SBGrid: Optimizing your software and compute environment to accelerate structural biology/cryoEM research.

Thomas Nicholson², Jason Key¹, Peter A Meyer¹, Shaun Rawson¹, Alex Ho¹, Kurt L Krause², and Piotr Sliz¹

¹Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA
²Department of Biochemistry, University of Otago, Dunedin, New Zealand

piotr_sliz@hms.harvard.edu, http://sbgrid.org

Keywords: cryoEM, software database, high-performance computing

The Structural Biology Grid (SBGrid) [1,2] is an open consortium of research and development (R&D) laboratories that cooperate to create and maintain a homogeneous structural biology research software infrastructure. Established in 1999, SBGrid initially focused on providing access to macromolecular X-ray crystallography applications. Over the years, SBGrid, which is located at Harvard Medical School, has expanded to encompass macromolecular structure determination by cryogenic electron microscopy (cryoEM), microcrystal electron diffraction (MicroED), cryogenic electron tomography (cryoET) and nuclear magnetic resonance (NMR). SBGrid has a growing number of consortium members, and all 456 partners are involved in the structural biology R&D community at large. They include academic research laboratories, and extend to a growing number of pharmaceutical companies. To support the increasing use of SBGrid in Australasia, SBGrid recently established a satellite site at University of Otago in New Zealand, which coordinates regional software installations.

SBGrid supports structural biology laboratories and biomedical researchers with a tested and refined software infrastructure that includes a large library of scientific applications. The SBGrid team configures, compiles, tests, and continuously upgrades a comprehensive and growing collection of about 500 built-to-run structural biology applications, which is augmented by an additional 600+ computational biology applications. SBGrid members utilize the SBGrid installation manager (Figure 1) to locally install the SBGrid supported applications on Mac and Linux workstations, High Performance Computing (HPC) clusters and in various cloud environments. Once the SBGrid software is installed no further configuration is needed. All applications are packaged by SBGrid with required libraries and supporting software.

SBGrid facilitates rapid access to the newest versions of software and routinely disseminates applications within days of the initial release by developers. Compute-intensive applications are optimized, and instructions for effective execution are provided in the SBGrid web-based software directory (sbgrid.org/software). To advise on the most cost-effective hardware configurations, SBGrid benchmarks CryoEM software with a widely ranging types of datasets. In Figure 2 we present a standard CryoEM structure determination workflow, which consists of 11 steps that need to be completed in sequence. The workflow was tested on dozens of various hardware configurations. Depending on the dataset size, the computational runtimes for a single iteration of structure determination on a single workstation varied from 43 minutes to over 42 hours. In general, the total runtime was proportional to the number of images and storage space of the dataset. Patch Motion Correction (PMC) and Non-Uniform Refinement were the most time-consuming steps and as the number of GPUs increased, the runtime for PMC drastically decreased.