Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Direct-method-aided phasing of MAD data

Y. X. Gu, Y. D. Liu, Q. Hao, S. E. Ealick and H. F. Fan

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Direct-method-aided phasing of MAD data

Y. X. Gu,[a] Y. D. Liu,[a] Q. Hao,[a,b] S. E. Ealick[c] and H. F. Fan[a]*

[a]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, [b]Department of Chemistry, De Montfort University, Leicester LE1 9BH, England, and [c]Department of Chemistry and Chemical Biology, The Baker Laboratory, Cornell University, Ithaca, NY 14853, USA

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

Received 20 September 2000
Accepted 23 November 2000

The direct methods of breaking the phase ambiguity intrinsic in one-wavelength anomalous scattering (OAS) data and MAD phasing are powerful methods in their own rights. In a different context, in addition to their success in phasing OAS data, direct methods can also be useful in the treatment of MAD data. The idea has been tested with the MAD data at 2.5 Å resolution from the protein human adenosine kinase [Mathews *et al.* (1998), *Biochemistry*, **37**, 15607–15620]. The results showed that the incorporation of direct methods in MAD phasing led to a significant improvement of phases over those obtained from the conventional MAD phasing method alone, as indicated by improved map correlation coefficients (with the existing model), reduced phase errors by 4.5° and improved map connectivity.

1. Introduction

The direct method of breaking the phase ambiguity intrinsic in one-wavelength anomalous scattering (OAS) data was initiated by Fan (1965) and further improved by Fan & Gu (1985). The method has been successfully tested with a number of known proteins (Fan *et al.*, 1990; Sha *et al.*, 1995; Zheng *et al.*, 1996). It has also been applied to solve an unknown protein structure (rusticyanin) with a molecular weight of 16.8 kDa at 2.1 Å resolution (Harvey *et al.*, 1998; Liu *et al.*, 1999). A recent study by Hao (2000) showed that a number of proteins originally solved by the MAD method could have been solved as successfully by direct-method phasing of OAS data. A program *OASIS* (Hao *et al.*, 2000) based on the procedure used by Fan *et al.* (1990) is now available in the latest version of the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) for phasing OAS and SIR protein data. In a different context, besides their success in phasing OAS data, direct methods can also be useful in the treatment of MAD data. In this paper, we provide an example which shows that the incorporation of direct methods in MAD phasing led to initial phases better than those obtained from the conventional MAD phasing method alone.

2. Phasing strategy

A set of MAD data consists of several sets of OAS data at different wavelengths. The conventional phasing of MAD data is to combine the bimodal phase distribution of the corresponding OAS data sets to give a unique phase indication for individual reflections. Since the anomalous scattering effect is relatively weak, the resultant phase indications are also weak. Typically, for a protein of moderate size with diffraction data at 2–3 Å resolution, about a third of the total reflections may

have MAD phases with a figure of merit less than 0.3. This explains why in some cases MAD phases are not good enough to initiate a phase-improvement procedure such as solvent flattening. On the other hand, direct-method phasing of OAS data within a MAD data set yields phases independent of conventional MAD phases. Although the indications would not be very strong, they provide additional information which may help to improve the quality of MAD phases. Furthermore, a set of MAD phases as a whole may not be very good; however, they may contain a subset, say those with the figure of merit ≥ 0.9, which are accurate enough to serve as starting phases in a direct-method procedure. As has been shown in the single isomorphous replacement (SIR) case (Zheng *et al.*, 1997), a small subset of reliable starting phases can greatly strengthen the direct-method procedure of breaking the phase ambiguity. In view of the above, the phasing strategy will consist of the following steps.

(i) A set of MAD phases is obtained using the conventional MAD phasing method. The subset of these with figures of merit ≥ 0.9 is extracted and used as starting phases in the next step.

(ii) Each OAS set of the MAD data is treated separately by a direct-method procedure similar to that used by Fan *et al.* (1990).

The phase ambiguity in the OAS case is expressed as

$$\varphi_{\mathbf{h}} = \varphi''_{\mathbf{h},A} \pm |\Delta\varphi_{\mathbf{h}}|. \tag{1}$$

Here, $\varphi_{\mathbf{h}}$ denotes the phase of

$$\mathbf{F}_{\mathbf{h}} = (\mathbf{F}^{+}_{\mathbf{h}} + \mathbf{F}^{-*}_{\mathbf{h}})/2. \tag{2}$$

$\mathbf{F}^{-*}_{\mathbf{h}}$ is the complex conjugate of $\mathbf{F}^{-}_{\mathbf{h}}$.

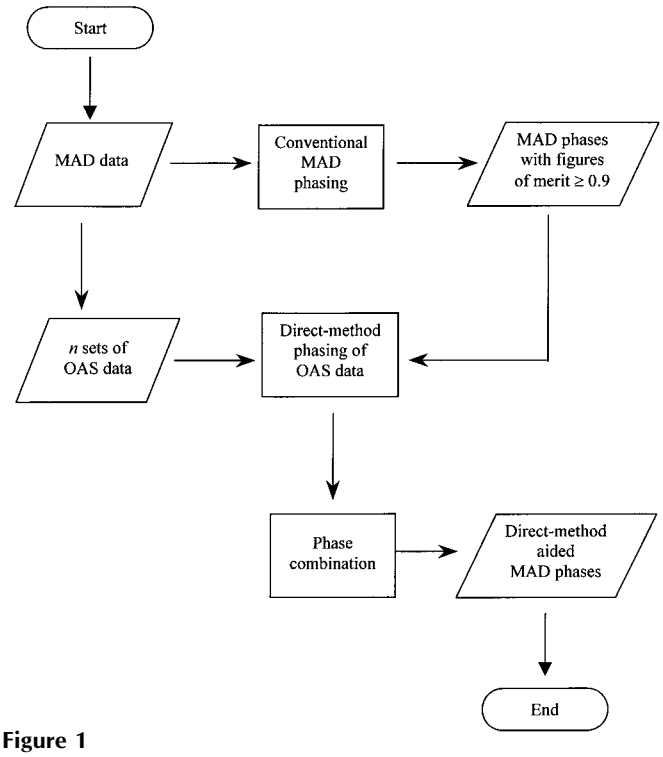

Figure 1
Flowchart of the process of direct-method-aided MAD phasing.

Table 1
Summary of the test data from the protein adenosine kinase.

The abbreviation a.s.u. stands for asymmetric unit.

Space group	$P2_12_12$		
Unit-cell parameters (Å)	a = 64.528, b = 109.841, c = 49.231		
Non-H atoms in a.s.u.	2695 (3093 including water molecules)		
Number of Se sites in a.s.u	6		
Source	X-12C at National Synchrotron Light Source		
Data-collection protocol	Inverse beam		
Wavelength (Å)	0.9790	0.9787	0.9500
f' (electrons)	−7.381	−6.830	−2.739
f'' (electrons)	3.840	3.838	3.605
Resolution (Å)	20–2.5		
Unique reflections	12525		
Completeness (%)	99.0		

On the right-hand side of (1), $\varphi''_{\mathbf{h},A}$ is the phase of

$$\mathbf{F}''_{\mathbf{h},A} = \sum_{A=1}^{N_A} i\Delta f''_A \exp(i2\pi\mathbf{h}\cdot\mathbf{r}_A), \tag{3}$$

$$\Delta\varphi_{\mathbf{h}} = \pm\cos^{-1}[(F^{+}_{\mathbf{h}} + F^{-}_{\mathbf{h}})/2F''_{\mathbf{h},A}], \tag{4}$$

where $F''_{\mathbf{h},A}$ is the amplitude of $\mathbf{F}''_{\mathbf{h},A}$ (see Kartha, 1975).

The sign of $\Delta\varphi_{\mathbf{h}}$ is now derived using the formulae

$$P_{+}(\Delta\varphi_{\mathbf{h}}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\left\{\sin|\Delta\varphi_{\mathbf{h}}| \times\left[\sum_{\mathbf{h}'} m_{\mathbf{h}'} m_{\mathbf{h}-\mathbf{h}'}\kappa_{\mathbf{h},\mathbf{h}'}\sin(\Phi'_3 + \Delta\varphi_{\mathbf{h}'\,\mathrm{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}'\,\mathrm{best}}) + x\sin\delta_{\mathbf{h}}\right]\right\}, \tag{5}$$

where $\Phi'_3 = -\varphi''_{\mathbf{h},A} + \varphi''_{\mathbf{h}',A} + \varphi''_{\mathbf{h}-\mathbf{h}',A}$ and $x\sin\delta_{\mathbf{h}}$ is the Sim weight (Sim, 1959),

$$\Delta\varphi_{\mathbf{h}\,\mathrm{best}} = \varphi_{\mathbf{h}\,\mathrm{best}} - \varphi''_{\mathbf{h},A}. \tag{6}$$

During the derivation of P_{+}, reflections are divided into two categories: (i) those having the MAD phase with the figure of merit ≥ 0.9 and (ii) the remaining reflections. For reflections of category (i), values of $\Delta\varphi_{\mathbf{h}\,\mathrm{best}}$ are obtained by replacing $\varphi_{\mathbf{h}\,\mathrm{best}}$ in (6) with the MAD phase, while values of $m_{\mathbf{h}}$ are set to 1.0. Phases of these reflections are kept fixed in the process. For reflections of category (ii), values of $\Delta\varphi_{\mathbf{h}\,\mathrm{best}}$ and $m_{\mathbf{h}}$ are calculated each cycle using (7) and (8) with the initial P_{+} set to $\frac{1}{2}$.

$$\tan(\Delta\varphi_{\mathbf{h}\,\mathrm{best}}) = 2(P_{+} - \tfrac{1}{2})\sin|\Delta\varphi_{\mathbf{h}}|/\cos\Delta\varphi_{\mathbf{h}}, \tag{7}$$

$$m_{\mathbf{h}} = \exp(-\sigma^2_{\mathbf{h}}/2)\{[2(P_{+} - \tfrac{1}{2})^2 + \tfrac{1}{2}] \times (1 - \cos 2\Delta\varphi_{\mathbf{h}}) + \cos 2\Delta\varphi_{\mathbf{h}}\}^{1/2}, \tag{8}$$

where $\sigma_{\mathbf{h}}$ is related to the experimental error and can be calculated from the mean square of the 'lack-of-closure error' (Blow & Crick, 1959). For the theory behind and the practical application of (5) to (8), the reader is referred to Fan & Gu (1985), Fan *et al.* (1990) and Hao *et al.* (2000). Upon completion of this step, there will be a number of direct-method-resolved OAS phase sets.

(iii) The direct-method resolved OAS phase sets are combined using the formulae

$$\varphi_{\rm Combined} = \tan^{-1}\left[\frac{\sum_{j=1}^{n}(m_{\mathbf{h}}\sin\varphi_{\rm best})_j}{\sum_{j=1}^{n}(m_{\mathbf{h}}\cos\varphi_{\rm best})_j}\right], \quad (9)$$

$$(m_{\mathbf{h}})_{\rm Combined} = \frac{\left\{\left[\sum_{j=1}^{n}(m_{\mathbf{h}}\sin\varphi_{\rm best})_j\right]^2 + \left[\sum_{j=1}^{n}(m_{\mathbf{h}}\cos\varphi_{\rm best})_j\right]^2\right\}^{-1/2}}{n}, \quad (10)$$

where n is the number of phase sets (one at each wavelength) involved in the combination. Such a combination can be considered as a reciprocal-space equivalent of calculating a sum function of Fourier maps corresponding to n sets of direct-method-resolved OAS phases.

Fig. 1 shows a flowchart of the phasing process.

3. Test results and discussion

The method was tested with the MAD data set of the SeMet-incorporated crystals of human adenosine kinase (Mathews *et al.*, 1998; see Table 1 for details). The structure was originally solved by the conventional MAD phasing and density-modification method implemented in the programs *MLPHARE* and *DM* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). A total of 12 525 reflections at 2.5 Å resolution were used in the calculation as described in the flowchart (Fig. 1). The results are listed in Table 2(*a*). The direct-method-aided MAD phasing (column 4) was significantly better than MAD phasing alone (column 3). The difference in terms of phase errors after density modification (columns 5 and 6) appeared to be small, but as the two phasing methods have different error sources the combined MAD plus dm and direct-method-aided MAD plus dm phases (column 7) using (9) and (10) were significantly better than the MAD plus dm phases alone. It is interesting to see that the most substantial benefit of the DMAD treatment was in the low $\rm FOM_{mad}$ range (Table 2*b*). The improvement in phases led to higher map correlation coefficients against the refined structure (Mathews *et al.*, 1998), as shown in Fig. 2. The stronger reflections, which include most low-resolution ones, benefitted more from the improvement (Table 2); therefore, the connectivity of the electron-density map was improved significantly (Fig. 3).

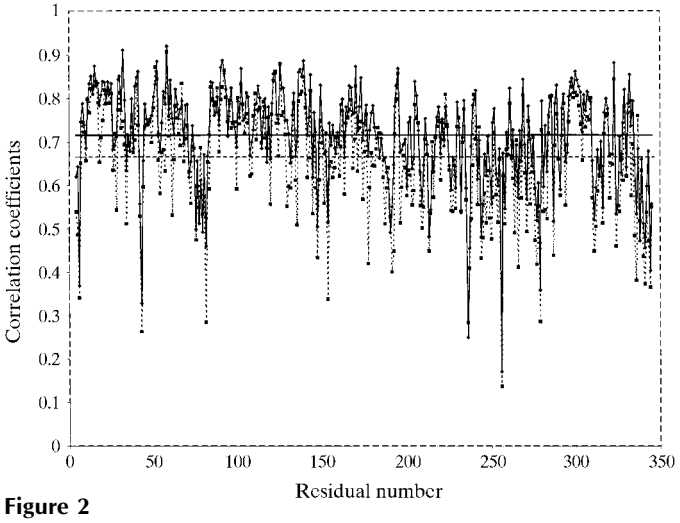

Figure 2
Map correlation coefficient curves [dotted line, MAD phasing plus density modification (dm); solid line, combined MAD + dm and direct-method-aided MAD + dm] calculated against the final refined model structure (Mathews *et al.*, 1998) for individual residuals. The horizontal lines represent mean values.

Table 2
Phase-error analysis.

MAD, conventional MAD phasing using *MLPHARE* in *CCP*4; DMAD, direct-method-aided MAD phasing; dm, density-modification (Zhang & Main, 1990) package in *CCP*4.

(*a*) Cumulative mean phase errors with reflections sorted by $F_{\rm obs}$.

		Mean phase error (unweighted)				
$\langle F_{\rm obs}\rangle$	No. of reflections	MAD	DMAD	MAD + dm	DMAD + dm	Combined MAD + dm and DMAD + dm
844.7	1000	57.78	40.91	31.68	27.51	26.42
725.3	2000	58.69	44.55	36.84	33.16	31.54
651.6	3000	59.79	48.54	40.10	38.10	35.65
597.2	4000	60.96	51.10	42.89	41.00	38.73
553.5	5000	62.42	53.57	46.19	44.49	42.55
516.6	6000	63.94	55.37	48.82	47.21	45.20
484.4	7000	64.82	56.93	51.02	49.73	47.45
455.4	8000	65.43	58.39	53.01	51.75	49.48
428.8	9000	66.16	59.82	54.80	54.13	51.54
403.7	10000	66.97	61.15	56.36	56.02	53.32
379.4	11000	68.00	63.00	58.31	58.16	55.45
354.8	12000	69.44	65.17	60.61	60.63	57.90
341.6	12516	70.46	66.00	61.97	61.90	59.27

(*b*) Mean phase errors with reflections grouped and sorted by the MAD figure of merit.

		Mean phase error (unweighted)				
$\langle \rm FOM_{mad}\rangle$	No. of reflections	MAD	DMAD	MAD + dm	DMAD + dm	Combined MAD + dm and DMAD + dm
0.99	1000	39.25	39.25	39.24	39.21	39.33
0.95	1000	48.19	48.19	47.42	48.11	47.24
0.92	1000	53.63	53.63	51.78	52.40	51.36
0.87	1000	57.52	67.63	54.42	63.76	54.20
0.80	1000	61.51	71.11	59.26	66.57	58.43
0.70	1000	65.80	71.81	61.09	65.57	59.76
0.57	1000	75.44	79.17	68.13	68.60	67.41
0.41	1000	78.74	79.84	70.48	72.16	69.46
0.24	1000	84.43	83.36	74.36	75.47	73.34
0.06	1000	87.14	80.77	75.38	70.37	71.87
0.00	1000	92.35	58.93	67.90	56.41	56.05
0.00	1000	89.77	59.13	69.47	57.56	57.56
0.00	516	92.52	54.38	67.24	60.52	60.87

4. Concluding remarks

MAD phasing and direct-method OAS phasing are powerful methods in their own rights. Here, we have demonstrated that the combination of the two can provide an even better tool. This is particularly useful when MAD data at different

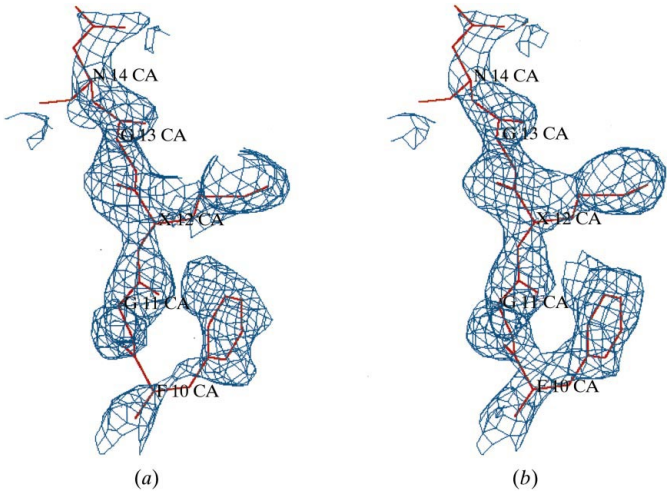

Figure 3
A region (residuals 10–14) of electron-density map calculated from (*a*) MAD phasing plus density modification, (*b*) combined MAD + dm and direct-method-aided MAD + dm. The final refined model structure (Mathews *et al.*, 1998) is superimposed.

wavelengths are correlated, *i.e.* when $\Delta f''$ are small, as direct-method OAS phasing does not depend on f''.

We are very grateful to Dr I. I. Mathews for providing the human adenosine kinase MAD data and useful discussions. This project is supported by the Chinese Academy of Sciences and by NKBRSF, People's Republic of China, No. G1999075604.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
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.
Hao, Q. (2000). *J. Synchrotron Rad.* **7**, 148–151.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
Harvey, I., Hao, Q., Duke, E. M. H., Ingledew, W. J. & Hasnain, S. S. (1998). *Acta Cryst.* D**54**, 629–635.
Kartha, G. (1975). *Anomalous Scattering*, edited by S. Ramaseshan & S. C. Abrahams, p 369. Copenhagen: Munksgaard.
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., Erion, M. D. & Ealick, S. E. (1998). *Biochemistry*, **37**, 15607–15620.
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.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.
Zheng, X. F., Fan, H. F., Hao, Q., Dodd, F. E. & Hasnain, S. S. (1996). *Acta Cryst.* D**52**, 937–941.
Zheng, X. F., Zheng, C. D., Gu, Y. X., Mo, Y. D., Fan, H. F. & Hao, Q. (1997). *Acta Cryst.* D**53**, 49–55.