# A procedure for *ab initio* phasing of one-wavelength anomalous scattering data from proteins – combining direct methods with the solvent flattening technique

X. F. Zheng, Y. X. Gu, C. D. Zheng, Y. D. Mo, H. F. Fan*

Chinese Academy of Sciences Institute of Physics, Beijing 100080, P.R. China

and Q. Hao

De Montfort University, Department of Applied Physics, Leicester LE1 9BH, England and Chinese Academy of Sciences, Institute of Physics, Beijing 100 080, P.R. China

Received April 9, 1996; accepted August 6, 1996

**Abstract.** Both direct methods and the solvent flattening technique have been proved capable of breaking the phase ambiguity intrinsic in one-wavelength anomalous scattering (OAS) data. However, in theory, while direct methods are better in deriving starting phases from OAS data, the solvent flattening technique is more powerful in the subsequent phase refinement. A procedure is described in this paper which combines direct methods with the solvent flattening technique for phasing of OAS data. The procedure has been tested with the known protein RNase Sa. This protein is of moderate size, and the diffraction data were collected routinely without taking special care for the recording of Friedels. The test proved that the combination of the direct method and the solvent flattening leads to a much better result than that from either method alone.

## Introduction

There have been reported a number of successful tests on direct-method phasing of one-wavelength anomalous scattering (OAS) data from proteins. The OAS data of the Hg-derivative of aPP, a small protein (Blundell, Pitts, Tickle, Wood and Wu, 1981), has been phased by a direct method leading to an interpretable Fourier map (Fan, Hao, Gu, Qian, Zheng and Ke, 1990). The combination of direct method with phase-refinement techniques had been used to derive phases of the OAS data from a moderate-size protein, selenobiotin-binding core of streptavidin (Hendrickson, Pähler, Smith, Satow, Merritt and Phizackerley, 1989), intensities of which were collected with synchrotron radiation at a wavelength close to and slightly shorter than that of the absorption edge of selenium. Special arrangement was made to ensure that all reflections were collected under nearly the same condition as their Friedel partners. The resultant Fourier map allowed backbone tracing (Sha, Liu, Gu, Fan, Ke, Yao and Woolfson 1995). A similar procedure was also successful in phasing the OAS data of the native protein azurin containing copper atoms as anomalous scatterers (Zheng, Fan, Hao, Dodd and Hasnain, 1996). In this case, however, the data were collected with synchrotron radiation at a wavelength far shorter than that of the absorption edge of copper. In the present paper, we describe a further test on phasing OAS data of RNase Sa (Dodson, Sevcik, Dodson and Zelinka, 1987), a protein of medium size. The data were measured with Cu$K_\alpha$ radiation, without special provisions for collecting the Friedel pairs. This is the method commonly used in protein crystallography. The success of the present test proves the flexibility and feasibility of our procedure. In a different context, the solvent flattening technique (Wang, 1981, 1985) has been widely used in protein crystallography mainly for phase refinement and extension, but it has also been succeeded in breaking the OAS phase ambiguity in the determination of bovine neurophysin II dipeptide complex (Chen, Rose, Breslow, Yang, Chang, Furey, Sax and Wang, 1991) and of Cd, Zn metallothionein (Robbins, McRee, Williamson, Collett, Xuong, Furey, Wang and Stout, 1991). In theory, direct methods are better for resolving the phase ambiguity, while the solvent flattening technique is better for subsequent phase refinement. Hence, it is expected that the combination of these two methods will give better results than those obtained by either of them alone. The present work shows that it is true in practice.

## Method

The solvent flattening technique used in our test is that typically applied in protein crystallography (Wang, 1981, 1985). The main point of the technique is to distinguish, on a given Fourier map, the ordered protein region from that of the surrounding disordered solvent. Following that, the densities inside the protein envelope are raised by a

* Correspondence author (e-mail: fan@aphy02.iphy.ac.cn)

constant value and then densities lower than a certain value are removed. Outside the protein region, the density is smoothed to a constant level. By this process some of the major noise in the Fourier map is filtered and a rough structural model is constructed, which can be used to resolve the phase ambiguity. The procedure is made iterative. To apply the solvent flattening technique, it is necessary to assume the solvent content of the crystal. In the present test, the solvent content was assumed to be 50%. When a stable mask is obtained during the iteration, this mask will be applied back to the original Fourier map calculated with averaged OAS phases, and then the iteration will be repeated until it converges to a new stable mask. By doing this, it is often possible to improve the final phases.

A brief description of the direct method used to break the phase ambiguity is given below.

The phase doublets inherent in the OAS method are expressed as

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

Where $\varphi_H$ denotes the phase of structure factors of the protein; $\varphi''_H$ is the phase of

$$\mathbf{F}''_{ano} = \sum_{j=1}^{N_{ano}} i\,\Delta f''_j \,\exp\,(i2\pi\mathbf{H}\cdot\mathbf{r}_j)\,, \tag{2}$$

which can be calculated from the known positions of the anomalous scatterers and the known value of $\Delta f''$; $\Delta\varphi_H$ is the difference between $\varphi_H$ and $\varphi''_H$, its absolute value can be approximated by (Blundell and Johnson, 1976)

$$\cos\,\Delta\varphi_H = (F_H^+ - F_H^-)/2\,|\mathbf{F}''_{ano}|\,. \tag{3}$$

In practice, values of $|\cos\,(\Delta\varphi_H)|$ obtained from equation (3) may exceed 1 due to the error in data measurement and scaling. If there are only a few reflections, say less than 1%, with $\cos\,(\Delta\varphi_H)$ outside the range of $-1$ to $+1$, we can just simply reduce values greater than $+1$ to $+1$ and increase values smaller than $-1$ to $-1$. However, if there are too many reflections with $\cos\,(\Delta\varphi_H)$ outside the range of $-1$ to $+1$, the whole set of $\cos\,(\Delta\varphi_H)$ will not be used directly in subsequent calculations. Instead they are first sorted in descending order and then modified to fit into a uniform distribution between $+1$ and $-1$. New values of $\cos\,(\Delta\varphi_H)$ obtained in this way are then used in the following calculations.

The phase problem in the OAS case is in fact a sign problem according to equation (1). The probability for $\Delta\varphi_H$ to be positive is given by Fan and Gu (1985):

$$P_+(\Delta\varphi_h) = \tfrac{1}{2} + \tfrac{1}{2}\,\tanh\left\{\sin|\Delta\varphi_h|\left[\sum_{H'} m_{H'} m_{H-H'}\kappa_{H,H'}\times\right.\right.$$
$$\times\,\sin\,(\Phi'_3 + \Delta\varphi_{H',best} + \Delta\varphi_{H-H',best}) +$$
$$\left.\left. + \chi\,\sin\,\delta_H\right]\right\}. \tag{4}$$

where

$$m_H = \exp\,(-{\sigma_H}^2/2)\,\{[2(P_+ - \tfrac{1}{2})^2 + \tfrac{1}{2}]\times$$
$$\times\,[1 - \cos\,(2\Delta\varphi_h)] + \cos\,(2\Delta\varphi_H)\}^{1/2}\,, \tag{5}$$

$$\tan\,(\Delta\varphi_{H,best}) = 2(P_+ - \tfrac{1}{2})\,\sin\,|\Delta\varphi_H|/\cos\,\Delta\varphi_H\,, \tag{6}$$

$$\kappa_{H,H'} = 2\sigma_3/{\sigma_2}^{3/2}\,|E_{-H}E_{H'}E_{H-H'}|\,,\quad \sigma_n = \sum_j Z_j^n\,, \tag{7}$$

$$\Phi'_3 = \varphi''_{-H} + \varphi''_{H'} + \varphi''_{H-H'}\,, \tag{8}$$

$$\chi = 2\,|E_H E_{H,ano}|/\sigma_u\,,\qquad \sigma_u = \sum_u Z_u^2/\sigma_2\,, \tag{9}$$

the subscript $u$ denotes atoms of the non-anomalous scatterers,

$$\delta_H = \varphi_{H,ano} - \varphi''_H\,, \tag{10}$$

$E_{H,ano}$ and $\varphi_{H,ano}$, respectively, are the magnitude and phase of the normalized structure factor of

$$\mathbf{F}'_{ano} = \sum_{j=1}^{N_{ano}} (f_j + \Delta f'_j)\,\exp\,(i2\pi\mathbf{H}\cdot\mathbf{r}_j)\,. \tag{11}$$

The factor $\exp\,(-{\sigma_H}^2/2)$ in equation (5) is related to the "lack of closure error" (Blow and Click, 1959) and is calculated according to Hendrickson and Lattman (1970). Equation (4) can be used in connection with equations (5) and (6) to break the phase ambiguity. At the beginning, $P_+$ is set to 1/2 for all reflections. Equations (5) and (6) are then used to calculate values of $m_H$ and $\Delta\varphi_{H,best}$ for each reflection. These results are then used in equation (4) to obtain a new set of $P_+$. In our test, the minimum value of $\kappa_{H,H'}$ for accepting $\Sigma_2$ relationships was set to 0.03. With the values of $P_+$ so obtained, equations (5) and (6) yield new values of $m_H$ and $\Delta\varphi_{H,best}$, which are used to calculate the original Fourier map for use in the solvent flattening procedure described above. Programs for solvent flattening were available in the CCP4 suit (Collaborative computational project, number 4, 1994). Programs for direct-method phasing were written in our laboratory. Fourier maps for evaluating the results were calculated using the program FRODO (Jones, 1978).

## Test data

The data used in the present test were collected with Cu$K_\alpha$ radiation from the platinum derivative of ribonuclease Sa (RNase Sa) at 2.5 Å resolution (Dodson, Sevcik, Dodson and Zelinka, 1987). The crystals belong to space group $P2_12_12_1$ with unit cell parameters $a = 64.90$ Å, $b = 78.32$ Å and $c = 38.79$ Å. There are two molecules in the asymmetric unit, each with 96 amino-acid residues. Five platinum positions were found in the asymmetric unit, but being only partially occupied so that there are effectively only six platinum atoms in the whole unit cell. The value for the imaginary anomalous-scattering correction of Pt, $\Delta f'' = 6.925$, was used in the following calculations. No correction for the real anomalous scattering was applied.

## Result and discussion

A total of 6995 reflections were used in the test. OAS doublets were calculated from the intensities of Bijvoet pairs ($F_H^+$ and $F_H^-$) and the known substructure of anoma-

**Table 1.** Comparison of mean phase errors resulting from different phasing procedures.

| Number of reflections | $F_{obs}$-weighted mean phase error (°) | | | |
|---|---|---|---|---|
| | Averaged OAS phases | Direct-method resolved OAS phases | Solvent flattening based on averaged OAS phases | Solvent flattening based on direct-method phases |
| 1000 | 68.77 | 55.56 | 51.88 | 43.43 |
| 2000 | 68.14 | 57.16 | 55.52 | 46.76 |
| 3000 | 68.58 | 59.10 | 57.98 | 49.67 |
| 4000 | 68.88 | 60.61 | 59.86 | 52.09 |
| 5000 | 69.65 | 62.23 | 61.69 | 54.51 |
| 6000 | 70.56 | 63.71 | 63.18 | 56.37 |
| (6766) 6995 | (71.03) | (64.34) | 64.00 | 57.42 |

Reflections were sorted in descending order of the observed structure-factor amplitudes and then cumulated into seven groups. The number of reflections in each group is listed in the first column. Values of averaged phase errors were calculated with respect to the final structure model.

**Table 2.** Comparison of map correlation coefficients for the two phasing procedures.

| Correlation coefficient | Number of amino-acid residues | | | |
|---|---|---|---|---|
| | Main chain | | Side chain | |
| | Direct method + Solvent flattening | Solvent flattening alone | Direct method + Solvent flattening | Solvent flattening alone |
| $\geq 0.8$ | 8 | 7 | 5 | 0 |
| $\geq 0.7$ | 38 | 20 | 16 | 7 |
| $\geq 0.6$ | 82 | 61 | 48 | 23 |
| $\geq 0.5$ | 134 | 97 | 80 | 49 |
| $\geq 0.4$ | 162 | 130 | 109 | 84 |

Correlation coefficients of the resultant electron density maps were calculated with respect to the "final" electron density map for individual amino-acid residues. The coefficients were then sorted in descending order and cumulated into five groups as indicated in the first column. Number of residues in each cumulated group was counted for the main chain and the side chain separately.

lous scatterers according to equations (1), (2), and (3). Resultant mean phase errors of different phasing procedures are listed in Table 1. Obviously, the OAS phase ambiguity was effectively broken by either direct method or solvent flattening alone indicated by the great improvement to the averaged OAS phases. However, the combination of direct method and solvent flattening led to much better results than those obtained by either method alone. In Table 2, map correlation coefficients calculated for individual amino-acid residues are compared for the two resultant Fourier maps, one from "direct method + solvent flattening", and another from "solvent flattening alone". The map correlation coefficient is defined as

$$r = (\overline{\varrho_e \varrho_t} - \overline{\varrho}_e \overline{\varrho}_t)/\sigma_{\varrho_e}\sigma_{\varrho_t},$$

where $\varrho_e$ and $\varrho_t$ are the electron densities derived from estimated phases and true phases, respectively, and $\sigma_{\varrho_e}$ and $\sigma_{\varrho_t}$ are their standard deviations. It is clear from the comparison that "direct method + solvent flattening" is definitely superior to "solvent flattening alone". Identical portions of these two Fourier maps are shown in Figure 1. It is concluded that the solvent flattening technique for *ab initio* phasing of OAS data can be greatly strengthened by the incorporation of direct methods. Since the present test was done under conditions common in protein structure analysis, our procedure could be of use in practice.

*Acknowledgments*. The authors should like to thank Dr. Eleanor Dodson and Professor Guy Dodson for making available the diffraction data of RNase Sa. Thanks are also due to Professor Liang Dong-cai for helpful discussions and to Drs. Jiang Tao and Ding Jin-hui for their help on using the program FRODO. Thanks are also due to the referees, whose suggestions improved the presentation of this paper. The project is supported by the Chinese Academy of Sciences.

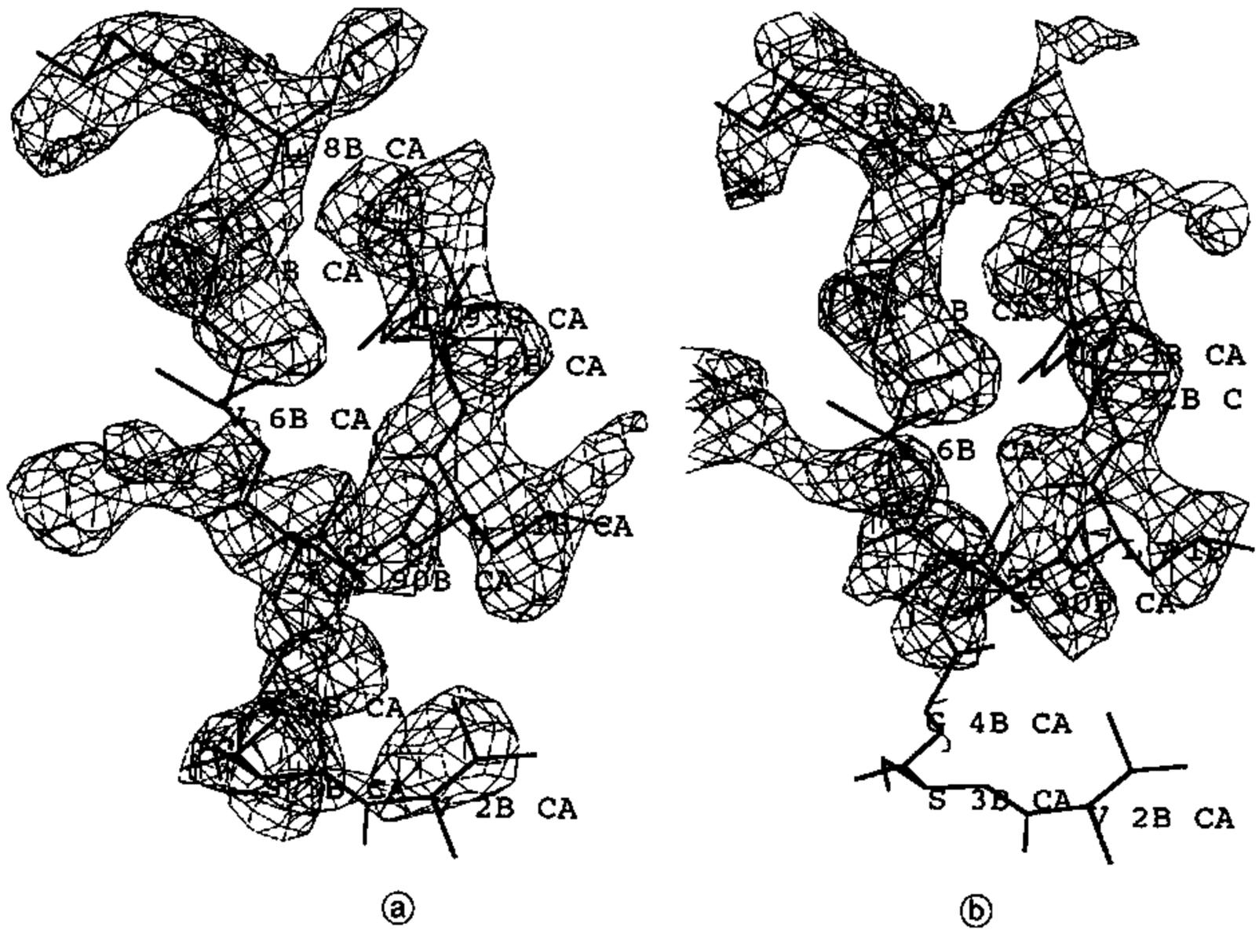


**Fig. 1.** The same portion from the two electron density maps resulting from (**a**): "direct method + solvent flattening"; (**b**): "solvent flattening alone". The maps are contoured at $1.5\sigma$. Some distant contours projected onto the skeleton have been erased manually. The superimposed model is based on the final parameters from the original authors.

## References


Blow, D. M.; Crick, F. H. C.: The Treatment Errors in the Isomorphous Replacement Method. Acta Crystallogr. **12** (1959) 794–802.

Blundell, T. L.; Johnson, L. N.: In: *Protein Crystallography*, p. 338. New York, Academic Press 1976.

Blundell, T. L.; Pitts, J. E.; Tickle, I. L.; Wood, S. P.; Wu, C. W.: X-ray analysis (1.4-Å resolution) of avian pancreatic polypeptide: Small globular protein hormone. Proc. Natl. Acad. Sci. USA **78** (1981) 4175–4179.

Chen, L. Q.; Rose, J. P.; Breslow, E.; Yang, D.; Chang, W. R.; Furey, W. F.; Sax, M.; Wang, B. C.: Crystal structure of a bovine neurophysin II dipeptide complex at 2.8 Å determined from the single-wavelength anomalous scattering single of an incorporated iodine atom. Proc. Natl. Acad. Sci. USA **88** (1991) 4240–4244.

Collaborative Computational Project, Number 4: The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. **D50** (1994) 760–763.

Dodson, G. G.; Sevcik, J.; Dodson, E.; Zelinka, J.: Metabolism of Nucleic Acids, Including Gene Manipulation, p. 33–36. Bratislava, Slovak Academy Science 1987.

Fan, H. F.; Gu, Y. X.: Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. III. The Incorporation of Partial Structure Information. Acta Crystallogr. **A41** (1985) 280–284.

Fan, H. F.; Hao, Q.; Gu, Y. X.; Qian, J. Z.; Zheng, C. D.; Ke, H.: Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. VII. Ab Inition Phasing of One-Wavelength Anomalous Scattering Data from a Small Protein. Acta Crystallogr. **A46** (1990) 935–939.

Hendrickson, W. A.; Lattman, E. E.: Representation of Phase Probability for Simplified Combination of Independent Phase Information. Acta Crystallogr. **B26** (1970) 136–143.

Hendrickson, W. A.; Pähler, A.; Smith, J. L.; Satow, Y.; Merritt, E. A.; Phizackerley, R. P.: Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. Proc. Natl. Acad. Sci. USA **86** (1989) 2190–2194.

Jones, T. A.: A graphics model building and refinement system for macromolecules. J. Appl. Crystallogr. **11** (1978) 268–272.

Robbins, A. H.; McRee, D. E.; Williamson, M.; Collett, S. A.; Xuong, N. H.; Furey, W. F.; Wang, B. C.; Stout, C. D.: Refined Crystal Structure of Cd, Zn Metallothionein at 2.0 Å Resolution. J. Mol. Biol. **221** (1991) 1269–1293.

Sha, B. D.; Liu, S. P.; Gu, Y. X.; Fan, H. F.; Ke, H.; Yao, J. X.; Woolfson, M. M.: Direct Phasing of One-wavelength Anomalous Scattering Data of the Protein Core Streptavidin. Acta Crystallogr. **D51** (1995) 342–346.

Wang, B. C.: Protein Structure Determination By The Single Isomorphous Replacement Method With A Phase Selection And Refinement Process. Acta Crystallogr. **A37** (1981) Suppl. C-11.

Wang, B. C.: Diffraction Methods for Biological Macromolecules. In: *Methods in Enzymology* (Eds. H. Wyckoff, C. H. W. Hirs and S. N. Timasheff), Vol. 115, pp. 90–112. New York, Academic Press, 1985.

Zheng, X. F.; Fan, H. F.; Hao, Q.; Dodd, F.; Hasnain, S. S.: Direct-Method Structure Determination of the Native Azurin II Protein Using One-wavelength Anomalous Scattering Data. Acta Crystallogr. **D52** (1996) 937–941.