A Rapid and Robust Network-Based Approach to Reveal the 3D Discrete Conformations of Protein Using Cryo-EM

鍾思齊
國立中山大學應用數學系

摘要

Recently, computational and learning methods are becoming popular to determine the 3D structure of a protein. The most remarkable achievement is arguably the recent release of AlphaFold 2, which can perform high accuracy structure prediction from chains of amino acid sequences. However, these methods are mainly used to predict a static structure that can not reveal the conformational changes of the protein. It is noted that the functionality of protein is highly related to the underlying conformational change, and describing it is still a challenging task. Among the tools available, cryo-EM is a promising computational technique with high efficiency to perform conformation analysis. However, the data characteristics include heavy noise, huge dimension, a large number of unlabeled samples (no clean target is available for training) with unknown orientations, making it very challenging to reach a robust computation conclusion for heterogeneity analysis. Traditional approaches address this problem at the 3D level, which is computationally expensive and may not be applicable to datasets whose conformations differ significantly. Therefore, there is a need to develop a new algorithm that preserves accuracy while solving the scalability issue due to the fast-growing data acquisition rate. In this talk, I will first introduce the importance of 3D structure determination of protein and the related background of cryo-EM image processing. Second, I will elaborate on our approach to the heterogeneity problem, which utilizes network analysis to partition the dataset into several homogeneous communities. Specifically, I will discuss several novel criteria to measure the conformation similarity between cryo-EM images. Finally, I will discuss some ongoing research directions based on self-supervised learning.