J. Synchrotron Rad. (2000). **7**, 148–151

Direct phasing of one-wavelength anomalous-scattering data

Quan Hao

Department of Chemistry, De Montfort University, Leicester LE1 9BH, UK, and Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China. E-mail: qhao@dmu.ac.uk

(*Received 29 January 2000; accepted 23 February 2000*)

This paper presents a brief survey of methods in *ab initio* phasing of one-wavelength anomalous-scattering data. In particular, the method implemented in the computer program *OASIS* has been tested using two new data sets from orotidine 5′-monophosphate decarboxylase (OMPDC) [Appleby *et al.* (2000). *Proc. Natl Acad. Sci. USA*. In the press] and PurE [Mathews *et al.* (1999). *Structure*, **7**(11), 1395–1406]. The Se atoms were located by the small-molecule program *SAPI*. The electron density maps after *OASIS* and density modification for both structures clearly revealed the $C\alpha$ trace and, in the case of PurE, most side-chains. The test with the OMPDC data demonstrated that, by exploiting the anomalous signal at a single wavelength, direct methods can be used to determine phases at moderate (∼2.5 Å) macromolecular crystallographic resolution for a large-size protein (5663 non-H atoms in the asymmetric unit). The exceptionally good quality of the electron map shown in the case of PurE suggested that fully automatic model fitting is possible.

Keywords: one-wavelength anomalous scattering; direct methods.

1. Introduction

In view of the mounting evidence that one-wavelength anomalous scattering (OAS) may be sufficient to solve protein structures (Brünger, 1999), attempts have long been made to resolve the phase ambiguity arising from one-wavelength anomalous scattering without using additional multiwavelength or isomorphous derivative diffraction data. This is of importance in protein crystallography since 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 multiwavelength anomalous diffraction (MAD) approach generally requires a minimum of three wavelengths and thus the development of OAS is highly significant given the explosion of synchrotron-based structural biology research. The 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 special experimental protocols. In comparison, an OAS experiment is straightforward, where data can be collected in the standard way. Ramachandran & Raman (1956) proposed that for the two possible phases of each reflection one can always make that choice 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 OAS phase distribution with the Sim distribution (Sim, 1959) calculated from the known positions of anomalous scatterers. Wang's density modification 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 in trying to break the OAS 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 successfully tested with experimental OAS data from proteins (Fan *et al.*, 1990; Sha *et al.*, 1995; Zheng *et al.*, 1996). This development has led to the first example of solving an unknown protein structure, rusticyanin, with the OAS data at 2.1 Å resolution from a native crystal by a procedure which combines direct methods and density modification (Harvey *et al.*, 1998). A comparison of the direct-methods approach with the Sim distribution approach and *MLPHARE* demonstrated the superior phases and map from the direct-methods approach (Liu *et al.*, 1999). To further test this method, two data sets are used in the current study: orotidine 5′-monophosphate decarboxylase (OMPDC) (Appleby *et al.*, 2000) and PurE (Mathews *et al.*, 1999). The OAS data (at the wavelength for which f'' is the largest) were taken from the original MAD data sets (see Table 1 for details).

2. Locating the selenium sites

The Se anomalous scatterers for both structures were located by the conventional direct-methods program *SAPI91* (Fan *et al.*, 1990) using magnitudes of anomalous differences,

$$|\Delta\mathbf{F}(\mathbf{H})| = \big||\mathbf{F}(\mathbf{H})| - |\mathbf{F}(-\mathbf{H})|\big|,$$

for reflections within 3.0 Å. The solution was selected by a default run of the program. The largest 800 (OMPDC) and 400 (PurE) normalized structure factors E's were used in tangent formula phase refinement. The resultant electron density map produced a group of 18 and 4 highest peaks for OMPDC and PurE, respectively; there was a clear gap between this group and other peaks in terms of peak height. The absolute configuration of these sites was determined by the P_s-function-based method (Woolfson & Yao, 1994). These Se sites formed the basis for the next phasing step.

3. Evaluation of phase doublets

The phase doublets inherent in the OAS method are expressed as

$$\varphi_H = \varphi'_H \pm \big|\Delta\varphi_H\big|, \tag{1}$$

where φ'_H is the phase of

$$F''_{\rm ano} = \sum_{j=1}^{N} i f''_j \exp(i\,2\pi H.r_j), \tag{2}$$

which can be calculated from the known positions of the anomalous scatterers and the known value of f''; $\big|\Delta\varphi_H\big|$ is obtained from (Blundell & Johnson, 1976)

$$\cos\Delta\varphi_H = (F_H^+ - F_H^-)/2\big|F''_{\rm ano}\big|. \tag{3}$$

The phase problem in the OAS case is in fact a sign problem according to (1). The probability for $\Delta\varphi_H$ positive is given by Fan & Gu (1985),

$$P_+(\Delta\varphi_H) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left\{\sin|\Delta\varphi_H|\left[\sum_{H'} m_{H'} m_{H-H'} \kappa_{H,H'} \times \sin\big(\Phi'_3 + \Delta\varphi_{H',\rm best} + \Delta\varphi_{H-H',\rm best}\big) + \chi\sin\delta_H\right]\right\}. \tag{4}$$

The procedure of using (4) for *ab initio* phasing of the OAS data was implemented in the computer program *OASIS* (Hao *et al.*, 2000). All Friedel pairs (including centric reflections) were evaluated using *OASIS*.

Density modification using the CCP4 program *DM* (Collaborative Computational Project, Number 4, 1994) was then applied to the resulting phase sets. Phase error analysis and figures of merit before and after *DM* are given in Table 2. The electron density maps after *OASIS* and density modification (Figs. 1 and 2) for both structures clearly revealed the *Cα* trace. In the case of PurE, most side-chains were well defined – the exceptionally good map quality was due to the high resolution and high redundancy of the diffraction data. The MAD + *DM* phased electron density maps in the region are also shown for comparisons. A correlation coefficient between the *OASIS* + *DM* phased map, the MAD + *DM* phased map and the final refined structure was 0.611, 0.753, respectively, for OMPDC, and 0.841, 0.876 for PurE.

Table 1
OAS data.

Values in parentheses refer to the highest resolution shell. The abbreviation a.s.u. stands for asymmetric unit.

	OMPDC	PurE
Space group	$P2_12_12$	$I422$
Unit cell	a = 78.93 Å	a = 112.43 Å
	b = 89.94 Å	b = 112.43 Å
	c = 105.47 Å	c = 49.16 Å
Non-H atoms in a.s.u.	5663	1367
Content of a.s.u.	78.0 kDa	17.6 kDa
Number of Se sites in a.s.u.	18	4
Source	APS beamline 19-ID	CHESS beamline F2
Data-collection protocol	Inverse beam	Inverse beam
Wavelength	λ = 0.9790 Å	λ = 0.9790 Å
f'' (in electrons)	3.879	3.879
Resolution	20–2.5 Å	20–1.4 Å
Unique reflections	26338	24277
Completeness	95.0%	94.7%
Redundancy	3.5	8.0
$I/\sigma(I)$	18.64 (4.74)	9.2 (2.2)
$R_{\rm sym}$	7.7%	5.5%

4. Discussion

Recently there has been a tremendous interest in the use of direct methods for phase determination for macromolecules. This surge of interest has primarily resulted from two factors: one has been the ability to obtain atomic-resolution (<1.3 Å) data in favourable cases and the other has been the development of some powerful methods including traditional direct methods ('shake and bake'), so-called 'half-baked' and combinations of direct methods with isomorphous replacement and/or anomalous scattering. The ultimate potential of the traditional direct methods is still unknown but one limit appears to be certain and that is the requirement for atomic-resolution data. Here we demonstrate that, by exploiting the anomalous signal at a single wavelength, direct methods can be used to determine phases at moderate (∼2.5 Å) macromolecular crystallographic resolution for a large-size protein. The method provides a powerful alternative in solving a *de novo* protein structure without either preparing isomorphous heavy-atom derivative crystals or collecting multi-wavelength diffraction data. It is worth noting that these data were originally intended for MAD phasing and not optimized for one-wavelength phasing. It has been suggested that the *OASIS* approach could be used for proteins with molecular weights of up to 33 kDa per Se by exploring the 'white line' at the Se absorption edge

Table 2
Phase error analysis and figure of merit.

Reflections were sorted in descending order of F_{obs} and cumulated into groups. Phase errors were calculated against the refined models (Appleby *et al.*, 2000; Mathews *et al.*, 1999) weighted by F_{obs}. As a comparison, SAD phases were calculated with the CCP4 program *MLPHARE* (Collaborative Computational Project, Number 4, 1994).

	Phase errors (°)					
	OMPDC			PurE		
Number of reflections	*MLPHARE*	*OASIS*	*OASIS* + *DM*	*MLPHARE*	*OASIS*	*OASIS* + *DM*
4000	68.3	52.3	37.0	59.0	36.3	21.6
8000	68.8	55.0	41.7	58.4	39.6	24.4
12000	70.3	57.9	45.5	58.7	41.6	26.3
16000	71.3	59.8	48.3	59.0	43.1	27.9
20000	72.2	61.5	50.7	59.6	44.6	29.5
24000	73.0	62.9	52.5	60.3	46.0	31.0
26338	73.4	63.4	53.2			
Mean figure of merit	0.25	0.68	0.76	0.38	0.75	0.85

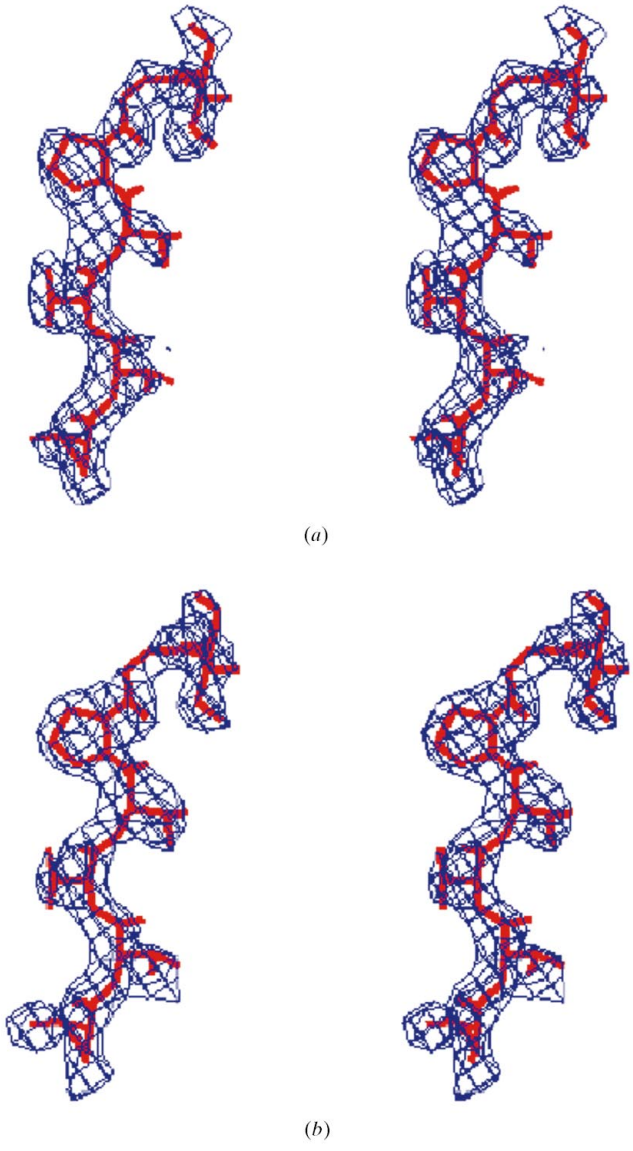

Figure 1
Stereoview of the representative electron density (residuals 178–184) of OMPDC: (*a*) *OASIS* + *DM* phasing and (*b*) MAD + *DM* phasing, with final coordinates superimposed (both contoured at 1σ).

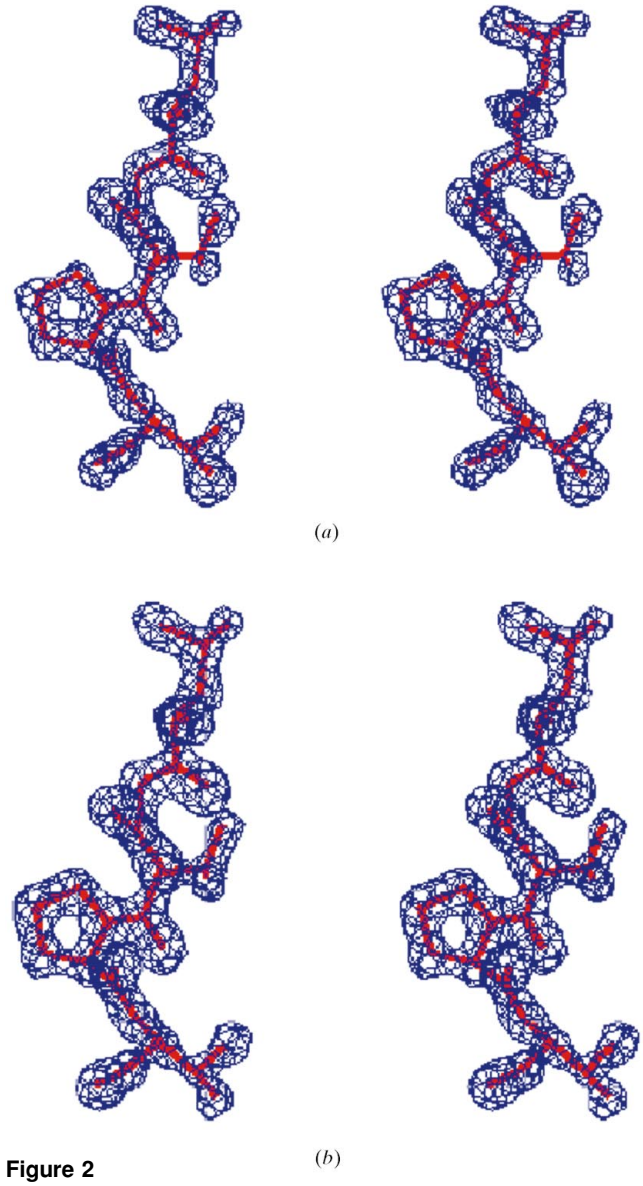

Figure 2
Stereoview of the representative electron density (residuals 89–93) of PurE: (*a*) *OASIS* + *DM* phasing and (*b*) MAD + *DM* phasing, with final coordinates superimposed (both contoured at 1σ).

(Harvey *et al.*, 1998). It has also been proposed that it might be more efficient to collect very highly redundant single-wavelength data than to collect multiple-wavelength data, from a point of view of phasing. Indeed, there is little difference in terms of map quality between *OASIS* and MAD phased maps in the case of PurE where the data redundancy is high. The exceptional quality of the electron map suggests that fully automatic model fitting is possible.

I would like to thank Professor S. Ealick, Drs I. Mathews and T. Appleby for making available PurE and OMPDC data. I am also grateful to De Montfort and Cornell Universities for sponsoring my sabbatical leave. Professors S. Hasnain and H. Fan are thanked for useful discussions. This project is supported by the National Key Basic Research Special Funds of China, No. G1999075604 and the Royal Society.

References

Appleby, T. C., Kinsland, C. L., Begley, T. P. & Ealick, S. E. (2000). *Proc. Natl Acad. Sci. USA*. In the press.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 338. New York: Academic Press.
Brünger, A. (1999). XVIIIth IUCr Congress and General Assembly, Glasgow, UK, 4–13 August 1999. M13.BB.002.
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., 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. (1993). *Acta Cryst.* A**39**, 585–592.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* In the press.
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.
Liu, Y. D., Harvey, I., Gu, Y. X., Zheng, C. D., He, Y. Z., Fan, H. F., Hasnain, S. S. & Hao, Q. (1999). *Acta Cryst.* D**55**, 1620–1622.
Mathews, I. I., Kappock, T. J., Stubbe, J. & Ealick, S. E. (1999). *Structure*, **7**(11), 1395–1406.
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.
Woolfson, M. M. & Yao, J. X. (1994). *Acta Cryst.* D**50**, 7–10.
Zheng, X. F., Fan, H. F., Hao, Q., Dodd, F. E. & Hasnian, S. S. (1996). *Acta Cryst.* D**52**, 937–941.