

Testing and implementing a live processing workflow at the New York Structural Biology Center

Eugene Chua¹, Joshua Mendez², Reza Paraan², Huihui Kuang², Kashyap Maruthi², Edward Eng¹, Shaker Krit², Anchi Cheng², Clint Potter¹ and Bridget Carragher¹

¹New York Structural Biology Center, New York, New York, United States, ²New York Structural Biology Center, United States

The Simons Electron Microscopy Center (SEMC) and National Center for Cryo-EM Access and Training (NCCAT) are located at the New York Structural Biology Center (NYSBC) in Manhattan, New York. SEMC serves researchers from the nine New York member institutions, while NCCAT provides access to cryo-EM training and instrumentation for researchers across the country. Together, we house seven Titan Krios microscopes, most of which are equipped with a K3 camera and an energy filter. In addition, two Krioses have Falcon 3 cameras, and one has a spherical aberration (Cs) corrector. Lastly, our newest Krios is equipped with a Falcon 4 camera. We also have a Glacios screening microscope, multiple side-entry electron microscopes, and a tabletop scanning electron microscope (SEM). Each Krios microscope collects 5,000 to 10,000 high magnification images per day and serves a different user every 2 to 7 days, all year round.

We currently use Appion (Lander et al, 2009) as our live preprocessing tool. Appion performs motion correction and dose weighting through MotionCor2 (Zheng et al, 2017), and CTF estimation through Ctfind4 (Rohou & Grigorieff, 2015). Appion runs on computational servers that intake micrographs from the microscope camera, preprocesses them, then outputs the motion corrected files along with their drift and CTF information into a database that can be conveniently viewed through a web browser. This preprocessing pipeline and database structure is tightly integrated with our data collection software, Leginon (Suloway et al, 2005). This pipeline allows users and microscopists to track the quality of the micrographs being collected in real time. This real-time pre-processing information about the data being collected can then be used to make adjustments to the microscope settings to maximize data quality and acquisition throughput, or to change sample as required.

Additional on-the-fly feedback during data collection through 2D classes and 3D reconstructions can provide valuable information on particle and reconstruction quality that goes beyond the quality metrics provided by motion correction and CTF estimation. This will allow users and microscopists to maximize microscope time, for instance by changing to a different grid once the collection reaches a target resolution, or by assessing from 2D classes if there is preferred orientation, then changing to a region with different ice thickness, or to a grid with different conditions. Because of our large daily volume of data collection and wide user base, the ideal live processing package should be fast, easy to use, and robust.

There are now several software packages that have been developed to provide live feedback during data collection. These include Warp (Tegunov & Cramer, 2019), RELION Live (Zivanov et al, 2018), and CryoSPARC Live (Punjani et al, 2017). Here, we compare each software package using different data that vary in protein composition, image size, and other data collection parameters including number of frames and cameras used. To benchmark the performance of each package, we use metrics including processing time at each step, resources consumed, number of human decisions, and ease-of-use and integration into our existing workflow. We compare the pros and cons of each program, and show how the results from our testing and implementation help improve the quality and quantity of cryo-EM data collection at NYSBC.

References

- Lander GC, Stagg SM, Voss NR, Cheng A, Fellman D, Pulokas J, Yoshioka C, Irving C, Mulder A, Lau PW, Lyumkis D, Potter CS, Carragher B. *Journal of Structural Biology* Apr;166(1):95-102 (2009).
- Punjani A, Rubinstein JL, Fleet DJ, & Brubaker MA. *Nature Methods* 14, 290-296 (2017).
- Rohou A & Grigorieff N. *Journal of Structural Biology* Nov;192(2):216-221 (2015).
- Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, Stagg S, Potter CS, Carragher B. *Journal of Structural Biology* Jul;151(1):41-60 (2005).
- Tegunov D, Cramer P. *Nature Methods* 16,1146-1152 (2019).
- Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. *Nature Methods* 14, 331-332 (2017).
- Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E, Scheres SHW. *eLife* 2018;7:e42166