CCP4 Beyond Structure Solution

Eugene Krissinel

CCP4, STFC Research Complex at Harwell
Didcot, United Kingdom

eugene.krissinel@stfc.ac.uk

Macromolecular Crystallography School in Brazil, Sao Carlos, 8-16 April 2014

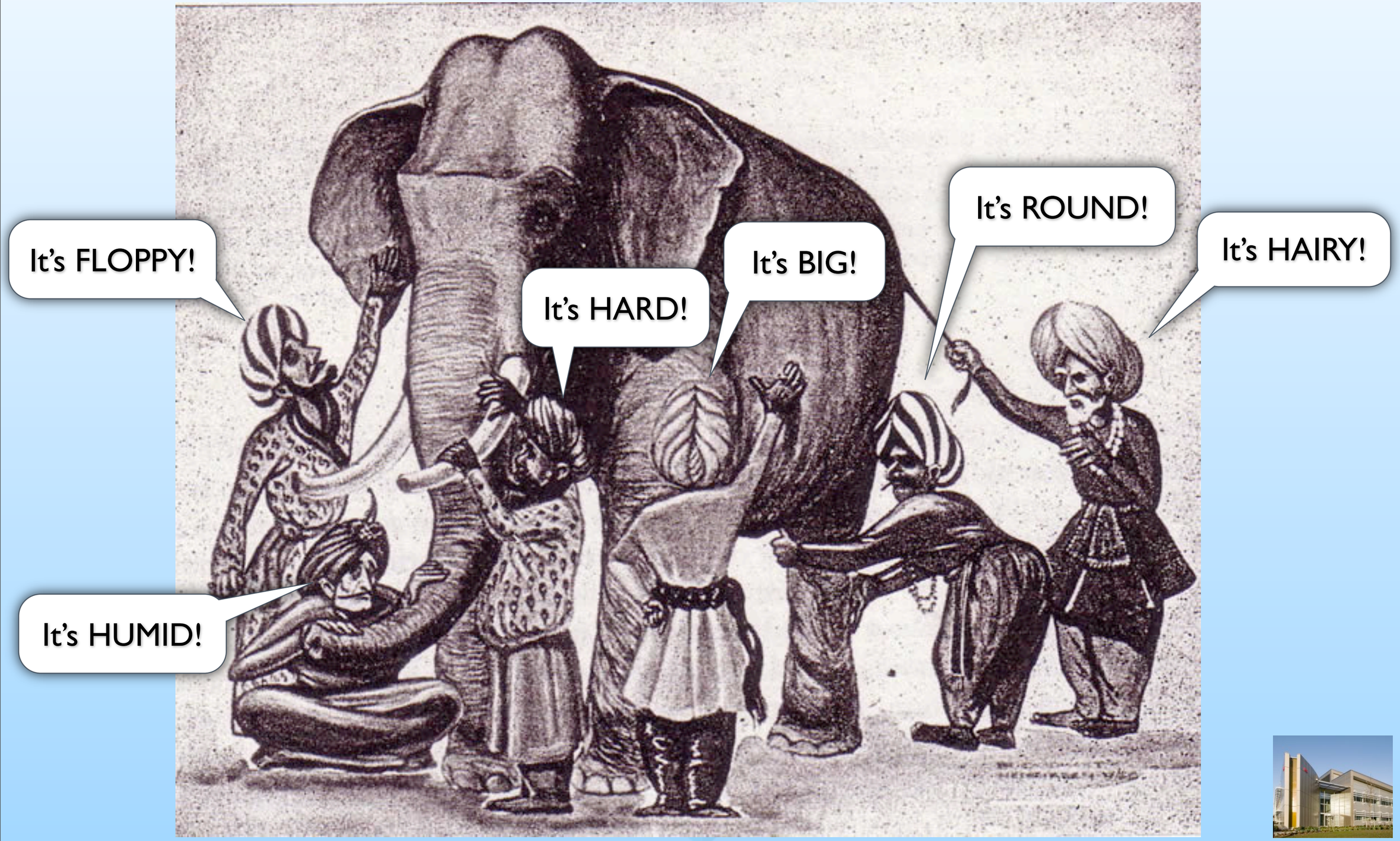
Research Complex at Harwell

What is CCP4?

- It is the Collaborative Computational Project No. 4 in Protein Crystallography
- 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

Research Complex at Harwell

So, what is CCP4?



Research Complex at Harwell

CCP4 is a Structure Solution Suite

- CCP4 contains about 250 program components
  - ~30% components directly related to structure solution

    MOSFLM, XIA2, POINTLESS, SCALA, MOLREP, AMORE, BEAST, REFMAC, BUCCANEER, PIRATE, DM, PARROT, SOLOMON, MRBUMP, BALBES, CRANK, PHASER, COOT etc.

    It's HARD!
  - ~45% components assisting structure solution

    CHAINSAW, HGEN, WATERTIDY, CTRUNCATE, PROFESS, LIBCHECK, SFALL, SKETCHER, JLIGAND, ECALC, RFCORR, DETWIN, SAPI, WATPEAK, SEQWT etc.

    It's BIG!
  - ~20% related to file transforms and various jiffies

    COORD_FORMAT, PDBSET, PDBCUR, PDB2TO3, COORDCONV, GENSYM, SYMCONV, PHISTATS, STEREO, MTZDUMP, PDB_EXTRACT, F2MTZ, MTZ2CIF, MTZ2VARIOUS, MTZUTILS, SORTMTZ etc.

    It's ROUND!
  - ~5% post-structure solution analysis

    A relatively small subset of the Suite that may be used by structural biologists to aid their research when they have their structure solved or otherwise obtained

    It's useful?

Research Complex at Harwell

Structure Analysis in CCP4

Coordinate analysis

ACT	*general coordinate analysis: contacts, statistics, selection on B-factors etc*
CONTACT/NCONT	*calculation of various atomic contacts, h-bond candidates etc*
DISTANG	*calculation of interatomic distances and angles*
GEOMCALC	*calculation of various structure features such as planes, deviations from planes, interatomic distances, inter-plane angles, torsions, coordinate transforms and many more*

Domain identification

DYNDOM	*identification of dynamic domains when two protein conformations are available*

Quality checks

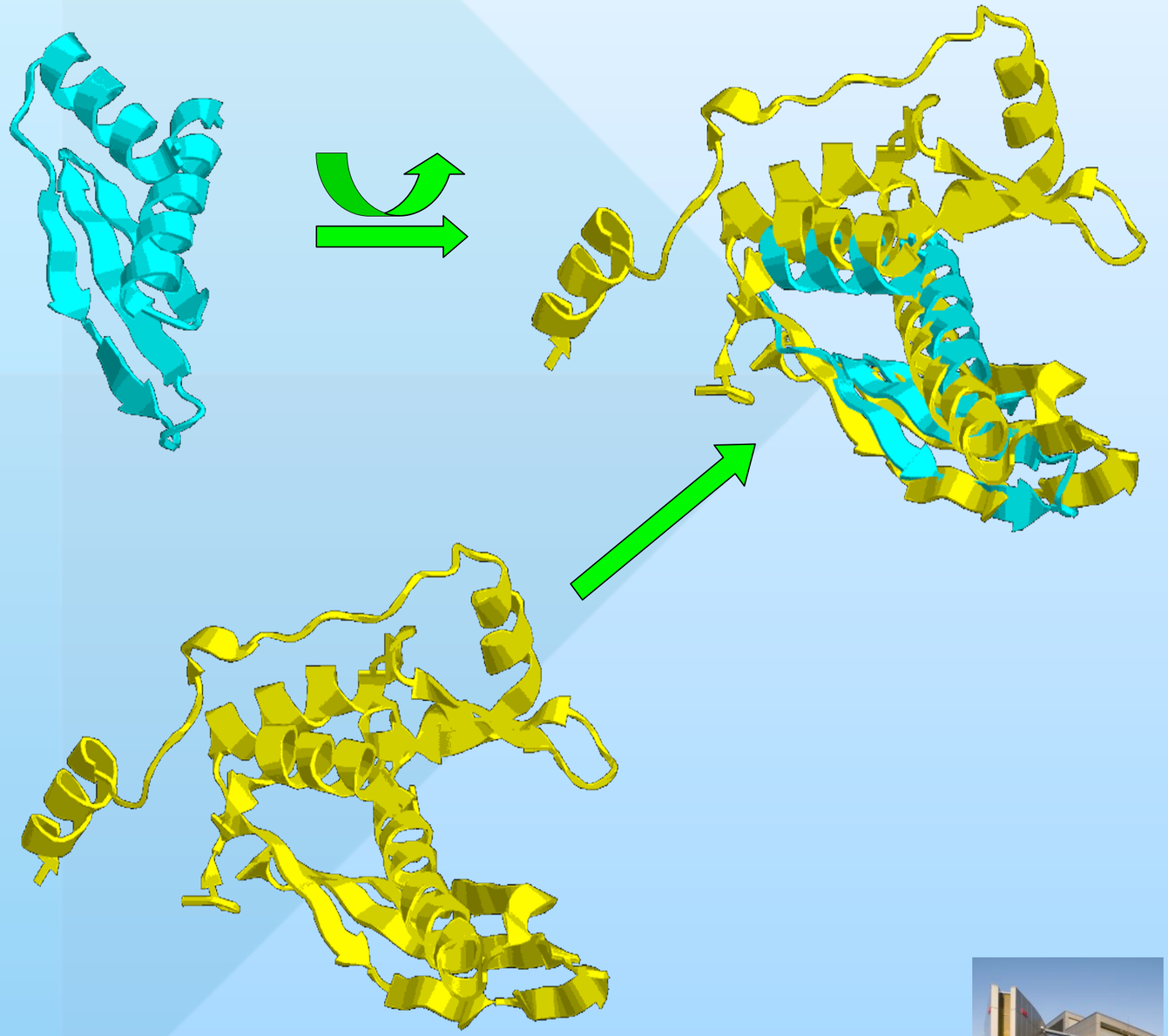
PROCHECK	*checks stereochemical quality of protein structurs*
ROTAMER	*identification of unusual rotamer configurations*
SFCHECK	*assessment of the agreement between structure and x-ray data*

Research Complex at Harwell

Structure Analysis in CCP4

- Structure alignment, superpositon and comparison

LSQKAB	*superposition of given coordinate subsets*
POLYPOSE	*multiple superposition of same-length structures (conformers)*
TOPP	*topological analysis of protein structures*
SUPERPOSE	*generic structure aligner and superposer, with automatic identification of superposing subsets, based on SSM algorithm*
GESAMT	*generic structure aligner and superposer, with automatic identification of superposing subsets, free from SSM limitations*
PROSMART	*structure aligner and superposer for the calculation of external crystallographic restraints*



Research Complex at Harwell

Superpose is also known as SSM@EBI

Structure alignment

http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver

Home RAL Mail SSC Oracle CCP4 wiki Google Mail Yahoo! Mail Staff Central Google Google Maps Wikipedia

EMBL-EBI Enter Text Here Find Help | Feedback

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

Home > Databases > PDBe > Services > SSM | contact PDBe

Submission Form for ⊙ **pairwise** ○ multiple **3D alignment**

explanation of input

Query	Target
Source: PDB entry PDB code 1sar view	Source: All PDB archive
Select chains Find chains Chains: *(all)	
Lowest acceptable match (%) 70	Lowest acceptable match (%) 70

☑ match individual chains ☑ best matches only
☑ match connectivity ☑ unique matches only
☑ if no matches within acceptibility limits found, show *some* of the close ones

Precision: normal Sort by: Q-score Viewer: Jmol

Submit your query Back to Home Page

SSM v2.36 << 13/03/2009 >>

last modified: 13/03/2009

Terms of Use | EBI Funding | Contact EBI | © European Bioinformatics Institute 2011. EBI is an Outstation of the European Molecular Biology Laboratory.

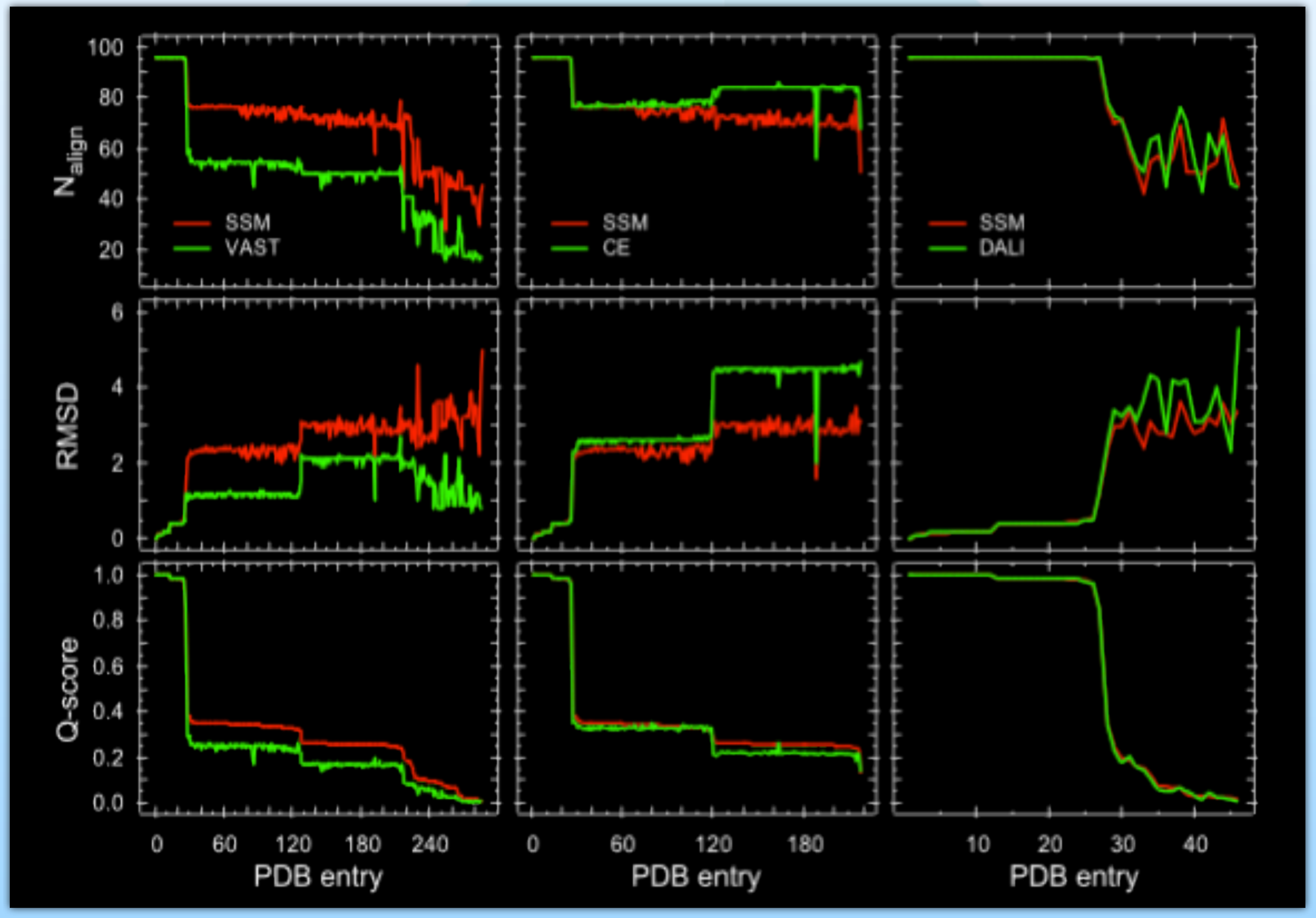
Research Complex at Harwell

Quality of Structure Alignment

- Often not very well known
- Depends on measure
- No agreement on measure

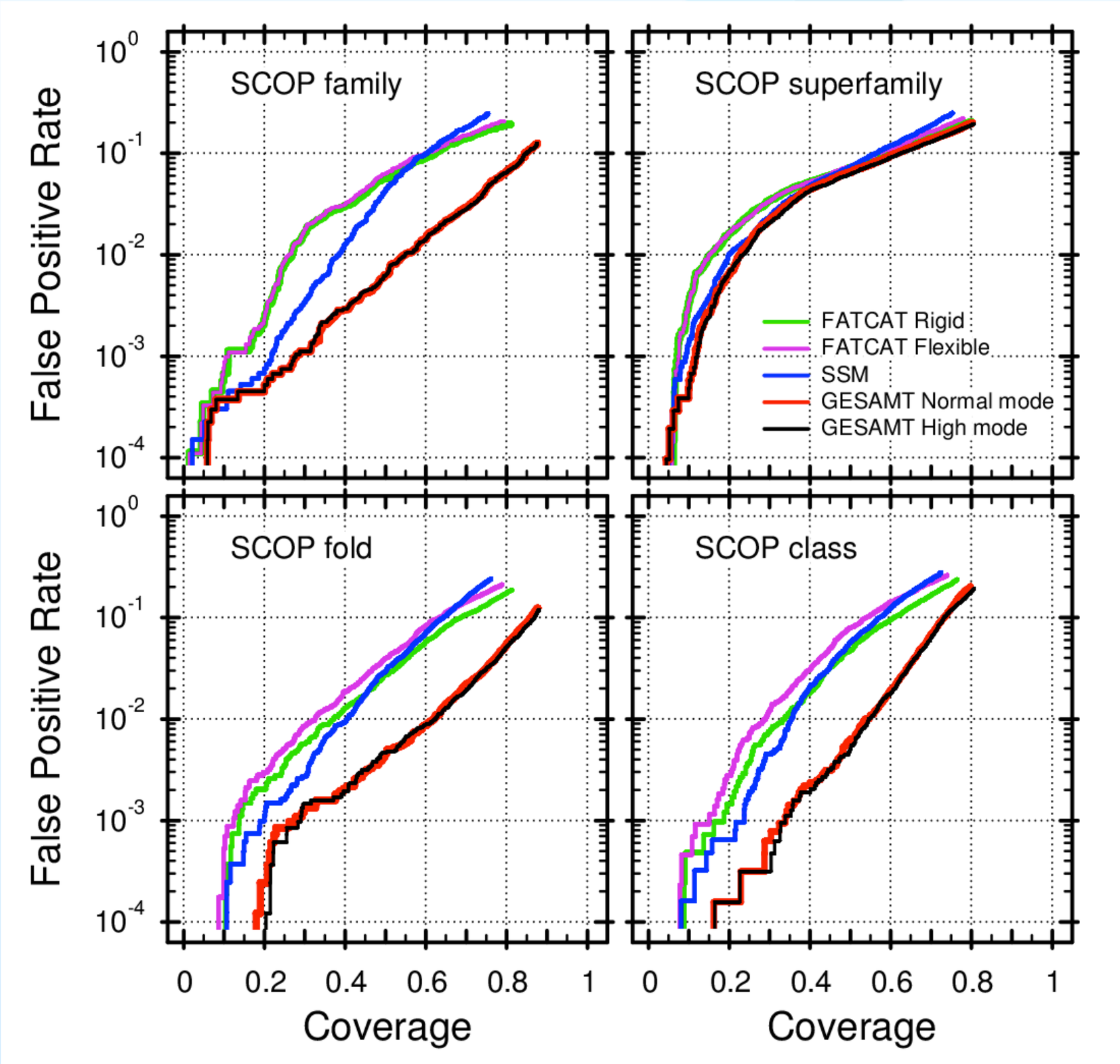
$$Q = \frac{N_{align}^2}{\left(1+\left(rmsd/R_0\right)^2\right)N_A N_B}$$

Acta Cryst D (2004), **60,** 2256-68



Research Complex at Harwell

Quality of Structure Alignment



$$FPR = \frac{FP}{TN + FP}$$

$$Coverage = \frac{TP}{TP + FN}$$

TP : True Positives
TN : True Negatives
FP : False Positives
FN : False Negatives

J. Mol. Biochem. (2012), **1,** 76-85

Research Complex at Harwell

Structure Analysis in CCP4

- Surface analysis

CAVENV	*"visualisation" of cavities (through the generation of a "cavity map")*
SC	*assessment of shape complementarity*
AREAIMOL	*surface and interface area calculations*
SURFACE	*assessment of surface accessibility*
VOLUME	*polyhedral volume of selected part of a protein*

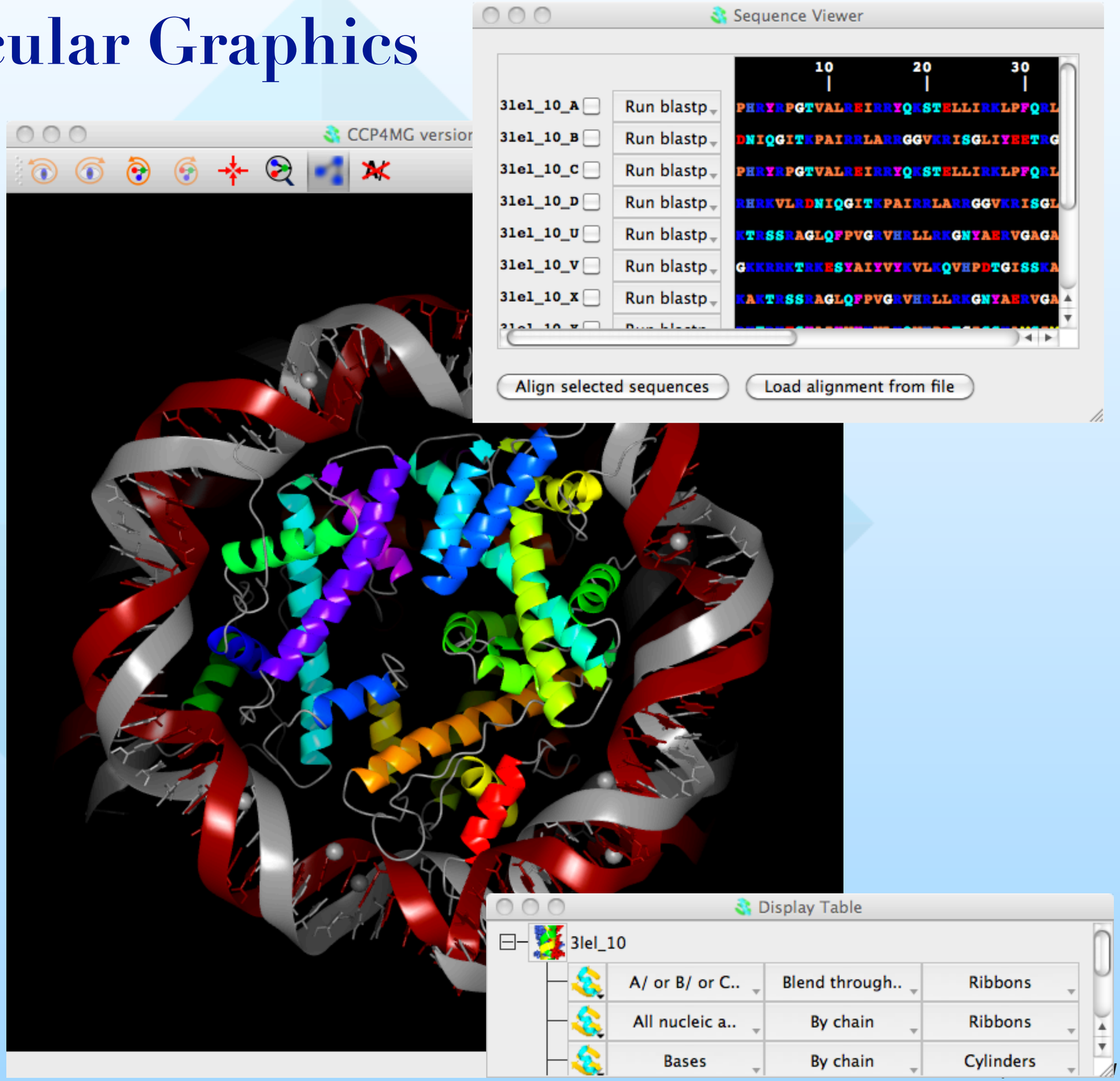
- Visualisation

TOPDRAW	*sketchpad for protein topology diagrams*
CCP4 MG	*advanced Molecular Graphics viewer*

Research Complex at Harwell

CCP4 Molecular Graphics

- Version 2.8.1: Top-Standard Molecular Imaging
  - fast and (almost) stable
  - highest publication-quality rendering
  - flexible image composition
  - rich set of tools and options
  - developed selection capabilities
  - integrated with other CCP4 components, such as SSM and PISA
  - connected to external web-services from NIH, PDB, Uppsala and EBI
  - stereo capabilities
  - integrated structure analysis
  - integrated sequence and structure alignment
  - normal mode analysis
  - movies
  - advanced Picture Wizard
  - open architecture with PyQt framework



CCP4 Molecular Graphics – Tutorial Contents

http://www.ccp4.ac.uk/MG/ccp4mg_help/tutorial/index.html

Home RAL Mail SSC Oracle CCP4Forge Google STFC CCP4 wiki Yahoo! Mail Google Mail Google Drive Google Maps Wikipedia Apple YouTube News Popular

CCP4 Molecular Graphics Tutorials

Tutorial Contents

Documentation Contents | On-line Documentation | Tutorials | CCP4mg Home

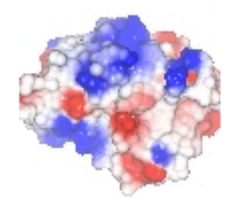
A Guided Tour of the Program

This is a tutorial covering most of the basic features of the program.

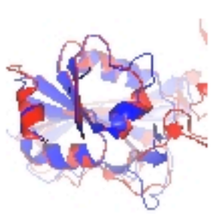
CCP4MG tutorial (PDF version)

Alternatively, there is a set of mini tutorials covering slightly more material. It is recommended that you follow the Introduction tutorial first, but you can pick and choose from the subsequent ones. They will probably take about ten minutes each.

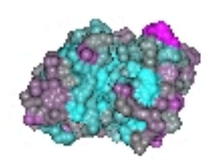
Introduction - loading and viewing PDB or mmCIF files. Essential reading.



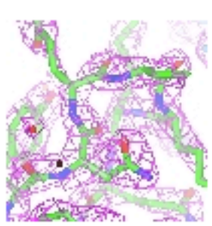
Surfaces - selecting and colouring and electrostatic surfaces



Superpose proteins - instamagic protein superposition



Importing Model Analysis Data



Electron density maps - better representation for images and packing diagrams

Adding Text and Extra Images

Structure Analysis in CCP4

- Protein interactions and complexes

  PISA *analysis of Protein Interactions, Surfaces and Assemblies*

... to be continued in next presentation

Research Complex at Harwell