Experimental phasing of serial femtosecond crystallography data

Ilme Schlichting*

Max Planck Institute for Medical Research, Heidelberg, Germany.
*Correspondence e-mail: ilme.schlichting@mpimf-heidelberg.mpg.de

X-ray free-electron lasers (XFELs) provide ultra-bright femtosecond X-ray pulses that are intense enough to enable data collection of small weakly scattering objects and short enough to outrun most radiation damage effects. These beam features allow not only to study highly radiation-sensitive systems such as metalloproteins and tiny crystals but also to capture fleeting reaction intermediates in time-resolved studies. They therefore expand the structural biologist's toolbox by a great deal (Schlichting, 2015; Spence, 2017).

De novo phasing of XFEL data, however, remains problematic, for several reasons. First, there is the stochastic nature of the XFEL beam parameters (pulse and photon energy, including the spectral distribution of the latter, pointing, beam profile, *etc.*). Second, there is the serial femtosecond crystallography (SFX) data collection scheme that only once probes randomly oriented microcrystals intersecting the XFEL beam and gives a still image. Thus, it had long been doubted that the resulting integrated Bragg intensities are accurate enough for *de novo* phasing to be successful. A number of XFEL-specific phasing approaches have been suggested (Spence *et al.*, 2011; Son *et al.*, 2011; Ayyer *et al.*, 2016), but so far none have demonstrated success in *de novo* phasing of experimental data.

Nevertheless, indications that phasing might indeed work came from the observation of the weak anomalous signal of sulfur (Barends *et al.*, 2013), which was soon afterwards demonstrated by the first *de novo* phasing of lysozyme using the strong anomalous signal of bound gadolinium ions (Barends *et al.*, 2014). This was followed by using mercury for single isomorphous replacement with anomalous scattering (SIRAS) (Yamashita *et al.*, 2015, 2017), iodine substitutions in a detergent molecule to phase SFX data of the model membrane protein bacteriorhodopsin (Nakane *et al.*, 2016), exploiting the anomalous signal of endogenous sulfur atoms (Nakane *et al.*, 2015; Nass *et al.*, 2016; Batyuk *et al.*, 2016) or metal centres (Fukuda *et al.*, 2016) and inserting selenomethionine in recombinantly expressed protein (Yamashita *et al.*, 2017; Hunter *et al.*, 2016). All these studies used model systems that diffract strongly to high resolution and relied on phasing approaches (SAD, SIR, SIRAS) that are commonly used at synchrotron sources. A novel approach, again using a well established model system, was demonstrated recently by exploiting a dual colour mode of the XFEL source SACLA delivering two XFEL pulses widely separated in photon energy simultaneously (Hara *et al.*, 2013), thereby enabling efficient SFX data collection for multi-wavelength anomalous dispersion (MAD) phasing while saving sample and beam time (Gorel *et al.*, 2017).

Given all this progress it would almost seem as if XFEL *de novo* phasing were coming of age. However, so far only one previously unknown structure has been revealed by *de novo* phasing of XFEL data. The structure of the mosquito larvicide BinAB was solved by multiple isomorphous replacement with anomalous scattering (MIRAS) using *in vivo* grown nanocrystals diffracting to better than 2.5 Å resolution (Colletier *et al.*, 2016). Indeed, having high-resolution data was important for all successful examples solved so far. In contrast to the atomic form factor which decreases strongly with scattering angle, the anomalous scattering signal f'' does not depend on it. Therefore, in principle, the contribution of the anomalous signal increases with resolution. However, since high-resolution data are typically measured less accurately, using crystals that diffract strongly to high resolution makes a big difference for phasing. Indeed, in a systematic study describing selenium- and mercury-based anomalous phasing approaches published in this issue of *IUCrJ*, Yamashita *et al.* (2017) show that significantly higher multiplicity of measurements for individual reflections and thus many more diffraction
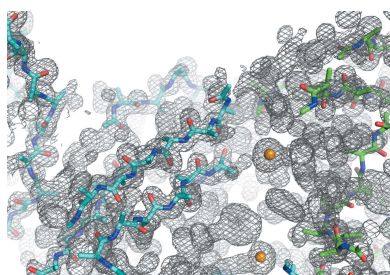
images are needed when reducing the resolution of a mercury derivative dataset for SIRAS phasing from 1.5 Å to 2.6 Å (400000 patterns instead of 11000). Importantly, this large increase in multiplicity is required despite the fact that the data are excellent and very strong at 2.6 Å (unlike for systems that really stop diffracting at 2.6 Å). They partly attribute this to the generally poor quality of low-resolution SFX data.

Knowing whether one can solve a structure by SAD phasing and when to stop collecting more data because the anomalous signal in SAD phasing is high enough is still an active area of research [see, for example, Terwilliger *et al.* (2016) and references therein] despite the fact that SAD phasing is, and has been for quite some time, the method of choice of structure determination if no related structure is available in the Protein Data Bank. In contrast to conventionally collected SAD data where the anomalous correlation CC_{ano} is a good metric for the quality of the anomalous differences, this is rarely the case for SFX data, as also pointed out by Yamashita *et al.* (2017) in this issue. Thus, better analysis programs are needed urgently, improving SFX data quality such that CC_{ano} (or an alternative measure) becomes meaningful and fewer images (and thus less sample and beam time) are needed to obtain accurate Bragg intensities.

In fact, data processing programs have improved significantly in recent years. This is reflected in the finding that fewer and fewer diffraction patterns of a given dataset are needed for phasing. For example, the very first demonstration of *de novo* phasing required 60000 indexed patterns for fully automatic model building (Barends *et al.*, 2014), but recently 10000 patterns were sufficient (Nass *et al.*, 2016). Similarly, two years ago, Yamashita *et al.* (2015) could not solve their structure using only the anomalous signal of their mercury derivative, and now they can (Yamashita *et al.*, 2017). The future will show how far this can go but it seems likely that more than the optimization of details is needed before real systems diffracting weakly to medium resolution can be phased with reasonable requirements on sample and beam time.

References

Ayyer, K., Yefanov, O. M., Oberthür, D., Roy-Chowdhury, S., Galli, L., Mariani, V., Basu, S., Coe, J., Conrad, C. E., Fromme, R., Schaffer, A., Dörner, K., James, D., Kupitz, C., Metz, M., Nelson, G., Xavier, P. L., Beyerlein, K. R., Schmidt, M., Sarrou, I., Spence, J. C. H., Weierstall, U., White, T. A., Yang, J., Zhao, Y., Liang, M., Aquila, A., Hunter, M. S., Robinson, J. S., Koglin, J. E., Boutet, S., Fromme, P., Barty, A. & Chapman, H. N. (2016). *Nature (London)*, **530**, 202–206.

Barends, T. R., 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 (London)*, **505**, 244–247.

Barends, T. R. M., Foucar, L., Shoeman, R. L., Bari, S., Epp, S. W., Hartmann, R., Hauser, G., Huth, M., Kieser, C., Lomb, L., Motomura, K., Nagaya, K., Schmidt, C., Strecker, R., Anielski, D., Boll, R., Erk, B., Fukuzawa, H., Hartmann, E., Hatsui, T., Holl, P., Inubushi, Y., Ishikawa, T., Kassemeyer, S., Kaiser, C., Koeck, F., Kunishima, N., Kurka, M., Rolles, D., Rudek, B., Rudenko, A., Sato,

T., Schroeter, C.-D., Soltau, H., Strueder, L., Tanaka, T., Togashi, T., Tono, K., Ullrich, J., Yase, S., Wada, S., Yao, M., Yabashi, M., Ueda, K. & Schlichting, I. (2013). *Acta Cryst.* D**69**, 838–842.

Batyuk, A., Galli, L., Ishchenko, A., Han, G. W., Gati, C., Popov, P. A., Lee, M.-Y., Stauch, B., White, T. A., Barty, A., Aquila, A., Hunter, M. S., Liang, M., Boutet, S., Pu, M., Liu, Z., Nelson, G., James, D., Li, C., Zhao, Y., Spence, J. C. H., Liu, W., Fromme, P., Katritch, V., Weierstall, U., Stevens, R. C. & Cherezov, V. (2016). *Sci. Adv.* **2**, e1600292.

Colletier, J. P., Sawaya, M. R., Gingery, M., Rodriguez, J. A., Cascio, D., Brewster, A. S., Michels-Clark, T., Hice, R. H., Coquelle, N., Boutet, S., Williams, G. J., Messerschmidt, M., DePonte, D. P., Sierra, R. G., Laksmono, H., Koglin, J. E., Hunter, M. S., Park, H. W., Uervirojnangkoorn, M., Bideshi, D. K., Brunger, A. T., Federici, B. A., Sauter, N. K. & Eisenberg, D. S. (2016). *Nature (London)*, **539**, 43–47.

Fukuda, Y., Tse, K. M., Nakane, T., Nakatsu, T., Suzuki, M., Sugahara, M., Inoue, S., Masuda, T., Yumoto, F., Matsugaki, N., Nango, E., Tono, K., Joti, Y., Kameshima, T., Song, C., Hatsui, T., Yabashi, M., Nureki, O., Murphy, M. E. P., Inoue, T., Iwata, S. & Mizohata, E. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 2928–2933.

Gorel, A. *et al.* (2017). *Nat. Commun.* In the press.

Hara, T., Inubushi, Y., Katayama, T., Sato, T., Tanaka, H., Tanaka, T., Togashi, T., Togawa, K., Tono, K., Yabashi, M. & Ishikawa, T. (2013). *Nat. Commun.* **4**, 2919.

Hunter, M. S., Yoon, C. H., DeMirci, H., Sierra, R. G., Dao, E. H., Ahmadi, R., Aksit, F., Aquila, A. L., Ciftci, H., Guillet, S., Hayes, M. J., Lane, T. J., Liang, M., Lundström, U., Koglin, J. E., Mgbam, P., Rao, Y., Zhang, L., Wakatsuki, S., Holton, J. M. & Boutet, S. (2016). *Nat. Commun.* **7**, 13388.

Nakane, T., Hanashima, S., Suzuki, M., Saiki, H., Hayashi, T., Kakinouchi, K., Sugiyama, S., Kawatake, S., Matsuoka, S., Matsumori, N., Nango, E., Kobayashi, J., Shimamura, T., Kimura, K., Mori, C., Kunishima, N., Sugahara, M., Takakyu, Y., Inoue, S., Masuda, T., Hosaka, T., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Inoue, T., Nureki, O., Iwata, S., Murata, M. & Mizohata, E. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 13039–13044.

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.

Schlichting, I. (2015). *IUCrJ*, **2**, 246–255.

Son, S.-K., Chapman, H. N. & Santra, R. (2011). *Phys. Rev. Lett.* **107**, 218102.

Spence, J. C. H. (2017). *IUCrJ*, **4**, 322–339.

Spence, J. C., Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., White, T., Barty, A., Chapman, H. N., Marchesini, S. & Holton, J. (2011). *Opt. Express*, **19**, 2866–2873.

Terwilliger, T. C., Bunkóczi, G., Hung, L.-W., Zwart, P. H., Smith, J. L., Akey, D. L. & Adams, P. D. (2016). *Acta Cryst.* D**72**, 346–358.

Yamashita, K., Kuwabara, N., Nakane, T., Murai, T., Mizohata, E., Sugahara, M., Pan, D., Masuda, T., Suzuki, M., Sato, T., Kodan, A., Yamaguchi, T., Nango, E., Tanaka, T., Tono, K., Joti, Y., Kameshima, T., Hatsui, T., Yabashi, M., Manya, H., Endo, T., Kato, R., Senda, T., Kato, H., Iwata, S., Ago, H., Yamamoto, M., Yumoto, F. & Nakatsu, T. (2017). *IUCrJ*, **4**, 639–647.

Yamashita, K., Pan, D., Okuda, T., Sugahara, M., Kodan, A., Yamaguchi, T., Murai, T., Gomi, K., Kajiyama, N., Mizohata, E., Suzuki, M., Nango, E., Tono, K., Joti, Y., Kameshima, T., Park, J., Song, C., Hatsui, T., Yabashi, M., Iwata, S., Kato, H., Ago, H., Yamamoto, M. & Nakatsu, T. (2015). *Sci. Rep.* **5**, 14017.