

Solving crystal structures from two-wavelength X-ray powder diffraction data – breaking the phase ambiguity in the noncentrosymmetric case

Y. X. Gu,^a Y. D. Liu,^a Q. Hao^{a,b} and H. F. Fan^{a*}

^aInstitute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, and ^bDepartment of Chemistry, De Montfort University, Leicester LE1 9BH, England. Correspondence e-mail: fan@aphy.iphy.ac.cn

Direct methods of breaking phase ambiguities in protein crystallography have been introduced in powder diffraction analysis. This is aiming at *ab initio* solution of noncentrosymmetric structures using two-wavelength anomalous powder diffraction data. The known structure of the hydrogen bromide salt of leotidine ($C_{14}H_{20}O_2N_2 \cdot HBr$) in space group $P2_12_12_1$ was used for simulating two-wavelength anomalous powder diffraction with the Br atom as anomalous scatterer. X-ray wavelengths are selected at $\lambda_1 = 0.920$ and $\lambda_2 = 1.500$ Å. Unique reflections from the diffraction pattern of λ_2 were able to locate the Br atom accurately. All overlapping diffraction peaks were uniformly partitioned to decompose into single reflections. Structure-factor amplitudes were then extracted. With these and the substructure of Br atoms, unique phases for centric reflections ($hk0$, $h0l$ and $0kl$) and phase doublets for noncentric reflections were obtained. The direct method was used to break the phase ambiguity leading to an interpretable electron-density map, from which five cycles of Fourier iteration yielded the complete structure.

© 2000 International Union of Crystallography
Printed in Great Britain – all rights reserved

1. Introduction

The anomalous scattering effect has been used for *ab initio* solution of crystal structures from X-ray powder diffraction (Prandl, 1990, 1994). Unlike that in single-crystal diffraction, a Bijvoet pair, $F(\mathbf{h})$ and $F(-\mathbf{h})$, will always overlap in powder diffraction patterns. Hence there will be no use of the imaginary part of anomalous scattering, f'' . But the real part of anomalous scattering f' can be used to simulate isomorphous replacement data. This means powder diffraction data from two different wavelengths can be treated as those from the native crystal and its isomorphous heavy-atom derivative. For centrosymmetric crystals, diffraction data from two suitable wavelengths will be sufficient to solve the phase problem. However, for noncentrosymmetric crystals, phase ambiguity will occur with reflections not in the centric zone. Although in theory it is possible to solve the problem when the sample contains two different kinds of suitable anomalous scatterers and if three different wavelengths are used, this is not quite feasible in practice. Burger & Prandl (1999) proposed a method to overcome this difficulty. They used the ambiguous but restricted phases as a constraint condition in an iterative maximum-entropy-method calculation, which eventually revealed the test structure. In the present paper, an alternative approach is proposed, in which the phase ambiguity is broken directly leading to an interpretable electron-density map.

2. Simulating structure-factor amplitudes

The known structure of the hydrogen bromide salt of leotidine ($C_{14}H_{20}O_2N_2 \cdot HBr$) was used to simulate two-wavelength synchrotron X-ray powder diffraction. The model structure has a unit cell of $a = 10.75$, $b = 7.91$ and $c = 16.59$ Å in the space group $P2_12_12_1$. There is only one molecule in the asymmetric unit, including 1 Br atom and 18 non-hydrogen light atoms. The structure was solved originally by the Patterson method with single-crystal X-ray diffraction data (Fan & Li, 1965). The two wavelengths used in the present simulation are $\lambda_1 = 0.920$ and $\lambda_2 = 1.500$ Å. The corresponding anomalous corrections (in electrons) for the Br atom are $f'_1 = -8.539$, $f''_1 = 3.822$ and $f'_2 = -0.816$, $f''_2 = 1.221$. Structure factors were calculated within the range $10 < 2\theta < 125^\circ$ for the two wavelengths using atomic coordinates obtained from the single-crystal analysis. The half-height width of a single diffraction peak is assumed to be $\Delta\theta = 0.02^\circ$. Reflections far from their nearest neighbors with $\Delta\theta > 0.02^\circ$ are assumed to be unique reflections. Others are regarded as overlapping. There are 285 unique reflections out of the total of 4820 at the wavelength 0.920 Å and 558 unique reflections out of the total of 1416 at the wavelength 1.500 Å. When extracting structure-factor amplitudes from a multiple diffraction peak, the total intensity is uniformly partitioned to give intensities of the component reflections. In this way, two

sets of powder diffraction data were simulated yielding two sets of structure-factor amplitudes; one corresponds to the wavelength of 0.920 Å and the other corresponds to that of 1.500 Å. We assumed that experimental errors due to preferred orientation, absorption *etc.* have been treated properly with available techniques. Hence they were not considered in the simulation.

3. Locating the heavy atom

With the program *SAPI* (Fan *et al.*, 1991), it was straightforward to locate the Br atom using only the unique reflections from the data set at the 1.500 Å wavelength. The resulting positional parameters are $x = 0.9014$, $y = 0.0614$ and $z = 0.5997$ in comparison with the true values $x = 0.9021$, $y = 0.0624$ and $z = 0.6004$.

4. Breaking the phase ambiguity

Structure factors corresponding to the two wavelengths are expressed as

$$F_1(\mathbf{h}) = \sum_{j=1}^N f_j^0 \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) + \sum_{j=1}^{N_{\text{Br}}} (f'_{\lambda_1} + if''_{\lambda_1}) \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) \quad (1)$$

$$F_2(\mathbf{h}) = \sum_{j=1}^N f_j^0 \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) + \sum_{j=1}^{N_{\text{Br}}} (f'_{\lambda_2} + if''_{\lambda_2}) \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j), \quad (2)$$

where N is the total number of atoms in the unit cell, while N_{Br} is that of Br atoms. Subtracting (1) from (2), we obtain

$$F_2(\mathbf{h}) - F_1(\mathbf{h}) = \Delta F(\mathbf{h}) = \sum_{j=1}^{N_{\text{Br}}} [(f'_{\lambda_2} - f'_{\lambda_1}) + i(f''_{\lambda_2} - f''_{\lambda_1})] \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j). \quad (3)$$

Since we have located the Br atom, $\Delta F(\mathbf{h})$ can be calculated from (3). However, since we know only the amplitude and not the phase of both $F_1(\mathbf{h})$ and $F_2(\mathbf{h})$, there will be two ways to draw the triangle defined by $F_1(\mathbf{h})$, $F_2(\mathbf{h})$ and $\Delta F(\mathbf{h})$. This leads to the phase ambiguity as shown in Fig. 1. The phase ambiguity can be expressed as

$$\varphi = \varphi' \pm |\Delta\varphi|, \quad (4)$$

where φ' is the phase of ΔF . According to Fig. 1, we have

$$\Delta\varphi_1 = \pm \cos^{-1}(F_2^2 - F_1^2 - \Delta F^2 / 2F_1\Delta F) \quad (5)$$

and

$$\Delta\varphi_2 = \pm \cos^{-1}(F_2^2 - F_1^2 + \Delta F^2 / 2F_2\Delta F). \quad (6)$$

In order to break the phase ambiguity, we need to determine the sign of $\Delta\varphi$. This is similar to the case of single isomorphous replacement in protein crystallography (see Woolfson & Fan, 1995). A direct method has been developed for finding the

Table 1

Results on breaking the phase ambiguity.

Cycle	Average phase error (°)	F_{obs} -weighted average phase error (°)
1	60.15	51.39
2	37.96	30.18
3	37.18	29.47
Background phase error	34.77	

sign of $\Delta\varphi$ (Fan & Gu, 1985). The method is implemented by the program *OASIS* (Hao *et al.*, 2000). The main point of the method is to calculate the probability for $\Delta\varphi$ to be positive:

$$P_+(\Delta\varphi_{\mathbf{h}}) = \frac{1}{2} + \frac{1}{2} \tanh \left\{ \sin|\Delta\varphi_{\mathbf{h}}| \left[\sum_{\mathbf{h}'} m_{\mathbf{h}'} m_{\mathbf{h}-\mathbf{h}'} \kappa_{\mathbf{h},\mathbf{h}'} \times \sin(\Phi_3' + \Delta\varphi_{\mathbf{h}',\text{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}',\text{best}}) + \chi \sin \delta_{\mathbf{h}} \right] \right\}. \quad (7)$$

For the theory behind and the procedure to use (7), the reader is referred to Fan *et al.* (1984), Fan & Gu (1985) and Fan *et al.* (1990). In the present test, from the two sets of structure-factor amplitudes and the coordinates of the Br atom, a set of $|\Delta\varphi_2|$ and a set of ΔF were calculated. The signs of $\Delta\varphi_2$ and the associated figures of merit were derived by an iterative calculation based on (7). Results are summarized in Table 1. Ten cycles of iteration were performed. The results converged after the third cycle. Phase errors were calculated for all 1416 reflections within the range $10 < 2\theta < 125^\circ$. The background error was calculated assuming the correct signs of $\Delta\varphi_2$. Hence

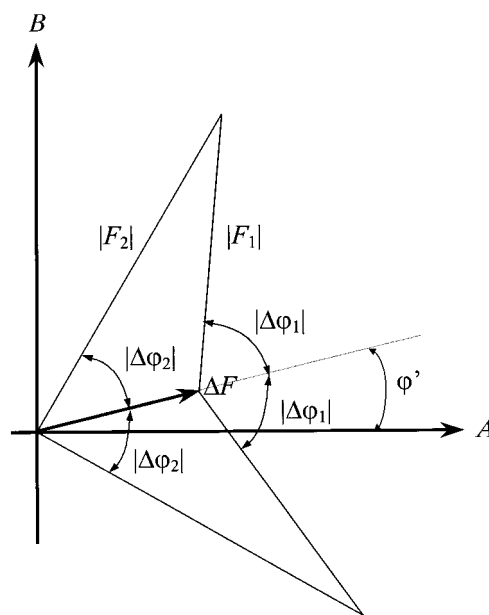


Figure 1

Phase ambiguity from two-wavelength powder diffraction. $|F_1|$ is the structure-factor amplitude corresponding to λ_1 . $|F_2|$ is the structure-factor amplitude corresponding to λ_2 . ΔF is the structure-factor difference between λ_2 and λ_1 for Br atoms. φ' is the phase angle of ΔF . $\Delta\varphi_1$ is the phase difference between F_1 and ΔF . $\Delta\varphi_2$ is the phase difference between F_2 and ΔF .

the error is mainly due to the uniform decomposition of overlapping reflections, which brought errors into the values of $|\Delta\varphi_2|$. It is seen from Table 1 that the iteration converged rapidly and the phase ambiguity was broken effectively. It is also seen that the F_{obs} -weighted errors are smaller than the corresponding unweighted ones, which indicates that phases of stronger reflections have been estimated better. A Fourier map was calculated using phases after resolving the ambiguity. In addition to the Br atom, a structure fragment (see Fig. 2*a*) was obtained by a default automatic interpretation with the program *SAPI* (Fan *et al.*, 1991). The fragment contains 13 of the total 18 light atoms and one ghost peak (marked with an X in the figure).

5. Completing the structure

Starting from the fragment, five cycles of the Fourier iteration led to the complete structure (Fig. 2*c*). The atomic types were identified according to the peak heights and the chemical contents of the unit cell. *R* factors in the Fourier recycling are listed in Table 2. They are relatively large. This is due to the uniform decomposition of the overlapping reflections.

6. Comparison with the heavy-atom method

It is reasonable to expect that the sample structure used in this paper could also be solved by the heavy-atom method using

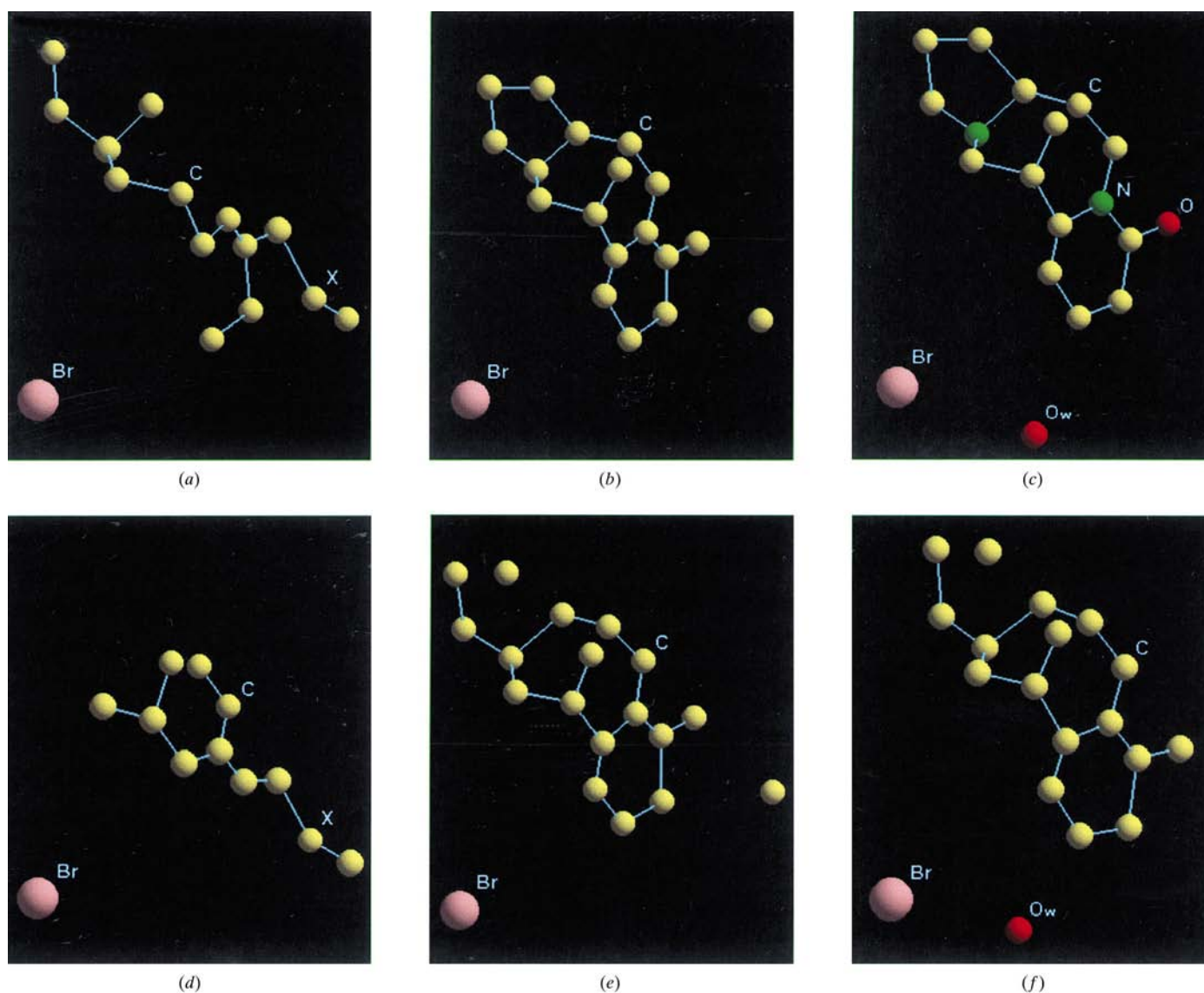


Figure 2

Results of Fourier recycling based on phases from the two-wavelength direct method (*a*, *b* and *c*) and the heavy-atom method (*d*, *e* and *f*). (*a*) Structure fragment obtained from automatic interpretation of the Fourier map phased by the two-wavelength direct method. (*b*) Structure fragment obtained after two cycles of Fourier iteration based on (*a*). (*c*) The complete structure obtained after five cycles of Fourier iteration based on (*a*). (*d*) Structure fragment obtained from automatic interpretation of the Fourier map phased by the heavy-atom method using the 1.500 Å wavelength data alone. (*e*) Structure fragment obtained after four cycles of Fourier iteration based on (*d*), the automatic interpretation failed to include one of the chemically bonded atoms (at the top of the figure) into the whole organic molecule. (*f*) Structure fragment obtained after five cycles of Fourier iteration based on (*d*), the automatic interpretation still failed to include one of the chemically bonded atoms (at the top of the figure) into the organic molecule.

Table 2*R* factors in Fourier recycling.

Cycle	<i>R</i> factor
0	29.20
1	24.99
2	23.93
3	23.16
4	22.30
5	22.09

the 1.500 Å wavelength data alone. While this is true, a comparison between the heavy-atom method and the two-wavelength direct method showed that the latter is superior. In the comparison, we started with the same heavy-atom site using the same diffraction data and the same strategy of Fourier recycling. The two starting Fourier maps are phased respectively by the two-wavelength direct method and the heavy-atom method. Automatic interpretation of these two maps with the program *SAPI* (Fan *et al.*, 1991) led respectively to Figs. 2(*a*) and (*d*). Fig. 2(*a*) is obviously better than Fig. 2(*d*). Furthermore, two cycles of Fourier iteration resulted in Fig. 2(*b*), revealing all the independent atoms in the unit cell. On the other hand, to obtain from Fig. 2(*d*) a similar but a little worse result (Fig. 2(*e*)), it needed four rather than two cycles of Fourier iteration. Finally, as described above, with the two-wavelength direct method, the Fourier recycling converged to the complete structure in five cycles (Fig. 2(*c*)). However, with the heavy-atom method, the convergence is much slower and incomplete (see the result of the fifth cycle in Fig. 2(*f*)). In short, while the heavy-atom method can solve the sample structure, it needs more cycles of iteration and more experience and skill from the user. It is reasonable to expect that the two-wave-

length direct method can deal with some structures that the heavy-atom method would fail to solve.

7. Conclusions

The direct method of breaking phase ambiguities in protein crystallography has been introduced into X-ray powder diffraction analysis. This leads to a straightforward approach of using two-wavelength anomalous powder diffraction data in solving complicated noncentrosymmetric crystal structures.

This project is supported by the NSF of China and by NKBRSE, People's Republic of China: No. G1999075604. QH is indebted to the Royal Society for supporting his visit to China. The authors would like to thank the referee whose comments improved this paper.

References

- Burger, K. & Prandl, W. (1999). *Acta Cryst.* **A55**, 719–728.
 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., Hao, Q., Gu, Y. X., Qian, J. Z., Zheng, C. D. & Ke, H. (1990). *Acta Cryst.* **A46**, 935–939.
 Fan, H. F. & Li, P. F. (1965). Unpublished results.
 Fan, H. F., Yao, J. X., Zheng, C. D., Gu, Y. X. & 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, People's Republic of China.
 Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
 Prandl, W. (1990). *Acta Cryst.* **A46**, 988–992.
 Prandl, W. (1994). *Acta Cryst.* **A50**, 52–55.
 Woolfson, M. M. & Fan, H. F. (1995). *Physical and Non-physical Methods of Solving Crystal Structures*, pp. 125–127, 157–166. Cambridge University Press.