 1994). The two molecules have identical amino acids in 65% of the sequence, and the mean displacement in main-chain atomic positions between the two trypsins is 0.7 Å. The underlying structural basis for the observed low-temperature adaptation, is still under debate. In this study we will try to illuminate this issue by discussing the high-resolution crystal structure of the cationic salmon trypsin, a non-psychrophilic enzyme (Outzen *et al.*, 1996) originating from the cold-blooded Atlantic salmon.

## 3. Experimental

### 3.1. *Purification and crystallization*

Cationic salmon trypsin was isolated from pancreatic tissue by affinity, hydrophobic interaction, ion-exchange and gel-filtration chromatography (Outzen *et al.*, 1996). Crystals were obtained by the vapour-diffusion technique using a protein solution of 15 mg ml$^{-1}$ protein and 60 m$M$ benzamidine, and 6 µl drops containing a 1:1 mixture of protein and reservoir solution. At 310 K well shaped crystals grew within a period of a few days to three weeks for wells containing ammonium sulfate in the range of 1.60–1.75 $M$ and 0.1 $M$ acetate buffer at pH 4.6. Crystallization parameters were found by improving the conditions initially obtained from the sparse-matrix screening method described by Jancarik & Kim (1991). Crystals used for data collection were about 0.2 × 0.1 × 0.1 mm, they were found to be orthorhombic, space group $P2_12_12_1$ and cell dimensions $a$ = 65.91, $b$ = 83.11 and $c$ = 154.79 Å.

### 3.2. *Data collection and processing*

Data were collected in two rounds at the Swiss–Norwegian Beamline (D1) at the European Synchrotron Radiation Facility (ESRF) in Grenoble using a MAR image-plate system, a wavelength of 0.873 Å and processed with the *XDS* (Kabsch, 1988) and *DENZO* (Otwinowski, 1993) program packages. Scaling and merging were carried out using *SCALA* and *AGROVATA* in the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Crystals used for data collection diffracted beyond 1.7 Å. In order to achieve satisfactory completeness of the final data, three scans from the first and five scans from the second data-collection round were merged yielding an $R_{\rm merge}$ of 12.3% in the 28–1.7 Å resolution range. The completeness of the data set was 99% to 3 Å, and reduced to 81% for data to 1.7 Å resolution. A Wilson plot (Wilson, 1949) calculated from the final data set gave an overall $B$ factor of 19.6 Å$^2$. A summary of crystallographic information and data processing is given in Table 1.

### 3.3. *Structure determination*

The structure was solved by the molecular-replacement method using *AMoRe* (Navaza, 1994).

3.3.1. *Search model.* The search model was built from bovine trypsin (Marquart *et al.*, 1983; entry 3PTB), mutating all residues according to the gene sequence of CST (Male *et al.*, 1995). These coordinates were superimposed onto anionic salmon trypsin (Smalås *et al.*, 1993; entry 2TBS) which was used as a second search model. Searching with either of the two models gave comparable results, and thus only the solution from the prebuilt model with the CST sequence will be presented here.

3.3.2. *Self rotation.* The self-rotation map was calculated with *POLARRFN* (Collaborative Computational Project, Number 4, 1994) using reflections from 8.0 to 4.0 Å. However, the self-rotation function failed to reveal any non-crystallographic symmetry related molecules in the asymmetric unit. The crystal volume indicates three or four molecules per asymmetric unit

Table 2. *Results of the molecular-replacement search using AMoRe*

$\alpha$, $\beta$, $\gamma$ are eulerian angles, **Tx**,**Ty**,**Tz** are translation vectors in fractional coordinates [for the prebuilt model superimposed on the anionic salmon trypsin structure (2TBS)], and CC is the correlation coefficient. The improvement of the CC and $R$ value while fixing Mol $A$, then $A$ and $B$ and so on, are shown along with the final values after rigid-body refinement in *AMoRe*.

| | $\alpha$ | $\beta$ | $\gamma$ | **Tx** | **Ty** | **Tz** | | CC | $R$ value |
|---|---|---|---|---|---|---|---|---|---|
| Mol $A$† | 48.75 | 64.88 | 142.24 | 0.0486 | 0.7341 | 0.5682 | | 8.5 | 55.6 |
| Mol $B$ | 26.73 | 84.36 | 215.84 | 0.1493 | 1.0638 | 0.3382 | FIX $A$ | 23.3 | 51.5 |
| Mol $C$ | 97.82 | 77.33 | 40.54 | 1.3432 | 0.4635 | 0.4272 | FIX $A + B$ | 35.5 | 47.7 |
| Mol $D$ | 111.37 | 50.35 | 137.49 | −0.1743 | 0.4087 | 0.3122 | FIX $A + B + C$ | 48.5 | 43.0 |
| | | | | | | | Rigid body | 57.7 | 39.3 |

† Mol $A$ was further moved $(-x + 0.5, -y + 1.0, z - 0.5)$ in order to pack the cell appropriately.

Table 3. *Statistics for each round of crystallographic X-PLOR refinement of cationic salmon trypsin*

| | Resolution range (Å) | No. of reflections | NCS restraints | No. of waters | $R_{working}$ | $R_{free}$ | Remarks |
|---|---|---|---|---|---|---|---|
| 1 | 15–2.3 | 37285 | Y | 0 | 26.7 | 32.5 | Molecular dynamic simulations with slow-cool (3000 K), grouped temperature-factor refinement, restrained NCS. |
| 2 | 15–2.1 | 47741 | Y | 88 | 25.2 | 29.5 | Simulated annealing (3000 K), individual temperature factors were included in all subsequent steps. 4 $Ca^{2+}$ ions were introduced. |
| 3 | 15–1.9 | 61024 | Y | 194 | 23.7 | 28.0 | Slow-cool (3000 K), energy minimization, 6 $SO_4^{2-}$ ions were included. In later stages 4 more $SO_4^{2-}$ ions were added. From here on poor defined residues/atoms were refined with occupancy 0. |
| 4 | 15–1.7 | 76153 | Y | 199 | 24.1 | 27.6 | Slow-cool (3000 K), energy minimization. Some loops and residues were not regarded with NCS restraints. |
| 5 | 15–1.7 | 76153 | Y | 293 | 22.8 | 27.2 | Slow-cool (2000 K), energy minimization. |
| 6 | 15–1.7 | 68897 | N | 373 | 20.9 | 24.2 | Minimization, slow-cool (2000 K). From here on **no** NCS restraints were used, 2 $Ca^{2+}$ ions removed. $3\sigma_F$ cut-off was introduced. |
| 7 | 8–1.7 | 68062 | N | 396 | 17.7 | 22.1 | Energy minimization and focus on the water structure. |
| 8 | 8–1.7 | 68062 | N | 405 | 17.3 | 21.8 | Energy minimization, only minor changes. |
| 9 | 8–1.7 | 68062 | N | 419 | 17.2 | 21.3 | Energy minimization, 1 $Ca^{2+}$ ion removed. |
| 10 | 8–1.7 | 68062 | N | 379 | 17.3 | 21.5 | Energy minimization. Waters with inappropriate density were deleted. |

with resulting $V_m$ values of 2.9 and 2.2 Å$^3$ Da$^{-1}$, respectively, yielding a water content of 44–58% which is considered reasonable. Further searches were therefore carried out in order to find three or four independent rotation and translation solutions which would give 12 or 16 molecules in the cell.

3.3.3. *Cross-rotation and translation search.* Application of the fast rotation function, using reflections in the resolution range 15.0–4.0 Å and an integration radius of 20 Å, gave five major peaks of which four corresponded to the correct rotation solutions. Since the self-rotation search failed to reveal the number of molecules in the asymmetric unit, all possible cross-rotation solutions were tried in the translation search using the same resolution range. Four different rotations gave translation vectors with high correlation coefficients, indicating four molecules in the asymmetric unit. A two-body translation search was started by fixing molecule $A$ (Mol $A$) at the position obtained from the first translation search, while searching for the translation of Mol $B$. The result from this translation search was input to the three-body search, by fixing both Mol $A$ and $B$, and finally a four-body search was carried out with $A$, $B$ and $C$ fixed, searching for a translation solution for Mol $D$. This process increased the correlation coefficient (CC) from 9 to 49% as listed in Table 2. Since the CC is the main parameter to consider for translation search (Navaza & Vernoslova, 1995), this high improvement by introducing more molecules in the cell, supported the assumption of four independent molecules in the asymmetric unit. After rigid-body optimization in *AMoRe*, a correlation coefficient of 57.7% and an $R$ value of 39.3% was obtained (Table 2). None of the four located molecules are related by non-crystallographic symmetry that are close to crystallographic symmetry.

### 3.4. *Refinement and model building*

Crystallographic refinement was carried out using *X-PLOR* version 3.8 (Brünger, 1992*a*,*b*). The starting model was based on the bovine trypsin 3PTB (Marquart *et al.*, 1983) with all side chains mutated according to the gene sequence of CST (Male *et al.*, 1995), rotated and translated according to the molecular replacement solution in Table 2. All temperature factors were reset to 20.0 Å$^2$ and all solvent molecules, the calcium ion and

Table 4. *Final refinement statistics, including all atoms in the asymmetric unit of the cationic salmon trypsin crystals*

| | |
|---|---|
| Resolution range (Å) | 8.0–1.7 |
| No. of reflections $>3\sigma_F$ | 68062 |
| Final $R$ value, all data included (%) | 17.4 |
| $R_{\text{working}}$ (%) | 17.3 |
| $R_{\text{free}}$ (%) | 21.5 |
| Total no. of atoms | 7106 |
| Non-H protein atoms | 6640 |
| Water molecules | 379 |
| Benzamidine molecules | 4 |
| Sulfate ions | 10 |
| Calcium ion | 1 |
| Luzzati (1952) coordinate error (Å) | 0.18 |
| Read (1986) $\sigma_A$ coordinate errors (Å) | 0.200 |
| R.m.s. deviations from standard | |
| Bond lengths (Å) | 0.013 |
| Angles (°) | 2.697 |
| Dihedral angles (°) | 26.356 |
| Improper angles (°) | 1.087 |
| Mean $B$ factors (Å$^2$) | |
| All atoms | 22.93 |
| All water molecules | 33.23 |
| Benzamidines | 15.52 |
| $Ca^{2+}$ ion | 46.81 |
| Sulfate ions | 43.08 |
| R.m.s. $B$-factor difference | |
| Along bonds | 2.307 |
| Along angles | 3.455 |
| Correlation coefficients | |
| All protein atoms | 0.912 |
| Main chain | 0.926 |
| Side chain | 0.880 |

the benzamidine molecule were removed prior to the first refinement round. Initial refinement was carried out using reflections in the 15–2.3 Å resolution shell. Further on, the resolution was expanded to 15–2.1, 15–1.9, 15–1.7 and 8–1.7 Å in subsequent cycles. Throughout the refinement a test set including 5% randomly selected reflections were used for $R_{\text{free}}$ calculations. In the final refinement round all reflections from 8.0 to 1.7 Å with $|F| > 3\sigma_F$ were included. The progress of the refinement is summarized in Table 3.

$\sigma_A$-weighted (Read, 1986) $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated with *X-PLOR* and visualized using *O* (Jones *et al.*, 1991). The model, including four molecules, was refined with restrained non-crystallographic symmetry (NCS) in the first five rounds with more data being gradually included (Table 3). For the last five rounds, all molecules were refined individually. In total, ten rounds of refinement and manual rebuilding were carried out, the first six including simulated annealing, followed by four rounds of conjugate-gradient minimization and restrained temperature-factor refinement until the $R_{\text{working}}$ and $R_{\text{free}}$ converged. The correctness of the stereochemistry of the model was checked using *PROCHECK* (Laskowski *et al.*, 1993), calculations of r.m.s. deviations from ideal values for bonds, angles, dihedral and improper angles were performed in *X-PLOR* and residue based correlation coefficients (CC) were calculated by *O* (Zou & Mowbray, 1994). For each round of the refinement, new solvent molecules were added if they formed at least one stereochemically reasonable hydrogen bond and were above the $3\sigma$ level in the difference-density map. The solvent water molecules were included as O atoms with occupancy 1, and ten sulfate ions and one calcium ion with various occupancies were also interpreted. Water molecules were labelled *A–D* depending on their nearest trypsin molecule. Water molecules found at similar positions in all four CST structures are numbered from residue 300 and those unique for Mol *A*, *B*, *C* and *D* are numbered from 400, 500, 600 and 700, respectively. Water molecules were deleted if they failed to reappear at $1\sigma$ in the $\sigma_A$-weighted $2F_o - F_c$ density map after a new refinement round. The model was refined to an $R_{\text{free}}$ of 21.5% and $R_{\text{working}}$ of 17.3%. Inclusion of all data in the final refinement round yielded a conventional $R$ value of 17.4% for 68 062 reflections with $|F| > 3\sigma_F$ in the 8.0–1.7 Å resolution shell. Analysis of intermolecular contacts, water-accessible surface areas and r.m.s. differences were carried out using *CONTACT*, *AREAIMOL* and *LSQKAB*, respectively, from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994).

## 4. Results

### 4.1. *Quality of the refined structure*

The final model of CST consists of 6640 non-H protein atoms distributed on 4 × 223 amino acids, 379

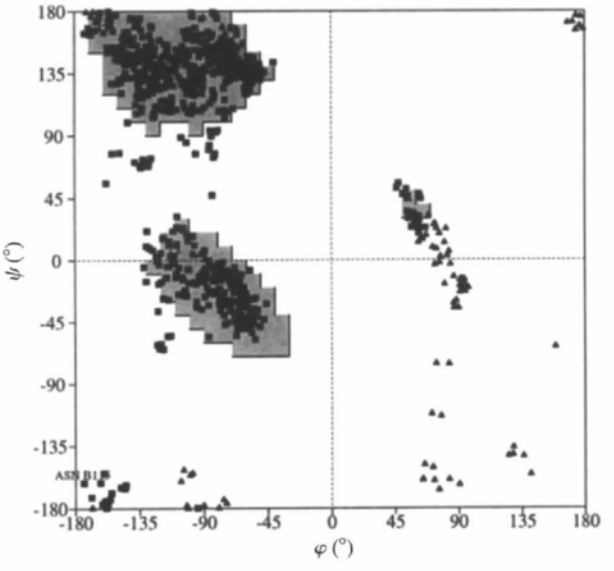



Fig. 1. Ramachandran plot of main-chain dihedral angles of cationic salmon trypsin with 89.7% of the 760 non-glycine and non-proline residues in the energetically favourable core regions. Residue Asn*B*115, with conformational angles outside the allowed regions, is labelled. Glycine residues are shown as triangles.

water O atoms, four benzamidine molecules, one calcium ion and ten sulfate ions. The stereochemistry of the final model is good, with deviations from ideal values within ranges found for other comparable protein structures as given in Table 4. A Ramachandran plot (Ramachandran *et al.*, 1963) of the refined structure

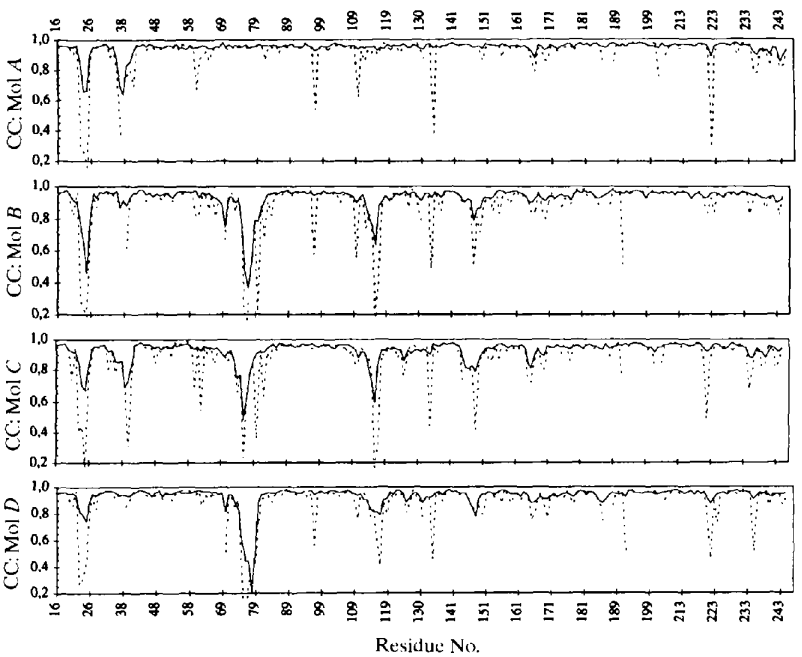



Fig. 2. Variation in real-space correlation coefficients (CC) along the poly polypeptide chains, averaged over main-chain atoms (bold lines) and side-chain atoms (broken lines). The top plot is for Mol *A* followed by corresponding plots for Mol *B*, *C* and *D*, respectively.

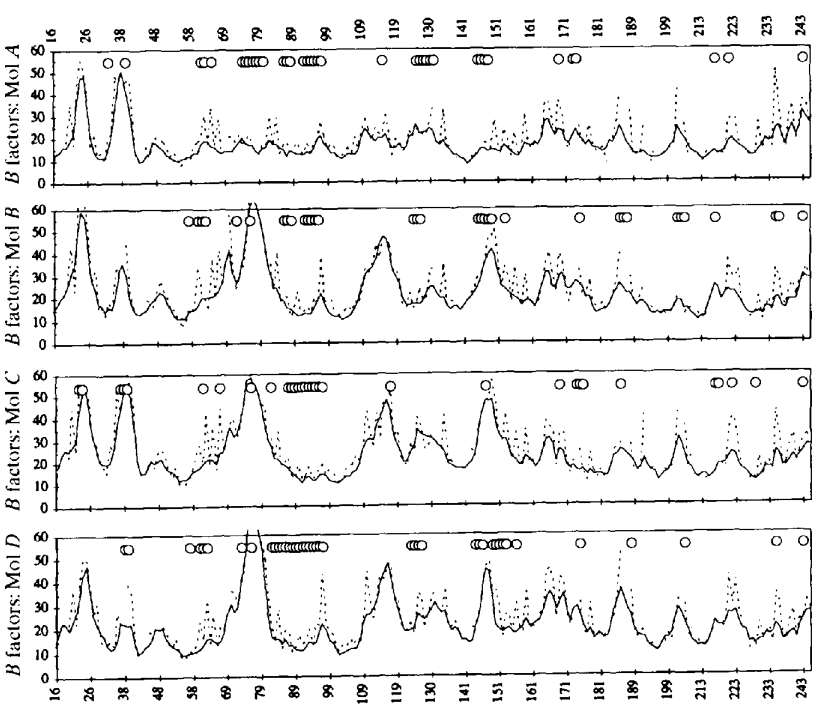



Fig. 3. Residue-based *B* factors (Å$^2$) plotted along the four peptide chains. Main-chain values are given in bold lines and side-chain values in dotted lines. Residues involved in packing interactions <3.4 Å are marked with (○).

Table 5. *Atoms and residues not fully defined in the 1σ $2F_o - F_c$ map and thus refined with zero occupancy*

| Residue | Mol *A* | Mol *B* | Mol *C* | Mol *D* |
|---|---|---|---|---|
| 23–25 | Side chains | Side chains | Side chains | Side chains |
| Ser37 | Cα, Cβ, Oγ | – | – | – |
| Tyr39 | Cδ1, Cε2 | Cδ2, Cε2, Cζ | N, Beyond Cα | – |
| His40 | – | – | Beyond Cβ | – |
| Phe41 | Cδ2, Cε2 | – | – | – |
| Lys60 | Cδ, Cε, Nζ | – | Cε, Nζ | – |
| Arg62 | – | – | Beyond Cβ | – |
| Glu70 | – | Cβ, Cγ | Beyond Cβ | Beyond Cα |
| 74–80 | – | Many side chains | Many side chains | Side chains, |
| | – | Main-chain 77 | Main-chain 76 | Main-chain 76–78 |
| Phe82 | – | Cδ2, Cε2 | Beyond Cγ | – |
| Arg97 | Beyond Cβ | Beyond Cδ | Cγ, Cδ, Nη1, Nη2 | Beyond Cβ |
| Lys110 | Cγ, Cε, Nζ | Beyond Cγ | – | – |
| Asn115 | | – | C, Cβ | – |
| Ser116 | – | Cα, Cβ, Oγ | Cβ, Oγ | Oγ |
| Tyr117 | – | Beyond Cγ | Beyond Cα | Beyond Cα |
| Arg135 | Beyond Cγ | Beyond Cβ | Beyond Cβ | Beyond Cγ |
| Gln192 | – | Cδ, Oε1, Nε2 | – | Beyond Cβ |
| Tyr217 | – | – | – | Beyond Cγ |
| Arg222 | Beyond Cβ | – | Beyond Cβ | Beyond Cβ |
| Arg235 | – | – | Cδ, Nη1 | Beyond Cβ |

generated by *PROCHECK* (Laskowski *et al.*, 1993), shows that the majority of the main-chain dihedral angles (89.7%) fall into the most favoured regions (Fig. 1). Asn115 from Mol *B* has adopted a less favoured conformation as can be explained from its poorly defined electron density and its location in the rather flexible interdomain loop (residues 109–133). See Fig. 10 for nomenclature of the structural regions. The remaining 10.3% of the non-glycine and non-proline residues are in the additionally allowed regions. Analysis of the side-chain parameters $\chi_1$ and $\chi_2$ shows that 90% are clustered around the ideal *trans* or *gauche* conformations. From a Luzzati plot (Luzzati, 1952) the mean positional error is estimated to about 0.18 Å and the $\sigma_A$ plot method by Read (1986) gives an r.m.s. coordinate error of 0.20 Å. The mean temperature factor of 22.9 Å$^2$ for all atoms in the model is somewhat higher than the value of 19.6 Å$^2$ calculated from the Wilson plot (Table 1). The mean temperature factors for main-chain and side-chain atoms are 20.7 and 23.7 Å$^2$, respectively, while *B* values for water O atoms range from 9.8 to 59.9 Å$^2$, with an average of 33.2 Å$^2$.

Examination of the electron density of CST shows that all well defined residues agree with the reported gene sequence by Male *et al.* (1995). In some cases two alternative conformations were clearly visible in the electron density, but no attempt was made to model more than one conformation for the side chains. The variation in average correlation coefficients and temperature factors along the four polypeptide chains are shown in Figs. 2 and 3, and overall most of the residues are well defined with continuous electron density and high correlation coefficients. The exceptions are in some surface loops and in the calcium-binding loop of Mol *B*, *C* and *D*. 63 of the 892 residues (7.1%) have been classified as poorly defined with main-chain or side-chain CC lower than 0.8 (for further details see Table 5). The benzamidine inhibitors located in the active-site cleft are all well defined with a mean *B* factor of 15.2 Å$^2$, and seven water molecules are found in or near the active site in each of the four molecules.

### 4.2. *Crystal packing interactions*

The crystals of cationic salmon trypsin are tightly packed with 44% solvent and a $V_m$ of 2.2 Å$^3$ Da$^{-1}$, which is probably one reason for the high-diffraction capability of the present crystal form. As can be seen from Table 6, many of the packing interactions are intermolecular hydrogen bonds, while there are relatively few hydrophobic interactions. One hydrophobic packing area worth mentioning is the aromatic–aromatic stacking of the side chains of Tyr*A*217 and Tyr*C*217, for which the main-chain atoms form a part of the walls of the specificity pockets. Residue 217 is also in an environment close to Trp215 and Tyr172 in both molecules. Of the 165 unique intermolecular hydrogen bonds, 65 are between non-crystallographically related molecules and 100 are between symmetry-related molecules (Table 6). The packing environments shown in Fig. 4, are highly different among the four independent molecules. The number of close contacts and regions involved in such interactions are not identical, and give rise to significantly different loop conformations and flexibility among the four molecules, as discussed in the following sections.

### 4.3. *Comparison of the four independent molecules in the asymmetric unit*

The four molecules in the asymmetric unit were refined individually for the last five refinement rounds, and the differences in CC, *B* factors and conformations

Table 6. *Number of unique intermolecular interactions and potential hydrogen bonds in the crystal of cationic salmon trypsin*

| Residue | Contacts < 4 Å | Hydrogen bonds | Contact with | Symmetry operation |
|---|---|---|---|---|
| (*a*) NCS-related contacts | | | | |
| *A*59–*A*62 | 27 | 7 | *B*90–95 | NCS |
| *A*86–*A*97 | 63 | 21 | *B*37, *B*59–62, *B*86–93, *C*192 | NCS |
| *A*143–150 | 48 | 15 | *D*235–239, *D*243, *C*97 | NCS |
| *A*173–192 | 18 | 4 | *C*97, *C*146–149, *C*219–221, *D*114 | NCS |
| *A*217–224 | 31 | 5 | *C*175, *C*217–219 | NCS |
| *B*39–41, *B*57, *B*60, *B*74 | 27 | 2 | *C*34–39, *C*66–75 | NCS |
| *B*149–153 | 12 | 4 | *C*64, *C*76–82 | NCS |
| *C*97 | 1 | 1 | *D*235 | NCS |
| *C*166–178 | 27 | 6 | *D*202–204, *D*122–127 | NCS |
| SUM | 254 | 65 | | |
| (*b*) Symmetry-related contacts | | | | |
| *A*32–41 | 36 | 3 | *D*39–41, *D*57–60 | 2 |
| *A*64–A66 | 7 | 2 | *D*39, *D*149–*D*151 | 2 |
| *A*73–*A*76, *A*82 | 34 | 8 | *D*38–40, *D*73–74, *D*149–153 | 2 |
| *A*76–80 | 42 | 16 | *B*125–127, *B*202–204, *B*235 | 3 |
| *A*110–111, *A*115–117 | 13 | 3 | *B*114–115, *B*120, *B*202–203 | 3 |
| *A*122–134 | 25 | 10 | *B*222, *B*186–188B, *B*145–148 | Ident |
| *A*141, *A*149–153 | 6 | 1 | *D*38–39, *D*62 | 2 |
| *A*169–174 | 9 | 2 | *C*23–24 | 2 |
| *A*202–204 | 8 | 3 | *B*186, *B*221–222 | Ident |
| *A*243 | 2 | 1 | *D*188 | 1 |
| *B*72–77 | 17 | 4 | *C*185–186, *C*222 | 2 |
| *B*124–125 | 3 | 2 | *D*150 | 1 |
| *B*153 | 5 | 1 | *C*186 | 2 |
| *B*173–175 | 10 | 4 | *D*173–*D*175, *D*217 | 2 |
| *B*217 | 4 | 1 | *D*97 | 2 |
| *B*235–243 | 28 | 10 | *D*17, *D*144–151, *D*156, *D*221B | 1 |
| *C*59–62, *C*76 | 16 | 6 | *D*90–96, *C*186 | 2 |
| *C*86–96 | 51 | 14 | *D*59–62, *D*87–93, *D*245 | 2 |
| *C*117 | 3 | 1 | *C*147 | 2 |
| *C*239–243 | 17 | 8 | *D*76–82 | Ident |
| SUM | 336 | 100 | | |

Symmetry operators: Ident = $x, y, z$; 1 = $-x + 0.5, -y, z + 0.5$; 2 = $x + 0.5, -y + 0.5, -z$; 3 = $x, y + 0.5, -z + 0.5$.

are thus mainly due to different packing interactions. Mol *A* seems to be the best defined molecule, with the highest fraction of well defined atoms (Table 5), lowest mean *B* value and highest CC, and is therefore used as reference in this comparison. Superposition of Mol *B*, *C* and *D* onto Mol *A*, gives mean r.m.s. differences for all non-H atoms of 1.1, 0.6 and 1.2 Å, respectively (Table 7). The corresponding values for main-chain and C$\alpha$ atoms only are ranging from 0.3 to 0.8 Å. These values are within the ranges claimed to be reasonable differences between NCS-related molecules refined to high resolution (see *e.g.* Kleywegt, 1996).

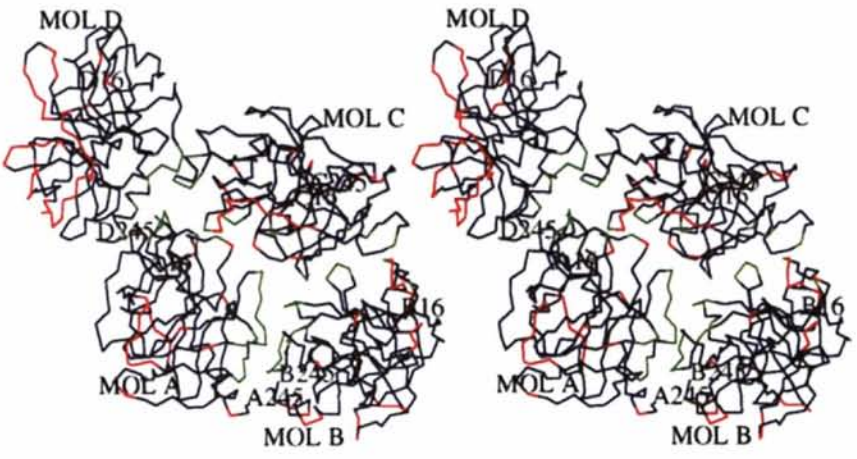



Fig. 4. Packing regions in the asymmetric unit of the cationic salmon trypsin crystals including Mol *A*–*D*. Green denotes regions involved in intermolecular contacts shorter than 3.4 Å between non-crystallographically related molecules, while red denotes such contacts between symmetry-related molecules. Produced by *O/oplot* (Jones *et al.*, 1991).

Table 7. *R.m.s. differences between the non-crystallographically related CST molecules for $C_\alpha$, main-chain, side-chain and all protein atoms*

Only atoms refined with full occupancy are included. R.m.s. differences for all atoms excluding residues 37–41 and 69–79 are also shown.

| | *A–B* | *A–C* | *A–D* | *B–C* | *B–D* | *C–D* |
|---|---|---|---|---|---|---|
| $C_\alpha$ atoms | 0.719 | 0.312 | 0.763 | 0.785 | 0.376 | 0.800 |
| Main-chain atoms | 0.721 | 0.312 | 0.768 | 0.715 | 0.388 | 0.737 |
| Side-chain atoms | 1.437 | 0.903 | 1.580 | 1.322 | 0.794 | 1.417 |
| All protein atoms | 1.097 | 0.641 | 1.194 | 1.027 | 0.601 | 1.087 |
| All ÷ 2 loops† | 0.576 | 0.537 | 0.664 | 0.631 | 0.463 | 0.699 |

† Excluding residues 37–41 and 69–79.

A superposition of the four molecules is shown in Fig. 5 and r.m.s. differences as a function of residues for Mol *B*, *C* and *D* are plotted in Fig. 6. The largest differences for main-chain atoms (>1.3 Å) are confined to two loops, namely the N$\beta$1–N$\beta$2 loop (residues 37–41) and the calcium-binding loop (residues 69–80). For both these regions, Mol *A* and *C* have adopted one conformation, while *B* and *D* have another, as also reflected in the r.m.s. differences in Table 7. Residues 37–42 are poorly defined for both Mol *A* and *C*, whereas this range is well defined in Mol *B* and *D* (Fig. 2), reflecting an inherent conformational variability of this region. In fact, the N$\beta$1–N$\beta$2 loop and the calcium-binding loop are to a high extent responsible for the large overall differences among the molecules, and by omitting these loops, the calculations give similar r.m.s. deviations for all pairs of molecules (Table 7). Many large differences in side-chain conformation among the four molecules coincide with residues with low CC, but many side chains are forced into different conformations due to different intermolecular packing environments.

### 4.4. *Temperature factors and intermolecular contacts*

The variation in loop mobility arising from different crystal packing environments is illustrated in Fig. 7,

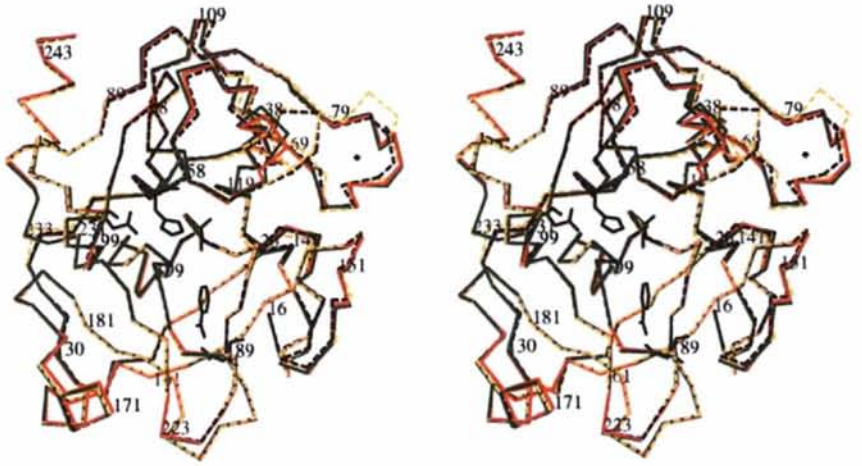



Fig. 5. Superposition of the $C_\alpha$ atoms of the four independent molecules in the asymmetric unit. Mol *A* (red) and *C* (green) are shown in bold lines and Mol *B* (blue) and *D* (orange) in broken lines. The two most striking differences among the four molecules can be seen in the upper right of the molecules (residues 37–41 and 69–79). The three catalytic residues (Asp102, His57 and Ser195), benzamidine inhibitors and the calcium ion are also included. Every 20th residue is labelled.

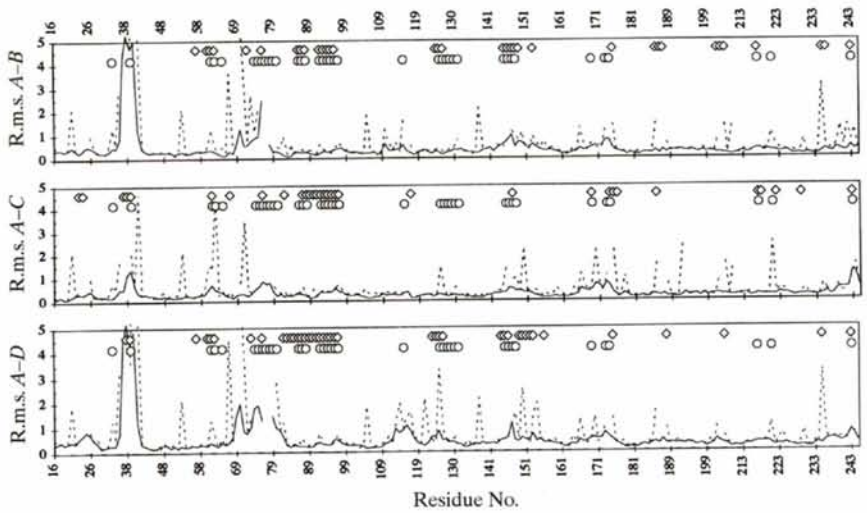



Fig. 6. Residue averaged r.m.s. differences (Å) between Mol *A* and *B* (top), Mol *A* and *C* (middle) and Mol *A* and *D* (bottom). Differences for main-chain atoms (bold line) and side-chain atoms (dashed line) of each residue are plotted as a function of residue number. Residues involved in intermolecular contacts shorter than 3.4 Å are marked with circles (○) for Mol *A* and diamonds (◇) for the other molecules. Only atoms with occupancy 1.0 are included.

where the backbones are colour coded according to the normalized temperature factors of the respective molecules. The mean temperature factors of all protein atoms in Mol *A*, *B*, *C* and *D* are 18.4, 23.4, 24.6 and 22.5 Å$^2$, respectively. The lower *B* value of Mol *A* is mainly due to reduced flexibility of three loop regions (Figs. 3 and 7); the calcium-binding loop, residues 114–117 of the interdomain loop and residues 146–150 of 'the autolysis' loop. Significant differences in temperature factors among Mol *B*, *C* and *D* are only seen for the N$\beta$1–N$\beta$2 loop. The N$\beta$1–N$\beta$2 loop of Mol *A* packs against the same symmetry-equivalent loop in Mol *D*, and by superimposing Mol *B* onto Mol *A*, it becomes clear that the loop conformation in Mol *B* would not fit the packing arrangement found for Mol *A*.

The motion of external regions of Mol *A* is in general more restricted compared to the other three molecules due to a higher number of intermolecular contacts. Mol *A* forms 369 contacts ($\leq$4.0 Å) to neighbouring molecules in the cell, which is about 100 more than for Mol *B*, *C* and *D*. The calcium-binding loop (residues 69–80) is responsible for a large portion of the additional contacts, which could be one reason for a more rigid conformation of Mol *A*. Even though Mol *A* is involved in more intermolecular contacts than the other molecules in the asymmetric unit, this is not reflected in the water-accessible surface (WAS) of the crystal-bound molecules (data not shown). In fact, while the variation in WAS among Mol *A*, *B* and *C* is within 1%, the WAS of Mol *D* is 7–8% higher.

### 4.5. *The calcium-binding loop*

A calcium ion, in the classical binding loop of trypsin, could only be assigned for molecule *A*. The high *B* factor (46.8 Å$^2$) of the calcium ion probably means that the occupancy of the ion is relatively low, but it appears with clear electron density throughout the refinement. The five calcium ligands found in Mol *A* of CST (Fig. 8*a*) are the same, with a similar octahedral arrangement, as

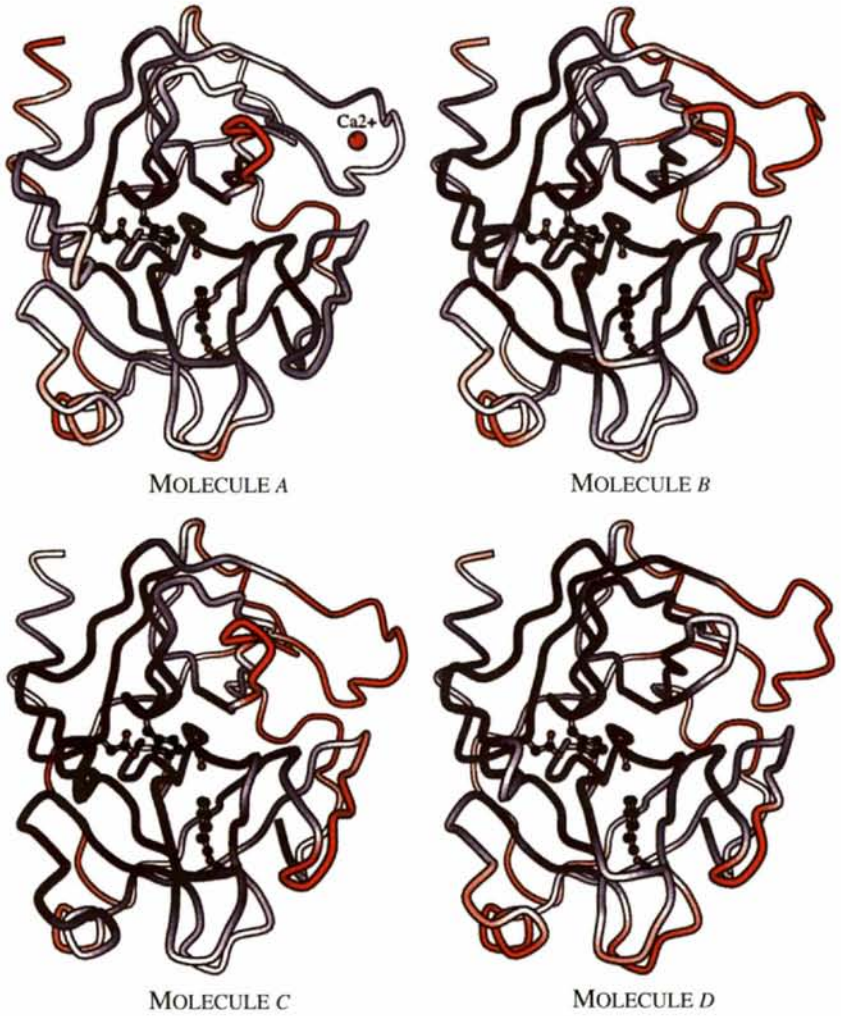
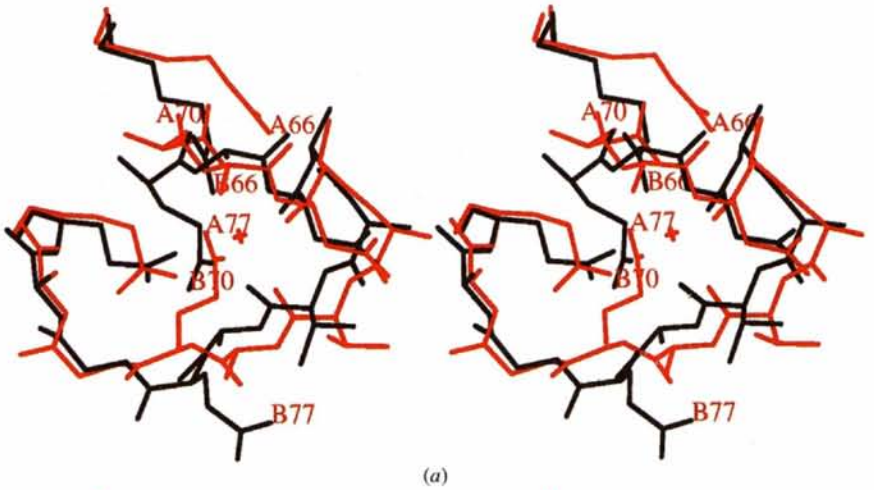
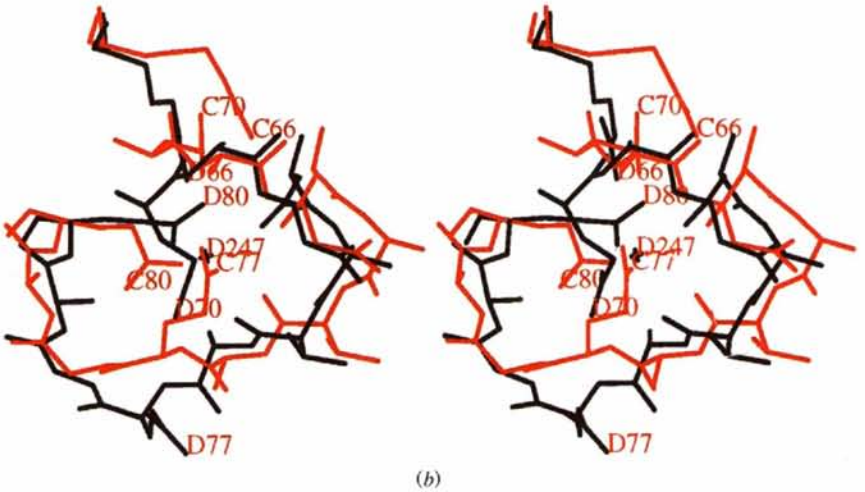



Fig. 7. Ribbon-style representation of the four cationic salmon trypsin molecules, colour coded according to their normalized *B* factors. The mean *B* factors are: Mol *A* 18.4, Mol *B* 23.4, Mol *C* 24.6 and Mol *D* 22.5 Å$^2$. Blue denotes *B* factors lower than 10 Å$^2$ below the mean value, while red denotes values higher than 10 Å$^2$ above the mean values, for each molecule. Catalytic residues, benzamidine inhibitors and the calcium ion are also drawn. The figure was created by *BobScript* (Esnouf, 1997).

corresponding ligands in anionic salmon trypsin (AST). The calcium-to-ligand distances are: 2.94 Å (Glu*A*70 $O^{\varepsilon 1}$), 2.44 Å (Asn*A*72 O), 2.39 Å (Val*A*75 O), 2.91 Å (Glu*A*77 $O^{\varepsilon 1}$) and 2.86 Å (Glu*A*80 $O^{\varepsilon 2}$). The sixth ligand in AST, a water molecule, did not appear with interpretable density for Mol *A*. Superposition of




(*a*)




(*b*)

Fig. 8. The arrangement of the calcium-binding loop including residues 66 and 70–80. In (*a*) Mol *A* (red) and *B* (blue), and in (*b*) Mol *C* (red) and *D* (blue) are pairwise superimposed. The calcium ion in Mol *A* and the water O atom bound in a similar position in Mol *D*, are drawn as crosses. Produced by *O/oplot* (Jones *et al.*, 1991).

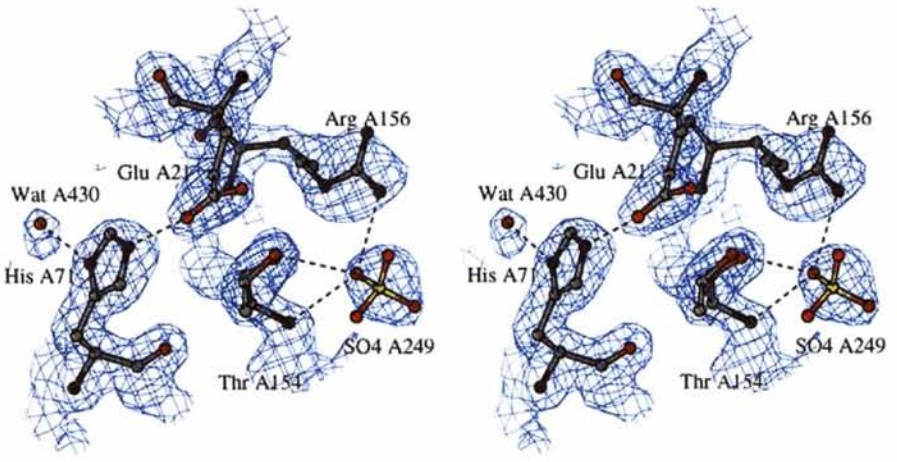



Fig. 9. A $\sigma_A$-weighted $2F_o - F_c$ map at 1.3$\sigma$ of the sulfate ion $SO_4$249 of Mol *A* bound to Arg*A*156 and Thr*A*154, and the neighbouring ion pair in Mol *A*: $Glu^{21}$–$His^{71}$. Created by *BobScript* (Esnouf, 1997).

the calcium-binding loop of the four molecules, Figs. 8(*a*) and 8(*b*), shows that Mol *A* and *C* have adopted one conformation and Mol *B* and *D* a different one. In Mol *C* residues 75–80 are more or less disordered, but even with a loop conformation similar to Mol *A*, no metal ion could be interpreted. For Mol *B* and *D* an apparent 'chain-reaction' seems to have occurred, initiated by Gln34 at the beginning of the N$\beta$1–N$\beta$2 loop. This residue is to our knowledge an Asn in all other trypsins, and due to the longer side chain and close packing contacts, Gln34 of Mol *B* is forced to move towards Arg66. The guanido group of Arg*B*66 has in turn replaced the position held by the carboxyl group of Glu*A*70 in Mol *A*, forcing the side chain of Glu*B*70 to a position similar to the side chain of Glu*A*77 (Mol *A*), resulting in a totally disordered Glu*B*77 (Fig. 8*a*). For Mol *B*, the side chain of Arg66 is well defined in the $2F_o - F_c$ map, forming a hydrogen bond with Glu*B*70 O and a potential salt bridge with Glu*B*80 $O^{\varepsilon 1}$. The final result of the 'chain-reaction' in *B* and *D* is that the new position of Arg66 disrupts calcium binding and imposes disorder. In Mol *D* a water molecule with a *B* value of 30.0 Å$^2$ could be interpreted in a similar position as the $Ca^{2+}$ ion in Mol *A*, although with totally different ligands (Figs. 8*a* and 8*b*). The water O atom is bound to Asn*D*72 O, Val*D*75 O and Arg*D*66 $N^{\eta 2}$, and the close interaction with the positively charged arginine, rules out the possibility of a low-occupancy calcium ion in Mol *D*.

### 4.6. *The solvent structure*

From the total of 379 water molecules in the asymmetric unit of CST and based on a 'closest to' criteria, 120, 84, 78 and 97 are associated with Mol *A*, *B*, *C* and *D*, respectively. 34 water molecules form hydrogen bonds to two trypsin molecules, where 25 are bridging water molecules from Mol *A*, 13 from Mol *B*, 16 from Mol *C* and 14 from Mol *D*. The higher number of water molecules in Mol *A*, probably also reflects the more ordered structure. The 36 water molecules found at similar positions in all four molecules, have a low WAS and all of them could be classified as internal. Internal water molecules are also to a high extent conserved among various trypsins, as 26 and 22 of the 36 common water molecules in CST, are found at corresponding positions in AST and BT, respectively.

### 4.7. *Sulfate ions*

The final model of CST includes ten sulfate ions, two located at similar positions for all four molecules, and two more at the surface of Mol *D*. The first conserved sulfate, $SO_4$248, is bound in the 'oxyanion hole' of the substrate binding site at an equivalent position as in the second crystal form of anionic salmon trypsin (PDB entry 1BIT, Berglund *et al.*, 1995*a*), bovine trypsin (1TLD; Bartunik *et al.*, 1989) and porcine pancreatic elastase (3EST; Meyer *et al.*, 1988). The sulfate molecules in all the mentioned structures form hydrogen bonds to His57 $N^{\varepsilon 2}$, Ser195 $O^{\gamma}$, Gly193 N and a water O atom. The second sulfate ion, $SO_4$249, found in all molecules of CST, is located at the surface of the protein, forming a salt bridge to the guanido group of Arg156, and hydrogen bonds to Asp153 N, Thr154 N and Thr154 $O^{\gamma 1}$ (except Mol *D*) as shown in Fig. 9. A similar sulfate is found in the crystal structure of bovine guanidino-benzoyl trypsin (1GBT; Mangel *et al.*, 1990) forming a salt bridge to Lys156. $SO_4$*D*250 is bound to Arg*D*62 and two other arginines of two different symmetry-related molecules (Arg*C*97 and Arg*D*235). Ion *D*251 with occupancy 0.6, is bound to Lys*D*60, Ser*D*61 N, a water molecule, two symmetry-related water molecules and a symmetry-generated serine residue.

## 5. Discussion

### 5.1. *Comparison with other trypsins*

The recent characterization of cationic and anionic salmon, bovine and porcine trypsin as described in §2 (Outzen *et al.*, 1996), showed that the cationic salmon and the cationic bovine trypsin are very similar in activity and thermostability. These and other observed differences are the basis for the subsequent comparison between the reported crystal structure of cationic salmon trypsin (CST) and the previously determined structures of anionic salmon (AST) and bovine trypsin (BT). For the sake of the comparison, it is important to have in mind that the total stability of a protein is the sum of all possible interactions observed, and that the stabilizing effects are additive. This is confirmed through mutation studies of T4 lysozyme (Zhang, Baase *et al.*, 1995), barnase (Serrano *et al.*, 1993), and other enzymes (Watanabe *et al.*, 1994; Eijsink *et al.*, 1995).

### 5.2. *Sequence and overall structure*

The alignment of primary structures of CST, AST and BT in Fig. 10, shows that the sequence identities are high; CST–BT 73.5%, CST–AST 68.9% and AST–BT 65.3%. The corresponding values for internal residues only are: 92.8, 91.3 and 88.4%, respectively. In spite of the high degree of conservation, there are some striking differences. For example the number of methionine residues are notably different between the fish and the mammalian enzymes, with a total of five, six and two methionines in CST, AST and BT, respectively. The positions of the methionines in anionic salmon trypsin and the high number of methionines ($\geq$5) are, to a high extent, conserved in other fish trypsins, like dogfish (Titani *et al.*, 1975), Antarctic fish (Genicot *et al.*, 1996) and two cod trypsins (Gudmundsdóttir *et al.*, 1993). In addition to methionine 104 and 180 which are conserved for all three sequences, Met242 in the $\alpha$-helical tail

region of the C-terminal is the only one found for both the salmon enzymes. Met242 is also reported in Antarctic fish elastase (Aittaleb *et al.*, 1997), two cod elastases (Gudmundsdóttir *et al.*, 1996), salmon elastase (Berglund *et al.*, 1995*b*) and in mouse elastase (Stevenson *et al.*, 1986). For cod elastase, the authors suggest that Met242 would increase the flexibility of the helix and thus weaken the overall stability of the enzyme. This could be the case for the native anionic salmon trypsins, salmon elastase (entries 2TBS, 1BIT and 1ELT) and anionic salmon trypsin–inhibitor complexes (Helland *et al.*, 1998), where higher flexibility in terms of increasing *B* values and disorder is observed for the last two or three residues of the C-terminal α-helix. For the cationic salmon trypsin reported here, Met242 and a well defined α-helix co-exist (see Fig. 13), which does not support the assumption that Met242 is in part responsible for the decreased thermal stability of the psychrophilic fish enzymes.

The three enzymes also differ in the number of charged residues, resulting in a net charge of +4 and +6 for CST and BT, respectively, and −4 for AST. These differences are confirmed through the isoelectric points of above 10 for the two cationic trypsins and 4.7 for AST (Outzen *et al.*, 1996; Bier & Nord, 1951). Electrostatic effects have been found to be important for the stability of many enzymes (White *et al.*, 1996; Chung *et al.*, 1996), but the charges might also be important for the enzymatic activity. The more negatively charged anionic salmon trypsin could be more efficient in guiding the positively charged Lys or Arg of the substrates into the specificity pocket than the cationic enzymes.

Superposition of the cationic and anionic salmon trypsin (Fig. 11) shows that eight loop regions deviate by

| | N-terminal | | | | | | | N-terminal loop | | | | | | | Nβ1 | | | | | Nβ1-Nβ2 loop | | | | Nβ2 | | | | | | Nβ2-Nβ3 loop | | | | Nβ3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 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 | I | E | E | E | E | I | I | 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 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 |
| CST | I | V | G | G | Y | E | C | R | K | N | S | A | S | Y | Q | A | S | L | Q | S | G | Y | H | F | C | G | G | S | L | I | S | S | T | W | V | V | S | A |
| AST | I | V | G | G | Y | E | C | K | A | Y | S | Q | A | H | Q | V | S | L | N | S | G | Y | H | F | C | G | G | S | L | V | N | E | N | W | V | V | S | A |
| BT-C | I | V | G | G | Y | T | C | G | A | N | T | V | P | Y | Q | V | S | L | N | S | G | Y | H | F | C | G | G | S | L | I | N | S | Q | W | V | V | S | A |

| | Nβ3-Nβ4 loop | | | | | | | | Nβ4 | | | | Ca binding loop | | | | | | | | | | | Nβ5 | | | | | | | | | | - | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | E | I | E | E | E | E | I | E | I | E | I | I | 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 |
| | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 |
| CST | A | H | C | Y | K | S | R | I | Q | V | R | L | G | E | H | N | I | A | V | N | E | G | T | E | Q | F | I | D | S | V | K | V | I | M | H | P | S | Y |
| AST | A | H | C | Y | K | S | R | V | E | V | R | L | G | E | H | N | I | K | V | T | E | G | S | E | Q | F | I | S | S | S | R | V | I | R | H | P | N | Y |
| BT-C | A | H | C | Y | K | S | G | I | Q | V | R | L | G | E | D | N | I | N | V | V | E | G | N | E | Q | F | I | S | A | S | K | S | I | V | H | P | S | Y |

| | Nβ5-Nβ6 loop | | | | | | | | | Nβ6 | | | | | Interdomain loop | | | | | | | | | | | | | | | | | | | | | | | - |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | E | E | E | E | E | E | I | I | I | I | I | E | I | E | E | E | E | E | E | E | E | E | I | E | E | I | E | E | E | E | E | E | E | E | E | E | E |
| | 95 | 96 | 97 | 98 | 99 | 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 | 128 | 129 | 130 | 132 | 133 | 134 |
| CST | N | S | R | N | L | D | N | D | I | M | L | I | K | L | S | K | P | A | S | L | N | S | Y | V | S | T | V | A | L | P | S | S | C | A | S | S | G | T |
| AST | S | S | Y | N | I | D | N | D | I | M | L | I | K | L | S | K | P | A | T | L | N | T | Y | V | Q | P | V | A | L | P | T | S | C | A | P | A | G | T |
| BT-C | N | S | N | T | L | N | N | D | I | M | L | I | K | L | K | S | A | A | S | L | N | S | R | V | A | S | I | S | L | P | T | S | C | A | S | A | G | T |

| | Cβ1 | | | | | | Autolysis loop | | | | | | | | | | | | | | | Cβ2 | | | Cβ2-Cα1 | | | Cα1 helix | | | | | | | - |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | I | E | I | I | I | I | I | E | E | E | E | E | E | E | E | E | E | E | E | I | E | I | I | E | I | E | E | E | E | E | I | E | E | E | E | I |
| | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 |
| CST | R | C | L | V | S | G | W | G | N | L | S | G | S | S | S | N | Y | P | D | T | L | R | C | L | D | L | P | I | L | S | S | S | S | C | N | S | A | Y |
| AST | M | C | T | V | S | G | W | G | N | T | M | S | S | T | A | D | | S | N | K | L | Q | C | L | N | I | P | I | L | S | Y | S | D | C | N | N | S | Y |
| BT-C | Q | C | L | I | S | G | W | G | N | T | K | S | S | G | T | S | Y | P | D | V | L | K | C | L | K | A | P | I | L | S | D | S | S | C | K | S | A | Y |

| | Cα1- Cβ3 loop | | | | | | | Cβ3 | | | | | Loop 1 | | | | | | Specificity pocket | | | | | | | Cβ4 | | | | | Cβ4-Cβ5 | | | Cβ5 - | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E | E | E | E | E | E | E | I | E | I | I | I | E | E | E | E | E | E | I | I | I | E | E | I | E | I | I | I | I | E | I | E | E | E | I | I | I | I |
| | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184A | 184B | 185 | 186 | 187 | 188A | 188B | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 209 | 210 | 211 | 212 |
| CST | P | G | Q | I | T | S | N | M | F | C | A | G | F | M | E | G | G | K | D | S | C | Q | G | D | S | G | G | P | V | V | C | N | G | Q | L | Q | G | V |
| AST | P | G | M | I | T | N | A | M | F | C | A | G | Y | L | E | G | G | K | D | S | C | Q | G | D | S | G | G | P | V | V | C | N | G | E | L | Q | G | V |
| BT-C | P | G | Q | I | T | S | N | M | F | C | A | G | Y | L | E | G | G | K | D | S | C | Q | G | D | S | G | G | P | V | V | C | S | G | K | L | Q | G | I |

| | Cβ5 | | | Specifi.pocket | | | | Loop 2 | | | | | | Cβ6 | | | | | Cβ6- Cα2 loop | | | | C terminal helix (Cα2) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | I | E | E | E | E | E | E | E | E | E | E | I | I | I | I | I | E | I | E | E | I | E | E | E | I | E | E | E | E | E | E | E |
| | 213 | 214 | 215 | 216 | 217 | 219 | 220 | 221A | 221B | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 |
| CST | V | S | W | G | Y | G | C | A | Q | R | N | K | P | G | V | Y | T | K | V | C | N | Y | R | S | W | I | S | S | T | M | S | S | N |
| AST | V | S | W | G | Y | G | C | A | E | P | G | N | P | G | V | Y | A | K | V | C | I | F | N | D | W | L | T | S | T | M | A | S | Y |
| BT-C | V | S | W | G | S | G | C | A | Q | K | N | K | P | G | V | Y | T | K | V | C | N | Y | V | S | W | I | K | Q | T | I | A | S | N |

Fig. 10. Amino-acid sequence of cationic salmon trypsin (CST), anionic salmon trypsin (AST) and bovine trypsin (BT). Nomenclature of the various structural regions are from Heimstad (1996). Residues are classified internal (I) when they have a water-accessible surface less than 10 $Å^2$ in the trypsin structures 1BIT, 1BRA, 2TBS, 2TRM and 3PTB, others are external (E). Residues differing from CST are shaded.

Table 8. *R.m.s. difference between the CST structures, AST and BT calculated for all protein atoms and for backbone only*

| | | CST | | | | |
|---|---|---|---|---|---|---|
| | | Mol *A* | Mol *B* | Mol *C* | Mol *D* | BT |
| AST | All atoms | 1.836 | 1.717 | 1.832 | 1.777 | 1.572 |
| | Main-chain | 1.775 | 1.697 | 1.789 | 1.713 | 1.710 |
| AST | All atoms ÷ (145–153)† | 1.238 | 0.969 | 1.198 | 1.122 | 0.696 |
| | Main-chain ÷ (145–153)† | 0.963 | 0.670 | 0.957 | 0.774 | 0.553 |
| BT | All atoms | 1.131 | 0.783 | 1.074 | 1.008 | — |
| | Main-chain | 0.895 | 0.588 | 0.908 | 0.794 | — |

† R.m.s. calculations excluding residues 145–153 of the autolysis loop.

≥1.0 Å, while corresponding comparison between CST and BT show deviations in only six loops (Figs. 12*a* and 12*b*). Of the four non-crystallographically related CST molecules, Mol *A* and *C* differ from both AST and BT in the N$\beta$1–N$\beta$2 loop, while Mol *B* and *D* fold more similarly. Alignment of the main-chain atoms results in r.m.s. differences of: CST-AST 1.7–1.8 Å, CST-BT 0.6–0.9 Å and AST-BT 1.7 Å (Table 8). A large portion of the high deviations, especially for AST compared to the cationic variants, is caused by the autolysis loop, which is known to be very different in AST and BT (Smalås *et al.*, 1994). The two loop orientations can be viewed in Fig. 11, showing a similar conformation for CST and BT. Leaving this loop out of the calculations gives more similar r.m.s. deviations (Table 8).

### 5.3. *Internal residues*

Based on water-accessible surfaces, 69 of the 223 residues are considered internal and thus labelled I in Fig. 10. 60 of the internal residues are conserved in all three sequences. Of the nine non-conserved internal residues, six are different between CST and AST, five are different between CST and BT, and eight internal residues are replaced in AST compared to BT. R.m.s. differences between internal residues among the three trypsin structures are in the range 0.42–0.57 and 0.26–0.29 Å for all atoms and main-chain atoms, respectively. Two hydrogen bonds involving internal residues (Thr229 $O^{\gamma 1}\cdots$Asp102 O and Tyr234 $O^{\eta}\cdots$Asn101 $N^{\delta 2}$) are conserved in both cationic trypsins, but absent in AST due to non-polar residues at position 229 and 234. These interactions could be important as shown for two barnase mutants, Tyr→Phe and Ser→Ala, where the hydrogen bonds involving Tyr $O^{\eta}$ and Ser $O^{\gamma}$ were found to stabilize the structure by 1.4 and 1.9 kcal $mol^{-1}$, respectively (Chen *et al.*, 1993). The two additional hydrogen bonds in CST and BT could impose more rigidity as compared to AST and in part explain the observed difference in thermostability and/or catalytical efficiency.

Examination of the other differences in internal residues (residues 31, 63, 121, 138, 160, 212 and 238), shows that most of these are valines and isoleucines in CST, predominantly valines in AST and a majority of isoleucines in BT. Both salmon trypsins have three Val, and BT has three Ile at 121, 138 and 212, giving more free space in the core of the fish trypsins. In total, the

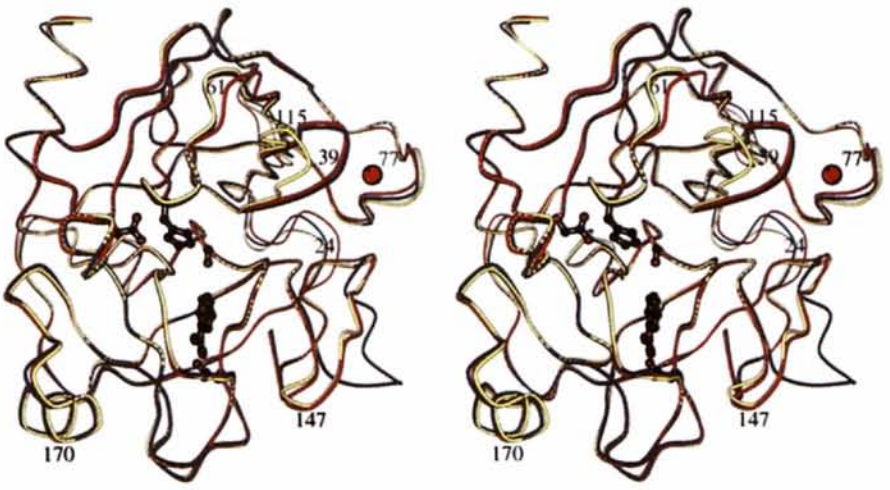



Fig. 11. Superposition of the $C_\alpha$ traces of cationic salmon trypsin Mol *A* in yellow, anionic salmon trypsin in blue and bovine trypsin in red. The largest differences is in the autolysis loop (residues 140–155) in the lower right of the molecules.

interior of BT seems to be more compact and contains two more hydrogen bonds than AST. CST seems to fall somewhere in-between, with the two internal hydrogen bonds as in BT, but with a less optimized core packing similar to AST.

### 5.4. *Hydrophobicity*

The hydrophobicity indices for internal residues of the three enzymes, using the *Protean* program (DNASTAR Inc), the scale of Kyte & Doolittle (1982) and a window size of five residues, were found to be 24.7, 24.6 and 27.7 for CST, AST and BT, respectively. Corresponding hydrophobicity indices for external residues (labelled *E* in Fig. 10), show that the surface of the fish trypsins possesses a more hydrophilic nature compared to bovine trypsin, with indices of −64.3, −53.7 and −42.1 for CST, AST and BT, respectively. The more hydrophilic surface of the fish trypsins is also consistent with observations for Antarctic fish trypsin (Genicot *et al.*, 1996), cold-adapted subtilisin (Davail *et al.*, 1994) and cold-adapted α-amylase (Feller *et al.*, 1992). For the complete fish trypsin sequences, a significantly lower hydrophobicity is observed, as also reported for cod elastase compared with porcine elastase (Gilberg & Øverbø, 1990). The overall decrease in hydrophobicity for the fish enzymes as compared to their mammalian counterparts could contribute to the thermosensitivity and flexibility as proposed for the cold-adapted α-amylase (Feller *et al.*, 1994). However, since the non-psychrophilic cationic salmon trypsin has an even more hydrophilic amino-acid sequence than its psychrophilic anionic counterpart, there seems to be no direct interconnection between low hydrophobicity and psychrophilicity.

### 5.5. *Ion pairs*

The stabilizing effect from ion-pair interactions is still under debate. In proteins from extreme thermopiles several networks of ion-pair interactions have been found, but their exact role is not fully clear (See review Goldman, 1995). By defining an ion-pair by a distance criterion of <4 Å, as carried out by Barlow & Thornton (1983), there are five or six salt bridges in CST and five in both AST and BT. Based on the amino-acid sequence and expected structural vicinity, additional potential salt bridges could be present in both CST ($Arg^{135}$–$Asp^{159}$) and AST (C-terminal $COO^- - Arg^{87}$ and $Lys^{107}$). These are not considered observed due to poor electron density for some residues in the regions. Two ion pairs are highly conserved in all trypsins and are important for the enzymatic activity, namely $N-Ile^{16}$–$Asp^{194}$ and $His^{57}$–$Asp^{102}$. For the CST molecules, Mol *A* is the only one including the $Glu^{21}$–$His^{71}$ interaction (see Fig. 9) as in the AST structure, while another interaction, $Glu^{21}$–$Arg^{156}$, is present for Mol *B*, *C* and *D*. Differences in the calcium-binding loop explain why the $Arg^{66}$–$Glu^{70}$ ion-pair interaction is unique to Mol *A* and *C*. AST has two salt bridges not observed in any of the other molecules namely $Glu^{64}$–$Arg^{66}$ and $Asp^{150}$–$Lys^{154}$, and in BT there is one such unique interaction, namely $Asp^{165}$–$Lys^{169}$. In contrast to AST, the two cationic structures include well defined ion pairs from the C-terminal carboxyl group to

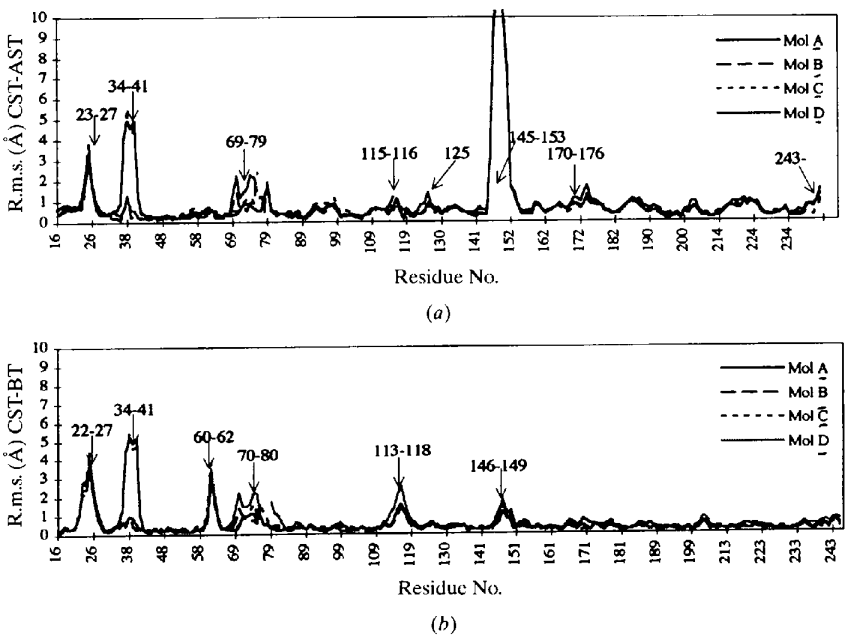



Fig. 12. Residue-averaged r.m.s. deviations for main-chain atoms of (*a*) anionic salmon trypsin (AST) and (*b*) bovine trypsin (BT) compared to each of the four NCS-related cationic salmon trypsin molecules.

Lys[87] and Lys[107] (Fig. 13). By also including the potential C-terminal ion-pairs of the cold-adapted AST structure, there are actually more salt bridges in anionic salmon trypsin than in the non-psychrophilic cationic ones. This is not consistent with observations for psychrophilic subtilisin (Davail *et al.*, 1994) and Antarctic fish trypsin (Genicot *et al.*, 1996) where fewer salt bridges are found in the cold-adapted enzymes. However, no definite conclusions should yet be drawn at this point, since well defined ion pairs in crystal structures could be highly affected by crystal packing as demonstrated for the independently refined CST molecules.

### 5.6. *The active-site region*

The active site and the enzyme mechanism of trypsin are well established from a large number of experiments (*e.g.* Singer *et al.*, 1993; Sprang *et al.*, 1987). In all the crystal structures examined here, the small competitive benzamidine inhibitor is bound in the active site with well defined electron density and low $B$ values. A comparison of the hydrogen-bonding pattern in the active-site region, as performed by Smalås *et al.*, (1994), shows no significant difference in hydrogen-bonding distances among the four CST molecules, AST, AST-II (1BIT) and BT. Superposition and r.m.s. calculations based on atoms of the catalytic residues (His57, Asp102, Ser195) and the walls of the specificity pocket (189–193 and 214–220), result in deviations in the range 0.16–0.62 Å. Residue 217, a serine in BT and a tyrosine in CST and AST, is the only active-site residue which is not identical in the three trypsins (Fig. 10).

Not only the primary binding site but also loops 1 and 2 (residues 184*B*–188*B*, 221*A*–225), which are the extensions of the substrate binding pocket, are found to be important for substrate specificity (Hedstrom, Farr-Jones *et al.*, 1994; Hedstrom, Perona *et al.*, 1994). The amino-acid sequences of loop 2 are highly different, where CST and BT both include two positively charged residues (Arg/Lys222, Lys224) while AST has one negatively charged residue (Glu221). If the charge in this loop affects the substrate binding, a difference of 3 units, +2 in the cationic and −1 in the anionic enzyme, could account for some of the differences in activity. In addition, Gly223 in AST could add substantially more freedom to the loop and thereby affect the activity.

### 5.7. *The autolysis loop*

The most pronounced structural distinction among the three trypsin enzymes is in the so-called autolysis loop (residues 141–155) where AST folds differently to both CST and BT. It should be pointed out that this loop is not an autolysis loop in either of the two salmon trypsins, but since the term is so well established for trypsins, we continue to use the expression. The loop region which folds differently (residues 145–153), comprises many amino-acid substitutions among the three trypsins, and one deletion of Tyr151 in AST (Fig. 10). Tyr151 and Pro152 (Ser152 in AST) present in both CST and BT, are probably responsible for the different loop conformations. Tyr151 is also suggested to be important for substrate activation and for facilitating the catalytic activity (Oliveria *et al.*, 1993), and it was found phosphorylated in the crystal structure of human trypsin 1 (Gaboriaud *et al.*, 1996). However, since the comparison of kinetic data for CST, AST and BT (Outzen *et al.*, 1996) indicates a better substrate affinity and more efficient catalysis for AST which lacks Tyr151, the role of this residue is still unclear.

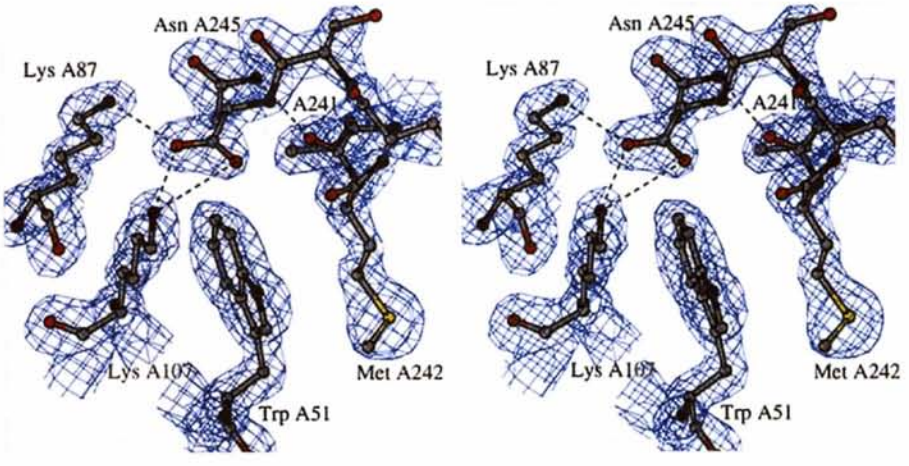



Fig. 13. Electron-density map ($2F_o - F_c$) of the ion-pair interactions between the C-terminal carboxyl group (Asn*A*245), Lys*A*87 and Lys*A*107 in Mol *A* of cationic salmon trypsin. The hydrogen bonding from Asn245 $N^{\delta 2}$ to Thr241 O and residues 242–244 and Trp*A*51 are also shown. Contour level is 1.5$\sigma$.

### 5.8. *The C-terminal α-helix*

The structure reported here is the first salmon trypsin where the complete C-terminal α-helix is well interpreted in the electron density. Fig. 13 shows the interactions and electron-density map for the last residues in the helix. For the native and trypsin–inhibitor complexes of anionic salmon trypsin reported so far (Helland *et al.*, 1998; Smalås *et al.*, 1993; Berglund *et al.*, 1995*a*), high *B* factors and poor electron density are observed for the three terminal residues (243–245) of the helix. For CST, AST and BT the main-chain dihedral angles of the helices are close to predicted values, and the intrahelical hydrogen bonds are similar for all the three trypsin structures. Based on the N-capping residues (235–238) of the C-terminal helix and the helix propensities reported by Muñoz & Serrano (1995) and Fersht & Serrano (1993), the AST structure should actually be more stable compared to the cationic structures. AST contains Asn235 and Asp236, where the negative charge stabilizes the helical macrodipole, while the N-cap of CST and BT (Arg and Val) includes destabilizing residues. For the C-capping residues of the helix the opposite is found. Tyr245 in AST is known to destabilize, whereas Asn245 in CST and BT could stabilize the C-terminal. This is also consistent with observations in the crystal structures, were the side chain of Asn245 stabilizes the helix by hydrogen-bonding to the main-chain O atom of residue 241 as shown in Fig. 13. Tyr245 of AST is unable to form this interaction, and thus imposes disorder. In addition, the Ser243 $O^{\gamma}\cdots$Ser239 O hydrogen bond in CST is not present in AST due to substitution to Ala at position 243 of the latter. These interactions might partly explain the lack of salt bridges between the C-terminal carboxyl group and the basic side chains of residues 87 and 107 in AST. The salt bridges are present in both CST (Fig. 13) and BT, and the tightly bound helix could also act as a lock for the double-domain structure of trypsin, probably restricting the relative mobility of the two β-barrels and increasing the stability of the cationic enzymes.

## 6. Concluding remarks

### 6.1. *The four NCS-related molecules*

The model of cationic salmon trypsin, comprising an asymmetric unit of four independent molecules, has been refined to a final *R* value of 17.4% with data to 1.70 Å resolution. In the initial stages of the refinement, NCS restraints were applied, but it became clear that the model could not be satisfactorily adjusted by this protocol. For the refinement without NCS restraints, the ratio between number of observations (68 062) and refined parameters (28 424) is about 2.4 which is considered adequate (Kleywegt, 1996; Kleywegt & Jones, 1995). The overall r.m.s. difference between the four independent molecules range from 0.6 to 1.2 Å, when all atoms are included. Close examination of the structures showed that these high values mainly arise from differences in a few external regions that undoubtedly are folded differently among the four trypsin molecules. Exclusion of two such loops from the calculations resulted in differences in the order of 0.4–0.6 Å for all remaining atoms. Furthermore, a comparison of equivalent main-chain dihedral angles of the NCS-related molecules as suggested by Kleywegt (1996), has been carried out. The largest deviation was between Mol *A* and *D*, where 14.8 and 16.1% of the residues have $|\Delta\varphi|>10^{\circ}$ and $|\Delta\psi|>10^{\circ}$, respectively. This is within the range found for other comparable structures. Kleywegt (1996) have reported that the average $\varphi$ deviation for NCS-related molecules from crystals diffracting beyond 2 Å, is lower than 15°. For CST the maximum $\langle|\Delta\varphi|\rangle$ value is 6.9° (Mol *C*–*D*), indicating that the refinement without NCS restraints has been satisfactory.

For almost all vertebrate trypsin structures reported so far, a structurally bound calcium ion has been located at a specific site. The presence of calcium is also found to stabilize the enzyme against thermal inactivation (see *e.g.* Osnes *et al.*, 1985). For the new crystal structure presented here, only one of the four independent molecules is refined with a structurally bound calcium ion. We do observe distinct differences, both locally at the binding sites and for the overall structures, depending on whether calcium is bound or not. The calcium-binding loop of Mol *A*, for which calcium is present, is heavily involved in intermolecular contacts, while such contacts are hardly present for the calcium-binding loops of the other molecules. Judged from the *B*-factor distribution along the polypeptide chains of the four molecules (Fig. 7), the lack of calcium destabilizes not only the binding region, but also the surrounding loops. The non-calcium-binding molecules seem to be generally more flexible, as reflected in higher mean *B* values, lower correlation coefficients and fewer surrounding water molecules. This is most likely due to the lack of calcium, but effects from different packing environments cannot be ruled out. The low occupancy of calcium could not be increased by adding calcium to the protein solution prior to the crystallization, supporting an assumption of the crystal packing environments being responsible for the poor metal binding sites in Mol *B* and Mol *D*.

The second region with large deviations among the four NCS-related molecules is the Nβ1–Nβ2 loop (Fig. 5). The loop possesses a conformation similar to AST and BT for Mol *B* and *D*, while a totally different conformation is observed for Mol *A* and *C*. In the complex between trypsin and BPTI, residues of this loop constitute some of the secondary binding sites (Marquart *et al.*, 1983). Since CST also is strongly inhibited by BPTI (Outzen *et al.*, 1996), corresponding

interactions are probably present in this complex too, requiring loop conformations equivalent to those of AST and BT. Mol *A* and *C* thus probably do not represent 'the low-energy' conformation of CST, but demonstrates that the trypsin molecule in fact is relatively adaptable. Adaptation to the environment to form adequate interactions, is also displayed in many crystal forms of T4 lysozyme (Zhang, Wozniak *et al.*, 1995).

### 6.2. *Comparison to anionic salmon and bovine trypsin*

The cationic trypsin secreted by the psychrophilic Atlantic salmon, shares most of its biochemical properties with mammalian trypsins, and is therefore not characterized as a cold-adapted enzyme. The structure of cationic salmon trypsin enables us to point out features that might be common to fish trypsin, but not necessarily connected to psychrophilicity, and *vice versa*. Common features of both cationic and anionic salmon trypsin is the high methionine content (five in CST, six in AST) compared to BT (two). Mutation studies on T4 lysozyme have revealed a close relation between an increased number of methionines and reduced stability (Baldwin *et al.*, 1996; Gassner *et al.*, 1996). As high methionine content is also observed for many other enzymes from cold-adapted fish species, *e.g.* salmon and cod elastase (Berglund *et al.*, 1995*b*; Gudmundsdóttir *et al.*, 1996) and cod chymotrypsin (Gudmundsdóttir *et al.*, 1994), it might play a functional role in the fish enzymes, but the present study on CST shows that there is no direct link between thermal instability and high number of methionines.

Other similarities between the salmon enzymes are the reduced size of three internal residues and reduced hydrophobicity of the core, compared to BT. Common to the two fish trypsins is also a more hydrophilic surface and higher overall hydrophilicity, which are also observed for cold-adapted subtilisin (Davail *et al.*, 1994), $\alpha$-amylase (Feller *et al.*, 1992), Antarctic fish trypsin (Genicot *et al.*, 1996) and cod elastase (Gilberg *et al.*, 1990). The effect of increased hydrophilicity is proposed to be improved solvent interactions and reduced compactness of the molecule (Davail *et al.*, 1994). This could explain the destabilization of the psychrophilic AST, but does not seem to destabilize the even more hydrophilic CST enzyme. One should also bear in mind that electrostatic interactions are formed exothermically and are therefore stabilized by decreased temperature, whereas hydrophobic interactions are formed endothermically and are destabilized at low temperature (Davail *et al.*, 1994). The consequence of these thermodynamic considerations, is that there should be fewer salt bridges, hydrogen bonds and weakly polar (aromatic) interactions in the cold-adapted enzymes to preserve protein flexibility at low temperature, as in AST. The C-terminal salt bridge, the two internal hydrogen bonds and the intrahelical hydrogen bonds that are not present in AST, could be factors responsible for the increased thermostability, reduced flexibility and lower catalytic efficiency of CST and BT, compared to AST. The balance between endo- and exothermically formed interactions and not necessarily the number of bonds, is probably also critical for cold-adapted enzymes.

From the work presented here, there seems to be more similarities between CST and BT than between CST and AST. CST and BT exhibit high sequence similarity (73.5%) and similar overall three-dimensional structures. The autolysis loop is folded similarly, including Tyr151 and Pro152, where the proline seems to force the entire loop into a different conformation than in AST. The presence of Tyr151 and Pro152 in CST, is not reported in any of the other fish trypsin sequences (Titani *et al.*, 1975; Gudmundsdóttir *et al.*, 1993; Male *et al.*, 1995; Genicot *et al.*, 1996). Lack of Tyr151 gives more freedom to Gln192 and thereby increased accessibility to the binding pocket as in AST. Furthermore, the interior of the two cationic trypsins, CST and BT, comprises two additional internal hydrogen bonds, which could be important for the stability of the enzymes (Thr229···Asp102, Tyr234···Asn101). Since both bonds are close to the active site and the catalytic residue Asp102, they could be responsible for higher rigidity in this region and thus make the enzymes less adaptable during catalysis.

In conclusion the internal binding pattern of the non-psychrophilic cationic salmon trypsin seems to be more similar to bovine trypsin than to anionic salmon trypsin. However similar to the cold-adapted proteins cationic salmon trypsin does have more methionines and a more hydrophilic sequence than the mammalian enzymes. The importance of the different interactions and which are accountable for the cold-adapted behaviour is not clear, but it seems to be a sum of many and perhaps small differences.

We would like to thank the organizers of the Swiss–Norwegian Beamline (SNBL) at the European Synchrotron Radiation Facility (ESRF) for kindly supplying us with beamtime, and Professor Edward Hough, Bjørn Riise and Ronny Helland for help during the data collection. We also are indebted to Professor Lars Kr. Hansen for the critical reading and important remarks on the manuscript. This work has been supported by the Norwegian Research Council (AOS, HKS).†

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1a0j). At the request of the authors, the atomic coordinates will remain privileged until 1 December 1998 and the structure factors will remain privileged until 1 December 1999.

## References


Aittaleb, M., Hubner, R., Lamotte-Brasseur, J. & Gerday, Ch. (1997). *Protein Eng.* **10**, 475–477.

Arpigny, J.-L., Feller, G. & Gerday, C. (1993). *Biochim. Biophys. Acta*, **1171**, 331–333.

Ásgeirsson, B., Fox, J. W. & Bjarnason, J. B. (1989). *Eur. J. Biochem.* **180**, 85–94.

Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1996). *J. Mol. Biol.* **259**, 542–559.

Barlow, D. J. & Thornton, J. M. (1983). *J. Mol. Biol.* **168**, 867–885.

Bartunik, H. D., Summers, L. J. & Bartsch, H. H. (1989). *J. Mol. Biol.* **210**, 813–828.

Berglund, G. I., Smalås, A. O., Hordvik, A. & Willassen, N. P. (1995*a*). *Acta Cryst.* D**51**, 725–730.

Berglund, G. I., Smalås, A. O., Hordvik, A. & Willassen, N. P. (1995*b*). *Acta Cryst.* D**51**, 925–937.

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.

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

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

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

Chen, Y. W., Fersht, A. R. & Henrick, K. (1993). *J. Mol. Biol.* **234**, 1158–1170.

Chung, H. O., Tomizawa, K., Kato, T., Wakabayashi, K. & Kato, Y. (1996). *Zool. Sci.* **13**, 915–919.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984). *J. Biol. Chem.* **259**, 14255–14264.

Davail, S., Feller, G., Narinx, E. & Gerday, Ch. (1994). *J. Biol. Chem.* **269**, 17448–17453.

Eijsink, V. G. H., Veltman, O. R., Aukema, W., Vriend, G. & Venema, G. (1995). *Nature Struct. Biol.* **2**, 374–379.

Esnouf, R. M. (1997). *J. Mol. Graphics*, **15**, 133–138.

Feller, G., Lohienne, T., Deroanne, C., Libioulle, C., Van-Beeumen, J. & Gerday, Ch. (1992). *J. Biol. Chem.* **267**, 5217–5221.

Feller, G., Payan, F., Theys, F., Qian, M., Haser, R. & Gerday, Ch. (1994). *Eur. J. Biochem.* **222**, 441–447.

Fersht, A. R. & Serrano, L. (1993). *Curr. Opin. Struct. Biol.* **3**, 75–83.

Fletcher, T. S., Alhadeff, M., Craik, C. S. & Largman, C. (1987). *Biochemistry*, **26**, 3081–3086.

Gaboriaud, C., Serre, L., Guy-Crotte, O., Forest, E. & Fontecilla-Camps, J. C. (1996). *J. Mol. Biol.* **259**, 995–1010.

Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12155–12158.

Genicot, S., Feller, G. & Gerday, Ch. (1988). *Comp. Biochem. Physiol. B*, **90**, 601–609.

Genicot, S., Rentier-Delrue, F., Edwards, D., VanBeeumen, J. & Gerday, Ch. (1996). *Biochim. Biophys Acta*, **1298**, 45–57.

Gilberg, A. & Øverbø, K. (1990). *Comp. Biochem. Physiol. B*, **97**, 775–782.

Goldman, A. (1995). *Structure*, **3**, 1277–1279.

Gudmundsdóttir, A., Gudmundsdóttir, E., Óskarsson, S., Bjarnason, J. B., Eakin, A. K. & Craik, C. S. (1993). *Eur. J. Biochem.* **217**, 1091–1097.

Gudmundsdóttir, A., Óskarsson, S., Eakin, A. E., Craik, C. S. & Bjarnason, J. B. (1994). *Biochim. Biophys. Acta*, **1219**, 211–214.

Gudmundsdóttir, E., Spilliaert, R., Yang, Q., Craik, C. S., Bjarnason, J. B. & Gudmundsdóttir, A. (1996). *Comp. Biochem. Physiol. B*, **113**, 795–801.

Hedstrom, L., Farr-Jones, S., Kettner, C. A. & Rutter, W. J. (1994). *Biochemistry*, **33**, 8764–8769.

Hedstrom, L., Perona, J. J. & Rutter, W. J. (1994). *Biochemistry*, **33**, 8757–8763.

Heimstad, E. S. (1996). Doctoral thesis, Institute of Mathematical and Physical Sciences, Department of Chemistry, University of Tromsø, Norway.

Helland, R., Berglund, G. I., Otlewski, J., Apostoluk, W., Andersen, O. A., Willassen, N. P. & Smalås, A. O. (1998). *Acta Cryst.* D**54**. In the press.

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

le Huerou, I., Wicker, C., Guilloteau, P., Toullec, R. & Puigserver, A. (1990). *Eur. J. Biochem.* **193**, 767–773.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.

Kim, H. R., Meyers, S. P., Pyeun, J. H. & Godber, J. S. (1994). *Comp. Biochem. Physiol. B*, **107**, 197–203.

Kleywegt, G. J. (1996). *Acta Cryst.* D**52**, 842–857.

Kleywegt, G. J. & Jones, T. A. (1995). *Structure*, **3**, 535–540.

Kobori, H., Sullivan, C. W. & Shizuya, H. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6691–6695.

Kyte, J. & Doolittle, R. F. (1982). *J. Mol. Biol.* **157**, 105–132.

Laskowski, R. A., 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.* **232**, 677–685.

Mangel, W. F., Singer, P. T., Cyr, D. M., Umland, T. C., Toledo, D. L., Stroud, R. M., Pflugrath, J. W. & Sweet, R. M. (1990). *Biochemistry*, **29**, 8351–8357.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.

Martínes, A., Olsen, R. L. & Serra, J. L. (1988). *Comp. Biochem. Physiol. B*, **91**, 677–684.

Meyer, E., Cole, G., Radhakrishnan, R. & Epp, O. (1988). *Acta Cryst.* B**44**, 26–38.

Monera, O. D., Kay, C. M. & Hodges, R. S. (1994). *Protein Sci.* **3**, 1984–1991.

Muñoz, V. & Serrano, L. (1995). *J. Mol. Biol.* **245**, 275–296.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. & Vernoslova, E. (1995). *Acta Cryst.* A**51**, 445–449.

Oliveira, M. G. A., Rogana, E., Rosa, J. C., Reinhold, B. B., Andrade, M. H., Greene, L. J. & Mares-Guia, M. (1993). *J. Biol. Chem.* **268**, 26893–26903.

Osnes, K. Kr. & Mohr, V. (1985). *Comp. Biochem. Physiol. B*, **82**, 607–619.

Otwinowski, Z. (1993). *DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography*. Yale University, New Haven, Connecticut, USA.

Outzen, H., Berglund, G. I., Smalås, A. O. & Willassen, N. P. (1996). *Comp. Biochem. Physiol. B*, **115**, 33–45.

Ramachandran, G. N., Ramakrishan, G. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Rentier-Delrue, F., Mande, S. C., Monyens, S., Terpstra, P., Manifroid, V., Goraj, K., Lion, M., Hol, H. G. W. & Martial, J. (1993). *J. Mol. Biol.* **229**, 85–93.

Serrano, L., Day, A. G. & Fersht, A. (1993). *J. Mol. Biol.* **233** 305–312.

Simpson, B. K. & Hard, N. F. (1984). *Comp. Biochem. Physiol. B*, **79**, 613–622.

Singer, P. T., Smalås, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *Science*, **259**, 669–673.

Smalås, A. O., Heimstad, E. S., Hordvik, A., Willassen, N. P. & Male, R. (1994). *Proteins*, **20**, 149–166.

Smalås, A. O. & Hordvik, A. (1993). *Acta Cryst.* D**49**, 318–330.

Sprang, S., Standing, T., Fletterick, R. J., Stroud, R. M., Finer-Moore, J., Xuong, N. H., Hamlin, R., Rutter, W. J. & Craik, C. S. (1987). *Science*, **237**, 905–908.

Stevenson, B. J., Hagenbuchle, O. & Wellauer, P. K. (1986). *Nucleic Acids Res.* **14**, 8307–8330.

Titani, K., Ericsson, L. H., Neurath, H. & Walsh, K. A. (1975). *Biochemistry*, **14**, 1358–1366.

Torrisen, K. R (1987). *Aquaculture*, **62**, 1–10.

Torrisen, K. R (1991). *Aquaculture*, **93**, 299–312.

Uchida, N., Anzai, H. & Nishide, E. (1986). *Bull. Jpn Soc. Sci. Fish*, **52**, 731–735.

Vekovski, V., Schatter, D. & Zuber, H. (1990). *Biol. Chem. Hoppe-Seyler*, **371**, 103–110.

Wang, K., Gan, L., Lee, I. & Hood, L. (1995). *Biochem. J.* **307**, 471–479

Watanabe, K., Masuda, T., Ohashi, H., Mihara, H. & Suzuki, Y. (1994). *Eur. J. Biochem.* **226**, 277–283.

White, P. J., Squirrell, D. J., Arnaud, P., Lowe, C. R. & Murray, J. A. (1996). *Biochem. J.* **319**, 343–350.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Zhang, X.-J., Baase, W. A., Shoichet, B. K., Wilson, K. P. & Mattews, B. W. (1995). *Protein Eng.* **8**, 1017–1022.

Zhang, X.-J., Wozniak, J. A. & Matthews, B. W. (1995). *J. Mol. Biol.* **250**, 527–552.

Zou, J.-Y. & Mowbray, S. L. (1994). *Acta Cryst.* D**50**, 237–249.