

Single-wavelength anomalous dispersion (SAD) phasing with native anomalous scatterers using serial femtosecond crystallography

Calculation of the electron density requires complex structure factors. While the absolute values of structure factors are experimentally measurable, their phases are not. This is known as *the phase problem* of crystallography. When the structures of homologous proteins for molecular replacement are unavailable, we have to rely on experimental phasing, where phases are derived from isomorphous difference signals between native and derivative crystals, or anomalous dispersion effects of anomalous scatterers. Sulfur is the lightest element that provides useful anomalous differences at wavelengths commonly used for protein crystallography. Since almost all proteins contain sulfurs from methionines and cysteines, experimental phasing with sulfur does not require protein derivatization, which can be tedious and tricky. However, the signals from sulfurs are much weaker than those from heavier atoms and great care must be taken to accurately measure them.

In serial femtosecond crystallography (SFX), independent diffraction patterns from tens of thousands of microcrystals are combined to yield a dataset. This is because pulses from an X-ray free electron laser (XFEL) are so powerful that a single shot completely destroys a crystal. As each crystal provides only one diffraction image, indexing and orientation determination are more challenging than by using the traditional rotation method. Moreover, every observation is partial since crystals are virtually stationary during the femtosecond X-ray exposure. Scaling and outlier rejection are also problematic because of inhomogeneities in the crystal size and quality, and fluctuations of the beam intensity and spectrum. These issues could potentially limit the quality of SFX datasets.

Since single wavelength anomalous dispersion phasing with sulfur atoms (S-SAD) requires highly accurate data, it provides a unique opportunity to test the quality of SFX datasets. In addition, SAD phasing shares some similarities with time-resolved studies in that both depend on the accurate measurement of tiny differences between Friedel pairs or between excited and reference datasets, respectively. Thus, the study of the data quality and how to optimize it will benefit not only experimental phasing but also other SFX experiments.

In this study, we solved the structure of lysozyme by S-SAD (Fig. 1) [1]. Although lysozyme is a model protein, it is no easier to solve by S-SAD; lysozyme contains eight cysteines and two methionines in 129 residues (Fig. 1), giving rise to the expected anomalous signal, or a Bijvoet ratio $\langle |\Delta F| \rangle / \langle |F| \rangle$ of only 1.6%. Before this study, the gadolinium SAD phasing of lysozyme [2] carried out at

LCLS, an XFEL facility at Stanford, was reported. In this case, the Bijvoet ratio was larger than 10%.

Lysozyme microcrystals (7–10 μm) were suspended in a grease medium [3]. The experiments were performed at SACLA BL3 on the Diverse Application Platform for Hard X-ray Diffraction in SACLA (DAPHNIS) [4]. The X-ray wavelength was 1.77 Å (7 keV) and each X-ray pulse delivered $\sim 7 \times 10^{10}$ photons within a 10 fs duration (FWHM) to the grease matrix. The X-ray beam was focused to $1.5 \times 1.5 \mu\text{m}^2$. The crystals in the grease matrix were serially loaded using a high-viscosity micro-extrusion injector system installed in a helium chamber. Diffraction images were collected using a 4 M pixel detector built with eight panels of multi-port charge-coupled device (MPCCD).

Diffraction images were filtered by a data processing pipeline developed for SACLA [5] based on *Cheetah*. Out of about 700,000 images collected, about 450,000 images with more than 20 spots were retained as hits and processed by *CrystFEL*. About 180,000 images were indexed by *DirAx*. Integrated intensities were scaled by per-image scale factors (but not B factors) before merging. No sigma cutoff or partiality correction was applied.

Because the data quality improves with the multiplicity, we attempted to phase the dataset from varying numbers of indexed images. As shown in Fig. 2, the data quality indicators improved with the number of merged images. It turned out that at least 150,000 indexed images were necessary to solve this structure. Interestingly, the data precision (CC_{ano}) remained low even when the accuracy (peak height in the anomalous difference map) was sufficiently high to solve the structure.

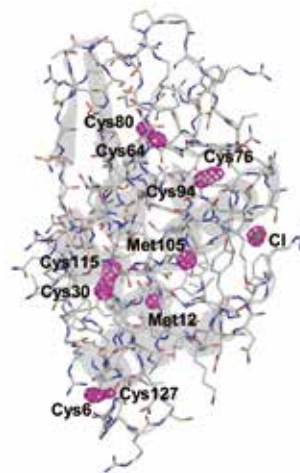


Fig. 1. Anomalous difference Fourier map (contoured at 6.0σ) calculated by *ANODE*, showing sulfur and chlorine atoms. [1]

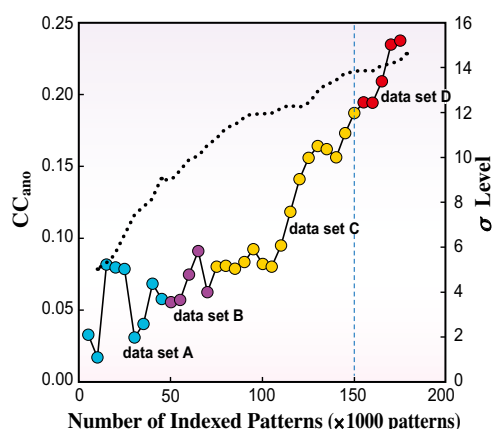


Fig. 2. Data quality statistics with varying number of merged images. The solid line indicates CC_{ano} . See the original paper [1] for the details of the four datasets. The peak height of the Met105 sulfur atom in the anomalous difference map is shown as the dotted black line. 150,000 images (blue dashed line) had to be merged for successful phasing.

Experimental phasing was performed with the *SHELX* suite. Systematic trials of parameter combinations were essential for the success. First, up to 500,000 trials of *SHELXD* were executed at various high-resolution limits between 2.0 and 3.0 Å to locate anomalous scatterers. When only 150,000 indexed images were used, reflections to 2.2 Å had to be used

and the best solution only appeared after 320,000 trials. In contrast, solutions could be found more easily at a lower resolution with fewer trials when more images were merged. Next, the experimentally phased map was calculated and improved by iterative autotracing and density modification in *SHELXE*. Here, the number of sites, the high-resolution cutoff and the solvent content were systematically varied. Although the initial map was noisy and fragmented, the iterations improved the phase and *SHELXE* eventually traced 90 out of 129 residues (Fig. 3). *Buccaneer* automatically completed the model in the experimental map from *SHELXE*.

After the publication of this research, we also reported the Cu-SAD phasing of a metalloenzyme with a Bijvoet ratio of 1.7% [6]. Other groups reported the S-SAD phasing of thaumatin [7] and A2a G-protein-coupled receptor [8], with Bijvoet ratios of 2.1% and 1.9%, respectively. Although these results collectively establish that data from SFX are sufficiently accurate to detect small differences, it is noteworthy that all of these studies required more than 100,000 indexed images. To reduce the number of necessary images, and thus sample consumption and beam time, further improvements in data processing algorithms are under way.

The raw data (diffraction images) have been deposited in CXIDB (ID#33; <http://cxidb.org/id-33.html>). Readers are encouraged to reprocess our data to learn SFX and try their new ideas.

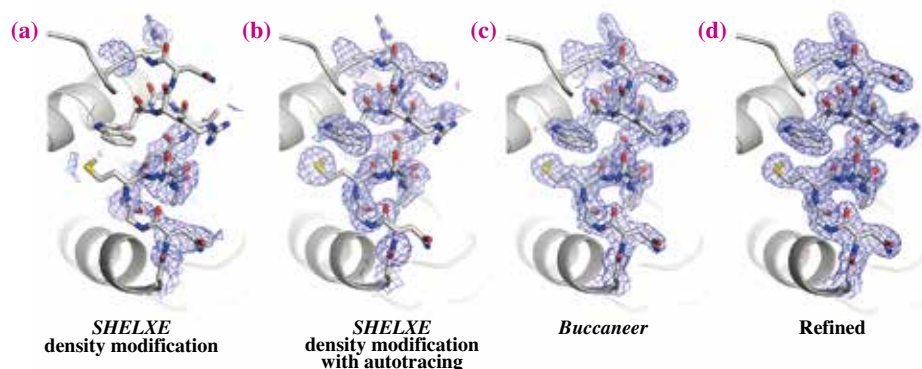


Fig. 3. $2F_o - F_c$ electron density maps contoured at the 1.0σ level from the various steps of the phasing process. SAD phasing was performed by (a) *SHELXE* with density modification, followed by (b) autotracing of the main chain in *SHELXE* and (c) automatic modeling of side chains and remodeling of the main chain by *Buccaneer*. (d) shows the final refined map. [1].

Takanori Nakane^a, So Iwata^{b,c} and Michihiro Sugahara^{b,*}

^a Graduate School of Science, The University of Tokyo

^b RIKEN SPring-8 Center

^c Graduate School of Medicine, Kyoto University

*Email: msuga@spring8.or.jp

References

- [1] T. Nakane, C. Song, M. Suzuki, E. Nango, J. Kobayashi, T. Masuda, S. Inoue, E. Mizohata, T. Nakatsu, T. Tanaka, R. Tanaka, T. Shimamura, K. Tono, Y. Joti, T. Kameshima, T. Hatsui, M. Yabashi, O. Nureki, S. Iwata and M. Sugahara: *Acta Crystallogr. D Biol. Crystallogr.* **71** (2015) 2519.
- [2] T. Barends *et al.*: *Nature* **505** (7482) (2014): 244-247.
- [3] M. Sugahara *et al.*: *Nat. Methods* **12** (2015) 61.
- [4] K. Tono *et al.*: *J. Synchrotron Rad.* **22** (2015) 532.
- [5] T. Nakane *et al.*: *J. Appl. Crystallogr.* **49** (2016) 1035.
- [6] Y. Fukuda *et al.*: *Proc. Natl. Acad. Sci. USA* **113** (2016) 2928.
- [7] K. Nass *et al.*: *IUCrJ* **3** (2016) 180.
- [8] A. Batyuk *et al.*: *Sci. Adv.* **2** (2016) e1600292.