# Direct-method-aided phasing of MIR diffraction data from proteins

**Y. X. Gu,[a] W. R. Chang,[b] T. Jiang,[b] C. D. Zheng[a] and H. F. Fan[a]***

[a]Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, and [b]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China. Correspondence e-mail: fan@aphy.iphy.ac.cn

Direct methods have successfully been used to break the phase ambiguity intrinsic in the single isomorphous replacement (SIR) data of proteins. Based on this, the procedure 'direct-method-aided MIR phasing' (DMIR) has been proposed and applied to the four-derivative multiple isomorphous replacement (MIR) data of a known protein containing 682 amino acid residuals in the asymmetric unit. The data set consists of 14500 unique reflections at 3 Å resolution with $F$(obs.) greater than $2\sigma$. Test calculation showed that the phases from conventional MIR phasing could be significantly improved by direct methods leading to obvious improvement in the quality of the resultant Fourier maps.

## 1. Introduction

Attempts have long been made to break the phase ambiguity intrinsic in protein SIR data. This is important since it is not always easy to prepare suitable multiple isomorphous derivatives of proteins for diffraction data collection. The solvent-flattening method by Wang (1981, 1985) has been a successful procedure in practice. However, the method works only in the case that the replacing heavy atoms are in a noncentrosymmetric arrangement. Different kinds of direct methods have been proposed to break the phase ambiguity (Coulter, 1965; Fan, 1965; Karle, 1966; Hendrickson, 1971; Hauptman, 1982; Fan & Gu, 1985; Fortier et al., 1985; Klop et al., 1987; Giacovazzo et al., 1988; Kyriakidis et al., 1993). Tests with experimental protein data showed that the combination of the direct method (Fan & Gu, 1985) and the solvent-flattening procedure (Wang, 1981, 1985) leads to results better than that from solvent flattening alone (Zheng et al., 1997; Liu et al., 1999). Based on this technique, we propose here the direct-method-aided MIR phasing (DMIR) procedure.

## 2. Direct-method-aided MIR phasing

A set of MIR data consists of several sets of SIR data, which correspond to different pairs of isomorphous derivatives. The conventional phasing of MIR data is to combine the bimodal phase distributions of different SIR data sets to give a unique phase indication for individual reflections. Since the contribution of replacing heavy atoms to the structure factors need not always be strong, the indication of phase for a considerable number of reflections may be rather weak. Besides, imperfect isomorphism increases the phase error of reflections in the high-angle region. Hence it would be better to have

some ways to improve MIR phases before using them to calculate a traceable Fourier map. Direct methods provide independent phase information without requiring additional experimental data. It can be used to improve phases from conventional MIR phasing. The DMIR (direct-method-aided MIR phasing) procedure is designed for this purpose. The technique is based on both the conventional MIR phasing and the direct-method treatment of SIR data. For details of the latter, the reader is referred to the original papers (Fan & Gu, 1985; Zheng et al., 1997; Liu et al., 1999). The points of the present method are as follows.

(i) A few hundreds of reliable MIR phases are selected as the starting phases for direct-method SIR phasing. For conventional MIR phasing, it is not difficult to pick up some hundreds of reliable phases among thousands of reflections. On the other hand, these starting phases can greatly enhance the phasing power of direct methods.

(ii) Direct-method phasing is first applied separately to each set of SIR data. The phasing result will be less affected by the imperfect isomorphism in comparison with MIR phasing.

(iii) Sources of error in phases resulting from each separate direct-method SIR phasing and from MIR phasing are rather different. On the other hand, either direct-method SIR phasing or MIR phasing is followed by the same density-modification treatment. Resultant phases from each SIR data set and the MIR data set have comparable figures of merit. Hence they can be combined [see equations (1) and (2) below] to give a better estimation of phases.

### 2.1. Phasing strategy

The flow chart of the DMIR procedure is shown in Fig. 1. The main points are:

**Table 1**
Summary of the test data from the protein (*R*)-phycoerythrin.

| | |
|---|---|
| Space group | *R*3 |
| Unit cell | *a* = 189.8, *c* = 60.1 Å |
| No. of residues in the a.u. | 682 |
| No. of derivatives | 4 |
| Replacing atoms | Au, Pt, Hg, Hg |
| Resolution limit | 3.0 Å |
| No. of unique reflections | 14500 |

**Table 2**
Comparison of MIR phasing and DMIR phasing *via* phase error [weighted by the product of *F*(obs.) and the corresponding figure of merit] for cumulative reflection groups.

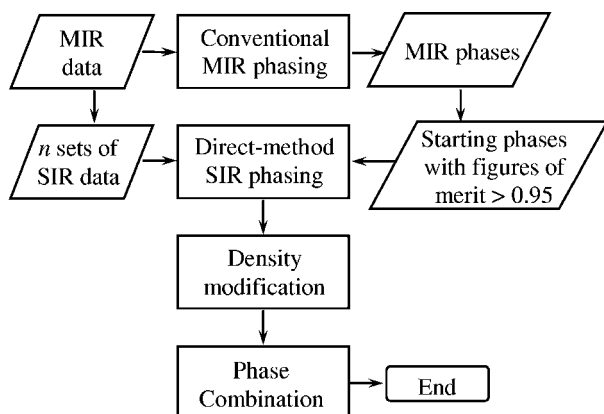| No. of reflections in the cumulative group | Averaged phase error (°) | | |
|---|---|---|---|
| | MIR+*dm* (CCP4) | DMIR+*dm* (CCP4) | Change from MIR to DMIR |
| 500 | 38.57 | 34.79 | −3.78 |
| 1000 | 42.04 | 37.47 | −4.57 |
| 3000 | 48.28 | 43.69 | −4.59 |
| 5000 | 50.32 | 46.72 | −4.05 |
| 7000 | 51.65 | 48.05 | −3.60 |
| 9000 | 52.96 | 49.86 | −3.10 |
| 11000 | 53.94 | 51.13 | −2.81 |
| 13000 | 54.68 | 52.15 | −2.53 |
| 14500 | 55.11 | 52.72 | −2.39 |

51 (i) The conventional MIR phasing is first applied to a set of MIR data.

(ii) MIR phases with figures of merit larger than a certain limit, say 0.95, are used as starting phases in the direct-method phasing.

(iii) The MIR data are divided into *n* sets of SIR data; *n* is the number of isomorphous heavy-atom derivatives.

(iv) Direct methods are used to break the phase ambiguity of each set of SIR data based on the starting phases from the conventional MIR phasing.

(v) Density modification is applied to the direct-method-phased SIR Fourier map.

(vi) Combine the resultant phases after density modification.

(vii) Combine the combined SIR phases with the conventional MIR phases.

The phase combination is performed according to the following formulas:

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

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

where *n* is the number of phase sets involved in the combination. Such a combination can be regarded as a reciprocal-space equivalent of calculating a sum function of Fourier maps corresponding to the *n* sets of phases.
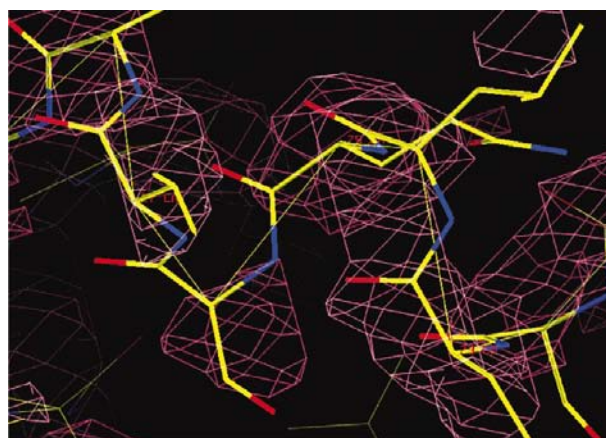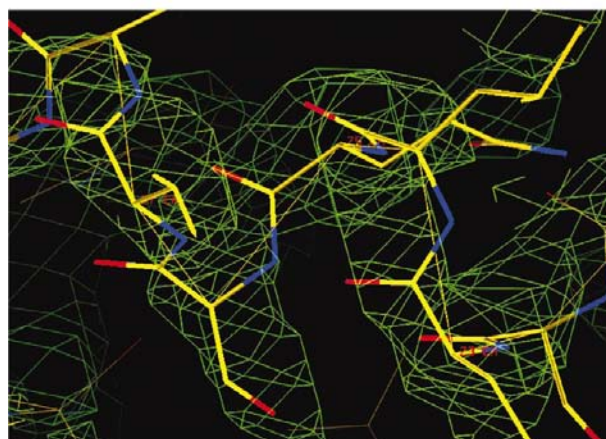




**Figure 2**
A portion of the Fourier map of *R*-PE around residue 25. (*a*) From conventional MIR phasing; (*b*) from DMIR phasing. Contour level = 1.2σ. The final structure model from the PDB file is superimposed.



**Figure 1**
The flow chart of the DMIR procedure.

**Table 3**
Variation of averaged phase error [weighted by the product of $F$(obs.) and the corresponding figure of merit] against $d$ spacing.
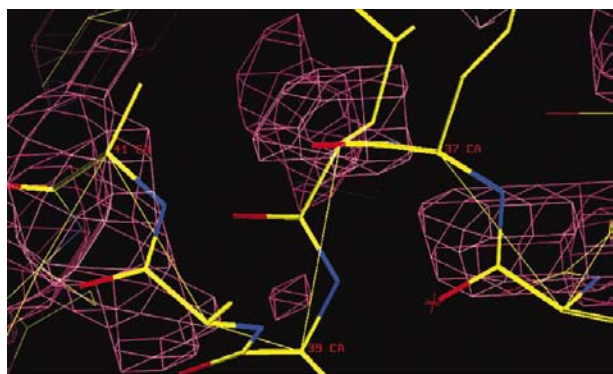
| $d$ spacing (Å) | MIR + $dm$ (CCP4) | | DMIR + $dm$ (CCP4) | | |
| | No. of reflections | Averaged phase error (°) | No. of reflections | Averaged phase error (°) | Change from MIR to DMIR (°) |
| --- | --- | --- | --- | --- | --- |
| 20.00–5.00 | 3374 | 42.17 | 3403 | 42.22 | +0.05 |
| 4.99–4.60 | 985 | 49.55 | 993 | 45.69 | −3.86 |
| 4.59–4.30 | 966 | 46.89 | 979 | 41.83 | −5.06 |
| 4.29–4.00 | 1266 | 56.48 | 1290 | 49.24 | −7.24 |
| 3.99–3.80 | 1080 | 60.15 | 1104 | 53.45 | −6.70 |
| 3.79–3.65 | 949 | 61.86 | 971 | 56.64 | −5.22 |
| 3.64–3.50 | 1088 | 65.96 | 1123 | 62.00 | −3.96 |
| 3.49–3.40 | 823 | 64.40 | 861 | 62.48 | −1.92 |
| 3.39–3.30 | 902 | 63.37 | 953 | 61.78 | −1.59 |
| 3.29–3.20 | 998 | 67.35 | 1069 | 66.66 | −0.69 |
| 3.19–3.10 | 1037 | 67.24 | 1128 | 66.50 | −0.74 |
| 3.09–2.99 | 1106 | 71.99 | 1240 | 72.05 | +0.06 |

In our test, the starting phases for direct-method phasing were selected from the original MIR phases according to their figures of merit. The density modification was performed using the program $dm$ in the CCP4 suite (Collaborative Computational Project, Number 4, 1994). Resultant phases were cumulated in descending order of $F$(obs.). Average phase errors [weighted by the product of $F$(obs.) and the corresponding figure of merit] of the cumulative groups are listed in Table 2. As is seen, MIR phases are clearly improved by the DMIR procedure especially for the top 5000 strongest reflections, which is about one third of the total reflections. This ensures significant improvement in the quality of the resultant Fourier maps. Table 3 shows the variation of the averaged phase error against $d$ spacing (resolution). Significant improvement of MIR phases through DMIR phasing is observed at the resolution range of 5.0 to 3.5 Å. This led to obvious improvement in the connectivity of the resultant electron-density map. Comparison of the Fourier map resulting from MIR phasing and that from DMIR phasing
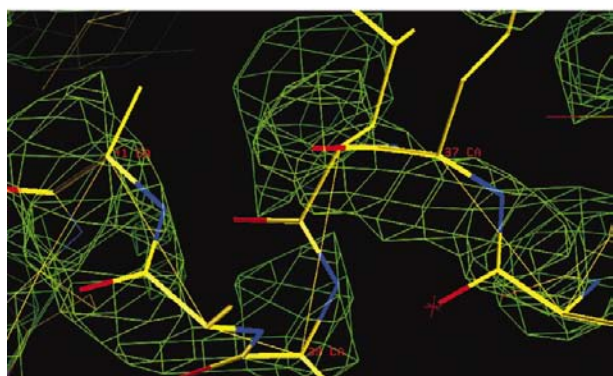
## 2.2. Data and test results

The known protein ($R$)-phycoerythrin ($R$-PE) (Chang *et al.*, 1996) was used for testing the DMIR procedure. The structure was originally solved using programs in the CCP4 suite (Collaborative Computational Project, Number 4, 1994). A summary of the data is shown in Table 1. The abbreviation a.u. in the table stands for asymmetric unit.



**Figure 3**
A portion of the Fourier map of $R$-PE around residue 38. (*a*) From conventional MIR phasing; (*b*) from DMIR phasing. Contour level = 1.2$\sigma$. The final structure model from the PDB file is superimposed.
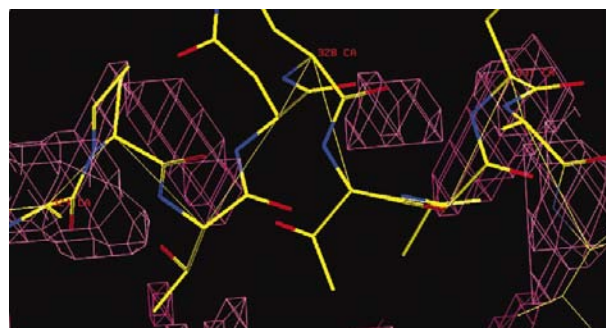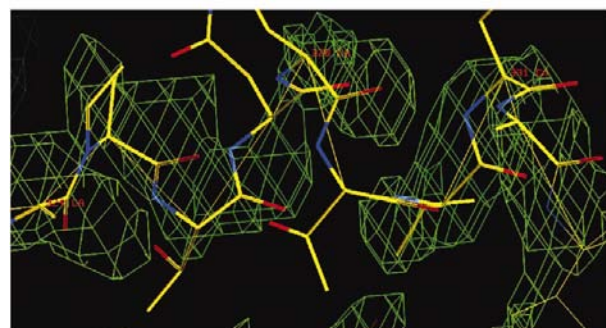


**Figure 4**
A portion of the Fourier map of $R$-PE around residue 328. (*a*) From conventional MIR phasing; (*b*) from DMIR phasing. Contour level = 1.2$\sigma$. The final structure model from the PDB file is superimposed.

# research papers

shows that within the asymmetric unit there are more than 20 places where the DMIR map is significantly better than the MIR map. Four examples are shown in Figs. 2 to 5, respectively. In contrast, there are only about 5 places in the asymmetric unit where the MIR map is better. An example is shown in Fig. 6.

## 3. Concluding remarks

The direct-method SIR phasing used in this paper is different from any kind of existing MIR phasing. It is also different from various kinds of density modification. Hence, direct-method SIR phasing can provide independent phase indications and there exists the opportunity by incorporating the direct method to improve a set of MIR phases obtained by other methods. The present test proved that it is true at least for one example. The test was done with a known protein of moderate size, which is the biggest one so far treated by direct methods at a resolution as low as 3 Å. The result showed that the DMIR procedure could significantly improve the MIR phases of the test sample leading to much better electron densities in the resultant Fourier map. Hopefully, in some cases when the conventional MIR technique failed to solve the structure, the DMIR procedure may be capable of producing a traceable Fourier map.
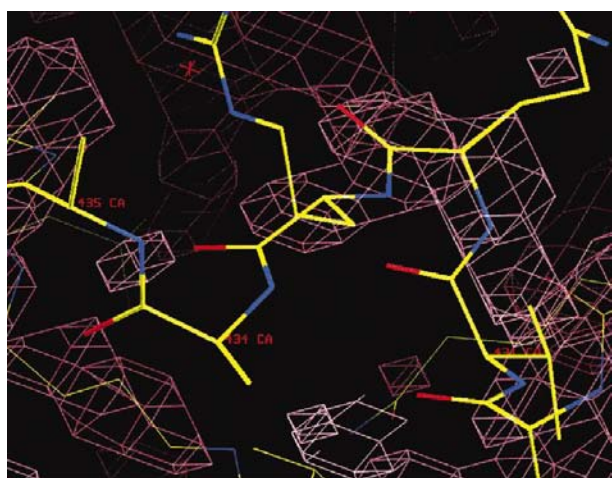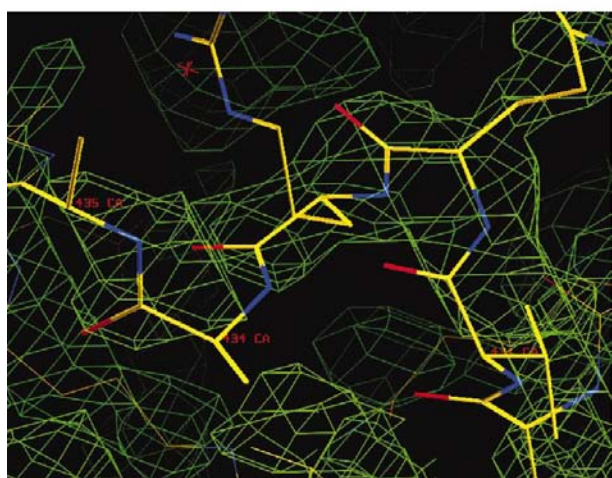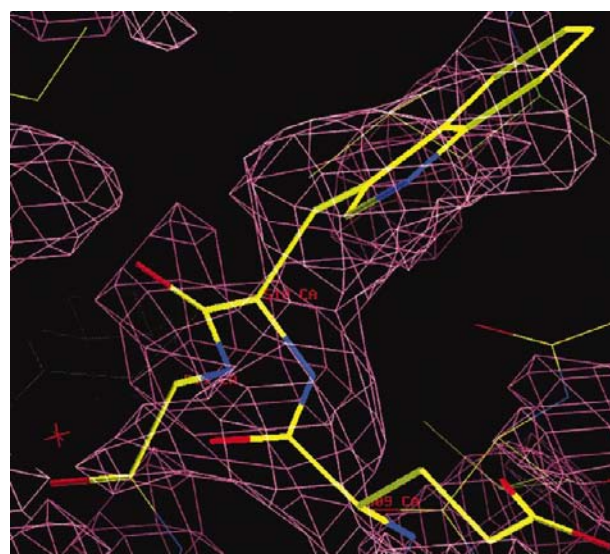
**Figure 5**
A portion of the Fourier map of *R*-PE around residue 432. (*a*) From conventional MIR phasing; (*b*) from DMIR phasing. Contour level = 1.2σ. The final structure model from the PDB file is superimposed.
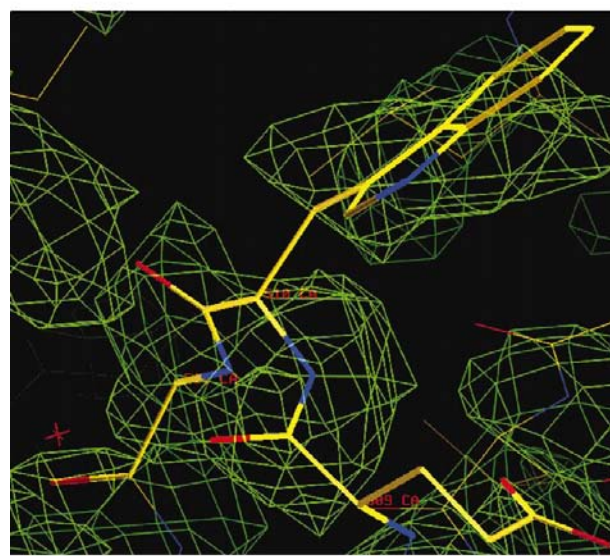


**Figure 6**
A portion of the Fourier map of *R*-PE around residue 512. (*a*) From conventional MIR phasing; (*b*) from DMIR phasing. Contour level = 1.2σ. The final structure model from the PDB file is superimposed.

## References

Chang, W. R., Jiang, T., Wan, Z. L., Zhang, J. P., Yang, Z. & Liang, D. C. (1996). *J. Mol. Biol.* **262**, 721–731.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coulter, C. L. (1965). *J. Mol. Biol.* **12**, 292–295.

Fan, H. F. (1965). *Acta Phys. Sin.* **21**, 1114–1118 (in Chinese); Engl. transl: *Chinese Phys.* pp. 1429–1435.

Fan, H. F. & Gu, Y. X. (1985). *Acta Cryst.* A**41**, 280–284.

Fortier, S. Moore, N. J. & Fraser, M. E. (1985). *Acta Cryst.* A**41**, 571–577.

Giacovazzo, C., Cascarano, G. & Zheng, C. (1988). *Acta Cryst.* A**44**, 45–51.

Hauptman, H. (1982). *Acta Cryst.* A**38**, 289–294.

Hendrickson, W. A. (1971). *Acta Cryst.* B**27**, 1474–1475.

Karle, J. (1966). *Acta Cryst.* **21**, 273–276.

Klop, E. A. Krabbendam, H. & Kroon, J. (1987). *Acta Cryst.* A**43**, 810–820.

Kyriakidis, C. E., Peschar, R. & Schenk, H. (1993). *Acta Cryst.* A**49**, 557–569.

Liu, Y. D., Gu, Y. X., Zheng, C. D., Hao, Q. & Fan, H. F. (1999). *Acta Cryst.* D**55**, 846–848.

Wang, B. C. (1981). *Acta Cryst.* A**37**, C11.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Zheng, X. F., Zheng, C. D., Gu, Y. X., Mo, Y. D., Fan, H. F. & Hao, Q. (1997). *Acta Cryst.* D**53**, 49–55.