Welcome!

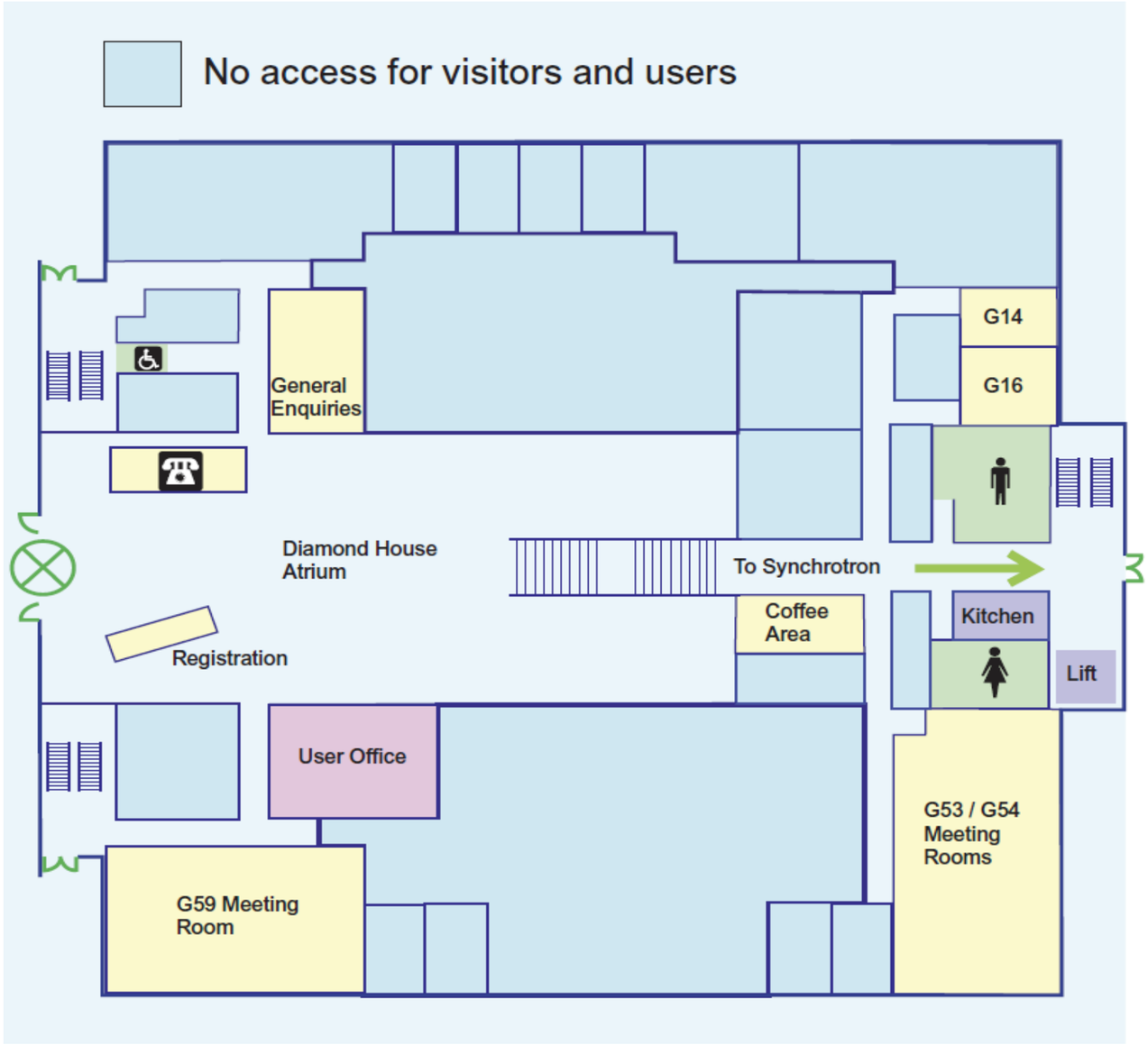
6th DLS-CCP4 Data Collection and Structure Solution Workshop

Fire Alarm

If the fire alarm sounds, there will be an alarm which alternates with a voice message stating that "there is an emergency in the building".
You should leave the building by the closest fire exit and congregate in the carpark in front of Diamond House.

Emergency Contacts

In case of emergency within Diamond House, please call the RAL emergency line on **2222** from any internal phone or **01235 778888** from a mobile.



How to get the most out of the workshop

- **Ask questions!** The tutors are experts and developers of the software you will use. They are not all available for the full week, please approach them if you need help.
- **Talk to your neighbours**. The attendees at the course have various levels of expertise and scientific backgrounds. You may learn from each other and not just the tutors.
- **Take your time**. You will get the most out of the tutorials if you work through each step carefully and think about what you are doing
- **Course materials** will be uploaded to the website http://www.ccp4.ac.uk/schools/DLS-2019, so don't feel you have to copy every slide. You will have them available at the end of the workshop.

If progress you make during the workshop leads to published results we ask that you please acknowledge this in your papers. We will link to them on our website. This helps us to keep delivering similar workshops in the future.

We wish you every success with your projects!

DLS-CCP4 Data Collection and Structure Solution Workshop



This page lists some of the publications where the research involved benefited from the assistance given at the DLS-CCP4 workshop. We are grateful to each of the authors for acknowledging the help they recieved at the workshop.

- 2017 Publications
- 2015 Publications
- 2014 Publications

2017 Publications

The authors of the following publications have expressed their gratitude for the help they received in solving their structures at the DLS-CCP4 school in 2017.

- **"Crystal structure of human Acinus RNA recognition motif domain"** –
  Humberto Fernandes, Honorata Czapinska, Katarzyna Grudziaz, Janusz M. Bujnicki and Martyna Nowacka. **PeerJ** (2018).
- **"Crystal Structure and Directed Evolution of Specificity of NlaIV Restriction Endonuclease"** –

DLS-CCP4 Data Collection and Structure Solution Workshop

December 1- 8, 2019, Diamond Light Source Ltd., UK

Introduction to CCP4: The Structure Solution Roadmap

Eugene Krissinel & Ronan Keegan

CCP4, Research Complex at Harwell
Didcot, Oxfordshire, UK

Science & Technology Facilities Council

CCP4

Science & Technology Facilities Council

Research Complex at Harwell

Outline

- What is CCP4?
- Educational Outreach
- The CCP4 Software Suite
  - What is there in CCP4
  - Using CCP4 – the interface(s)
  - The structure solution roadmap
- Other useful information
- CCP4 in the cloud

Science & Technology Facilities Council

Research Complex at Harwell

What is CCP4?

Science & Technology Facilities Council
Research Complex at Harwell

- CCP4 stands for *"Collaborative Computational Project No. 4 in Protein Crystallography"*
- One of several CCPs set up in the UK to advance and support scientific software developments
- CCP4 was formally set up in ***1979*** to support collaboration between researchers working on MX software in the UK, and to assemble a comprehensive collection of software to satisfy the computational requirements of the relevant UK groups
- Funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and industrial licencing
- The project is coordinated at Scientific Computing Department of the Science and Technology Facilities Council (STFC)

- Today CCP4's main activities are:
  - Supporting the development and distribution of a comprehensive suite of software for performing structure solution in macromolecular crystallography
  - Providing educational and training events to users of the software including schools/workshops where users can get access to expert assistance with structure solution problems
  - Providing other resources to the user community including the CCP4BB mailing list, Wiki's and access to web/cloud-based computational resources

Science & Technology Facilities Council
Research Complex at Harwell

Main CCP4 Groups

- *Oxford*
  - CCP4 core team – software development, maintenance, distribution, educational outreach and CCP4 online resources
  - Diamond – data processing developments (DIALS)
- *Cambridge*
  - MRC/LMB – software development for refinement (REFMAC) and data processing (MOSLFM, Aimless, etc.)
  - Uni. of Cambridge – Phaser development team for phasing software
- *University of York*
  - Software developments (CCP4mg, Buccaneer, Privateer)
  - CCP4 GUI-2 developments



Science & Technology Facilities Council
Research Complex at Harwell

Main CCP4 Groups

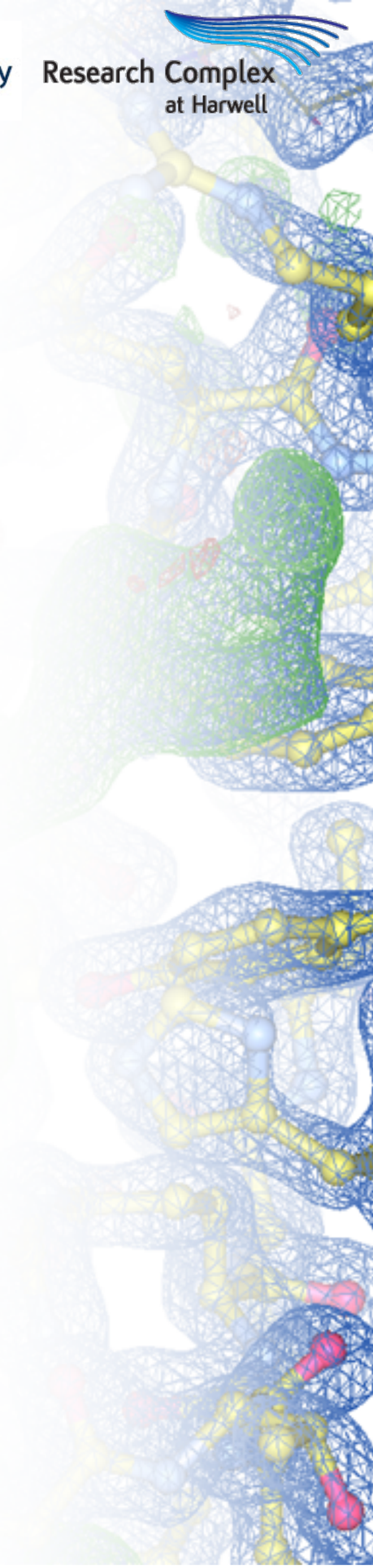
- *University of Liverpool*
  - ~ AMPLE (ab-initio MR)
  - ~ SIMBAD (sequence-less MR)

- *Associated projects*
  - ~ ARP/wARP (Model building, EMBL-Hamburg)
  - ~ PDB redo (Validation and completion, NKI, the Netherlands)
  - ~ SHELX (Experimental phasing, Goettingen and Barcelona)
  - ~ Arcimboldo (Fragment-based MR, Barcelona)
  - ~ Crank-2 (Automated experimental phasing, University of Leiden, The Netherlands)



Science & Technology Facilities Council

Research Complex at Harwell



Educational Outreach

CCP4 Study Weekend

- Takes place every January in the UK
- Each year tackles a different topic:
  - ~Data processing
  - ~Experimental phasing
  - ~Molecular replacement
  - ~Model building and refinement
  - ~Related topics: Complementary methods (EM, NMR, SAX etc.), Structure analysis, Complexes, Ligands, Databases
- Lectures streamed live on the internet

CCP4 Study Weekend

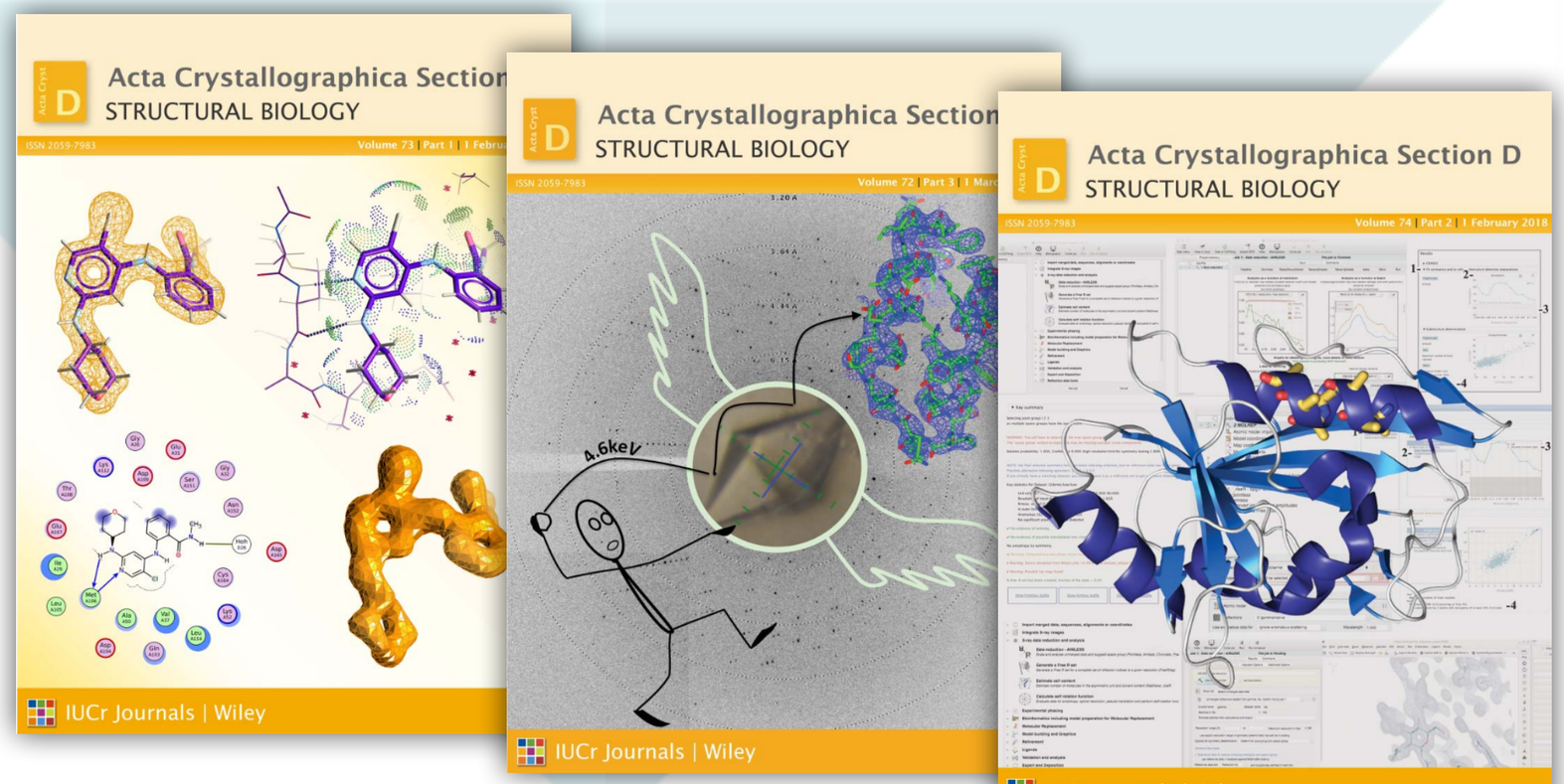
- Proceedings published in special edition of **Acta Cryst. D**

  2017: *"From crystals to structure with CCP4"*

  2018: *"Multi and Serial Crystal Data Collection and Processing"*

  2019: "Molecular Replacement"

  2020: "Model Building"



Science & Technology Facilities Council

Research Complex at Harwell

CCP4 users around the world



Science & Technology Facilities Council
Research Complex at Harwell

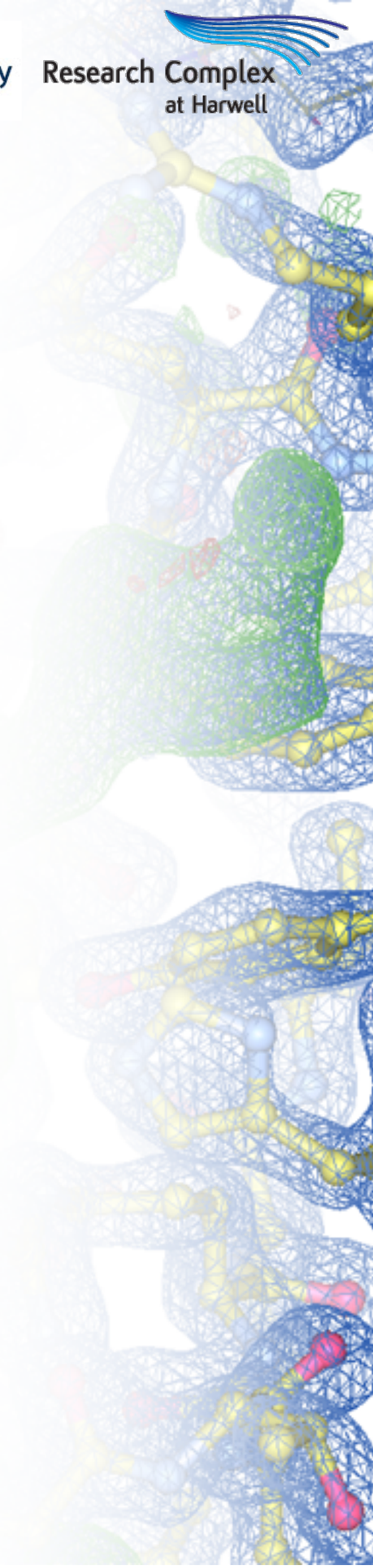
CCP4 Resources

- CCP4BB mailing list — bulletin board, news, expert help and advice
- CCP4 help desk — *ccp4@ccp4.ac.uk* — bug reports, problems with software
- CCP4 wiki — *http://ccp4wiki.org* — program documentation and tutorials
- CCP4 wiki @ Konstanz — *http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Crystallography*

  (thanks to Kay Diederichs)
- CCP4 web site — *http://www.ccp4.ac.uk*

Science & Technology Facilities Council

Research Complex at Harwell



The CCP4 Software Suite

CCP4 Roadmap

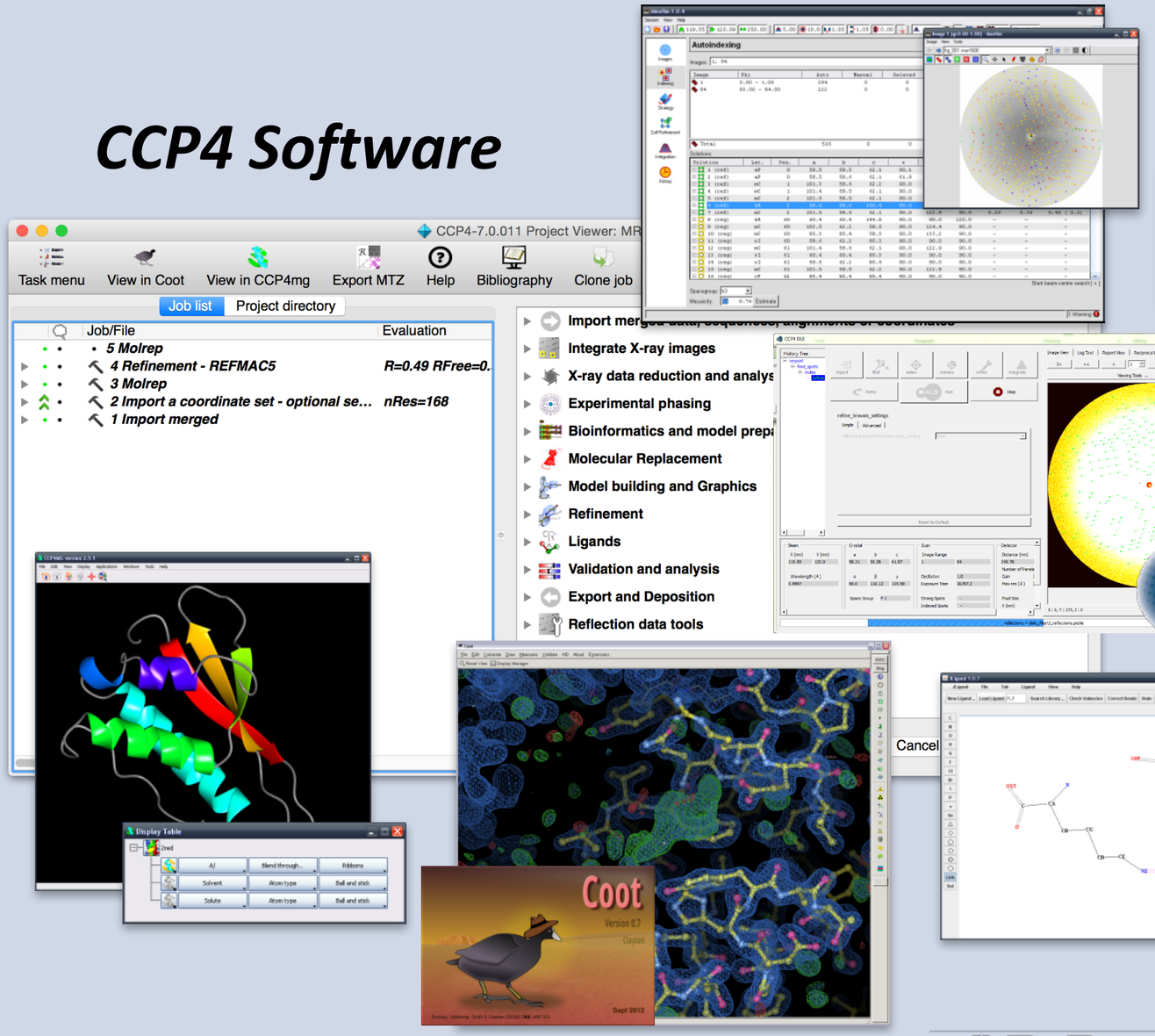
- The CCP4 Suite is a comprehensive suite of software for macromolecular crystallography and it contains about 250 program components
- New versions of the suite are released about every 12 months
  - New programs
  - Major updates to existing programs
  - Other new features such as changes to CCP4i interface
- Revisions to the current release are made available through the CCP4 updates manager

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Scope

CCP4 Software



Crystallisation

Data Collection

Data Processing and Reduction

Experimental Phasing

Molecular Replacement

Density Modification

Model Building

Refinement

Structure Analysis

Deposition

SHELX

ARP wARP 7.3

DIALS

PDB_REDO

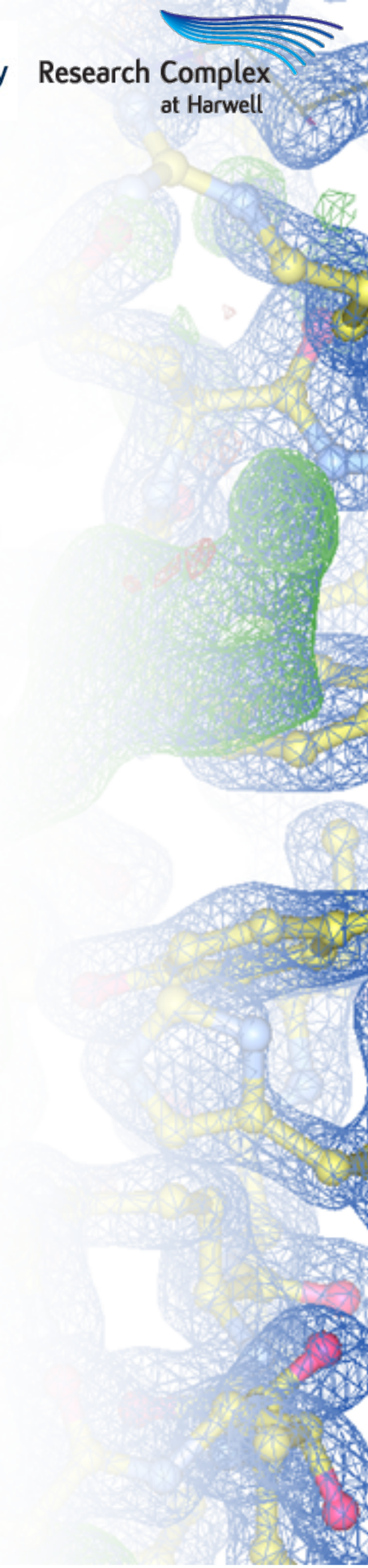
Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Releases

- Current release version — CCP4 7.0 (this workshop: 7.1)
- Recent programs:
  - SHELX, DIALS, Arcimboldo, CCP4i2
  - QtPISA, Aimless, Dimple, AMPLE, SIMBAD, Zanuda, ProSMART, Nautilus, AceDRG, Privateer, Feckless and other
  - In collaboration with the EMBL-Hamburg, ARP/wARP model building software is now rolled with CCP4 Suite
- Continuing development of major components:
  - Phaser, Refmac, Mosflm/iMosflm, Crank-2, MrBUMP
- Newer infrastructure
  - Build system, Setup Manager and Updates Manager

Science & Technology Facilities Council

Research Complex at Harwell



Using CCP4

Science & Technology Facilities Council

Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

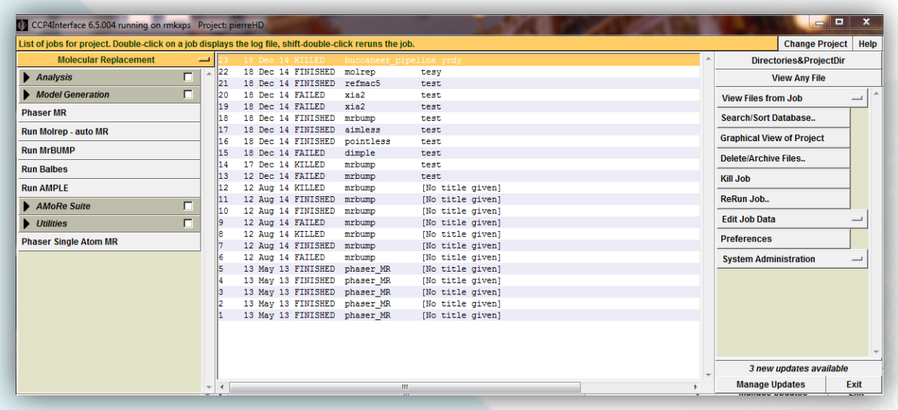
Science & Technology Facilities Council

Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

1. **CCP4i** – original interface developed around 2000



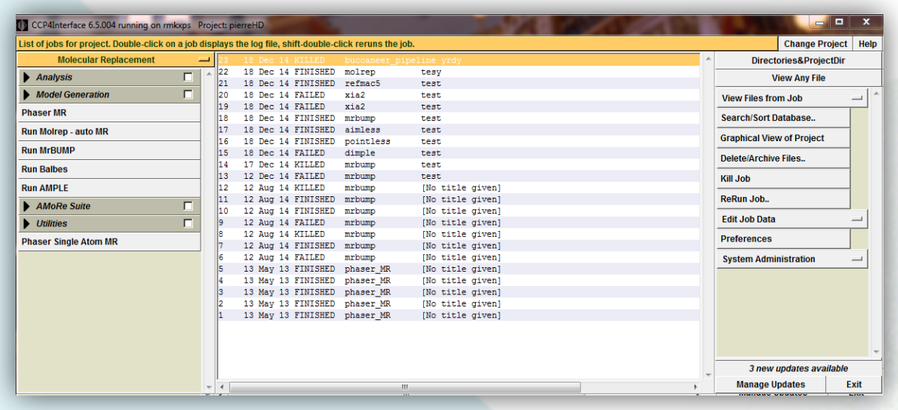
Science & Technology Facilities Council

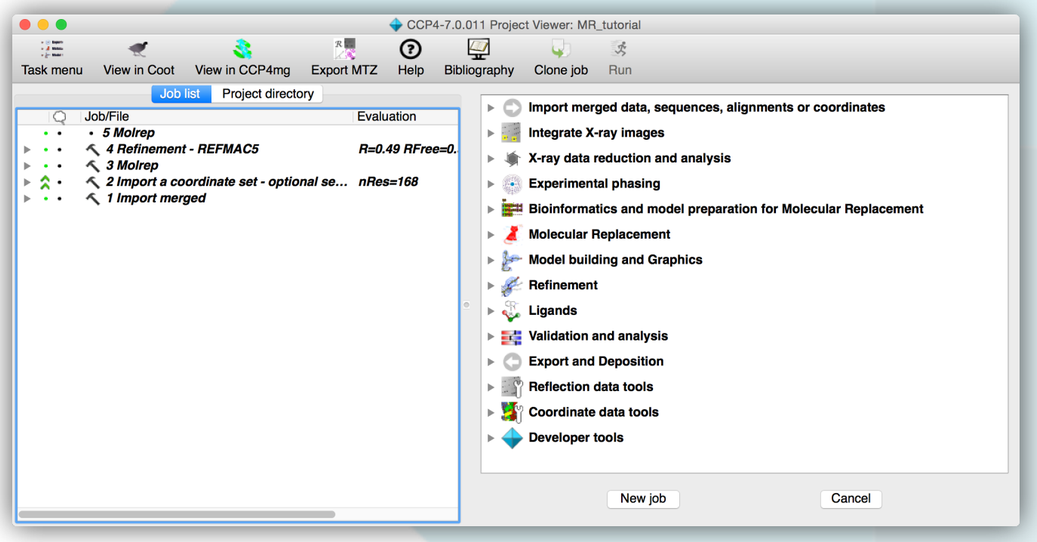
Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

1. **CCP4i** – original interface developed around 2000



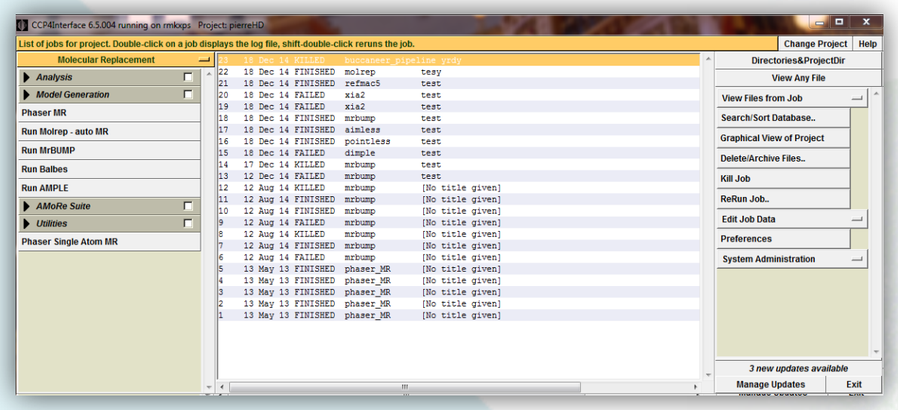


2. **CCP4i2** – new graphical desktop interface (2016)

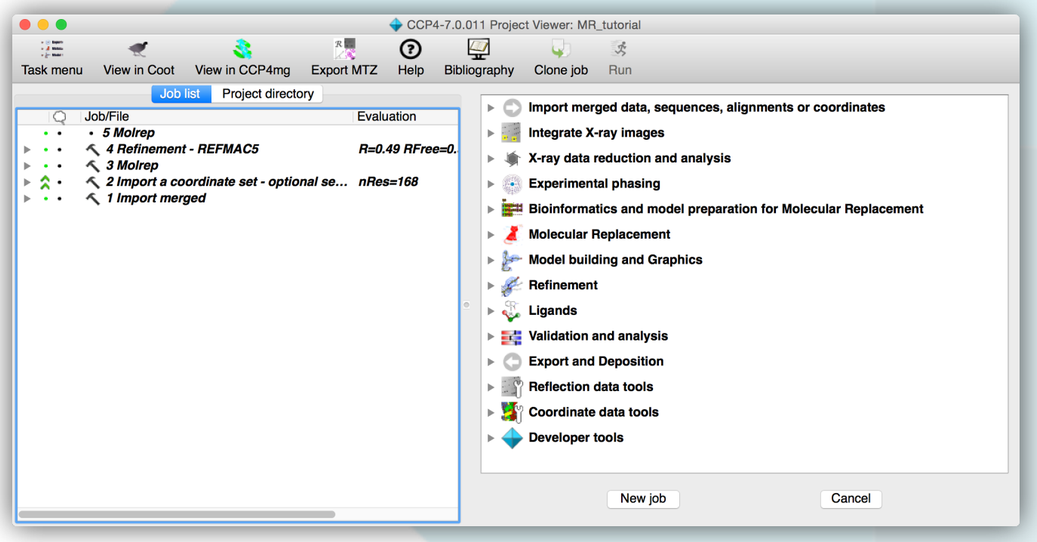
The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

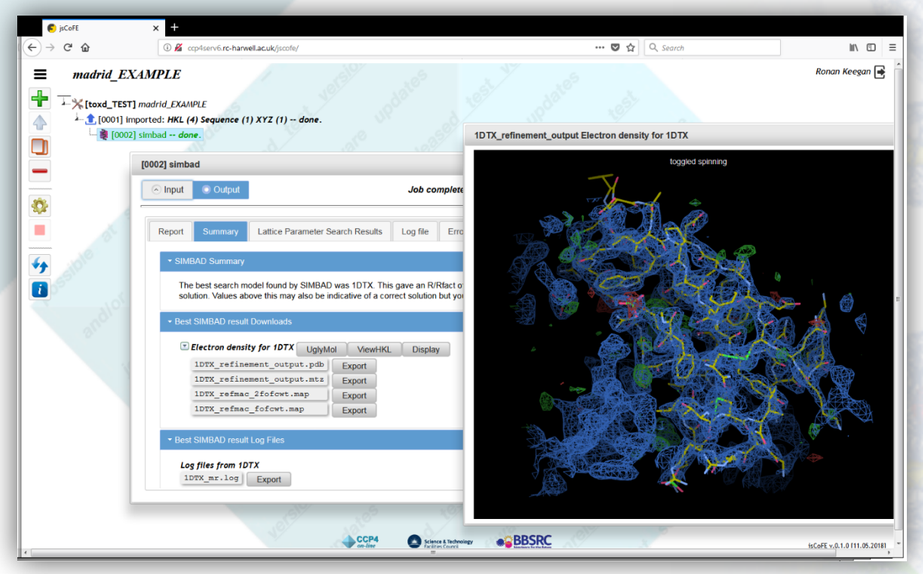
1. **CCP4i** – original interface developed around 2000



2. **CCP4i2** – new graphical desktop interface (2016)



3. **CCP4 Cloud**, browser based (2019)

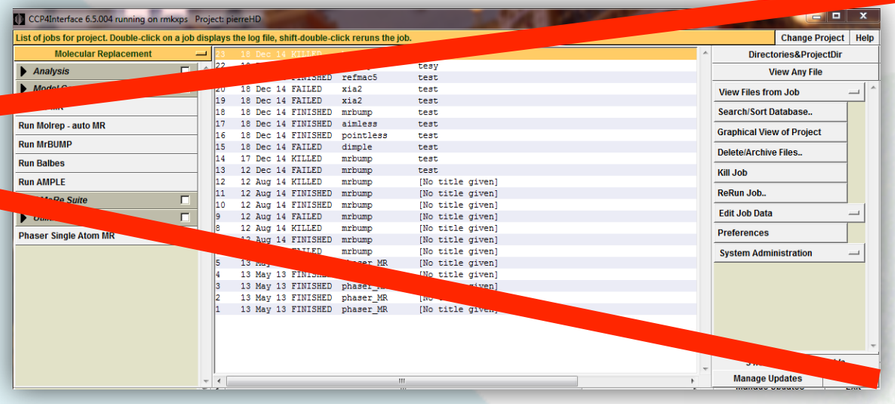


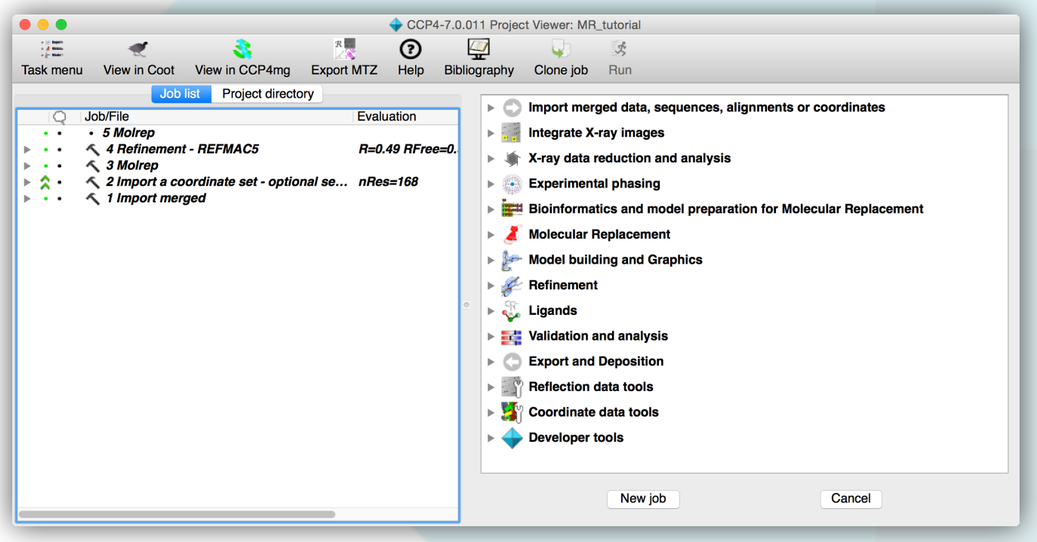
The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

1. **CCP4i** – original interface developed around 2000

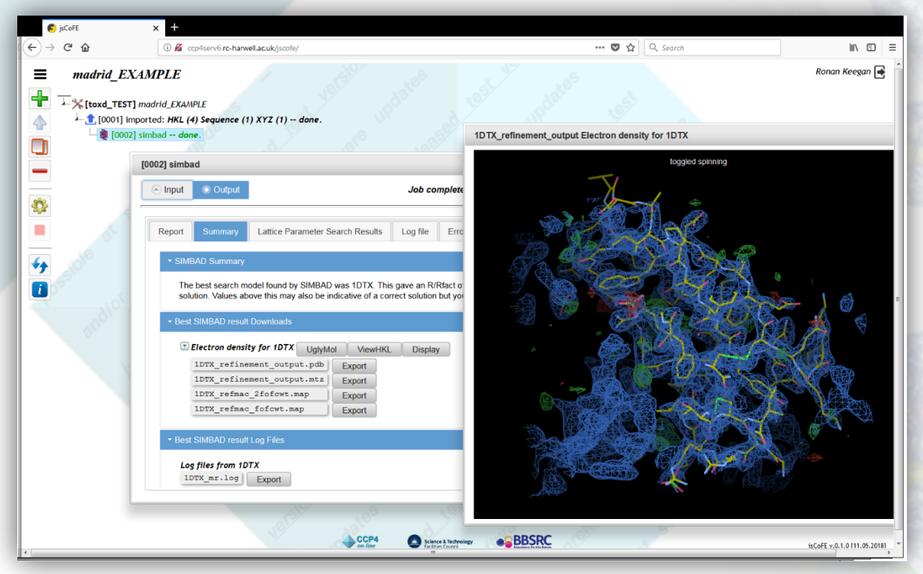
to be discontinued





2. **CCP4i2** – new graphical desktop interface (2016)

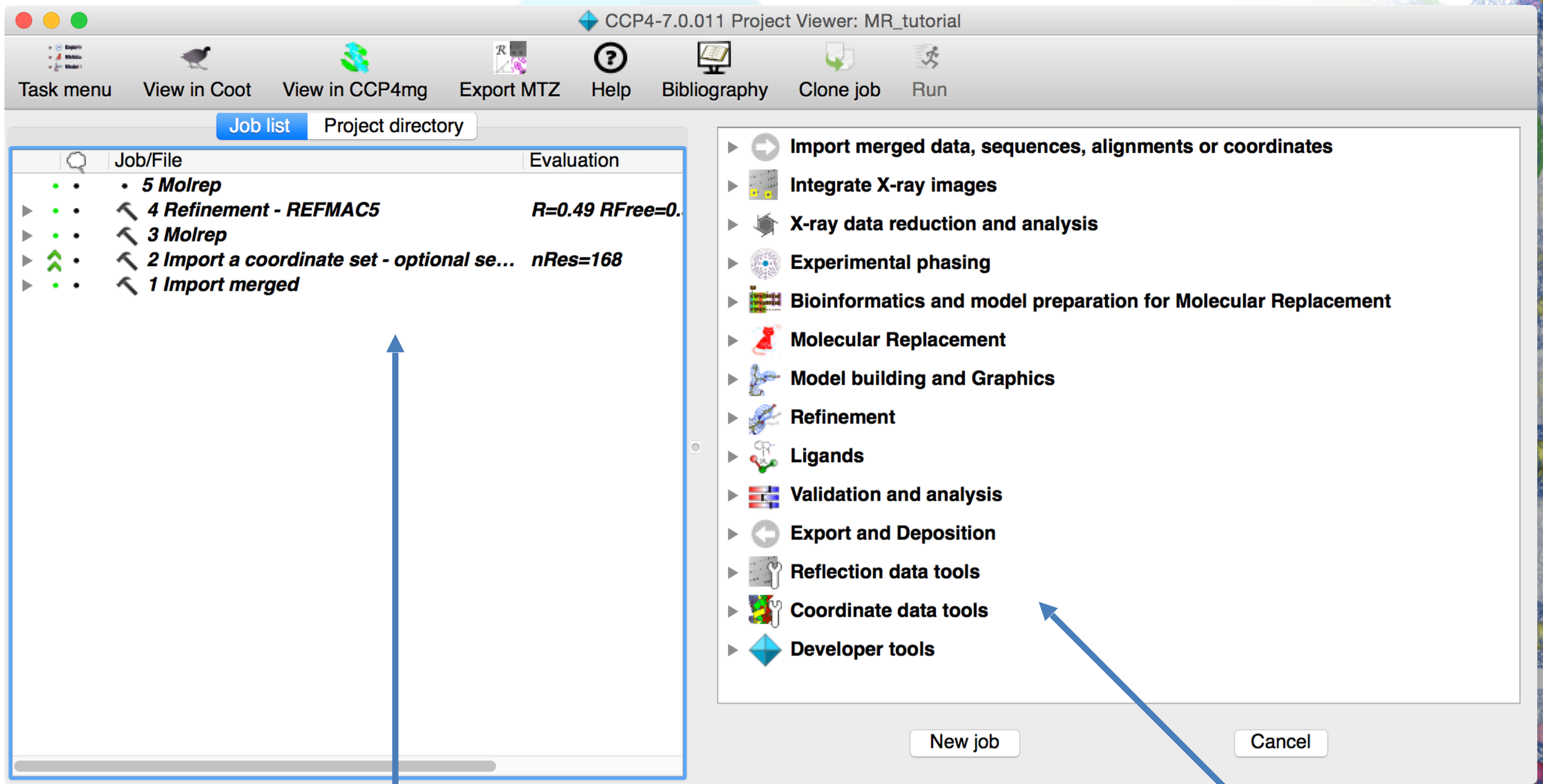
3. **CCP4 Cloud**, browser based (2019)



Science & Technology Facilities Council

Research Complex at Harwell

The CCP4i2 Interface



Job list

Task menu

Science & Technology Facilities Council

Research Complex at Harwell

The CCP4i2 Interface

CCP4-7.0.011 Project Viewer: MR_tutorial

Task menu | View in Coot | View in CCP4mg | Export MTZ | Help | Bibliography | Clone job | Run

Job list | Project directory

Job/File	Evaluation
6 Autobuild - BUCCANEER	
5 Molrep	
4 Refi	R=0.49 RFre
3 Mol	
2 Imp	ptional se... nRes=168
1 Imp	

Delete
Clone
Edit label
Open ▶
View ▶
Export ▶
Next task.. ▶

Input
Output
Comments

Job actions (right-click)

Job report

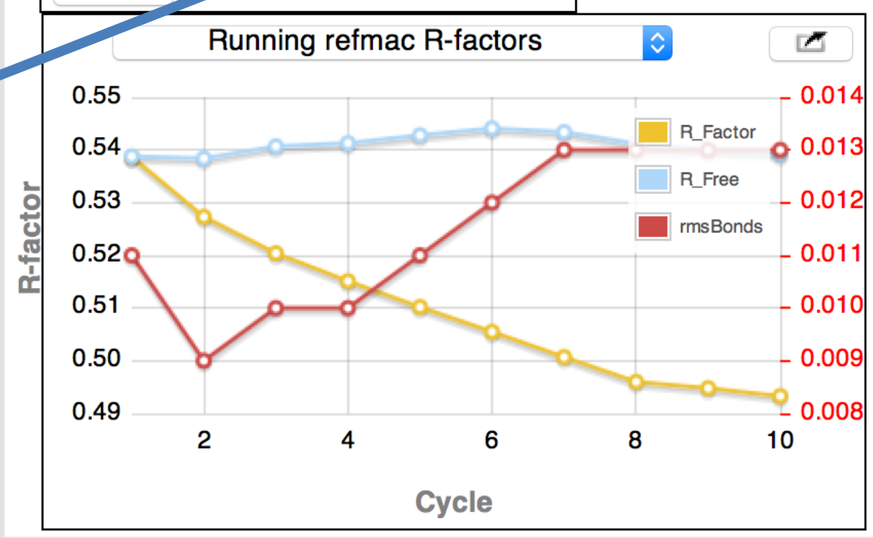
Job 4: Refinement - REFMAC5 **The job is Finished**

Input | Results | Comments

Default weight | Per cycle | Picture | Outliers | Other | Biblio | Run

No. reflections all/free	21135 / 106
R-factor/R-free	0.492 / 0.53
RMS Deviations	
Bonds	0.0135
Angles	1.782
Chain mean B (No. atoms)	
AA	17.7914 (1
Aa	42.3363 (6

Download

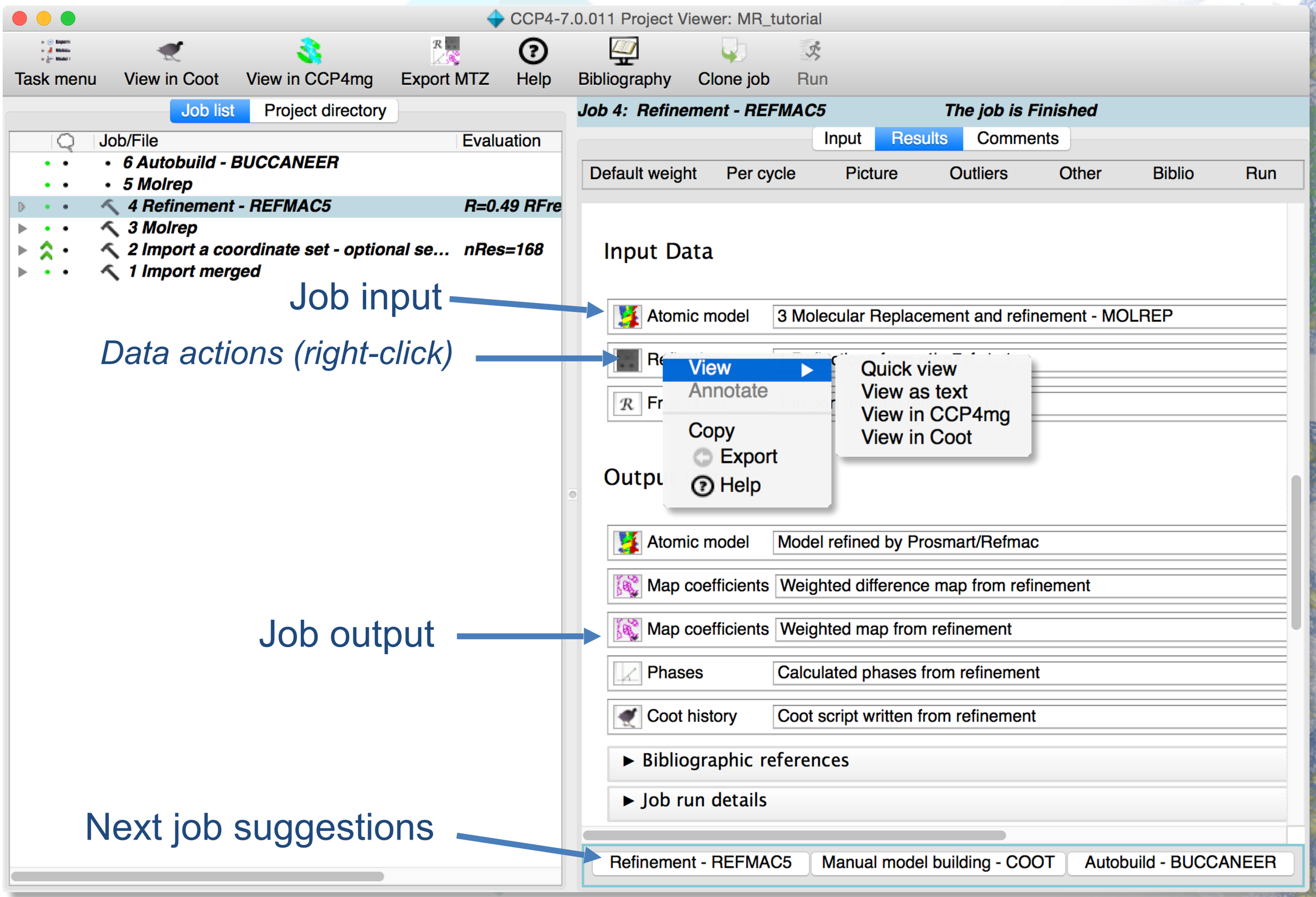


Refinement - REFMAC5 | Manual model building - COOT | Autobuild - BUCCANEER

Science & Technology Facilities Council

Research Complex at Harwell

The CCP4i2 Interface



Science & Technology Facilities Council

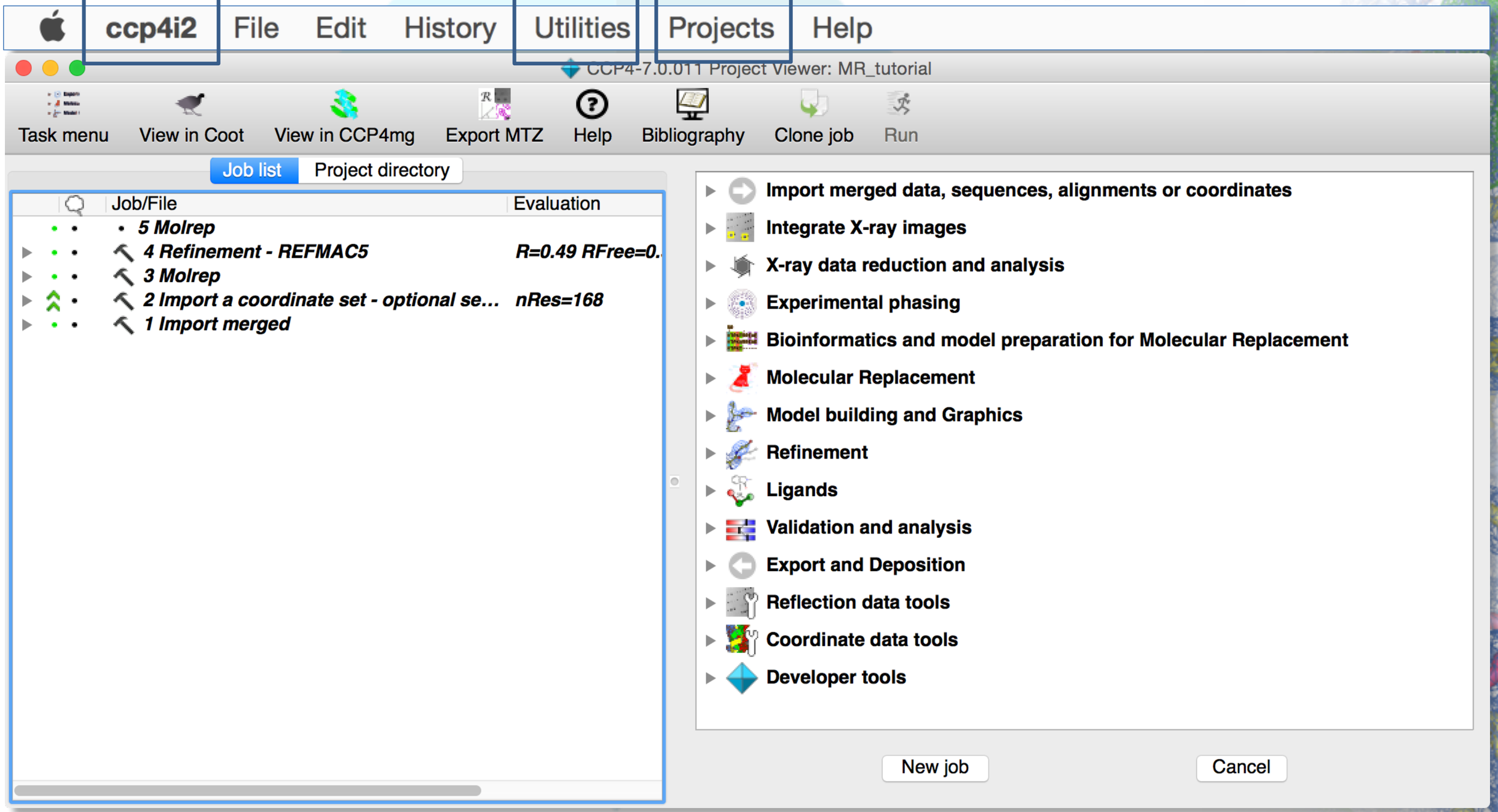
Research Complex at Harwell

The CCP4i2 Interface

Preferences and Settings

Managing Software Updates

Project Management

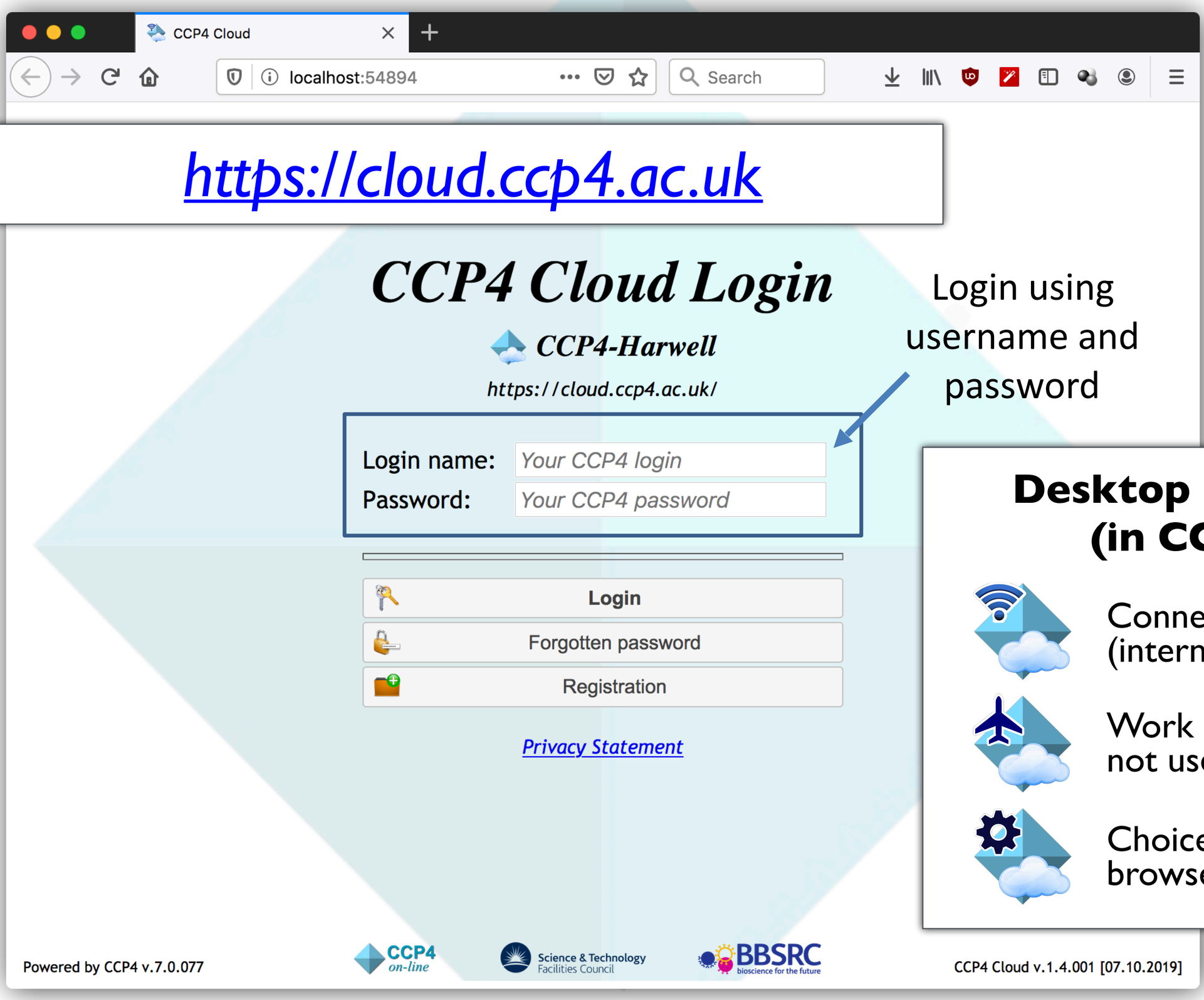


Science & Technology Facilities Council

Research Complex at Harwell

The CCP4 Cloud Interface

https://cloud.ccp4.ac.uk



Login using username and password

Desktop integration (in CCP4 7.1)

- Connect to remote server (internet required)
- Work locally (internet is not used)
- Choice of server and browser, and updates

Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Cloud: Projects

CCP4 Cloud

localhost:54894

Search

My Projects

20% 0%:0% Eugene Krissinel

Open | Add | Rename | Delete | Export | Import | Demo projects | Help

ID	Name	Disk (MBytes)	CPU (hours)	Date Created	Date Last Opened
05.ep-insulin	Insulin structure solution with EP on sulphur	215.8	1.9793	2019-04-25	2019-11-27
04.mr-enant	MR in enantiomorphic space groups with ligand fitting	169.4	0.4526	2019-04-25	2019-11-30
03.ccp4go	Fully automated solution with CCP4go pipeline	442.4	0.4665	2019-04-25	2019-11-22
02.auto-ep	Simple Auto-EP with Crank-2	154.9	0.6343	2019-04-25	2019-11-30
01.auto-mr	Simple Auto-MR with MORDA	15.2	0.2227	2019-04-24	2019-11-30
import-revision	How to continue projects started elsewhere	24.8	0.0149	2019-08-08	2019-11-22
gere-SAD	GerE: SAD and MR-SAD Phasing	84	0.3242	2019-05-02	2019-11-22
refinement-2	Twin Refinement - Heme Oxygenase	20.8	0.0226	2019-05-07	2019-05-13
gere-MAD	GerE: Auto-EP with Crank-2	268.6	0.8298	2019-05-05	2019-10-07
4gos	4GOS Demo	325.2	5.6599	2019-01-10	2019-11-30
gere	GerE	264.9	1.5733	2019-05-06	2019-11-21
tests	Tests	263.3	2.4375	2019-07-04	2019-11-12
rnase	RNase	212.4	149.6485	2019-07-04	2019-11-20
toxd	TOXD	28.9	0.1495	2019-07-08	2019-10-27
mdm2	MDM2 linked to Nutlin	1046	8.5762	2019-07-18	2019-11-09
insulin	Insulin phasing on Sulphur	254.2	1.7115	2019-07-30	2019-10-26
deposition	Deposition	17.3	0.0228	2019-08-06	2019-09-18

Powered by CCP4 v.7.0.077

CCP4 on-line | Science & Technology Facilities Council | BBSRC bioscience for the future

CCP4 Cloud v.1.4.001 [07.10.2019]

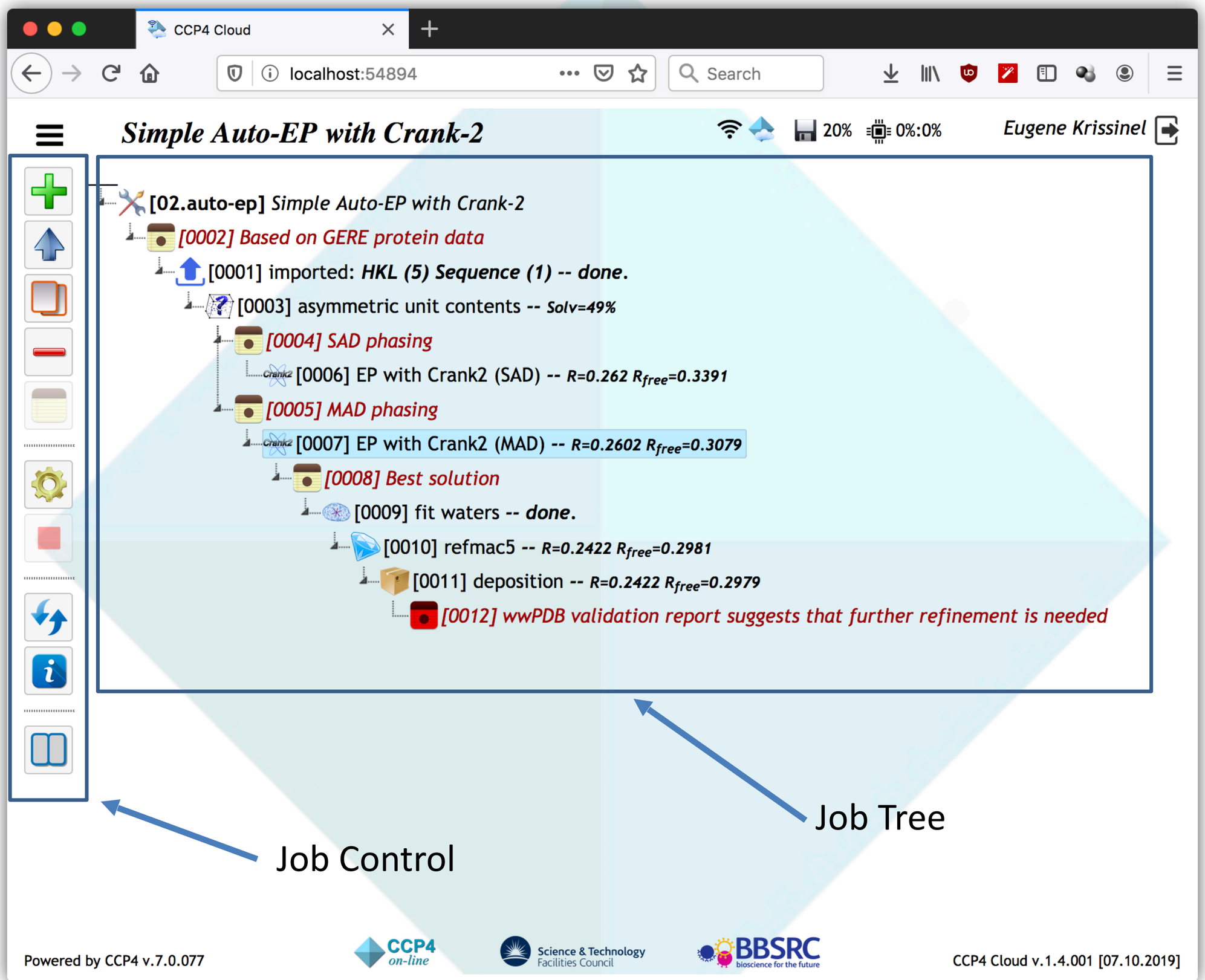
Project control

List of Projects

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Cloud: Jobs



Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Cloud: Reports

CCP4 Cloud

localhost:56718

Search

Simple Auto-EP with Crank-2

20% 0%:0% *Eugene Krissinel*

Job Report tabs – logs, summary, etc.

[0007] EP with Crank2 (MAD)

Input | Output | *Job completed* | Export | Ref. | Help | Close

Report | Main Log | Service Log | Errors

[0007] Crank-2 Automated Experimental Pl

Job Summary

Majority of the model should be successfully built!
Number of residues in the built model: 407
Final R factor: 0.2602
Final Rfree factor: 0.3079

Output:

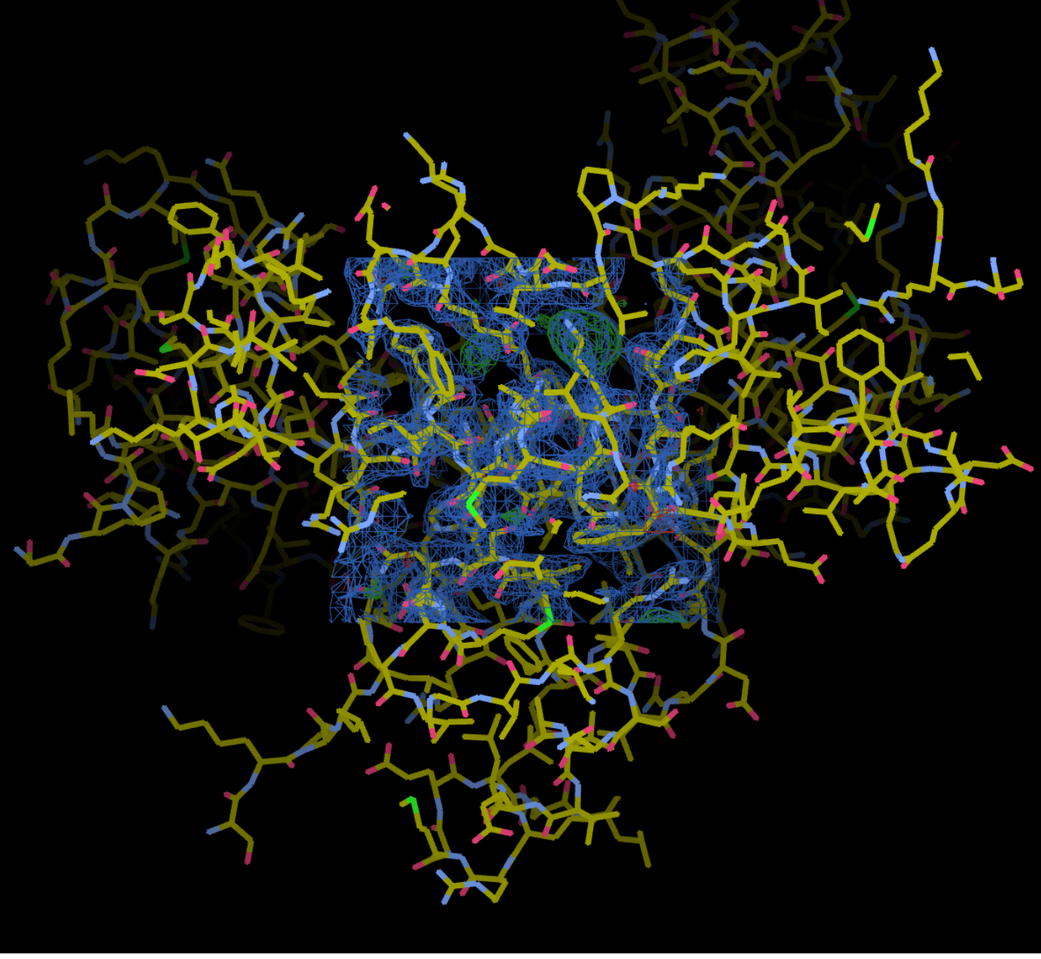
Built model and map | UglyMol | ccp4mg | ViewHKL

0007-01_crank2.pdb	Export
0007-01_crank2.mtz	Export
0007-01_crank2.mtz.map	Export
0007-01_crank2.mtz_diff.map	Export

Job output files

0007-01_crank2 <small>Built model and map</small>

Nothing here. Press H for help.



Powered by CCP4 v.7.0.077

CCP4 on-line

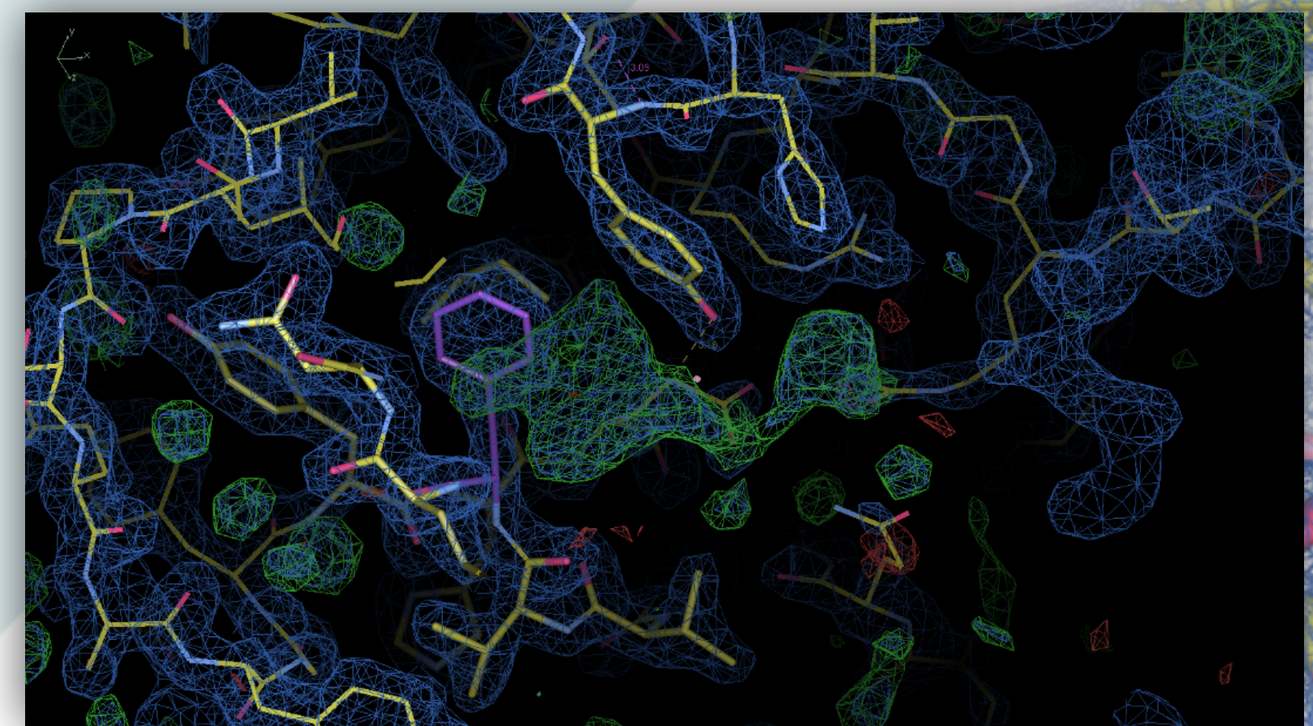
Science & Technology Facilities Council

Research Complex at Harwell

Structure solution with CCP4: a Roadmap

Structure Solution in Essence

- As an ultimate goal, one would like to generate an atomic model for their target
- This could be achieved by applying the inverse Fourier transform to experimental structure factors (which are Fourier transform of the structure's atomic coordinates) if they were known or could be indeed obtained in an experiment
- In reality, the X-ray diffraction experiment gives only moduli, ***but not phases***, of structure factors
- Therefore, phases need to be elicited by other means



Science & Technology Facilities Council
Research Complex at Harwell

Structure Solution Pathway

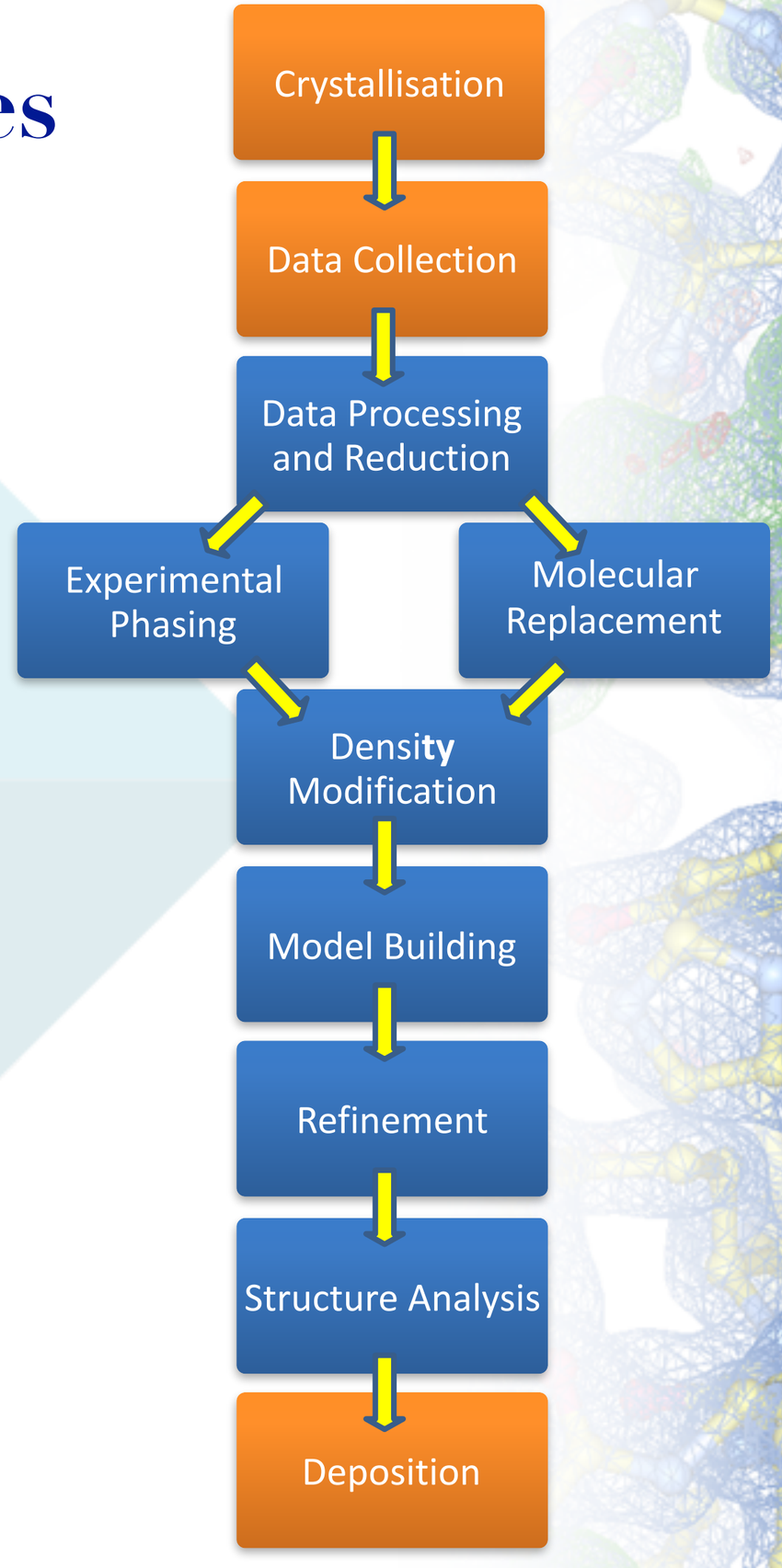
- As a first step, one need to derive structure factor amplitudes from the intensities of spots in X-ray images
- Next, phases need to be approximated by using methods such as
  - ~ Experimental Phasing: SAD, MAD, SIRAS etc.
  - ~ Molecular Replacement
  - ~ Combination of the above
- Then, structure factors can be converted into electron density via *inverse Fourier transform*
- The obtained electron density map needs to be interpreted via model building and refinement to produce a model for target structure
- The model needs to be analysed to answer biochemical questions and deposited in the PDB

sounds like we know what to do …

Science & Technology Facilities Council

Research Complex at Harwell

The Structure Solution Stages

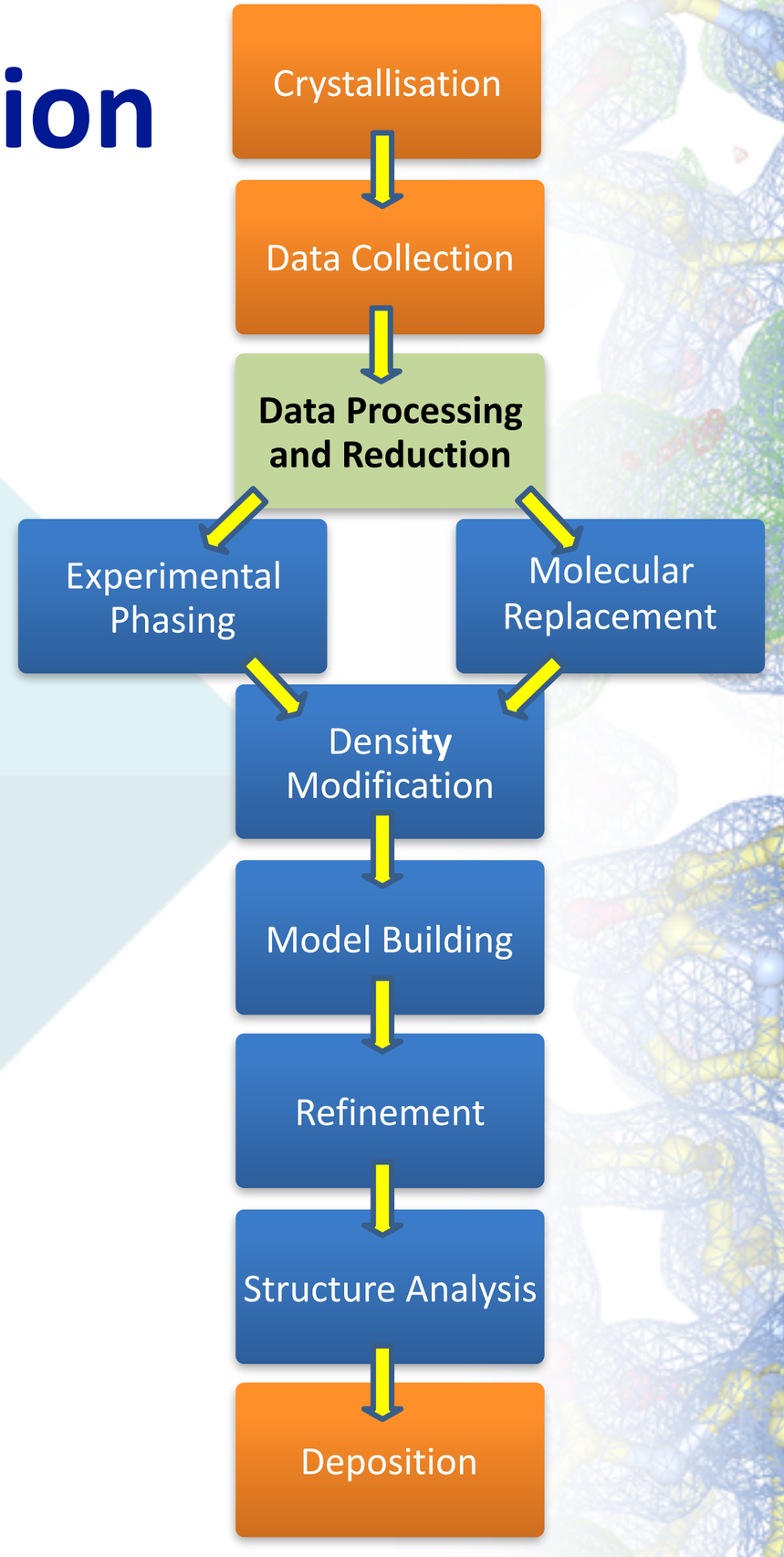


Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction

- Data Processing:
  - ~ Calculate intensities from images
  - ~ Initial estimation of space group and cell parameters (indexing)
  - ~ Look for problems in images
    - ✦ multiple lattices,
    - ✦ weak reflections
    - ✦ mosaicity
    - ✦ anisotropy



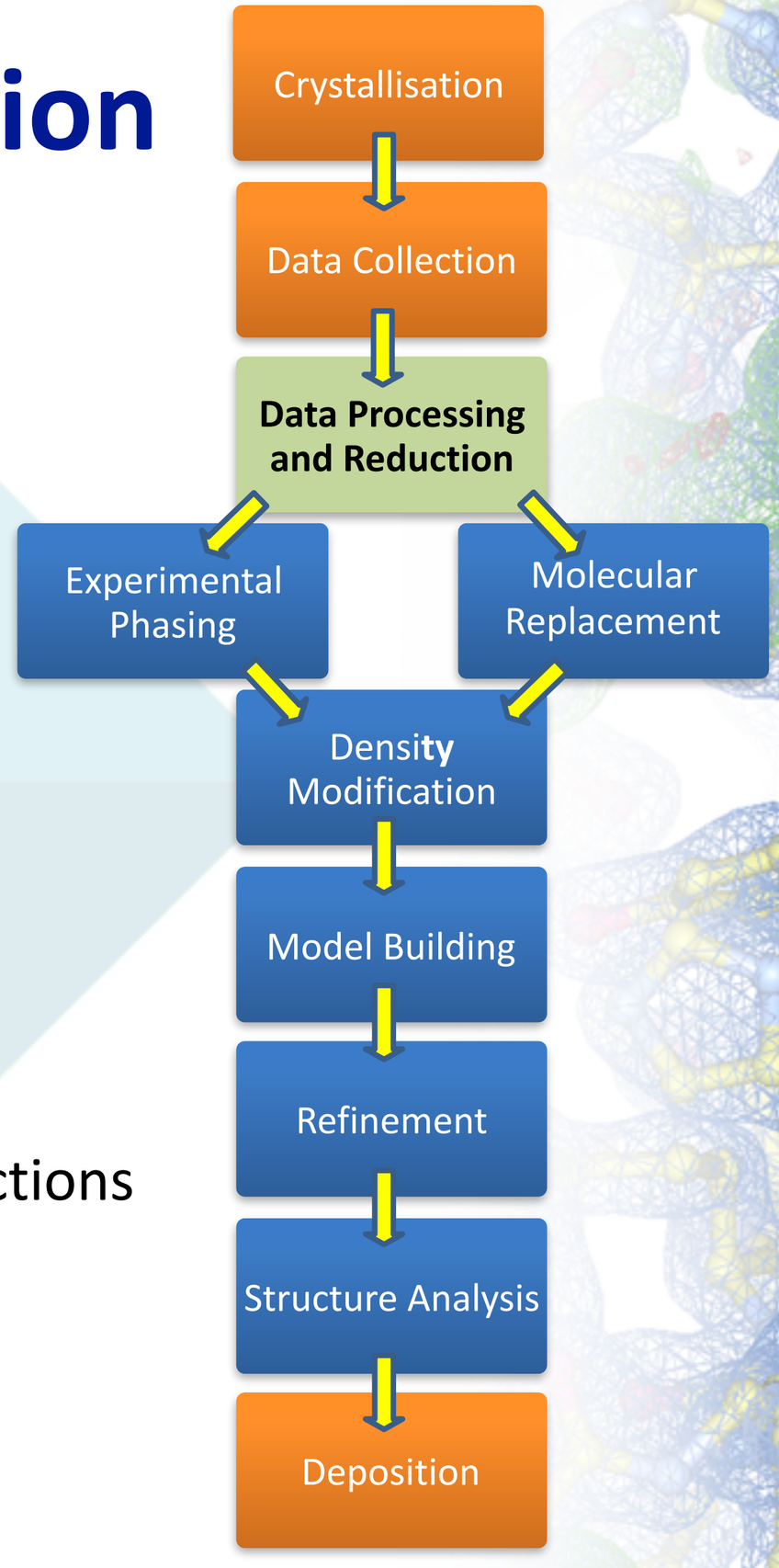
Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction

- Data Processing:
  - ~ Calculate intensities from images
  - ~ Initial estimation of space group and cell parameters (indexing)
  - ~ Look for problems in images
    - ✦ multiple lattices,
    - ✦ weak reflections
    - ✦ mosaicity
    - ✦ anisotropy

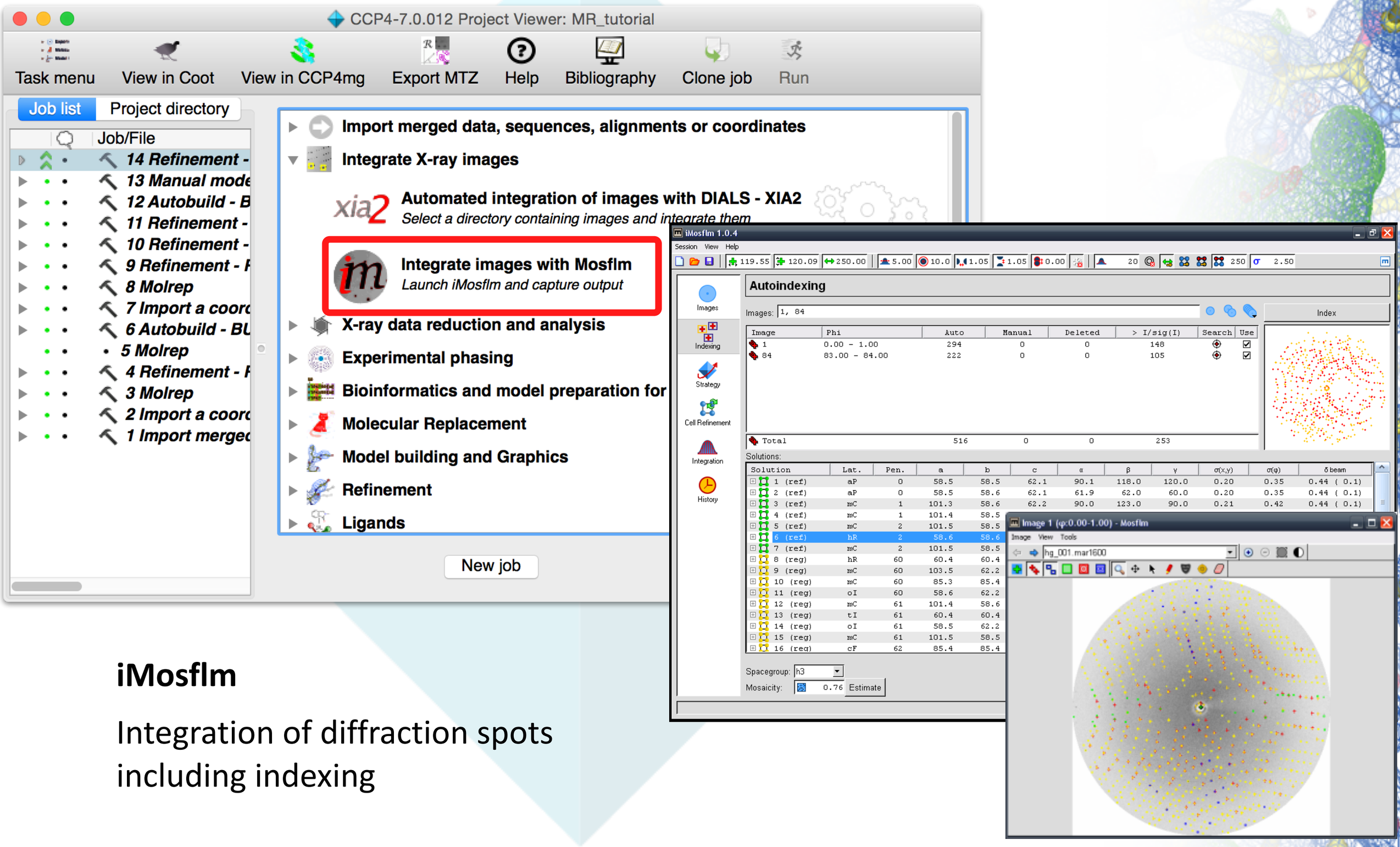
- Data Reduction: Scaling and merging
  - ~ Scale and merge multiple observations of reflections
  - ~ Assignment of point group/space group
  - ~ Look for problems in images
  - ~ Calculate structure factor amplitudes



Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction



iMosflm

Integration of diffraction spots including indexing

Data Processing and Reduction

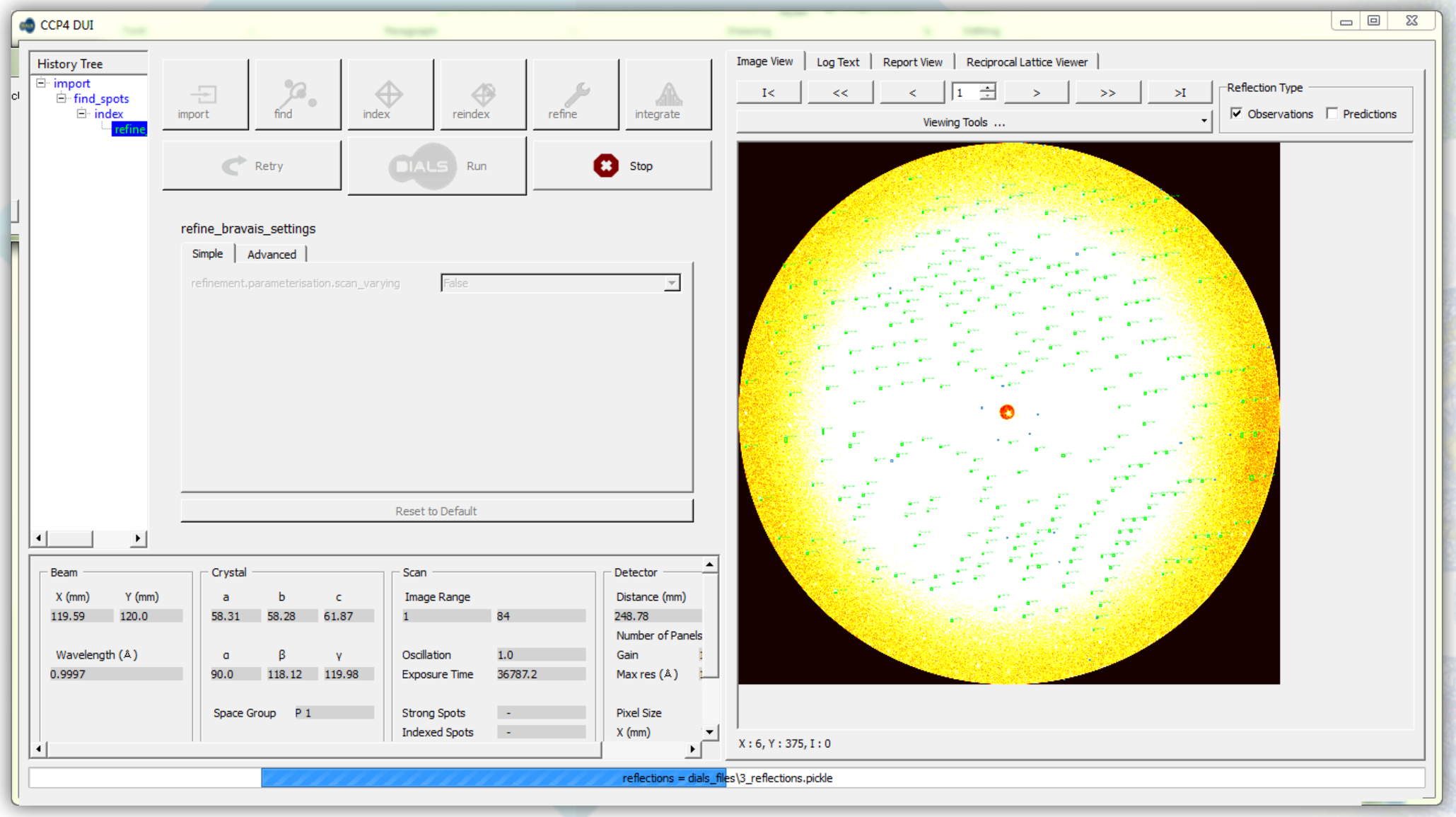
- ***D***iffraction ***I***ntegration for ***A***dvanced ***L***ight ***S***ources
- New data processing software designed for: speed, flexibility and accuracy, esp. with challenging data
- Will cater for both synchrotron and XFEL experiments
- Main philosophy: build a comprehensive toolkit. Thus implementing both 3D and 1D FFT indexing methods, 2D and 3D integration methods, multiple optimisation engines and allowing extension through an algorithm plug-in system
- In active development by CCP4, Diamond Ltd and Berkeley
- Can be run from command-line and under *Xiao*

Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction

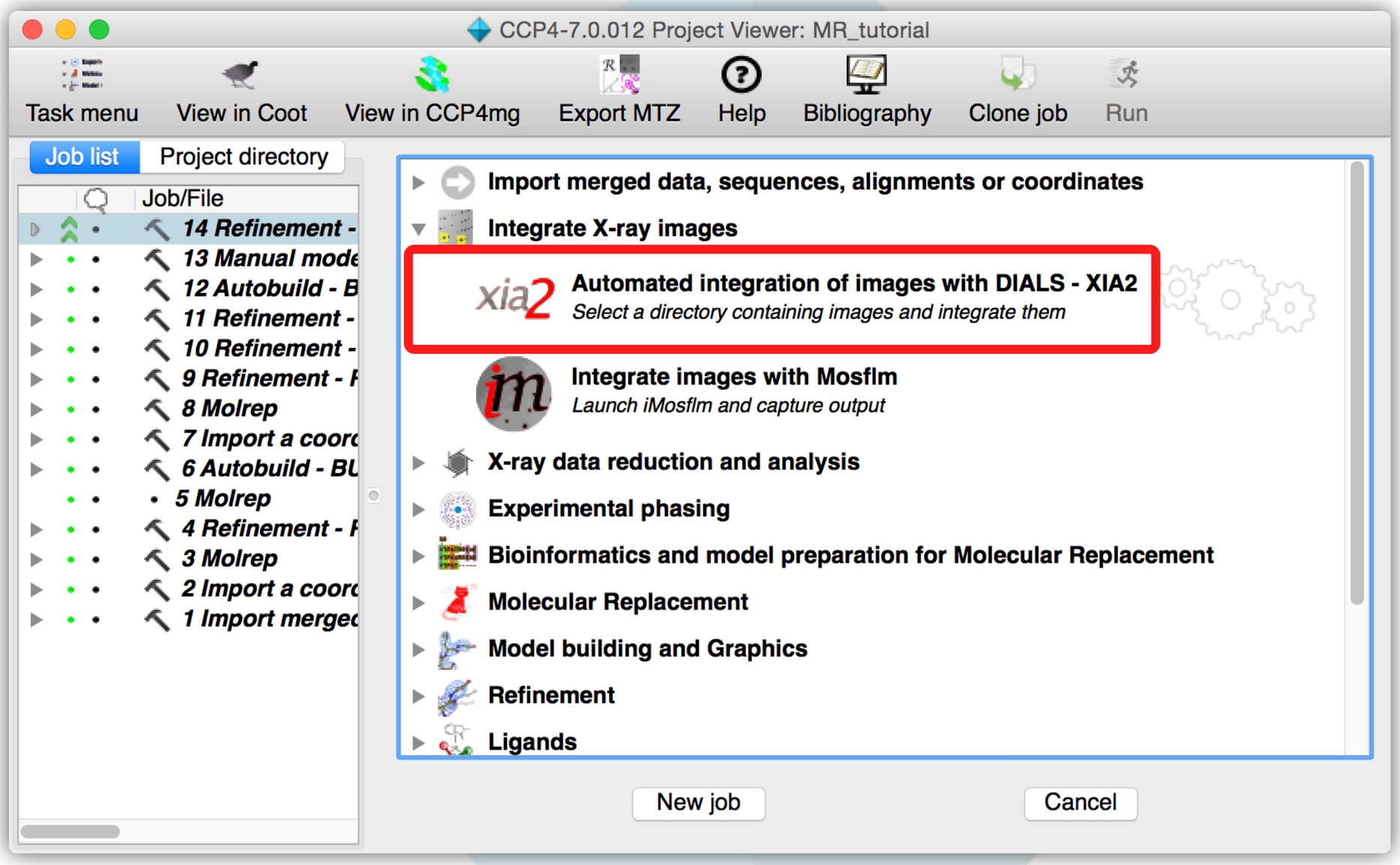
- ***D****iffraction* ***I****ntegration for* ***A****dvanced* ***L****ight* ***S****ources*
- Recently released: DUI the DIALS graphical interface



Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction



Xia-2: Expert knowledge-based, decision-making system for automated data processing from images to merged MTZ file. Uses DIALS and XDS.

Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction

Pointless

- Determines point-group (and space group)

Aimless

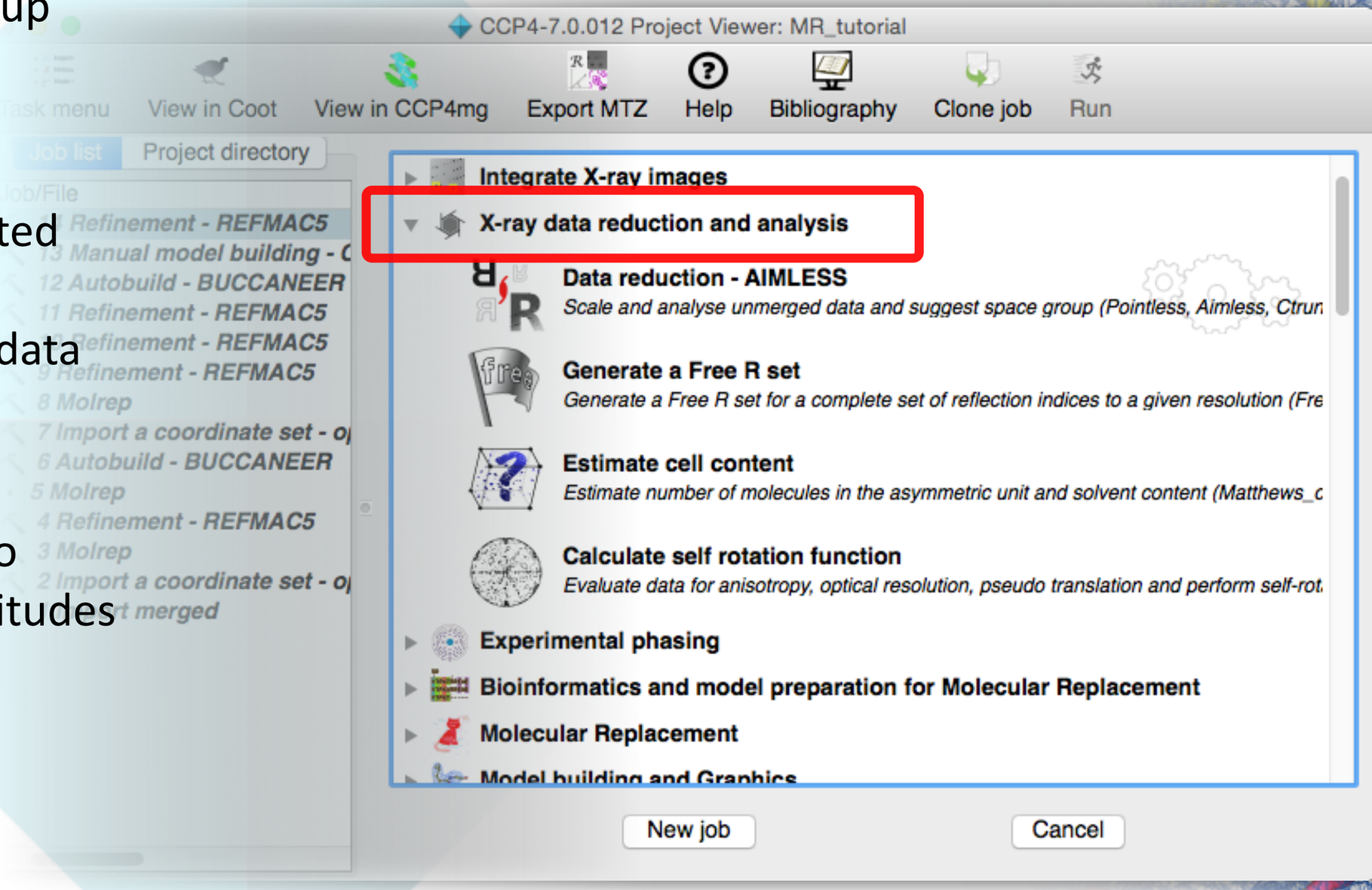
- Scales symmetry-related intensities together
- Produce statistics on data quality

Ctruncate

- Converts intensities to structure factor amplitudes
- Checks for twinning, anisotropy, pseudo symmetry

SFcheck

- Analyses data and check for problems – postscript report



Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction

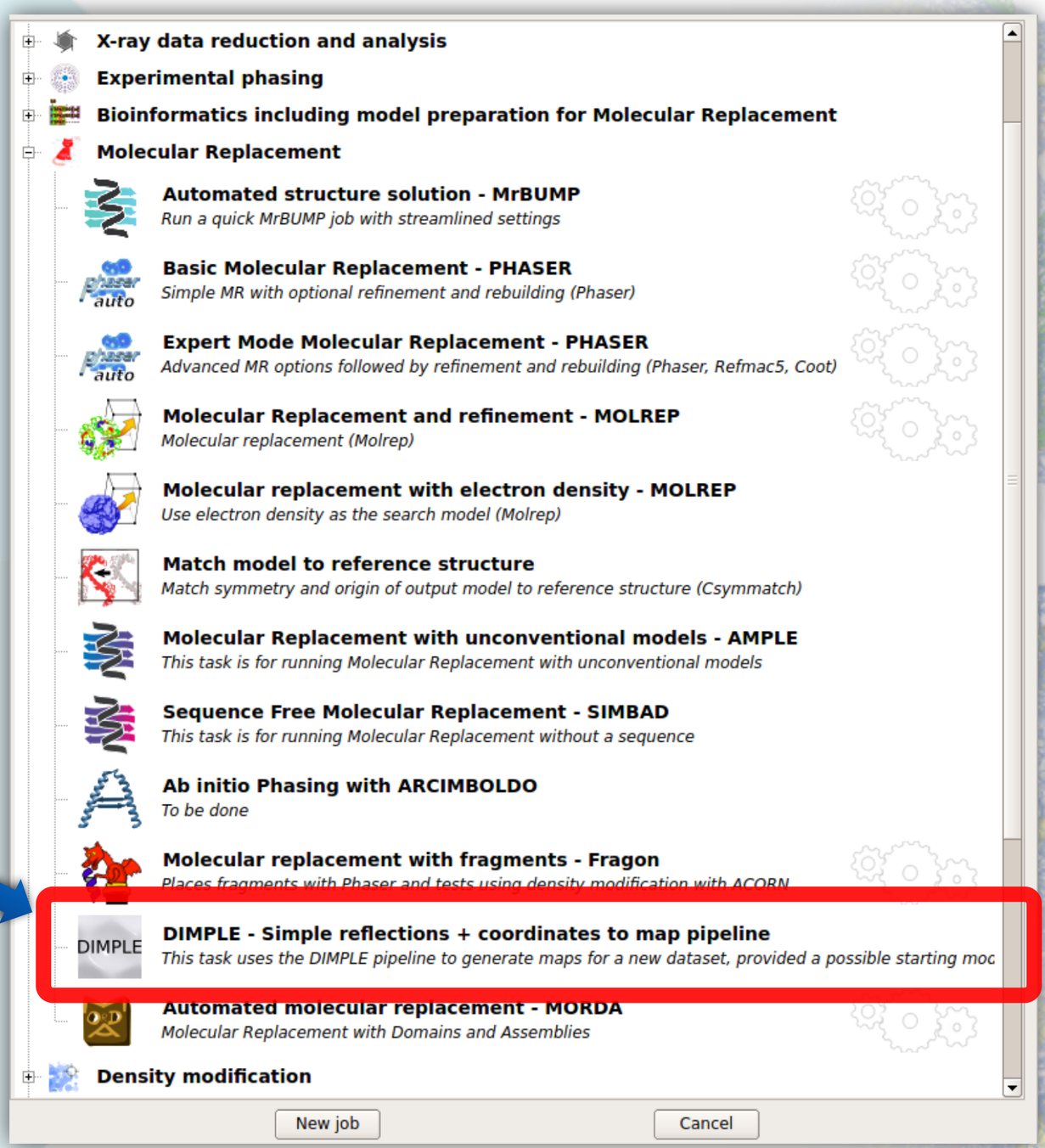
BLEND

Merging partial data sets from multiple crystals

Particularly useful where only small crystals are available

Dimple

On-beamline rapid calculation of difference maps to ascertain success or failure of ligand binding



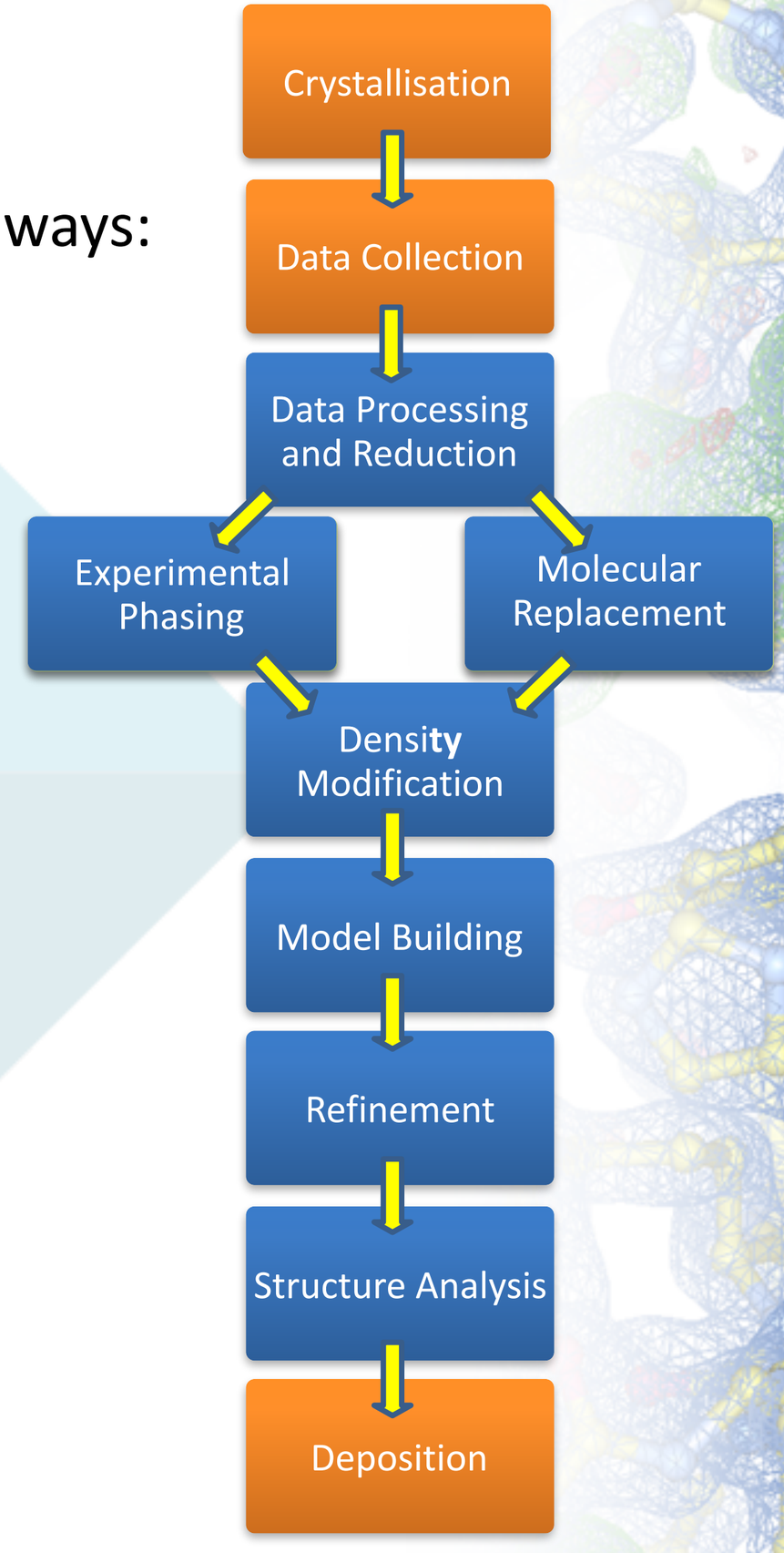
ccp4i2

Science & Technology Facilities Council

Research Complex at Harwell

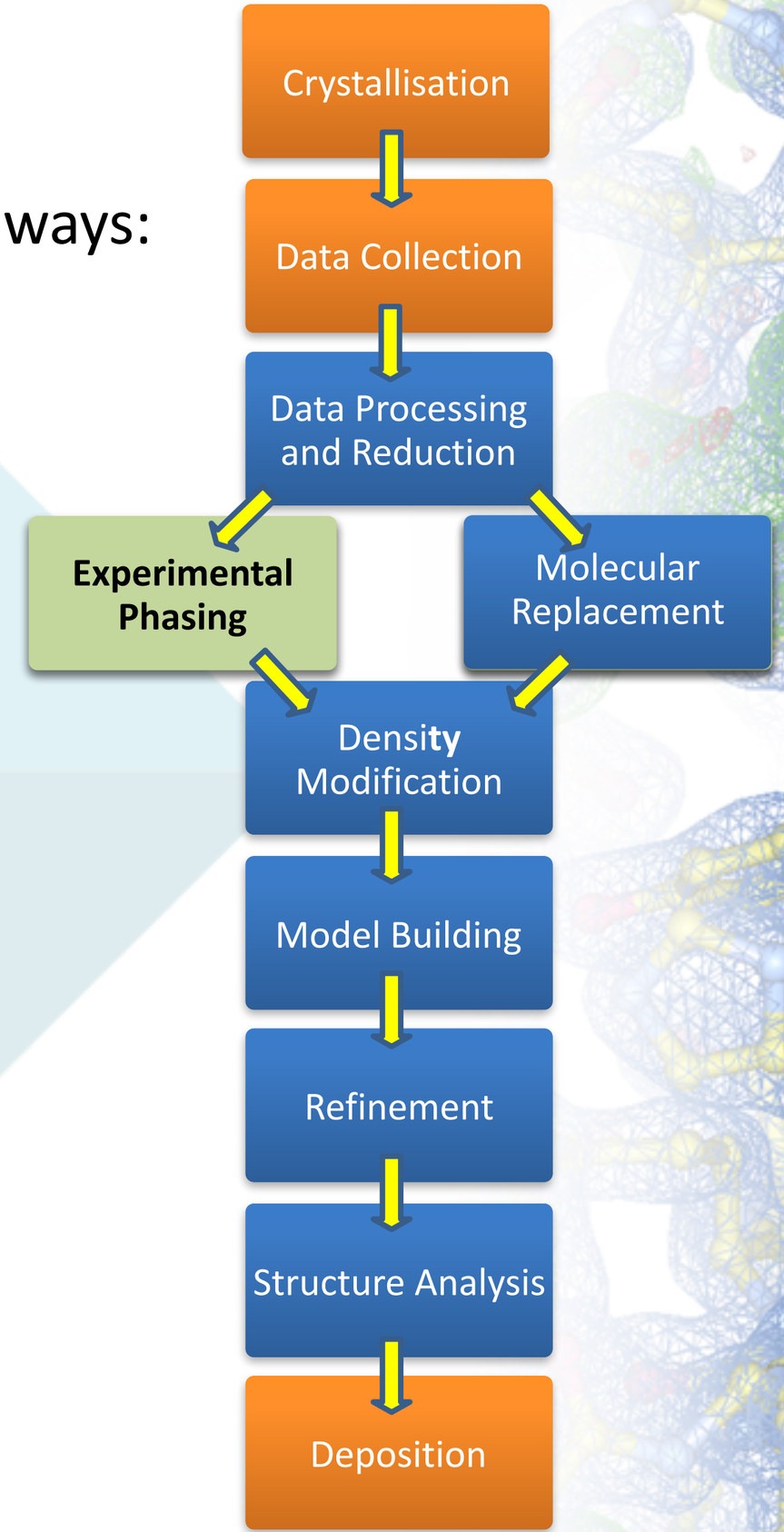
Phasing

- Phase information may be derived in two main ways:



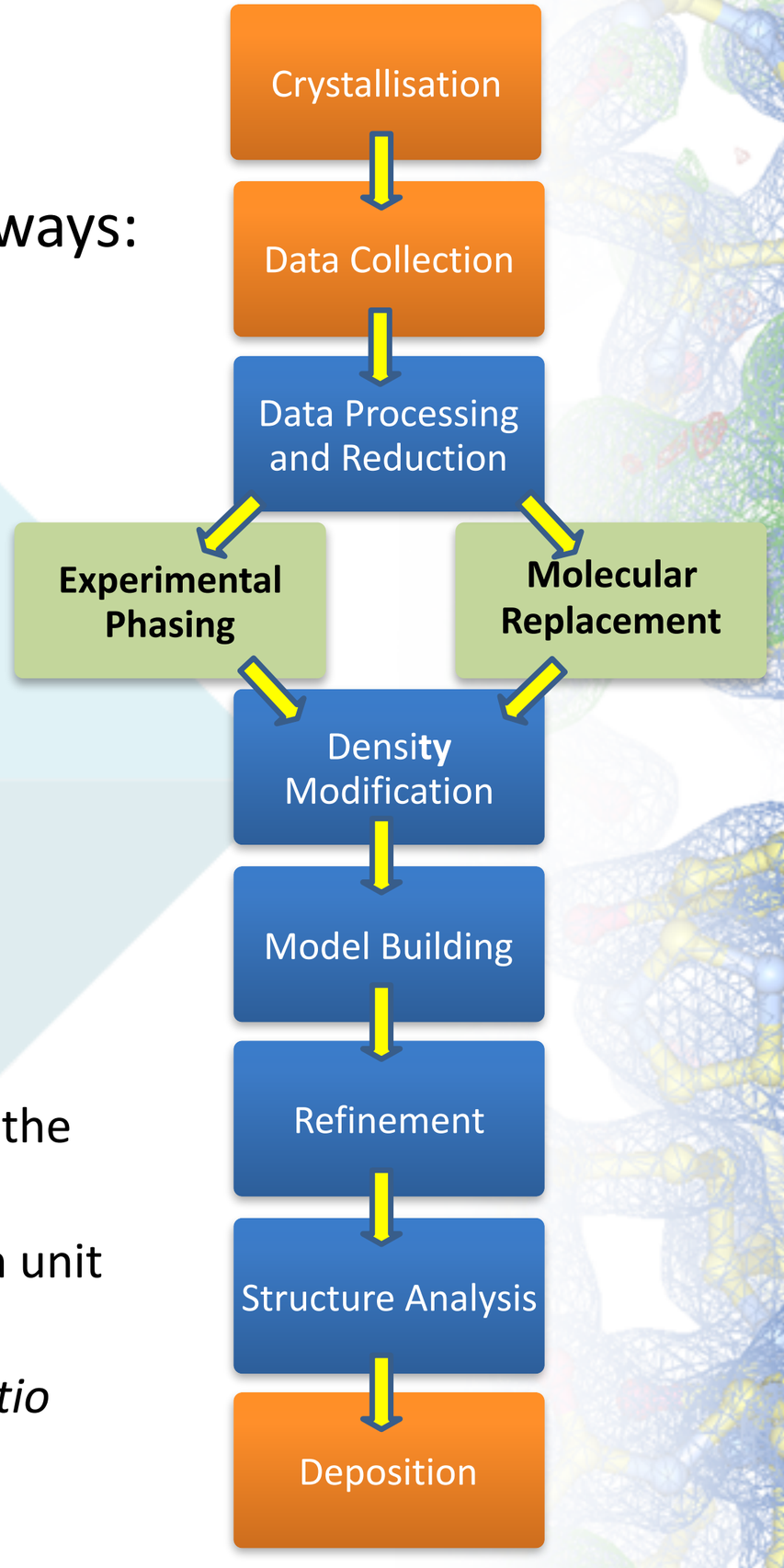
Phasing

- Phase information may be derived in two main ways:
- *Experimental phasing:*
  - collect multiple data sets with anomalous or isomorphous differences or differences produced by radiation damage
  - Determine the differences in atomic structure which lead to the differences in observations — usually a heavy atom substructure
  - Refine the parameters of the substructure and use the observations and the substructure to estimate phases



Phasing

- Phase information may be derived in two main ways:
- *Experimental phasing:*
  - collect multiple data sets with anomalous or isomorphous differences or differences produced by radiation damage
  - Determine the differences in atomic structure which lead to the differences in observations — usually a heavy atom substructure
  - Refine the parameters of the substructure and use the observations and the substructure to estimate phases
- *Molecular Replacement:*
  - Use the structure of an already solved, homologous macromolecule as a first approximation for the phases of the target
  - Requires correct repositioning of homologous structure in unit cell of target
  - Can also generate homologue from sequence using *ab initio* modelling



Science & Technology Facilities Council

Research Complex at Harwell

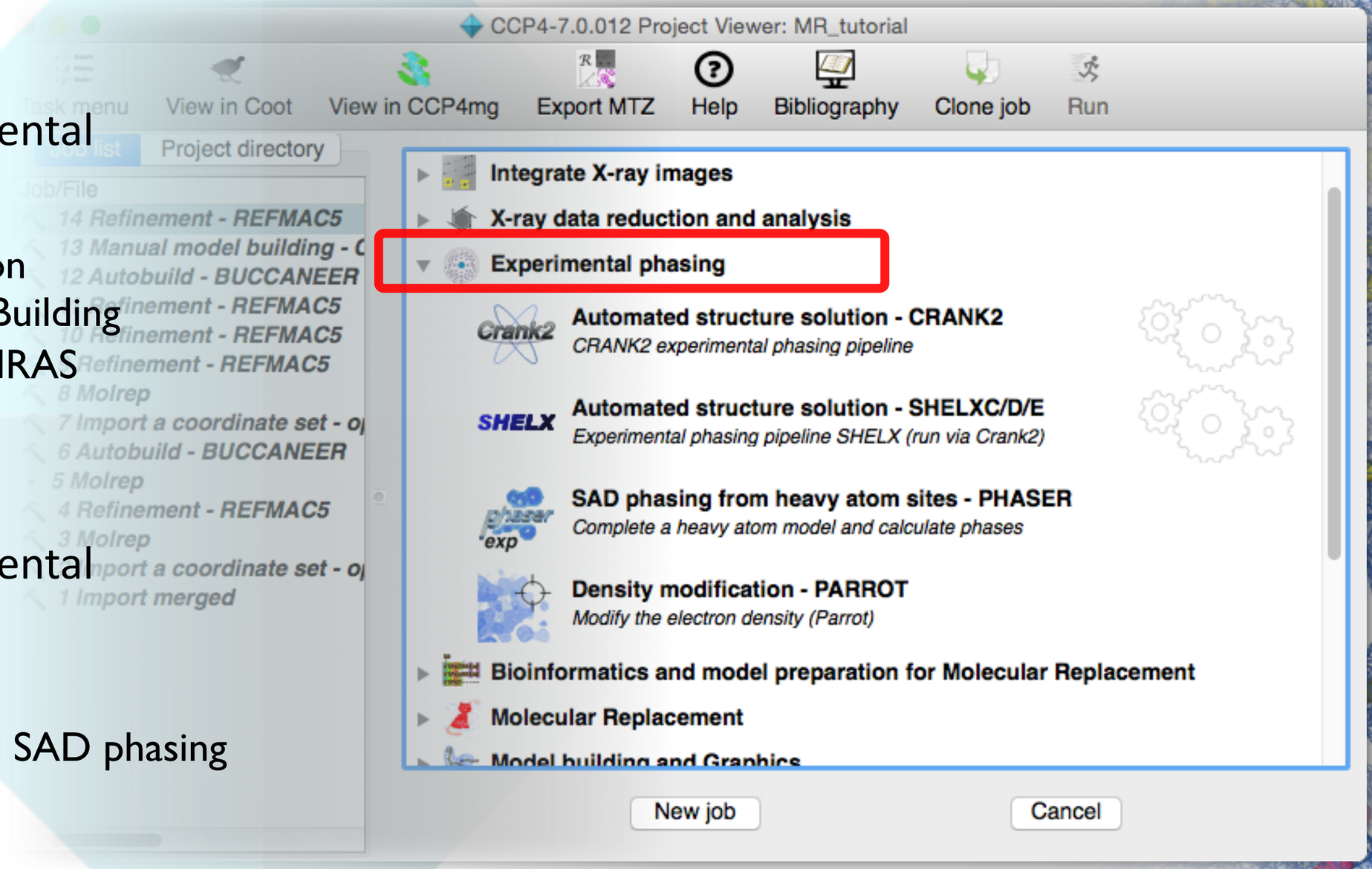
Experimental Phasing

Crank2 & SHELX

- Automated experimental phasing pipelines
  - ~ Heavy atom location through to Model Building
  - ~ SAD(-MR), MAD, SIRAS

Phaser EP

- Automated experimental phasing
  - ~ SAD phasing
  - ~ Combined MR and SAD phasing

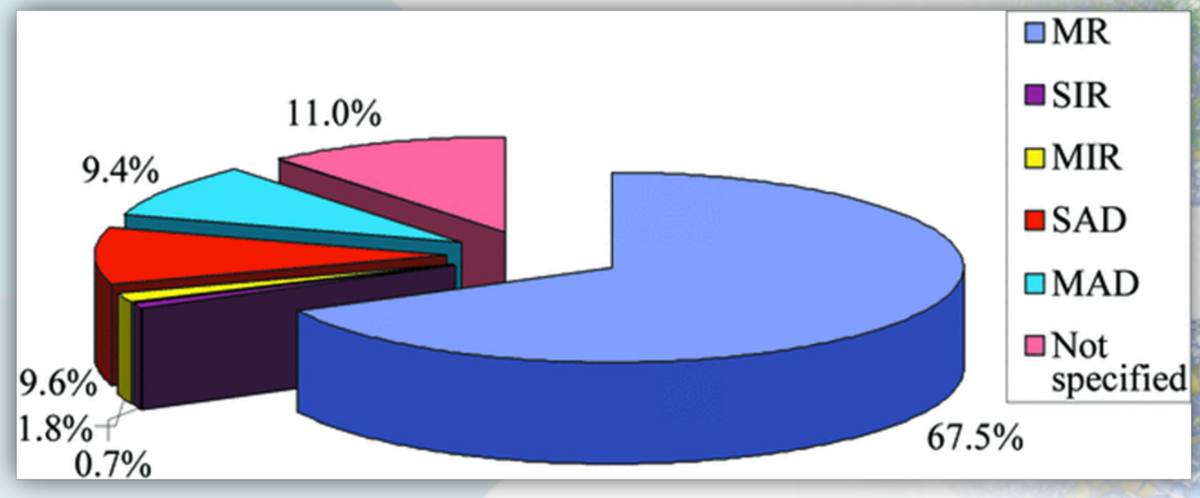


Science & Technology Facilities Council

Research Complex at Harwell

Molecular Replacement

- Today, the vast majority of entries in the PDB have been phased using MR
- Increasing success rate of MR has been due to:
  - Improving methods in software e.g. maximum likelihood in Phaser
  - Increasing availability of suitable search models



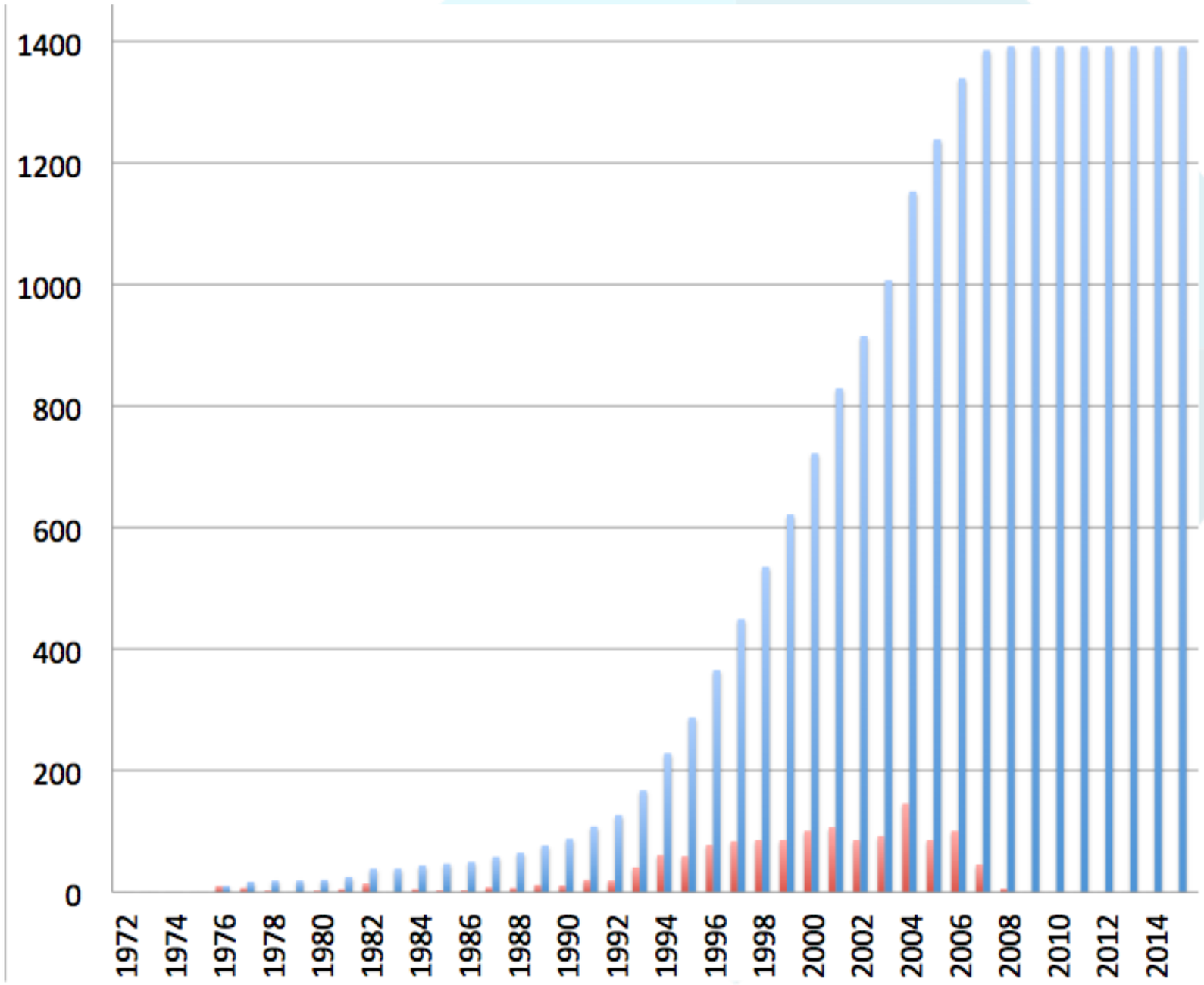
(image from Fei Long)

Science & Technology Facilities Council

Research Complex at Harwell

The Protein Data Bank (PDB)

- Currently over 140,000 structures (125,000 X-ray structures)
- For an X-ray crystallographer, a treasury of models for MR



Chothia C: *One thousand families for the molecular biologist.* **Nature** 1992, **357**: 543-544.

- MR is the dominating phasing method in today's protein crystallography

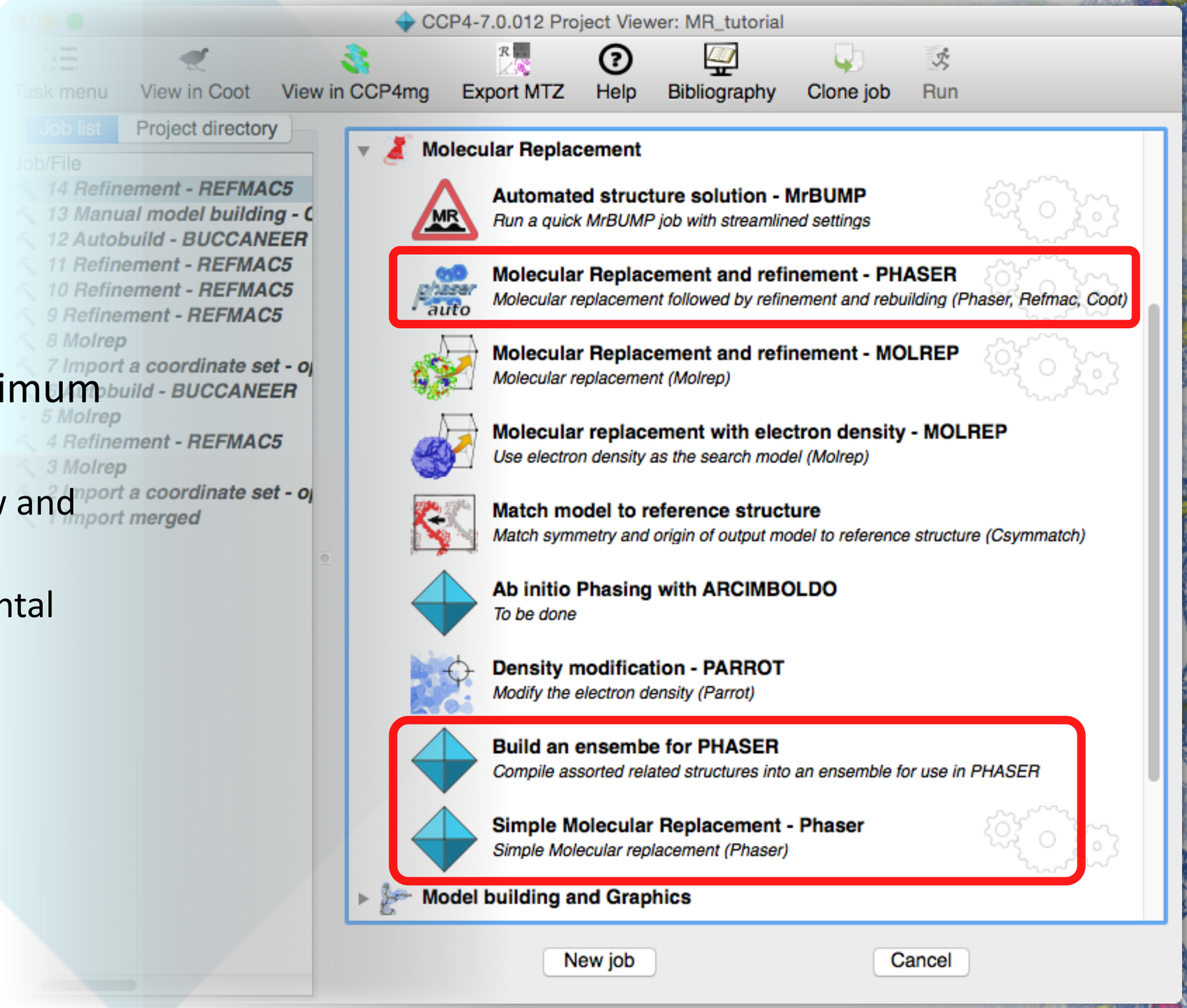
Science & Technology Facilities Council

Research Complex at Harwell

Molecular Replacement

Phaser MR

- Automated molecular replacement using Maximum Likelihood method
  - ~ Sophisticated workflow and decision making
  - ~ Accounts for experimental and model errors



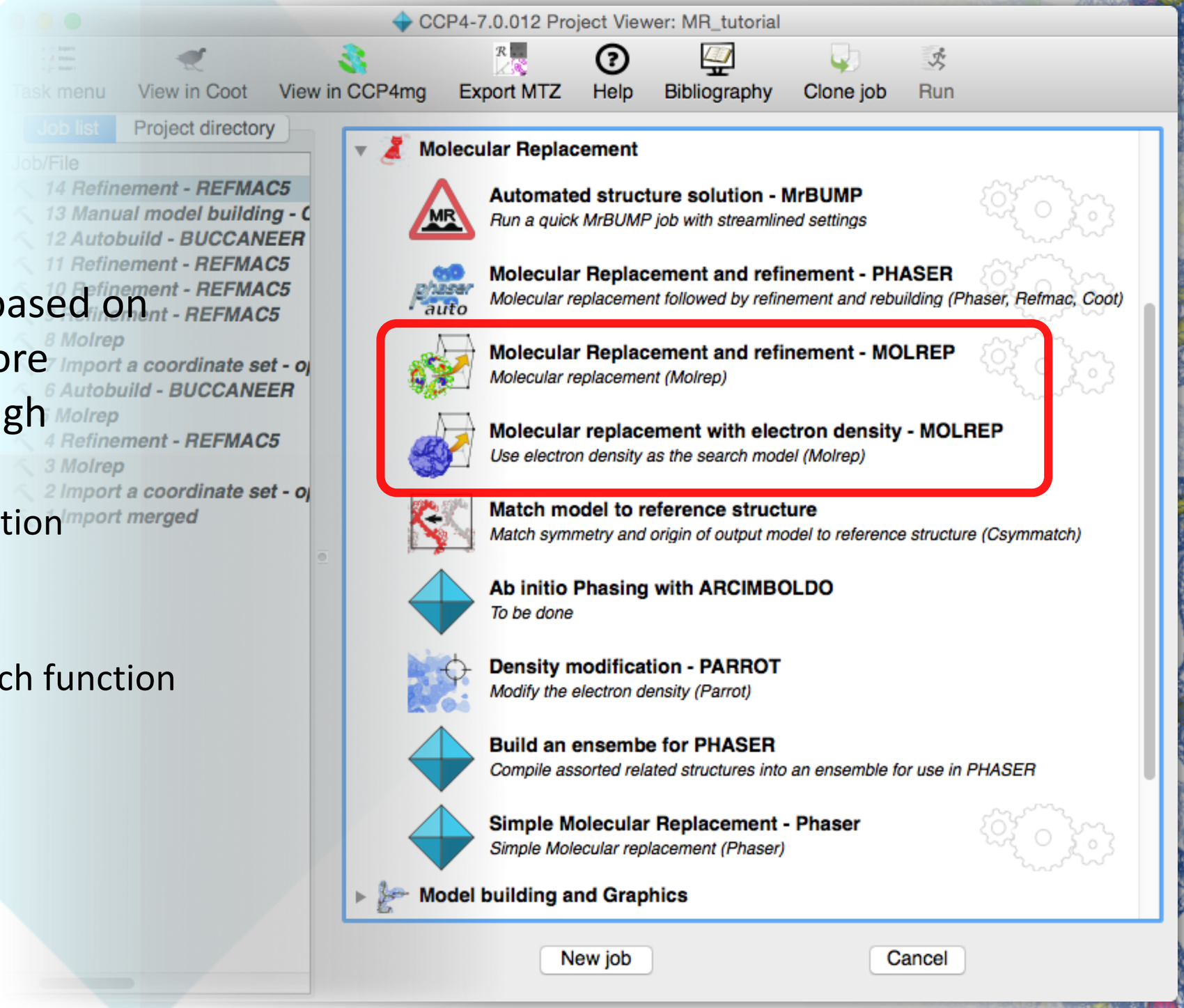
Science & Technology Facilities Council

Research Complex at Harwell

Molecular Replacement

Molrep

- Molecular replacement based on Patterson search with more manual interaction through specialized functions
  - ~ Phased Translation/Rotation functions
  - ~ Multi-copy search
  - ~ Spherical averaging search function



Science & Technology Facilities Council

Research Complex at Harwell

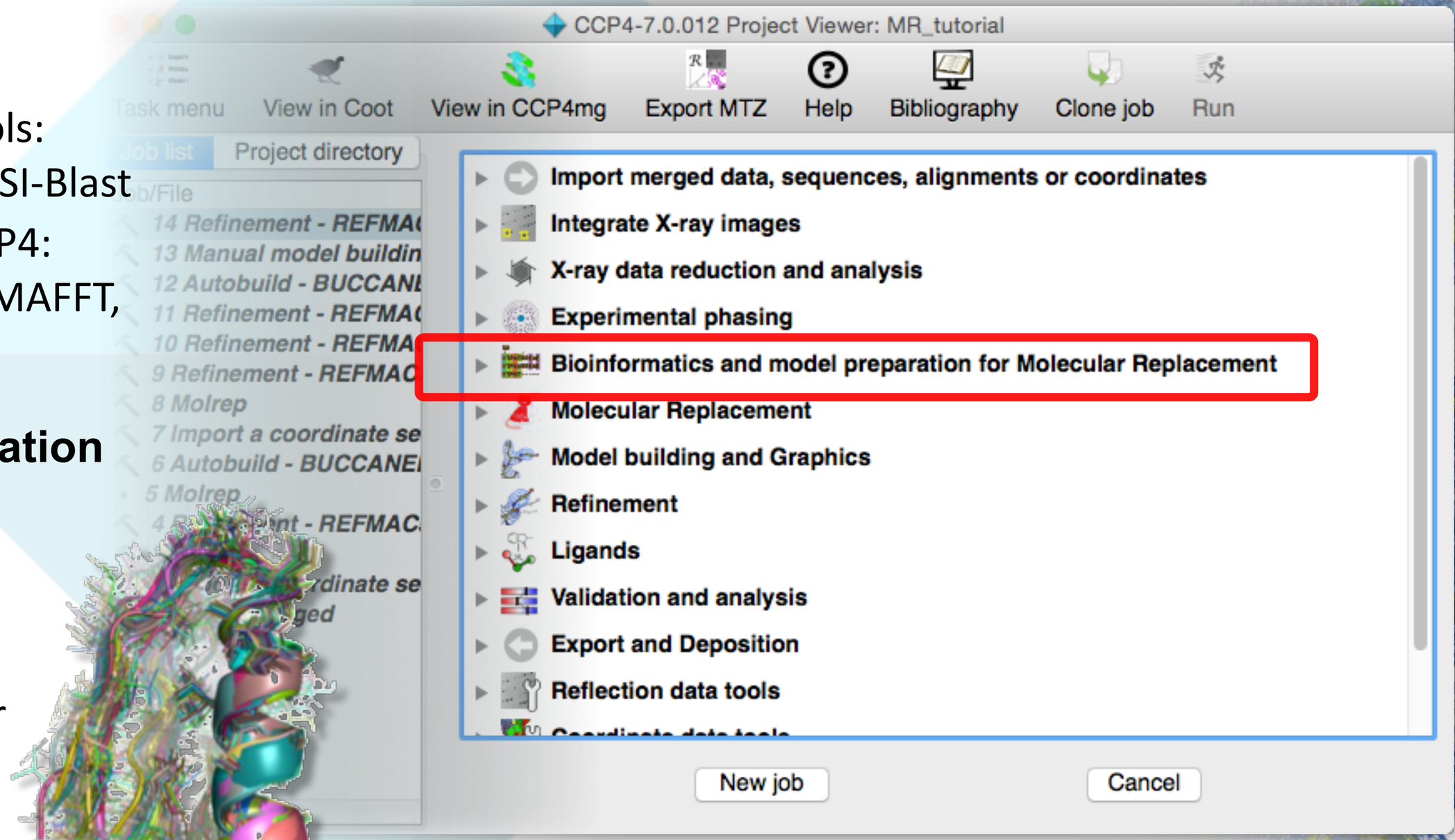
MR Search model preparation

Model search

- Online tools: HHpred, PSI-Blast
- Within CCP4: Phmmer, MAFFT, Clustalw

Model preparation

- Molrep
- Chainsaw
- Sculptor
- Ensembler



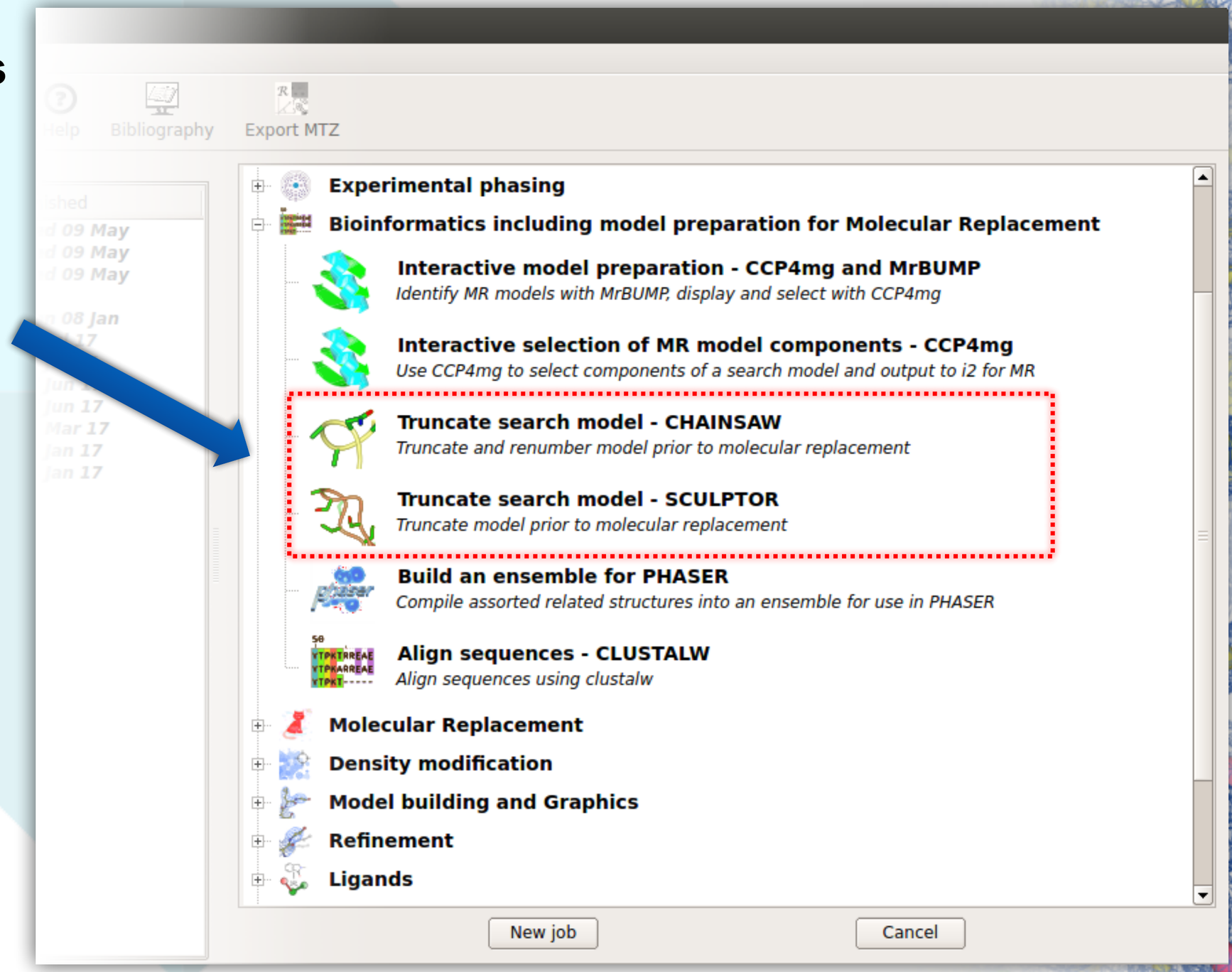
Science & Technology Facilities Council

Research Complex at Harwell

MR Search model preparation

Model preparation tools

- **Chainsaw** & **Sculptor**
  - both take input sequence alignment and model and truncate accordingly



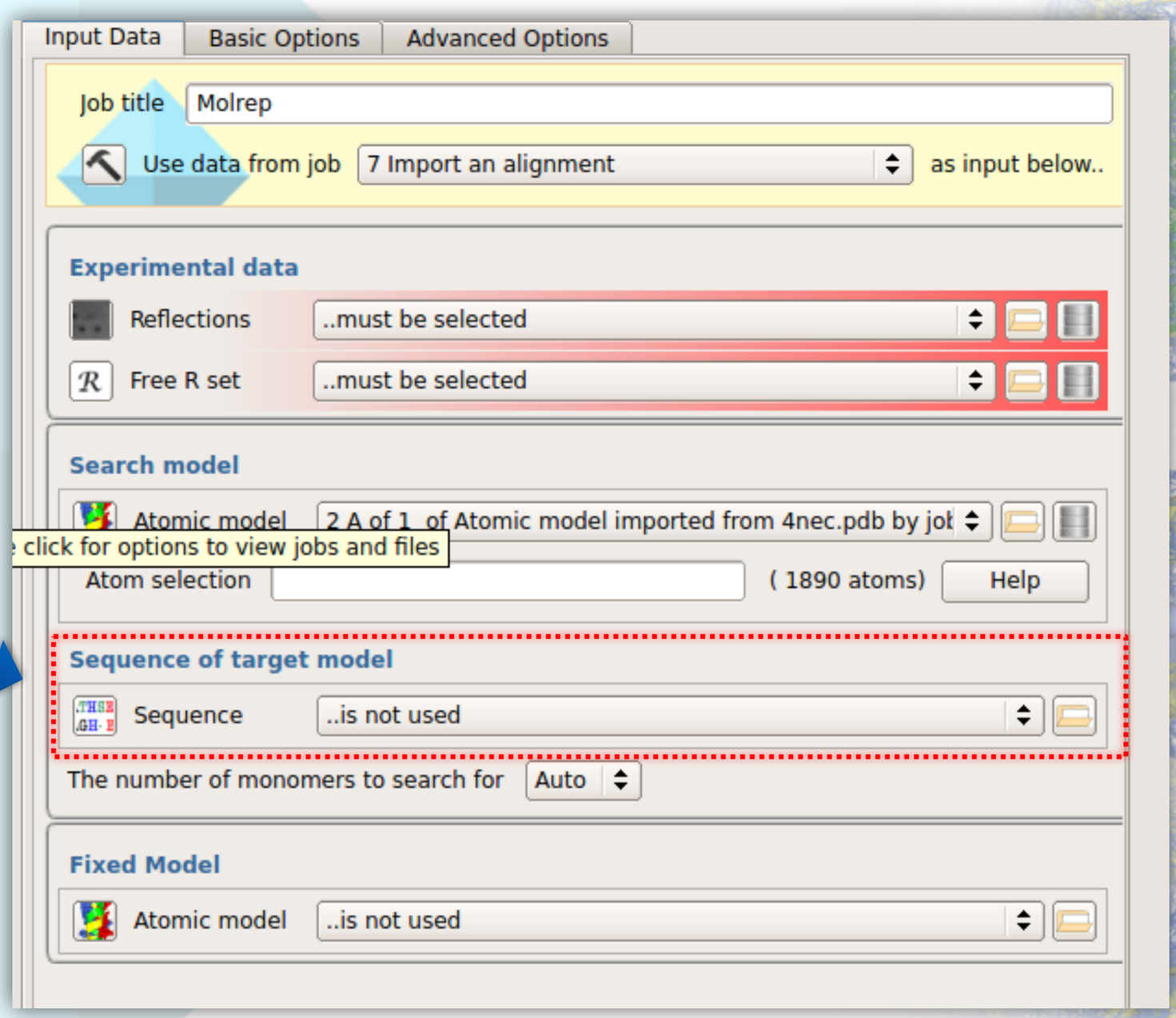
Science & Technology Facilities Council

Research Complex at Harwell

MR Search model preparation

Model preparation tools

- **Chainsaw** & **Sculptor**
  - both take input sequence alignment and model and truncate accordingly
- **Molrep**
  - provide sequence and Molrep will align and truncate sequence of search model



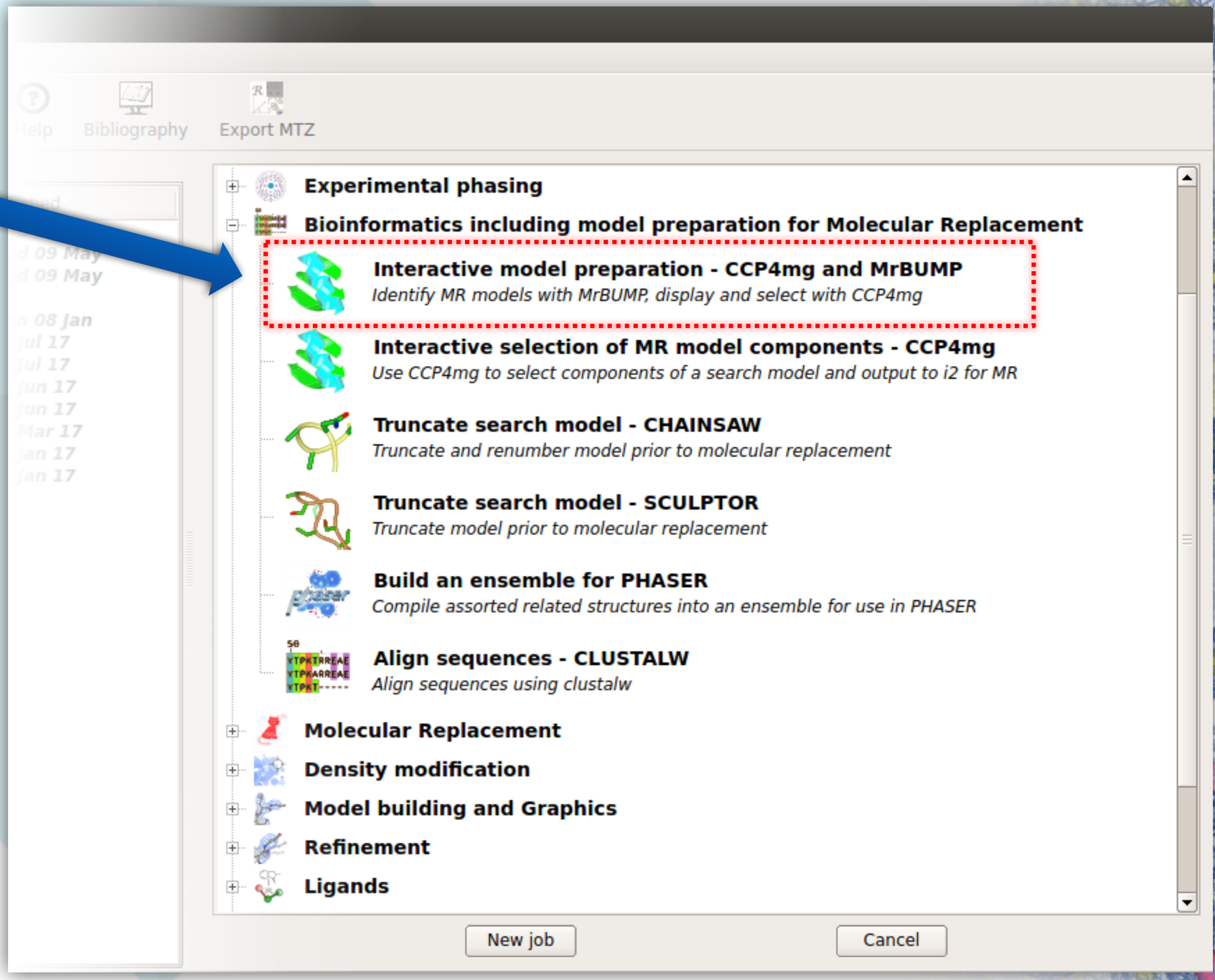
Science & Technology Facilities Council

Research Complex at Harwell

MR Search model preparation

Model preparation tools

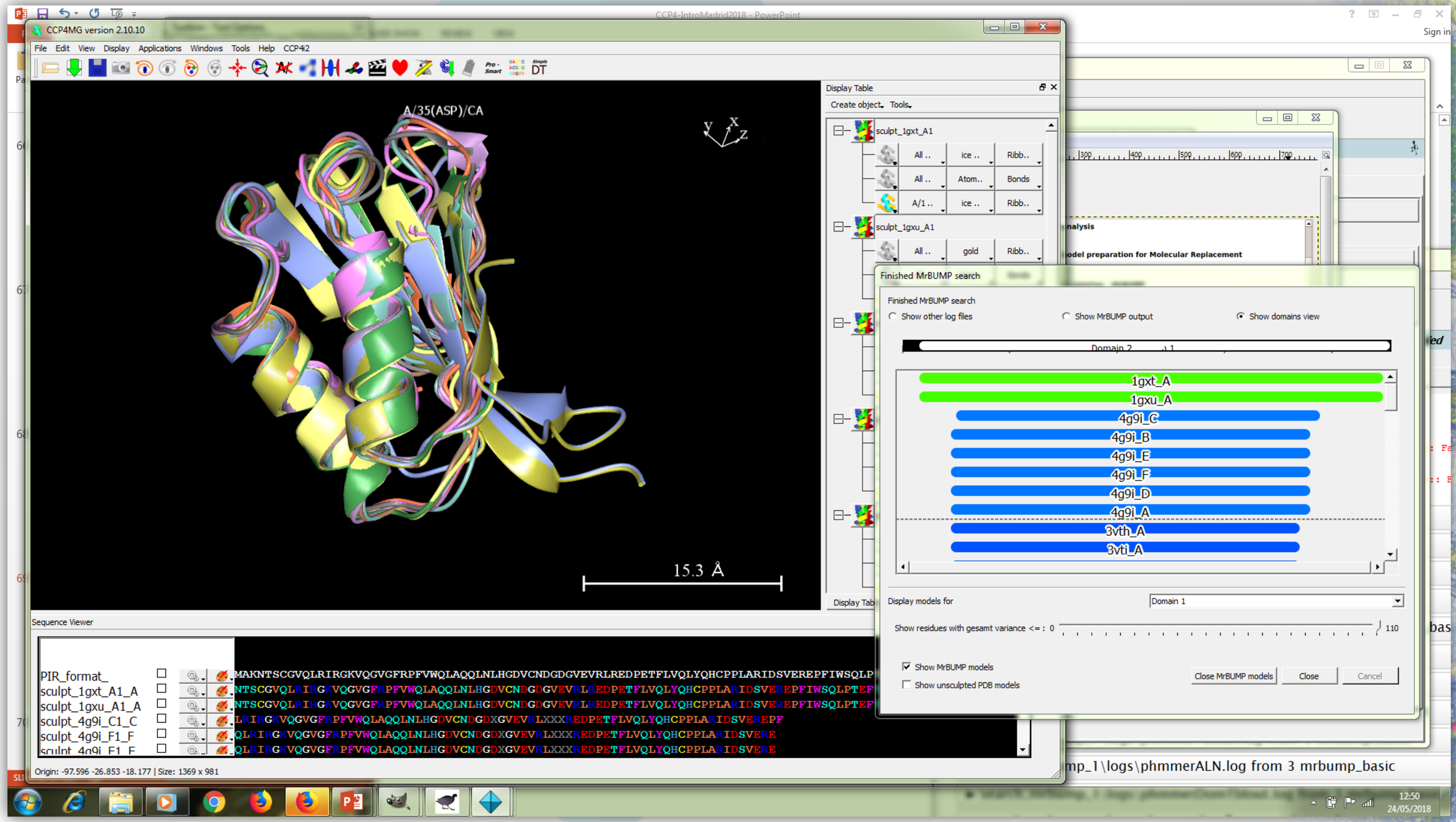
- **CCP4mg** – graphical aided model search and preparation using MrBUMP
- Can be used to create ensemble search model give availability of homologues



Science & Technology Facilities Council

Research Complex at Harwell

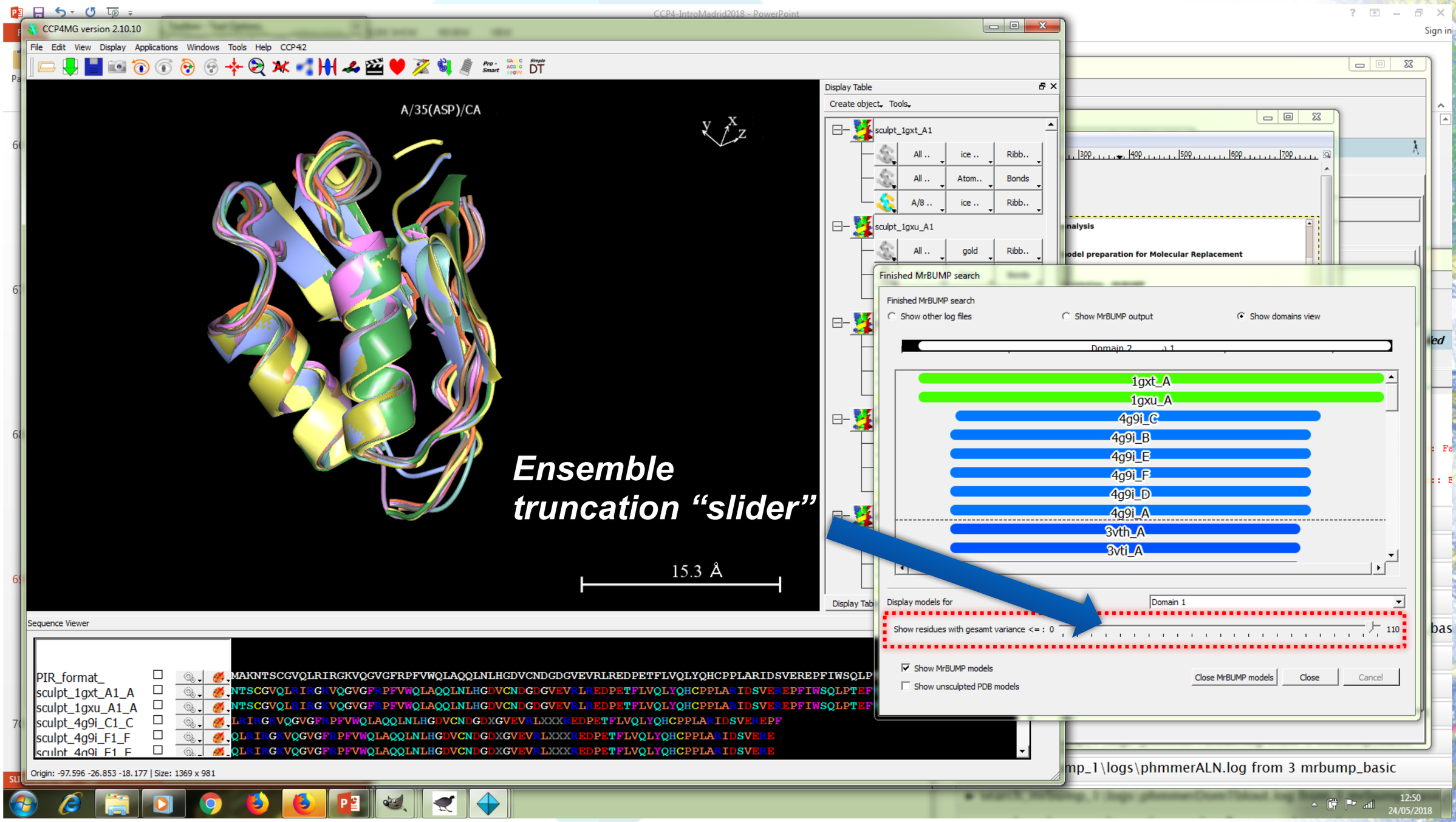
MR Search model preparation



Science & Technology Facilities Council

Research Complex at Harwell

MR Search model preparation



Science & Technology Facilities Council

Research Complex at Harwell

Automatic Molecular Replacement

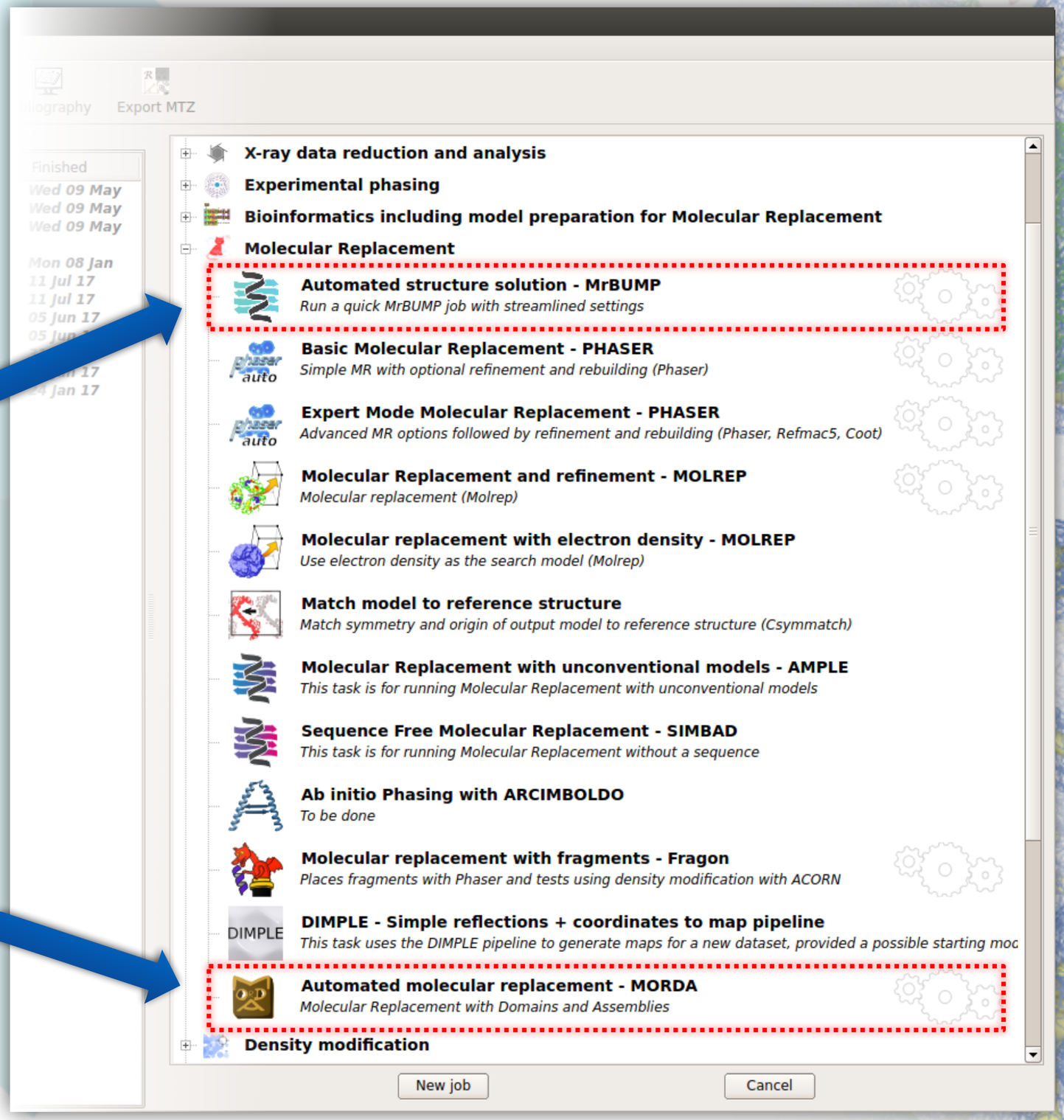
From model search and preparation through to initial refinement and model building

MrBUMP

- Sequence search
- Multiple model preparation
- Uses Phaser and Molrep
- Refinement & model building

MoRDa

- Custom database with search models
- Capable of making multimeric models (uses PISA)
- Uses Molrep and Refmac



Science & Technology Facilities Council

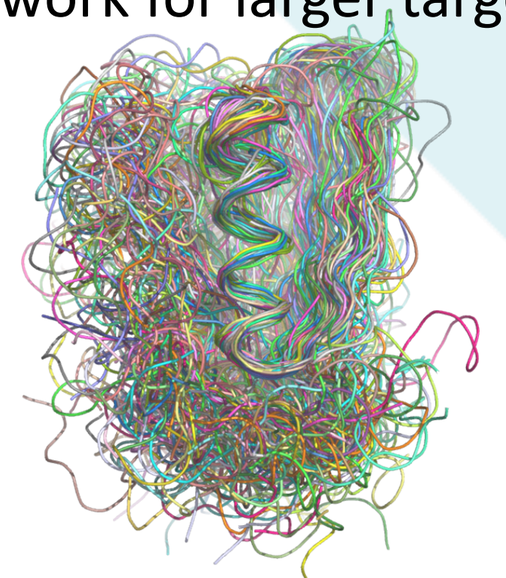
Research Complex at Harwell

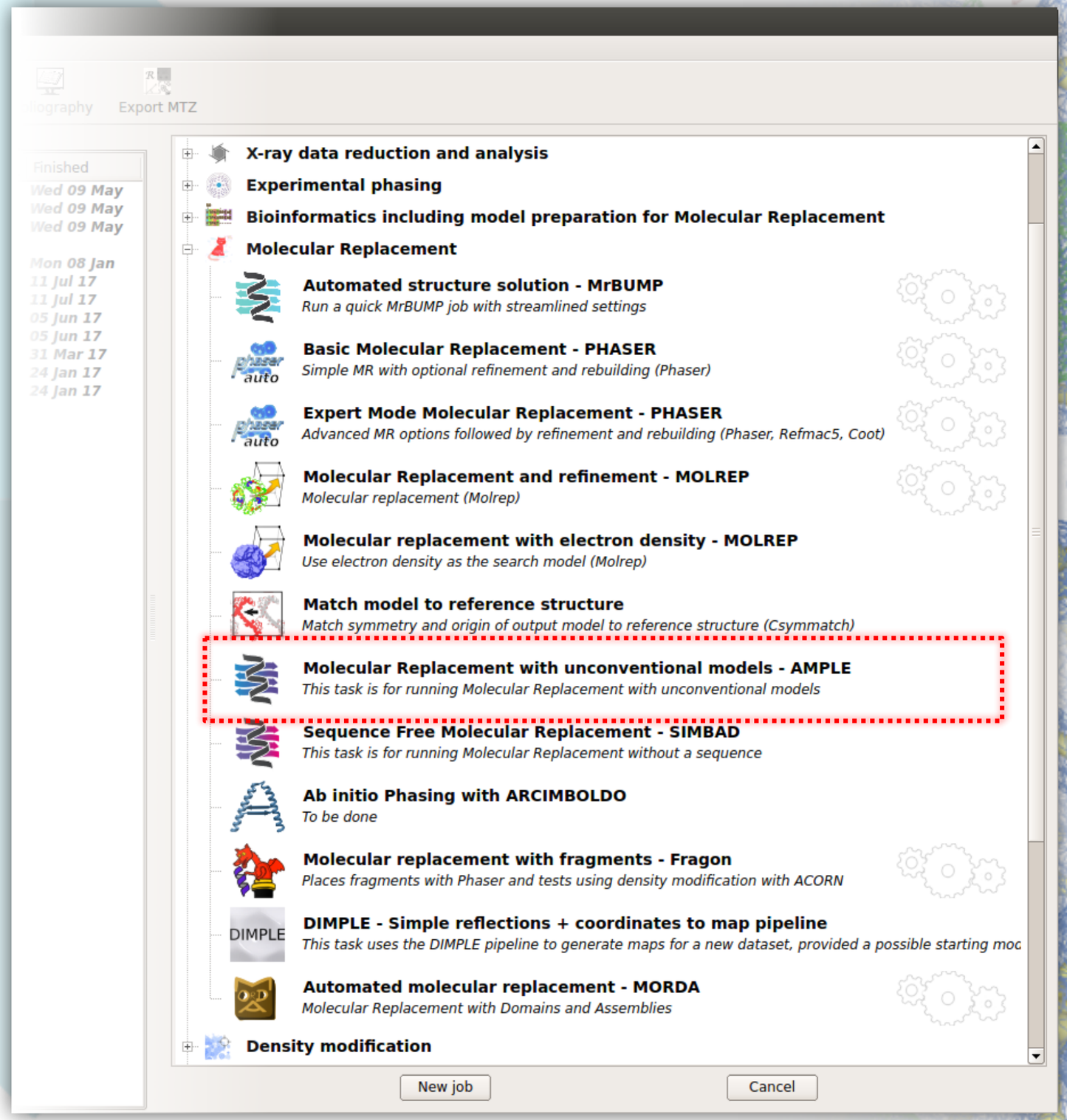
Ab initio Modelling for Molecular Replacement

UNIVERSITY OF LIVERPOOL

AMPLE

- Generate and prepare *ab initio* or *de novo* models using Rosetta or Quark software for use as search models in molecular replacement
- Works well for smaller, α-helical proteins (<120 residues), but has been known to work for larger targets





Science & Technology Facilities Council

Research Complex at Harwell

Crystal contaminants

THE PROTEIN SOCIETY

METHODS AND APPLICATIONS

Protein purification and crystallization artifacts: The tale usually not told

Ewa Niedzialkowska,[1,2,3] Olga Gasiorowska,[1,3] Katarzyna B. Handing,[1,3] Karolina A. Majorek,[1,3,4] Przemyslaw J. Porebski,[1,3] Ivan G. Shabalin,[1,3,4,5] Ewelina Zasadzinska,[6] Marcin Cymborowski,[1,3] and Wladek Minor[1,3,4,5]*

[1]Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, 1340 Jefferson Park Avenue, Jordan Hall, Room 4223, Charlottesville, Virginia, 22908

[2]Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Niezapominajek 8, Krakow, 30-239, Poland

[3]Midwest Center for Structural Genomics (MCSG), Argonne, Illinois, 60439

[4]Center for Structural Genomics of Infectious Diseases (CSGID), Chicago, Illinois, 60611

[5]New York Structural Genomics Research Consortium (NYSGRC), Bronx, New York, 10461

[6]Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, 1340 Jefferson Park Avenue, Jordan Hall, Room 6044, Charlottesville, Virginia, 22908

Received 1 October 2015; Accepted 2 December 2015
DOI: 10.1002/pro.2861
Published online 13 December 2015 proteinscience.org

Abstract: The misidentification of a protein sample, or contamination of a sample with the wrong protein, may be a potential reason for the non-reproducibility of experiments. This problem may occur in the process of heterologous overexpression and purification of recombinant proteins, as well as purification of proteins from natural sources. If the contaminated or misidentified sample is used for crystallization, in many cases the problem may not be detected until structures are deter-

---

Abbreviations: CSGID, Center for Structural Genomics of Infectious Diseases; GNAT, Gcn5-related N-acetyltransferase; IMAC, immobilized metal affinity chromatography; MAD, multi-wavelength anomalous diffraction; MR, molecular replacement; MCSG, Midwest Center for Structural Genomics; NYSGRC, New York Structural Genomics Research Consortium; PDB, Protein Data Bank; RMSD, root mean square deviation; SAD, single-wavelength anomalous dispersion; SEC, size exclusion chromatography; TEV, tobacco etch virus.

Additional Supporting Information may be found in the online version of this article.

Ewa Niedzialkowska and Olga Gasiorowska have contributed equally to this work.

The authors declare that there is no conflict of interest.

Description of Supporting Information material: Summary of data collection and refinement statistics for the deposited structures and the list of deposits used to identify crystallization artifacts by MR. Filename: Supplementary Materials.

Structural data are available in PDB database under accession numbers 4TNN, 4YYC, and 4ZNZ.

This work focuses on a particular difficulty that may occur as a result of accidental purification or contamination of the sample with a

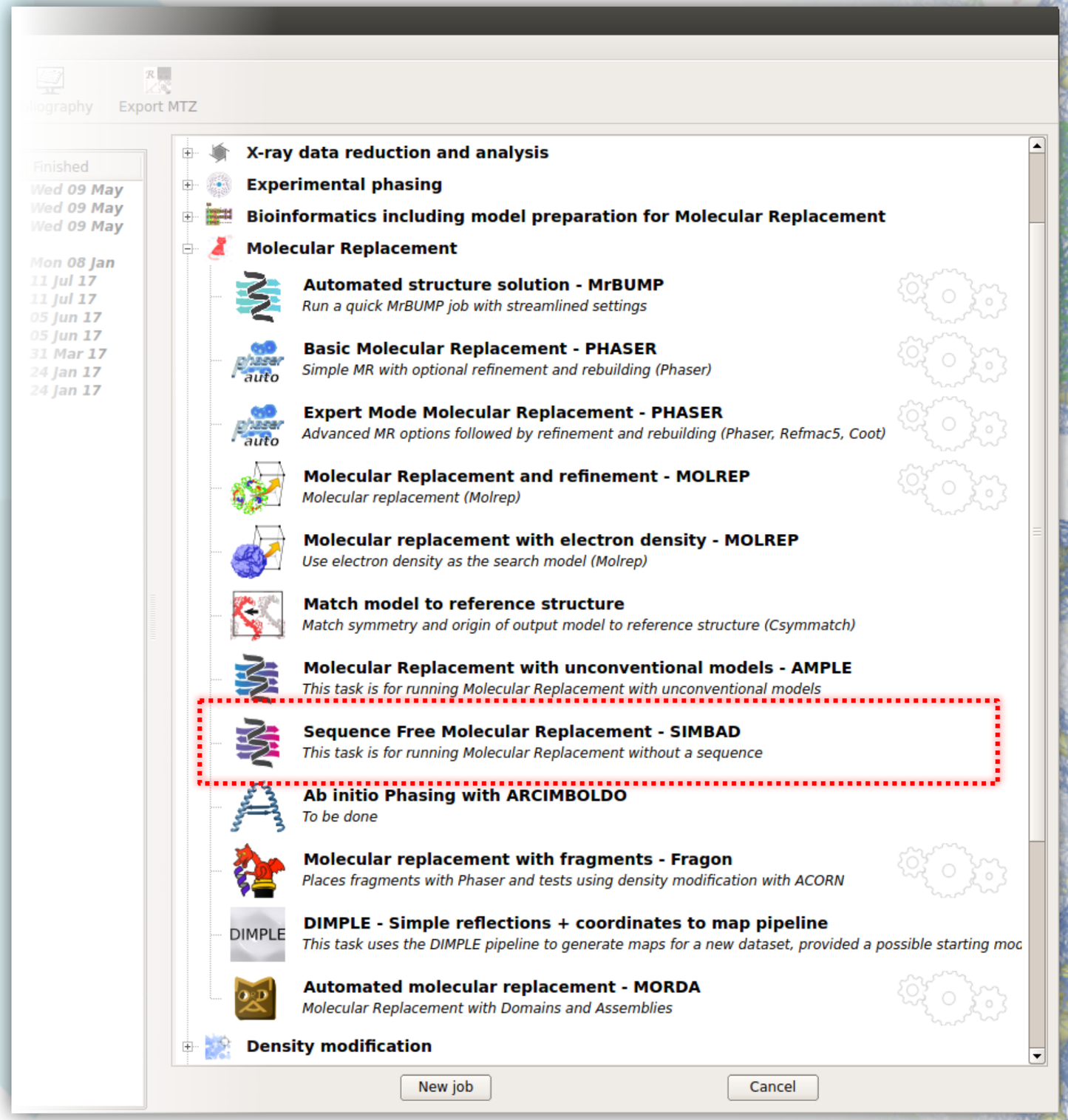
Crystal contaminants

UNIVERSITY OF LIVERPOOL

SIMBAD

- Sequence independent Molecular Replacement

- 3 Steps:
  1. nearest cell (lattice)
  2. contaminants
  3. entire non-redundant domain database (90000 models)



Science & Technology Facilities Council

Research Complex at Harwell

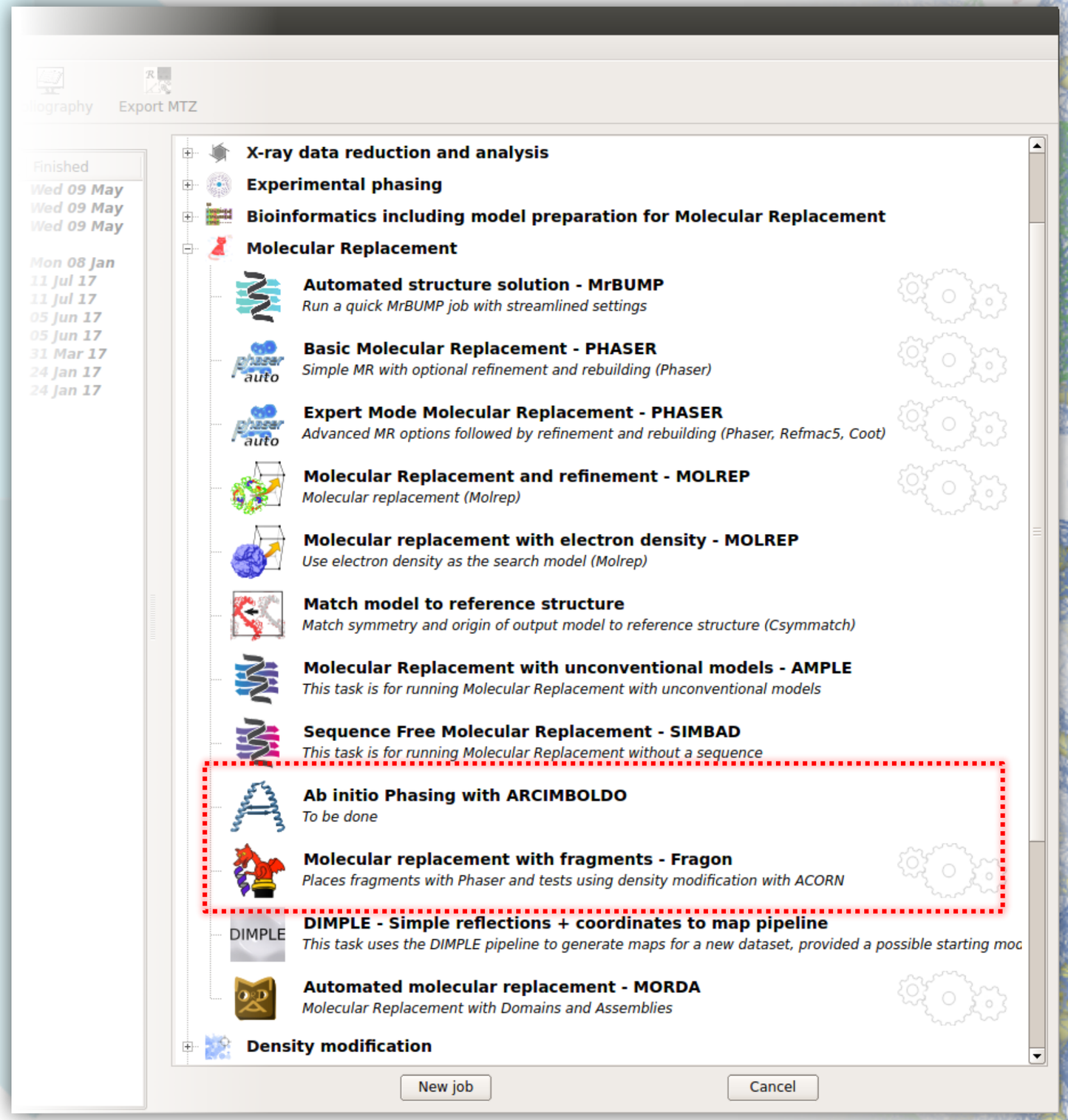
Other MR options

ARCIMBOLDO

- *Ab initio* phasing from fragment models

FRAGON

- MR using helical models at high resolution – using ACORN & Phaser



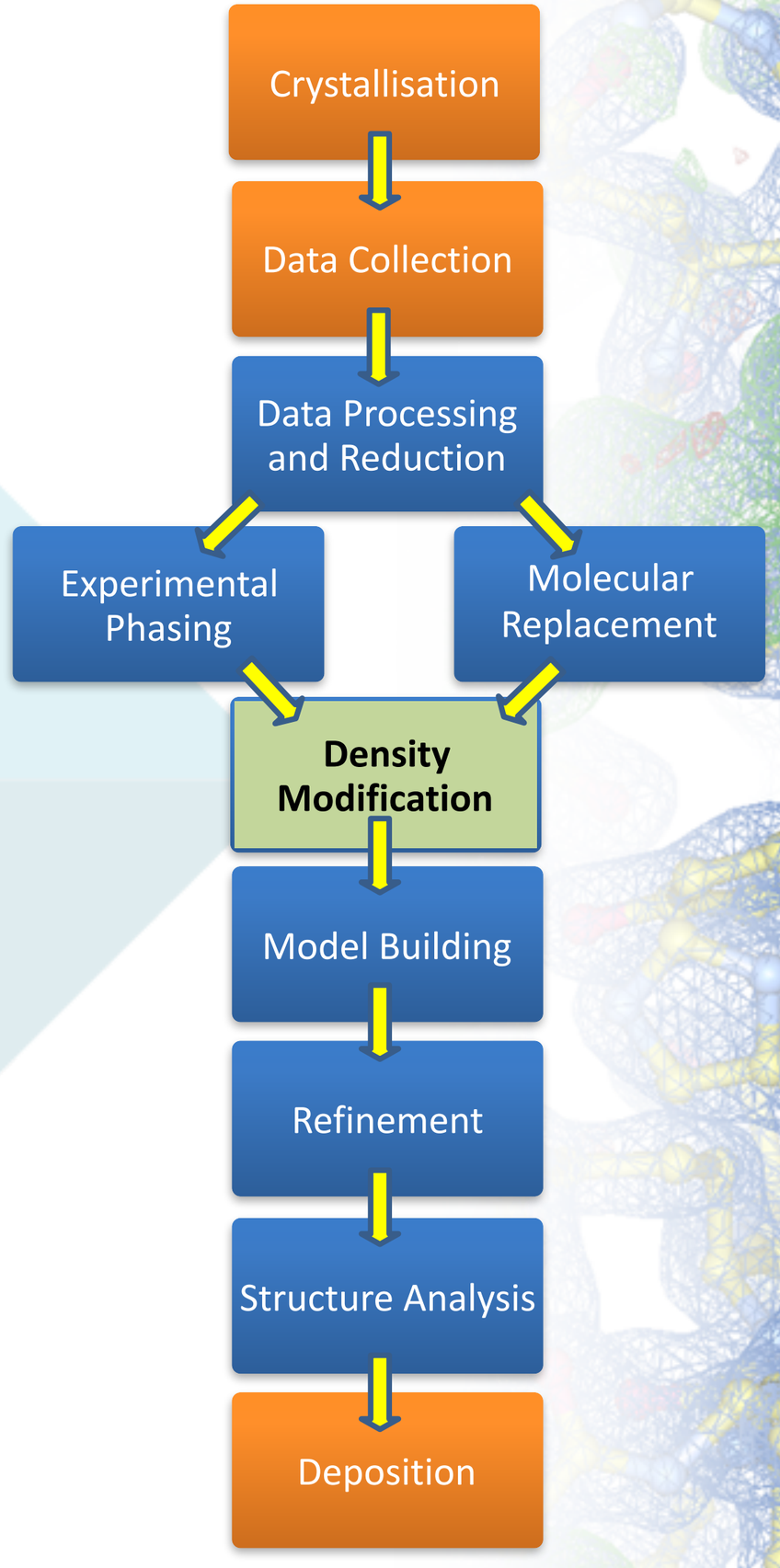
Science & Technology Facilities Council

Research Complex at Harwell

Back to the roadmap..

Density Modification

- A tool for phases improvement
- *Traditional approaches:*
  - NCS averaging
  - Histogram matching
  - Solvent flattening
  - Phase combination
    - Experimental phases
    - Phases from modified map
  - Iterative procedure
- *Statistical approaches:*
  - Improved phase probability distributions
  - Removes bias
  - Slower than traditional approaches



Science & Technology Facilities Council
Research Complex at Harwell

Density Modification





- DM can be perfect if noise is genuinely statistical (never the case)
- Otherwise, DM is equivalent to bringing in additional data
- Imperfectness can potentially lead to artefacts
- Existing programs are robust and DM is known to work

Science & Technology Facilities Council

Research Complex at Harwell

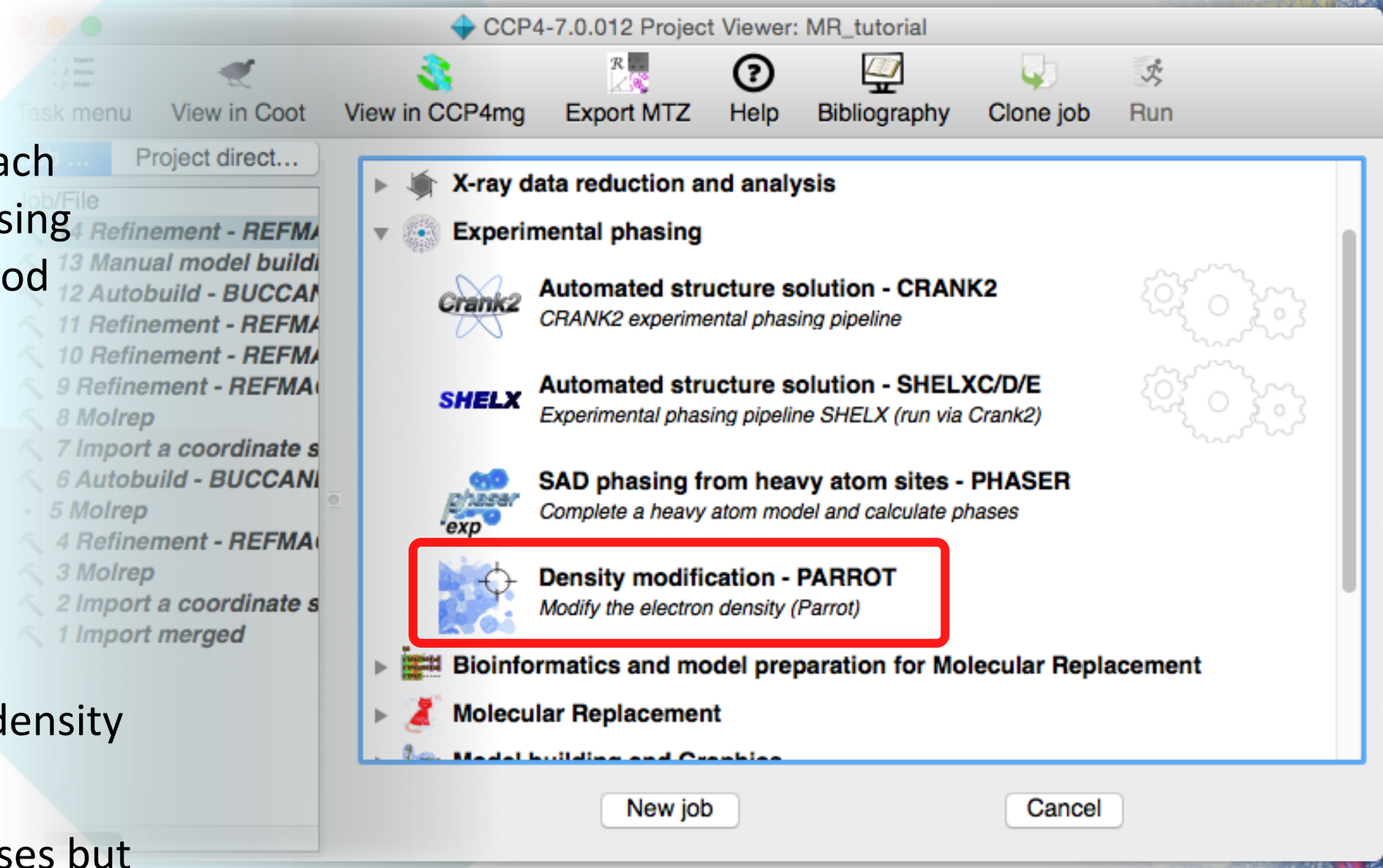
Density Modification Tools

Parrot

- Traditional approach benefiting from using maximum likelihood methods
- Fast and fully automated

Pirate

- Statistical-based density modification
- Better in some cases but slower than Parrot



Science & Technology Facilities Council

Research Complex at Harwell

Density Modification Tools

SHELXE

- Density modification combined with C-alpha tracing
- Works well starting from partial MR solution or from Experimental phases
- Usually works best with resolution better than 2.5 Å but can work at lower resolutions

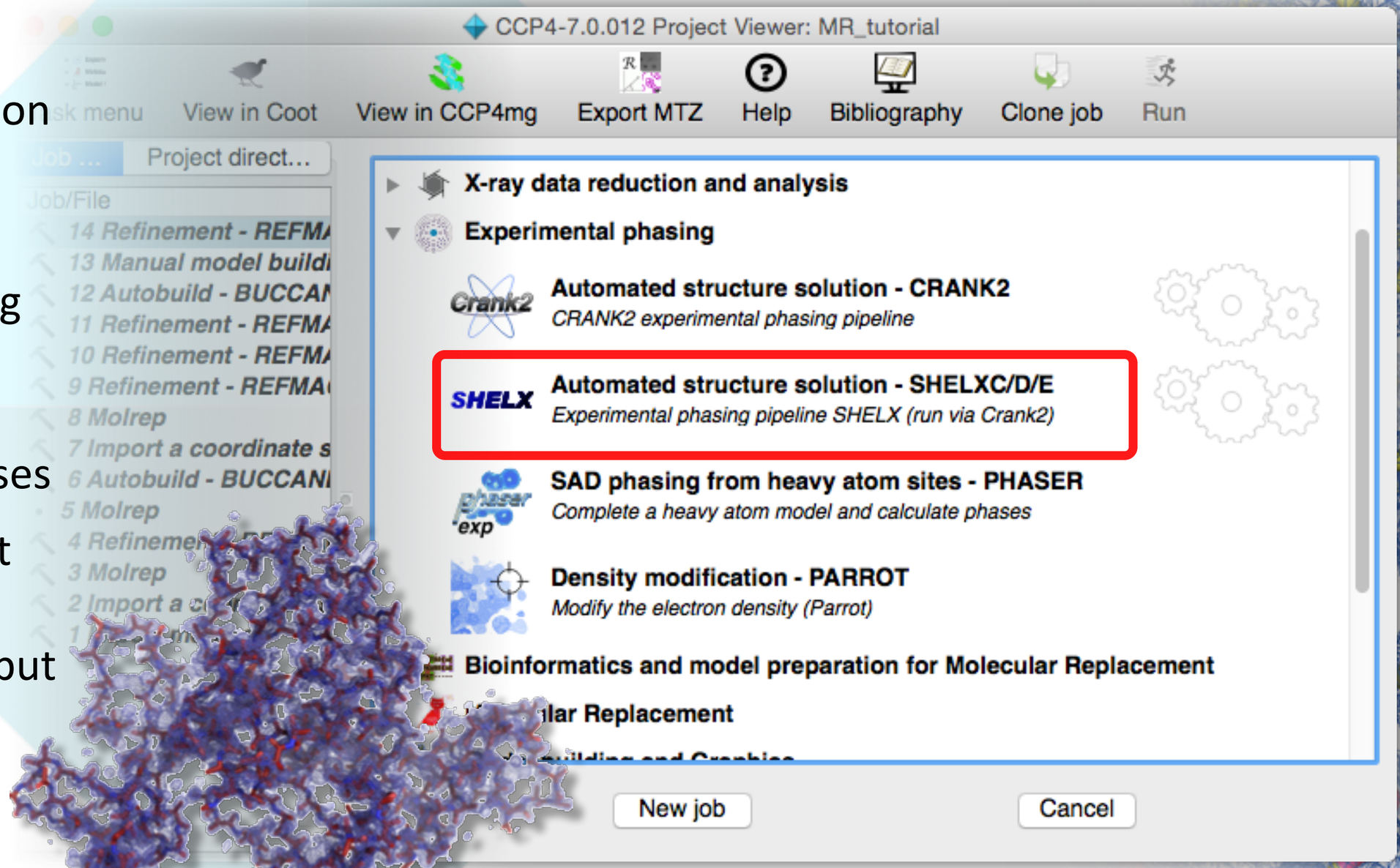
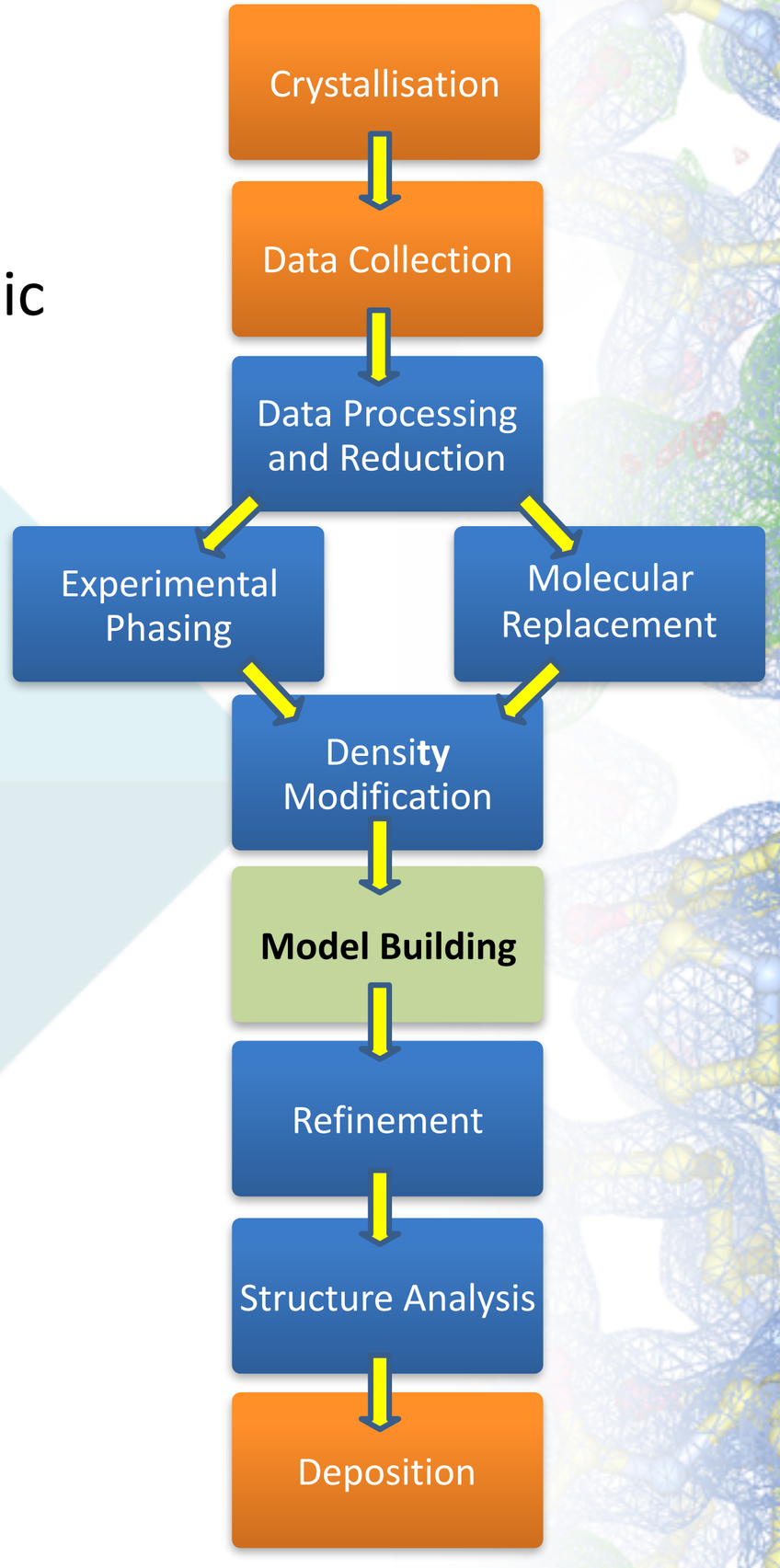


Figure courtesy of Andrea Thorn

Model Building

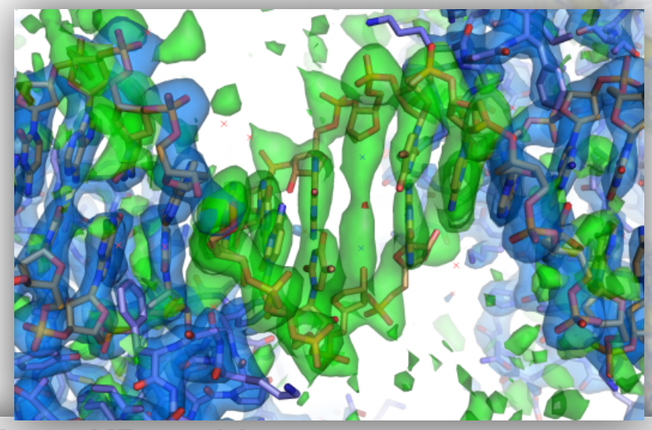
- Interpret electron density map in terms of atomic coordinates
  - X-ray experiment and phasing yield the electron density map
  - Model's atom need to be put in the correct locations of the map
  - The process is aided by the chemical knowledge of residue structure, as well as secondary structure assignments
  - Validation tools are essential for choosing the right electron density interpretation:
    - Quality of density fit
    - Ramachandran restraints
    - Geometry analysis
- Try to account for as much of the density as possible including solvent molecules



Science & Technology Facilities Council

Research Complex at Harwell

Model Building Tools



Buccaneer

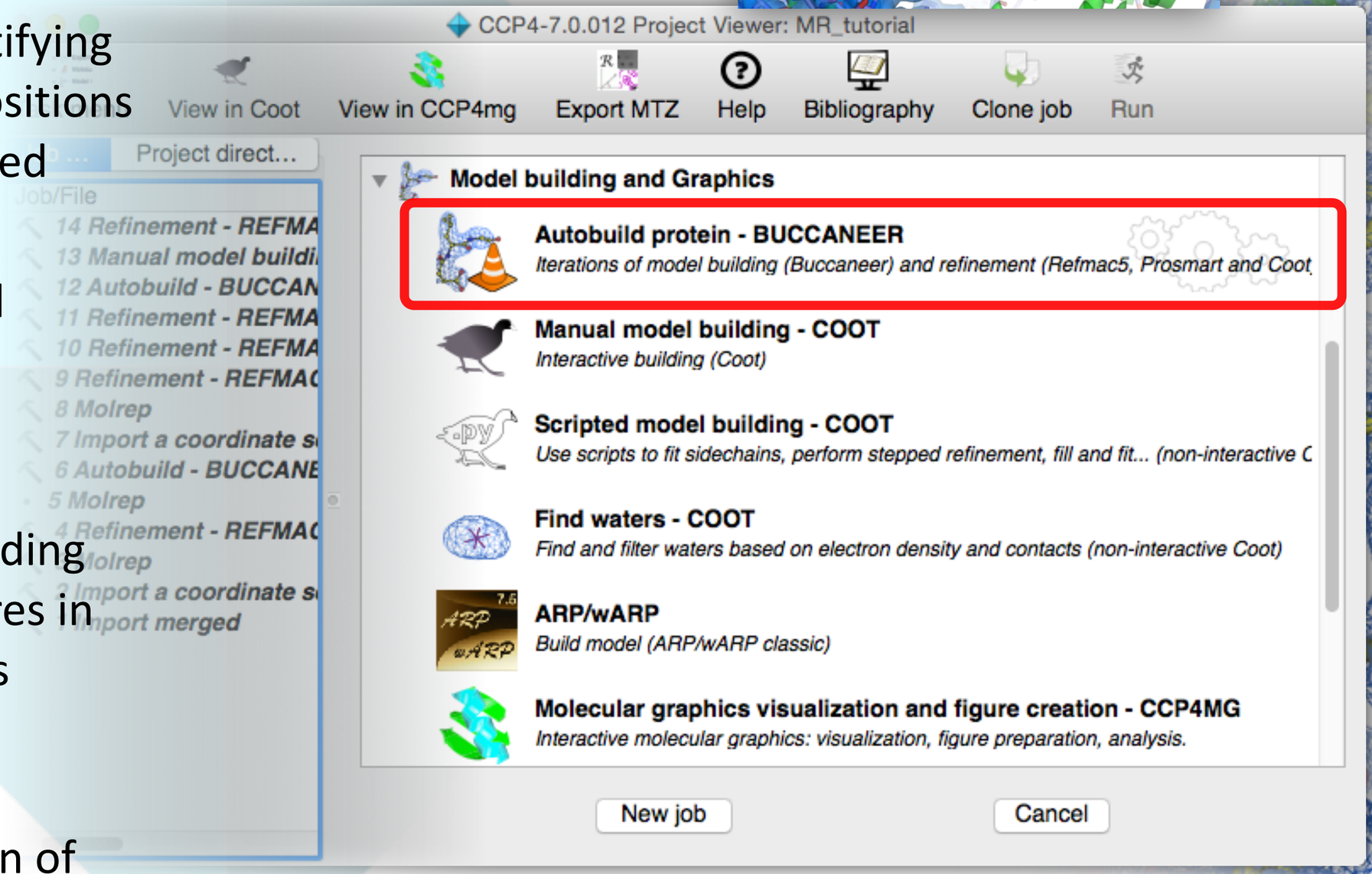
- Chain tracing by identifying connected C-alpha positions using a likelihood-based density target
- Low-resolution model building

Nautilus

- Automatic model building of nucleotide structures in electron density maps

Privateer

- Building and validation of carbohydrates



Science & Technology Facilities Council

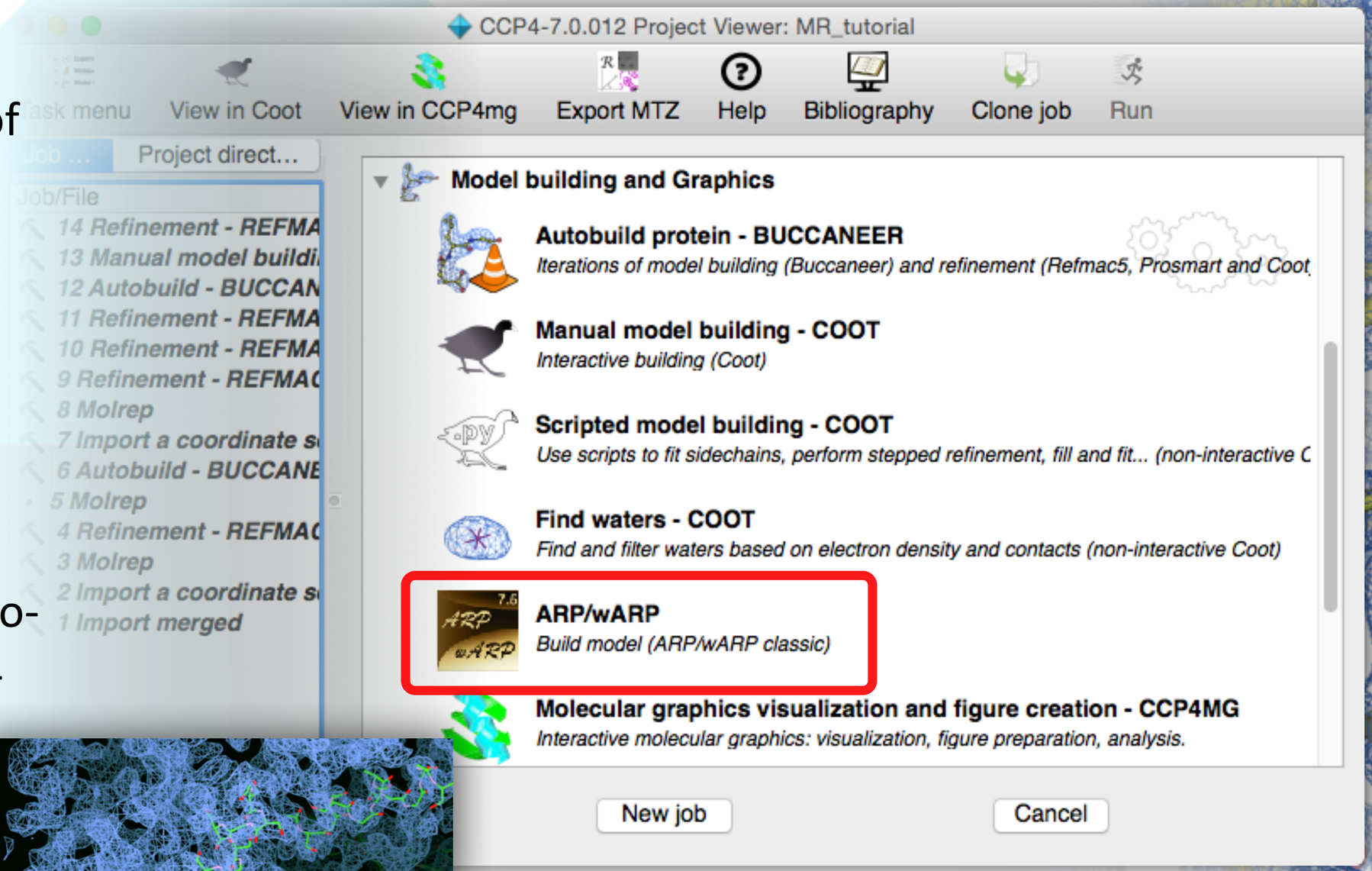
Research Complex at Harwell

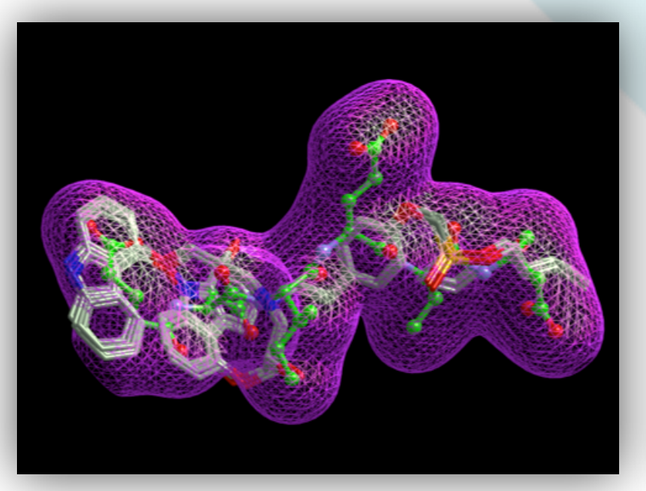
Model Building Tools

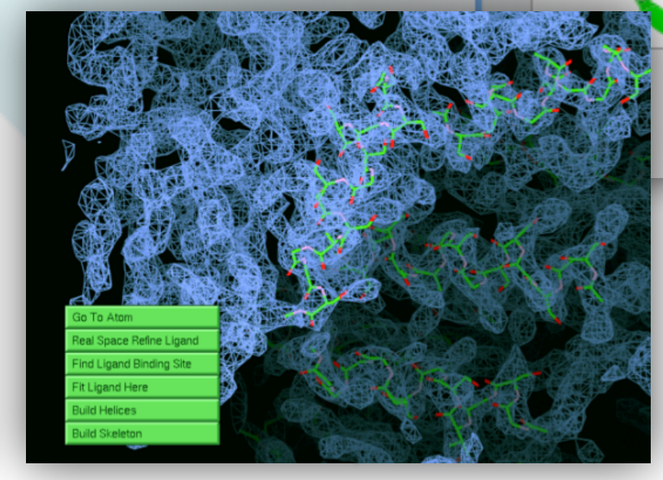
EMBL

ARP/wARP

- Automated building of
  - proteins
  - RNA/DNA
  - secondary structure
  - side chains
  - loops
  - solvent
  - ligands
- Now integrated and co-distributed with CCP4







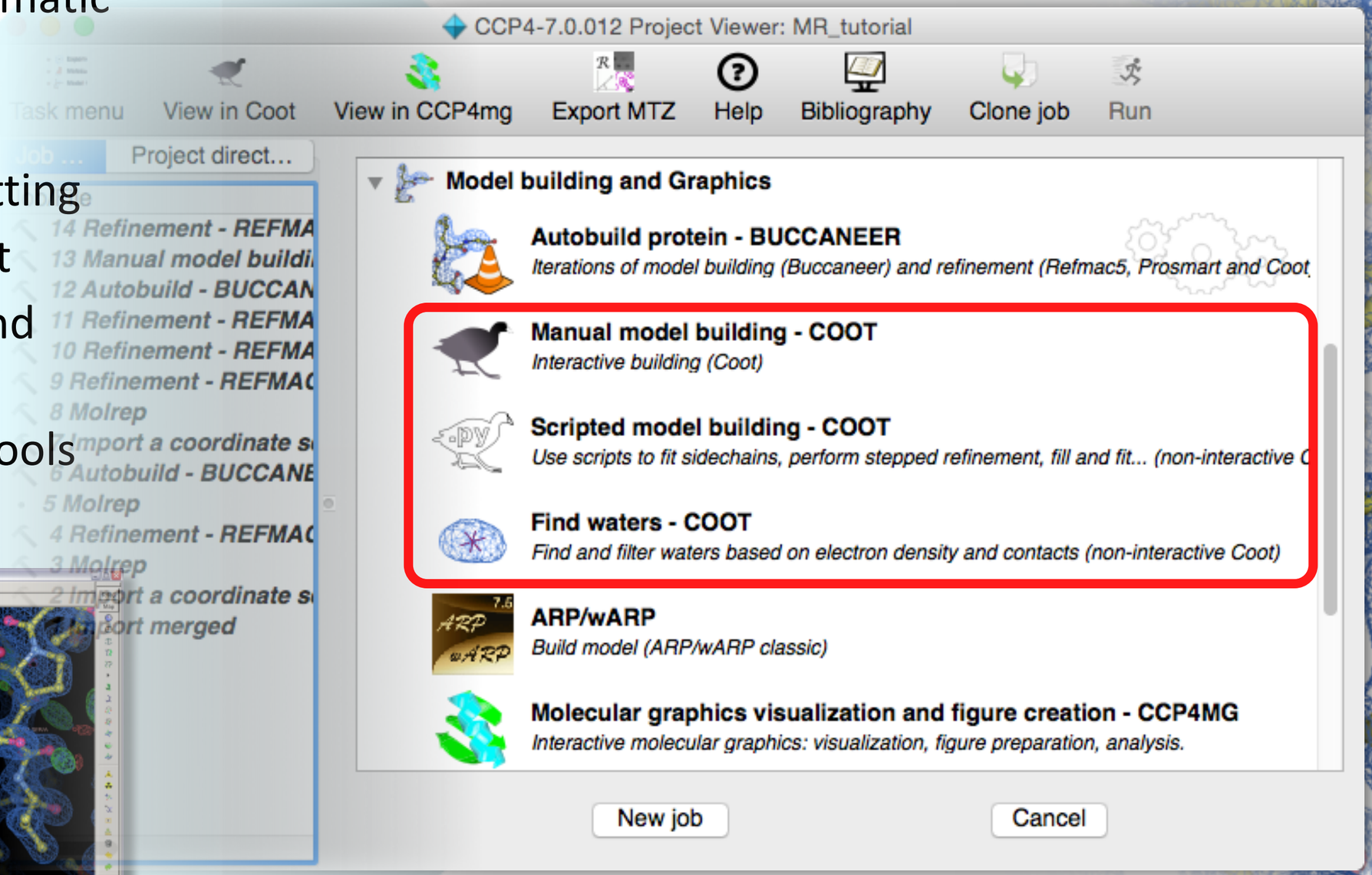
Science & Technology Facilities Council

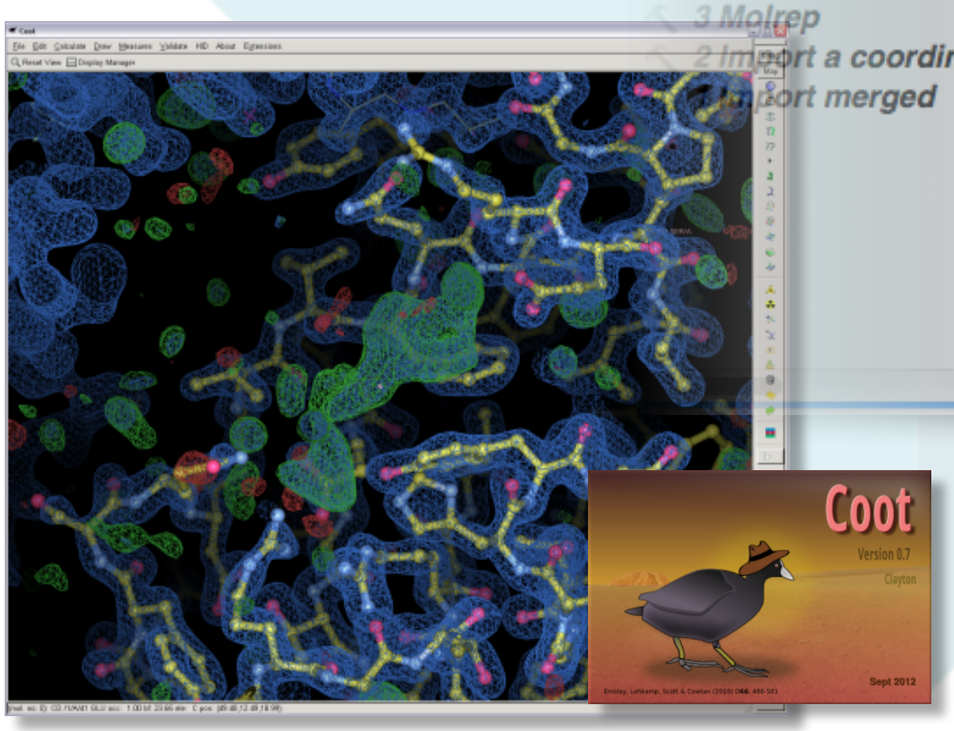
Research Complex at Harwell

Model Building Tools

Coot

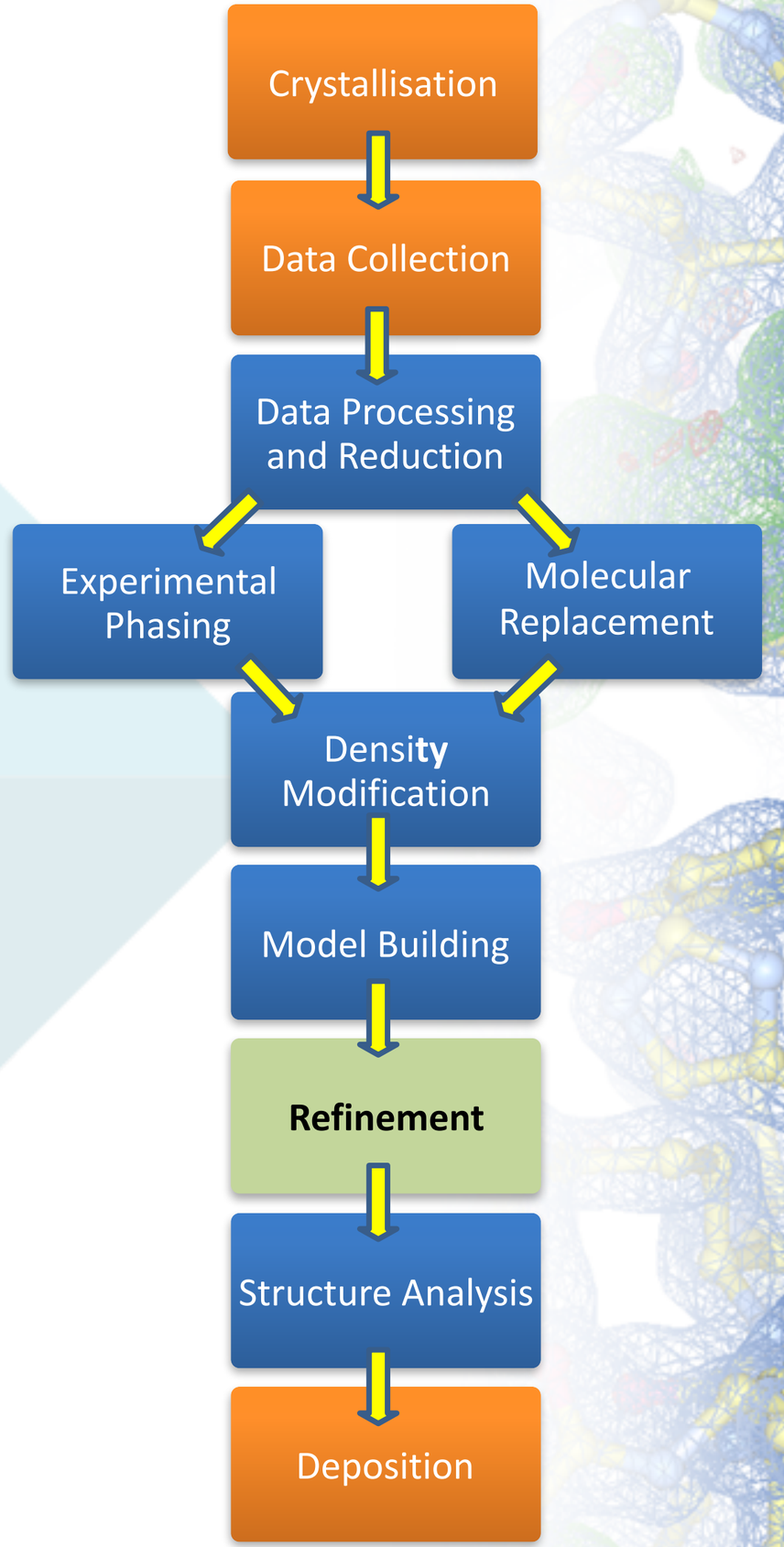
- Manual and semiautomatic model building
- Interactive graphics
- Ligand building and fitting
- Real-space refinement
- Paired with Refmac and jLigand
- Wealth of validation tools
- Integrated with CCP4





Refinement

- Refinement is one of most important procedures in structure solution process
- Refinement is used mainly to
  - produce a structure which is properly balanced between chemistry restraints and experimental observations
  - improve density maps for further cycle of model building
- Refinement often uses additional information from various sources
  - Phases
  - Twinning (if detected)
  - External restraints
  - NCS restraints
  - TLS groups



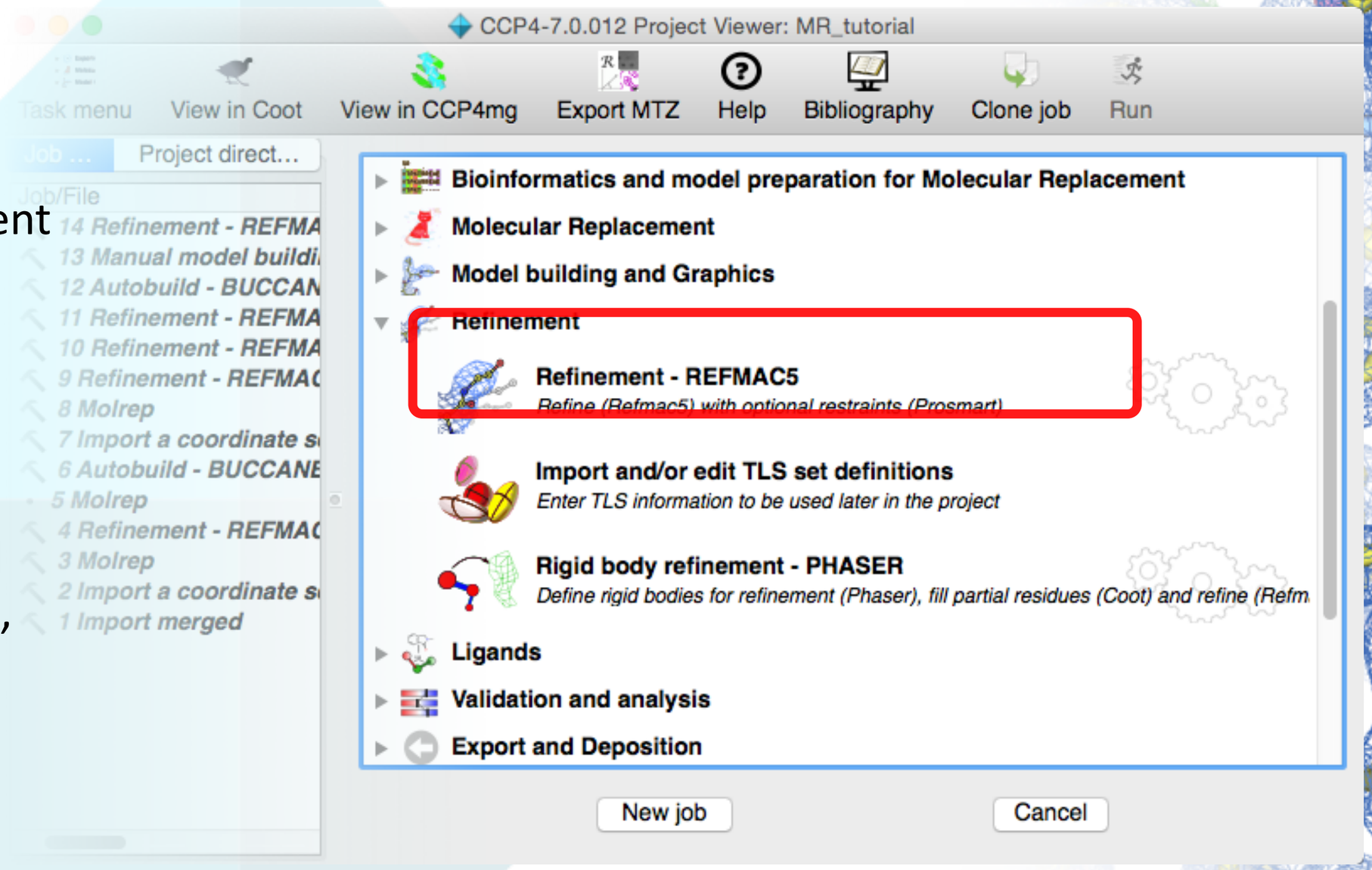
Science & Technology Facilities Council

Research Complex at Harwell

Refinement

Refmac 5.8

- CCP4's main refinement program
- Implements several likelihood target functions for experimental data (amplitudes, twinned data, external phases, Lattman coefficients)
- Restraints
  - ~ CCP4 monomer library
  - ~ Custom library
  - ~ External restraints



- Several refinement protocols
  - ~ Rigid body
  - ~ Restrained
  - ~ Jelly-body
- NCS refinement
  - ~ local
  - ~ global

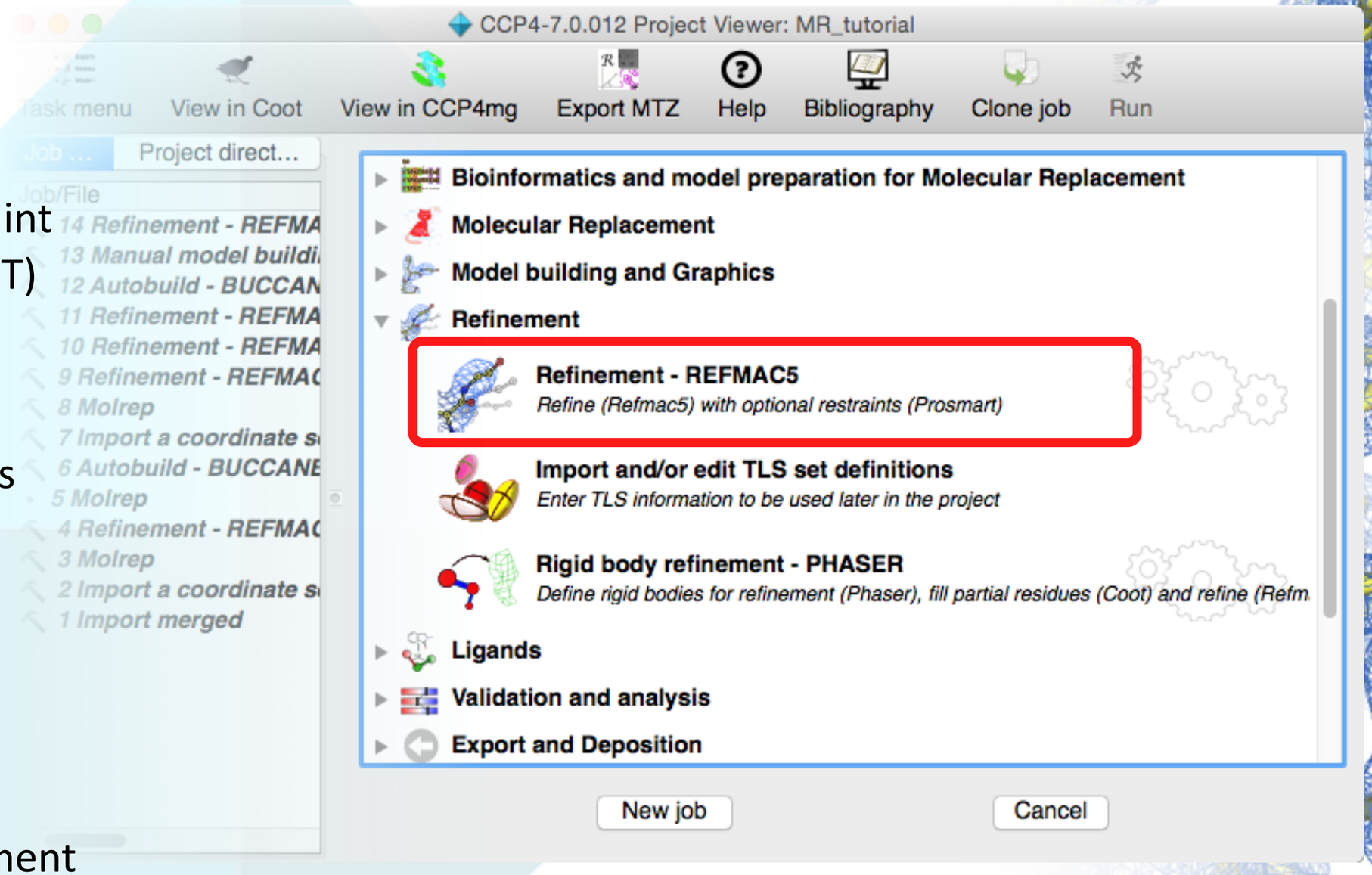
Science & Technology Facilities Council

Research Complex at Harwell

Refinement

Refmac 5.8

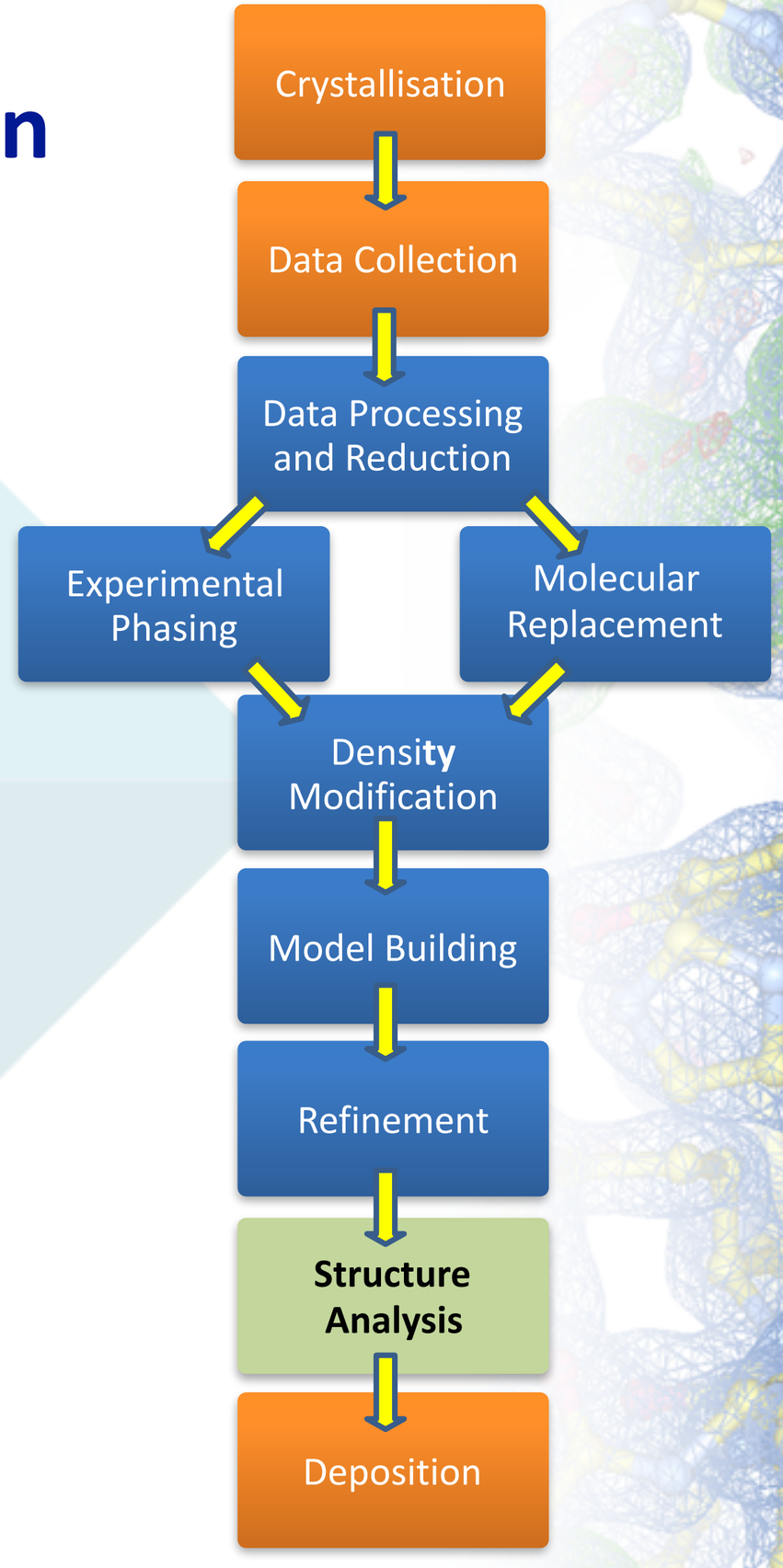
- Improved jelly body refinement
- Improved use of restraint information (ProSMART)
- DNA/RNA base pair restraints
- Sugar/pucker restraints
- Simultaneous density modification and refinement
- Better anomalous difference maps
- Beta ensemble refinement
- Multi-imputation and averaging of structure factors to desired resolution



- Sampling of conformational space – Gibbs sampling
- Estimation of individual errors of atoms – less biased electron density calculation

Structure Analysis & Visualisation

- To find out structural features which may shed light on chemical activity or structure function
- Identification of complexes (quaternary structures)
- Finding and comparing homologues
- Visualise and inspect the target structure
- Can also be used earlier on in structure solution process
  - ~ Provide information on complexes, domains, mobility of structure for use in molecular replacement or for construct design



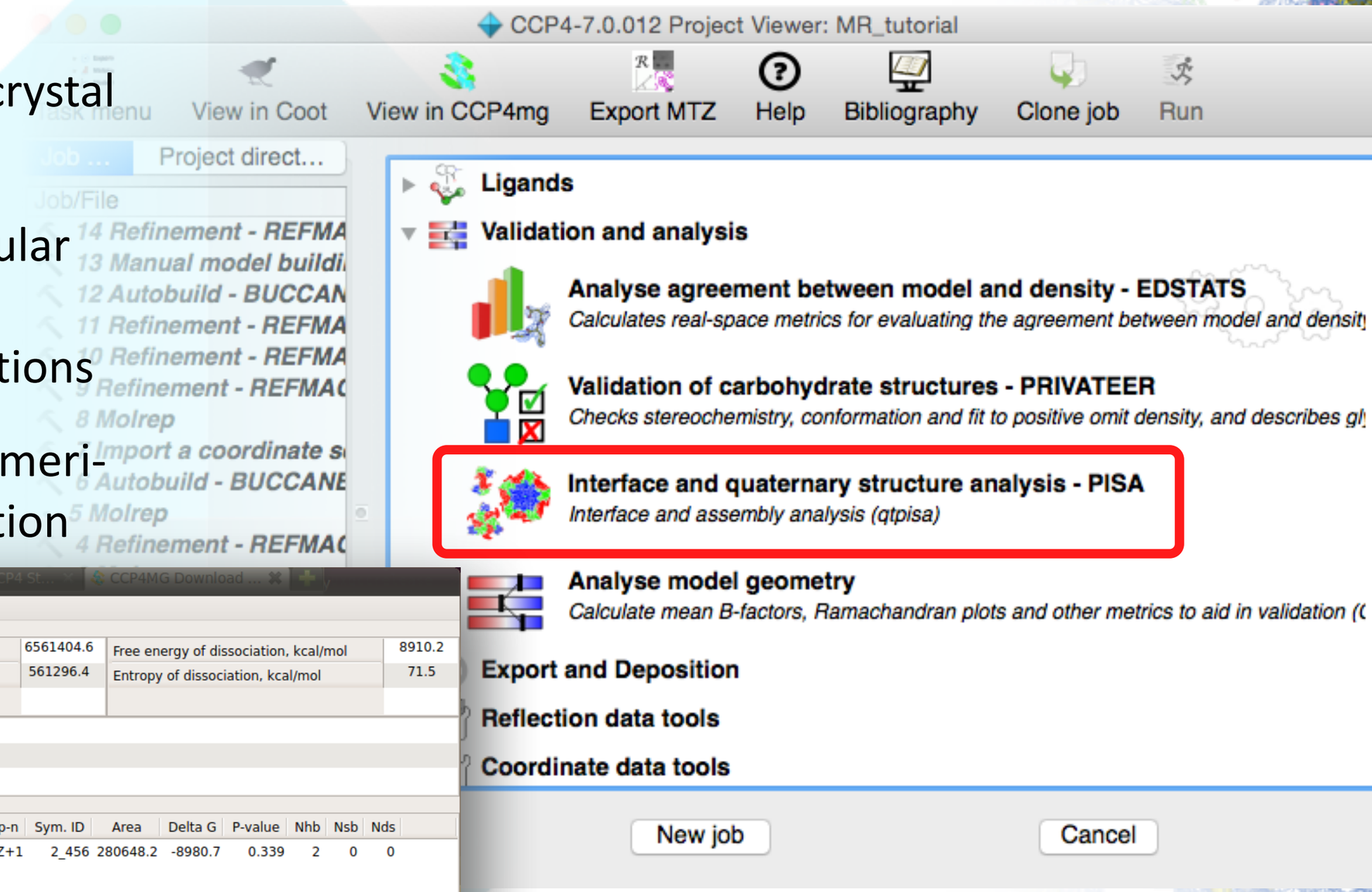
Science & Technology Facilities Council

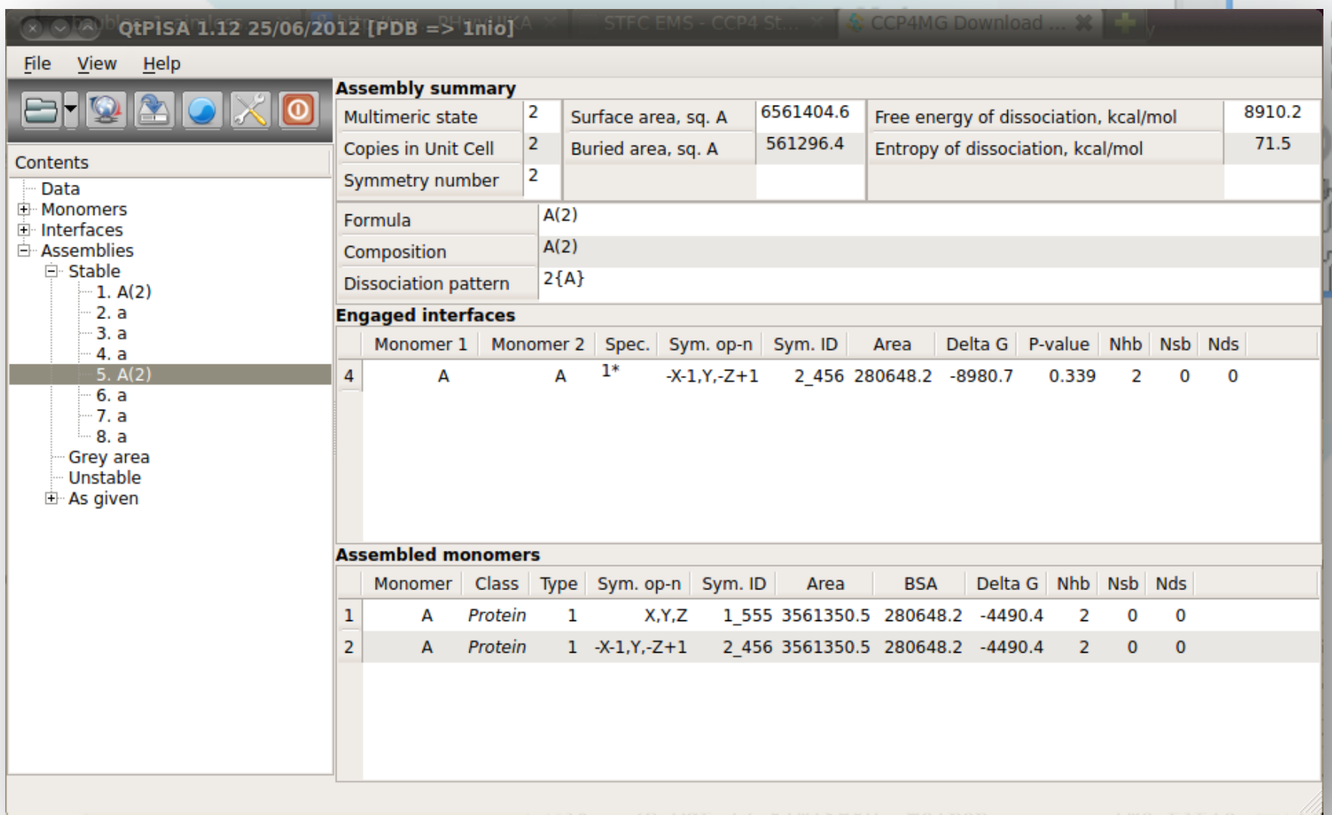
Research Complex at Harwell

Structure Analysis

PISA

- Automatic inference on multimeric states from crystal packing
- Analysis of macromolecular interfaces and macromolecular interactions
- Analysis of protein oligomeri-sation versus concentration





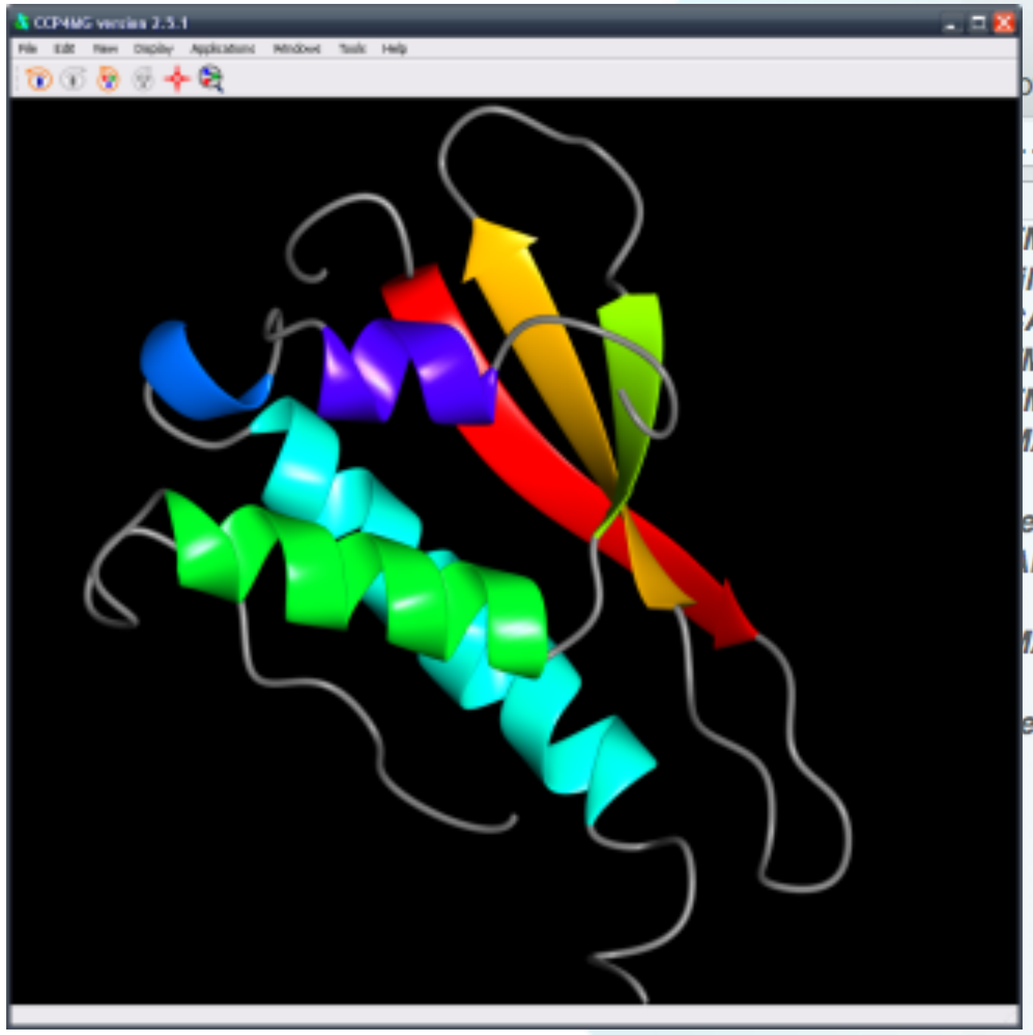
- QtPisa – QT based interface

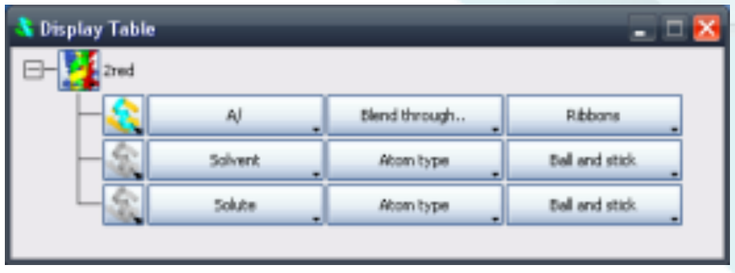
Science & Technology Facilities Council

Research Complex at Harwell

Structure Analysis

CCP4MG





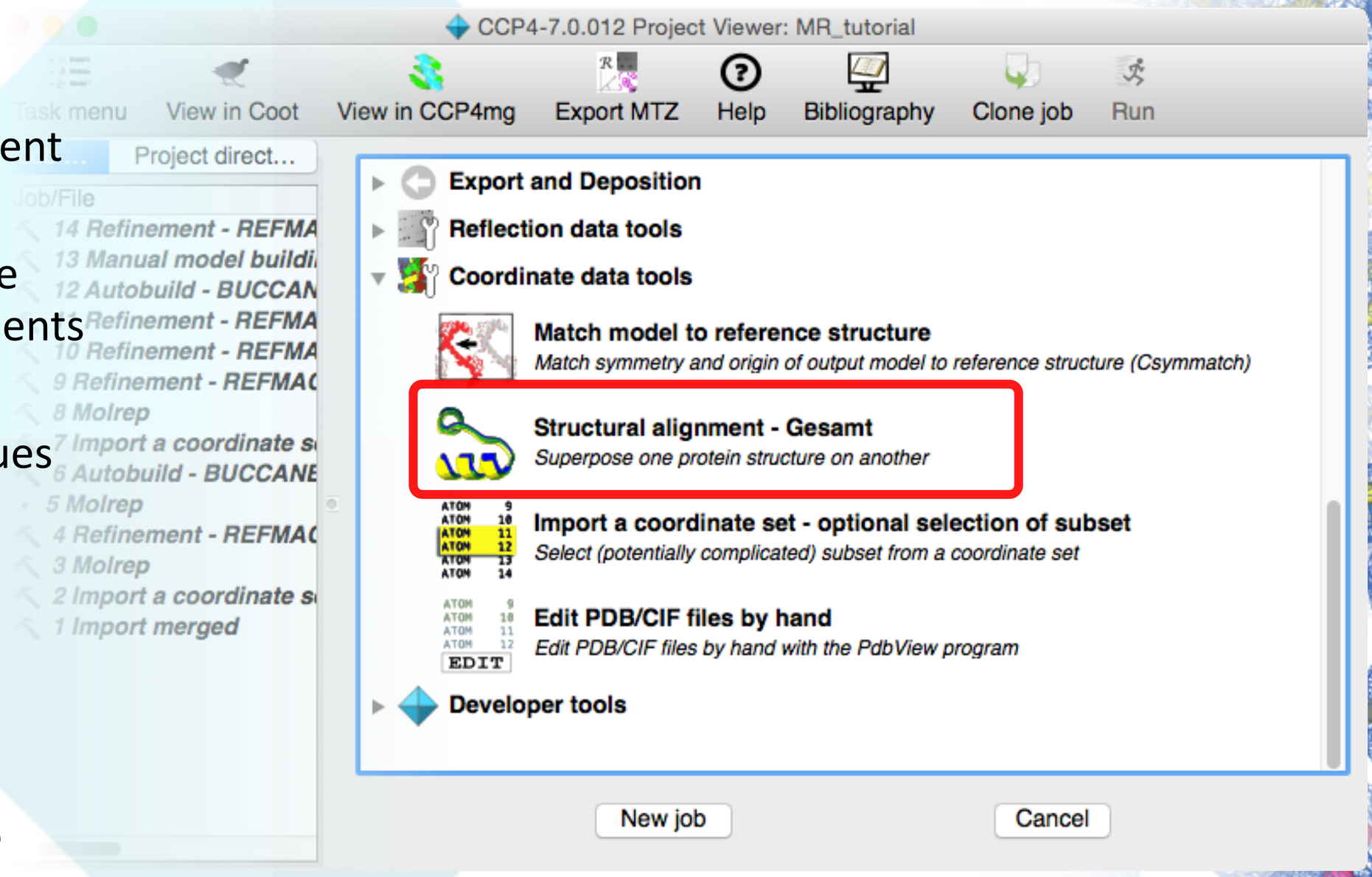
Science & Technology Facilities Council

Research Complex at Harwell

Structure Analysis

Gesamt

- 3D structure alignment and superposition
- Makes both pairwise and multiple alignments
- Fast PDB scans for structural homologues
- Identification of structural domains
- Replacement for Superpose (SSM)
- Doesn't rely on secondary structure



Validation and Deposition

Coot

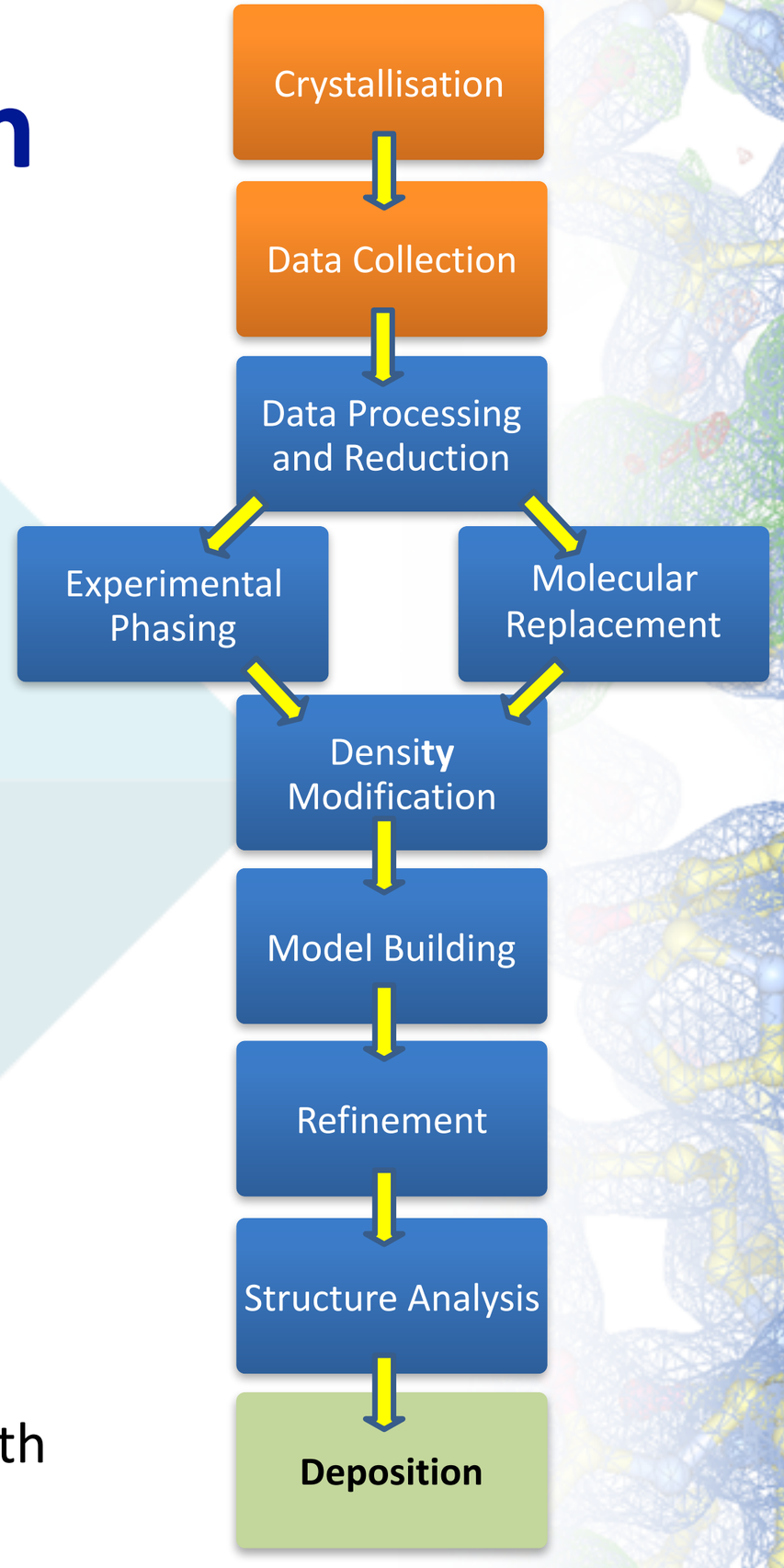
- Comprehensive set of validation tools: Ramachandran plots, density fit analysis, geometry analysis, unmodelled blobs etc.

PDB_REDO

- Optimisation of refinement parameters: weighting of X-ray terms, B-factors
- Checking rotomers
- Comparison of refinement with rest of PDB statistics

Deposition

- Dedicated deposition preparation modules in both CCP4i2 and CCP4 Cloud



Science & Technology Facilities Council

Research Complex at Harwell

Other useful information

Science & Technology Facilities Council
Research Complex at Harwell

Command Line Operation

Example:

refmac5 HKLIN my.mtz HKLOUT output.mtz XYZIN my.pdb XYZOUT output.pdb

Keywords:

- Most CCP4 programs take “keywords” as command options through standard input to the program
- Gives access to all options of the program and the most fine-grained control
- Can be provided through ccp4i using “**Run & view com file**”

Scripting:

- Repetitive tasks can be scripted
- Pipelines of programs can be created in a script

Command line

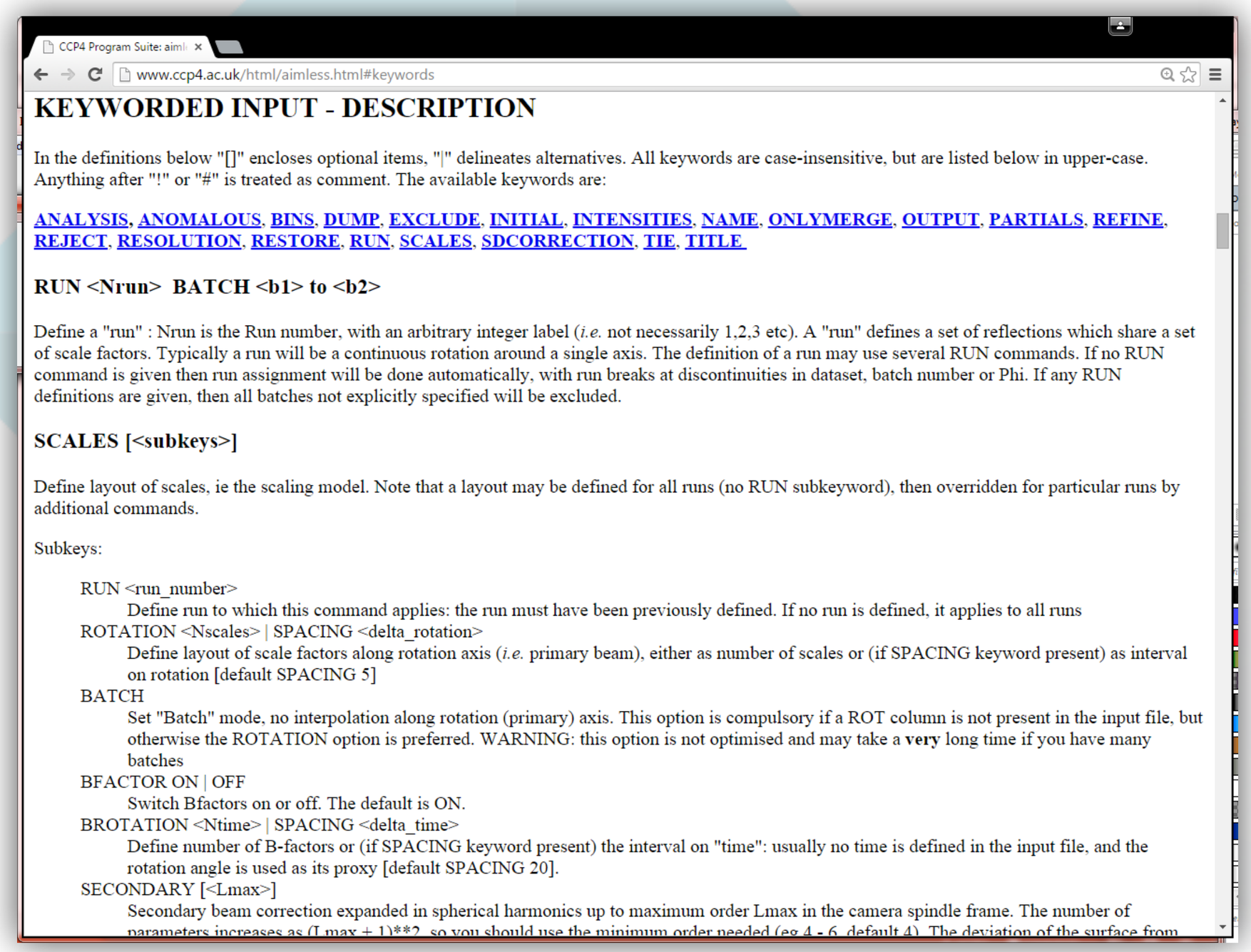
Keywords

```
f2mtz hklin ${1} hklout shelxe-output.mtz << eof
TITLE   ample
cell    ${cell}
symm    ${symm}
labout  H K L F FOM PHI SIGF
CTYPOUT H H H F W P Q
pname   ample
dname   ample
END
```

Science & Technology Facilities Council
Research Complex at Harwell

Command Line Operation

- *Documentation for CCP4 program keywords provided by documentation pages online or in the suite ($CCP4/html/INDEX.html):*

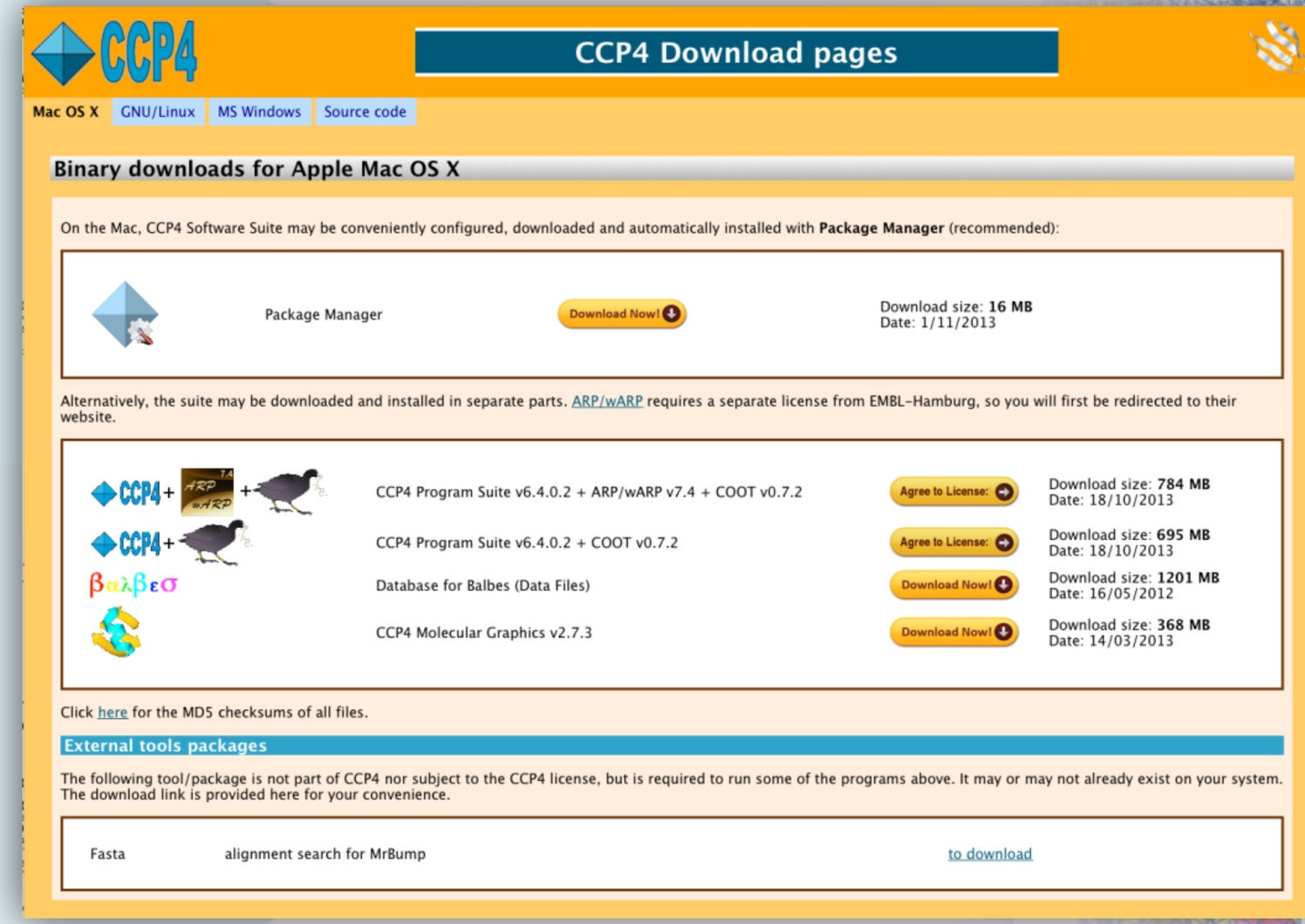


Science & Technology Facilities Council

Research Complex at Harwell

Downloading the Suite

- Download CCP4 for Windows, Mac or Linux
- Co-distribution of ARP/wARP
- Package Manager – easy installation
- Only 64-bit version for Mac OSX, both 32-bit and 64-bit versions for other systems
- Additional software dependencies also available
- Download page accessible from http://www.ccp4.ac.uk

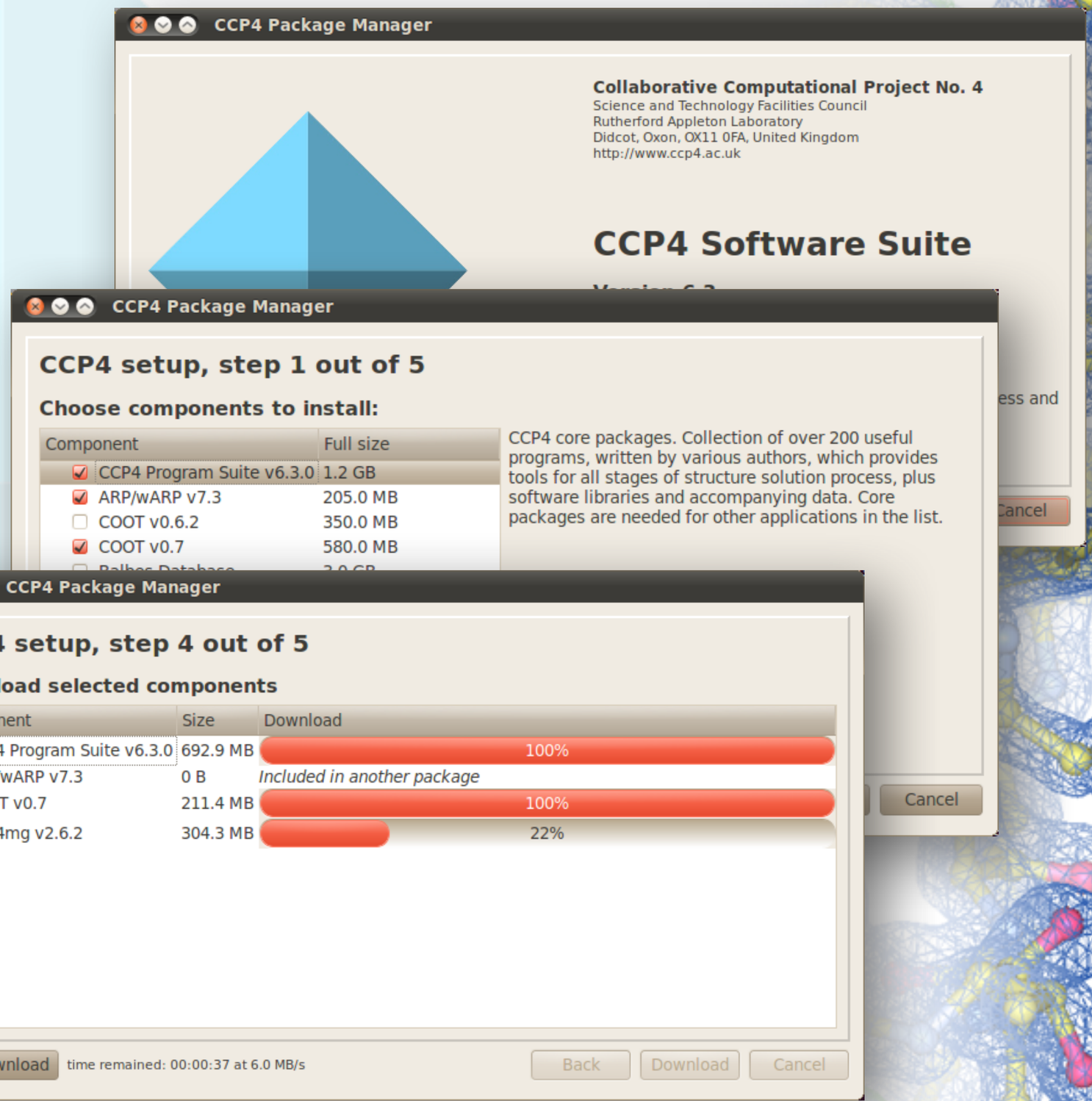


Science & Technology Facilities Council

Research Complex at Harwell

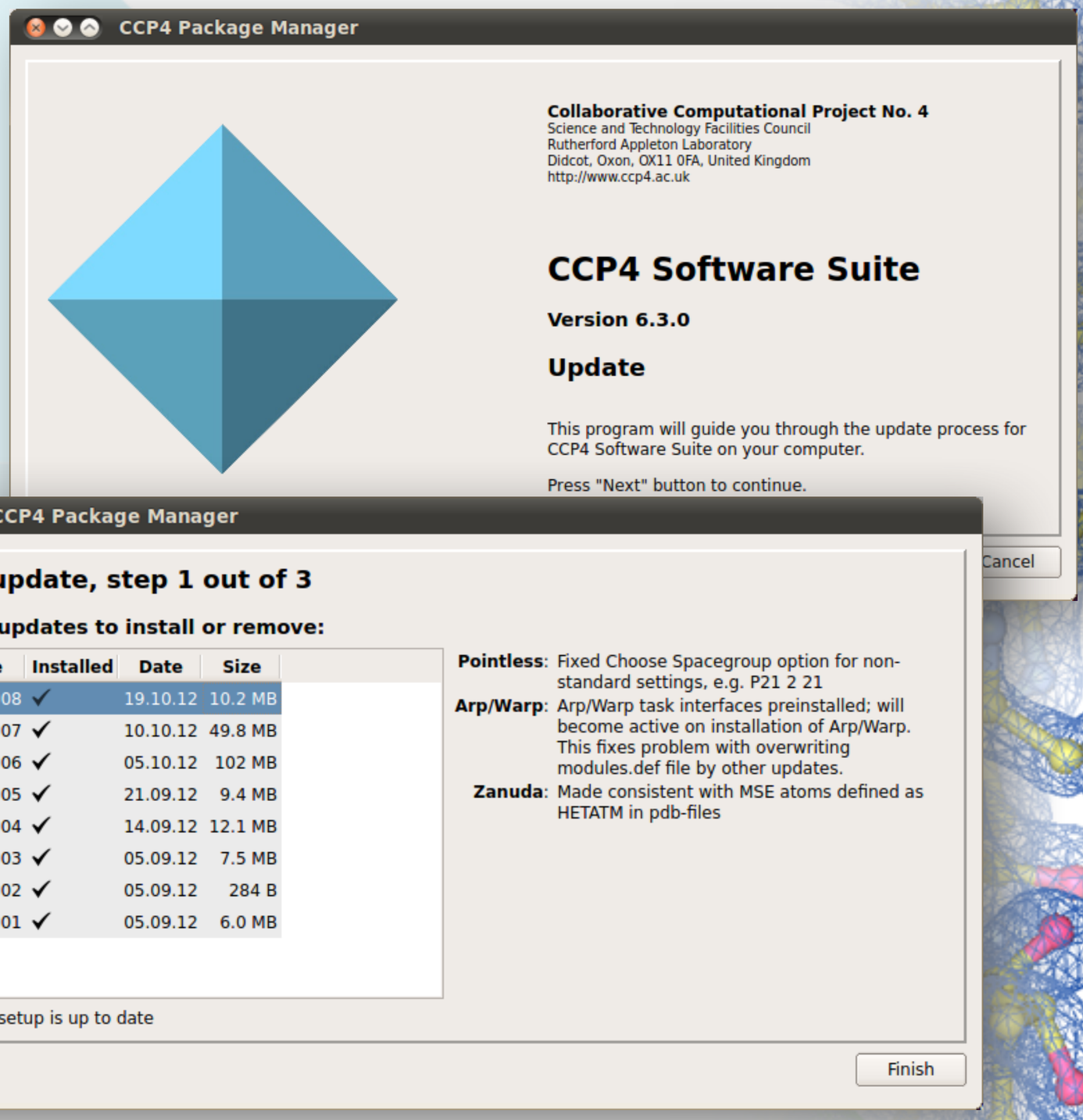
CCP4 Setup Manager

- Easy installation mechanism for CCP4 and associated packages (Coot, ARP/wARP, CCP4mg)
- Available on Linux and Mac OSX platforms
- Sets up system variables automatically if you wish



Update Manager

- Automatically install latest fixes and new features
- Allows for rolling back to previous versions
- Updates made available once every two weeks or as and when they are needed
- Available on Linux, Mac OSX and Windows



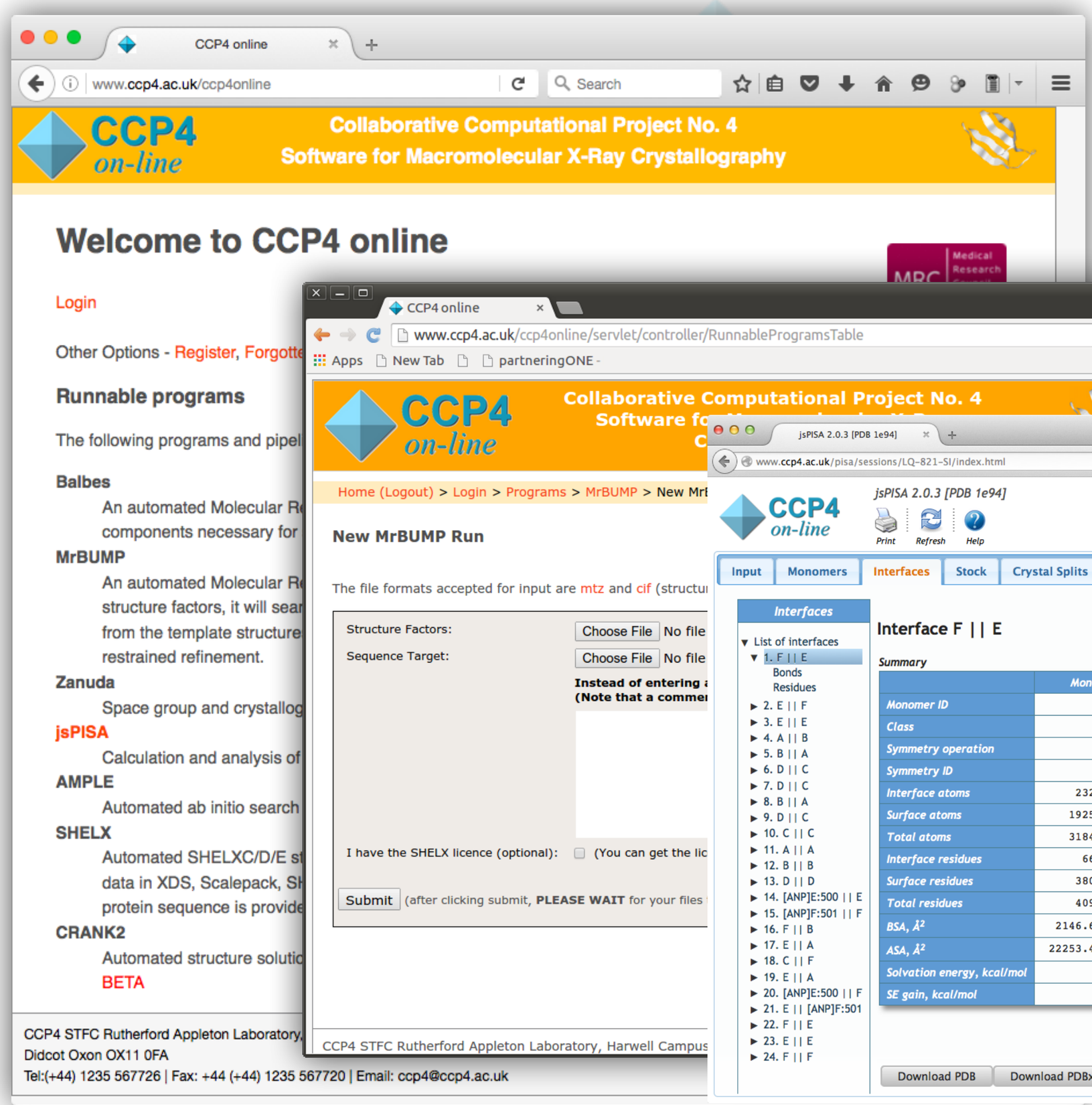
CCP4 Roadmap

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 on-line services

http://www.ccp4.ac.uk/ccp4online



Acknowledgements

CCP4 Core group: Andrey Lebedev, Eugene Krissinel, Charles Ballard, David Waterman, Marcin Wojdyr, Ville Uski, Karen McIntyre, Kyle Stevenson, Martyn Winn

LMB/MRC: Andrew Leslie, Phil Evans, Garib Murshudov, Rob Nicholls, Harry Powell, Owen Johnson, Fei Long, Paul Emsley, Andrea Thorn

Phaser Group: Airlie McCoy, Randy Read, Rob Oeffner, Gabor Bunkoczi

YSBL York: Keith Wilson, Kevin Cowtan, Liz Potterton, Stuart McNicholas, Eleanor Dodson

Diamond: Gwyndaf Evans, Graeme Winter, Luis Fuentes-Montero

Leiden: Raj Pannu, Pavol Skubak

Others: Bernhard Lohkamp, Clemens Vonrhein, Ruslan Sanishvili, Frank Von Delft, Martin Noble, Jaclyn Bibby, Daniel Rigden, Jens Thomas, Alun Ashton, David Brown, Arwen Pearson, Tim Gruene, George Sheldrick and many more..

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 People (core and associated developers)



Martyn Winn



Charles Ballard



Eugene Krissinel



Ronan Keegan



David Waterman



Andrey Lebedev



Keith Wilson



Eleanor Dodson



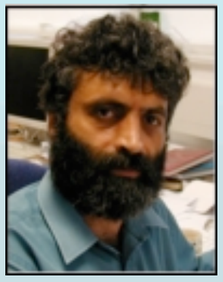
Liz Potterton



Stuart McNicholas



Kevin Cowtan



Garib Murshudov



Alexei Vagin



Fei Long



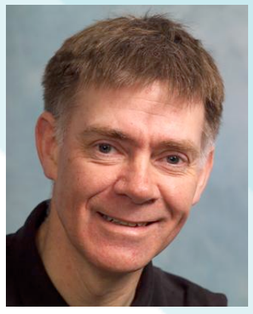
Phil Evans



Andrew Leslie



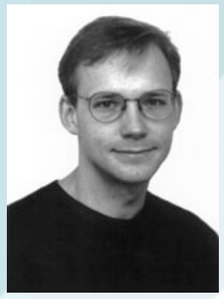
Harry Powell



Owen Johnson



Randy Reed



Martin Noble



Paul Emsley



Raj Pannu



Chris Morris



Graeme Winter