DING, J., KOELLNER, G., GRUNERT, H.-P. & SAENGER, W. (1991). *J. Biol. Chem.* In the press.
EGAMI, F., OSHIMA, T. & UCHIDA, T. (1980). *Mol. Biol. Biochem. Biophys.* **32**, 250–277.
FINZEL, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.
HEINEMANN, U. & HAHN, U. (1989). *Protein–Nucleic Acid Interaction*, edited by W. SAENGER, & U. HEINEMANN, pp. 111–141. London: Macmillan.
HEINEMANN, U. & SAENGER, W. (1982). *Nature (London)*, **299**, 27–31.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
HUBBARD, R. & EVANS, P. (1985). *FRODO*. Version E4.4. MRC Laboratory of Molecular Biology, Cambridge, England.
ITAYA, M. & INOUE, Y. (1982). *Biochem. J.* **207**, 357–362.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
KELLER, E. (1988). *SCHAKAL88*. Kristallographisches Institut der Albert-Ludwigs-Univ., Freiburg, Germany.
KOEPKE, J., MASLOWSKA, M., HEINEMANN, U. & SAENGER, W. (1989). *J. Mol. Biol.* **206**, 475–488.
KOSTREWA, D., CHOE, H.-W., HEINEMANN, U. & SAENGER, W. (1989). *Biochemistry*, **28**, 7592–7600.
KUNDROT, C. E. & RICHARDS, F. M. (1987). *Acta Cryst.* B**43**, 544–547.
LENZ, A., CORDES, F., HEINEMANN, U. & SAENGER, W. (1991). *J. Biol. Chem.* **266**, 7661–7667.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MARTINEZ-OYANEDEL, J., CHOE, H.-W., HEINEMANN, U. & SAENGER, W. (1991). *J. Mol. Biol.* **222**, 335–352.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
PACE, C. N. & GRIMSLEY, G. R. (1988). *Q. Rev. Chem. Soc.* **22**, 3242–3246.
PACE, C. N., HEINEMANN, U., HAHN, U. & SAENGER, W. (1991). *Angew. Chem.* **103**, 351–369.
QUAAS, R., MCKEOWN, Y., STANSSENS, P., FRANK, R., BLÖCKER, H. & HAHN, U. (1988). *Eur. J. Biochem.* **173**, 617–622.
SHERIFF, S. (1987). *J. Appl. Cryst.* **20**, 55–57.
SMITH, J. L., HENDRICKSON, W. A., HONZATKO, R. B. & SHERIFF, S. (1986). *Biochemistry*, **25**, 5018–5027.
SUSSMAN, J. L. (1985). *Methods Enzymol.* **115**, 271–303.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Proteins Struct. Funct. Genet.* **1**, 370–375.
TAKAHASHI, K. & MOORE, S. (1982). *The Enzymes*, Vol. 15, pp. 435–468. New York: Academic Press.
TAKEUCHI, H., SATOH, Y. & HARADA, I. (1991). *J. Mol. Struct.* **242**, 49–59.
UCHIDA, T. & EGAMI, F. (1971). *The Enzymes*, 3rd ed., Vol. 4, pp. 205–227. New York: Academic Press.
YU, H.-A., KARPLUS, M. & HENDRICKSON, W. A. (1985). *Acta Cryst.* B**41**, 191–201.

Acta Cryst. (1992). B**48**, 191–199

Structure Determination and Refinement of Homotetrameric Hemoglobin from *Urechis caupo* at 2.5 Å Resolution

BY PRASANNA R. KOLATKAR,* STEPHEN R. ERNST AND MARVIN L. HACKERT†

Clayton Foundation Biochemical Institute, Department of Chemistry and Biochemistry, University of Texas, Austin, TX 78712, *USA*

CRAIG M. OGATA AND WAYNE A. HENDRICKSON

Howard Hughes Medical Institute and Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, *USA*

AND ETHAN A. MERRITT‡ AND R. PAUL PHIZACKERLEY

Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, CA 94309, *USA*

(*Received* 25 *April* 1991; *accepted* 15 *October* 1991)

Abstract

A 5 Å resolution multiple isomorphous replacement solution for hemoglobin isolated from *Urechis caupo* revealed a previously unobserved quaternary structure for tetrameric hemoglobin [Kolatkar, Meador, Stanfield & Hackert (1988). *J. Biol. Chem.* **263**(7), 3462–3465]. We report here the structure of *Urechis* hemoglobin in the cyanomet state refined to 2.5 Å resolution by simulated annealing yielding $R = 0.148$ for reflections $F > 3\sigma$ between 5.0 and 2.5 Å resolution. The starting model was fitted to a map originally derived from multiple-wavelength anomalous-dispersion phases to 3 Å resolution that was then subjected to cyclic twofold molecular

* Present address: Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.
† Author for correspondence.
‡ Present address: Department of Biological Structure, University of Washington, Seattle, WA 98105, USA.

0108-7681/92/020191-09$03.00

averaging and solvent flattening. Structural analysis of the resultant model shows that the unique quaternary assemblage is possible due to several favorable interactions between subunits, including salt links, hydrophobic pockets and interactions mediated by bound water. The tetramer is stabilized by subunit–subunit interactions between the *G/H* turns and *D* helices within the crystallographic dimer, and the *A/B* turn regions and *E* helices between subunits related by a molecular twofold axis. Interestingly, each subunit has one cysteine residue (Cys21) located in the *A/B* turn. These twofold-related cysteinyl residues are near enough to one another to form a disulfide bridge but do not.

Introduction

Oxygen transport in the echiuran innkeeper worm, *Urechis caupo*, is supported by a homotetrameric hemoglobin (62700 daltons) found in red cells (Garey & Riggs, 1984). This hemoglobin exhibits unusual properties when compared to other tetrameric hemoglobins in that cooperativity of oxygen binding is minimal and there is an absence of the Bohr effect. These properties are present in most hemoglobins from vertebrate as well as many other invertebrate sources. Sequence comparison with human β-hemoglobin shows 19% identity (Garey & Riggs, 1986). There is a similar lack of identity when compared to hemoglobin sequences from other invertebrates.

Crystals of *U. caupo* hemoglobin belong to the orthorhombic space group $C222_1$ ($a = 104.8$, $b = 54.9$, $c = 110.6$ Å) and contain one dimer per asymmetric unit. In this paper we report the structure determination of *U. caupo* hemoglobin in the cyanomet state with data refined at 2.5 Å resolution. The 5.0 Å resolution multiple isomorphous replacement (MIR) electron density map showed that although the standard myoglobin fold is conserved, the quaternary structure is different relative to all previously reported hemoglobin structures (Kolatkar, Meador, Stanfield & Hackert, 1988).

MIR data collected to 2.5 Å did not yield electron density maps of the quality necessary for accurate fitting of side chains and subsequent successful refinement. Therefore, multiple-wavelength anomalous-dispersion (MAD) data associated with the *K* absorption edge of the heme iron were collected to 3.0 Å Bragg spacings. A MAD Patterson map ($|^0F_A|^2$) clearly revealed the location of the two iron atoms in the asymmetric unit. The resulting MAD phased electron density map was clearer than the previous 3.0 Å resolution MIR map, but still limited for fitting amino-acid side chains. Cyclic map averaging with solvent flattening was subsequently applied. The resultant map allowed positioning of side chains. Phases were extended to 2.5 Å resolution during refinement using simulated annealing (*XPLOR*) (Brunger, Kuriyan & Karplus, 1987).

Experimental

U. caupo hemoglobin was isolated as described by Garey & Riggs (1984). Crystallization conditions used were the same as those reported previously (Kolatkar *et al.*, 1988). Data for native and heavy-atom derivatives were collected to 5.0 Å resolution using a four-circle automatic diffractometer. This facility consisted of a Syntex four-circle goniostat, Krisel automation package, Picker X-ray generator, and a PDP 11/04 computer. Reduction of data to correct for absorption, decay and Lorentz–polarization effects was carried out as described by Parks, Ernst, Hamlin, Xuong & Hackert (1985). Difference Patterson maps were used to locate heavy-metal sites. Difference electron density maps were used to check the positions of the metal atoms (to minimize bias, phases for the Fourier synthesis included all metal-atom types except the one being checked). Several cycles of least-squares refinement of the metal positions were carried out using phases from all metal types except the one being refined. Five heavy-metal derivatives proved to be useful for phase generation. The final phases resulting from all five low-resolution derivatives had a figure-of-merit of 0.79. Solvent flattening was carried out in conjunction with the iterative Wang algorithm to enhance the MIR map (Wang, 1985).

Higher-resolution (2.5 Å) data sets for native and four of the heavy-atom derivatives were collected using two Hamlin–Xuong multiwire area detectors (Xuong, Nielsen, Hamlin & Anderson, 1985) and Cu $K\alpha$ radiation from an Elliott GX-20 rotating-anode X-ray generator fitted with a graphite monochromator. The detectors were mounted on a table, built by Blake Industries, which enabled computer control of the θ_c position of each detector. Reduction of the area-detector data was carried out as described by Howard, Nielsen & Xuong (1985). Analysis of metal positions was carried out as before for the low-resolution data. Table 1* summarizes the heavy-atom derivative conditions, location of metal ions, their relative occupancies, and amino-acid residue neighbors of the metals.

MAD data were collected to 3.0 Å Bragg spacings at the Stanford Synchrotron Radiation Laboratory

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory, and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37056 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 1. *Heavy-atom derivatives*

Heavy metal	*x*	*y*	*z*	Occ.	Site of attachment
$K_2Pt(NO_2)_4$,	0.260	0.590	0.0160	0.0780	Cys163
1 m*M*, 2 days	0.485	0.842	0.262	0.0990	Lys183
	0.343	0.427	0.249	0.208	Met239
	0.382	0.654	0.236	0.126	Met282/Met229
	0.227	0.544	0.0300	0.138	Cys21
Di-Hg,*	0.0720	0.984	−0.0177	0.133	His45
2 m*M*, 4 days	0.0369	0.977	−0.0123	0.145	His45
	0.217	1.007	0.0367	0.0370	Heme142
cis-$[(NH_3)_2PtCl_2]$,	0.224	0.528	0.0316	0.141	Cys21
1.5 m*M*, 1 day	0.257	0.561	0.0200	0.168	Cys163
$CH_3HgOOCCH_3$,†	0.0720	0.986	−0.0176	0.134	His45
3 m*M*, 2 days	0.0178	0.639	0.257	0.0930	Amino1
$NaAuCl_4$,	0.128	0.973	0.0108	0.174	His45/Heme142
0.3 m*M*, 7 days					

* 3,6-Bis(mercurimethyl)dioxane acetate.

† Data for methylmercuric acetate collected to 5.0 Å.

(SSRL) in 1987 on beamline I-5 AD using the area-detector facility designed by Phizackerley, Cork & Merritt (1986). The experimental design and operating procedures were essentially the same as those reported for MAD measurements on lamprey hemoglobin (Hendrickson, Smith, Phizackerley & Merritt, 1988). Two crystals were used for data collection, one being mounted with the *b* axis along the capillary and the other having its *c* axis along the capillary. The crystals were oriented for rotation about these axes so that Bijvoet mates would be recorded simultaneously by mirror symmetry. The crystals were maintained at 277 K. The crystal-to-detector distance was set to 480 mm with $\theta_c = 18°$ for the measurements.

Since these crystals of *Urechis* hemoglobin are in the cyanomet state, as was the case for the lamprey hemoglobin experiments, the chemical environment of the iron atoms is expected to be virtually identical in the two situations. Moreover, since the X-ray optics were retained in the same configuration, the same energy resolution was expected. X-ray absorption measurements made from one of the *Urechis* hemoglobin crystals confirmed that the spectra were indeed very similar to those reported for cyanomet lamprey hemoglobin (Hendrickson *et al.*, 1988). As before, the X-ray energy was calibrated at the *K* absorption edge from an iron foil. The four X-ray wavelengths for diffraction measurements were chosen to be the same as those used in the lamprey hemoglobin experiments. These wavelengths are 1.738 (maximal f''), 1.740 (minimal f'), 1.800 (remote pre-edge point) and 1.500 Å (remote high-energy point). The values of f' and f'' used in the data analysis were taken to be the same as those derived from the lamprey hemoglobin absorption spectra since differences between the two spectra were judged to be too slight to have any significant effect on the phase assignment. Consequently, the *Urechis* absorption spectra were not processed.

Table 2. *Anomalous signal values for dispersive (off diagonal) and Bijvoet (diagonal) differences*

Numbers in parentheses are anomalous signal values from centric (background) terms.

λ	1.8000 Å	1.7400 Å	1.7382 Å	1.5000 Å
1.8000 Å	0.035 (0.022)	0.039	0.034	0.031
1.7400 Å		0.042 (0.025)	0.024	0.049
1.7382 Å			0.049 (0.026)	0.044
1.5000 Å				0.034 (0.031)

The reduction of MAD data proceeded in the following way. The MAD data were initially brought to a common scale using the programs *ANOSCL* and *SCALE* 2 which scale together Bijovet pairs related by mirror symmetry (Hendrickson *et al.*, 1988). Data at the four different wavelengths were then brought to a common scale by applying *WVLSCL*. At this point, spurious dispersive outliers were removed. The program *MADLSQ* (Hendrickson, 1985) was used to derive structure-factor amplitudes, $|^0F_T|$ and $|^0F_A|$, which represent the normal (wavelength-invariant) contributions from all atoms and the anomalous scatterers, respectively. *MERGIT* (Yang, Hendrickson, Crouch & Satow, 1990) allowed merging of symmetry-related redundant data. After reducing the MAD data, a clear anomalous signal (2–3%) was seen (Table 2) which was similar to that seen for lamprey hemoglobin (Hendrickson *et al.*, 1988).

A MAD Patterson ($|^0F_A|^2$) map (Fig. 1) clearly showed the two iron sites after appropriate cutoffs were used (2117 reflections used, 1819 reflections rejected). The cutoffs included discarding reflections with large outliers for the anomalous component and reflections which had a low anomalous signal ($< 2.0\sigma$). Least-squares refinement of the iron positions using *ANOLSQ* resulted in well-behaved occupancies and $R = 0.359$. Reflections used for refinement were those which had an anomalous signal greater than 1.0σ and excluded those reflections with measured anomalous differences which were too large (3010 reflections used, 926 reflections omitted). The iron positions were within 0.6 Å of the positions determined from the earlier 5.0 Å MIR electron density map. *MADFAZ* was subsequently used to calculate φ_T (native) to 3.0 Å resolution. *MADABCD* (Pähler, Smith & Hendrickson, 1990) was used to generate phase probability distributions defined by four phase (*A*, *B*, *C*, *D*) coefficients (Hendrickson & Lattman, 1970).

Molecular averaging was carried out about the twofold vector calculated using the coordinates of the iron atoms and assuming that there is no component along the *a* axis [a self-on-self rotation func-

tion showed that there is a twofold axis 35° from *c* in the *bc* plane (Kolatkar *et al.*, 1988)]. The procedures used for averaging maps are described by Bricogne (1976). Solvent flattening was applied using a mask (42% protein) derived from generating density about the atoms previously positioned in the 5.0 Å resolution map. The resultant map was back transformed to yield structure factors and phases. Phase probability distributions described by the four phase coefficients were generated using Sim weighting as described by Bricogne (1976). Phase combination employed summation of individual phase coefficients for the two phase probability distributions from the MAD and averaged/flattened map. Ten cycles of twofold averaging/solvent flattening were applied beginning with the 3.0 Å resolution electron density map derived from MAD phases. An attempt was made to extend phases to 2.5 Å resolution before model building by increasing the extent of data included in the cyclic process in small steps (0.1 Å), but the resulting maps were of poor quality.

Model building was carried out using the molecular graphics program *FRODO* written by T. A. Jones and modified by J. S. Sack, J. W. Pflugrath, M. A.

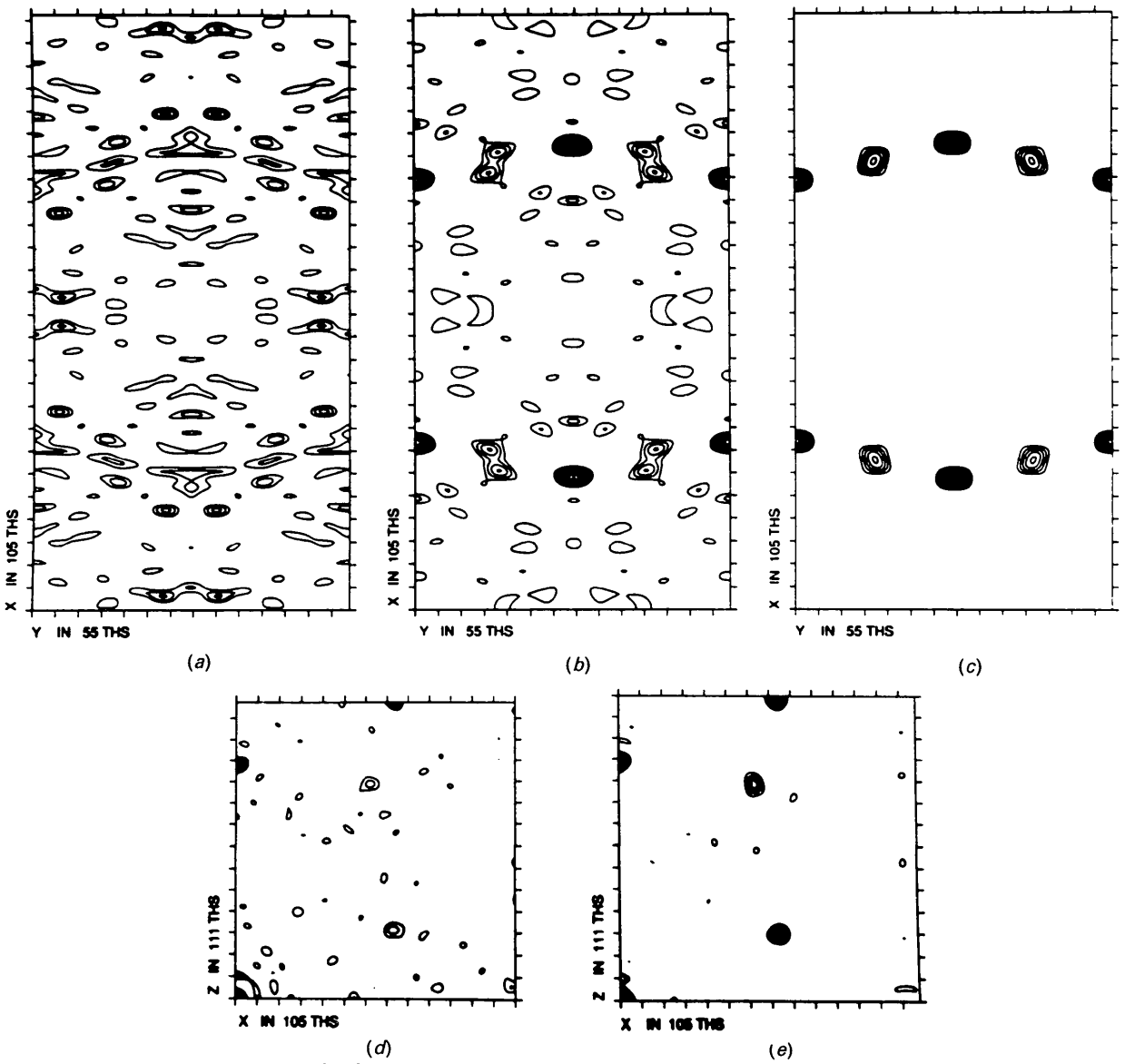

Fig. 1. Harker sections of the MAD ($|^0F_A|^2$) Patterson: (*a*) Harker section ($w = 55/111$) without cutoffs, (*b*) after application of cutoffs, (*c*) calculated ($w = 55/111$) Harker section, (*d*) Harker section ($v = 0/55$) after application of cutoffs, (*e*) calculated ($v = 0/55$) Harker section.

Table 3. *Summary of refinement*

Model	*A*	*B*	*C*	*D*	*E*	*F*	*G*
Resolution of shell (Å)	5–3	5–2.5	5–2.5	5–2.5	5–2.5	5–2.5	5–2.5
Cutoff $>n\sigma F$	–	–	–	3	3	3	3
No. of reflections	5166	9877	9877	8350	8350	8350	8350
Type of refinement	SA[a]	SA, *B* min[c]	SA, *B* min	SA, *B* min	*xyz* min[b], *B* min	*xyz* min, *B* min	*xyz* min, *B* min
No. of water molecules	–	–	–	–	108	133	154
Start R[d]	0.53	0.46	0.286	0.235	0.203	0.204	0.158
Final R	0.24	0.25	0.206	0.168	0.168	0.155	0.148
R.m.s. deviation from ideal geometry							
Bond lengths (Å)	0.043	0.016	0.015	0.014	0.014	0.016	0.015
Bond angles (°)	5.047	3.472	3.299	3.124	3.310	3.546	3.330
Dihedral angles (°)	25.141	21.937	21.121	20.640	21.147	21.453	21.046
Improper angles (°)	2.983	1.923	1.805	1.733	1.792	1.949	2.027

Notes: (*a*) simulated annealing (SA) consisted of (1) 100–150 steps of conjugate gradient minimization (charges on), (2) 'slow cooling' (charges off) *i.e.* 4000 to 300 K in steps of 25 K and at each temperature 25 steps × 0.0005 ps step^{-1}, (3) 100–120 steps of conjugate gradient minimization (charges off); (*b*) conjugate gradient minimization usually 80 steps (charges on); (*c*) isotropic temperature-factor refinement usually 20 steps; (*d*) $R = (\sum_{hkl}||F_o^{hkl}| - |F_c^{hkl}||)/(\sum_{hkl}|F_o^{hkl}|)$.

Saper, B. L. Bush & T. A. Jones (Jones, 1979). Human β-hemoglobin with the *Urechis* amino-acid sequence was aligned onto the electron density of one subunit in the averaged map. Insertions, deletions, and adjustments of helices, turns and side chains were made. A duplicate of this fitted monomer was then rotated and used as a starting model for the second subunit. The model dimer was then subjected to simulated annealing using data to 3.0 Å resolution with temperature factors being set to 15.0 Å^2 for all atoms (Brunger, 1988; Brunger, Karplus & Petsko, 1989). Simulated annealing consisted of conjugate gradient minimization, 'slow cooling', followed with conjugate gradient minimization (Weis, Brunger, Skehel & Wiley, 1990). The two subunits in the asymmetric unit were refined independently without any symmetry restraints.

The resultant model was used to build a non-existent density (NED) map. An NED map is similar to a complete OMIT map but is based on omitting slabs of electron density, each slab consisting of several sections of the electron density map, instead of a run of atoms. In this process the NED map is assembled from slabs of a $(2w|F_{obs}| - |F_{calc}|)\varphi_{calc}$ map, where the $|F_{calc}|$ and φ_{calc} terms are derived from a modified map that omits those sections currently being calculated (approximately 10% of the volume) and w is the Sim weight as described by Bricogne (1976). Reconstituting the new electron density for the omitted volume using the back-transformed $|F_{calc}|$ and φ_{calc} terms from the remaining electron density helps diminish model bias (Rice, 1981). The atoms used in the structure-factor calculation are not further refined in the absence of the missing atoms or 'shaken' to reduce 'memory' bias. However, this could easily be incorporated into the NED map generation process. This process is iterated to calculate the missing electron density for all such slabs, which are used to construct the resultant NED map. Some overlap is used in determining the sections for each slab so that the outer sections can be discarded. This type of map seems to work well once the structural model is good enough to produce a model with an R value of <0.35, and is especially useful in later stages of model building when the 'straight' $(2|F_{obs}| - |F_{calc}|)\varphi_{calc}$ map has little new information due to the close agreement of the observed and calculated structure-factor amplitudes. The NED map also has an advantage compared to the more conventional type of OMIT map by always omitting neighboring atoms as well as adjacent residues.

The NED map allowed another round of model building and the resultant model was refined by simulated annealing including data to 2.5 Å resolution. The temperature factors (B) were also refined for individual atoms at this time. Cutoffs were applied to exclude badly measured reflections in later cycles. Seven cycles of model building and refinement resulted in a well defined electron density map at 2.5 Å resolution. Table 3 summarizes the details of the refinement.

A difference map $(|F_{obs}| - |F_{calc}|)\varphi_{calc}$ was calculated to locate potential solvent molecules. A conservative approach was taken to place solvent molecules. A peak-finding program used in conjunction with another program which determines the hydrogen-bonding suitability of neighboring residues (Sheriff, Hendrickson & Smith, 1987) allowed the positioning of 154 water molecules. Peaks included as solvent had to be at least at a level of 3.0σ and would only be placed at a peak location if it would not conflict with nearby residues. The positions of these water molecules were checked by visual inspection to ascertain proper hydrogen-bonding geometry. These water molecules were subsequently included in four cycles of conventional refinement where

positions and temperature factors were refined in alternating cycles until convergence was reached.

A Ramachandran plot of the resultant model is shown in Fig. 2 (Ramachandran & Sasisekharan, 1968). A plot of *B versus* residue was made to check for regions of high mobility as well as to gauge the quality of refinement (Fig. 3). Differences in main-chain atom positions between the two subunits after superpositioning (root-mean-square deviations) were plotted *versus* residue to check for locations in the structure where the coordinates deviate about the twofold and also to check for possible problems in the model (Fig. 4).

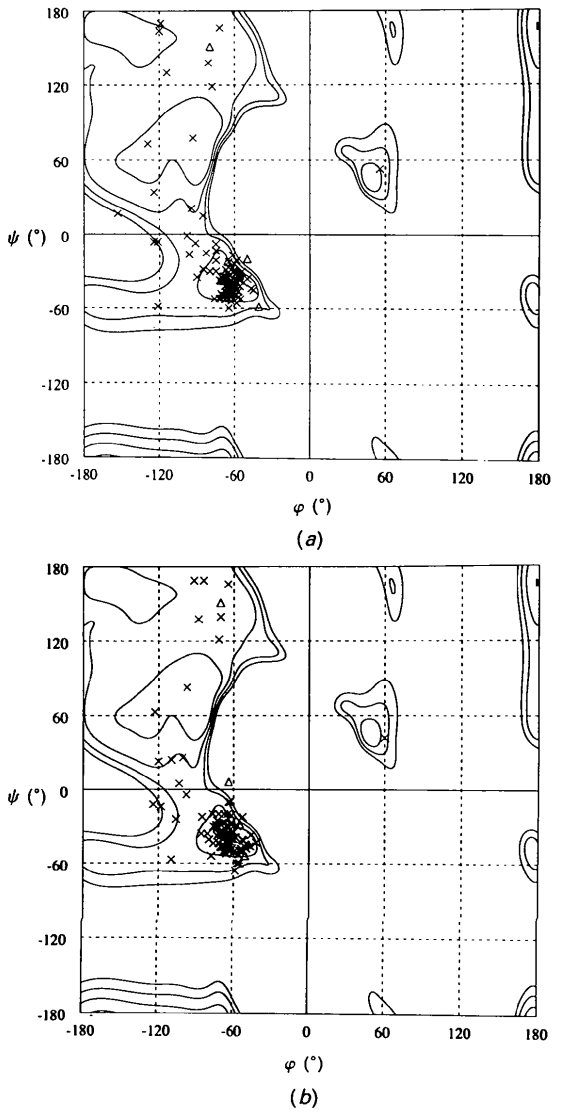

Fig. 2. Ramachandran plots displaying main-chain conformational angles for the two independent subunits: (*a*) subunit 1, (*b*) subunit 2. Energy surface contours for allowed dihedral regions for alanylalanine are also drawn. The triangles represent proline residues. Glycine residues are not included in the plot.

To verify the existence of free sulfhydryl groups at Cys21, samples of *Urechis* hemoglobin were run under anaerobic SDS gel electrophoresis conditions. Anaerobic SDS gel electrophoresis was carried out in a chamber filled with argon to prevent autooxidation of the cysteine residues during electrophoresis. Previous runs under aerobic conditions had shown significant amounts of dimer as well as monomer species (two bands, 15 and 30 kDa), although the dimer component completely disappeared if pretreated with iodoacetamide (one band, 15 kDa). The dimer species present in aerobic gels is likely the result of disulfide links formed during electrophoresis between Cys21 residues of two subunits. The anaerobic gel clearly showed that no disulfide bond was present.

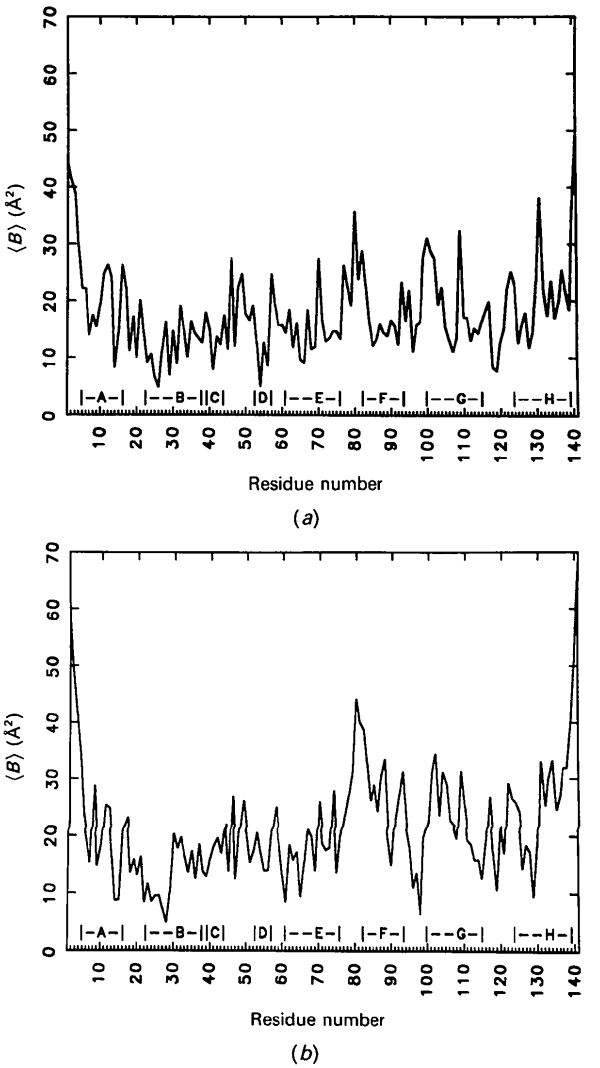

Fig. 3. Plot of average temperature factor as a function of amino-acid residue. The locations of the eight helices are indicated. (*a*) Subunit 1, (*b*) subunit 2.

Results and discussion

Initially, five heavy-metal derivatives had facilitated the calculation of a low-resolution electron density map. The map had allowed an unambiguous chain trace as reported previously (Kolatkar *et al.*, 1988). The chain trace showed that *U. caupo* hemoglobin possesses an unusual quaternary structure although the subunit tertiary structure displays the myoglobin fold characteristic of all hemoglobins.

An attempt to use these same heavy-metal derivatives for higher-resolution information (2.5 Å resolution) proved unsuccessful. Even after applying twofold averaging/solvent flattening, the map lacked the quality necessary for accurately fitting side chains or even always affirming the general fit of individual helices (Fig. 5*a*). The poor quality of the map at higher resolution likely resulted from a lack of isomorphism and/or the fact that several metal atoms were located at special symmetry positions. The figure-of-merit for MIR phases to 2.5 Å resolution was 0.65. Therefore MAD phasing was employed because it would be independent of either derivative information or model bias.

A 3.0 Å electron density map was calculated once the positions of the iron atoms were refined (3936 reflections, 62% of data to 3.0 Å). The resulting map clearly showed the main-chain path for the protein, but electron density for most side chains was still poorly defined. There was also some noise near the iron sites, perhaps due to its large anomalous component. Thus, the higher-resolution $|F|$ values collected with an area detector at the University of Texas using a rotating-anode source with Cu $K\alpha$ radiation ($\lambda = 1.5418$ Å) were used since they might be expected to be better data than the derived $|^0F_T|$ values obtained from the treatment of the MAD data. This map (Fig. 5*b*) was indeed more interpretable, particularly near the iron sites. Much ambiguity still remained, however, with regard to side-chain placement.

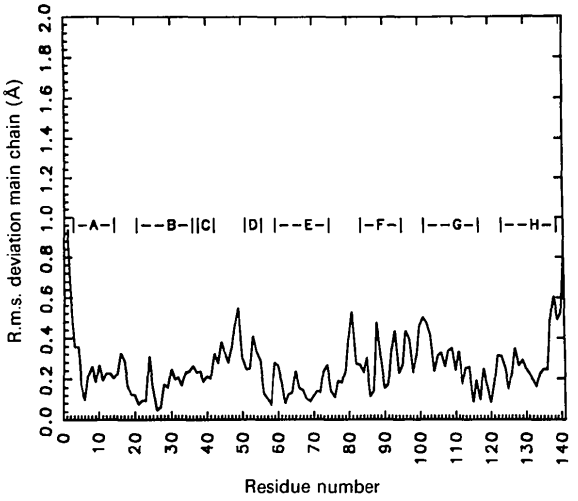

Fig. 4. Plot of root-mean-square deviation for main-chain atoms of all residues of *Urechis* hemoglobin after superimposing the coordinates of the second subunit onto the first subunit.

Refinement of the MAD phases by molecular averaging about the local twofold in conjunction with solvent flattening for ten cycles greatly enhanced the map and most side-chain density became readily apparent (Fig. 5*c*). There was an average phase change of 18.8° from the start of averaging to the last cycle. The phase change from the penultimate to the ultimate cycle was less than 2° indicating that the cyclic averaging had converged. Attempts to extend the phases to 2.5 Å proved unsuccessful. Owing to the possibility of small differences between the two subunits, averaging was not used at higher resolution.

The averaging process had also created a bridge of density between the sole cysteine residue of one subunit and its twofold-related mate suggesting a disulfide bridge. However, an NED $(2w|F_{obs}| - |F_{calc}|)\varphi_{calc}$ map at higher resolution (2.5 Å resolution) showed that there was no connecting density between the cysteine residues, in fact, the sulfhydryl groups were seen to be pointed away from each other. In addition, anaerobic SDS gel electrophoresis showed that no disulfide was present. The absence of a disulfide bond is consistent with the character of sulfhydryl groups in other intracellular proteins. In general, disulfide bonds are observed in extracellular proteins while intracellular proteins have free thiol groups (Torchinsky, 1981). This is likely the result of the reducing environment inside the cell (Ziegler, 1985).

Phase extension was facilitated by simulated annealing using data from 5.0 to 2.5 Å resolution for refinement. The starting model, obtained from refinement using data between 5.0 and 3.0 Å resolution, showed no major changes upon refinement using data to 2.5 Å resolution. The final R value for two independent monomers and 154 waters was 0.148 for those reflections with $|F| > 3.0\sigma$ (8350 reflections, 85% of the data) between 5.0 and 2.5 Å resolution and retention of good geometry (Table 3). $R = 0.172$ for those reflections $|F| > 3.0\sigma$ (9634, 86% of the data) between 10.0 and 2.5 Å resolution.

The phases from the refined model were compared to experimental phases from MIR, MAD, MAD (*ABCD*), and the averaged map. Experimental phases were also compared with each other. The average differences in phases are listed in Table 4. It is clear that the MAD data set was superior for obtaining phase information relative to MIR phases and the phases resulting from the averaged MAD (*ABCD*) map were closer in value to the phases calculated from the final refined model. The average

phase discrepancies found here for both the MAD and average MAD phase sets (55.0 and 51.4°) are slightly better than those obtained for streptavidin (56.9 and 54.4°) (Hendrickson, Pähler, Smith, Satow, Merritt & Phizackerley, 1989).

A Ramachandran plot (Fig. 2) of the structure shows that nearly all residues are in the allowable geometric regions. Ser120 is in an allowed zone but with somewhat strained dihedral angles ($\varphi = 56°$, $\psi = 53°$). Ser120 is located in a loop involved in one of the crystallographic contacts. Temperature-factor (B) values increase near the amino-terminal and carboxy-terminal ends (Fig. 3). Density is weak for the last two amino-terminal residues but is well-defined for all other residues. Residues near the surface have B values higher than residues in the interior of the protein. The well-behaved values for B are another indication of the quality of refinement.

A comparison of main-chain atoms from the two independent subunits superimposed shows that the two subunits are very similar (Fig. 4). The magnitude of the differences between subunits appears to correlate directly with the value of temperature factor for that particular residue. Most notably high temperature factors and large differences between subunits are observed for amino-terminal and carboxy-terminal residues as well as the turn regions of the subunits. The main-chain atoms of the two subunits are related by an angle of 179.8° ($t_x = 0$).

There is a considerable amount of contact between the two independent subunits, involving hydrogen bonds, van der Waals contacts, salt links and water-

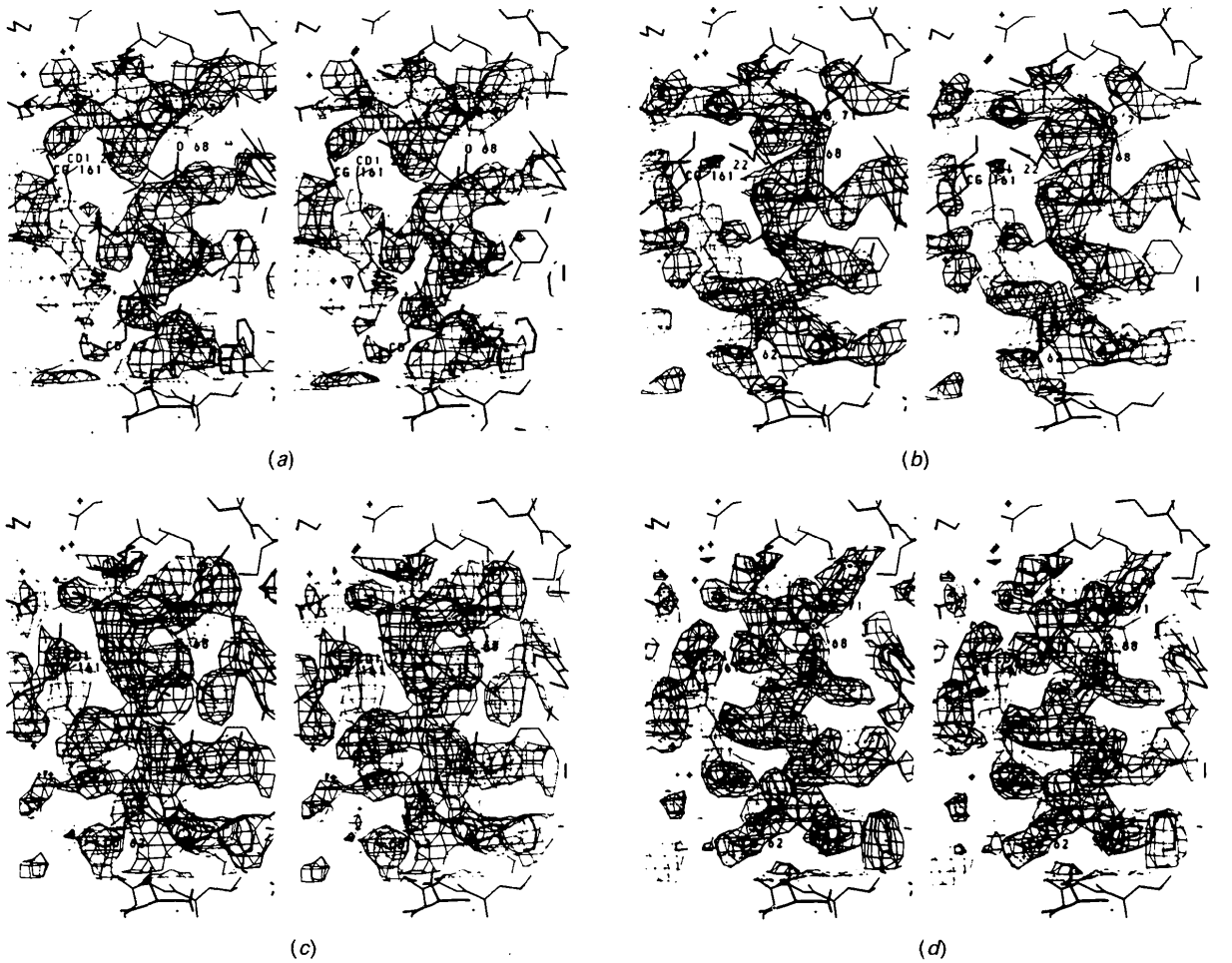

Fig. 5. Comparison of electron density maps: (*a*) 3.0 Å resolution MIR map, (*b*) 3.0 Å resolution map generated using MAD phases, (*c*) 3.0 Å resolution map generated after applying ten cycles averaging/solvent flattening to the 3.0 Å resolution MAD map, and (*d*) 2.5 Å resolution ($|2F_{obs}| - |F_{calc}|)\varphi_{calc}$ NED map using the refined coordinates. Region shown is part of the E helix. A discontinuous region of electron density through the main chain is observed in the MIR map (*a*) as well as missing electron density for side chains including a Tyr71 above the heme and Lys62 on the lower left. The MAD map (*b*) has continuous density for main-chain atoms but excludes many side chains including the tyrosine and lysine. The averaged map (*c*) has most of the atoms in density although the electron density is still sparse for the side chain of Lys62 compared to the final NED map (*d*).

Table 4. *Average differences between phases obtained from several sources*

	Model	MAD	*ABCD*	Averaged
MAD	55.0			
*ABCD**	55.4	11.0		
Averaged†	51.4	23.4	18.8	
MIR	66.2	73.4	73.4	74.3

* Phases derived by formulating the phase distribution from MAD in terms of four coefficients (*A*, *B*, *C*, *D*).

† Phases obtained from averaging the MAD (*ABCD*) map about the local twofold axis.

mediated hydrogen bonds (Fig. 6). The region of contact for subunits related by the molecular twofold is the *A*/*B* turn and *E* helix. Similar interactions involving the *G*/*H* turn and *D* helix are present between subunits related by a crystallographic twofold at the other subunit interface, although there are fewer salt links around the crystallographic twofold contact.

It is seen from this work that many pitfalls in the high-resolution structure determination were circumvented by using a combination of methods at our disposal. MIR allowed a low-resolution 'structure determination' but was not useful at high resolution. MAD phasing provided an independent estimate of the phases. Subsequently, cyclic averaging and simulated annealing allowed a structure determination to be made at 'high' resolution. A detailed discussion of the structure and comparison to other hemoglobin structures will be published elsewhere.

We wish to thank Drs J. Garey and A. Riggs for providing material for these studies, Dr R. Stanfield, Dr J. Smith and W. Meador for earlier contributions to this project, and Dr D. Ziegler, Dr L. Poulson and Katy Kuo for their assistance and discussions involving sulfur chemistry. We also thank Aaron Wernham for assistance with preliminary MAD data reduction. Synchrotron facilities used at SSRL are supported by the Department of Energy and by the National Institutes of Health. The early steps in refinement were carried out at the Pittsburgh Supercomputing Center (grant No. DMB880059P). The UT System at the Center for High-Performance Computing (CHPC) facility was used for most of the later stages of refinement. This work was supported in part by grants No. GM30105 (to MLH) and GM34102 (to WAH) from the National Institutes of Health and a grant from the Foundation for Research (to MLH).

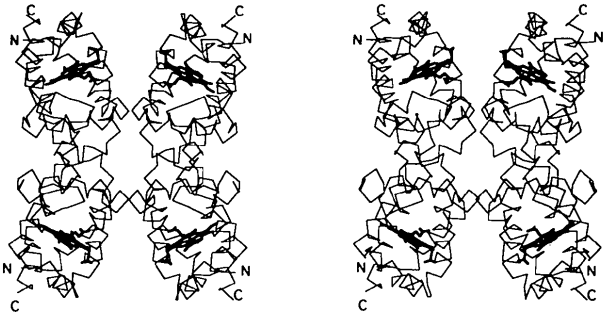

Fig. 6. An α-carbon representation of *Urechis* tetramer. Contact region at the *A*/*B* turn and *E* helices (horizontal contacts across the page) is observed as well as contact between the *D* and *G*/*H* helices (vertical contacts).

References

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

BRUNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

BRUNGER, A. T., KARPLUS, M. & PETSKO, G. A. (1989). *Acta Cryst.* A**45**, 50–61.

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

GAREY, J. R. & RIGGS, A. F. (1984). *Arch. Biochem. Biophys.* **228**(1), 320–331.

GAREY, J. R. & RIGGS, A. F. (1986). *J. Biol. Chem.* **261**, 16446–16450.

HENDRICKSON, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

HENDRICKSON, W. A., PÄHLER, A., SMITH, J. L., SATOW, Y., MERRITT, E. A. & PHIZACKERLEY, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.

HENDRICKSON, W. A., SMITH, J. L., PHIZACKERLEY, R. P. & MERRITT, E. A. (1988). *Proteins*, **4**, 77–88.

HOWARD, A. J., NIELSEN, C. & XUONG, N. H. (1985). *Methods Enzymol.* **114**, 452–472.

JONES, T. A. (1979). *J. Appl. Cryst.* **11**, 268–272.

KOLATKAR, P. R., MEADOR, W. E., STANFIELD, R. L. & HACKERT, M. L. (1988). *J. Biol. Chem.* **263**(7), 3462–3465.

PÄHLER, A., SMITH, J. L. & HENDRICKSON, W. A. (1990). *Acta Cryst.* A**46**, 537–540.

PARKS, E. H., ERNST, S. R., HAMLIN, R., XUONG, N. H. & HACKERT, M. L. (1985). *J. Mol. Biol.* **182**, 455–465.

PHIZACKERLEY, R. P., CORK, C. W. & MERRITT, E. A. (1986). *Nucl. Instrum. Methods*, A**246**, 579–595.

RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Adv. Protein Chem.* **23**, 283–437.

RICE, D. W. (1981). *Acta Cryst.* A**37**, 491.

SHERIFF, S., HENDRICKSON, W. A. & SMITH, J. L. (1987). *J. Mol. Biol.* **197**, 273–296.

TORCHINSKY, YU. M. (1981). *Sulfur in Proteins*, translated by W. WITTENBERG, edited by D. METZLER, pp. 199–217. Oxford: Pergamon Press.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

WEIS, W. I., BRUNGER, A. T., SKEHEL, J. J. & WILEY, D. C. (1990). *J. Mol. Biol.* **212**, 737–761.

XUONG, N. H., NIELSEN, C., HAMLIN, R. & ANDERSON, D. (1985). *J. Appl. Cryst.* **18**, 342–350.

YANG, W., HENDRICKSON, W. A., CROUCH, R. J. & SATOW, Y. (1990). *Science*, **249**, 1398–1405.

ZIEGLER, D. M. (1985). *Annu. Rev. Biochem.* **54**, 305–329.