TomoDRGN: resolving heterogeneous protein complexes in situ using cryo-ET and deep learning

Barrett Powell[1], Joey Davis[2]
[1]MIT [2]MIT
bmp@mit.edu

Proteins and their complexes routinely change their conformation and composition when performing essential cellular functions. The range of heterogeneity accessible to a given protein between its synthesis and degradation defines its structural landscape, and a key goal of structural biology is to determine this full ensemble of structures. Here, we present tomoDRGN, a machine learning based tool designed to resolve ensembles of highly heterogeneous structures in native cellular environments, as imaged by cryo-ET.

Cryo-EM single particle reconstruction and cryo-ET sub-tomogram averaging algorithms traditionally rely on the simplifying assumption that all images are derived from projections of one (or k) distinct underlying volumes. These approaches are ill suited when analyzing complexes undergoing continuous motions, and often fail to reveal functionally important but low abundance discrete structural states. To complement these traditional tools, our lab recently developed the software cryoDRGN which employs a variational autoencoder (VAE) framework to produce an ensemble of continuously heterogeneous volumes using single particle cryo-EM images. Here we extend the cryoDRGN VAE framework to subtomographic data, thereby enabling analysis of molecular machines undergoing continuous structural changes in their native, spatially contextualized cellular environments. Using simulated data, we first assess the efficacy of the tomoDRGN approach. We then apply tomoDRGN to analyze ribosomes extracted from the M. pneumoniae dataset EMPIAR-10499, finding that this framework produces volumes at nearly equivalent resolutions to traditional sub-tomogram averaging. Additionally, we find tomoDRGN readily resolves an ensemble of functionally relevant states of the ribosome. TomoDRGN inherits a series of analysis and particle filtering methods from the existing cryoDRGN toolkit, and benefits from new tools that map structural heterogeneity back to the underlying cellular tomograms. We anticipate tomoDRGN will be broadly useful to structural biologists interested in analyzing dynamic protein complexes in their native cellular environment.