Advances in modelling continuous heterogeneity from single particle cryo-EM data

Ali Punjani[1], David Fleet[2]
[1]University of Toronto [2]University of Toronto
alipunjani@cs.toronto.edu

Single particle cryo-EM excels in determining static structures of biological macromolecules such as proteins. However, many proteins are dynamic, with their motion inherently linked to their function. Recovering the continuous motion and detailed 3D structure of flexible proteins from cryo-EM data has remained an open challenge. In this talk, we describe two new algorithms that allow both motion and structure of flexible proteins to resolved from cryo-EM data.

First, 3D variability analysis (3DVA), an algorithm that fits a linear subspace model of conformational change to cryo-EM data at high resolution. 3DVA enables the resolution and visualization of detailed molecular motions of both large and small proteins, revealing new biological insight from single particle cryo-EM data. Experimental results demonstrate the ability of 3DVA to resolve multiple flexible motions of α-helices in the sub-50 kDa transmembrane domain of a GPCR complex, bending modes of a sodium ion channel, five types of symmetric and symmetry-breaking flexibility in a proteasome, large motions in a spliceosome complex, and discrete conformational states of a ribosome assembly. 3DVA is implemented in the cryoSPARC software package.

Second, 3D Flexible Refinement (3DFlex), a motion-based deep neural network model of continuous heterogeneity. 3DFlex directly exploits the knowledge that conformational variability of a protein is often the result of physical processes that transport density over space and tend to conserve mass and preserve local geometry. From 2D image data, the 3DFlex model jointly learns a single canonical 3D map, latent coordinate vectors that specify positions on the protein's conformational landscape, and a flow generator that, given a latent position as input, outputs a 3D deformation field. This deformation field convects the canonical map into appropriate conformations to explain experimental images. Applied to experimental data, 3DFlex learns non-rigid motion spanning several orders of magnitude while preserving high-resolution details of secondary structure elements. Further, 3DFlex resolves canonical maps that are improved relative to conventional refinement methods because particle images contribute to the maps coherently regardless of the conformation of the protein in the image. Together, the ability to obtain insight into motion in macromolecules, as well as the ability to resolve features that are usually lost in cryo-EM of flexible specimens, will provide new insight and allow new avenues of investigation into biomolecular structure and function.