

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).

Using supercomputing resources for structural biology can be hard. PReSTO aims to provide ISB software in a user-friendly high-performance compute environment.

Available Software

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

Serial MX: CrystFEL, Cheetah, nXDS, NanoPeakCell

Cryo-EM: Scipion, Relion, CisTEM, EMAN2, Gctf, Gautomatch, Motion2Co, Ctffindr, Xmipp, CryoSPARC, CryoDRGN

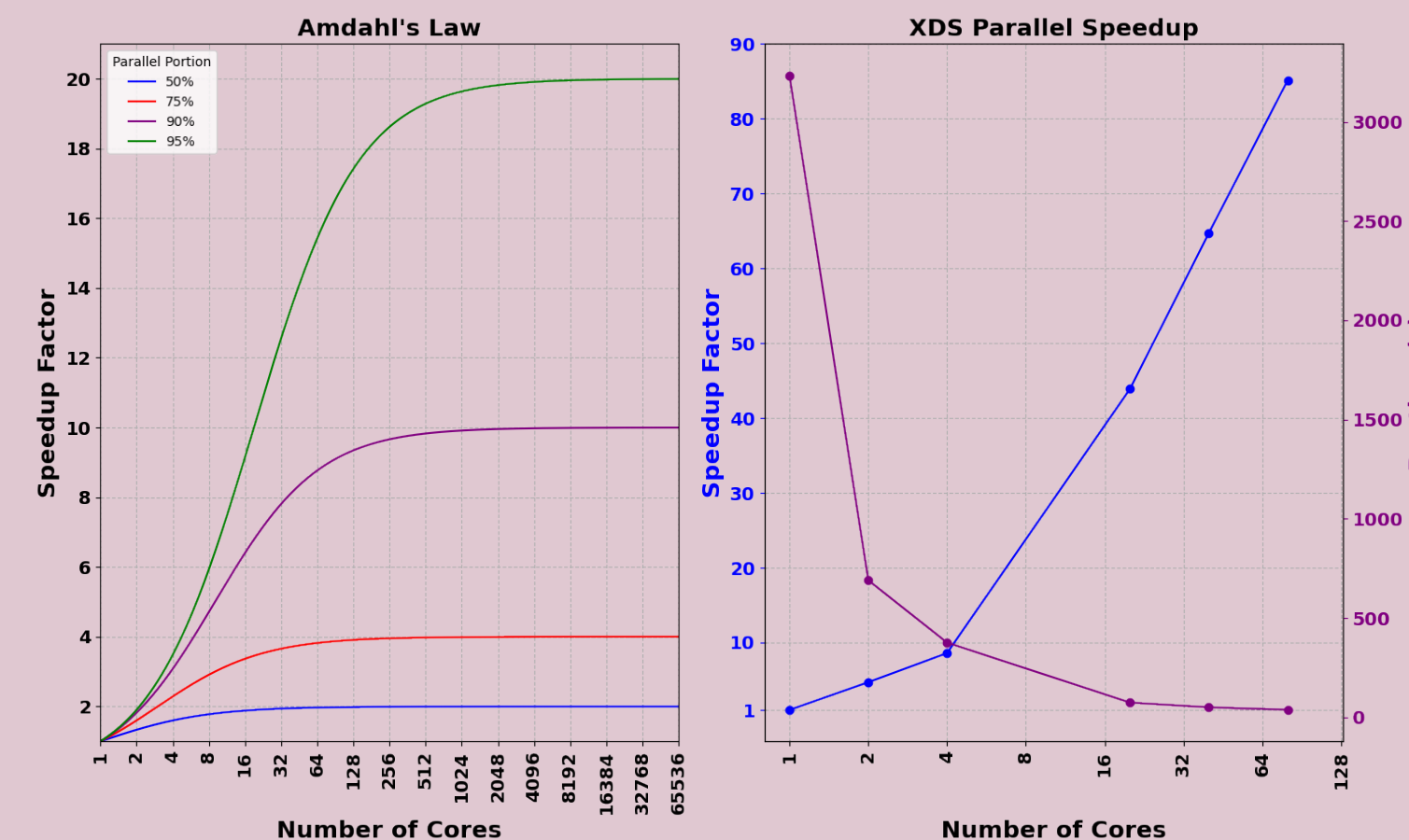
NMR: NMRPipe, MddNMR, qMDD

Our supported software library is constantly growing. Please ask PReSTO team for add-ons.

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 recently launched an Integrated Structural Biology Platform (ISB) 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 (DDL) 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 molecular dynamics simulations.

Highly Parallel Computing Benefits MX Data Processing!



Cores	Speedup	Runtime (s)
1	1	3233.8
2	4.7	690.4
4	8.6	376.5
20	43.9	73.7
40	64.7	50.0
80	85.1	38.0

EIGER 16M test data processing at LUNARC Aurora 1.3 Å EIGER_16M_Nov2015.tar.bz2 data, from DECTRIS (900 frames)

Programs such as XDS take advantage of parallel computing, greatly speeding up the processing speed of datasets. PReSTO allows users to take advantage of supercomputing facilities' parallelization facilities in structural biology programs. Here we show how a test dataset from BioMAX (900 frames, EIGER X 16M detector) can be processed in under a minute with ~40 cores.

PReSTO for Integrated Structural Biology.

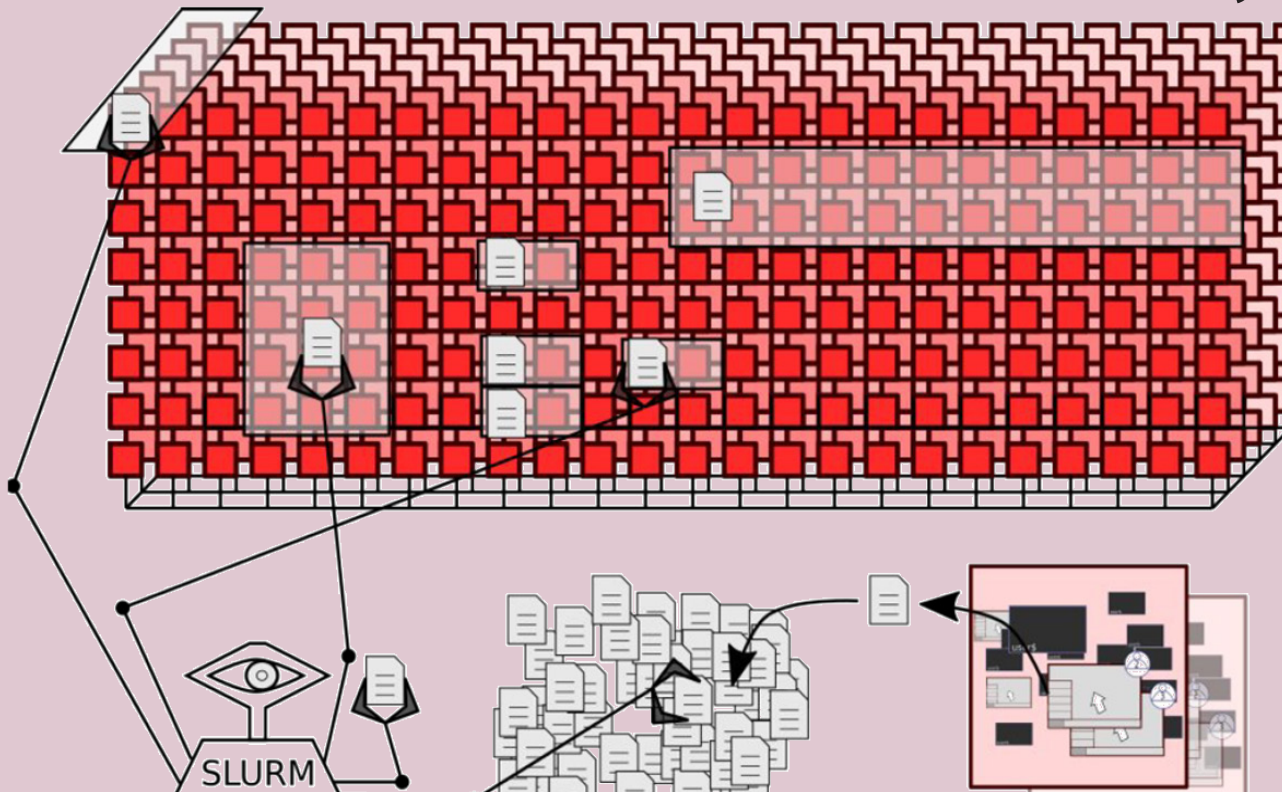
Compare GPU enriched compute nodes at "PReSTO resources"

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
MAX IV						
Online	12	4x V100	48	2x Xeon Gold 6240	400 GB	240 GB SATA SSD

Table 1. GPU enriched compute nodes on PReSTO resources. ¹Berzelius compute node have eight (8) GPUs. NSC Berzelius designed for AI/machine learning lack storage to support the CryoEM community. Upcoming EuroHPC resource Arrhenius in Linköping must resolve this bottleneck.

Queue System: Slurm

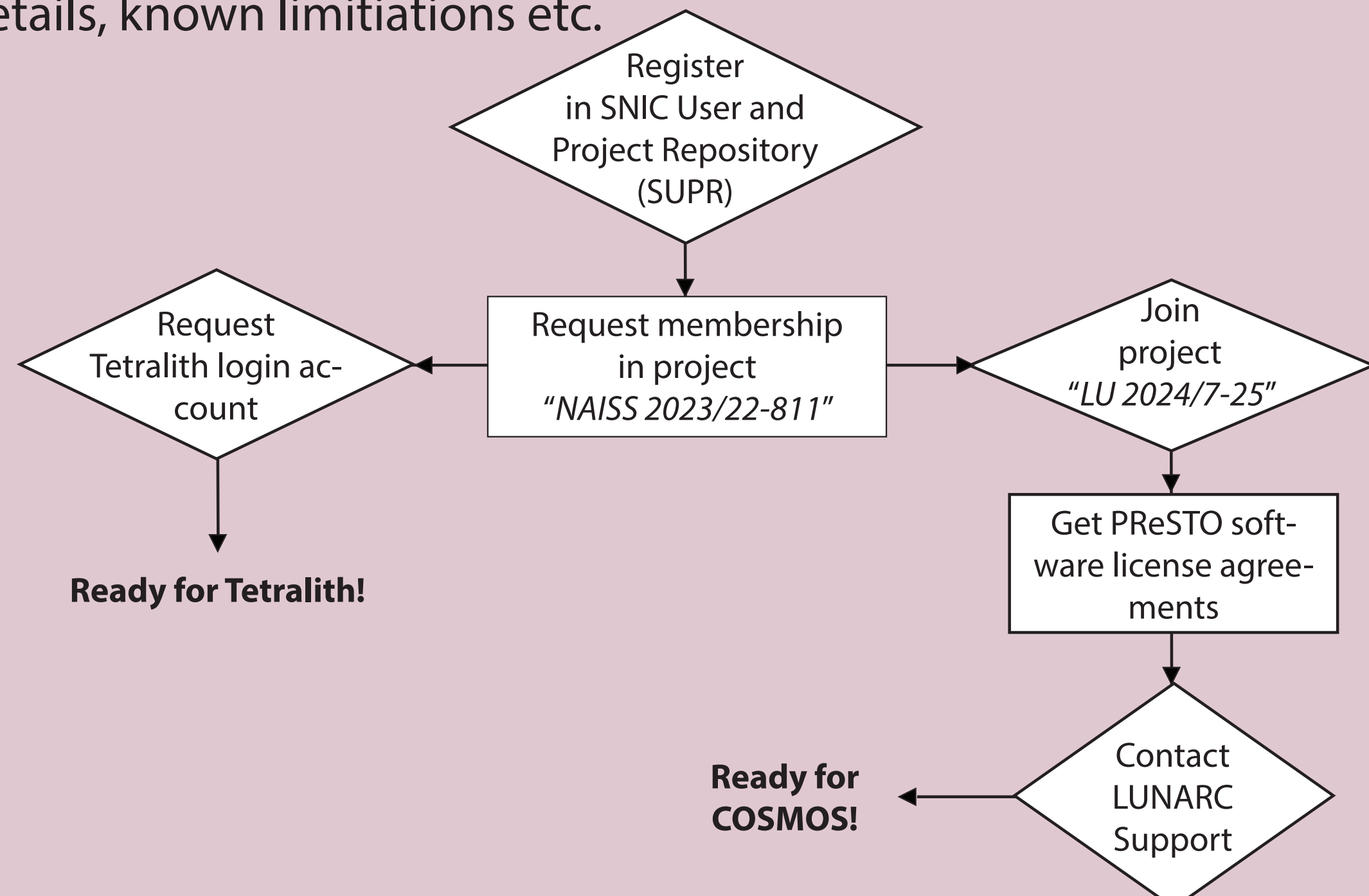
Slurm is a resource allocation queuing system used to share compute resources by many simultaneous users that submit "jobs" to the queuing system and Slurm allocates computing resources for jobs automatically, based on users' parameters, number of CPUs, memory, etc. The centralized slurmctld manager 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.

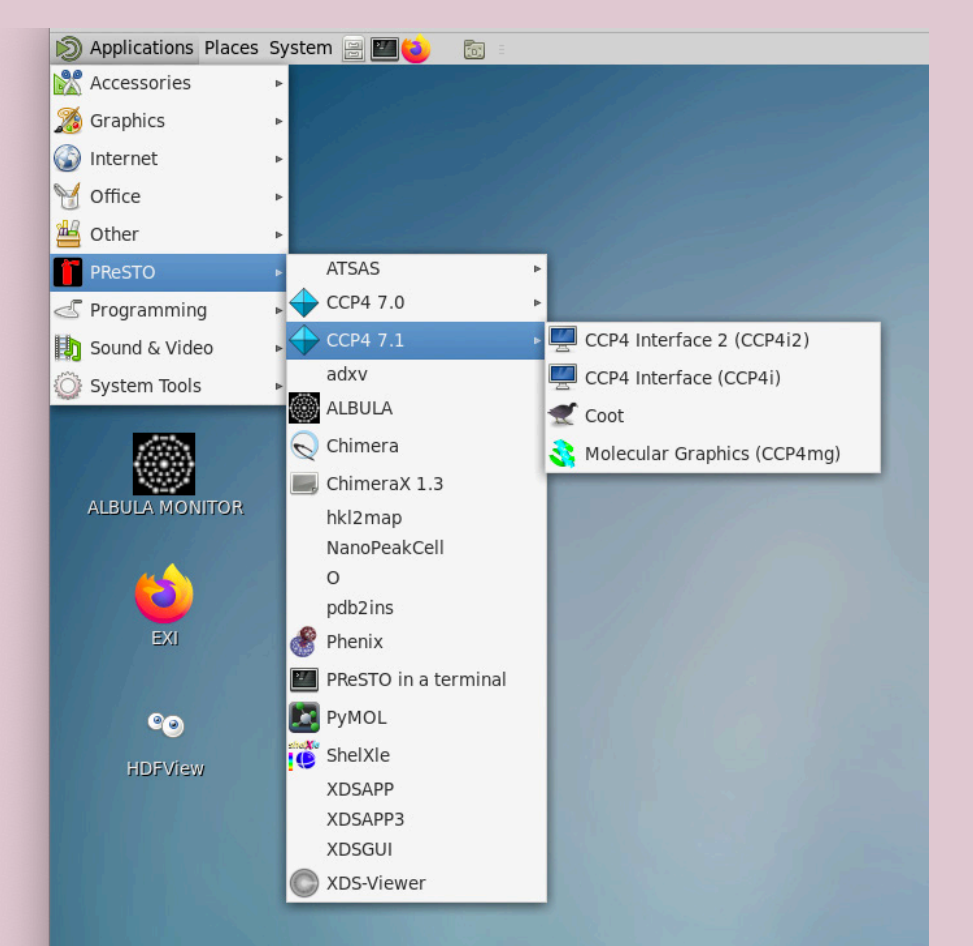
Accessing PReSTO at NSC/LUNARC

Access to the PReSTO installation is via the NAISS, the National Academic Infrastructure for Supercomputing in Sweden, funded by the Swedish Research Council. Thinlinc is a remote desktop server that supports the integrated structural biology workflow by enabling graphic applications to run smoothly from a remote computer. 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.



PReSTO at MAX IV

- Apply for beamtime
- Beamtime accepted: grants access to HPC cluster
- Download ThinLinc Desktop and VPN client
- Access PReSTO software on online cluster (via VPN or on-site) during beamtime
- **MAX IV data available at LUNARC**



The PReSTO Team

Marcos Acebes¹, Alexandra Ahlner², Oskar Aurelius³, Piotr Draczkowski⁶, Ana Gonzalez³, Karl Hörnell⁴, Daniel Larsson⁶, Christian Luckey⁴, Zdenek Matej³, Martin Moche⁵, Jie Nan³, Torben Rasmussen⁴, Tim Schulte⁶, Sebastian Thorarensen⁴

¹Lund University, ²Linköping University, ³MAX IV Laboratory, ⁴National Supercomputing Centre, ⁵Karolinska Institutet, ⁶SciLifeLab CryoEM

Contact: martin.moche@ki.se

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

