CCP4 Workshop at DLS, Harwell, UK, Dec 13–20, 2016

CCP4 Roadmap

Eugene Krissinel
(with slides from Ronan Keegan)

CCP4, Research Complex at Harwell
Didcot, Oxfordshire, UK

eugene.krissinel@stfc.ac.uk

Science & Technology Facilities Council
Research Complex at Harwell

CCP4 History and Background

- 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 is not a single group in RCaH-Rutherford but a world-wide collaboration of developers and users
- 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
- CCP4 was originally supported by the UK Science and Engineering Research Council (SERC), and is now supported by the Biotechnology and Biological Sciences Research Council (BBSRC).
- The project is coordinated at Computational Science and Engineering Department of STFC

Science & Technology Facilities Council

Research Complex at Harwell

Main CCP4 Groups

- Core Group (Oxford)
  - ~ maintain and support software suite
  - ~ application and infrastructural software developments
  - ~ collaborate with Diamond on beam line deployment of CCP4 software and joint developments
  - ~ educational outreach
  - ~ maintain CCP4 resources such as the CCP4 bulletin board
- MRC/LMB (Cambridge)
  - ~ data processing software (Mosflm, Aimless)
  - ~ refinement software (Refmac)
  - ~ model building (Coot)
- Phaser group (University of Cambridge)
- University of York
  - ~ Software developments (CCP4mg, Buccaneer, Privateer)
  - ~ CCP4 GUI-2 developments
- Others
  - ~ Experimental Phasing from Leiden (Crank-2)
  - ~ Ab-initio MR, University of Liverpool (AMPLE)
- Associated projects
  - ~ Arp/wArp (EMBL-Hamburg)
  - ~ PDB Redo (NKI, the Netherlands)

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Users Geography

Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Study Weekend

- Annual CCP4 users meeting taking place in 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
- Proceedings published in special edition of Acta Cryst. D
- 2016: *"Protein-Ligand Complexes: Understanding Biological Chemmistry"*
- *2017: "From crystals to structure with CCP4"*
- *2018: "Serial crystallography"*

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Schools and Workshops

- Main aim is to give education and expert help to researchers working with collected crystal data
- Some schools also provide expert help with collecting crystal data (APS, Hamburg, Diamond)
- Applicants with challenging crystals and/or data are given strong consideration but this is not mandatory
- Software developers from CCP4, Phenix (APS Workshop), SHELX, Arp/wArp, XDS and HKL as well as others give lectures, tutorials and are also available to help with data processing and data solution throughout the meetings

http://www.ccp4.ac.uk – Courses & Events

Science & Technology Facilities Council

Research Complex at Harwell

Public CCP4 Resources

- CCP4BB mailing list — http://www.jiscmail.ac.uk/lists/ccp4bb.html — 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 web site — *http://www.ccp4.ac.uk* — documentation, developers info, news, links to the above resources and CCP4-online services for structure solution.

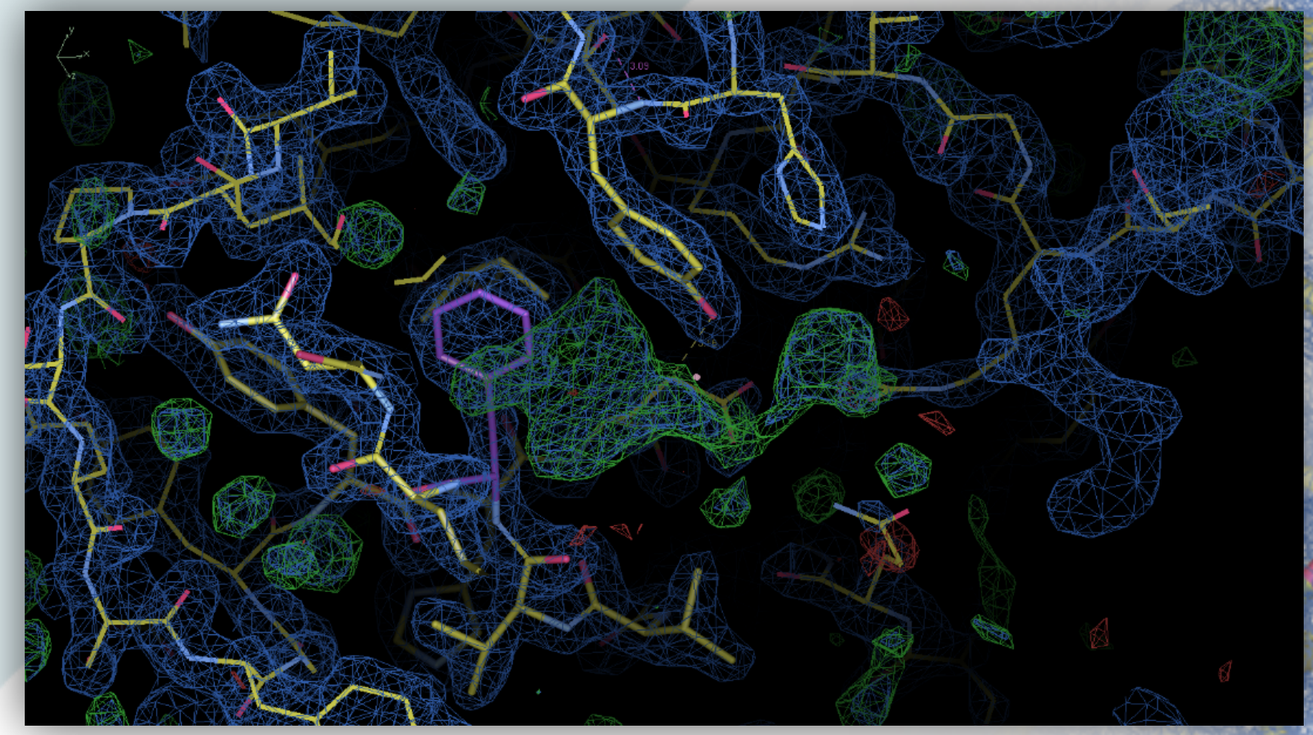
Science & Technology Facilities Council
Research Complex at Harwell

The CCP4 Software Suite

- 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 (CCP4 7.0) are made available through the CCP4 updates manager

Structure Solution in Essence

- X-ray diffraction as such produces a reciprocal image of a target structure, or Fourier transform of it's atomic coordinates. As an ultimate goal, one would like to find an atomic model for their target from X-ray diffraction images and other information available (such as sequence)
- 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 recorded in an experiment
- In reality, the X-ray diffraction experiment produces 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 need to be interpreted via model building and refinement to produce a model for target structure
- The model need to be analysed to answer biochemical questions and deposited in the PDB

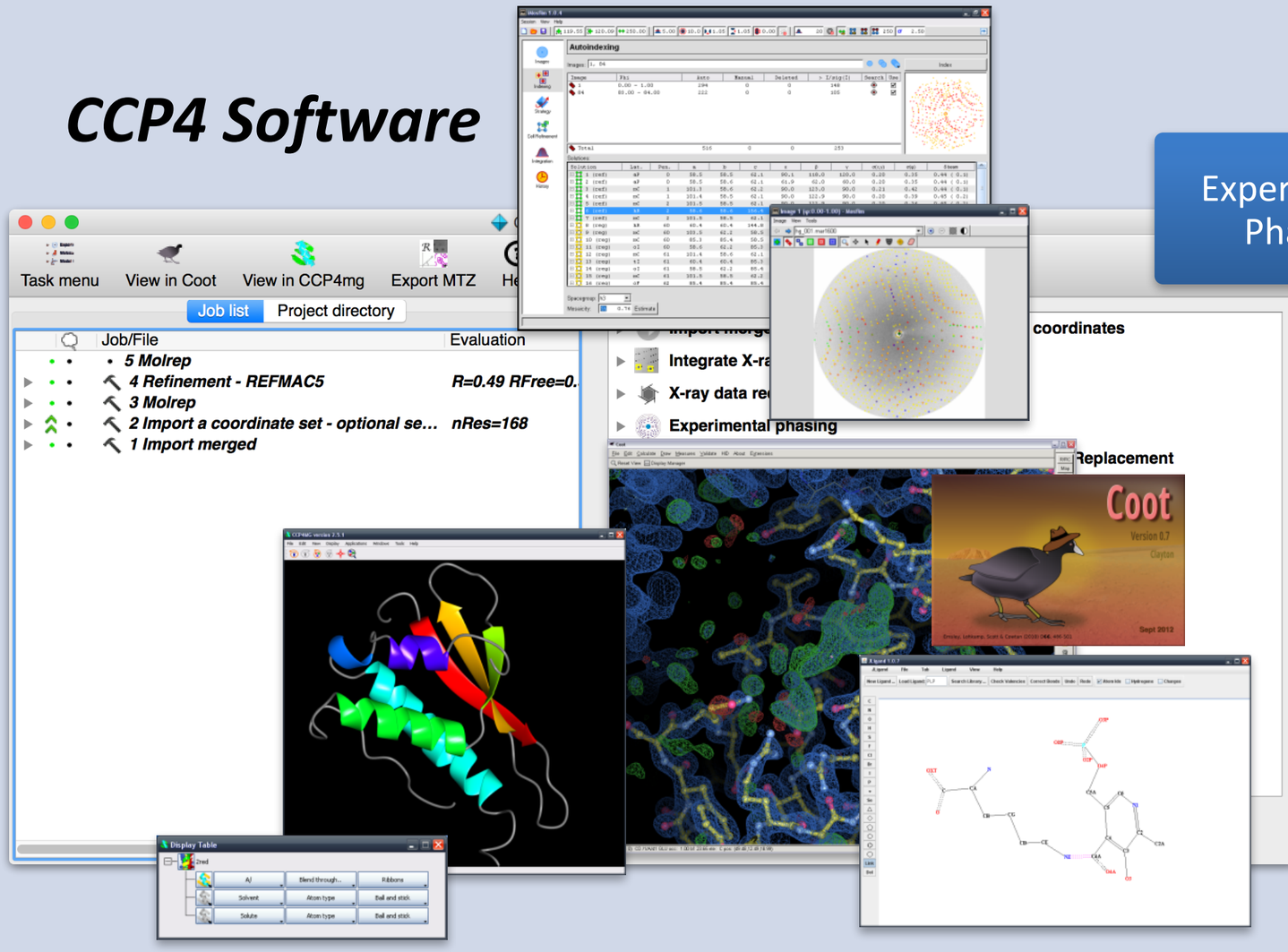
sounds like we know what to do …

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Scope

CCP4 Software



PiMS

Crystallisation

Data Collection

Data Processing and Reduction

DIALS

Experimental Phasing

Molecular Replacement

Density Modification

SHELX

Model Building

ARP wARP 7.3

Refinement

Structure Analysis

PDB_REDO

Deposition

CCP4 Roadmap

Science & Technology Facilities Council

Research Complex at Harwell

External components

SHARP/ autoSHARP

XDS & autoPROC

PiMS

Crystallisation

Data Collection

Data Processing and Reduction

DIALS

Experimental Phasing

Molecular Replacement

Density Modification

Model Building

Refinement

Structure Analysis

Deposition

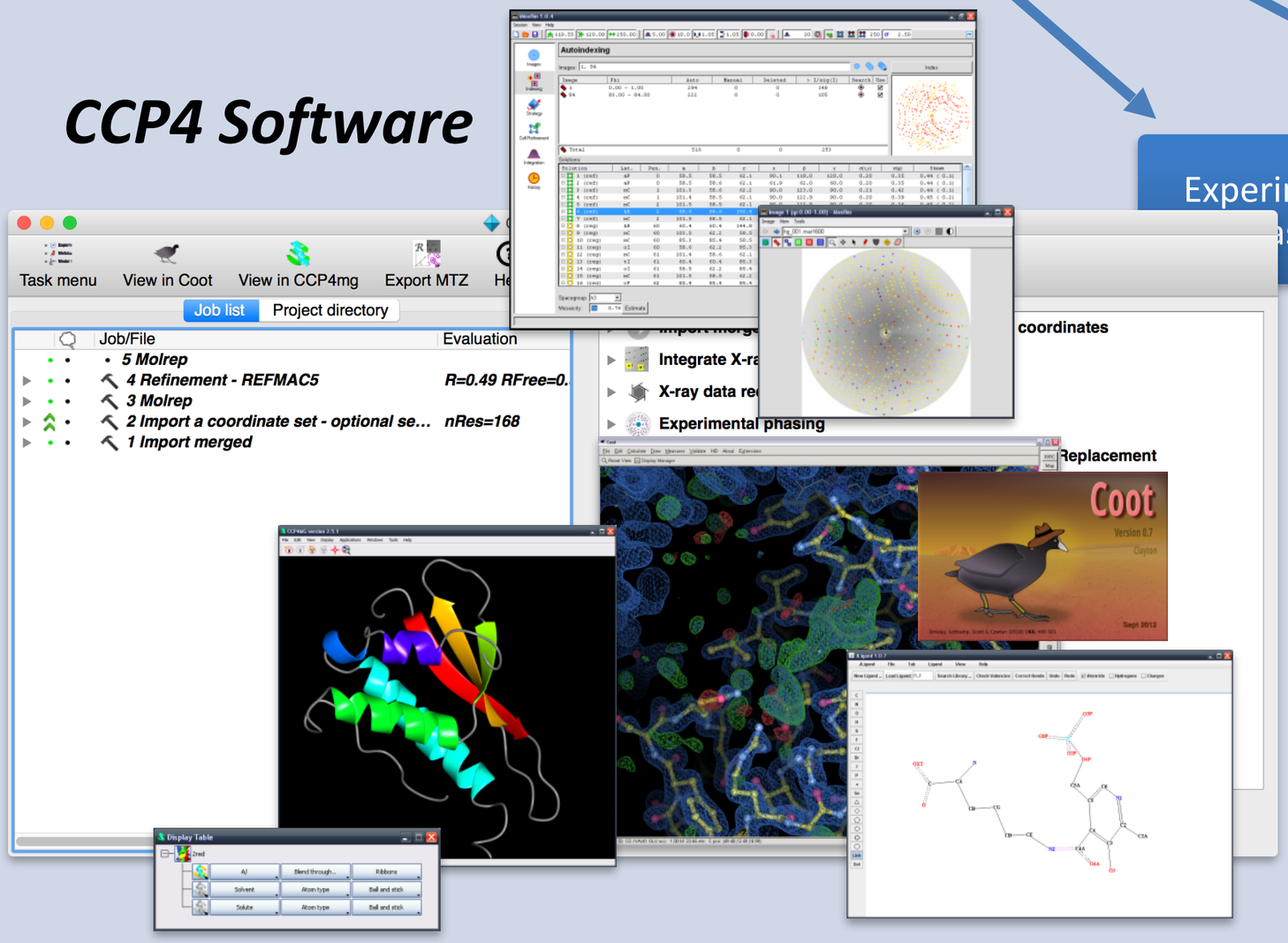
BUSTER

SHELX

7.3 ARP wARP

PDB_REDO

CCP4 Software



Science & Technology Facilities Council

Research Complex at Harwell

The CCP4i2 Interface

Preferences and Settings

Managing Software Updates

Project Management

ccp4i2 File Edit History Utilities Projects Help

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
5 Molrep	
4 Refinement - REFMAC5	*R=0.49 RFree=0.*
3 Molrep	
2 Import a coordinate set - optional se…	*nRes=168*
1 Import merged	

Job list (Project)

- Import merged data, sequences, alignments or coordinates
- Integrate X-ray images
- X-ray data reduction and analysis
- Experimental phasing
- Bioinformatics and model preparation for Molecular Replacement
- Molecular Replacement
- Model building and Graphics
- Refinement
- Ligands
- Validation and analysis
- Export and Deposition
- Reflection data tools
- Coordinate data tools
- Developer tools

Task menu

New job | Cancel

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 Refinement - REFMAC5	*R=0.49 RFre*
3 Molrep	
2 Import a coor... se...	*nRes=168*
1 Import merge	

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

Input
Output
Comments

Job actions (right-click)

Job report

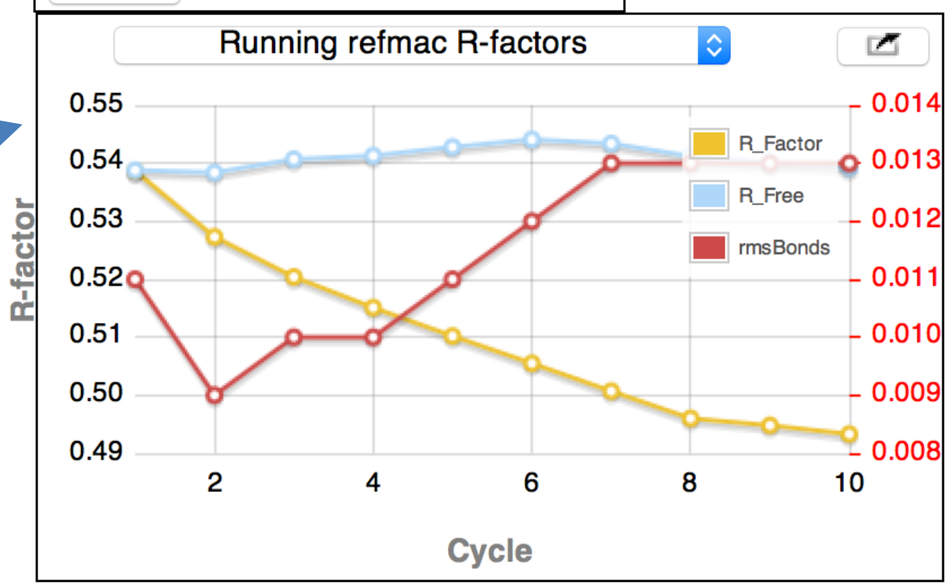
Job 4: Refinement - REFMAC5 ***The job is Finished***

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

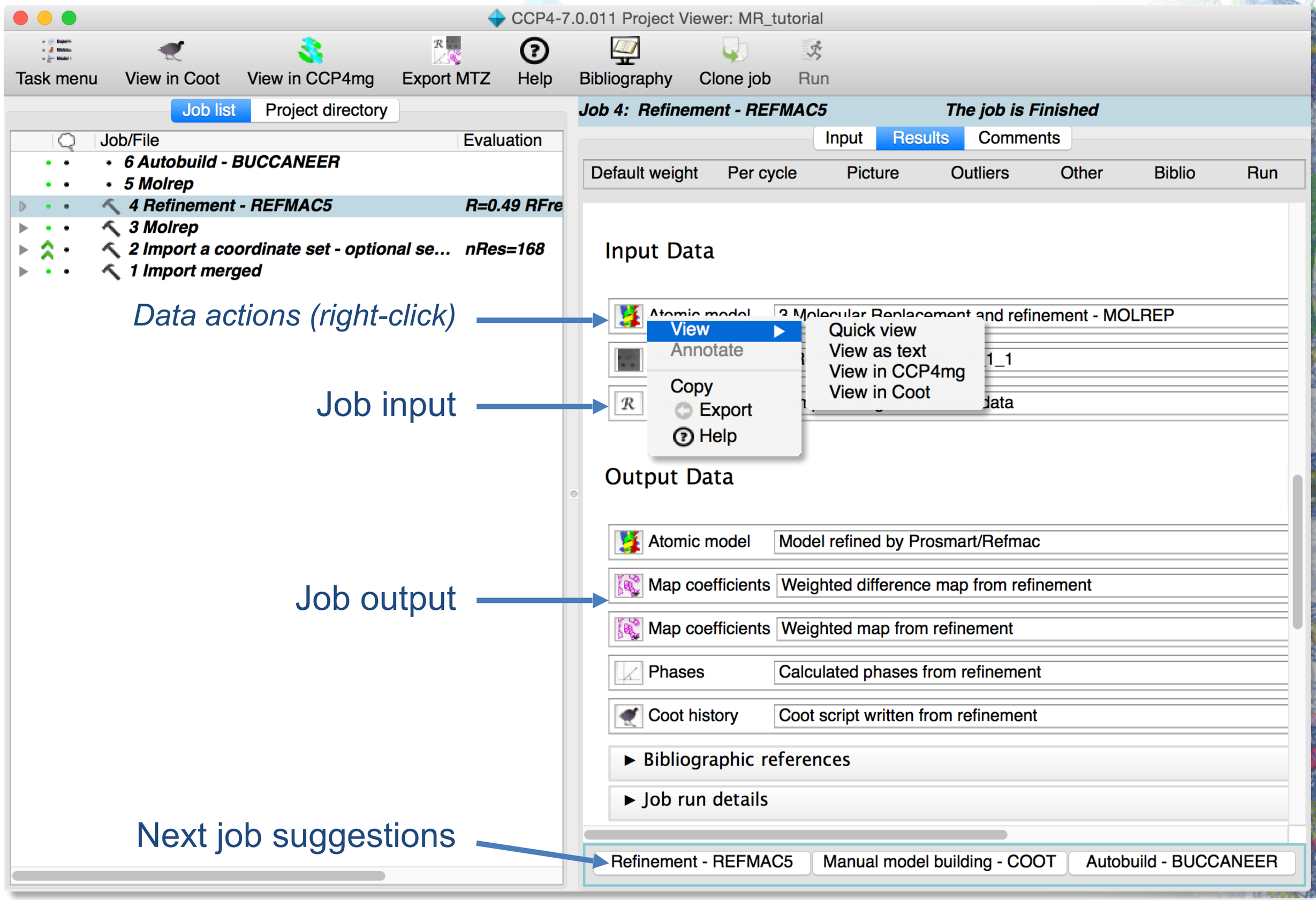


Refinement - REFMAC5 | Manual model building - COOT | Autobuild - BUCCANEER

Science & Technology Facilities Council

Research Complex at Harwell

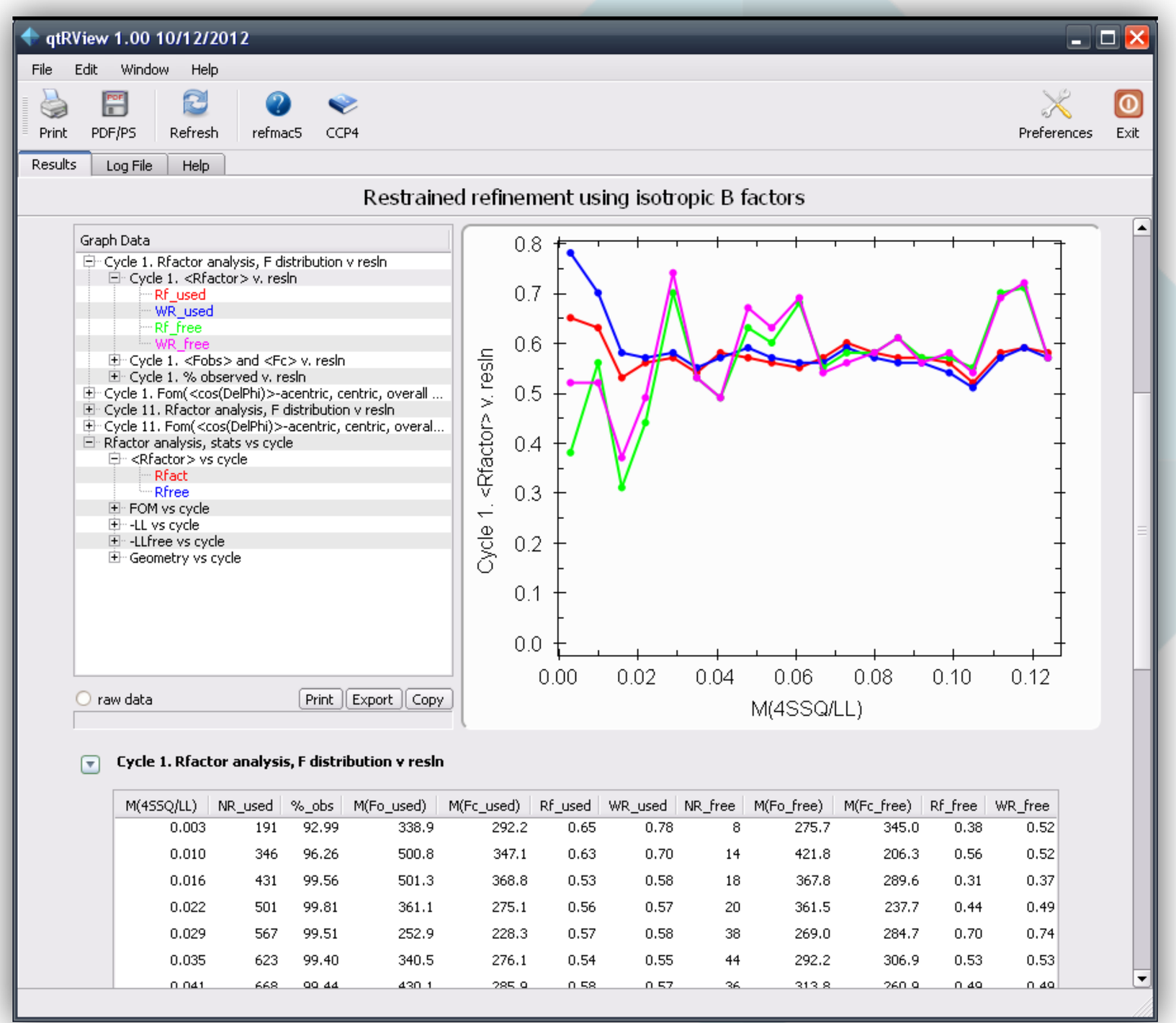
The CCP4i2 Interface



Science & Technology Facilities Council

Research Complex at Harwell

Report and Log File Viewer QtRView



M(4SSQ/LL)	NR_used	%_obs	M(Fo_used)	M(Fc_used)	Rf_used	WR_used	NR_free	M(Fo_free)	M(Fc_free)	Rf_free	WR_free
0.003	191	92.99	338.9	292.2	0.65	0.78	8	275.7	345.0	0.38	0.52
0.010	346	96.26	500.8	347.1	0.63	0.70	14	421.8	206.3	0.56	0.52
0.016	431	99.56	501.3	368.8	0.53	0.58	18	367.8	289.6	0.31	0.37
0.022	501	99.81	361.1	275.1	0.56	0.57	20	361.5	237.7	0.44	0.49
0.029	567	99.51	252.9	228.3	0.57	0.58	38	269.0	284.7	0.70	0.74
0.035	623	99.40	340.5	276.1	0.54	0.55	44	292.2	306.9	0.53	0.53
0.041	668	99.44	430.1	285.9	0.58	0.57	36	313.8	260.9	0.49	0.49

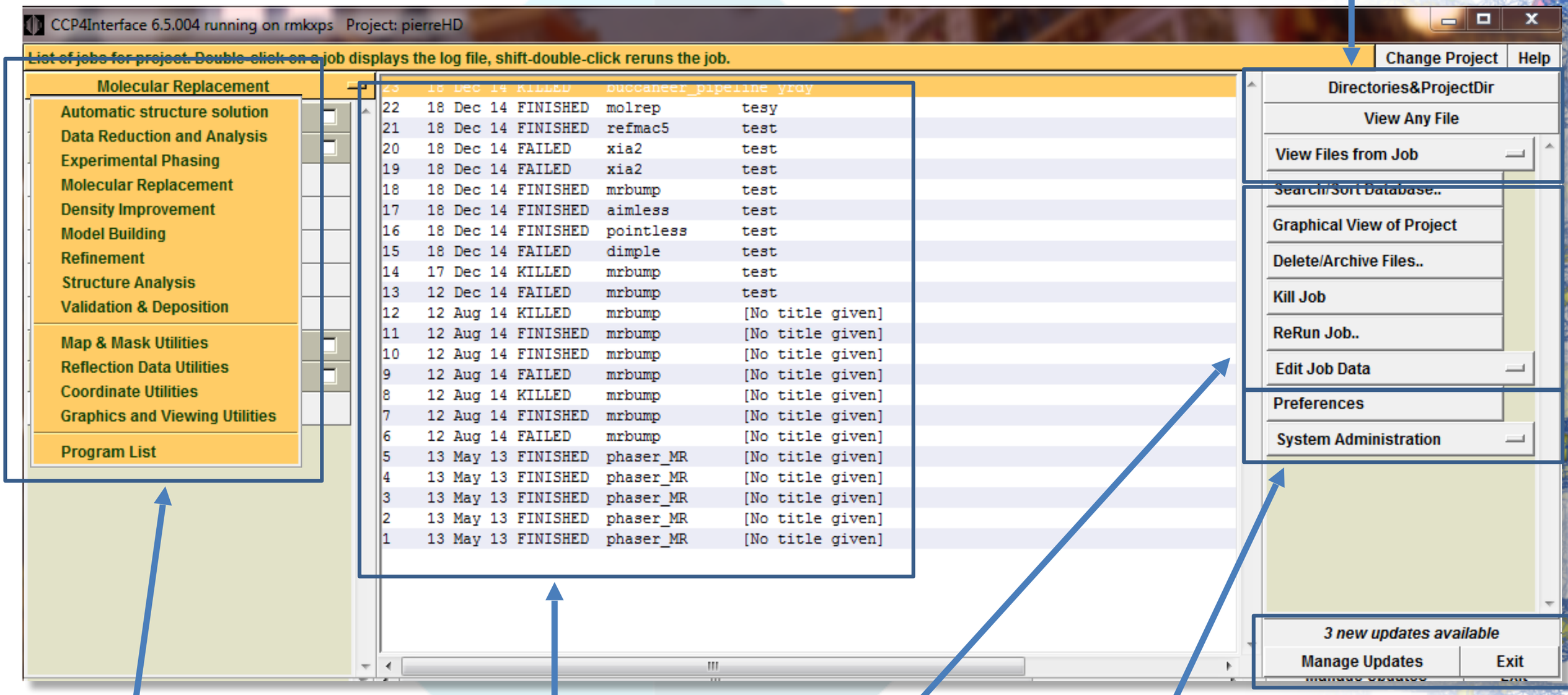
- Graphical log file viewer
- View graphs, tables and job summary information
- Launch Coot, CCP4mg and ViewHKL to work with the content of data files
- Export result files
- Can still view the original (plain text) log file

Science & Technology Facilities Council

Research Complex at Harwell

The CCP4i(1) Interface

Project Management



Task/program menu

Job list

Job actions: clone, kill, delete, view files, etc.

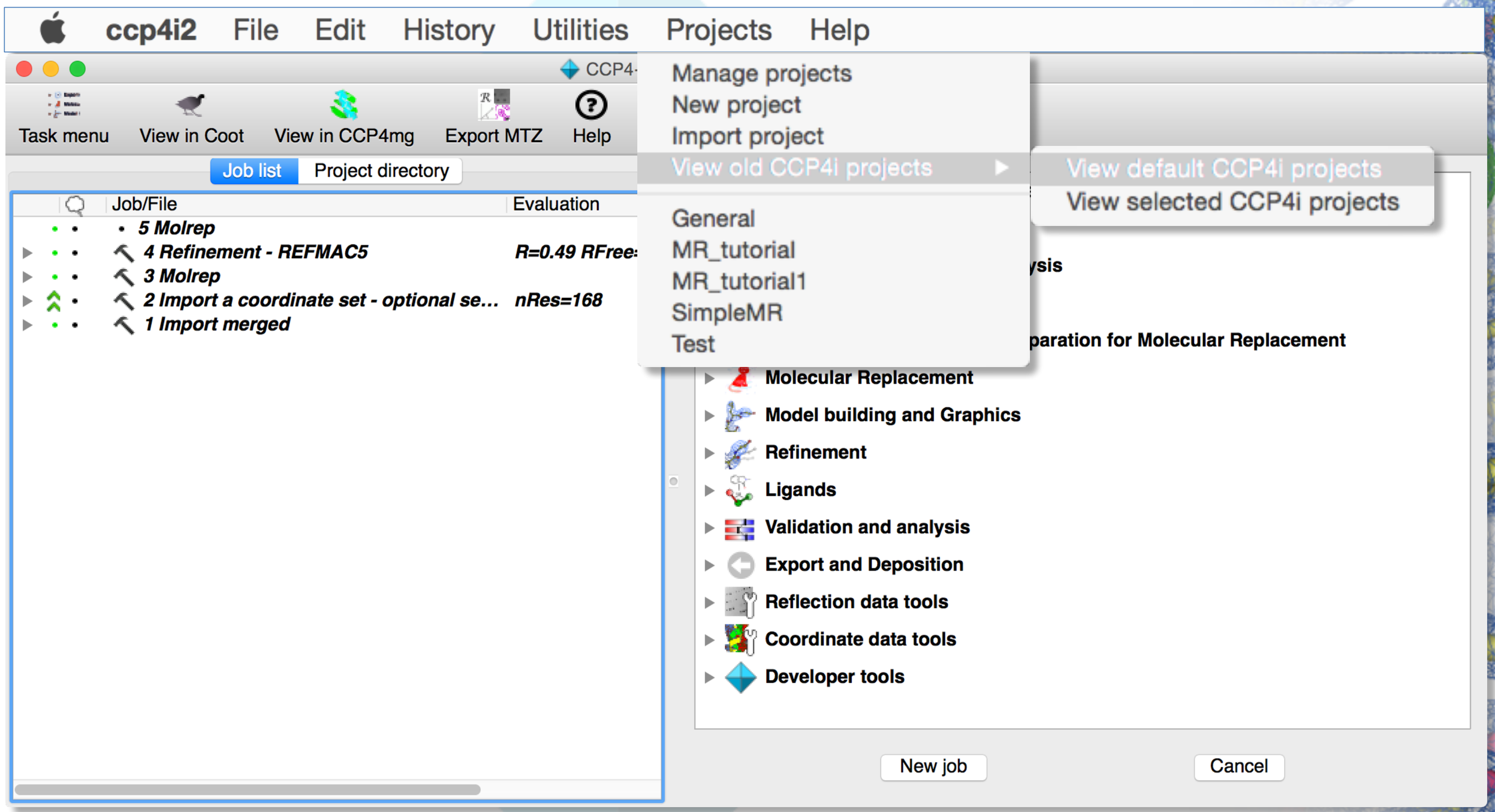
Preferences and Settings

CCP4 Updates

Science & Technology Facilities Council

Research Complex at Harwell

From CCP4i(1) to CCP4i2

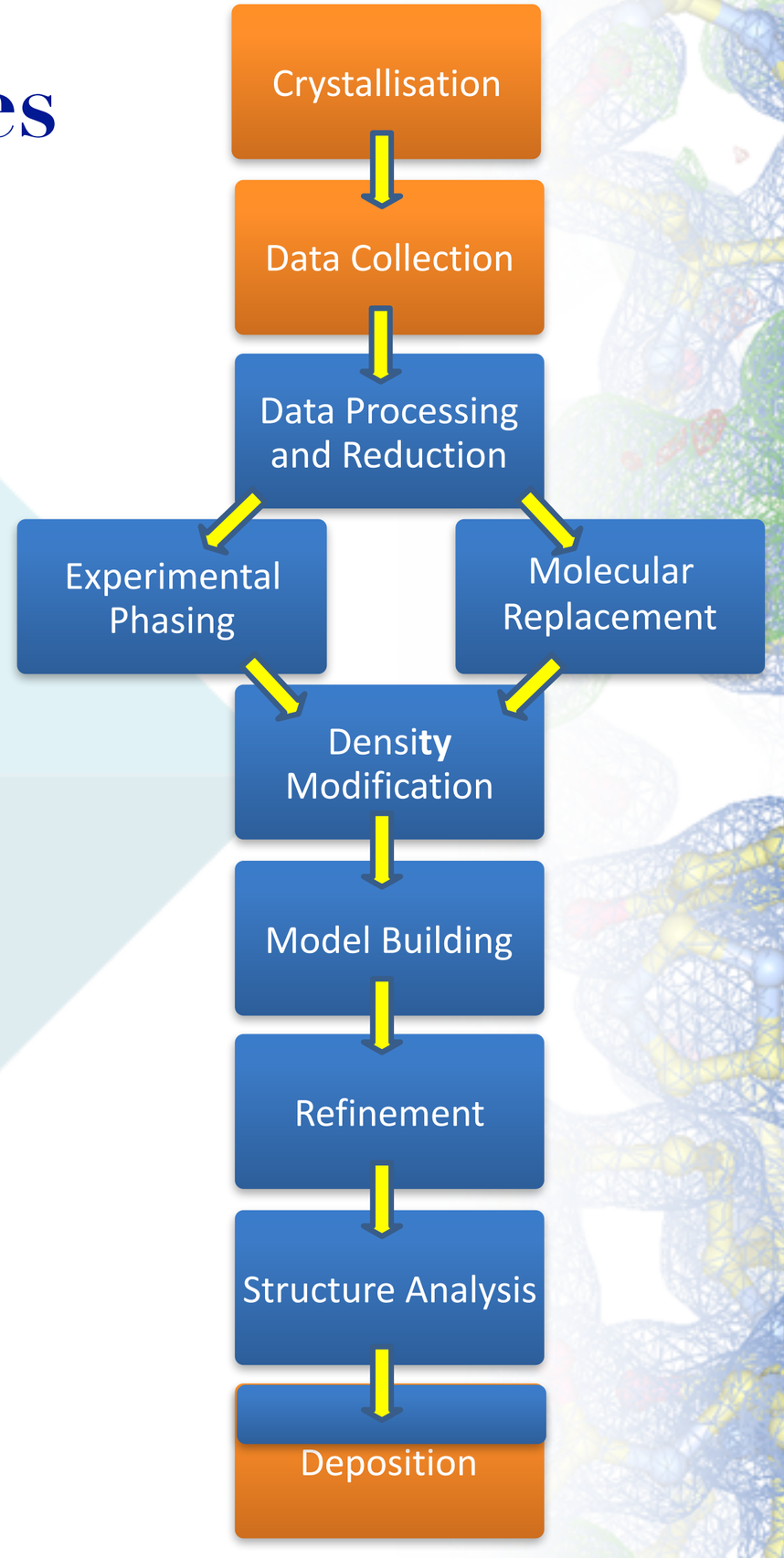


… one-way road

Science & Technology Facilities Council

Research Complex at Harwell

The Structure Solution Stages



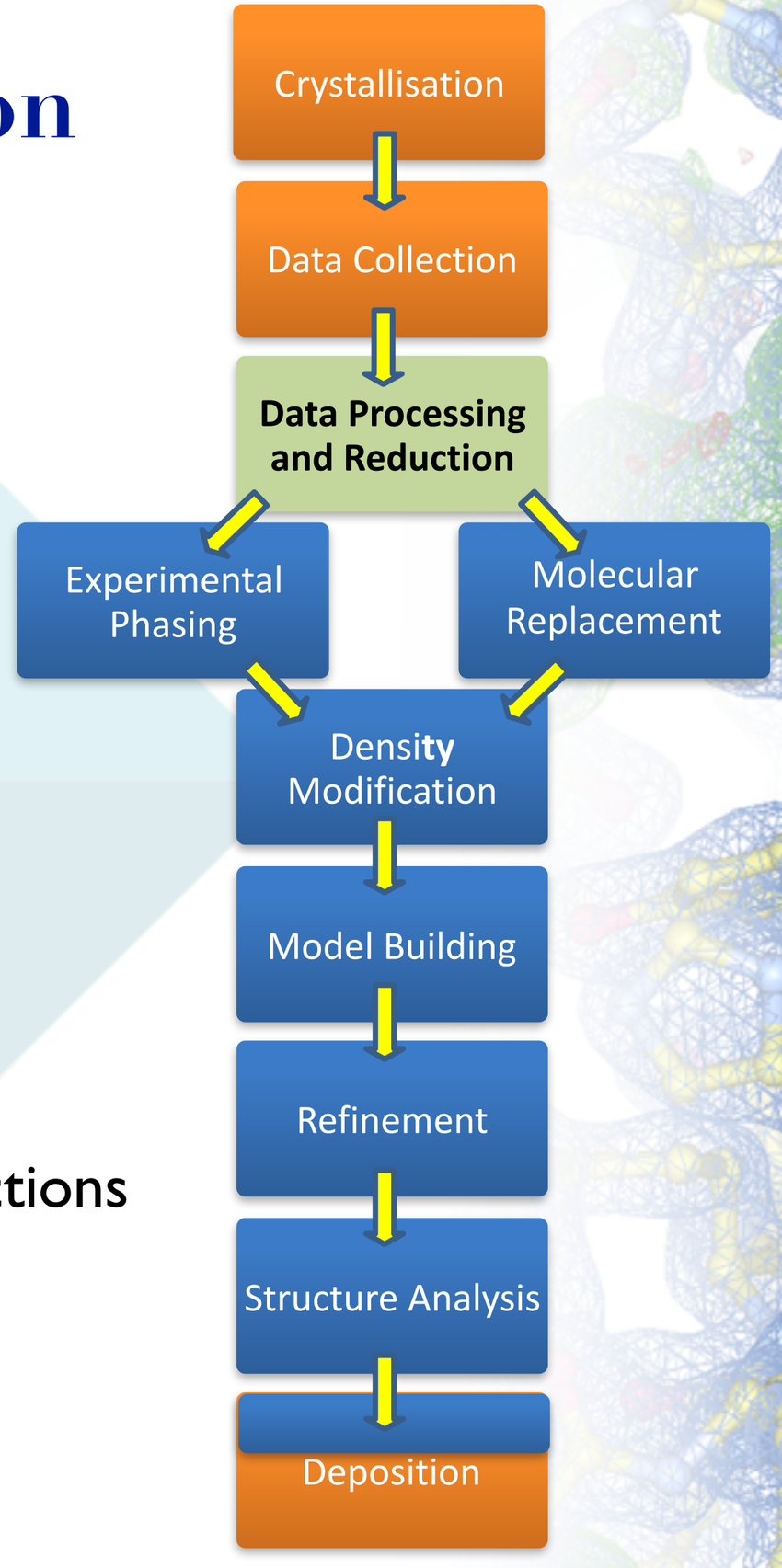
Science & Technology Facilities Council

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

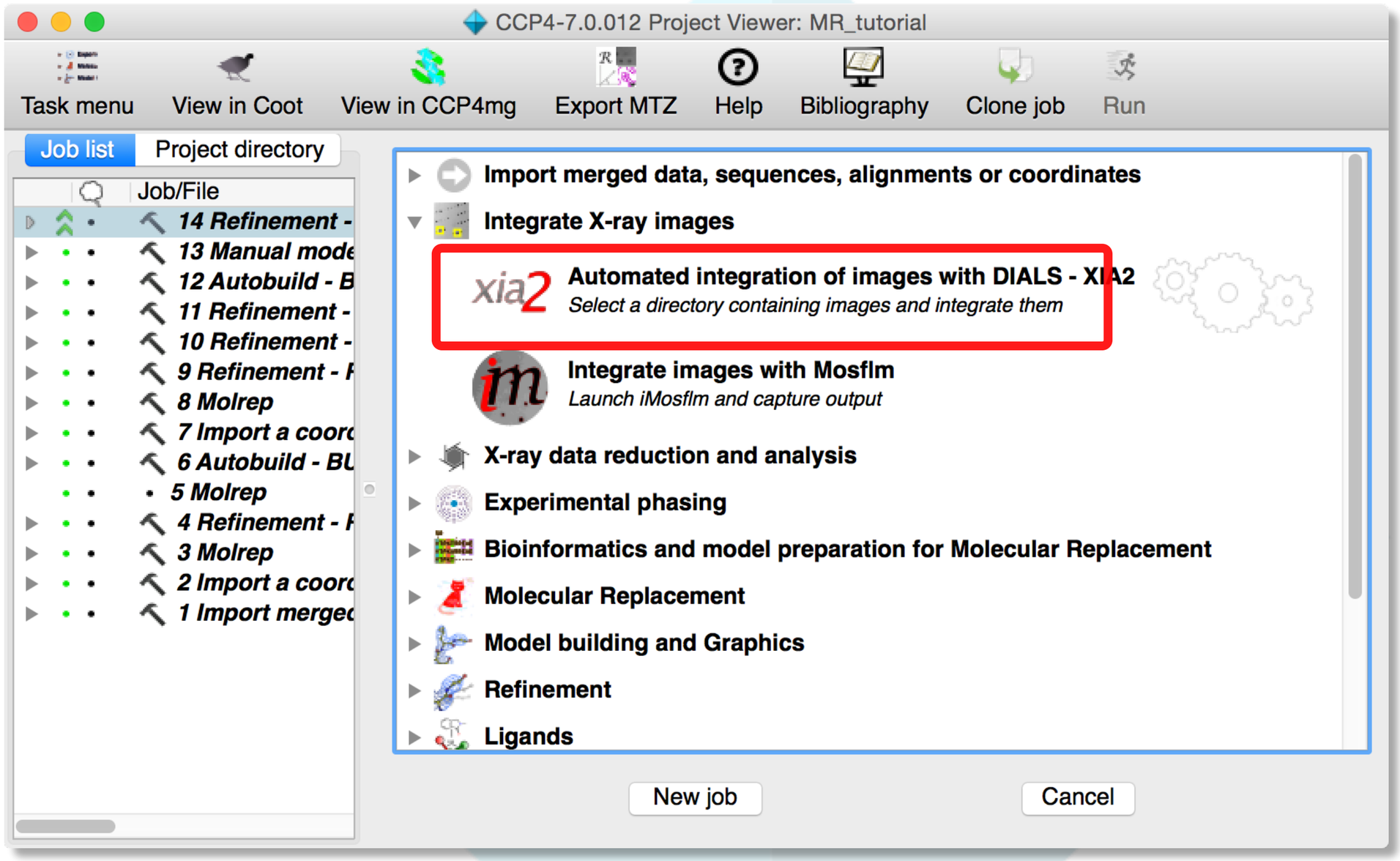
- Data Reduction: Scaling and merging
  - ~ Scale and merge multiple observations of reflections
  - ~ Assignment of point group/space group
  - ~ Look for problems in images
  - ~ Calculate structure factor amplitudes



Science & Technology Facilities Council

Research Complex at Harwell

Data Processing and Reduction



DIALS
Diffraction Integration for Advanced Light Sources

xia2

DIALS: New data processing and data reduction tool, available through Xia-2 pipeline

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

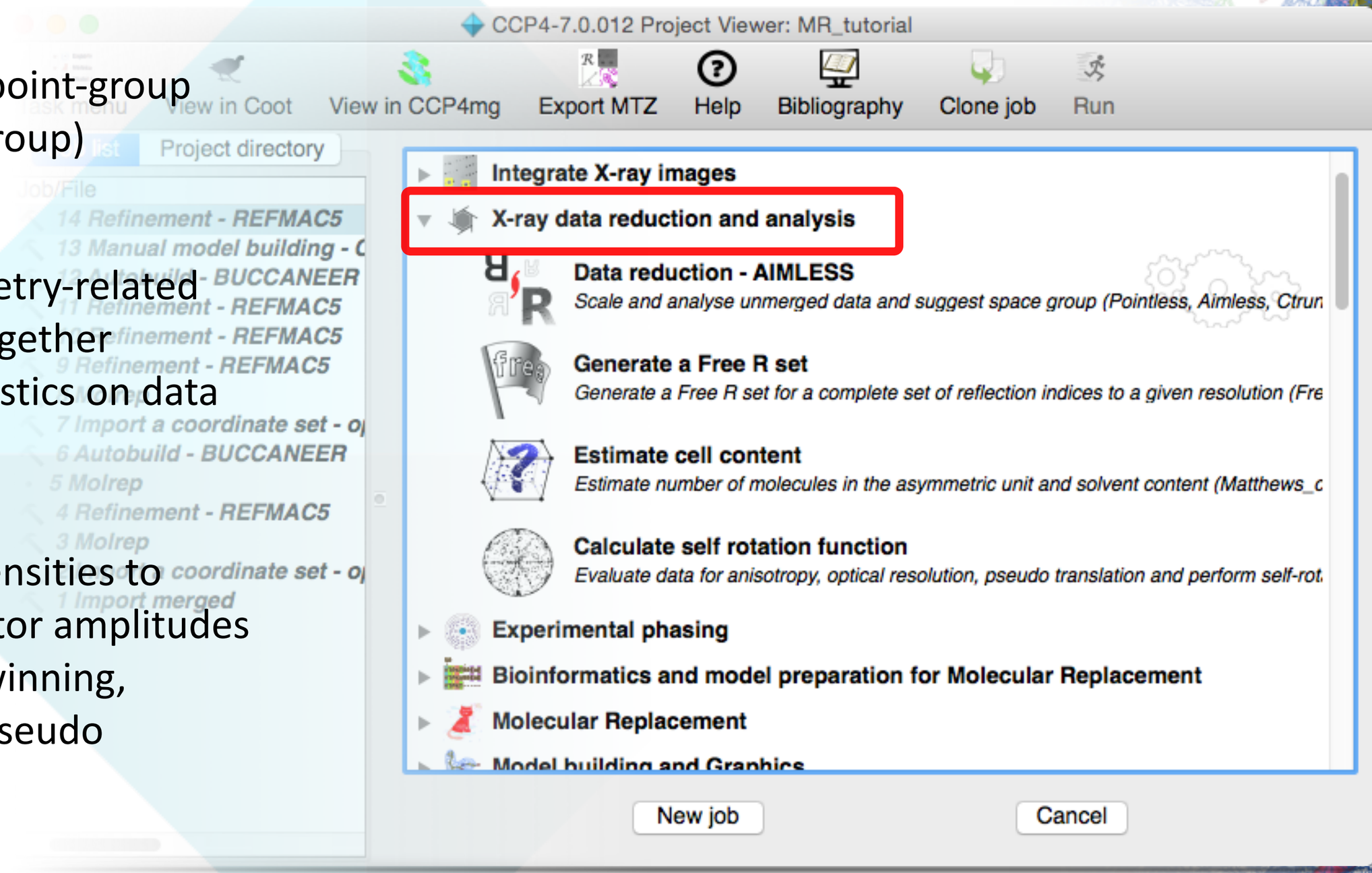
- Scales symmetry-related intensities together
- Produce statistics on data quality

Ctruncate

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

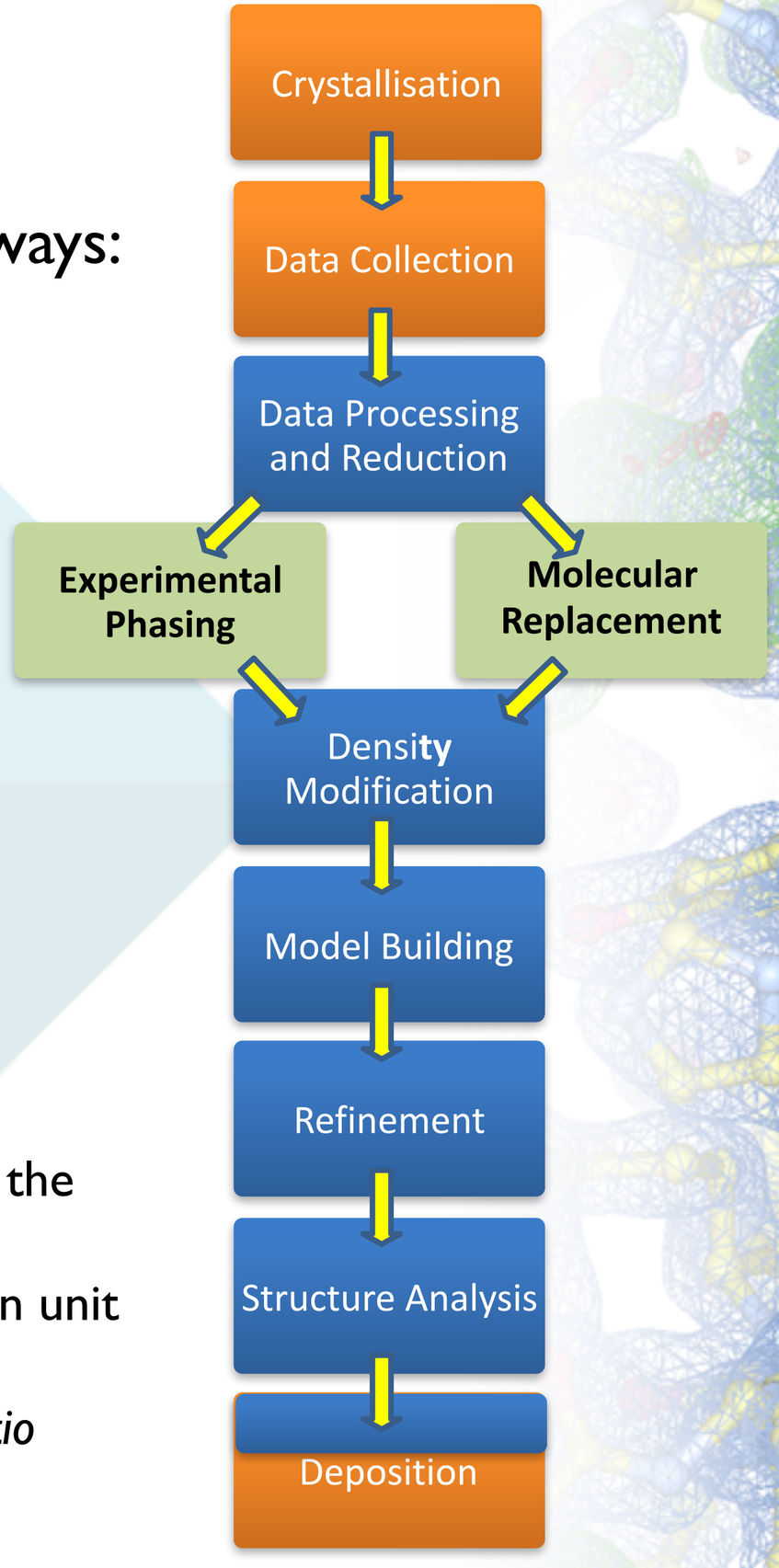
SFcheck

- Analyses data and check for problems – postscript report



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

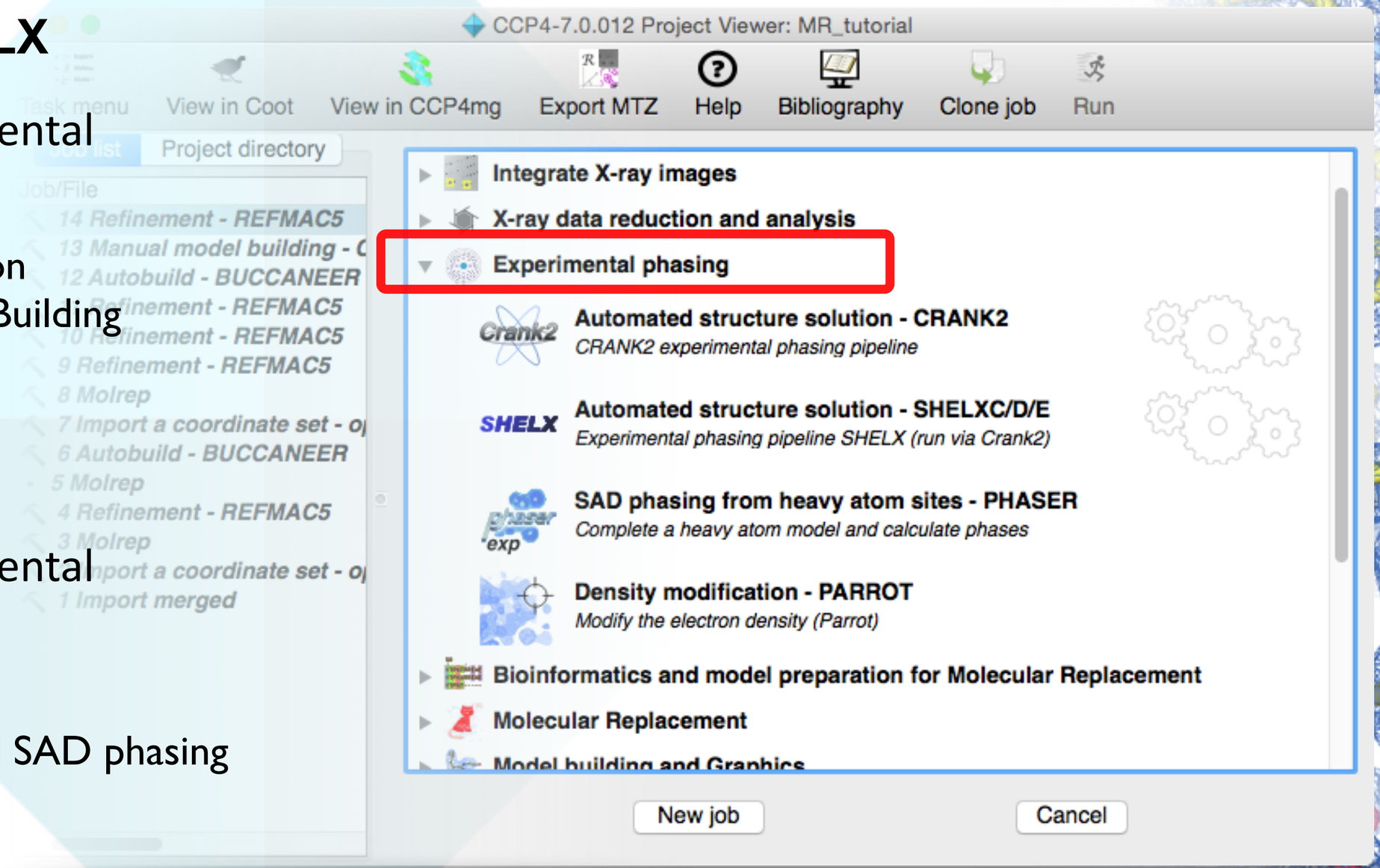
Experimental Phasing

Crank, Crank2 & SHELX

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

Phaser EP

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



Science & Technology Facilities Council

Research Complex at Harwell

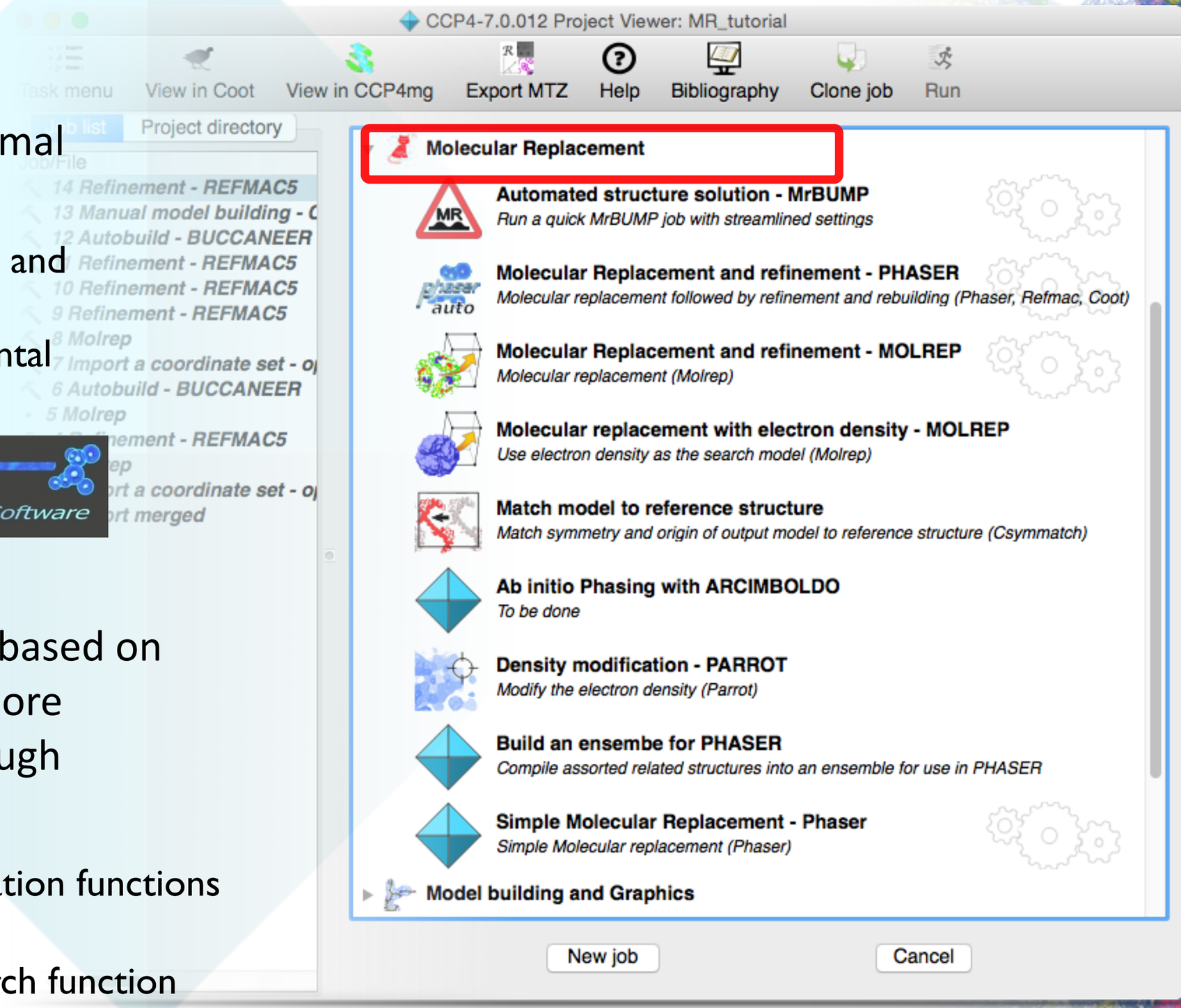
Molecular Replacement

Phaser MR

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

Molrep

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



Science & Technology Facilities Council

Research Complex at Harwell

Automatic Molecular Replacement

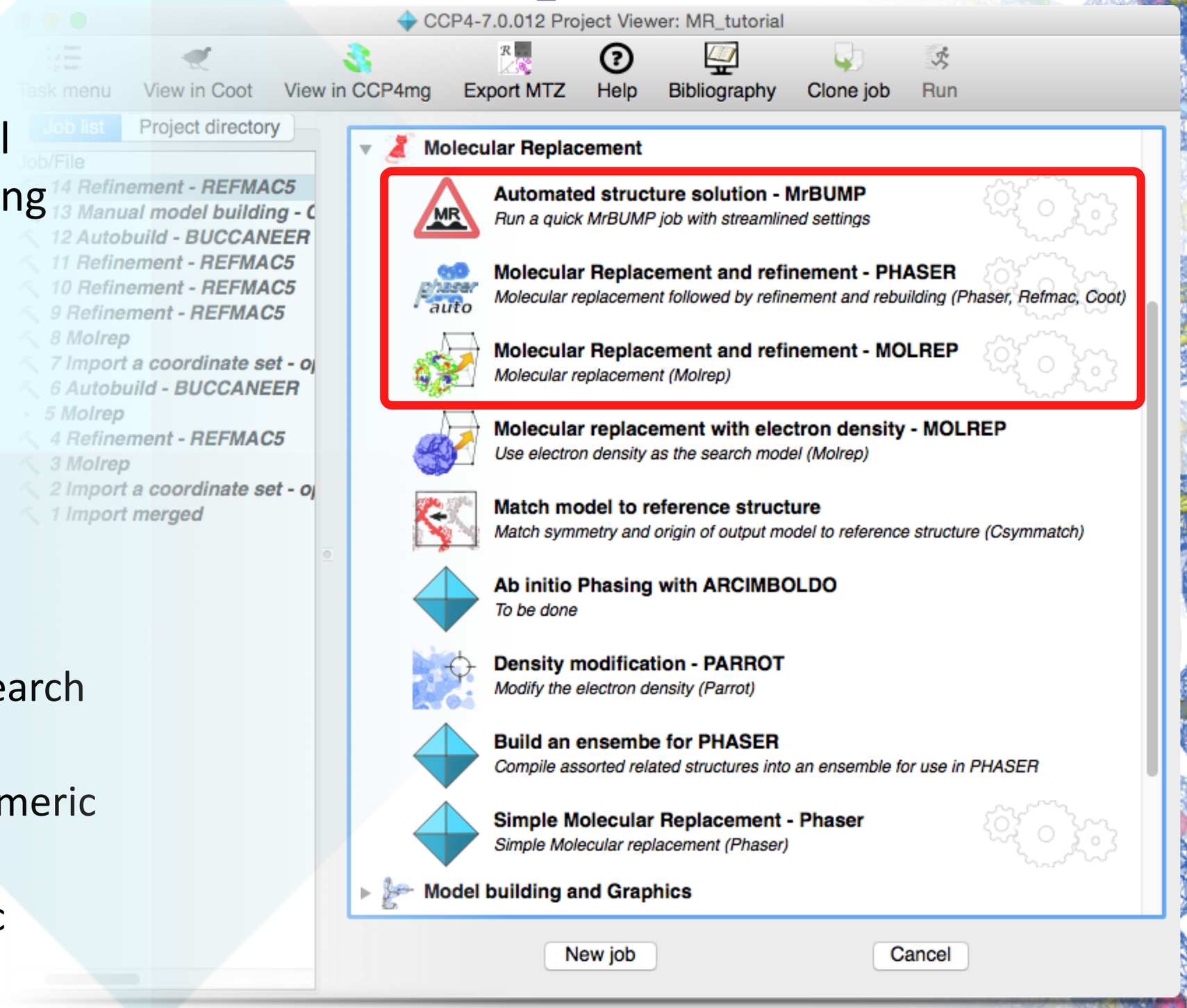
From model search and preparation through to initial refinement and model building

MrBUMP

- Online model search
- Uses Phaser and Molrep

BALBES

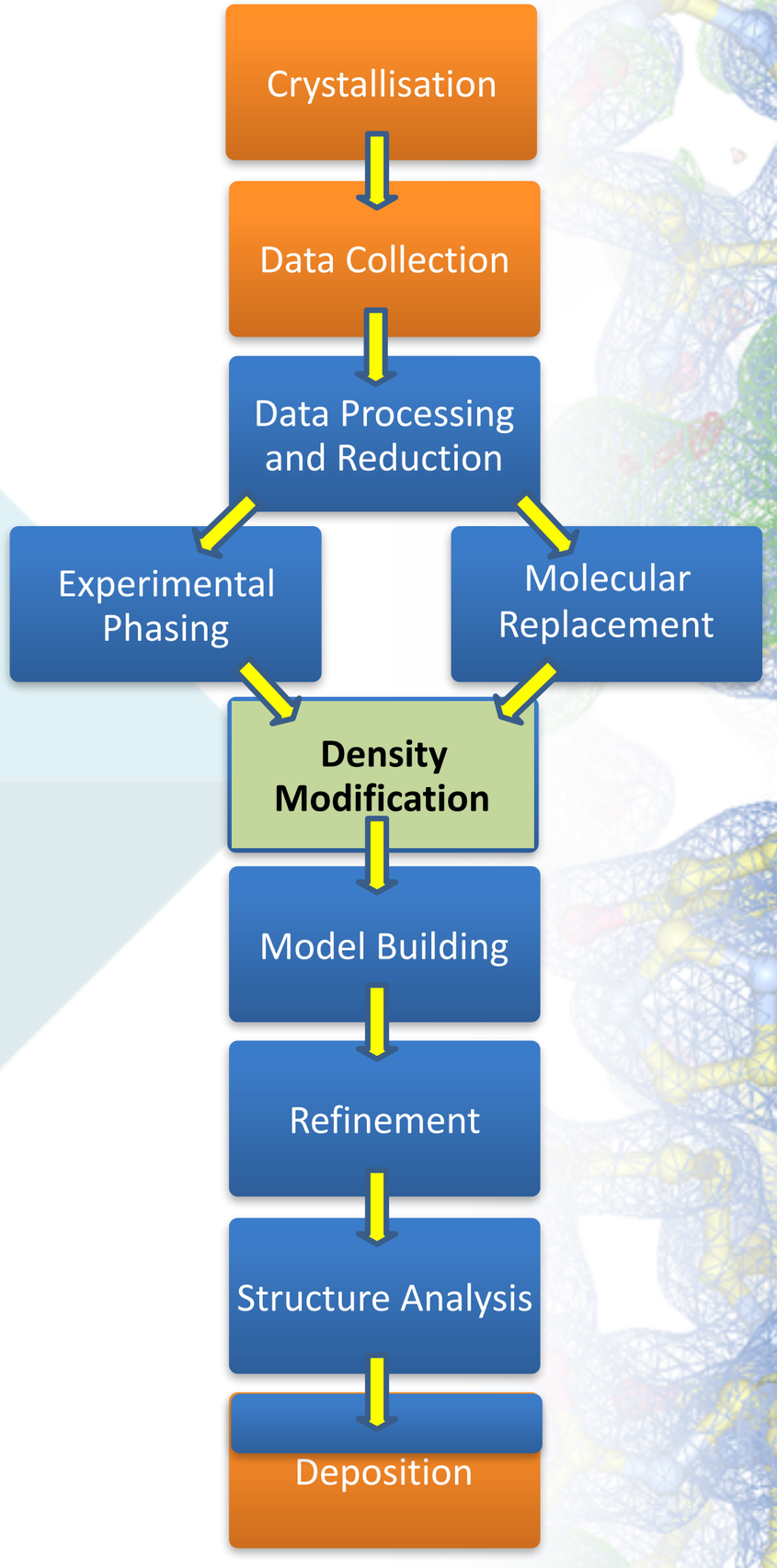
- Custom database with search models
- Capable of making multimeric models (uses PISA)
- Uses Molrep and Refmac



Science & Technology Facilities Council
Research Complex at Harwell

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

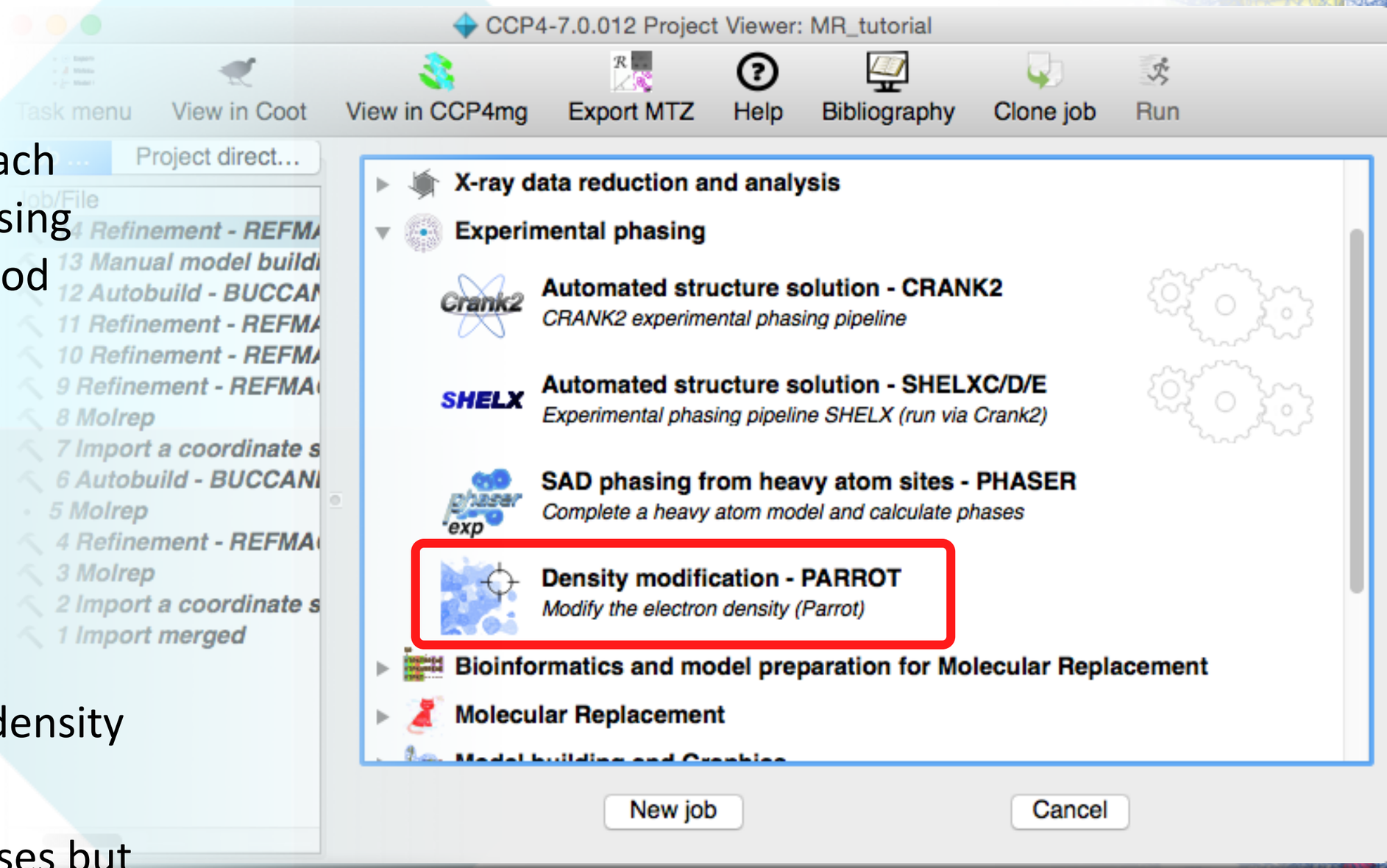
Density Modification Tools

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

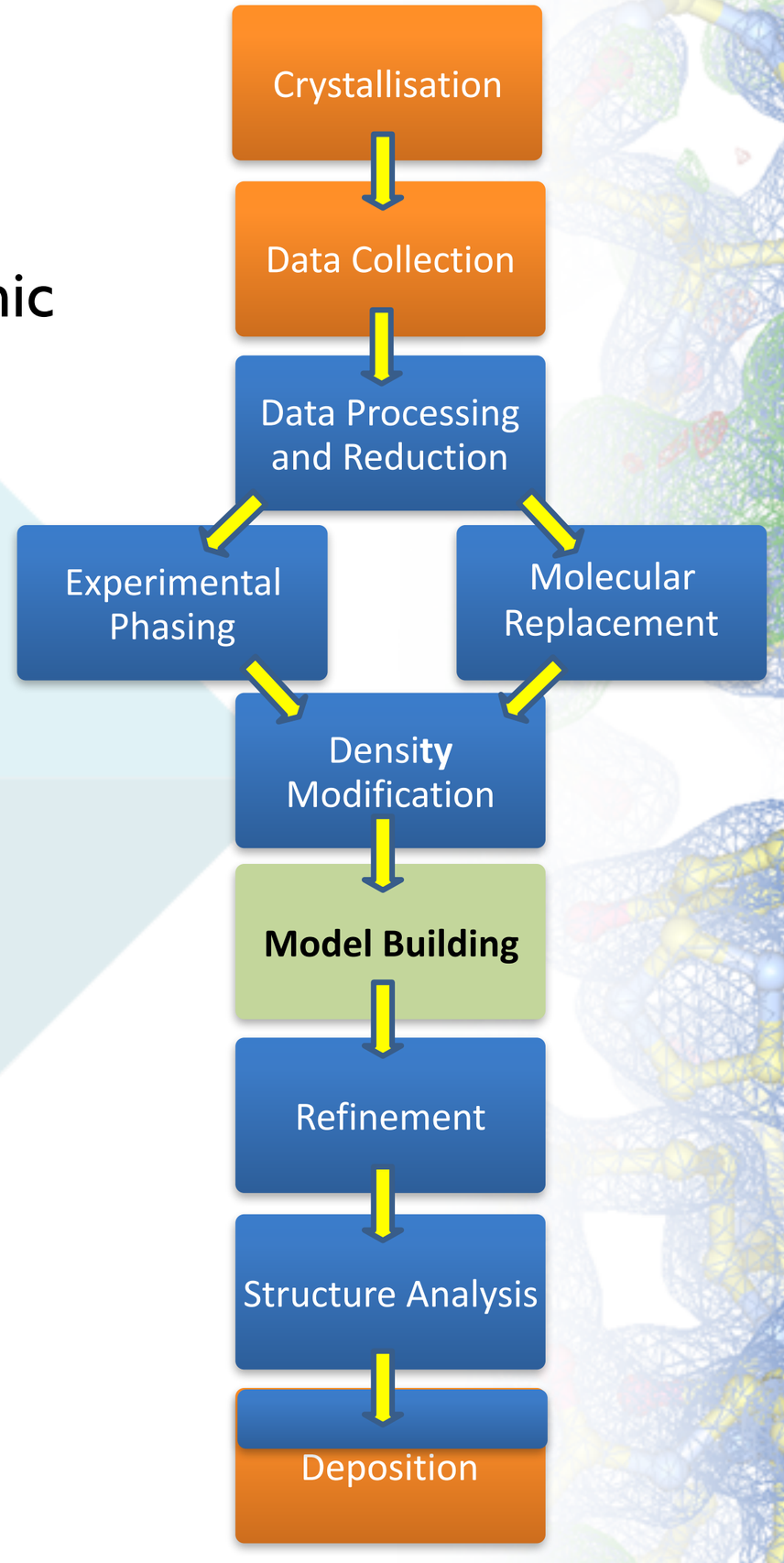


Science & Technology Facilities Council
Research Complex at Harwell

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

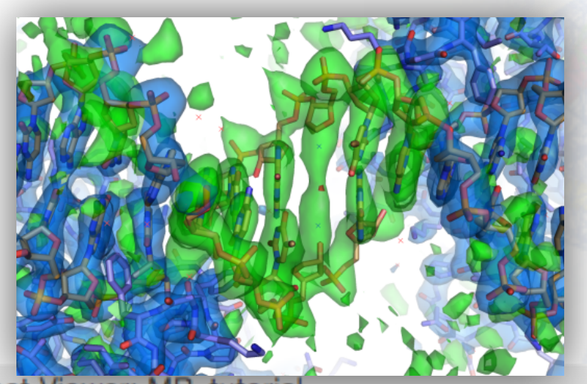
- Try to account for as much of the density as possible including solvent molecules



Science & Technology Facilities Council

Research Complex at Harwell

Model Building Tools



Buccaneer

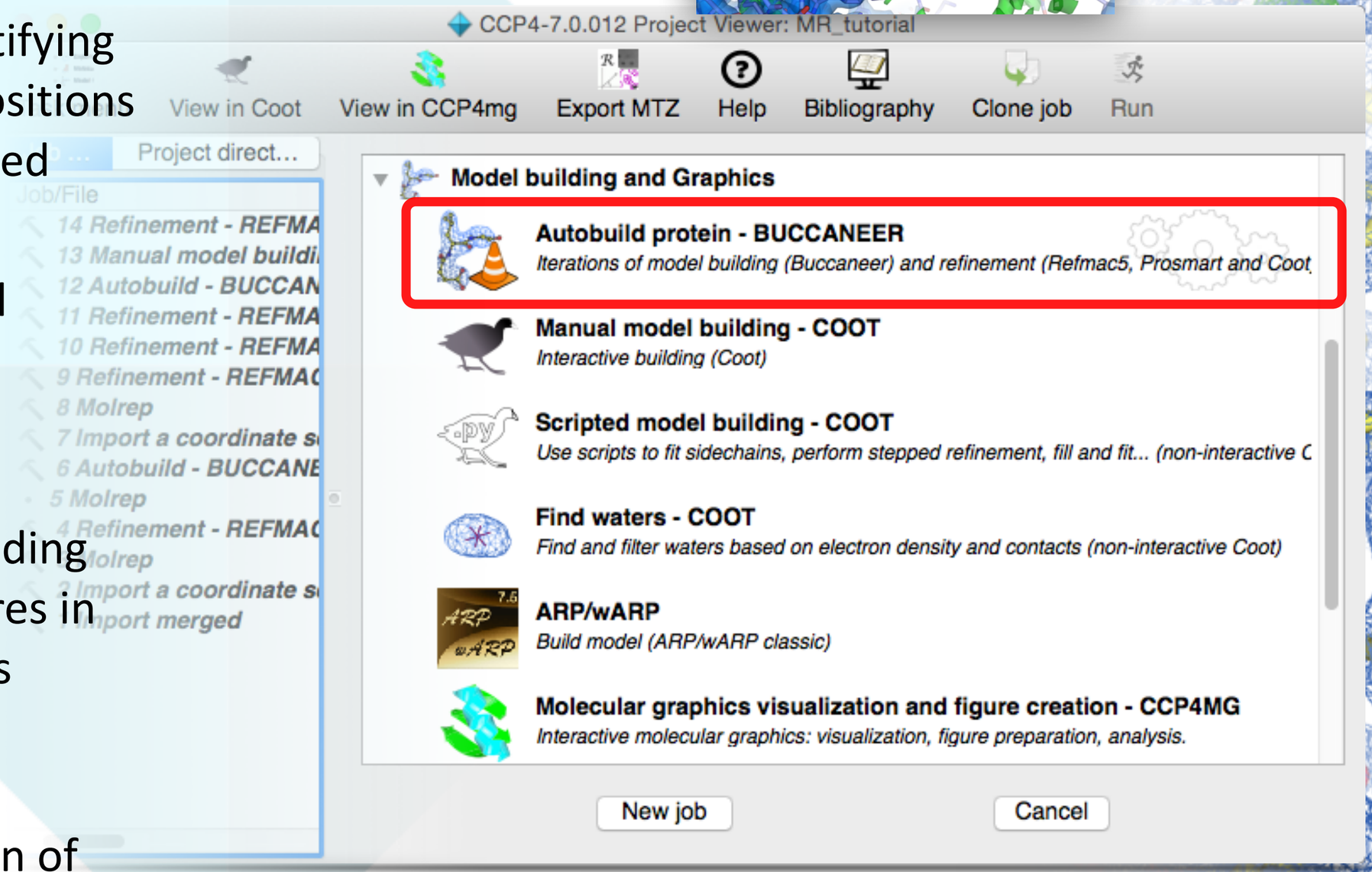
- Chain tracing by identifying connected C-alpha positions using a likelihood-based density target
- Low-resolution model building

Nautilus

- Automatic model building of nucleotide structures in electron density maps

Privateer

- Building and validation of carbohydrates



Science & Technology Facilities Council

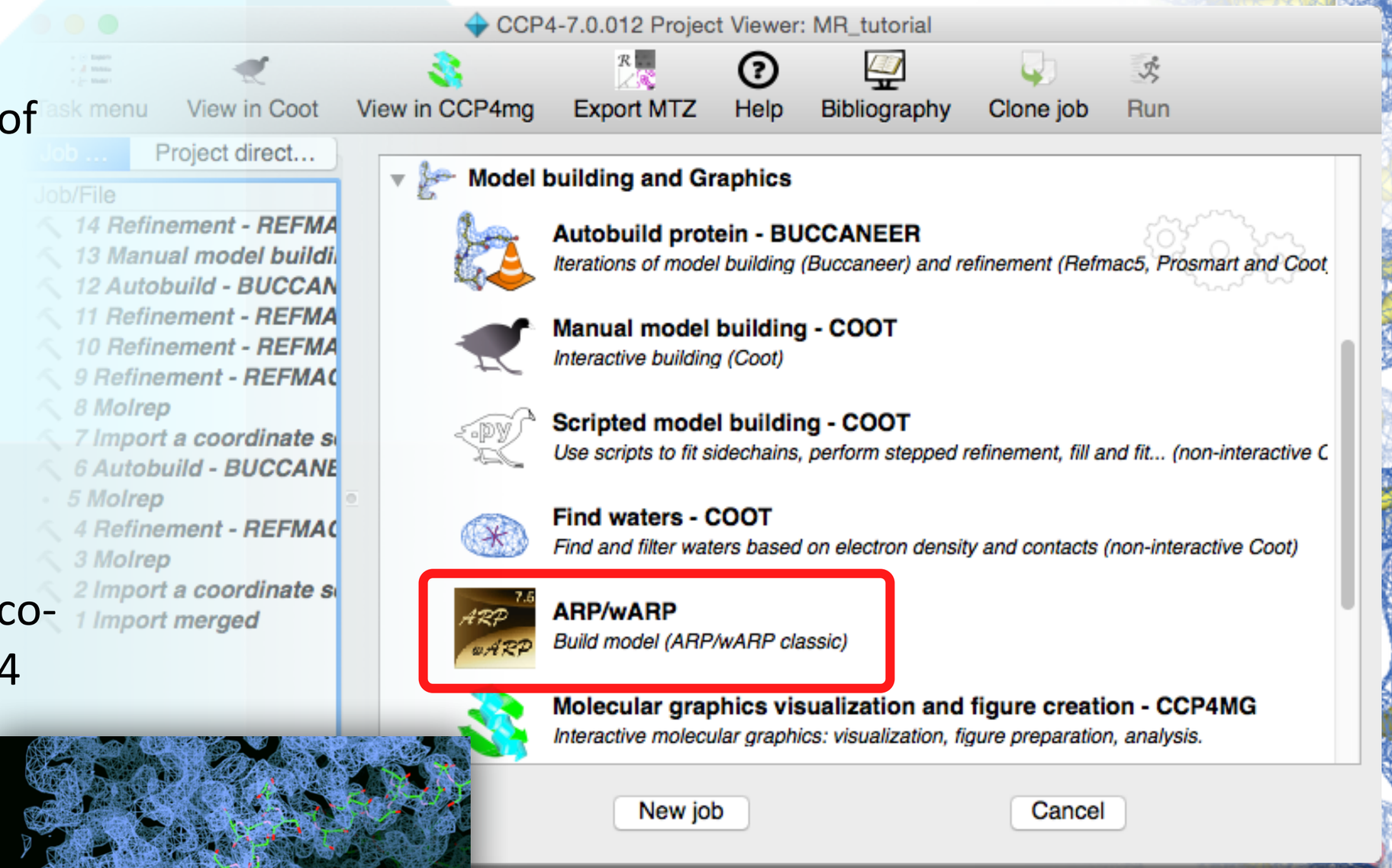
Research Complex at Harwell

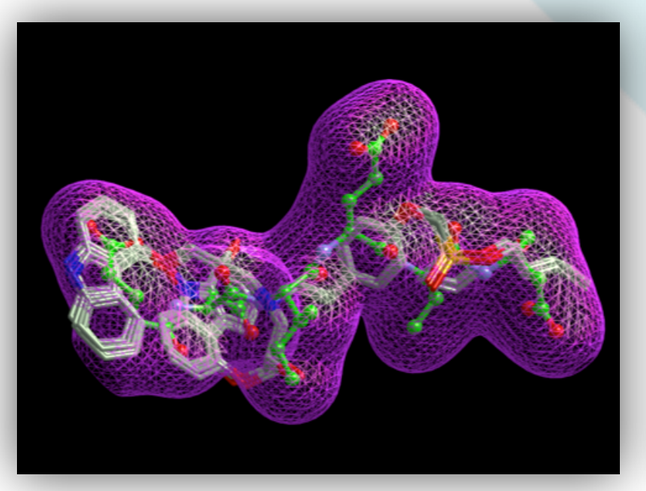
Model Building Tools

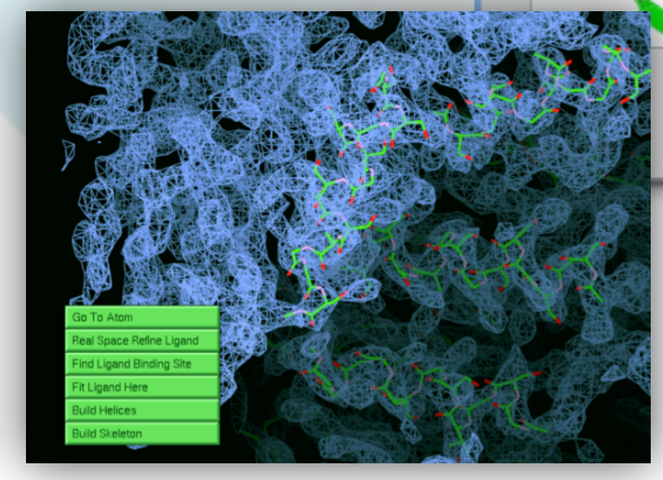
EMBL

Arp/wArp 7.6

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







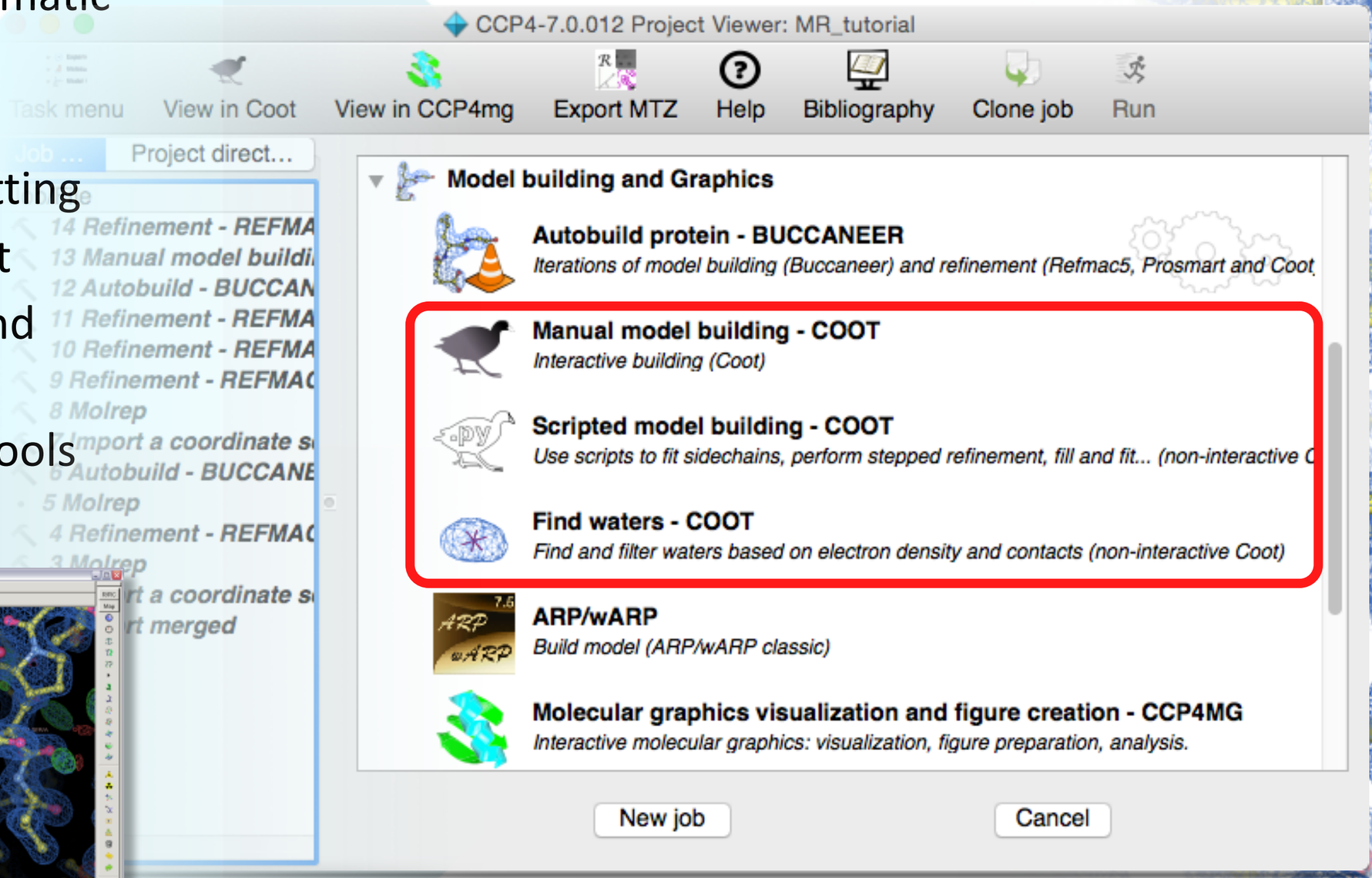
Science & Technology Facilities Council

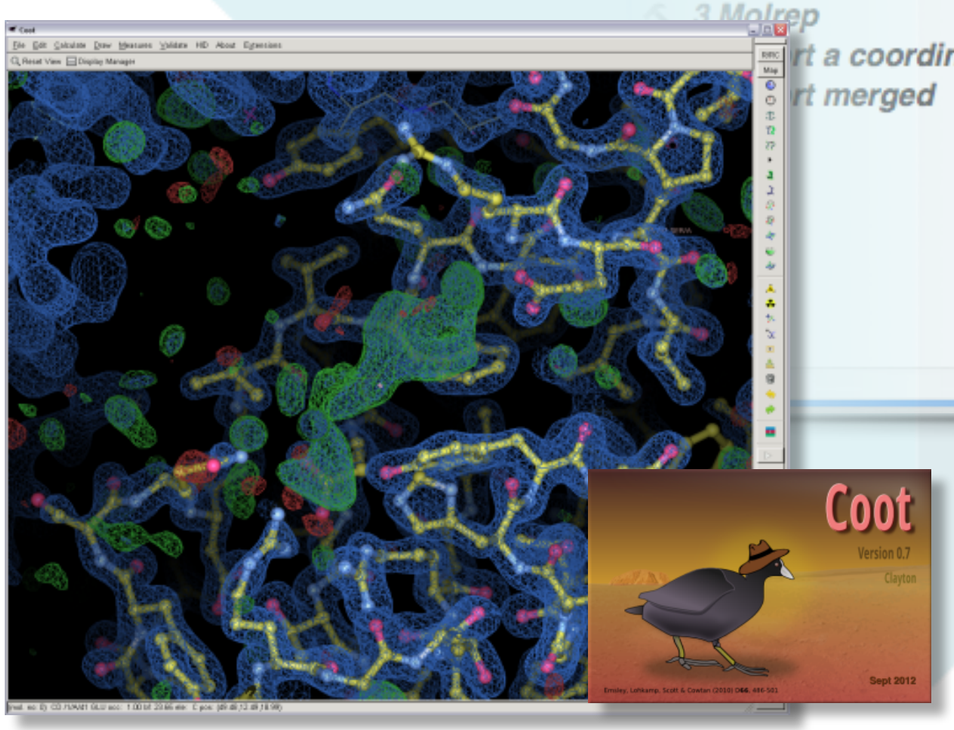
Research Complex at Harwell

Model Building Tools

Coot 0.8.3

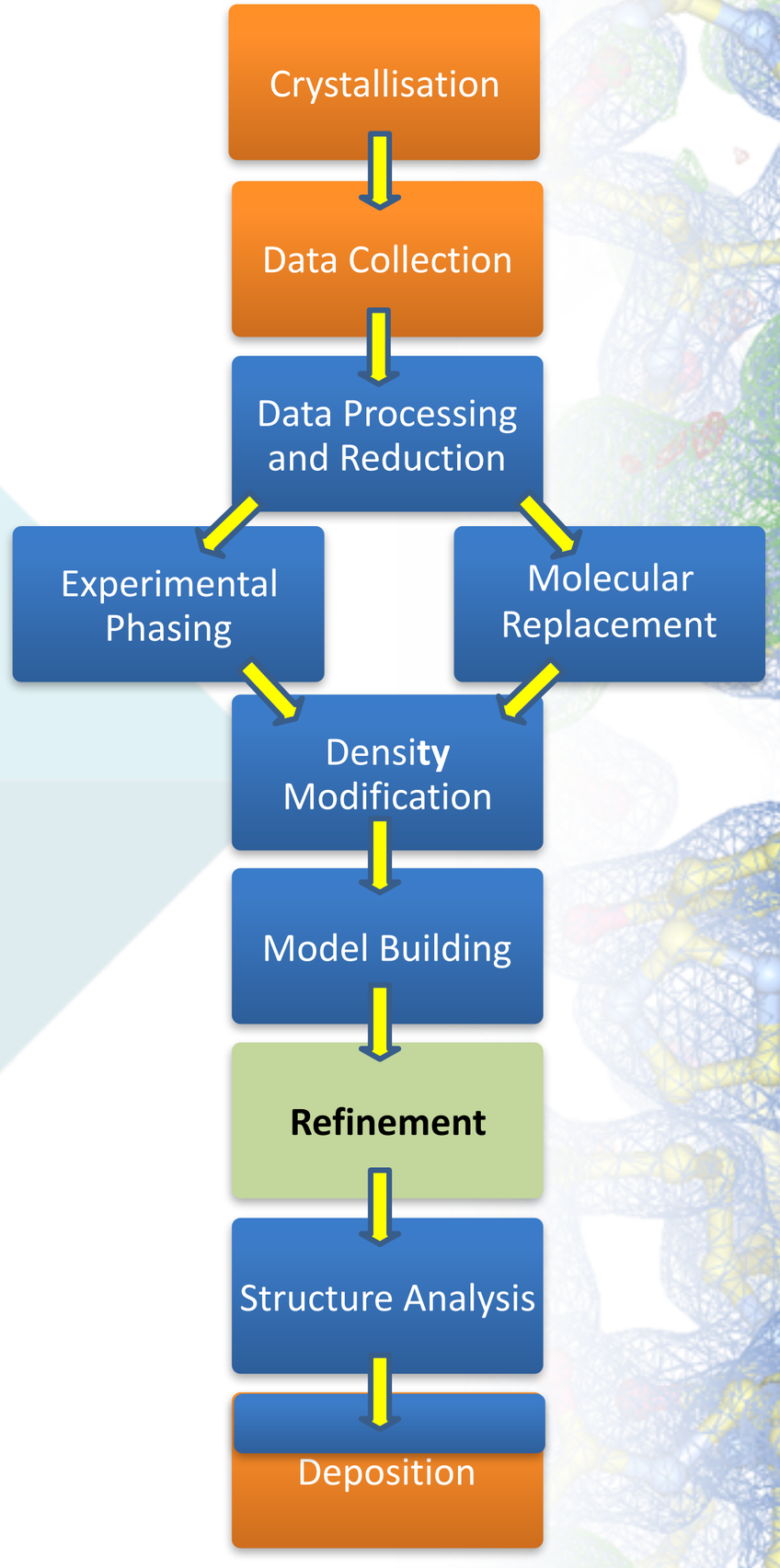
- Manual and semiautomatic model building
- Interactive graphics
- Ligand building and fitting
- Real-space refinement
- Paired with Refmac and jLigand
- Wealth of validation tools
- Integrated with CCP4





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



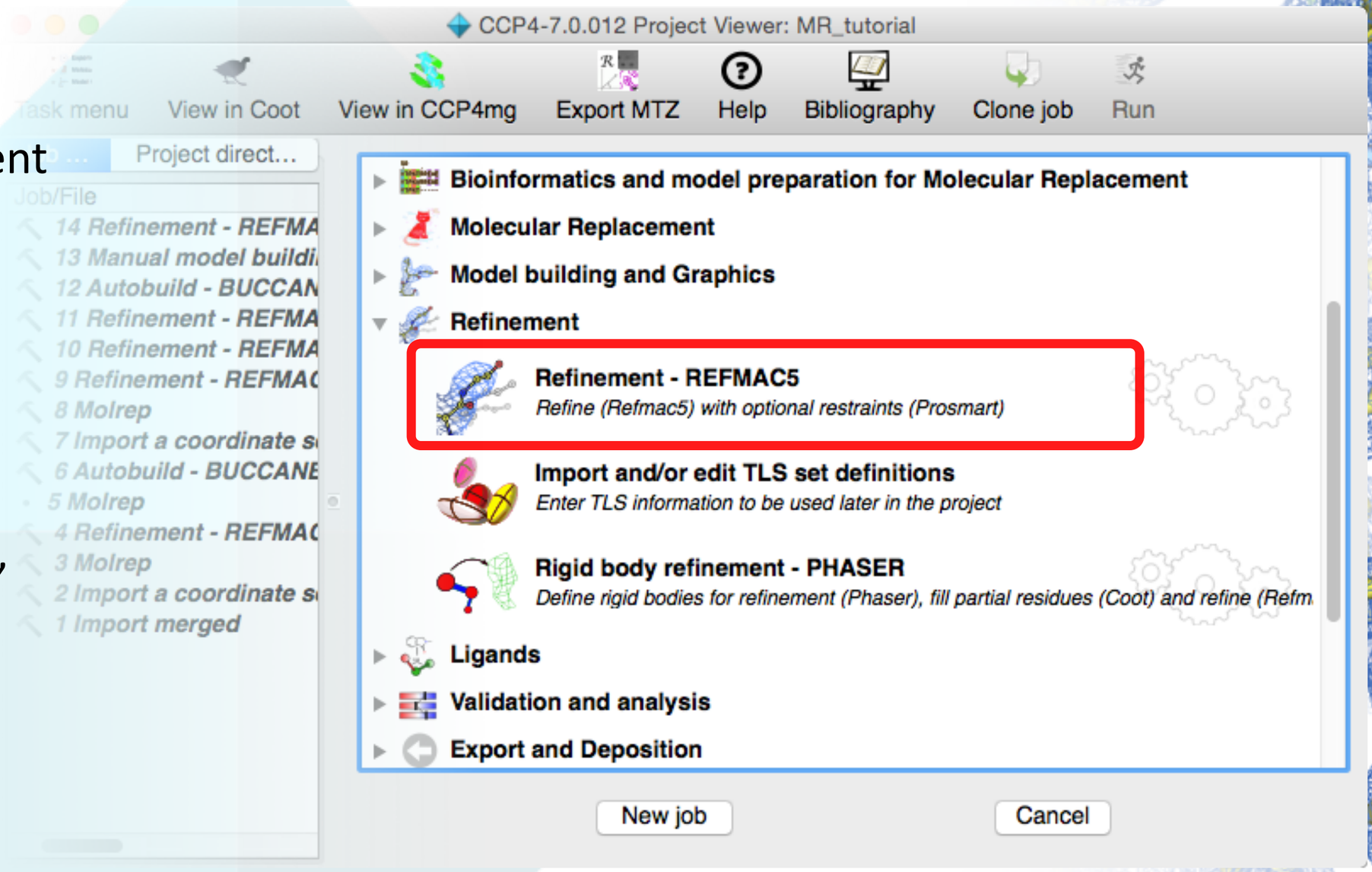
Science & Technology Facilities Council

Research Complex at Harwell

Refinement

Refmac 5.8

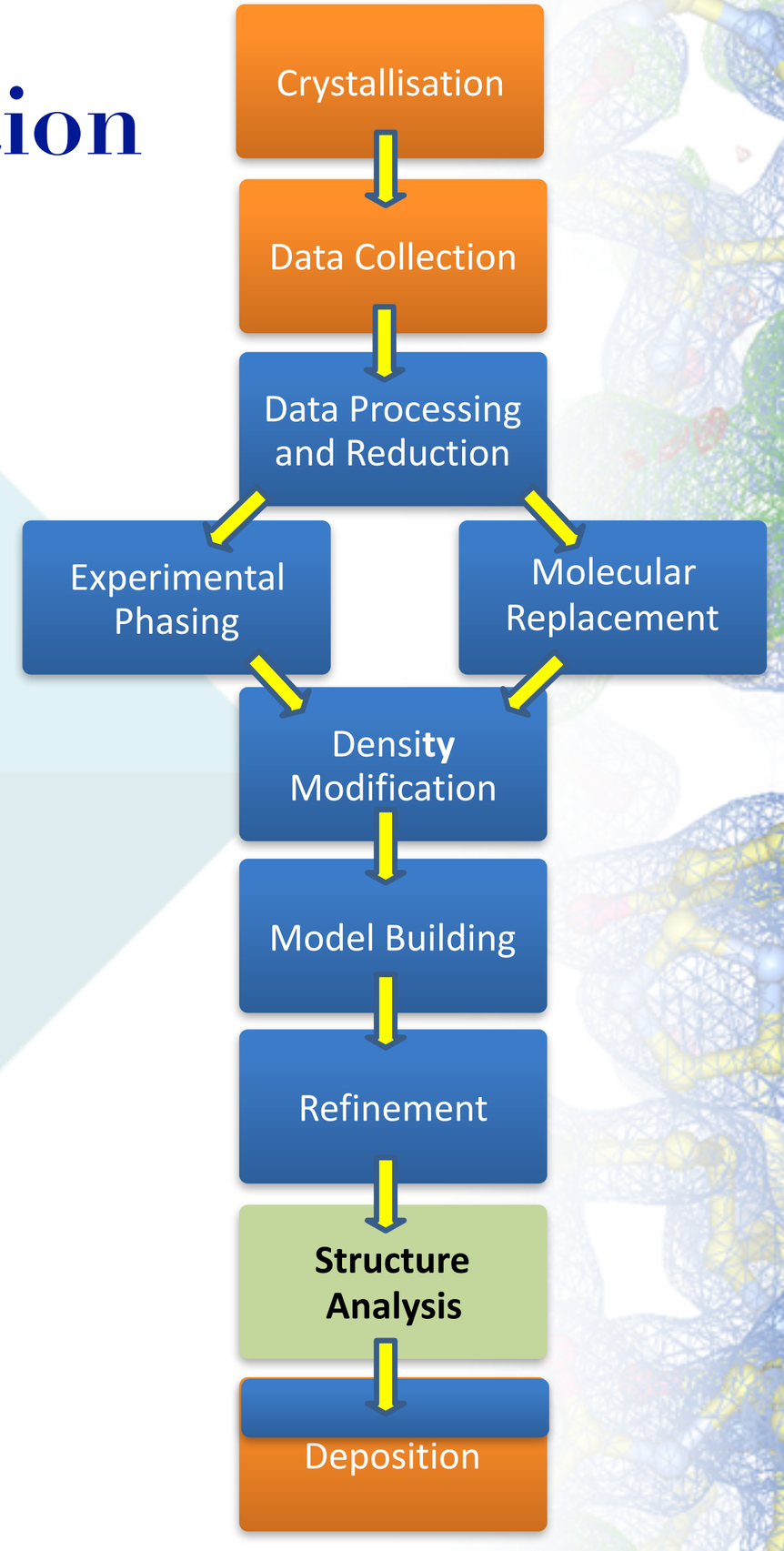
- CCP4's main refinement program
- Implements several likelihood target functions for experimental data (amplitudes, twinned data, external phases, Lattman coefficients)
- Restraints
  - CCP4 monomer library
  - Custom library
  - External restraints



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

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



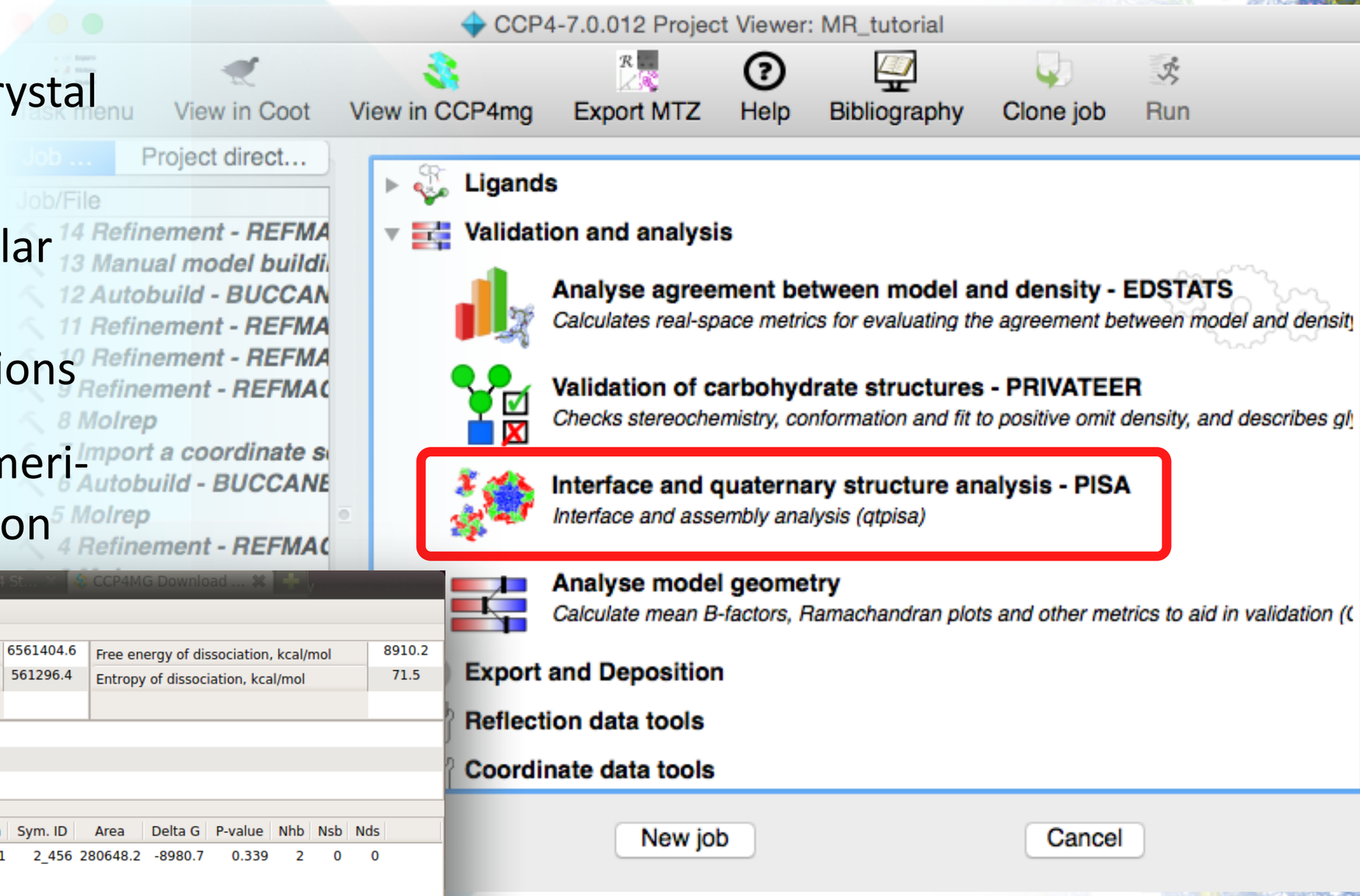
Science & Technology Facilities Council

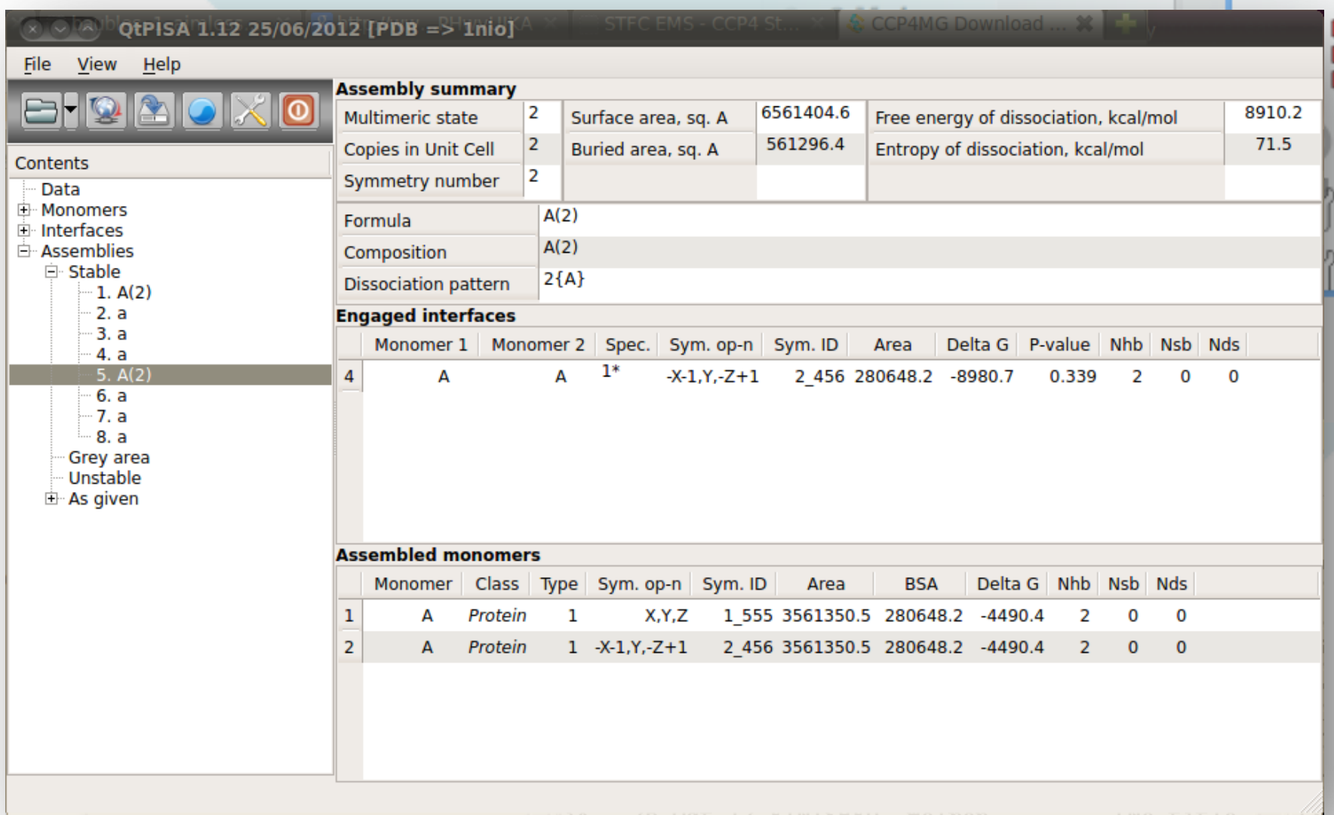
Research Complex at Harwell

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

Validation and Deposition

Coot

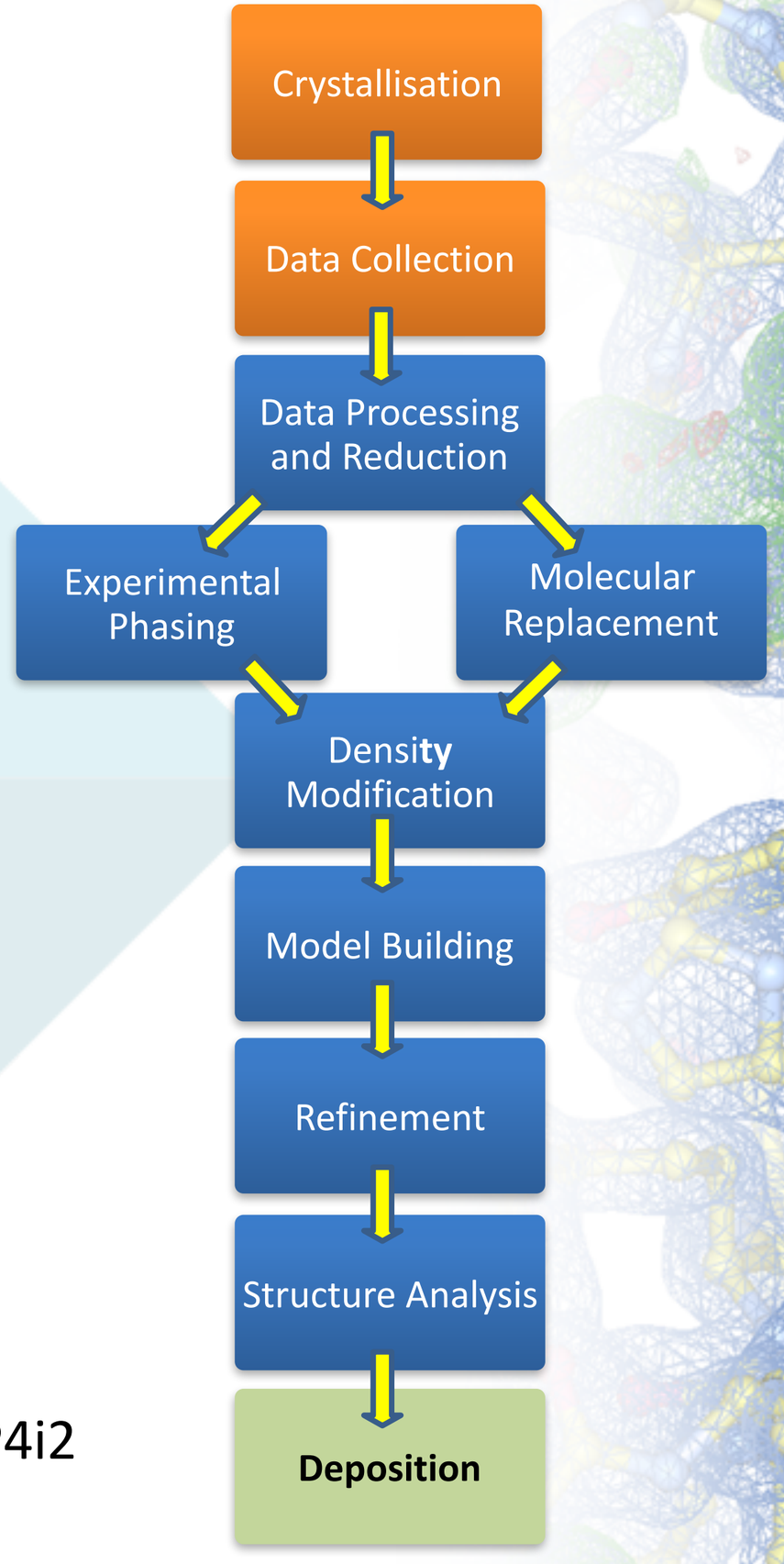
- Comprehensive set of validation tools: Ramachandran plots, density fit analysis, geometry analysis, unmodelled blobs etc.

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 module in CCP4i2



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

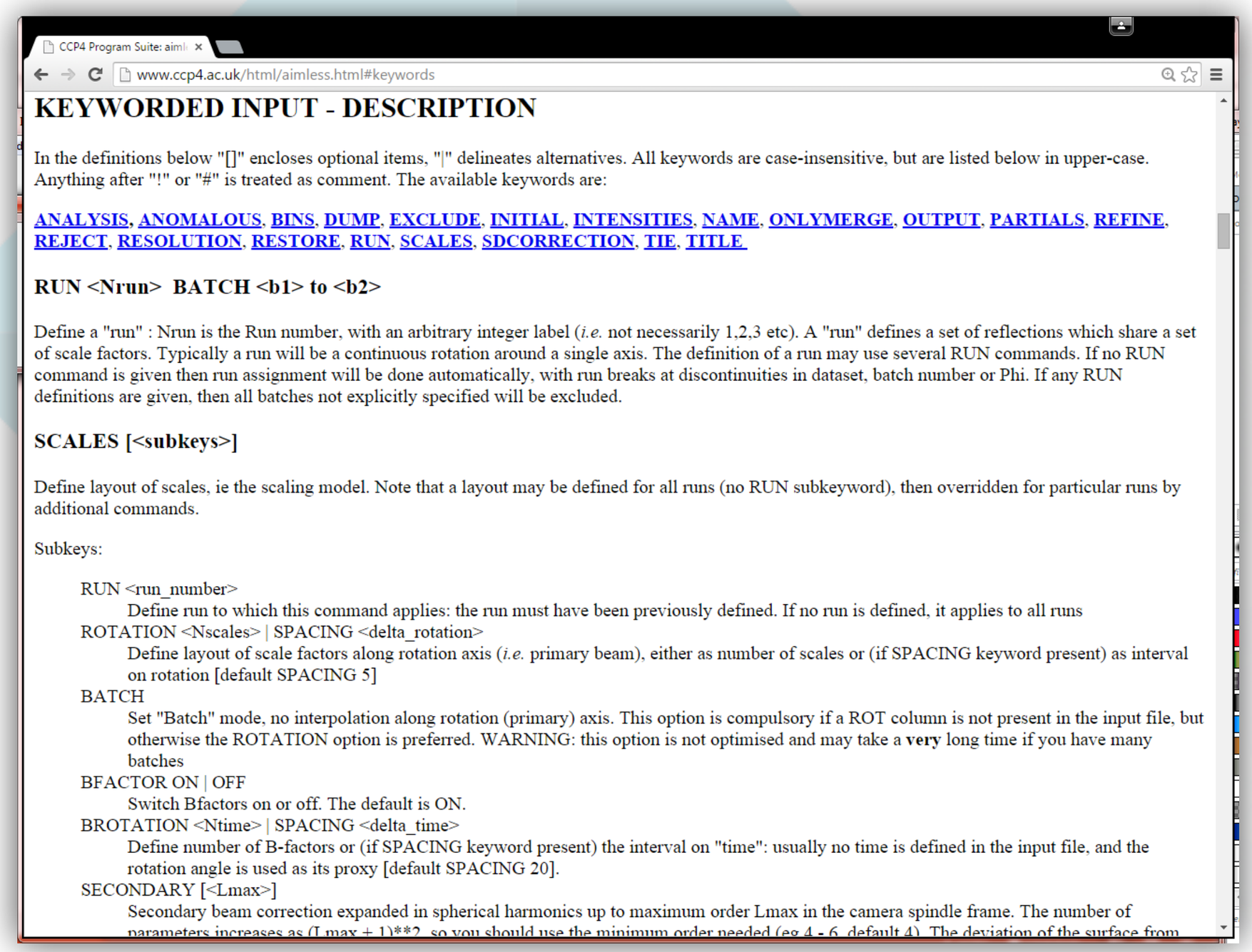
Keywords

```
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):*

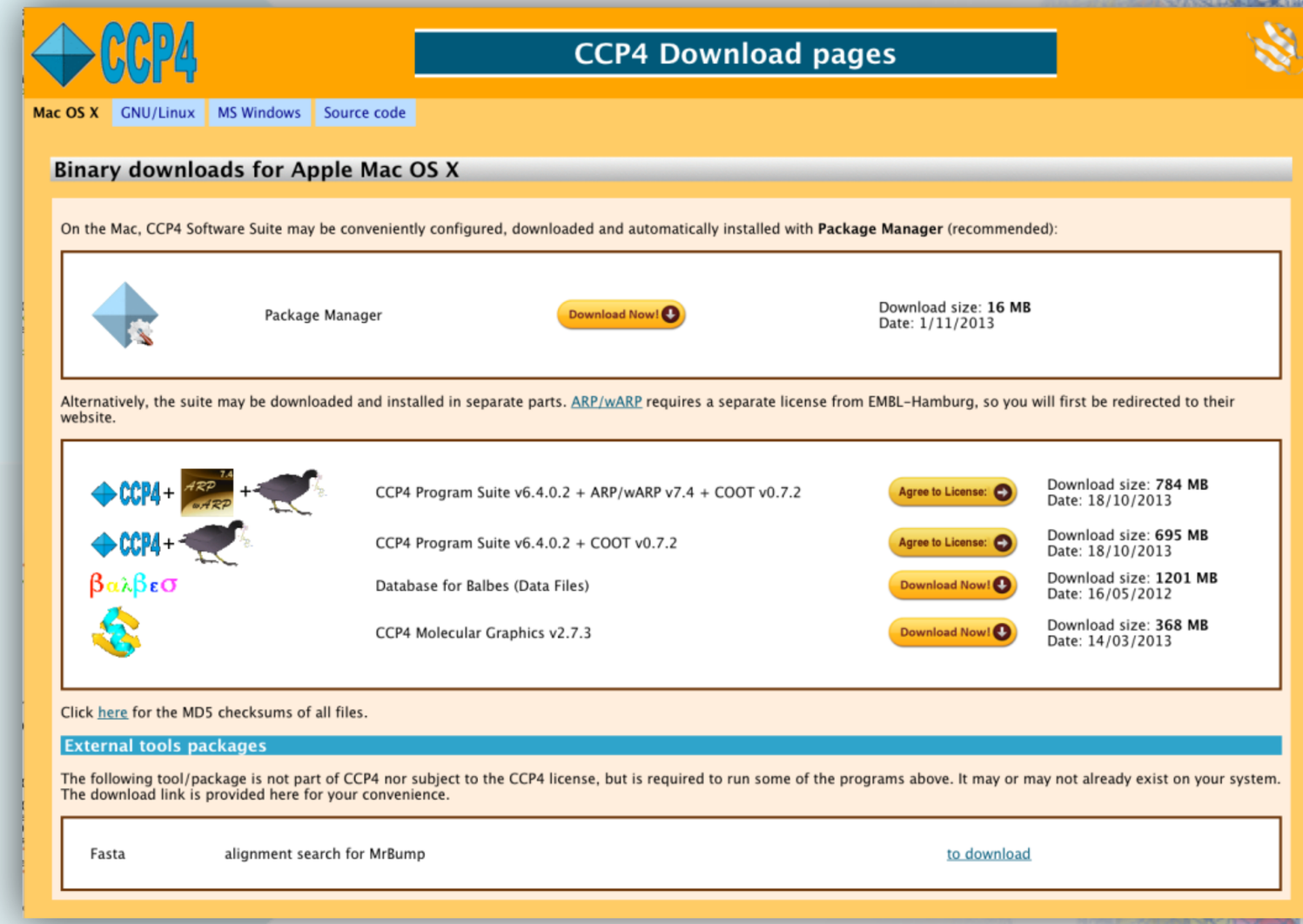


Science & Technology Facilities Council

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

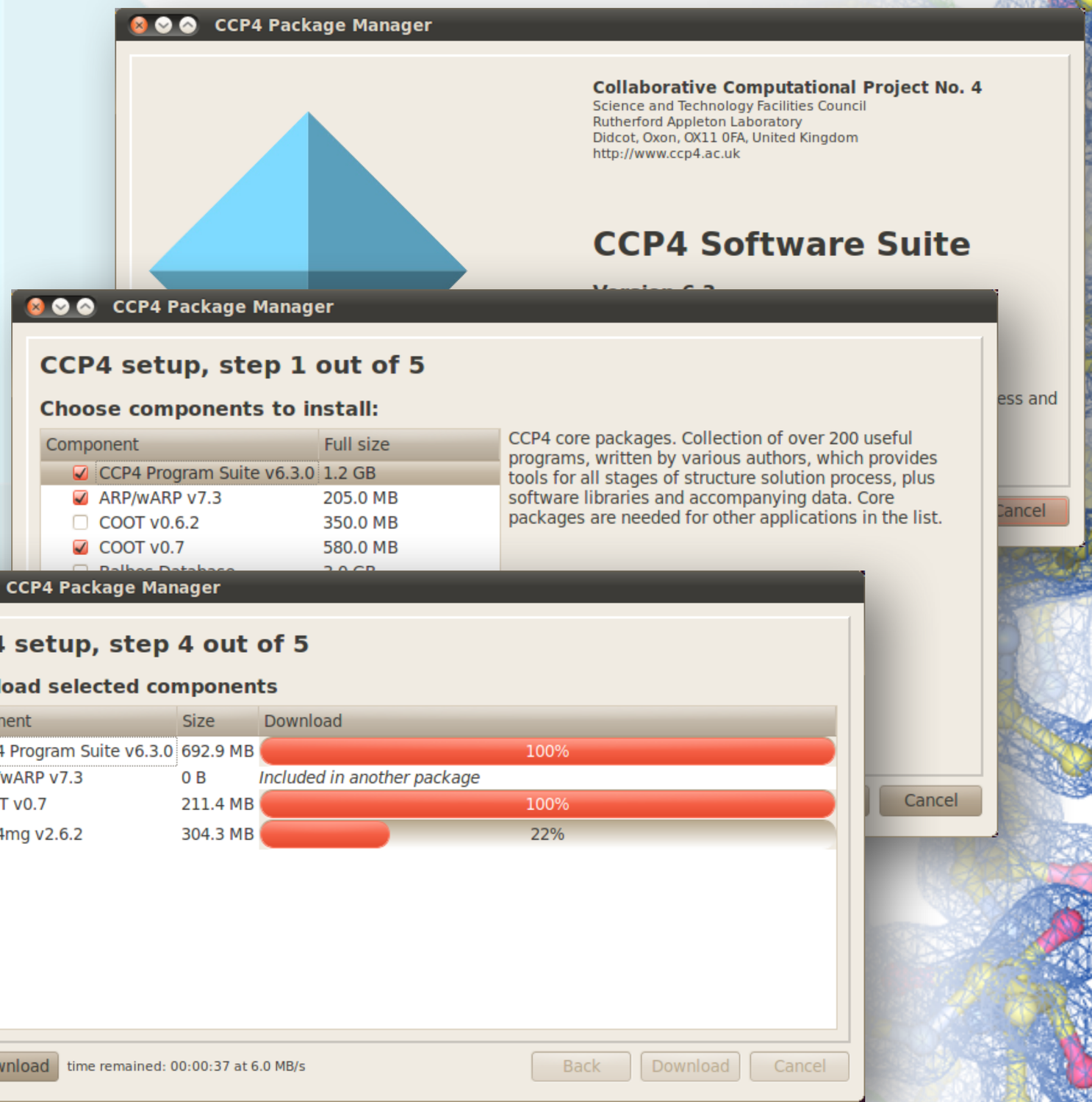


Science & Technology Facilities Council

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

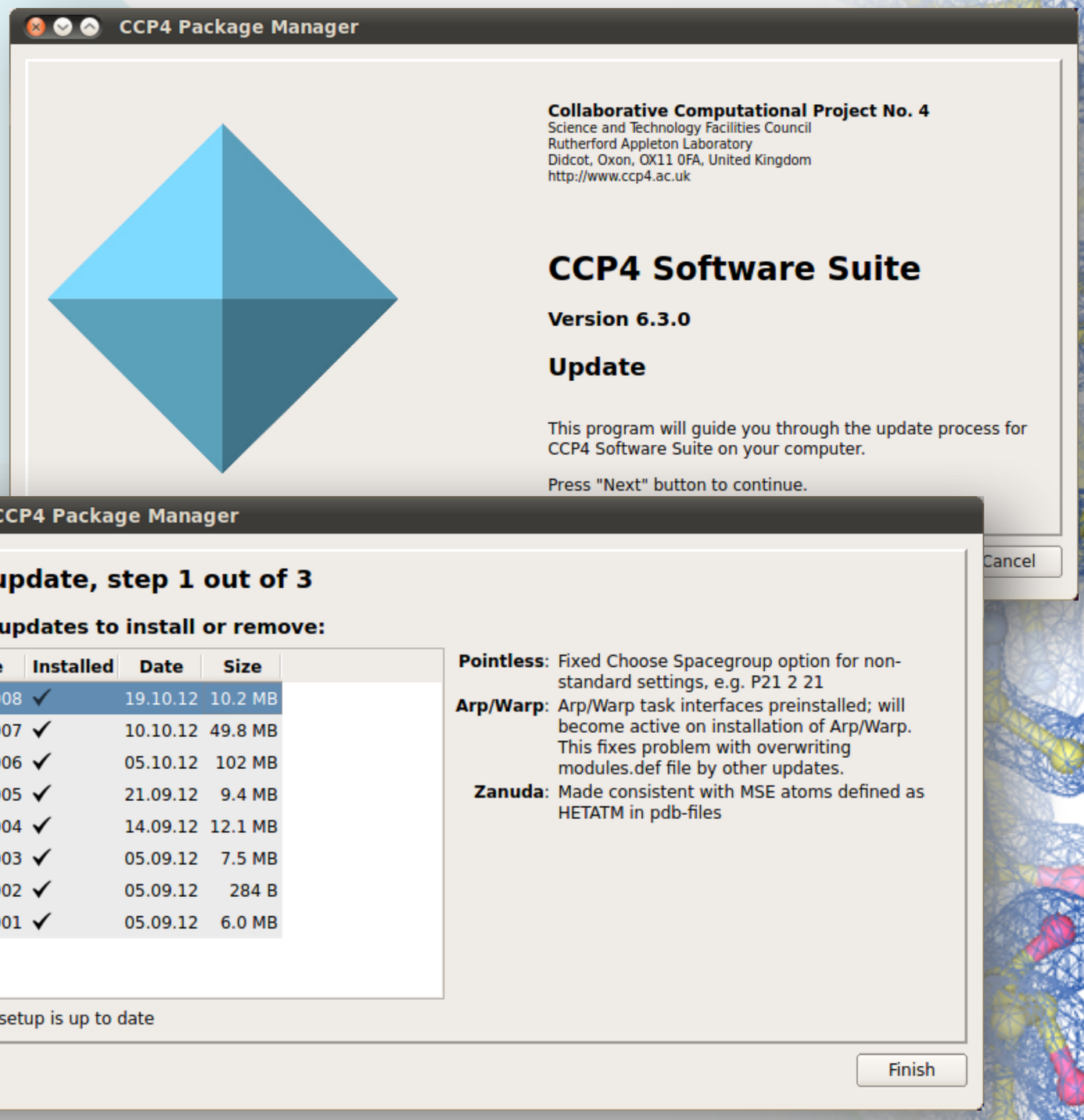


Science & Technology Facilities Council

Research Complex at Harwell

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

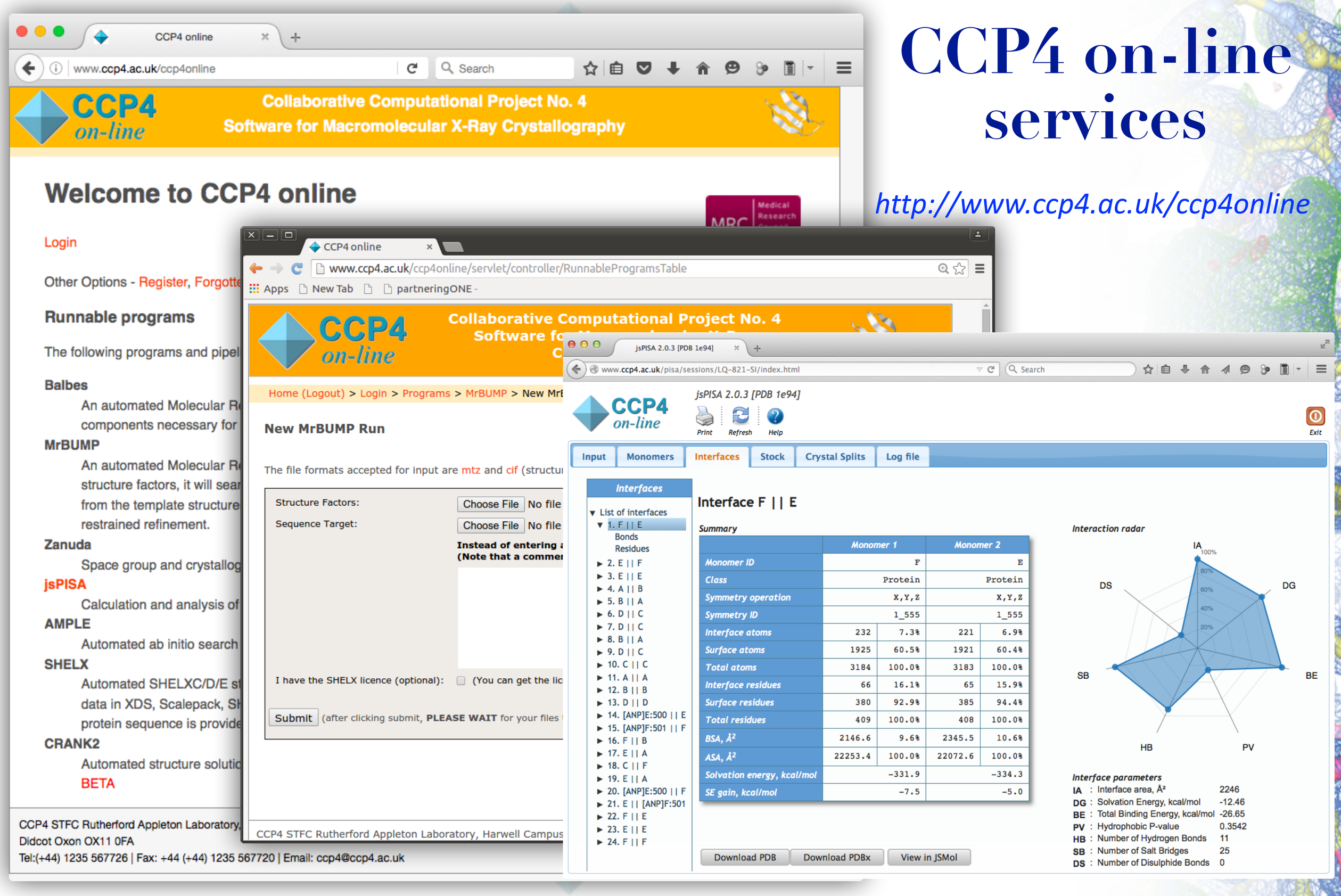
CCP4 Roadmap

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 on-line services

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



Acknowledgements

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

LMB/MRC: Andrew Leslie, Phil Evans, Garib Murshudov, Rob Nicholls, Oleg Kovalevskyi, 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, Jon Agirre, Eleanor Dodson

Diamond: Gwyndaf Evans, Graeme Winter

Leiden: Raj Pannu, Pavol Skubak

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