

About PReSTO

PReSTO is a software stack for integrated structural biology (ISB) adapted to high performance computing resources at the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the local MAX IV compute cluster. Our aim is to support integrative structural biologists evaluating their data from macromolecular X-ray crystallography (MX), X-ray free electron lasers (XFEL), nuclear magnetic resonance spectroscopy (NMR), cryo-electron microscopy (Cryo-EM), neutron scattering (NMX), and small-angle X-ray scattering (SAXS).

One Install to rule them all, One Login to find them, One Module to bring them all and on Virtual Desktop run them

Software in PReSTO

MX: XDS, CCP4, PHENIX, PyMOL, autoPROC / BUSTER / SHARP from Global Phasing, Chimera, ChimeraX, hkl2map, O, etc.

SAXS: ATSAS and SasView

SSX/XFEL: CrystFEL, Cheetah, nXDS, NanoPeakCell

Cryo-EM: Scipion, RELION, CistEM, EMAN2, Gctf, Gautomatch, Motion2Cor, Ctffind, Xmipp, CryoSPARC, CryoDRGN

NMR: NMRPipe, MddNMR, qMDD

Please approach PReSTO team for additional software

Integrated Structural Biology

The multi-purpose and multi-technique approach to characterize how macromolecules and their assemblies interact in space and time, is known as integrated structural biology. SciLifeLab has an integrated structural biology platform and PReSTO connects data generating structural biology infrastructures such as MAX IV, Swedish NMR Centre, and SciLifeLab CryoEM, with NAISS compute infrastructures.

Structural biology is a part of the large Swedish effort known as Data Driven Life Science (DDLs) that will benefit science when addressing complex research questions involving machine learning and artificial intelligence. Modern structural biology stretches towards cellular length scales and *in-situ* studies of life capturing dynamic processes via NMR, electron and synchrotron X-ray tomography, correlative microscopy, and molecu-

Job ID	Type	Berzelius mig	Tetralith gpu	Cosmos gpua 100i	Berzelius safe
J27	Extensive Validation	01 hr. 17 min.	02 hr. 53 min.	49 min. 01 sec.	48 min. 40 sec.
J28	Import Movies	22 sec.	56 sec.	24 sec.	23 sec.
J29	Patch Motion Correction	06 min. 05 sec.	06 min. 34 sec.	03 min. 41 sec.	03 min. 15 sec.
J30	Patch CTF Estimation	02 min. 06 sec.	02 min. 48 sec.	02 min. 26 sec.	02 min. 15 sec.
J32	Blob Picker	16 sec.	52 sec.	17 sec.	18 sec.
J34	Extract From Micrographs (CPU)	17 sec.	21 sec.	18 sec.	17 sec.
J35	2D Classification (NEW)	03 min. 36 sec.	02 min. 37 sec.	01 min. 45 sec.	01 min. 45 sec.
J37	Template Picker	20 sec.	53 sec.	25 sec.	23 sec.
J39	Extract From Micrographs (CPU)	33 sec.	39 sec.	33 sec.	33 sec.
J40	2D Classification (NEW)	06 min. 07 sec.	05 min. 04 sec.	03 min. 58 sec.	04 min. 11 sec.
J43	Ab-Initio Reconstruction	12 min. 38 sec.	11 min. 30 sec.	09 min. 06 sec.	11 min. 30 sec.
J44	Homogeneous Refinement	12 min. 24 sec.	08 min. 55 sec.	05 min. 09 sec.	06 min. 52 sec.
J45	Non-uniform Refinement	21 min. 57 sec.	18 min. 15 sec.	07 min. 13 sec.	07 min. 06 sec.
J46	3D Classification	02 min. 16 sec.	02 min. 41 sec.	02 min. 19 sec.	02 min. 06 sec.
J47	3D Variability	04 min. 53 sec.	05 min. 14 sec.	04 min. 14 sec.	04 min. 22 sec.

CryoSPARC extensive validation test job comparison at Tetralith, Cosmos and Berzelius. Cosmos and the Berzelius safe lane is faster than the Berzelius mig lane using more resources. The mig lane does not meet the minimum CryoSPARC system requirements.

PReSTO for Integrated Structural Biology.



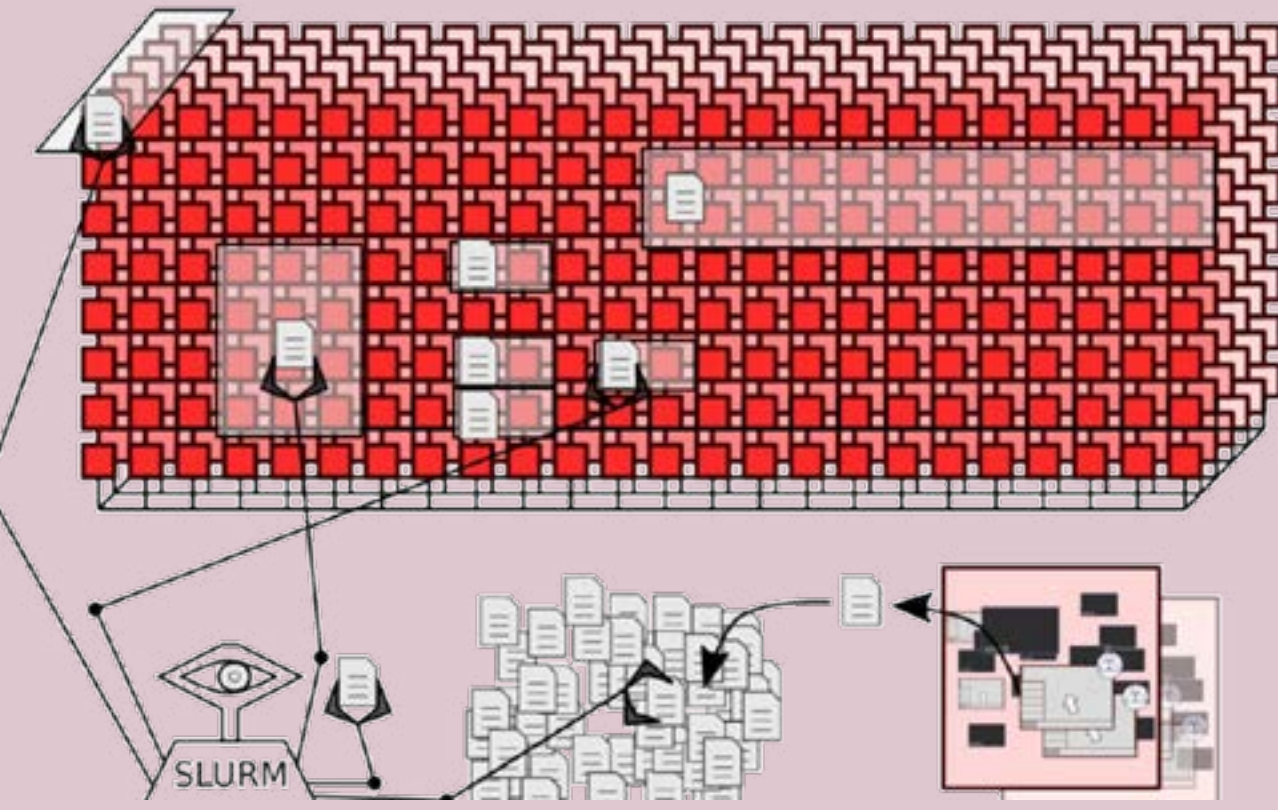
Compare GPU enriched compute nodes at “PReSTO resources”

NSC	Nodes	NVIDIA GPUs	CPUs	Intel/AMD CPUs	RAM	/scratch/local
Tetralith	170	1x Tesla T4	32	2x Xeon Gold 6130	96 GB	2 TB NVMe SSD
Berzelius-fat	34	8x ¹ A100 TC	128	2x Epyc 7742	2 TB	30 TB NVMe SSD
Berzelius-thin	60	8x ¹ A100 TC	128	2x Epyc 7742	1 TB	15 TB NVMe SSD
LUNARC						
COSMOS	6	1x A100	48	2x Epyc 7713	512 GB	2 TB NVMe SSD

GPU enriched compute nodes on PReSTO resources. ¹Berzelius compute node have eight (8) GPUs, however popular CryoSPARC is NOT using all resources leading to job termination by Berzelius GPU Usage Efficiency Policy. Two workarounds: NVIDIA Multi Instance GPU (mig) lane at Berzelius allows GPU partition into smaller instances e.g. 1/7 A100 GPU capability, 10 GB VRAM and 2 cores/4 threads, and Berzelius safe lane; one full A100 GPU, 40 GB VRAM and 16 cores/32 threads. The safe lane is underprovisioned in slurm, however does not terminate.

Queue System: Slurm

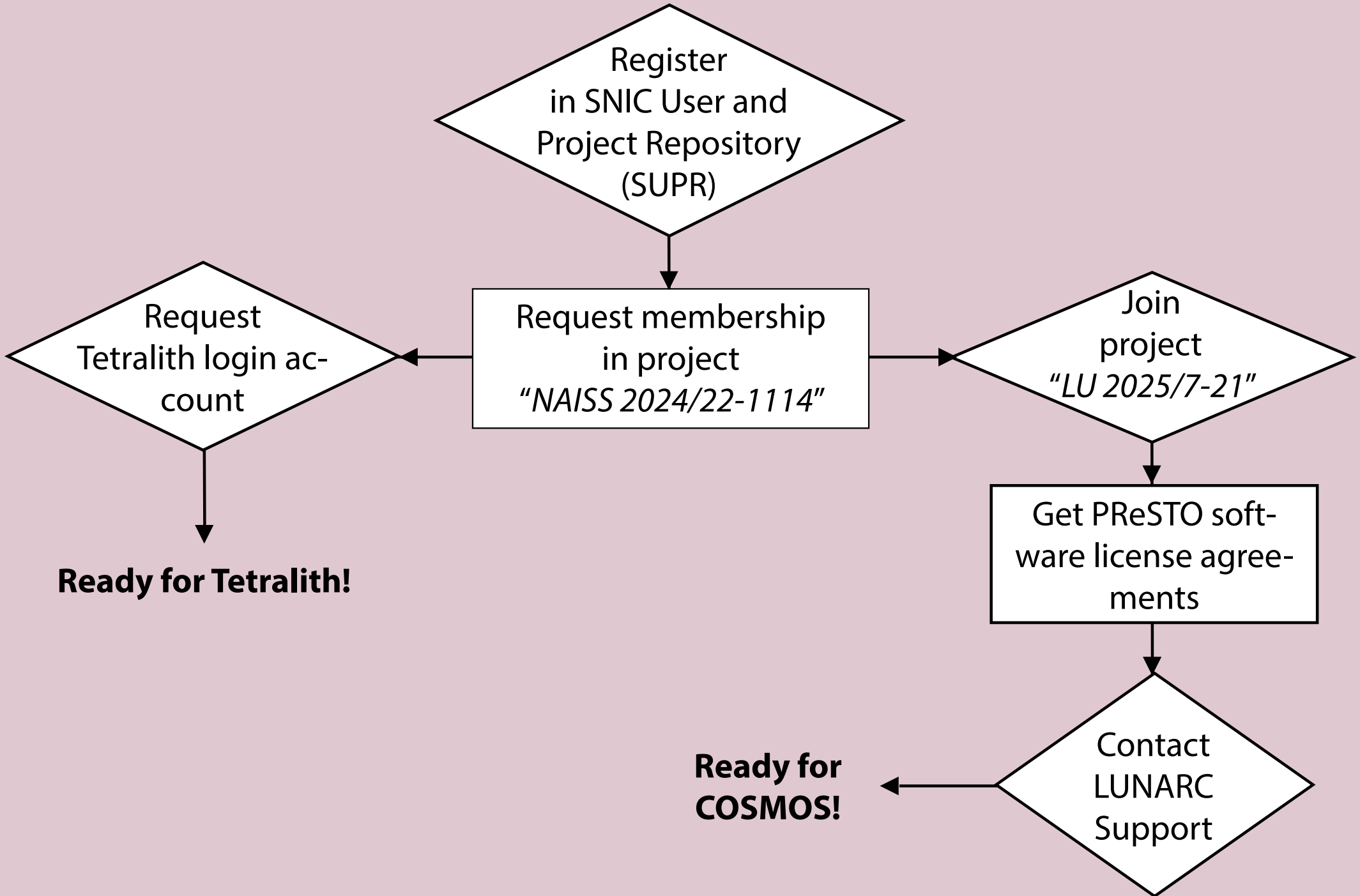
Slurm is a resource allocation queuing system used to share compute resources by many simultaneous users that submit “jobs” to the queuing system and Slurm allocates computing resources for jobs automatically, based on users’ parameters, number of CPUs, memory, etc. The centralized slurmctld manager and compute node slurm (a daemon) are user controlled by srun, scancel, sinfo, sacct, squeue etc.



Some ISB software, such as Phenix and CrystFEL have built-in Slurm support, enabling job submission from their GUIs to the Slurm queue. For other software, that benefits from many cores, a sbatch script is used and the PReSTO website has tutorials available for this task.

Accessing PReSTO at NSC/LUNARC

The homepage (<https://www.nsc.liu.se/support/presto>) is written for structural biologists new to high-performance computing environments, and provides software startup and configuration instructions, training opportunities, access details, known limitations etc. Most structural biologists can access the “PReSTO pilot” project as pictured below. However CryoEM researchers requires more resources than provided by the pilot project and need to apply for compute time and storage space at Tetralith and Berzelius via SUPR (<https://supr.naiss.se/>).



PReSTO at MAX IV

- Apply for beamtime at MAX IV
- Beamtime accepted: grants access to HPC cluster
- Download ThinLinc Desktop and VPN client
- Access MAX IV offline cluster via Thinlinc



The PReSTO Team

Marcos Acebes¹, Ana Gonzalez², Piotr Draczkowski³, Stefan Fleischmann³, Björn Forsberg⁴, Karl Hörnell⁴, Daniel Larsson⁵, Zdenek Matej², Martin Moche⁶, Jie Nan², Torben Rasmussen⁴, Tim Schulte³, Tanvir Shaikh⁷, Sebastian Thorarensen⁴

Lund University ¹LUNARC ²MAX IV, ³SciLifeLab CryoEM, ⁴Linköping University NSC, ⁵Uppsala University, ⁶Karolinska Institutet, ⁷Umeå University

Contact: martin.moch@ki.se

Website: <https://www.nsc.liu.se/support/presto>

