*Direct Methods for Solving Macromolecular Structures*
Ed. by S. Fortier,
Kluwer Academic Publishes, The Netherlands, 1998, pp. 479-485.

# APPLICATIONS OF DIRECT METHODS WITH SINGLE ISOMORPHOUS REPLACEMENT OR ONE-WAVELENGTH ANOMALOUS SCATTERING DATA

H. F. Fan [a] , Q. Hao [b, c, a] , I. Harvey [b, c], S. S. Hasnain [c] , Y. D. Liu [a],
Y. X. Gu [a], C. D. Zheng [a] and H. Ke [d],
[a] *Institute of Physics, Chinese Academy of Sciences,*
*Beijing 100080, P.R. China*
[b] *School of Applied Sciences, De Montfort University,*
*Leicester LE1 9BH, England*
[c] *CCLRC Daresbury Laboratory, Warrington WA4 4AD, England*
[d] *Department of Biochemistry & Biophysics, University of North Carolina,*
*Chapel Hill, NC 27599, USA*

## 1. Introduction

Recent developments of our study have been concentrated on the following topics: 1) the accuracy of our Cochran-distribution-based probability formula: it has been shown that the results of the formula are much more accurate than what people would usually predict; 2) an iterative procedure of combining direct methods with the solvent flattening technique: test on the experimental single isomorphous replacement (SIR) data of the known protein RNase Sa showed considerable improvement to the previous procedure; 3) direct-method phasing of the one-wavelength anomalous scattering (OAS) data of an unknown protein: the result leads to a traceable Fourier map and thus to the solution of the structure.

## 2. On the accuracy of the Cochran-distribution-based probability formula

Equation (1) is the probability formula we used to break the SIR or OAS phase ambiguity.

$$P_{+}(\Delta\varphi_h) = \frac{1}{2}+\frac{1}{2}\tanh\left\{\sin(|\Delta\varphi_h|)\left[\sum_{H'} m_{h'}m_{h-h'}\kappa_{h,h'}\sin(\Phi'_3+\Delta\varphi_{h',best}+\Delta\varphi_{h-h',best})+\chi\sin\delta_h\right]\right\} \qquad (1)$$

For details of the formula the reader is referred to reference [1]. This formula is based

on the Cochran distribution [2]

$$P(\phi_3) = \frac{1}{2\pi I_0(\kappa_{h,h'})} \exp(\kappa_{h,h'} \cos\phi_3) \tag{2}$$

For equal-atom structures,

$$\kappa_{h,h'} = 2N^{-1/2} \left| E_{-h} E_{h'} E_{h-h''} \right| \tag{3}$$

where N is the number of atoms in the unit cell. The reliability of phase estimation based on (2) depends on the 'sharpness', and thus the quantity $\kappa_{h,h'}$ of the distribution. For protein structures since N is large, the value of $\kappa_{h,h'}$ will be small. This might affect the accuracy of the phasing results. A comparison of the accuracy of estimating three-phase structure invariants by different probability formulas was reported [3]. Part of the results are shown in columns 1 to 3 of Table 1. As is seen the accuracy of estimation by the Cochran distribution is the lowest. This is true for equation (2). However the situation may be quite different for equation (1). Firstly, in equation (1) large number of three-phase structure invariants can be used jointly to predict a single phase leading usually to a much accurate estimation. Secondly, the task of equation (1) is just to make choice between the two possible signs of $\Delta\varphi$, while that of equation (2) is to predict a value in the range of 0 to $2\pi$, the latter is surely much more complicated and inaccurate. Finally, the figure of merit $m_h$ in equation (1) is very efficient in minimising the error of estimation. For a comparison we used a set of calculated OAS data of the protein aPP at 2Å resolution. The condition is the same as that of reference [3] for aPP. Individual phases were derived by making use of equation (1). Errors of the three-phase structure invariants calculated accordingly were than cumulated as shown in columns 4 and 5 of Table 1.

Table 1. Cumulative statistics of the three-phase structure invariants for the protein aPP

| Cochran 250 largest E's | | PS (true) 250 largest E's | | SDs (true) 250 largest E's | | Equation (1) 250 largest E's | | Equation (1) total 2108 E's | |
|---|---|---|---|---|---|---|---|---|---|
| NTR | ERR | NTR | ERR | NTR | ERR | NTR | ERR | NTR | ERR |
| 315 | 140 | 255 | 52 | 197 | 11 | 600 | 21 | 80,000 | 37 |
| 906 | 141 | 1021 | 57 | 1049 | 12 | 1000 | 25 | 400,000 | 55 |
| 3750 | 153 | 3750 | 98 | 3750 | 52 | 2288 (total) | 32 | 807,262 | 71 |

NTR — number of triplet relations (three-phase structure invariants);
ERR — average error of three-phase structure invariants in mc (1000 mc = $2\pi$ rad)
ERR = $<|\phi_{3\,true} - \phi_{3\,est}|>$

We concluded that the accuracy of estimation by equation (1) is far better than what

people would usually predict. This explains why equation (1) has been so successful in dealing with the experimental OAS and SIR data of protein crystals.

## 3. An iterative procedure of combining direct methods with the solvent flattening technique

The major experimental error in OAS or SIR data comes from the measurement of the Bijvoet differences or the structure-factor-magnitude differences between a pair of isomorphous crystals. A considerable number of the observed differences will be incorrect not only in their magnitudes but also in their signs, which can in some circumstances have a very large effect on the estimate of the phase angle. When direct methods are used in such a case, large errors, usually not randomly distributed, will then be transferred to the solvent flattening and affect significantly the result. A new procedure [4] has been proposed for dealing with this difficulty. The procedure differs from the previous one [5] in that the direct method not only provides input phases to, but also accepts feed back phases from solvent flattening, thus forming an iterative process for breaking the ambiguities and refining the values of phases. The flow-chart is shown in *figure* 1. The new procedure was tested with the experimental SIR data of the known structure ribonuclease (RNase) Sa. For the strongest 1000 of the total 7264 reflections, the mean Fobs-weighted phase error is respectively 7.5 and 9.4 degrees lower than that of the previous procedure and that of the solvent flattening alone. Some portions of the resultant Fourier map are shown in *figure* 2 in comparison with that obtained from the previous procedure.

## 4. Solving an unknown protein by direct-method phasing of the OAS data

Rusticyanin is a copper-containing electron transfer protein with a very high redox potential. No similar structures are available to enable a molecular replacement solution. The crystallisation of the protein was carried out at Daresbury Laboratory. The crystals belong to the space group $P2_1$ with unit cell dimensions: $a$=32.43, $b$=60.68, $c$=38.01Å and $\beta$=107.82$^o$. There is one molecule (about 1300 non-H atoms including one Cu atom) in the asymmetric unit. X-ray diffraction data were collected on station 9.5 at the Daresbury Synchrotron Radiation Source. The wavelength of 1.3765Å near the Cu K-edge was chosen to maximise the value of $\Delta f''$ (=3.87). Only one crystal was used in the experiment and it was aligned carefully so that Friedel pairs were collected on the same image plate.

The copper atom was located by the conventional direct-method program *SAPI*91 [6] using the magnitudes of anomalous differences. The direct method [7 - 9] was then used to break the phase ambiguity intrinsic to OAS data. The noise level of the direct-method map was reduced by a minimum function of the direct-method map and a Ps-function map [10]. Further improvement was obtained by using the electron density modification program DM in CCP4 suite. The resultant electron density map (*figure* 3) was clearly traceable and the backbone of the structure was then built by using the modelling program O [11]. This work was carried out at De Montfort University and Daresbury Laboratory (Harvey, Hao, Ingledew and Hasnain, unpublished results).

The same protein was also studied independently at Cornell University using the multiwavelength anomalous diffraction (MAD) method. They have reported the structure in late 1996 [12]. However till now neither atomic positions nor diffraction phases have been released. Besides, our direct-method phases were derived before the Cornell group's publication of the structure. Hence it is fair to say, the structure was solved independently by direct phasing of the OAS data.

**References**


1. Fan, H. F. and Gu, Y. X. (1985) Combining direct methods with isomorphous replacement or anomalous scattering data III. The incorporation of partial structure information, *Acta Cryst.* A**41**, 280-284. See also Fan, H. F. (1997) Direct methods of solving phase ambiguities in macro-molecular structures, in this *Proceedings.*
2. Cochran, W. (1955) Relations between the phase of structure factors, *Acta Cryst.* **8**, 473-478.
3. Kyriakidis, C. E., Peschar, R. and Schenk, H. (1993) On the direct methods phase information from differences between isomorphous structure factors, *Acta Cryst.* A**49**, 557-569.
4. Gu, Y. X., Zheng, C. D., Zhao, Y. D., Ke, H. and Fan, H. F. (1997) A New Procedure of Combining Direct Methods with Solvent Flattening — Dealing with the Phase Ambiguity in Protein Crystallography, *Acta Cryst.* D**53** (submitted).
5. Zheng, X. F., Zheng, C. D., Gu, Y. X., Mo, Y. D., Fan, H.F. and Hao, Q. (1997) Use of single isomorphous replacement data of proteins — resolving the phase ambiguity and a new procedure for phase extension, *Acta Cryst.* D**53**, 49-55.
6. Fan, H. F., Yao, J. X., Zheng, C. D., Gu, Y. X. and Qian, J. Z. (1991) *SAPI* 91: a computer program for automatic solution of crystal structures from X-ray diffraction data, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, P. R. China.
7. Fan, H. F. and Gu Y. X. (1985) Combining direct methods with isomorphous replacement or anomalous scattering data III. The incorporation of partial structure information, *Acta Cryst.* A**41**, 280-284.
8. Fan H. F., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. and Ke, H. (1990) Combining direct methods with isomorphous replacement or anomalous scattering data VII. Ab initio phasing of the OAS data from a small protein, *Acta Cryst.* A**46**, 935-939.
9. Hao, Q., Gu, Y. X., Zheng, C. D. and Fan, H. F. (1996) OASIS: a computer program for breaking the phase ambiguity in OAS or SIR protein data, School of Applied Sciences, De Montfort University, Leicester LE1 9BH, England and Institute of Physics, Chinese Academy of Sciences, Beijing 100080, P. R. China.
10. Hao, Q. and Woolfson, M.M. (1989) Application of the Ps-function method to macromolecular structure determination, *Acta Cryst.* A**45**, 794-797.
11. Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Cryst.* A**47**, 110-119.
12. Walter, R.L., Ealick, S.E., Friedman, A.M., Blake, R.C., Proctor, P. and Stoham, M. (1996) Multiple wavelength anomalous diffraction (MAD) crystal structure of rusticyanin: a highly oxidising cupredoxin with extreme acid stability, *J. Mol. Biol.*, **263**, 730-751.

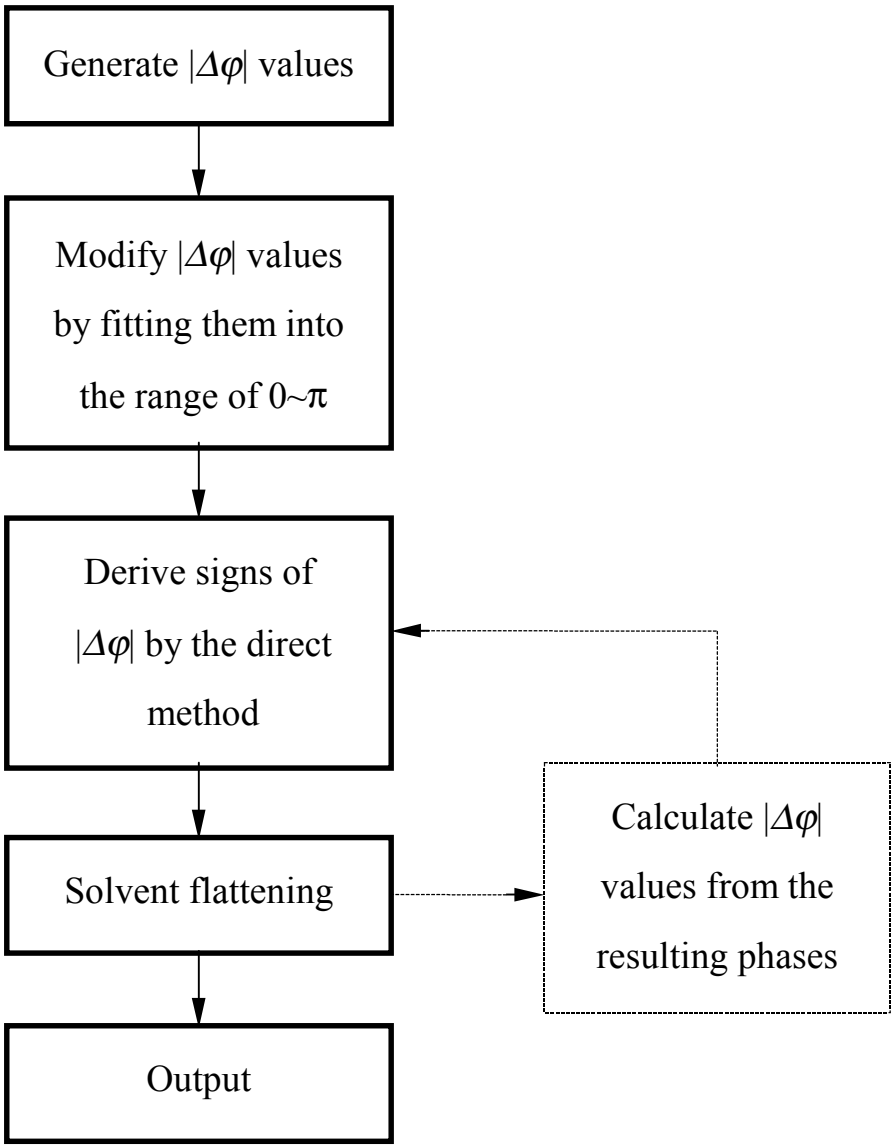


*Figure* 1. Flowchart of the iterative "direct method + solvent flattening". The portion shown with dashed lines is new to the previous simple "direct method + solvent flattening".

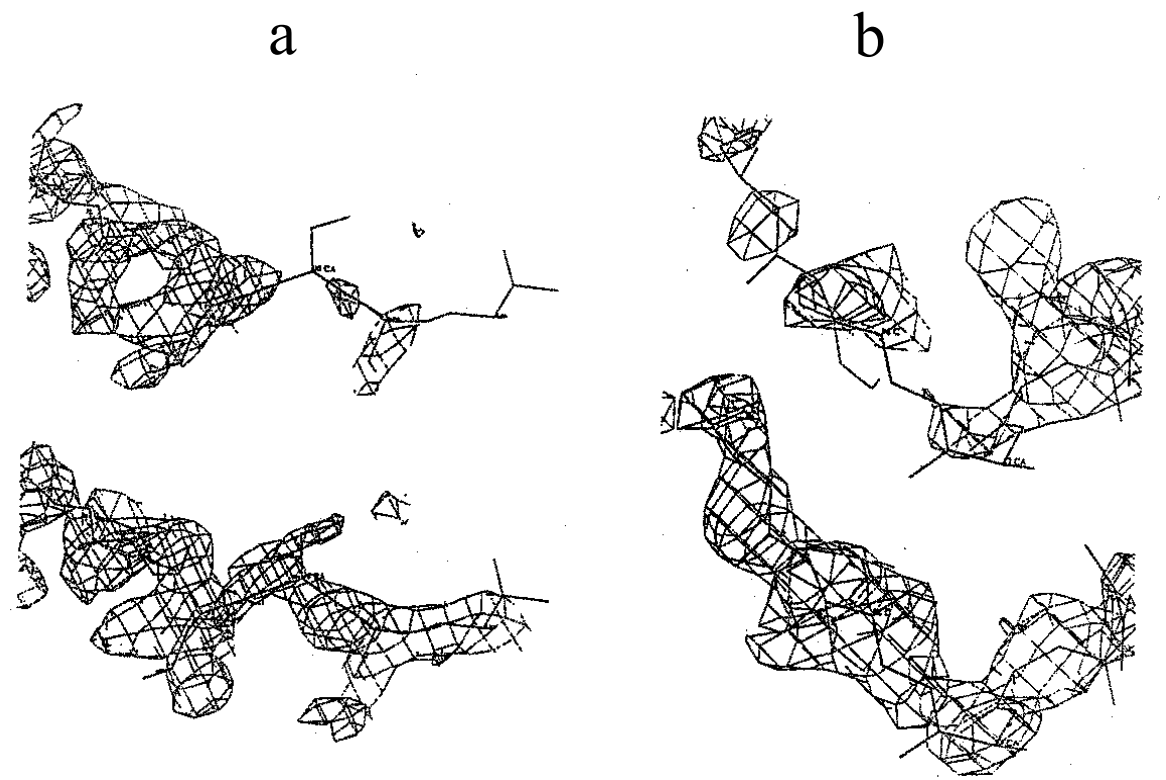


*Figure* 2. Comparison of Fourier maps on two different portions (a) and (b). Maps from the simple "direct method + solvent flattening" are shown on the upper part, while the corresponding maps from the iterative "direct method + solvent flattening" are shown on the lower part. All maps are contoured at 1σ.

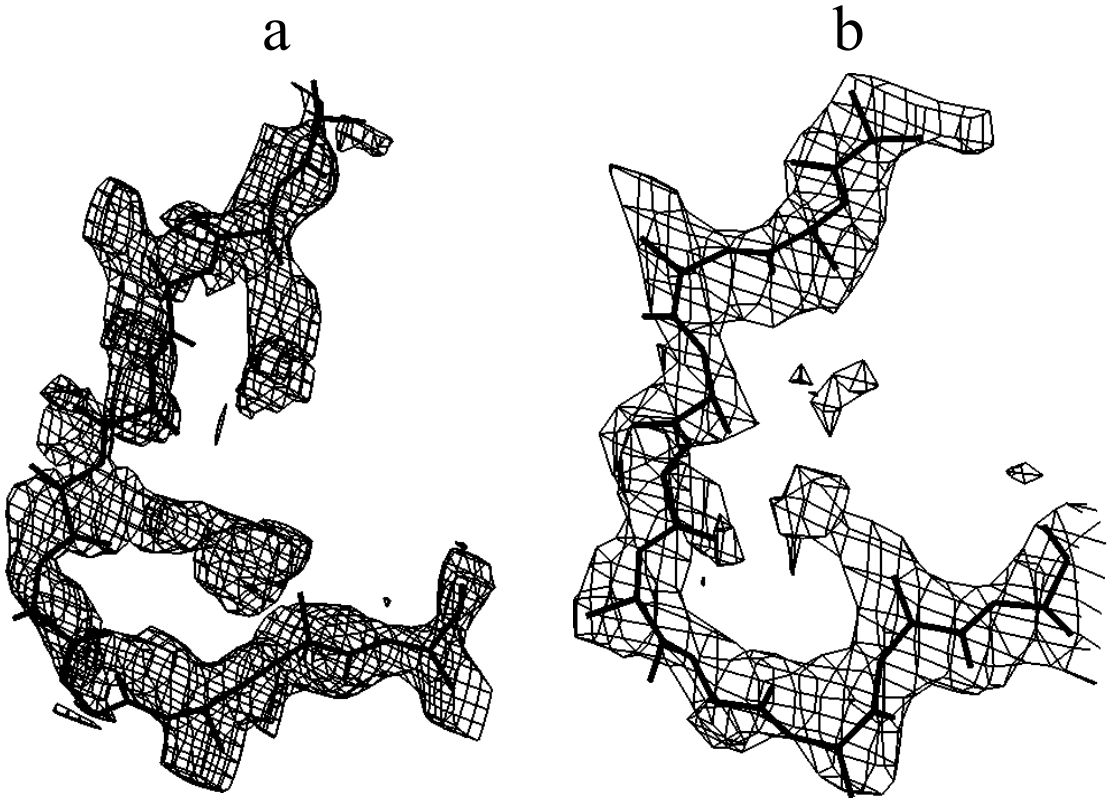


*Figure* 3. A section (residues 126-135) of the rusticyanin electron density map: (a) after a minimum function of direct method and Ps-function maps + density modification; (b) after one cycle of simulated annealing refinement.