Welcome!

8th DLS-CCP4 Data Collection and Structure Solution Workshop

Welcome to the Virtual Space!

In order to make the Workshop most efficient, help tutors and each other, please:

- **Use your full name** when on Zoom/Slack, so that others know who you are
- **During a talk/presentation:**
  - ***Ensure your video/audio are off***
  - ***Post questions in the chat***
- In any other moment:
  - *Feel free/you are encouraged to keep your video and audio on*
  - *Just speak up to ask a question*
- What we expect – our **code of conduct**:
  - The Diamond-CCP4 Data Collection and Structure Solution Workshop is an inclusive event where people should feel comfortable sharing their work, opinions, and perspectives. All of us commit to engaging with each other mindfully to ensure an environment that promotes shared learning and collaboration.

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 each other**. 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

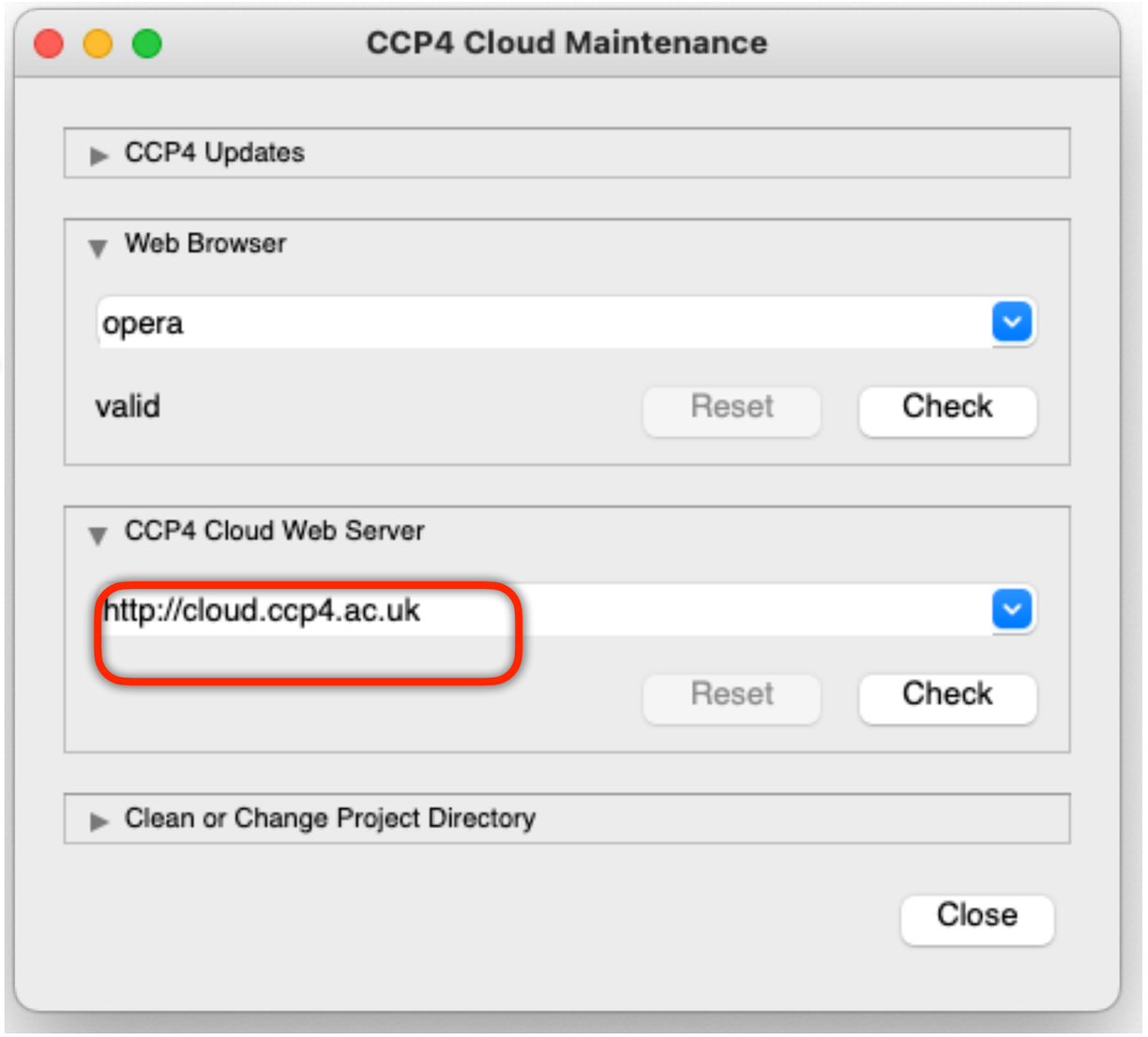
- **Course materials** will be uploaded to the website https://www.ccp4.ac.uk/schools/DLS-2021/course_material.php, so don't feel you have to copy every slide. You will have them available at the end of the workshop.

A few more points of the moment

- **If you still have problems with software setup and data upload:** post a message to the # help-computer channel on the Slack and we will help you out.

- **Your posters:** please upload them to *your own Slack channel*. Each upload will start a thread, in which the poster discussion will then take place asynchronously by Slack chat

- **CCP4 Cloud:** set http protocol in local configuration:



https can be used when connecting directly from browser. This will be corrected in next CCP4 update.

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!

CCP4

Collaborative Computational Project No. 4
Software for Macromolecular X-Ray Crystallography

diamond

DLS-CCP4 Data Collection and Structure Solution Workshop *November 29th - December 10th 2021*

Workshop Announcement

We are very pleased to announce the eighth joint Diamond-CCP4 MX workshop for graduate students, postdocs and early career scientists. As last year, the event remains online due to the COVID-19 pandemic. Data collection will take place at Diamond Light Source, on the Harwell Oxford campus in Oxfordshire, UK.

All aspects of structure solution will be covered during the workshop, from data collection through to phasing, refinement and validation.

- Lectures and tutorials will be delivered by experts in the field
- One day of data collection time will be provided at Diamond's excellent MX beamlines

DLS-CCP4 Data Collection and Structure Solution Workshop

November 30 - December 10, 2020, Diamond Light Source Ltd., UK

Introduction to CCP4: Project and Software

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

A Project with History

Science & Technology Facilities Council
Research Complex at Harwell

What is CCP4?

- 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 licensing
- The project is coordinated at Scientific Computing Department of the Science and Technology Facilities Council (STFC)

Science & Technology Facilities Council
Research Complex at Harwell

Main Activities

- 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

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

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 Main Event: the 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

Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Study Weekend

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

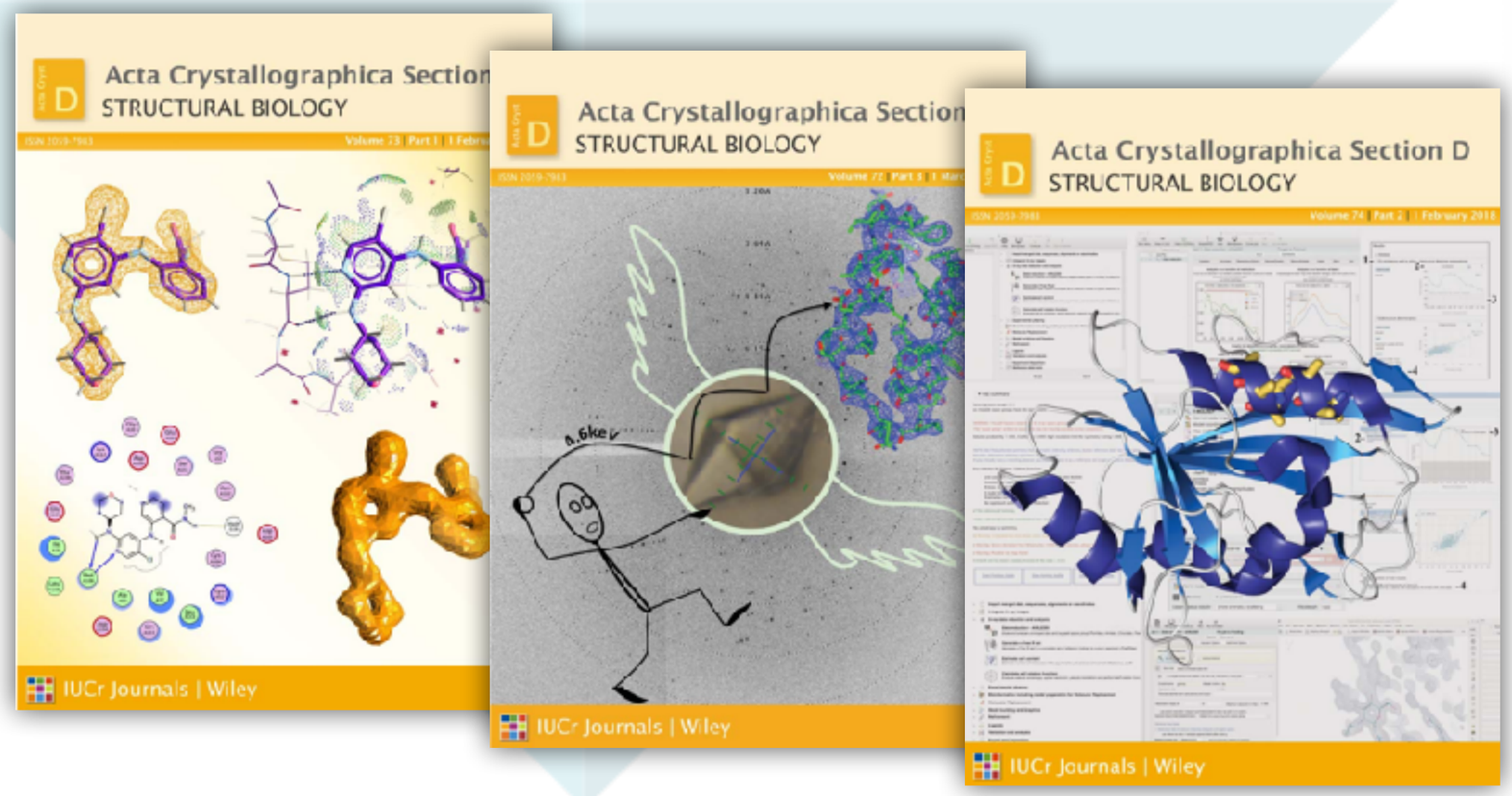
  2017: *"From crystals to structure with CCP4"*

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

  2019: "Molecular Replacement"

  2020: "Model Building"

  2021: "Integrating Structural Biology"



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

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

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 users around the world

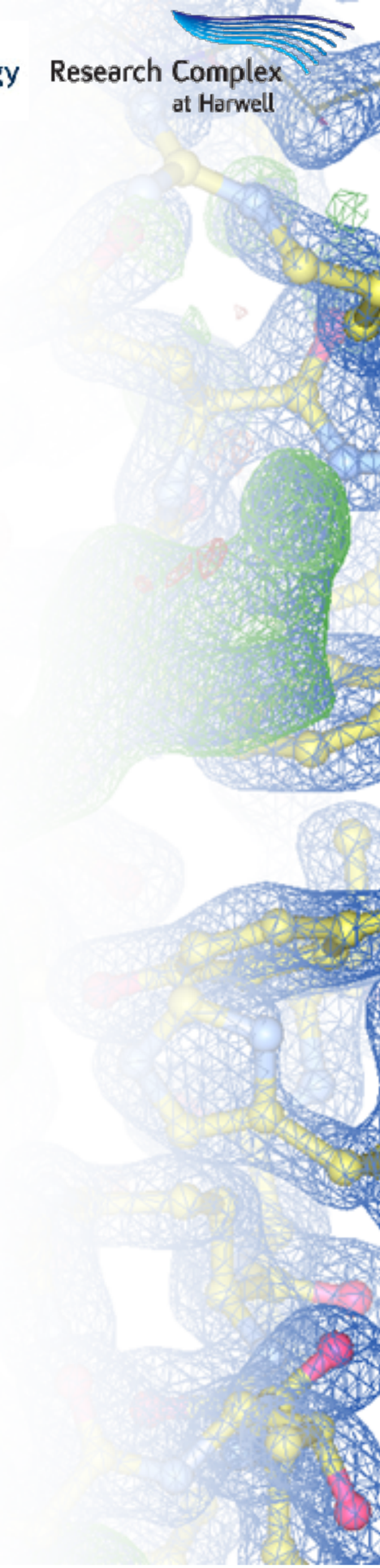
CCP4 Resources

- CCP4BB mailing list — bulletin board, news, expert help and advice
- CCP4 help desk — *ccp4@ccp4.ac.uk* — bug reports, problems with software
- CCP4 wiki @ Konstanz: — *http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Crystallography* (thanks to Kay Diederichs)
- CCP4 web site — *https://www.ccp4.ac.uk* (downloads, documentation, news and more)
- CCP4 Cloud manuals:
  https://cloud.ccp4.ac.uk/manuals/html-taskref/index.html (task manual)
  https://cloud.ccp4.ac.uk/manuals/html-userguide/index.html (user guide)

Documentation is a public resource and contributions are welcome!

Science & Technology Facilities Council

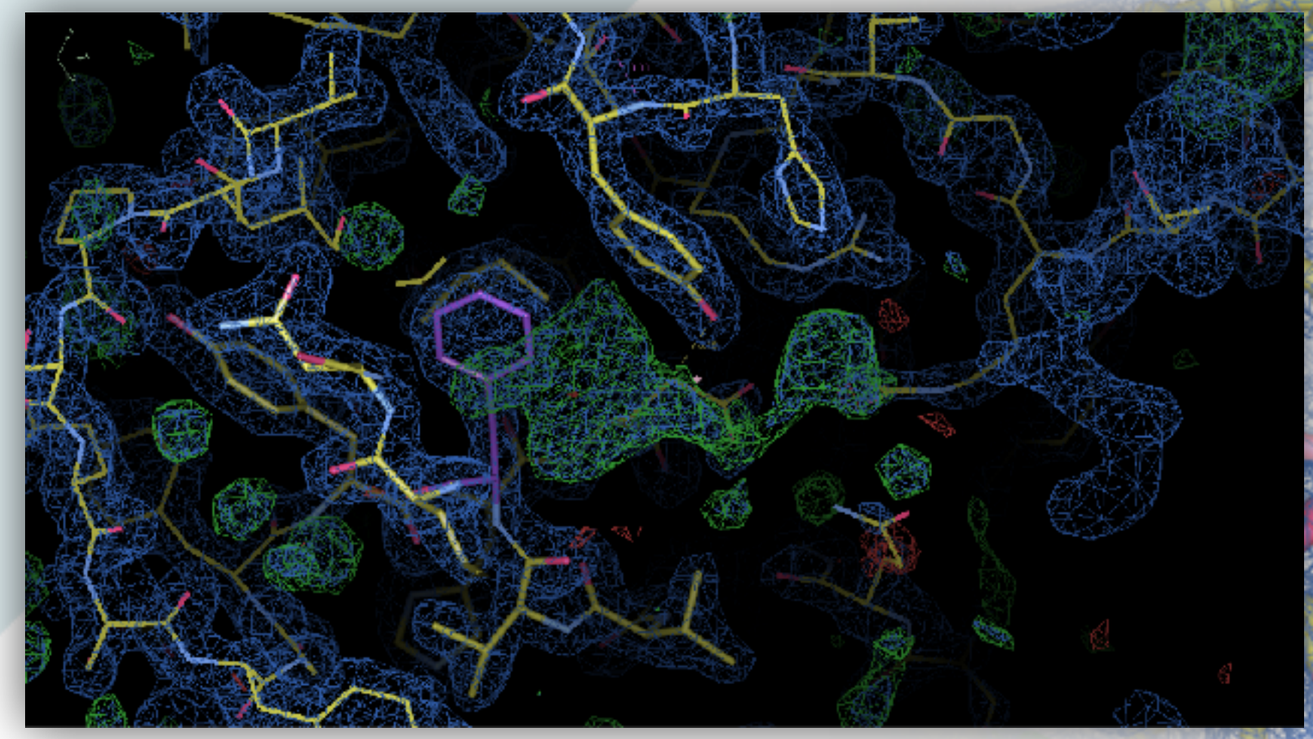
Research Complex at Harwell



Crystallography in nutshell

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

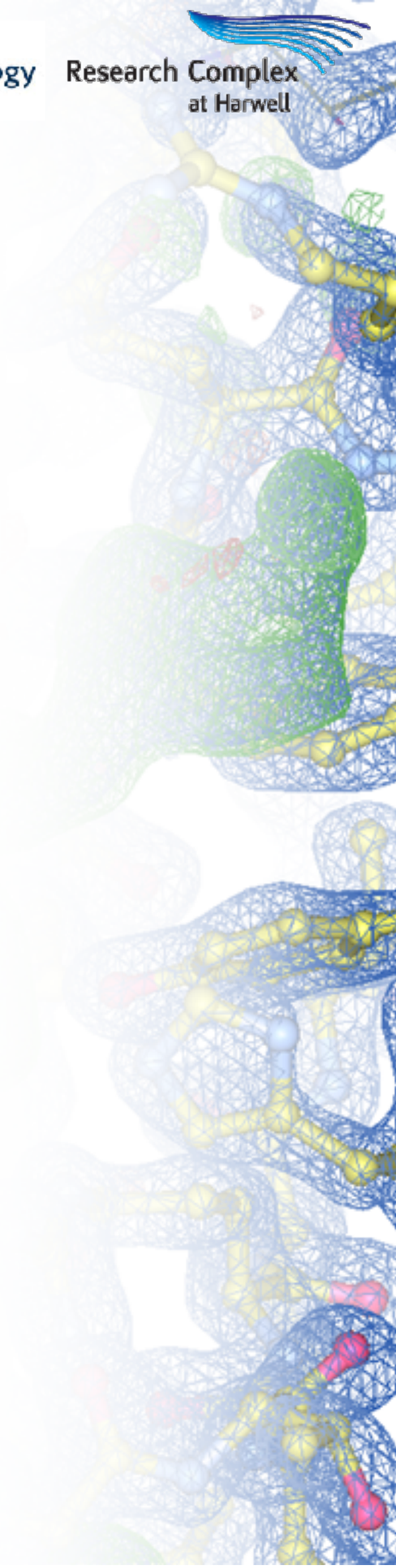


Structure Solution Pathway

- 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

Science & Technology Facilities Council

Research Complex at Harwell



The Software

Science & Technology Facilities Council
Research Complex at Harwell

The CCP4 Software Suite

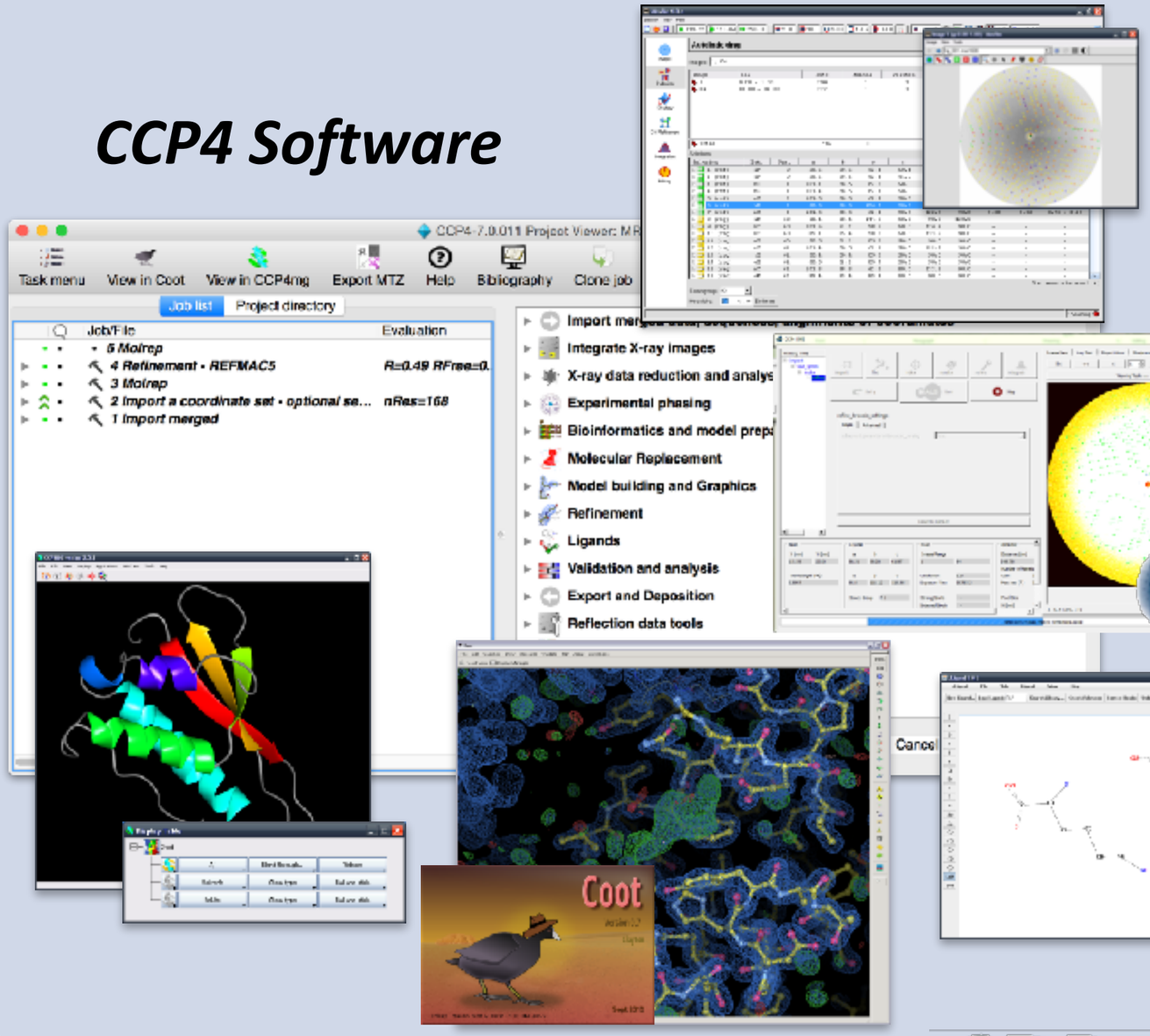
- The CCP4 Suite is a comprehensive suite of software for macromolecular crystallography and it contains about 400 program components
- New versions of the suite are released about every 12 months
  - New programs
  - Major updates to existing programs
  - Other new features such as changes to CCP4i 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

Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Releases

- Current release version — CCP4 7.1.008
- Recent programs:
  - SHELX, DIALS, MorDA, Arcimboldo, CCP4i2, CCP4 Cloud, CCP4Build
  - 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

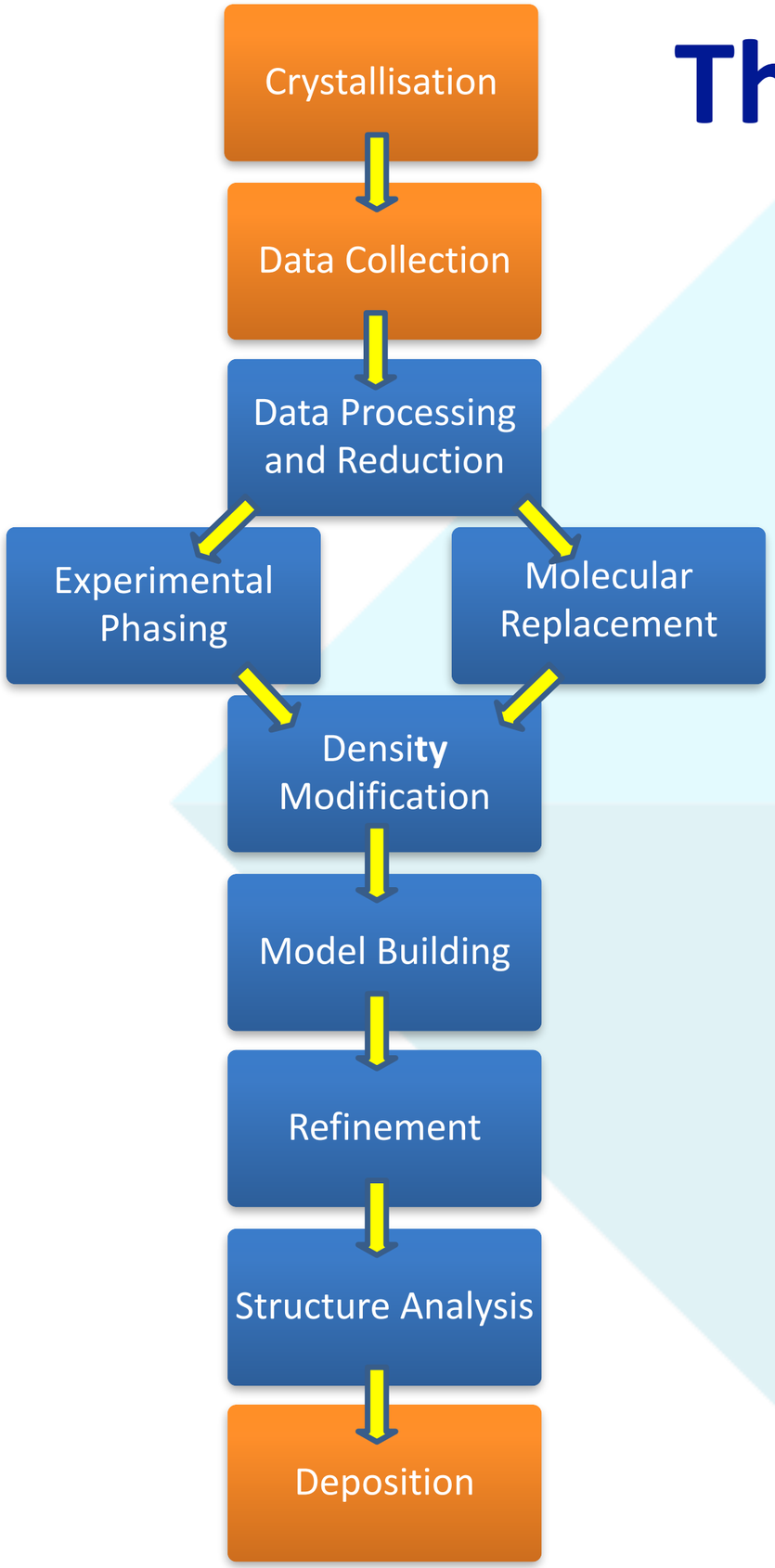
Research Complex at Harwell

Using CCP4

Science & Technology Facilities Council

Research Complex at Harwell

The Structure Solution Stages



	Automated*	Fundamental*
Data Processing and Reduction	Xia-2	XDS, DIALS, Aimless, DUI, XDSGui, iMosflm
Molecular Replacement	MrBump, Balbes, MorDA, Crank-2, Simbad, Ample	Phaser, Molrep, ShelxCD, Sculptor, Chainsaw
Density Modification		Parrot, Acorn, ShelxE
Model Building	Buccaneer, CCP4Build, Arp/wArp	Coot, AceDrg, jLigand
Refinement	Lorestr, Dimple	Refmac, Prosmart
Structure Analysis	PDB Deposition module	Zanuda, PISA, Gesamt

* *only selected representative programs are shown*

Programs' use and Projects

- Crystallographic projects may be very complex
  - ~ over 100 programs may need to be used
  - ~ sometimes, 100s jobs may need to be run
- All CCP4 programs can be used from command line
  - ~ [+] gives full control of programs and access to all parameters
  - ~ [-] difficult to memorise all options and parameters
  - ~ [-] difficult to keep 100s and 1000s files organised
  - ~ [-] may be difficult to analyse the structure solution pathway *post factum*
- Most frequently used programs are available through graphical interfaces
  - ~ [+] greatly simplifies manipulation with data and program parameters
  - ~ [+] arranges work in ***Projects***, easier to navigate, analyse and make decisions
  - ~ [+] take care of many petty jobs such as routine format conversions, map preparations and other
  - ~ [-] may give access only to a subset of most frequently used program parameters
  - ~ [-] may not exist for a particular program of interest

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

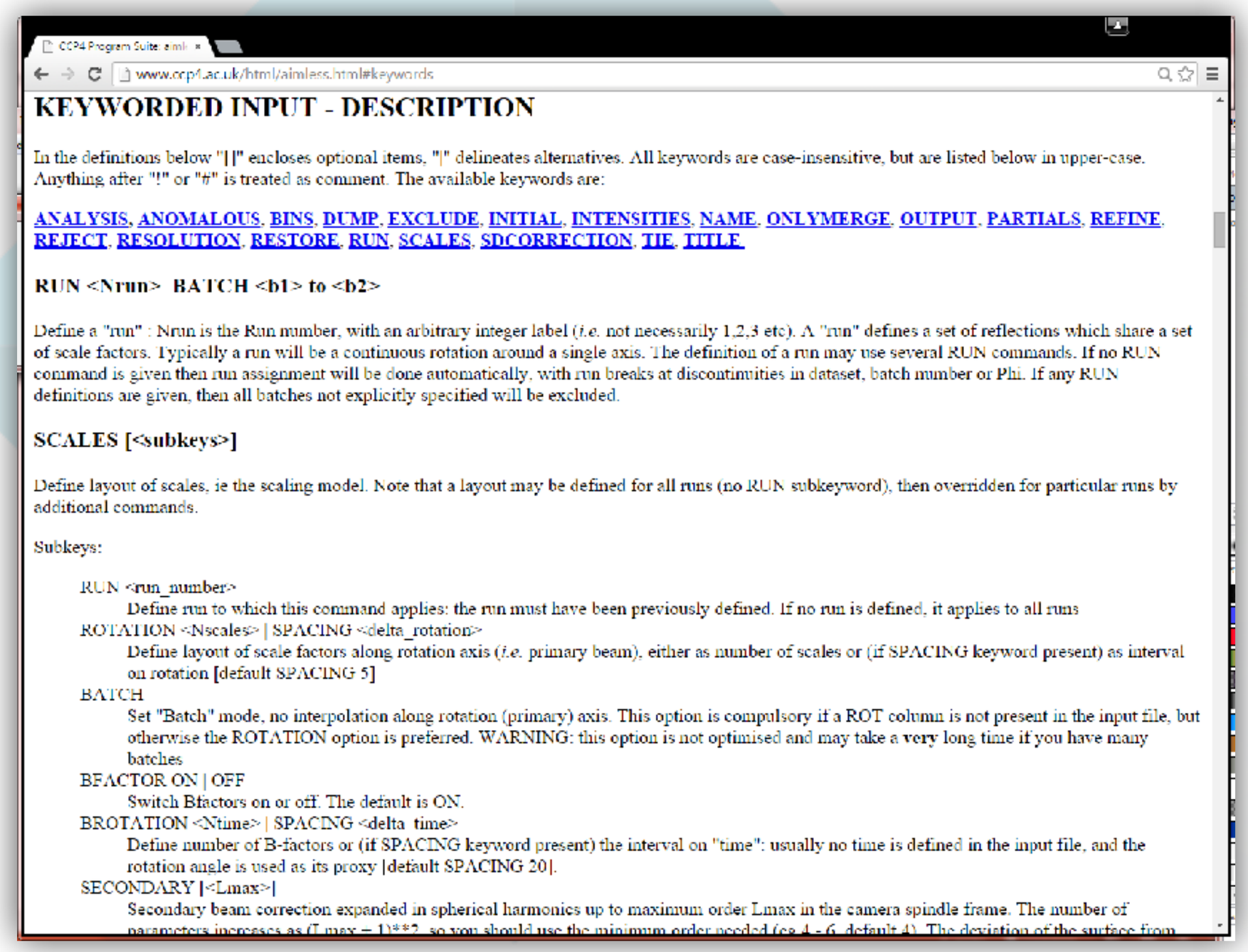
```
f2mtz hklin ${1} hklout shelxe-output.mtz << eof
TITLE   ample
cell    ${cell}
symm    ${symm}
labout  H K L F FOM PHI SIGF
CTYPOUT H H H F W P Q
pname   ample
dname   ample
END
```

Science & Technology Facilities Council

Research Complex at Harwell

Command Line Operation

- *CCP4 program keyword descriptions are found in documentation pages online or in the suite ($CCP4/html/INDEX.html):*



Science & Technology Facilities Council

Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

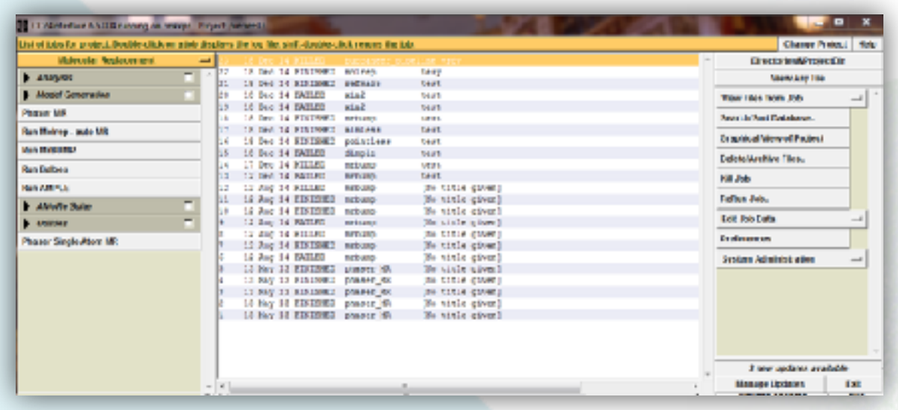
Science & Technology Facilities Council

Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

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



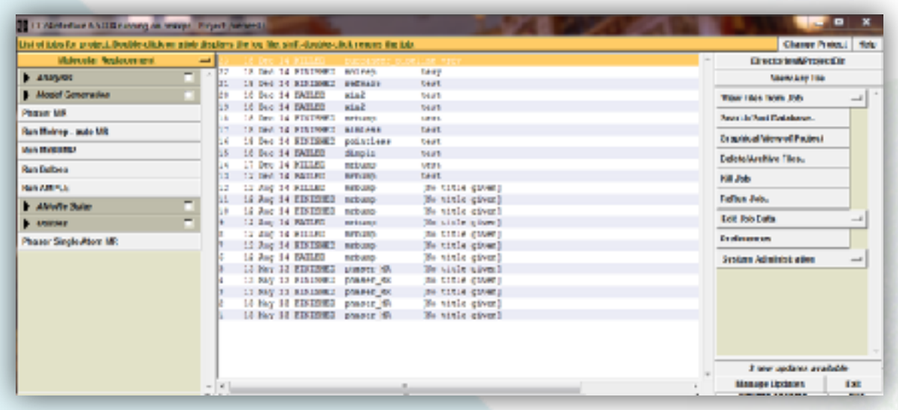
Science & Technology Facilities Council

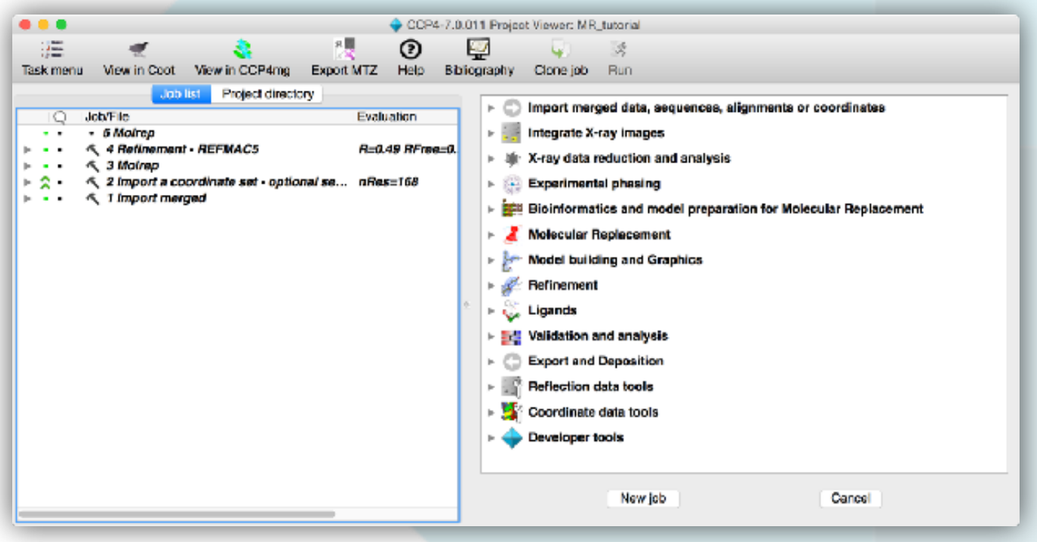
Research Complex at Harwell

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)

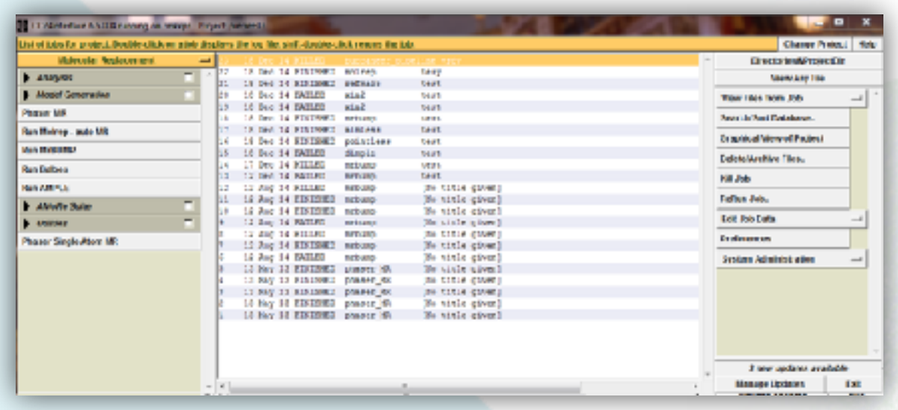
Science & Technology Facilities Council

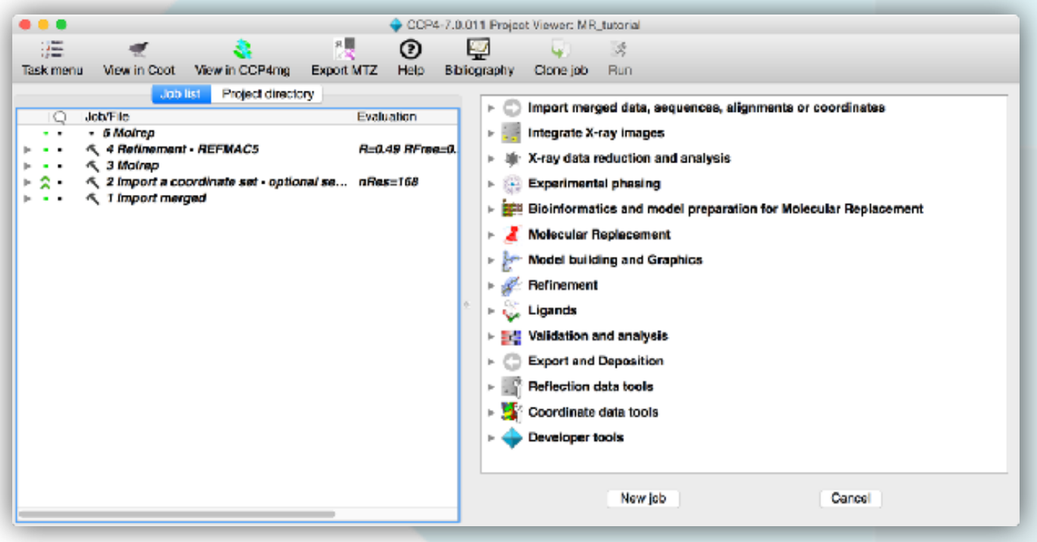
Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

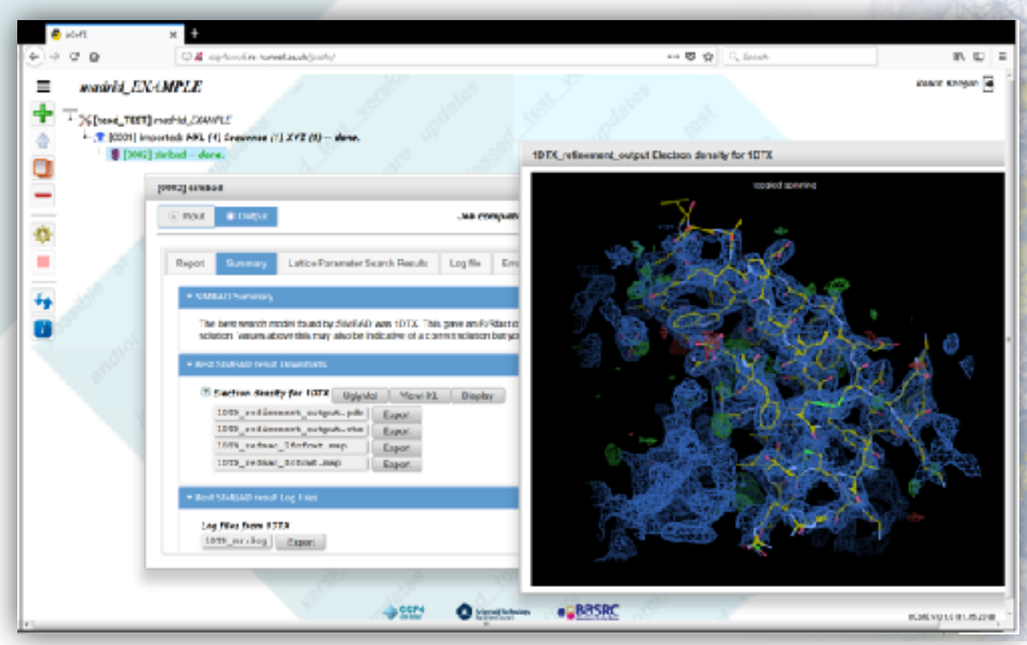
1. **CCP4i** – original interface developed around 2000





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

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



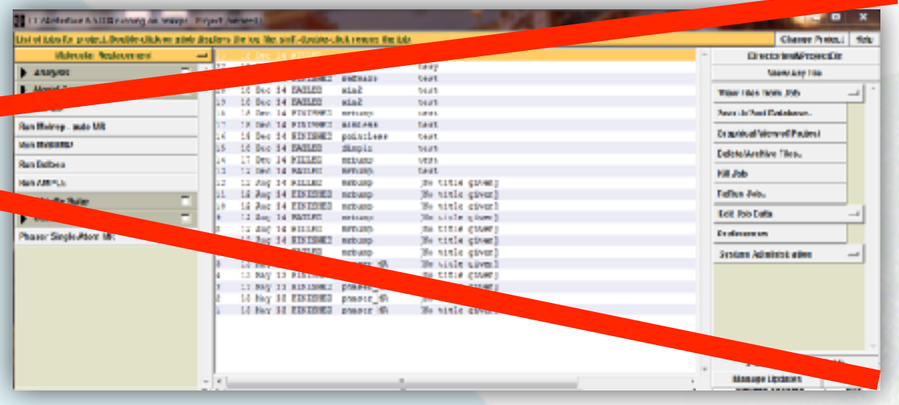
Science & Technology Facilities Council

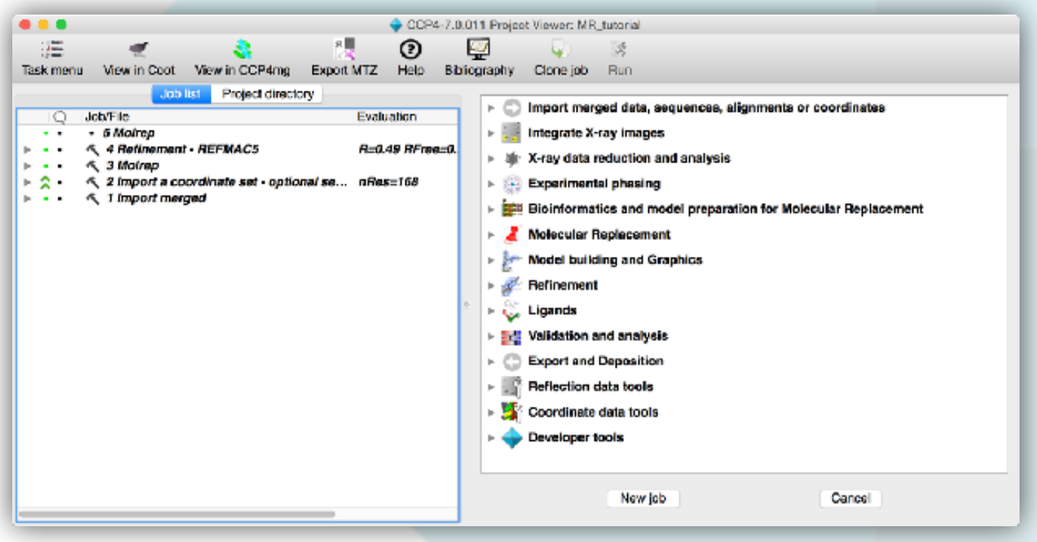
Research Complex at Harwell

The CCP4 user interface(s)

- Three main interfaces to the CCP4 suite

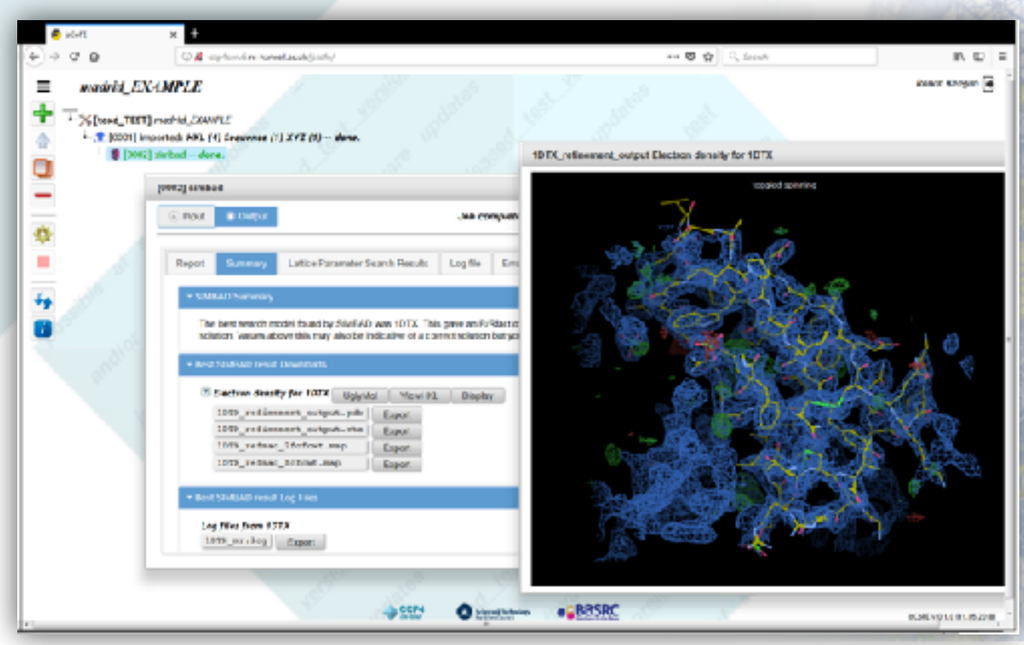
1. **CCP4i** – original interface developed around 2000
   deprecated





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

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



Science & Technology Facilities Council

Research Complex at Harwell

Documentation in CCP4i2

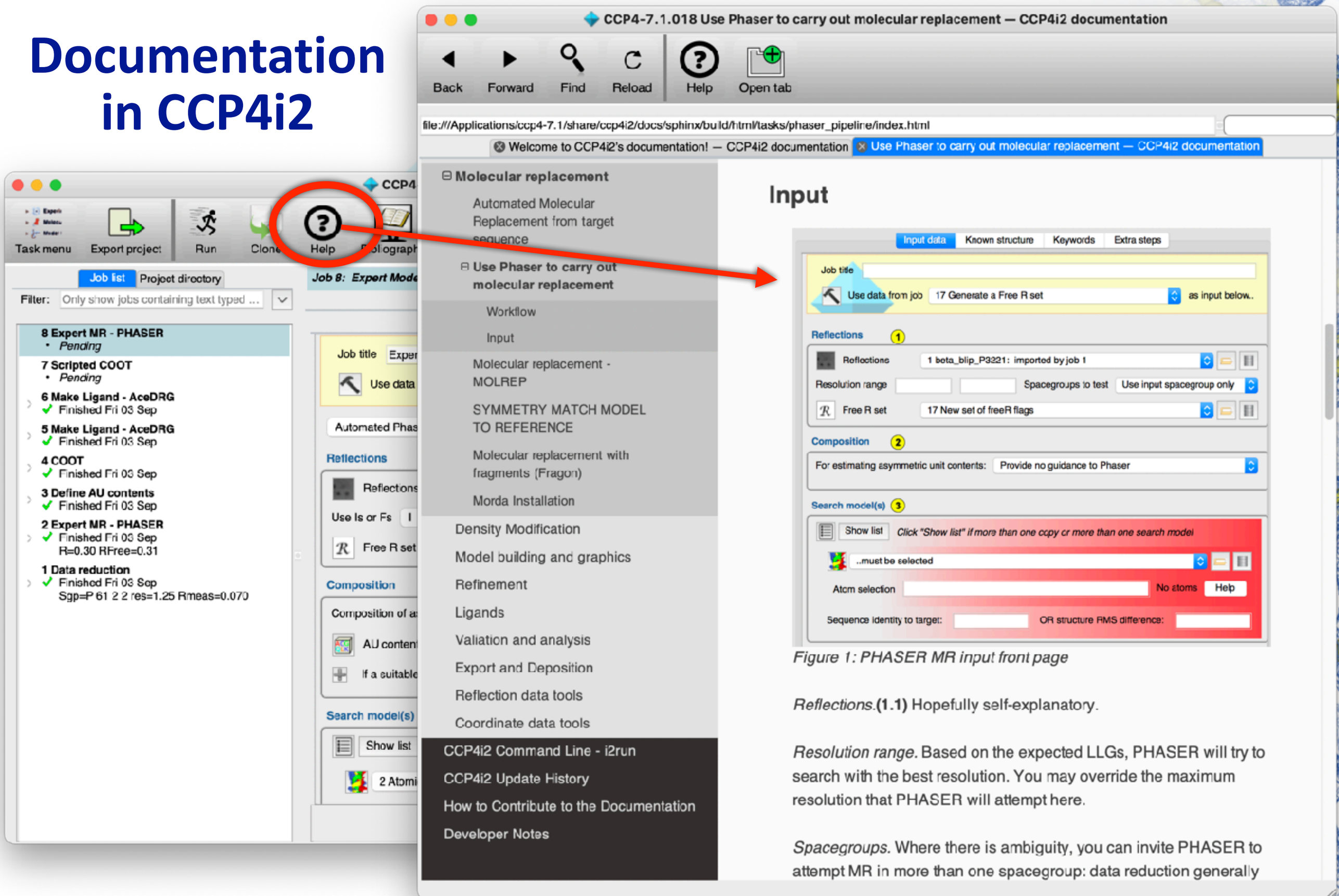


Figure 1: PHASER MR input front page

*Reflections.***(1.1)** Hopefully self-explanatory.

Resolution range. Based on the expected LLGs, PHASER will try to search with the best resolution. You may override the maximum resolution that PHASER will attempt here.

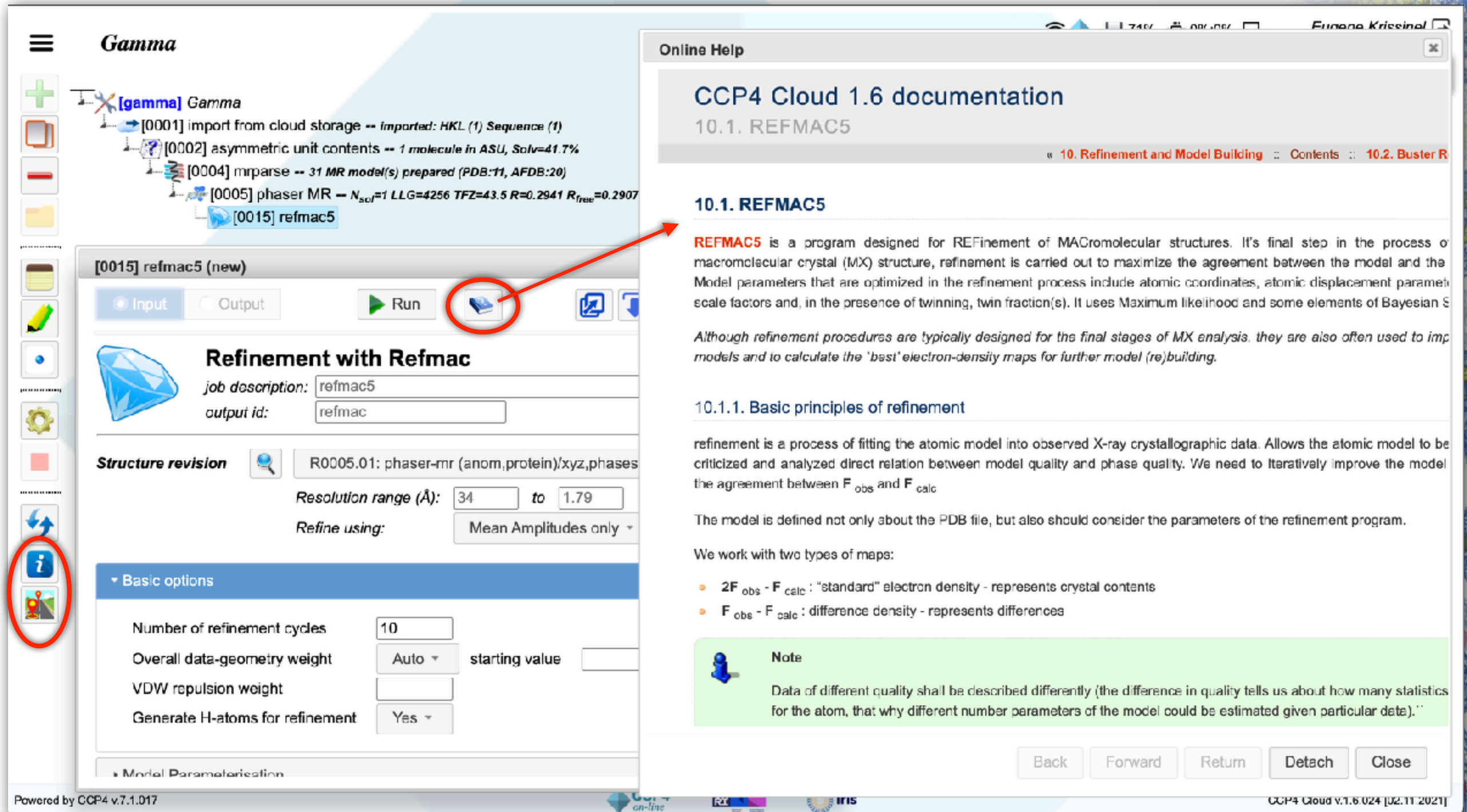
Spacegroups. Where there is ambiguity, you can invite PHASER to attempt MR in more than one spacegroup: data reduction generally

Science & Technology Facilities Council

Research Complex at Harwell

Documentation in CCP4 Cloud

- Trying to focus on the scientific content of the task, rather than interface features



Science & Technology Facilities Council
Research Complex at Harwell

Documentation in CCP4

Community input and effort

GitLab

Online Help

- 13.3. X-ray cross sections and anomalous scattering factors
- 13.4. Coordinate Utilities
- 13.5. Gemmi
- 13.6. Pairwise Structural Alignment with GESAMT
- 13.7. Structure Superposition with LsqKab
- 13.8. Sequence Alignment with ClustalW
- 13.9. Symmetry Match to Reference Structure with CSymMatch

Search

- Search Page

Authors and Contributors

Task Reference for CCP4 Cloud was prepared by

- *Maria Fando (Institute of Protein Research, Pushchino, Russian Federation)*
- *Eugene Krissinel (CCP4, STFC, Harwell, UK)*

with help and contributions from

- *Andrey Lebedev (CCP4, STFC, Harwell, UK)*
- *Robert Nicholls (MRC-LMB, Cambridge, UK)*
- *Garib Murshudov (MRC-LMB, Cambridge, UK)*
- *Pavol Skubak (Leiden University Medical Centre, Leiden, The Netherlands)*

Please send documentation-related questions and suggestions to: **CCP4 Cloud Reception** .

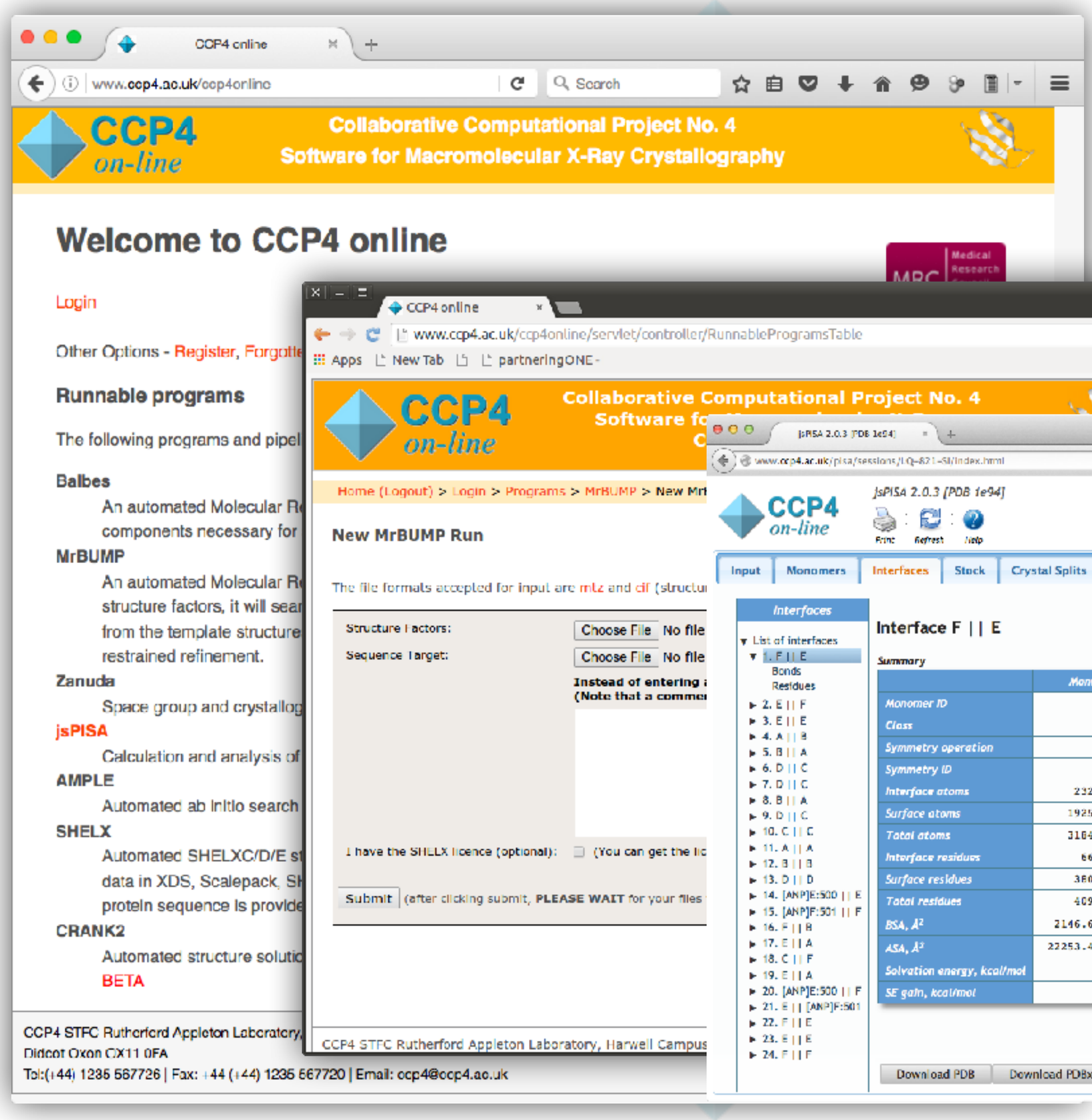
Back | Forward | Return | Detach | Close

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 on-line services

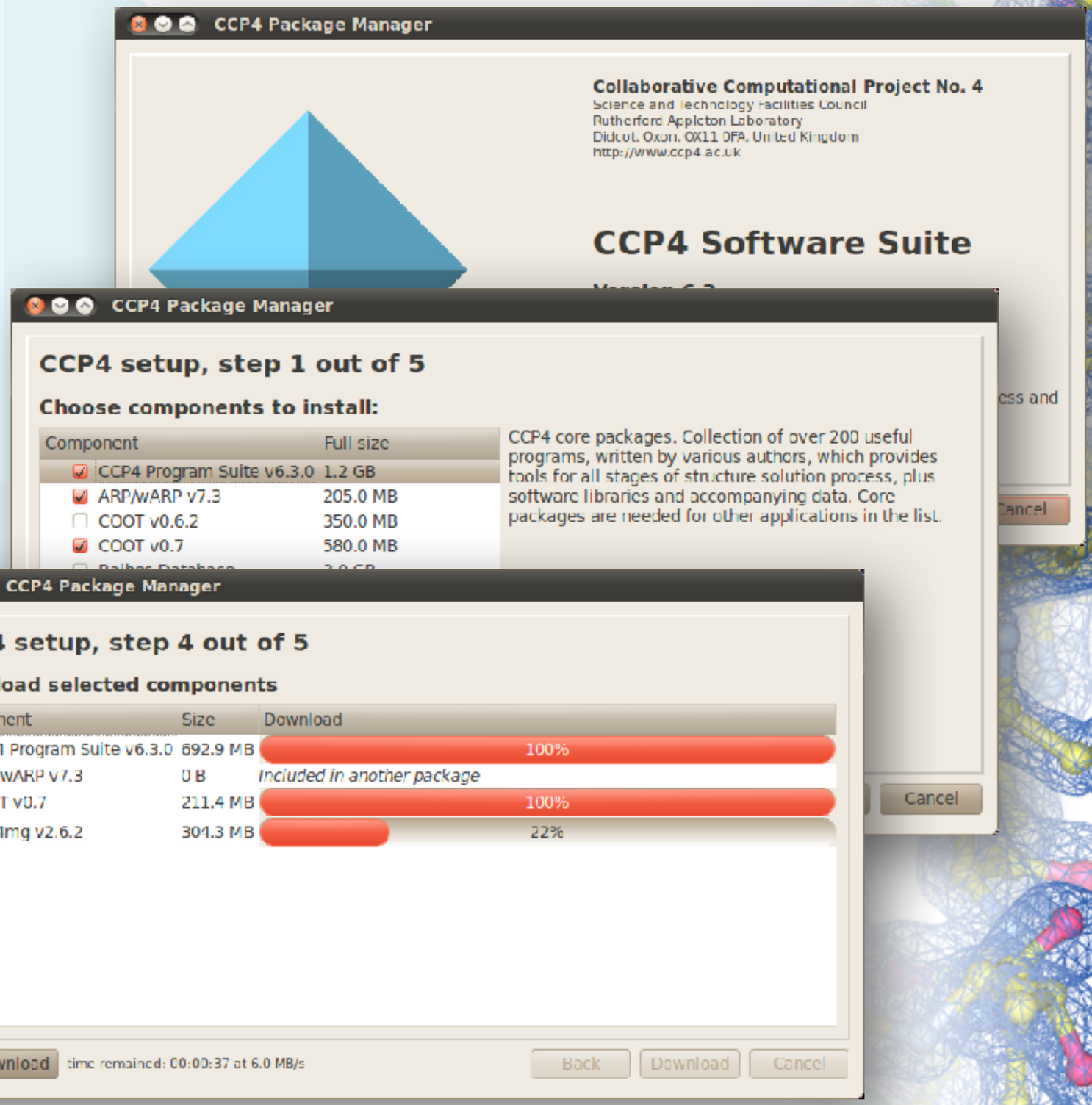
http://www.ccp4.ac.uk/ccp4online



Science & Technology Facilities Council
Research Complex at Harwell

CCP4 Setup Manager

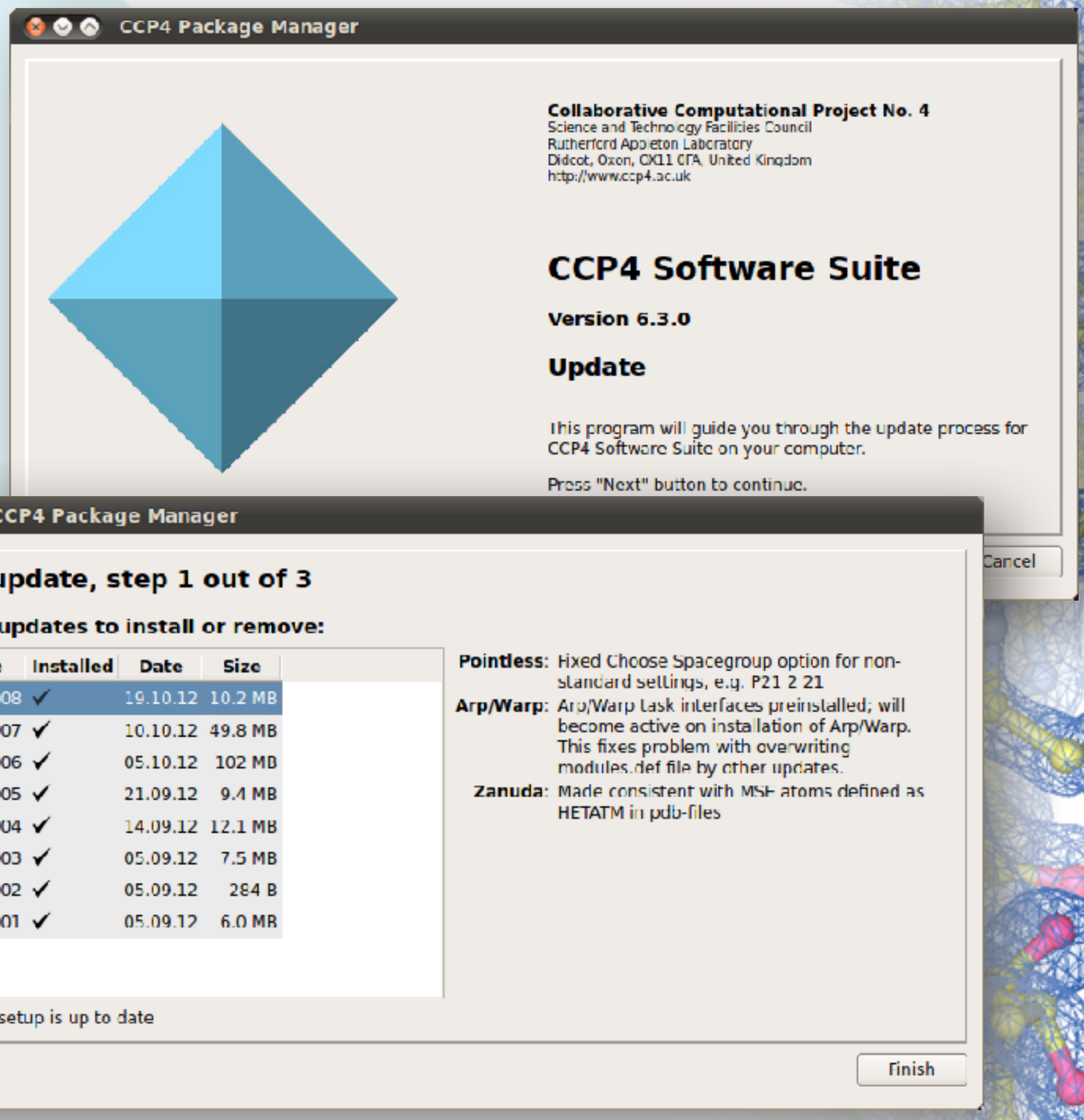
- Easy installation mechanism for CCP4 and associated packages (Coot, ARP/wARP, CCP4mg)
- Available on Linux and Mac OSX platforms
- Sets up system variables automatically if you wish



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



Acknowledgements

CCP4 Core group: Andrey Lebedev, Charles Ballard, Ronan Keegan, David Waterman, Marcin Wojdyr, Ville Uski, Karen McIntyre, Oleg Kovalevskyi, Maria Fando, Kyle Stevenson, Tarik Drevon, 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, Massimo Sammito

YSBL York: Keith Wilson, Eleanor Dodson, Kevin Cowtan, Liz Potterton, Stuart Mcnicholas, Paul Bond, Jon Agirre, Huw Jenkins

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

Global Phasing Ltd: Gerard Bricogne, Clemens Vonrhein

Leiden: Raj Pannu, Pavol Skubak

Others: Bernhard Lohkamp, Ruslan Sanishvili, Frank Von Delft, Martin Noble, Jaclyn Bibby, Daniel Rigden, Isabel Uson, Jens Thomas, Alun Ashton, David Brown, Arwen Pearson, Tim Gruene, George Sheldrick, Ivo Tews, Robbie Joosten, James Foadi, Arnaud Basle, Michael Isupov, Nick Keep, Paula Salgado, Judit Debreczeni, Dave Lawson, Claudia Millan Nebot, Victor Lamzin, Grzegorz Chojnowski, Andy Purkiss ***and many, many more …***

Science & Technology Facilities Council

Research Complex at Harwell

CCP4 People (core and associated developers)


Martyn Winn


Eugene Krissinel

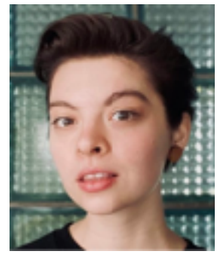

Charles Ballard


Ronan Keegan


David Waterman

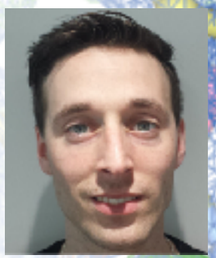

Andrey Lebedev


Oleg Kovalevskyi


Maria Fando


Kyle Stevenson


Ville Uski


Tarik Drevon

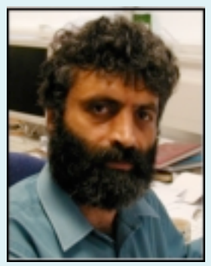

Keith Wilson


Eleanor Dodson


Liz Potterton


Stuart McNicholas


Kevin Cowtan

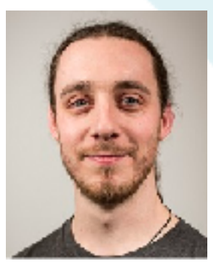

Garib Murshudov


Alexei Vagin


Fei Long


Hew Jenkins


Adam Simpkin


Paul Bond


Pavol Skubak


Robert Nichols


Daniel Rigden


Graeme Winter


Gwyndaf Evans


Phil Evans


Andrew Leslie


Harry Powell


Owen Johnson


Jon Agirre


Robbie Joosten


Randy Reed


Isabel Uson


Martin Noble


Paul Emsley


Raj Pannu


Luis Fuentes-Montero


Chris Morris

Melanie Vollmar