

Tom Burnley

An overview of CCP-EM

EMBO Image Processing Course 2017



Science & Technology
Facilities Council



What is a CCP?

- ‘Collaborative Computational Project’
- UK initiative running since 1978
- Aim to tackle large-scale scientific software development, maintenance and distribution... encouraging widespread and long term use
- Provide testing, distribution, dissemination and training...

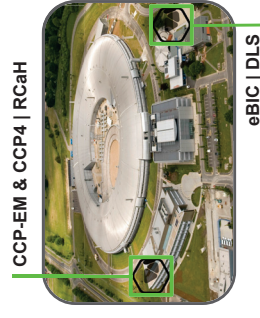
Current CCPs

CCP4	Macromolecular Crystallography
CCP5	The Computer Simulation of Condensed Phases
CCP9	Computational Electronic Structure of Condensed Matter
CCPEngSci	High Performance Computing in Engineering
CCP-BioSim	Biomolecular Simulation at the Life Sciences Interface
CCP-EM	Electron Cryo-Microscopy
CCPI	Tomographic Imaging
CCPN	NMR
CCP-NC	NMR Crystallography
CCP-Plasma	Computational Plasma Physics
CCPQ	Quantum Dynamics in Atomic, Molecular and Optical Physics
CCP-SAS	Analysis of Structural Data in Chemical Biology and Soft Condensed Matter
CCPForte	Collaborative Software Development Environment Tool
CCPET/IR	Positron Emission Tomography (PET) / Magnetic Resonance (MR) Imaging
CCP CoDiMa	Computational Discrete Mathematics
CCP-WSI	A Collaborative Computational Project in Wave/Structure Interaction
CCPmag	Computational Magnetism



Collaborative Computational Project for
Electron cryo-Microscopy

STFC located at Research Complex at
Harwell



Tom
Burnley



Colin
Palmer



Martyn
Winn



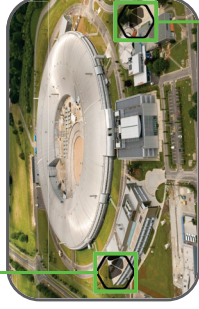
Research Complex
at Harwell

Funded by Medical Research Council since 2012

Support users and developers in computational aspects of biological EM

1. Nurture UK EM community
2. Support software users
3. Support developers

CCP-EM & CCP4 | RCaH



eBIC | DLS



Collaborative Computational Project #4

Focused on macromolecular crystallography

STFC Scientific Computing Department, RCaH

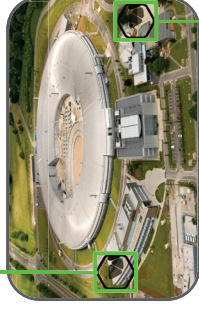
Formed in 1979

Supported 200+ applications

~4000+ citations

140 industrial licenses

CCP-EM & CCP4 | RCaH



eBIC | DLS

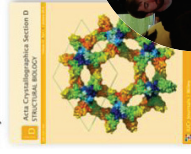


CCP-EM Spring Symposium

- **Aim: Building EM community**
- **UK cryoEM annual conference since 2015**
- **John Briggs, Yifan Cheng, Rado Danev, Kay Grunewald, Erik Lindahl, Piet Gros, Paul Adams, Maya Topf, Greg McMullan, Lori Passmore, Tanmay Bharat, Carolyn Moores, Ausra Domanska, Rob Nicholls, David Waterman, Tristan Croll, Morgan Beeby, Becky Thompson**

June 2017 issue

Proceedings of the CCP-EM Spring Symposium
Guest editors: Tom Burnley and Paula da Fonseca
Early view articles

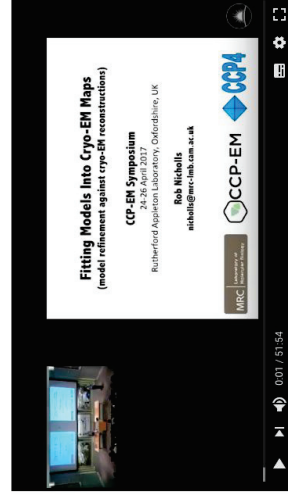


The revolution will be televised

- 2015-2017 talks on YouTube
- STFC CCP-EM channel

ccpem rob nicholls

ccpem rob nicholls



CCP-EM Rob Nicholls | Spring Symposium 2017

27 VIEWS

1

0

SHARE

...

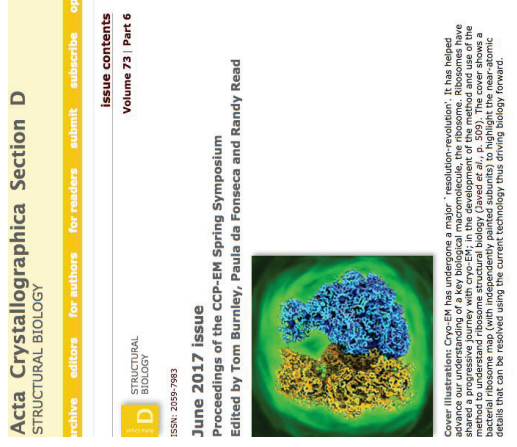


Science and Technology Facilities Council
Published on 13 Jun 2017

SUBSCRIBE 8.1K

The revolution will be televised

- Proceedings published in Acta Cryst. D
- 2016 edition in print
- 2017 edition in progress



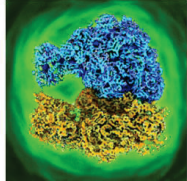
Acta Crystallographica Section D
STRUCTURAL BIOLOGY

archive editors for authors for readers submit subscribe or

Issue contents
Volume 73 | Part 6

ISSN: 2059-7953

June 2017 issue
Proceedings of the CCP-EM Spring Symposium
Edited by Tom Burnley, Paula da Fonseca and Randy Read



Cover illustration: Cryo-EM has undergone a major "resolution-revolution". It has helped to understand the structure of a wide range of biological macromolecules. The cover shows a progressive journey with cryo-EM, in the development of the method and use of the method to understand ribosome structural biology (Javed et al., p. 509). The cover shows a ribbon diagram of the ribosome structure, with the cryo-EM density map overlaid. The density map is shown in a color gradient from blue to red, indicating the resolution of the map. The ribbon diagram is shown in yellow and orange, indicating the atomic structure of the ribosome. The cover also features a 3D molecular model of a protein structure, which is shown in a color gradient from blue to red, indicating the resolution of the map.

Save the data...



2018 9-11th April | Keele University

CCP-EM Training Workshops

- *Aim: Support users*
- 10+ training workshop since 2014
 - SPR
 - Model building
 - Subtomogram averaging
- See CCP-EM mailing list for announcements
 - jiscmail.ac.uk/ccpem
 - 1600 subscribers
 - Main resource for RELION guidance
 - Forum for general EM related questions



CCP-EM Next Events

- **Dynamo subtomogram processing w/ Daniel Castano-Diez**
 - 27-29th November
 - RAL / DLS, UK
- **Annual Icknield high resolution model building workshop**
 - Spring 2018
 - RAL / DLS, UK
- **Spring Symposium IV 9-11th April**
 - Keele University, UK



Integration with DLS & eBIC

Supporting users and staff at Electron
Bio-Imaging Centre



7 workshop hosted at DLS with support from
Data Analysis software group

Provide software support for eBIC

- Software installation and maintenance
- RELION benchmarking and GPU setup
- Hot desk at eBIC Wednesday afternoon



CCP-EM Software Suite

Collaborating Developers



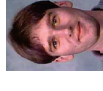
Garib
Murshudov



Maya
Topf



Agnel
Joseph



Alan
Roseman



Ardan
Patwardhan



Alexei
Vagin



Paul
Emsley



Sjors
Scheres



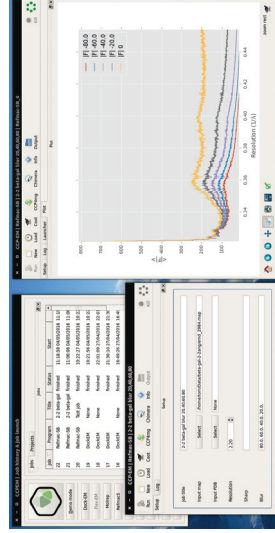
Jude
Short

- *Aim: supporting developers*
- *Unified access to many cryo-EM programs*
- *Linux / Mac / Windows...*
- *Close compatibility with CCP4*
- *Rapid testing and distribution*
- *Easy for users / easy for developers*



CCP-EM Beta release

- **Beta release for Linux and Mac now available**
- **ccpem.ac.uk**
- **Free for academic use**
- **Requires CCP4 license**
- **Bugs & requests:**
 - ccpem@stfc.ac.uk



Tom
Burnley



Colin
Palmer

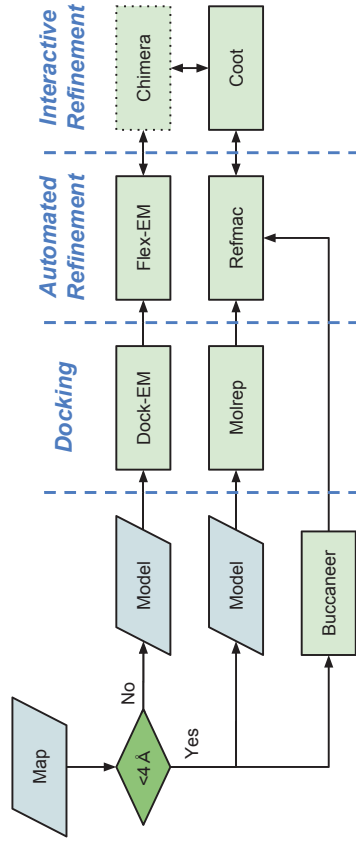


Martyn
Winn

CCP-EM Install

```
> cp ~/Downloads/ccpem-180517ap.tgz .
> tar xvf ccpem-180517ap.tgz
> cd ccpem
> ./CCPEM-BINARY.setup
> source ./bin/ccpem.setup-sh
> ccpem
```

CCP-EM: Model building pipeline



'Recent developments in the CCP-EM software suite' | Acta D | 2017

Docking and model preparation

CCP-EM: DockEM

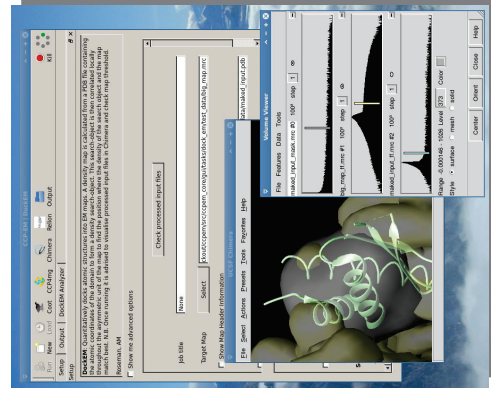
- Docking atomic models at medium to low resolution
- Exhaustive 6D rigid body search
- Target region of interest
- Solutions ranked by cross-correlation coefficient (CCC)
- View best hits in Chimera



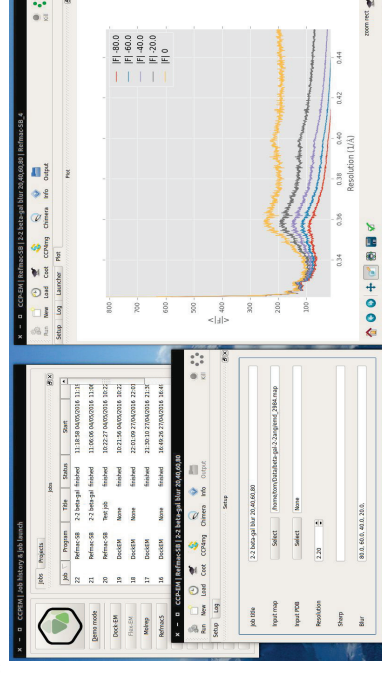
Alan Roseman



Sony Malhotra

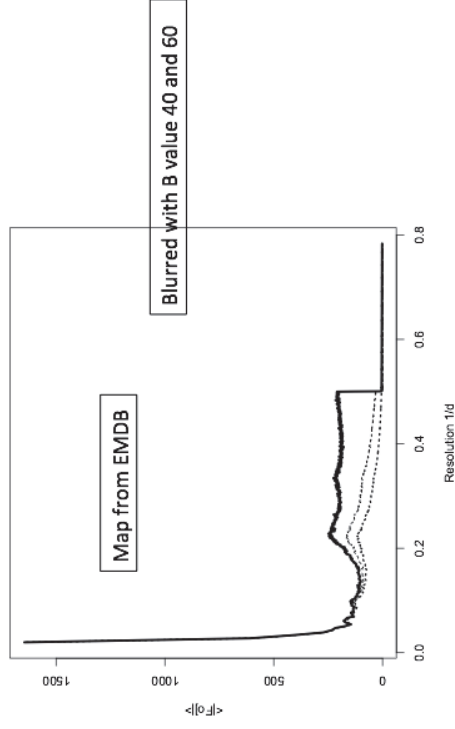


Optimise map sharpening

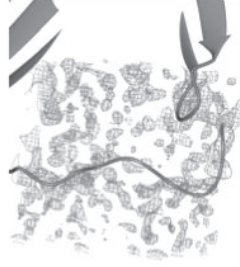


- MRC to MTZ interface
- Enter an array of sharpening factors
- Inspect Wilson plot and visualize maps in Coot
- Expect local variation

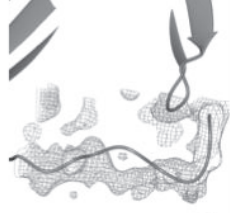
$\langle |F| \rangle$ vs resolution



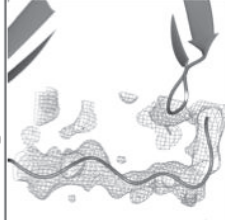
Map from PDB



Blurred: B = 60

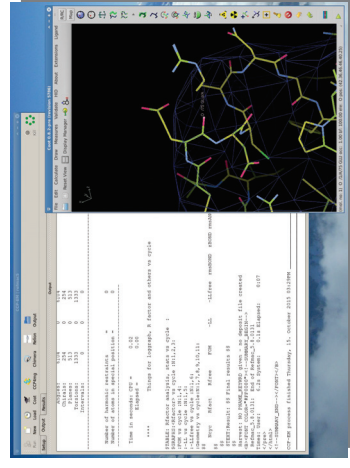


Model refined against blurred map



CCP-EM: Molrep

- Fast docking of molecular models into high resolution maps
- Originally developed for MX molecular replacement
- Optimised for use in EM
- Require map and model
- Find multiple copies
- Adjust map scale

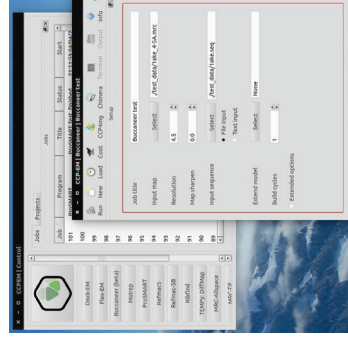


Alexei
Vagin



CCP-EM: Buccaneer & Nautilus

- Automated model building
 - Buccaneer: amino acid
 - Nautilus: nucleic acid
- Requires map and sequence
- Statistical chain tracing
- Identifies connected alpha-carbon positions using a likelihood-based density target
- Suitable for resolution $< 4 \text{ \AA}$



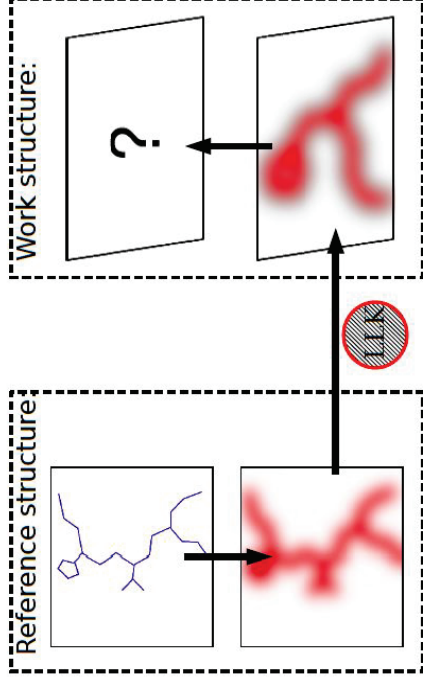
Scott
Hoh



Kevin
Cowtan

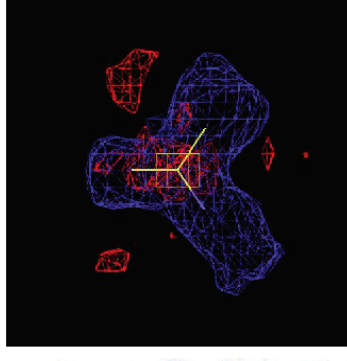
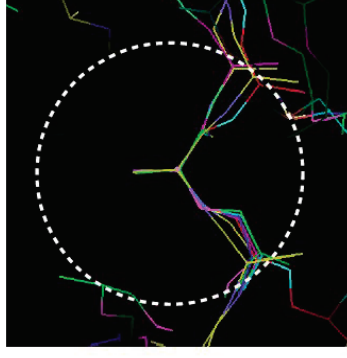
CCP-EM: Buccaneer Method

- Compare simulated map and known model to obtain likelihood target, then search for this target in the unknown map.



CCP-EM: Buccaneer Method

- Compile statistics for reference map in 4Å sphere about C α => LLK target.



- Use mean/variance.

4Å sphere about Ca also used by 'CAPRA' loeger et al. (but different target function).

CCP-EM: Buccaneer Method

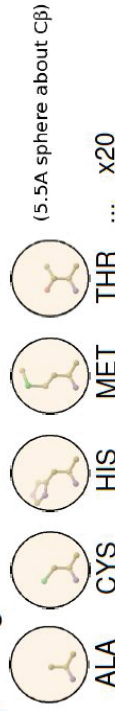
Use a likelihood function based on conserved density features.

The same likelihood function is used several times. This makes the program very simple (<3000 lines), and the whole calculation works over a range of resolutions.

Finding, growing: Look for C-alpha environment



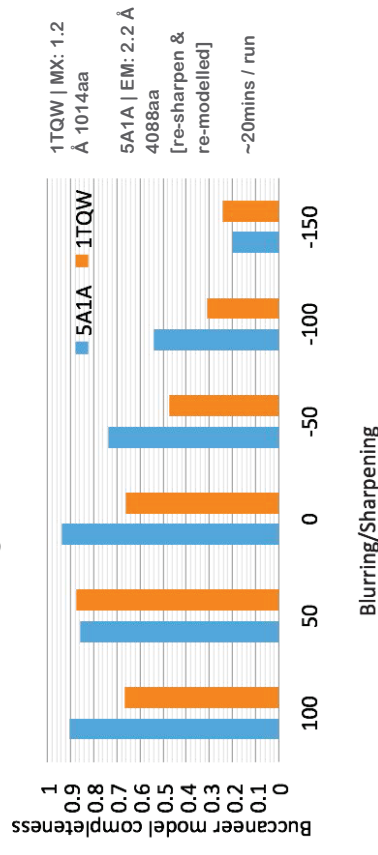
Sequencing: Look for C-beta environment



Haemoglobin test case

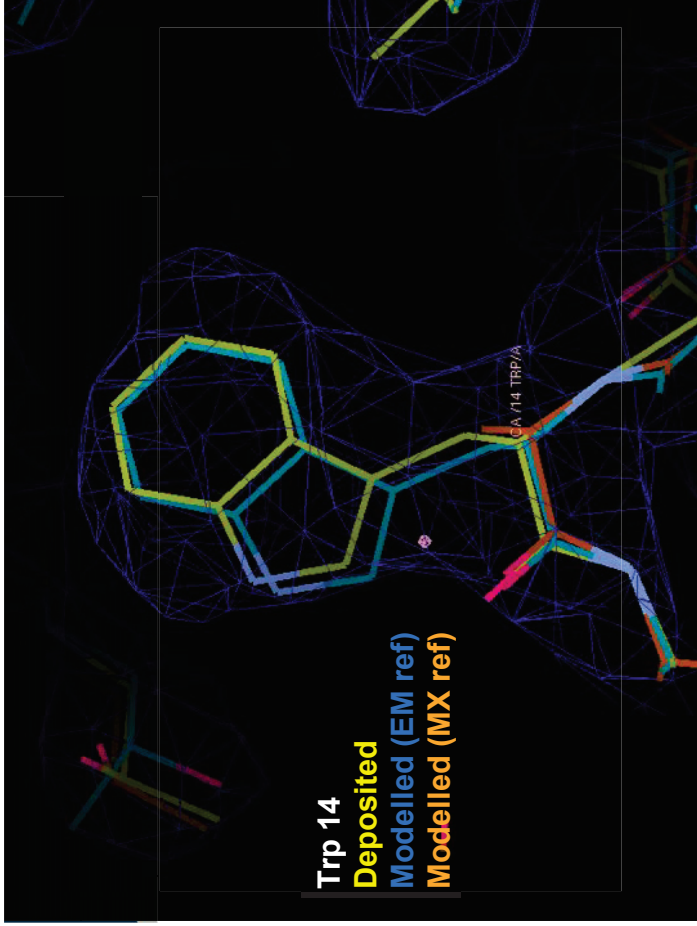
Rebuilding know structure: Danev haemoglobin map 3.2 Å (5N11)

Matching residues built

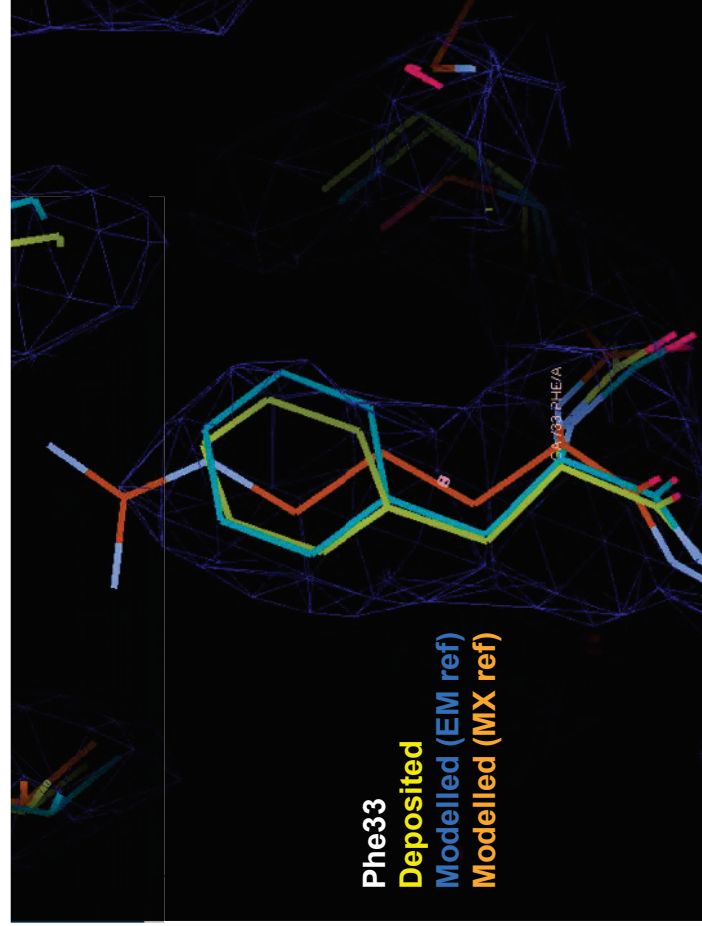


Trp 14

Deposited
Modelled (EM ref)
Modelled (MX ref)



Phe33
Deposited
Modelled (EM ref)
Modelled (MX ref)



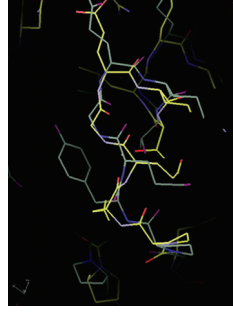
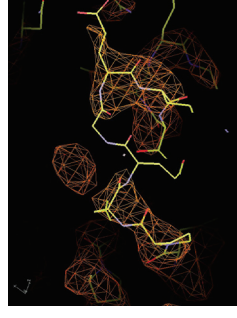
Phe33
Deposited
Modelled (EM ref)
Modelled (MX ref)

CCP-EM: Buccaneer Tidy up

Afterward running: inspect and tidy up
with Coot

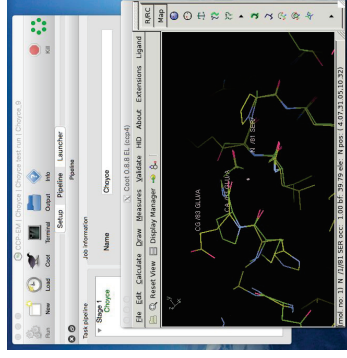
Typical Coot steps:

- Connect up any broken chains
- Use density fit and rotamer analysis to check rotamers
- Check Ramachandran, molprobrity, etc
- Add waters, ligands, check un-modeled blobs



CCP-EM: Choyce

- Homology modelling
- Modeller 'automodel' class
- Requires template structure and target sequence
- Align, build, energy minimize
- Convenience feature for high-resolution modelling
- Future: multiple templates, automated docking and refinement



Maya
Topf

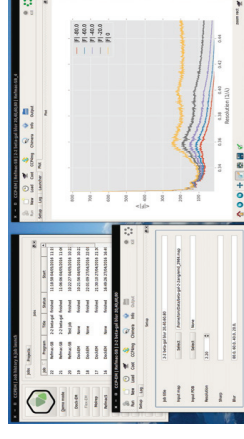
Modeller

Program for Comparative Protein Structure Modelling by Satisfaction of Spatial Restraints

Automated model refinement

CCP-EM: Refmac5 EM Pipeline

- High resolution $< \sim 5 \text{ \AA}$
- Original developed in CCP4 for MX
- Vector difference refinement
- Global or local refinement modes
- Map to mtz conversion
- Find in map (Molrep)
- Map sharpening



Garib
Murshudov



Oleg
Kovalevskiy



Rob
Nicholls

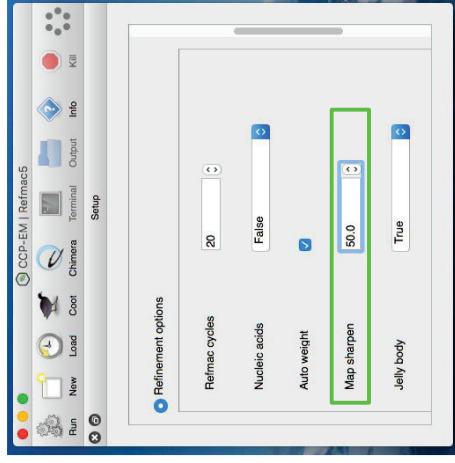
'Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions' | Acta D | 2017

**Model should agree with
the observed data**

**Model must be chemically
and structurally sensible**

Map Sharpening

- Set overall sharpening factor in Refmac
- MRC to MTZ conversion
- Sharpening / blurring can have dramatic effect on map quality



Half map validation

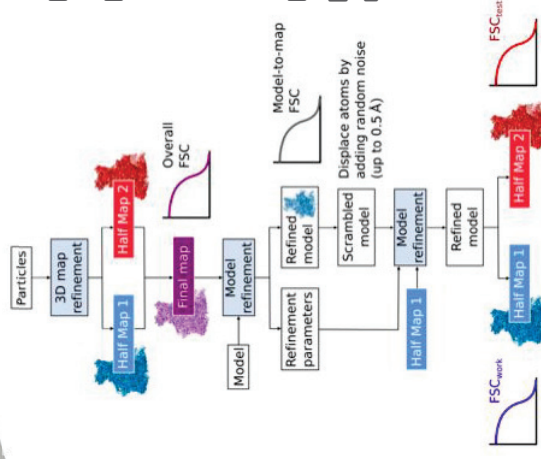
Refmac validation pipeline

Requires 3 input maps

- Final map
- Half map 1
- Half map 2

Refinement protocol

performed twice; 1) vs Final map, 2) vs HM1

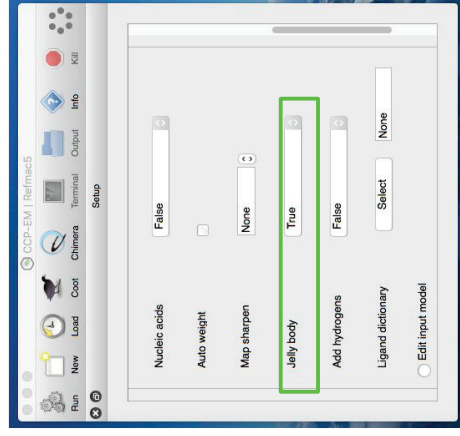


Additional restraints

- Jelly body restraints

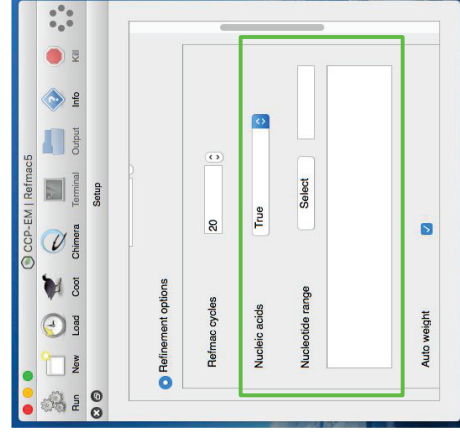
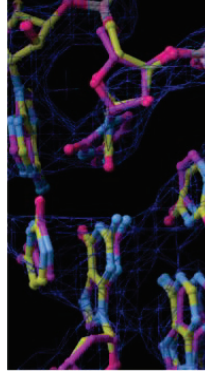
$$\sum_{\text{pairs}} w(|d| - |d_{\text{current}}|)^2$$

- During refinement inter-atom distances do not change significantly
- Distance pair 'springs'
- ~'implicit normal modes'



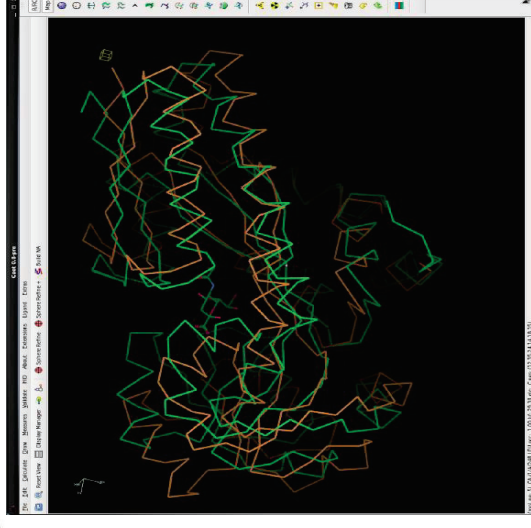
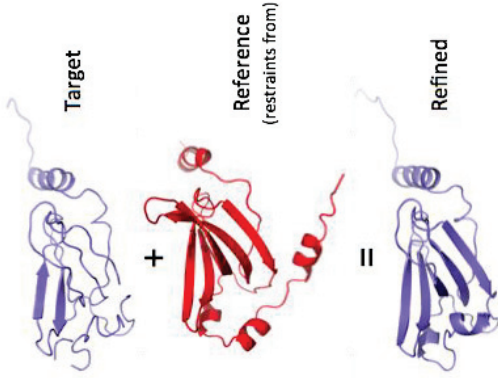
Additional restraints

- Nucleic acid restraints: LIBG
- Base-pair, parallelization and pucker restraints for nucleic acids



Additional restraints

- External restraints: ProSMART
- Requires 'good' reference structure
 - High resolution
 - Good stereochemistry
 - High structural homology
- CCP-EM ProSMART interface / Coot

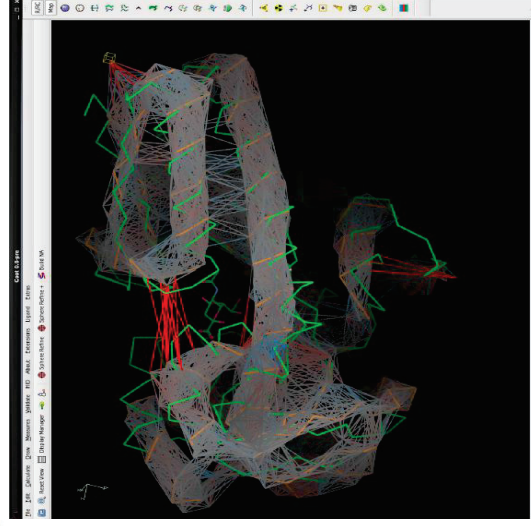


- View / edit in Coot

Additional restraints

Additional restraints

- View / edit in Coot



Additional restraints

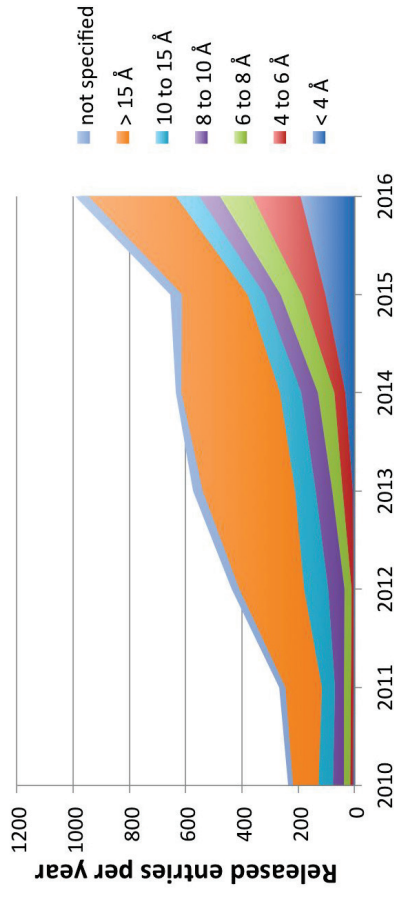
More information

CCP-EM Spring Symposium Lectures

- **Rob**
<https://www.youtube.com/watch?v=Lm5X6B10wIU>
- **Garib**
<https://www.youtube.com/watch?v=DTMPiZHAzPk>

Not all maps are high resolution...

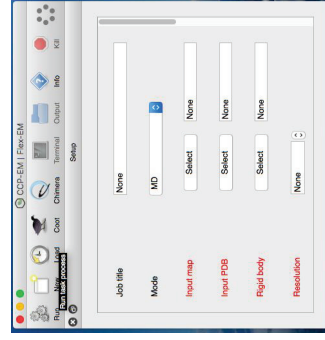
Resolution trends of released EMDB entries



Ardan Patwardhan, EMDDB

CCP-EM: Flex-EM

- Flexible fitting of rigid body domains to EM maps
- Medium to low resolution datasets
- MD simulated annealing with real space experimental restraints
- Requires map, pdb and rigid body definitions



Agnel Joseph



Maya Topf

"Refinement of atomic models in high resolution EM reconstructions using Flex-EM and local assessment" | Methods | 2016

CCP-EM: Ribfind

- Generate rigid body definitions for Flex-EM
- Requires PDB
- DSSP to identify secondary structure elements
- Inspect and manually adjust cluster cut-off to control the number of SSE in each rigid body

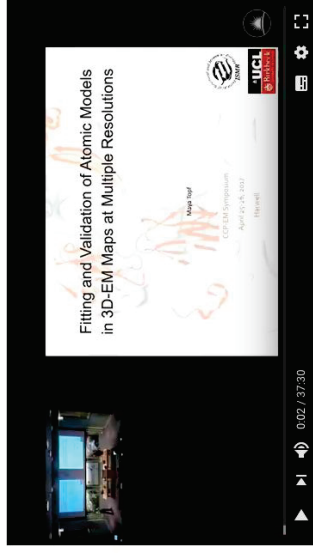


More information

CCP-EM Spring Symposium Lectures

Fitting and validation and multiple resolutions

- Maya
- <https://www.youtube.com/watch?v=v898HMTDh8>



CCP-EM Maya Topf | Spring Symposium 2017
24 views

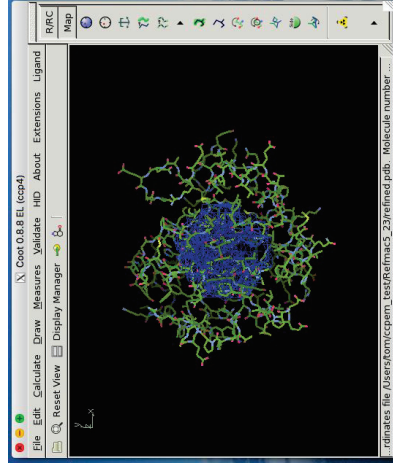


Science and Technology Facilities Council
Published on 13 Jun 2017

SUBSCRIBE 8.1K

Coot: Interactive refinement

- Molecular graphics tool for interactive refinement and validation of models
- Extremely useful for high-resolution maps (2-4 Å)
- Visual inspection of fit to density
- Integrated into CCP-EM

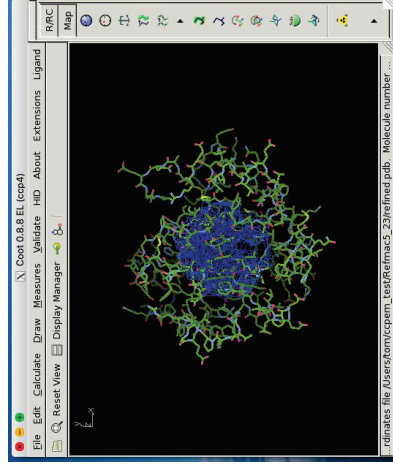


Paul Emsley

Interactive model refinement

Coot Utilities

- Manual local real space refinement
- Add / modify residues
- Search and fit side chain rotamers
- Find and fix Ramachandran outliers
- Atomic clash markup
- Check B-factor distribution
- Please see *Icknield Coot* lecture and tutorial for more details



Paul Emsley

High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective

Alexander Wlodawer,^{1,4,*} Mi Li,^{1,2} and Zbigniew Dauter³

Another observation common to almost all the deposited models based on high-resolution maps is that they seem to lack the final quality control. The presence of very doubtful multiple conformations of the side chains, poor geometry of the model in comparatively clear regions of the maps, location of the side chains outside of the clear density, or the occurrence of interatomic clashes may indicate the difficulty of manual inspection of these very large structures....

Nevertheless, more attention needs to be paid to such problems that are not easily solved by purely automated means.

UK Validation Working Group

6 sites received Wellcome trust funding to start 2018

Map validation

Model validation

Map to model validation



welcometrust

CCP-EM: Other tools

```
13 import read_mrc_header
14
15 class ClipperXMap(object):
16     ...
17     XMap = Xray map w/ symmetry, info
18     ...
19     def __init__(self, filename):
20         self.filename = filename
21         self.clipper_xmap = mrc_map
22     ...
23     def print_stats(self):
24         ...
```

- clipper-python
 - Stuart McNicholas
 - Jon Agirre
 - Kevin Cowtan
 - University of York
- ConKit
 - Dan Rigden
 - Felix Simkovic
 - University of Liverpool
- TEMPY
 - Maya Topf
 - Birkbeck
- mrcfile
 - Colin Palmer
 - STFC
- MRC Image Processing System
 - Tony Crowther, Richard Henderson & Jude Short
 - MRC-LMB

Core library development

CCP-EM responsible for MRC data format (aka CCP4 map format)
mrcfile: stand alone open-source Python library for easy access to MRC files

NumPy integration

- pip install mrcfile
- <https://github.com/ccpem/mrcfile>
- <http://mrcfile.readthedocs.org>

```
>>> import mrcfile
>>> with mrcfile.open('filename.mrc') as mrc:
...     mrc.data
array([[ 2.58179283,  3.14060021, ...,  3.61837363],
...
...
...
... [ 3.63812474,  3.64495087, ...,  1.92690361]], dtype=float32)
```

Wrap-up

Download: ccpem.ac.uk

Mailing list: jiscmail.ac.uk/ccpem

Dynamo Workshop: Registration open

Spring Symposium: 2018 9-11th April

*Fellowship opportunity: STFC Rutherford
International Fellowship Programme*



Science & Technology
Facilities Council
10 Years of Impact and Inspiration



Science & Technology
Facilities Council

Acknowledgements

- CCP-EM core team
 - Martyn Winn
 - Colin Palmer
- CCP4 core team
- STFC SCD
 - Steven Lamerton
- DLS / eBIC
 - Alun Ashton
 - Peijun Zhang
- Birkbeck
 - Maya Topf
 - Agnel Joseph
- University of Manchester
 - Alan Roseman
- EBI
 - Gerard Kleywegt
 - Ardan Patwardhan
- MRC-LMB
 - Garib Murshudov
 - Sjors Scheres
 - Paul Emsley
 - Paula da Fonseca
- University of York
 - Kevin Cowtan
 - Soon Wen 'Scott' Hoh
 - Jon Agirre
- Francis Crick Institute
 - Peter Rosenthal



Science & Technology
Facilities Council