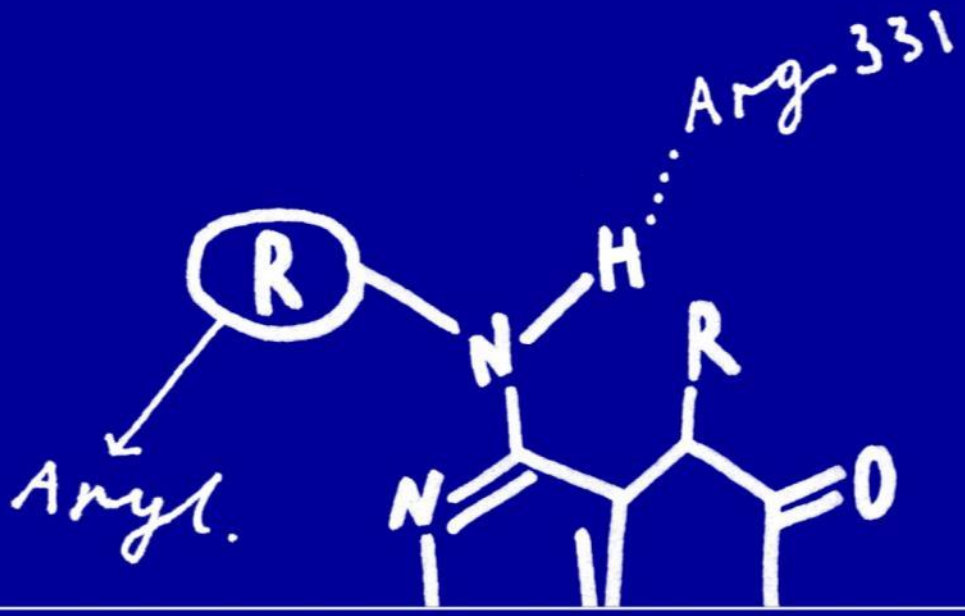
evotec

'RESEARCH NEVER STOPS'
Drug Discovery Solutions
& MORE

Introduction to COOT

Paul McEwan Ph.D.



evotec

- Introduction
- The basics
- Refinement
- More Refinement
- Ligands (briefly)
- Summary

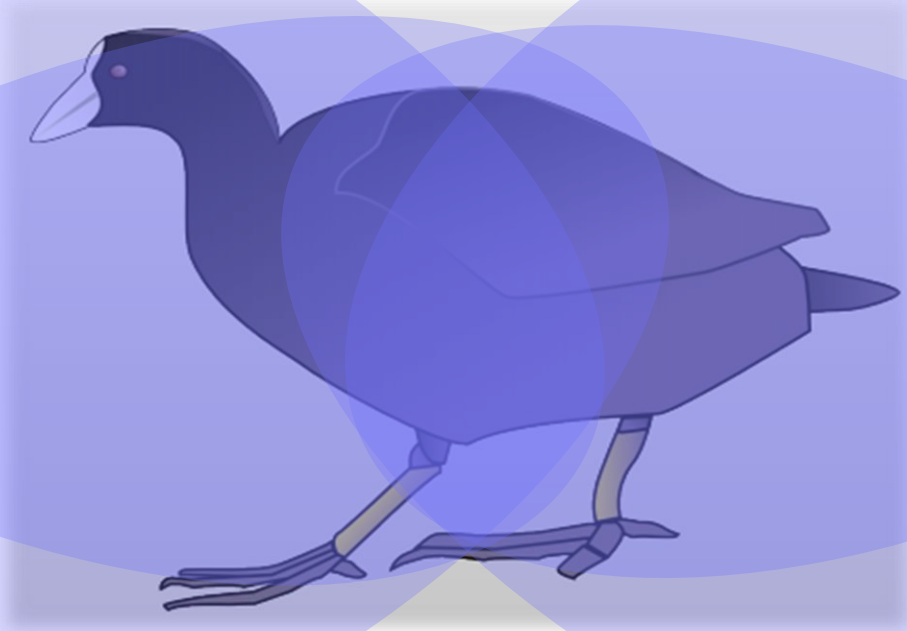
Coot- Crystallographic Object-Oriented Toolkit

- A crystallographers workhorse interface
- Integration of manual intervention and computational refinement tools
- Primarily developed by Paul Emsley (LMB)
  - Contributions from numerous other authors
  - Open source and designed to be easily learned
  - Scriptable
- Huge capability and this presentation will only skim the surface

Refinement (restraints & data) and validation should go hand-in-hand

Refinement
Refmac
Shelx
Buster
Phenix

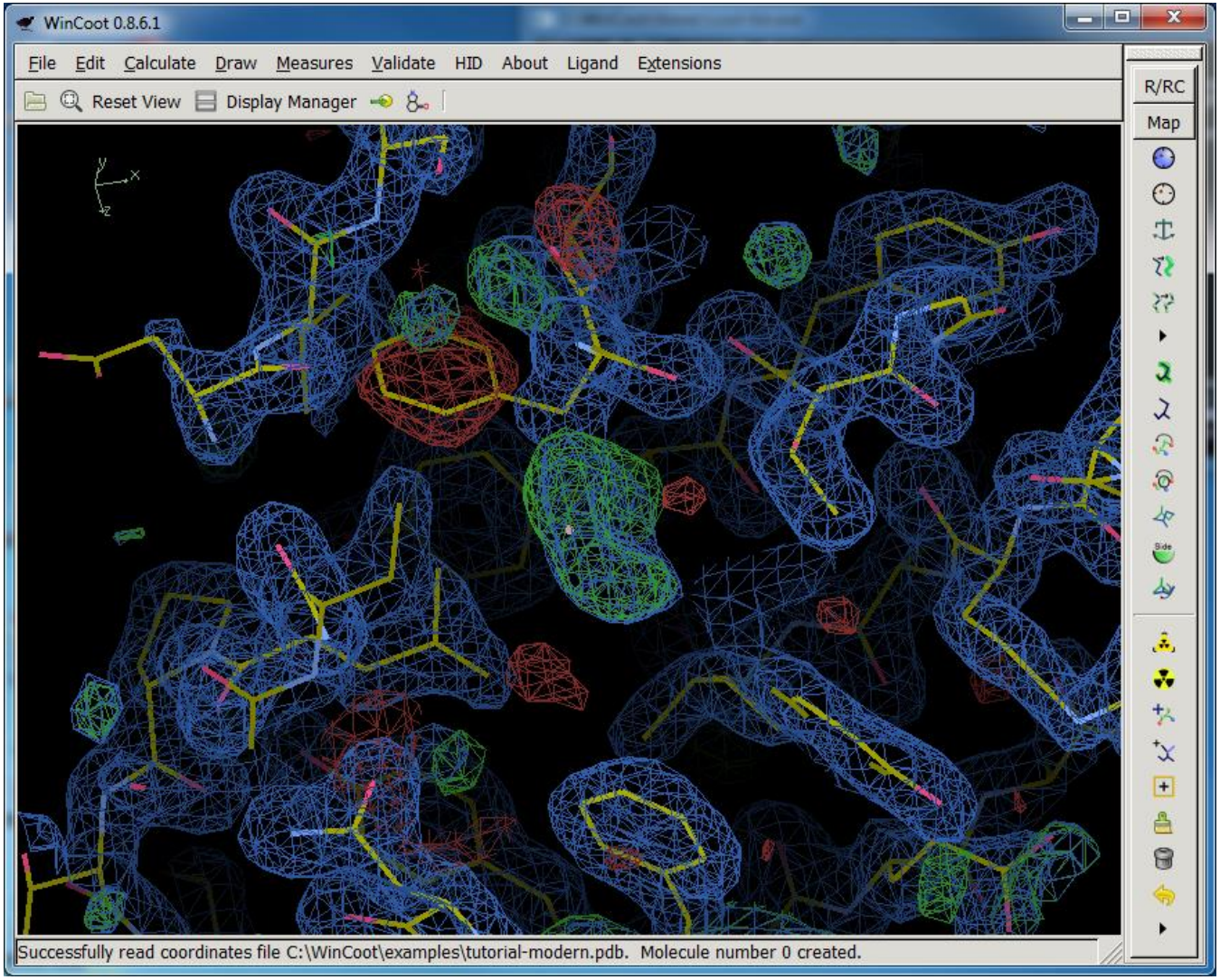
Restraints
libcheck
Acedrg
cProdrg
Grade



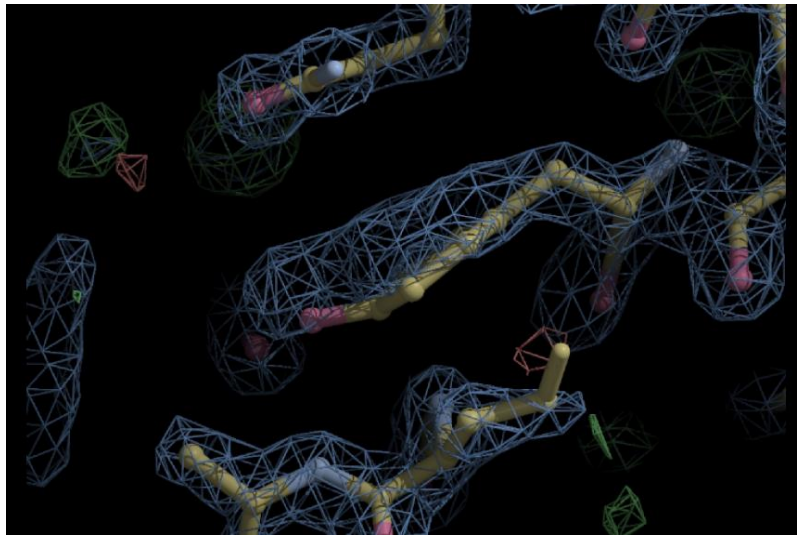
Data
Structure factors
intensities

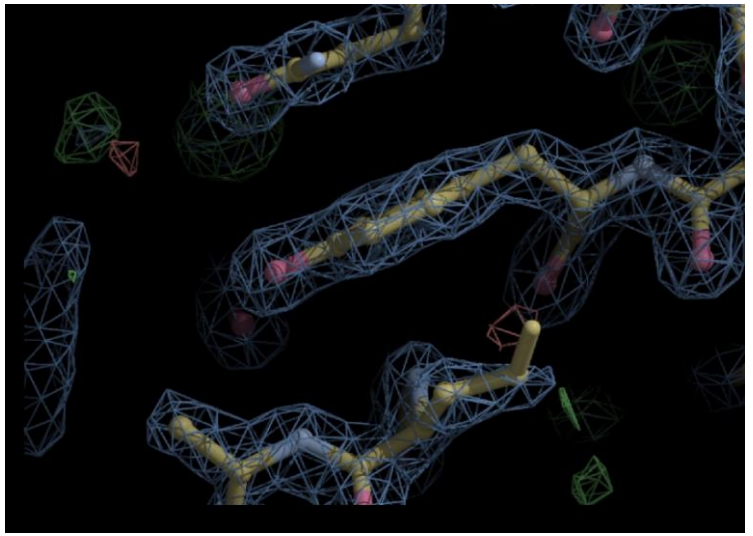
Validation
Molprobity
Mogul

The Interface

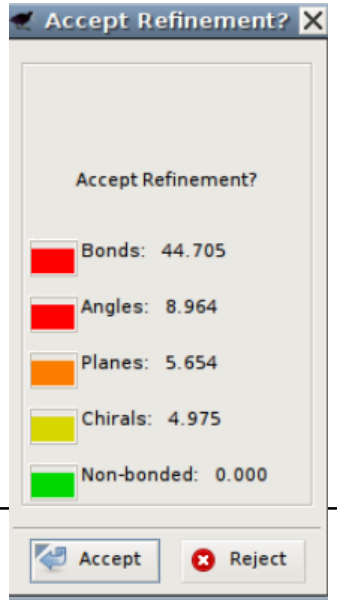


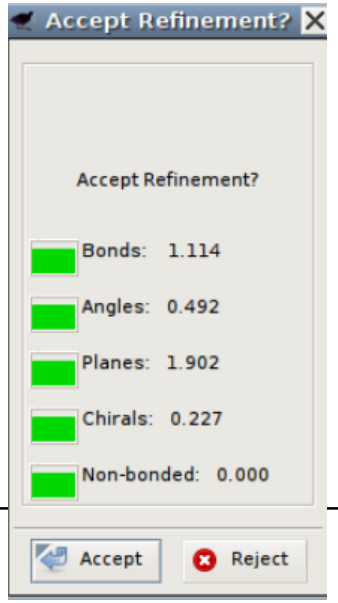
The Tool Bar





- Which map?
- Additional restraints?
- Grab and drag
- Traffic light warnings
- Interactive refinement in real space





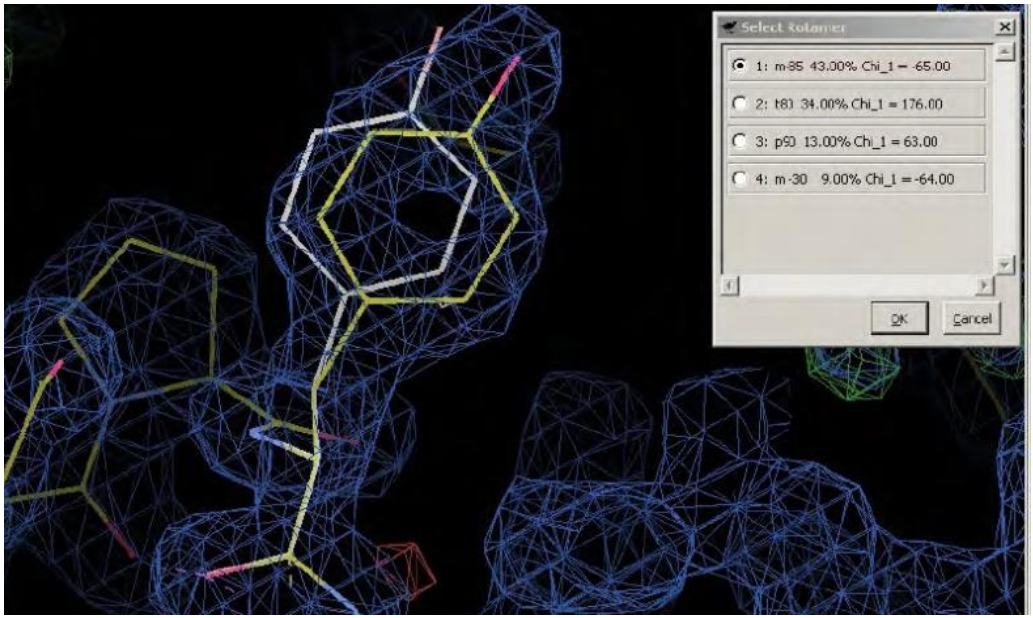
The Tool Bar

Rotamers (low resolution “back rub” method)
Extensions → Modelling → Rotamer Search

“mutate” /change residue

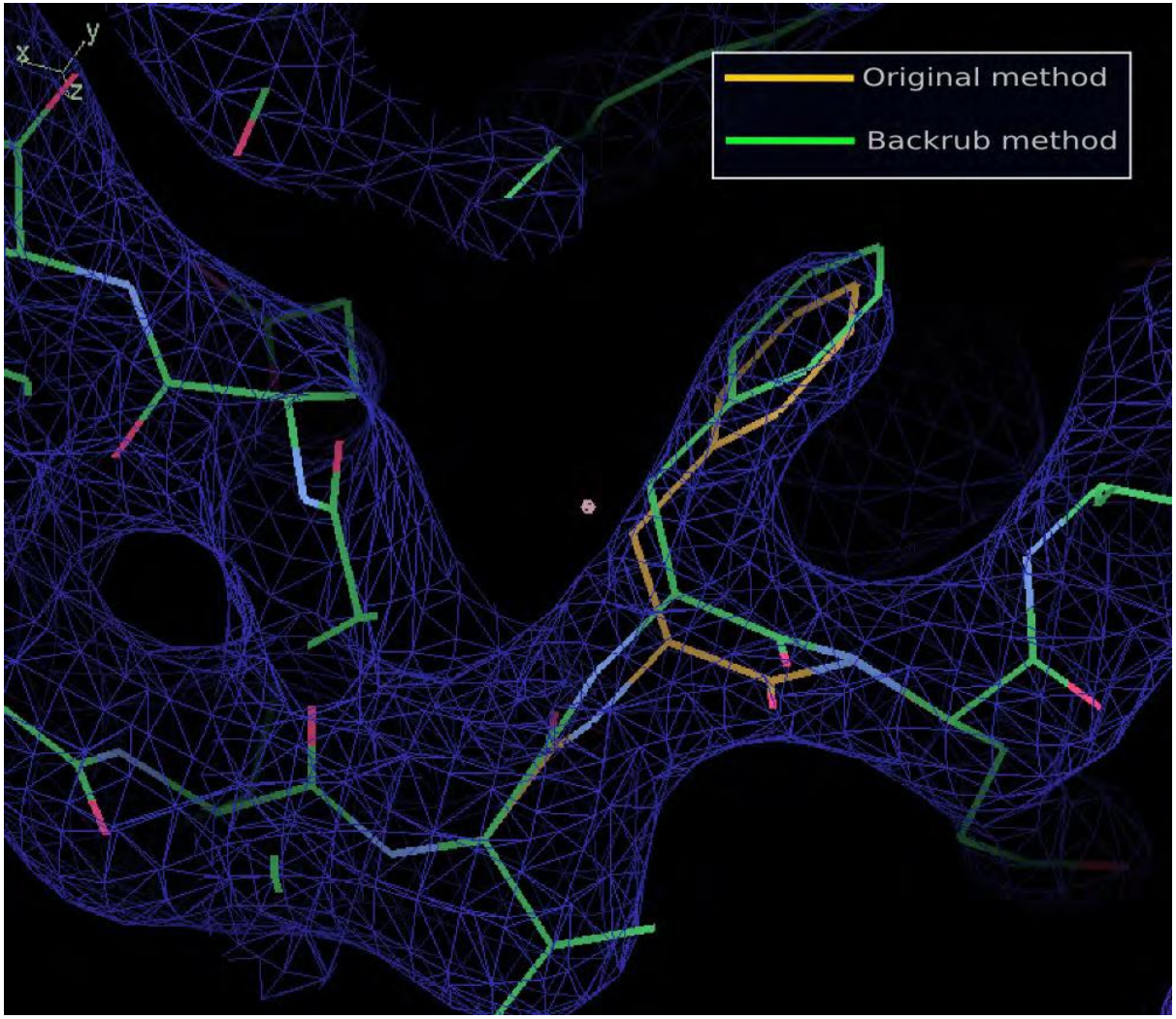
Add terminal residue

Add conformation

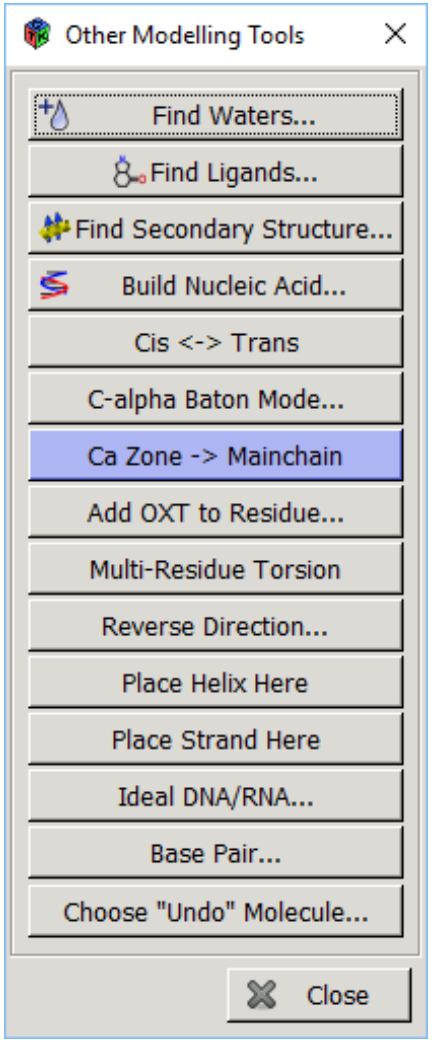


Low Resolution Rotamer search

- Where Resolution >2.9 Å automatically turned on
- Where resolution is higher the back rub adds little additional benefit



Automation for bulk fitting tasks



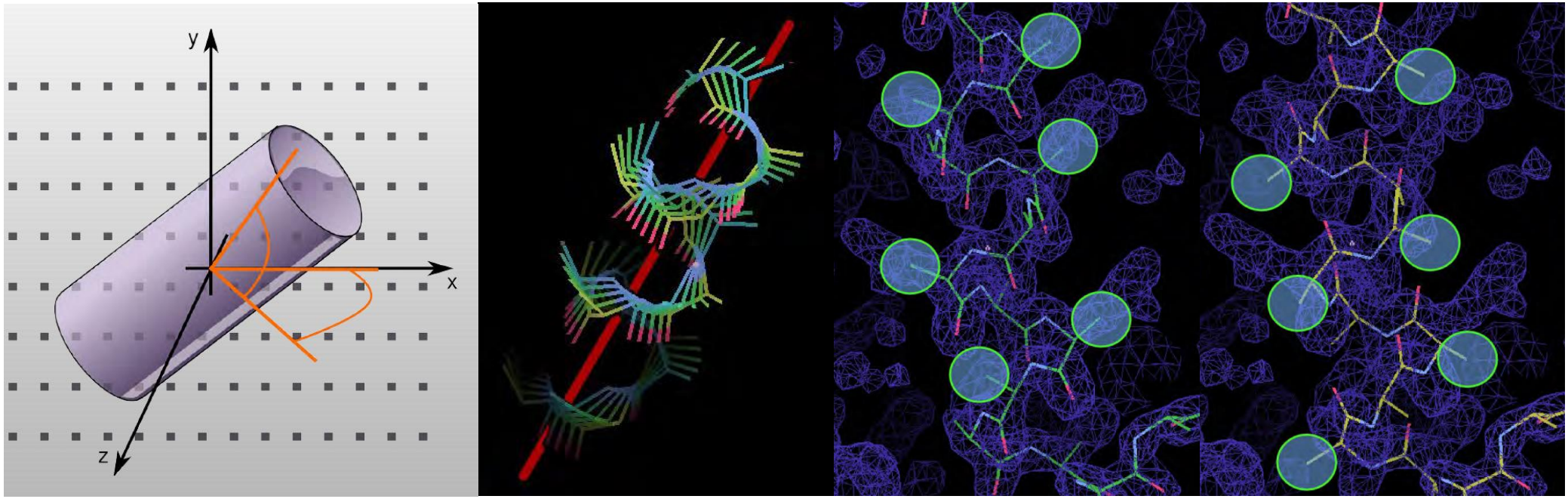
Secondary structures

De novo modelling… Should you use Buccaneer?

Waters

Ligands

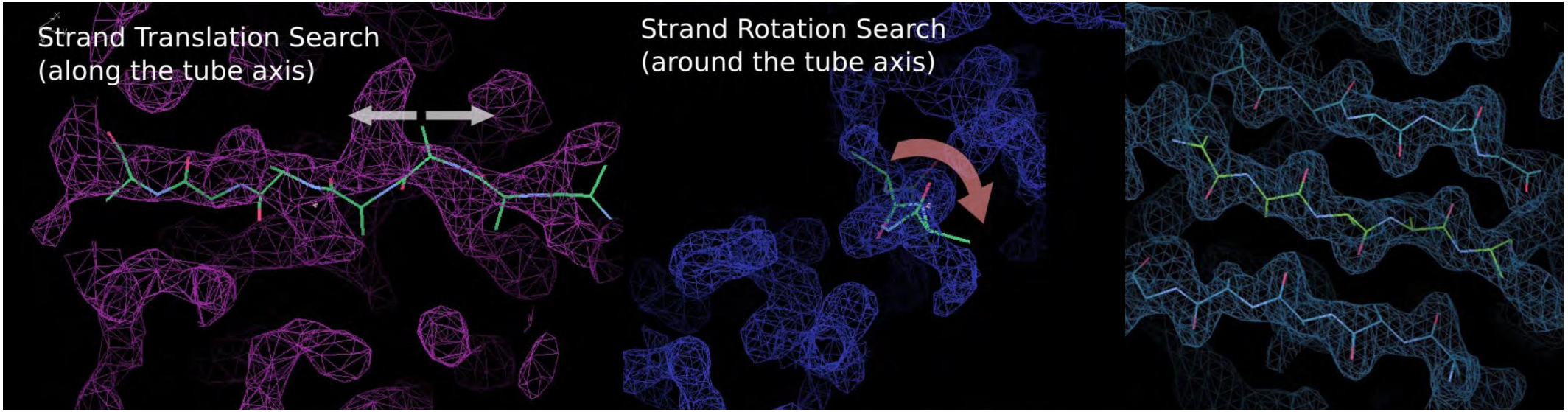
Alpha Helices



- Automated alpha helix placement
- Move cursor to “tube” of alpha helix
- Automated search of orientation, register and polarity
- Robust even at modest resolutions

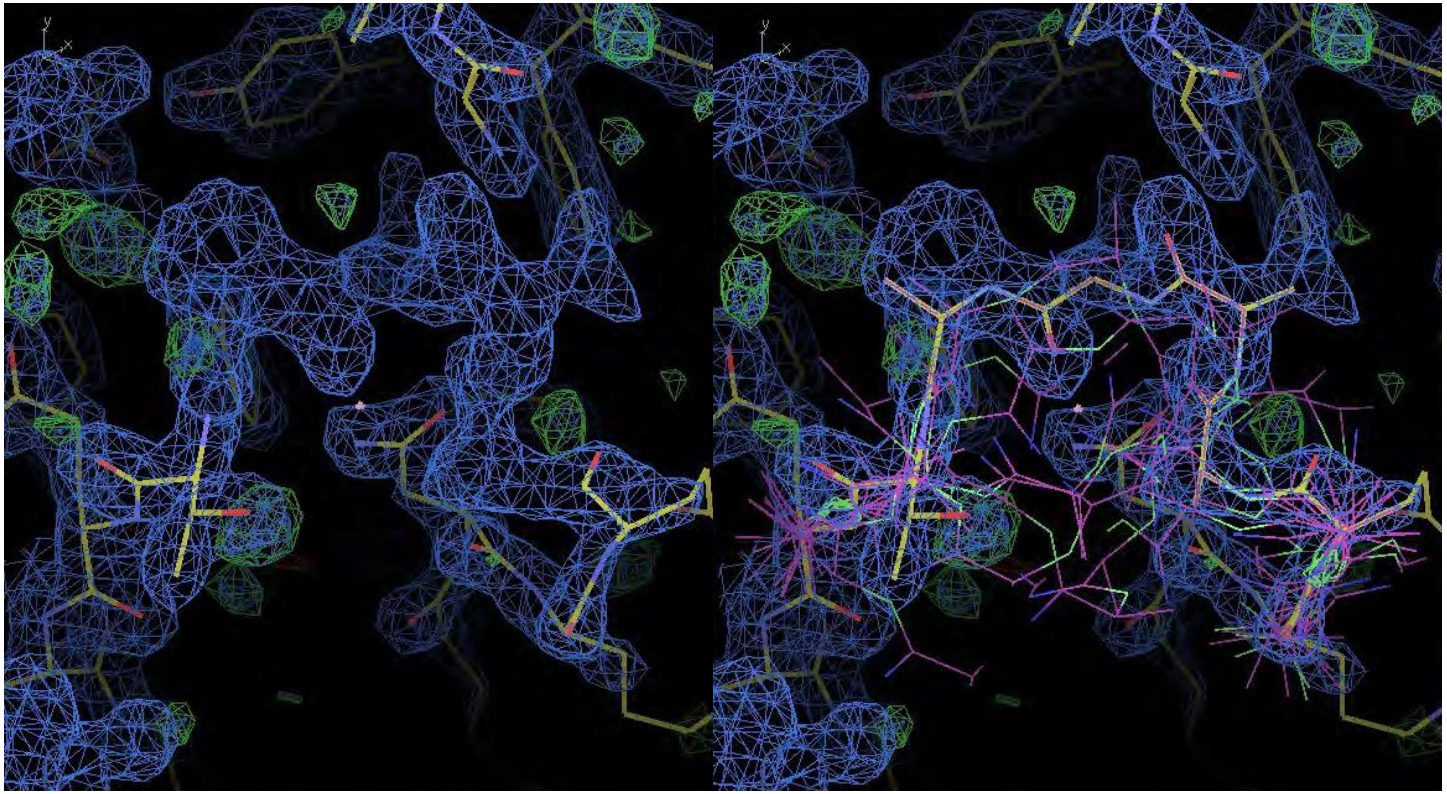
Beta sheets

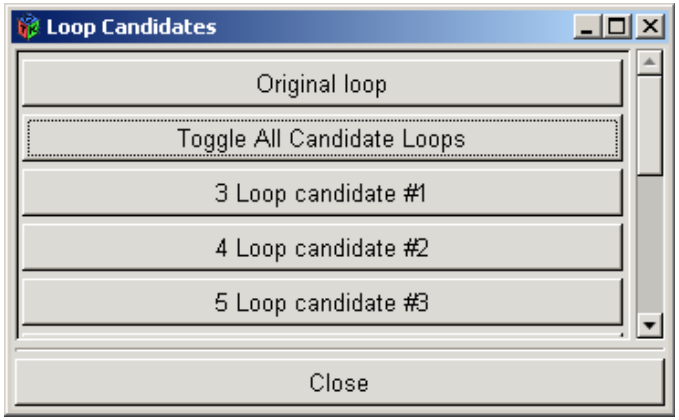
- More complicated than alpha Helices
- Non-ideal to accommodate frequent curvature of strands
- Start with cylinder search and move to strand extension followed by rigid body and real-space refinement
- Detection of strand polarity



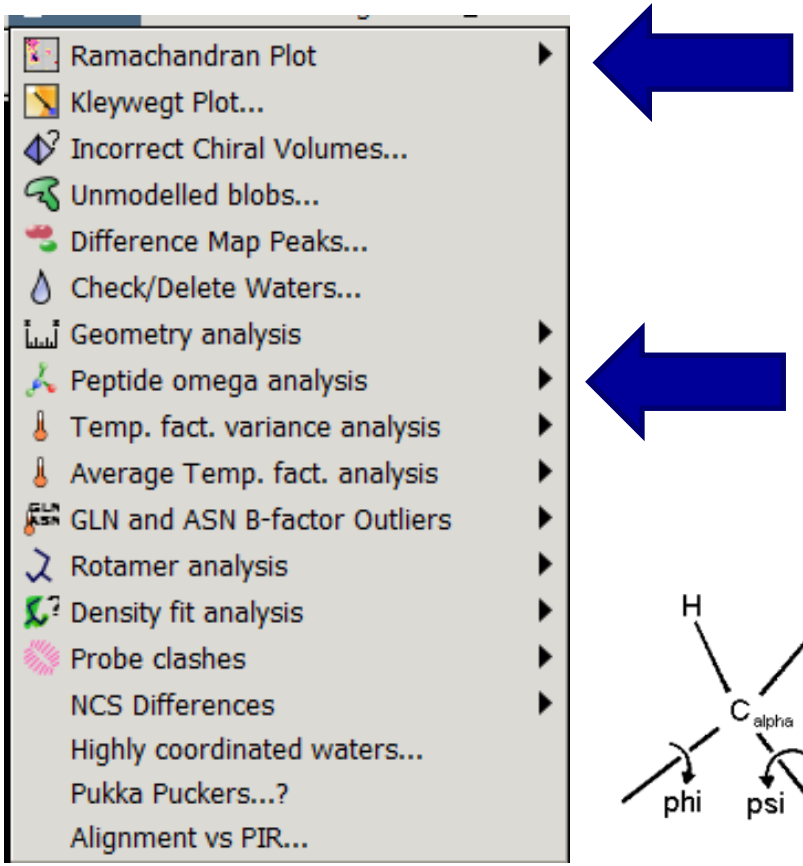
loops

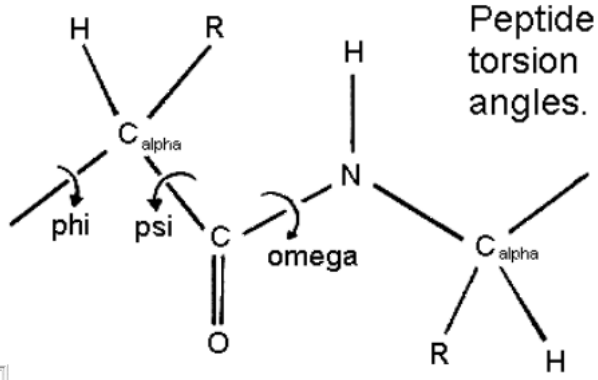
- Simple manual fitting of single residues
- Database search using existing information



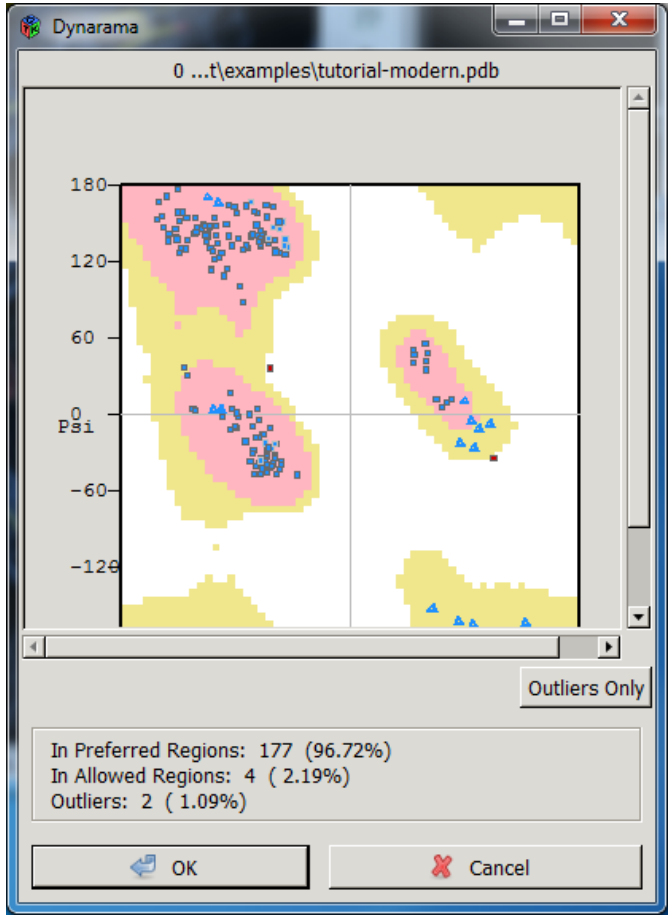


Validation Tools

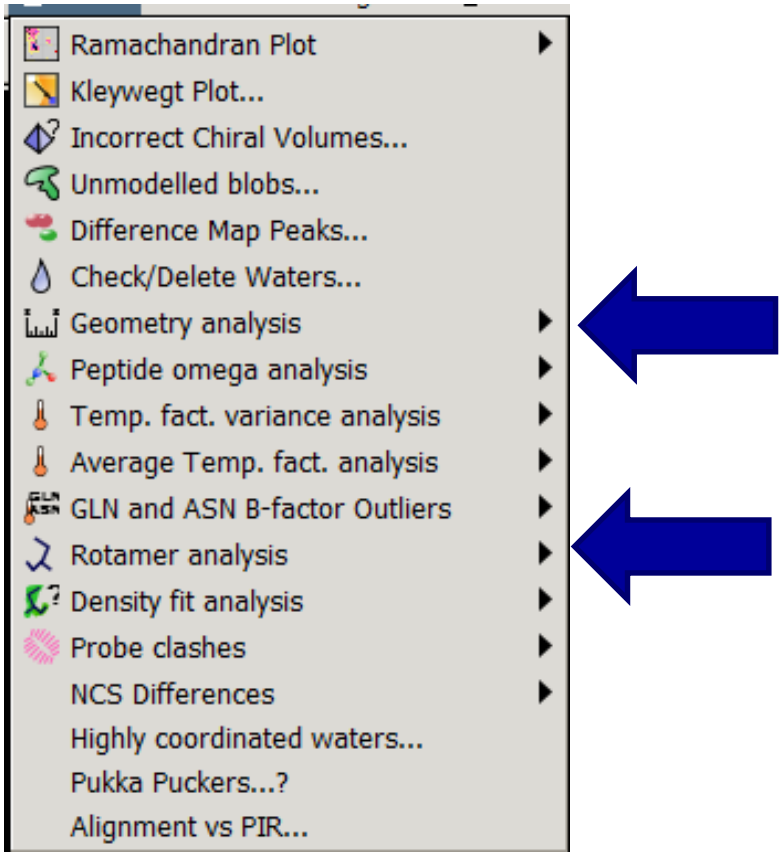


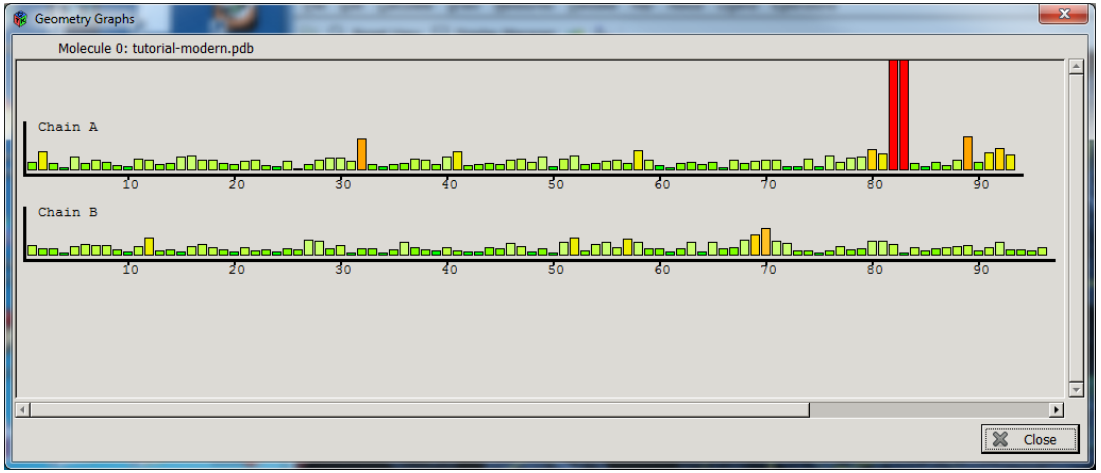


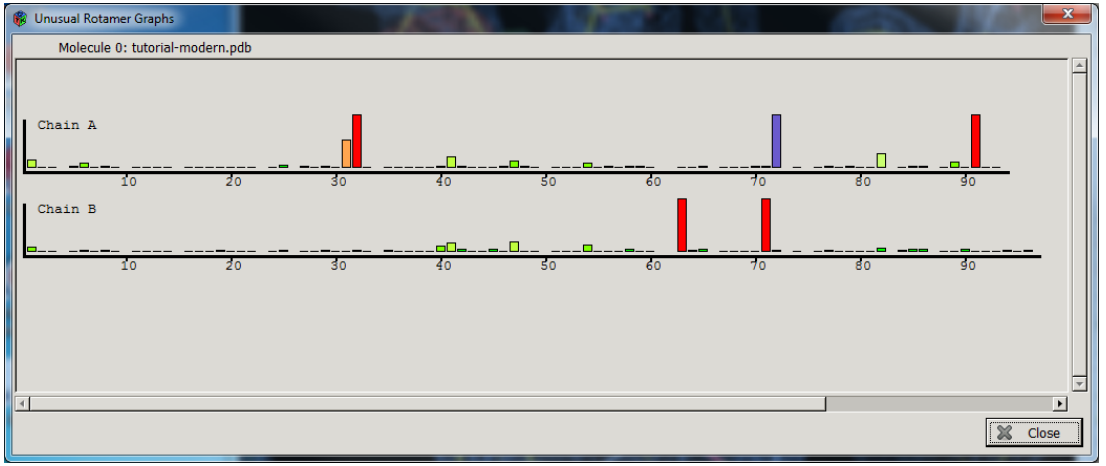
Peptide torsion angles.



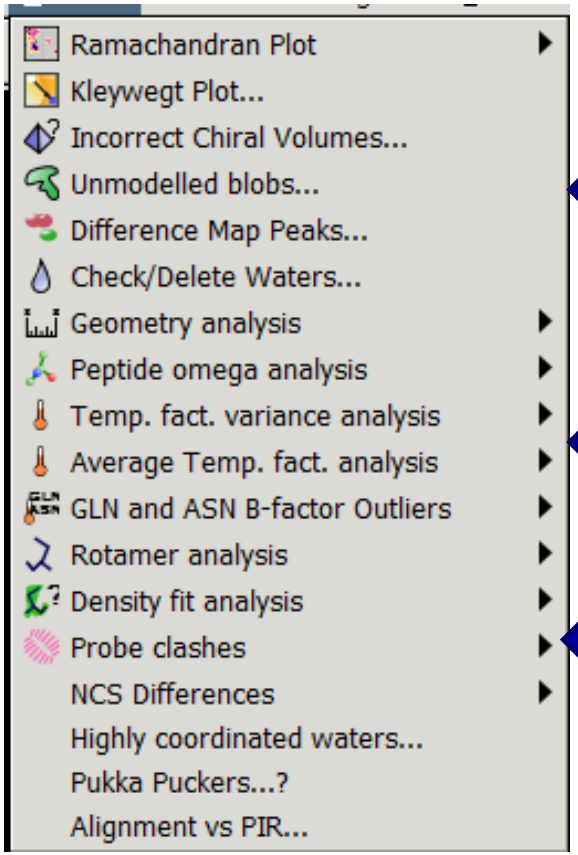
Validation Tools







Validation Tools



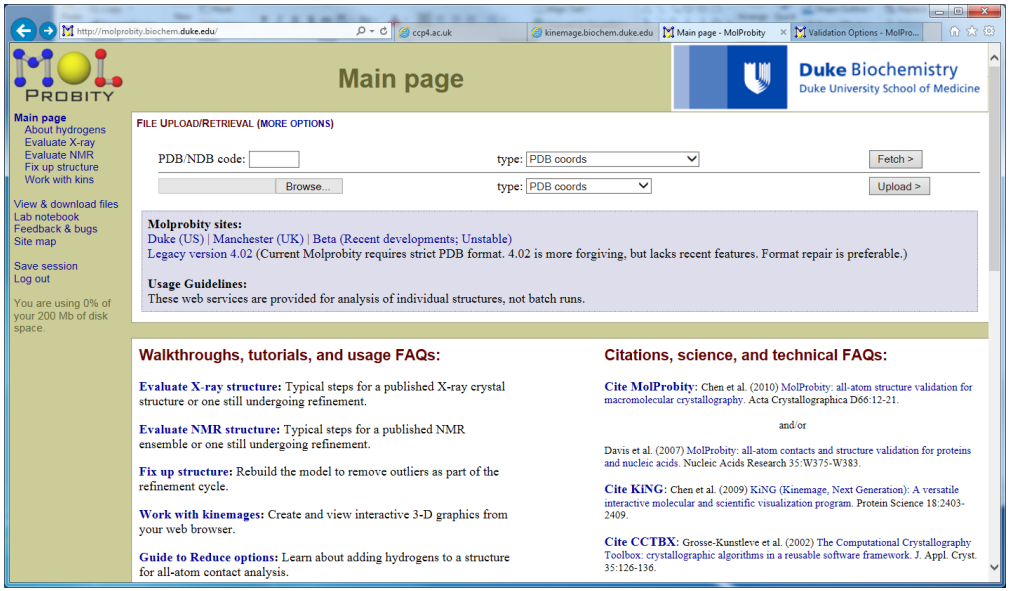
Ligands? Unmodelled protein? Incomplete model

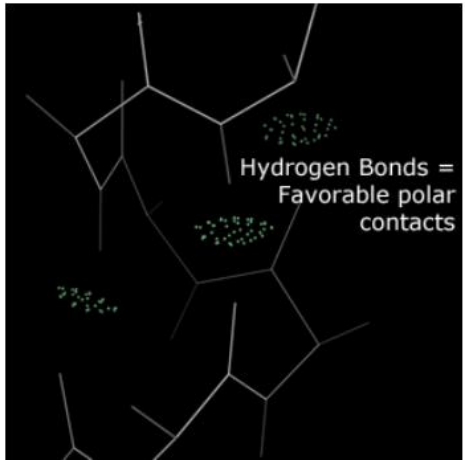
B- factors, weak density or poor fit?

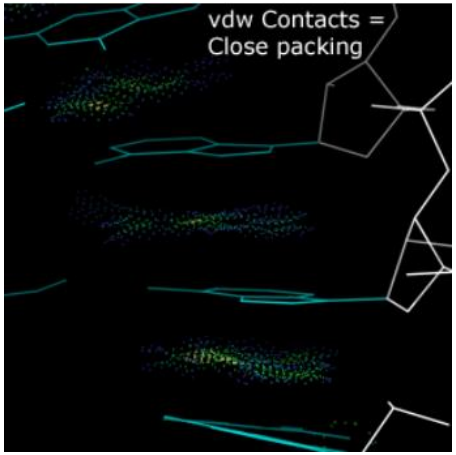
Probe/Reduce and Molprobity

Molprobity

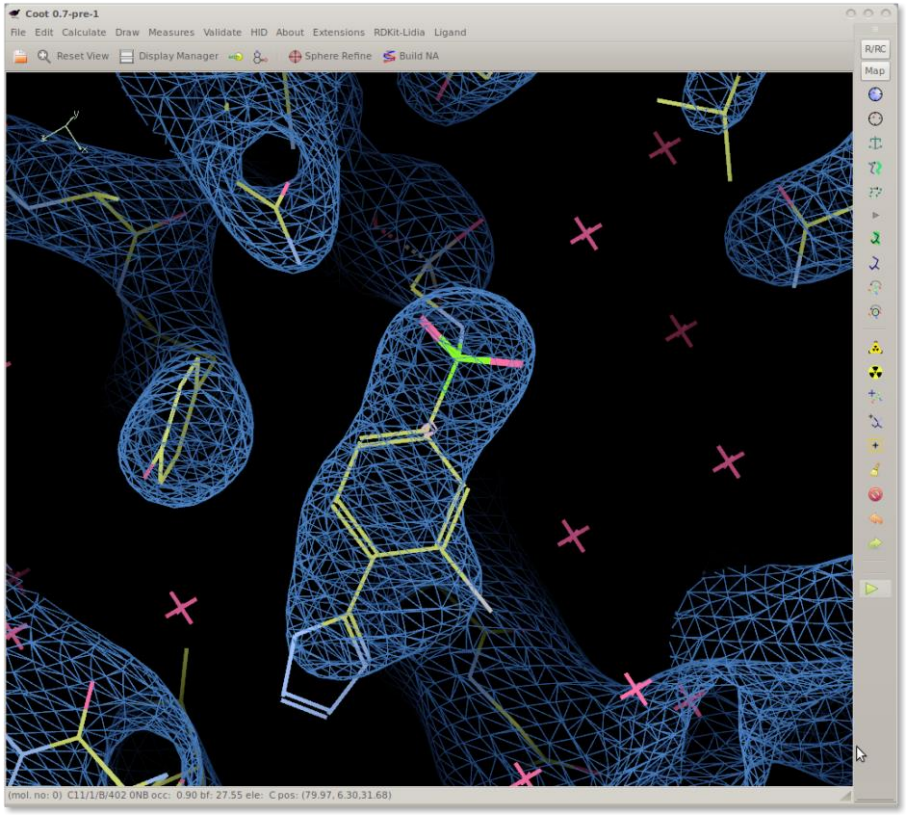
- Amazing tool to look at geometry
- Coot plugin to interpret output and to make “problem-fixing” easier
- List of specific problems
- Coot visualisation

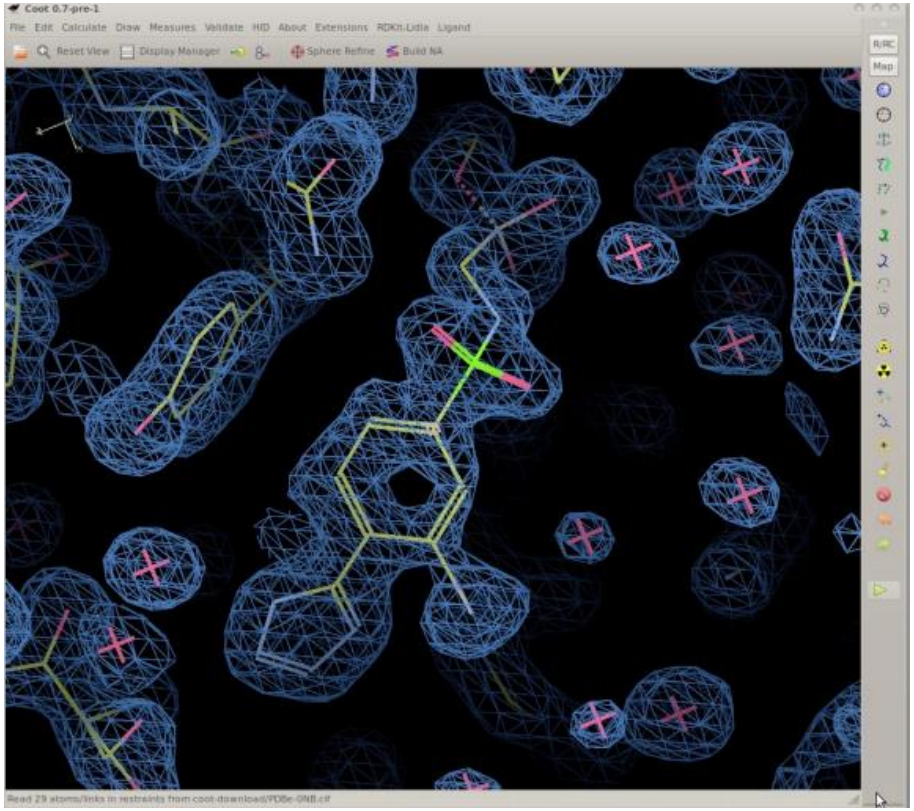






Ligand and Density...

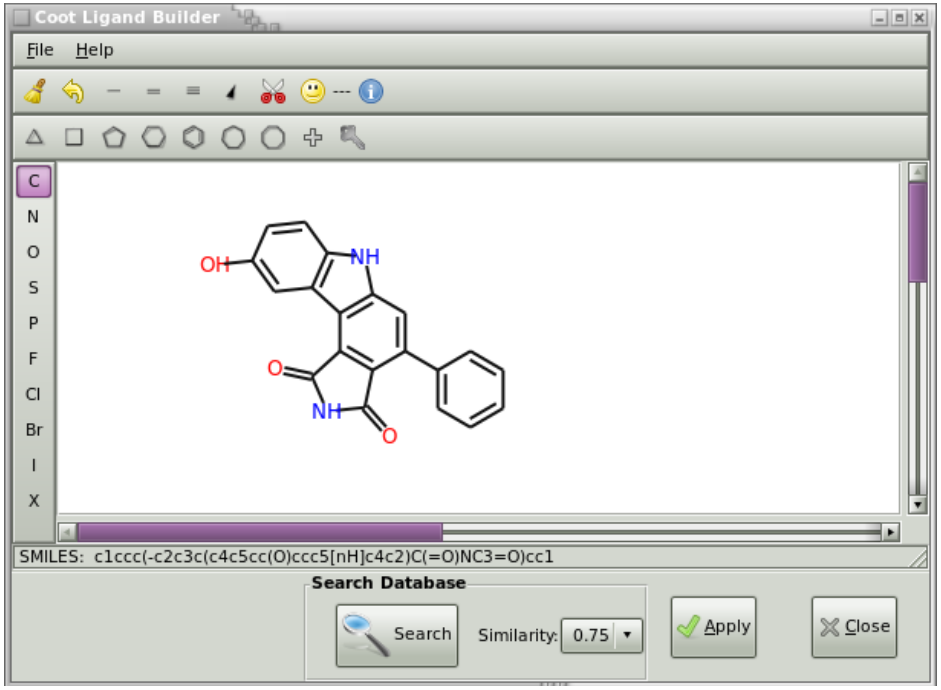


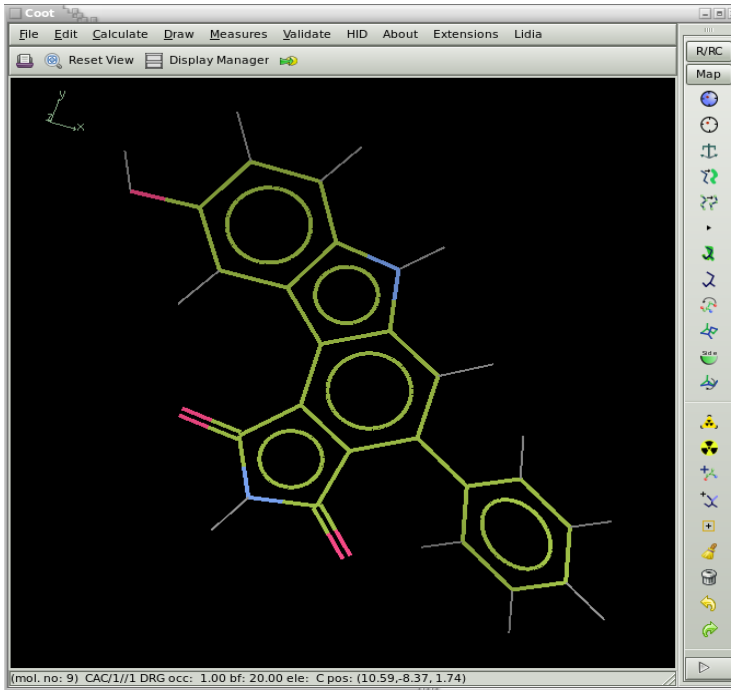


- Protein-ligand complex models are often a result of subjective interpretation
- Combination of map quality and ligand restraints
- Judit will be giving a practical session later in the week

2D Ligand Builder

- External tools
- COOT
  - Free sketch
  - SBase search





QED Score

Quantitative Evaluation of Drug-likeness

ARTICLES
PUBLISHED ONLINE: 24 JANUARY 2012 | DOI: 10.1038/NCHEM.1243
nature chemistry

Quantifying the chemical beauty of drugs

G. Richard Bickerton[1], Gaia V. Paolini[2], Jérémy Besnard[1], Sorel Muresan[3] and Andrew L. Hopkins[1]*

Drug-likeness is a key consideration when selecting compounds during the early stages of drug discovery. However, evaluation of drug-likeness in absolute terms does not reflect adequately the whole spectrum of compound quality. More worryingly, widely used rules may inadvertently foster undesirable molecular property inflation as they permit the encroachment of rule-compliant compounds towards their boundaries. We propose a measure of drug-likeness based on the concept of desirability called the quantitative estimate of drug-likeness (QED). The empirical rationale of QED reflects the underlying distribution of molecular properties. QED is intuitive, transparent, straightforward to implement in many practical settings and allows compounds to be ranked by their relative merit. We extended the utility of QED by applying it to the problem of molecular target druggability assessment by prioritizing a large set of published bioactive compounds. The measure may also capture the abstract notion of aesthetics in medicinal chemistry.

The concept of drug-likeness provides useful guidelines for early-stage drug discovery[1,2]. Analysis of the observed distribution of some key physicochemical properties of approved drugs, including molecular mass (M_r), hydrophobicity and polarity, reveals that they occupy preferentially a relatively narrow range of possible values[3]. Compounds that fall within this range are described as 'drug-like'. This definition holds in the absence of any obvious structural similarity to an approved drug. It has been shown that the preferential selection of drug-like compounds increases the likelihood of surviving the well-documented high rates of attrition in drug discovery[4].

Drug-likeness can be rationalized by considering how simple physicochemical properties impact molecular behaviour *in vivo*, with particular respect to solubility, permeability, metabolic stability and transporter effects. Indeed, drug-likeness is often used as a proxy for oral bioavailability. However, drug-likeness provides a broad composite descriptor that implicitly captures several criteria,

Paradoxically, since the publication of the seminal paper by Lipinski *et al.*[5] there appears to be a growing epidemic, which Hann has termed 'molecular obesity'[8], among new pharmacological compounds (Supplementary Fig. S1). Compounds with higher relative M_r and lipophilicity have a higher probability of attrition at each stage of clinical development[4,9–11]. Thus, the inflation of physicochemical properties that increases the risks associated with clinical development may explain, in part, the decline in productivity of small-molecule drug discovery over the past two decades[4]. However, the mean molecular properties of new pharmacological compounds are still considered Lipinski compliant, even though their property distributions are far from historical norms.

Although the Ro5 is predictive of oral bioavailability, 16% of oral drugs violate at least one of the criteria and 6% fail two or more (although this does include natural products and substrates of transporters) (Supplementary Fig. S2a and Supplementary Table S1). High-profile drugs, such as atorvastatin (Lipitor) and montelukast

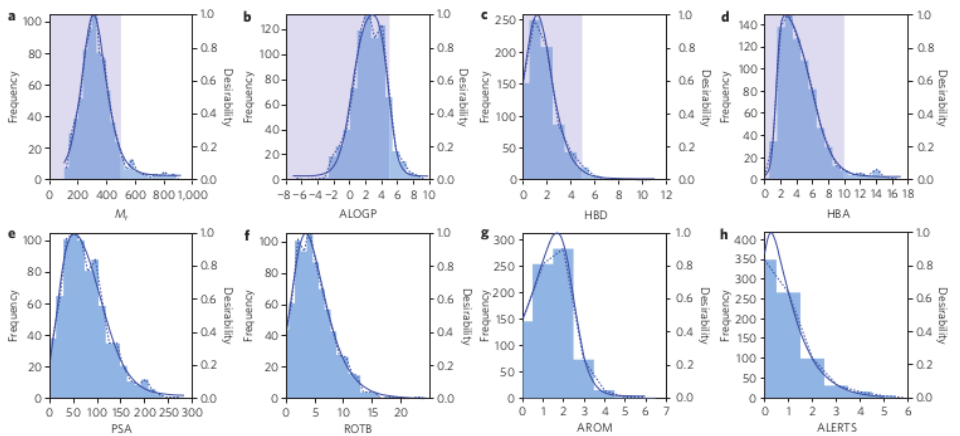


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h, Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**), number of AROMs (**g**) and number of ALERTS (**h**). The Lipinski-compliant areas are shown in pale blue in (**a**), (**b**), (**c**) and (**d**). The solid blue lines describe the ADS functions (equation (2)) used to model the histograms. The parameters for each function are given in Supplementary Table S1.

design[17,18], prioritization of molecular targets, penetration of the central nervous system[19] and estimating the reliability of screening data[20]. The concept was introduced originally by Harrington[15] in the area of process engineering and further refined by Derringer and Suich[21]. Desirability takes multiple numerical or categorical parameters measured on different scales and describes each by an individual desirability function. These are then integrated into a single dimensionless score. In the case of compounds, a series of desirability functions (d) are derived, each of which corresponds to a different molecular descriptor. Combining the individual desirability functions into the QED is achieved by taking the geometric

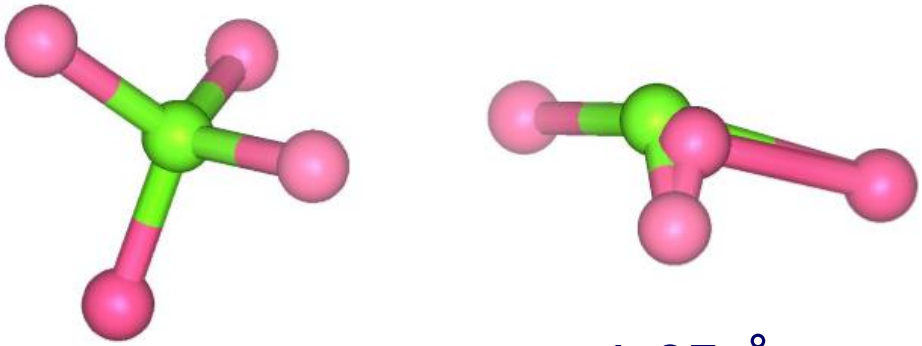
asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

$$d(x) = a + \frac{b}{\left[1+\exp\left(-\frac{x-c+\frac{d}{2}}{\quad}\right)\right]}\left[1-\frac{1}{\left[1+\exp\left(-\frac{x-c-\frac{d}{2}}{\quad}\right)\right]}\right]$$

Why bother?

- 11,000 chemical compounds in the PDB!

- Analysis of structures in the public domain
- Iridium: A Highly Trustworthy Protein-Ligand Structure Database
- *Drug Disc. Today*, **12**, 1270 (2012)



1.65 Å

- 19% of the ligands in this data set contained errors

- wwPDB: Greater focus on validation tools for PDB

evotec

Sources of help

The Homepage, F.A.Q, documentation
http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/

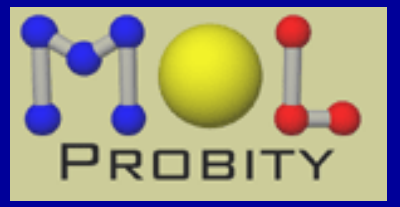
Coot Wiki
http://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Coot

The mailing list
http://www.jiscmail.ac.uk/lists/coot.html

Ligand validation (recommendations for best practise)
Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop
Structure. 2016 Apr 5;24(4):502-8. doi: 10.1016/j.str.2016.02.017.

evotec

'RESEARCH NEVER STOPS'
Drug Discovery Solutions
& MORE



Thanks To:

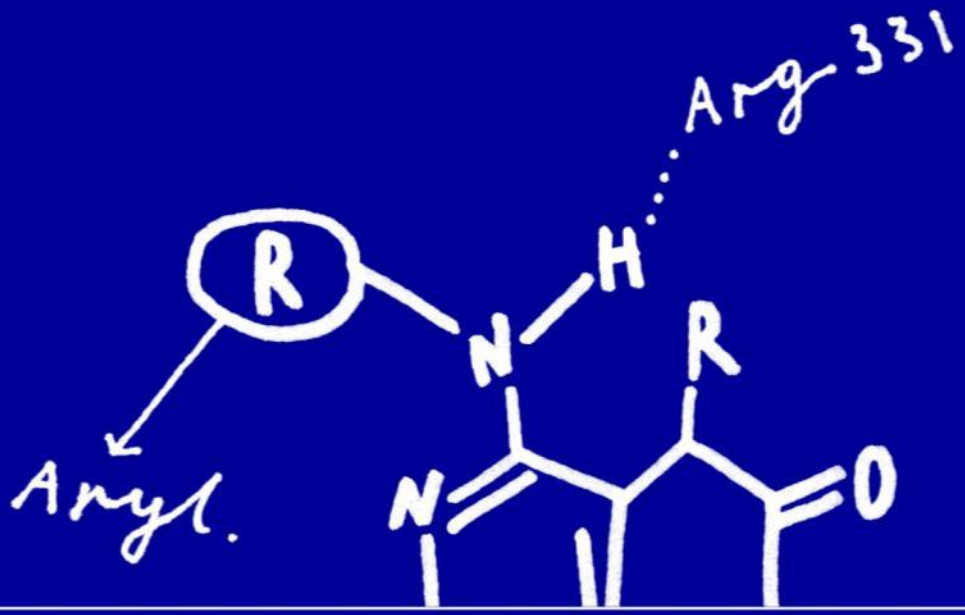
COOT
Paul Emsley
Kevin Cowtan
Eleanor Dodson
Keith Wilson

Libraries & Dictionaries
Alexei Vagin
Eugene Krissinel
Stuart McNichols

Your contact:

Paul McEwan Ph.D.

Principal Scientist, Structural Biology



evotec

‘RESEARCH NEVER STOPS’

Building innovative drug discovery alliances