

Cryo-EM Information Management System and Sample Evaluation at Stanford-SLAC Cryo-EM Center

Htet Khant¹, Yee T. Li², Murali Shankar², Corey Hecksel¹, Patrick Mitchell¹, Yan Liu¹, Lydia-Marie Joubert¹, Chensong Zhang¹, Lisa Dunn¹, Michael Schmid¹, Britt Hedman¹ and Wah Chiu^{1*}

¹ Division of CryoEM and Bioimaging, Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA.

² Scientific Computing Services, SLAC National Accelerator Laboratory, Menlo Park, CA, USA.

* Corresponding author: wah@slac.stanford.edu

The Stanford-SLAC Cryo-EM Center (S²C²) [1] is a national service center funded under the NIH Transformative High-Resolution Cryoelectron Microscopy Program. Our mission is to provide access to high resolution cryo-EM data collection and training for biomedical research. We have implemented a data management workflow from cryo-specimen preparation to data collection using a Cryo-EM Information Management System (CiMS) eLogBook. The eLogbook is a data management software platform used in multiple facilities at SLAC, providing a file catalog, laboratory notebook, sample management and other functionality. Its composable architecture supports pluggable components to enable features specific to the cryo-EM workflow. A Python based web services application programming interface allows users to record experimental meta data as well as automation that seamlessly transfers large volumes of data. It enables users to manage experiments, provides a web-based interface to automatically move data from microscopes to the data center, sets up permission to the data, manages remote access to microscopes via FastX, provides real-time previews of motion-corrected images and corresponding contrast transfer functions (CTF) via Slack messaging app, provides summary information from MotionCorr2 [2] and CTFIND4 [3] programs on a web browser, and provides notebook functionality to record experiments. In addition, it also manages data subsystems at the data center and removes old data from the microscope's local data servers. Data management workflow tasks described above are managed by Apache Airflow and it is kept under revision control on github [4]. Our users can also process the data by requesting compute cycles reserved for cryo-EM at SLAC Shared Scientific Data Facility.

Such software and hardware infrastructure at our center play an important role in evaluating cryo-EM samples by users carrying out experiments on site, as well as for remote users who are shipping samples to us. Users shipping samples have the option of sending their vitrified specimen grids to us or preparing them onsite with help of our staff members. Irrespective of where the users are, our setup described above allows them to carry out data processing in parallel to data collection, significantly improving the turnaround time in sample evaluation. Offsite users can remotely operate our microscopes via FastX and process their data using computing resources at Shared Scientific Data Facility. During a typical 24-hour cryo-EM screening session, users can screen up to 24 specimens during the day, followed by an overnight data collection session on a sample of their choice, using one of our 300 kV Titan Krios electron microscopes, which includes various options of electron detectors, energy filters and fringe-free imaging. Our eLogBook automatically transfers all movies to the data center and provides real-time motion-corrected and CTF-based data quality analysis through our website and via the Slack messaging app. CryoSPARC Live [5] is integrated to give live updates of 2D class averages. High performance computer clusters allow our users to process the data in real-time (CryoSPARC/ CryoSPARC Live or Relion) [5,6] to provide immediate feedback on particle orientation, distribution, and structural

homogeneity. This on-the-fly feedback enables an efficient use of our resources and enhances the users' productivity. The eLogBook archives all metadata and images, allowing the data to be accessed by the users with proper authentication. Recently, we have also expanded our support in providing users access for cryo-grid screening and sample optimization, which appear to be a major bottleneck for many researchers. A typical sample screening session of multiple cryo-EM grids is followed by collecting a small set of data on selected grid(s). A screening session consists of identifying grids with reasonable ice thickness and searching for areas to evaluate particle concentration and distribution. When a candidate grid has been identified, a small-scale data collection is carried out, generally overnight. Once the data collection has begun, automated on-the-fly data processing can be started on the computer cluster to generate live streaming of 2D classes by using CryoSPARC Live and 3D reconstruction using CryoSPARC. For off-site users, a center staff member will load the grids in the microscope so that they can do screening, data collection and data processing remotely. In each of the above steps, our staff can provide expert advice to facilitate the users' discovery process [7].

References:

- [1] <https://cryoem.slac.stanford.edu/s2c2/>
- [2] S Zheng et al., *Nat Methods* **14** (2017), p. 331. <https://doi.org/10.1038/nmeth.4193>
- [3] A Rohou and N Grigorieff, *J Struct Biol* **192**(2) (2015), p. 216-21. doi: 10.1016/j.jsb.2015.08.008.
- [4] Github, <https://github.com/slaclab/cryoem-airflow> (accessed 02 24, 2022).
- [5] A Punjani et al., *Nat Methods* **14** (2017), p. 290. <https://doi.org/10.1038/nmeth.4169>
- [6] S Scheres, *Journal of Structural Biology* **180**(3) (2012), p. 519. <https://doi.org/10.1016/j.jsb.2012.09.006>
- [7] The authors acknowledge funding from the National Institutes of Health Common Fund Transformative High-Resolution Cryoelectron Microscopy Program (U24 GM129541).