ten neighbouring molecules are placed in the $x0z$ (or $\bar{x}0z$) orientation, the origin molecule can rotate, without hindrance, from the $x0z$ to the $\bar{x}0z$ equilibrium position (according to the previous criterion on the H—H distance). The high steric hindrance along the **b** axis explains the low contraction of the **b** parameter with temperature, compared with that along **a** (Fig. 2).

Concluding remarks

By careful refinements, probability distributions and steric hindrance considerations we have proved that the orientational disorder of 1-iodoadamantane can be described with a Frenkel model. The rotation between the two equilibrium positions is quasi-free.

An antiferroelectric order has been pointed out similar to that locally encountered in the glassy phase of 1-cyanoadamantane. Some geometrical and energy features have been qualitatively examined to explain the lack of plastic phase in comparison to other adamantane derivatives.

An extension of this study will be to investigate the behaviour of mixtures of 1-cyano- and 1-iodoadamantane. If these two compounds syncrystallize, complete thermodynamic, structure and molecular-motion analyses should allow the part played by physical parameters in a glassy crystalline phase (*i.e.*, molecular geometry, dipolar interaction, slow molecular tumbling and particularly the local order) to be quantified.

The authors are grateful to M. Muller and D. Prevost for preparing the sample.

References

ANDRÉ, D., FOURME, R. & RENAUD, M. (1971). *Acta Cryst.* B**27**, 2371–2380.
BEE, M., FOULON, M., AMOUREUX, J. P., CAUCHETEUX, C. & POINSIGNON, C. (1987). *J. Phys. C.* In the press.
CLARK, T., MCKNOX, O., MACKLE, H. & MCKERVEY, M. A. (1977). *J. Chem. Soc. Faraday. Trans.* **73**, 1224–1231.
COPPENS, P., LEISEROWITZ, L. & RABINOVICH, D. (1965). *Acta Cryst.* **18**, 1035–1038.
CRUICKSHANK, D. W. J. (1956). *Acta Cryst.* **9**, 757–758.
CUVELIER, P. & FOULON, M. (1986). Private communication.
DESCAMPS, M., CAUCHETEUX, C., ODOU, G. & SAUVAJOL, J. L. (1984). *J. Phys. Lett.* **45**, L719–727.
FOULON, M., AMOUREUX, J. P., SAUVAJOL, J. L., CAVROT, J. P. & MULLER, M. (1984). *J. Phys. C*, **17**, 4213–4229.
FOULON, M., LEFEBVRE, J., AMOUREUX, J. P., MULLER, M. & MAGNIER, D. (1985). *J. Phys. (Paris)*, **46**, 919–926.
LEFEBVRE, J., ROLLAND, J. P., SAUVAJOL, J. L. & HENNION, B. (1985). *J. Phys. C*, **18**, 241–255.
MAGNIER, D. (1986). Thesis, Univ. of Lille, France.
MAIN, P., LESSINGER, L., WOOLFSON, M. M., GERMAIN, G. & DECLERCQ, J.-P. (1977). *MULTAN77. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.
SCHOMAKER, V. & TRUEBLOOD, K. N. (1968). *Acta Cryst.* B**24**, 63–76.
SHELDRICK, G. M. (1976). *SHELX76.* Program for crystal structure determination. Univ. of Cambridge, England.
VIRLET, J., QUIROGA, L., BOUCHER, B., AMOUREUX, J. P. & CASTELAIN, M. (1983). *Mol. Phys.* **48**, 1289–1303.

Acta Cryst. (1988). B**44**, 163–172

Synchrotron X-ray Data Collection and Restrained Least-Squares Refinement of the Crystal Structure of Proteinase K at 1·5 Å Resolution

BY CH. BETZEL,* G. P. PAL AND W. SAENGER

Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, *D*-1000 *Berlin* 33, *Federal Republic of Germany*

(*Received* 2 *June* 1987; *accepted* 22 *September* 1987)

Abstract

The structure of the serine endopeptidase proteinase K (279 amino acid residues; 28 790 daltons) has been refined by restrained least-squares methods to a conventional R value of 16·7% employing synchrotron film data of 30 812 reflections greater than 3σ in the 5·0 to 1·5 Å resolution range. During refinement, the molecular structure was restrained to known stereochemistry, with root-mean-square (r.m.s.) deviation of 0·015 Å from ideal bond lengths. The average atomic temperature factor, B, is 11·1 Å^2 for all atoms. The final model comprises 2020 protein atoms and 174 solvent molecules (which were given unit occupancies). Four corrections to the amino acid sequence were made, which were confirmed later by sequence analysis of the proteinase K gene: a deletion of one glycine in position 80; a change of sequence in position 207–208 and insertions of the dipeptide 210–211 and of residue 270. The r.m.s. deviation in the α-C atomic positions between the final refined model and the initial model

* Present address: European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Federal Republic of Germany.

0108-7681/88/020163-10$03.00

built on the basis of a 3·3 Å mini-map is 1·72 Å for 227 out of 266 residues, which were originally traced in the mini-map without sequence information. The positions of the remaining 39 residues deviate by more than 8 Å from the original ones and are located in regions where extensive revision of the structural model was necessary.

Introduction

Proteinase K (E.C. 3.4.21.14) isolated from the fungus *Tritirachium album* Limber is an endopeptidase consisting of 279 amino acids with known sequence (Jany & Mayer, 1985; Jany, Lederer & Mayer, 1986) and has a molecular weight of 28 790. At its active site, it carries the triad Asp39, His69, Ser224, which classify proteinase K as a serine protease (Ebeling, Hennrich, Klochow, Metz, Orth & Lang, 1974). In the degradation of native proteins, it is the most potent known proteolytic enzyme (Ebeling *et al.*, 1974; Kraus, Kiltz & Femfert, 1976) and therefore has found application in the preparation of protein-free nucleic acids (Wiegers & Hilz, 1981). Moreover, in contrast to all known proteases, proteinase K even hydrolyses native keratin, therefore the letter K.

Proteinase K was crystallized and its main-chain folding determined on the basis of a 3·3 Å resolution electron density map phased by multiple isomorphous replacement methods (MIR) without amino acid sequence information (Pähler *et al.*, 1984). The three-dimensional structure shows a high degree of homology to the bacterial subtilisins. Since the subtilisins are of prokaryotic origin while proteinase K is eukaryotic, it was of particular interest to obtain more detailed information concerning questions relating to enzyme–substrate interactions, to structure–activity correlations, and to evolutionary aspects.

Because the crystals of proteinase K diffract X-rays to high resolution (Dattagupta *et al.*, 1975), it provides a good system for studying in detail enzyme–inhibitor complexes and in this way to map out the active-site geometry (Betzel, Pal, Struck, Jany & Saenger, 1986; Pal, Betzel, Jany & Saenger, 1986). In this context, it was essential to study the structure of the native enzyme at high resolution. This was done with synchrotron X-ray diffraction data, and in the following data collection and structure refinement will be described.

Crystallization and data collection

Proteinase K was kindly provided in pure form from Drs N. Hennrich and H. D. Orth (Merck, Darmstadt, Federal Republic of Germany). Crystals of 0·7 × 0·7 × 1·0 mm were grown by microdialysis from a 10% protein solution containing 50 m*M* Tris buffer [2-amino-2-(hydroxymethyl)-1,3-propanediol], 10 m*M* $CaCl_2$, 1 *M* $NaNO_3$ and 0·02% NaN_3 adjusted to pH 6·5 (Dattagupta *et al.*, 1975). Space group, cell constants and other crystallographic details are given in Table 1.

Table 1. *Proteinase K crystal data*

Space group	$P4_32_12$
Crystal habit	Fully developed or truncated tetragonal bipyramids
Cell dimensions	$a,b = 68{\cdot}17\ (6), c = 108{\cdot}26\ (6)$ Å
Unit-cell volume	$5{\cdot}03 \times 10^5$ Å^3
Molecular mass of proteinase K	28 790 daltons
Number of amino acids	279
Average size of crystals used in the collection of data	0·7 × 0·7 × 0·9 mm
Crystal density	1·285 g cm^{-3}
Crystal volume/dalton	2·17 Å^3/dalton
Z	8

Because of its efficiency in recording the large number of X-ray data required for high-resolution protein crystal structure determination, we used the Arndt–Wonacott oscillation film camera (Arndt & Wonacott, 1977). Two film data sets were collected, the first to 2·2 Å resolution using an in-house rotating-anode Elliott GX20 generator and the second to 1·5 Å resolution using the X-rays emitted from the storage ring DORIS at DESY, Hamburg.

In all X-ray experiments, crystals of proteinase K were mounted in thin-walled quartz capillaries with some mother liquor. The orientation of the crystals in the capillaries was such that rotation was about the c axis so as to make use of the inherent fourfold symmetry which required rotation in the range $0 \leq \varphi \leq 45°$. Most of the reflections were measured four times, as Friedel equivalents (hkl and $hk\bar{l}$) on the same film and as symmetry-related equivalents on different films.

2·2 Å rotating-anode data collection and processing

For the determination of crystal orientation, still photographs with spindle settings $\varphi = 0$, 45 and 90° were taken. Following this, data were collected with sequential non-overlapping oscillation ranges which allows summation of terms partially recorded at the end of one exposure and at the beginning of the next. There were three films per pack. The crystal was stable enough in the X-ray beam so that the whole data set could be collected from one specimen using the operating conditions given in Table 2. The first photograph was repeated after data collection was completed (56 h). It showed no visible decrease in reflection intensities.

The film orientation matrix and the crystal-to-film distance were determined and refined on the basis of the positions of the fiducial spots and of the positions of 40 strong well defined reflections on the still photographs. Data films were digitized using an Optronics Photoscan P-1000 drum scanner operated in an off-line

Table 2. *Data collection and processing: experimental conditions*

	To 2·2 Å resolution	To 1·5 Å resolution
X-ray source	Elliott GX20 rotating-anode generator operated at 40 kV, 70 mA (2·8 kW), focal size 0·2 × 2·0 mm	Beam-line X11/EMBL Hamburg storage ring DORIS, 5·1 GeV, 30–40 mA
Detector	Film, Kodak DEF-2	Film, Kodak DEF-2
Collimation	Standard pinhole, 0·6 mm diameter	0·35 mm diameter
Film cassettes	Flat plate, radius 60 mm	Flat plate, radius 60 mm
Crystal-to-film distance	70 mm	90 mm
Total rotation (45° required for *c*-axis rotation) for a data set	50°	47°
for each exposure	2°	0·5°
Number of films per data set	75 (25 three-film packs)	282 (94 three-film packs)
Time for each exposure	4000 s deg^{-1}; 2·2 h; 10 oscillations	~8 min
Wavelength	1·54182 Å Cu $K\alpha$/graphite monochromator	0·8655 Å
Polarization corrections for synchrotron radiation		
horizontal divergence		0·17–0·20°
vertical divergence		0·013–0·03°
dispersion		0·0015–0·0023°
Results of data processing		
Total number of reflections	41304	116040
Number of independent merged reflections	11520	36039
$R(I)$ merge (%)	8·9	8·8
Number of reflections rejected with $\lvert I-\bar{I}\rvert < 4\sigma$	394	627
Number of reflections with negative intensity	204	1058
Number of reflections $> 1\sigma$	11510	36014
$> 2\sigma$	11299	35349
$> 3\sigma$	10876	32096

mode. The reflections on the films are about 0·9 mm in diameter, and were scanned with 100 μm grid width. The optical density values in the range 0·0–3·0 OD units were output as integers in the range 0–255. The further integration and reduction of the data were subsequently carried out on the in-house VAX 11/750 computer using the *MOSCO* program system (Machin, Wonacott & Moss, 1983).

A typical 2·2 Å resolution film contains some 2200 reflections, including partials. Only one to three strong reflections are 'overloaded' on the third films of the packs.

The integrated X-ray intensities were corrected for Lorentz–polarization and oblique-incidence effects (Cox & Shaw, 1930; Whittaker, 1953). The inter-film scaling was performed with the scale-sort-merge program (*SSM*; Wilson & Yeates, 1979), and symmetry-equivalent terms were averaged to yield a unique data set. The scale and relative temperature factors between the film sets, which to some extent correct residual radiation-damage effects and small differences in absorption, were determined by the method of Fox & Holmes (1966). The data were finally passed through the program *TRUNCATE* (French & Wilson, 1978) which employs the principles of Bayesian statistics to correct for negative and very weak intensities. The final data set covered about 95% of the total data to 2·2 Å resolution, with statistics given in Table 2.

1·5 Å synchrotron data collection and processing

The film data collection at the EMBL outstation at DORIS was carried out during 'parasitic' beam time in March 1983, when DORIS was operated in single-bunch mode at 5·1 GeV and about 30–40 mA with approximately 1 h between injections. The Arndt–Wonacott camera was installed on the 'old' beam line X11 on the positron side of the storage ring with a collimator system especially constructed and optimized for synchrotron radiation (Bartunik, Gehrmann & Robrahn, 1984). Meanwhile, this beam line has been reconstructed and now provides about ten times higher intensity, *i.e.* a factor 400 more brilliant compared to X-ray generators with rotating anodes (Wilson & Betzel, 1986, private communication).

On the X11 synchrotron line, the attainable limit of resolution of the proteinase K crystals extended to 1·0 Å. Attempts to collect data at this resolution were, however, not successful since the diffraction pattern decreased in the 1·0 Å range after about 30 min, and the exposure time needed to produce measurable reflections caused strong background scattering which decreased the overall quality of the data.

Therefore, we decided to collect a data set with a nominal resolution of 1·5 Å. To estimate the orientation matrix and to refine crystal-to-film distance and wavelength, still photographs were recorded at spindle settings 0 and 90°, with exposure times of 10 s. X-ray intensity data were collected in oscillation steps of 0·5° in the range of 0 to 47° from one crystal, which was moved three times across the beam during the experiment to expose different new volumes. The parameters and conditions for data collection are summarized in Table 2.

Because the average spot size was 0·7 mm diameter, the films could be scanned with the 100 μm raster. The optical density was digitized in the 0–2 OD range owing to the high proportion of weaker intensities near the

resolution limit of 1·5 Å. The data were finally processed as described for the 2·2 Å set. A typical film contained around 2100 reflections of which about 40% were partially recorded. In addition to the geometrical corrections described for the data collection with the rotating-anode X-ray generator, the expression for polarization was applied as given by Kahn, Fourme, Gadet, Janin, Dumas & Andre (1982), because the synchrotron X-radiation is considerably polarized.

Statistics describing the quality of this data set are given in Table 2. For the whole data set, 89% of the reflections had intensities greater than 1σ, where the standard deviation σ is derived from background statistics. The accuracy of the refined crystal orientation matrix can be assessed by comparing the intensities of symmetry-equivalent reflections which were fully recorded on one film and partially recorded on two adjacent film packs (Fig. 1). For both data sets, a marginal systematic discrepancy is observed which is probably caused by small residual errors in the orientation matrices. It leads to an underestimation of apparently fully recorded reflections, *i.e.* some reflections are flagged as fully recorded when they are, in reality, partials.

The merging $R(I)$ factor for the 1·5 Å synchrotron and the 2·2 Å rotating-anode data sets is 8·1% for the 10 694 common reflections. The quality of the high-resolution data can be estimated from the Wilson distribution which is illustrated for the total of 36 039 intensity data in Fig. 2. The fit to the theoretical straight line is very satisfactory for a protein data set except for the range 4 to 6 Å where contributions due to secondary-structure elements come into play. The overall temperature factor obtained from the slope of the least-squares line, 12·2 Å^2, was later confirmed by the refinement.

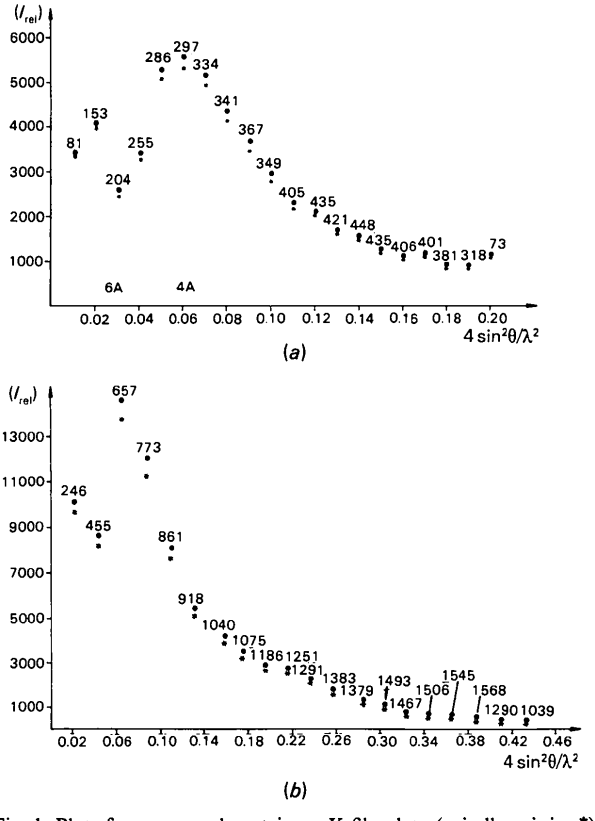

Fig. 1. Plots for processed proteinase K film data (spindle axis is c^*) (*a*) at 2·2 Å resolution (rotating anode) and (*b*) at 1·5 Å (synchrotron radiation). Plotted is the average intensity *vs* $4\sin^2\theta/\lambda^2$ for fully recorded reflections (*) and their equivalent reflections measured as the sum of two partially recorded reflections on adjacent films (•). The numbers above each point refer to the number of reflections in that range.

Course of model building and refinement

Because the data-parameter ratio is not sufficient even at 1·5 Å resolution to refine atomic coordinates independently, the method of restrained least squares was applied as programmed by Hendrickson & Konnert (1981). In recent years, the original program *PROLSQ* has been intensively improved especially with the introduction of space-group-specific subroutines for the most time-consuming step involving structure-factor calculations (Machin, Wonacott & Moss, 1984). Therefore, all refinement runs were performed on the in-house VAX 11/750 computer where a typical cycle near the end of the procedure, with 30 812 data and 8777 variables needed about 10 h CPU time, and 3·5 h at the 2·5 Å resolution stage.

(*a*) *Model building*

The initial structure model of proteinase K was constructed according to a manual tracing of 266 α-C

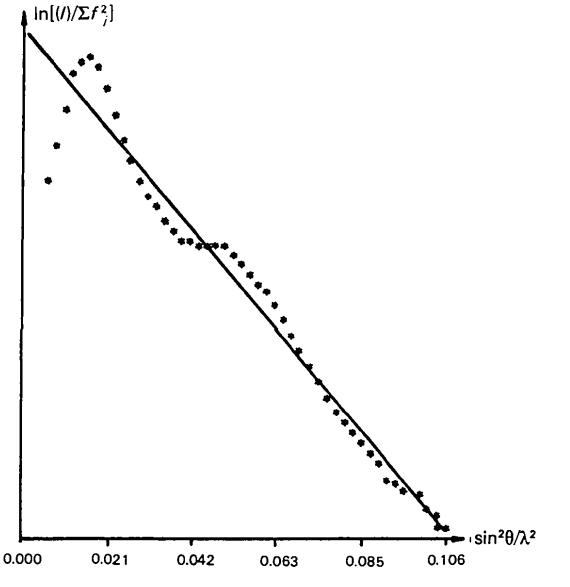

Fig. 2. Wilson plot for proteinase K using 1·5 Å resolution synchrotron film data.

atoms carried out at 3·3 Å resolution on the basis of a multiple isomorphous replacement (MIR) electron density mini-map drawn at a scale of 1 cm/3·8 Å (Pähler *et al.*, 1984). At that time, the amino acid sequence was only available for some short unconnected fragments whose position relative to the total sequence was not known except for the active-site region where comparison with the homologous subtilisin allowed the assignment of some amino acids according to the electron density maps. At this stage an Evans & Sutherland PS 300 color graphics display became available which, in combination with the program system *FRODO* (Jones, 1978; modified by J. Pflugrath) was instrumental in the improvement of the model fitting.

In the first stage of model building a polyalanine chain was fitted to the MIR electron density map calculated at 2·8 Å resolution. Some of the more obvious side chains were added so that 1120 out of about 2000 protein atoms were located which belonged to 230 peptide units. In order to introduce as small as possible errors in subsequent phase-angle calculations, the newly included atoms were chosen very conservatively. In this starting model, the amino and carboxy termini were omitted as well as some peripheral loops which were ill defined. The geometry of this partial model was idealized and a crystallographic R factor was calculated to 52% for data between 5 and 2·8 Å resolution. Data in the 5 Å to ∞ region were not included because they are affected by contributions due to solvent and protein secondary structure.

Attempts to improve the proteinase K model on the basis of an electron density map calculated with coefficients $2F_o - F_c$ and phase angles α_c failed because the quality of the α_c was still too poor. Therefore, the α_c were combined with the MIR phase angles α_{MIR} according to the algorithm developed by Hendrickson & Lattman (1970). The mean phase-angle difference obtained was 58°, a value in a typical range also observed by other authors at comparable stages of protein refinement (Holmes & Matthews, 1982; Higuchi, Kusanoki, Matsuura, Yasuoka & Kakudo, 1984). The calculated phase angles α_c were weighted according to a scheme developed by Sim (1959) which takes into account their relative contribution to the total crystal structure. The interpretability of the obtained 'combined' electron density map was considerably improved especially in regions which previously were not so clearly defined.

When the complete amino acid sequence of the protein became available (Jany & Mayer, 1985; Jany, Lederer & Mayer, 1986), the interpretation of the electron density map was speeded up considerably. There were, however, four errors in the sequence which became obvious at high resolution and involved the inclusion of two additional amino acids to yield a total of 279 residues. The modifications which had to be made to the amino acid sequence were a deletion of glycine in position 80, an interchange of sequence 207–208 and insertion of a dipeptide in position 210–211, and another insertion of residue 270. All these changes were later confirmed when the gene of proteinase K was cloned and sequenced (Gassen, 1986, private communication). In order to fit all main-chain and side-chain atoms to the electron density map, electron densities with coefficients $F_o - F_c$ and $2F_o - F_c$ were calculated as well as 'omit' maps (Bhat & Cohen, 1984), where sequences of up to ten amino acids were omitted from the calculated electron density and checked in a subsequent $F_o - F_c$ difference map. In several cases these 'omit' maps proved more valuable than $2F_o - F_c$ maps because phase angles calculated with a partially incorrect model can lead to a false electron density map. In extreme cases this can worsen the calculated phase angles especially for the high-resolution structure factors so that the effective resolution is reduced.

After each model-building session, the structure model was subjected to several cycles of combined idealization and least-squares refinement. This continued until the polypeptide chain was reasonably well defined with 275 amino acids and 1917 atoms. The positions of 17 side chains which were located at the periphery and of four amino acids at the carboxy terminus were not yet obvious at this stage but clearly showed up when higher-resolution data were included near the end of the refinement.

(*b*) *Refinement procedure*

At the beginning of the refinement some test calculations were carried out to derive an optimum weighting scheme. The best weighting factor $\sigma(F)$ for the structure amplitudes was set as a function of resolution, $\sigma(F) = A + B(\sin\theta/\lambda - \frac{1}{6})$, with $A = \frac{1}{2}(\langle |F_o| \rangle - \langle |F_c| \rangle)$, and B was set to -45. These resolution-dependent weighting parameters were found satisfactory and were maintained during the whole refinement process. The refinement was carried out in eight steps with a total of 113 least-squares cycles including data with $F > 3\sigma(F)$. At the end of each of the eight steps, the model was checked carefully on the basis of $2F_o - F_c$ and $F_o - F_c$ electron density maps. A summary of the refinement is given in Tables 3 and 4.

In *step* 1 of the refinement, the initial proteinase K model consisting of 1917 atoms was refined with 7369 film intensity data in the 5·0 to 2·5 Å range collected with the rotating-anode X-ray generator. The R factor reduced from 45·4 to 29% where it converged. Besides atomic coordinates (torsion angles), the scale factor and the overall temperature factor were refined. In addition to the stereochemistry of the amino acids, the then known N—H···O hydrogen bonds of the secondary-structure elements and the S—S distances in the two disulfide bridges were restrained to canonical values. At

Table 3. *Description of the course of refinement*

Step	1	2	3	4	5	6	7	8
Number of refinement cycles	17	12	22	13	19	7	12	11
Number of parameters varied	5753	6101	8133	8133	8517	8829	8797	8777
Number of reflections	7369	9647	13 697	20 427	20 427	20 427	20 427	30 812
Number of atoms	1917	2033	2033	2033	2129	2207	2199	2194
Solvent molecules (water)	(0)	(76)	(76)	(76)	(123)	(185)	(176)	(174)
Resolution range (Å)	5·0–2·5	5·0–2·3	5·0–2·0	5·0–1·8	5·0–1·8	5·0–1·8	5·0–1·8	5·0–1·5
Mean temperature factor B (Å^2)	8·2	8·0	9·3	10·8	11·1	11·6	11·4	11·1
$(\lvert F_o\rvert - \lvert F_c\rvert)$	132·5	122·4	90·5	78·0	63·0	57·0	55·2	48·3
R (start of refinement) (%)	45·4	41·1	32·1	25·5	30·1	21·9	17·1	18·5
R (end of refinement) (%)	29·0	29·1	23·1	23·2	18·8	16·6	16·1	16·7
Standard deviations from ideal values								
Bond lengths (1–2 distances) (Å)	0·031	0·030	0·034	0·029	0·023	0·015	0·020	0·015
(1–3 distances) (Å)	0·111	0·089	0·067	0·060	0·052	0·038	0·041	0·034
Planar groups (Å)	0·023	0·019	0·024	0·024	0·020	0·021	0·019	0·012
Chiral volumes (Å^3)	0·37	0·30	0·32	0·28	0·26	0·19	0·19	0·15
Torsion angle (°) of peptide plane (ω)	3·9	7·8	5·5	7·0	9·5	7·0	6·3	4·9

Table 4. *Weighting scheme, results and standard deviations after the last cycle of refinement*

	σ*	Standard deviation	No. of parameters
Bond lengths (1–2 neighbours) (Å)	0·017	0·015	2058
Angles corresponding to (1–3 neighbours) (Å)	0·027	0·034	2798
Planes corresponding to (1–4 neighbours) (Å)	0·037	0·043	753
Planar groups (Å)	0·020	0·012	357
Chiral volumes (Å^3)	0·100	0·146	312
van der Waals contacts (Å)			
Single torsion contacts	0·400	0·166	714
Multiple torsion contacts	0·400	0·309	907
Possible hydrogen-bonding contacts	0·400	0·173	232
Torsion angles (°)			
Peptide plane (ω)	2·5	4·9	288
Staggered torsion contacts (60/120°)	15·0	18·7	305
Orthonormal torsion contacts (90°)	20·0	23·1	29
Number of *cis*-prolines: 1			
Isotropic temperature factors (B in Å^2)			
Main chain (1–2 neighbours)	1·00	0·849	1152
Main chain (1–3 neighbours)	1·50	1·234	1456
Side chain (1–2 neighbours)	1·00	1·452	906
Side chain (1–3 neighbours)	1·50	2·195	1344

Final R factor in the resolution range 5·0–1·5 Å

For 30 812 reflections	$> 3\sigma(F)$	$R = 16{\cdot}7\%$
34 211 reflections	$> 2\sigma(F)$	$R = 17{\cdot}7\%$
34 876 reflections	$> 1\sigma(F)$	$R = 18{\cdot}0\%$

*The weight on each restraint corresponds to $1/\sigma$.

the end of the refinement, 76 solvent molecules (water O atoms) and 40 additional protein atoms were located. *Step* 2: The resolution limit was extended to 2·3 Å, and included 9647 reflection data. Because the synchrotron data were used from step 3 on, special weight was put on the parameters defining the stereochemistry of the model. The refinement converged at $R = 29{\cdot}1\%$. The calculated electron density did not allow the location of more protein atoms but indicated differences with the chemically determined amino acid sequence in the regions 78–82, 205 to 212, and at the carboxy terminus. *Step* 3: At this point we changed to synchrotron data in the resolution range 5·0 to 2·0 Å including 13 697 data. After ten cycles, temperature factors were refined individually, and convergence was reached after two cycles with $R = 23{\cdot}1\%$. Some of the side chains had B values around 30 Å^2, and their positions had to be redefined. *Step* 4: The resolution was extended to 1·8 Å, refinement converged after 12 cycles at $R = 23{\cdot}2\%$ for 10 427 data. The model was checked thoroughly by calculating 30 different 'omit' maps. A glycine had to be deleted in position 80, the sequence 205–211 had to be rebuilt with interchange of amino acids 207–208 and insertion of a peptide in position 210–211, and an extra Asn had to be inserted in position 270. The main chain of the carboxy terminus was traced, side chains were still unclear in this region. 46 more solvent molecules were located and a cation, which was provisionally identified as Na^+. *Step* 5: The newly included solvent atoms were assigned a B factor of 20 Å^2, all rebuilt regions of the protein structure were given a B factor of 10 Å^2. Refinement converged after 19 cycles at $R = 18{\cdot}8\%$. 63 more solvent molecules were located and the cation was redefined as Ca^{2+} based on the strong residual electron density in the $F_o - F_c$ maps and on distances to ligands coordinated to it. *Step* 6: After eight more cycles, refinement converged at $R = 16{\cdot}6\%$. Amino acid side chains of the carboxy terminus were located, except for position 278. Nine of the solvent molecules had $B > 60$ Å^2 and were omitted from the atoms list. Because occupancies and temperature factors are strongly correlated, the former were fixed at 1·0 for all solvent molecules and for all protein atoms; none of the main-chain or side-chain atoms was found to be significantly disordered. *Step* 7: Refinement converged at $R = 16{\cdot}1\%$ after 12 cycles. Two more solvent molecules were omitted and the terminal residues 278, 279 were added. The difference electron density showed no significant peaks above 0·35 e Å^{-3}. *Step* 8: All restraints in hydrogen-bond and S–S distances were released, the number of coplanar main-chain atoms was

reduced from five (α-C, C, O, N, α-C) to four (α-C, C, O, N) in order to obtain a more 'free' refinement of the structure model. All 30812 data in the 5·0 to 1·5 Å resolution range with $F > 3\sigma(F)$ (comprising 80% of the measurable data) were included in the refinement which converged after ten cycles at $R = 16{\cdot}7\%$. The quality of the refined structure is obvious from the $2F_o - F_c$ electron density sections illustrated in Fig. 3.*†

Quality of refinement

Table 4 shows the standard deviations of the stereochemical parameters. They were obtained from the correlation matrix after the last cycle of refinement. Only 7·2% of all distances between next, second-next and third-next neighbours deviate by more than two standard deviations from the corresponding ideal value. Not more than 21 distances, *i.e.* 0·4% of all distances exceed four standard deviations. The average coordinate shift was only 0·011 Å in the last cycle of refinement and the mean change in B factors was 0·10 Å^2.

In a final difference electron density, there were 20 positions where the residual density was above 0·30 e Å^{-3}. Of these, five could be interpreted as weakly occupied solvent positions and the others were close to side chains of serine, leucine or isoleucine and indicated possible minor disorder which we neglected. The side chains of glutamines 103 and 278 were not detectable in this difference map and we conclude that they are highly disordered.

The quality of the refined structure is also obvious from the Ramachandran plot, Fig. 4, in which torsion angles Φ, Ψ along the polypeptide chain are entered. The sterically allowed regions for all amino acids except glycine are indicated by dashed lines (Ramakrishnan & Ramachandran, 1965). Within these regions we find almost all amino Φ, Ψ values except for glycines and Asp39 (indicated by an arrow), which is located at the end of a pleated-sheet structure and belongs to the active site. The amino acids Asn5, Glu50, Asn163, Asn276 and some glycines form short stretches of left-handed α-helix (α_L) which are stabilized by strong hydrogen bonds involving the polar side chains.

The torsion angles ω of the peptide planes have a mean deviation of 4·2° with respect to the ideal value of 180°. This is slightly larger than observed for other protein structures but should correspond more closely to reality since the restraints on the peptide plane were reduced from five to four atoms at the end of the refinement. There is one *cis*-proline, Pro171, all other prolines are in the more common *trans* form.

The isotropic temperature factors will be presented in detail in a subsequent publication. The B values are in the range 5 to 30 Å^2 for protein atoms and up to 60 Å^2 for water molecules. The final mean B value for all atoms is 11·1 Å^2, in close agreement with the 12·2 Å^2 deduced from the Wilson statistics.

In comparison with other protein structures the temperature factors are on the lower side which we correlate with the unusually tight network of hydrogen

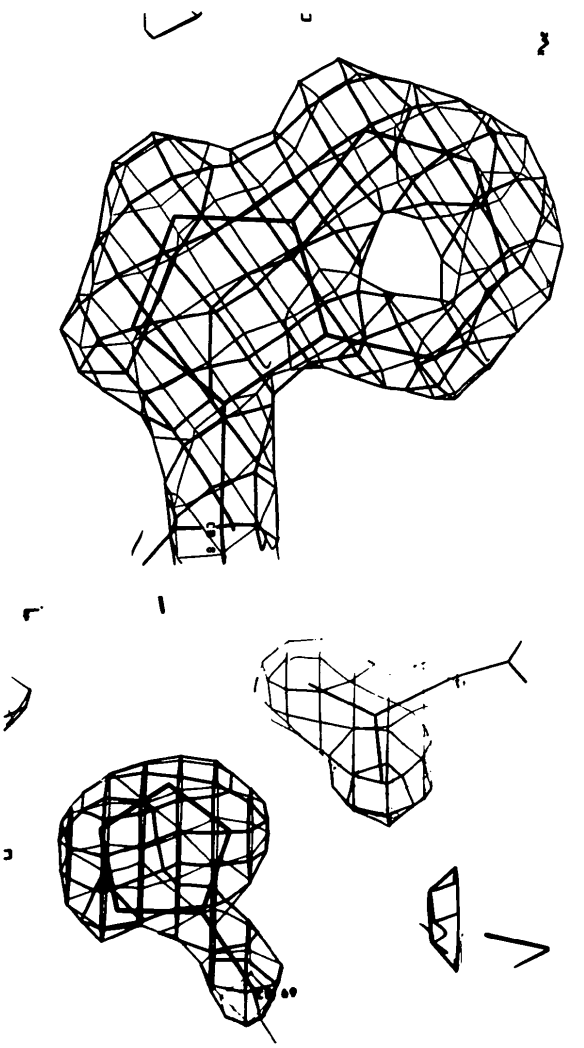

Fig. 3. Two sections of the final electron density map calculated at 1·5 Å resolution, showing (top) the indole ring of Trp8 and (bottom) the active-site residues His69, Asp39. The two pictures are drawn at different scales.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2PRK, R2PRKSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37022 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 31 December 1991.

† During a later check of the water structure, we identified an additional tightly bound Ca^{2+} ion based on the Ca^{2+}—O distances. This ion was previously thought to be a solvent water.

bonds which stabilize the secondary and tertiary structure of proteinase K, see Fig. 5. The B values are low in helical and pleated-sheet regions and are higher, corresponding to more flexibility, in regions more exposed at the surface of the molecule.

An upper limit for the average coordinate error can be assessed according to Luzatti (1952) by plotting the R factor as a function of resolution, see Fig. 6. Except for the low-resolution range, the data points follow satisfactorily the line drawn for a mean coordinate error of 0·15 Å. However, since the Luzatti plot overestimates the error in the model because it assumes that the disagreement between F_o and F_c is only caused by coordinate errors, we believe that the actual mean coordinate error in the refined proteinase K model is no more than 0·12 Å.

The initial 266 α-C positions based on the 3·3 Å MIR electron density were compared with the 279 α-C positions obtained after the final refinement cycle according to the least-squares method of Rossmann & Argos (1975). 39 α-C positions were beyond a cut-off value of 8 Å owing to erroneous chain tracing in regions where the electron density at low resolution was not well defined. If the 227 α-C atoms in comparable positions are considered, the mean deviation is 1·72 Å. The actual difference will be larger because the method optimizes the fit between the two coordinate sets.

The difference in mean phase angles $\langle |\alpha_{MIR} - \alpha_c| \rangle$, calculated on the basis of the refined atomic coordinates (α_c) and derived according to the MIR procedure

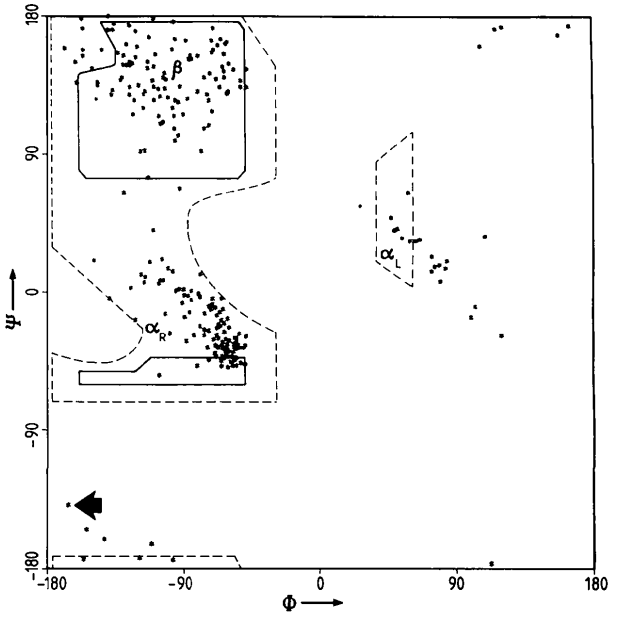

Fig. 4. Ramachandran (Φ,Ψ) plot for proteinase K. Torsion-angle regions typical for right-handed α-helix, β-pleated sheet and left-handed α-helix are indicated by α_R, β, α_L. The dashed lines contour the allowed regions (Ramakrishnan & Ramachandran, 1965). Glycines and Asp39 (indicated by arrow) are outside these regions.

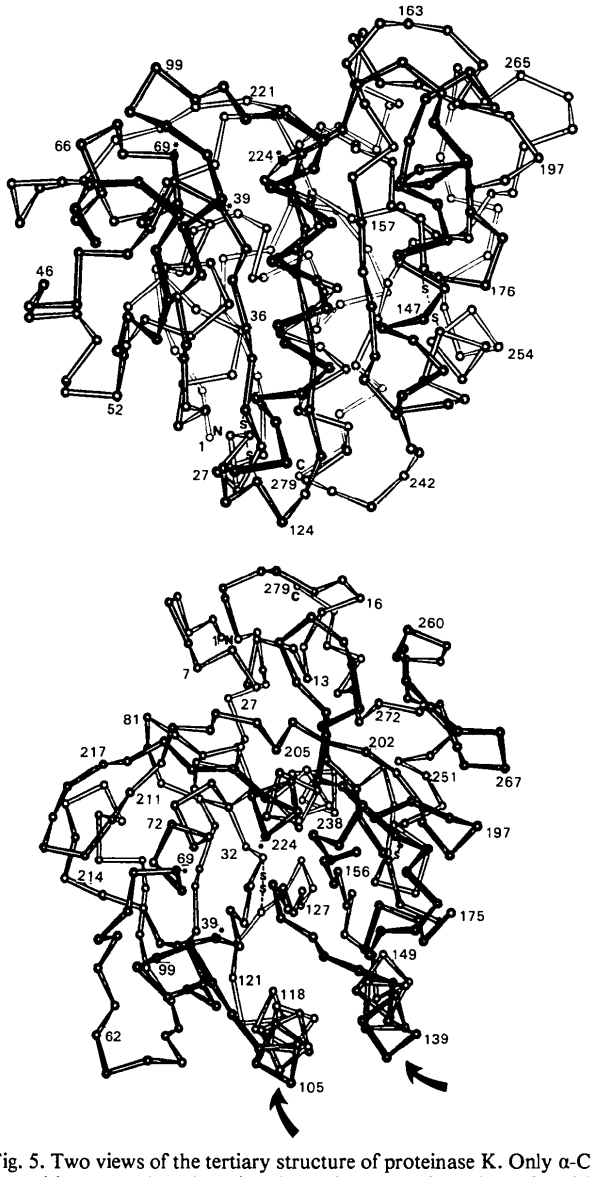

Fig. 5. Two views of the tertiary structure of proteinase K. Only α-C positions are plotted, active-site residues Asp39, His69, Ser224 are marked by asterisks and the two disulfide bridges are indicated by S—S. The arrows point to the two peripheral α-helices which interact intermolecularly in the crystals.

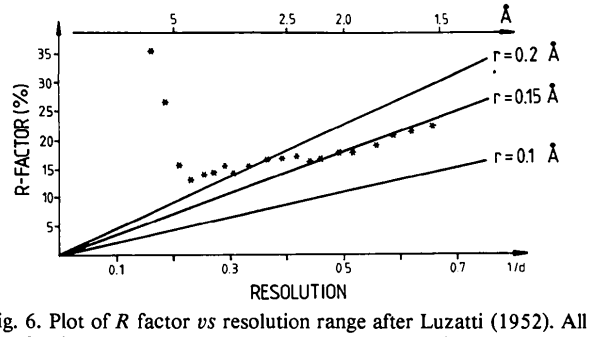

Fig. 6. Plot of R factor *vs* resolution range after Luzatti (1952). All reflections $F > 3\sigma_F$ were used, the atom-coordinate error r is below 0·15 Å, see text.

(α_{MIR}), reduced in the resolution range 100 to 2·8 Å from a starting root-mean-square value of 68 to 44° at the end of the refinement. The mean error in phase angles calculated on the basis of the 'figure of merit' of 84·6% was 32° (Pähler *et al.*, 1984), *i.e.* estimated 12° better than actually found. Nevertheless, the MIR phase angles were well determined because the phase angles calculated from the coordinates of the structure model approached the MIR phases during the course of refinement.

Crystal packing

The packing of the proteinase K molecules in the unit cell is illustrated by the stereo diagram, Fig. 7. The arrangement is such that two molecules occupy opposite corners on the diagonal of the *ab* plane giving rise to a string of molecules oriented along the [110] diagonal line. The 4_3 screw operation produces the next string rotated −90° and shifted by $\frac{1}{4}$ in *z etc.* This kind of packing gives rise to hexagonally close-packed sheets parallel to the (112) plane which, according to Fig. 7, are occupied more densely than the space between these planes.

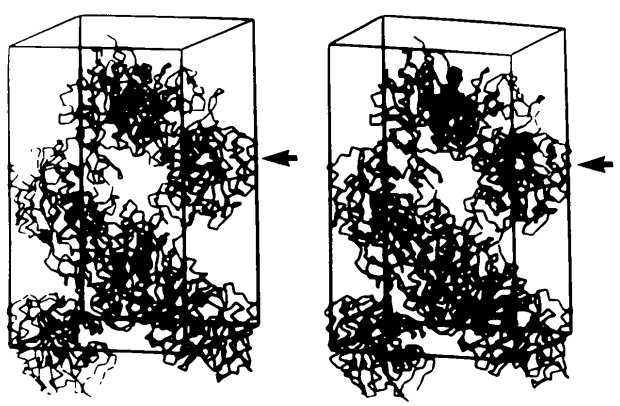

Fig. 7. Stereo plot of the molecular packing in the crystal unit cell. The arrows point to the two α-helices which are engaged in the dimer formation.

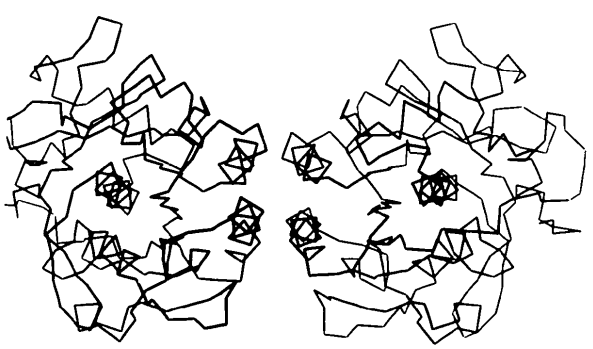

Fig. 8. 'Dimer' formed by proteinase K molecules in the crystal lattice. The α-helices forming the contact area are indicated by arrows in Fig. 5 and 7.

Table 5. *Hydrogen-bonding contacts between symmetry-related proteinase K molecules*

Group in (*x*, *y*, *z*)		Symmetry-related group		Distance (Å)	Symmetry element*
Arg121	NH1	Ala279	C=O	2·99	4
Arg188	NH1	Gly241	C=O	2·97	4
Arg218	NH1	Arg167	C=O	3·13	8
Gln3	OE1	Tyr61	O–H	3·35	3
Gln3	OE1	Tyr61	O–H	3·37	4
Tyr24	OH	Asn122	OD1	3·14	7
Asn163	ND2	Asp254	OD2	3·10	4
Asn163	ND2	Asp254	OD1	3·30	3
Asn168	OD1	Arg218	NH1	3·01	4

* Symmetry operations: (1) *x*, *y*, *z*; (2) $-x$, $-y$, $\frac{1}{2}+z$; (3) $\frac{1}{2}-y$, $\frac{1}{2}+x$, $\frac{3}{4}+z$; (4) $\frac{1}{2}+y$, $\frac{1}{2}-x$, $\frac{1}{4}+z$; (5) *y*, *x*, *z*; (6) $-y$, $-x$, $\frac{1}{2}-z$; (7) $\frac{1}{2}-x$, $\frac{1}{2}+y$, $\frac{3}{4}-z$; (8) $\frac{1}{2}+x$, $\frac{1}{2}-y$, $\frac{1}{4}-z$.

Within the strings, the most prominent packing motif is a 'dimer' formation in which the peripheral α-helices (residues 103 to 121 and 138 to 153; see Fig. 5) of two molecules related by the diad in [1$\bar{1}$0] are packed against each other, see Fig. 8. In spite of this close packing, however, there are no direct hydrogen-bonding contacts between these helices.

The intermolecular hydrogen bonds formed by the proteinase K molecules are given in Table 5. Two of these contacts are close together, NH1 Arg121–O Ala279 and OD1 Asn122–OH Tyr24, all the others are more evenly distributed over the surface of the molecules. Two of the side chains form multiple contacts with symmetry-related molecules, *viz* OD1 Asp254–ND2 Asn163 (symmetry 4) and OD2 Asp254–ND2 Asn163 (symmetry 3); NH1 Arg218–OD1 Asn168 (symmetry 4) and –O Arg167 (symmetry 8). All the symmetry operations between hydrogen-bonded proteinase K molecules involve only screw rotations, and no pure rotations or translations.

If crystals of proteinase K are soaked with substrate, it is turned over, and if they are soaked with small molecular inhibitors, the active site is blocked (Betzel *et al.*, 1986). This observation suggests that the active site of the enzyme in the crystalline form is freely accessible. In fact, it opens to the wide solvent channels, as illustrated in Fig. 7.

Concluding remarks

One of the main results of this study is that we were able to collect three-dimensional X-ray diffraction data of high quality at the synchrotron which in practice would not have been accessible with conventional rotating-anode X-ray generators. This illustrates that synchrotron radiation is an ideal source for the collection of data extending to the resolution limit of a (protein) crystal.

The refined structure model of proteinase K combines strictly idealized molecular geometry with good agreement of observed and calculated structure factors to a resolution of 1·5 Å. This renders proteinase K one

of the best presently available refined native enzymes of the subtilisin family and the free cysteine Cys72 suggests, in fact, that it is a member of a subgroup of the subtilisin family (Jany, Lederer & Mayer, 1986; Betzel *et al.*, 1986). It serves as a starting point for crystallographic studies with synthetic dipeptide inhibitors of the chloroketone type (Betzel *et al.*, 1986; Betzel, Wilson, Bellemann, Pal, Bajorath & Saenger, 1988) and work is in progress to solve the structure of the homologous enzyme mesentericopeptidase and thermitase by using molecular replacement, based on the molecular structure of proteinase K (Wilson, 1987, private communication).

We are most grateful to Dr N. Hennrich and Dr H. D. Orth (Merck, Darmstadt), for providing pure samples of proteinase K, and to Dr K. Bartels and Dr H.-D. Bartunik (EMBL/DESY) for providing us with the facilities to collect the synchrotron data. We thank Professor K. D. Jany for making available the amino acid sequence of proteinase K prior to publication and Professor Gassen for communicating changes of the sequence based on the nucleotide sequence of the proteinase K gene. This study has been funded by the German Federal Minister for Research and Technology (BMFT) under contract No. FKZ: B72C050, by Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 9, Teilprojekt A 7), and by Fonds der Chemischen Industrie.

References

ARNDT, U. W. & WONACOTT, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

BARTUNIK, H. D., GEHRMANN, T. & ROBRAHN, B. (1984). *J. Appl. Cryst.* **17**, 120.

BETZEL, CH., PAL, G. P., STRUCK, M.-M., JANY, K. D. & SAENGER, W. (1986). *FEBS Lett.* **197**, 105–110.

BETZEL, CH., WILSON, K. S., BELLEMANN, M., PAL, G. P., BAJORATH, J. & SAENGER, W. (1988). *Proteins*. Submitted.

BHAT, T. N. & COHEN, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

COX, E. G. & SHAW, W. F. B. (1930). *Proc. R. Soc. London*, **127**, 71.

DATTAGUPTA, J. K., FUJIWARA, T., GRISHIN, E. V., LINDNER, K., MANOR, P. C., PIENIAZEK, N. J., SAENGER, W. & SUCK, D. (1975). *J. Mol. Biol.* **97**, 267–271.

EBELING, W., HENNRICH, N., KLOCHOW, M., METZ, H., ORTH, H. D. & LANG, H. (1974). *Eur. J. Biochem.* **47**, 91–97.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

FRENCH, S. & WILSON, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

HENDRICKSON, W. A. & KONNERT, J. H. (1981). In *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. SRINIVASAN, Vol. 1, pp. 43–57. Oxford: Pergamon Press.

HENDRICKSON, W. A. & LATTMAN, L. E. (1970). *Acta Cryst.* B**26**, 136–142.

HIGUCHI, Y., KUSANOKI, M., MATSUURA, Y., YASUOKA, N. & KAKUDO, M. (1984). *J. Mol. Biol.* **172**, 109–139.

HOLMES, M. A. & MATTHEWS, B. W. (1982). *J. Mol. Biol.* **160**, 623–639.

JANY, K. D., LEDERER, G. & MAYER, B. (1986). *FEBS Lett.* **199**, 139–144.

JANY, K. D. & MAYER, B. (1985). *Hoppe-Seyler's Z. Physiol. Chem.* **366**, 485–492.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KAHN, R., FOURME, R., GADET, A., JANIN, J., DUMAS, C. & ANDRE, D. (1982). *J. Appl. Cryst.* **15**, 330–337.

KRAUS, E., KILTZ, H. H. & FEMFERT, U. E. (1976). *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 233–237.

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

MACHIN, P. A., WONACOTT, A. J. & MOSS, D. (1983). *Daresbury Lab. News*, **10**, 3–9.

MACHIN, P. A., WONACOTT, A. J. & MOSS, D. (1984). *Daresbury Lab. News*, **13**, 17–19.

PÄHLER, A., BANERJEE, A., DATTAGUPTA, J. K., FUJIWARA, T., LINDNER, K., PAL, G. P., SUCK, D., WEBER, G. & SAENGER, W. (1984). *EMBO J.* **3**, 1311–1314.

PAL, G. P., BETZEL, CH., JANY, K. D. & SAENGER, W. (1986). *FEBS Lett.* **197**, 111–113.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

WHITTAKER, E. J. W. (1953). *Acta Cryst.* **6**, 218.

WIEGERS, U. & HILZ, H. (1981). *Biochem. Biophys. Res. Commun.* **44**, 513–519.

WILSON, K. & YEATES, D. (1979). *Acta Cryst.* A**38**, 146–157.

Acta Cryst. (1988). B**44**, 172–178

The Use of Pseudosymmetry in the Rotation Function of γIVa-Crystallin

BY H. E. WHITE, H. P. C. DRIESSEN, C. SLINGSBY, D. S. MOSS, W. G. TURNELL AND P. F. LINDLEY

Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

(*Received* 20 *April* 1987; *accepted* 28 *October* 1987)

Abstract

Bovine lens γIVa-crystallin crystallizes in space group $C222_1$ with cell dimensions $a = 35{\cdot}1$, $b = 46{\cdot}2$, $c = 186{\cdot}2$ Å, and contains one molecule in the asymmetric unit. The structure was determined at 3·0 Å resolution using cross-rotation functions and R-factor searches with the bovine lens protein γII-crystallin as the model structure. The rotation function appears to be very sensitive to the resolution range and type of coefficient employed; the use of normalized structure-factor amplitudes gave the best results. The

0108-7681/88/020172-07$03.00