# Tools for Cryo-EM Map Fitting

Paul Emsley
MRC Laboratory of Molecular Biology

April 2017

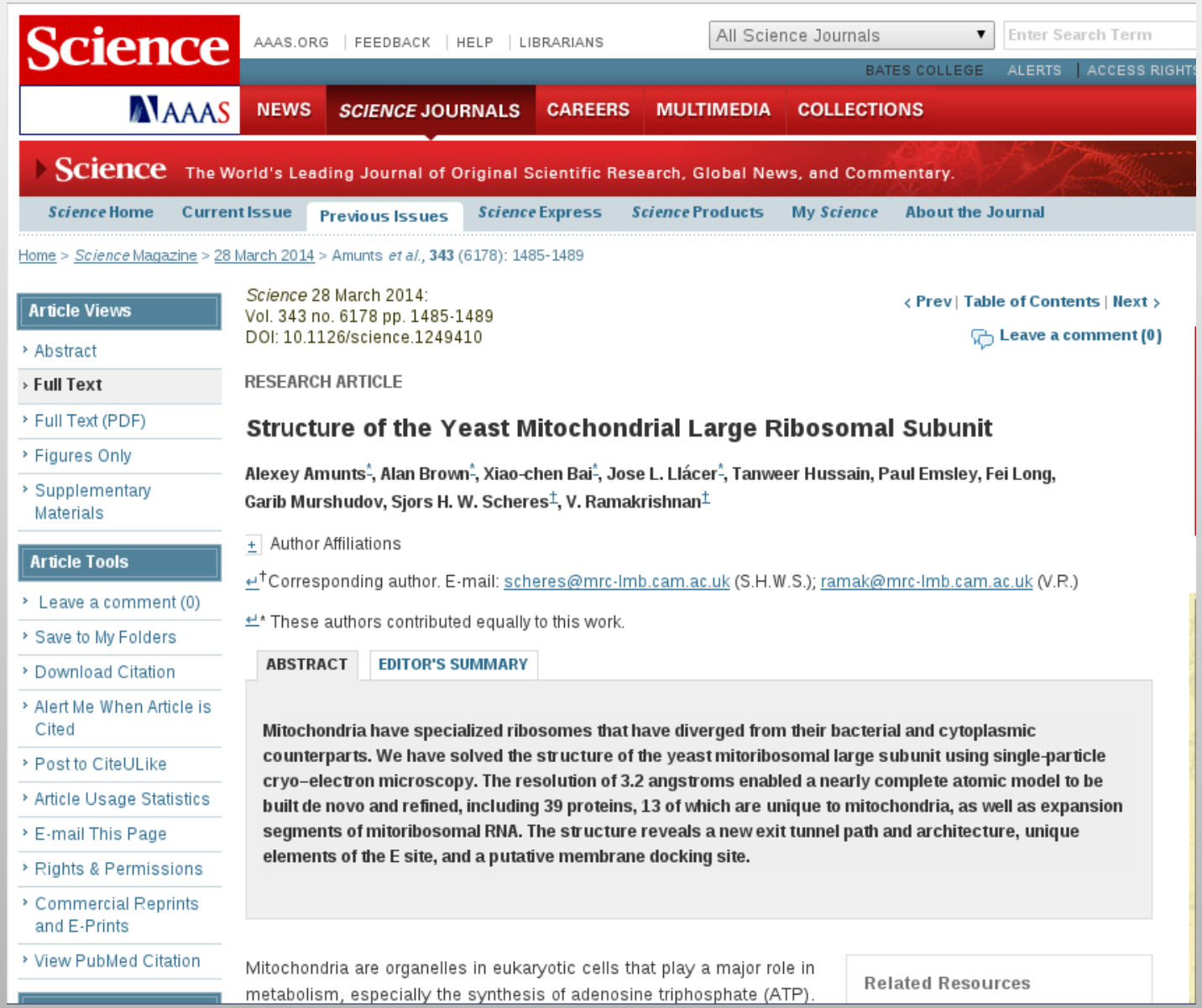
# Cryo-EM model-building

- typically need to move more atoms that one does for crystallography
- the maps are lower resolution and the starting model is further from where you want them to be – usually systematically so
- addressing these needs has been the focus of my work extending/changing coot for cryo-EM

# Cryo-EM Model-building

- autozone multi-residue
- sphere refine, sphere refine+
- sphere regularize, sphere regularize +
- Geman-McClure distance restraints
- Multi-threading/parallel processing
- backrub rotamers

# 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).

# 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

Acta Crystallographica Section D
**Biological Crystallography**
ISSN 1399-0047

# Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions

**Alan Brown, Fei Long, Robert A. Nicholls, Jaan Toots, Paul Emsley and Garib Murshudov***

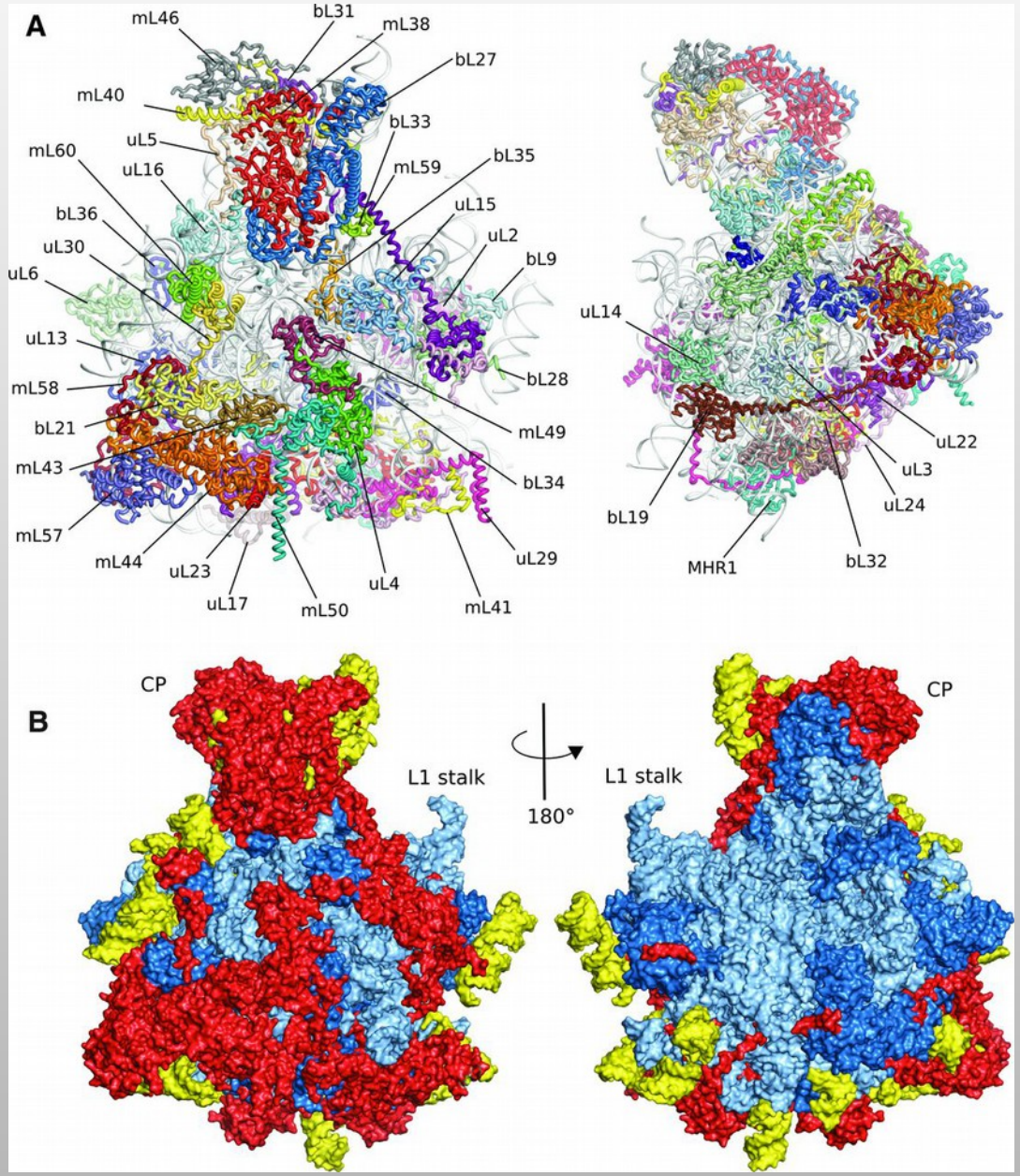
MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, England

Correspondence e-mail: garib@mrc-lmb.cam.ac.uk

Received 3 June 2014
Accepted 1 October 2014

The recent rapid development of single-particle electron cryo-microscopy (cryo-EM) now allows structures to be solved by this method at resolutions close to 3 Å. Here, a number of tools to facilitate the interpretation of EM reconstructions with stereochemically reasonable all-atom models are described. The *BALBES* database has been repurposed as a tool for identifying protein folds from density maps. Modifications to *Coot*, including new Jiggle Fit and morphing tools and improved handling of nucleic acids, enhance its functionality for interpreting EM maps. *REFMAC* has been modified for optimal fitting of atomic models into EM maps. As external structural information can enhance the reliability of the derived atomic models, stabilize refinement and reduce overfitting, *ProSMART* has been extended to generate interatomic distance restraints from nucleic acid reference structures, and a new tool, *LIBG*, has been developed to generate nucleic acid base-pair and parallel-plane restraints. Furthermore, restraint generation has been integrated with visualization and editing in *Coot*, and these restraints have been applied to both real-space refinement in *Coot* and reciprocal-space refinement in *REFMAC*.

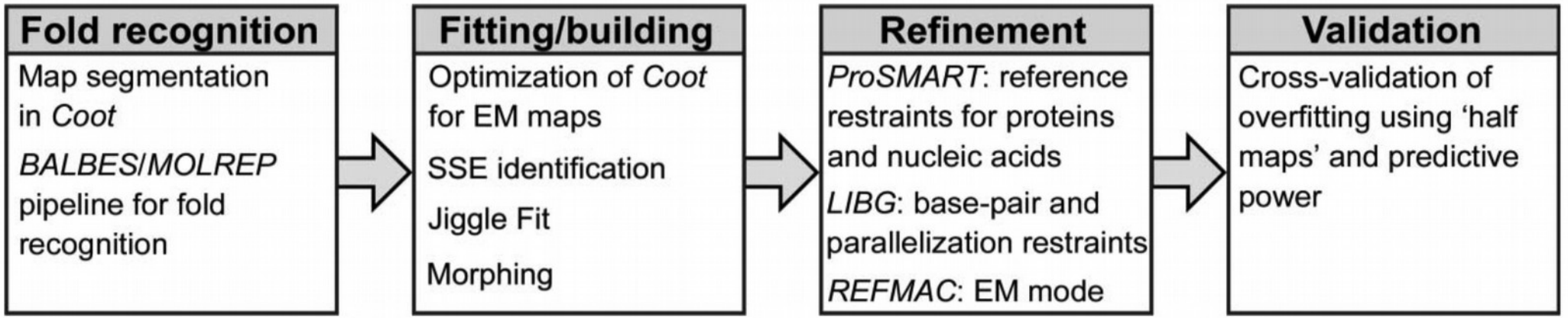
# New Components




blue: conserved
red: found in both
yellow: yeast only

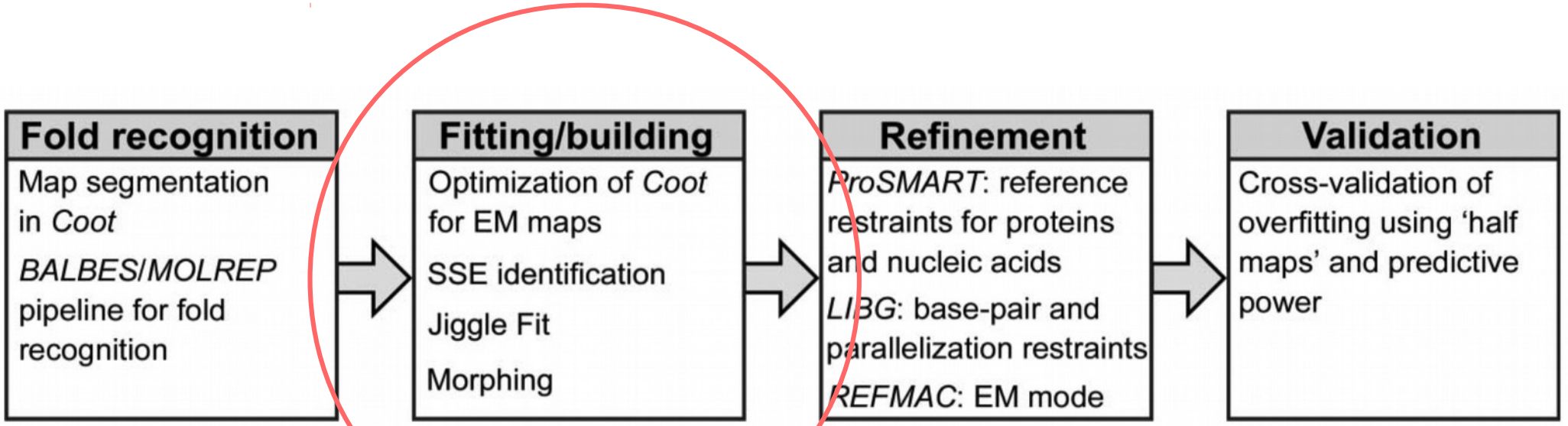
# Model-Building Tools

## Recent Developments



# Model-Building Tools

## Recent Developments

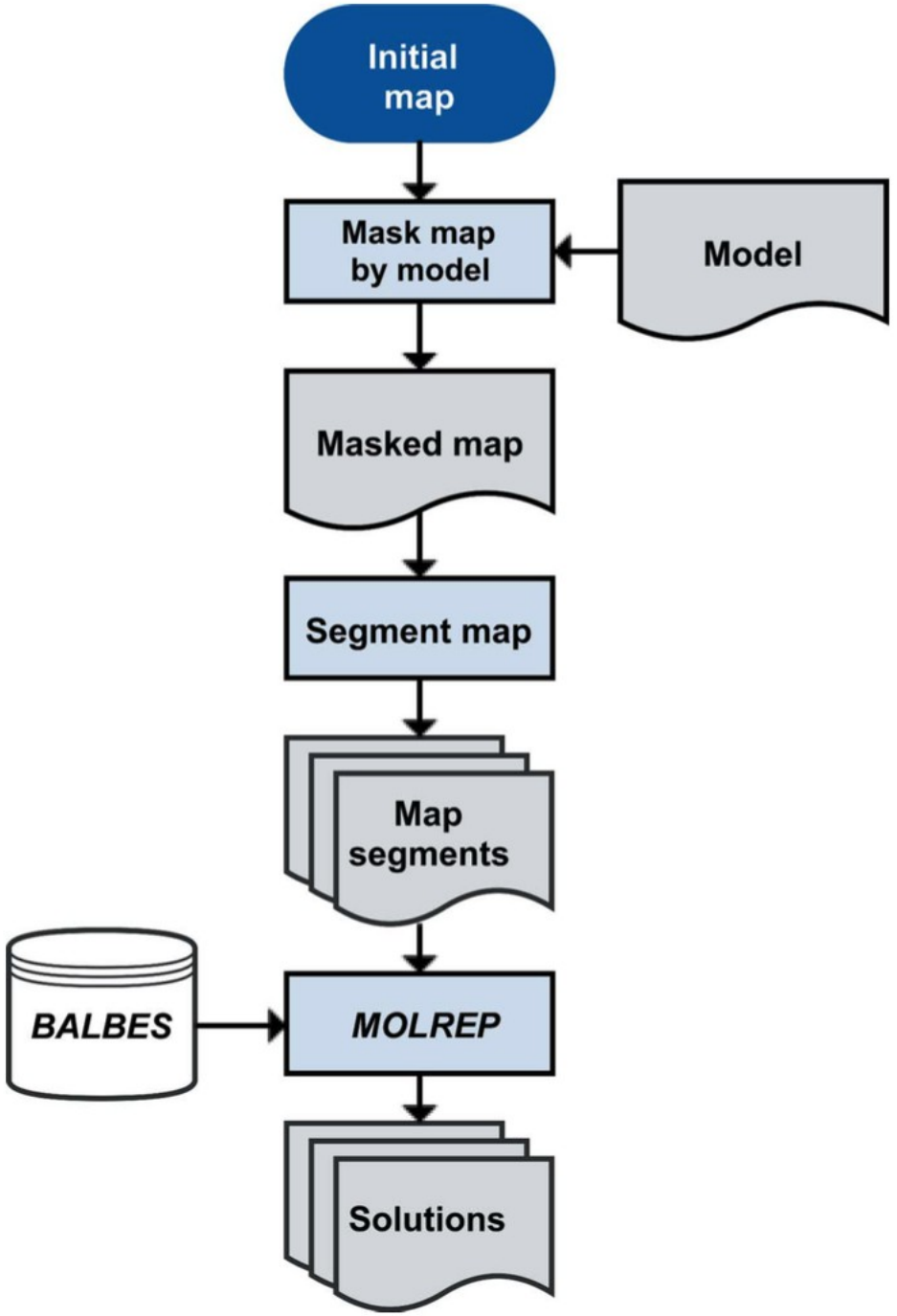


# Cryo-EM data

- Ability to collect data from native sources
  - in such cases, the composition of the complex may not be known
  - Mere “docking” of high resolution structures/fragments cannot work
  - At 4A, it may be possible to trace the backbone
  - At better than 4A it may be possible to assign the amino-acid sequence
    - thus search a sequence database for possible matches

# Cryo-EM data

- Alternatively, use fold recognition
- Using the BALBES database
  - (originally design for molecular replacement)
  - screen domains against unknown density


Initial map
Mask map by model
Model
Masked map
Segment map
Map segments
BALBES
MOLREP
Solutions

# Fold Recognition

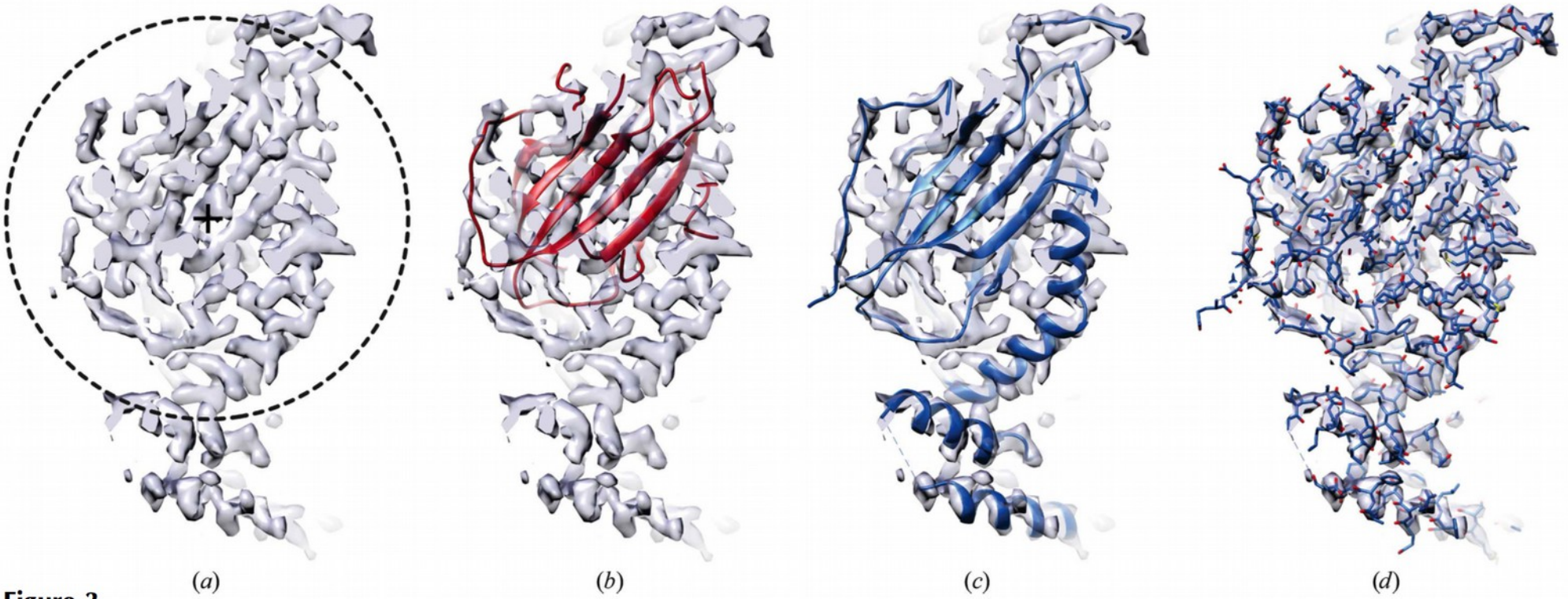



**Figure 3**
Fold recognition can identify template molecules for model building. (*a*) Density map corresponding to the final model of the mitoribosomal protein mL38 with the segmented search map indicated. (*b*) Top solution from the *BALBES–MOLREP* pipeline. (*c*, *d*) Final refined model of mL38 in (*c*) cartoon and (*d*) full-atom representation.

# For What is Coot Useful?

What resolution ranges for cryo-EM?

- “Resolution Revolution” maps
- 2-3.8A is the strength
- seeing side-chains and purines vs pyrimidines

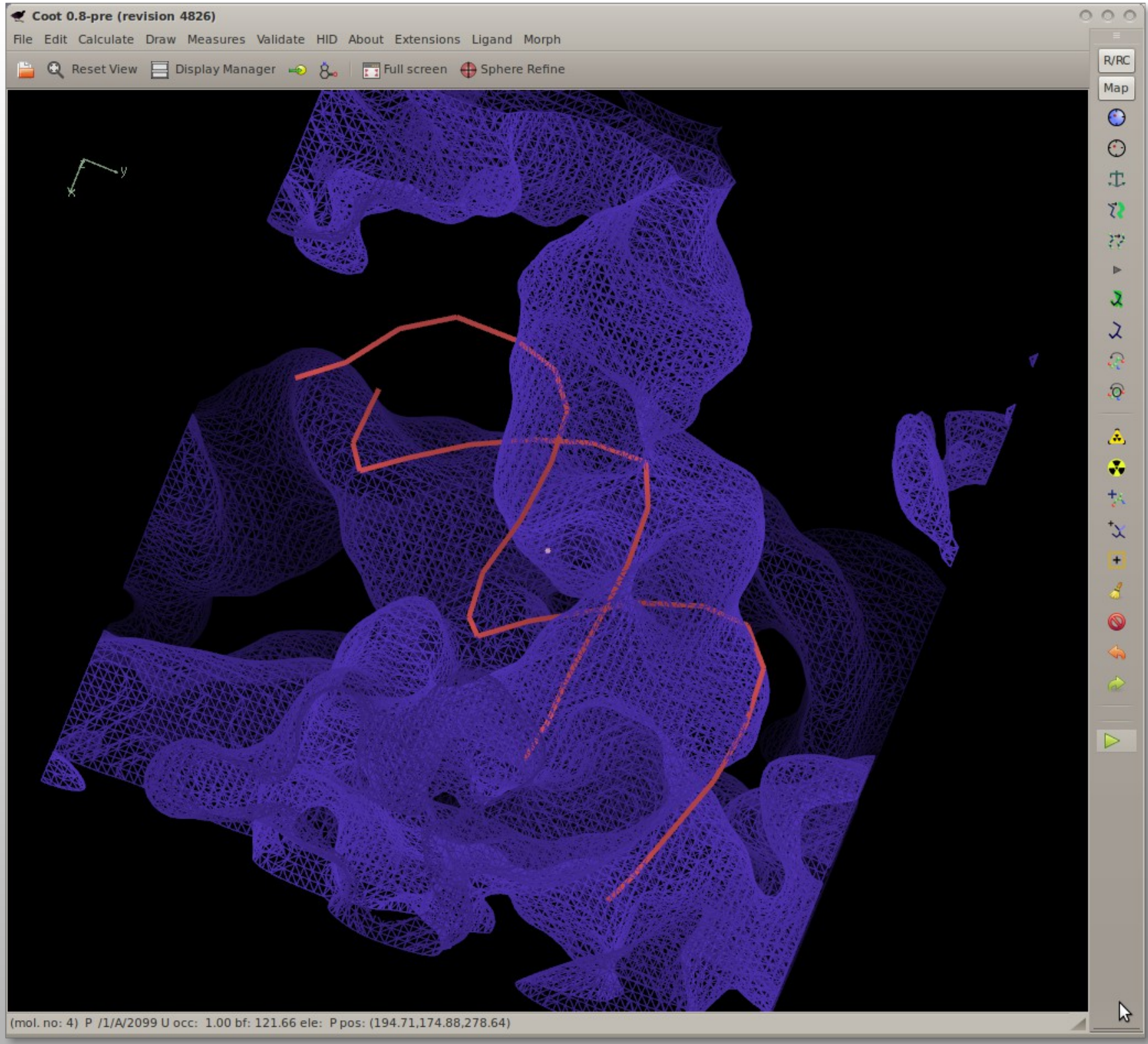
Local good fit of model to density

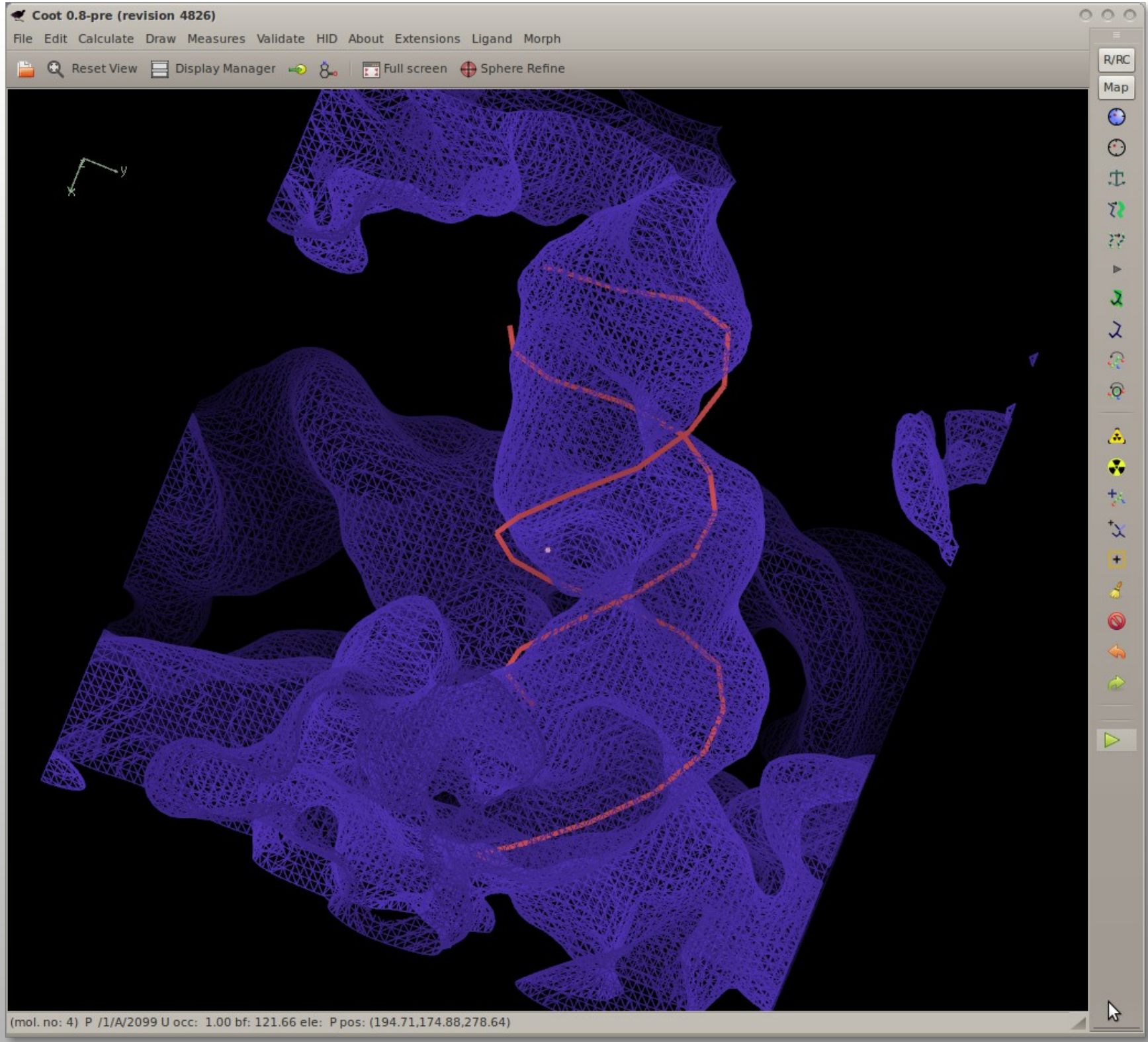
# Jiggle Fit

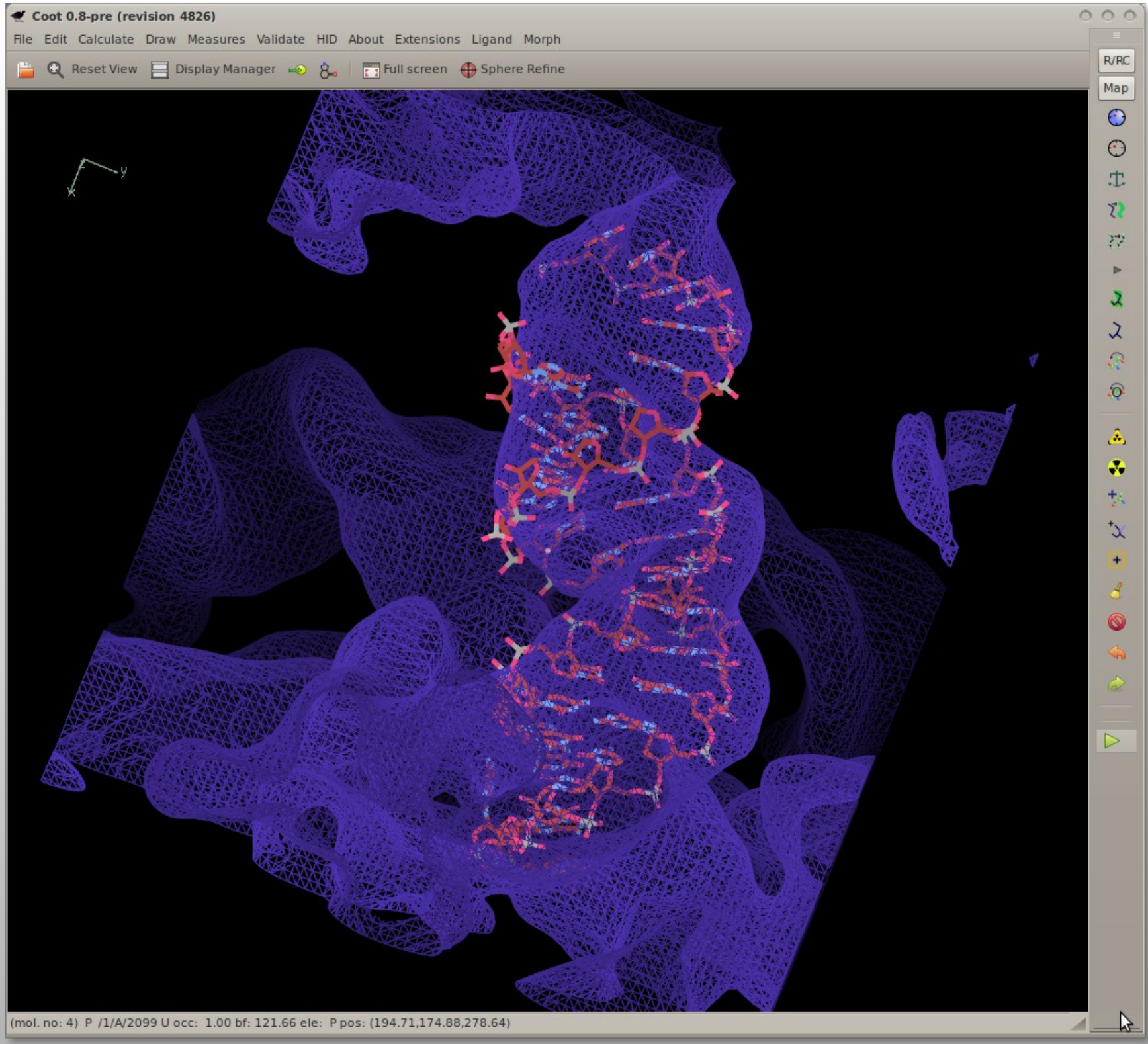
- How do I rotate and translate these atoms to fit the density?
  - 6-dimensional problem
- Originally used to fit simple ligands/solvent molecules to blobs of density
- Now extended to fit arbitrary atom selections
  - *e.g.* by Chain

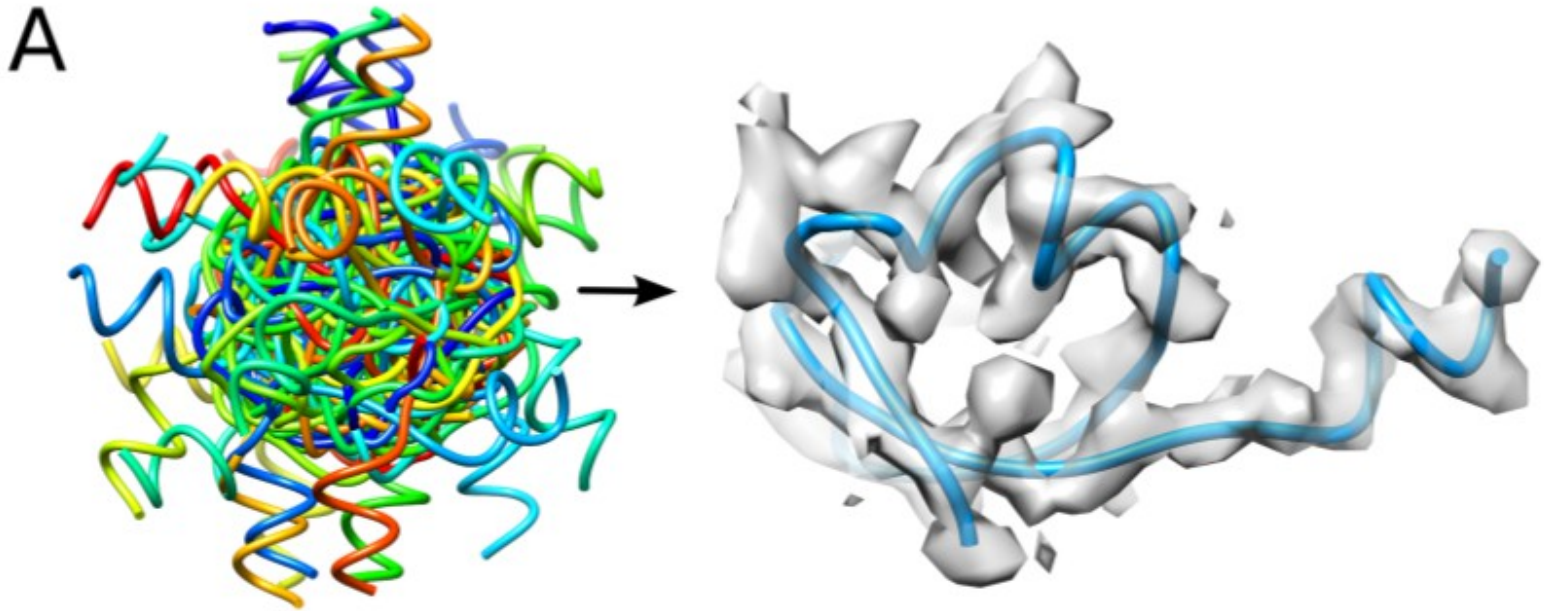
# Jiggle Fit: How it Works

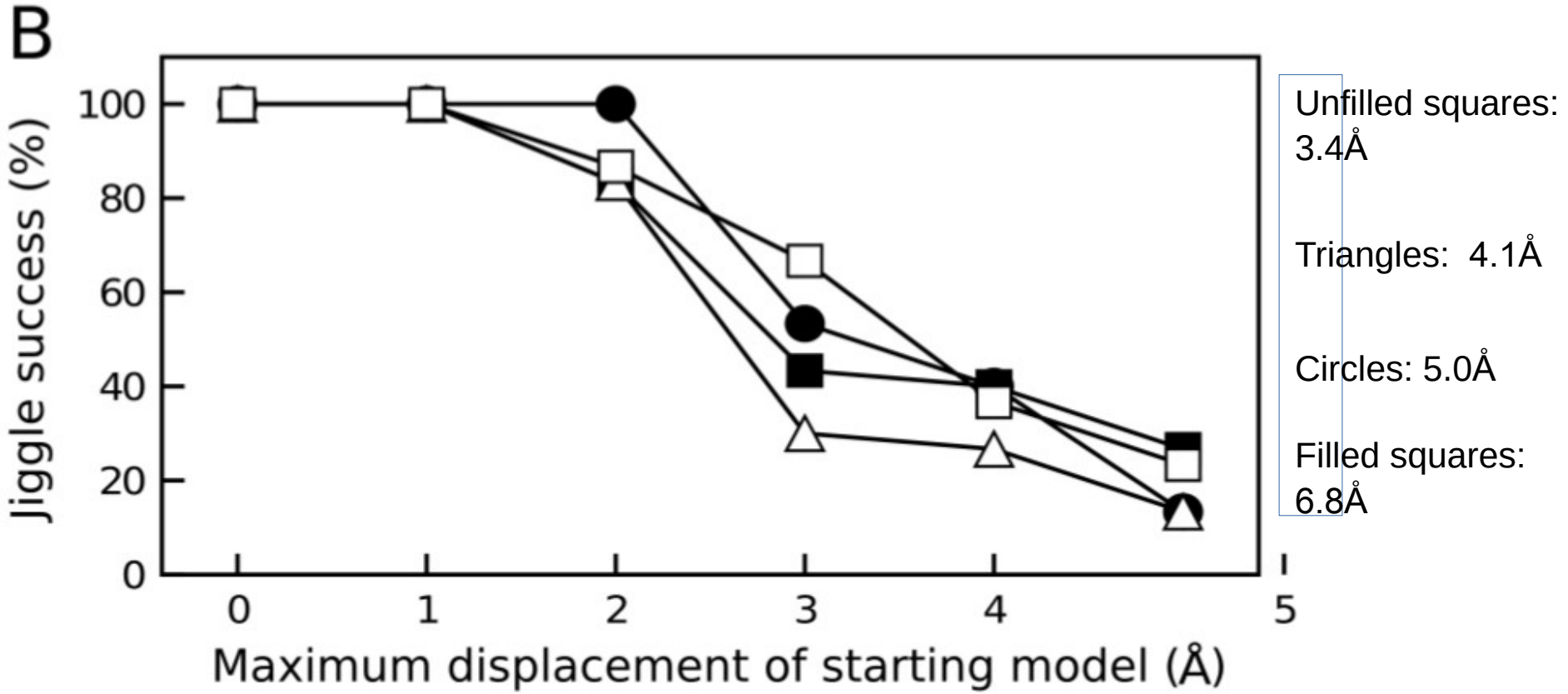
- Loop *n* (say 1000) times:
  - Generate random angles and translations
  - Transform atom selection by these rotations and translation
  - Score and store the fit to density
- Rank density fit scores,
  - Pick top 20 solution, for each of them
    - Rigid body fit and score solutions
    - Pick the 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)


A


B
Jiggle success (%)
100
80
60
40
20
0
0
1
2
3
4
5
Maximum displacement of starting model (Å)
Unfilled squares: 3.4Å
Triangles: 4.1Å
Circles: 5.0Å
Filled squares: 6.8Å

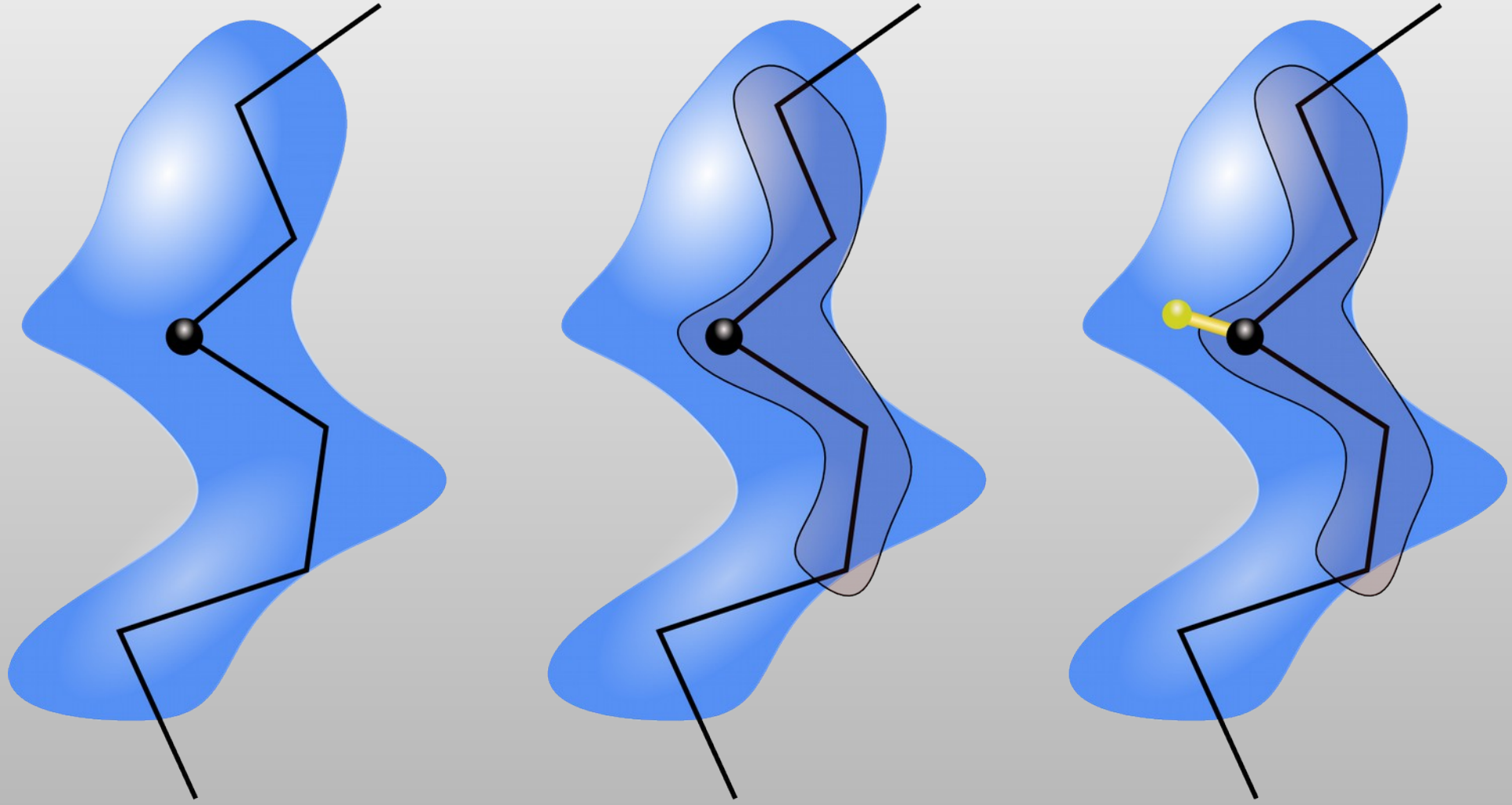
So we have our ideal RNA or homologous protein sitting roughly in the density

(not a great fit)

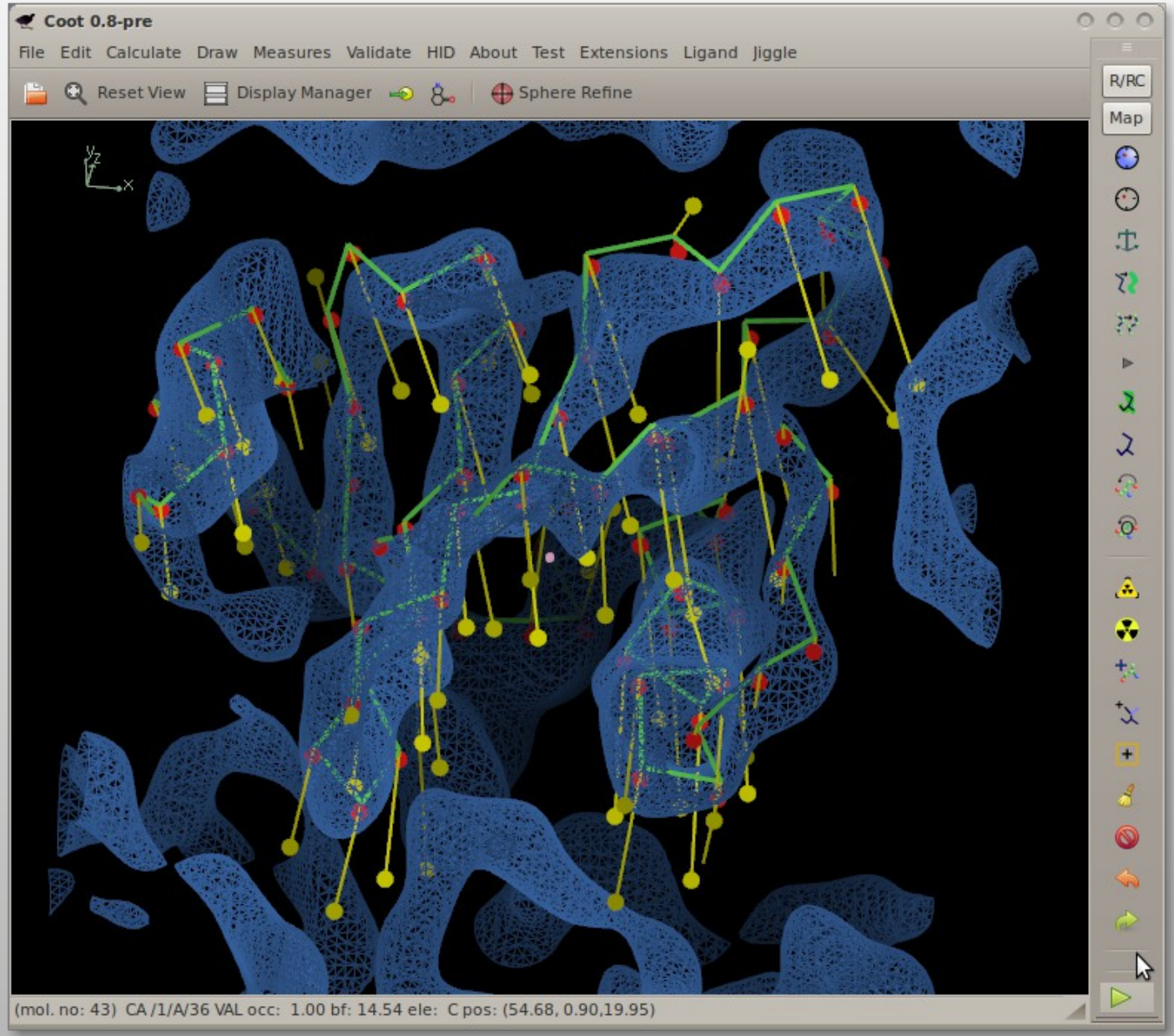
# Model Morphing: How it Works

- 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

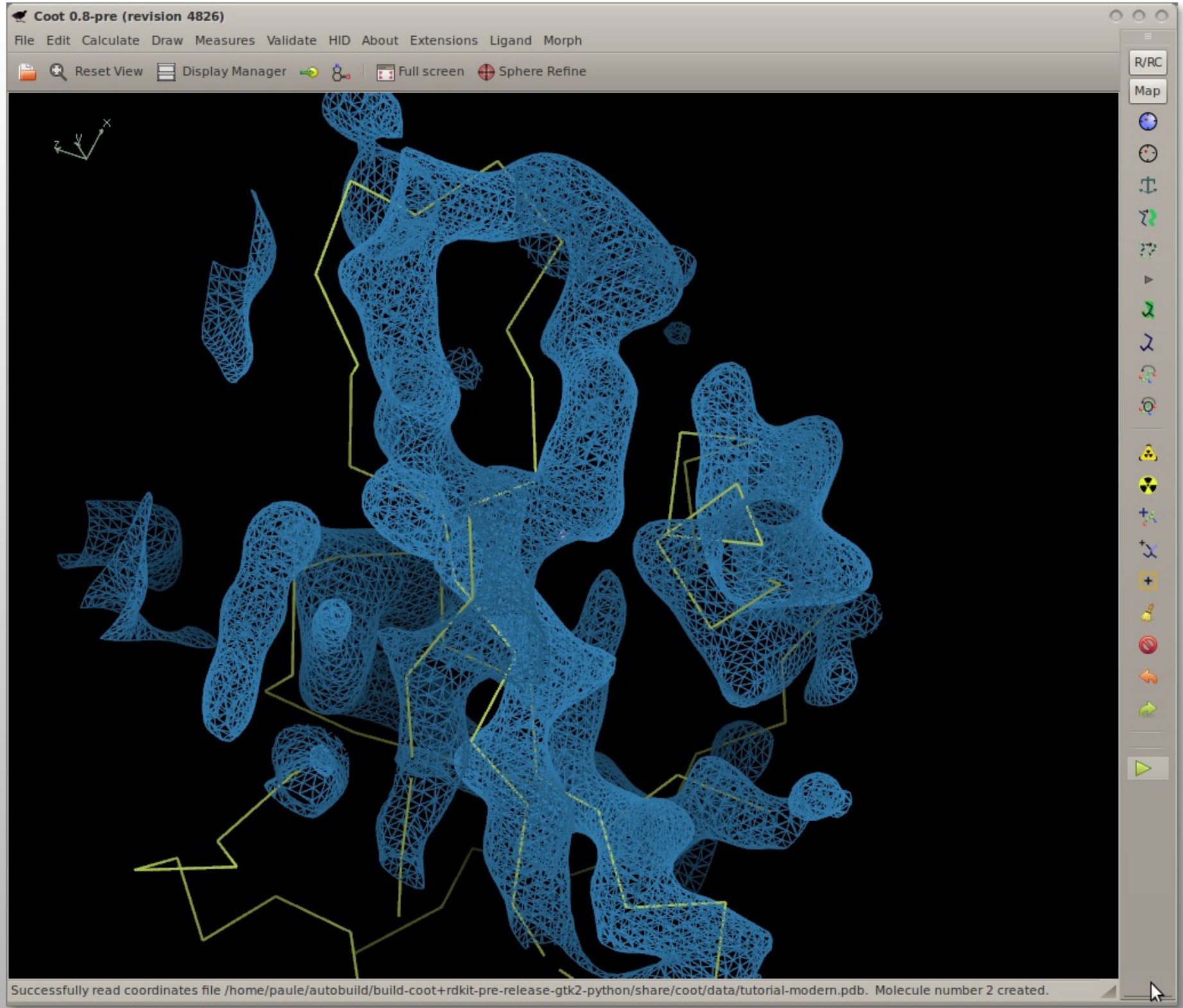


# Model Morphing: Example



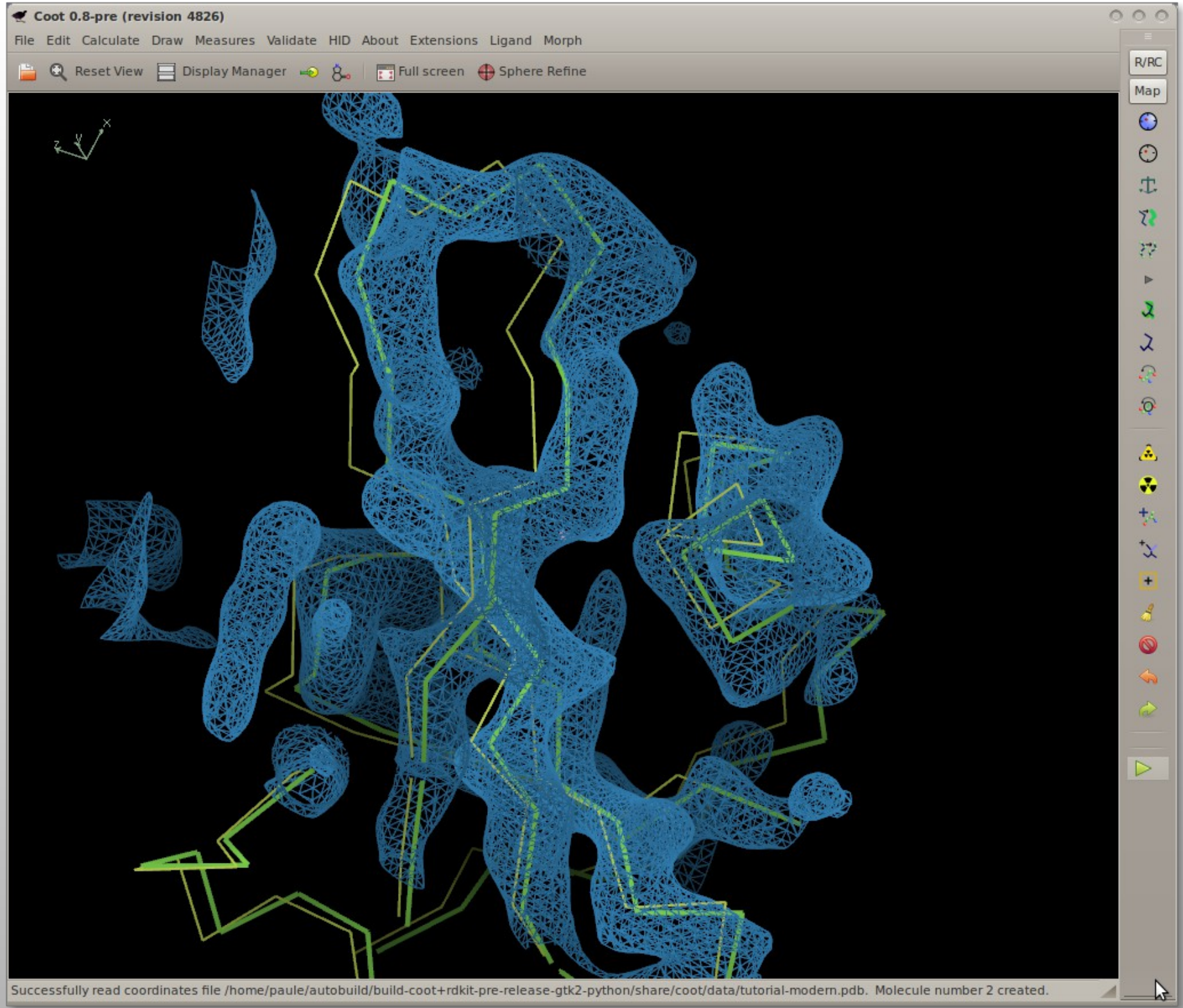
# Model Morphing: Robust Averaging

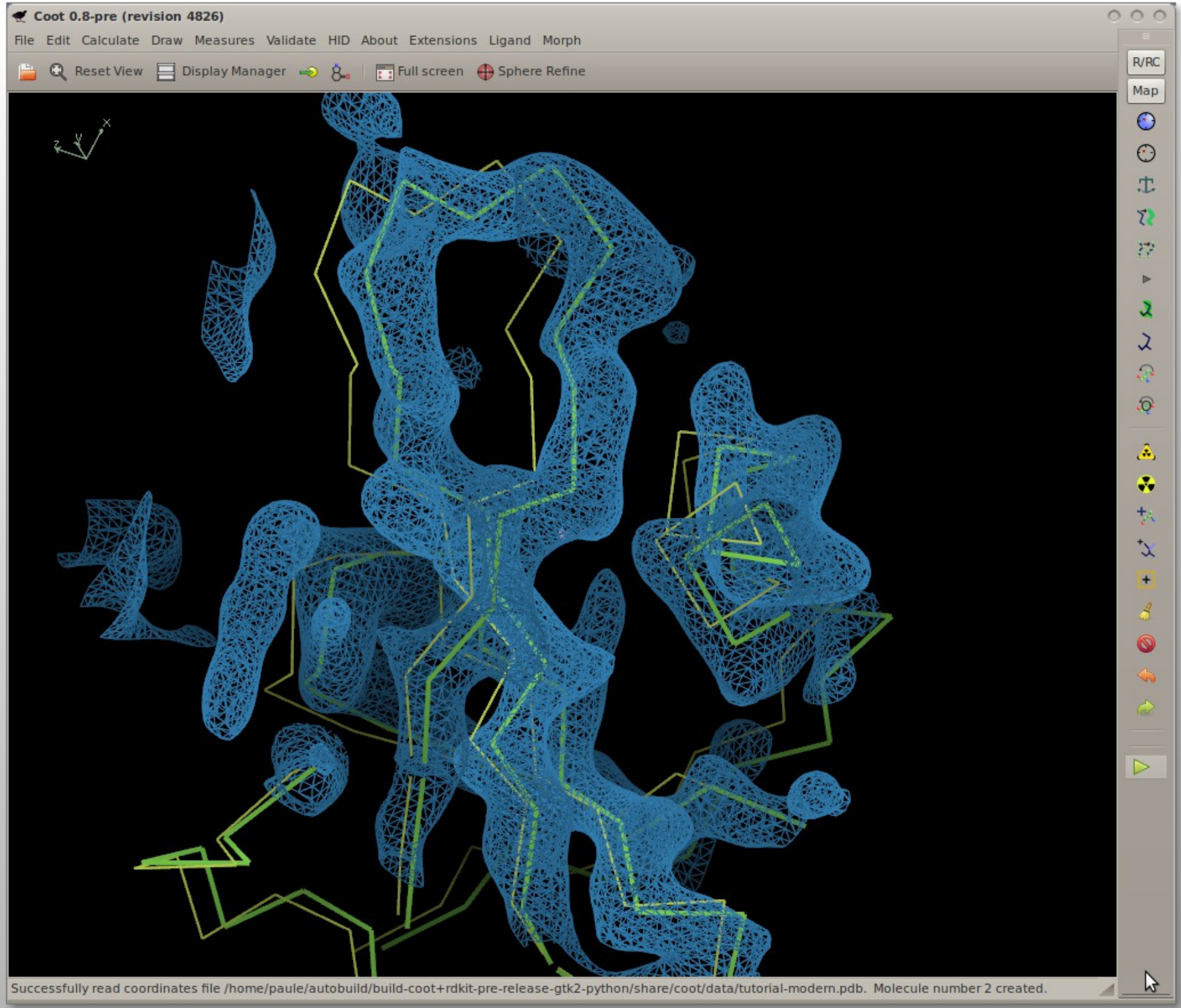
- What are the residues in the environment of a residue?
  - What are their RTs?
  - Create a metric 'distance', sort on that
  - Discard the top and bottom 25%
  - Use remaining RTs to generate average
  - ...which is then applied to central residue
- Repeat for all residues
- Larger environment radii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

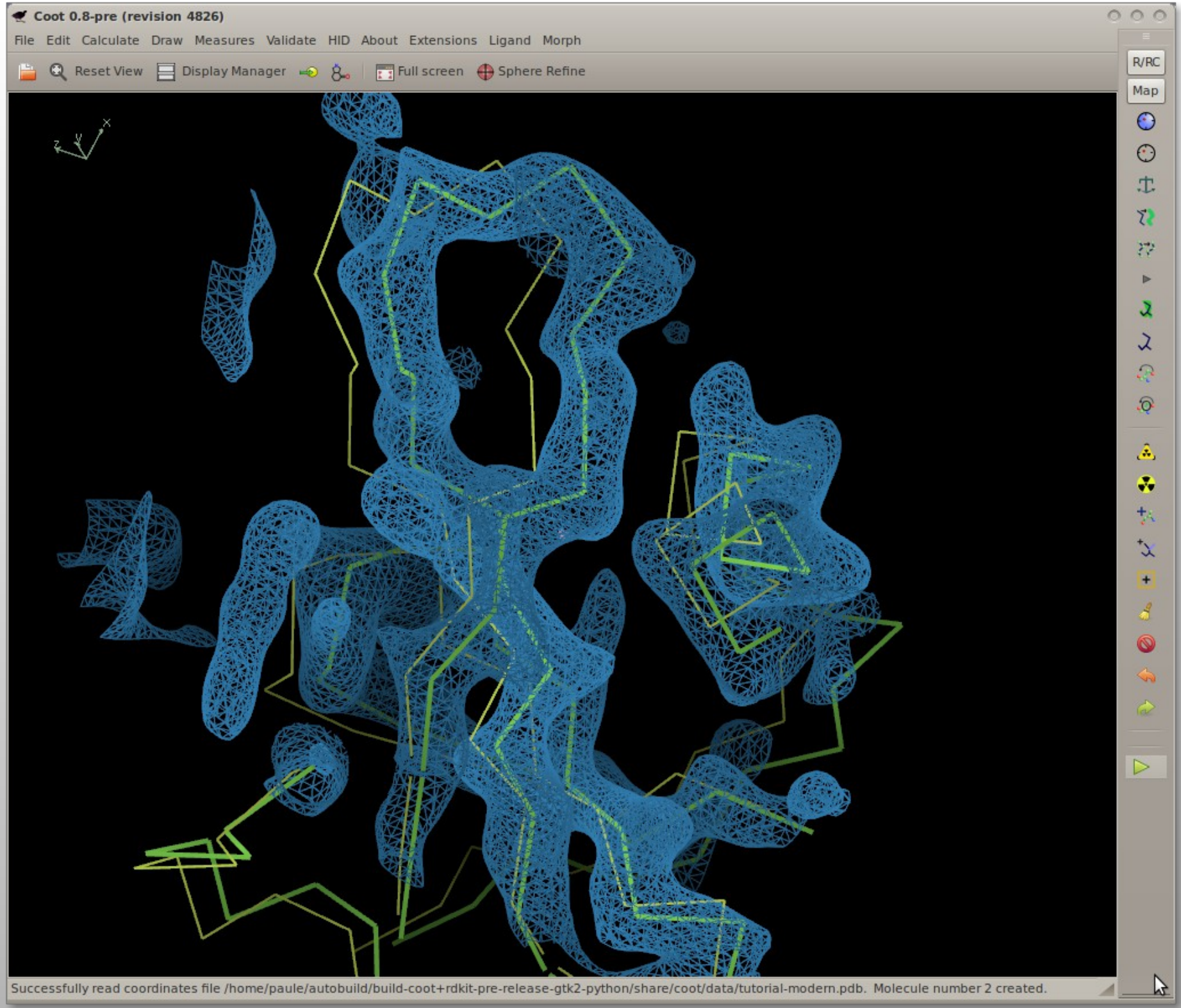

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb.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

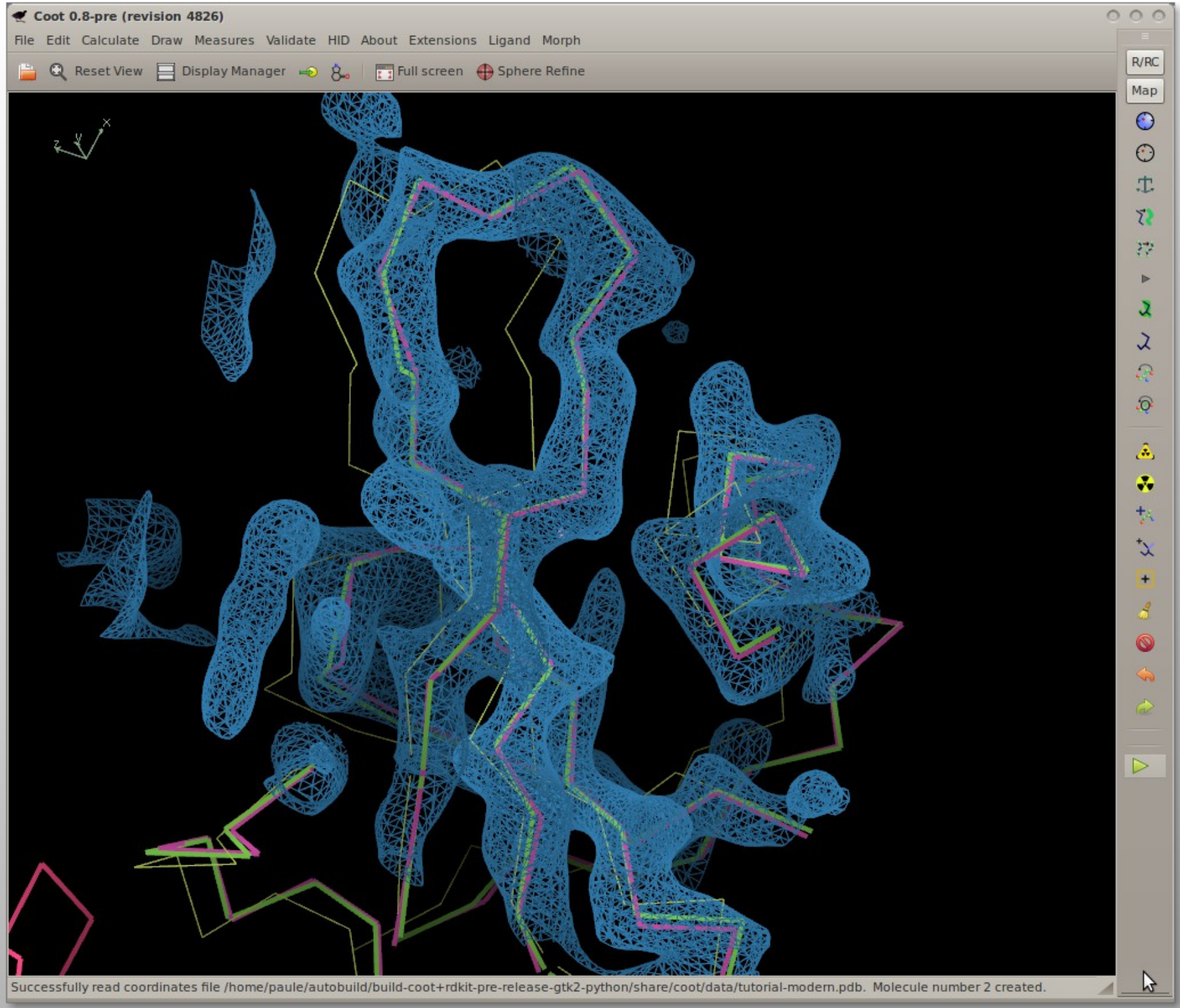
# Model Morphing: Robust Averaging

- 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

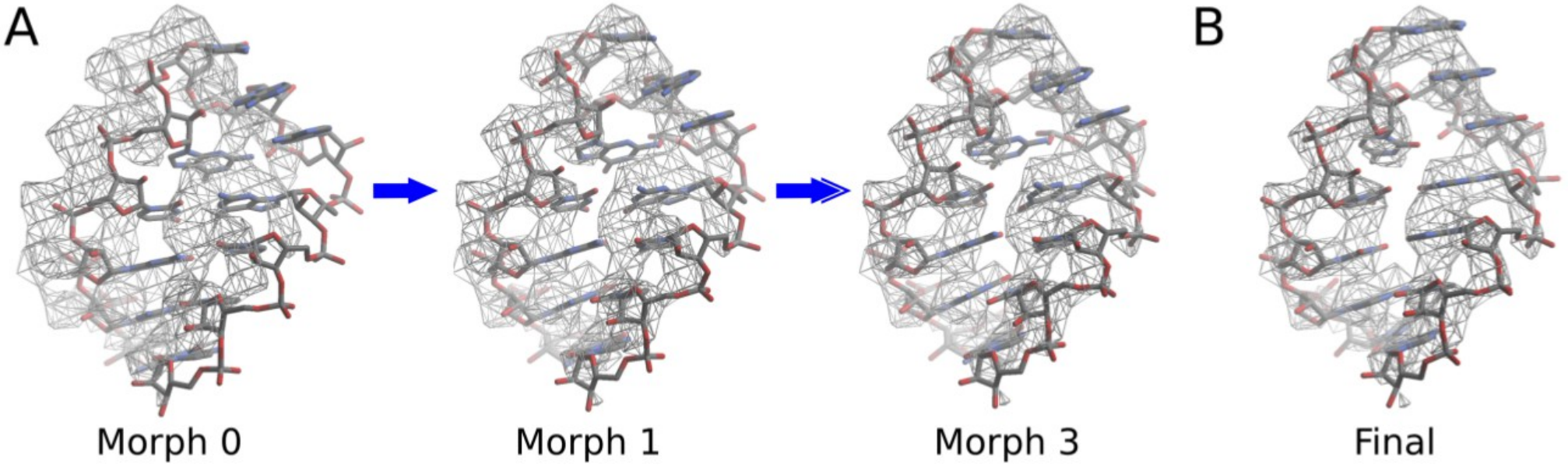

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb.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

# Model Morphing



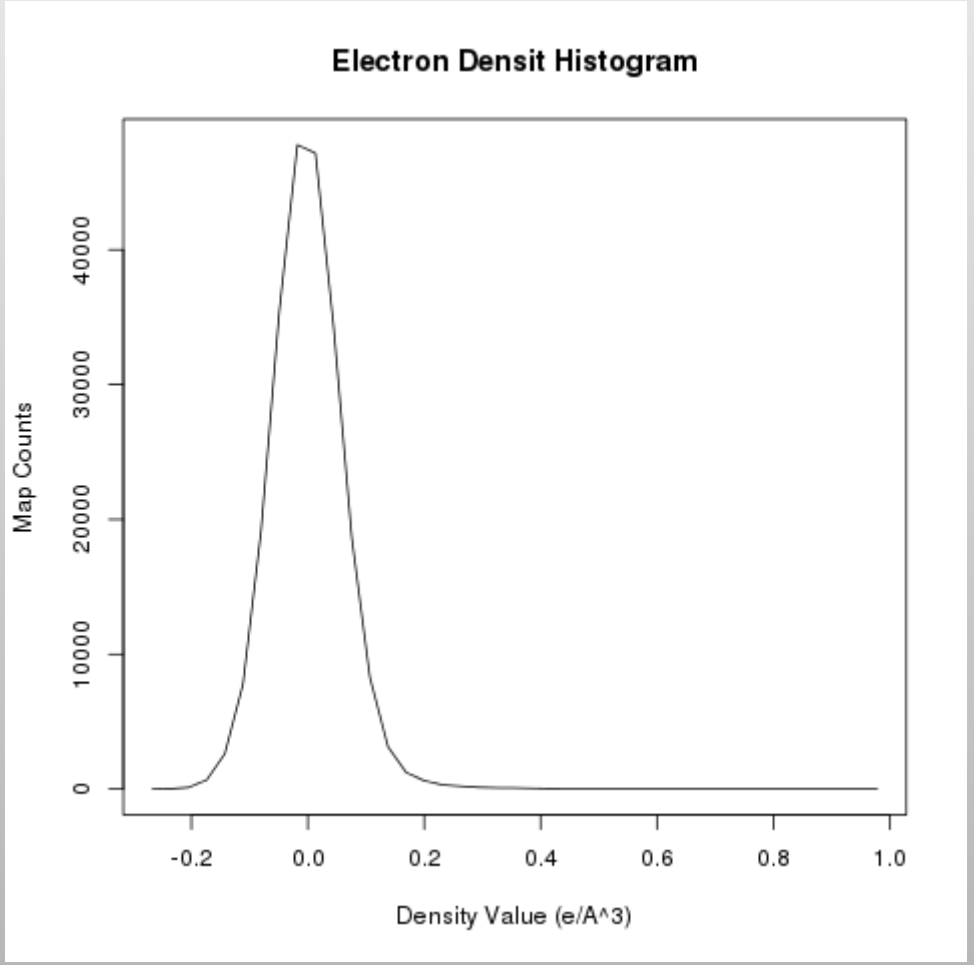
# Helix Fitting

- The distribution of electron density is quite unlike that of x-ray maps
  - e.g. You don't see main-chain atoms at 4 rmsd in x-ray maps
    - regions of dense electron density contribute negatively to helix score

# Helix Fitting

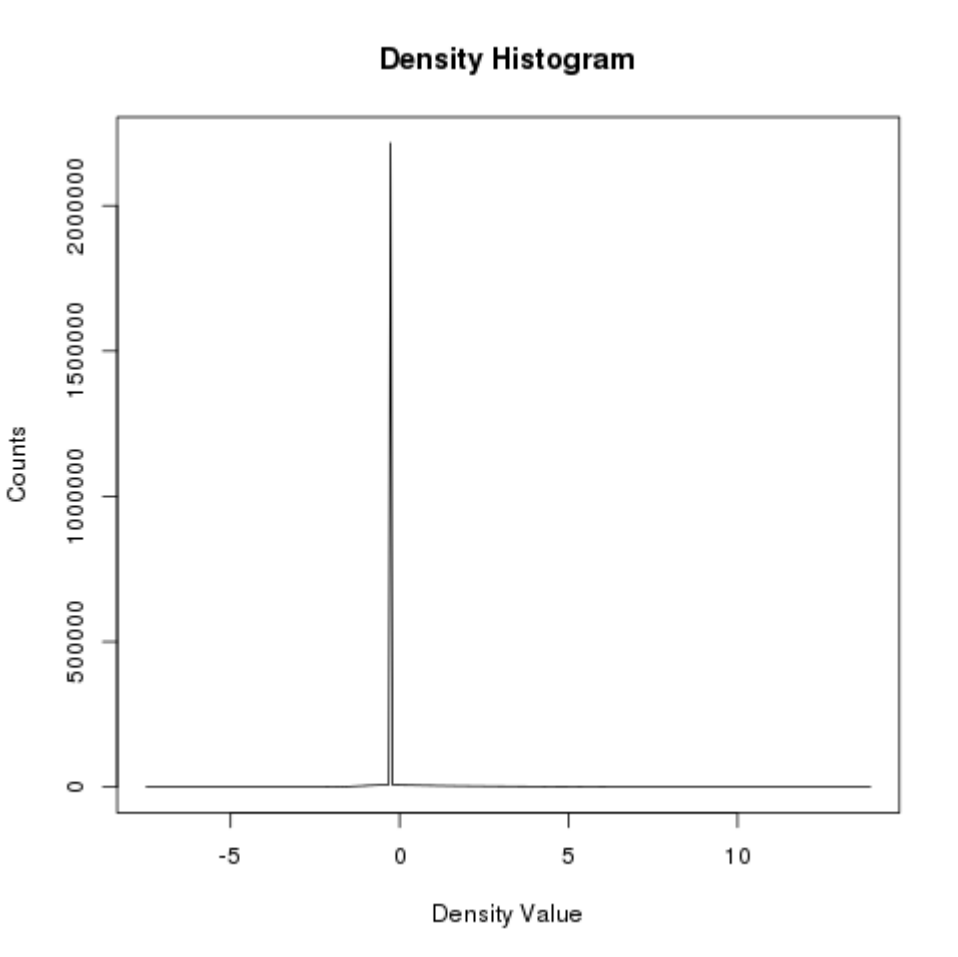
- The distribution of electron density is quite unlike that of x-ray maps

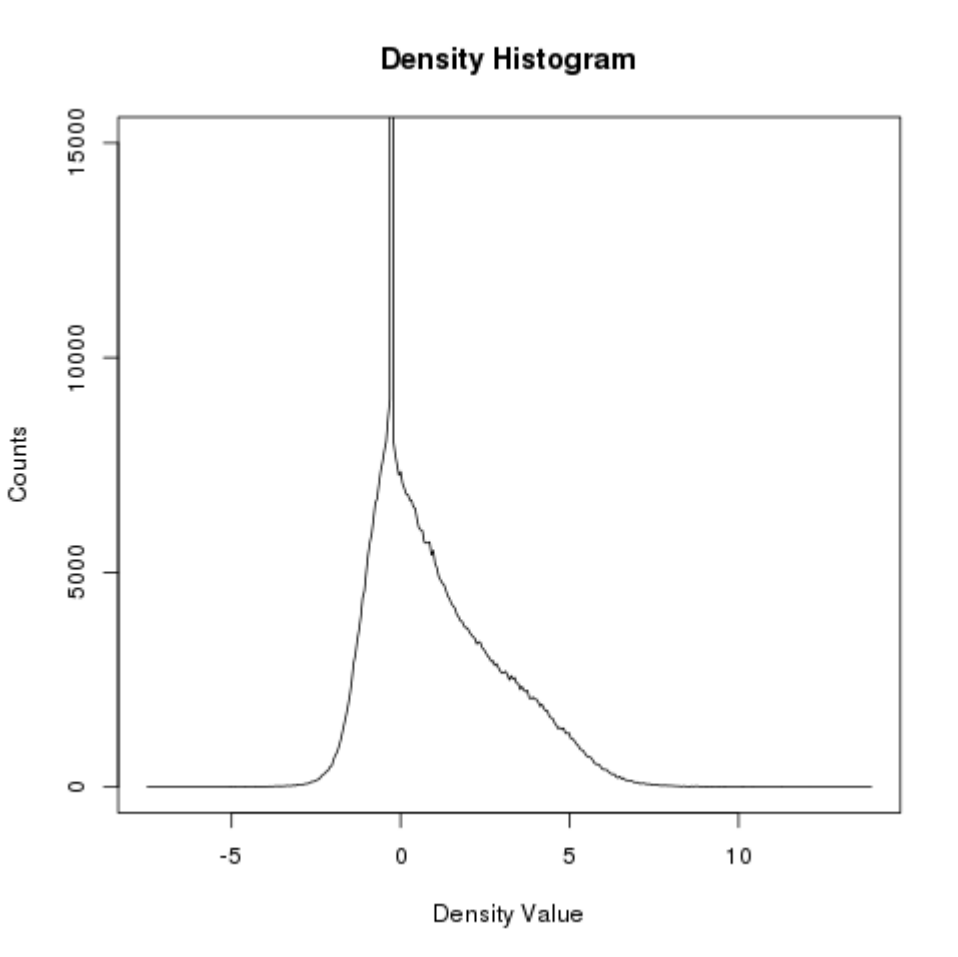
Typical Density Histogram from an X-ray map



# Helix Fitting

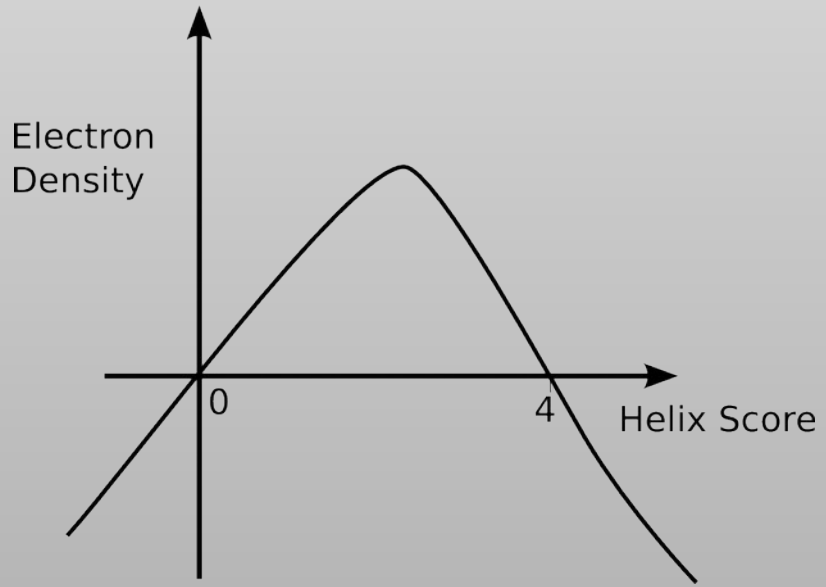
- The distribution of electron density is quite unlike that of x-ray maps



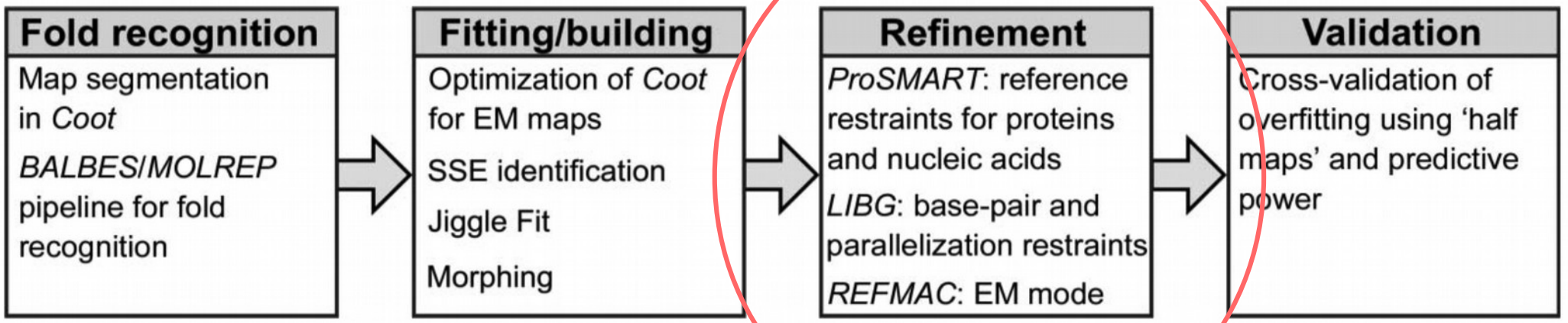
# Helix Fitting

- The distribution of electron density is quite unlike that of x-ray maps
  - e.g. You don't see main-chain atoms at 4 rmsd in x-ray maps
    - regions of dense electron density contribute negatively to helix score
  - These EM maps were sharpened and in a big box of mostly nothing
  - Lots to see at 4 rmsd



# Model-Building Tools

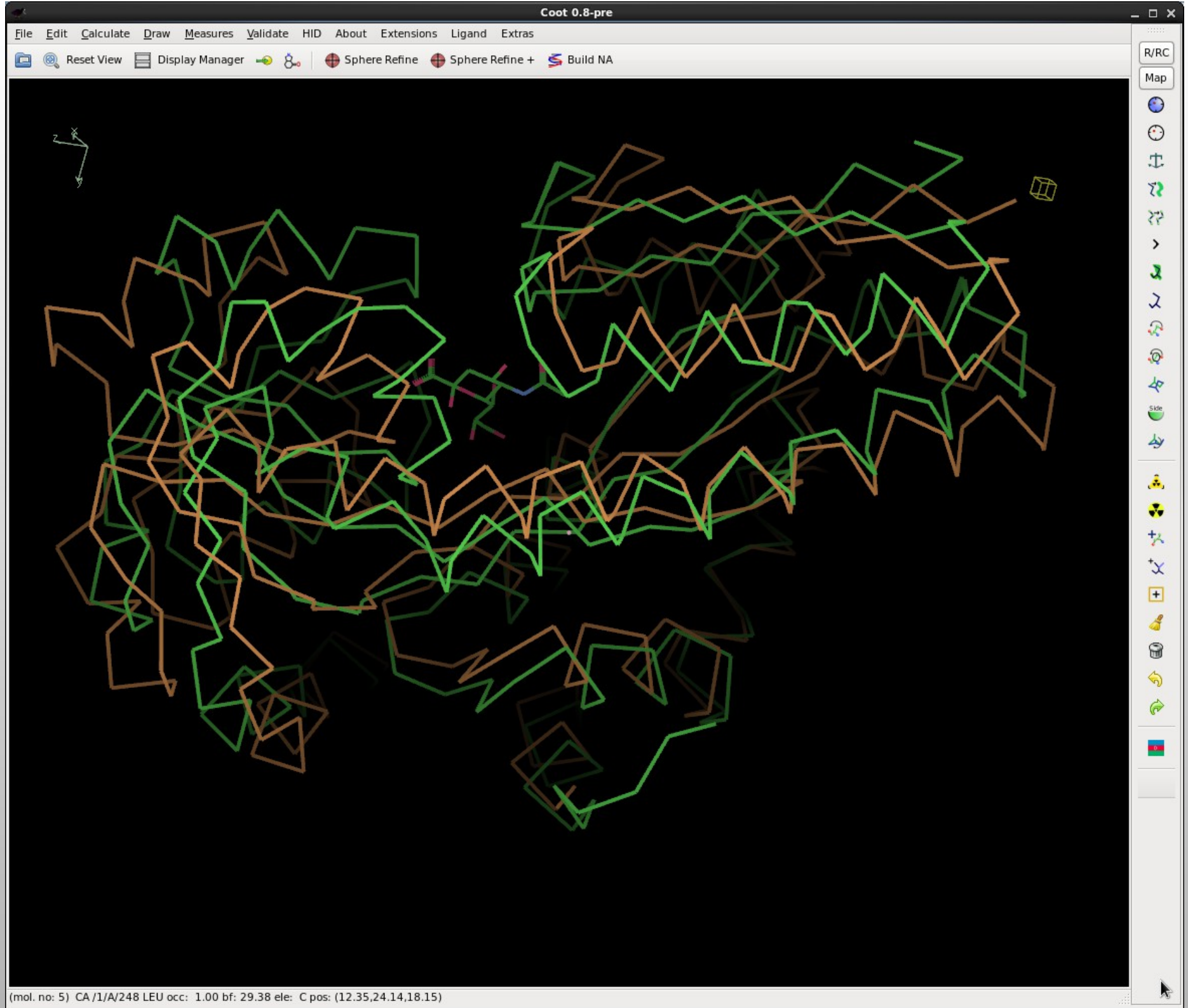
## Recent Developments

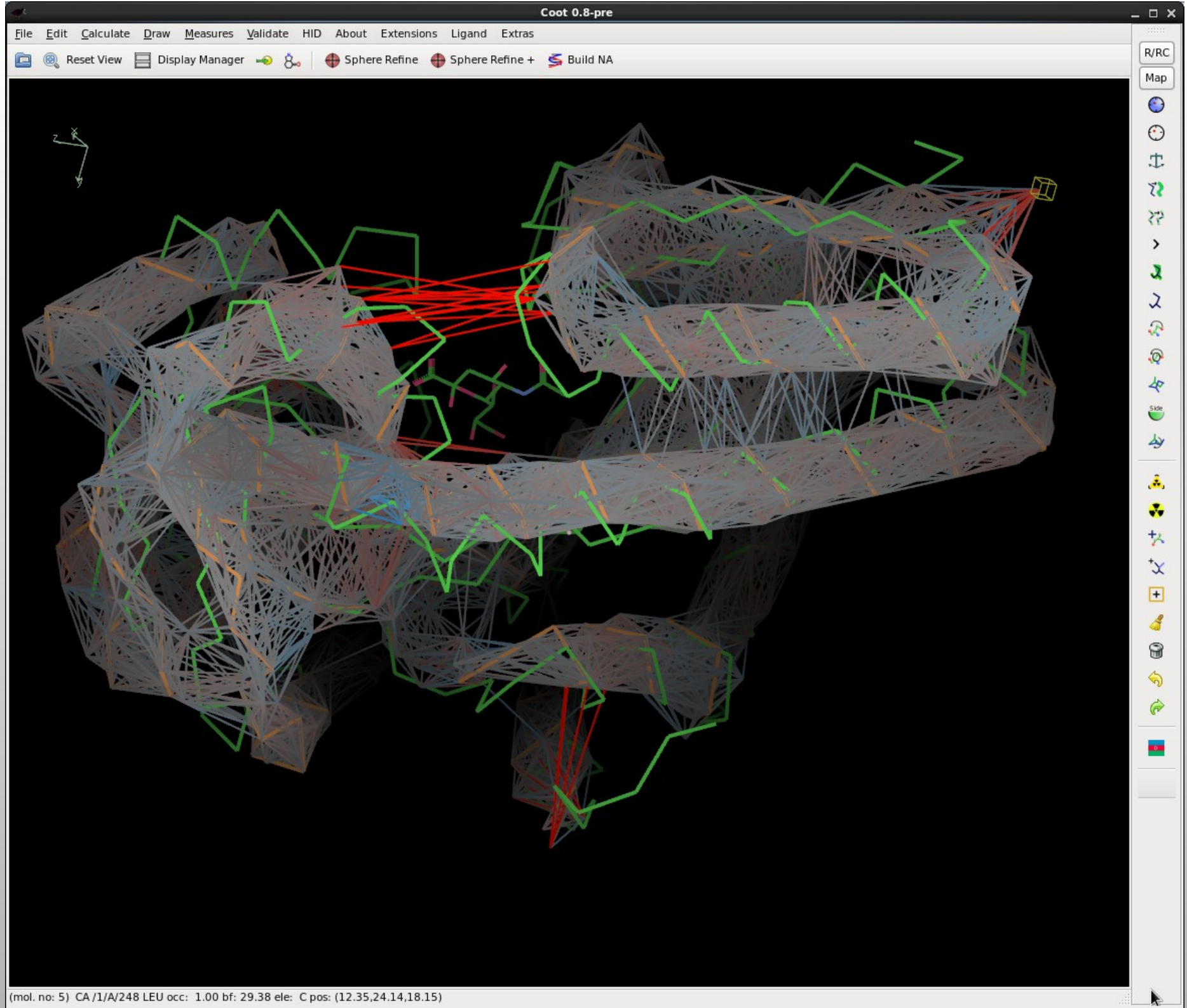


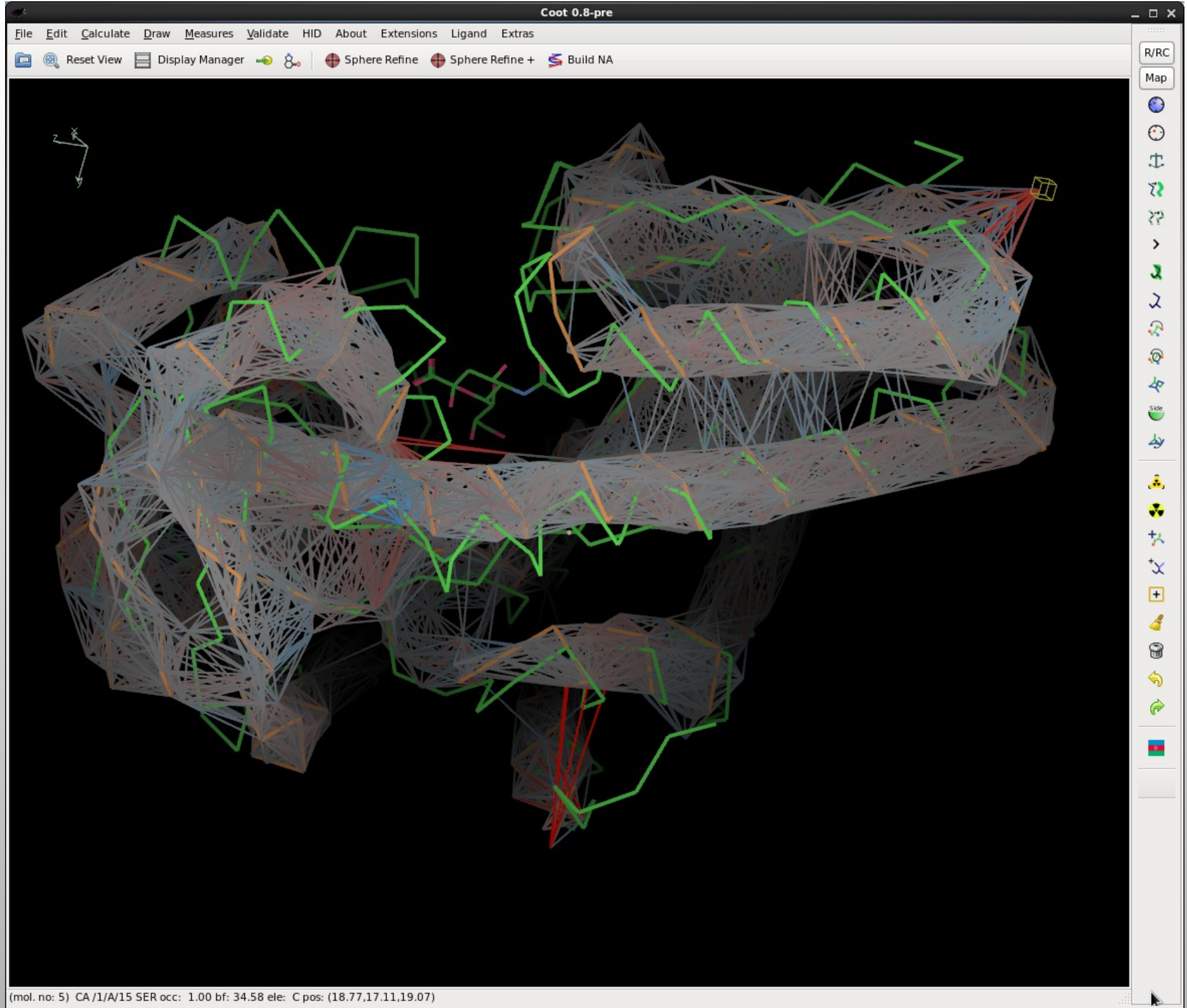
# ProSMART Interface

- 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
(mol. no: 5) CA /1/A/248 LEU occ: 1.00 bf: 29.38 ele: C pos: (12.35,24.14,18.15)

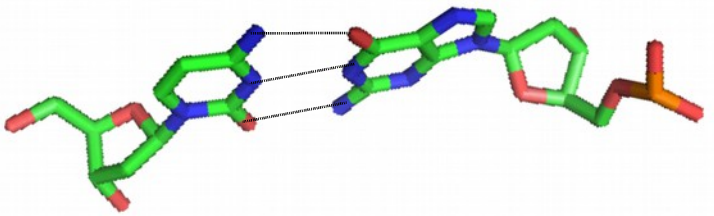

Coot 0.8-pre
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Extras
Reset View
Display Manager
Sphere Refine
Sphere Refine +
Build NA
R/RC
Map
(mol. no: 5) CA /1/A/248 LEU occ: 1.00 bf: 29.38 ele: C pos: (12.35,24.14,18.15)


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)

# LIBG Restraints for DNA/RNA

**LIBG** – for the generation of nucleic acid restraints

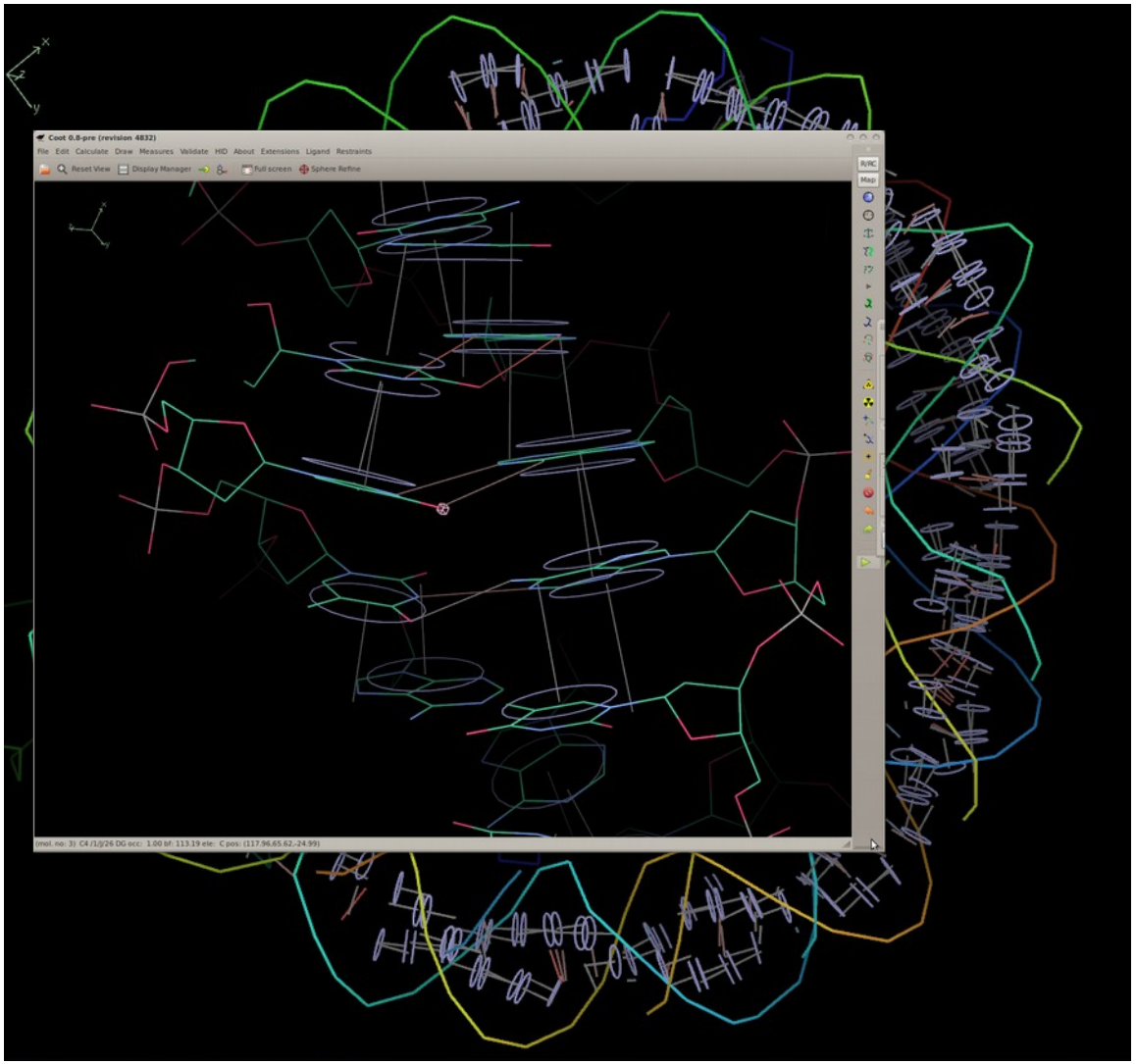
Base-pair restraints:

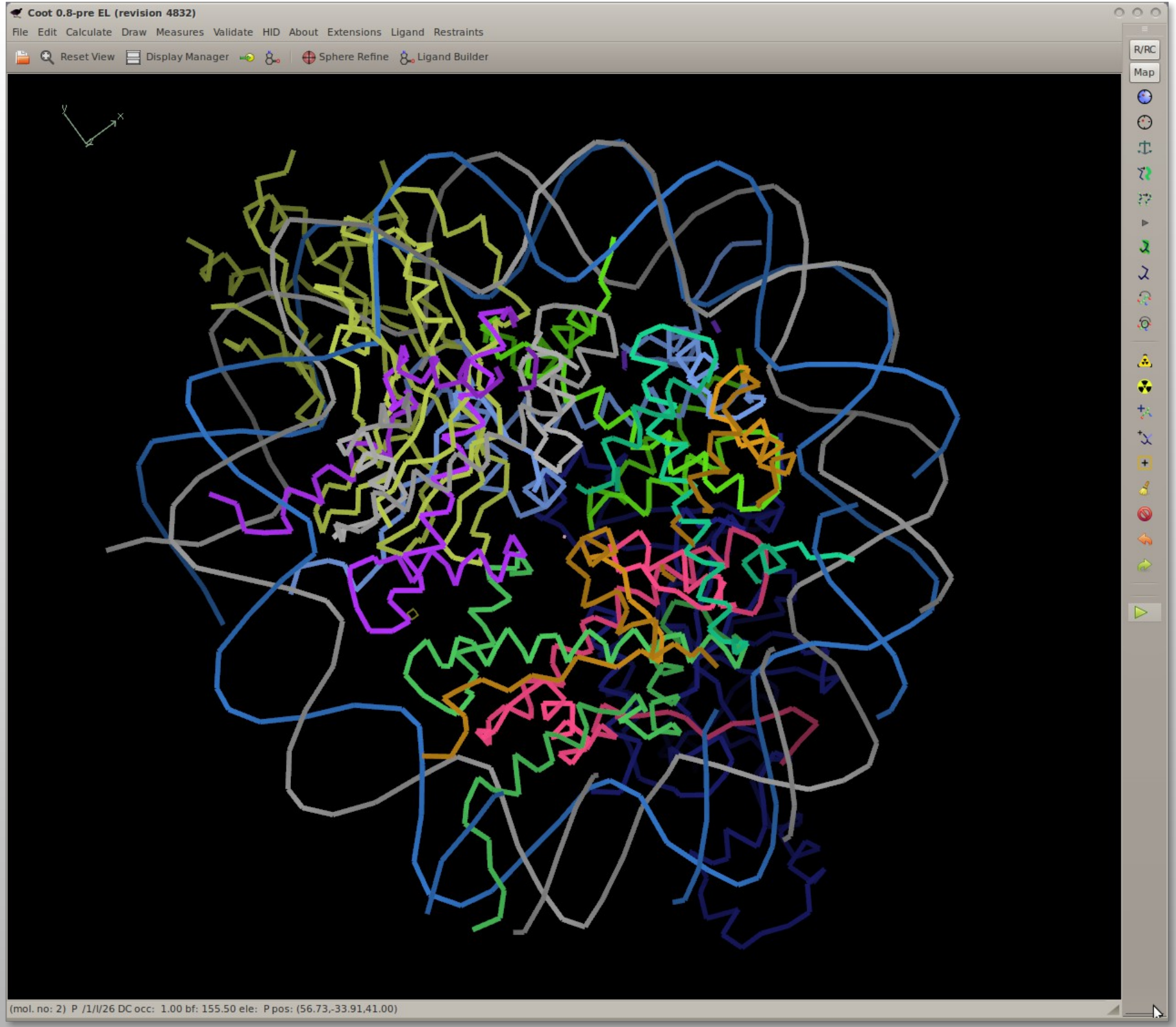


# LIBG Restraints for DNA/RNA

**LIBG** – for the generation of nucleic acid restraints

Base-stacking restraints:
(parallel plane 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)

# Purpose of Refinement

**Crystallographic refinement has two purposes:**

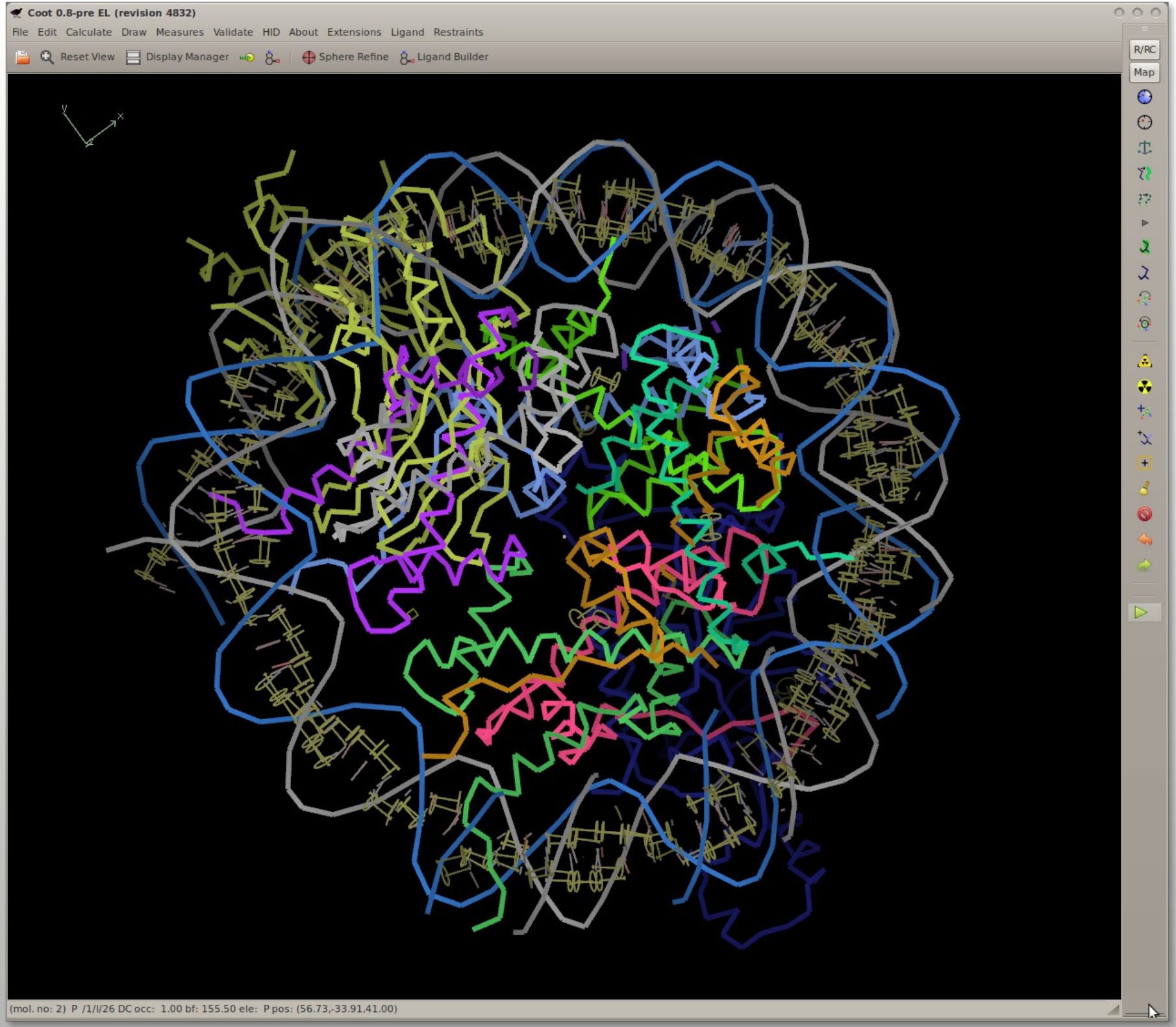
1. Fit atomic model into observed X-ray crystallographic data

   *Model should agree with the observed data*

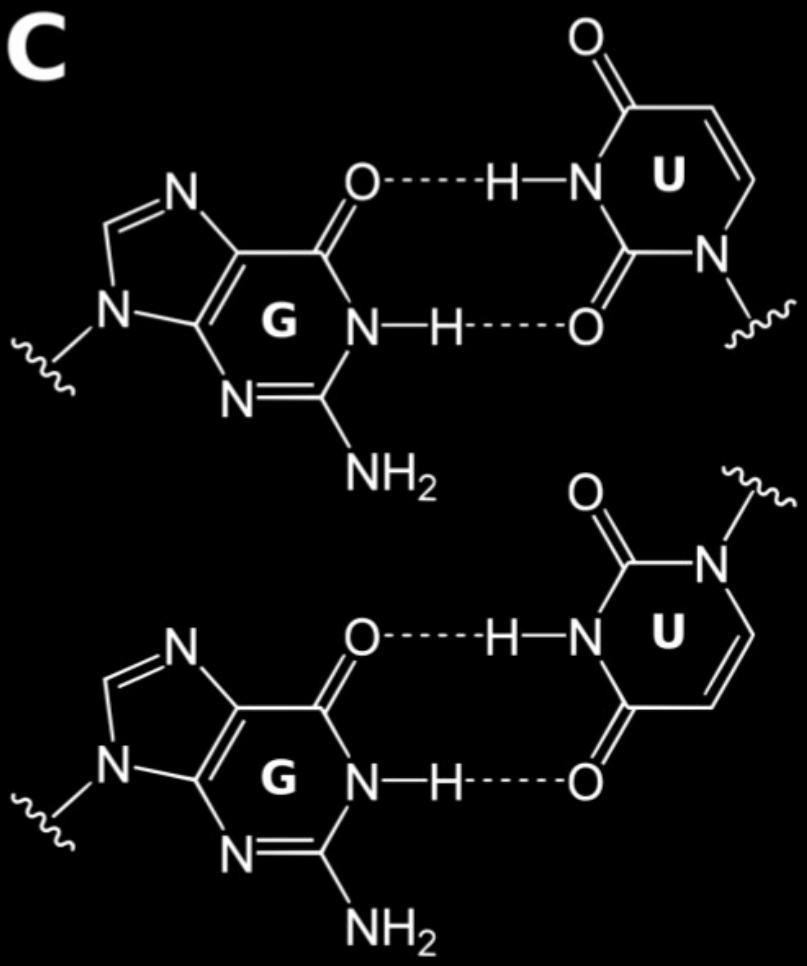
   *Model must be chemically and structurally sensible*

2. Calculate best possible electron density map

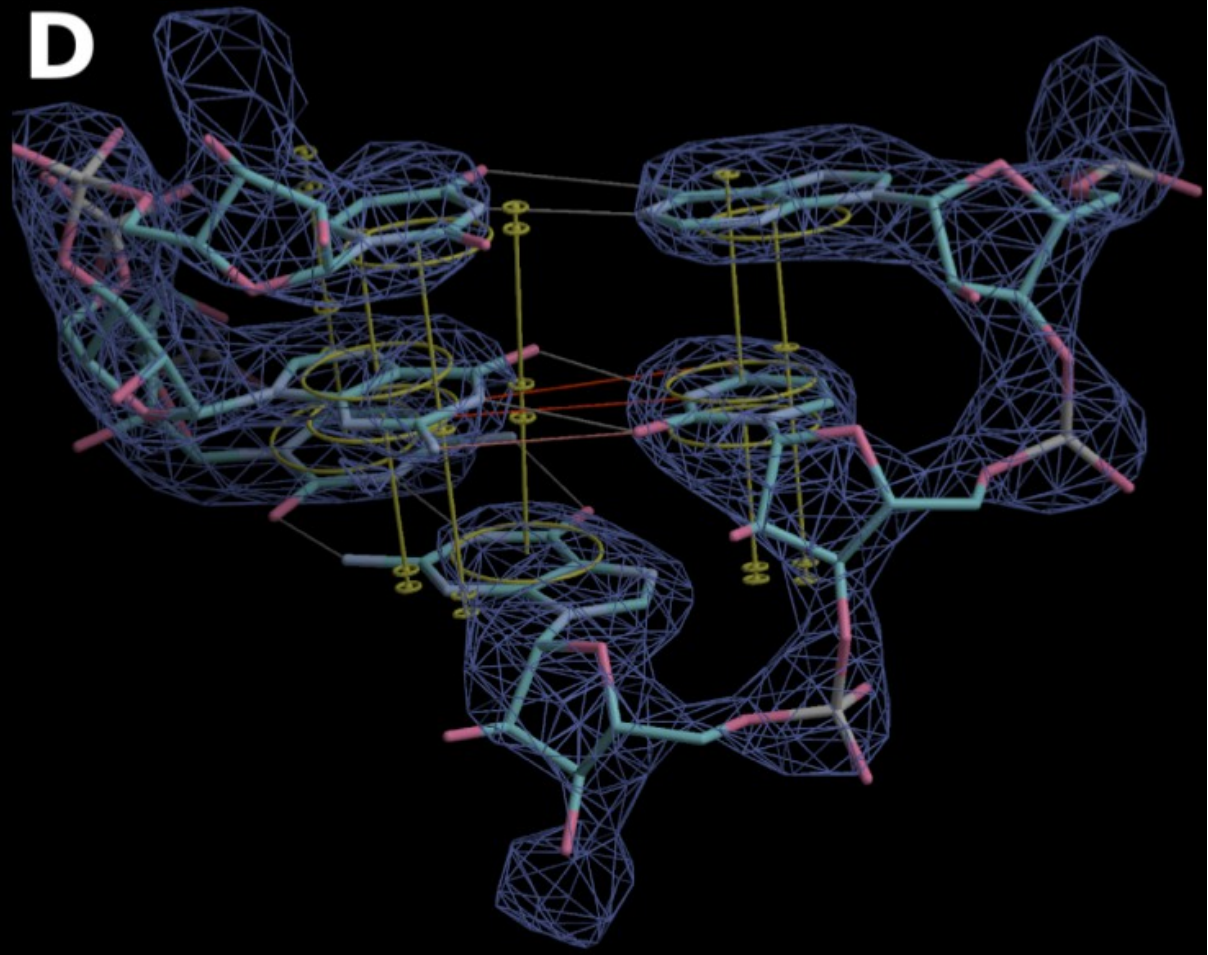
   *Allowing the atom model to be visualised, criticised and analysed*


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)

# Libg restraints




(Watson Crick and)
Wobble, Reverse Wobble




Representation in Coot

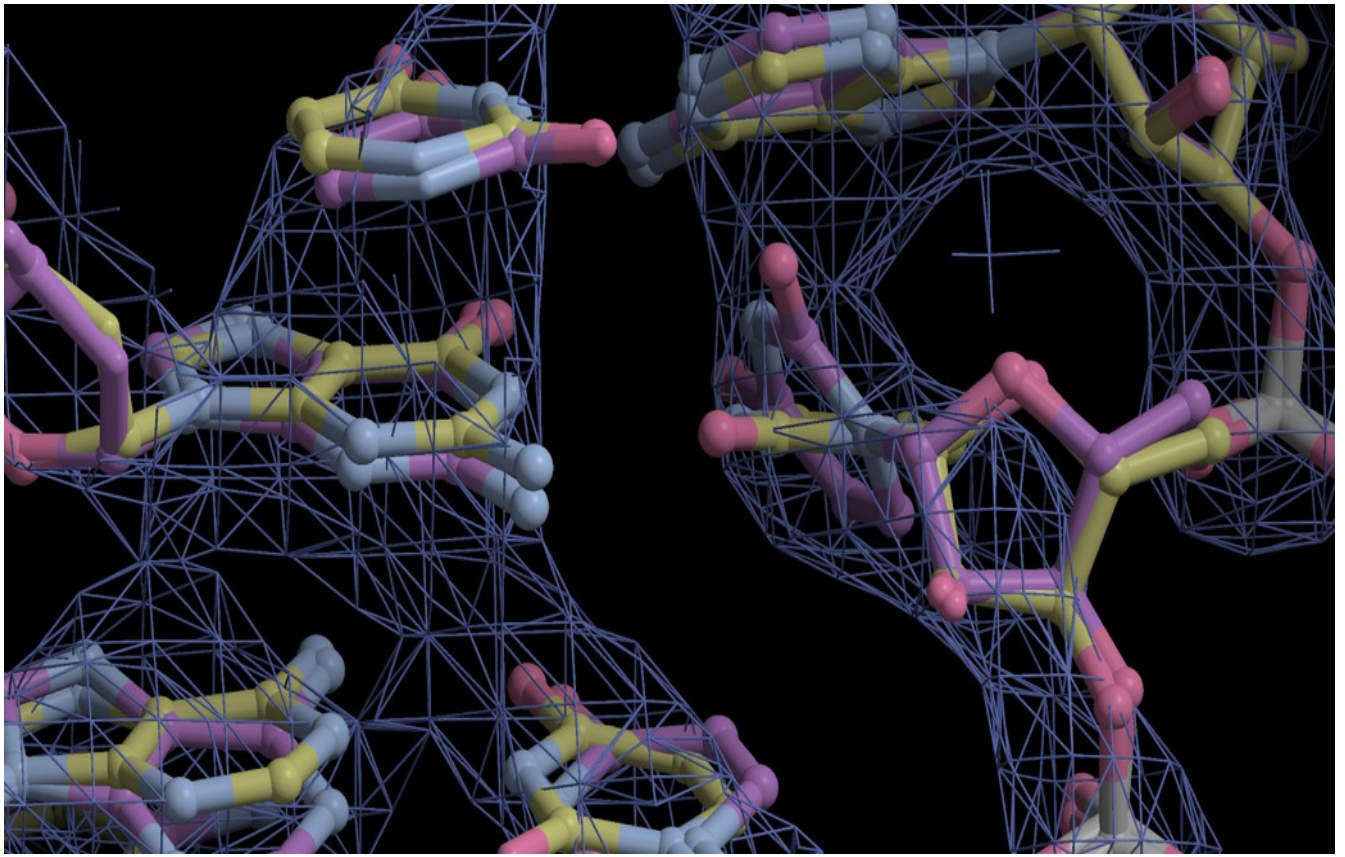
# LIBG Restraints for DNA/RNA

**LIBG** – for the generation of nucleic acid restraints

Base-stacking restraints:
(parallel plane restraints)

Purple – refined without LIBG
Yellow – refined with LIBG

Example:
3.2Å cryo-EM



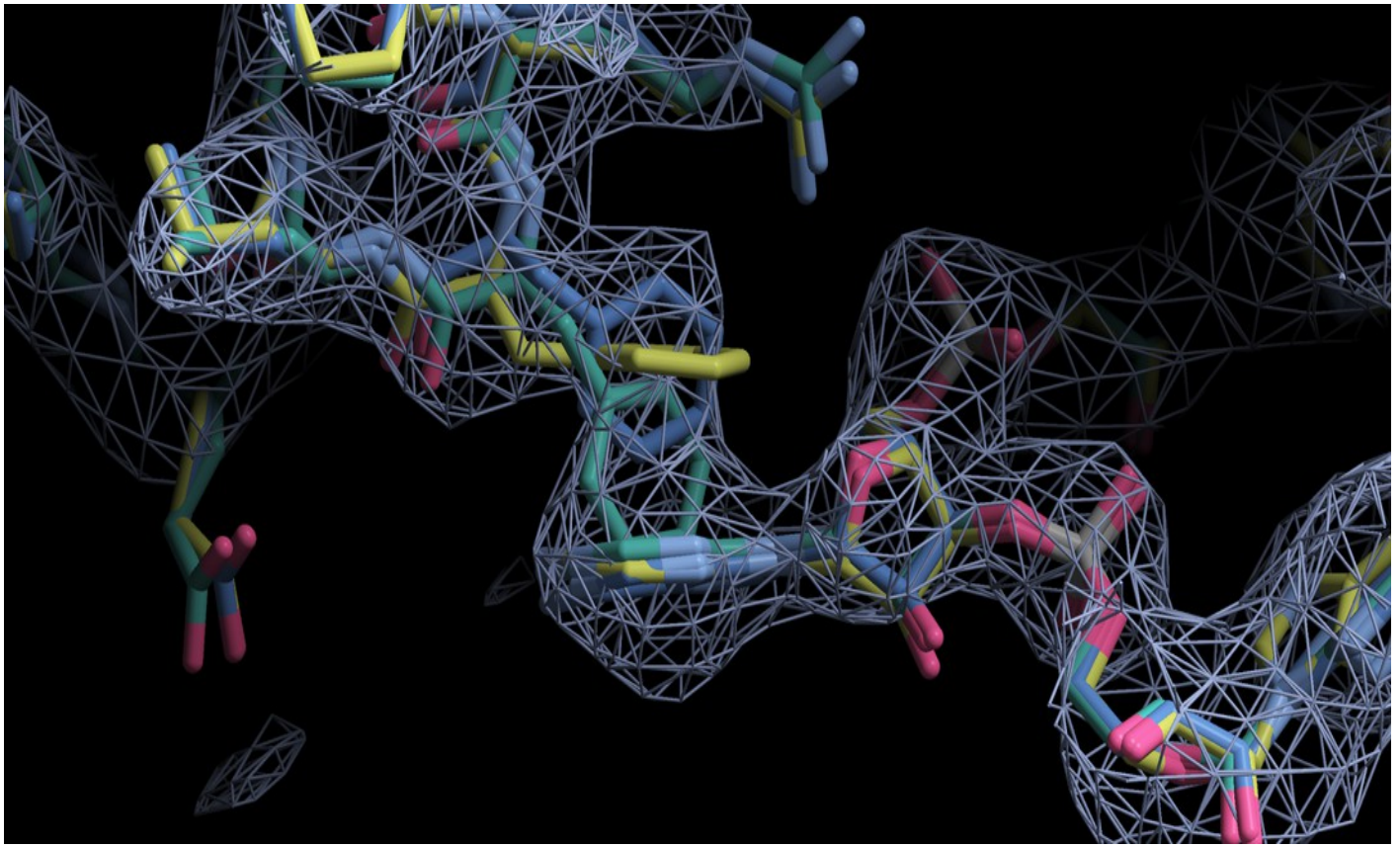
# LIBG Restraints for DNA/RNA

**LIBG** – for the generation of nucleic acid restraints

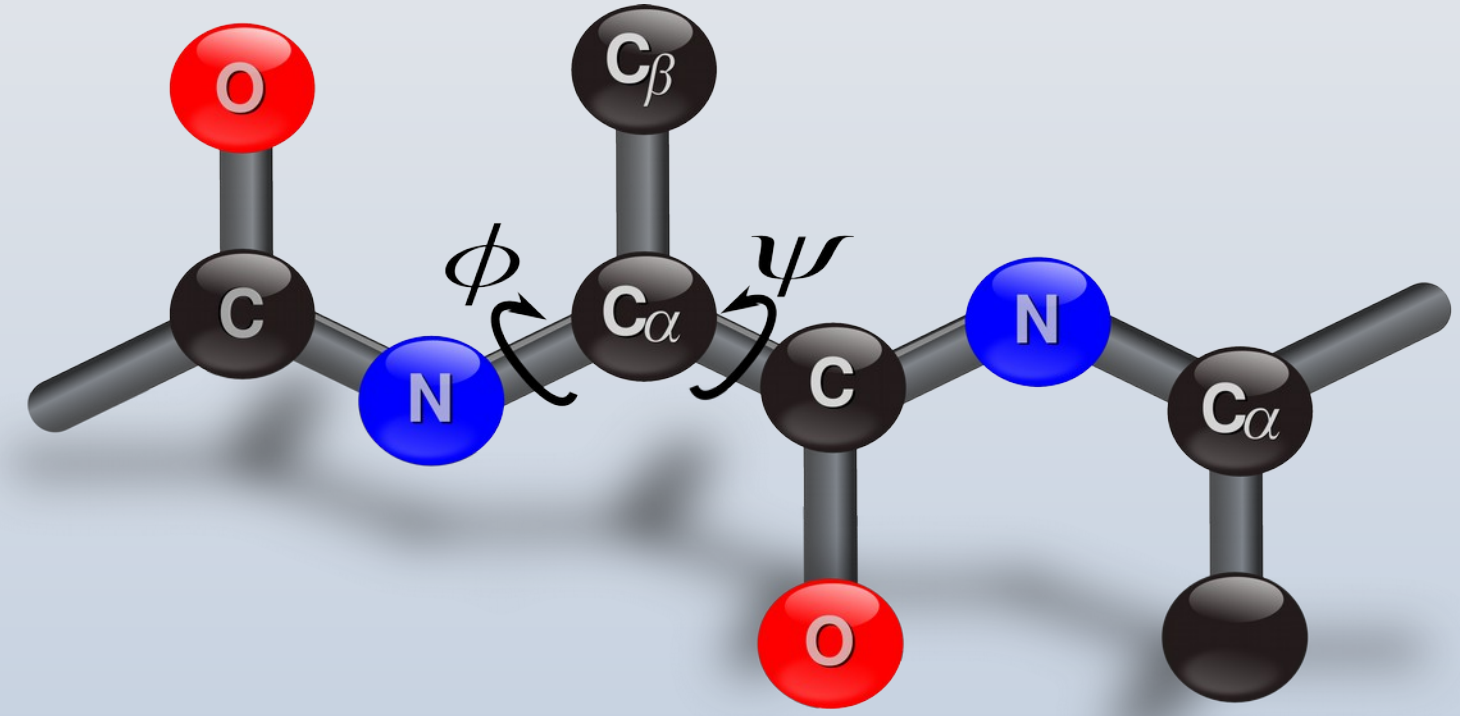
Base-stacking restraints:
(parallel plane restraints)

Green – before refinement
Blue – refined without LIBG
Yellow – refined with LIBG

Example:
3.2Å cryo-EM



# Peptide Backbone Geometry



# Low Resolution Model-Building

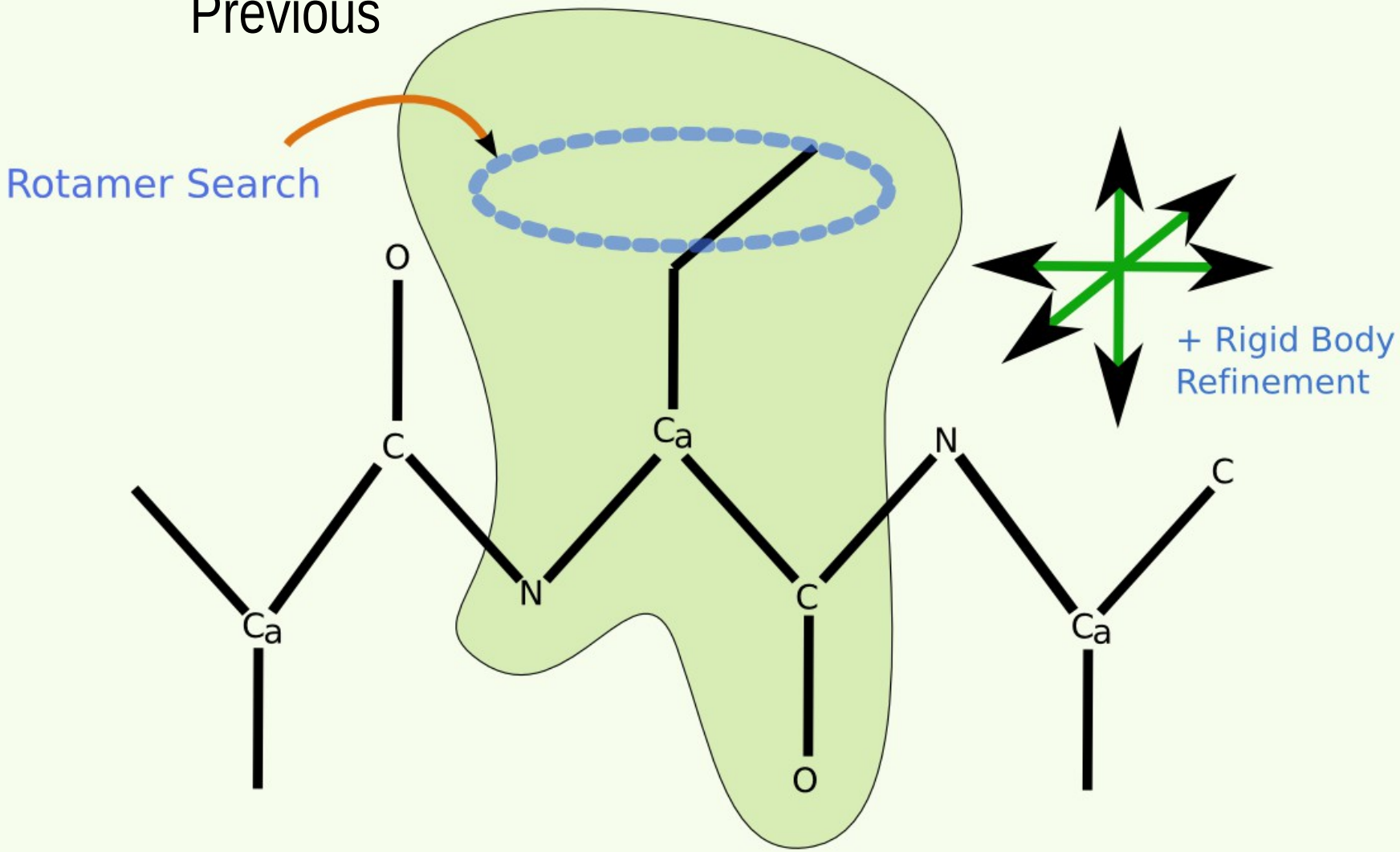
- “Backrub” rotamers

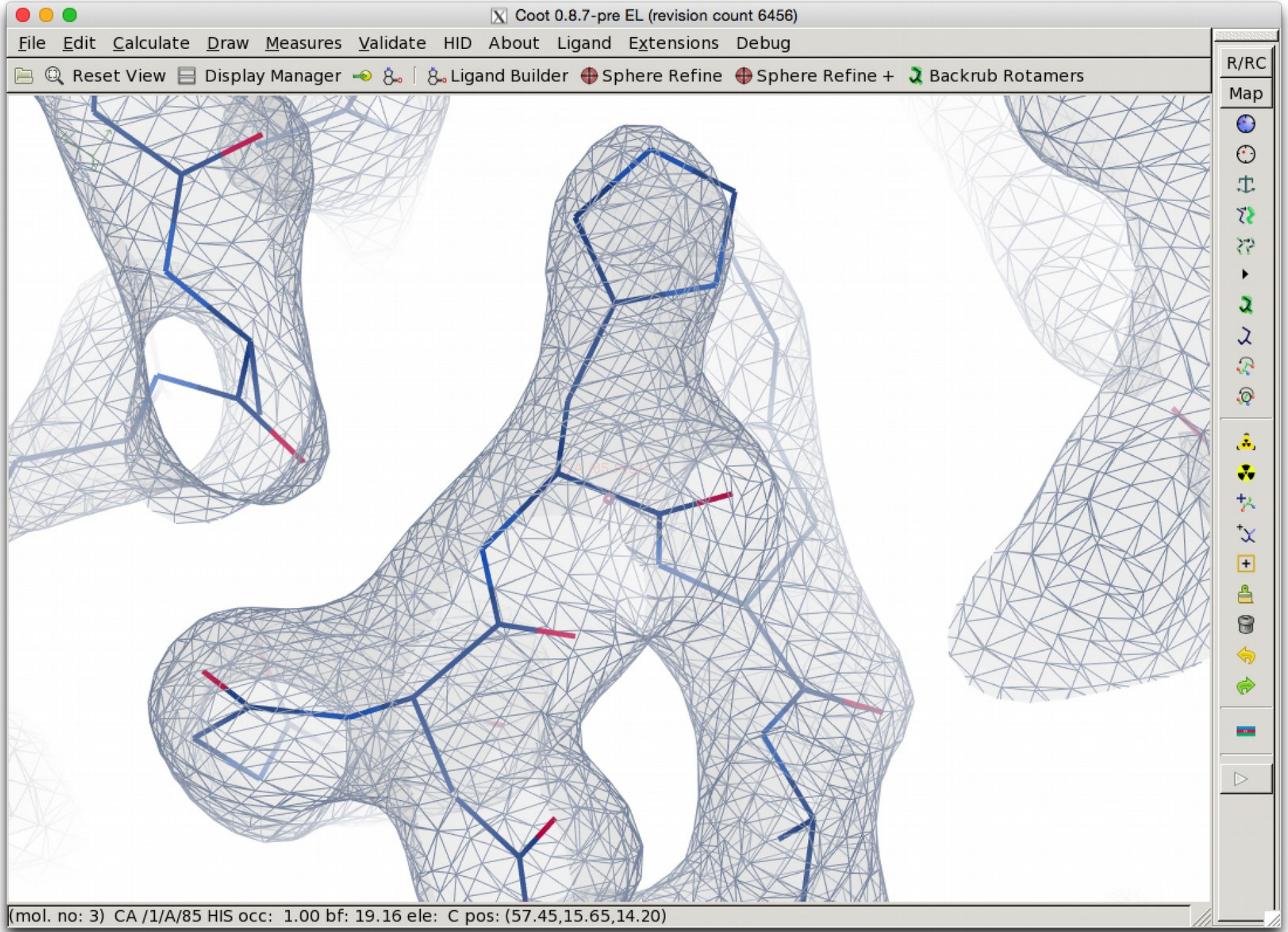
# Rotamer Searching

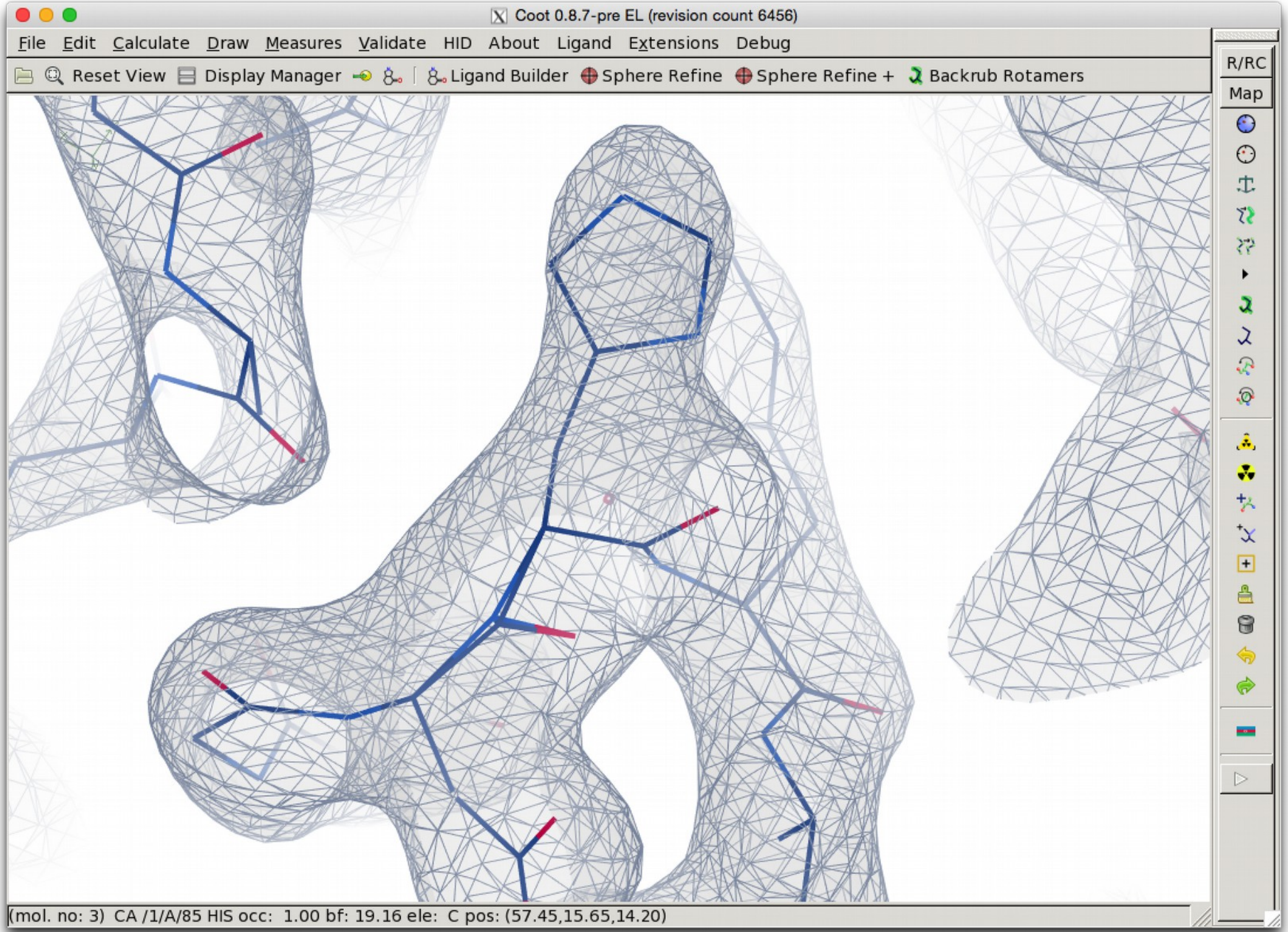
- Two methods
  - Traditional
  - Backrub

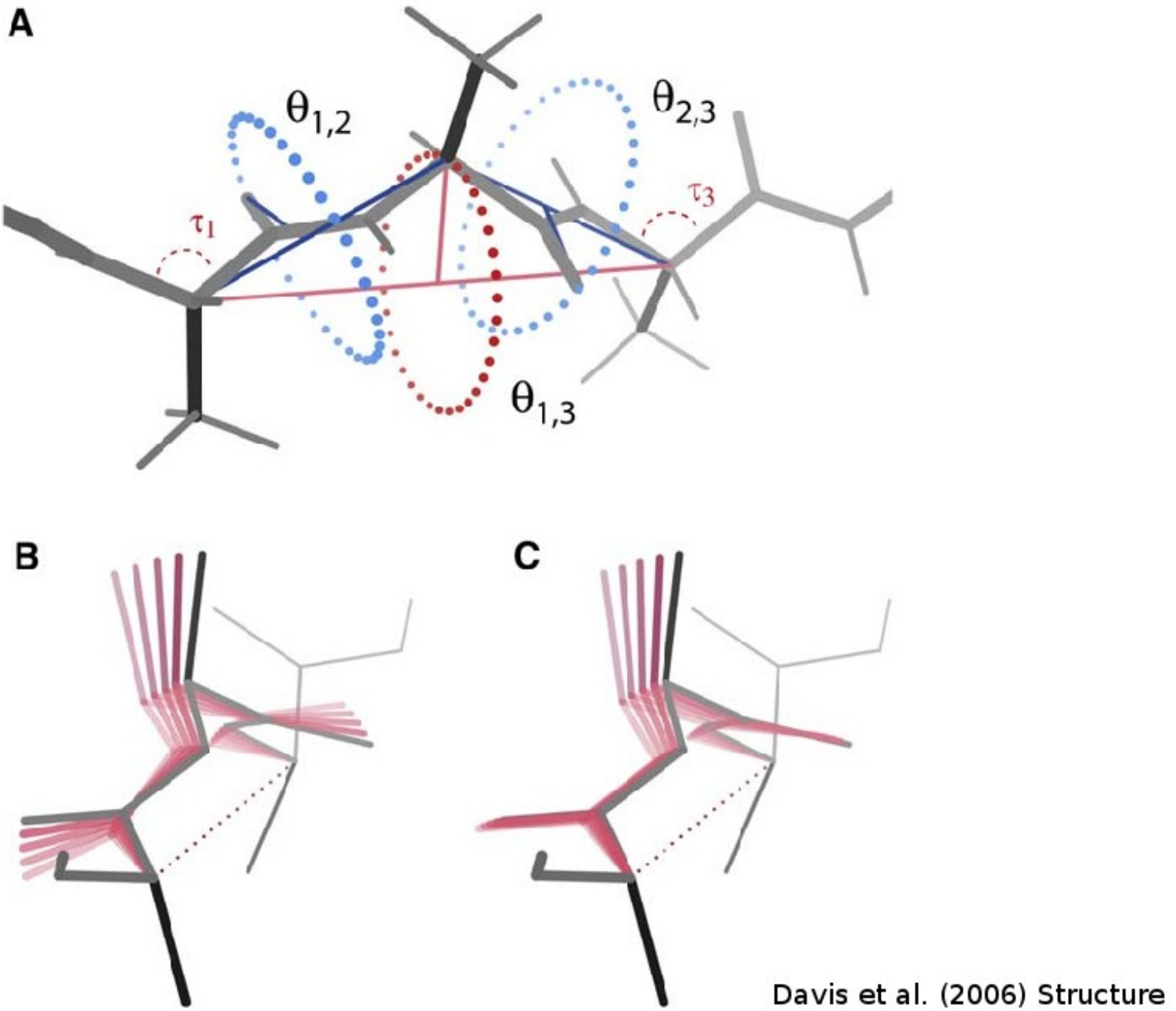
~~Current~~ Low Resolution Rotamer Search

Previous



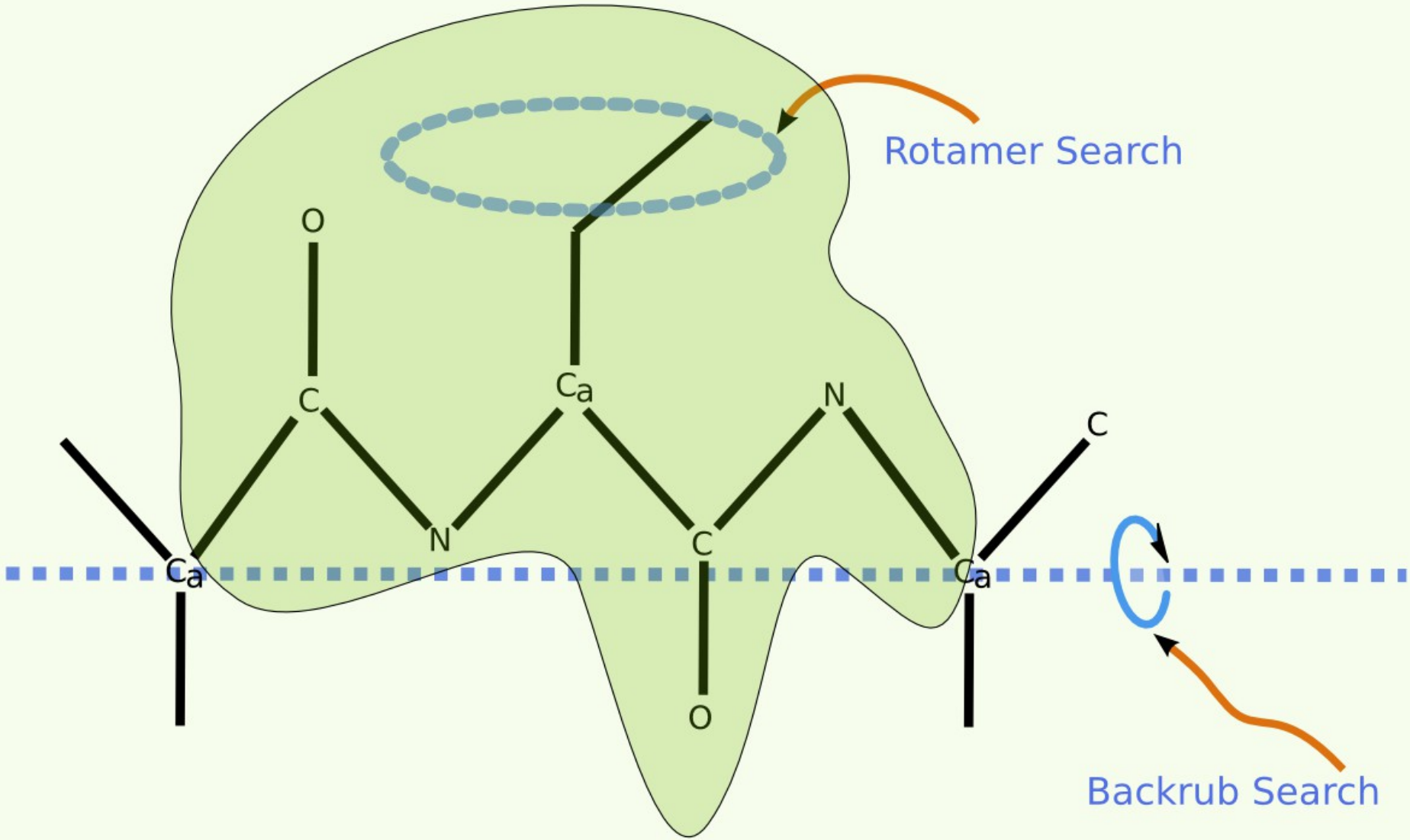

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


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

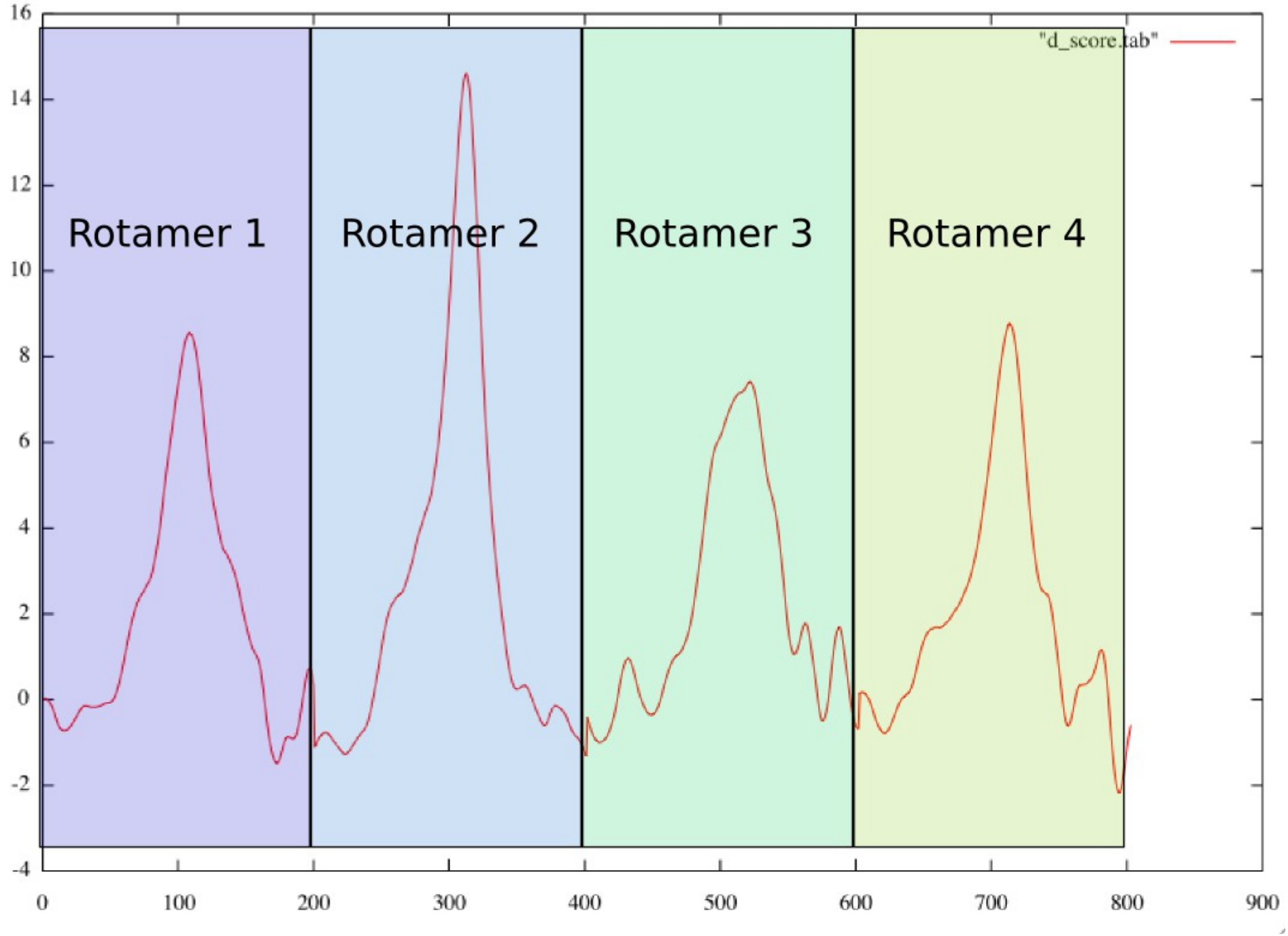


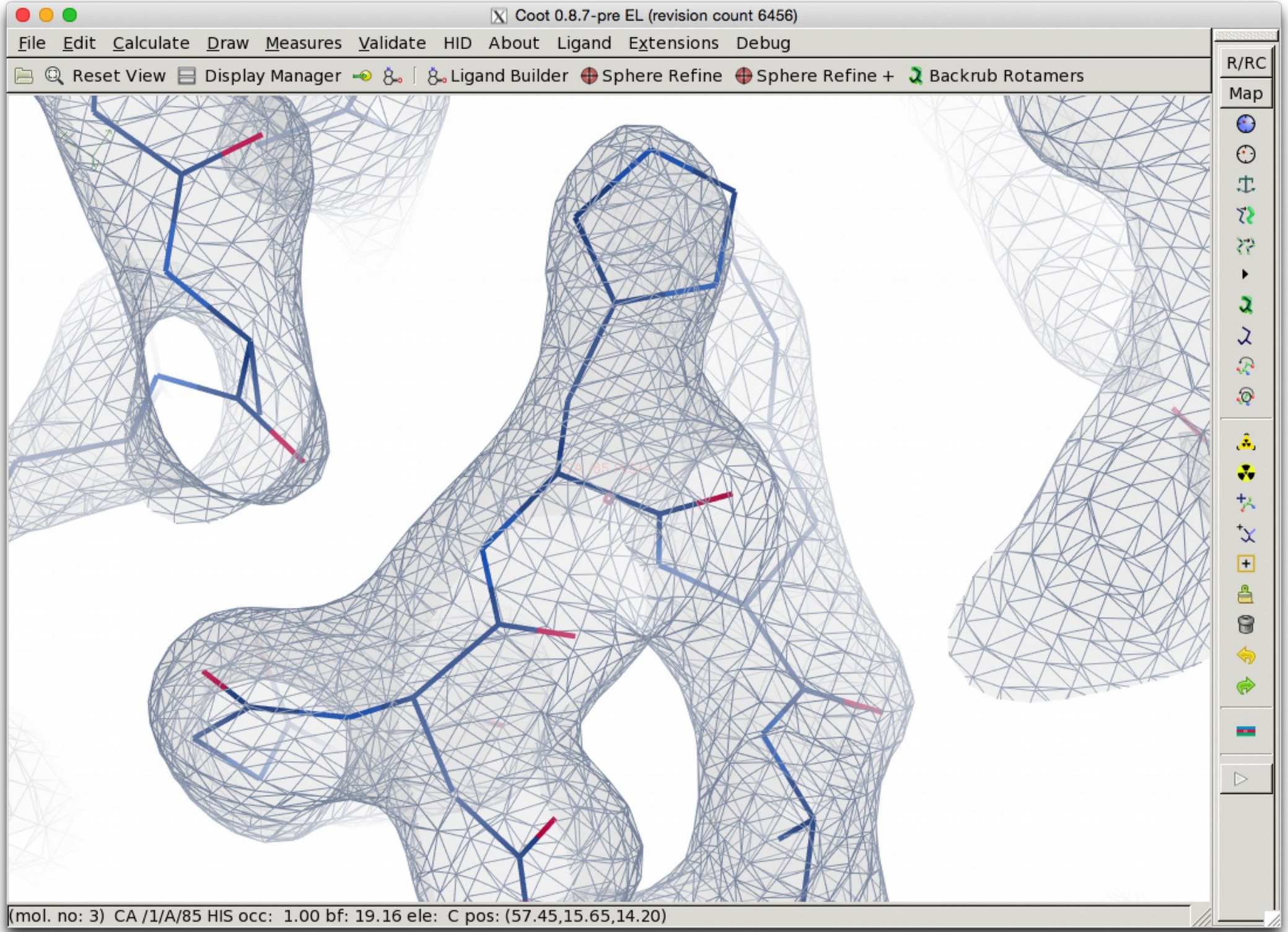

Davis et al. (2006) Structure

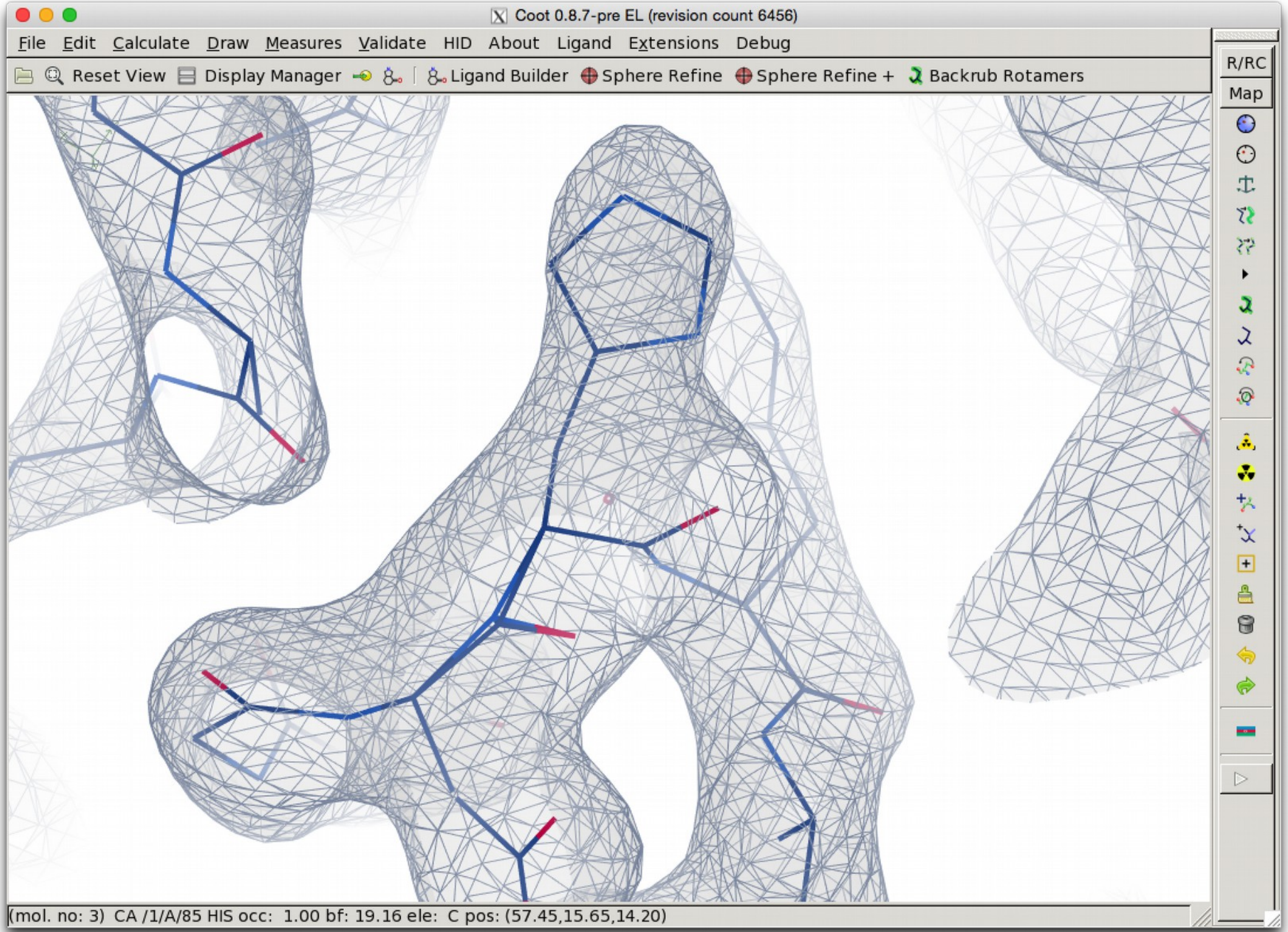
# New Low Resolution Rotamer Search

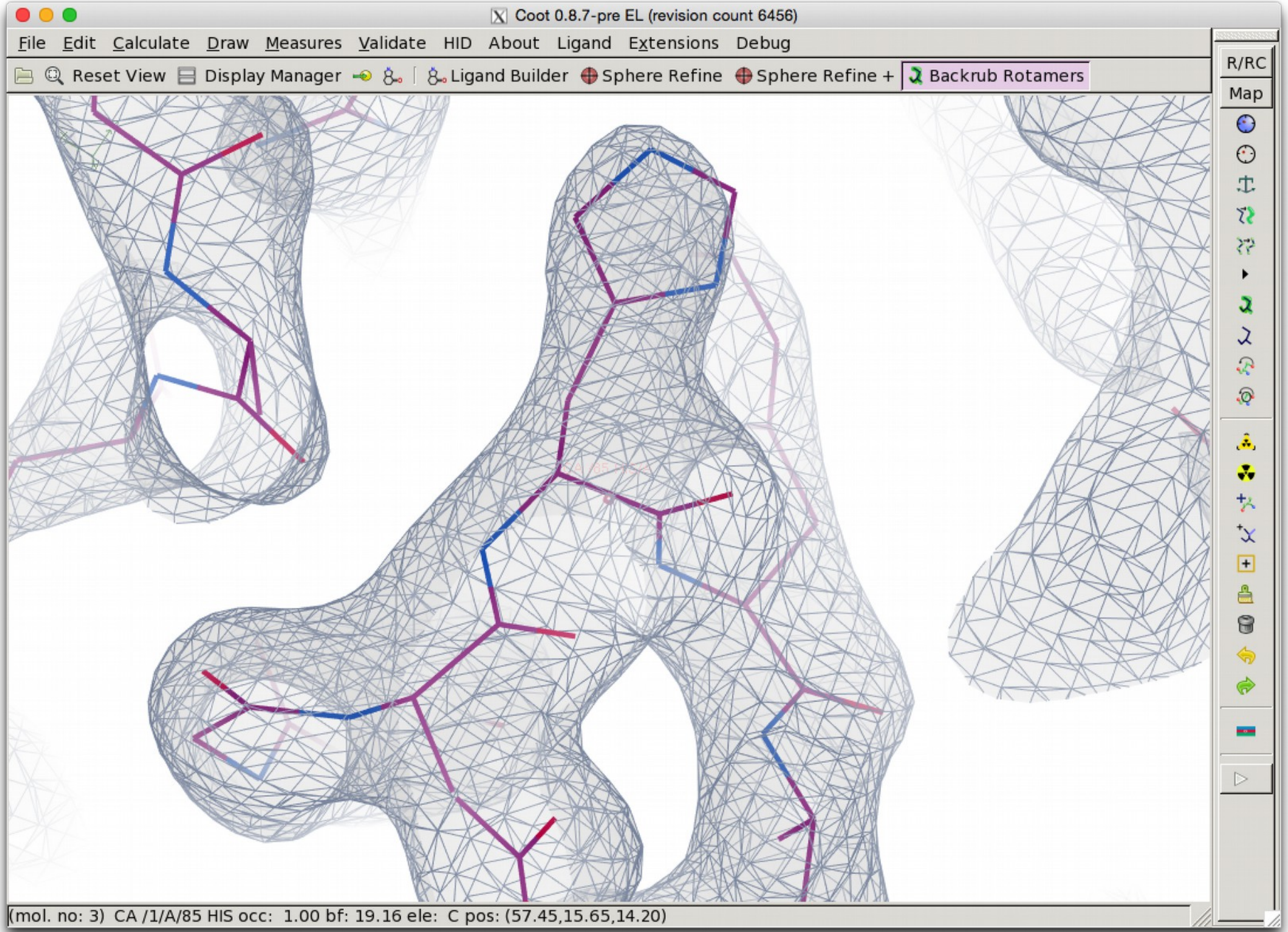


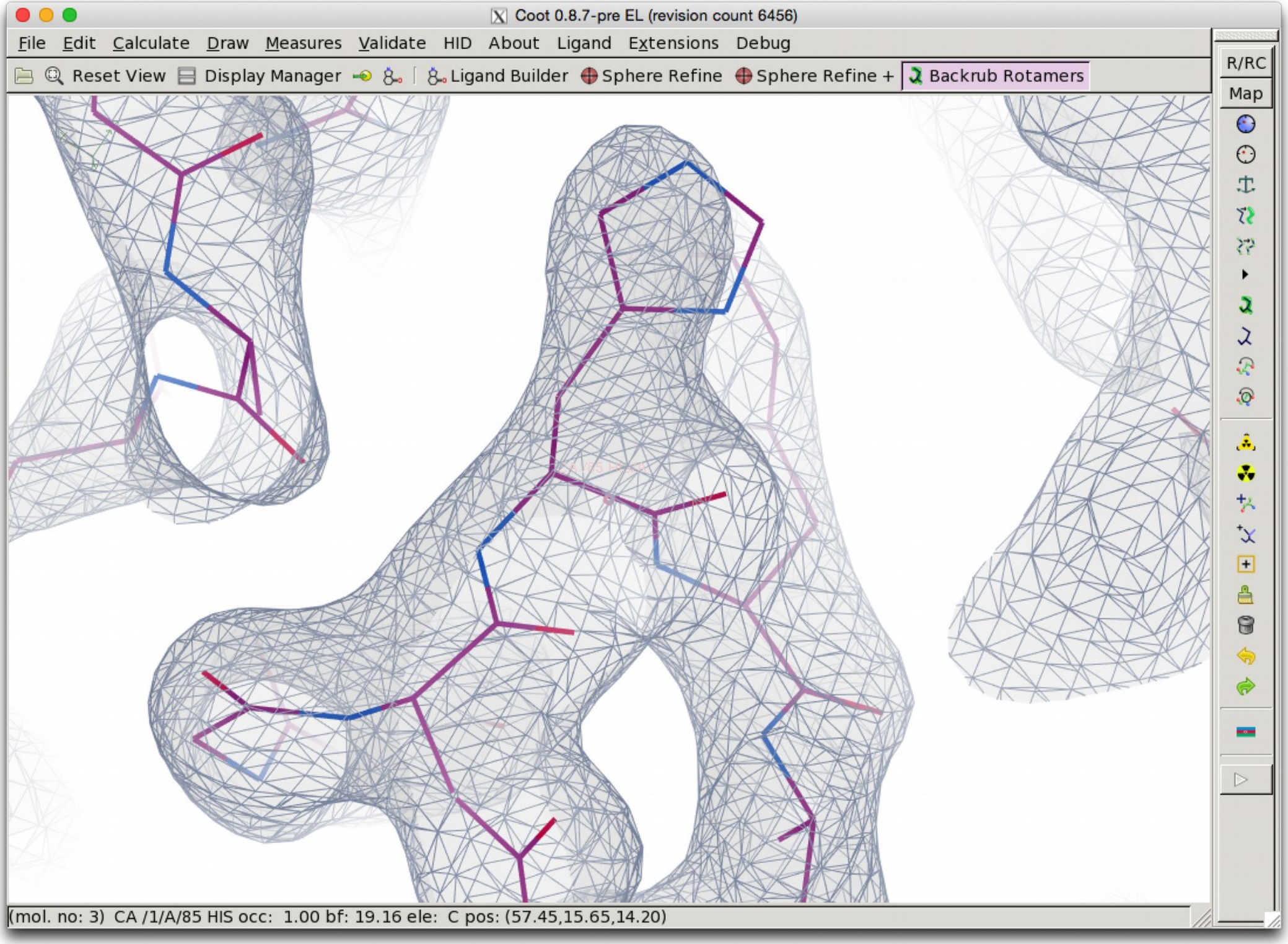

After Fitting Tools in KING/Molprobity

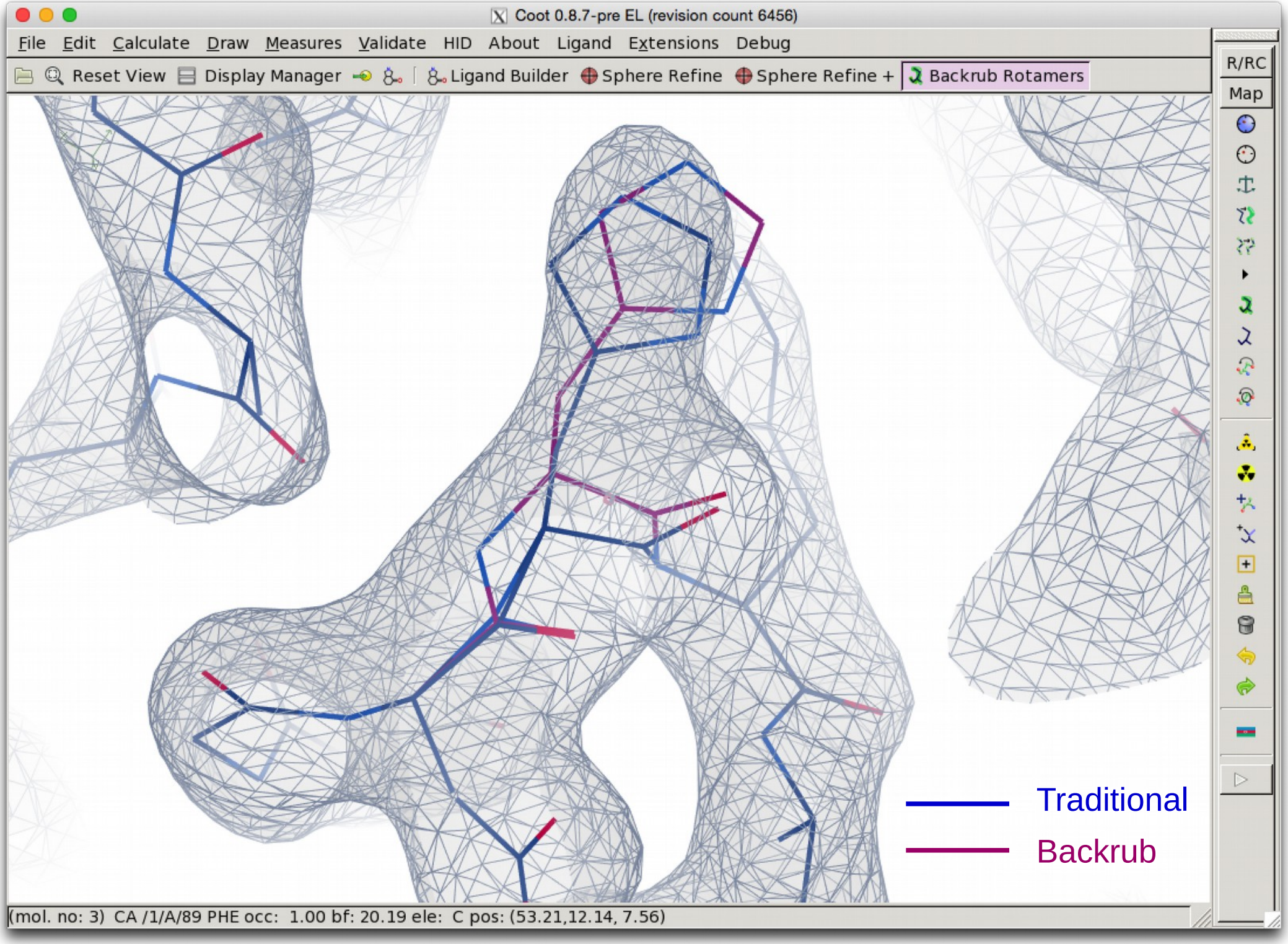

"d_score.tab"
Rotamer 1
Rotamer 2
Rotamer 3
Rotamer 4
16
14
12
10
8
6
4
2
0
-2
-4
0
100
200
300
400
500
600
700
800
900

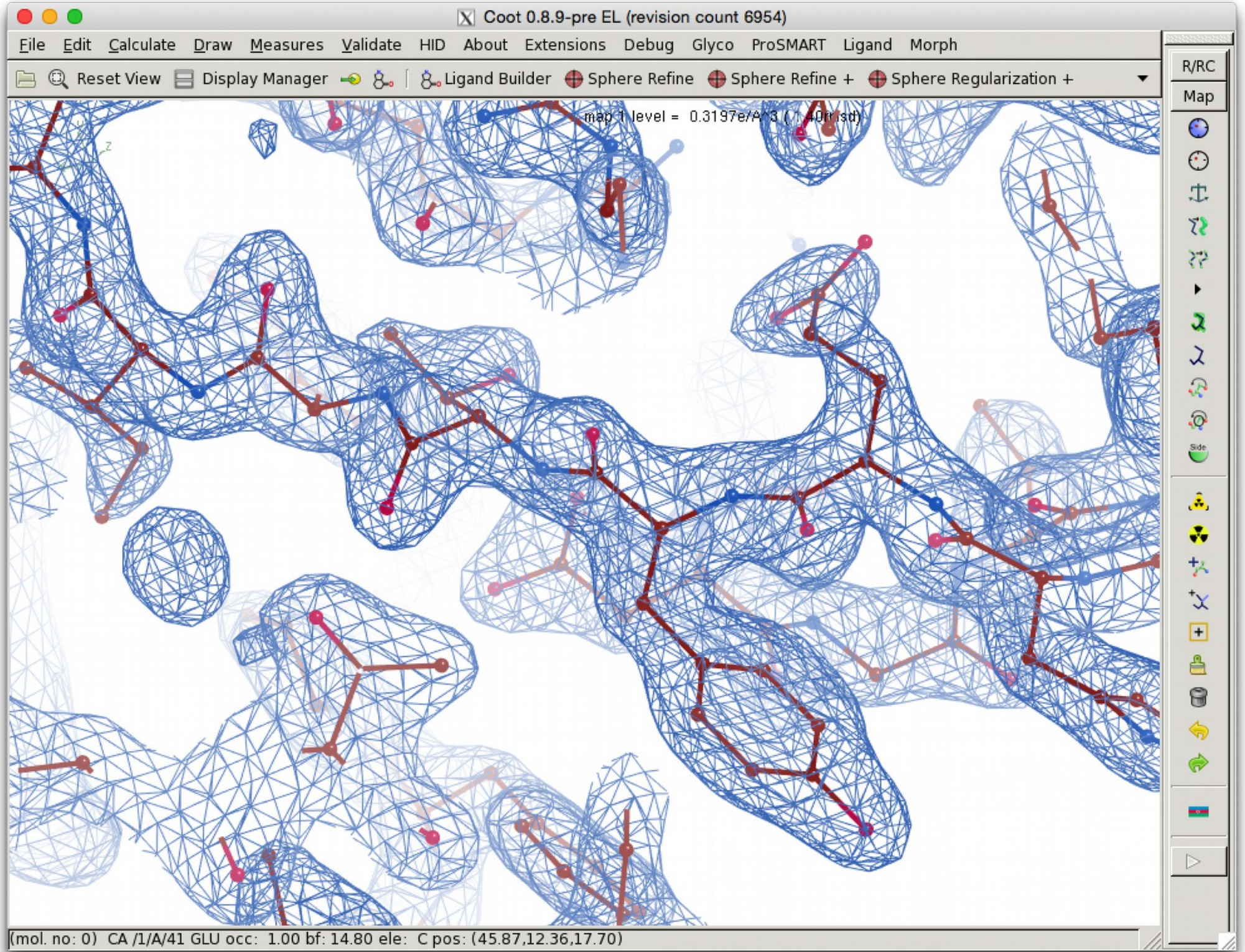

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)

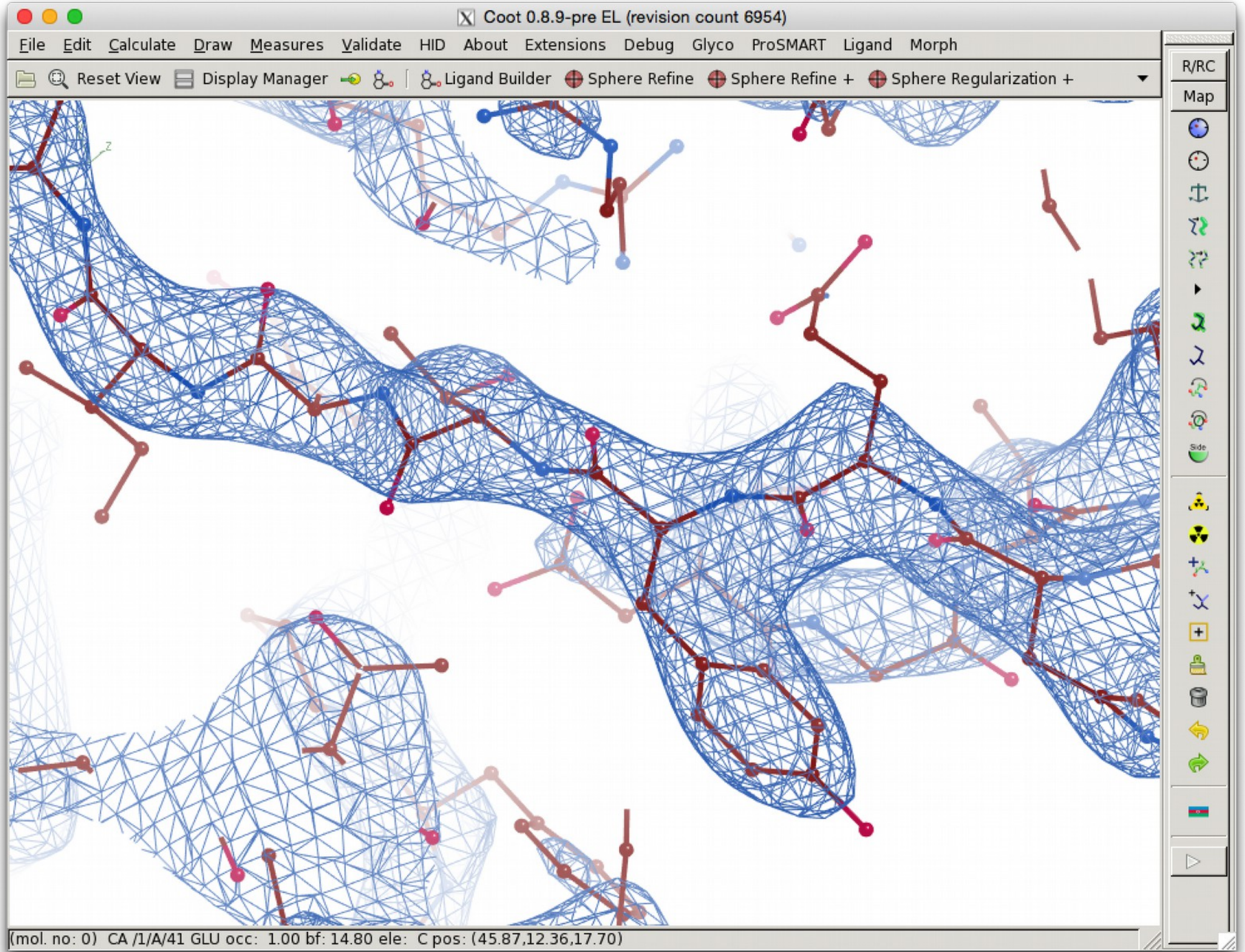

Coot 0.8.7-pre EL (revision count 6456)
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Backrub Rotamers
R/RC
Map
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)


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


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


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

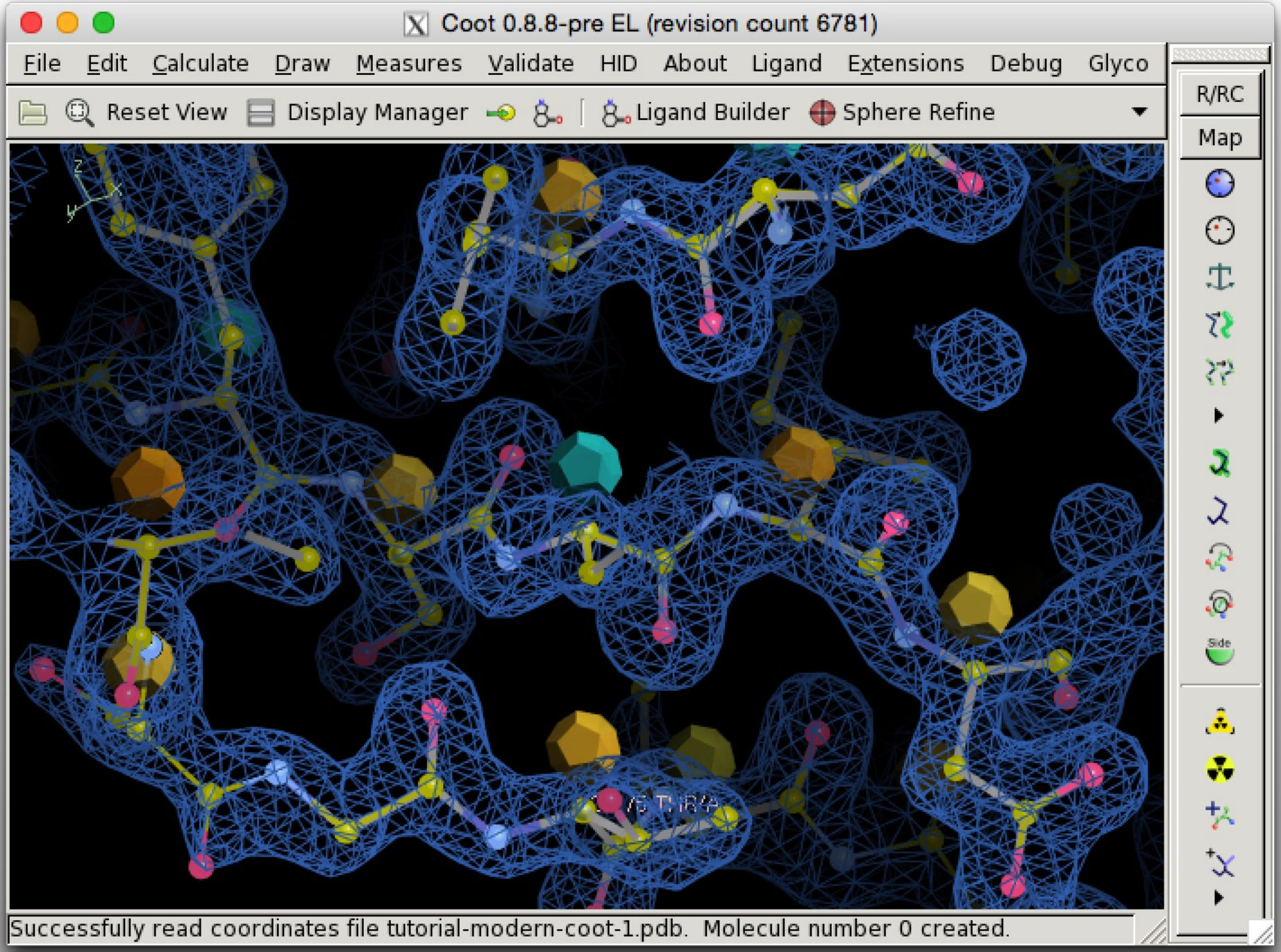

Coot 0.8.9-pre EL (revision count 6954)
File Edit Calculate Draw Measures Validate HID About Extensions Debug Glyco ProSMART Ligand Morph
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Sphere Regularization +
R/RC
Map
map 1 level = 0.3197e/A^3 ( 1.40rms)
(mol. no: 0) CA /1/A/41 GLU occ: 1.00 bf: 14.80 ele: C pos: (45.87,12.36,17.70)


Coot 0.8.9-pre EL (revision count 6954)
File Edit Calculate Draw Measures Validate HID About Extensions Debug Glyco ProSMART Ligand Morph
Reset View
Display Manager
Ligand Builder
Sphere Refine
Sphere Refine +
Sphere Regularization +
R/RC
Map
Side
(mol. no: 0) CA /1/A/41 GLU occ: 1.00 bf: 14.80 ele: C pos: (45.87,12.36,17.70)

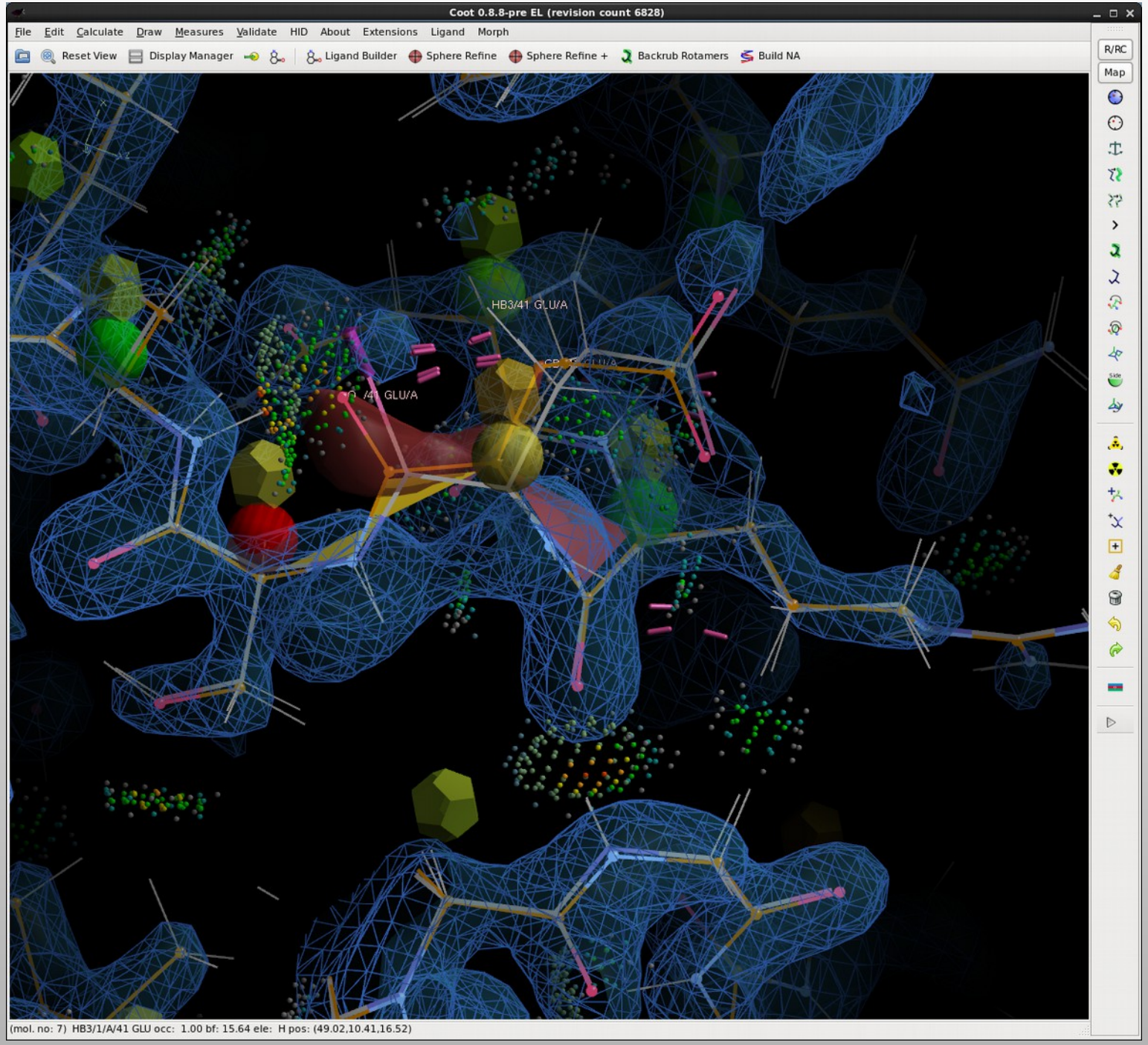
# Cryo-EM Model-building

- autozone multi-residue
- sphere refine, sphere refine+
- sphere regularize, sphere regularize +
- Geman-McClure distance restraints
- Multi-threading/parallel processing
- backrub rotamers
- interactive ramachandran, rotamer and clash markup

# Interactive Rotamer Goodness



# Multi-Criteria Markup



# Acknowledgements


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

# Aesop Libs Interface (Martin Noble)