School of Computer Science and Engineering
Presentation

Adil Al-Azzawi

Ph.D. in Electrical Engineering and Computer Science
University of Missouri-Columbia

CSE Faculty Candidate
Wednesday, January 29, 2020
Time: 10:30am-11:30am
JB-359

CRYO-EM BASED PROTEIN STRUCTURE MODELING

Protein represents 17% of the human body, but it is one of the most important components. It is a key component of all cells used for building and repairing tissues, making enzymes, and hormones. Protein is the essential building block of bones, muscles, cartilages, skin, and blood. Therefore, a large quantity of protein is always needed. However, protein cannot be stored in the body as it is the case for fat. Instead, every time a protein is needed it is created. A copy of every protein needed by the body is stored in the DNA. For space efficiency, proteins are stored in the form of a sequence of nucleotides that can easily be converted into a sequence of amino acids which is known as the protein primary structure. Every two amino acids are connected through the backbones atoms N (Nitrogen), C-α (carbon- α) and C (Carbon).

For a protein to performs its job, it needs to be in a three-dimensional structure. Also, known as the protein tertiary structure. The human body can readily convert a protein from its primary structure to its tertiary structure. However, this process, (operation) of the inferencing protein tertiary structure from its primary structure, has preoccupied scientists for more than fifty years. This problem is known as protein structure prediction. In between primary structure and tertiary structure there exist a secondary structure, which mainly consists of complex protein tertiary structure that may contain α-helix and β-sheet. The protein structure prediction problem is the inference of the protein three-dimensional (tertiary) structure from its amino acid sequence (primary structure). Initially, scientists tried to solve this problem experimentally. Several methods were developed for this purpose. The most important ones among them are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and recently Electron Microscopy (EM). Each of these methods has achieved some success. However, they share some common disadvantages. These methods require complicated procedures that are hard to implement except by well-trained specialists. They are time-consuming and labor-intensive. Results can easily be contaminated. The detection of one protein structure can take years which costs a lot of money and involves the use of expensive equipment that needs to be maintained which in place cost more money. Therefore, an alternative approach is desperately needed. This new approach needs to be easy to use, does not consumes a lot of time and money, and does not necessitate the use of complex and expensive equipment. Solving this problem can lead to breakthroughs in medicine and biotechnology. In medicine new and better drugs can be designed, which will increase its efficiency and reduce its side effects. Whereas for biotechnology, new and more efficient enzymes can be designed which impact many areas of our daily life such as detergents, textiles, food and beverages, leather, and bioethanol.

Machine learning (ML) based unsupervised, supervised, and deep learning approaches have had a big impact on many areas such as video and audio applications, healthcare, commercial, education, and many more. It is regarded by many as the solution of today's problems. Implementation of Machine Learning approaches is easy, and it does not require the design and implementation of complex design. Instead, all design non-learnable and learnable models such as unsupervised and supervised learning are needed.

In general, first, an essential step of the molecular (protein) structure determination process is the single-particle picking (2D particles). Protein particle shapes in the most cryo-EM datasets are either common shape – circle (top view) or square (side view), in addition to the noise, contaminants, and ice object particle shapes. Top and side-view particles are even overlapped, or some additional objects are attached to the original particles. Another common protein particle shape in very low SNR cryo-EM images is either complex or irregular shapes that face two main problems. First, particles in the cryo-EM appear in non-structural object shapes. That makes template matching algorithms unable to distinguish between the objects and the background. In additional to that the particles in the very low SNR cryo-EM images have almost the same intensity level of the background. The single-particle picking step is a labor-intensive step in the computational molecular reconstruction procedure and is a major obstacle for the automated cryo-EM pipeline. In the past, particles from cryo-EM micrographs are often selected manually. A manual picking process is usually a laborious, tedious, and time-consuming task that inevitably requires a considerable amount of human effort to obtain a sufficient number of good-quality particles. To overcome this issue, first, various machine learning models based unsupervised and supervised learning are designed for semi-automated or semi fully automated single-particle picking, such as RELION 3.0 [1] and EMAN 2.21 [2]. Recently, full automated approaches for single-particle picking are proposed such as AutoCryoPicker [3], a fully automated particle picking approach based on image preprocessing, unsupervised clustering and shape detection. SuperCryoEMPicker [4], a fully automated super particle clustering method for picking particles of complex and irregular shapes in cryo-EM images. DeepCryoPicker, a fully automated deep neural network for single-particle picking in cryo-EM [5].

Second, to ensure a high-resolution 3D density map from 2D particle images, a large number of single-particle images are extracted. According to the different biological macromolecules data, good particle selection become more difficult. The reason is that the selected particles are similar copies of different views, different conformation, or either some of them are partially damaged. Also, single-particle images are very noisy, so image averaging is the process that is used to improve the quality of the particle images based on improving the single-to-noise-ratio. For this reason, the class average is used to represent a group of particle images that all particles look the same (2D multiclass). This stage requires two steps; particle image classification step for grouping the particle images into homogenous subsets (same particle view), and particle alignment step to make each subset have the same view. Most of the existing tools such as RELION 3.0 [1] and EMAIN 2.21 [2] are based on the reference-based manual selection to build a reliable 3D density map. Recently, DeepCryoMap [6] a fully automated approach is proposed. It is a fully automated cryo-EM particle alignment and 3D density maps reconstruction based deep supervised and unsupervised learning approaches.

[1] Jasenko Zivanov, Takanori Nakane, Björn O Forsberg, Dari Kimanius,Wim JH Hagen, Erik Lindahl, and Sjors HW Scheres, “New tools for automated high-resolution cryo-EM structure determination in RELION-3”, eLife. 2018; 7: e42166. Published online 2018 Nov 9. doi: 10.7554/eLife.42166

[2] James M. Bell, 2016, “High Resolution Single Particle Refinement in EMAN2.21”, Methods.

[3] Adil Al-Azzawi, Anes Ouadou, John J. Tanner, and Jianlin Cheng, “AutoCryoPicker: an unsupervised learning approach for fully automated single particle picking in cryo-EM images”, BMC Bioinformatics, accepted, (2019).

[4] Adil Al-Azzawi, Anes Ouadou, Jianlin Cheng, “A Super Clustering Approach for Fully Automated Single Particle Picking in Cryo-EM”, Genes (Basel). 2019 Aug 30;10(9). pii: E666. doi: 10.3390/genes10090666.

[5] Adil Al-Azzawi, Anes Ouadou, Highsmith Max R, John J. Tanner, and Jianlin Cheng, “DeepCryoPicker: Fully Automated Deep Neural Network for Single Protein Particle Picking in cryo-EM”, bioRxiv preprint first posted online Sep. 10, 2019; doi: http://dx.doi.org/10.1101/763839.

[6] Adil Al-Azzawi, Anes Ouadou, Ye Duan, John J. Tanner, and Cheng, Jianlin, “DeepCryoMap: Fully Automated cryo-EM Particles Alignment Approach for 3D Density Maps Reconstruction Based Deep Supervised and Unsupervised Learning Approaches”, (To be submitted Feb/March. 2020).