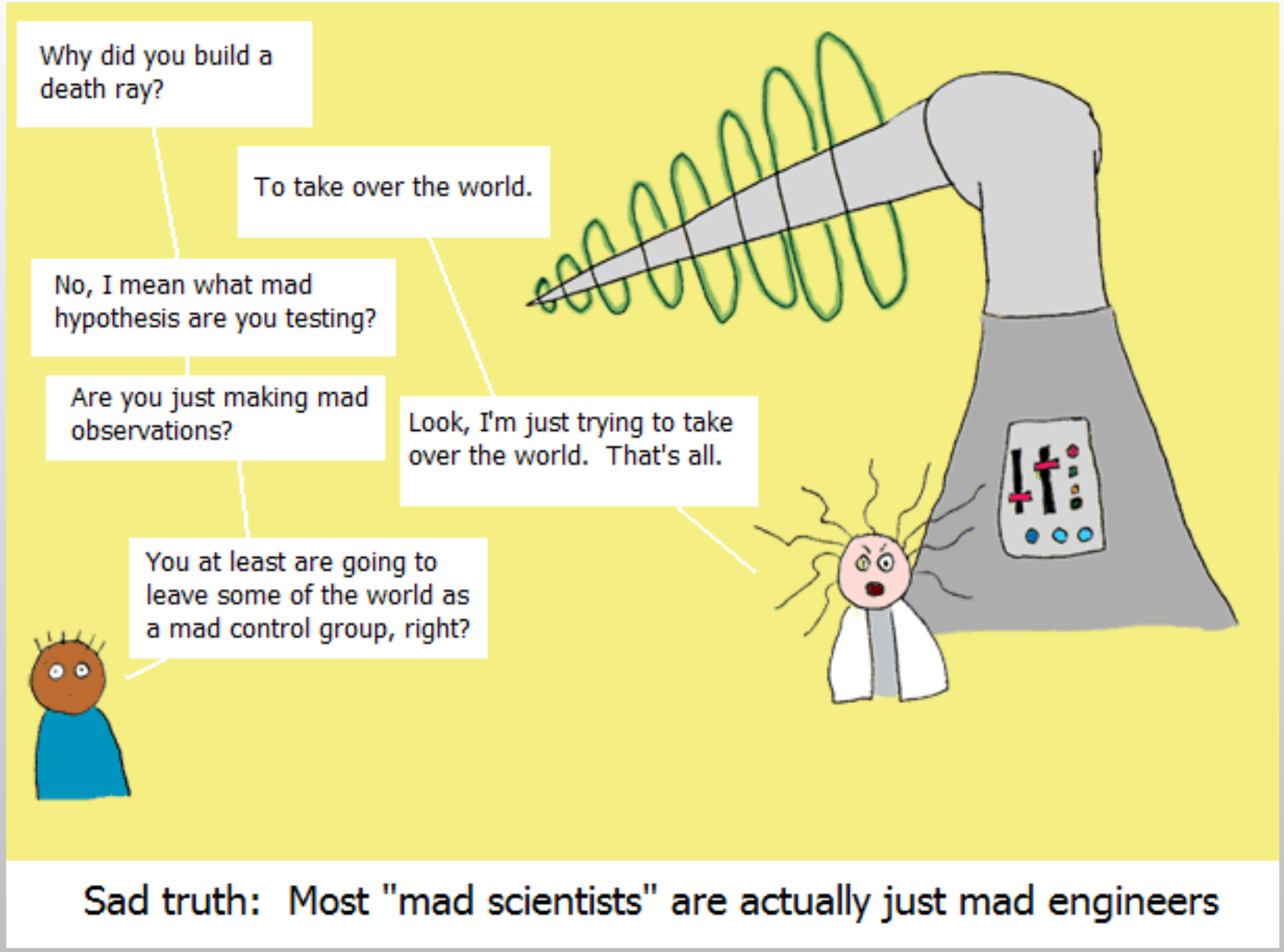
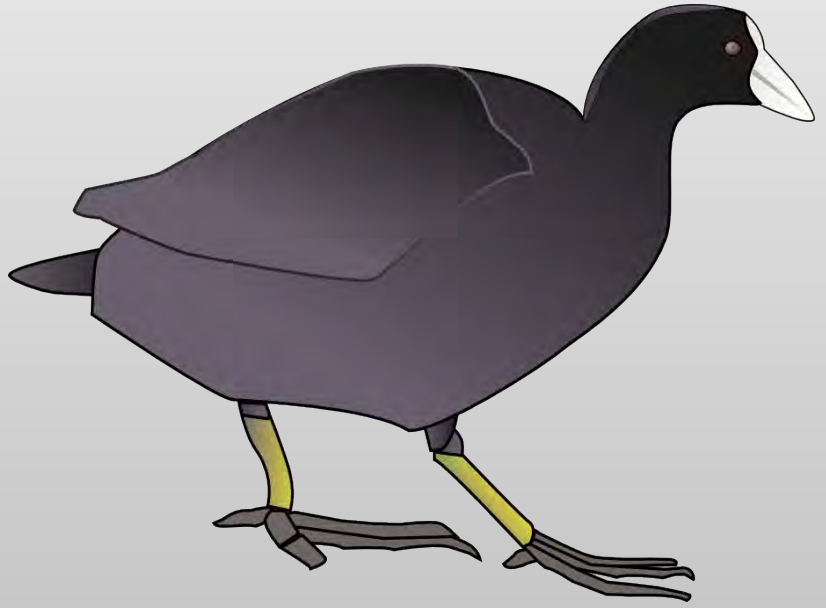


Sad truth: Most "mad scientists" are actually just mad engineers

cowbirdsinlove

Using *Coot* Tools for Protein-Ligand Analysis



Paul Emsley

MRC Laboratory of Molecular Biology

July 2013

Using *Coot* Tools for Protein-Ligand Analysis

- Automated Scoring of Protein-Ligand Complexes
- Tools for editing, design of ligands
- Tools for Presentation and Navigation
- (Maybe) Fitting (N-linked) Carbohydrate

Coot Tools for Protein-Ligand Complexes

- Aim is the provide tools that analyse the ligand complex under investigation
- ... combining existing software with new tools
- ... to give it the “Green Lights” (if appropriate)
- ... judge the rank as compared to other protein-ligand complexes

Scoring Protein-Ligand Complexes

- Score all PDB protein-ligand complexes
  - The (first) biggest complete Het-group
  - No covalent link to protein
  - No alt confs
  - Het-groups with more than 6 atoms
    - Glycerol included
  - Only use accession codes with (readable) data
    - 2007-2012
  - Only those het-groups for which I could construct a molecule with sane chemistry and an MDL molfile using Refmac restraints dictionary

Scoring Protein-Ligand Complexes

- Score – 3 Metrics:
  - Correlation of maps: omit *vs*. calculated
    - around the ligand
  - Clash-score
    - *c.f.* Molprobity tool
  - Mogul distortion
    - z-worst

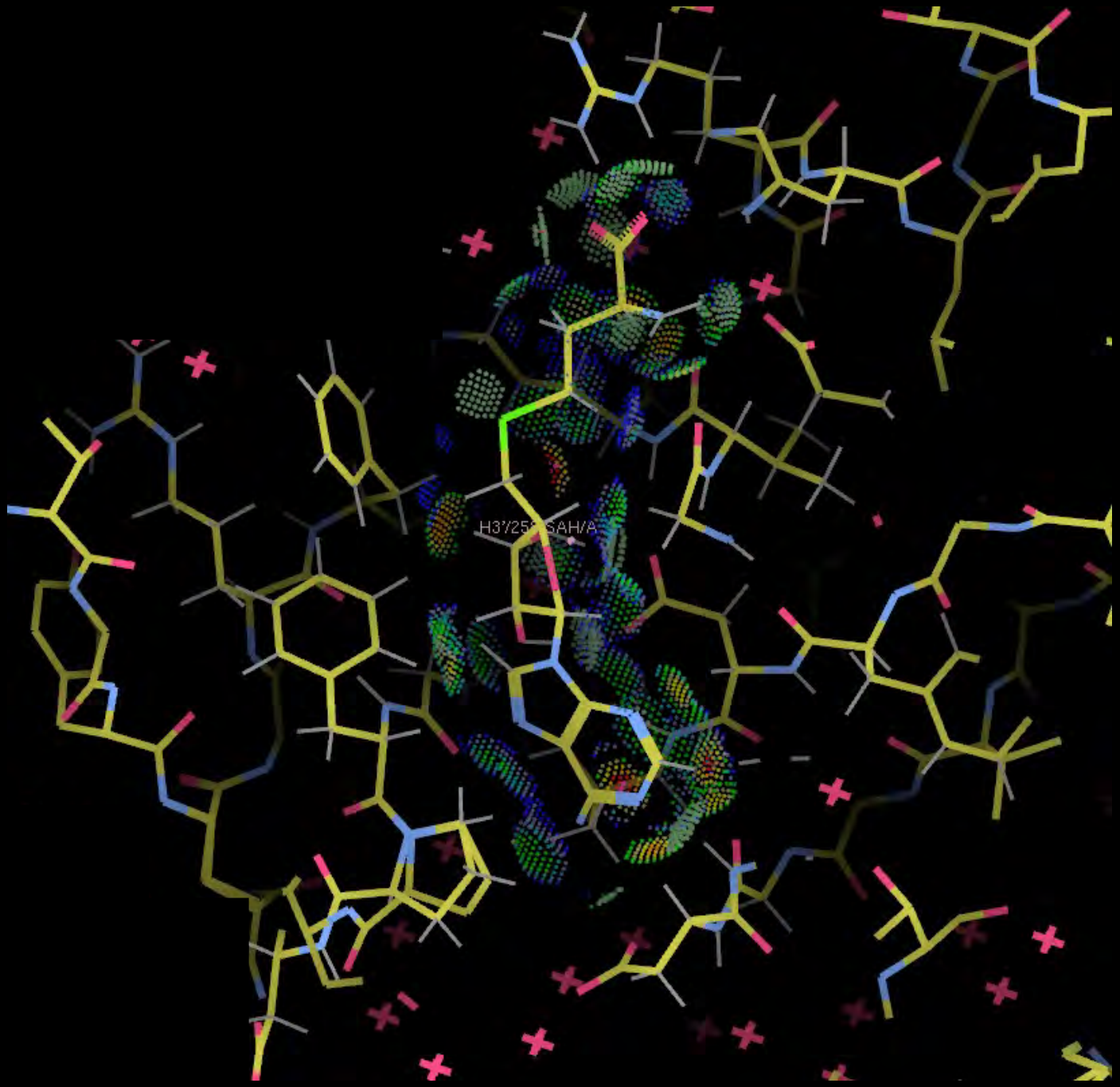
Density Correlation Metric

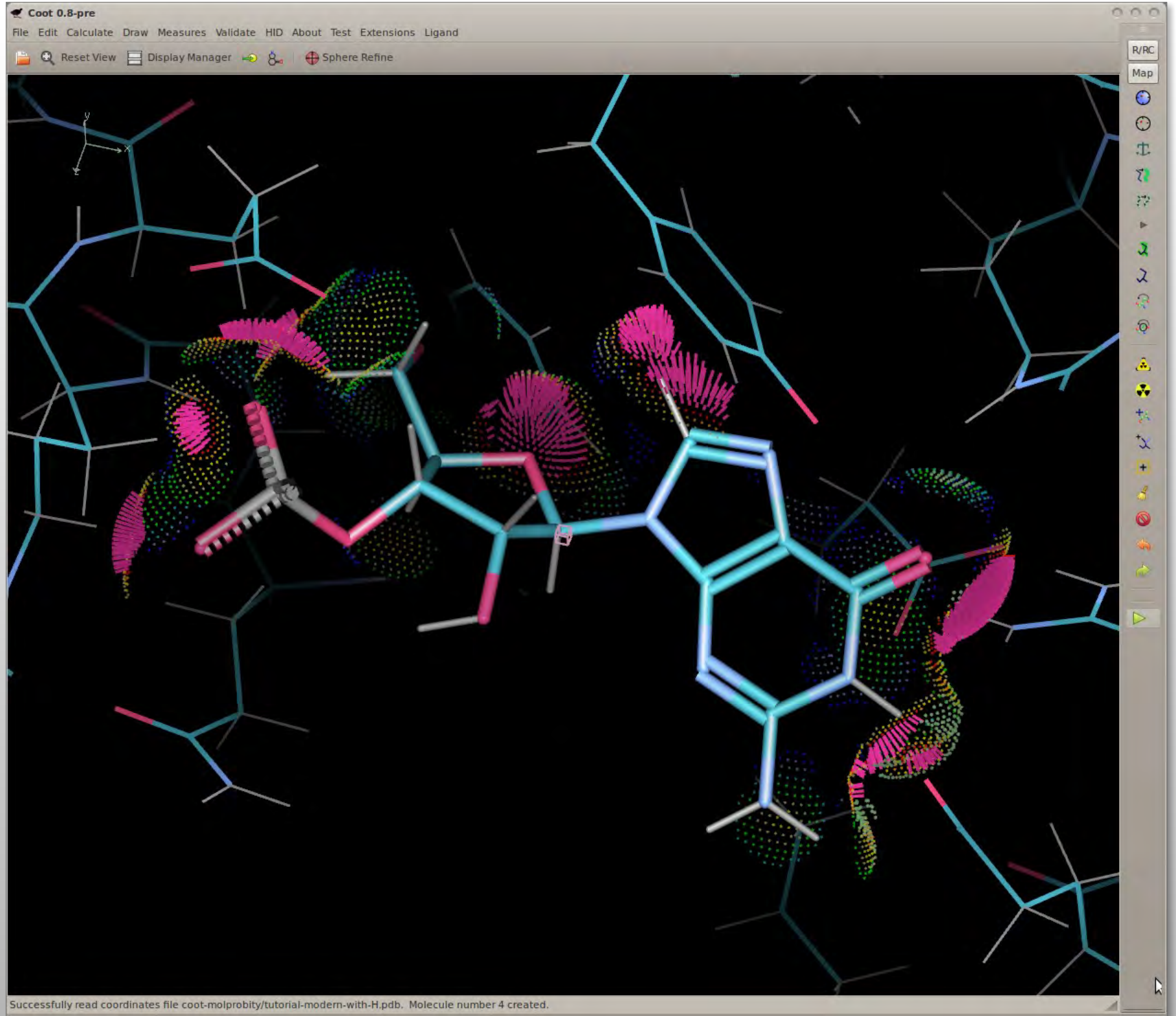
- Identify ligand of interest
- Construct an MDL molfile?
- Remove ligand
- Run Refmac to calculate structure factors
  - omit map
- Identify correlation coefficient
  - omit map *vs.* calc map
  - in the region of the ligand

Probe Score Metric

- Using Reduce and Probe
  - Richardsons and co-workers
- Consider only protein-ligand interactions
- Count the number of “bad overlap” atom pairs

Probe Contacts




Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Test Extensions Ligand
Reset View
Display Manager
Sphere Refine
R/RC
Map
Successfully read coordinates file coot-molprobity/tutorial-modern-with-H.pdb. Molecule number 4 created.

Probe Score Metric

- Using Reduce and Probe
  - Richardsons and co-workers
- Consider only protein-ligand interactions
- Count the number of “bad overlap” atom pairs

Probe Score Metric

- Using Reduce and Probe
  - Richardsons and co-workers
- Consider only protein-ligand interactions
- Count the number of “bad overlap” atom pairs

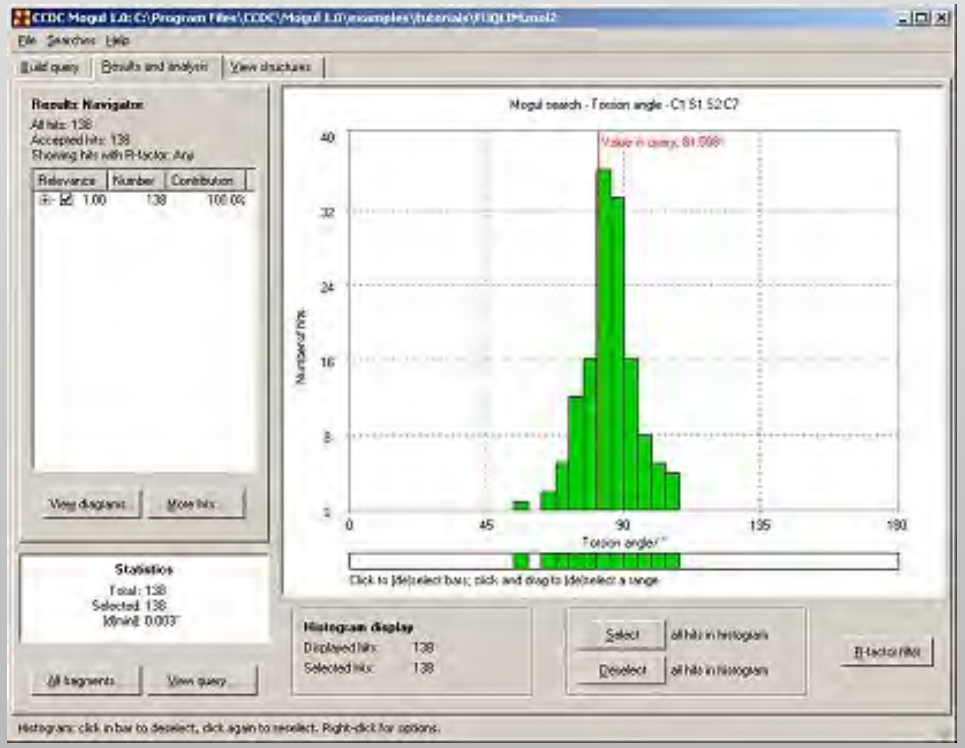
Mogul Score Metric

- Use CSD Mogul
  - MDL query
    - Coordinates from PDB ligand, bond orders from Refmac restraints
  - customized csv output
    - parsed and (interactive mode) represented in Coot

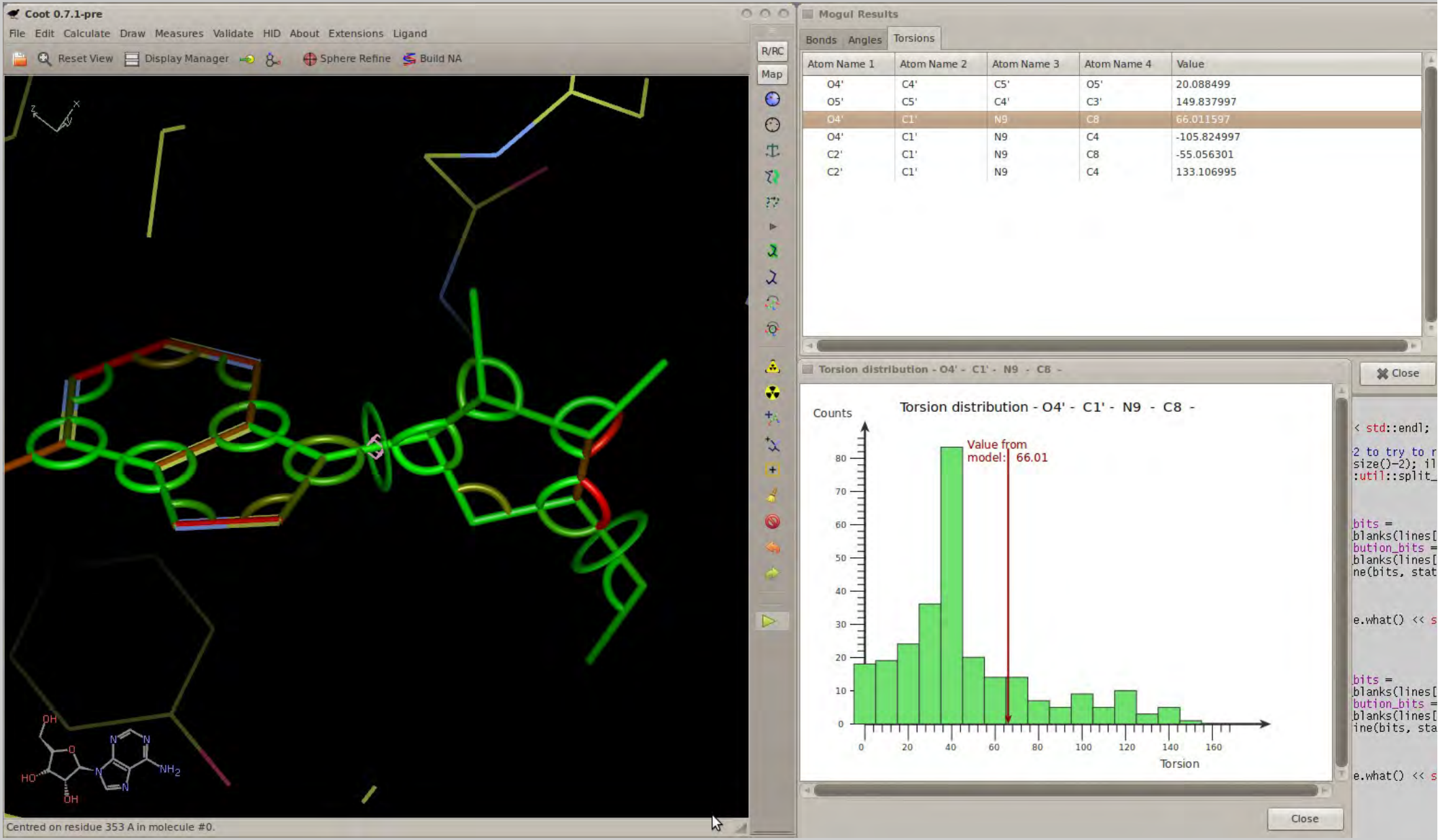
CSD Mogul

Knowledge-base of geometric parameters based on the CSD

- Can be run as a “batch job”
- Mean, median, quartiles, Z-scores.
- Query constructed as an MDL file using bond orders from the Refmac monomer library
- Histograms



Mogul Results Representation

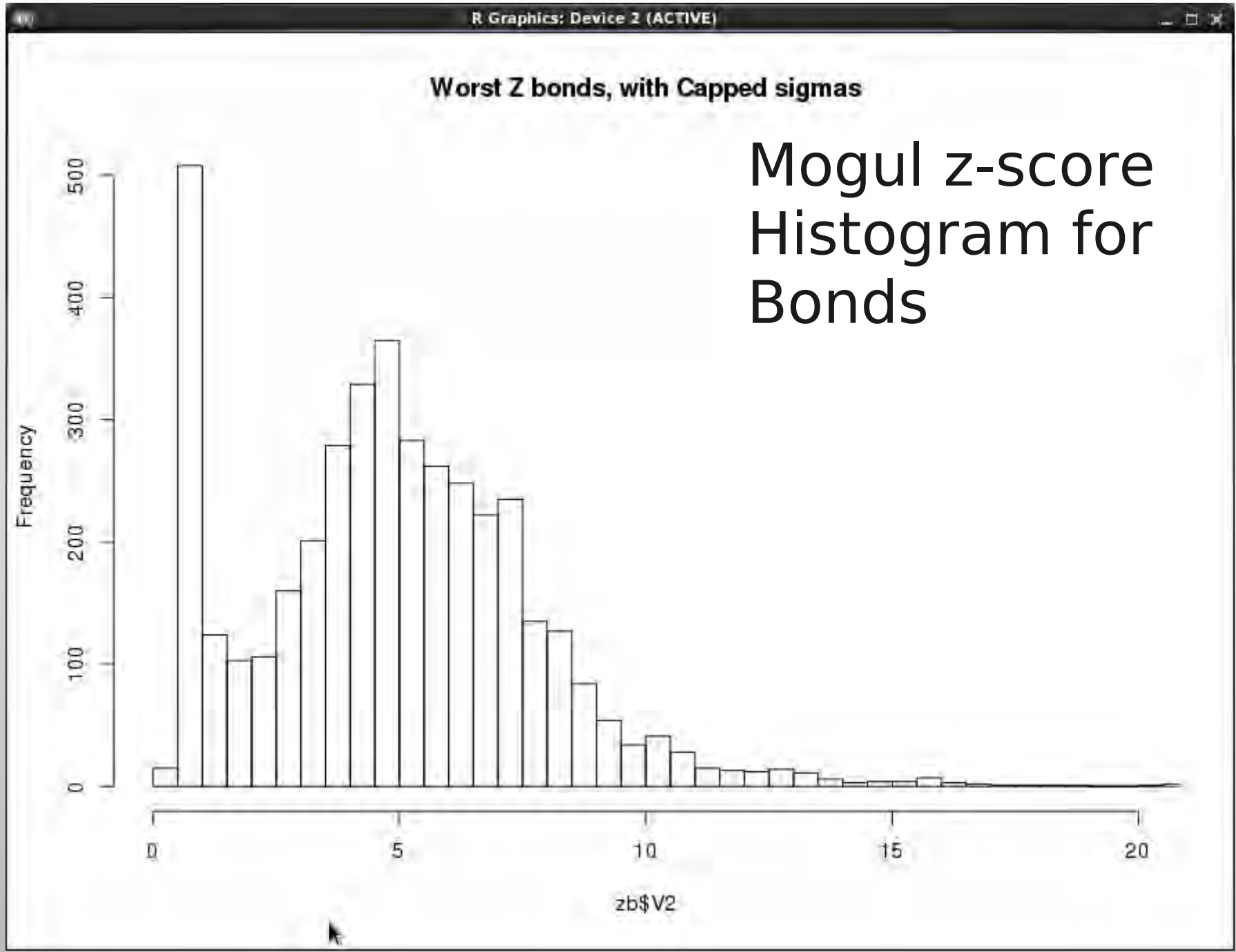


Mogul Score Metric

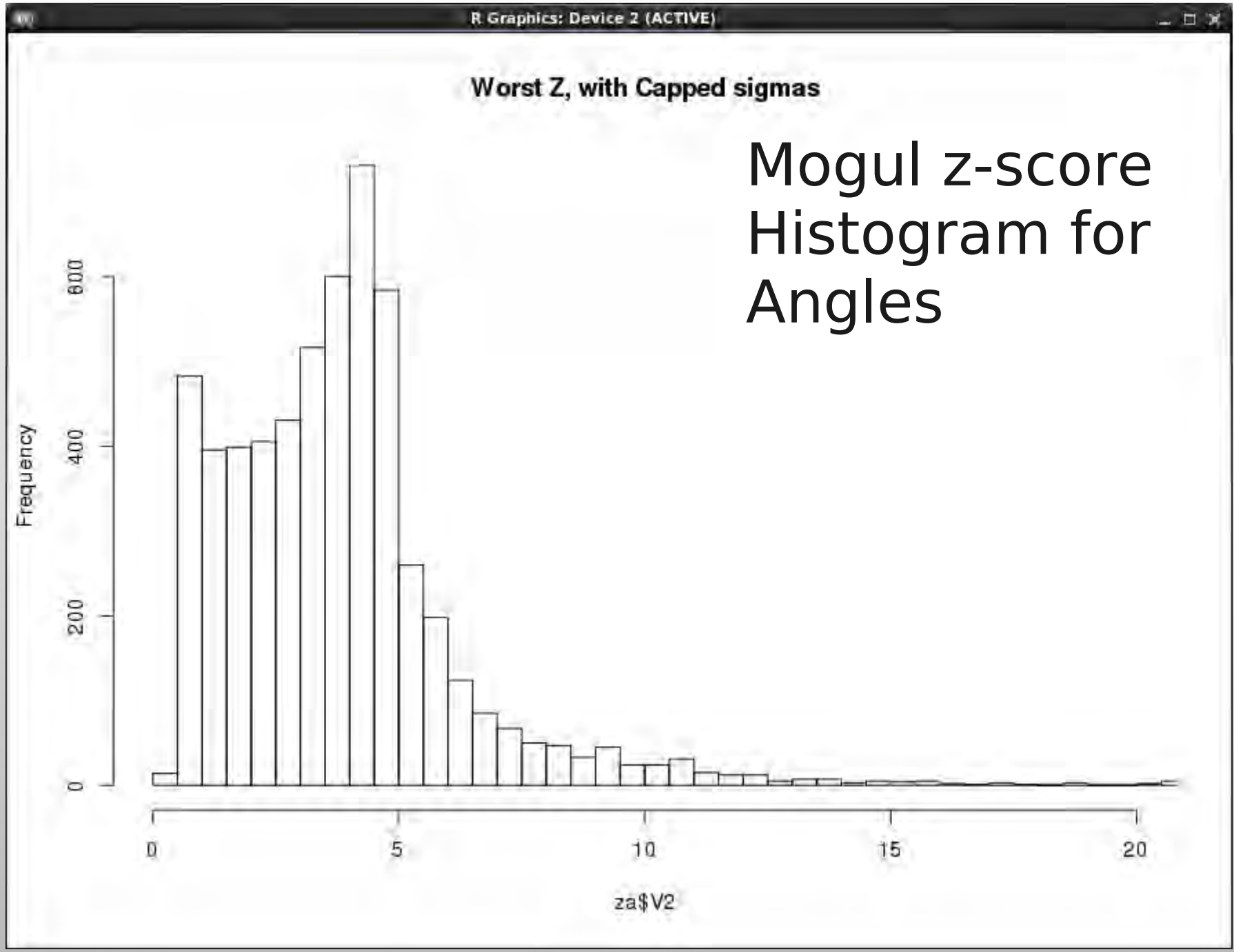
- How do you score a ligand with distorted geometry?
- Average badness?
  - A highly distorted bond in an otherwise adequate large ligand will be hidden
- Worst z-score outlier for bonds and angles
  - Bad chemistry is bad chemistry no matter how big the ligand
- Using modified standard deviations
  - *i.e.* not simply those from describing the distribution of the data from the crystal structures
  - Lower-bounds caps (every bond and angle checked)

Mogul-Based Ligand Validation

- Mogul plugin in Coot
  - Run mogul in “non-interactive” mode
  - graphical display of results
  - Update restraints (target and esds for bonds and angles)
  - CSD data not so great for plane, chiral and torsion restraints
    - remain unexploited for automated validation to date



Mogul z-score Histogram for Bonds



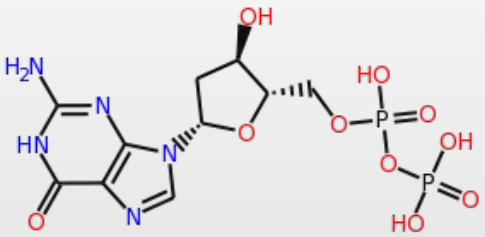
Mogul z-score Histogram for Angles

Let's Rank Comp-ids

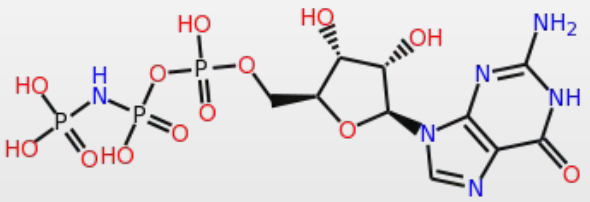
- By Average Mogul Z score
  - to identify the most distorted group types in the PDB

The Most Distorted Groups in the PDB

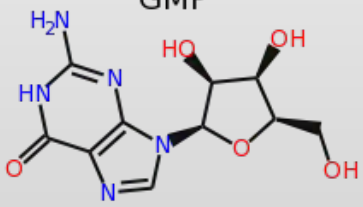
DGI



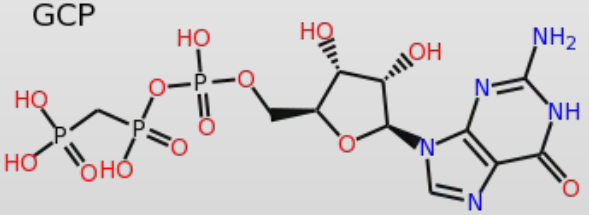
GNP



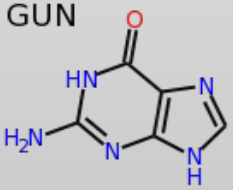
GMP



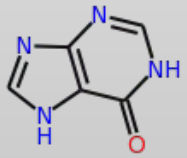
GCP



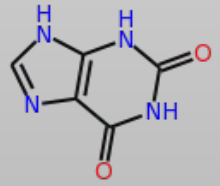
GUN



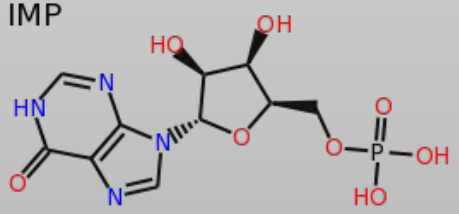
HPA



XAN



IMP



(with more than one structure)

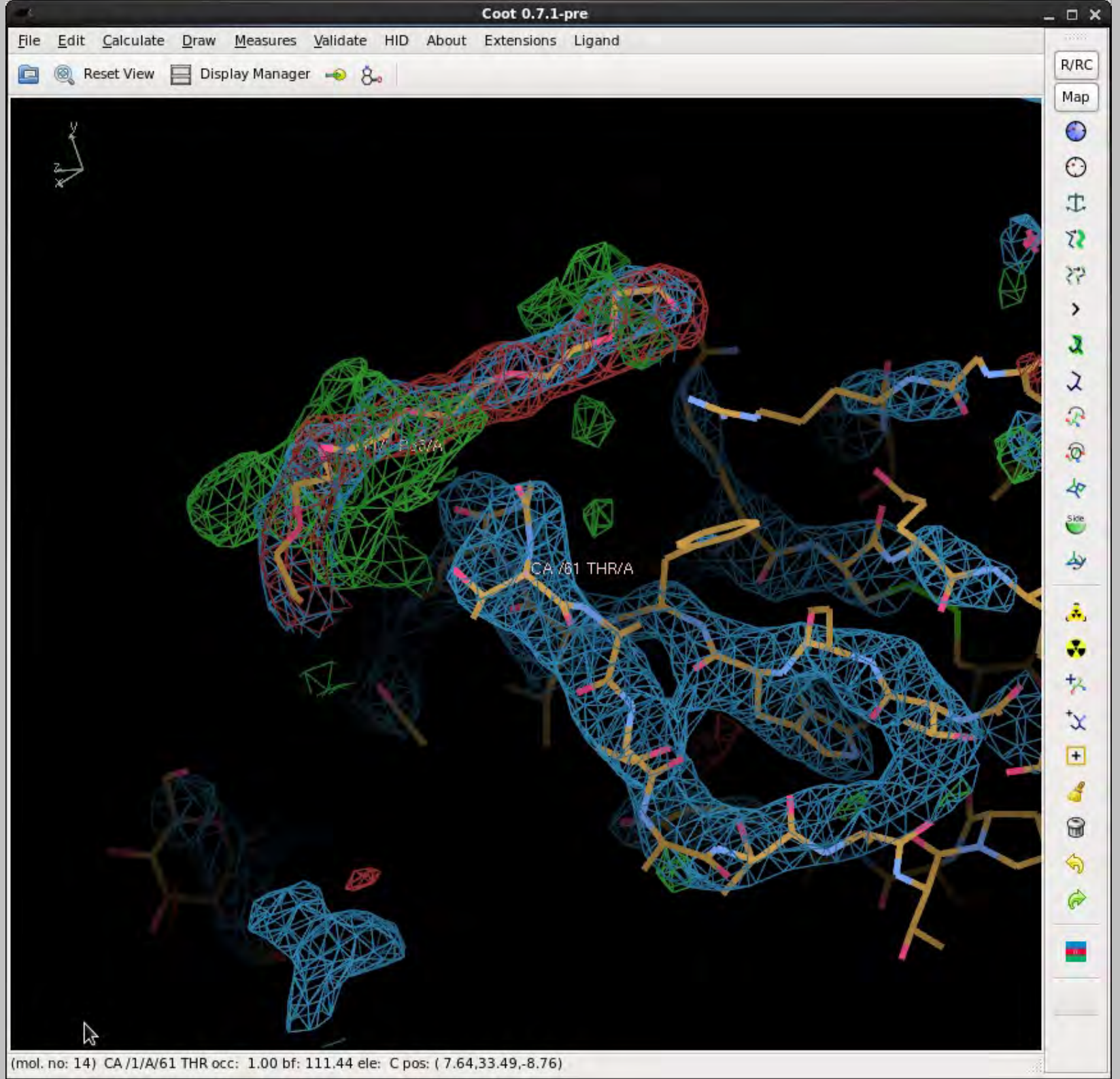
Additional Criterion:

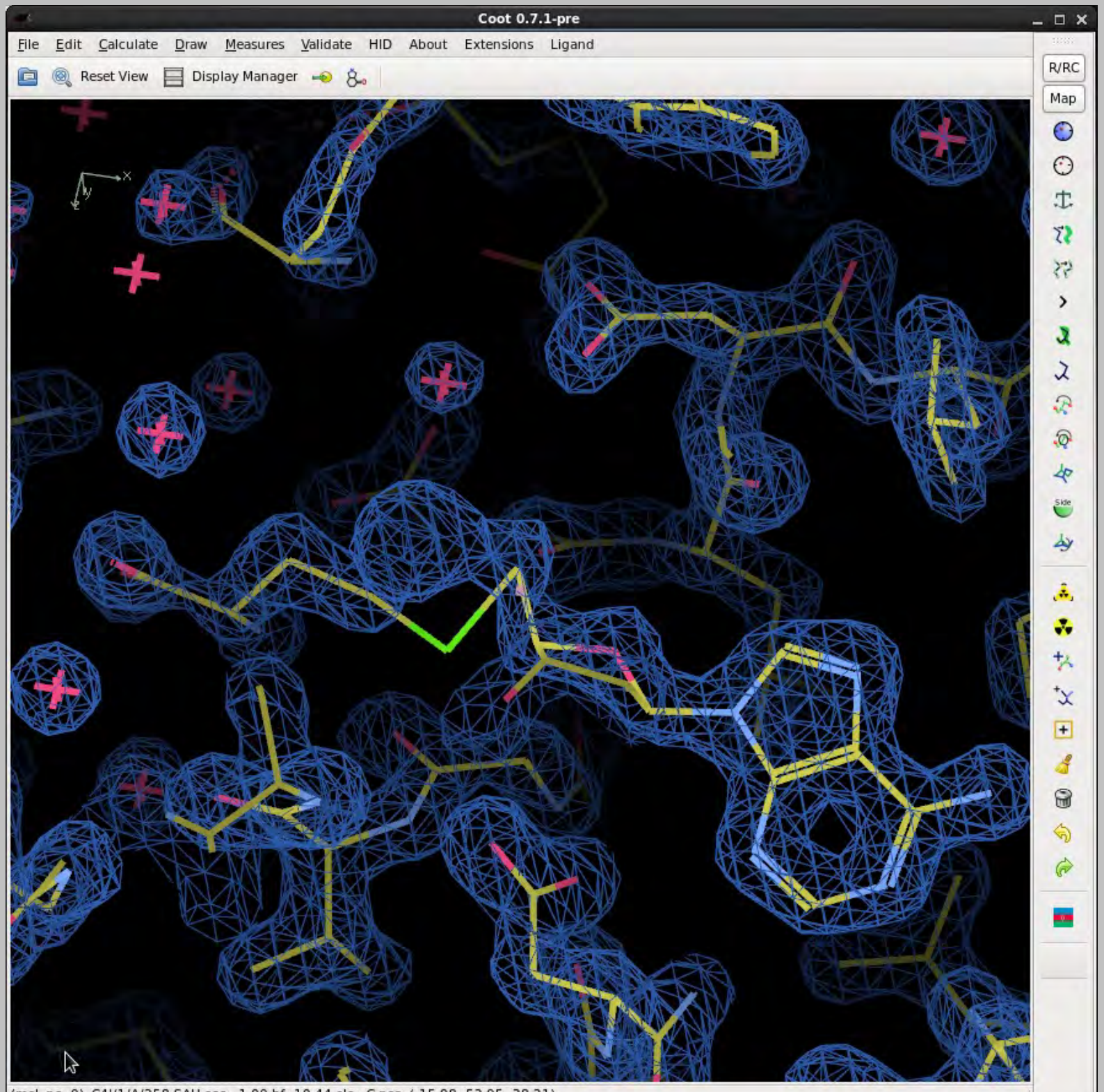
- Are the temperature factors of the ligand atoms drawn from the same distribution as the surrounding atoms?
  - Use Kolmogorov-Smirnov distribution test
  - Not build in to over-all score
    - But interesting outliers...

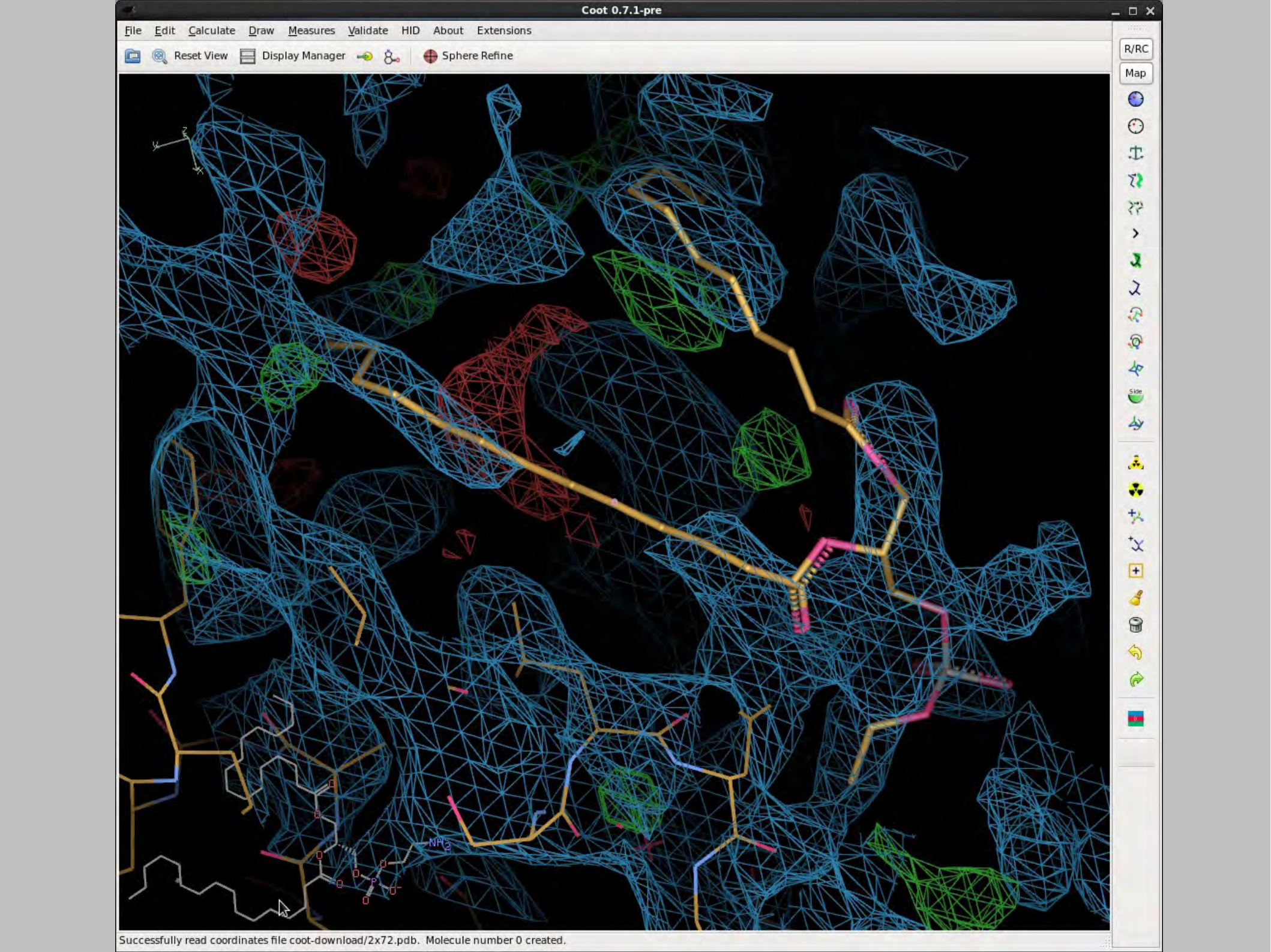
Combining Scores

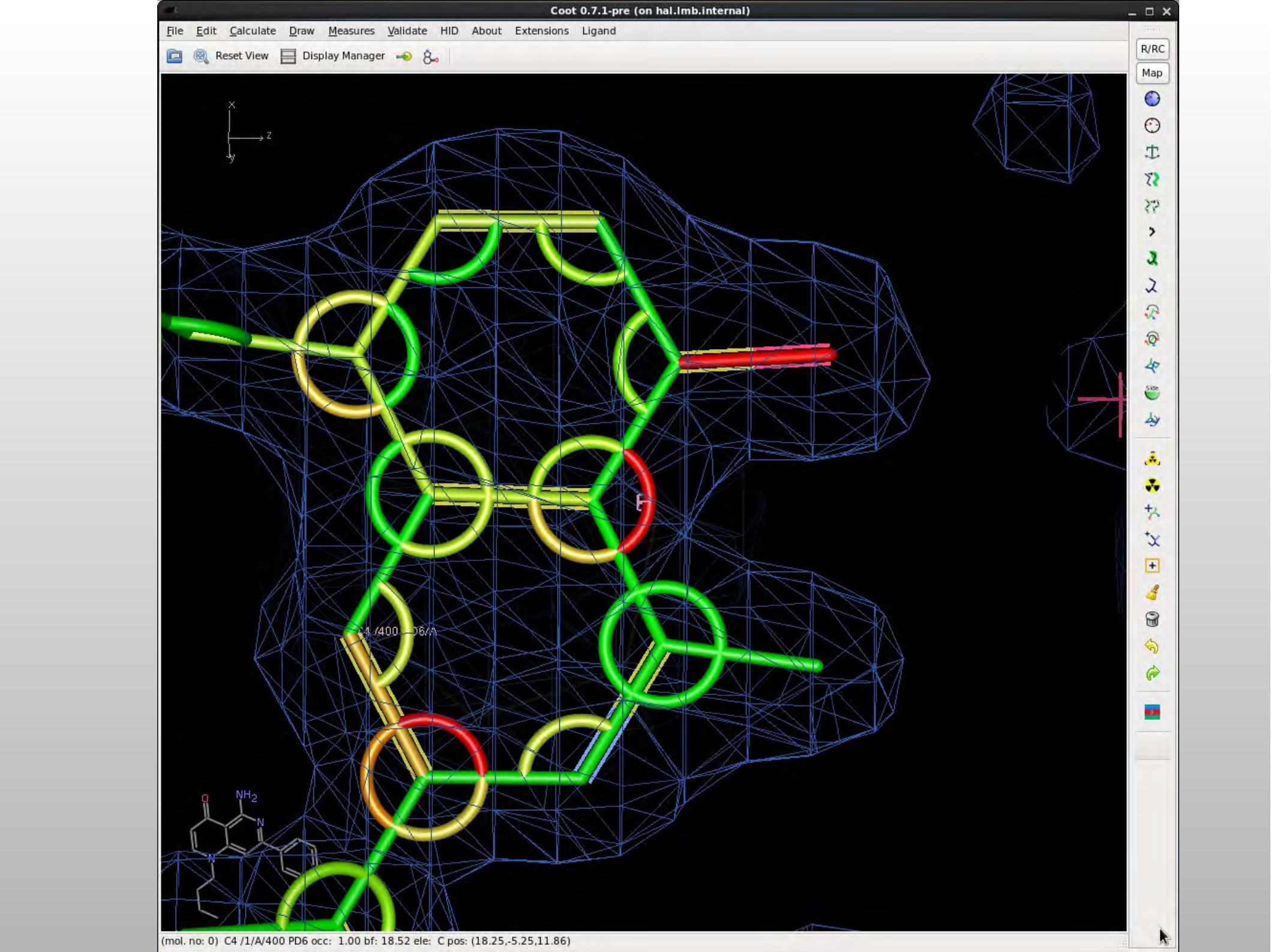
- Ranking the density correlation, Mogul and bump scores gives us individual ranks:
  - R_{corr}, R_{mogul} , R_{bump}
- Combined into total score:
  - $S_T = R_{corr}^2 + R_{mogul}^2 + R_{bump}^2$
  - *Coot* LIgand Toolkit Score
- Which can then be ranked...

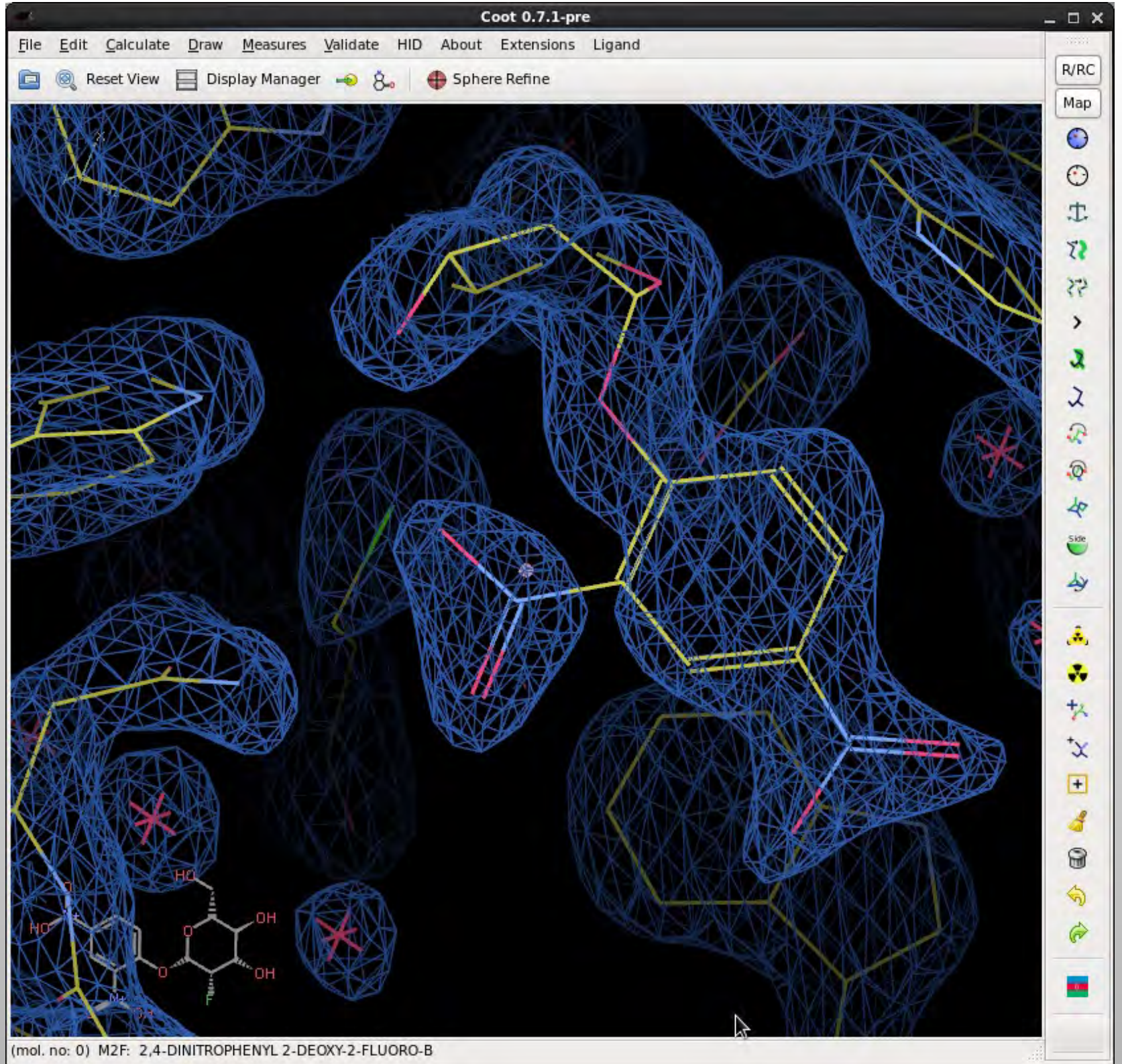
A Gallery of Outliers

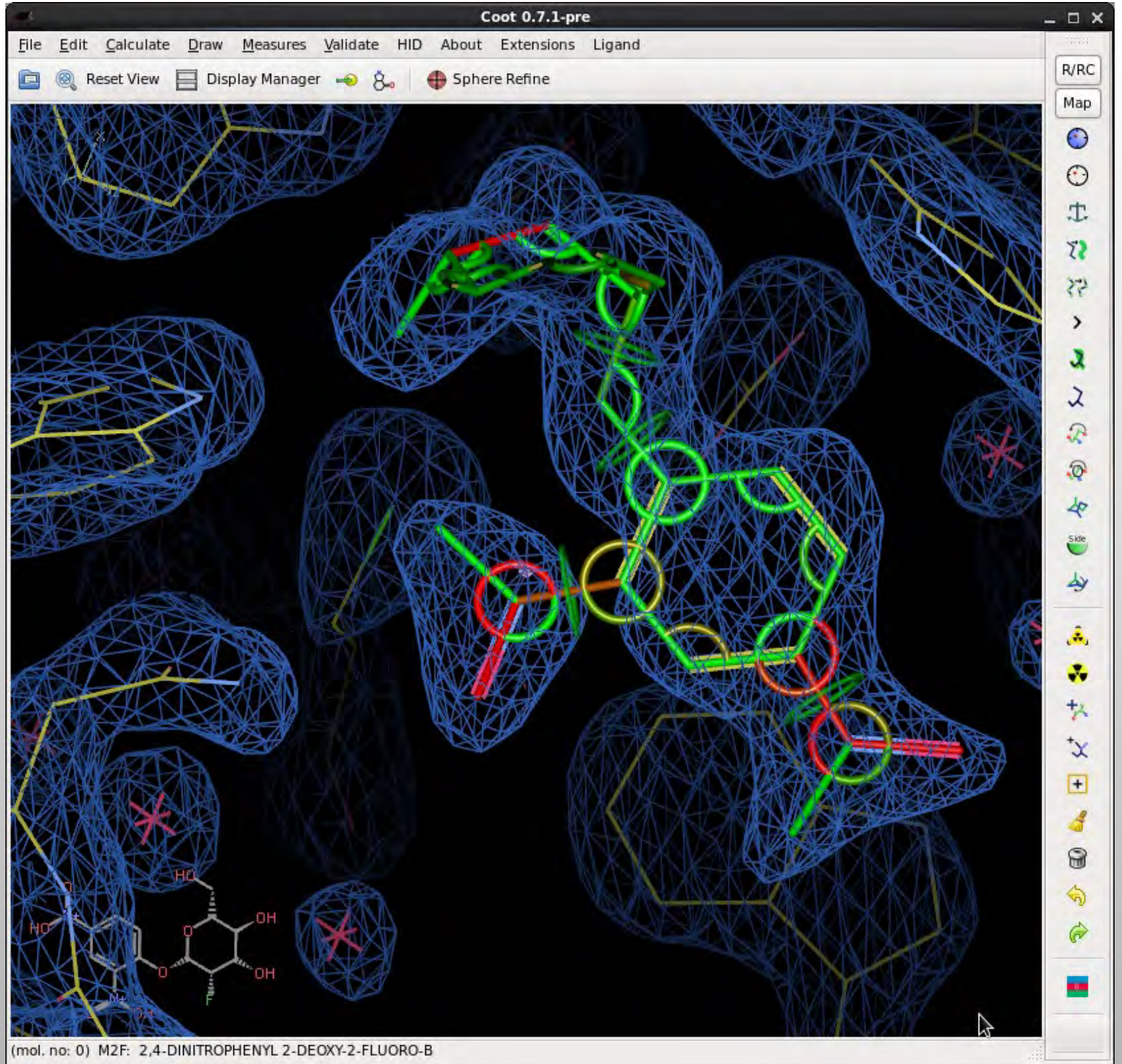

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
R/RC
Map
CA /61 THR/A
(mol. no: 14) CA /1/A/61 THR occ: 1.00 bf: 111.44 ele: C pos: (7.64,33.49,-8.76)


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

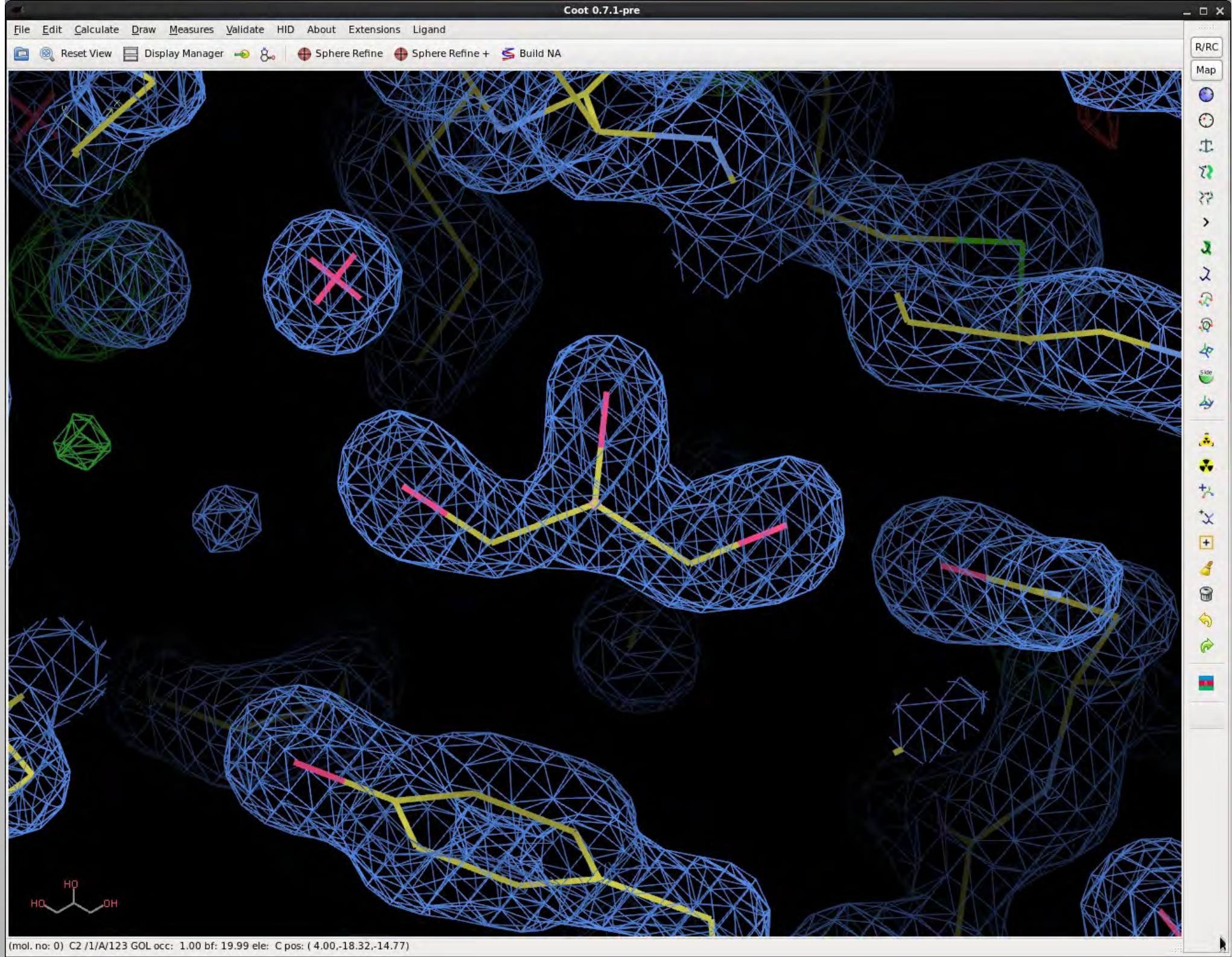

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
Sphere Refine
R/RC
Map
NPP
Successfully read coordinates file coot-download/2x72.pdb. Molecule number 0 created.

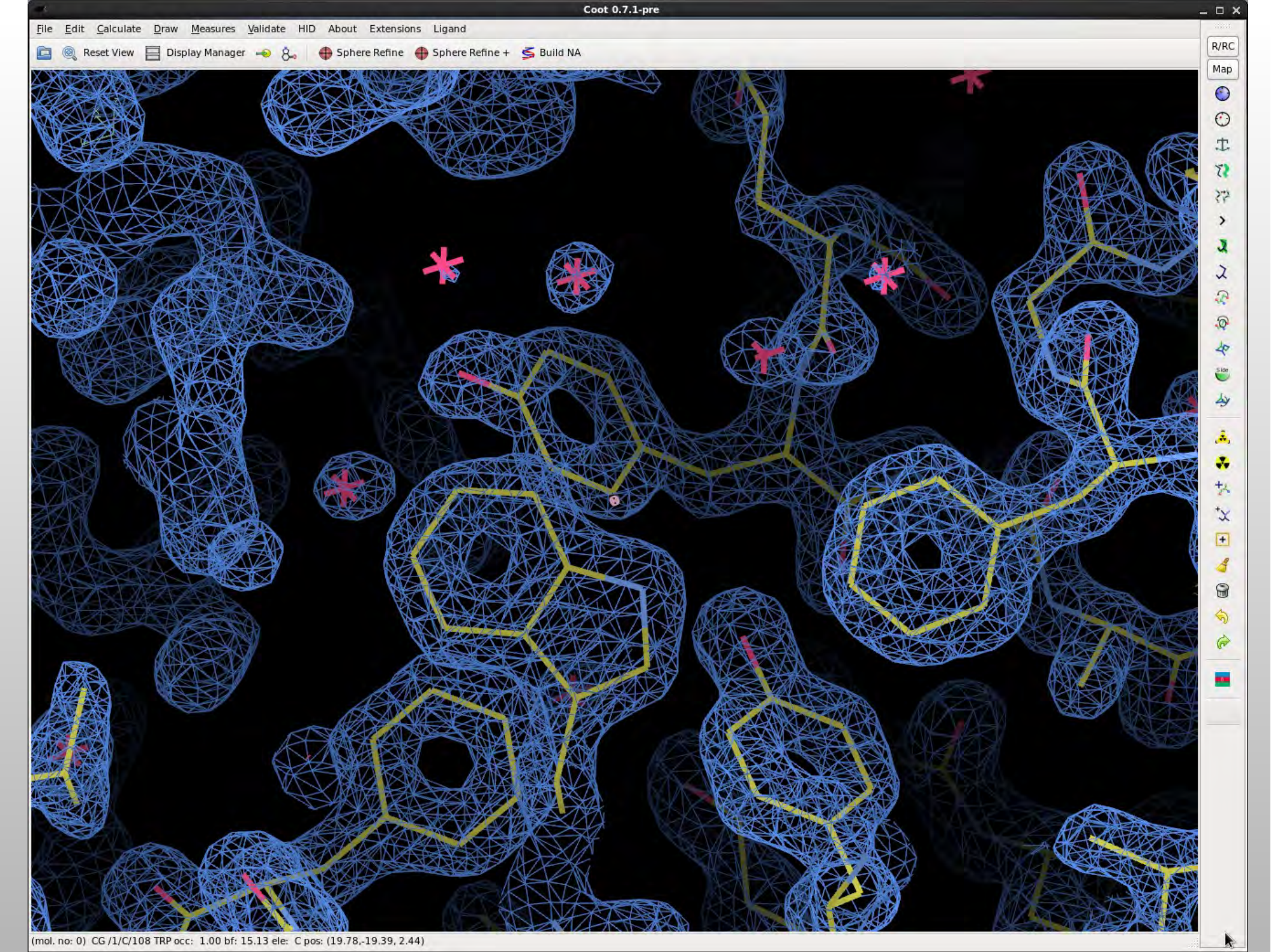

Coot 0.7.1-pre (on hal.lmb.internal)
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
R/RC
Map
NH2
O
N
N
(mol. no: 0) C4 /1/A/400 PD6 occ: 1.00 bf: 18.52 ele: C pos: (18.25,-5.25,11.86)

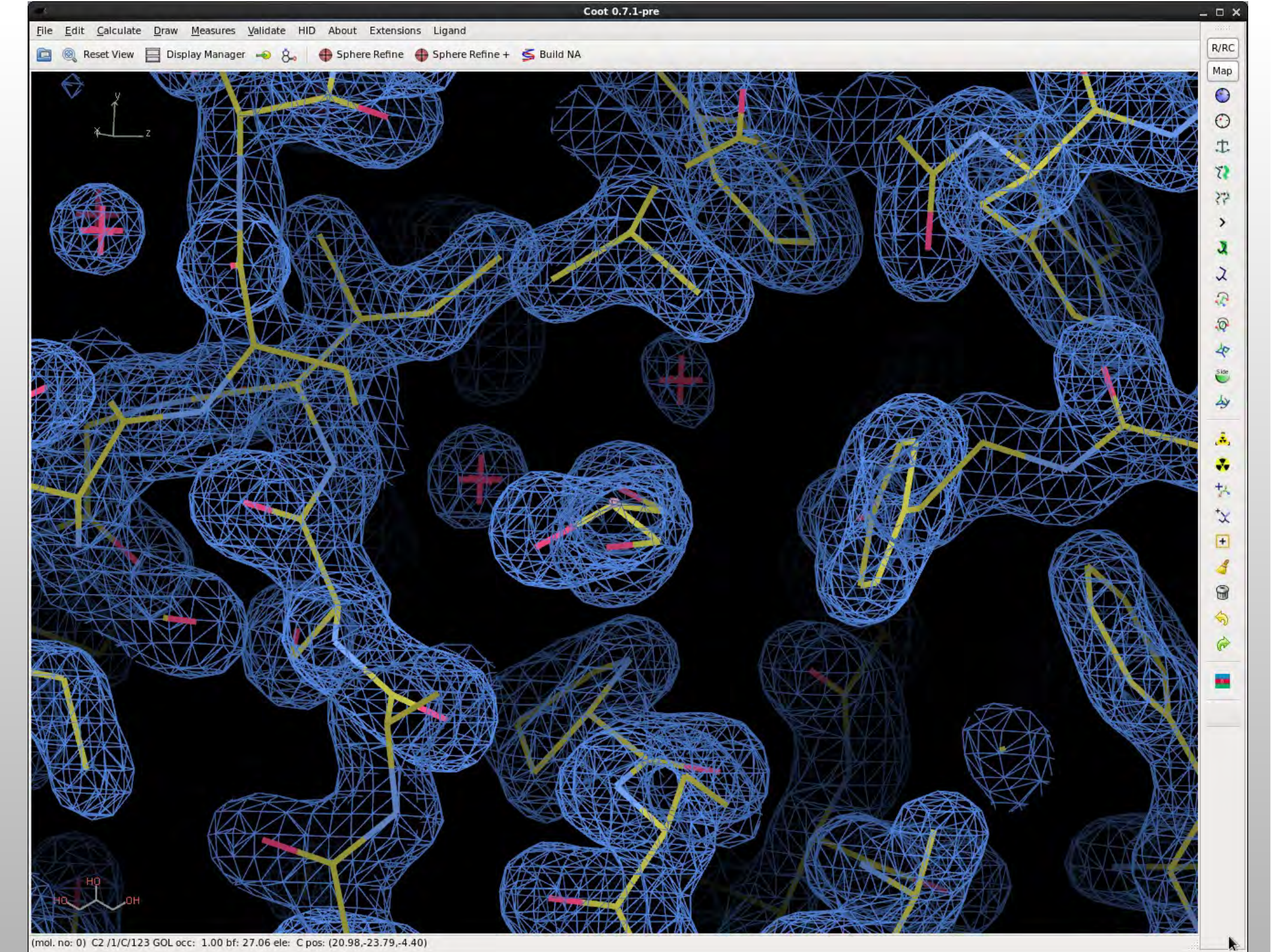

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
Sphere Refine
R/RC
Map
Side
HO
OH
O
OH
HO
O
F
(mol. no: 0) M2F: 2,4-DINITROPHENYL 2-DEOXY-2-FLUORO-B

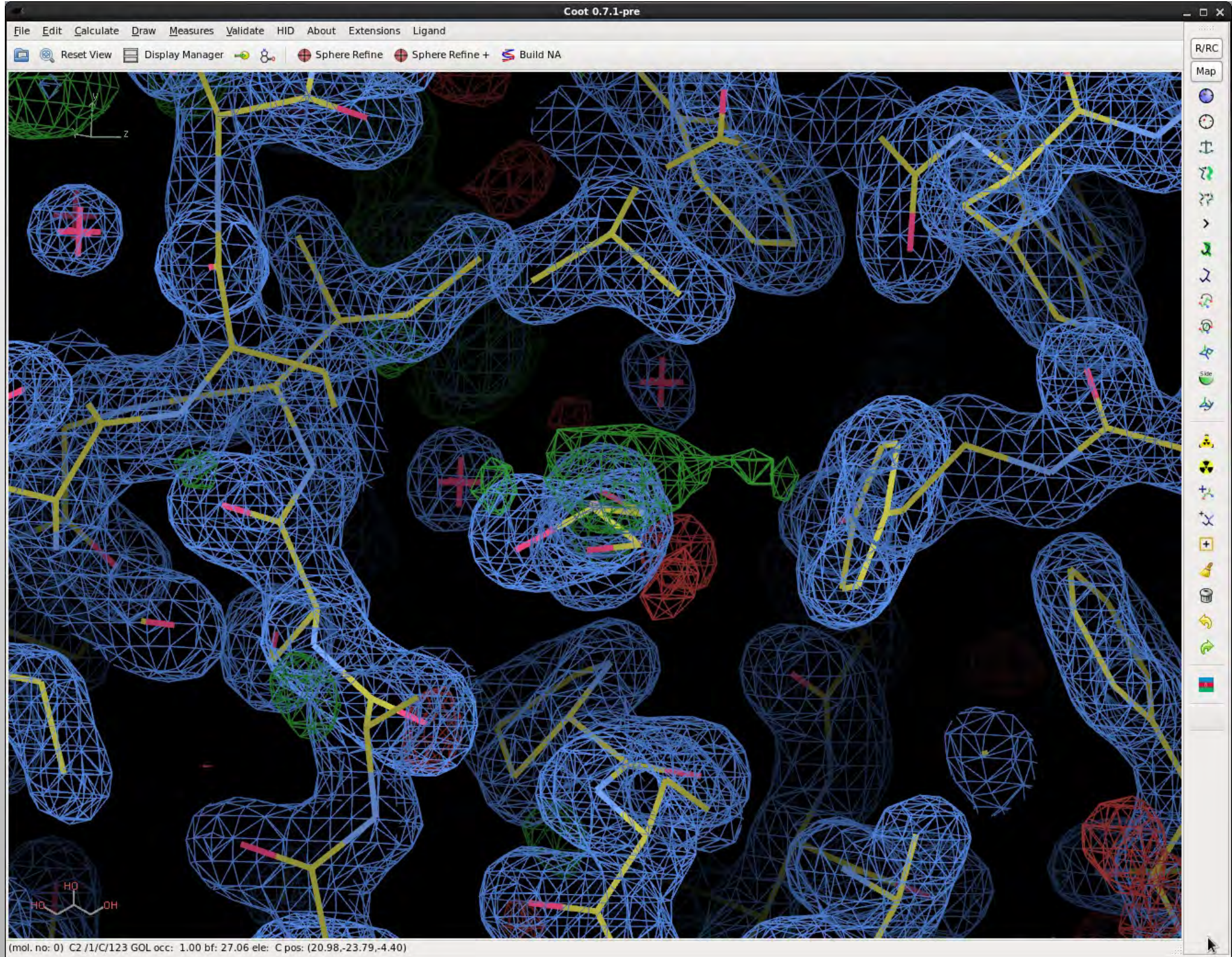

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
Sphere Refine
R/RC
Map
Side
(mol. no: 0) M2F: 2,4-DINITROPHENYL 2-DEOXY-2-FLUORO-B

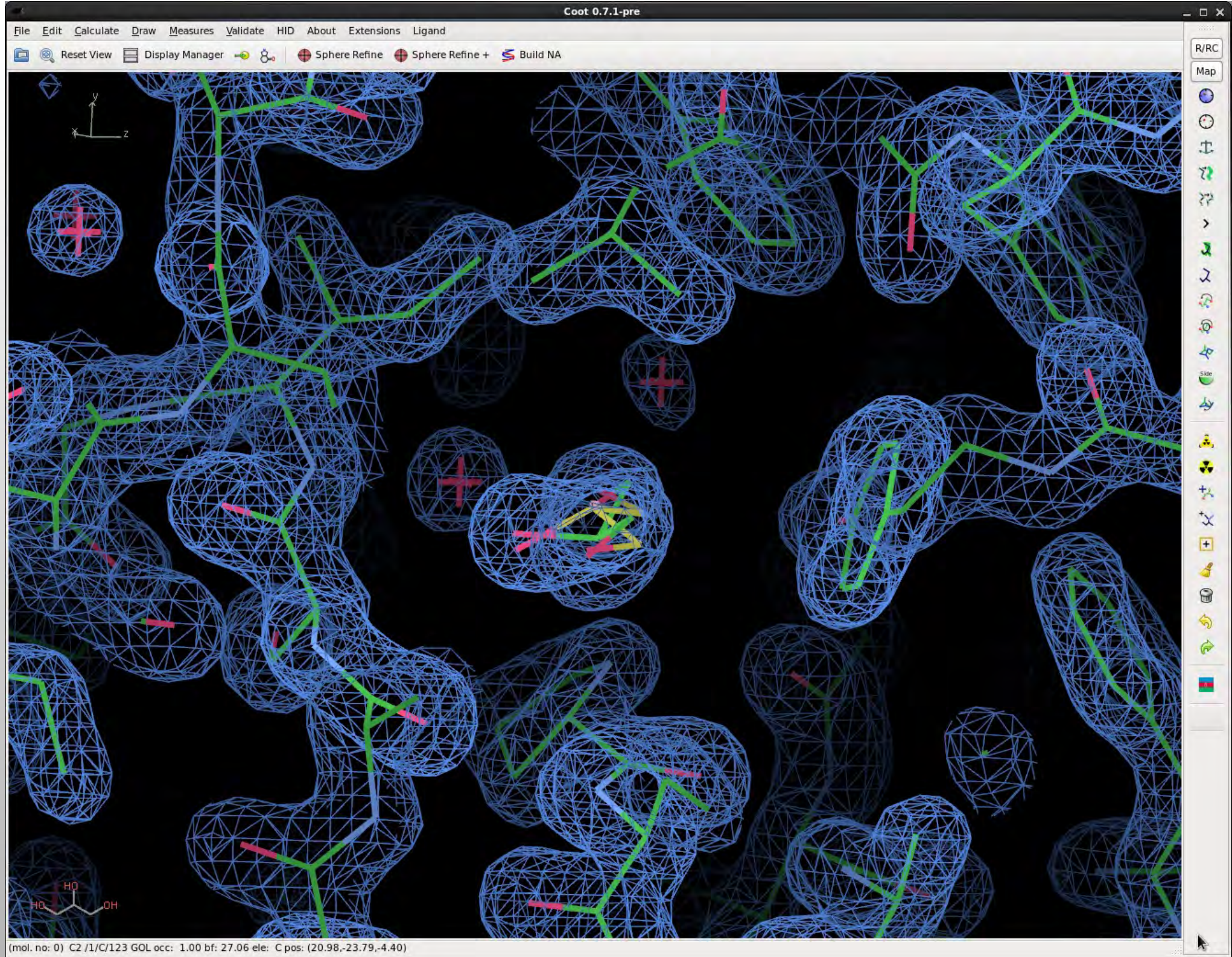
Ligand Ranked #1 of 8470

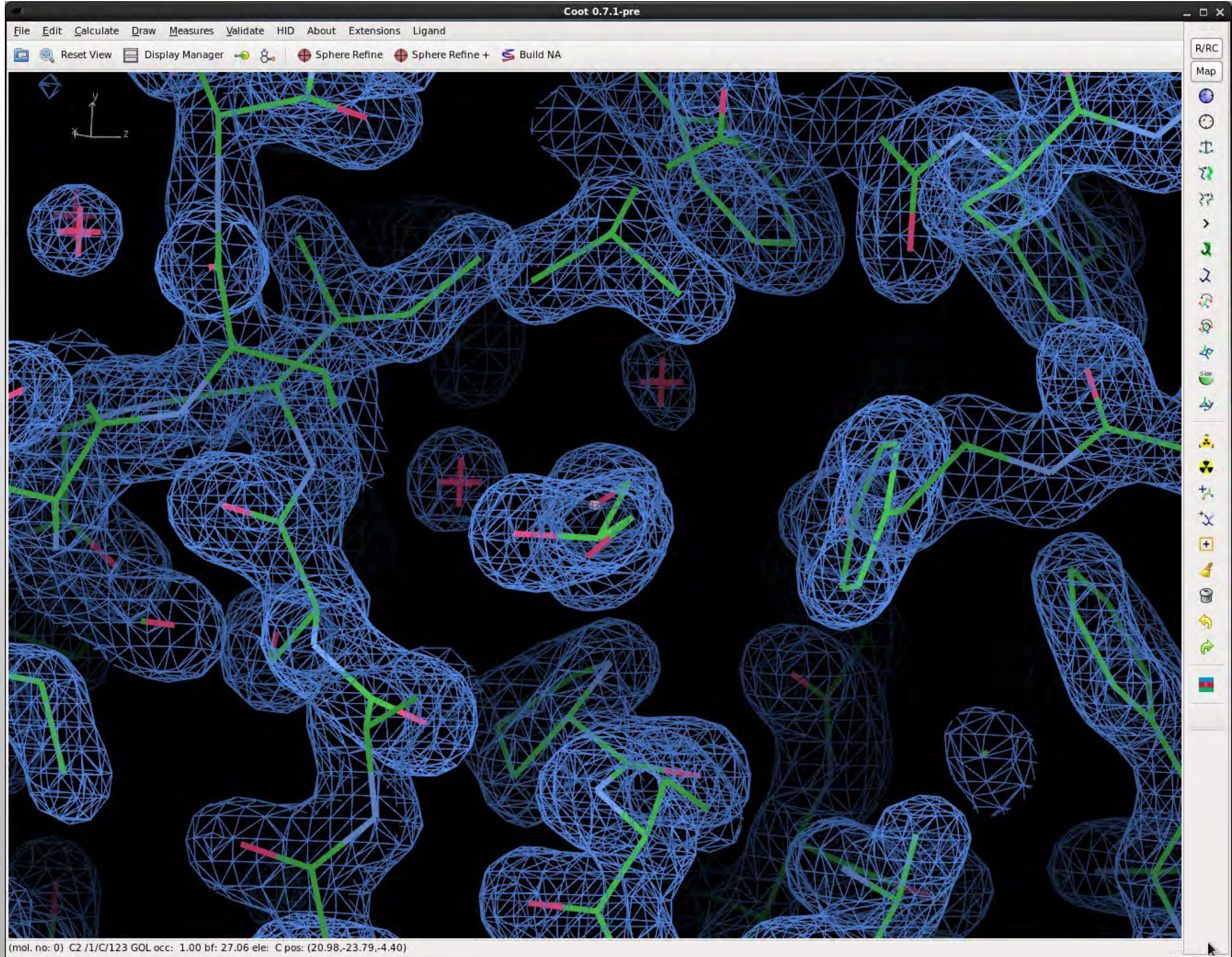

Coot 0.7.1-pre
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
HO
HO
OH
(mol. no: 0) C2 /1/A/123 GOL occ: 1.00 bf: 19.99 ele: C pos: (4.00,-18.32,-14.77)

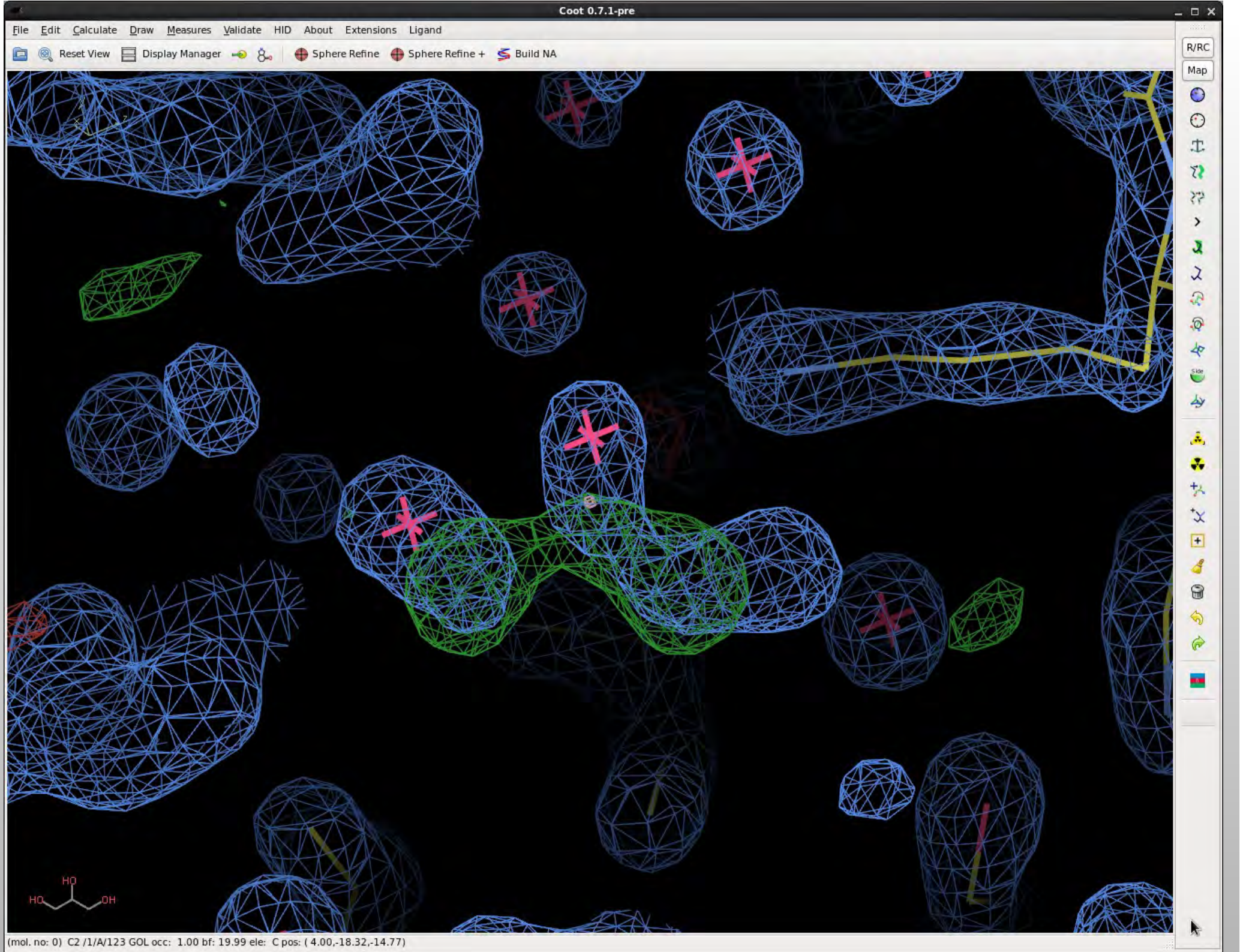

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 0) CG /1/C/108 TRP occ: 1.00 bf: 15.13 ele: C pos: (19.78,-19.39, 2.44)


Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
HO
HO
OH
(mol. no: 0) C2 /1/C/123 GOL occ: 1.00 bf: 27.06 ele: C pos: (20.98,-23.79,-4.40)

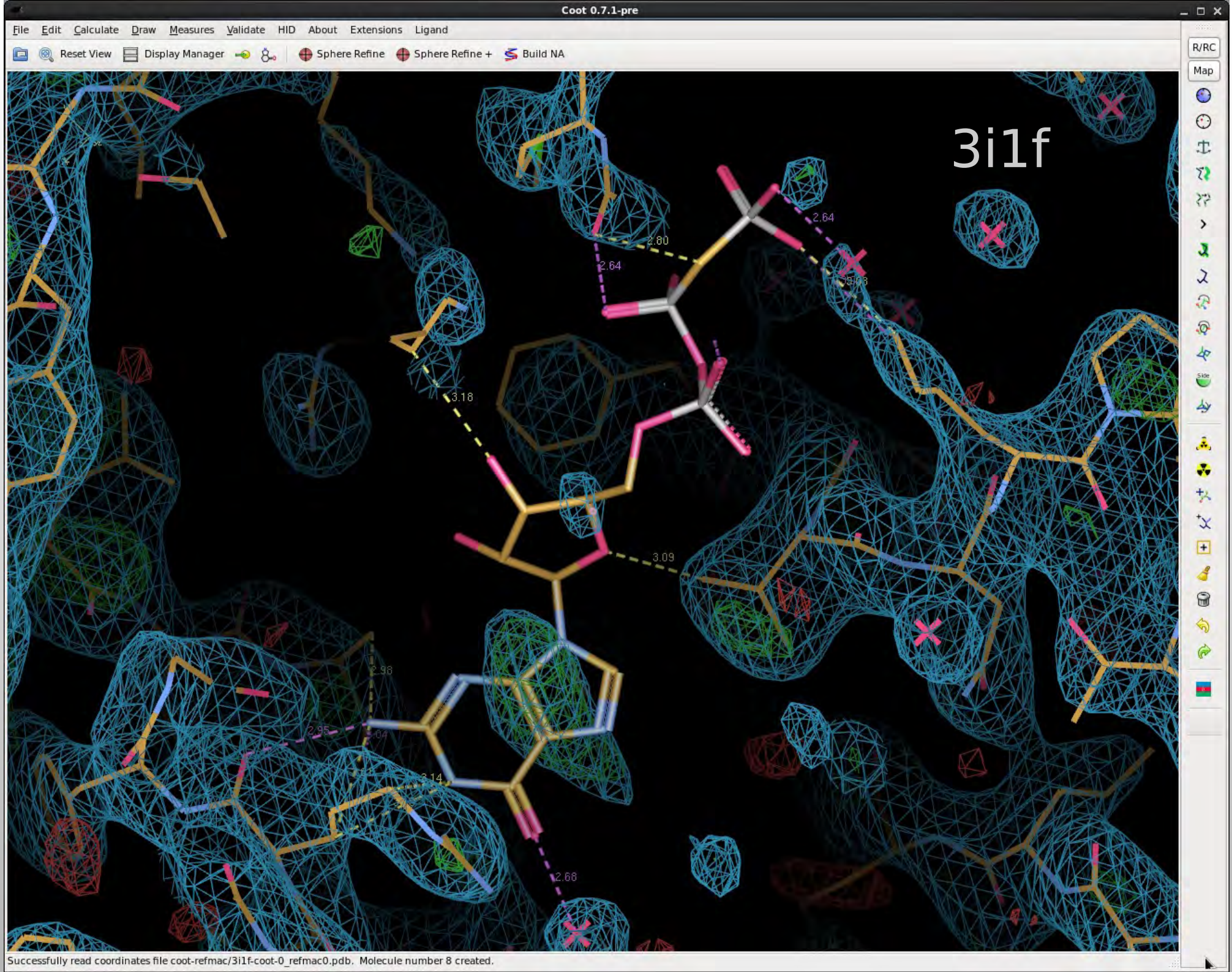

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
HO
HO
OH
(mol. no: 0) C2 /1/C/123 GOL occ: 1.00 bf: 27.06 ele: C pos: (20.98,-23.79,-4.40)


Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
HO
HO
OH
(mol. no: 0) C2 /1/C/123 GOL occ: 1.00 bf: 27.06 ele: C pos: (20.98,-23.79,-4.40)

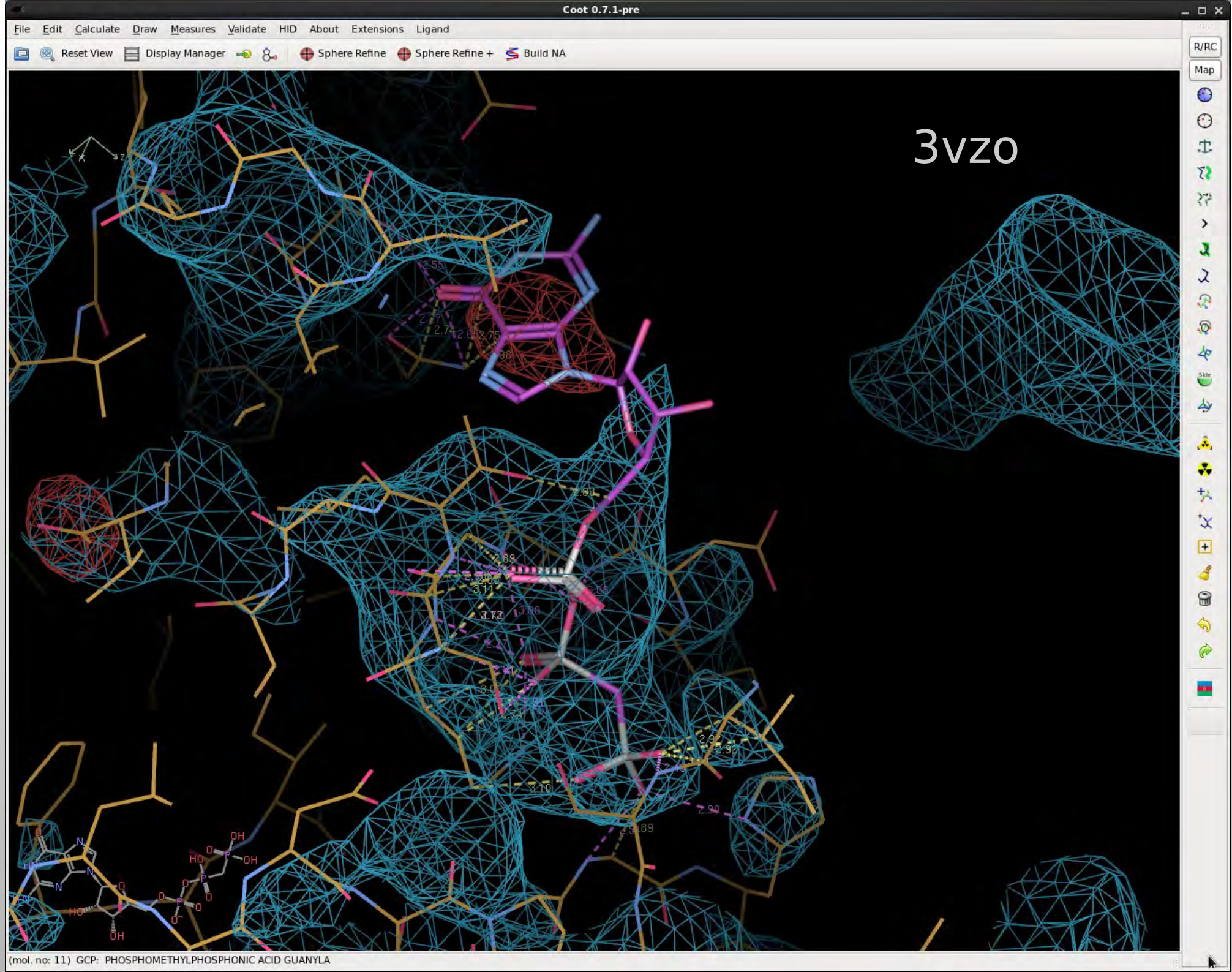

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 0) C2 /1/C/123 GOL occ: 1.00 bf: 27.06 ele: C pos: (20.98,-23.79,-4.40)


Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
HO
HO
OH
(mol. no: 0) C2 /1/A/123 GOL occ: 1.00 bf: 19.99 ele: C pos: (4.00,-18.32,-14.77)

Ligand Ranked #8470 of 8470

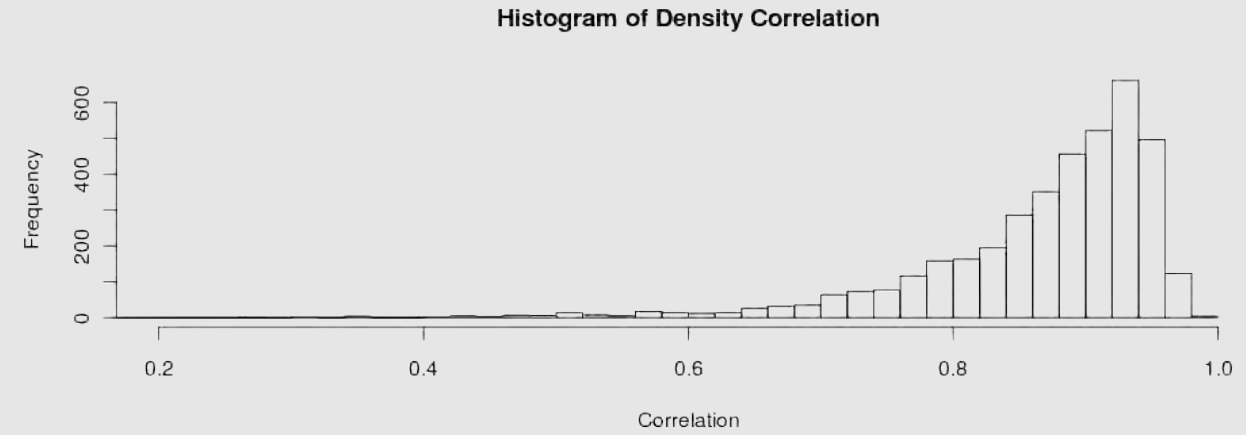

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
3i1f
2.64
2.80
2.64
3.18
3.09
2.98
2.95
3.14
2.68
Successfully read coordinates file coot-refmac/3i1f-coot-0_refmac0.pdb. Molecule number 8 created.

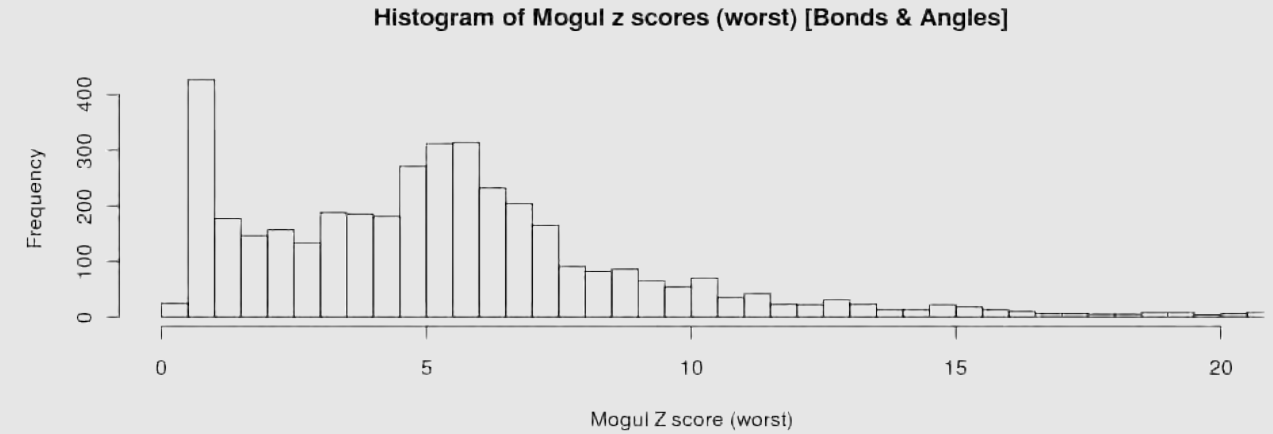
Rank #8469 of 8470

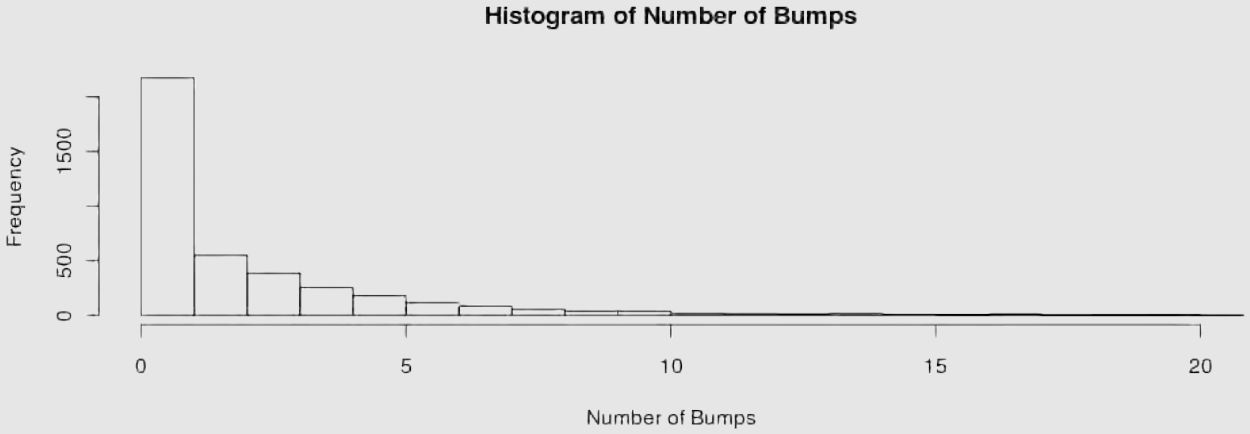

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
3vzo
(mol. no: 11) GCP: PHOSPHOMETHYLPHOSPHONIC ACID GUANYLA

Score Histograms

- Density Correlation
- Mogul z-score
- # Bumps/ligand


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

Ligand Scoring

Preliminary recommendatation...

Scoring Ligands:
To Be Better Than The Median:

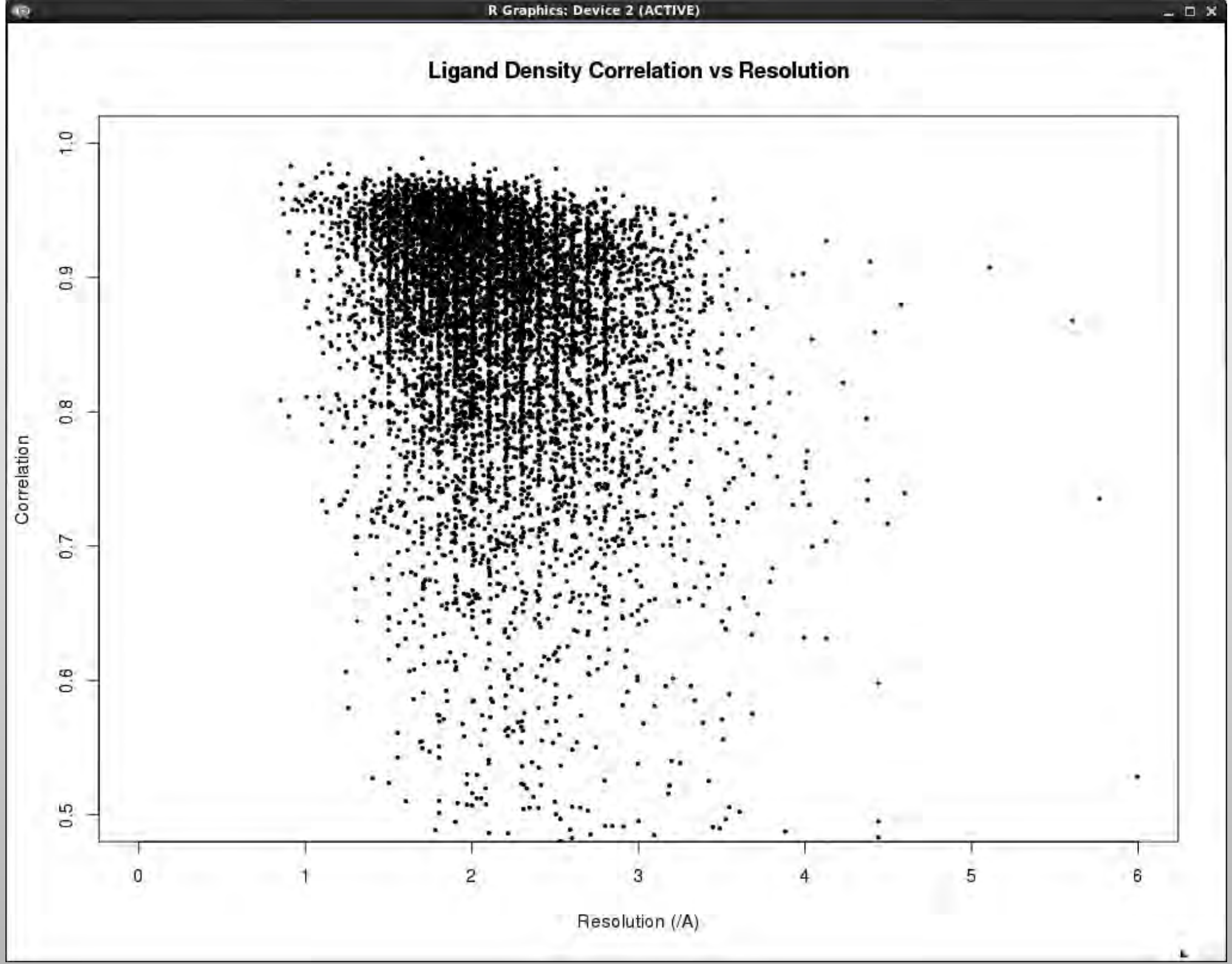
- 0 bumps
- Mogul z(worst) < 6.3
  - (note: query errors may be encoded in this value)
- Resolution Independence:
  - Density correlation > 0.9

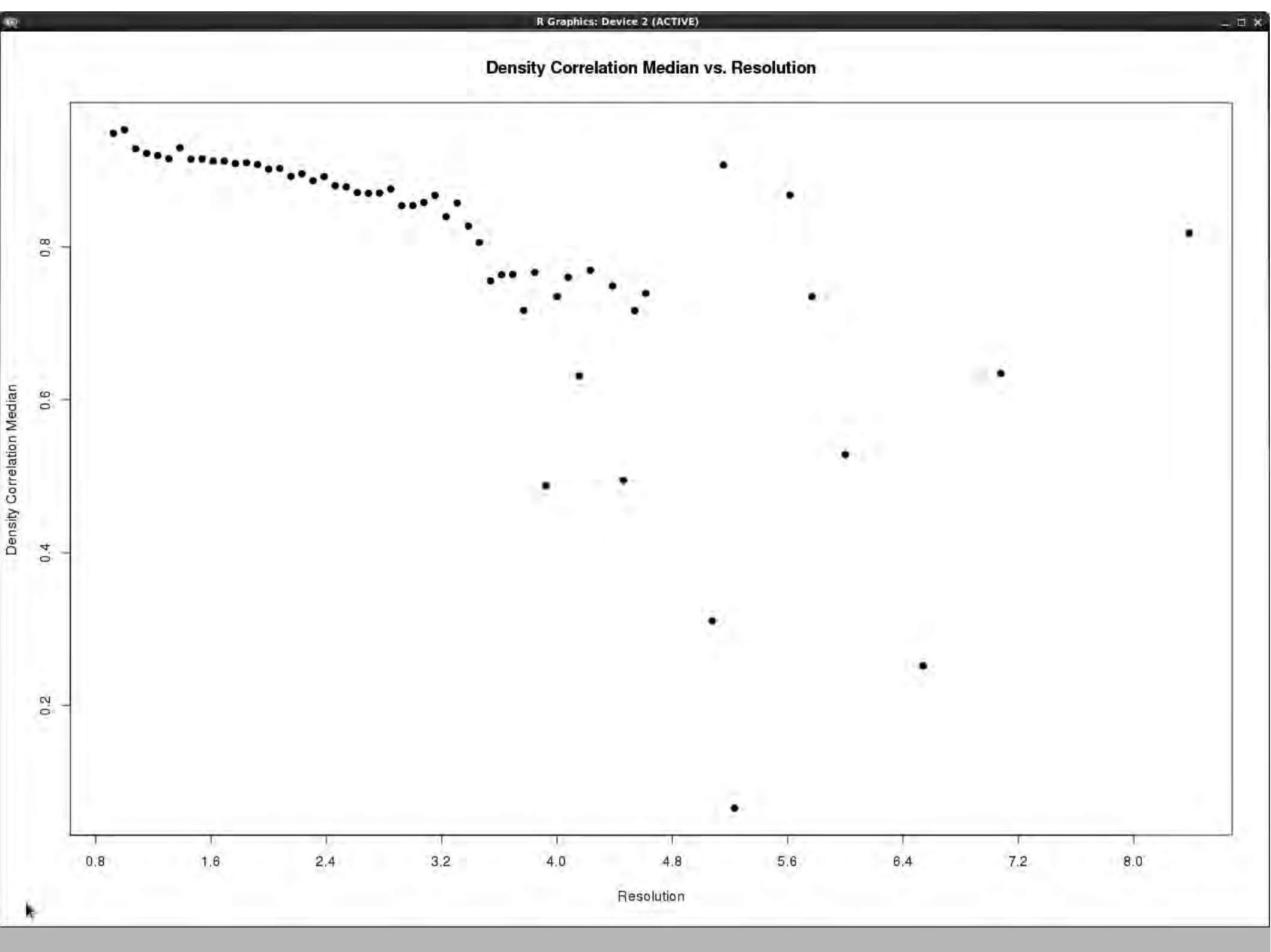
Effective Resolution

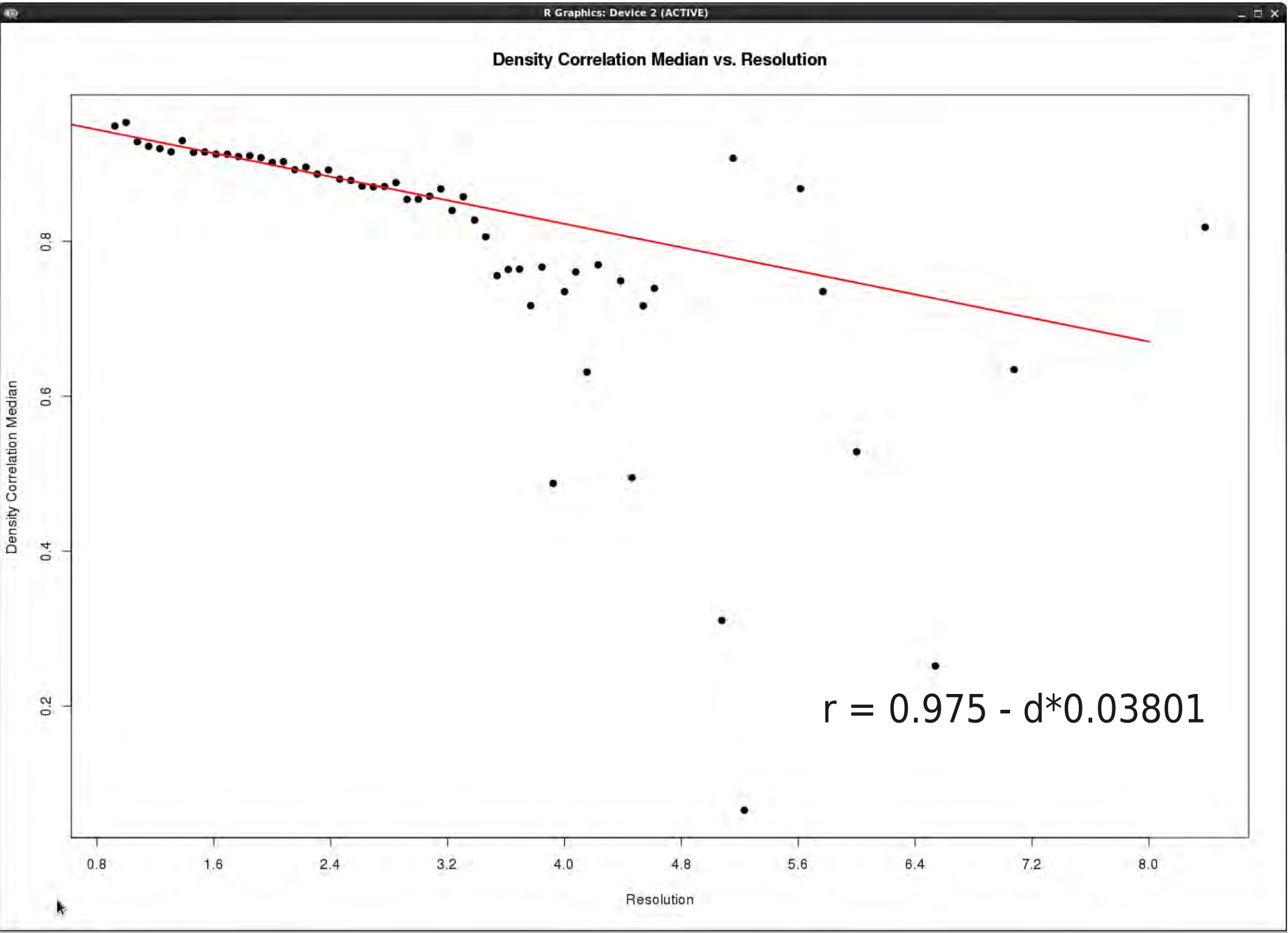
- Use standard deviations in the assessment of the data resolution

$$R_{eff} = \frac{R_{nom}}{\left(\frac{1}{N}\sum_i \sqrt{\frac{\varepsilon_i <F^2>}{\sigma_{Fi}^2 + \varepsilon_i <F^2>}}\right)^{1/3}}$$

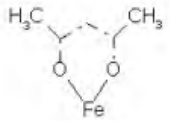
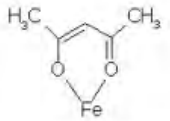
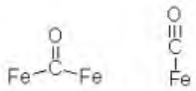
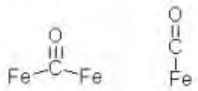
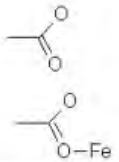
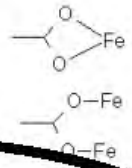
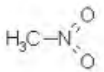
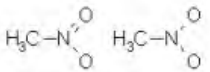
$$< F^2 >= \frac{1}{n_{bin}} \sum_{bin} \frac{F_j^2}{\varepsilon_j}$$


R Graphics: Device 2 (ACTIVE)
Ligand Density Correlation vs Resolution
Correlation
1.0
0.9
0.8
0.7
0.6
0.5
0
1
2
3
4
5
6
Resolution (/A)


R Graphics: Device 2 (ACTIVE)
Density Correlation Median vs. Resolution
Density Correlation Median
0.2
0.4
0.6
0.8
0.8
1.6
2.4
3.2
4.0
4.8
5.6
6.4
7.2
8.0
Resolution

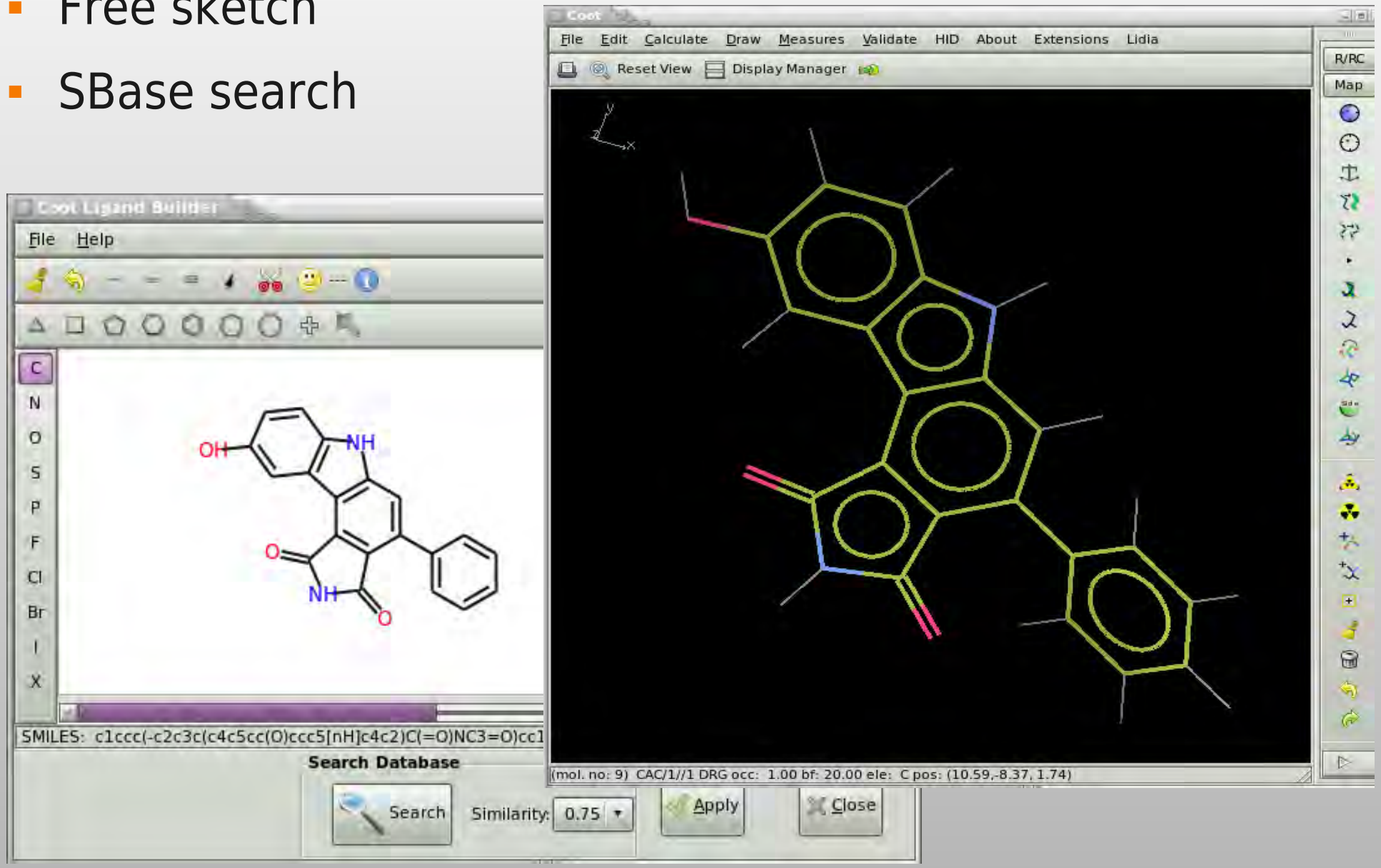

R Graphics: Device 2 (ACTIVE)
Density Correlation Median vs. Resolution
Density Correlation Median
0.2
0.4
0.6
0.8
0.8
1.6
2.4
3.2
4.0
4.8
5.6
6.4
7.2
8.0
Resolution
r = 0.975 - d*0.03801

Difficult Functional Groups and Ions

Group	Guideline	Correct Example	Incorrect Example
Acetylacetonato ion, coordinated to metal.	**Use delocalised bond type for the carbon-oxygen and carbon-carbon bonds is the conjugated system.**		
Carbonyl.	**Use a double C=O bond if the group is bridging metal atoms, but use a triple bond if it is bonded to only one metal atom.**		
Carboxylate ion, uncoordinated or coordinated via only one of the oxygen atoms. Or thio equivalent.	**Use one single C-O bond and one double C=O bond.**		
Carboxylate ion, bidentate to one or two metals. Or thio equivalent.	**Use the delocalised bond type for both carbon-oxygen bonds.**		
Nitro and nitrate.	**Use two double N=O bonds (an uncoordinated nitrate ion would have two double bonds and one single).**		
Perchlorate ion.	**Use three double bonds and one single bond.**		

2D Ligand Builder

- Free sketch
- SBase search

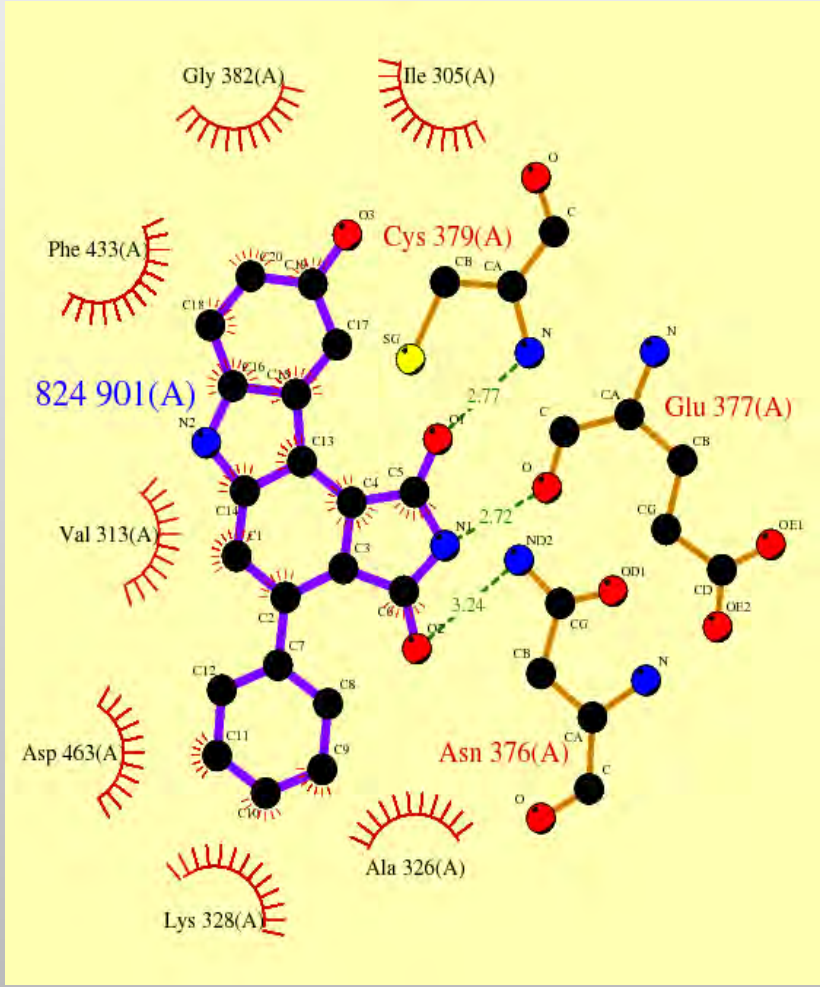


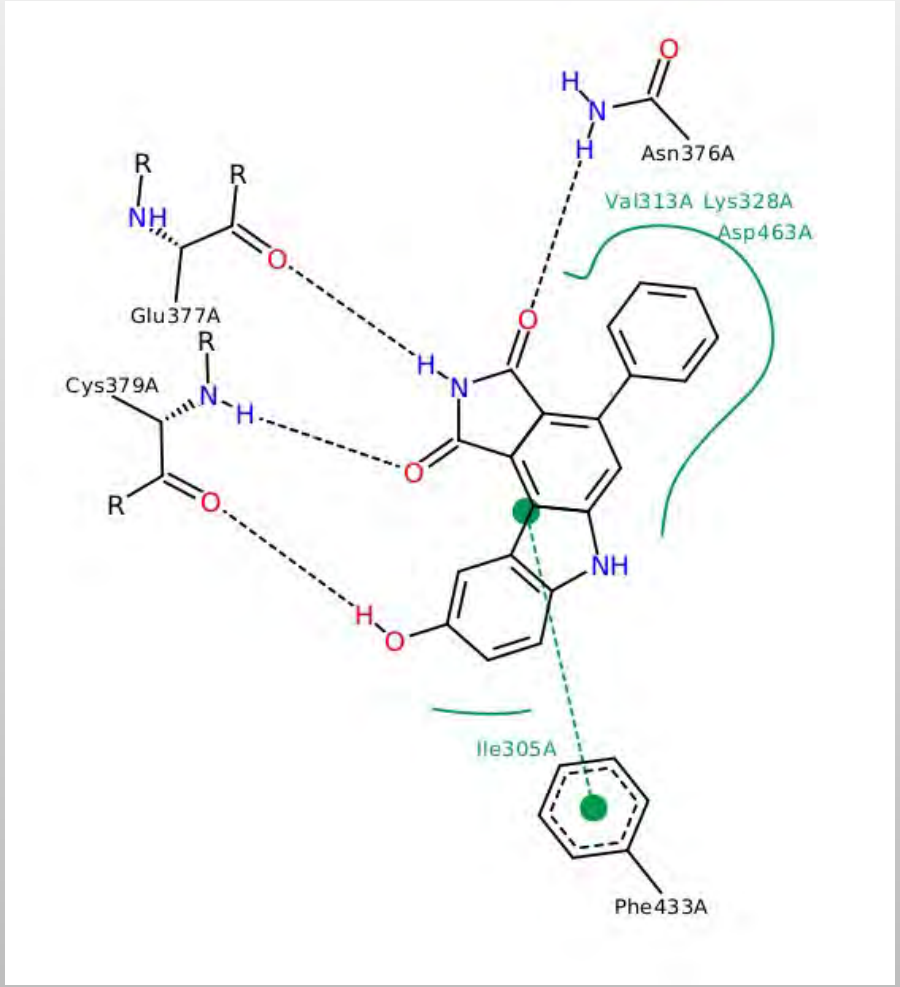
Ligand Represenation

- Bond orders (from dictionary restraints)

Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





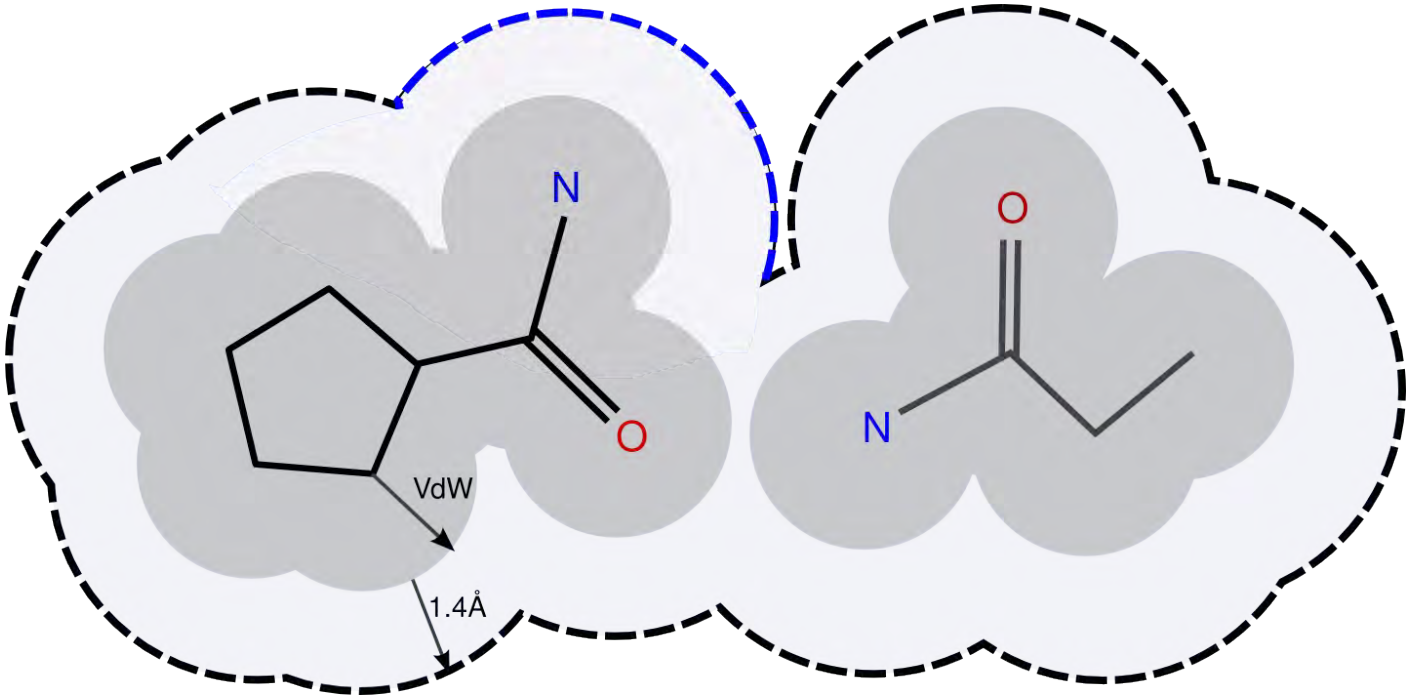
Can we do better? - Interactivity?

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

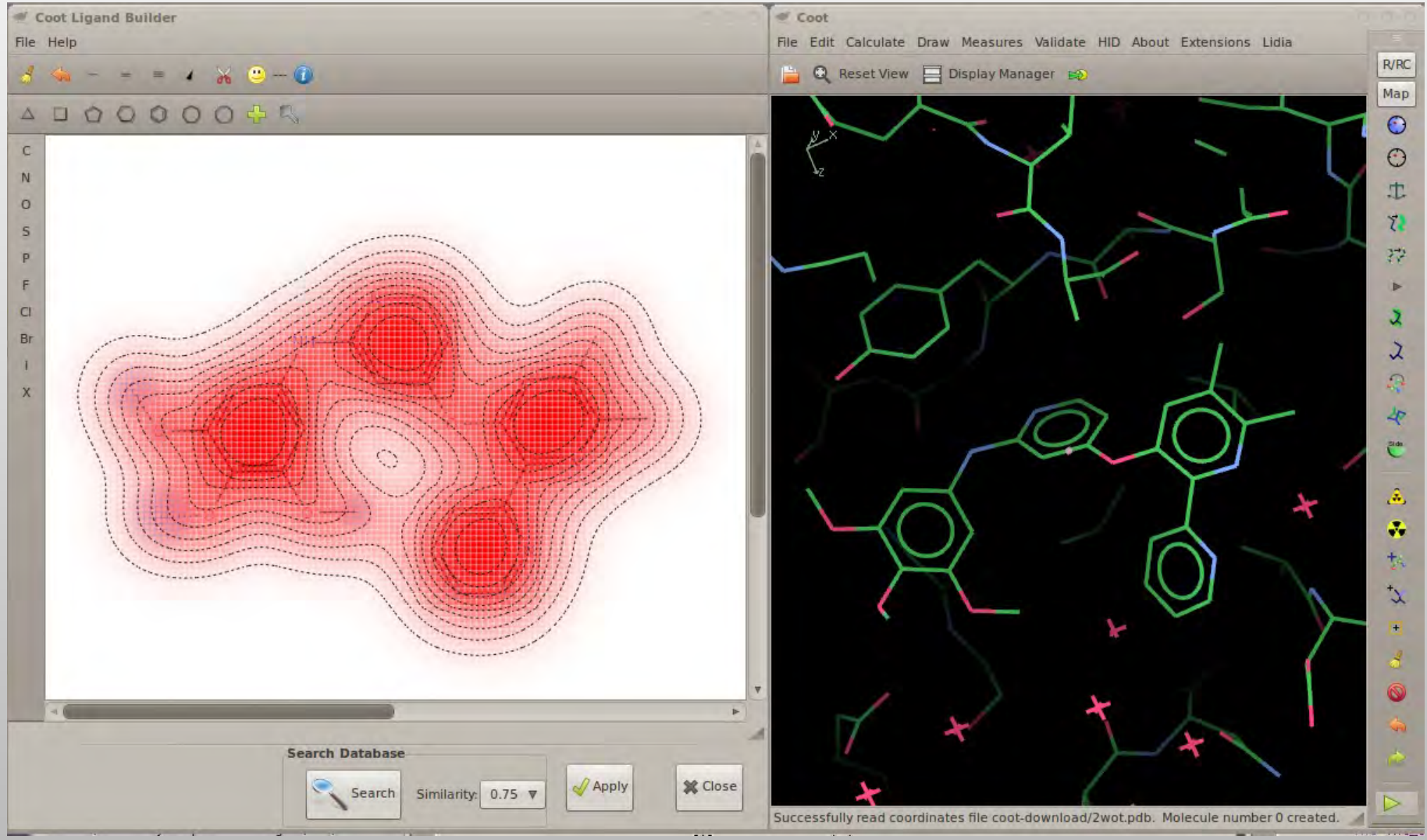
$$\sum\sum (d_{ik}^2 - D_{ik}^2) +$$

Residues are close to H-bonding ligand atoms

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

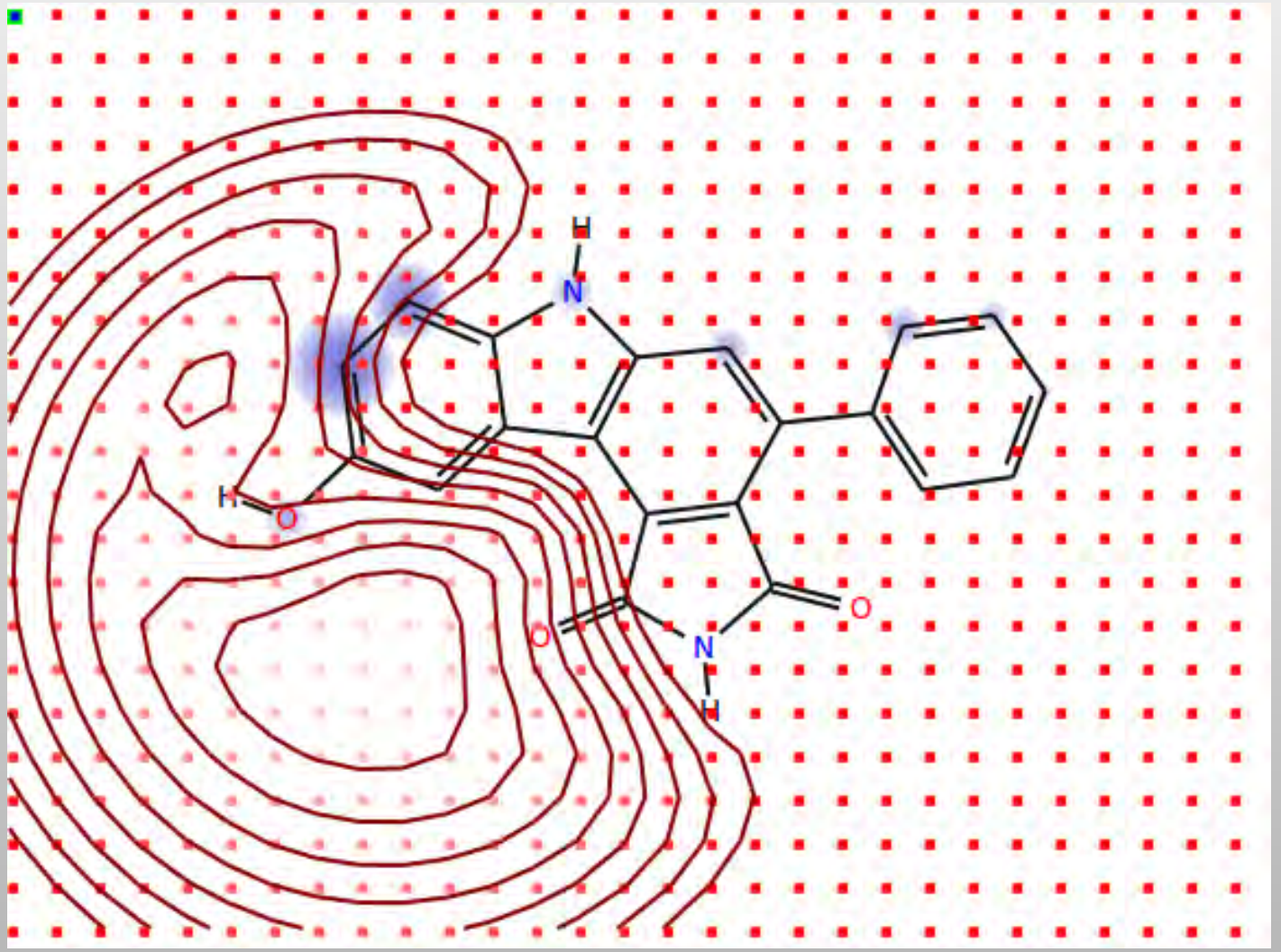
Residues don't overlap ligand

”Don't overlap the ligand”



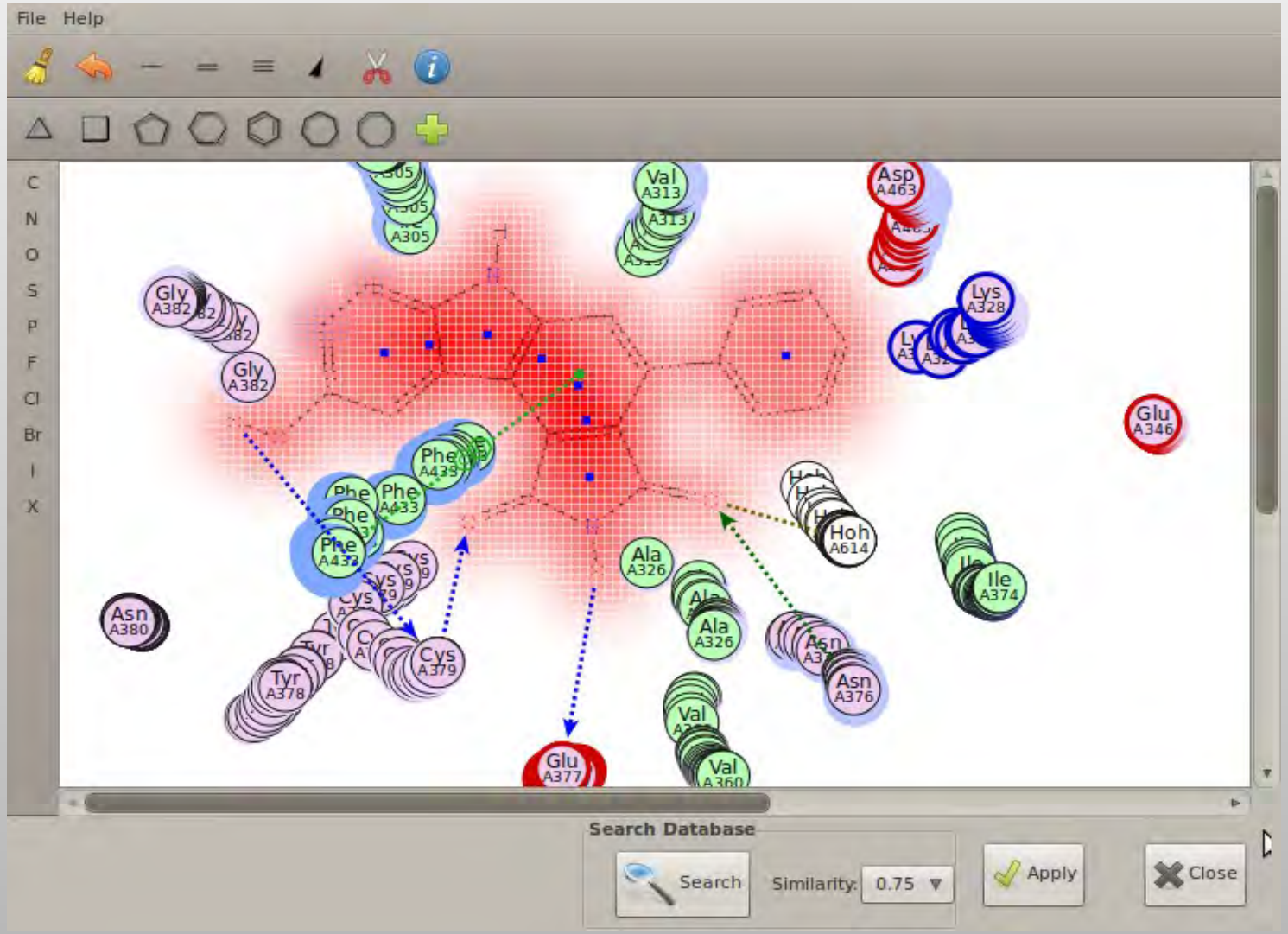
Ligand Environment Layout

- Initial residue placement



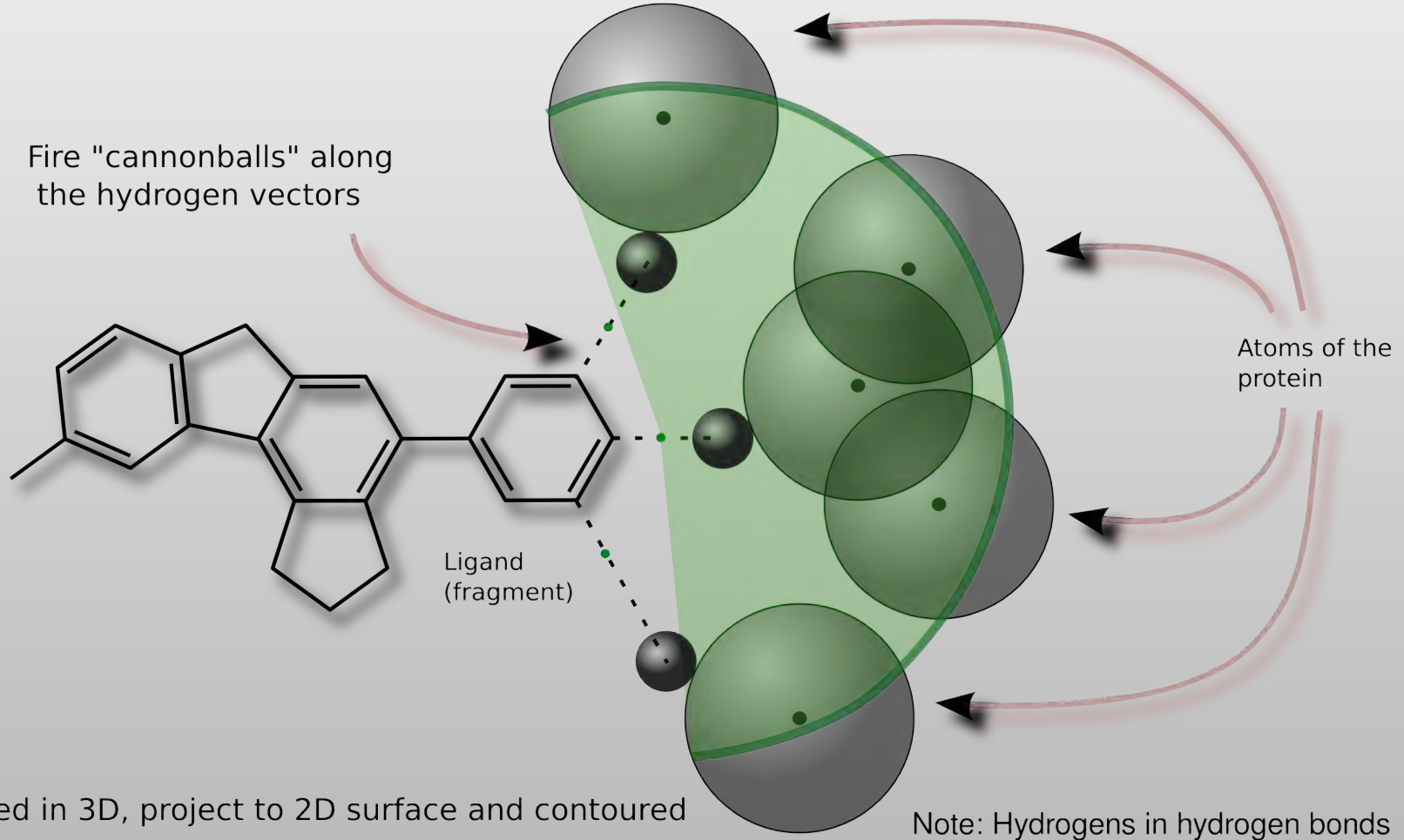
Ligand Environment Layout

- 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?



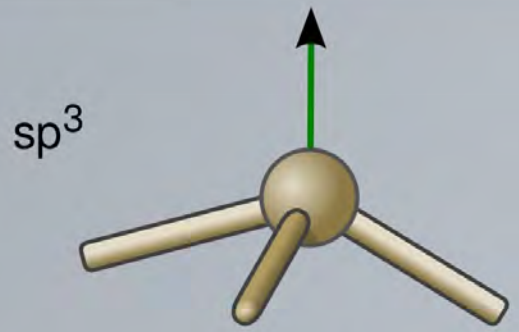
Determined in 3D, project to 2D surface and contoured

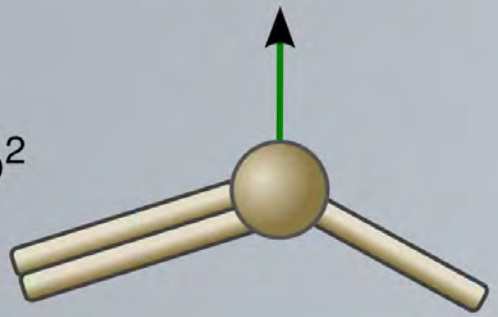
c.f. Clarke & Labute (2007)

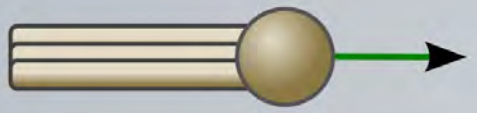
Note: Hydrogens in hydrogen bonds are a confounding factor

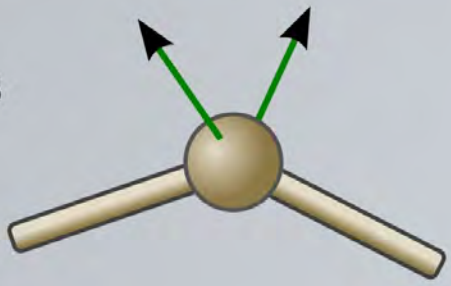
Substitution Contour: Extending along Hydrogens

Riding Hydrogens



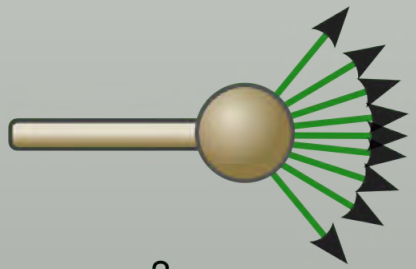


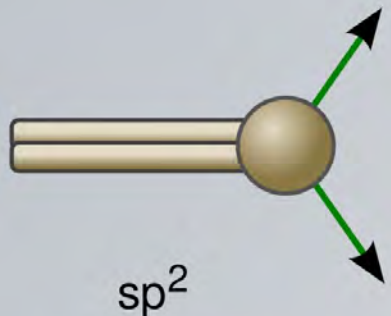


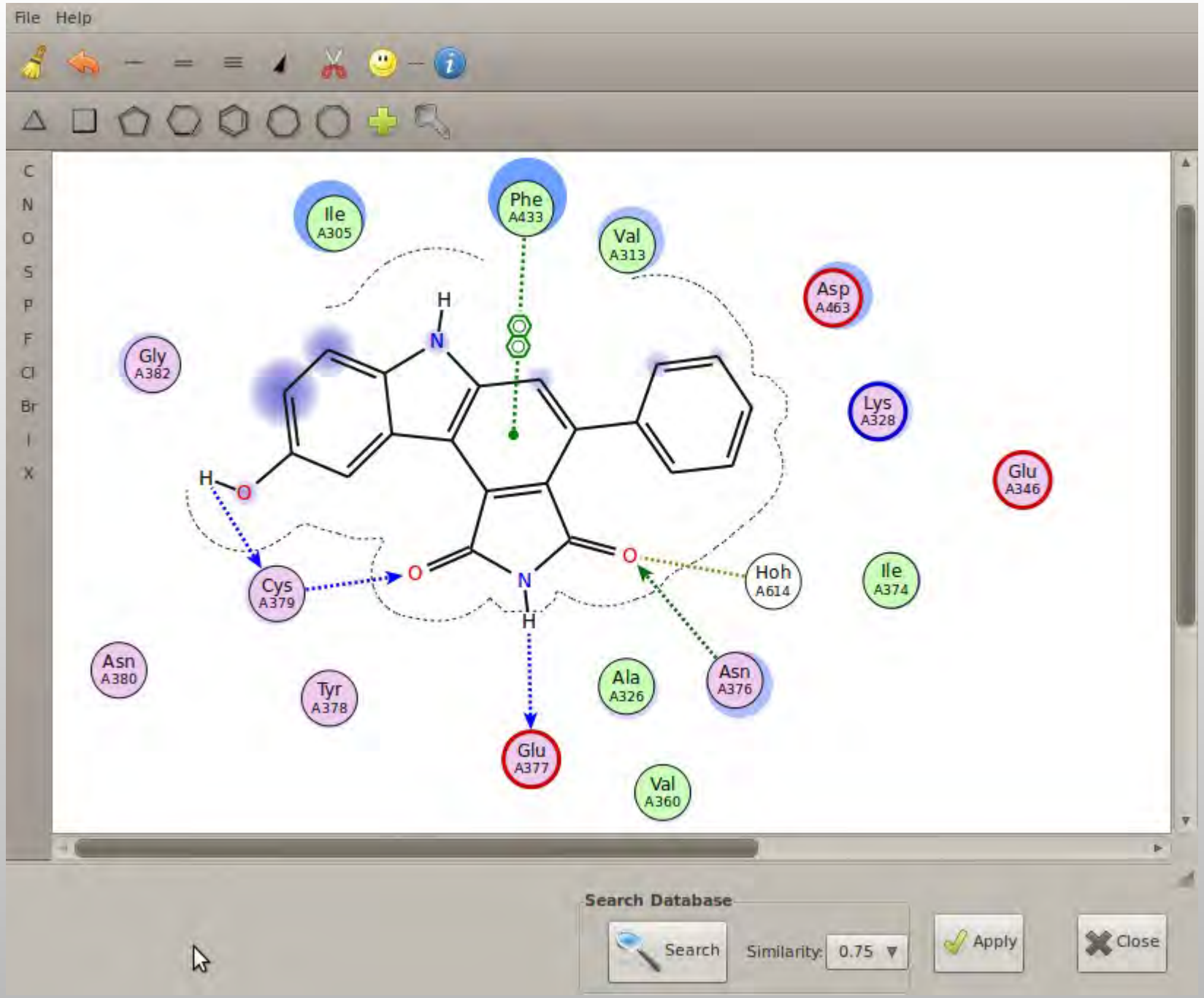


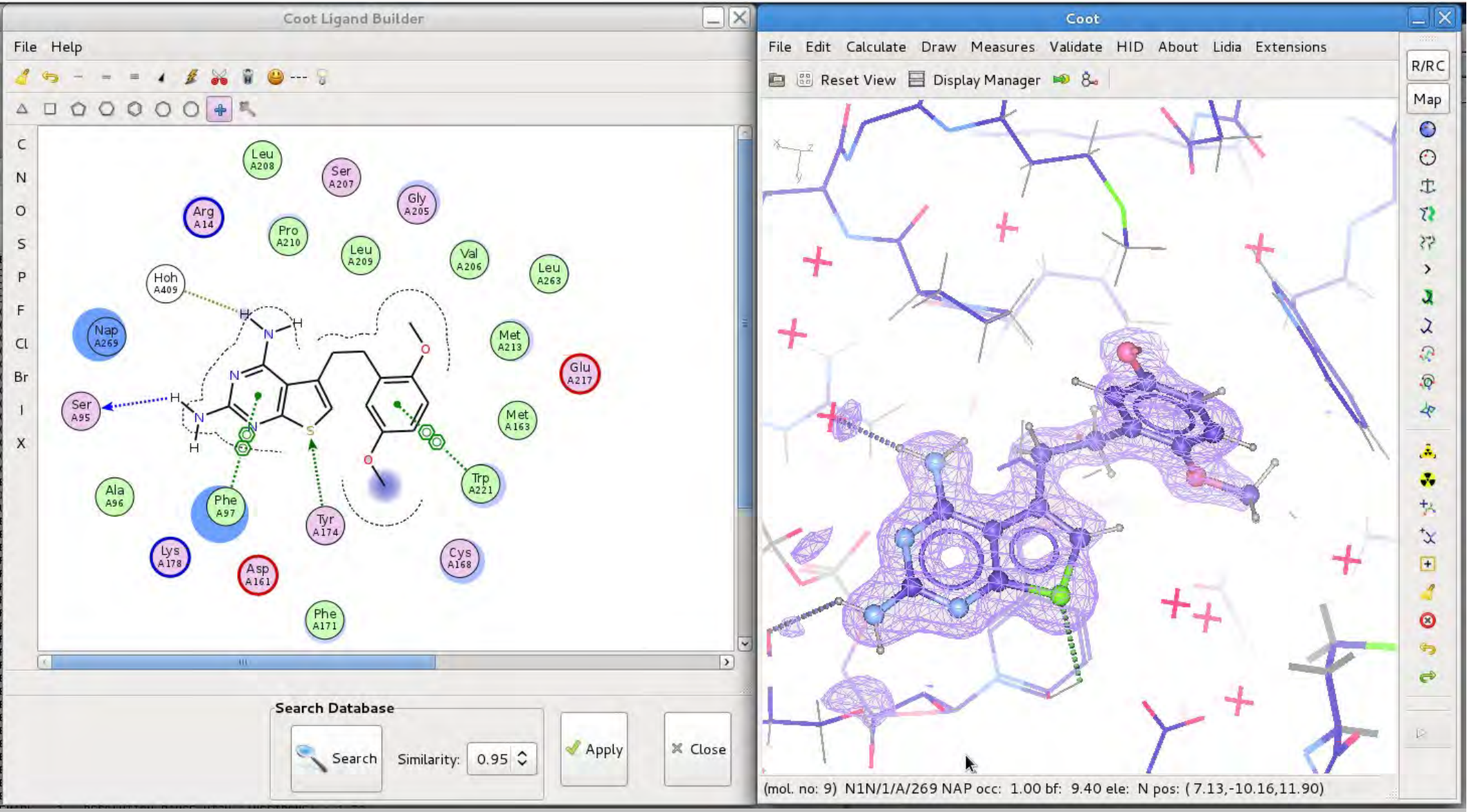
Torsionable Hydrogens

(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


Coot Ligand Builder
File Help
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)

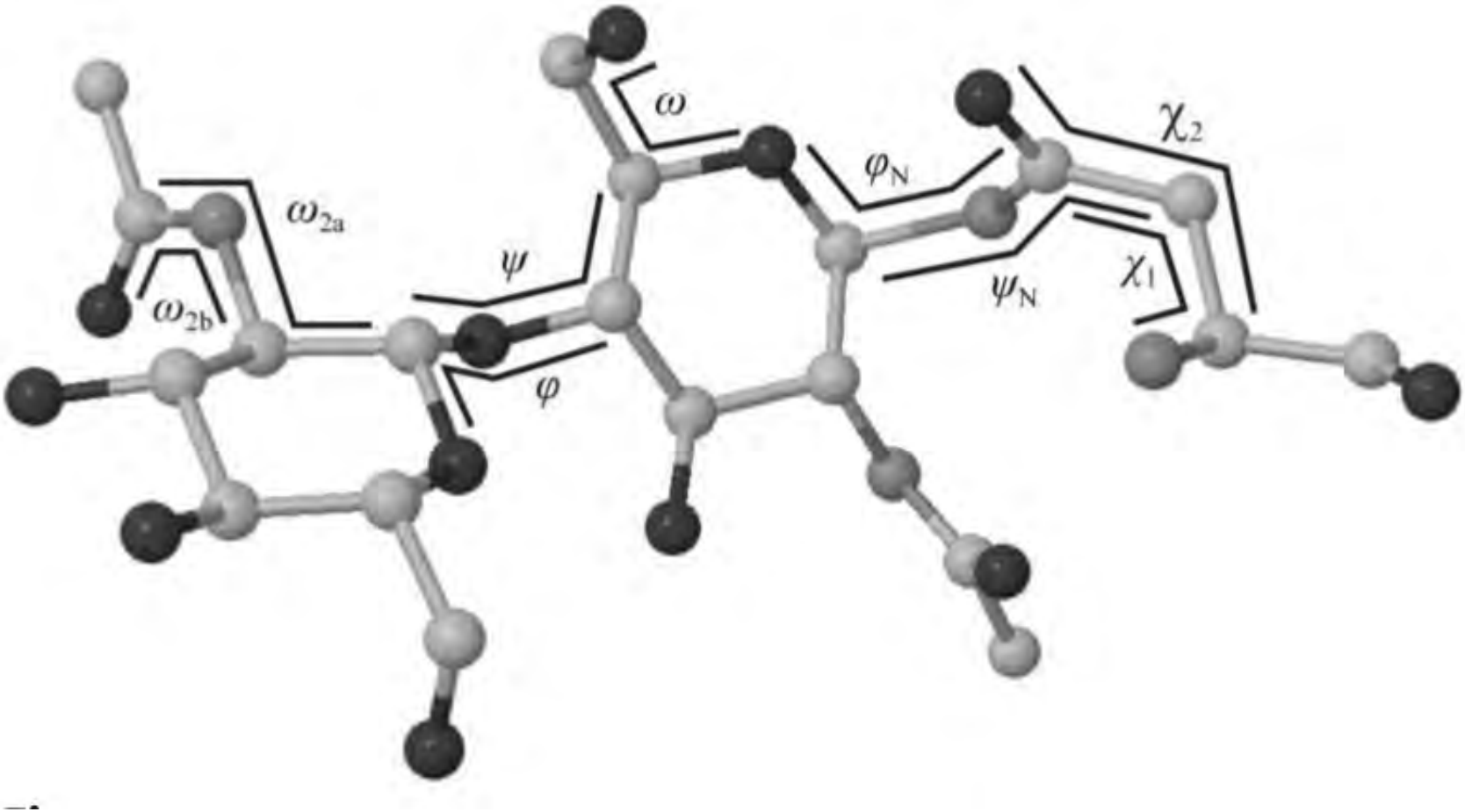
Modelling Carbohydrates

- Validation,
- Model-building,
- Refinement

Problematic Glycoproteins

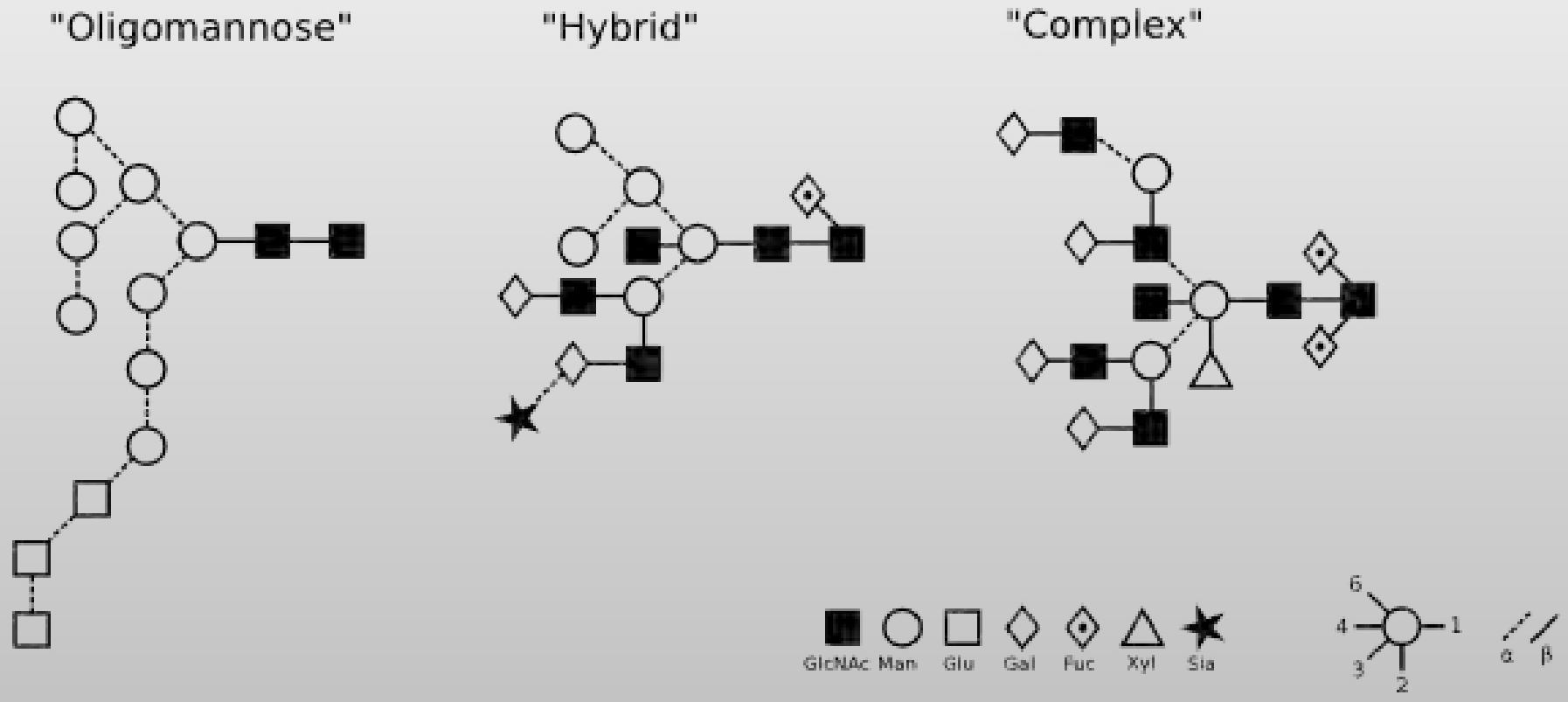
- Crispin, Stuart & Jones (2007)
  - NSB Correspondence
  - “one third of entries contain significant errors in carbohydrate stereochemistry...”
  - “carbohydrate-specific building and validation tools capable of guiding and construction of biologically relevant stereochemically accurate models should be integrated into popular crystallographic software. Rigorous treatment of the structural biology of glycosylation can only enhance the analysis of glycoproteins and our understanding of their function”
  - PDB curators concur

Carbohydrate Links



Thomas Lütteke (2007)

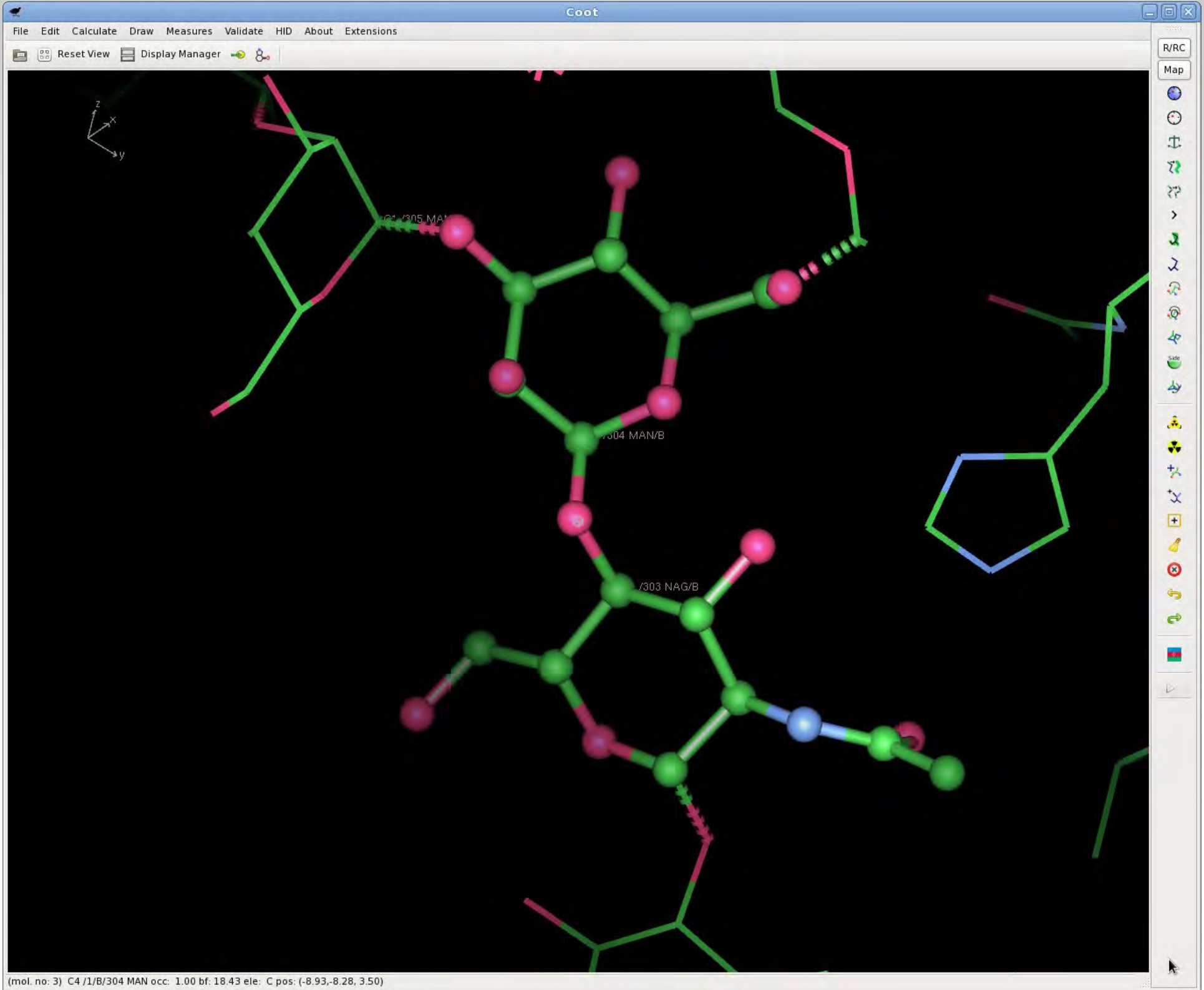
Validate the Tree: N-linked carbohydrates

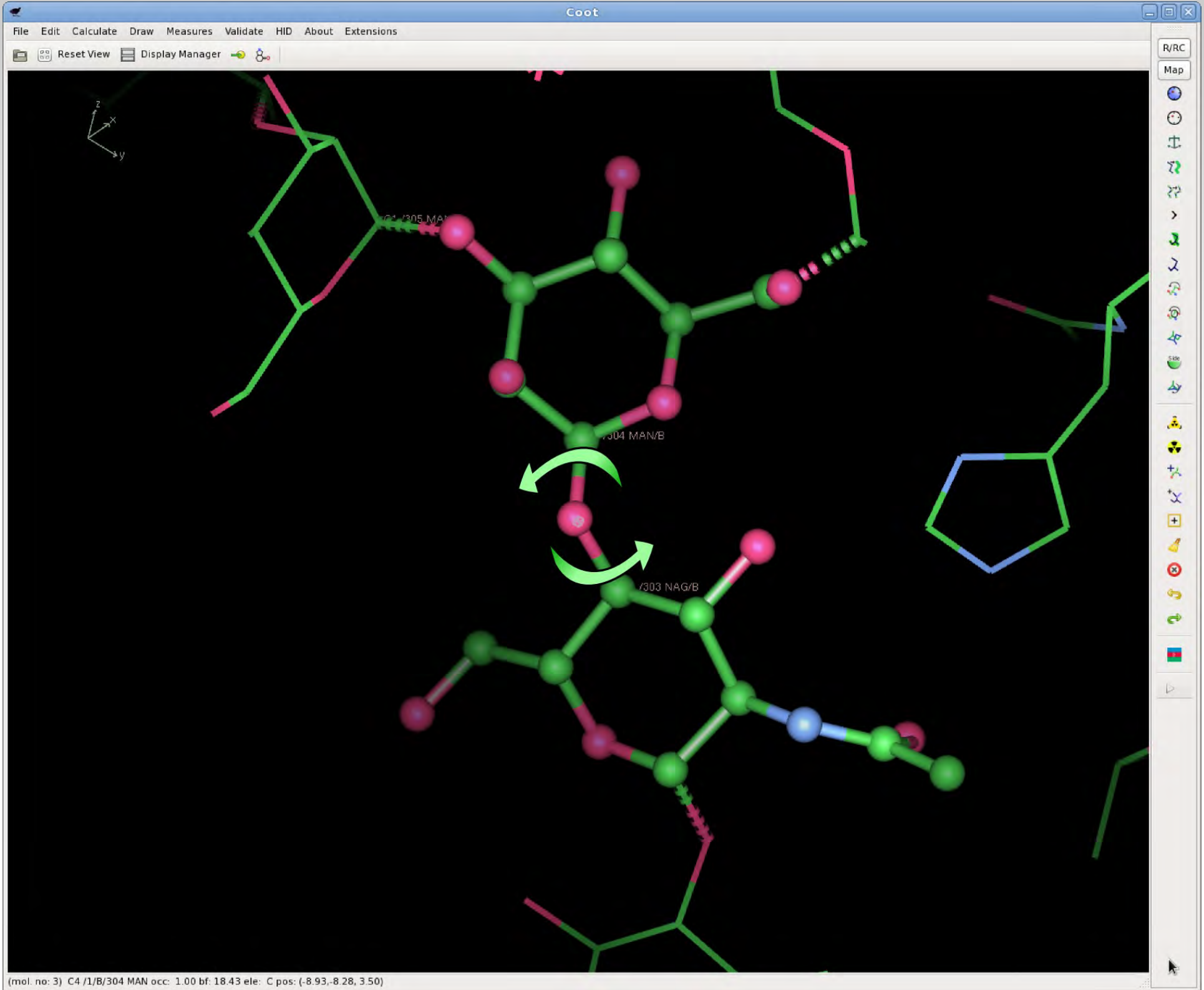


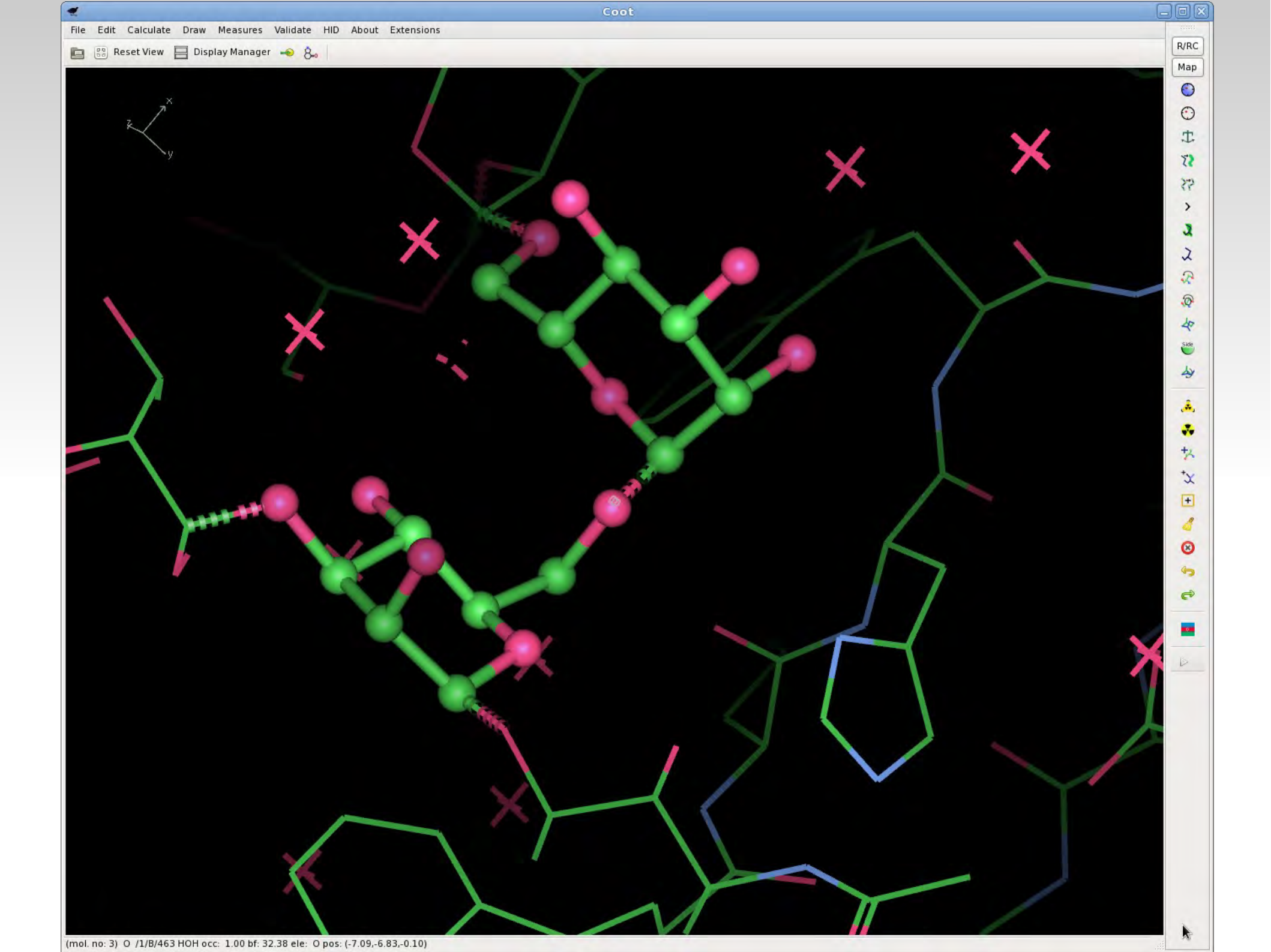
Linking Oligsaccharides/Carbohydrates:

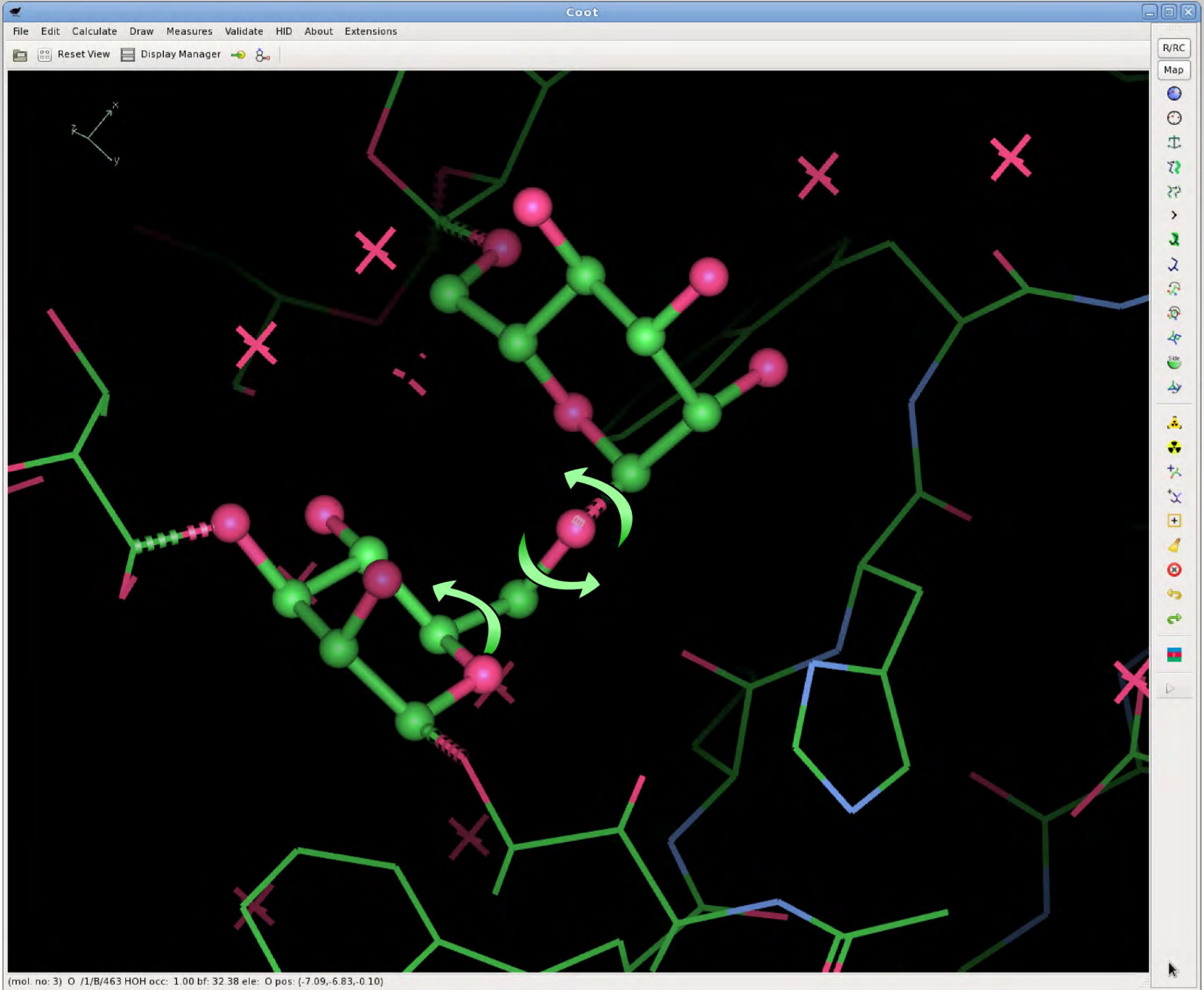
LO/Carb

- Complex carbohydrate structure
  - from a dictionary of standard links
  - and monomers
  - torsion-angle refinement

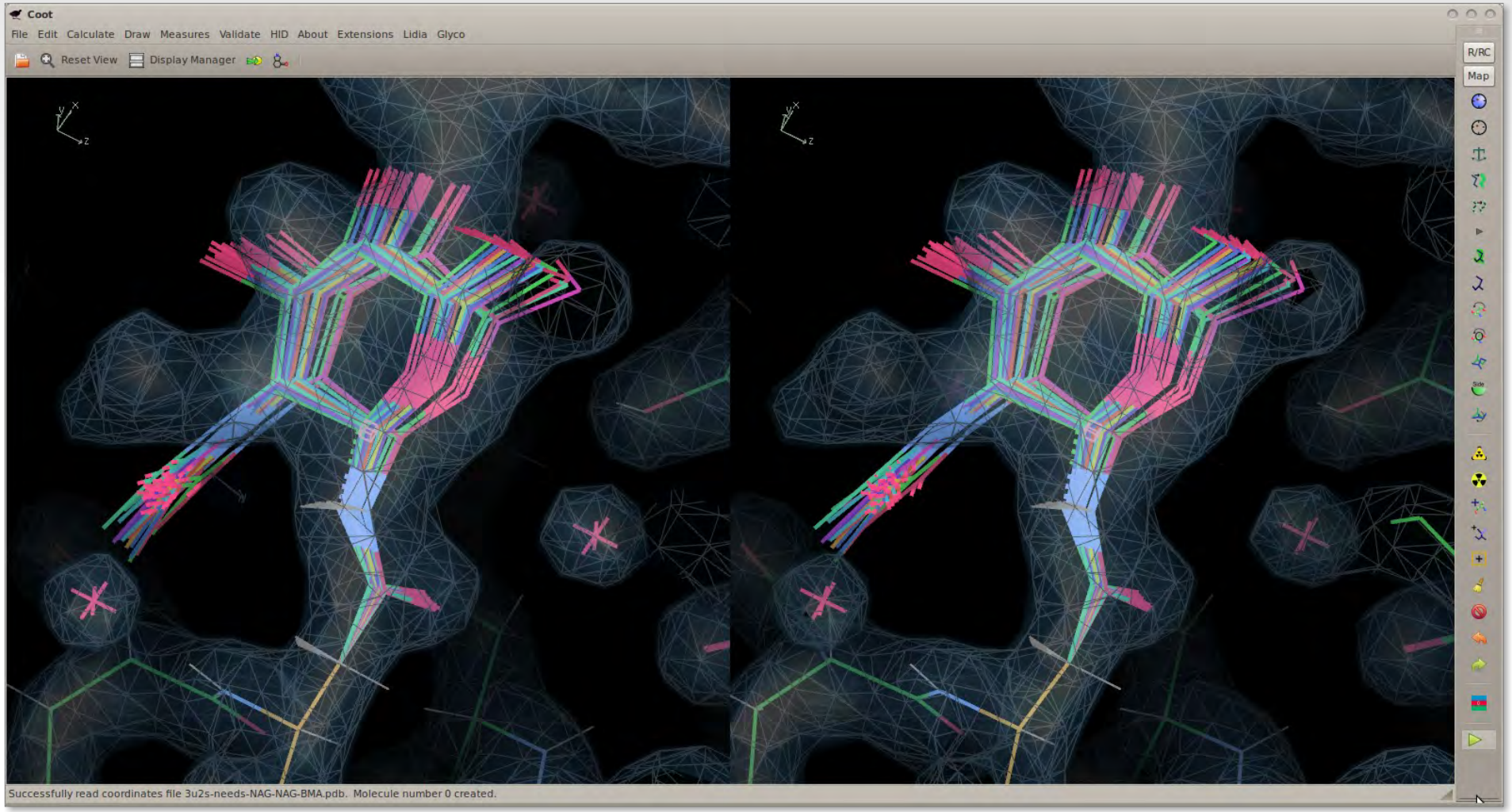

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

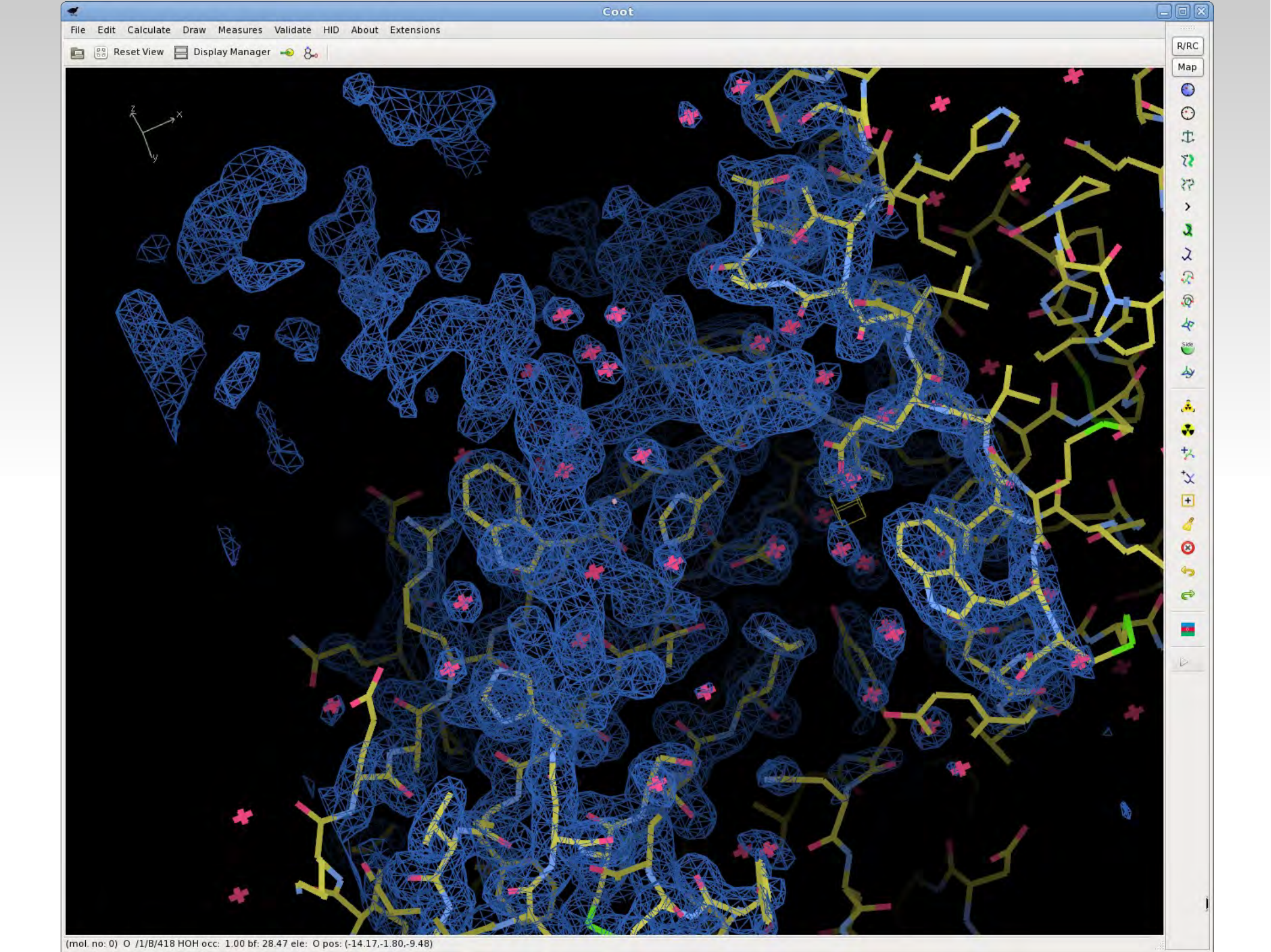

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

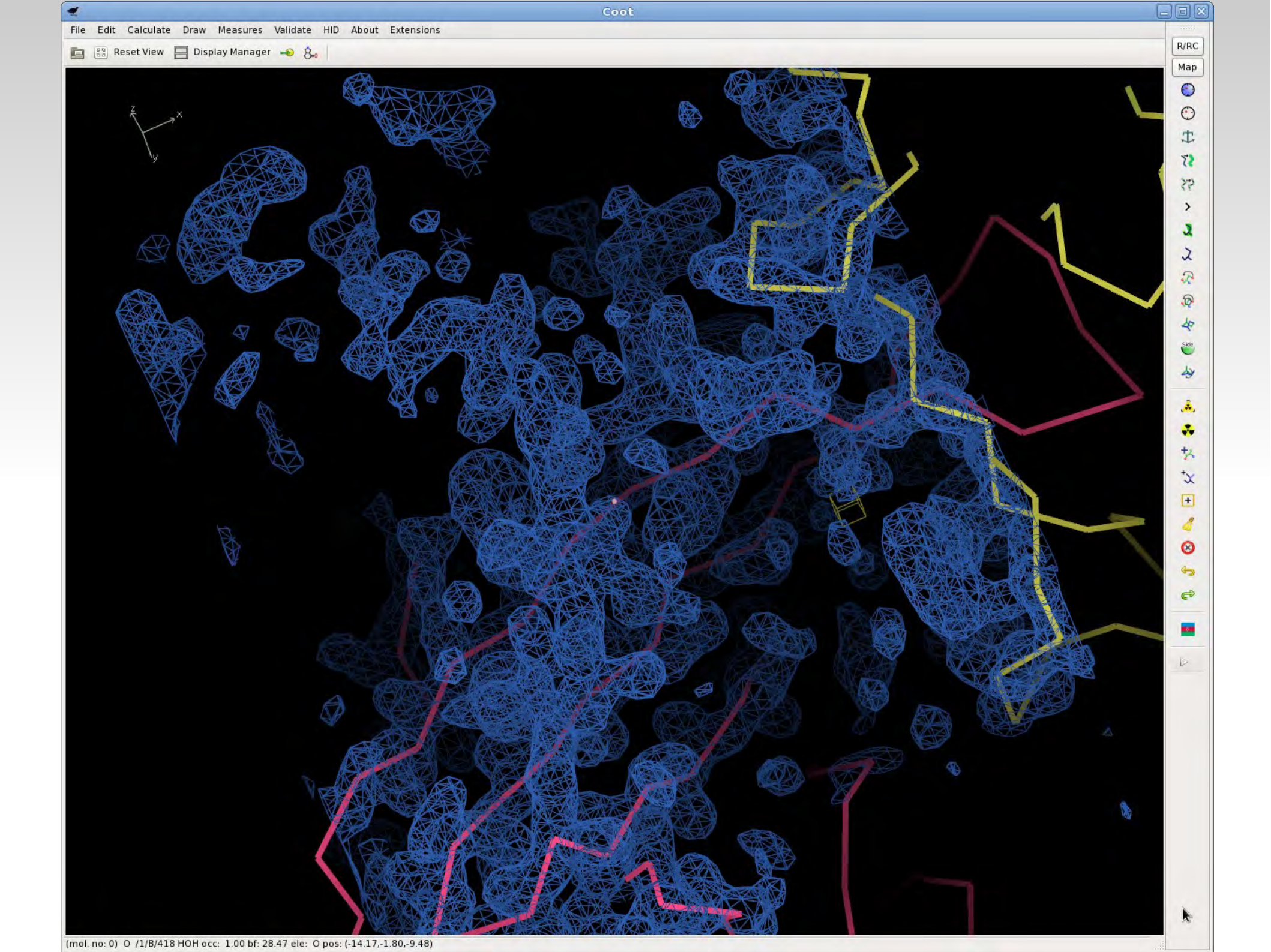

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 3) O /1/B/463 HOH occ: 1.00 bf: 32.38 ele: O pos: (-7.09,-6.83,-0.10)


Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 3) O /1/B/463 HOH occ: 1.00 bf: 32.38 ele: O pos: (-7.09,-6.83,-0.10)

Refinement Trials (NAG-ASN example)

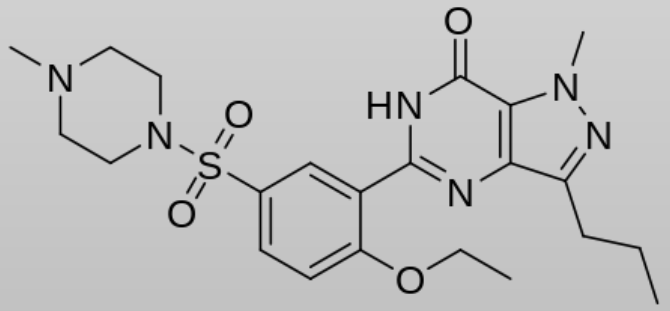



Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 0) O /1/B/418 HOH occ: 1.00 bf: 28.47 ele: O pos: (-14.17,-1.80,-9.48)


Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 0) O /1/B/418 HOH occ: 1.00 bf: 28.47 ele: O pos: (-14.17,-1.80,-9.48)

Coot Release 0.7.1

- Fixes to ligand fitting
- Fixes to Sequence View
- Retrive PDBe ligand description
  - (for new ligands)
- Improvements to Mogul Interface
- Lidia
  - Keyboard accelerators
    - target sildenafil in 20 seconds



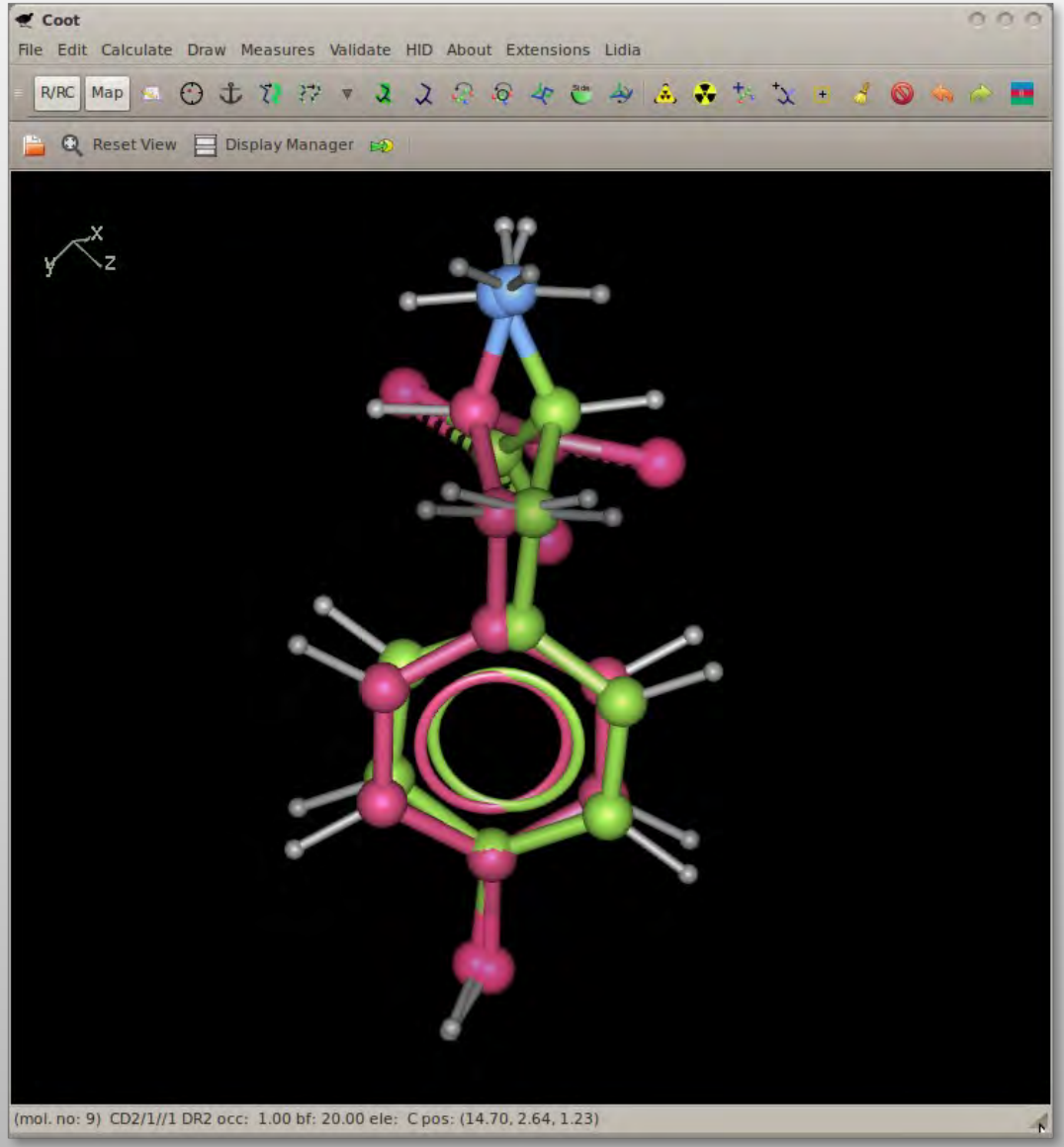
Acknowledgements

- Kevin Cowtan
- Bernhard Lohkamp

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

- Funding:
  - BBSRC & CCP4

Chiral Centre Inversion

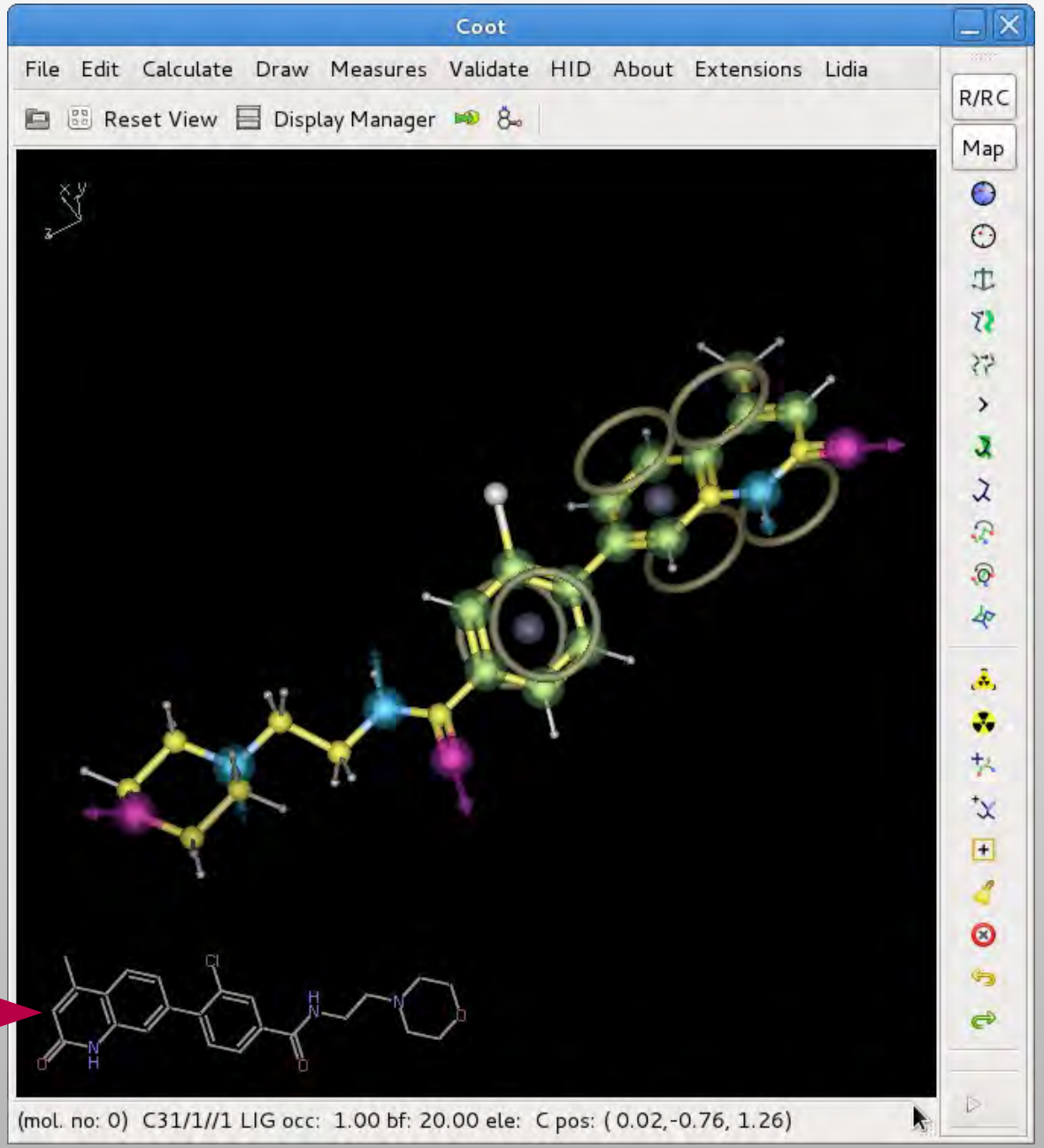


Inverted chiral centre refinement pathology detection

Hydrogen tunnelling

Chemical Features

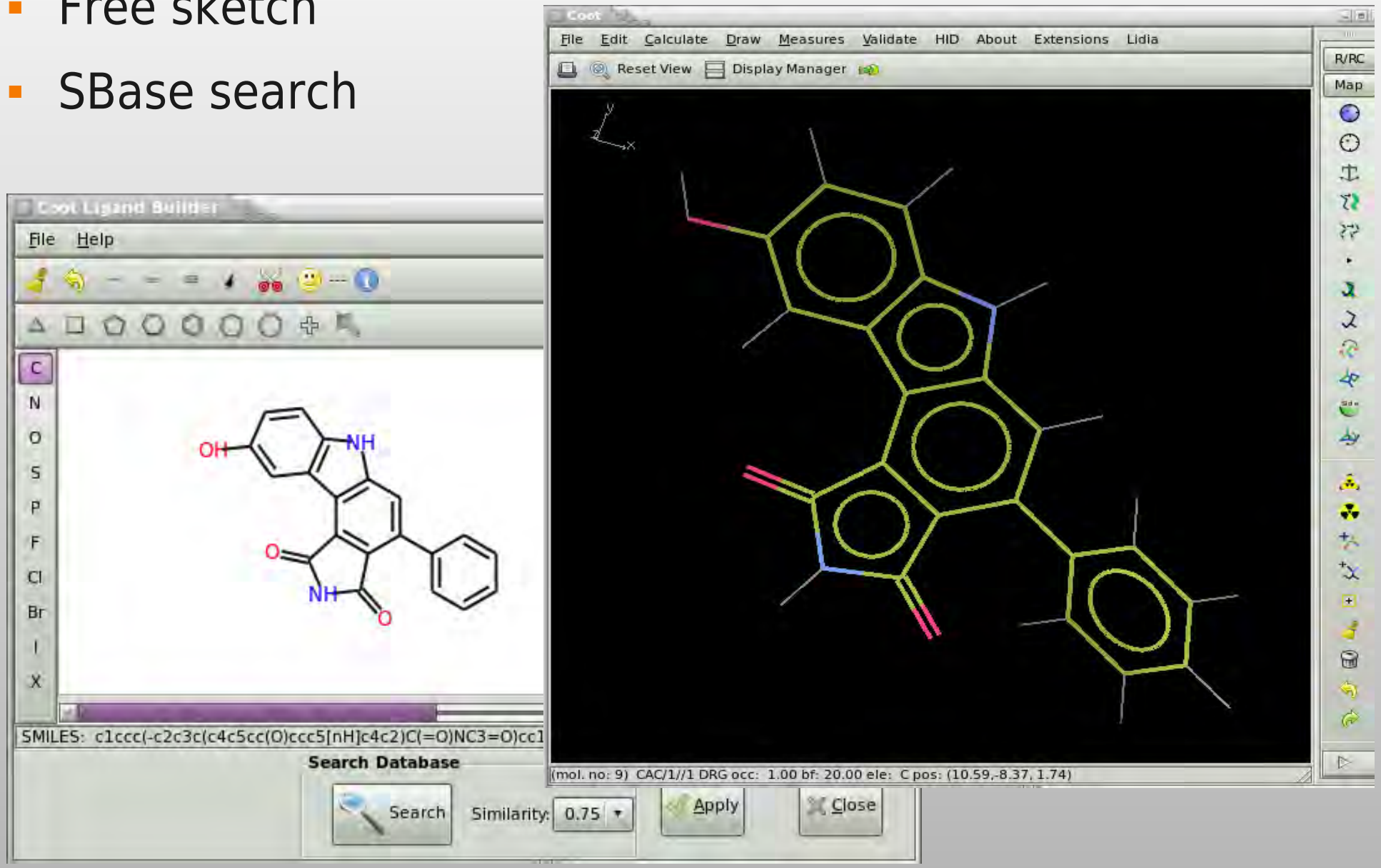
Uses built-in FeatureFactory



...and on the fly thumbnailing

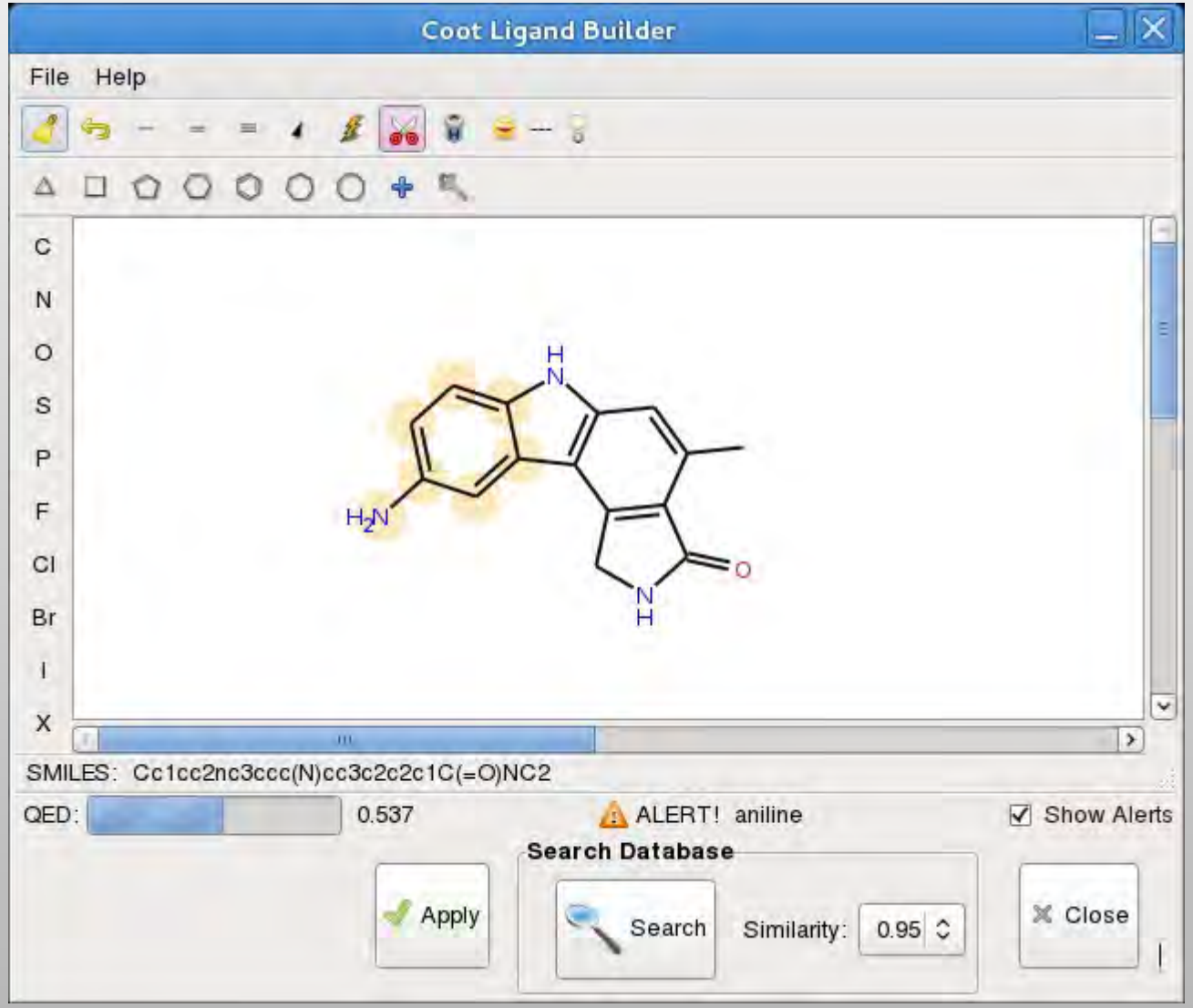
2D Ligand Builder

- Free sketch
- SBase search



2D Sketcher

- Structural Alerts



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

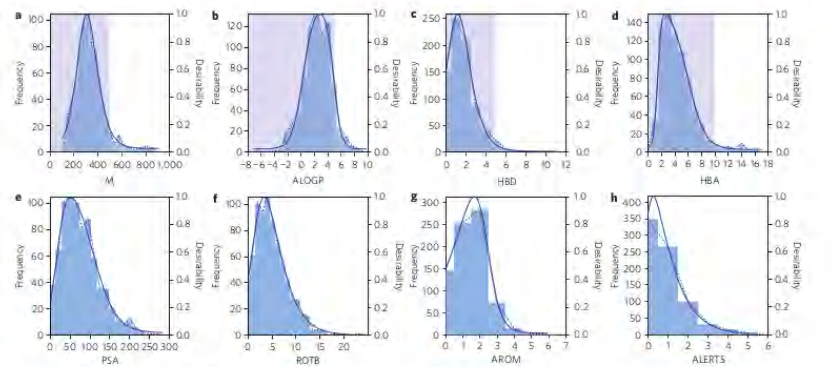


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h. Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**),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

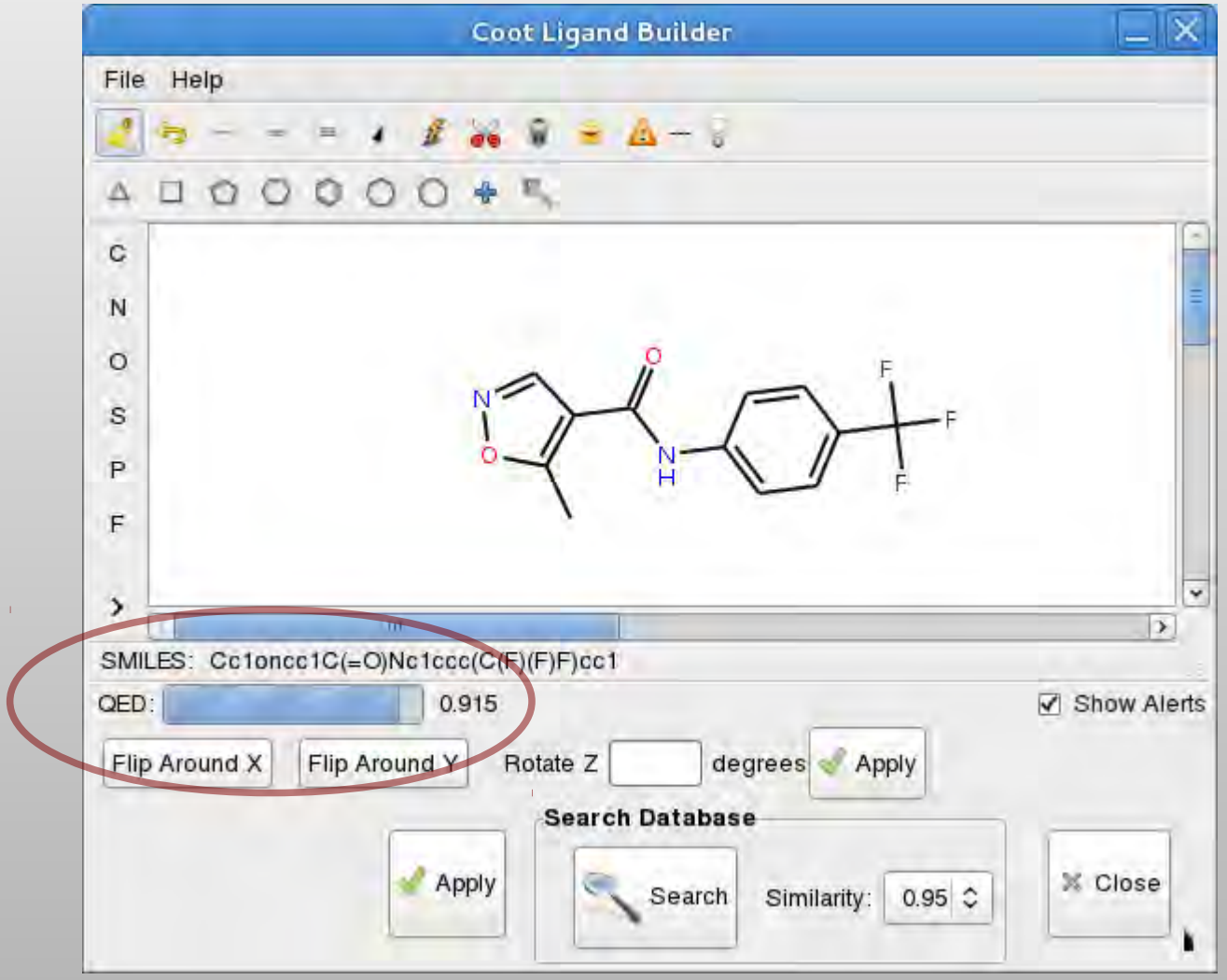
asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

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

Bickerton et al (2012) *Nature Chemistry*

2D Sketcher

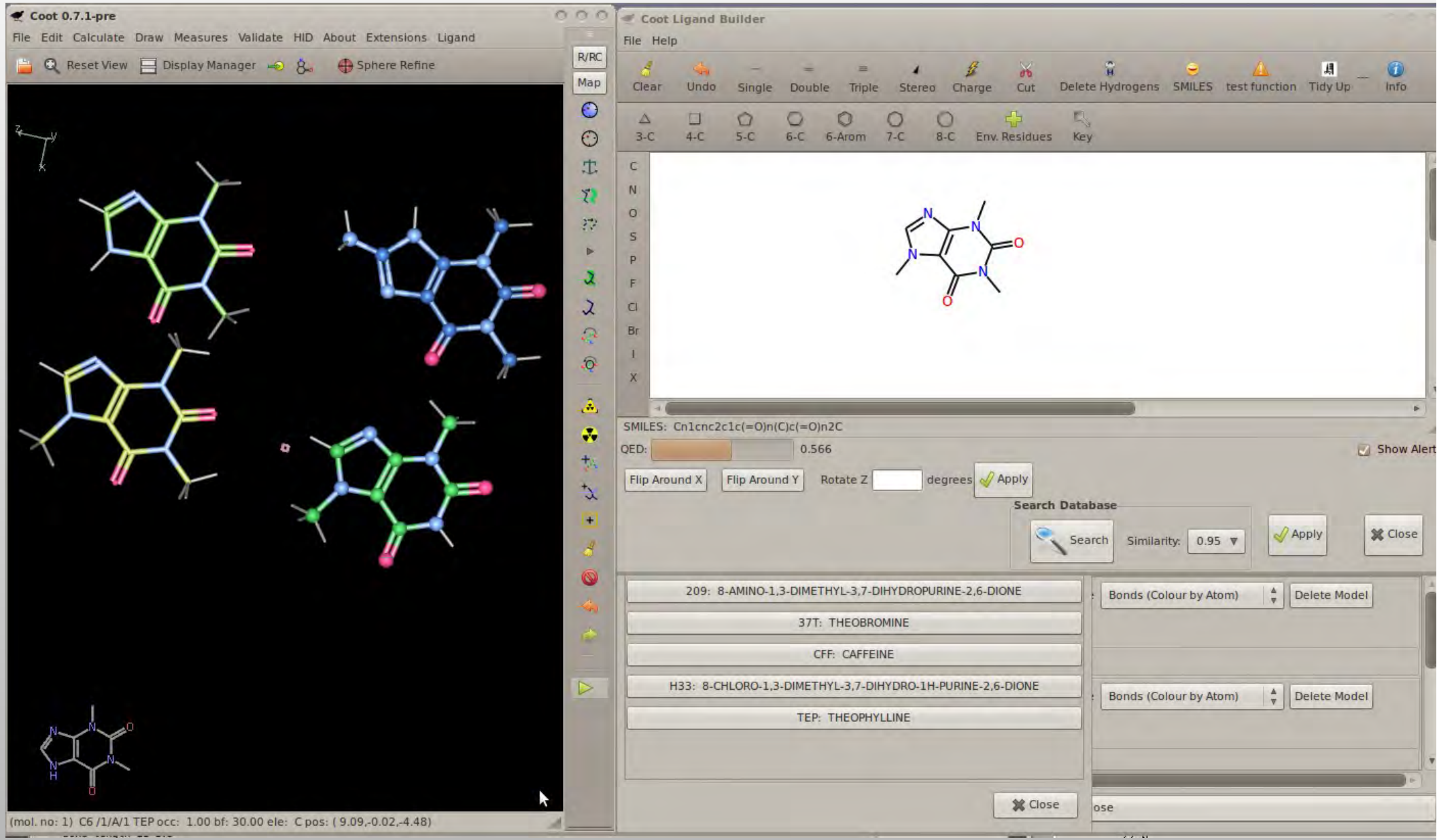
- QED score



Silicos-it's Biscu-it™

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

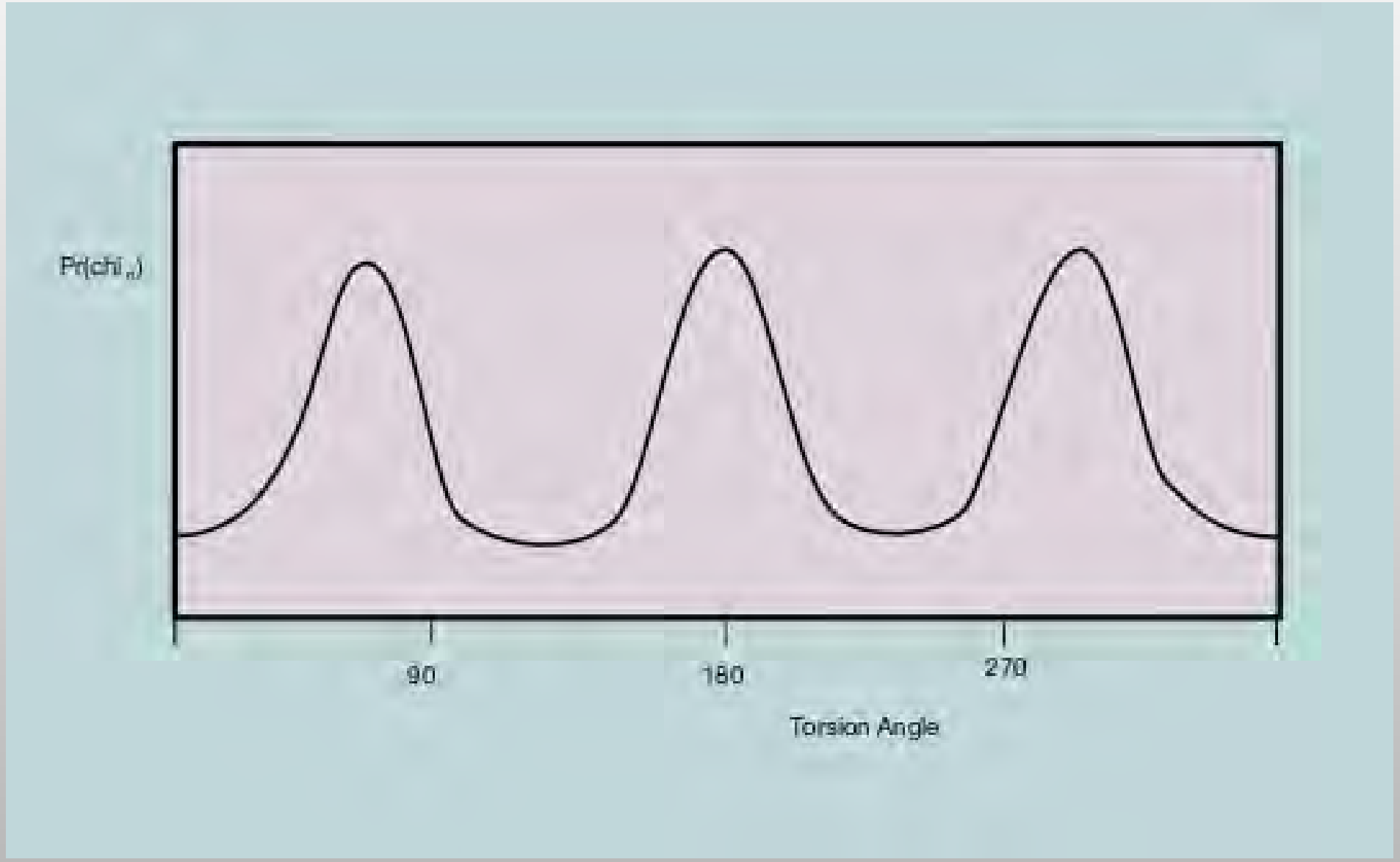
Ligand Utils – CCP4 SRS



REFMAC Monomer Library
chem_comp_tor

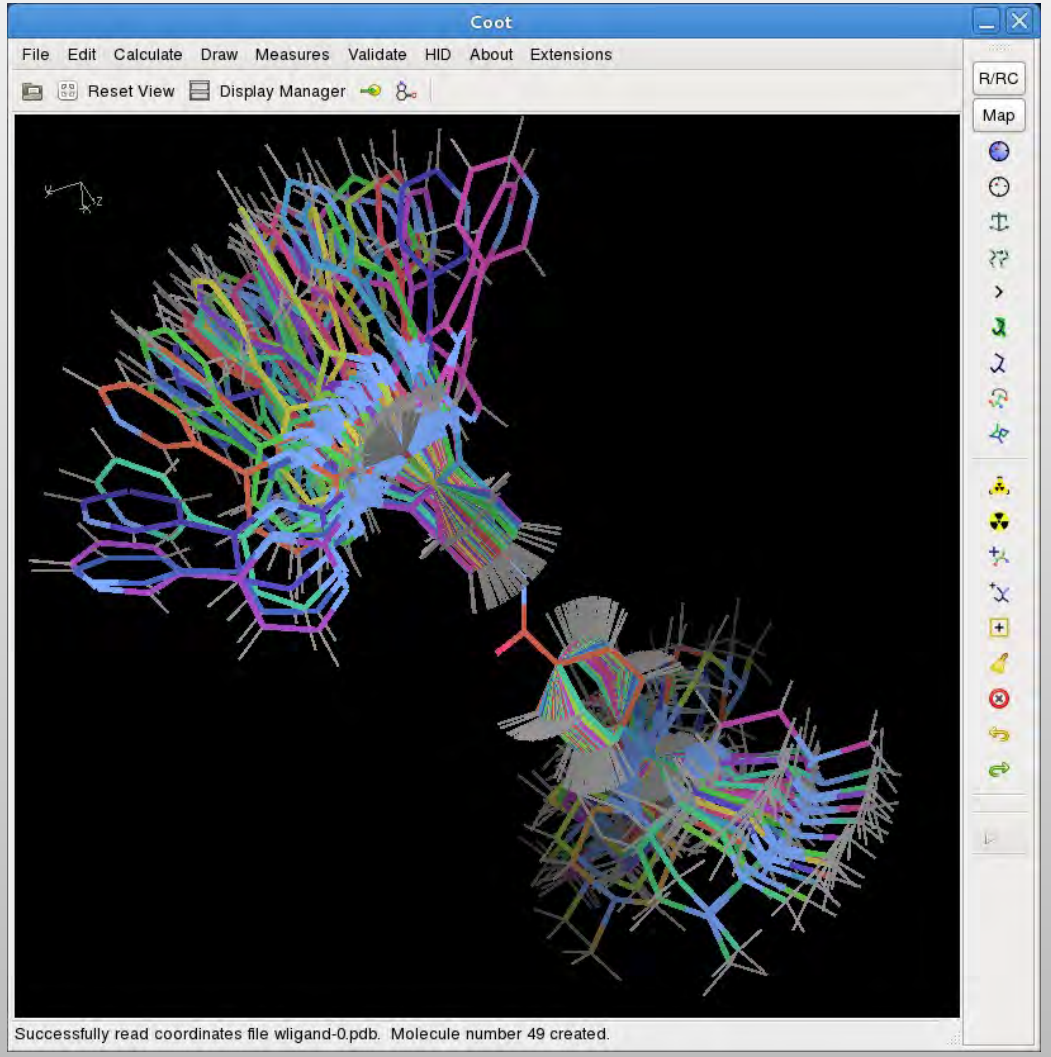
```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB        60.005   20.000   1
 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

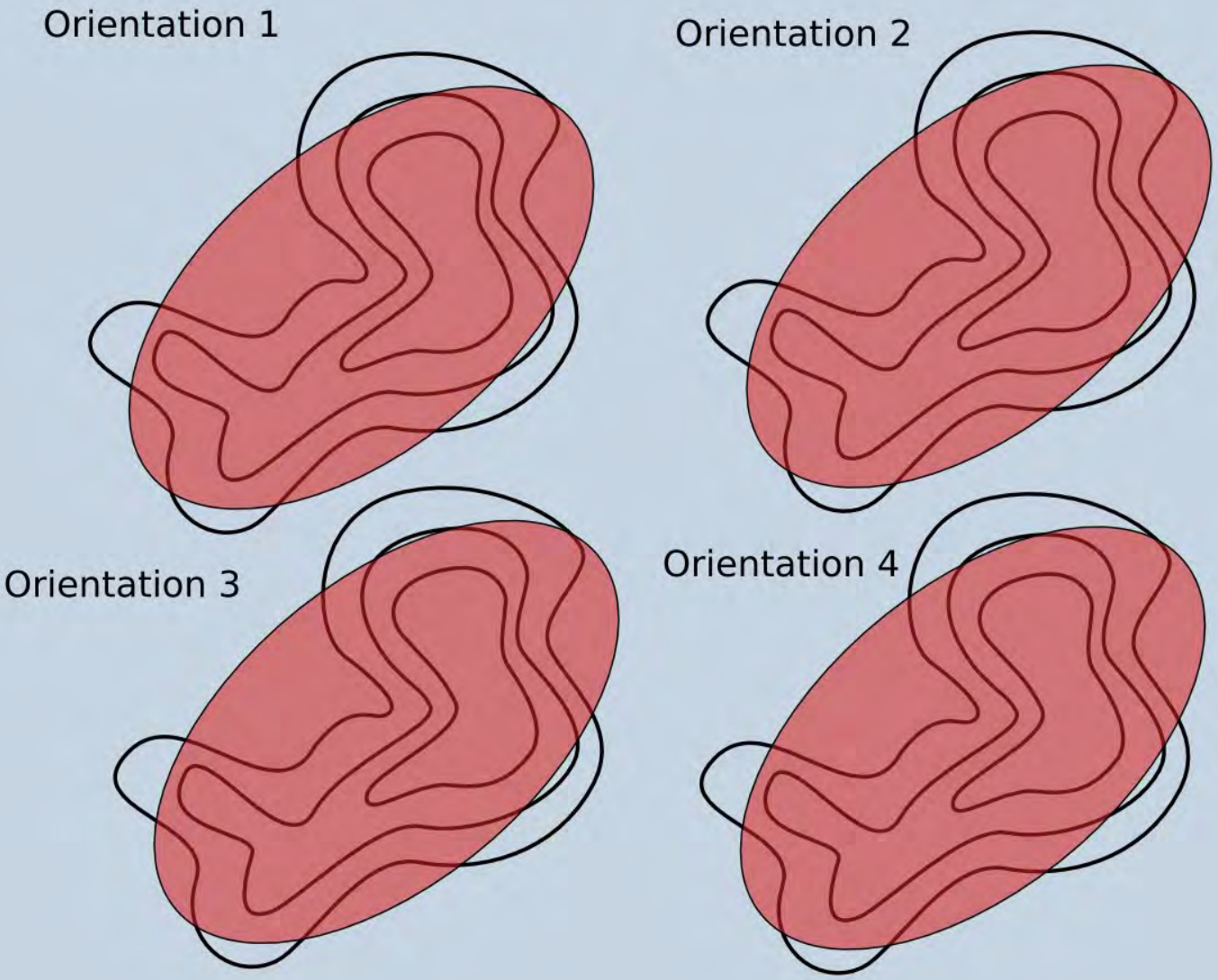


Conformer Generation

Non-Hydrogen
Non-CONST
Non-Ring



Orienting the Ligand



Orienting the Ligand