A Sponsored Supplement to ***Science***, pp. 42-43 (2018)

Phase problem in X-ray crystallography and electron microscopy

Hai-Fu Fan*

Since 1980, a research group led by Hai-Fu Fan at the Institute of Physics (IOP) has been dedicated to the "phase problem" in X-ray crystallography and electron microscopy (EM) (http://cryst.iphy.ac.cn). The group collaborated for two papers published in 1965, both of which focus on "direct methods" in crystallography (*1*, *2*). The first paper (*1*) extended the Sayre equation (*3*) by considering the existence of heavy atoms with known positions. A subsequent systematic study on pseudosymmetries and phase ambiguities (*4*) led to the development of the program SAPI (structure analysis programs with intelligent control—or, read backwards, "Institute of Physics, Academia Sinica") (*5*–*7*), which is the first direct-method computer program that can detect and solve superstructures. SAPI was adopted by the Rigaku Corporation (Japan) and the Molecular Structure Corporation (United States), from 1986 until the early 1990s, as the main structure-solving program for small crystal structures. The second paper (*2*) proposed a method to break the phase ambiguity intrinsic in SAD (single-wavelength anomalous diffraction) or SIR (single isomorphous replacement) experiments, marking one of the earliest attempts to introduce direct methods to protein crystallography. The results of the research have been successfully applied in numerous disciplines, including materials and life sciences. Since its inception, the group has focused on exploring new applications for direct methods outside of their traditional fields.

1. Direct methods as a tool of image processing in EM

A two-stage image processing procedure that combines information from an EM image and the corresponding electron diffraction (ED) pattern has been proposed (*8*). Tests using the EM image of the high-critical-temperature (Tc) superconductor Bi-2212 showed that the resolution of the original EM image is enhanced by a factor of two (*9*), as shown in Figure 1. Hopefully, this technique may also be useful in improving the resolution of cryo-EM images. The program package VEC (visual computing in electron crystallography) was written for the automation of two-stage EM image processing (*10*).

2. Incommensurate modulated structures and composite structures solved by multidimensional direct methods

Multidimensional direct methods (MDDM) were proposed for the determination of incommensurate crystal structures without relying on a preassumed model (*11*, *12*). The application of MDDM to the *0klm* ED pattern of the high-Tc superconductor Bi-2223 revealed incommensurate modulation, which was reported by Zhong-Xian Zhao during the Nobel Jubilee Symposium of 1991 (*13*). The program DIMS (direct methods in multidimensional space) was written for the implementation of these methods (*14*), and it was later incorporated in the VEC package (*15*, *16*).

3. Combining direct methods with protein crystallographic methods

The P_+ probability formula (*17*) was derived by

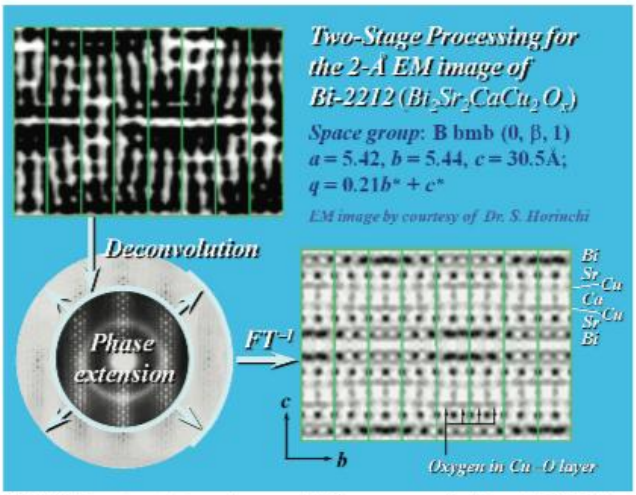

FIGURE 1. Two-stage electron microscopy (EM) image processing. Upper left: 2-Å EM image of the high-critical-temperature (Tc) superconductor Bi-2212. **Lower left:** 1-Å electron diffraction pattern plus 2-Å diffraction phases; the latter resulted from stage 1 (deconvolution). **Lower right:** 1-Å structure image, the inverse Fourier transform (FT) of the result from stage 2 (phase extension).

Beijing National Laboratory for Condensed Matter Physics and Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China.
*Corresponding author: fan@iphy.ac.cn

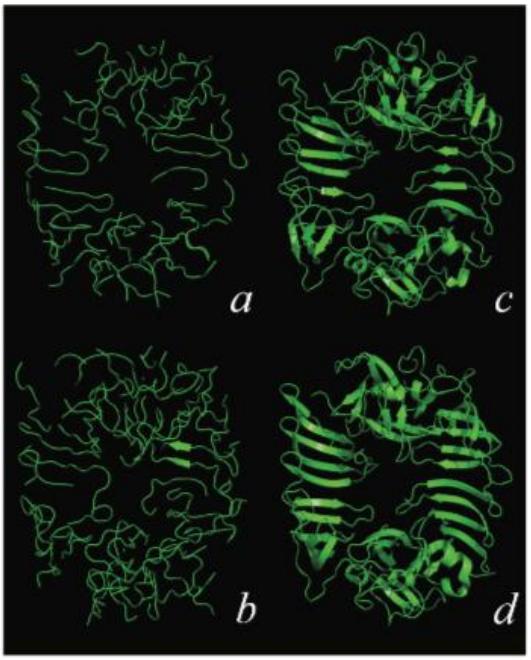

FIGURE 2. Direct-method-aided model completion. (*a*) The starting model obtained by PHENIX.AutoBuild with 5.3-Å 1h3i (PDB code) MAD data. (*b*) Result obtained by IPCAS running in Model Extension mode with 2.8-Å 1h3i SAD (at high-remote wavelength) data. (*c*) Result obtained by IPCAS running in Phase/Model Extension mode with 2.8-Å 1h3i SAD high-remote data. (*d*) The final 1h3i model. IPCAS, Iterative Protein Crystal structure Automatic Solution; MAD, multiwavelength anomalous diffraction; PDB, Protein Data Bank; SAD, single-wavelength anomalous diffraction. (Models in this figure were plotted by PyMOL in the cartoon mode.)

combining the Cochran distribution in the direct methods, the bimodal distribution of SAD and SIR experiments in protein crystallography, and the Sim distribution in small-molecule crystallography. The program OASIS (one-wavelength anomalous scattering and single isomorphous substitution) (*18*) was written based on the P_+ formula. OASIS has been adopted by CCP4 (Collaborative Computational Project No. 4, United Kingdom) since 2000 as the only direct-method program capable of breaking the SAD/SIR phase ambiguity for diffraction data far below the 1.2-Å atomic resolution. The pipeline IPCAS (Iterative Protein Crystal structure Automatic Solution) (*19*) has been developed, in which OASIS and several world-renowned programs are involved. Two features of IPCAS are iterative direct-method SAD phasing and direct-method-aided model completion (*20, 21*). An example of the latter is shown in Figure 2, where a set of 5.3-Å MAD (multiwavelength anomalous diffraction) phases has been successfully extended to a set of 2.8-Å SAD (at high-remote wavelength) data, leading to more than 90% of the complete structure. With this technique, a set of cryo-EM phases at about 6 Å resolution could readily be extended to a set of X-ray native protein data at a resolution higher than 3 Å.

References

1. H. F. Fan, *Acta Phys. Sin.* **21**, 1105–1114 (1965); in Chinese; English translation available in *Chinese Phys.* 1418–1428 (1965).
2. H. F. Fan, *Acta Phys. Sin.* **21**, 1114–1118 (1965); in Chinese; English translation available in *Chinese Phys.* 1429–1435 (1965).
3. D. Sayre, *Acta Cryst.* **5**, 60–65 (1952).
4. H. F. Fan, *Rigaku Journal* **1**, 15–21 (1984).
5. H. F. Fan, J. X. Yao, J. Z. Qian, *Acta Cryst.* **A44**, 688–691 (1988).
6. H. F. Fan *et al.*, *Acta Cryst.* **A44**, 691–692 (1988).
7. H. F. Fan *et al.*, *Acta Cryst.* **A46**, 99–103 (1990).
8. H. F. Fan, F. H. Li, in *Direct Methods, Macromolecular Crystallography and Crystallographic Statistics*, Proceedings of the IUCr Winter School, Madras, India, December 9–19, 1985, H. Schenk, A. J. C. Wilson, and S. Pathasarathy, Eds. (World Scientific, Singapore, 1987), pp. 400–409.
9. Z. Q. Fu *et al.*, *Ultramicroscopy* **54**, 229–236 (1994).
10. Z. H. Wan *et al.*, *Z. Krist.* **218**, 308–315 (2003).
11. Q. Hao, Y. W. Liu, H. F. Fan, *Acta Cryst.* **A43**, 820–824 (1987).
12. H. F. Fan, S. van Smaalen, E. J. W. Lam, P. T. Beurskens, *Acta Cryst.* **A49**, 704–708 (1993).
13. Y. D. Mo *et al.*, in *Low Dimensional Properties of Solids*, Proceedings of the Nobel Jubilee Symposium, Gothenburg, Sweden, December 4–7, 1991, M. Jonson, T. Claeson, Eds., Physica Scripta, The Royal Swedish Academy of Sciences (World Scientific, Singapore, 1992), pp. 18–19.
14. Z. Q. Fu, H. F. Fan, *J. Appl. Cryst.* **27**, 124–127 (1994).
15. H. F. Fan, *CompComm Newsletter* **5**, 16–23 (2005).
16. H. F. Fan, *CompComm Newsletter* **5**, 24–27 (2005).
17. H. F. Fan, Y. X. Gu, *Acta Cryst.* **A41**, 280–284 (1985).
18. Q. Hao, Y. X. Gu, C. D. Zheng, H. F. Fan, *J. Appl. Cryst.* **33**, 980–981 (2000).
19. T. Zhang *et al.*, IPCAS 1.0, Institute of Physics, Chinese Academy of Sciences, People's Republic of China (2012).
20. D. Q. Yao *et al.*, *Acta Cryst.* **D70**, 2686–2691 (2014).
21. H. F. Fan *et al.*, *Acta Cryst.* **A70**, 239–247 (2014).