Acta Cryst. (1995). D**51**, 925–937

Structure of Native Pancreatic Elastase from North Atlantic Salmon at 1.61 Å Resolution

BY GUNNAR I. BERGLUND

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

NILS P. WILLASSEN

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

AND ASBJØRN HORDVIK AND ARNE O. SMALÅS*

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

(*Received* 30 *December* 1994; *accepted* 5 *April* 1995)

Abstract

The crystal structure of native salmon pancreatic elastase (SPE) has been solved by molecular-replacement methods, and refined by conventional conjugate-gradient methods and simulated-annealing techniques. The final R value is 17.2% for 21 389 reflections between 8.0 and 1.61 Å, and the corresponding free R value is 23.9%. The overall tertiary structure of SPE is remarkably similar to that of porcine pancreatic elastase I (PPE), to which it shows about 67% sequence identity. The primary structure of SPE is determined from the electron-density maps, and only about 15 side chains are somewhat uncertain. Interesting differences between SPE and PPE, are one sequence deletion assigned to position 186, the residue 192 at the entrance of the specificity pocket is substituted from a Gln in PPE to Asn in SPE, and one of the calcium ligands is different. Furthermore, electron density is missing in SPE for the last three residues of the C-terminal helix. A comparison of the present amino-acid sequence of SPE with other sequences available indicates that SPE belongs to the class I pancreatic elastases.

1. Introduction

Elastases are endopeptidases belonging to the large and diverse family of serine proteinases characterized by a common catalytic mechanism involving three essential residues, namely serine, histidine and aspartate, known as the catalytic triad [for reviews, see Bode & Huber (1986); Kraut (1977)]. Like chymotrypsins and trypsins, pancreatic elastases are synthesized in cells of pancreatic tissue as inactive precursors which on entering the duodenum are activated by tryptic cleavage of a single peptide bond.

Elastases are rather diverse pancreatic endopeptidases with isoelectric points in the alkaline as well as acidic range. Four classes of pancreatic elastases have so far been identified using sequence information and biochemical characteristics. Two classes of cationic elastases are characterized by their ability to hydrolyse elastin. These two elastases, designated pancreatic elastases I and II, respectively, have been isolated from several mammalian species (Largman, 1983; and references therein). Elastases I cleave preferentially synthetic substrates with alanine in P_1 position (Schechter & Berger, 1967). While, elastases II have a preference for cleavage at residues with medium to large hydrophobicity. In this respect, elastases II resemble the chymotrypsins. The third class of pancreatic elastases, named elastases III (proteinase E), has specificity for alanine in P_1. This class comprises anionic elastases which show little or no elastolytic activity (Shen, Fletcher & Largman, 1987; Shirashu *et al.*, 1988; Tani, Ohsumi, Mita & Takiguchi, 1988). A fourth class has recently been proposed based upon identified cDNA from rat (Kang, Wiegand & Müller-Hill, 1992), but so far no biochemical data are available.

Judged by k_{cat}/K_m, elastases from cold-adapted fish species have been reported to hydrolyse specifically synthetic elastase substrates more efficiently than homologous mammalian enzymes (*e.g.*, Gildberg & Øverbø, 1990; Ásgeirsson & Bjarnason, 1993). This is probably so as to compensate for the lower body temperature of ectothermic organisms (Ásgeirsson & Bjarnason, 1993). In addition to the higher catalytic efficiency, cold-adapted elastases have also been shown to be less stable at low pH values and high temperatures than their mammalian equivalents. These features are not restricted to elastases, but have also been observed for other fish serine proteinases like trypsins (*e.g.* Ásgeirsson, Fox & Bjarnason, 1989; Raae & Walther, 1989; Berglund *et al.*, in preparation) and chymotrypsins (*e.g.* Raae & Walther,

* To whom correspondence should be addressed.

Table 1. *Data-collection statistics for the native data collected on the FAST system*

d_{min} (Å)	No. of observations	No. of independent observations	Multiplicity	Completeness	% of $I > 3\sigma_I$	R_{sym}	$\langle I/sd \rangle$
4.82	1327	973	1.4	94.7	96.9	0.016	56.24
3.52	3706	1610	2.3	100.0	97.6	0.017	75.55
2.90	4651	1984	2.3	99.2	97.2	0.019	64.99
2.53	5314	2249	2.4	97.7	95.4	0.028	49.85
2.27	5204	2408	2.2	93.5	93.4	0.034	40.56
2.07	3745	2522	1.5	89.3	91.0	0.039	28.44
1.92	3551	2631	1.3	86.4	86.8	0.052	21.31
1.80	3649	2730	1.3	83.8	79.6	0.083	14.18
1.70	3684	2807	1.3	81.4	69.9	0.136	9.17
1.61	2002	1660	1.2	46.1	59.0	0.174	6.08
Totals	36833	21574	1.7	83.6	85.9	0.024	32.89

1989; Ásgeirsson & Bjarnason, 1991; Kristjánsson & Nielsen, 1992).

As part of a larger project with the long-term goal of elucidating features which can explain structure/function relationships for psychrophilic enzymes, we have initiated structure studies on pancreatic elastase from Atlantic salmon to facilitate comparative studies with elastases from warm-blooded species. In general, little structure information of enzymes from marine sources is available. The only protein structure for a fish enzyme, reported so far, is trypsin from Atlantic salmon. This structure was recently solved in our laboratory (Smalås & Hordvik, 1993; Smalås, Heimstad, Hordvik, Willassen & Male, 1994). The three-dimensional structure of PPE and human leukocyte elastase have been determined to atomic resolution [see Bode, Meyer & Powers (1989) for review], as well as the metalloproteinase *Pseudomonas aeruginosa* elastase (Thayer, Flaherty & McKay, 1991).

We report here the first molecular structure of a fish elastase, and a comparison of this structure with that of porcine pancreatic elastase.

2. Experimental

2.1. *Crystallization and X-ray data collection*

Crystallization conditions and preliminary crystal information have been reported (Berglund, Smalås, Hansen & Willassen, 1995).

A single crystal of approximately 0.7 × 0.4 × 0.3 mm was mounted in a thin-walled glass capillary with a plug of mother liquor to prevent the crystal from drying out. Diffraction data were collected at 291 K on an Enraf–Nonius FAST area detector using graphite-monochromatized Cu $K\alpha$ radiation provided by an Enraf–Nonius FR-571 rotating-anode generator operated at 2.8 kW (40 kV, 70 mA). The crystal diffracted to 1.61 Å, and was stable in the X-ray beam during the time of data collection. The detector was positioned at a distance of 65 mm from the crystal at θ angles −35 and −18°. At the first θ angle, two orientations of the crystal were chosen. Data were collected during a 50° scan about the crystallographic c axis, and further during a 65° scan about an arbitrary axis 38° away from the c axis. With the detector in the latter position, data were collected during a 80° scan about the same arbitrary axis as described above. Each frame, resulting from a 0.1° rotation of the ω-circle, was recorded for 45 s. Unit-cell parameters have been refined to $a = b = 68.00\,(3)$ and $c = 83.98\,(5)$ Å, and the space group was identified as $P4_12_12$ or $P4_32_12$ based upon systematic absences. Data collection and initial data processing were performed by the FAST version of the *MADNES* program package (Messerschmidt & Pflugrath, 1987). Profiles and integrated intensities were determined by profile fitting using the *PROCOR* program (Kabsch, 1988). The data were scaled and merged by the programs *ROTAVATA* and *AGROVATA* of the *CCP*4 (Collaborative Computational Project, Number 4, 1994) program suite to give 21 574 unique reflections from 36 833 usable observations. The overall R_{sym} was 2.4% for 12 128 reflections measured more than once [$R_{sym} = (\sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |\langle I_h \rangle|) \times 100\%$, where I_{hi} is the intensity of the ith measurement of the same reflection, $\langle I_h \rangle$ being mean observed intensity for that reflection]. Overall completeness was 90.6 and 83.6% for the resolution ranges 14.0–1.75 and 14.0–1.61 Å, respectively. Structure-factor amplitudes were calculated by *TRUNCATE* of the *CCP*4 package. Selected statistics for assessment of data quality are given in Table 1. A Wilson plot (Wilson, 1949) for the final data gives an overall temperature factor of 13.4 Å^2.

2.2. *Structure solution by molecular replacement*

The structure was solved by molecular replacement (MR) using the *MERLOT* program package, version 2.3 (Fitzgerald, 1988).

2.2.1. *Search model.* The starting model was the complete structure of native porcine pancreatic elastase (Meyer, Cole & Radhakrishnan, 1988) obtained from the Protein Data Bank (Bernstein *et al.*, 1977; entry 3EST). Non-protein atoms were not included.

2.2.2. *Rotation searches.* The cross-rotation function was initially calculated by the fast-rotation function

Table 2. *Representative results from the Crowther & Blow* (1967) *translation searches using data in the resolution range* 8–4 Å.

Relative peak height means peak height relative to the second highest peak in the map, or peak height relative to the highest peak (value below 1.0). The rotation angles $\alpha = 53.0$, $\beta = 64.0$ and $\gamma = 61.0°$ calculated by the Lattman rotation function, were input to the calculations.

Translation vector*		Peak position			Relative peak height	Molecular centre		
From	To	T_a	T_b	T_c		x	y	z
Molecule 2	Molecule 1	0.14	0.60	0.50	1.59	0.07	0.30	—
Molecule 3	Molecule 1	0.86	0.72	0.25	2.25	0.07	0.29	—
Molecule 4	Molecule 1	0.28	0.86	0.75	2.25	0.07	0.29	—
Molecule 5	Molecule 1	0.78	0.22	0.98	1.85	—	—	0.495
Molecule 6	Molecule 1	0.36	0.36	0.48	1.34	—	—	0.495
Molecule 7	Molecule 1	0.64	0.50	0.74	1.75	0.07	—	0.49
Molecule 8	Molecule 1	0.50	0.10	0.24	0.83	—	0.30	0.49

* Molecules are numbered according to their order of appearance in *International Tables for X-ray Crystallography* (1969, Vol. I).

(Crowther, 1972). The calculations were carried out for several different resolution ranges on a coarse grid with 5° samplings in the eulerian angles β and γ, and 1.25° samplings in α, with maximum resolutions between 4.0 and 6.0 Å, and minimum resolutions between 8.0 and 12.0 Å.

An unambiguous and well resolved peak was found for almost all resolution ranges. Data from 8 to 4 Å gave the best result. The peak of 6.6 r.m.s. was located at $\alpha = 53.75$, $\beta = 65.00$, $\gamma = 60.00°$. The Lattman rotation function (Lattman & Love, 1970) was calculated around this peak. In these calculations the samplings were 1° in all three eulerian angles. The peak was located at $\alpha = 53.00$, $\beta = 64.00$, $\gamma = 61.00°$.

2.2.3. *Translation searches.* The Crowther & Blow translation function (1967) was calculated for both possible space groups using data in the resolution range 8–4 Å. After rotating the model according to the Lattman rotation angles given above, the translation function was sampled in fractions of 0.02 in each axial direction. As would be expected the same and well resolved peaks occurred for both enantiomorphic space groups. Solving for the molecular translation vectors (x,y,z) a consistent set of vectors was found for space group $P4_12_12$, *cf.* Table 2, with the centre of the molecule located at $x = 0.070$, $y = 0.295$, $z = 0.490$. The resulting crystal packing showed only two unique main-chain–main-chain contacts shorter than 2.8 Å to neighbouring molecules. Due to the fact that the amino-acid sequence of the salmon elastase was not available, intermolecular contacts involving side chains were not considered. The rotation and translation solutions found in $P4_32_12$ gave a crystal packing with overlapping molecules. $P4_12_12$ was therefore assumed to be the correct space group.

The initial R factor of the rotated and translated model in the cell of salmon elastase was 48.3% (8–2.3 Å) considering space group $P4_12_12$, decreasing to 44.5% after rigid-body refinement by *X-PLOR* (Brünger, 1992*a*).

2.3. *Model building and refinement*

Refinement was carried out by the simulated-annealing method as well as by the restrained least-squares procedure using the programs *X-PLOR* (Brünger, 1992*a*) and *PROLSQ* (Hendrickson, 1985). The model was systematically improved through iterative cycles of crystallographic refinement and manual rebuilding, using step-wise increased resolution of data, *cf.* Table 3. The model fitting was carried out as judged by inspection of electron-density maps calculated with both $2F_o - F_c$ and $F_o - F_c$ coefficients. Omit maps were used throughout for uncertain regions of the structure. Such areas were checked carefully and built into omit density whenever possible after being omitted from refinement. Refined temperature factors were also used regularly together with electron-density information to locate erroneously placed atoms. In particular, the unrestrained B-factor option in *PROLSQ* was useful to locate atoms with high and fluctuating B values. The program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993) was used throughout to provide overall assessment of the stereochemistry of the in-progress models and to highlight regions that should be further investigated on the graphical terminal.

Before inspection of electron-density maps and manual intervention, the model was subjected to one round of initial refinement by both restrained least-squares and simulated annealing (SA) methods (initial round, Table 3). For the SA, the model was initially heated to 3000 K and slow-cooled to 300 K by reducing the temperature in steps of 25 K after every 50 steps of molecular-dynamics simulation, each representing 0.5 fs (Brünger & Krukowski, 1990). All further refinement (rounds 2–9, Table 3) were carried out by the conventional least-squares procedure by alternating *X-PLOR* and *PROLSQ* refinements, and by step-wise increasing the resolution of the data to the maximum 1.61 Å. The refinement was completed by *X-PLOR* using force-field parameters derived from the Cambridge Structural Database (Engh

Table 3. *Progress of the combined X-PLOR and PROLSQ refinement*

R.m.s. deviations from ideal values are relative to the PROTIN. DAT dictionary supported by the *CCP*4 package.

Round	Initial	2	3	4	5	6	7	8	9
Resolution range (Å)	8.0–2.3	8.0–2.0	8.0–2.0	8.0–1.9	8.0–1.75	8.0–1.61	8.0–1.61	8.0–1.61	8.0–1.61
No. of atoms refined	1822	1729	1752	1785	1829	1871	1903	1917	1922
No. of reflections (no σ cut-off used)	8687	12769	12769	14671	18270	21389	21389	21389	21389
Water molecules	0	0	55	85	119	154	158	156	156
Ca^{2+} ion	0	0	0	0	1	1	1	1	1
Mean temperature factors (Å^2)									
All atoms	8.3	12.7	13.2	13.8	14.6	15.3	15.7	15.3	15.1
Water molecules	—	—	15.0	19.5	24.4	28.7	29.4	28.2	28.0
R start (%)	45.4	30.5	24.3	22.4	21.5	20.1	17.8	17.6	17.6
R end (%)	25.0	24.6	20.5	18.7	17.6	16.5	16.0	16.1	16.1
R.m.s. deviation in coordinates from previous round (Å)*	0.84	0.53	0.42	0.22	0.25	0.32	0.15	0.14	0.10
Ramachandran plot† residues (%)									
Most favoured regions	81.1	86.2	85.7	86.0	86.5	86.4	87.4	88.8	88.8
Additional allowed regions	18.4	12.8	13.3	12.5	13.0	13.1	11.6	11.2	11.2
Generously allowed regions	0.5	1.0	1.0	1.5	0.5	0.5	1.0	0.0	0.0
R.m.s. deviations (Å)									
Bond distances	0.025	0.016	0.012	0.011	0.010	0.010	0.009	0.009	0.009
Angle distances	0.061	0.053	0.044	0.037	0.032	0.033	0.032	0.032	0.032
Planar distances	0.076	0.054	0.047	0.039	0.037	0.036	0.036	0.035	0.036
No. of distances deviating more than 3σ from ideal	297	97	34	25	8	3	4	5	6
R.m.s. deviation from plane (Å)	0.027	0.027	0.024	0.022	0.021	0.017	0.017	0.017	0.017
R.m.s. deviation from chiral volume (Å^3)	0.156	0.109	0.000	0.097	0.094	0.039	0.039	0.039	0.040

* The r.m.s. deviation is based on least-squares superposition of all protein atoms of the in-progress models.
† Produced with *PROCHECK* (Laskowski *et al.*, 1993).

& Huber, 1991), resulting in a final *R* factor of 17.2% (Table 4).

Two side chains, both located at the surface of the protein, were ultimately modelled as disordered. In the final model, the side chain of Gln23 was refined occupying two discrete conformations with occupancy 0.5, *cf.* Fig. 1. Ser113 was refined with its O_γ atom distributed over two sites with occupancy 0.5. However, the refined temperature factors for O_γ in the two positions indicate a higher occupancy for one of the positions (See §3.2).

Water molecules were carefully added when the *R* factor had dropped below 25% and a considerable part of the side chains had been modelled into their positions.

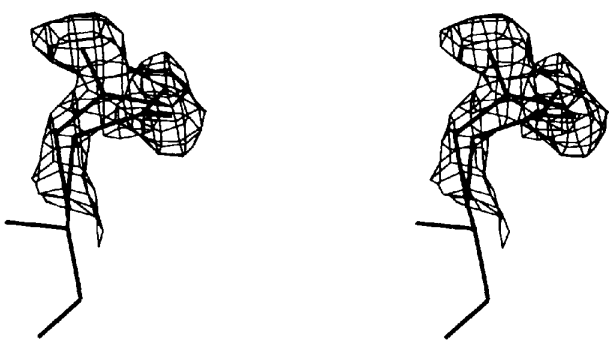

Fig. 1. The final $2F_o - F_c$ electron density at Gln23 contoured at 1σ showing two alternative side-chain conformations superimposed on the density.

The potential sites were only included in the model if they corresponded to positive peaks in both $F_o - F_c$ (3σ) and $2F_o - F_c$ (1σ) difference maps, made reasonable hydrogen bonds to either protein atoms or other water molecules, and did not interact with any part of the structure where there were doubts about the conformation at that stage. After each round, water molecules that failed to reappear in $2F_o - F_c$ (1σ) or had a refined temperature factor greater than 70 Å^2 were excluded from further calculations. The present model includes 156 solvent molecules treated as water O atoms. Two water molecules located close to a crystallographic twofold axis were added to the model with half occupancy after completion of the refinement. All other water molecules are given unit occupancy. Table 3 gives an overview of the progress of refinement, and final *X-PLOR* refinement statistics are summarized in Table 4.*

2.4. *Free R factor calculations*

The free *R* factor measures the agreement between observed and calculated structure factors for a test set of reflections that is not included in the refinement (Brünger, 1992*b*). The final model was first modified

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1ELT). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: SE0169). At the request of the authors, the atomic coordinates and the structure factors will remain privileged until 1 January 1996.

Table 4. *Summary of X-PLOR refinement parameters for the final set of coordinates*

This table represents the results for the final *X-PLOR* refinement carried out after round 9, *cf.* Table 3.

Parameter file used	Parhcsdx.pro
Resolution range (Å)	8.0–1.61
Final R value	0.172
R_{free} value	0.239
Real-space fit correlation*	0.937
No. of atoms	1920
No. of reflections (no σ cut)	21389
Minimum B value (Å^2)	4.7
Maximum B value (Å^2)	69.0
Mean B values (Å^2)	
All atoms	14.6
Protein atoms	13.3
Water molecules	28.7
Ca^{2+} ion	23.1
Model	
Protein residues†	236
Ca^{2+} ion	1
Water molecules	156
Mean positional error (Å) from Luzzati plot	0.15–0.20
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.008
Bond angles (°)	1.5
'Improper' angles (°)	1.2
Dihedral angles (°)	26.3

*Correlation coefficient between model and $2F_o - F_c$ map is averaged over all residues. The calculation includes all atoms per residue.

†The last three residues in the C-terminal helix are missing in the map, and one intermediate amino acid is deleted in SPE relative to the PPE sequence (Shotton & Hartley, 1970).

by deleting all water molecules and resetting of all B values to 15 Å^2. A random selection of approximately 10% of the data was assigned for calculation of the free R factor. As a first step, the model was refined by simulated annealing in *X-PLOR* against the remaining 90% of the data between 8.0 and 1.61 Å to remove some of the memory towards the test set. Water molecules were then included and the model was subjected to restrained positional and individual B-factor refinement in *X-PLOR*. The free R factor converged at 23.9%.

3. Results

3.1. *Quality of model and electron density*

The final model contains 1920 atoms comprising 236 amino acids, 156 water molecules and one Ca^{2+} ion, and the R value is 17.2% for 21 389 reflections (no σ cut-off) in the resolution range 8.0-1.61 Å. The free R factor is 23.9% which compares favourably with values quoted by Brünger (1992*b*) for well refined structures. The good agreement between observed and calculated structure factors for SPE has not been achieved at the expense of model geometry. The r.m.s. deviations from standard geometry values are 0.008 Å for bond distances, 1.5° for bond angles, 1.2° for 'improper' angles, and 26.3 for dihedral angles. Refinement statistics are given in Table 4. The estimated coordinate error according to Luzzati (Luzzati, 1952) is in the range 0.15-0.20 Å. The geometry of the main chain and side chains of SPE have been analysed using the program *PROCHECK* (Laskowski *et al.*, 1993). A Ramachandran plot (Ramachandran & Sasisekharan, 1968) of the main-chain conformational torsion angles, ψ and φ, *cf.* Fig. 2, shows that all residues lie within allowed regions, with 88.8% being within the most favoured regions and 11.2% in additionally allowed regions. Analysis of the side-chain χ_1 and χ_2 dihedral angles shows that 88% of the side chains cluster around the ideal geometry, *i.e.* they are within 2.5σ of the *gauche* and *trans* conformations. The temperature factors also provide a measure of the quality of the structure. The average temperature factors for main-chain and side-chain atoms are 12.1 and 14.7 Å^2, respectively. The average B factor for water molecules is 28.7 Å^2, ranging from 7.7 to 69.0 Å^2. The refined average B factor for all atoms of 14.6 Å^2, *cf.* Table 4, agrees well with the overall B factor of 13.4 Å^2 from the Wilson plot (Wilson, 1949).

The electron-density maps are generally clear, *cf.* Fig. 3. The fit of the model to the present $2F_o - F_c$ electron density has been quantified by calculating real-space coefficients (Jones & Kjeldgaard, 1993), which provide an objective quality measure of the fit to the electron density. The average values for main-chain and side-chain atoms are 0.945 and 0.924, respectively. The uniformly high correlation coefficients, *cf.* Fig.

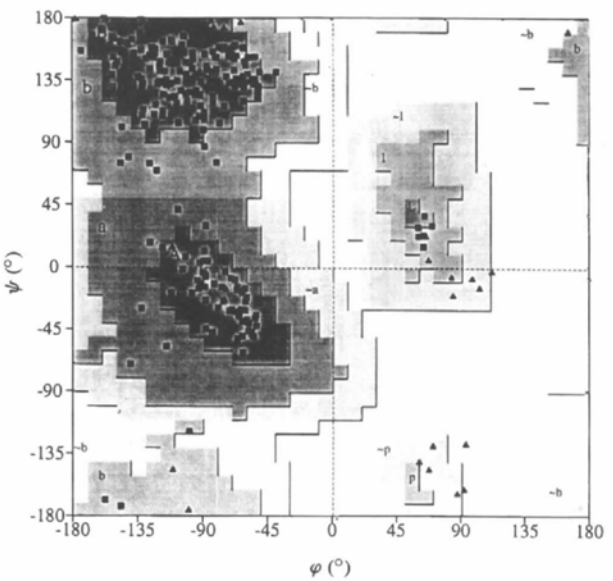

Fig. 2. Ramachandran plot for the main-chain torsion angles (φ,ψ) of the final SPE model. Glycine residues are shown as triangles and all other residues as squares. The plot is produced by the program *PROCHECK* (Laskowski *et al.*, 1993).

4(*a*), reflects the clear and continuous electron density observed throughout the structure. Only for residues 36*B*–37, 204–205, and 217*A*–218, are weaker electron density and lower real-space correlation values observed, *cf.* Fig. 4(*a*).

Residues 217*A* and 218 have the lowest main-chain correlation coefficients, 0.81 and 0.76, respectively. At these positions the density is weak and smeared out, making a precise interpretation difficult. By inspection, 15 out of 236 side-chains (6.4%) have been classified as poorly defined in the present $2F_o - F_c$ map contoured at 1σ. These residues are marked in Fig. 4(*a*) for comparison with the real-space correlation values, see also Table 5. Side chains with well defined densities generally have correlations greater than 0.8, while the values for those with poor defined density are generally lower. Using a 0.8 as cut-off for the real-space fit correlation, 12 out of 236 (5%) residues fall into the category of poorly defined residues. A similar cut-off for the correlation has been suggested as appropriate by Zou & Mowbray (1994). It is not surprising that residues possessing several degrees of rotational freedom like arginines, lysines and glutamines located at the surface of the protein comprise almost half of the residues with poor electron-density fit (Arg20, Gln23, Arg48, Arg107, Gln119, Gln202, Lys223), *cf.* Table 5. The variation in average main-chain and side-chain temperature factors along the polypeptide chain is shown in Fig. 4(*b*) together with corresponding water-accessible surface areas for the individual residues in the crystal lattice. As would be expected, regions showing the highest mobility according to their temperature factors are located in solvent-exposed loop regions, and the overall patterns of the real-space fit and *B*-factor plots are similar, *cf.* Figs. 4(*a*) and 4(*b*).

Table 5. *Residues with poor electron-density fit in final map at* 1σ *contour level*

These residues cannot be interpreted with confidence.

Residue	Comment
Arg20	Good density up to C_δ
Gln23	Weak density for $C_\alpha - C_\beta$ bond, modelled into two conformations, *cf.* Fig. 1
Lys36	Good density up to C_ε
Ser36C	Weak density for main-chain carbonyl group and O_γ due to possible dynamic disorder
Arg48	Good density up to N_ε
Asn74	Weak density for $N_\delta 2$
Thr75	Weak density for C_γ and O_γ
Arg107	Good density up to N_ε
Gln119	Weak density beond C_δ
Asn127	Weak density, but clear branch at C_α
Gln202	Weak density beond C_β
Ser217A	Noisy and rater 'smeared out' density
Ser218	
Lys223	Weak density beyond C_γ
Asn239	Weak density for O_γ

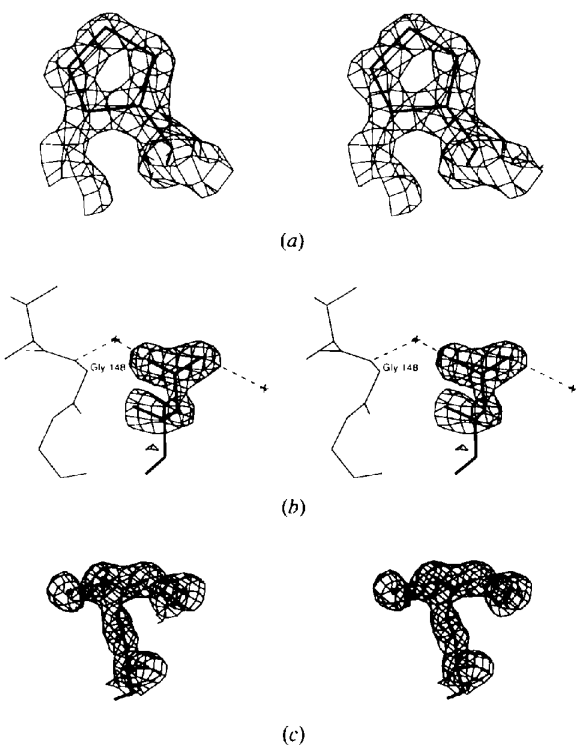

Fig. 3. Selected residues in SPE different from those in PPE. (*a*) $2F_o - F_c$ electron density at residue Pro24 (Arg in PPE) contoured at 1.7σ. (*b*) $2F_o - F_c$ electron density for residue Asn192 (Gln in PPE) at the entrance of the specificity pocket contoured at 1.5σ. Hydrogen bonds are indicated by dashed lines. Crosses represent water molecules. (*c*) $2F_o - F_c$ electron density for Glu77 (Asp in PPE) contoured at 1σ.

3.2. *Deduction of the primary structure of SPE from map interpretation*

The structure of SPE has been determined without knowing the amino-acid sequence. However, the combination of well defined electron density at 1.61 Å resolution and sequence information available from different mammalian pancreatic elastases made structure determination and refinement possible. To our knowledge, no complete elastase sequence from fish is available. Two N-terminal sequences have been reported, namely those from African lungfish (de Häen & Gertler, 1974) and Atlantic cod (Ásgeirsson & Bjarnason, 1993).

A number of possible sequence disagreements between the PPE sequence and the experimental map occurred during the refinement process. The choice of side chains at these positions were guided by selected sequence information from mammalian species, *cf.* Fig. 5. However, the shape of the electron density was always decisive for our choice when the alternative was not accounted for among the different possibilities. Although the crystallographically derived sequence for SPE shows high agreement with the electron density in terms of real-space fit correlation, *cf.* Fig. 4(*a*), there are still some uncertainties.

Some of these uncertainties are side chains located on the surface of the protein, *cf.* Table 5. Others comprise the ambiguity between amidic and acidic forms and the ambiguity between threonine, valine, and serine. Considering amidic or acidic forms, our first choice was based on the most conserved alternative, *e.g.* Asp153 and Gln202, *cf.* Fig. 5. In cases where neither alternatives were conserved, we chose the amide to preserve a basic character of SPE (Berglund *et al.*, in preparation). However, at residue 134 and 188*A*, the hydrogen-bonding pattern favoured the choice of asparagine. Considering a threonine and valine ambiguity, we allowed the environment to be decisive for the choice, *e.g.* Thr84 and Val88, *cf.* Fig. 5. At position 21, a threonine was our first choice due to the fact that this residue is located on the surface of the molecule. In this case, the *B* factor for the O_γ atom refined to a higher value than the C_γ atom (20 *versus* 11 Å^2). Due to the fact that both atoms have similar van der Waals interactions with a symmetry-related molecule, a closely related *B* value for these atoms would be expected. It was, therefore, interpreted as a valine, *cf.* Fig. 5. This example highlights the physical information which can be drawn from refined temperature factors at high resolution. The choice of a valine in this position is also supported by the fact that the N-terminal sequence from cod elastase has a valine in this position (Åsgeirsson & Bjarnason, 1993). At position 113, we find clear evidence for a disordered serine, with two alternative positions for O_γ, instead of threonine which is conserved among elastase I sequences, *cf.* Fig. 5. Both alternative positions for O_γ are acceptable due to hydrogen-bonding distances. A threonine at this site is not reasonable due to resulting short distances between C_γ and neighbouring atoms.

75 amino-acid substitutions have been made in the PPE sequence to establish the present proposed sequence for SPE. Two possible substitutions where found in so-called conserved positions, namely at residues 127 and 148. The conserved residues in these positions are glycine and asparagine, respectively, *cf.* Fig. 5. A branch in the omit density is observed at residue 127. This residue has been interpreted as an asparagine, although the density is of substantial lower quality than the well defined Asx-type residues in the structure. At

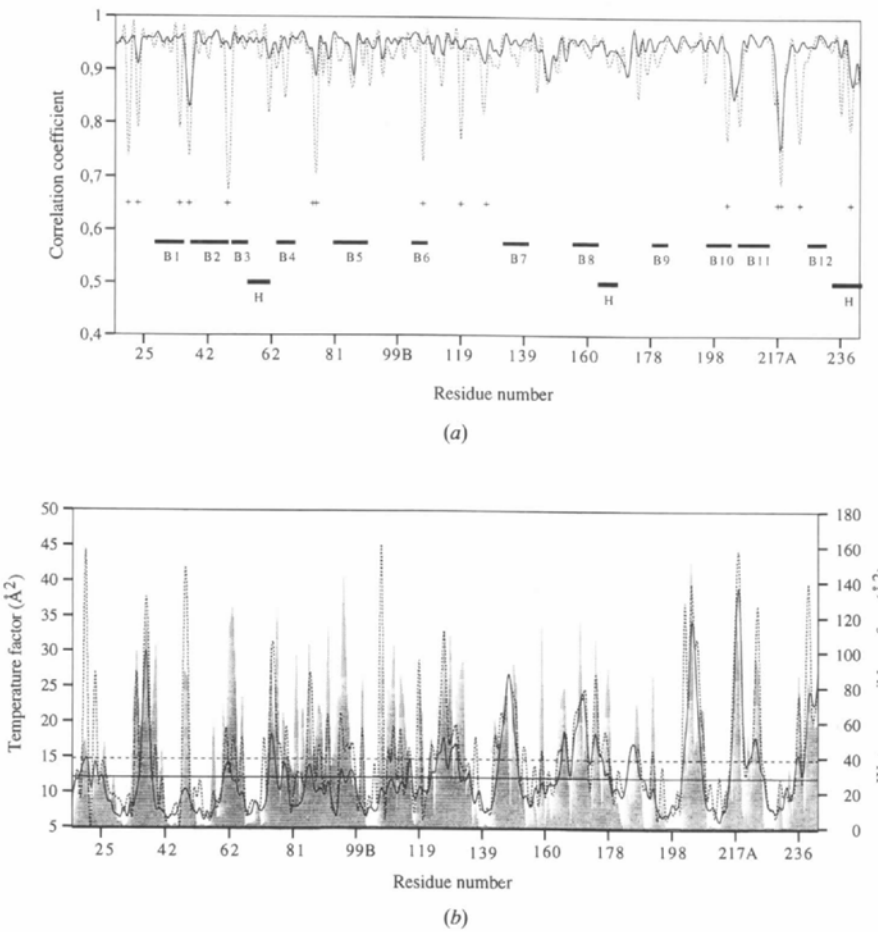

Fig. 4. Variation of correlation coefficients along the polypeptide chain of SPE. (*a*) Real-space correlation coefficients. The main-chain values are shown as a solid line, and the side-chain values as a dashed line. Values of A_o = 0.90 and an overall temperature factor of 14.0 Å^2 were used. The values for *c* and the integration radius were 1.04 and 4 Å, respectively (Jones & Kjeldgaard, 1993). Below the plot, residues with poor electron density as judged by inspection at the graphics display, are indicated as + signs. Secondary structure is indicated by thick bars: *H* for helix, and *B*1–*B*12 for the 12 β-strands. (*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. The overall mean temperature factors for main-chain atoms of 12.1 Å^2, and side-chain atoms of 14.7 Å^2 are represented by a solid and dashed line, respectively.

residue 148, we observe no sign of branching at the C_α position even at lower contour levels, suggesting a glycine in this position. 75 residues in SPE are classified as internal based on their water-accessible surface areas of less than 10 Å^2. Nine of the buried residues are different in SPE and PPE (residues 45, 50, 68, 83, 112, 117, 162, 200 and 238), and most of them represent almost complementary substitutions. In positions 50 and 117 more dramatic substitutions occur. Residue 50 is changed from asparagine in PPE to glycine in SPE, and residue 117 is changed from tyrosine in PPE to alanine in SPE. However, in PPE these side chains project into solvent. Another amino-acid substitution at position 79 from threonine in PPE to lysine in SPE leaves no space for a tyrosine in position 117. As compared to PPE, the three last residues are not visible in the C terminus of SPE. This feature is also seen in salmon trypsin (Smalås *et al.*, 1994).

The present sequence of SPE shows that SPE belongs to class I elastases represented by the porcine pancreatic elastase I structure (Meyer *et al.*, 1988). This is supported by the fact that SPE hydrolyses the elastase I specific substrate *N*-succinyl-(Ala)$_3$-pNA at high rate (Berglund *et al.*, in preparation). The elastase I specificity for small uncharged non-aromatic residues, especially alanine in *P*1 position may be partially due to the presence of Val216 and Thr226, *cf.* Fig. 5. These residues are responsible for the obstructed substrate-

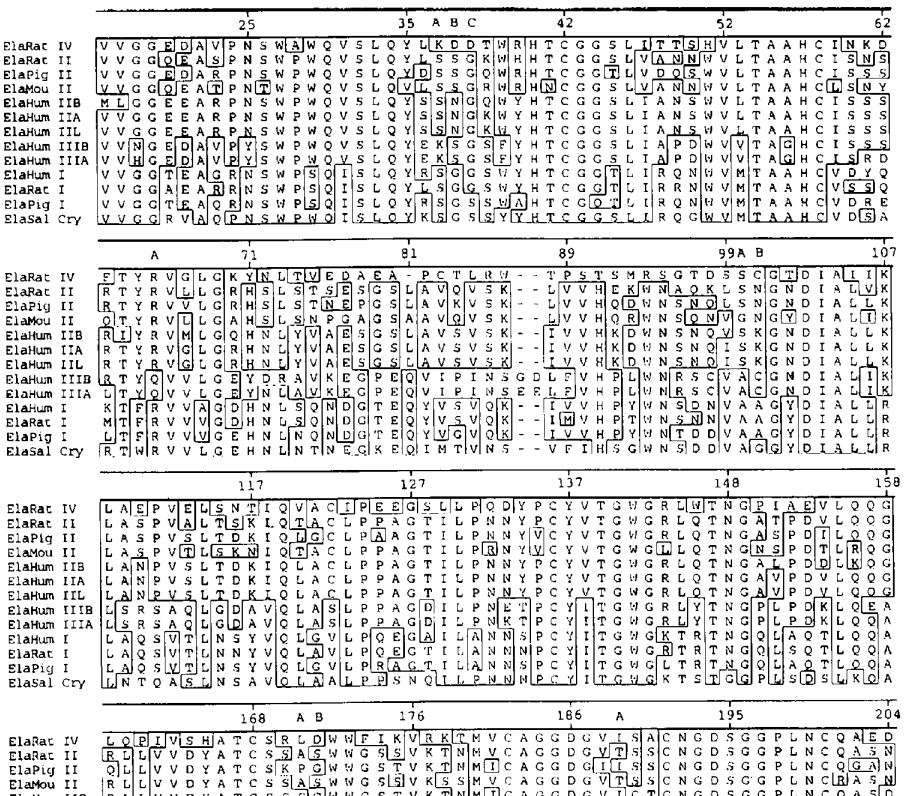

Fig. 5. Amino-acid sequences for pancreatic elastases. The alignment is based on amino-acid data of porcine (ElaPig I, ElaPig II), human (ElaHum I, ElaHum IIA, ElaHum IIB), and rat (ElaRat I, ElaRat II) elastase I and II, and data from human pancreatic endopeptidase E (ElaHum IIIA, ElaHum IIIB) are taken from Tani *et al.* (1988). The data of human elastase II (ElaHum IIL), mouse elastase II (ElaMou II), rat elastase IV (ElaRat IV) are from Shirasu *et al.* (1987), Stevenson, Hagenbüchle & Wellauer (1986) and Kang *et al.* (1992), respectively. ElaSal Cry is the sequence of SPE derived from the present work. Residue types representing the majority at each position is framed. The alignment was carried out by the Clustal V method (Higgins & Sharp, 1989) using the *MegAlign* program (DNASTAR Inc., 1992). The numbering system is that adopted from chymotrypsinogen A (Hartley & Kauffman, 1966).

binding pocket, originally observed for porcine elastase I, that makes binding of large amino-acid side chains sterically impossible (Shotton & Watson, 1970).

The sequence identity between SPE and PPE is about 67%. The identities between SPE and other elastase I sequences available are 69% to rat pancreatic elastase I,, and 68% to human pancreatic elastase I (see Fig. 5 for references). The percentage identities with class II, III and IV elastases are 55–56, 57–58 and 49%, respectively.

Despite SPE being most related to class I elastases, it shares some characteristics with the other classes too. This is particularly evident at some sites where the mammalian class I elastases are clearly distinguished from the other classes. For example, the salmon enzyme was found to contain a proline in position 24 (Fig. 3*a*) and a tryptophan in position 29, both conserved characteristics of class II, III and IV elastases, see also position 171, 192, and 200). Ásgeirsson & Bjarnason (1993) report a tryptophan at position 29 in cod elastase.

3.3. *Coordinate differences between SPE and PPE*

The overall folding of SPE is similar to other pancreatic serine proteinases in general, and to PPE in particular. A detailed description of the elastase molecule is given by Sawyer *et al.* (1978), of trypsin by Stroud, Kay & Dickerson (1971, 1974), and of chymotrypsin by Birktoft & Blow (1972). Superposition of SPE onto PPE after least-squares fitting of 944 common main-chain atoms gives an overall r.m.s. deviation in main-chain atoms of 0.55 Å. The main chains of SPE and PPE are shown superimposed in Fig. 6. Alignment of the two structures based on main-chain atoms for residues building up the core of the proteins, gives an overall r.m.s. difference of 0.30 Å. The core residues are marked in Fig. 7. Despite good overall agreement between the two structures, a more detailed comparison reveals strong variations along the polypeptide chain, *cf.* Fig. 7. The largest differences are found in the N-terminal loop, residues 24–26, and at the extension of the primary binding site, residues 186–188, where the largest displacements are 2.8 and 3.9 Å, respectively. Other differences greater than or close to one r.m.s. are located in the segments 18–19, 36*B*–36*C*, 60–62, 75–78, 204–205 and 221, respectively. Most of these differences occur at positions where differences in sequence between SPE and PPE seem to allow or require different main-chain conformations. For example, at position 24 in the N-terminal loop a proline substitution in SPE forces the chain to take a different and more rigid conformation. Furthermore, a deletion of residue 186 in SPE is responsible for a large deviation in structure, *cf.* Figs. 8(*a*) and 8(*b*). At Gly18 a peptide flip is accounted for by a new hydrogen-bond interaction made possible after a lysine substitution at position 156. N_ζ of Lys156 and the carbonyl O of Gly18 forms the new hydrogen bond. The carbonyl at Gly18 of PPE do not form hydrogen bonds. However, a water molecule displaced by the lysine forms a hydrogen bond to NH of Gly18 in PPE. A replacement of an alanine by a tyrosine at residue 39 in the SPE sequence seems to effect the conformation at the insertion loop, 36*A*–36*C*. The tyrosine points towards Ser36*A* making another conformation necessary to avoid close contacts. Differences are also seen in loop regions which appear to be flexible according to their *B* values, and therefore able to adopt different conformations. Differences in crystal packing may also be responsible for some of the differences.

3.4. *The active-site region*

The backbone atoms of the catalytic triad, Ser195, His57, and Asp102, and the residues forming the central core of the binding site, segments 189–195, 213–216, 226–228, and residue 41; see Bode *et al.*, 1989, can be superimposed with an r.m.s. deviation of only 0.17 Å, which is well below the overall r.m.s. value of 0.55 Å for the entire molecule. Some minor differences in the relative side-chain conformation of His57 and Ser195 are observed. This is most likely due to the presence of a sulfate ion at the catalytic site of PPE (Meyer *et al.*, 1988). The maximum displacement occurs at residue 192 where the main-chain positions differ about 0.36 Å. Residue 192 has been interpreted as an asparagine in SPE, *cf.* Fig. 3(*b*), while a glutamine occupies the equivalent position in PPE. As seen from Fig. 5, an asparagine is conserved at this position in class II, III and IV elastases. Gln192 has been thought to facilitate binding of substrates with small side chains in the P_1 position by movement over the binding cleft after substrate binding (Shotton, White & Watson, 1971). MacDonald *et al.* (1982) suggested that the substitution of asparagine at this position in elastases II may cause the enzyme to be less effective in binding substrates with small side chains. Comparative kinetic studies on SPE and PPE do not support this suggestion. Activity measurements using the elastase specific substrates *N*-Suc-$(Ala)_3$-pNA and MeOSuc-Ala-Ala-Pro-Val-pNA on SPE and PPE, respectively (Berglund *et al.*, in preparation) show that the K_m values for both substrates are lower for SPE than for PPE in spite of such a substitution. However, no finite conclusion on this point can be drawn from such measurements without testing it out on a relevant mutant.

3.5. *Calcium-binding site*

A metal-binding site, presumably occupied by a calcium ion is found in a similar position as reported for PPE (Dimicoli & Bieth, 1977; Meyer *et al.*, 1988) formed by the residues 69–80, *cf.* Fig. 9. However, one of the ligands (residue 77) proved to be clearly one C—C bond longer than the corresponding ligand in PPE, fitting perfectly as Glx, *cf.* Fig. 3(*c*). It has been taken as a Glu, which is the only alternative given

by the different sequences, *cf.* Fig. 5. To maintain a proper coordination to calcium at this site, a somewhat different main-chain conformation is required. The calcium coordination geometry is approximately octahedral with the following calcium-to-ligand distances; Glu70 $O_\varepsilon 2 = 2.4$ Å; Asn72 O = 2.3 Å; Thr75 O = 2.2 Å; Glu77 $O_\varepsilon 2 = 2.6$ Å; Glu80 $O_\varepsilon 2 = 2.5$ Å; OW355 = 2.7 Å. The values agree with those given for PPE, except for a somewhat longer calcium-to-water distance in SPE. The calcium coordination in the present structure is equal to that found in salmon trypsin (Smalås & Hordvik, 1993).

3.6. *Solvent structure*

A total of 156 well ordered solvent molecules were localized in the crystal structure. Due to the close packing in the cell, only 14 of these make contacts to other waters only, and may thus be characterized as second layer water molecules. 12 water molecules form direct hydrogen bonds to two SPE molecules, and 16 more form hydrogen bonds to a first layer water molecule of a neighbouring SPE molecule in the cell.

3.6.2. *Conserved water sites.* The solvent model of SPE have been compared with those of available models of porcine pancreatic elastases in general, and with native PPE (entry 3EST) in particular since both structures have been refined to similar resolution. For comparison, the PPE molecules and their waters were superimposed on the SPE model using only main-chain atoms to define the transformation. Water molecules were treated as common among the structures if their positions were within 1.5 Å of each other. Using this criterion, 72 water sites in SPE are common with corresponding sites in at least one of the PPE structures. The best fit in conserved water sites (67) were between SPE and native PPE, entry 3EST.

3.6.3. *Internal water molecules.* Using 5 $Å^2$ as a threshold value for the accessible area, 26 water molecules are buried in the interior of the SPE molecule compared to 27 in PPE. There are 19 common internal water sites between the two enzymes. Additional three of the internal waters in SPE have common sites in PPE, but can not be characterized as internal in the latter due to structural differences making these waters exposed

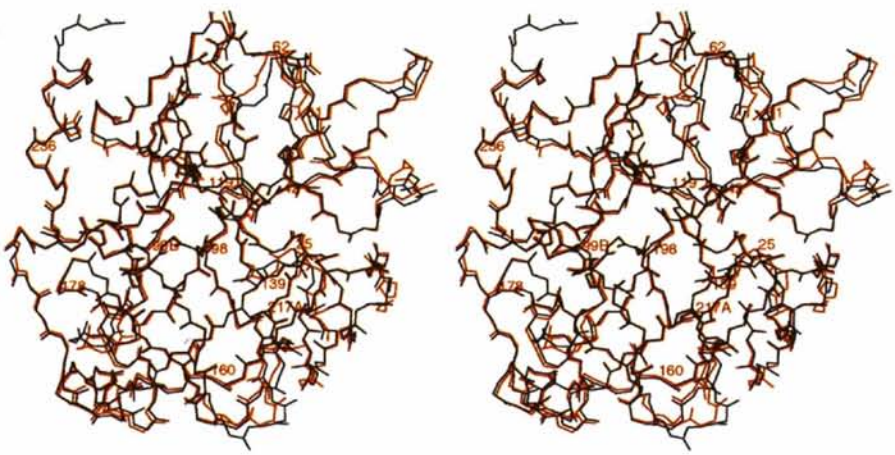

Fig. 6. Main-chain traces of SPE (red) and PPE (dark blue) after superposition of the two structures. Every 20th amino acid is labelled.

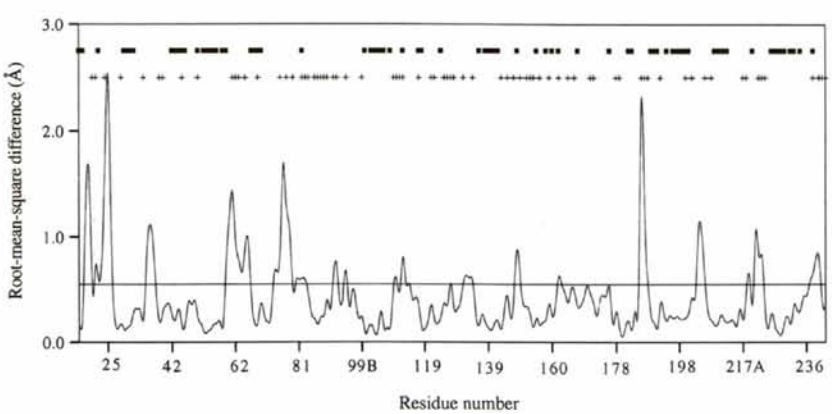

Fig. 7. R.m.s. differences (Å) in equivalent main-chain atomic positions between SPE and PPE, plotted for each residue along the chain. Average deviation of 0.55 Å is shown by a horizontal line. Above the plot, the positions for amino-acid substitutions in SPE relative to PPE are indicated as + signs. Closed legends denote residues corresponding to the core of the protein.

to the surface. Correspondingly, three common water sites may be characterized as internal in PPE but not in SPE. Both SPE and PPE have two internal waters, OW304 and OW329 in SPE, associated with the loop of residue 185–188. Due to the structural differences in this region, these waters have different positions and hydrogen-binding partners in the two enzymes, and can thus not be characterized as equivalent. Two internal water molecules (Sol310 and Sol311) in PPE are not present in SPE due to a different conformation of the side chain of Asn25. However, this is giving rise to a new internal water in SPE, OW382, which is not present in PPE. Finally, the replacement of residue 50 from an asparagine in PPE to a glycine in SPE, *cf.* Fig. 5, gives rise to an internal water in SPE (OW380) at the position occupied by this side chain in PPE.

3.6.4. *Solvent clusters.* About 70 water molecules (43%) in SPE are bound in clusters composed of at least three water molecules, and 35 water molecules form hydrogen bonds to one other water molecule. 51 water molecules are bound to protein atoms only, and eight of these are buried in the interior of the molecule. 18 internal water molecules are involved in clusters running from the surface of the molecule to the interior. These water tunnels are similar to those described in detail for PPE (Meyer *et al.*, 1988), but seem to be even more extended in SPE.

3.7. *Crystal packing*

The SPE molecule makes a total of 216 intermolecular contacts shorter than 4.0 Å to six neighbouring molecules. Only 38 of these are shorter than 3.4 Å and 28 are possible hydrogen bonds. The 14 unique intermolecular hydrogen bonds are listed in Table 6. One is a main-chain–main-chain hydrogen bond, while the others are main-chain–side-chain and side-chain–side-chain interactions.

4. Concluding remarks

The high diffraction quality of the crystals of native SPE is probably the main reason for the good model accuracy obtained after refinement. Temperature factors and real-space fit diagrams reveal that most of the structure is well ordered and therefore show low static and dynamic

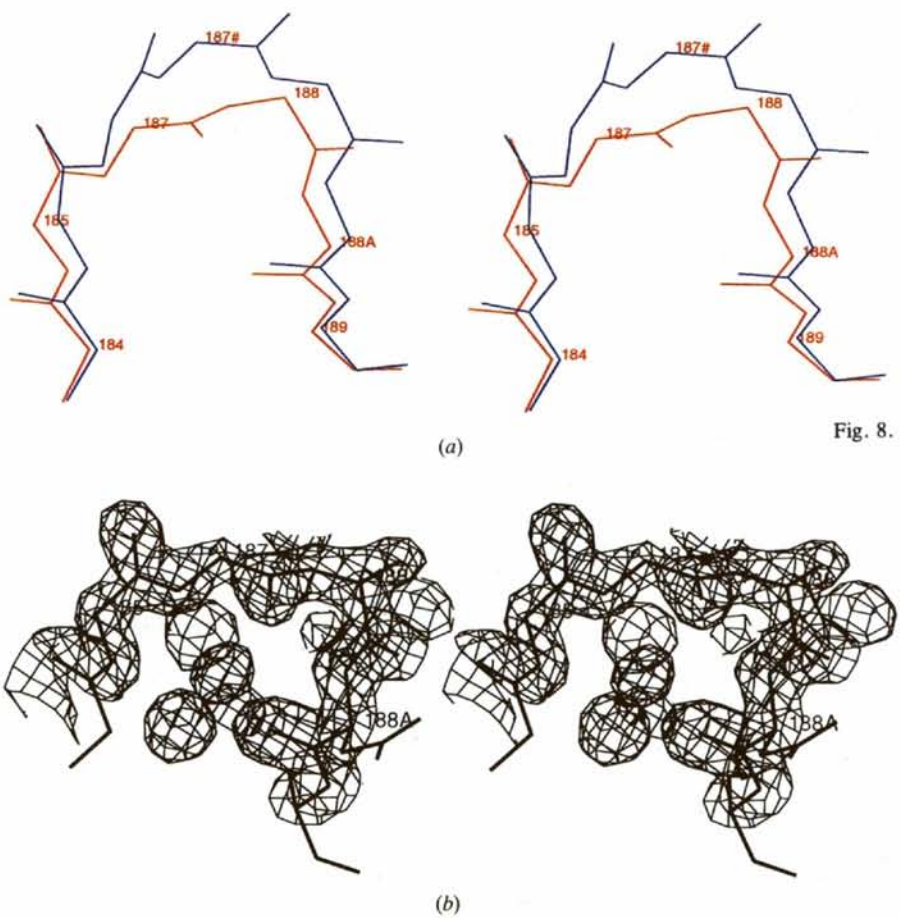

Fig. 8. (*a*) Comparison of peptide conformation around the site of amino-acid deletion (186) in SPE (red) with that of PPE (blue). (*b*) Present $2F_o - F_c$ electron density for residues 185 to 188*A* in SPE contoured at 1σ.

Table 6. *Unique intermolecular hydrogen bonds between symmetry-related molecules in the cell of SPE*

Molecules*	First molecule		Second molecule		Distance (Å)
Mol I–II	$N_{\eta}1$	Arg20	O	Gly148	3.34
	$O_{\varepsilon}1$	Glyn23	$O_{\gamma}1$	Thr41	2.69
	$N_{\varepsilon}2$	Gln23	O	Thr41	2.93
	OH	Tyr137	O	Gly148	3.39
Mol I–III	OH	Tyr35	$N_{\varepsilon}2$	His71	2.80
	$O_{\delta}1$	Asp60	N_{ζ}	Lys79	2.84
	N	Ala62	O	Glu77	2.93
	N_{ε}	Arg63	$O_{\varepsilon}1$	Glu77	2.87
	$N_{\eta}2$	Arg63	$O_{\varepsilon}1$	Glu77	3.36
	$O_{\delta}1$	Asp97	N	Ser116	3.04
Mol I–IV	O–δ1	Asn76	O	Ile90	3.24
	O	Ile82	$N_{\delta}2$	Asn95	2.95
Mol I–V	$O_{\delta}2$	Asp97	$O_{\gamma}1$	Thr113	3.25
Mol I–VI	$N_{\eta}2$	Arg107	O	Ser169	3.34

* Symmetry-related molecules are designated as x, y, z (Mol I), $\frac{1}{2}-y, \frac{1}{2}+x, \frac{1}{4}+z$ (Mol II), $\frac{1}{2}+y, \frac{1}{2}-x, \frac{3}{4}+z$ and translation $-A$, $-C$ (Mol III), $\frac{1}{2}+x, \frac{1}{2}-y, \frac{3}{4}-z$ and translation $-A$, $+B$, $+C$ (Mol IV), $\frac{1}{2}+x, \frac{1}{2}-y, \frac{3}{4}-z$ and translation $+B$, $+C$ (Mol V), $\frac{1}{2}+y, \frac{1}{2}-x, \frac{3}{4}+z$ and translation $-A$, $+B$, $-C$ (Mol VI).

disorder, *cf.* Fig. 4. This is consistent with a crystal density V_m of 1.95 Å^3 Da^{-1}, a low value for typical protein crystals (Matthews, 1968), and corresponding to a solvent content of 36.8% (Berglund *et al.*, 1995).

The convincing values of the various reliability indexes after refinement, *cf.* Table 4, are also strong indications of a mainly correct amino-acid sequence interpretation for SPE. The results from the present crystallographic study show that only about 15 residues are somewhat uncertain and the crystallographically derived sequence for SPE shows about 67% identity with PPE (Meyer *et al.*, 1988). The overall folding of the two structures are similar. The main-chain atoms of SPE and PPE can be superimposed with an overall r.m.s. deviation of 0.55 Å. For comparison, superposition of backbone atoms comprising the catalytic triad and the central core of the binding site (Bode *et al.*, 1989) gives an r.m.s. deviation of 0.17 Å.

Interesting differences between SPE and PPE comprise one amino-acid deletion assigned to position 186, substitution of residue Gln192 in PPE with an asparagine in SPE, and one different calcium-ligand. Furthermore, no density is seen for the last three residues in the C-terminal helix.

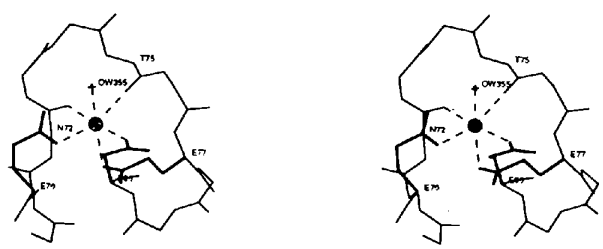

Fig. 9. Residues of the calcium-binding site in SPE, comprising residues 69–80. Side-chains interacting with the calcium (filled sphere) are included, otherwise only main-chain atoms are shown. A water ligand is indicated by a cross. Calcium–ligand interactions are indicated by dashed lines. See §3.5 for details.

This work has been supported by the Norwegian Research Council (GIB, AOS, AH).

References

ÁSGEIRSSON, B. & BJARNASON, J. B. (1991). *Comp. Biochem. Physiol. B*, **99**, 327–335.

ÁSGEIRSSON, B. & BJARNASON, J. B. (1993). *Biochim. Biophys. Acta*, **1164**, 91–100.

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

BERGLUND, G. I., SMALÅS, A. O., HANSEN, L. K. & WILLASSEN, N. P. (1995). *Acta Cryst.* D**51**, 393–394.

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.

BIRKTOFT, J. J. & BLOW, D. M. (1972). *J. Mol. Biol.* **68**, 187–220.

BODE, W. & HUBER, R. (1986). *Molecular and Cellular Basis of Digestion*, edited by P. DESNUELLE, H. SJÖSTRÖM & O. NORÉN, pp. 213–234. Amsterdam: Elsevier.

BODE, W. MEYER, E. & POWERS, J. C. (1989). *Biochemistry*, **28**, 1951–1963.

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

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

BRÜNGER, A. T. & KRUKOWSKI, A. (1990). *Acta Cryst.* A**46**, 585–593.

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

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–185. New York: Gordon & Breach.

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

DIMICOLI, J. L. & BIETH, J. (1977). *Biochemistry*, **16**, 5532–5537.

DNASTAR Inc. (1992). *MegAlign*, DNASTAR Inc., London.

ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* A**47**, 392–400.

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

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

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

DE HÄEN, C. & GERTLER, A. (1974). *Biochemistry*, **13**, 2673–2677.

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

HIGGINS, D. G. & SHARP, P. M. (1989). *Comput. Appl. Biosci.* **5**, 151–153.

JONES, T. A. & KJELDGAARD, M. (1993). *O, The Manual, Version* 5.9.1. Uppsala, Sweden.

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

KANG, J., WIEGAND, U. & MÜLLER-HILL, B. (1992). *Gene*, **110**, 181–187.

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

KRISTJÁNSSON, M. M. & NIELSEN, H. H. (1992). *Comp. Biochem. Physiol. B*, **101**, 247–253.

LARGMAN, C. (1983). *Biochemistry*, **22**, 3763–3770.

LASKOWSKI, R., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.

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

MACDONALD, R. J., SWIFT, G. H., QUINTO, C., SWAIN, W., PICTET, R. L., NIKOVITS, W. & RUTTER, W. J. (1982). *Biochemistry*, **21**, 1453–1463.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Crys.* **20**, 306–315.

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

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

RAAE, A. J. & WALTHER, B. T. (1989). *Comp. Biochem. Physiol. B*, **93**, 317–324.

SAWYER, L., SHOTTON, D. M., CAMPBELL, J. W., WENDELL, P. L., MUIRHEAD, H., WATSON, H. C., DIAMOND, R. & LADNER, R. C. (1978). *J. Mol. Biol.* **118**, 137–208.

SCHECHTER, I. & BERGER, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157.

SHEN, W., FLETCHER, T. S. & LARGMAN, C. (1987). *Biochemistry*, **26**, 3447–3452.

SHIRASU, Y., TAKEMURA, K., YOSHIDA, H., SATO, Y., IIJIMA, H., SHIMADA, Y., MIKAYAMA, T., OZAWA, T., IKEDA, N., ISHIDA, A., TAMAI, Y., MATSUKI, S., TANAKA, J., IKENAGA, H. & OGAWA, M. (1988). *J. Biochem.* **104**, 259–264.

SHIRASU, Y., YOSHIDA, H., MATSUKI, S., TAKEMURA, K., IKEDA, N., SHIMADA, Y., OZAWA, T., MIKAYAMA, T., IIJIMA, H., ISHIDA, A., SATO, Y., TAMAI, Y., TANAKA, J. & IKENAGA, H. (1987). *J. Biochem.* **102**, 1555–1563.

SHOTTON, D. M. & WATSON, H. C. (1970). *Nature (London)*, **225**, 811–816.

SHOTTON, D. M., WHITE, N. J. & WATSON, H. C. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 91–105.

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

SMALÅS, A. O., HEIMSTAD, E. S., HORDVIK, A. WILLASSEN, N. P. & MALE, R. (1994). *Proteins Struct. Funct. Genet.* **20**, 149–166.

STEVENSON, B. J., HAGENBÜCHLE, O. & WELLAUER, P. K. (1986). *Nucleic Acids Res.* **14**, 8307–8330.

STROUD, R. M., KAY, L. M. & DICKERSON, R. E. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 125–140.

STROUD, R. M., KAY, L. M. & DICKERSON, R. E. (1974). *J. Mol. Biol.* **83**, 185–208.

TANI, T., OHSUMI, J., MITA, K. & TAKIGUCHI, Y. (1988). *J. Biol. Chem.* **263**, 1231–1239.

THAYER, M. M., FLAHERTY, K. M. & MCKAY, D. B. (1991). *J. Biol. Chem.* **266**, 2864–2871.

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

ZOU, J. & MOWBRAY, S. L. (1994). *Acta Cryst.* D**50**, 237–249.