Acta Cryst. (1995). D51, 575–589

Refined 3.2 Å Structure of Glycosomal Holo Glyceraldehyde Phosphate Dehydrogenase from *Trypanosoma brucei brucei*

BY FRED M. D. VELLIEUX*

BIOSON Research Institute, University of Groningen, Nijenborgh 16, 9747 *AG Groningen, The Netherlands*

JÁNOS HAJDU

Laboratory of Molecular Biophysics, Oxford University, South Parks Road, Oxford OX1 3QU, England

AND WIM G. J. HOL†

BIOSON Research Institute, University of Groningen, Nijenborgh 16, 9747 *AG Groningen, The Netherlands*

(*Received* 31 *August* 1994; *accepted* 1 *March* 1995)

Abstract

The three-dimensional crystal structure of the enzyme glyceraldehyde phosphate dehydrogenase from the kinetoplastid *Trypanosoma brucei brucei* has been determined at 3.2 Å resolution from a 37% complete data set collected using the Laue method. The crystals used in the structure determination contain one and a half tetrameric enzyme molecules in the asymmetric unit, *i.e.* six identical subunits. Initial phasing was carried out by the method of molecular replacement using the refined coordinates of holo glyceraldehyde phosphate dehydrogenase from *Bacillus stearothermophilus* as a search model. The initial electron-density distribution, obtained from the molecular-replacement solution, was greatly improved by a procedure consisting of 36 cycles of iterative non-crystallographic density averaging. During the averaging procedure, the missing reflections (63% of the data) were gradually introduced as map-inversion structure factors. At completion of the procedure, the *R* factor between averaged map-inversion amplitudes and observed structure-factor amplitudes was 19.0% for all data between 7.0 and 3.2 Å resolution, and that between the map-inversion amplitudes and later recorded structure-factor amplitudes was 41.9%. After model building into the resulting averaged electron-density map, refinement by molecular-dynamics procedures with *X-PLOR* provided the current model, which has an *R* factor of 17.6% for 34 835 reflections between 7.0 and 3.2 Å resolution. The refined model, comprising 2735 protein atoms plus one NAD^+ molecule and two sulfate ions per subunit, has r.ms. deviations from ideality of 0.02 Å for bond lengths and 3.6° for bond angles. All subunits, located either within the tetrameric molecule or within the half tetramer present in the asymmetric unit, are related to each other by almost exact twofold symmetry. The overall structure of the glycosomal glyceraldehyde phosphate dehydrogenase subunit and its quaternary arrangement in the tetrameric molecule are similar to that of the enzyme of lobster and *Bacillus stearothermophilus* (with r.m.s. differences between equivalent Cα positions of 0.71 and 0.64 Å, respectively). The main differences between the structures is the presence of three insertions, plus the substitution of a β-strand by a short α-helix, both occurring at the surface of the glycosomal enzyme subunit.

1. Introduction

D-Glyceraldehyde-3-phosphate dehydrogenase (E.C. 1.2.1.12, GAPDH) is a tetrameric enzyme composed of four identical subunits comprizing *ca* 330 residues each ($M_r \simeq 36$ kDa). It catalyzes the oxidative phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphosphoglycerate in the presence of inorganic phosphate, with the concomitant reduction of NAD^+. This is a key reaction in the glycolytic conversion of glucose to pyruvate. The enzymatic, kinetic and physicochemical properties of this enzyme are all well documented (see *e.g.*, Harris & Waters, 1976). To date, high-resolution structures of holo and apo GAPDH from *Bacillus stearothermophilus* have been reported (Skarżyński, Moody & Wonacott, 1987; Skarżyński & Wonacott, 1988), which have shed light on the type of molecular rearrangements taking place upon cofactor binding. Lower resolution models of the lobster enzyme have also been described (Moras *et al.*, 1975; Murthy, Garavito, Johnson & Rossman, 1980), a high-resolution study of the human enzyme is in progress (Read *et al.*, unpublished work), and the three-dimen-

* Author to whom all correspondence should be addressed. Current address: LCCP, IBS JP Ebel CEA CNRS, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 1, France.

† Current address: Departments of Biological Structure and Biochemistry, SM-20, Biomolecular Structure Program and HHMI, School of Medicine, University of Washington, Seattle, Washington 98195, USA.

sional structure of the *E. coli* enzyme has recently been determined (Olivier *et al.*, 1995).

From these studies, the generic architecture of this enzyme is well known: in all GAPDHases studied thus far, the four identical subunits are arranged with molecular 222 symmetry (Fig. 1). Each subunit is composed of two domains: the N-terminal or NAD-binding domain, and the C-terminal or catalytic domain. The NAD-binding domain (*ca* 150 residues) is folded into a central six-stranded β-sheet surrounded on both sides by helices. This topology, also known as the 'Rossmann fold', is homologous to that found in the dinucleotide-binding domains of other dehydrogenases (Rossmann, Liljas, Brändén & Banaszak, 1975). The catalytic domain (*ca* 180 residues) has an α/β topology containing an extended mixed β-sheet, and is concluded by a C-terminal helical segment which interacts with both domains. The active site is located in this latter domain and contains a cysteine and a histidine, both residues being essential for catalytic activity (Harris & Waters, 1976).

The unicellular kinetoplastid *Trypanosoma brucei* is the causating agent of sleeping sickness in humans and of nagana in cattle. In this parasite, an essential part of glycolysis takes place in a unique peroxisome-like organelle, the glycosome (Opperdoes & Borst, 1977), in which nine of the enzymes involved in glucose and glycerol metabolism are sequestered. This probably accounts for the extremely high rate of glycolysis observed in trypanosomes (Misset, Bos & Opperdoes, 1986). In its bloodstream form the trypanosome is entirely dependent upon glycolysis for ATP production. As a consequence, inhibition of the glycolytic pathway leads to the rapid death of the parasite (Fairlamb, Opperdoes & Borst, 1977). Thus, the enzymes of the glycolytic cascade have been selected as a potential target for selective protein structure-based inhibitor design (Hol, 1986; Opperdoes, 1987).

Glycosomal enzymes exhibit several peculiarities with respect to the cytosolic isoenzymes: eight out of the nine glycosomal enzymes have higher isoelectric points than the cytosolic counterparts, which is because of the presence in their amino-acid sequences of a marked excess of positively charged residues (Misset *et al.*, 1986). Also, glycosomal enzymes have higher subunit M_r than the cytosolic enzymes. This is because of the presence of N- or C-terminal extensions, and of specific insertions in the amino-acid sequences (Opperdoes *et al.*, 1990, and references therein). The glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) subunit from *T. brucei* comprises 358 residues giving a total M_r of 35.4 kDa, and has a sequence identity of *ca* 50% with other GAPDHases (Michels *et al.*, 1986).

In this paper, we wish to focus on several non-standard aspects of the three-dimensional (3-D) structure determination of this glycosomal enzyme. We also compare the *T. brucei* enzyme to the model of the holo-enzyme from *B. stearothermophilus* (Skarżyński *et al.*, 1987).

2. Experimental procedures

2.1. *Crystallization, data collection and processing*

Purified *Trypanosoma brucei brucei* gGAPDH (Misset *et al.*, 1986) was kindly provided by Dr F. R. Opperdoes and coworkers (ICP, Brussels). The crystals used in this study were grown using the previously described protocol (Read *et al.*, 1987): samples of 5 µl of the purified enzyme (6 mg ml^{-1}), dissolved in 25 m*M* Tris–HCl, pH 7.8, 0.6 *M* ammonium sulfate, 2 m*M* NAD^+, 1 m*M* dithiothreitol, 1 m*M* sodium azide, 1 m*M* EDTA, were mixed with the same volume of precipitant solution [52% saturated ammonium sulfate, 0.2 *M* *N*,*N*-bis(2-hydroxyethyl)-2-aminoethane sulfonic acid, pH 6.5, 1 m*M* dithiothreitol, 1 m*M* sodium azide, 1 m*M* EDTA, 20 m*M* 3-phosphoglycerate]. The resultant droplets were suspended above 1 ml of this precipitant solution in a hanging-drop setup. Although crystals grow quite readily, the shortage of purified protein material, together with the propensity of the enzyme to give several small-sized crystals per drop rather than only one or two larger crystals, meant that only few suitable

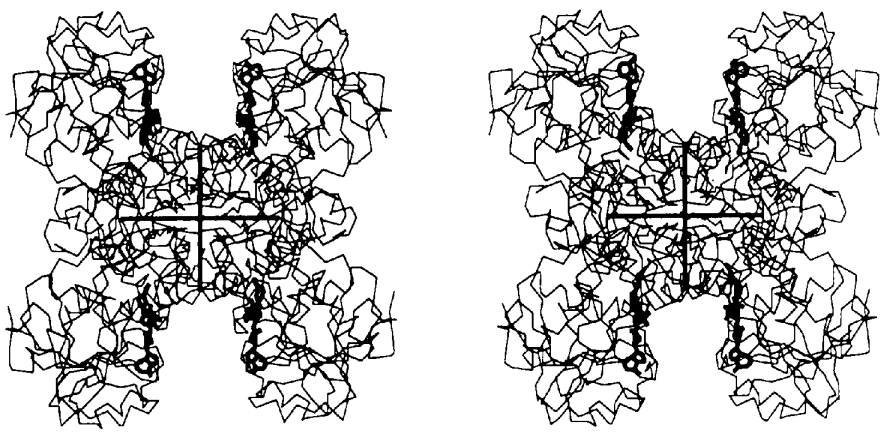


Fig. 1. The gGAPDH tetramer, viewed down the molecular *P* axis. The molecular twofold axes (*Q* and *R*) are shown in thick lines. The Cα tracings of the gGAPDH subunits are in thin lines, and the NAD cofactors are displayed in thick lines.

Table 1. *Laue data-collection statistics*

Data set	Resolution (Å)	No. of reflections	No. of independent reflections	% of unique data	R_{merge}* (%)
Native	26.24–3.20	35082	22226	33.4 (26.2–3.2 Å) 36.7 (7.0–3.2 Å)	13.5
Pentalenolactone	6.42–3.20	47022	26626	45.8 (6.5–3.2 Å)	12.6†
Combined	7.40–3.20	80481	34998	57.3 (7.4–3.2 Å) 58.0 (7.0–3.2 Å)	14.7

* $R_{merge} = \sum_i (I_i - \langle I \rangle) / \sum \langle I \rangle$.
† After anisotropic scaling, the pentalenolactone data set had an R factor of 12.6% *versus* the native structure-factor amplitudes.

crystals were available for our studies. Of the two crystal forms available (Read *et al.*, 1987), this study has focussed on form I gGAPDH crystals, of space group $P2_12_12_1$ with $a = 135.5$, $b = 256.3$ and $c = 114.9$ Å. Typical crystals used for data collection have size *ca* $0.6 \times 0.15 \times 0.1$ mm, and diffract initially to *ca* 2.3 Å resolution using monochromatic synchrotron radiation. However, their short lifetime in the X-ray beam would have resulted, with a monochromatic data collection strategy, in the use of 20–30 crystals to collect the 90° of oscillation which are required per data set. With the scarcity of available crystals, we have resorted to the use of the crystal-efficient Laue method with white synchrotron radiation to collect all X-ray data.

For the inhibitor-binding studies, three gGAPDH crystals were soaked in mother liquor containing 0.5 *M* of the inhibitor pentalenolactone (generously provided by Professor D. Mecke, Tübingen). Reaction of the inhibitor with the enzyme was monitored by following the colour change taking place in the crystals, which gradually turned from colourless to brown.

For data collection, three native gGAPDH crystals and, later, three crystals containing the inhibitor were exposed to the white radiation (wavelength range: 0.25–2.1 Å) at station 9.7 of the SRS (Daresbury, England). The effect of the spectral range on the quality of Laue data sets has been discussed (Clifton, Elder & Hajdu, 1991; Hajdu *et al.*, 1991; Sweet *et al.*, 1993). The Laue film packs collected from three of these crystals had to be discarded, due to the inadequate quality of the data from two crystals, and the presence, for the third crystal, of a third unknown crystal form. Thus, all data used for this work was collected from a total of three crystals, as described previously (Vellieux *et al.*, 1993): two native gGAPDH crystals were used for the native data set, and one crystal containing the inhibitor pentalenolactone for the additional data set.

All film packs were processed and merged using the Daresbury Laue processing software (Helliwell *et al.*, 1989) as described previously (Vellieux *et al.*, 1993). At the time the data was processed the software did not allow us to obtain absolute cell dimensions but only ratios of these. Since then, methodological and software development has allowed the determination of exact cell-parameter values from Laue data alone (Carr, Cruickshank & Harding, 1992; Carr, Dodd & Harding, 1993). In our case, the refined unit-cell dimensions were thus obtained starting from those measured by Read *et al.* (1987), with the shortest cell axis, the c axis, assumed to be known with highest precision from the rotation photographs and, therefore, kept constant. The native data set, consisting of 22 226 unique reflections, has an overall R_{merge} of 13.5%, where,

$$R_{merge} = \sum_{hkl} \left[\sum_i (I^i_{hkl} - \langle I_{hkl} \rangle) \right] \Big/ \left(\sum_{hkl} \langle I_{hkl} \rangle \right).$$

The resolution limits of this native data set are 26.2 and 3.2 Å, with most of the data between 7.0 and 3.2 Å, resulting in an overall completeness of 37% in this resolution range.

The pentalenolactone data set, consisting of 26 626 unique reflections between 26.2 and 3.2 Å resolution, has an R_{merge} value of 12.4% and is 45.8% complete in that resolution range. After refinement of the molecular model (*vide infra*), with the limited resolution of the data and the low completeness of the resulting data set, no clear density was observed which could be assigned to the pentalenolactone inhibitor. In order to help overcome the problems created by the limited completeness of the two individual data sets, a combined data set was also generated by merging all native and pentalenolactone film packs together. This combined data set (34 998 unique reflections between 7.4 and 3.2 Å, with an overall completeness of 57.3%) has $R_{merge} = 14.7\%$, and was used for the refinement of the gGAPDH structure. Data-collection statistics are summarized in Table 1.

2.2. *Rotation function*

Rotation-function calculations were carried out using the program *MERLOT* (Fitzgerald, 1988). Self-rotation functions, calculated with Tanaka's modification of Crowther's fast rotation function (Tanaka, 1977; Crowther, 1972) with native data between 7.0 and 3.2 Å resolution and an integration radius of 23 Å, gave inconclusive results: the only significant peak observed was the origin peak.

For the cross-rotation function calculations, the search model was the refined *B. stearothermophilus* GAPDH tetramer (Skarżyński *et al.*, 1987), whose sequence is 54% identical to that to *T. brucei* glycosomal GAPDH

Table 2. *Results of the cross-rotation function calculations*

Model *A*: the refined *B. stearothermophilus* GAPDH tetramer (Skarżyński *et al.*, 1987), aligned with its *P*, *Q*, *R* molecular twofold axes coincident with the *x*, *y*, *z* axes of the artificial orthogonal unit cell; model *B*: an arbitrary rotation of (20,20,20) was applied to model *A*; model *C*: one half of the *B. stearothermophilus* GAPDH tetramer, *i.e.* a dimer positioned in the same orientation as for model *A*.

	Eulerian angles (°)			Peak height
Search model	α	β	γ	(arbitrary units)
A	55.0	90.0	90.0	100.0
	22.5	90.0	90.0	76.6
	90.0	90.0	90.0	53.0
B	36.0	83.3	51.0	100.0
	2.5	83.5	50.6	77.7
	71.9	83.6	51.1	54.4
C	54.5	87.4	89.9	100.0
	18.3	85.9	89.9	84.1
	87.5	90.0	90.0	58.2

(Walker, Carne, Runswick, Brigden & Harris, 1980; Michels *et al.*, 1986). The NAD cofactors were removed from this model in order to obtain an unbiased electron density in this part of the molecule at various stages of the structure determination. The search model was oriented with its three mutually perpendicular *P*, *Q* and *R* molecular twofold axes coincident with the *x*, *y* and *z* axes of an artificial unit cell of dimensions 110.0 × 90.0 × 100.0 Å. The cross-rotation function was calculated with data from 6.0 to 4.0 Å and an integration radius in Patterson space of 23.0 Å, with the rotation space sampled with a grid spacing of 2.5° along α and 5.0° along β and γ. The map showed two prominent peaks at Eulerian angle values of (55, 90, 90°) and (20, 90, 90°), with respective peak heights of 100.0 and 76.6 (Table 2). The next highest peak (at 90, 90, 90°), observed with a height of 53.0, is inconsistent with space-group symmetry since it would place the *P* and *R* molecular twofold axes along the *X* and *Y* axes, which are twofold screw axes in space group $P2_12_12$. Attempts were made to test the correctness of the orientation parameters by repeating the calculations using either a search model with a different starting orientation, or by using as a search model half of a tetramer. All these tests gave results which were in accordance with the previously found solutions (Table 2), and also indicated the presence in the asymmetric unit of two distinct molecular orientations.

Eulerian angle values of (α, 90, 90°) indicate that the molecular *Q* axis is parallel to the *z* axis, which is a twofold axis in space group $P2_12_12$. In this case, the translation function is expected to contain two peaks related by exact twofold symmetry along the *z* axis. However, if the molecular twofold axis is only approximately parallel to the crystallographic twofold axis, the translation function will only exhibit pseudo twofold symmetry. In order to distinguish between these two possibilities, the rotation parameters for each molecular orientation were refined prior to the translation-function calculations. This was carried out using the program *BRUTE* (Fujinaga & Read, 1987) with 5.0 to 4.0 Å data, by searching for the maximum value of the correlation coefficient between calculated and observed intensities, in the vicinity of the two rotation function maxima {correlation coefficient, $C = [\sum(I_{o_1} - \langle I_o \rangle)] \times [\sum(I_c - \langle I_c \rangle)]/[\sum(I_o - \langle I_o \rangle)^2]^{\frac{1}{2}}[\sum(I_c - \langle I_c \rangle)^2]^{\frac{1}{2}}$}. Refined rotation parameters were (53.0, 89.18, 90.27°) and (19.0, 90.0, 90.0°), with correlation coefficients of 0.0987 and 0.0742, respectively. These results indicate that the asymmetric unit contains one gGAPDH molecule oriented with its *Q* axis approximately parallel to the crystallographic twofold axis, plus a second molecule oriented with its *Q* axis exactly parallel to the *z* axis.

Table 3. *Results of the translation function*

The search model used during these calculations was the refined *B. stearothermophilus* GAPDH tetramer (Skarżyński *et al.*, 1987).

	Molecular orientation (Eulerian angles α, β, γ, °)							
	(α, β, γ = 53.00, 89.18, 20.27)				(α, β, γ = 19.00, 90.00, 90.00)			
Harker section	*u*	*v*	*w*	Peak height*	*u*	*v*	*w*	Peak height*
$u = \frac{1}{2}$	$\frac{1}{2}$	1.00	0.34	99	$\frac{1}{2}$	0.50	0.98	99
$v = \frac{1}{2}$	0.80	$\frac{1}{2}$	0.34	94	0.50	$\frac{1}{2}$	0.98	99
	0.20	$\frac{1}{2}$	0.34	54				
$w = 0$	0.30	0.50	0	99				
	0.70	0.50	0	91				

* Peak heights are given in arbitrary units. Values of 99 correspond to the highest peak in the section.

2.3. *Translation function*

In space group $P2_12_12$, the translation function (Crowther & Blow, 1967) contains Harker sections at ($u = \frac{1}{2}$), ($v = \frac{1}{2}$) and ($w = 0$), where maxima should be located. The translation-function calculations were carried out for both molecular orientations using the correctly oriented *B. stearothermophilus* tetramer (Skarżyński *et al.*, 1987) with native data between 7.0 and 5.0 Å resolution. The highest peaks, located in the expected sections, were prominent and obeyed both the exact and the pseudo twofold symmetries (Table 3). The corresponding translation parameters were (0.15, 0.25, 0.17) and (0.0, 0.0, 0.49) for the two molecular orientations, respectively (Fig. 2). The translation parameters corresponding to the second molecular orientation indicates that the molecular twofold *Q* axis is coincident with the crystallographic *z* axis. Hence, the asymmetric unit contains one gGAPDH tetramer situated in a general position, plus half a tetramer, located in a special position. With a total of six gGAPDH monomers in the unit cell, of total $M_r \simeq 849.6$ kDa, this crystal form has a V_M of 4.70 Å^3 Da^{-1} (Matthews, 1968) and a large solvent content of *ca* 73%.

The relative origin between the two molecules was determined using the program *BRUTE* (Fujinaga & Read, 1987) by calculating correlation coefficients with 5.0–4.0 Å data after translating the half tetramer to the eight permissible origins of the space group, while keeping the

location of the tetramer fixed. Translation parameters of (0.5, 0.0, 0.49) gave a correlation coefficient between calculated and observed structure-factor intensities of 0.37, far above the values obtained for the seven alternative origins (ranging between 0.27 and 0.29). Afterwards, the rotation and translation parameters were further improved by maximizing the correlation coefficient. This was carried out by varying separately the rotation parameters (with an increment of 0.1°) and the translation parameters (with an increment of 0.1 Å). The final rotation and translation parameters are given in Table 4.

Table 4. *Final molecular-replacement rotation and translation parameters*

	gGAPDH tetramer (general position)	gGAPDH dimer (special position)
Rotation*	53.40, 89,26, 90.26	19.00, 90.00, 90.00
Translation†	0.154, 0.249, 0.169	0.500, 0.000, 0.489

* Eulerian angles α, β, γ (°).

† Translation parameters along x, y, z, expressed in fractions of the crystallographic axes.

This properly rotated and translated *B. stearothermophilus* GAPDH model gave an R factor of 43.1% *versus* the 22022 native structure-factor amplitudes in the resolution range 7.0–3.2 Å.

2.4. *Energy minimization of the molecular-replacement solution*

In order to check further the validity of the molecular-replacement solution, which had been obtained with a 37% complete native data set, the properly oriented and translated *B. stearothermophilus* model, lacking the NAD cofactors, was subjected to 250 steps of steepest descent energy minimization. The calculations, which included X-ray terms in the resolution range 7.0–4.0 Å, were performed using the program *GROMOS* (Fujinaga, Gros & van Gunsteren, 1989; van Gunsteren & Berendsen, 1987). The weight w given to the X-ray terms, which is applied in the form,

$$1/w^2 \sum_{hkl}(|F_{obs}| - k|F_{calc}|)^2,$$

was set equal to half the average r.m.s. discrepancy between observed and calculated structure-factor amplitudes. This procedure reduced the R factor to 31.2% for 7.0–4.0 Å data, and gave an R factor of 36.5% for all data between 7.0 and 3.2 Å.

2.5. *Preliminary model building*

The coordinates of the properly oriented and translated *B. stearothermophilus* GAPDH molecule were used as a starting point for initial model building of the gGAPDH structure. All model building was carried out for the R subunit of the full tetramer only (subunit names are as described by Skarżyński *et al.*, 1987) in the 7.0–3.2 Å σA $2mF_o - DF_c$ (Read, 1986) electron-density distribution, with the program *FRODO* (Jones, 1985). After model building, the remaining five subunits were generated from the R subunit by application of the non-crystallographic symmetry operations. We used the sequence alignment given by Michels *et al.* (1986) for this preliminary model building: mutations of dissimilar side chains were carried out using the sub-program *SAM* of *FRODO*. Wherever appropriate, the conformation of the polypeptide chain and of side chains were modified for a good fit in the electron density. Otherwise, the side chains were left in the locations found in the *B. stearothermophilus* GAPDH model, or they were set so that the overlap between atoms of the new and old side chains was maximized. The electron density corresponding to the insertions present in the gGAPDH sequence (Michels *et al.*, 1986) was of insufficient quality to determine the course of the polypeptide chain. Hence, these regions were introduced at the surface of the subunit by lengthening the existing secondary-structure elements, while preventing clashes with atoms of symmetry-related subunits. This preliminary model

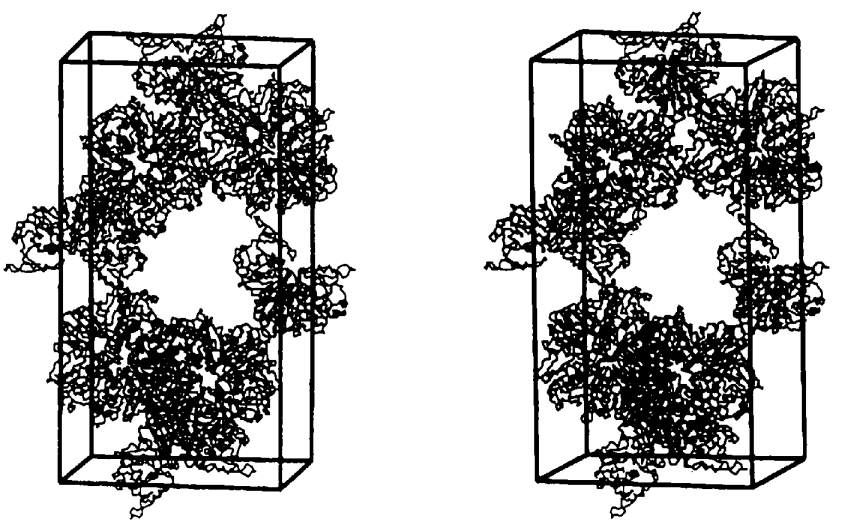


Fig. 2. Molecular packing within the $P2_12_12$ unit cell. The plots show both the gGAPDH molecules present as tetramers in the asymmetric unit, and the gGAPDH molecules present as half tetramers in the asymmetric unit. These are located in special positions within the unit cell, with their molecular Q axes coincident with the crystallographic z axis. The unit cell is shown in thick lines. Only Cα tracings are shown. Note the very large solvent channels between molecules, which are responsible for the very large solvent content (*ca* 74%) of this crystal form.

Table 5. *Iterative electron-density modification*

Solvent flattening (resolution limits: 7.0–3.2 Å)

Cycle No.	Map-inversion *R* factor	Average Sim weight	Average phase shift* (°)	Phase shift from previous cycle (°)
1	0.326	0.829	33.7	—
6	0.140	0.986	43.3	3.9
11	0.127	0.989	48.2	2.4

Sixfold averaging (resolution limits: 7.0–3.2 Å)

Cycle No.	Extension limits (Å)	Map-inversion *R* factor	Average Sim weight	Average phase shift* (°)	Phase shift from previous cycle (°)
1	—	0.340	0.859	42.3	—
4	4.60 4.42	0.302	0.894	49.6	4.6
7	4.79 4.26	0.297	0.897	53.0	2.7
10	5.00 4.10	0.305	0.885	54.3	2.1
13	5.22 3.96	0.302	0.877	54.9	1.7
16	5.47 3.83	0.296	0.875	55.2	1.5
19	5.75 3.71	0.291	0.876	56.0	1.4
22	6.05 3.59	0.280	0.882	56.0	1.3
25	6.38 3.48	0.262	0.891	55.9	1.3
28	6.76 3.38	0.240	0.901	55.7	1.2
31	7.00 3.28	0.217	0.916	55.6	1.1
34	7.00 3.20	0.199	0.933	55.5	0.8
36	7.00 32.0	0.190	0.942	55.4	0.7

* Phase shift from the phases obtained after energy minimization of the molecular replacement solution. These values are computed only for measured reflections.

building step was required in order to defined a proper envelope for density modification (*vide infra*).

2.6. *Density modification*

In order to overcome the effects caused by the abscence of 63% of the native data, which results in noisy and discontinuous electron-density maps, the native map was improved by an iterative procedure of density modification using software from the *DEMON/ANGEL* software suite (Vellieux *et al.*, 1995, in the press) as described previously (Vellieux *et al.*, 1993). Briefly, molecular envelopes corresponding to the six identical subunits were generated separately starting from the atomic coordinates of the preliminary gGAPDH model, by taking as part of the envelope all points situated within 4.0 Å from any atomic position. Care was taken at this stage to include the atoms corresponding to the NAD cofactors for the generation of the envelopes, in order to avoid accidental flattening of the electron density in the cofactor-binding region. The envelopes were then merged and further processed to take care of overlapping segments. The resulting envelope was used throughout the density-modification procedure. During this procedure, no phase combination was carried out at any stage, so that the starting phases (obtained after energy minimization of the molecular-replacement model) were used only once, to compute the initial σA $2mF_o - DF_c$ (Read, 1986) electron-density map.

The electron density was first subjected to 11 cycles of solvent flattening (Wang, 1985), with a solvent content in the envelope of 47%. Solvent flattening was carried out using the measured data between 7.0 and 3.2 Å, and no attempts were made at this stage to complete the phasing set by the use of map inversion structure factors in place of the missing reflections. Also, in contrast to the original solvent-flattening procedure (Wang, 1985), negative-density truncation was not carried out in the protein regions. Solvent flattening reduced the map inversion *R* factor from an initial value of 32.6% in the initial cycle to 12.7% at the ultimate cycle, while the average Sim weight (Sim, 1959, 1960) rose from an initial value of 0.829 to 0.989 (Table 5). The average phase difference between the phases obtained after energy minimization, which were used to compute the initial electron-density map, and the map-inversion phases obtained in the last cycle was 50.6°.

Next, the resulting electron-density map was further improved by iterative sixfold density averaging and solvent flattening. In this process, the measured data between 7.0 and 3.2 Å were used throughout, and the missing reflections (63% of the data in this resolution range) were gradually introduced as map-inversion structure-factor amplitudes and phases: after an initial three cycles of averaging using only observed amplitudes, map-inversion structure factors were introduced in place of the missing reflections in a narrow resolution shell (4.60 to 4.42 Å) where the completeness of the native data is high. Further phase expansion was then gradually carried out towards both high and low resolution, in steps of 1 reciprocal lattice point along the *b* axis (Table 5). After each extension step, phase refinement was carried out for two cycles before the next extension step. Thus, the whole averaging procedure

Table 6. *Simulated-annealing refinement protocol using X-PLOR*

Round number	Data set*	Non-crystallographic symmetry	Description†
1	Native	Constrained	30 steps of rigid-body refinement (RBREF), with each subunit defined as an independent rigid group; determination of weight W_A (747720 kcal mol^{-1}); 70 steps of energy minimization with X-ray constraints (EMX), with harmonic repulsion; 130 EMX steps without harmonic repulsion; 1 ps molecular dynamics MDX, $\tau = 0.5$ fs, with $T = 2000$ K; 1.7 ps MDX, $\tau = 0.5$ fs, with T decreasing from 2000 to 300 K in steps of 25 K; 44 EMX steps; 20 steps of temperature factor refinement (BFREF).
2	Combined	Constrained	Update of the non-crystallographic symmetry operations with 30 RBREF steps. The protocol was similar to that of round 1, with $W_A = 222350$ kcal mol^{-1}, and included 120 EMX steps; 3.7 ps MDX, $\tau = 0.5$ fs, with T decreasing from 4000 to 300 K; 120 steps of EMX, and 20 BFREF steps.
3	Combined	Constrained	Identical to round 2, with $W_A = 300000$ kcal mol^{-1}
4	Combined	Constrained	Similar to refinement rounds 2 and 3, without the initial RBREF steps. Otherwise, $W_A = 339350$ kcal mol^{-1}, and the slow cooling MDX steps were carried out for a total of 3.5 ps, with T decreasing from 4000 to 475 K.
5*	Combined	Restrained	After the initial RBREF steps, 120 EMX steps were performed with non-crystallographic symmetry constraints ($W_A = 310000$ kcal mol^{-1}). Afterwards, non-crystallographic symmetry restraints were used, with a positional restraint force constant of 250 kcal mol^{-1}, a target deviation for temperature factors of 2 Å^2 and $W_A = 1710000$ kcal mol^{-1}. During slow cooling, the temperature was reduced from 4000 to 1550 K in 2.5 ps.

* Resolution range 7.0–3.2 Å.
† For details of the protocol, see Brünger (1990).

required three initial cycles at constant resolution, followed by 11 phase and amplitude-extension steps, and 22 phase regularization steps (Table 5). Map coefficients of the form $m_{\rm sim}|F_o|[\exp(i\varphi_{\rm inv})]$ for the observed reflections and of the form $0.9\langle m_{\rm sim}\rangle|F_{\rm inv}|[\exp(i\varphi_{\rm inv})]$ for the unobserved reflections were used in the course of the iterations, where $\langle m_{\rm sim}\rangle$ is the average Sim weight in a given resolution bin. During this procedure, the average Sim weight rose from an initial value of 0.859 to a final value of 0.942, while the map inversion R factor dropped from 34.0% in the initial cycle to 19.0% in the final cycle. The average phase difference between the final map-inversion phases and the phases obtained after energy minimization of the molecular-replacement model was 55.4° (Table 5). That between the phase set obtained after solvent flattening and the phases obtained at completion of the averaging procedure was 56.0°.

2.7. *Refinement*

The electron density obtained at completion of the iterative density-modification procedure was of good quality, allowing rebuilding of the model. In particular, unambiguous density corresponding to the insertions present in the gGAPDH sequence was observed at the surface of the protein. Model rebuilding was initiated starting from the initial gGAPDH model which had been built prior to the density-modification procedure. *A posteriori*, seeing that the overall topology of the gGAPDH molecule (not taking into account the regions were insertions are located) is very similar to that of the *B. stearothermophilus* GAPDH subunit, it would have been preferable to start model rebuilding from the latter model. The model of holo-gGAPDH, which included the NAD cofactors, was refined against the available 7.0–3.2 Å resolution Laue data in five rounds of simulated annealing (Table 6) using the program *X-PLOR* (Brünger, Kuriyan & Karplus, 1987). In the first four refinement rounds, sixfold non-crystallographic symmetry constraints were used, whereas tight non-crystallographic symmetry restraints were used in the ultimate round. Initial isotropic temperature factors of the starting model were identical for equivalent atoms of the six gGAPDH subunits: the original temperature factors of the *B. stearothermophilus* holo-GAPDH model were

Table 7. *Refinement statistics*

Progress of the refinement			
Round number	Refinement stage	R factor (%)	R.m.s. coordinate shift (Å)*
1	—	39.5	—
	RBREF	39.1	0.196
	EMX	25.7	0.955
	MDX	24.0	0.893
	EMX	22.4	0.056
	BFREF	21.0	—
2	—	29.2	2.958
	RBREF	28.9	0.126
	EMX	23.8	0.512
	MDX	22.7	0.636
	EMX	21.9	0.110
	BFREF	20.8	—
3	—	21.9	1.400
	RBREF	21.9	0.029
	EMX	20.4	0.244
	MDX	20.4	0.384
	EMX	19.9	0.085
	BFREF	19.4	—
4	—	20.7	0.433
	EMX	19.5	0.196
	MDX	19.7	0.333
	EMX	19.4	0.099
	BFREF	19.3	—
5	—	20.7	0.704
	RBREF	20.7	0.023
	EMX	19.5	0.219
	MDX	19.0	0.276
	EMX	18.0	0.110
	BFREF	17.6	—

* For main-chain atoms only.

used for the identical residues in the gGAPDH structure. These temperature-factor values were also used for equivalent atomic positions in dissimilar residues. All other atoms were given identical arbitrary B-factor values of 12.0 Å^2. After each refinement round, a molecular envelope was generated from the refined coordinates, which was used for sixfold averaging of electron densities calculated with refined model phases. Both once-averaged and unaveraged 7.0–3.2 Å resolution σA $2mF_o$–DF_c electron-density maps (Read, 1986) were used for model rebuilding after each refinement round. The criteria used to determine the problem regions in the models included *e.g.* a poor fit in electron density, unfavourable steric contacts or interactions, disallowed φ–ψ combinations and high temperature-factor values. At completion of the first round of refinement, the composition of the model was updated to include two sulfate ions per subunit, located in the active site, in similar positions to those found in *B. stearothermophilus* (Skarżyński *et al.*, 1987) and lobster (Olsen, Garavito, Sabesan & Rossmann, 1976) GAPDH subunits. Table 7 summarizes the results of this refinement protocol, which reduced the R factor from an initial value of 36.5 to 17.4% for all data in the combined data set in the resolution range 7.0–3.2 Å. With the very weak restraints imposed on temperature factors, a few poorly defined segments of the current model have very high B factors and, therefore, contribute very little to the calculated stucture factors.

3. Results of data completion by iterative density modification

Due to the lack of completeness of the native Laue data set, the initial electron-density distribution, obtained after energy minimization of the molecular-replacement model followed by solvent flattening, was of a poor quality. In particular, it contained numerous breaks along the polypeptide chain, as can be judged from a plot of the real-space correlation coefficient (not shown; Jones, Zou, Cowan & Kjeldgaard, 1991) computed for the main-chain atoms only. The procedure which we used to improve the map relied on the introduction of map-inversion amplitudes and phases in place of the missing data, which were almost randomly distributed throughout reciprocal space (Vellieux *et al.*, 1993). No phase combination was carried out at any stage of the density-modification procedure.

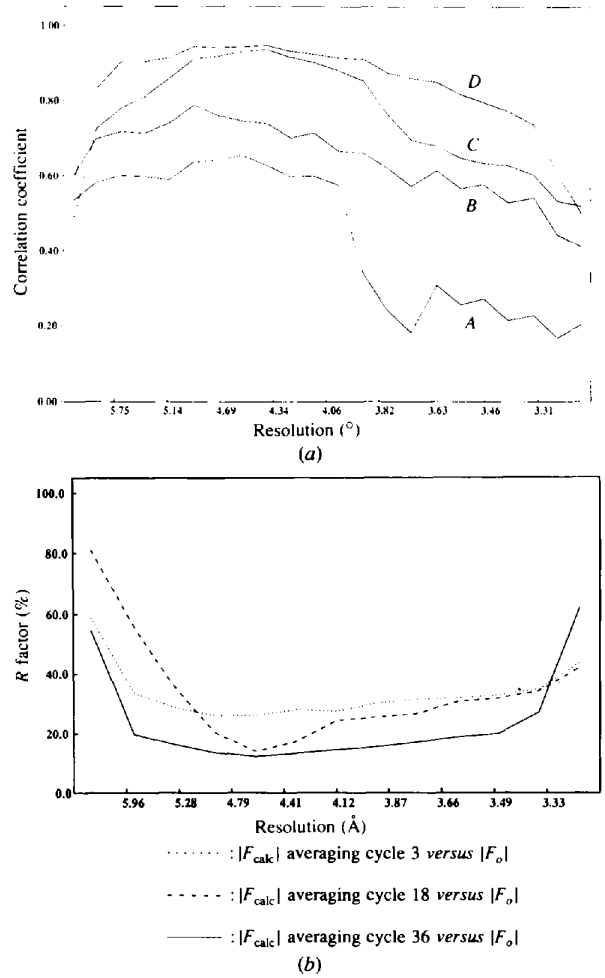


Fig. 3. Results of the procedure of data completion by iterative sixfold density averaging. Plots of the R factors and correlation-coefficient values (CC) between observed and calculated structure-factor amplitudes, calculated in 22 bins within the resolution range 7.0–3.2 Å, where $R = \sum_h(|F_o^h| - |F_c^h|)/\sum_h|F_o^h|$ and $CC = \{\sum_h[(\langle F_o\rangle - |F_o^h|)(\langle F_c\rangle - |F_c^h|)]\}/\{[\sum_h[(\langle F_o\rangle - |F_o^h|)^2(\langle F_c\rangle - |F_c^h|)^2]^{\frac{1}{2}}\}$. (*a*) Plots of the correlation coefficients at various stages of the procedure. Curve *A*, using calculated structure-factor amplitudes after energy minimization of the molecular-replacement model; curve *B*, using map-inversion structure-factor amplitudes, after the initial cycle of averaging; curve *C*, idem, after 18 cycles of the averaging procedure; curve *D*, idem, at completion of the iterative density averaging procedure (cycle 36). (*b*) Plots of the map-inversion R factors at various stages of the procedure; · · ·, $|F_c|$ obtained after cycle 3 of the procedure; – – –, idem, after cycle 18; ——, idem, after cycle 36. (*c*) Plots of the map inversion R factors *verus* different types of structure-factor amplitudes; – – –, $|F_c|$ obtained after cycle 36 of the procedure *versus* missing $|F_o|$, which were recorded later in the pentalenolactone data set; ——, $|F_c|$ obtained after cycle 36 of the procedure *versus* observed $|F_o|$.

The success of this procedure may be judged by analyzing plots of the correlation coefficient and R factor between map-inversion and native structure-factor amplitudes, which were computed at different stages of the procedure (Figs. 3*a* and 3*b*): the agreement between amplitudes improves once estimated values for the missing reflections have been introduced. Also, since additional Laue data was recorded at a later stage for crystals soaked in the inhibitor pentalenoactone (this data being used during refinement), it is possible to compare *a posteriori* the estimated amplitudes used in place of the missing data obtained at completion of the averaging procedure with the structure-factor amplitudes in the inhibitor data set, absent from the native Laue data set, which were recorded later (Fig. 3*c*). The overall R factor between these two sets of amplitudes is 41.9%, and the corresponding curves have the same shape as those between map-inversion and observed structure-factor amplitudes obtained at completion of the averaging procedure. This indicates that the estimated amplitudes provided for these missing data are non-random.

Another way of analyzing the success of the procedure is by means of comparing several phase sets. With a refined molecular model now available, the phases from this refined model can be taken as a reference to which several phase sets can be compared. These comparisons

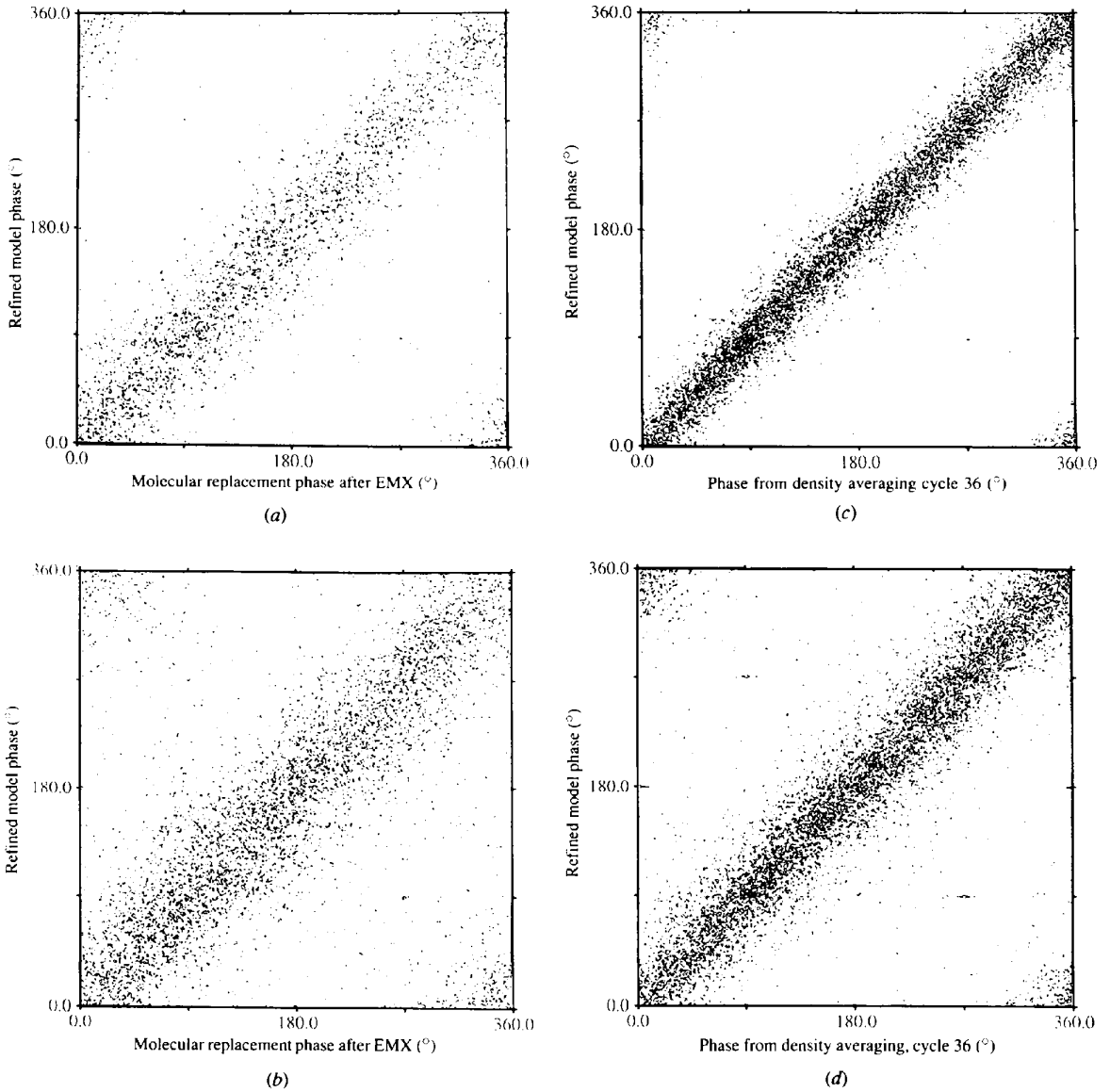
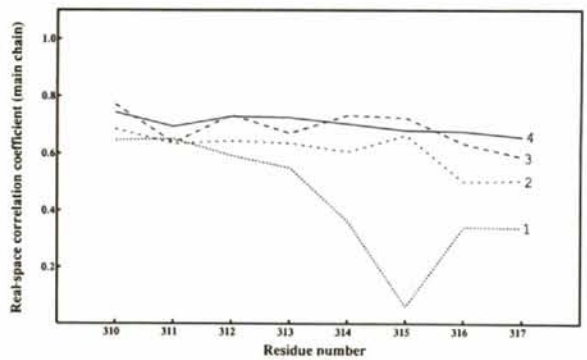
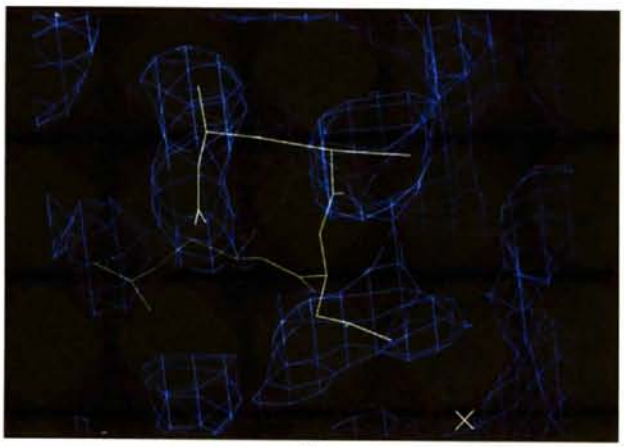


Fig. 4. Analysis of the accuracy of the phases generated by the density-modification procedure. Phase scatter plots comparing different phase sets to the phases computed from the refined molecular model. (*a*) Phase set computed from the molecular-replacement model, for the observed reflections only; (*b*) phase set obtained at completion of the iterative averaging procedure, for the observed reflections only; (*c*) as in (*a*), for the unobserved, missing reflections; (*d*) as in (*b*), for the unobserved, missing reflections.



(*a*)



(*b*)

(*c*)

Fig. 5. Improvement in the quality of the electron density in the course of the iterative averaging procedure. (*a*) Plots of the real-space correlation coefficients (Jones *et al.*, 1991) in the region of the gGAPDH subunit where a short β-strand is replaced by a short α-helical segment, computed at different stages of the density-modification procedure. 1, ρ_{model} *versus* $\rho_{\text{em}\times 5}$ ($\langle CC\rangle = 0.38$); 2, ρ_{model} *versus* ρ_{avg4} ($\langle CC\rangle = 0.52$); 3, ρ_{model} *versus* ρ_{avg19} ($\langle CC\rangle = 0.57$); 4, ρ_{model} *versus* ρ_{avg36} ($\langle CC\rangle = 0.59$). (*b*) Electron density corresponding to the additional α-helix, $\alpha 7$, obtained after energy minimization of the molecular-replacement solution. (*c*) Averaged electron density corresponding to helix $\alpha 7$ obtained at cycle 36 of the density-modification procedure. Electron densities are contoured at 1.0σ level; the refined molecular model is shown in green.

Table 8. *R.m.s. deviations from ideal geometry after completion of the refinement*

These calculations were carried out using the program *TNT* (Tronrud, Ten Eyck & Matthews, 1987)

Bond lengths (Å)	0.020
Bond angles (°)	3.579
Torsion angles (°)	25.506
Trigonal non-planarity (Å)	0.006
Non-planarity (Å)	0.028
Non-bonded interactions (Å)	0.028

have been made separately for the observed reflections (which had been used in the averaging procedure) and for the unobserved data (Fig. 4). Comparing the results obtained with the phases calculated from the energy-minimized molecular-replacement model to those obtained after completion of the density-modification procedure show the same trend, both for the observed (Figs. 4*a* and 4*b*) and the missing reflections (Figs. 4*c* and 4*d*). In all cases, the phases form a cloud scattered along the main diagonal indicating their non-randomness. This cloud contracts along the diagonal at completion of the procedure, indicating an improvement in the quality of the resulting phase set, also for the missing reflections. For the observed reflections, the average phase difference between the refined model phases and the phases from the molecular-replacement solution after energy minimization is 53.3°, and that between the refined model phases and those obtained at completion of the density-modification procedure is 29.6°. For the unobserved reflections, these figures are 62.2 and 50.2°, respectively. Thus, the estimates for the

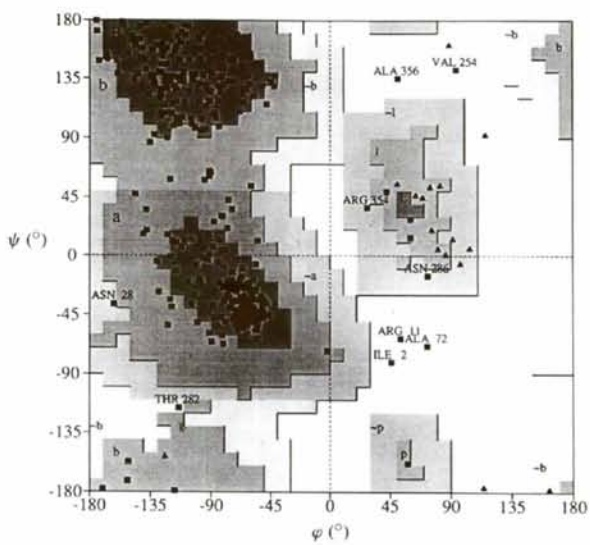


Fig. 6. The (φ, ψ) plot of one of the gGAPDH subunit. Glycine residues are indicated by black triangles and other residues by black squares. The regions of the plot shown in dark grey indicate the fully allowed (φ, ψ) combinations, those in lighter grey indicate the additionally allowed (φ, ψ) combinations, and the regions shaded in light grey indicate the generously allowed regions of conformational space (Laskowski *et al.*, 1993). The (φ, ψ) plots for the other subunits are virtually identical.

phases of the missing reflections provided by the averaging procedure are also not random. The inclusion of non-random estimated values for the missing structure factors led to a dramatic improvement in map quality.

This latter point can be judged from the electron density observed in a surface portion of the molecule, in which a short β-strand present in the *B. stearothermophilus* enzyme model, which was used for the molecular-replacement phasing, is replaced by a small α-helical segment, $\alpha 7$, comprising residues 311–316. Plots of the real-space correlation coefficients, computed in this region of the molecule (Fig. 5*a*), indicate an improvement in the quality of the electron density in the course of the averaging procedure. The initial electron-density map is discontinuous, with little evidence for a helical region (Fig. 5*b*), while the map obtained at completion of the density-modification procedure shows very clear density for this α-helical segment (Fig. 5*c*).

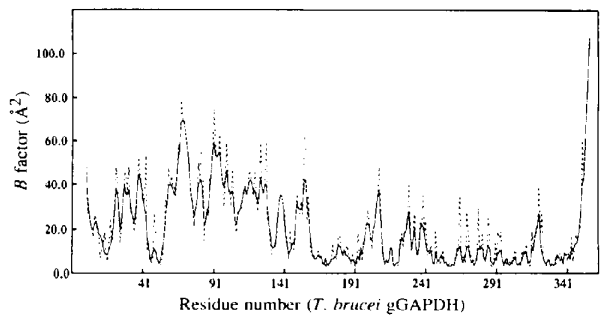


Fig. 7. Plot of the *B* factor for subunit *P versus* the sequence number. Continuous line: main-chain atoms; broken line: side-chain atoms. The lines at the top of the plot indicate the secondary-structure elements found in the gGAPDH subunit.

Table 9. *Molecular symmetry within the gGAPDH molecules*

Molecular axis	Eulerian angles (°) κ	φ	ψ
P	180.00	180.24	53.78
Q	180.00	88.62	89.08
R	180.00	358.44	36.21
P'	180.00	180.00	19.00

Notes: the definition of the molecular axes is as defined by Skarżyński *et al.* (1987). The axes *P*, *Q* and *R* apply to the complete gGAPDH tetramer, located in a general position within the asymmetric unit, and the *P'* axis applies to the gGAPDH molecule which has its *Q* axis coincident with the crystallographic *z* axis. The Eulerian angle values were calculated by superimposing subunits across the molecular axis using the method of Rao & Rossman (1973), with the refined gGAPDH coordinates obtained after rigid-body refinement with *X-PLOR*, at round 5 of the refinement procedure.

4. The overall structure

4.1. *Quality of the model*

The refined model contains 2735 protein atoms, plus one NAD cofactor and two sulfate ions per subunit. In view of the limited resolution of the data used in this

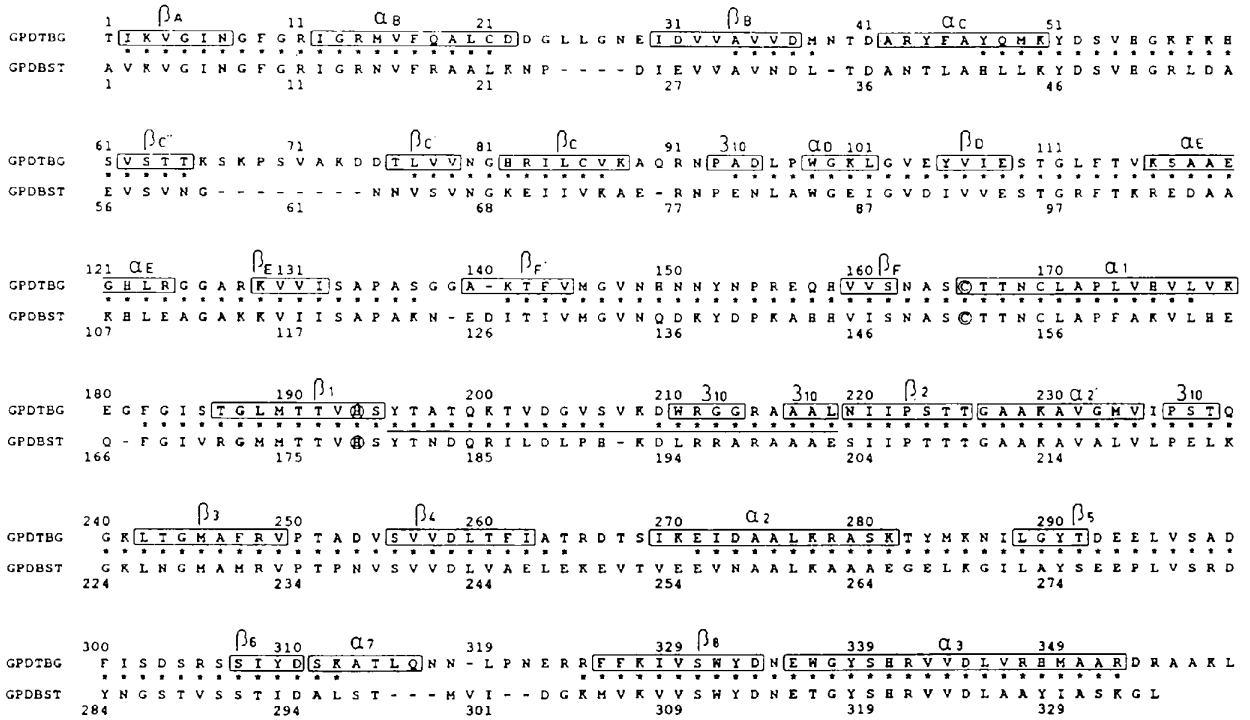


Fig. 8. Aligned sequences of the *T. brucei* gGAPDH and *B. stearothermophilus* GAPDH subunits based on the superposition of their three-dimensional structures. GPDTBG: sequence of the *T. brucei* gGAPDH subunit; GPDBST: sequence of the *B. stearothermophilus* GAPDH subunit. The stars indicate the residues used for the structure superposition. Boxed residues show the secondary-structure elements present in the gGAPDH subunits, and the horizontal line indicates the residues present in the S-loop.

study, no water molecules were built into this model, although convincing density corresponding to several potential water sites was observed at the surface of the protein. The r.m.s. deviations from ideal stereochemical parameters are well within the range expected in view of the 3.2 Å resolution of the data (Table 8).

The stereochemical quality of this refined model was analyzed with the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). The φ, ψ plot (Fig. 6, Ramakrishnan & Ramachandran, 1965) shows most residues in the allowed regions of conformational space. The most favoured regions contain 82% of the residues, with an additional 15% present in the additional allowed regions (Laskowki *et al.*, 1993). Five residues out of a total of 358 residues are in the fully disallowed regions of conformational space. These are Ile2, Arg11, Ala72, Val254 and Ala356, for which the electron-density distribution, obtained either after completion of the real-space averaging procedure, or computed using phases from the refined model, shows no convincing density. All stereochemical parameters for the main-chain and side-chain atoms are better than those obtained for protein structures refined to similar resolutions. This is probably because of the specific use which was made of the sixfold non-crystallographic redundancy, both to generate an electron-density distribution of high quality by real-space averaging, and later during the refinement process. The r.m.s. coordinate error, estimated from a Luzzati plot (Luzzati, 1952) or from a σA plot (Read, 1986) is 0.37 Å.

Another interesting point to analyze is the distribution of the temperature-factor values. The average B is 21.7 Å^2, with a very high maximum of 114.0 Å^2. The plot of the average temperature factor along the polypeptide chain (Fig. 7) shows two regions, one at the N-terminal half of the sequence having higher temperature factors, the other corresponding to the C-terminal half with lower B factors. Interestingly, this division of the amino-acid sequence corresponds to the subdivision of the subunit into two structural domains (*vide infra*).

4.2. *Symmetry of the molecule*

As is the case for the holo-enzyme from *B. stearothermophilus*, the *T. brucei* glycosomal GAPDH molecule exhibits ideal 222 symmetry (Table 9). This is the case for both types of molecules present in the asymmetric unit. Nonetheless, the lack of deviation from this ideal molecular symmetry may be caused, at least in part, by the constrained then highly restrained refinement procedure which was used.

4.3. *Intersubunit contacts*

The subunit–subunit interactions holding together the gGAPDH tetramer are much more extensive than the interactions between the molecules which form the crystal lattice. These interactions can be classified according to the accessible surface which is buried in each subunit/subunit interface across the molecular twofold axes. The smallest interactions take place across the molecular Q axis (notations of the molecular axes are as in Skarżyński *et al.*, 1987): the surface area buried within the OQ dimer is 516 Å^2 per subunit. This surface area is 553 Å^2 in the *B. stereothermophilus* enzyme. The other two molecular axes bury larger and similar areas of subunit surface per gGAPDH subunit: 1428 Å^2 for the R-axis interface, and 1994 Å^2 for the P-axis interface (with corresponding surface areas of 1432 and 2124 Å^2 buried in the *B. stearothermophilus* enzyme, respectively).

4.4. *Overall architecture and topology of the gGAPDH subunit*

The *T. brucei* gGAPDH subunit has amino-acid sequence indentity of 57% with the sequence of lobster GAPDH, and of 54% with that of *B. stearothermophilus* GAPDH (Michels *et al.*, 1986). It is, therefore, not

Fig. 9. *MOLSCRIPT* (Kraulis, 1991) ribbon drawing of the *T. brucei* gGAPDH subunit, with indications of the secondary-structure elements. The NAD cofactor and the two sulfate ions present in the subunit are also displayed.

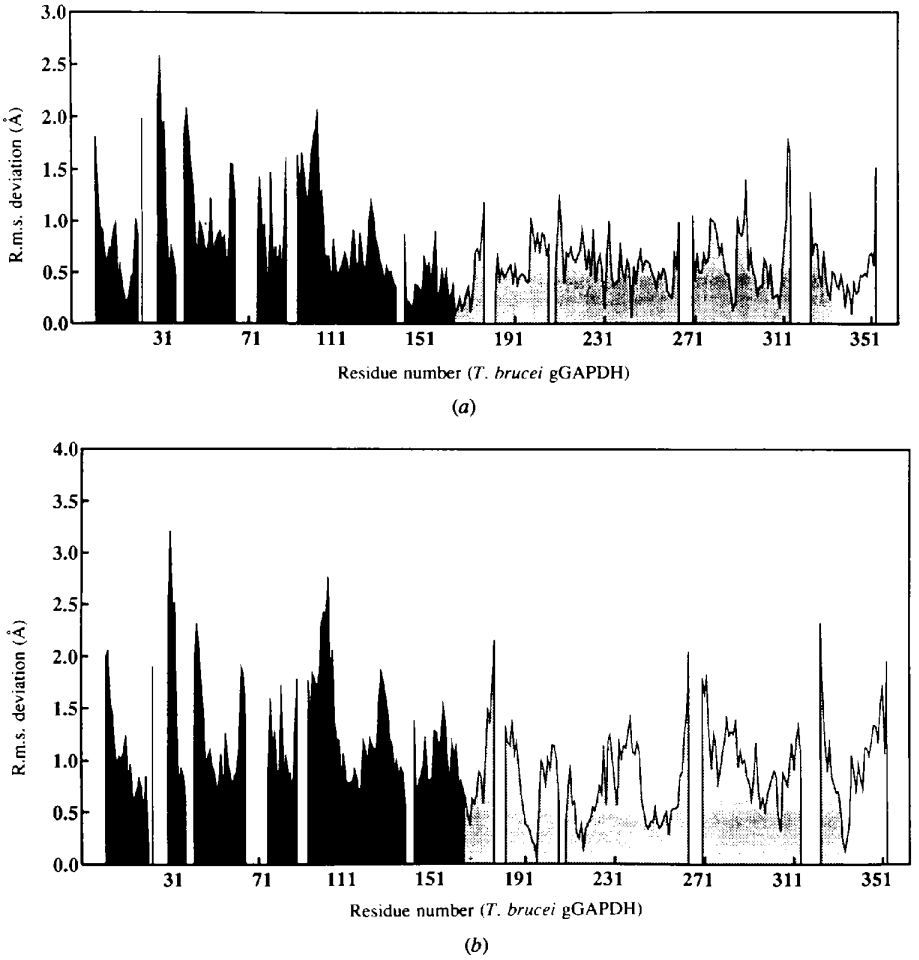


Fig. 10. Plots of the r.m.s. deviations between equivalent atomic positions in the superimposed *T. brucei* gGAPDH and *B. stearothermophilus* GAPDH molecules. (*a*) The superposition was carried out using one subunit only; (*b*) The superposition of the GAPDH molecules was carried out by means of superimposing their respective 222 molecular axes.

Fig. 11. The superimposed subunits of the GAPDHases from *T. brucei* (thick lines) and *B. stearothermophilus* (thin lines). Only the Cα tracings of the two subunits are shown; also displayed are the NAD cofactor and the two sulfate ions located in the active site of the trypanosomal enzyme.

unexpected that the overall fold of the *T. brucei* gGAPDH subunit resembles that of these two dehydrogenases of known three-dimensional structure, with r.m.s. coordinate differences of 0.71 Å for 268 equivalent Cα positions and of 0.64 Å for 268 equivalent positions, respectively. Fig. 8 shows the sequence alignment, obtained from the superimposition of three-dimensional structures, between the *T. brucei* gGAPDH subunit and that of the enzyme from *B. stearothermophilus*. It also displays the secondary-structure elements present in the gGAPDH subunit. The overall fold of the gGAPDH subunit is shown, in the form of a *MOLSCRIPT* ribbon drawing (Kraulis, 1991), in Fig. 9.

5. Comparison with the enzyme from *B. stearothermophilus*

The *T. brucei* gGAPDH molecule, just as is the case for the molecule from *B. stearothermophilus* (Skarżyński *et al.*, 1987), displays nearly perfect molecular 222 symmetry, at least within the limitations because of the limited resolution (3.2 Å) and the refinement procedure used. The two structures have been compared both by superimposing their subunits, and by superimposing the two tetrameric molecules by means of superimposing their respective 222 molecular axes. The subunits of the two structures have very similar folds, with an r.m.s. deviation of 0.64 Å for 268 equivalent Cα positions. Interestingly, a plot of the r.m.s. deviation *versus* the residue number (Fig. 10*a*) shows larger deviations for the N-terminal domain than for the C-terminal catalytic domain. The plots made using both GAPDH molecules compared by superimposing their respective molecular 222 axes show a comparable pattern (Fig. 10*b*). This is very similar to the variation of the temperature factors as function of the residue number (*vide supra*). Thus, the more mobile NAD-binding domain is also more dissimilar to the equivalent domain in the *B. stearothermophilus* enzyme.

The two superimposed subunits (Fig. 11) clearly show that the *T. brucei* subunit occupies more volume than the *B. stearothermophilus* subunit. This is in accordance with their respective amino-acid sequences (Fig. 8, Michels *et al.*, 1986) in which the former is longer than the latter by 24 residues. The gGAPDH subunit contains four insertions of three residues or larger. These are located in surface regions (Fig. 11), where the chain tracing between the two subunits differs most. These sequence segments differing most in their three-dimensional structures are residues 22–29, 66–74, 179–181, 313–322 and 354–358 in the *T. brucei* subunit, corresponding to residues 22–25, 61, 165–166, 297–302 and 334 in the bacterial enzyme.

We thank Dr R. J. Read for his contributions to the initial stages of this project. We are most grateful to Dr F. R. Opperdoes and colleagues in Brussels for providing glycosomal GAPDH isolated from the sleeping sickness parasite. We also acknowledge support from the WHO/UNDP/World Band Special Program for Tropical Diseases (TDR0) and the Science and Technology for Development (STD) Programme for the EEC.

References

BRÜNGER, A. T. (1990). *X-PLOR Version* 2.1 *Manual*. Yale Univ., New Haven, CT, USA.

BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **235**, 458–460.

CARR, P. D., CRUICKSHANK, D. W. J. & HARDING, M. M. (1992). *J. Appl. Cryst.* **25**, 294–308.

CARR, P. D., DODD, I. M. & HARDING, M. M. (1993). *J. Appl. Cryst.* **26**, 384–387.

CLIFTON, I. J., ELDER, M. & HAJDU, J. (1991). *J. Appl. Cryst.* **24**, 267–277.

CROWTHER, R. A. (1972). In *The Molecular Replacement Method* edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon & Breach.

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

FAIRLAMB, A. H., OPPERDOES, F. R. & BORST, P. (1977). *Nature (London)*, **265**, 270–271.

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

FUJINAGA, M., GROS, P. & VAN GUNSTEREN, W. F. (1989). *J. Appl. Cryst.* **22**, 1–8.

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

VAN GUNSTEREN, W. F. & BERENDSEN, H. J. C. (1987). *GROMOS. Groningen Molecular Simulation Library*. BIOMOS bv, Groningen, The Netherlands.

HAJDU, J., ALMO, S. C., FARBER, G. K., PRATER, J. K., PETSKO, G. A., WAKATSUKI, S., CLIFTON, I. J. & FULOP, V. (1991). In *Crystallographic Computing* 5: *From Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 29–49. Oxford Univ. Press.

HARRIS, I. J. & WATERS, M. (1976). In *The Enzymes*, edited by P. D. BOYER, Vol. 13, pp. 1–49. New York: Academic Press.

HELLIWELL, J. R., HABASH, J., CRUICKSHANK, D. W. J., HARDING, M. M., GREENHOUGH, T. J., CAMPBELL, J. W., CLIFTON, I. J., ELDER, M., MACHIN, P. A., PAPIZ, M. Z. & ZUREK, S. (1989). *J. Appl. Cryst.* **22**, 483–487.

HOL, W. G. J. (1986). *Angew. Chem. Int. Ed. Engl.* **25**, 767–778.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

JONES, T. A., ZOU, J.-Y., COWAN, S. W. & KJELDGAARD, M. (1991). *Acta Cryst.* A**47**, 110–119.

KRAULIS, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

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

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

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

MICHELS, P. A. M., POLISZCZAK, A., OSINGA, K. A., MISSET, O., VAN BEEUMEN, J., WIERENGA, R. K., BORST, P. & OPPERDOES, F. R. (1986). *EMBO J.* **5**, 1049–1056.

MISSET, O., BOS, O. J. M. & OPPERDOES, F. R. (1986). *Eur. J. Biochem.* **157**, 441–453.

MORAS, D., OLSEN, K. W., SABESAN, M. N., BUEHNER, M., FORD, G. C. & ROSSMAN, M. G. (1975). *J. Biol. Chem.* **250**, 9137–9162.

MURTHY, M. R. N., GARAVITO, R. M., JOHNSON, J. E. & ROSSMANN, M. G. (1980). *J. Mol. Biol.* **138**, 859–872.

OLIVIER, L., BUISSON, G., FANCHON, E., CORBIER, C., BRANLANT, G. & DIDEBERG, O. (1995). *Acta Cryst.* D**51**, 247–249.

OPPERDOES, F. R. (1987). *Ann. Rev. Microbiol.* **41**, 127–151.

OPPERDOES, F. R. & BORST, P. (1977). *FEBS Lett.* **80**, 360–364.

OPPERDOES, F. R., WIERENGA, R. K., NOBLE, M. E. M., HOL, W. G. J., WILLSON, M., KUNTZ, D. A., CALLENS, M. & PERIÉ, J. (1990). *Parasites: Molecular Biology, Drug and Vaccine Design*, edited by T. CERAMI & N. AGABIAN, pp. 233–246, New York: Wiley–Liss.

OLSEN, K. W., GARAVITO, R. M., SABESAN, M. N. & ROSSMANN, M. G. (1976). *J. Mol. Biol.* **107**, 571–576.
RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.
RAO, S. T. & ROSSMANN, M. G. (1973). *Acta Cryst.* **29**, 241–256.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
READ, R. J., WIERENGA, R. K., GROENDIJK, H., LAMBEIR, A., OPPERDOES, F. R. & HOL, W. G. J. (1987). *J. Mol. Biol.* **194**, 573–575.
ROSSMANN, M. G., LILJAS, A., BRÄNDÉN, C.-I. & BANASZAK, L. J. (1975). *The Enzymes*, edited by P. D. BOYER, Vol. 11, pp. 61–102. New York: Academic Press.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
SKARŻYŃSKI, T., MOODY, P. C. E. & WONACOTT, A. J. (1987). *J. Mol. Biol.* **193**, 171–187.
SKARŻYŃSKI, T. & WONACOTT, A. J. (1988). *J. Mol. Biol.* **203**, 1097–1118.
SWEET, R. M., SINGER, P. T. & SMALÅS, A. (1993). *Acta Cryst.* D**49**, 305–307.
TANAKA, N. (1977). *Acta Cryst.* A**33**, 191–193.
TRONRUD, D. E., TEN EYCK, L. & MATTHEWS, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
VELLIEUX, F. M. D., HAJDU, J., VERLINDE, C. L. M. J., GROENDIJK, H., READ, R. J., GREENHOUGH, T. J., CAMPBELL, J. W., KALK, K. H., LITTLECHILD, J. A., WATSON, H. C. & HOL, W. G. J. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2355–2359.
VELLIEUX, F. M. D. A. P., HUNT, J. F., ROY, S. & READ, R. J. (1995). *J. Appl. Cryst.* In the press.
WALKER, J. E., CARNE, A. F., RUNSWICK, M. J., BRIGDEN, J. & HARRIS, J. I. (1980). *Eur. J. Biochem.* **108**, 549–565.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.