

Unsupervised Determination of the Number of Conformations in Single-particle Cryo-EM

Ye Zhou¹, Amit Moscovich², Priyamvada Acharya¹ and Alberto Bartesaghi¹

¹Duke University, Durham, North Carolina, United States, ²Princeton University, Princeton, New Jersey, United States

Single-particle cryo-EM is a powerful imaging modality used to determine the structure of biomedically important macromolecular complexes. The underlying principle for three-dimensional reconstruction consists in combining projection views corresponding to multiple copies of the target of interest taken from random orientations. While well-behaved and homogeneous samples can readily yield high-resolution reconstructions [1], most realistic targets are often heterogeneous, flexible or dynamic, making it challenging to obtain high-quality density maps. Successful analysis of such datasets relies on the use of laborious protocols for 2D and 3D image classification aimed at identifying homogeneous subpopulations of particles that when combined in 3D can yield high-resolution structures. This supervised process of image sorting is not only prone to bias, difficult to reproduce and computationally inefficient, but also incompatible with modern strategies for high-throughput data collection and analysis used in single-particle cryo-EM. We have previously addressed the problem of unsupervised structure refinement in the case of homogeneous samples [2], and showed that high-resolution structures of even challenging samples like HIV-1 Env can be obtained without user supervision [3]. Here, we extend these ideas to the case of samples that display discrete heterogeneity, i.e., where a finite but unknown number of conformations are present in the data. To address this problem, we propose a strategy to automatically determine the number of distinct conformations present in a mix using a simple statistical test based on particle scores measured against multiple 3D references. Results obtained on several challenging datasets demonstrate that our strategy can be used to successfully recover the correct number of discrete conformations present in heterogeneous datasets without supervision, paving the way for the development of high-throughput strategies for single-particle cryo-EM data analysis.

References

- [1] A. Bartesaghi *et al.*, “Atomic Resolution Cryo-EM Structure of β -Galactosidase,” *Structure*, vol. 26, no. 6, pp. 848-856.e3, 2018.
- [2] Y. Zhou, A. Moscovich, T. Bendory, and A. Bartesaghi, “Unsupervised particle sorting for high-resolution single-particle cryo-EM,” *Inverse Probl.*, 2019.
- [3] R. Henderson *et al.*, “Disruption of the HIV-1 Envelope allosteric network blocks CD4-induced rearrangements,” *Nat. Commun.*, vol. 11, no. 1, p. 520, 2020.

