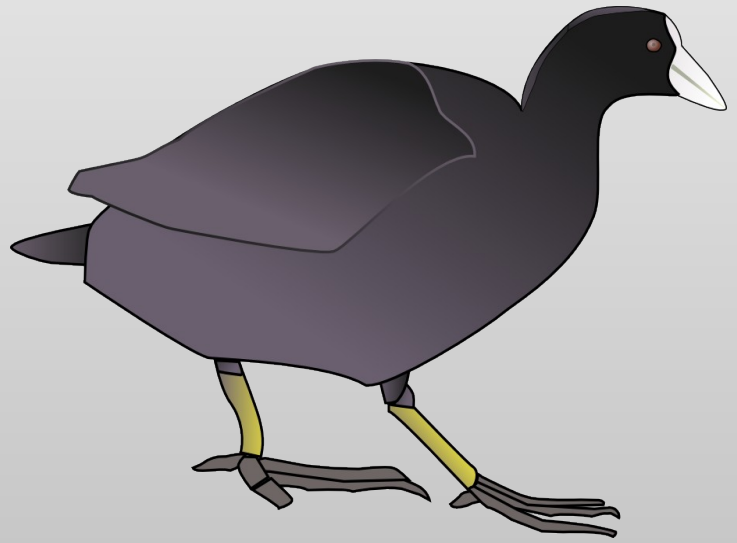
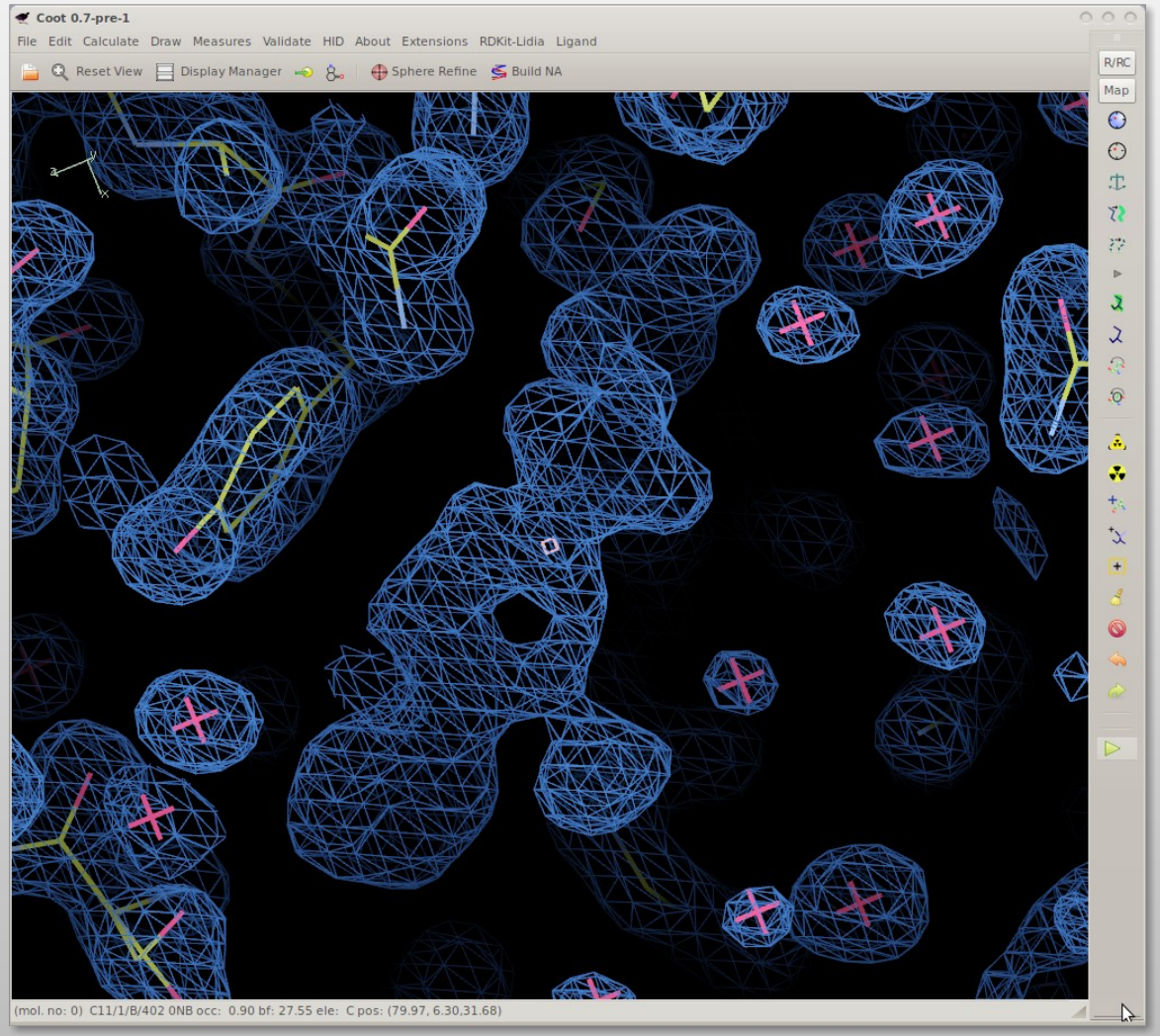
Manipulating Ligands Using *Coot*



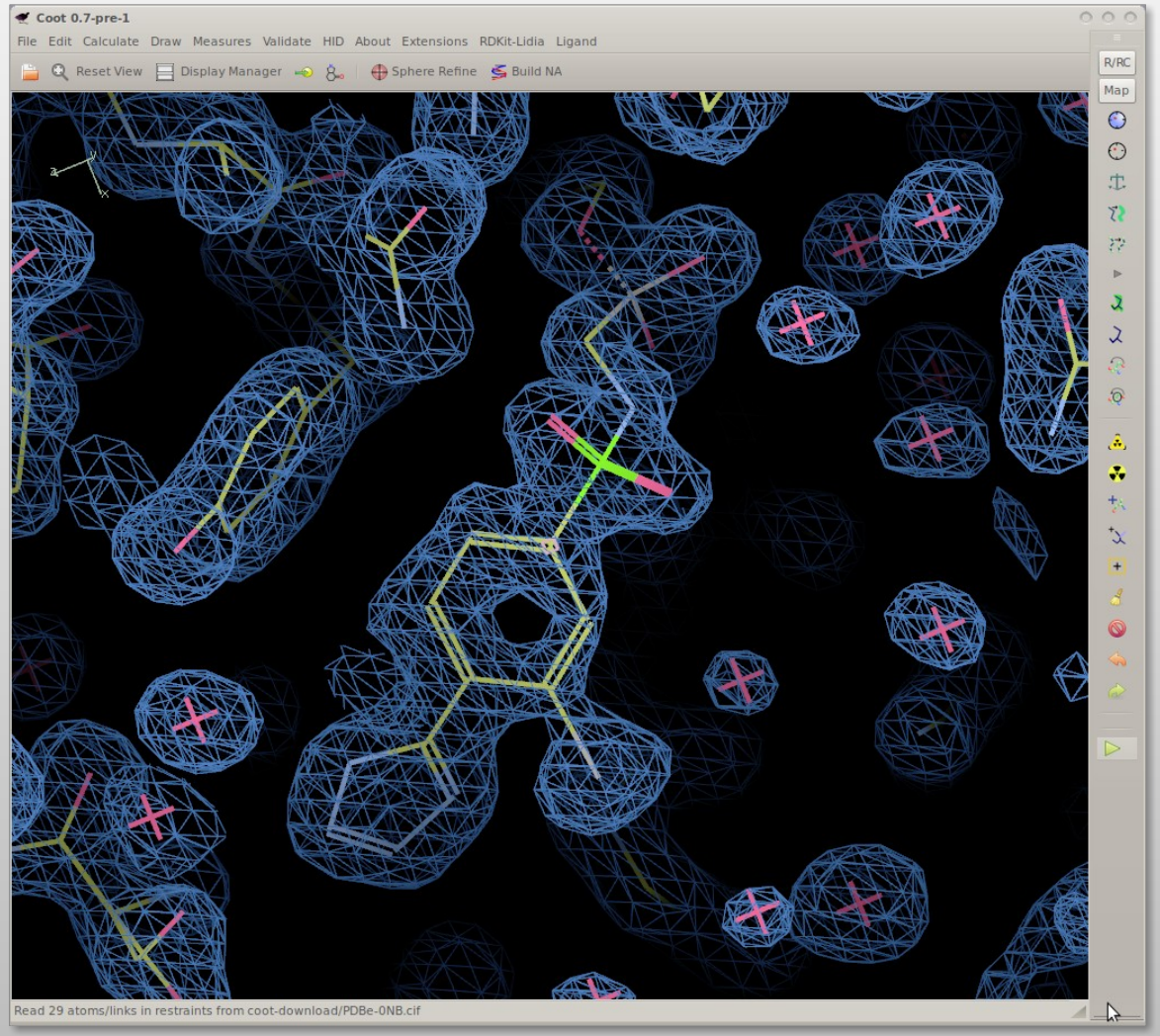
Paul Emsley

December 2014

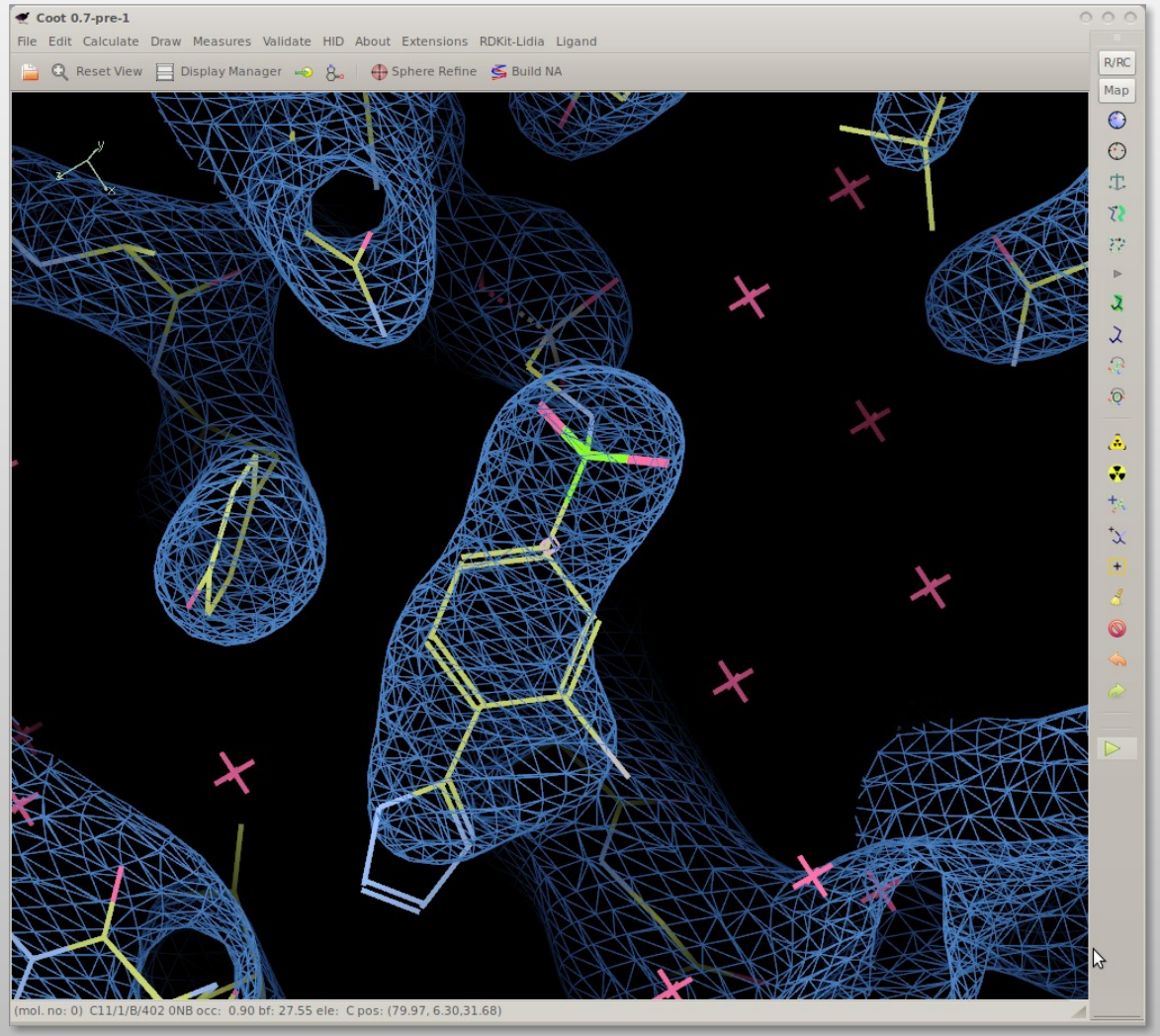
Ligand and Density...



Ligand and Density...



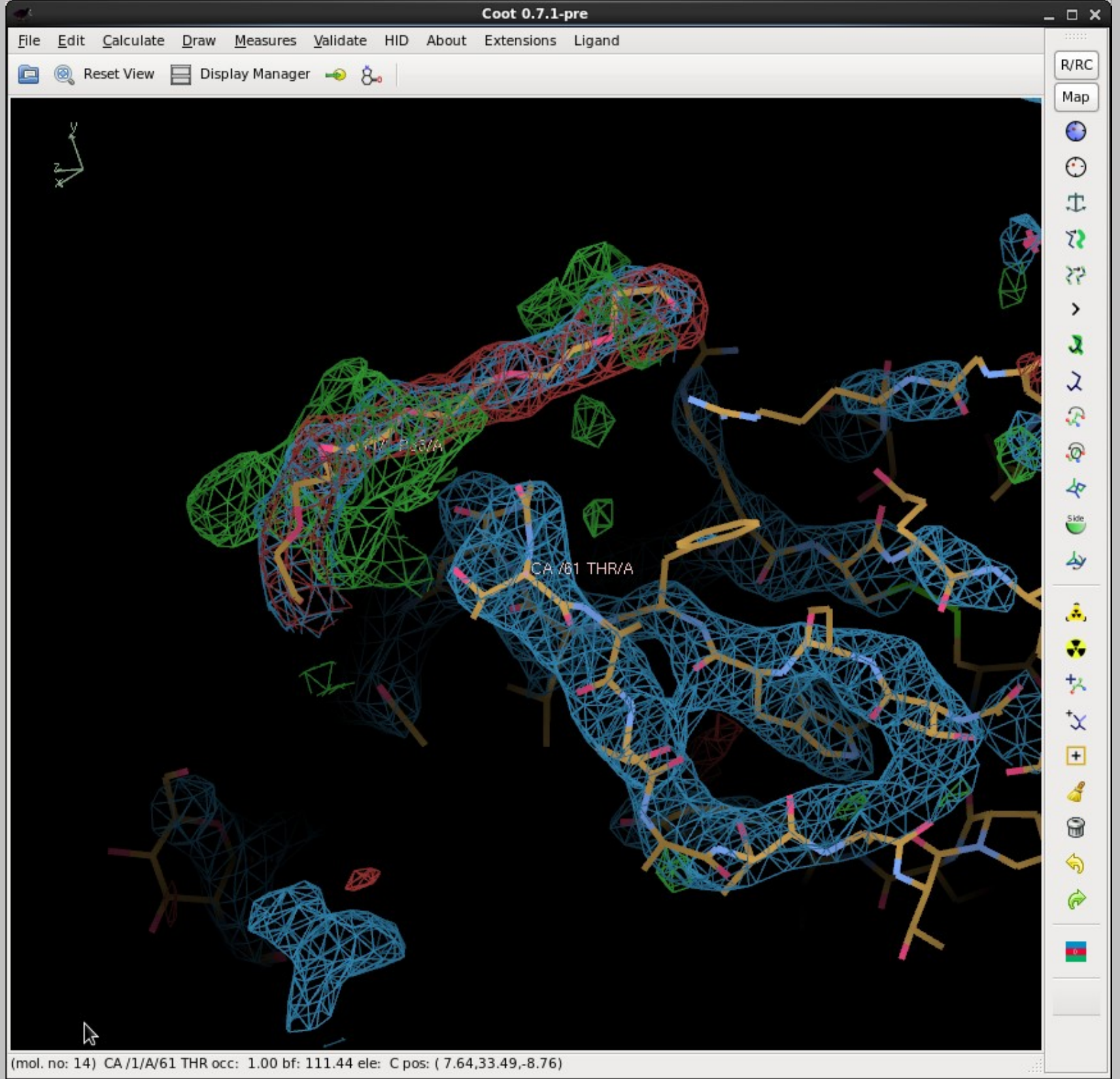
Ligand and Density...

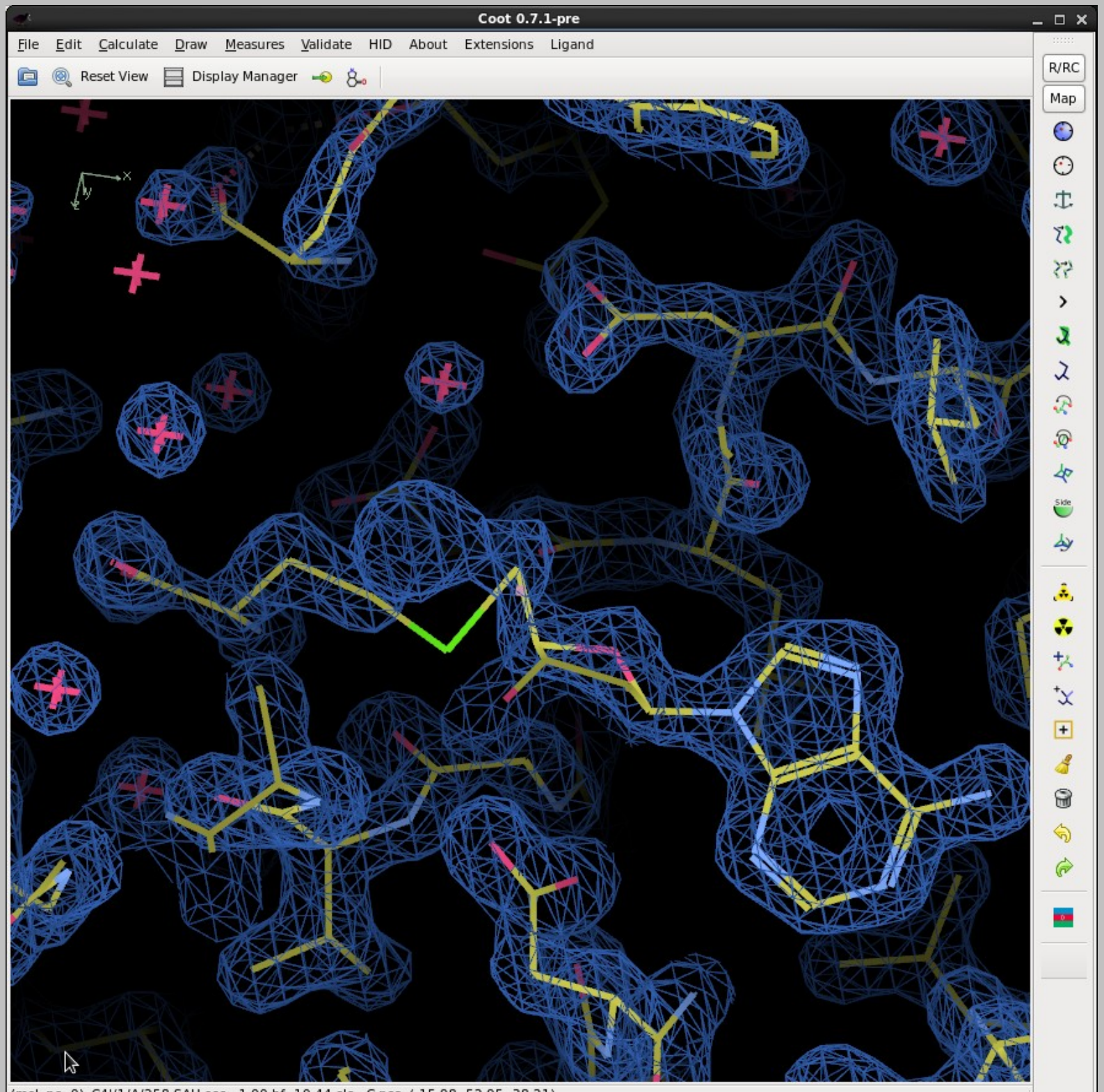


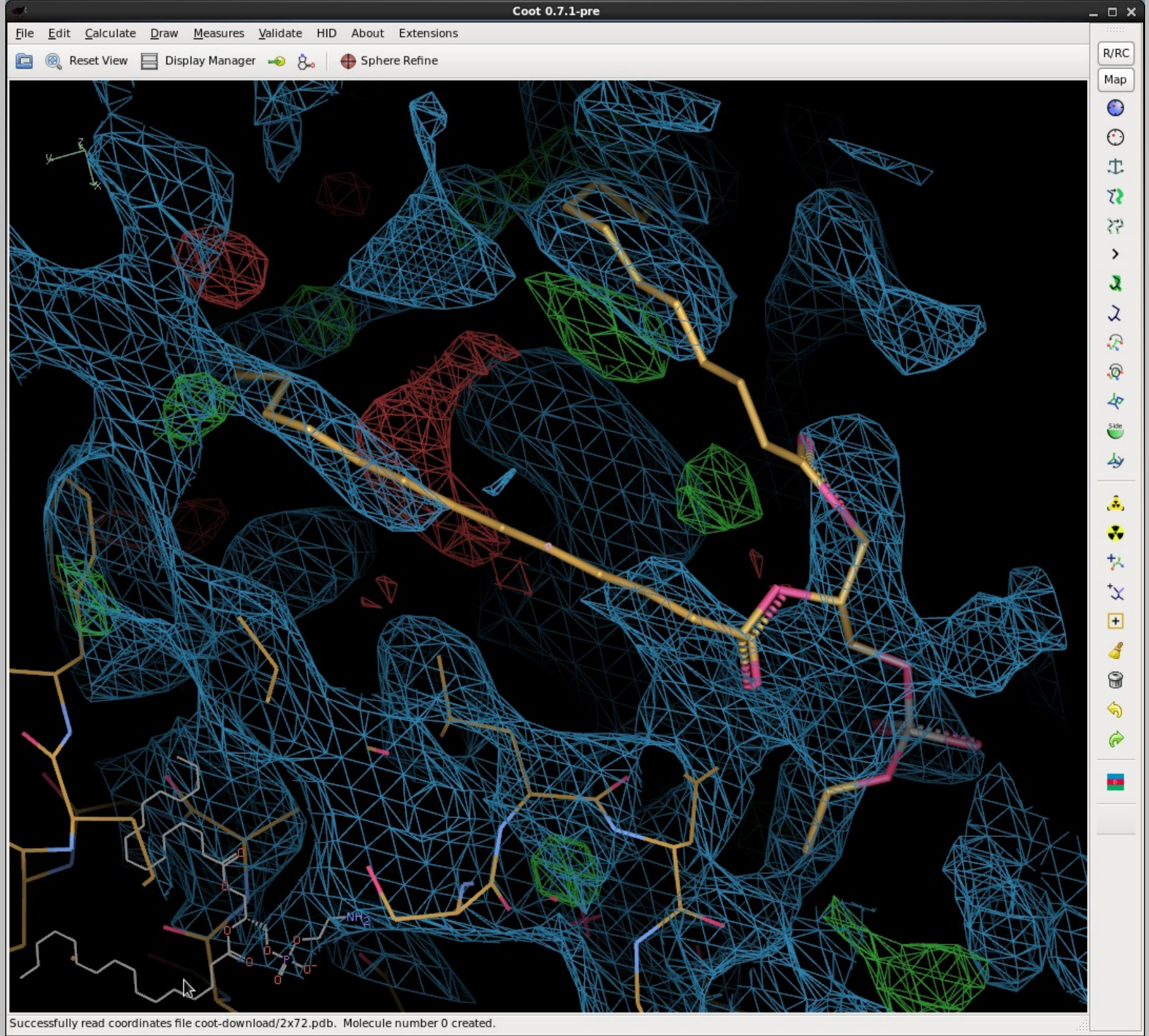
Protein-ligand complex models are often a result of subjective interpretation

Scoring Protein-Ligand Complexes

- Score all PDB protein-ligand complexes
  - No covalent link to protein
  - No alt confs
  - Hetgroups with more than 6 atoms
- Score:
  - Correlation of maps: omit vs calculated
    - around the ligand
  - Mogul distortion
    - z-worst
  - Clash-score
    - c.f. Molprobity tool

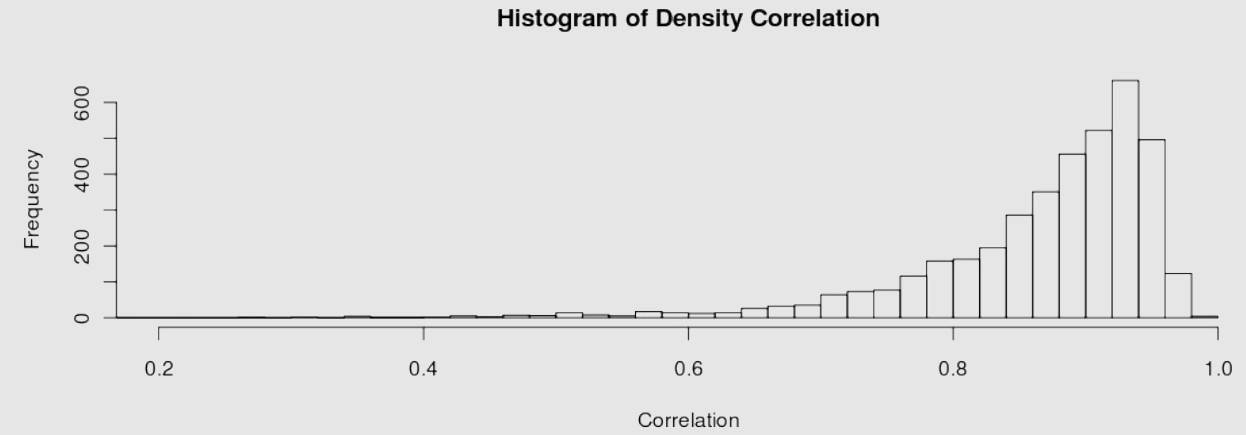

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
R/RC
Map
Side
CA /61 THR/A
(mol. no: 14) CA /1/A/61 THR occ: 1.00 bf: 111.44 ele: C pos: (7.64,33.49,-8.76)

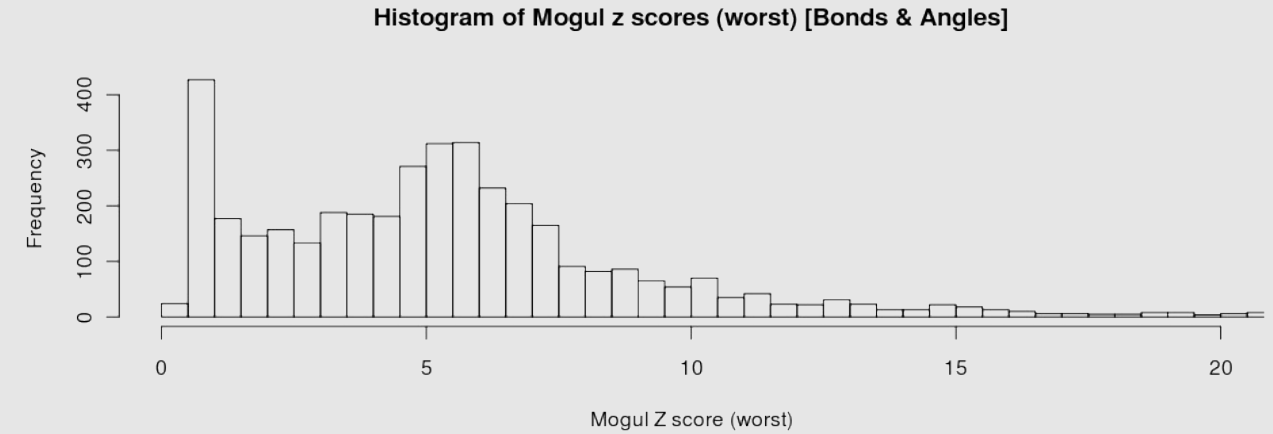

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
R/RC
Map
Side

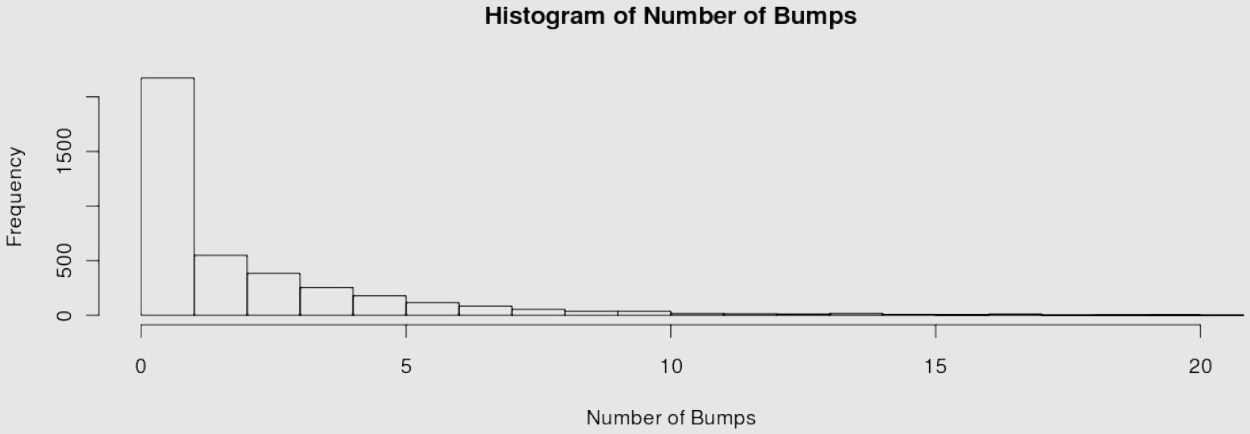

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
Sphere Refine
R/RC
Map
Side
NPP
Successfully read coordinates file coot-download/2x72.pdb. Molecule number 0 created.

Ligand Scoring

Preliminary results & conclusions...


Histogram of Density Correlation
Frequency
0
200
400
600
0.2
0.4
0.6
0.8
1.0
Correlation


Histogram of Mogul z scores (worst) [Bonds & Angles]
Frequency
0
100
200
300
400
0
5
10
15
20
Mogul Z score (worst)

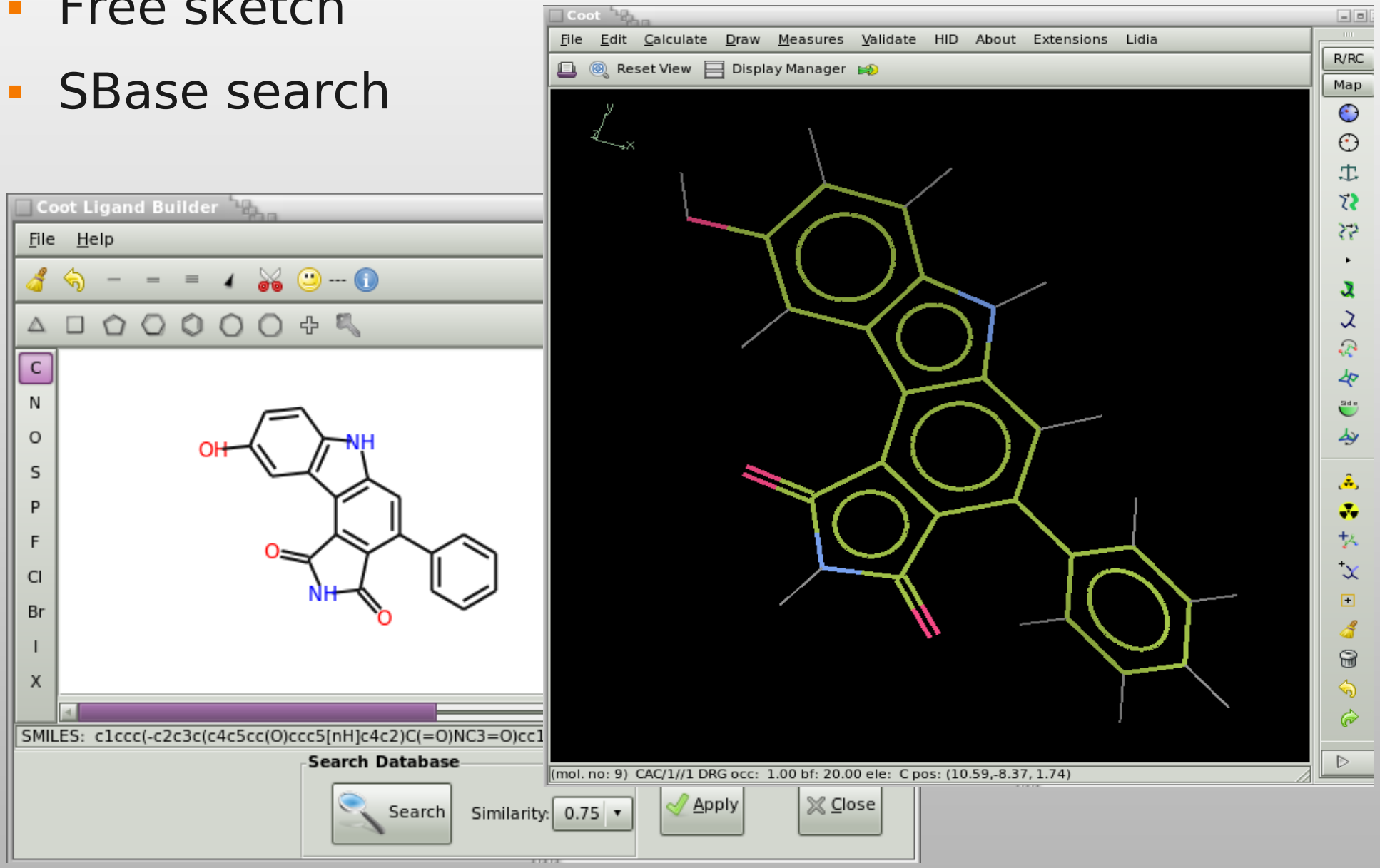

Histogram of Number of Bumps
Frequency
0
500
1500
0
5
10
15
20
Number of Bumps

Scoring Ligands:
To Be Better Than The Median:

- 0 bumps
- Mogul z(worst) < 5.4
  - But probably < 3.0 (when geometry is fixed)
- Density correlation > 0.9
  - resolution dependence?
  - do we believe the resolution in the data file?

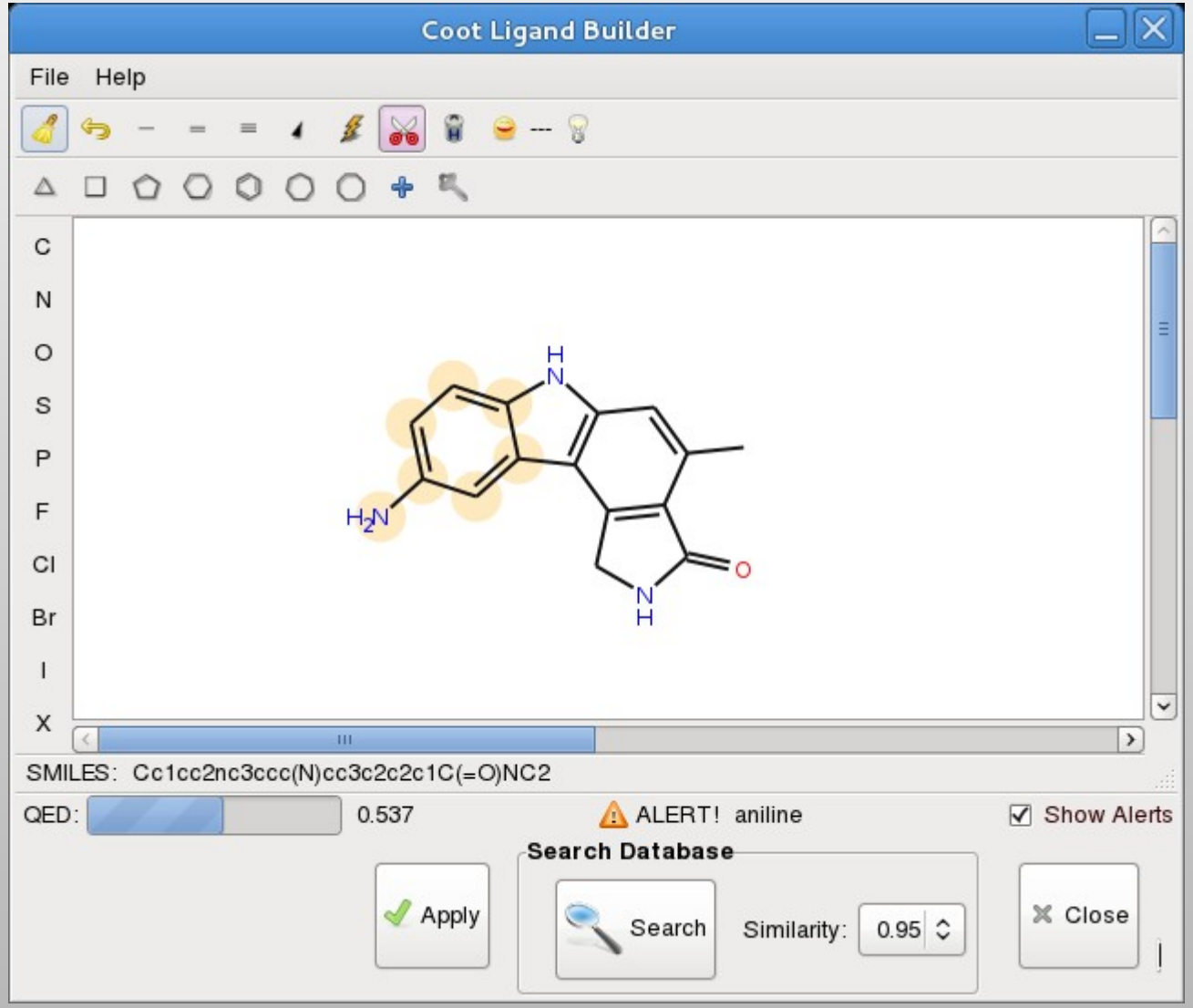
2D Ligand Builder

- Free sketch
- SBase search



2D Sketcher

- Structural Alerts



•On the fly ROMol creation
•Check *vs*. vector of SMARTS
  - (from Biscu-it)
  - And user-defined (python variable) list

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'[6], 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

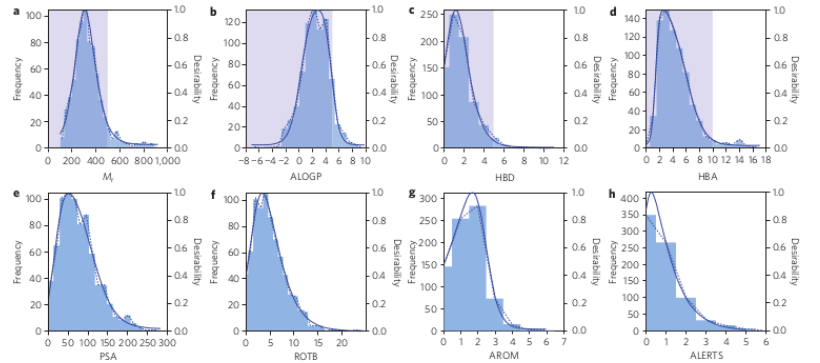


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h. Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**), 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

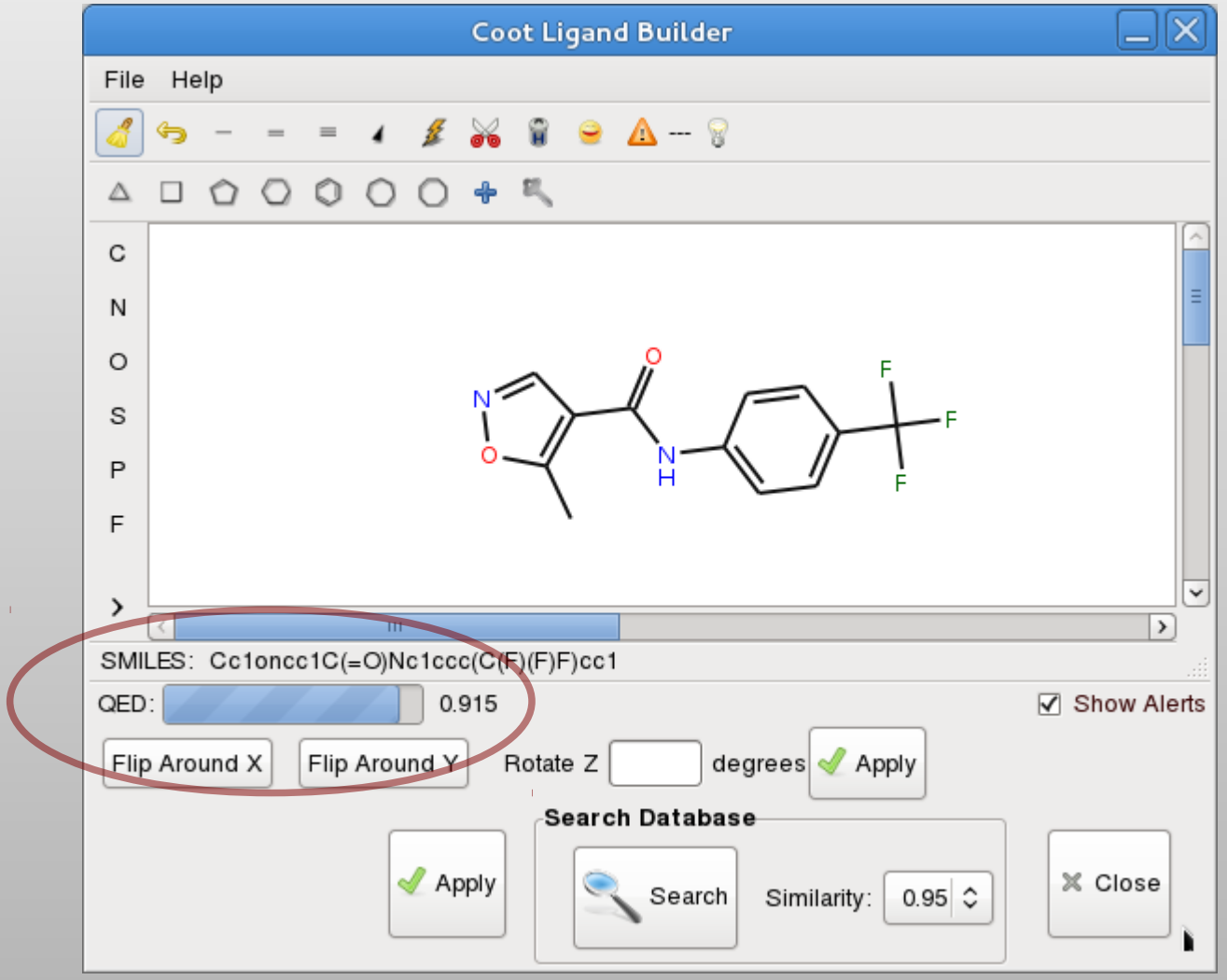
asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

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

Bickerton et al (2012) *Nature Chemistry*

2D Sketcher

- QED score



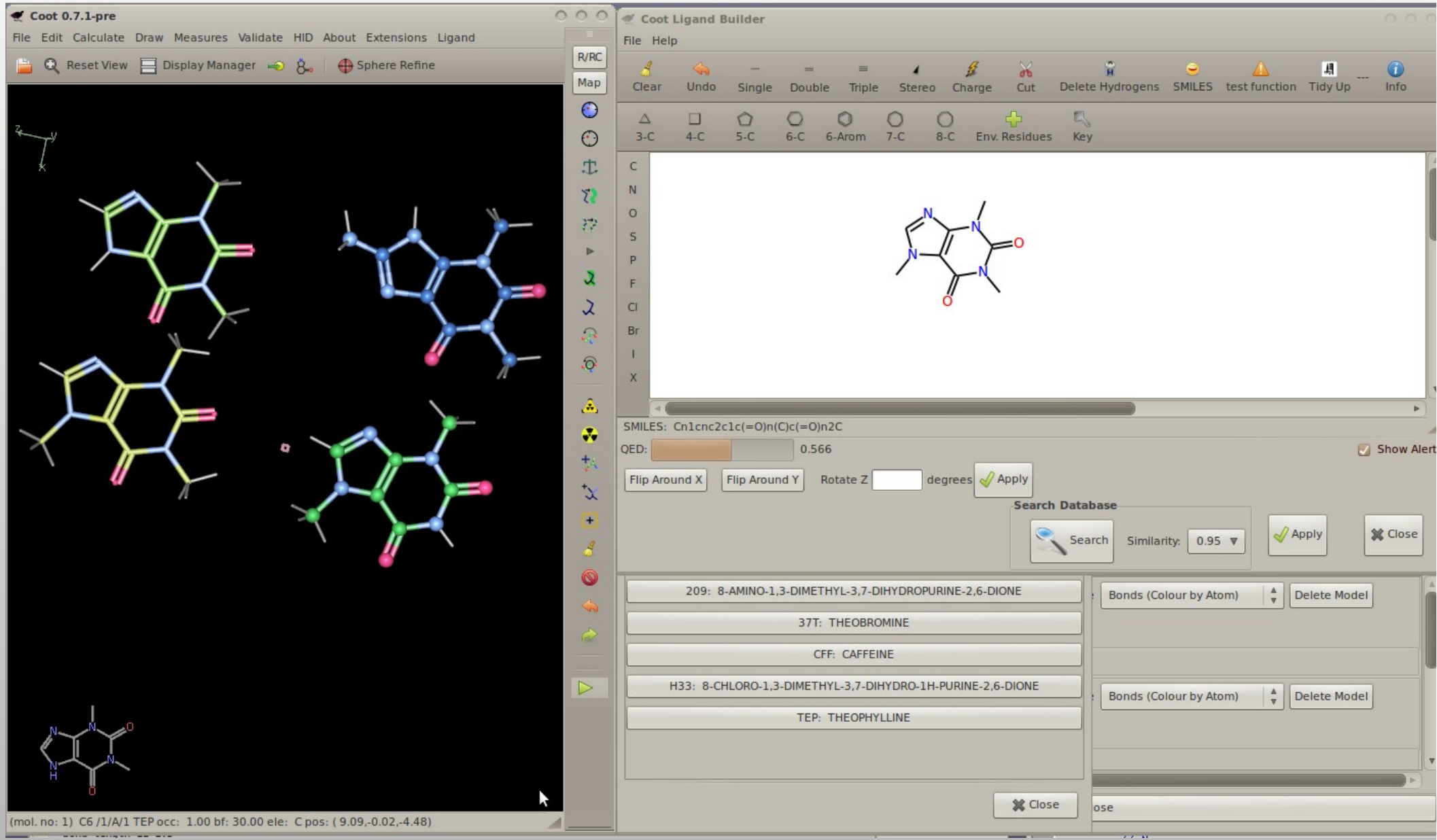
Silicos-it's Biscu-it™

Look up the function with PyModule_GetDict() and PyModule_GetItem()

Ligand Utils

- “Get Drug”
  - Uses network connection to Wikipedia
- Get *comp-id* ligand-description from PDBe
  - downloads and reads (*e.g.*) AAA.cif
    - (extracted from chemical component library)
- Drag and drop
  - Uses network connection to get URLs
  - or file-system files
- pyrogen
  - restraints generation

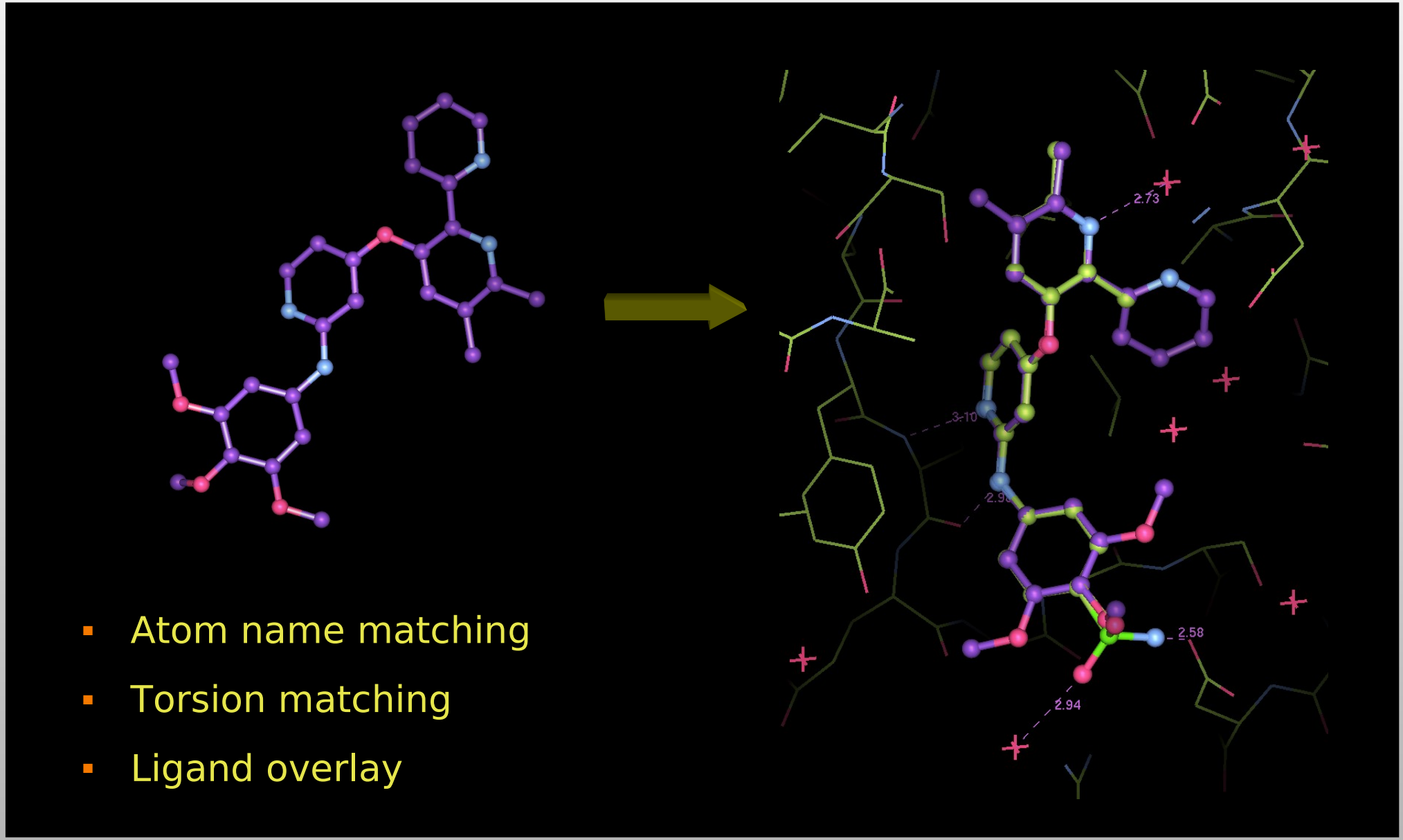
Ligand Utils – CCP4 SRS



Manipulating Ligands

Using ”Yesterday's” Ligand

Common subgraph isomorphism, Krissinel & Henrick (2004)



- Atom name matching
- Torsion matching
- Ligand overlay

Generating Conformers

- Using restraint information...

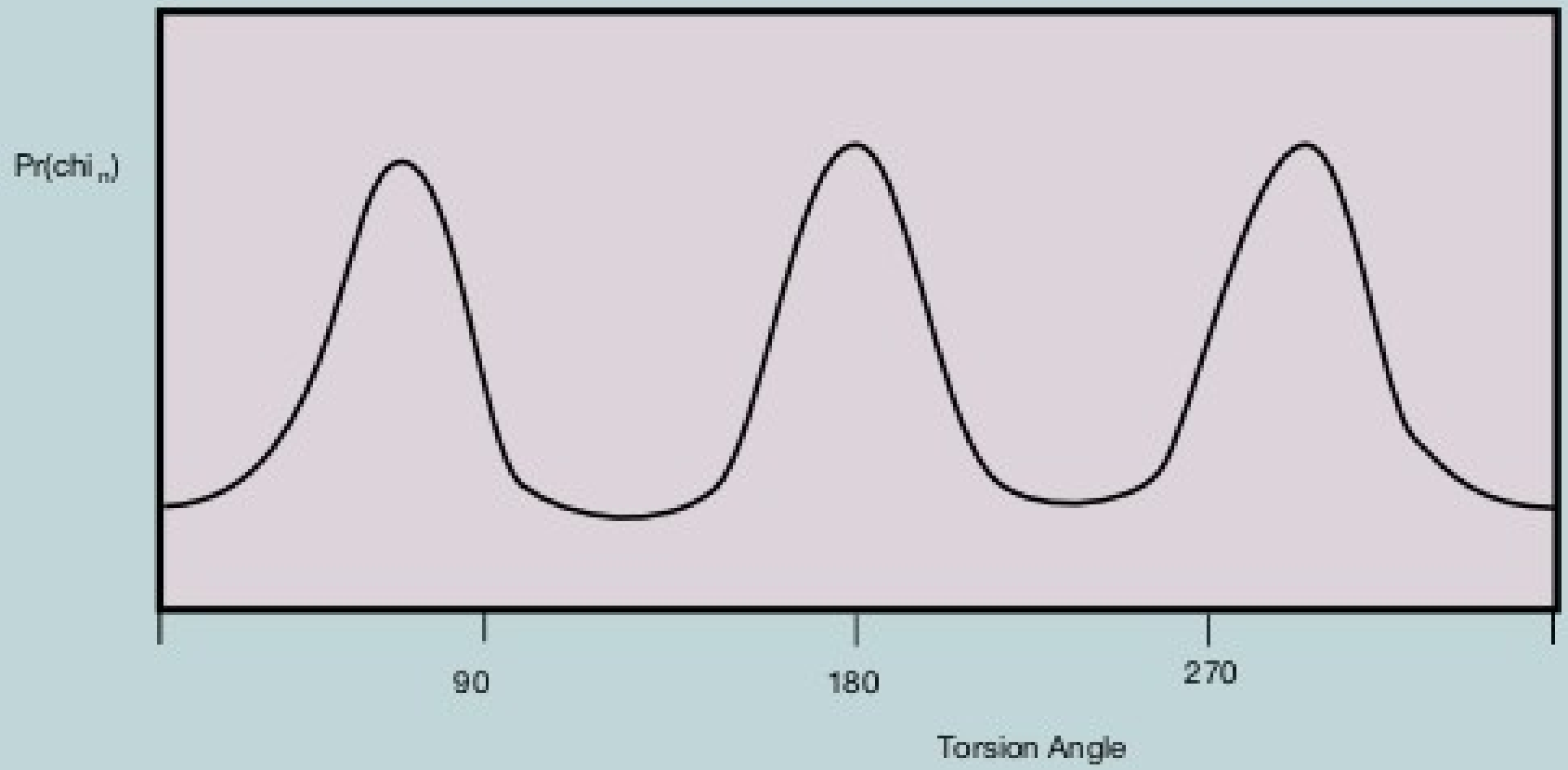
REFMAC Monomer Library
chem_comp_bond

```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 ALA      N      H         single      0.860    0.020
 ALA      N      CA        single      1.458    0.019
 ALA      CA     HA        single      0.980    0.020
 ALA      CA     CB        single      1.521    0.020
 ALA      CB     HB1       single      0.960    0.020
 ALA      CB     HB2       single      0.960    0.020
```

REFMAC Monomer Library
chem_comp_tor

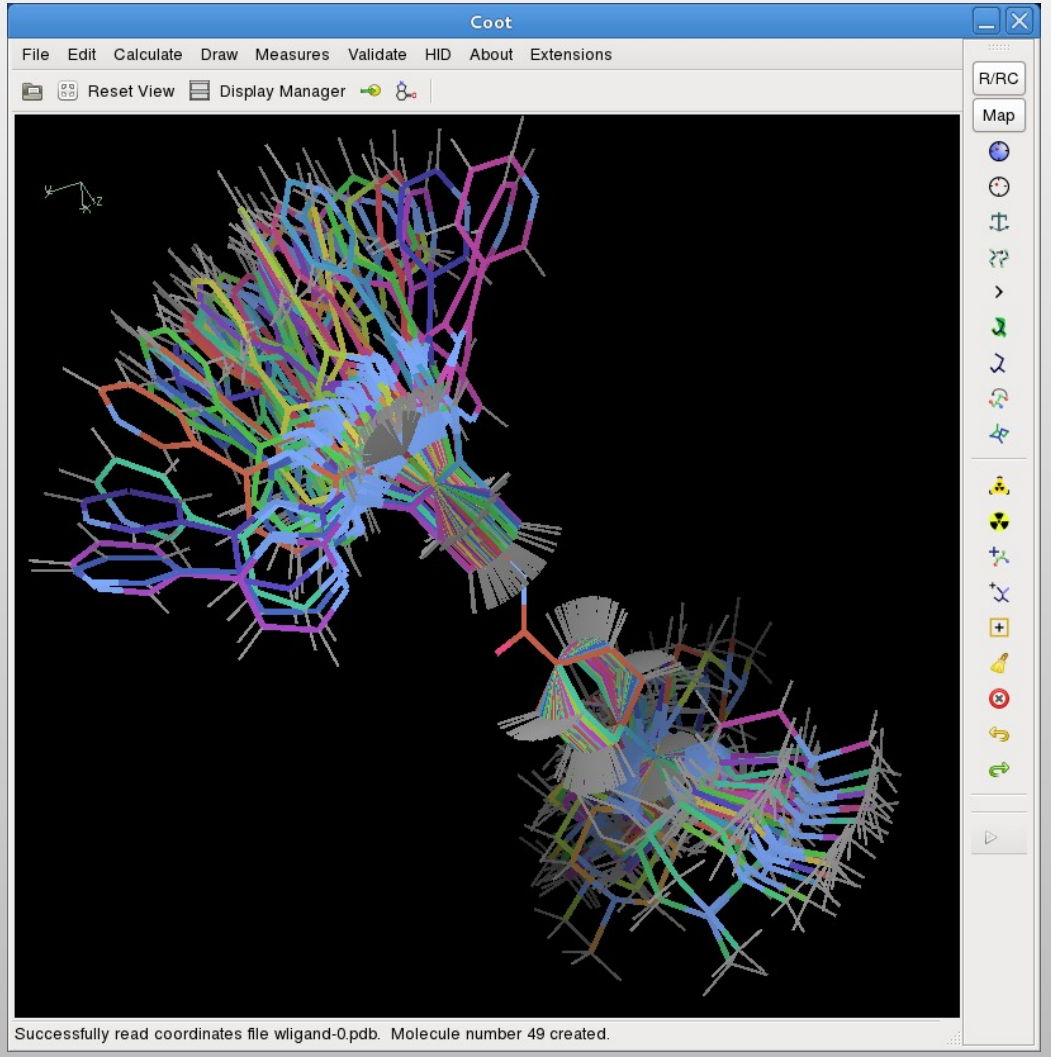
```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB        60.005   20.000   1
 ADP      var_2    PA     O3A    PB     O1B       59.979   20.000   1
 ADP      var_3    O2A    PA     "O5'"  "C5'"    -59.942   20.000   1
 ADP      var_4    PA     "O5'"  "C5'"  "C4'"    179.996   20.000   1
 ADP      var_5    "O5'"  "C5'"  "C4'"  "C3'"    176.858   20.000   3
 ADP      var_6    "C5'"  "C4'"  "O4'"  "C1'"    150.000   20.000   1
 ADP      var_7    "C5'"  "C4'"  "C3'"  "C2'"   -150.000   20.000   3
```

Ligand Torsionable Angle Probability from CIF file

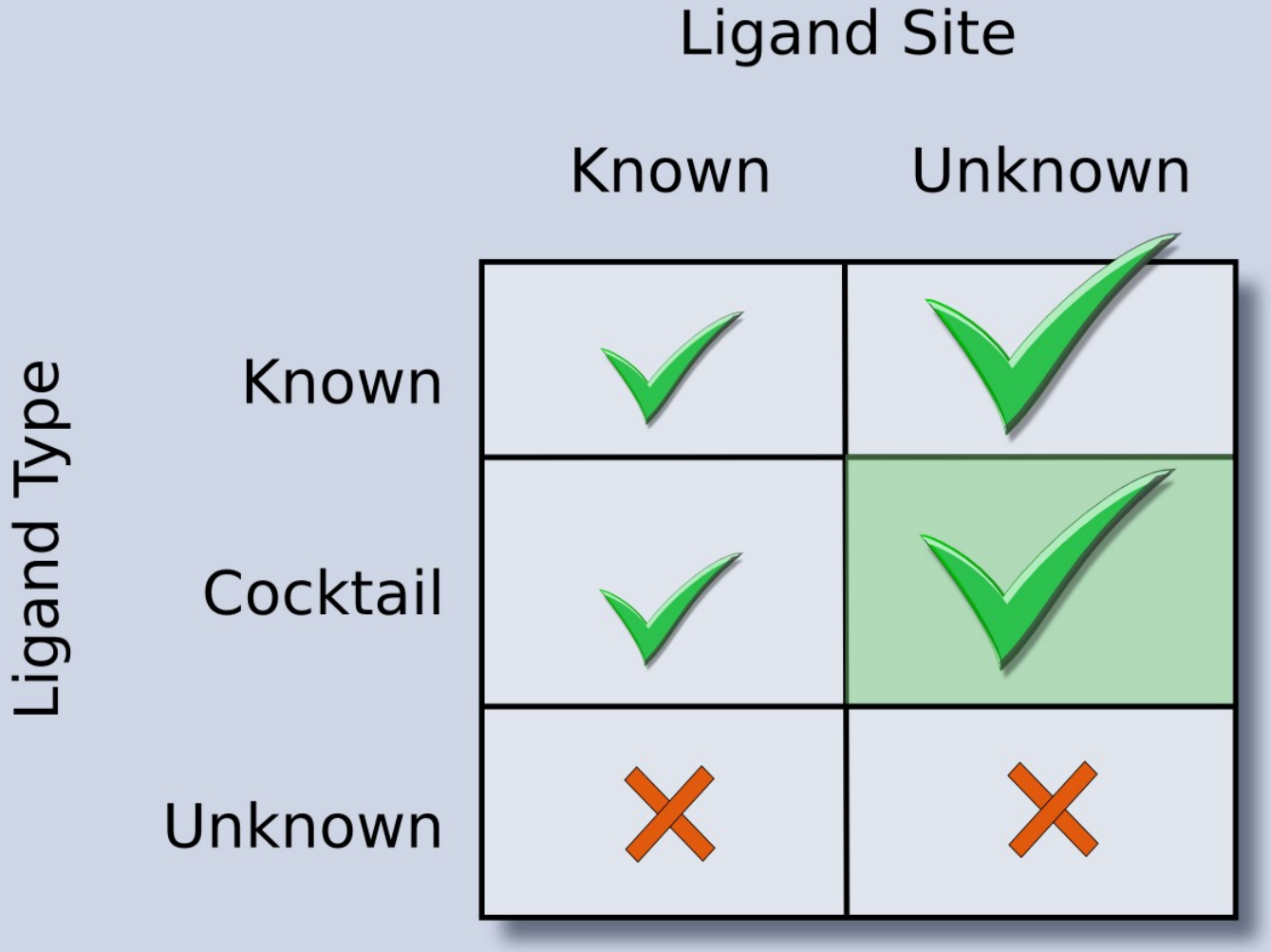


Conformer Generation

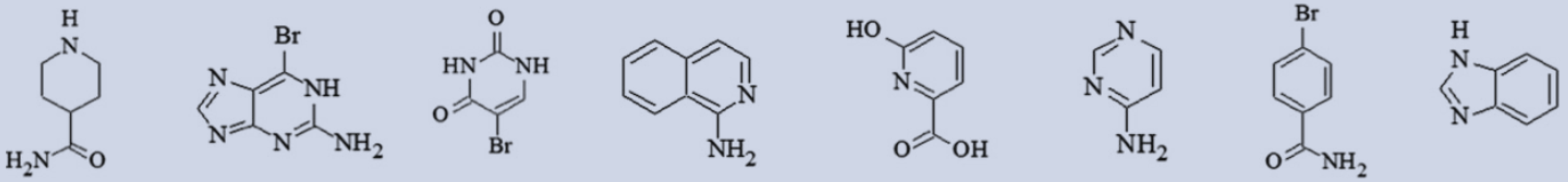
Non-Hydrogen
Non-CONST
Non-Ring



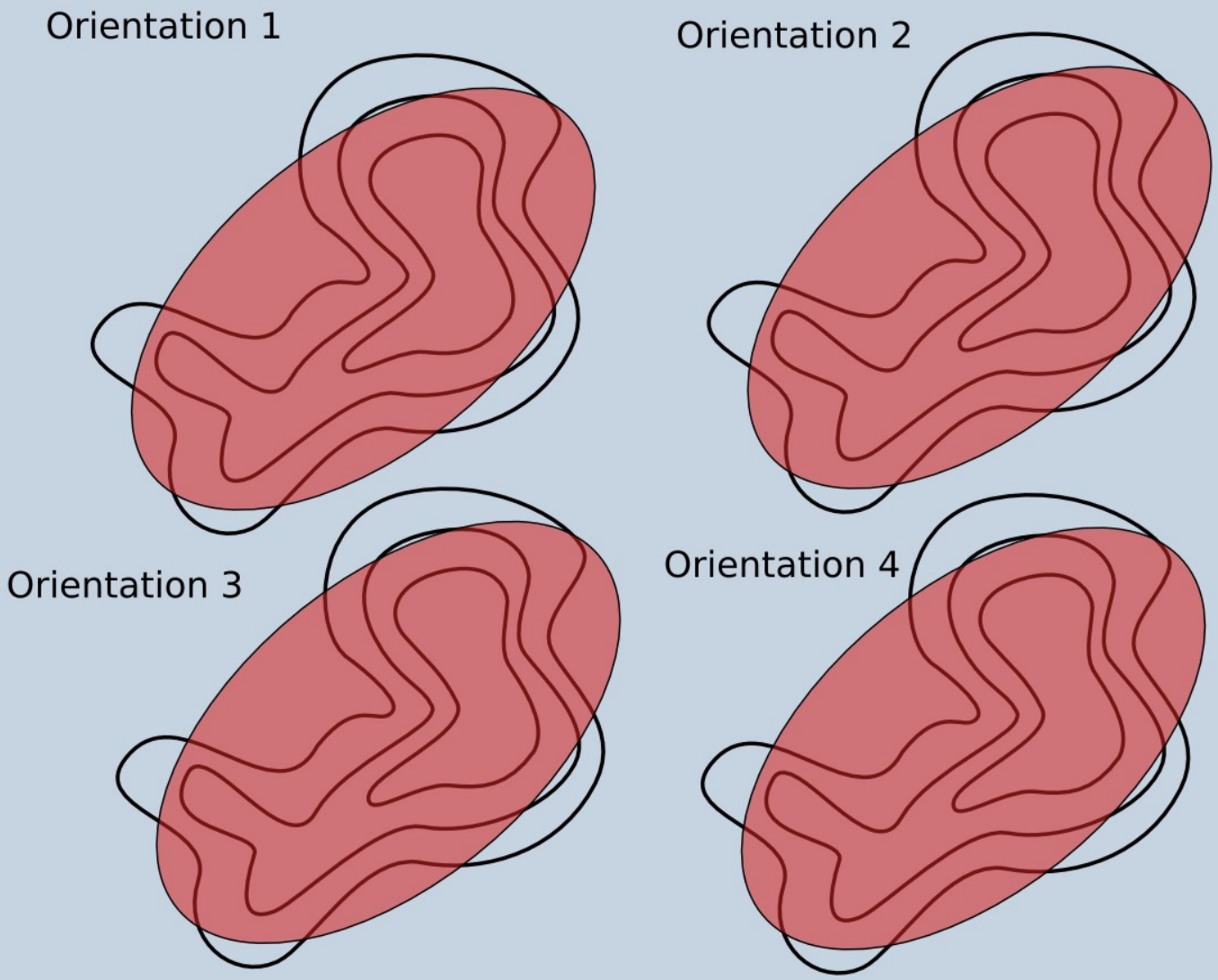
Fitting Ligands



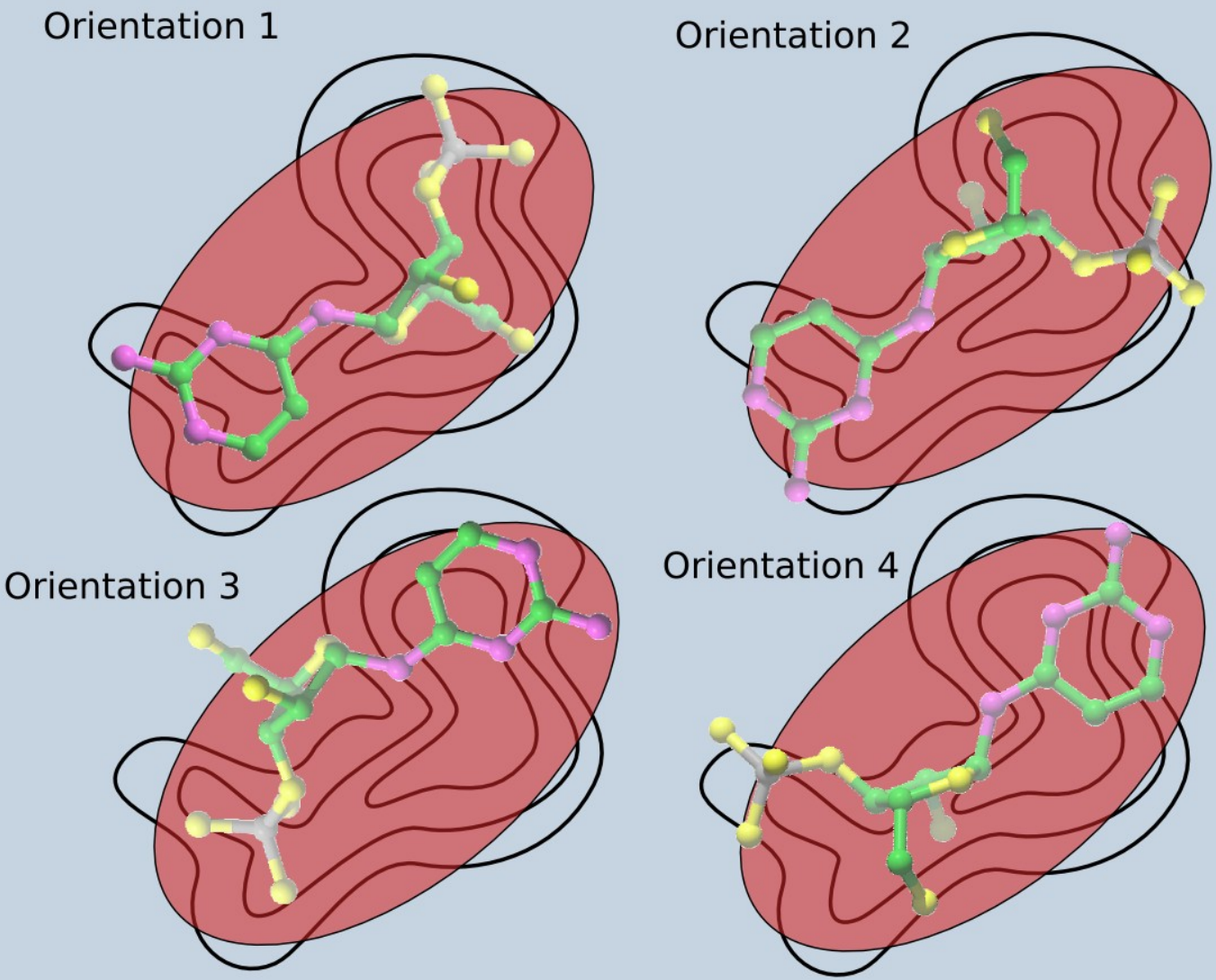
Cocktail Examples



Orienting the Ligand

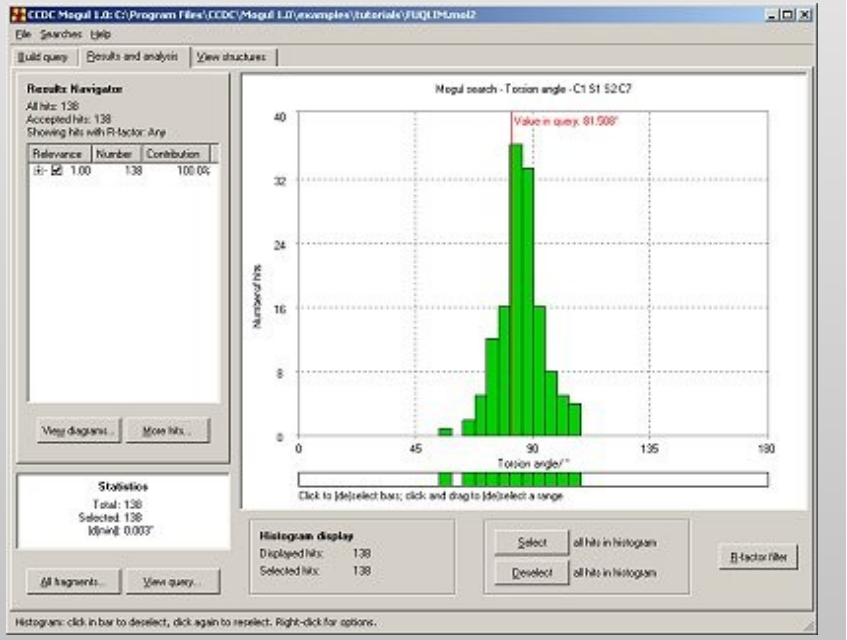


Orienting the Ligand



Parmatisation issues... (what if they are wrong?)

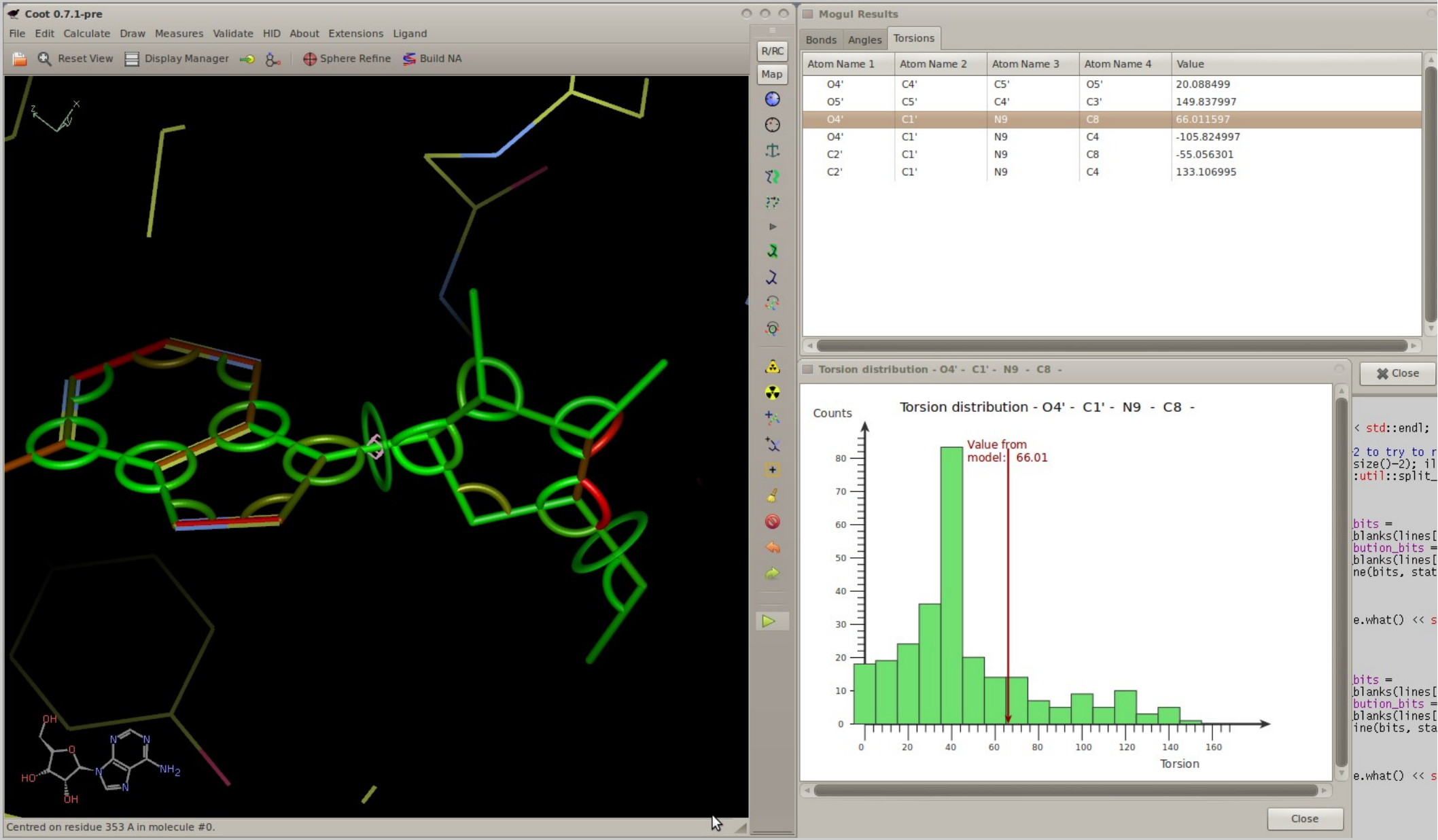
- Perfect refinement with incorrect parameters
  → distorted structure
- CSD's Mogul
  - Knowledge-base of geometric parameters based on the CSD
  - Can be run as a “batch job”
  - Mean, median, mode, quartiles, Z-scores.



Ligand Validation

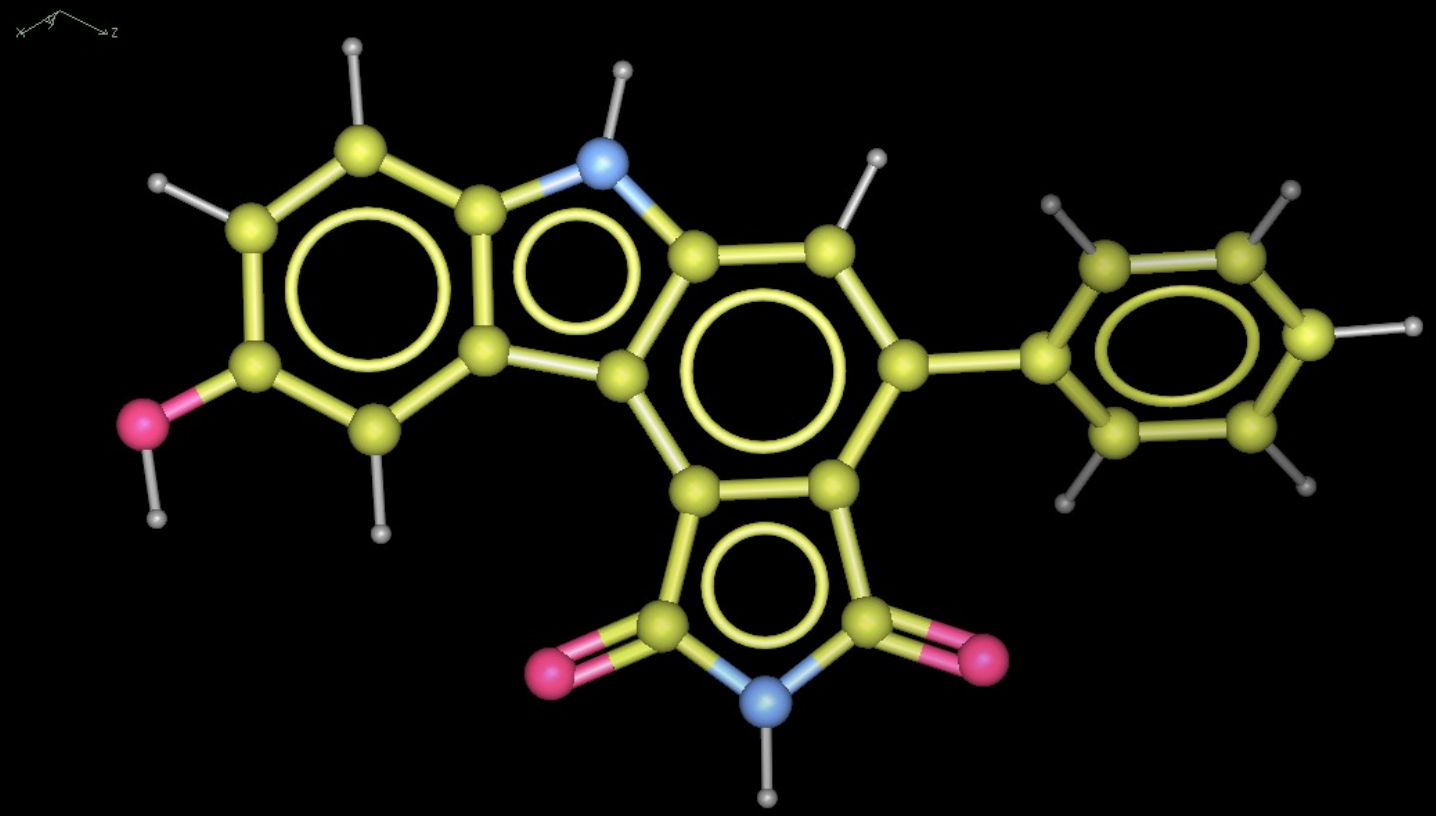
- Mogul plugin in Coot
  - Run mogul, graphical display of results
  - Update restraints (target and esds for bonds and angles)
  - CSD data not so great for plane, chiral and torsion restraints
    - (not by me, anyway)

Mogul Results Representation

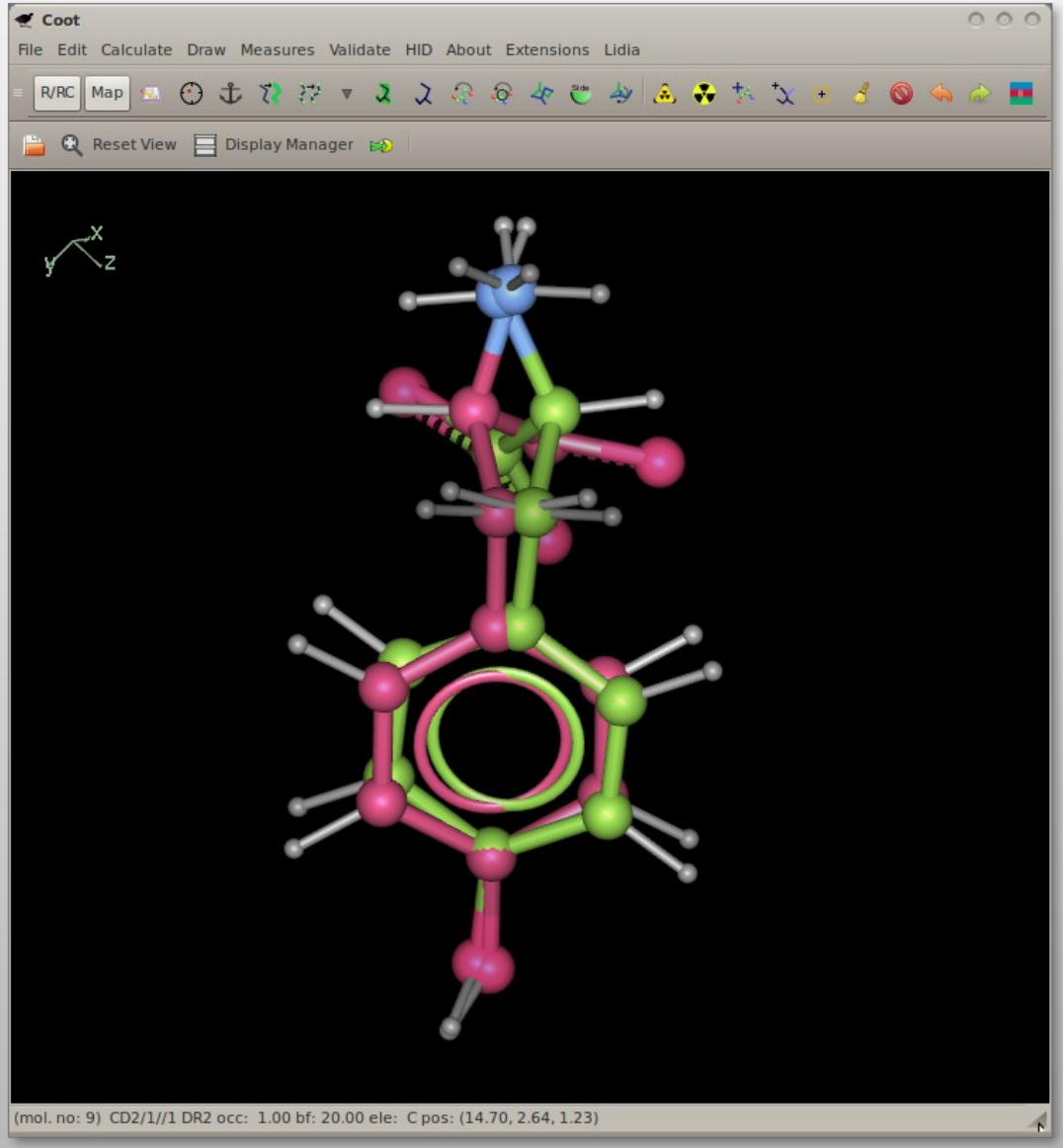


Ligand Represenation

- Bond orders (from dictionary restraints)



Chiral Centre Inversion

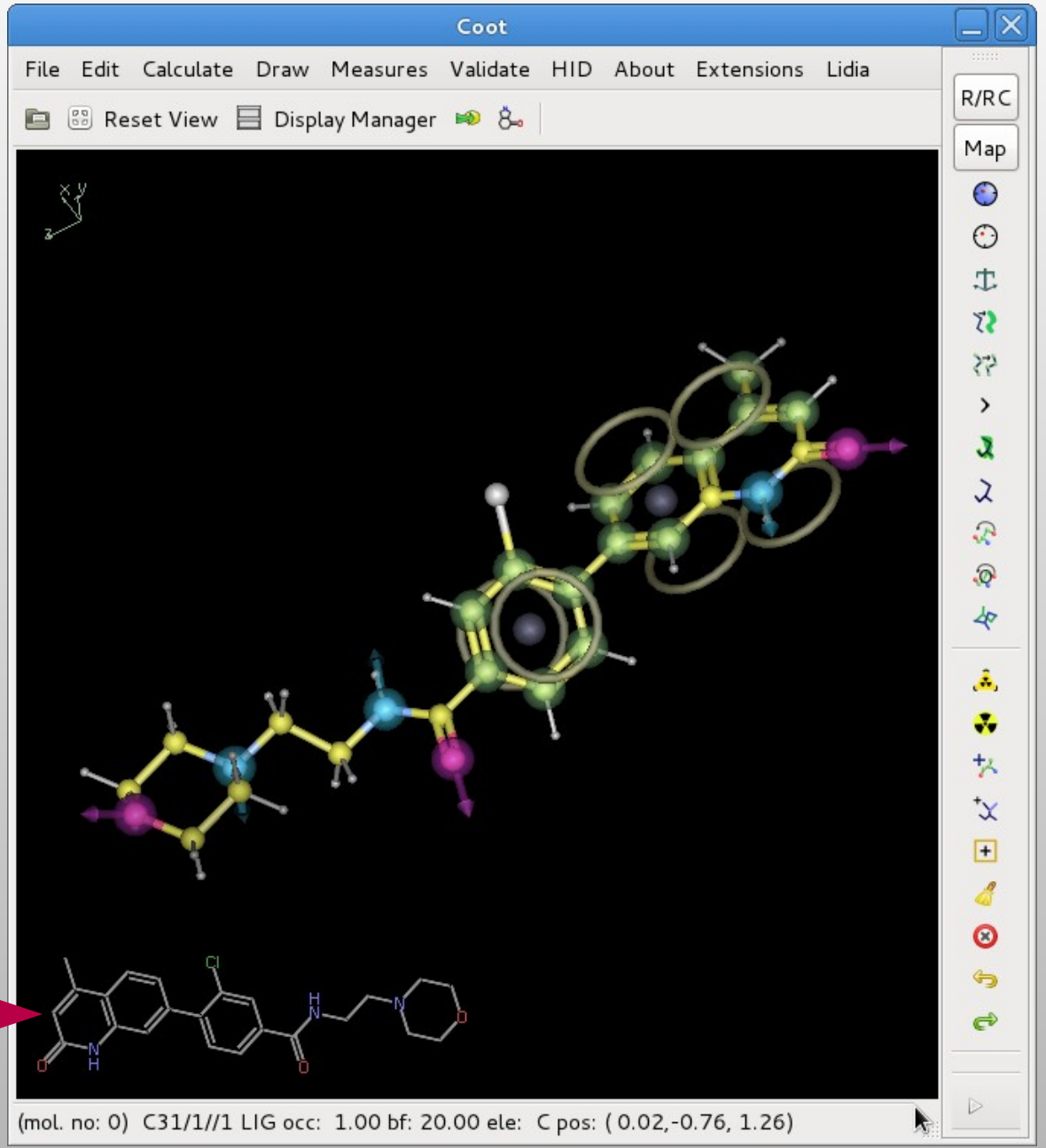


Inverted chiral centre refinement pathology detection

Hydrogen tunnelling

Chemical Features

Uses built-in FeatureFactory



...and on the fly thumbnailing

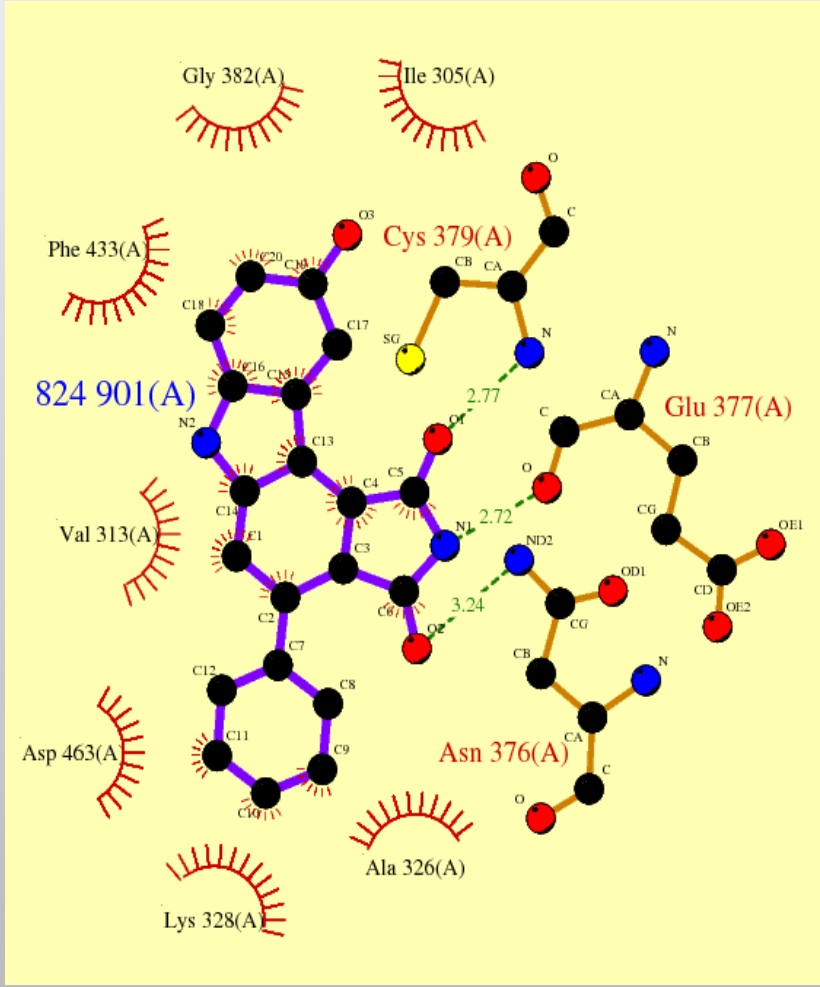
Pyrogen: *Coot*'s Ligand Generation Program

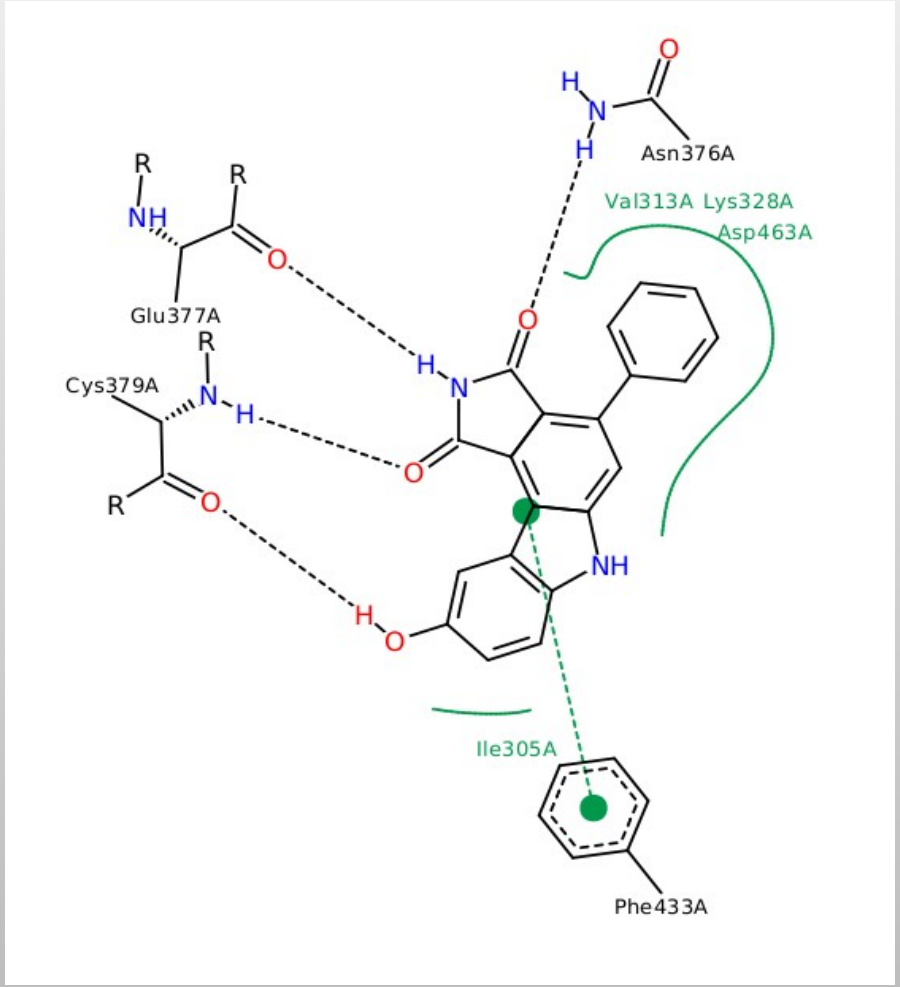
- Generate low energy conformer
- and mmCIF dictionary with full description
- Uses refmac atom types for (fast) lookup of geometry from the refmac library
- Updated by CCDC's mogul (if you have it)
- Or MMFF94 if not

- CCP4 has another dictionary generator: Acedrg

Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





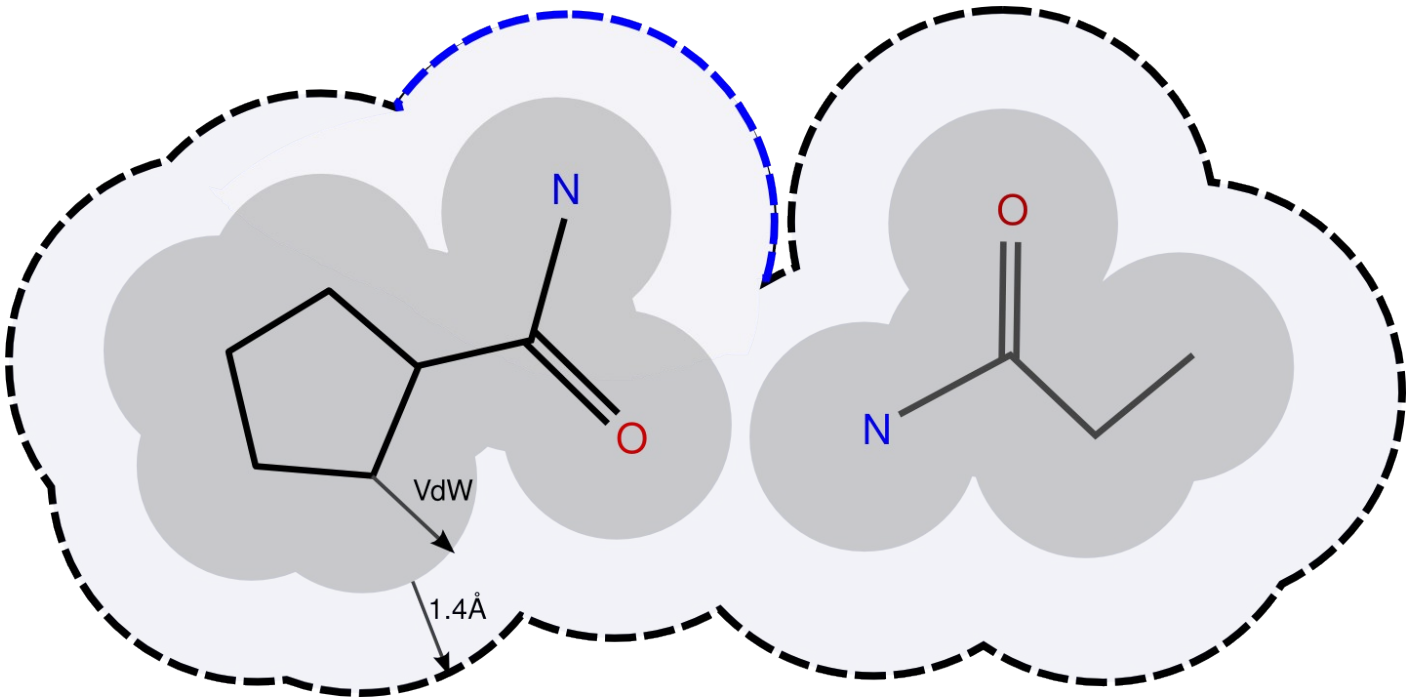
Can we do better? - Interactivity?

Ligand Environment Layout

- Binding pocket residues
- Interactions
- Substitution contour
- Solvent accessibility halos
- Solvent exclusion by ligand

Solvent Exposure

- Identification of solvent accessible atoms



Ligand Enviroment Layout

- Considerations
  - 2D placement and distances should reflect 3D metrics (as much as possible)
    - H-bonded residues should be close the atoms to which they are bonded
  - Residues should not overlap the ligand
  - Residues should not overlap each other
  - *c.f.* Clark & Labute (2007)

Layout Energy Terms

$$E = \sum\sum w_{ij}(d_{ij}^2 - D_{ij}^2) +$$

Residues match 3D Distances

$$\sum\sum \exp(-\frac{1}{2}d_{ij}^2) +$$

Residues don't overlay each other

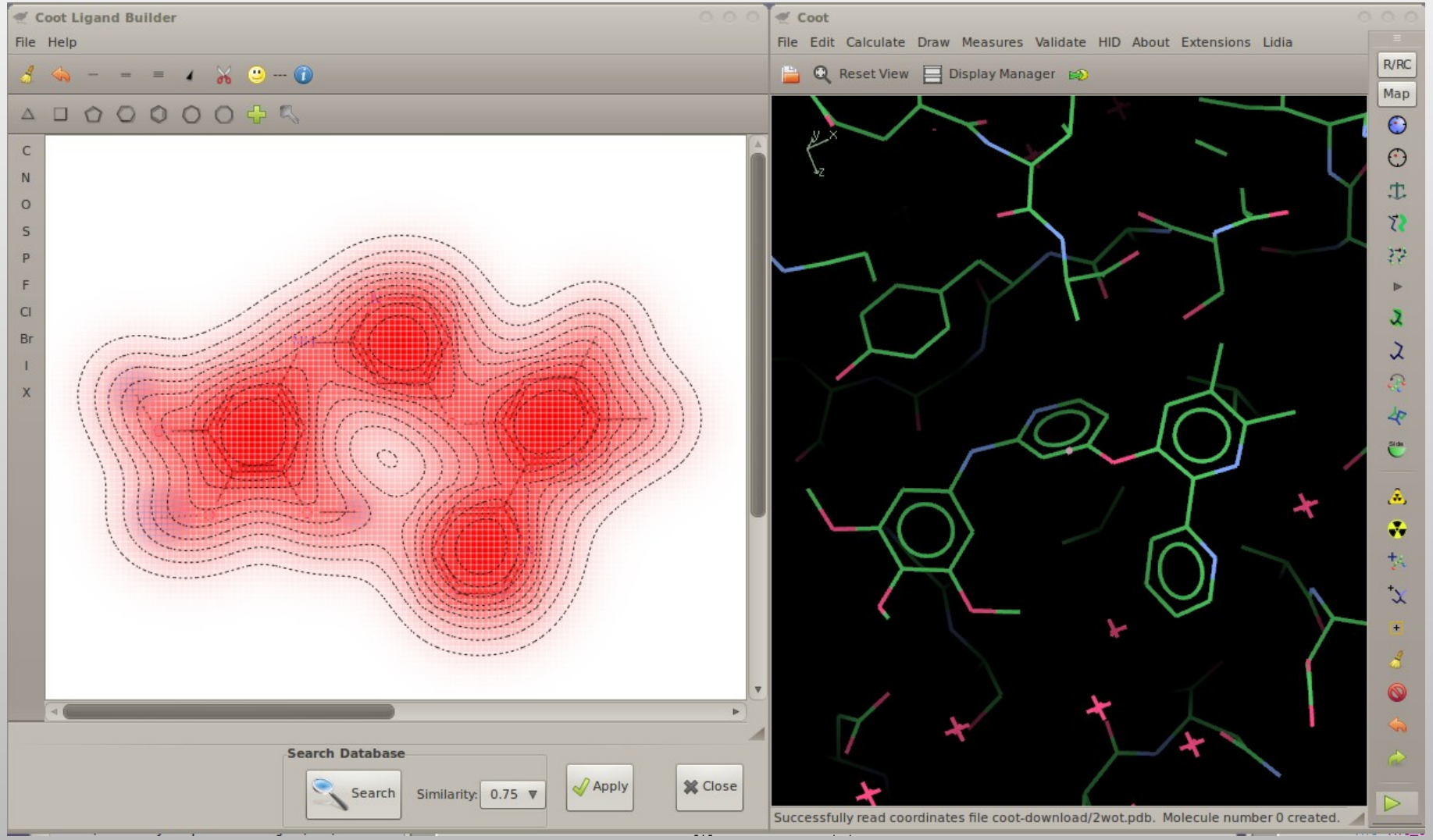
$$\sum\sum (d_{ik}^2 - D_{ik}^2) +$$

Residues are close to H-bonding ligand atoms

$$\sum\sum \exp(-\frac{1}{2}d_{ik}^2)$$

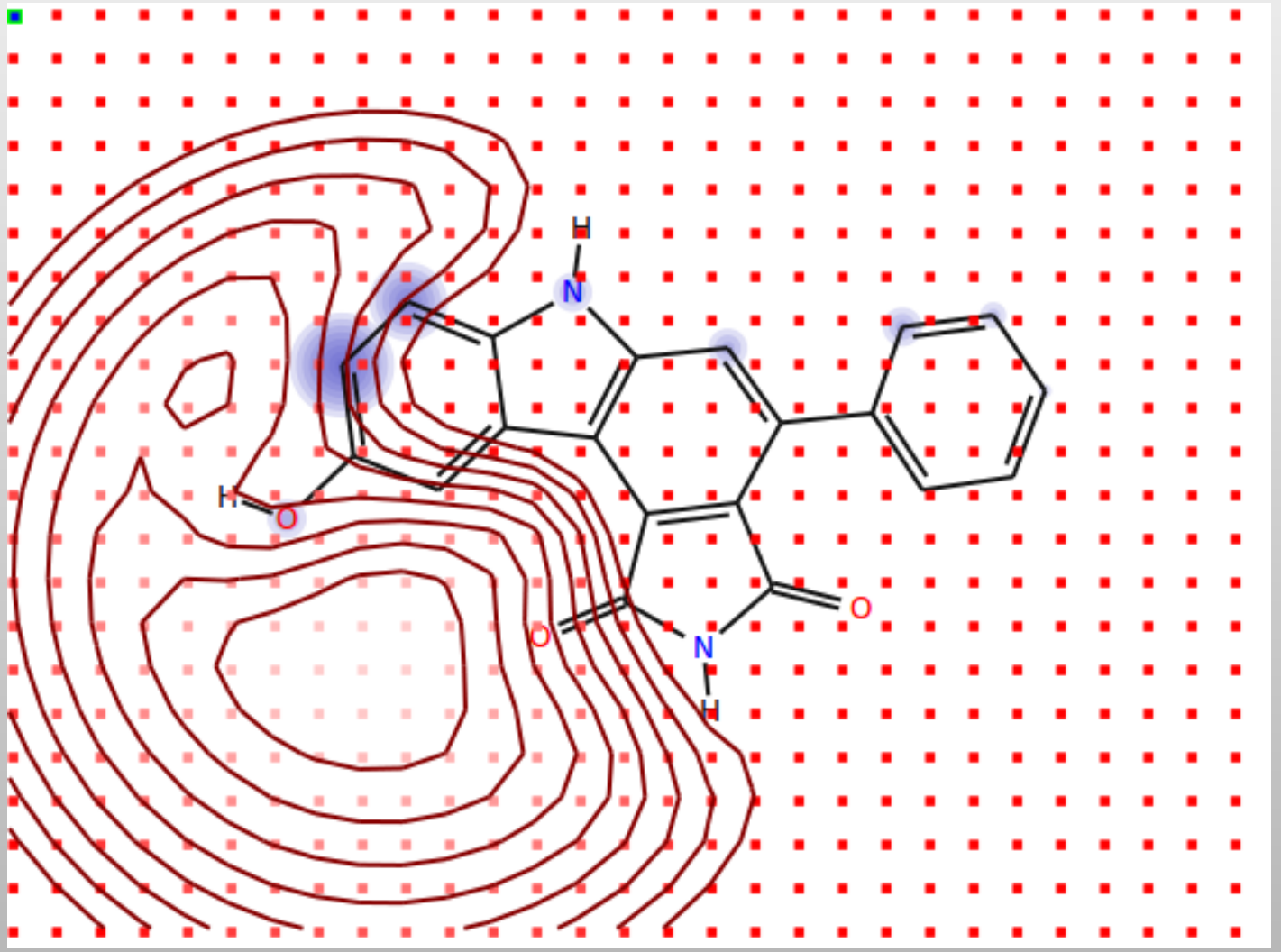
Residues don't overlap ligand

”Don't overlap the ligand”



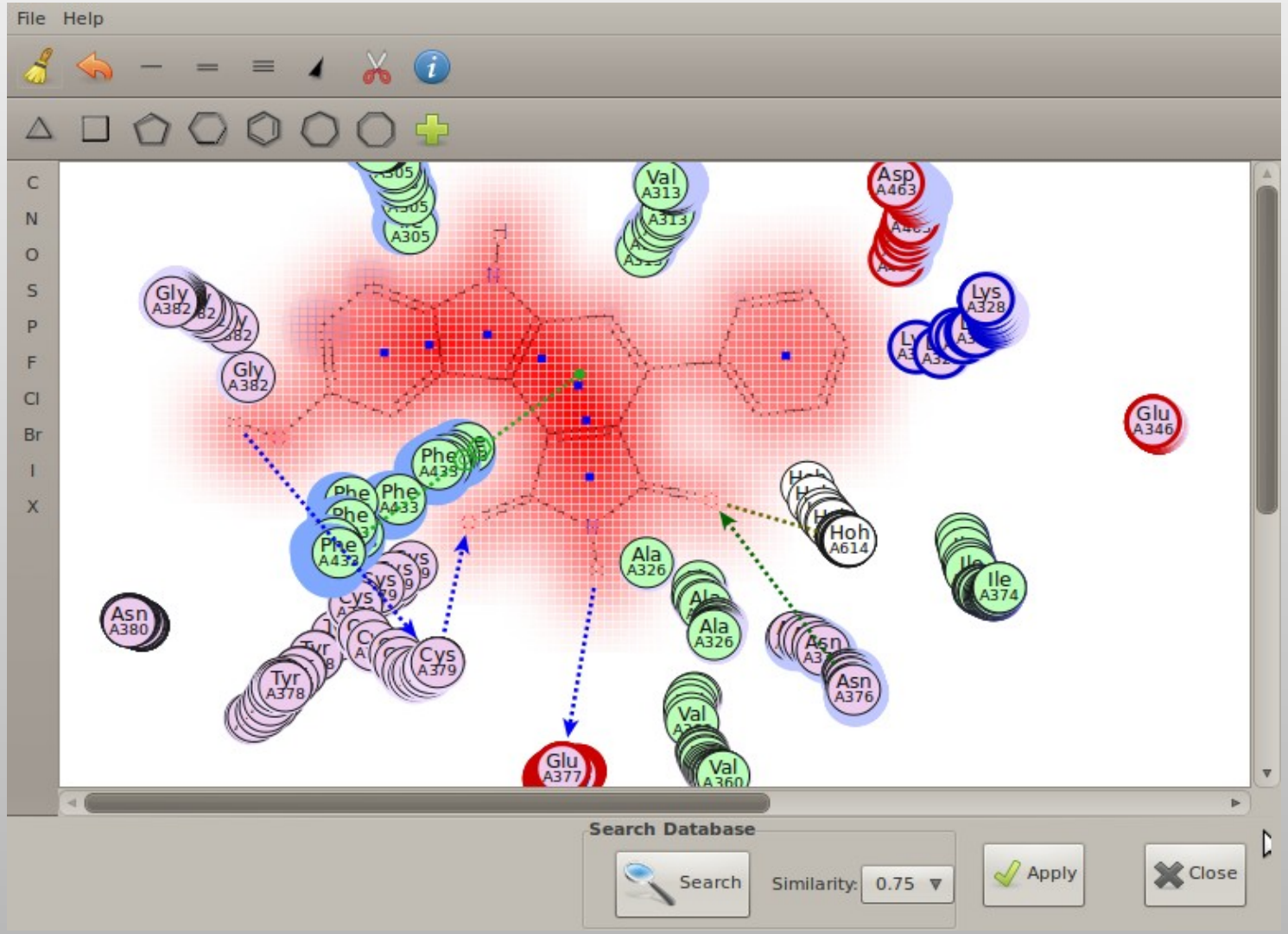
Ligand Environment Layout

- Initial residue placement



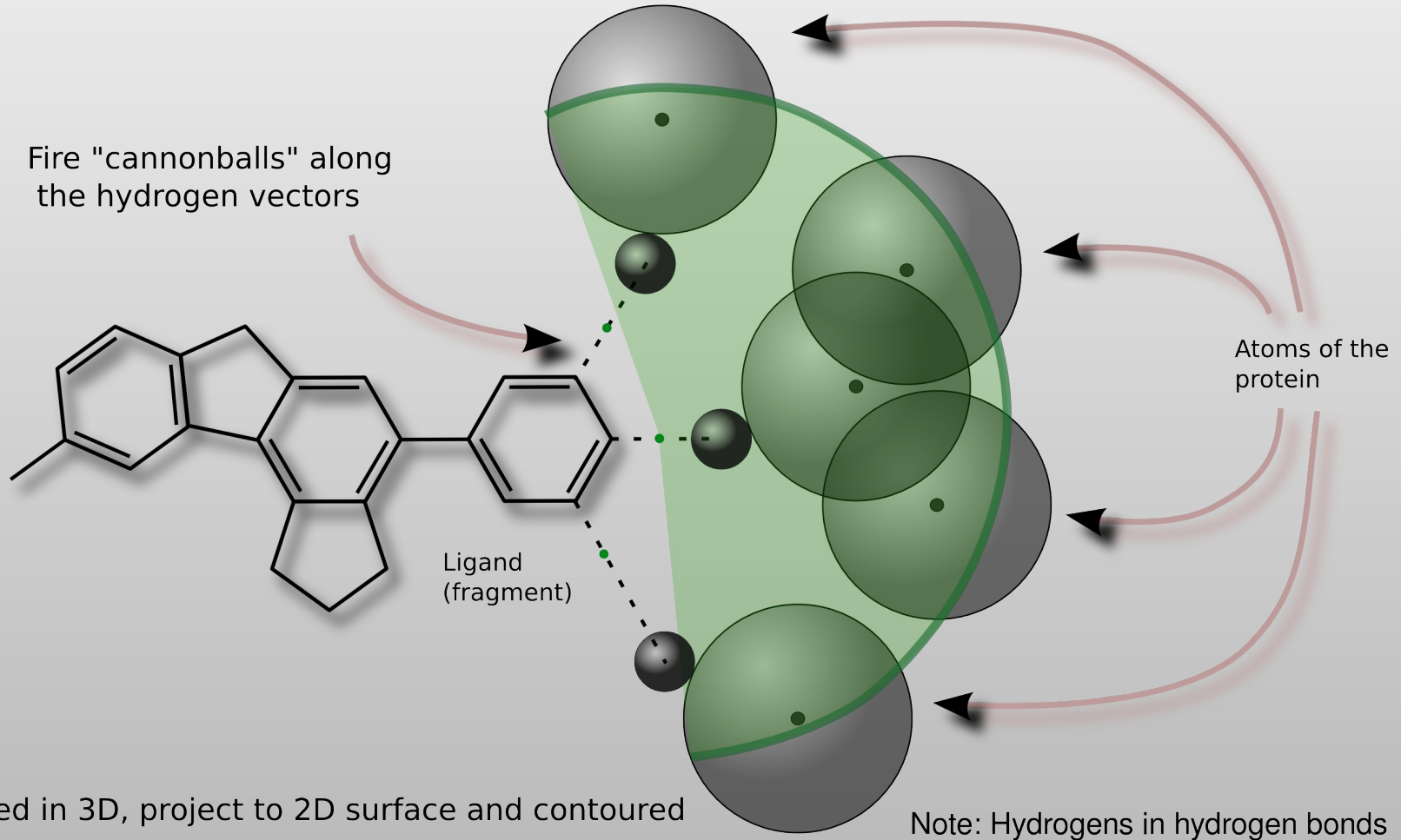
Ligand Environment Layout

- Residue position minimisation



Determination of the Substitution Contour

How far can we go (in the direction of the hydrogens) before hitting atoms of the protein?



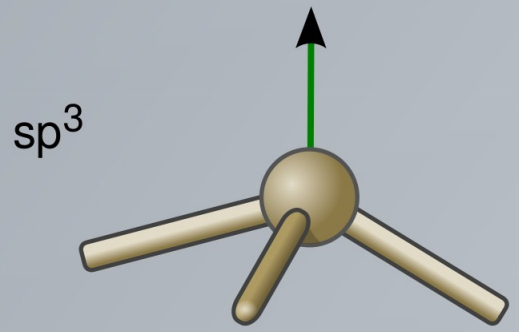
Determined in 3D, project to 2D surface and contoured

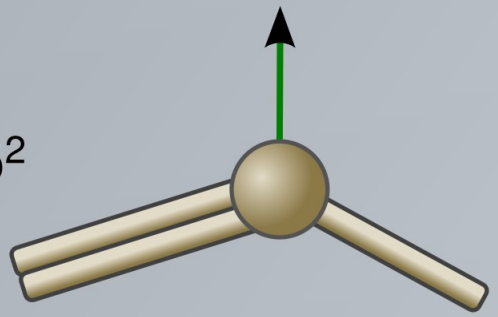
c.f. Clarke & Labute (2007)

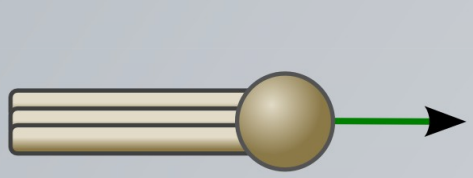
Note: Hydrogens in hydrogen bonds are a confounding factor

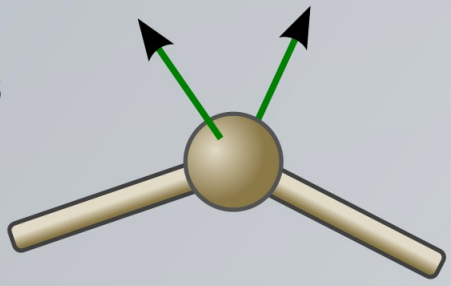
Substitution Contour: Extending along Hydrogens

Riding Hydrogens



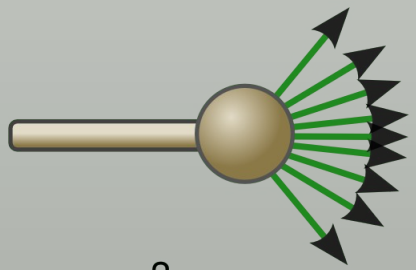


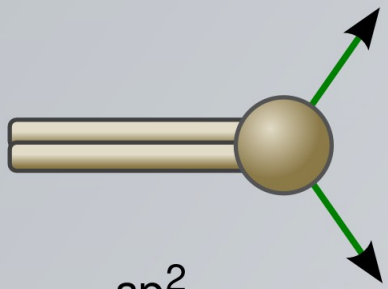


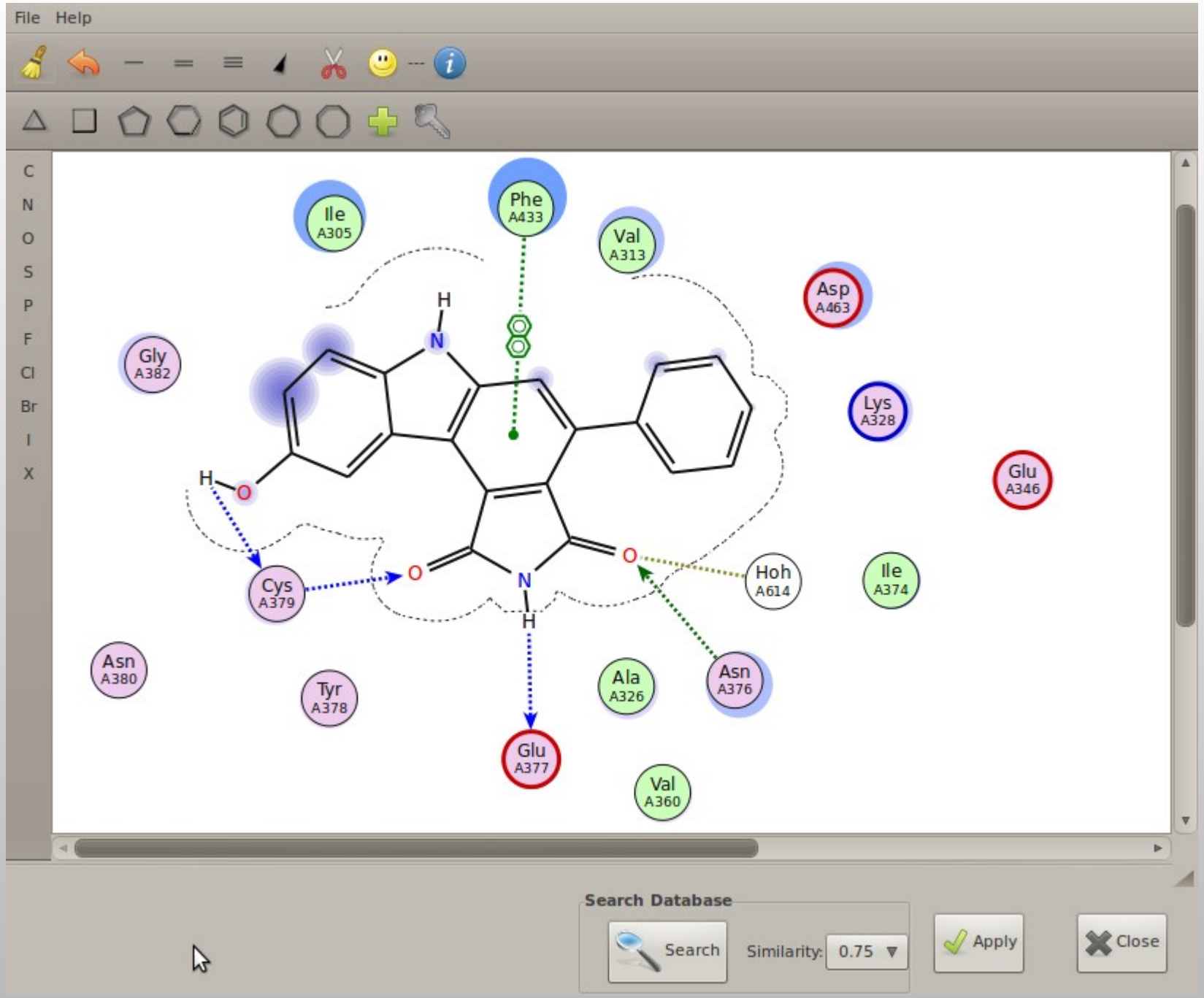


Torsionable Hydrogens

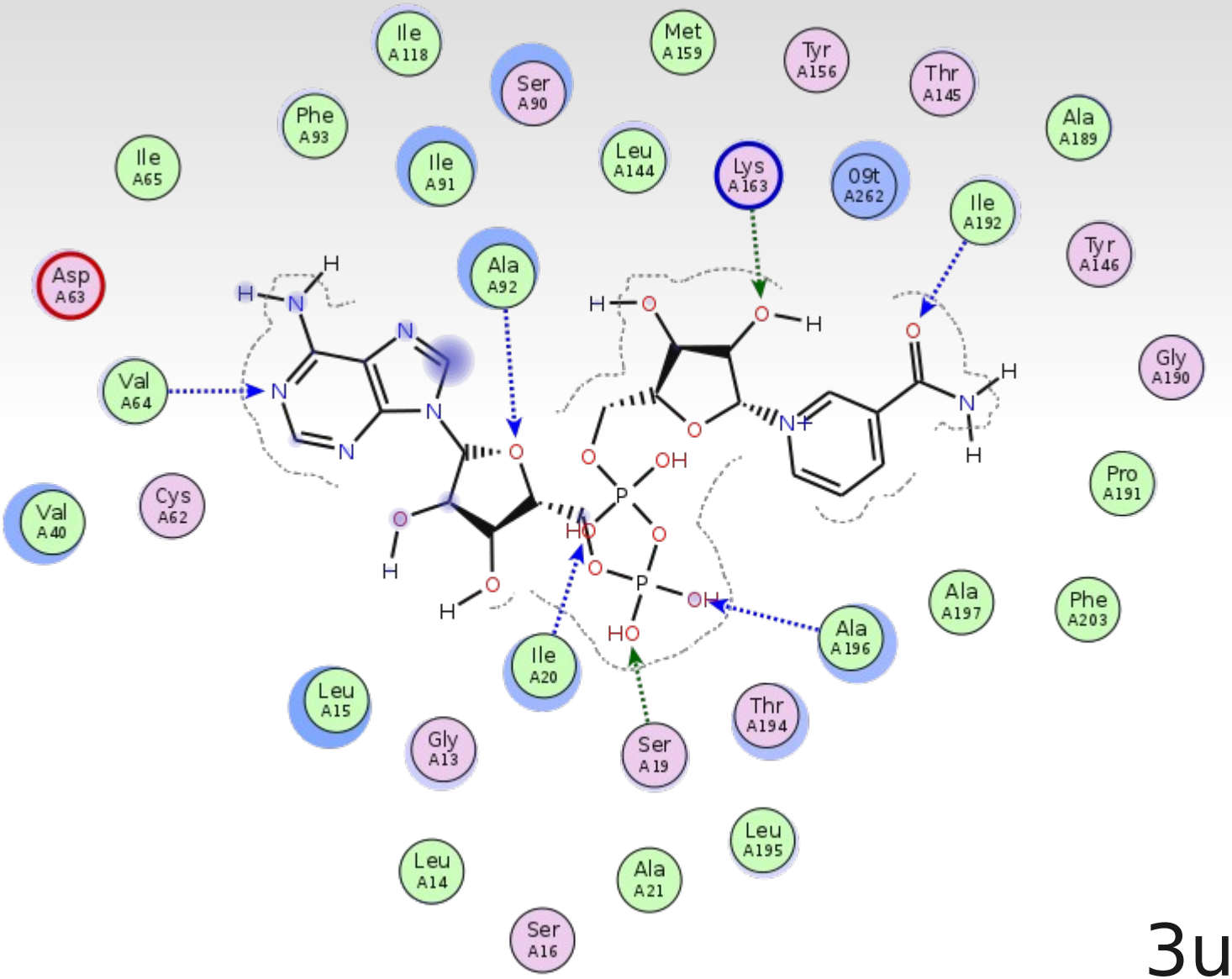
(test multiple directions)



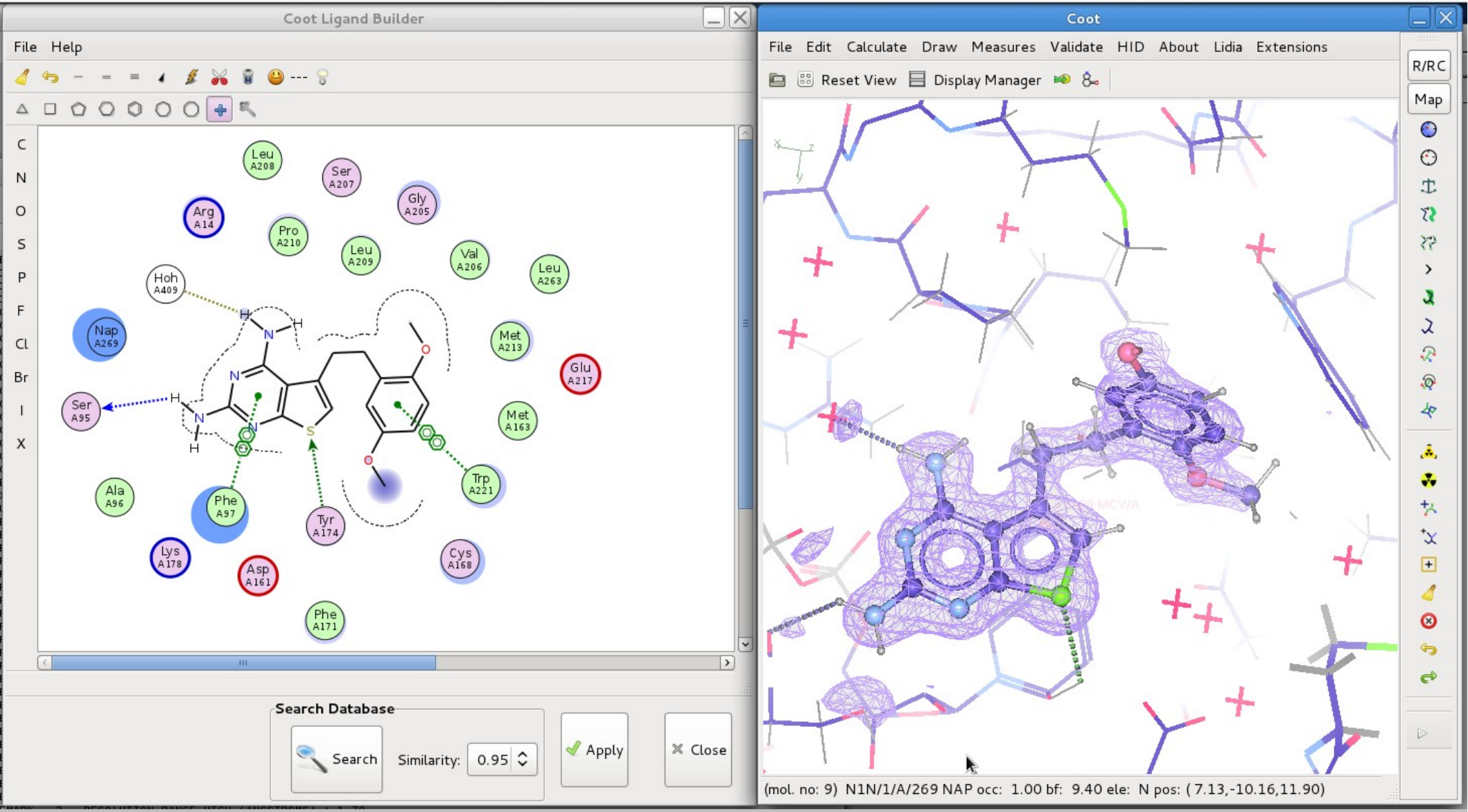



File Help
C
N
O
S
P
F
Cl
Br
I
X
Ile A305
Phe A433
Val A313
Asp A463
Gly A382
Lys A328
Glu A346
Cys A379
Hoh A614
Ile A374
Asn A380
Tyr A378
Ala A326
Asn A376
Glu A377
Val A360
Search Database
Search
Similarity: 0.75
Apply
Close

Layout Examples



3uic


Coot Ligand Builder
File Help
C
N
O
S
P
F
Cl
Br
I
X
Leu A208
Ser A207
Gly A205
Arg A14
Pro A210
Leu A209
Val A206
Leu A263
Hoh A409
Nap A269
Met A213
Glu A217
Ser A95
Met A163
Trp A221
Ala A96
Phe A97
Tyr A174
Lys A178
Asp A161
Cys A168
Phe A171
Search Database
Search
Similarity: 0.95
Apply
Close
Coot
File Edit Calculate Draw Measures Validate HID About Lidia Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 9) N1N/1/A/269 NAP occ: 1.00 bf: 9.40 ele: N pos: (7.13,-10.16,11.90)

Acknowledgements

- Kevin Cowtan
- Bernhard Lohkamp

- Libraries, Dictionaries
  - Alexei Vagin, Garib Murshudov
  - Eugene Krissinel
  - Greg Landrum

- Funding:
  - BBSRC & CCP4