MS02-P12

New tools for serial crystallography: SPIND - sparse pattern auto-indexing and DatView - exploring and optimizing large multi-crystal datasets

Nadia Zatsepin[1], Chufeng Li[1], Natasha Stander[2], Xuanxuan Li[3], Richard Kirian[1], Haiguang Liu[4], Petra Fromme[2], John Spence[1]

1. Department of Physics, Arizona State University, Tempe, United States of America
2. School of Molecular Sciences and Biodesign Center for Applied Structural Discovery, Arizona State University , Tempe, United States of America
3. Complex Systems Division, Beijing Computational Science Research Center and Department of Engineering Physics, Tsinghua University, Beijing, China
4. Complex Systems Division, Beijing Computational Science Research Center, Beijing, China

email: nadia.zatsepin@asu.edu

We present two new tools for making efficient use of serial femtosecond crystallography (SFX) data. A large fraction of SFX patterns from microcrystals of large macromolecules contain fewer than 15 Bragg spots and are often discarded by existing autoindexing algorithms. SPIND is a reference-based algorithm for auto-indexing sparse, snapshot serial crystallography data with as few as 5 Bragg spots, using a known unit cell [1]. We demonstrate the suitability of SPIND for indexing sparse inorganic crystal data and for improving the quality of SFX data from two G protein-coupled receptors.

We have also developed DatView, a tool for exploring correlations in, and optimizing merged reflection lists from, large serial crystallography datasets. Stochastic fluctuations in microcrystals and X-ray free electron laser pulse parameters lead to orders of magnitude variability in measured intensities from equivalent reflections, which necessitated large SFX datasets for accurate merged structure factors. Numerous improvements in sample delivery, detectors and, critically, data analysis (including geometry optimization, orientation refinement, scaling and post-refinement), have significantly decreased the required number of indexed SFX patterns for structure determination. However, large SFX datasets (collected from thousands of microcrystals at room temperature) now allow users to not only select an optimum subset from potentially anisomorphous microcrystals for more accurate structure factors, but, importantly, it enables the quantification of microcrystal variability and sensitivity to experimental environment such as temperature and humidity, and offer insights into protein flexibility.

SPIND and DatView are written in Python and available via zatsepinlab.atlassian.net

References:

[1] C. Li, X. Li, R. Kirian, J. Spence, H. Liu, and N. Zatsepin. 2018. IUCrJ. In press.

Keywords: serial femtosecond crystallography, dataset optimization, clustering

MS02-P13

Collecting best data for meaningful structures

Andreas Förster[1], Claudio Klein[2], Clemens Schulze-Briese[1]

1. DECTRIS Ltd., Switzerland
2. marXperts GmbH, Germany

email: andreas.foerster@dectris.com

Data collection is the last experimental step in a crystallographic project. All subsequent procedures (phasing, model building, model refinement) critically depend on the quality of the collected data. No amount of computational wizardry will turn poor data into a great model that can provide reliable biological insight.

With their intrinsic lack of noise, high dynamic range, small point spread and fast frame rates, modern Hybrid Photon Counting (HPC) detectors like PILATUS3 and EIGER enable the collection of data of exceptional quality. To fully exploit the power of these detectors, the user has to design her collection strategy based on the underlying technology.

In this presentation, we will explain why HPC data should always be collected with a small crystal oscillation increment per image (relative to the crystal mosaicity). This strategy is known as fine phi-slicing. We will also explain why HPC technology favors high multiplicity experiments. Native SAD structures determined at room temperature and from a crystal in space group P1 will illustrate the power of these two approaches.

Small oscillation increments minimize the background measured with each reflection, thus increasing signal to noise. High multiplicity decreases the systematic errors of the experiment and brings integrated intensities closer to the real value. Both strategies have long been known in theory to lead to the most accurate data possible. In practice, technological hurdles like detector readout noise and dark current and detector readout time have forced the user into compromises in the past, balancing what's best for the crystal with the constraints of the detector.

With HPC detectors, everyone can collect data optimally. This doesn't mean every experiment is a success. It still has be set up right. Keeping the points in this presentation in mind, trained crystallographers will not have to think twice to get the most from their crystals.

Keywords: Accurate data, Experimental strategy, Hybrid Photon Counting