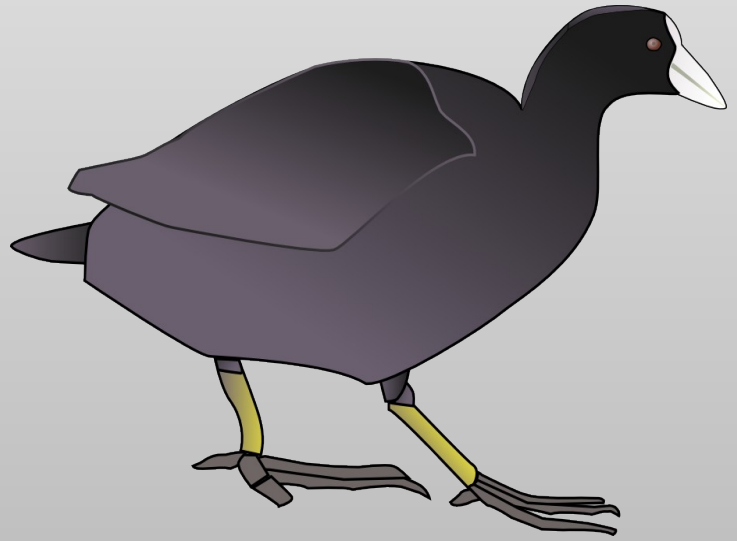
Beyond the Basics with *Coot:*

Tools for Low Resolution, EM Map Fitting & Validation

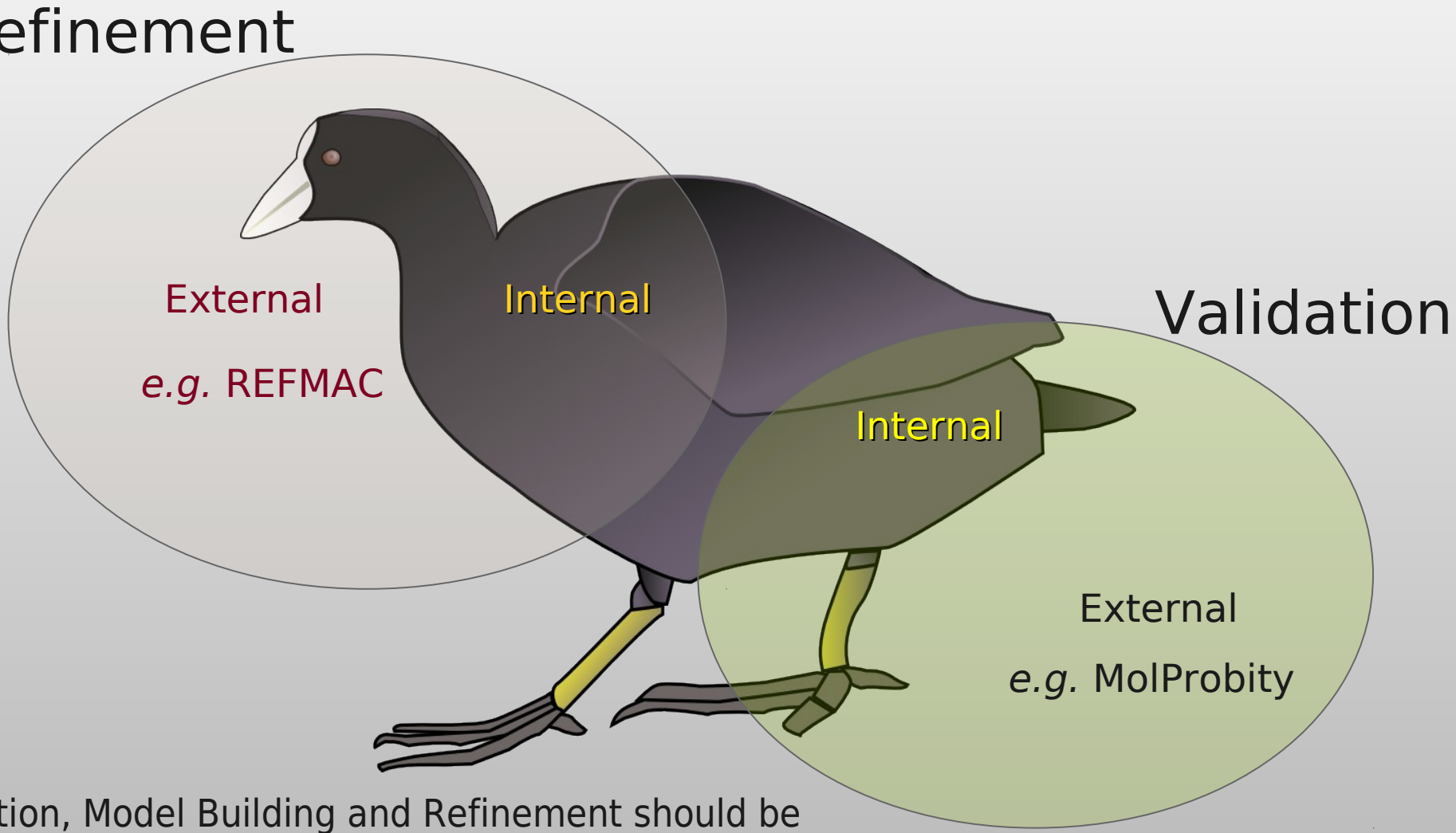


Paul Emsley
MRC Laboratory of Molecular Biology
December 2014

Coot

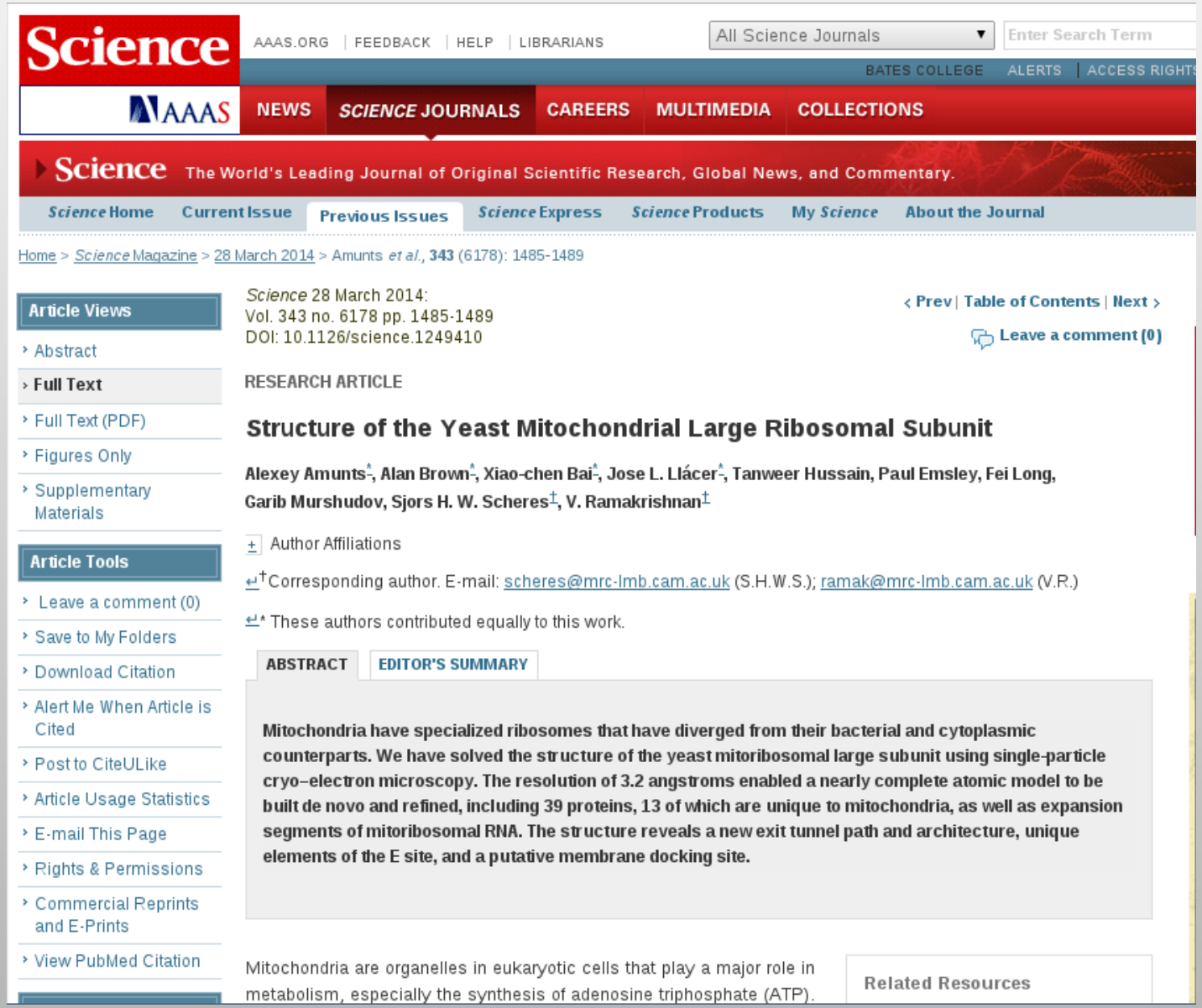
- Crystallographic Object-Oriented Tool-kit
- Primarily a tool for the interpretation of electron density generated from X-ray data
  - with tools for modelling:
    - rotate/translate, rotamers,
    - refinement & regularization
    - add, delete
    - ligand fitting
  - A “workhorse”, not a show-pony

Feature Integration



Validation, Model Building and Refinement should be used together

Yeast Mitochondrial Large Ribosomal Subunit



Science 28 March 2014:
Vol. 343 no. 6178 pp. 1485-1489
DOI: 10.1126/science.1249410

RESEARCH ARTICLE

Structure of the Yeast Mitochondrial Large Ribosomal Subunit

Alexey Amunts[*], Alan Brown[*], Xiao-chen Bai[*], Jose L. Llácer[*], Tanweer Hussain, Paul Emsley, Fei Long, Garib Murshudov, Sjors H. W. Scheres[†], V. Ramakrishnan[†]

[+] Author Affiliations

[†]Corresponding author. E-mail: scheres@mrc-lmb.cam.ac.uk (S.H.W.S.); ramak@mrc-lmb.cam.ac.uk (V.R.)

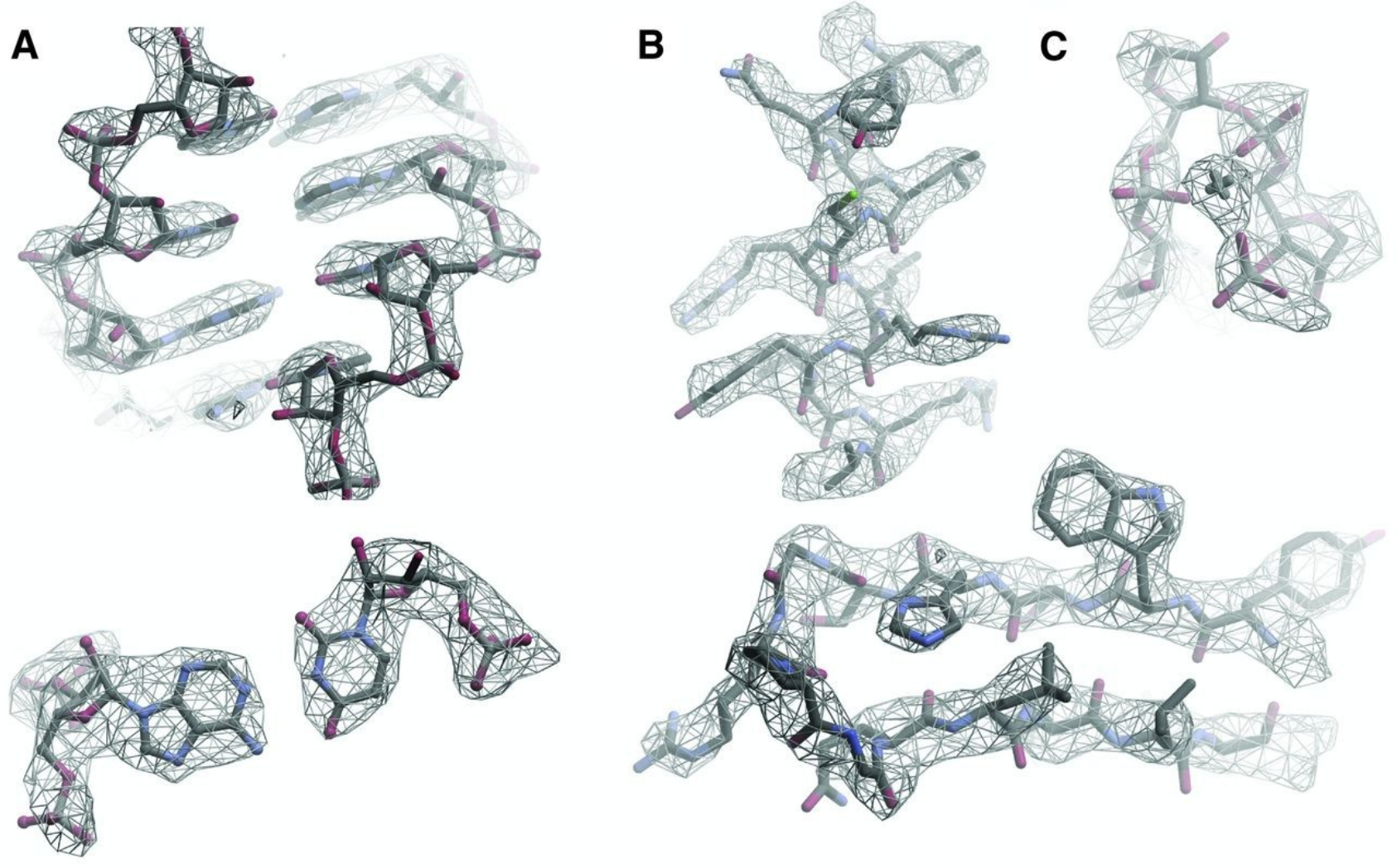
[*] These authors contributed equally to this work.

ABSTRACT | EDITOR'S SUMMARY

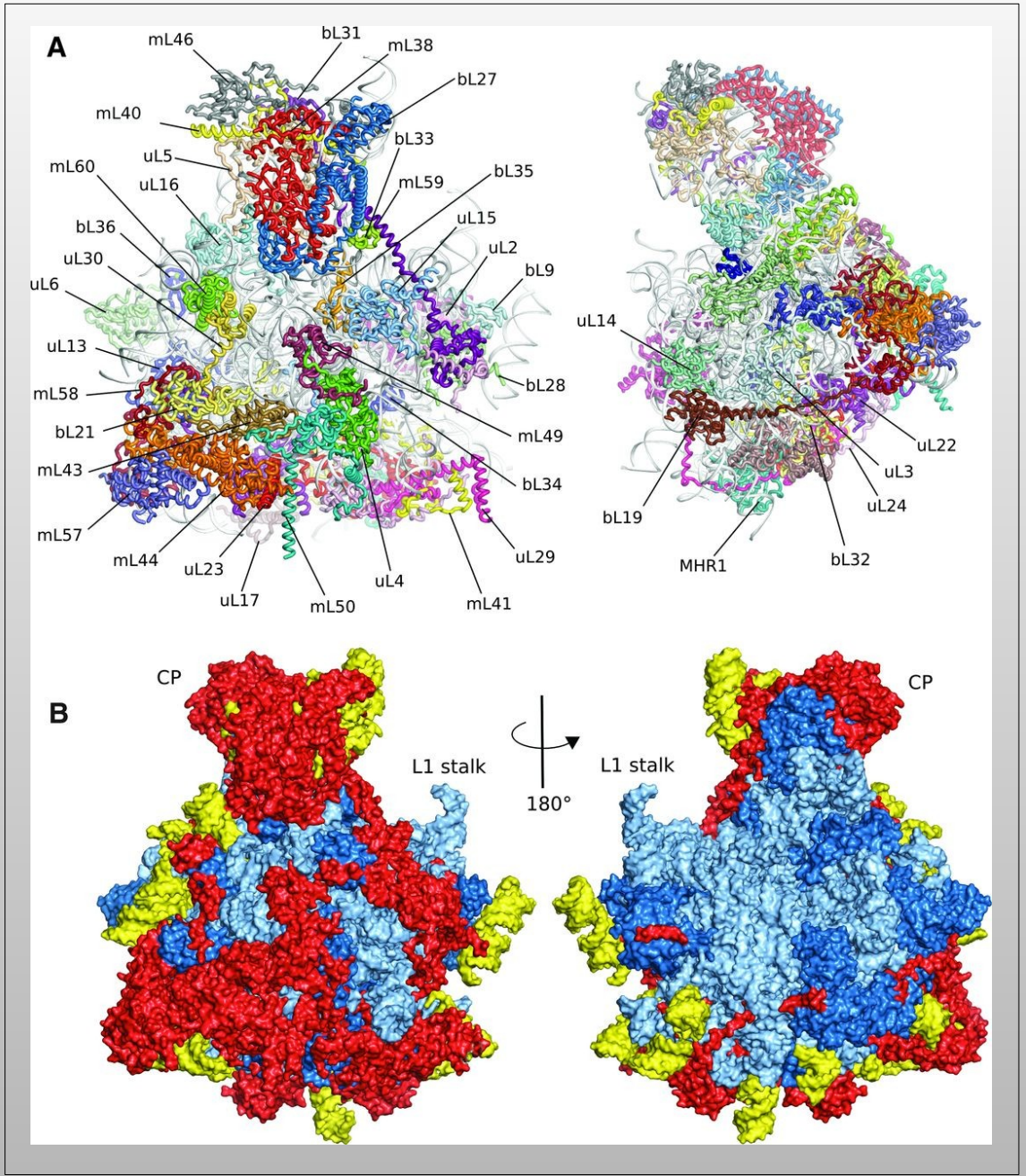
Mitochondria have specialized ribosomes that have diverged from their bacterial and cytoplasmic counterparts. We have solved the structure of the yeast mitoribosomal large subunit using single-particle cryo–electron microscopy. The resolution of 3.2 angstroms enabled a nearly complete atomic model to be built de novo and refined, including 39 proteins, 13 of which are unique to mitochondria, as well as expansion segments of mitoribosomal RNA. The structure reveals a new exit tunnel path and architecture, unique elements of the E site, and a putative membrane docking site.

Mitochondria are organelles in eukaryotic cells that play a major role in metabolism, especially the synthesis of adenosine triphosphate (ATP).

Density Examples



New Components



Low Resolution Tools Overview

- Backrub Rotamers
- “Jiggle Fit”
  - (interactive map sharpening if needed)
- Map fragments for Molecular Replacement
- Morphing
- Parallel Plane Restraints

Real Space Refinement

- Major Feature of Coot
  - Gradient-based minimiser (BFGS derivative)
  - Geometry library is the standard mmCIF-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
  - e.g. Sphere Refine
    - e.g. Metal sites

Refinement Techniques

- Dragging an atom with Sphere Refine...
  - too much moves (I need to address this)
- so for now, use:
  - Single-Atom Drag
    - Over-dragging
  - Key-bindings:
    - Triple Refine
    - Single Residue Refine with Auto-accept
- Parallel Plane Restraints

”Backrub Rotamers”

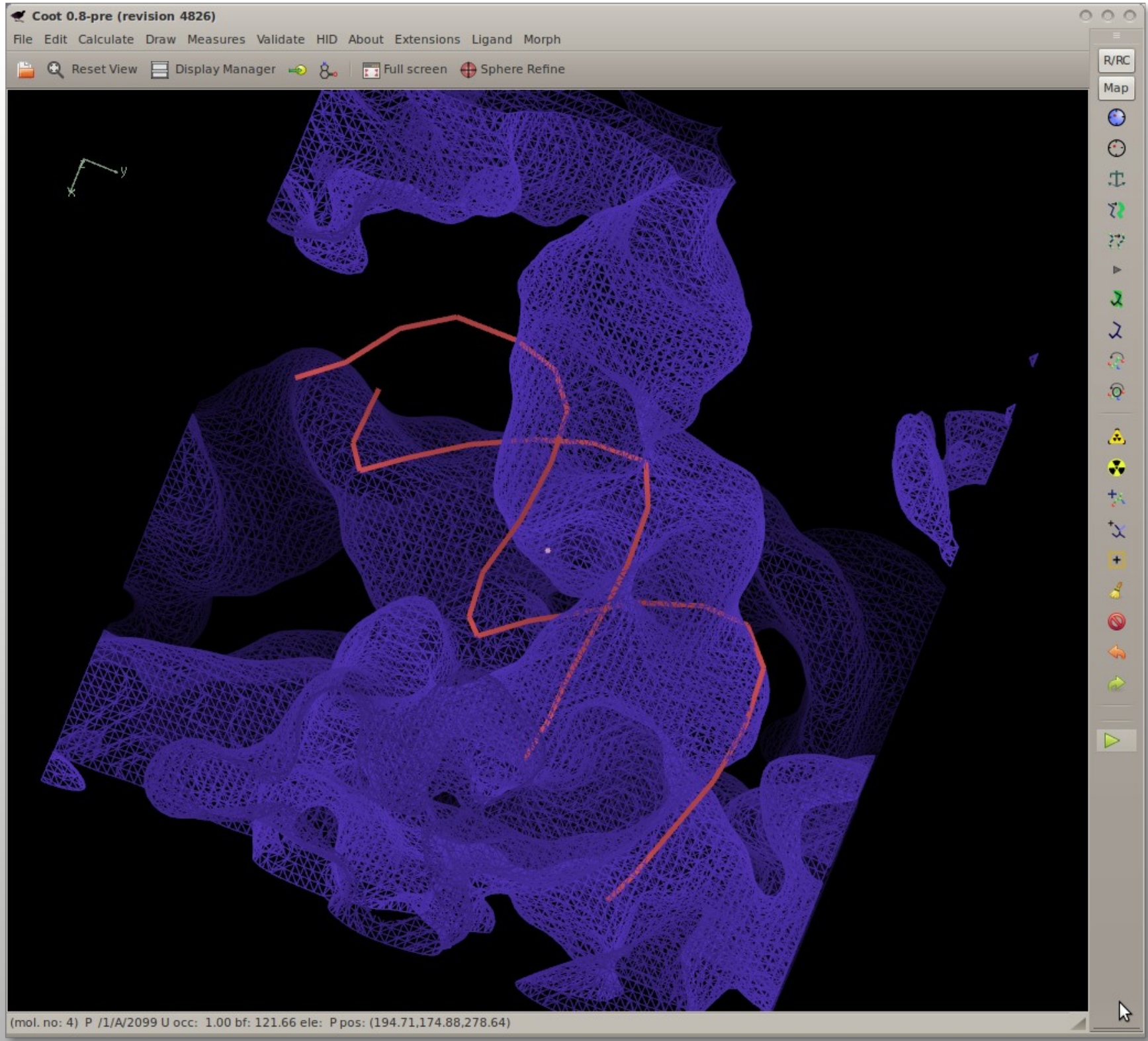
- High probability models with low resolution data

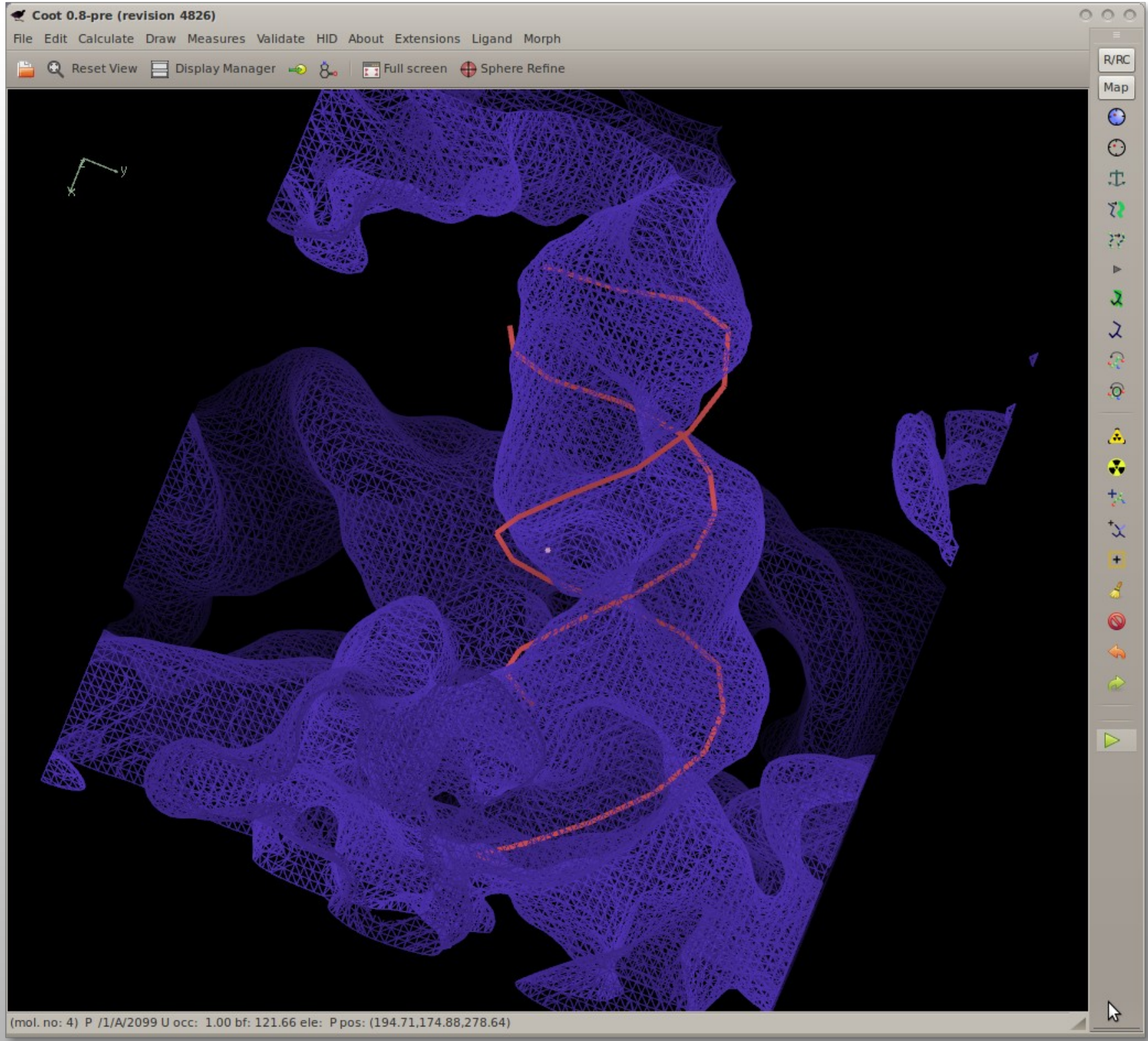
Jiggle Fit

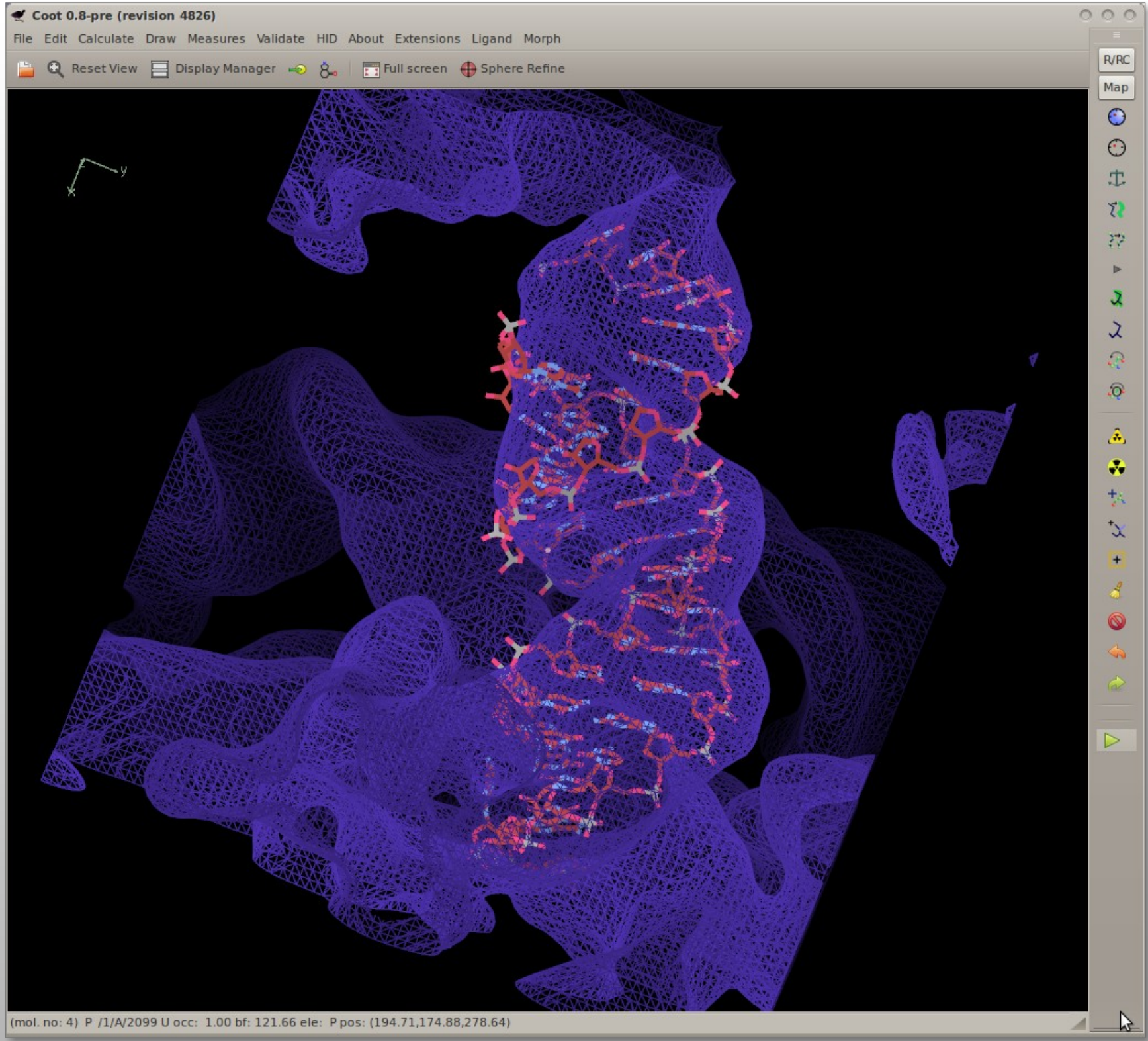
- How do I rotate and translate these atoms to fit the density?
  - 6-dimensional problem
- Originally used to fit simple ligands/solvent molecules to blobs of density
- Now extended to fit arbitrary atom selections
  - *e.g.* by Chain

Jiggle Fit: How it Works

- Loop *n* (say 1000) times:
  - Generate random angles and translations
  - Transform atom selection by these rotations and translation
  - Score and store the fit to density
- Rank density fit scores,
  - Pick top 20 solution, for each of them
    - Rigid body fit and score solutions
    - Pick the highest scoring solution if it's better than the starting model)
- Radius of Convergence is larger when using a low-pass map


Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
y
(mol. no: 4) P /1/A/2099 U occ: 1.00 bf: 121.66 ele: P pos: (194.71,174.88,278.64)


Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
y
(mol. no: 4) P /1/A/2099 U occ: 1.00 bf: 121.66 ele: P pos: (194.71,174.88,278.64)

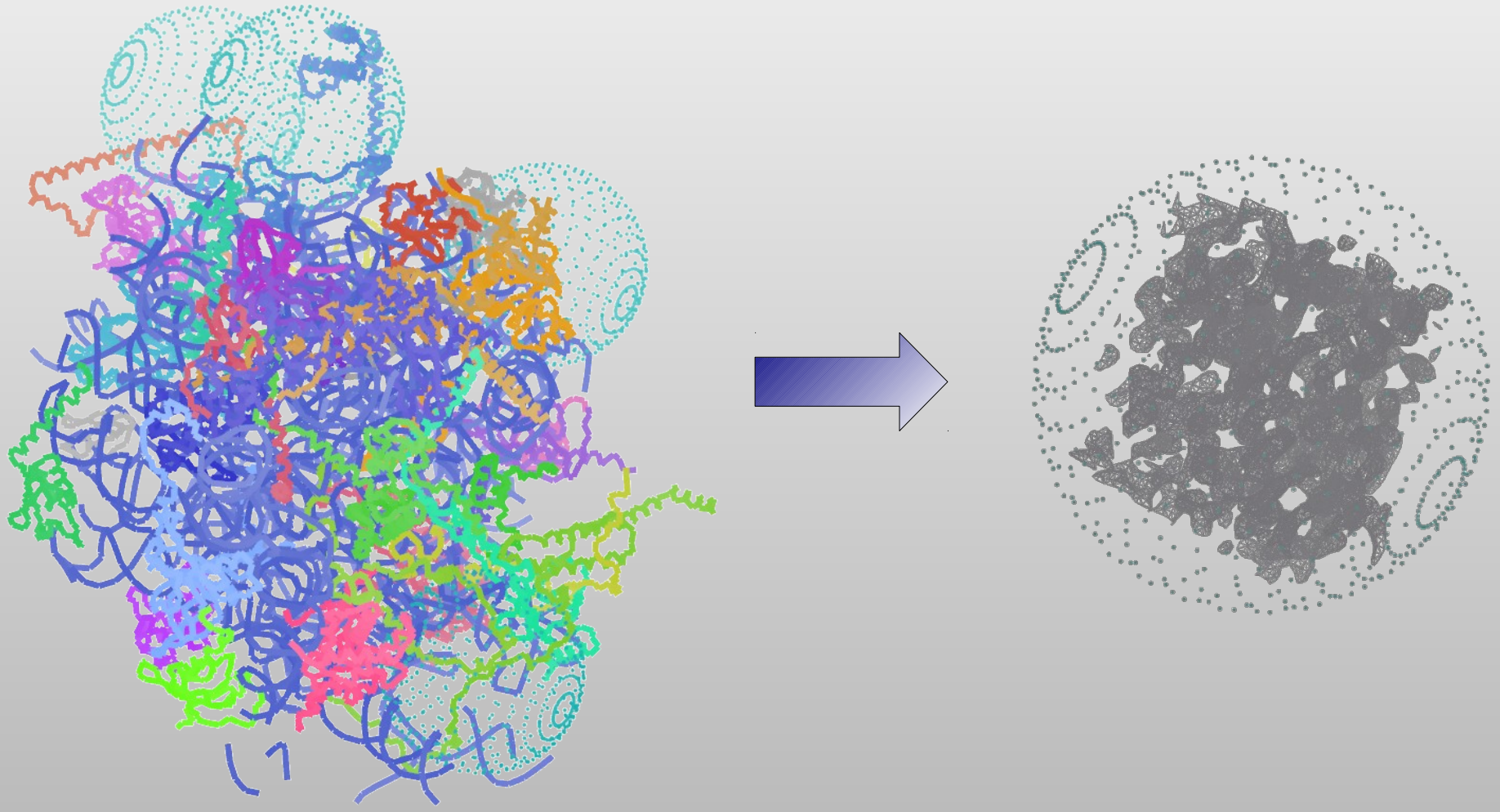

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
y
(mol. no: 4) P /1/A/2099 U occ: 1.00 bf: 121.66 ele: P pos: (194.71,174.88,278.64)

Molecular Replacement

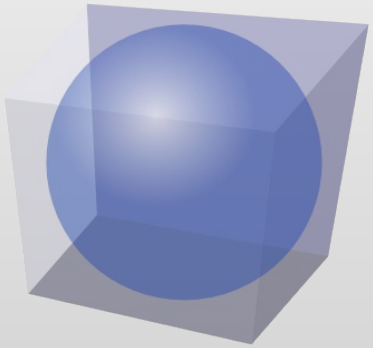
- What's this blob?

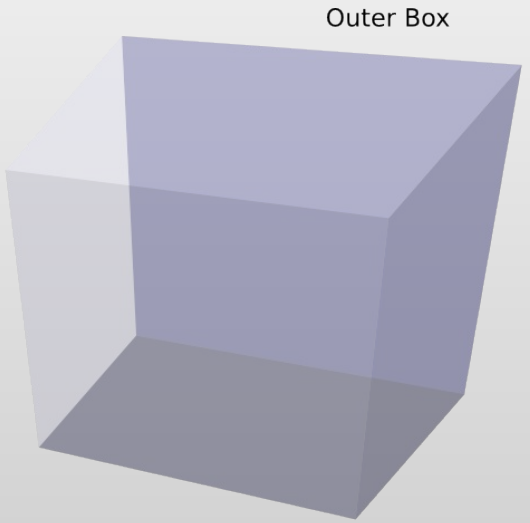
- What was known?
  - Some additional proteins that had some “homology” with known structures
  - “Let's make models and see if they fit the density”

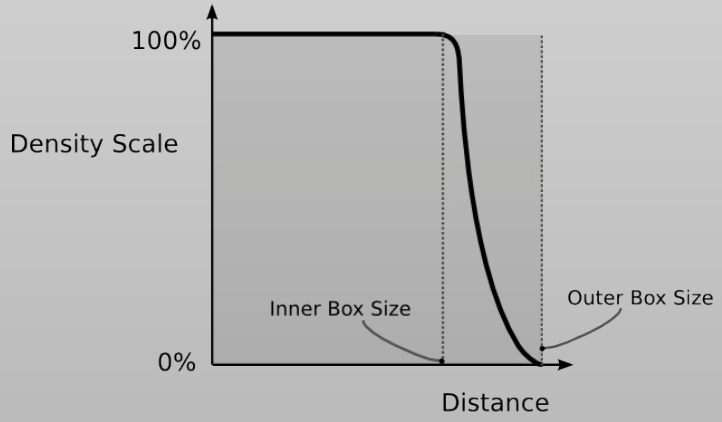
Molecular Replacement by Blob Extraction

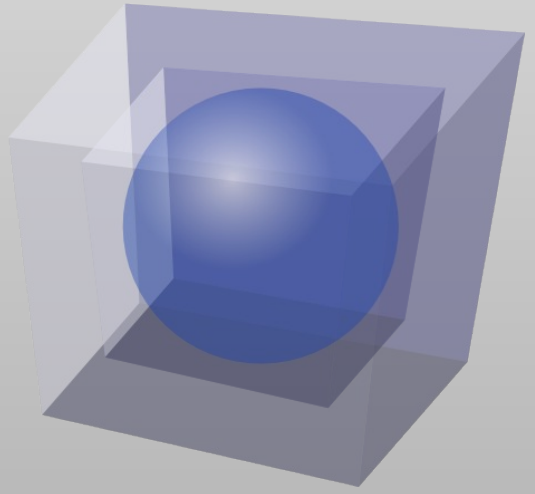


Density Preparation
for Molecular Replacement





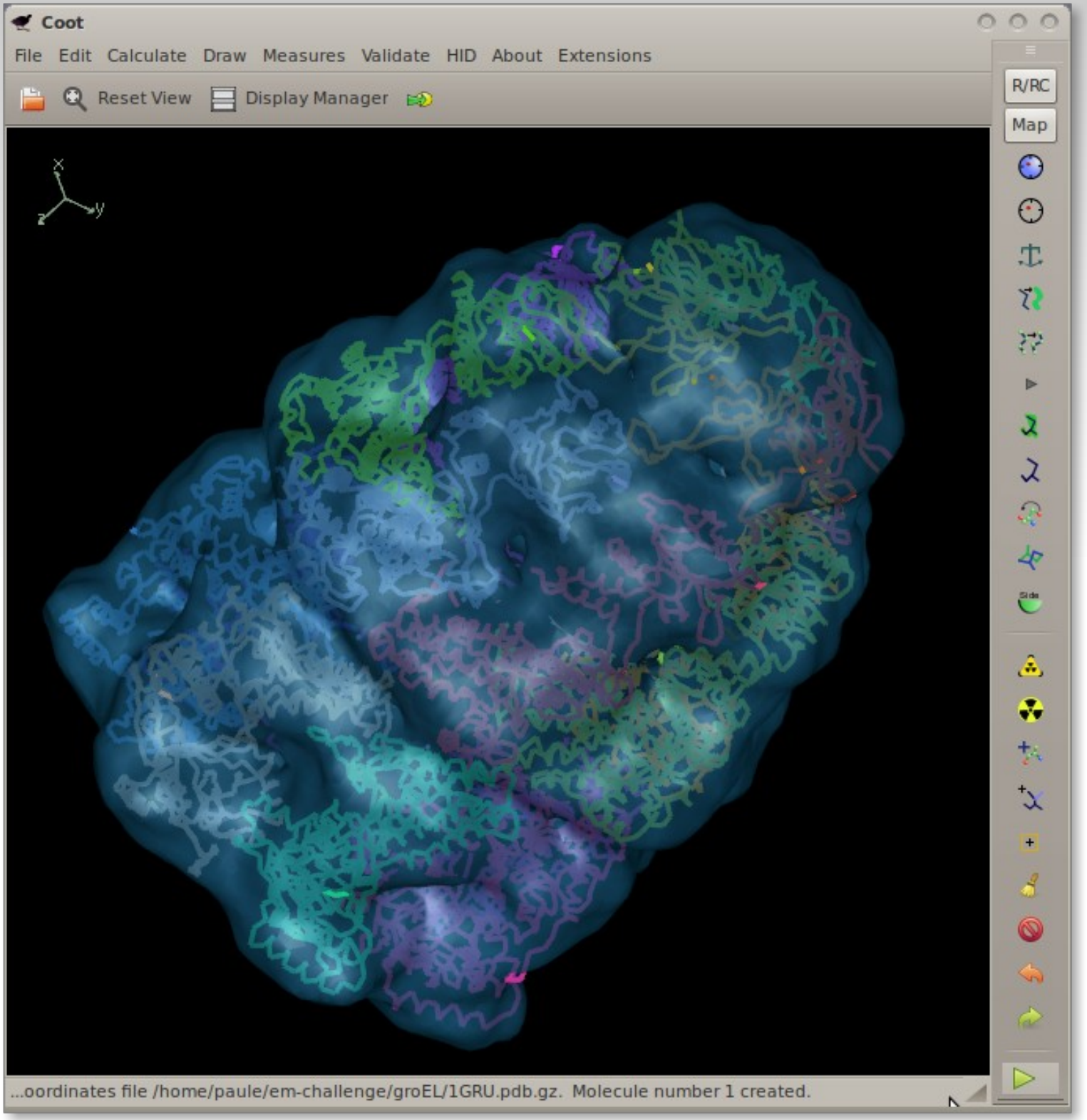


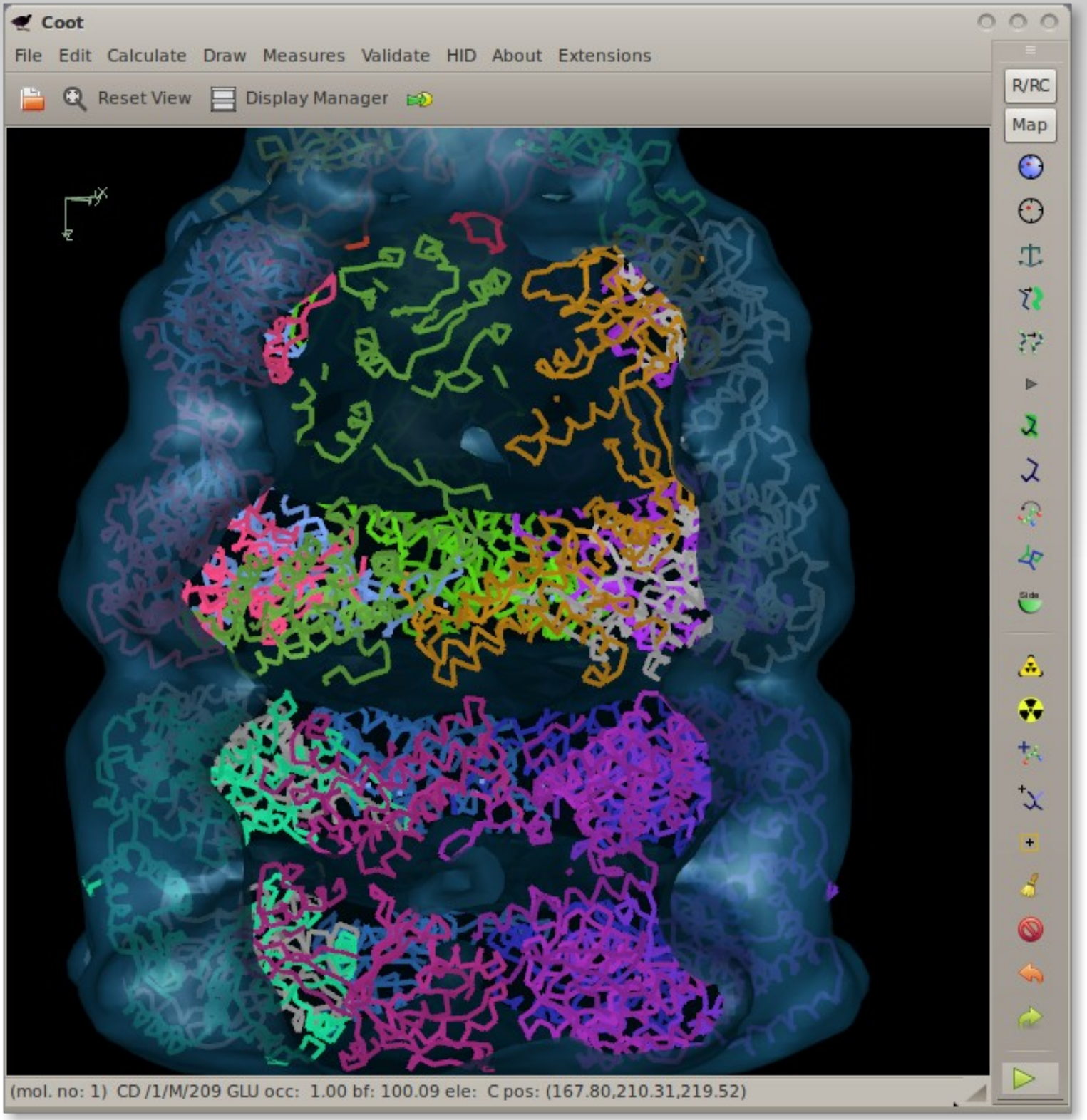


& translate to origin

Molecular Replacement

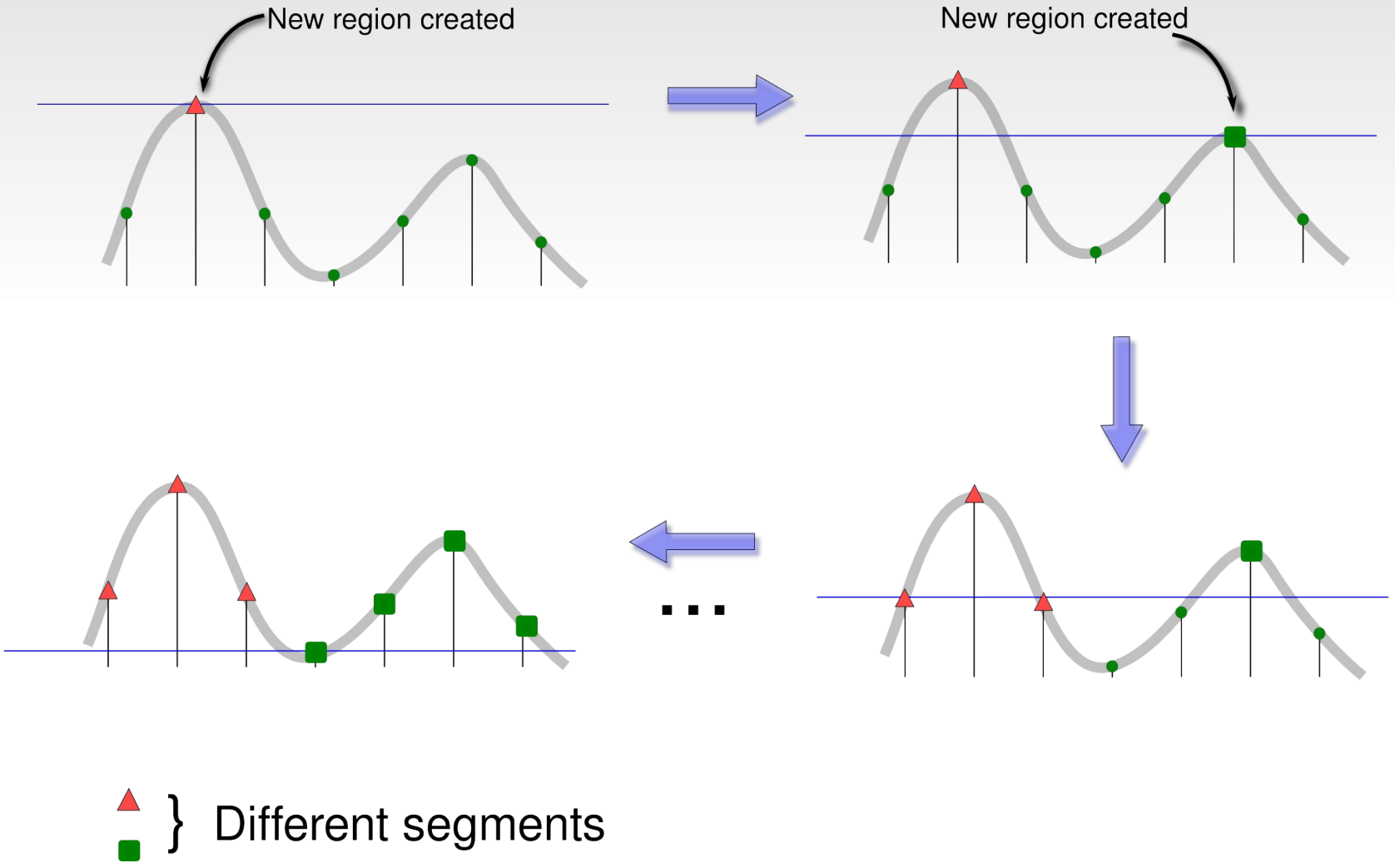
- We had to tweak the Grid Sampling/Box Radius
  - so that `molrep` would read the map
- Convert to Structure Factors
- Straightforward when there is a set of candidate models
  - But there were some regions of the map that could not be assigned to a previously know homologous structure
  - So search against the BALBES Domain Database
  - ... use the cluster and Sun Grid Engine
    - (take care not to set everything off at once, otherwise the NFS server is brought to its knees)

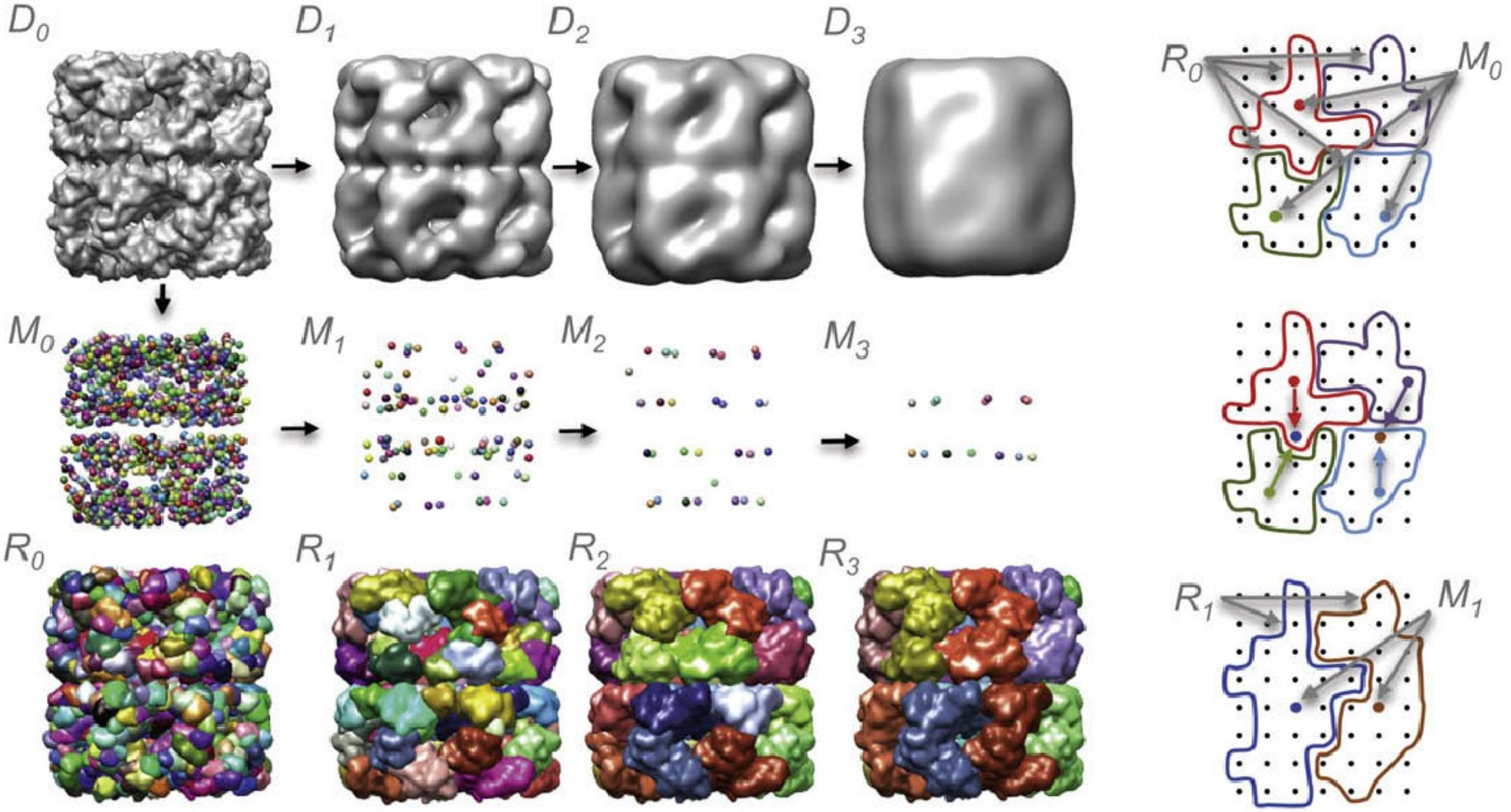

Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
R/RC
Map
...oordinates file /home/paule/em-challenge/groEL/1GRU.pdb.gz. Molecule number 1 created.


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)

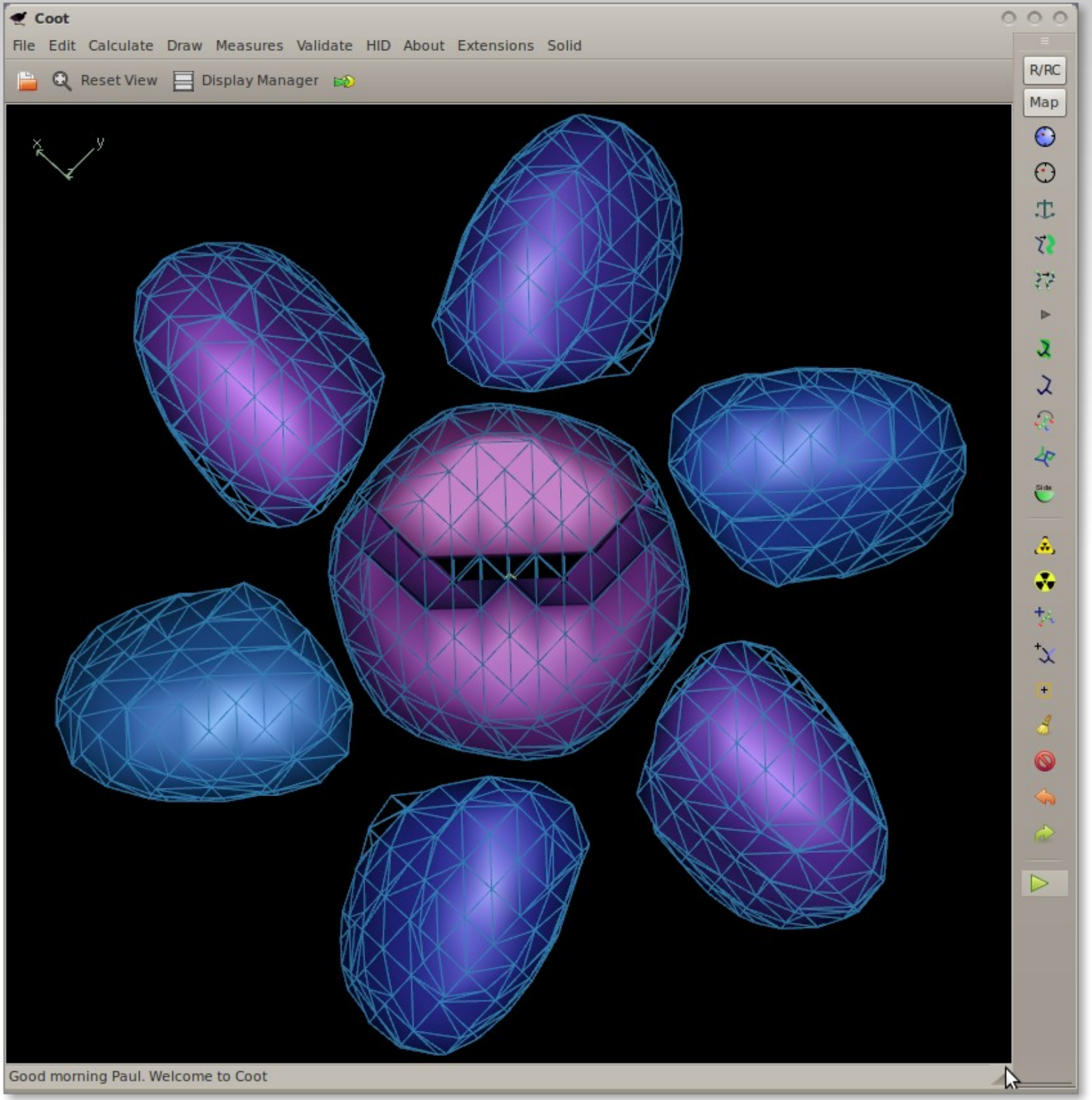
Partioning Maps: Watershed Algoritm

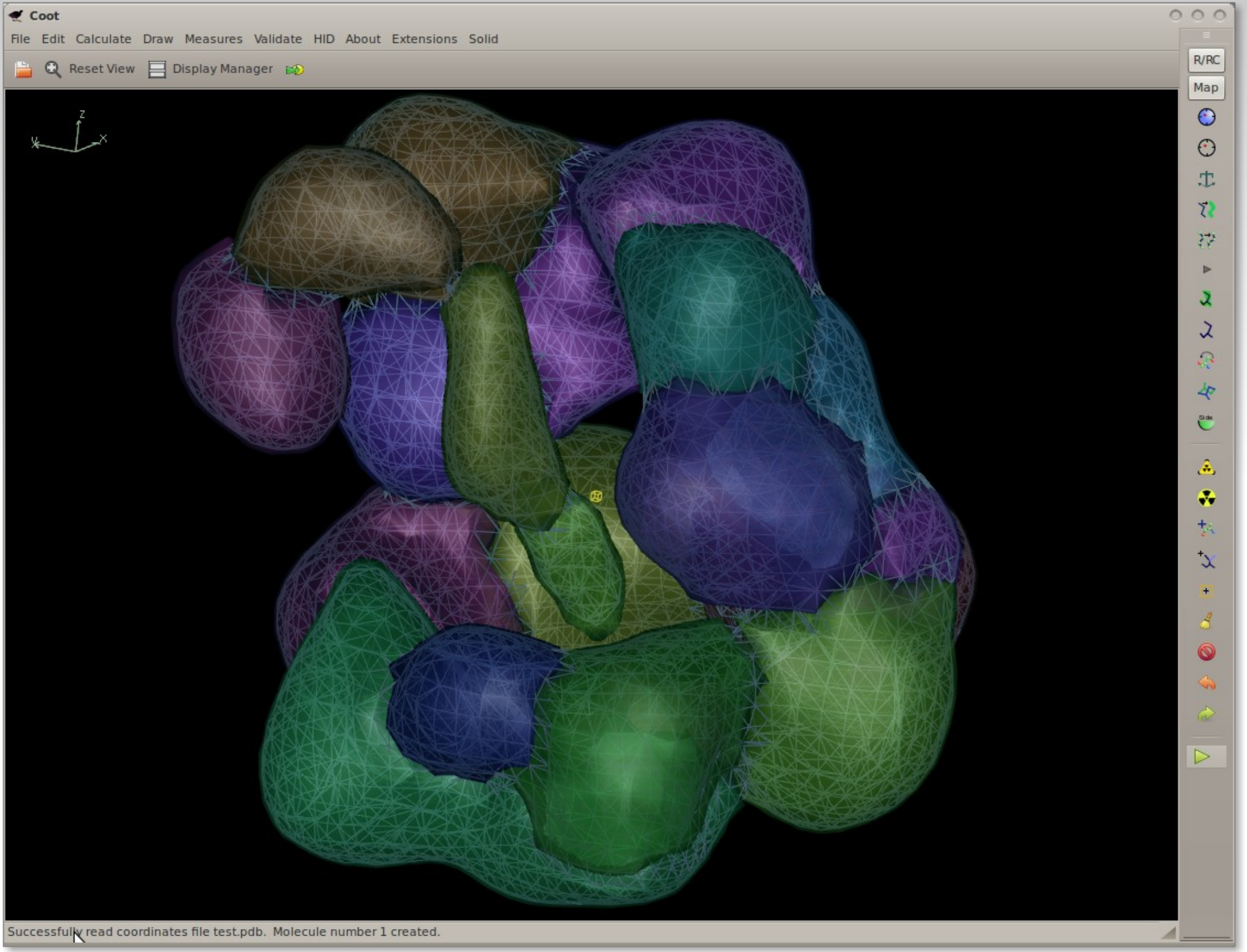
1D-analog

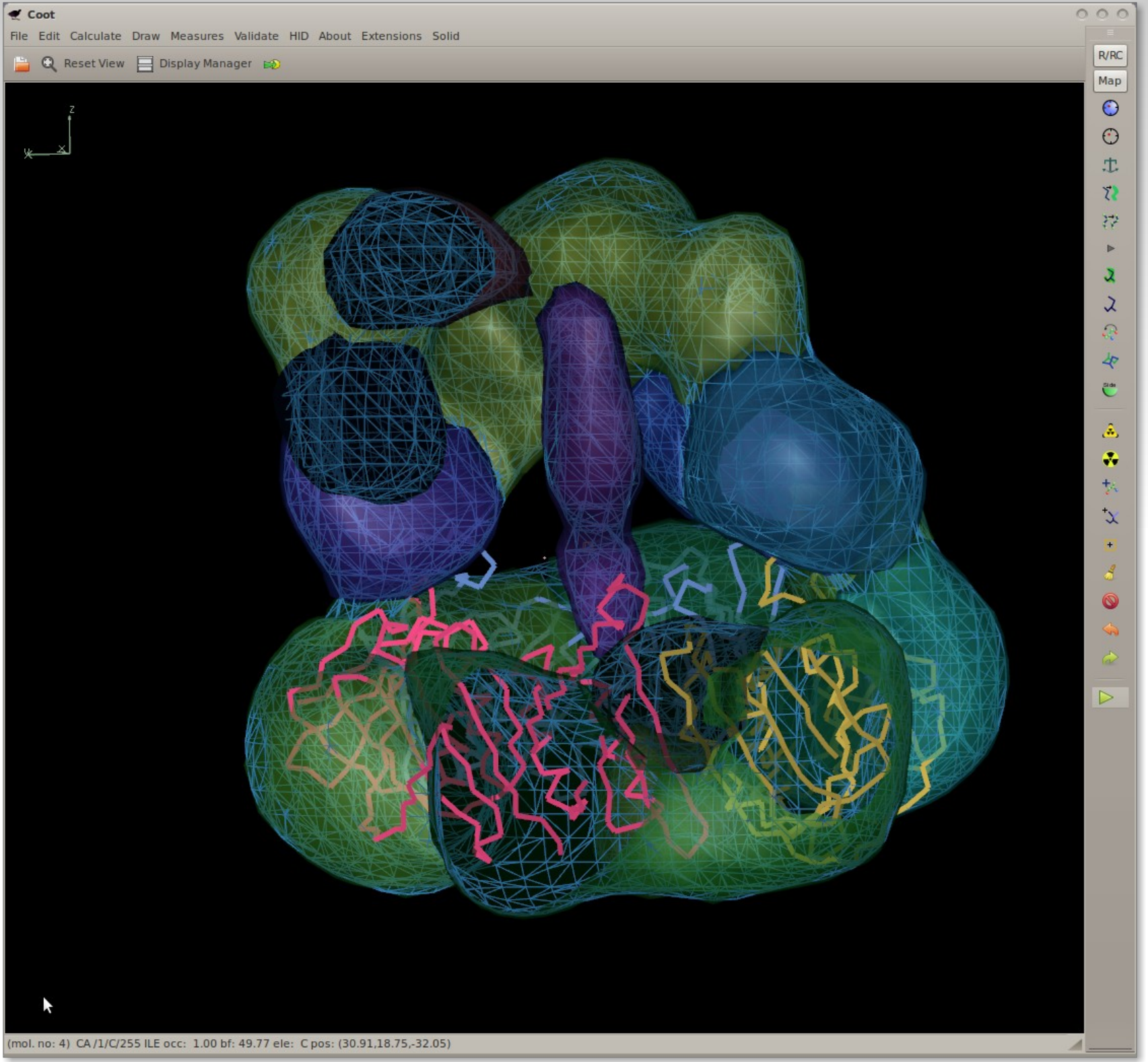




Pintilie *et al.* (2010)
J.Struct.Biol.


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?

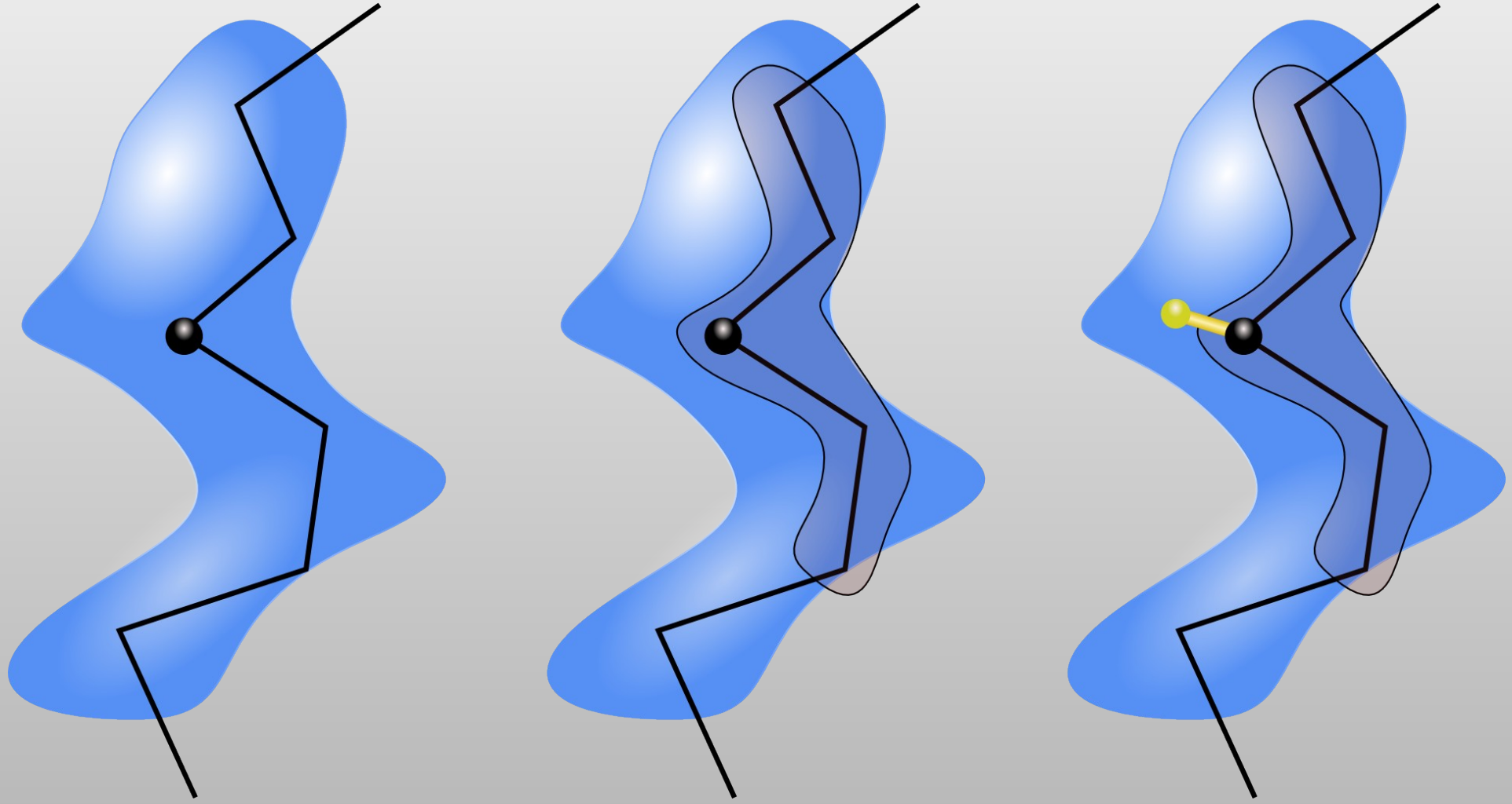
So we have our ideal RNA or homologous protein sitting roughly in the density

(not a great fit)

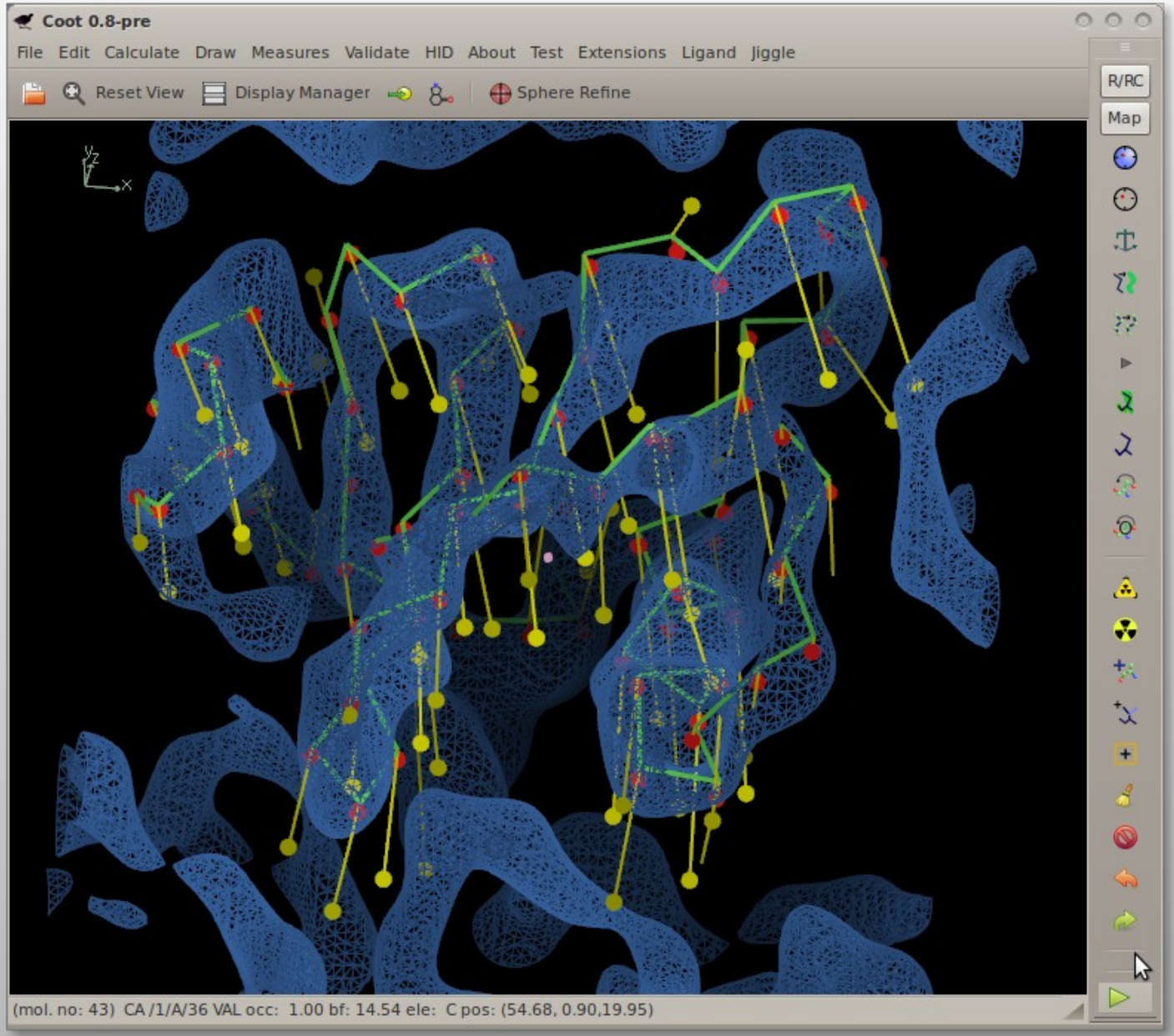
Model Morphing: How it Works

- For each residue in a chain, we ask:
  - where does a small fragment centred on this residue want to go?
  - (Robust) average the transformations and apply them on a per-residue basis
- Repeat

Model Morphing: Generating the Raw RTs

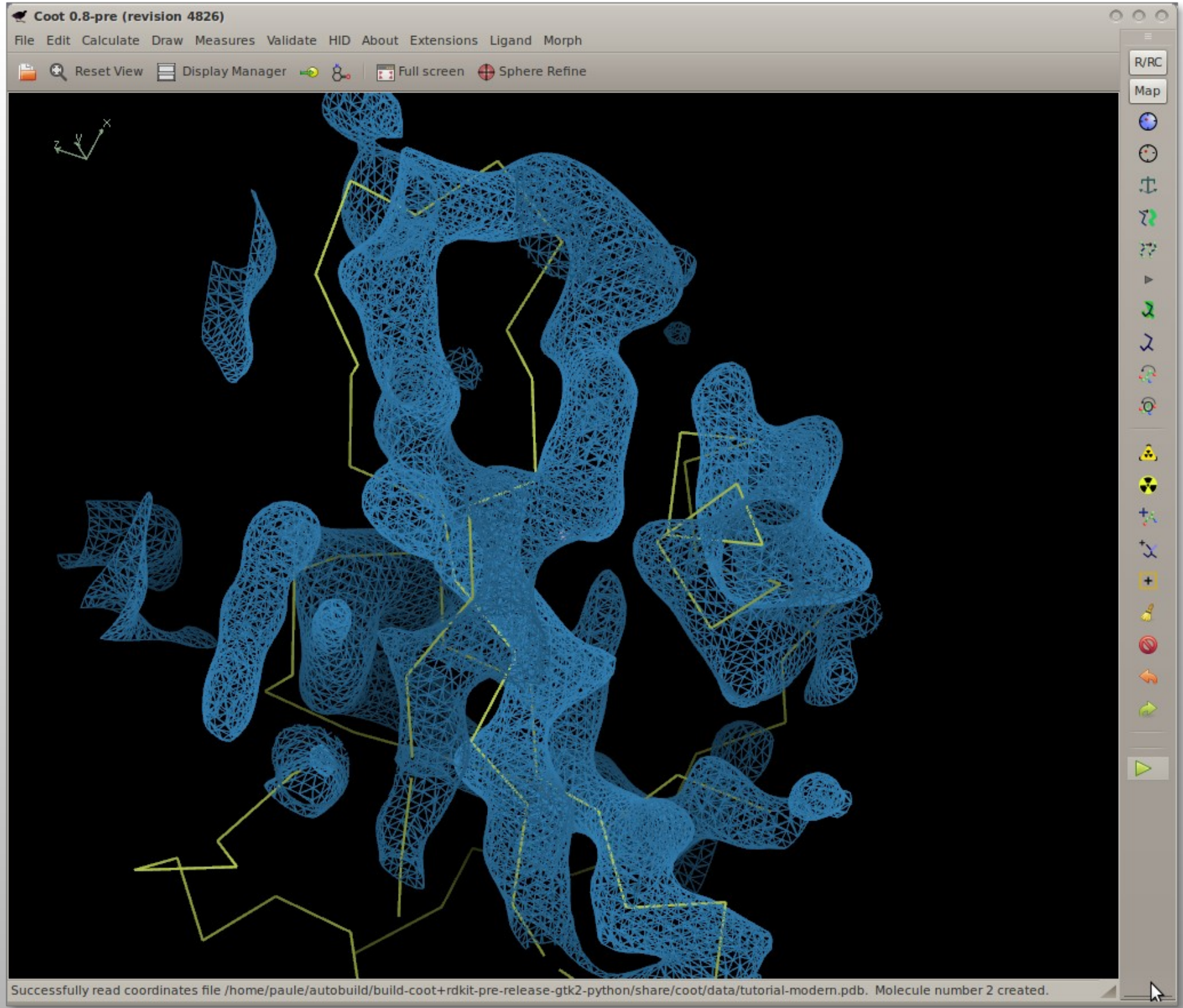


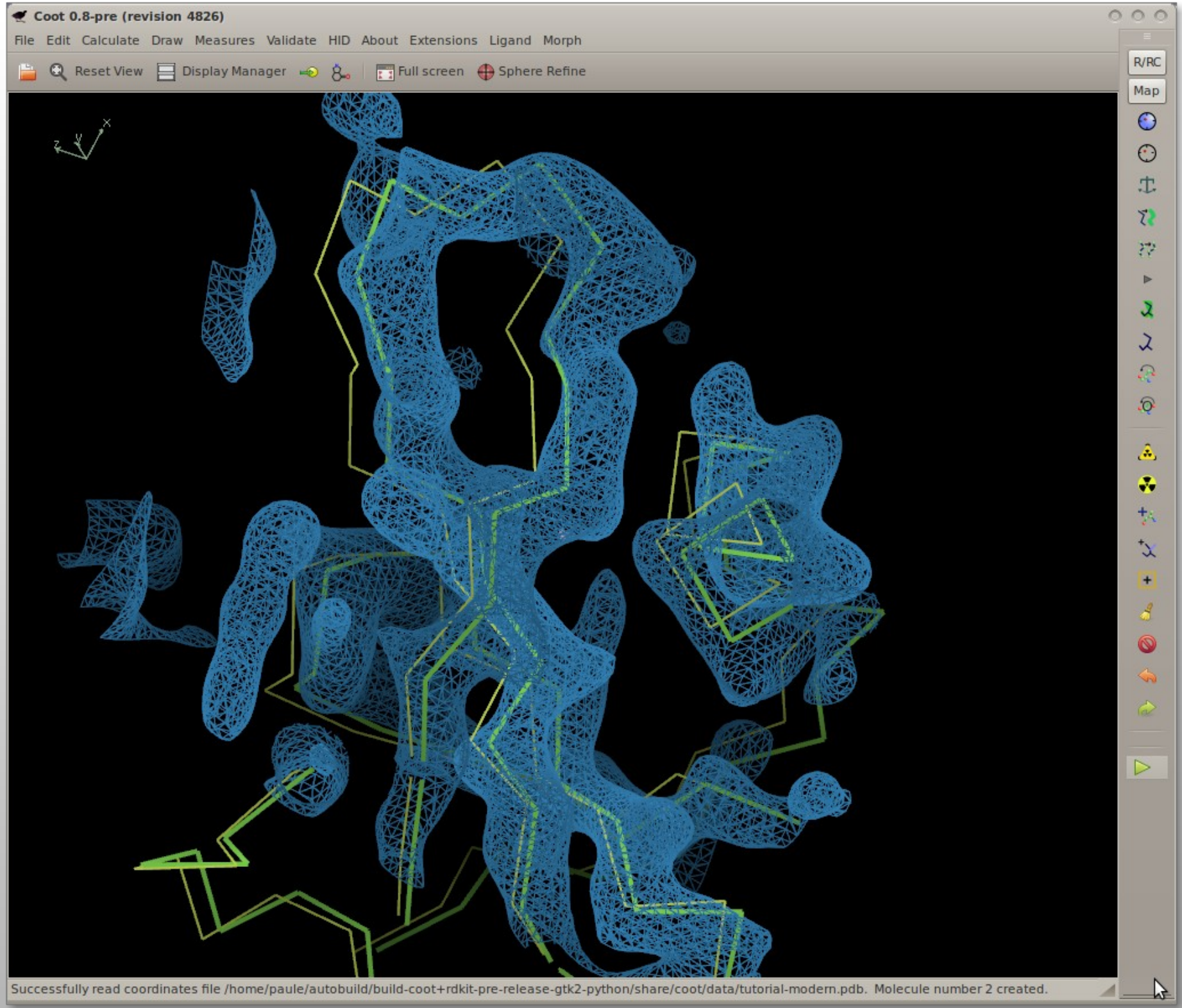
Model Morphing: Example

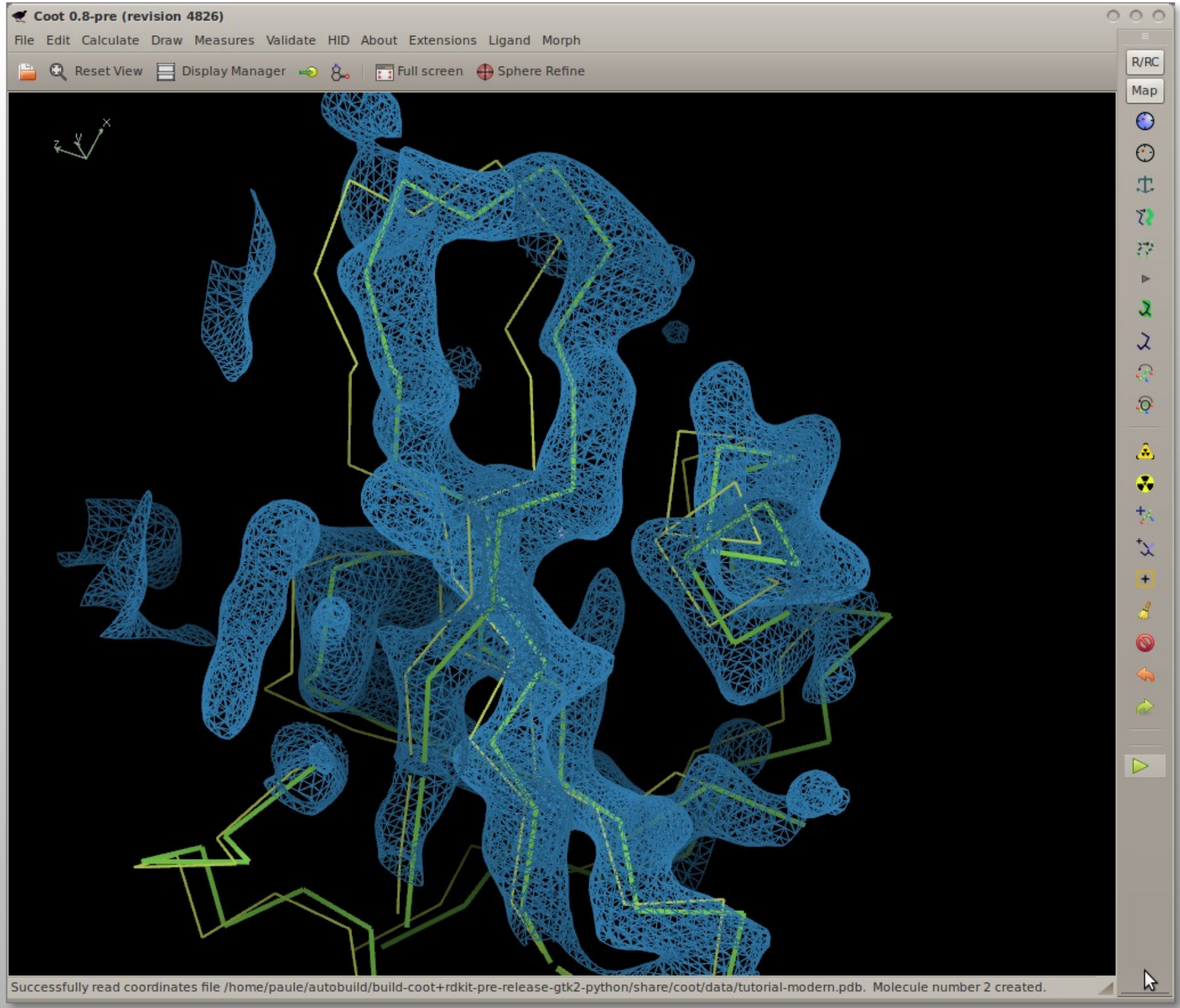


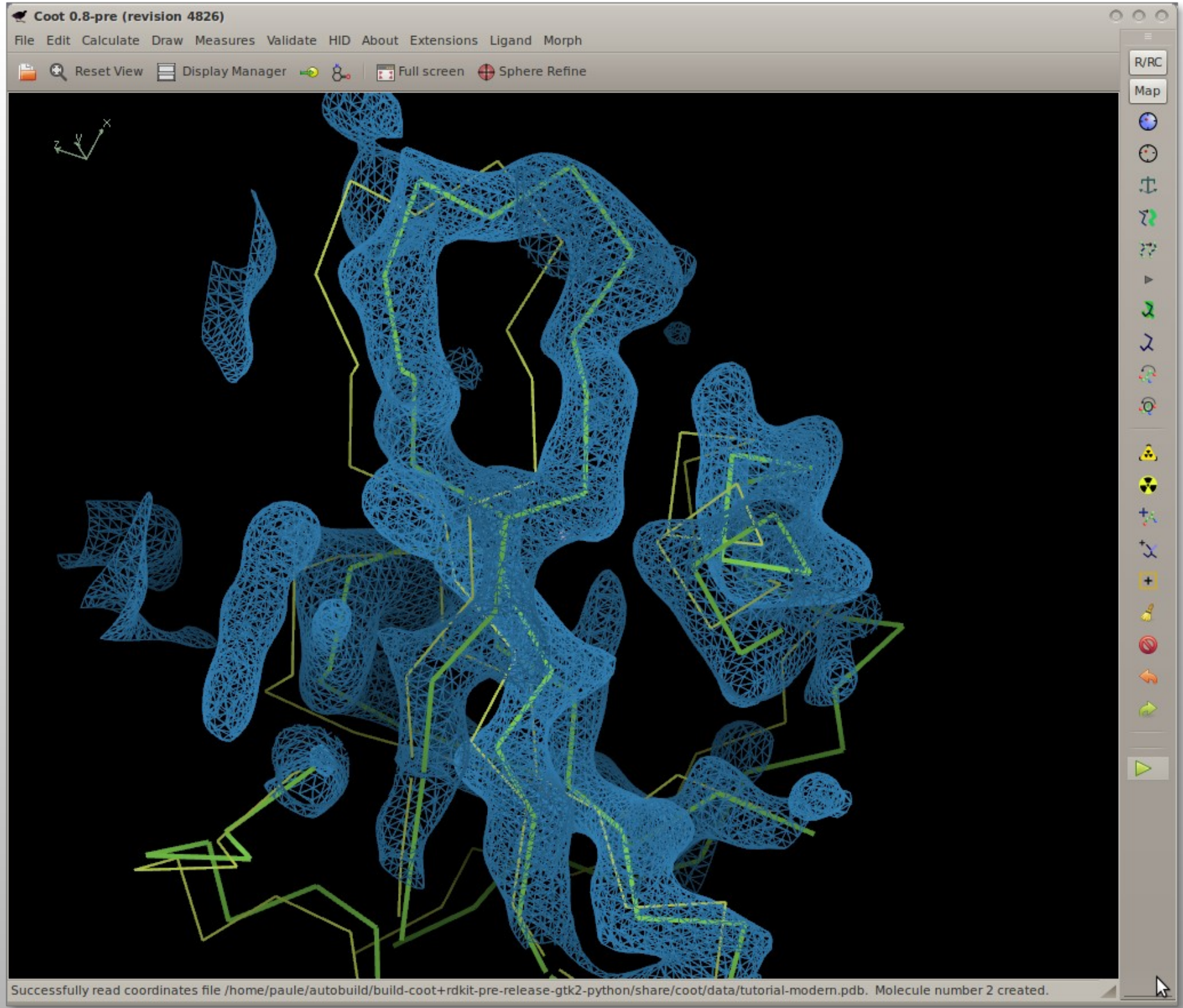
Model Morphing: Robust Averaging

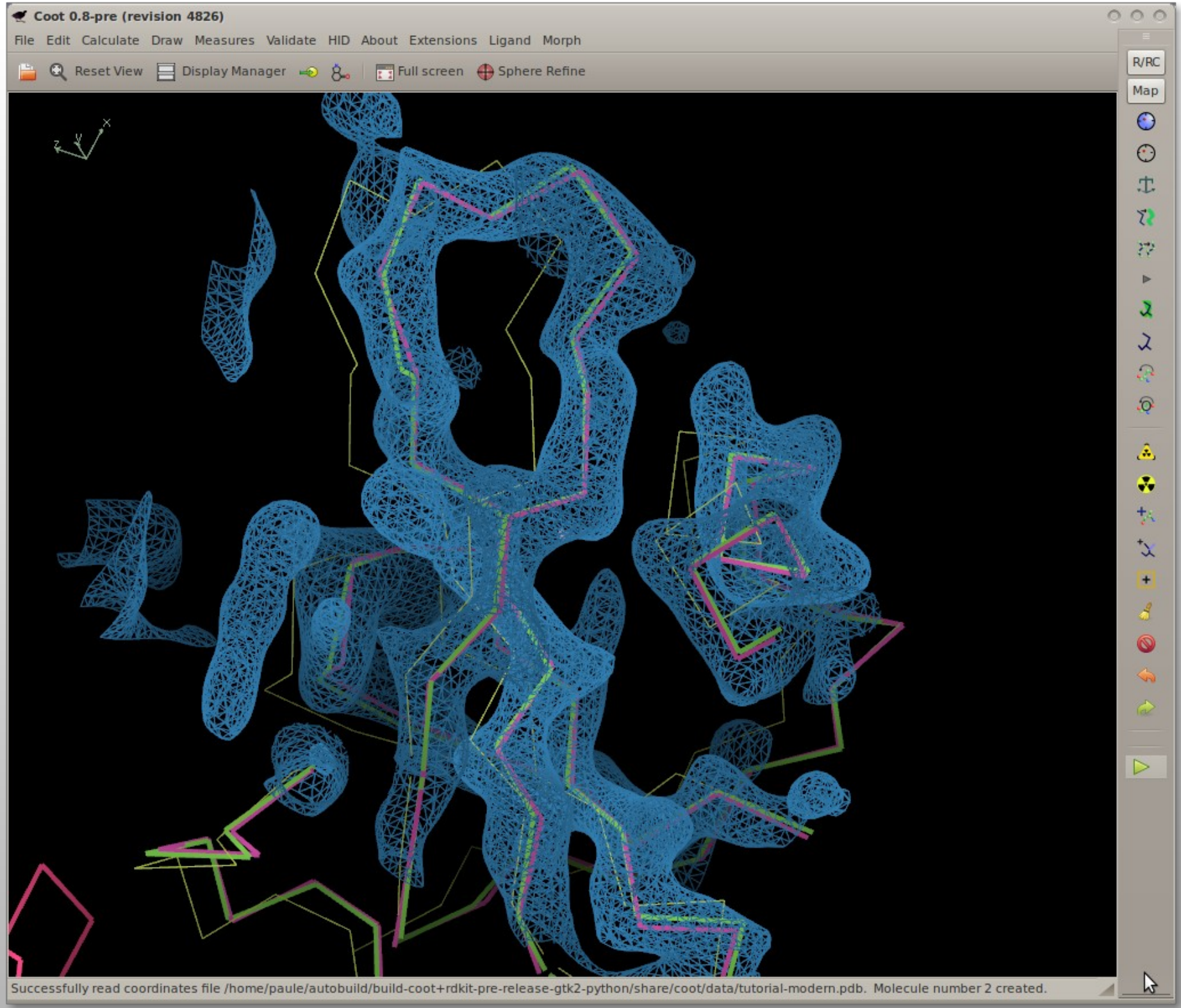
- What are the residues in the environment of a residue?
  - What are their RTs?
  - Create a metric 'distance', sort on that
  - Discard the top and bottom 25%
  - Use remaining RTs to generate average
  - ...which is then applied to central residue
- Repeat for all residues
- Larger environment radii make the shifts smaller/more conservative
  - More cycles needed


Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.


Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.


Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.

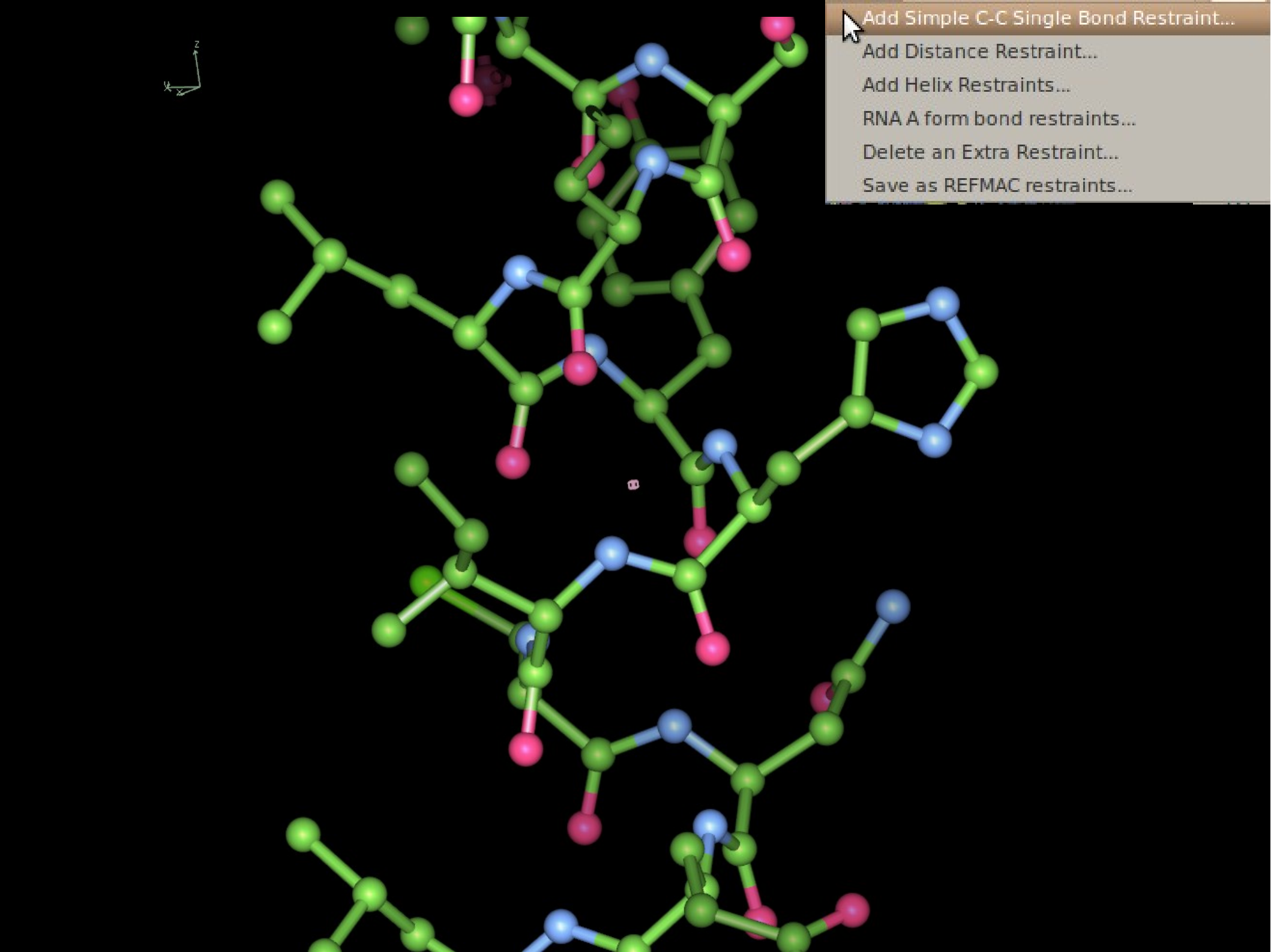

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.

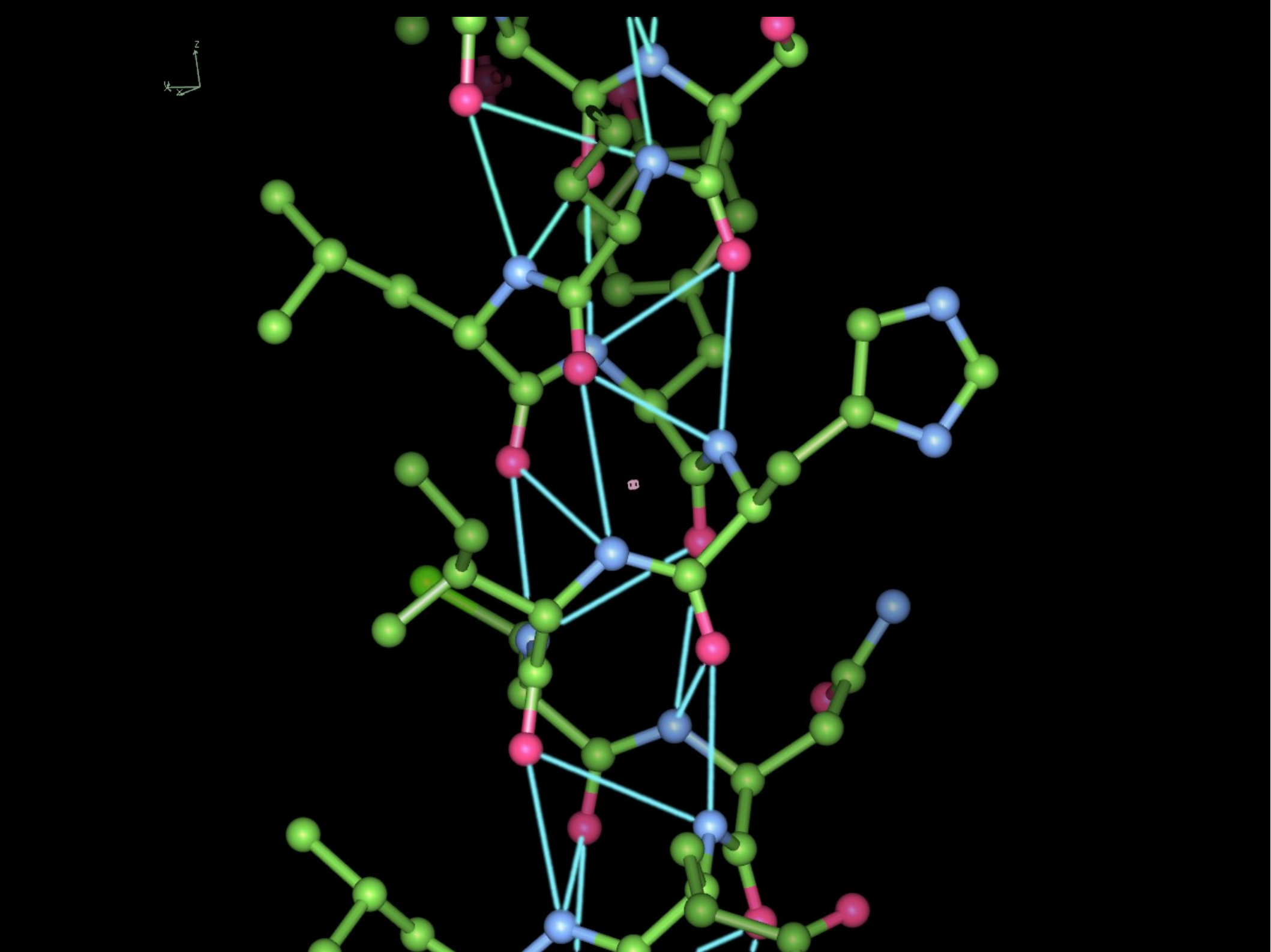

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.

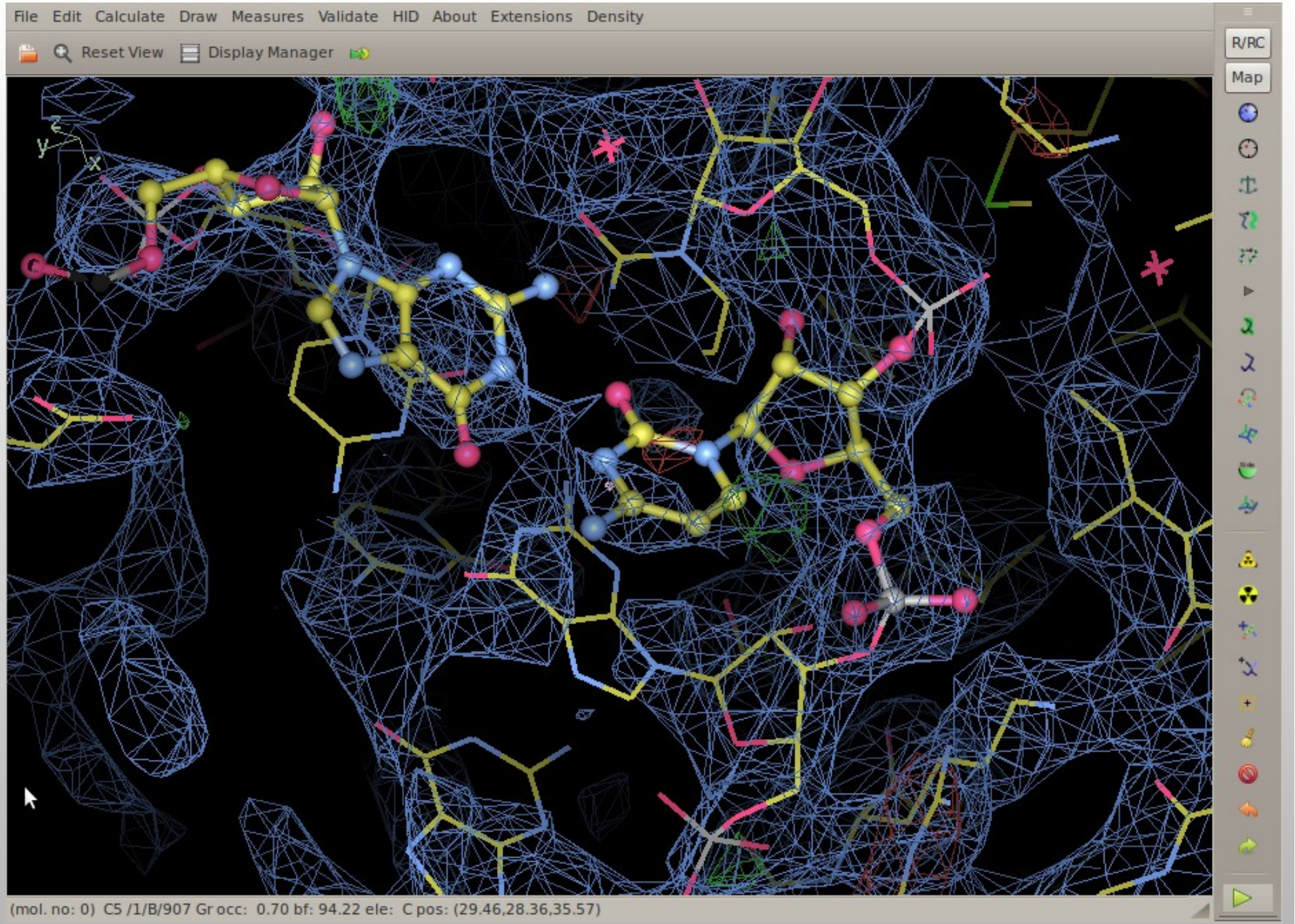
Additional Restraints

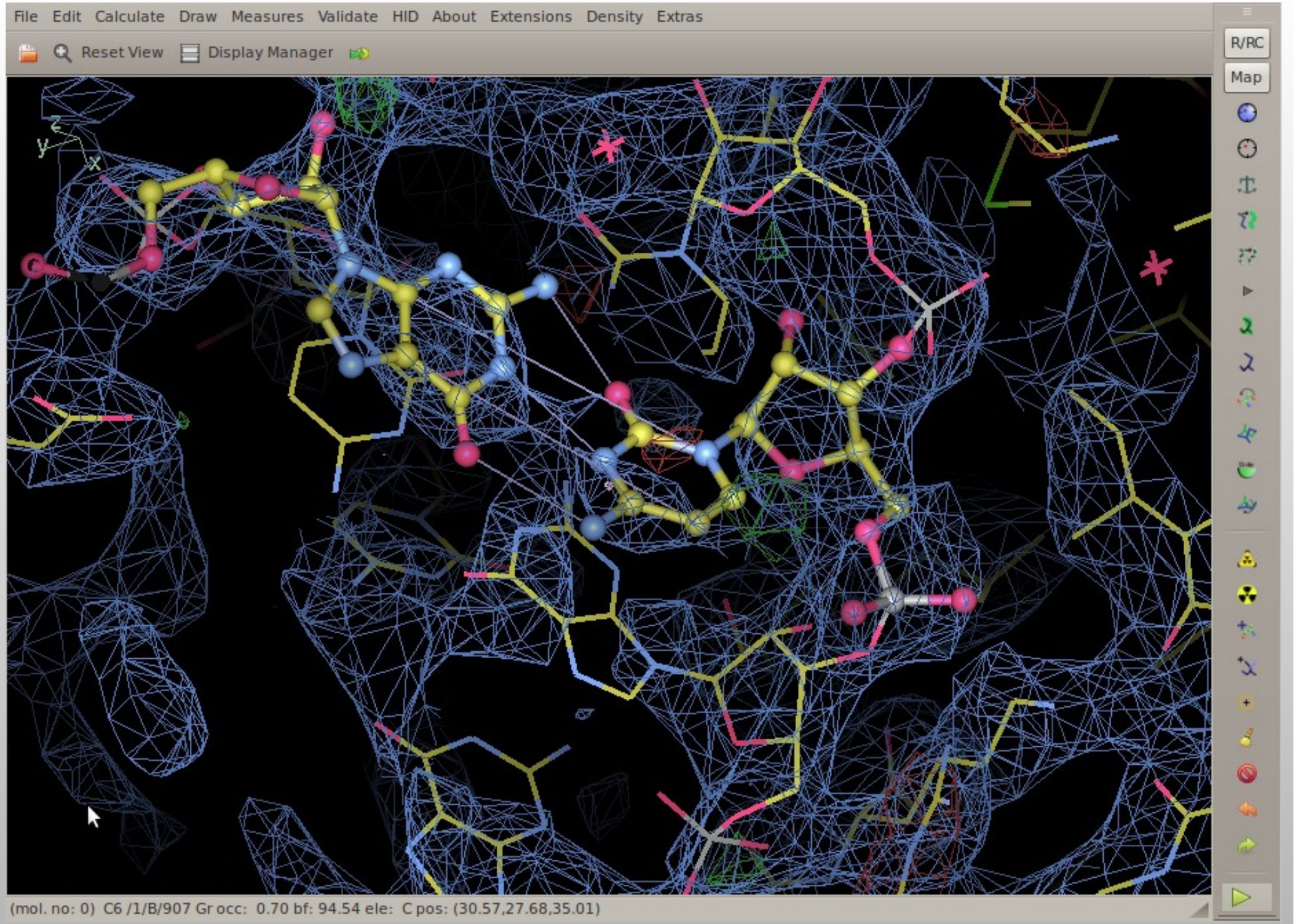
Restraints Editing in *Coot*

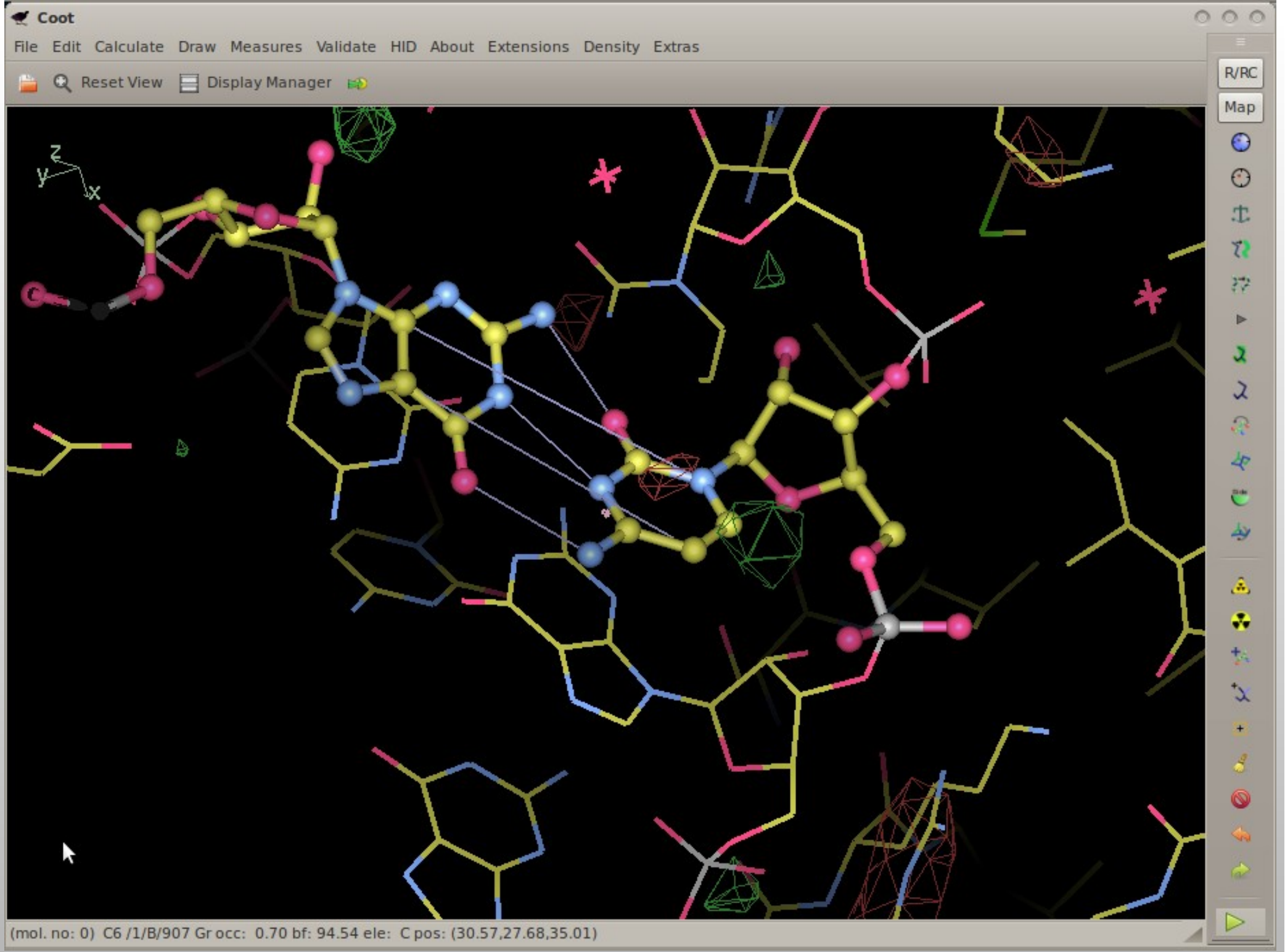
- Distance Restraints:
  - Alpha helices, A-form RNA, B-form DNA
- Add and delete individual restraints
  - User-selectable sigma
- Select 2 residues for range
- User-defined torsion restraints
- Input from ProSMART
- Output to Refmac

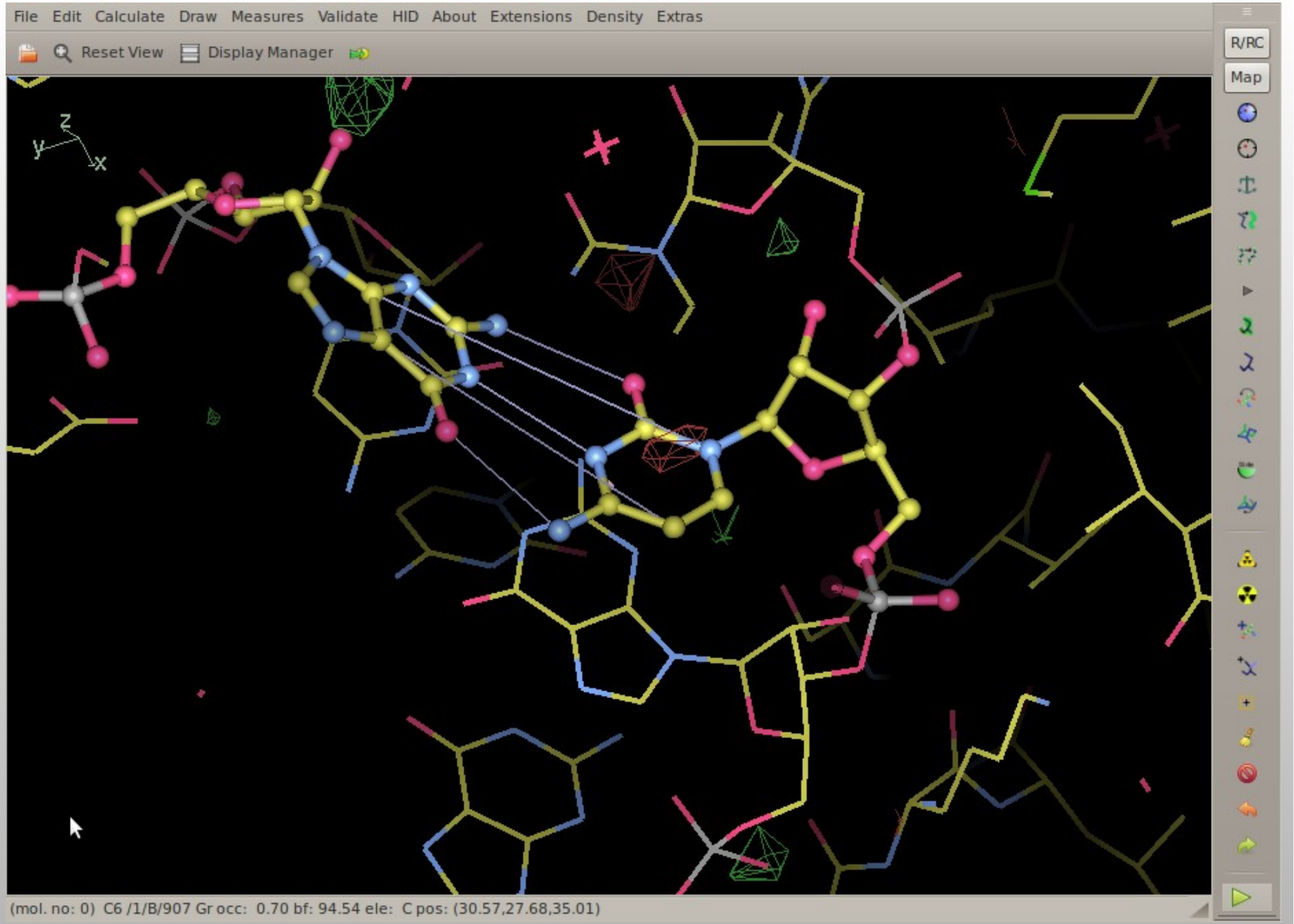

Add Simple C-C Single Bond Restraint...
Add Distance Restraint...
Add Helix Restraints...
RNA A form bond restraints...
Delete an Extra Restraint...
Save as REFMAC restraints...

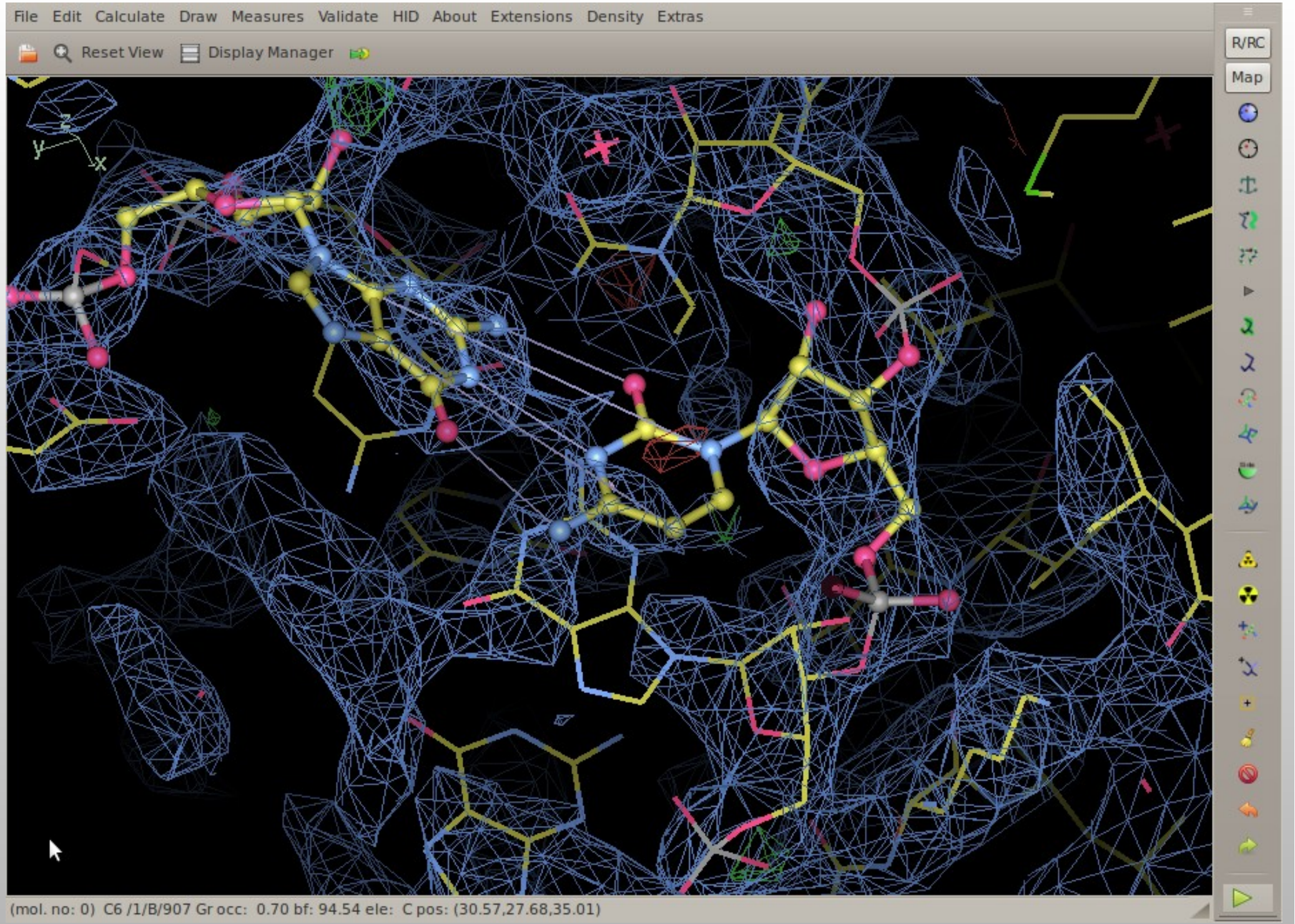



File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
(mol. no: 0) C5 /1/B/907 Gr occ: 0.70 bf: 94.22 ele: C pos: (29.46,28.36,35.57)


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)

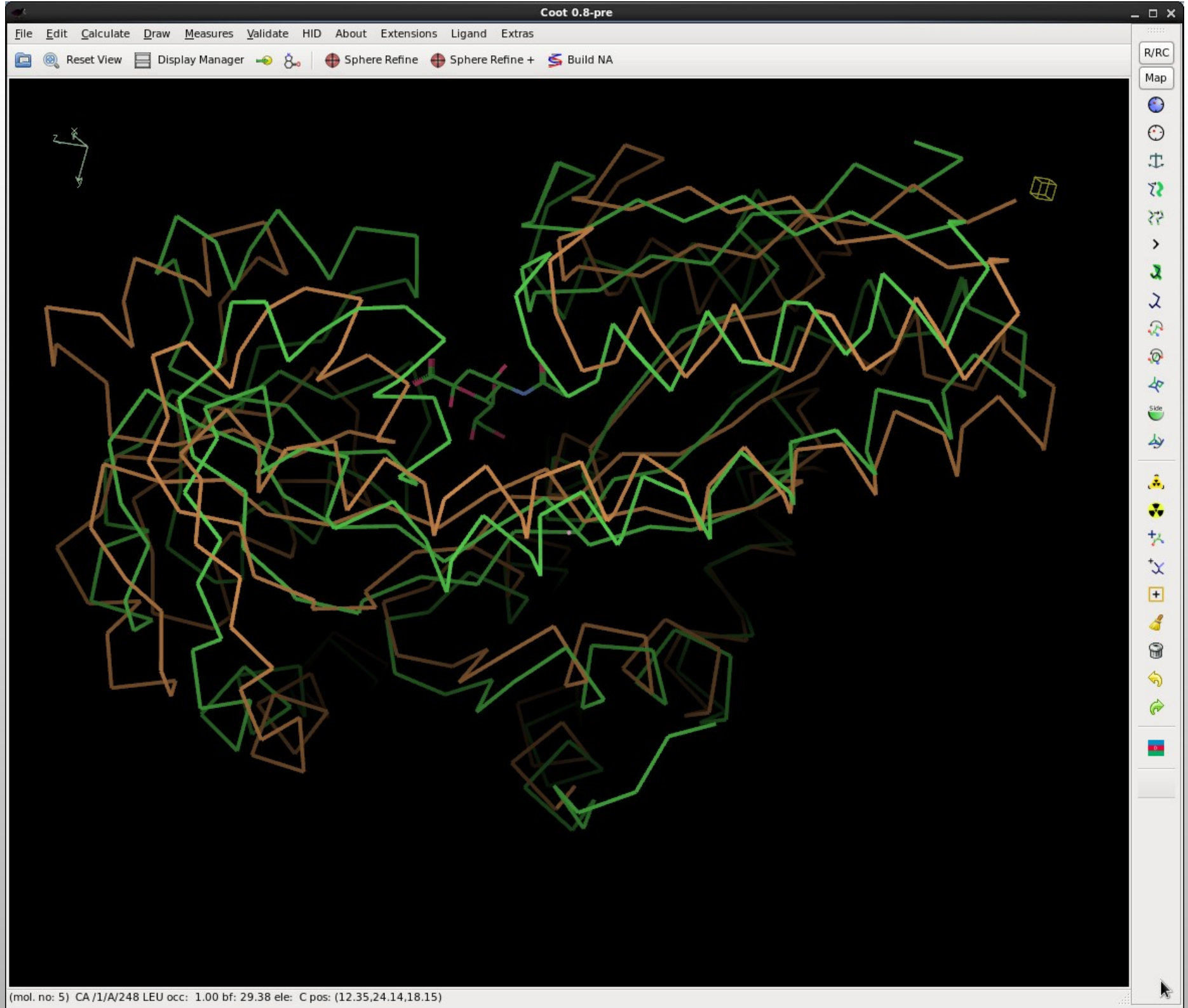
Export as Refmac Restraints:

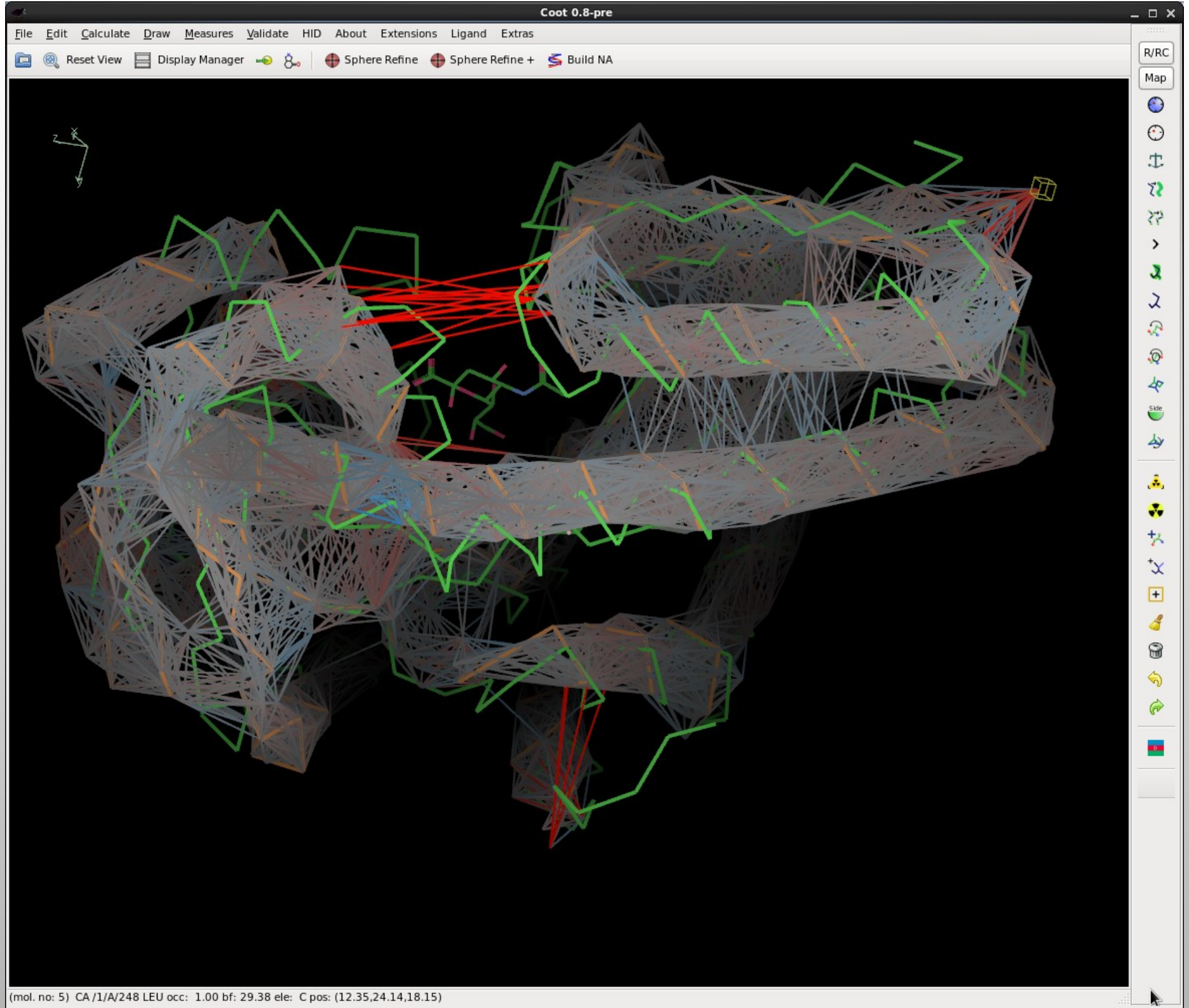
- EXT DIST FIRST CHAIN A RESI 55 INS . ATOM CA   SECOND CHAIN A RESI 55 INS . ATOM C   VALUE 1.54 SIGMA 0.05

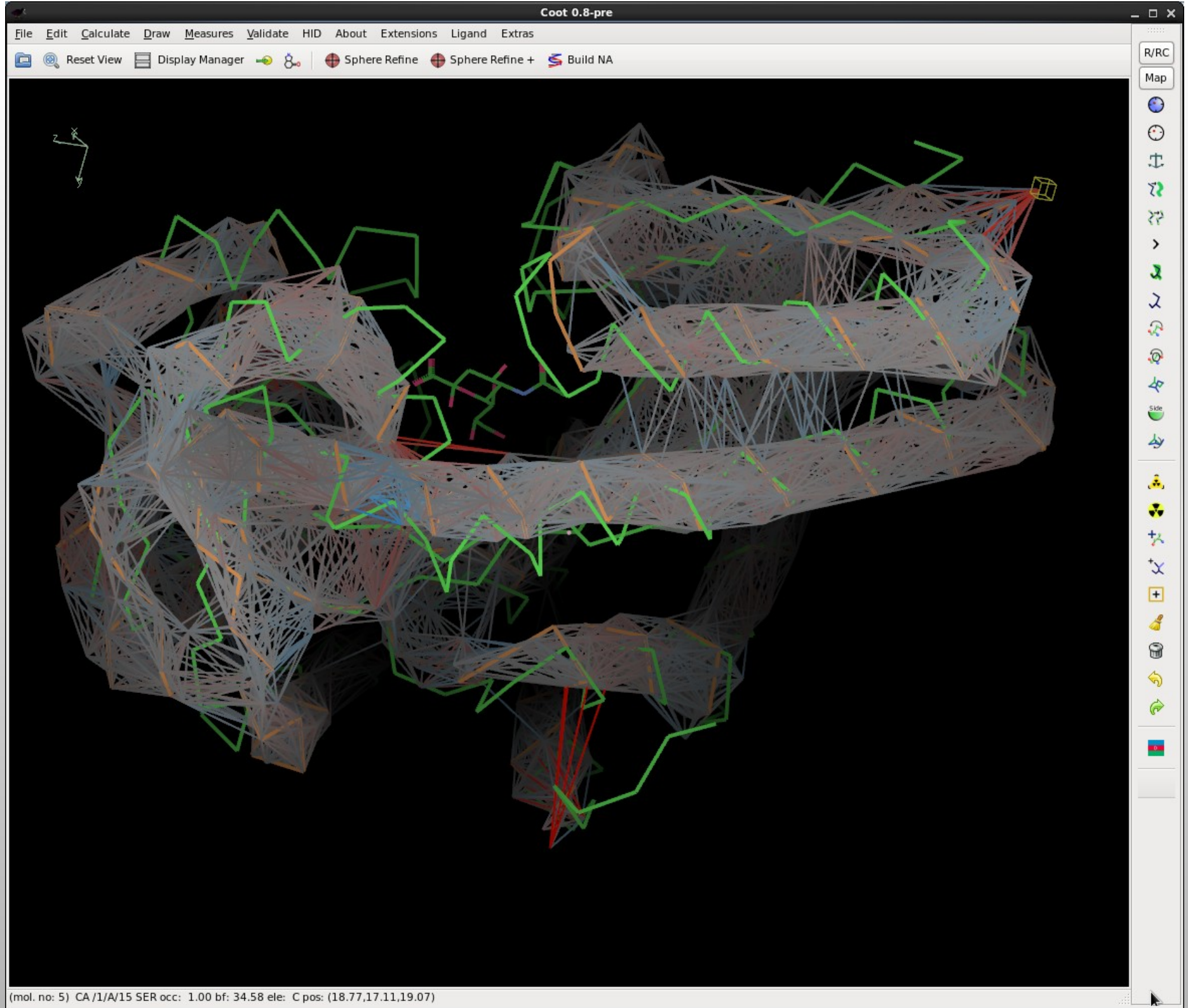
ProSMART Interface

- Use previous-solved “template” structures to inform the refinement of the (low resolution) target protein
- Conformation-independent structural comparison/superposition
- and restraint generation


Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Extras
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 5) CA /1/A/248 LEU occ: 1.00 bf: 29.38 ele: C pos: (12.35,24.14,18.15)


Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Extras
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
Side
(mol. no: 5) CA /1/A/248 LEU occ: 1.00 bf: 29.38 ele: C pos: (12.35,24.14,18.15)

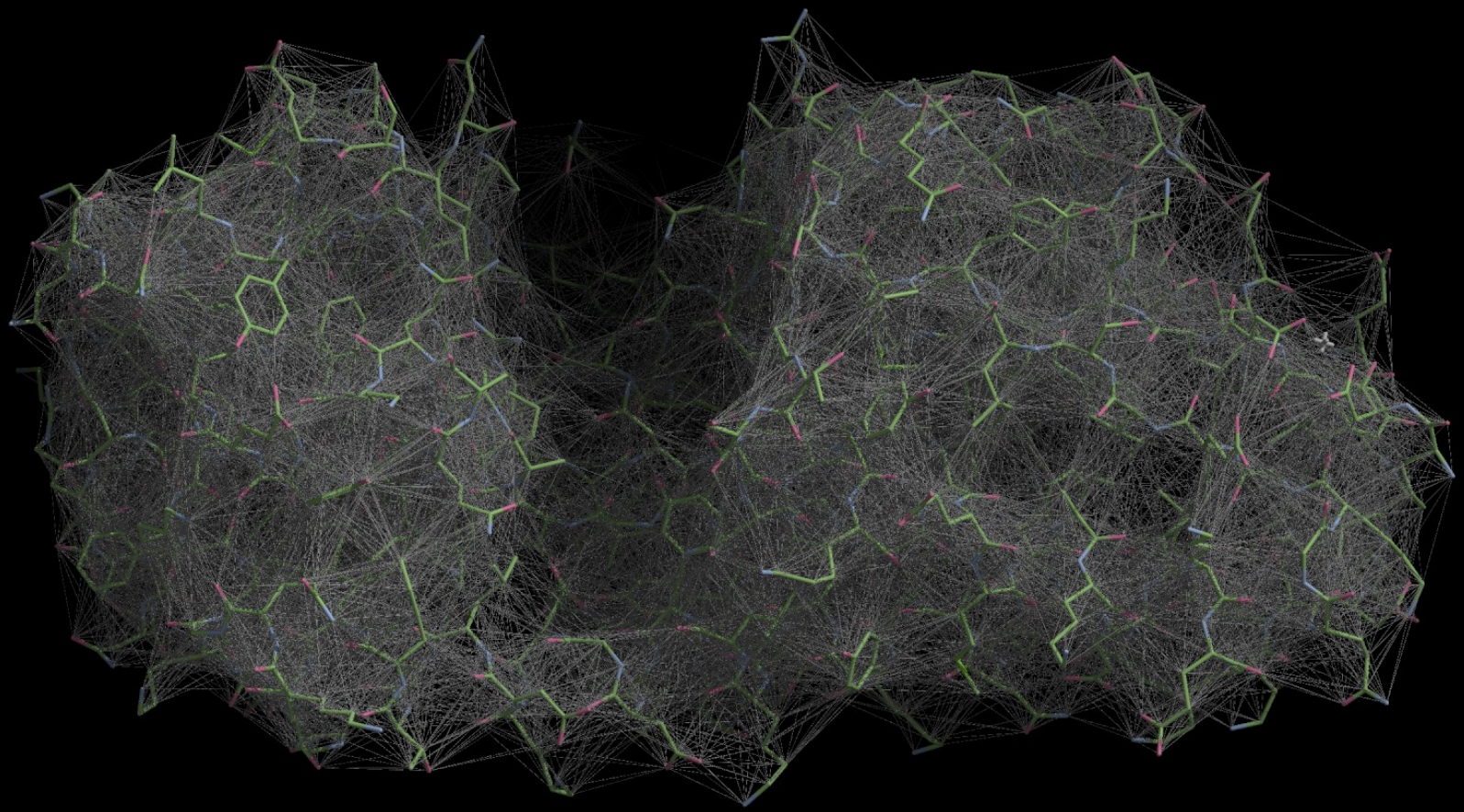

Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Extras
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 5) CA /1/A/248 LEU occ: 1.00 bf: 29.38 ele: C pos: (12.35,24.14,18.15)


Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Extras
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 5) CA /1/A/15 SER occ: 1.00 bf: 34.58 ele: C pos: (18.77,17.11,19.07)

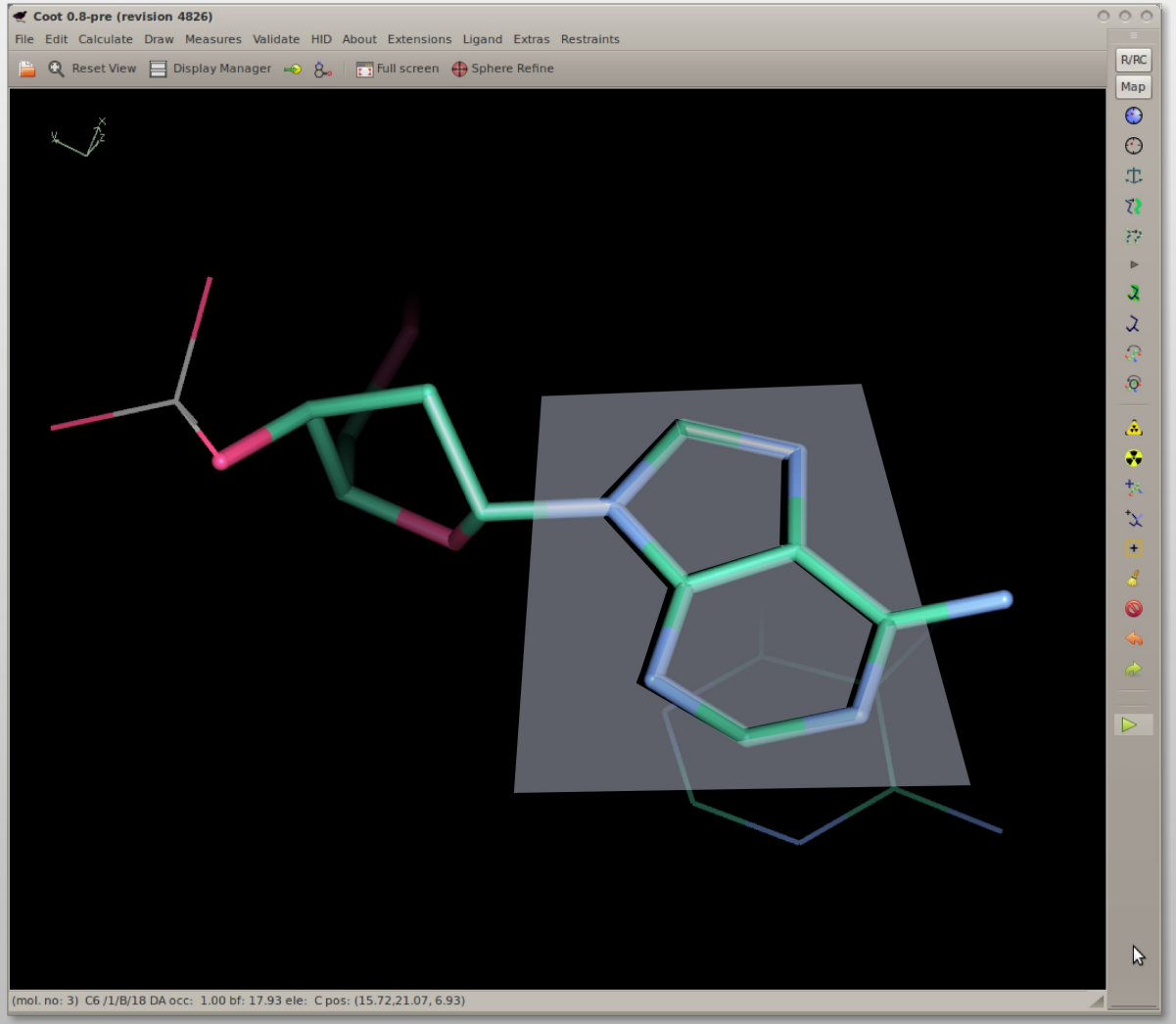
ProSMART integration

- ProSMART generates distance restraints from homologous structures
  - to be applied to current model for refinement
  - now available in *Coot*

ProSMART Restraints

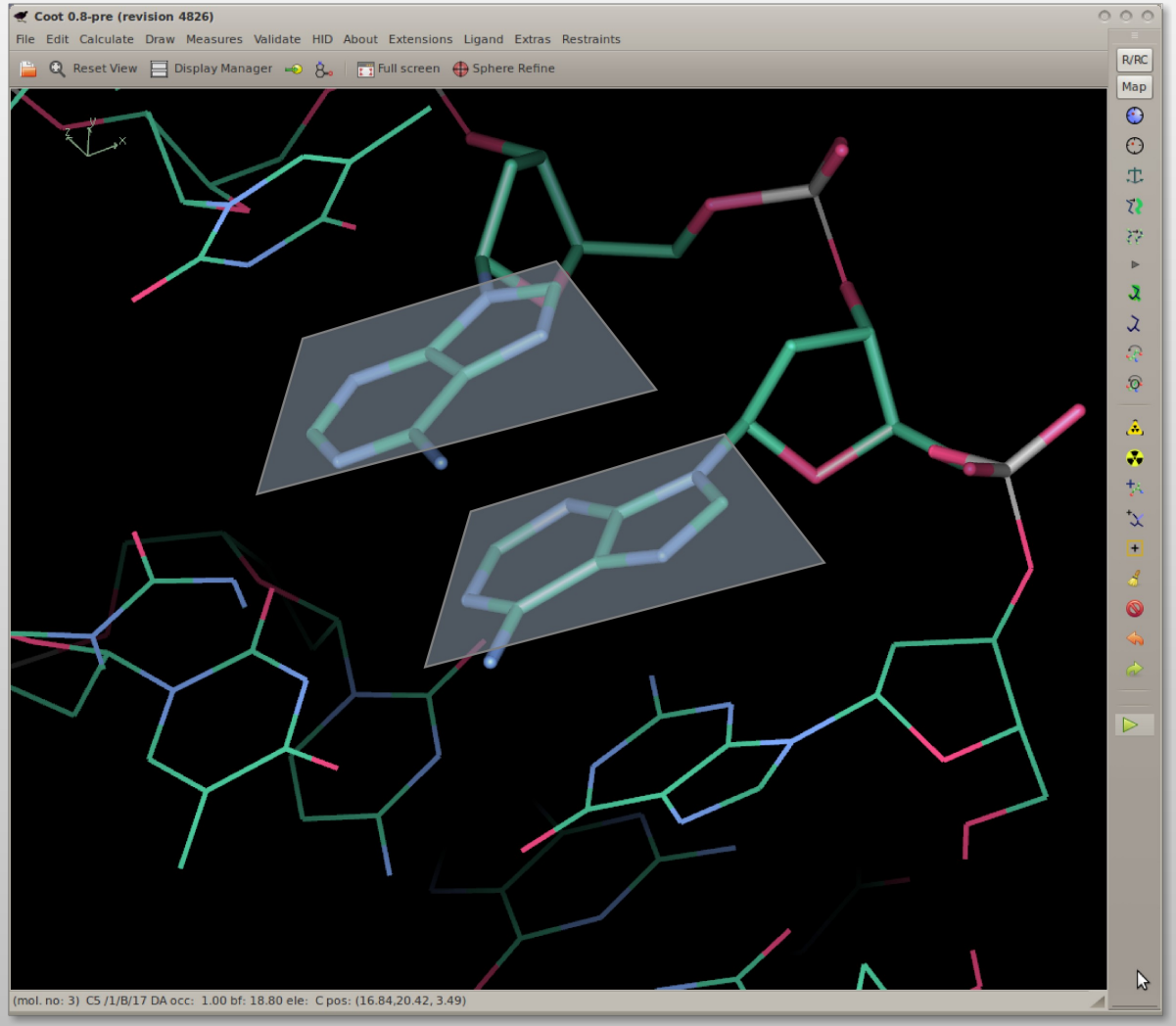


Plane Restraints



Derivativaties are an eigenvector scaled by out-of-plane distance

Parallel Planes Restraints



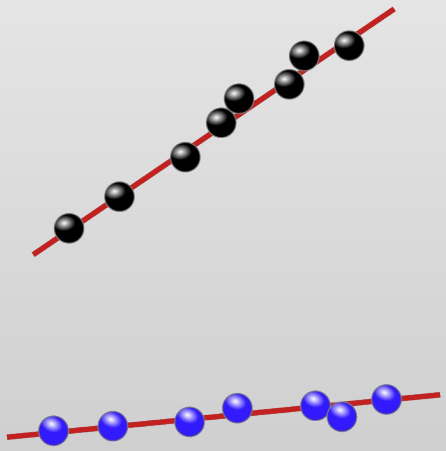
$$S = (a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2$$

Not easy to use in Coot

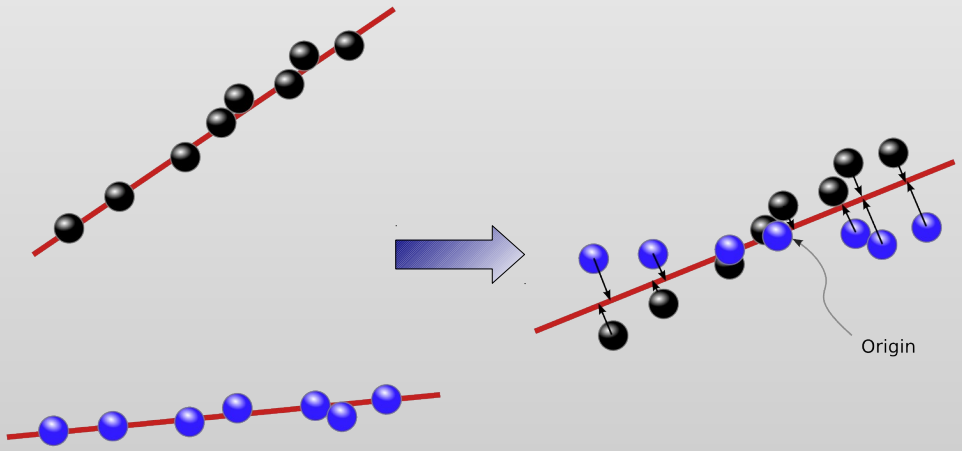
Parallel Planes Restraints

- Also, we have considered parallel-planes distance restraints
  - More tricky still to implement
  - Not implemented yet (not in *Coot*, anyway)

Parallel Planes Restraints

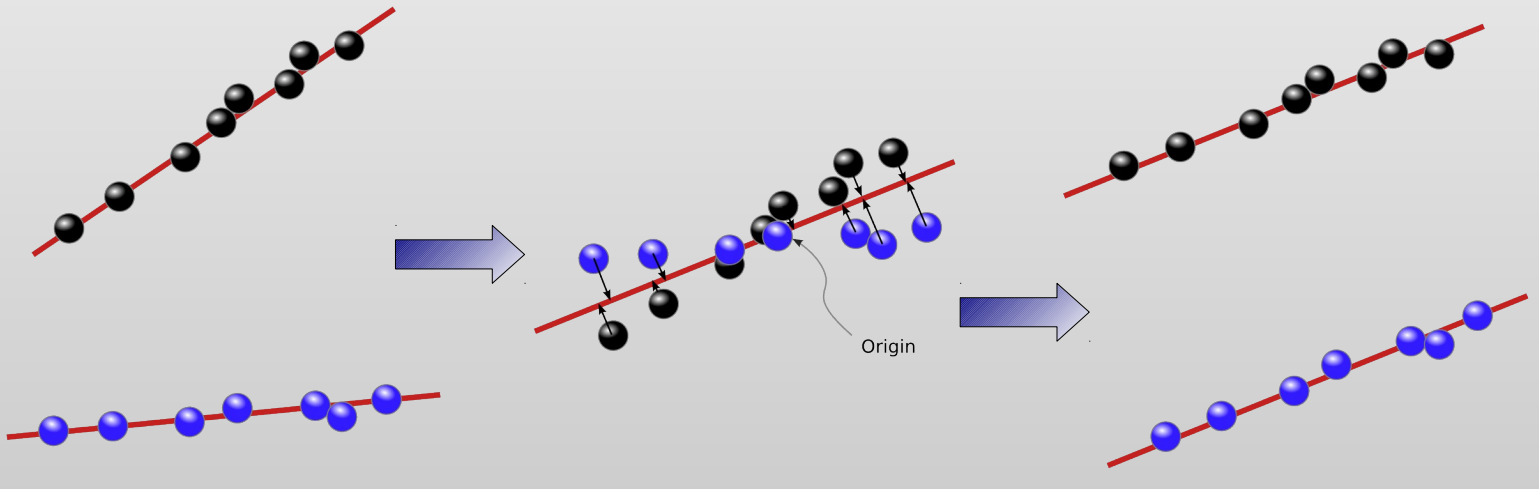


Parallel Plane Restraints



Shift to Origin

Parallel Planes Restraints

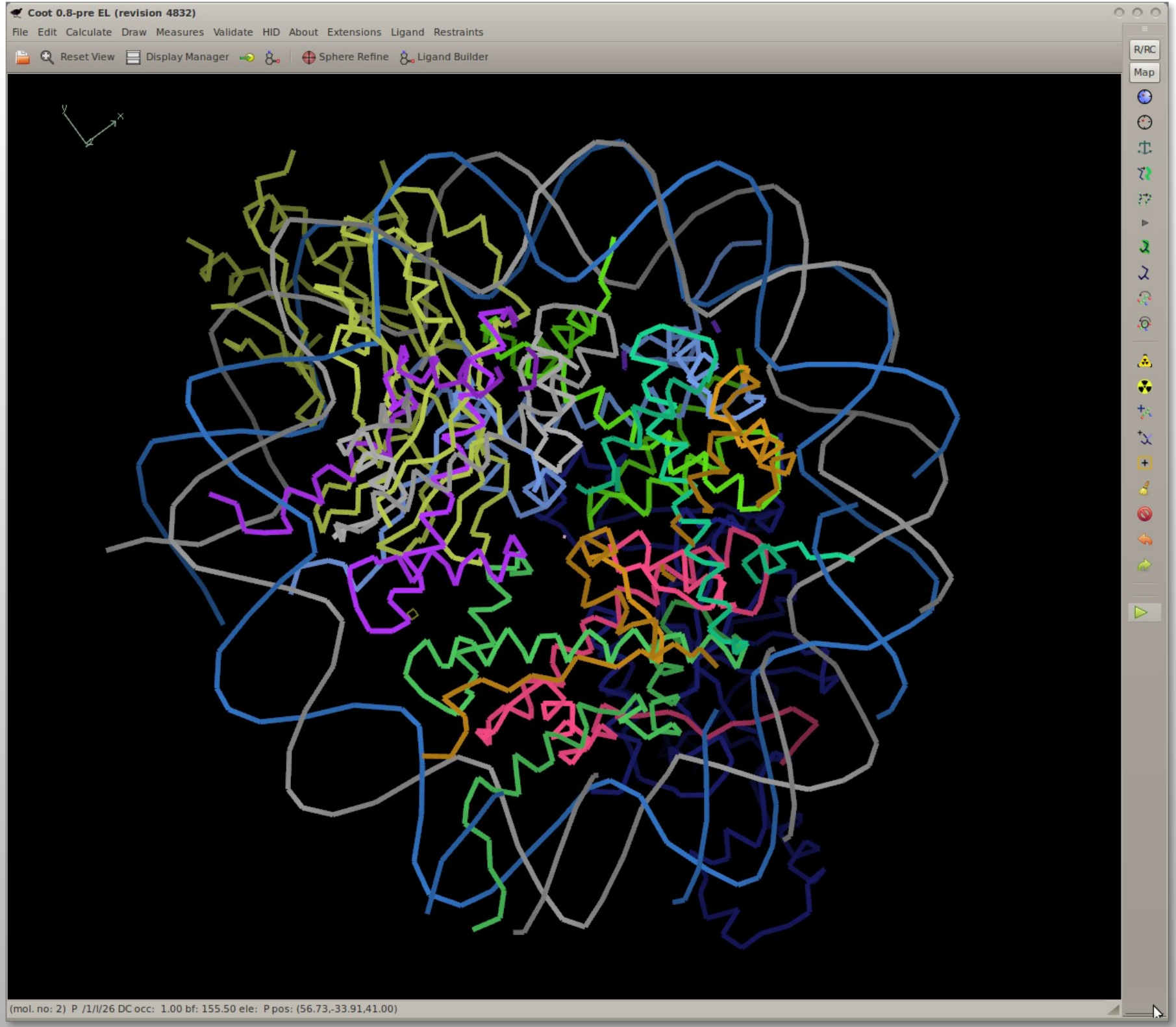


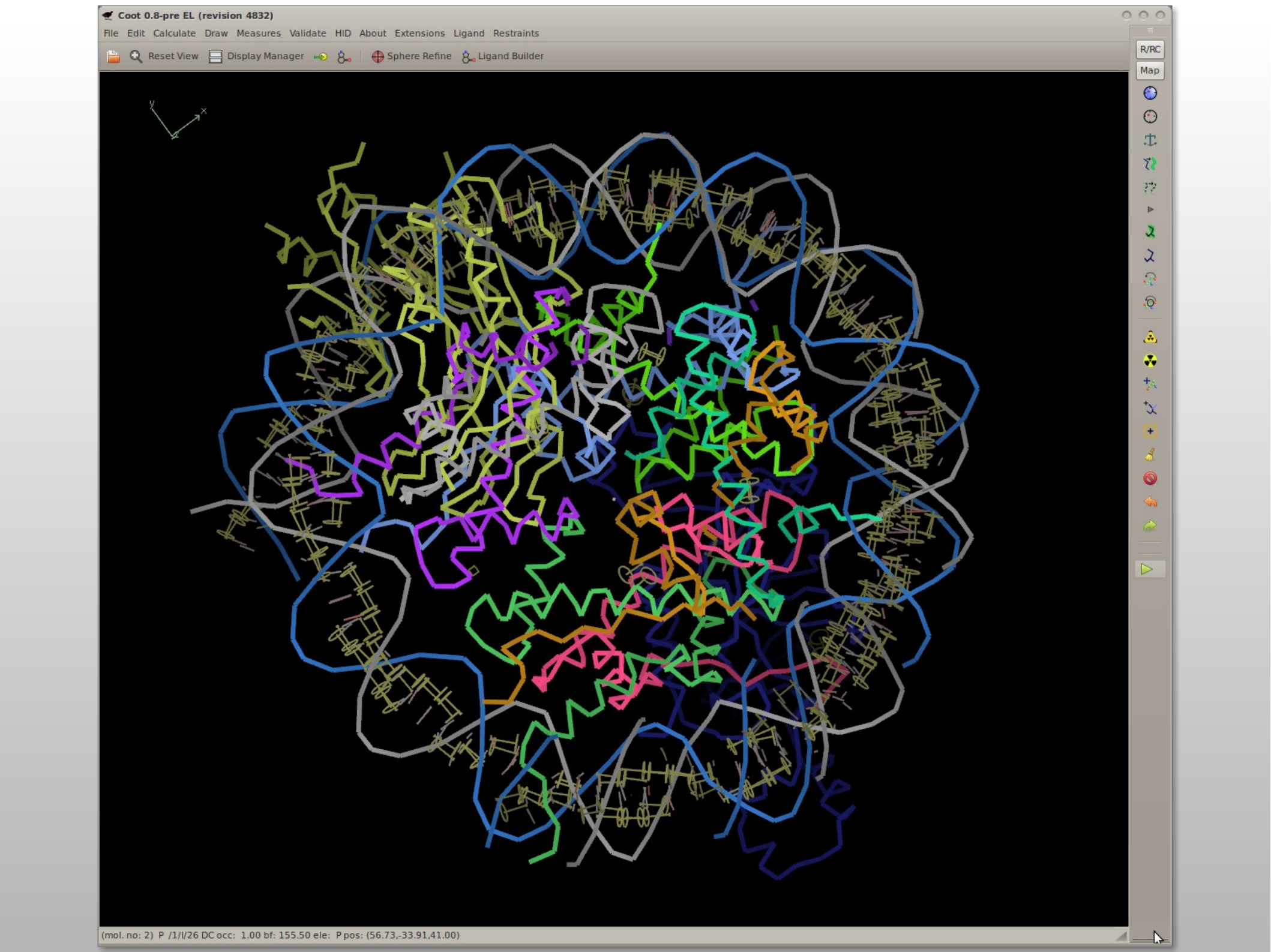
Shift to Origin

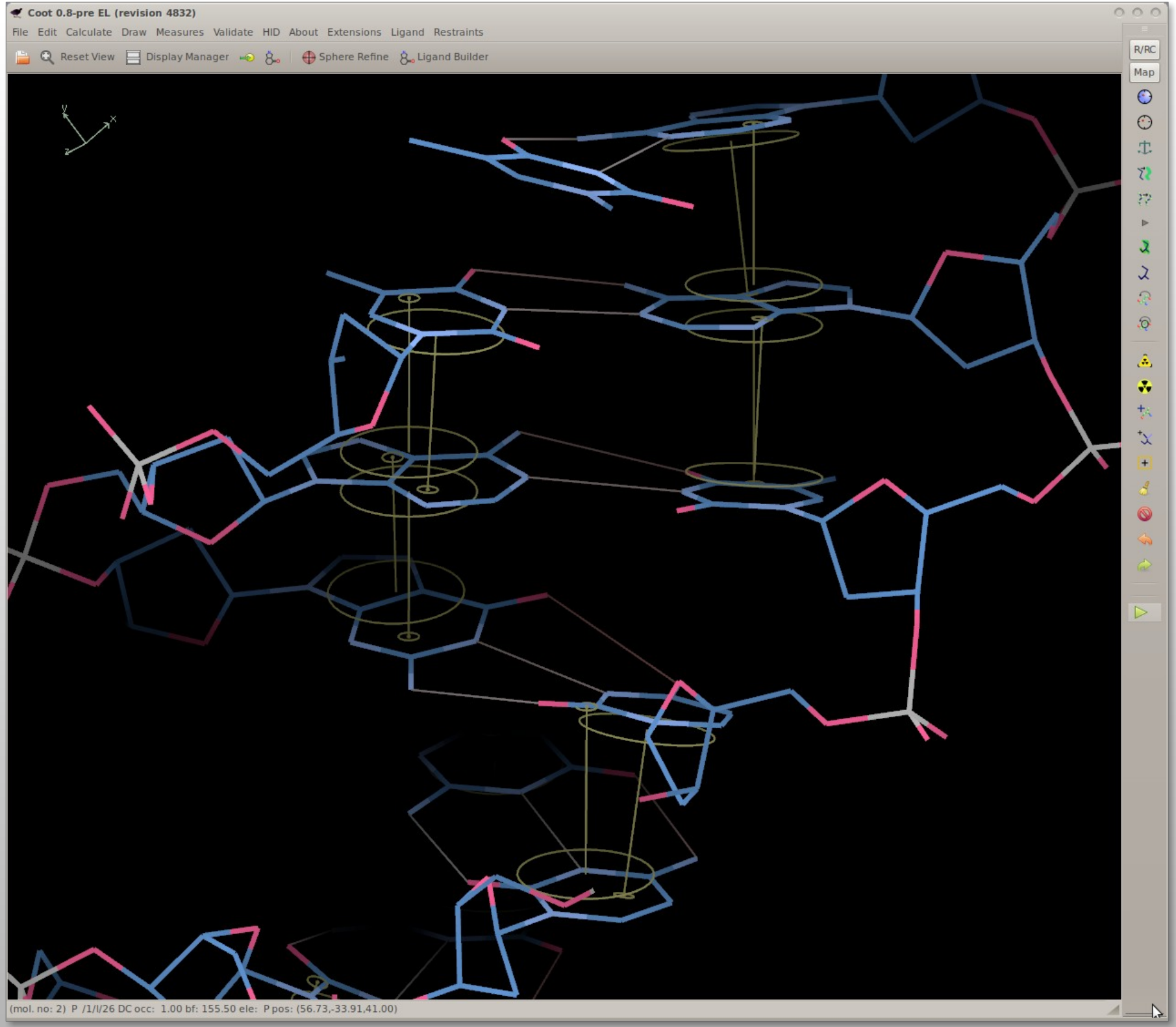
Move Back to Molecule

Automatic Generation of Base Pairing and Stacking Restraints

- Fei Long's `libg_d`
  - Provide it with a model and it writes out Refmac restraint descriptions
  - ... which *Coot* can also read
  - *Coot* can also create user-define base-paring and stacking restraints

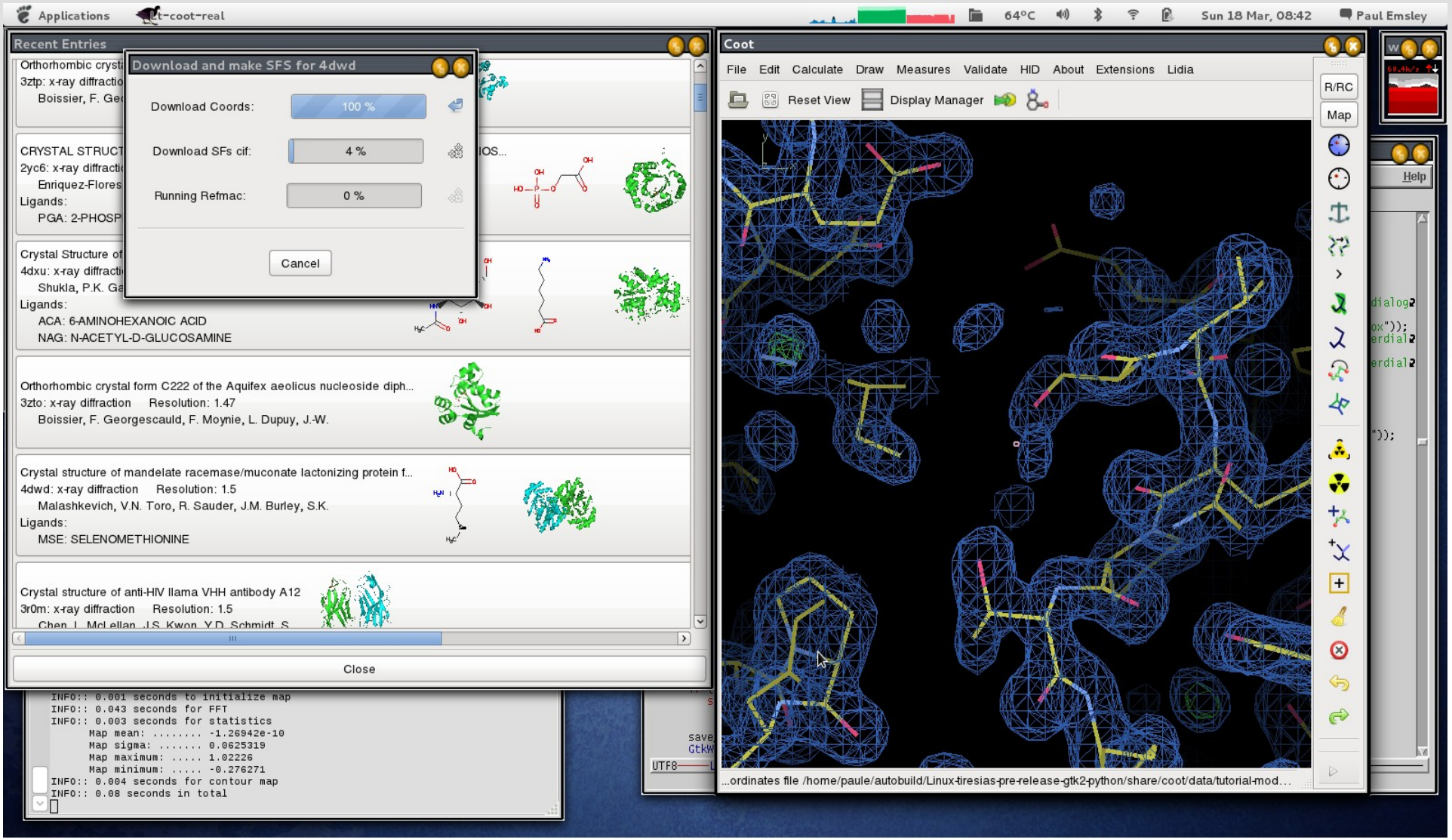

Coot 0.8-pre EL (revision 4832)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Restraints
Reset View
Display Manager
Sphere Refine
Ligand Builder
R/RC
Map
(mol. no: 2) P /1/I/26 DC occ: 1.00 bf: 155.50 ele: P pos: (56.73,-33.91,41.00)


Coot 0.8-pre EL (revision 4832)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Restraints
Reset View
Display Manager
Sphere Refine
Ligand Builder
R/RC
Map
(mol. no: 2) P /1/I/26 DC occ: 1.00 bf: 155.50 ele: P pos: (56.73,-33.91,41.00)


Coot 0.8-pre EL (revision 4832)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Restraints
Reset View
Display Manager
Sphere Refine
Ligand Builder
R/RC
Map
(mol. no: 2) P /1/I/26 DC occ: 1.00 bf: 155.50 ele: P pos: (56.73,-33.91,41.00)

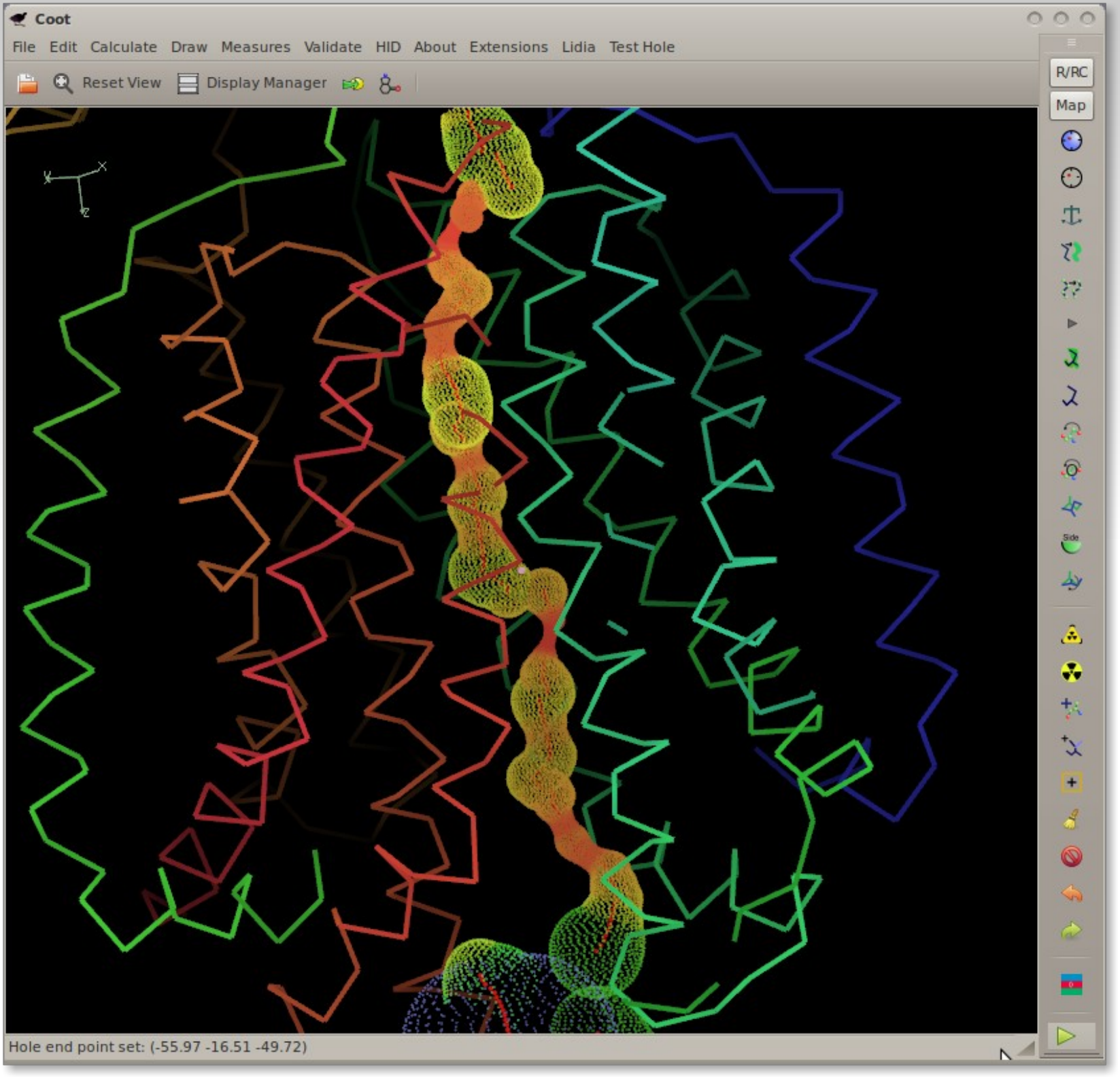
PDBe Recent Structures

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
Side
Hole end point set: (-55.97 -16.51 -49.72)

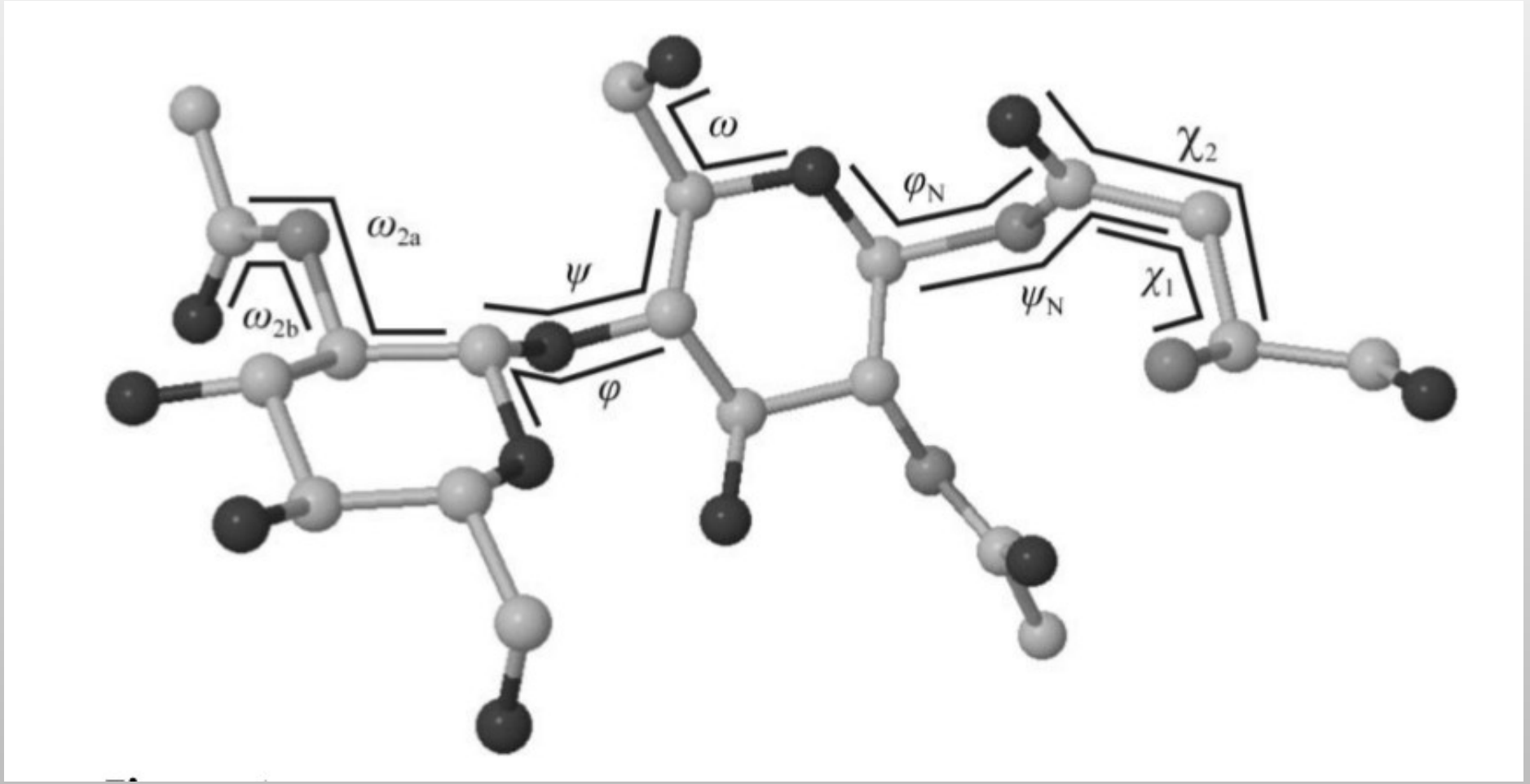
Modelling Carbohydrates

- Validation,
- Model-building,
- Refinement

Problematic Glycoproteins

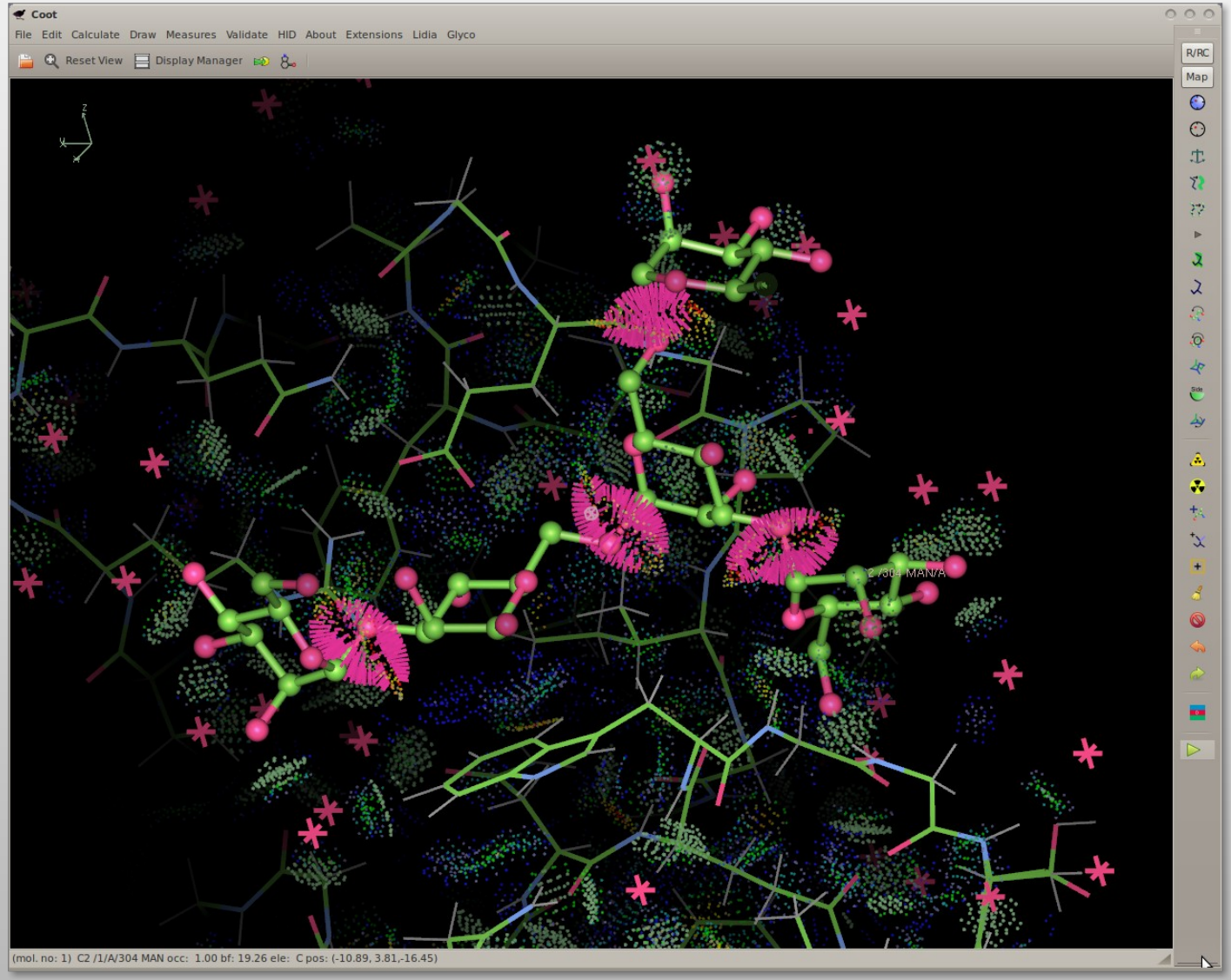
- Crispin, Stuart & Jones (2007)
  - NSB Correspondence
  - “one third of entries contain significant errors in carbohydrate stereochemistry...”
  - “carbohydrate-specific building and validation tools capable of guiding and construction of biologically relevant stereochemically accurate models should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

Carbohydrate Links

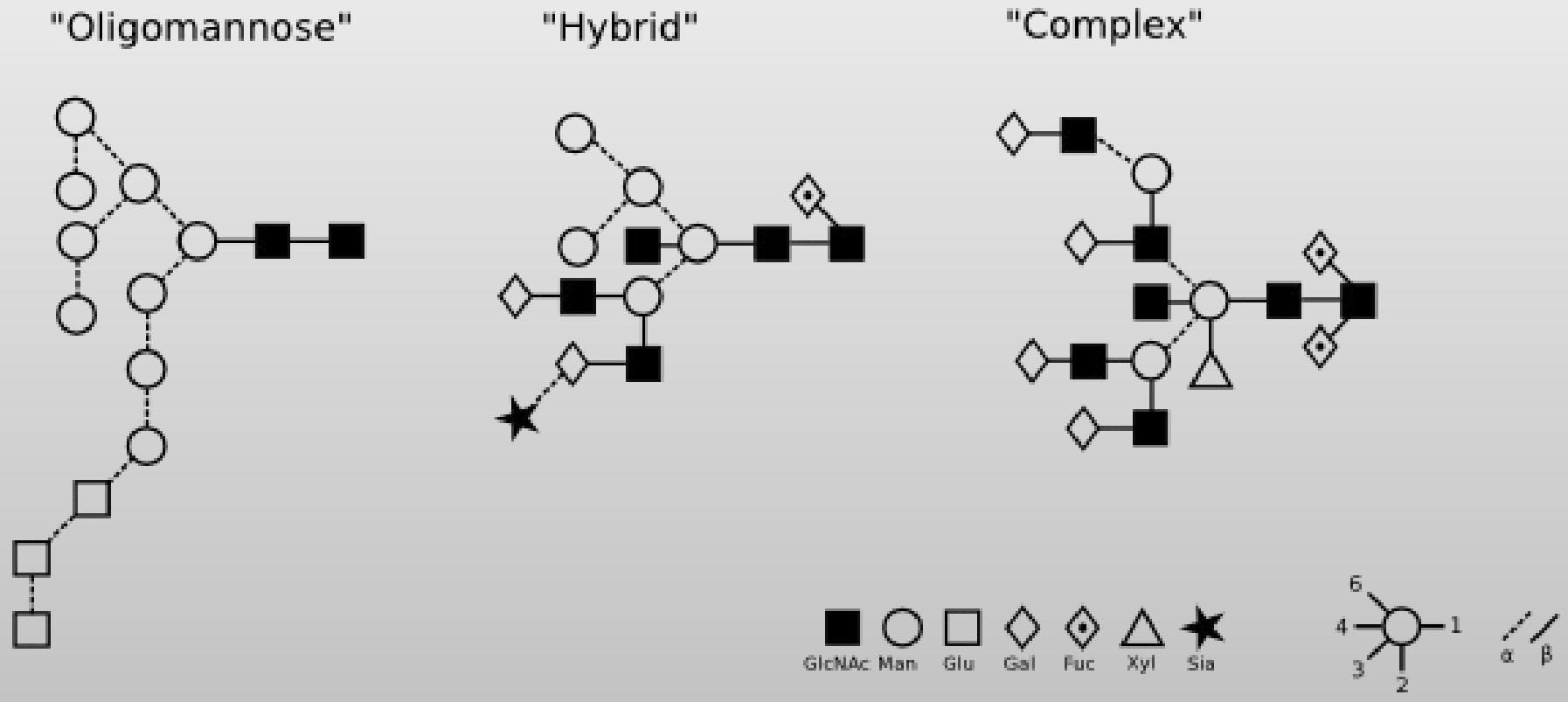


Thomas Lütteke (2007)

Validate the Links



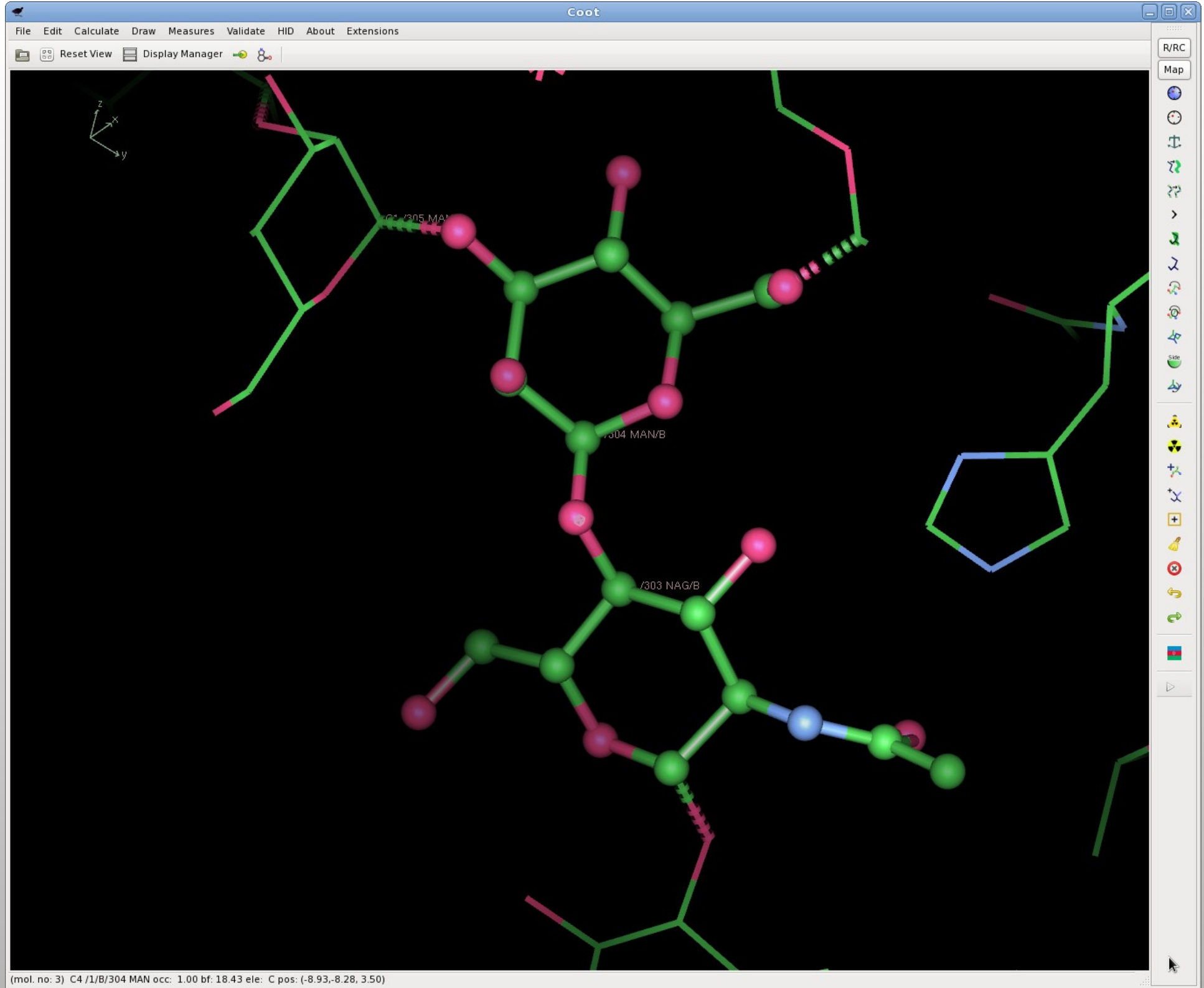
Validate the Tree: N-linked carbohydrates

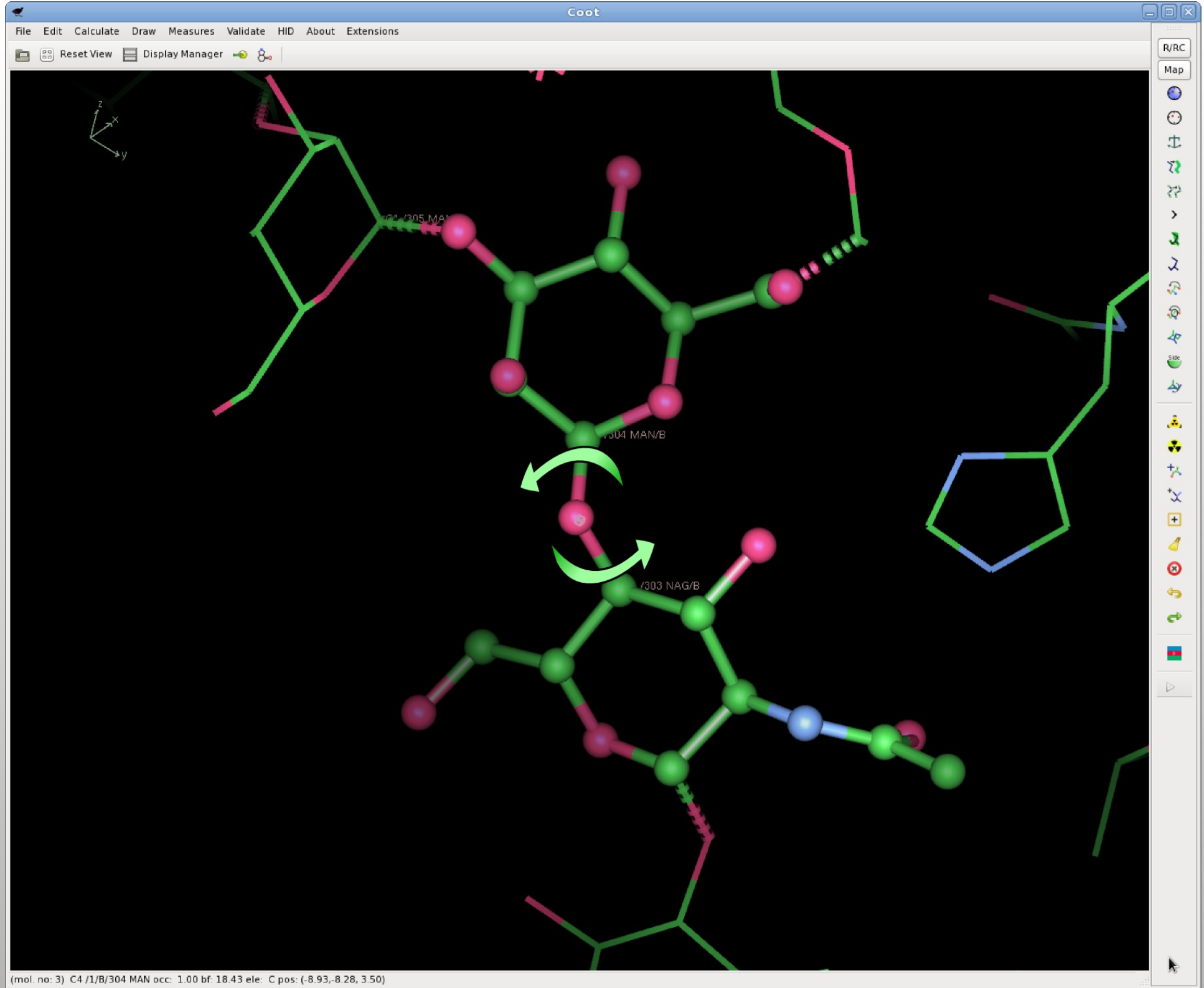


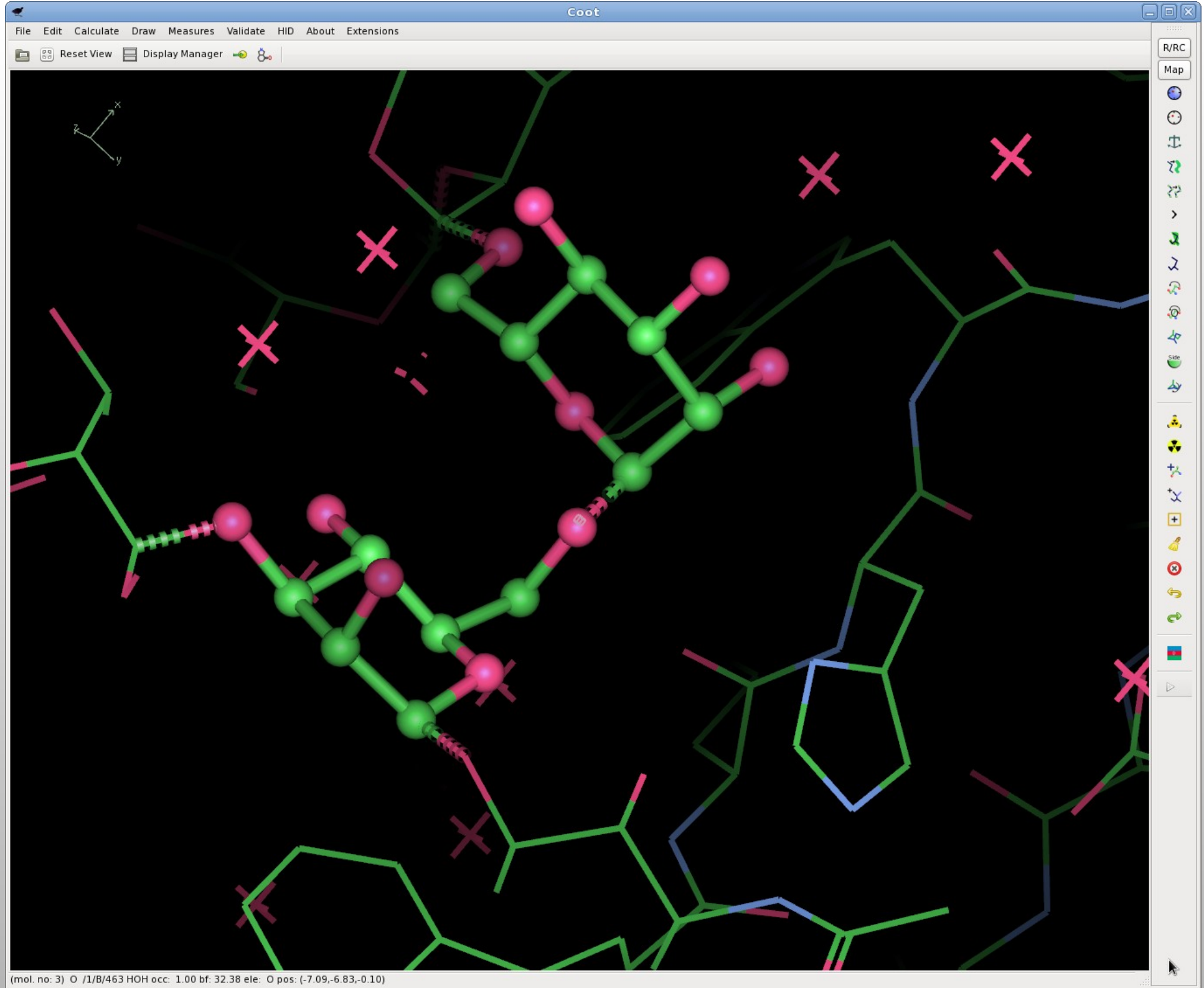
Linking Oligsaccharides/Carbohydrates:

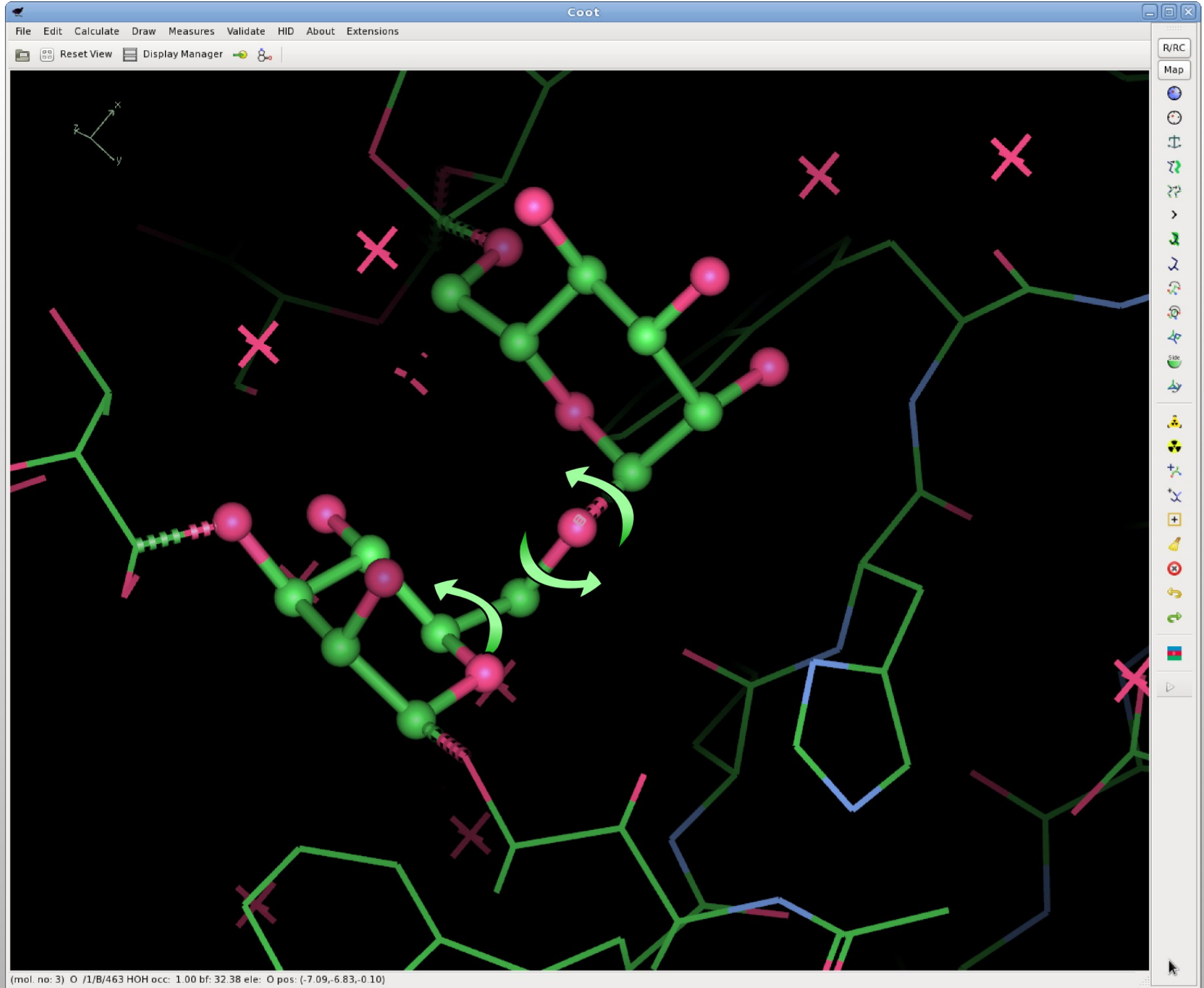
LO/Carb

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

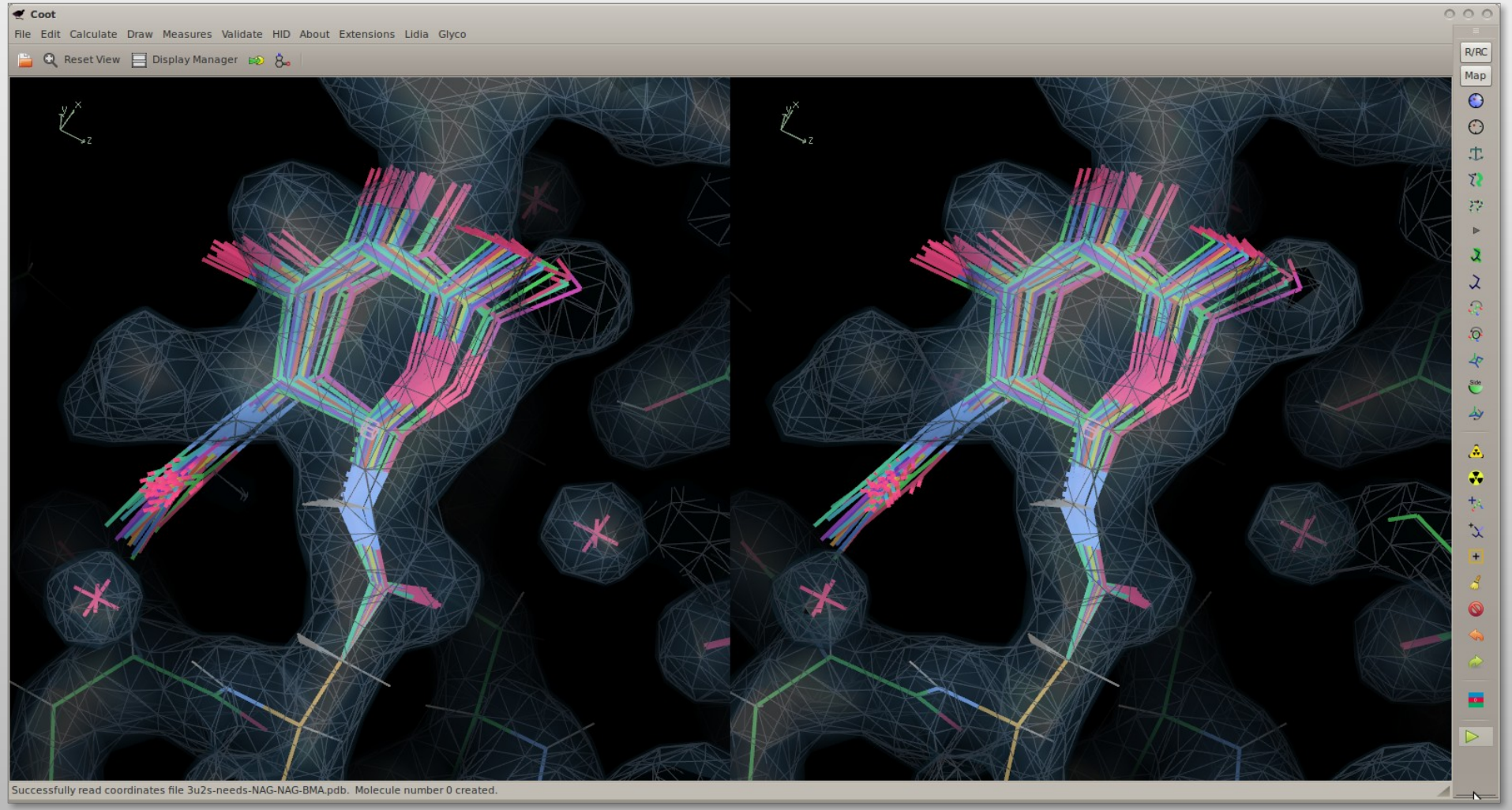

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

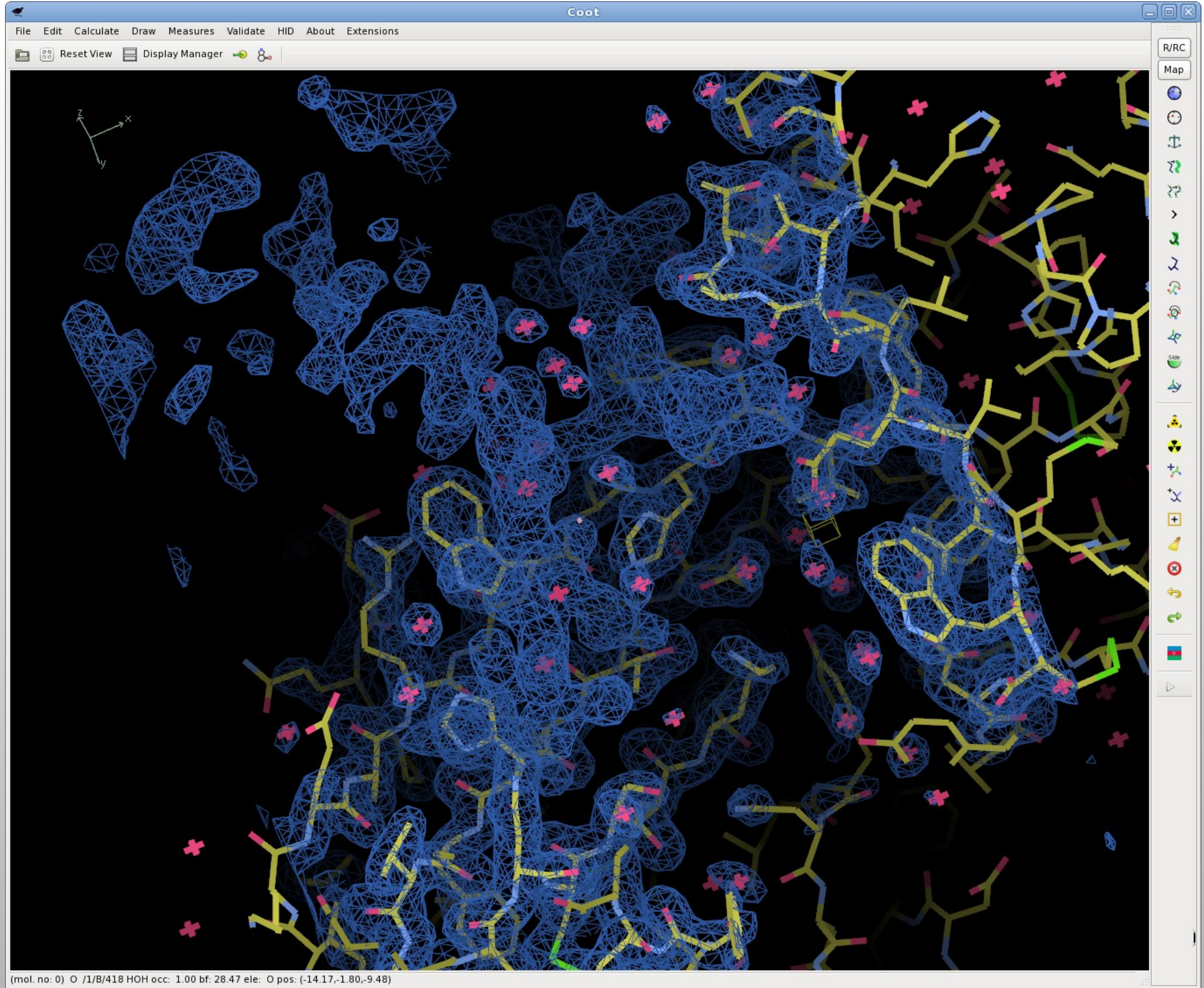

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)

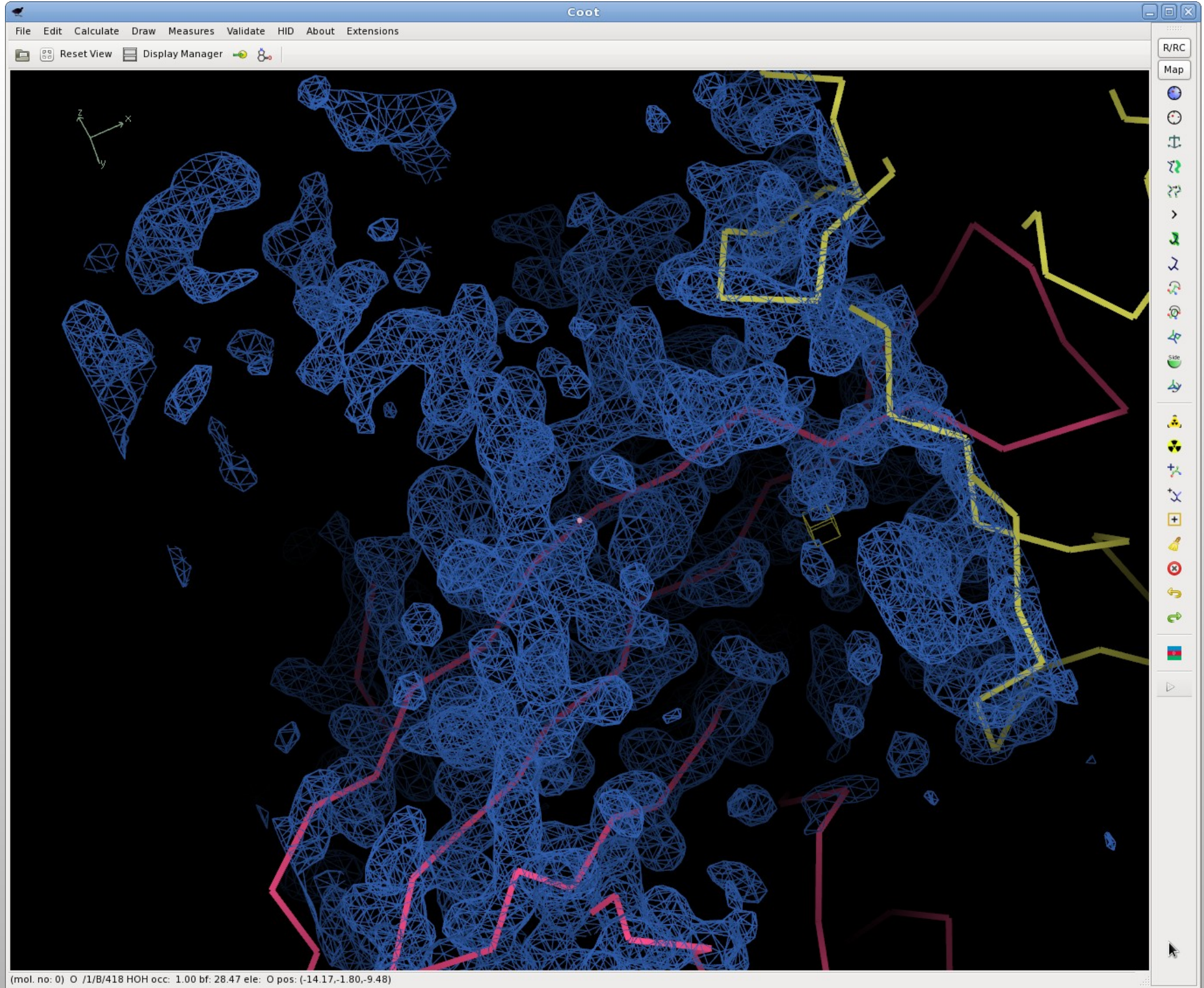

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 3) O /1/B/463 HOH occ: 1.00 bf: 32.38 ele: O pos: (-7.09,-6.83,-0.10)


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

Refinement Trials (NAG-ASN example)




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


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

What is Non-Crystallographic Symmetry?

- 2 or more copies of a molecule in the unit cell not related by crystallographic symmetry
- Crystallographic copies of molecules are (of course) treated as if they were exactly the same across the unit cell – and indeed across the whole crystal
- Non-crystallographically related molecules provide different representations of the same molecule
  - This can be useful for model-building
  - But difficult to use in practice

Handling NCS

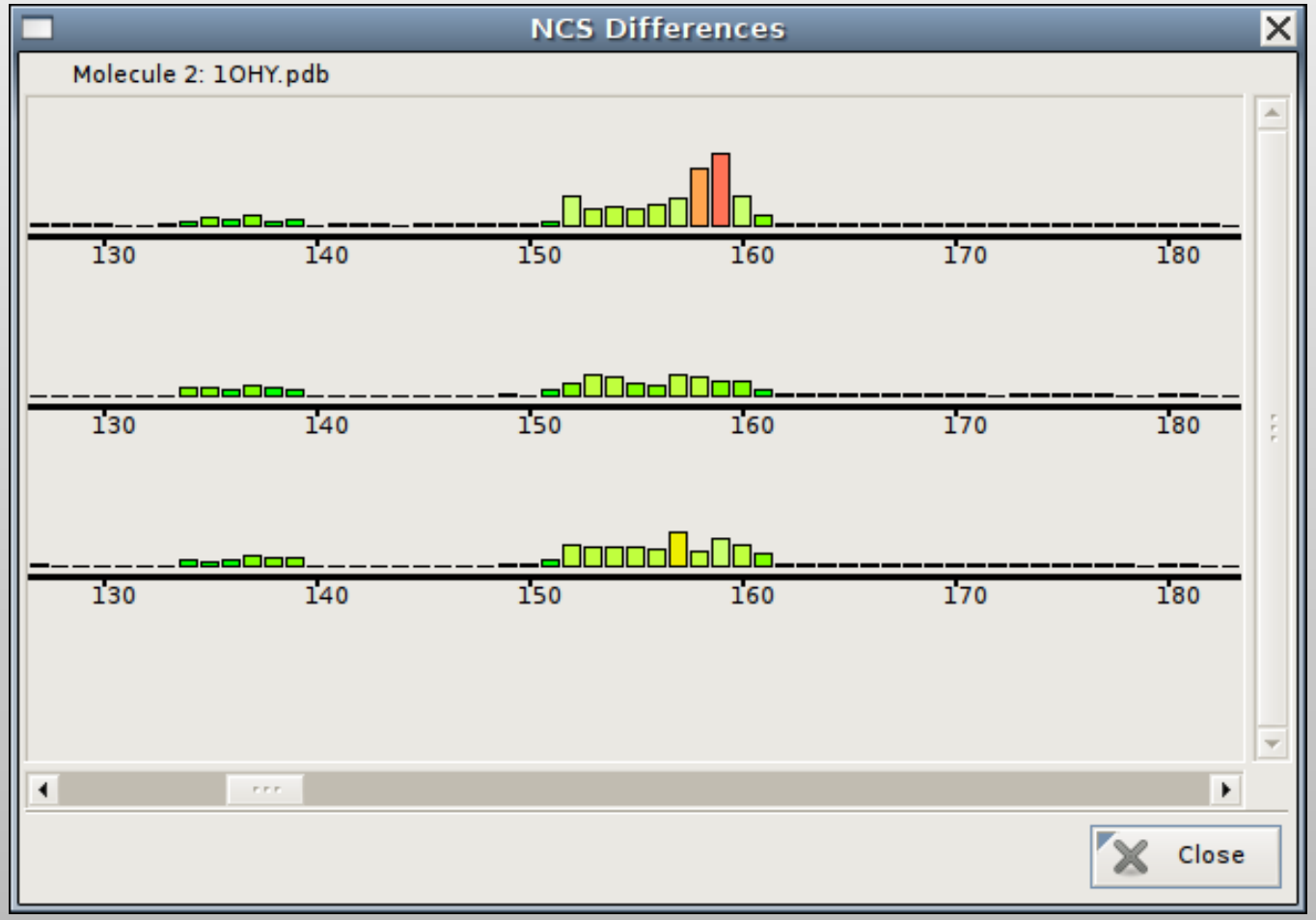
- What are the Problems?
- Strict NCS:
  - NCS should appear like crystallographic symmetry does [exact copies]
- Non-Strict NCS:
  - Molecules are different
  - How to cope with differences, but minimize unnecessary rebuilding?

Handling NCS

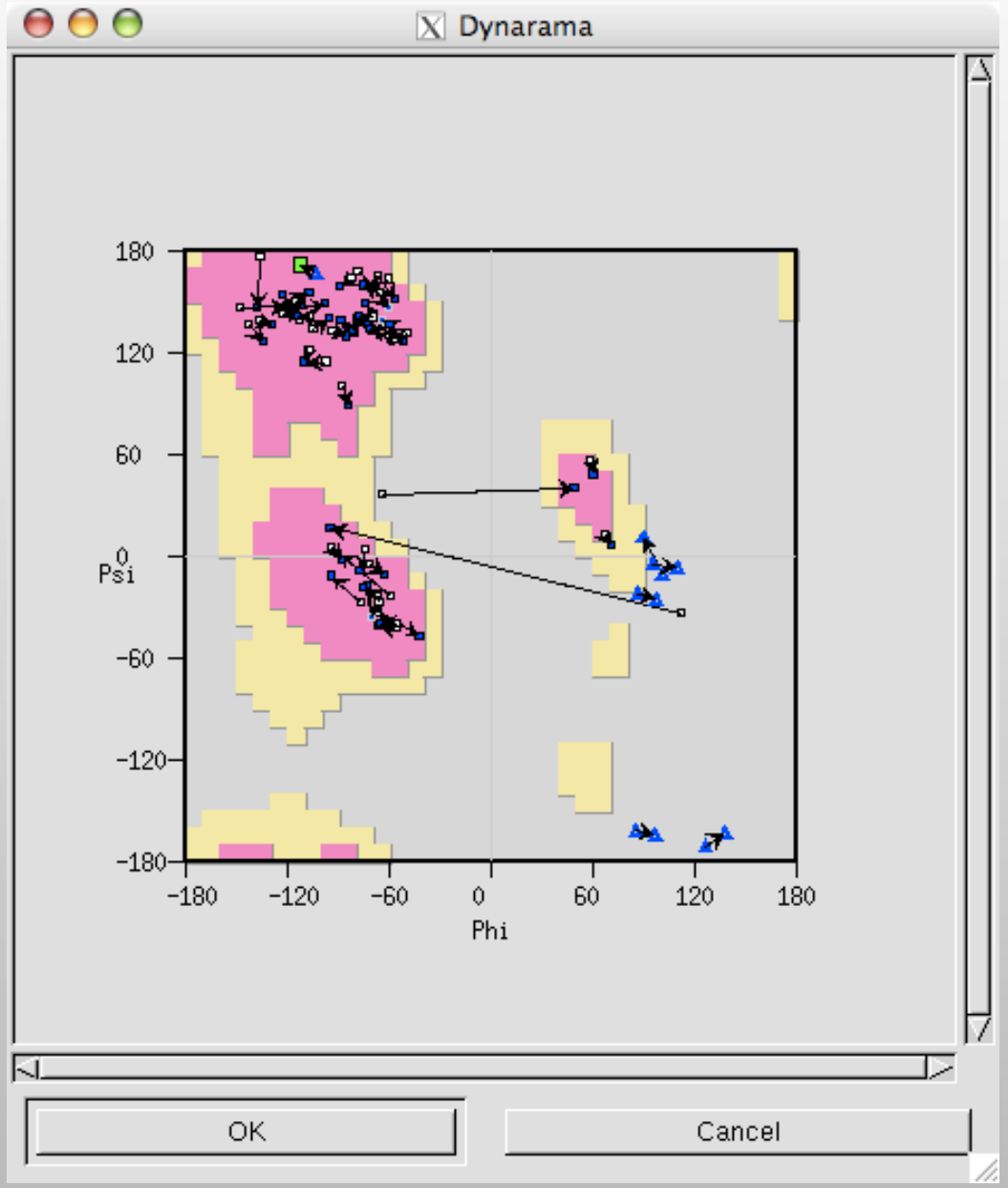
Typical Scenario:

- I have done an LSQ overlap of my NCS-related molecules and from the graph, have seen significant deviations in the positions of some side-chains.
- Why are they different?

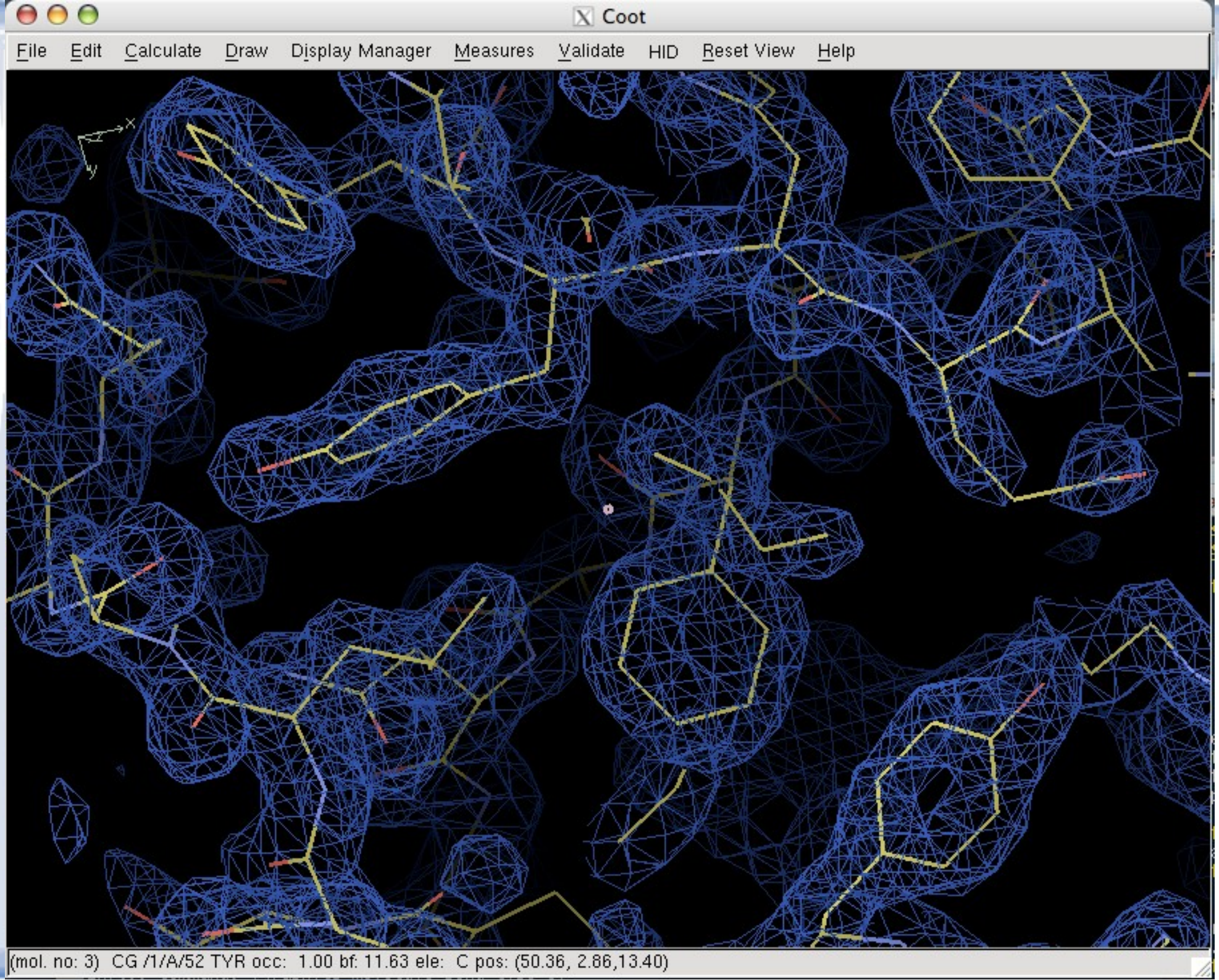
...or new NCS Differences graph



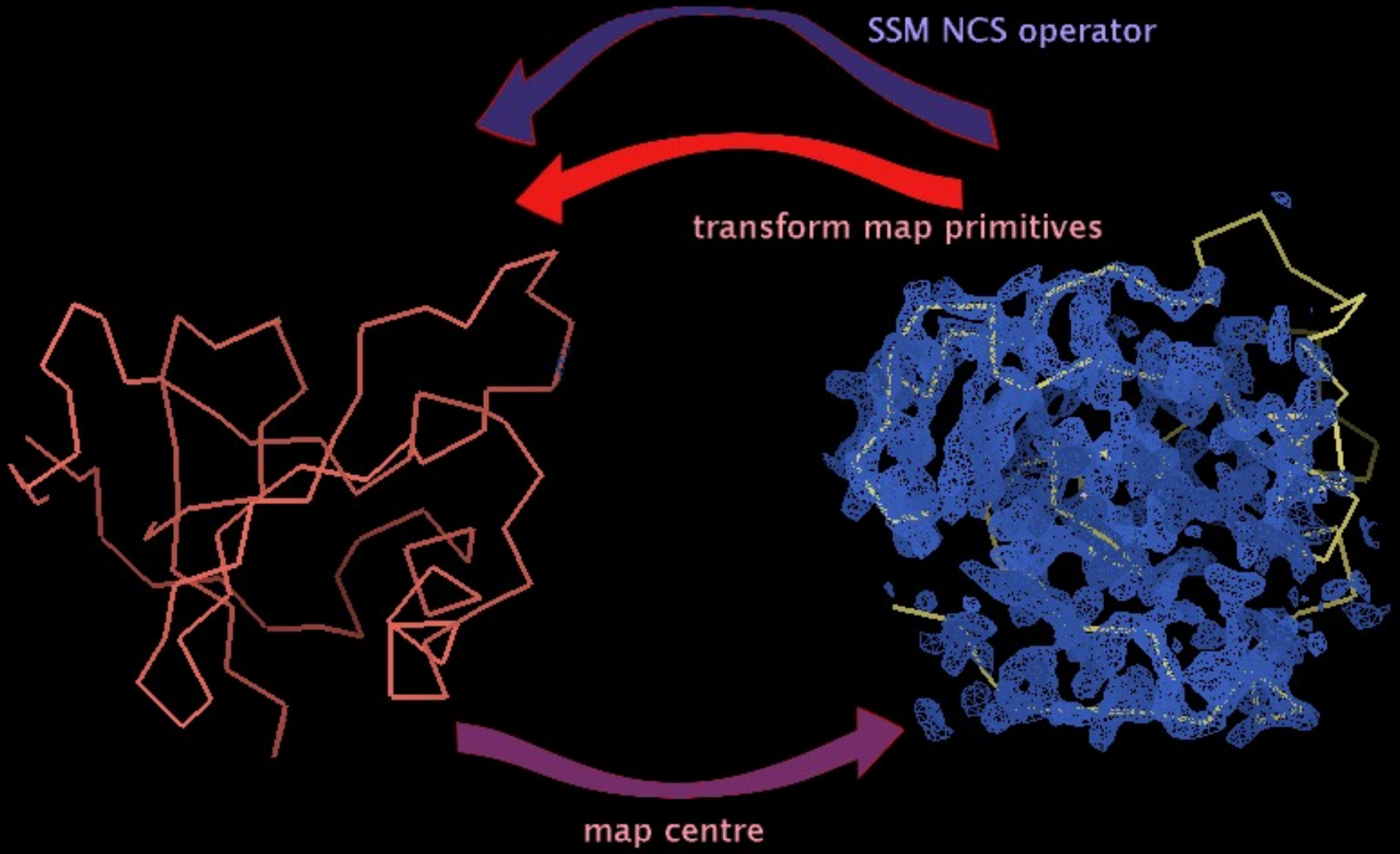
...or Kleywegt Plots[*]

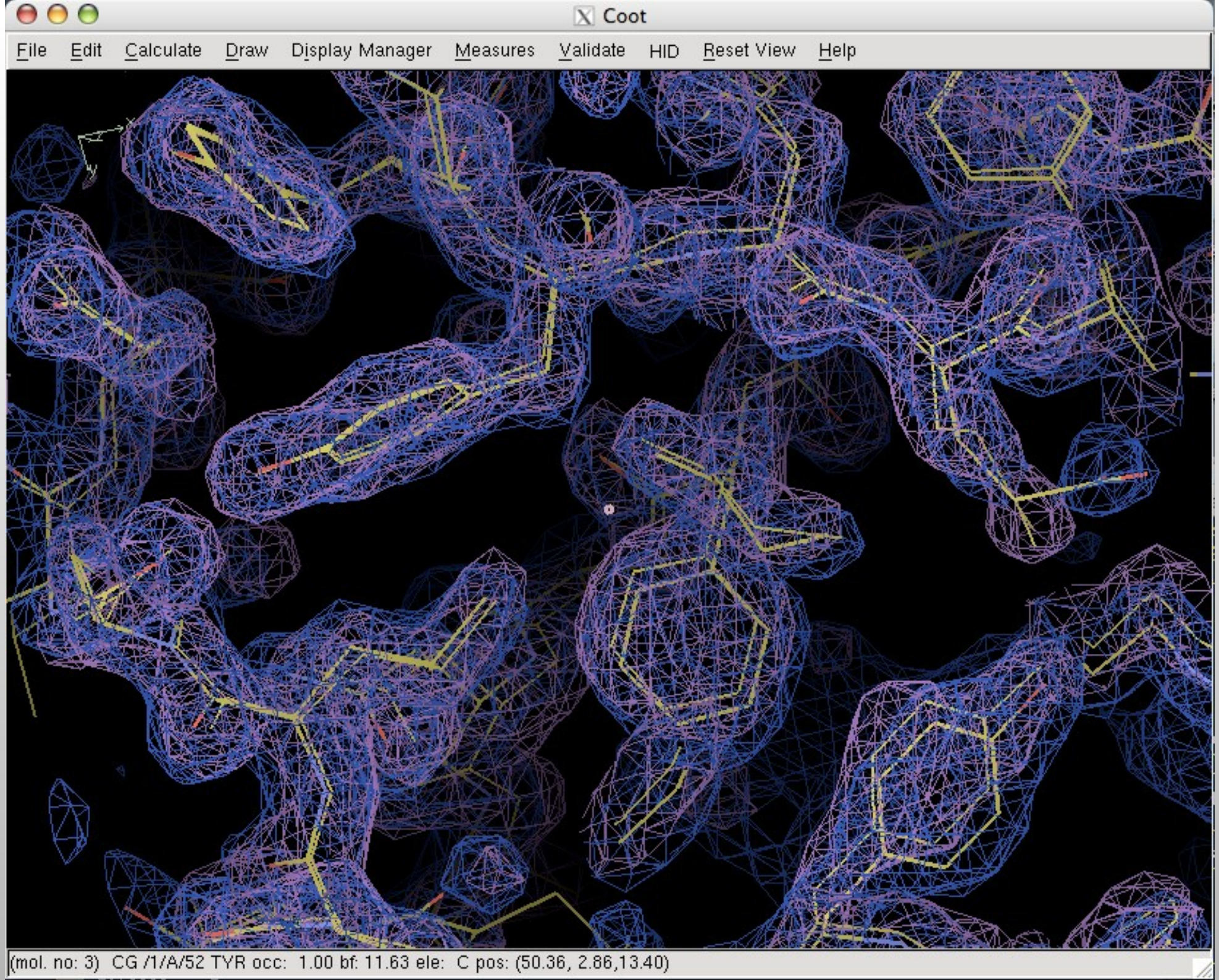


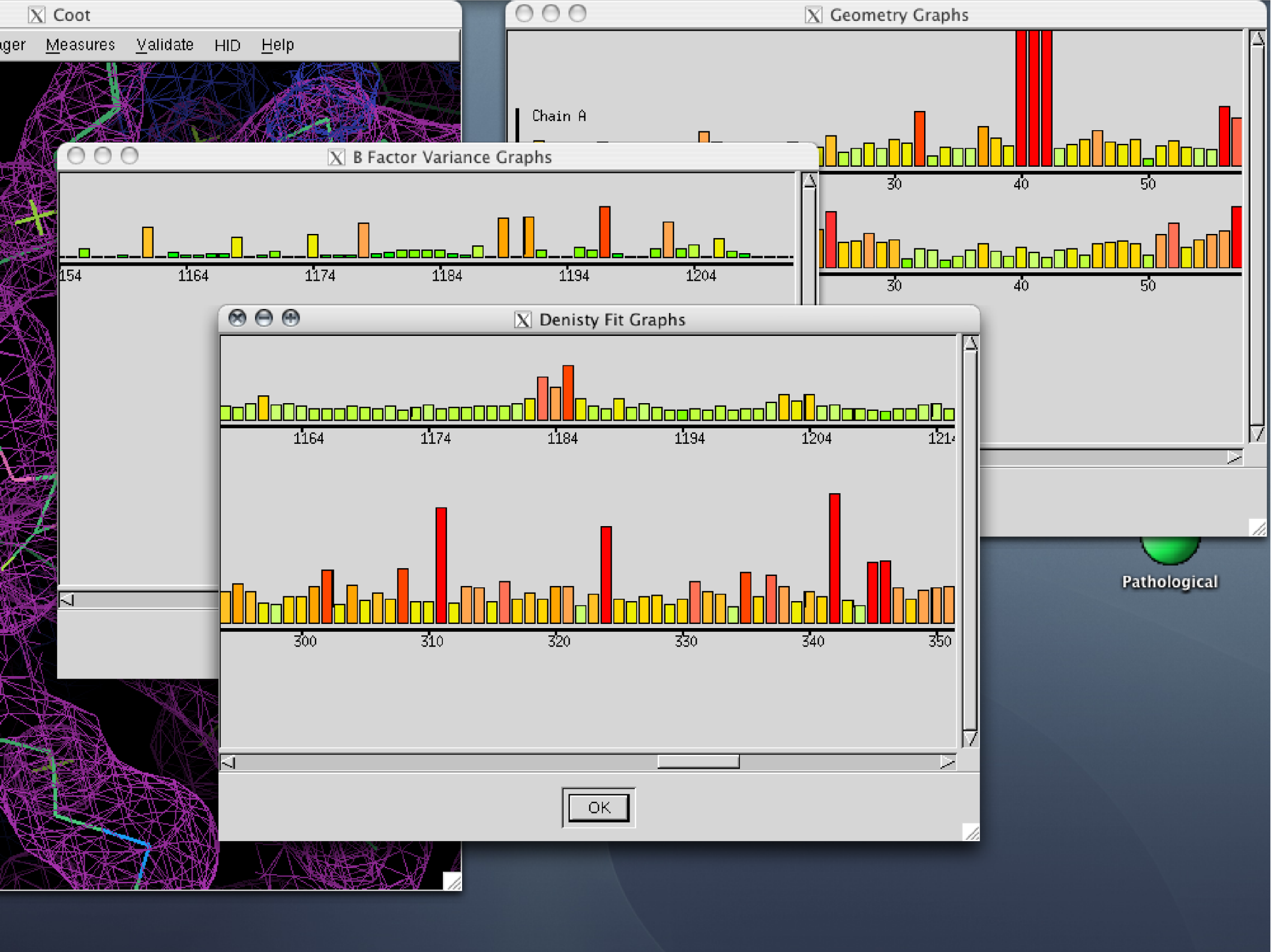
[*] Named by George Sheldrick


Coot
File
Edit
Calculate
Draw
Display Manager
Measures
Validate
HID
Reset View
Help
(mol. no: 3) CG /1/A/52 TYR occ: 1.00 bf: 11.63 ele: C pos: (50.36, 2.86,13.40)

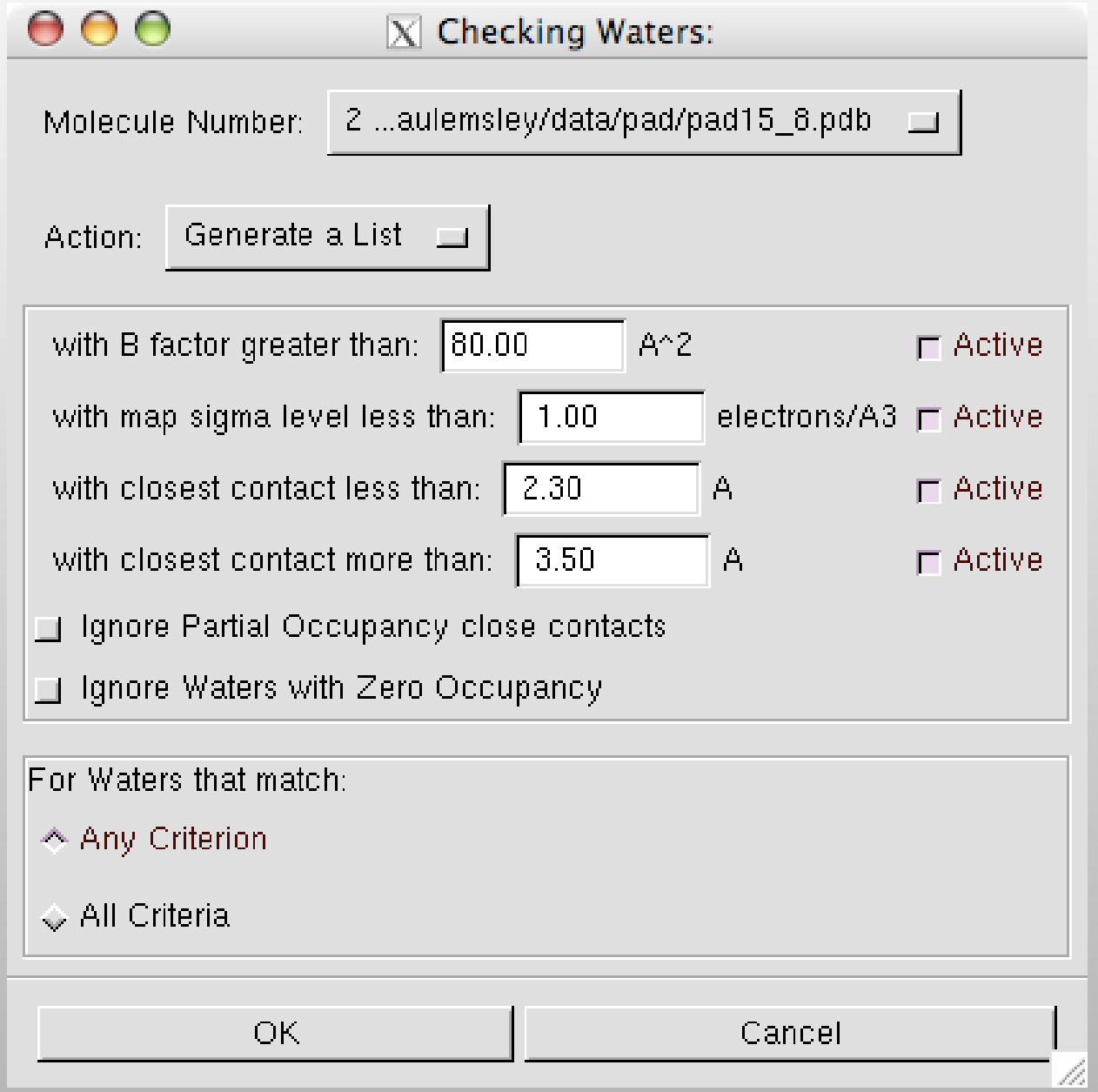
NCS Overlays




Coot
File Edit Calculate Draw Display Manager Measures Validate HID Reset View Help
(mol. no: 3) CG /1/A/52 TYR occ: 1.00 bf: 11.63 ele: C pos: (50.36, 2.86,13.40)


Coot
ger Measures Validate HID Help
Geometry Graphs
Chain A
30
40
50
B Factor Variance Graphs
154
1164
1174
1184
1194
1204
Denisty Fit Graphs
1164
1174
1184
1194
1204
300
310
320
330
340
350
OK
Pathological

Check/Delete Waters



Highly Coordinated waters

- Waters with 4 or more contacts
  - Purely geometry based
- Tested by electronic shielding for candidates as K+, Na+, Ca2+, Mg2+
  - Nayal & Di Cera (1996)
  - a.k.a. “WASP”
  - Includes symmetry search

A Sample of Tools

- A few tools and tricks have been described here
- Also, validation and ligand fitting
- There are several interfaces to other programs/suites/web-services...
  - e.g. EBI, EDS, CCP4, Refmac, Libcheck, Molprobity, What_check, Raster3D, SHELXL...

Acknowledgements

- Group Murshudov
  - Fei Long & Rob Nicholls
- Kevin Cowtan
- Bernhard Lohkamp

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