PReSTO for Integrated Structural Biology

About *PReSTO*

PReSTO is a software stack for integrated structural biology 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 makes using the software you need as simple as if it were on your own home computer.

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. Swedish SciLifeLab recently launched a new Integrated Structural Biology Platform (ISB) and *PReSTO* will connect data generating structural biology infrastructures such as MAX IV, SciLifeLab Cryo-EM and Swedish NMR Centre with the National Academic Infrastructure for Supercomputing in Sweden (NAISS). Structural biology is a part of the large Swedish effort known as Data Driven Life Science (DDLS) that will benefit protein scientists when addressing complex structural biology questions involving machine learning and artificial intelligence when stretching towards cellular length scales and dynamic events of life via synchrotron X-ray tomography, correlative microscopy, and molecular dynamics simulations.

Available Software

MX: XDS, CCP4, PHENIX, PyMOL, autoPROC / BUSTER / SHARP from Global Phasing, ChimeraX, hkl2map, O, ***Alphafold (new!)***

Serial MX: CrystFEL, Cheetah, nXDS, NanoPeakCell

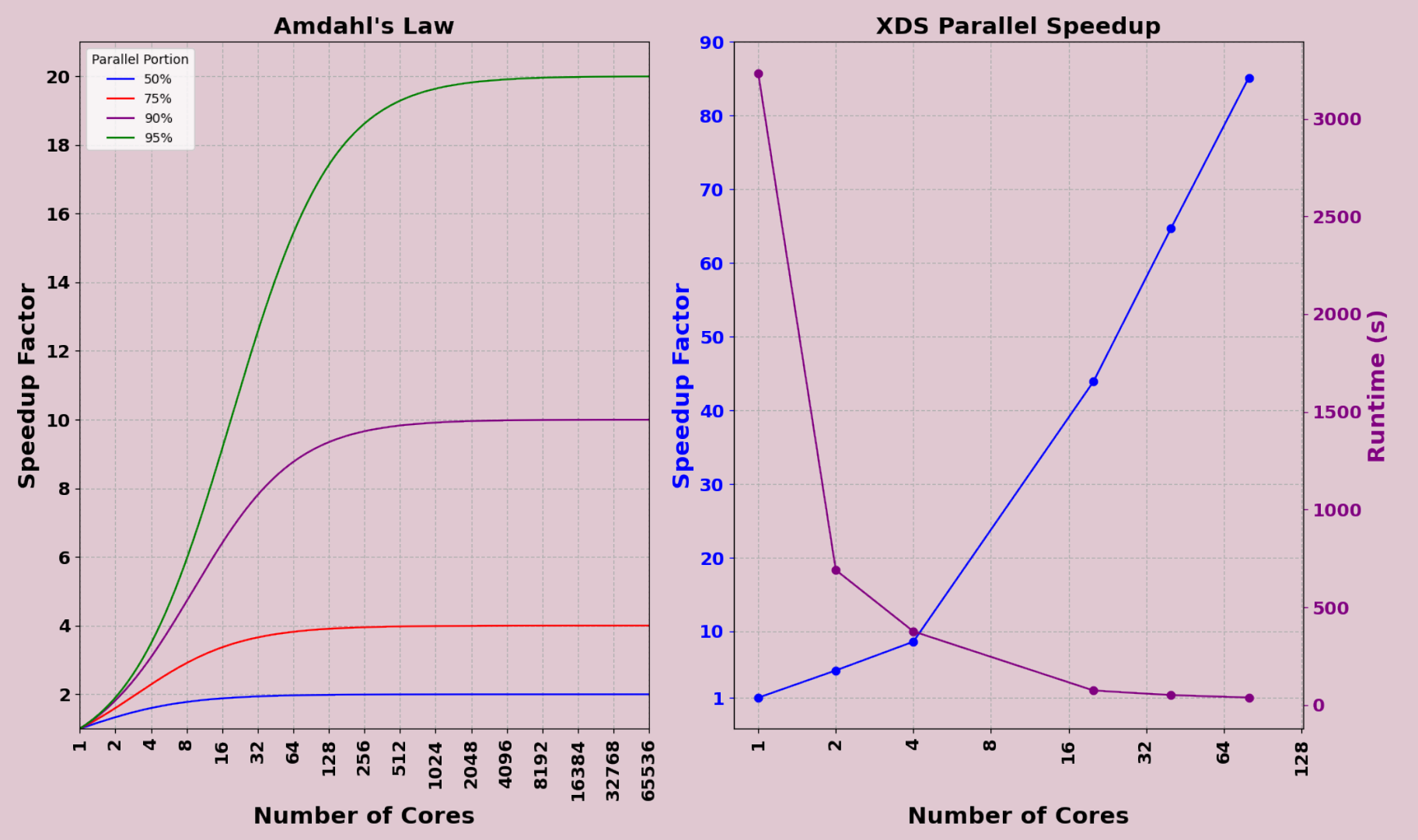
Cryo-EM: Scipion, Relion, CisTEM, EMAN2, Gctf, Gautomatch, Motion2Co, Ctffindr, Xmipp, CryoSPARC

NMR: NMRPipe, MddNMR, qMDD, CYANA, CCPN **SAXS:** ATSAS

Our supported software library is constantly growing.

If you have any other software needs that the *PReSTO* team can assist you with, just ask!

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

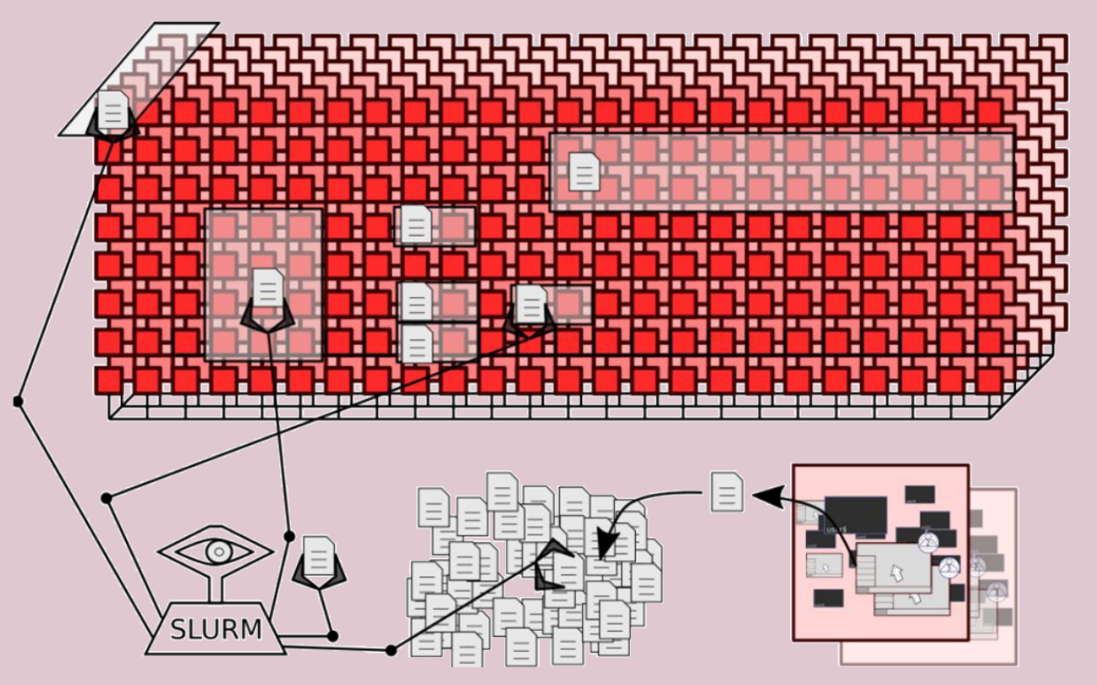
EIGER 16M test data processing at LUNARC Aurora 1.3 Å EIGER_16M_Nov2015.tar.bz2 data, from DECTRIS (900 frames)

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 typical dataset from BioMAX (900 frames, EIGER X 16M detector) can be processed in under a minute with ~40 cores.

Computing Centres Hardware Information

NSC	Nodes	CPU Type	CPUs	RAM	GPU Nodes	/scratch/local
Tetralith	170	2x Intel Xeon Gold 6130	32	96 GiB	1x NVIDIA T4	2 TB NVMe SSD
Berzelius	60	2 AMD Epyc 7742	128	1 TB	8x NVIDIA A100 Tensor Core	15 TB NVMe SSD
LUNARC						
COSMOS	182	2x AMD Epyc 7413	48	256 GiB	8x NVIDIA A100 Tensor Core 12x NVIDIA A40	1 TB NVMe SSD
MAX IV						
Online Cluster	96	Intel Xeon E5-2650	48	200 GiB	12 NVIDIA V100	1 TB NVMe SSD

Queue System: Slurm

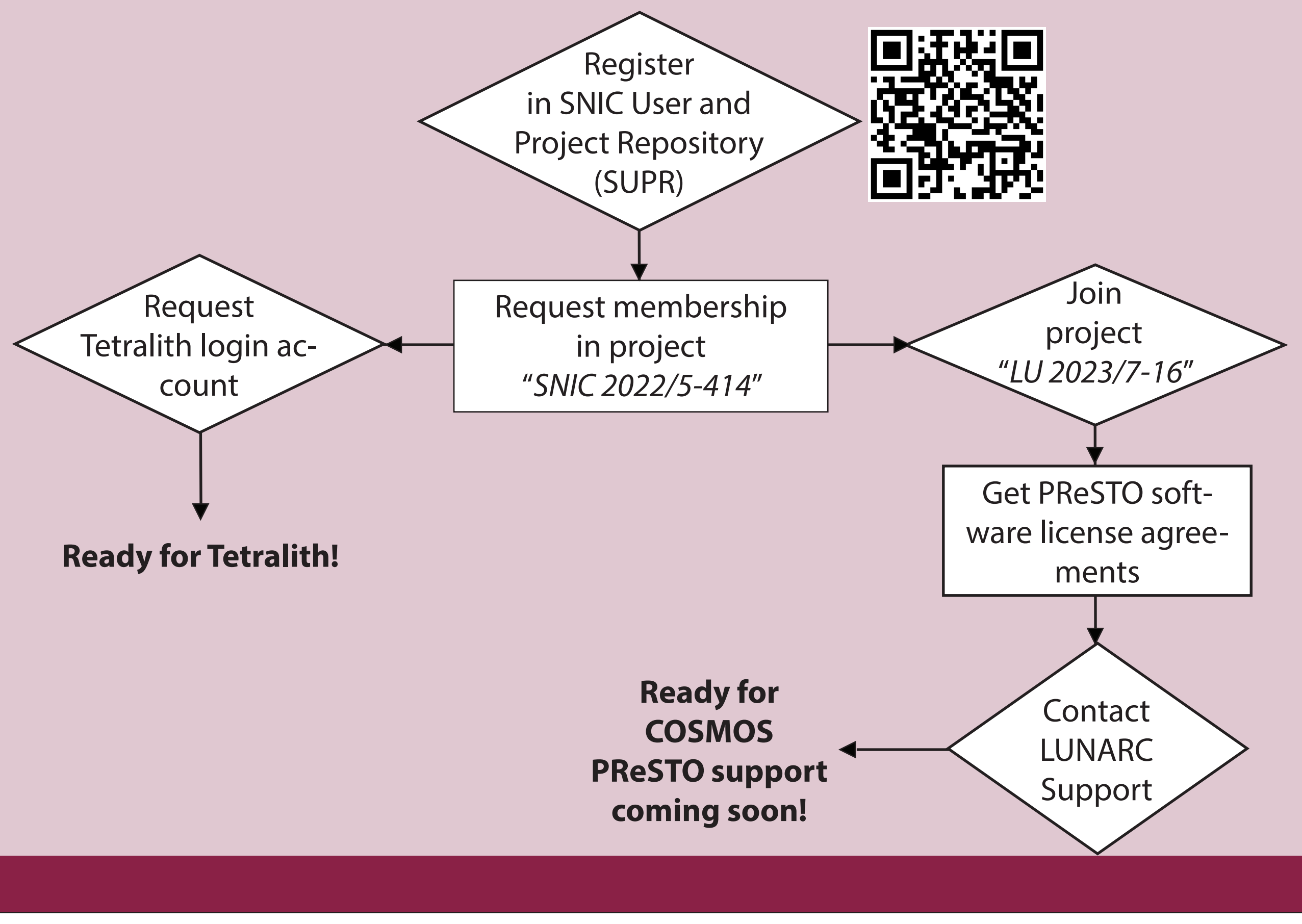


Slurm is a resource allocation queuing system used at all supported computing centres. Users submit "jobs" to the queue system and Slurm allocates computing resources for jobs automatically, based on users' parameters (number of CPUs, memory, etc.).

Many programs supported by *PReSTO* have built-in Slurm support, which can automatically submit jobs to the queue. Jobs can also be submitted manually, or interactively via the Terminal. The PReSTO website has tutorials available for supported programs.

Accessing *PReSTO* at NSC/LUNARC

Access to the *PReSTO* installation is via the National Academic Infrastructure for Supercomputing in Sweden and starts at supr.naiss.se. 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 step-by-step instructions for all supported programs.



Register in SNIC User and Project Repository (SUPR)

Request membership in project *"SNIC 2022/5-414"*

Request Tetralith login account

Ready for Tetralith!

Join project *"LU 2023/7-16"*

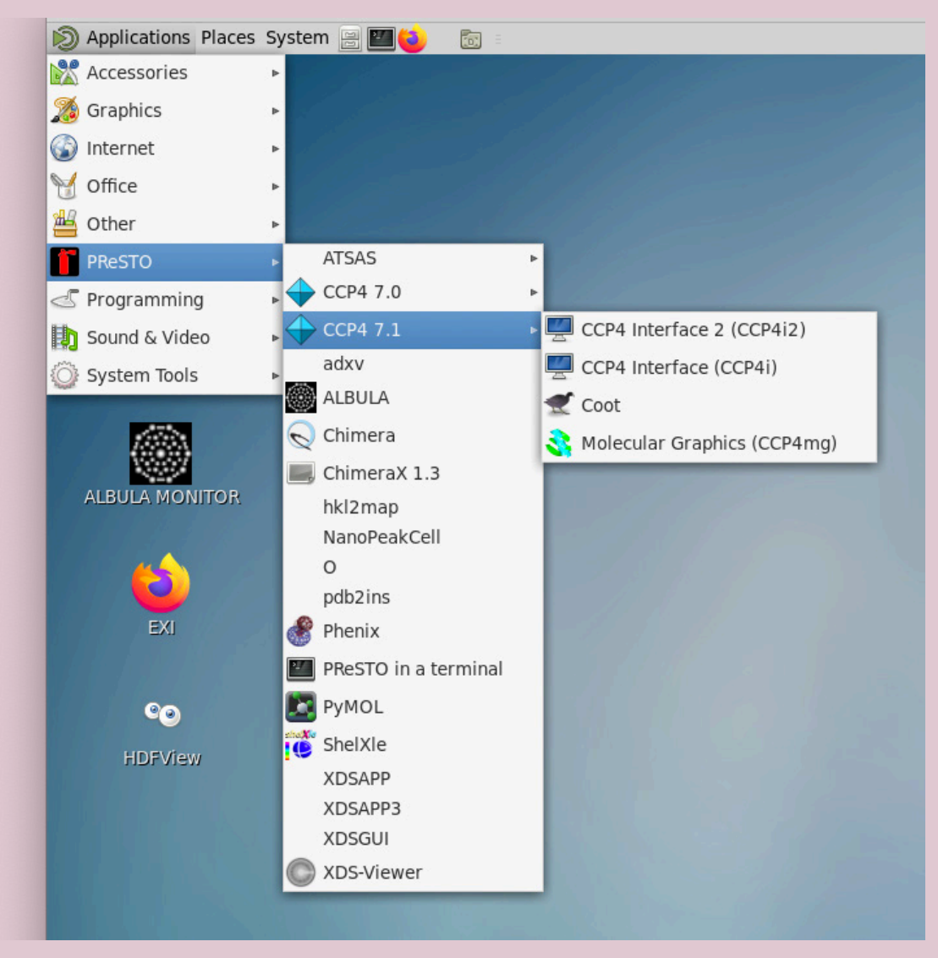
Get PReSTO software license agreements

Contact LUNARC Support

Ready for COSMOS PReSTO support coming soon!

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 is available at LUNARC as well**



The PReSTO Team

Marcos Acebes,[1] Alexandra Ahlner,[2] Oskar Aurelius,[3] Aaron Finke,[3] Ana Gonzalez,[3] Karl Hörnell,[4] Christian Luckey,[4] Zdenek Matej,[3] Martin Moche,[5] Jie Nan,[3] Torben Rasmussen,[4] Anders Sjöström,[1] Maria Sunnerhagen,[2] Sebastian Thorarensen[4]

[1]Lund University, [2]Linköping University, [3]MAX IV Laboratory, [4]National Supercomputing Centre, [5]Karolinska Institutet

Contact: aaron.finke@maxiv.lu.se, martin.moche@ki.se, torbenr@nsc.liu.se

Website: https://www.nsc.liu.se/support/presto →

NSC MAX IV LINKÖPING UNIVERSITY Karolinska Institutet