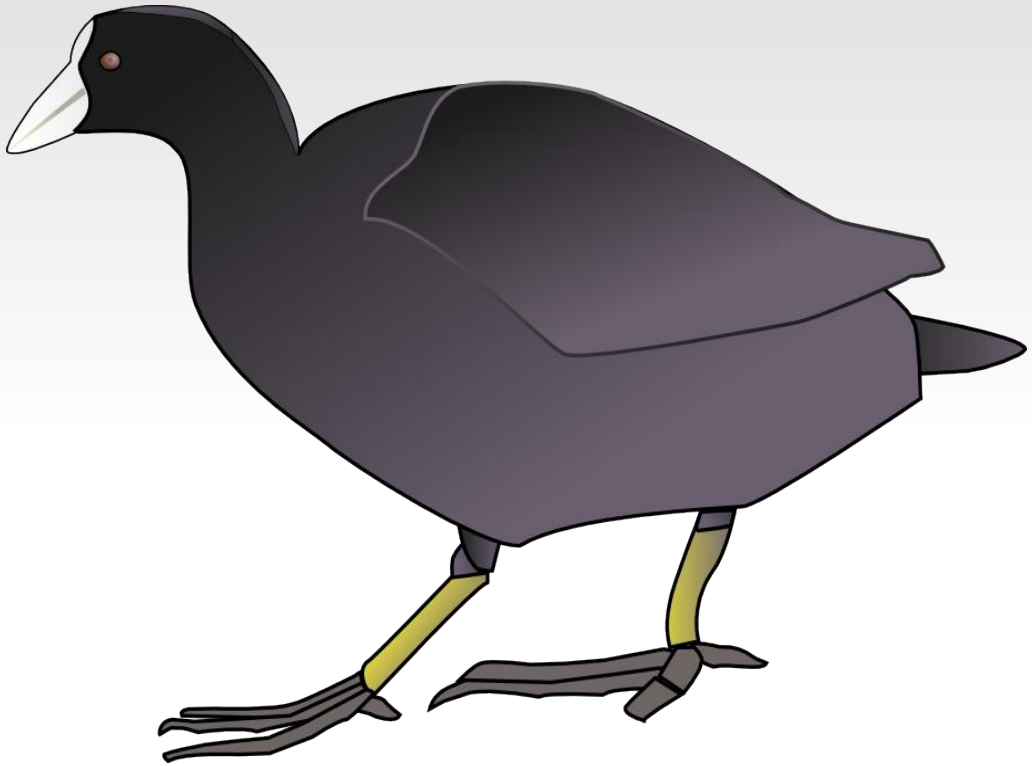
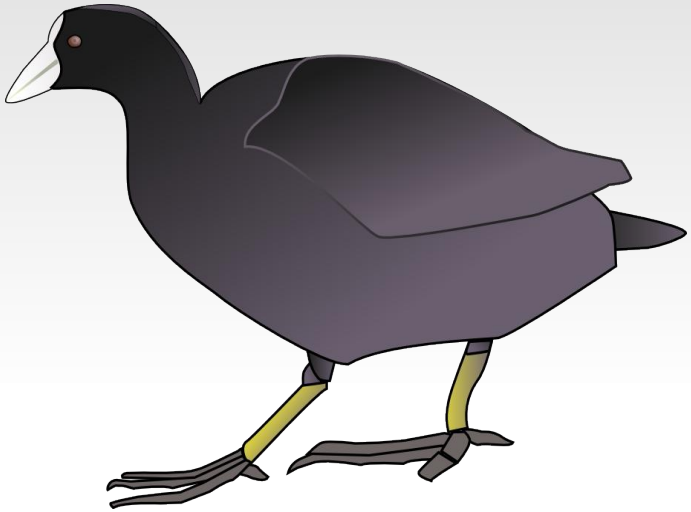
Model-Building with Coot



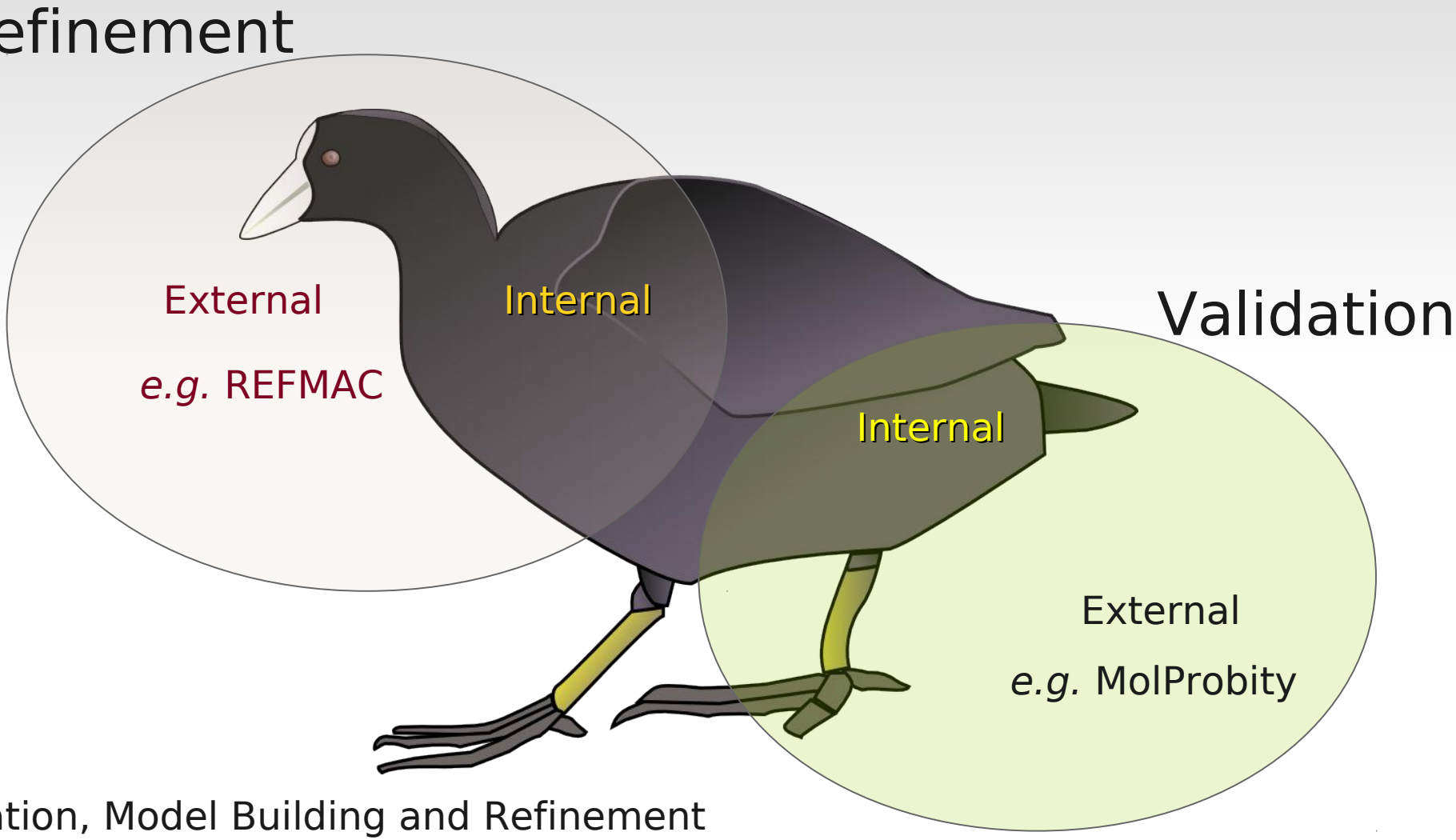
Paul Emsley
University of Oxford
June 2011

Model-Building with Coot



- Overview of Presentation:
  - New algorithms and features
  - Tips on Usage
  - Focus on:
    - Tools for Ligand-handling
    - Tools for Low Resolution X-ray data
    - Tools for EM
    - Future Directions

Feature Integration



Validation, Model Building and Refinement should be used together

Real Space Refinement

Diamond, R. (1971). *Acta Cryst*. A 27, 436-452.

- 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 since the start

Representation of Results:

File Edit View Terminal Help

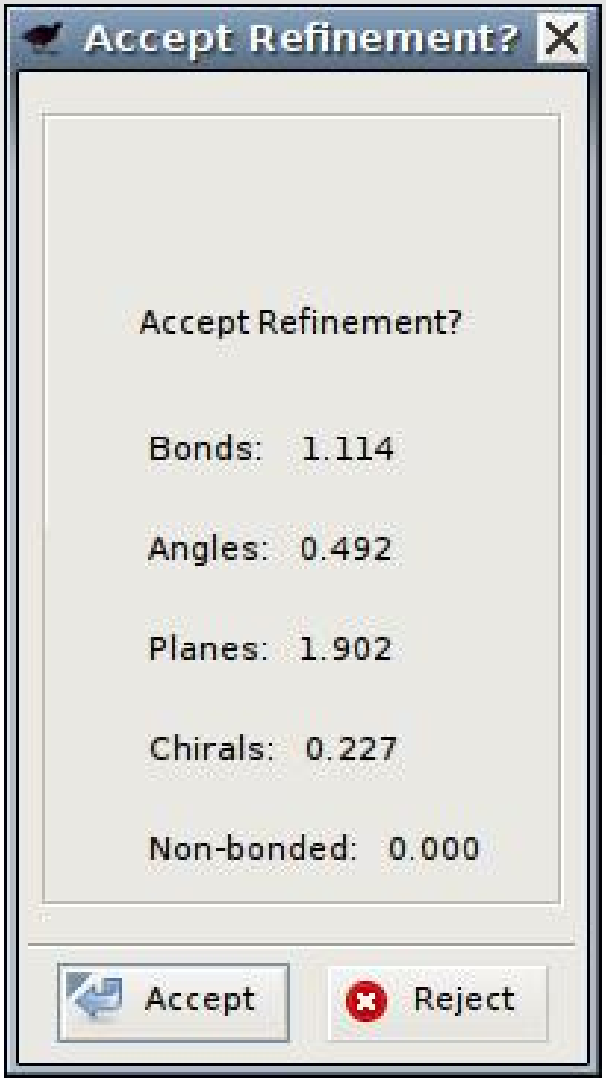
```
created 32 bond       restraints
created 38 angle      restraints
created 1 plane      restraints
created 5 chiral vol restraints
created 76 restraints

   INFO:: [spec: "A" 45 ""] [spec: "A" 46 ""] link_type :TRANS:
   INFO:: [spec: "A" 45 ""] [spec: "A" 44 ""] link_type :TRANS:
Link restraints:
   2 bond    links
   6 angle   links
   4 plane   links
Flanking residue restraints:
   4 bond    links
   12 angle   links
   8 plane   links
INFO:: made 668 non-bonded restraints
initial distortion_score: -16033.2
    Initial Chi Squareds
bonds:      1.15701
angles:     0.847832
torsions:   N/A
planes:     1.6176
non-bonded: 0
chiral vol: 0.705728
rama plot:  N/A
Minimum found (iteration number 67) at -16275.9
    Final Estimated RMS Z Scores:
bonds:      1.19412
angles:     0.713337
torsions:   N/A
planes:     1.05134
non-bonded: 0
chiral vol: 0.522415
rama plot:  N/A
 SUCCESS
 TIME:: (dragged refinement): 332.657
```

The first attempt

Student Reaction:

“Oh, I don't look at that window…”
(I maximise the window immediately)

Representation of Results:



Second attempt...

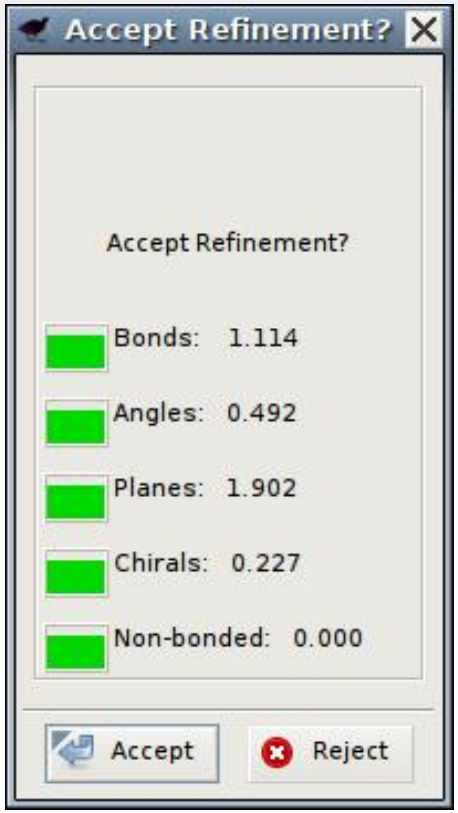
Student Reaction:

“Oh, box of meaningless numbers.

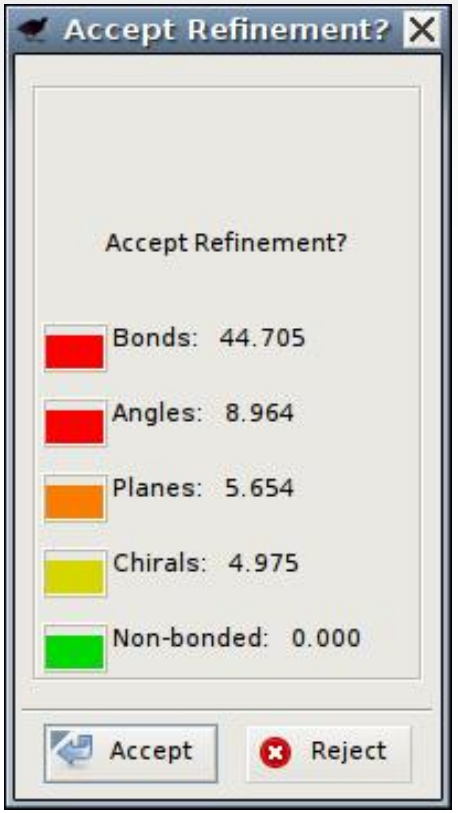
Go away”

Representation of Results: “Traffic Lights”

“Traffic Lights” represent the RMSd values for each of the refined geometry types



Good refinement



Bad refinement

Refinement Techniques

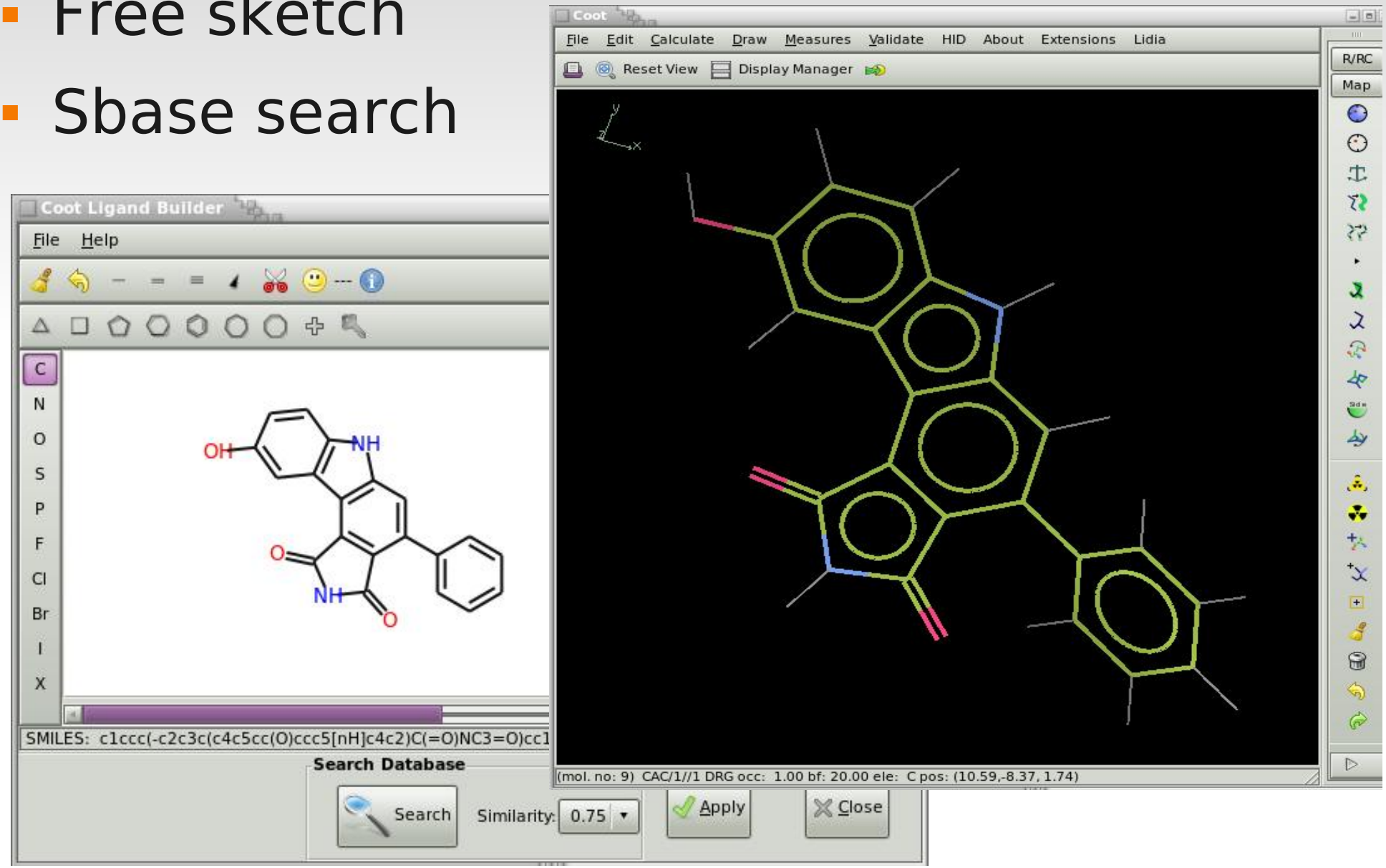
- Single-Atom Drag
  - Over-dragging
- Key-bindings:
  - Triple Refine
  - Single Residue Refine with Auto-accept
- Ramachandran Refinement

Ligands in *Coot*

- Importing and building ligand from scratch
  - PRODRG, LIBCHECK
- Ligand Fitting
- Validation
  - Mogul
- Representation
  - Bond orders
  - Surfaces
- Analysis
  - Molprobity, LIDIA

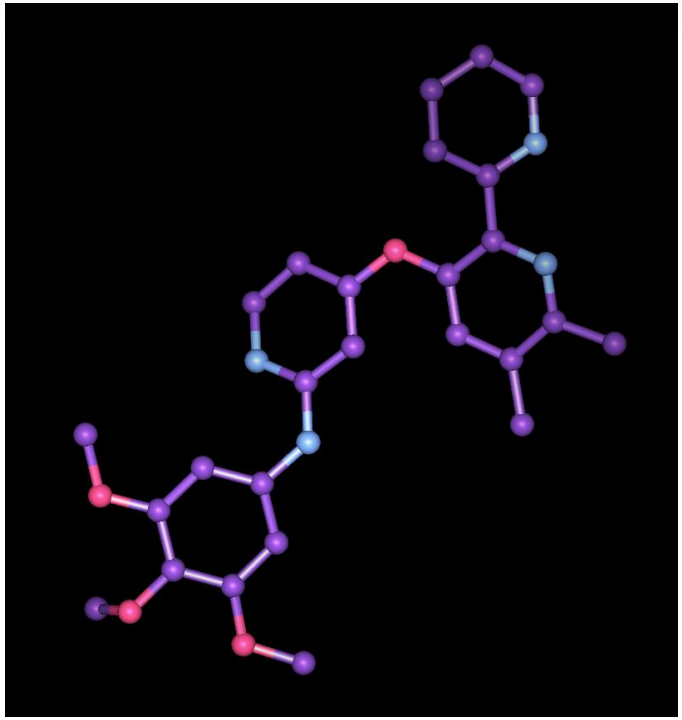
2D Ligand Builder

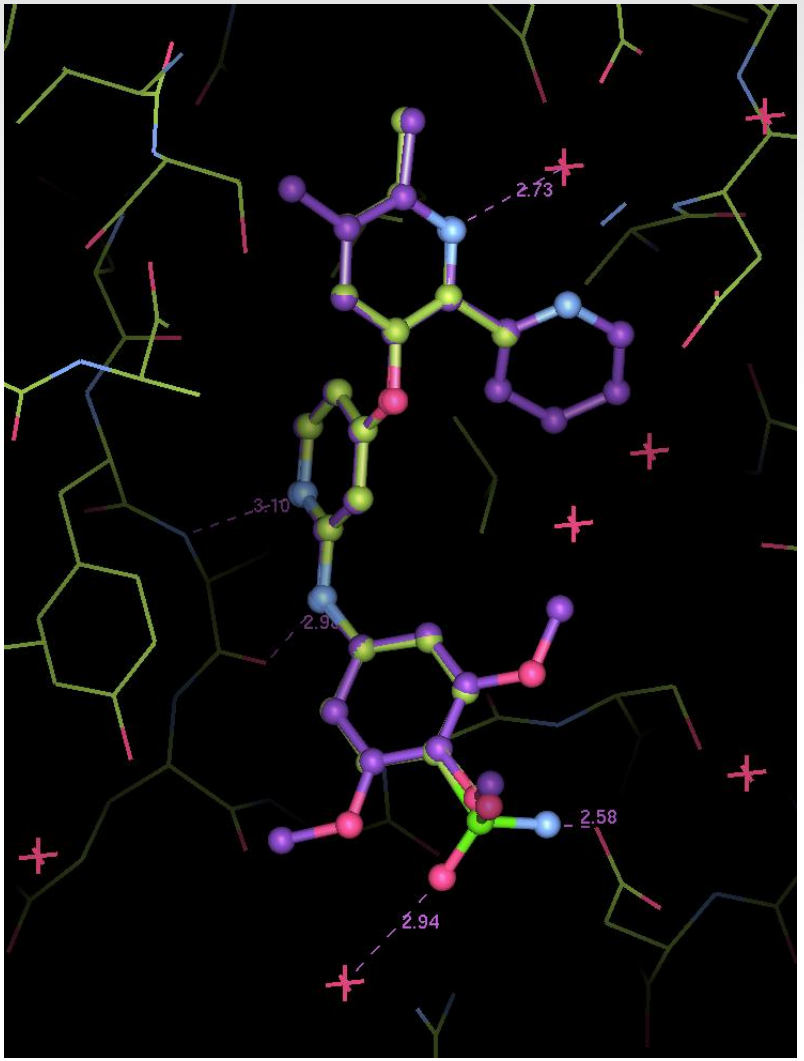
- Free sketch
- Sbase search



”Yesterday's” Ligand

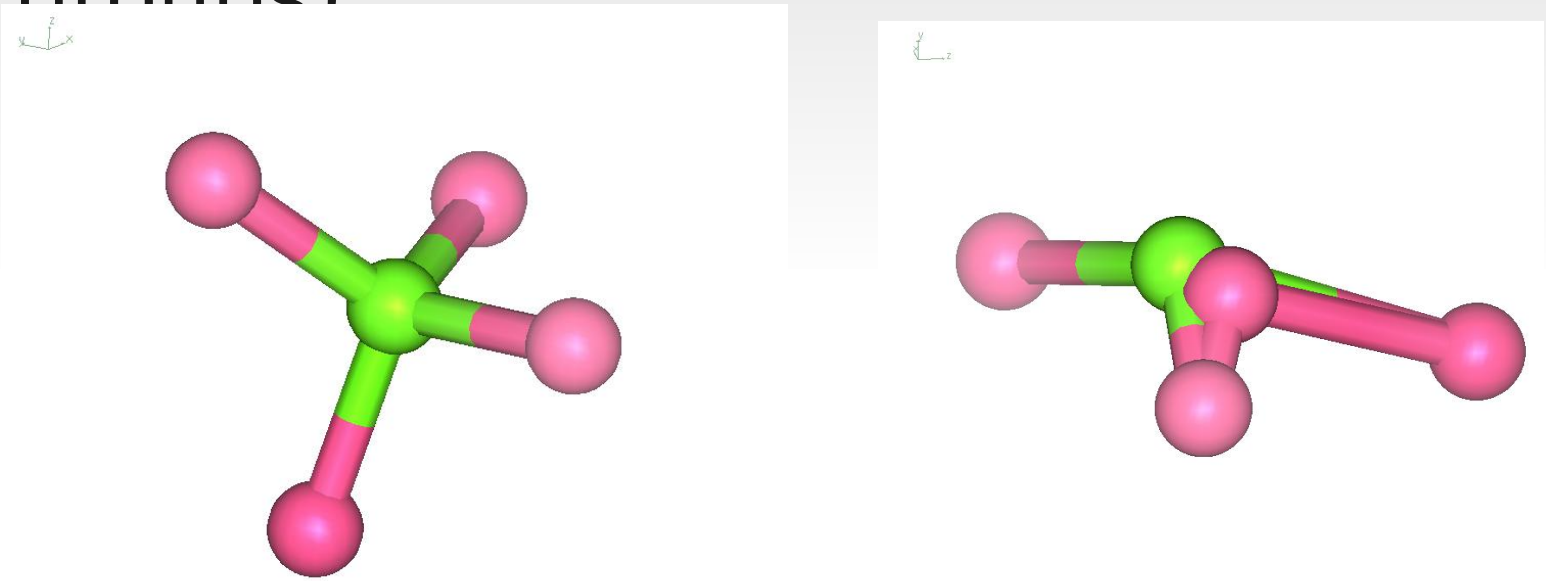
- Atom name matching
- Torsion matching
- Ligand overlay





Why Validate?

- 11,000+ chemical structures (Het-groups)

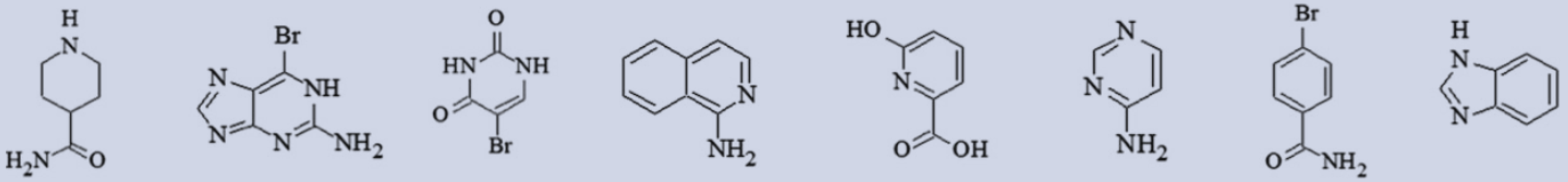


Sulfate ions in 1DW9 (1.65Å resolution)

Ligand Site

Ligand Type	Known	Unknown
Known	✓	✓
Cocktail	✓	✓
Unknown	✗	✗

Cocktail Examples

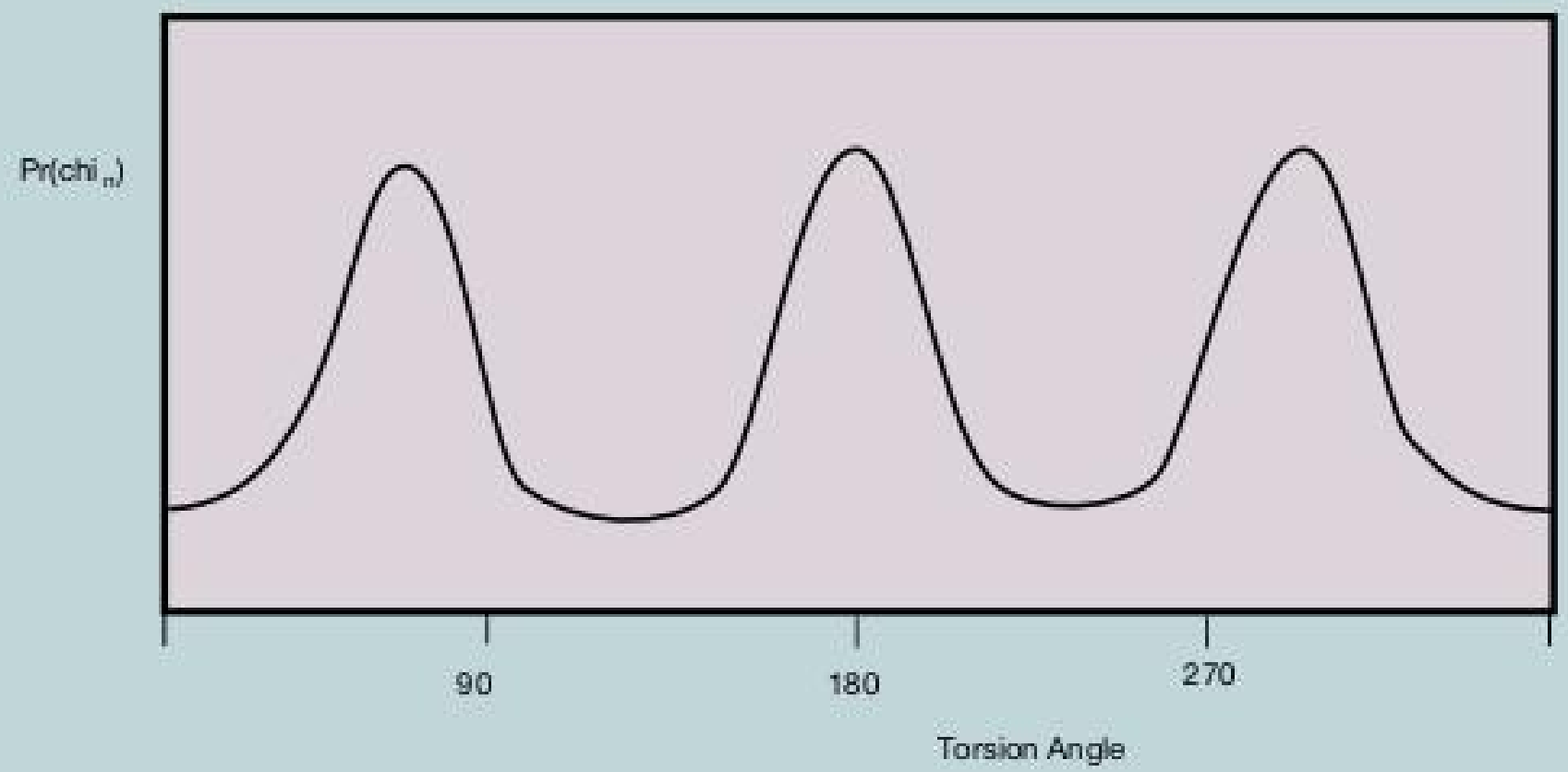


Conformer Generation

REFMAC Monomer Library
chem_comp_tor

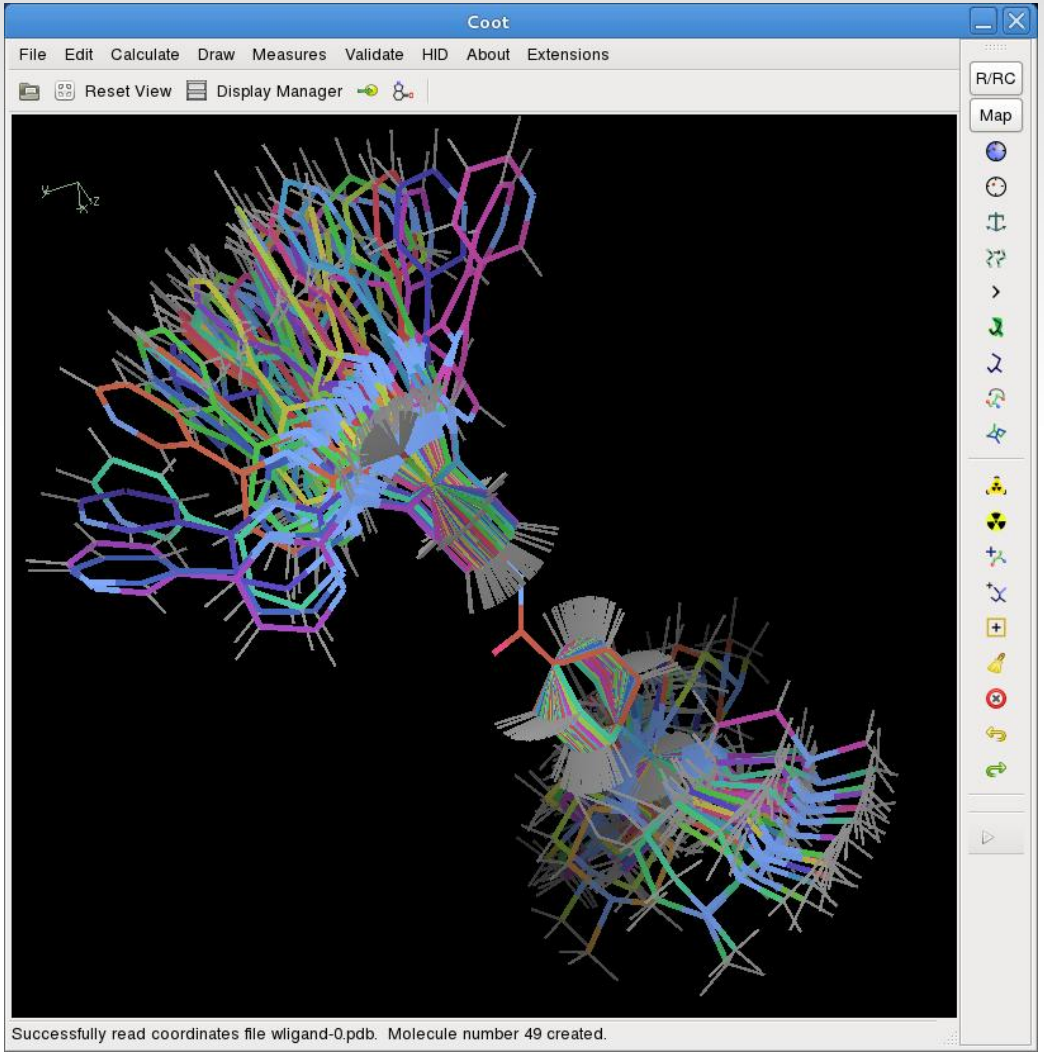
```
loop_
_chem_comp_tor.comp_id
_chem_comp_tor.id
_chem_comp_tor.atom_id_1
_chem_comp_tor.atom_id_2
_chem_comp_tor.atom_id_3
_chem_comp_tor.atom_id_4
_chem_comp_tor.value_angle
_chem_comp_tor.value_angle_esd
_chem_comp_tor.period
 ADP      var_1    O2A    PA     O3A    PB        60.005   20.000   1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

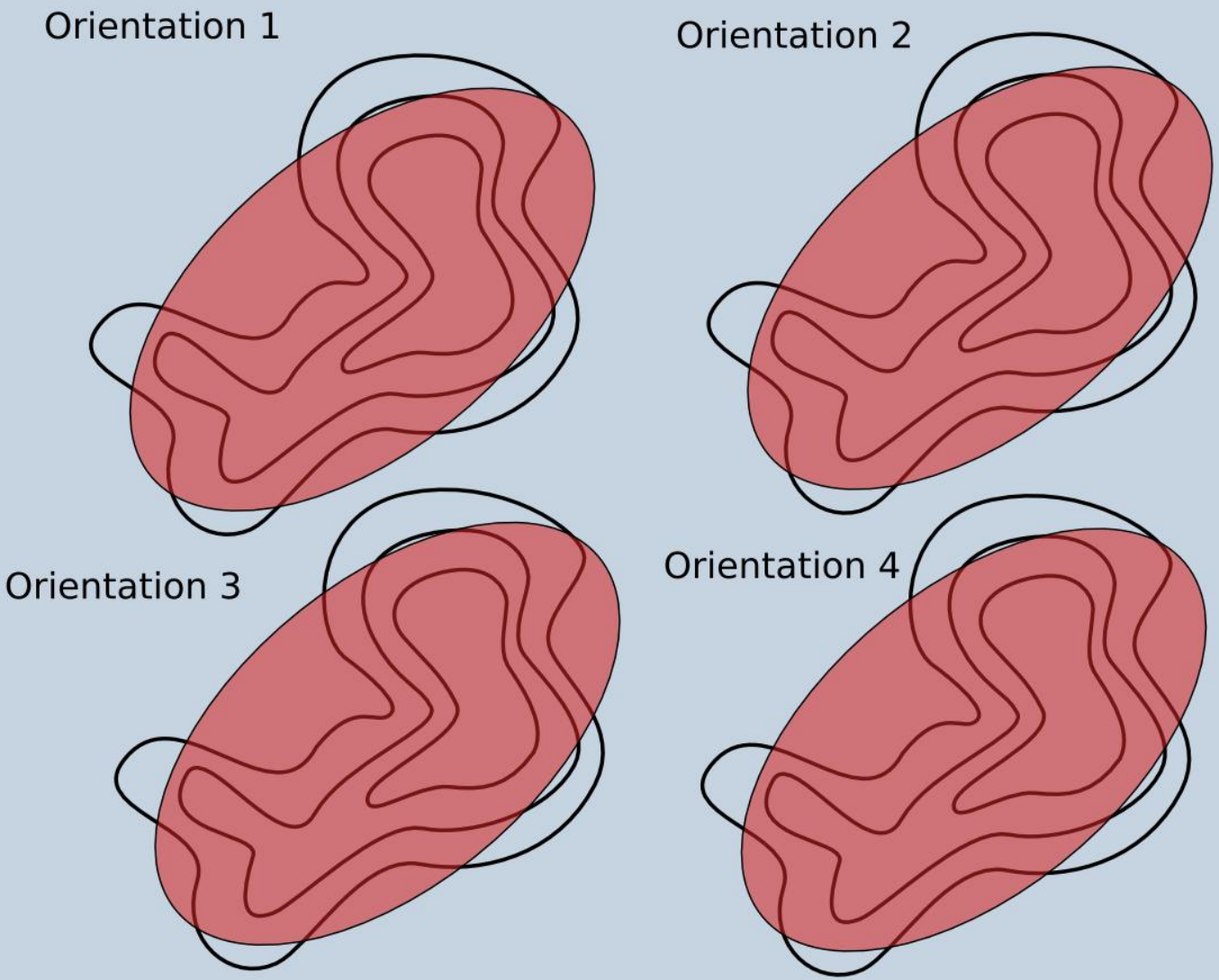


Conformer Generation

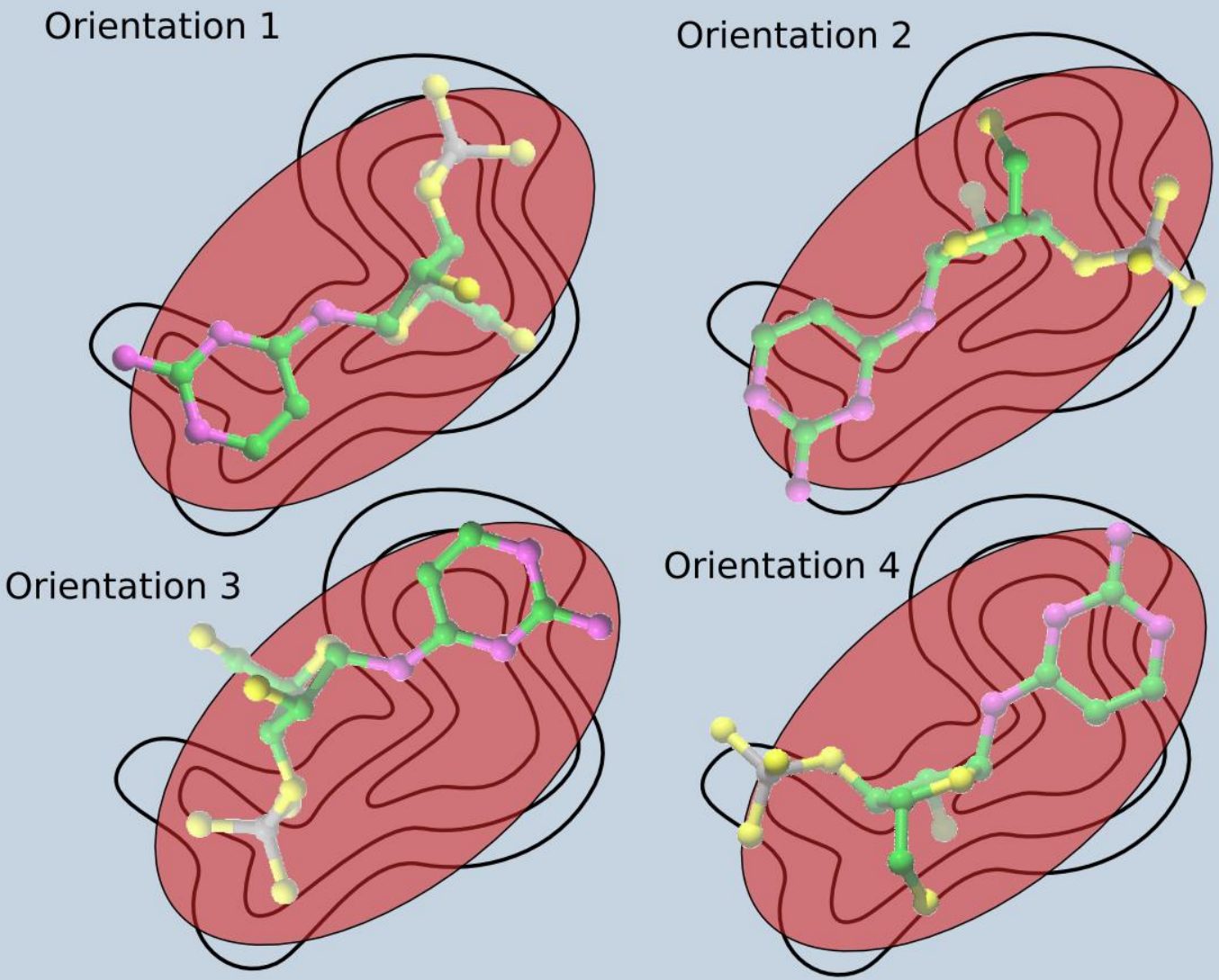
Non-Hydrogen
Non-CONST
Non-Ring



Orienting the Ligand



Orienting the Ligand

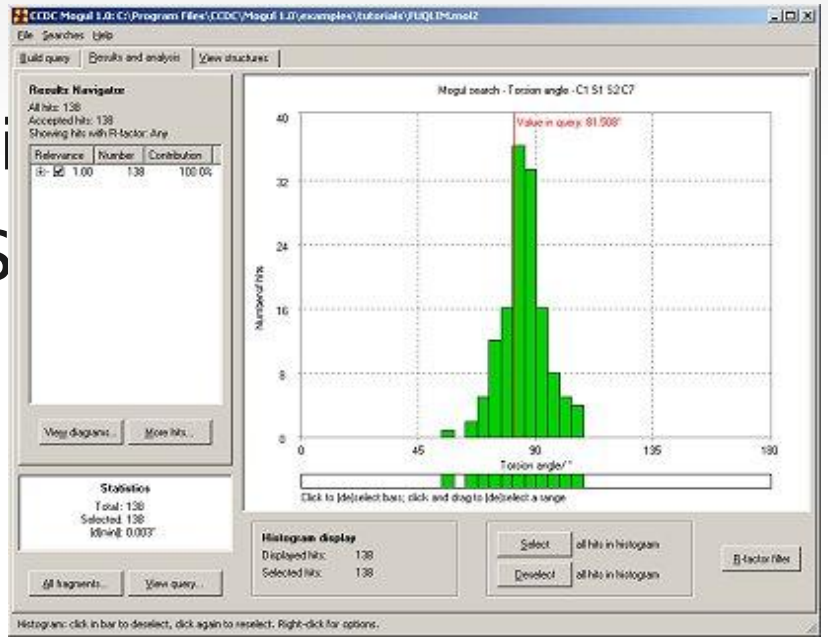


Ligand Validation

- Compare the observed structure to the restraints
  - In *Coot*: Validate → Geometry analysis

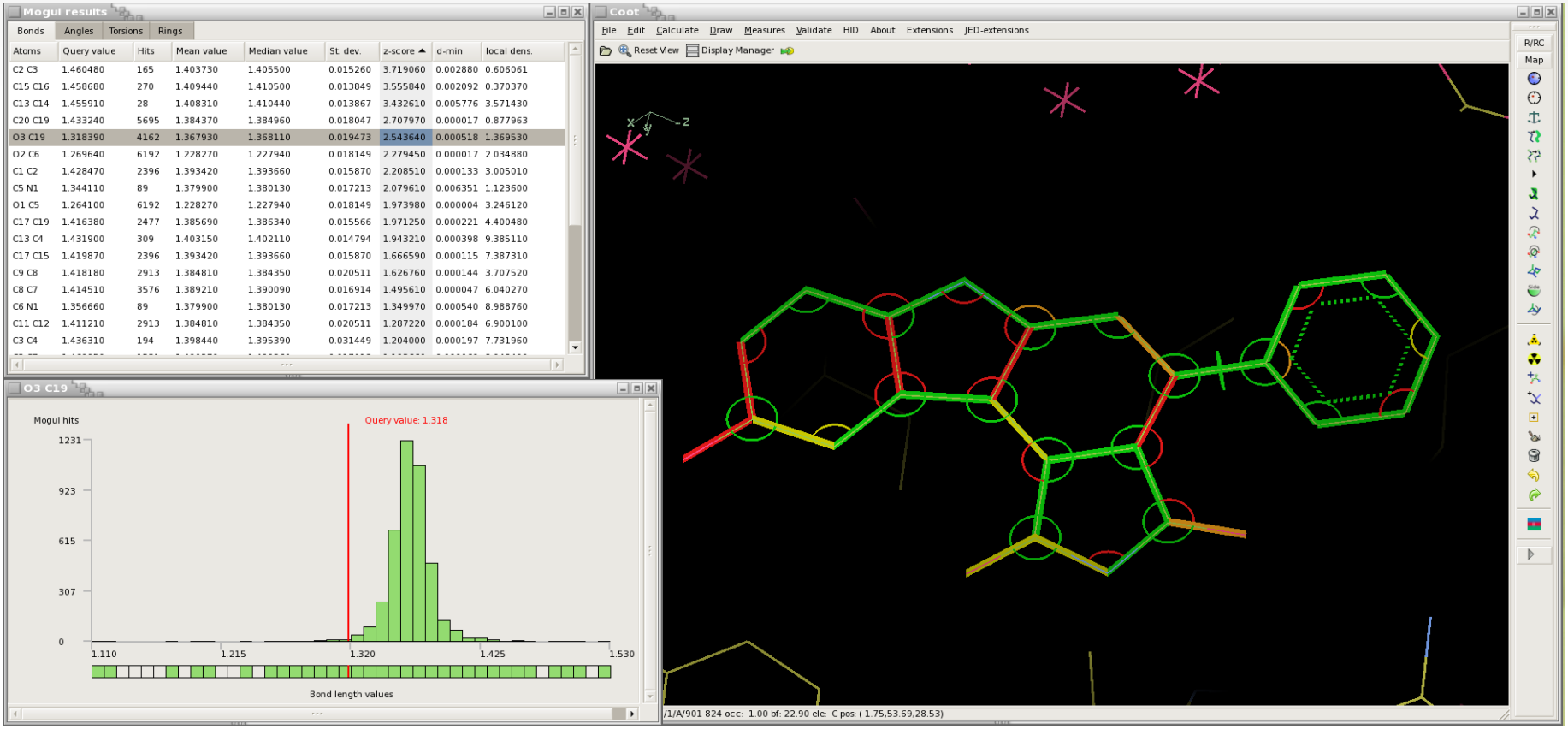
Parmatisation issues... (what if they are wrong?)

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



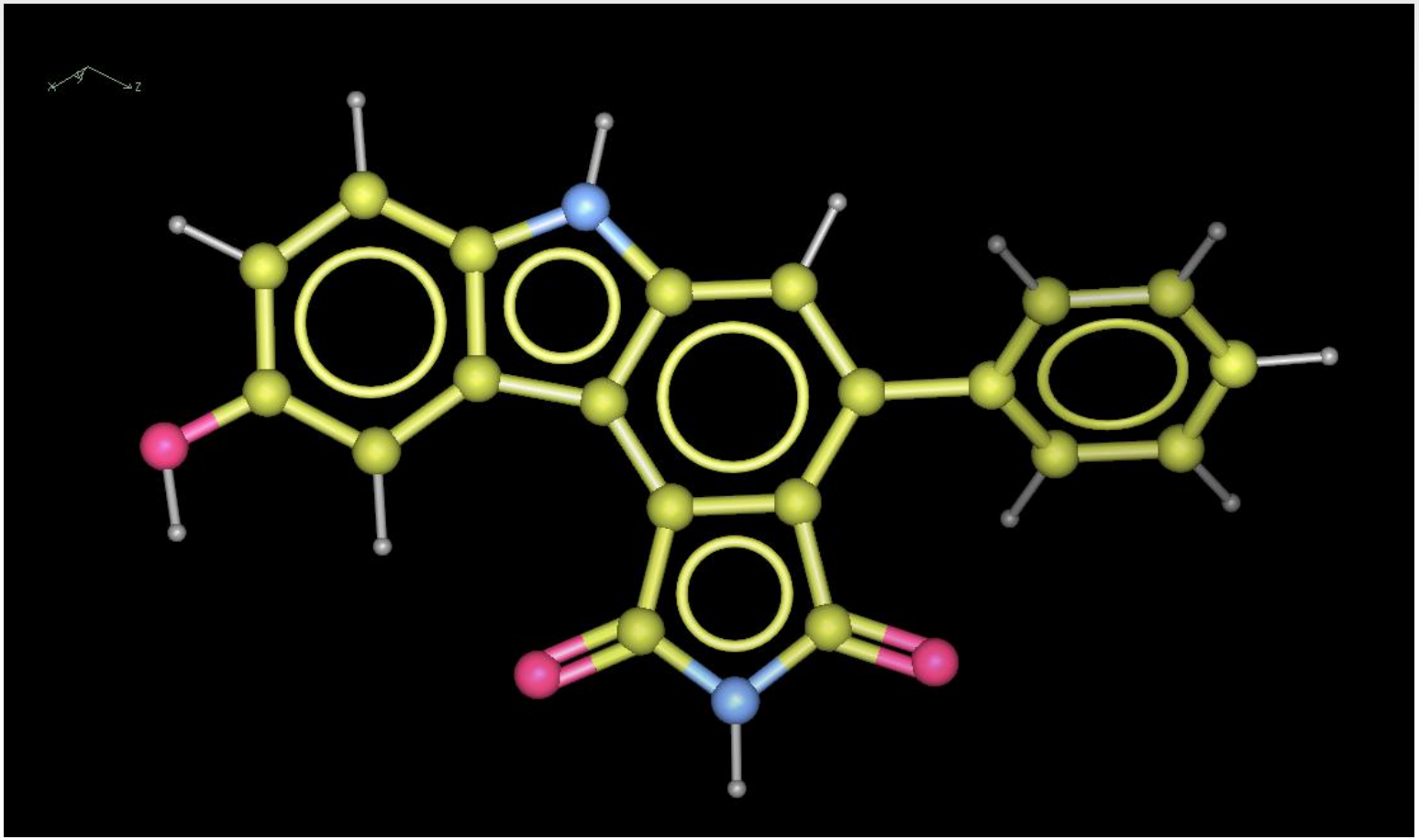
Ligand Validation

- Mogul plugin in Coot
  - Run mogul, graphical display of results
  - Update restraints (target and esds for bonds and angles)



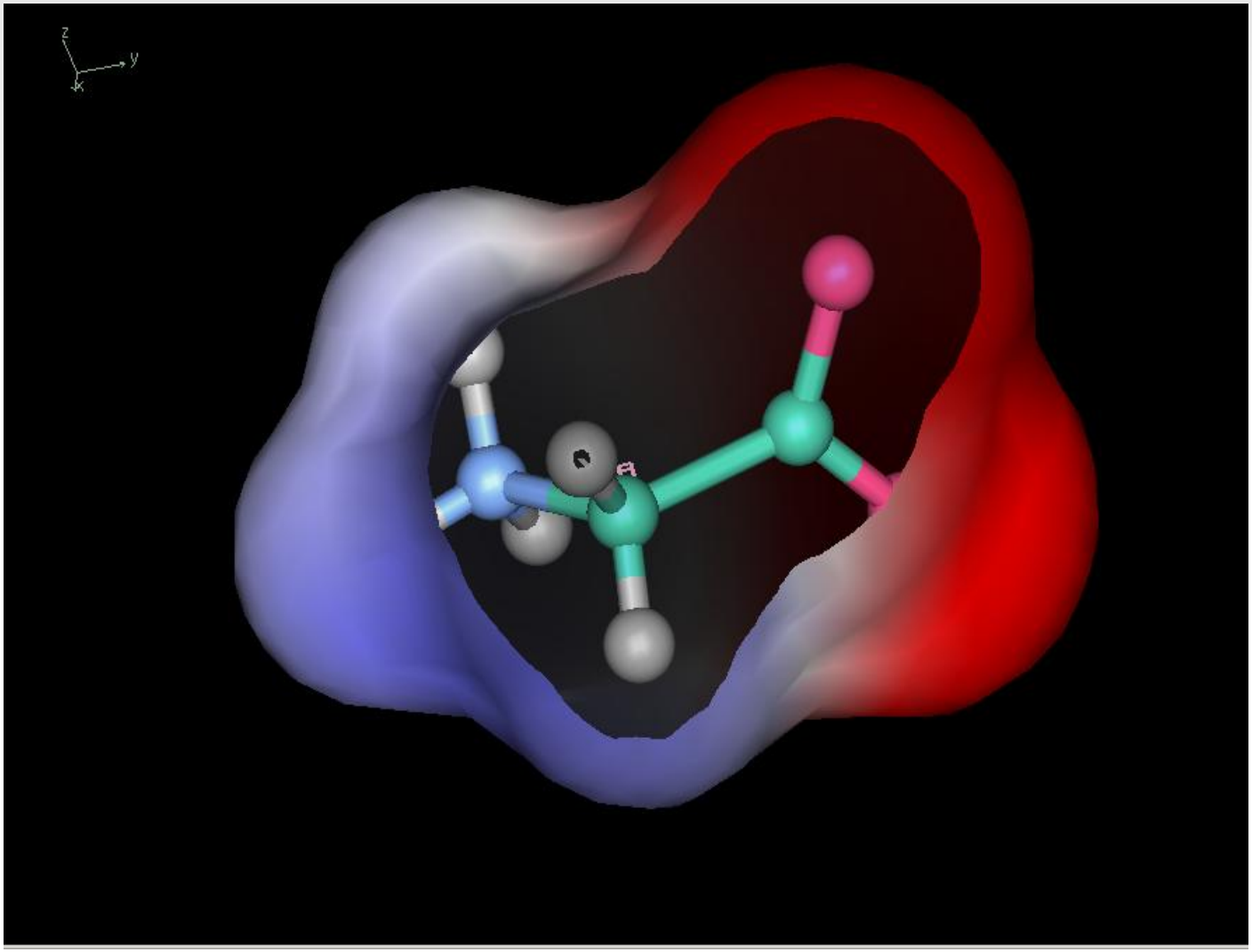
Ligand Represenation

- Bond orders (from dictionary)



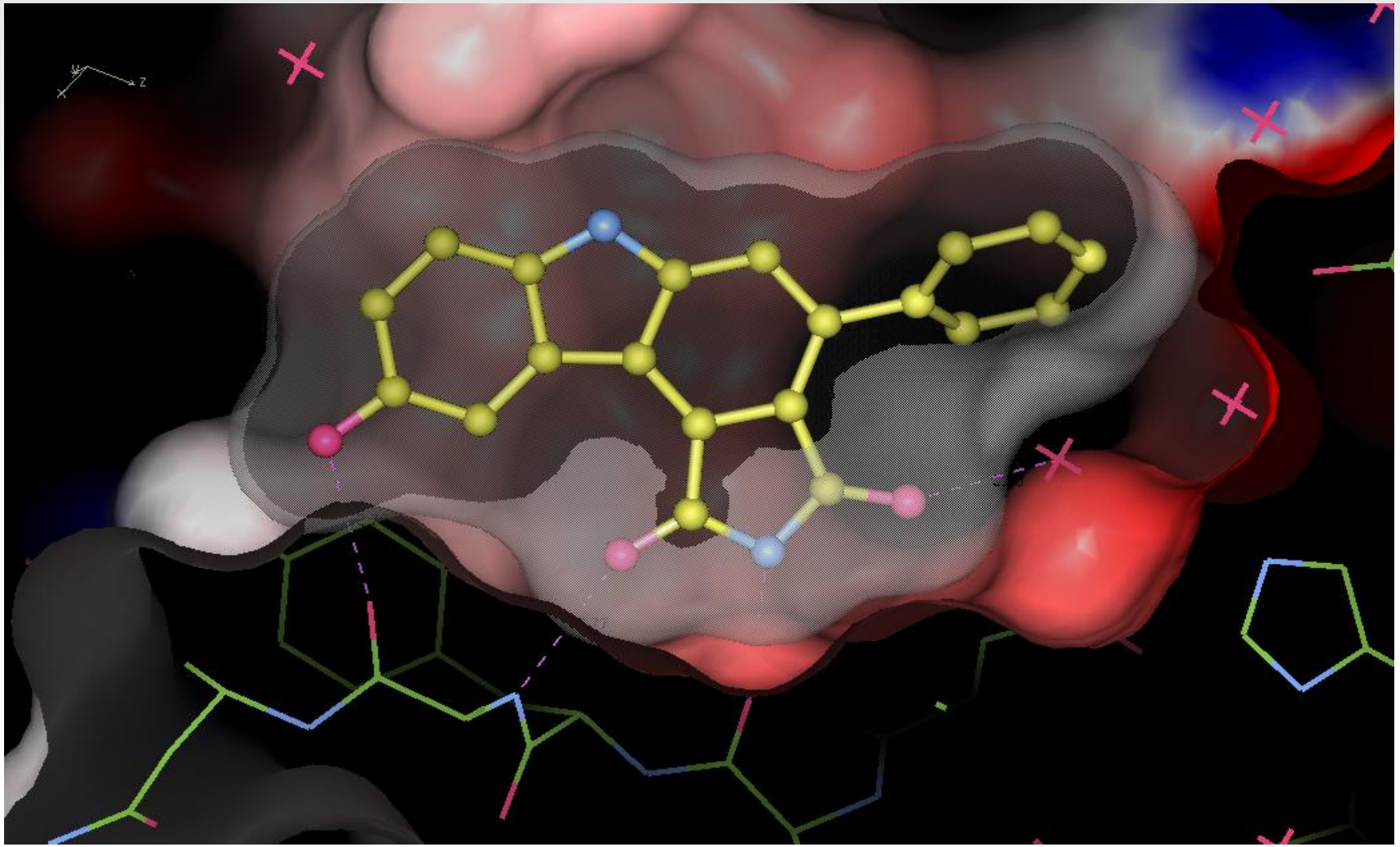
Ligand Representation

- Surfaces using Partial Charges



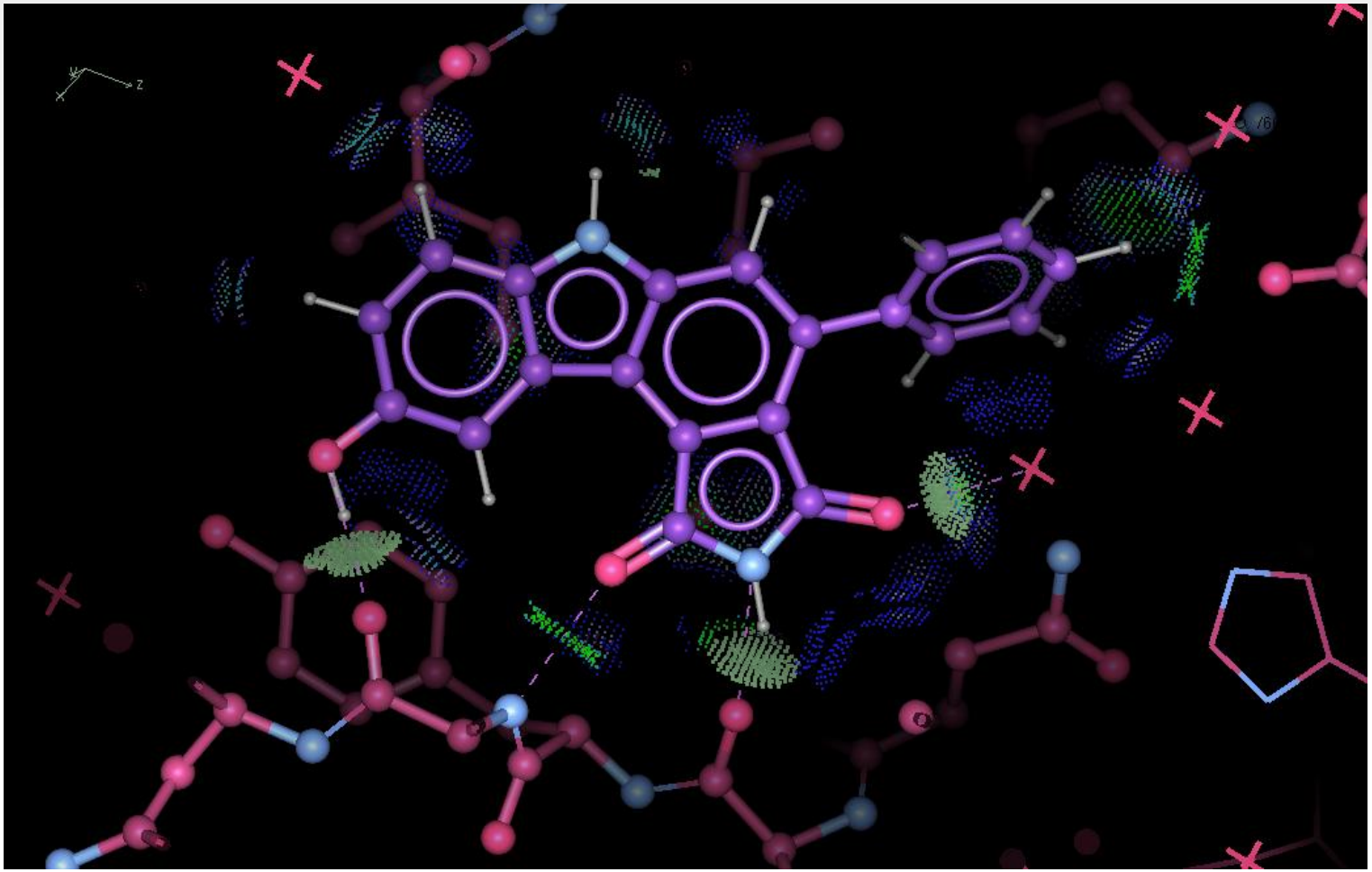
Ligand Representation

- Transparent surfaces



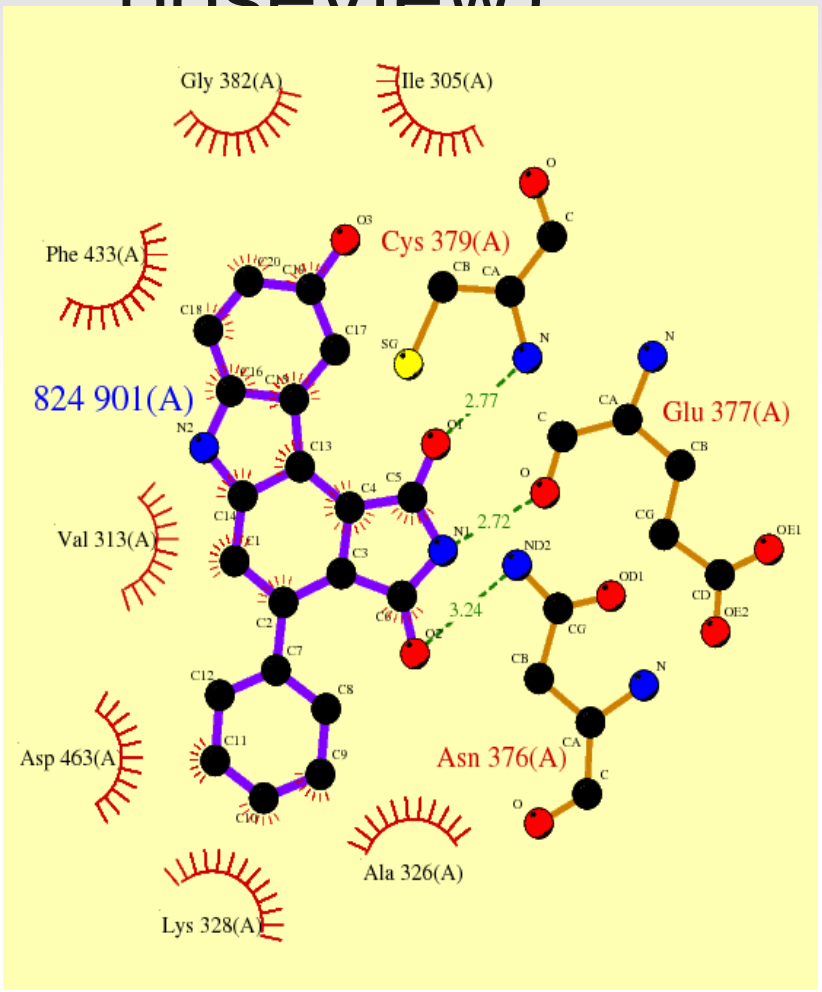
Binding mode analysis

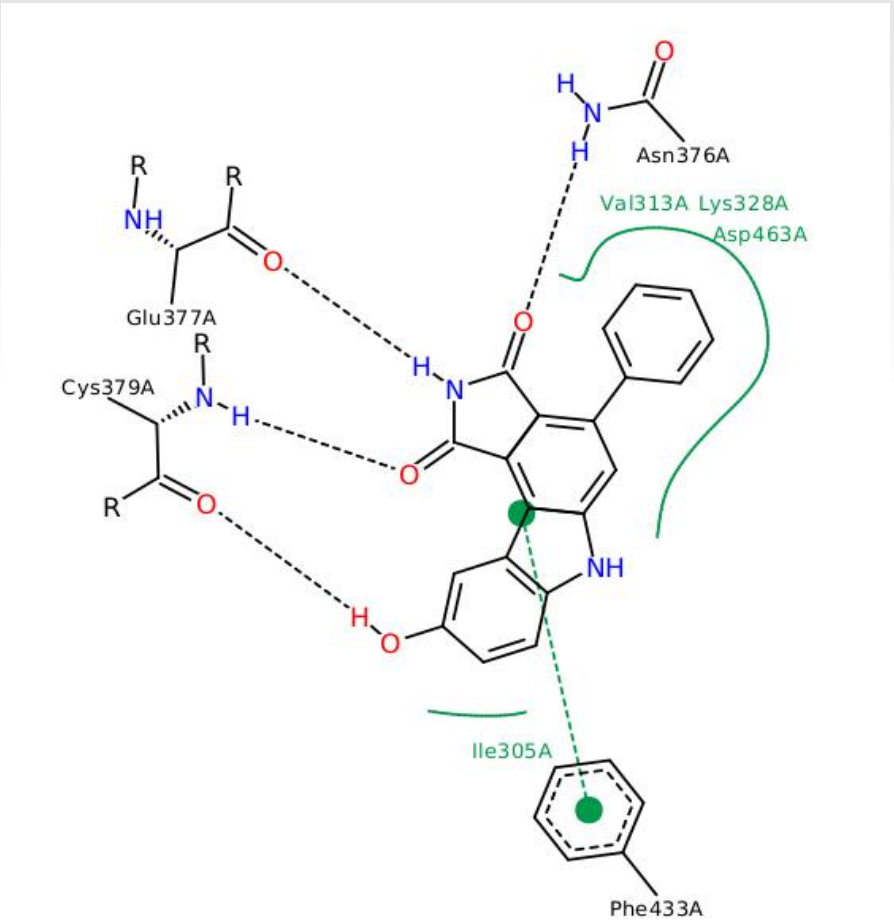
- Binding site highlighting,
- Isolated Molprobity dots



Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





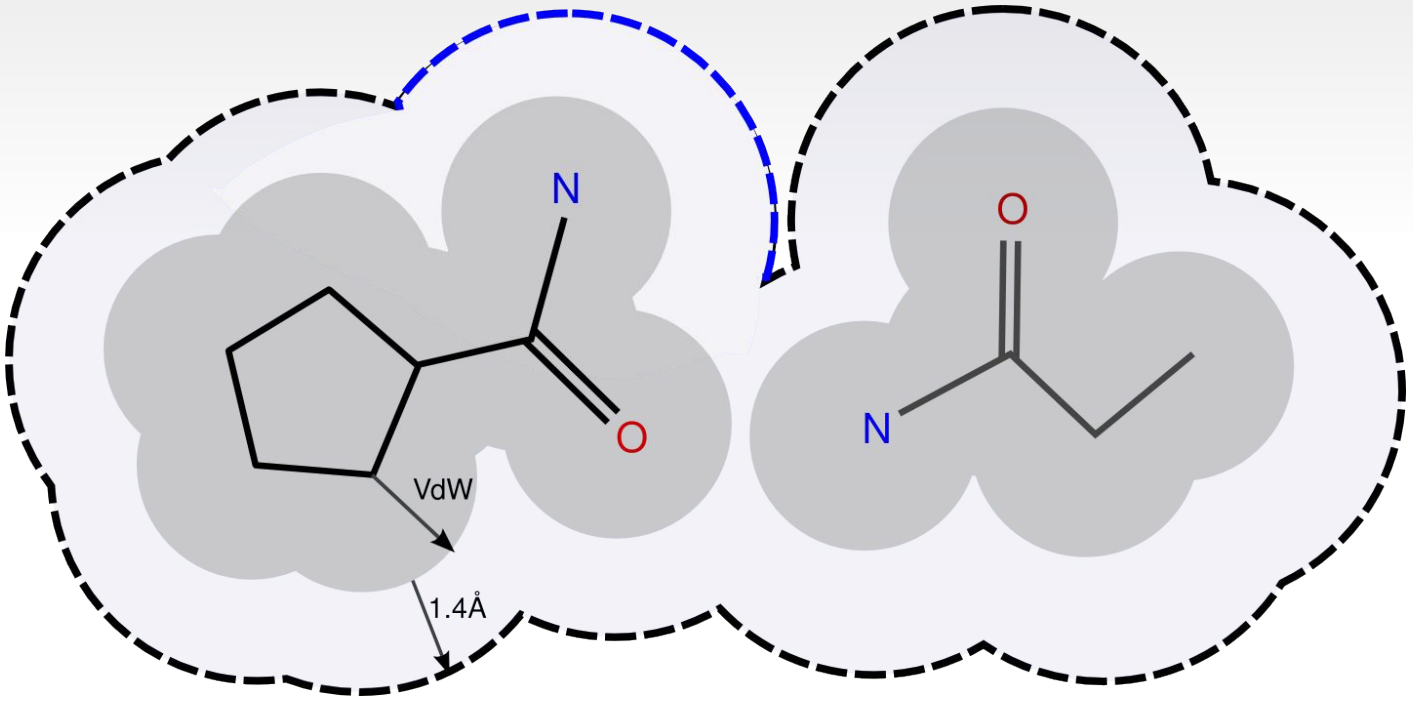
Can we do better? - Interactivity?

Ligand Environment Layout

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

Solvent Exposure

- Identification of solvent accessible atoms



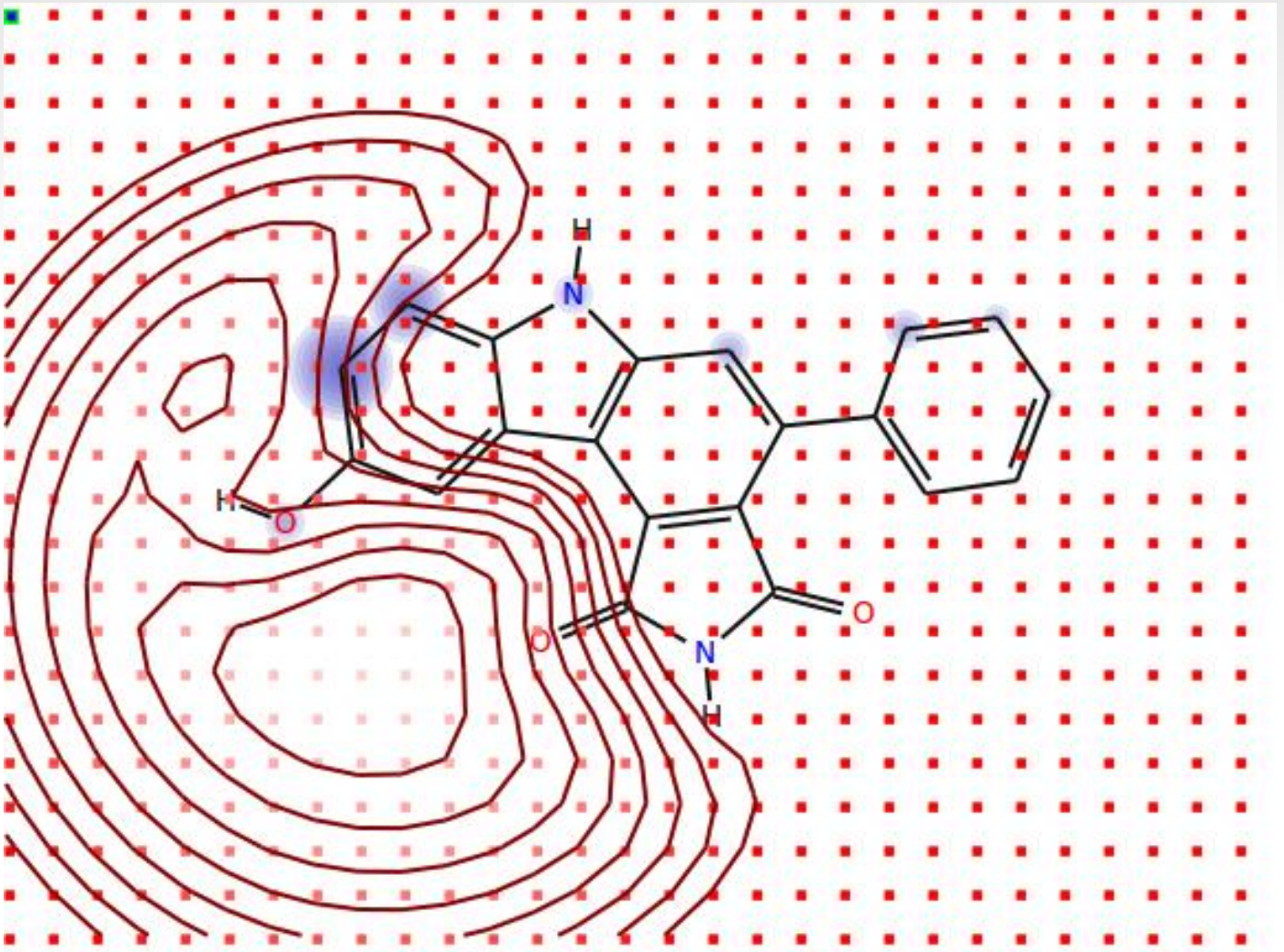
Ligand Enviroment Layout

- 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)

    (work in progress)

Ligand Environment Layout

- Initial residue placement



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

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

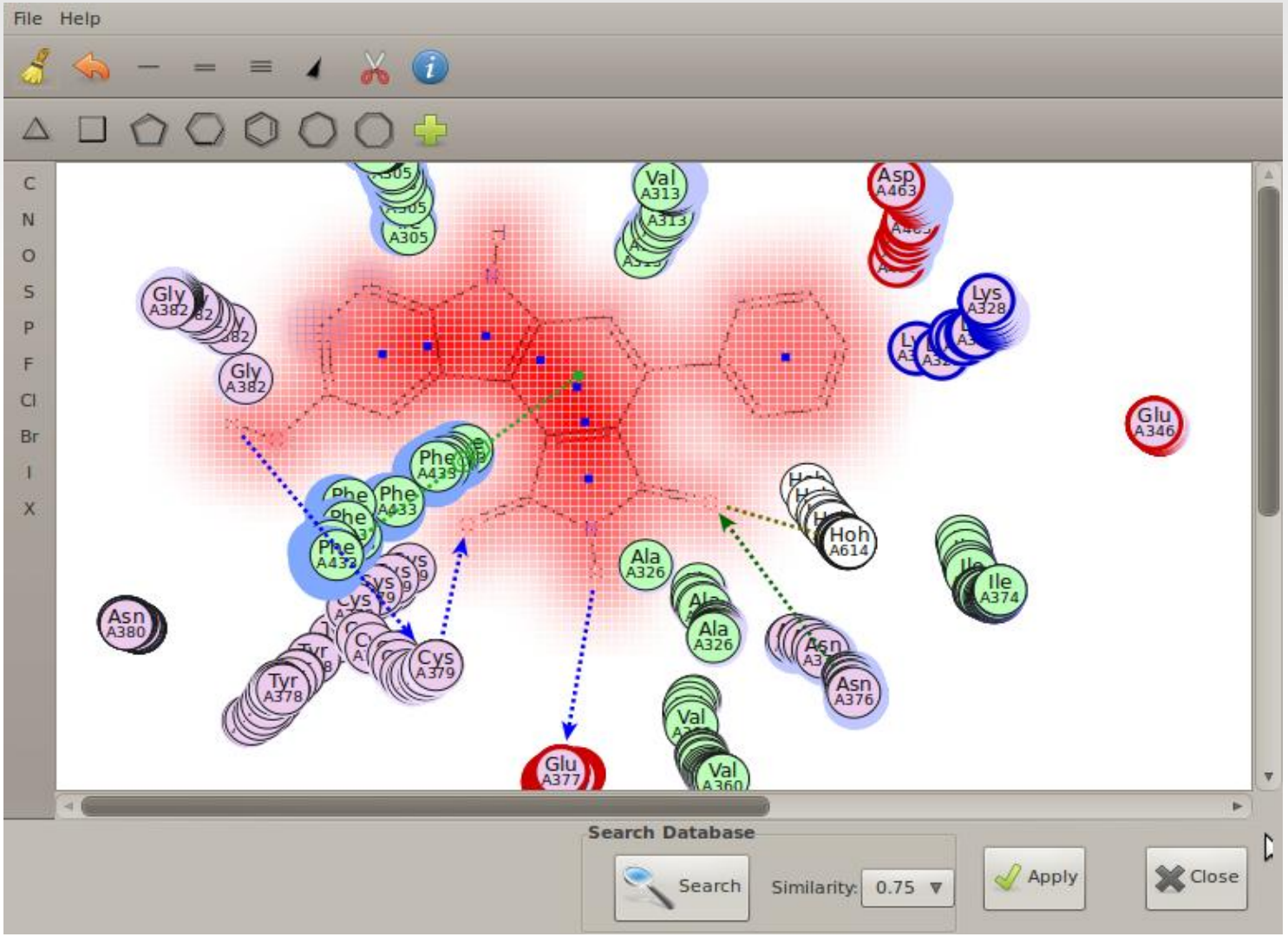
Residues are close to H-bonding ligand atoms

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

Residues don't overlap ligand

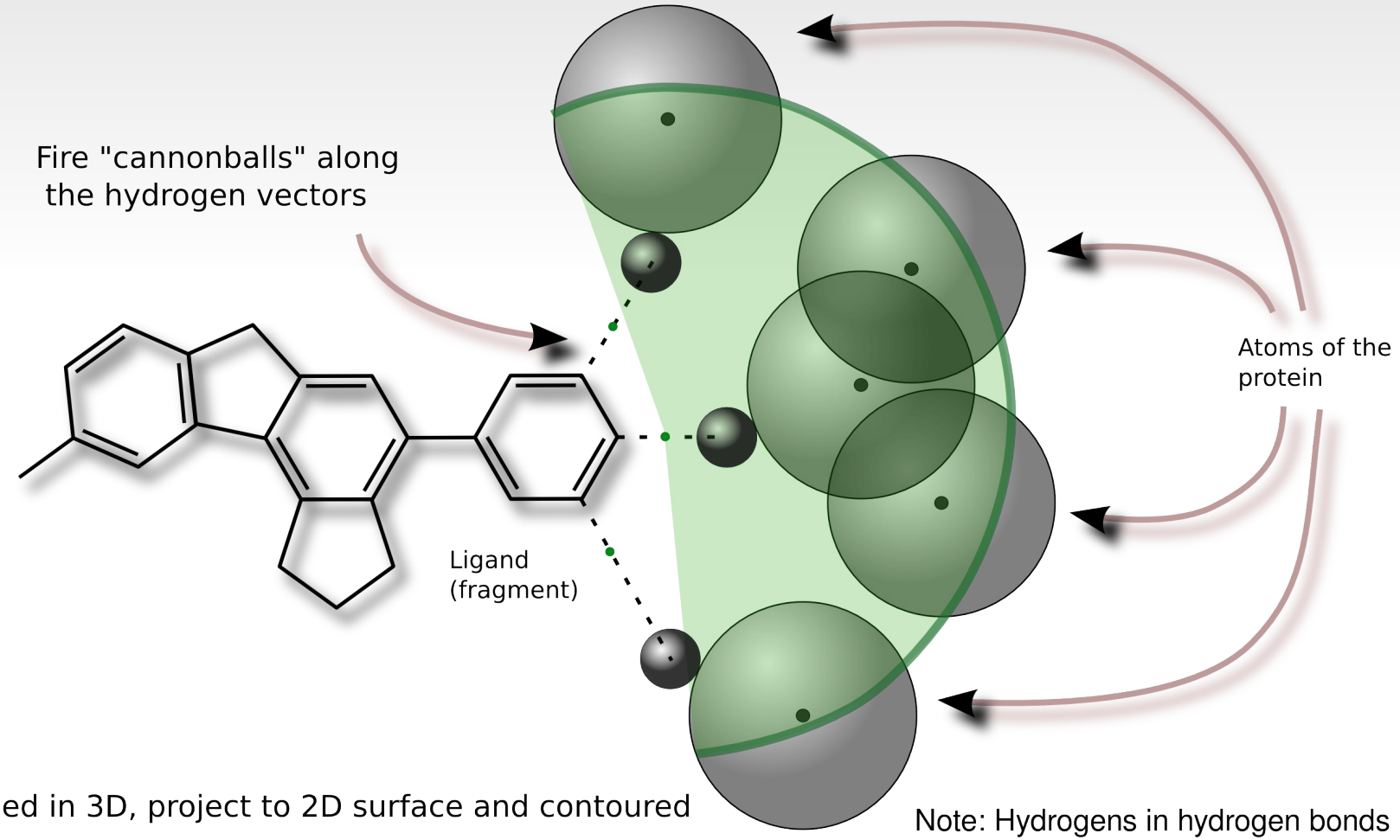
Ligand Environment Layout

- 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?



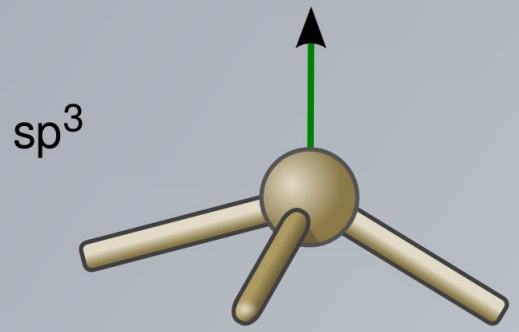
Determined in 3D, project to 2D surface and contoured

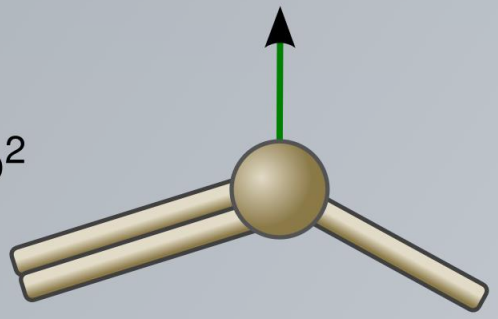
c.f. Clarke & Labute (2007)

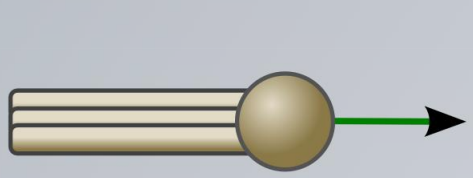
Note: Hydrogens in hydrogen bonds are a confounding factor

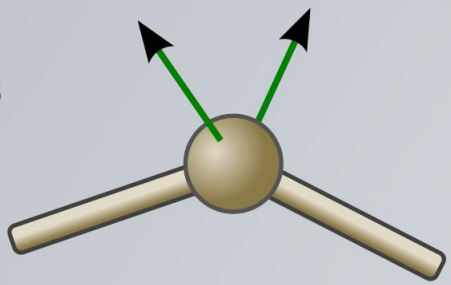
Substitution Contour: Extending along Hydrogens

Riding Hydrogens



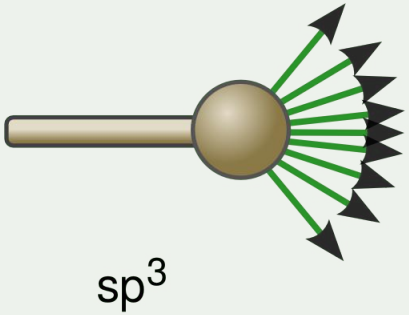


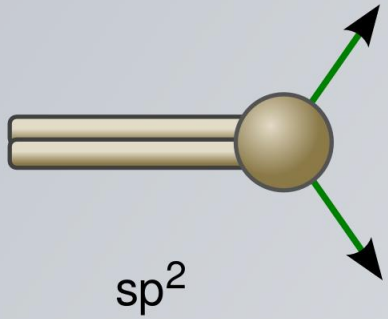


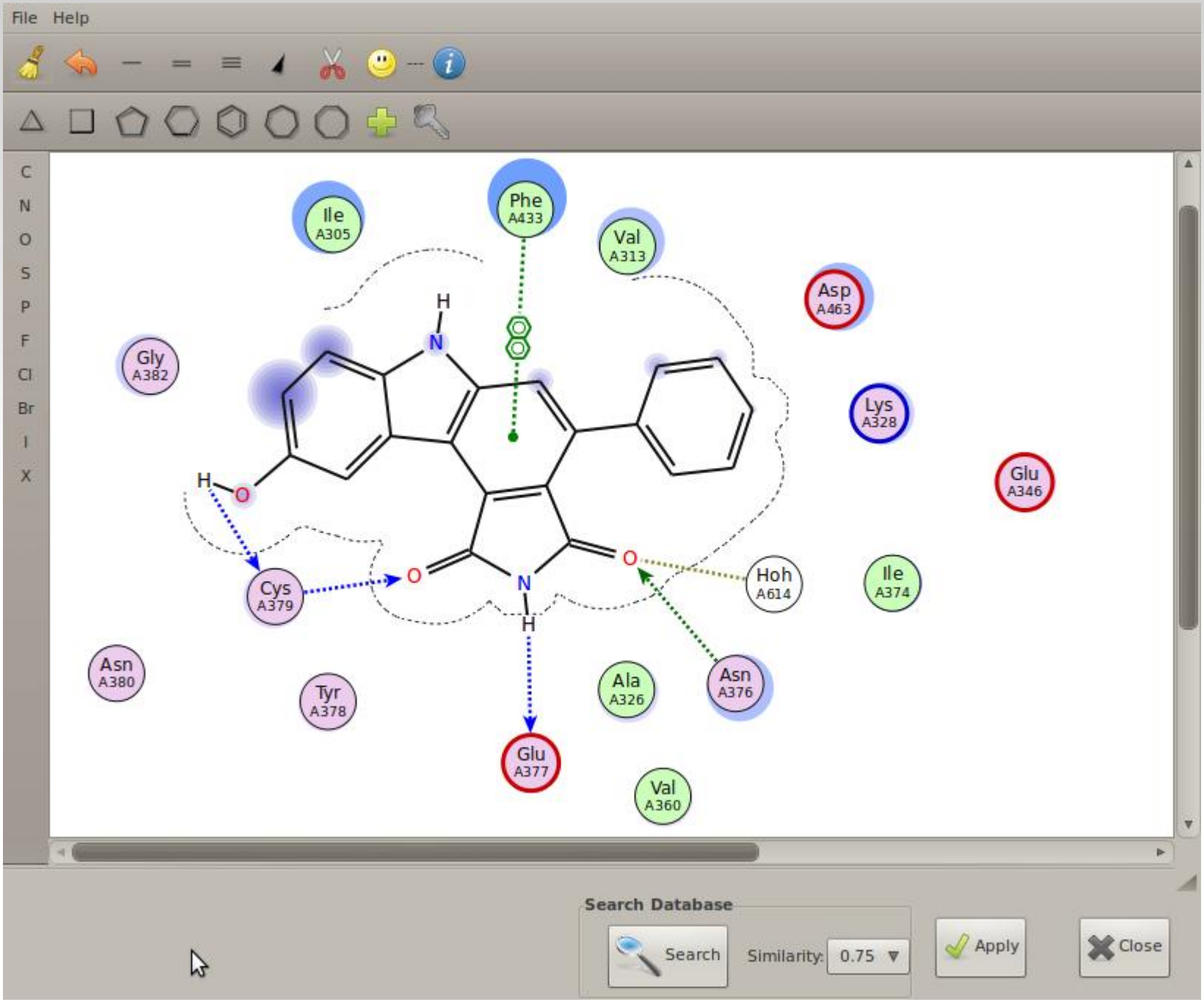


Torsionable Hydrogens

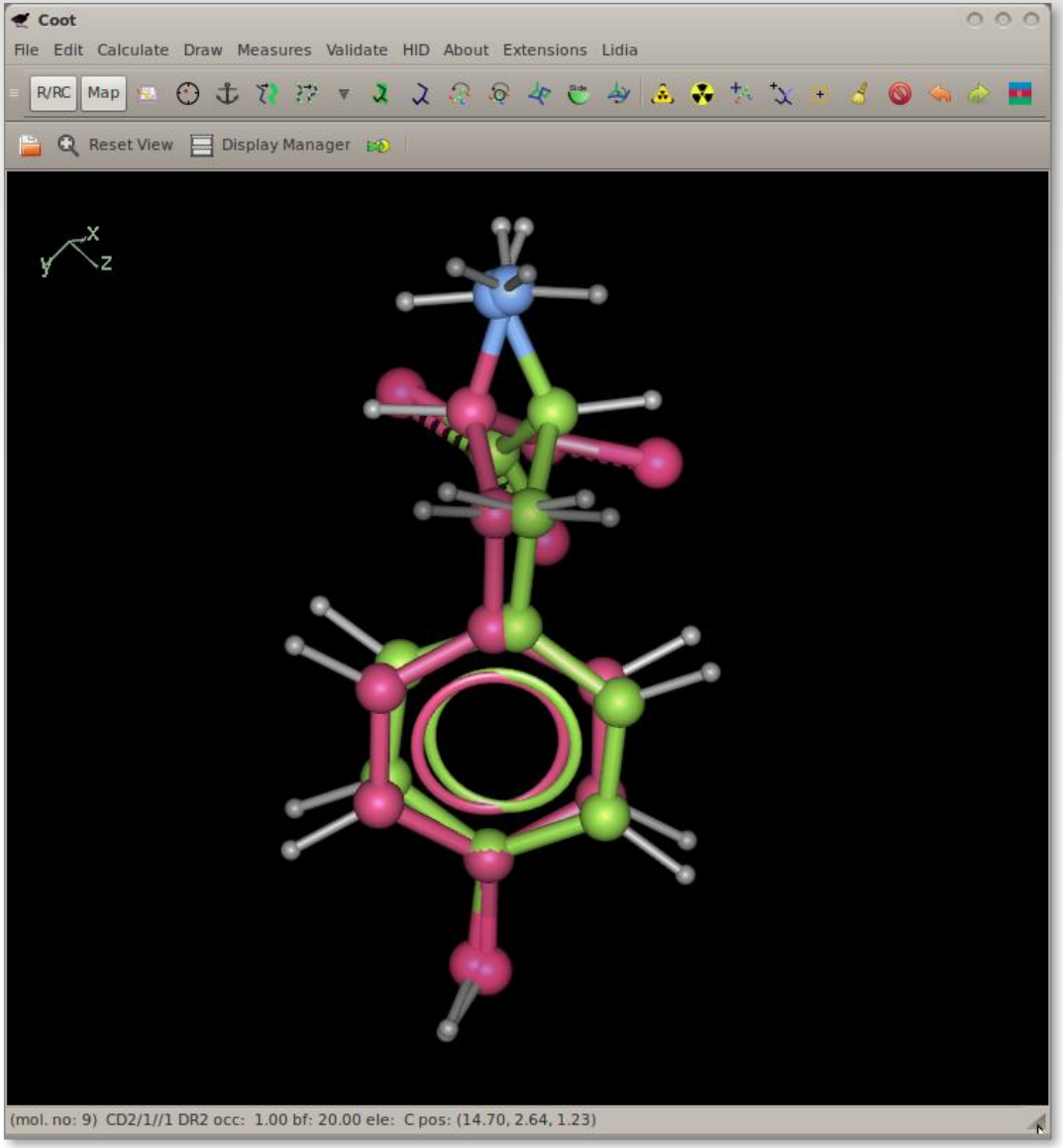
(test multiple directions)






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

Chiral Centre Inversion



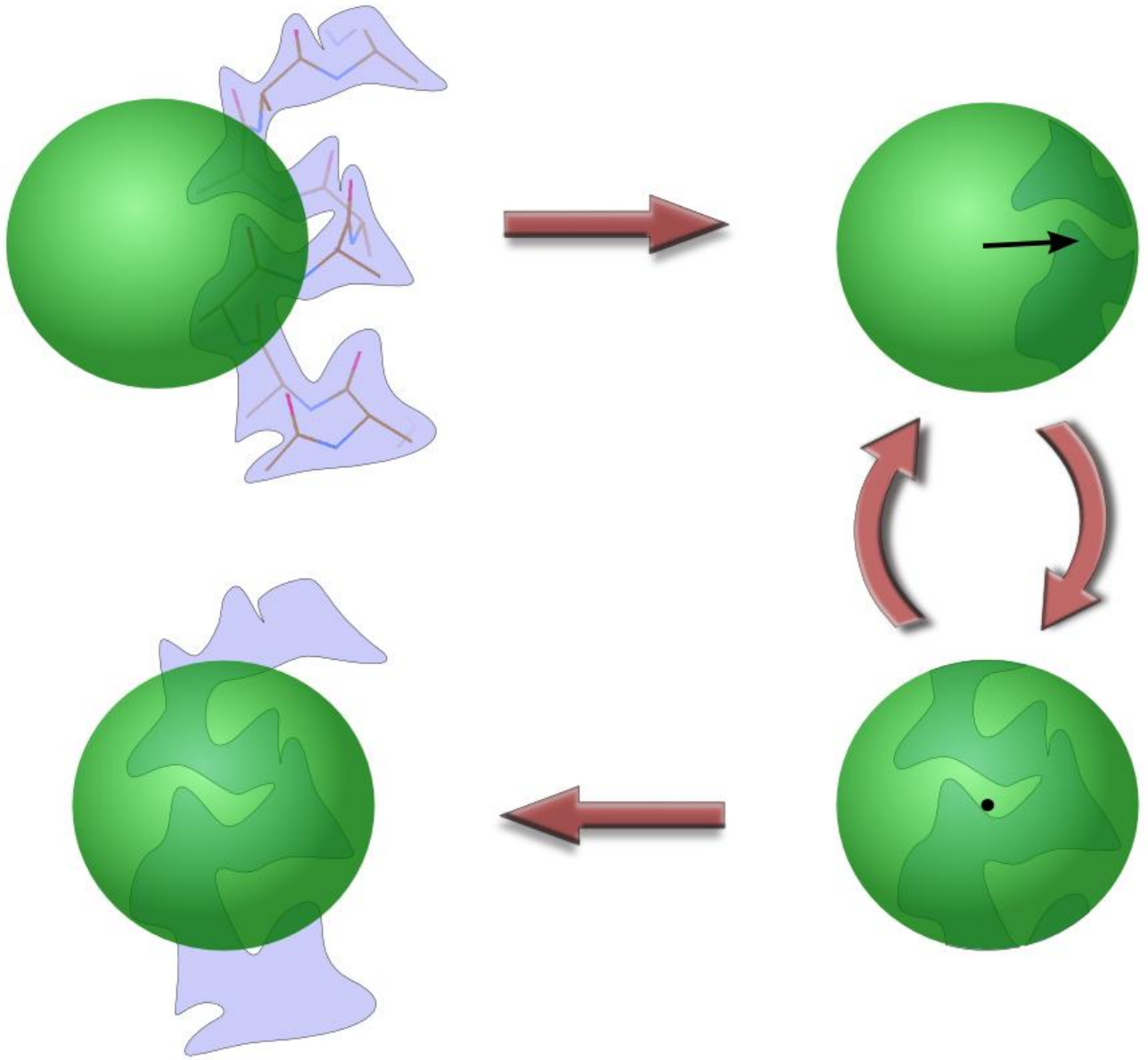
Inverted chiral centre refinement pathology detection

Hydrogen tunnelling

Alpha Helix Placement

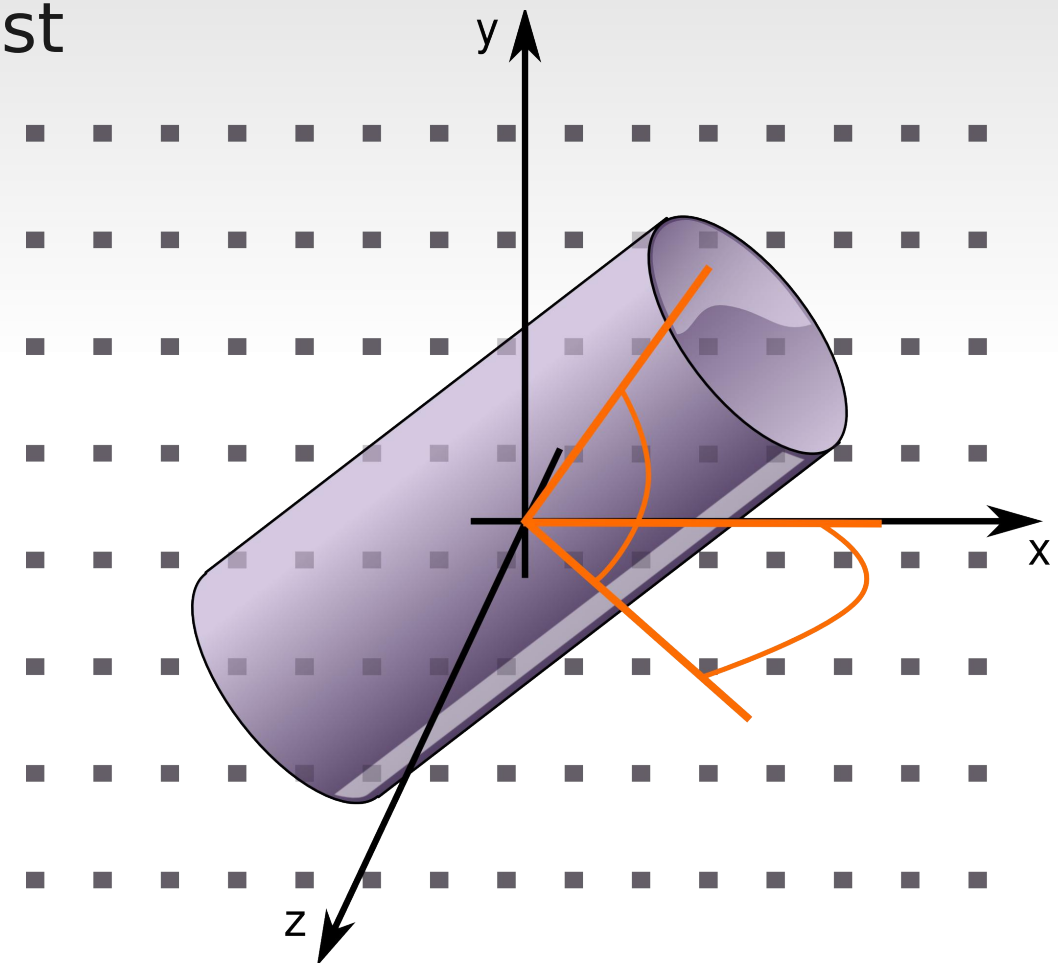
- Scenario: Looking at a new map, not built with automatic tools:
  - “I can see that there’s a helix here - build it for me!”
- From a given point:
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)
  - 1D Rotation search to find best fit
  - Score based on density at CB positions
  - Trim ‘n Grow

Centering the Rotation point

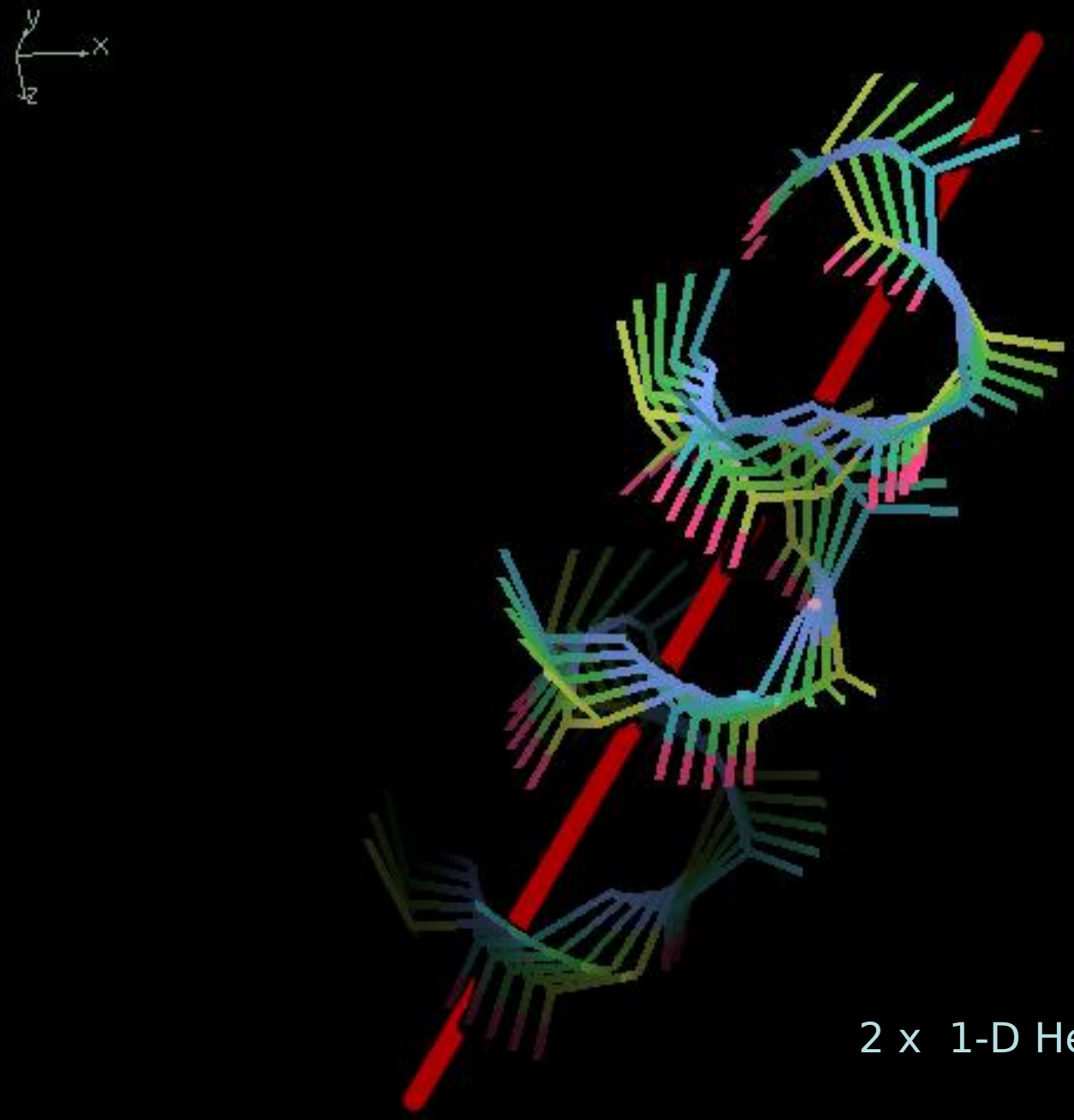


Cylinder Search

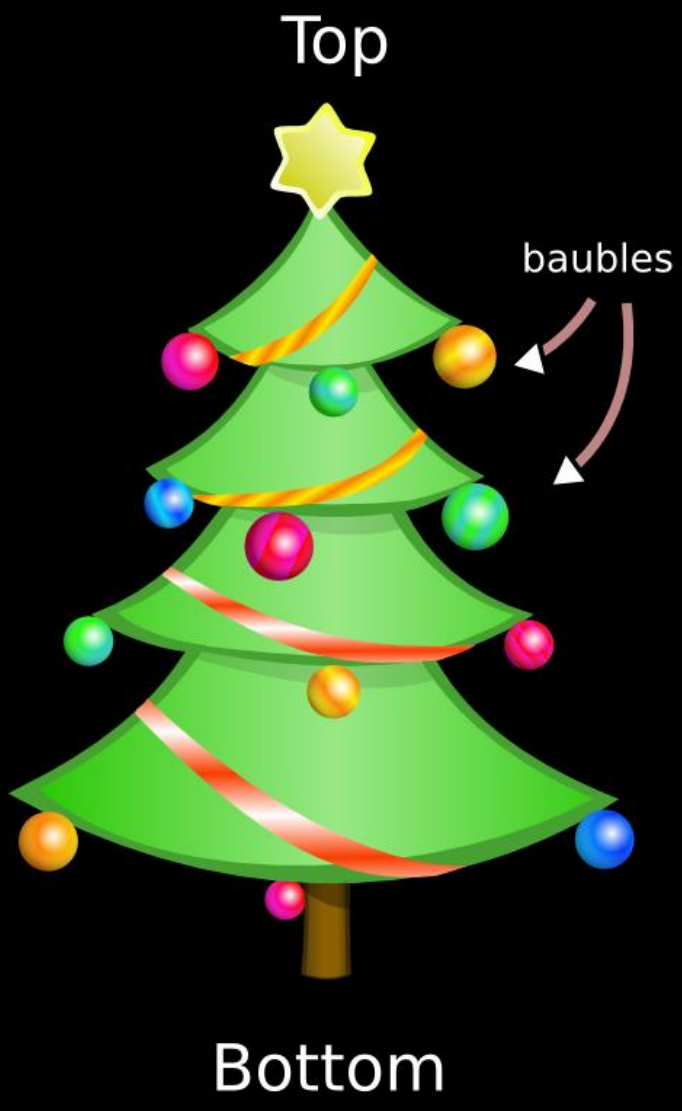
- Pick the orientation that encapsulates the most electron density

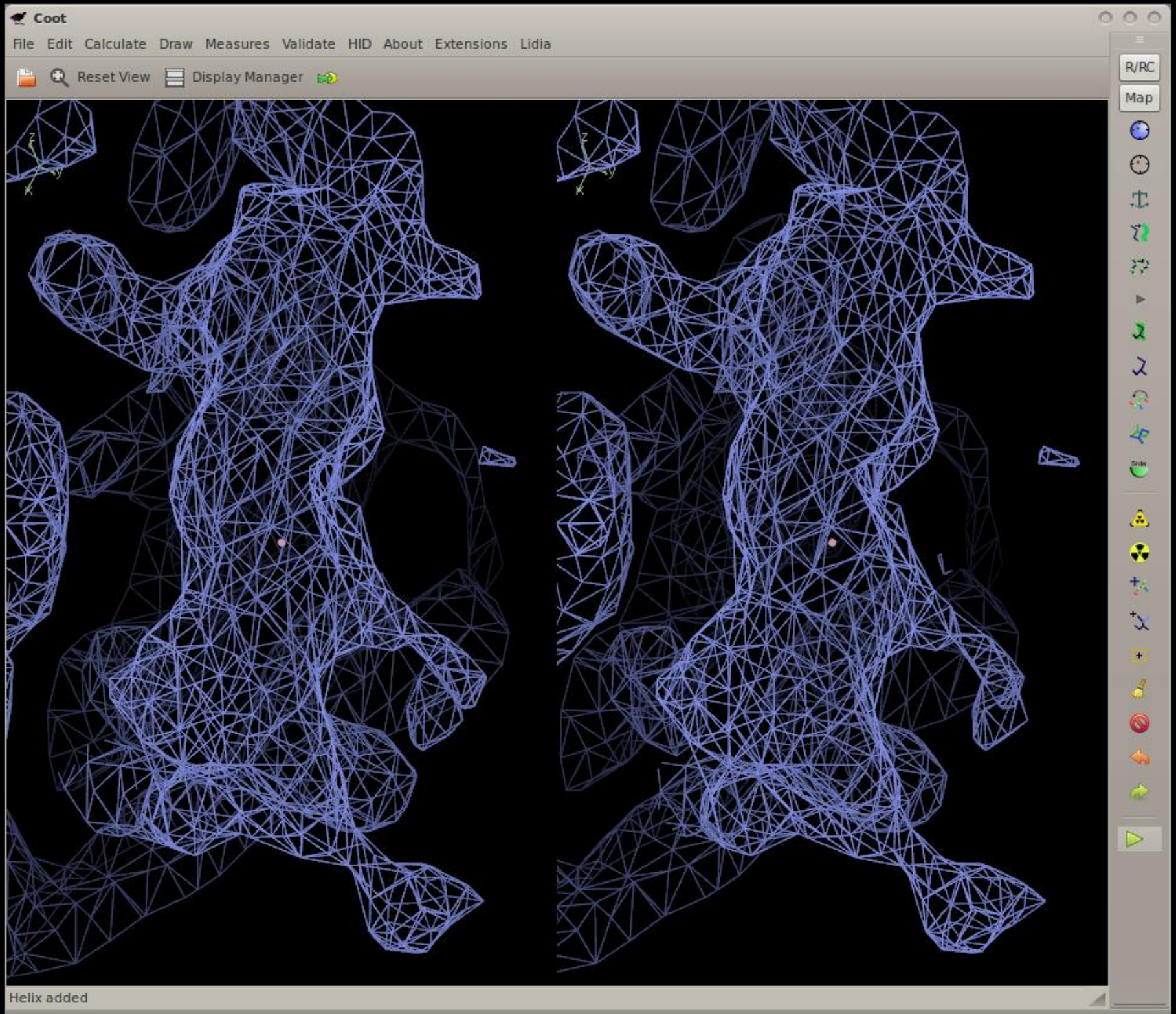


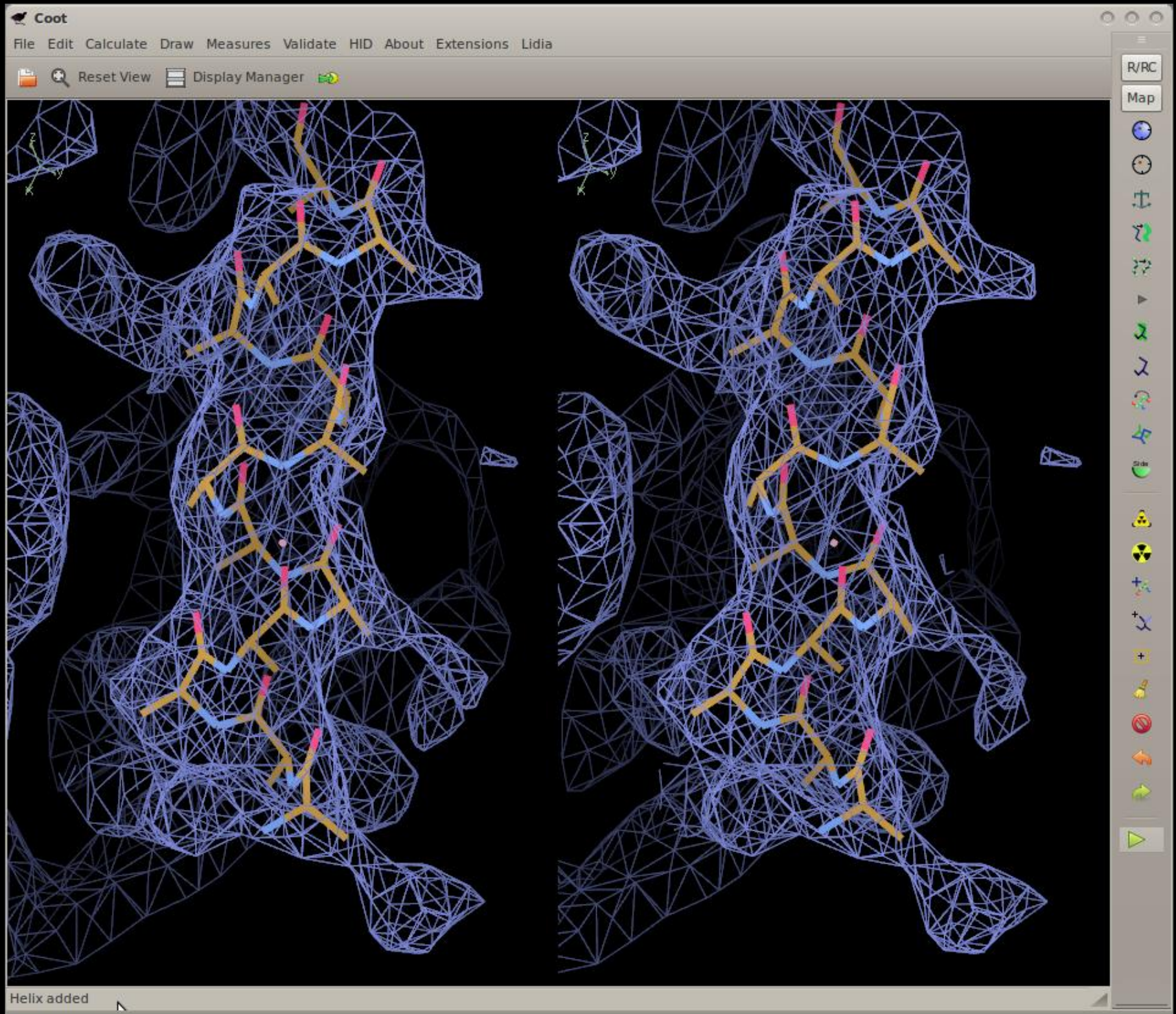
Using 2 rotation axes

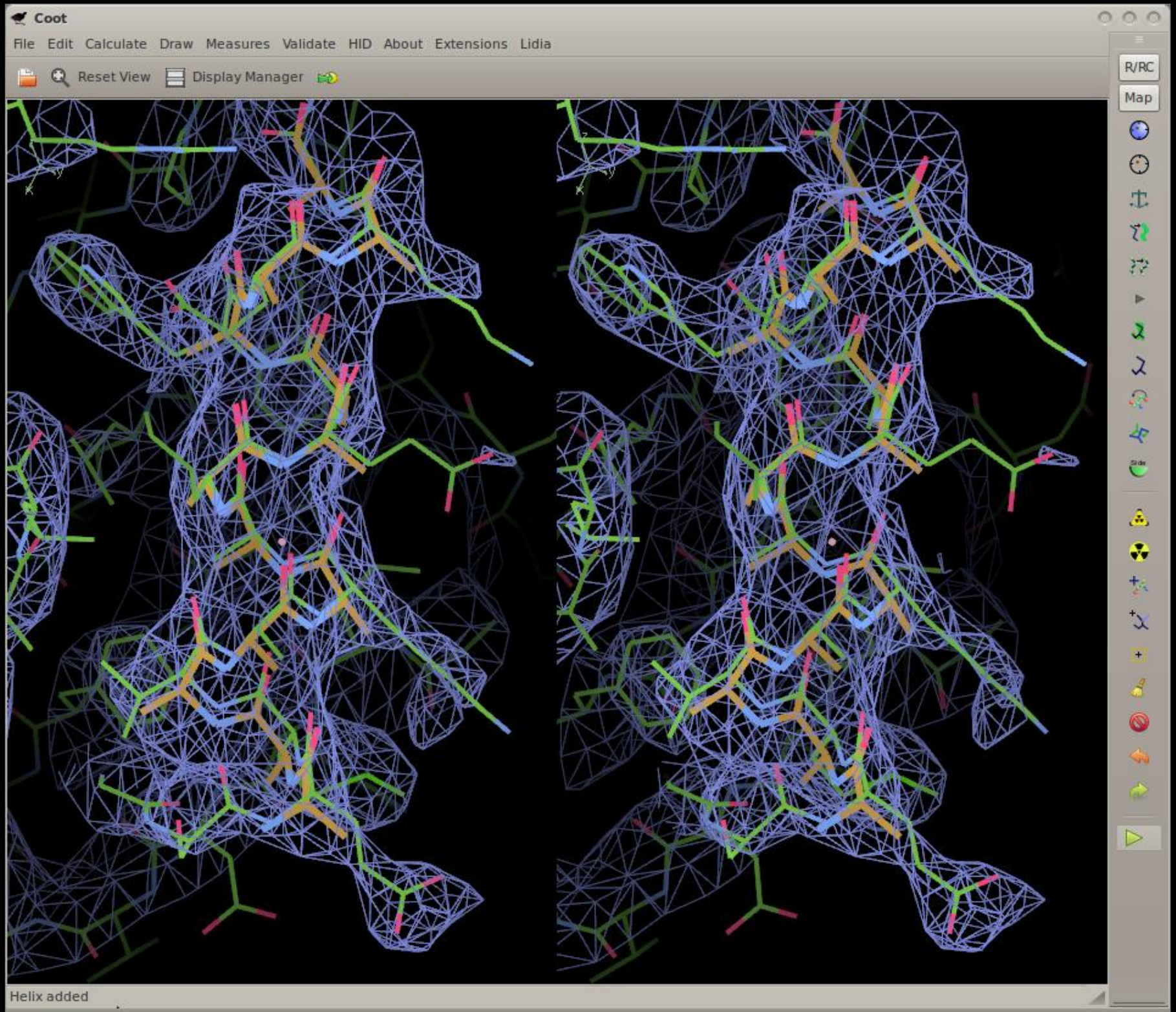


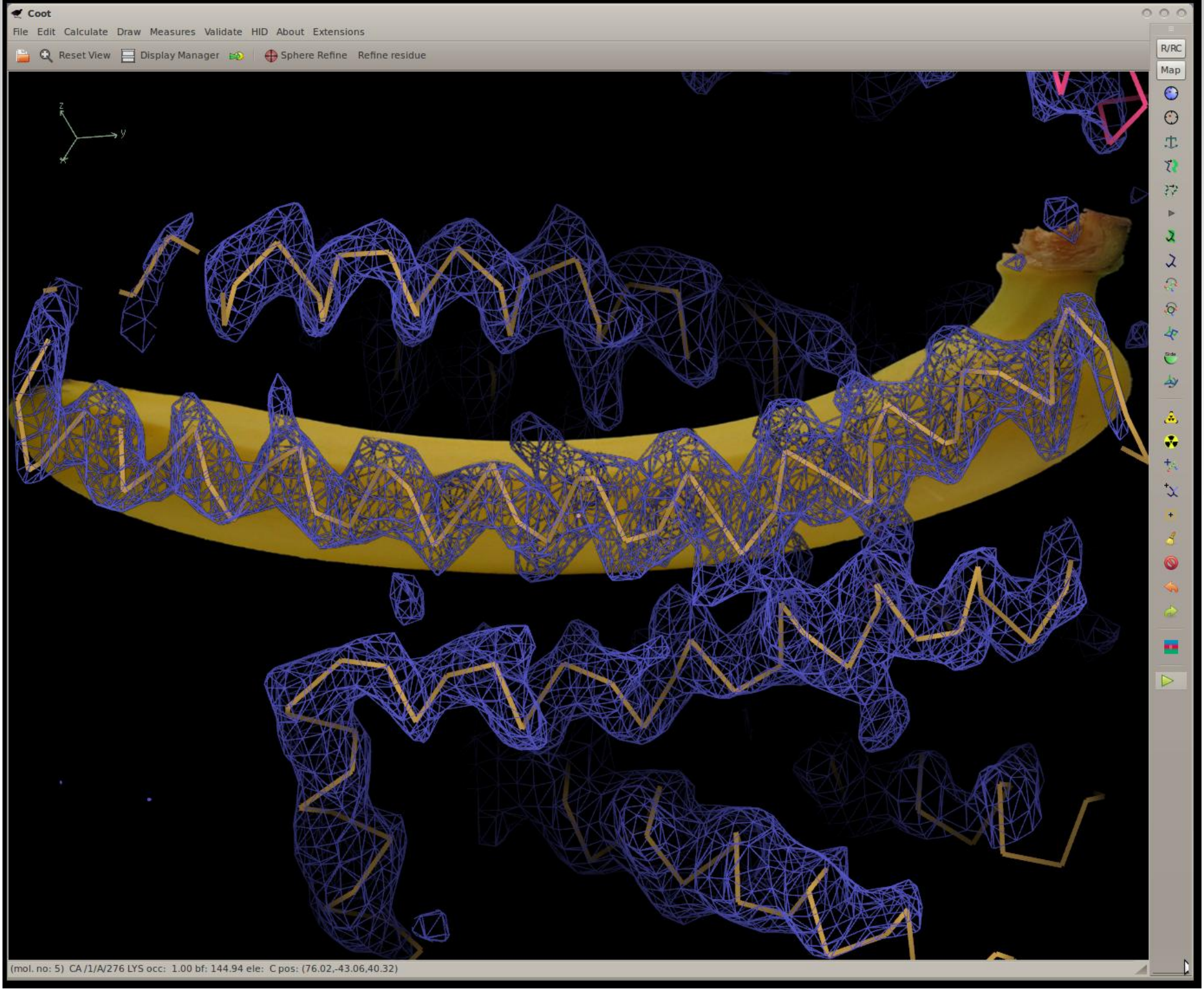
2 x 1-D Helix orientation searches


Top
baubles
Bottom

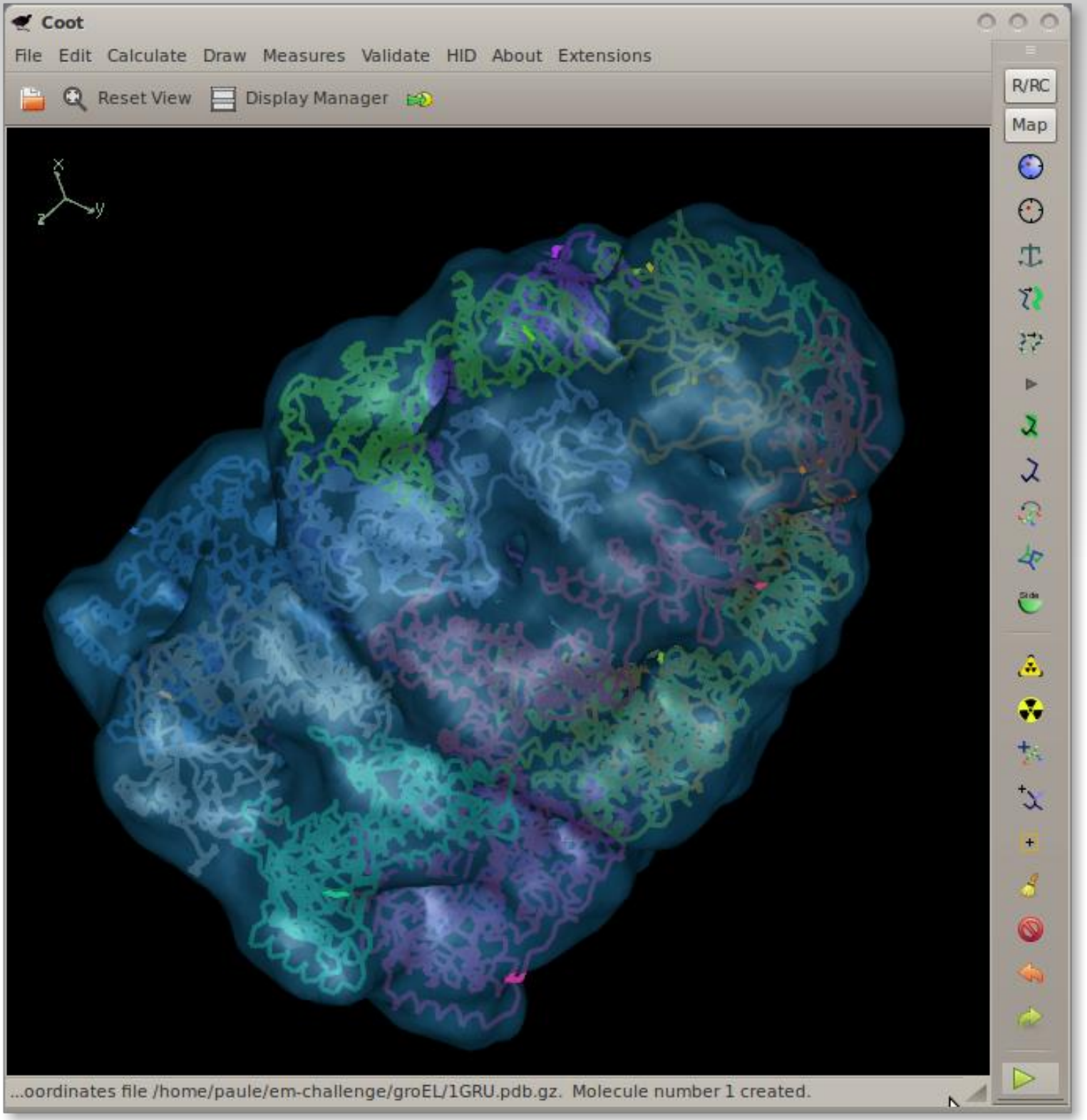

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added

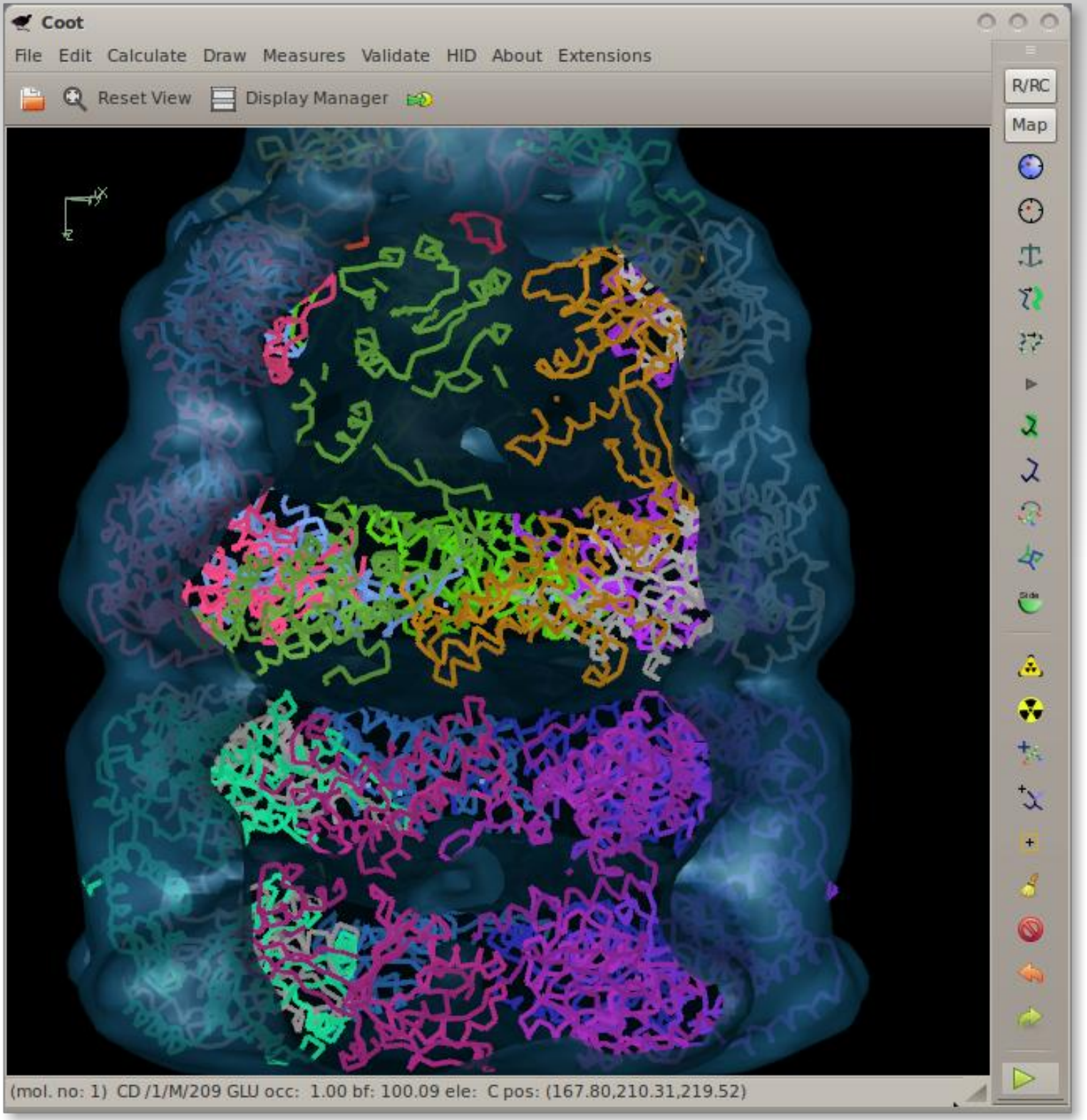

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added


Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
Sphere Refine
Refine residue
R/RC
Map
(mol. no: 5) CA /1/A/276 LYS occ: 1.00 bf: 144.94 ele: C pos: (76.02,-43.06,40.32)

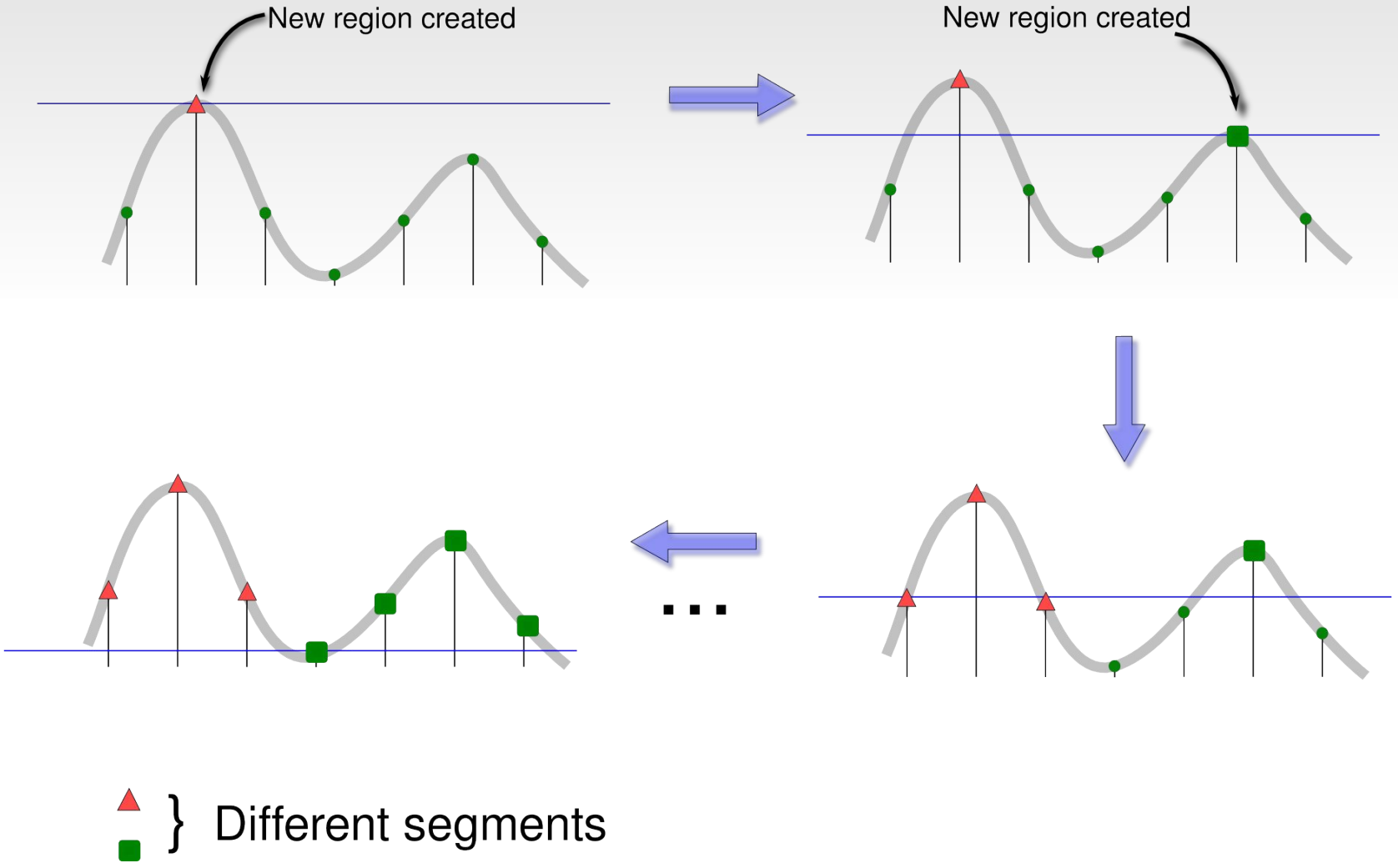
Handling EM maps

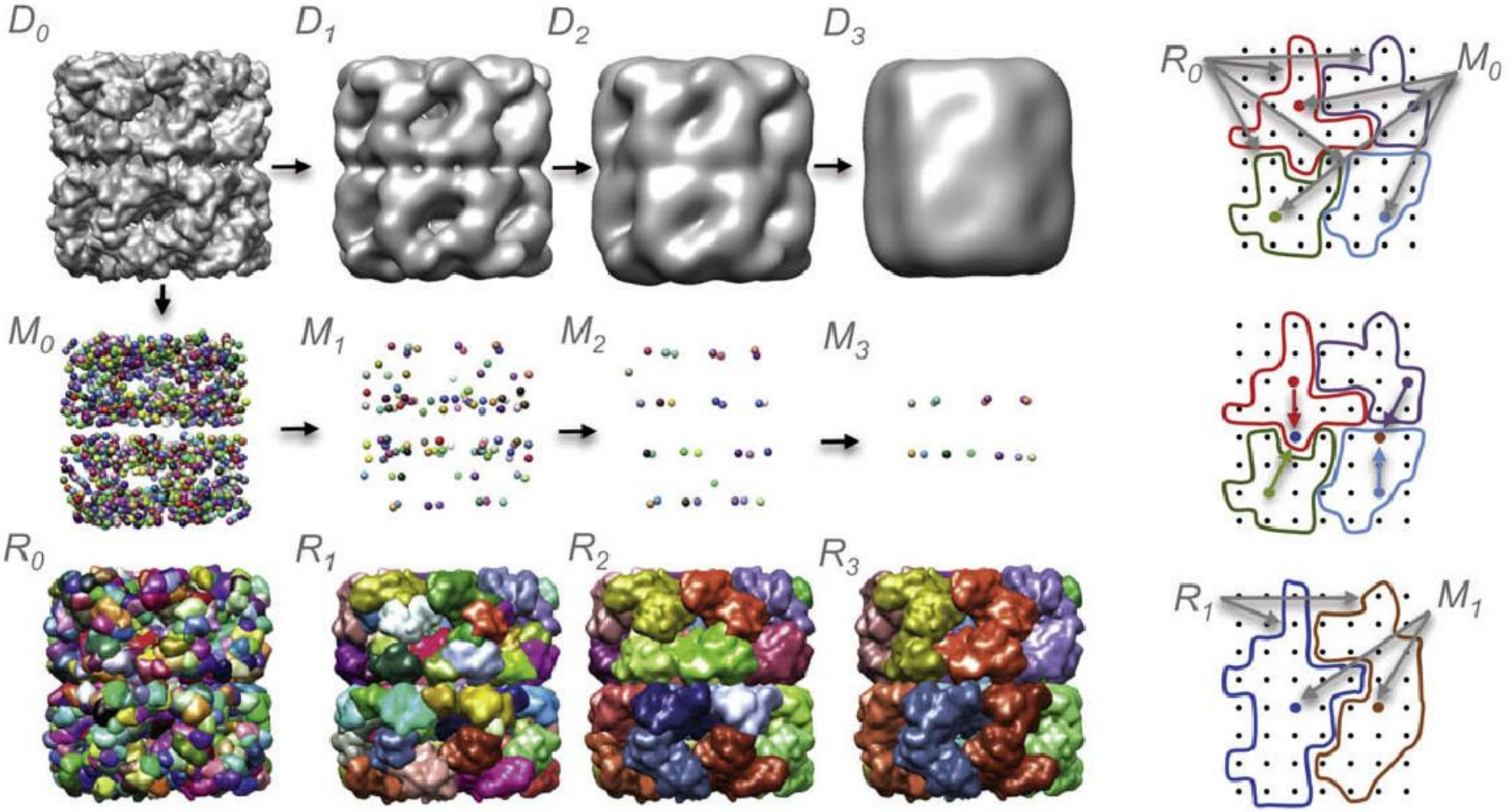

Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
...oordinates file /home/paule/em-challenge/groEL/1GRU.pdb.gz. Molecule number 1 created.


Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 1) CD /1/M/209 GLU occ: 1.00 bf: 100.09 ele: C pos: (167.80,210.31,219.52)

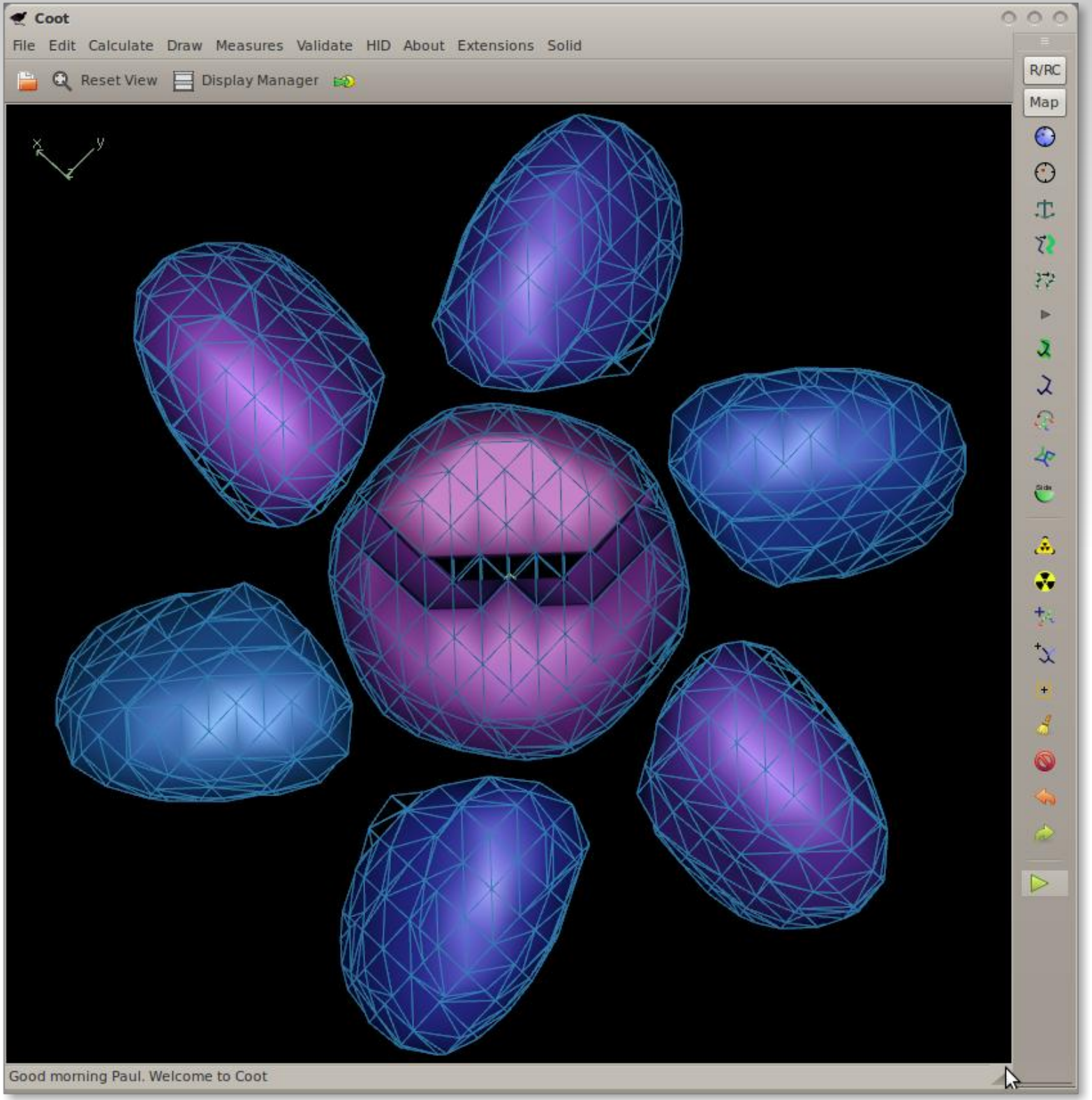
Partioning Maps: Watershed Algoritm

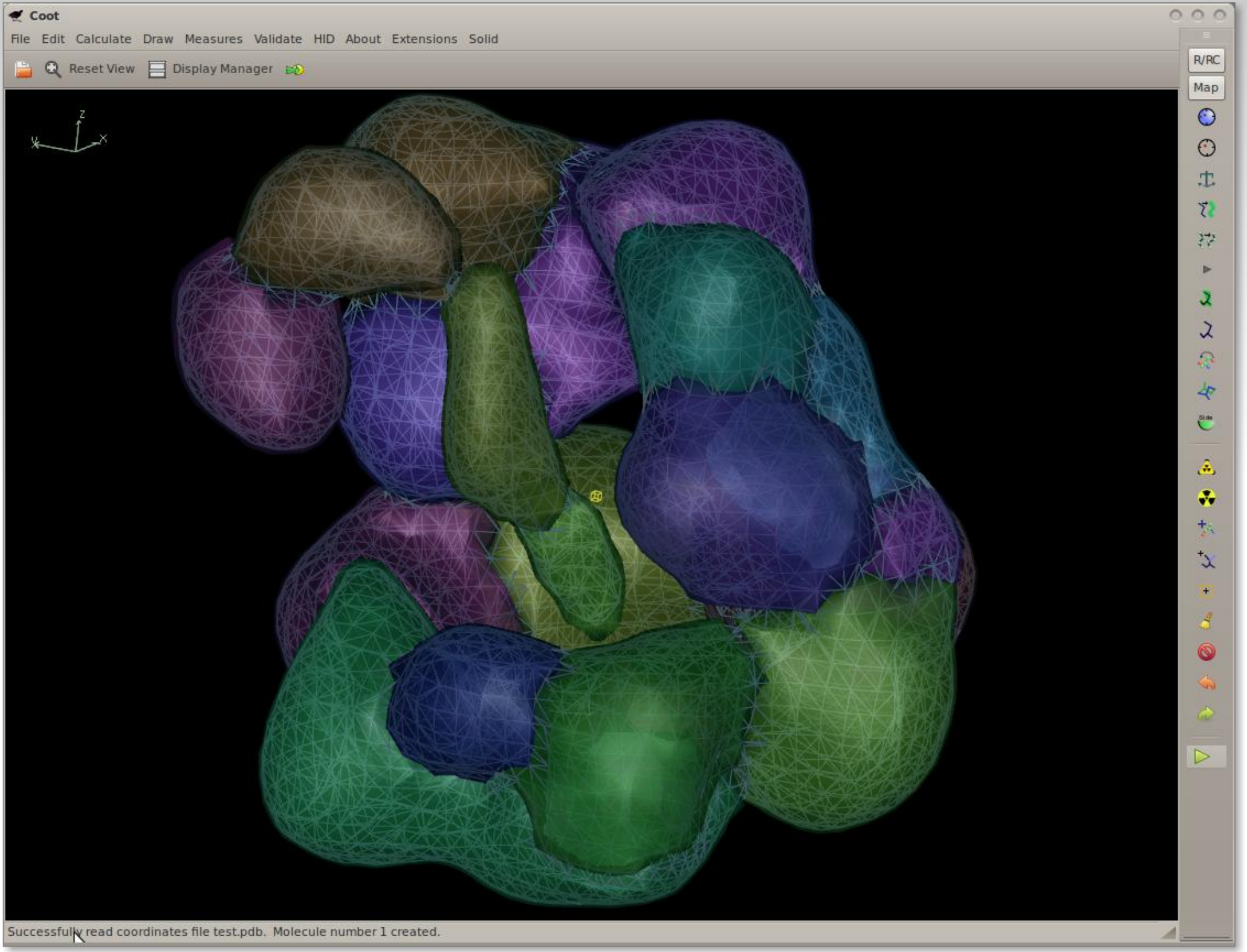
1D-analog

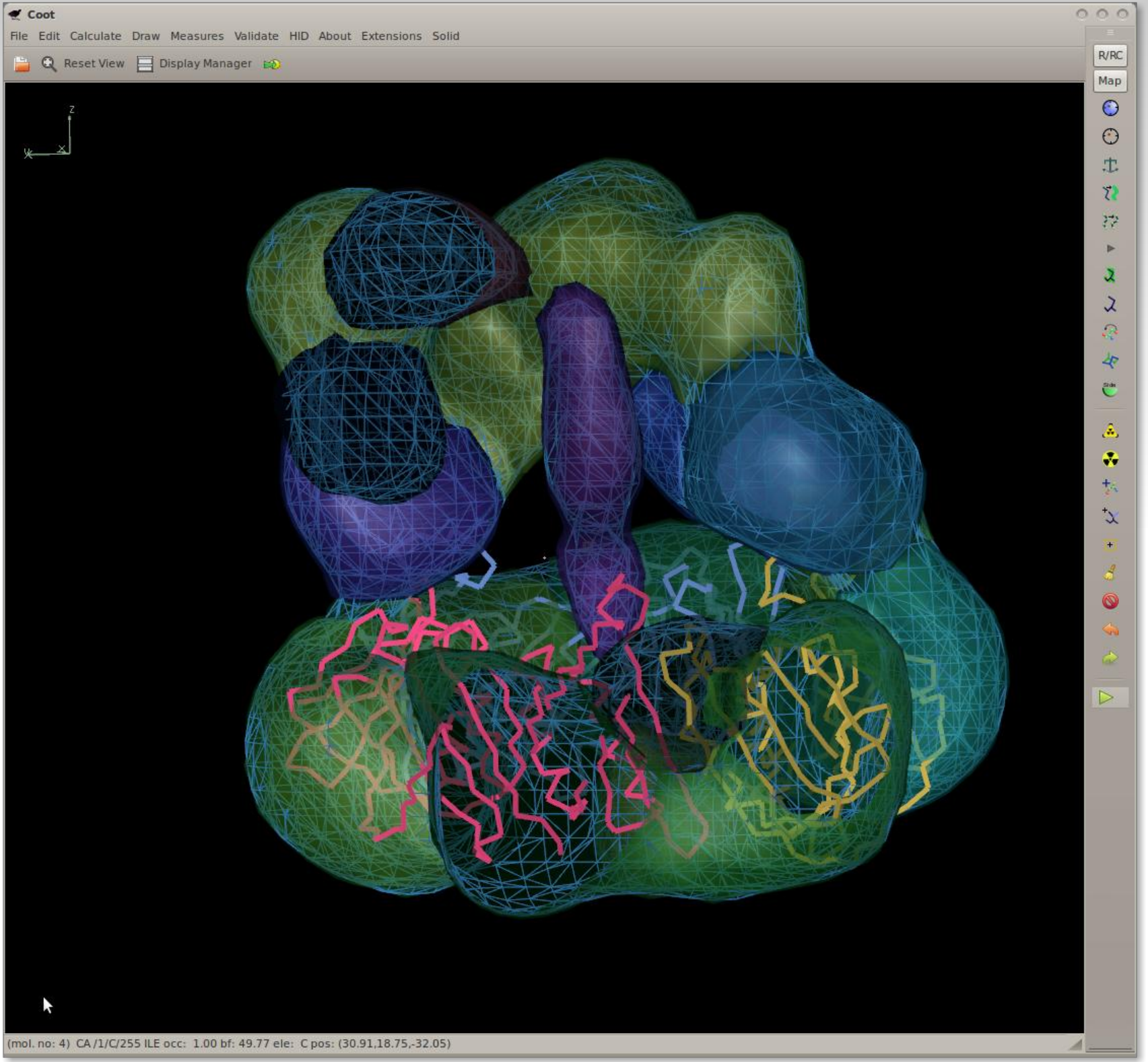




Pintilie *et al.* (2010)
J.Struct.Biol.

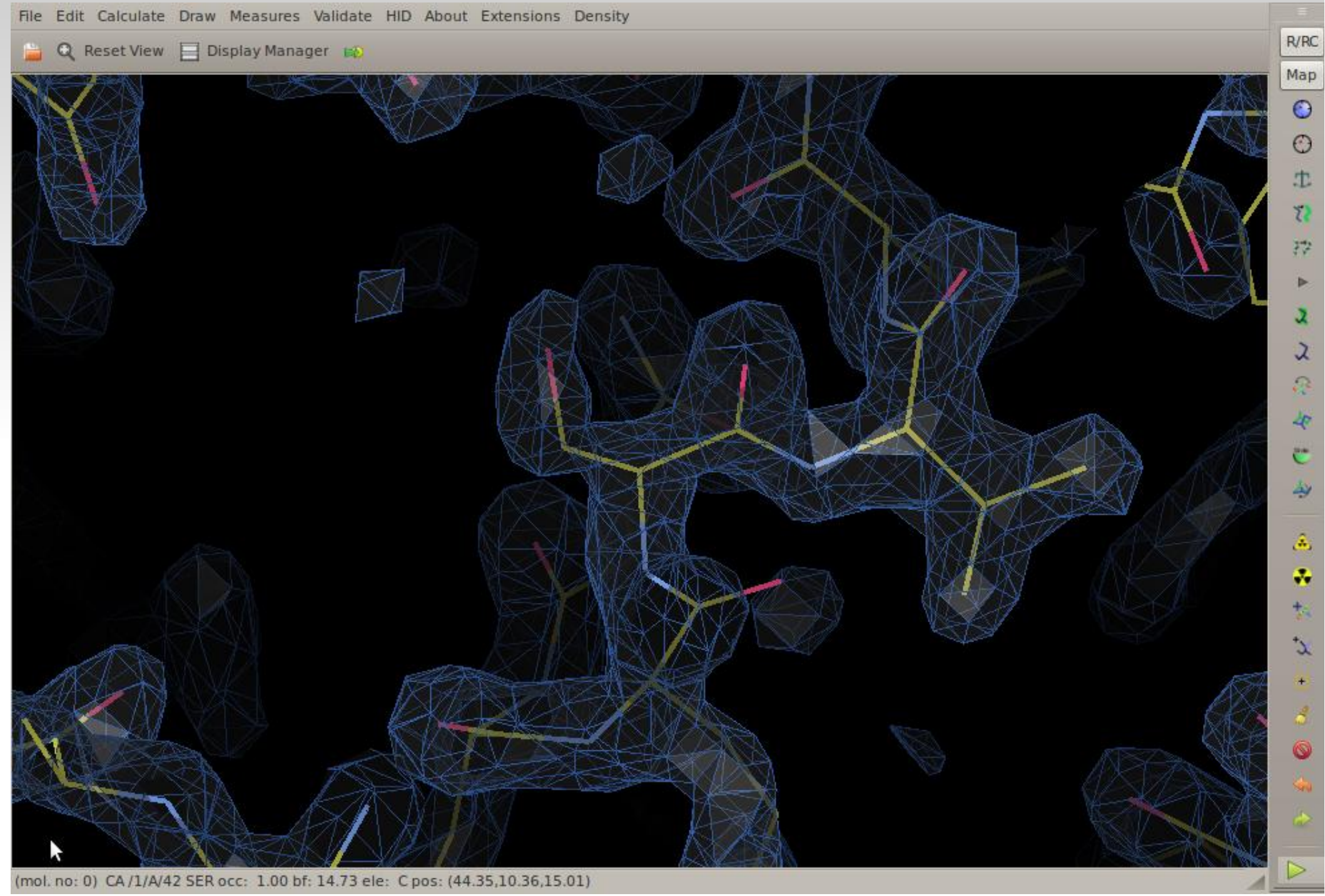

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Solid
Reset View
Display Manager
R/RC
Map
Good morning Paul. Welcome to Coot

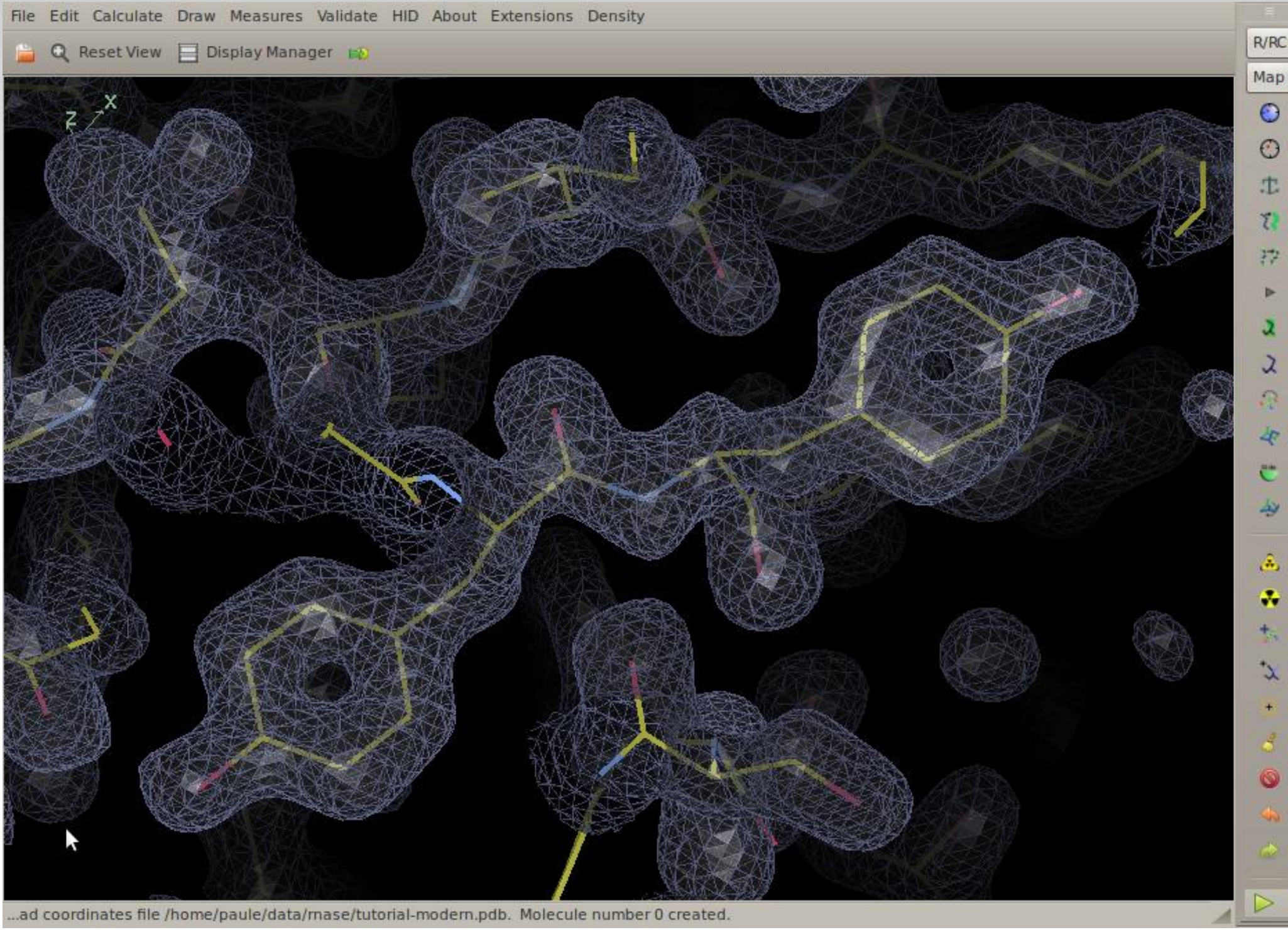

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Solid
Reset View
Display Manager
R/RC
Map
Z
Y
X
Successfully read coordinates file test.pdb. Molecule number 1 created.


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Solid
Reset View
Display Manager
R/RC
Map
(mol. no: 4) CA /1/C/255 ILE occ: 1.00 bf: 49.77 ele: C pos: (30.91,18.75,-32.05)

EM Futures

- Use Fast Fourier Feature recognition
  - Positioning protein domains or fragments in complete or segmented maps
  - Extremely parallelizable(?) and non-graphical
  - Map scoring with protein-protein docking hypotheses?
    - Score complexes with PISA?

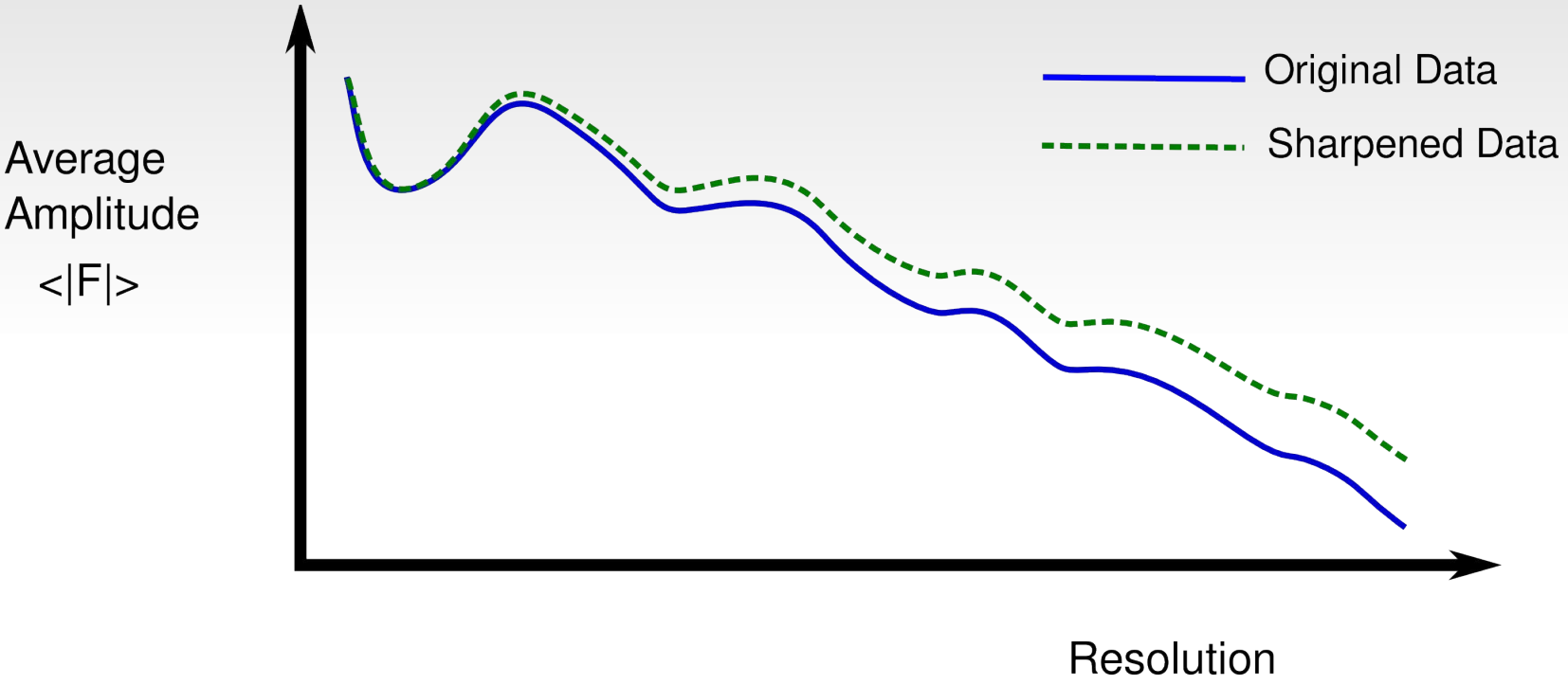

File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
(mol. no: 0) CA /1/A/42 SER occ: 1.00 bf: 14.73 ele: C pos: (44.35,10.36,15.01)


File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
Z
X
...ad coordinates file /home/paule/data/rnase/tutorial-modern.pdb. Molecule number 0 created.

Low Resolution Model-Building

- Interactive Map Sharpening
- User-define distance restraints
  - And torsion restraints
- Ramachandran Restraints
- “Backrub” rotamers
- Protein-db loop database (from Kevin Cowtan)

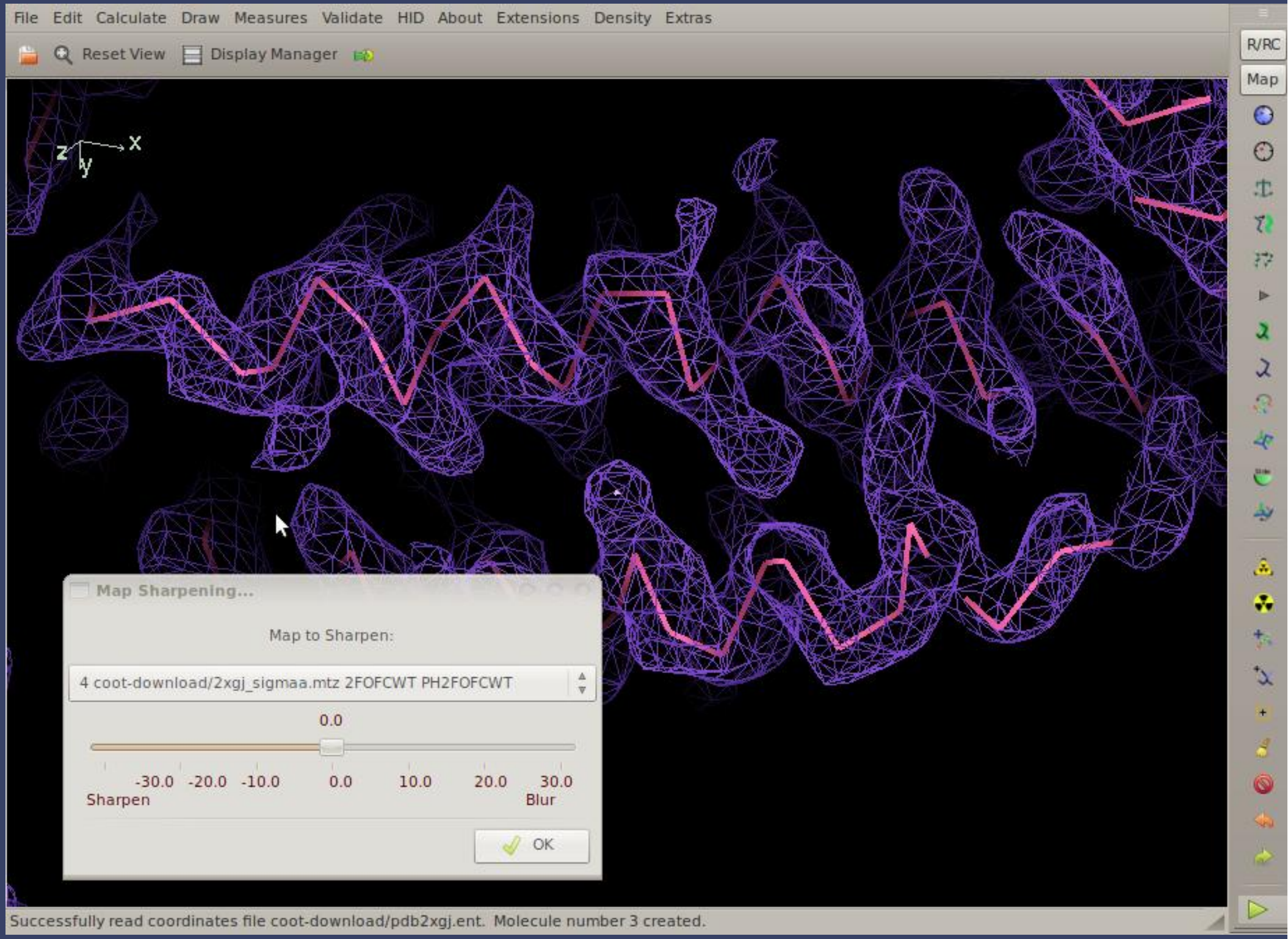
Sharpening the Data

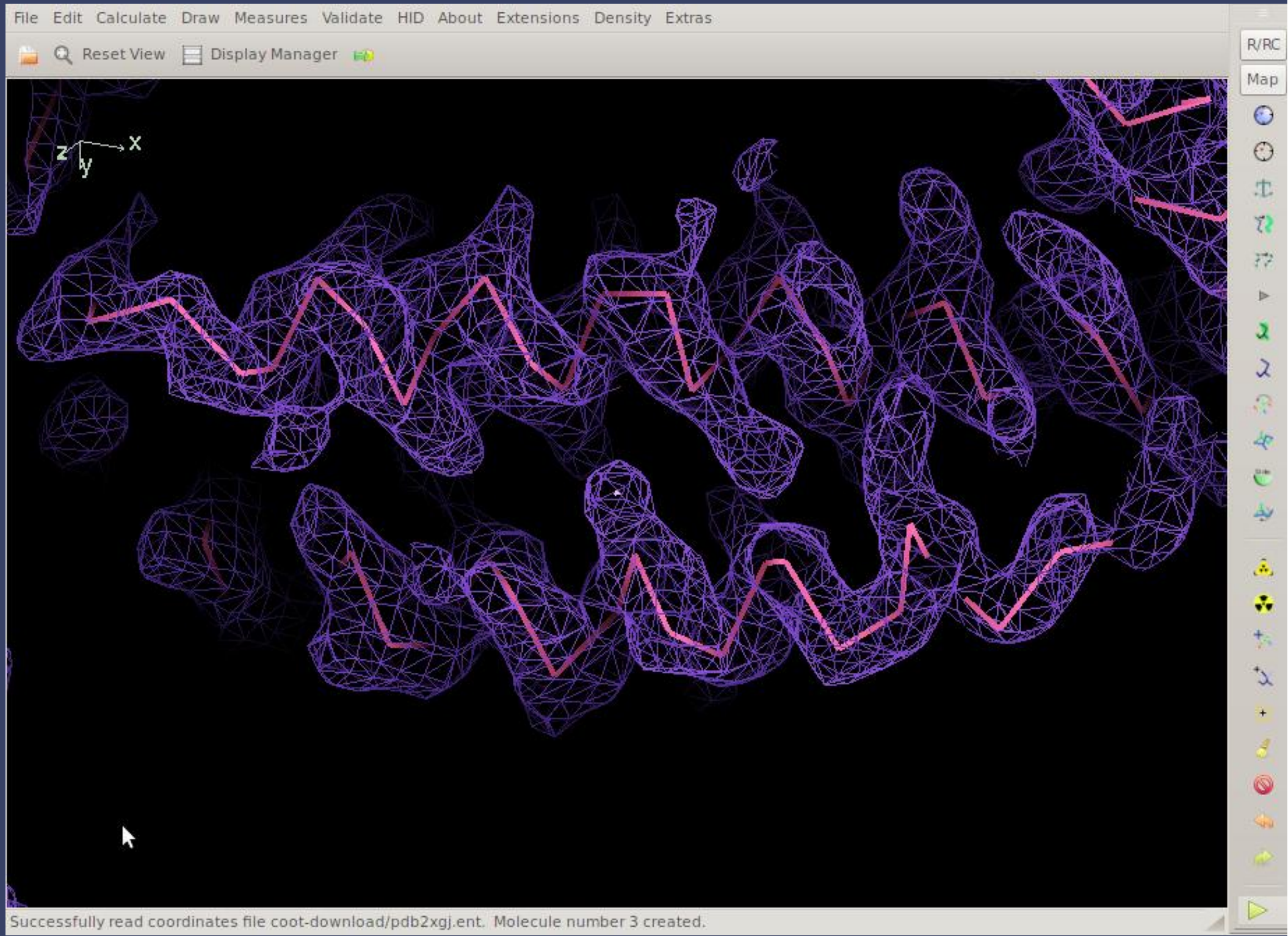


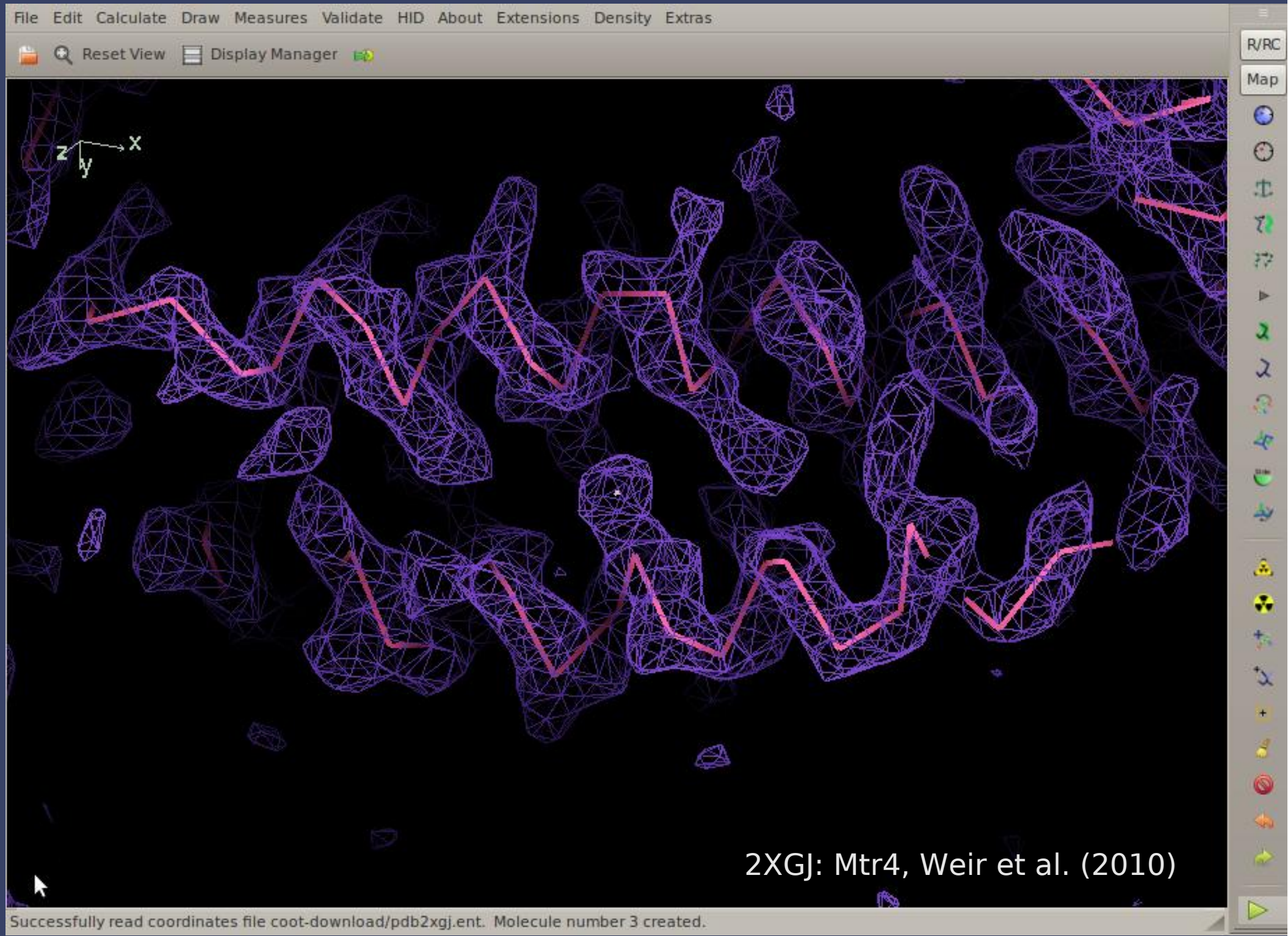
Sharpening the Data

Which B-factor shall I use to get the most interpretable map?

Interactively adjust the structure factor amplitudes and re-generate the map with FFT and recontouring...


File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
z
x
y
Map Sharpening...
Map to Sharpen:
4 coot-download/2xgj_sigmaa.mtz 2FOFCWT PH2FOFCWT
0.0
-30.0 -20.0 -10.0 0.0 10.0 20.0 30.0
Sharpen
Blur
OK
Successfully read coordinates file coot-download/pdb2xgj.ent. Molecule number 3 created.

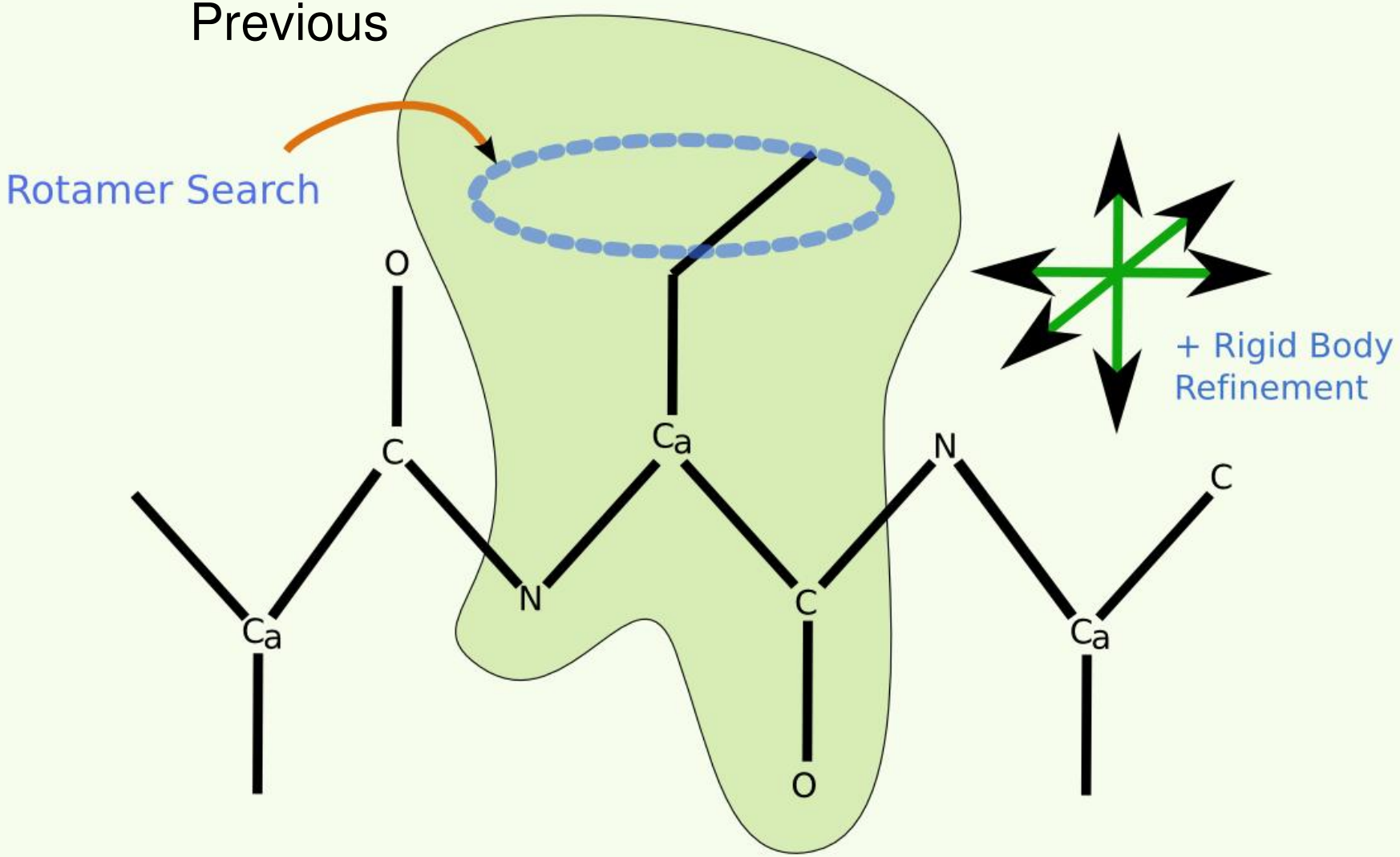

File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
z
x
y
Successfully read coordinates file coot-download/pdb2xgj.ent. Molecule number 3 created.

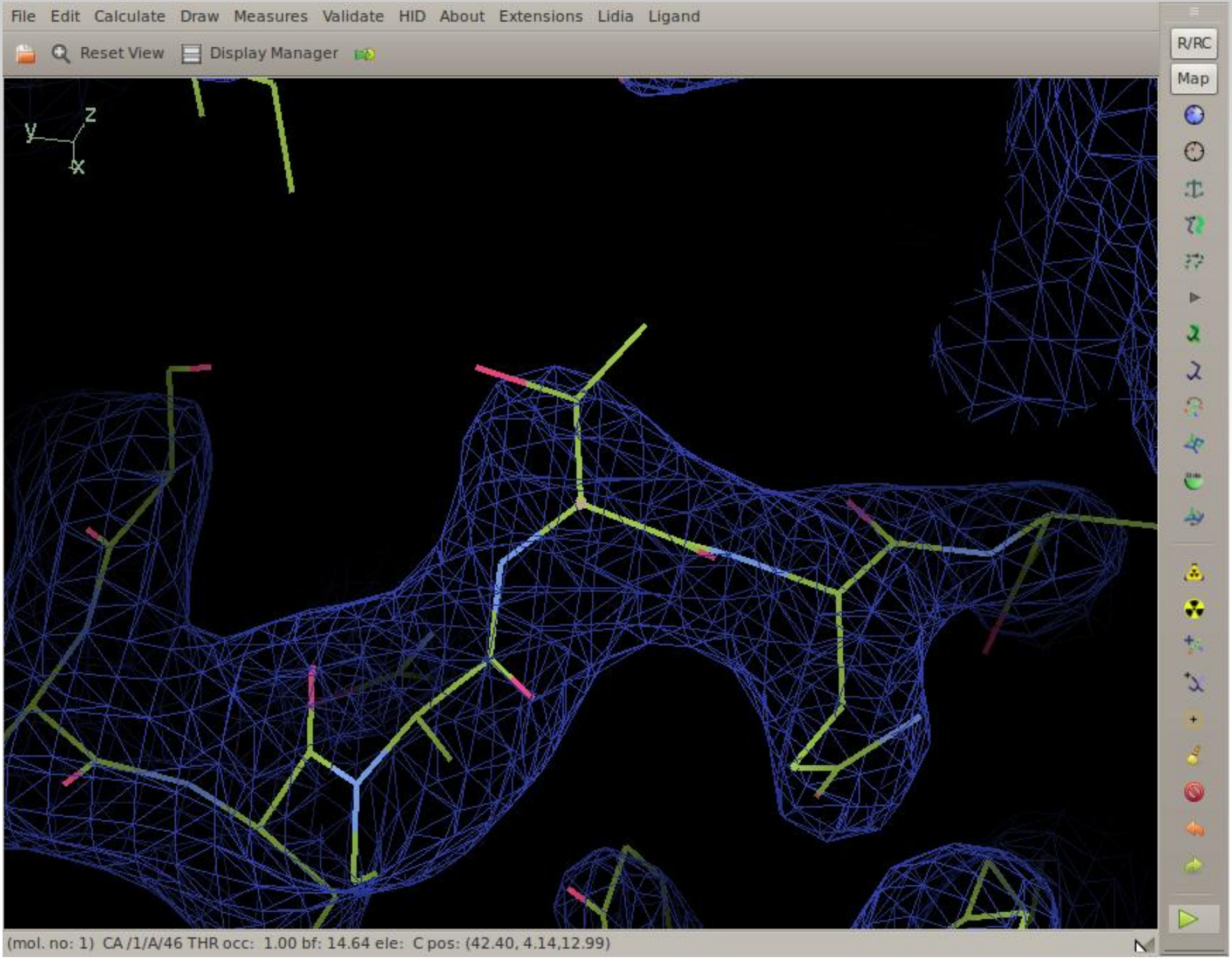

File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
z
x
y
2XGJ: Mtr4, Weir et al. (2010)
Successfully read coordinates file coot-download/pdb2xgj.ent. Molecule number 3 created.

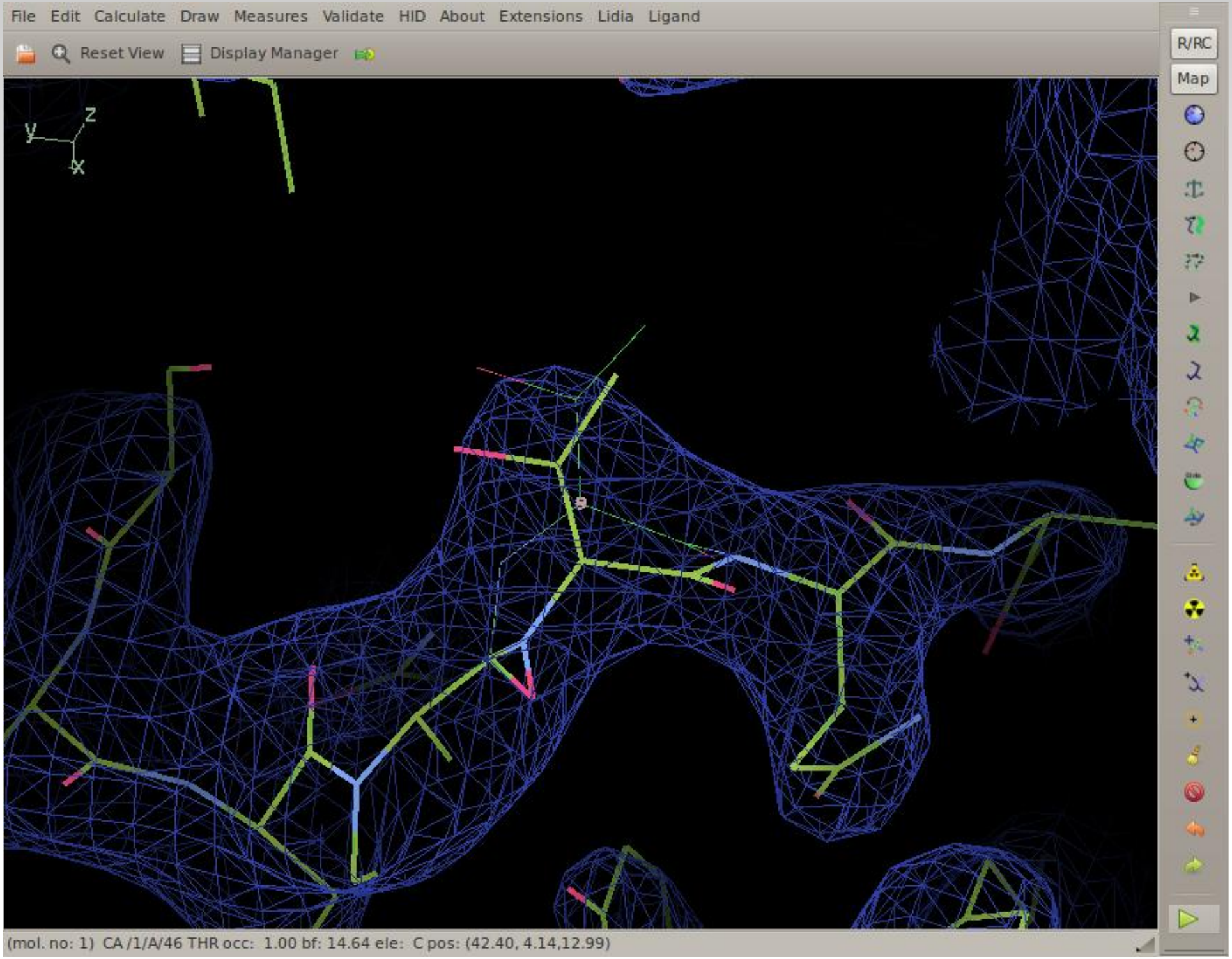
”Backrub Rotamers”

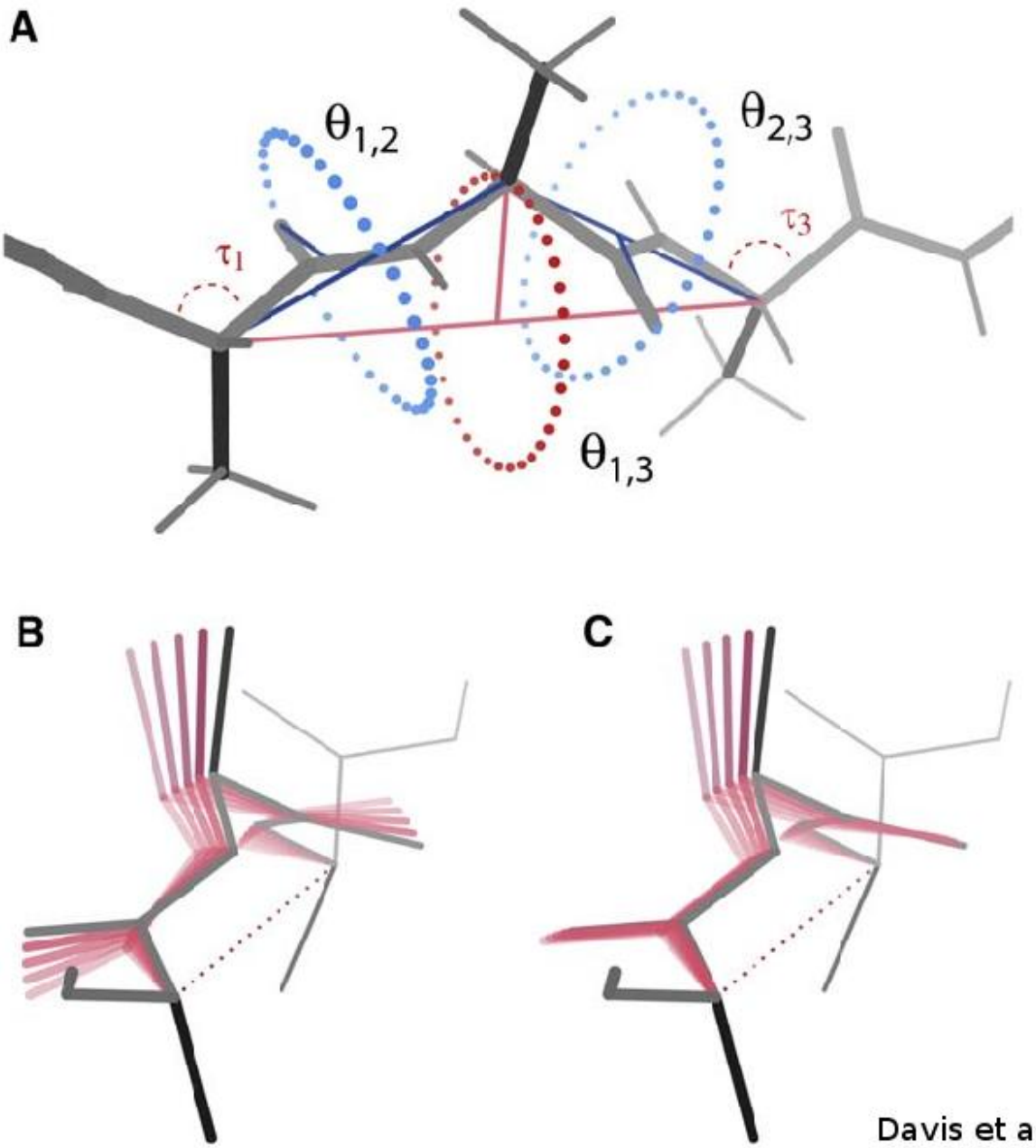
- High probability models with low resolution data

Current Low Resolution Rotamer Search



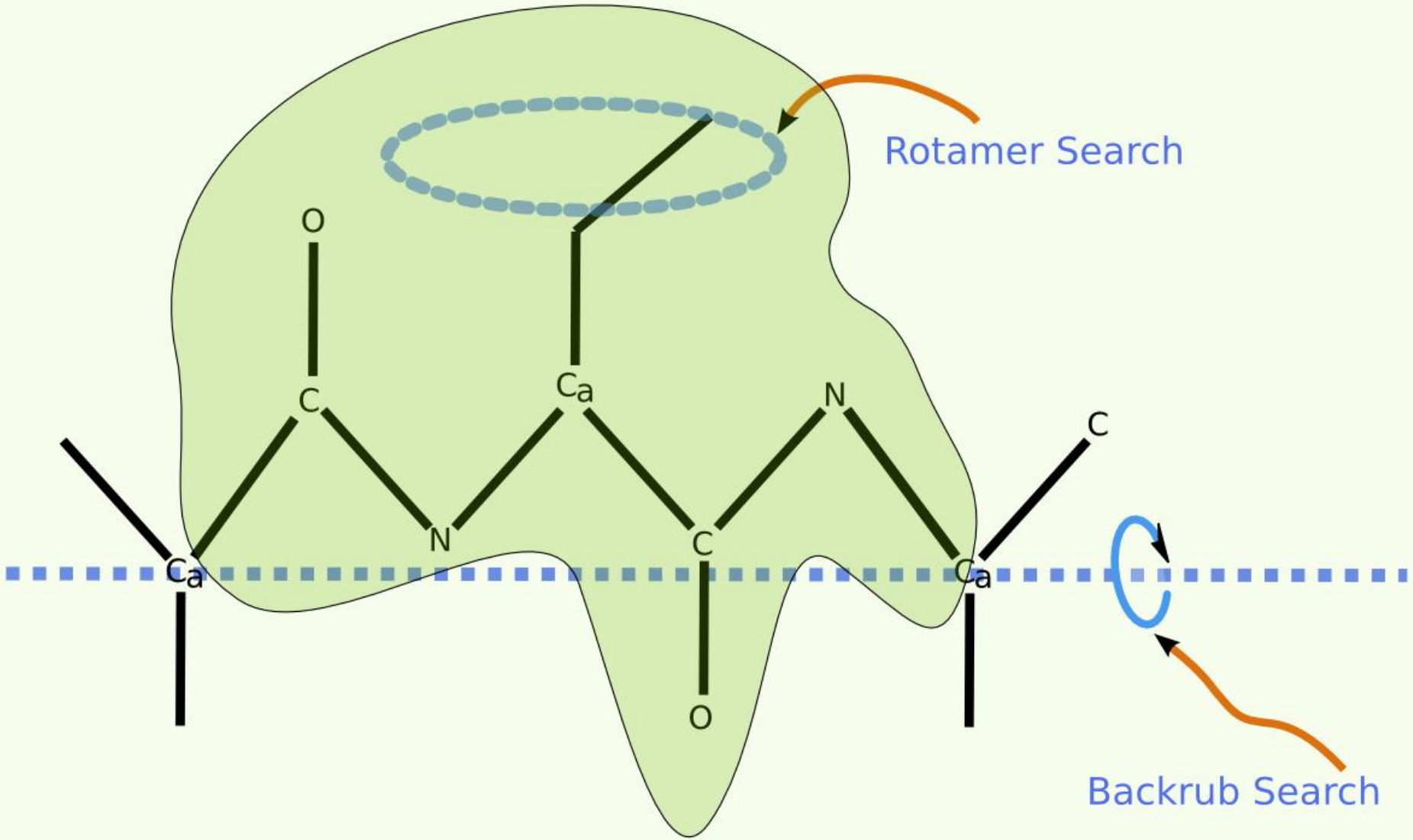

File Edit Calculate Draw Measures Validate HID About Extensions Lidia Ligand
Reset View
Display Manager
R/RC
Map
Z
Y
X
(mol. no: 1) CA /1/A/46 THR occ: 1.00 bf: 14.64 ele: C pos: (42.40, 4.14,12.99)


File Edit Calculate Draw Measures Validate HID About Extensions Lidia Ligand
Reset View
Display Manager
R/RC
Map
Z
Y
X
(mol. no: 1) CA /1/A/46 THR occ: 1.00 bf: 14.64 ele: C pos: (42.40, 4.14,12.99)

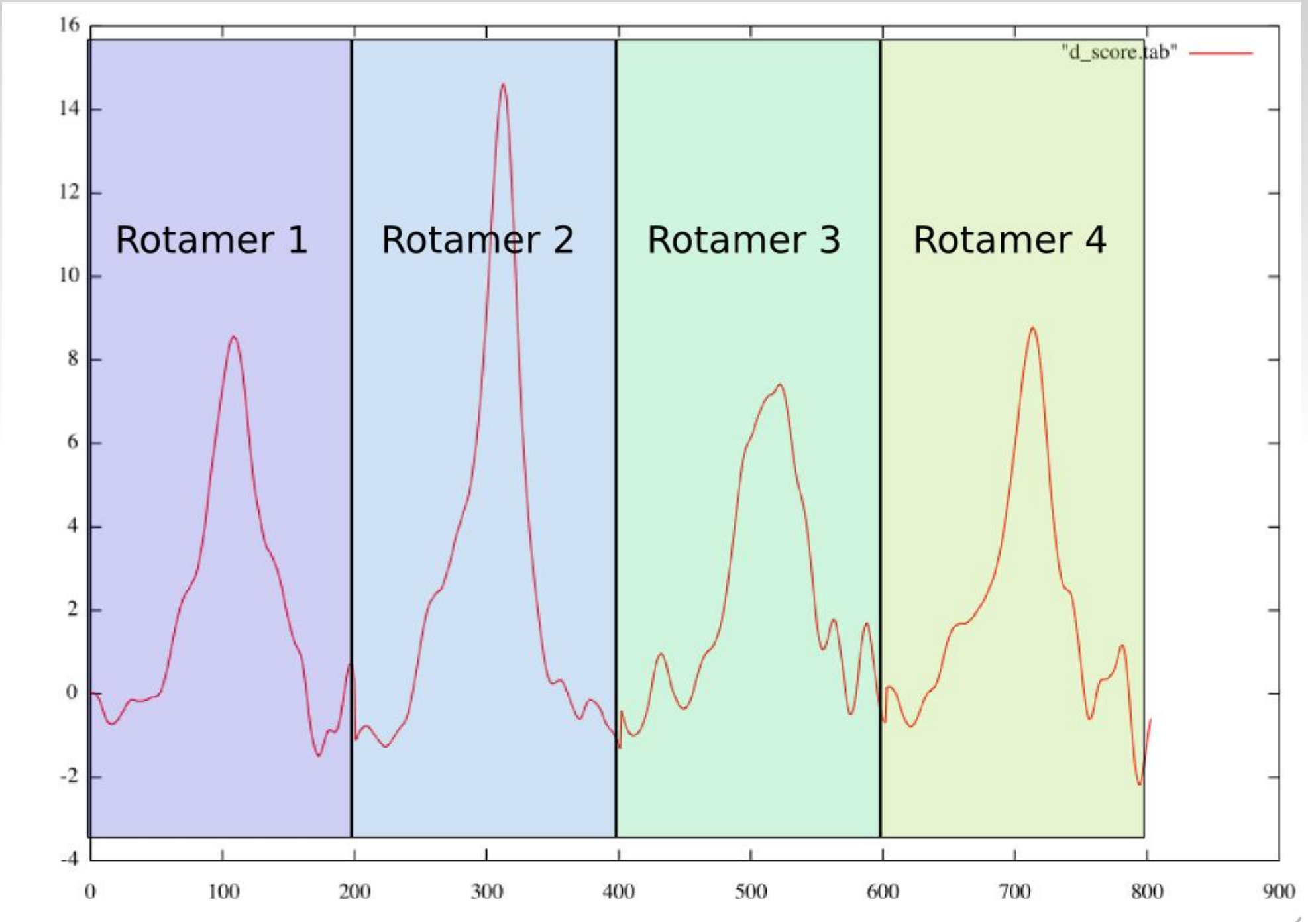


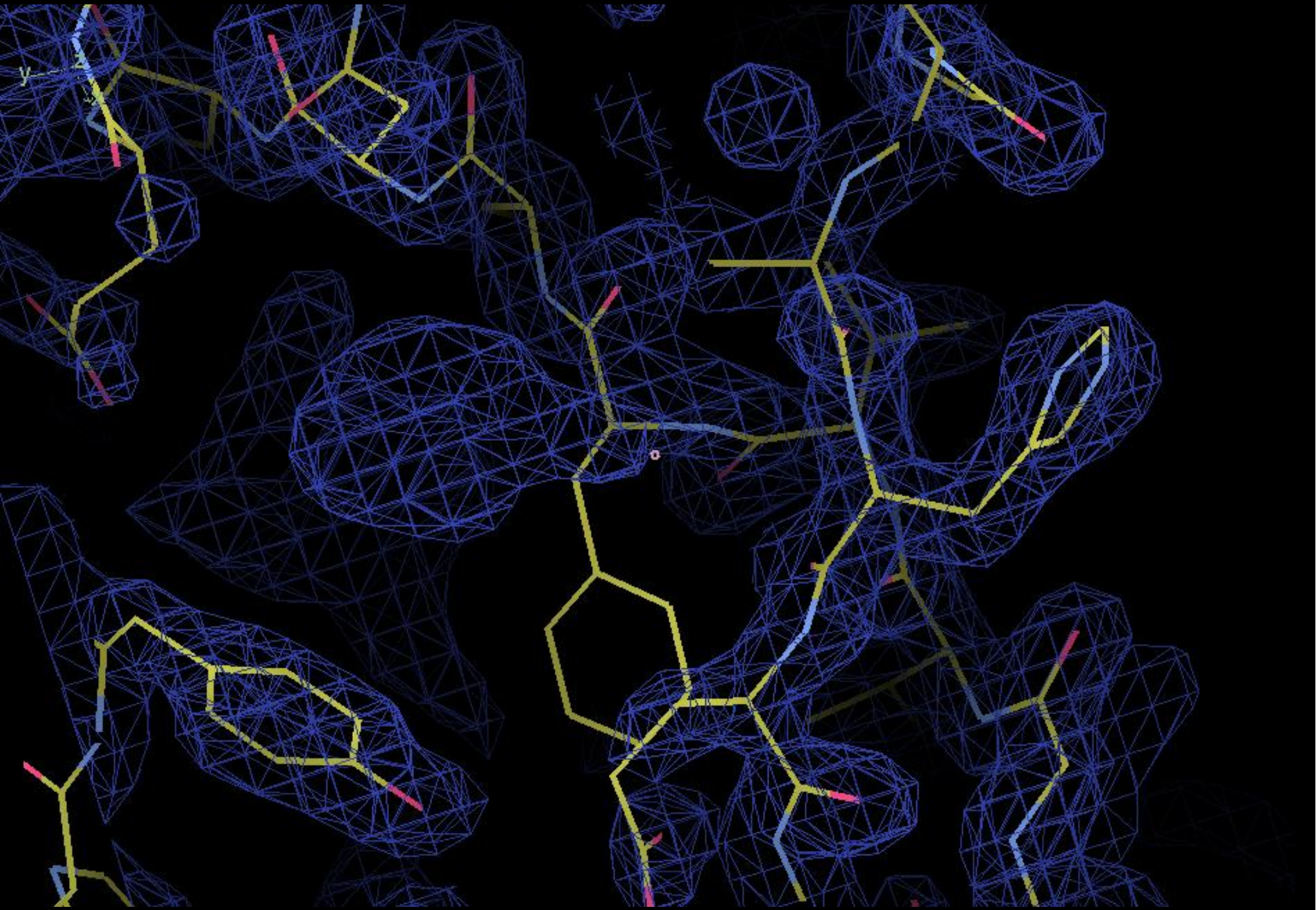
Davis et al. (2006) Structure

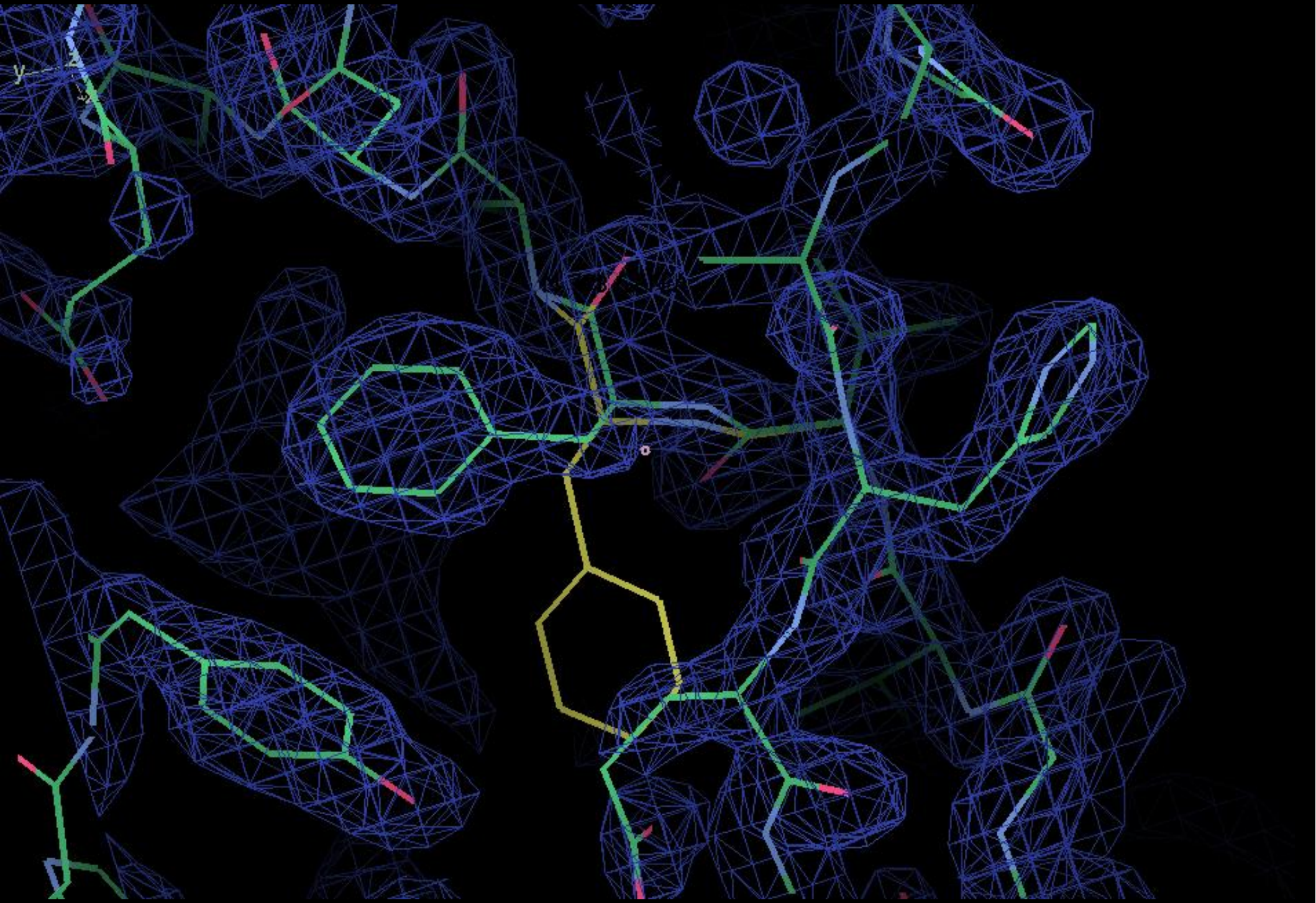
New Low Resolution Rotamer Search

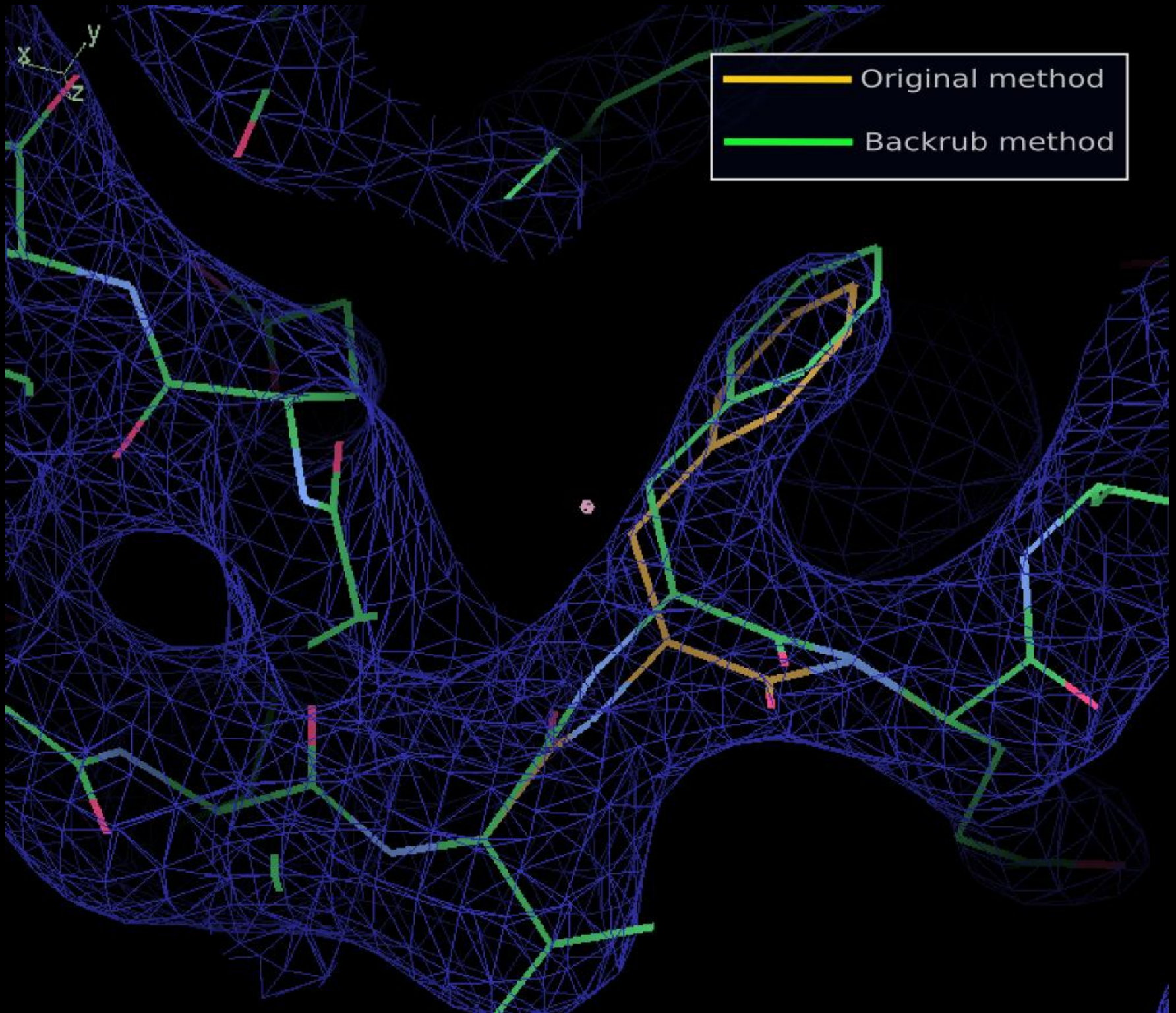


After Fitting Tools in KING/Molprobity


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






x
y
z
Original method
Backrub method

To turn it on…

- (ROTAMERSEARCHLOWRES)

To turn it on…

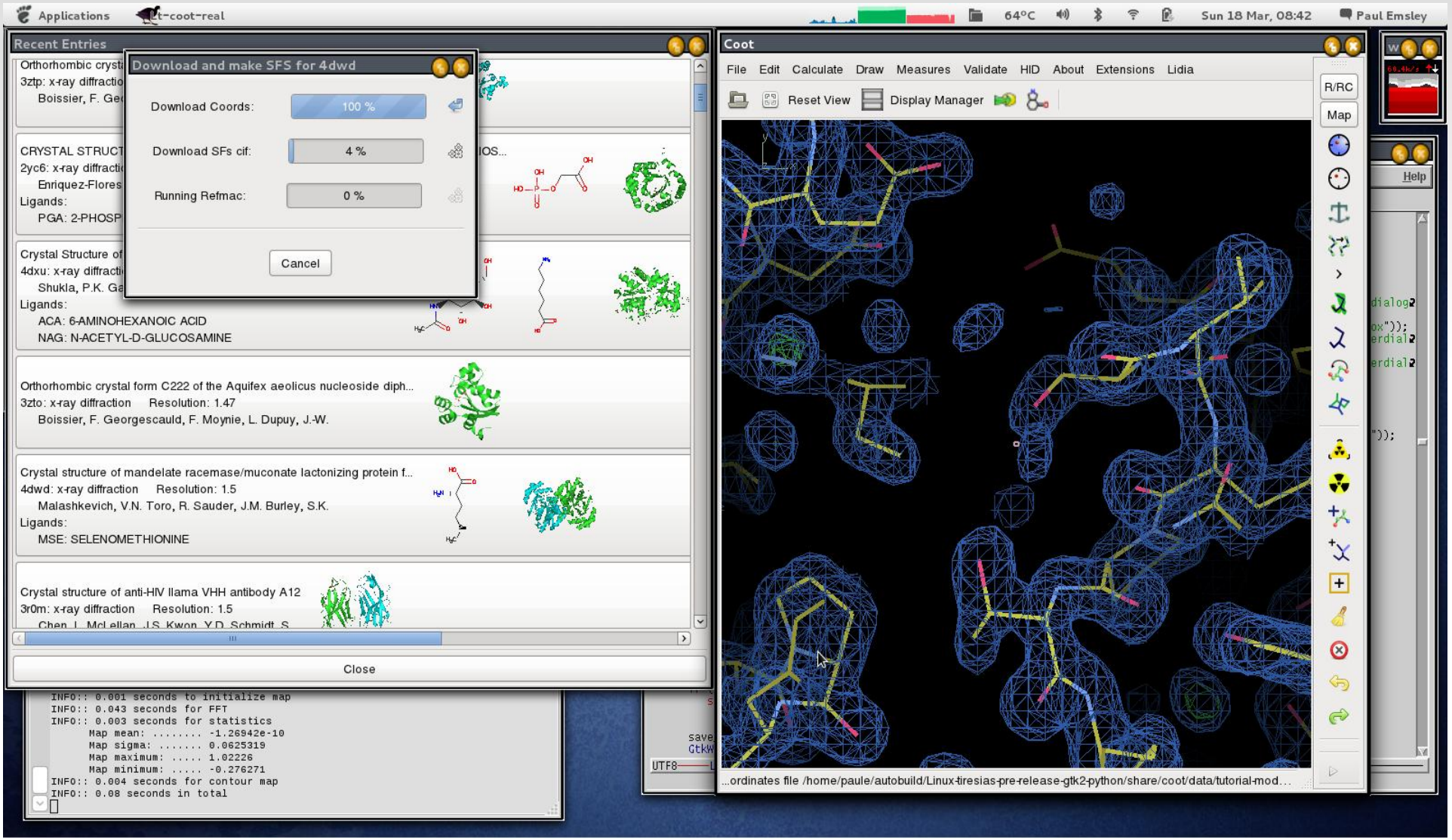
- (ROTAMERSEARCHLOWRES)

Networking...

- PDBe interface...
- Drag and drop
  - Also with drugbank

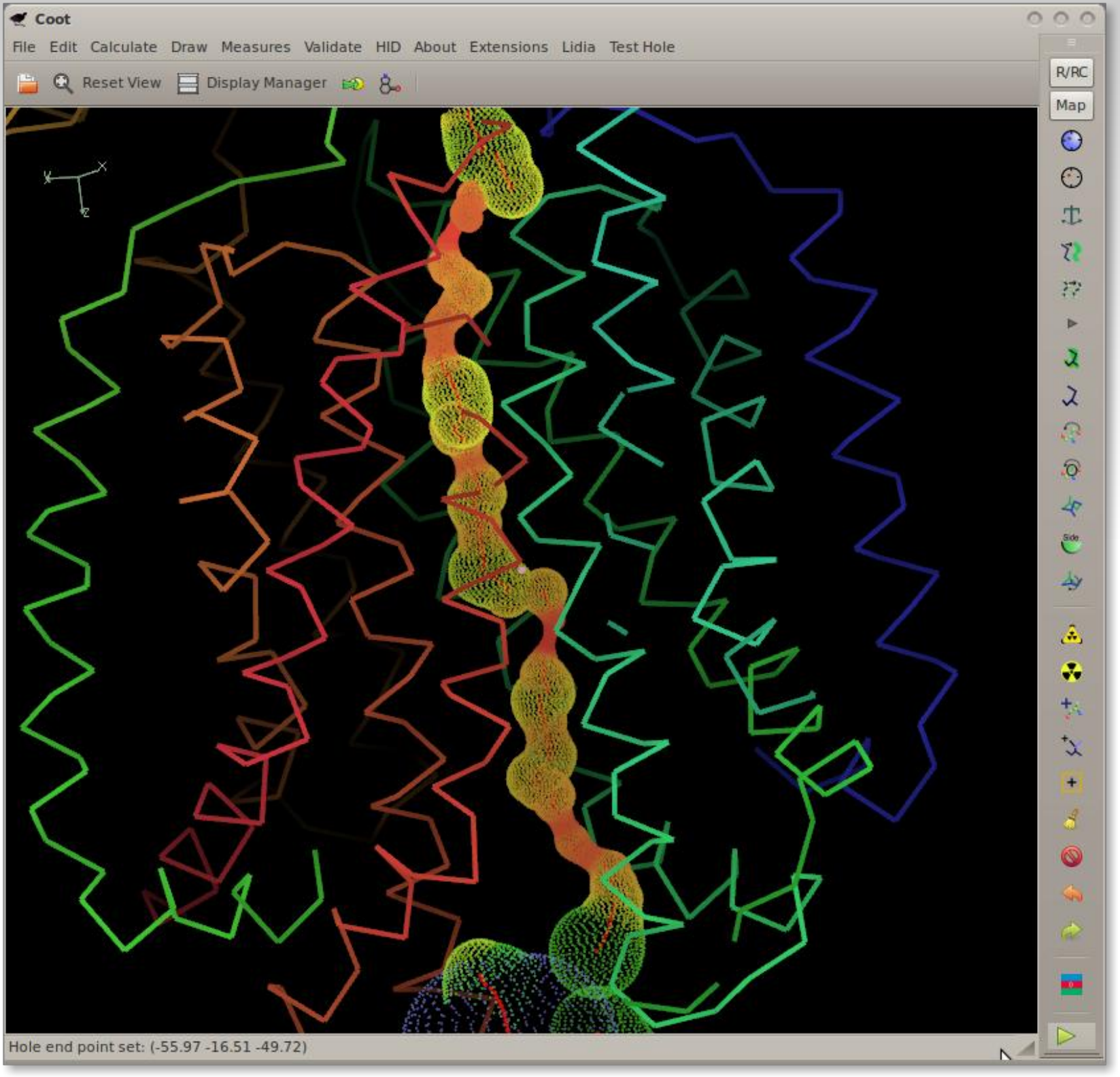
PDBe Recent Structures

JSON parser, network threaded code



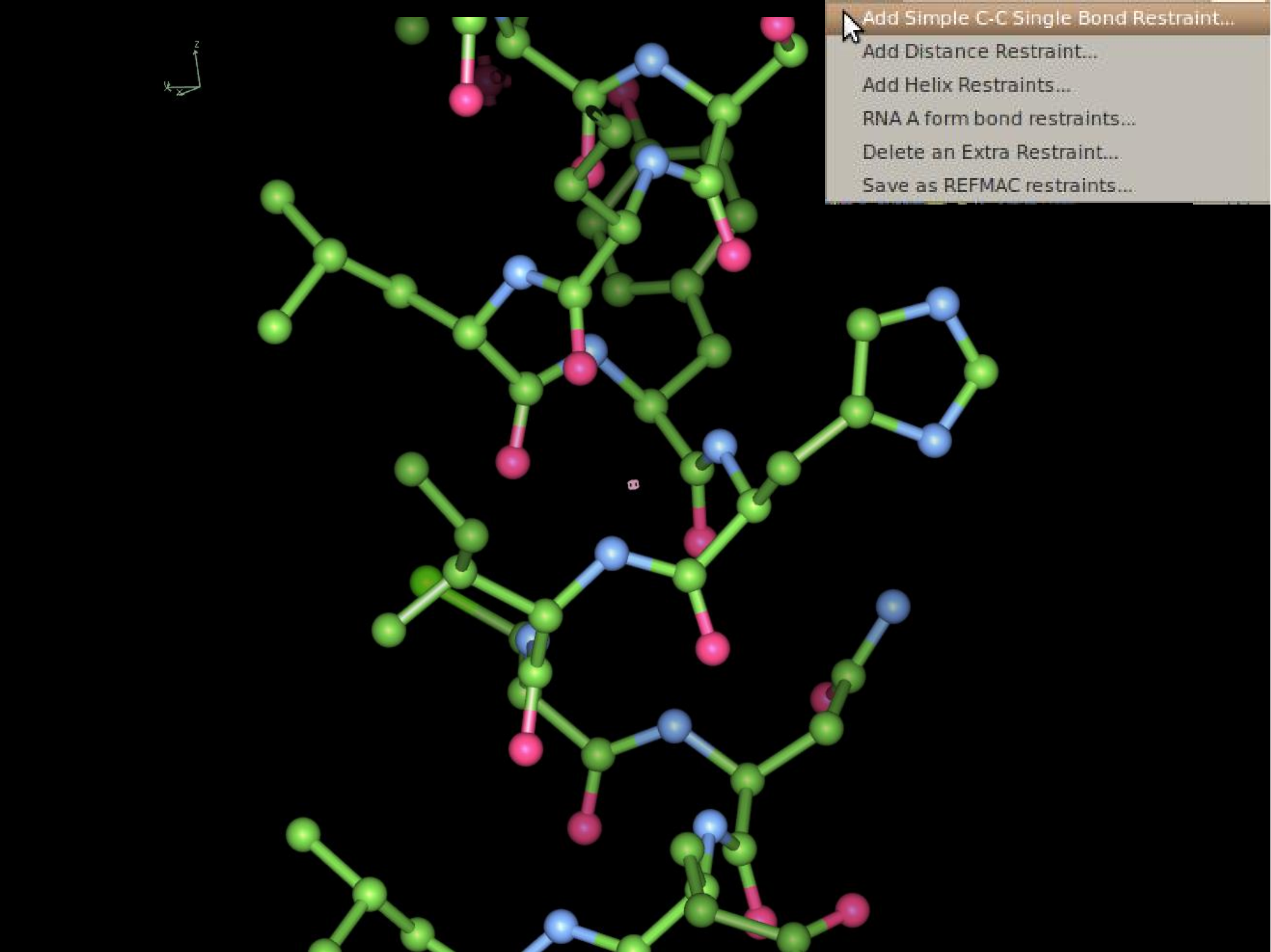
Finding Holes

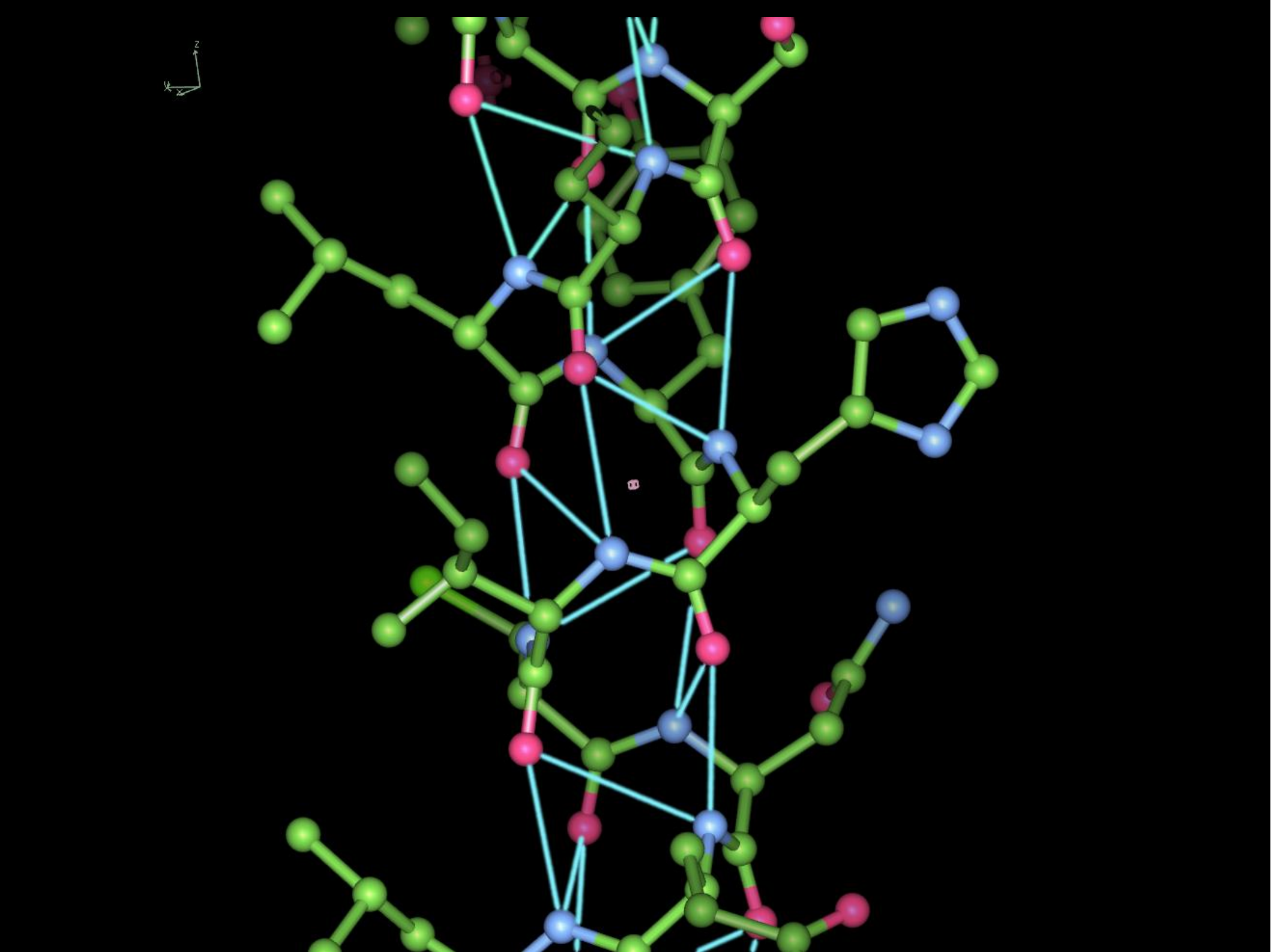
- An implementation of
  - Smart, Goodfellow & Wallace (1993) Biophysics Journal **65**, 2455
  - Atomic radii from AMBER
  - I used
    - radii from CCP4 monomer library
    - sans simulated annealing

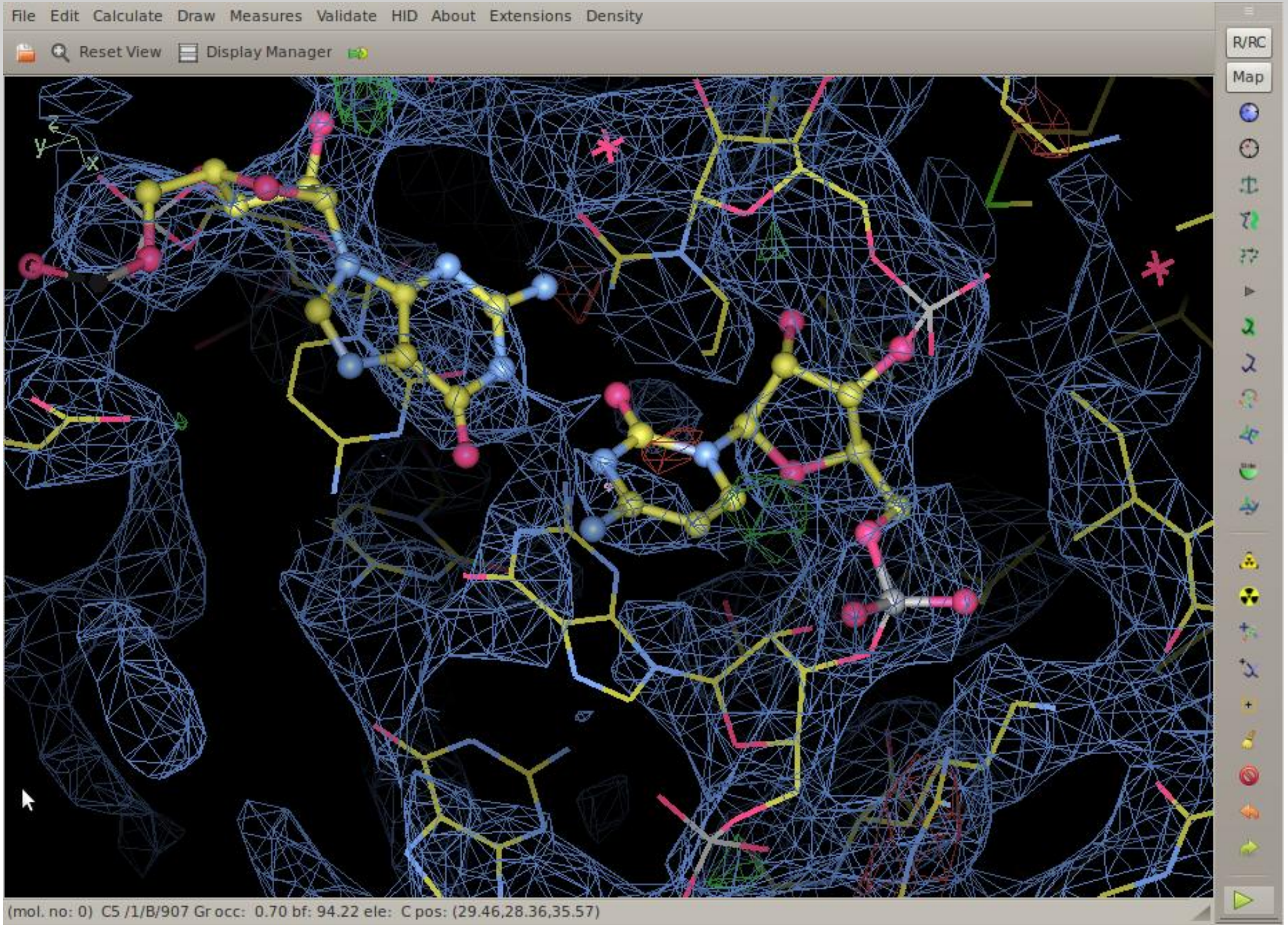

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia Test Hole
Reset View
Display Manager
R/RC
Map
Hole end point set: (-55.97 -16.51 -49.72)

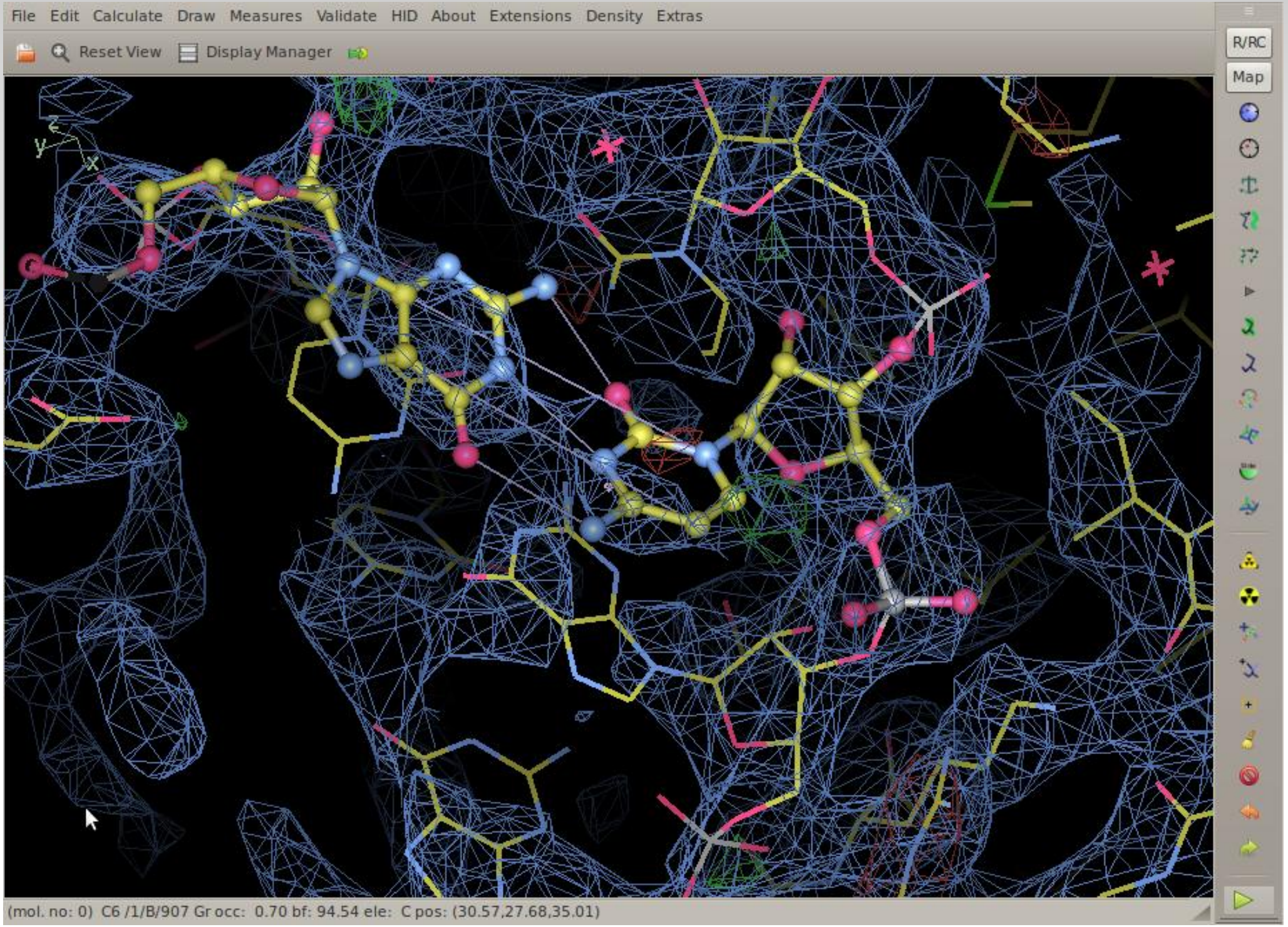
Restraints Editing in *Coot*

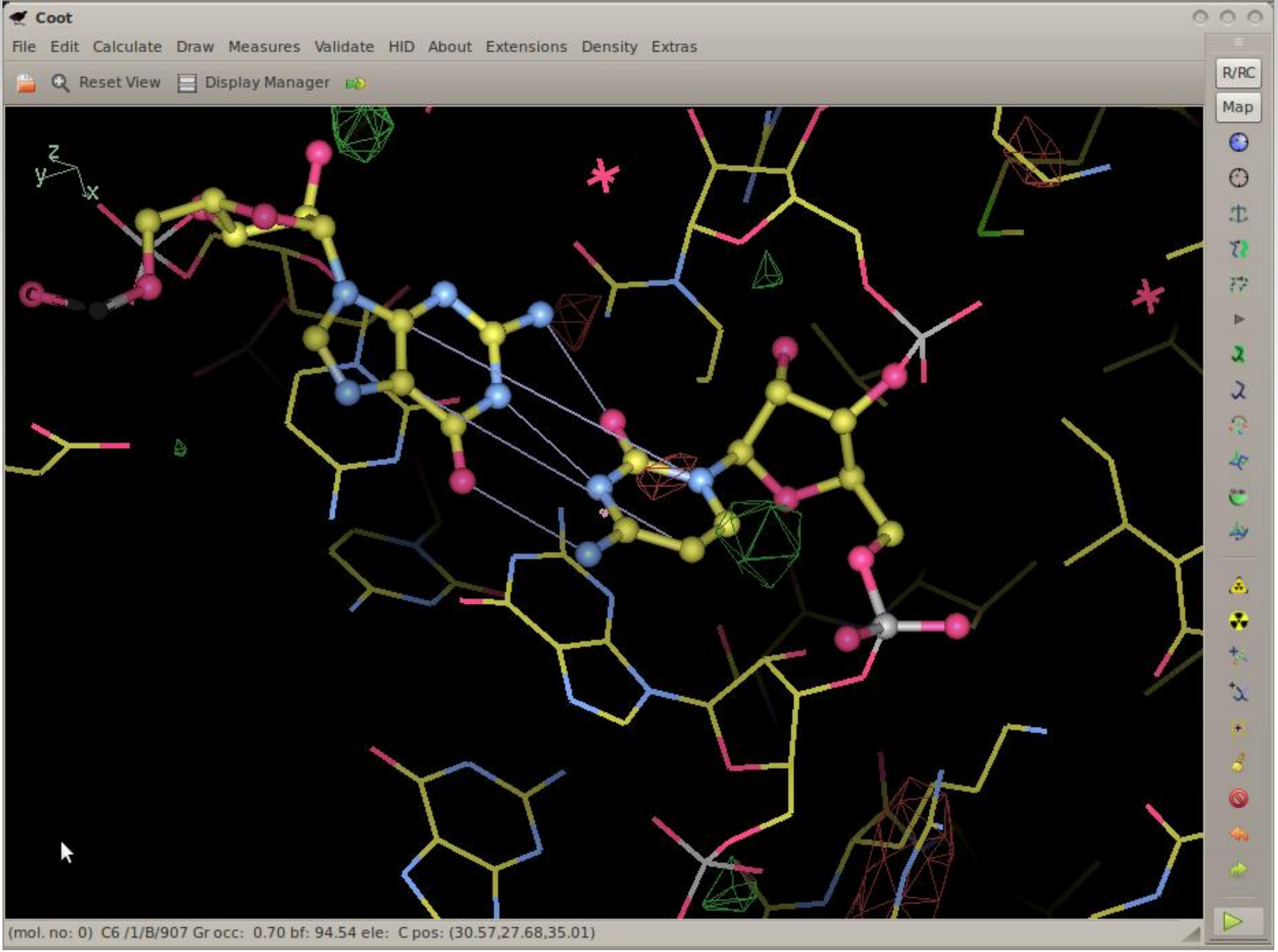
- Distance Restraints:
  - Alpha helices, A-form RNA
- Add and delete individual restraints
  - User-selectable sigma
- Select 2 residues for range
- User-defined torsion restraints
- Input from ProSMART
- Output to Refmac

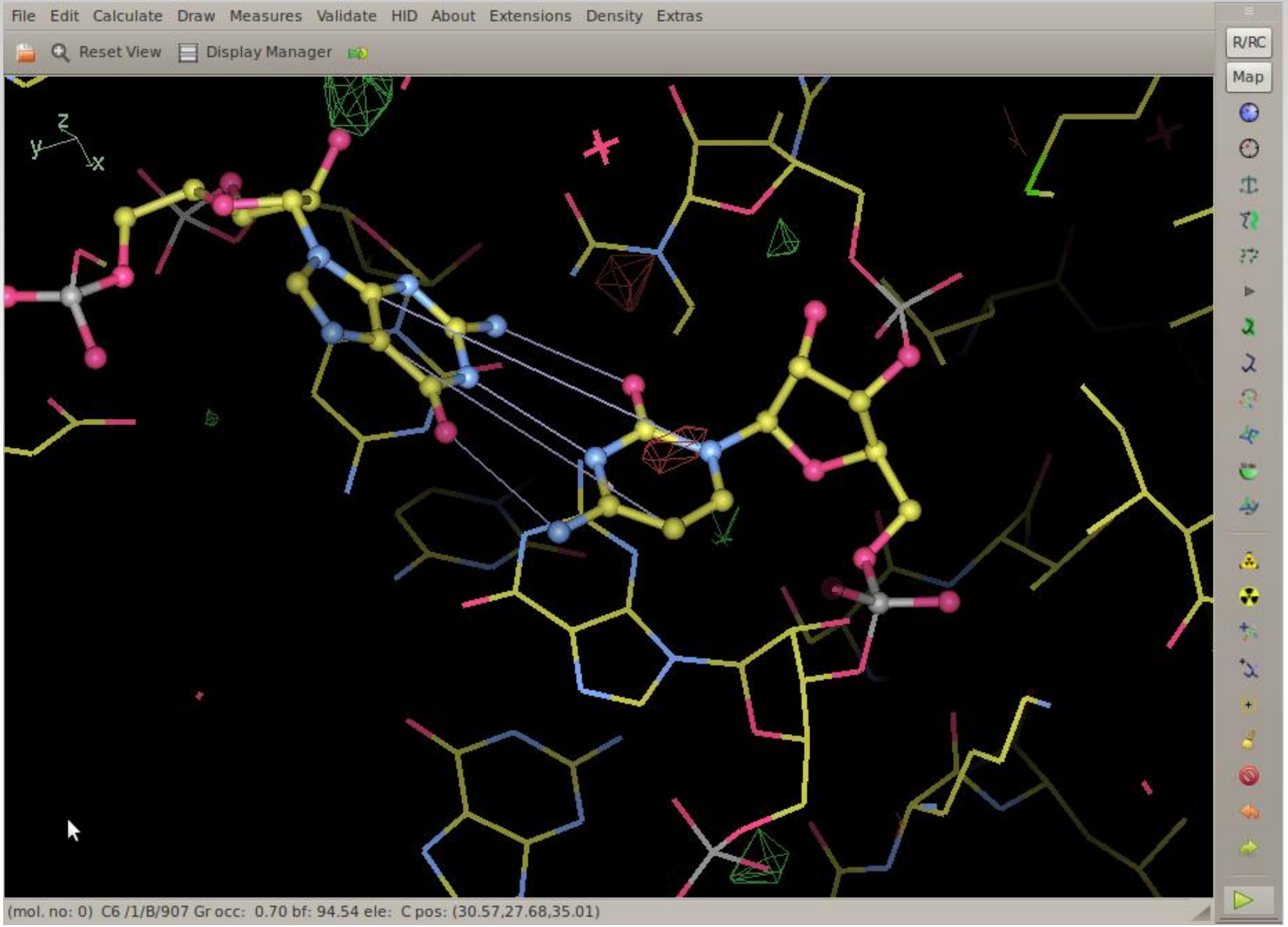

Add Simple C-C Single Bond Restraint...
Add Distance Restraint...
Add Helix Restraints...
RNA A form bond restraints...
Delete an Extra Restraint...
Save as REFMAC restraints...

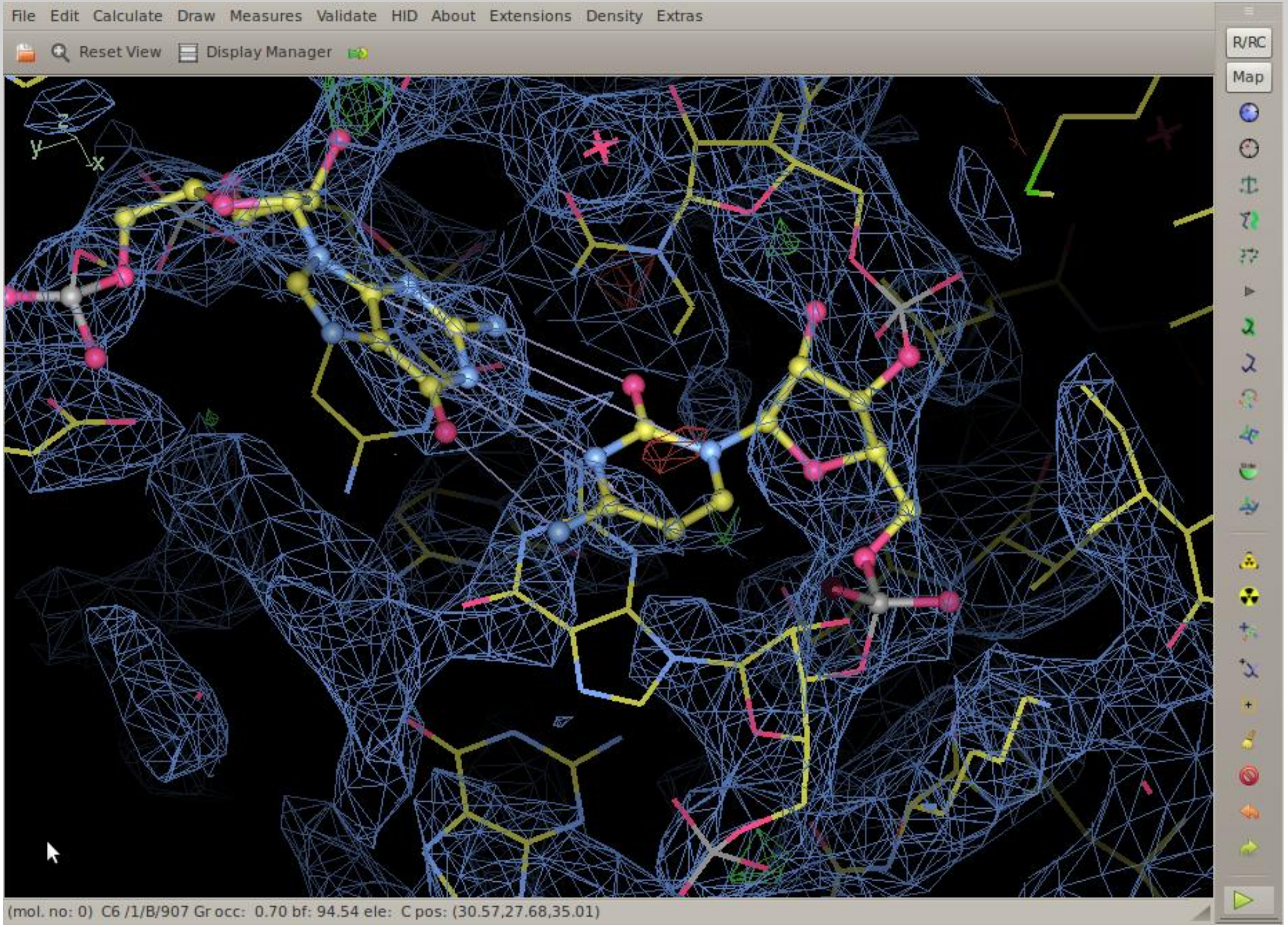



File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
(mol. no: 0) C5 /1/B/907 Gr occ: 0.70 bf: 94.22 ele: C pos: (29.46,28.36,35.57)


File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
(mol. no: 0) C6 /1/B/907 Gr occ: 0.70 bf: 94.54 ele: C pos: (30.57,27.68,35.01)


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
(mol. no: 0) C6 /1/B/907 Gr occ: 0.70 bf: 94.54 ele: C pos: (30.57,27.68,35.01)


File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
Z
Y
X
(mol. no: 0) C6 /1/B/907 Gr occ: 0.70 bf: 94.54 ele: C pos: (30.57,27.68,35.01)


File Edit Calculate Draw Measures Validate HID About Extensions Density Extras
Reset View
Display Manager
R/RC
Map
(mol. no: 0) C6 /1/B/907 Gr occ: 0.70 bf: 94.54 ele: C pos: (30.57,27.68,35.01)

Export as Refmac Restraints:

- `EXT DIST FIRST CHAIN A RESI 55 INS . ATOM  CA   SECOND CHAIN A RESI 55 INS . ATOM  C   VALUE 1.54 SIGMA 0.05`

- (And Nat Echols has been working a the phenix interface)

ProSMART Interface

- <demo>