Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Improving experimental phasing: the role of strongest reflections

Yanina Vekhter

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375-5000, USA

Correspondence e-mail: vekhter@ccs.nrl.navy.mil

Received 20 October 2004
Accepted 17 March 2005

Very strong reflections have a dominant impact on the initial phasing and model-building stages of structure determination. However, experimental phasing (MIR, SAD or MAD) fails on some of the strongest reflections when the heavy-atom contribution to scattering is relatively weak or absent. It is shown that when just a few (∼50–100) of these reflections are assigned low-error phases, the entire set of isomorphous replacement phases becomes significantly improved after density modification. This improvement is indicated by higher map correlation coefficients and reduced mean phase errors of the updated data. The problem of phasing the strongest reflections may be solved by the direct measurement of triplet phases in a three-beam diffraction experiment. The analysis shows that merging isomorphous replacement data with a limited number of highly accurate phases from the reference-beam diffraction experiment would significantly improve conventional experimental phasing.

1. Introduction

Isomorphous replacement phasing relies on scattering by heavy atoms that occur naturally in the protein sample or have been incorporated into it. In a conventional experimental phasing procedure, once a model of the substructure of heavy atoms has been obtained, phasing information is extracted either from the registered changes in structure-factor amplitudes for the native and derivative crystals or from the Bijvoet differences. A statistical treatment of the data is required owing to errors in the heavy-atom model, errors in the measurements of structure factors and lack of isomorphism. Following Blow & Crick (1959), initial estimates for protein phases are assigned from the phase probability curves based on the assumption of a Gaussian distribution of errors. The probability measure of a phase $P(\alpha_P)$ is evaluated as

$$P(\alpha_P) \propto \exp(-\varepsilon^2/2E^2), \qquad (1)$$

where ε is the combined error, called the lack-of-closure error, defined as the discrepancy between the measured and calculated structure-factor amplitudes and E is the mean-square error, the mean-square value of the 'lack-of-closure' residual. When information from various sources is combined, the resultant probability for a phase $P(\alpha)$ is found as the product of individual probabilities. Modern maximum-likelihood-based phasing permits more accurate phase probability-based estimates to be obtained by directly taking into account the lack of isomorphism and measurement errors (de La Fortelle & Bricogne, 1997; Read, 2003).

Table 1
Test structures and data statistics.

Test structure/ PDB code	Space group	Unit-cell parameters				Protein size (kDa)	Wavelength (Å)	Highest resolution native/derivative data (Å)	Unique reflections	'Heavy' atoms, phasing experiment, references
		a (Å)	*b* (Å)	*c* (Å)	*β* (°)					
Lysozyme, 1lz8	$P4_32_12$	71.81	71.81	36.80		13.3	1.54	1.53	17964	10 S + 7 Cl, SAD; Dauter *et al.* (2002)
2Zn insulin, 4ins	*R*3	80.92	80.92	33.50	120	10.8	0.93	1.0	46849	1 Zn, SAD; Dauter *et al.* (2002)
Ca subtilisin, 1svn	$P2_12_12_1$	75.06	47.37	60.94		25.0	1.54	1.75	22601	3 Ca, SAD; Betzel *et al.* (1988)
Porcine elastase, 1lvy	$P2_12_12_1$	51.64	57.97	75.32		26.95	0.863	1.87	17598	1 Kr, SIRAS; Schiltz *et al.* (1997)
Actino-xanthin, 1acx	$P2_12_12_1$	30.9	48.8	64.1		11.8	1.54	2.0/2.8	2589	1 U, SIR; Pletnev *et al.* (1982)

The phase probabilities are in turn used in the so-called 'best' electron-density maps that are computed with the 'best' phases that correspond to the centroid of the phase probability distribution and with weighted Fourier coefficients $F_{\rm best}(H) = mF_P$. The weight m is a 'figure of merit' (FOM) for a reflection. It depends on the shape of the phase probability curve and is a measure of reliability of phase determination. To reach the model-building and refinement stage (Murshudov *et al.*, 1997; Lamzin *et al.*, 2001), the 'best' maps are further improved by a variety of density-modification procedures (Wang, 1985; Leslie, 1987; Terwilliger, 2000; Lunin, 1988; Zhang & Main, 1990; Rossmann & Blow, 1963) that have been programmed in *PHASES* (Furey & Swaminathan, 1997), *DM* (Cowtan & Zhang, 1999), *RESOLVE* (Terwilliger, 2002) and *CNS* (Brünger *et al.*, 1998). The major contribution to the 'best' mean phase error arises from two distinctive groups of reflections with low or zero FOMs.

(i) Non-centric reflections with bimodal phase probability curves having two maxima further apart than 90°. For these reflections, the registered changes between the 'best' and the 'most probable' phases may be ∼45–90°.

(ii) Both centric and non-centric reflections with high 'lack of closure'. Such centric rellections have almost equal probabilities for the two restricted phases, while non-centric reflections have unimodal and flat phase probability curves. The width of a phase probability curve reflects the combined effect of the errors in the heavy-atom model, weak contribution from heavy atoms and measurement errors or the lack of isomorphism and indicates an error in the phase determination.

Very strong measured reflections may belong to one of these two groups. The reflections are assigned high error phases when the signal from heavy atoms is relatively weak or when there is no contribution from heavy atoms to scattering (as with centric reflections in a SAD experiment). So in the end these strong reflections do not contribute to the electron-density maps and some reflections retain high phase error even after density modification.

This study addresses the following questions.

(i) What impact on phasing would the strongest reflections have if they were assigned error-free or low-error phases?

(ii) How many accurately phased strongest reflections would be enough to obtain an improvement over conventional phasing procedures?

(iii) How can we determine unknown phases?

PDB code 'BEST' phases, FOMs with program	Reflection number at cutoff resolution (Å)	'BEST' phases + high *F*, low FOM model phases	'BEST'map correlation coefficient CC	'BEST' map correlation coefficient CC_{DM}	'BEST' + MODEL map correlation coefficient CC	'BEST' + MODEL map correlation coefficient CC_{DM}	'BEST' mean phase error $\langle\Delta\Phi\rangle$	'BEST' mean phase error $\langle\Delta\Phi\rangle_{DM}$	'BEST'+ MODEL mean phase error $\langle\Delta\Phi\rangle$	'BEST'+ MODEL mean phase error $\langle\Delta\Phi\rangle_{DM}$
1lz8 *PHASES*	6161 2.0	+34 model phases (*F* > 150, FOM < 0.2)	0.389	0.497	0.44	0.535	63.83	57.17	63.83	55.63
		+95 model phases (*F* > 120, FOM < 0.2)			0.49	0.575			62.59	53.8
4ins *SOLVE*	5514 2.0	+58 model phases (*F* > 1000, FOM < 0.05)	0.258	0.293	0.335	0.395	74.83	72.66	73.94	69.3
		+113 model phases (*F* > 900, FOM < 0.05)			0.395	0.499			73.09	65.2
1svn *SOLVE*	15157 2.0	+71 model phases (*F* > 700, FOM < 0.05)	0.286	0.453	0.38	0.513	68.62	62.11	68.16	59.8
		+131 model phases (*F* > 600, FOM < 0.05)			0.4	0.558			67.84	57.5
1lvy *PHASES*	4256 3.0	+48 model phases (*F* > 1000, FOM < 0.6)	0.362	0.44	0.4	0.497	59.8	57.4	59.2	51.6
		+94 model phases (*F* > 900, FOM < 0.6)			0.434	0.54			58.5	48.8
1acx *PHASES*	2589 2.8	+35 model phases (*F* > 500, FOM < 0.6)	0.378	0.44	0.416	0.54	69.7	66.4	68.5	60.6
		+91 model phases (*F* > 400, FOM < 0.6) added			0.46	0.67			66.68	55.97

Figure 1
Density modification with 'BEST'+ MODEL and with 'BEST' phases

2. Tests and results

Standard isomorphous replacement phasing was performed on the test SIR, SIRAS and SAD structures presented in Table 1. 'BEST' phases and FOMs were obtained using the programs *PHASES* and *SOLVE*. Density modification was performed with *DM*. Mean phase errors ($\langle\Delta\Phi\rangle$) and map correlation coefficients (CC) were evaluated before and after density modification and are presented in Fig. 1 as $\langle\Delta\Phi\rangle_{\text{'BEST'}}$, $\langle\Delta\Phi\rangle_{DM\_\text{'BEST'}}$, $\text{CC}_{\text{'BEST'}}$ and $\text{CC}_{DM\_\text{'BEST'}}$. This was compared with the phasing on the improved data, which is referred to as 'BEST' + MODEL. The objective was to select a limited number of say 50, 100 *etc*. strongest reflections with the lowest FOMs. Firstly, the cutoffs for the FOMs were applied and all reflections with the lowest FOMs were picked. Next, the cutoffs for the structure-factor magnitudes were determined so that about 50 or 100 strongest reflections with the lowest FOMs were singled out. Selected reflections have a very high mean phase error of ~80–90° that was reduced after density modification to 50–77°. Some of the phases, however, were not improved after density modification. In the tests, selected reflections were assigned error-free ('MODEL') phases and FOMs equal to one and combined with the rest of the data. Cutoffs applied to the structure-factor magnitudes F and to FOMs are listed in Fig. 1. Density modification was performed on the 'BEST' + MODEL data. Mean phase errors and map correlation coefficients before and after density modification are listed in Fig. 1 as $\langle\Delta\Phi\rangle_{\text{'BEST'+MODEL}}$, $\langle\Delta\Phi\rangle_{\text{DM\_'BEST'+MODEL}}$, $\text{CC}_{\text{'BEST'+MODEL}}$ and $\text{CC}_{\text{DM\_'BEST'+MODEL}}$. In Fig. 2 mean phase errors and map correlation coefficients are shown for the 'BEST' (B), 'BEST' + MODEL (BM), 'BEST' after density modification (B_DM) and 'BEST' + MODEL after density modification (BM_DM) cases. Color-coded areas in Figs. 1 and 2 illustrate the performance of the 'BEST' (standard procedure) and 'BEST' + MODEL data. The result of adding phased strongest reflections to the 'BEST' maps is quite striking. As shown in Fig. 1, with ~50 phased strongest reflections added to a set of 'BEST' phases, the map correlation coefficients improve for the 'BEST' + MODEL maps by 0.05–0.1. This is approximately the effect that conventional density modification has on the 'BEST' maps. However, the maps already improved by the addition of the MODEL phases become much better after density modification. For example, with ~100 error-free phases, after density modification the 'BEST' + MODEL map correlation coefficients are improved by 0.08–0.2 and mean phase errors by 5–10° compared with the corresponding 'BEST' maps and phases (Figs. 1 and 2).

Next, the effect of the strongest reflection errors on phasing was investigated. Random errors of 0, 20, 30, 40 and 50 were imposed on about 100 strongest reflections. The rest of the reflections were assigned the 'BEST' phases and FOMs. Density modification was applied to each of the five sets. As illustrated in Fig. 3 for the example of 1acx and 1svn test structures, with a mean phase error for the strongest reflections of ~20–30° the maps and phases after solvent flattening are still significantly better compared with conventional phasing.

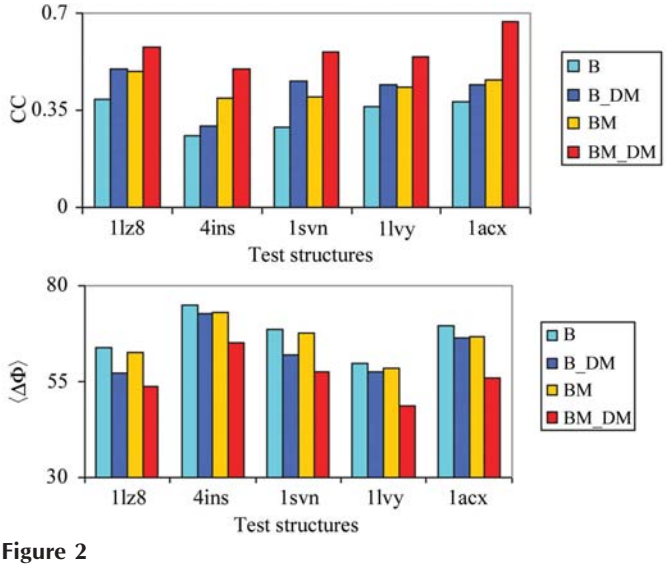

Figure 2
Map correlation coefficients (CC) and mean phase errors ($\langle\Delta\Phi\rangle$) with 'BEST' (B), 'BEST' after density modification (B_DM), ' BEST' + MODEL (BM) and 'BEST' + MODEL after density modification (BM_DM) phases.

3. Conclusions

This study was initiated by an observation that SAD, MIR and MAD maps were improved when a very small number of phased strongest reflections was combined with isomorphous replacement data. The problem is that those very strong reflections are frequently poorly phased when the signal from heavy atoms is relatively weak or absent. The focus of this

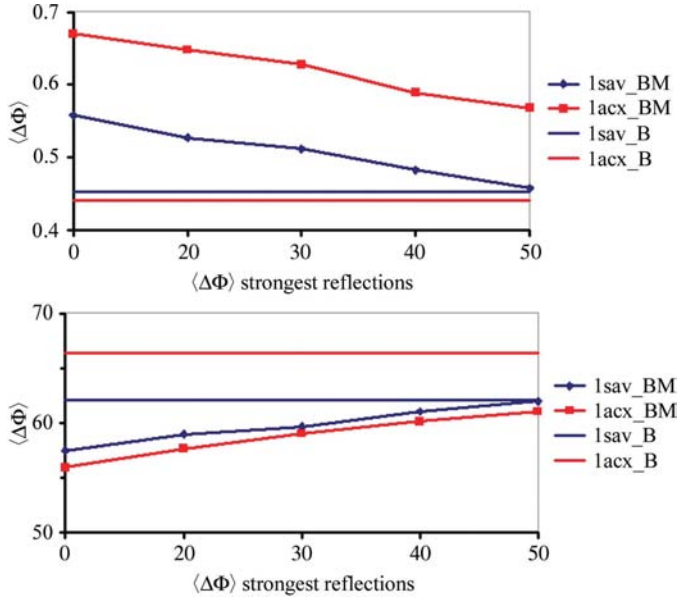

Figure 3
Map correlation coefficients (CC) and mean phase errors ($\langle\Delta\Phi\rangle$) with 'BEST' + MODEL (BM) phases as a function of error level of the strongest reflections for test structures 1acx and 1sav. The corresponding CC and $\langle\Delta\Phi\rangle$ for the 'BEST' (B) phases are shown for comparison as straight lines.

study was to investigate how the strongest reflections affect phasing and how many phased strongest reflections would be enough to obtain a significant improvement over standard experimental phasing. The analysis shows that it is sufficient to assign the phases with mean phase error of 20–30° for about 100 strongest reflections. This is only about 0.05–3% of a typical protein data set at 2–3 Å resolution and it is remarkable that a small percentage of data has such a profound effect on phasing. The problem remains concerning how unknown phases may be determined. This problem may be solved with the direct measurement of triplet phases in a reference-beam diffraction experiment (Chang *et al.*, 1991; Weckert & Hümmer, 1997; Shen, 1998*a*,*b*). In a three-beam diffraction experiment, during a ψ-scan around a strong reflection **h**, another strong reflection with the reciprocal-lattice plane **g** may be excited simultaneously. The primary wave **h** with the phase $\varphi(\mathbf{h})$ interferes with the wave that is diffracted twice by the reciprocal-lattice planes **g** and $\mathbf{h} - \mathbf{g}$ and has phase $\varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g})$. The interference between the two waves depends on the difference between their phases and the intensity of reflection **h** changes owing to the interference. Therefore, the triplet phase $\Phi = \varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) - \varphi(\mathbf{h})$ can be experimentally determined. The triplet measurements are most reliable for very strong reflections. This is where isomorphous replacement phasing usually fails. Therefore, triplet measurements may be used to provide valuable information complementary to the isomorphous replacement data. Very strong reflections are usually involved in many triplets. To determine an unknown phase, all measured triplets involving that unknown phase should be considered. At this point, the other two remaining phases involved with the unknown phase in any of the triplets should be already accurately phased in the SAD or MAD experiments as indicated by their high FOMs. The unknown phase is then determined algebraically by minimizing the discrepancy between the measured and calculated triplet phases.

The group of phased reflections is then expanded and additional triplets are considered so that more reflections may be phased. Preliminary analysis of the measured triplet data shows that unknown phases may be determined with $\sim$30° mean phase error. That would be sufficient to obtain an improvement over conventional phasing. In summary, an improved phasing procedure is proposed that combines the data from three-beam diffraction and isomorphous replacement experiments. This procedure would be of use, in particular, for phasing SAD data. The applications of this procedure will be described in forthcoming papers.

I wish to thank Drs Jerome Karle and John Konnert for many stimulating discussions and comments on phasing the strongest reflections.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chang, S. L., King, H. E. Jr, Huang, M. T. & Gao, Y. (1991). *Phys. Rev. Lett.* **67**, 3113–3116.
Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.
Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 720–722. Dordrecht: Kluwer Academic Publishers.
Leslie, A. G. W. (1987). *Proceedings of the CCP4 Daresbury Study Weekend.* Warrington: Daresbury Laboratory.
Lunin, V. Y. (1988). *Acta Cryst.* A**44**, 144–150.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Pletnev, V., Kuzin, A., Trakhanov, S. & Popovich, V.(1982). *Chemistry of Peptides and Proteins*, edited by W. Voelter, E. Wuensch, J. Ovchinnikov & V. Ivanov, Vol. 1, pp. 429–433. Berlin: Walter de Gruyter.
Read, R. J. (2003). *Acta Cryst.* D**59**, 1891–1902.
Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.
Schiltz, M., Shepard, W., Fourme, R., Prangé, T., de La Fortelle, E. & Bricogne, G. (1997). *Acta Cryst.* D**53**, 78–92.
Shen, Q. (1998*a*). *Phys. Rev. Lett.* **80**, 3268–3271.
Shen, Q. (1998*b*). *Science*, **280**, 828.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1937–1940.
Wang, B.-C. (1985) *Methods Enzymol.* **115**, 90–112.
Weckert, E. & Hümmer, K. (1997). *Acta Cryst.* A**53**, 108–143.
Zhang, K. Y. D. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.