Volume 312, number 2,3, 127–131 FEBS 11667 November 1992

Three-dimensional structure of catalase from *Micrococcus lysodeikticus* at 1.5 Å resolution

G.N. Murshudov[a], W.R. Melik-Adamyan[a], A.I. Grebenko[a], V.V. Barynin[a], A.A. Vagin[a], B.K. Vainshtein[a], Z. Dauter[b] and K.S. Wilson[b]

[a]*Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, Moscow 117333, Russia* and [b]*European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, Hamburg 52, Germany*

Received 28 July 1992; revised version received 7 September 1992

The three-dimensional crystal structure of catalase from *Micrococcus lysodeikticus* has been solved by multiple isomorphous replacement and refined at 1.5 Å resolution. The subunit of the tetrameric molecule of 222 symmetry consists of a single polypeptide chain of about 500 amino acid residues and one haem group. The crystals belong to space group $P4_22_12$ with unit cell parameters $a = b = 106.7$ Å, $c = 106.3$ Å, and there is one subunit of the tetramer per asymmetric unit. The amino acid sequence has been tentatively determined by computer graphics model building and comparison with the known three-dimensional structure of beef liver catalase and sequences of several other catalases. The atomic model has been refined by Hendrickson and Konnert's least-squares minimisation against 94,315 reflections between 8 Å and 1.5 Å. The final model consists of 3,977 non-hydrogen atoms of the protein and haem group, 426 water molecules and one sulphate ion. The secondary and tertiary structures of the bacterial catalase have been analyzed and a comparison with the structure of beef liver catalase has been made.

Catalase; X-Ray structure; *Micrococcus lysodeikticus*

1. INTRODUCTION

Catalase (EC 1.11.1.6) is an enzyme with a molecular mass of 230–300 kDa that is present in the cells of all aerobic organisms. It decomposes hydrogen peroxide to molecular oxygen and water:

$$2H_2O_2 \rightarrow 2H_2O + O_2$$

The properties of catalase have been reviewed by Deisseroth and Dounce [1] and Schonbaum and Chance [2]. The most intensely investigated catalases are the tetrameric ones with a haem group in the active site of each subunit. Crystals of catalases from a variety of sources have been described and some of them have been used for structure investigation by X-ray diffraction.

To date, the three-dimensional structures of two catalases have been determined and refined: beef liver catalase (BLC) at 2.5 Å resolution [3] and that of the fungus, *Penicillium vitale* (PVC), at 2.0 Å resolution [4,5]. Previously we have investigated the polypeptide chain folding in two bacterial catalases, the non-haem hexameric Mn-catalase from *Thermus thermophilus* with a different subunit organization [6], and the tetrameric haem-containing catalase from *Micrococcus lysodeikticus* (MLC) with spatial organization closely resembling that of BLC [7,8]. Preliminary crystallographic results have been reported for two other haem-containing bacterial catalases, from *E. coli* [9] and from *Proteus mirabilis* PR [10]. Comparison of the three-dimensional structures of the catalases from evolutionarily distant sources suggests regions of the enzyme molecule that are most important for activity, folding and quaternary interactions. Here we report the structure of MLC refined at 1.5 Å resolution.

Correspondence address: W.R. Melik-Adamyan, Institute of Crystallography, Russian Academy of Sciences, Leninsky pr. 59, 117333 Moscow, Russia. Fax: (7) (095) 135 1011.

2. MATERIALS AND METHODS

2.1. *Crystallization and preparation of heavy atom derivatives*

Catalase from *Micrococcus lysodeikticus* (MLC) was prepared as described by Herbert and Pinsent [11]. Crystals of MLC were obtained at room temperature by hanging-drop vapour diffusion using ammonium sulphate as precipitant. The most suitable crystals for X-ray crystallography were obtained in 15–18 μl droplets of solution containing 20–30 mg/ml MLC, 0.6–0.8 M ammonium sulphate and 0.05 M sodium acetate buffer at pH 5.2. The equilibrating solution consisted of 0.6 ml of 1.2–1.4 M ammonium sulphate buffered to the same pH. Crystals of dimensions 0.5–1.0 mm grew after one week.

X-Ray studies of the MLC crystals showed that they belonged to the tetragonal space group, $P4_22_12$, with unit cell dimensions $a = b = 106.7$ Å, $c = 106.3$ Å, and that there was one subunit of the tetramer per asymmetric unit. This was in agreement with results obtained previously by Marie et al. [12]. MLC crystals were stable in the conventional X-ray beam for about 100 h and diffracted up to 1.5 Å resolution.

Heavy atom derivatives were obtained by soaking native protein crystals in solutions of heavy atom compounds in the above precipitating liquor at concentrations of 0.002–0.005 M for 1–2 weeks.

2.2. *X-Ray data collection*

X-Ray diffraction data from crystals of native protein and isomorphous heavy atom derivatives were collected up to 2.5 Å resolution on a KARD-4 area detector diffractometer [13]. High resolution data to 1.5 Å were collected from a native crystal using the EMBL synchrotron radiation beam-line X31 at the DORIS storage ring, DESY, Hamburg, with an image plate as detector. This data set was processed using the MOSFLM program package [14]. Each data set was collected from a single crystal. A summary of the data collection is given in Table I.

2.3. *Phasing*

Initial phases for the protein structure factors up to 2.5 Å resolution were determined by multiple isomorphous replacement (MIR). The major heavy atom binding site was determined for each derivative independently from the Harker sections of the difference Patterson synthesis. The positional parameters of the heavy atoms were refined [15] and phase information from the various derivatives was combined using Hendrickson and Lattman coefficients [16]. Phases obtained from these coefficients were used for calculation of difference Fourier maps to find minor sites which were used in the next refinement cycle. This procedure was repeated several times and gave a final mean figure of merit for the MIR phases of 0.64. Final statistics for the heavy atom refinement are listed in Table II. For further improvement of the phases, solvent flattening [17,18] was used. After 5 cycles of this procedure the best electron density map was calculated and used for interpretation.

2.4. *Model building and refinement*

A partial model was built using the interactive graphics program, FRODO [19], implemented on a Tektronix-4129 by D. Vassylyev [20]. The quality of the electron density map was quite good and allowed the tracing of the polypeptide chain and the fitting of some side chains. The sequence of MLC was unknown and at this stage the knowledge of sequences of five other catalases from different species [21] was very useful. The initial model included 481 amino acid residues with 3,372 non-hydrogen atoms and one haem group per subunit. This model was refined by the Hendrickson and Konnert stereochemically restrained least-squares minimization procedure [22]. The initial crystallographic R-factor was 0.395 in the resolution range from 5 to 3 Å. After 25 cycles of refinement with progressive extension of resolution to 2.5 Å (the limit of the isomorphous phases) the R-factor fell to 0.264. During this refinement, only positional atomic coordinates and the overall scale and temperature factor were refined.

Phases calculated from this partially refined model were combined with MIR phases and the 'best' electron density map calculated. The revised model was refined using restrained least-squares but now with individual atomic temperature factors and using FFT subroutines in the structure factor and gradient calculation [23], as implemented by us for all space groups except cubic (unpublished results). After 125 refinement cycles with gradual extension of resolution from 2.5 to 1.5 Å, five manual rebuildings and addition of water molecules the R-factor dropped to 0.150 for all 94,315 reflections in the resolution range 8–1.5 Å. Water molecules were located using the (F_o-F_c) difference map. Peaks in this map were assigned as water molecules if they were situated in a position to give at least one hydrogen bonding contact to a protein atom.

The secondary structure of the subunit was determined by the algorithm of Kabsch and Sander [24] using the program DSSP.

Table I

Statistics of data sets

Crystal	Resolution (A)	Number of reflections		R_{merge} (%)	R_{si} (%)
		Measured	Unique		
Native-1	2.5	122,125	20,868	8.7	6.3
Native-2	1.5	362,868	95,355	7.4	6.1
MMA	2.5	177,879	20,732	9.5	5.5
$UO_2(NO_3)_2$	2.5	137,727	20,576	10.7	7.2
$KAu(CN)_2$	2.5	124,501	21,386	10.5	9.1

$R_{merge} = \Sigma\Sigma|I_{h,i}-\langle I\rangle_h|/\Sigma\langle I\rangle_h$, where $\langle I\rangle_h = \Sigma I_{h,i}/N_h$.
$R_{si} = \Sigma\Sigma\sigma_{h,i}/\Sigma\Sigma I_{h,i}$, where h = (h,k,l), N_h is the number of measured equivalents, and i = 1, N_h.
$I_{h,i}$ and $\sigma_{h,i}$ are the measured intensity and its standard deviation.

3. RESULTS AND DISCUSSION

The tetrameric molecule of MLC with 222 symmetry has the dimensions, 66 × 90 × 93 Å. Each subunit of MLC contains one polypeptide chain and one haem group (Fig. 1a). The refined atomic model of the subunit consists of 3,977 non-hydrogen atoms of protein and haem, 426 water molecules and one sulphate ion. The mean value of the isotropic temperature factor for the protein atoms is 10.0 Å^2, and for solvent atoms is 23.6 Å^2. The final refinement statistics are shown in Table III. The mean error in the atomic coordinates estimated as by Luzzati [25] is 0.10–0.15 Å

An 'X-ray sequence' of MLC was determined from the 1.5 Å resolution electron density. During refinement and model rebuilding it was possible to locate 497 amino acid residues in the MLC subunit. All residues with aromatic side chains were clearly determined and residues situated in the inner parts of the molecule could be defined easily. However, ambiguity remains in such residues as Asn or Asp, Gln or Glu and Val or Thr. In addition there are some doubts in several side chains situated on the surface of the molecule with high solvent accessibility and high temperature factors. In the final electron density map there was some unexplained electron density at the N-terminal region of the polypeptide chain, and this is the reason why number 3 was assigned to the first residue.

A stereo plot of the MLC monomer is shown in Fig. 1b. The polypeptide chain of the MLC subunit forms

Table II

Derivatives and phase refinement parameters

Compound	Resolution (Å)	R_c (%)	R_s (%)	Phasing power	No. of sites	Figure of merit
MMA	2.5	11.4	64.5	1.7	4	0.37
$UO_2(NO_3)_2$	2.5	12.4	62.7	1.8	3	0.39
$KAu(CN)_2$	2.5	14.9	68.7	1.5	1	0.36
Overall figure of merit						0.64

$R_c = \Sigma||Fp_o| - |Fph_o||/\Sigma|F_o|$, $R_s = \Sigma||Fph_o| - |Fph_c||/\Sigma||Fph_o| - |Fp_o||$, and phasing power = f_h/E.
Fp_o and Fph_o are observed structure factors for the native and heavy atom derivatives, respectively; Fph_c is the calculated structure factor for a derivative; f_h is the root mean square contribution of the heavy atom; and E is the rms lack of closure error.

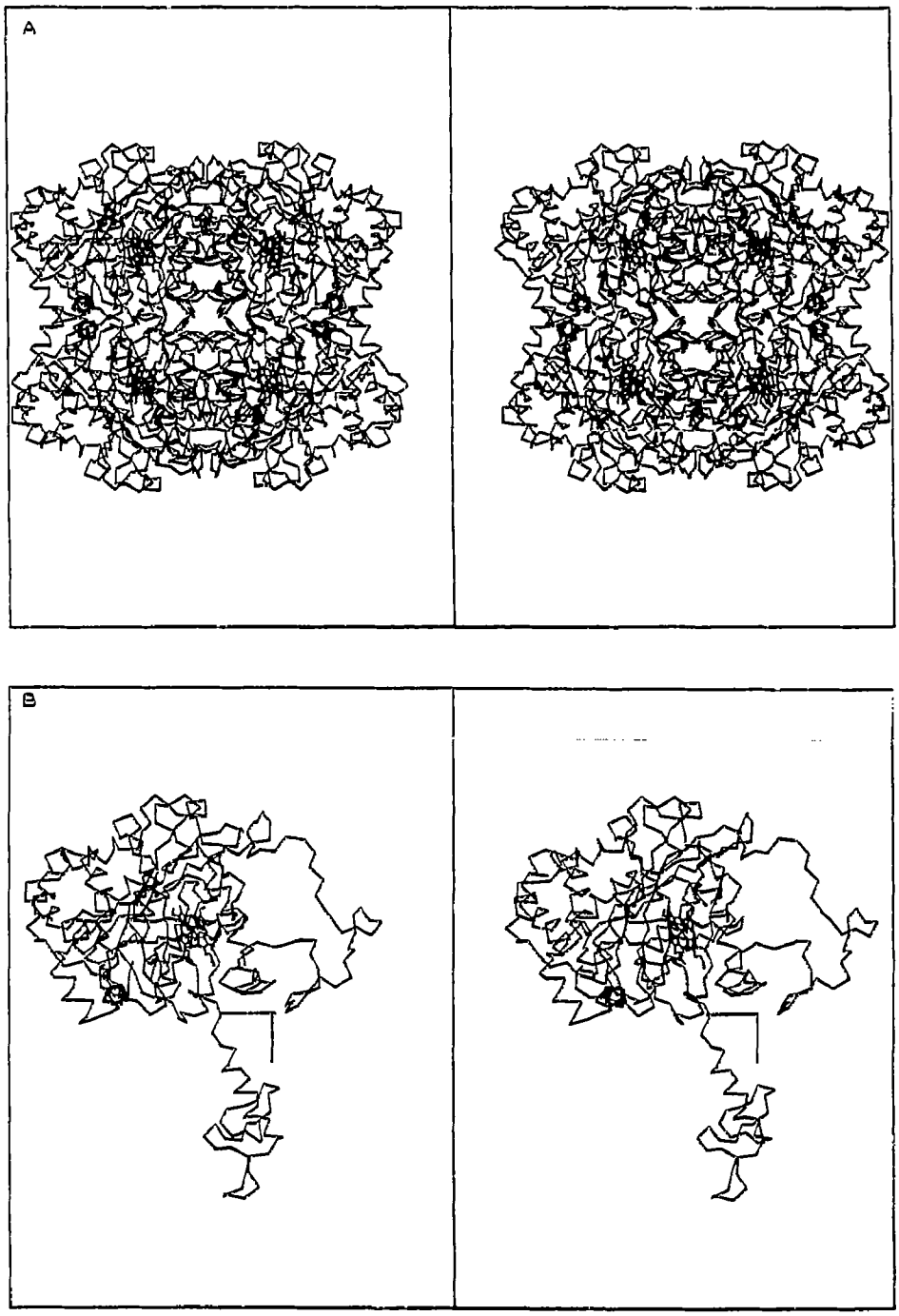

Fig. 1. A stereo plot of the C_α atoms and haem group for (a) the whole molecule and (b) one subunit. The orientations of molecule and subunit are the same. P, Q and R are molecular symmetry axes.

two domains of different length and organization. The large, haem-containing (α+β) domain is composed of 290 residues (59–348) and contains a β-barrel of 8 antiparallel β-strands and 8 α-helices. The second smaller domain consists of 76 residues (422–497) and contains 4 α-helices (422–497). The N-terminal part of the subunit consists of 56 residues (3–58) and contains one α-helix. This part of the structure is entirely buried in a neighbouring subunit within the tetramer. The first (α+β) domain and the four helix domain are connected by a long polypeptide loop which wraps around the surface of the molecule. There are 73 residues (349–421) in this segment and the first part of it is involved in many intersubunit contacts. Analysis of hydrogen

Table III

Summary of refinement statistics

Parameters	Target (σ)	Standard deviation	Number of parameters
R-factor (%)		15.01	94,315
Bond distances (Å)			
Bond distances (1–2 neighbours)	0.020	0.012	4,095
Angle distances (1–3 neighbours)	0.040	0.032	5,578
Planar distances (1–4 neighbours)	0.050	0.062	1,501
Planar groups (Å)	0.020	0.016	3,667
Chiral volumes (Å^3)	0.150	0.143	570
Non-bonded contacts (Å)			
Single torsion contacts	0.250	0.170	1,106
Multiple torsion contacts	0.250	0.173	929
Possible hydrogen bonds	0.250	0.163	221
Torsion angles (°)			
Planar	4.0	3.6	528
Staggered (± 60,180)	15.0	12.7	644
Orthonormal (± 90)	20.0	31.2	71

σ is the inverse of the square root of the least-squares weight used for the parameters.
R – factor = $\Sigma||F_o|-F_c||\Sigma|F_o|$

bonds shows that there are 7 small part of a polypeptide chain identified as 3_{10} helices. Three of them are situated at the beginning or end of α-helices.

Comparison of the three-dimensional structures of MLC and BLC (coordinates of BLC were taken from the Protein Data Bank at the Brookhaven National Laboratory [26]) shows that the overall spatial organisation of MLC and BLC is very similar. The main differences between the two molecules are in the N-terminal part, where the first helix of BLC is absent in MLC and in the connection segment.

The haem group in these two catalases lies in the same position and orientation. The haem in MLC has a curved conformation as in BLC. The deviation of the iron atom from the haem plane (NA-NB-NC-ND plane) is 0.12 Å. The distance between the iron atom and OH oxygen of the proximal Tyr-339 is 1.9 Å. The sixth coordination position of iron is occupied by a water molecule. The distance between the water oxygen and the iron atom is 2.28 Å. One more water molecule is situated in the haem pocket and forms hydrogen bonds with the iron coordinating water and with the distal histidine important for catalase activity (His-57).

BLC tightly binds NADPH [3]. In the equivalent position in the electron density map of MLC at 3 Å resolution there was some electron density which allowed us to suppose that MLC also binds NADPH [7,8]. However, in this place in the (F_o-F_c) map at 1.5 Å resolution there were several separate peaks which were difficult to explain as a NADPH molecule. The highest electron density peak situated in the position of the NADPH

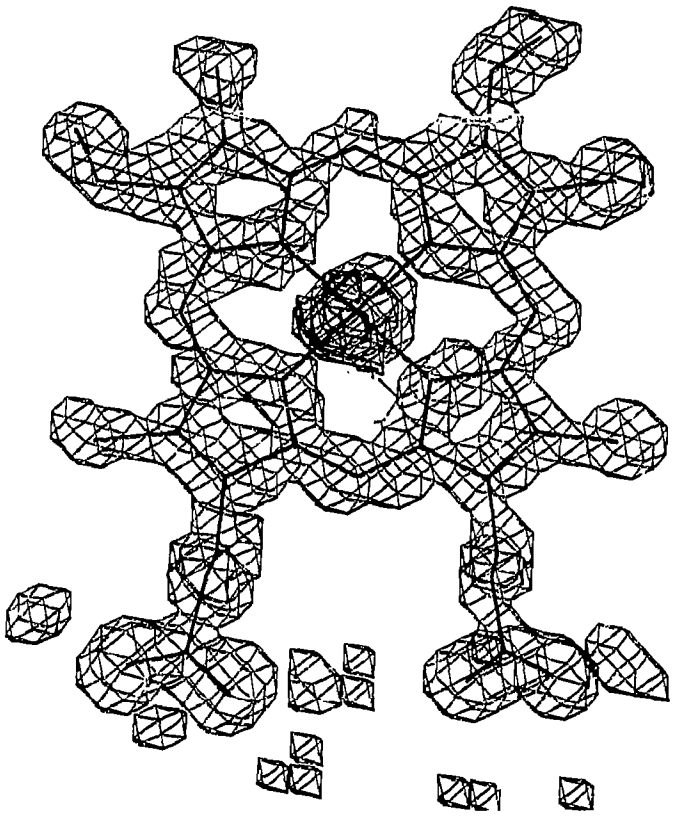

Fig. 2. The electron density map for the haem group.

phosphate group in BLC was assigned as a sulphate ion (Fig. 3) and other peaks were assigned as water. Nevertheless in the structure of MLC there is space for a NADPH molecule, and the possibility that MLC could bind NADPH cannot be excluded. No biochemical data were obtained about the presence of NADPH in MLC prior to crystallization.

The sequence of MLC determined during the X-ray

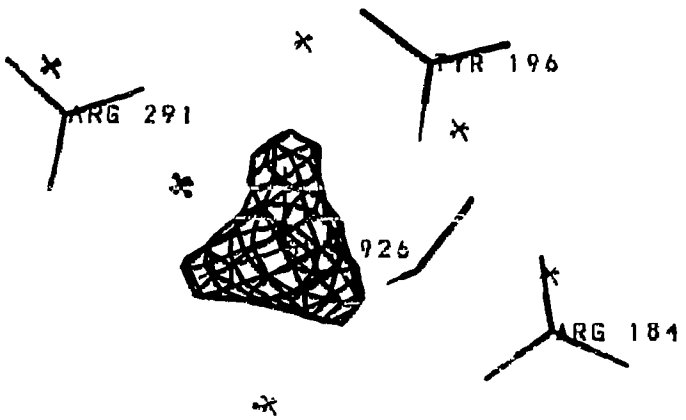

Fig. 3. The difference (F_o-F_c) electron density peak assigned as sulphate ion. The model used for F_c and ϕ_c calculation included a water molecule in this position.

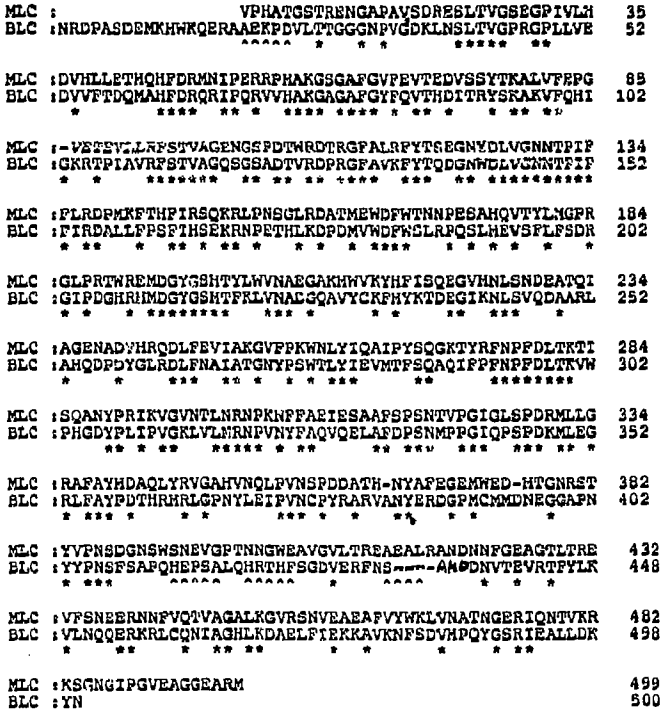

Fig. 4. X-Ray sequence of MLC aligned with the BLC sequence. *indicates identical residues,' indicates residues which are not equivalent (distances between C_α atoms more than 2.2 Å). Of 459 equivalent residues 204 are found to be identical.

study was aligned by a dynamic alignment algorithm [27] with the chemical sequence of BLC and corrected by three-dimensional superposition (Fig. 4). The main differences between the two catalases are at the ends of the polypeptide chain. The first domain is much more conserved than other parts of the molecule. There are more changes of amino acid residues in the second domain but the spatial similarity of these parts is nevertheless quite good. The rms deviation between C_α atoms for 459 equivalent residues in MLC and BLC is 1.0 Å. The residues were assumed to be equivalent if their C_α atom positions were within 2.2 Å after superposition of these structures.

The environment of the haem, including the active site, is very conserved except for a few residues. One of the most surprising differences between these two catalases is around the haem group, before the active distal His-57, (His-74 in BLC). Val-73 of BLC is replaced by Pro-56 which has the *cis* conformation, and the small hydrophobic Val-72 is replaced by the large charged residue Arg-55.

In spite of the fact that only the X-ray sequence of MLC is available at present the refined structure of MLC shows that cloning and sequencing of the MLC gene could make this catalase a good object for enzyme engineering experiments to check hypotheses on catalase activity and molecule assembly.

A more detailed analysis of the structure will be published later.

Acknowledgements: We thank Dr. A.A. Antson for assistance with high resolution data collection and processing, and Dr. D.V. Vassylyev for assistance with using the Tektronix version of FRODO.

REFERENCES

[1] Deisseroth, A. and Dounce, A.L. (1970) Physiol. Rev. 50, 319–375.
[2] Schonbaum, G.R. and Chance, B. (1976) Enzymes 13, 363–408.
[3] Fita, I., Silva, A.M., Murthy, M.R.N. and Rossmann, M.G. (1986) Acta Crystallogr. B42, 497–515.
[4] Vainshtein, B.K., Melik-Adamyan, W.R., Barynin, V.V., Vagin, A.A., Grebenko, A.I., Borisov, V.V., Bartels, K.S., Fita, I. and Rossmann, M.G. (1986) J. Mol. Biol. 188, 49–61.
[5] Melik-Adamyan, W.R., Barynin, V.V., Vagin, A.A., Borisov, V.V., Vainshtein, B.K., Fita, I., Murthy, M.R.N. and Rossmann, M.G. (1986) J. Mol. Biol. 188, 63–72.
[6] Vainshtein, B.K., Melik-Adamyan, W.R., Barynin, V.V., Vagin, A.A. and Grebenko, A.I. (1985) J. Biosci. 8, 471–479.
[7] Melik-Adamyan, W.R., Grebenko, A.I., Barynin, V.V., Yusifov, E.F., Vagin, A.A., Murshudov, G.N., Rubinski, C.V. and Vainshtein, B.K. (1989) Dokl. AN. SSSR 306, 620–623 (in Russian).
[8] Yusifov, E.F., Grebenko, A.I., Barynin, V.V., Murshudov, G.N., Vagin, A.A., Melik-Adamyan, W.R. and Vainshtein, B.K. (1989) Sov. Phys. Crystallogr. 34, 870–874.
[9] Tormo, J., Fita, I., Switala, J. and Loewen, P.C. (1990) J. Mol. Biol. 213, 219–220.
[10] Jouve, H.-M., Gouet, P., Boudjada, N., Buisson, G., Kahn, R. and Duee, R. (1991) J. Mol. Biol. 221, 1075–1077.
[11] Herbert, D. and Pinsent, J. (1948) J. Biochem. 43, 193–202.
[12] Marie, A.L., Parak, F. and Hoppe, W. (1979) J. Mol. Biol. 129, 675–676.
[13] Kheiker, D.M., Popov, A.N. and Andrianova, M.E. (1989) Methods of Structural Analysis, pp. 125–140, Moscow, Nauka (in Russian).
[14] Leslie, A.G.W., Brick, P. and Wonacott, A.J. (1986) CCP4 Newslett. 18, 33–39.
[15] Dickerson, R.E., Weinzierl, J.E. and Palmer, R.A. (1968) Acta Crystallogr. B24, 997–1003.
[16] Hendrickson, W.A. and Lattman, E.E. (1970) Acta. Crystallogr. B26, 136–143.
[17] Wang, B.C. (1985) Methods Enzymol. 115, 90–112.
[18] Leslie, A.G.W. (1988) Improving Protein Phases, pp. 25–31, Proceedings of the Daresbury Study Weekend.
[19] Jones, T.A. (1978) J. Appl. Crystallogr. 11, 268–272.
[20] Vassylyev, D.G. and Adzhubei, A.A. (1992) J. Mol. Graphics (in press).
[21] Murray, W.W. and Rachubinski, A.R. (1987) Gene 61, 401–413.
[22] Konnert, J.H. and Hendrickson, W.A. (1980) Acta Crystallogr. A36, 344–350.
[23] Agarwal, R.C. (1978) Acta. Crystallogr. A34, 791–809.
[24] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[25] Luzzatti, V. (1952) Acta Crystallogr. 5, 802–810.
[26] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Mayer Jr. E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Simanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[27] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 442–453.