Multi-wavelength anomalous diffraction *de novo* phasing of serial femtosecond crystallography data using two-color pulses at SACLA

Serial femtosecond crystallography (SFX) at XFELs offers unprecedented possibilities for macromolecular structure determination of systems prone to radiation damage. However, *de novo* structure determination, i.e. without prior structural knowledge, is complicated by the inherent inaccuracy of SFX data. By its very nature, SFX data collection entails shot-to-shot fluctuations in X-ray wavelength and intensity as well as variations in crystal size and quality that must be averaged out. Hence, to obtain accurate diffraction intensities for *de novo* phasing, very large numbers of diffraction patterns are required — the multiplicity of measurements for a given reflection being typically several 100- to 1000-fold depending on the phasing method and signal strength — and, concomitantly large amounts of sample and long XFEL beamtimes. Since both are typically precious and often limited, improved use of one or both is essential to future evolution of XFEL-based structural biology.

The recently established two-color operation of SACLA [1] opened up a novel possibility of collecting two SFX datasets simultaneously, without doubling the amount of sample used. Due to the unprecedentedly large energy separation of the two tunable colors of the XFEL beam [1], two distinct and spatially well separated diffraction patterns can be recorded simultaneously on one diffraction image of the same crystal. The simultaneous arrival of the two XFEL pulses precludes damage effects from the first pulse affecting the diffraction of the second pulse. This allows simultaneous same-crystal acquisition of two-wavelength datasets for multiple wavelength anomalous dispersion (MAD) phasing. (This is in marked contrast to data collection at synchrotron sources where they are typically collected sequentially.)

In principle, given the availability of more information, MAD phase angles are expected to be more accurate than those from single wavelength anomalous dispersion (SAD) experiments. To explore whether this can be put to use for XFEL-based *de novo* phasing with the added benefit of halved sample consumption and beamtime, we performed a two-color SFX experiment at SACLA **BL3**. We used microcrystals of the well-established model system lysozyme, in complex with a lanthanide compound and collected SFX diffraction data in the DAPHNIS chamber using a multiport charge coupled device (MPCCD) detector (see Fig. 1). SACLA operated at

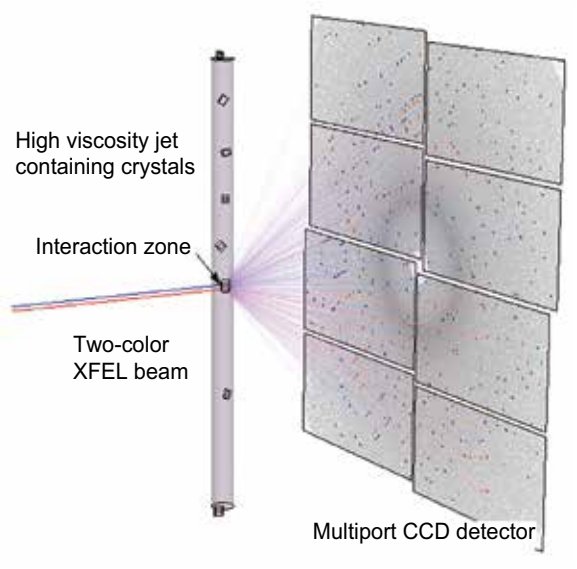

Fig. 1. Two-color serial femtosecond crystallography experiment. Experimental setup.

30 Hz and simultaneously delivered two-color X-ray pulses of nominally 7.0 keV and 9.0 keV photon energy, which are above the *M*- and *L*-edges of gadolinium, respectively. We collected 570,000 diffraction patterns in ~12 hours. Online data analysis was performed with CASS [2] and the Graphic User Interface to the offline data processing pipeline Cheetah Dispatcher was used to identify 208,373 hits (37% hit rate). A wide range inline spectrometer was used to simultaneously record the spectral information for the 7 keV and 9 keV colors for each XFEL pulse.

The analysis of two-color SFX data is not straightforward. In fact, direct processing with CrystFEL [3] was unsuccessful, since only a minute fraction of the hits could be indexed in both colors. Despite aiming for similar pulse energies for the two colors and compensating for the difference in detector quantum efficiency by inserting an aluminium filter, the intensity distribution of the two patterns in the diffraction image varied. One diffraction pattern typically dominated and could be indexed in one color, but indexing of the weaker second diffraction pattern in the other color typically failed. To index the weaker diffraction pattern, the threshold for identifying peaks had to be lowered and the previously indexed peaks were eliminated

from the list of all identified peaks. Using this approach we successfully indexed and integrated 11.1 % of the hits in both colors (Fig. 2). Phases were determined automatically using data to 1.9 Å resolution (Fig. 3).

We deliberately used a model system with an unusually strong anomalous signal. In spite of this we see a significant increase in the data information content of the two-color data used for MAD phasing, as evidenced by the higher figure of merit indicating more accurate initial phases, and easier model building compared to the single color data SAD phasing approach [4]. This difference is particularly striking at 5000 images, which is a comparatively low number for SFX data collection. Hence, these diffraction intensities are of lower precision than those derived from a larger number of images, as evidenced by the data statistics. At 5000 images, the first round of automated building essentially failed in the SAD case, whereas in the MAD case most of the structure was built. It has been suggested that for suboptimal data, density modification might more easily improve even inaccurate phases provided by MAD, which are unimodal, rather than SAD phases which are additionally compromised by a handedness ambiguity. This could help explain the superiority of the MAD phases during the later stages of structure determination. We expect the difference between SAD and MAD to be even larger for more challenging cases with weaker anomalous signals.

In conclusion, we have demonstrated that XFEL-based two-color phasing is not only feasible but also advantageous [4,5]. Using a well-characterized model system we show that significantly fewer indexed patterns are required for *de novo* phasing using two-color data compared to single color data. This should reduce the requisite amounts of sample and beamtime. We expect two-color data collection to be particularly useful for difficult-to-phase projects where it may make the crucial difference between being able to solve the structure and not.

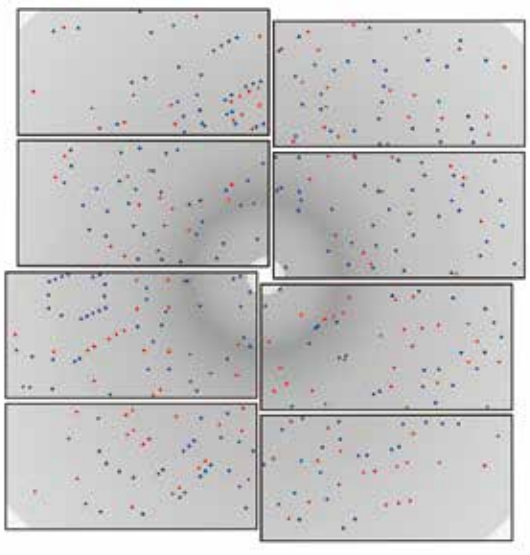

Fig. 2. Indexing of the two-color diffraction patterns.

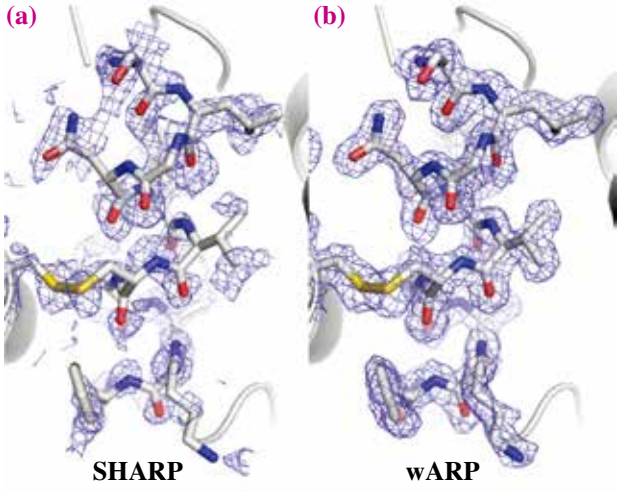

Fig. 3. Progression of the MAD phasing process with 5000 images. **(a)** Initial phases obtained with SHARP. **(b)** Phases after automatic building and -refinement by ARP/wARP. All maps are contoured at 1 sigma and are superimposed onto the final, refined structure (PDB code 5OER).

Alexander Gorel and Ilme Schlichting*

Max-Planck-Institut für medizinische Forschung, Germany

*Email: Ilme.Schlichting@mpimf-heidelberg.mpg.de

References

[1] T. Hara *et al.*: Nat. Commun. **4** (2013) 2919.
[2] L. Foucar: J. Appl. Cryst. **49** (2016) 1336.
[3] T.A. White: J. Appl. Cryst. **45** (2012) 335.
[4] *A. Gorel, K. Motomura, H. Fukuzawa, R.B. Doak, M.L. Gruenbein, M. Hilpert, I. Inoue, M. Kloos, G. Kovácsová, E. Nango, K. Nass, C.M. Roome, R.L. Shoeman, R. Tanaka, K. Tono, Y. Joti, M. Yabashi, S. Iwata, L. Foucar, K. Ueda, T.R.M. Barends and I. Schlichting: Nat. Commun. **8** (2017) 1170.*
[5] A. Gorel *et al.*: Scientific Data **4** (2017) 170188.