POSTER ABSTRACT PRESENTATIONS

SESSION TITLE: ADVANCES IN EM TECHNOLOGY AND PROCESSING

Abstract P-5: Neural Network Approaches to Classify 3D Protein Structures from the Data of X-ray Laser Radiation Diffraction from Single Particles

Valery Novoseletsky, Mikhail Lozhnikov, Grigoriy Armeev, Aleksandr Kudriavtsev, Alexey Shaytan, Georgiy Kobelkov, Konstantin Shaitan

M. V. Lomonosov Moscow State University, Moscow, Russia

Background: Protein structure determination using X-ray free-electron laser (XFEL) includes analysis and merging a large number of snapshot diffraction patterns. Convolutional neural networks are widely used to solve numerous computer vision problems, e.g. image classification, and can be used for diffraction pattern analysis. But the task of protein structure determination with the use of CNNs only is not yet solved.

Methods: We simulated the diffraction patterns using the Condor software library and obtained more than 1000 diffraction patterns for each structure with simulation parameters resembling real ones. To classify diffraction patterns, we tried two approaches, which are widely known in the area of image classification: a classic VGG network and residual networks.

Results: 1. Recognition of a protein class (GPCRs vs globins). Globins and GPCR-like proteins are typical α-helical proteins. Each of these protein families has a large number of representatives (including those with known structure) but we used only 8 structures from every family. 12,000 of diffraction patterns were used for training and 4,000 patterns for testing. Results indicate that all considered networks are able to recognize the protein family type with high accuracy. 2. Recognition of the number of protein molecules in the liposome. We considered the usage of lyposomes as carriers of membrane or globular proteins for sample delivery in XFEL experiments in order to improve the X-ray beam hit rate. Three sets of diffractograms for liposomes of various radius were calculated, including diffractograms for empty liposomes, liposomes loaded with 5 bacteriorhodopsin molecules, and liposomes loaded with 10 bacteriorhodopsin molecules. The training set consisted of 23625 diffraction patterns, and test set of 7875 patterns. We found that all networks used in our

study were able to identify the number of protein molecules in liposomes independent of the liposome radius. Our findings make this approach rather promising for the usage of liposomes as protein carriers in XFEL experiments.
Conclusion: Thus, the performed numerical experiments show that the use of neural network algorithms for the recognition of diffraction images from single macromolecular particles makes it possible to determine changes in the structure at the angstrom scale.

Key Words: X-ray crystallography • diffraction pattern • protein structure • neural networks.

Corresponding author: Valery Novoseletsky. E-mail: valery.novoseletsky@yandex.ru