

Evaluation of Automated Particle Picking for Cryo-EM Using High Precision TEM Simulation by Utilizing a Multi-slice Method

Masataka Ohashi¹, Fumio Hosokawa¹, Takao Shinkawa¹ and Kenji Iwasaki²

¹BioNet Lab. Inc., Tachikawa, Tokyo, Japan, ²University of Tsukuba, Tsukuba, Ibaraki, Japan

We report an automated particle picking software for cryo-electron microscopy (cryo-EM) and the evaluation of this software using high precision transmission electron microscope (TEM) images simulated by utilizing a multi-slice method. Cryo-EM single-particle analysis is a method to determine three-dimensional (3D) structures of proteins where the structures are estimated from observed TEM images of proteins embedded in thin vitreous ice. In single-particle analysis, the particles are picked from each of the acquired micrographs. Template matching [1] is a popular approach to automatically pick particles in cryo-EM micrographs since a large set of particle images are needed to estimate 3D structures of proteins. However, it has certain limitations. In template matching, the particles in the acquired micrographs need to be visible to the eye as the templates are created by picking particles manually. Micrographs are generally acquired at a large defocus setting (e.g. a few micrometers) where particles are easily visible owing to the fact that the larger is the defocus, the easier it is to observe particles. However, in the process of deconvolution of the contrast transfer function, which is necessary to estimate 3D structures, a large defocus will cause major errors in the estimated structures than a small defocus. Moreover, an additional issue is that the templates only have projected images from certain orientations, which is either due to a preferred orientation of particles or because only visible particles are picked preferentially.

In this report, we suggest an automated particle picking method to address the aforementioned issues. In this method, automated particle picking is performed by repeatedly processing compensation of positions of a particle, 2D classification, and the selection of class-average images. The composition of our method is quite simple. This method is available for particles that are not clearly visible and is able to pick particles irrespective of their orientations. In addition, if our method could not pick any of the particles, it can be concluded that those micrographs are unsuitable for 2D classification. Hence, we can avoid wasting time at later stages of the analyses.

In the case of evaluation of an algorithm related to cryo-EM with various conditions, it is not easy to acquire the desired evaluation data. Hence, for evaluation purposes, we generated simulated data by using elbis (BioNet Lab. Inc.), which is a software used to generate precise TEM images with Graphics Processing Unit (GPU) by utilizing a multi-slice method and transmission cross coefficients [2]. We generated test data with a defocus of 15000 Å, which is commonly used for acquiring micrographs, and a defocus of 2500 Å, which is a small enough value to generate projected images of particles with low visibility. Then, we randomly tiled the images that were generated by elbis with various projected orientations and created a simulated micrograph. We used an atomic model of hemoglobin to generate the projected images. Moreover, we picked particles by our method from the simulated micrographs and performed 2D classification, 3D classification, and 3D refinement in this evaluation. We used RELION [3] as a 3D structure determination software.

Fig. 1 shows the result of the evaluation. Figs. 1A and 1B are examples of particle images simulated with defocus values of 2500 Å and 15000 Å, respectively. The particle image with a defocus of 15000 Å (Fig. 1B) is visible and the particle image with a defocus of 2500 Å (Fig. 1A) is not clearly visible. Figs. 1C and 1D are 2D class-averaged images with defocus values of 2500 Å and 15000 Å, respectively. Figs. 1E

and 1F are the estimated 3D structures from datasets with defocus values of 2500 Å and 15000 Å, respectively. These results show that our method could clearly pick particles with a small defocus (2500 Å) properly. The estimated 3D structure of the small defocus dataset (2500 Å) can be more detailed than the estimated 3D structure of the large defocus dataset (15000 Å). Moreover, we will show the evaluation using a dataset in which the projected orientation is ununiformed.

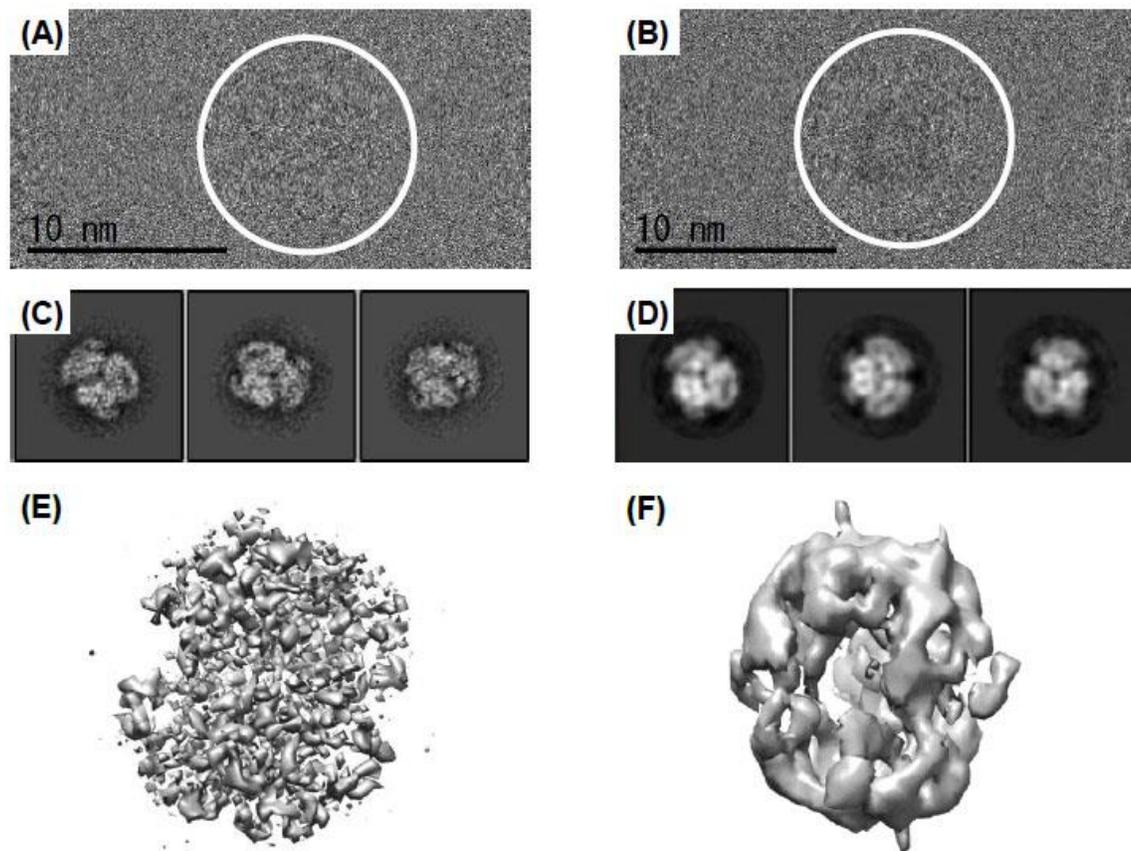


Figure 1. Simulated particles of hemoglobin (A, B), 2D class-averaged images (C, D), and estimated 3D structures (E, F): (A), (C), and (E) are related to a defocus value of 2500 Å; (B), (D), and (F) are related to a defocus value of 15000 Å. There are particles in white circles of (A) and (B).

References

- [1] J. Frank et al., *Ultramicroscopy* 12 (1983) 169-175.
- [2] F. Hosokawa et al., *Ultramicroscopy* 158 (2015) 56-64.
- [3] S. Scheres, *J Struct Biol*, 180(3) (2012) 519-530.