

PReSTO for Integrated Structural Biology

Martin Moche¹, Sebastian Thorarensen², Karl Hörnell², Christian Luckey², Daniel Larsson³, Jie Nan⁴, Oscar Aurelius⁴, Ana Gonzales⁴, Alexandra Ahlner⁵, Michael Hall⁶, Piotr Draczkowski⁷, Tim Schulte⁷, Stefan Fleischmann⁷, and Torben Rasmussen².

¹Protein Science Facility Karolinska Institutet, ²National Supercomputer Centre Linköping University, ³Cell och molekylärbioologi Uppsala University, ⁴MAX IV laboratory Lund University, ⁵Fysik, Kemi och Biologi Linköping University, ⁶SciLifeLab CryoEM Umeå University, ⁷SciLifeLab CryoEM Stockholm University.

The PReSTO project makes structural biology software available in the high-performance-compute (HPC) resources of NSC Tetralith, NSC Berzelius, LUNARC Cosmos and the MAX IV cluster with technical support from National Academic Infrastructure for Supercomputing in Sweden (NAISS). PReSTO started in 2013 as an outreach activity of the National Supercomputer Centre (NSC) in Linköping, funded by MAX IV as a satellite from 2015 – 2018, and by the Swedish Research Council 2018-2022. In 2023 the PReSTO project applied to join SciLifeLab National Bioinformatics Infrastructure in Sweden (NBIS) from 2025-2028 that besides funding will benefit from support and training activities developed by NBIS. PReSTO have small manpower resources summing up to less than 1 FTE now in 2024.

Currently and in the foreseeable future, researchers using Cryo Electron Microscopy, serial synchrotron X-ray crystallography, free electron lasers and AI-based prediction methods need access to advanced GPU computation. NSC Berzelius with its 94 NVIDIA DGX-A100 compute nodes provides the required computational power. Indeed, our implementation of Cryo-EM software via PReSTO on Berzelius and other clusters has already facilitated more than ten Cryo-EM projects. Through our workshops and training initiatives, the most recent given in April 2024, we aim to attract more users and broaden the research community's access to these resources. However, Berzelius was tailored primarily towards the AI and machine-learning community and current NSC Berzelius usage policies, such as a power consumption watchdog, limit the effectiveness of this resource and we therefore created specific allocation reservations (lanes) to circumvent this problem. Additionally, our community faces a significant need for data storage, i.e. a single Cryo-EM project easily amounts to at least 10-20 TB. Therefore, the upcoming Cryo-EM facility in Lund will, together with LUNARC, take storage into account and likely purchase a login-node (frontend) for CryoSPARC thereby designing a resource for Cryo-EM data processing from scratch.

To create an optimal user experience, PReSTO leverages Thinlinc, Gitlab, and Easybuild to modify and patch structural biology software, thereby facilitating the use of SLURM scheduling with applications such as XDS and CryoSPARC, among others. PReSTO requires help with testing of updated software, while keeping the old versions active as default versions. At MAX IV we install an entire suite of updated MX software's in one go (release), wait for MAX IV staff to test the newly installed versions, keeping the old versions as default, and make the new versions default once MAX IV staff has tested them. The same procedure is in place at NSC Tetralith/Berzelius, and LUNARC Cosmos, so having assistance from the large Swedish structural biology infrastructures is important to PReSTO. Several structural biology courses organized by Lund University, the integrated structural biology course across Sweden, and the recent MicroMAX summer schools all used PReSTO for training 2021-2024.