Crystal structure of a complex between thermitase from *Thermoactinomyces vulgaris* and the leech inhibitor eglin

Zbigniew Dauter, Christian Betzel, Wolfgang-Ernst Höhne*, Margareta Ingelman and
Keith S. Wilson

European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, FRG and
*Institut für Biochemie, Humboldt-Universität Berlin, Hessische Strasse 3-4, DDR-1040 Berlin, GDR

Thermitase, the thermostable alkaline protease from *Thermoactinomyces vulgaris*, has been crystallised in a 1:1 complex with eglin, the inhibitor from the medical leech. Two large crystals were grown, with cell dimensions of $a = 49.3$ Å, $b = 67.3$ Å, $c = 90.5$ Å and space group $P2_1 2_1 2_1$. The crystals are relatively tightly packed with $V_m = 2.1$ Å^3/Da. Three-dimensional data to 1.9 Å have been recorded from one of these crystals. The orientation and position of the complex in the unit cell have been established using the subtilisin Carlsberg-eglin structure as a model. The structure of the complex is being refined by restrained least-squares. The present crystallographic R factor ($= \Sigma \| F_o - F_c \| / \Sigma | F_o \|$) is 26% at 2.5 Å resolution.

Thermitase; Serine proteinase; Crystallography; Thermostability; Molecular replacement; Ca^{2+} binding

1. INTRODUCTION

Thermitase is the thermostable extracellular alkaline protease secreted by *Thermoactinomyces vulgaris* [1–3]. The sequence of its 279 amino acids has been determined [4], and the calculated molecular mass is 28 369 Da. The sequence confirms the suggestions from a number of physical and biochemical observations that thermitase is a member of the subtilisin family of serine proteases, with an Asp-His-Ser triad involved in the active site.

Complete amino acid or gene sequences have been published for several members of this family from bacilli, including subtilisins from *Bacillus subtilis* strain I 168 [5], *B. amyloliquefaciens* (subtilisin NOVO or BPN') [6], *B. licheniformis* (subtilisin Carlsberg) [7], *B. amylosacchariticus* [8], *B. mesentericus* [9] and *B. subtilis* strain DY [10]. These enzymes are highly homologous to one another. The three-dimensional structures have

Correspondence address: Z. Dauter, European Molecular Biology Laboratory, c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, FRG

been determined for the proteins from *B. amyloliquefaciens* [11] and from *B. licheniformis* [12,13], the latter in complex with the inhibitor eglin from the leech *Hirudo medicinalis*. The tertiary structures show a near identity of the folding of the proteins. Deletions and insertions are very small in number and occur in the loops at the surface of the molecules, as do the majority of the non-conservative amino acid substitutions. None of these proteins contains a cysteine residue.

A subgroup of the subtilisin family does contain a single free cysteine residue, at the same point in the sequence where this is known, and quite close to the active site in the three-dimensional structure. This subgroup includes thermitase and also proteinase K from *Tritirachium album* [14] for both of which the sequences have been published, the protease from *Malbranches pulchella* for which the N-terminal sequence confirms its similarity to proteinase K [15], and the subtilisins from *B. cereus* [16] and *B. thuringiensis* [17]. The three-dimensional structure of native proteinase K has been refined at 1.5 Å resolution [18], and an initial analysis of native thermitase reported at 2.5 Å [19].

The subtilisins possess activity as proteinases, esterases and amidases. They have extended substrate binding regions, with at least 4 or 5 subsites on the N-terminal side of the scissile bond, and a preference for hydrophobic side chains, particularly phenylalanine, at the P_1 site (the first substrate site on the N-terminal side) [20]. Thermitase, as well as proteinase K and subtilisin BPN', has additional specificity sites on the C-terminal side of the scissile bond: there appear to be at least 2 such sites (P_1' and P_2') for subtilisin and thermitase, and three (P_3' also) for proteinase K. All three enzymes have a preference for small residues (Gly or Ala) at P_1', show slower hydrolysis for large residues such as Phe or Leu, and no hydrolysis for Pro at P_1'.

The subtilisin family generally require Ca^{2+} for maximum stability toward thermal or urea induced denaturation and autolysis, and for optimal activity. In thermitase biochemical studies suggest that there are two tightly bound Ca^{2+}, which remain attached during extraction, and one more weakly bound Ca^{2+} with a dissociation constant of about 10^{-4} M at 25°C, pH 3.5–7.5 [3]. The removal of the weakly bound Ca^{2+} slightly reduces the activity of the enzyme, and decreases significantly its stability to denaturation and autolysis. Removal of at least one of the two tightly bound ions does not appear to destabilise the enzyme further. The coordination geometry of these Ca^{2+} and their role in stabilising the three-dimensional structure are of interest.

We here report the crystallisation of thermitase in a complex with the inhibitor eglin from the leech *Hirudo medicinalis*, and the preliminary structure analysis at 2.5 Å resolution. Analysis of the thermitase structure should shed light on its thermostability and the details of the Ca^{2+}-binding sites in comparison to other refined subtilisin structures, especially in comparison with proteinase K, which is missing the Ca^{2+}-binding segment (amino acids 81–87) common to all other subtilisins for which structures are available, but has one additional Ca^{2+}-binding site [18]. The study will also provide accurate data on the geometry of the substrate recognition site and its relation to the activity profile, and give insight into the evolutionary and functional aspects of this enzyme in relation to other members of the family.

2. MATERIALS AND METHODS

Thermitase was prepared as described previously [2]. In brief, a crude preparation obtained by sodium sulphate precipitation from the culture medium was desalted by dialysis and then purified to homogeneity in a single step by preparative flat-bed isoelectric focusing.

Eglin was kindly provided by Dr Schnebli of Ciba-Geigy, Basel. Crystals were grown by the hanging drop vapour diffusion technique [21]. Drops were set up in Linbro plastic tissue culture dishes. With the limited amount of material available, two and one half dishes were set up in all, with a total of 60 drops. A 1:1.3 enzyme/inhibitor solution was made up from the lyophilised samples of thermitase and eglin, the relative concentrations being estimated by weight from the molecular masses. The final composition of the protein solution was 5 mg·ml^{-1} of thermitase, 1.7 mg·ml^{-1} of eglin c, 37.5 mM sodium acetate, pH 5.2, 5 mM $CaCl_2$ and 5% PEG 4000. A 15 μl drop of this solution was placed at the centre of a microslide coverslip. This was inverted and sealed over the reservoirs with silicon vacuum grease. The reservoirs contained 500 μl of an aqueous solution of PEG 4000 in concentration steps from 8 to 22%: half of the reservoirs also contained 0.5 M NaCl. The dishes were left to equilibrate at a constant temperature of 20 ± 1°C.

For X-ray studies a single crystal was mounted in 1.0 mm diameter thin-walled glass capillary tubes and characterised using an Enraf-Nonius precession camera mounted on an Elliot GX18 rotating anode generator operating at 2.8 kW. Three-dimensional data were recorded from the second single crystal using synchrotron radiation from the EMBL beam line X31 at the DORIS storage ring, DESY, Hamburg. Data were recorded at room temperature, on 3-film packs of CEA-25 reflex X-ray film, with flat-plate cassettes on a modified Arndt-Wonacott oscillation camera.

Data were first recorded with radiation of wavelength 1.15 Å and a crystal-to-film distance of 56 mm. The rate of rotation of the crystal was controlled by the current observed in an ionisation chamber monitoring the direct X-ray beam. 90° of data were recorded with an oscillation range of 1.25° per film, with the crystal rotating about the c axis. Typical exposure times were 10–15 min per pack. The same crystal was then used to record 90° of low resolution data, with a wavelength of 1.488 Å, crystal-to-film distance of 90 mm, an oscillation range of 2.25° per film, and a limiting resolution of 2.8 Å. Typical exposure times were 1–2 min. This was collected to overcome the problem of saturated reflections at low angle on the high resolution photographs. No attempt has been made to record those reflections lying in the blind region.

Films were digitised using an Optronics P-1000 photoscan microdensitometer interfaced to a VAX 11/750 computer, with the absorbance range 0–3 *A* represented by the integers 0–255; and a raster size of 50 μm. The digitised images were evaluated using the MOSCO programme package [22] to provide integrated intensities. Later calculations were performed on a micro VAX-II computer using the CCP4 programme suite. Both the CCP4 and MOSCO programmes were provided by the SERC, Daresbury Laboratory. Fitting of the electron density was carried out on an Evans and Sutherland PS330 interactive graphics system, using FRODO [23].

3. RESULTS

In most drops there were no indications of crystals. In each of two drops a single large crystal grew on the surface of the glass. The two crystals were about 0.8 mm in each dimension, with roughly tetrahedral morphology. Both crystals grew over reservoirs containing 10% PEG 4000 without sodium chloride. Time for maximum growth was 6–8 weeks. The total amount of thermitase used to produce the two crystals of the complex and allow a solution of the structure was less than 5 mg.

One of the crystals was used for the characterisation of the space group on the precession camera. Precession photographs allowed the identification of the space group as $P2_12_12_1$ with cell dimensions $a = 49.3$ Å, $b = 67.3$ Å and $c = 90.5$ Å. The volume of the unit cell, with the known molecular masses of 28309 Da for thermitase and 7990 Da for eglin, give a volume per Da, V_m, of 2.1 Å³/Da, assuming one molecule of the complex per asymmetric unit. With the very limited availability of crystals, no chemical proof for the existence of the complex in the crystal was sought: the electron density reported below directly supports this assumption however. The value of V_m, at the lower end of the scale usually observed for protein crystals [24], is typical of other subtilisin proteinases and their complexes.

The second large crystal was used to collect three-dimensional data to 1.8 Å using synchrotron radiation. The data presented considerable difficulty in evaluation as the crystal slipped at several points during data collection: this was in retrospect due to a cautionary quantity of mother liquor being left around the crystal in the capillary to ensure that the single existing large crystal did not dry out. Data were recorded to two limiting resolutions: 1.8 Å and 2.7 Å. In spite of the problem with crystal slipping, the merging factor, defined as $R_{sym} = \sum\sum |I - I_i| / \sum\sum I_i$ where I is the mean value of i intensity observations I_i, for the 1.8 Å data was 8.2%. For the 2.7 Å data, with no crystal slippage – the crystal was dried out before collecting these data, the R_{sym} was 4.2%. For the final merging of all the data R_{sym} was 7.2%. There were 26664 reflections in the reduced data, making up 93.0% of the complete data to 1.8 Å. 88.4% of these had an intensity greater than three standard

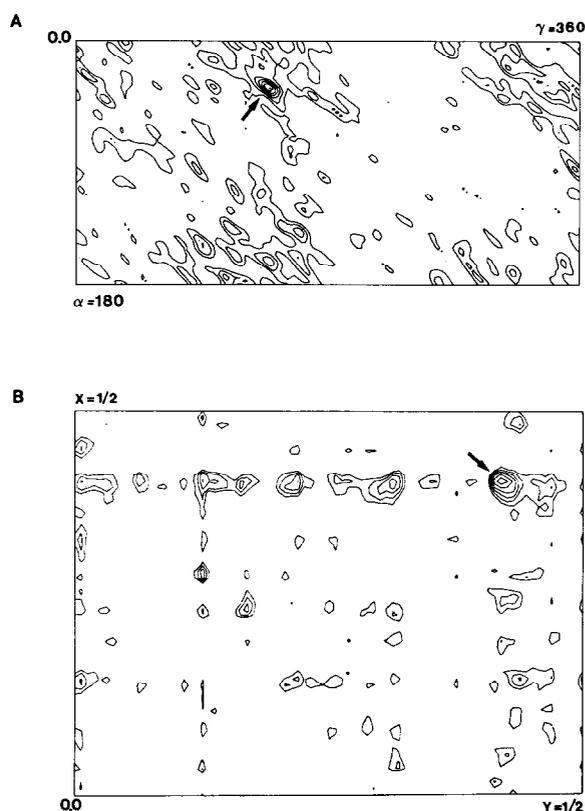
Fig.1. Results of the molecular replacement method with subtilisin Carlsberg-eglin as a model using the resolution range between 3 and 10 Å: (a) section $\beta = 51°$ of the rotation map, (b) section $z = 0.33$ of the R search map. The function plotted is the number of standard deviations of the point below the average R factor for the whole asymmetric unit of the synthesis.

deviations. Details of the data processing and of the refinement outlined below will be published elsewhere.

The phase problem was solved using the refined coordinates of the subtilisin Carlsberg-eglin complex [13] kindly provided by Professor M. James, Edmonton. The orientation and position of the complex in the unit cell were readily determined using the fast rotation function and the R factor translation search programmes from the CCP4 package (fig.1), in spite of the relatively low sequence homology of thermitase and subtilisin Carlsberg. Alignment of the sequences of the two proteases gives identical residues at 48% of a possible total of 267 positions in the two chains, with insertions or deletions at 6 points.

The model of subtilisin Carlsberg-eglin was now placed in the correct position in the thermitase-eglin unit cell, and refined at 3.0 Å resolution by restrained least-squares minimisation for 20 cycles. The R factor fell from 51.5% to 30.2%. At this stage the $2|F_o| - |F_c|$ map was inspected using FRODO, and the subtilisin side chains changed to those of thermitase. It was possible to fit most of the side chains into the electron density, with some problems on the surface and in the loops. A further 16 cycles of least-squares minimisation were performed using data from 5.0 to 2.8 Å, and the R factor dropped to 25%. An additional 30 cycles of refinement using data up to 2.5 Å resulted in the current R factor of 26%. The position of almost all the residues can now be clearly seen in the electron density, with only a few ambiguities remaining in the surface loops, particularly where there are deletions/insertions between the two proteins and there are indications of some changes to the published sequence. The remaining ambiguities will be clarified during the refinement of the structure at high resolution. A representative view of the present $2|F_o| - |F_c|$ map is shown in fig.2 and a plot of the C_α positions in the complex is shown in fig.3. The eglin electron density is generally less well defined than that of thermitase, and may benefit from a rigid-body refinement of the eglin relative to the protease position: the N-terminal segment is cleaved off and there is only defined electron density from Phe-10 onwards.

4. DISCUSSION

The sequences of a representative set of proteases from the subtilisin family are aligned in fig.4. The *Bacillus* sequences shown are those from *B. mesentericus*, a mesophilic species, and from *B. amyloliquefaciens* (BPN) and *B. licheniformis*, the two thermophilic species for which three-dimensional structures have been determined. These sequences are highly homologous, with identical residues at more than 70% of the aligned positions. There is only one deletion of a single residue in the *B. licheniformis* sequence, corresponding to position 55 in the other two. There are no additional amino acids at the C- or N-termini. No free cysteines or disulphide bridges are to be found in any of the three proteins. The secondary structural elements of the two three-dimensional *Bacillus* structures are shown in the figure, and the positions of the catalytic residues are marked.

The sequences of the two thermophilic fungal proteases, thermitase and proteinase K, are also aligned with those of the *Bacillus* subtilisins in fig.4, and the secondary structural elements of the refined proteinase K are shown. These two proteins have a much lower homology with the *Bacillus* enzymes than the internal *Bacillus* homology given above. There are roughly 45% identical residues between thermitase and the *Bacillus* enzymes, 40% between proteinase K and the latter, and indeed only 34% between ther-

Fig.2. A superposition of several sections of the $2|F_o| - |F_c|$ electron density map at the current stage of refinement. The active site is shown. The free cysteine is marked by an arrow. The sulphurs in the cysteine 75 and methionine 226 are labelled with S.
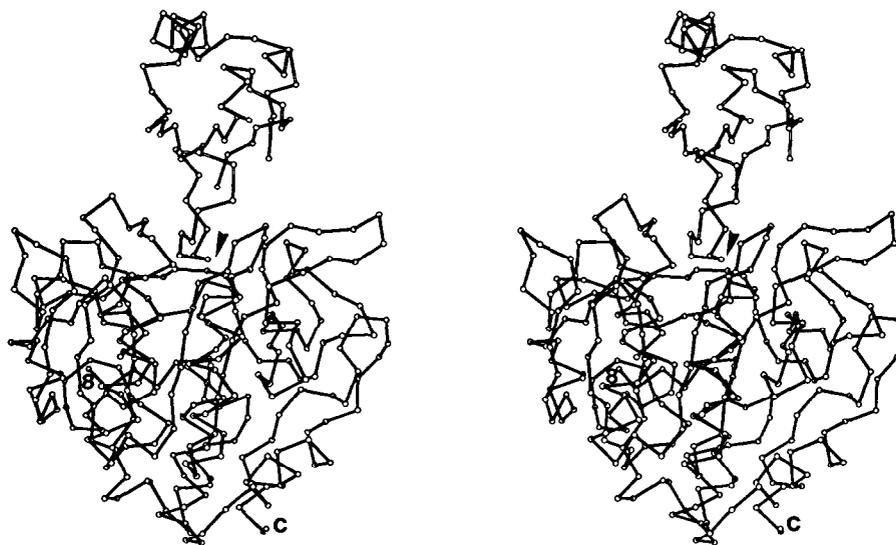
Fig.3. An α-carbon backbone drawing of thermitase (filled lines) complexed with eglin C (open lines). The active site is indicated by an arrow. The N-terminal start of the interpretable density is labelled at residue 8 and the C-terminus is also labelled.

mitase and proteinase K. There are between 6 and 8 deletions or insertions required to align proteinase K or thermitase with the *Bacillus* enzymes. Moreover these do not fall consistently in the same positions in the two sequences. Only four deletions/insertions are required to align the thermitase and proteinase K sequences with one another. As would be expected the deletions and insertions occur systematically on the surface of the proteins, in the loops between the α-helices and β-strands, allowing the overall subtilisin fold to be maintained.

A least-squares superposition of the structurally equivalent α-carbon atoms of thermitase and proteinase K gives at this stage of refinement an rms deviation of 0.89 Å. This value agrees with the results of a study of the relationship between sequence homology and similarity of three-dimensional structure of proteins [25] for the observed homology of 34%. The best agreement is in the active site region with an rms distance for 50 catalytically important atoms including the free Cys-75 and Met-226 of only 0.53 Å. However the close similarity of three-dimensional folding, as for all known subtilisins, is not restricted to the catalytically essential residues, but also extends to regions which are known to be involved in the substrate binding.

Thermitase and proteinase K together with thermomycolin form part of that subgroup of the subtilisins which contain a free cysteine SH group, Cys-75 in thermitase, at the same position in the sequence. The sequences of the mesophilic enzymes from *B. cereus* and *B. thuringiensis* containing such an SH group have unfortunately not yet been completed; however it is clear that the presence of the SH group is not restricted to the fungal enzymes. The free cysteine is located in the immediate vicinity of the active site and could influence the substrate cleavage. In this context it is of interest that in recent experiments with subtilisin BPN' the Met-222 (which corresponds to Met-226 in thermitase, fig.2) lying near the active Ser was substituted by each of the other 19 amino acids. All the mutants obtained were less active than the native enzyme, except for Cys-222 which increased the activity by 38% [26]. Thus the nearby SH group has a positive influence on the activity of subtilisin BPN'. The mechanism of this phenomenon in the mutant BPN' and the influence of the Cys-75 SH-group on the activity of thermitase are up to now not fully established.

The cysteine in thermitase is placed 'below' the functional His-71, with an S-NE2 distance of about 4 Å (fig.2) and appears to be practically inaccessible to the solvent. However HgCl₂ is a

Fig.4. Sequence comparison of thermitase (TH) with proteinase K (PRTK) and the subtilisins BPN' and Carlsberg (CA). The numbering above the top line corresponds to thermitase, that below the bottom line to subtilisin Carlsberg. Homologous amino acids are boxed; the catalytically active amino acids are indicated by arrows. The free cysteine residue is indicated by an asterisk. Proteinase K is the only member of the family with two disulphide bridges. The secondary structural elements of proteinase K and Carlsberg are shown above and below the sequences respectively.

known inhibitor of thermitase through its binding to the free SH group. Some flexibility in this region of the protein would be required to allow $HgCl_2$ access to the cysteine thiol group. To identify the coordination and binding geometry of the Hg^{2+} we have grown crystals of thermitase in the presence of $HgCl_2$ and plan to determine its structure.

At this stage of the refinement we were able to make a preliminary identification of two calcium cations in the structure, on the basis of peaks in the difference Fourier syntheses, the distance of these peaks to neighbouring waters and protein oxygens and the relatively high electron density of the two peaks. These we assume correspond to the two tightly bound calcium ions defined in the biochemical studies. One of these tightly bound calcium ions is located within the extended loop segment Ala-81 to Ile-87 which protrudes out of the surface of the enzyme. The mean distances of the proposed calcium to the chelating water and protein oxygen atoms is about 2.4 Å. This loop is absent in proteinase K, which is highly resistant to autolysis. This calcium site in thermitase is thus likely to be that which confers protection to autolysis, the autolysis initiating at this exposed loop if no calcium is bound.

The second proposed site is located in a rather small cleft on the surface of thermitase in the region Pro-172 to Ile-179. It has roughly bipyramidal coordination to protein and 3–4 water oxygen atoms. We assume this site is the second tightly bound calcium necessary for activity of the enzyme by stabilising the geometry of the substrate binding region, and also present in proteinase K [27] and other subtilisins.

A detailed analysis of the geometry of the calcium sites will be published later after refinement of the two forms of the complex at high resolution (see below). It appears however that at least two of the three calcium sites identified from solution studies are occupied in the present crystals, where the calcium concentration is 5 mM.

At the time when we were collecting the high resolution data on the thermitase-eglin complex, we became aware that P. Gros and co-workers in Groningen had also obtained orthorhombic crystals of the complex (Hol, W.G.J., personal communication). The high resolution analysis of the structure will therefore be carried out jointly by the two groups. We have, however, since then

established that the unit cells of the Groningen and Hamburg crystals, though both are in space group $P2_12_12_1$, are quite different. The comparison between the two structures will allow us to identify more accurately those surface groups which are perturbed by crystal contacts in the two crystal forms and which may deviate from the solution structure of the complex.

Future work thus includes the refinement of the structure of the complex, the comparison of the inhibited thermitase structure with that of the native enzyme, and with other members of the subtilisin family.

REFERENCES

[1] Kleine, R. and Rothe, U. (1977) Acta Biol. Med. Germ., K27–K31.
[2] Frömmel, C., Hausdorf, G., Höhne, W.E., Behnke, U. and Ruttlof, H. (1978) Acta Biol. Med. Germ. 37, 1193–1204.
[3] Frömmel, C. and Höhne, W.E. (1981) Biochim. Biophys. Acta 670, 25–31.
[4] Meloun, B., Baudys, M., Kostka, V., Hausdorf, G., Frömmel, C. and Höhne, W.E. (1985) FEBS Lett. 183, 195–200.
[5] Stahl, M.L. and Fessari, E. (1984) J. Bacteriol. 158, 411–418.
[6] Markland, F.S. and Smith, E.L. (1967) J. Biol. Chem. 242, 5198–5211.
[7] Smith, E.L., De Lange, R.J., Evans, W.H., Landon, M. and Markland, F.S. (1968) J. Biol. Chem. 243, 2184–2191.
[8] Schechter, I. and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162.
[9] Svendsen, I., Genov, N. and Idakieva, K. (1986) FEBS Lett. 196, 228–232.
[10] Nedkov, P., Oberthür, W. and Braunitzer, O. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 1537–1540.
[11] Wright, C.S., Alden, R.A. and Kraut, J. (1969) Nature 221, 235–242.
[12] Bode, W., Papamokos, E., Musil, D., Seemüller, U. and Fritz, H. (1986) EMBO J. 5, 813–818.
[13] McPhalen, C.A., Schnebli, H.P. and James, M.N.G. (1985) FEBS Lett. 188, 55–58.
[14] Jany, K.D., Lederer, G. and Mayer, B. (1986) FEBS Lett. 199, 139–144.
[15] Gaucher, G.M. and Stevenson, K.J. (1976) Methods Enzymol. 45, 414–433.
[16] Chestukina, G.G., Epremyan, A.S., Gaida, A.V., Osterman, A.L., Khodova, O.M. and Stepanov, V.M. (1982) Bioorg. Chim. 8, 1649–1658.
[17] Stepanov, V.M., Chestukhina, G.G., Radenskaya, G.N., Epremyan, A.S., Osterman, A.L., Khodova, O.M. and Belyanva, L.P. (1987) Biochem. Biophys. Res. Commun. 100, 1680–1687.

[18] Betzel, C., Pal, G. and Saenger, W. (1988) Acta Cryst. B44, 163–172.

[19] Teplyakov, A.V., Strokopytov, V., Kuranova, I.P., Popov, A.N., Harutunyan, E.G., Vainstein, B.K., Frömmel, K. and Höhne, W. (1986) Kristallografiya 31, 931–936.

[20] Brömme, D., Peters, K., Fink, S. and Fittkau, S. (1986) Arch. Biochem. Biophys. 244, 439–446.

[21] McPherson, A.C. (1982) Preparation and Analysis of Protein Crystals, Wiley, New York.

[22] Machin, P.A., Wonacott, A. and Moss, P. (1983) Daresbury Laboratory News Lett. 10, 3–9.

[23] Jones, T.A. (1983) Computational Crystallography (Sayre, D. ed.) pp.303–317, Oxford University Press, Oxford.

[24] Matthews, B.W. (1968) J. Mol. Biol. 33, 491–497.

[25] Chothia, C. and Lesk, A.M. (1986) EMBO J. 5, 823–826.

[26] Estell, D.A., Graycar, T.P. and Wells, J.A. (1985) J. Biol. Chem. 260, 6518–6521.

[27] Betzel, Ch., Pal, G.P. and Saenger, W. (1988) Eur. J. Biochem., submitted.