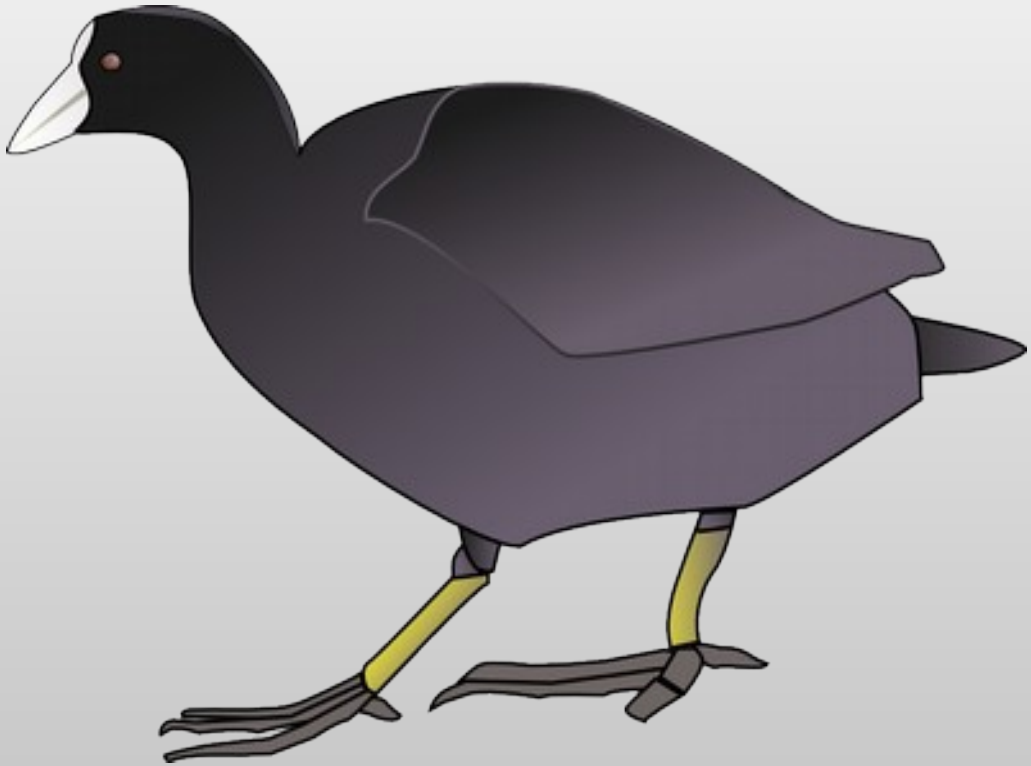
Coot (not a duck)

An Introduction to *Coot*



Paul Emsley
MRC Laboratory of Molecular Biology
Nov 2018

Model-building with *Coot*

- Tools to help improve the quality of the macromolecular model
  - Refinement tools/residue-based tools
  - Tools for ligand analysis and presentation
  - Tools for Cryo-EM fitting

About this Presentation

- Refinement
- Rotamers
- cis-peptides
- Helices
- Representation
- Using *Coot*

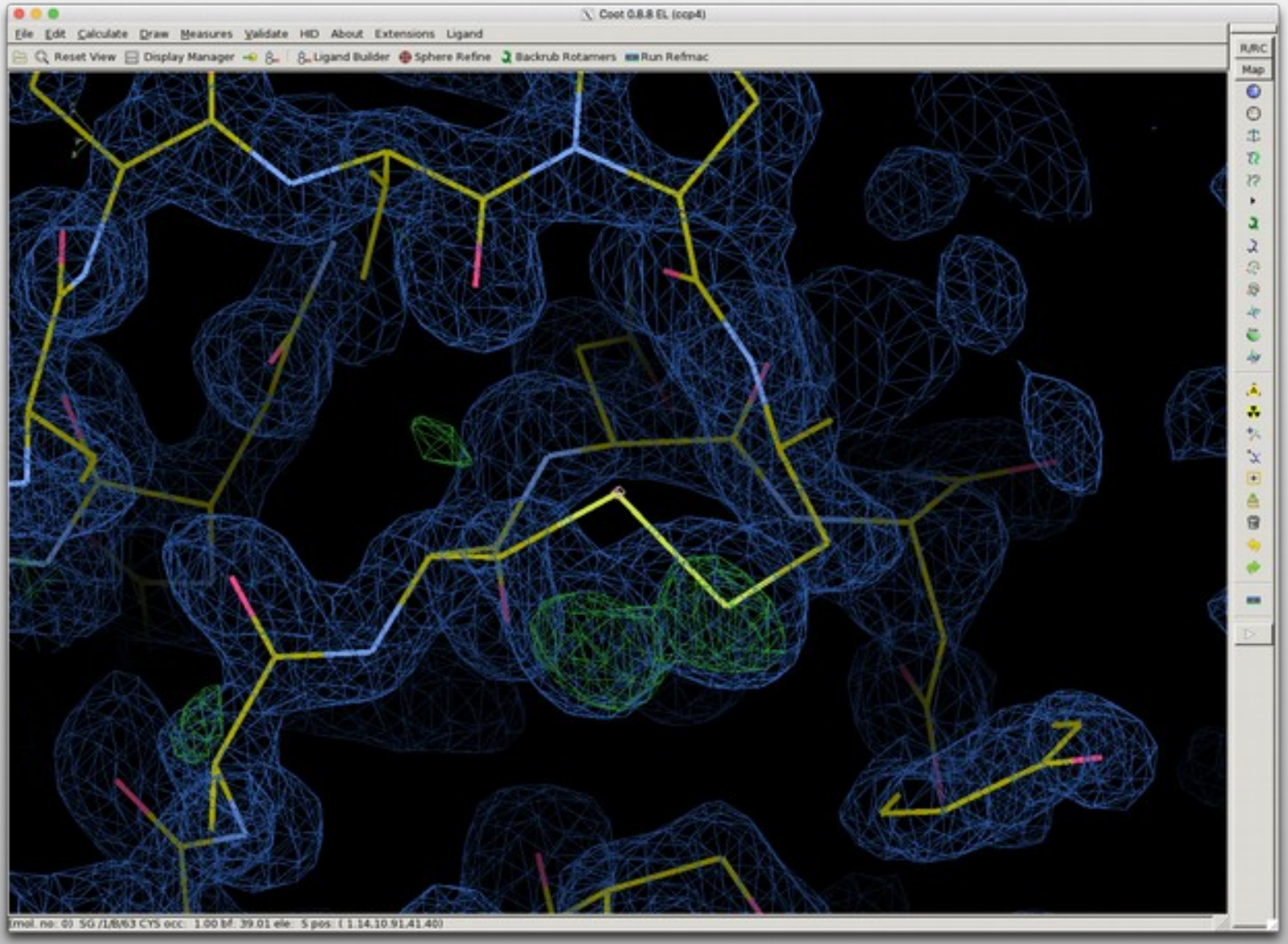
Coot

- Molecular Graphics application
  - Protein Crystallographic model-building tools
  - Designed to “fill the gap” where automatic methods fail
    - (generally, we don't use molecular graphics programs to do what automatic methods can do)
- Interface to other programs: SHELXL, Refmac, Libcheck, Probe&Reduce (Molprobity), EBI, EDS, Povray… and others

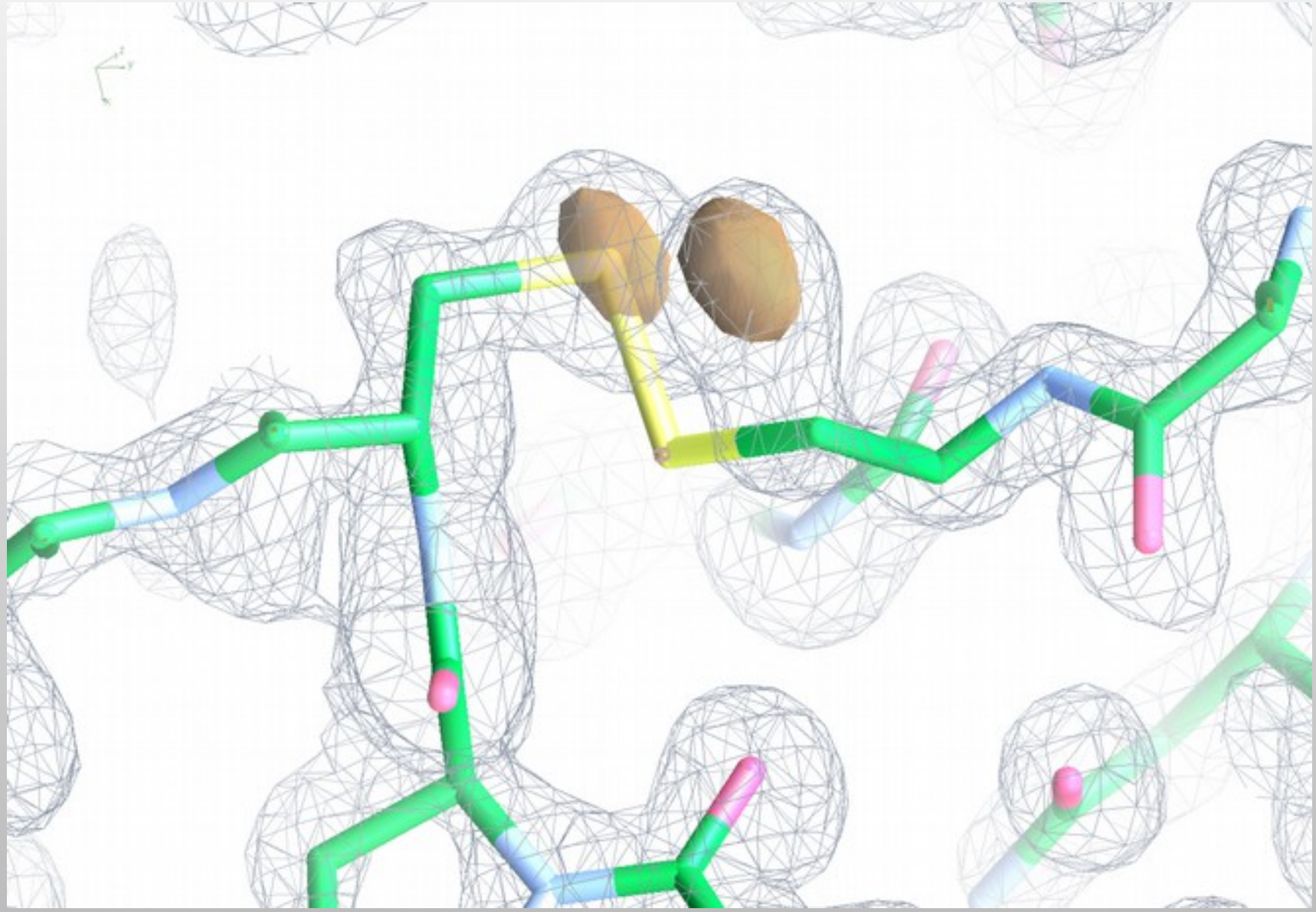
But Why Bother?

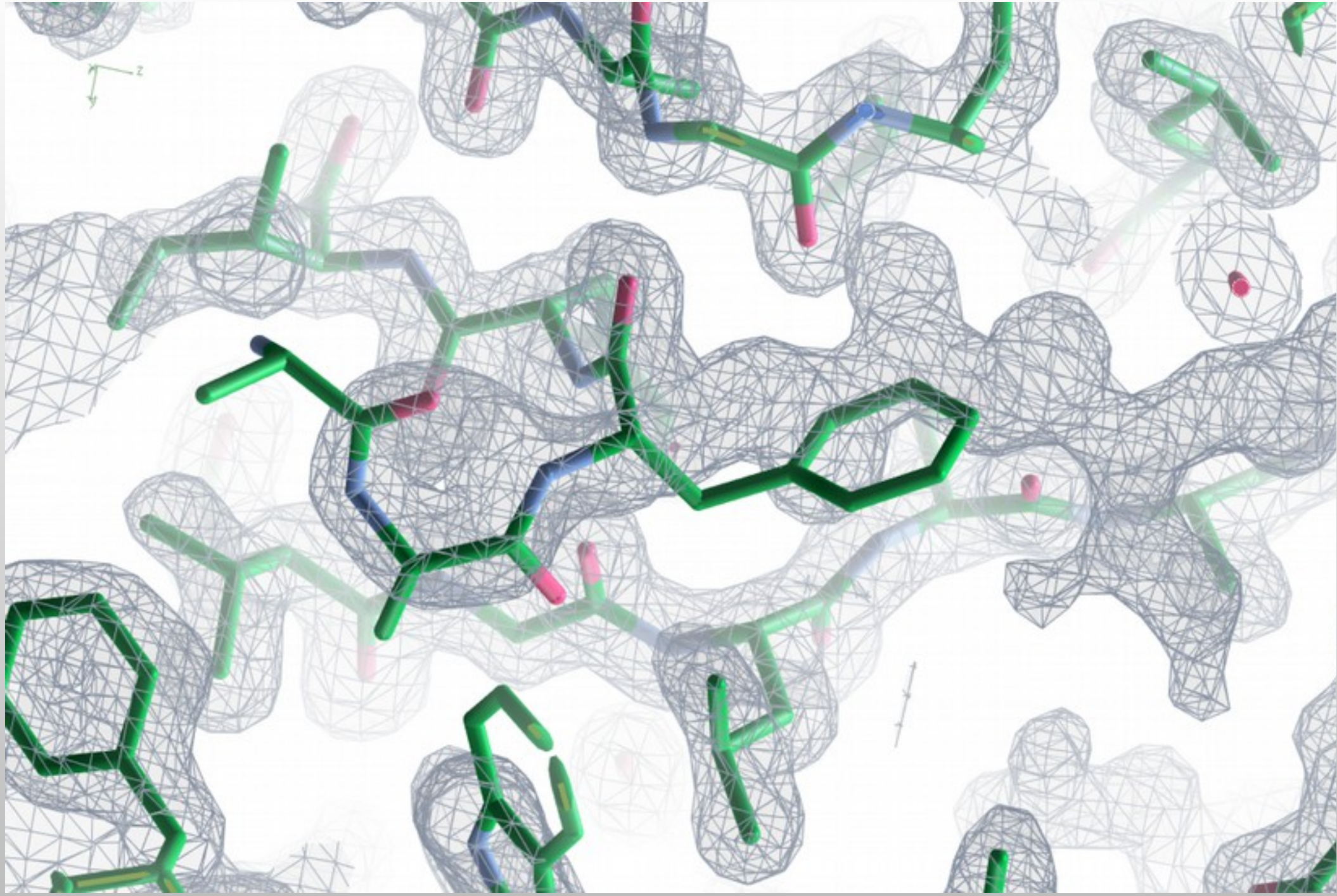
- Automated model-building for complete models is still impossible
  - It takes a brain to validate
- Concerted correction/improvement of a model is difficult on the larger scale

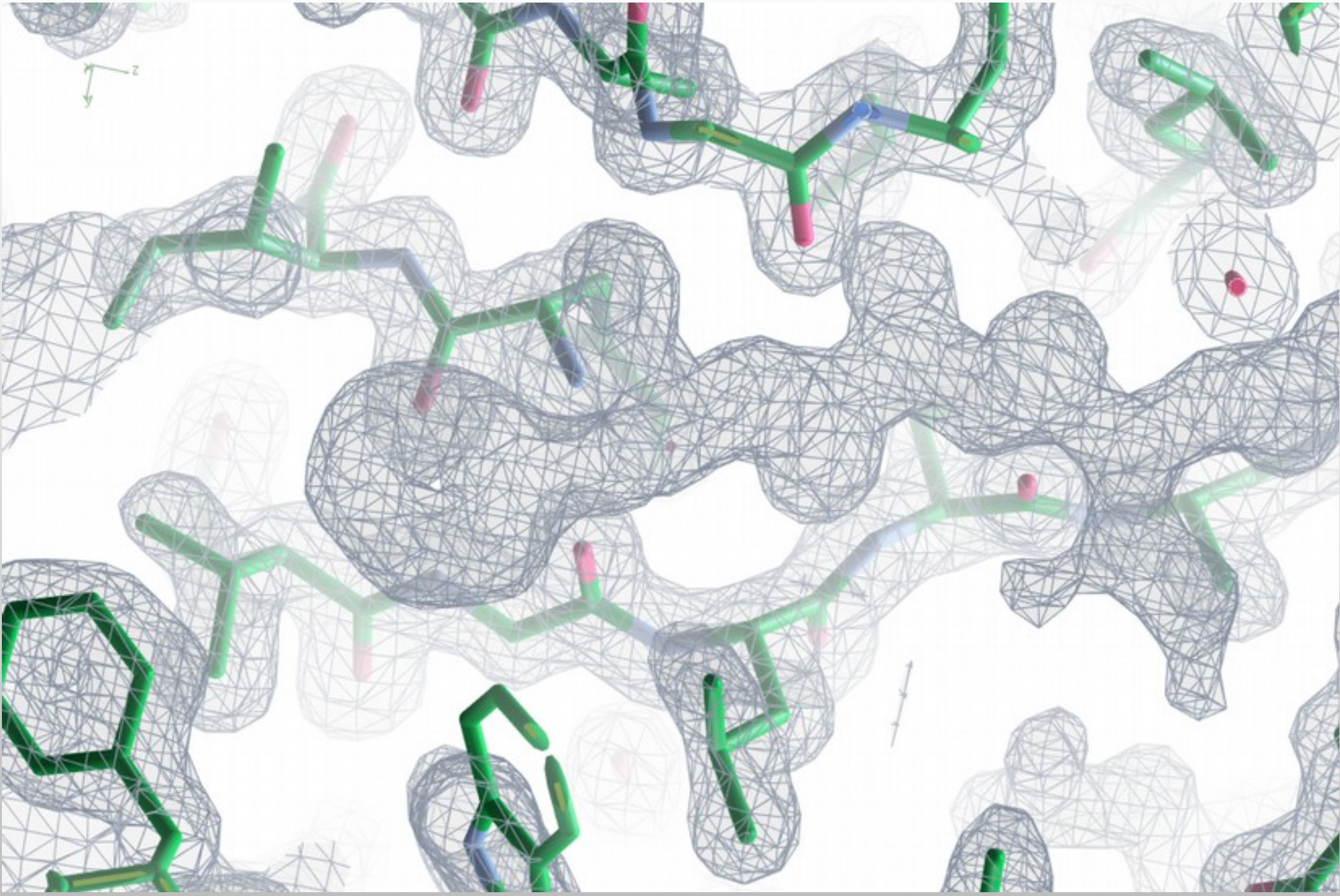
Fixing what auto-building doesn't get right



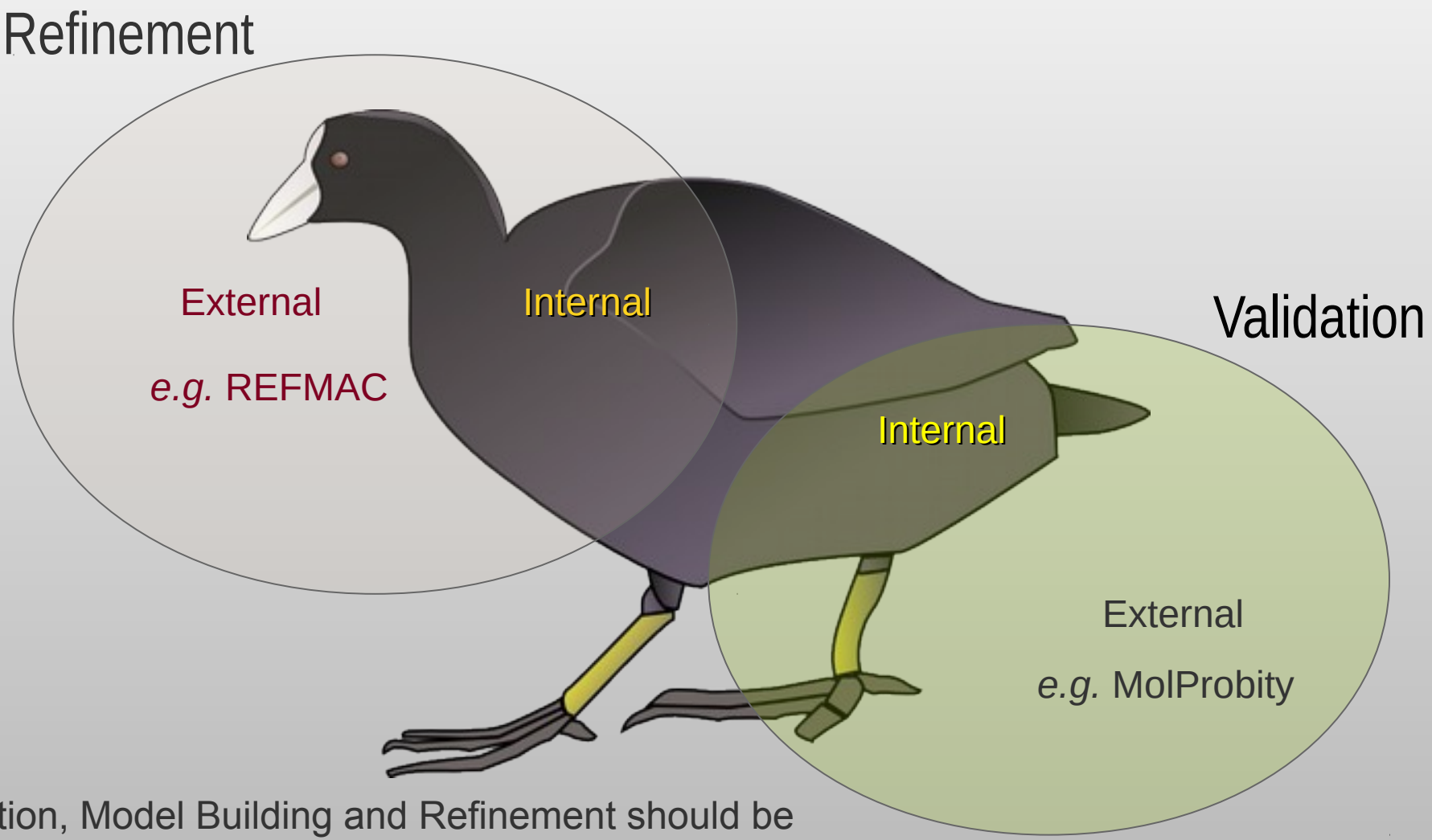
Fixing what auto-building doesn't get right







Feature Integration



Validation, Model Building and Refinement should be used together

What is “Refinement”?

- The adjustment of model parameters (co-ordinates) so that the calculated structure factors match the observations as nearly as possible
  - In “one-shot” real-space refinement, such as in *Coot*, this translates to:
    - move the atoms into as high density as possible while minimizing geometrical distortions

Real Space Refinement

- Major feature of Coot
  - Gradient minimizer (BFGS derivative)
  - Based on mmCIF standard dictionary
  - Minimizing bonds, angles, planes, non-bonded contacts, torsions, chiral volumes
- Provides “interactive refinement”

Refinement has been extended in several ways...

What prior geometric information do we have?

- We know chemistry....
  - We know bond lengths and uncertainties
  - We know bond angles and uncertainties
  - We know the chiral centres
  - We know which atoms should lie in a plane
  - We know (more or less) about torsions
- We combine the gradients from the data with those from molecular mechanics in the minimisation

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    0.033
 ALA      CA   C       single     1.525    0.021
 ALA      C    O       double     1.231    0.020
```

APPENDIX *A*
Regularization and refinement derivatives

The function that we are trying to minimize is S, where

$$S = S_{\text{bond}} + S_{\text{angle}} + S_{\text{torsion}} + S_{\text{plane}} + S_{\text{nbc}} + S_{\text{chiral}}$$

A1. Bonds

$$S_{\text{bond}} = \sum_{i=1}^{N_{\text{bonds}}} (b_i - b_{0_i})^2,$$

where b_{0_i} is the ideal length (from the dictionary) of the ith bond, $\mathbf{b}_i$ is the bond vector and b_i is its length.

$$\frac{\partial S_i}{\partial x_m} = \frac{\partial S_i}{\partial b_i}\frac{\partial b_i}{\partial x_m} = [2(b_i - b_{0_i})]\frac{\partial b_i}{\partial x_m},$$

$$b_i = [(x_m - x_k)^2 + (y_m - y_k)^2 + (z_m - z_k)^2]^{1/2}.$$

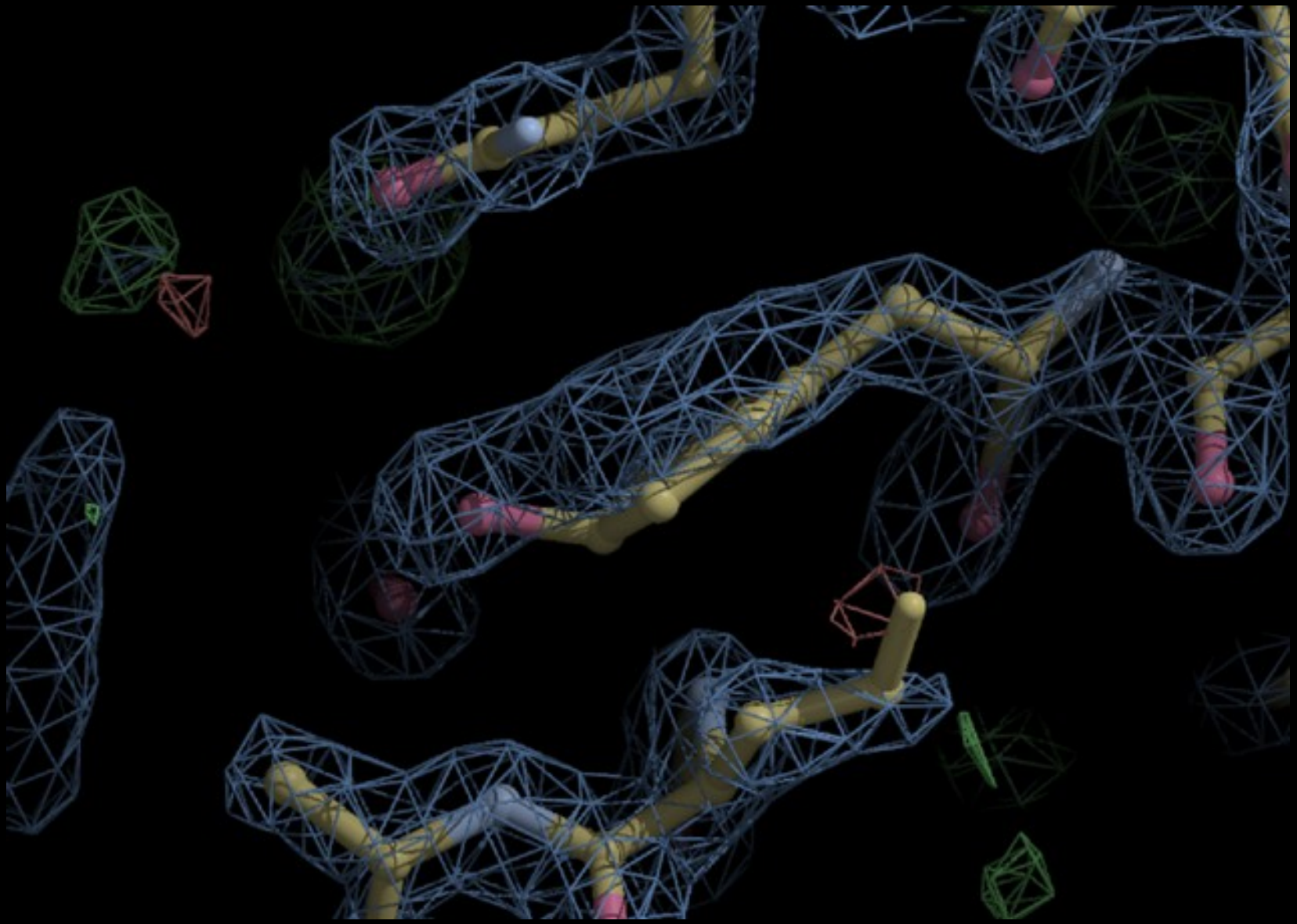
Therefore

$$\frac{\partial b_i}{\partial x_m} = \left(\frac{1}{2}\frac{1}{b_i}\right)2(x_m - x_k) = \frac{(x_m - x_k)}{b_i}$$

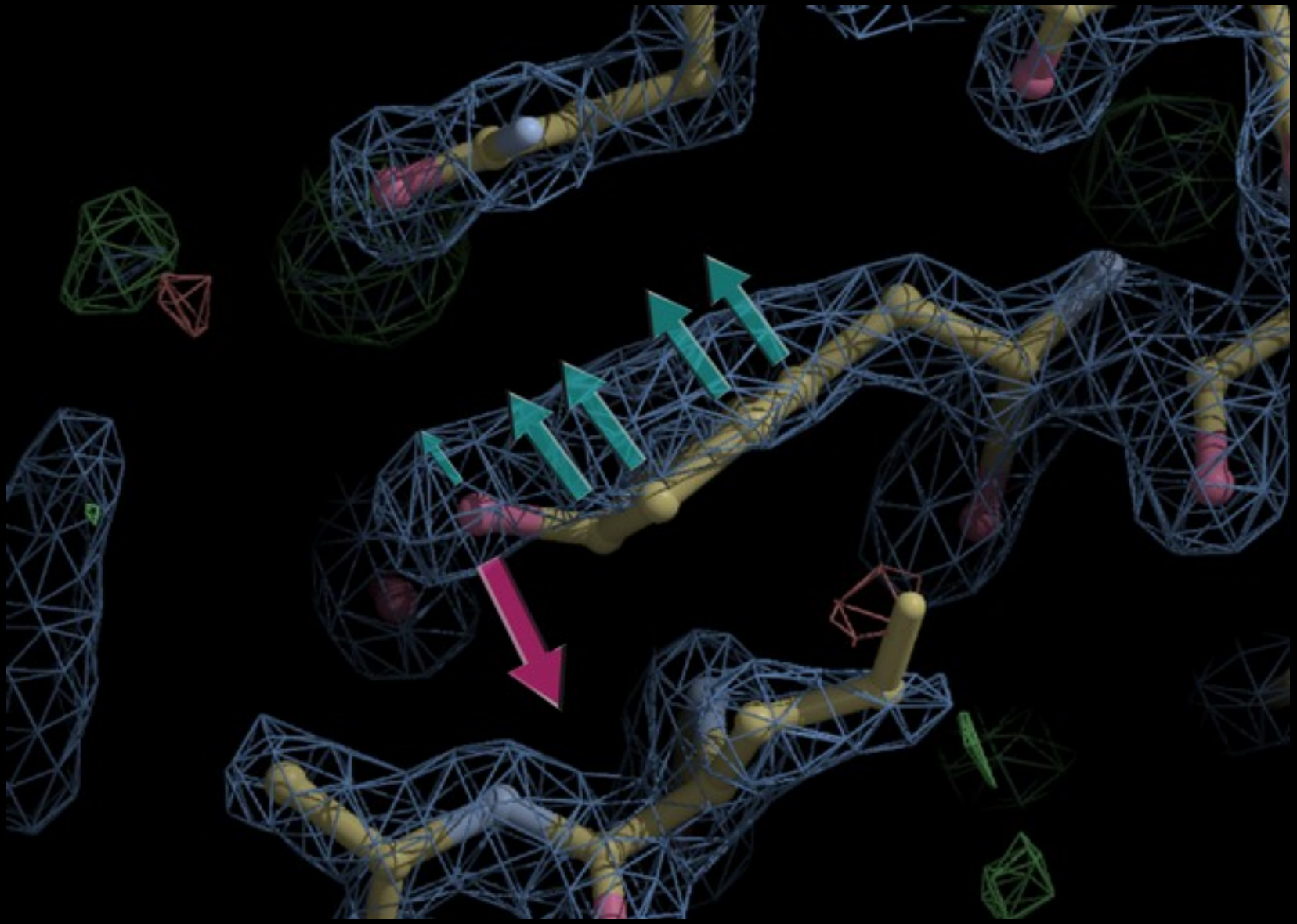
and

$$\frac{\partial S_i}{\partial x_m} = 2[b_i - b_{0_1}]\frac{(x_m - x_k)}{b_i}.$$

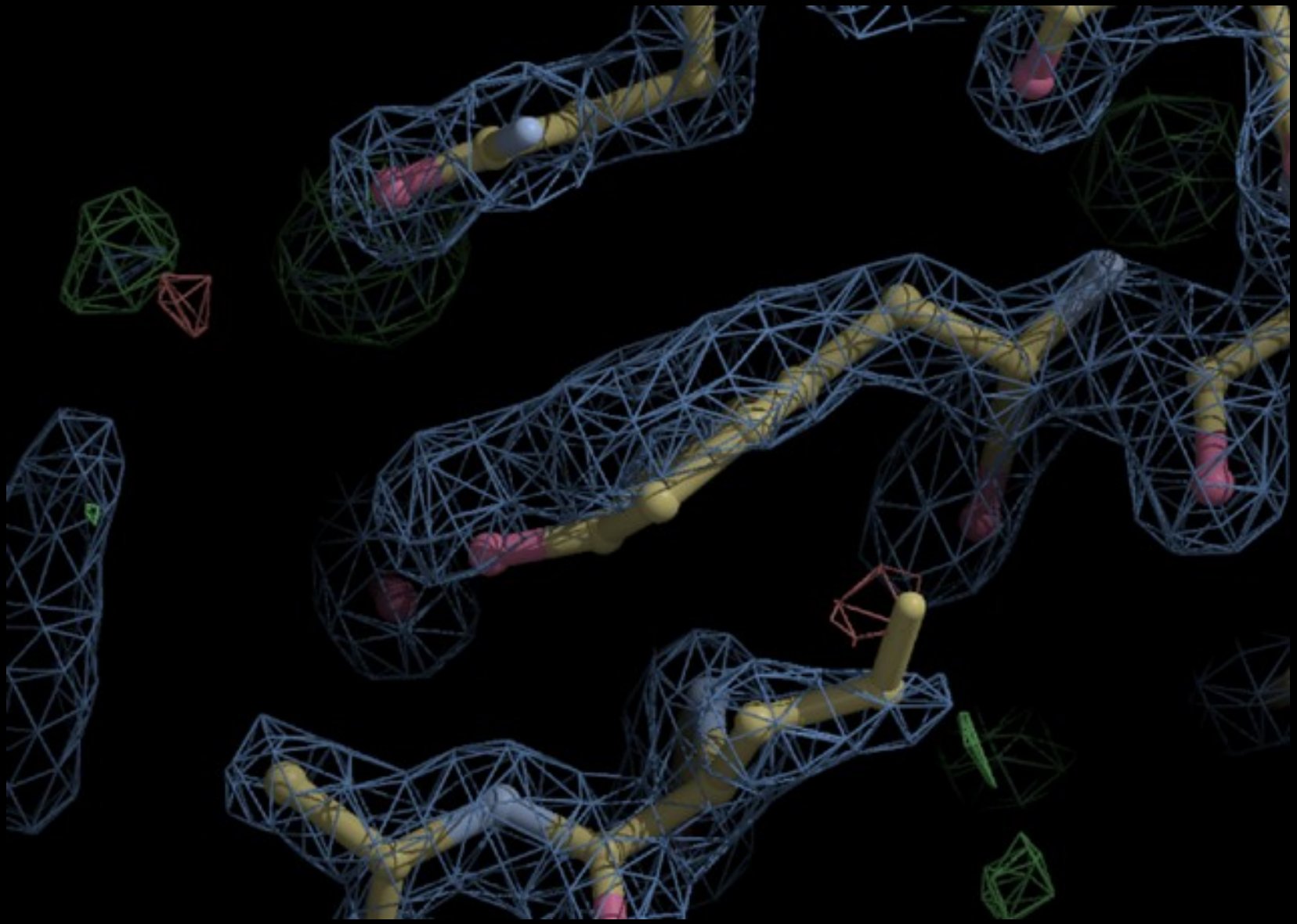
Distorted Geometry Pre-Refinement



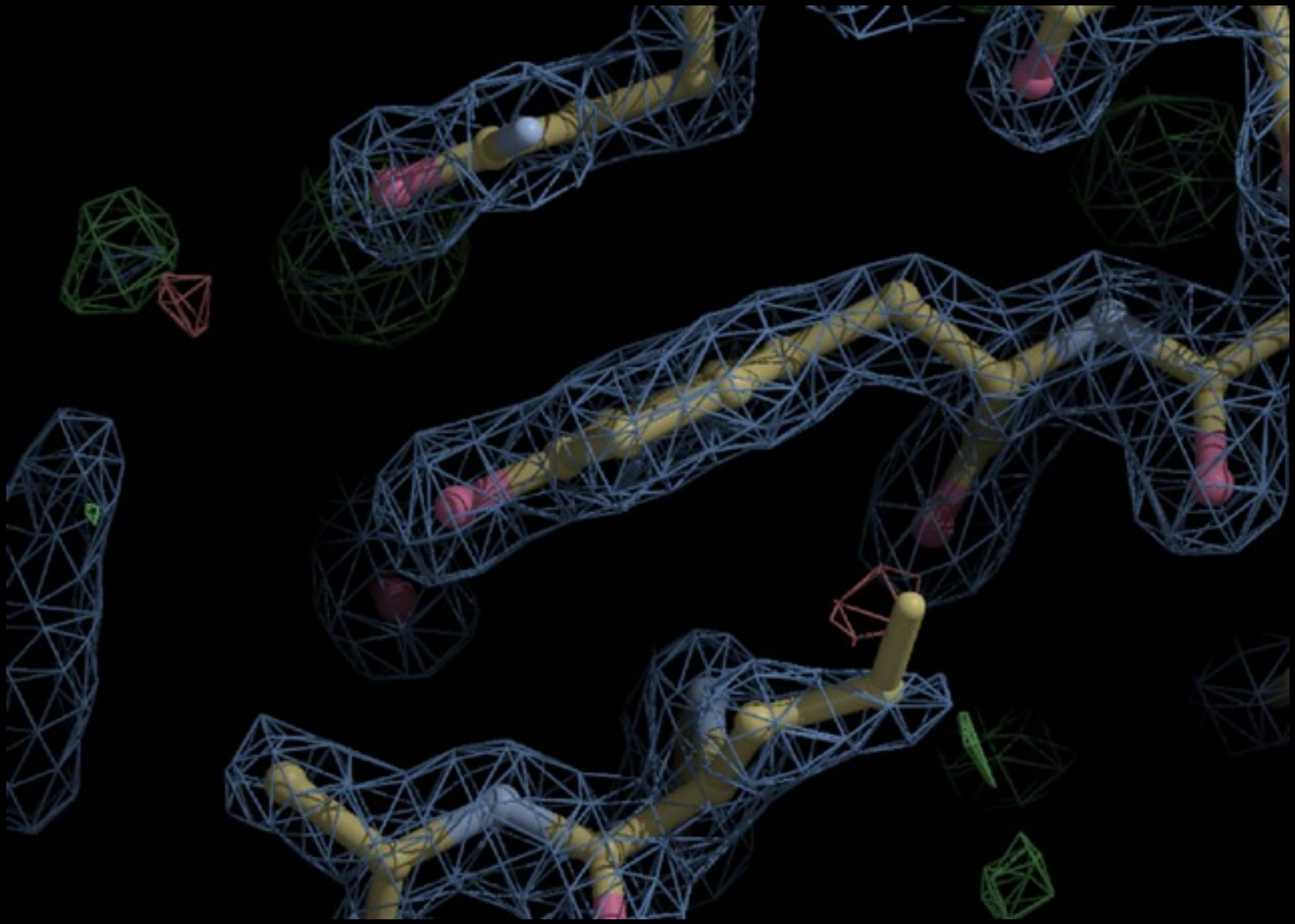
Refinement Gradients



Refinement: Cycle 3

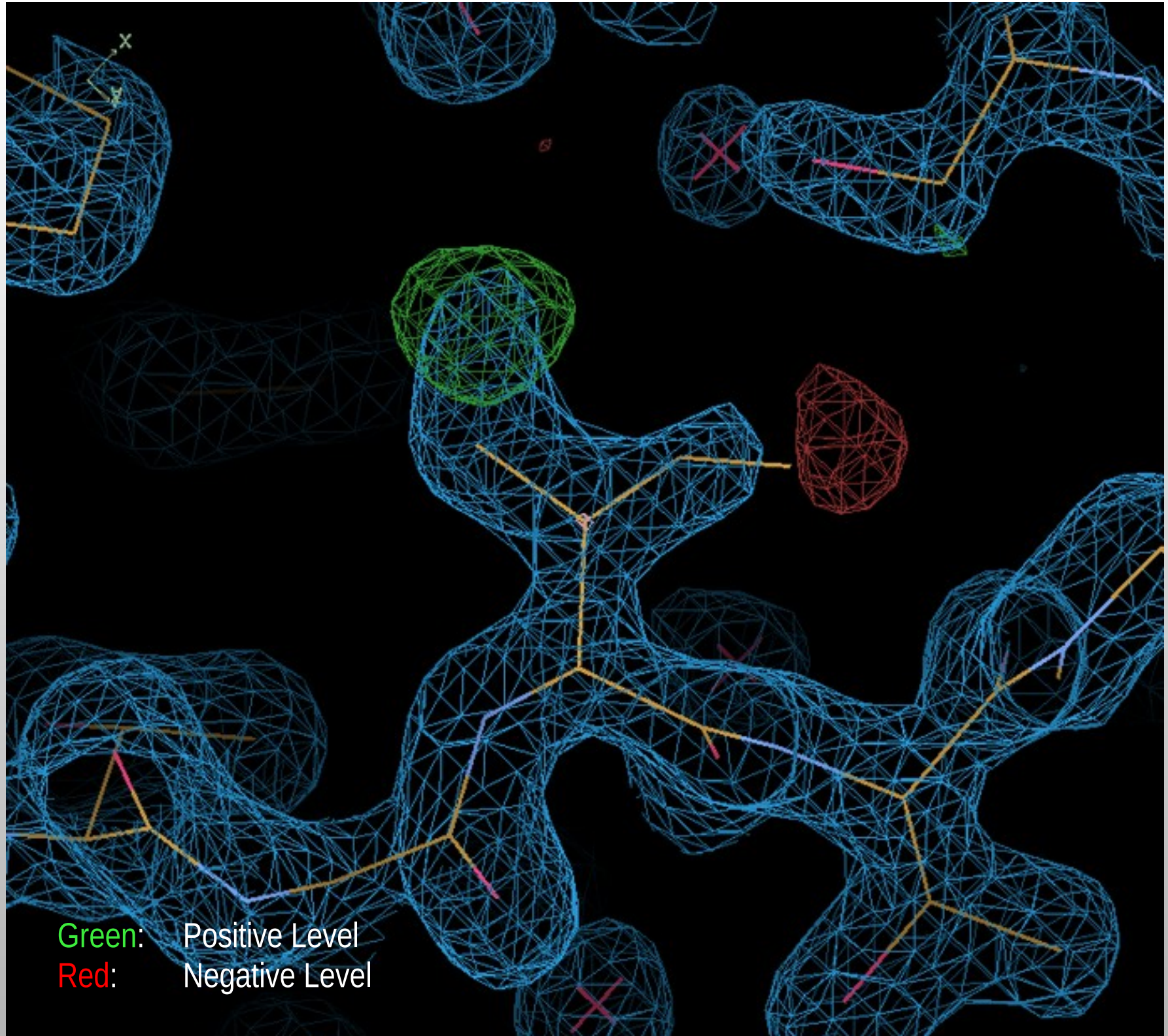


Refinement Cycle 200: Minimized



Different types of electron density maps

- "Experimental" maps
  - maps that result directly from the crystallographic data analysis: MIR, MAD, SAD
- Direct Maps:
  - where the atoms are
- Coefficients F_o-F_c ("difference map")
  - Identifies errors in the model. Locations in space where there should be atoms show positive peaks, while locations where the model contains atoms that should not be there show negative peaks.


Green: Positive Level
Red: Negative Level

Representation of Results:

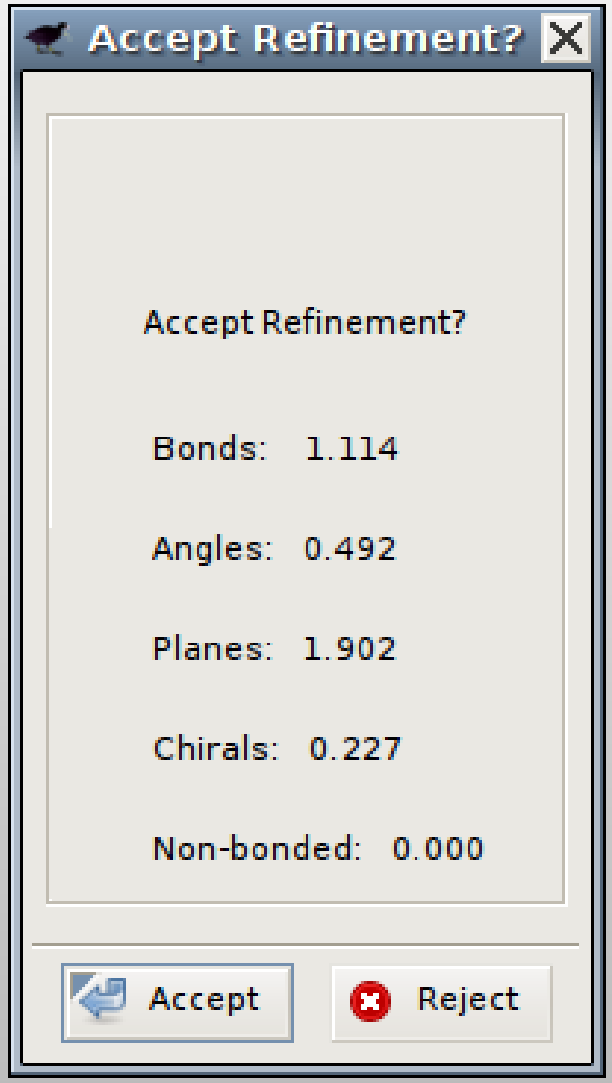
```
File  Edit  View  Terminal  Help
created 32 bond       restraints
created 38 angle      restraints
created 1 plane      restraints
created 5 chiral vol restraints
created 76 restraints

   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...”

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

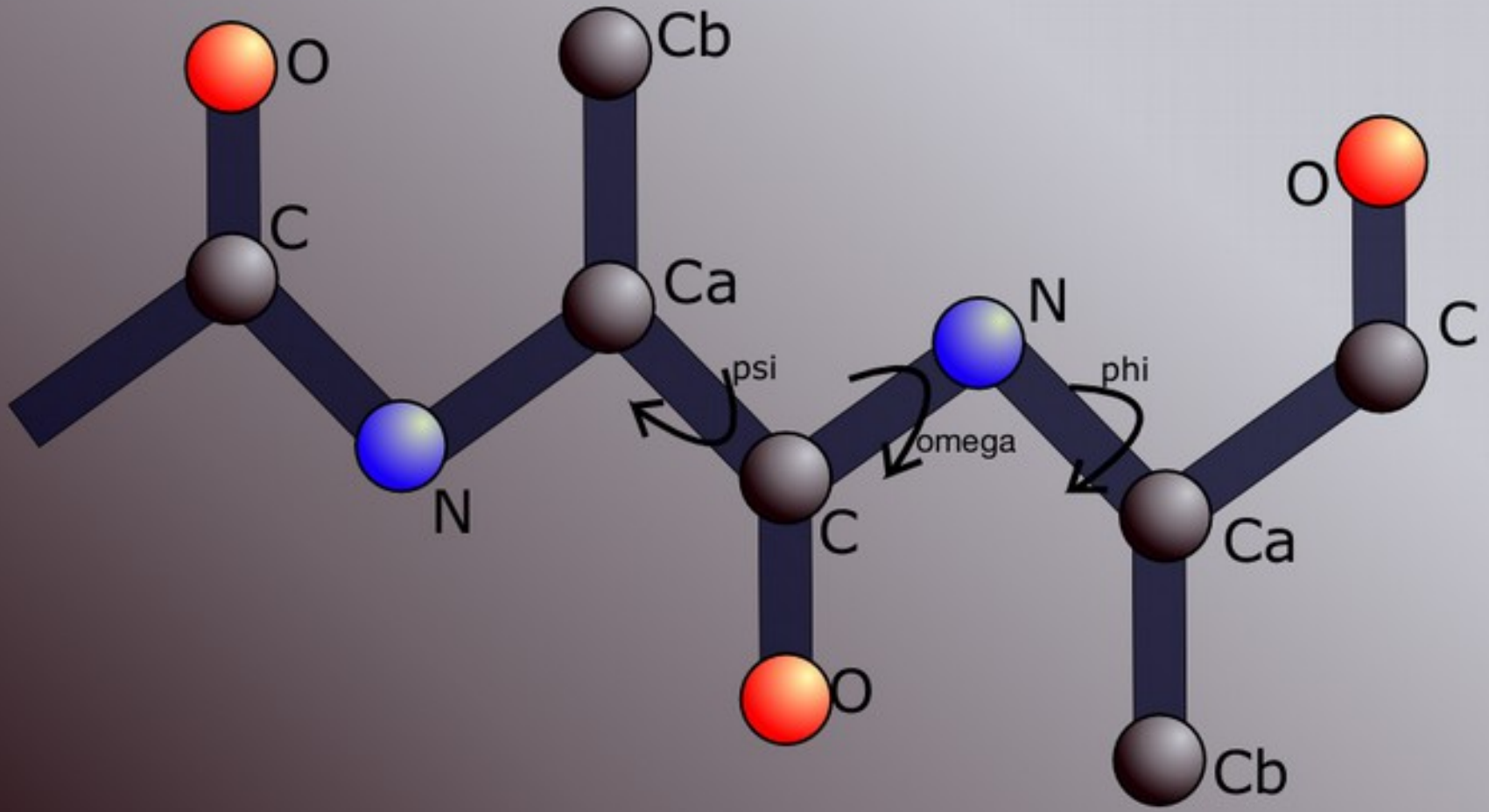
Refinement Techniques

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

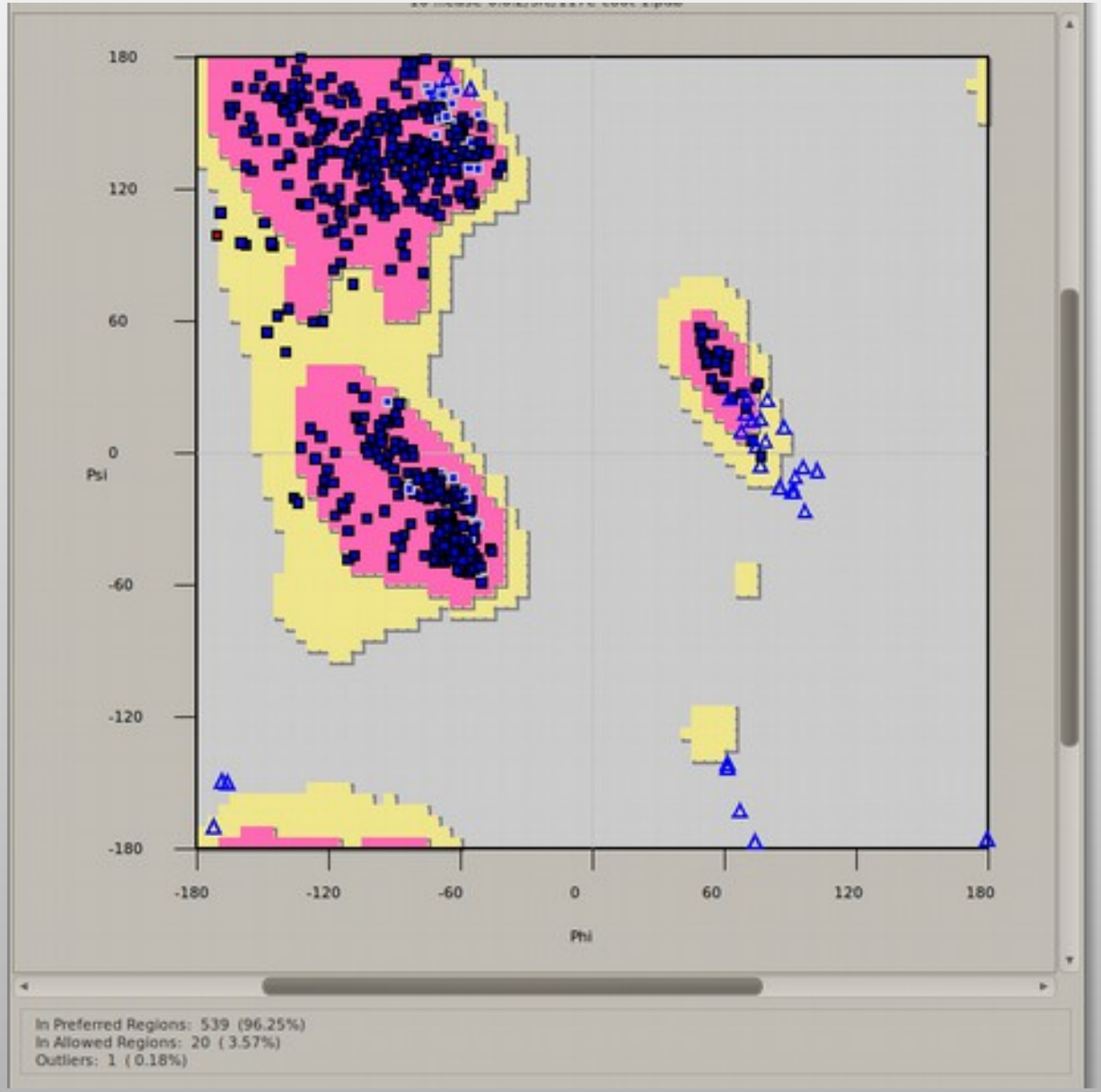
Ramachandran Plot...

- What do we know about protein structures...
  - (that can help in model-building?)

Peptide Torsion Angles



Typical 2D Projection of Ramachandran Plot



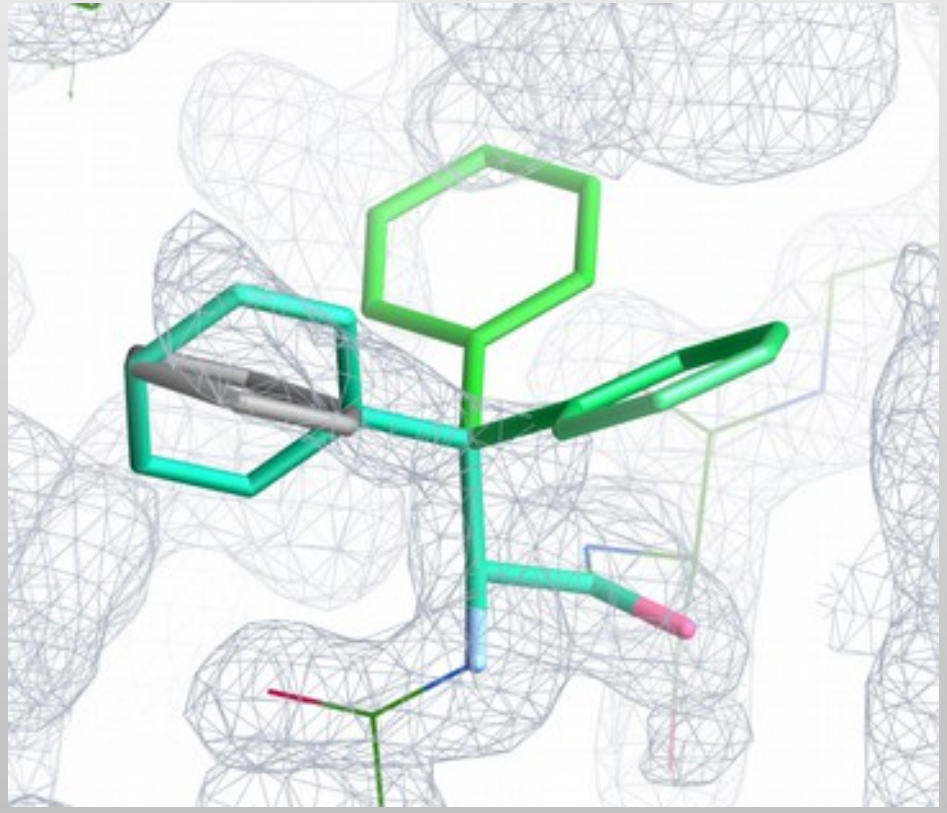
Rotamer Searching

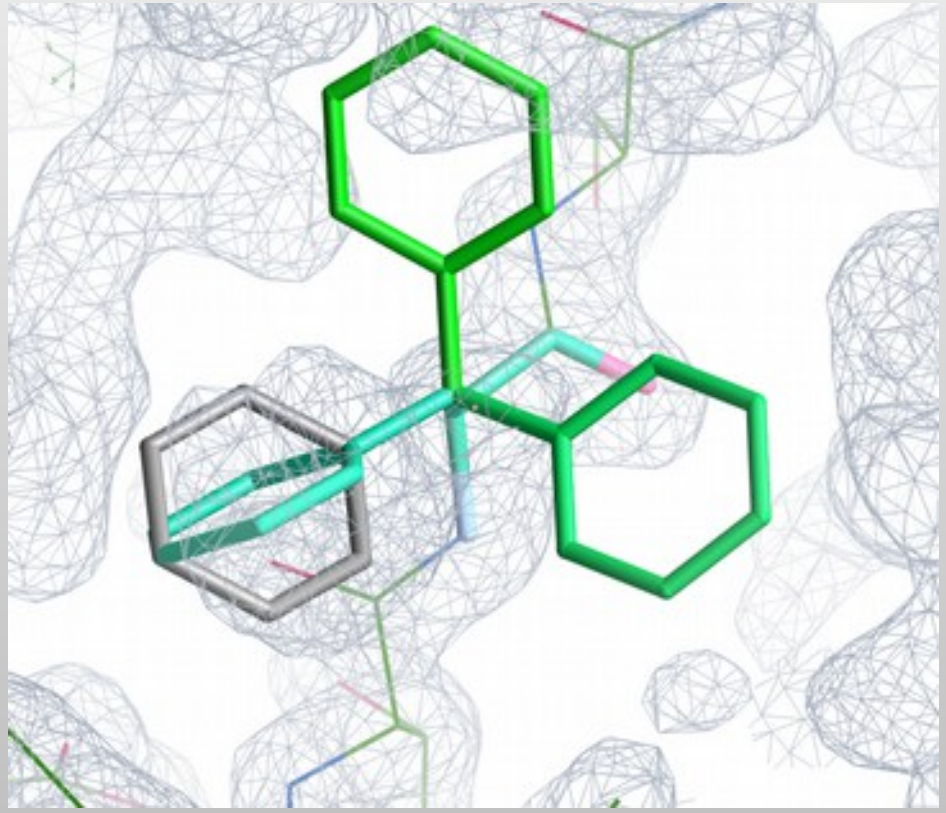
- Two methods
  - Traditional
  - Backrub

Rotamers

- Rotamers are preferred configurations of a side-chains rotatable bonds
  - where “preferred” means these configurations occur more frequently in a set of reference protein structures
  - “preferred” because they are low-energy conformations
- Several Rotamer “databases” exist
  - best: (Son of) Penultimate Rotamer Library

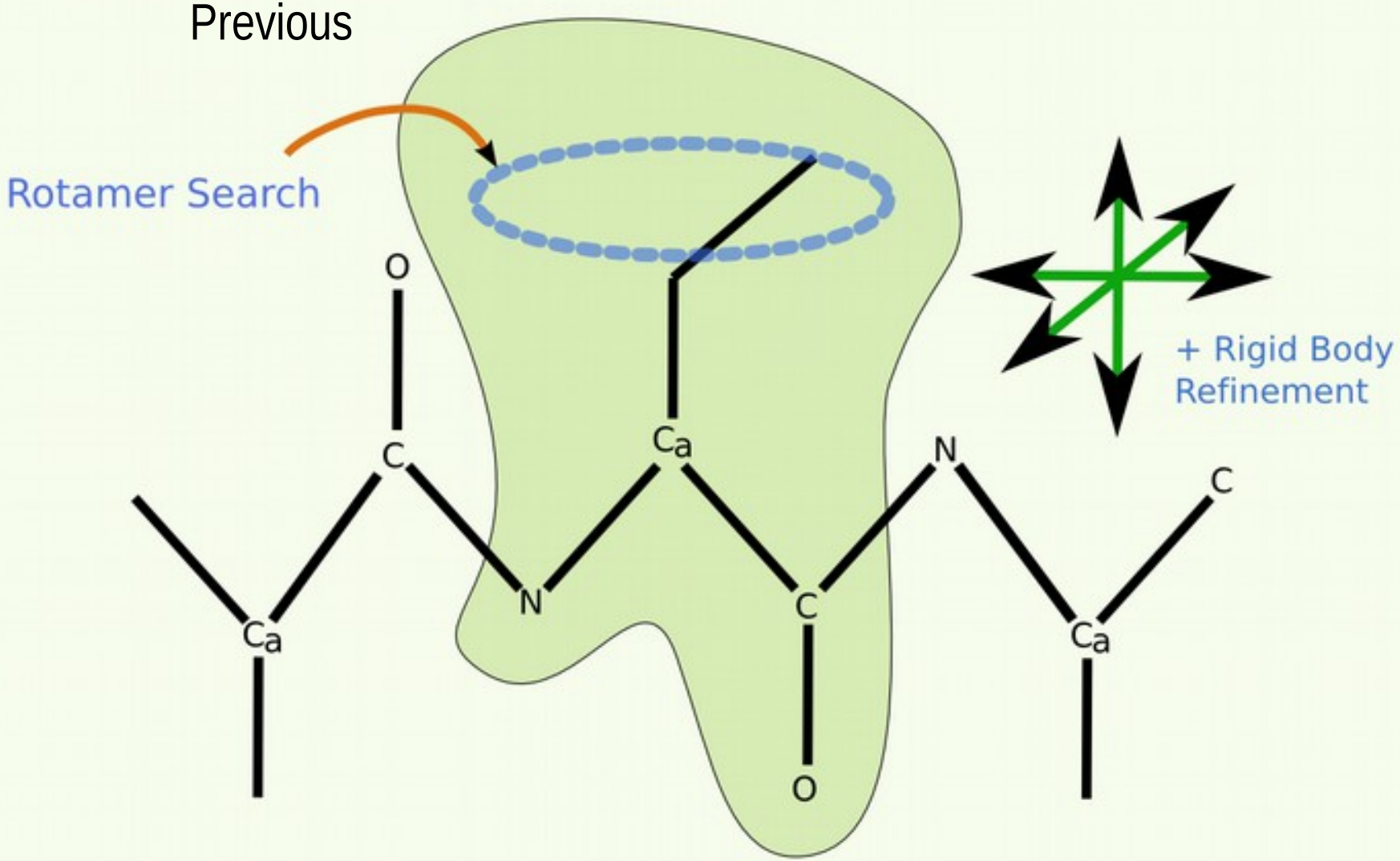
4 PHE Rotamers

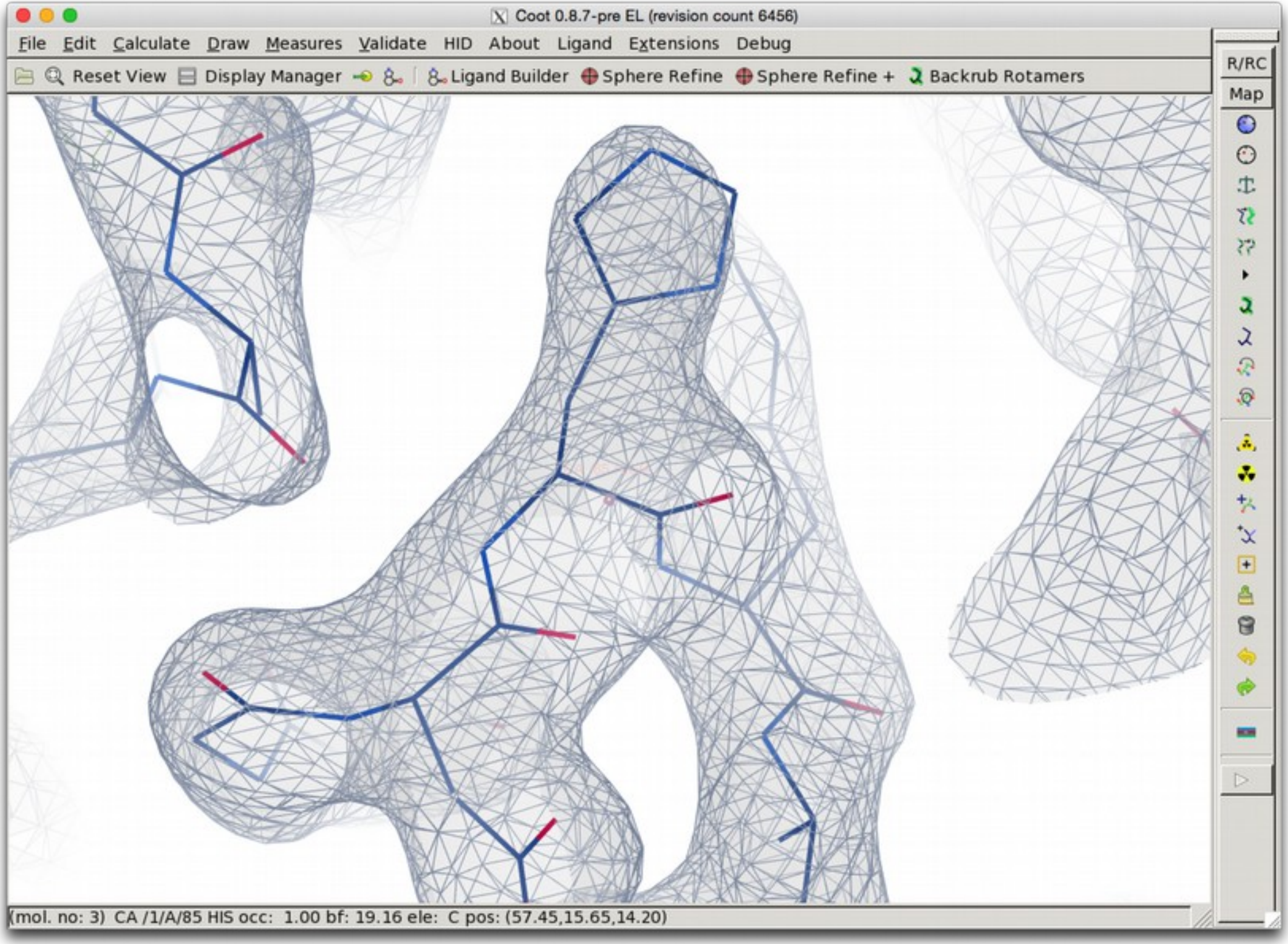


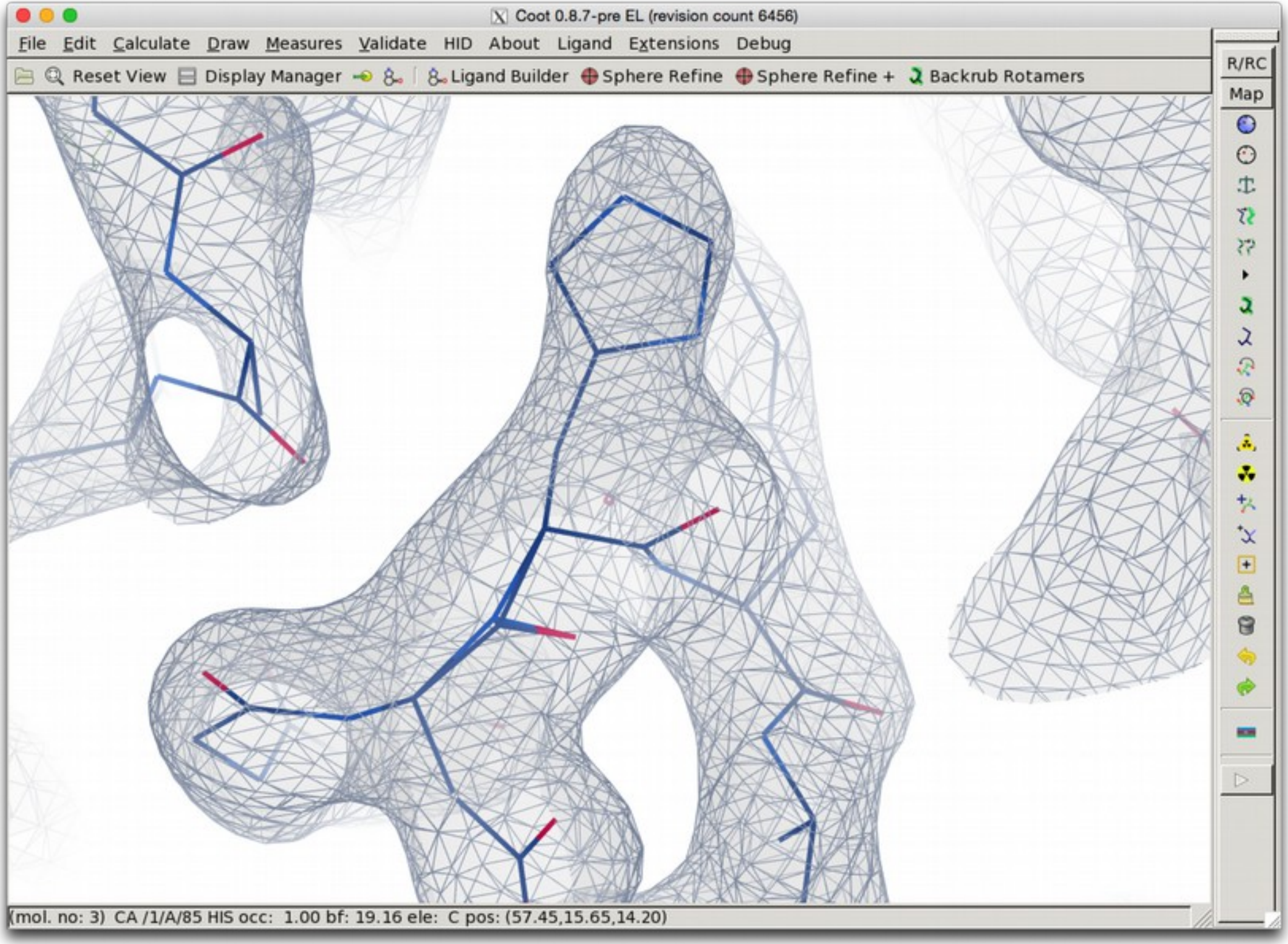


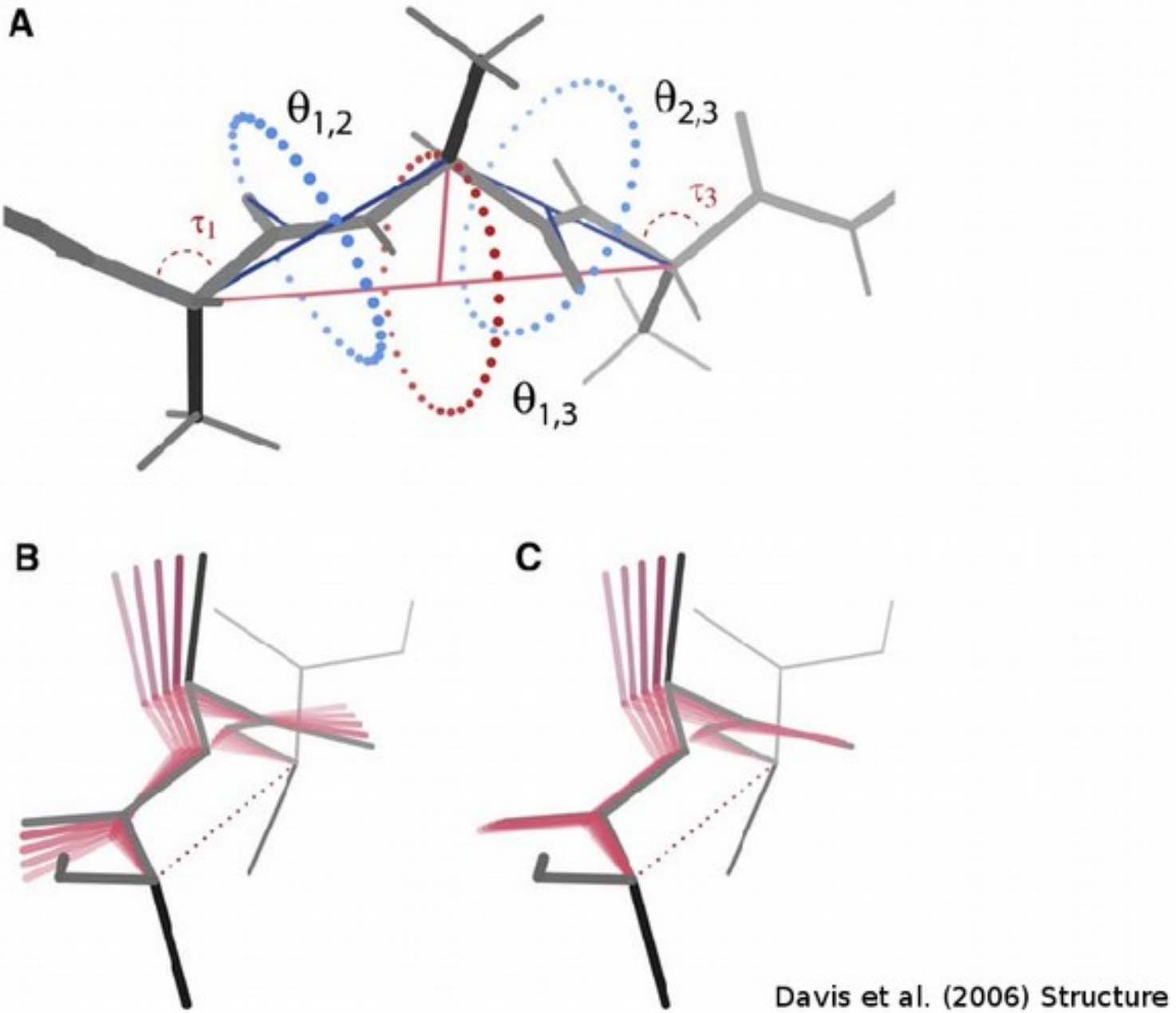
Current Low Resolution Rotamer Search

Previous



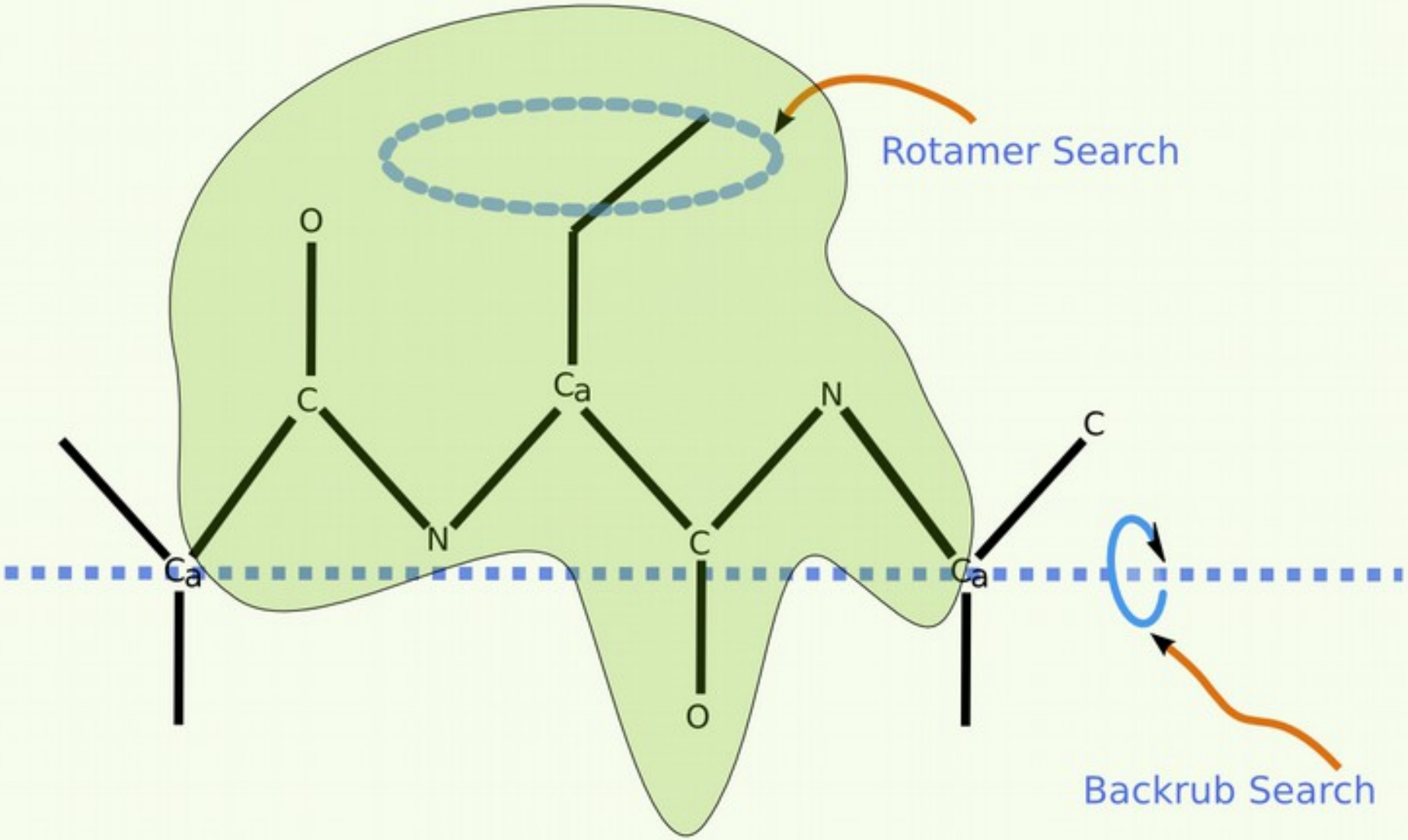

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


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)

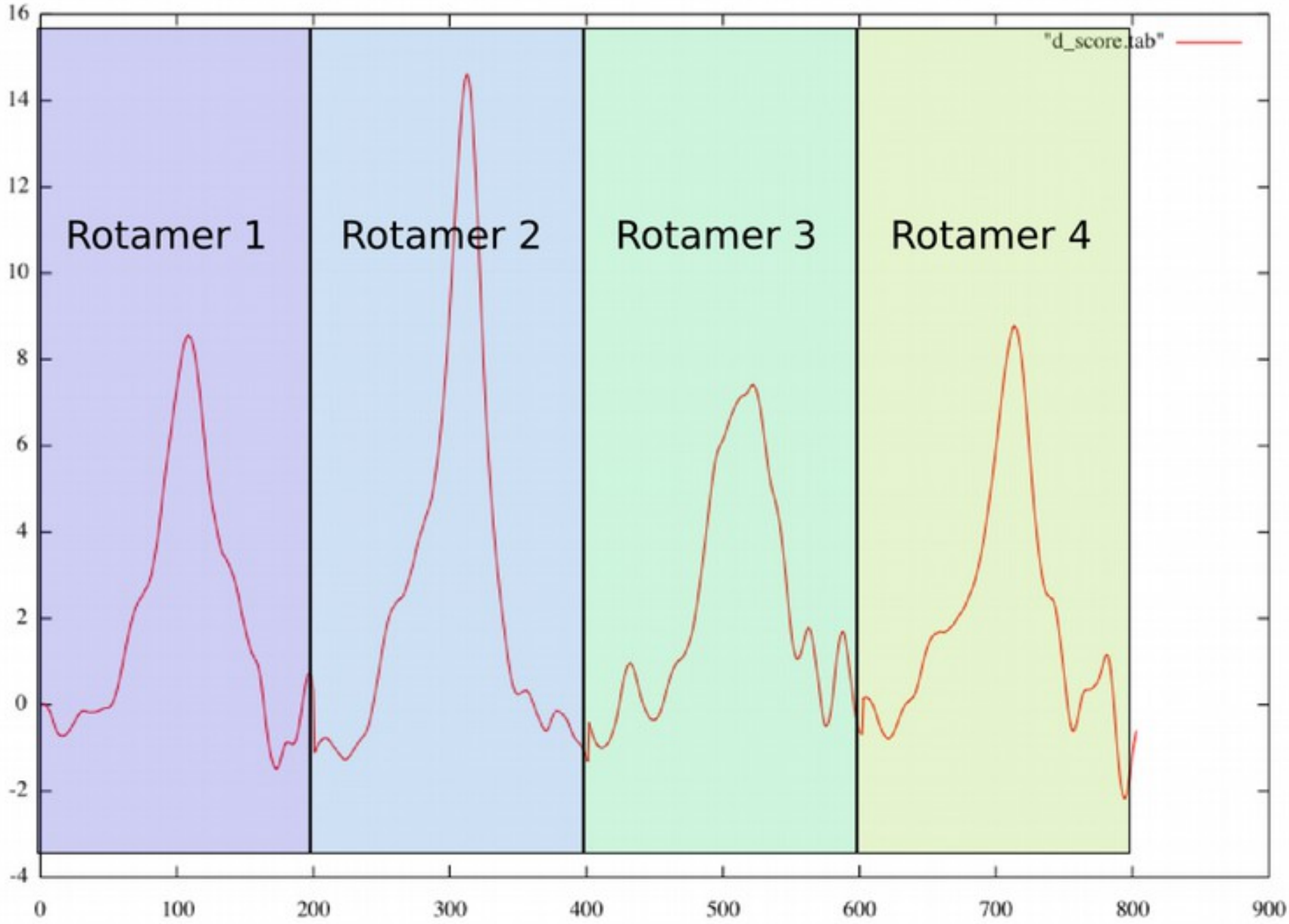


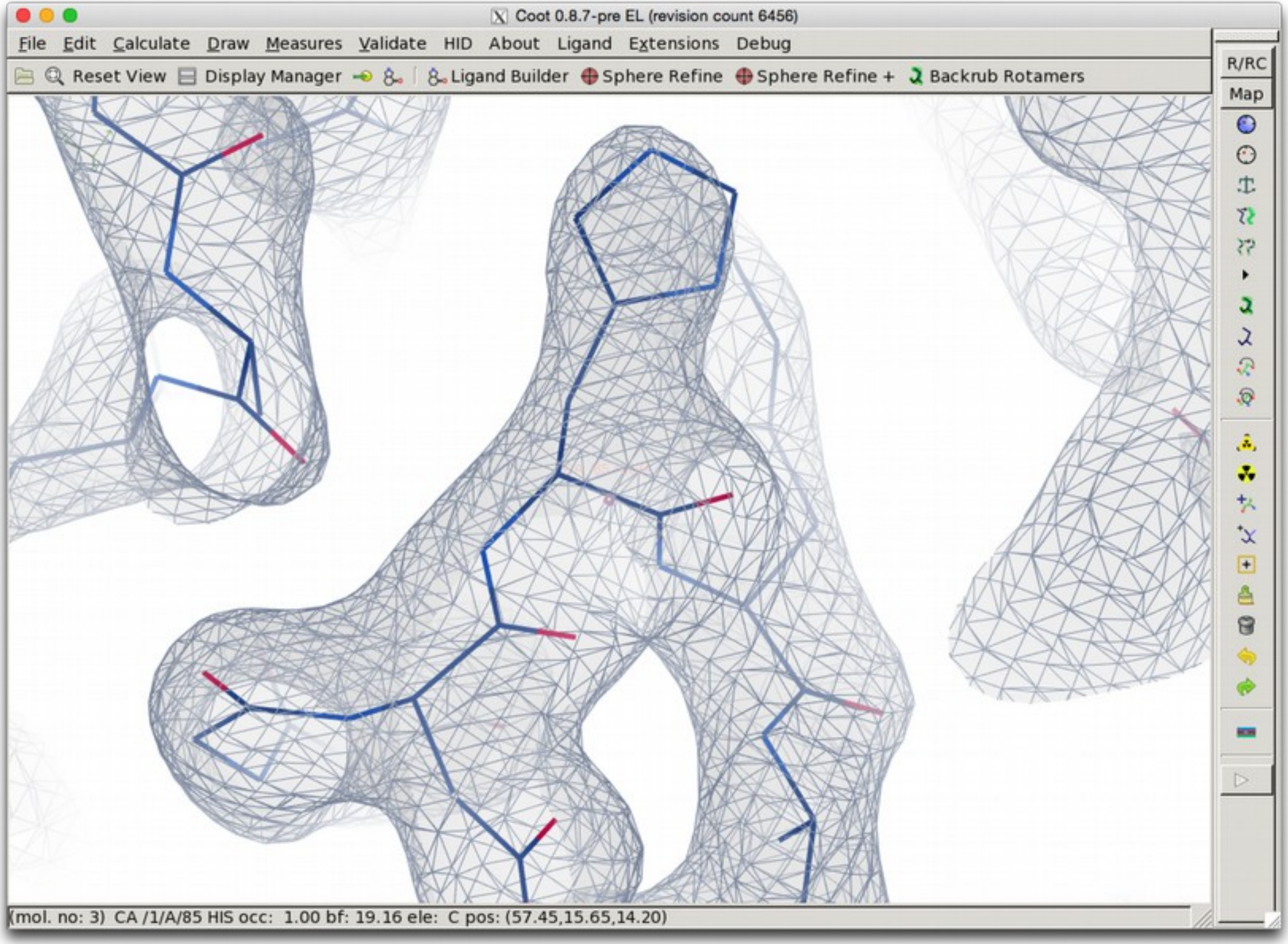
Davis et al. (2006) Structure

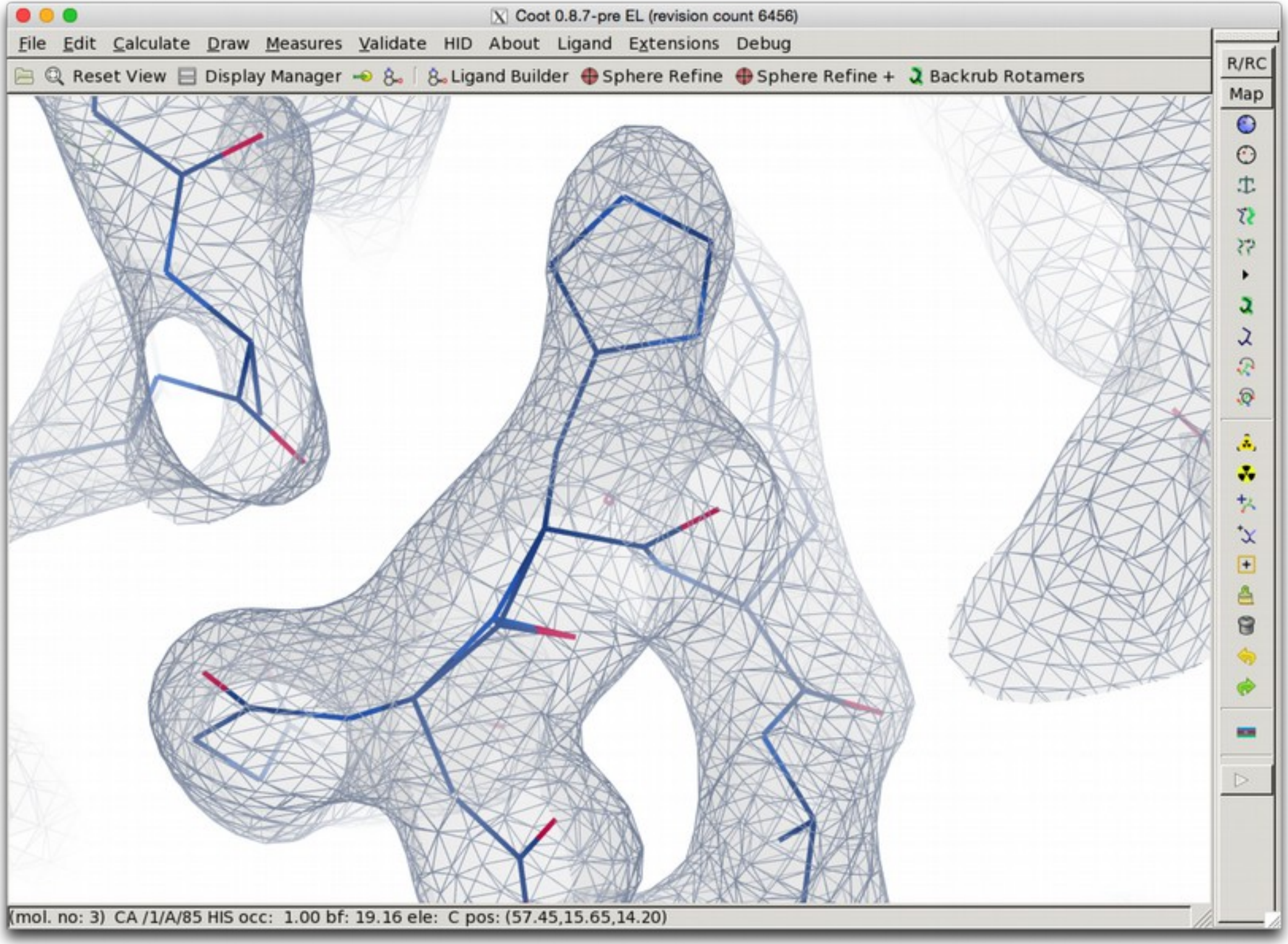
New Low Resolution Rotamer Search

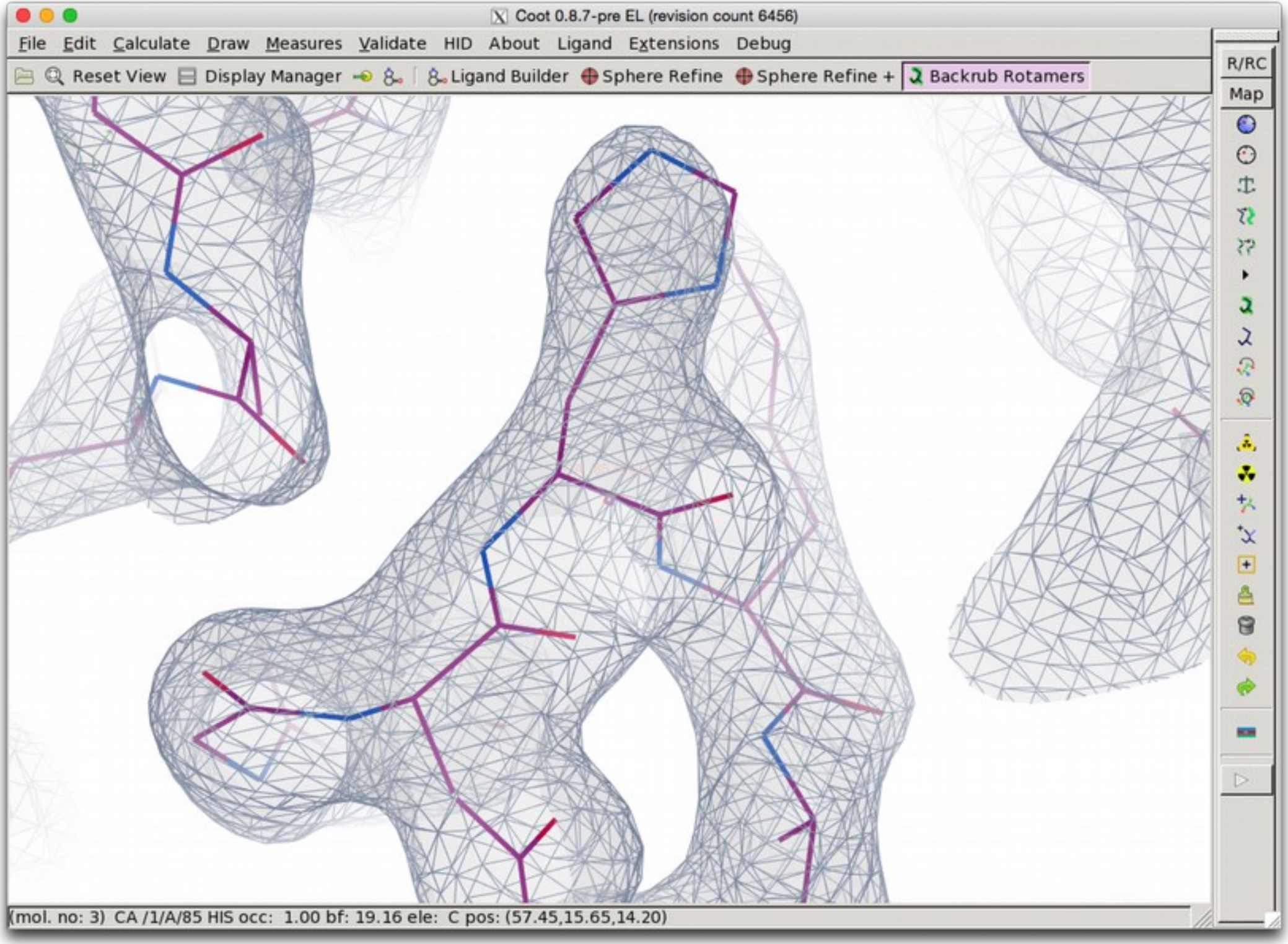


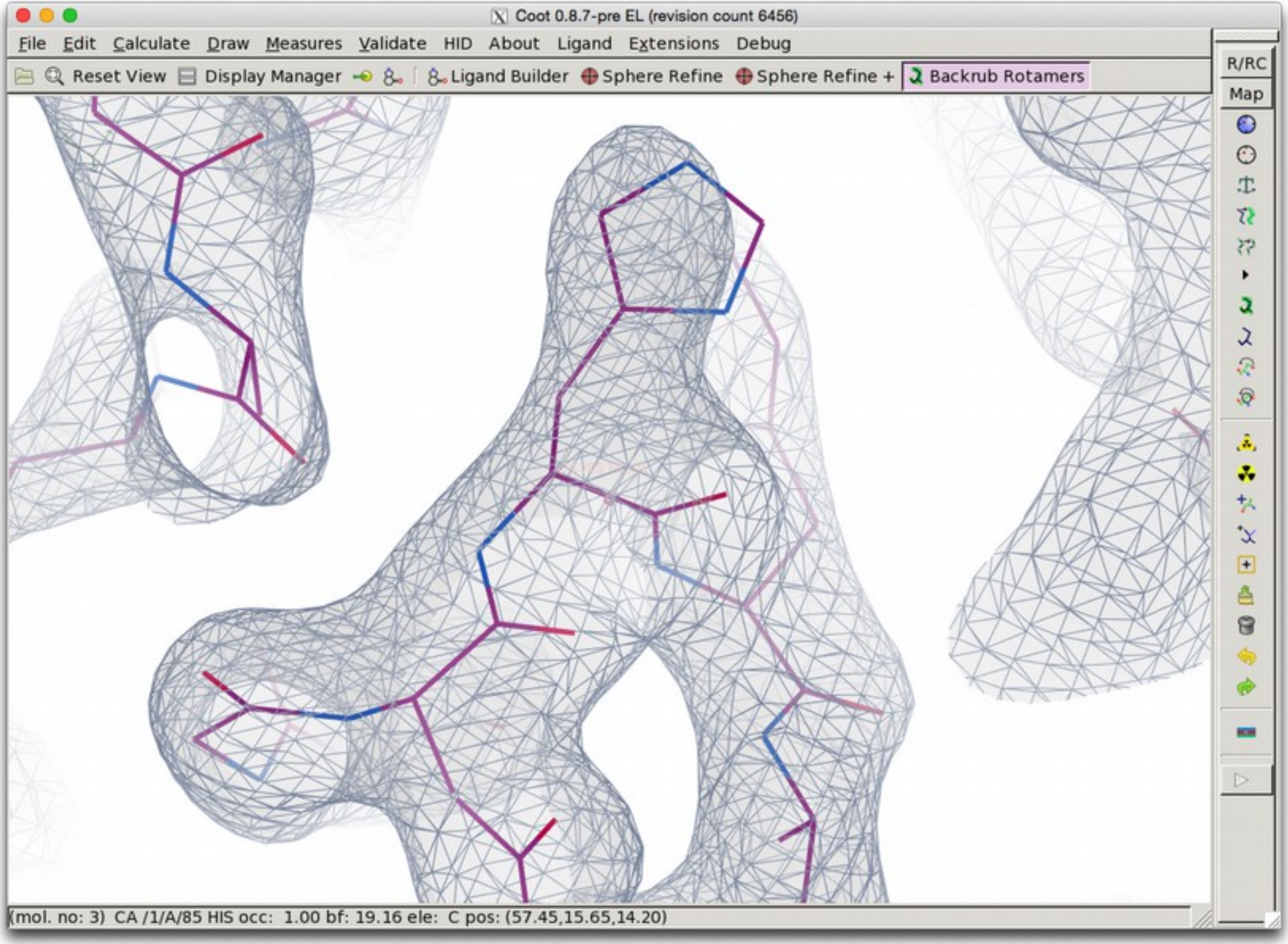
After Fitting Tools in KING/Molprobity

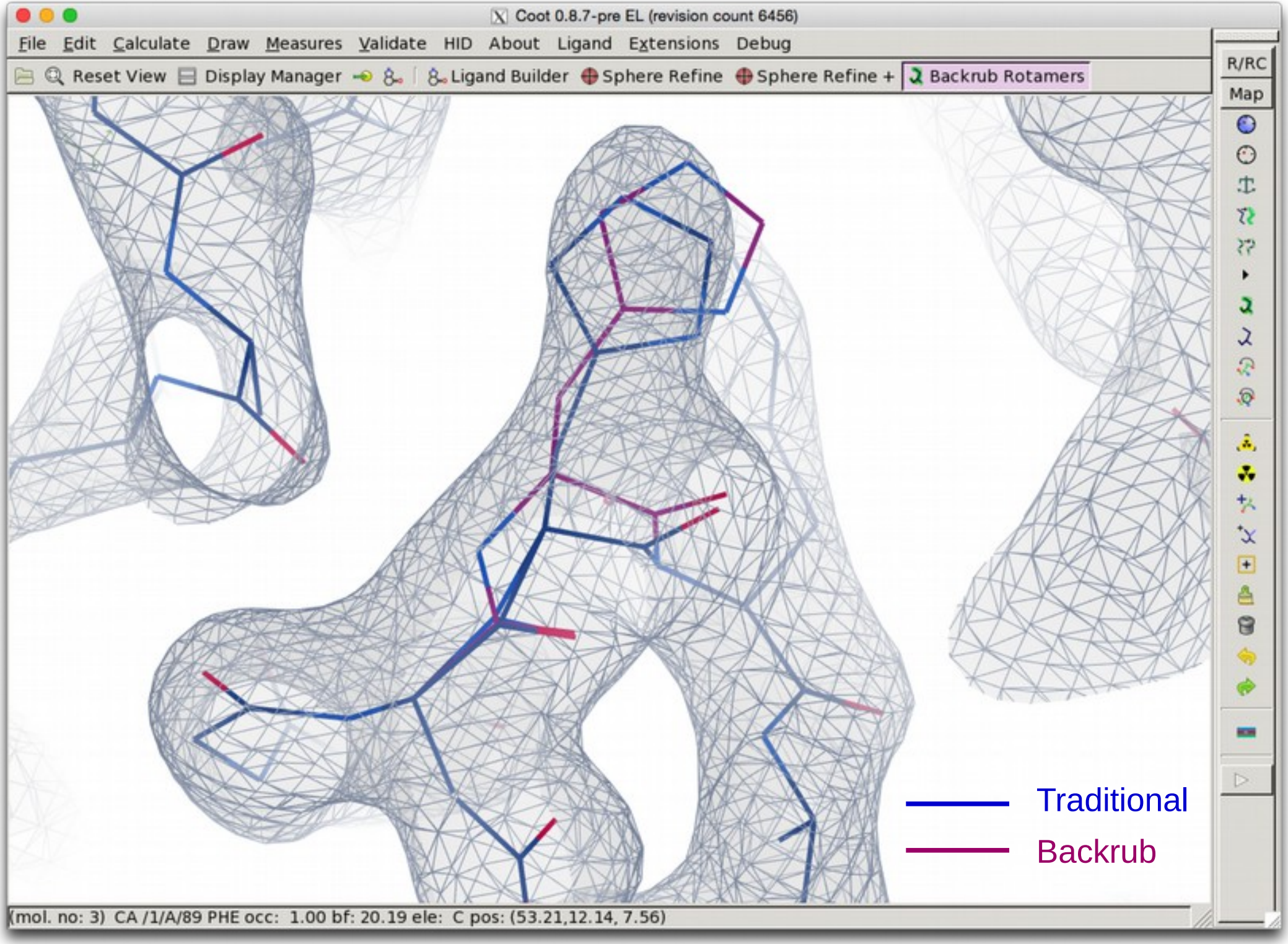

"d_score.tab"
Rotamer 1
Rotamer 2
Rotamer 3
Rotamer 4
16
14
12
10
8
6
4
2
0
-2
-4
0
100
200
300
400
500
600
700
800
900


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)

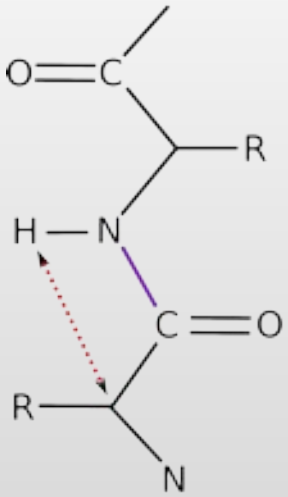
cis-Peptides

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

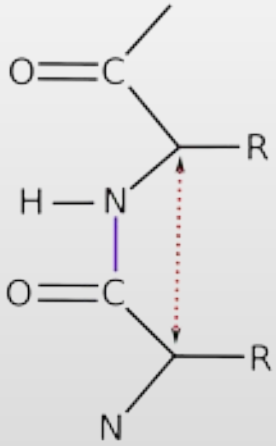
cis-Peptides

- 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

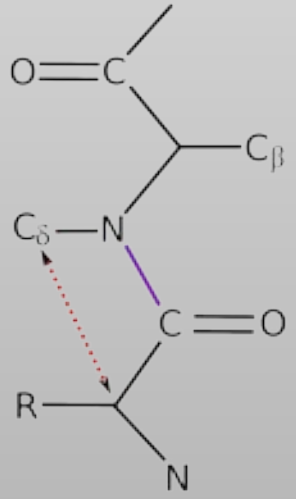
cis-Peptides



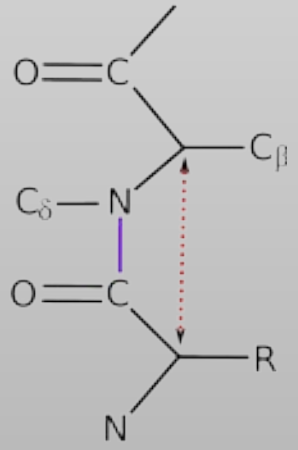
trans-peptide



cis-peptide

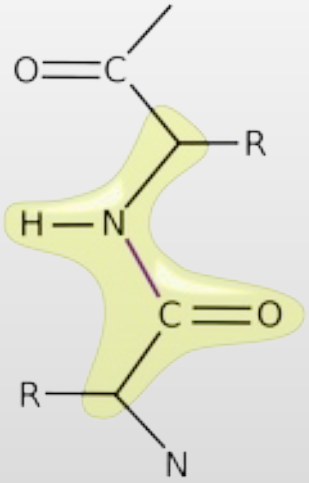


PRO *trans-peptide*

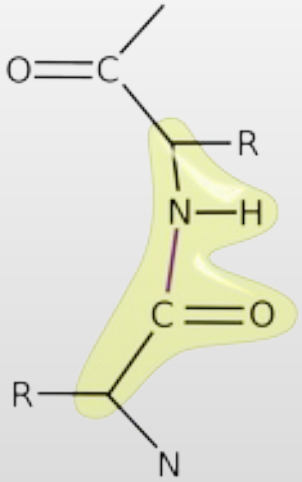


PRO *cis-peptide*

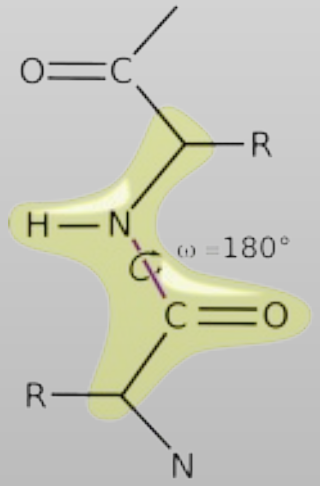
cis-Peptides



trans-peptide
with plane restraints



cis-peptide
with plane restraints



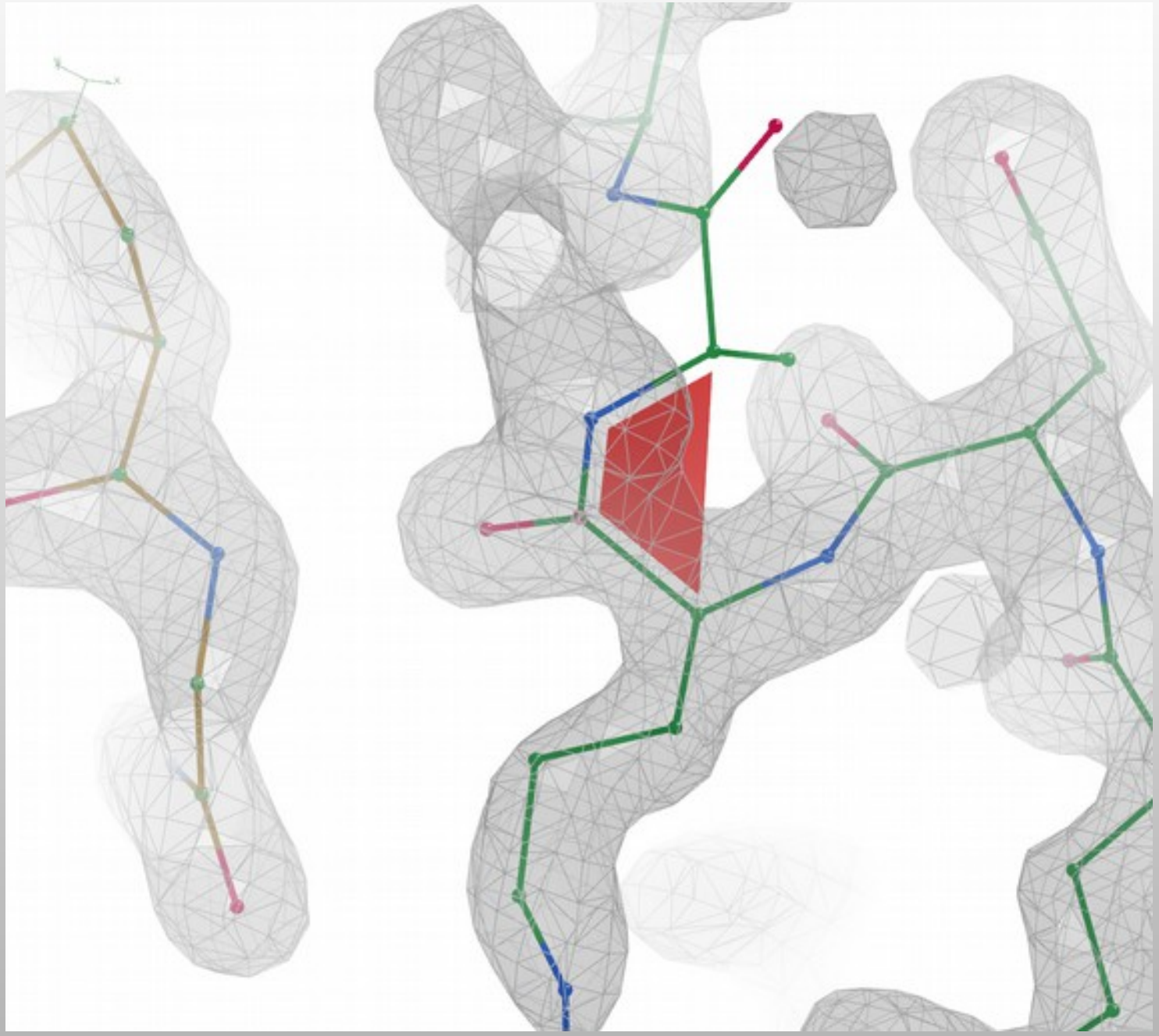
trans-peptide
with plane and trans restraints

Non-Crystallographic Symmetry

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

cis-peptide Representation



Pre-PRO

Twisted-trans

Non-pre-PRO

Handling NCS

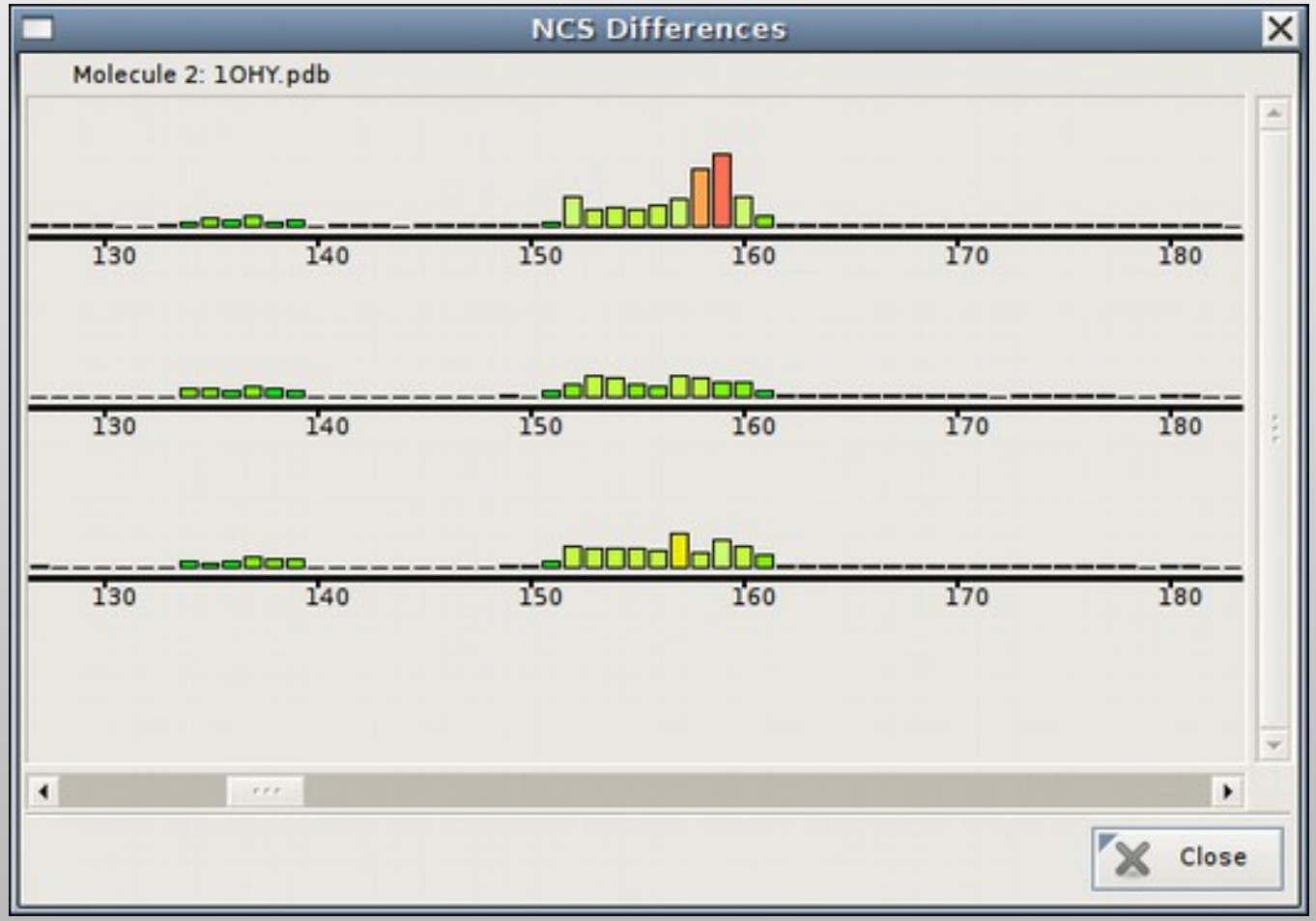
- What are the Problems?
- Strict NCS:
  - NCS should appear like crystallographic symmetry does [exact copies]
- Non-Strict NCS:
  - Molecules are different
  - How to cope with differences, but minimize unnecessary rebuilding?

Handling NCS

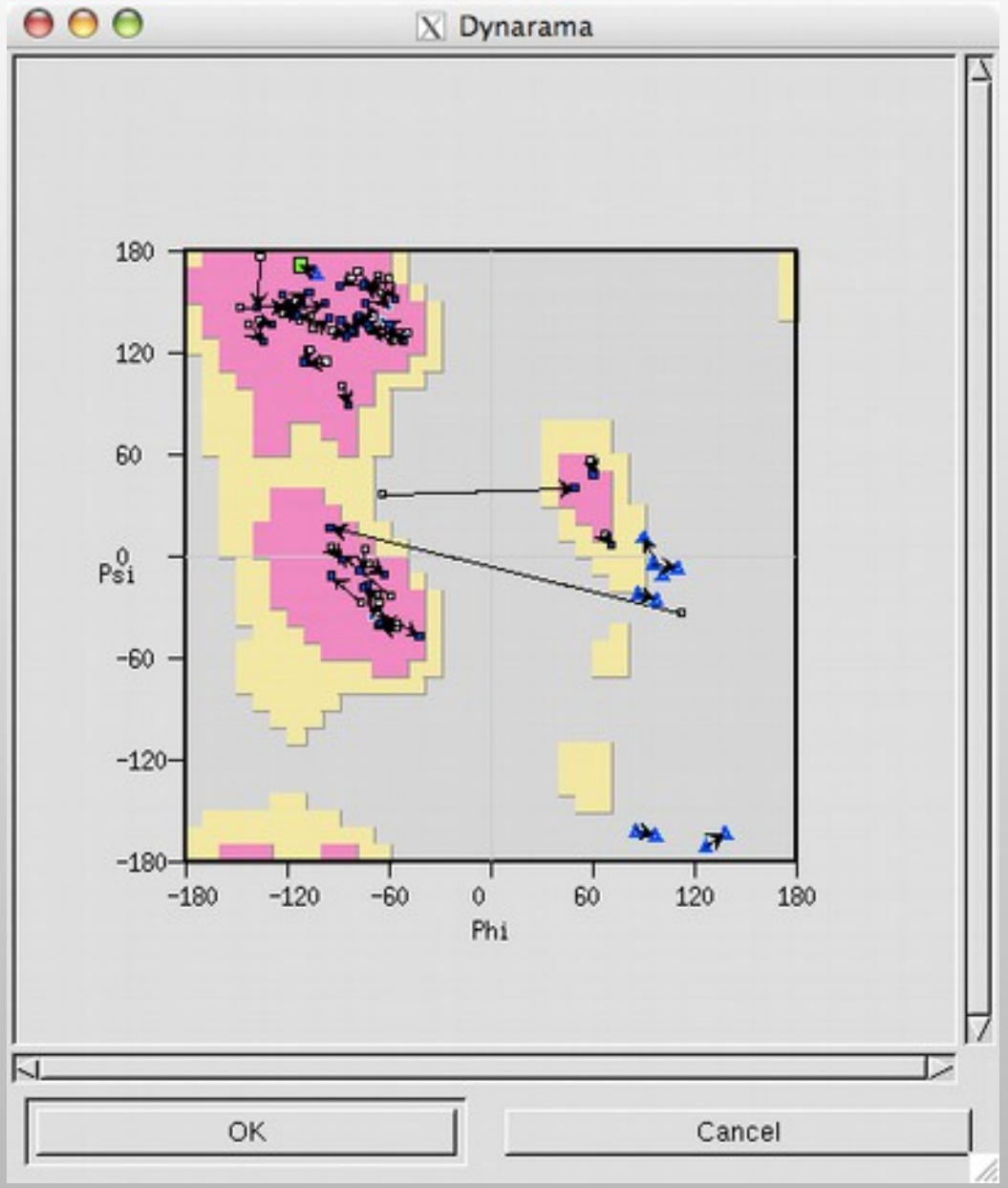
Typical Scenario:

- 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?

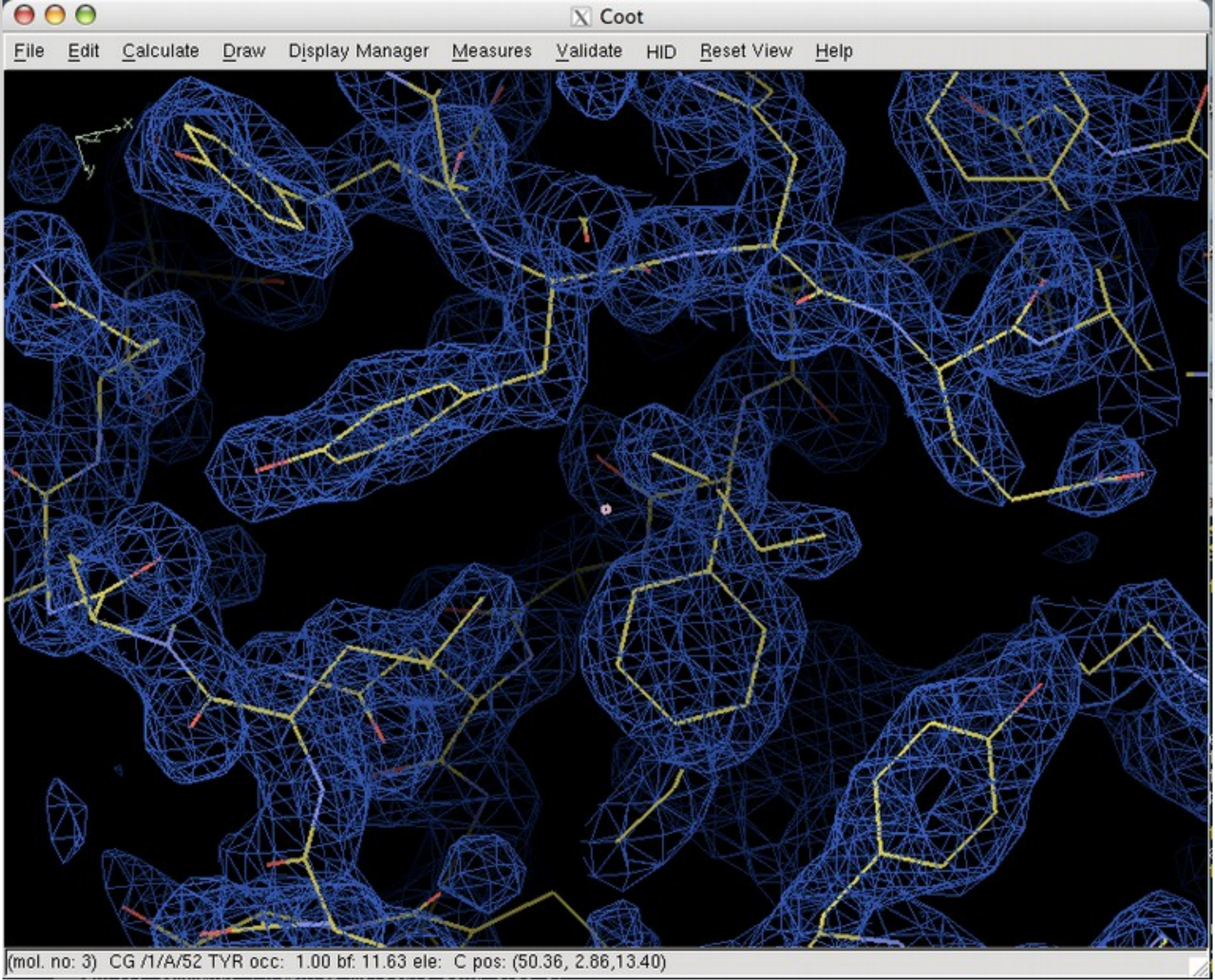
...or new NCS Differences graph



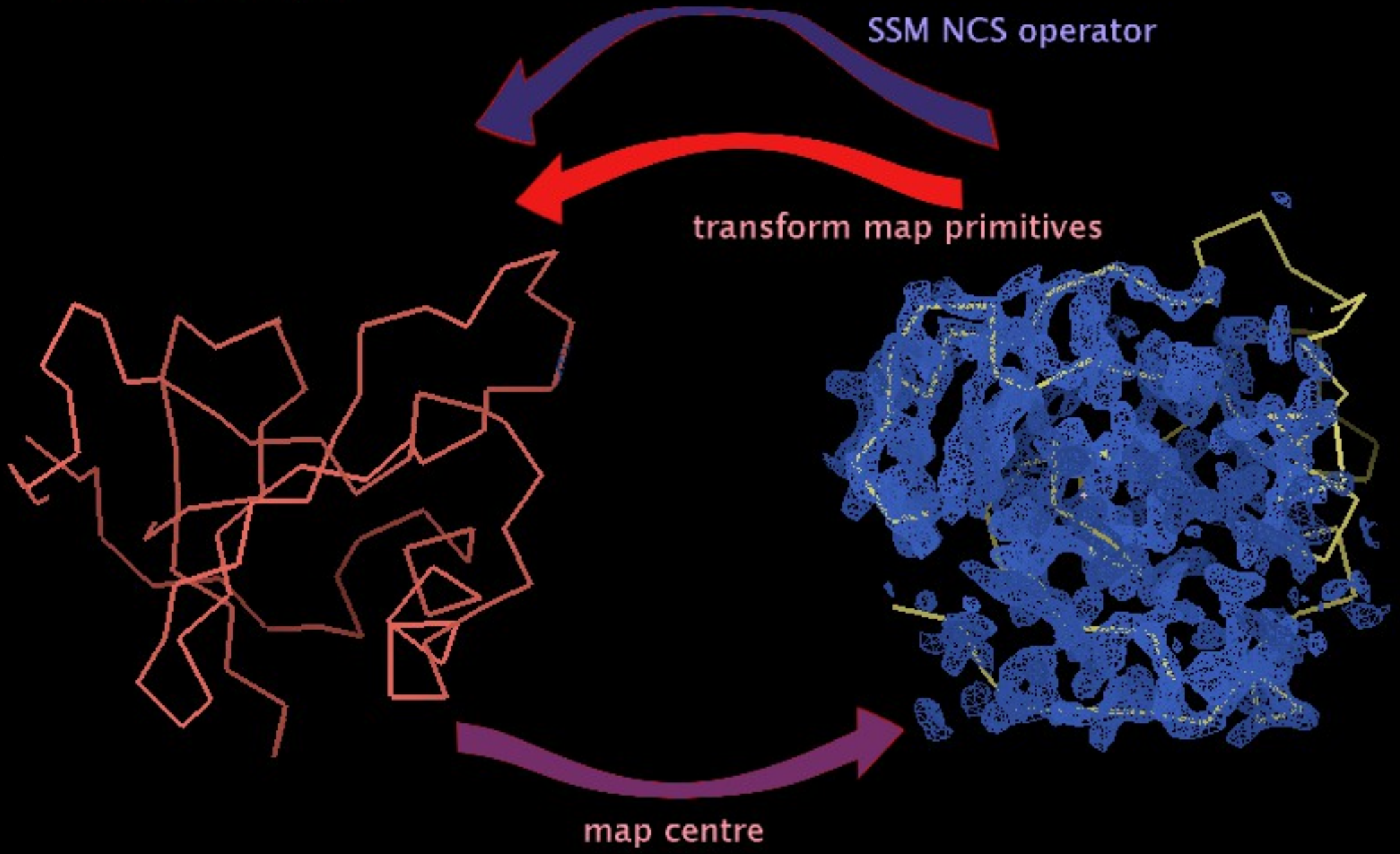
...or Kleywegt Plots[*]

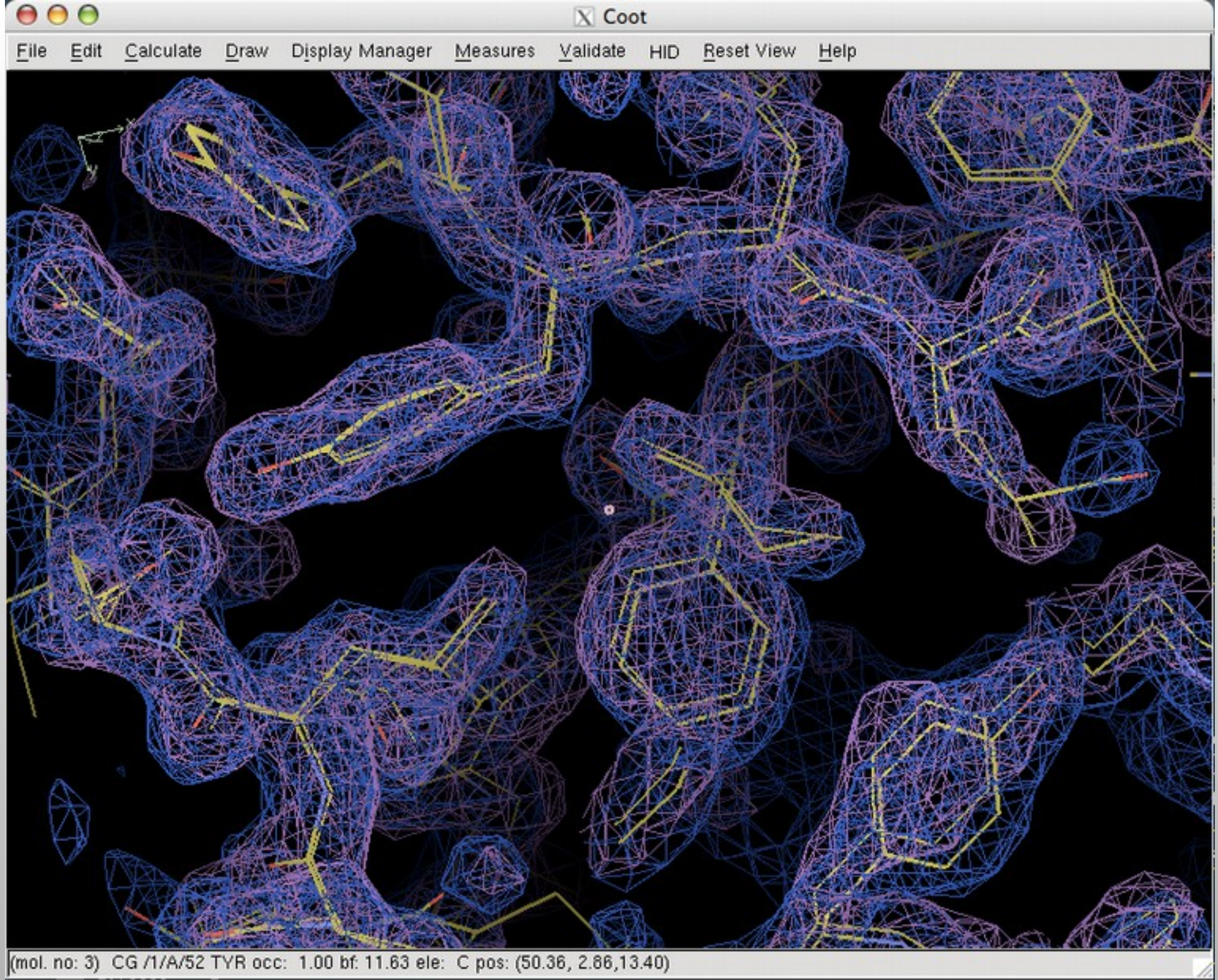


[*] Named by George Sheldrick


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



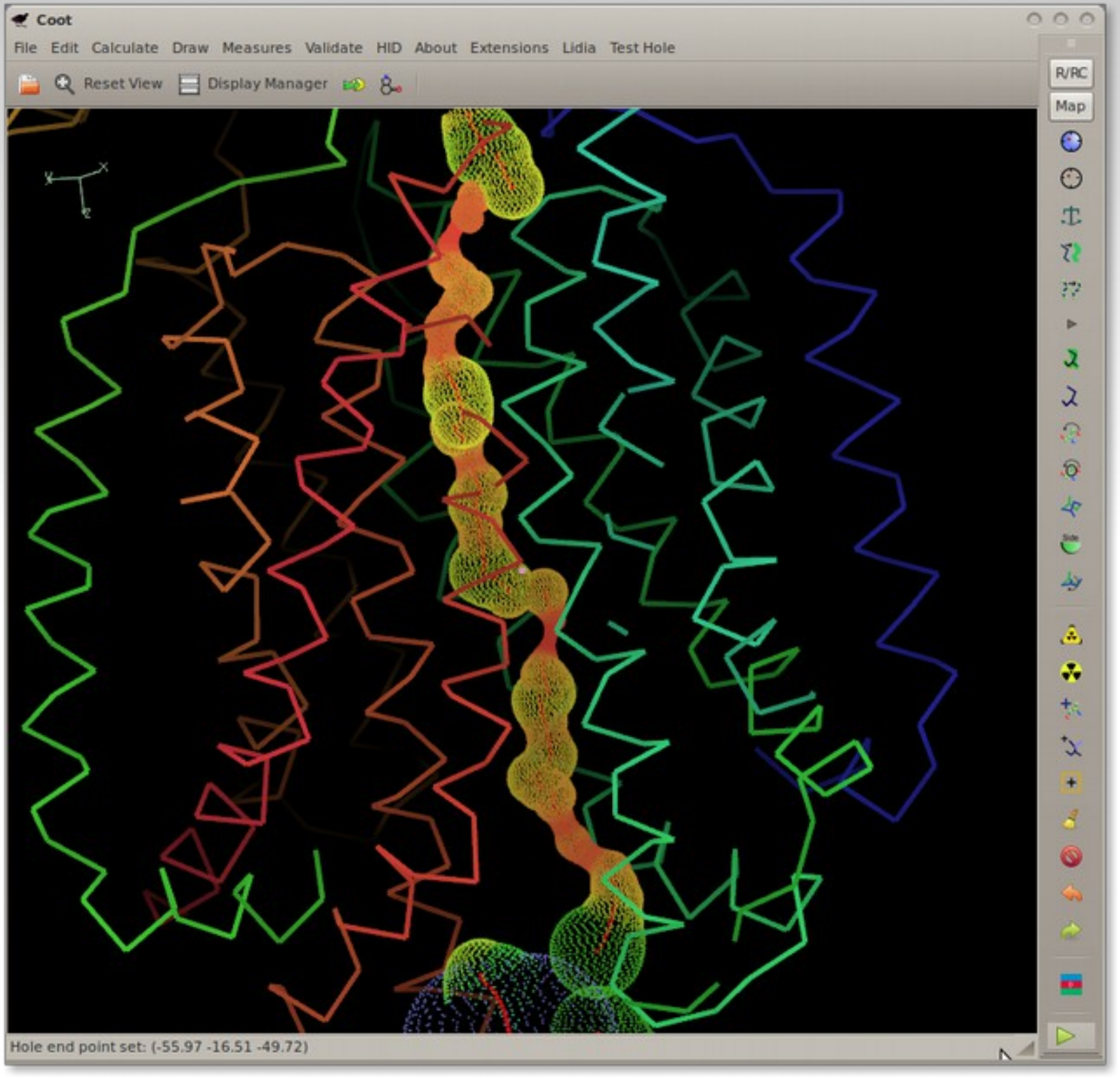

Coot
File Edit Calculate Draw Display Manager Measures Validate HID Reset View Help
(mol. no: 3) CG /1/A/52 TYR occ: 1.00 bf: 11.63 ele: C pos: (50.36, 2.86,13.40)

Alternatively...

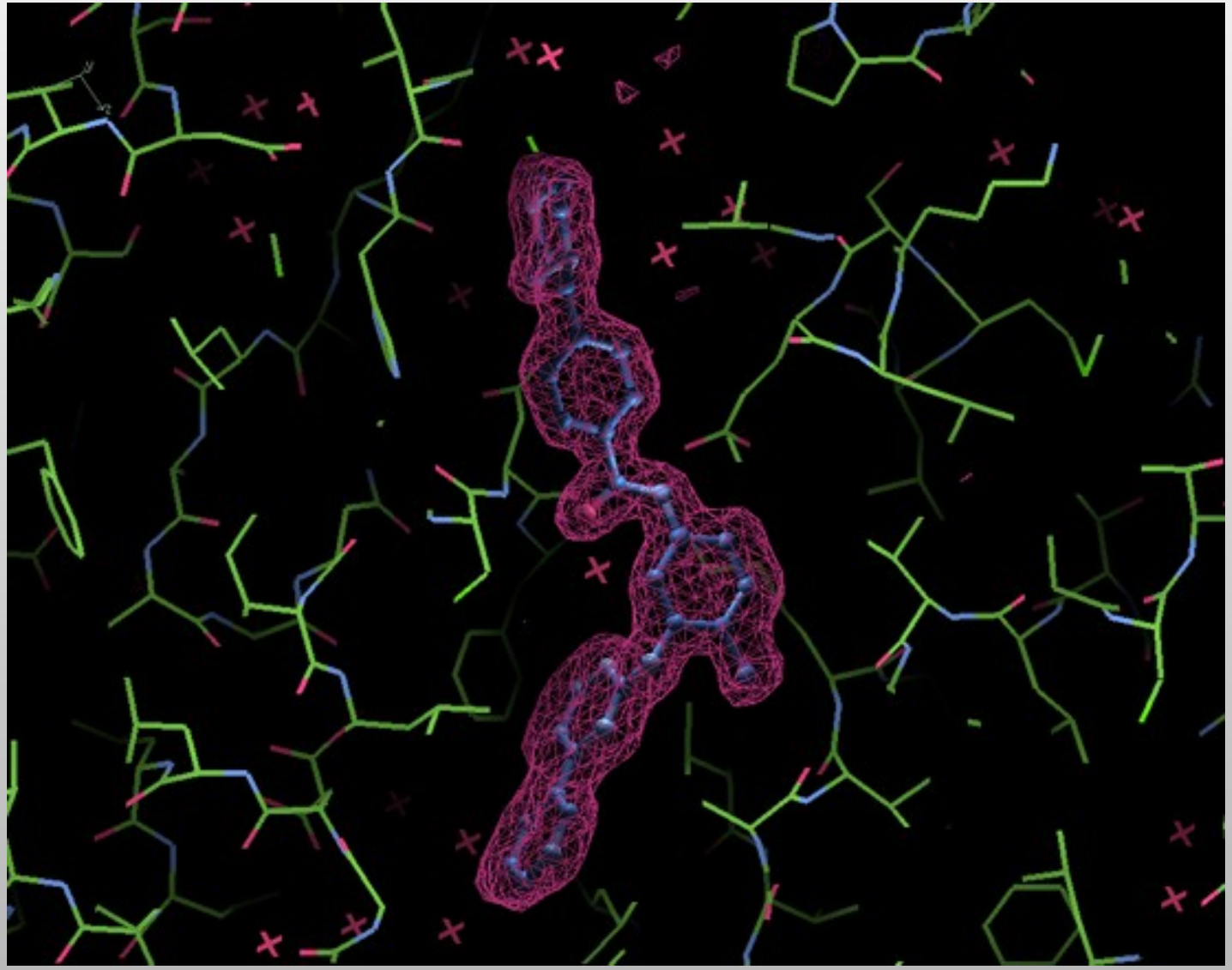
- We can handle NCS using “NCS Jumping”
- *<see the tutorial>*

Finding Holes

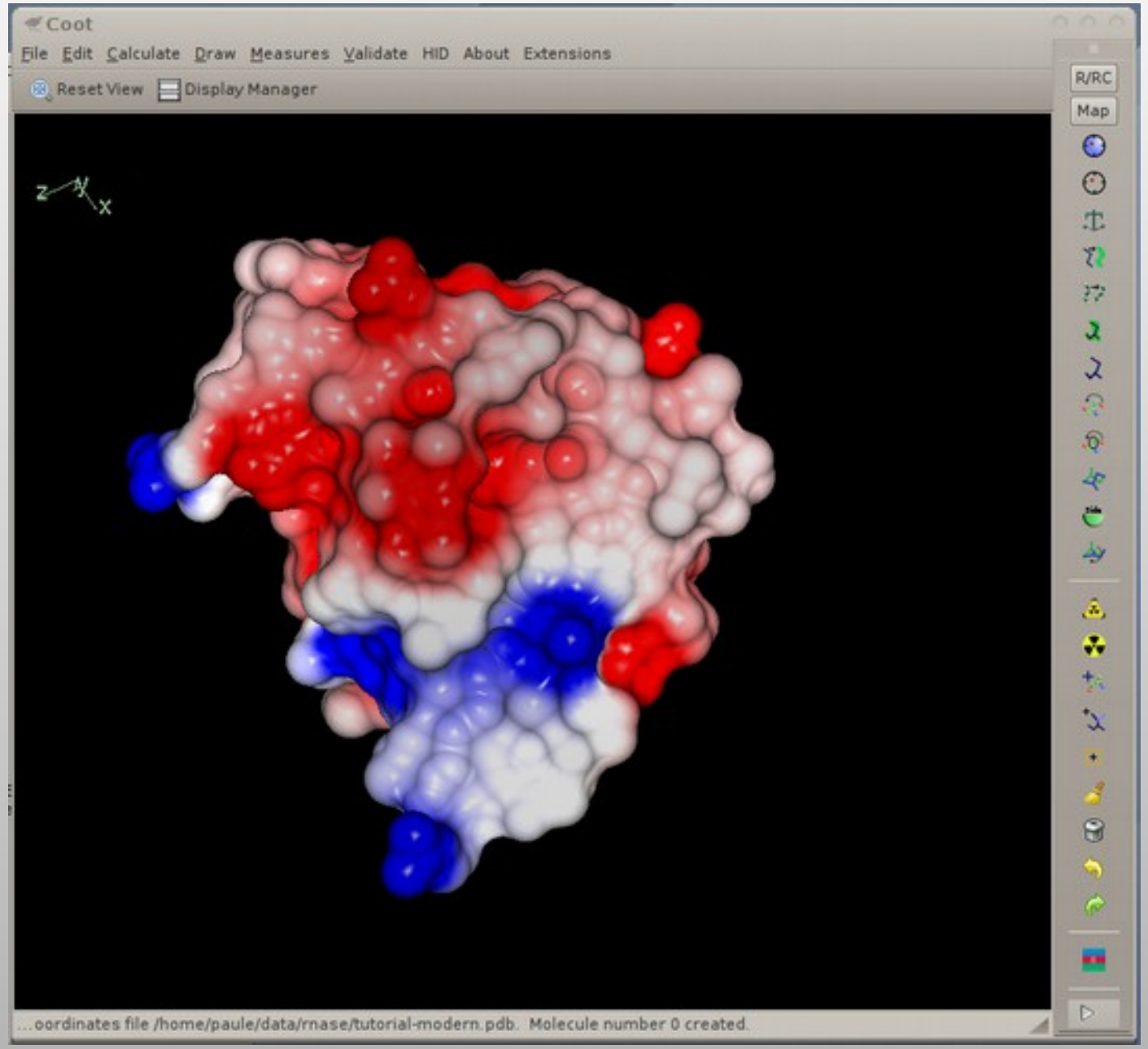
- An implementation of
  - Smart, Goodfellow & Wallace (1993) Biophysics Journal **65**, 2455
  - Atomic radii from AMBER
  - I used
    - radii from CCP4 monomer library
    - sans simulated annealing


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)

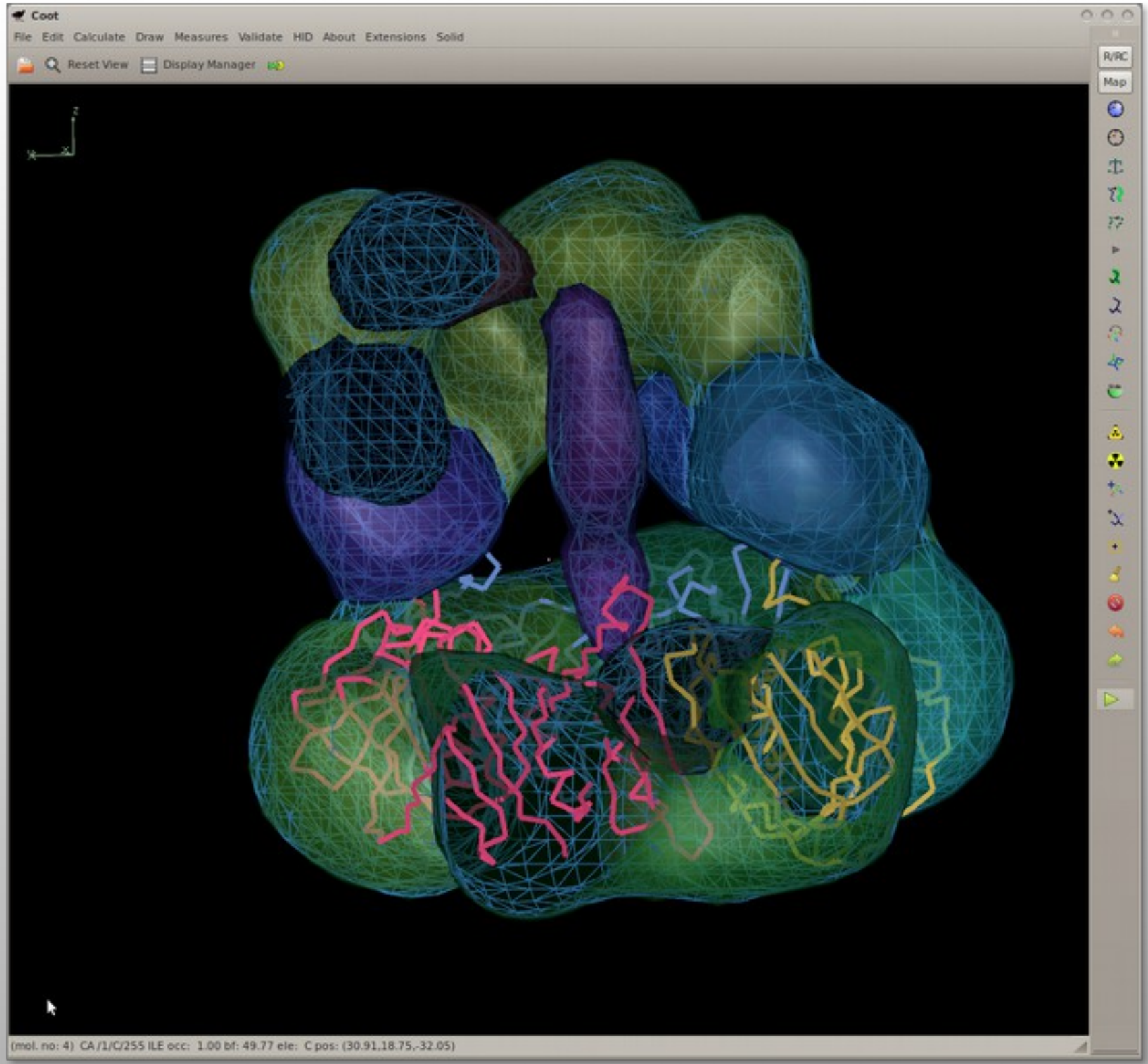
Some Representation Tools

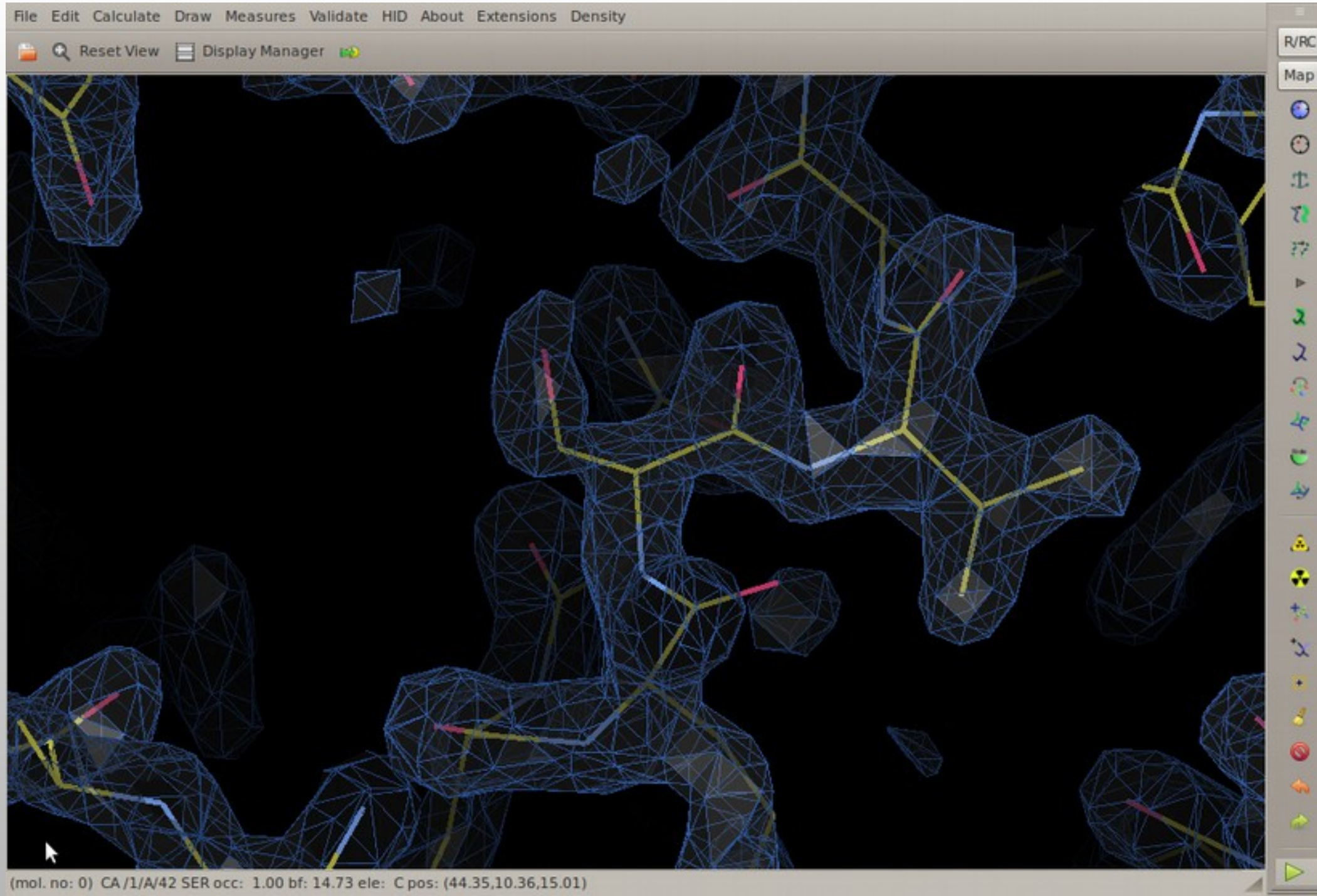


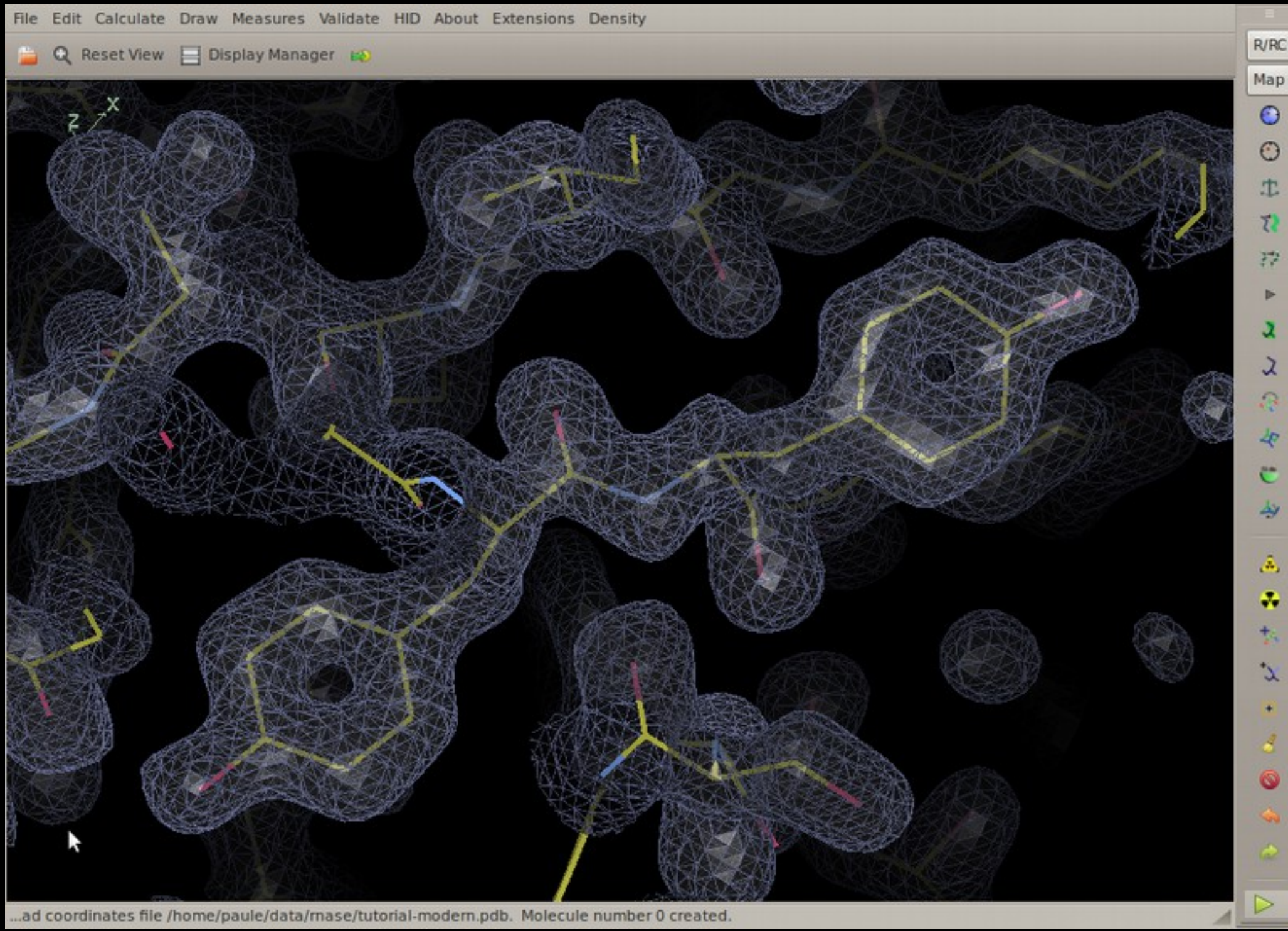
Some Representation Tools



Gruber & Noble (2007)


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)


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.

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...

Using *Coot*

Mouse clicks and motion

- Left-mouse click and drag
  - → rotate the view
- Right-mouse click and drag
  - → zoom in
- Middle-mouse click
  - → label atom
- Middle-mouse scroll
  - → change map contour level

More mousing

- Left-mouse double click
  - → label atom
- Ctrl left-mouse drag
  - → drag view/translate

Ctrl Shift scroll middle-mouse

- → change representation style

Ctrl Right-mouse drag

- change depth cue (up/down)
- translate in screen z (left/right)

Button presses...

- c: toggle cross-hairs
- d & f: depth cueing
- i: toggle spin/rock
- <Shift> l: label atom
- m & n: zoom
- o and <Shift> O: Other NCS chain
- p: “intelligent” nearest atom
- v: undo symmetry view

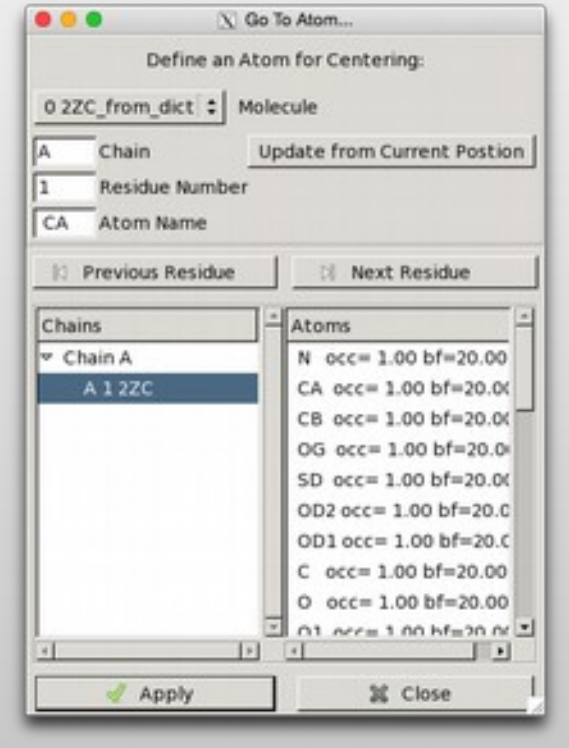
Ctrl Button presses

- Ctrl s: quick save-as
- Ctrl z: undo model modification
- Ctrl g: go to residue
  - Opens a small dialog box
  - type residue number (and chain id)
  - or residue triple *e.g.* HGR

Ctrl + <Arrow key>

- Ctrl + Arrow key:
  - Translates active residue
- Ctrl + Shift + Arrow key:
  - Rotates active residue

IISTDTIDIW



- If I See This Dialog Then I'm Doing It Wrong

Setup X11 for MacOSX

- In the X11 settings:
  - Emulate 3-Button mouse
  - Focus Follows Mouse

Usage Tips: Navigation

- (Type) Ctrl-G <*resno*>: → recentres on <*resno*>
- “G”: go to blob
- “P”: go to nearest atom (or CA if the residue if the residue is a standard AA)
- “L” (unlabel) the atom at the screen centre
- “Go To Ligand” (icon) → Jump to (next) ligand
- “V”: Undo symmetry view

Standard extra button presses

- e: flip residue
- g: go to blob
- h, <shift> h, r, <shift r>, t, x: forms of refine and regularize
- j: auto-fit rotamer
- k and <shift> k: kill and fill side-chain
- q: flip peptide
- y: add peptide

Usage Tips: Use Keybindings

(Noob → Pro)

- Built-in: “A”, “B”, “C”, “D” “F”, “Ctrl-Z”, “G”, “I”, “M”, “N”, “O”, “P”, “S”, “U”,“Ctrl-Z”, “<space>”
- (My personal set): “Shift A”, “Shift B”, “E”, “Shift E”, “H”, “J”, “Shift J”, “K”, “Shift K”, “Shift M”, “Shift P”, “Q”, “Shift Q”, “R”, “Shift R”, “T”, “Shift T”, V”, “Shift V”, “W”, “Shift W”, “X”, “Y”, “Shift 4”, “<bar>”, “<stroke>
  - (You can download these from the Coot Wiki)

Using NVIDIA Cards (with a Linux kernel)

- For antialiasing:
  - setenv __GL_FSAA_MODE *n*
  - export __GL_FSAA_MODE=*n*
  - where *n* is 5 (or so)

Acknowledgements

- LMB:
  - Garib Murshudov, Rob Nicholls, Fei Long
- Kevin Cowan, Bernhard Lohkamp
- Libraries & Dictionaries:
  - Jane & Dave Richardson
  - Alexei Vagin
  - Eugene Krissinel