Optimizing data quality in injector-based serial millisecond crystallography

N. Zatsepin[a,b], S. Botha[a,b], J. Martin-Garcia[b,c], H. Hu[b,c], U. Weierstall[b,c], W. Shi[d], B. Andi[d], J. Skinner[d], H. Bernstein[e], P. Fromme[b,c], M. Fuchs[d].

[a] Department of physics, Arizona State University, Tempe AZ 85287, USA, nzatsepi@asu.edu
[b] Biodesign Center for Applied Structural Discovery, Arizona State University, Tempe AZ 85287, USA
[c] School of Molecular Sciences, Arizona State University, Tempe AZ 85287, USA
[d] Energy & Photon Sciences Directorate, Brookhaven National Laboratory, Upton NY 11973, USA
[e] School of Chemistry and Materials Science, Rochester Institute of Technology, Rochester NY, USA

Serial millisecond crystallography (SMX) is an effective way to mitigate radiation damage in room temperature structure determination by spreading the dose over many crystals. Recent upgrades at synchrotron microfocus beamlines coupled with high framerate, low noise detectors have enabled high-resolution, room-temperature structure determination by merging data from many very small micron-scale protein crystals. In "snapshot" SMX, microcrystals are delivered using a microfluidic injector for viscous media, most commonly LCP, lipidic cubic phase [1], in which the randomly-oriented microcrystals rotate negligibly during their single, brief exposures of 30 ms or less [2]. LCP-SMX has been used for room-temperature structure determination from microcrystals of radiation hypersensitive membrane proteins [3], G protein-coupled receptor structure determination using pink-beam SMX [4], native single-wavelength anomalous dispersion (SAD) phasing [3,5,6] and pump-probe time-resolved studies [7].

In order for LCP-SMX to become routine, however, one of the challenges that needs to be overcome is the large data volume (high reflection multiplicity) necessary for accurate merged SMX datasets. We have carried out a systematic study of experimental parameters and data metrics to guide data collection for successful and efficient SAD phasing from LCP-SMX data, in part to help address the data volume challenge.

Meanwhile, funding agencies and journals are increasingly requiring the sharing of datasets, which will no doubt accelerate development of analysis methods and improve reliability of published results. However, what constitutes useful "raw data" for snapshot SMX (or serial femtosecond crystallography), is up for debate – a debate I wish to liven with some provocative questions based on our SAD studies.

References
1. Weierstall, U. *et al.* (2014) *Nat. Commun.* **5**(3309)**,** 1–6.
2. Martin-Garcia, J. M. *et al.* (2017) *IUCrJ* **4,** 1–16.
3. Weinert, T. *et al.* (2017) *Nat. Commun.* **8**(542)**,** 1–11.
4. Martin-Garcia, J. M. *et al.* (2019) *IUCrJ* **6,** 1–14.
5. Botha, S. *et al.* (2015) *Acta Cryst. D* **71,** 387–397.
6. Huang, C.-Y. *et al.* (2018) *Communications Biology* **1** (124), 1–8.
7. Weinert, T. *et al.* (2019) bioRxiv **doi:** https://doi.org/10.1101/576629