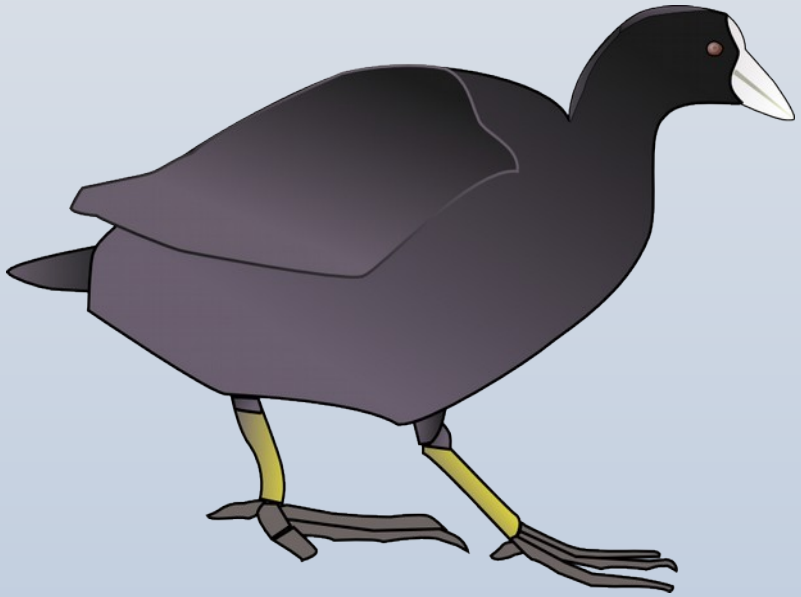
Model-Building Using X-ray Data With *Coot*

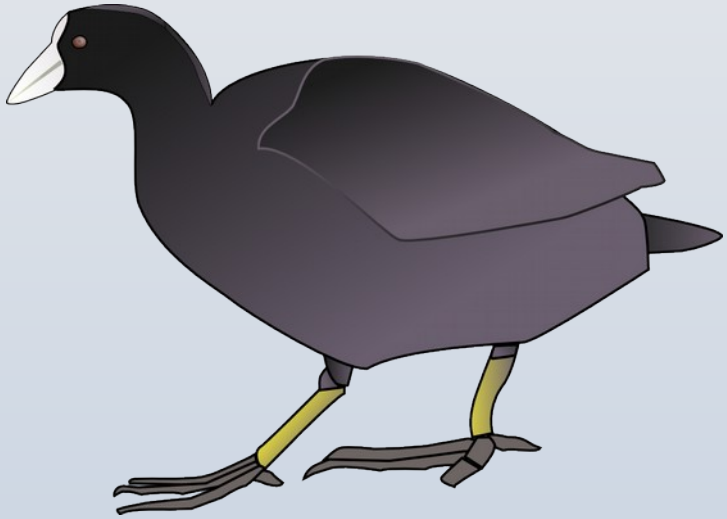


Paul Emsley
MRC Laboratory of Molecular Biology
Jan 2017

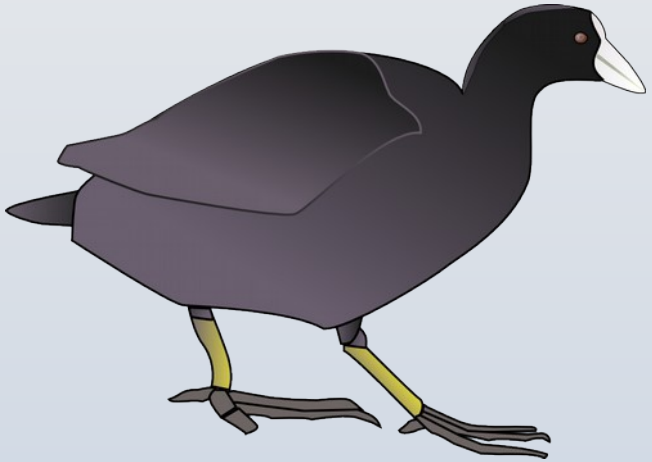
Modelling Proteins with *Coot*

About this presentation:

- (Quite) New tools
- “Bonbons pour les yeux”
- Backrub Rotamers
- Ligands
- N-linked carbohydrates
- cis-peptides
- pdf available if needed

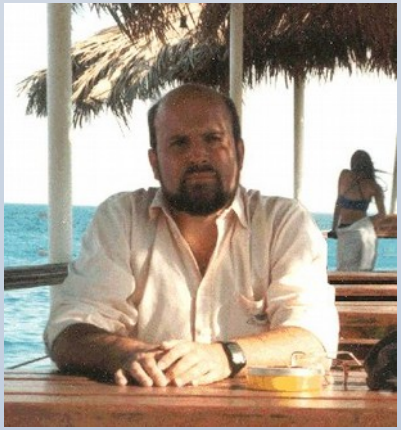


Acknowldegments, Collaborators



Bernhard Lohkamp

Kevin Cowtan



Eugene Krissinel

Stuart McNicholas

Martin Noble

Alexei Vagin

A Brief History of *Coot*

- Released in 2004, Coot was designed primarily for model-building protein models into maps from x-ray data
  - Torsions: Rotamers, Ramachandran plots
  - Several optimisers, including Real Space Refinement
- Used typically after automated model-building or refinement
- Since:
  - Nucleic Acids, Ligands & Cryo-EM
- It's never been pretty...

  - Not the best tool for presentation graphics and animations

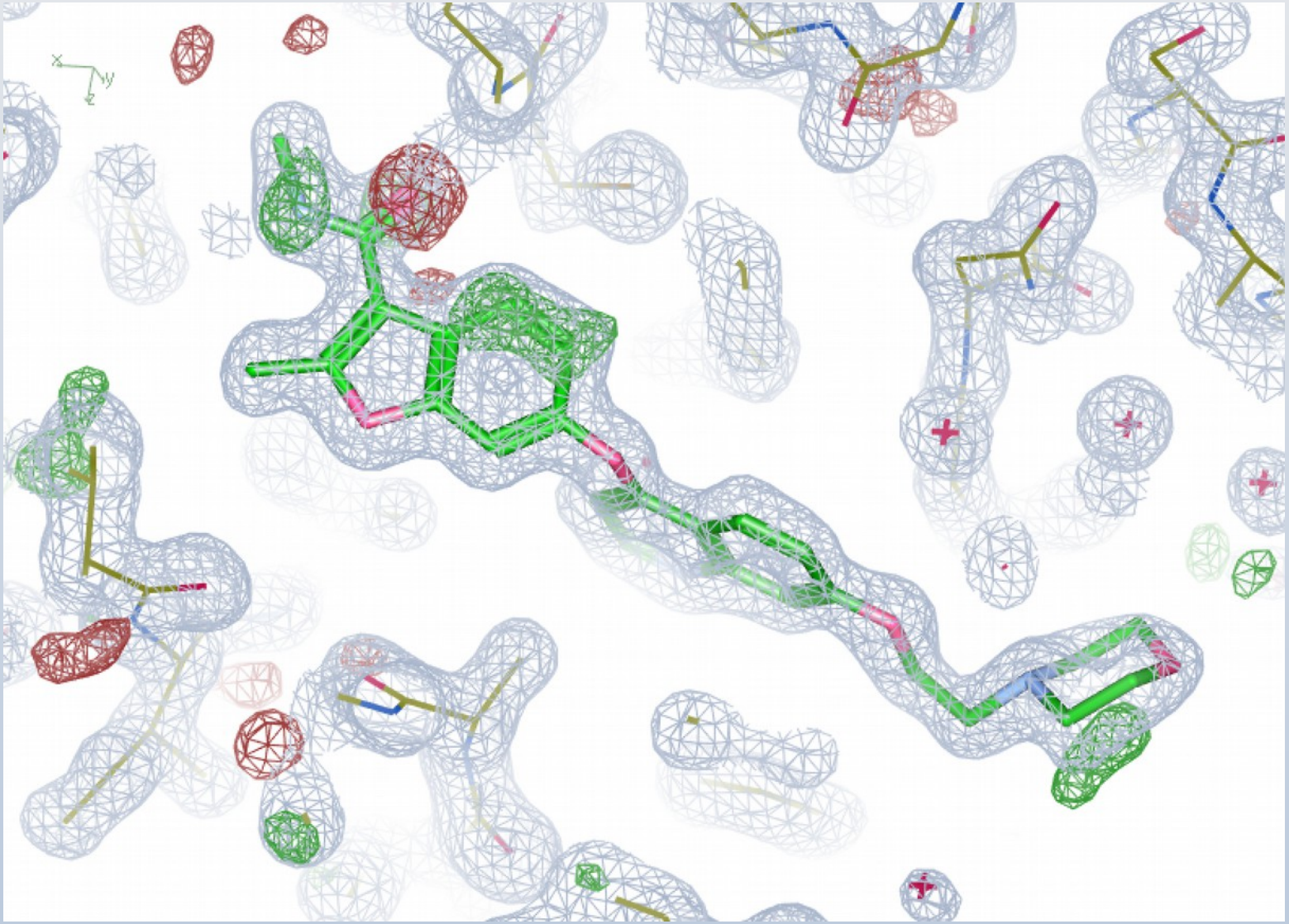
Coot Key-bindings

- Many hundreds of functions available in *Coot*'s API
  - available via scheme or python
- *Coot*'s gui doesn't help much to learn key-bindings
  - they are "off" by default
  - so that you can program your own
- If you are more than a casual/occasional users of *Coot*... are probably worth learning

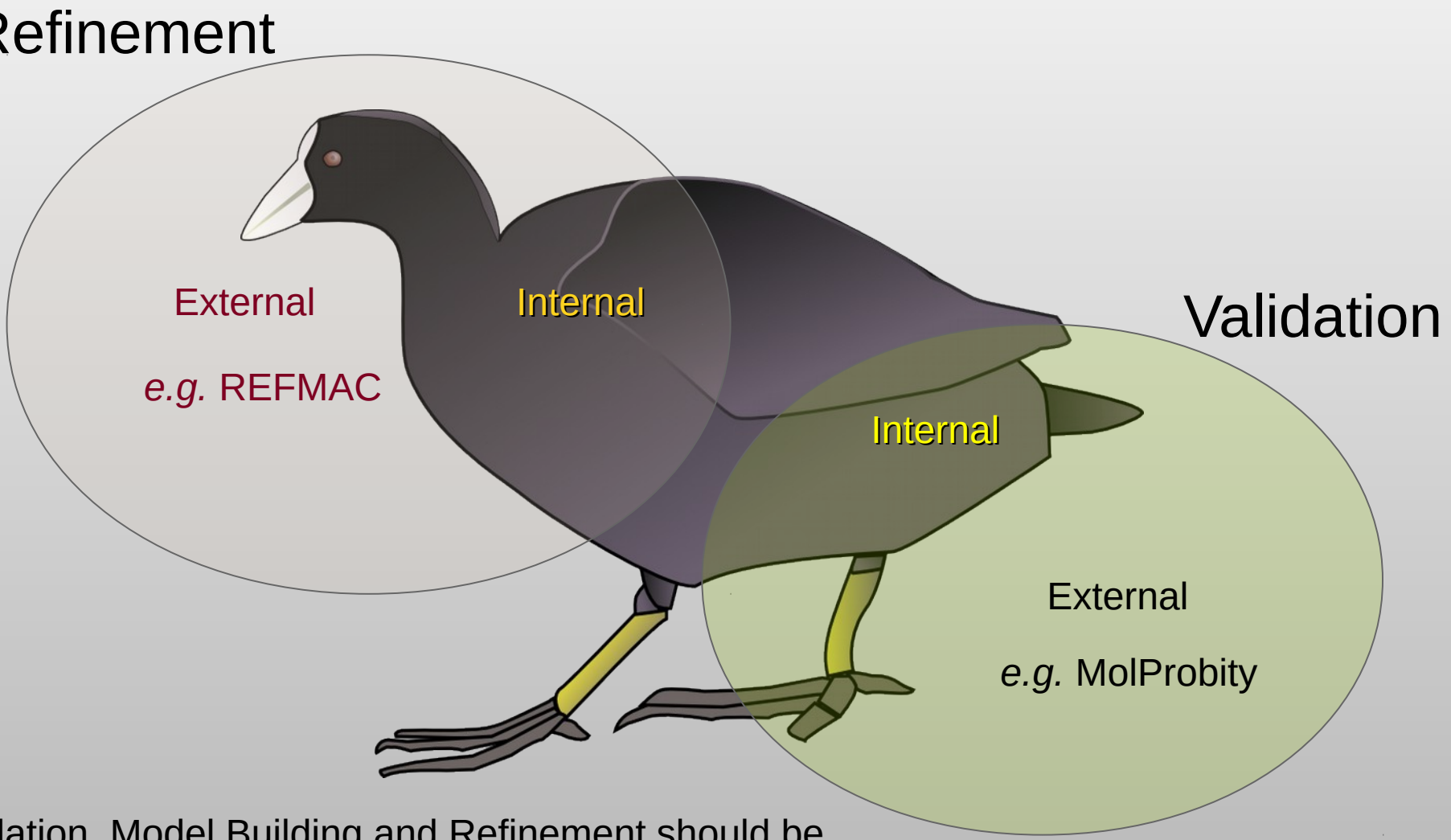
Making Density Slides with *Coot*

- White background
- “High” Oversampling (2.3x)
- Pale gray (or very pastel) density colour
- Enable Cut-glass mode 5-10%
- Anti-aliased Coot
  - `$ setenv __GL_FSAA_MODE 5`
  - 0.8.3 will do a better job of anti-aliasing out the box

Example Density Slide



Feature Integration

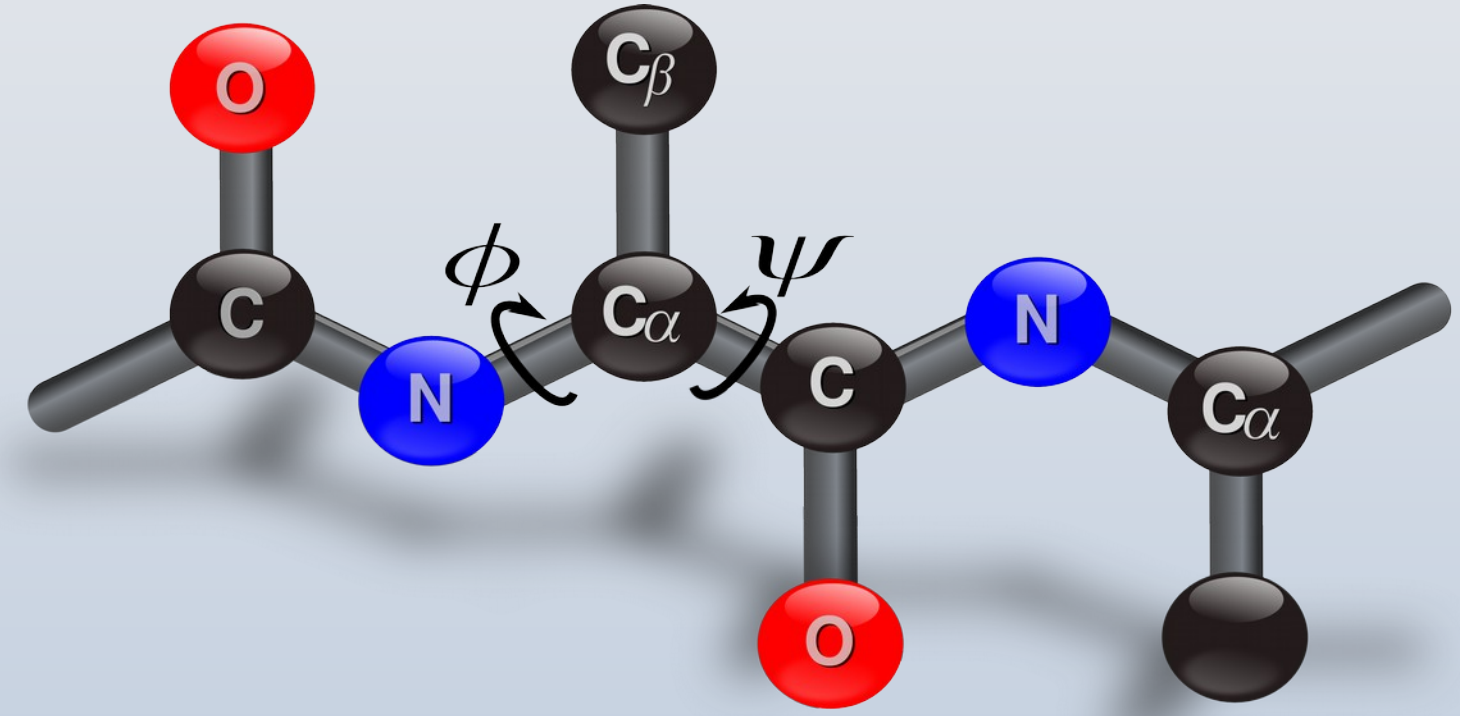


Validation, Model Building and Refinement should be used together

Real Space Refinement

- Major Feature of Coot
  - Gradient-based minimiser (BFGS derivative)
  - Geometry library is the standard CIF-based Refmac dictionary
    - Minimise deviations in bond length, angles, torsions, planes, chiral volume, non-bonded contacts
    - Including links and modifications
- Provides “interactive” refinement
- Subject to substantial extension

Peptide Backbone Geometry



Low Resolution Model-Building

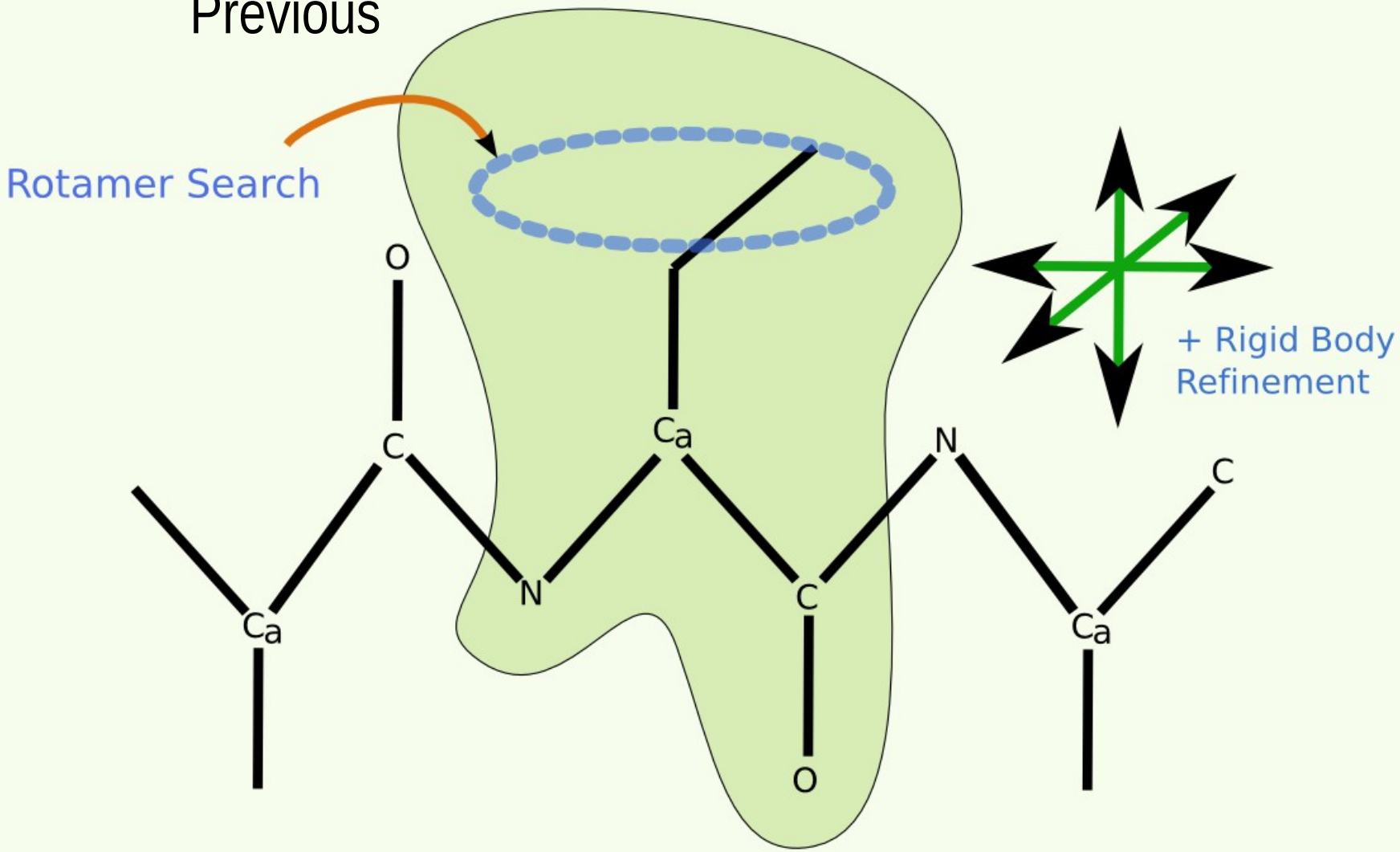
- “Backrub” rotamers

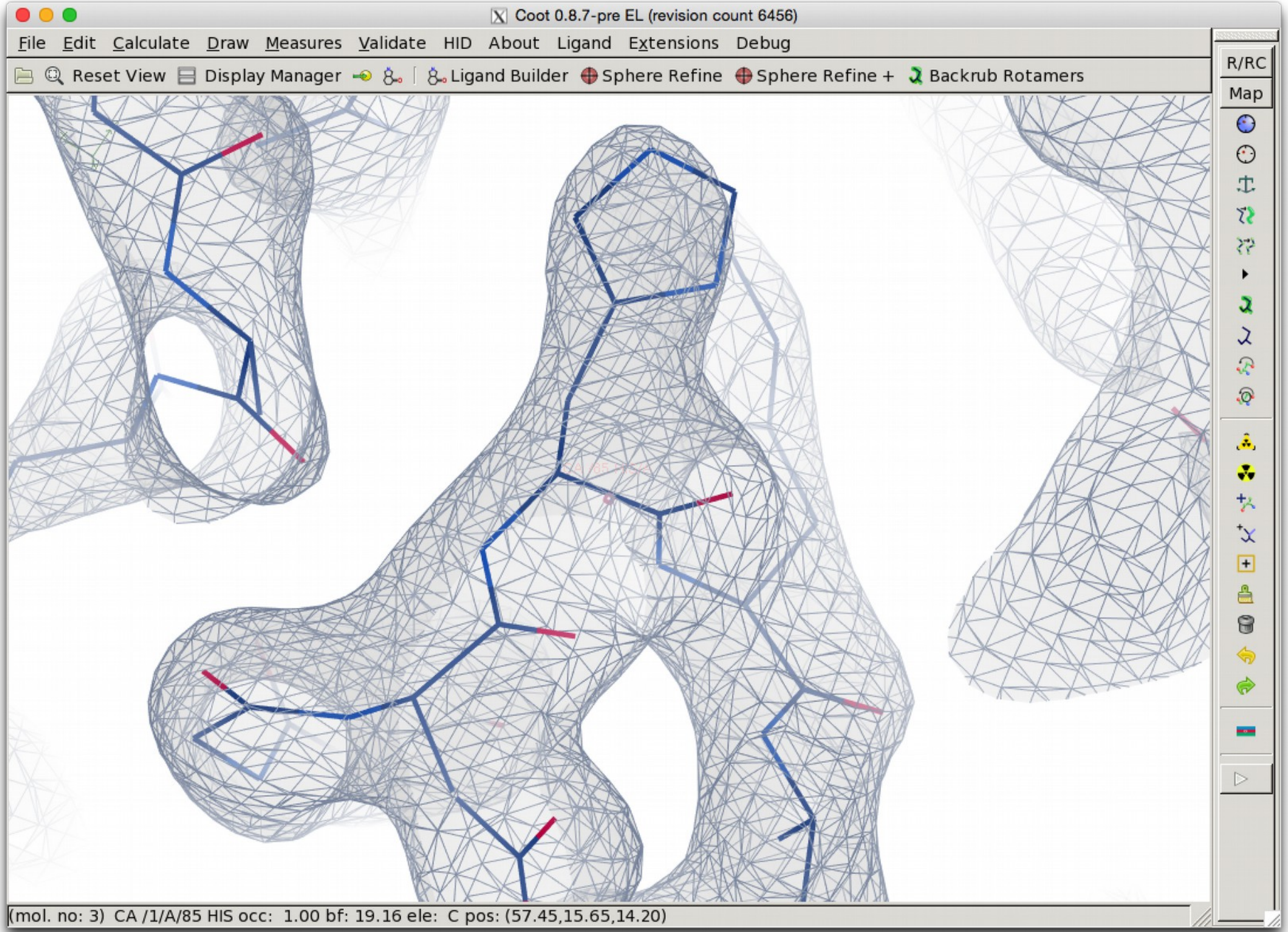
Rotamer Searching

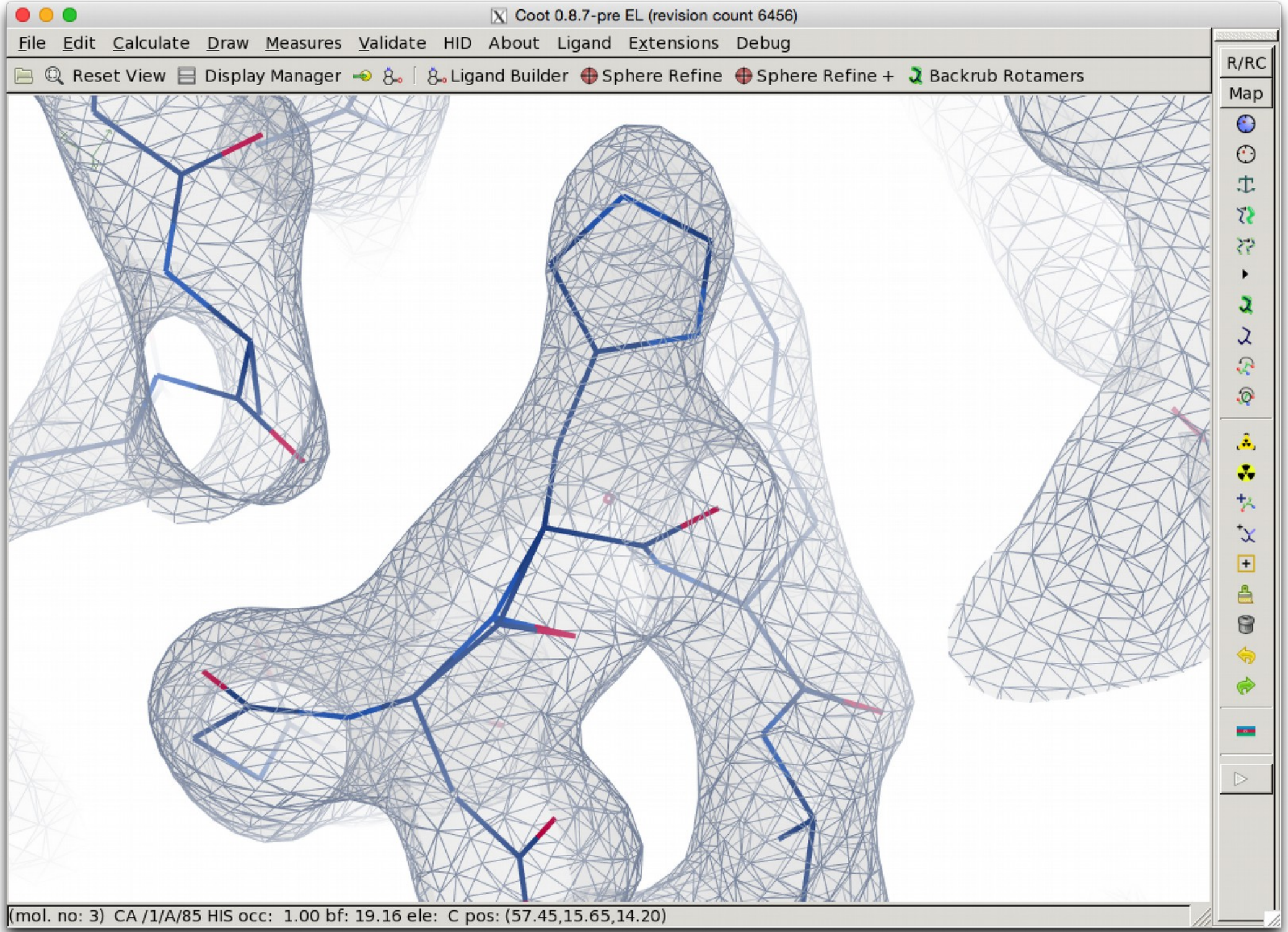
- Two methods
  - Traditional
  - Backrub

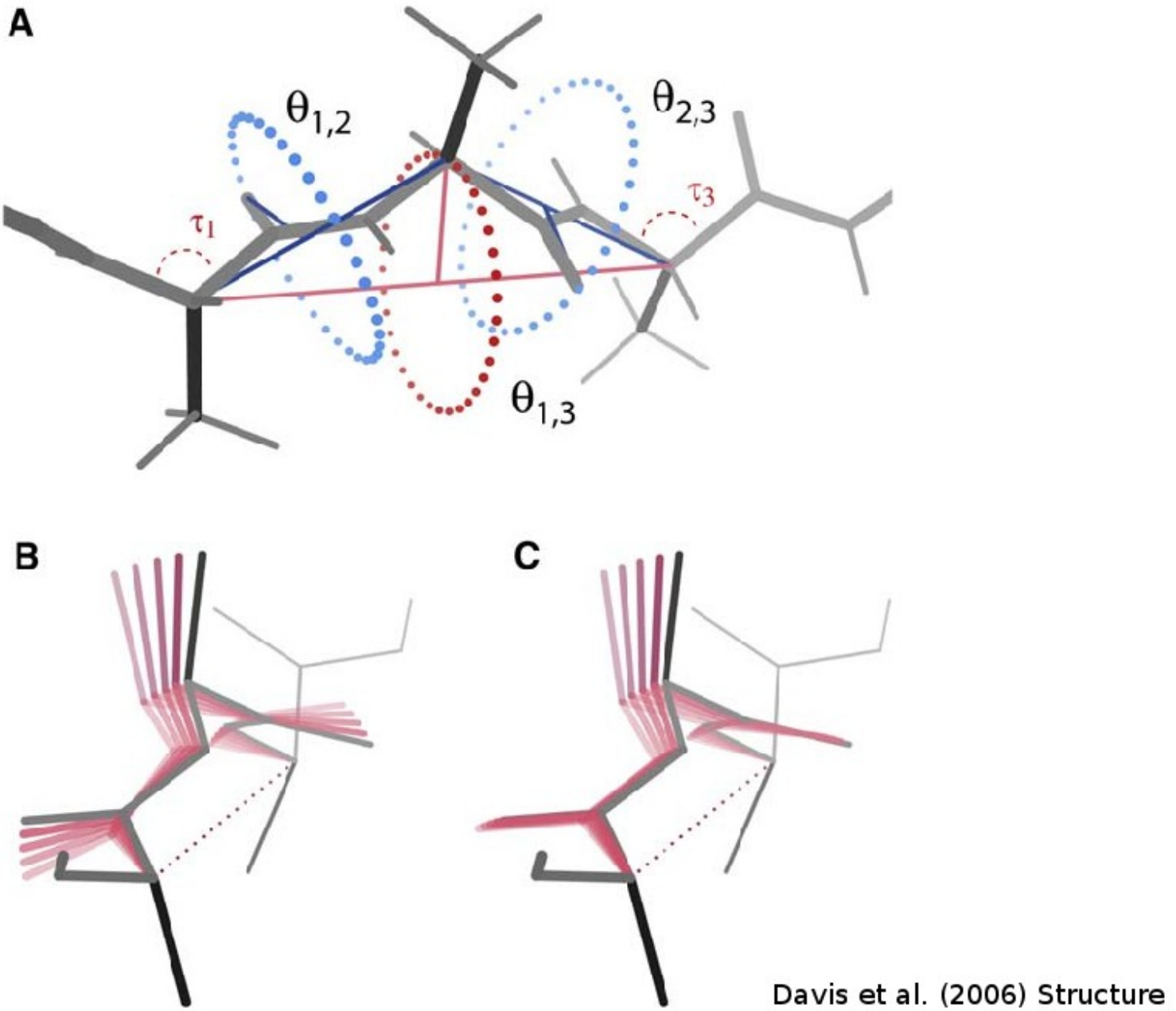
~~Current~~ Low Resolution Rotamer Search

Previous



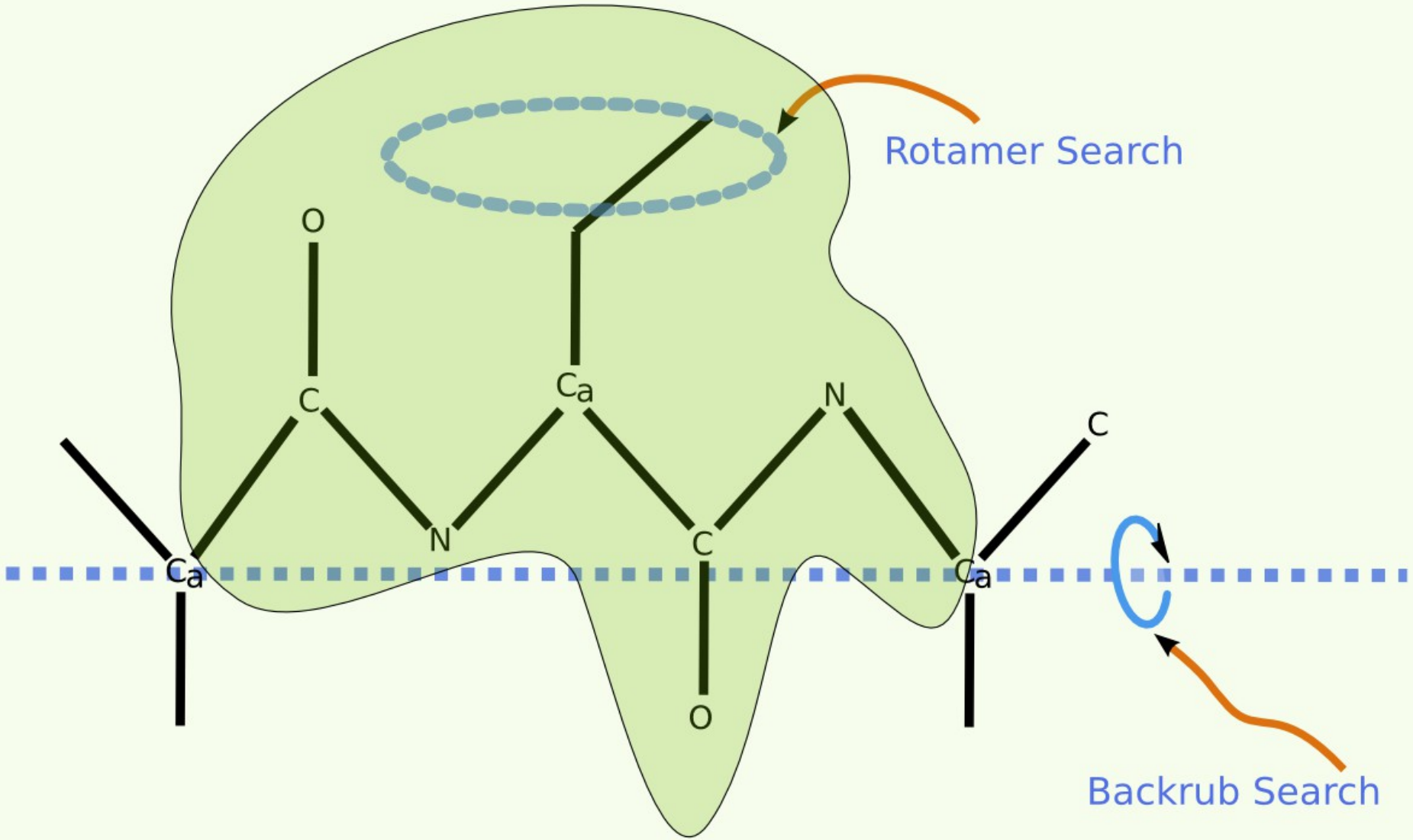

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)

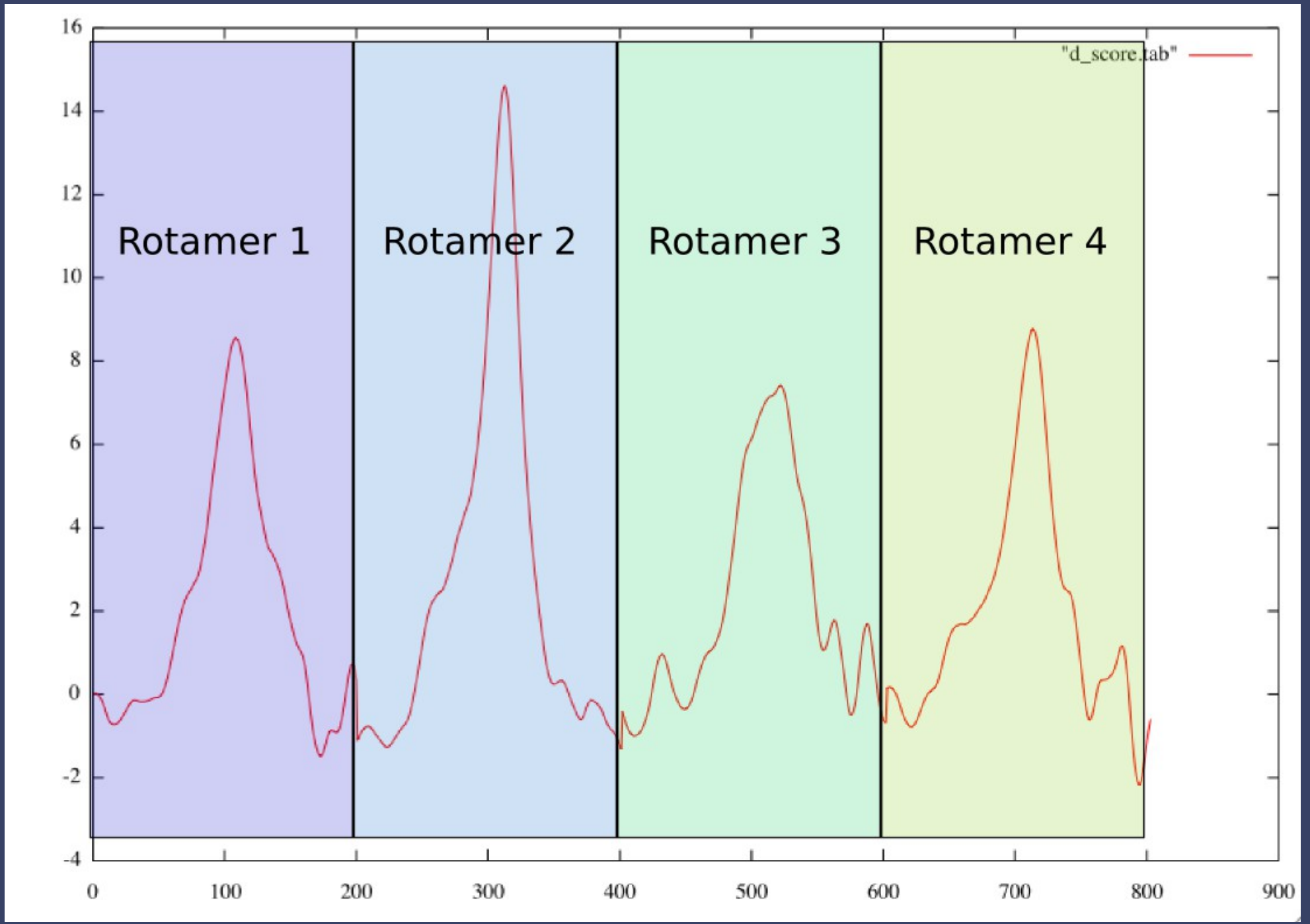


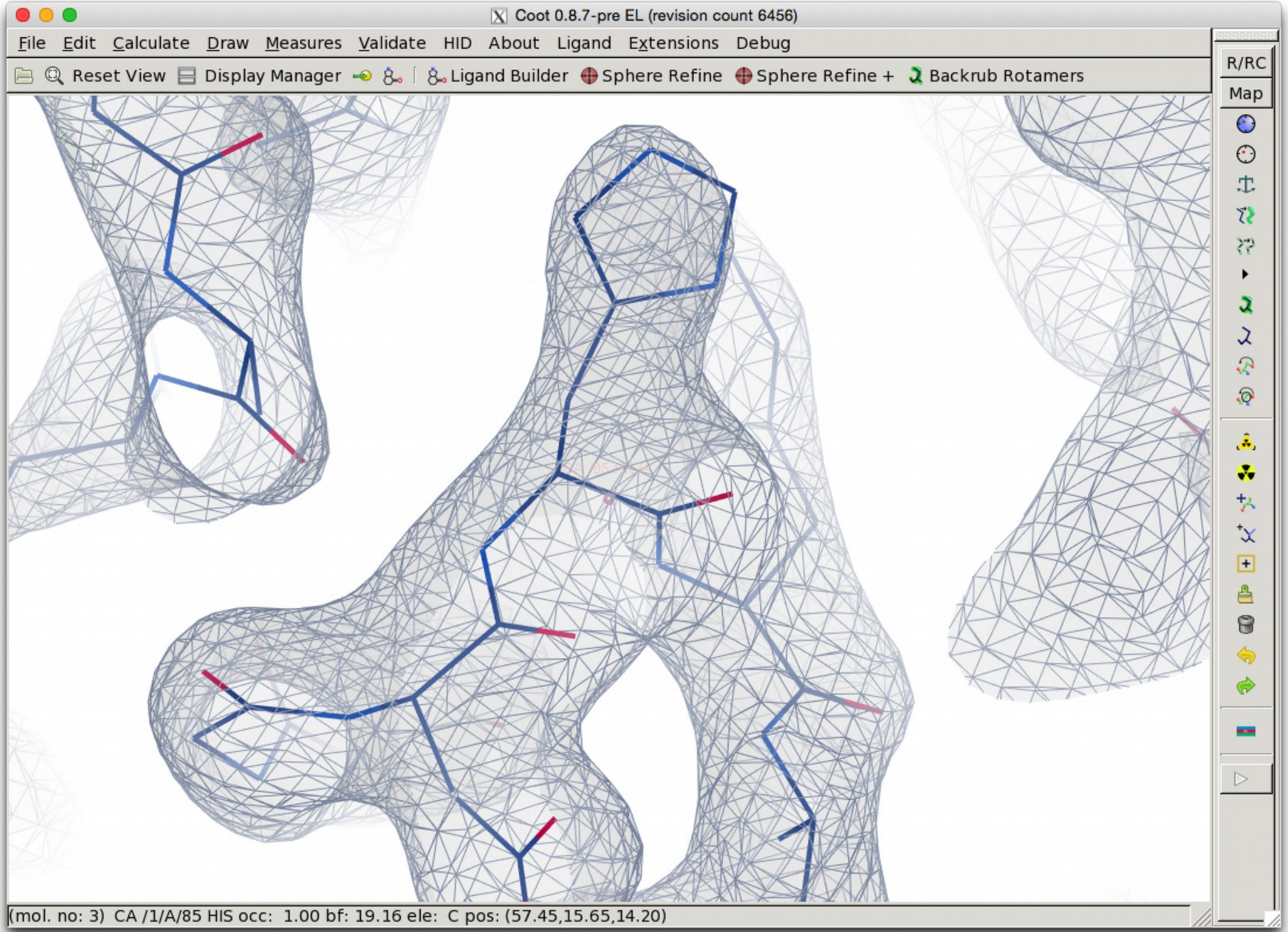
Davis et al. (2006) Structure

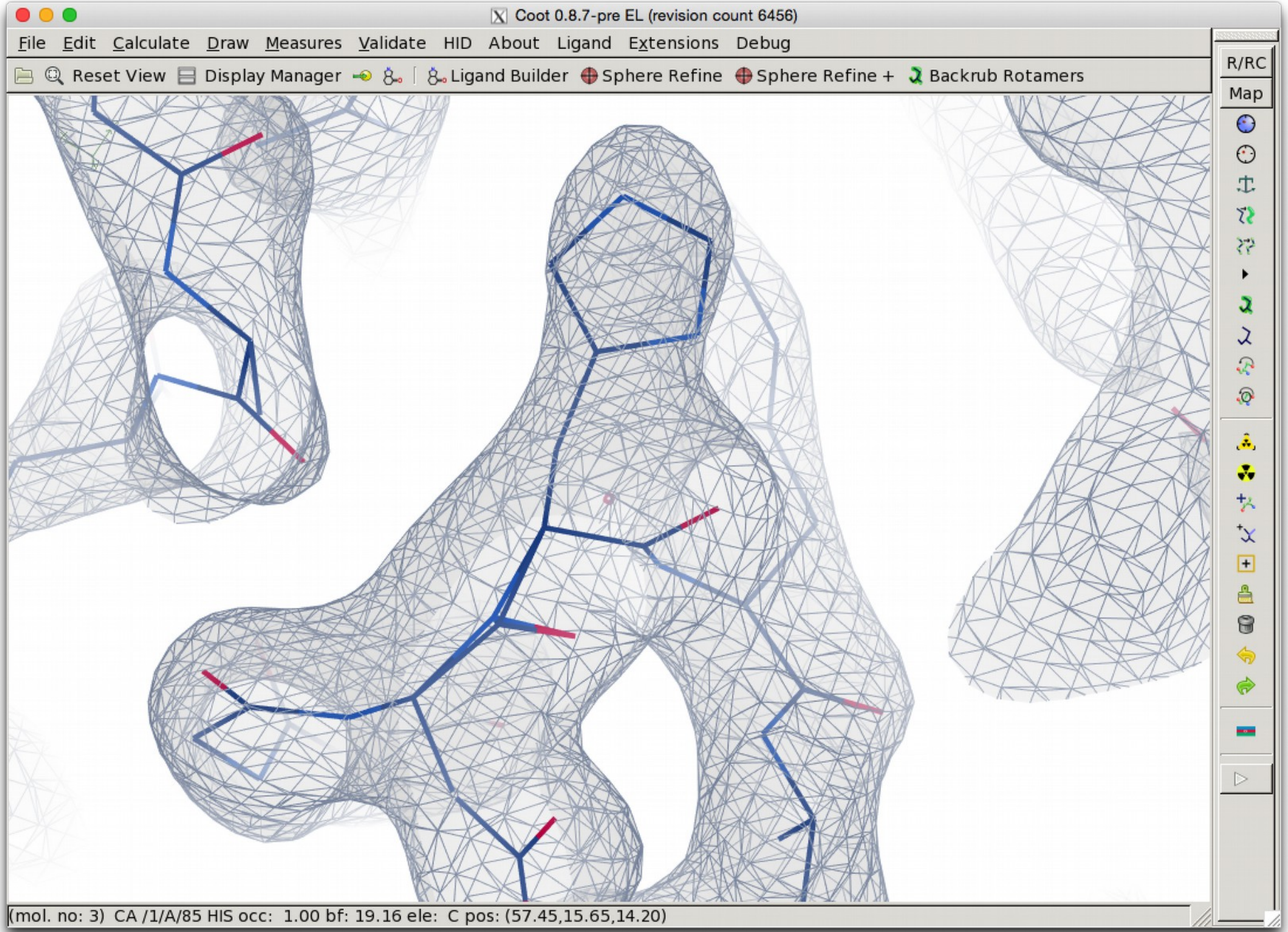
New Low Resolution Rotamer Search

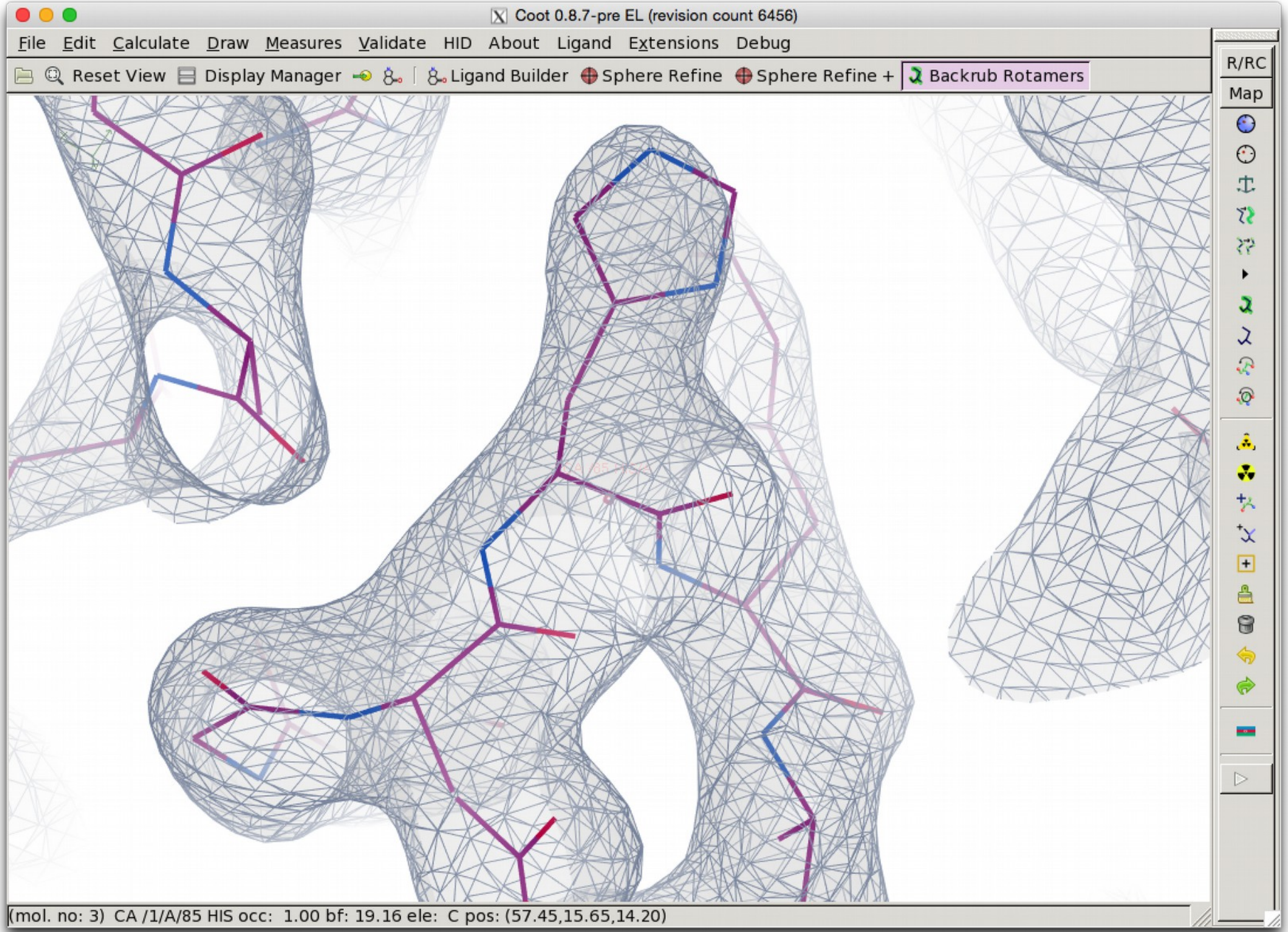


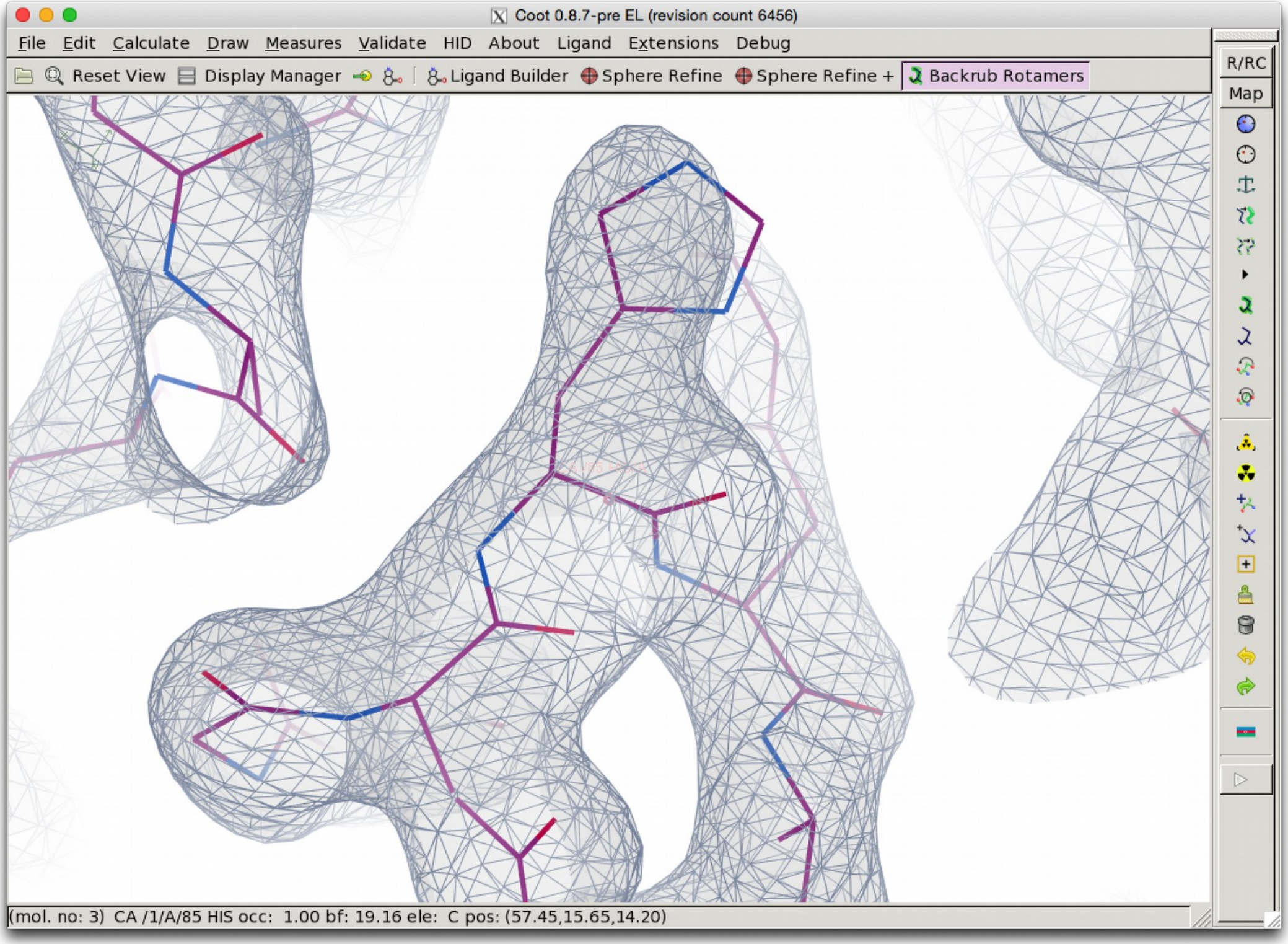
After Fitting Tools in KING/Molprobity

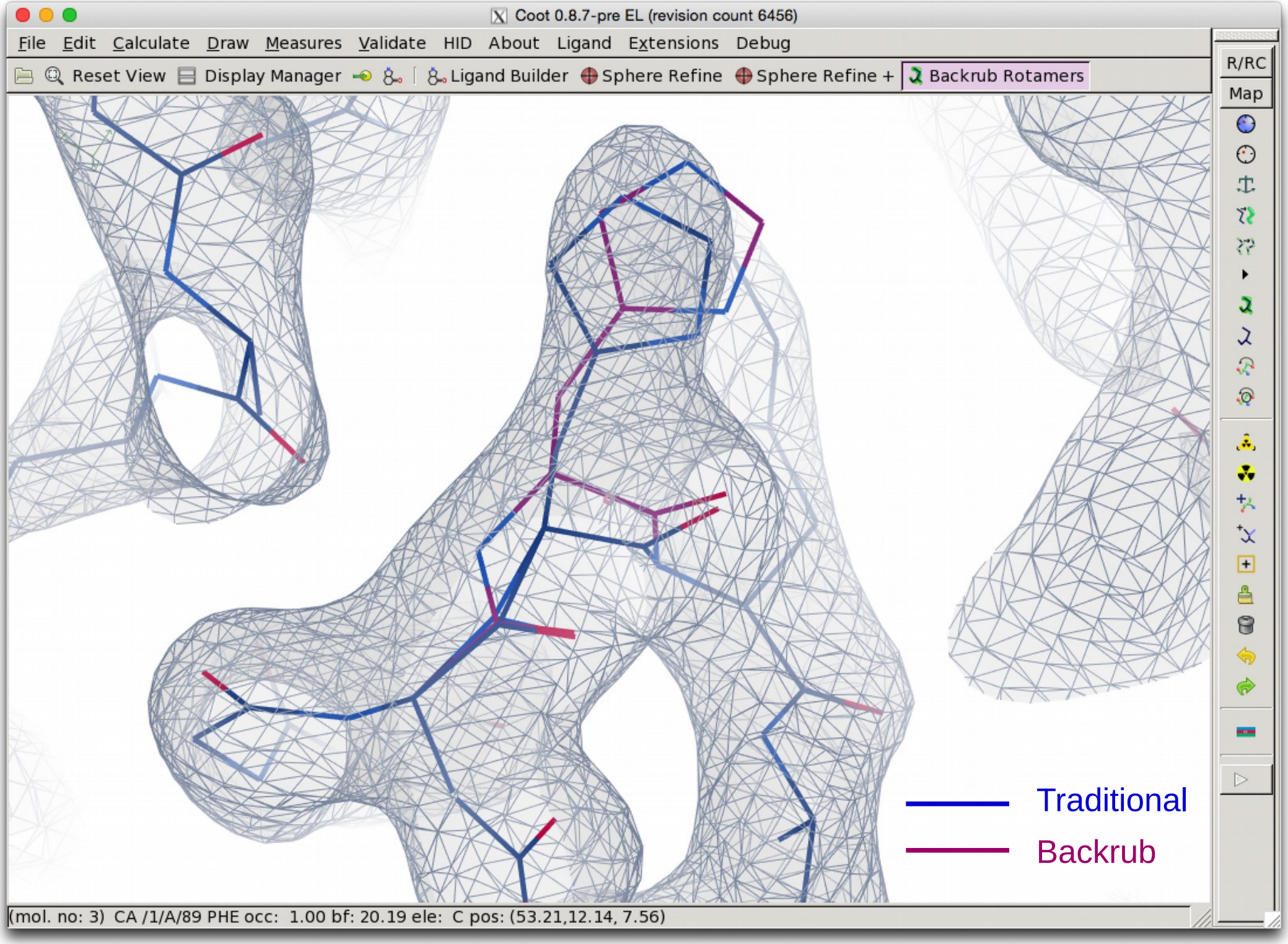

"d_score.tab"
Rotamer 1
Rotamer 2
Rotamer 3
Rotamer 4
16
14
12
10
8
6
4
2
0
-2
-4
0
100
200
300
400
500
600
700
800
900


Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)

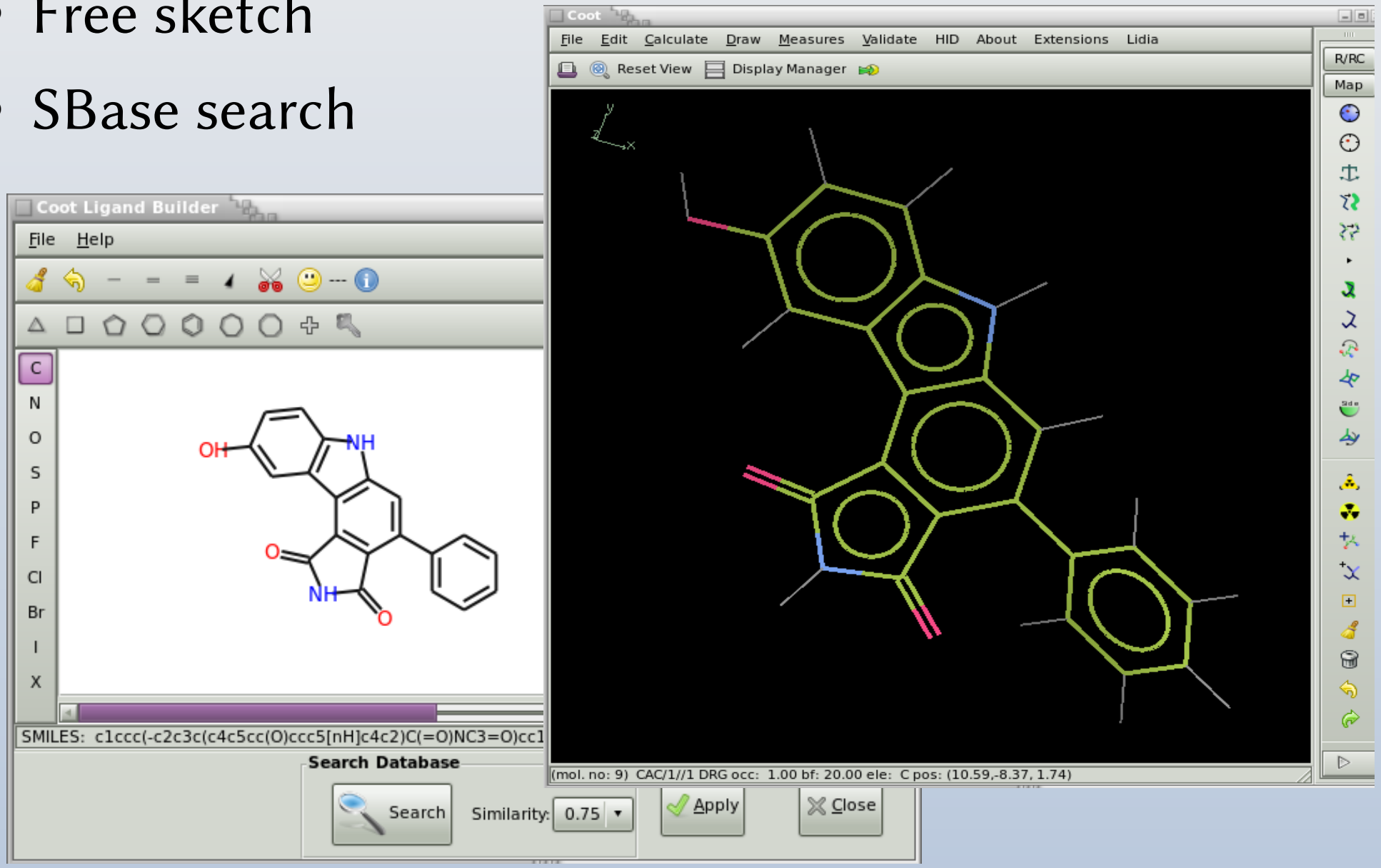

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
Traditional
Backrub
(mol. no: 3) CA /1/A/89 PHE occ: 1.00 bf: 20.19 ele: C pos: (53.21,12.14, 7.56)

Ligands

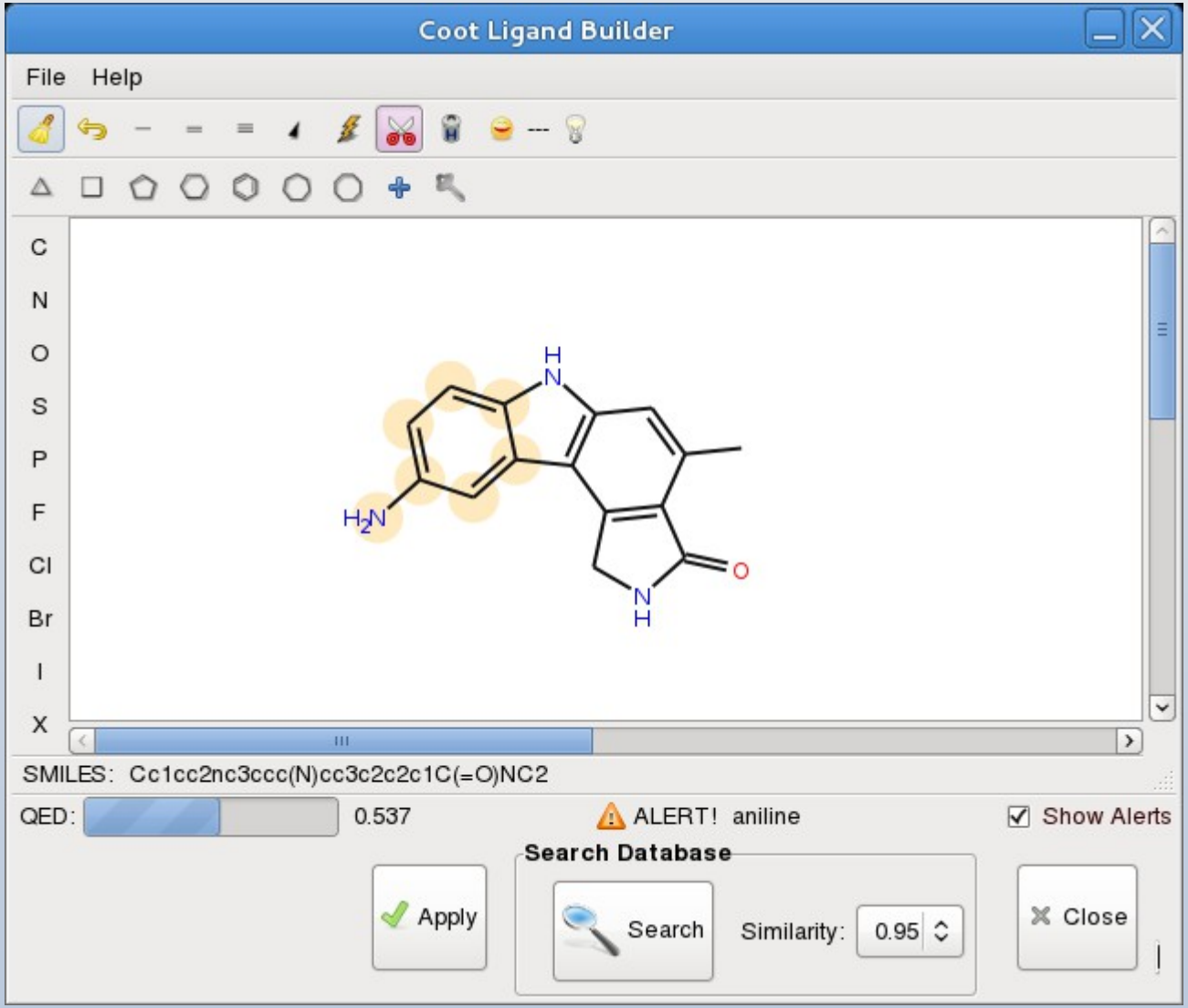
2D Ligand Builder

- 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

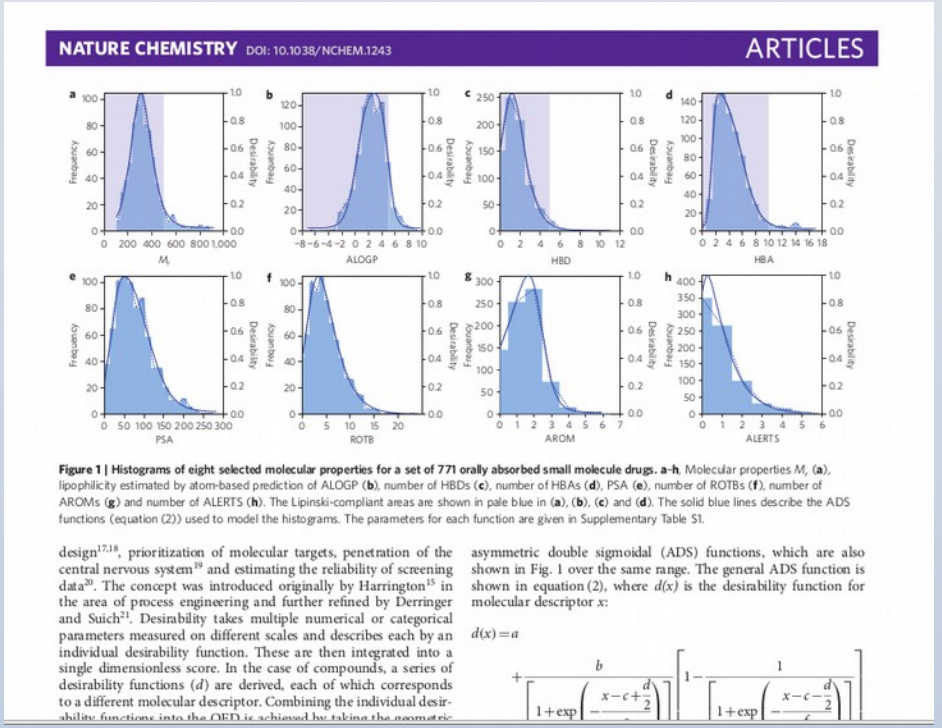


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h. Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**),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

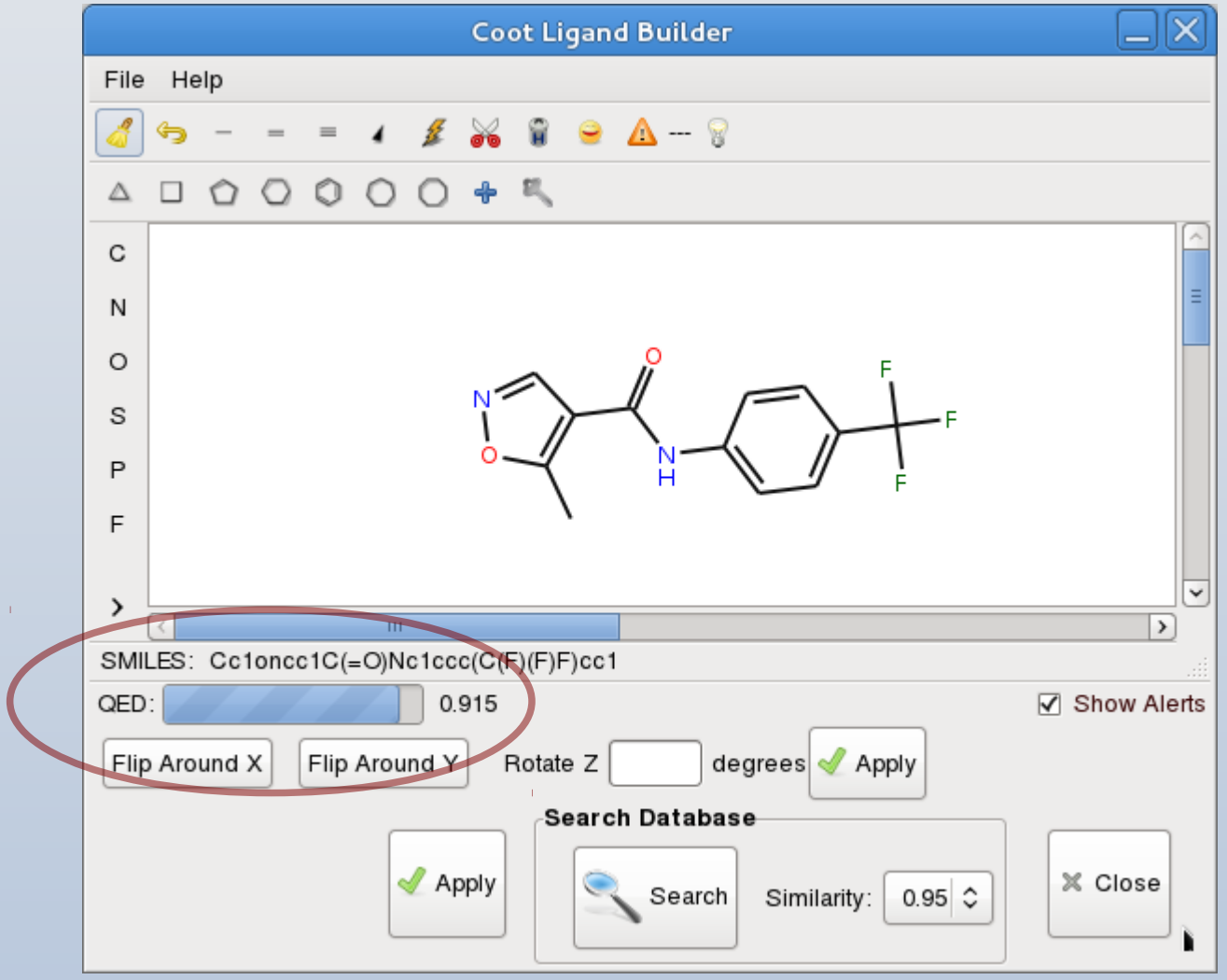
asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

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

Bickerton et al (2012) *Nature Chemistry*

2D Sketcher

- QED score



Silicos-it's Biscu-it™

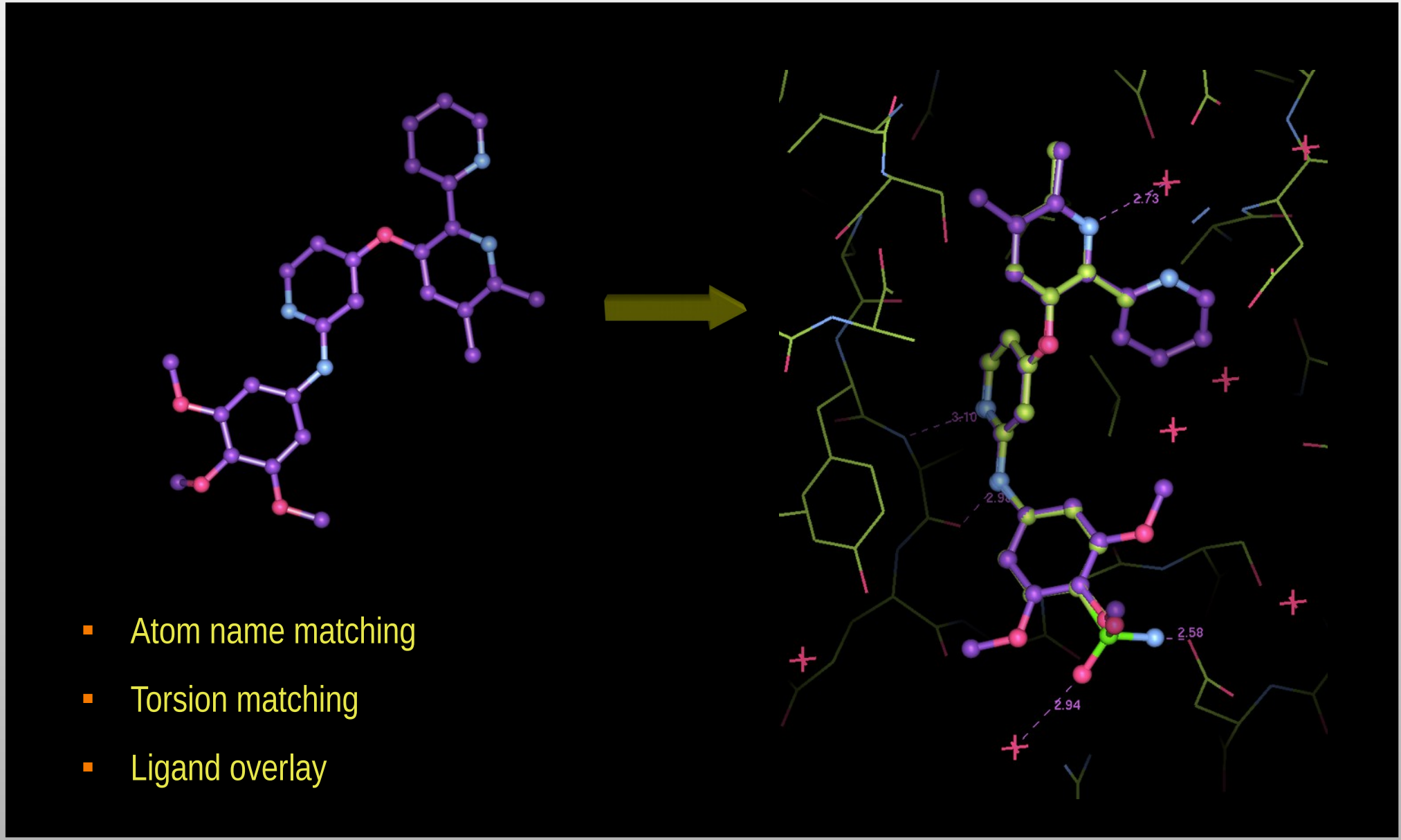
Look up the function with PyModule_GetDict() and PyModule_GetItem()

Ligand Utils

- "Fetch Molecule"
  - 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

Using ”Yesterday's” Ligand

Common subgraph isomorphism, Krissinel & Henrick (2004)



- Atom name matching
- Torsion matching
- Ligand overlay

Generating Conformers

- Using restraint information...

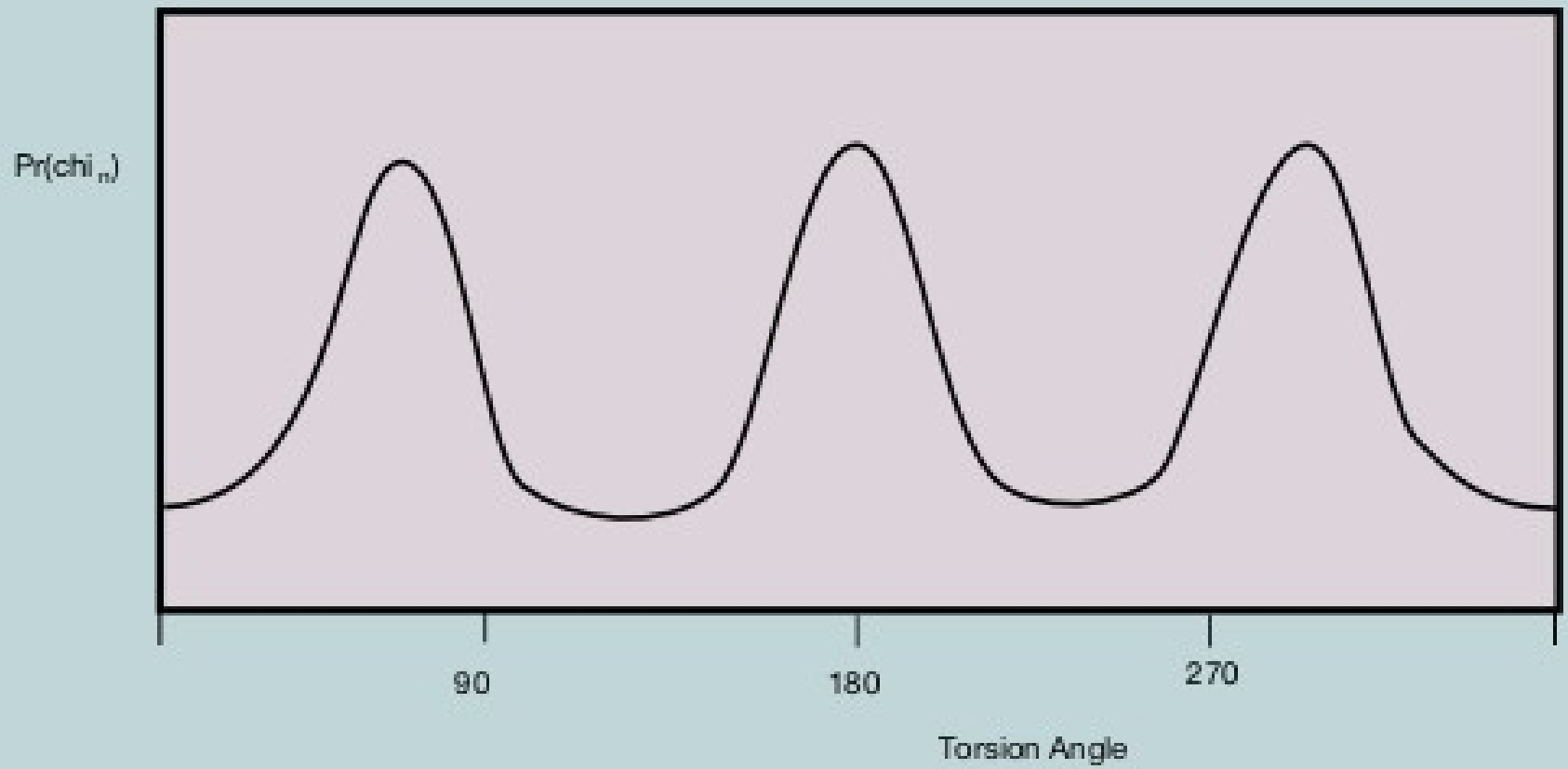
REFMAC Monomer Library
chem_comp_bond

```
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`

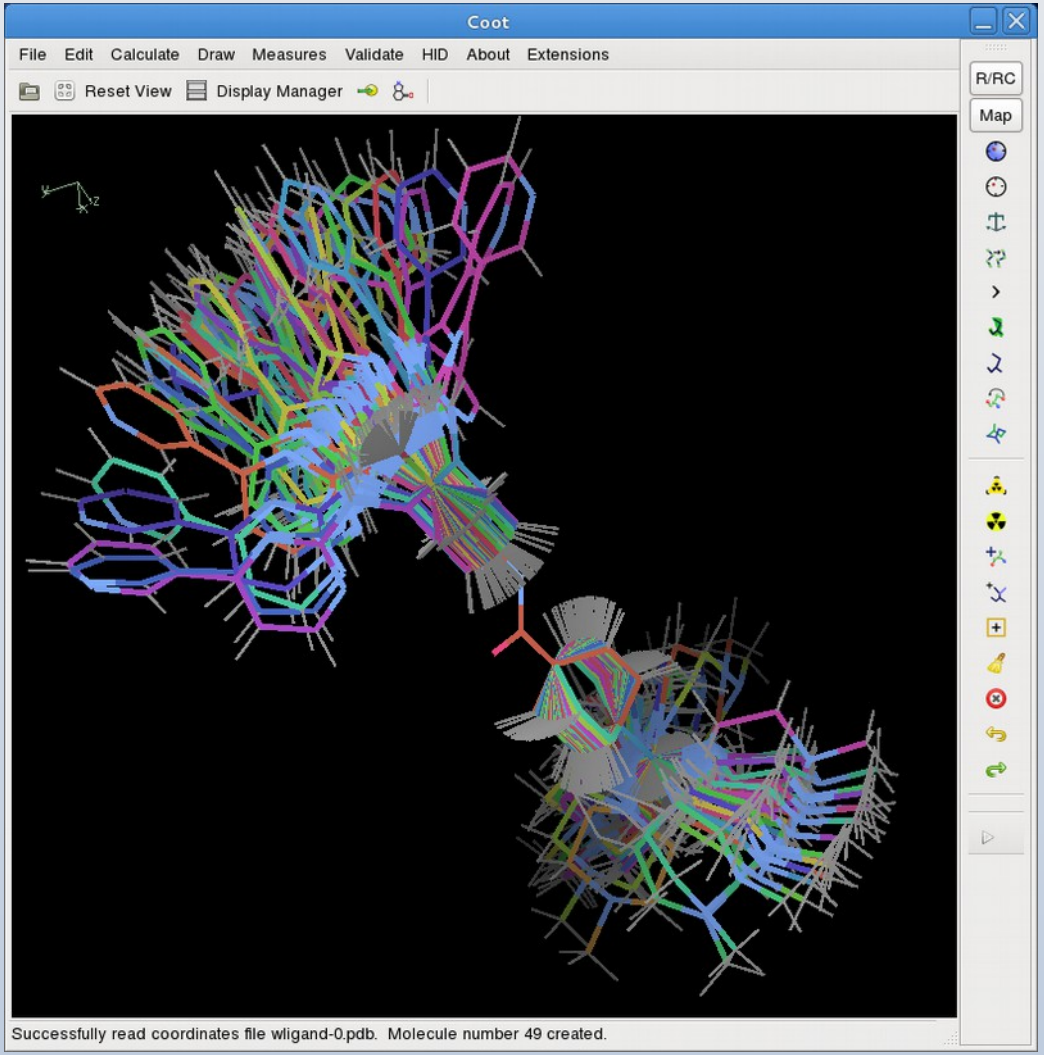
```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB       60.005   20.000   1
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

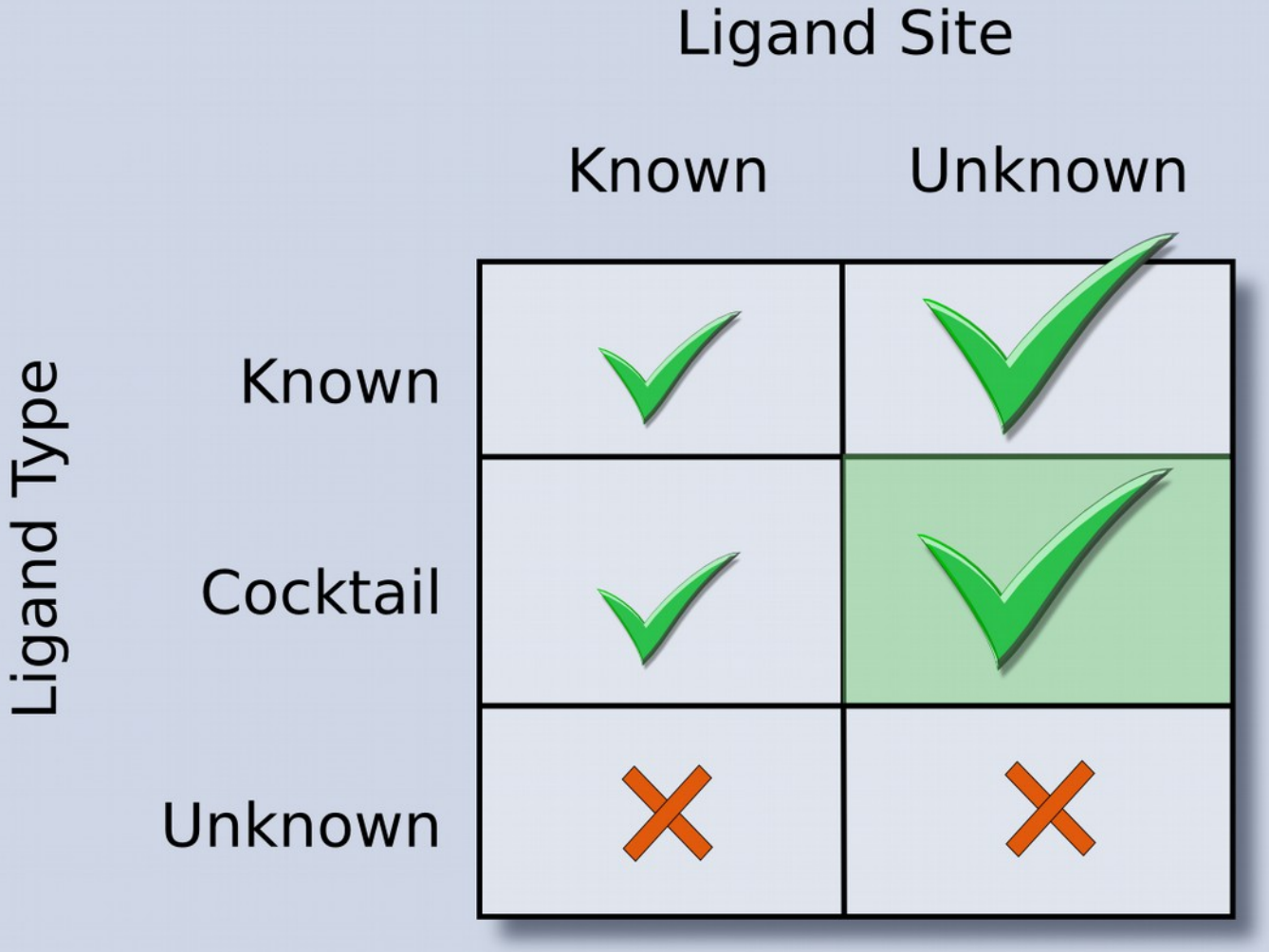


Conformer Generation

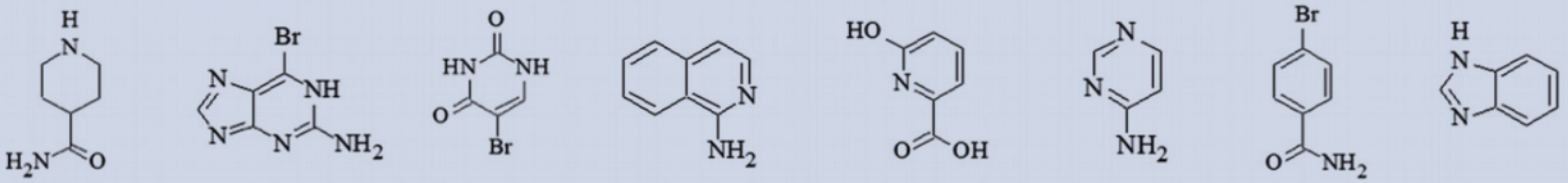
Non-Hydrogen
Non-CONST
Non-Ring



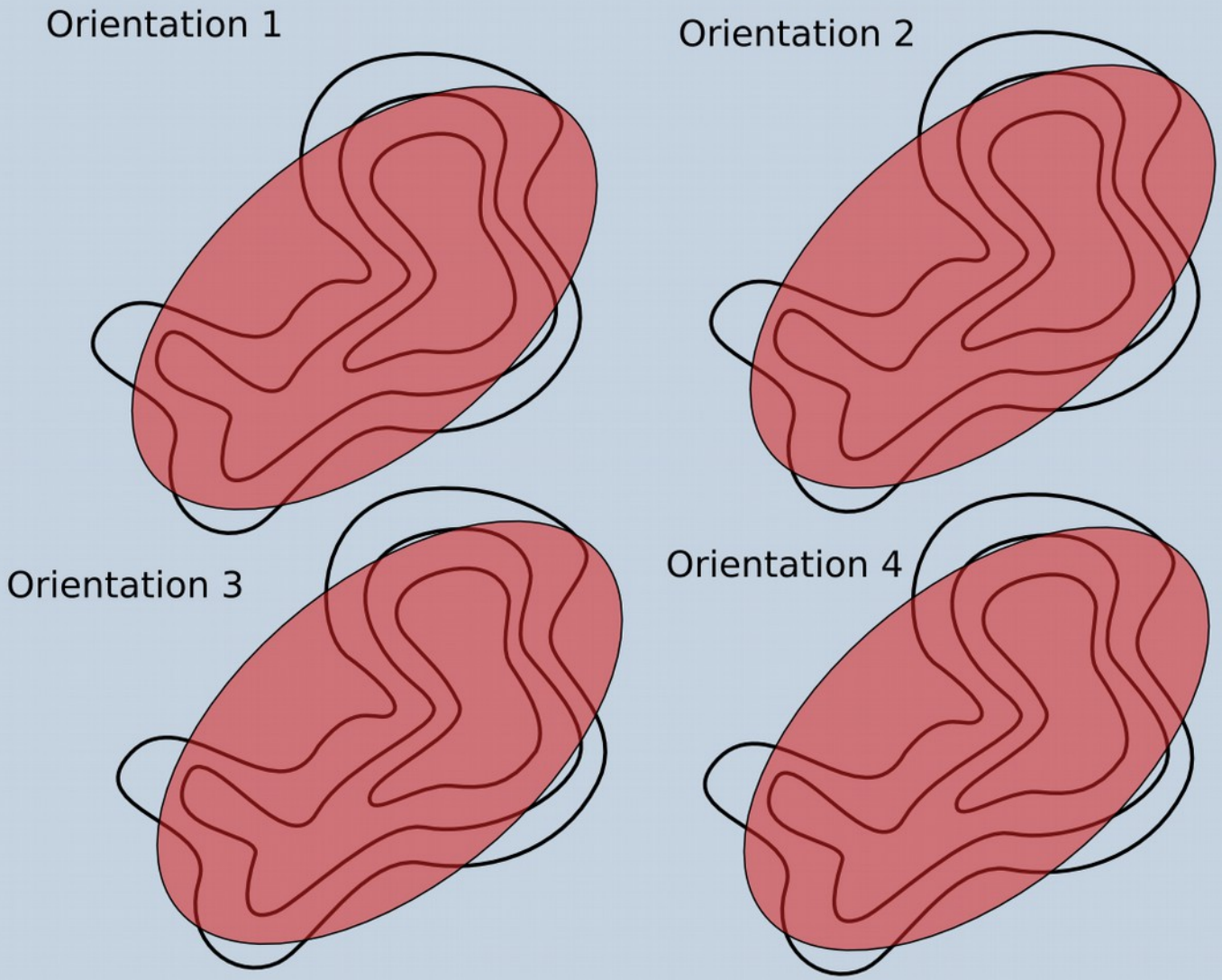
Fitting Ligands



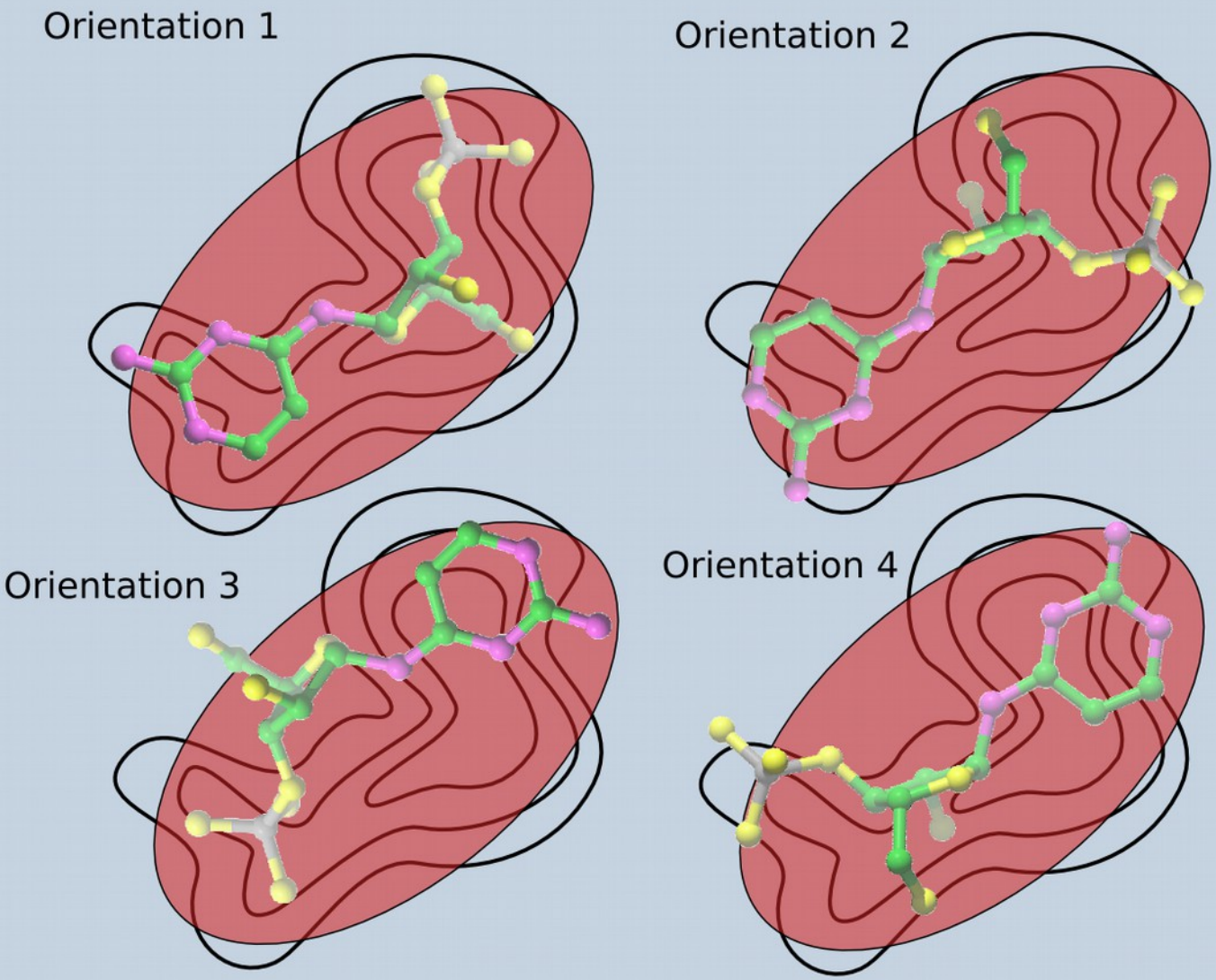
Cocktail Examples



Orienting the Ligand



Orienting the Ligand



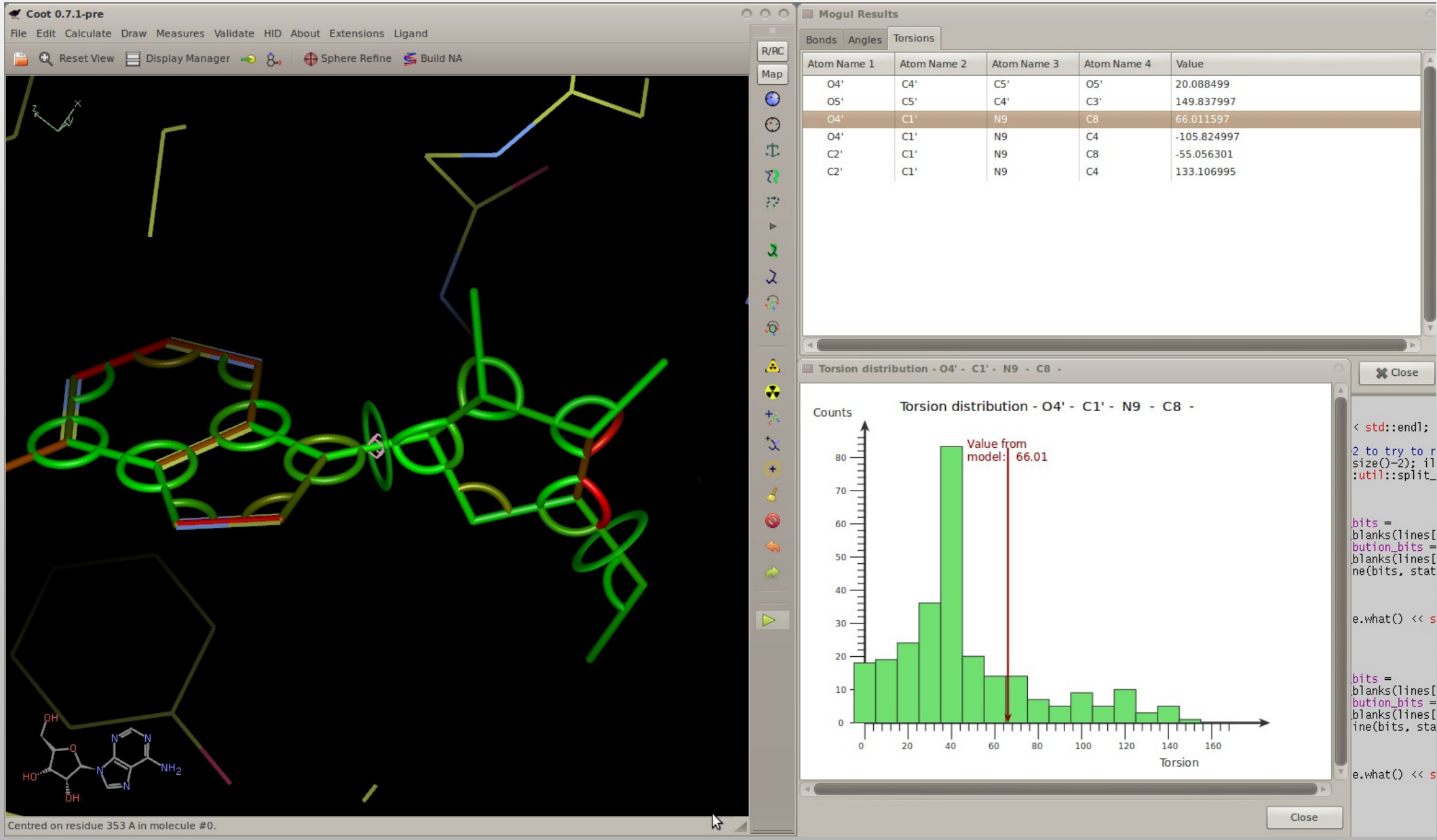
Ligand Validation

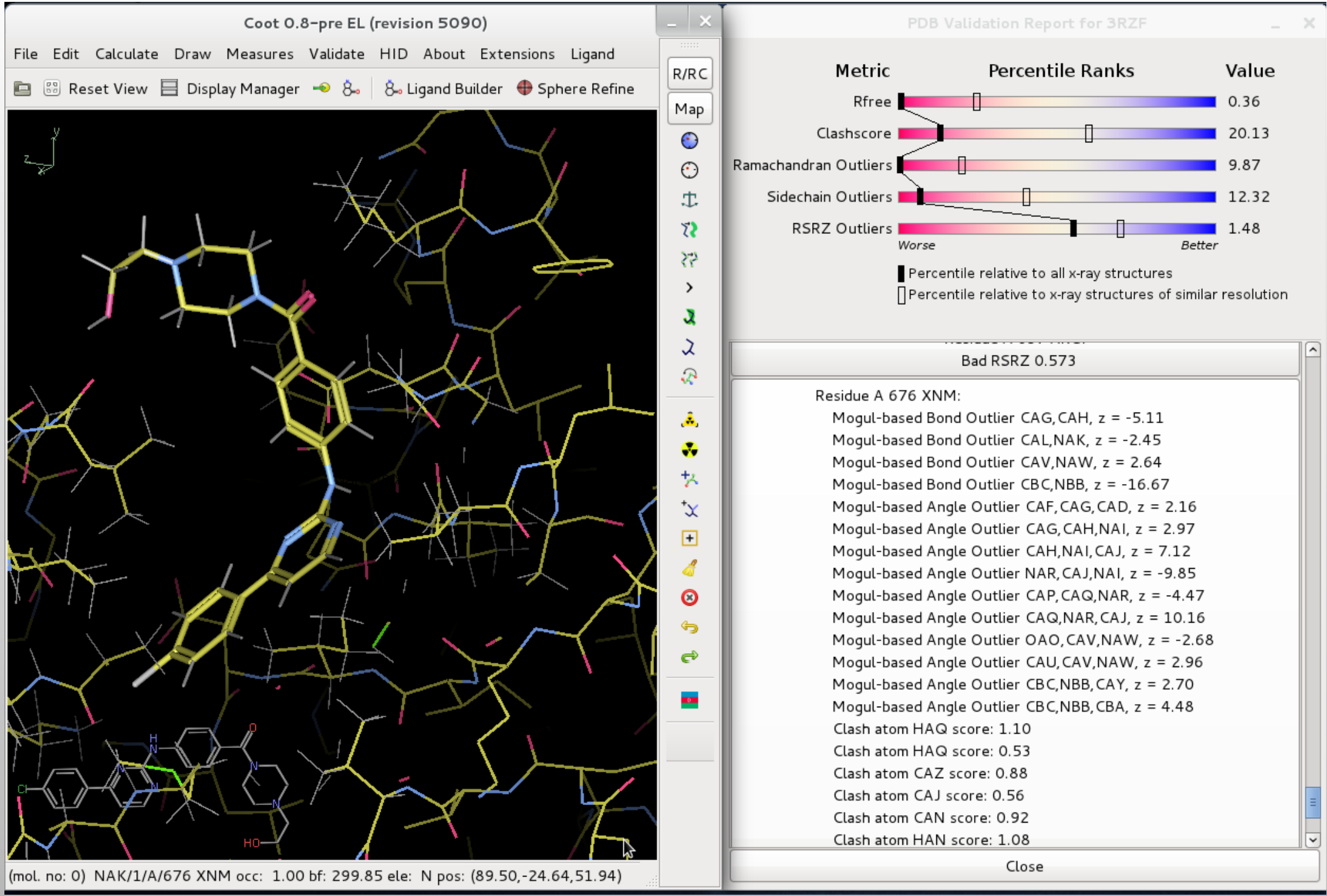
- Mogul plugin in Coot
  - Run mogul, graphical display of results
  - Update restraints (target and esds for bonds and angles)
  - CSD data not so great for plane, chiral and torsion restraints
    - (not by me, anyway)

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
```

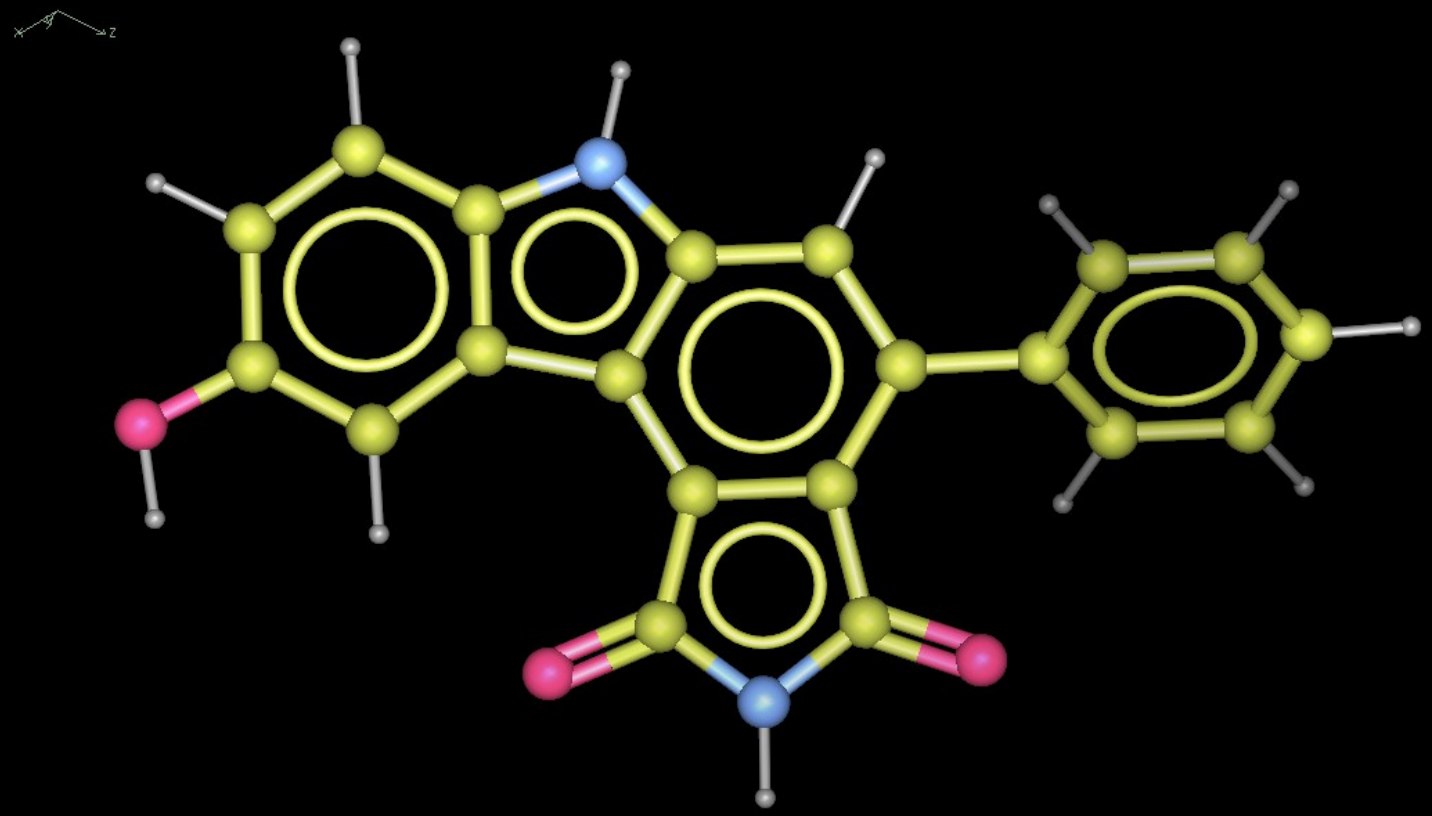
Mogul Results Representation



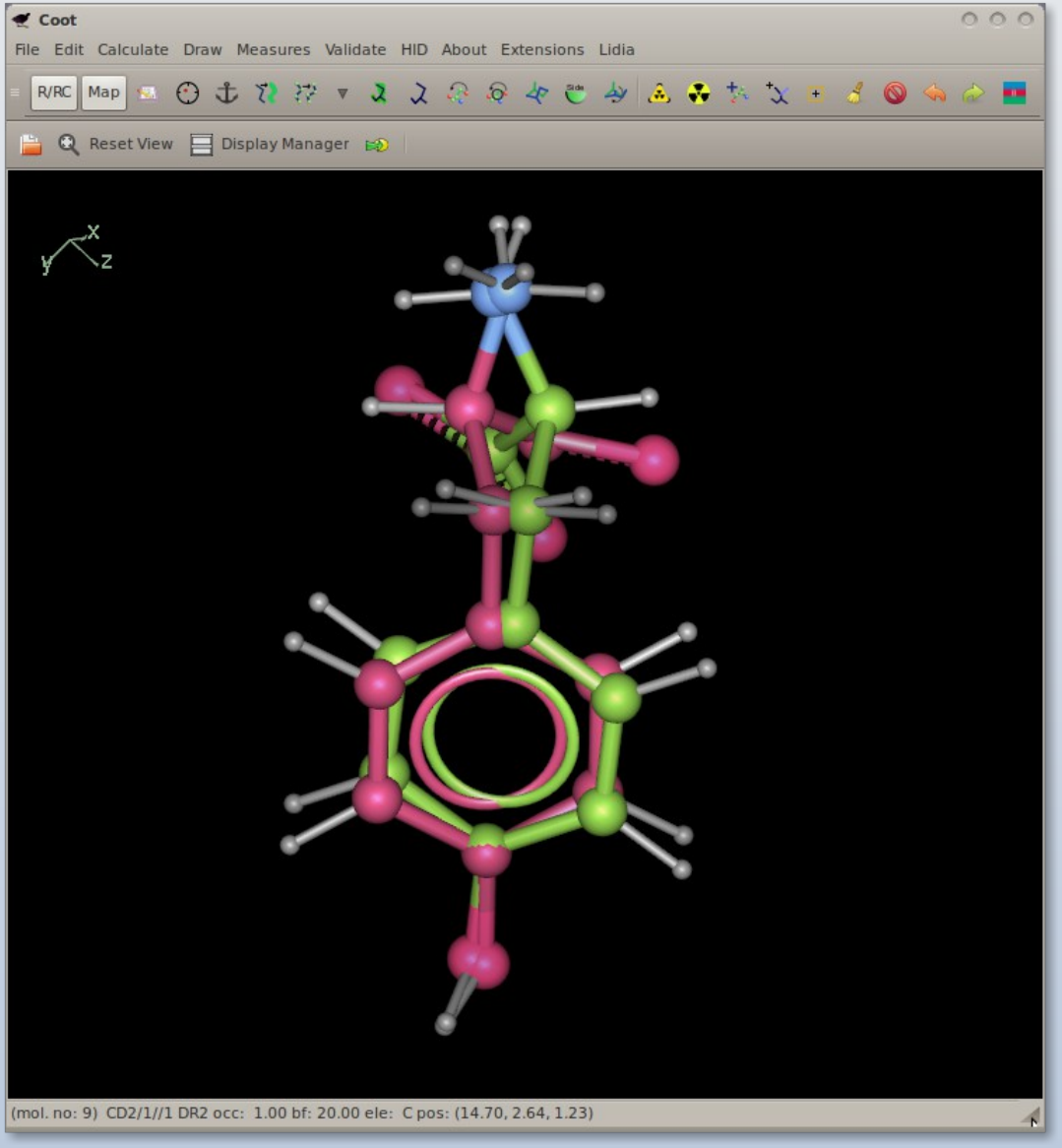

Coot 0.8-pre EL (revision 5090)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand
Reset View
Display Manager
Ligand Builder
Sphere Refine
R/RC
Map
(mol. no: 0) NAK/1/A/676 XNM occ: 1.00 bf: 299.85 ele: N pos: (89.50,-24.64,51.94)
PDB Validation Report for 3RZF
Metric
Percentile Ranks
Value
Rfree
0.36
Clashscore
20.13
Ramachandran Outliers
9.87
Sidechain Outliers
12.32
RSRZ Outliers
1.48
Worse
Better
Percentile relative to all x-ray structures
Percentile relative to x-ray structures of similar resolution
Bad RSRZ 0.573
Residue A 676 XNM:
Mogul-based Bond Outlier CAG,CAH, z = -5.11
Mogul-based Bond Outlier CAL,NAK, z = -2.45
Mogul-based Bond Outlier CAV,NAW, z = 2.64
Mogul-based Bond Outlier CBC,NBB, z = -16.67
Mogul-based Angle Outlier CAF,CAG,CAD, z = 2.16
Mogul-based Angle Outlier CAG,CAH,NAI, z = 2.97
Mogul-based Angle Outlier CAH,NAI,CAJ, z = 7.12
Mogul-based Angle Outlier NAR,CAJ,NAI, z = -9.85
Mogul-based Angle Outlier CAP,CAQ,NAR, z = -4.47
Mogul-based Angle Outlier CAQ,NAR,CAJ, z = 10.16
Mogul-based Angle Outlier OAO,CAV,NAW, z = -2.68
Mogul-based Angle Outlier CAU,CAV,NAW, z = 2.96
Mogul-based Angle Outlier CBC,NBB,CAY, z = 2.70
Mogul-based Angle Outlier CBC,NBB,CBA, z = 4.48
Clash atom HAQ score: 1.10
Clash atom HAQ score: 0.53
Clash atom CAZ score: 0.88
Clash atom CAJ score: 0.56
Clash atom CAN score: 0.92
Clash atom HAN score: 1.08
Close

Ligand Represenation

- Bond orders (from dictionary restraints)



Chiral Centre Inversion

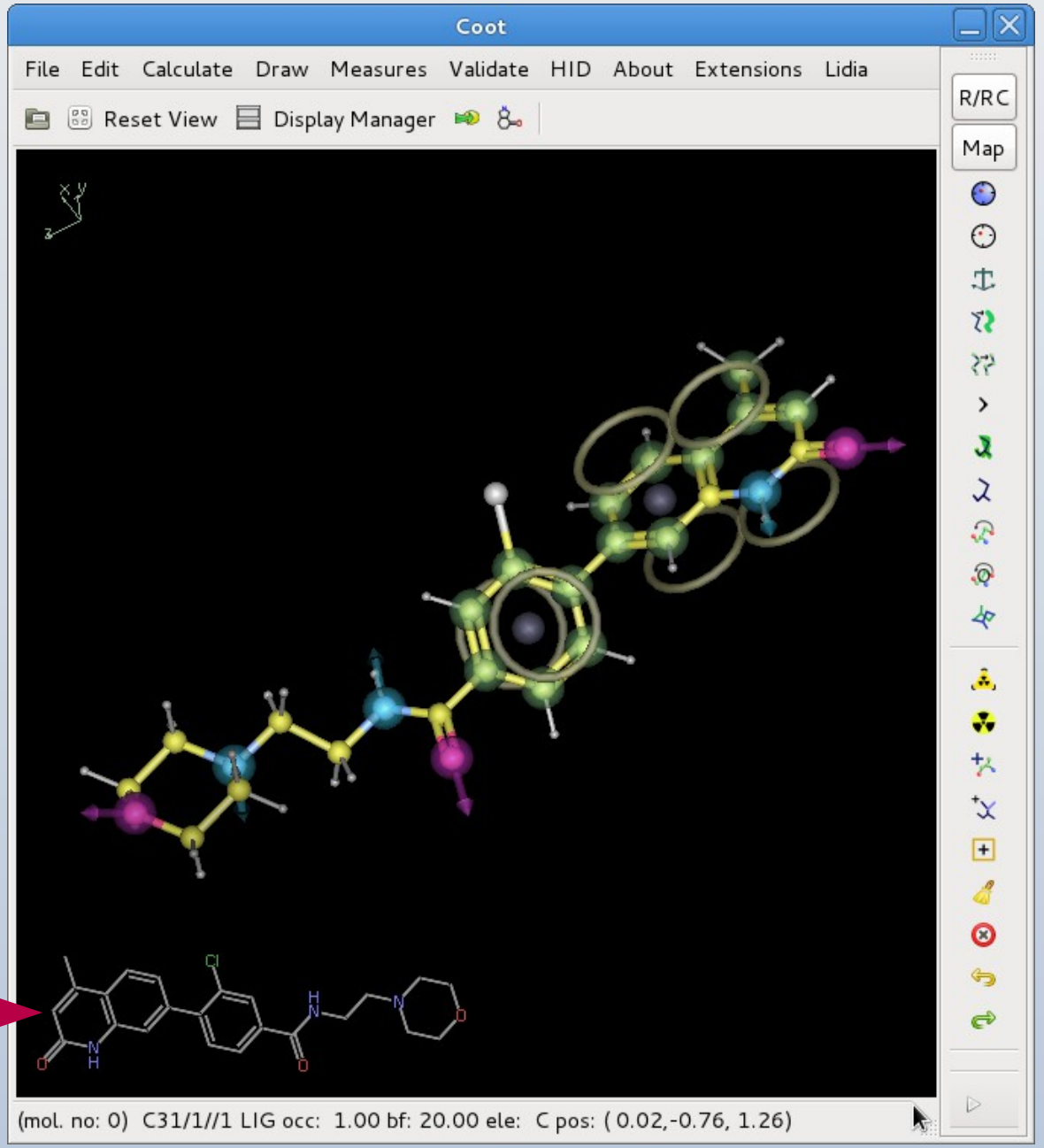


Inverted chiral centre refinement pathology detection

Hydrogen tunnelling

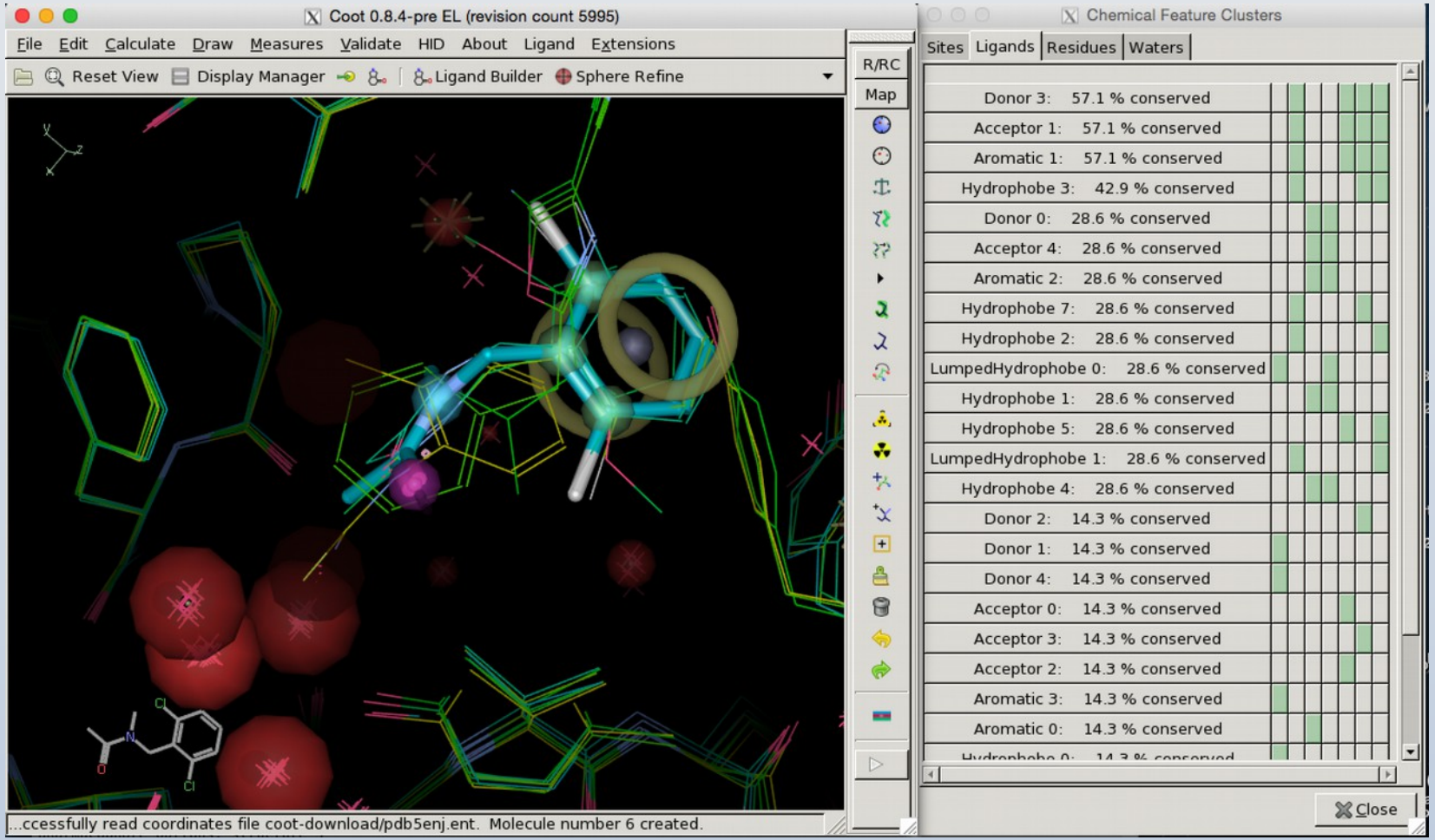
Chemical Features

Uses built-in FeatureFactory



...and on the fly thumbnailing

Conserved Pharmacophores

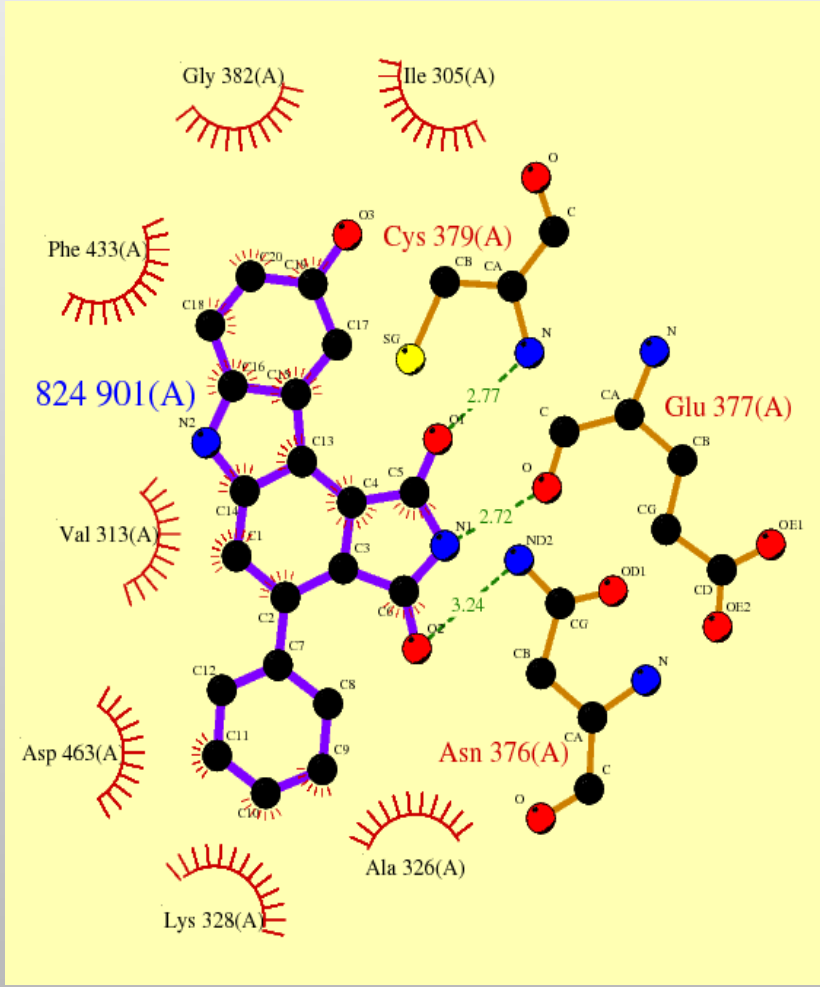


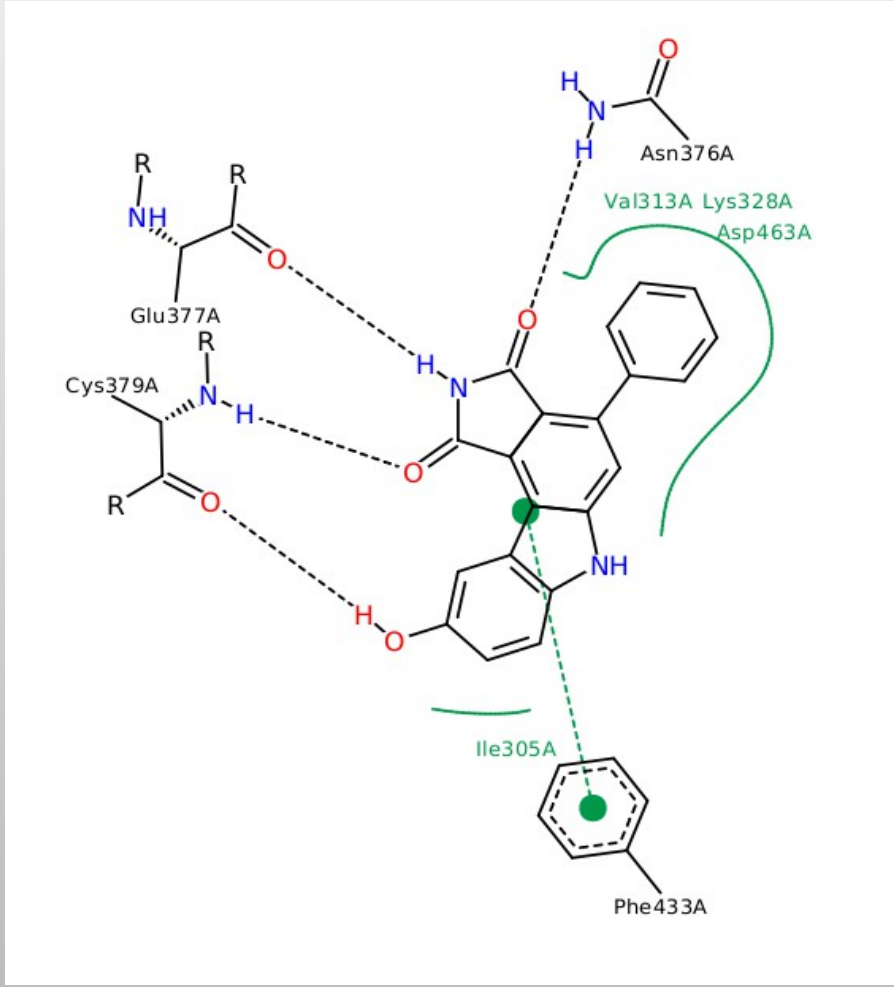
- **Acedrg:**
  - Structural database is the Crystallography Online Database
  - Bond and angle table generation
  - Use tables to generate dictionaries
    - Given a molecular description (input MDL mol, mol2, SMILES)
  - Fei Long (Murshudov Group)

- **Pyrogen:**
  - Based on:
    - Refmac Monomer Library Base Tables
    - MMFF94s Forcefield
    - CCDC Mogul
  - Available with *Coot*

Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





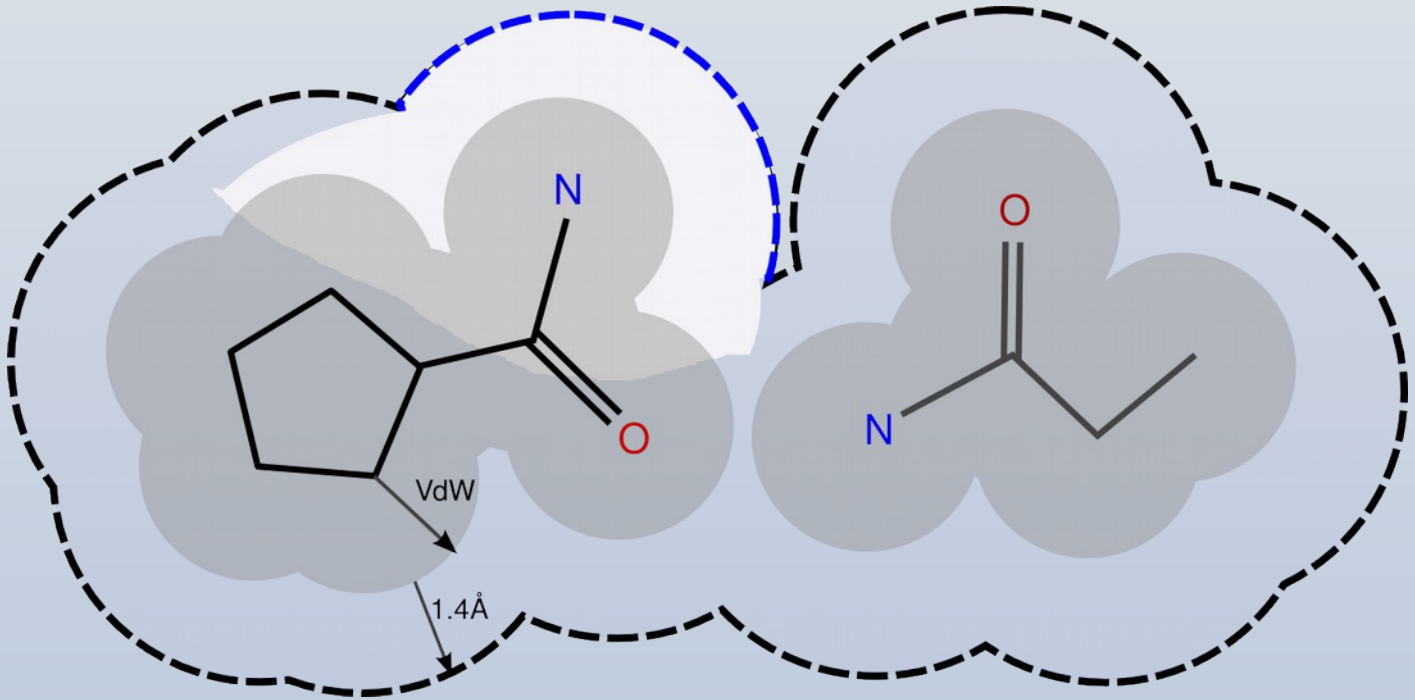
Can we do better? - Interactivity?

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

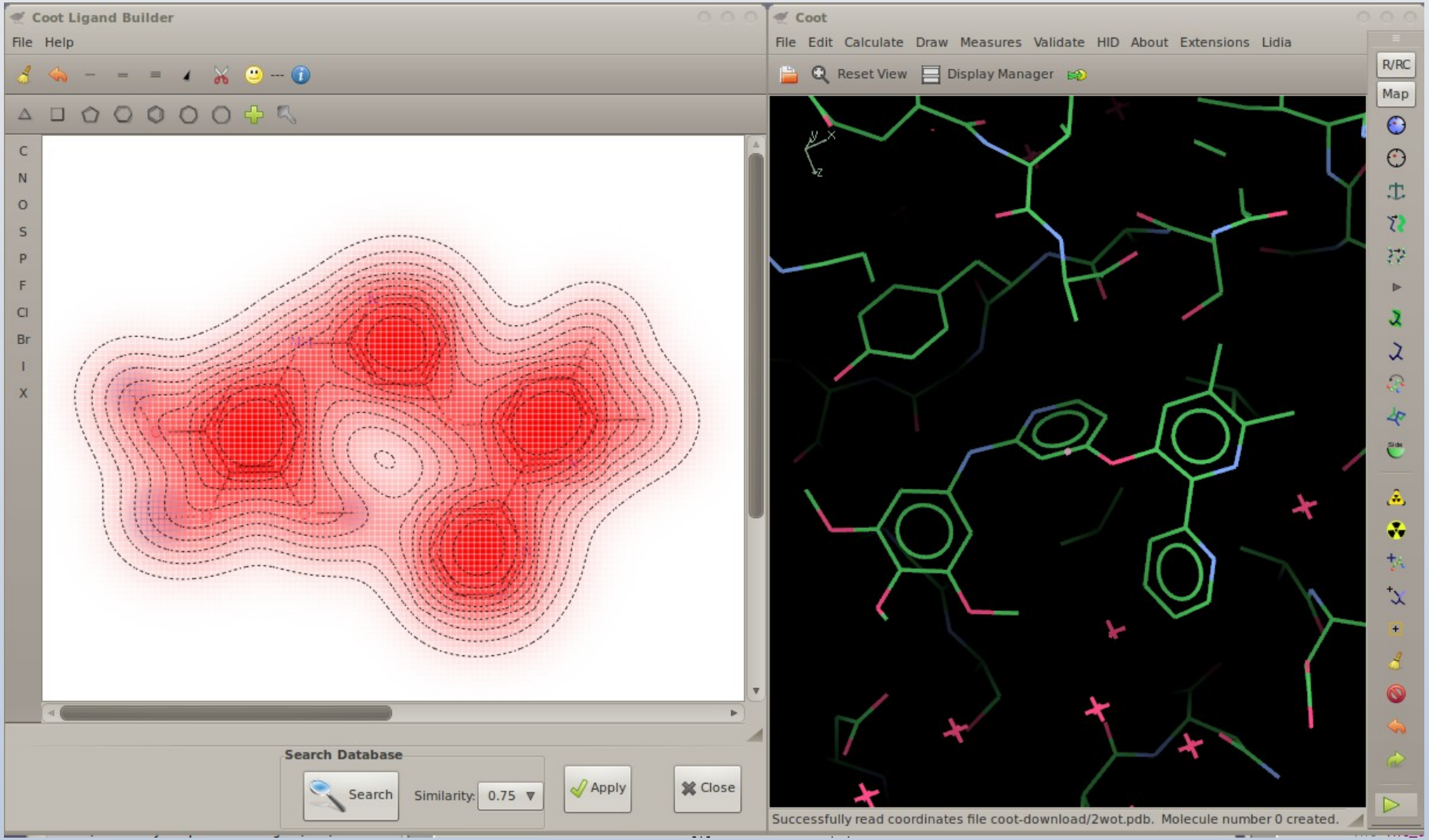
$$\sum\sum (d_{ik}^2 - D_{ik}^2) +$$

Residues are close to H-bonding ligand atoms

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

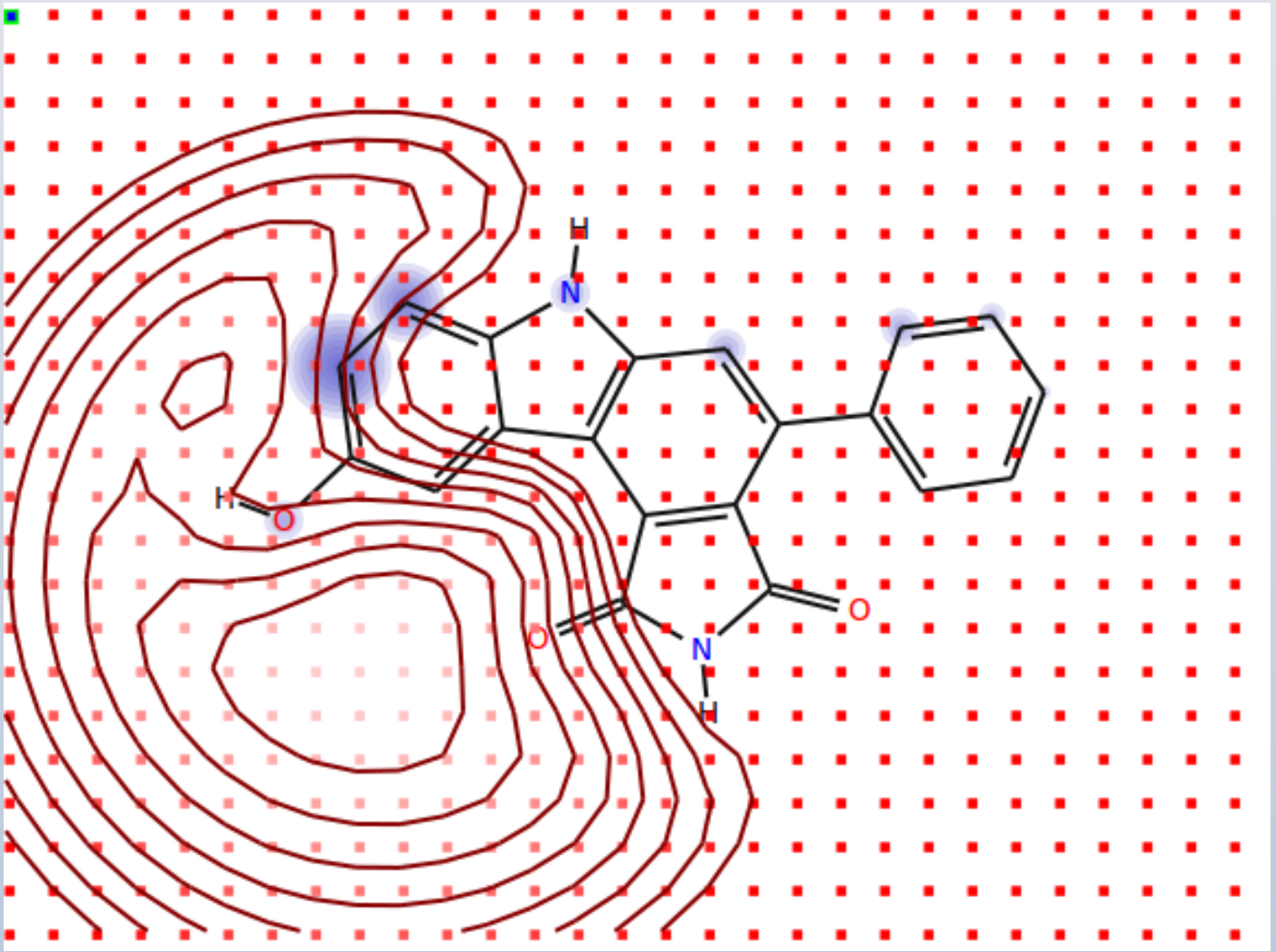
Residues don't overlap ligand

”Don't overlap the ligand”



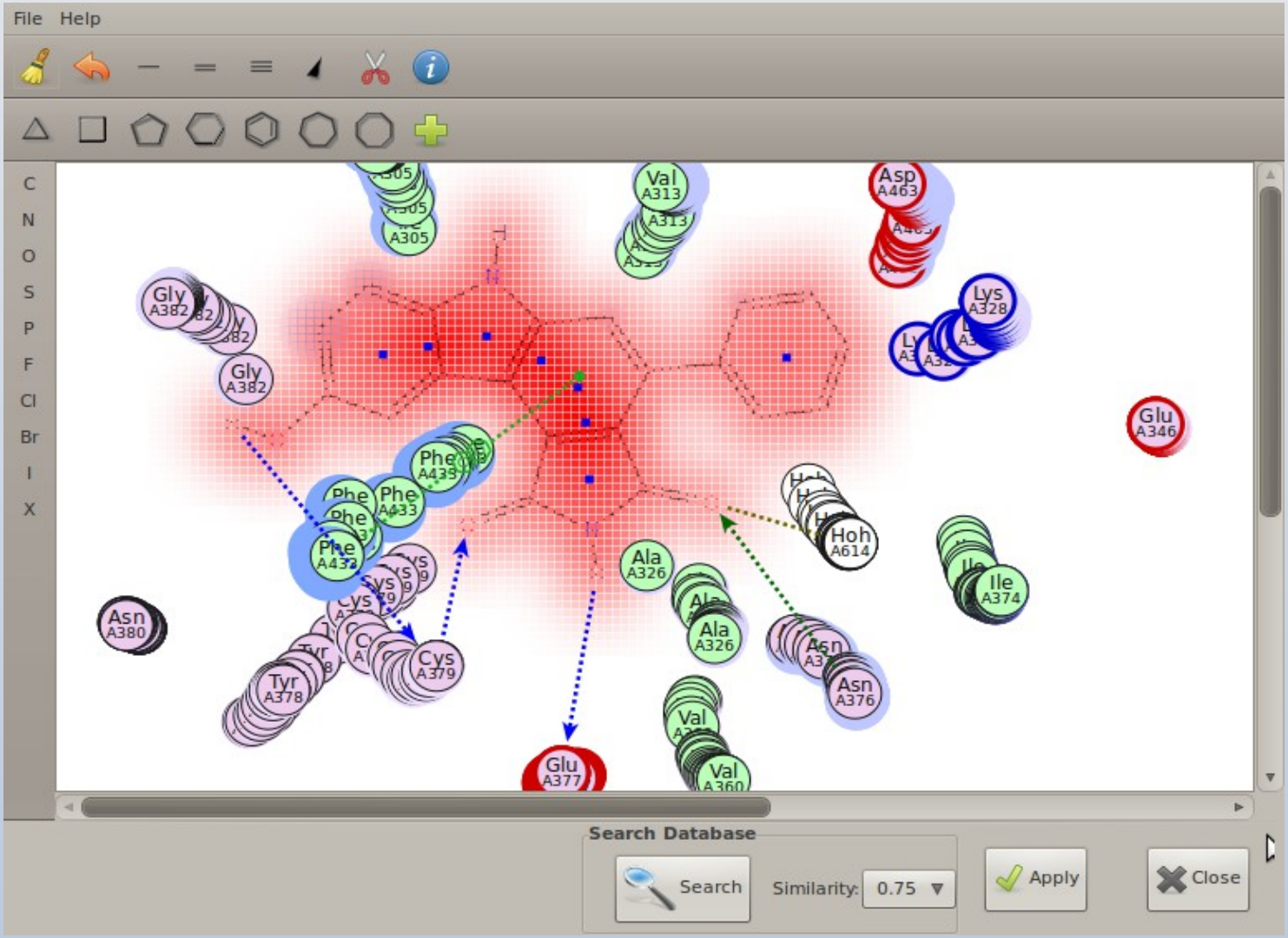
Ligand Environment Layout

- Initial residue placement



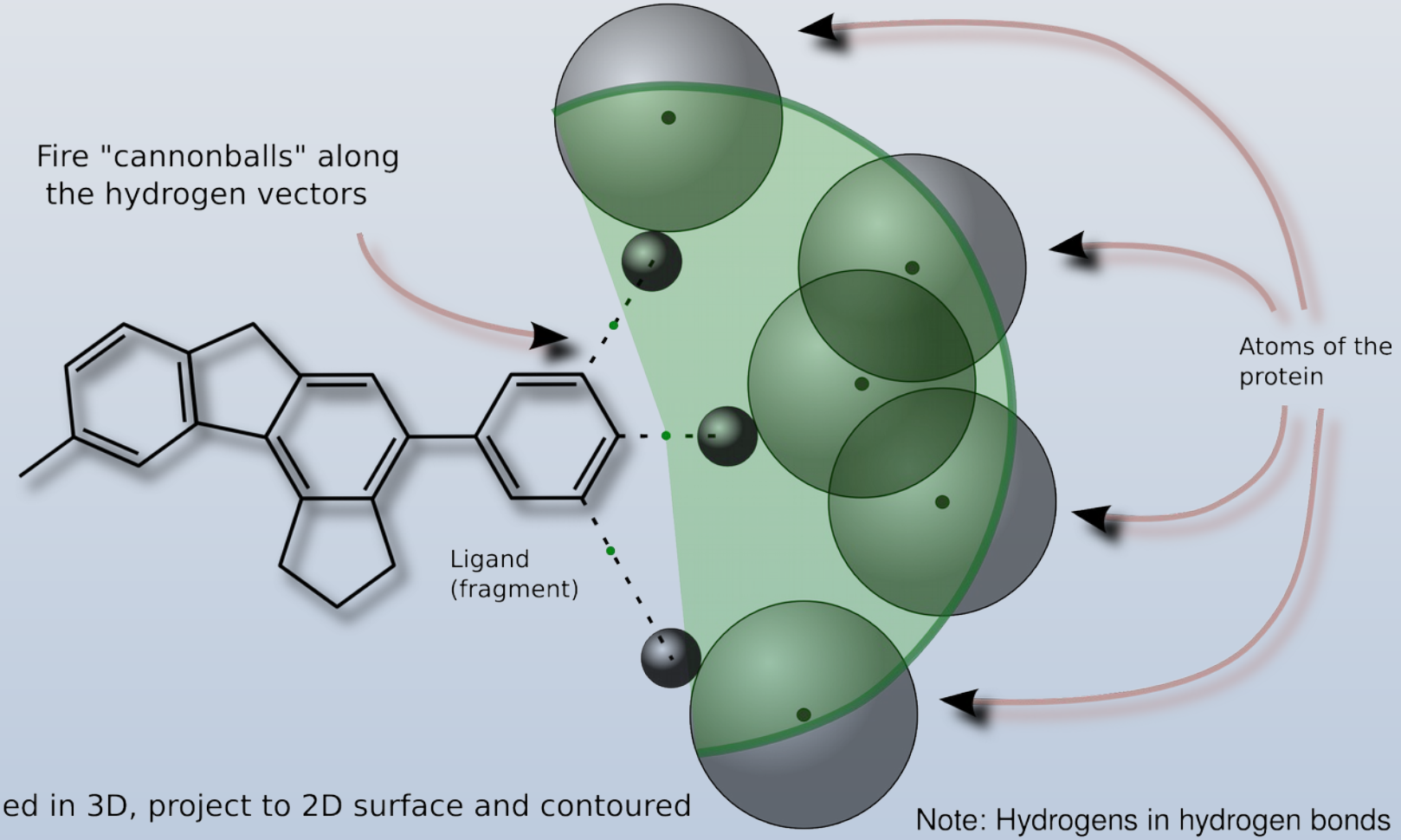
Ligand Environment Layout

- 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?



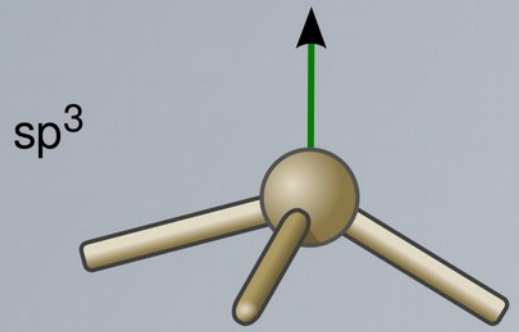
Determined in 3D, project to 2D surface and contoured

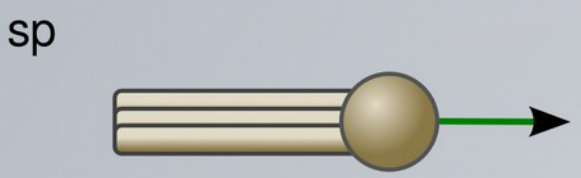
c.f. Clarke & Labute (2007)

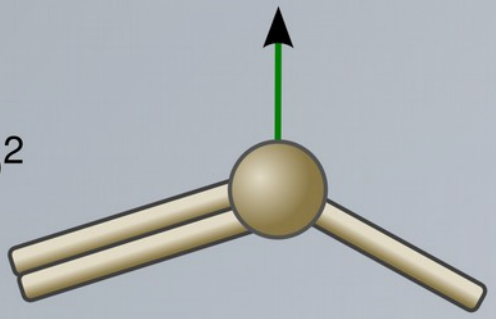
Note: Hydrogens in hydrogen bonds are a confounding factor

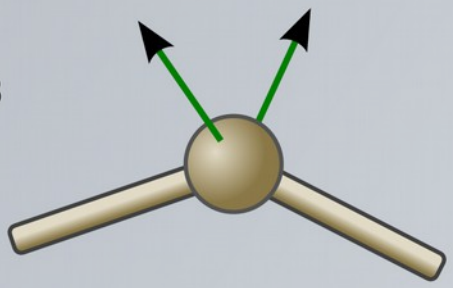
Substitution Contour: Extending along Hydrogens

Riding Hydrogens



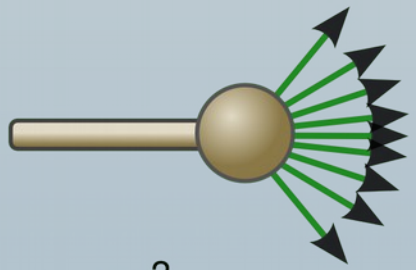


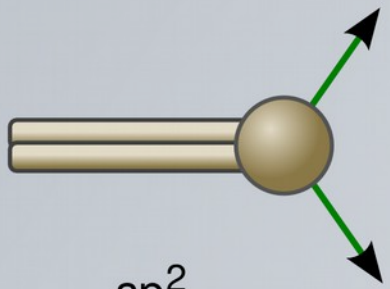


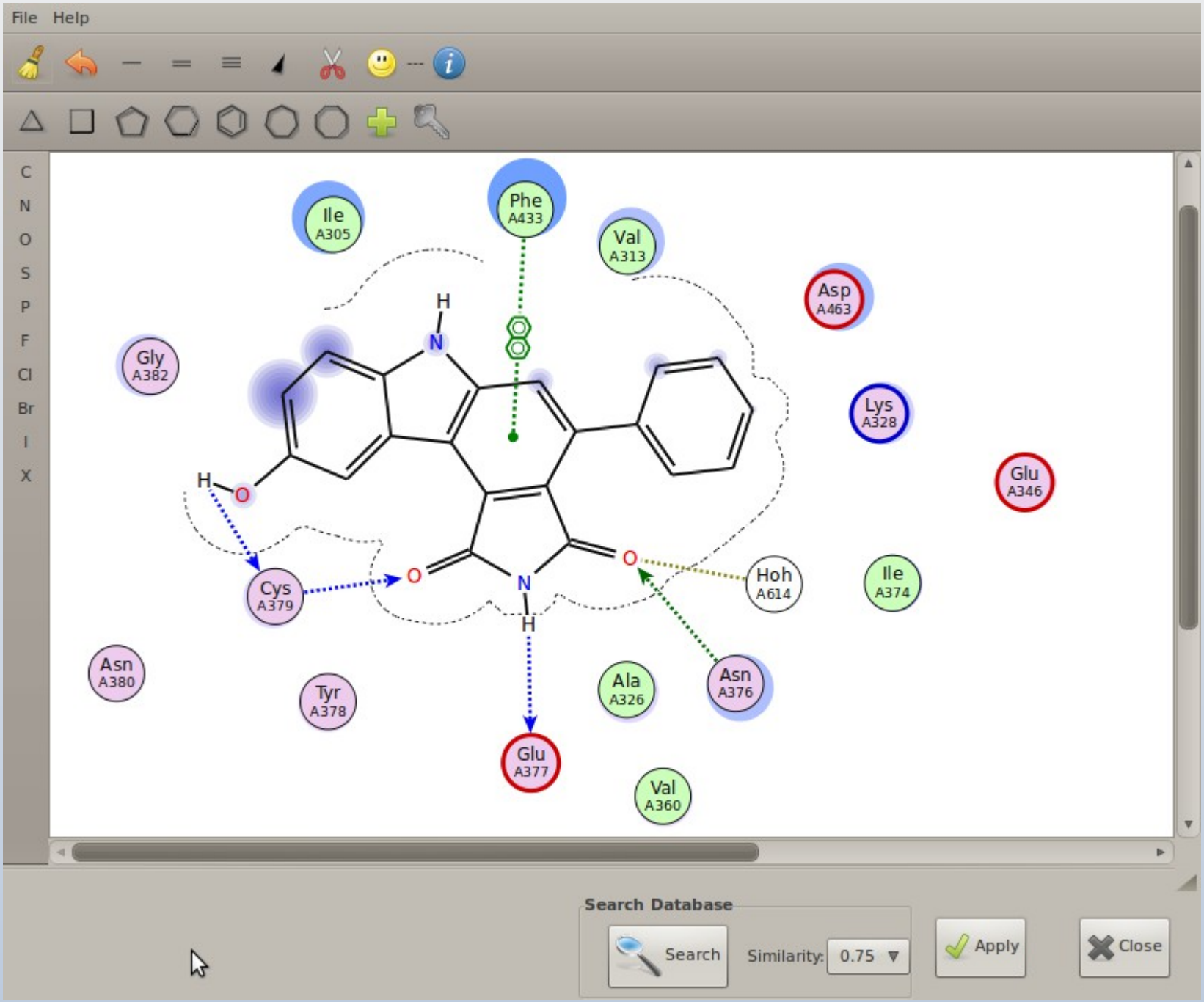


Torsionable Hydrogens

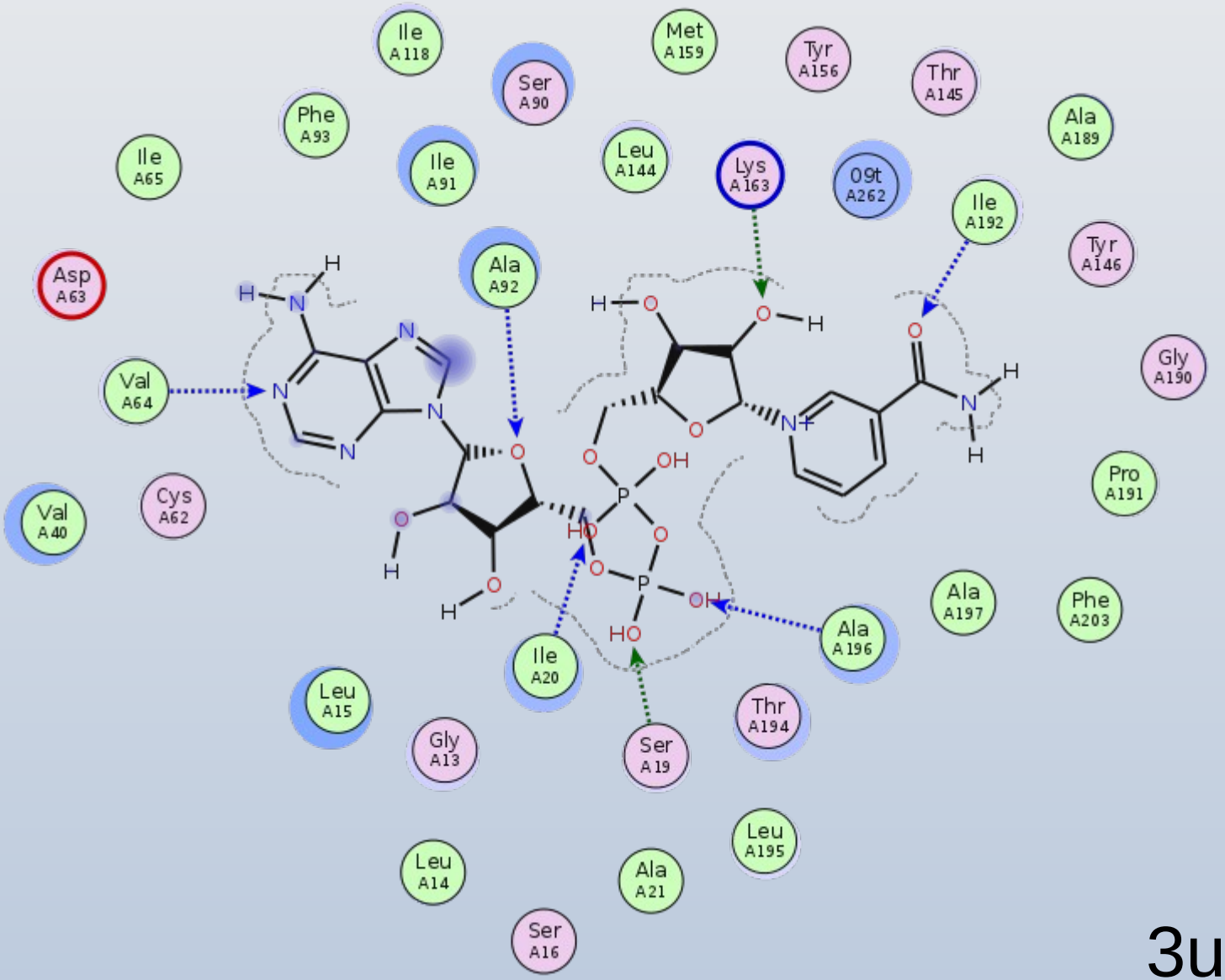
(test multiple directions)



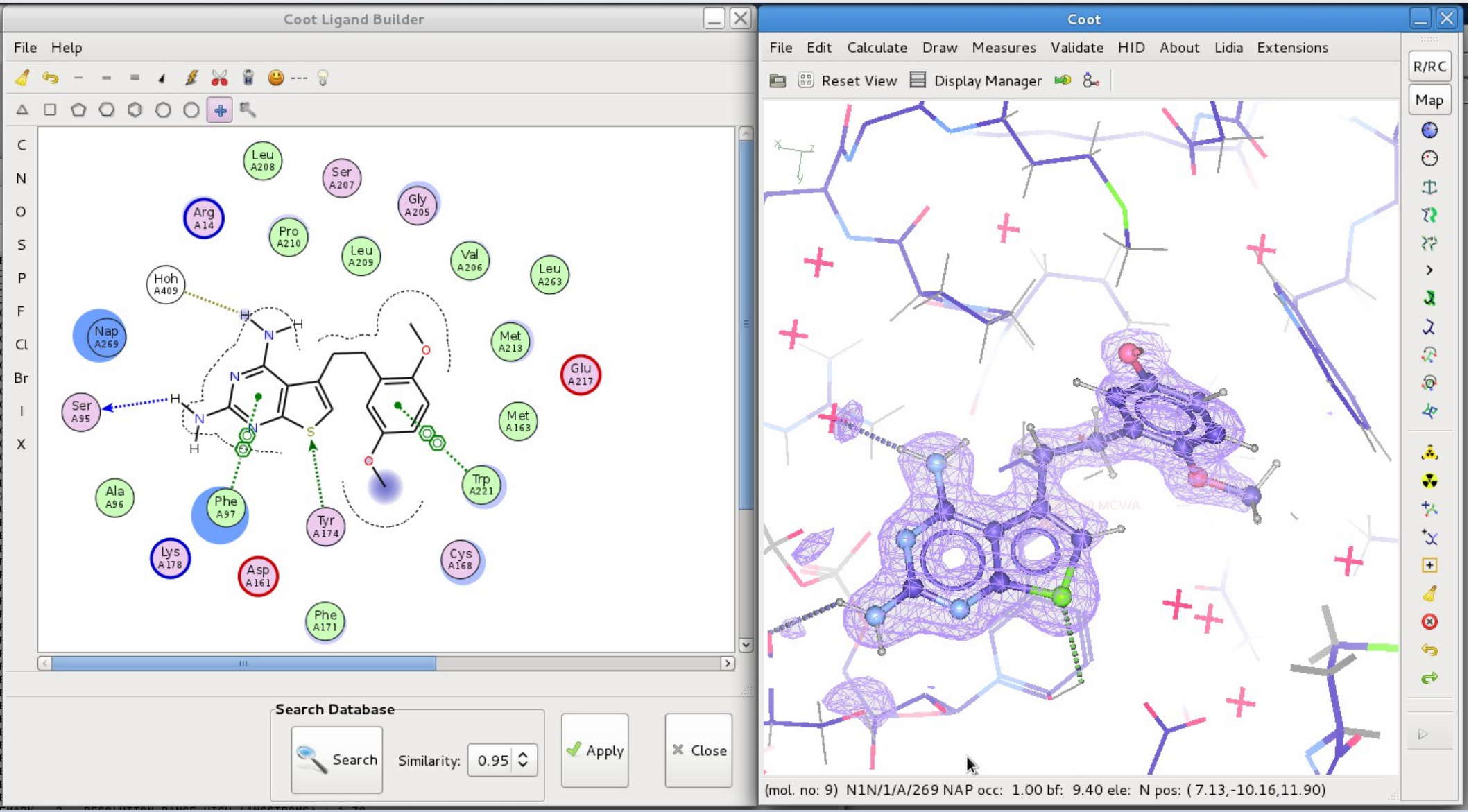



File Help
C
N
O
S
P
F
Cl
Br
I
X
Ile A305
Phe A433
Val A313
Asp A463
Gly A382
Lys A328
Glu A346
Cys A379
Hoh A614
Ile A374
Asn A380
Tyr A378
Ala A326
Asn A376
Glu A377
Val A360
Search Database
Search
Similarity: 0.75
Apply
Close

Layout Examples



3uic

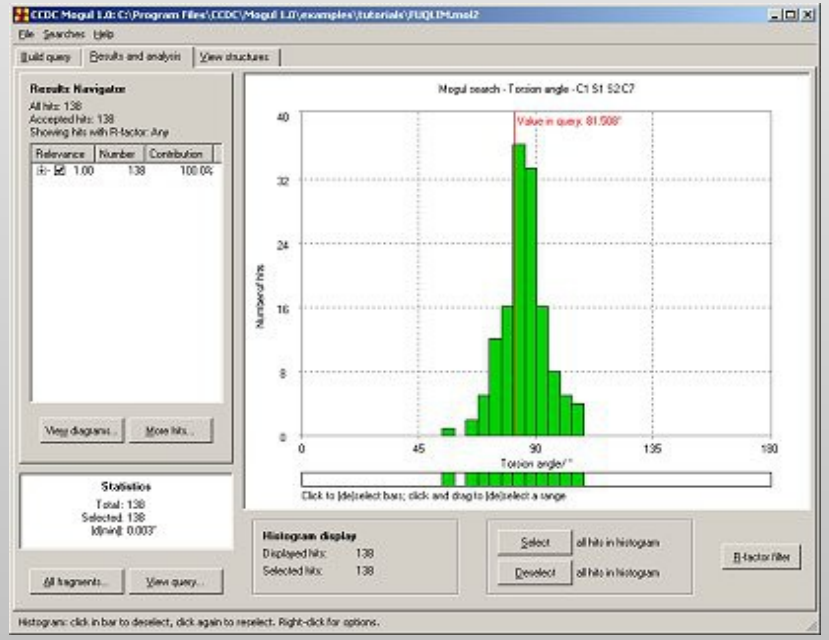

Coot Ligand Builder
File Help
C
N
O
S
P
F
Cl
Br
I
X
Leu A208
Ser A207
Gly A205
Arg A14
Pro A210
Leu A209
Val A206
Leu A263
Hoh A409
Nap A269
Met A213
Glu A217
Ser A95
Met A163
Trp A221
Ala A96
Phe A97
Tyr A174
Lys A178
Asp A161
Cys A168
Phe A171
Search Database
Search
Similarity: 0.95
Apply
Close
Coot
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)

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

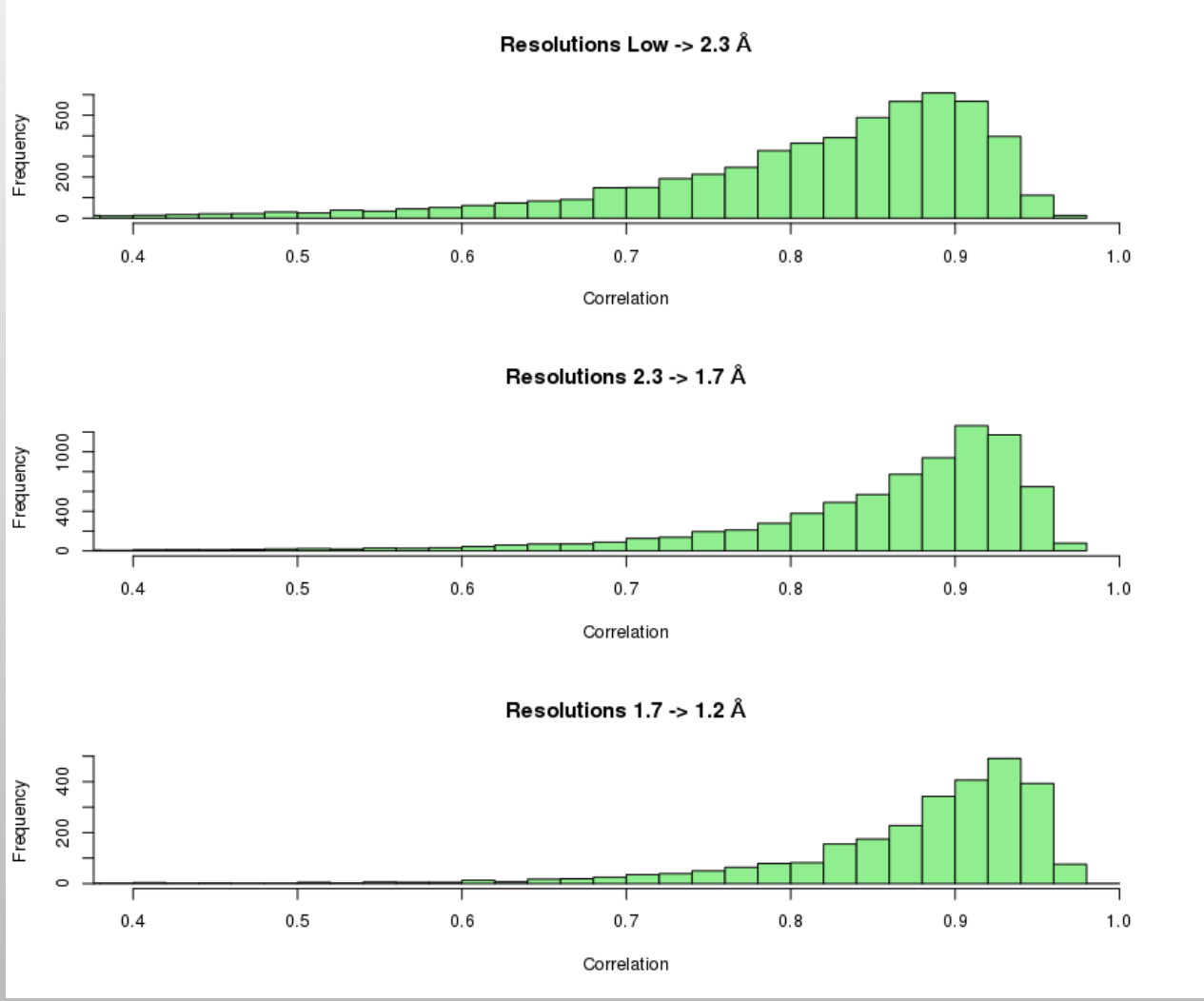
- CSD's Mogul
  - Knowledge-base of geometric parameters based on the CSD
  - Can be run as a “batch job”
  - Mean, median, mode, quartiles, Z-scores.



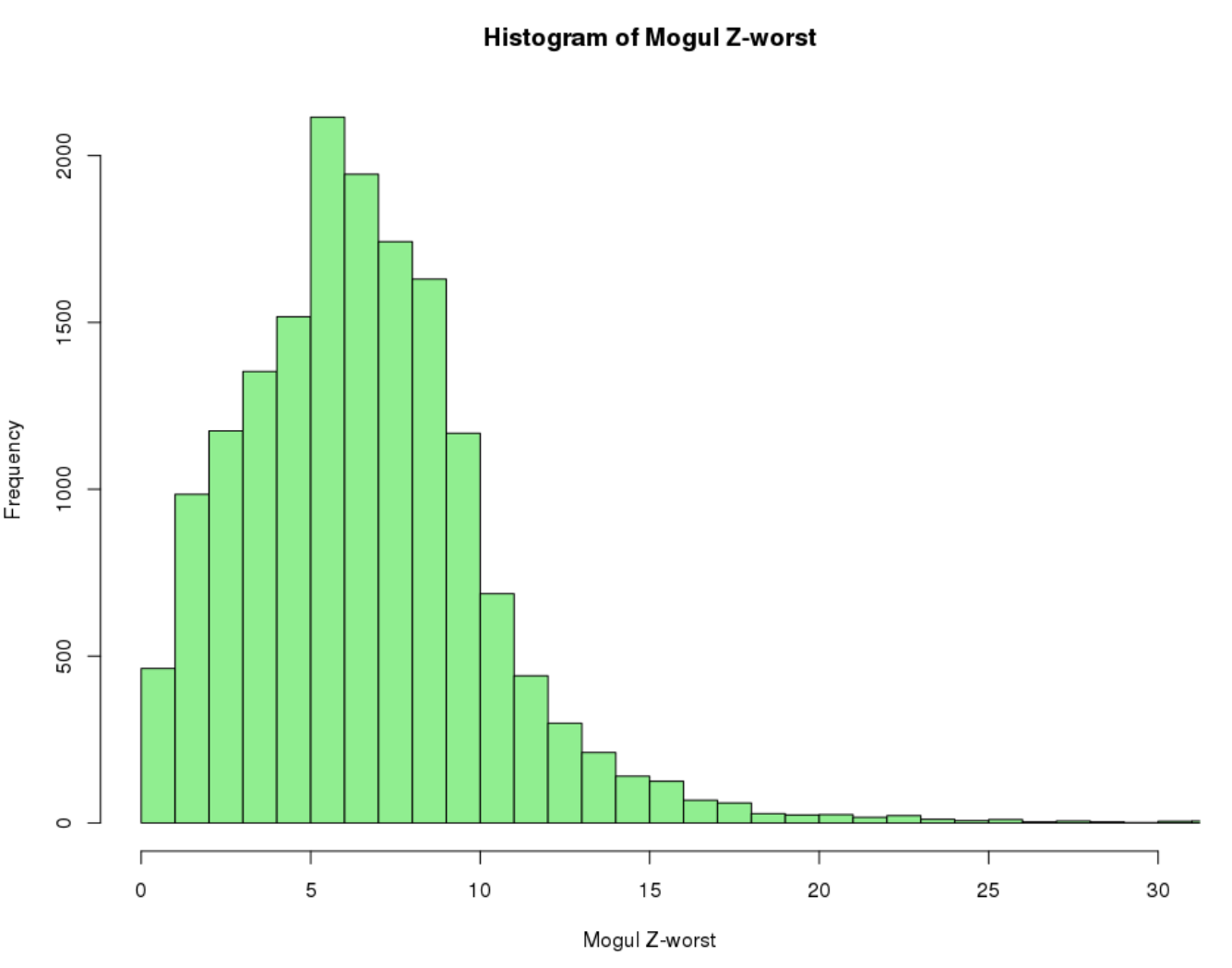
Score Histograms

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

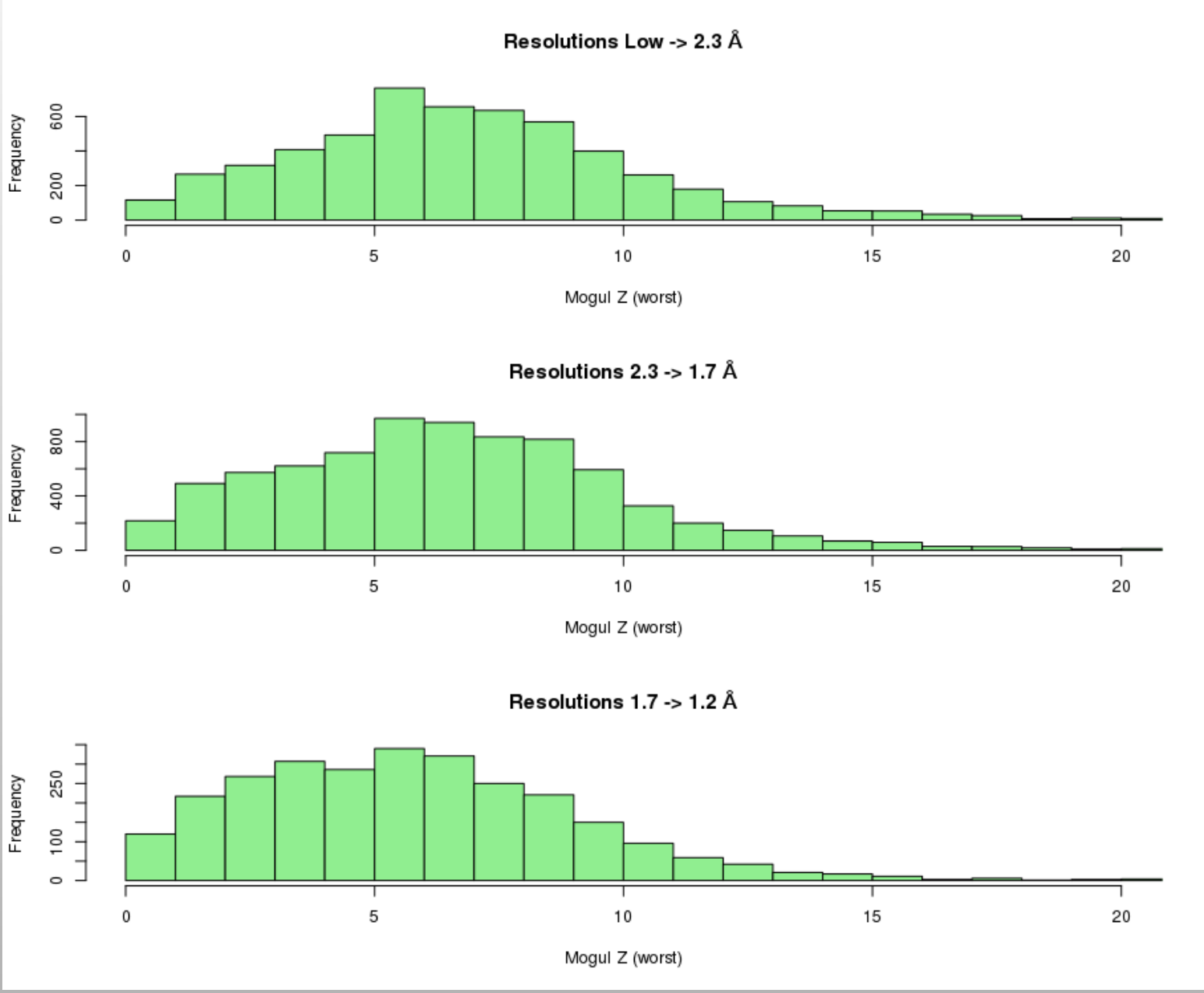
Resolution dependence of Density Correlation



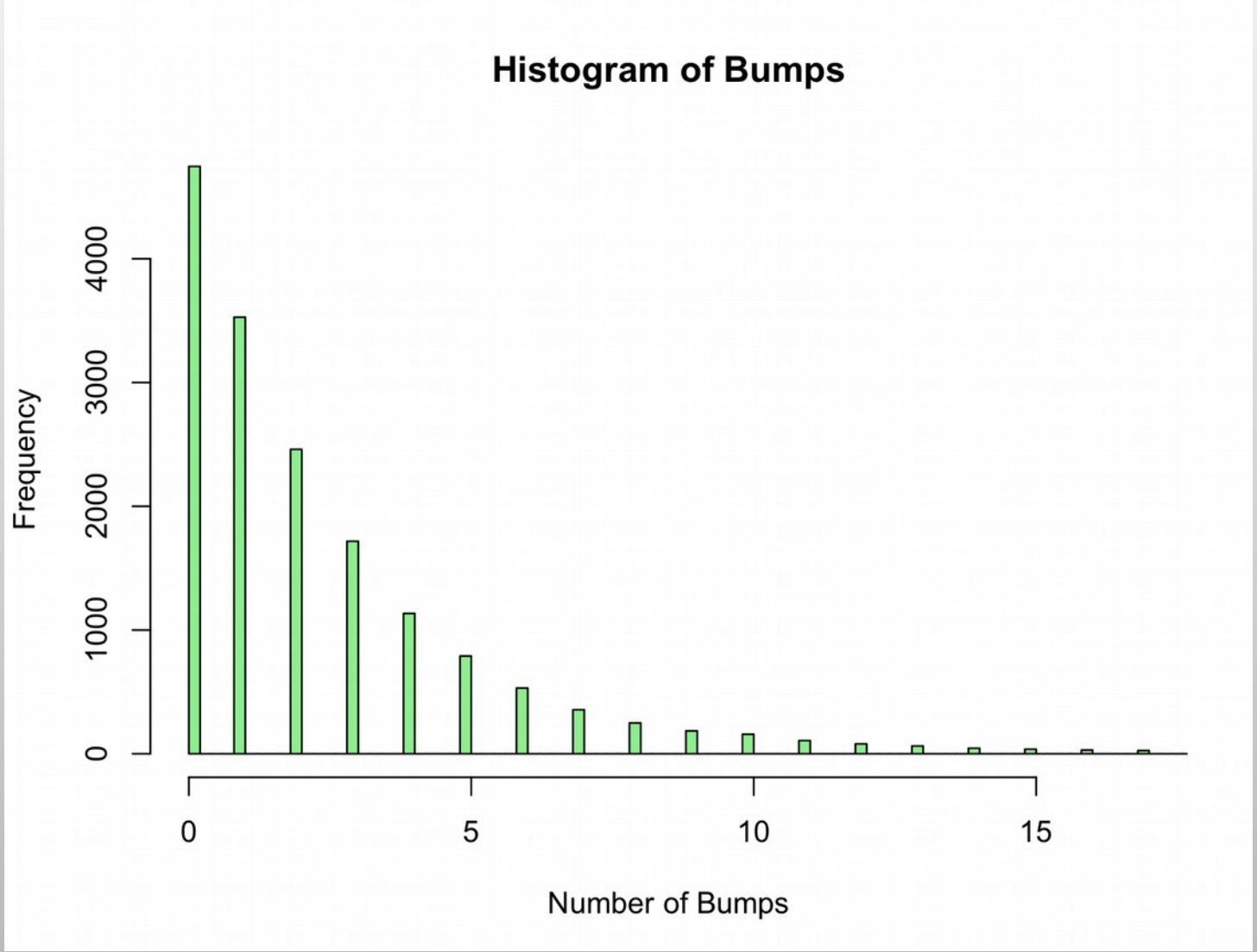
Overall Histogram of Mogul Z-worst of wwPDB Ligands



Resolution Dependence of Mogul Z-worst

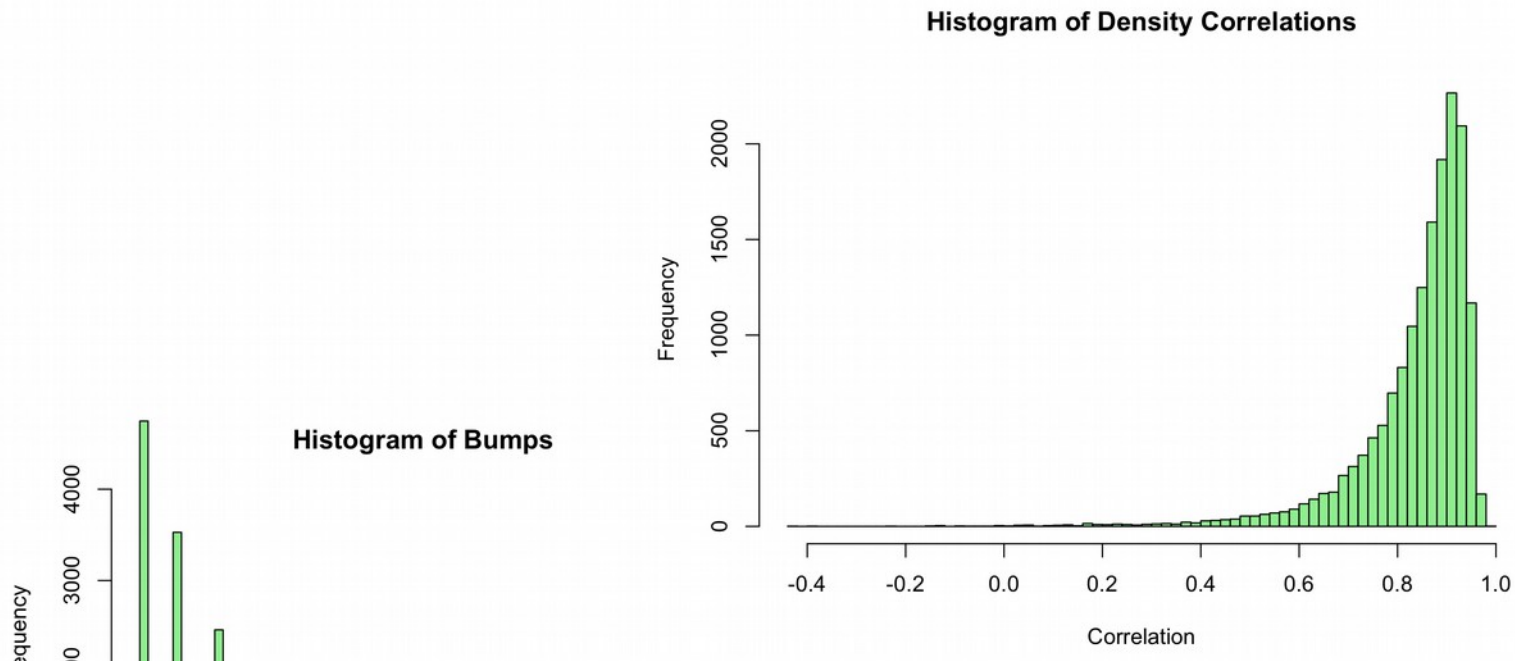


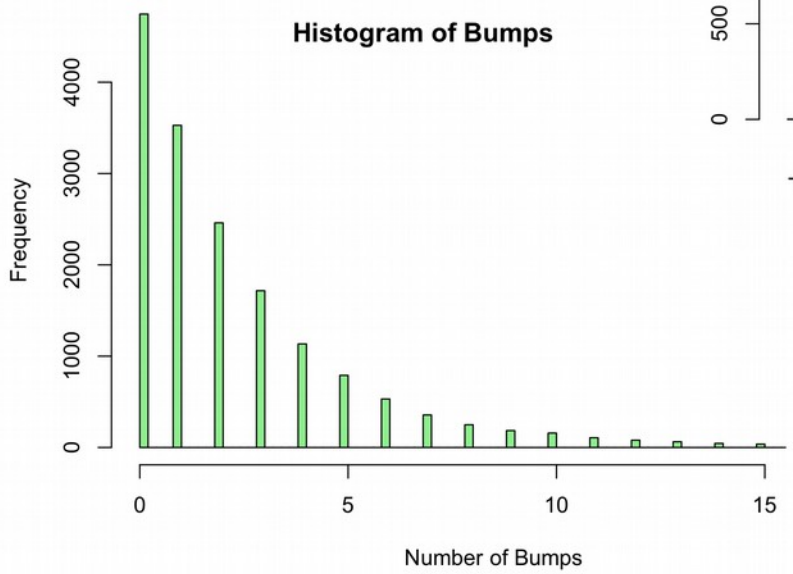
Histogram of Bad Contacts

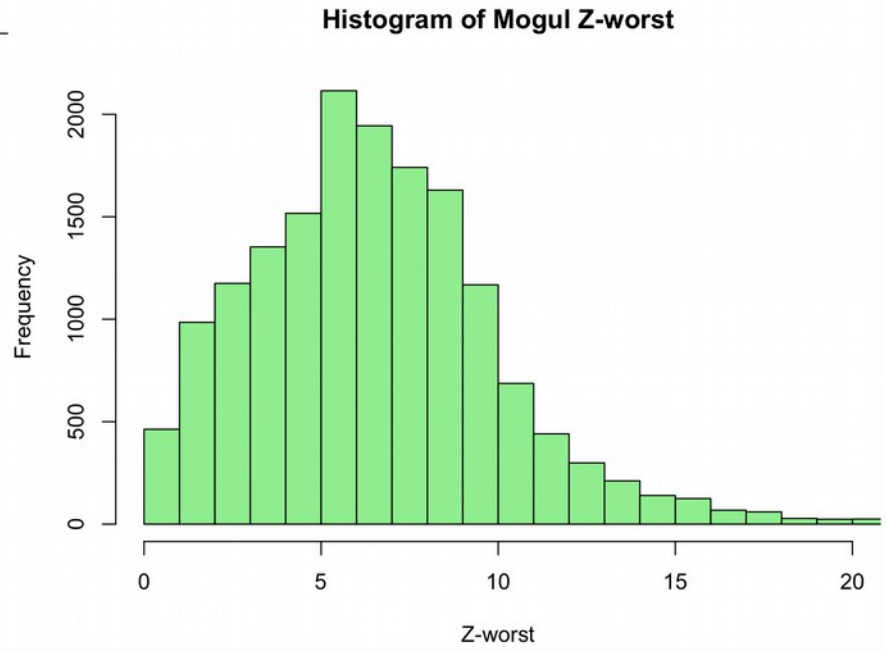


Ligand Scoring

Preliminary recommendatation...

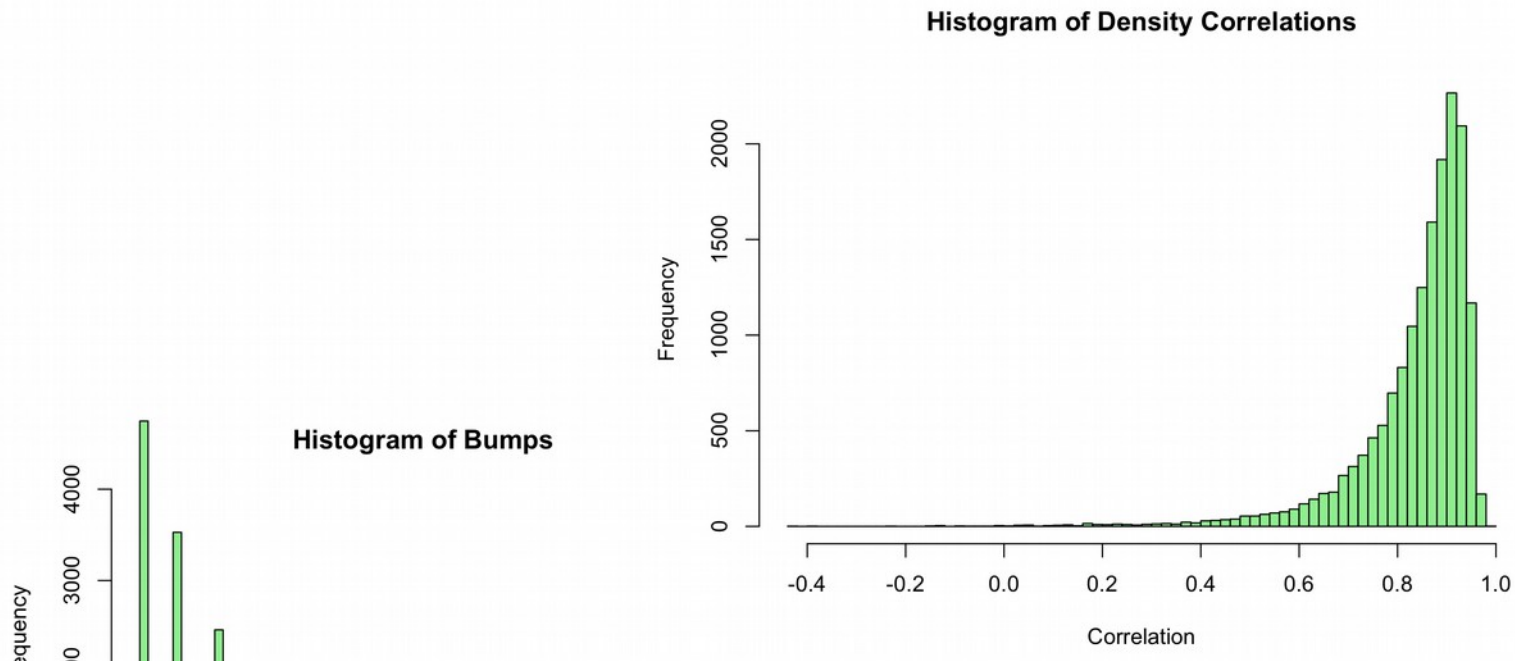

Histogram of Density Correlations
Frequency
0
500
1000
1500
2000
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Correlation

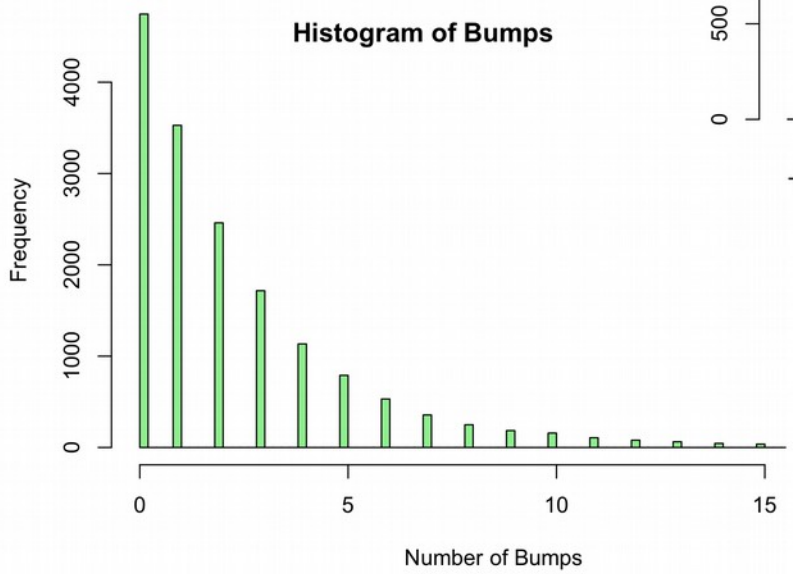

Histogram of Bumps
Frequency
0
1000
2000
3000
4000
0
5
10
15
Number of Bumps

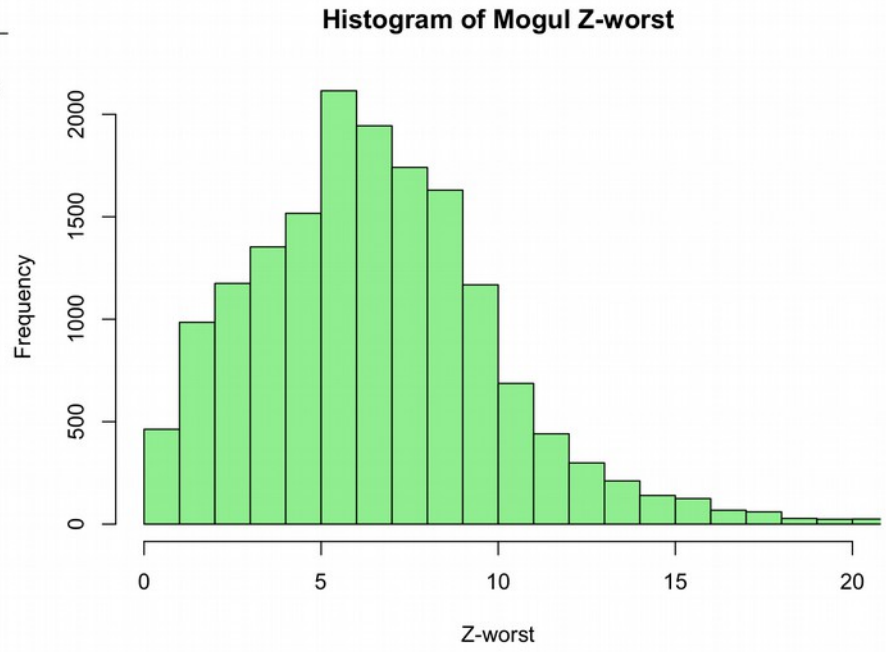

Histogram of Mogul Z-worst
Frequency
0
500
1000
1500
2000
0
5
10
15
20
Z-worst

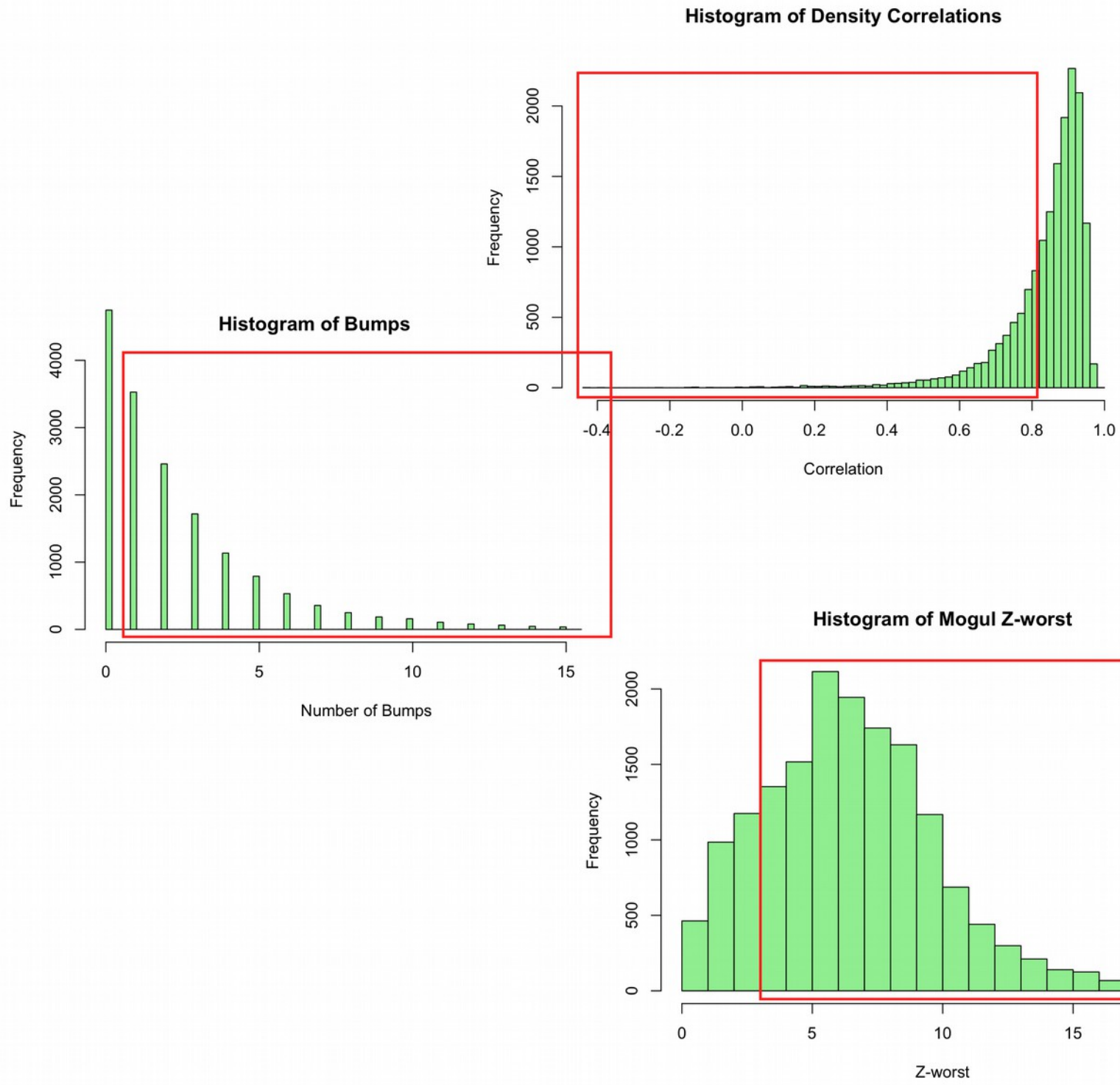
Scoring Ligands:
To Be Better Than The Median:

- 1 or 0 bumps
- Mogul z(worst) < 6.3
- Density correlation > 0.88


Histogram of Density Correlations
Frequency
0
500
1000
1500
2000
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Correlation


Histogram of Bumps
Frequency
0
1000
2000
3000
4000
0
5
10
15
Number of Bumps


Histogram of Mogul Z-worst
Frequency
0
500
1000
1500
2000
0
5
10
15
20
Z-worst

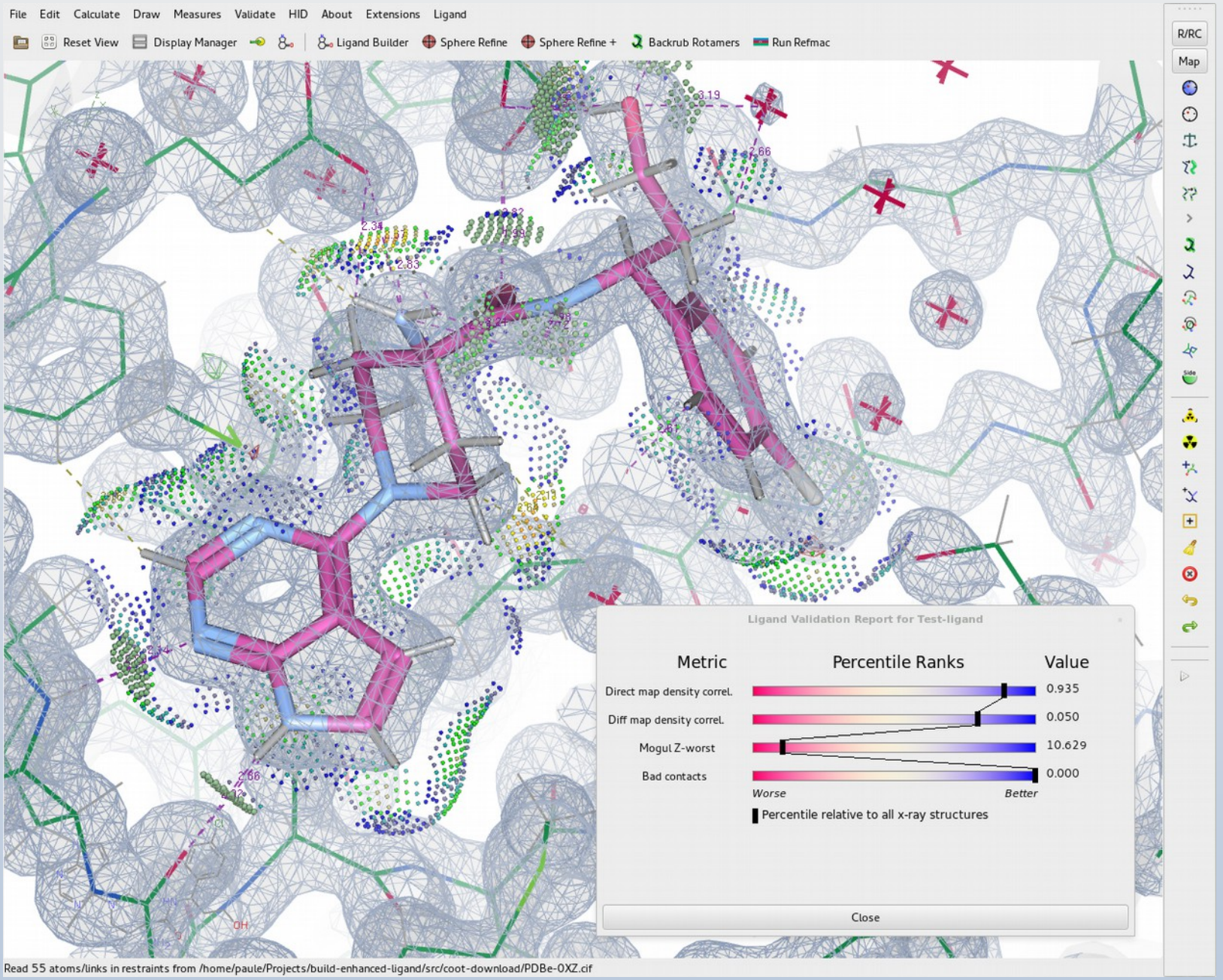

Histogram of Density Correlations
Frequency
0
500
1000
1500
2000
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
Correlation
Histogram of Bumps
Frequency
0
1000
2000
3000
4000
0
5
10
15
Number of Bumps
Histogram of Mogul Z-worst
Frequency
0
500
1000
1500
2000
0
5
10
15
20
Z-worst

Sliders

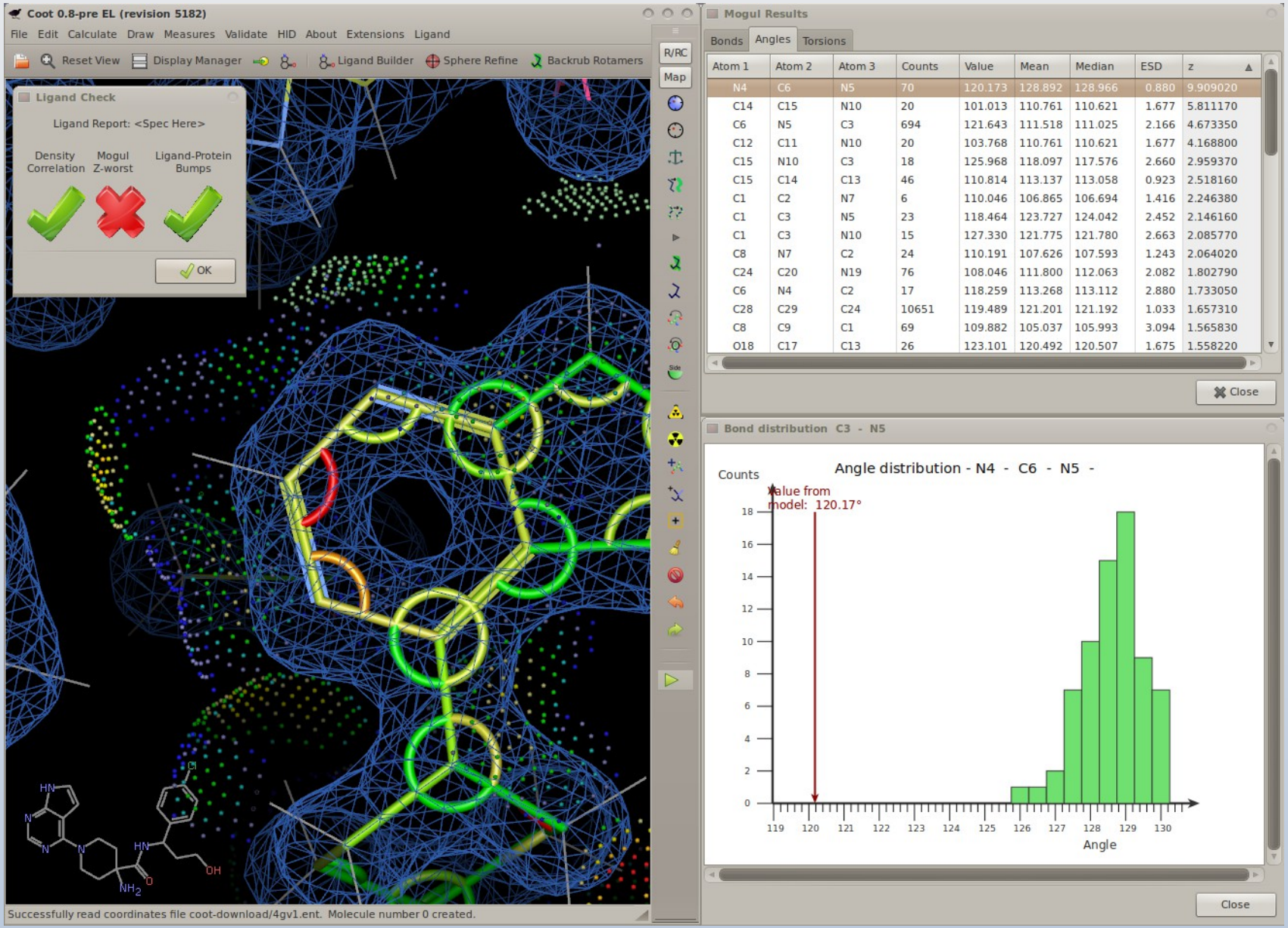
or

Yes/No?

Ligand Validation Sliders



Coot Ligand Validation Metrics Screenshot



4gv1

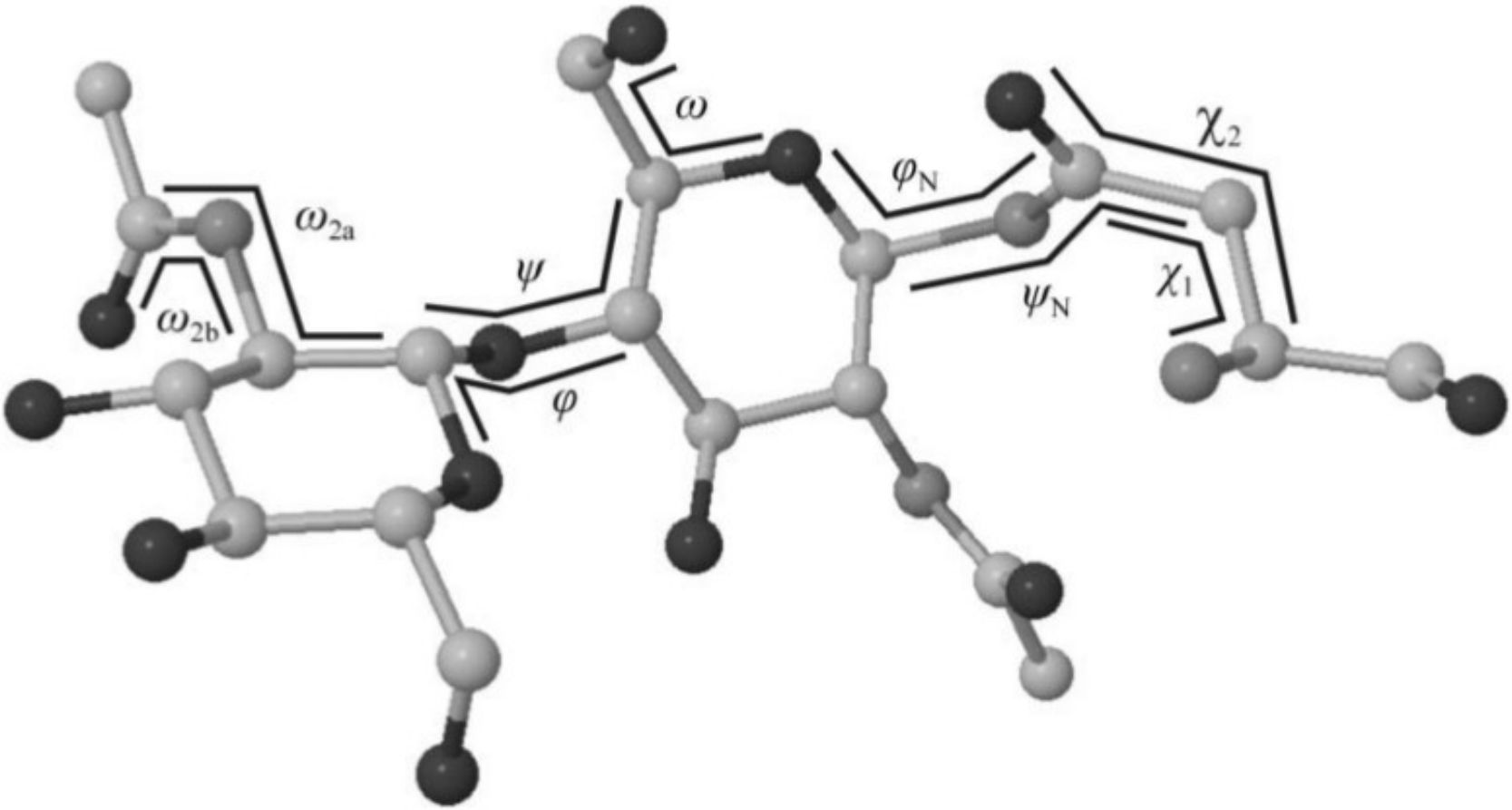
Problematic Glycoproteins

- 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
  - Also Joosten & Lűtteke (2017), Agirre *et al.* (2017)

Problematic Glycosylation

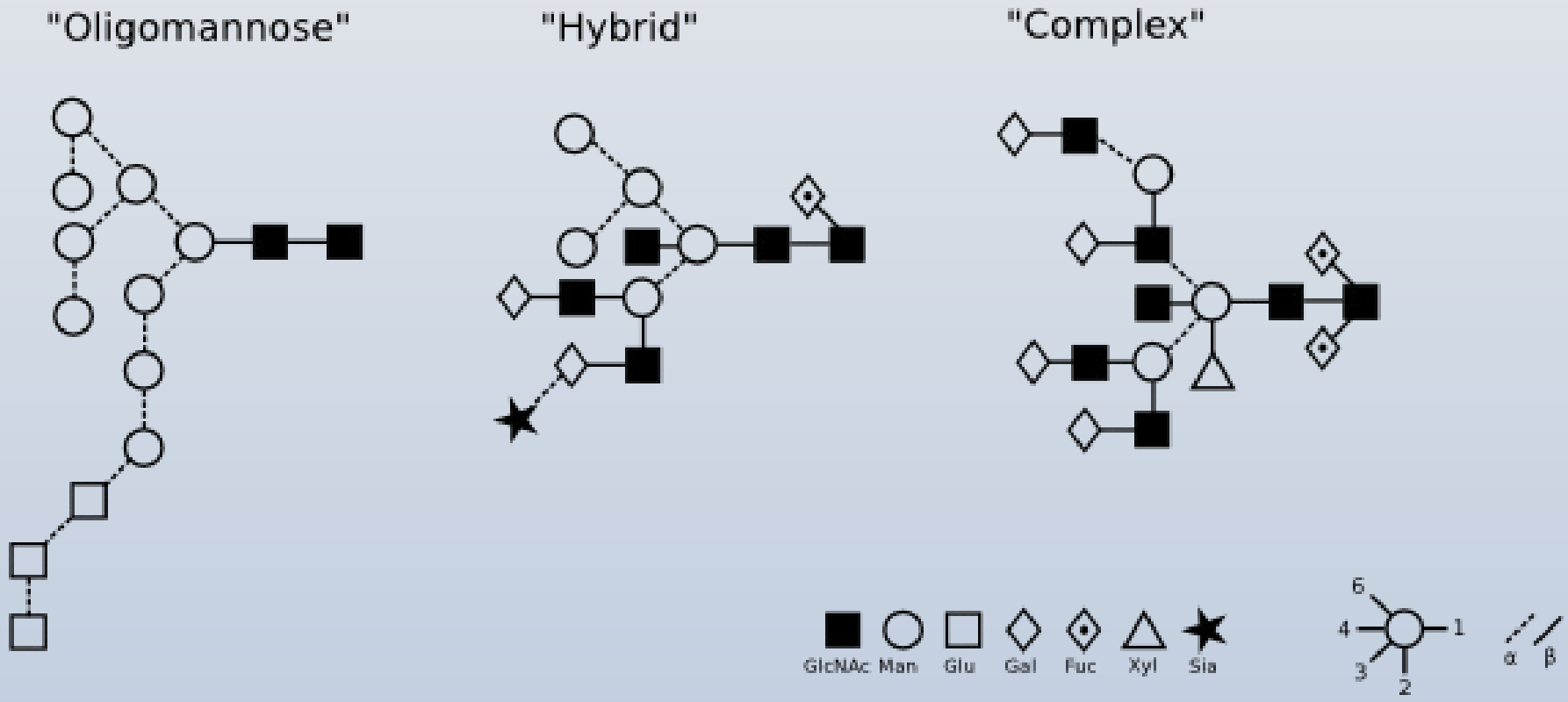
- In the case of carbohydrates, their inherent complexity [and] conformational flexibility [] are causing massive experimental problems which hinder the determination of the exact tertiary structures of these biomolecules
  - Engelsen *et al.* (2014) "Biopolymers"

Carbohydrate Links



Thomas Lütteke (2007)

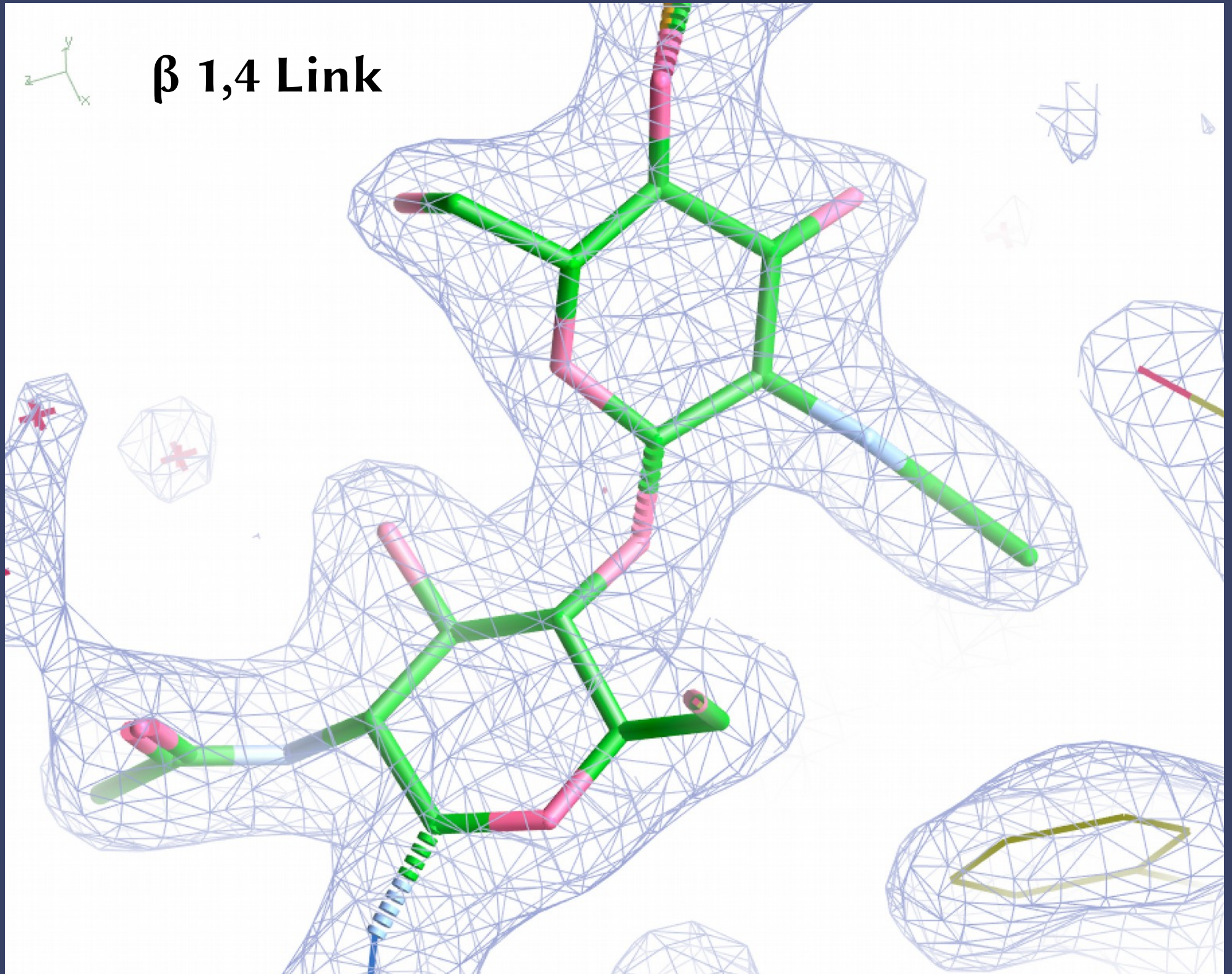
Validate the Tree:
N-linked carbohydrates



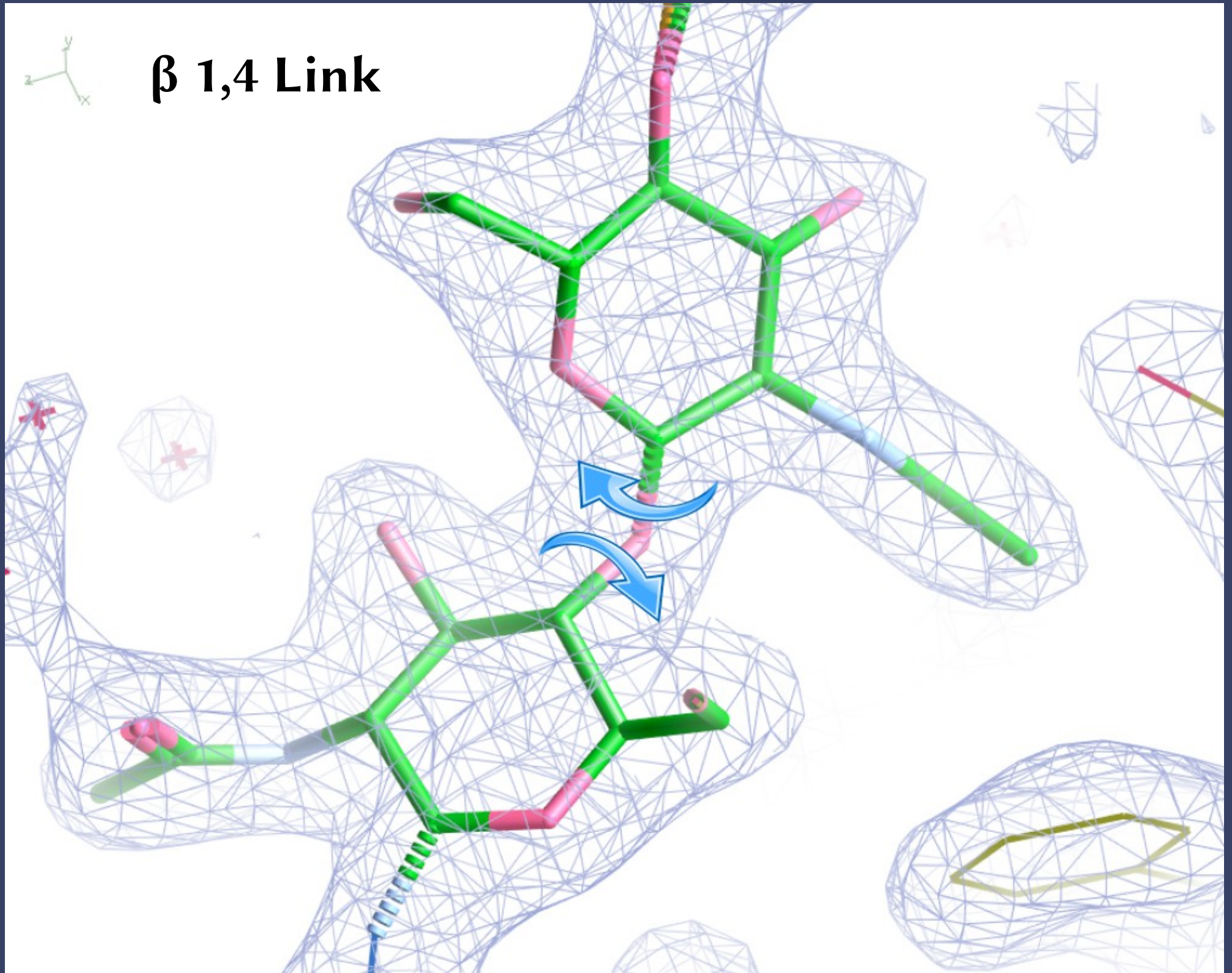
Linking Oligosaccharides/Carbohydrates: LO/Carb

- One can fully define carbohydrate structure by the primary structure and a set of torsion angles
- Build complex carbohydrate structure
  - from a dictionary of standard links
  - and monomers
  - torsion-angle refinement
    - by simulated annealing

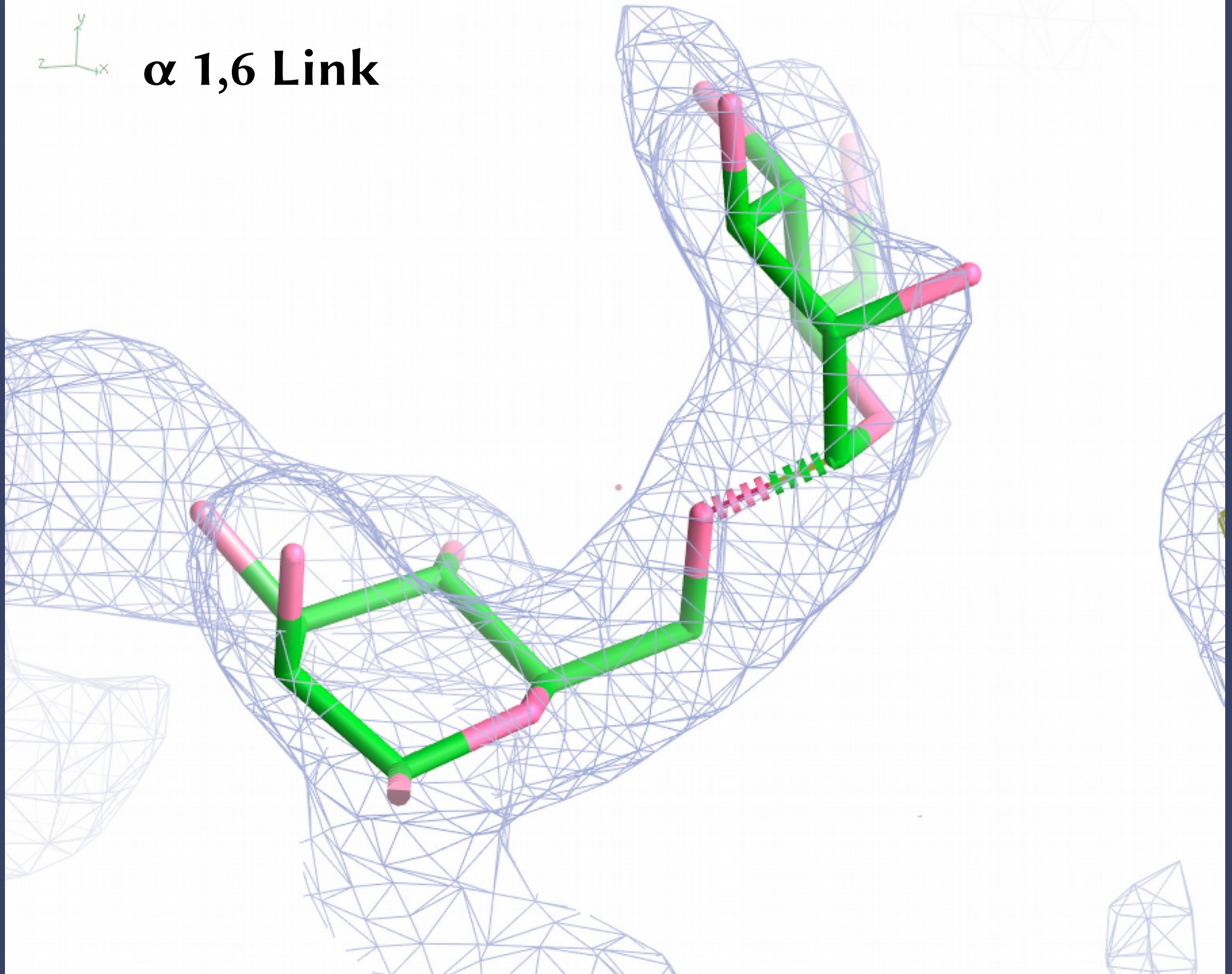
β 1,4 Link



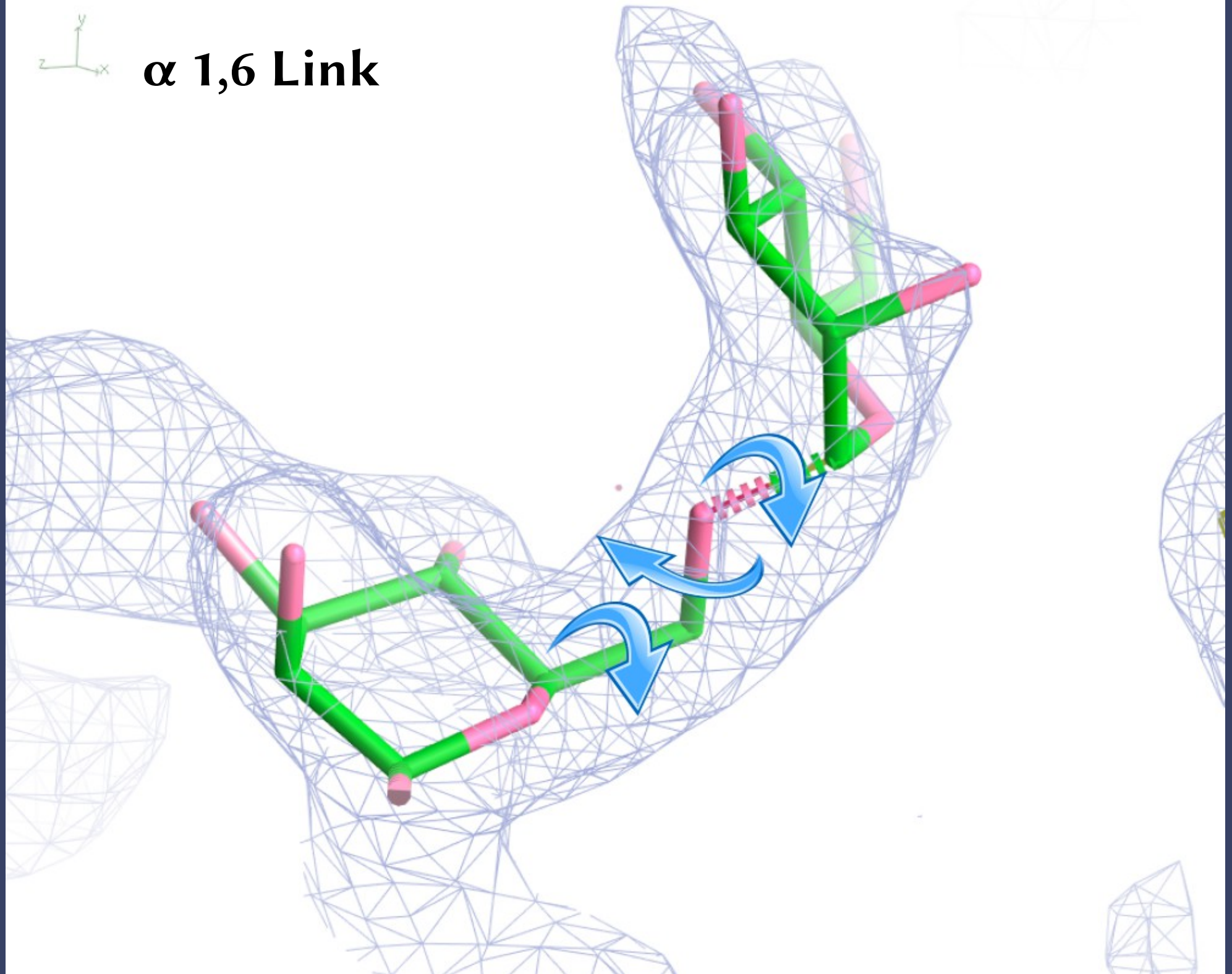
β 1,4 Link



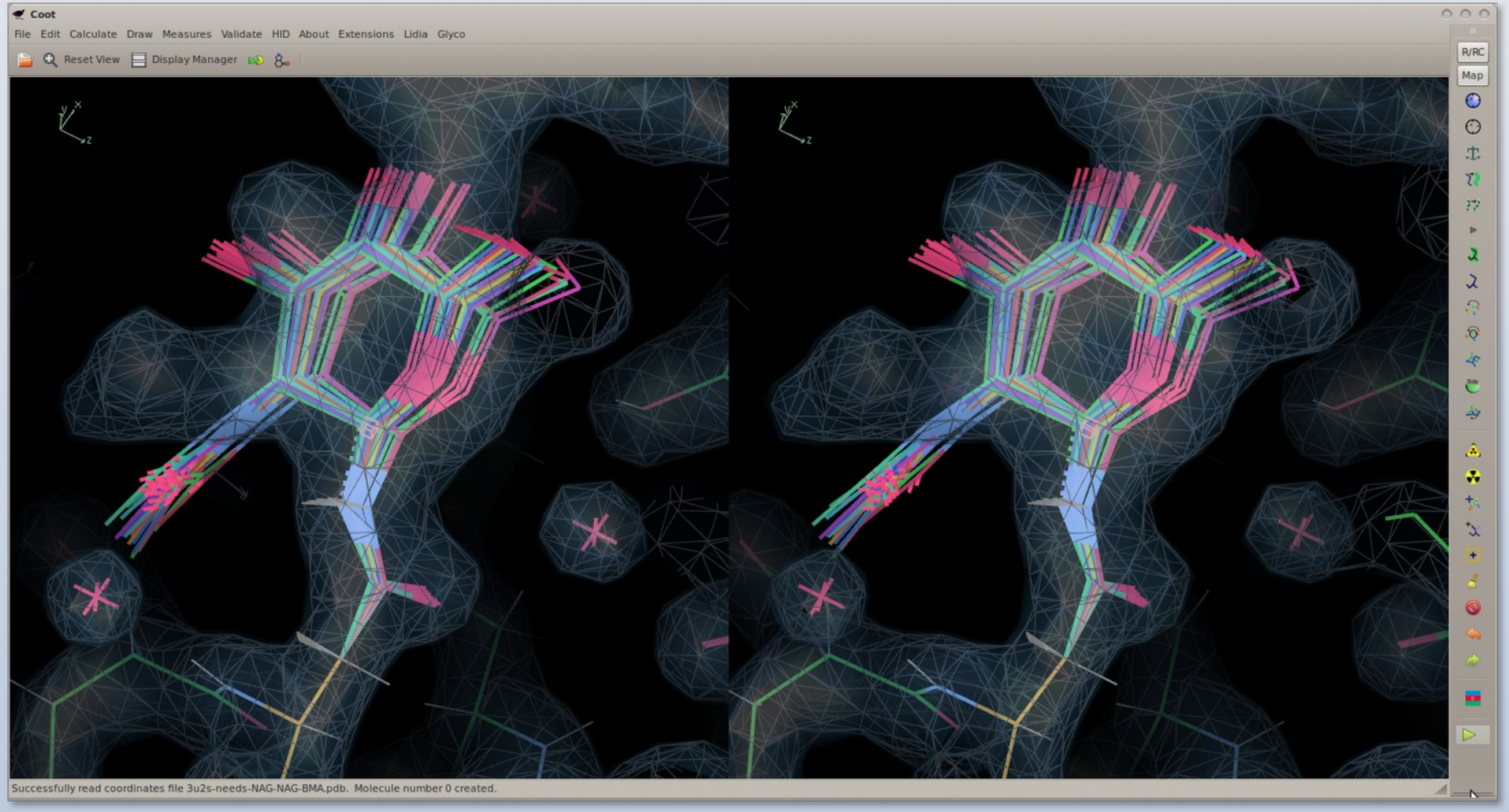
α 1,6 Link



α 1,6 Link



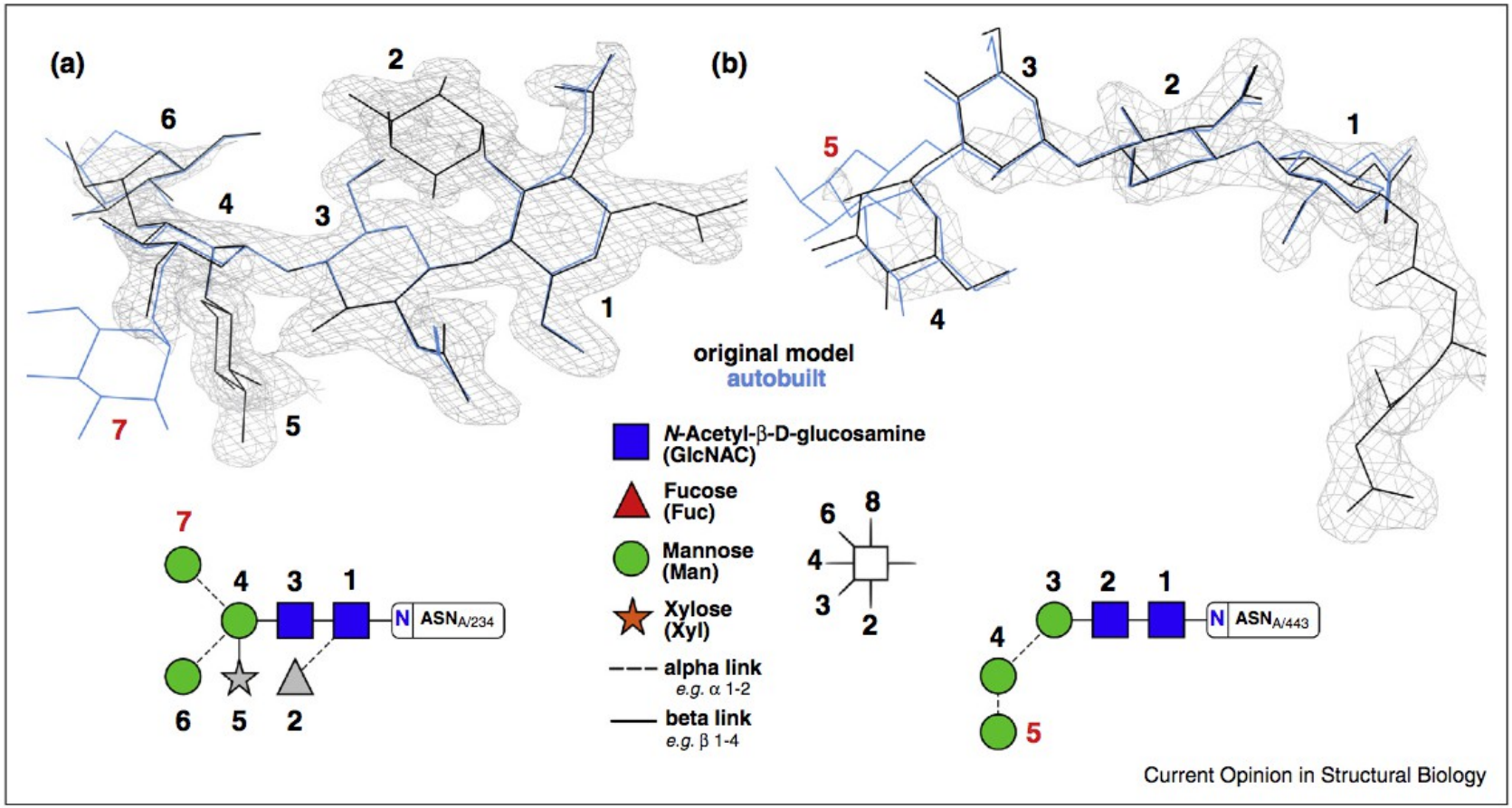
Refinement Progress (NAG-ASN example)



Problematic Glycosylation

Agirre *et al.* (2017) The Rocky Road to Automation

Figure 2



Linking Fucose: Fuc-α1,3

- Add a menu item to wrap the command
  - `add_linked_residue("FUC", "ALPHA1-3")`

Added into a new N-linked tree:

- paucimannose

Xyl-β1,2

- Xyl - β1,2 – Man
  - using XYP (beta D xylosepyranose)
  - was not in the Refmac Monomer Library list of links
  - It has been added and will be available to CCP4 shortly

Building Models "Wrongly" (judging by density)

	Good Density	Poor/Bad density
Model built	✓	False Positive
No Model	False Negative	✓

Adding PRIVATEER for Model Validation

- 2016-*Coot* had no validation for carbohydrate geometry
  - (only fit to density was used)
- Now the model is validated (and filtered) by tree
  - using the output of PRIVATEER
  - both GUI interface and built into the auto-builder
- New Interface
  - needs debugging?

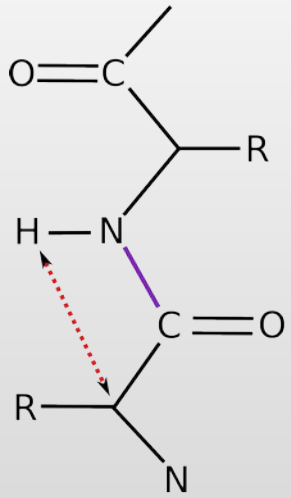
cis-Peptides

- What is a cis-peptide?
- Peptide restraints in Coot 2004-2015

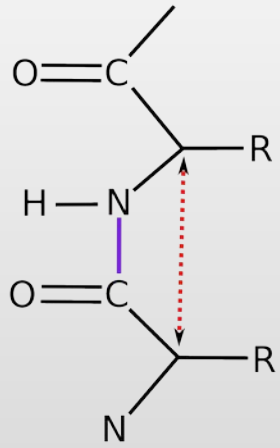
cis-Peptides

- A number of paper have been published recently highlighting the unusually large number of cis-peptides in some structures:
  - Croll: The rate of cis-trans conformation errors is increasing in low-resolution crystal structures *Acta Cryst.* (2015). **D**71, 706-709
  - Touw *et al*.: Detection of trans–cis flips and peptide-plane flips in protein structures *Acta Cryst.* (2015). **D**71, 1604-71614

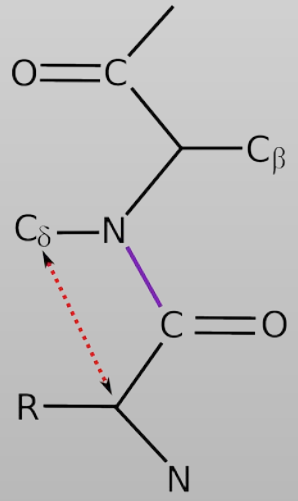
cis-Peptides



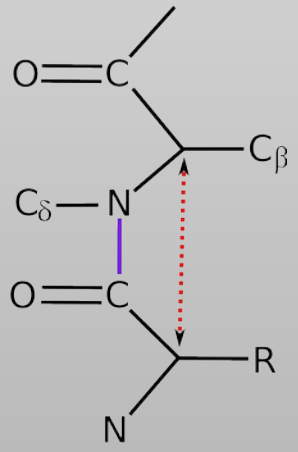
trans-peptide



cis-peptide

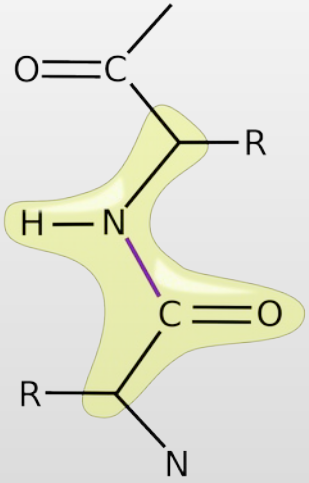


PRO *trans-peptide*

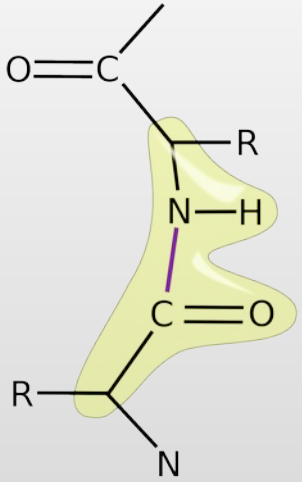


PRO *cis-peptide*

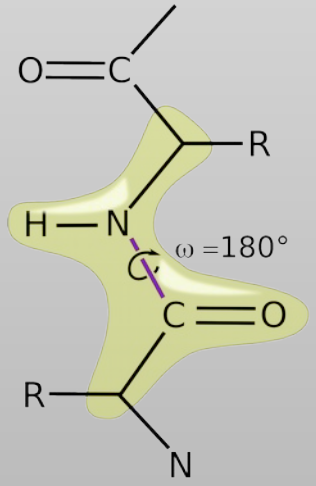
cis-Peptides



trans-peptide
with plane restraints

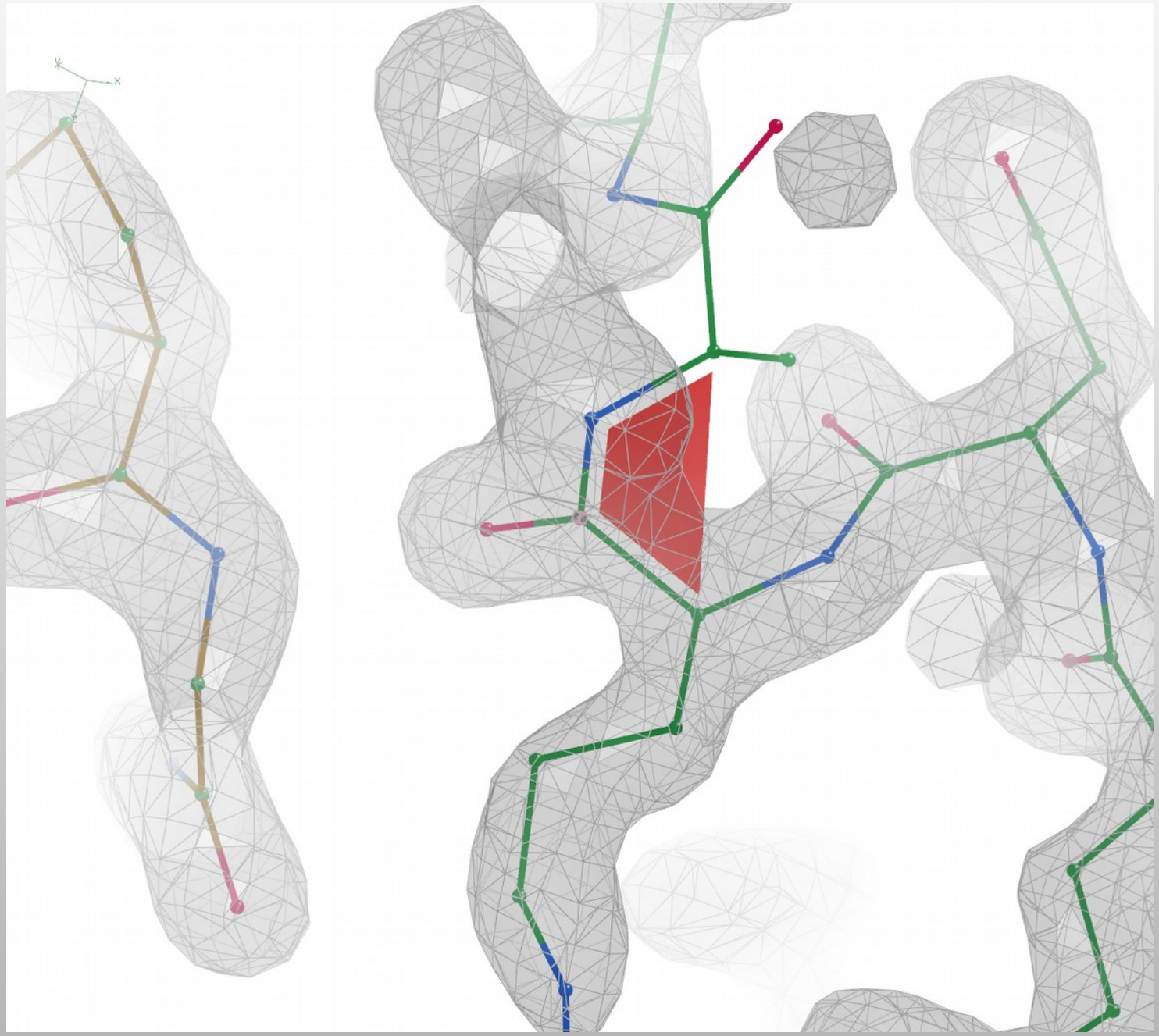


cis-peptide
with plane restraints



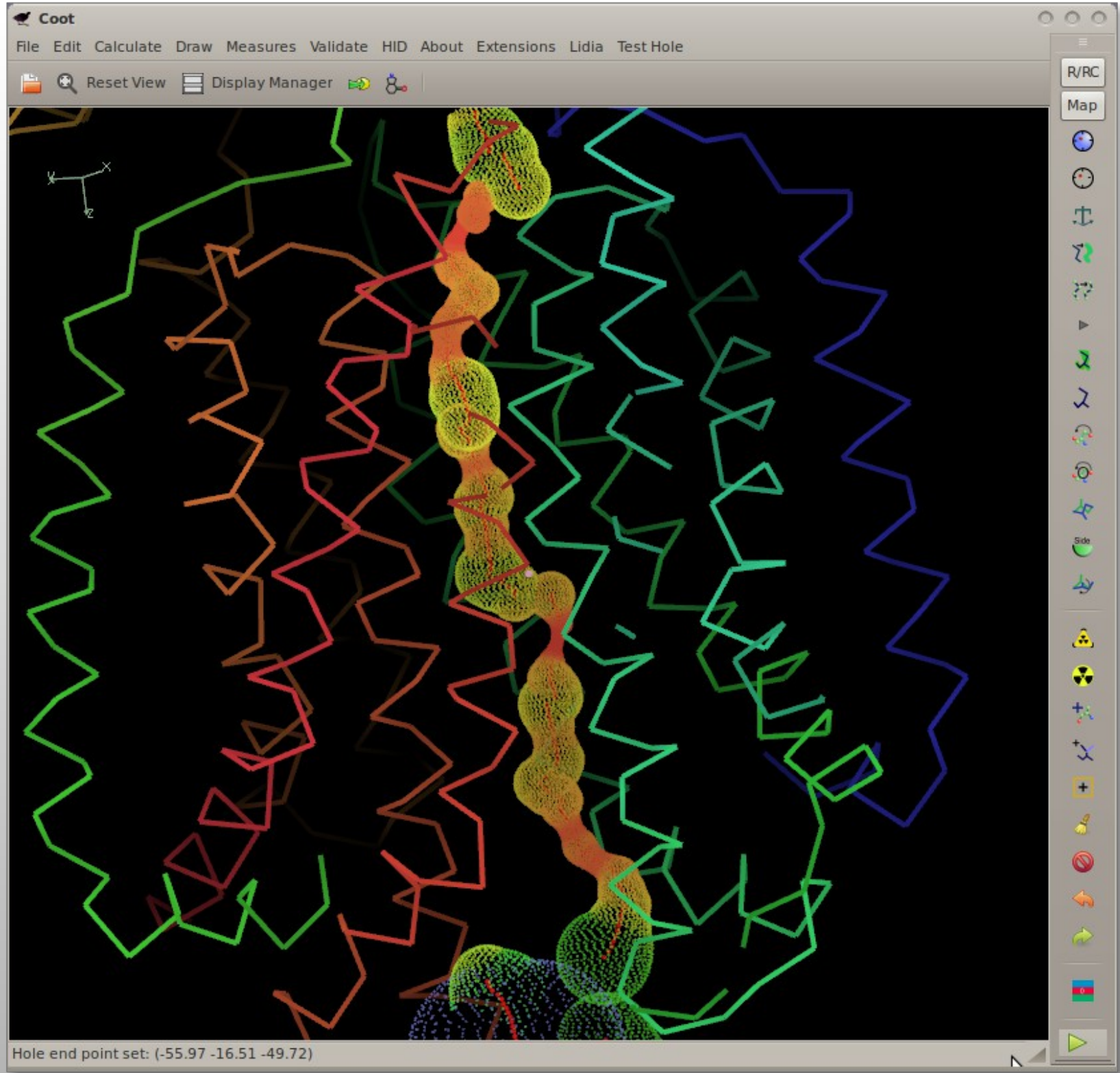
trans-peptide
with plane and trans restraints

cis-peptide Representation



Finding Holes

- An implementation of
  - Smart, Goodfellow & Wallace (1993) Biophysics Journal **65**, 2455
  - Atomic radii from AMBER
  - I used
    - radii from CCP4 monomer library
    - sans simulated annealing


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia Test Hole
Reset View
Display Manager
R/RC
Map
Side
Hole end point set: (-55.97 -16.51 -49.72)

Acknowledgements

- Kevin Cowtan
- Bernhard Lohkamp
- Eleanor Dodson
- Keith Wilson

- Libraries, dictionaries
  - Alexei Vagin, Eugene Krissinel
  - Richardsons (Duke)

- Funding
  - BBSRC, CCP4 & MRC