THE PROBLEM OF PHASE AMBIGUITY IN SINGLE CRYSTAL STRUCTURE ANALYSIS

FAN HAI-FU

Research group on crystal structure analysis,
Institute of Physics, Chinese Academy of Sciences, Beijing, China

Introduction

It is well known that phase ambiguities cause troubles in the determination of crystal structures. Many difficult structures have such a common property that some kind of phase ambiguities often occurs in the process of structure determination. Phase ambiguity can also appear as an inherent property of certain techniques such as single isomorphous replacement (SIR) and one-wave-length anomalous scattering (OAS).

On the other hand phase ambiguity may be rather an indication of success than a portent of failure. Because when phase ambiguity occurs, the phase problem in fact has been partially solved.

Phase ambiguities may be classified into two categories:

1. Translational phase ambiguity

In this case, phases of reflections belonging to certain index group(s) will remain systematically undetermined after a conventional phase developing process. The task of solving this problem is to find some way to derive the undetermined phases by making use of the known ones.

2. Enantiomorphous phase ambiguity

This can only occur in the determination of non-centrosymmetric structures. In the case that this ambiguity occurs, we can find two possible phases (one true and one fault) associated with each reflection. The remaining task of the structure determination is to make choice between the two possibilities for each reflection. Thus the phase problem reduces to a matter similar to that of solving a centrosymmetric structure.

In our group, direct methods for breaking various kinds of phase ambiguities are in continuously developing. Some of them were successfully applied to a number of difficult small structures. Others are now tested on protein structures and the results are very encouraging.

Phase Ambiguities due to Pseudo Symmetry in Real Space

For detail discussion on this topic, the reader is referred to the paper by Fan & Zheng (1982) and the paper cited therein. Here a brief account is given.

1. Translational phase ambiguity due to pseudo translational symmetry

If in a crystal structure there exists a pseudo translational vector $t = T/n$, where T is the shortest exact translation vector in the structure parallel to t and n is an integer, then there will exist some group(s) of reflections having systematically weak intensities. The reciprocal vector H of these reflections will satisfy the relation $H \cdot t \neq n$. In the process of the structure determination, phases of reflections belonging to the 'strong' group(s) with $H \cdot t = n$ can be obtained by using conventional methods, but the phases of reflections belonging to the 'weak' group(s) with $H \cdot t \neq n$ will be rather difficult to determine. This results in an n-fold superimposed image, which contains the true structure together with $n-1$ translational images of the structure. Normally it would be very difficult to pick up the true structure from such a multiple image. Fan (1975) proposed a method for solving this problem by making use of a modified Sayre equation. The method has been verified with the crystal structure of SHAS, $C_5H_6O_5N_3K$, which crystallizes in space group $P2_12_12_1$ with $a = 7.51$ Å, $b = 9.95$ Å, $c = 10.98$ Å and $Z = 4$. The K atom in the
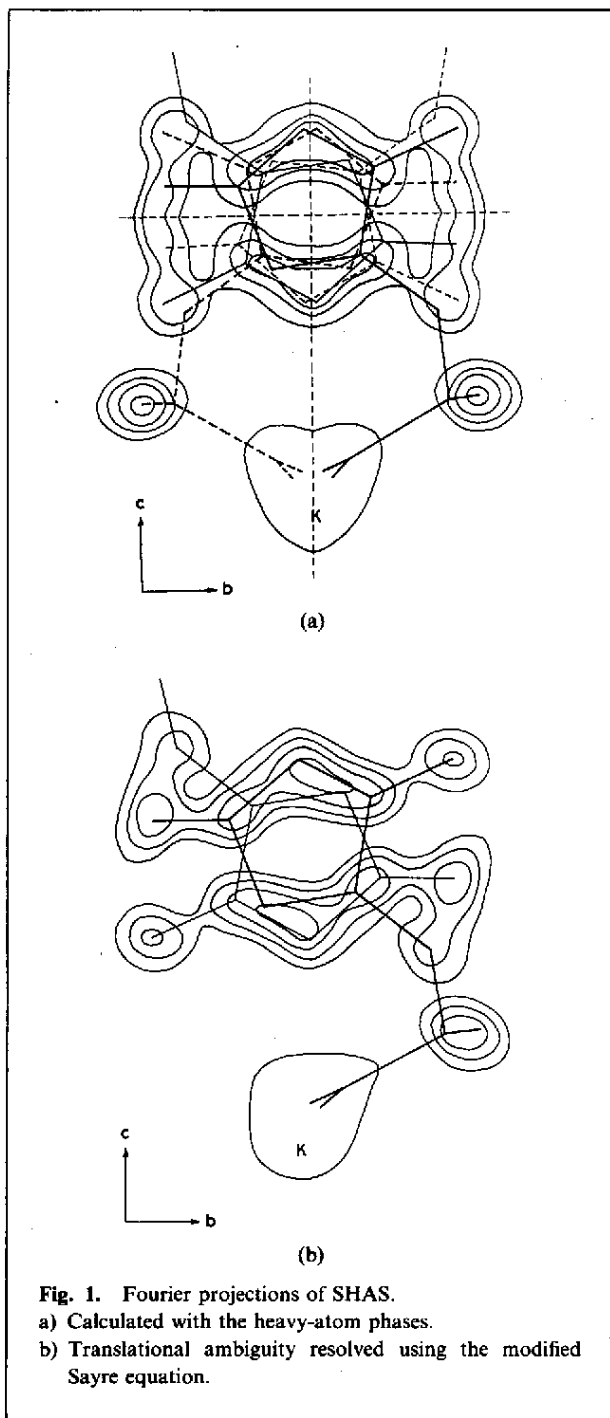
Fig. 1. Fourier projections of SHAS.
a) Calculated with the heavy-atom phases.
b) Translational ambiguity resolved using the modified Sayre equation.

as two dotted lines parallel to the b and c axes respectively, which are originated from the pseudo translational symmetry t. According to Fan (1965a, 1975) we have

$$F_H^0 = (f/V)\sum_{H'} F_{H'}^0 F_{H-H'}$$
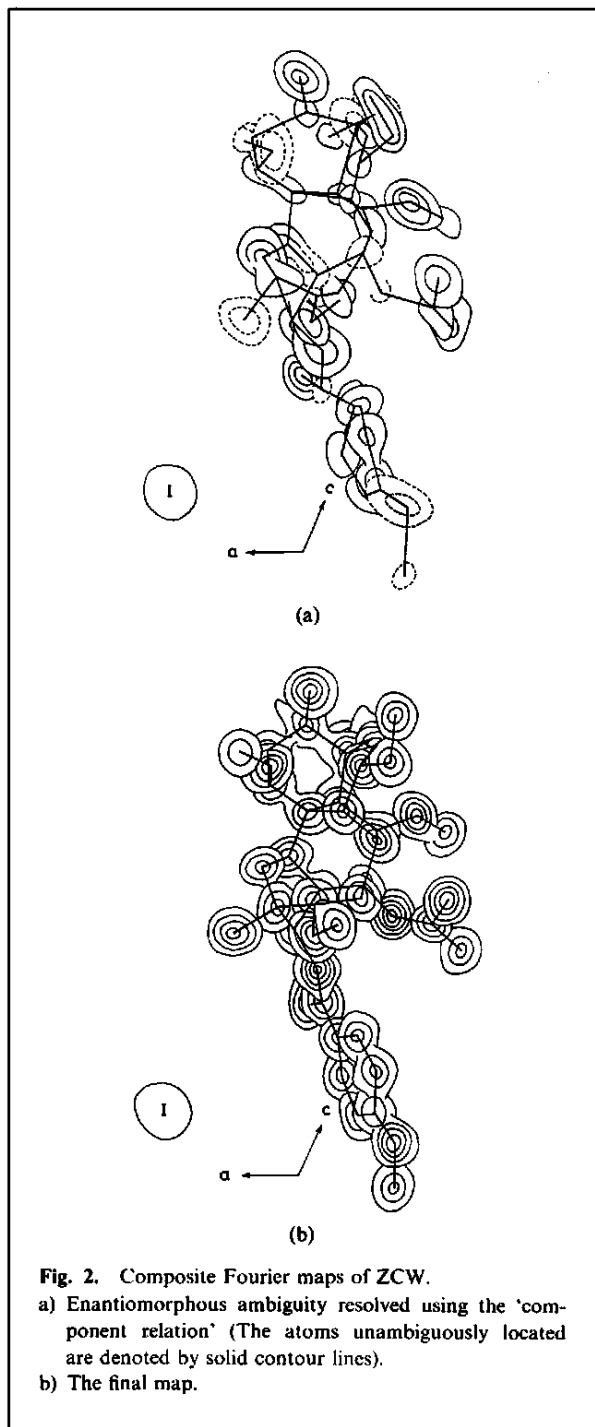
(1)

where F_H^0 denotes the structure factor belonging to the 'weak' group(s). With (1) the phases of the 'weak' reflections were derived making use of those from the 'strong' ones. The resulting Fourier projection is shown in Fig. 1b, in which the pseudo symmetry has been effectively eliminated. The method has further developed to tackle the phase problem in solving superstructures (Fan, He, Qian & Liu, 1978) and finally incorporated into the MULTAN program (Fan, Yao, Main & Woolfson, 1983). Studies on the application of direct methods to structures having pseudo translational symmetry have also been presented by other authors (Gramlich, 1975, 1978, 1984; Boehme, 1982; Prick, Beuskens & Gould, 1983; Giacovazzo, 1984)

2. Enantiomorphous phase ambiguity due to pseudo centrosymmetry

Non-centrosymmetric structures can reveal pseudo centrosymmetry if they contain some dominating heavy-atoms in centrosymmetric arrangement. In this case either Patterson or conventional direct methods would result in a pseudo centrosymmetric image, in which the true structure and its enantiomorph are superimposed. This obstructs the solution of the structure. Special methods have been proposed to overcome this obstacle (Van den Hark, Prick & Beuskens, 1976; Prick, Beuskens & Gould, 1978; Hull & Irwin, 1978; Prick, Beuskens & Gould, 1983). With these methods a set of difference structure factors are first calculated by subtracting the heavy-atom contribution from the observed structure factors, then the light atoms can be found by solving the 'difference structure'. From another point of view, the existence of the centrosymmetric arragement of the heavy-atoms might facilitate rather than obstruct the solution of the structure. One can calculate easily the real part of the structure facrors from a map containing both enantiomorphs. Then the absolute value of the imaginary part of the structure factors can be obtained as

$$|B_H| = (|F_H|^2 - A_H^2)^{1/2},$$

where A_H and B_H are the real and imaginary parts of the structure factors F_H respectively. Now it can be seen that, owing to the existence of the centrosymmetric arrangement of heavy-atoms, the phase problem of a non-centrosymmetric structure reduces to a problem of finding the signs of B_H's. This can be

asymmetric unit is situated at $x=0.000$, $y=0.250$, $z=0.890$. Hence the arrangement of the heavy-atoms possess a subperiodicity of translation, $t=(a+b+c)/2$. Consequently the heavy-atoms have no contribution to the reflections with $h+k+l$ odd. Thus the phases of these reflections were difficult to determine by either heavy-atom or conventional direct methods. Figure 1a shows the Fourier projection along the a axis calculated with the heavy-atom phases. In this projection there exist two pseudo mirror planes, shown

Fig. 2. Composite Fourier maps of ZCW.
a) Enantiomorphous ambiguity resolved using the 'component relation' (The atoms unambiguously located are denoted by solid contour lines).
b) The final map.

parameters $a = 12.58$ Å, $b = 14.38$ Å, $c = 11.00$ Å, $\beta = 114.6°$ and $Z = 2$. Neither Patterson method nor direct methods as MULTAN-80 (Main *et al.*, 1980) and RANTAN (Yao, 1981) could solve the structure. From the resulting ambiguous Fourier map, 331 largest A_H and 79 largest $|B_H|$ were obtained. The signs of B_H's were then derived using (2). A Fourier map calculated with the 331 A_H and 79 B_H revealed 33 of the total 46 light-atoms unambiguously (Fig. 2a). The final Fourier map is shown in Fig. 2b.

Phase Ambiguities due to the Phasing Procedure in Reciprocal Space

1. Phase degeneration in direct methods

A serious problem in the application of direct solved without much difficulties by making use of the 'component relation' (Fan, 1965b).
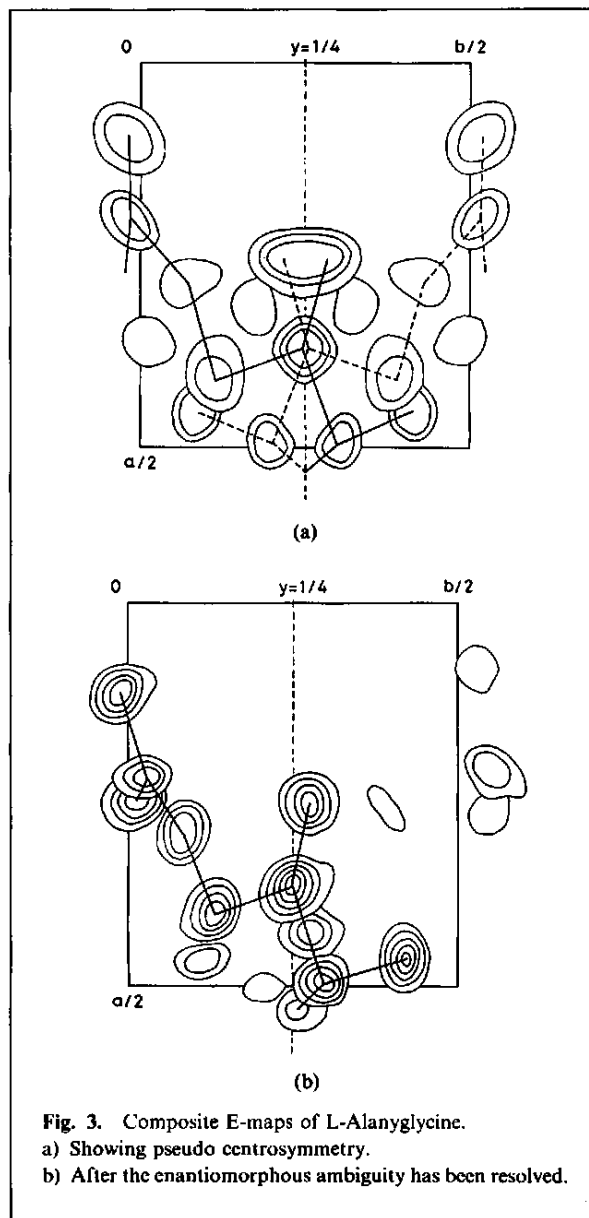
$$B_H = (2f/V)\sum_{H'} A_{H'} B_{H-H'} \qquad (2)$$

The procedure has been verified (Fan & Zheng, 1978) by a typical difficult small structure ZCW, $C_{34}O_{11}NH_{47} \cdot HI$, with space group $P2_1$ amd unit cell

Fig. 3. Composite E-maps of L-Alanyglycine.
a) Showing pseudo centrosymmetry.
b) After the enantiomorphous ambiguity has been resolved.

methods is the problem of phase degeneration, i.e. the phases derived for a non-centrosymmetric structure converge to a set of centrosymmetric phases. This often occurs when the crystal belongs to a polar space group such as $P2_1$, $C2$, $P4_1$, etc. When phase degeneration occurs, the resulting E-map will contain both enantiomorphs. Some authors considered this problem as the instability of enantiomorph fixation. Accordingly a number of procedures strengthening the enantiomorph specification and stabilizing the enantiomorph fixation have been proposed (Duax & Hauptman, 1972; Busetta, 1976; Gilmore, 1977; Hull & Irwin, 1978; Olthof & Schenk, 1981). However, the problem can be solved by an alternative method (Fan & Qian, 1981), which makes use of the superimposed image resulting from phase degeneration to calculate a set of A_H and $|B_H|$, then solves the 'sign problem' of B_H by the component relation (2). The efficiency of the method has been test with the structure of L-Alanyglycine, $C_5H_{10}N_2O_3$, which was solved originally by Koch & Germain (1970). The crystals belong to space group $P2_1$ with $a = 5.283$ Å, $b = 11.810$ Å, $c = 5.510$ Å, $\beta = 101.58°$ and $Z = 2$. The test calculation was divided into two stages. In the first stage, the structure was assumed to be centrosymmetric with space group $P2_1/m$, the signs rather than phases were derived by a conventional direct method. This resulted in an E-map containing both enantiomorphous with a few ghost peaks (Fig. 3a). In the second stage, a set of A_H and $|B_H|$ was calculated according to the above E-map and the signs of B_H were derived by making use of (2). In the resulting E-map (Fig. 3b) one of the enantiomorphs has been eliminated and the true structure clearly revealed.

2. Enantiomorphous phase ambiguities in SIR and OAS methods

In the case of single isomorphous replacement (SIR), for a given reciprocal vector H, we have

$$F_{H,N} = F_{H,D} - F_{H,R} \tag{3}$$

where $F_{H,N}$ is the structure factor of the native protein, $F_{H,D}$ is that of the heavy-atom derivative and $F_{H,R}$ is the contribution of the replacing-atoms to $F_{H,D}$. The magnitudes of $F_{H,N}$ and $F_{H,D}$ can be obtained from experiment. Accordingly the parameters of the replacing-atoms can be found and $F_{H,R}$ be calculated. Consequently, we have two ways for drawing the triangle of (3) leading to an enantiomorphous phase doublet for both $F_{H,N}$ and $F_{H,D}$ in the phase-vector diagram as shown in Fig. 4.

In the case of one-wave-length anomalous scattering (OAS), we have

$$F_H^+ = F_H + F_{H,A}'' \tag{4}$$

and

$$F_H^{-*} = F_H - F_{H,A}'' \tag{5}$$

Here F_H is the contribution of both the normal scattering and real part of the anomalous scattering, $F_{H,A}''$ is the contribution of the imaginary part of the anomalous scattering, F_H^{-*} denotes the conjugate of F_H^-. It follows from (4) and (5) that,

$$F_H^+ - F_H^{-*} = 2F_{H,A}'' \tag{6}$$

The magnitudes of F_H^+ and F_H^{-*} can be obtained from experiment and then $F_{H,A}''$ can be derived. Hence we also have two ways for drawing the triangle of (6) leading to an enantiomorphous phase doublet for F_H, as shown in Fig. 5.

If there is some way to break the above phase amiguities, then SIR and OAS method will be of great
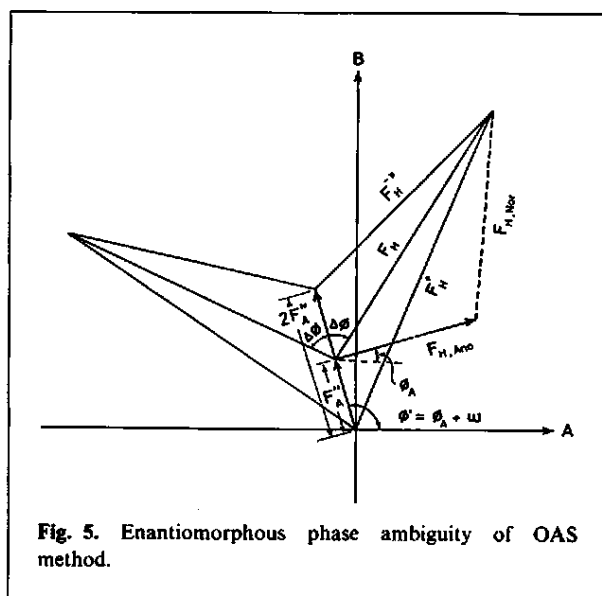
Fig. 5. Enantiomorphous phase ambiguity of OAS method.

importance in protein structure analysis. Because it is then possible to determine a protein structure from a single pair of isomorphous crystals or just the heavy-atom derivative. Even more, if some suitable anomalous scatterers are contained in the native protein, then by OAS method, the structure might be determined from the native protein alone. Real space method was first proposed to slove the phase ambiguity of SIR method (Blow & Rossmann, 1961). New development has recently been made. With the so-called ISIR method (Wang, 1981; 1984) a dozen of unknown protein structures have been solved. However, the method is subject to the limitation that it would not be applicable when the replacing-atoms are in a centrosymmetric arrangement. Reciprocal space method using partial structure information to resolve the phase ambiguity of OAS method has been proposed by Hendrickson & Teeter (1981). The method has successfully solved some unknown protein structures. On the other hand, attempts have been made since 1960's to solve the phase ambiguities of SIR and OAS method by making use of direct methods (Fan, 1965b; Coulter, 1965; Karle, 1966). Recently, Hauptman (1982a, b), Giacovazzo (1983) and Karle (1983, 1984) have succeeded in deriving large number of three-phase structure invariants from the error free data of a model protein structure. An alternative procedure has been proposed in our group (Fan, 1983; Fan, Han, Qian & Yao, 1984; Fan, Han & Qian, 1984; Fan & Gu, 1984). The method is capable of deriving individual phases instead of three-phase invariants and has been tested with error free and experimental data of a number of proteins. Further description will be given in the following paragraph. Owing to the capacity of linking together phase information from various sources, direct method would eventually not be an opponent but rather a companion of other methods for breaking enantiomorphous phase ambiguities.

3. A generalized method for breaking enantiomorphous phase ambiguities and the application to protein structures

All the enantiomorphous phase ambiguities described in this paper can be expressed by the phase doublet

$$\varphi_H = \varphi'_H \pm |\Delta\varphi_H| \ .$$

In the case of SIR:

$$\varphi'_H = \varphi_{H.R}$$

and

$$\Delta\varphi_H = \pm\cos^{-1}[(F^2_{H.D} - F^2_{H.R} - F^2_{H.N})/2F_{H.R}F_{H.N}] \ .$$

In the case of OAS:

$$\varphi'_H = \varphi''_{H.A} \ ,$$

where $\varphi''_{H.A}$ is the phase of $F''_{H.A}$.

$$\Delta\varphi_H = \pm\cos^{-1}[(F^+_H - F^-_H)/2F''_{H.A}] \ .$$

In the case that the phase doublets are due to a centrosymmetric partial structure or a pseudo centrosymmetric E-map:

$$\varphi'_H = 2\pi H \cdot r_0 \ ,$$

where r_0 is the positional vector of the pseudo inverse centre with respect to the origin of the unit cell. $|\Delta\varphi_H|$ is calculated by

$$\tan\Delta\varphi_H = \pm(F^2_H - C^2_H)^{1/2}/C_H \ ,$$

where

$$C_H = \sum_j f_j \cos 2\pi H \cdot (r_j - r_0) \ ,$$

r_j denotes the positional vector of any one of the two possible positions, which corresponding to the jth atom in the unit cell.

Now the problem of resolving any kind of enantiomorphous phase ambiguity is reduced to that of finding the signs of $\Delta\varphi_H$. According to Fan (1983), the probability for $\Delta\varphi_H$ to be positive is given as

$$P_+(\Delta\varphi_H) = \frac{1}{2}$$

$$+ \frac{1}{2}\tanh\left[\sin|\Delta\varphi_H|\sum_{H'}K_{HH'}\cdot\sin(\Phi'_3 + \Delta\varphi_{H'} + \Delta\varphi_{H-H'})\right] ,$$

(7)

where

$$K_{HH'} = 2\sigma_3\sigma_2^{-3/2}E_HE_{H'}\cdot E_{H-H'} \ ,$$

$$\Phi'_3 = -\varphi'_H + \varphi'_{H'} + \varphi'_{H-H'} \ .$$

Using the concept of 'best phase relationship' (Fan, Han & Qian, 1984) and incorporating the partial structure information (Fan & Gu, 1984), equation (7) can be modified to give

$$P_+(\Delta\varphi_H) = \frac{1}{2} + \frac{1}{2}\tanh\left\{\sin|\Delta\varphi_H|\left[\sum_{H'}m_{H'}m_{H-H'}K_{HH'}\right.\right.$$

$$\left.\left.\times\sin(\Phi'_3 + \Delta\varphi_{H'best} + \Delta\varphi_{H-H'best}) + x\sin\delta_H\right]\right\}$$

(8)

with

$$m_H = \exp(-\sigma^2_H/2)\left\{\left[2\left(P_+ - \frac{1}{2}\right)^2\right.\right.$$

$$\left.\left. + \frac{1}{2}\right](1 - \cos 2\Delta\varphi_H) + \cos 2\Delta\varphi_H\right\}^{1/2} ,$$

(9)

$$\tan(\varDelta\varphi_{H'best}) = 2\left(P_+ - \frac{1}{2}\right)\sin|\varDelta\varphi_H| \Big/ \cos\varDelta\varphi_H \quad (10)$$

and

$$x = 2\sigma_2 E_H E_{H,p} \Big/ \sum_u Z_u^2,$$

$$\delta_H = \varphi_{H,p} - \varphi'_H.$$

The subscripts p and u denote the known and un-known part of the structure respectively. Applications of (8), (9) and (10) to protein structures are elucidated in the following examples:

Example 1. Application to error free SIR data of insulin and its Pb-derivative with the replacing-atoms in non-centrosymmetric arrangement

Error free SIR data were calculated according to the model structures of insulin and its Pb-derivative. Insulin crystallizes in space group R3 with $a = 82.5\,\text{Å}$, $c = 34.0\,\text{Å}$, $\gamma = 120°$ and $Z = 9$. There are ~ 6400 independent reflections at 1.9 Å resolution. In the test calculation only 1000 largest E's were used and 60000 of the total $\sim 75000\ \Sigma_2$ relationships were involved.

Starting with $P_+ = 1/2$, values of m_H and $\varDelta\varphi_{H best}$ were calculated using (9) and (10) respectively and

Table 1. Test result on the error free SIR data of insulin.

Group	%	ER
1	93.5	7
2	89.8	11
3	82.5	16
4	77.1	18

The reflections were arrangeged in descending order of $P = |P_+ - 1/2|$ and then cumulated into 4 groups. The groups numbered 1, 2, 3 and 4 contain the top 200, 400, 600 and 800 reflections respectively.

%: the percentage of reflections with the signs of $\varDelta\varphi_H$ correctly determined.

ER: the averaged error of phases in degree.

then substituted into (8) to calculate the probabilities. Most of the P_+ so obtained differed greatly from 1/2, the phase ambiguities were thus broken. With the newly calculated P_+, one more cycle of iteration led to further improvement on the reliability. The results are listed in Table 1. It shows that the method is capable of deriving large number of initial phases with high reliability.

Example 2. Application to error free SIR data of APP and its Hg-derivative with the replacing-atoms in centrosymmetric arrangement

APP (avian pancreatic polypeptide) crystallizes in space group C2 with $a = 34.18\,\text{Å}$, $b = 32.92\,\text{Å}$, $c = 28.44\,\text{Å}$, $\beta = 105.30°$ and $Z = 4$. There are ~ 2100 independent reflections at 2 Å resolution. 1000 of them with largest E's were used and 60000 of the total $\sim 130000\ \Sigma_2$ relationships were involved in the test.

Two kinds of phase ambiguities simultaneously occured in this example. One is inherent in the SIR method, this can be resolved as in the above example. The other comes from the special arrangement of the replacing-atoms, this causes the term $\sin(\varPhi'_3 - \varDelta\varphi_{H'best} + \varDelta\varphi_{H-H'best})$ to be identical with zero at the begining of iteration. In order to over-come this difficulty a multi-solution procedure using random starting sign sets was applied (Yao & Fan, 1984). Results from the first ten random starting sets are listed in Table 2 from left to right in descending order of the figures of merit. It can be seen that the results having the highest figures of merit are in good agreement with the theorectical values.

Example 3. Application to experimental OAS data of the Hg-derivative of APP

In this example OAS data of the Hg-derivative of APP were treated such as in solving an unknown structure. 1000 largest E's and 60000 largest Σ_2 relationships were used. The result is listed in Table 3. From the result, it is reasonable to expect that the structure can actually be solved by the method used here.

One of the present trends on the development of

Table 2. Test result on the error free SIR data of APP.

FOM	.4136		.4091		.3995		.3918		.3848		.3690		.3575		.3518		.3446		.3362	
Set	1		8		9		2		3		10		5		4		7		6	
Group	%	ER	%	ER	%	ER	%	ER	%	ER	%	ER	%	ER	%	ER	%	ER	%	ER
1	92.0	7	90.5	9	95.0	6	89.5	10	87.0	16	80.0	21	68.5	37	59.5	49	69.0	39	66.5	42
2	86.3	15	83.3	18	88.5	11	81.8	18	82.3	19	75.3	26	63.3	42	57.0	49	63.5	40	60.3	44
3	78.5	20	78.0	20	79.8	18	74.8	22	74.8	23	68.7	30	62.5	39	56.0	46	62.7	38	60.2	42
4	72.4	22	72.8	21	74.3	20	70.6	23	70.6	25	64.1	30	60.8	37	53.6	43	58.9	37	56.9	39

Table 3. Test result on the experimental OAS data of APP.

Group	%	ER
1	95.0	24
2	90.8	30
3	91.3	29
4	89.0	32

direct methods is to collaborate with other methods rather than to supplant them. The combination of direct methods with SIR or OAS method may be an entrance for direct methods to the *ab-initio* phasing of protein diffraction data. This in turn may improve the art of protein structure analysis.

The author is indebted to Drs. G. Dodson and E. Dodson and to Prof. T. L. Blundell for making available the insulin and APP data respectively.

References

Blow, D. M. and Rossmann, M. G. (1961), *Acta Cryst.*, 14, 1195–1202.

Böhme, R. (1982), *Acta Cryst.*, A38, 318 326.

Busetta, B. (1976), *Acta Cryst.*, A32, 139 143.

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

Duax, W. L. and Hauptman, H. (1972), *Acta Cryst.*, B28, 2912–2916.

Fan, H.-F. (1965a), *Acta Physica Sinica*, 21, 1105 1113.

Fan, H.-F. (1965b), *Acta Physica Sinica*, 21, 1114–1118.

Fan, H.-F. (1975), *Acta Physica Sinica*, 24, 57–60.

Fan, H.-F. (1983), in IUCr Summer School on Crystallographic Computing, Kyoto, Japan.

Fan, H.-F. and Gu, Y.-X. (1984), *Acta Cryst.* (Sumbitted for publication)

Fan, H.-F., Han, F.-S. and Qian, J.-Z. (1984), *Acta Cryst.*, A40 (to be published).

Fan, H.-F., Han, F.-S., Qian, J.-Z. and Yao, J.-X. (1984), *Acta Cryst.*, A40 (to be published).

Fan, H.-F., He, L., Qian, J.-Z. and Liu, S.-X. (1978), *Acta Physica Sinica*, 27, 554–558.

Fan, H.-F. and Qian, J.-Z. (1981), *Acta Physica Sinica*, 30, 594–601.

Fan, H.-F., Yao, J.-X., Main, P. and Woolfson, M. M. (1983), *Acta Cryst.*, A39, 566 569.

Fan, H.-F. and Zheng, Q.-T. (1978), *Acta Physica Sinica*, 27, 169–174.

Fan, H.-F. and Zheng, Q.-T. (1982), *Acta Physica Sinica*, 31, 191–198.

Giacovazzo, C. (1983), *Acta Cryst.*, A39, 585–592.

Giacovazzo, C. (1984), private communication.

Gilmore, C. J. (1977), *Acta Cryst.*, A33, 712–716.

Gramlich, V. (1975), *Acta Cryst.*, A31, S90.

Gramlich, V. (1978), *Acta Cryst.*, A34, S43.

Gramlich, V. (1984), private communication.

Gauptman, H. (1982a), *Acta Cryst.*, A38, 289–294.

Hauptman, H. (1982b), *Acta Cryst.*, A38, 632 641.

Hendrickson, W. A. and Teeter, M. M. (1981), *Nature* (London), 290, 107–113.

Hull, S. E. and Irwin, M. J. (1978), *Acta Cryst.*, A34, 863–870.

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

Karle, J. (1983), *Acta Cryst.*, A39, 800–805.

Karle, J. (1984), *Acta Cryst.*, A40, 4–11.

Koch, M. H. J. and Germain, G. (1970), *Acta Cryst.*, B26, 410.

Main, P., Fiske, S. J., Hull, S. E., Lessinger, L., Germain, G., Declercq, J. P. and Woolfson, M. M. (1980), A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data. Universities of York, England, and Louvain, Belgium.

Olthof, G. J. and Schenk, H. (1981), *Acta Cryst.*, A37, 684–689.

Prick, P. A. J., Beuskens, P. T. and Gould, R. O. (1978), *Acta Cryst.*, A34, S42.

Prick, P. A. J., Beuskens, P. T. and Gould, R. O. (1983), *Acta Cryst.*, A39, 570–576.

Van den Hark, Th. E. M., Prick, P. A. J. and Beuskens, P. T. (1976), *Acta Cryst.*, A32, 816–821.

Wang, B. C. (1981), *Acta Cryst.*, A37, Suppl. C-11.

Wang, B. C. (1984), in "Diffraction Methods for Biological Macromolecules" a volume of "Methods in Enzymology" Ed. by H. W. Wyckoff. London: Academic Press. (In press)

Yao, J.-X. (1981), *Acta Cryst.*, A37, 642–644.

Yao, J.-X. and Fan, H.-F. (1984), *Acta Cryst.*, (Submitted for publication)