

Recent algorithmic advances for single-particle cryo-EM

Ali Punjani^{1,2}

¹. University of Toronto, Department of Computer Science, Toronto, Canada.

². Structura Biotechnology Inc., Toronto, Canada.

alipunjani@cs.toronto.edu

Single particle cryo-EM (cryo-electron microscopy) allows high-resolution imaging of macromolecular complexes in close-to-native state, at near-atomic resolutions. Cryo-EM has undergone several technological breakthroughs in microscopy, electron detectors, and image processing that have enabled its use recently in solving high-resolution structures of difficult proteins and complexes [1, 2]. As cryo-EM makes rapid progress, several key challenges remain in the quest for higher resolutions on challenging targets, as well as in the widespread and routine use of cryo-EM. Here, we introduce new algorithmic advances in the cryo-EM data processing pipeline, along with implementations within the cryoSPARC software system [3], to address some of these challenges.

The image processing pipeline in single particle cryo-EM is required to solve the 3D electron density of a target molecule, in potentially many conformational states, from noisy 2D images collected using cryo-TEM. Each collected image is a movie of dose-fractionated frames that require motion estimation and correction. The corrected images (micrographs) are then used to estimate the microscope CTF during the exposure, as well as to find and pick out single particles. The single particles are extracted from the micrographs, and then are sorted and filtered using 2D classification methods. The resulting filtered particle stacks are used to perform *ab initio* 3D structure determination of potentially multiple discrete states or targets. These coarse structures are then further classified and refined in 3D to yield interpretable molecular density maps and achieve state of the art resolutions.

Here we introduce new robust methods for multiple stages in the processing pipeline, along with fast implementations of these new algorithms in the cryoSPARC software system.

Conformational heterogeneity in a sample is a common and interesting phenomenon that can be detected and exploited by cryo-EM algorithms in order to resolve multiple conformational states of a target molecule. Often, protein molecules that have a mechanism of action that involves motion will have flexible parts, inducing continuous conformational heterogeneity rather than discrete heterogeneity. Computationally dealing with flexible heterogeneity is a cutting-edge area of research within the community, and we introduce new tools for diagnosing and recovering flexibility information from cryo-EM single particle images. These new algorithms are able to provide insight into the mechanism and coupling of different parts of a flexible protein molecule.

References:

[1] D Cressey and E Callaway, *Nature* **550** (2017), p. 167

[2] X Bai, G McMullan, and S Scheres, *Trends in Biochemical Sciences* **40** Issue 1 (2017), p. 49-57

[3] A Punjani *et al*, *Nature Methods* **14** (2017), p. 290–296