1292
doi:10.1017/S1431927619007190
Microsc. Microanal. 25 (Suppl 2), 2019

Computational Methods to Process Highly Heterogeneous Cryo-EM Samples

Josue Gomez-Blanco[1], Satinder Kaur[1], Joaquin Ortega[1] and Javier Vargas[1*]

[1.] Department of Anatomy and Cell Biology, McGill University, Montreal, Canada.
* Corresponding author: javier.vargasbalbuena@mcgill.ca

Currently, cryo-electron microscopy (cryo-EM) has the potential to obtain quasi-atomic 3D reconstructions of macromolecules showing different physiological conformations. However, current image processing tools fall short in dealing with datasets presenting extensive heterogeneity. This work focuses on developing new computational tools to visualize and deconvolute the entire landscape of conformations sampled by immature ribosomes in bacteria.

Existing classification methods require *a priori* assumptions by the user, including the number of classes or the number of refinement iterations. They also require an initial volume to prime the algorithm [1]. The use of a unique initial volume in samples affected by extensive heterogeneity is incorrect. Note that 3D classification approaches are local optimizers [2]. Thereby, the outcome of the classification highly depends on the suitability of the starting model. Inaccurate initial volumes will lead to incorrect results. Our preliminary data shows that the selection of multiple accurate initial volumes is essential to fully sort out all the conformers present in the dataset. As shown in Fig 1, we first classified a dataset of ~300,000 particles containing a mixture of immature 50S ribosomal subunits (44S particles) using the mature 50S subunit structure as initial model. We obtained the structure of only one assembly intermediate (Class III in Fig. 1A) from 16% particles that refined to a moderate resolution (~5Å). The remaining particles (84%) comprised a heterogeneous mixture of earlier assembly intermediates. However, their refined structure could not be obtained. Next, we attempted to compute three initial maps from the dataset by intensive 2D reclassification strategies, multiple rounds of our RANSAC method [3], as well as visual evaluation and validation. We then used these maps to prime the classification algorithm obtaining ten distinct assembly intermediates representing 94% of the particles in the dataset. These classes refined to 3-4Å resolution (Fig 1B). These results illustrate the need for methodologies that can automatically produce multiple accurate initial maps in samples affected by extensive heterogeneity.

Additionally, state-of-the art 3D classification approaches [1-2,4] are affected by the "attractor" problem [5-6] that limits the number of different classes that can be obtained from the data, irrespectively of the number of classes existing in the dataset. This limitation comes because major classes, with higher SNR, "attract" particles from other classes, thus, low abundancy particle subpopulations cannot be extracted by standard classification procedures [5-6]. In addition, in samples affected by massive heterogeneity the remanent heterogeneity in the obtained 3D classes limits the resolution of these reconstructions.

In this work, we present novel image processing strategies to obtain multiple accurate *ab initio* initial volumes and to reconstruct many different 3D classes, overcoming the attraction problem, that can be used to process cryo-EM samples affected by massive heterogeneity. We have applied these methods to the problem of capturing the entire landscape of conformations sampled by immature ribosomes in bacteria. These methods were able to face the massive heterogeneity of these datasets [7] as show in Fig. 2 [8].

References:

[1] SWH Scheres, J Struct Biol **180** (2012), p. 519.
[2] SWH Scheres, Methods Enzymol **579** (2012) p. 125.
[3] J Vargas et al., Bioinformatics **30** (2014) p. 2891.
[4] A Punjani et al., Nature methods **14** (2017) p. 290.
[5] COS Sorzano et al., J Struct Biol **171** (2010) p. 197.
[6] J Wu et al., PLoS One **12** (2017) p. 197.
[7] A Razi et al., Bioarxiv (2019), http://dx.doi.org/10.1101/525360. T
[8] The authors acknowledge funding from NSERC Discovery Grant (RGPIN-2018-04813) and FRQNT New University Researchers Start-Up (NC-253837).

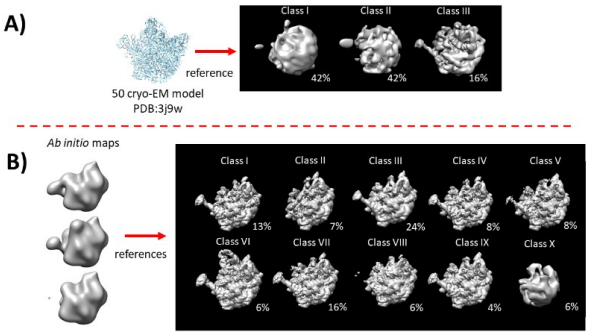

Figure 1. Classification results obtained when using an initial map derived from PDB 3J9W (A) or from ab initio RANSAC method (B).

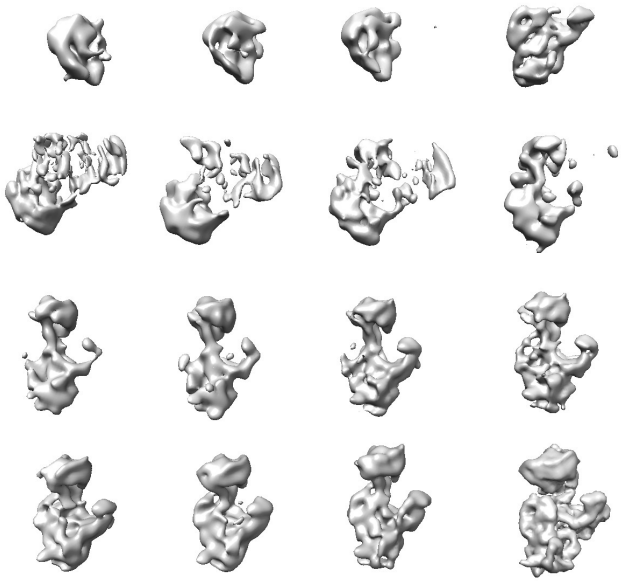

Figure 2. Cryo-EM maps obtained from a sample containing purified 30SEra-depleted particles using our proposed 3D image classification strategy. This result has been pre-published in [7].

https://doi.org/10.1017/S1431927619007190 Published online by Cambridge University Press