Structure Analysis and Deposition Tools in CCP4

Eugene Krissinel

CCP4, STFC Research Complex at Harwell
Didcot, United Kingdom

eugene.krissinel@stfc.ac.uk

CCP4/DLS Workshop on Macromolecular Crystallography
Nov 28 - Dec 6, 2022, Harwell, UK

Research Complex at Harwell

CCP4 is a Structure Solution Suite

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

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

  - ~45% components assisting structure solution

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

  - ~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.

  - ~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 solved or otherwise obtained structures

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

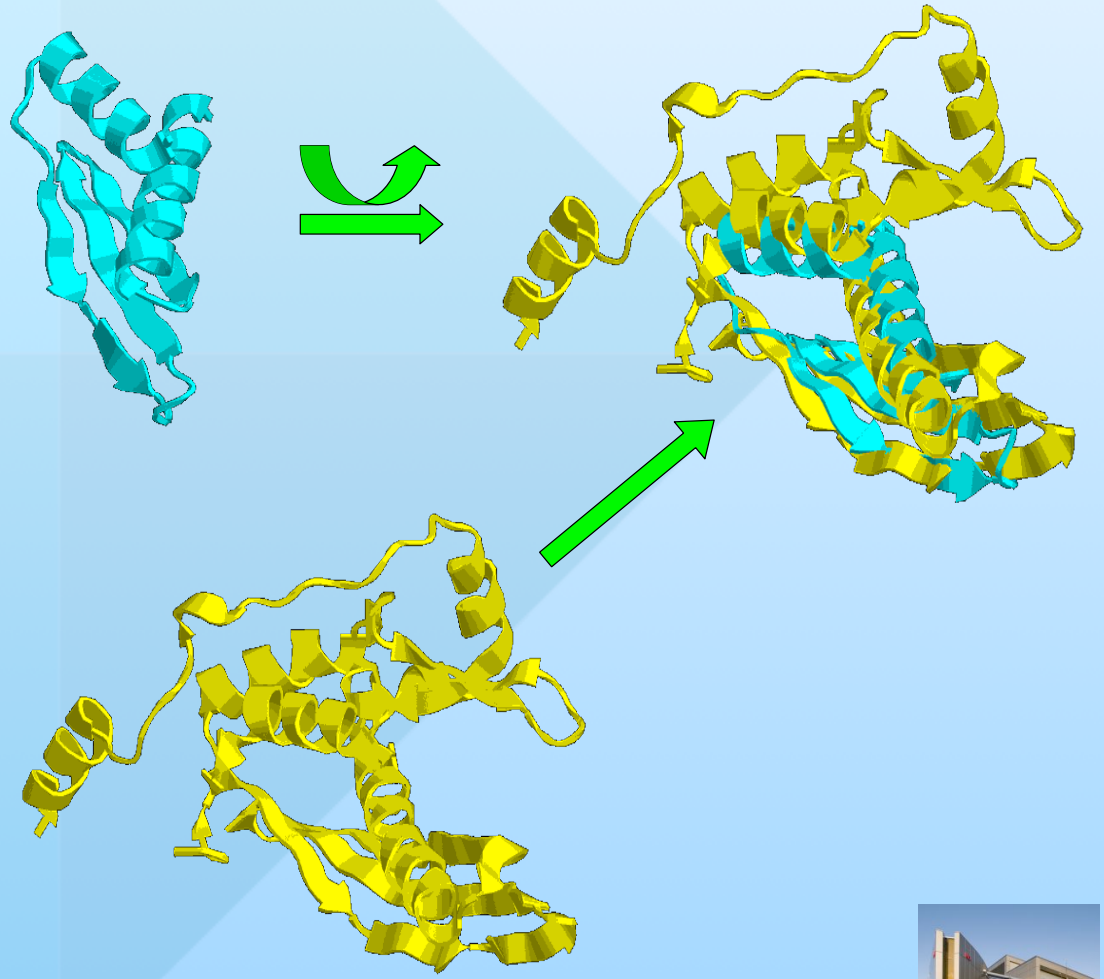
PROCHECK	*checks stereochemical quality of protein structures*
ROTAMER	*identification of unusual rotamer configurations*
SFCHECK	*assessment of the agreement between structure and x-ray data*

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*



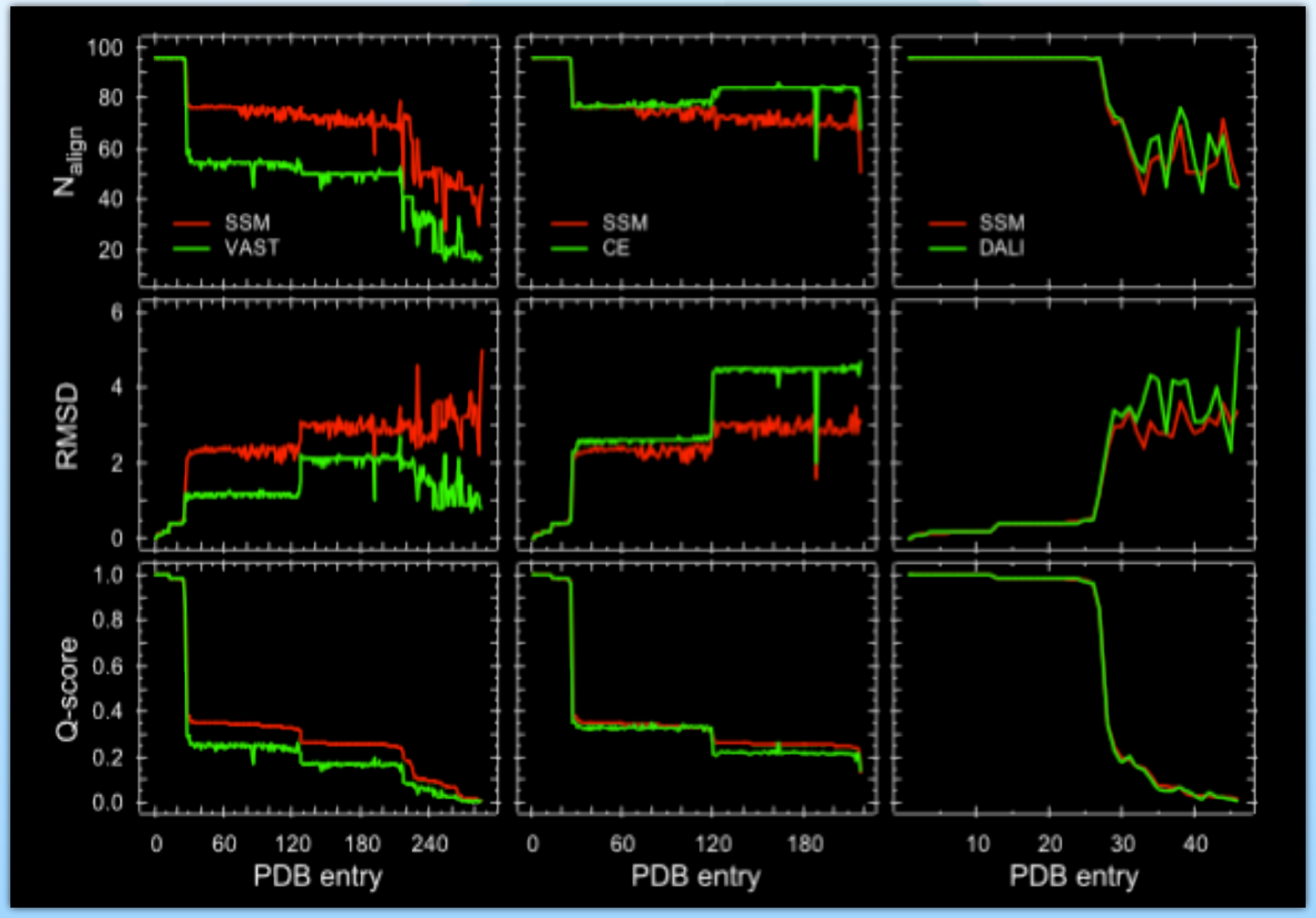
Research Complex at Harwell

Quality of Structure Alignment

- Often not very well understood
- Depends on measure
- No agreement on the measure

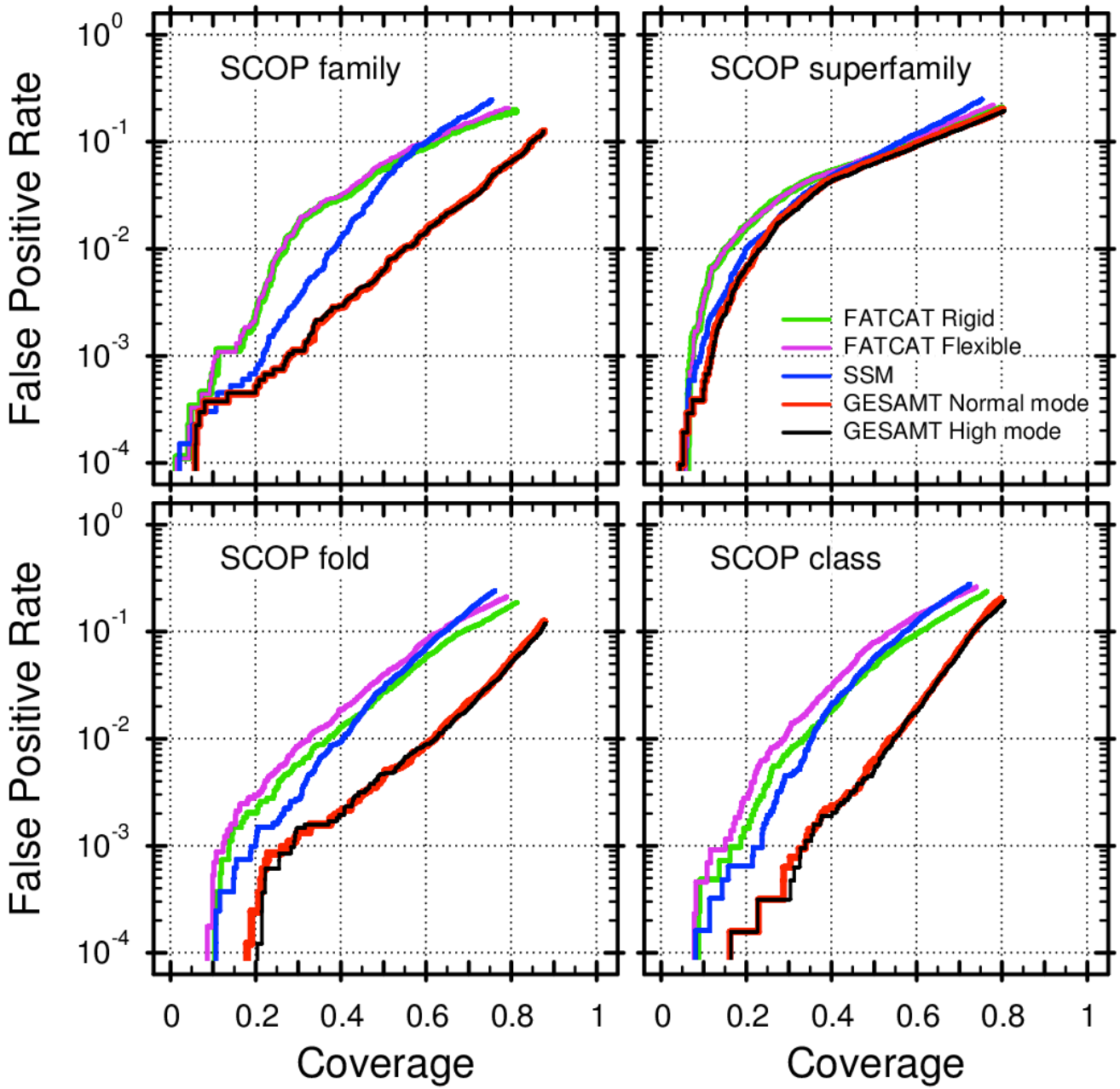
$$Q = \frac{N_{align}^2}{\left(1 + \left(rmsd/R_0\right)^2\right) N_A N_B}$$

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

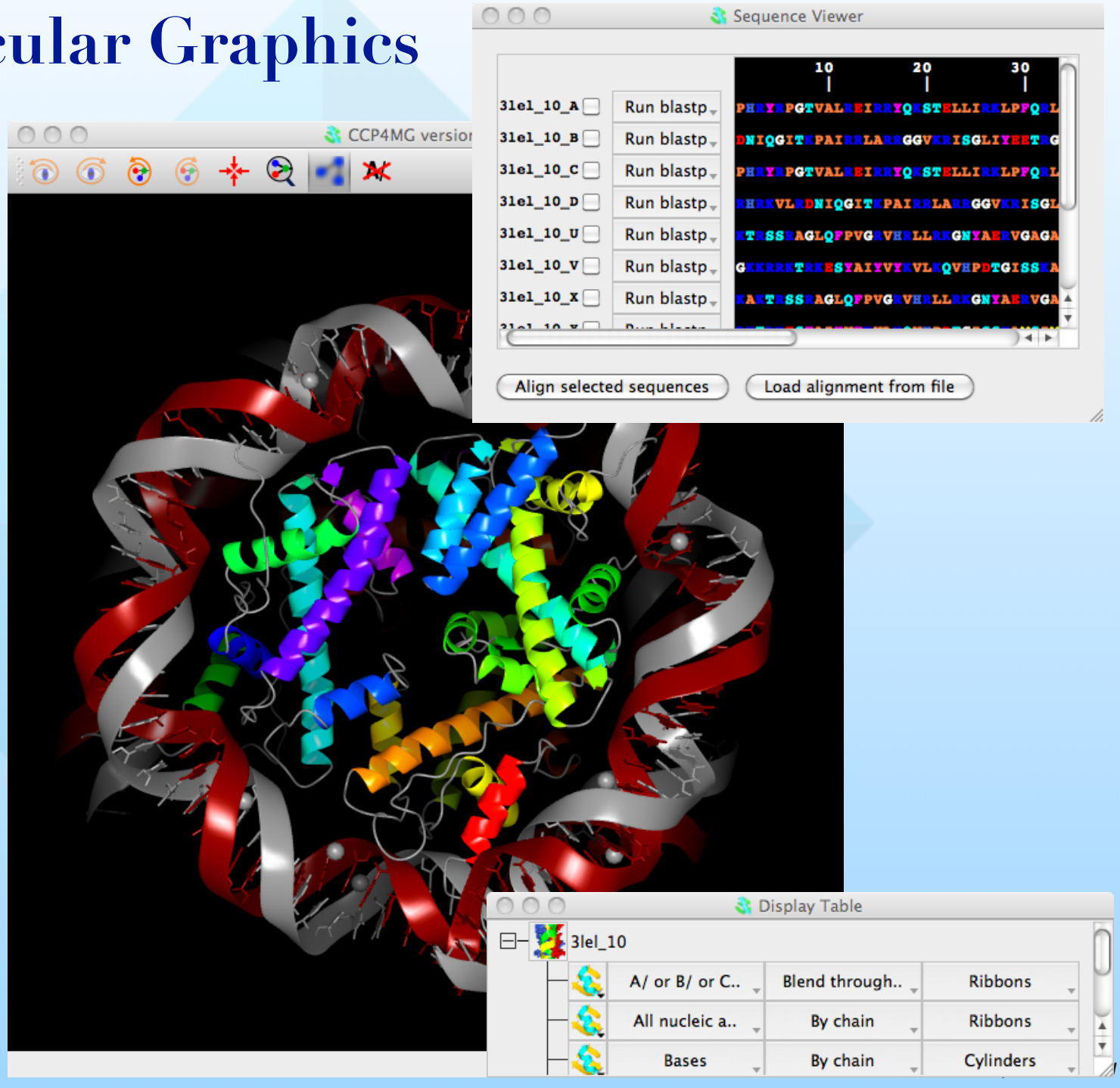
TOPDRAW	*sketchpad for protein topology diagrams (obsoleted)*
CCP4 MG	*advanced Molecular Graphics viewer*

Research Complex at Harwell

CCP4 Molecular Graphics

Version 2.11.0: 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



Structure Analysis in CCP4

- Protein interactions and complexes

  PISA *analysis of Protein Interactions, Surfaces and Assemblies*

Research Complex at Harwell