Expanding the Computational Toolbox for Single-Particle Cryo-Electron Microscopy

Alberto Bartesaghi[1,2,3,4,5*]

[1.] Department of Computer Science, Trinity College of Arts & Sciences
[2.] Department of Biochemistry, Duke University School of Medicine
[3.] Department of Electrical and Computer Engineering, Pratt School of Engineering
[4.] Duke Institute for Brain Sciences, Computational Biology and Bioinformatics Program
[5.] Duke University, Durham, NC, USA.
* Corresponding author: alberto.bartesaghi@duke.edu

Recent technological advances have sparked a revolution in single-particle cryo-Electron Microscopy (EM) enabling the study of important biomedical targets that were previously inaccessible to other technologies. The structures of an impressive number of targets, including large dynamic assemblies, membrane proteins, and complexes bound to small molecules have been successfully analyzed by cryo-EM at unprecedented levels of detail. The resolutions achieved in some cases rival those obtained by X-ray crystallography and are high enough to reveal features that may be key for drug discovery [1], [2]. The combination of automated strategies for data collection with recent advances in camera technology, have dramatically improved the rate of data production at modern microscopy facilities, driving the development of computational tools that can significantly facilitate the process of structure determination. Several software packages exist in the field today that allow users to successfully convert the noisy image data into high-resolution structures by providing effective tools for data analysis that can characterize and sort out the intrinsic conformational heterogeneity present in single-particle cryo-EM samples.

In this tutorial, I will present an overview of the different steps in the single-particle cryo-EM processing pipeline required to convert the raw movie data into density maps and describe how user decisions may affect the outcome of the structure determination procedure. In the era of fully automated data collection where fluctuations in data quality are unavoidable, the first critical step is to identify the subset of usable micrographs that will be used for the downstream processing. I will describe robust and effective strategies to sort electron micrographs and automatically select particle images which can then be subjected to 2D and 3D classification, **Figure 1**. I will also present methods for frame alignment and CTF estimation of individual molecular images and describe the use of data-driven approaches for exposure-weighting to optimize the extraction of high-resolution information from raw electron micrographs, **Figure 2**. I will also highlight some of shortcomings and remaining challenges in terms of computational data analysis that will need to be addressed in order to expand the capabilities of single-particle cryo-EM.

References:

[1] A Bartesaghi et al., Structure **26 (6)** (2018), p. 848.
[2] S Banerjee et al., Science **351 (6275)** (2016), p. 871.
[3] Y Kang et al., Nature **558** (2018), p. 553.
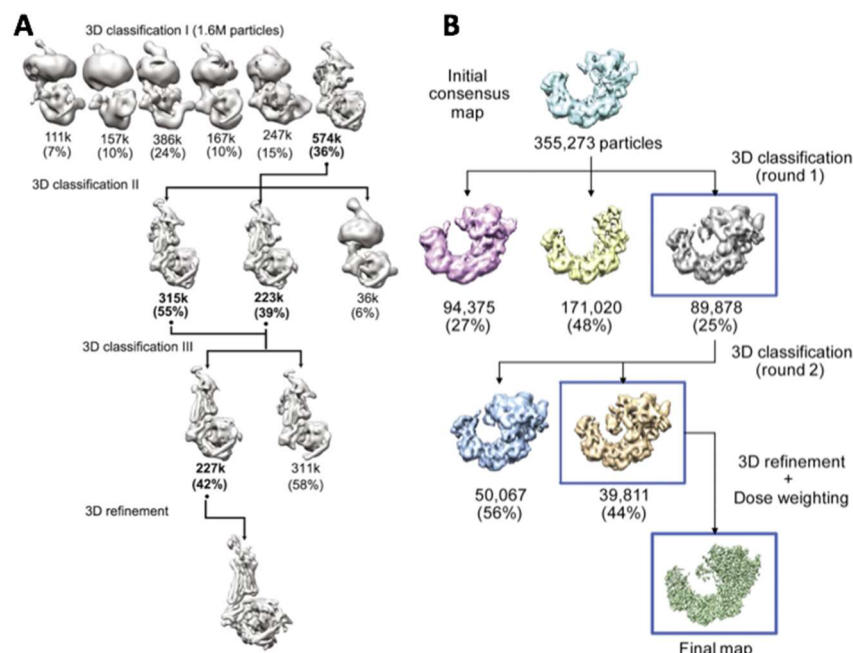[4] T Guo et al., Cell **171 (2)** (2017), p. 414.

Figure 2. Identification of homogeneous particle populations using supervised image classification. A) Successive rounds of 3D classification used to determine the cryo-EM structure of the rhodopsin–G_i–Fab complex at a nominal global resolution of 4.5 Å [3]. B) Overview of the image refinement protocol used to obtain the structure of Csy surveillance complex bound to dsDNA. An initial consensus map was obtained from the starting set of 355,273, followed by two successive rounds of 3D classification yielding a final homogeneous set of 39,811 particles resulting in a final density map at 2.9 Å resolution [4].
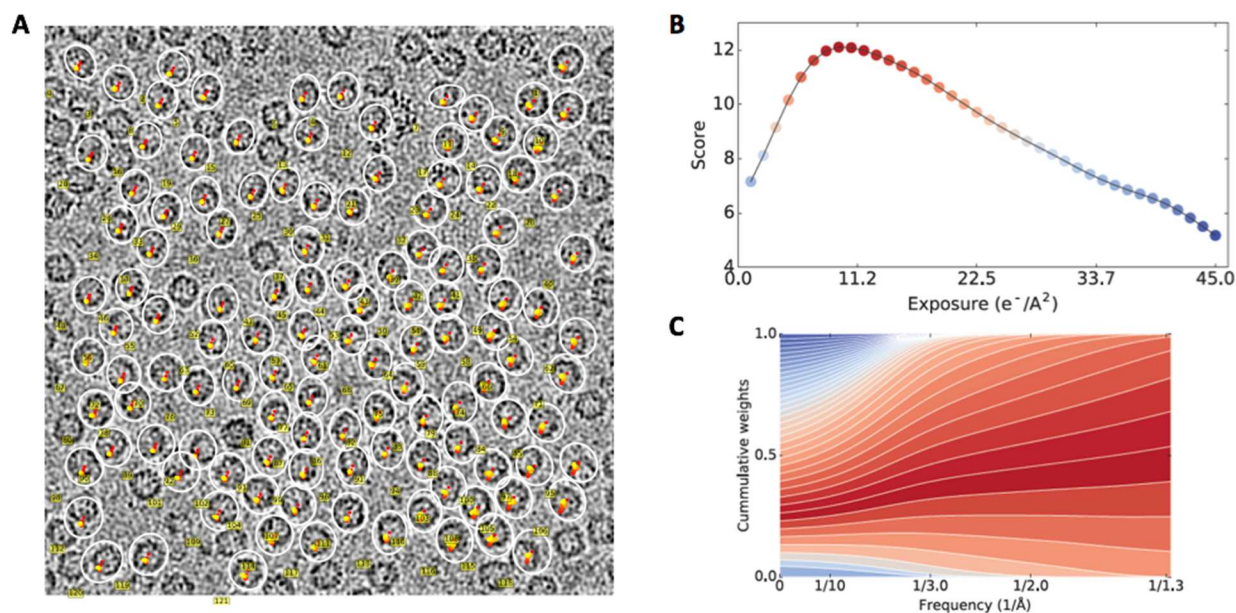

Figure 1. Data-driven refinement strategies to improve the resolution of cryo-EM maps. A) Image of apoferritin from EMPIAR-10216 showing the location of individual particles with overlaid drift-trajectories and defocus changes across the field of view. B) Plot of average particle-scores vs. cumulative electron dose. Scores are computed by matching each particle-frame against re-projections of the map [1]. C) Corresponding frequency-space weights showing the relative contribution of individual frames to the weighted averages used during refinement.