

Solving protein structure using highly realistic diffraction photographs generated from MD trajectory of crystalline lysozyme

Y. Xue¹, N. Liu², N. Skrynnikov³

¹Tsinghua University ²School of Life Sciences, Tsinghua University, ³St. Petersburg State University, St. Petersburg, Russia; Purdue University, West Lafayette, IN, USA

yixue@mail.tsinghua.edu.cn

We have designed and implemented an MD-based pipeline to emulate the entire process of solving crystallographic structures of biomolecules. As a starting point, we have used the structure of lysozyme solved in-house (tetragonal lattice, resolution 2.1 Å). Using this structure, we constructed a supercell containing $5 \times 5 \times 5 = 125$ crystal unit cells and explicit interstitial solvent (SPC/E water). The supercell was used as a periodic boundary box to record 0.3 μs MD trajectory of crystalline lysozyme at room temperature employing Amber 18 software with ff14SB force field. A series of snapshots from this trajectory were then used to simulate a set of diffraction photographs. Normally, crystallographic diffraction pattern is calculated using FFT-based methods or, alternatively, direct summation methods, producing an array of structure factors rather than diffraction photographs per se. In our study we take a different approach. Specifically, we treat each atom in the system (including bulk water) as a scattering center, applying the fundamental Huygens-Fresnel principle to calculate the complete diffraction image of the crystal (not limited to the diffraction spots). The example of such highly realistic MD-based diffraction photograph is shown in Figure 1.

Having generated an array of 180 diffraction images, we then used the standard suite of crystallography programs, HKL2000/XDS, Coot and Phenix, to process these images, extract structure factors and ultimately calculate protein coordinates using the experimental crystal structure as a molecular replacement model. The procedure has been successful, resulting in $R_{\text{work}}=0.205$ and $R_{\text{free}}=0.264$. The recovered lysozyme coordinates are within 0.57 Å of the original structure (see Figure 2) and within 0.25 Å of the average MD coordinates. Thus, we have demonstrated that long state-of-the-art MD trajectory of protein crystal can be used to emulate the entire process of crystallographic structure determination at the unparalleled level of realism, leading to an accurate structural model. We envisage that in the future this procedure can be used to interrogate the relationship between protein internal dynamics and crystallographic variables (e.g. B-factors), to improve computational tools used in the field of biomolecular crystallography and to validate and benchmark different MD force fields used for biomolecular simulations.

This study is supported by the joint NSFC-RSF grant to Yi Xue (award 32061133011) and Nikolai Skrynnikov (award 21-44-00033).

Cyan: initial crystal structure

