Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Is single-wavelength anomalous scattering sufficient for solving phases? A comparison of different methods for a 2.1 Å structure solution

Liu Yu-dong,[a] I. Harvey,[b] Gu Yuan-xin,[a] Zheng Chao-de,[a] He Yi-zong,[a] Fan Hai-fu,[a]* S. S. Hasnain[b,c] and Q. Hao[b]

[a]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, [b]Department of Chemistry and Physics, De Montfort University, Leicester LE1 9BH, England, and [c]CCLRC Daresbury Laboratory, Warrington WA4 4AD, England

Correspondence e-mail: fan@aphy.iphy.ac.cn

Received 8 March 1999
Accepted 4 June 1999

The structure of rusticyanin is the largest unknown structure (M_r = 16.8 kDa) which has been recently solved by the direct-methods approach using only single-wavelength anomalous scattering (SAS) data from the native protein [Harvey *et al.* (1998). *Acta Cryst.* D**54**, 629–635]. Here, the results of the Sim distribution approach [Hendrickson & Teeter (1981). *Nature (London)*, **290**, 107–113] and of the *CCP*4 procedure *MLPHARE* [Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763] are compared with those from direct methods. Analysis against the final refined model shows that direct methods produced significantly better phases (average phase error 56°) and therefore significantly better electron-density maps than the Sim distribution and *MLPHARE* approaches (average phase error was around 63° in both cases).

1. Introduction

Attempts have long been made to resolve the phase ambiguity arising from single-wavelength anomalous scattering (SAS) without using additional multi-wavelength or isomorphous derivative diffraction data. This is of importance in protein crystallography as most protein crystals are sensitive to X-ray irradiation and isomorphous derivatives are not always easy to prepare. In addition, despite tremendous growth in synchrotron-radiation beam time, it remains a highly valuable resource and any time saving is highly desirable. The now traditional multi-wavelength anomalous diffraction approach (MAD) generally requires a minimum of three wavelengths, and the development of SAS is, therefore, highly significant given the explosion of synchrotron-based structural biology research. MAD experiments have generally been successful (Fourme & Hendrickson, 1990) when performed on specialized instruments, where equivalent segments of data for different wavelengths are acquired sequentially, and as such have required sophisticated experimental protocols. In comparison, an SAS experiment is straightforward and data can be collected in the standard way. Ramachandran & Raman (1956) proposed that for the two possible phases of each reflection, one should always choose the one which has a phase closer to that of the heavy-atom contribution. Hendrickson & Teeter (1981) used a similar but improved method in the structure determination of the hydrophobic protein crambin. Their method combines the bimodal SAS phase distribution with the Sim distribution (Sim, 1959) calculated from the known positions of anomalous scatterers. Wang's solvent-flattening technique (Wang, 1985) uses the same information as input but incorporates the treatment of the 'lack-of-closure error' (Blow & Crick, 1959). Another procedure, *MLPHARE*, is based on maximum-likelihood heavy-atom refinement and phase calculation (Collaborative Computational Project, Number 4, 1994). In a different context, direct methods have been used for many years to try to break the SAS phase ambiguity (Fan, 1965; Karle, 1966; Hazell, 1970; Sikka, 1973; Heinerman *et al.*, 1978; Hauptman, 1982; Giacovazzo, 1983; Fan & Gu, 1985; Kyriakidis *et al.*, 1993). So far, among the above-mentioned direct methods, only that of Fan & Gu (1985) has been tested successfully with experimental SAS data from proteins (Fan *et al.*, 1990; Sha *et al.*, 1995; Zheng *et al.*, 1996). This development has led to the first example of the solution of an unknown protein structure, rusticyanin, with the SAS data at 2.1 Å resolution from a native crystal by a procedure which combines direct methods and density modification (Harvey *et al.*, 1998). Here, a comparison of the direct-methods approach is made with the Sim-distribution approach and *MLPHARE*, clearly demonstrating the superior phases and map from the direct-methods approach.

Rusticyanin is a copper-containing electron-transfer protein with a very high redox potential. No similar structures were available to enable a molecular-replacement solution; molecular-replacement attempts with known cupredoxin structures were unsuccessful. The crystals belong to space group $P2_1$, with unit-

Table 1
Cumulative phase errors.

Reflections were sorted in descending order of $|F_{obs}|$ and cumulated into eight groups. Phase errors (in °) were calculated against the refined model (Harvey *et al.*, 1998) weighted by F_{obs}.

	Results before density modification			Results after density modification		
Number of reflections	Sim	*MLPHARE*	Direct methods	Sim	*MLPHARE*	Direct methods
1000	62.04	62.27	48.63	37.86	43.28	33.95
2000	61.81	61.77	50.97	41.63	45.71	37.43
3000	61.17	61.36	51.76	44.04	46.57	39.01
4000	61.46	61.47	52.89	45.87	48.12	40.79
5000	61.83	61.72	54.12	47.22	49.44	42.26
6000	62.05	62.04	54.73	48.43	50.37	43.36
7000	62.56	62.67	55.61	49.35	51.36	44.47
7898	62.98	63.07	56.09	49.95	51.99	45.17

cell dimensions a = 32.43, b = 60.68, c = 38.01 Å, β = 107.82°. There is one molecule (about 1300 non-H atoms, including one Cu atom) in the asymmetric unit. X-ray diffraction data were collected at a wavelength of 1.376 Å near the Cu K edge; this was chosen to maximize the value of f'' (= 3.87). Only one crystal was used in the experiment and it was aligned carefully so that Friedel pairs were collected on the same image plate.

2. Phasing methods

The phase doublets inherent in the SAS method are expressed as

$$\varphi_{\mathbf{h}} = \varphi'_{\mathbf{h}} \pm |\Delta\varphi_{\mathbf{h}}|, \quad (1)$$

where $\varphi'_{\mathbf{h}}$ and $|\Delta\varphi_{\mathbf{h}}|$ can be calculated from the known positions of the anomalous scatterers and experimental data (see Blundell & Johnson, 1976).

In principle, the phase ambiguity shown in (1) can be broken by many different methods. However, in practice only two kinds of methods have been successful for solving unknown proteins. The first uses the Sim distribution (Sim, 1959), which gives the probability distribution of the phase angle $\varphi_{\mathbf{h}}$ given the phase $\varphi_{\mathbf{h},p}$ of the known partial structure, which in our case consists of the anomalous scatterers,

$$P_{\mathrm{Sim}}(\varphi_{\mathbf{h}}) = [2\pi I_0(x)]^{-1} \times \exp[x\cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{h},p})]. \quad (2)$$

With the expression $\varphi_{\mathbf{h}} = \varphi'_{\mathbf{h}} + \Delta\varphi_{\mathbf{h}}$, (2) becomes

$$P_{\mathrm{Sim}}(\Delta\varphi_{\mathbf{h}}) = [2\pi I_0(x)]^{-1} \times \exp[x\cos(\Delta\varphi_{\mathbf{h}} - \delta_{\mathbf{h}})], \quad (3)$$

where

$$\delta_{\mathbf{h}} = \varphi_{\mathbf{h},p} - \varphi'_{\mathbf{h}}.$$

Notice that $\delta_{\mathbf{h}}$ is always negative in this case; the use of (3) would lead to negative values for all $\Delta\varphi_{\boldsymbol{h}}$ and could cause biased results. The alternative direct-methods approach was proposed by Fan & Gu (1985) and makes use of the product of the Sim distribution and the Cochran distribution (Cochran, 1955). The resulting $\Delta\varphi_{\mathbf{h}}$ could then be positive as well as negative, since the probability that $\Delta\varphi_{\mathbf{h}}$ is positive is given by

$$P_+(\Delta\varphi_{\mathbf{h}}) = 0.5 + 0.5\tanh\{\sin|\Delta\varphi_{\mathbf{h}}| \times [\sum_{\mathbf{h}'} m_{\mathbf{h}'} m_{\mathbf{h}-\mathbf{h}'} \kappa_{\mathbf{h},\mathbf{h}'} \times \sin(\Phi'_3 + \Delta\varphi_{\mathbf{h}',\mathrm{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}',\mathrm{best}}) + \chi\sin\delta_{\mathbf{h}}]\}. \quad (4)$$

For the theory behind and the procedure for using (4), the reader is referred to Fan *et al.* (1984, 1990) and Fan & Gu (1985). It has been pointed out by Kyriakidis *et al.* (1993) that the Cochran distribution is not a very strong criterion for use with protein data. However, (4) is very much strengthened by the following facts: firstly, in (4) a large number of three-phase structure invariants can be used jointly to predict a single phase, usually leading to a much more accurate estimation. Secondly, (4) is only used to make a choice between the two possible signs of $\Delta\varphi_{\mathbf{h}}$, while the Cochran distribution is used to predict a value in the range 0–2π; the latter is surely much more complicated and inaccurate. Finally, the figure of merit $m_{\mathbf{h}}$ in (4) is very efficient in minimizing the error of estimation (Fan *et al.*, 1984).

The density-modification program *DM* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) was used for the phase refinement. The program was running under default control in the recommended mode, which performs solvent flattening, histogram matching and multi-resolution density modification.

3. Calculations and results

A total of 7898 independent reflections at 2.1 Å resolution were used in the calculation. For a comparison, the phase ambiguities were broken by the three procedures described in the previous sections: the Sim distribution (3), *MLPHARE* (Collaborative Computational Project, Number 4, 1994) and direct methods (4). Cumulative phase errors calculated against the refined structure (Harvey *et al.*, 1998) are listed in the second, third and fourth column of Table 1. Density modification (*DM*) was then applied separately to the three resulting phase sets. The cumulative results are listed in the fifth, sixth and seventh column of Table 1. The phase errors for the Sim approach and *MLPHARE* were almost identical before *DM*. Map correlation coefficients were calculated; as the Sim and *MLPHARE* phases are very similar, only the Sim and direct-methods plots are shown in Fig. 1. Portions of the electron-density maps are shown in Figs. 2(*a*) and 2(*b*). It is evident

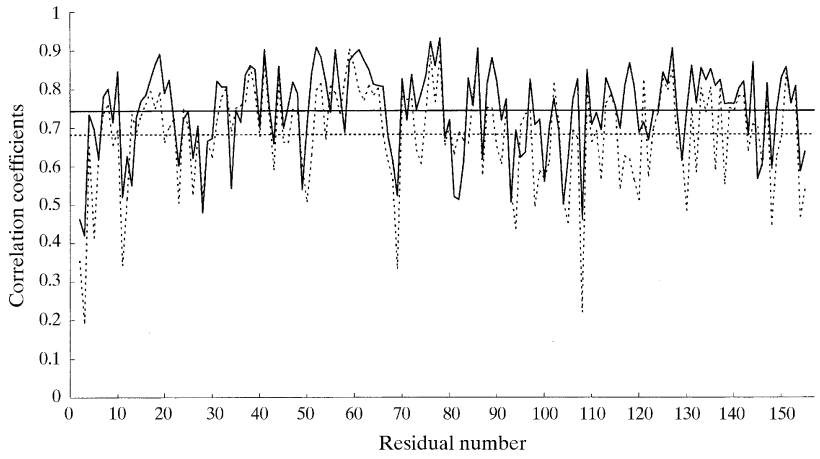

Figure 1
Map correlation coefficient curves (dotted line, Sim phasing plus density modification; solid line, direct-method phasing plus density modification) calculated against the final refined model structure (Harvey *et al.*, 1998) for individual residuals. The horizontal lines represent mean values.

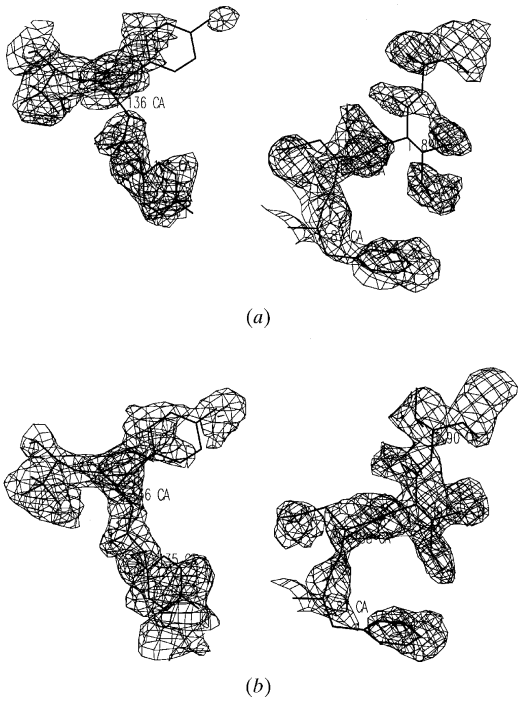

Figure 2
Regions of the electron-density map calculated from (*a*) Sim phasing plus density modification, (*b*) direct-method phasing plus density modification. The final refined model structure (Harvey *et al.*, 1998) is superimposed.

that direct methods produced significantly better phases and better electron-density maps than the Sim-distribution approach or *MLPHARE*. The superior quality of phases from direct methods was crucial in the interpretation and tracing of the density map and in leading to the final structure determination. We thus conclude that combining OAS data and direct methods is a powerful approach for resolving phases for protein structure determination; its wider adoption would result in a major saving of synchrotron-radiation experimental time.

This project is supported in part by the Chinese Academy of Sciences and the National Natural Science Foundation of China. We are grateful to the Chinese Academy of Sciences for awarding a Wang Kuan-Chen Award to QH. We thank Professor Barker, the Vice-Chancellor of De Montfort University for pump-prime funding of the project and CCLRC for provision of facilities. We also thank Dr E. Dodson of York University for writing a script to run *MLPHARE*/*DM*.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 749–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 338. New York: Academic Press.
Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fan, H.-F. (1965). *Acta Phys. Sin.* **21**, 1114–1118.
Fan, H.-F. & Gu, Y.-X. (1985). *Acta Cryst.* A**41**, 280–284.
Fan, H.-F., Han, F.-S. & Qian, J.-Z. (1984). *Acta Cryst.* A**40**, 495–498.
Fan, H.-F., Hao, Q., Gu, Y.-X., Qian, J.-Z., Zheng, C.-D. & Ke, H. (1990). *Acta Cryst.* A**46**, 935–939.
Fourme, R. & Hendrickson, W. A. (1990), *Synchrotron Radiation and Biophysics*, edited by S. S. Hasnain, pp. 156–175. Chichester: Horwood.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Harvey, I., Hao, Q., Duke, E. M. H., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.
Hauptman, H. (1982). *Acta Cryst.* A**38**, 632–641.
Hazell, A. C. (1970). *Nature (London)*, **227**, 269.
Heinerman, J. J. L., Krabbendam, H., Kroon, J. & Spek, A. L. (1978). *Acta Cryst.* A**34**, 447–450.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Karle, J. (1966). *Acta Cryst.* **21**, 273–276.
Kyriakidis, C. E., Peschar, R. & Schenk, H. (1993). *Acta Cryst.* A**49**, 557–569.
Ramachandran, J. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
Sha, B.-D., Liu, S.-P., Gu, Y.-X., Fan, H.-F., Ke, H., Yao, J.-X. & Woolfson, M. M. (1995). *Acta Cryst.* D**51**, 342–346.
Sikka, S. K. (1973). *Acta Cryst.* A**29**, 211–212.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–111.
Zheng, X.-F., Fan, H.-F., Hao, Q., Dodd, F. E. & Hasnain, S. S. (1996). *Acta Cryst.* D**52**, 937–941.