Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Optimizing the error term in direct-method SAD phasing

J. W. Wang, J. R. Chen, Y. X. Gu, C. D. Zheng, F. Jiang and H. F. Fan*

Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China

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

Received 10 June 2004
Accepted 23 September 2004

The probability formula of the direct-method SAD (single-wavelength anomalous diffraction) phasing proposed by Fan & Gu (1985, *Acta Cryst.* A**41**, 280–284) contains an error term which is related to the lack-of-closure error. This error term is used as a weighting function in the phase derivation and in the subsequent calculation of electron-density maps. Previously, there has been a constant in the error term that has had to be determined empirically for each particular case. It has been found that improper choice of the constant often leads to failure of the direct-method SAD phasing. The problem is resolved by introducing a modified error term and a method of automatically tuning the associated scaling factor.

1. Introduction

In the direct-method SAD (single-wavelength anomalous diffraction) phasing proposed by Fan and coworkers (Fan, Han & Qian, 1984; Fan, Han, Qian *et al.*, 1984; Fan & Gu, 1985), the phase of reflections is expressed as

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

where φ'' is the phase of imaginary-part scattering from the heavy-atom (anomalous-scatterer) substructure, *i.e.*

$$F''_{\mathbf{h}} = i\sum_{j=1}^{N} f''_j \exp(i2\pi\mathbf{h}\cdot\mathbf{r}_j). \tag{2}$$

Given the known heavy-atom substructure, the absolute value of $\Delta\varphi$ can be calculated as (see Blundell & Johnson, 1976)

$$|\Delta\varphi| \simeq \cos^{-1}[(|F^+| - |F^-|)/2|F''|] \equiv \cos^{-1}\left(\frac{\Delta F}{2|F''|}\right). \tag{3}$$

The sign of $\Delta\varphi$ is then estimated by the probability of $\Delta\varphi$ being positive

$$P_+(\Delta\varphi_{\mathbf{h}}) = \frac{1}{2} + \frac{1}{2}\tanh\left\{\sin|\Delta\varphi_{\mathbf{h}}|\left[\sum_{\mathbf{h}'} m_{\mathbf{h}'} m_{\mathbf{h}-\mathbf{h}'} \times \kappa\sin(\Phi'_3 + \Delta\varphi_{\mathbf{h}'\text{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}'\text{best}}) + \chi\sin\delta_{\mathbf{h}}\right]\right\}, \tag{4}$$

where

$$\kappa = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}|, \tag{5}$$

$$\Phi'_3 = \varphi''_{-\mathbf{h}} + \varphi''_{\mathbf{h}'} + \varphi''_{\mathbf{h}-\mathbf{h}'}, \tag{6}$$

$\chi\sin\delta_h$ is from the Sim distribution (Sim, 1959) related to the heavy-atom substructure, $\Delta\varphi_{\mathbf{h}\text{best}}$ is defined as

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

and $m_{\mathbf{h}}$ is the figure of merit for $E_{\mathbf{h}}$ and is expressed as

$$m_{\mathbf{h}} = \exp(-\sigma_{\mathbf{h}}^2/2)\big(\{2[P_+(\Delta\varphi_{\mathbf{h}}) - \tfrac{1}{2}]^2 + \tfrac{1}{2}\} \times (1 - \cos 2\Delta\varphi_{\mathbf{h}}) + \cos 2\Delta\varphi_{\mathbf{h}}\big)^{\frac{1}{2}}. \quad (8)$$

The value of $P_+(\Delta\varphi)$ in (7) and (8) is initially set to 1/2 and will be updated in each cycle during the phase iteration. The term $\exp(-\sigma_{\mathbf{h}}^2/2)$ in (8) is related to the lack-of-closure error, in which

$$\sigma_{\mathbf{h}}^2 = \frac{D^2}{4|F''_{\mathbf{h}}|^2 \sin^2 \Delta\varphi_{\mathbf{h}}} \quad (9)$$

and D is the lack-of-closure error defined by Blow & Crick (1959). Actually, $\exp(-\sigma_{\mathbf{h}}^2/2)$ is nothing other than a weighting function associated with individual reflections. For details of (4)–(9) the reader is referred to Woolfson & Fan (1995*a*,*b*). In the program *OASIS* (Hao *et al.*, 2000), the term $\sin^2\Delta\varphi_{\mathbf{h}}$ in (9) is replaced by its averaged value $\langle\sin^2\Delta\varphi_{\mathbf{h}}\rangle = 1/2$ and the lack-of-closure error D is replaced by a constant to be set empirically for each particular case. Thus,

$$\sigma_{\mathbf{h}}^2 = \frac{K^2}{2|F''_{\mathbf{h}}|^2}. \quad (10)$$

It has been found that failures of *OASIS* are often a consequence of an improper choice of the constant K. In this paper a modified expression is given, *i.e.*

$$\sigma_{\mathbf{h}}^2 = \frac{(n\sigma_{\Delta F_{\mathbf{h}}})^2}{2|F''_{\mathbf{h}}|^2}, \quad (11)$$

where $\sigma_{\Delta F_{\mathbf{h}}}$ is the measured standard deviation of the Bijvoet difference $\Delta F_{\mathbf{h}}$ and n is a scaling factor. A method for automatically tuning the scaling factor n is also proposed. The above modification avoids manual intervention and leads to good phasing results for a wide variety of sample proteins.

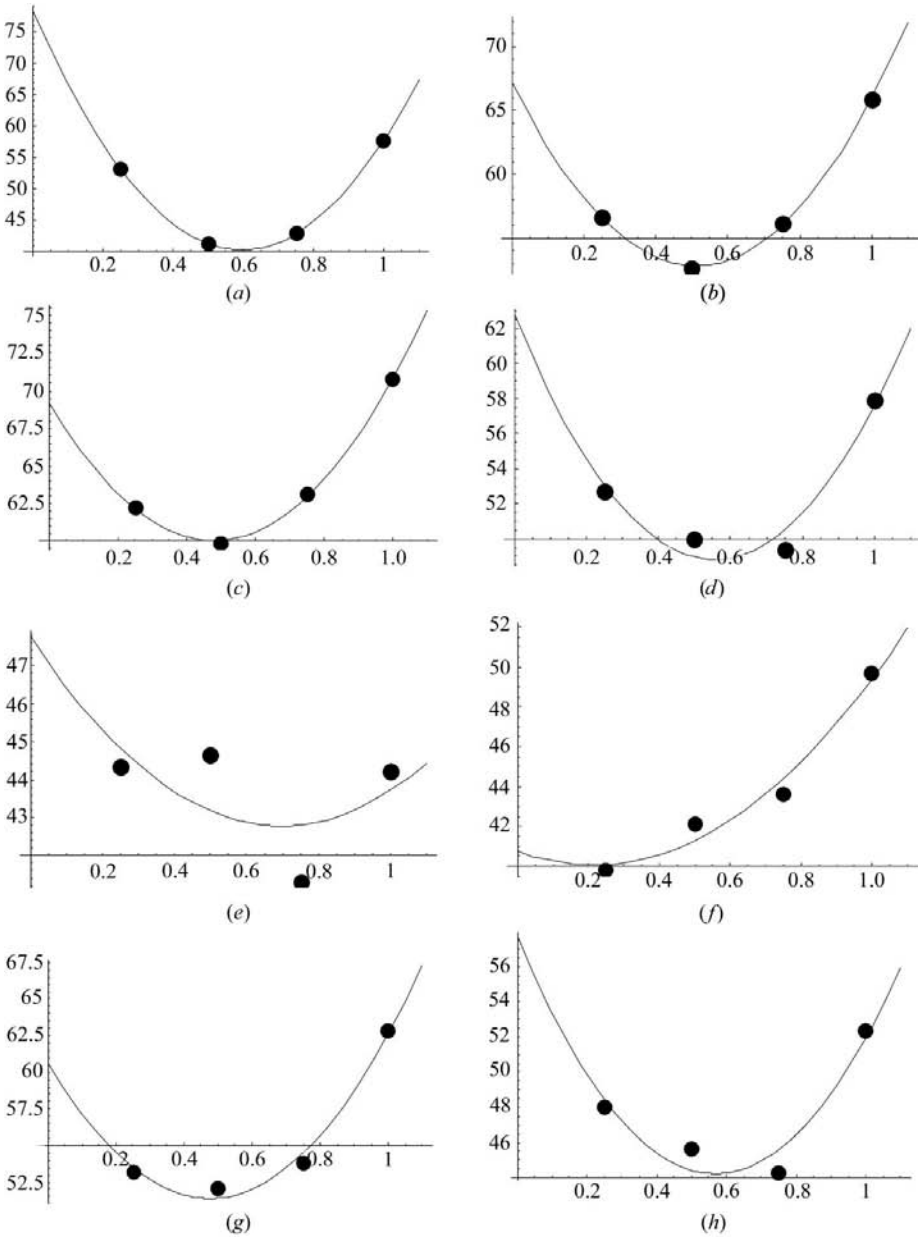

Figure 1
Overall average phase error (y axis) plotted against $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$ (x axis) for eight sample proteins. (*a*) Glucose isomerase, (*b*) xylanase, (*c*) azurin, (*d*) rusticyanin, (*e*) histone methyltransferase SET7/9, (*f*) KD93, (*g*) human acyl protein thioesterase and (*h*) porcine pancreatic elastase. The curve in each plot is the quadratic-equation fit to the discrete points.

2. Test samples

Experimental SAD data from eight known proteins (summarized in Table 1) were used in the test calculation. The size of proteins ranges from 129 to 560 residues per asymmetric unit. Anomalous scattering atoms in the sample proteins are S, Cu, Se, Br and Xe. The Bijvoet ratio changes from 0.68 to 8.78%.

3. Influence of $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$ on the phasing error

From (8) and (11) it is evident that the overall average value of $\exp(-\sigma_{\mathbf{h}}^2/2)$ would affect the resultant overall average phase error. In the following test, for each sample protein, by tuning the value n in (11) we set $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$ equal to 0.25, 0.50, 0.75 and 1.00 in turn. Phases were derived for each condition by *OASIS* and then improved using the program *DM* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Results are summarized in Table 2. We observe the following.

(i) For all sample proteins, the condition corresponding to the lowest overall average phase error is not

Table 1
Summary of the test data.

Protein	Residues in AU	Anomalous scatterer	λ (Å)	f'' (e)	$\langle\lvert\Delta F\rvert\rangle/\langle F\rangle$ (%)	Multiplicity	Reference
Glucose isomerase	388	S (9)	1.54	0.56	0.68	17.4	Ramagopal *et al.* (2003)
Xylanase	303	S (5)	1.74	0.70	0.69	12.0	Ramagopal *et al.* (2003)
Azurin	129	Cu (1)	0.97	2.21	1.44	10.0	Dodd *et al.* (1995)
Rusticyanin	155	Cu (1)	1.376	3.88	2.36	10.2	Harvey *et al.* (1998)
Histone methyltransferase SET7/9	560	Se (12)	0.9794	5.60	7.03	3.8	Wilson *et al.* (2002)
KD93	188	Se (5)	0.9712	3.79	6.60	23.0	Chen *et al.* (2004)
Human acyl protein thioesterase	464	Br (22)	0.9167	5.0	8.78	3.7	Devedjiev *et al.* (2000)
Porcine pancreatic elastase	240	Xe (1)	2.1	11.8	5.76	4.0	Mueller-Dieckmann *et al.* (2004)

Table 2
Summary of test results.

$\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$			0.25		0.50		0.75		1.00	
Protein	Anomalous scatterer	Bijvoet ratio (%)	n	Error (°)	n	Error (°)	n	Error (°)	n	Error (°)
Glucose isomerase	S (9)	0.68	1.05	53.07	0.55	41.21	0.27	42.80	0.00	57.60
Xylanase	S (5)	0.69	1.09	56.62	0.57	52.66	0.28	56.11	0.00	65.87
Azurin	Cu (1)	1.44	0.61	62.18	0.30	59.79	0.13	63.15	0.00	70.74
Rusticyanin	Cu (1)	2.36	1.42	52.70	0.77	49.99	0.35	49.42	0.00	57.92
Histone methyltransferase SET7/9	Se (12)	7.03	3.47	44.33	1.83	44.66	0.88	41.31	0.00	44.20
KD93	Se (5)	6.60	1.46	39.80	0.59	42.09	0.21	43.58	0.00	49.65
Human acyl protein thioesterase	Br (22)	8.78	1.77	53.22	0.90	52.08	0.43	53.77	0.00	62.82
Porcine pancreatic elastase	Xe (1)	5.76	3.23	48.01	1.70	45.67	0.83	44.31	0.00	52.34

† Bijvoet ratio = $\langle\lvert F^+ - F^-\rvert\rangle/\langle(F^+ + F^-)/2\rangle$.

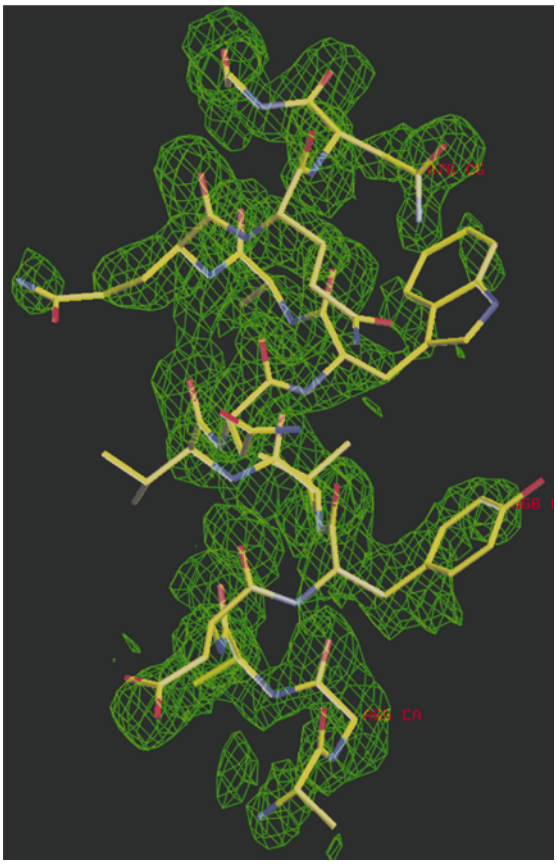

Figure 2
A portion of the electron-density map of the protein xylanase obtained by a default run of the modified *OASIS* followed by a default run of *DM*.

$\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle = 1$, *i.e.* $n = 0$. Furthermore, in seven of the eight cases the condition $n = 0$ corresponds to the largest overall average phase error. This means that the error term $\exp(-\sigma_{\mathbf{h}}^2/2)$ is inevitable in the direct-method SAD phasing in order to obtain more accurate phases.

(ii) The value n corresponding to the lowest overall average phase error changes from one sample to the other. This means that the value n should be fine-tuned in each case. In practice, this would be better performed automatically.

(iii) On average the condition $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle \simeq 0.5$ led to the best phasing result. This can be seen more clearly in Fig. 1, where the overall average phase error is plotted against $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$ for each sample protein.

4. Modification to the program *OASIS*

A simple algorithm has been added to the program *OASIS* that ensures that the condition $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle = 0.5$ is satisfied for various types of input SAD data by automatically tuning the scaling factor n. A keyword is also provided so that the user can set $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle$ equal to any value in the range 0–1. SAD data for all the eight sample proteins were successfully phased by a default run of the modified *OASIS* followed by a default run of *DM*. As pointed out by Ramagopal *et al.* (2003), SAD phasing with a default run of the program *OASIS* (Hao *et al.*, 2000) failed to solve the structure of xylanase. However in the present test, an easily traceable electron-density map can be obtained based on the SAD phasing with the modified *OASIS* (see Fig. 2).

5. Discussion

The weighting function $\exp(-\sigma_{\mathbf{h}}^2/2)$ was introduced by expressing the bimodal SAD phase distribution with the sum of two Gaussian functions taking into account the lack-of-closure error (see Fan, Han & Qian 1984). Introducing the lack-of-closure error into direct methods is important both in theory and in practice of the direct-method SAD phasing. As is seen from the test cases listed in Table 2, $\exp(-\sigma_{\mathbf{h}}^2/2)$ is essential to a successful direct-method SAD phasing. For the calculation of $\sigma_{\mathbf{h}}$ we need the experimentally measured standard deviation of Bijvoet differences. However, when we wrote the previous version of *OASIS*, protein diffraction data sets were not always provided with such measurements or they were not accurate enough. Hence, we used (10) instead of (9) in the program. Now with the technical advances in data collection and treatment we can replace (10) with (11), where the experimental standard deviation $\sigma_{\Delta F_{\mathbf{h}}}$ is included. The constant n preceding $\sigma_{\Delta F_{\mathbf{h}}}$ is for the compensation of some uncertain systematic errors. In Table 2, it is seen that in order to obtain the best phases, n should be set to different value for different sample. Tuning the value of n is thus inevitable in each particular phasing process. It is found from Fig. 1 that on average the condition $\langle\exp(-\sigma_{\mathbf{h}}^2/2)\rangle \simeq 0.5$ leads to the best phasing result. The phasing procedure has been made more efficient by automatically tuning the scaling factor n to satisfy the condition. The method proposed in this paper is also useful in dealing with single-isomorphous replacement (SIR) data.

FHF should like to thank Dr Z. Dauter, Dr M. S. Weiss, Dr S. J. Gamblin, Dr B. Xiao and Dr S. S. Hasnain for kindly providing experimental SAD data. This work is supported by the Innovation Project of the Chinese Academy of Sciences and the 973 Project (Grant Nos. G1999075604 and 2002CB713801) of the Ministry of Science and Technology of China.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 177. London: Academic Press.
Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F., Jiang, T., Liang, D. C. & Fan, H. F. (2004). In the press.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. Z. & Derewenda, Z. S. (2000). *Structure*, **8**, 1137–1146.
Dodd, F., Hasnain, S. S., Abraham, Z. H., Eady, R. R. & Smith, B. E. (1995). *Acta Cryst.* D**51**, 1052–1064.
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., Han, F. S., Qian, J. Z. & Yao, J. X. (1984). *Acta Cryst.* A**40**, 489–495.
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.
Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J. & Xiao, B. (2002). *Cell*, **111**, 105–115.
Woolfson, M. M. & Fan, H. F. (1995*a*). *Physical and Non-Physical Methods of Solving Crystal Structures*, pp. 157–163. Cambridge University Press.
Woolfson, M. M. & Fan, H. F. (1995*b*). *Physical and Non-Physical Methods of Solving Crystal Structures*, pp. 177–179. Cambridge University Press.