2328
doi:10.1017/S1431927618012126
Microsc. Microanal. 24 (Suppl 1), 2018

Single-particle CryoEM: Data Processing Techniques for Obtaining Optimal Results

Ali Punjani[1,2]

[1.] University of Toronto, Department of Computer Science, Toronto, Canada.
[2.] Structura Biotechnology Inc., Toronto, Canada.

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]. Cryo-EM is making rapid progress, both in the quest for higher resolutions on challenging targets, as well as in the widespread and routine use of cryo-EM. In this tutorial, you will learn about the principles and techniques used to process cryo-EM image data to obtain optimal map quality. The tutorial will demonstrate these techniques using the cryoSPARC software system [3].

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.

Each of the processing stages above can have parameters and inputs that requires decisions made by the scientist, using structural or other insight as prior knowledge to guide processing. Typical workflows on difficult molecules often involve multiple iterations of segments of the processing pipeline, each time with changes in the selection of input data or parameters, to yield optimal results. In particular, curating data and separating a sample *in silico* into multiple homogenous subsets representing different conformations can require knowledge both of the target and of the underlying algorithms. This tutorial will describe the critical parameters and their effects on the above processing stages, strategies for iteratively improving a processing workflow, and tips and tricks that can aid the user. The cryoSPARC software system will be described in the tutorial and includes fast and robust algorithms for all of the above steps along with tools to aid in automation and reproducibility of processing workflows.

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

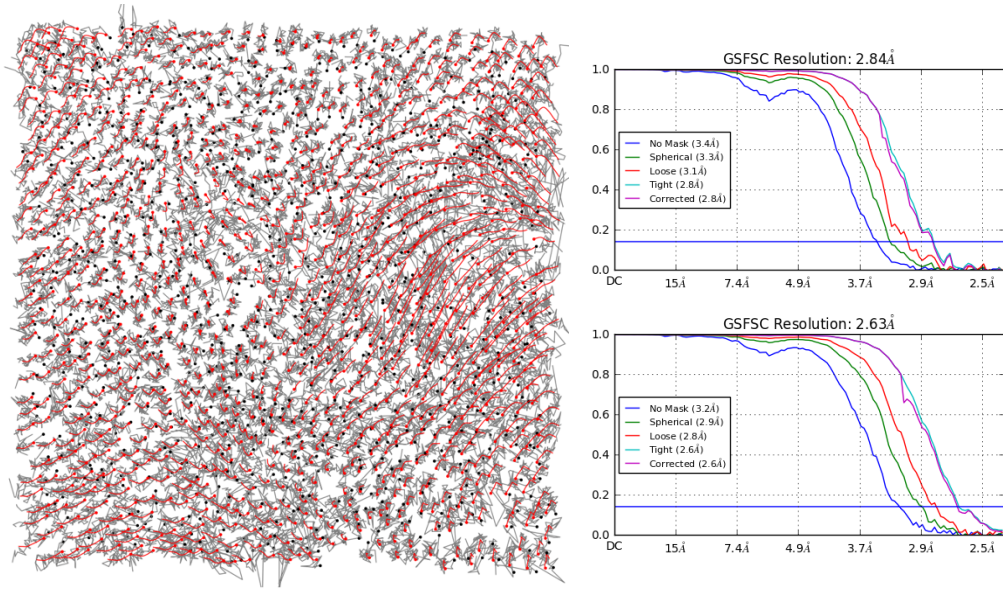

Figure 1. Left: Example of per-particle estimated motion trajectories on a typical single particle cryo-EM movie of 20S proteasome, showing both the raw estimated trajectories (gray) and the correctly smoothed trajectories (red), both shown at 40x scale. Right: Fourier Shell Correlation (FSC) plots for (top) a refinement done with over-smoothing, causing linear per-particle motion trajectories and (bottom) with correct smoothing, showing a clear improvement from 2.8A to 2.6A resolution, as an example of the sensitivity of result quality to parameter settings in cryo-EM data processing.

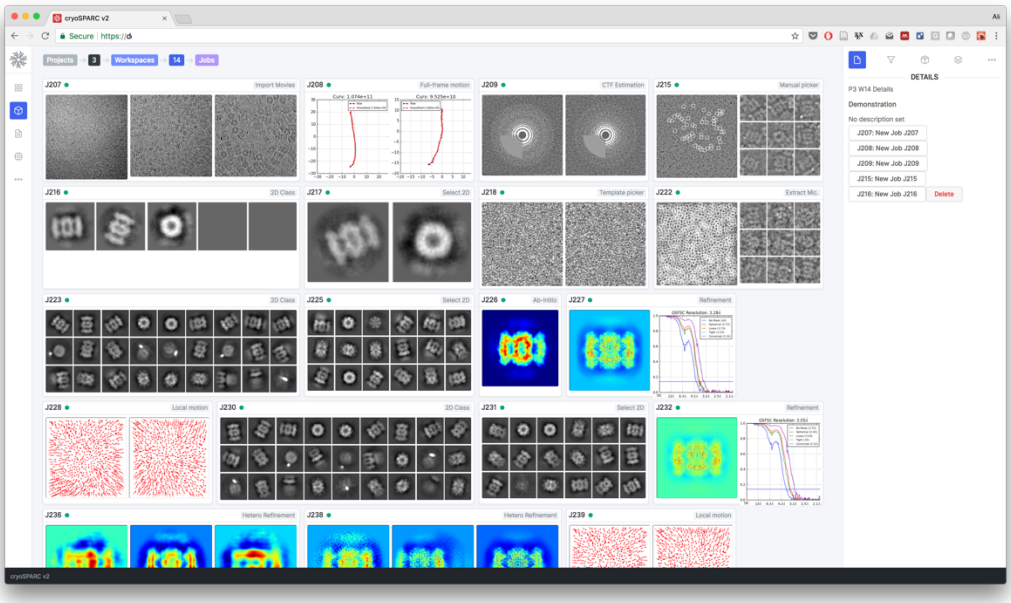

Figure 2. An example of a complete cryo-EM data processing workflow in the cryoSPARC software package, starting from motion correction of dose-fractionated movie frames and ending with a refined structure and resolution estimate.

https://doi.org/10.1017/S1431927618012126 Published online by Cambridge University Press