

De novo phasing with optimized XFEL data

Quan Hao*

School of Biomedical Sciences, L04-48 Laboratory Block, University of Hong Kong, 21 Sassoon Road, Hong Kong.

*Correspondence e-mail: qhao@hku.hk

Keywords: serial crystallography; XFEL; SAD phasing.

The unique capabilities of X-ray free-electron lasers (XFELs) have opened up new opportunities to collect damage-free diffraction data from randomly oriented sub-micron-sized crystals using serial femtosecond crystallography (SFX) (Chapman *et al.*, 2011). SFX data have been shown to be of sufficient quality to enable structure determination with the molecular replacement method (Kang *et al.*, 2015). However, *de novo* phasing with the SFX data is difficult because the fluctuations of intensity measurements may overwhelm the anomalous scattering signal owing to variation of crystal size, uneven photon spectral distribution, partial Bragg reflection and partial intersection of the crystal in the X-ray beam. In 2014, Schlichting and co-workers (Barends *et al.*, 2014) successfully utilized the strong anomalous signal of gadolinium to phase the single-wavelength anomalous diffraction (SAD) data collected at the Linac Coherent Light Source (LCLS) from crystals of a gadolinium derivative of lysozyme. In reality, however, finding a suitable derivative is often difficult. Therefore, use of endogenous atoms, such as sulfur, as a source of anomalous scattering would be a much more attractive approach. Nakane *et al.* (2015) demonstrated the native sulfur phasing of lysozyme from SFX data collected at SACLA.

In this issue of **IUCrJ**, Schlichting and co-workers (Nass *et al.*, 2016) report an optimized method to tackle the particular challenges of native sulfur phasing of SFX data collected at XFELs. They propose to (1) optimize the geometry (metrology) of the multi-tiled detector by translating and rotating individual tiles of the detector to minimize the distance between the predicted and observed spot locations, (2) refine the sample-to-detector distance by manually adapting the distance until the distributions of unit-cell lengths resemble single-Gaussian-like distributions, and (3) scale the integrated intensities from each diffraction image. They first tested their approach with the gadolinium-derivative lysozyme data and showed that the optimization improved the anomalous signal significantly: fully automated building of the structure with only 10 000 indexed images instead of the 60 000 images that were required in the original report (Barends *et al.*, 2014).

They then applied the same methodology to a thaumatin data set collected at the LCLS using the native sulfur as a source of anomalous scattering (Fig. 1). A photon energy of 6 keV was chosen to balance the quest for a high anomalous signal of the

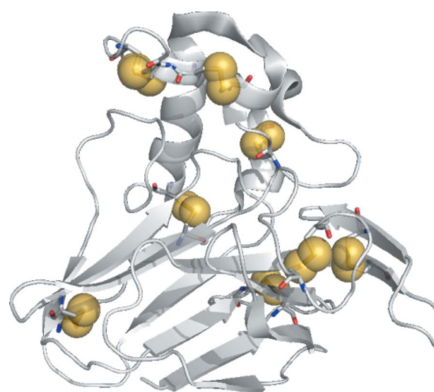


Figure 1

Single-wavelength anomalous diffraction from the sulfur atoms (colored in yellow) in native thaumatin was used to phase SFX data.

relatively light atoms (sulfur) and thus low photon energy, and the limitations set by beamline transmission, detector quantum efficiency, the resolution of the diffraction data and photon absorption. They were able to phase SFX data using the weak anomalous signal of sulfur and automatically build an almost complete molecular model using 125 000 indexed patterns of thaumatin microcrystals.

X-ray free-electron lasers have provided a new way to obtain SFX data that is advantageous over traditional synchrotron sources in terms of better data resolution, smaller sample size and minimal radiation damage (Kang *et al.*, 2015). However, *de novo* phasing of SFX data, particularly with only native crystals, is still far from trivial. Millions of crystals and many hours of beamtime are usually required to collect a sufficient amount of data for structure determination. Nass *et al.*'s study has demonstrated that with optimized data processing technology, single-wavelength anomalous diffraction data from the native sulfur atoms can allow protein structures to be solved *ab initio*. Improvements in XFEL instrumentation and processing software should further reduce the number of images required to phase SFX data.

Acknowledgements

Lingxiao Zeng is thanked for help with preparing the figure.

References

- Barends, T. R. M., Foucar, L., Botha, S., Doak, R. B., Shoeman, R. L., Nass, K., Koglin, J. E., Williams, G. J., Boutet, S., Messerschmidt, M. & Schlichting, I. (2014). *Nature*, **505**, 244–247.
- Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmeß, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Hömke, A., Reich, C., Pietschner, D., Strüder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kühnel, K., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Røcker, A., Jönsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schröter, C., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B. & Spence, J. C. H. (2011). *Nature*, **470**, 73–77.
- Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., de Waal, P. W., Ke, J., Tan, M. H. E., Zhang, C., Moeller, A., West, G. M., Pascal, B. D., Van Eps, N., Caro, L. N., Vishnivetskiy, S. A., Lee, R. J., Suino-Powell, K. M., Gu, X., Pal, K., Ma, J., Zhi, X., Boutet, S., Williams, G. J., Messerschmidt, M., Gati, C., Zatspein, N. A., Wang, D., James, D., Basu, S., Roy-Chowdhury, S., Conrad, C. E., Coe, J., Liu, H., Lisova, S., Kupitz, C., Grotjohann, I., Fromme, R., Jiang, Y., Tan, M., Yang, H., Li, J., Wang, M., Zheng, Z., Li, D., Howe, N., Zhao, Y., Standfuss, J., Diederichs, K., Dong, Y., Potter, C. S., Carragher, B., Caffrey, M., Jiang, H., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Ernst, O. P., Katritch, V., Gurevich, V. V., Griffin, P. R., Hubbell, W. L., Stevens, R. C., Cherezov, V., Melcher, K. & Xu, H. E. (2015). *Nature*, **523**, 561–567.
- Nakane, T., Song, C., Suzuki, M., Nango, E., Kobayashi, J., Masuda, T., Inoue, S., Mizohata, E., Nakatsu, T., Tanaka, T., Tanaka, R., Shimamura, T., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Nureki, O., Iwata, S. & Sugahara, M. (2015). *Acta Cryst. D* **71**, 2519–2525.
- Nass, K., Meinhart, A., Barends, T. R. M., Foucar, L., Gorel, A., Aquila, A., Botha, S., Doak, R. B., Koglin, J., Liang, M., Shoeman, R. L., Williams, G., Boutet, S. & Schlichting, I. (2016). *IUCrJ*, **3**, 180–191.