Science and Technology Facilities Council

Scientific Computing

What's new in CCP-EM?

Agnel Joseph and Joel Greer

23rd April 2026

CCP-EM Spring Symposium, Nottingham

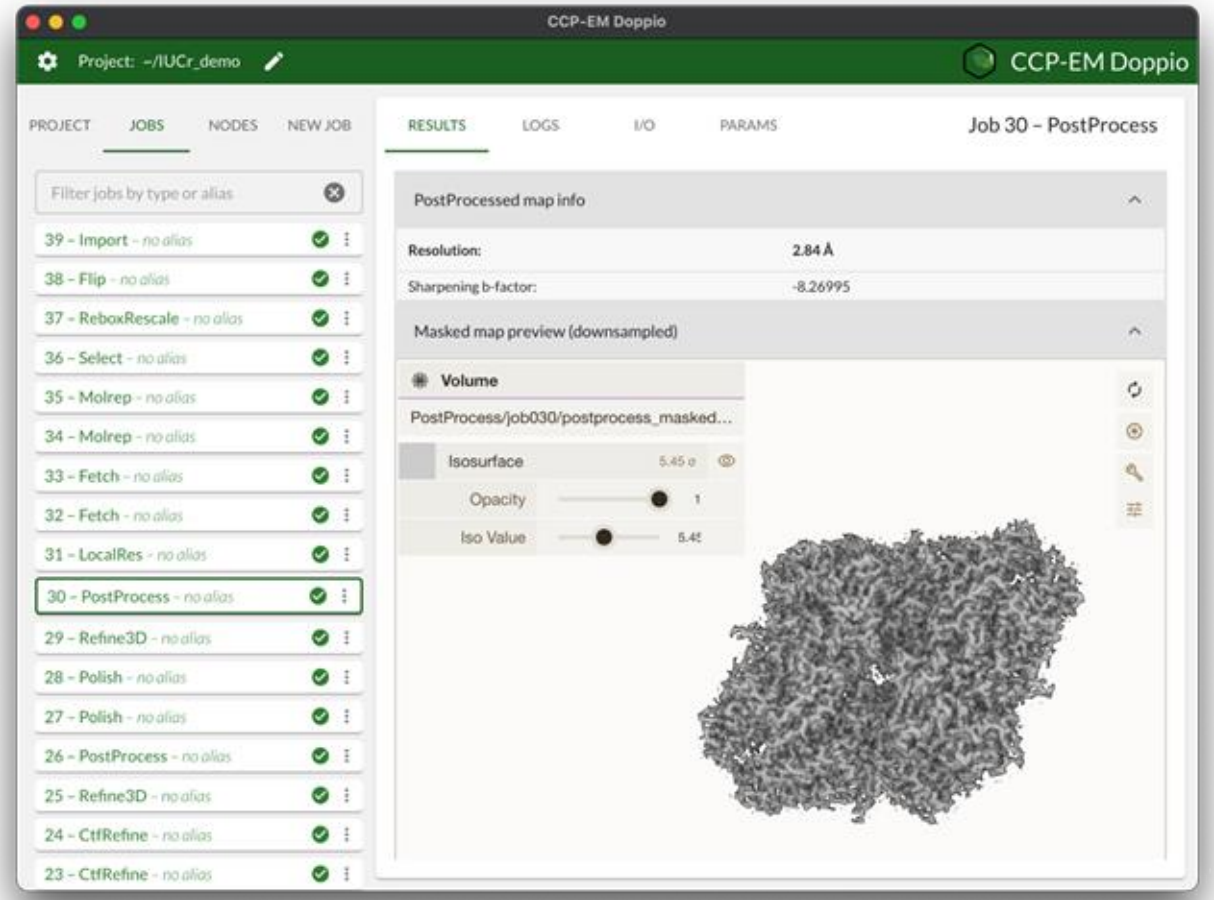
CCP-EM

What's new in CCP-EM?

CCP-EM

Workshops and events

Software updates



Science and Technology Facilities Council

Scientific Computing

CCP-EM Workshops 2025-2026

- UK EM facilities workshop, Warwick, June 2025
- Structure determination (MX and Cryo-EM). Thailand, July 2025
- Integrative structural biology, Italy, July 2025
- CCP4-BCA summer school. York, August 2025
- Shanghai School. September 2025.
- EMBO Cryo-EM image processing, Birkbeck, Sep 2025
- Icknield workshop, RAL, Oct 2025
- CCP4/CCP-EM School Santiago, Chile, Nov 2025
- Doppio workshop, St Andrews, March 2026
- Model building workshop, India, March 2026

- Various Doppio road shows around UK

- Since 2013 we have led or co-led ~48 workshops, teaching over 1,500 researchers

Science and Technology Facilities Council
Scientific Computing

CCP-EM upcoming Workshops 2026

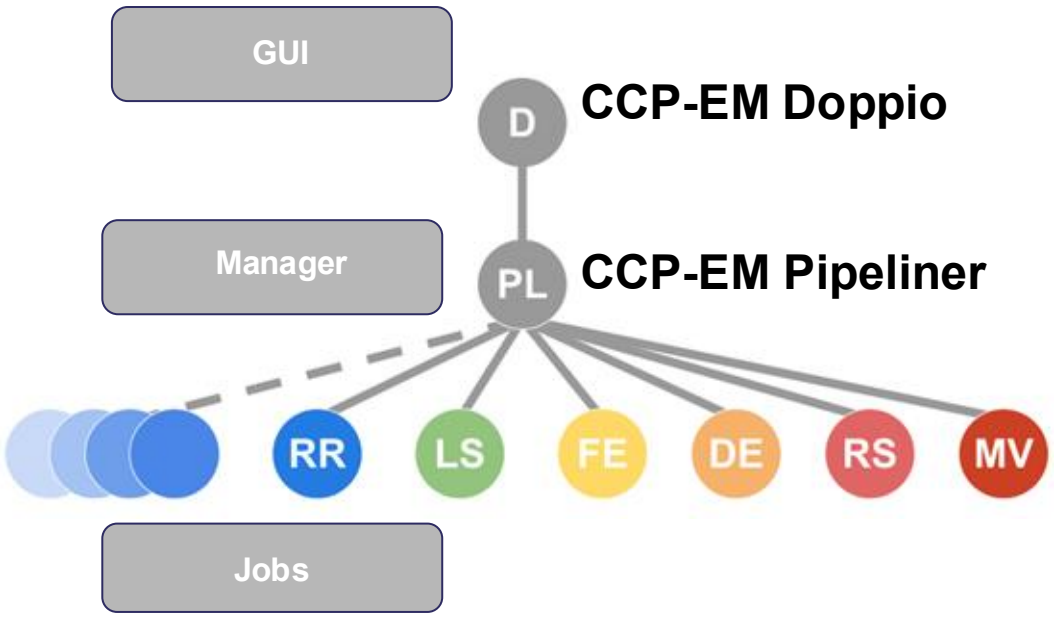
- Central European Workshop 2026, Czechia, April 2026
- International School Crystallography, May 2026
- Dutch cryoEM school, Leiden/Delft, June 2026
- CCP4-APS Chicago, July 2026
- IUCR Calgary, August 2026
- MSCDA cryoEM course, Exeter, Sep 2026
- CAHRA Heterogeneity challenge, Sept 2026
- China National Facility for Protein Science, Sep 2026
- Icknield workshop, Nov 2026
- King Mongkut's, Thailand, Nov 2026

Science and Technology Facilities Council
Scientific Computing

CCP-EM mailing list

ccpem@stfc.ac.uk

CCP-EM software overview



Doppio

- GUI layer
- JavaScript – web browser or standalone desktop app
- STFC licence, free for academic use

Pipeliner

- Project and job management layer
- Written in Python
- Project graph structure, fully compatible with RELION 4 & 5
- Open Source MPL 2.0 licence

Pipeliner jobs

- Python modules that call other software
- Plug-in architecture

Science and Technology Facilities Council
Scientific Computing

CCP-EM *Doppio* 1.6b

Last year: Doppio 1.3 release

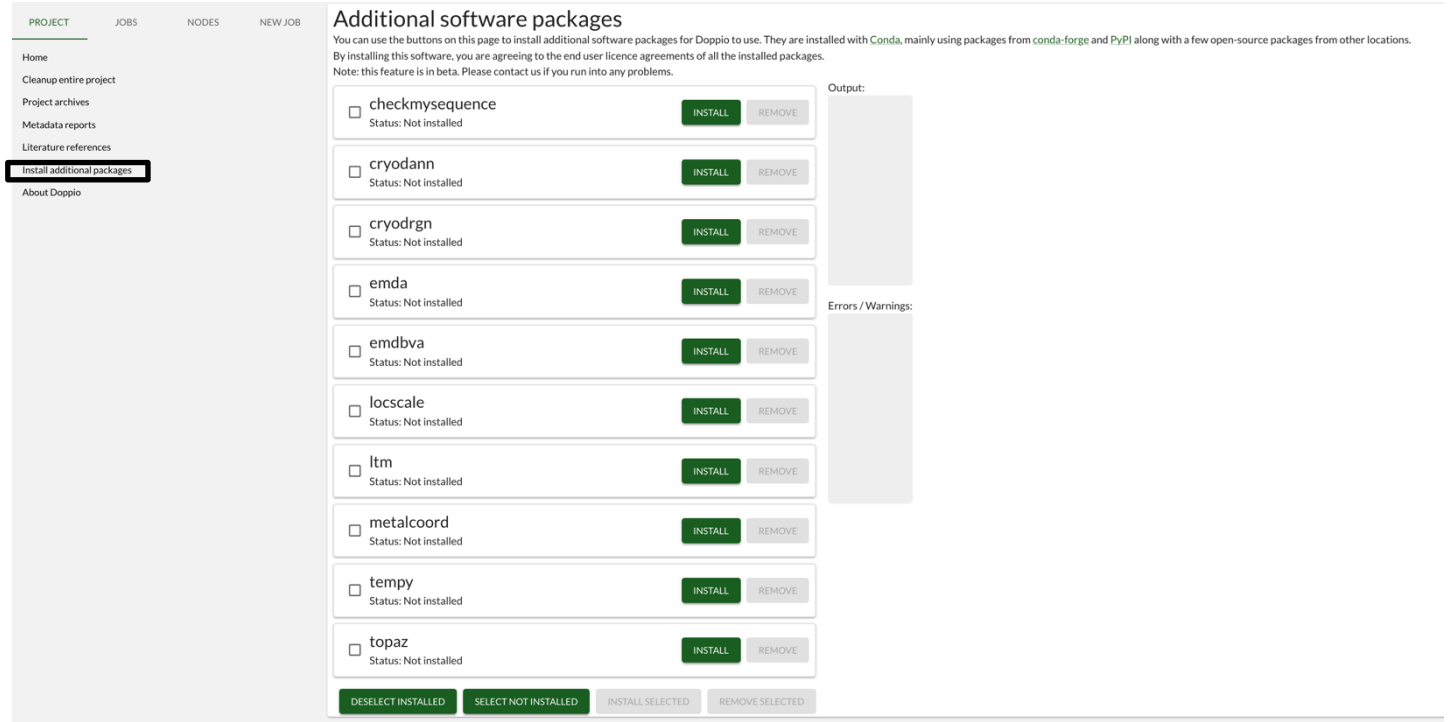
Now: Doppio 1.6 (beta version)

Download:
https://www.ccpem.ac.uk/download.php

User guide:
https://www.ccpem.ac.uk/docs/doppio/user_guide.html

Tutorials:
https://ccpem-tutorials.readthedocs.io
https://www.ccpem.ac.uk/tutorials.php

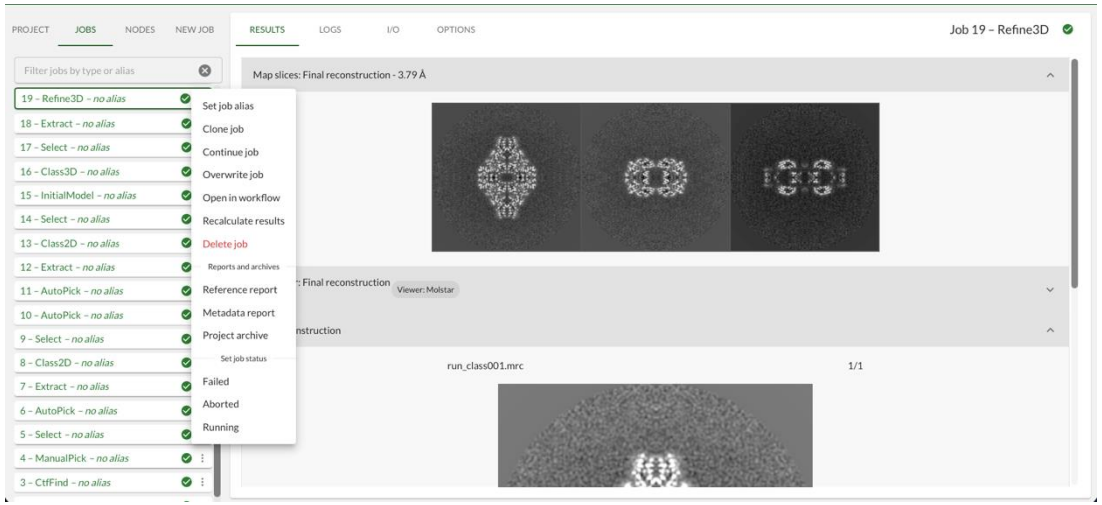
Additional package installation

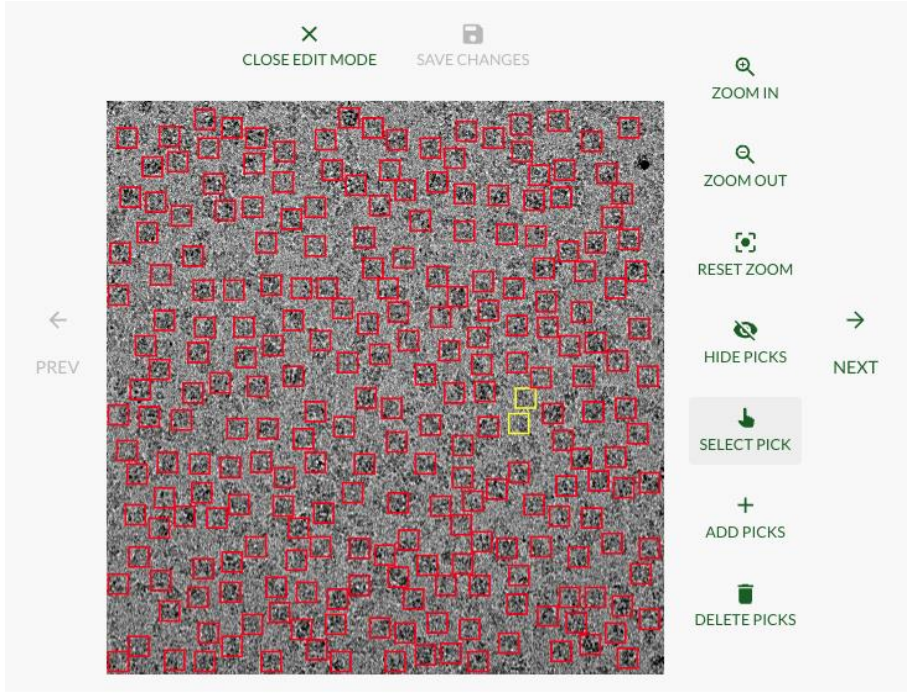


Science and Technology Facilities Council
Scientific Computing

CCP-EM *Doppio* 1.6b

- Job continuation and overwrite
- Edit and save particle picks
- Different Doppio instances can work on different projects
- Support for ChimeraX and Coot scripts for custom results display
- Improved B-factor distribution and FSC plots in Validation and Servalcat jobs
- Local Template Matching tool
- DPIscore (Protein interface assessment)
- Angular distribution plots and 3D .bild file display
- Metadata deposition tools
- Various dependency upgrades, improvements and bug fixes





Template Matching with LTM package

LTM (Local Template matching) is a Python package for 3D template matching in cryo-electron tomograms. It is GPU-accelerated using JAX and currently supports NVIDIA GPUs with CUDA drivers.

Rangana Warshamanage

What LTM Can Do

1. **Create templates** for template matching
2. **Create masks** (spherical or cylindrical) for template matching
3. **Template matching** (global exhaustive search across all orientations, or local search around know positions)
4. **Extract peaks** from the correlation map using Z-score thresholding
5. **Write outputs** as RELION 5 style .star files for downstream analysis
6. **Compute half-tomogram local CC** for quality-guided peak filtering
7. **Visualize** detected particles in Napari
8. Other helper tools

Documentation: *https://ltm.readthedocs.io/*

Science and Technology Facilities Council
Scientific Computing

University of Pittsburgh

Zachary Z Freyberg

LTM is available on Doppio

Template Matching

- **LC** Create Mask for TM
  ltm.create.mask – Create a mask for template matching using LTM
- **LC** Create Template for TM
  ltm.create.template – Create a template for template matching using LTM
- **LE** Extract Peaks from LTM Results
  ltm.extract.peaks – Extract peaks from LTM template matching results
- **LH** LTM Half-tomogram Local CC Calculation
  ltm.halftomo.localcc – Calculate local cross-correlation map for half-tomogram in LTM
- **LT** Template Matching with LTM
  ltm.template.match – Template matching using LTM

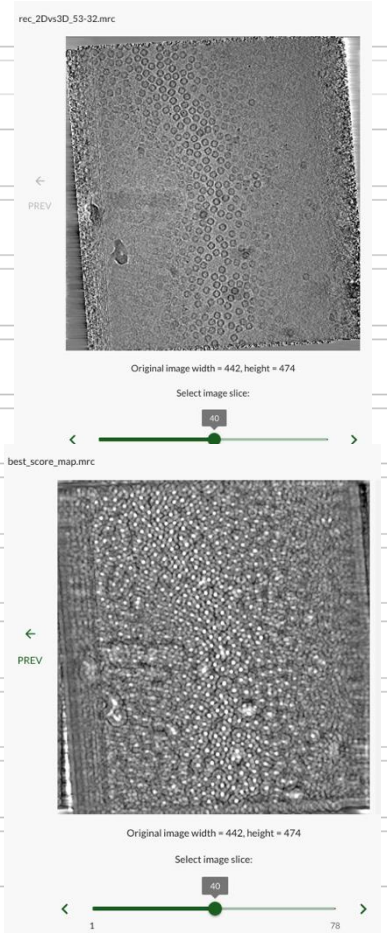
Template Matching with LTM

RUN | JOB INFO | RESET OPTIONS

Job alias:

Main

Tomogram file: *	required*
Template file: *	required*
Template mask file: *	required*
Minimum tilt angle (degrees, e.g. -40): *	-40
Maximum tilt angle (degrees, e.g. 40): *	40
Particle diameter (A): *	
Mask type: *	spherical
Angular step (degrees):	
Psi step (degrees):	
Chunk size (angles):	4
Which GPU ID to use (e.g. 0 or 1 etc):	0



Science and Technology Facilities Council
Scientific Computing

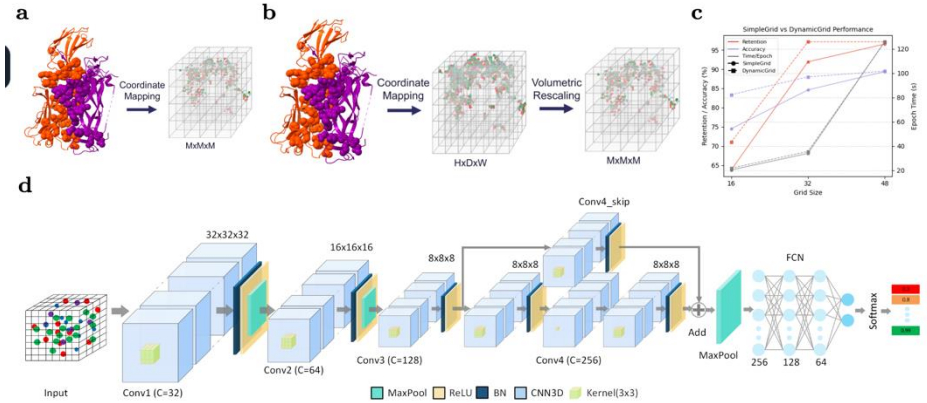
University of Pittsburgh

Zachary Z Freyberg

EMPIAR-10491 – search for apoferritin with template EMD-15854

Deep learning-based Protein Interface quality score (DPI-score)

Science and Technology Facilities Council
Ada Lovelace Centre



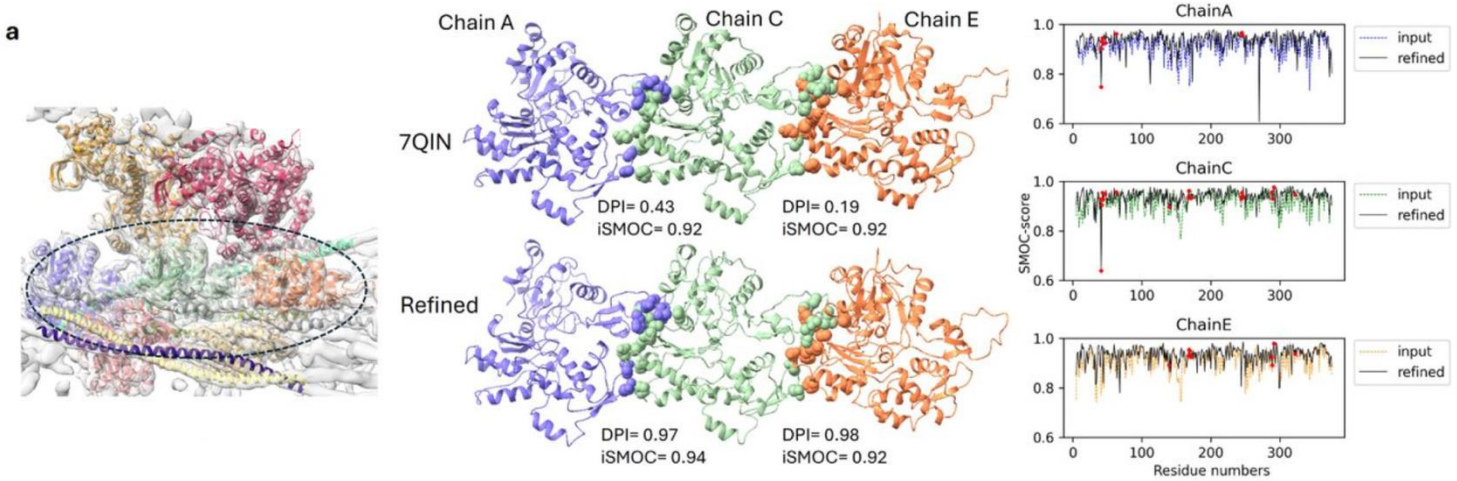
Validate quality of protein-protein interfaces


Sony Malhotra


Niraj Bhujel

6.6 Å resolution structure with a homo pentamer of Actin

CCP-EM software suite v2 (Doppio) was used to perform flexible fitting using TEMPy-REFF job, which was followed by Servalcat refinement.



	Submitted structure	Re-refined structure
Actomyosin complex (6.6Å)		
CCC-mask	0.54 (percentile: 49)	0.6 (percentile 57.9)
Average model map FSC	0.76 (percentile: 65)	0.89 (percentile 98.5)
iSMOC (AC, CE)	0.92, 0.92	0.94,0.92
DPI (AC, CE)	0.43, 0.19	0.97, 0.98

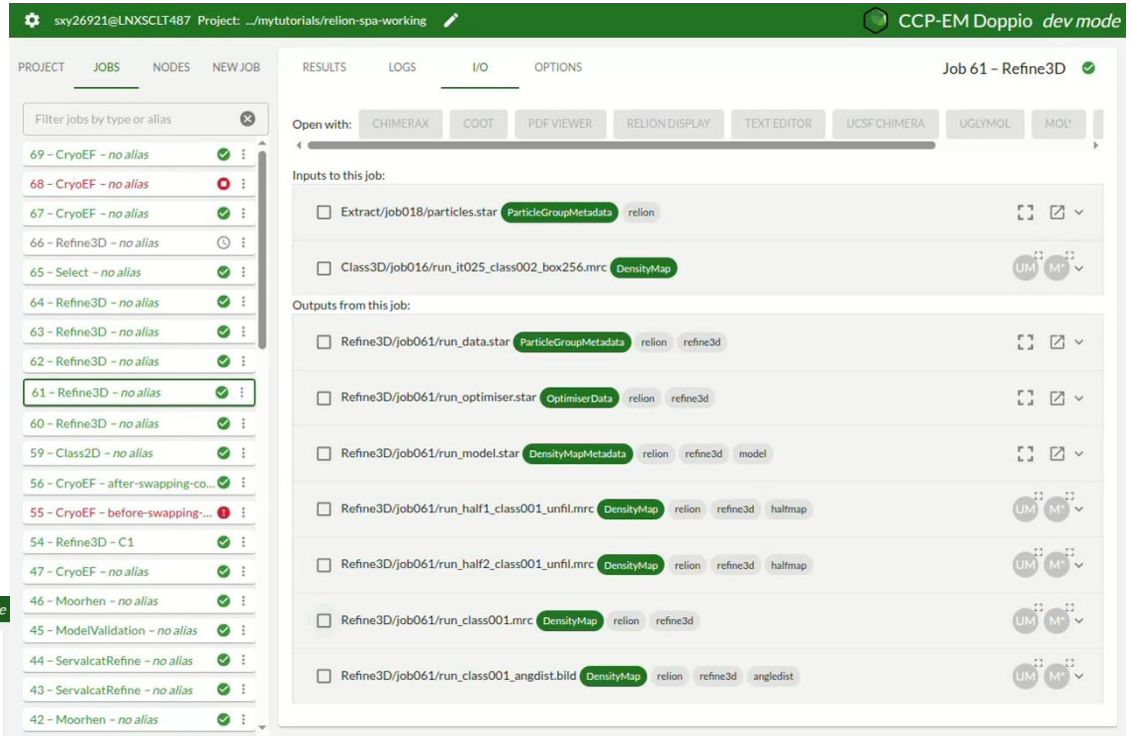
Science and Technology Facilities Council
Scientific Computing

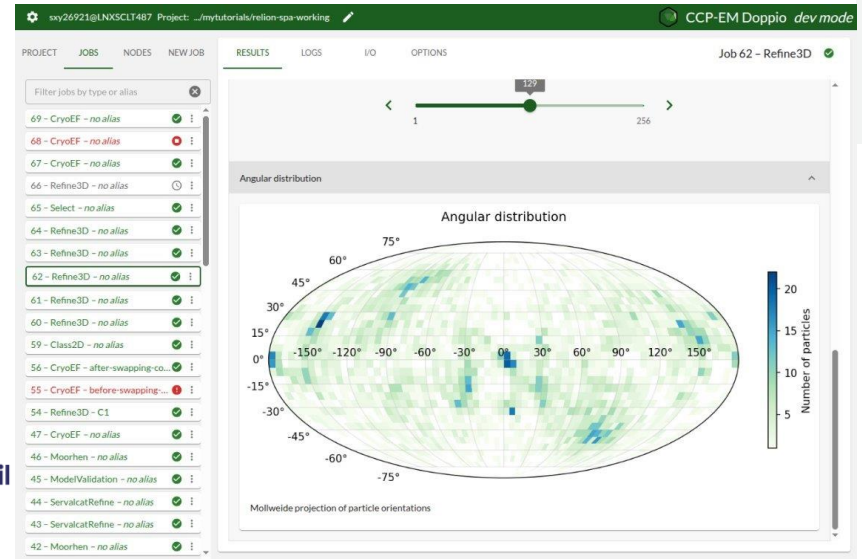
Bhujel et al., https://www.biorxiv.org/content/10.64898/2026.01.12.699015v1

Angular distribution and Mol* upgrades

Science and Technology Facilities Council
Ada Lovelace Centre

- Refine3D job now has preferred orientation visualisations
- Porcupine plots (bild file support) in Mol* 3D Viewer
- Additional Mollweide projection
- Mol* Menus have individual show / hide buttons for ease of navigation

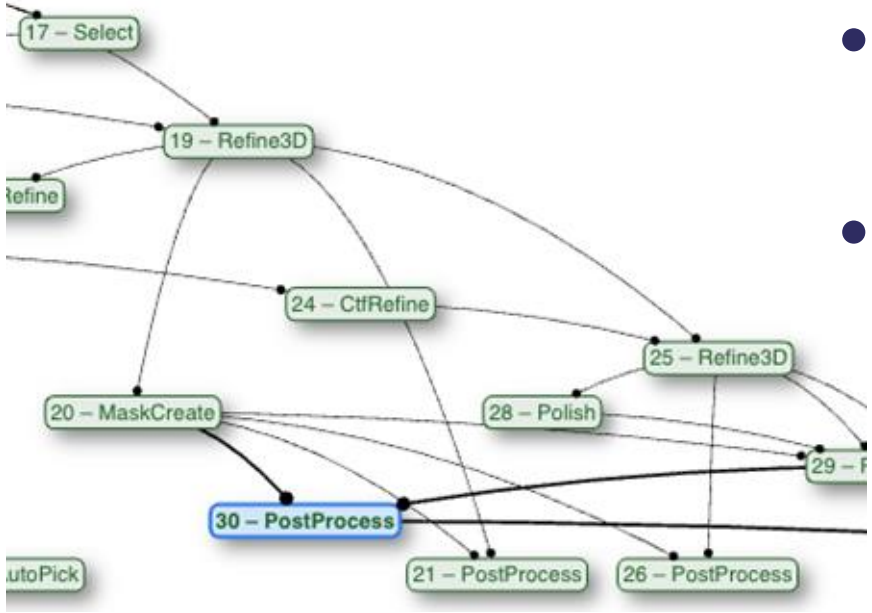




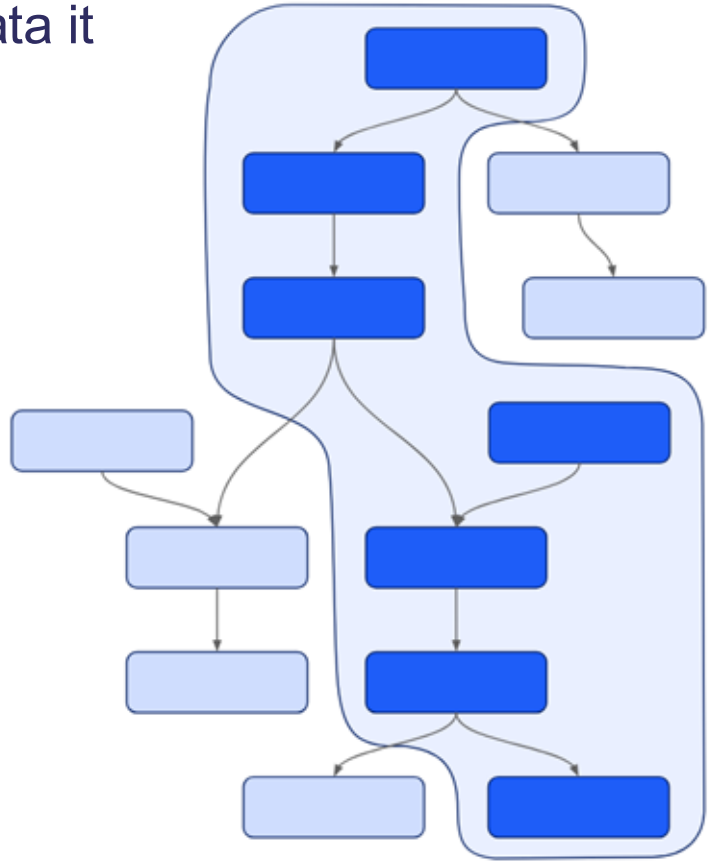
Jola Mirecka

Science and Technology Facilities Council
Scientific Computing

Metadata Collection and Analysis



- Every job records the complete history of all operations
- Every job defines the metadata it returns



- Generate a metadata report for an entire workflow upstream from any job
- Generate literature references for whole workflow - ***please cite all***

Preparing Depositions

Daniel Hatton

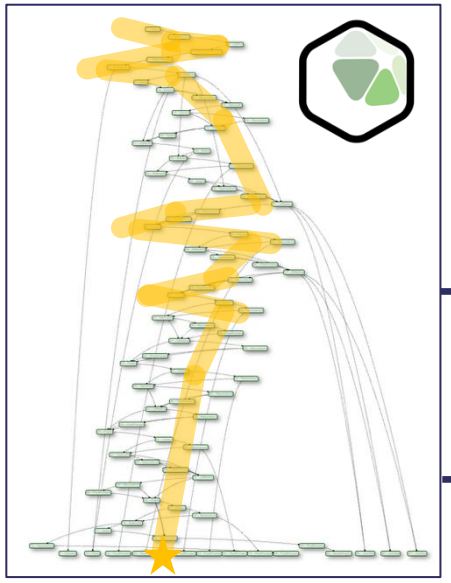
Science and Technology Facilities Council

Chan Zuckerberg Imaging Institute (CZII)

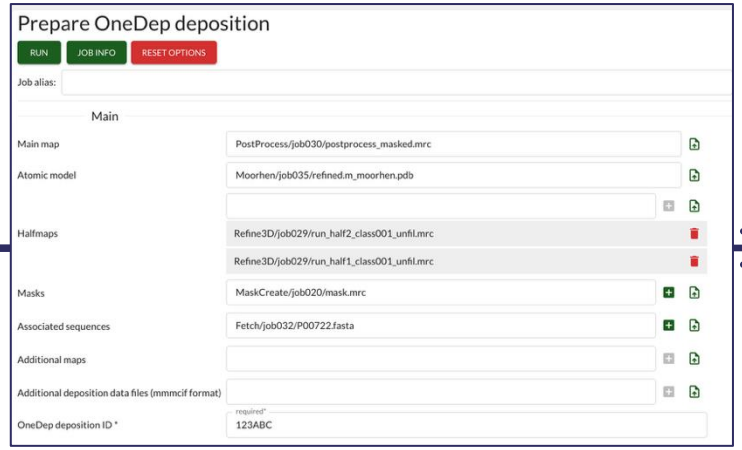
EmHarvester

diamond

Additional CIF files



CCP-EM Doppio



OneDep

PDBe
Protein Data Bank in Europe

EMDB

Matt Iadanza

CCP-EM

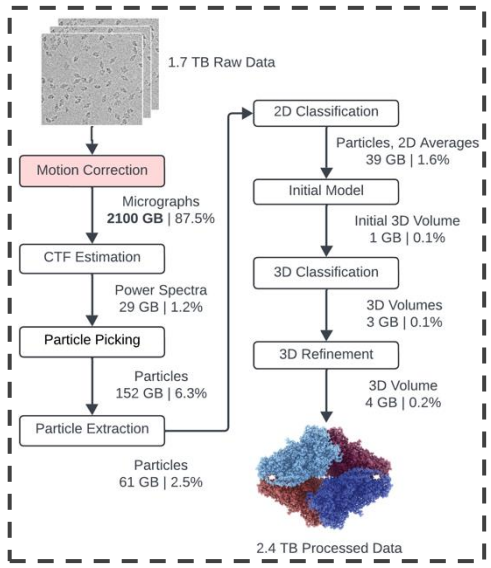
`empiar_depositor`

Deborah Harrus

Kyle Morris

EMPIAR

Cryo-EM data storage (eBIC)



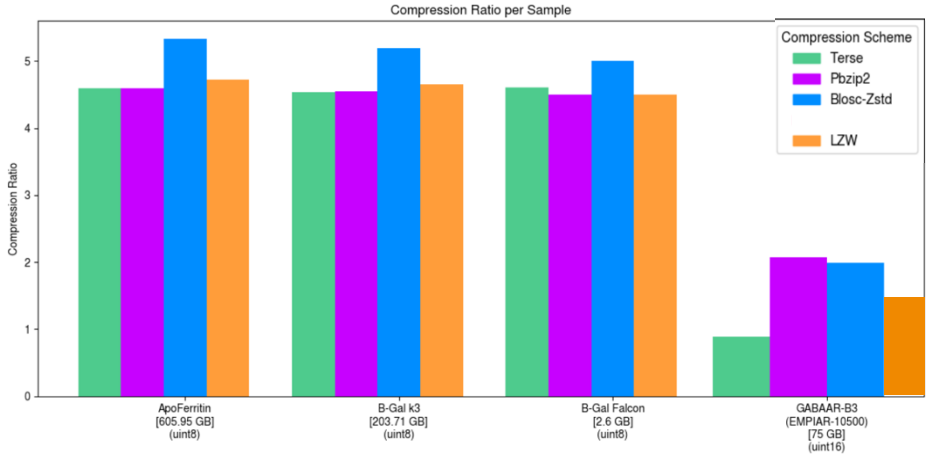

Mariam Demir

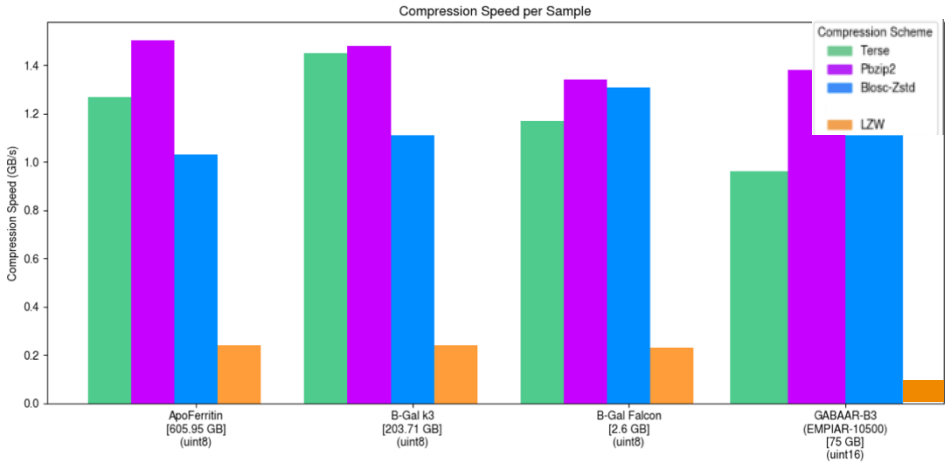

Stephen RIggs


Daniel Hatton

Data storage at eBIC (archive)

- ~10PB of data annually (~400PB worldwide)
- Raw movies: ~45% of total data
- MotionCorr output : ~47% of total data





Data reduction strategies tested

- **Benchmark new compression methods** for raw movies (micrographs and particle stacks).
  Better compression (e.g. using Blosc-Zstd) results in >5% (>0.5PB) reduction in storage

- Implement support for **generating corrected micrographs (RELION) using pre-calculated shift parameters**.
  Removing MotionCorr output from storage (and generating on-demand) results in ~47% reduction

UKRI Science and Technology Facilities Council

Ada Lovelace Centre

EMencode

Science and Technology Facilities Council
Ada Lovelace Centre

- Developed from a benchmarking framework to a more general purpose cryo-EM compression and data archival package.
- Various new features:
  - HDF5 archival workflow
  - Extra compression algorithms, namely lossy bit-truncation
  - File format conversion tools (tiff to mrc, and Ome-Zarr support)
  - Addition of some preliminary compression analysis tools
  - User friendly CLI, GUI and python API interfaces
  - Plugin system for supporting future compressors, file formats etc.

Willow Sparks

Mariam Demir

https://data-compression-framework-dbcb2f.gitlab.io

https://gitlab.com/ccpem/data-compression-framework

Science and Technology Facilities Council
Scientific Computing

Upcoming: Live processing in Doppio!

A collaboration between **CCP-EM** and **eBIC** (The UK's National Cryo-EM facility).

eBIC's automated processing pipeline integrated into Doppio. Run automated SPA processing (motion correction to post processing) as micrographs are produced. These automated processing 'recipes' can be customised.

Deployable on GPU workstations, clusters, and HPC. Integrated solution allows for easy deployment with Doppio at Cryo-EM facilities, labs, and industry around the world.

Run experimental jobs while data is processed, continue further processing, or start model building all in Doppio.

Live data dashboard (based on components from eBIC's PATo).

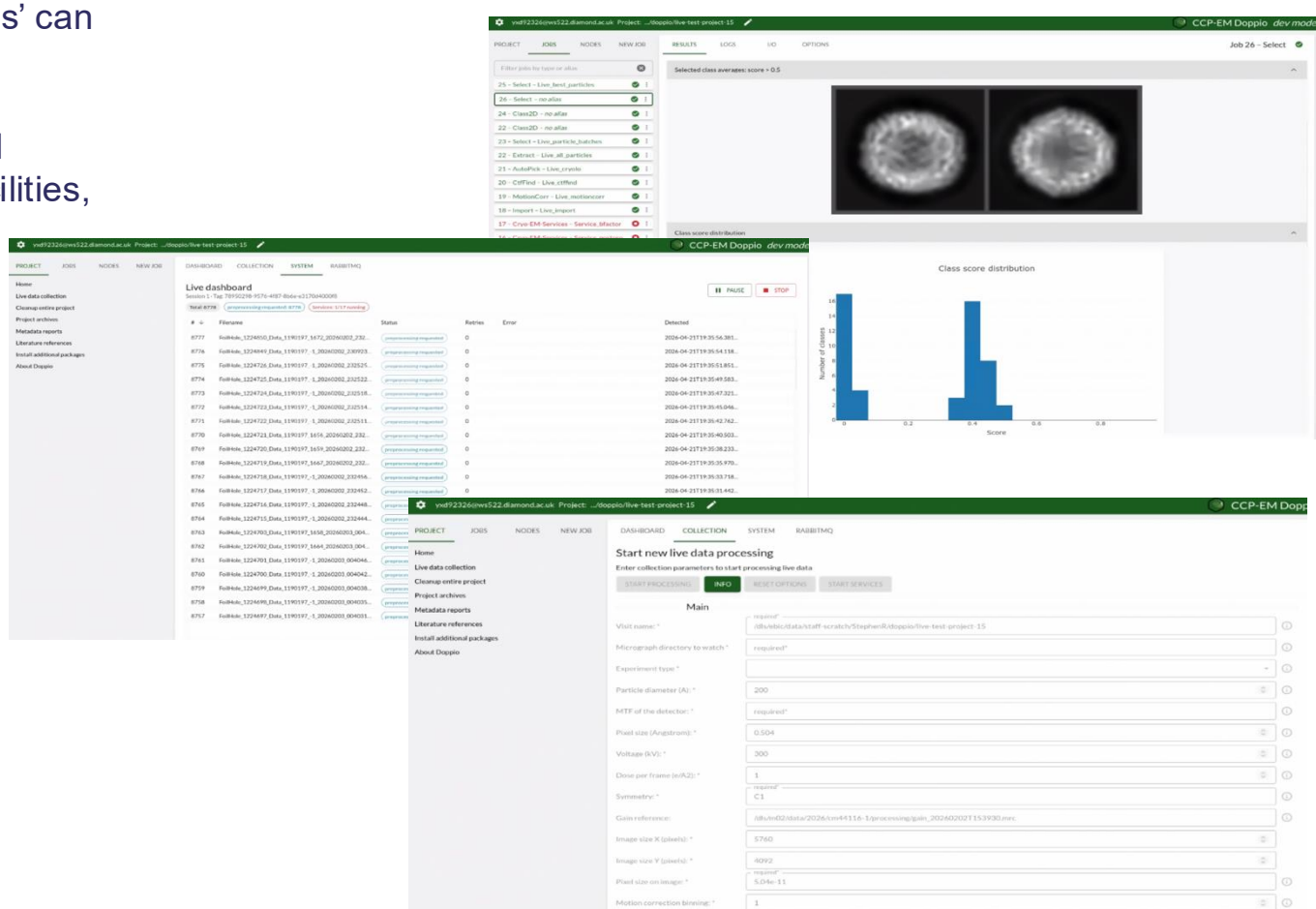
(Later this year) Live tomography processing pipeline in Doppio (motion correction through to alignment, denoising, and segmentation).


George Coldstream


Stephen RIggs


Daniel Hatton

Guilherme De Freitas

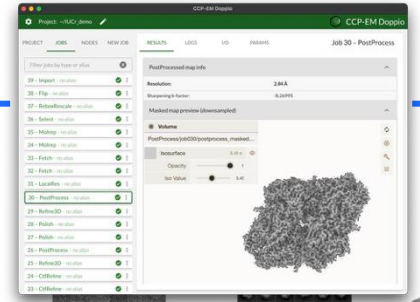


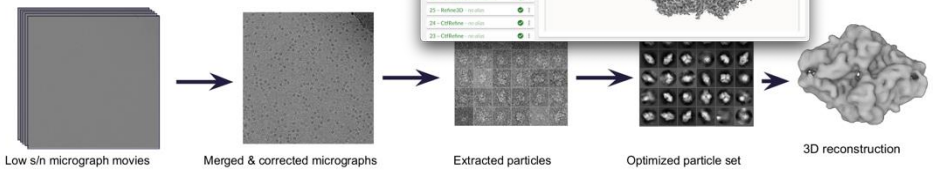
Science and Technology Facilities Council
Scientific Computing

Science and Technology Facilities Council
Ada Lovelace Centre

Doppio-Vivo

GUI

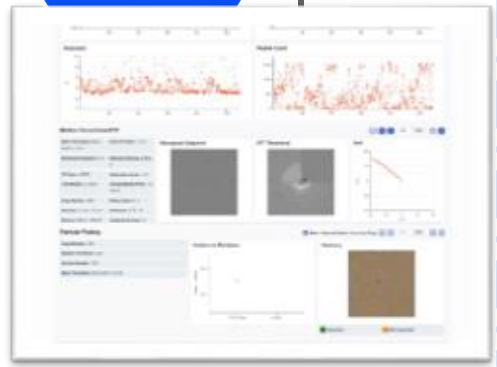




Doppio

Doppio

PATo



eBIC pipeline

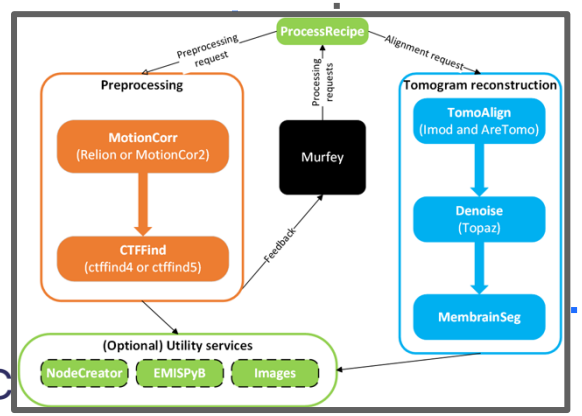
Managers

Murfey

CryoEM Services

Pipeliner

Tasks



Relion

crYOLO

Motion Corr2

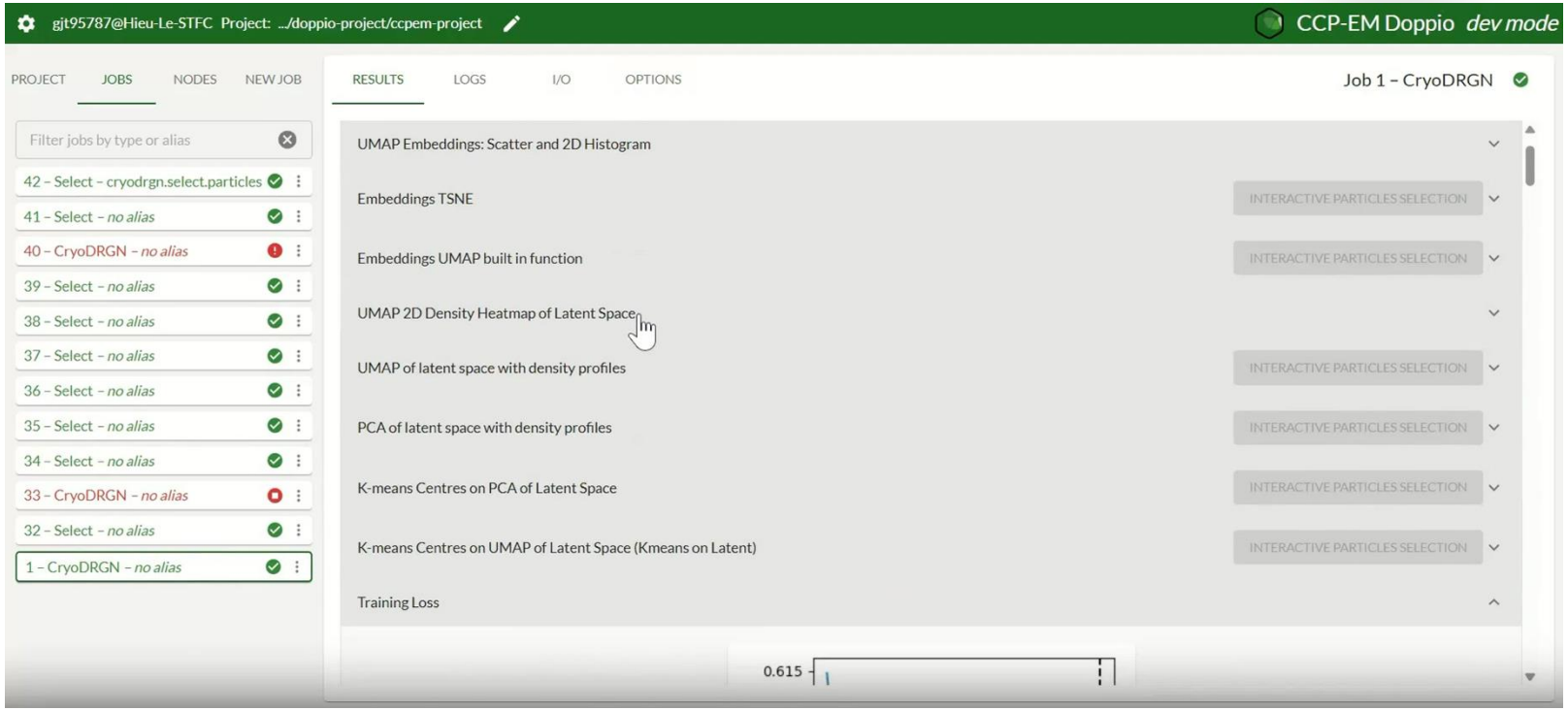
...

Scientific C

Stephen Riggs, George Coldstream, Daniel Hatton, Guilherme De Freitas

Upcoming: dynamic latent space visualisation

- Dynamic latent visualisations for heterogeneous reconstruction algorithms
- New LatentSpace class with routine latent space utility functions:
  - dimensionality reduction (PCA / UMAP / t-SNE)
  - clustering (k-means, GMM, etc.)



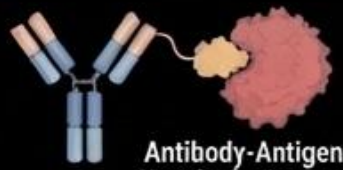
Science and Technology Facilities Council
Ada Lovelace Centre

Science and Technology Facilities Council
Scientific Computing

ADVANCING THE FRONTIERS OF STRUCTURAL BIOLOGY

Structural biologists: **WE NEED YOUR TARGETS FOR CASP17!**

1 IMMUNE COMPLEXES



Antibody-Antigen
Nanobody-Antigen
T-Cell Receptor complexes

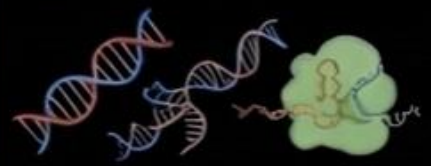
A MAJOR FAILURE AREA FOR CURRENT DEEP LEARNING.
We need rich & varied non-homologous sets.

2 ORGANIC LIGAND-PROTEIN COMPLEXES

Novel Ligand Chemistries
Affinity Rankings

RESULTS OFTEN FALL SHORT OF EXPERIMENTAL ACCURACY.
Sets of **3D** complexes for specific receptors.

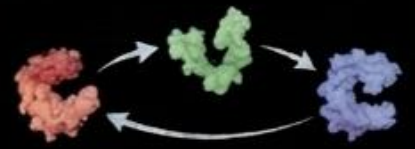
3 NUCLEIC ACIDS & COMPLEXES



Non-homologous RNA/DNA
Protein-Nucleic Acid Complexes

DEEP LEARNING USUALLY NO BETTER THAN CLASSICAL APPROACHES. Often fail without homologous info.

4 CONFORMATIONAL ENSENBLES



Multiple High-Resolution Conformations
Low-Res Data (cryo-ET, SAXS, NMR)

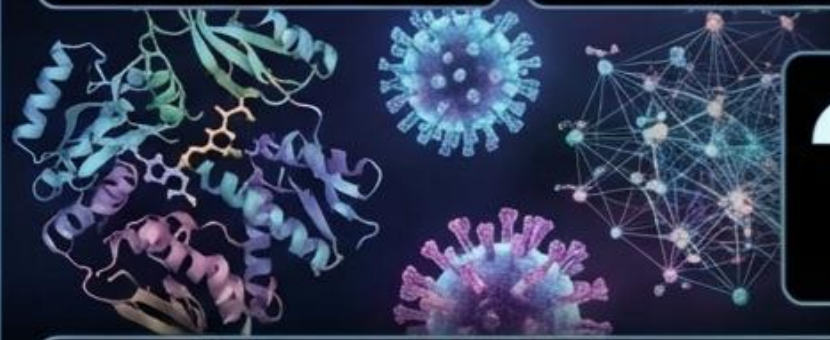
TESTING COMPUTATION OF MULTIPLE STATES.
Major expansion area in CASP.

5 DIFFICULT PROTEIN STRUCTURES & COMPLEXES

Membrane Proteins

Viral/Parasite, Weak Evolution Info
Complex Stoichiometry

CRITICAL WEAKNESSES IN DEEP LEARNING.
Large proteins and recently evolved interfaces.



? RULE OF THUMB:

If AlphaFold3 can generate a high-quality model, it is likely **NOT** a CASP-grade challenge.

IF IT STRUGGLES, WE WANT IT.

SUBMISSION GUIDELINES & DEADLINES

SUBMISSION WINDOW: Now through July 1, 2026.
DATA DELIVERY: Experimental coordinates/data by Sept 1, 2026.

HOW TO CONTRIBUTE:

Web Portal (Preferred): CASP17 Target Form
Direct Email: casp@ucdavis.edu
PDB Submission: Designate "on-hold" as CASP target

INCENTIVES:

Co-author papers in the special CASP issue.

The CASP Organizers: John Moult, Krzysztof Fidelis, Andriy Kryshtafovych, Torsten Schwede, Maya Topf

Science and Technology Facilities Council

Scientific Computing

The CAHRA Challenge

Joel Greer

23rd April 2026

CCP-EM Spring Symposium, Nottingham

CCP-EM FLATIRON INSTITUTE PRINCETON UNIVERSITY

Community-Wide Assessment of Cryo-EM Heterogeneous Reconstruction Algorithms

- Faithfully reconstructing the structural heterogeneity present in cryo-EM datasets remains a challenge

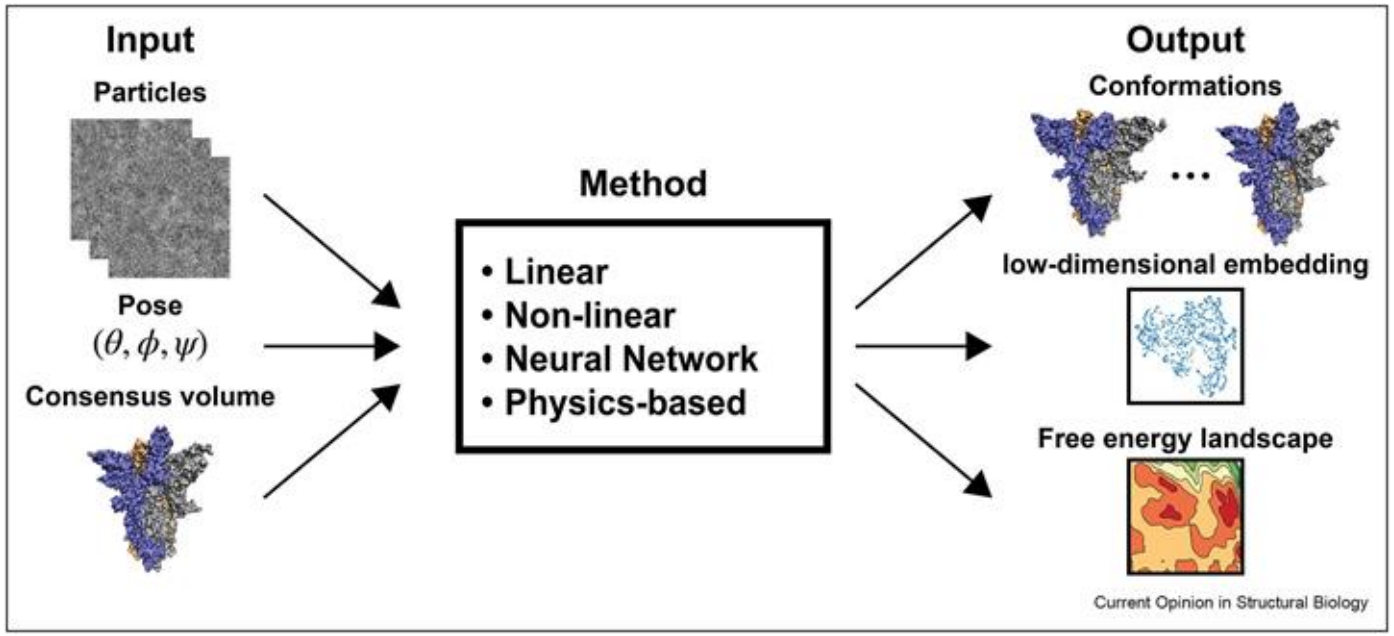


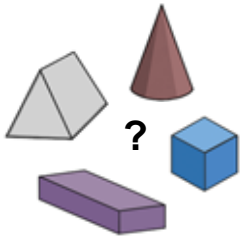
Fig 2, Tang et al., Current Opinion in Structural Biology, 2023

- To assess how successfully a given algorithm/workflow can recover the heterogeneity present, we can use metrics to assess their performance on benchmark datasets

The CAHRA Challenge Datasets

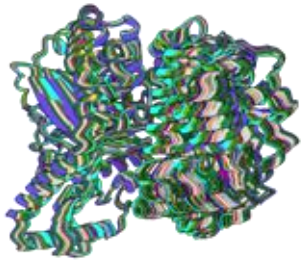
- Curated 3 datasets which address different aspects of heterogeneity present in cryo-EM datasets

1: Compositional Heterogeneity



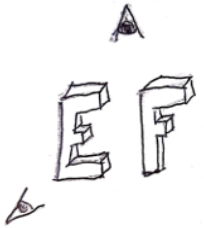
- Pose estimation and reconstruction of mixture of unknown particles
- Motivated by visual proteomics

2: Conformational Heterogeneity



- Conformations sampled via MD
- Tasked to reconstruct a representative ensemble of atomic models

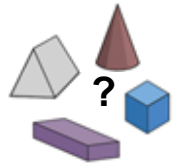
3: Pose Entanglement with Heterogeneity



- Non-uniform pose distribution
- Estimate poses and populations of two discrete states

- Facilitate the development and evaluation of new and existing HRAs
- Provide an interface for the cryo-EM community to address the challenge of interpreting the heterogeneity in cryo-EM datasets

1: Compositional Heterogeneity Dataset



Participants are given a "mystery" single particle dataset containing a mixture of samples:

- 1.6M particle stack, no poses
- Per-particle defoci
- Unknown sample composition

Goal: Assess particle classification accuracy and motivate the development of automated analyses

Task: Reconstruct as many complexes as possible and identify the best subset of particles for each class

Participants are asked to submit:

- Density maps for each identified complex
- Poses, CTF and class assignments
- Workflow overview

Example cryoSPARC workflow.

2: Conformational Heterogeneity Dataset

Participants are given a single particle dataset containing a range of conformations of a macromolecule:

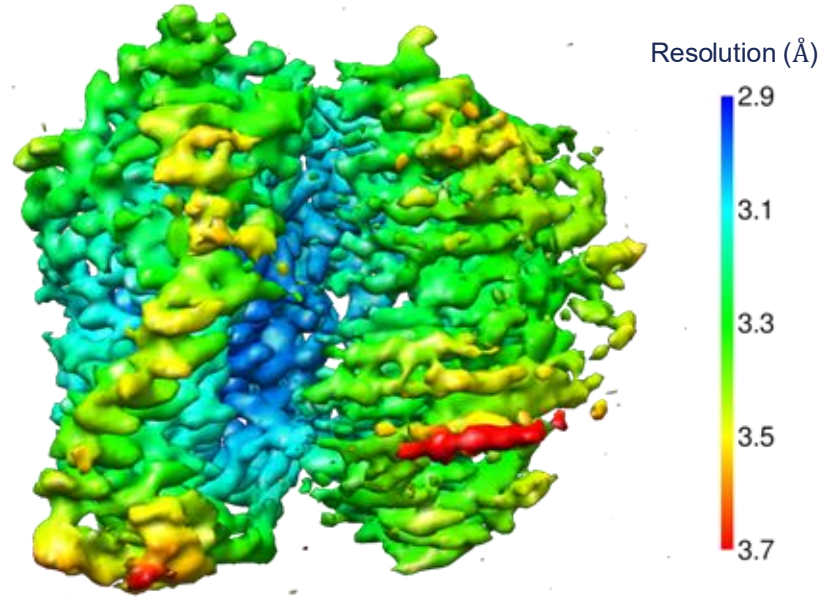
- 1.7M particle stack, poses from consensus reconstruction
- Per-particle defoci
- FASTA sequence of the molecule

Goal: Assess the capacity of workflows to faithfully extract the heterogeneity present

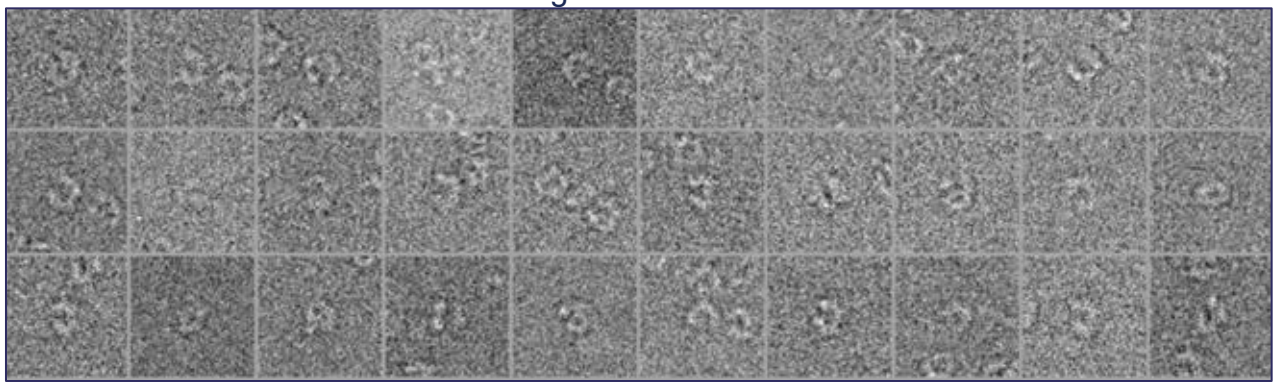
Task: Determine an ensemble of $10 \leq N \leq 100$ atomic models which best represent the conformational heterogeneity present in the particle stack

Participants are asked to submit:

- Ensemble of predicted atomic models
- Optionally:
  - Latent space representations
  - Density map representations
  - To build a community dataset for metrics development



Local resolution estimated from consensus reconstruction of the conformational heterogeneity challenge dataset



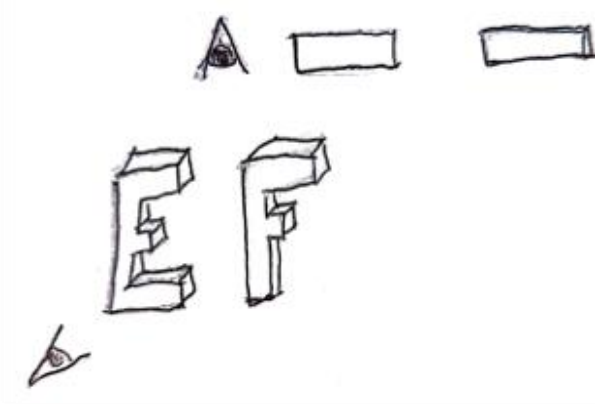
Montage of particles from the conf-het dataset

3: Pose Entanglement Dataset

In a single particle cryo-EM dataset, a non-uniform pose distribution can confuse the results of heterogeneity analysis, with different states being more difficult to identify from certain poses.

Participants are given:

- 30k particle stack, no poses
- Per-particle defoci
- Template maps for the 2 discrete states present

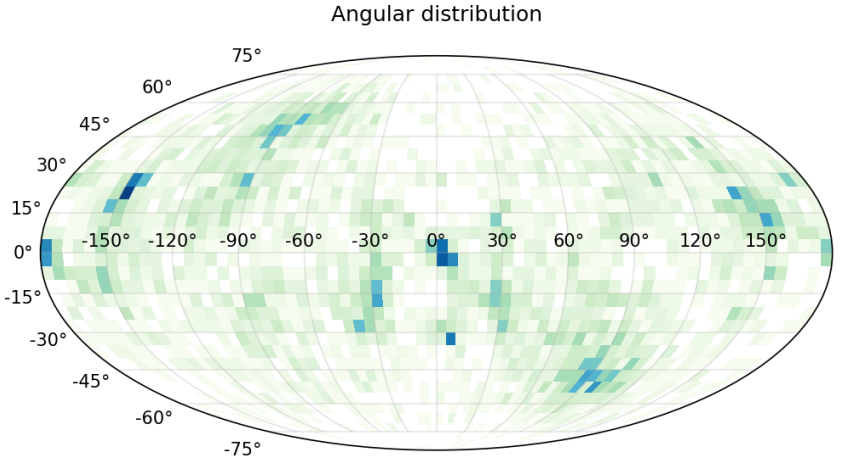


Goal: Assess the capacity of methods to correctly predict state populations despite a non-uniform pose distribution

Task: Determine the relative populations of the two states and estimate per-particle states and poses

Participants are asked to submit:

- Ensemble population of each state
- Optionally: Per image state and pose estimations



Challenge Timeline and Prizes

Datasets released, open for submissions	Nov 1st 2025
Launch Webinar	Nov 14th 2025
Deadline for preliminary submissions	March 14th 2026
Deadline for final submissions	**June 1st 2026**
Results announced	**July 2026**
CAHRA Challenge Review Workshop	**Sept 16th-18th 2026**

- Winner of each challenge will have access to a travel and accommodation bursary to attend the Challenge Review Workshop
- There is also a total of $10,000 in monetary awards for top-performing methods

- 11 preliminary entries across challenge datasets
- Rankings for preliminary entries will be released in the coming days!

Where To Find More Information

Visit the challenge website!

https://heterogeneity.notion.site/challenge

- Keep up-to-date:
  - Cahra-challenge@jiscmail.co.uk
  - Github discussions page
  - For private correspondence, each challenge has contact information
  - Leaderboard

- More details:
  - Links to the challenge datasets
  - Submission instructions, formats and examples
  - CAHRA Review Workshop details
  - Anonymity policy

Welcome entries from anyone – not just method developers!

CAHRA Acknowledgements

Dr. Jas Kalayan
(UKRI-STFC)

Geoffrey Woollard
(University of British Columbia)

Dr. James Gebbie-Rayet
(UKRI-STFC, CCPBioSim, HECBioSim)

David Silva-Sánchez
(Yale University)

Dr. James Parkhurst

CCP-EM Group Members

Maarten Joosten
(TU Delft)

Prof. Arjen Jakobi
(TU Delft)

The CAHRA Team


Dr. Pilar Cossio
(Flatiron Institute)


Dr. Sonya Hanson
(Flatiron Institute)


Dr. Tom Burnley
(CCP-EM, UKRI-STFC)


Dr. Joel Greer
(CCP-EM, UKRI-STFC)


Prof. Ellen Zhong
(Princeton)


Dr. Ryan Feathers
(Princeton)


Robert Heeter
(Princeton)

Science and Technology Facilities Council

CCP-EM

FLATIRON INSTITUTE

DRIIMB

Yale

Engineering and Physical Sciences Research Council

Research Complex at Harwell

CoSeC

HECBioSim

CCPBioSim

TU Delft

UBC

Website & Downloads: www.ccpem.ac.uk

Mailing list: www.jiscmail.ac.uk/ccpem

Bluesky: ccpem.bsky.social

Email: ccpem@stfc.ac.uk

CCP-EM Core Team

CCP4 Core Team

STFC Business & Innovation

CCP-EM Commercial License Holders

CCP-EM Collaborators

CCP-EM Users

MRC Core funding

https://heterogeneity.notion.site/challenge

Research Complex at Harwell

wellcome

University of Pittsburgh

Department for Science, Innovation & Technology

GΦL Global Phasing Limited

ada lovelace centre

CoSeC