
Development of graphical user interface for the three-dimensional merged intensity maps analysis

DESY Summer Student Programme, 2021

Anna Koriazhkina

Moscow Institute of Physics and Technology (National Research University)

Supervisor
O. Yefanov



September 9, 2021

1 Introduction

The technique of X-ray crystallography has remained the most reliable and productive method of structural biology for many decades. Since the first protein structure of myoglobin, which was determined in the late 1950s by Sir John Cowdery Kendrew, thousands of structures of proteins and other biomolecules have been obtained using X-ray crystallography. This data provided a fundamentally new level of understanding of the living organisms functioning. Structural biology also opened up a new era for drug development: high-resolution structures of proteins are now used for the rational search of candidate compounds with advanced potency and minimized side effects.[1]

Even wider horizons for protein structure determination have opened with the development of advanced X-ray sources. X-ray free-electron lasers (XFELs) have made serial protein crystallography possible. This method uses protein nanocrystals in a continuous flow of viscous liquid to collect multiple ‘snapshots’, each from a distinct crystal in a random orientation. The short pulses from XFEL enable the ‘diffraction-before-destruction’ data collection principle. Consequently, the diffraction data is collected before the radiation damage, as the pulse time scale is briefer than the characteristic time of the damage process. Therefore, protein crystals, which are either too small or too radiation-sensitive can be studied. Moreover, the technology opens up brand new possibilities in time-resolved crystallography.[2]

The common data analysis approach for serial crystallography usually involves indexing the diffraction pattern from each crystal individually. Therein, the intensity of every predicted reflection in the diffraction pattern is locally integrated, and then reflection intensities from many thousands of crystals are merged and averaged. At the same time, all the information about the intensity distribution around and between Bragg peaks is discarded.

However, this additional data can be preserved for direct-phasing strategies and other advanced analyzes. Consequently, a new approach for the inspection of diffraction patterns in the serial femtosecond crystallography has been suggested. The approach involves a ‘three-dimensional merge’ of the data from numerous nanocrystals by assembling diffraction patterns from individual crystals in three-dimensional reciprocal space using crystal orientation information derived from the autoindexing process.[3]

Besides novel phasing strategies, the three-dimensional merge approach opens up some other opportunities. First, it can be applied for the integration and refinement of datasets with partial reflections. Moreover, the three-dimensional intensity model can be used for the post-refinement of diffraction image orientation and scaling.

In this report, I would like to briefly describe my work on a graphical user interface (GUI) that can be used to manipulate the merged intensity map. This interface was prepared by me using the Python programming language and its modules. Several useful tools for serial femtosecond crystallography data analysis have been implemented in the GUI. Also in the first section of this report, I would talk about my introduction to the traditional data analysis pipeline for serial femtosecond crystallography using CrystFEL software.

2 CrystFEL software

First, I studied a tutorial to get familiar with the common pipeline for analysis of diffraction data using CrystFEL software. CrystFEL is adjusted to the needs of the serial femtosecond crystallography technique and is suitable for the analysis of diffraction patterns obtained using XFEL. Its functionality includes indexing, integrating, merging, viewing, and evaluating the quality of the data, and also simulating diffraction patterns. To accelerate the indexing and integrating routines for a large number of diffraction snapshots, the CrystFEL can use multi-core hardware. Another important feature of this software is the use of a generalized representation of the detector, which enables the work with more complex detector geometries.[4]

While following the tutorial, I made most of the main steps of using CrystFEL, using the freely available dataset, collected on the Linac Coherent Light Source facility.[5] Thus, the diffraction patterns from the dataset were examined, then the peak detection was optimized, the unit cell parameters were computed, the diffraction patterns were indexed, the detector geometry was optimized, and, finally, the intensities were merged, and some basic figures of merit were calculated and analyzed.

3 GUI for three-dimensional merging

Next, I started developing the GUI, which should facilitate the process of a merged three-dimensional intensity map inspection. Several useful tools were included in the GUI: first, it includes the adjustable level of intensity to build corresponding isosurfaces.

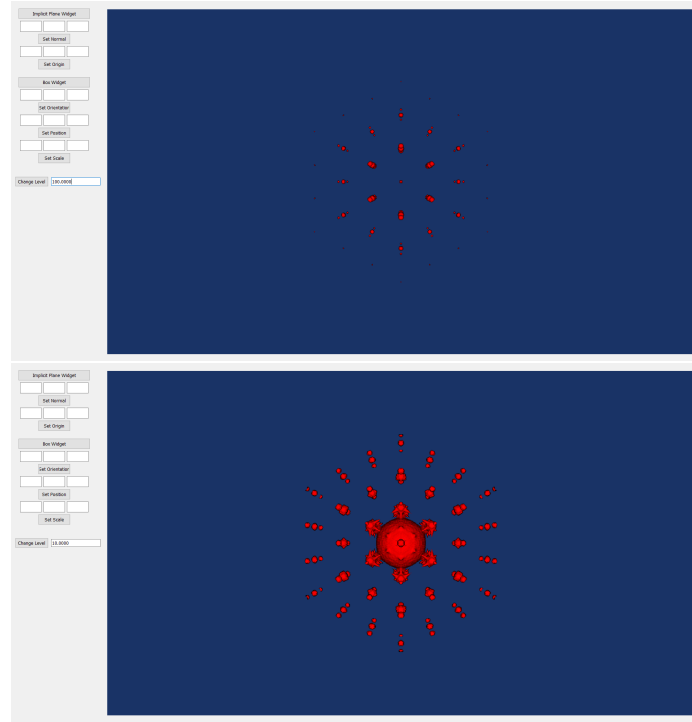


Figure 1: A tool for creating isosurfaces at the required level in the work

Second, the data can be sliced using the implicit plane widget and clipped using the box widget. The orientation and position of both widgets, as well as the scale of the box widget, are fully adjustable both manually and using text input fields. These tools allow inspection of individual slices of merged three-dimensional reciprocal space, as well as slices, averaged over a thin layer of reciprocal space, which also offer some advantages for more accurate data analysis.

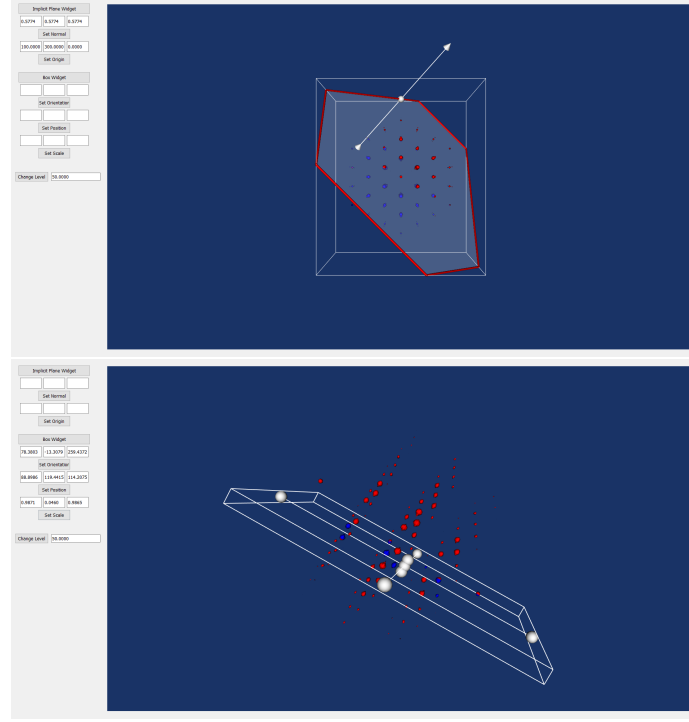


Figure 2: Implicit plane widget and box widget with adjustable orientation, position and scale in the work

Rotation and zoom of the scene using a mouse are also fully supported.

The GUI was implemented using the Visualization Toolkit (VTK), which is open-source software for displaying and managing scientific data. VTK is an object-oriented system, which consists of a compiled C++ class library and an “interpreted” wrapper layer that can be used to manipulate compiled classes using many languages, including Java, Tcl, and Python. We chose the VTK Python wrapper for our GUI. Basic GUI functionality, including buttons and text boxes implemented with the PyQt module.[6]

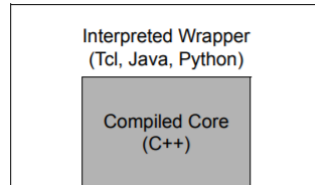


Figure 3: Visualization Toolkit (VTK) architecture from [6]

The implemented GUI will become a part of a new project on three-dimensional merge and the results of the project are being prepared for publication in the Journal of Applied Crystallography.

4 Results

- Familiarized with the main concepts of protein crystallography, and, particularly, with the serial femtosecond crystallography;
- Studied the differences between common approach for crystallography data analysis, and new method, which involves the three-dimensional merging of intensities;
- Dived into serial femtosecond crystallography data manipulation using CrystFEL software;
- Developed a code for GUI, which facilitates inspection of three-dimensional merged intensity maps, using VTK and PyQt modules.

References

- [1] Anderson A. C. The process of structure-based drug design. *Chemistry & biology*, 10(9):787–797, sep 2003.
- [2] Mishin A, Gusach A, Luginina A, Marin E, Borshchevskiy V, and Cherezov V. An outlook on using serial femtosecond crystallography in drug discovery. *Expert opinion on drug discovery*, 14(9):933–945, sep 2019.
- [3] Oleksandr Yefanov, Cornelius Gati, Gleb Bourenkov, Richard A. Kirian, Thomas A. White, John C. H. Spence, Henry N. Chapman, and Anton Barty. Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647), jul 2014.
- [4] Thomas A White, Richard A Kirian, Andrew V Martin, Andrew Aquila, Karol Nass, Anton Barty, and Henry N Chapman. CrystFEL : a software suite for snapshot serial crystallography. *J. Appl. Cryst*, 45:335–341, 2012.
- [5] Wei Liu, Daniel Wacker, Cornelius Gati, Gye Won Han, Daniel James, Dingjie Wang, Garrett Nelson, Uwe Weierstall, Vsevolod Katritch, Anton Barty, Nadia A. Zatsepin, Dianfan Li, Marc Messerschmidt, Sébastien Boutet, Garth J. Williams, Jason E. Koglin, M. Marvin Seibert, Chong Wang, Syed T.A. Shah, Shibom Basu, Raimund Fromme, Christopher Kupitz, Kimberley N. Rendek, Ingo Grotjohann, Petra Fromme, Richard A. Kirian, Kenneth R. Beyerlein, Thomas A. White, Henry N. Chapman, Martin Caffrey, John C.H. Spence, Raymond C. Stevens, and Vadim Cherezov. Serial femtosecond crystallography of G protein-coupled receptors. *Science*, 342(6165):1521–1524, 2013.
- [6] Lisa S. Avila, Utkarsh Ayachit, Sebastien Barre, Jeff Baumes, Francois Bertel, Rusty Blue, David Cole, David DeMarle, Berk Geveci, William A. Hoffman, Brad King, Karthik Krishnan, C. Charles Law, Kenneth M. Martin, William McLendon, Philippe

Pebay, Niki Russell, William J. Schroeder, Timothy Shead, Jason Shepherd, Andrew Wilson, and Brian Wylie. *The VTK User's Guide*. Kitware, Inc., 11th edition, 2010.