Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A direct-method *ab initio* phasing of a protein, pseudoazurin, at 1.55 Å resolution

Monika Mukherjee,[a] Susim Maiti[a] and M. M. Woolfson[b]*

[a]Department of Solid State Physics, Indian Association for the Cultivation of Science, Calcutta-700032, India, and [b]Department of Physics, University of York, York YO10 5DD, England

Correspondence e-mail: mmw1@york.ac.uk

Received 9 March 2000
Accepted 5 June 2000

The direct-method program *MULTAN*88 has been applied to solve a known protein, pseudoazurin, space group $P6_5$, unit-cell parameters $a = b = 50.0\,(1)$, $c = 98.5\,(3)$ Å, with 917 protein atoms, a Cu atom and 93 solvent water molecules in the asymmetric unit (>6000 non-H atoms in the unit cell) and data at 1.55 Å resolution. One of several trials with sets of initially random phases yielded phase estimates for 1000 reflections with a mean phase error of 68.3° that was recognized as the best available solution by the figure of merit being used. Phase extension to 2000 largest *E*s showed a distorted tetrahedral geometry around the Cu site. Density modification was applied to improve the phases and the quality of the maps, starting with phases calculated from the atomic positions indicated by the first map.

1. Introduction

In macromolecular structure determination, phases are generally estimated using either the multiple isomorphous replacement (MIR) or the multiple anomalous dispersion (MAD) method. Phases obtained using the MIR method sometimes suffer from lack of isomorphism, so that the electron-density maps calculated from phases thus estimated might not be good enough to indicate a model structure. In this paper, we have redetermined the structure of pseudoazurin from single-wavelength native data without the use of any known fragments or MIR phases.

The asymmetric unit of the redox protein pseudoazurin from *Alcaligenes facealies* contains 917 protein atoms, 93 ordered solvent water molecules and a Cu atom (123 amino-acid residues; molecular weight 13 kDa). The crystal system is hexagonal, space group $P6_5$, with unit-cell parameters $a = b = 50.0\,(1)$, $c = 98.5\,(3)$ Å. The pseudoazurin data used in the present analysis were obtained from the Protein Data Bank (PDB codes 1paz, r1pazsf), consisting of $|F|$ and $\sigma|F|$ for 19 659 unique reflections in the 9.0–1.55 Å resolution range. The experimental details of data collection using synchrotron radiation from the storage ring DORIS are given by Petratos *et al.* (1988). Knowledge of the structure was not used in the phase extension and refinement except for calculating the mean phase error by which progress was monitored.

Recently, the structures of rubredoxin (Mukherjee, 1999), RNAP1 (Mukherjee *et al.*, 1999) and HiPIP (Parisini *et al.*, 1999) have been solved from single native data sets by direct methods. The success of the *ab initio* phase determination in these cases relied on good-quality data extending to atomic resolution and, except for RNAP1, the presence of Fe atoms in the asymmetric unit.

It has been observed that there is marked deterioration in the interpretability of maps with a data resolution lower than 1.2 Å. In the present case of pseudoazurin with 1.55 Å resolution data, model building could not be carried out with the map from direct-method phases owing to the poor connectivity of the electron-density map. However, a combination of a direct method together with density modification was effective in solving the structure at 1.55 Å resolution.

2. *Ab initio* phase determination using *MULTAN*88

The version of *MULTAN*88 (Debaerdemaeker *et al.*, 1988) we used to solve pseudoazurin generates random numbers by a magic integer series (White & Woolfson, 1975) and uses the weighting scheme developed by Hull & Irwin (1978). Owing to the large number of non-H atoms in the unit cell (>6000), individual phase relationships have high variances and consequently obtaining a phase set with mean phase error (MPE) less than 75° was not straightforward. Three runs of *MULTAN*88 were made, each generating 1000 sets of phases, using different numbers of reflections with large $|E|$, different seeds for the random-number generator and with different values of KMIN, the quantity governing the lowest acceptable variance for a three-phase relationship. The results are summarized in Table 1. Only the first two runs gave a MPE < 75° and the best solution for the trial with 1000 reflections (set No. 326) could be identified by its having the best modified figures-of-merit (FOMs) proposed by Mukherjee & Woolfson (1993). For the 1200-reflection run the FOMs for the best solution in terms of MPE (set No. 402) were quite good, but not the best. It did have the highest value of ABSM, but only the fifth highest overall FOM, CFOM2, the higher values being 2.664, 2.652, 2.571 and 2.548. However, accepting that the FOMs are by no means perfect indicators, it would be normal strategy in the case of an unknown structure to investigate several phase sets at the top of the CFOM2 list. The FOMs for the two good phase sets are shown in Table 2.

In both Table 1 and Table 2, the column MPE2 gives the MPE for the enantiomorph structure (*i.e.* corresponding to all phases reversed in sign). Where structures contain a heavy atom, the phase sets from direct methods, even with acceptably low MPEs, sometimes correspond to a sum of the two enantiomorphs. The large difference between MPE1 and MPE2 in this case indicates that the phase estimates discriminated well between the enantiomorphs. This would also be evident if the structure was unknown through the value of DEVI, which indicates the mean magnitude of the deviation of three-phase invariants from the special values 0 and π.

3. Phase extension and refinement by the tangent formula

The results of extending and refining phase information are given in Table 3. Starting with 800 phases from set No. 326 (those indicated as most reliable by *MULTAN*88), phase extension to 2000 largest *E*s followed by phase refinement was carried out following the general procedure described by Woolfson & Yao (1988). During the refinement stage, the 800 initial phases were kept fixed with unit weight. This is a multi-solution process, but all the 100 sets of 2000 phases had similar MPEs of ∼69.9°. An *E* map was calculated at this stage with the 2000 estimated phases and was interpreted by the peak-search routine of *MULTAN*88. The Cu atom, with its tetrahedral coordination geometry, could easily be distinguished although distances and angles were somewhat distorted.

The same procedure of phase extension was repeated for set No. 402, again starting with 800 of the most reliable of the 1200 initial estimated phases. In this case, all 100 sets of 2000 phases had MPEs ∼72.4°. An *E* map (Fig. 1*a*) showed Cu as the strongest peak, together with four surrounding atoms with tetrahedral symmetry that had bond distances and angles very close to those expected around the Cu site. With the five peaks assigned as (Cu + 2S + 2N) atoms plus another 1000 of the

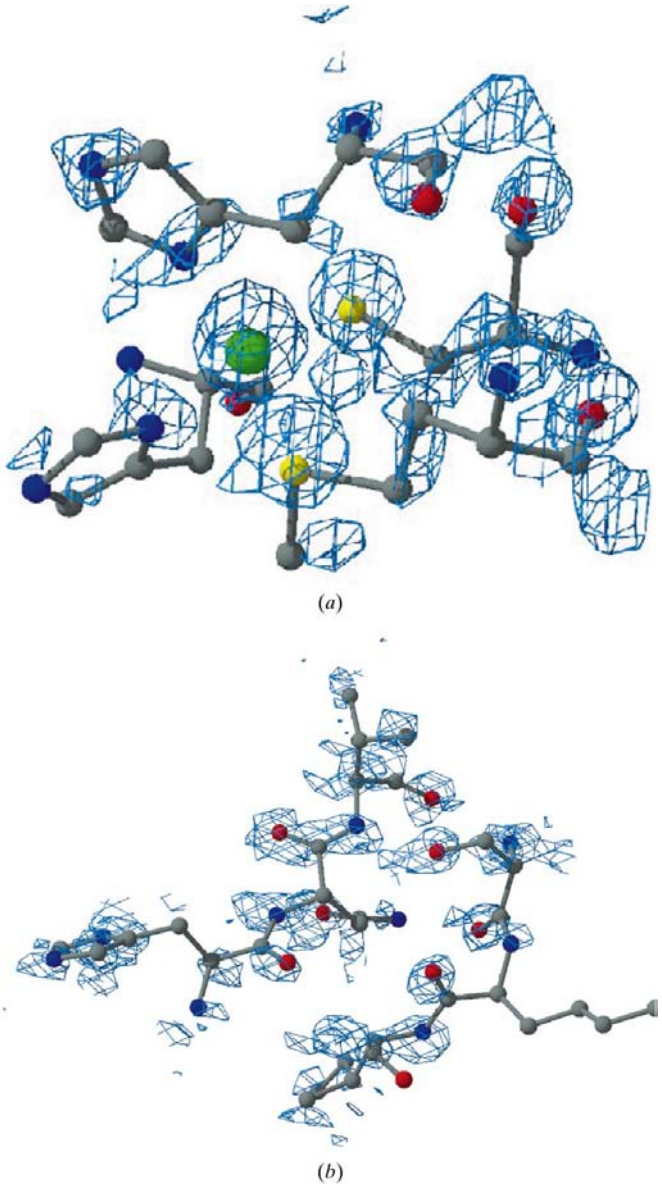

Figure 1
(*a*) A section of the *E* map around the copper protein. (*b*) A section of the *E* map showing two parts of the protein chain.

Table 1
Results of applying *MULTAN*88 to pseudoazurin for various NREF and KMIN with 1.55 Å data.

	NREF†	NREL‡	KMIN§	TOTSET¶	NG††	SET	MPE1 (MPE2)‡‡
(i)	1000	35470	0.25	1000	1	326	68.29 (79.38)
(ii)	1200	46728	0.50	1000	1	402	71.88 (81.97)
(iii)	1500	68461	0.50	1000	0	915	75.64 (83.93)

† Number of reflections with largest $|E|$. ‡ Number of three-phase relationships. § Minimum value of κ for any three-phase relationship, where $\kappa(\mathbf{h}, \mathbf{k}) = 2\sigma^3\sigma_2^{-3/2}|E(\mathbf{h})E(\mathbf{k})E(\mathbf{h}-\mathbf{k})|$, $\sigma_n = \sum_{j=1}^{N} z_j^n$ and where the jth of N atoms in the unit cell has atomic number z_j. ¶ Total number of phase sets generated. †† Number of phase set with mean phase error $\leq 75^\circ$. ‡‡ MPE1, mean phase error; MPE2, mean phase error for the enantiomorph.

strongest peaks taken as C, successive difference Fourier maps located a few missing atom sites and a plausible chain of 22 atoms could be recognized. Least-squares refinement using *SHELXL*93 (Sheldrick, 1993) using the conjugate-gradient algorithm (CGLS) option in the 9–1.55 Å resolution range lowered the R value to 0.30. However, owing to the poor connectivity of the electron-density map, attempts to interpret it were unsuccessful. This led us to try a procedure to improve the map quality by density modification.

4. Density modification

The density modification was carried out using the *PERP* package described by Refaat *et al.* (1996*b*). The individual density-modification processes contained in the package are SE, Sayre-equation refinement (Refaat *et al.*, 1995); LE, low-density elimination (Shiono & Woolfson, 1992; Refaat & Woolfson, 1993); HM, histogram matching (Zhang & Main, 1990); DM, density histogram using the local maximum density (Refaat *et al.*, 1996*a*); SQ, squaring of density, *i.e.* taking phases of the squared current density; DV, double histogram using the local density variance (Refaat *et al.*, 1996*a*). For the histogram-matching methods (HM, DM and DV) solvent flattening was included automatically.

Once initial phases are available to produce a first map then *PERP* automatically goes through cycles of modifying density and finding new phase estimates. In the version of *PERP* that we used, the original LE component was replaced by a much-improved process first described by Mutsugati & Shiono (1999). The steps in this process are as follows.

(i) A map is calculated from current phase estimates.

(ii) The density ρ is then modified to ρ' by

$$\rho' = 0, \quad \rho < 0$$
$$\rho' = \rho\{1 - \exp[-\tfrac{1}{2}(5\rho/\rho_c)^2]\}, \quad 0 \leq \rho \leq 2\rho_c$$
$$\rho' = 2\rho_c, \quad \rho > 2\rho_c,$$

where ρ_c is found as follows. (*a*) In each section along x, the maximum density $\rho_{\max}$ was calculated. The average for the sections is $\langle\rho_{\max}\rangle_x$. (*b*) This is repeated for sections along y and z to give $\langle\rho_{\max}\rangle_y$ and $\langle\rho_{\max}\rangle_z$, respectively. (*c*) $\rho_c = \min[\langle\rho_{\max}\rangle_x, \langle\rho_{\max}\rangle_y, \langle\rho_{\max}\rangle_z]$. This modification eliminates negative density, reduces low density in a progressive way and also truncates higher density in the map. This process, dynamic density modification, has been used by Foadi *et al.* (2000) in a procedure for solving protein structures from small fragments; they refer to it as DDM1.

(iii) Structure factors were calculated from the modified density using a FFT routine.

Table 2
Figures of merit for the 'good' sets 326 and 402.

MPEs were calculated against the refined model.

SET	ABSM†	PSIM‡	RESM§	CFOM2¶	DEVI††	MPE1 (MPE2)
326	0.949	0.729	34.22	2.417	43.41	68.29 (79.38)
402	0.899	0.716	34.46	2.543	43.16	71.88 (81.97)

† ABSM = $s/s_{\rm exp}$. ‡ PSIM = $\sum_l \left|\sum_k E(\mathbf{k})E(\mathbf{l}-\mathbf{k})\right|/s$. § RESM = $\sum_{\mathbf{h}} \{[\alpha(\mathbf{h})/s] - [\alpha(\mathbf{h})_{\rm exp}/s_{\rm exp}]\} \times 100$, where $\alpha(\mathbf{h}) = \left|\sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k})\right|$, $s = \sum_{\mathbf{h}}\alpha(\mathbf{h})$, the subscript exp corresponds to the value for the true phases and the summation over $\mathbf{h}$ is for large $|E|$s and that over $\mathbf{l}$ is for small $|E|$s. ¶ CFOM2 = w_1[(ABSM − $\text{ABSM}_{\min}$)]/[($\text{ABSM}_{\max}$ − $\text{ABSM}_{\min}$) + w_2(PSIM − $\text{PSIM}_{\min}$)]/[($\text{PSIM}_{\max}$ − $\text{PSIM}_{\min}$)] + w_3[($\text{RESM}_{\max}$ − RESM)/($\text{RESM}_{\max}$ − $\text{RESM}_{\min}$)], where the subscripts max and min correspond to the maximum and minimum values for the 1000 phase sets and the weights are set at $w_1 = w_2 = w_3 = 1.0$. †† DEVI = $\langle \min(|\varphi_{3,i}|, |180 - \varphi_{3,i}|)\rangle_i$, where $\varphi_{3,i}$ is the value in ° of the ith three-phase invariant.

Table 3
Phase extension from set Nos. 326 and 402.

MPEs were calculated against the refined model

Initial phases	From set	Extended to	NREL	KMIN	TOTSET	MPE1 (MPE2)
800	326	2000	113520	0.20	100	69.9 (81.6)
800	402	2000	113533	0.20	100	72.4 (82.4)

Several trials have been made with the *PERP* procedure, the results of which are given in Table 4. The *PERP* refinement cycles applied were given by the keywords REF 25 (SE 2 LE 1 HM 1 DM 1 SQ 1 LE 1 DV 1), which indicates a refinement sequences of 25 cycles consisting of two applications of SE followed by one application each of LE, HM, DM, SQ, LE, DV, where LE was actually DDM1 as described above. It was found in this case that DDM1 alone did not improve phases, but that the combination of *PERP* procedures was effective.

In the first trial, the initial phases were obtained from the Cu position only; in the next three trials, more light atoms were added – first (Cu, 2S, 2N), then (Cu, 2S, 2N, 2C) and finally 22 atoms (Cu, 2S, 2N, 17C). In these four trials, the phases of reflections with $|E| > 1.0$, 7168 in number, were refined. Starting MPEs from 74–72° dropped to ∼30° after 200 cycles of refinement with 7168 largest $|E|$s; the map correlation coefficient (MCC) also rose to a high value (Fig. 1*b*). The effect of taking a greater number of reflections by taking $|E| > 0.9$ (8587 reflections), $|E| > 0.8$ (10 164 reflections) and $|E| > 0.7$ (11 837 reflections) is also shown in Table 4. An analysis of the results in Table 4 shows that taking the Cu coordinates alone is very effective and that there is no gain in adding up to 21 light atoms. The effect of including more reflections in the *PERP* process is also seen not to give any

Table 4
Results of phase refinement using *PERP* with starting phases derived from coordinates obtained using *MULTAN*88.

MPEs and MCCs were calculated against the refined model. MCC is the standard map correlation.

No. of reflections	Starting phases with coordinates	MPEs Initial MPE1 (MPE2)	MPEs Final MPE1 (MPE2)	MCC Initial	MCC Final
7158 ($E > 1.0$)	Cu	83.5 (72.6)	85.7 (30.2)	0.259	0.819
7158 ($E > 1.0$)	Cu + 2S + 2N	81.9 (71.8)	85.9 (31.9)	0.272	0.814
7158 ($E > 1.0$)	Cu + 2S + 2N + 2C	81.6 (72.7)	85.6 (32.0)	0.257	0.802
7158 ($E > 1.0$)	Cu + 2S + 2N + 17C	82.3 (74.4)	85.8 (29.9)	0.231	0.821
8587 ($E > 0.9$)	Cu + 2S + 2N + 17C	83.1 (75.4)	86.4 (33.0)	0.223	0.787
10164 ($E > 0.8$)	Cu + 2S + 2N + 17C	83.7 (76.0)	86.8 (36.4)	0.218	0.760
11837 ($E > 0.7$)	Cu + 2S + 2N + 17C	84.5 (77.0)	87.3 (36.2)	0.212	0.750

Table 5
Results of phase refinement using *PERP* (starting phases generated from coordinates of Petratos *et al.*, 1988).

No. of reflections with	Starting phases with coordinates	MPEs Initial MPE1 (MPE2)	MPEs Final MPE1 (MPE2)	MCC Initial	MCC Final
7158 ($E > 1.0$)	Cu	83.1 (72.5)	85.6 (30.4)	0.265	0.818
7158 ($E > 1.0$)	Cu + 2S + 2N	82.1 (68.3)	85.7 (30.3)	0.328	0.818
7158 ($E > 1.0$)	Cu + 100 atoms	86.6 (60.5)	85.7 (30.2)	0.440	0.819

advantage. The MPE is higher, as one might expect, but judging from the values of MCC the final map quality was poorer. We also found that with a greater number of reflections, *i.e.* with $|E| > 0.9$, $|E| > 0.8$ and $|E| > 0.7$, the MPE oscillates, as does the MCC.

We repeated the *PERP* refinement but with coordinates taken from Petratos *et al.* (1988). It can be seen in Table 5 that once again, even with refined coordinates, Cu alone gave a final result almost identical to that of adding up to 100 extra light atoms, although the starting MPEs and MCCs were quite different. Starting with phases from known coordinates, both MPEs and MCC converge slowly, although the final MPEs and MCC do not differ appreciably from the corresponding values obtained with the starting phases from coordinates of *MULTAN*88. Throughout the refinement, enantiomorph discrimination is retained until the end.

Since the structure can be solved with initial phase information from the Cu atom alone, it might be thought that the *MULTAN*88 stage of the process was irrelevant and that the Cu could be found from the Patterson function. However, our attempt to find the Cu peaks from a Patterson function was unsuccessful. Perhaps a more assiduous search would have revealed the Cu position, but the *MULTAN*88 procedure was routine and straightforward and therefore preferred.

5. Concluding remarks

This structure could be solved from the heavy-atom position alone plus the use of *PERP* with the DDM1 modification. The addition of up to 21 extra peaks from the map in finding initial phase estimates did not improve matters – on the contrary, the estimates were slightly worse. This must be a consequence of the fact that the atoms represented by these peaks were almost certainly poorly located. It is surprising that the Cu atom alone, with less than 3% of the scattering power of the structure, could provide the first step in an objective procedure leading to a high-quality map at the less-than-atomic resolution of the data. Our experience indicates that presence of even a single heavy atom has a great influence on phase determination for proteins by direct methods (Woolfson & Yao, 1990; Mukherjee & Woolfson, 1993, 1995; Mukherjee, 1999) and also on phase refinement and extension, for example by low-density elimination (Shiono & Woolfson, 1992).

There is an often-expressed view that direct methods can only be applied to proteins when data at better than 1.2 Å resolution is available. The presence of a heavy atom seems to negate this conclusion. We previously found that we could solve 2Zn insulin at 1.5 Å resolution (Mukherjee & Woolfson, 1995) and that useful phase information could be obtained by objective procedures even at 2.25 Å resolution. The present example of pseudoazurin also supports the idea that lower resolution data can be used, although not perhaps with methods that require the detailed interpretation of low-resolution maps with large phase errors. We concluded that just from the heavy-atom positions, determined conveniently by direct methods, plus *PERP* with the DDM1 modification, at least some protein structures can be solved with data at less-than-atomic resolution. The limitations of the method have yet to be explored but since it is automatic and can be applied so easily it merits being regarded as a method of first choice.

Foadi *et al.* (2000) have described a very effective technique in which a protein structure may be solved from a small fragment of another known protein. The steps in their method involve first orienting the fragment and then density modification to refine the phase estimates found from the fragment. Another approach to obtaining an oriented fragment is from a run of a direct method. We shall be looking at the possibility of using this to obtain initial phase estimates when the structure contains no heavy atom.

We wish to express our appreciation to Dr Yao Jia-Xing and Mr James Foadi for their assistance with the *PERP* program, including DDM1.

References

Debaerdemaeker, T., Germain, G., Main, P., Refaat, L. S., Tate, C. & Woolfson, M. M. (1988). *MULTAN*88. *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Universities of York, England and Louvain-la-Neuve, Belgium.

Foadi, J., Woolfson, M. M., Dodson, E., Wilson, K. S. & Yao, J.-X. (2000). *Acta Cryst.* D**56**, 1137–1147.
Hull, S. E. & Irwin, M. J. (1978). *Acta Cryst.* A**34**, 863–870.
Matsugaki, N. & Shiono, M. (1999). *Abstracts of the XVIIIth International Union of Crystallography Congress*. Abstract P12.02.011.
Mukherjee, M. (1999). *Acta Cryst.* D**55**, 820–825.
Mukherjee, M., Ghosh, S. & Woolfson, M. M. (1999). *Acta Cryst.* D**55**, 168–172.
Mukherjee, M. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 9–12.
Mukherjee, M. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 626–628.
Parisini, E., Capozzi, F., Lubini, P., Lamzin, V., Luchinat, C. & Sheldrick, G. M. (1999). *Acta Cryst.* D**55**, 1773–1784.
Petratos, K., Dauter, Z. & Wilson, K. S. (1988). *Acta Cryst.* B**44**, 628–636.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 1036–1040.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1996*a*). *Acta Cryst.* D**52**, 252–256.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1996*b*). *Acta Cryst.* D**52,** 1119–1124.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Sheldrick, G. M. (1993). *SHELXL*93. *Program for the Refinement of Crystal Structures.* University of Göttingen, Germany.
Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.
White, P. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 53–56.
Woolfson, M. M. & Yao, J.-X. (1988). *Acta Cryst.* A**44**, 410–413.
Woolfson, M. M. & Yao, J.-X. (1990). *Acta Cryst.* A**46**, 41–46.
Zhang, K. Y.-J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.