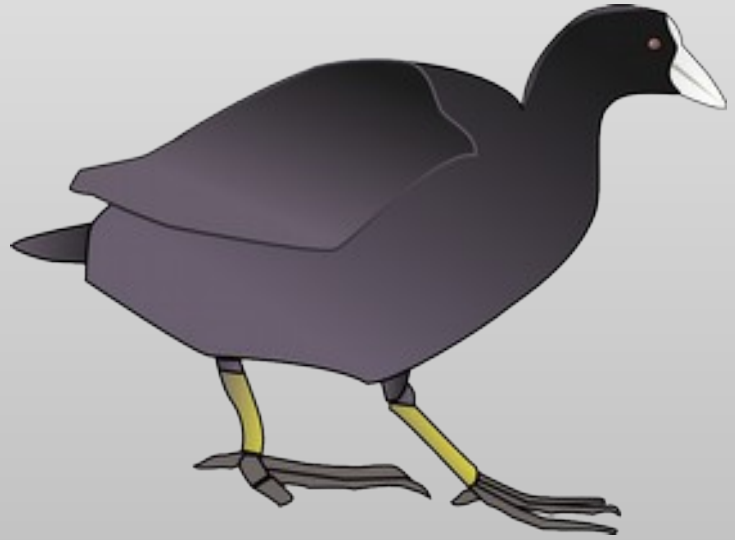
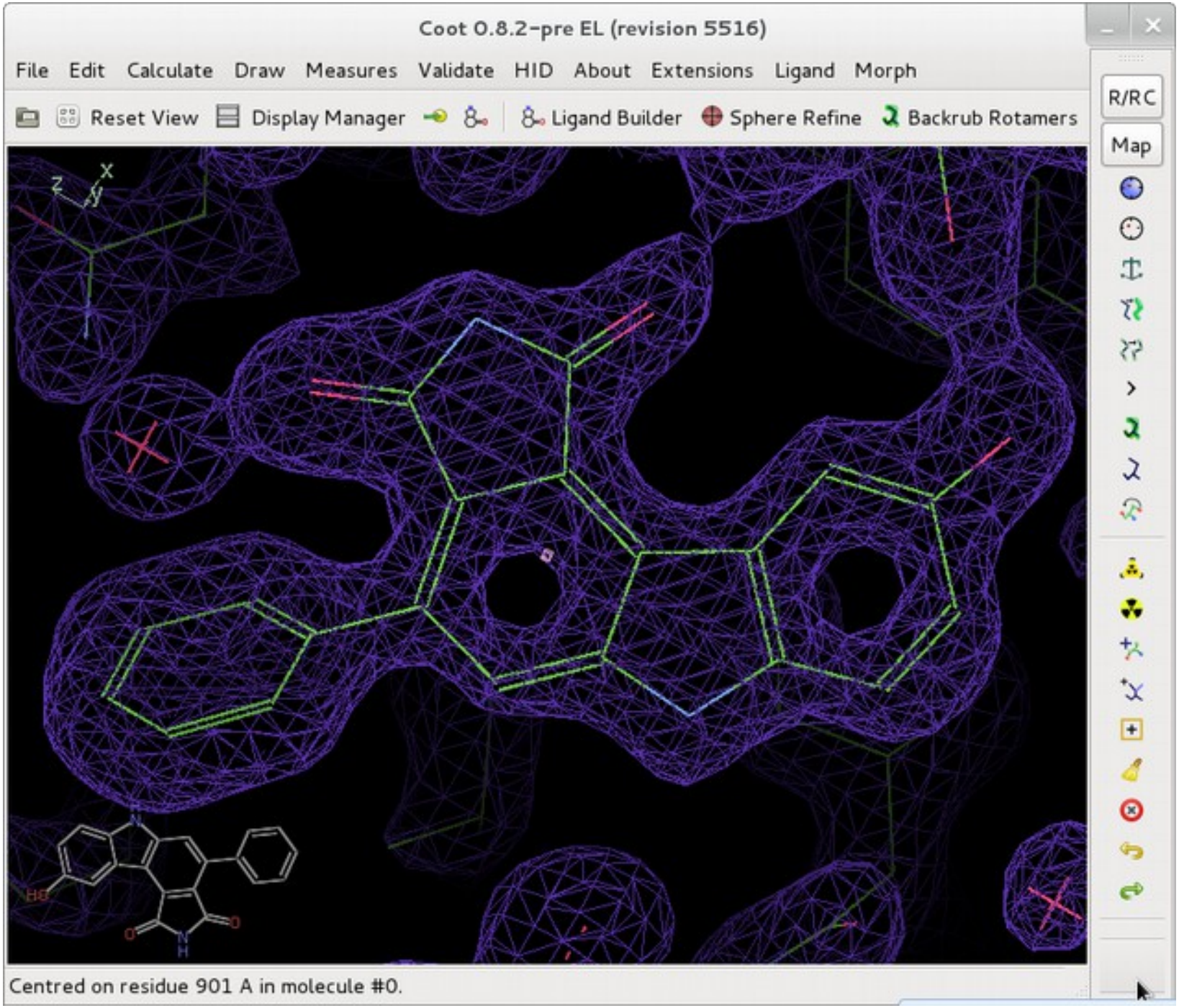
Validating Ligands Using *Coot*

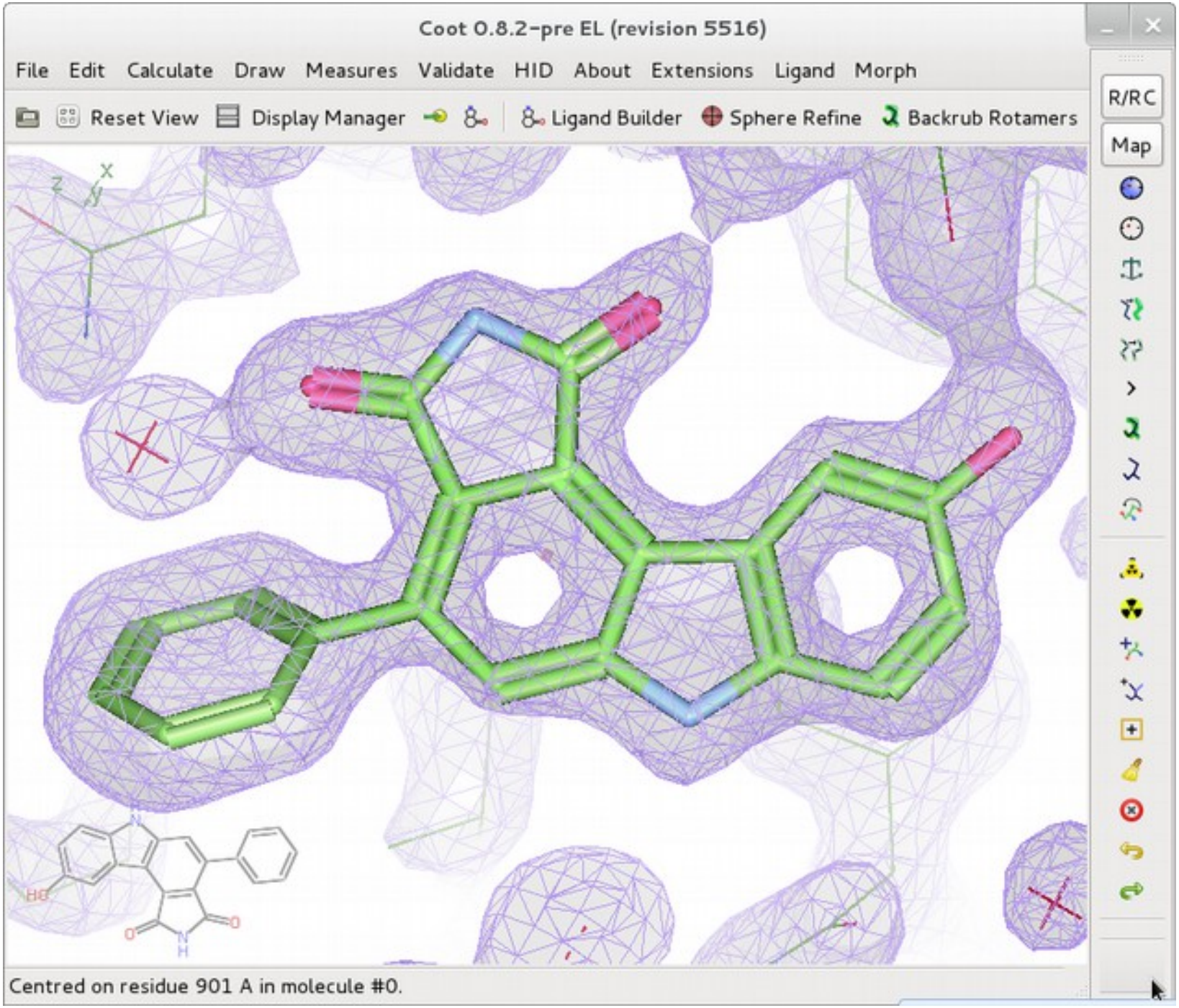
with a bit of carbohydrate building

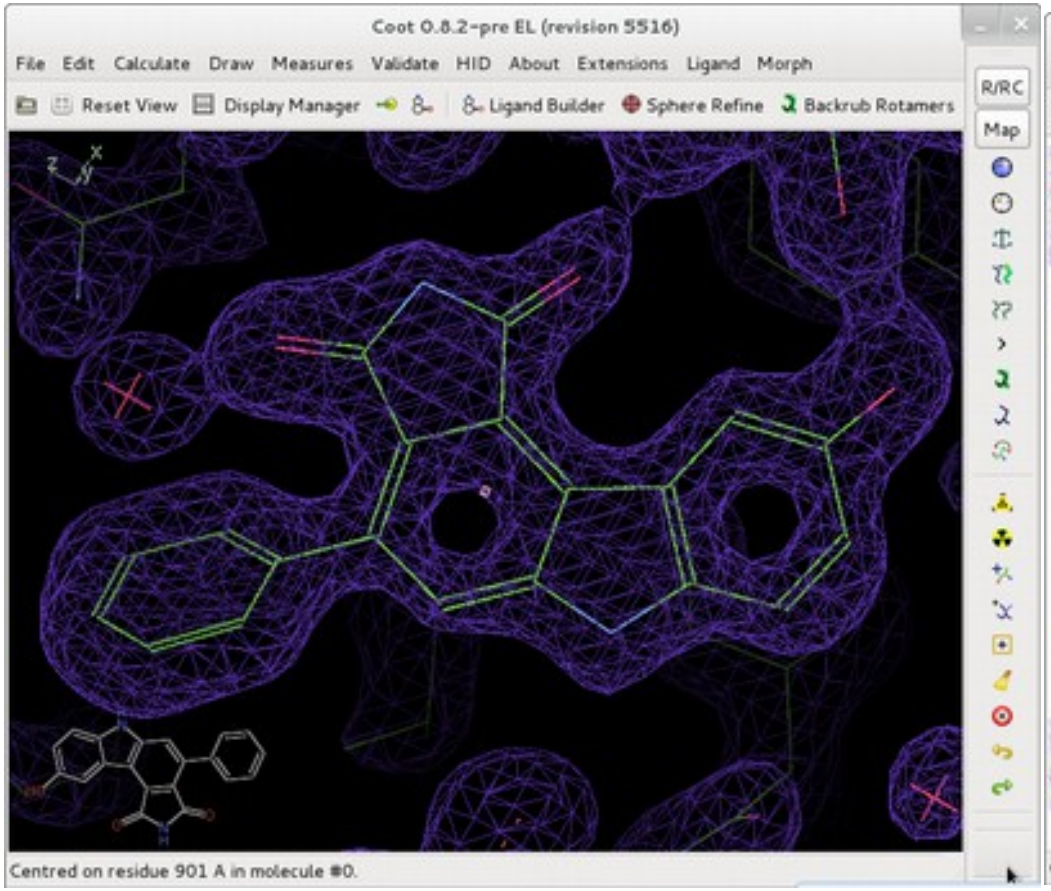


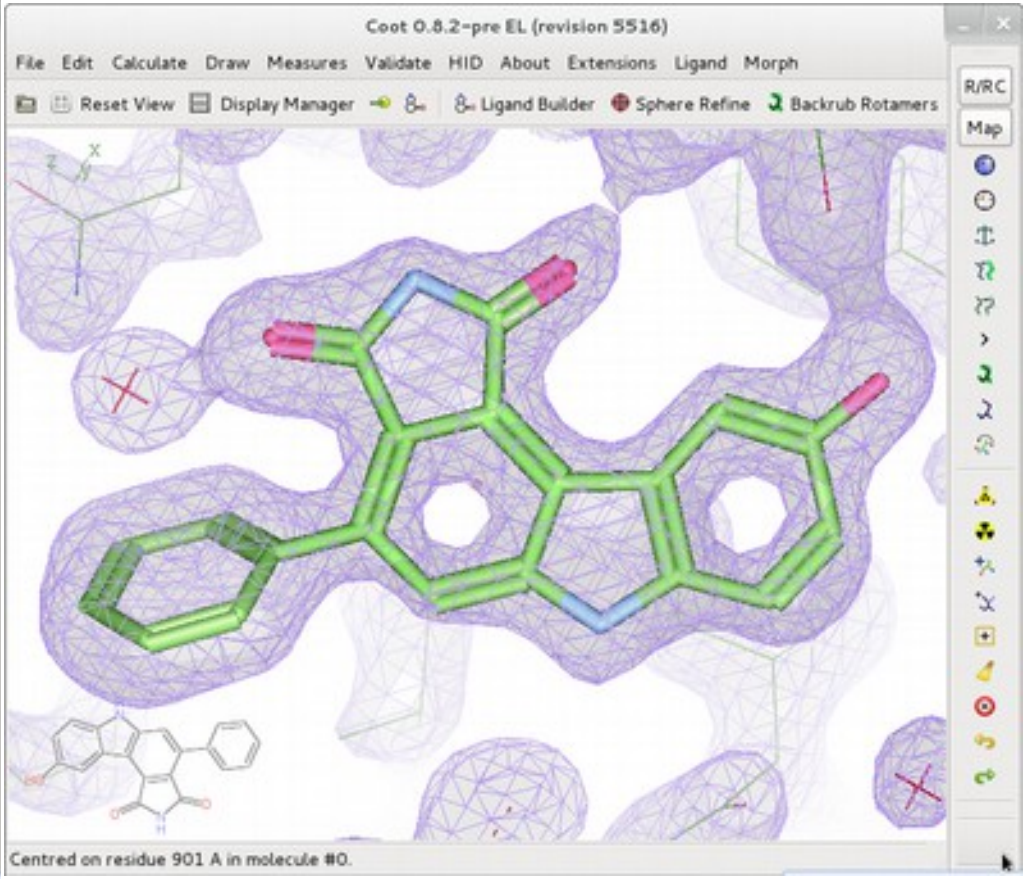
Paul Emsley

Nov 2018


Coot 0.8.2-pre EL (revision 5516)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Ligand Builder
Sphere Refine
Backrub Rotamers
R/RC
Map
Centred on residue 901 A in molecule #0.

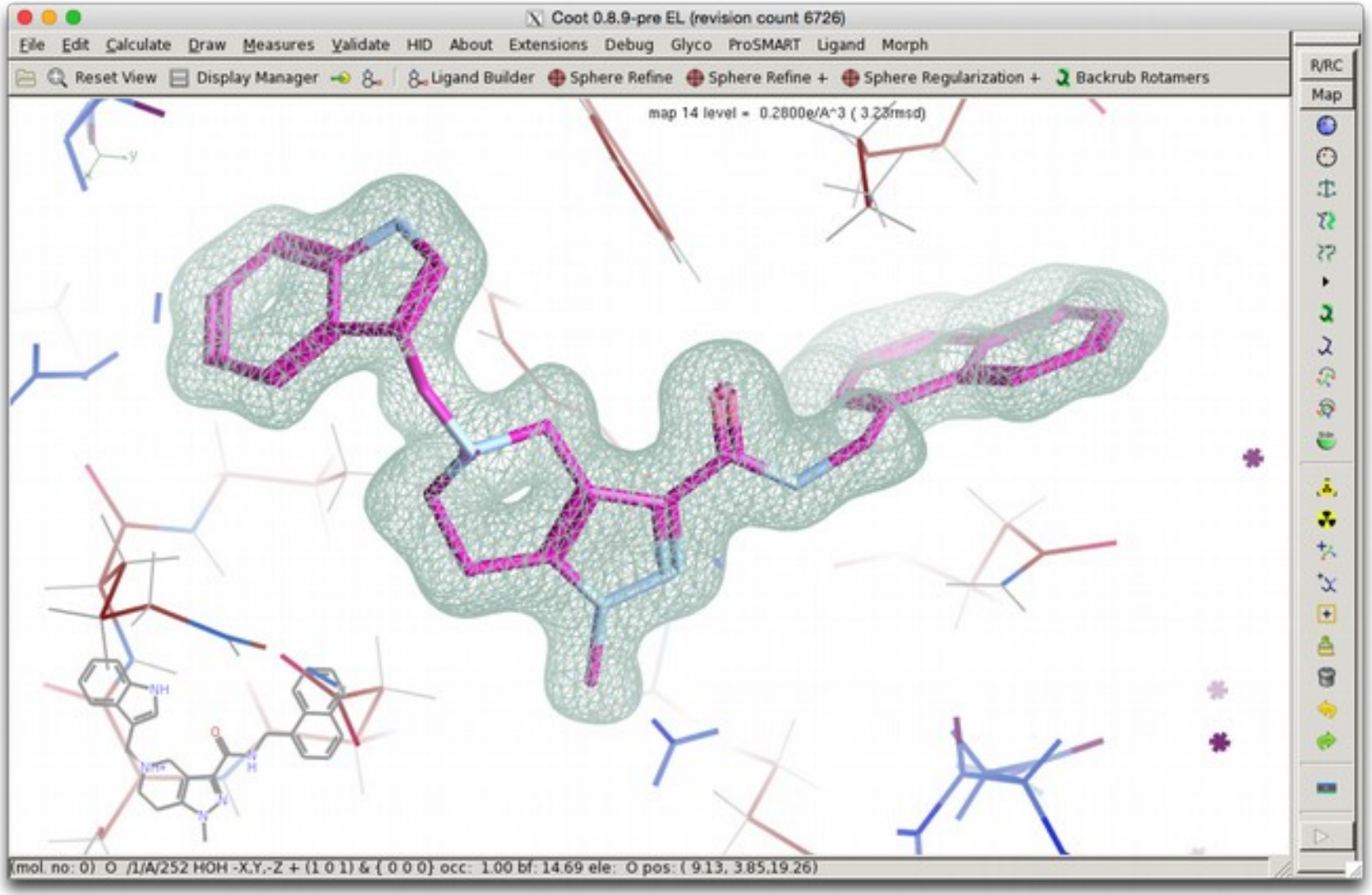

Coot 0.8.2-pre EL (revision 5516)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Ligand Builder
Sphere Refine
Backrub Rotamers
R/RC
Map
Centred on residue 901 A in molecule #0.





“So why don't you make a button to that for us, Paul?”

Ligands Representation in *Coot*

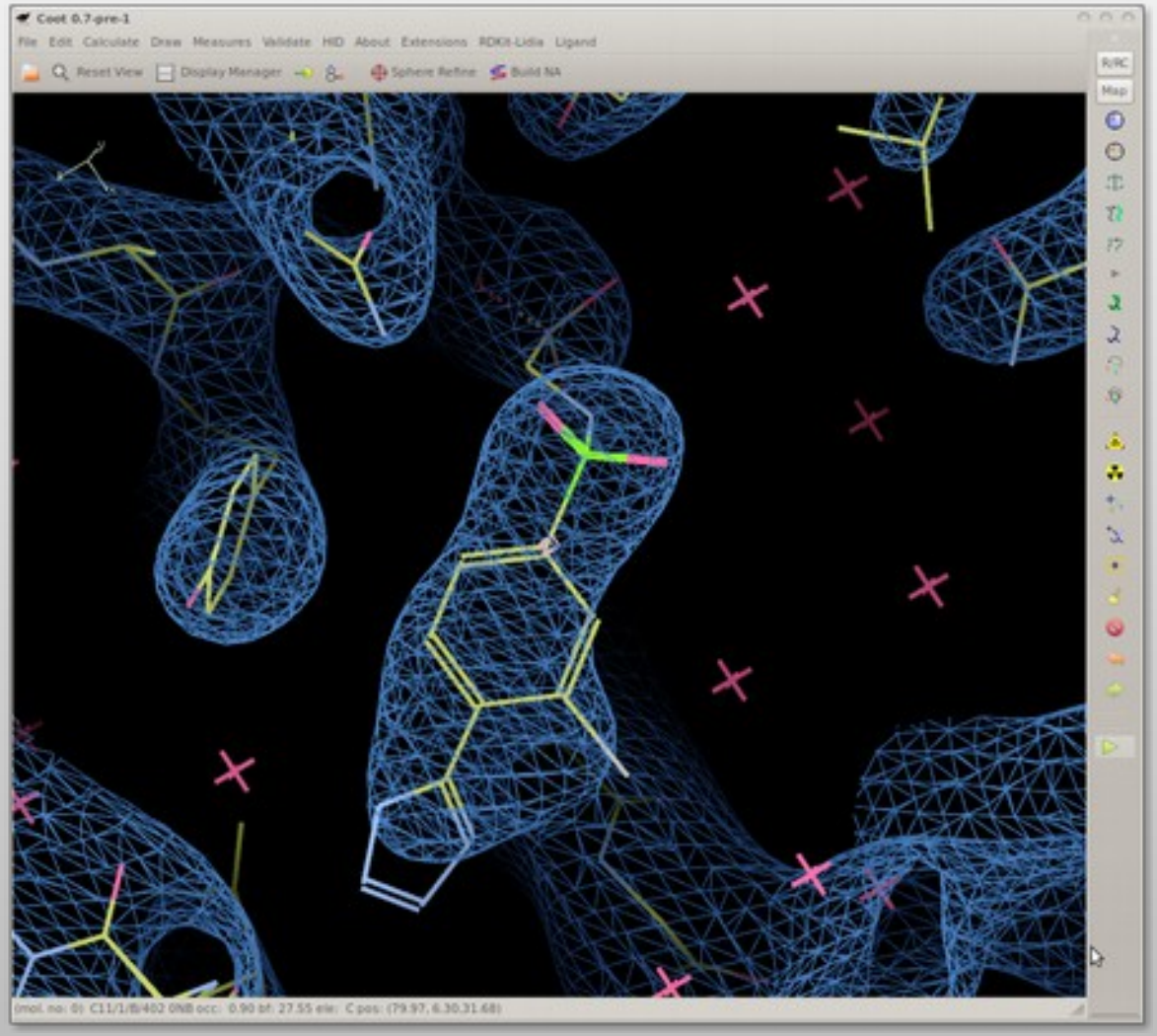


How To Make Electron Density Pictures with *Coot*

- Standard Screenshot
  - Density is too dark for high illumination seminar rooms
- Edit ⟶ Background Colour ⟶ White
- Draw ⟶ Additional Representations ⟶

  Ball & Stick ⟶ Add Representation
- Resample the map (1.8 ⟶ 2.2)
- Display Manager ⟶ Properties ⟶ Cut Glass ⟶ 10%
  - Map Colour ⟶ "Light Blue/Grey"

Ligand and Density...



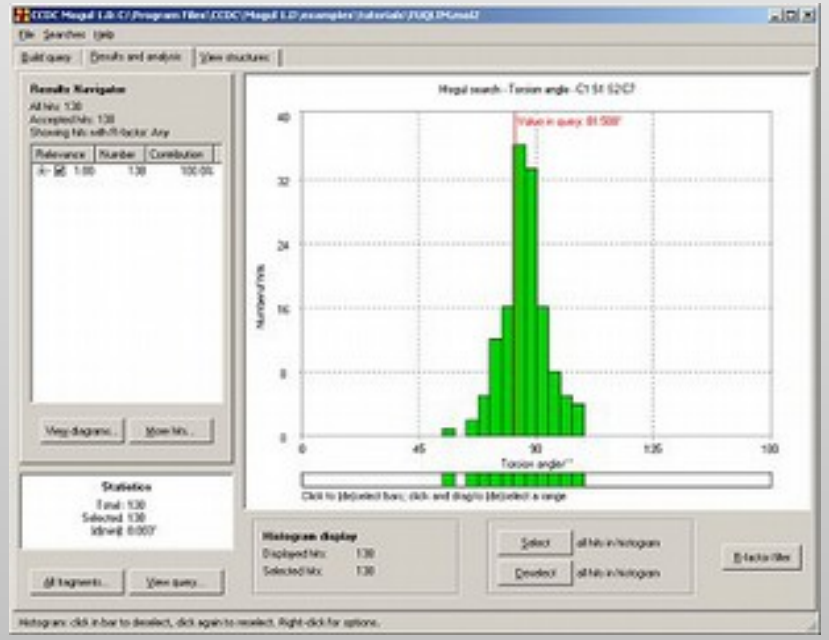
Protein-ligand complex models are often a result of subjective interpretation

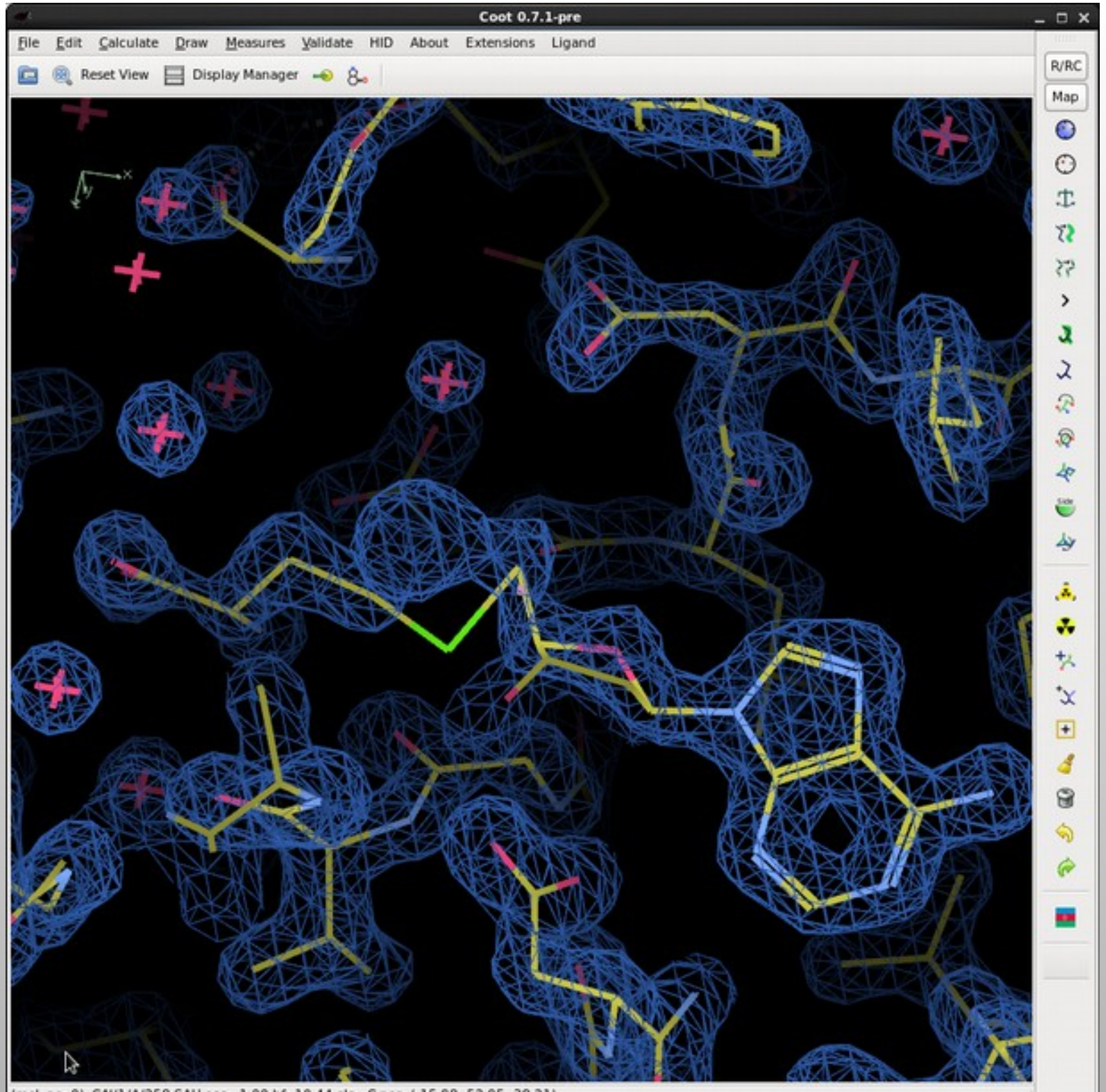
Scoring Protein-Ligand Complexes

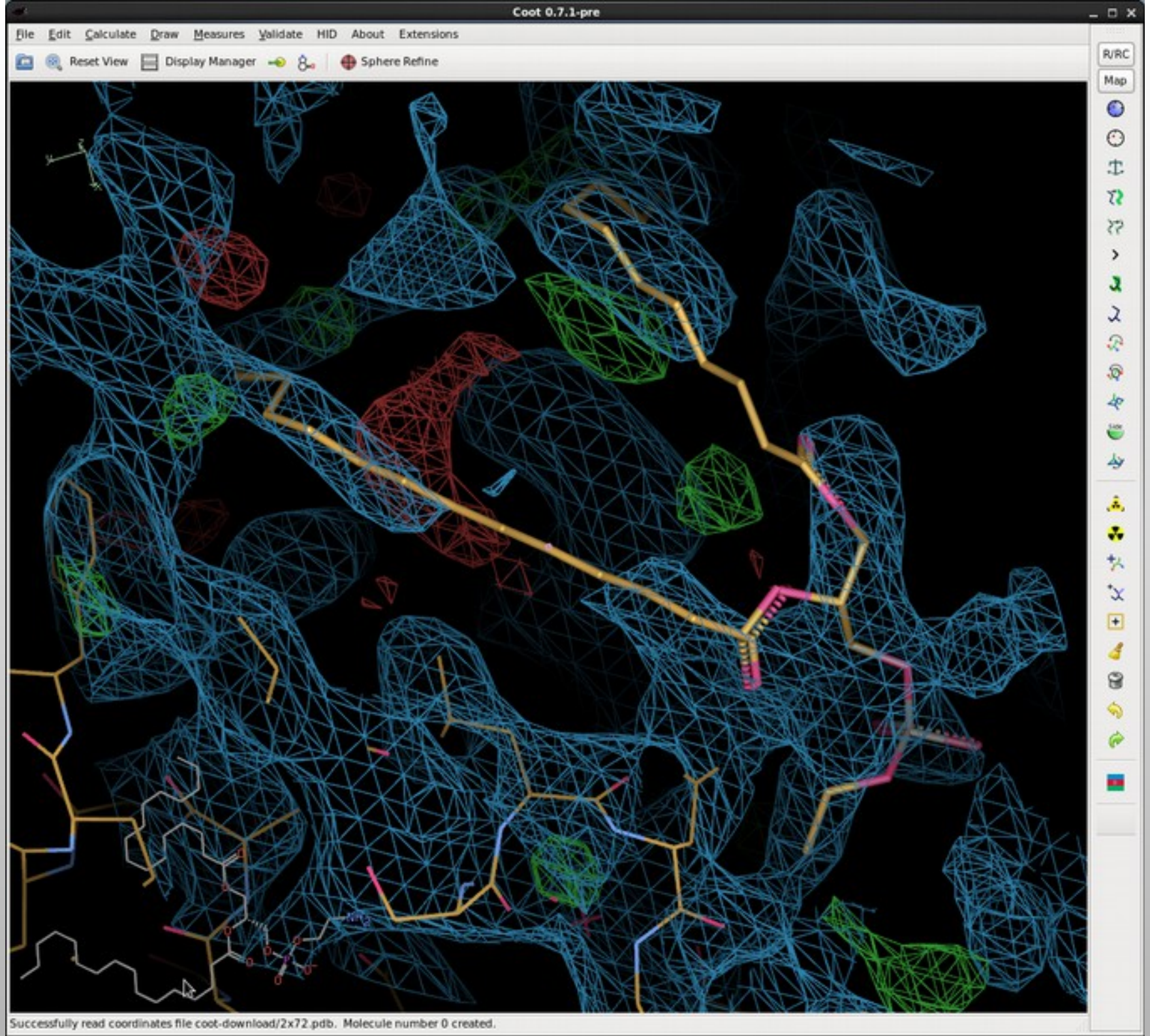
- Score all PDB protein-ligand complexes
  - No covalent link to protein
  - No alt confs
  - Hetgroups with more than 6 atoms
- Score:
  - Correlation of maps: omit vs calculated
    - around the ligand
  - Mogul distortion
    - z-worst
  - Clash-score
    - c.f. Molprobity tool

Assessing Ligand Geometry Accuracy

- 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.

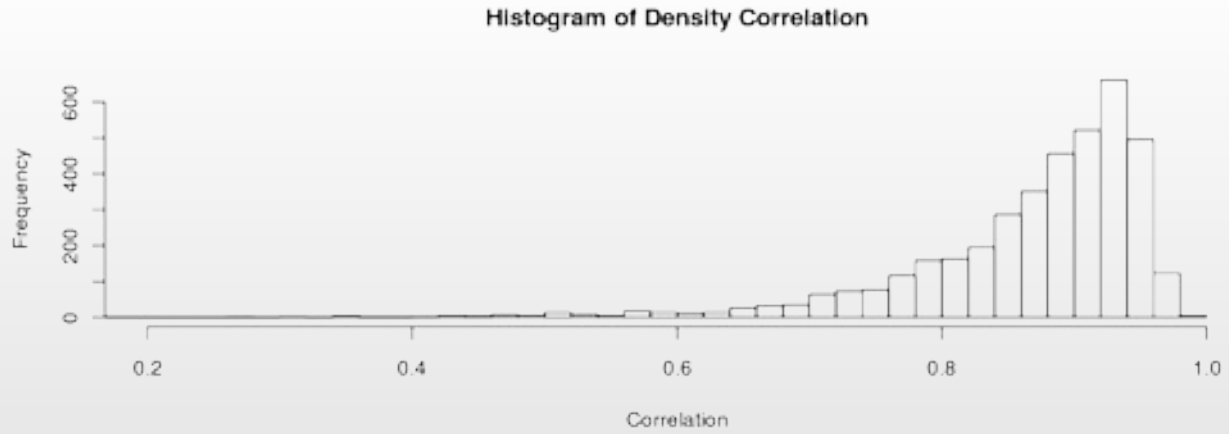


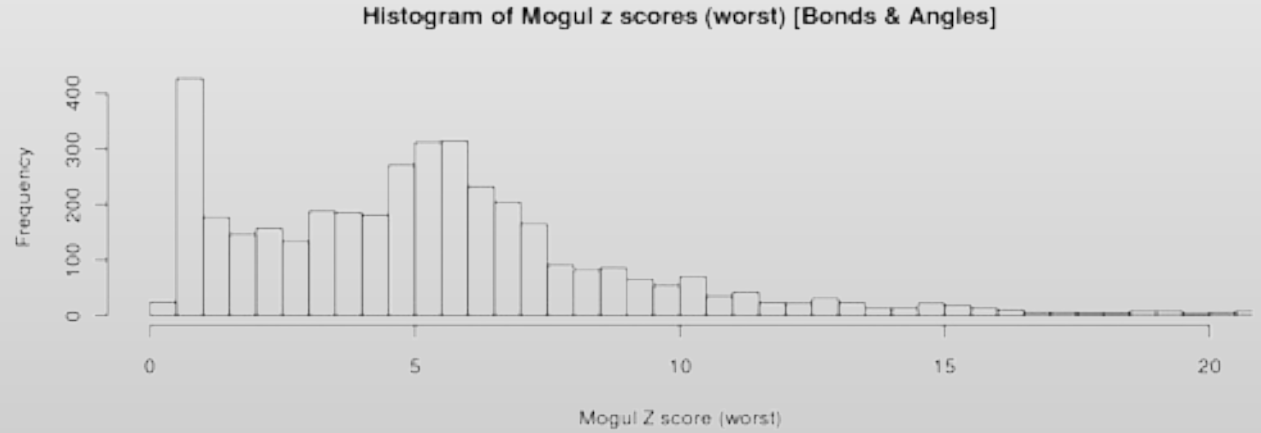

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
R/RC
Map

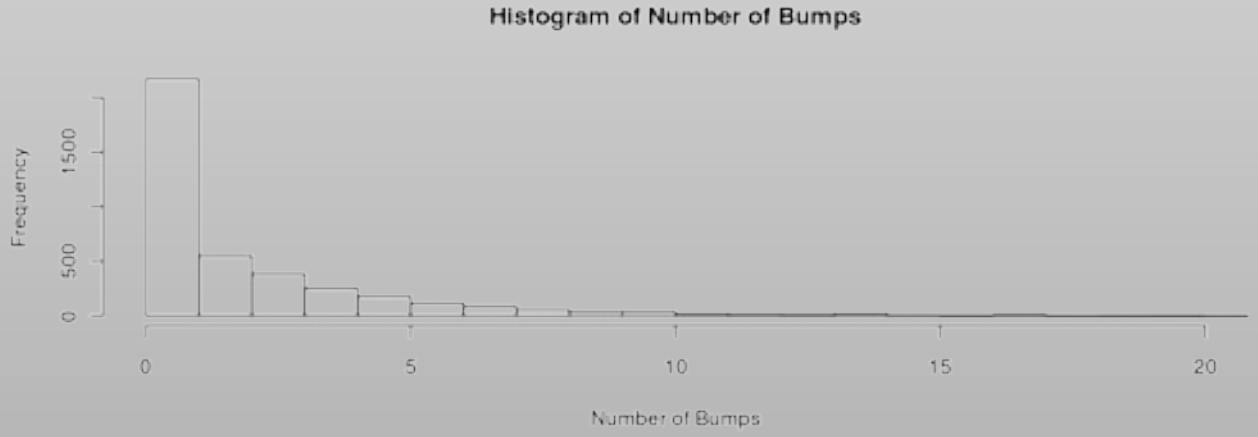

Coot 0.7.1-pre
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
Sphere Refine
R/RC
Map
Successfully read coordinates file coot-download/2x72.pdb. Molecule number 0 created.

Ligand Scoring

Preliminary results & conclusions...

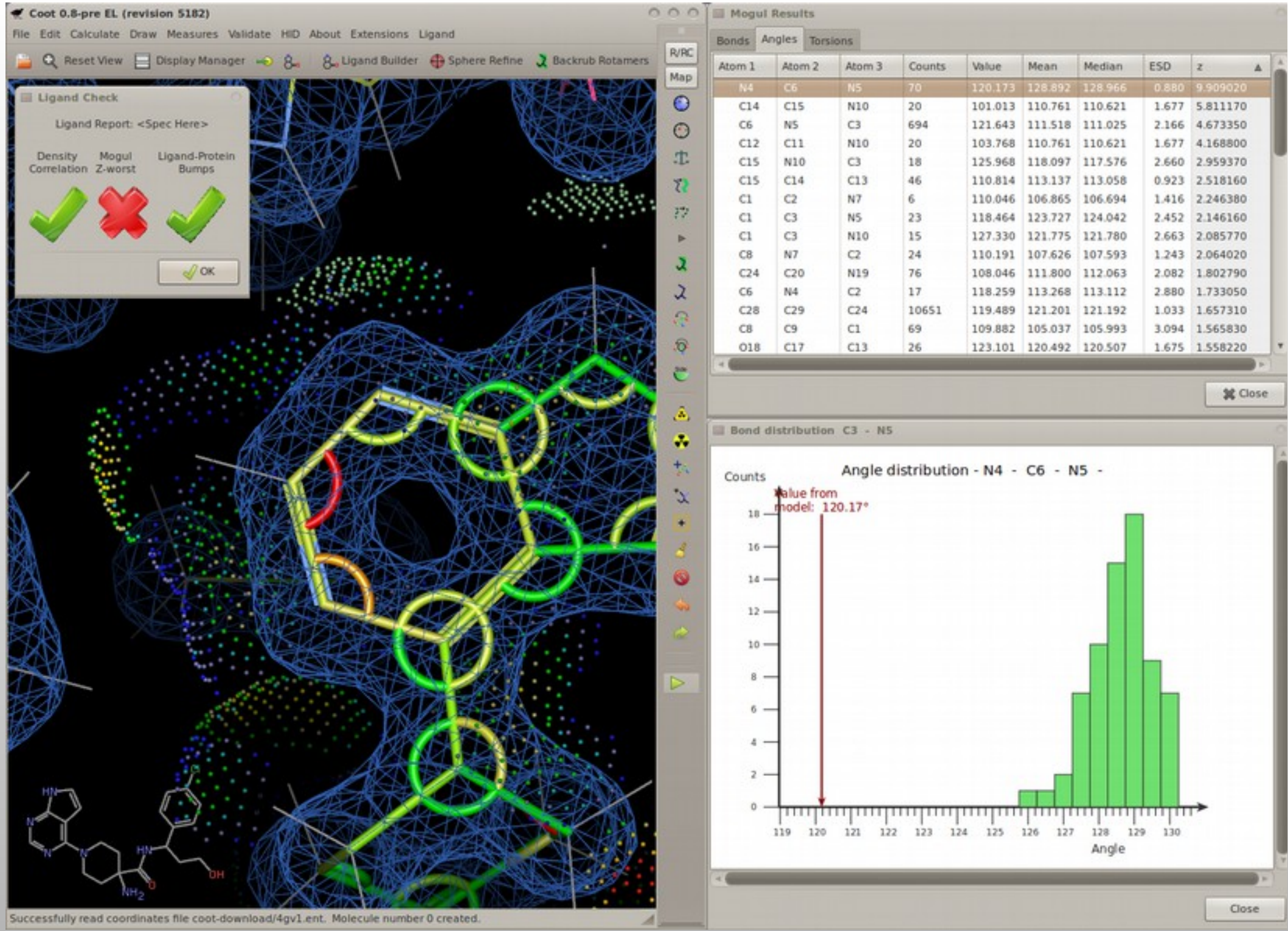

Histogram of Density Correlation
Frequency
0
200
400
600
0.2
0.4
0.6
0.8
1.0
Correlation


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?


Coot 0.8-pre EL (revision 5182)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Ligand Builder
Sphere Refine
Backrub Rotamers
R/RC
Map
Ligand Check
Ligand Report: <Spec Here>
Density Correlation
Mogul Z-worst
Ligand-Protein Bumps
OK
Mogul Results
Bonds
Angles
Torsions
Atom 1 Atom 2 Atom 3 Counts Value Mean Median ESD z
N4 C6 N5 70 120.173 128.892 128.966 0.880 9.909020
C14 C15 N10 20 101.013 110.761 110.621 1.677 5.811170
C6 N5 C3 694 121.643 111.518 111.025 2.166 4.673350
C12 C11 N10 20 103.768 110.761 110.621 1.677 4.168800
C15 N10 C3 18 125.968 118.097 117.576 2.660 2.959370
C15 C14 C13 46 110.814 113.137 113.058 0.923 2.518160
C1 C2 N7 6 110.046 106.865 106.694 1.416 2.246380
C1 C3 N5 23 118.464 123.727 124.042 2.452 2.146160
C1 C3 N10 15 127.330 121.775 121.780 2.663 2.085770
C8 N7 C2 24 110.191 107.626 107.593 1.243 2.064020
C24 C20 N19 76 108.046 111.800 112.063 2.082 1.802790
C6 N4 C2 17 118.259 113.268 113.112 2.880 1.733050
C28 C29 C24 10651 119.489 121.201 121.192 1.033 1.657310
C8 C9 C1 69 109.882 105.037 105.993 3.094 1.565830
O18 C17 C13 26 123.101 120.492 120.507 1.675 1.558220
Close
Bond distribution C3 - N5
Angle distribution - N4 - C6 - N5 -
Counts
Value from model: 120.17°
Angle
Close
Successfully read coordinates file coot-download/4gv1.ent. Molecule number 0 created.

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

- “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

Generating Conformers

- Using restraint information...

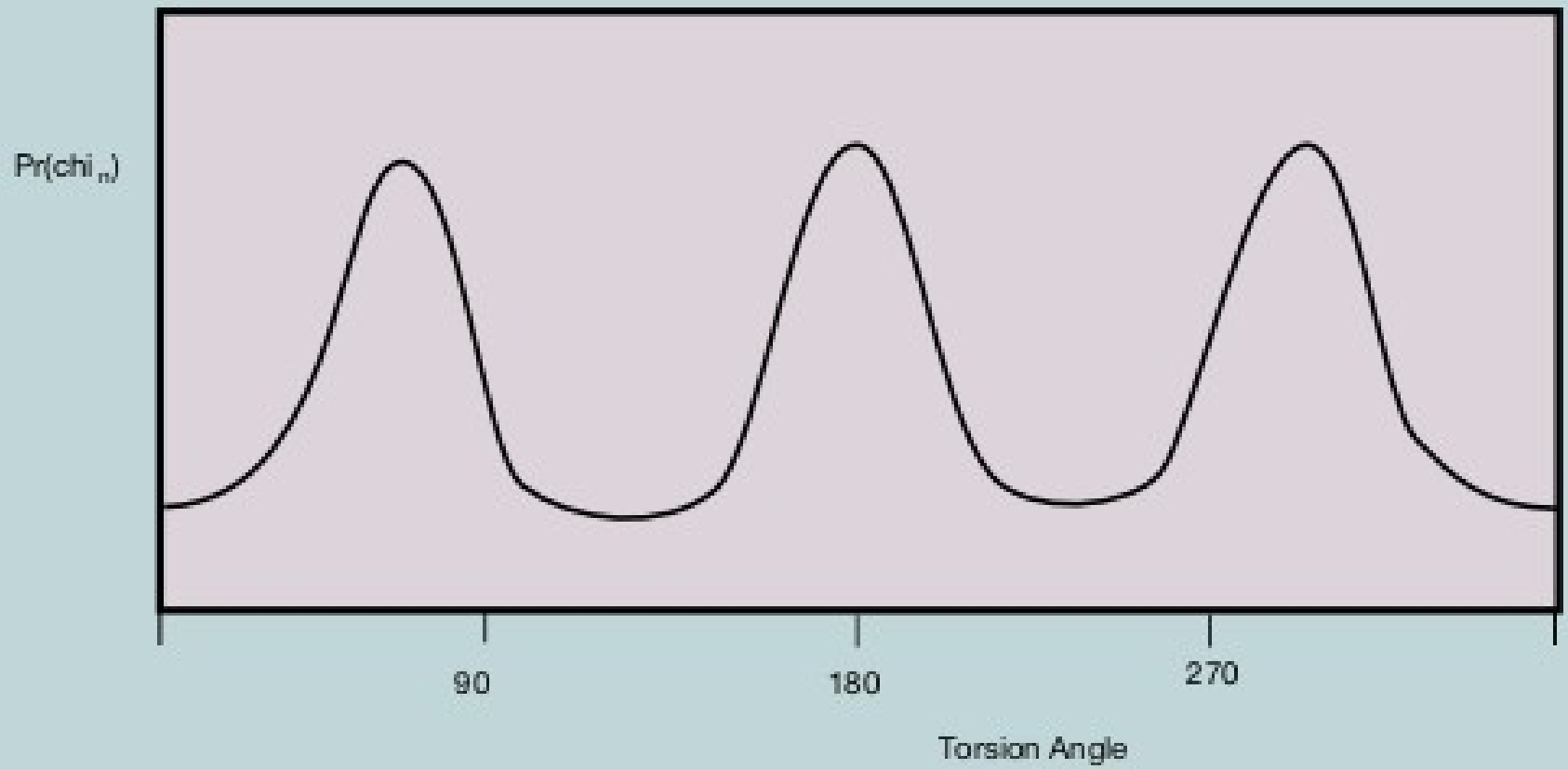
REFMAC Monomer Library
chem_comp_bond

```
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

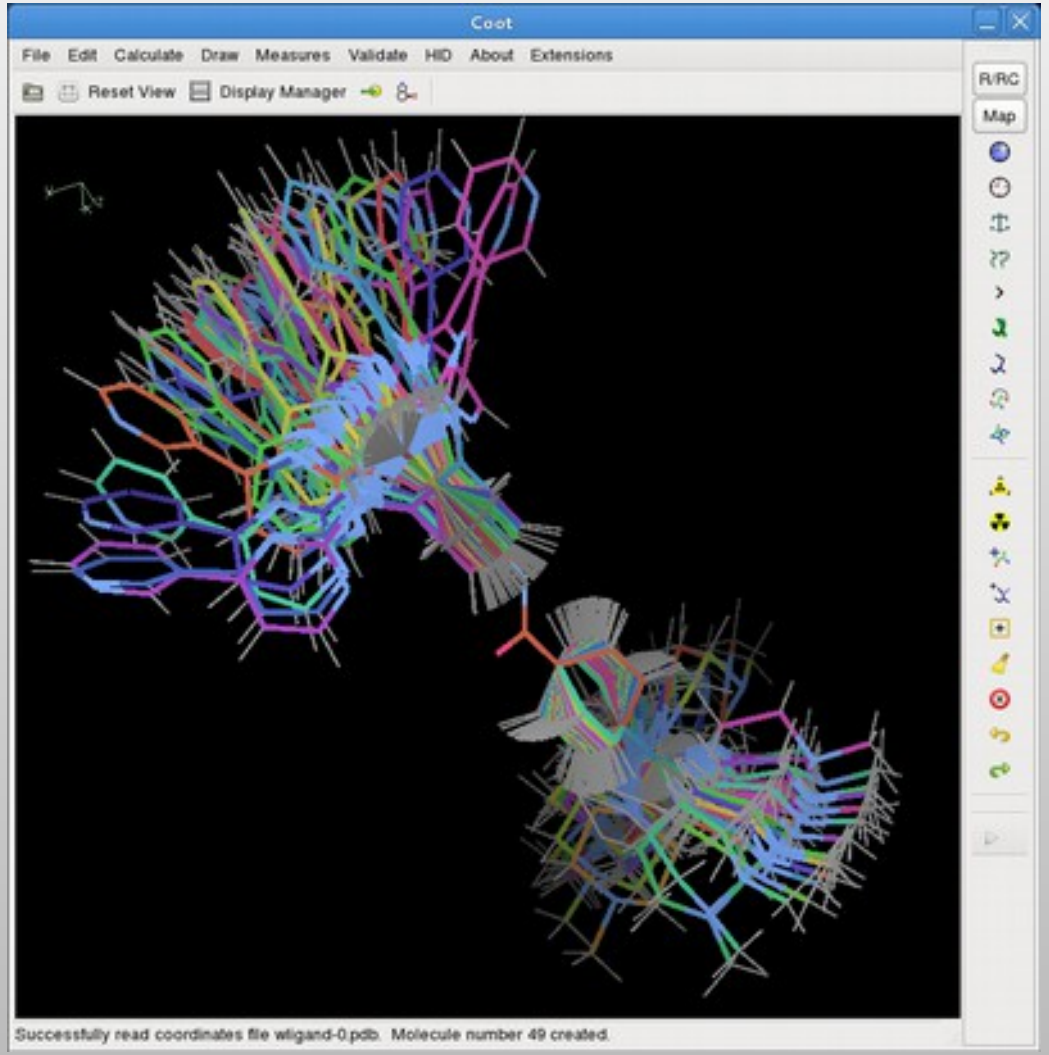
```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB        60.005   20.000   1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

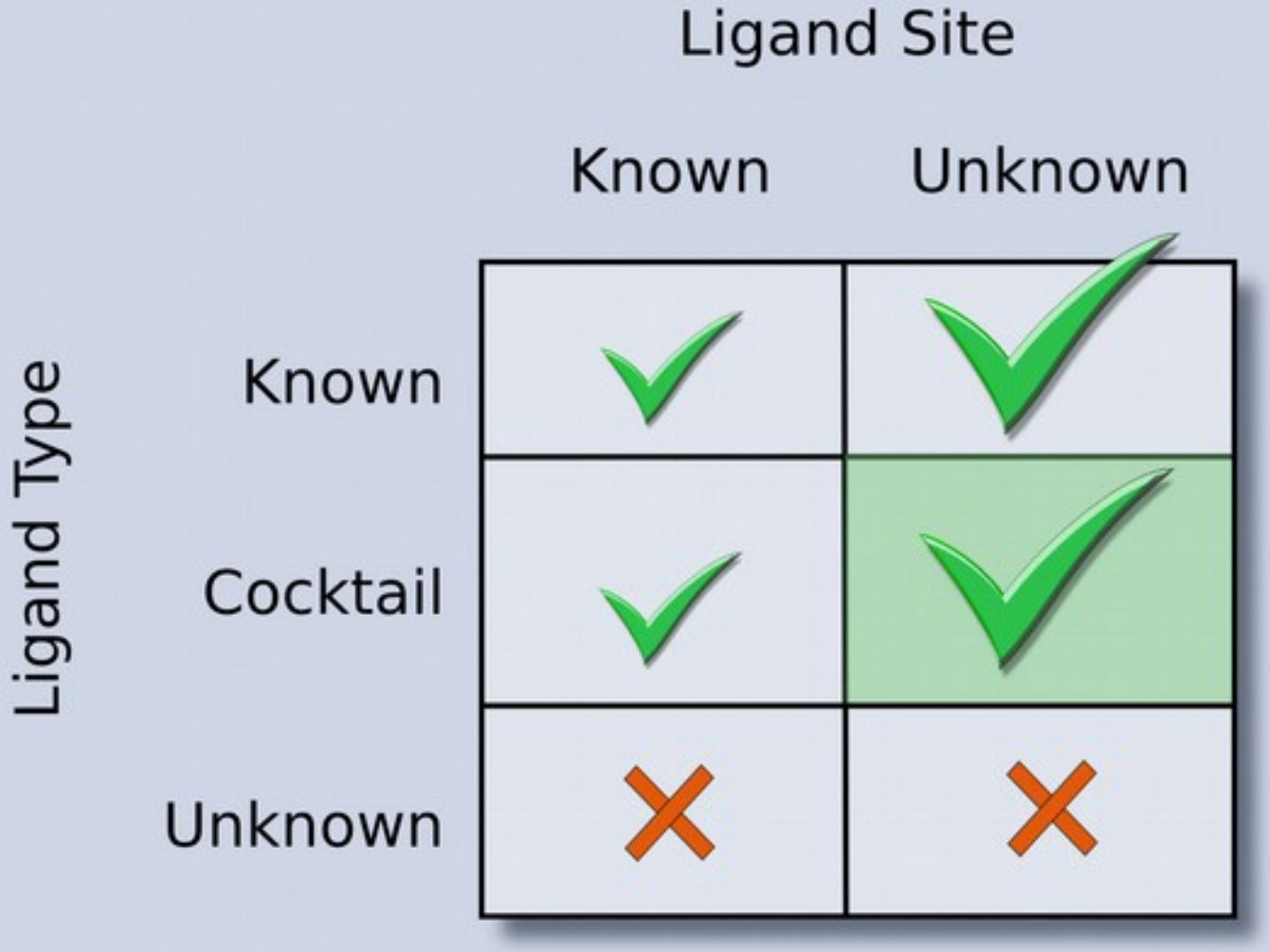


Conformer Generation

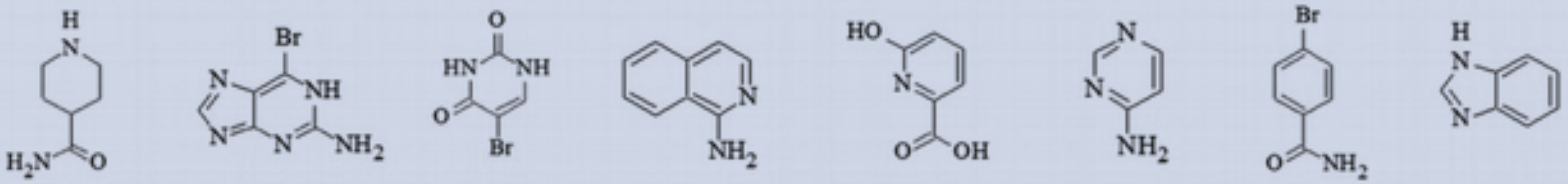
Non-Hydrogen
Non-CONST
Non-Ring



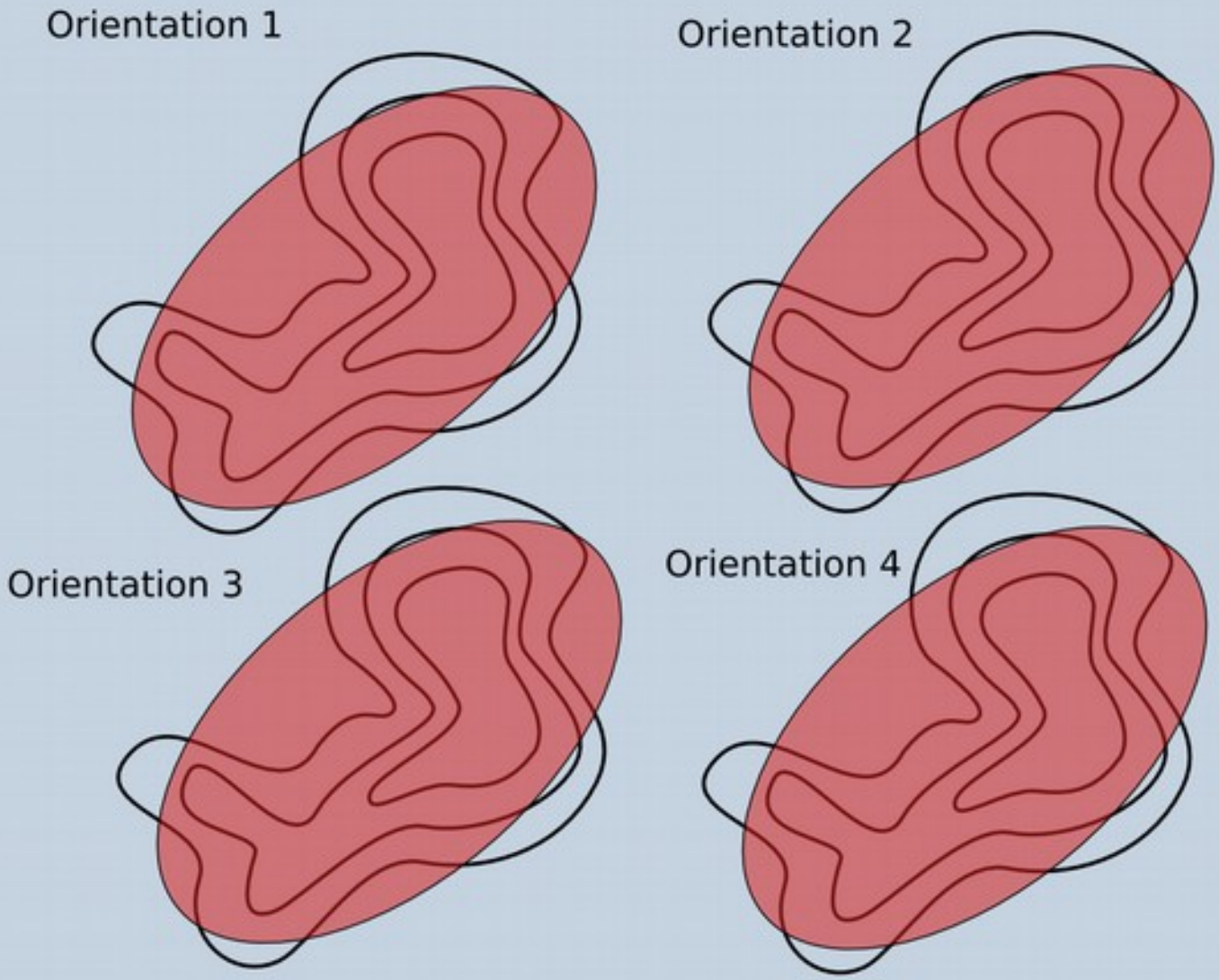
Fitting Ligands



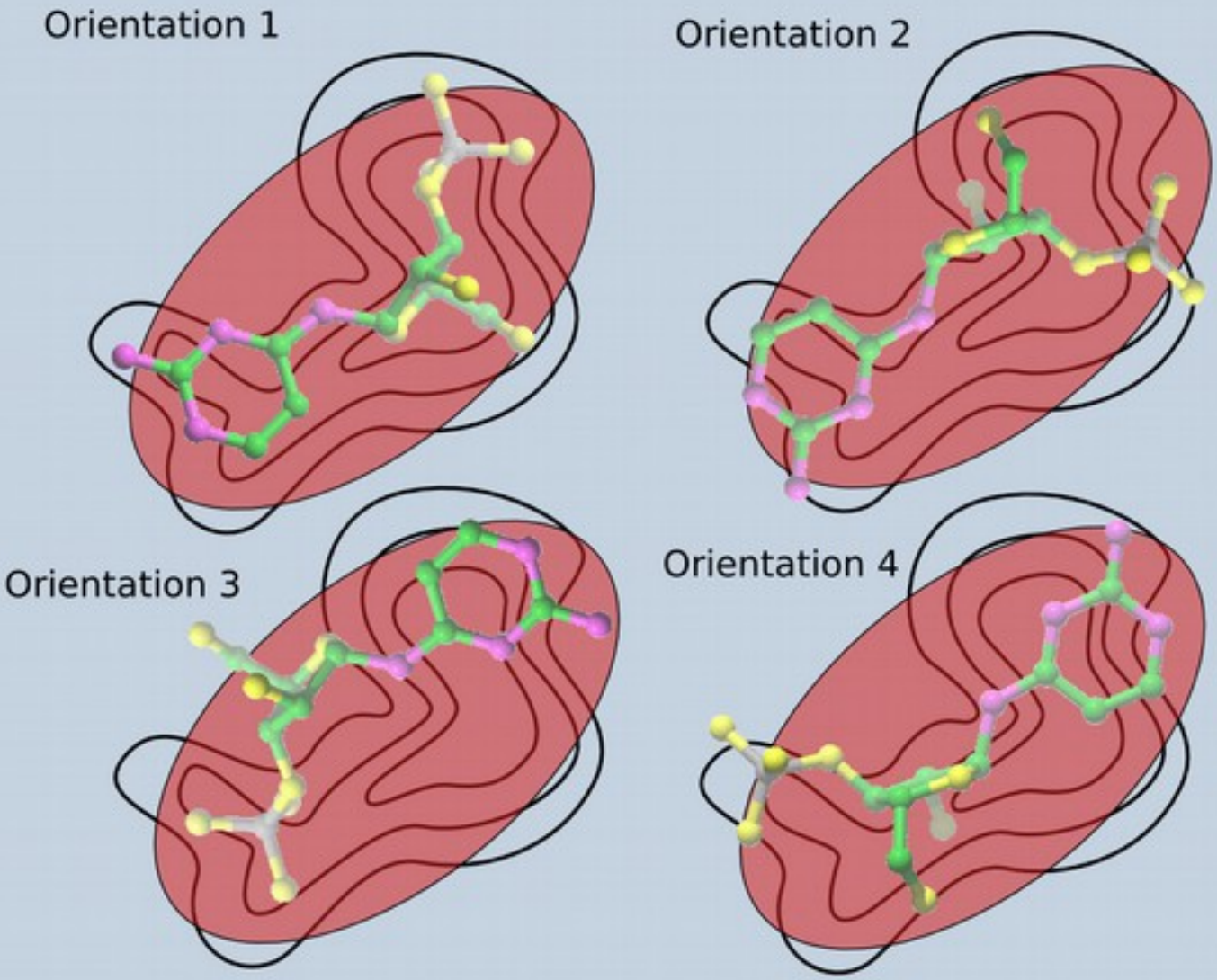
Cocktail Examples



Orienting the Ligand



Orienting the Ligand

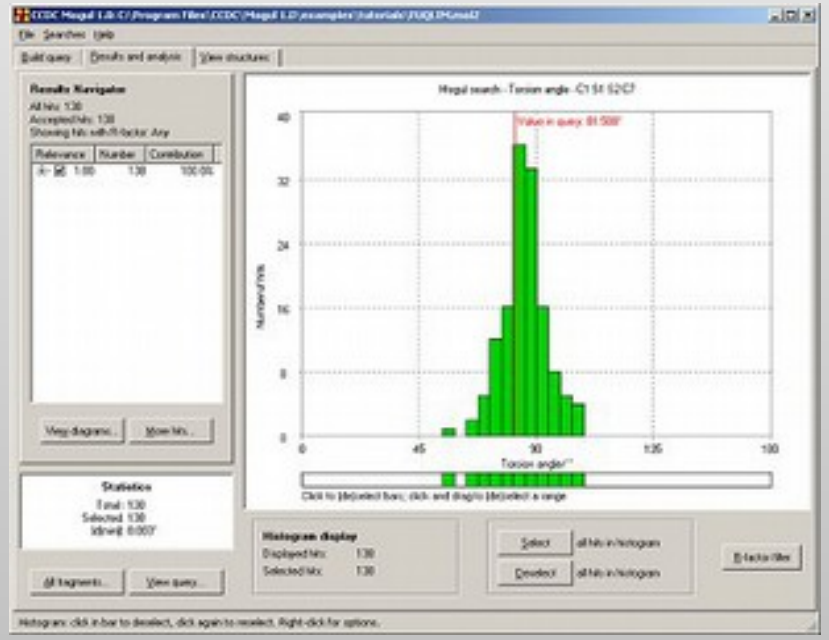


Ligand Validation

- 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)

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

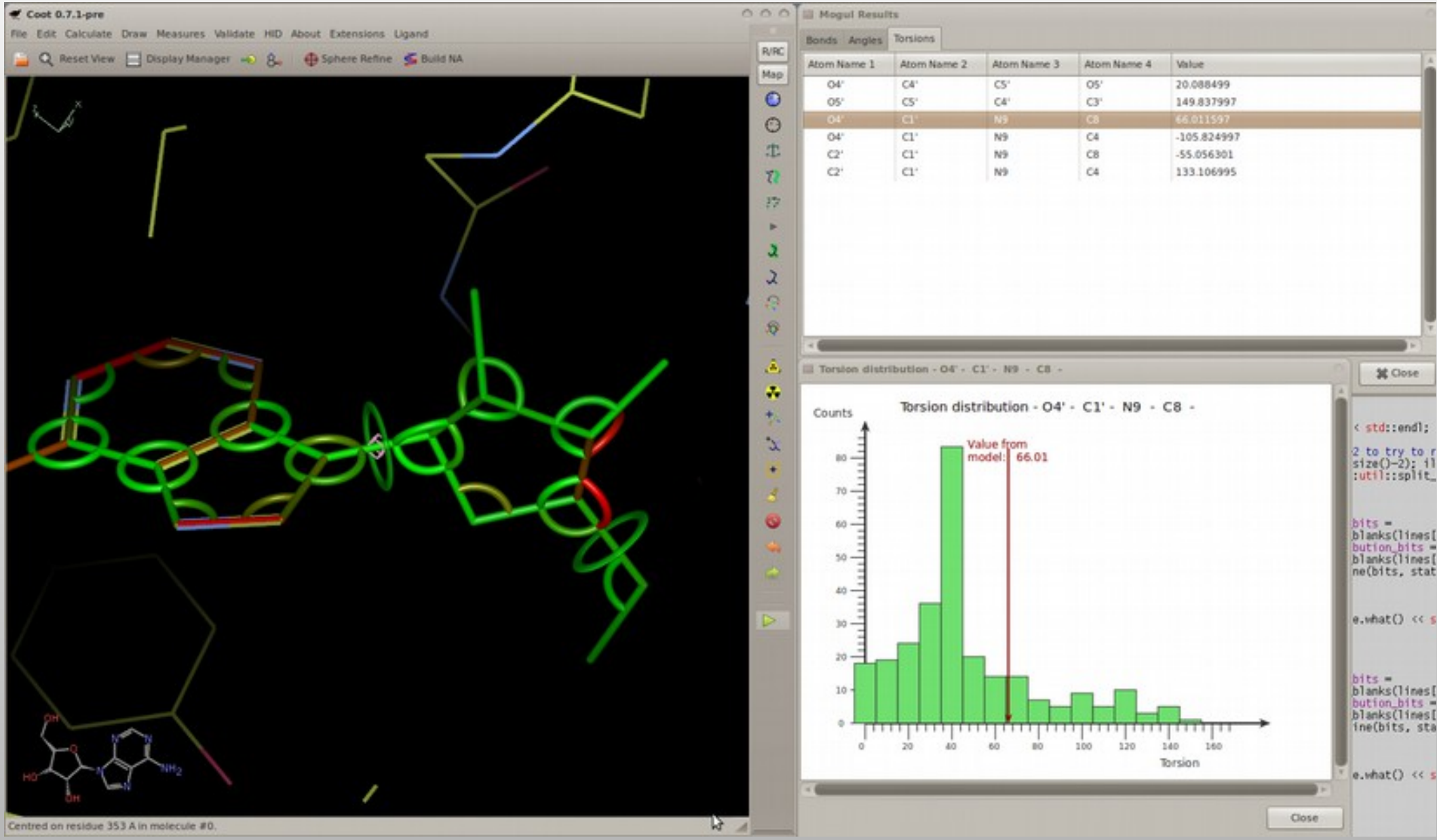
- Perfect refinement with incorrect parameters → distorted structure
- CSD's Mogul
- This time:
  - Display and Interactive



Example Coot Ligand Distortion Score

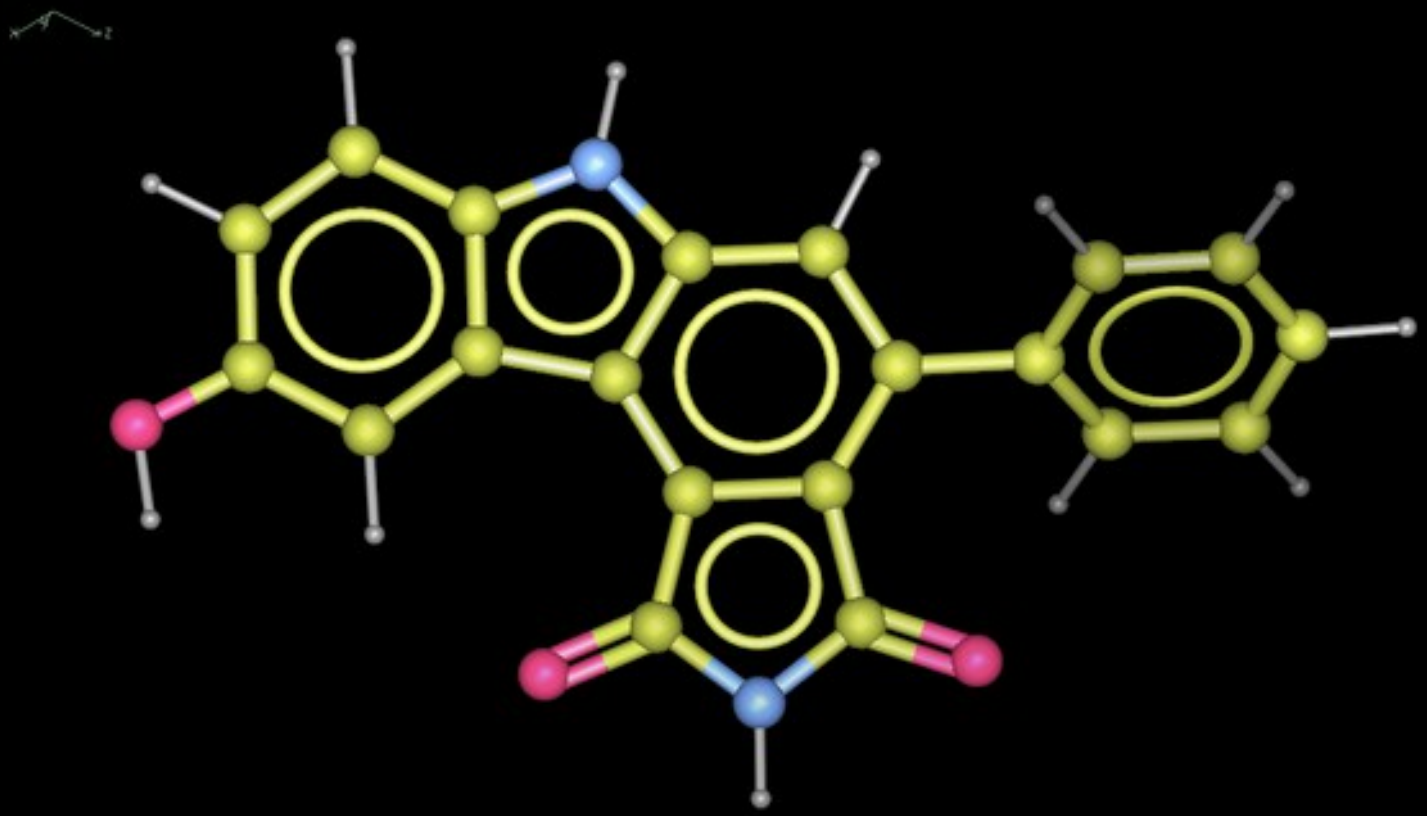
```
  Residue Distortion List:
   plane  O3   C19  C20  C18  C16  C15  C17  C13  C14  N2   C4   C5   O1   C3   C6   O2  penalty-score:  36.51
   plane  C2   C7   C8   C9   C10  C11  C12                                              penalty-score:   8.82
   bond  C13 to  C4  target_value:   1.490 d:   1.432 sigma:   0.020 length-devi  -0.058 penalty-score:   8.44
   bond  C4  to  C3  target_value:   1.490 d:   1.436 sigma:   0.020 length-devi  -0.054 penalty-score:   7.21
   bond  O3  to  C19 target_value:   1.362 d:   1.318 sigma:   0.020 length-devi  -0.044 penalty-score:   4.75
   bond  C19 to  C20 target_value:   1.390 d:   1.433 sigma:   0.020 length-devi   0.043 penalty-score:   4.67
   bond  C1  to  C2  target_value:   1.390 d:   1.428 sigma:   0.020 length-devi   0.038 penalty-score:   3.70
   bond  C4  to  C5  target_value:   1.490 d:   1.454 sigma:   0.020 length-devi  -0.036 penalty-score:   3.26
   bond  C13 to  C14 target_value:   1.490 d:   1.456 sigma:   0.020 length-devi  -0.034 penalty-score:   2.91
   bond  C15 to  C13 target_value:   1.490 d:   1.458 sigma:   0.020 length-devi  -0.032 penalty-score:   2.57
   bond  C16 to  C15 target_value:   1.490 d:   1.459 sigma:   0.020 length-devi  -0.031 penalty-score:   2.45
   angle  C13 -  C4  -  C5   target: 108.00 model_angle: 133.80 sigma:  3.00 angle-devi 25.80 penalty-score:  73.93
   angle  O1  -  C5  -  C4   target: 108.00 model_angle: 126.59 sigma:  3.00 angle-devi 18.59 penalty-score:  38.38
   angle  C13 -  C15 -  C16  target: 120.00 model_angle: 102.30 sigma:  3.00 angle-devi 17.70 penalty-score:  34.83
   angle  O2  -  C6  -  N1   target: 108.00 model_angle: 122.80 sigma:  3.00 angle-devi 14.80 penalty-score:  24.34
   angle  O2  -  C6  -  C3   target: 108.00 model_angle: 122.76 sigma:  3.00 angle-devi 14.76 penalty-score:  24.19
   angle  C13 -  C15 -  C17  target: 120.00 model_angle: 133.33 sigma:  3.00 angle-devi 13.33 penalty-score:  19.76
   angle  C4  -  C13 -  C15  target: 120.00 model_angle: 132.99 sigma:  3.00 angle-devi 12.99 penalty-score:  18.76
   angle  N1  -  C5  -  O1   target: 108.00 model_angle: 120.48 sigma:  3.00 angle-devi 12.48 penalty-score:  17.32
   angle  C15 -  C13 -  C14  target: 120.00 model_angle: 110.43 sigma:  3.00 angle-devi -9.57 penalty-score:  10.18
   angle  N1  -  C6  -  C3   target: 108.00 model_angle: 114.28 sigma:  3.00 angle-devi  6.28 penalty-score:   4.38
   angle  C6  -  C3  -  C4   target: 108.00 model_angle: 101.75 sigma:  3.00 angle-devi -6.25 penalty-score:   4.34
Residue Distortion Summary:
   29 bond restraints
   44 angle restraints
   sum of bond  distortions penalties:  59.5697
   sum of angle distortions penalties:  300.405
   average bond  distortion penalty:    2.05413
   average angle distortion penalty:    6.82739
   total distortion penalty:            405.304
   average distortion penalty:          4.93116
```

Mogul Results Representation

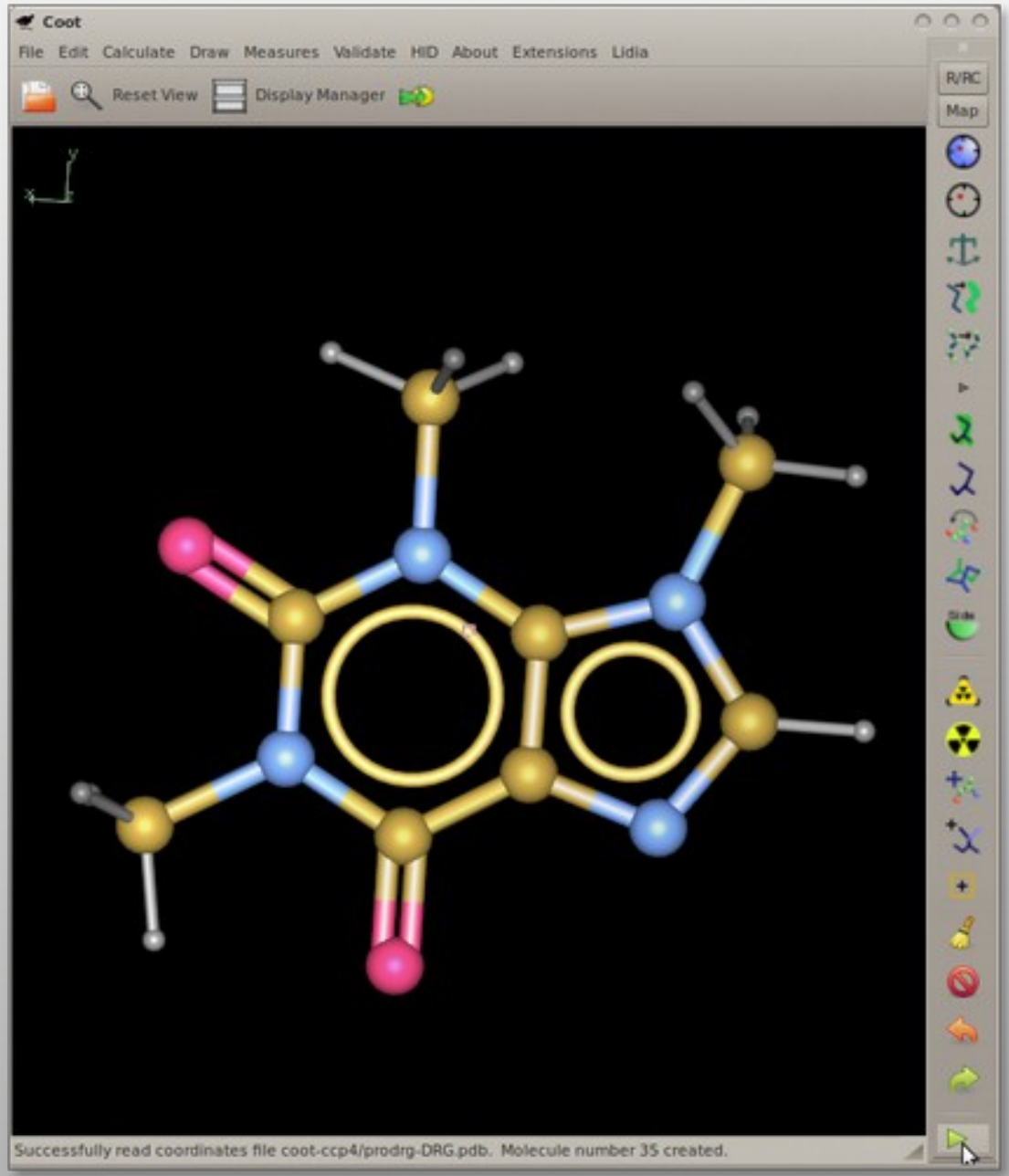


Ligand Represenation

- Bond orders (from dictionary restraints)

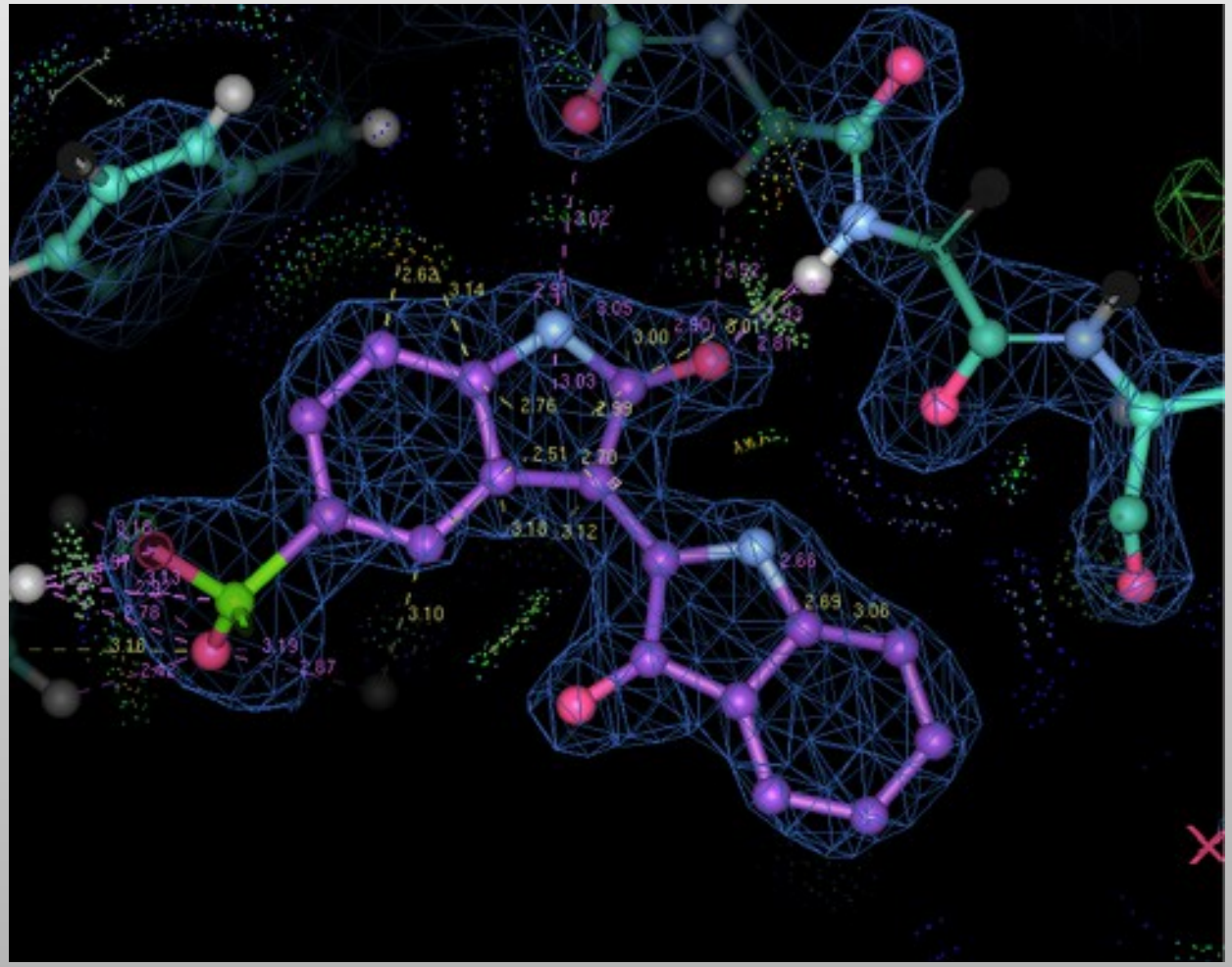


Representing Bond Orders



Other Tools

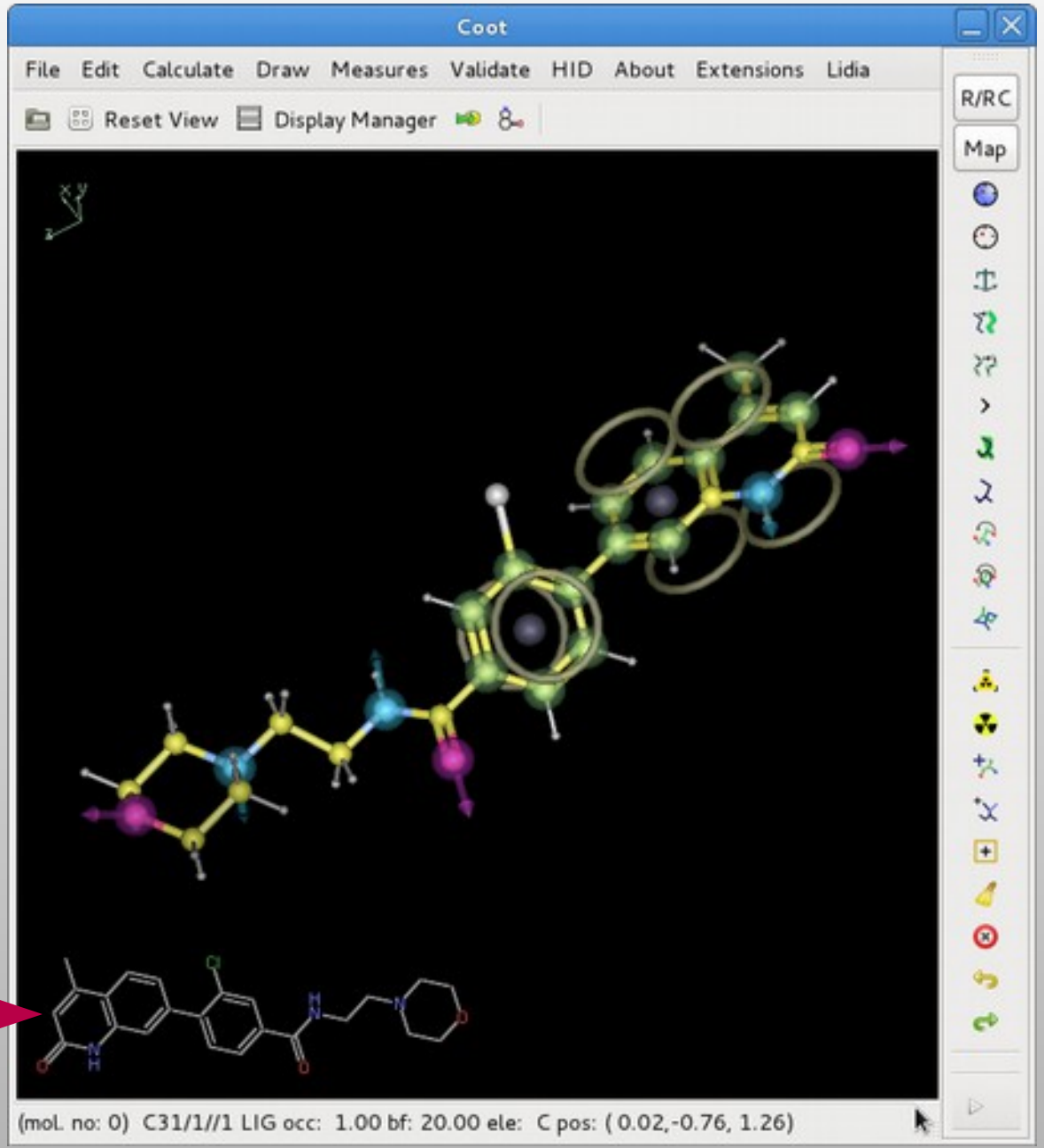
- Molprobity dots for ligands
  - Highlight interesting site



Chemical Features

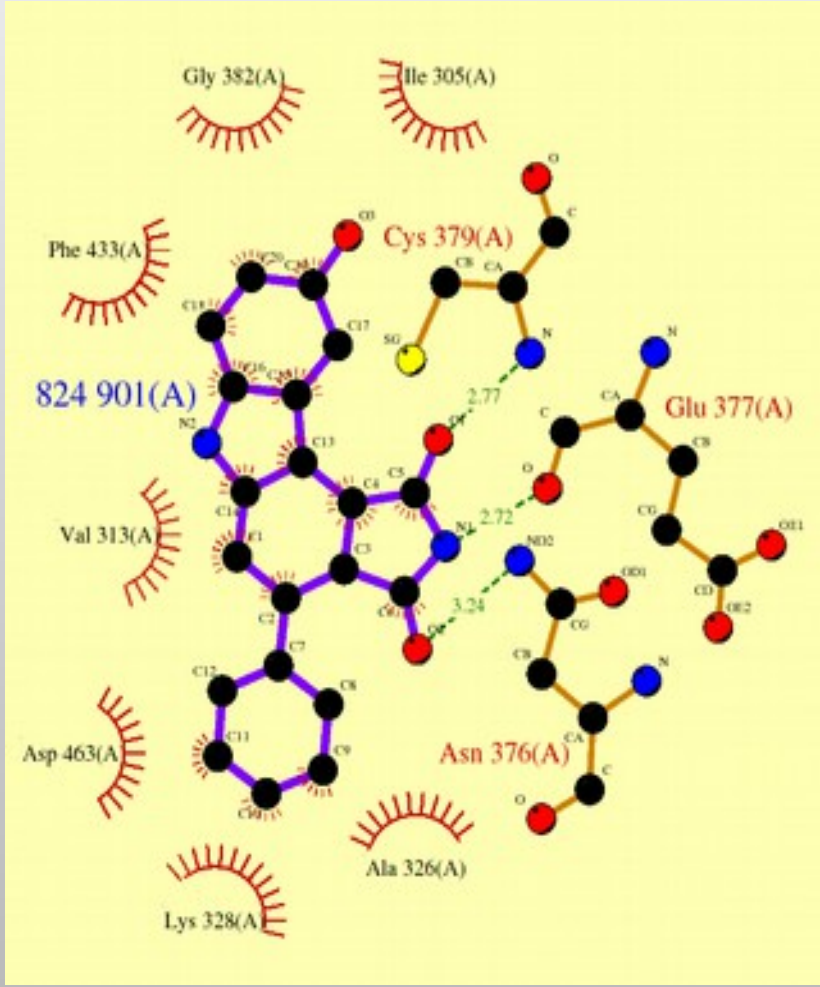
Uses built-in FeatureFactory

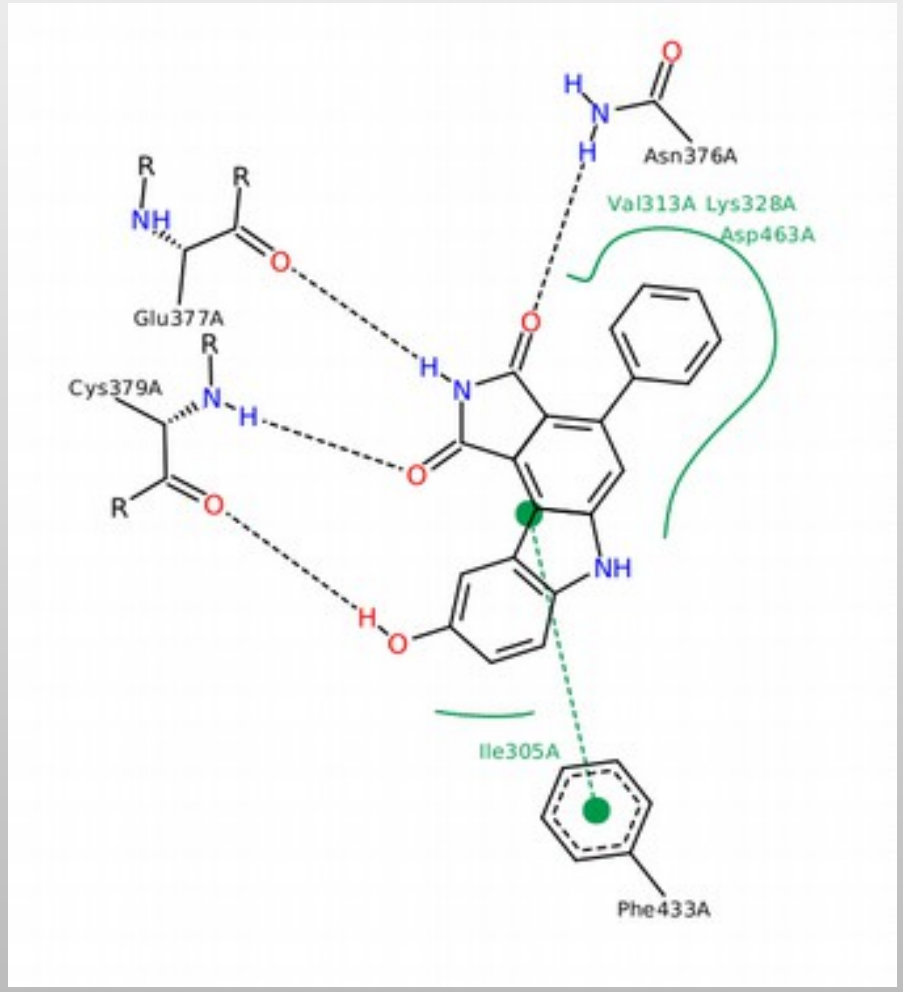
...and on the fly thumbnailing



Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





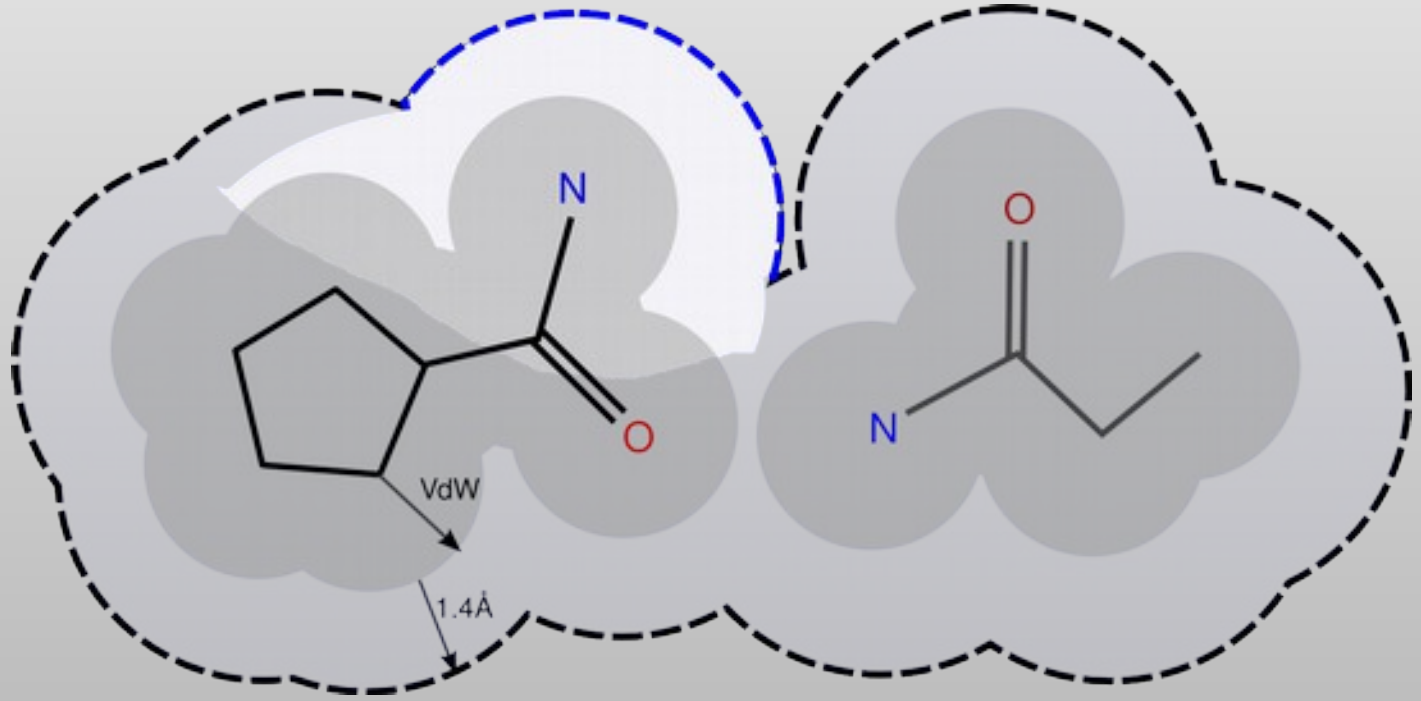
Can we do better? - Interactivity?

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

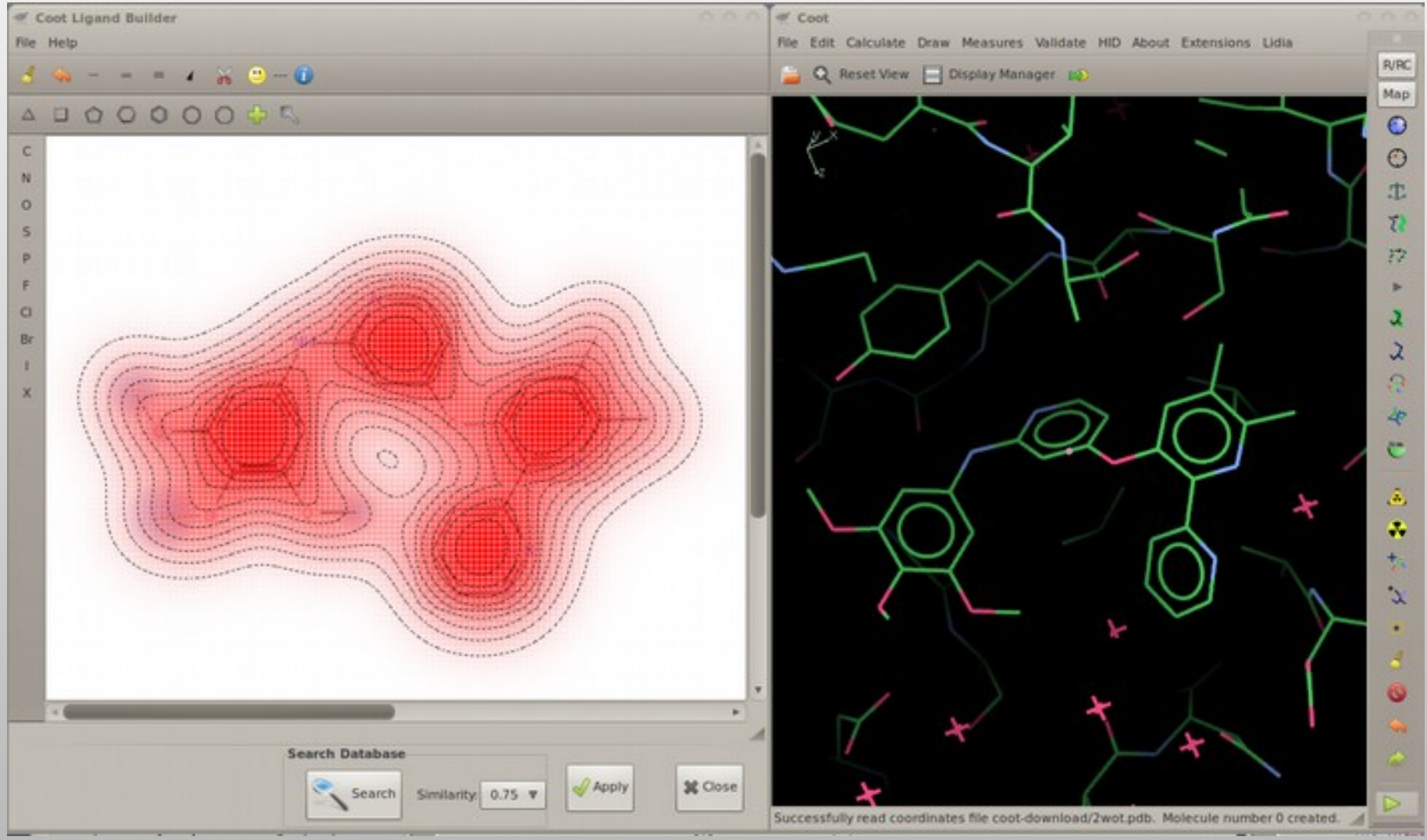
$$\sum\sum (d_{ik}^2 - D_{ik}^2) +$$

Residues are close to H-bonding ligand atoms

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

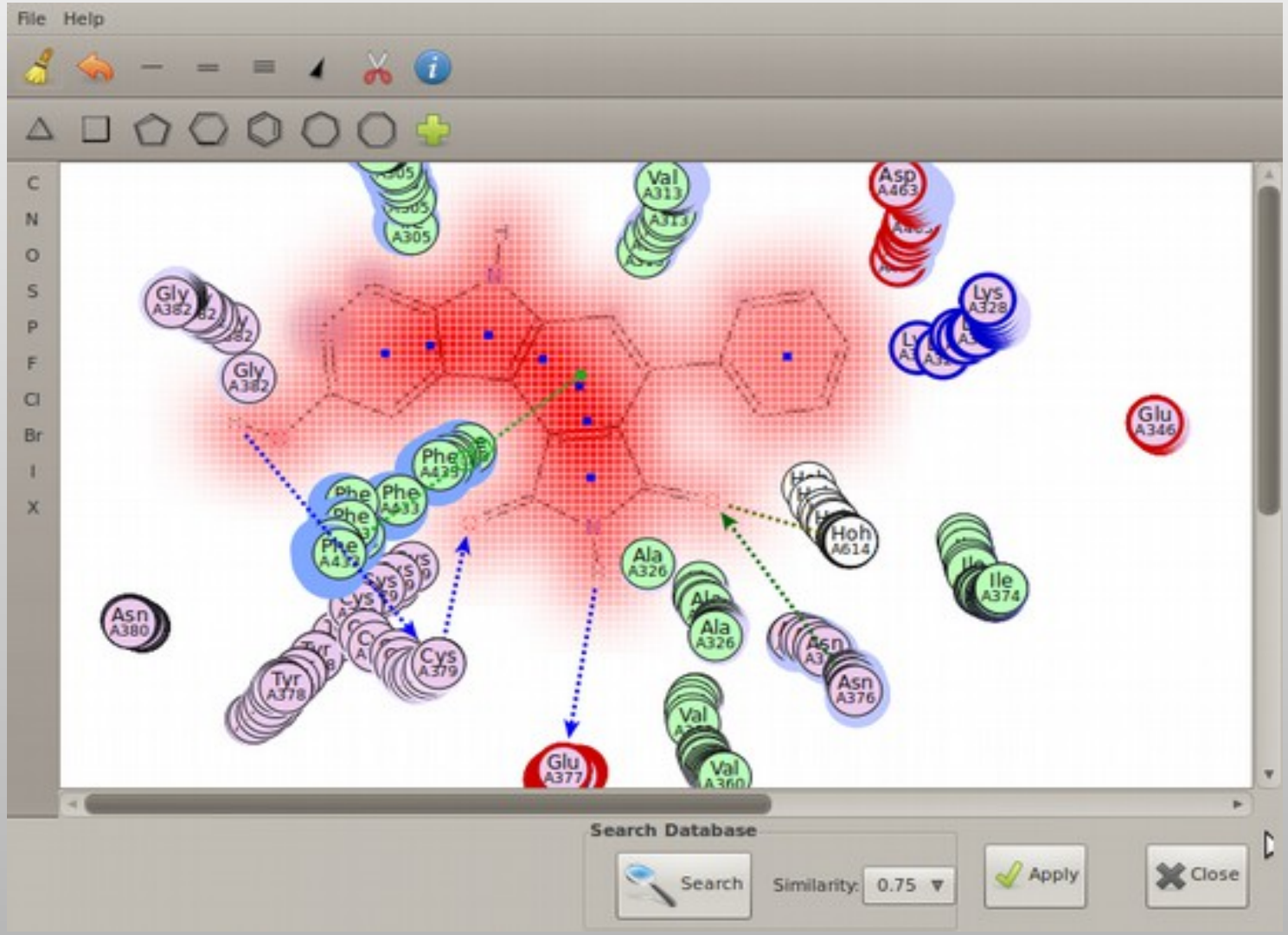
Residues don't overlap ligand

”Don't overlap the ligand”



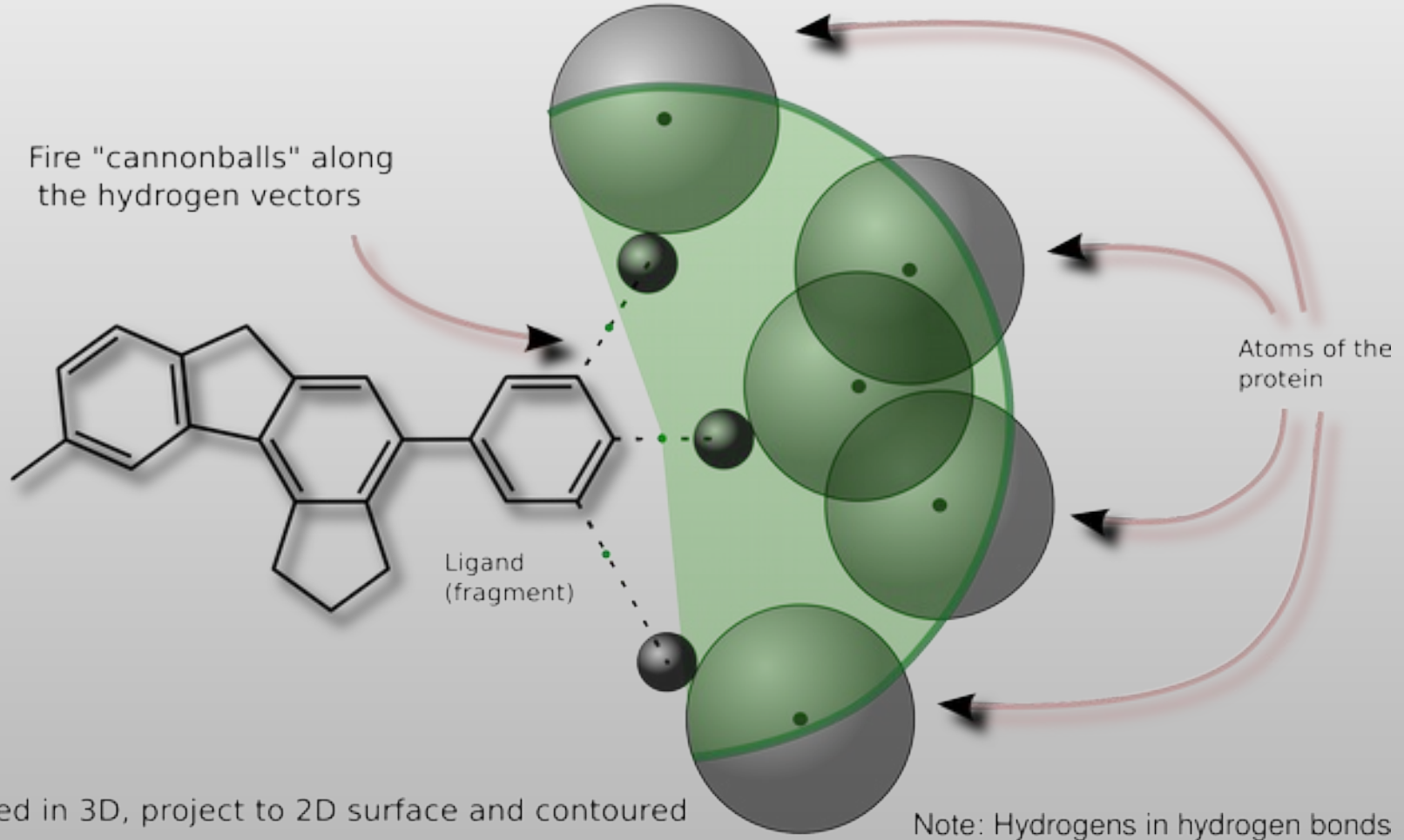
Ligand Environment Layout

- Residue position minimisation



Determination of the Substitution Contour

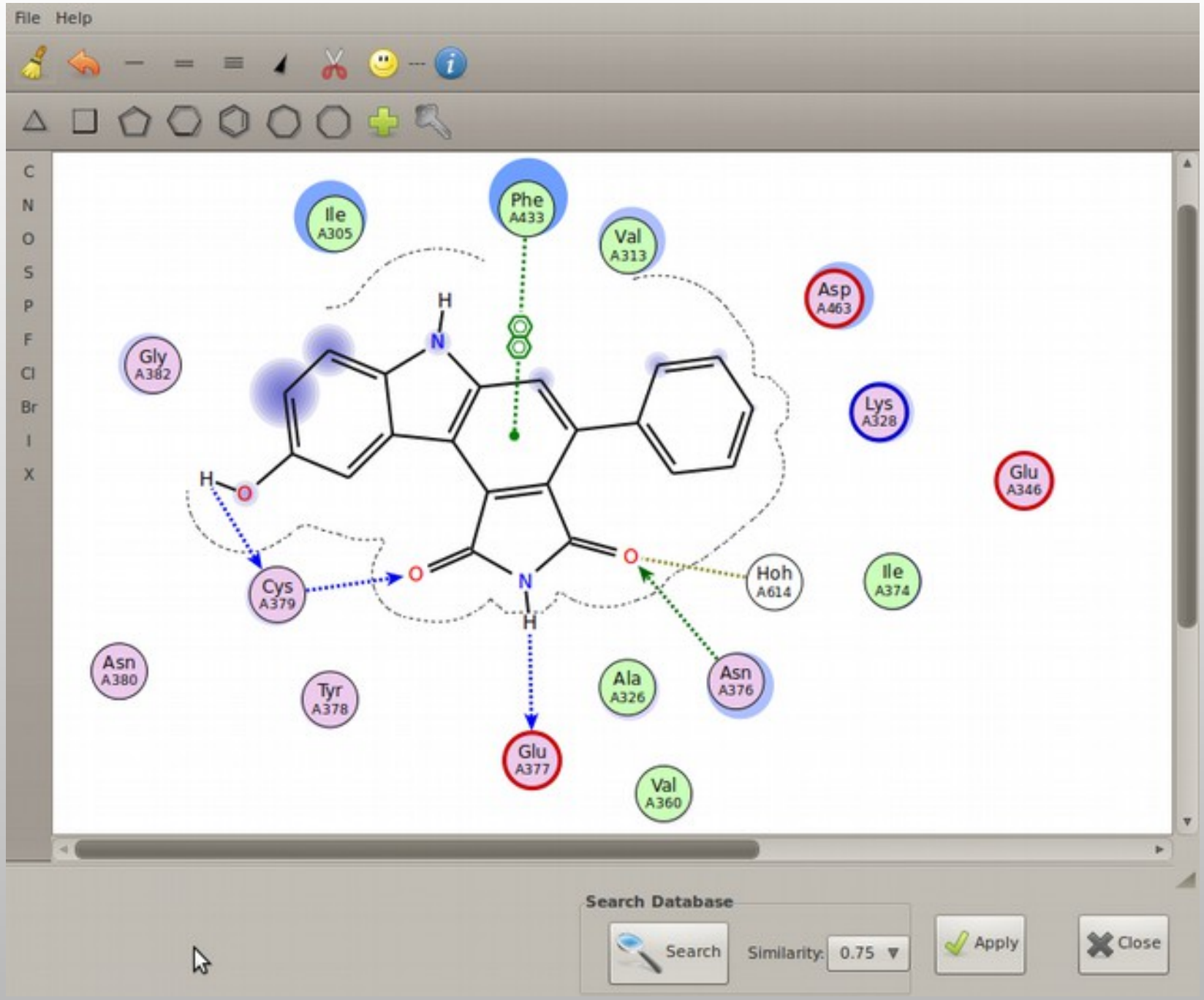
How far can we go (in the direction of the hydrogens) before hitting atoms of the protein?



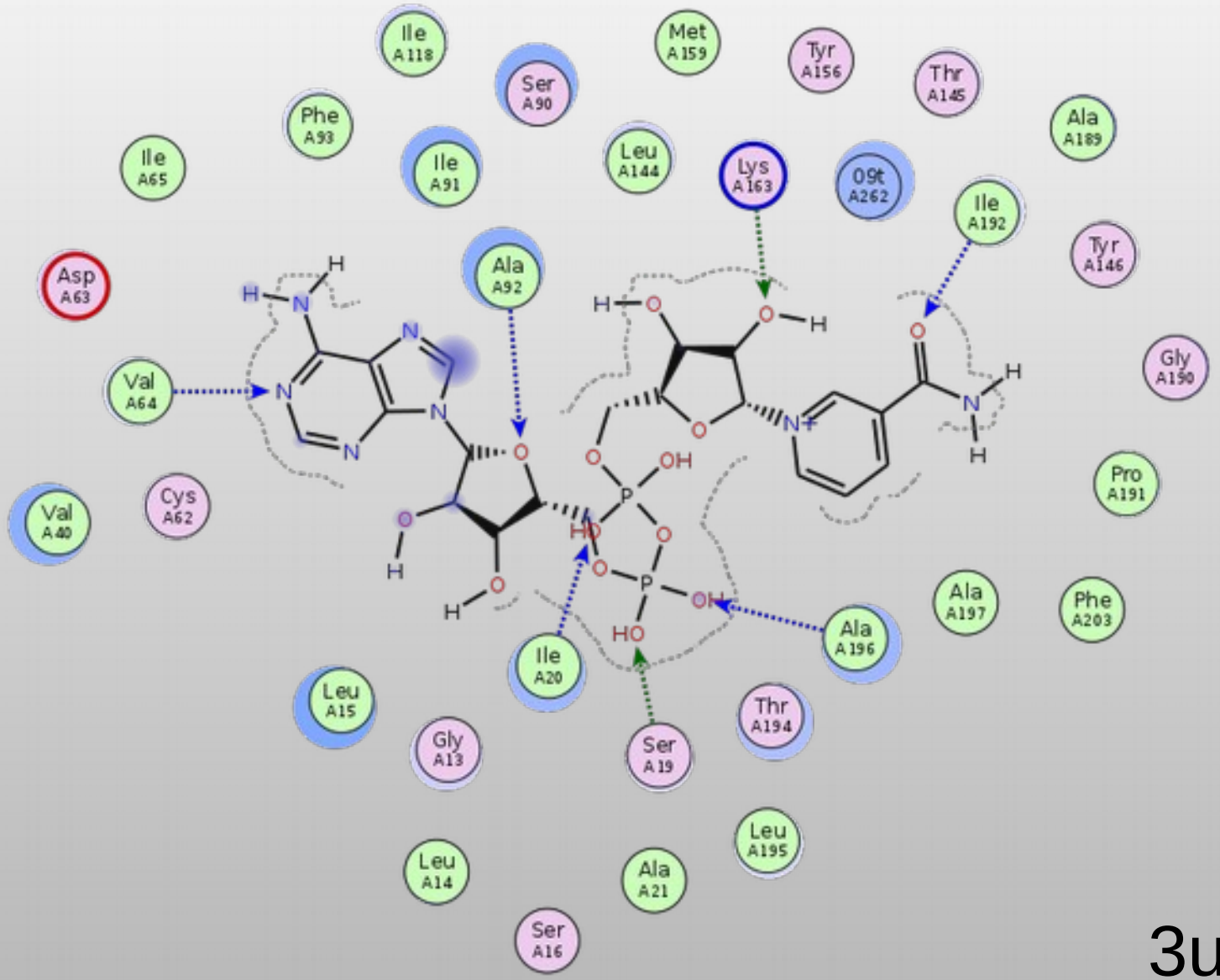
Determined in 3D, project to 2D surface and contoured

c.f. Clarke & Labute (2007)

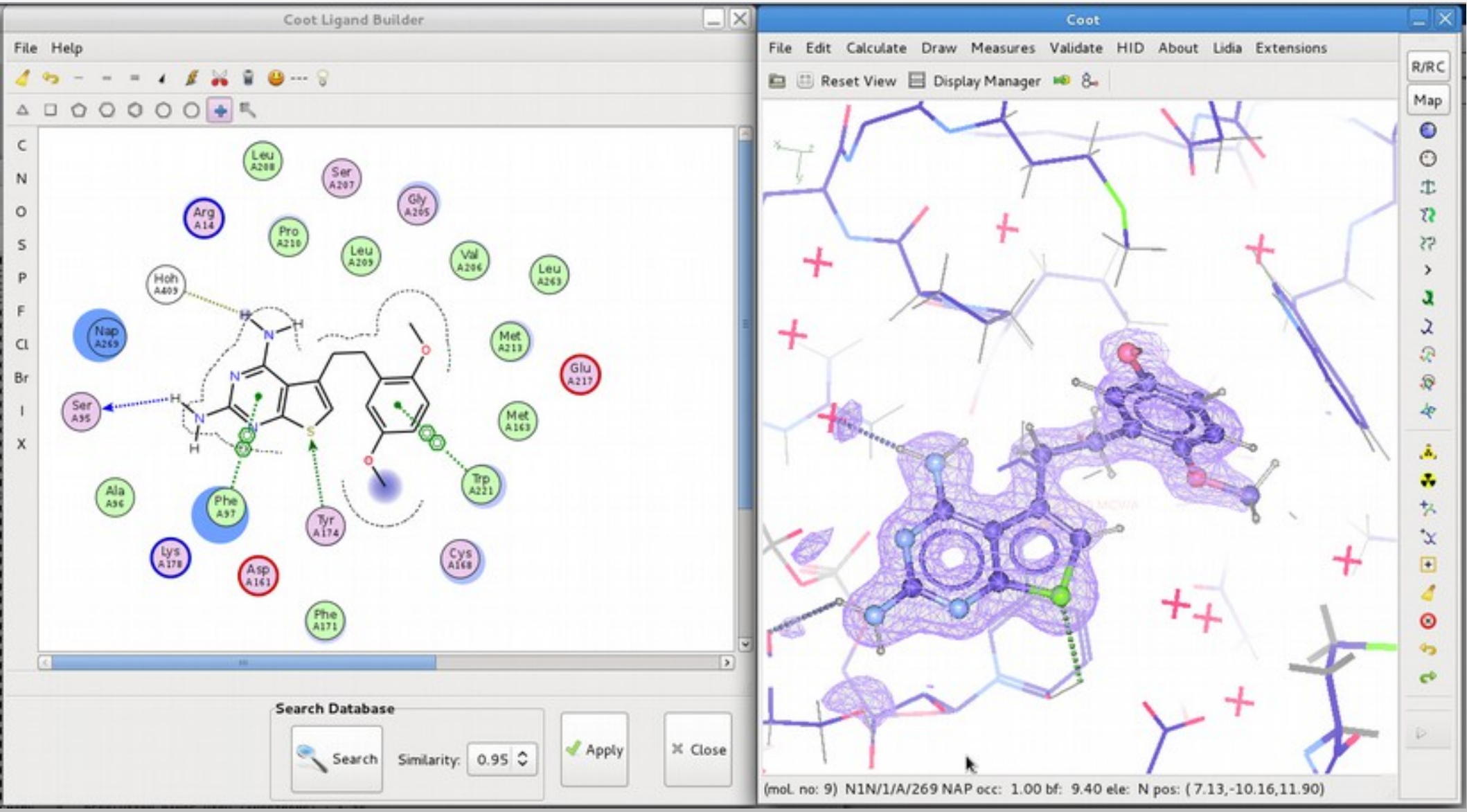
Note: Hydrogens in hydrogen bonds are a confounding factor


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

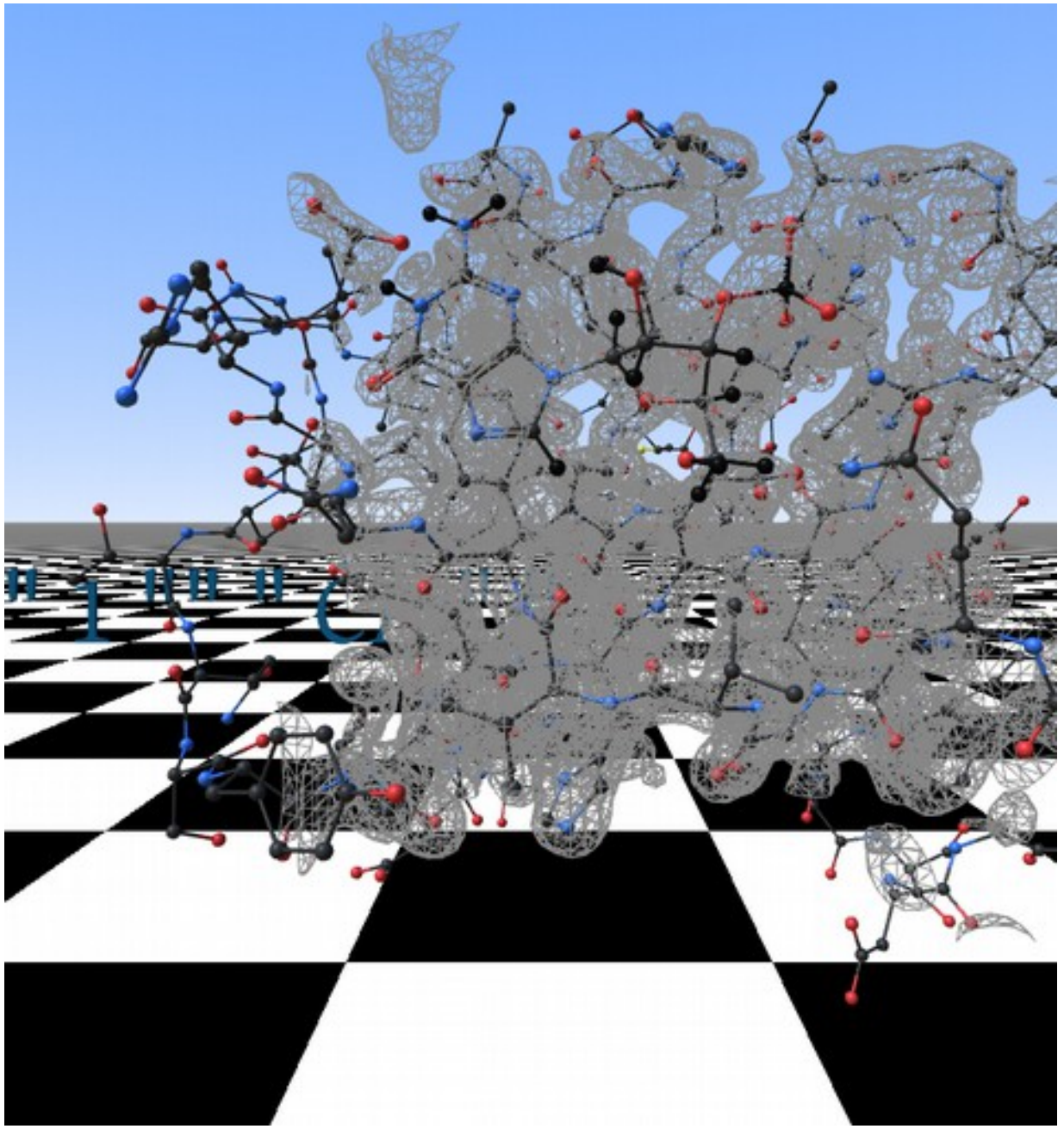


3uic


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)

Coot Futures: Virtual Reality

Hamish Todd



- **An Intuitive Interface:**
- Stereoscopic Representation
- Greater Field of View
- 2 Hands with Articulation
- However:
  - current tools are not immediately transferable
  - because: nausea

CootVR

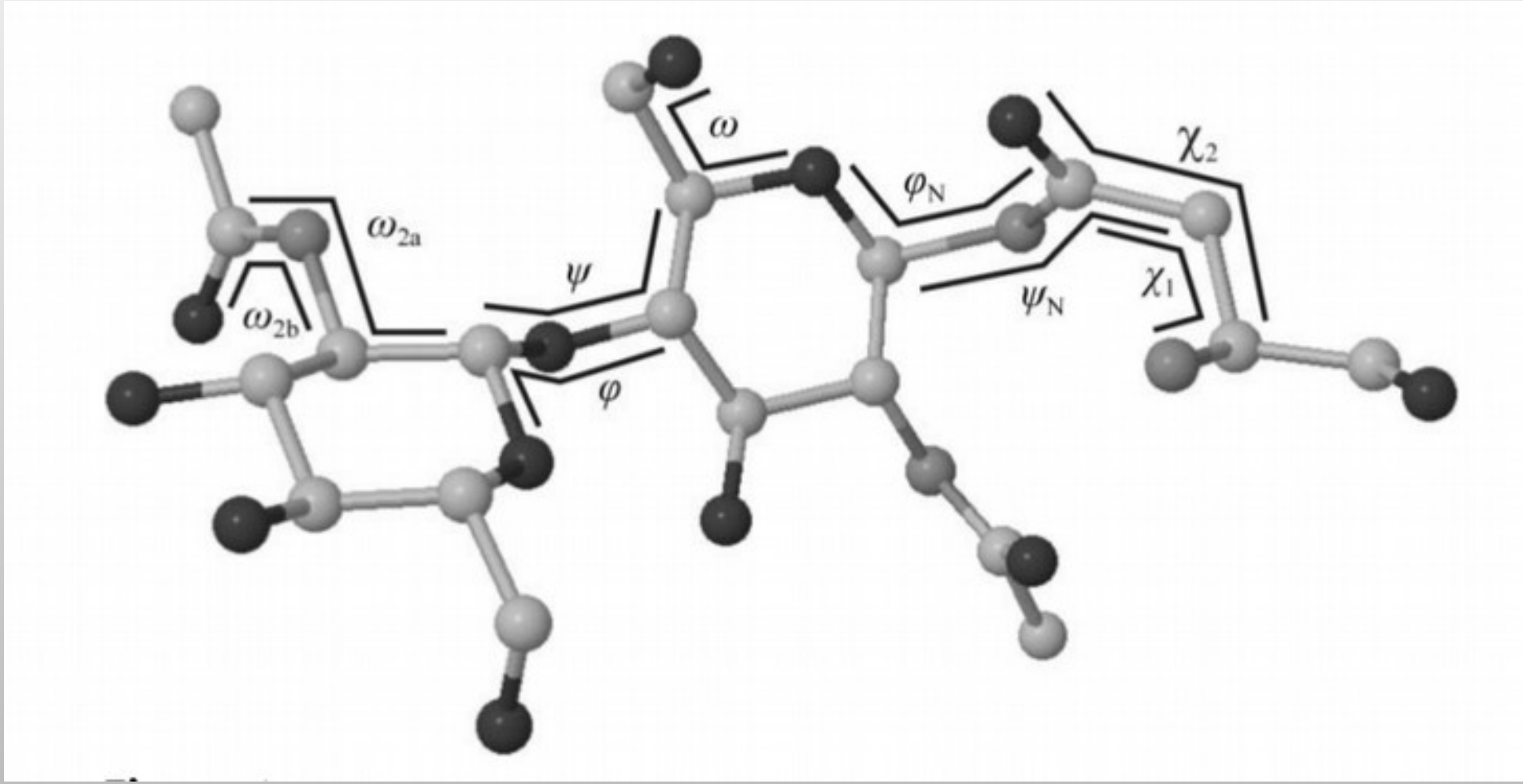
- Demonstrated at CCP-EM Meeting in Keele, UK in April

<demo video>

Problematic Glycoproteins

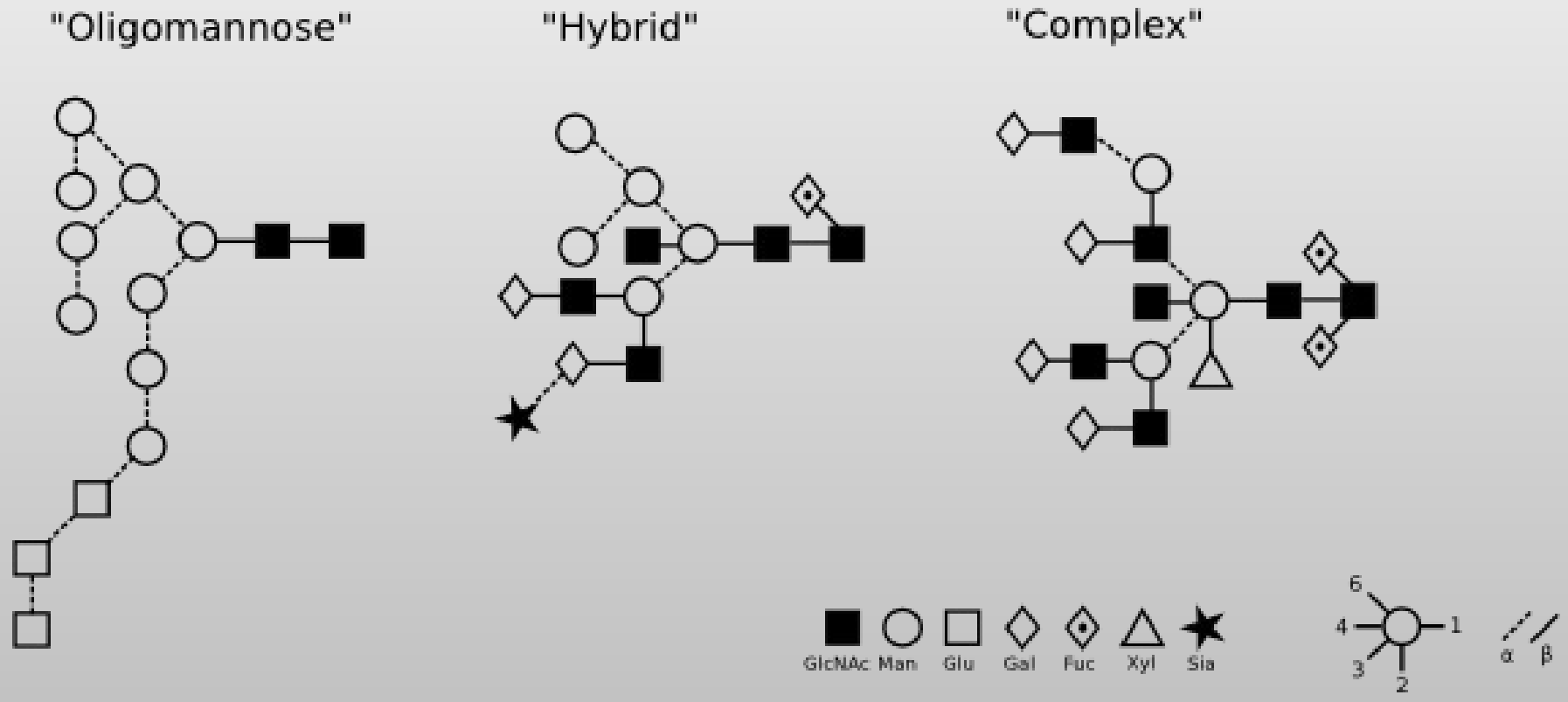
- Crispin, Stuart & Jones (2007)
  - NSB Correspondence
  - “one third of entries contain significant errors in carbohydrate stereochemistry...”
  - “carbohydrate-specific building and validation tools capable of guiding and construction of biologically relevant stereochemically accurate models should 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)

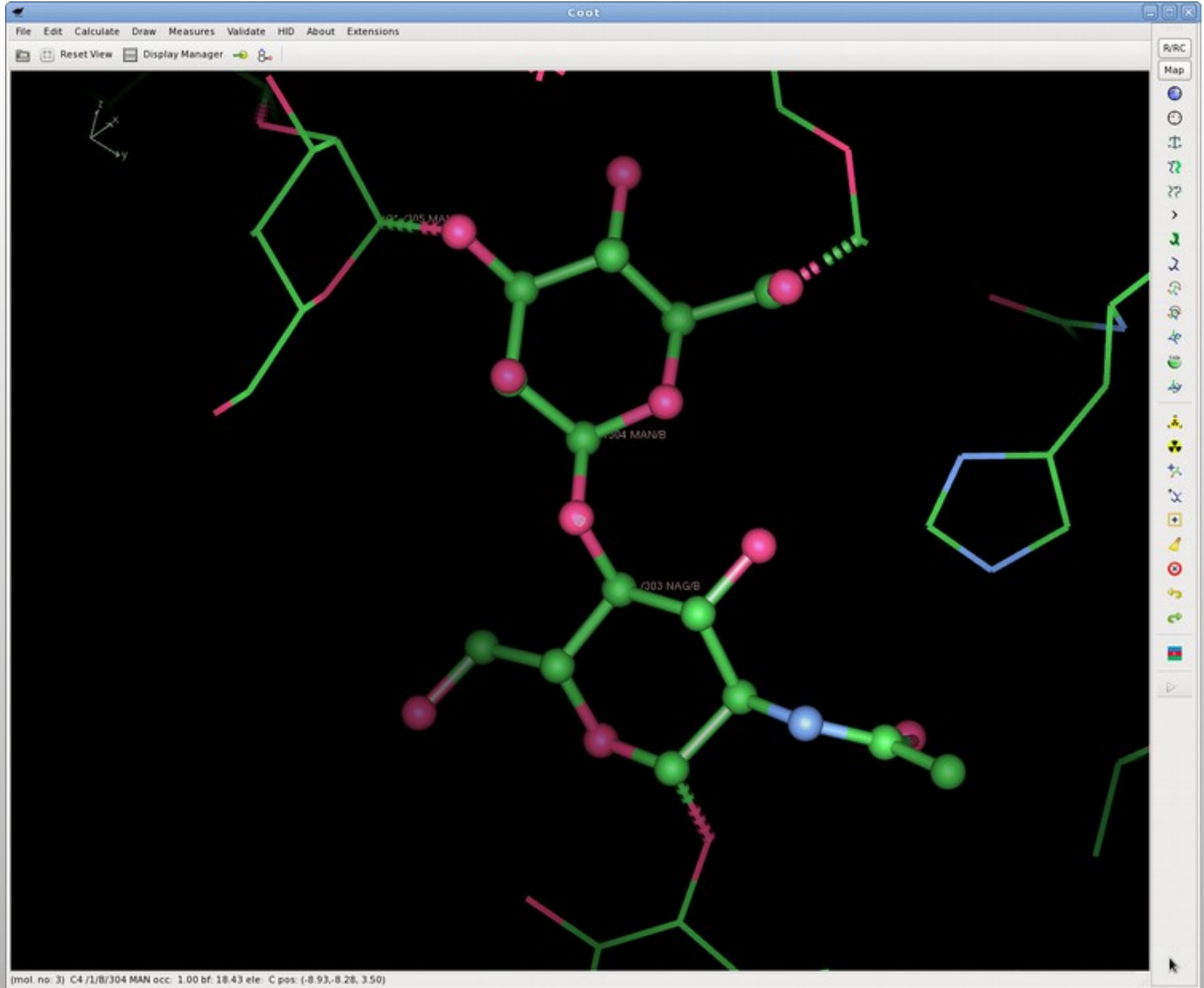
Validate the Tree: N-linked carbohydrates

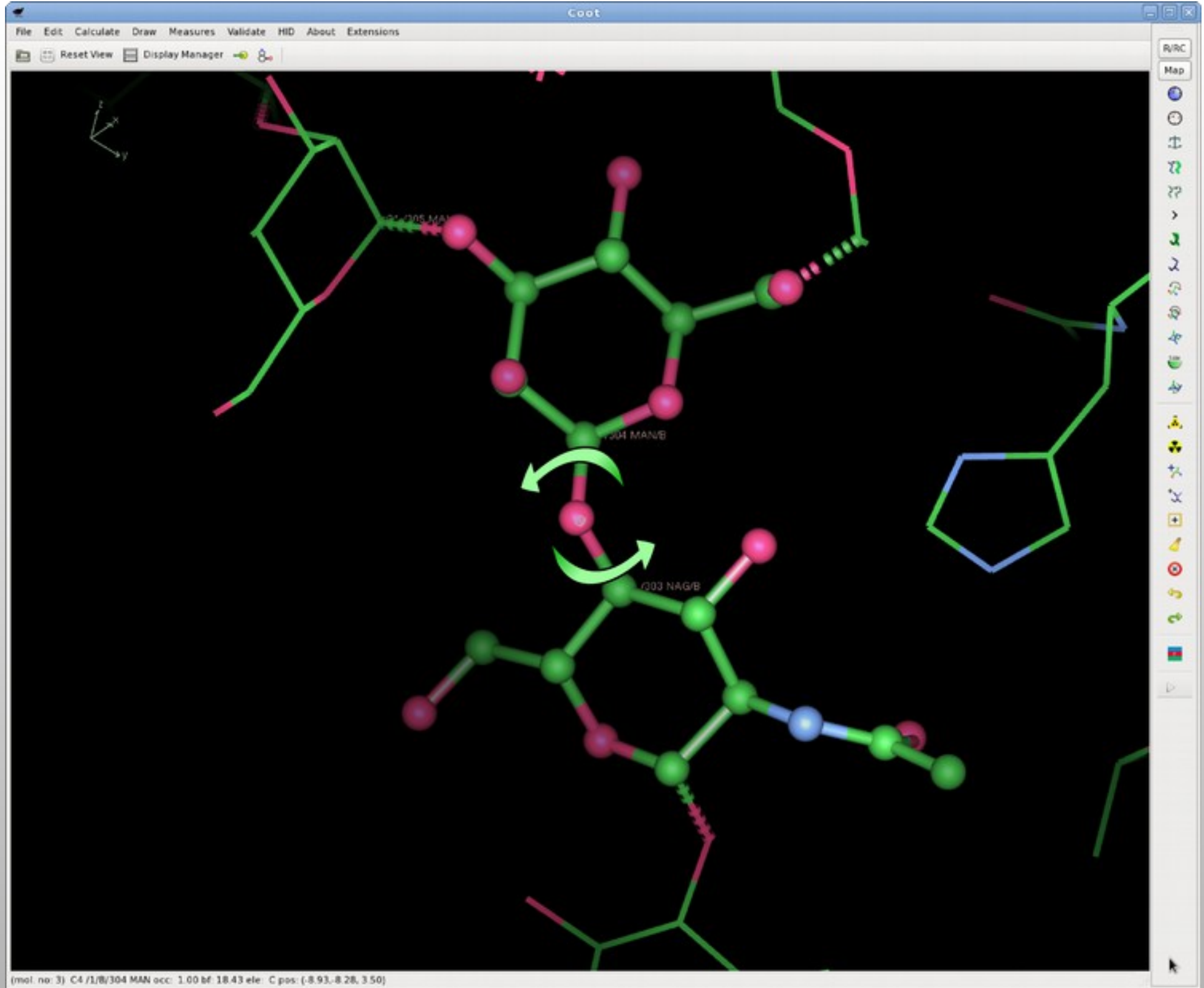


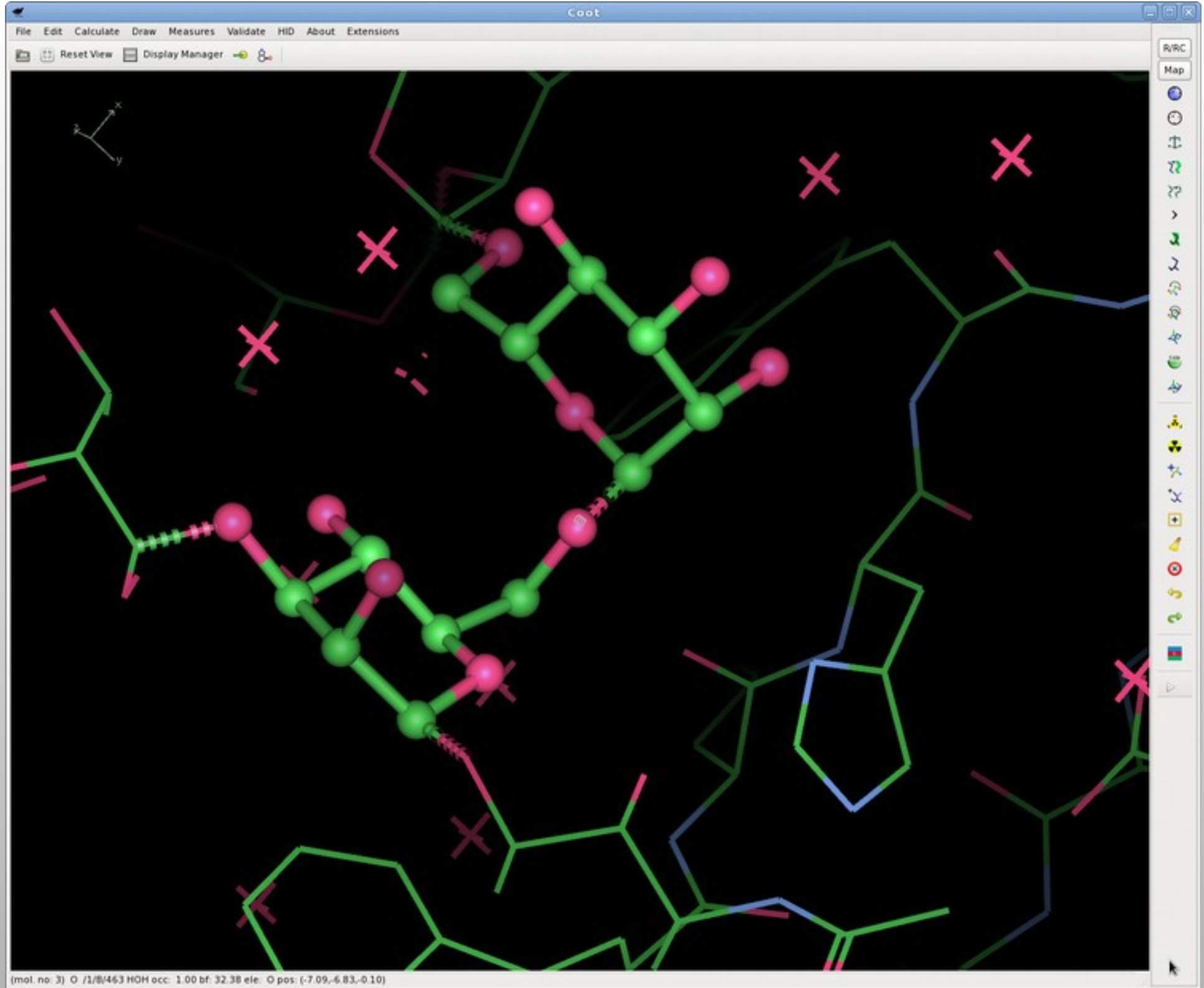
Linking Oligsaccharides/Carbohydrates:

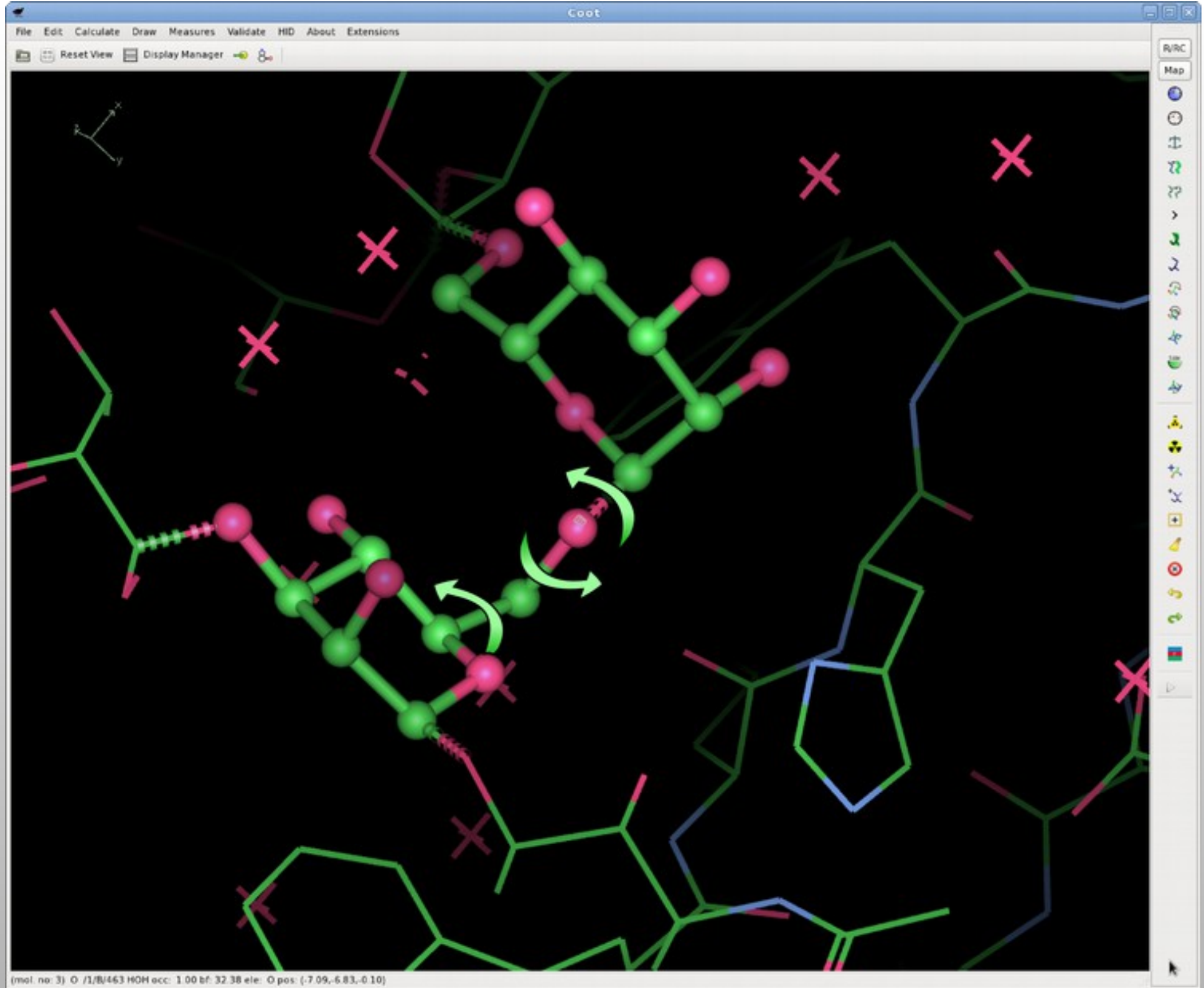
LO/Carb

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

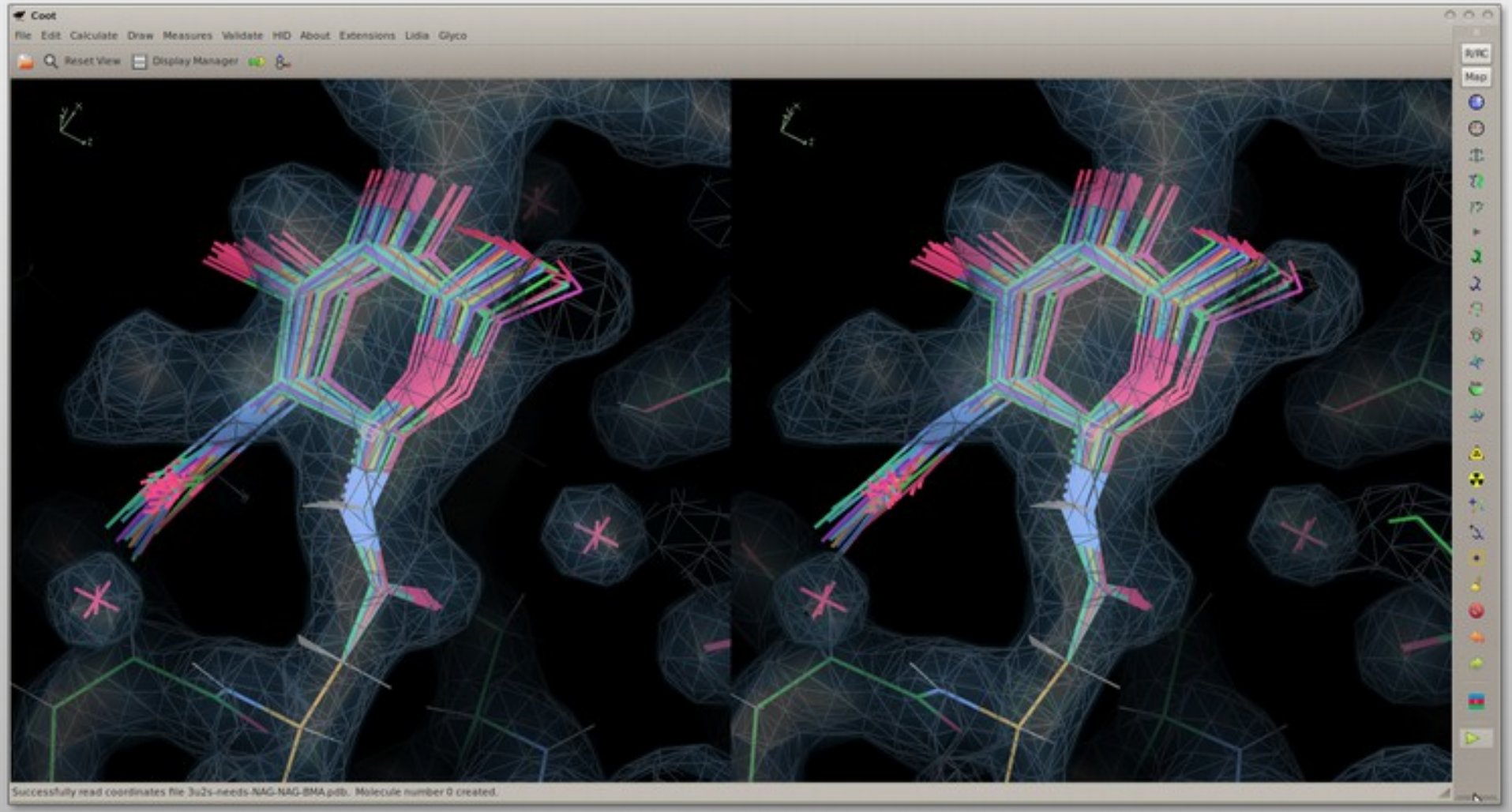

Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

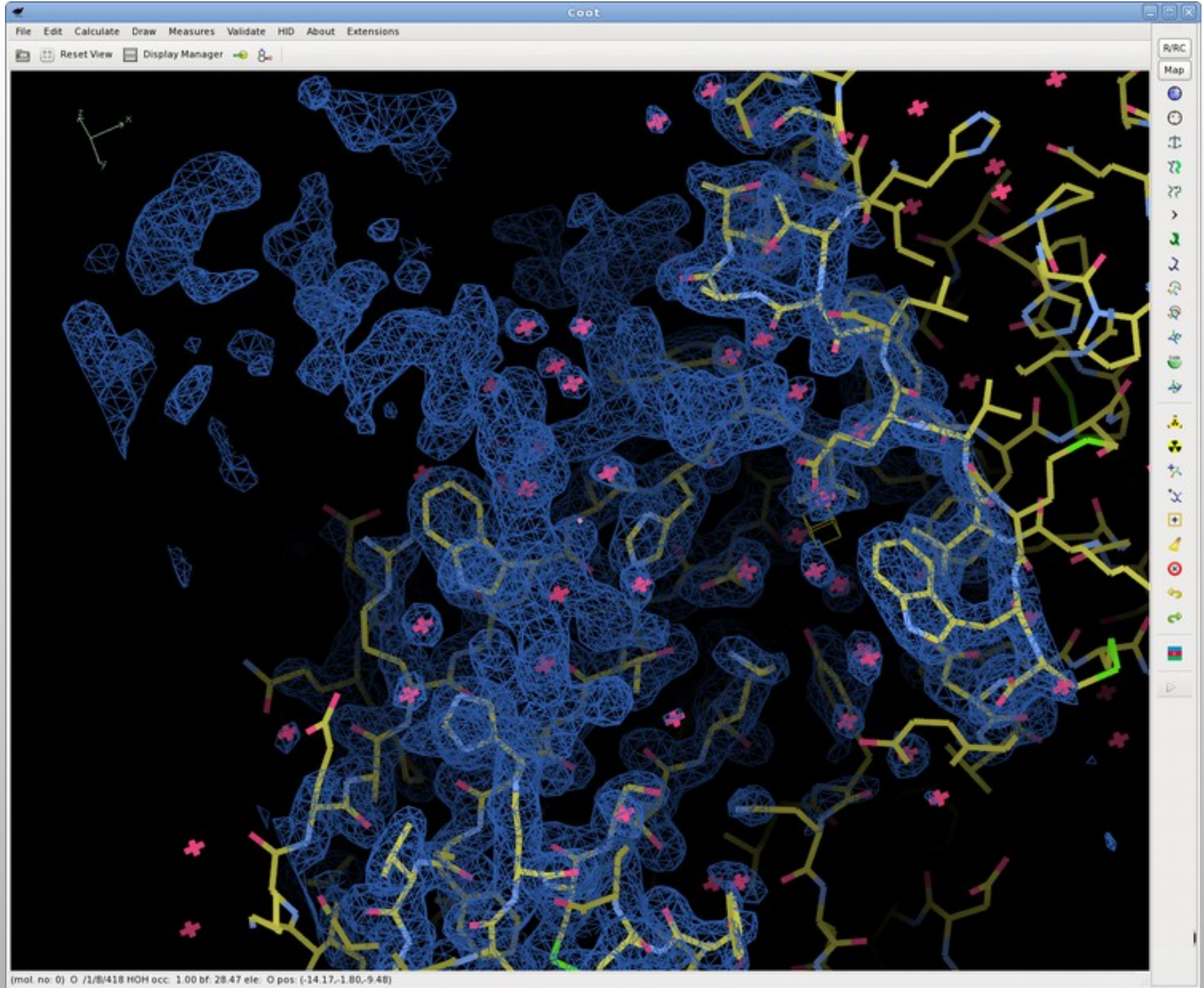

Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

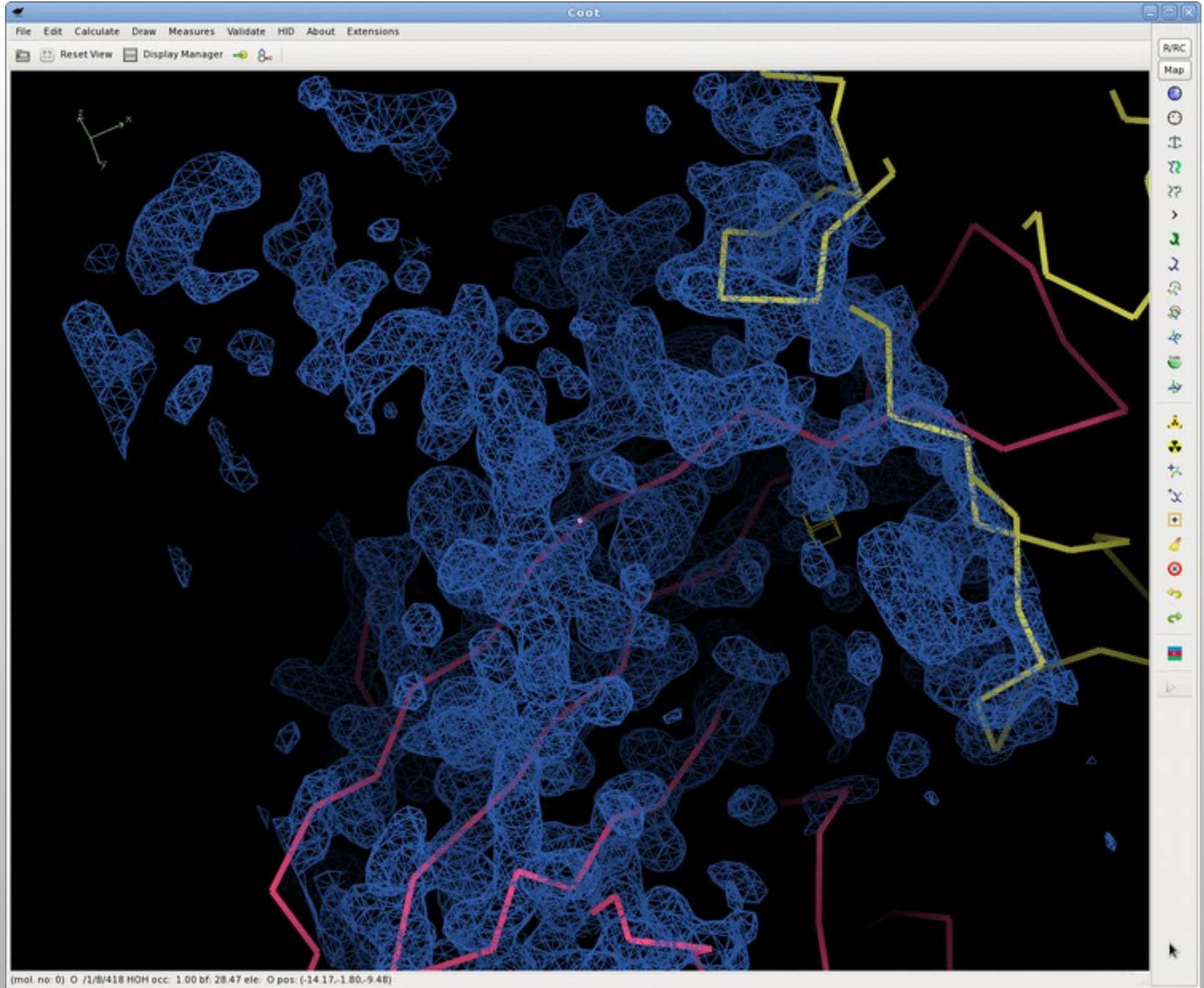

Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 3) O /1/B/463 HOH occ: 1.00 bf: 32.38 ele: O pos: (-7.09,-6.83,-0.10)


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)

Acknowledgements

- Kevin Cowtan
- Bernhard Lohkamp

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

- Funding:
  - BBSRC & CCP4

Representing Bond Orders