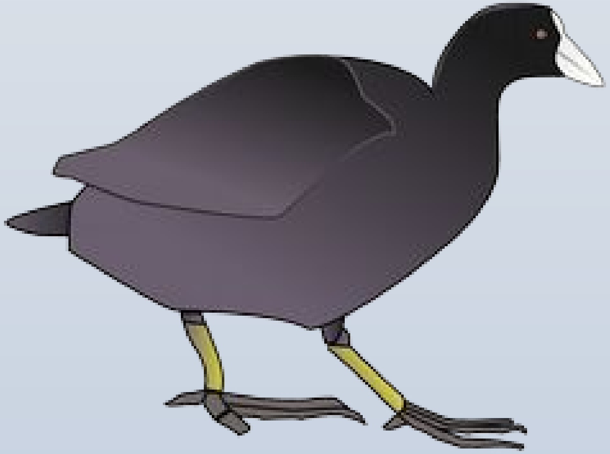
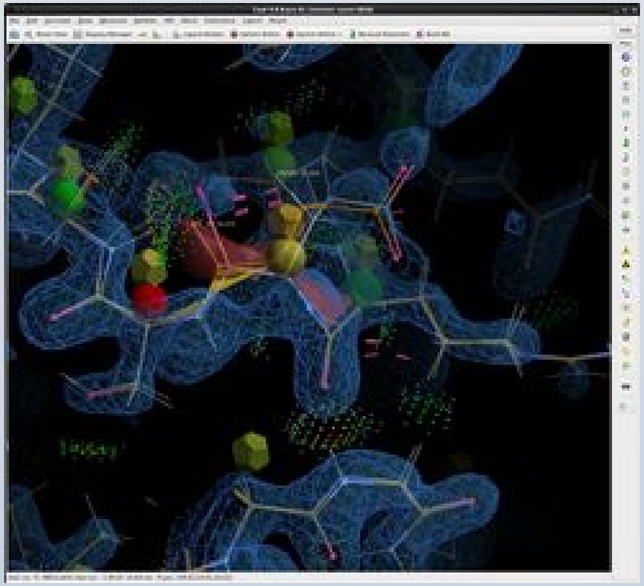
Cryo-EM Model-Building with Modern *Coot*



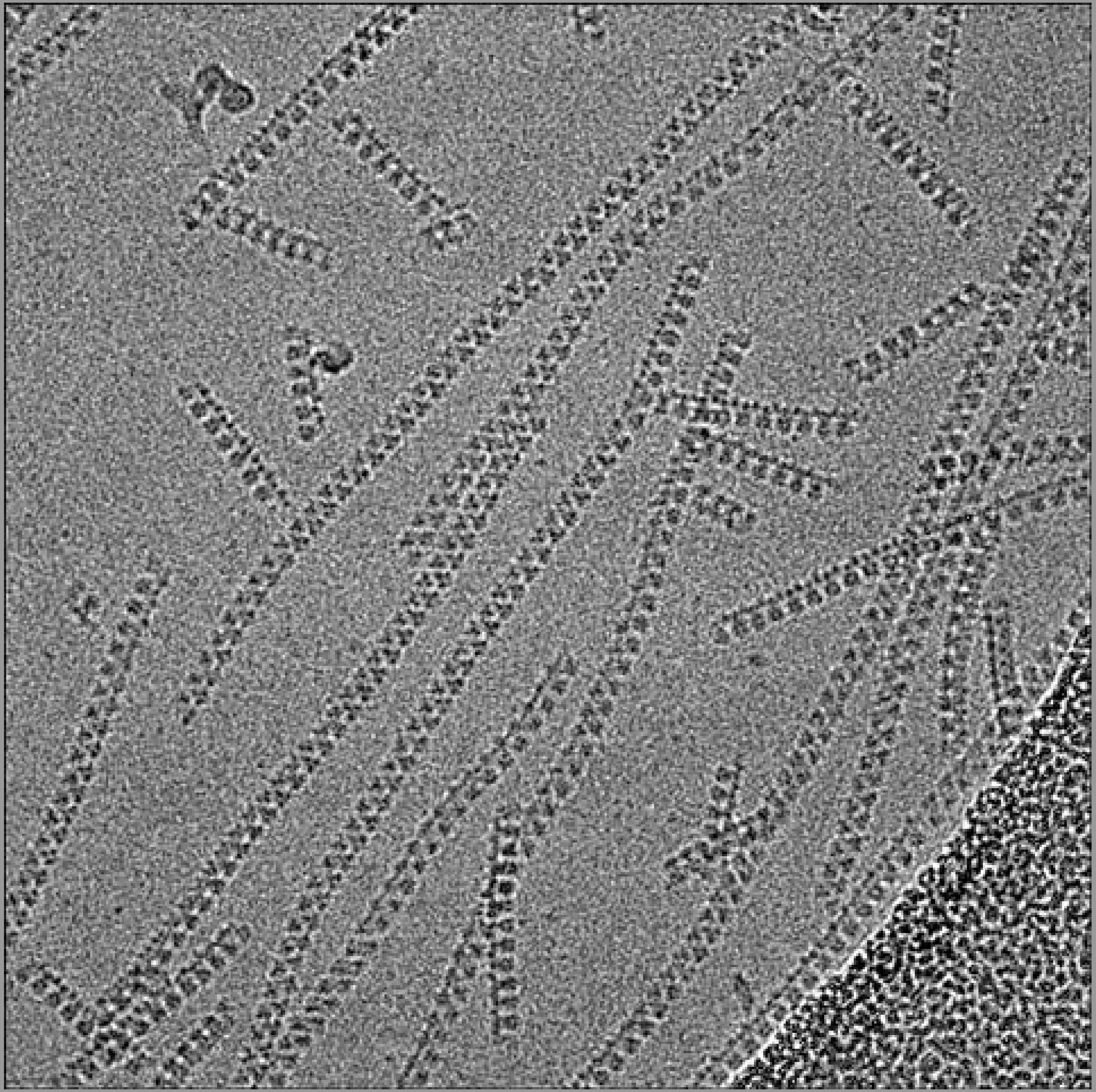


Paul Emsley
@
Ben Gurion University
March 2020

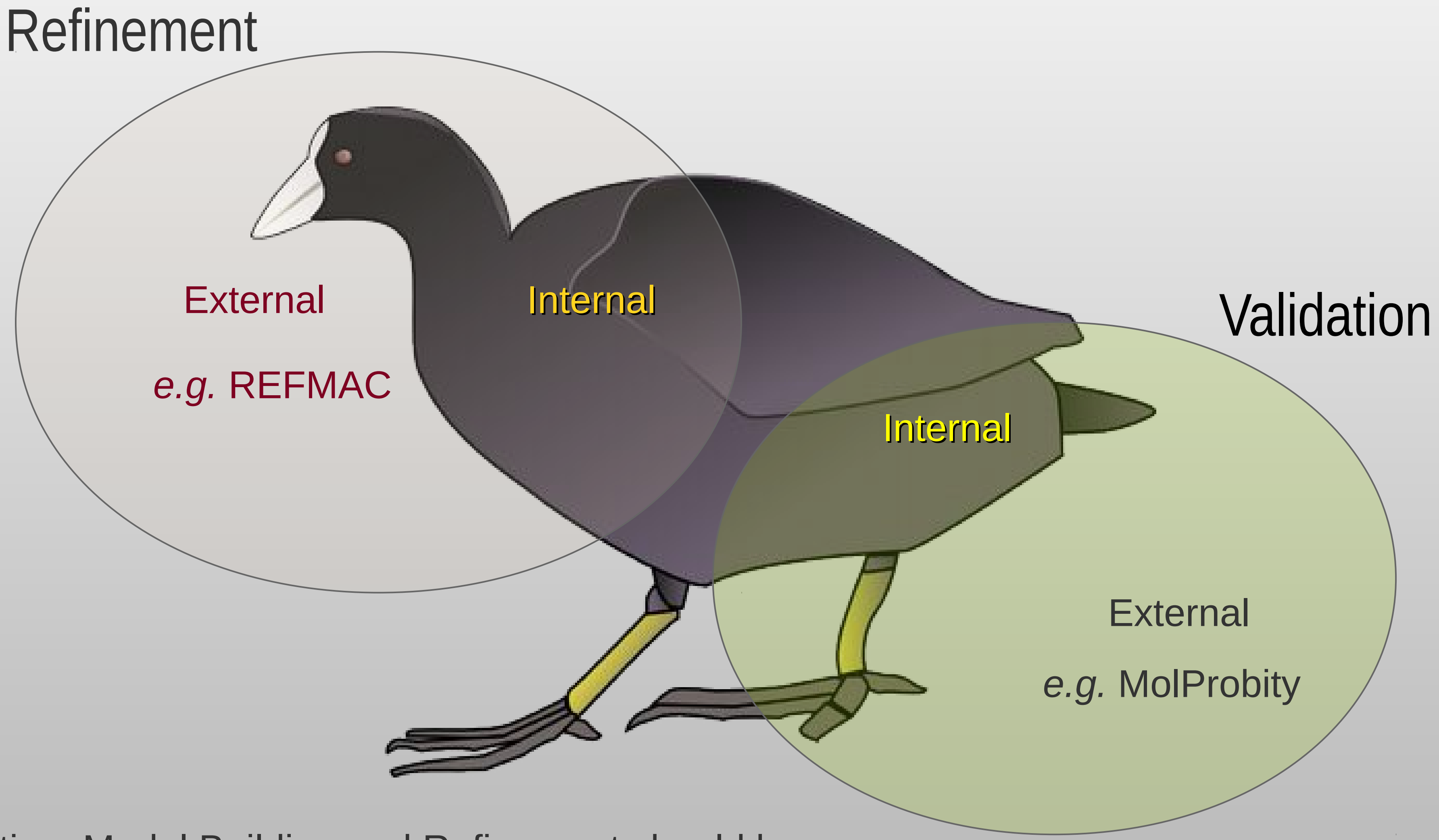
Overview

- Where does Cryo-EM Coot fit into the various builds/releases?
- What's new in 0.9 and how does it help?
  - Jiggle Fit
  - Real Space Refinement
  - Add terminal Residue
  - Fitting Helices, Merging Fragments
- Notes on Usage
- The Future: Version 1.0
  - Eye Candy

Joining the Fold



Coot: Feature Integration



Refinement

External

e.g. REFMAC

Internal

Validation

Internal

External

e.g. MolProbity

Validation, Model Building and Refinement should be used together

Comparing *Coots*

The Past

0.8 series

- Gtk+2
- Python 2
- PyGtk
- OpenGLv1

Linux ✓

Mac OS X ✓

Windows ✓

Today

0.9 series

- Gtk+2
- Python 2
- PyGtk
- OpenGLv1
- C++-11
- Boost

Linux ✓

Mac OS X ✓

Windows ✓?

The Future

1.x series

- Gtk+3
- Python 3
- OpenGLv3
- C++-11
- Boost

Linux ✓

Mac OS X ✓

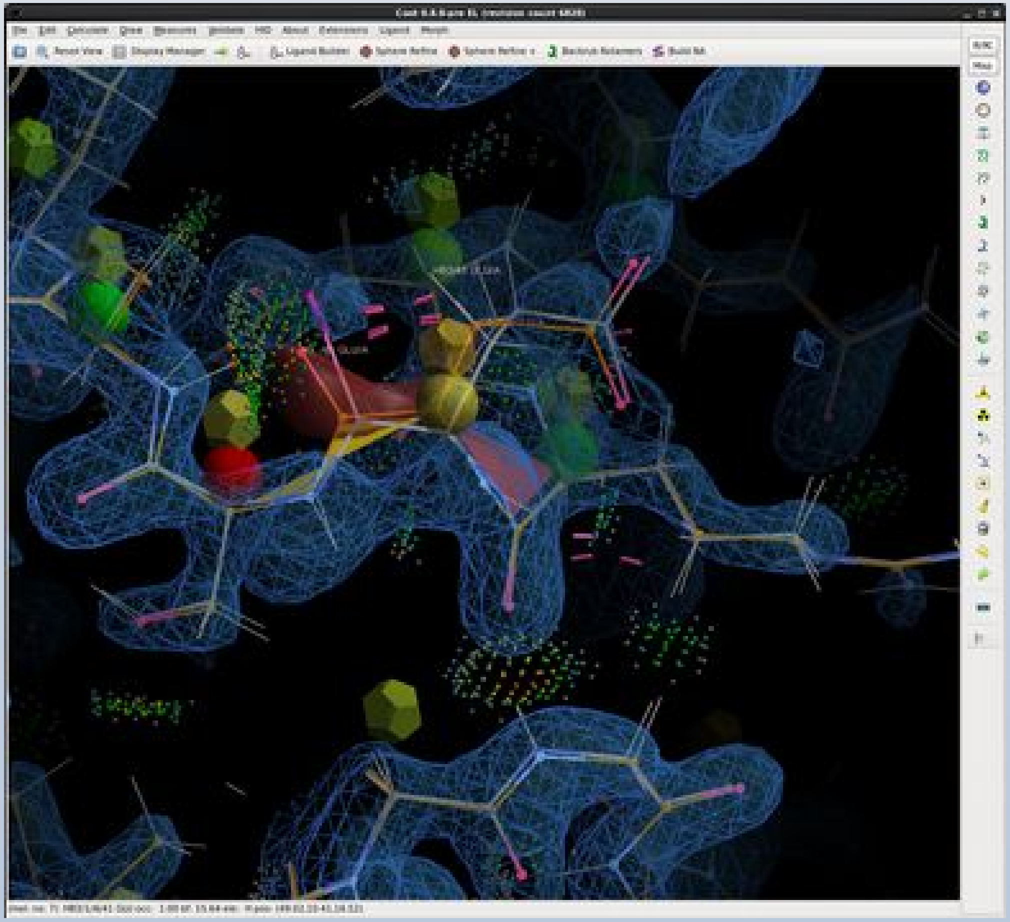
- Windows ✓?

What is *Coot* for?

- *Ab initio* model-building?
  - Yes, but it's interactive
- Low Resolution/Blobology?
  - Maybe, but that's not what it's designed for
  - (We will do something like this)
- Homology Modelling?
  - Maybe, but it uses the map when optimizing
- Average to good cryo-EM? A homolog domain that's similar to the target sequence?
  - This is the habitat of Coot:
  - 3.6Å or better

Multi-threaded Modules in 0.9

- Refinement
  - Target function and derivative evaluation, model and map all happen simultaneously now
  - Which means: more atoms, smoother updates and/or closer to the minimum
  - Refinement and Graphics updates/display are now separate threads
- Jiggle-Fit
- All-atom contact dots
- Ramachandran Score
- Rotamer Score
- Add Terminal Residue
  - ϕ,ψ hypothesis scoring

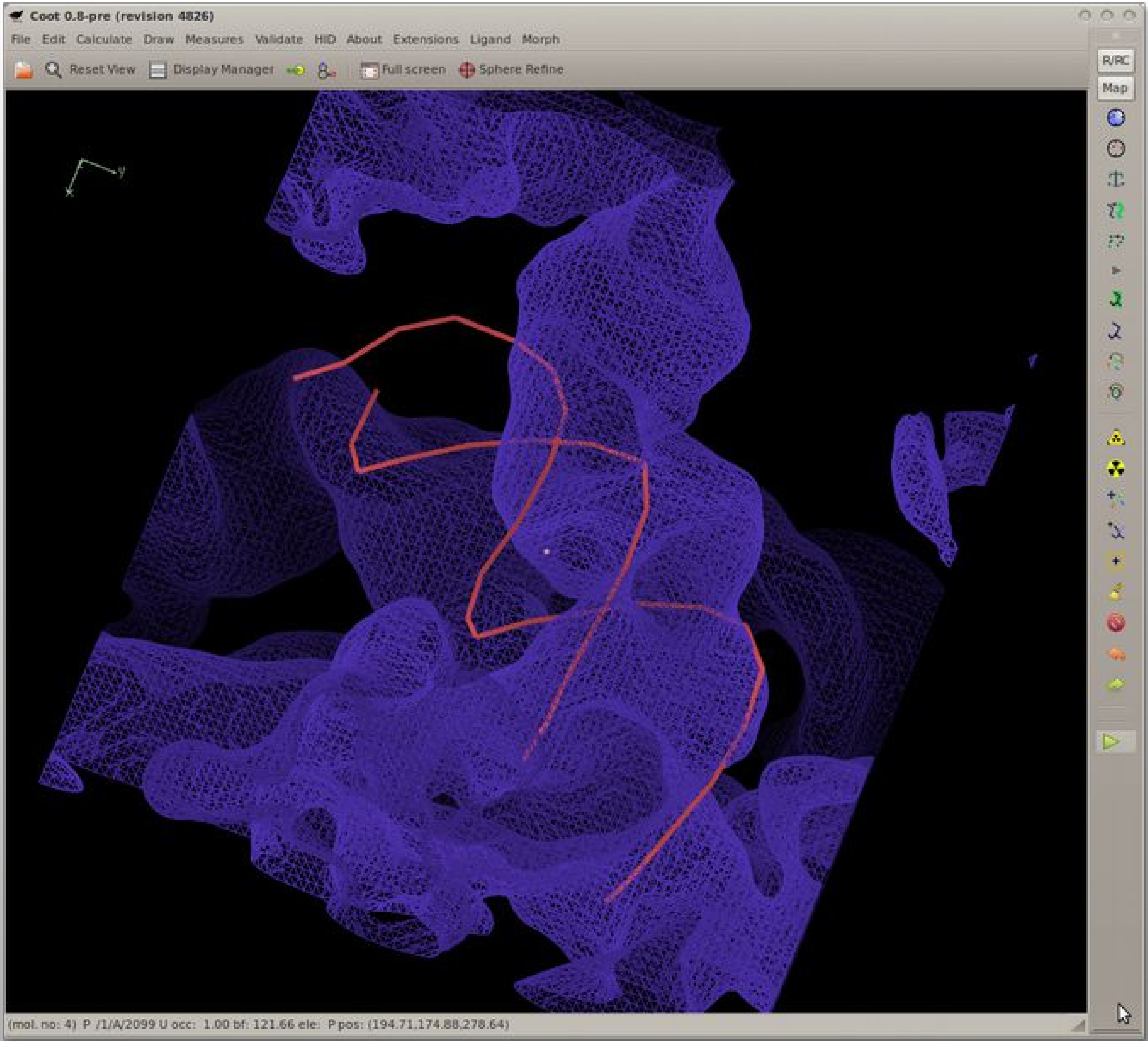


Jiggle Fit

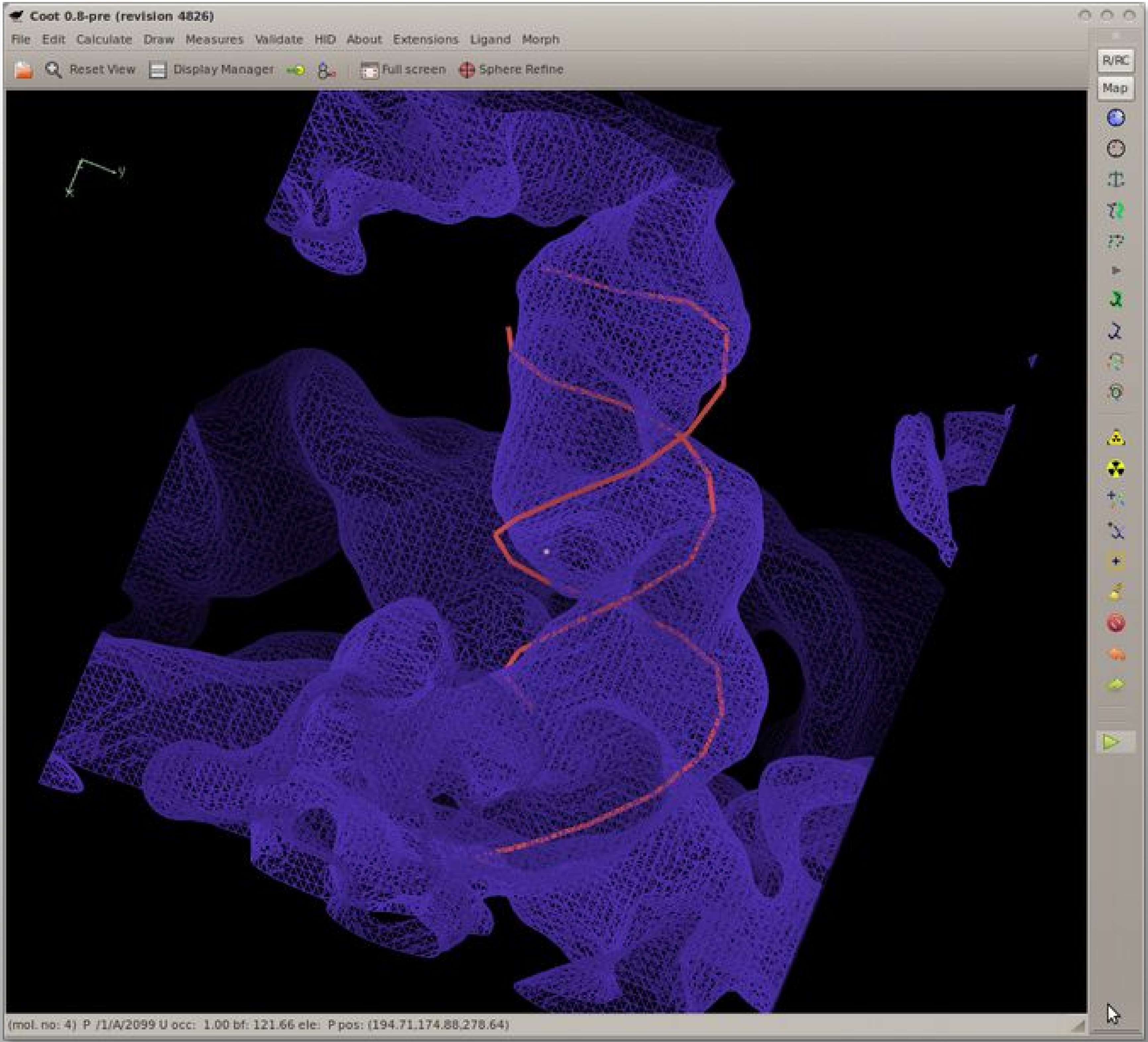
- How do I rotate and translate these atoms to fit the density?
  - 6-dimensional problem
- Originally used to fit simple ligands/solvent molecules to blobs of density
- Now extended to fit arbitrary atom selections
  - *e.g.* by Chain

Jiggle Fit: How it Works

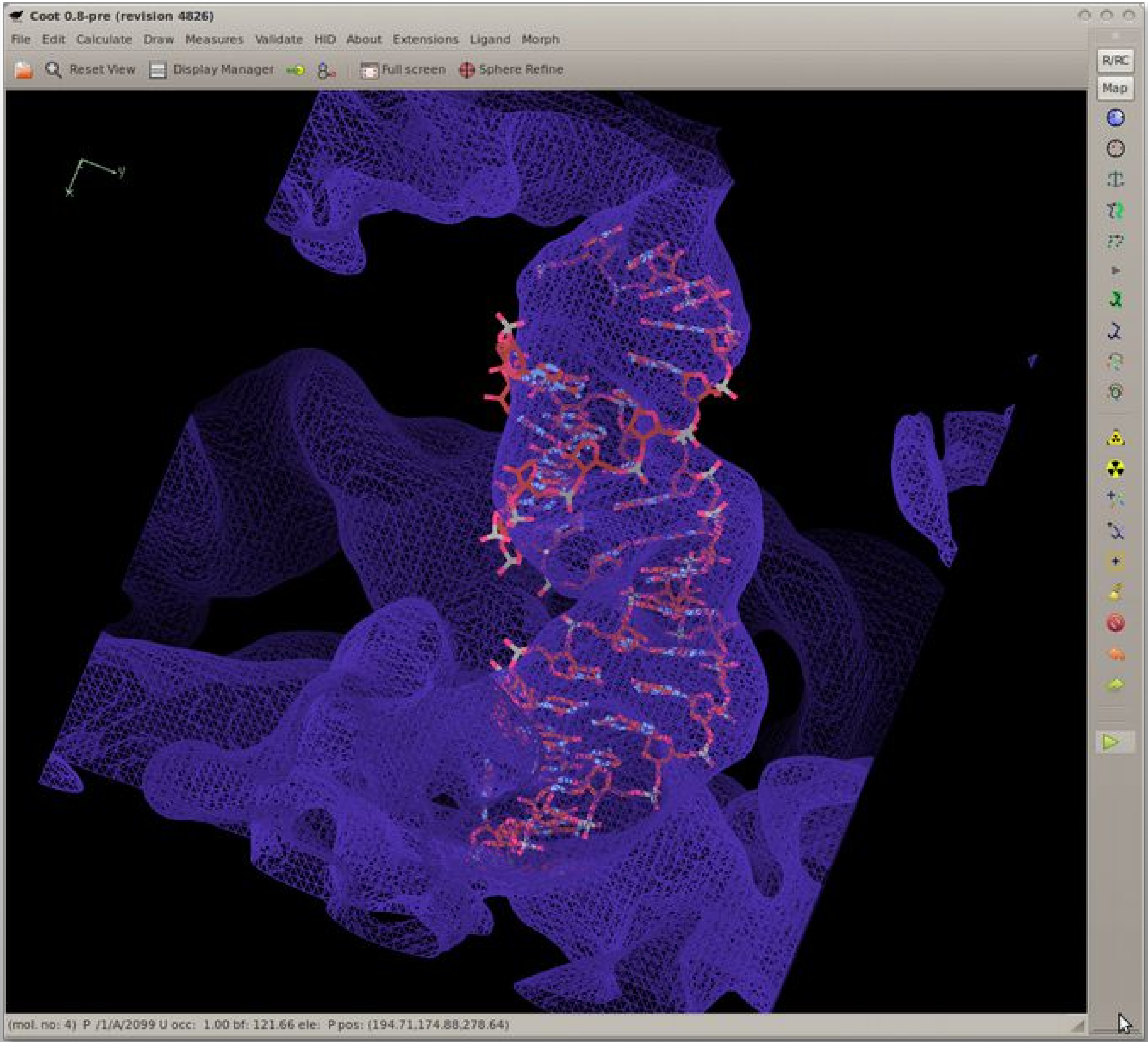
- Loop *n* (say 1000) times:
  - Generate random angles and translations
  - Transform atom selection by these rotations and translation
  - Score and store the fit to density
- Rank density fit scores,
  - Pick top 20 solution, for each of them
    - Rigid body fit and score solutions
    - Pick the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



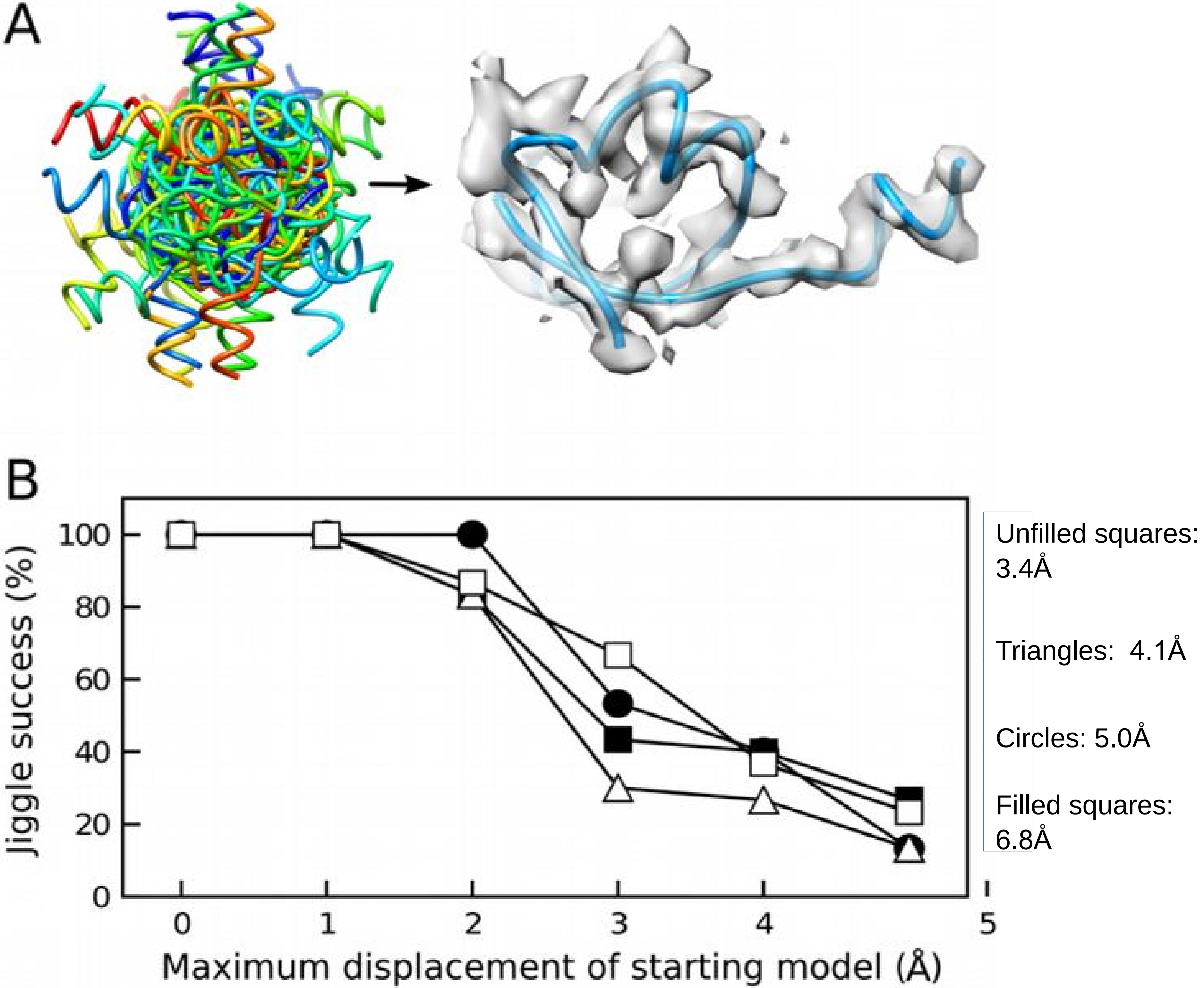
```
(mol. no: 4)  P  /1/A/2099 U occ:  1.00 bf: 121.66 ele:  P pos: (194.71,174.88,278.64)
```



```
(mol. no: 4) P /1/A/2099 U occ: 1.00 bf: 121.66 ele: P pos: (194.71,174.88,278.64)
```



```
(mol. no: 4)  P  /1/A/2099 U occ:  1.00 bf: 121.66 ele:  P pos: (194.71,174.88,278.64)
```



Unfilled squares: 3.4Å

Triangles: 4.1Å

Circles: 5.0Å

Filled squares: 6.8Å

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

(not a great fit)

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
  - Additional user-defined restraints,
    - secondary structure restraints
    - homologous protein local environment restraints
    - local self restraints
- Provides “interactive refinement”

Refinement in *Coot* 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

CCP4 Monomer Library

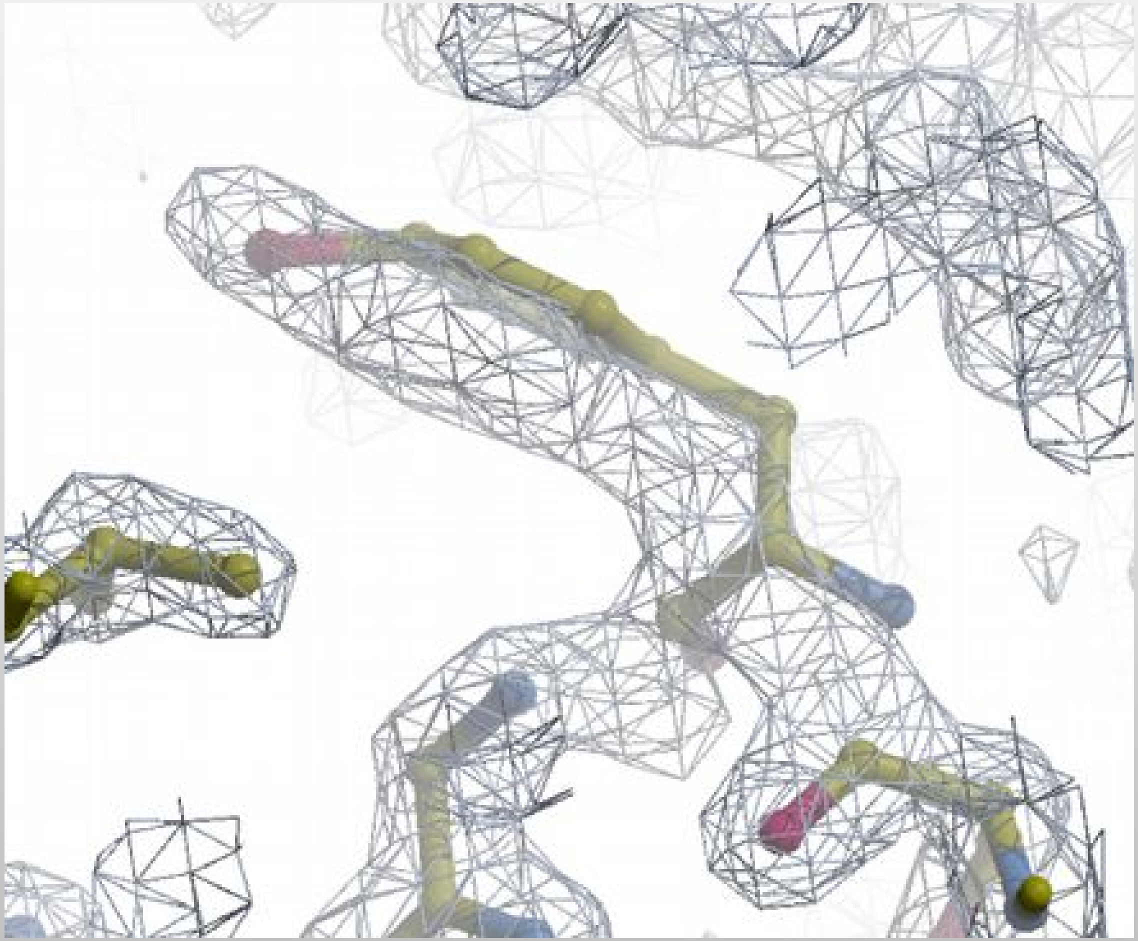
chem_comp_bond

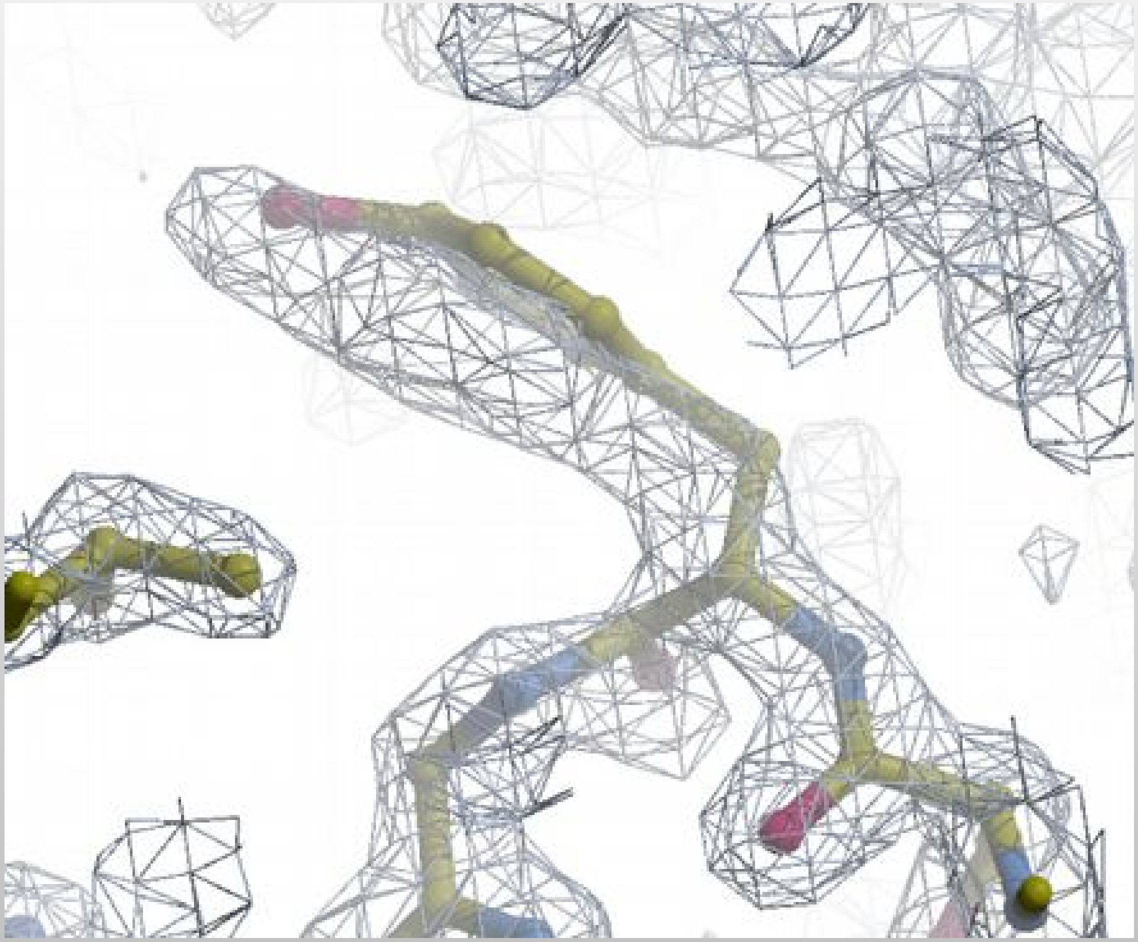
```
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd
 ALA      N    H       single     0.860    0.020
 ALA      N    CA      single     1.458    0.019
 ALA      CA   HA      single     0.980    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
```

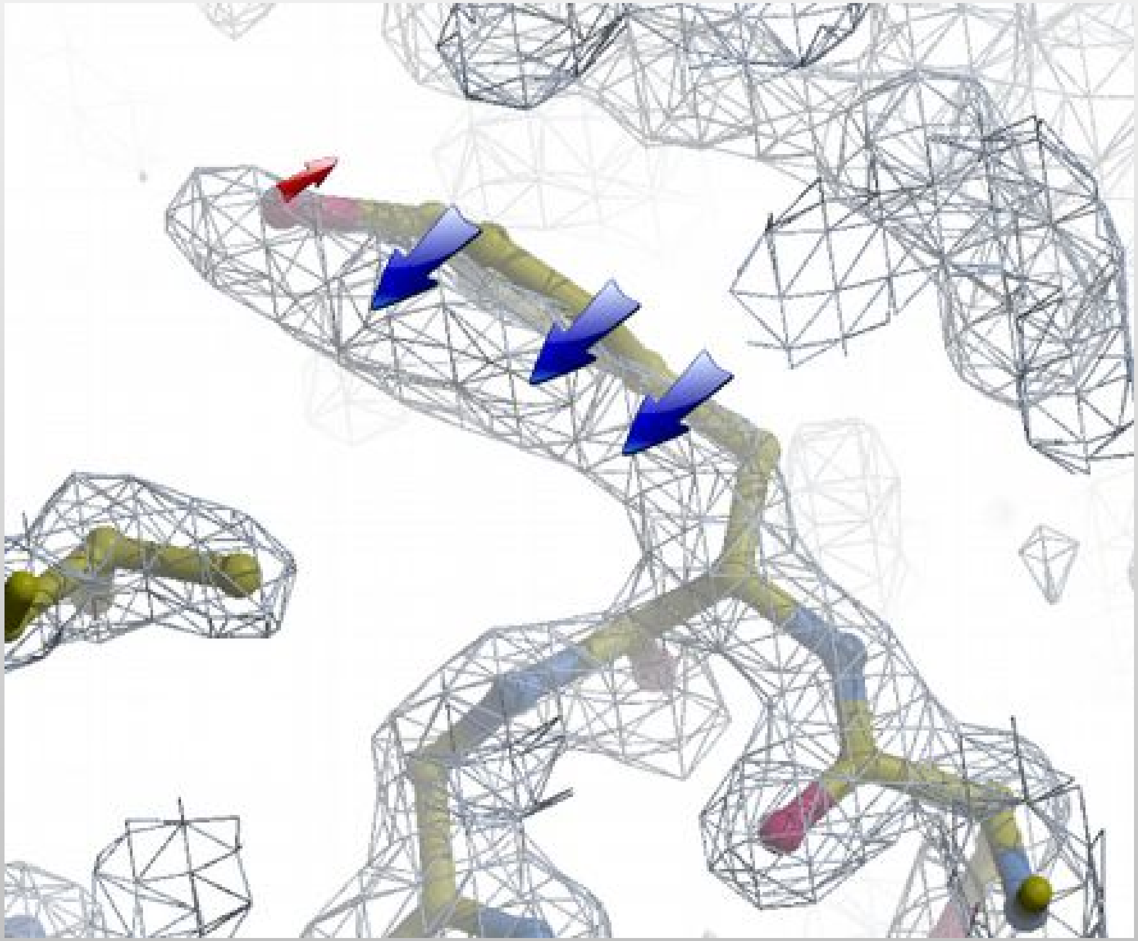
APPENDIX *A*
Regularization and refinement derivatives

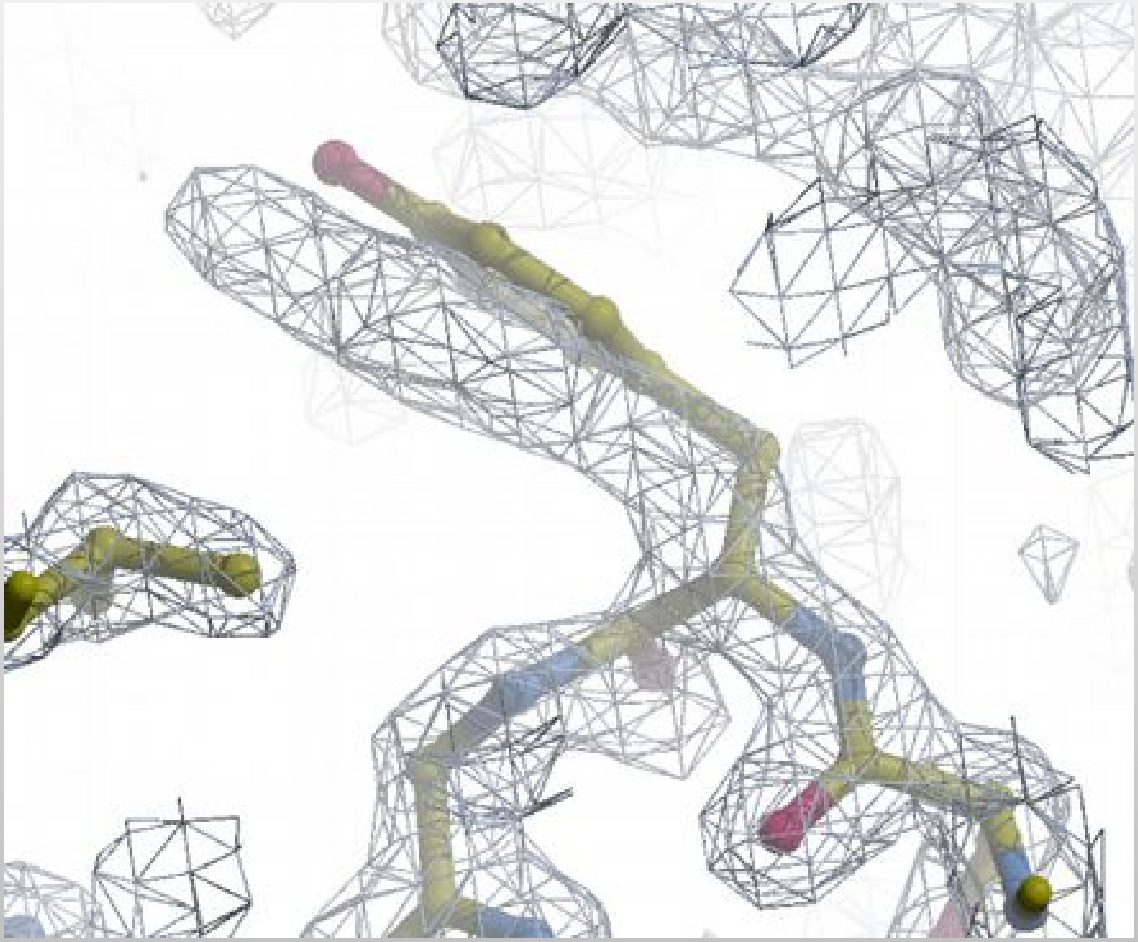
The function that we are trying to minimize is S, where

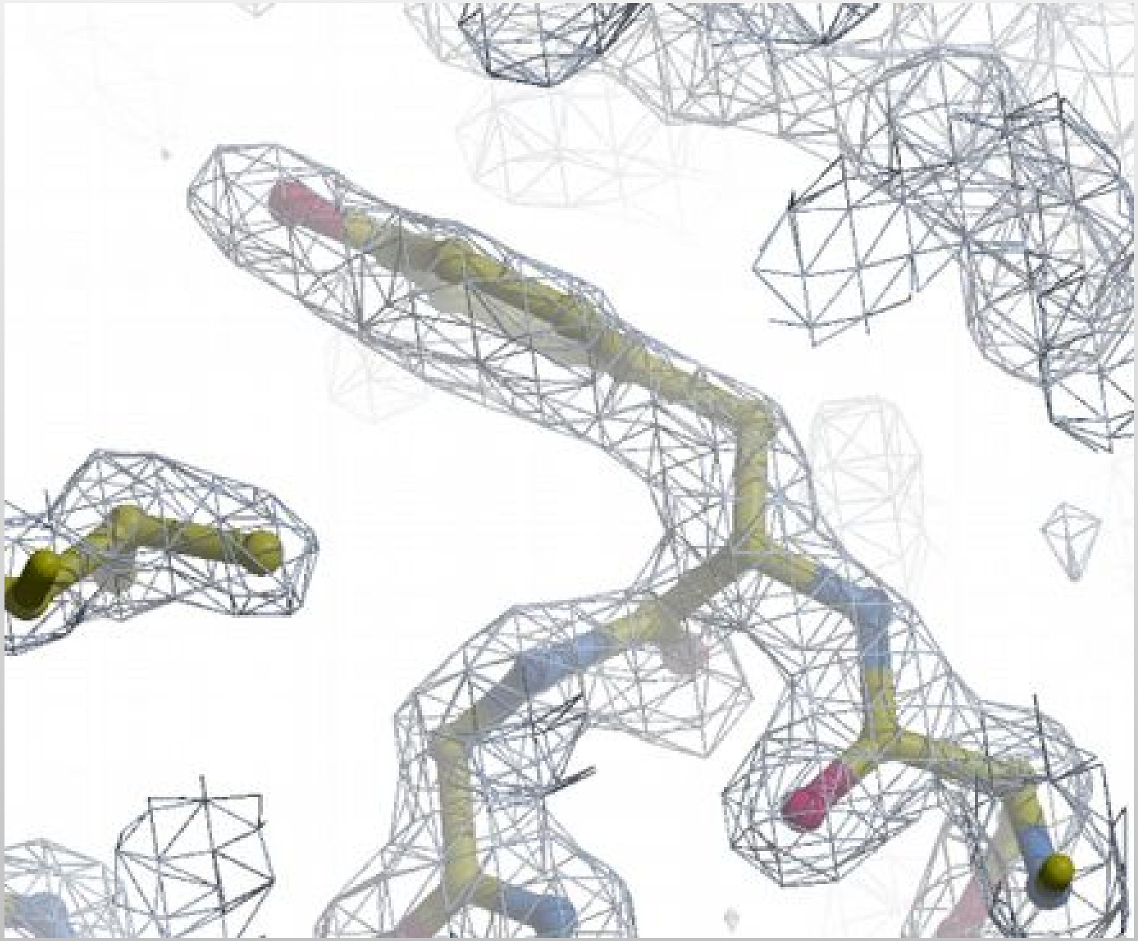
$$S = S_{\text{bond}} + S_{\text{angle}} + S_{\text{torsion}} + S_{\text{plane}} +$$
$$S_{\text{nbc}} + S_{\text{chiral}}$$
$$- S_{\text{density}}$$

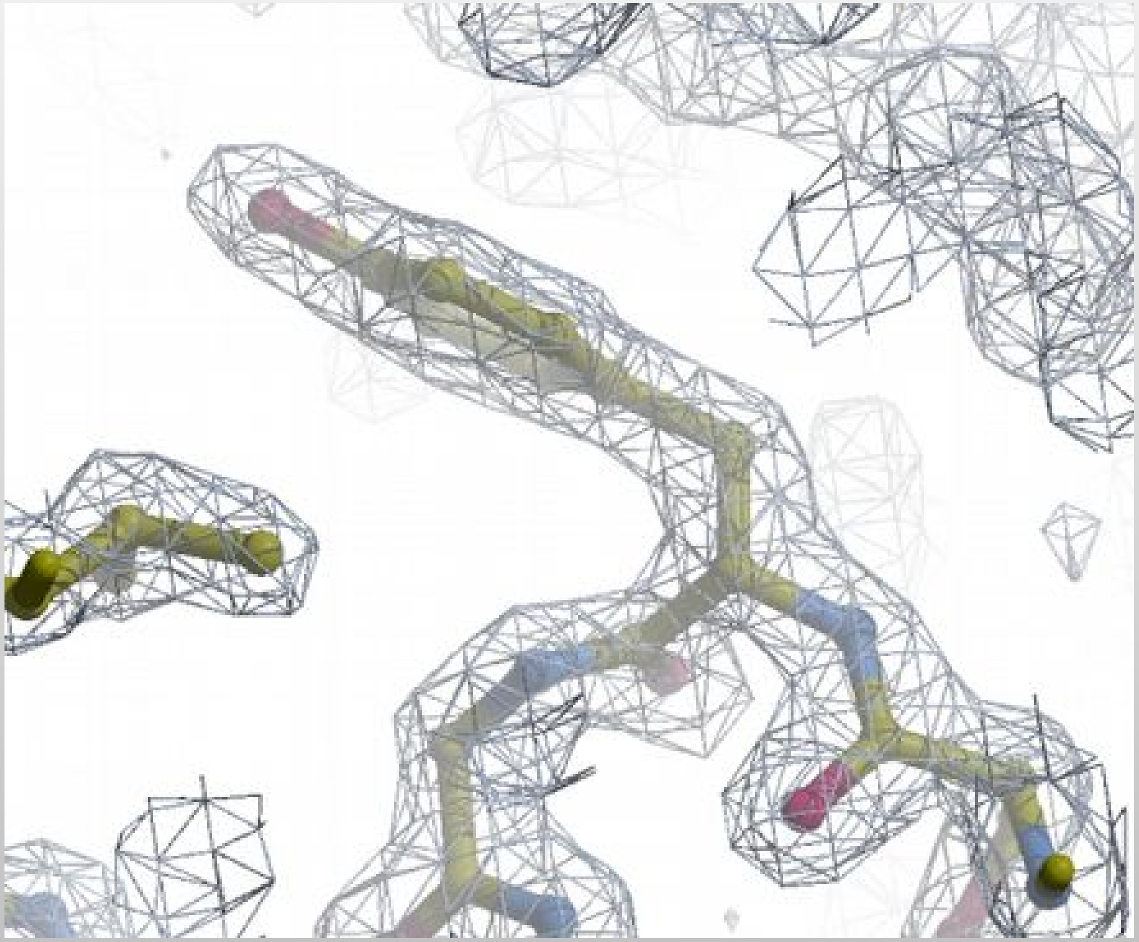






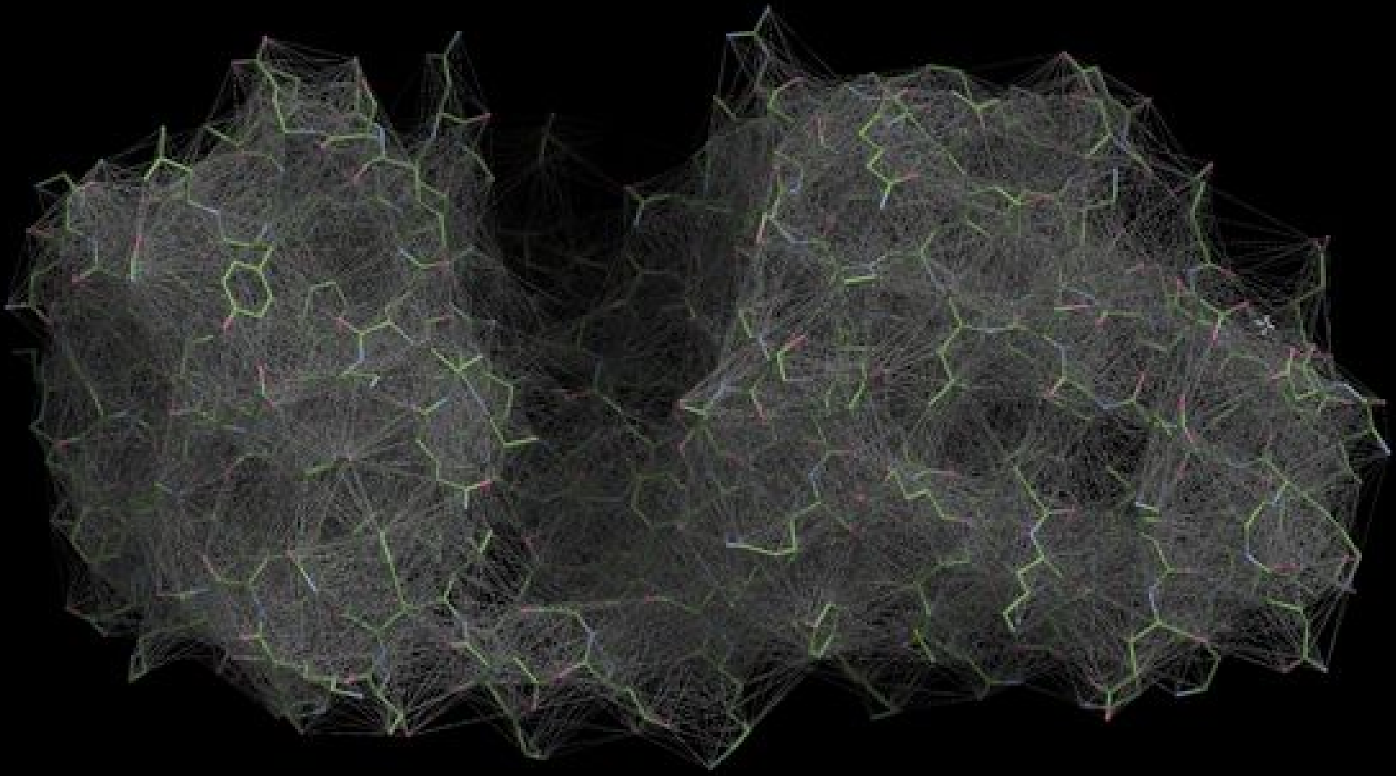




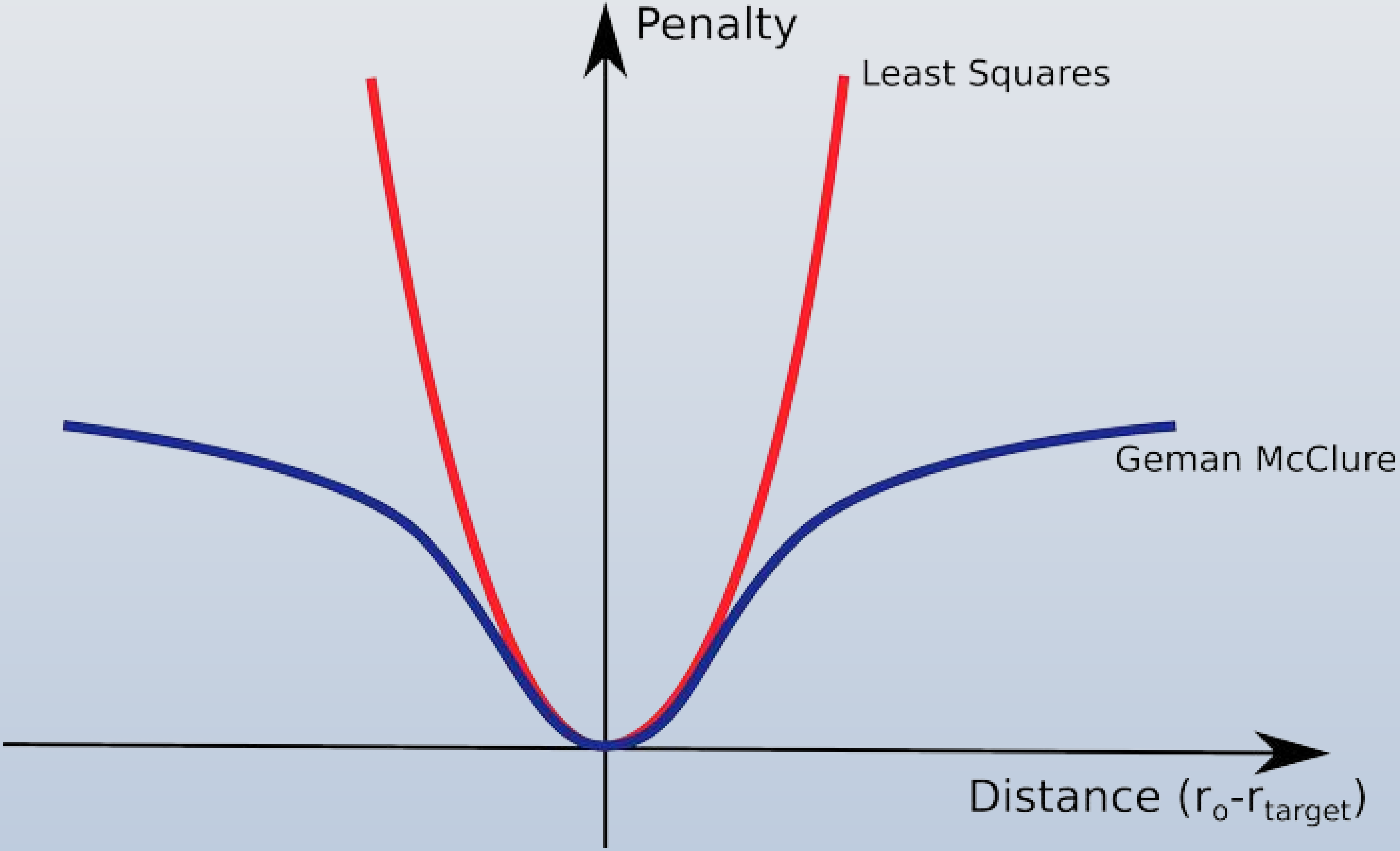


Additional Restraints

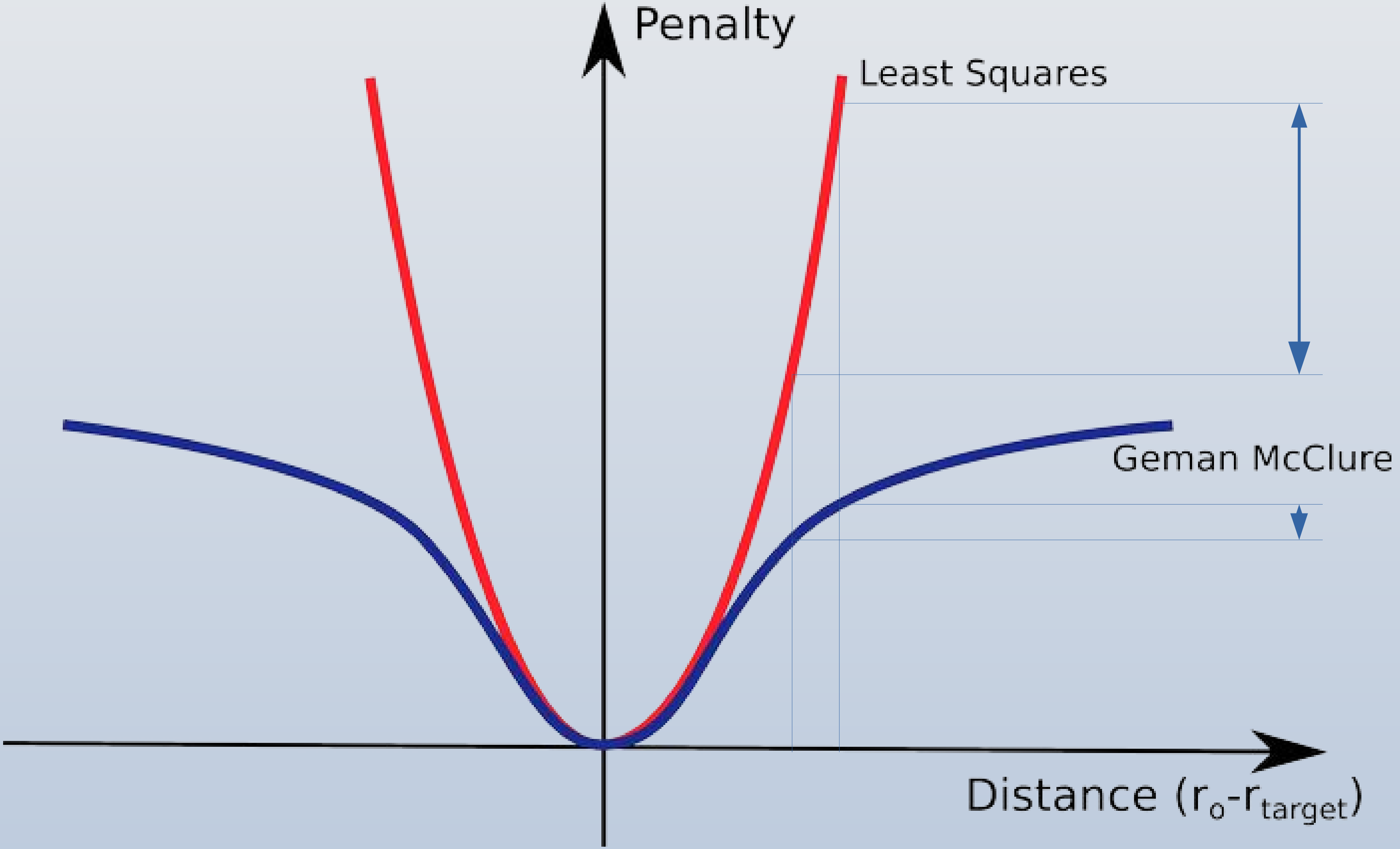
Local Distance Restraints



Modified Target Function



Modified Target Function



Cryo-EM Refinement: Using "ProSMART-like" Restraints

- As well as using a reference model, I often I use "Self" restraints
  - which can be calculated internally
  - the starting model is the "reference" from which the ideal distances are calculated
  - message to the refinement:
  - "keep the local environments similar to how they were when you started"

- The minimizer in *Coot* is a 1st order (derivative) based method
  - "Jelly body" stabilizer cannot work

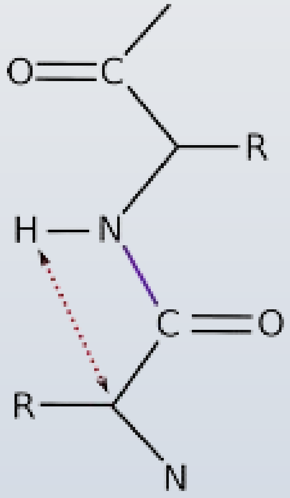
cis-Peptides

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

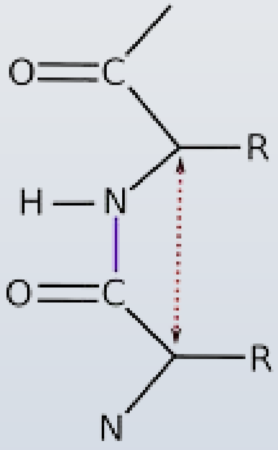
cis-Peptides

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

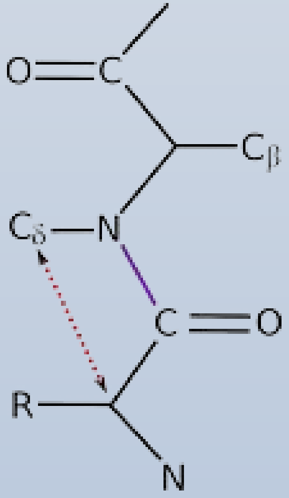
cis-Peptides



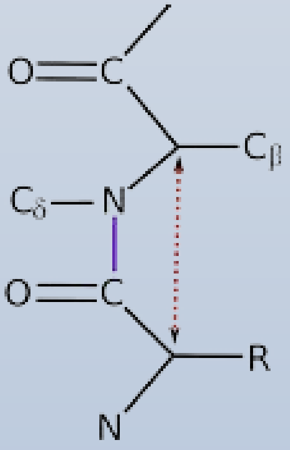
trans-peptide



cis-peptide

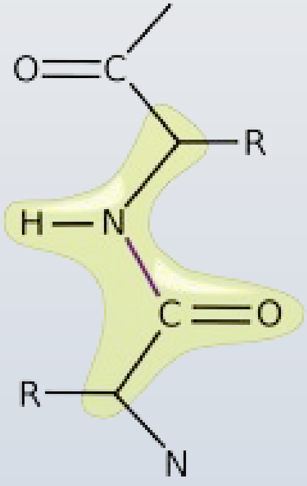


PRO *trans-peptide*

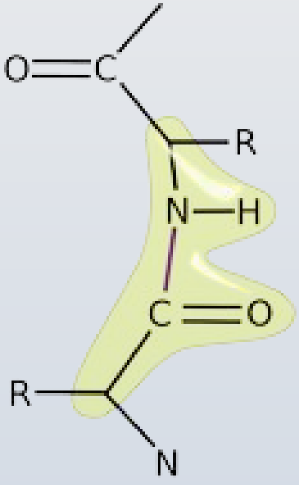


PRO *cis-peptide*

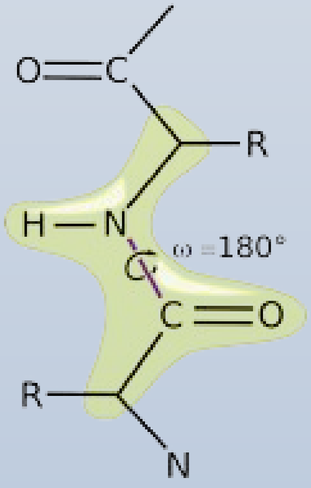
cis-Peptides



trans-peptide
with plane restraints

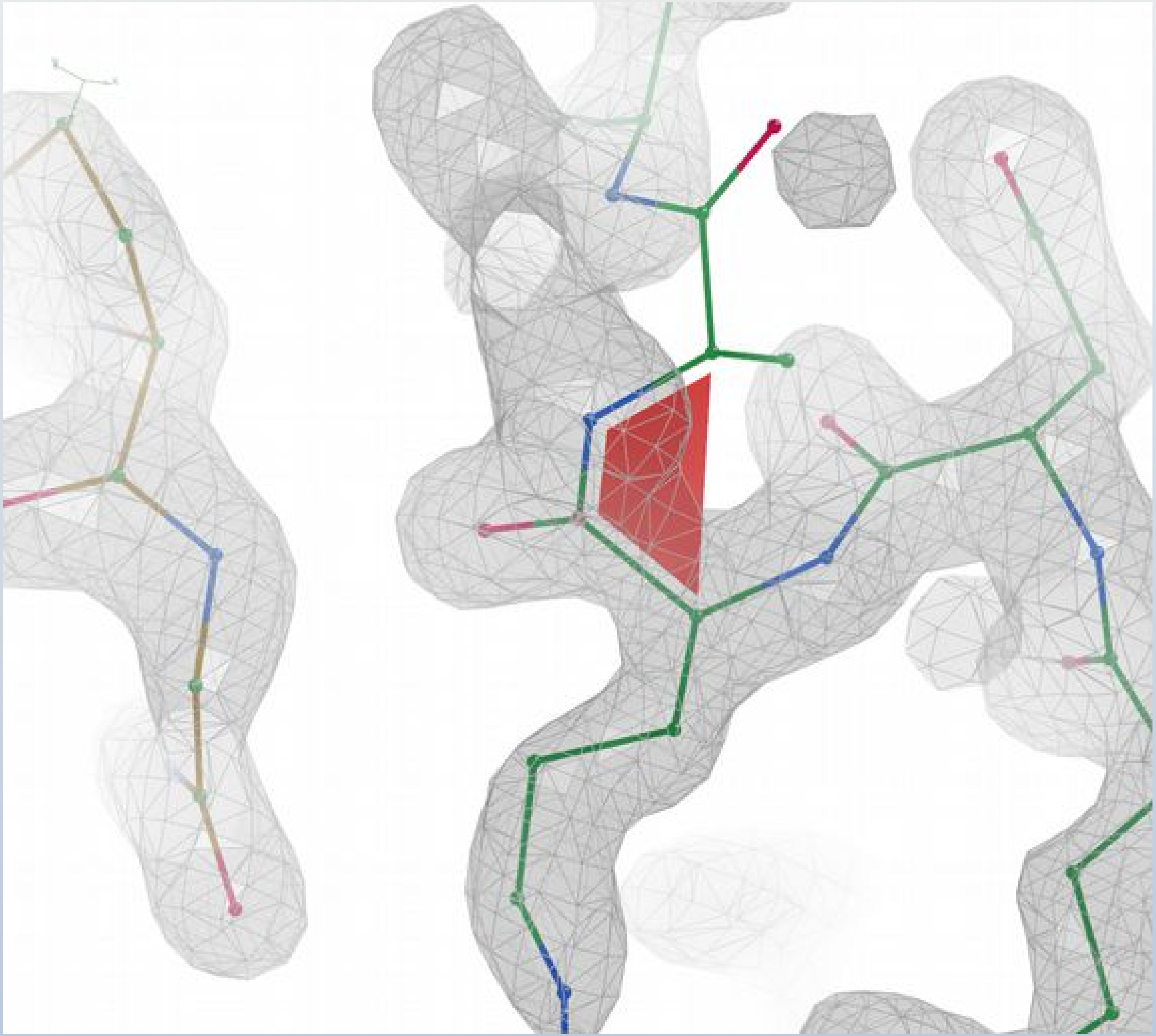


cis-peptide
with plane restraints



trans-peptide
with plane and trans restraints

cis-peptide Representation



Pre-PRO 🟩

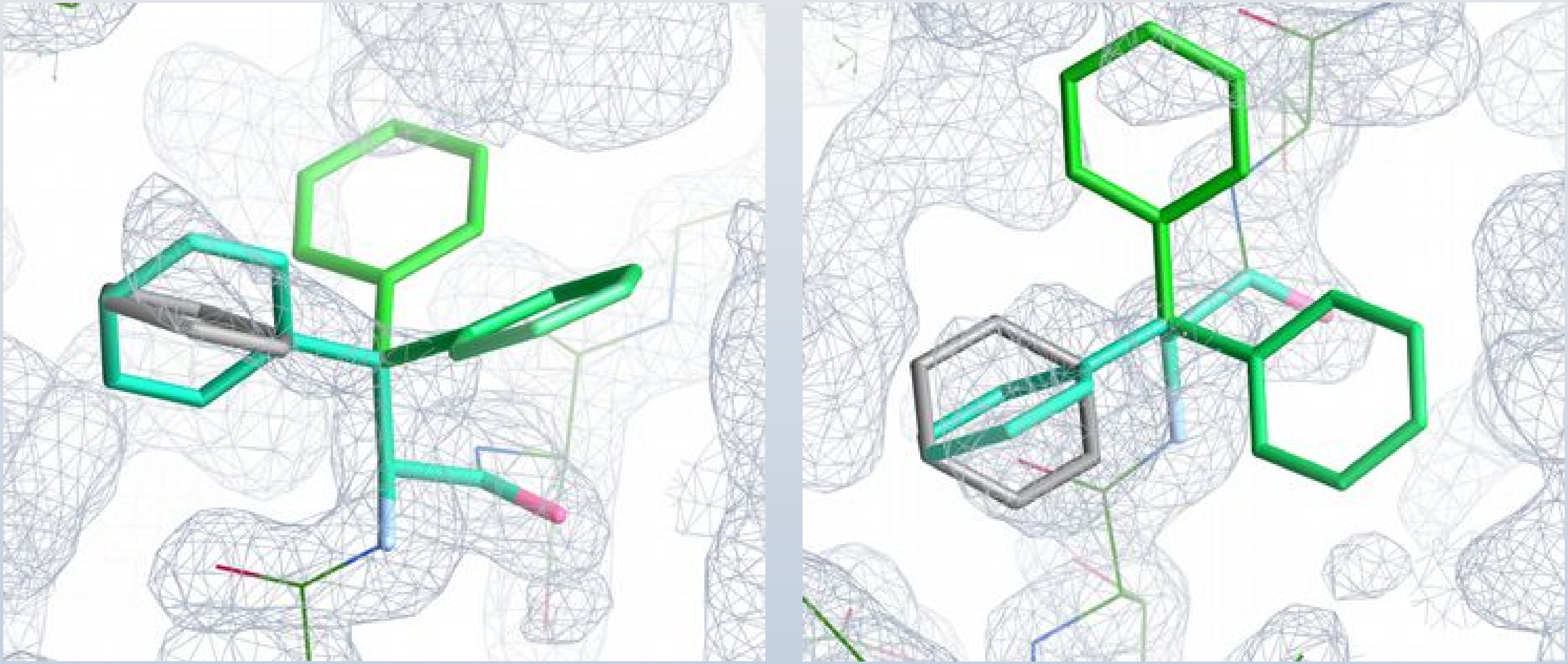
Twisted-trans 🟨

Non-pre-PRO 🟥

Rotamers

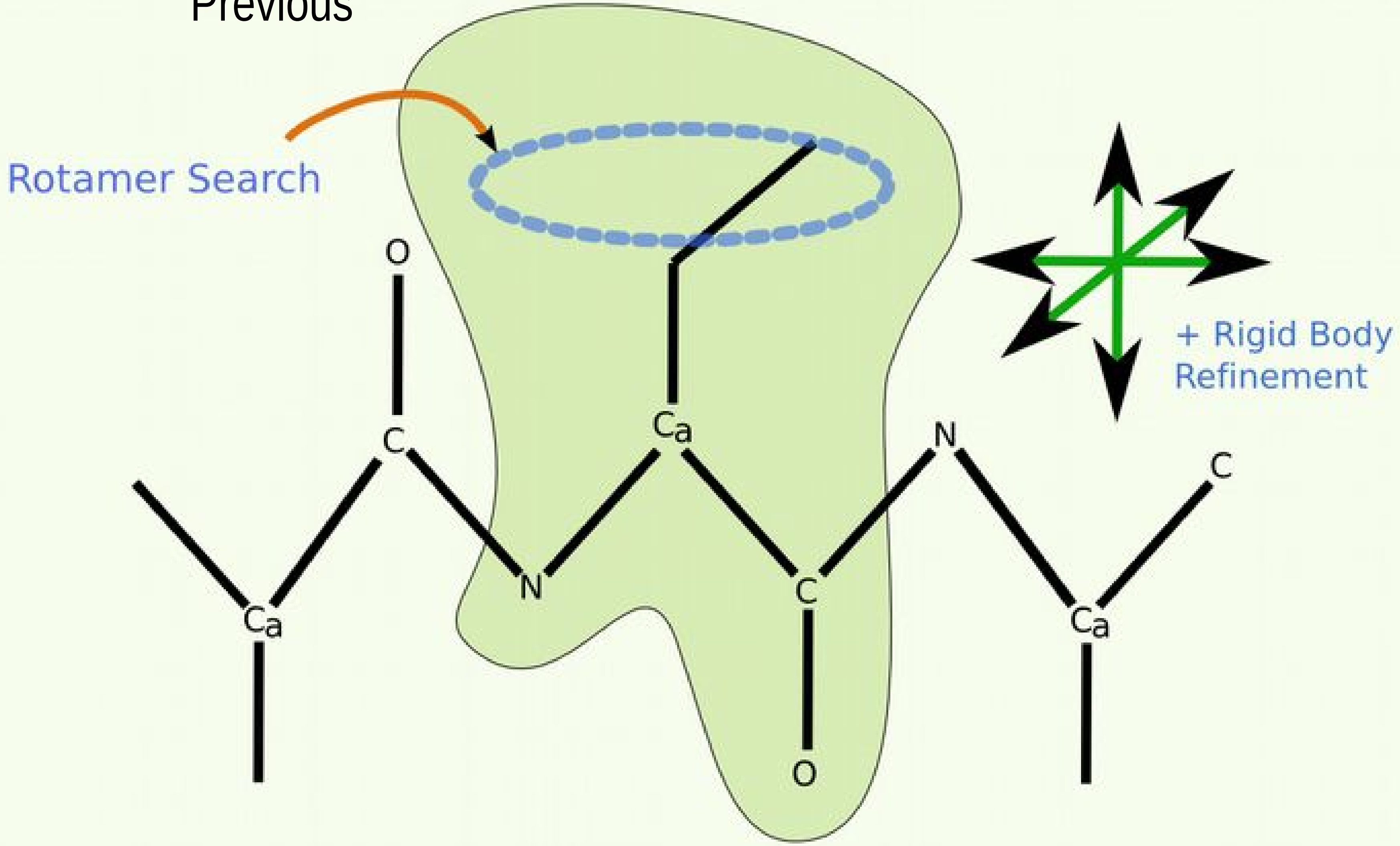
- Rotamers are preferred configurations of a side-chains rotatable bonds
  - where "preferred" means these configurations occur more frequently in a set of reference protein structures
  - "preferred" because they are low-energy conformations
- Several Rotamer "databases" exist
  - best: (Son of) Penultimate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

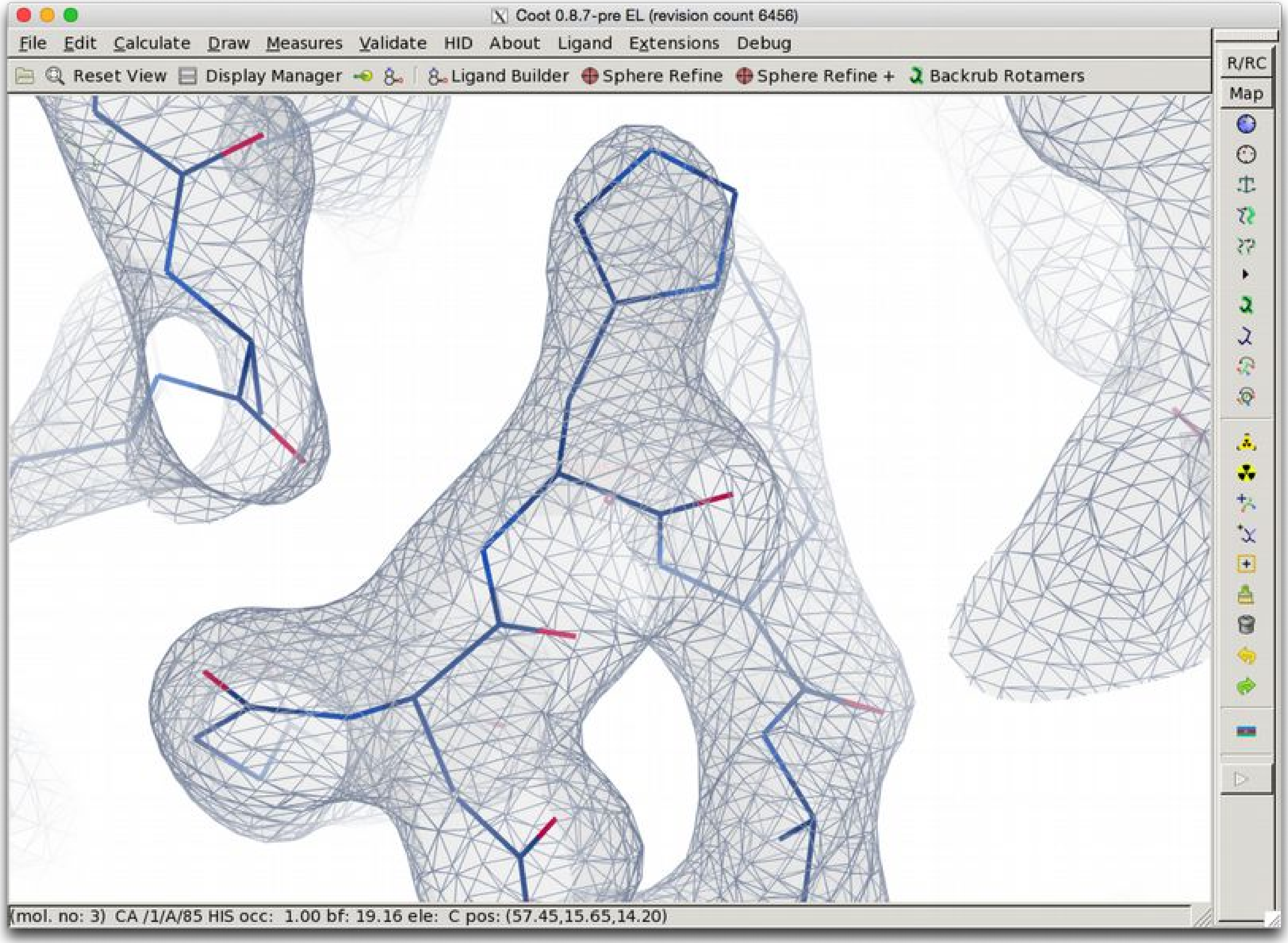
4 PHE Rotamers



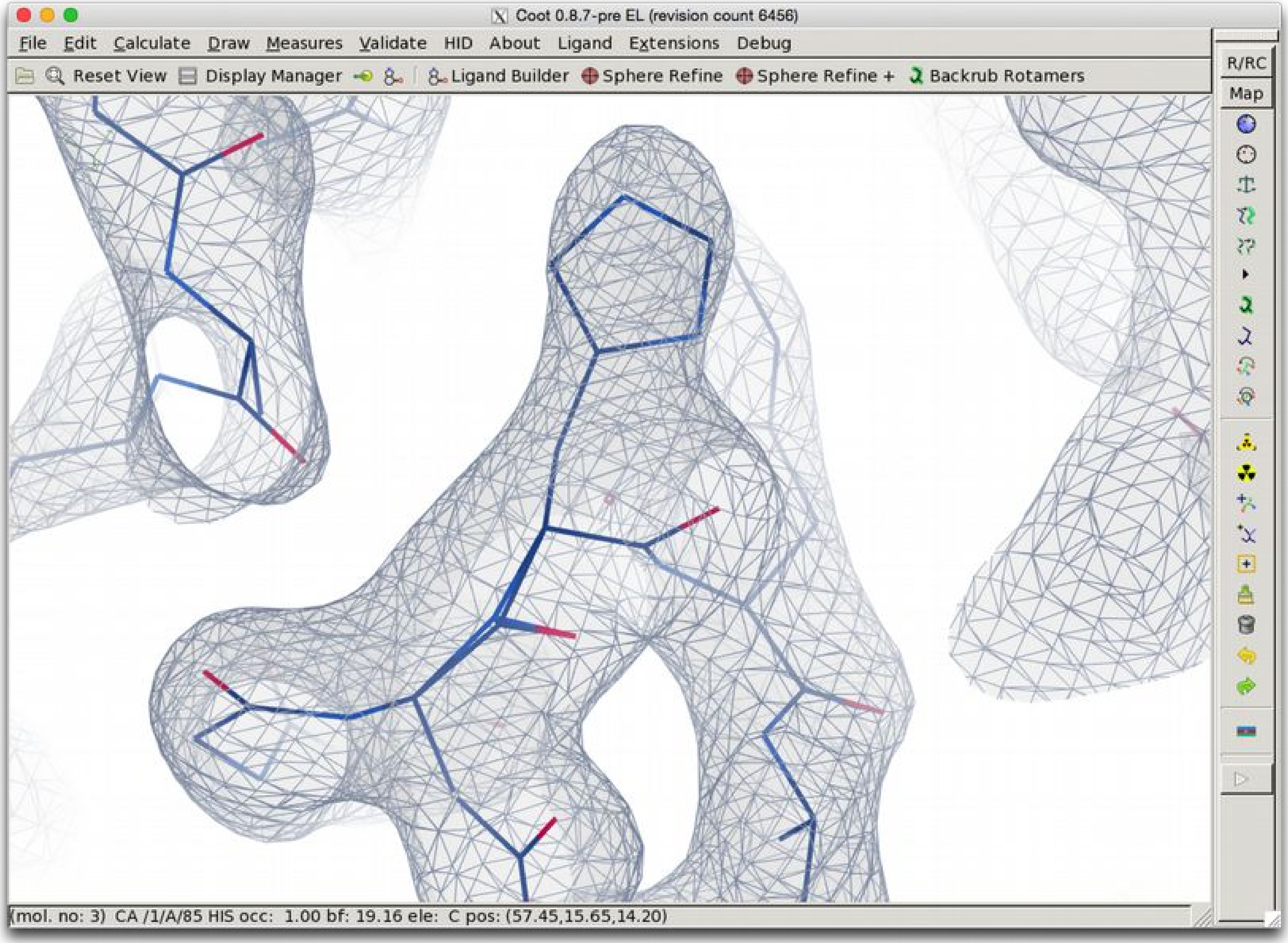
~~Current~~ Low Resolution Rotamer Search

Previous

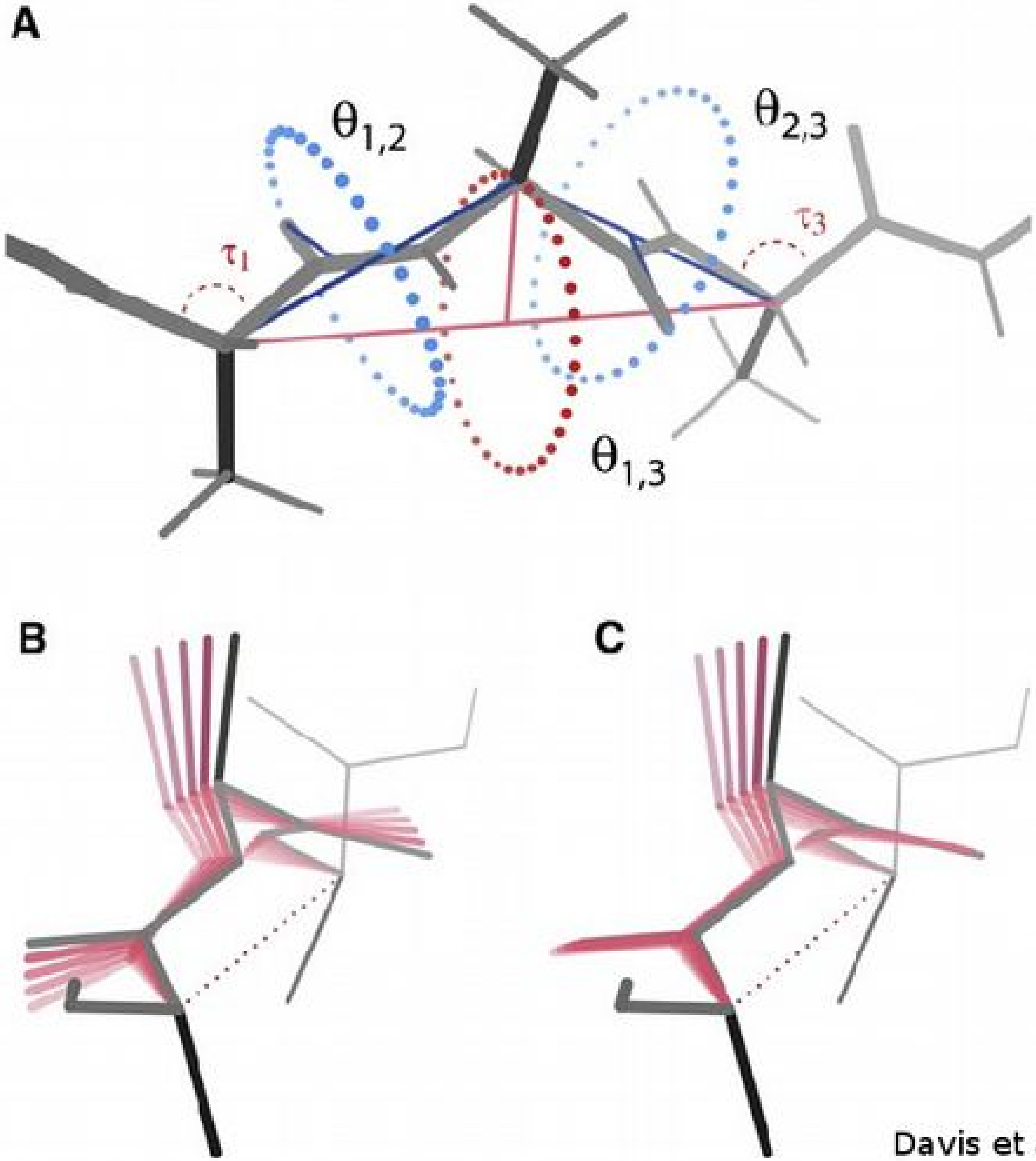




```
(mol. no: 3)  CA /1/A/85 HIS occ:  1.00 bf: 19.16 ele:  C pos: (57.45,15.65,14.20)
```

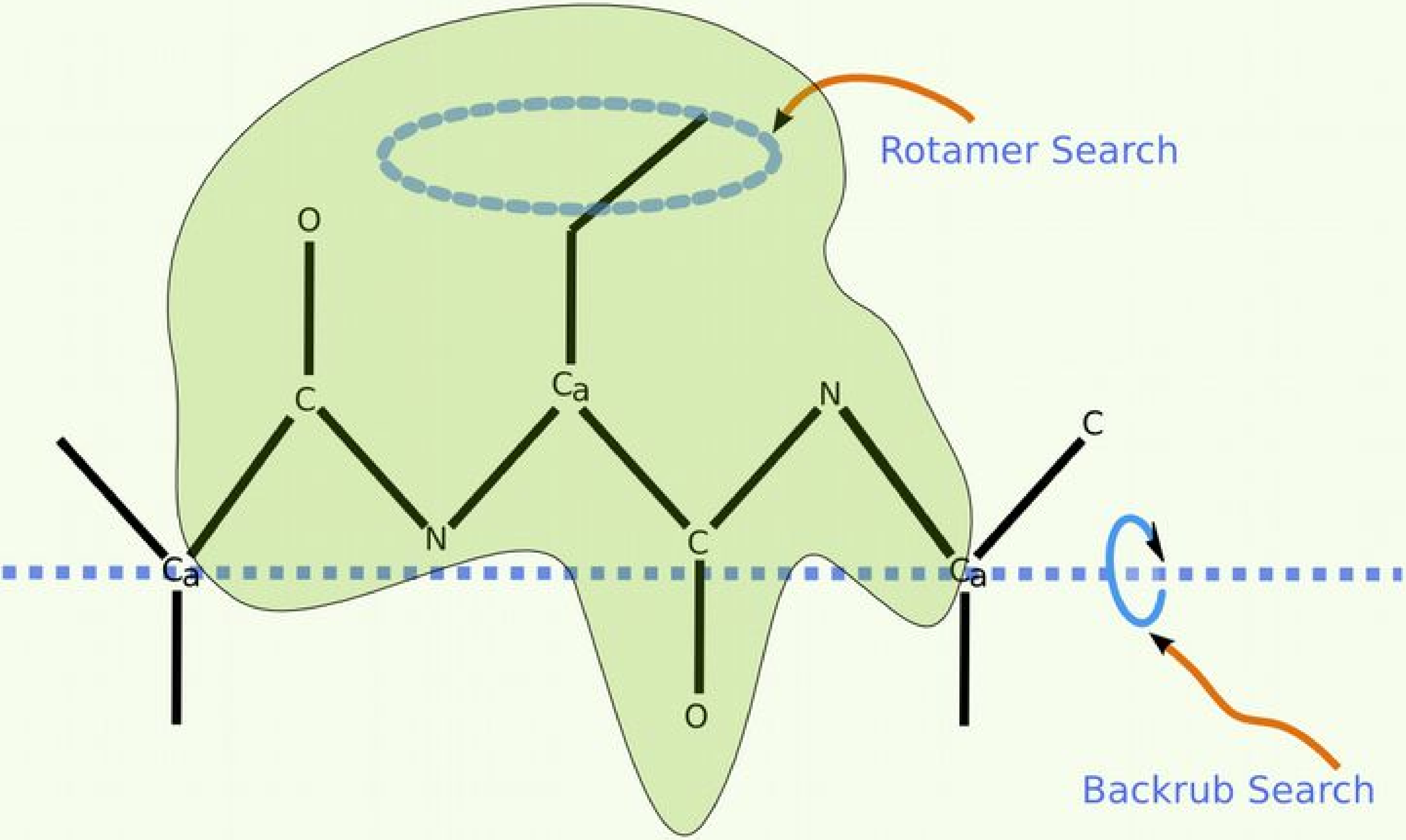


```
(mol. no: 3) CA /1/A/85 HIS occ:  1.00 bf: 19.16 ele:  C pos: (57.45,15.65,14.20)
```

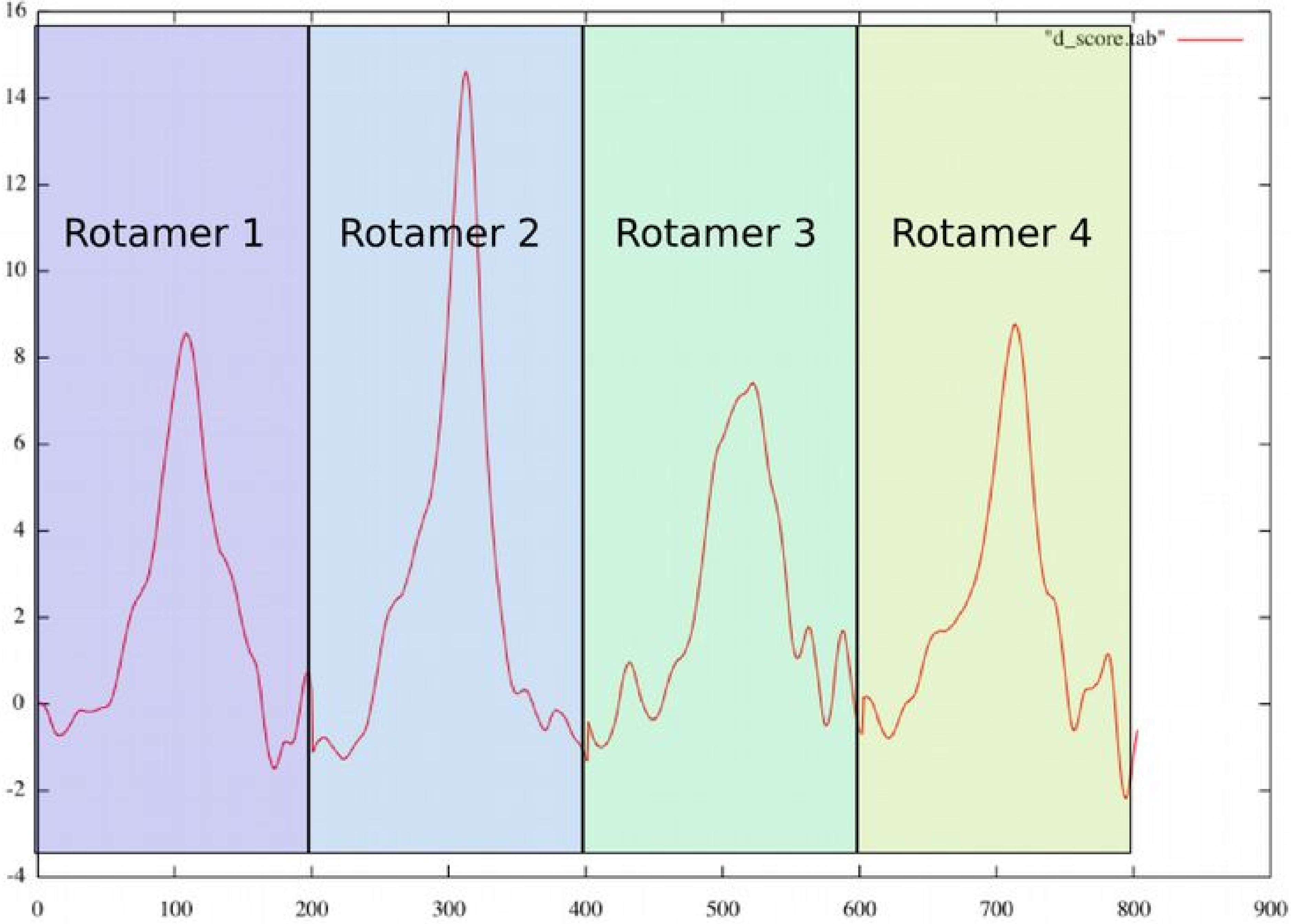


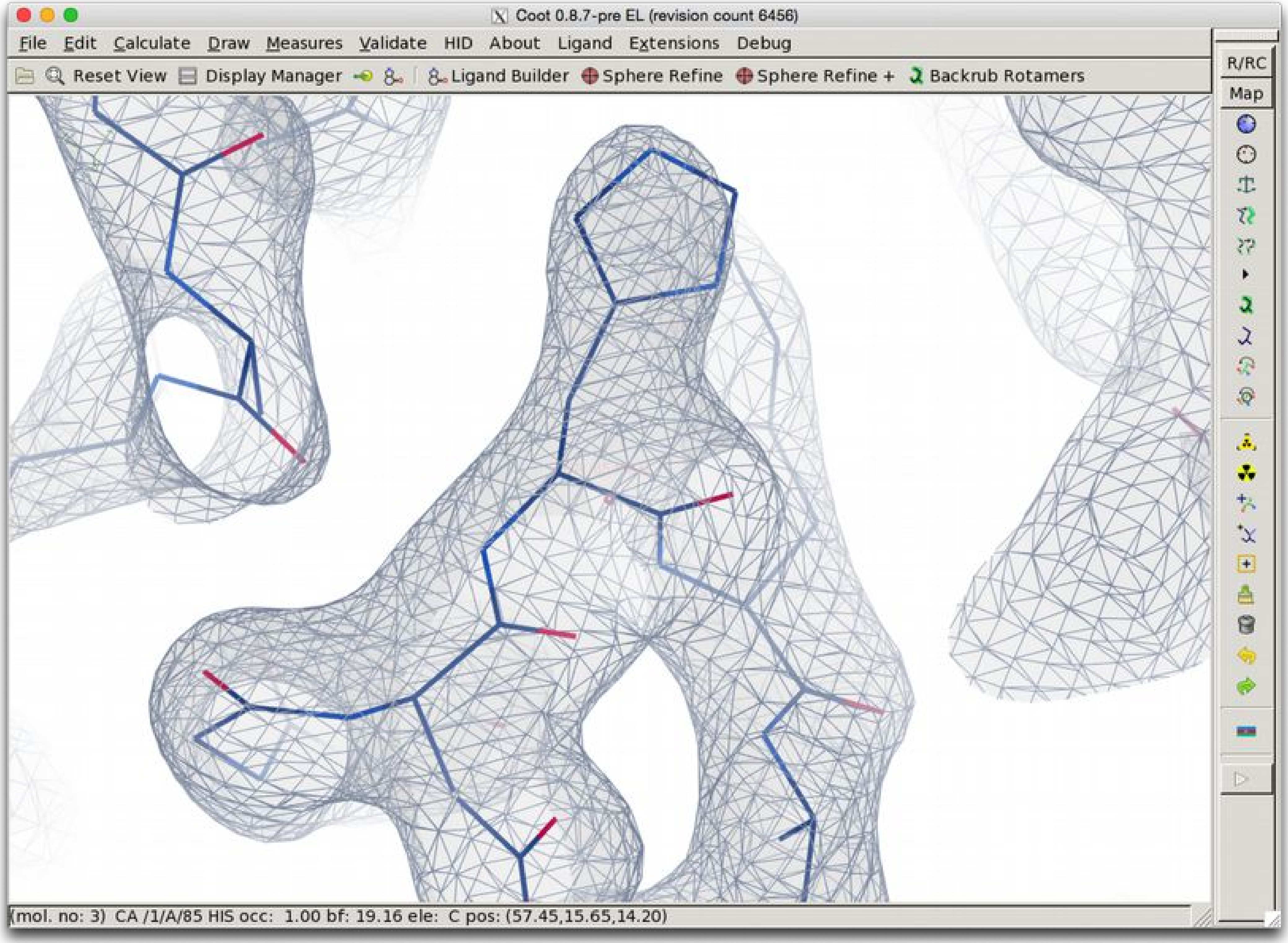
Davis et al. (2006) Structure

New Low Resolution Rotamer Search

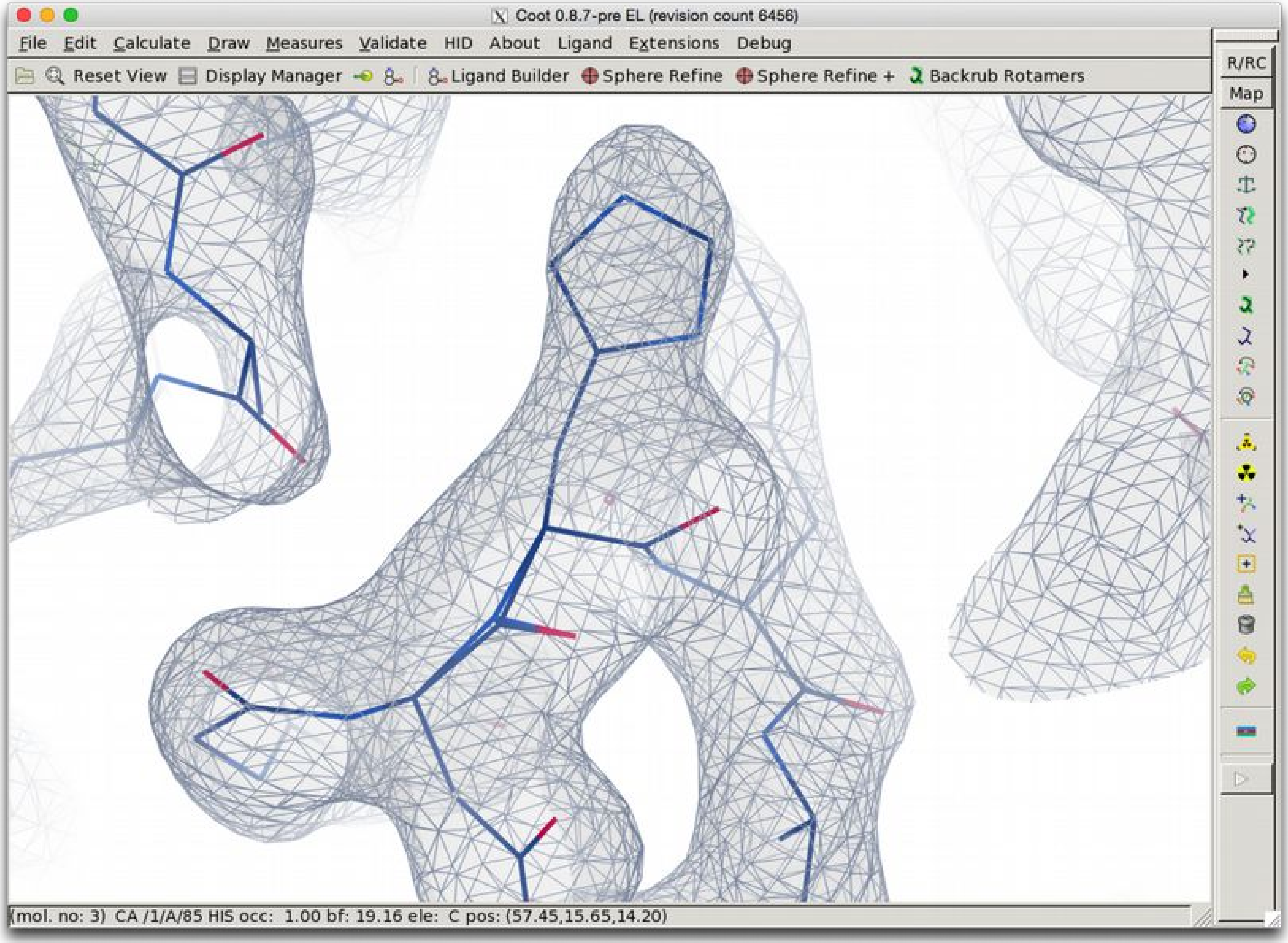


After Fitting Tools in KING/Molprobity

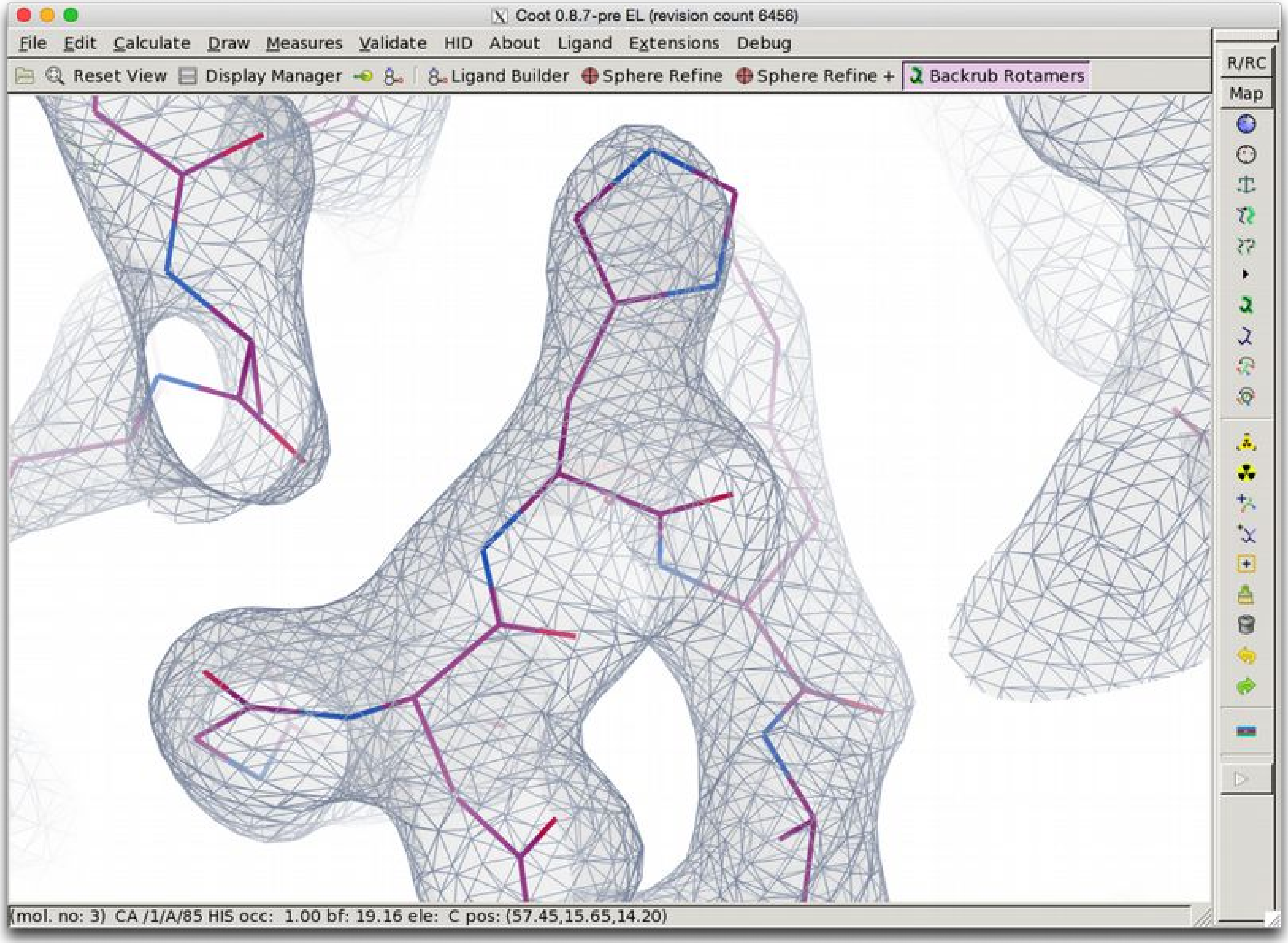

"d_score.tab"
Rotamer 1
Rotamer 2
Rotamer 3
Rotamer 4
16
14
12
10
8
6
4
2
0
-2
-4
0
100
200
300
400
500
600
700
800
900



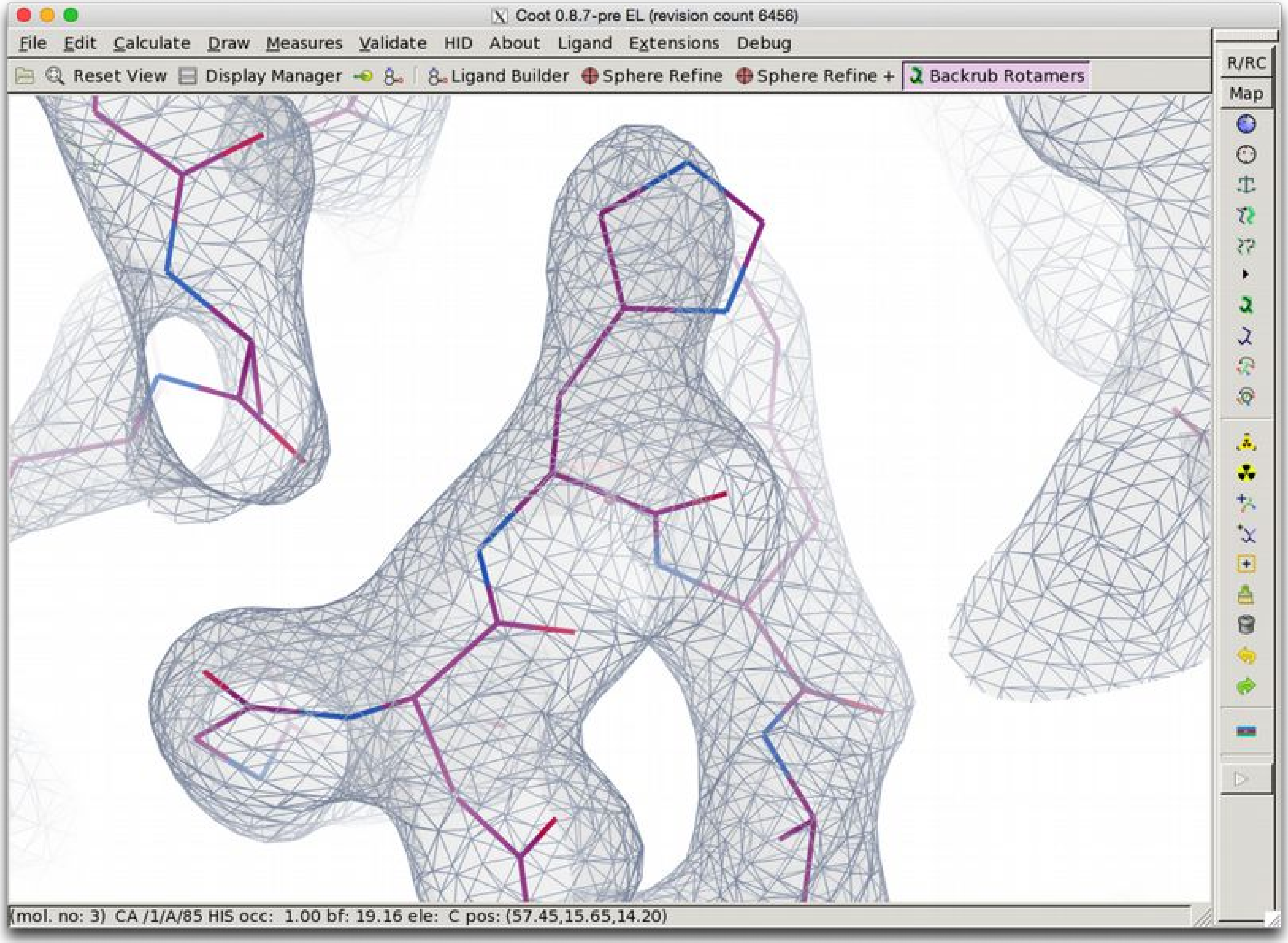
```
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)
```



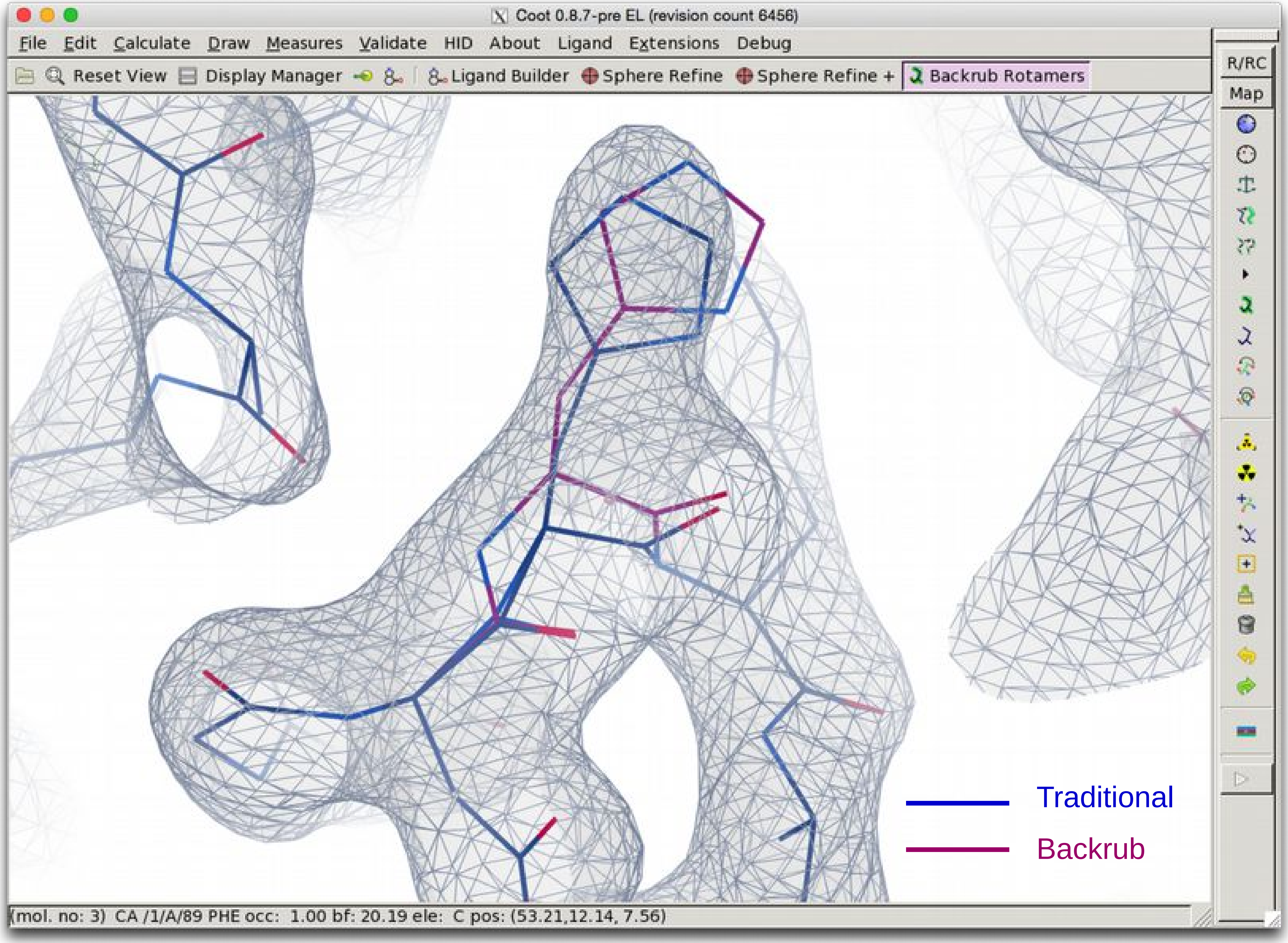
```
(mol. no: 3)  CA /1/A/85 HIS occ:  1.00 bf: 19.16 ele:  C pos: (57.45,15.65,14.20)
```



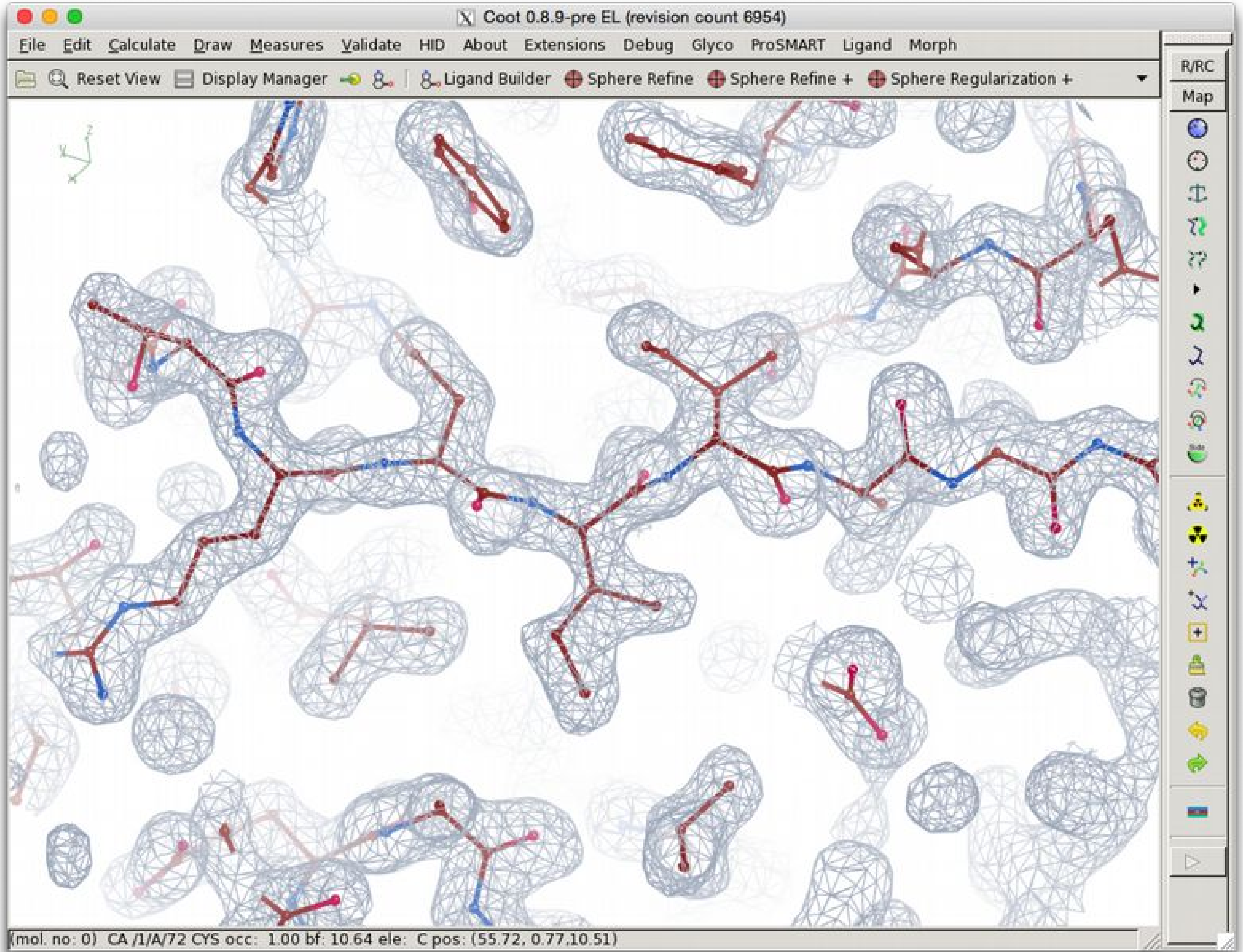
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)



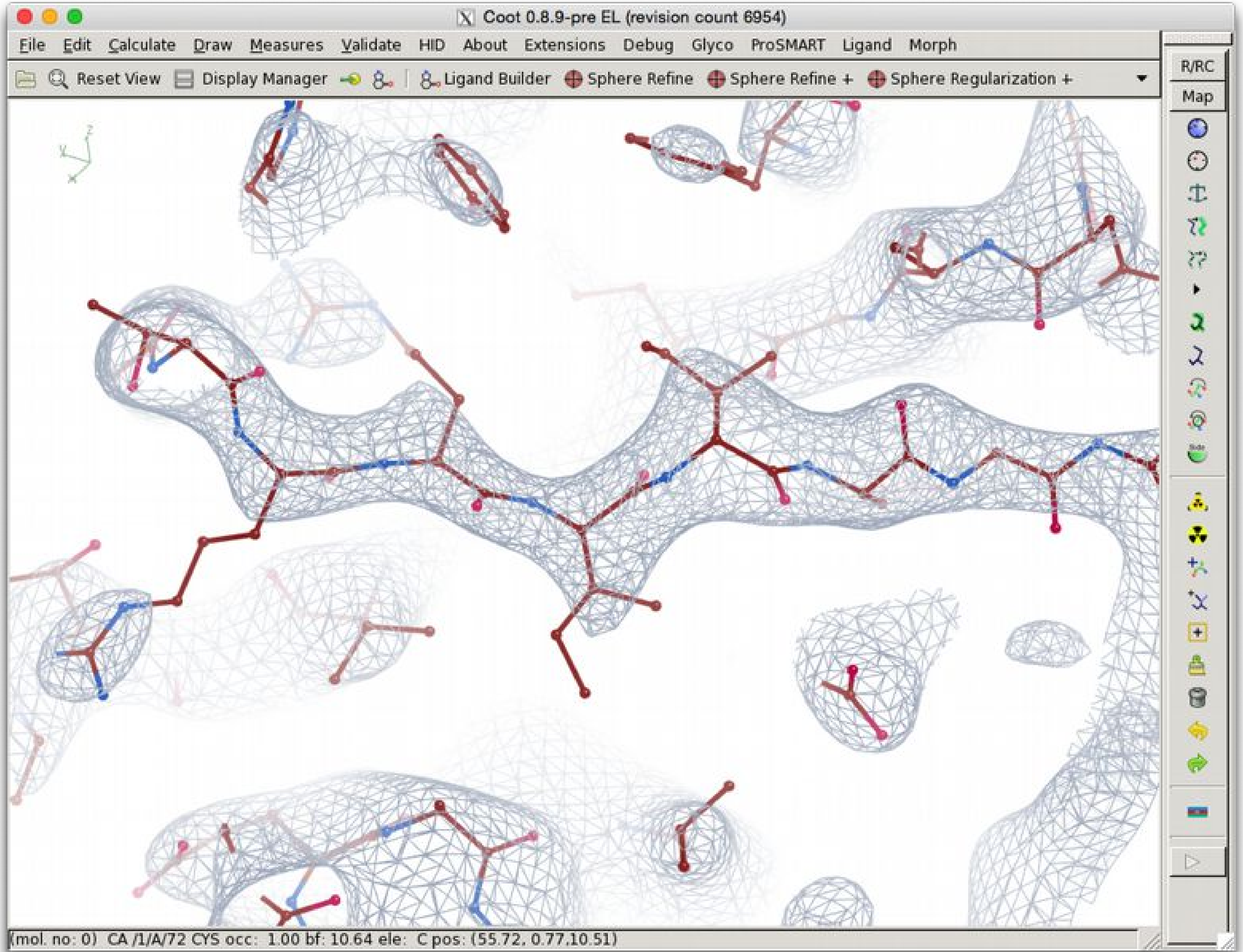
```
(mol. no: 3) CA /1/A/85 HIS occ:  1.00 bf: 19.16 ele:  C pos: (57.45,15.65,14.20)
```



```
(mol. no: 3) CA /1/A/89 PHE occ:  1.00 bf: 20.19 ele:  C pos: (53.21,12.14, 7.56)
```

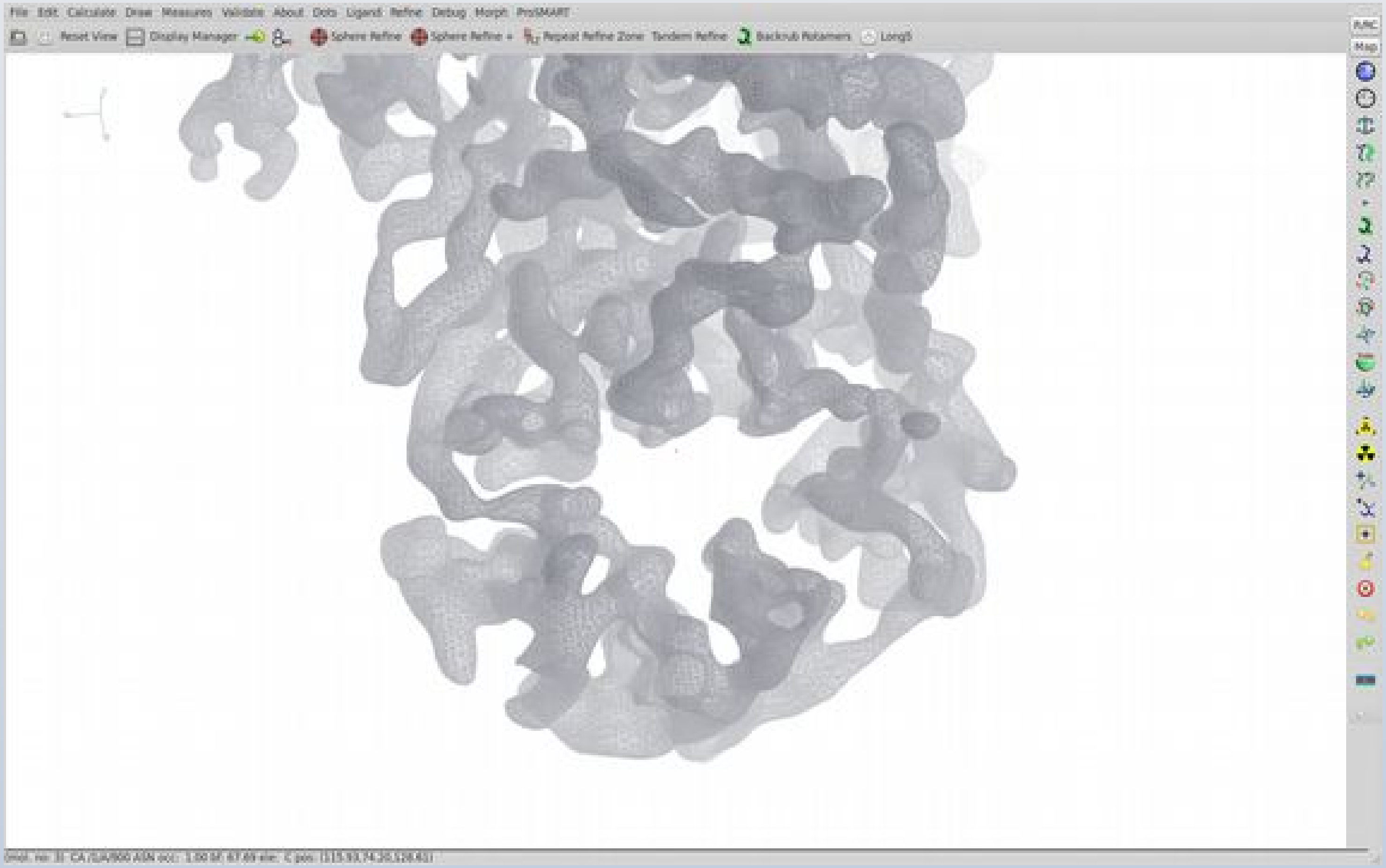


```
(mol. no: 0)  CA /1/A/72 CYS occ:  1.00 bf: 10.64 ele:  C pos: (55.72, 0.77,10.51)
```

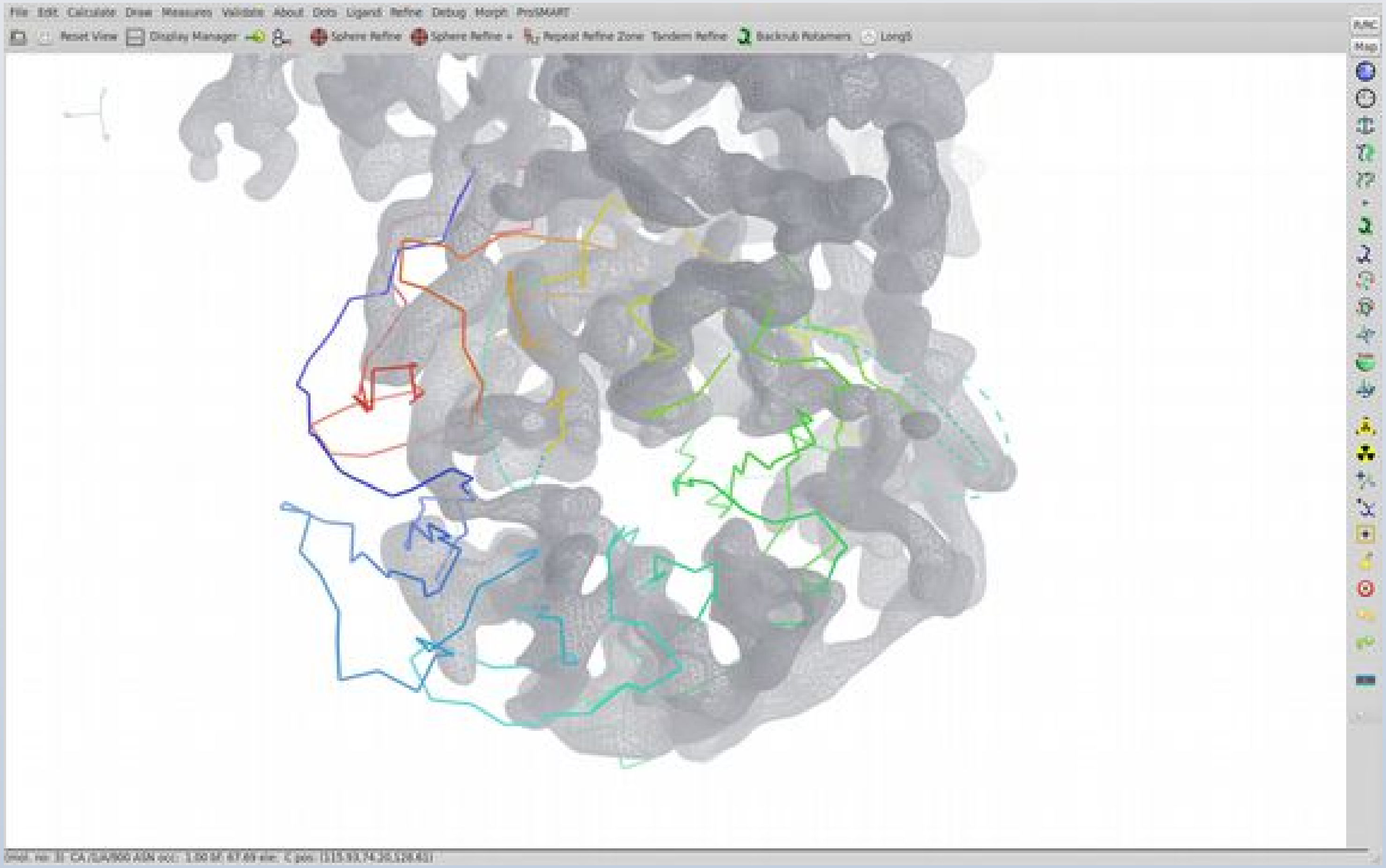


```
(mol. no: 0)  CA /1/A/72 CYS occ:  1.00 bf: 10.64 ele:  C pos: (55.72, 0.77,10.51)
```

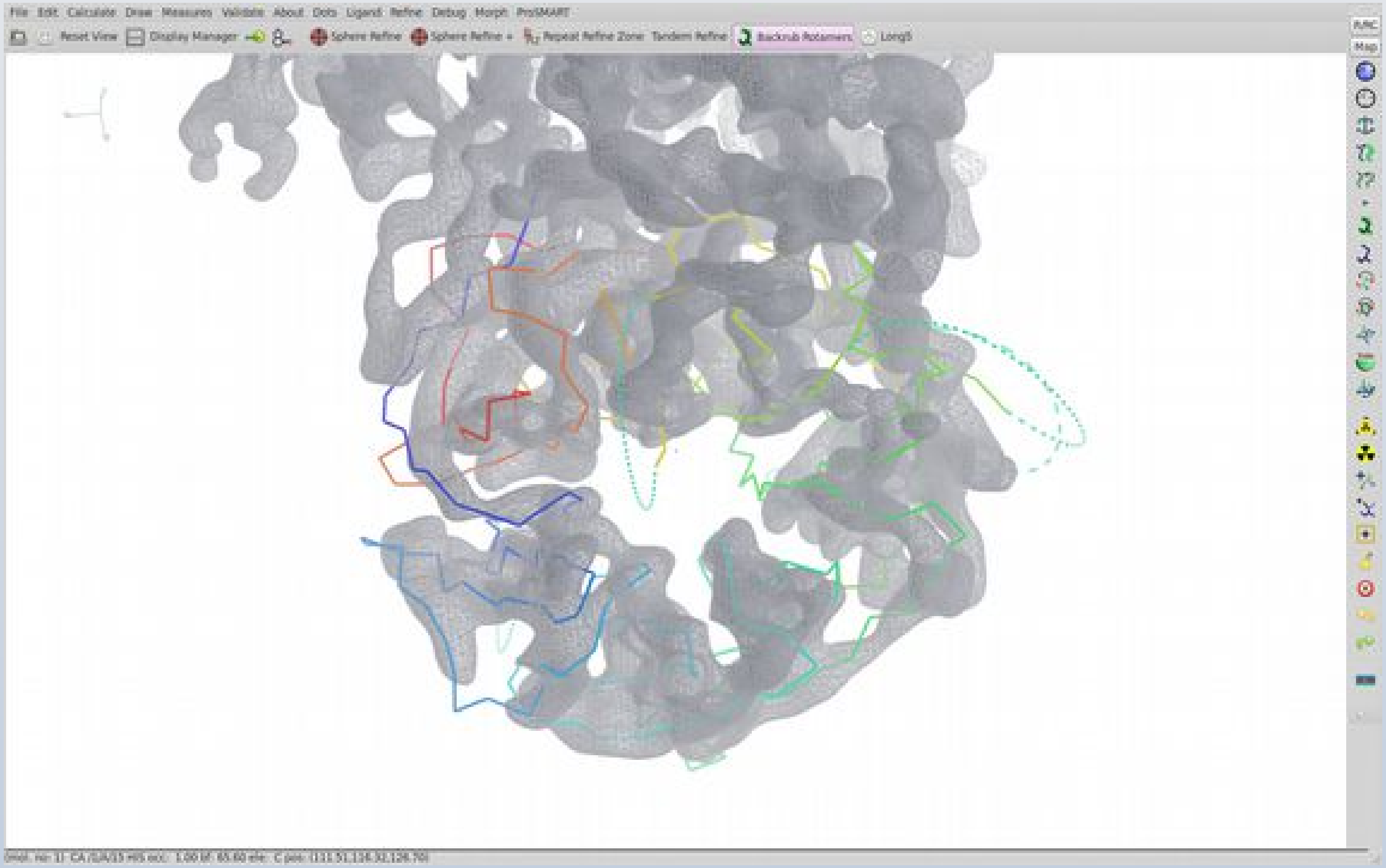
Tutorial Screenshot



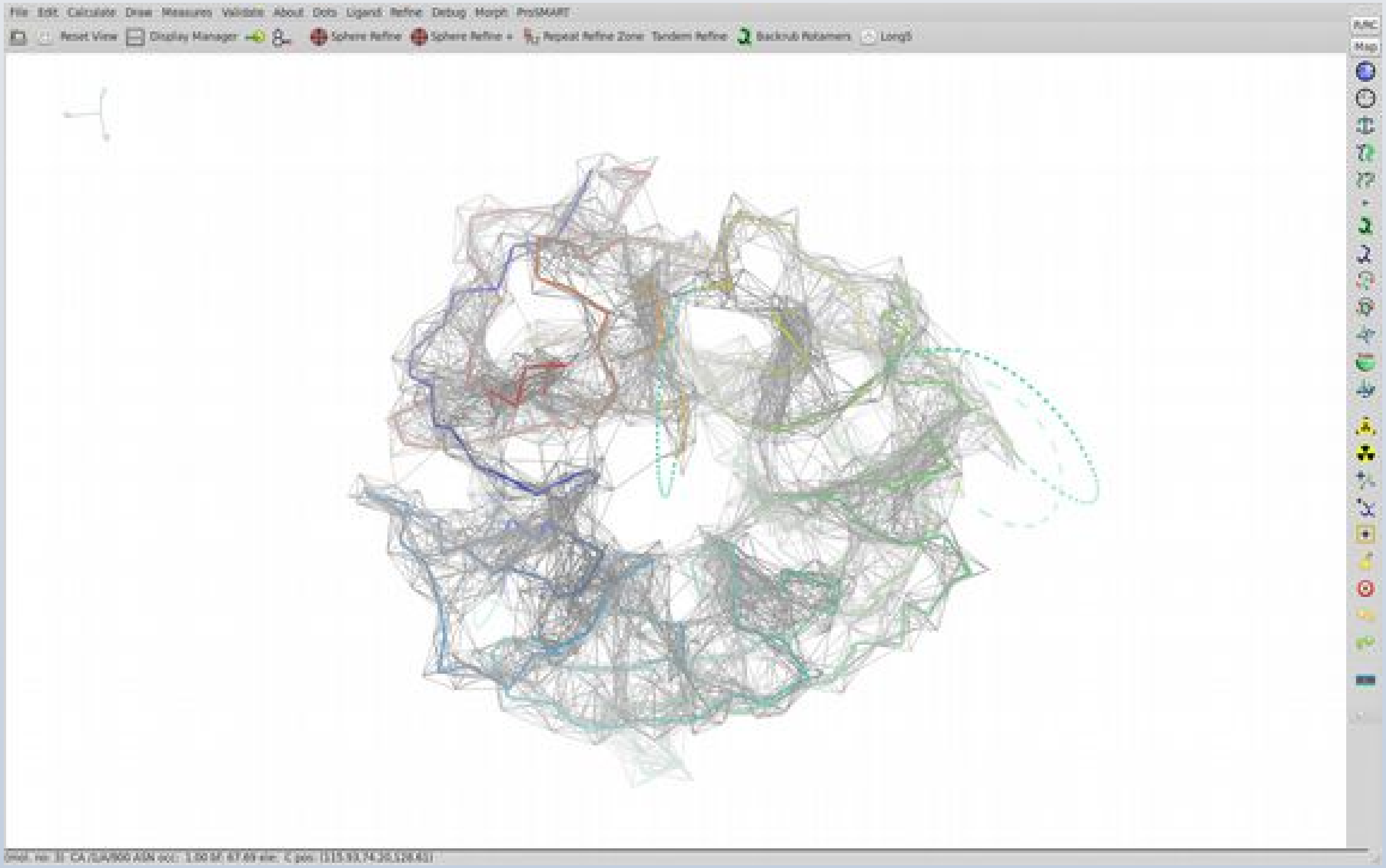
Tutorial Screenshot: Jiggle Fit and Selection



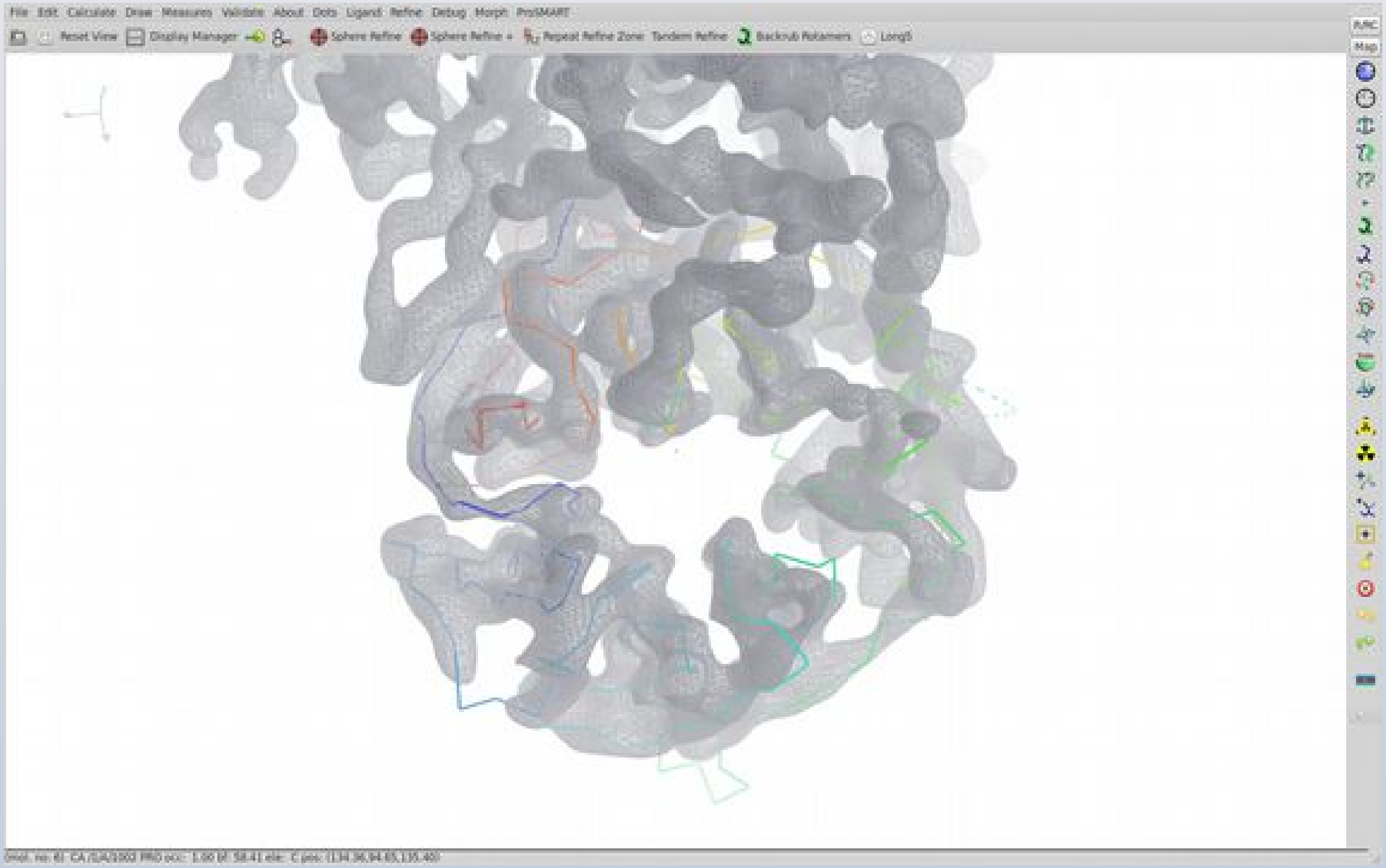
Tutorial Screenshot: Fitting Selection



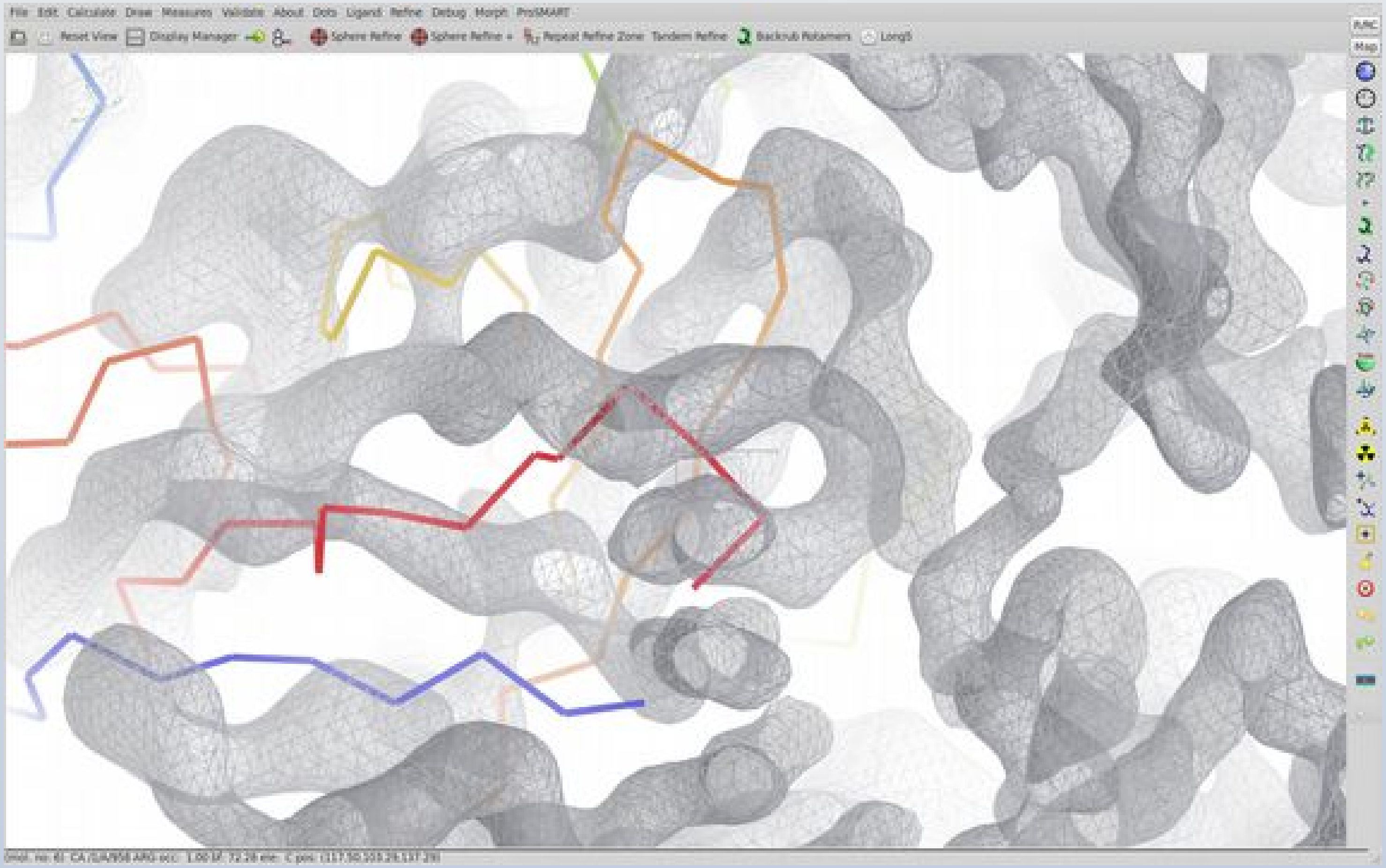
Generate GM Restraints



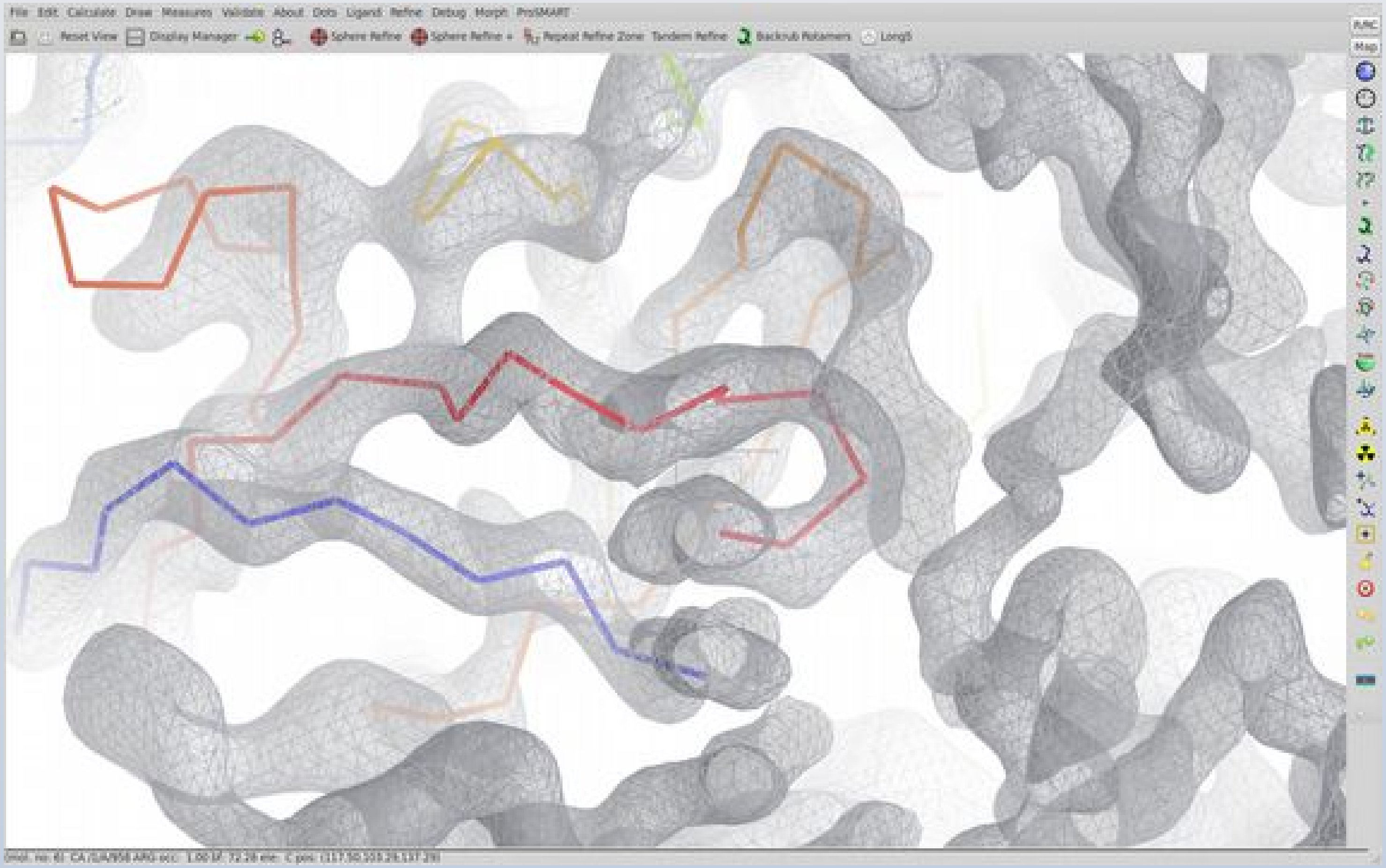
Post Real-Space Refinement



Zoom: Model Before RSR

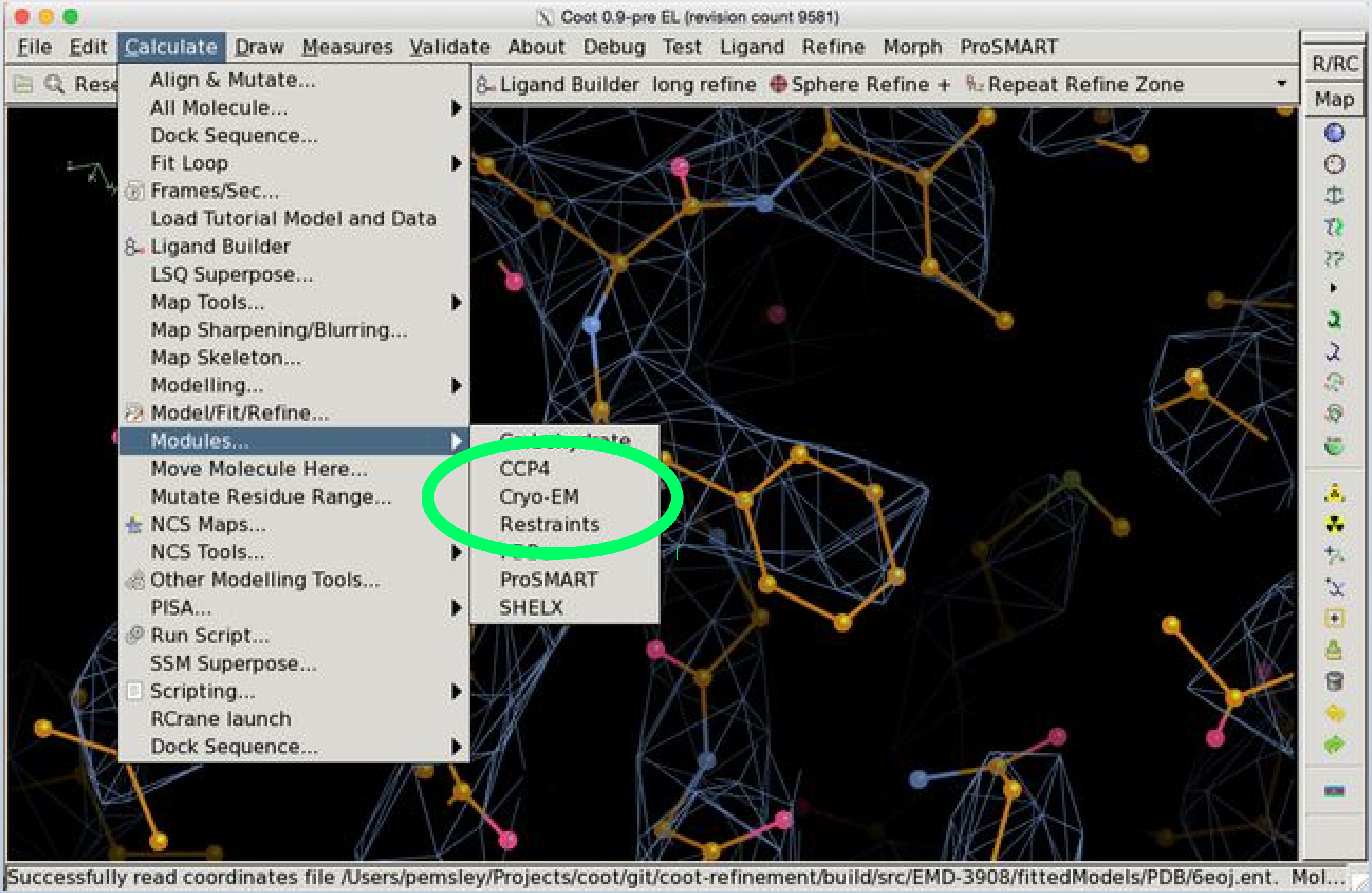


Zoom: Model After RSR



A Few Notes on Usage

Cryo-EM Module

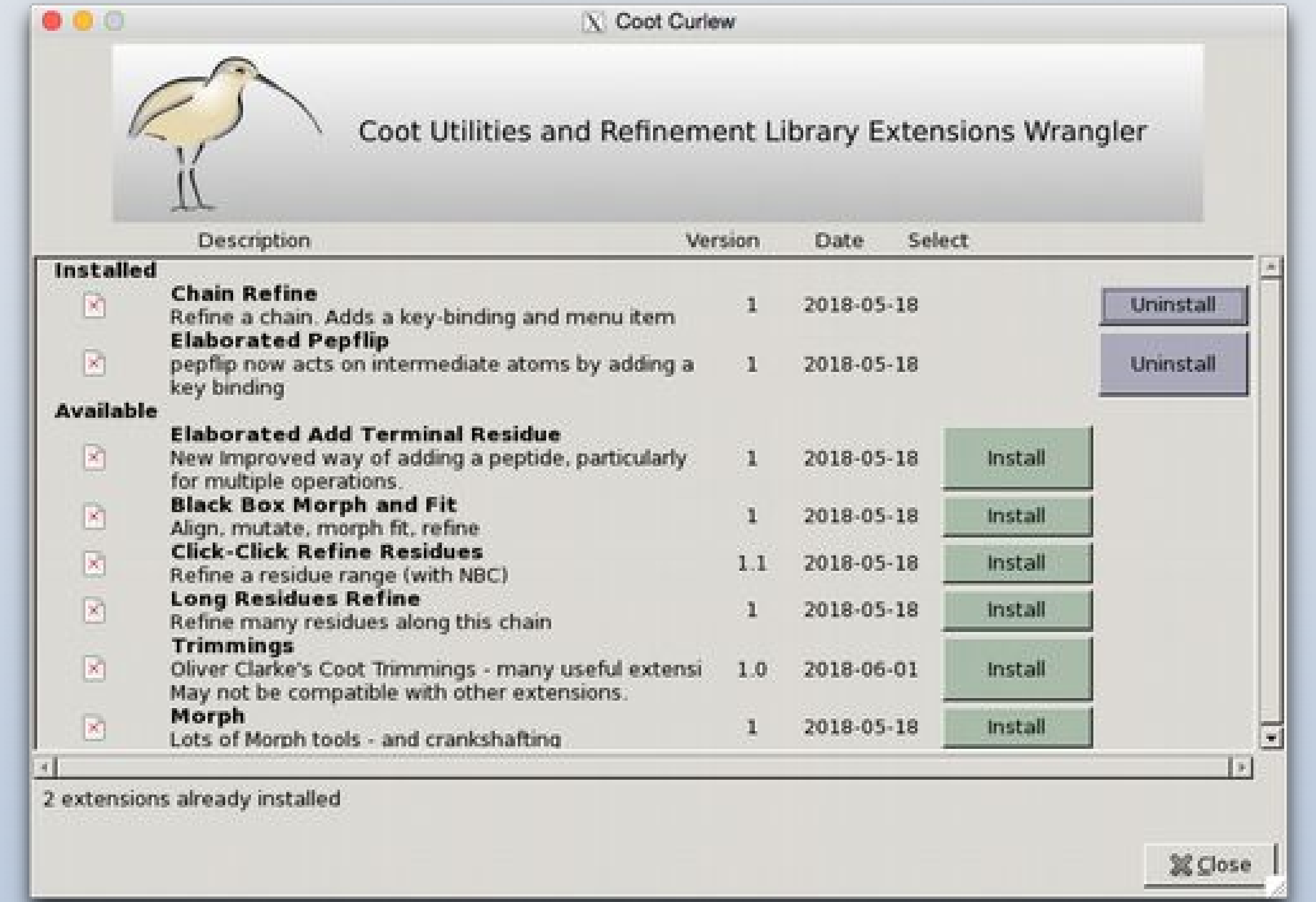


Successfully read coordinates file /Users/pemsley/Projects/coot/git/coot-refinement/build/src/EMD-3908/fittedModels/PDB/6eoj.ent. Mol...

CURLEW:

Coot Utilites and Refinement Library Extension Wranger

- Easy access to ”interesting” *Coot* scripts



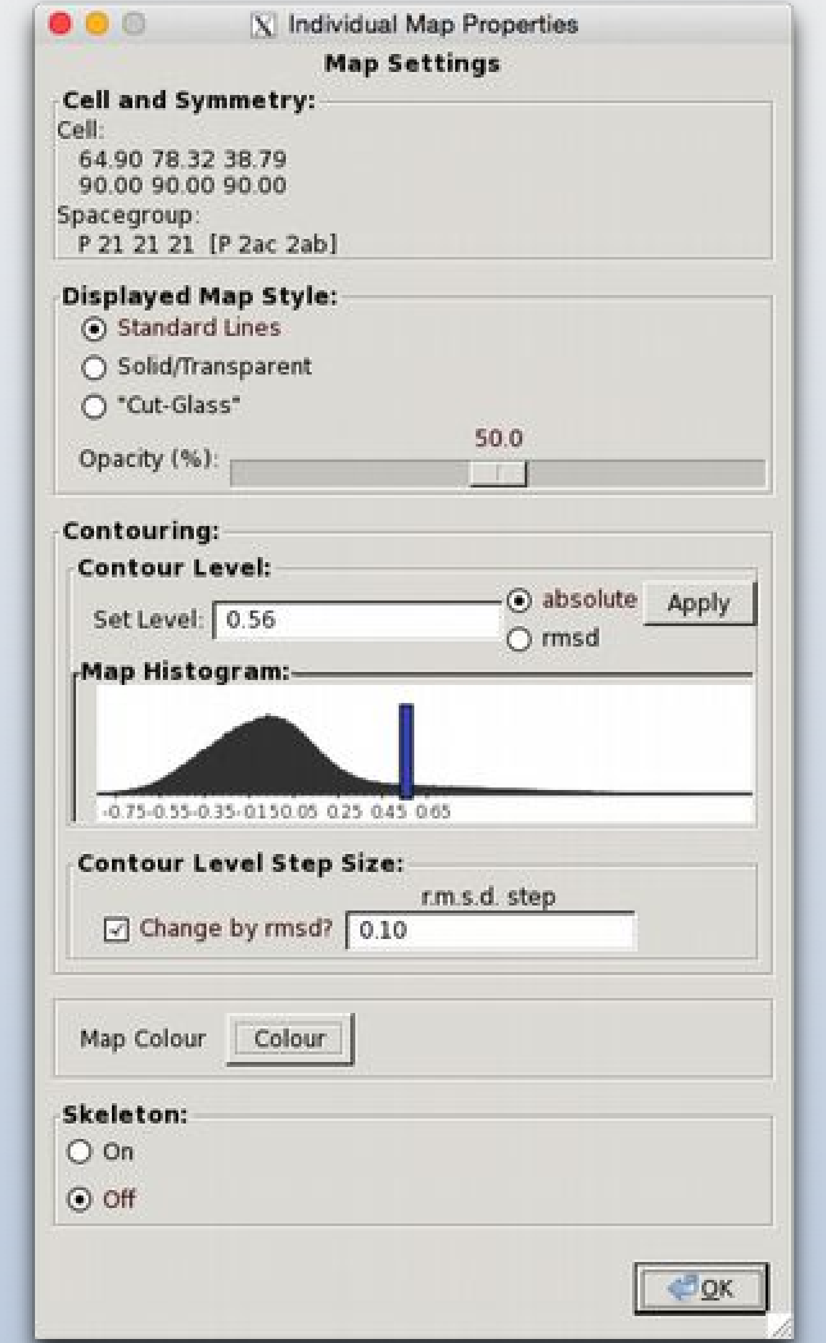
Coot Curlew

Coot Utilities and Refinement Library Extensions Wrangler

	Description	Version	Date	Select
Installed				
☒	**Chain Refine** Refine a chain. Adds a key-binding and menu item	1	2018-05-18	Uninstall
☒	**Elaborated Pepflip** pepflip now acts on intermediate atoms by adding a key binding	1	2018-05-18	Uninstall
Available				
☒	**Elaborated Add Terminal Residue** New Improved way of adding a peptide, particularly for multiple operations.	1	2018-05-18	Install
☒	**Black Box Morph and Fit** Align, mutate, morph fit, refine	1	2018-05-18	Install
☒	**Click-Click Refine Residues** Refine a residue range (with NBC)	1.1	2018-05-18	Install
☒	**Long Residues Refine** Refine many residues along this chain	1	2018-05-18	Install
☒	**Trimmings** Oliver Clarke's Coot Trimmings - many useful extensi May not be compatible with other extensions.	1.0	2018-06-01	Install
☒	**Morph** Lots of Morph tools - and crankshafting	1	2018-05-18	Install

2 extensions already installed

Map Properties



Individual Map Properties

Map Settings

Cell and Symmetry:

Cell:
64.90 78.32 38.79
90.00 90.00 90.00

Spacegroup:
P 21 21 21 [P 2ac 2ab]

Displayed Map Style:

- ● Standard Lines
- ○ Solid/Transparent
- ○ "Cut-Glass"

Opacity (%): 50.0

Contouring:

Contour Level:

Set Level: 0.56 ● absolute ○ rmsd

Map Histogram:

Contour Level Step Size:

☑ Change by rmsd? r.m.s.d. step 0.10

Map Colour

Skeleton:

- ○ On
- ● Off

Validation

- Update Ramachandran Data
  - Thanks to Andrea Thorn and team

- Interactive/Updating Validation

Improved Ramachandran Plot

All

General

ILE/VAL

GLY

PRE-PRO

PRO

Rotamer and Ramachandran Markup

- `set_show_intermediate_atoms_rota_markup(1)`
- `set_show_intermediate_atoms_rama_markup(1)`

Ramachandran Restraints

- Limited in scope
    - Sphere Refine
    - Tandem Refine
- Interactive
- Disapprove of "wholesale and blind"

Usage Note: The “Active Residue”

- Traditionally we used *Coot* this way:
  - <Tool Select> <Atom Select (pick)> <Atom Select (pick)>
  - But that, in retrospect, is cumbersome
- Increasingly, *Coot* tools are using the concept of the “Active Residue”
  - the residue at the centre of the screen
- For tools that act on one residue, that's straightforward
- For tools such as Sphere Refine and Residue Range Refine the residue selection is implicit
  - so we have to “eye-ball" it
  - Future *Coot*s will make the “Active Residue” more obvious

Acknowledgements

- LMB:
  - Garib Murshudov, Rob Nicholls

- Kevin Cowan, Bernhard Lohkamp

- Libraries & Dictionaries:
  - Jane & Dave Richardson
  - Alexei Vagin
  - Eugene Krissinel