Molecular Structure: Chemical Reactivity and Biological Activity

Edited by

John J. Stezowski

Universitat Stuttgart

Co-edited by

Jin-Ling Huang

Fuzhou University

and

Mei-Cheng Shao

Beijing University

INTERNATIONAL UNION OF CRYSTALLOGRAPHY
OXFORD UNIVERSITY PRESS
1988

55. New developments in direct methods

Hai-fu Fan

1. INTRODUCTION

Even in the early days of the development of direct methods, the investigators had to answer constantly the question: "Is the method now going to the end?". In reply to this, some people said "Yes" and withdrew from the competition, others said "No" and made further development of the method. Now direct methods have become routine works in the determination of small molecular structures. However it does not mean the development of direct methods should be finished. Contradictorily it signals that the time is now coming for direct methods to explore new fields of applications. In fact, new achievements have already been made in recent years. In this paper, three topics are to be discussed :

(1) Solving structures with pseudo-translational symmetry;

(2) Tackling the phase problem of macromolecular structures by direct phasing of the SIR or OAS data;

(3) Image processing in high resolution electron microscopy.

2. SOLVING STRUCTURES WITH PSEUDO-TRANSLATIONAL SYMMETRY

Structures with pseudo-translational symmetry are important in mineralogy, structural chemistry and solid state physics, However they are difficult to solve owing to the existence of certain kind of systematically weak reflections. Accordingly a direct method has been developed in order to get rid of this dif-

ficulty.A structure $\rho(\underline{r})$ possesses the pseudo-translational symmetry $\underline{t}$ of the order p, if there exists $\underline{t} = \underline{T}/p$ so as

$$\rho(\underline{r}) \sim \rho(\underline{r} + \underline{t}) \quad . \tag{2.1}$$

(2.1) means that

$$\rho(\underline{r}) \neq \rho(\underline{r} + \underline{t}) \quad \text{and} \quad \int_V |\rho(\underline{r}) - \rho(\underline{r} + \underline{t})| d\underline{r} << \int_V \rho(\underline{r}) d\underline{r} \quad .$$

Where $\underline{t}$ is a pseudo-translation vector; $\underline{T}$ is the shortest lattice vector parrallel to $\underline{t}$; p is an integer greater than 1; V is the volume of unit cell. In the above case, the structure factor can be written approaximately as:

$$F_H \sim \sum_{j=1}^{N/p} f_j \exp(i2\pi \underline{H}\cdot\underline{r}_j) + f_j \exp\{i2\pi \underline{H}\cdot(\underline{r}_j + \underline{t})\} +$$

$$+ f_j \exp\{i2\pi \underline{H}\cdot(\underline{r}_j + 2\underline{t})\} + \cdots + f_j \exp\{i2\pi \underline{H}\cdot(\underline{r}_j + (p-1)\underline{t})\}$$

$$= \sum_{j=1}^{N/p} f_j \exp(i2\pi \underline{H}\cdot\underline{r}_j) \times$$

$$\left[1 + \exp(i2\pi \underline{H}\cdot\underline{t}) + \exp(i2\pi \underline{H}\cdot 2\underline{t}) + \cdots + \exp\{i2\pi \underline{H}\cdot(p-1)\underline{t}\}\right] \tag{2.2}$$

The sum of the series in the bracket of (2.2) is given by

$$S = \{\exp(i2\pi \underline{H}\cdot p\underline{t}) - 1\} / \{\exp(i2\pi \underline{H}\cdot\underline{t}) - 1\}$$

$$= \begin{cases} p, & \text{if } \underline{H}\cdot\underline{t} = \text{integer} \quad ; \\ 0, & \text{if } \underline{H}\cdot\underline{t} \neq \text{integer} \quad . \end{cases}$$

Notice that $\underline{H}\cdot p\underline{t} = \underline{H}\cdot\underline{T}$ = integer. Hence all reflections with $\underline{H}\cdot\underline{t} \neq$ integer will be systematically weak leading to an effect of pseudo systematic extinction. In other words, all the strong reflections will satisfy the condition $\underline{H}\cdot\underline{t}$ = integer. Usually in solving a structure with pseudo-translational symmetry the phases of the systematically 'strong' reflections are relatively easy to determine. However the derivation of phases for the systematically 'weak' reflections will be extremely difficult. In order to solve the phase problem for the 'weak' reflections by making

use of the phases previously obtained for the 'strong' ones, a modified Sayre equation was derived (Fan, 1975; Fan, He, Qian & Liu, 1978; see also Fan, Yao, Main & Woolfson, 1983):

$$F_{H,wk} = 2\frac{\Theta}{V} \sum_{H'} F_{H',str} F_{H-H',wk} \quad . \tag{2.3}$$

An automatic procedure to use formula (2.3) was established under the collaboration between the direct methods group in York and that in Beijing. The main points of the procedure are as follows:

(1) Automatic search of the pseudo systematic extinction rule is first performed and the reflections are grouped accordingly;

(2) Each reflections group is normalized independently;

(3) The $\sum_2$ relationships involving three 'weak' reflections are eliminated and the phase development process is devided into two steps. In the first step only the phases of the 'strong' reflections are developed. Then in the second step the phases of the 'weak' reflections are derived by making use of the phases of the 'strong' ones.

This procedure has been incorporated in the structure analysis program system SAPI-85 (Yao, Zheng, Qian, Han, Gu & Fan, 1985; Fan, 1986). An example of automatically solving an unknown structure with pseudo-translational symmetry by SAPI-85 is given below:

RLD6, $C_{27}H_{31}NO_9$, belongs to space group $P2_1/c$ with a=24.420, b=8.363, c=24.747 Å, β=90.22°. There are 74 non-hydrogen atoms in the asymmetric unit which includes two independent molecules related by a pseudo-translation vector $\underline{t} = (\underline{a} + \underline{c})/2$. RANTAN-81 (Yao,1981) failed to solve the structure from 100 starting sets with the default control parameters. On the other hand SAPI-85 with default control yielded almost the complete structure from the 70th random starting set (see Fig.2.1). The diffraction data and refined atomic parameters of RLD6 were kindly provided by Professor T. C. W. Mak.

Studies on the application of direct methods to structures with pseudo-translational symmetry have also been reported by other authors (Gramlich, 1975, 1978, 1984; Böhme, 1982; Prick, Beurskens & Gould, 1983; Giacovazzo, 1984).

Among the structures having pseudo-translational symmetry, there is a typical class known as superstructure. Superstructure and incommensurate structure are both modulated structures. Up to now, there is no straightforward way to solve incommensurate structures, in spite of their significance in pure and applied sciences. Further development of the direct method described above may be of use in dealing with incommensurate structures.

Recently an interesting type of condensed materials called quasicrystal has been discovered and extensively investigated. The structure of quasicrystals, though not containing translational symmetry in 3-dimensional space, can be described in reciprocal space by a way similar to that used for incommensurate structures. It can be expected that direct methods can also find their use in solving the structures of quasicrystals.

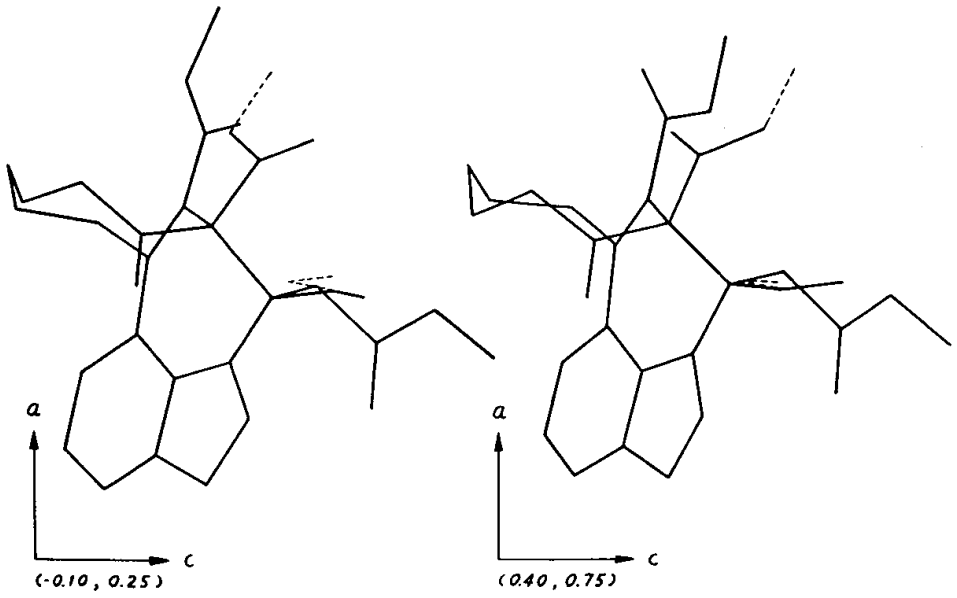

Figure 2.1

Two independent molecules of RLD6 projected along the b axis. The fragments shown with solid lines were obtained from the E-map output by SAPI-85 directly.

3. TACKLING THE PHASE PROBLEM OF MACROMOLECULAR STRUCTURES BY DIRECT PHASING OF THE SIR OR OAS DATA

Since the 1960's attempts have been made to combine direct methods with the single isomorphous replacement (SIR) and the one-wavelength anomalous scattering (OAS) methods (Coulter, 1965; Fan, 1965; Karle, 1966). Such a combination would be important for the structure analysis of proteins due to the possibilities of reducing the number of heavy-atom derivatives needed for solving a protein structure, saving the time in data collection thus enhancing the effective lifetime of the protein crystal and simplifying the process of the structure determination. However during the last decade, procedures proposed were not as successful as expected. Recently some progresses have been achieved. Hauptman (1982a,b) integrated the probabilistic theory of the triplet structure invariants with the SIR and OAS techniques. Giacovazzo (1983) reported a similar theory. Karle (1983; 1984a, b) proposed simple rules for the evaluation of triplet structure invariants from SIR or OAS data. All these methods have been successful in deriving large number of reliable triplet structure invariants using error free data. An alternative procedure has been proposed (Fan, Han, Qian & Yao, 1984; Fan, Han & Qian, 1984; Fan & Gu, 1985; Yao & Fan, 1985; Qian, Fan & Gu, 1985), which, by the test with experimental data, has been proved to be efficient in breaking the enantiomorphous phase ambiguities yielding large number of reliable individual phases.

3.1. Enantiomorphous phase ambiguities in SIR and OAS methods

In the SIR case, for a given reciprocal vector $\underline{H}$, we have

$$F_{H,N} = F_{H,D} - F_{H,R} \quad . \tag{3.1.1}$$

Where $F_{H,N}$, $F_{H,D}$ and $F_{H,R}$ are the structure factors of the native protein, the derivative and the replacing atoms respectively. The moduli of $F_{H,N}$ and $F_{H,D}$ can be obtained experimentally. 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 (3.1.1) leading to an phase doublet for both $F_{H,N}$ and $F_{H,D}$ in the phase vector diagram, Fig. 3.1.
In the case of OAS, we have

$$F_H^+ = F_H + F''_{H,A} \quad \text{and} \quad F_H^{-*} = F_H - F''_{H,A} \quad . \tag{3.1.2}$$

Here F_H is the contribution of both the normal and the real part anomalous scattering from the whole unit cell. F_H^{-*} denotes the conjugate of F_H^- . $F''_{H,A}$ is the contribution from the imaginary-part scattering of the anomalous scatterrers, i.e.

$$F''_{H,A} = \sum_{A=1}^{N_A} i\Delta f''_A \exp(i2\pi \underline{H}\cdot\underline{r}_A) \quad .$$

It follows from (3.1.2) that

$$F_H^+ = F_H^{-*} + 2F''_{H,A} \quad . \tag{3.1.3}$$

The moduli of F_H^+ and F_H^{-*} can be obtained experimentally and then $F''_{H,A}$ can be derived. Hence we have also two ways to draw the triangle of (3.1.3) leading to an enantiomorphous phase doublet for F_H as shown in Fig. 3.2. Both the phase doublets in SIR and OAS cases can be expressed in the generalized form

$$\phi_H = \phi'_H \pm |\Delta\phi_H| \quad , \tag{3.1.4}$$

In the case of SIR:

$$\begin{aligned} \phi'_H &= \phi_{H,R} \quad , \\ \Delta\phi_{H,N} &= \pm \cos^{-1}\{(F^2_{H,D} - F^2_{H,R} - F^2_{H,N})/2F_{H,R}F_{H,N}\} \\ \Delta\phi_{H,D} &= \pm \cos^{-1}\{(F^2_{H,D} + F^2_{H,R} - F^2_{H,N})/2F_{H,R}F_{H,D}\} \end{aligned} \tag{3.1.5}$$

Where $\Delta\phi_{H,N}$ and $\Delta\phi_{H,D}$ are the phase differences of the native and the derivative respectively.
In the case of OAS:

$$\phi'_H = \phi''_{H,A} \qquad \text{or} \qquad \phi'_H = \phi_{H,A} + \omega_H$$

where $\phi''_{H,A}$ is the phase of $F''_{H,A}$, $\phi_{H,A}$ is the phase of $F_{H,A}$, ω_H is the phase difference between $F''_{H,A}$ and $F_{H,A}$. If there is only one kind of anomalous scatterer in the unit cell, then $\omega_H = \pi/2$. We have for the OAS case (see Blundell & Johnson, 1976)

$$\Delta\phi_H = \pm \cos^{-1}\{(F_H^+ - F_H^-)/2F''_{H,A}\} \quad . \qquad (3.1.6)$$

Notice that ϕ_H now is the phase of $F_H = (F_H^+ + F_H^-)/2$.

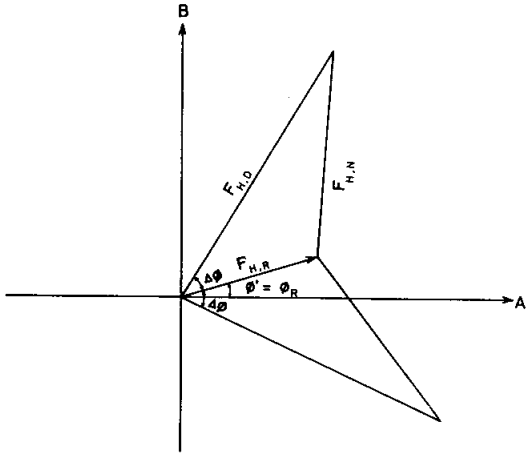

Figure 3.1
Enantiomorphous phase doublet in SIR case

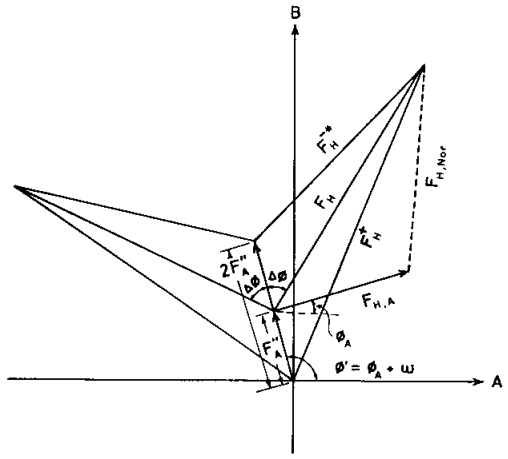

Figure 3.2
Enantiomorphous phase doublet in OAS case

3.2. Treatment of errors and the best starting phases in the presence of enantiomorphous phase ambiguity

The 'best phase relationship'

Following Blow & Crick (1959), we can introduce into direct methods the concepts of 'best phase' and 'figure of merit' for individual reflections. A best normalized structure factor is defined as

$$E_{H,best} = m_H |E_H| \exp(i\phi_{H,best}) \quad , \qquad (3.2.1)$$

where $\vec{m}_H = m_H \exp(i\phi_{H,best}) = \int \exp(i\phi_H) P(\phi_H) d\phi_H$.

A triplet phase relationship which consists of the best normalized structure factors is called the 'best' phase relationship'.

Expressions of the best phase and the figure of merit for a single reflection with enantiomorphous phase ambiguity

Defining $\Delta\phi_{H,best} = \phi_{H,best} - \phi'_H$, the $\Delta\phi_{H,best}$ and m_H can be expressed as follows (Fan, Han & Qian, 1984)

$$\tan(\Delta\phi_{H,best}) = 2\{P_+(\Delta\phi_H) - \tfrac{1}{2}\}\sin|\Delta\phi_H| / \cos\Delta\phi_H \quad (3.2.2)$$

$$m_H = \exp(-\sigma_H^2/2)\left[\{2(P_+ - \tfrac{1}{2})^2 + \tfrac{1}{2}\}(1-\cos2\Delta\phi_H)+\cos2\Delta\phi_H\right]^{\frac{1}{2}} \quad (3.2.3)$$

where $P_+(\Delta\phi_H)$ is the probability that $\Delta\phi_H$ has a positive sign, while σ_H is related to the experimental error and can be derived from the standard deviation D of the 'lack of closure error' (Blow & Crick, 1959). m_H may be regarded as a measure of reliability of $\Delta\phi_{H,best}$. As can be seen, there are three factors included in the expression of m_H :

$\exp(-\sigma_H^2/2)$	a measure of the sharpness of the experimental distribution of ϕ_H
$(P_+ - \frac{1}{2})^2$	a measure of the bias of $\Delta\phi_H$ towards positive or negative. It reaches the maximum when $P_+ = 0$ or 1
$\cos2\Delta\phi_H$	a measure of the closeness of the two possible phases, $\phi_H^+ = \phi'_H + \lvert\Delta\phi_H\rvert$ and $\phi_H^- = \phi'_H - \lvert\Delta\phi_H\rvert$. It reaches the maximum when $\Delta\phi_H = 0$ or π.

Either of the last two factors will have no effect on m_H when the other one reaches the maximum.

The best start

Substitute $P_+ = P_- = \frac{1}{2}$ into (3.2.2). It follows that

$$\Delta\phi_{H,best} = \begin{cases} 0 & , \text{ if } SIGN(\cos\Delta\phi_H) = 1 \\ \pi & , \text{ if } SIGN(\cos\Delta\phi_H) = -1 \end{cases}$$

or $$\exp(i\phi_{H,best}) = SIGN(\cos\Delta\phi_H)\exp(i\phi'_H) \quad (3.2.4)$$

Meanwhile, (3.2.3) reduces to

$$m_H = \exp(-\sigma_H^2/2)\lvert\cos\Delta\phi_H\rvert \quad . \quad (3.2.5)$$

Substituting (3.2.4) and (3.2.5) into (3.2.1), one obtains

$$E_{H,best} = \exp(-\sigma_H^2/2)\ \cos\Delta\phi_H |E_H| \exp(i\phi'_H) \qquad (3.2.6)$$

This is the 'best' normalized structure factor which could be obtained at the begining from the corresponding doublet. In other words, when enantiomorphous ambiguities are present, the best way to start a direct method process is to use the averaged value of the phase doublet as the starting phase and use the weight $\exp(-\sigma_H^2/2)|\cos\Delta\phi_H|$ for the corresponding E_H.

3.3. Incorporating the phase doublet information into direct method formulas

Use of modified Sayre equation

Denote the heavy atoms with known positions in the unit cell by p and the atoms of the unknown part by u. Then according to Fan (1965;1975) we have the so-called modified Sayre equation as

$$F_H = \frac{\Theta_{H,u}}{V} \sum_{H'} F_{H'} F_{H-H'} - \sum_p \left(\frac{\Theta_{H,u}}{\Theta_{H,p}} - 1\right) F_{H,p} \quad , \qquad (3.3.1)$$

where Θ is an atomic form factor. Replacing F_H by $|F_H|\exp(i\phi_H)$ and replacing ϕ_H by $\phi'_H + \Delta\phi_H$, (3.3.1) becomes

$$|F_H|\exp(i\Delta\phi_H) = \frac{\Theta_{H,u}}{V} \sum_{H'} |F_{H'} F_{H-H'}| \exp\{i(-\phi'_H+\phi'_{H'}+\phi'_{H-H'}+$$

$$+\Delta\phi_{H'}+\Delta\phi_{H-H'})\} - \sum_p \left(\frac{\Theta_{H,u}}{\Theta_{H,p}} - 1\right)|F_{H,p}|\exp\{i(\phi_{H,p}-\phi'_H)\} \ . \qquad (3.3.2)$$

Taking the imaginary part of (3.3.2) and denoting $-\phi'_H+\phi'_{H'}+\phi'_{H-H'}$ by Φ'_3 , we have

$$|F_H|\sin\Delta\phi_H = \frac{\Theta_{H,u}}{V} \{\sum_{H'} |F_{H'} F_{H-H'}| \sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})\}$$

$$- \sum_p \left(\frac{\Theta_{H,u}}{\Theta_{H,p}} - 1\right)|F_{H,p}|\sin(\phi_{H,p}-\phi'_H) \quad . \qquad (3.3.3)$$

Equation (3.3.3) can be used to refine the sign as well as the magnitude of $\Delta\phi_H$, once a large starting set of $\Delta\phi_H$ is available.

Use of the combination of Cochran's distribution and Sim's distribution

With the expression $\phi_H = \phi'_H + \Delta\phi_H$, the Cochran distribution (Cochran, 1955) can be modified to give (Fan et al., 1984):

$$P_{Cochran}(\Delta\phi_H) = \{2\pi I_o(\alpha')\}^{-1}\exp\{\alpha'\cos(\Delta\phi_H - \beta')\} \quad (3.3.4)$$

where $I_o(\alpha')$ is the modified Bessel function,

$$\alpha' = [\{\sum_{H'} \kappa_{HH'}\sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})\}^2+\{\sum_{H'} \kappa_{HH'}\cos(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})\}^2]^{1/2}$$

$$\tan\beta' = \sum_{H'} \kappa_{HH'}\sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'}) \;/\; \sum_{H'} \kappa_{HH'}\cos(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})$$

$$\Phi'_3 = -\phi'_H+\phi'_{H'}+\phi'_{H-H'} \quad , \quad \kappa_{HH'} = 2\sigma_3\sigma_2^{-3/2}|E_H E_{H'} E_{H-H'}| \quad .$$

In the same way, Sim's distribution (Sim, 1959) can be modified as

$$P_{Sim}(\Delta\phi_H) = \{2\pi I_o(x)\}^{-1}\exp\{x\cos(\Delta\phi_H-\delta_H)\} \quad , \quad (3.3.5)$$

where $x = 2|E_H E_{H,p}|/\{\sum_u Z_u^2/\sigma_2\}$, $\delta_H = \phi_{H,p} - \phi'_H$.

Combination of (3.3.4) and (3.3.5) gives the total probability distribution of $\Delta\phi_H$ (Fan & Gu, 1985):

$$P(\Delta\phi_H) = \{2\pi I_o(\alpha)\}^{-1}\exp\{\alpha\cos(\Delta\phi_H - \beta)\} \quad , \quad (3.3.6)$$

where

$$\alpha = [\{\sum_{H'} \kappa_{HH'}\sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})+x\sin\delta_H\}^2 + \sum_{H'} \kappa_{HH'}\cos(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})+x\cos\delta_H\}^2]^{1/2} \quad , \quad (3.3.7)$$

$$\tan\beta = \frac{\sum_{H'} \kappa_{HH'}\sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})+x\sin\delta_H}{\sum_{H'} \kappa_{HH'}\cos(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})+x\cos\delta_H} \quad . \quad (3.3.8)$$

Since $|\Delta\phi_H|$ is a known quantity when phase doublet information is available, the probability that $\Delta\phi_H$ has a positive sign can be derived from (3.3.6):

$$P_+(\Delta\phi_H) = \frac{1}{2} + \frac{1}{2}\tanh[\sin|\Delta\phi_H|\{\sum_{H'} \kappa_{HH'}\sin(\Phi'_3+\Delta\phi_{H'}+\Delta\phi_{H-H'})+x\sin\delta_H\}] . \quad (3.3.9)$$

With (3.3.9) the phase problem is now reduces to a sign problem.

On the other hand, by maximizing (3.3.6) we have $\Delta\phi_H = \beta$. Hence we can calculate the 'most probable' value of $\Delta\phi_H$ from (3.3.8). By replacing the $E_{H'}$ and $E_{H-H'}$ with their 'best' values, we can modify (3.3.8) and (3.3.9) to give

$$P_+(\Delta\phi_H) = \frac{1}{2} + \frac{1}{2}\tanh\left[\sin|\Delta\phi_H|\{\sum_{H'} m_{H'}, m_{H-H'}, \kappa_{HH'} \times \sin(\Phi_3' + \Delta\phi_{H'best} + \Delta\phi_{H-H'best}) + x\sin\delta_H\right] \quad (3.3.10)$$

and

$$\tan(\Delta\phi_H) = \frac{\sum_{H'} m_{H'}, m_{H-H'}, \kappa_{HH'}, \sin(\Phi_3' + \Delta\phi_{H'best} + \Delta\phi_{H-H'best}) + x\sin\delta_H}{\sum_{H'} m_{H'}, m_{H-H'}, \kappa_{HH'}, \cos(\Phi_3' + \Delta\phi_{H'best} + \Delta\phi_{H-H'best}) + x\cos\delta_H} \quad (3.3.11)$$

Formulas (3.2.2), (3.2.3), (3.3.10) and (3.3.11) can be used for ab initio phasing of the SIR or OAS data leading to unique estimates of individual phases.

Use of Hauptman's distribution

Hauptman (1982a,b) integrated the probabilistic theory of the three-phase structure invariants with SIR and OAS techniques leading to a series of formulas, which give unique estimates of the three-phase structure invariants. However, initial applications (Z.-B. Xu et al., 1984; Furey et al., 1986) showes that Hauptman's distribution tended to produce protein phases which were overly biased toward those of the heavy-atom substructure. Fortier (1985) incorporated the heavy-atom information into Hauptman's formulas resulting in the increase of accuracy of phase estimates. An alternative method was proposed recently by Hao (1986). He incorporated the phase doublet information into Hauptman's distribution leading to unique estimates of individual phases. Unlike Fortier's method, Hao makes use not only the magnitudes but also the phases of the heavy-atom substructure. On the other hand Hao's method is related to that of Fan et al. (1984) by replacing Cochran's distribution with Hauptman's. It is expected that Hauptman's distribution will give better result.

3.4. Direct phasing experimental OAS data of two known proteins

In a test calculation, the experimental OAS data of insulin and APP were directly phased by the procedure of Fan et al. (1984, 1985). Satisfactory results have been obtained.

Data

Insulin crystallizes in space proup R3 with a=82.5, c=34.0 Å, $\gamma=120^{\circ}$ and Z=9. There are ∿6400 independent reflections at 1.9 Å resolution. The data was kindly provided by Drs G. Dodson and E. Dodson. APP crystallizes in space group C2 with a=34.18, b=32.92, c=28.44 Å, $\beta=105.30^{\circ}$ and Z=4. There are ∿2100 independent reflections at 2.1 Å resolution. The data was kindly provided by Professor T. Blundell. 1000 largest E's and 60000 strongest $\sum_2$ relationships from each structure were used in the test calculation.

Test and results

Starting with $P_+ = \frac{1}{2}$, values of $\Delta\phi_{H,best}$ and m_H were calculated using (3.2.2) and (3.2.3) respectively and then substituted into (3.3.10) to calculate new values of P_+. Most of them differ from $\frac{1}{2}$ considerably. With the new set of P_+, one more cycle of iteration led to further improvement on the reliability. In order to examine the result statistically, the reflections were arranged in descending order of $|P_+ - \frac{1}{2}|$ and then cummulated into 5 groups, which contain the top 200, 400, 600, 800 and all reflections respectively. The average error in phase estimates and the percentage of reflections with the signs of $\Delta\phi_H$ correctly determined were calculated for each group. The results are listed in Table 3.1a and 3.2a for insulin and APP respectively. Multiple isomorphous replacement (MIR) phases of insulin and SIR-OAS phases of APP were also cummulated according to their figures of merit. The results are listed respectively in Tables 3.1b and 3.2b for comparision. In the case of insulin, the isomorphism of the sample crystals was not very good. The direct method phases are obviously better than those from MIR method,

at least for a thousand of reflections with large E values. On the other hand the isomorphism in APP was nearly perfect. The direct method phases are still comparable in quality with those of SIR-OAS method, so far as one half of reflections at 2.1 Å resolution are concerned. It can be concluded that the direct method under examine is very efficient in breaking the enantiomorphous phase ambiguity of protein OAS data. It can provide a very large number of reliable starting phases, which in turn can be the base of further phase development and refinement.

Table 3.1

Results on phasing the experimental data of insulin

(a) Direct phased OAS method

Group	%	Error
1	93.0	33°
2	94.5	35°
3	94.0	37°
4	93.5	38°
5	92.3	41°

(b) MIR method

Group	%	Error
1	70.0	41°
2	69.5	41°
3	63.5	44°
4	60.9	48°
5	58.8	54°

Table 3.2

Results on phasing the experimental data of APP

(a) Direct phased OAS method

Group	%	Error
1	95.0	24°
2	90.8	30°
3	91.3	29°
4	89.0	32°
5	86.3	36°

(b) SIR-OAS method

Group	%	Error
1	92.0	16°
2	88.8	20°
3	87.5	24°
4	84.1	30°
5	79.4	39°

4. IMAGE PROCESSING IN HIGH RESOLUTION ELECTRON MICROSCOPY

High resolution electron microscopy (HREM) is a powerful tool in the investigation of crystal structures, especially when the sample is not suitable for X-ray analysis. However in many cases ,high resolution electron micrographs without special processing can not reflect directly the true structure. On the other hand, direct methods are actually some kind of image processing technique, which may be of use in the processing of high resolution electron micrographs. Under the collaboration of the research group on crystal structure analysis and that on HREM in the Institute of Physics in Beijing, a new technique of image processing in HREM using direct methods has been proposed, which is a new junction of X-ray crystallography and electron microscopy. The procedure is devided into two parts, i.e. the image deconvolution and the resolution enhancement.

4.1. Image deconvolution

Let T(H) denotes the Fourier transform of the intensities of an electron micrograph (EM). Under the weak-phase-object approximation*, in which the dynamic diffraction effect is neglected, the structure factor F(H) is related to T(H) by the following formula:

$$F(H) = T(H) \;/\; 2\sigma \sin\chi_1(H) \cdot \exp\{-\chi_2(H)\} \quad . \qquad (4.1.1)$$

Here $\sigma = \pi/\lambda U$, λ is the electron wavelength and U the accelerating voltage. $\sin\chi_1(H)\cdot\exp\{-\chi_2(H)\}$ is the contrast transfer function. in which

$$\chi_1(H) = \pi \Delta f \lambda H^2 + \frac{1}{2}\pi C_s \lambda^3 H^4 \quad ,$$

$$\chi_2(H) = \frac{1}{2}\pi^2 \lambda^2 H^4 D^2 \quad .$$

---

* The applicability of the weak-phase-object approximation has been demonstrated by Unwin & Henderson (1975) for biological specimens and by Klug (1978/1979) for an inorganic compound.

Here Δf is the defocus value, C_s is the spherical aberration coefficient and D is the standard deviation of the Guassian distribution of defocus due to the chromatic aberration (Fijes, 1977). The task of image deconvolution is to find out F(H) from the corresponding T(H). For this purpose we have to know Δf, C_s and D in advance. Among these three factors, C_s and D can be determined experimentally without much difficulties. However it is usually very difficult to measure the value of Δf accurately. Hence in order to carry out image deconvolution, the main problem is to find out Δf. With the estimated values of C_s and D, we can calculate a set of F(H) from (4.1.1) for a given value of Δf. If this value is correct, the corresponding set of F(H) should obey the Sayre equation (Sayre, 1952)

$$F(H) = \frac{\Theta}{V} \sum_{H'} F(H')\, F(H-H') \quad . \qquad (4.1.2)$$

Hence the true Δf can be found by a systematic change of the trial Δf. The practical procedure should be as follows:

(1) Calculate a set of T(H) from an EM.

(2) Assign trial values of Δf in a wide range with a small interval, say 10 Å. For each trial Δf, a set of F(H) is calculated from T(H) using equation (4.1.1). Reflections with $|\sin\chi_1(H)\cdot\exp\{-\chi_2(H)\}| < 0.2$ will be neglected.

(3) Calculate the figure of merit S for each set of F(H) using the following formula (Debaerdemaeker, Tate & Woolfson, 1985)

$$S = \frac{\{\sum_H E^*(H) \sum_{H'} E(H')E(H-H')\}^2}{\sum_H |E(H)|^2 \sum_H |\sum_{H'} E(H')E(H-H')|^2} \quad .$$

S has a value between 0 and 1. The greater the value of S, the better the set of F(H) fits the Sayre equation.

(4) Find out the greatest S and then Fourier Transform the corresponding set of F(H) to deconvolute the image.

The procedure has been test by simulating calculations with a

series of theoretical EM of chlorinated phthalocyanine at 2 Å resolution taken under different Δf values. The other conditions are: Accelerating voltage = 500 kV; C_s = 1 mm; D = 150 Å . Some of the results are shown in Fig. 4.1. It can be seen that the deconvolution was quite successful. For further detail the reader is referred to Han, Fan, Zheng & Li (1986).

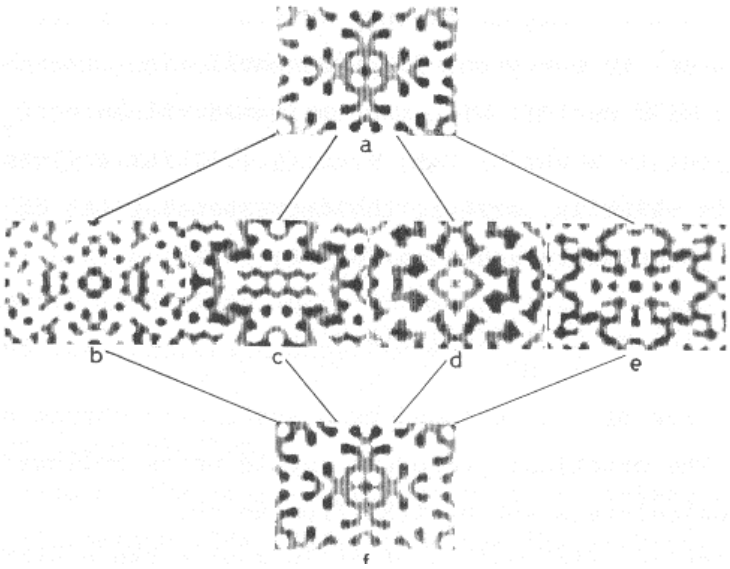

Figure 4.1

Deconvolution of theoretical EM's of copper chlorinated phthalocaynine

a — the expected image; b,c,d and e — theoretical EM's taken with Δf equal to -1000, +1000, -600 and +600 Å respectively; f — the deconvolution result.

4.2. Resolution enhancement

Phase extension

An electron diffraction (ED) pattern usually contains information up to 1 Å resolution, which is considerably higher than that which can be reached by an EM. In addition, the intensities of ED are independent of defocus and spherical aberration of the objective lens. Accordingly, under the weak-phase-object approximation a set of high resolution structure amplitudes of good quality can be obtained from an ED pattern. However, the struc-

ture analysis by ED alone is subject to the well known difficulty of the 'phase problem'. On the other hand, an EM after suitable deconvolution can provide phase information corresponding to about 2 Å resolution. This can greatly reduce the complexity of the solution of the phase problem. Hence an improved high-resolution image may be obtained by a phase extension procedure using the amplitudes of the structure factors from ED and starting phases from EM. Test of the procedure has been done using also the model structure of copper chlorinated phthalocaynine. The results are shown in Fig. 4.2, from which it can be seen that, while direct methods failed to solve the structure from the ED data alone, phase extension gave satisfactory result. For further detail see Fan, Zhong, Zheng & Li (1985).

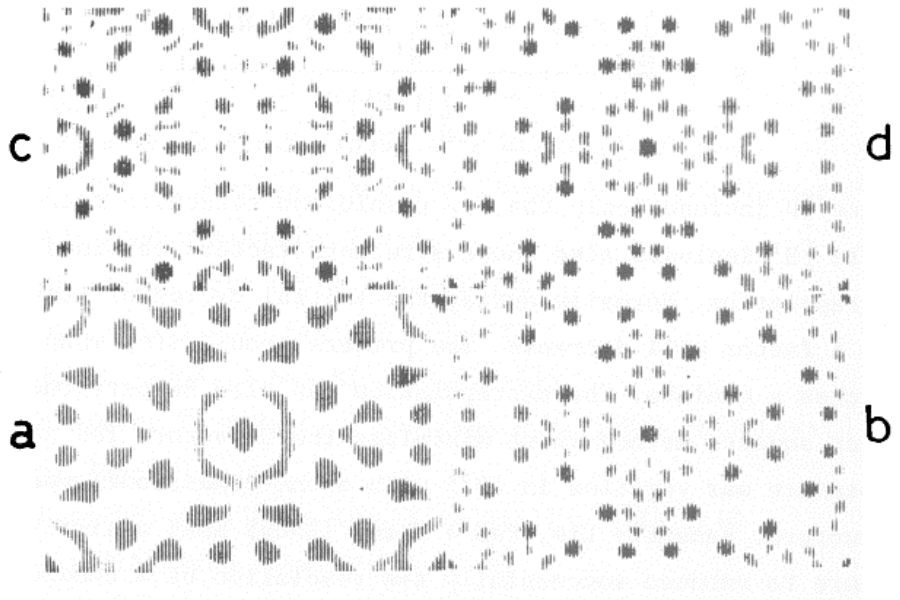

Figure 4.2

phase extension from 2 to 1 Å resolution

a — 2 Å EM after deconvolution; b — image resulted from phase extension using the 2 Å EM and the corresponding ED data; c — E-map from the direct method solution of ED data; d — the expected 1 Å image.

Structure factor extrapolation

While the technique of phase extension has been known for a long time, the extrapolation of structure factors on both magnitudes and phases is seldom used. Fan & Zheng (1975) pointed out that, even at very low resolution, there will be often enough reflections to set up simultaneous equations solving, at least in theory, the complete structure. Hence it is possible to extrapolate a set of low resolution structure factors to obtain that of high resolution ones. Sayre equation can be used for this purpose. Substituting a low resolution set of structure factors in the right hand side of Sayre equation (4.1.2), structure factors beyond the resolution limit can be obtained from the left hand side. This can be made iteratively to improve the result. For the indication of discrepancy, an R factor was defined as

$$R = \frac{\sum_{H} \left| F(H) - \frac{\Theta}{V} \sum_{H'} F(H')F(H-H') \right|}{\sum_{H} |F(H)|} , \qquad (4.2.1)$$

where $\sum_H$ includes only the low resolution structure factors, while $\sum_{H'}$ includes also those structure factors obtained from extrapolation. Normally during the initial cycles of iteration the R factor will decrease. The process should stop when R reaches a minimum. The extrapolation can also be performed using least squares method which minimizes the R factor. The above procedure was verified in 1975 with a one-dimensional model structure. Recently Liu, Fan & Zheng (1986) used a similar procedure to enhance successfully the resolution of a theoretical EM. The result is shown in Fig. 4.3.

In conclusion, direct methods have entered the field of HREM. The preliminary results were encouraging. We can expect that, direct methods will be as successful in HREM as they were in X-ray crystallography.

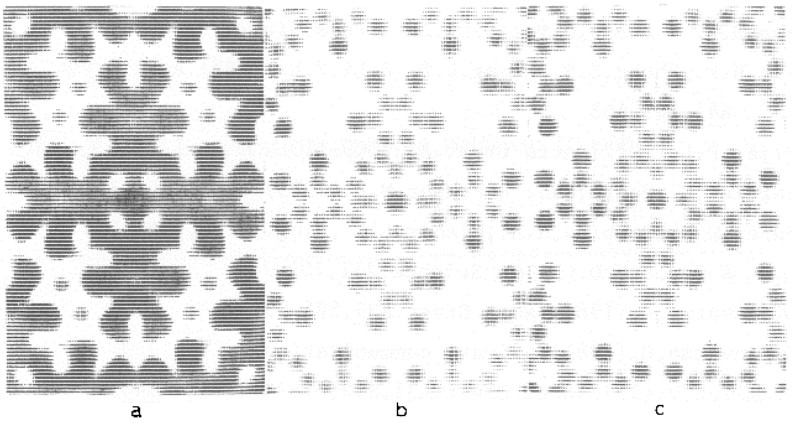

Figure 4.3

Structure factor extrapolation
of copper chlorinated phthalocaynine

a — 2 Å EM after deconvolution; b — resulted image of extrapolation from 2 - 1 Å resolution; c — expected image at 1 Å resolution.

REFERENCES

Blow,D.M. & Crick,F.H.C.(1959).Acta Cryst.12,794-802

Blundell,T.L. & Johnson,L.N.(1976)."Protein Crystallography" p177.London:Academic Press

Boehme,R.(1982).Acta Cryst.A38,318-326

Cochran,W.(1955).Acta Cryst.8,473-478

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

Debaerdemaeker,T.,Tate,C. & Woolfson,M.M.(1985).Acta Cryst. A41,286-290

Fan,H.F.(1965).Acta Phys.Sin.21,1105-1113; 1114-1118

Fan,H.F.(1975).Acta Phys.Sin.24,57-60

Fan,H.F.(1986).Rigaku Journal,3,25-30

Fan,H.F. & GU,Y.X.(1985).Acta Cryst.A41,280-284

Fan,H.F.,Han,F.S. & Qian,J.Z.(1984).Acta Cryst.A40,495-498

Fan,H.F.,Han,F.S.,Qian,J.Z. & Yao,J.X.(1984).Acta Cryst. A40,489-495

Fan,H.F.,He,L.,Qian,J.Z.,Liu,S.X.(19 78).Acta Phys.Sin.27,554-558

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

Fan,H.F. & Zheng,Q.T.(1975).Acta Phys.Sin.24,97-104

Fan,H.F.,Zhong,Z.Y.,Zheng,C.D. & Li,F.H.(1985).Acta Cryst. A41,163-165

Fijes,P.L.(1977).Acta Cryst.A33,109-113

Fortier,S.,Moore,N.J. & Fraser,M.E.(1985).Acta Cryst.A41,571-577

Furey,W.F.,Robbins,A.H.,Clancy,L.L.,Winge,D.R.,Wang,B.C. & Stout,C.D.(1986). Science,231,704-710

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

Giacovazzo,C.(1984).private communication

Gramlich,V.(1975)Acta Cryst.A31,S90;(1978).ibid.A34,S43

Gramlich,V.(1984).private communication

Han,F.S.,Fan,H.F.,Zheng,C.D. & Li,F.H.(1986).Acta Cryst.in press

Hao,Q.(1986).in this meeting

Hauptman,H.(1982a).Acta Cryst.A38,289-294;(1982b).ibid.632-641

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

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

Karle,J.(1984a).Acta Cryst.A40,4-11;(1984b).ibid.374-379

Klug,A.(1978/1979).Chem.Scr.14,245-256

Liu,Y.W.,Fan,H.F. & Zheng,C.D.(1986).in this meeting

Prick,P.A.J.,Beurskens,P.T. & Gould,R.O.(1983).Acta Cryst. A39,570-576

Qian,J.Z.,Fan,H.F. & Gu,Y.X.(1985).Acta Cryst.A41,476-478

Sayre,D.(1952).Acta Cryst.5,60-65

Sim,G.A.(1959).Acta Cryst.12,813-815

Unwin,P.N.T. & Henderson,R.(1975).J.Mol.Biol.94,425-440

Xu,Z.B. et al.(1984).in XIIIth IUCr Congress

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

Yao,J.X. & Fan,H.F.(1985).Acta Cryst.A41,284-285

Yao,J.X.,Zheng,C.D.,Qian,J.Z.,Han,F.S.,Gu,Y.X. & Fan,H.F.(1985) "SAPI-85: A Computer Program for Automatic Solution of Crystal Structures from X-ray Data" Institute of Physics, Academia Sinica, Beijing, China.