DATASHEET

Practical recommendations for cryo-EM single particle analysis reconstruction

Acquisition and processing of large data sets of EM images

Structure determination by cryo-EM single particle analysis (SPA) requires acquisition and processing of large data sets of electron microscopy images. The image processing represents a significant part of the workload and has a big impact on the scientific relevance of the workflow result.

In a publication by Y. Cheng, et al[1], the authors indicated multiple software packages, such as SPIDER, EMAN2, FREALIGN, RELION and SPARX, that contributed to near-atomic resolution structures published recently. They also indicated that RELION, with streamlined processing and fewer options, might be a good starting solution for new cryo-EM users.

An extensive and up-to-date list of software packages is maintained online:

https://en.wikibooks.org/wiki/Software_Tools_For_Molecular_Microscopy

The 3D reconstruction of macromolecules relies on the averaging of tens of thousands of particle views. For successful and efficient reconstruction, it is important to have appropriate computational resources available.

The computational landscape is rapidly evolving, as users have a wide choice of software packages and hardware architectures. These range from single workstations with CPUs or GPUs to large, institutionally-hosted, multi-node clusters or access to cloud-computing resources. Starting points for practical solutions are described in the IT infrastructure part of this document.

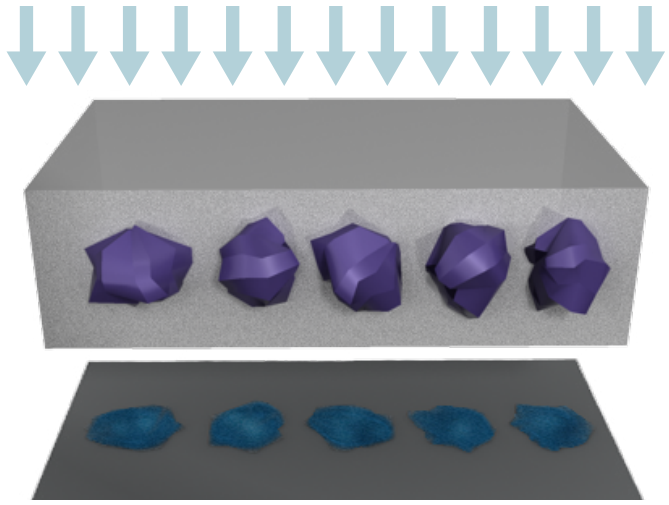

Image processing principle
The main steps in single particle analysis image processing include CTF correction, selection of particles and preparation of image stacks, generation of an initial structure and its refinement, treatment of structural heterogeneity, assessment of resolution, and interpretation of the final 3D density maps.

In single particle analysis, identical macromolecules occur in many thousands of copies with identical structure but different orientations within the ice layer.

The cryo-transmission electron microscope produces two-dimensional projections of these particles. Thus, provided the angular distribution is sufficiently uniform, a large series of micrographs—each showing a field with hundreds to thousands of particles—will yield all the information necessary to reconstruct the molecule.

[1] Y. Cheng, N. Grigorieff, P. A. Penczek, and T.S. Walz, *A Primer to Single-Particle Cryo-Electron Microscopy*. doi:10.1016/j.cell.2015.03.050. Extensive review explains the different steps and considerations involved in structure determination by single particle analysis cryo-EM.

ThermoFisher SCIENTIFIC

thermo scientific

For each of these steps, multiple software packages are available.

Image processing steps

Processing the raw microscope images into a useful three-dimensional model is realized through a sequence of different processing steps. Total time for computing a 3D structure can range from a few hours to days, depending on the size of the dataset, the processing protocols chosen and the specific hardware setup.

1. The **CTF correction** is needed for correction of defocus values for the recorded micrographs.
2. Once a good set of micrographs has been selected, image analysis starts with the process of **particle picking**. Particles can be selected in a manual, semi-automated or fully automated manner.
3. Structure determination continues with the analysis of the 2D image dataset, particularly the **alignment and 2D classification** of the particles into homogenous subsets.
4. Well-defined 2D class averages representing several different views of the particle can be used to generate an **initial 3D reconstruction** at low resolution. An initial 3D structure determination is necessary in cases in which no reasonable templates or guesses for the structure exist.
5. Prior to the refinement of a cryo-EM reconstruction, the particle projections can be further classified with **3D classification** routines that iteratively categorize the particles according to their similarities with one or multiple references. When all images are assigned, new 3D reconstructions are calculated and used as new references.
6. After obtaining an initial map, structure **model refinement** leads to the final three-dimensional density map.
7. High-resolution EM maps can be analyzed by performing *de novo* backbone tracing. At high resolution, docking of X-ray segments can be done with high precision, thus increasing apparent resolvability of the results and making it possible to detect atomic-scale conformational changes with respect to the X-ray results. Similarly, availability of high-resolution structures of multiple functional states of the complex makes cryo-EM single particle analysis a unique tool for the study protein dynamics.

IT infrastructure

Cryo-EM single particle analysis produces large amounts of data, up to, and sometimes even exceeding, one terabyte per day. Data transfer and storage recommendations:

- A 10G/InfiniBand connection between all systems from and to which data will be transferred.
- 100 TB of processing storage capacity is a good about to begin the first year of instrument use.

3D reconstruction

With the introduction of GPU-accelerated computing, reconstruction packages have become significantly faster. It is possible to achieve atomic-resolution reconstruction on a GPU workstation within days; within hours on a GPU cluster. A minimum of a four-GPU server or workstation will allow the processing of data in a very cost-effective and timely manner.

Reconstruction packages can also run on a CPU cluster. A minimum of 64 GB of RAM with a cluster of 100–200 cores is needed.

Several recommendations for hardware choices are available online, directly from academic software development experts:

- https://www2.mrc-lmb.cam.ac.uk/relion/index.php?title=Benchmarks_%26_computer_hardware#Computer_hardware_options (information about recommended hardware (CPU and GPU), as well as references to commercial offerings for ready-to-use machines)
- https://cistem.org/documentation#tab-1-17

Another commercial option is to run on the Amazon EC2™ cloud.

Find out more at **thermofisher.com/EM-Sales**

ThermoFisher SCIENTIFIC