Structure Determination of Aldose Reductase: Joys and Traps of Local Symmetry Averaging

By F. Tête-Favier and J.-M. Rondeau

Laboratoire de Cristallographie Biologique, IBMC, 15 rue Descartes, 67084 Strasbourg CEDEX, France, and Biostructure SA, Les Algorithmes, Bâtiment Euclide, Parc d'Innovation, 67400 Illkirch-Graffenstaden, France

and A. Podjarny and D. Moras

Laboratoire de Cristallographie Biologique, IBMC, 15 rue Descartes, 67084 Strasbourg CEDEX, France

Abstract

The structure of aldose reductase, a monomeric enzyme of 314 amino acids which crystallizes in space group $P1$ with four monomers per asymmetric unit, has been solved using a combination of single isomorphous replacement (SIR), solvent flattening and local symmetry averaging. The self rotation showed evidence of 222 local symmetry. The map calculated from the original single isomorphous replacement phases showed a clear solvent envelope but was uninterpretable. A first averaging attempt failed because the molecular envelope obtained from the SIR map weighted with monomer correlation was too small and the averaging was biased by low-resolution truncation. A second attempt with an enlarged envelope and including low-resolution reflections succeeded in refining phases at 3.5 Å resolution but failed to extend them correctly. Rigid-body refinement of a partial model based on the 3.5 Å map calculated from refined phases showed significant departures from the 222 symmetry. A third averaging attempt using the improved symmetry succeeded in producing a clear map with phases extended to 3.07 Å resolution. This map revealed a $(\beta/\alpha)_8$ fold, not previously found in NADPH-dependent enzymes. This work shows the importance of mask definition and local symmetry elements accuracy for averaging, and describes a method for improving these parameters.

1. Introduction

Local symmetry averaging is a very powerful technique for improving the quality of a phase set. First proposed in reciprocal space (Rossmann & Blow, 1962, 1963; Rossmann, 1972) it was subsequently implemented in real space (Buehner, Ford, Moras, Olsen & Rossmann, 1974; Bricogne, 1976) and it has been applied successfully to a number of cases, notably virus particles (Harrison, Olson, Schutt, Winkler & Bricogne, 1978; Rossmann, 1990; and references therein).

As already shown by Jones, Walker & Stuart (1991) for the case of tumor necrosis factor, the process of averaging

Table 1. *Crystallographic data and crystallization conditions for the triclinic crystal form*

Crystal form	Triclinic
Space group	$P1$
Unit-cell parameters	$a = 81.3, b = 85.9, c = 56.6$ Å
	$\alpha = 102.3, \beta = 103.3, \gamma = 79.0$
	$V = 371000$ Å^3
Number of monomers per asymmetric unit	4
Specific volume (Å^3 dalton^{-1})	2.65
Limit of diffraction (Å)	1.9
Drop composition (10 µl)	5 mg ml^{-1} AR
	2.5% Polyethylene glycol 6000
	15 mM Ci$_3$NH$_4^+$
	pH 6.2
Reservoir composition (750 µl)	20% Polyethylene glycol 6000
	15 mM Ci$_3$NH$_4^+$
	pH 6.2

depends strongly on the following parameters: (*a*) the local symmetry operations (both in number and accuracy); (*b*) the domain in which local symmetry acts (*i.e.* the molecular envelope); (*c*) the quality of the amplitude set (in resolution, accuracy and completeness); and (*d*) the errors in the initial phase set.

In some cases, such as the virus particles, the local symmetry is related to the crystal symmetry and is therefore defined with geometrical accuracy, and the molecular envelope is related to a known particle shape (*e.g.* icosahedral). This is not a general rule, since local symmetry operations and molecular envelopes can in principle be totally arbitrary. The importance of the completeness of the structure-factor set has already been reported (Rayment, 1983). Since current averaging processes are fixed-point iterations which do not allow a local minimum to be avoided, the radius of convergence is finite. Therefore the error of the initial phase set should be within this radius.

In a practical application, all the parameters listed above are related and improvement of one of them can lead to improvement of the others. The resolution of the structure of aldose reductase is an example of such a case. This structure was solved from one single derivative together with local symmetry averaging. Initially, all averaging parameters were roughly defined and it was necessary to implement a 'boot-strapping' procedure to improve them

Table 2. *Crystallographic data collection statistics for the two native and the PMS data sets*

The first native data set (NAT1) goes to higher resolution but the second native data set (NAT2) is more complete at resolutions lower than 10 Å.

Resolution d (Å)	Number of measured independent reflections			Completeness (%)			Averaged redundancy			R_{sym} (%)		
	NAT1	NAT2	PMS	NAT1	NAT2	PMS	NAT1	NAT2	PMS	NAT1	NAT2	PMS
9.86	529	783	695	64.9	96.1	85.3	1.6	4.8	2.6	5.4	6.7	3.6
6.99	1367	1469	1375	92.7	99.6	93.2	1.9	4.0	2.8	4.8	6.9	4.5
5.71	1743	1889	1800	92.0	99.7	95.0	1.9	3.6	3.0	4.9	6.2	5.5
4.95	2033	2263	2163	89.4	99.5	95.1	1.8	3.2	2.9	4.6	5.4	5.3
4.43	2283	2523	2381	90.3	99.8	94.2	1.8	3.1	2.9	4.6	5.4	5.0
4.04	2555	2805	2645	90.9	99.8	94.1	1.8	3.0	2.9	4.8	5.5	5.5
3.74	2742	3063	2859	88.8	99.2	92.6	1.8	2.9	2.8	5.0	6.2	6.7
3.50	2846	3190	2114	87.1	97.6	64.7	1.7	2.8	2.3	5.1	6.8	8.0
3.00	8627	9438	—	80.9	88.5	..	1.7	2.8	—	5.3	7.3	
2.50	14615	14132	—	69.6	67.3		1.7	2.8	.	6.1	9.3	..
Limit*	22072	6207		46.5	26.7	—	1.7	2.7	—	11.1	12.9	—
Total	61412	47762	16032	63.1	65.4	88.3	1.8	3.0	2.8	5.6	6.5	5.5

* Limit = 2.0 Å for NAT1 and 2.2 Å for NAT2.

and arrive at the correct solution. Such a procedure is described below.

Aldose reductase is a monomeric enzyme of 314 amino acids (MW = 36 kdalton) which catalyses the reduction of glucose into sorbitol, using NADPH as the coenzyme. For the following studies the enzyme isolated from pig lens (Reymann *et al.*, 1992) was used as it presents a very high sequence homology with human aldose reductase* (88% of strict identity). Since the intracellular accumulation of sorbitol is believed to be the cause of several complications of diabetes, the inhibition of the enzyme is clearly of pharmacological interest.

The enzyme crystallizes in three different forms (triclinic, tetragonal and hexagonal). For the purpose of this work, the triclinic form, which diffracts to 1.9 Å resolution, was used. The crystals were obtained from ammonium citrate buffer, and were easy to reproduce. Table 1 shows the crystallization conditions and the cell parameters.

2. Data analysis of the triclinic form and single isomorphous replacement phasing

2.1. Data collection

Two different native data sets were collected with a Siemens–Xentronics multiwire area detector at 263 K. The data sets were processed using *XENGEN* software (Nicolet Instrument Corporation). As shown in Table 2, the first set includes data to 2 Å but only 65% of the low-resolution reflections ($d > 10$ Å). The second set was collected to provide low-resolution completeness. It was 99% complete between 10 and 3.5 Å and 96% complete for the low-resolution reflections ($d > 10$ Å). It had a higher redundancy, but missed the high-resolution end (d

< 2.23 Å). These two data sets were not merged, since the relative R factor in amplitudes was 11%. They were, therefore, used separately as required.

2.2. Local symmetry

The volume of the unit cell agrees with the assumption of four molecules per asymmetric unit (V_m = 2.65 Å^3 dalton^{-1}). A 12–6 Å self-rotation function (Rondeau *et al.*, 1987) based on the native data was calculated using the program *PROTEIN* (Steigemann, 1974). Reflections with $I > 3\sigma(I)$ were used and only Patterson vectors between 30 and 8 Å were superimposed. The rotation space was sampled in steps of 5°. The resulting function clearly shows three orthogonal local twofold axes in the κ = 180° section (Fig. 1a). As a check, duplicate calculations were performed using *ALMN* (Dodson, 1985) and *MERLOT* (Fitzgerald, 1988), which gave similar results.

This calculation was repeated with data between 12 and 3 Å, using Patterson vectors of lengths between 30 and 5 Å. The rotation space was sampled around the previously found peaks in steps of 0.25°. Table 3(a) shows the result of this refinement. Even when peak 3 is somewhat extended, suggesting two values near 180° (Fig. 1b) that have coalesced, it was not possible to resolve this peak and obtain a more accurate κ value at this stage of the structure determination.

The self-rotation function was repeated using the program *ROTINTER*, based on the reciprocal space rotation function of Rossmann & Blow (1962). An initial search with data between 10 and 5 Å and a Patterson radius of 40 Å gave the same results as *PROTEIN*. A second search with 0.25° steps was conducted using data between 5 and 3 Å. The results are the same as obtained with *PROTEIN* to 0.4°, except for the value of φ_3, which differs by 1.6°, suggesting again that this rotation axis might be misplaced.

With this caveat, the result of both calculations is coherent with a 222 local symmetry, with screw translations to be determined (to differentiate this symmetry from a 222 point symmetry without translations, it will

* *Note added in proof*: After submission of this paper, a paper describing the 1.65 Å structure of the human aldose reductase holoenzyme and showing the very high similarity of the two structures appeared (Wilson, Bohren, Gabbay & Quiocho, 1992).

be called $2_12_12_1$ here). Since 222 point symmetry has been found in other dehydrogenases (Rossmann, Ford, Watson & Banaszak, 1972; Rossmann *et al.*, 1973), it seemed a plausible hypothesis at this stage.

2.3. *Derivative data*

A search for heavy-atom derivatives was conducted, and numerous compounds were tried. Since pig-lens aldose reductase has six cysteines, mercury derivatives were logical candidates. The only effective derivative was indeed a mercury salt, sodium phenylmercurysalicylate (PMS).* All other compounds either have no effect or crack the crystals. The fact that the mother liquor contains citrate

---

* The derivative is not EMTS, as mistakenly reported by Rondeau *et al.* (1992).
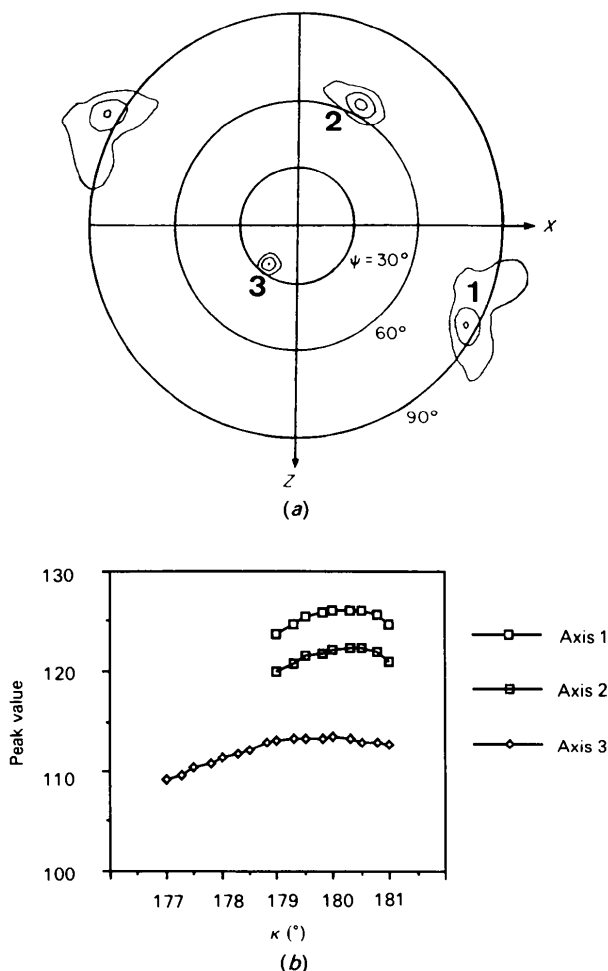
Table 3. *Comparison of* 222 *axes obtained from the self-rotation function and pseudo* 222 *axes obtained from rigid-body refinement*

The polar convention (Tanaka, 1977) is used. The screw translations *t* imply non-intersecting axes. The value *d* measures the distance of each axis to a middle point. The translation components are significant for axes 2 and 3 and the *κ* value for the third axis differs significantly from 180°, in agreement with the observation shown in Fig. 1(*b*).

	φ (°)	ψ (°)	κ (°)	t (Å)	d (Å)
(*a*) Self-rotation function					
Axis 1	330.15	86.45	180.0		
Axis 2	61.80	65.30	180.0		
Axis 3	232.50	25.00	180.0		
(*b*) Rigid-body refinement					
Axis 1	328.8	87.6	181.4	0.1	0.8
Axis 2	61.4	65.4	181.8	0.4	1.1
Axis 3	236.1	25.2	175.1	0.9	1.3

ions, which are strong chelating agents, is probably one cause of these difficulties. The classical solution of transferring crystals to another media did not work in this case.

Data for the PMS derivative were collected with an area detector at 263 K (see Table 2). Scale and temperature factors were applied to the derivative in order to minimize the isomorphous differences

$$\Delta_{\mathrm{iso}} = |F_{\mathrm{der}} - F_{\mathrm{nat}}|$$

with the first native data set. The mean value of $R_{\mathrm{mer}} = \Sigma\Delta_{\mathrm{iso}}/\Sigma F_{\mathrm{nat}}$ between 20 and 3.5 Å is 19.4 %, typical of a well substituted derivative. Lack of isomorphism appears at 4 Å, but it remains acceptable to 3.5 Å.

Since the space group P1 has no Harker sections, heavy-atom sites have to be identified from cross vectors. The difference Patterson map could not be manually interpreted, since there are 30 peaks above 3σ and 142 peaks above 2σ. This is characteristic of a multi-site derivative and, therefore, an automatic search procedure (*HASSP*; Terwilliger & Eisenberg, 1987) was used.

Table 4(*a*) shows the result of this search, in which nine correlated positions were found. All 36 cross vectors are in positive zones of the Patterson map and 50% of them are above 2σ, with a maximum of 5.4σ. A change of reference frame allowed the interpretation of these positions in terms of a $2_12_12_1$ local symmetry, with the rotation axes close to those observed in the self-rotation function. A residual difference Fourier map based on phases from these nine positions showed two additional peaks compatible with the $2_12_12_1$ symmetry, and a 12th position was assigned according to the local symmetry. This position was confirmed by a residual difference Fourier map based on the other 11 sites. These heavy-atom positions were refined using the program *HEAVY* (Terwilliger & Eisenberg, 1983) and the final result is shown in Table 4(*b*). Note that even when *HASSP* and *HEAVY* are both based on the use of an origin-removed difference Patterson map, the values of occupancies and *B* factors listed in Table 4(*b*)

Fig. 1. Self-rotation function. (*a*) Section $\kappa = 180°$ of the rotation function showing three perpendicular twofold axes. (*b*) Peak self-rotation function values of the three twofold axes shown in (*a*) corresponding to κ sections near 180°.

Table 4. *Description of heavy-atom location, refinement and phasing*

(a) Patterson cross peaks
The cross peaks were obtained by a systematic search with *HASSP* (Terwilliger & Eisenberg, 1987). The σ value is 1000. The numbering corresponds to the local symmetry classification shown in (b). Peaks 2, 3 and 9 were not shown in this search and were positioned as described in the text.

Site	1	4	5	6	7	8	10	11	12
1	—	2697	2422	4205	1671	5275	4432	2928	3637
4	2697	–	1393	2277	1242	68	1903	1377	1694
5	2422	1393		2043	1231	5418	1027	970	850
6	4205	2277	2043	—	4858	2484	3867	1679	2957
7	1671	1242	1231	4858	··	840	1496	1478	1385
8	5275	68	5418	2484	840	—	2698	1433	1717
10	4432	1903	1027	3867	1496	2698		3040	3847
11	2928	1377	970	1679	1478	1433	3040	—·	2235
12	3637	1694	850	2957	1385	1717	3847	2235	—

(b) Heavy-atom parameters
The parameters were obtained by the program *HEAVY* (OCC_{ini} and B_{ini}) and subsequent refinement against averaged phases (OCC_{fin} and B_{fin}) with data between 20 and 4 Å. A first set of 12 cycles refined the global scale, the positions and the occupancies, with B's fixed arbitrarily to 20 Å^2. This was followed by 12 cycles of refinement of global scale and B. As the origin is arbitrary for the space group $P1$, it was placed at the initial barycenter of the 12 sites, and kept fixed during the whole refinement, to avoid a continuous deviation of the procedure. The 12 sites have four groups A, B, C and D which are related by local symmetry and correspond to the protein monomers. There are four local symmetry-related strong sites (1, 4, 7 and 10) which are clearly distinguishable from the other eight.

Group	Site	x	y	z	OCC_{ini}	OCC_{fin}	B_{ini} (Å^2)	B_{fin} (Å^2)
A	1	−0.2220	0.2482	0.1686	13.1	13.7	21.3	17.8
	2	−0.1397	0.2489	0.1831	5.5	5.6	22.5	18.2
	3	−0.2362	0.2347	0.0028	3.5	3.4	16.7	15.0
B	4	0.0259	−0.2820	−0.3273	8.3	8.2	16.6	18.0
	5	−0.0744	0.2660	··0.2499	6.3	4.8	22.4	20.2
	6	−0.0517	−0.2785	−0.2370	6.5	6.5	15.0	14.5
C	7	·0.0027	0.3253	0.1426	9.3	11.1	18.3	17.3
	8	· 0.0565	0.3018	0.1214	5.4	4.6	17.5	20.0
	9	0.0422	0.2794	0.2580	3.6	4.2	20.7	17.5
D	10	0.1937	−0.2927	0.0102	9.9	6.8	21.3	21.4
	11	0.1257	−0.2747	−0.0575	5.1	5.3	20.2	20.0
	12	0.2401	−0.2257	·0.0538	3.3	5.6	14.5	17.7

(c) Superposition of equivalent heavy atoms
Root-mean-square (r.m.s.) values for the superposition were obtained using the positions in (b) calculated by the program *HEAVY*. Three types of non-crystallographic symmetry are analyzed: first, the 222 axes, which are used for the first averaging process, second, the 2,2,2, system corresponding to the best superposition of equivalent sites, and third, the final 2,2,2, system, which is obtained after rigid-body refinement.

Non-crystallographic symmetry	R.m.s. of superposition (Å)		
	Axis 1	Axis 2	Axis 3
222	0.56	1.23	1.26
Best 2,2,2,	0.49	0.93	0.94
Final 2,2,2,	0.58	1.33	1.36

(d) SIR phasing statistics as a function of resolution
Note that the figure of merit and the phasing power are rather uniform, in agreement with the curve shown in Fig. 2(a) which is quite uniform with the exception of the very highest resolution bin.

									Total
d_{max} (Å)	24.94	13.30	9.07	6.88	5.54	4.64	3.99	3.50	
Number of reflections	16	220	650	1266	2072	3034	4175	4516	15949
Figure of merit	0.46	0.43	0.50	0.47	0.47	0.46	0.45	0.44	0.45
Phasing power	1.57	1.98	2.03	1.94	1.88	1.76	1.60	1.53	1.70

(from *HEAVY*) do not correspond to the height of cross vectors (from *HASSP*) listed in Table 4(a). However, this is only an apparent inconsistency. *HASSP* does not refine the global agreement of the calculated and observed Patterson maps. It calculates the Patterson values at all interatomic vectors, identifies the minimum of these values and adjusts the positions of the sites to maximize this minimum value only. Thus, the values of the difference Patterson function which are listed for the other cross vectors cannot correspond to the maximum of the peaks, but only to their edge.

A systematic search of possible compact tetramers was conducted by combining the local symmetry obtained from the heavy-atom constellation with the lattice translations. This search produced 30 possibilities with at least one translation close to zero. Of these, one had all three translations close to zero ($t_1 = 0.51$, $t_2 = 0.49$, $t_3 = 0.42$ Å). When an exact 222 point symmetry with the orientation defined by the self-rotation function was fitted to this solution, the heavy-atom constellation agreed within an r.m.s. error of 0.56 for the first axis, 1.23 for the second and 1.26 Å for the third (see Table 4c). Since the best possible choice of

$2_12_12_1$ symmetry had a corresponding error of 0.49, 0.93 and 0.94 Å, it was decided that these two solutions were equivalent within the accuracy of the heavy-atom positions and, therefore, the local 222 symmetry without translations was chosen. This decision was later confirmed by analysis of the r.m.s. errors using the final $2_12_12_1$ symmetry (see Table 4c), which are comparable to those of the 222 symmetry and show, therefore, that an optimization of the symmetry based on the heavy-atom positions would have led to the wrong axes.

2.4. Single isomorphous replacement phasing

The heavy-atom data and parameters were used to calculate phases in the 30–3.5 Å range using the program PHARE from the CCP4 suite (SERC Daresbury Laboratory, 1986). Table 4(d) shows the figure of merit and phasing power as a function of resolution; they are typical of single isomorphous replacement phasing. The overall figure of merit of 0.45 suggests a 70° mean phase error. This was confirmed a posteriori by comparison of the single isomorphous replacement (SIR) phases with the current model, which shows that the mean phase error $\langle|\varphi_{SIR} - \varphi_{model}|\rangle$ is 70.6°. The resolution dependence is shown in Fig. 2 (curve a). Fig. 3(a) shows a global view of the Fourier transform of $mF_{obs}\exp(i\varphi_{SIR})$, where m is the SIR figure of merit. In this electron-density map the molecular and solvent regions can be distinguished. Fig. 4(a) shows a local view of the same map, superimposed with the final model in an α-helical region. The map had too many discontinuities to be interpretable, as can be expected from the large phase error. A procedure of phase improvement was therefore necessary.
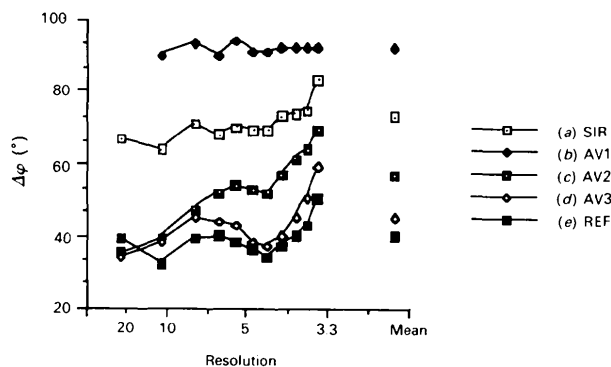
3. Real-space phase refinement

In order to improve the quality of the electron-density map, a procedure of real-space density modification was implemented. The constraints imposed on the density included the local symmetry inside the molecular envelope and the flatness of the solvent outside. The solvent content of the crystals, as estimated from Matthews' formula (Matthews, 1974), was 51.2%. The programs used were from the program package RMOL (Rees, Bilwes, Samama & Moras, 1990).

The determination of the correct local symmetry elements and tetramer envelope, which seemed clear at this
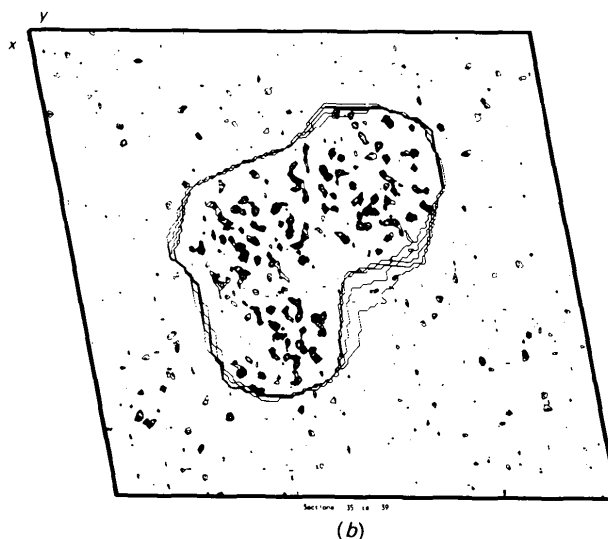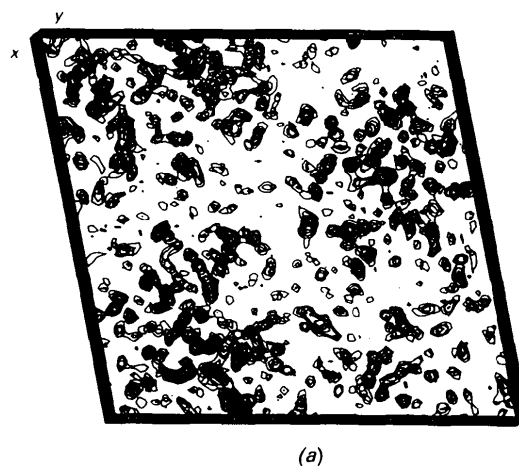

(a)

(b)

Fig. 3. (a) Slab of the SIR map calculated from Hg heavy-atom parameters refined by HEAVY. (b) Correlation-weighted map. The values of the SIR electron density are averaged and multiplied by the local correlation function. To show the boundaries clearly, eight unit cells are represented. The envelope obtained by Wang's procedure (Wang, 1985) from this correlation-weighted map is shown.

Fig. 2. Values of $\langle\varphi_{obs} - \varphi_{model}\rangle$ as a function of resolution for different stages of the structure solution. The mean value over the whole resolution range is also shown. The model was obtained from the 2.5 Å refinement. (a) φ_{obs} obtained from SIR phasing with the Hg derivative (SIR). (b) φ_{obs} obtained from the averaging described in §3.1 (AV1). (c) φ_{obs} obtained from the averaging described in §3.2 (AV2). (d) φ_{obs} obtained from the averaging described in §3.3 before heavy-atom re-refinement (AV3). (e) φ_{obs} obtained from the averaging described in §3.3 after heavy-atom re-refinement (REF).

stage, was very approximate. Therefore, several attempts were necessary until an accurate enough set was found. These attempts are described below and summarized in the flowchart shown in Fig. 5. The corresponding maps are shown in Fig. 4.
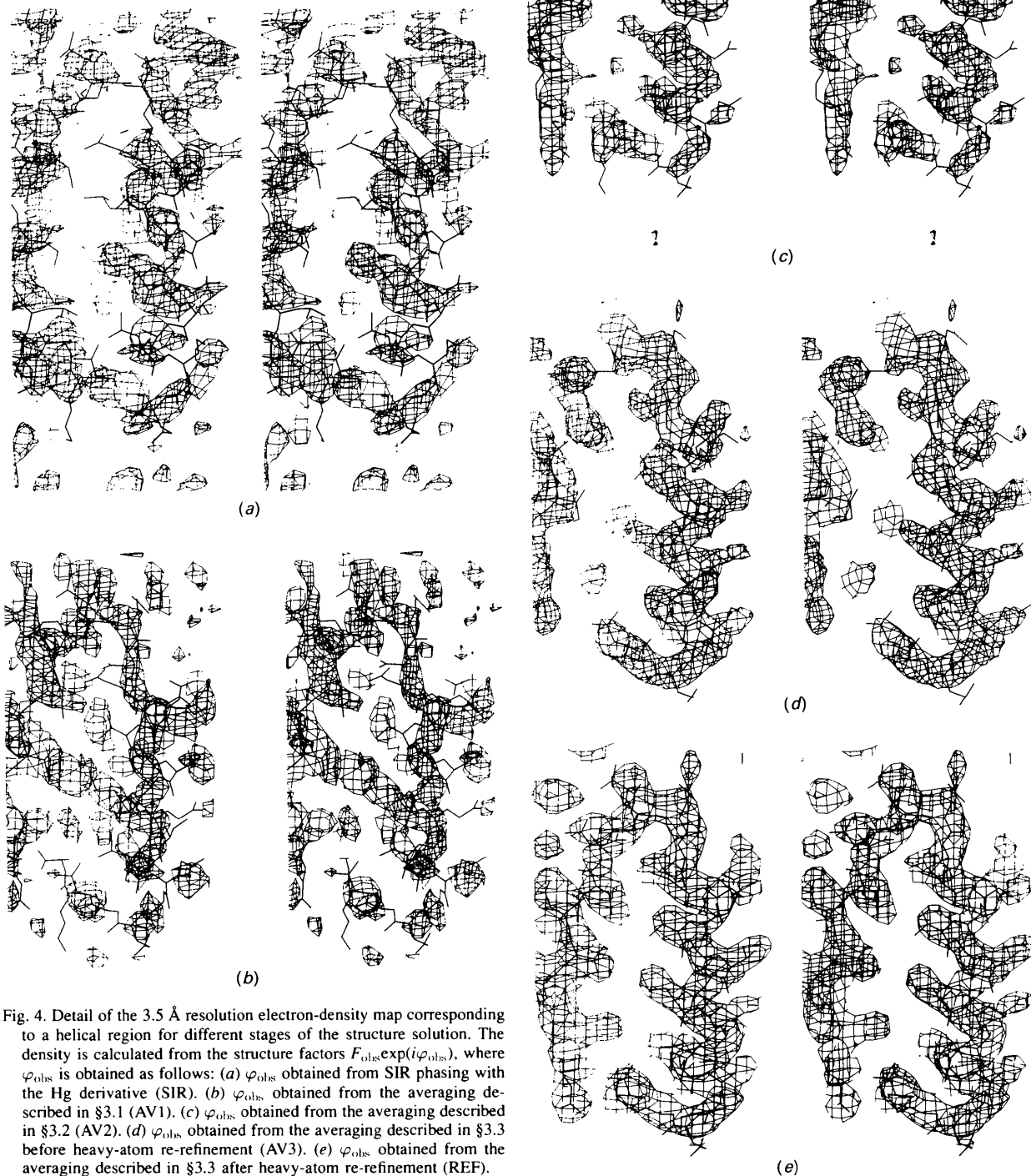

(a)

(b)

(c)

(d)

(e)

Fig. 4. Detail of the 3.5 Å resolution electron-density map corresponding to a helical region for different stages of the structure solution. The density is calculated from the structure factors $F_{obs}\exp(i\varphi_{obs})$, where φ_{obs} is obtained as follows: (a) φ_{obs} obtained from SIR phasing with the Hg derivative (SIR). (b) φ_{obs} obtained from the averaging described in §3.1 (AV1). (c) φ_{obs} obtained from the averaging described in §3.2 (AV2). (d) φ_{obs} obtained from the averaging described in §3.3 before heavy-atom re-refinement (AV3). (e) φ_{obs} obtained from the averaging described in §3.3 after heavy-atom re-refinement (REF).

3.1. *First attempt: 222 local symmetry, automatically determined molecular envelope and incomplete low-resolution shells*

Since the 222 symmetry obtained from the heavy-atom positions has no translations and extends to the whole tetramer, the distinction between monomers is not necessary for averaging purposes. A global envelope is therefore sufficient. Even when the SIR map shows the solvent regions needed for the calculation of this envelope, the boundaries are not sharp enough and there are ambiguities in the packing zones. In order to enhance the density of the molecular region and distinguish one single tetramer, the electron density was weighted by a correlation function based on the local symmetry* (Rees *et al.*, 1990)

$$\Gamma(X) = \left\{ \left[\sum_{i=1}^{n} \rho(r_i) \right]^2 / \sum_{i=1}^{n} \rho(r_i)^2 - 1 \right\} / (n - 1),$$

with $n = 4$. Fig. 3(*b*) shows the correlation-weighted map calculated over eight unit cells. This showed very clearly the molecular boundaries of the tetramer [for previous examples of the same principle, see Rossmann (1990)]. The map was then convoluted with a linear weighting sphere of 10 Å radius (Leslie, 1987) and a cut-off, corresponding to a solvent content of 40%, was applied. After 'tidying-up' of isolated density islands, this led to a solvent content of 45%.

---

* Note that trials to improve the local symmetry by monomer correlation in the SIR map did not show any clear signal.

SIR Phasing — 70° error

Averaging using :
222 Symmetry
45% solvent mask
Incomplete data — 90° error

**Mask improvement
Complete data**

Averaging using :
222 Symmetry
17% solvent mask
Extended data — 55° error

Exact symmetry

Averaging using :
Pseudo 222 Symmetry
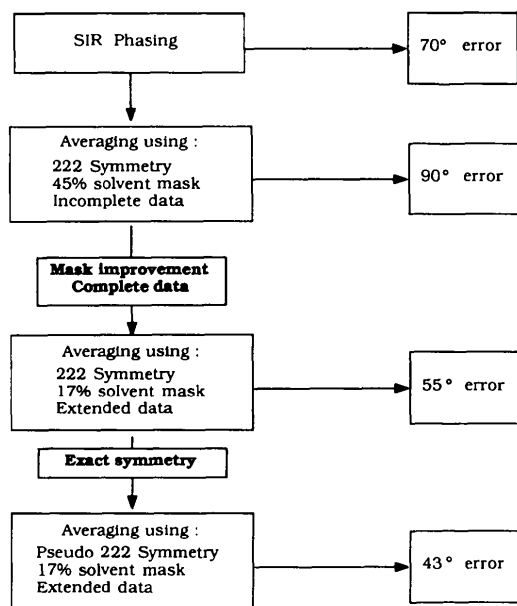17% solvent mask
Extended data — 43° error

Fig. 5. Flowchart describing the three different attempts at real-space refinement.

An averaging procedure was performed using native data set 1, incomplete at low resolution. At each cycle, only observed data between 12 and 3.5 Å were used in the calculation of electron-density maps. Exact 222 symmetry was applied to replace each density point inside the molecular envelope by the average of the four symmetry-related points. Densities inside the molecular envelope which were either too high or too low were attenuated. The average value of the solvent zone was very close to zero, and, therefore, all solvent points were set to zero. Calculated structure factors were obtained by Fourier inversion of averaged and solvent flattened maps, and a Sim weighting scheme was used to assign a probability to φ_{calc}. This phase was then merged with φ_{SIR} with equal weights, and the resulting phase φ_{mer} was used together with F_{nat} to calculate a new map. The procedure was reiterated until convergence of φ_{calc}.

Figs. 6(*a*) (curve 1) and 6(*b*) (curve 1) show the averaging statistics. Even when the correlation between monomers increases, the R factor between F_{calc} from the modified map and F_{nat} remains large (33%). A detailed view of the Fourier transform of $mF_{obs} \exp(i\varphi_{mer})$ where m is the merged figure of merit (Fig. 4*b*) shows that the electron-density map is clearly inferior. This can be confirmed *a posteriori* by calculating the mean phase difference with model phases, which increases from 70.6 to 90°, showing that the procedure has diverged. Fig. 2 curve (*b*) shows this phase error $\langle \varphi_{mer} - \varphi_{model} \rangle$ as a function of resolution.

An analysis of the possible reasons for this failure of phase improvement is given in §4. The conclusion was that the envelope was too small. Furthermore, a test case with Gaussian spheres of $B = 10\,000$ Å^2 (poor diffraction beyond 10 Å) centered at the monomer positions showed that a low-resolution cut-off led to poor monomer correlation (*e.g.* an 18 Å cut-off corresponded to 68% monomer correlation) leading to a lack of local symmetry in the truncated map. This showed the importance of including low-resolution terms. Both these conclusions were applied in the next attempt at phase improvement.

3.2. *Second attempt: 222 local symmetry, improved molecular envelope and low-resolution terms*

The second native data set, including more low-resolution terms, was used and a lower solvent level was imposed in an attempt to increase the size of the molecular region. This led to an envelope which corresponded to a solvent content of 31%. Because of the extreme sharpness of the local symmetry correlation, this number could not be further diminished by imposing an even lower solvent level. Therefore, the procedure of envelope enlarging implemented in *RMOL*, which consists of the addition of surface layers, was employed. By adding an extra 2 Å thick layer to the molecular region the solvent content was diminished to 16.9%.

Figs. 6(a) (curve 2) and 6(b) (curve 2) show the statistics of the refinement. Even when the R factor between F_{calc} from the modified map and F_{nat} remained large (34%), the a posteriori calculation of the phase error showed a decrease from 70.9 to 56°. Therefore, at this stage the procedure has begun to converge to the correct minimum.

To improve the conditions for averaging further (as discussed above), non-phased or non-measured low-resolution reflections (83 > d > 10 Å) were estimated (either in phase only or in both amplitude and phase) from the Fourier inverse of modified maps and included in a new averaging procedure. The combined effect of a better envelope and the low-resolution phasing clearly improved the map, as shown in a detailed view (Fig. 4c) of the Fourier transform of $mF_{obs}\exp(i\varphi_{mer})$, which is clearly interpretable in terms of a helix. Overall, the 3.5 Å averaged map clearly showed main-chain density for 250 residues, corresponding to about 80% of the total

number. In addition, clear side-chain densities enabled the positioning of 60% of the total number of atoms of one monomer. This model was extended by local symmetry to the full tetramer.

The quality of the electron-density map indicated a lower phase error, confirmed a posteriori by a $\langle\Delta\varphi\rangle$ of 54.8° (Fig. 2, curve c). However, attempts at phase extension failed to improve the quality of the map. This failure was later confirmed by phase-error analysis, which showed that the mean phase difference with model phases in the 3.5–3.07 Å range was 74°.

Several attempts to improve this phase extension failed, pointing to an error in the averaging parameters. This was solved by redefining the local symmetry axes, as described below.

3.3. Third attempt: pseudo 222 local symmetry, improved molecular envelope, low-resolution terms, heavy-atom re-refinement and phase extension

Using the partial model defined above as a rigid body, refinement of the four independent monomers was performed using data between 15 and 6 Å resolution. This refinement reduced the starting R factor of 43.8 to 40.9%, and, most importantly, it displaced the positions and orientations of the local axes significantly, (Table 3b), changing the symmetry from exact 222 to pseudo $2_t2_t2_t$. In this arrangement, the axes are not exactly twofold, they have a translation component t in the direction of the axis and they do not intersect.

A new averaging procedure was started with the new symmetry and original heavy-atom parameters from HEAVY. Since the symmetry is no longer 222 but pseudo $2_t2_t2_t$, the local symmetry is no longer a point group and each symmetry operation acts only on the corresponding monomer. This requires the definition of separate envelopes for each monomer. The envelopes were obtained from the partial model, and zones which were obviously empty and corresponding to uninterpreted portions of the map were filled with dummy atoms. The final tetramer envelope corresponded to a solvent content of 19.9%.

The averaging was carried out as described previously, with data between 83 and 3.5 Å and low-resolution unobserved reflections inserted from the inverse Fourier transform. The phases were clearly improved ($\langle\Delta\varphi\rangle$ = 43°), especially at medium and high resolution (Fig. 2, curve d), and this was reflected in clearly improved maps (Fig. 4d).

In order to improve the starting point, map averaging at 3.5 Å was combined with re-refinement of heavy-atom parameters against averaged phases. This changed the initial heavy-atom parameters, in particular the occupancies (see Table 4b), and led to an even better map (Fig. 4e) which was very close to the calculated one. Fig. 2 (curve e) shows the phase improvement ($\langle\Delta\varphi\rangle$ = 38°). Figs. 6(a)
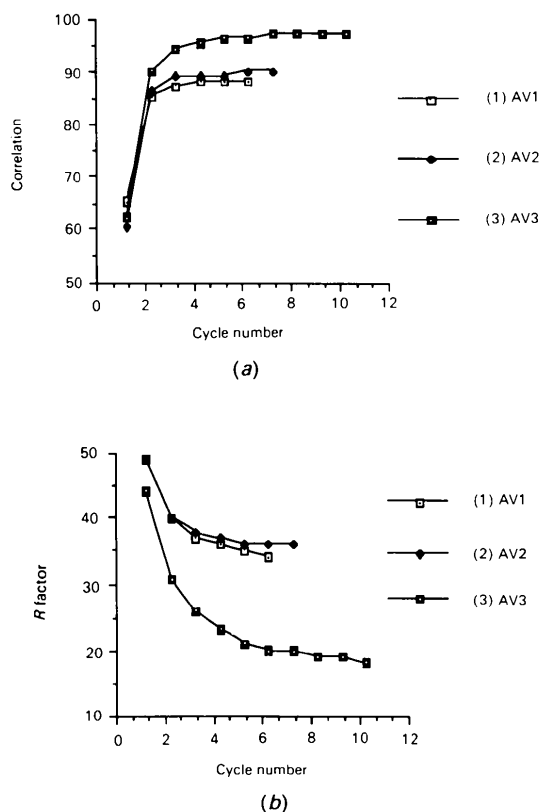

Fig. 6. Statistics of the averaging procedure as a function of cycle number during this procedure. (a) Mean value of the correlation between local symmetry-related monomers: (1) averaging procedure described in §3.1 (AV1), (2) averaging procedure described in §3.2 (AV2), (3) averaging procedure described in §3.3 (AV3). (b) R factor between F_{obs} and F_{calc} obtained from the inverse Fourier transform of the averaged map: (1) averaging procedure described in §3.1 (AV1), (2) averaging procedure described in §3.2 (AV2), (3) averaging procedure described in §3.3 (AV3).

(curve 3) and 6(b) (curve 3) show the effect of improved local symmetry and heavy-atom re-refinement in the averaging statistics; the most obvious effect of the improved symmetry is on the R factor between F_{calc} and F_{nat}, which dropped to 18%.

Based on the new symmetry elements and exact envelope a phase-extension procedure to 3.07 Å was performed. The reciprocal space between 3.5 and 3.07 Å was divided into 15 radial steps. Each step was subjected to an averaging procedure in which the corresponding amplitudes were phased. This led to an improved map as shown by a phase error of 44° in the 3.5–3.07 Å range as opposed to 74° with exact 222 symmetry. This map was interpreted leading to a complete model, which was properly aligned with the sequence. The model was refined to an R factor of 21.9% at 2.5 Å resolution. The identification of solvent molecules and the final refinement to 1.9 Å resolution is underway and will be published elsewhere. The model shows a $(\beta/\alpha)_8$ barrel fold not previously found in adenine dinucleotide dependent enzymes (Rondeau et al., 1992).

4. Discussion

During the process described above, two steps proved particularly difficult: definition of the molecular boundaries and accurate definition of the local symmetry.

4.1. Mask definition

The first averaging process diverged, and this was corrected by enlarging the tetramer mask. This mask seemed very clear from the correlation-weighted map, and it indicated a solvent content of 45%, compared with the predicted value of 51.2%. Since the envelope is a low-resolution object which cannot follow the exact shape of the surface to atomic resolution, it should include a molecular volume 20–30% larger than that indicated by Matthews' formula, and the solvent content in this case should then be between 20 and 30%. The solvent content of 45% used in the first averaging procedure was therefore too large. This can be confirmed a posteriori by superimposing the envelope with the final model (Fig. 7a). This superposition shows that in the model the external side chains are consistently outside the envelope. Therefore, the clear boundaries shown by the correlation-weighted map were misleading. This is most likely due to the error in the local symmetry axes. This error lowered the correlation at the borders, since the incorrect symmetry operations superimposed a border protein region with a solvent region. The net effect of this incorrect envelope was to cut the protein region, thus leading to the divergence of the phase refinement.

This problem was corrected by the addition of an extra layer to the molecular region. This layer needed to be large enough to prevent all cutting of the protein region. When a model was available, it was confirmed that indeed this new envelope does not cut the model (Fig. 7b). Since the main phasing power of the procedure arises from monomer averaging and not from the solvent flattening, the low solvent content of this extended mask did not prevent the success of the procedure.

### 4.2. Very low resolution data

The inclusion of very low resolution data, either observed or calculated from inverse Fourier transforms, is clearly helpful in the early stages of the phase refinement. It should, however, be noted that the extended mask is calculated without these terms and that extra tests have shown that if the exact symmetry is used the averaging procedure converges even without the low-resolution terms.

### 4.3. Calculation of the exact local symmetry (pseudo $2_12_12_1$)

The self-rotation function indicated a point 222 symmetry, confirmed by the interpretation of the heavy-atom constellation. This symmetry was only approximate, since the exact symmetry includes translations and rotation axes that are close to, but not exactly, twofold as shown in Table 3(b).

There were some indications of this problem. One of the self-rotation function peaks (peak 3) was elongated. The averaging R factor was higher than 30%, and phase extension beyond 3.5 Å failed. Even when all these factors signalled an error in the local axes, correlation methods either in Patterson or Fourier space failed to point towards the correct axes. The accuracy of Patterson superposition methods suffers in general from the increased radius of the Patterson peaks as compared to the Fourier peaks. In this particular case it also suffered from the superposition of symmetry-related peaks since the correct orientation is close to a mirror plane in Patterson space. Correlation methods in Fourier space suffered from the fact that SIR phases were calculated from a heavy-atom constellation which fitted the erroneous 222 symmetry. This is shown by the first value of the correlation during the different averaging procedures (Fig. 6) which is not sensitive to the improvement in the local symmetry axes. In contrast the final value is clearly better for the correct symmetry. Since the correlation methods were not accurate enough, it was only the rigid-body refinement that enabled the calculation of the exact symmetry.

The question of a correct method for obtaining the local symmetry axes, without prior knowledge of a model, remains open. A possibility could be a systematic search of possible orientations where a few averaging cycles are performed at each orientation and the R factor computed. It remains a large but not impossible computing task and the sensitivity of the R factor (which dropped in this case by almost half) to the correct axes suggests that it may work.
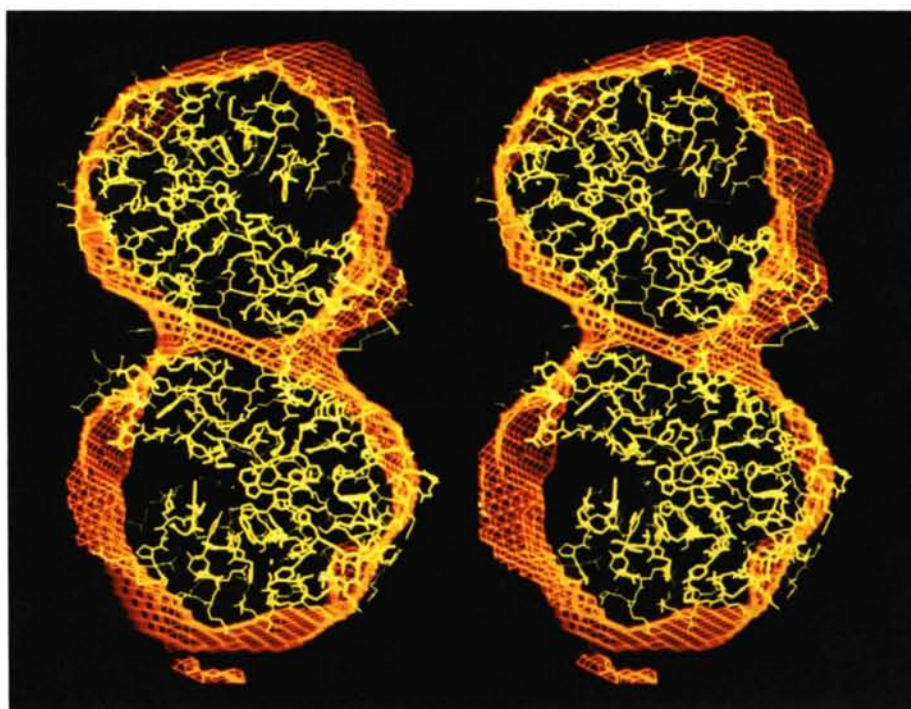
(a)
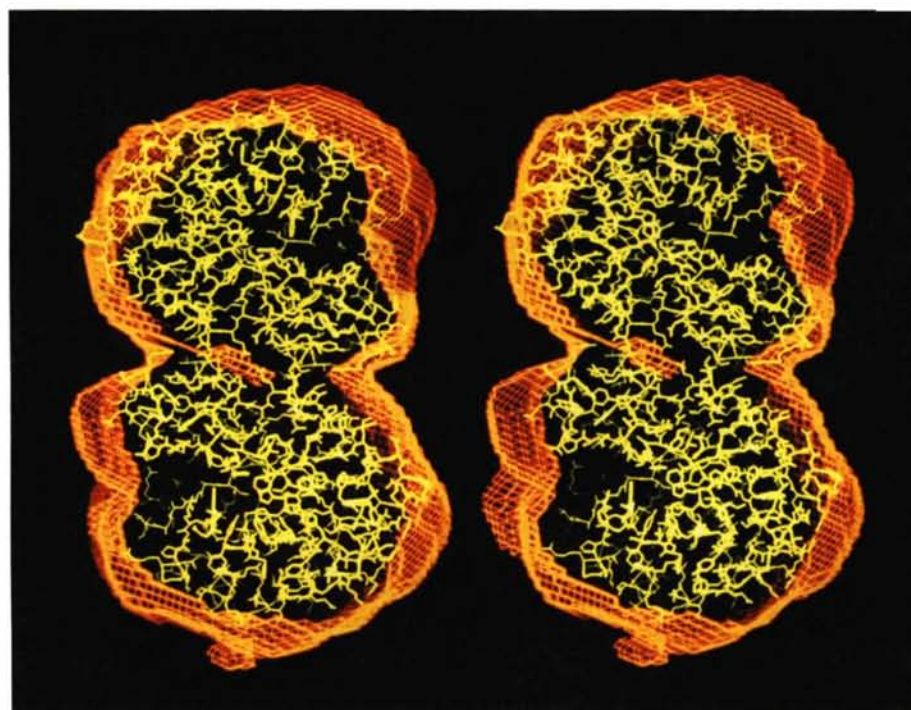
(b)

Fig. 7. Superposition of the tetramer envelope and the 2.5 Å refined model. (a) 45% solvent envelope utilized for the averaging procedure described in §3.1. (b) 20% solvent envelope utilized for the averaging procedure described in §3.2.

5. Concluding remarks

Local symmetry averaging is a very powerful technique, but its success depends on the quality of the initial parameters, such as the accuracy of the axis, the completeness of the amplitude data set, the quality of the initial phasing and the accuracy of the envelope. These parameters are correlated and they are only approximate at the beginning of the phasing procedure.

In the case described, the main problem was the local symmetry, which is approximately but not exactly 222. Use of an approximate symmetry in defining the envelope by local correlation led to a reduced-size envelope and to a diverging averaging procedure.

The problem was corrected by an expansion of the envelope, which led to a converging averaging procedure and to an adequate phase set to produce an initial partially interpretable map. Knowledge of the partial model enabled the definition of the correct symmetry elements and led to the final structure solution.

The problem of finding the correct symmetry elements directly is not straightforward. Neither Patterson methods, heavy-atom positions nor monomer density correlation in SIR maps seem to be accurate enough. The sensitive parameter is the R factor between F_{calc} and F_{nat} after averaging; therefore, a search procedure based on a few cycles of averaging for each symmetry element could be envisaged.

The solution of the structure of aldose reductase is an example of how the parameters defining non-crystallographic averaging can be changed during the procedure to obtain the best possible phase set.*

References

BRICOGNE, G. (1976). *Acta Cryst.* A32, 832–847.
BUEHNER, M., FORD, G. C., MORAS, D., OLSEN, K. W. & ROSSMANN, M. G. (1974). *J. Mol. Biol.* 82, 563–585.
DODSON, E. J. (1985). *Molecular Replacement*, edited by P. A. MACHIN, pp. 33–45. Warrington: SERC Daresbury Laboratory.
FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* 21, 273–278.
HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, 276, 368–373.
JONES, E. Y., WALKER, N. P. C. & STUART, D. I. (1991). *Acta Cryst.* A47, 753–770.
LESLIE, A. G. W. (1987). *Acta Cryst.* A43, 134–136.
MATTHEWS, B. W. (1974). *J. Mol. Biol.* 82, 513–526.
RAYMENT, I. (1983). *Acta Cryst.* A39, 102–116.
REES, B., BILWES, A., SAMAMA, J.-P. & MORAS, D. (1990). *J. Mol. Biol.* 214, 281–297.
REYMANN, J.-M., RONDEAU, J.-M., BARTH, P., JACQUINOD, M., VAN DORSSELAER, A. & BIELLMANN, J.-F. (1992). Submitted.
RONDEAU, J.-M., SAMAMA, J.-P., SAMAMA, B., BARTH P., MORAS, D. & BIELLMANN, J.-F. (1987). *J. Mol. Biol.* 195, 945–948.
RONDEAU, J.-M., TÊTE-FAVIER, F., PODJARNY, A. D., REYMANN, J.-M., BARTH, P., BIELLMANN, J.-F. & MORAS, D. (1992). *Nature (London)*, 355, 469–472.
ROSSMANN, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon and Breach.
ROSSMANN, M. G. (1990). *Acta Cryst.* A46, 73–82.
ROSSMANN, M. G., ADAMS, M. J., BUEHNER, M., FORD, G. C., HACKERT, M. L., LILJAS, A., RAO, S. T., BANASZAK, L. J., HILL, E., TSERNOGLOU, D. & WEBB, L. (1973). *J. Mol. Biol.* 76, 533–537.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* 15, 24–31.
ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* 16, 39–45.
ROSSMANN, M. G., FORD, G. C., WATSON, H. C. & BANASZAK, L. J. (1972). *J. Mol. Biol.* 64, 237–249.
SERC Daresbury Laboratory (1986). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington, England.
STEIGEMANN, W. (1974). PhD thesis, Technical Univ. Munich, Germany.
TANAKA, N. (1977). *Acta Cryst.* A33, 191–193.
TERWILLIGER, T. C. & EISENBERG, D. (1983). *Acta Cryst.* A39, 813–817.
TERWILLIGER, T. C. & EISENBERG, D. (1987). *Acta Cryst.* A43, 1–5.
WANG, B. C. (1987). *Methods Enzymol.* 115, 90–112.
WILSON, D., BOHREN, K. M., GABBAY, K. H. & QUIOCHO, F. A. (1992). *Science*, 257, 401–404.