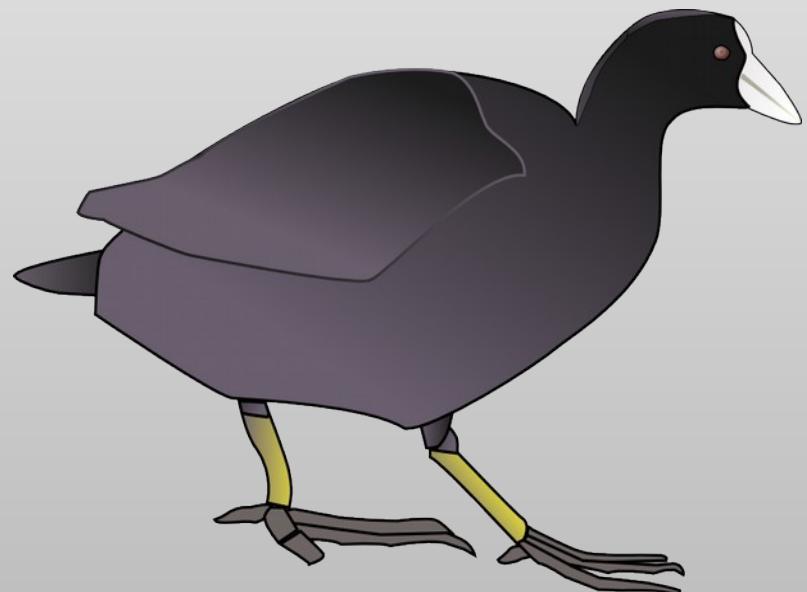




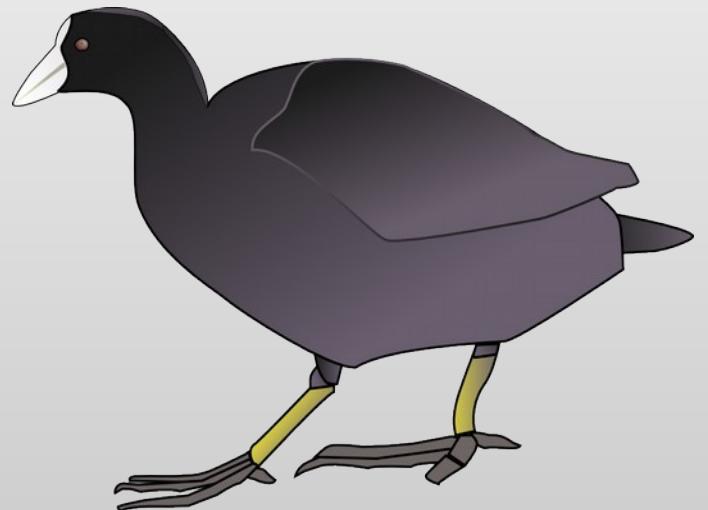
# Model-Building of Proteins Using X-ray and Cryo-EM Data With *Coot*



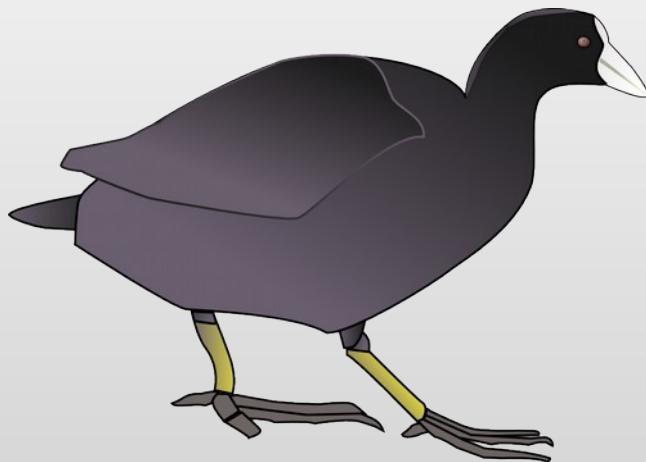
Paul Emsley  
May 2017

# Modelling Proteins with *Coot*

- Overview
  - Low resolution side-chains
  - Tools for EM
  - Tools for NCS



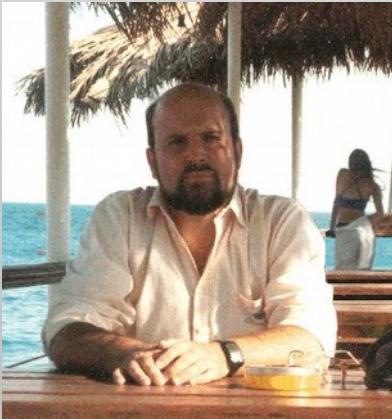
# Acknowledgments, Collaborators



Bernhard  
Lohkamp



Kevin  
Cowtan



Eugene  
Krissinel



Stuart  
McNicholas



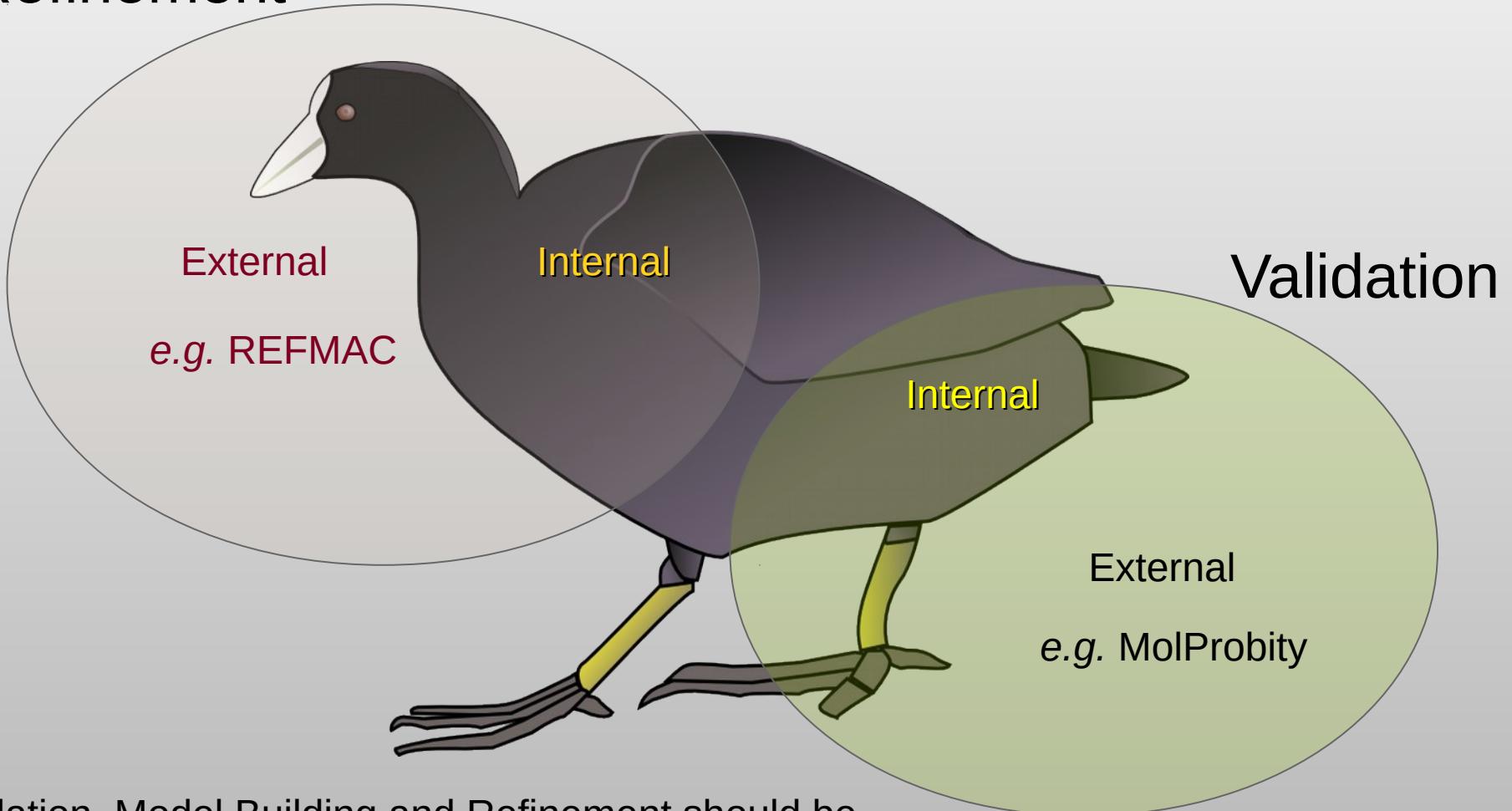
Martin  
Noble



Alexei  
Vagin

# Feature Integration

Refinement



Validation, Model Building and Refinement should be used together

# Real Space Refinement

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

- Major Feature of Coot
  - Gradient-based minimiser (BFGS derivative)
  - Geometry library is the standard CIF-based Refmac dictionary
    - Minimise deviations in bond length, angles, torsions, planes, chiral volume, non-bonded contacts
    - Including links and modifications
- Provides “interactive” refinement
- Subject to substantial extension

# Representation of Results:

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

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

The first attempt

Student Reaction:

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

# Representation of Results:



Second attempt...

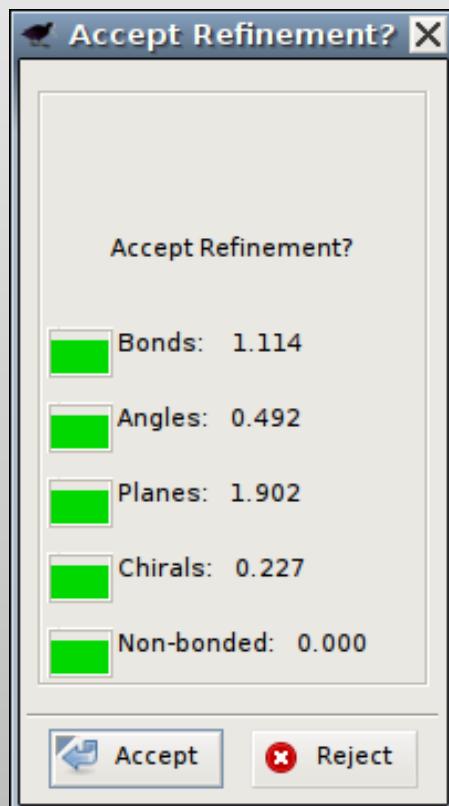
Student Reaction:

“Oh, box of meaningless numbers.

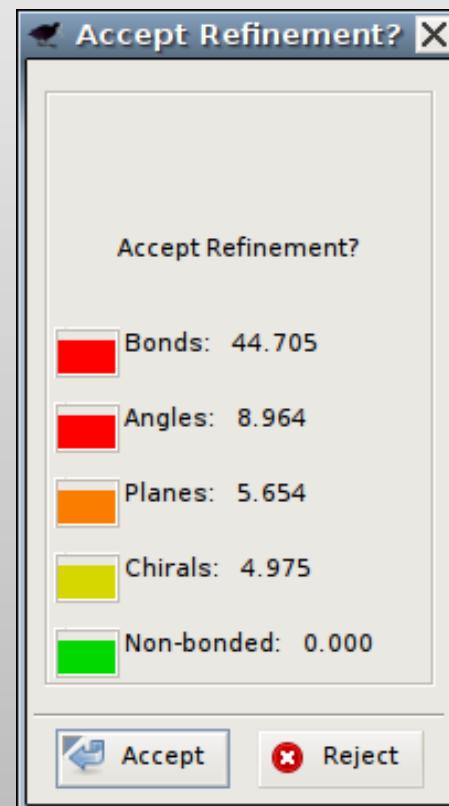
Go away”

# Representation of Results: “Traffic Lights”

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



Good refinement



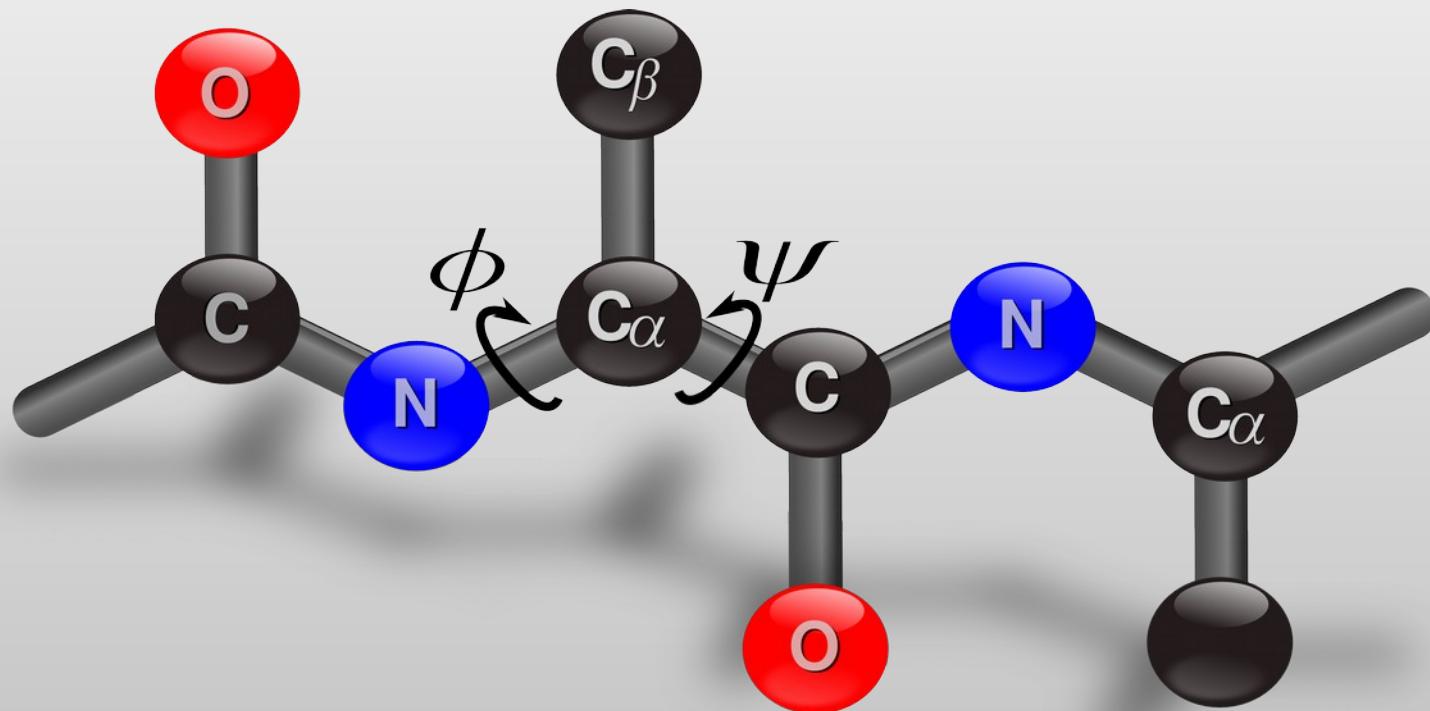
Bad refinement

9/100

# Refinement Techniques

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

# Peptide Backbone Geometry



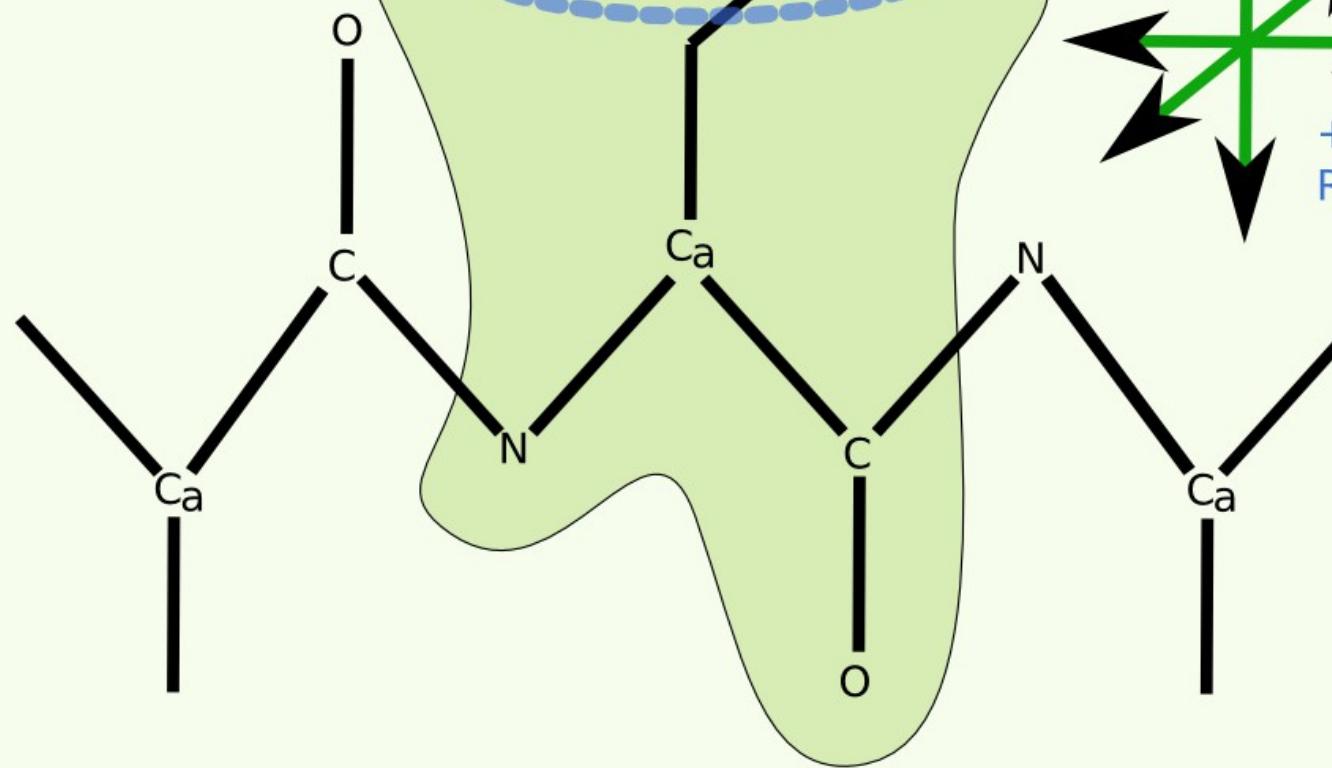
# Low Resolution Model-Building

- “Backrub” rotamers

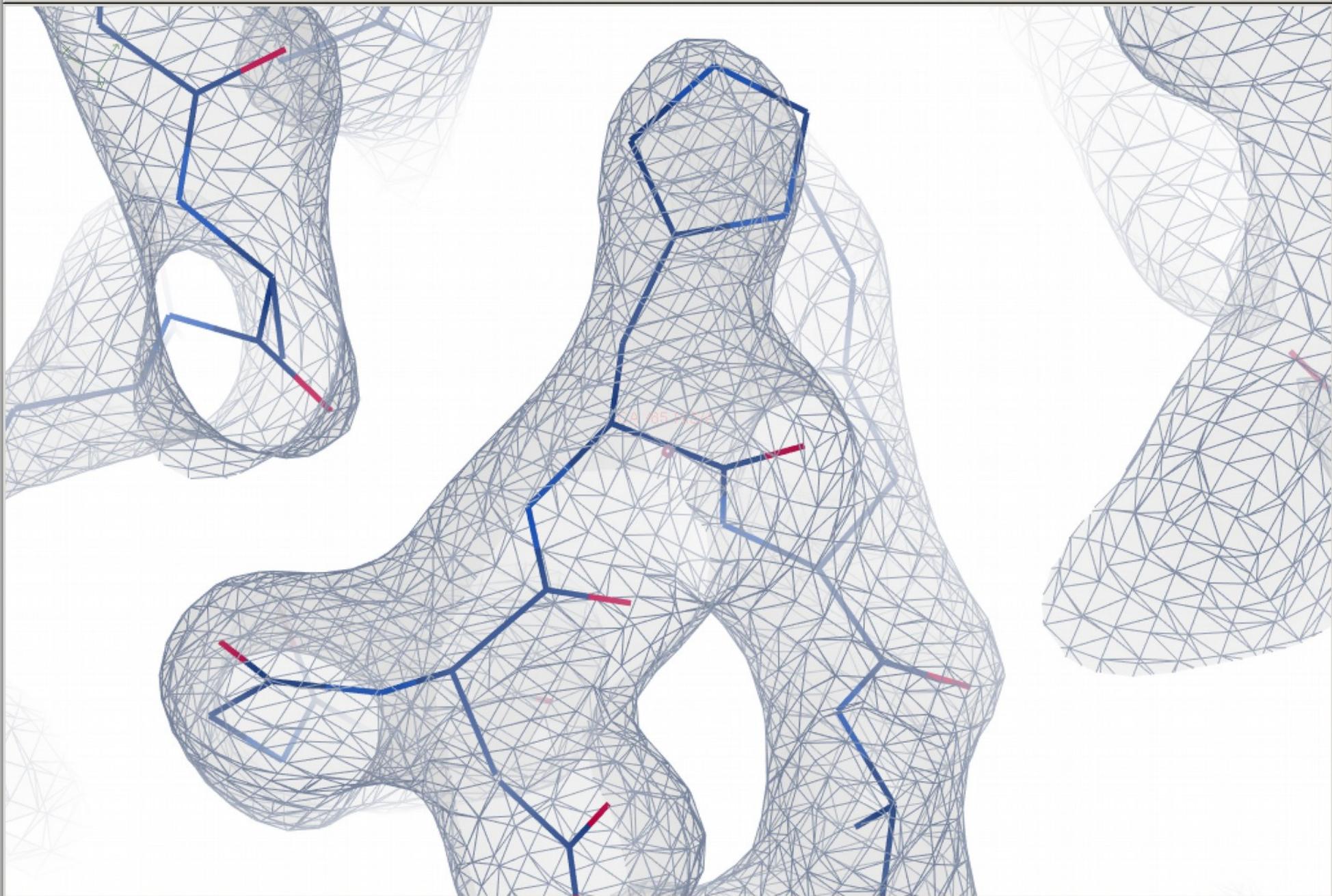
~~Current~~ Low Resolution Rotamer Search

Previous

Rotamer Search



+ Rigid Body  
Refinement



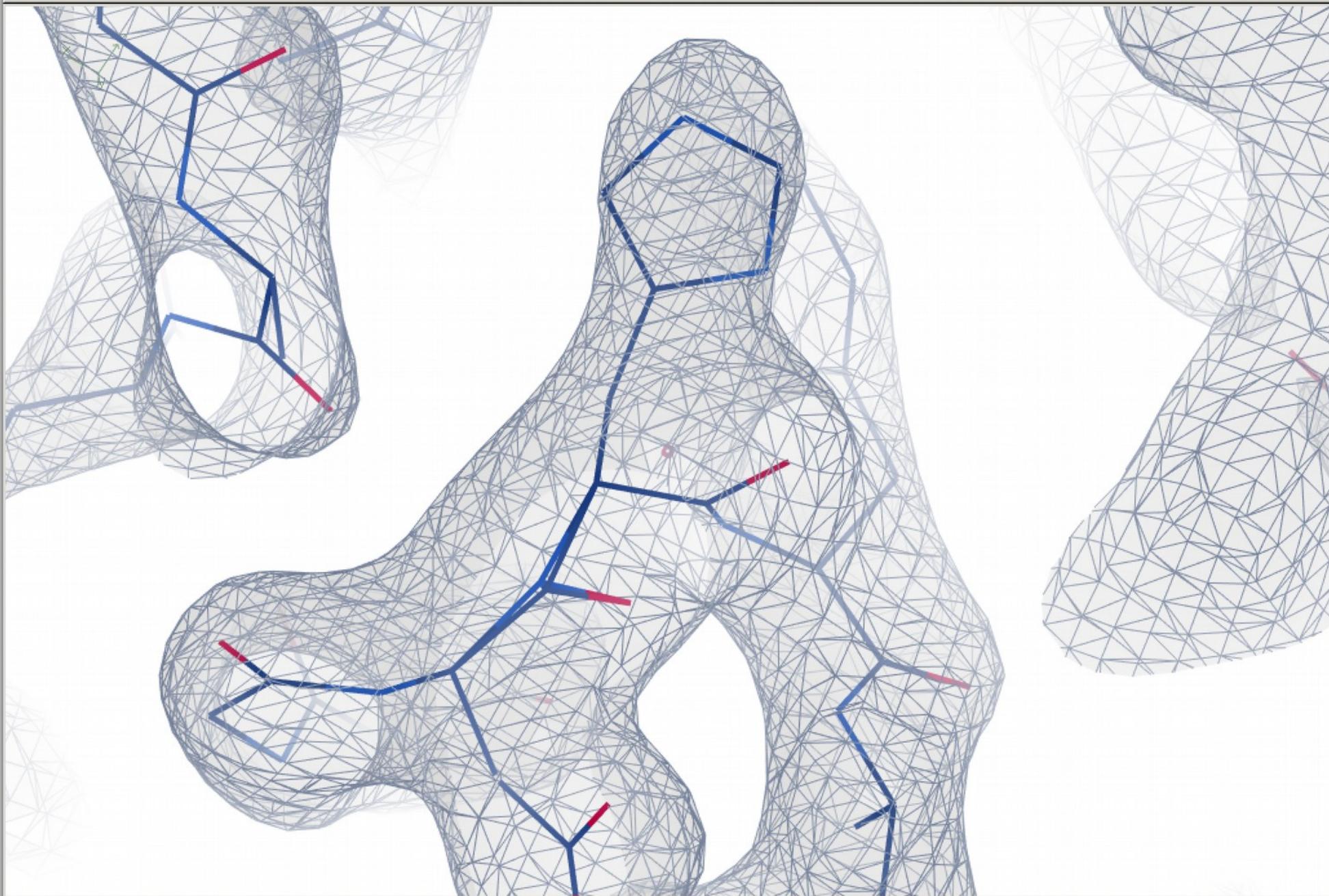
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)



Coot 0.8.7-pre EL (revision count 6456)

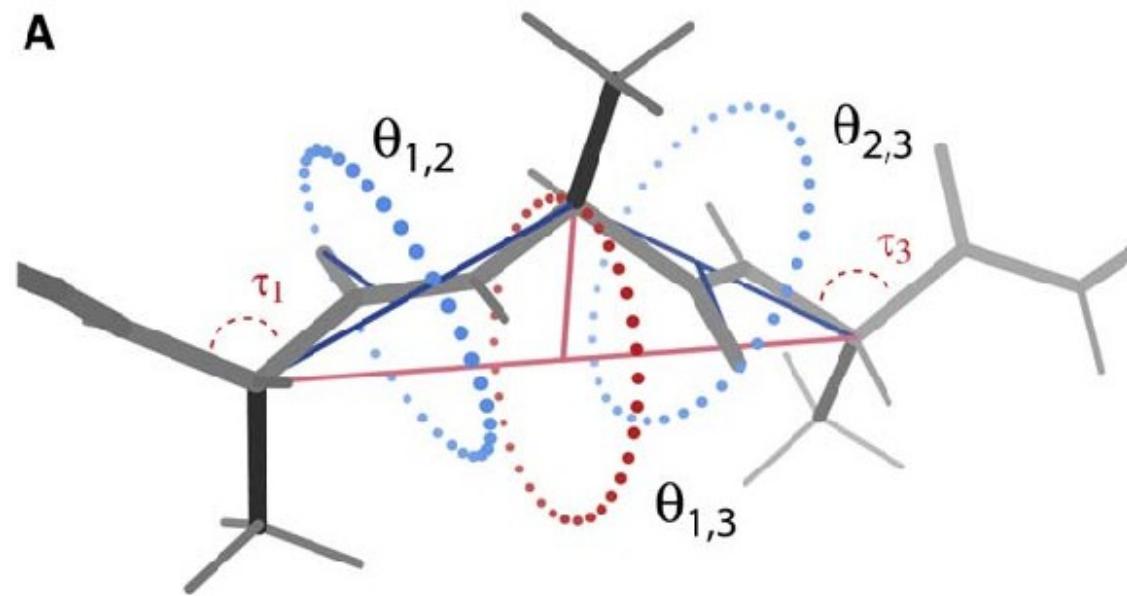
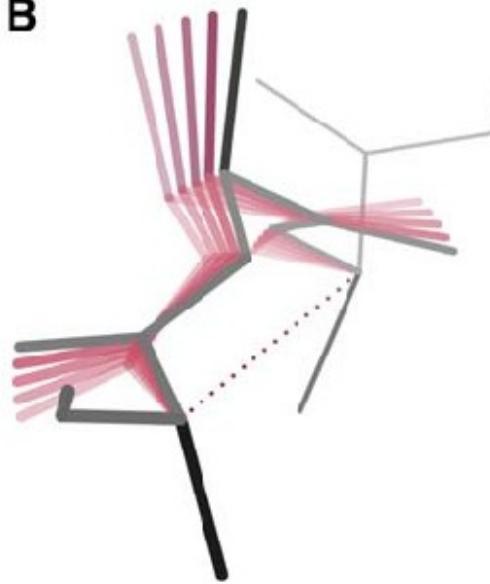
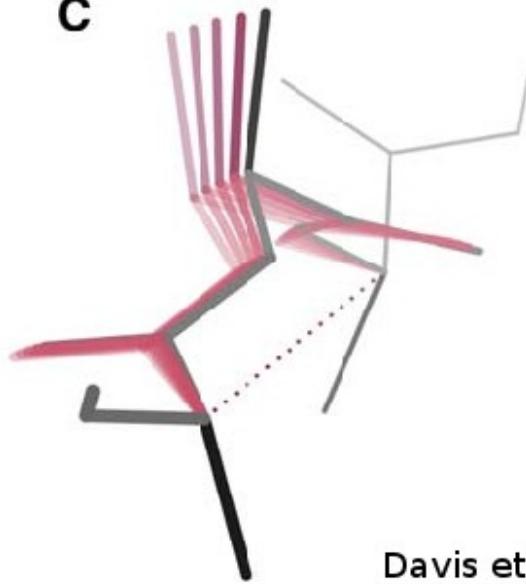
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug

Reset View Display Manager Ligand Builder Sphere Refine Sphere Refine + Backrub Rotamers



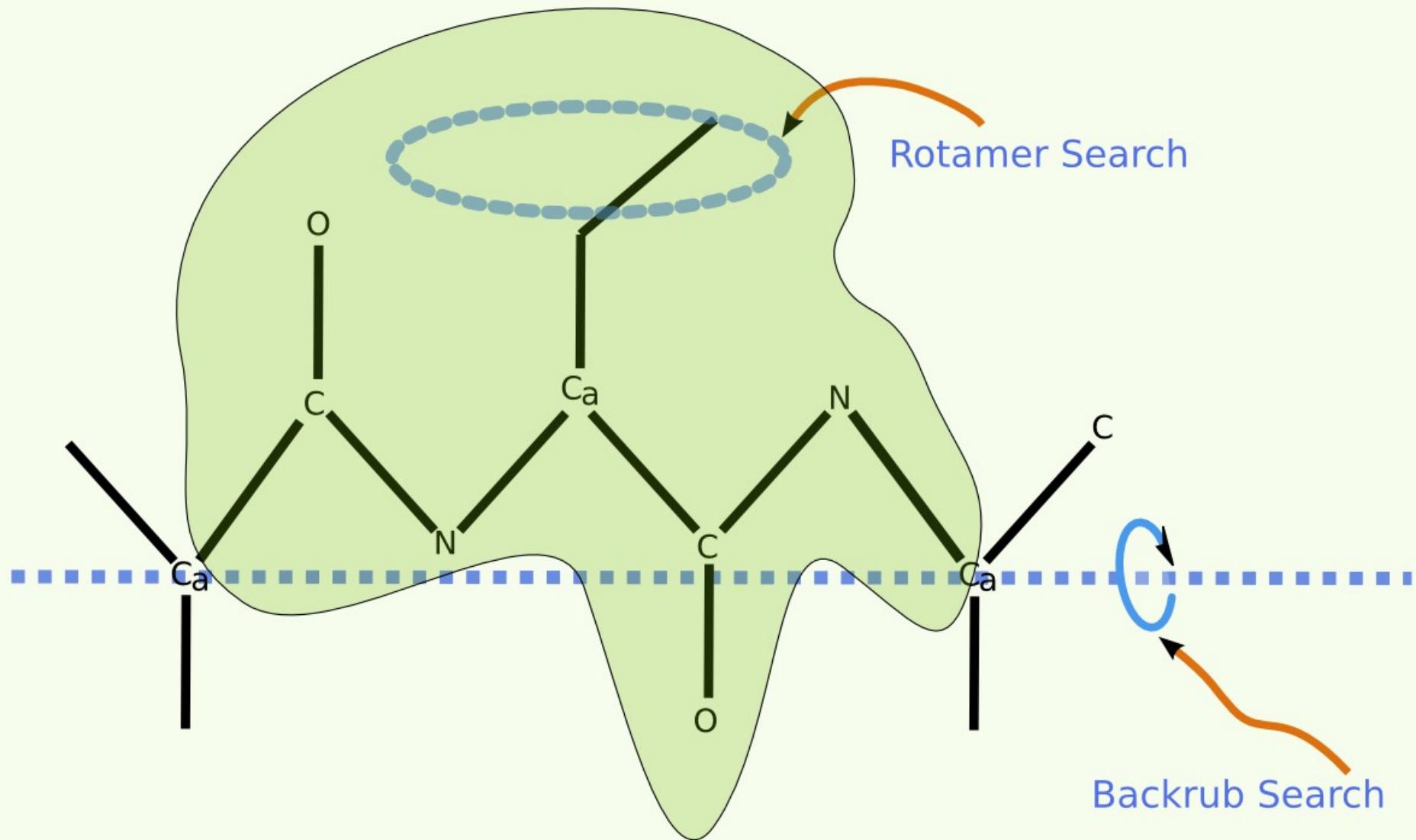
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)



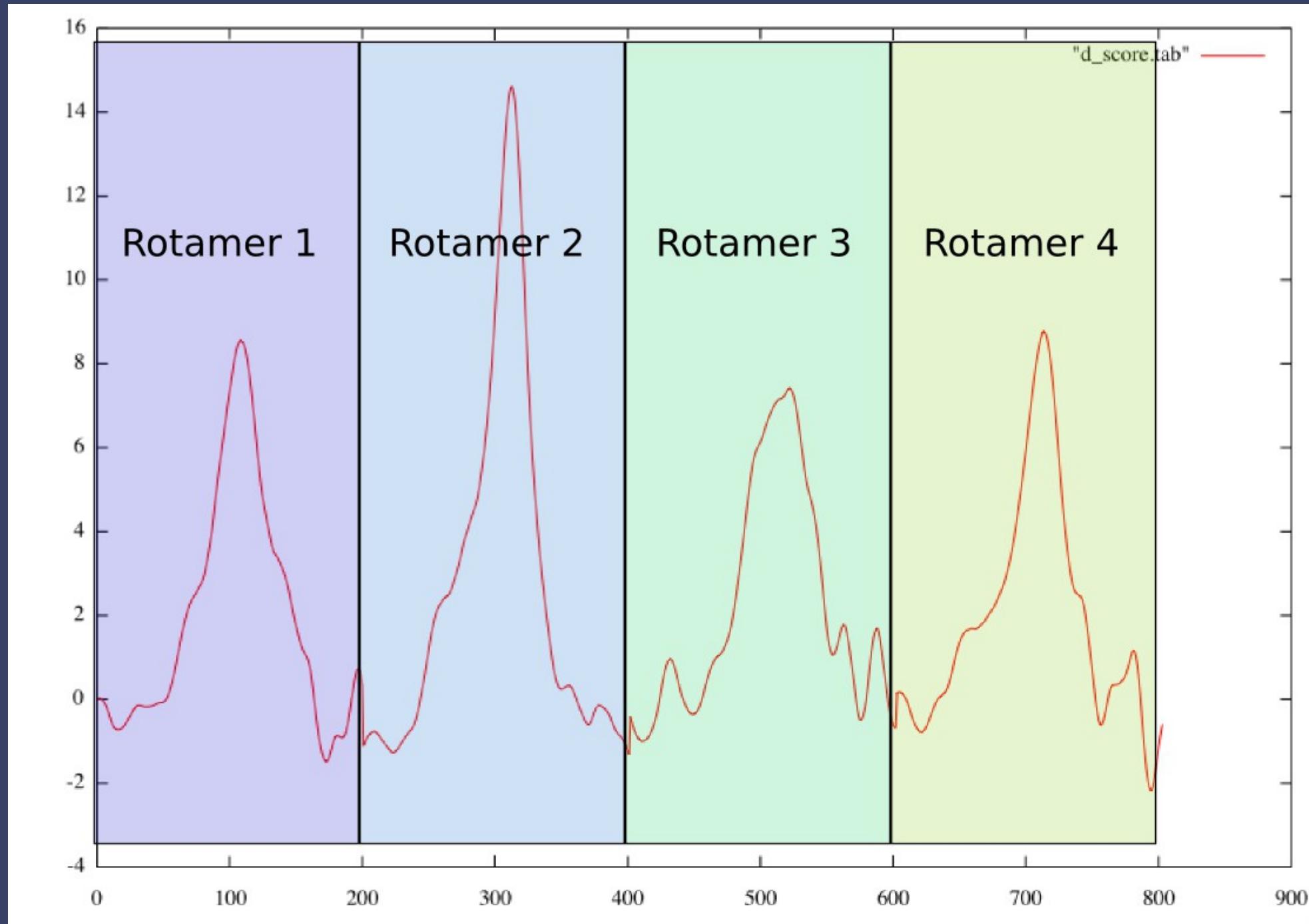
**A****B****C**

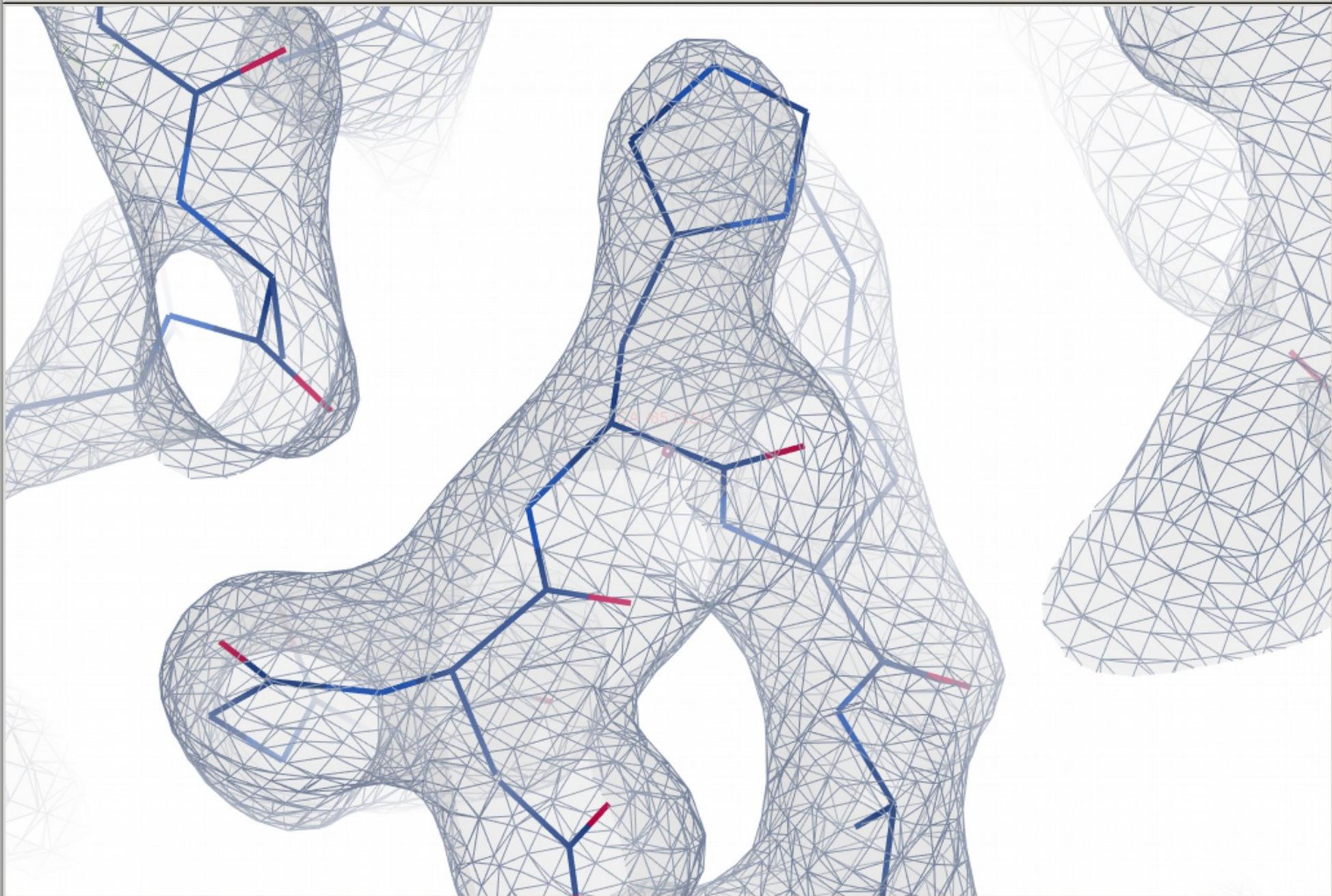
Davis et al. (2006) Structure

# New Low Resolution Rotamer Search



After Fitting Tools in KING/Molprobity





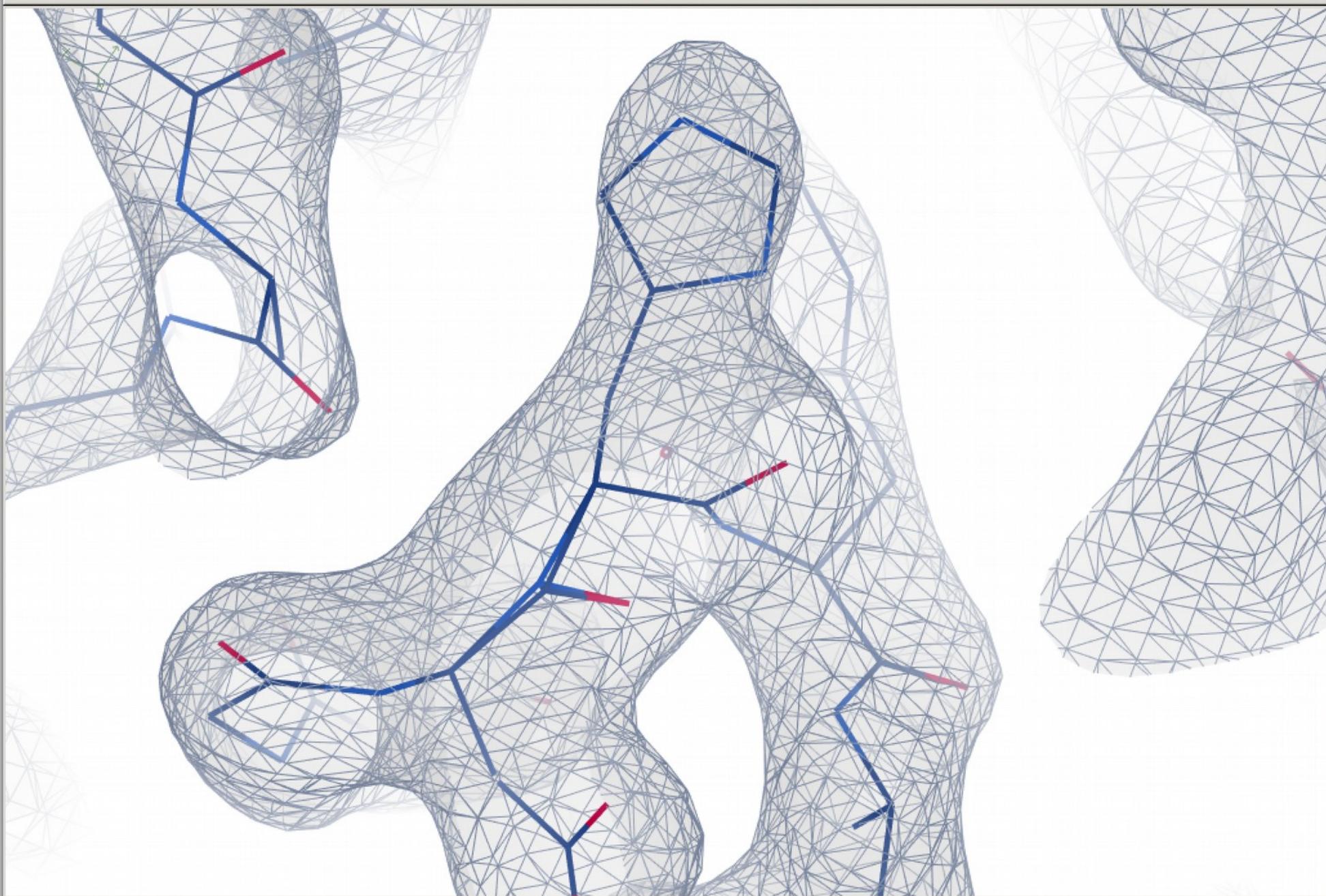
(mol. no: 3) CA /1/A/85 HIS occ: 1.00 bf: 19.16 ele: C pos: (57.45,15.65,14.20)



Coot 0.8.7-pre EL (revision count 6456)

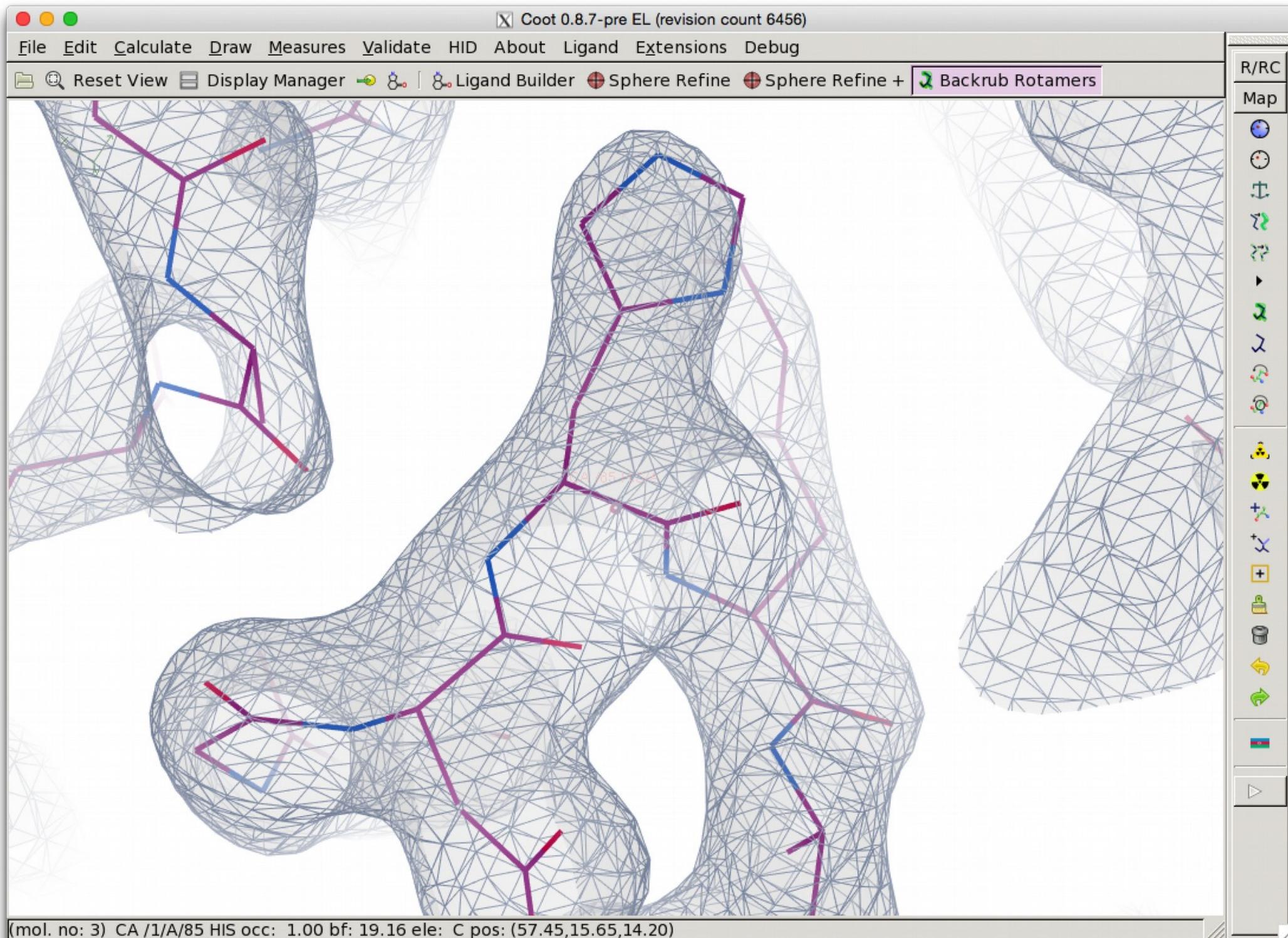
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug

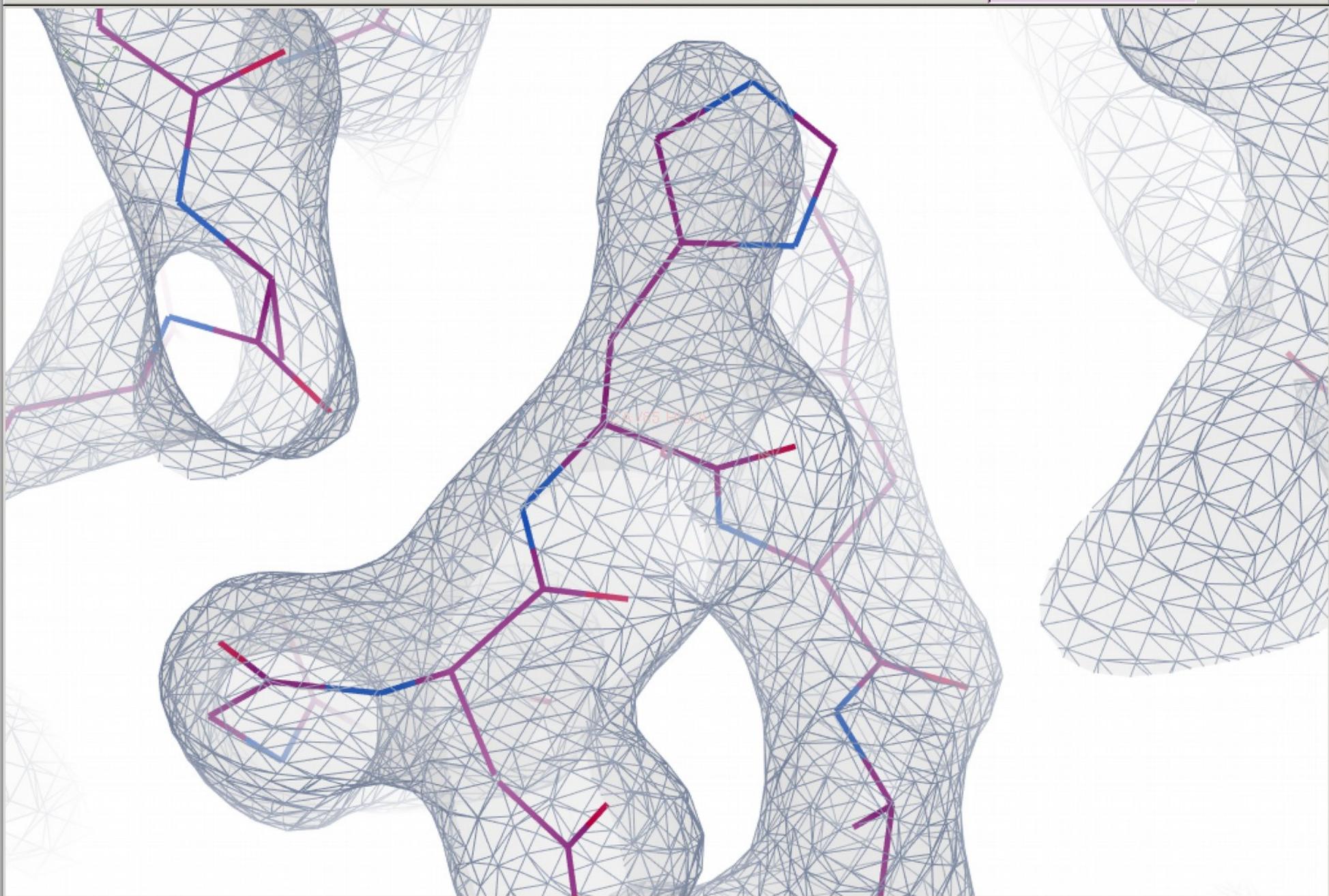
Reset View Display Manager Ligand Builder Sphere Refine Sphere Refine + Backrub Rotamers



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

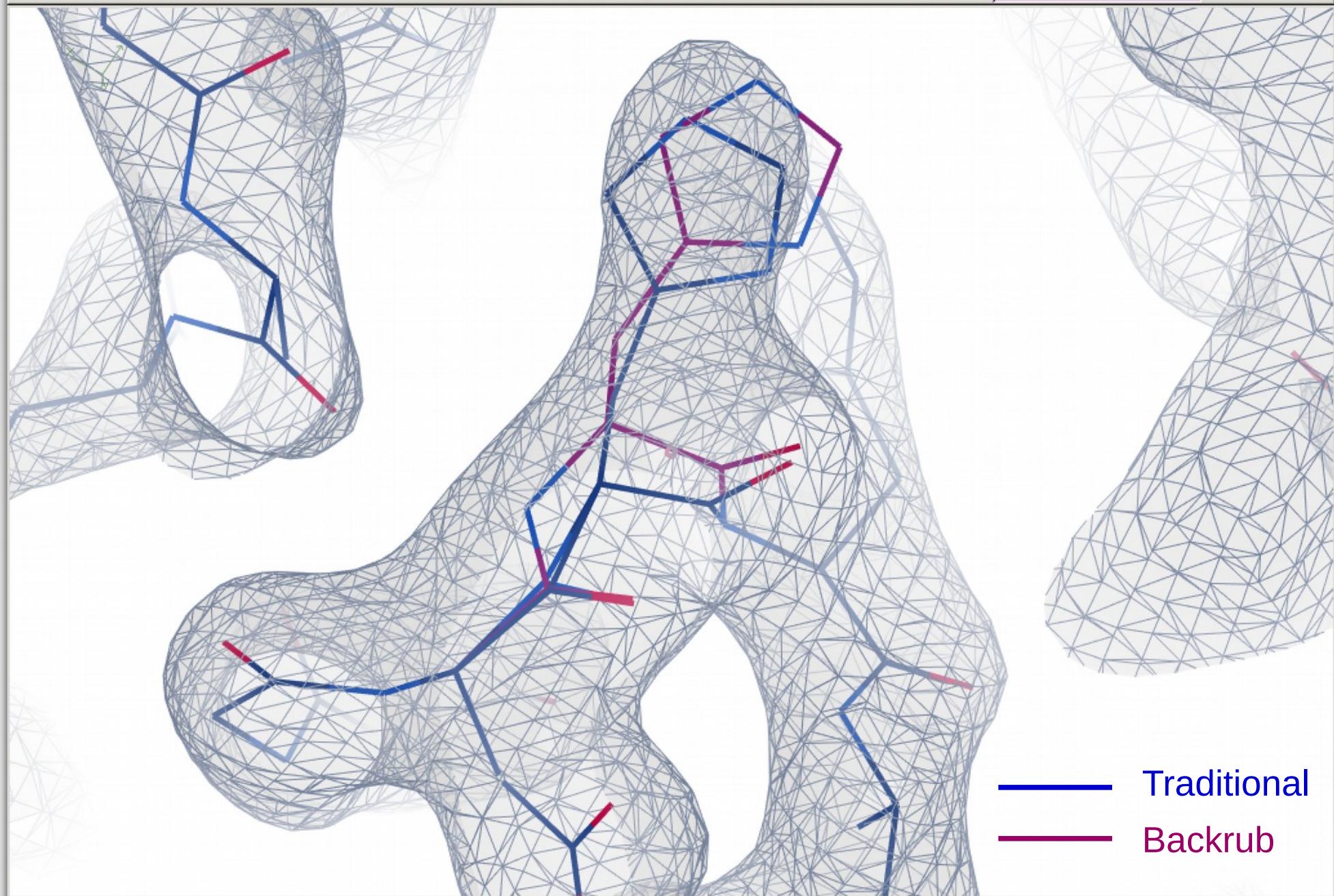








Coot 0.8.7-pre EL (revision count 6456)

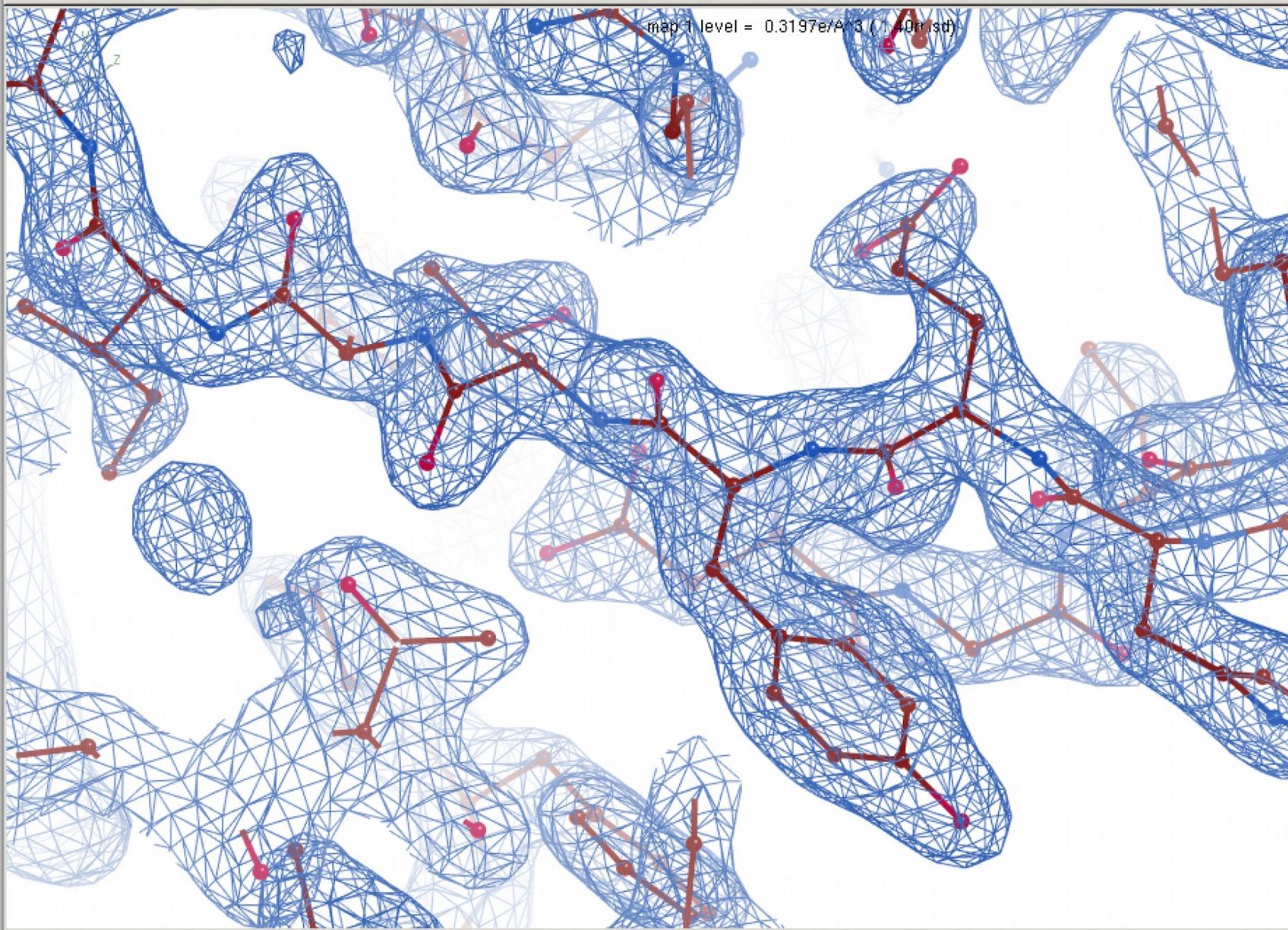
[File](#) [Edit](#) [Calculate](#) [Draw](#) [Measures](#) [Validate](#) [HID](#) [About](#) [Ligand](#) [Extensions](#) [Debug](#)[Reset View](#) [Display Manager](#) [Ligand Builder](#) [Sphere Refine](#) [Sphere Refine +](#) [Backrub Rotamers](#)R/RC  
Map

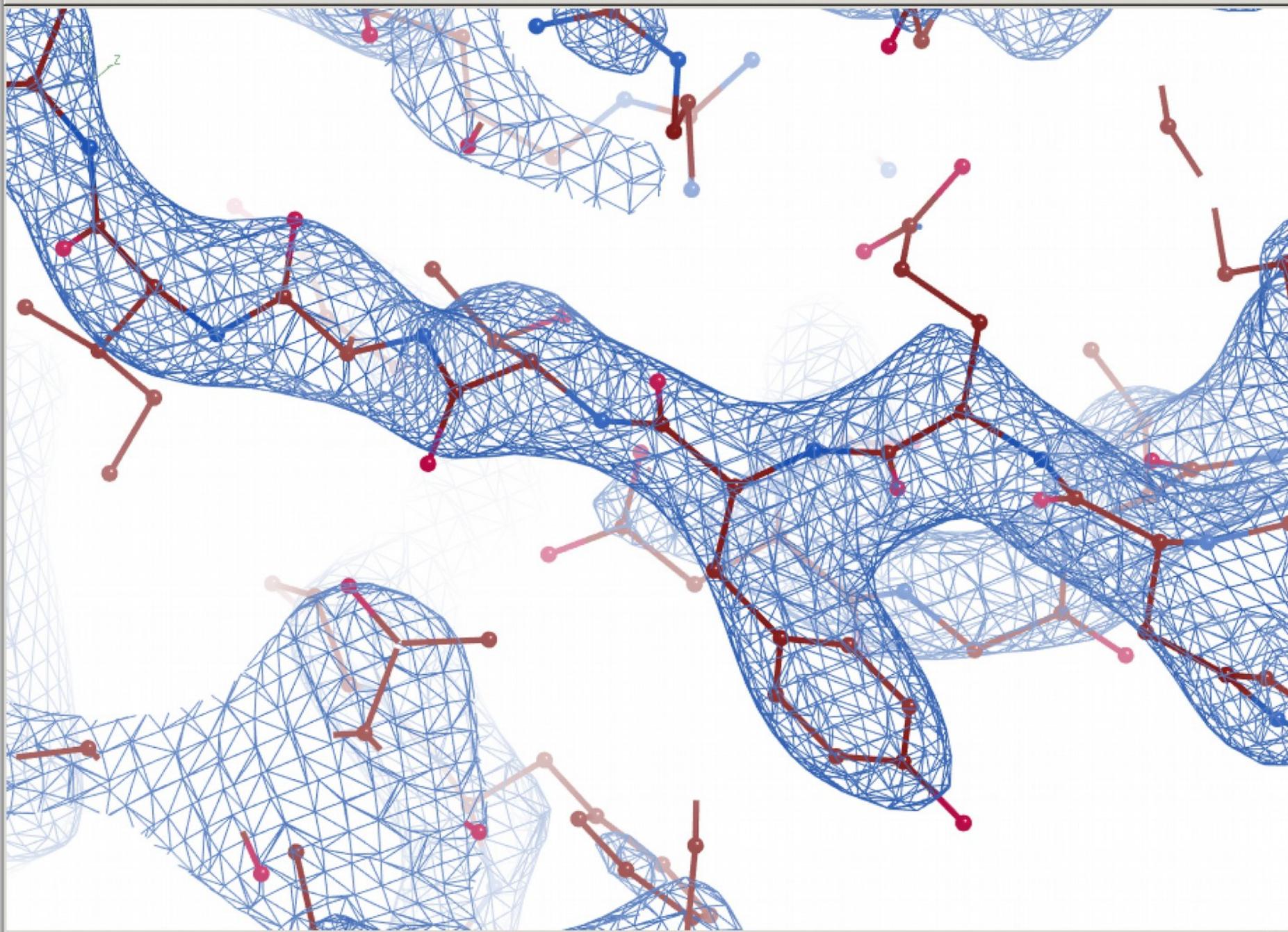
(mol. no: 3) CA /1/A/89 PHE occ: 1.00 bf: 20.19 ele: C pos: (53.21,12.14, 7.56)

X Coot 0.8.9-pre EL (revision count 6954)

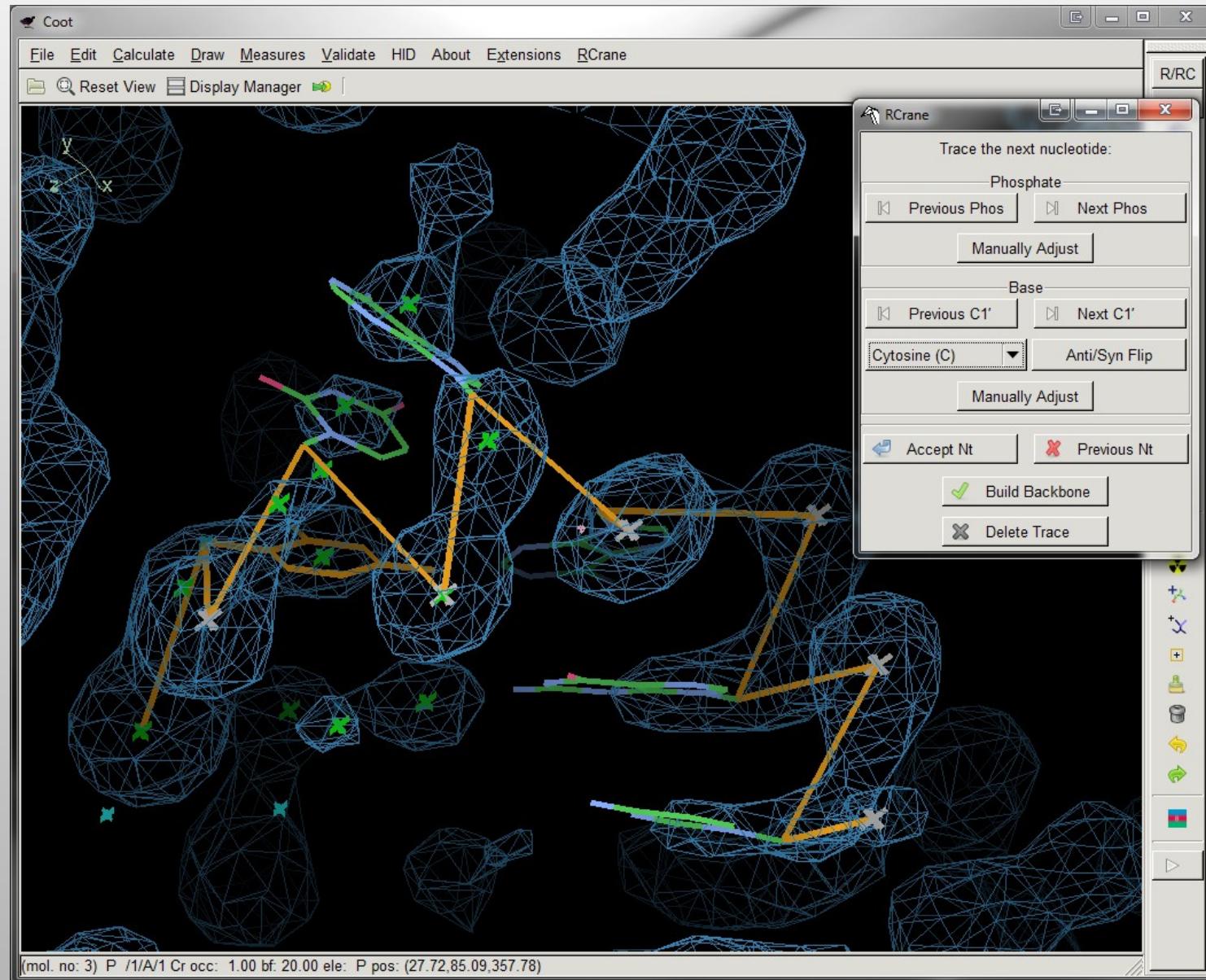
File Edit Calculate Draw Measures Validate HID About Extensions Debug Glyco ProSMART Ligand Morph

Reset View Display Manager Ligand Builder Sphere Refine Sphere Refine + Sphere Regularization +



R/RC  
Map

# RCrane: Semi-Automated Building of RNA



# Handling EM maps

Coot

File Edit Calculate Draw Measures Validate HID About Extensions

Reset View

Display Manager

R/RC

Map

○

○

+

○

○

○

○

○

○

○

○

○

○

○

○

○

○

○

○

○

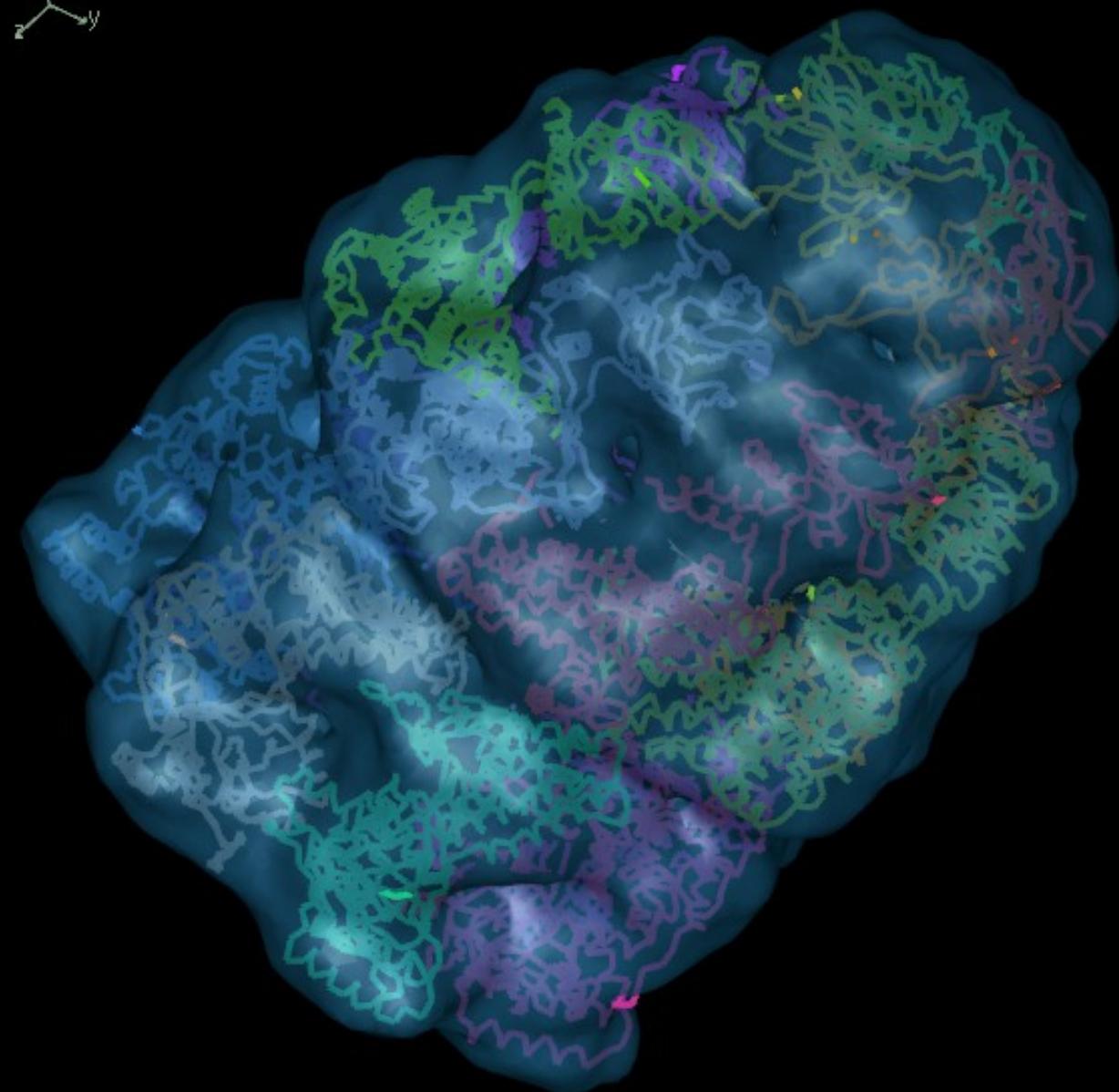
○

○

x

z

y



...coordinates 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

Atom

Site

Link

Region

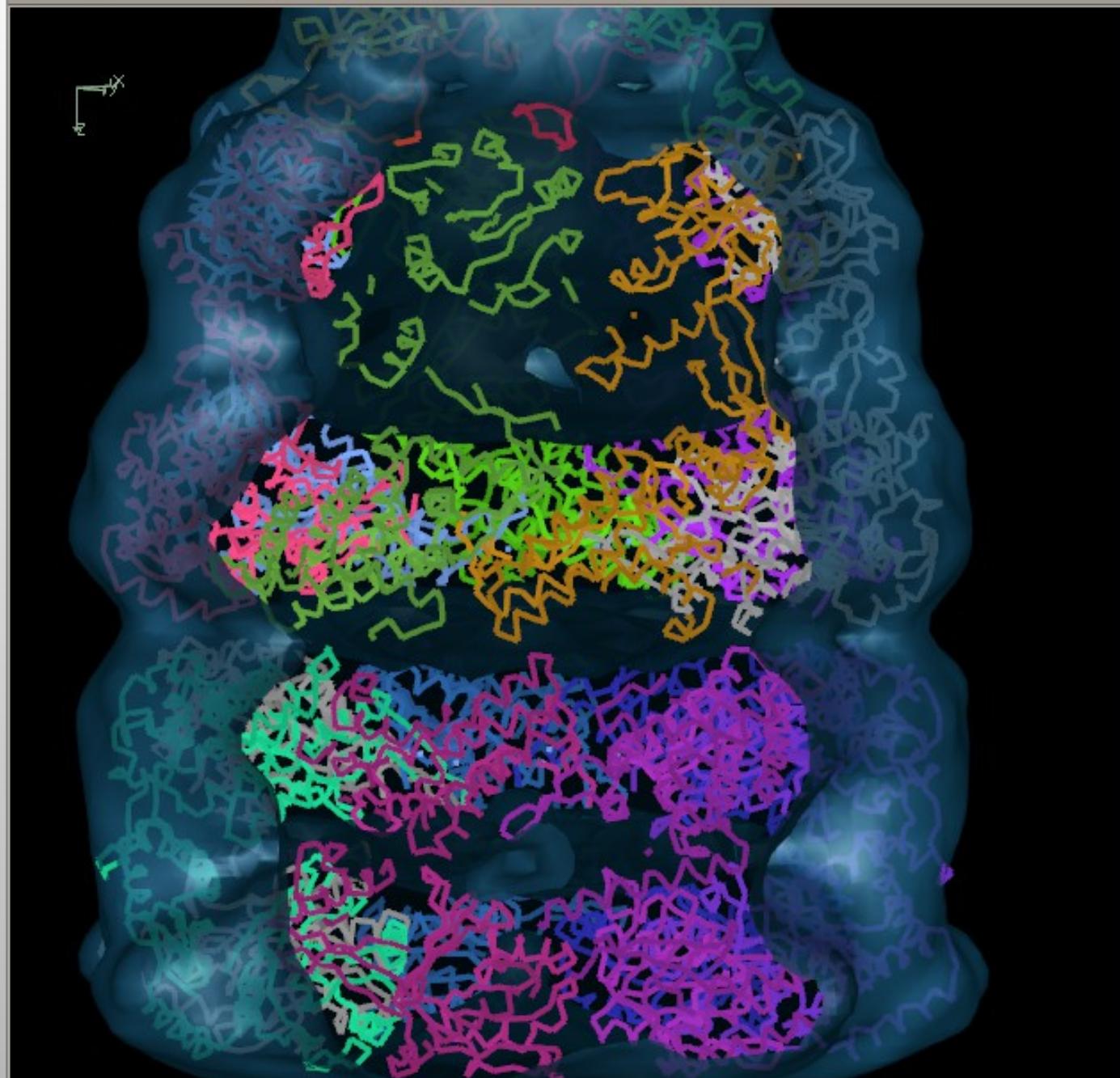
Surface

Vector

Angle

Diagonal

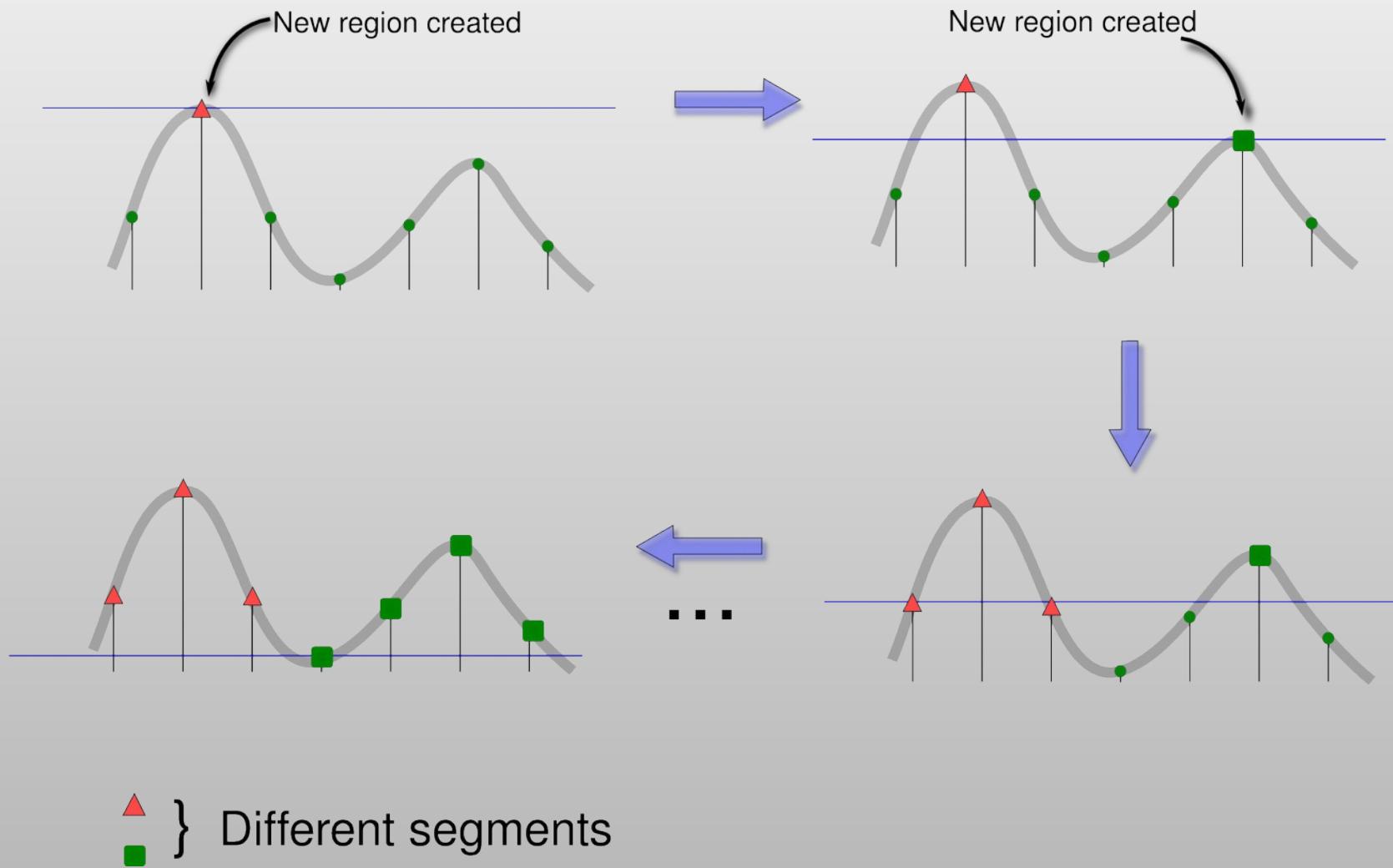
Plane

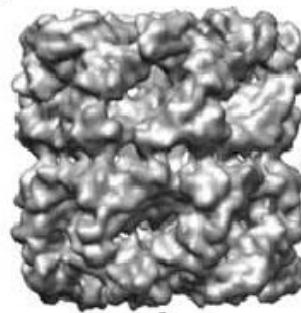
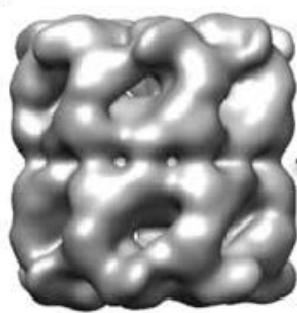
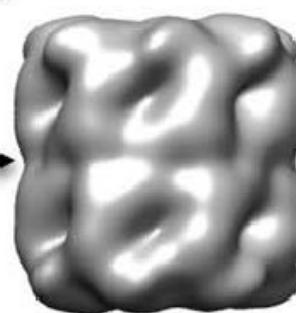
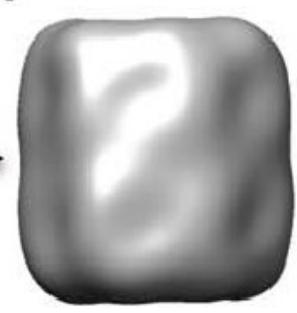
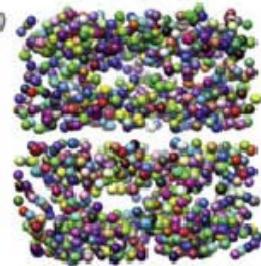
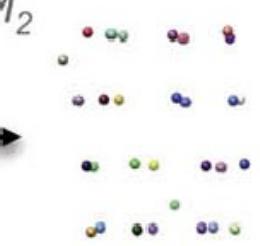
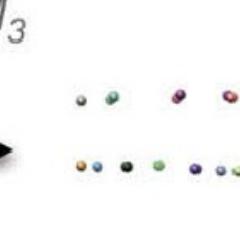
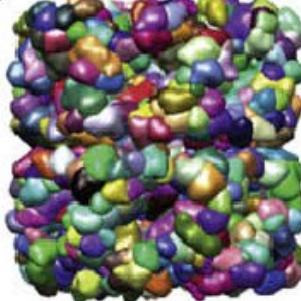
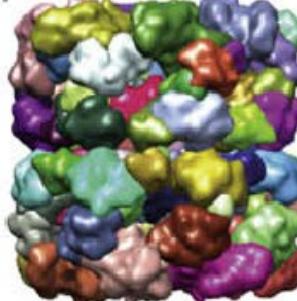
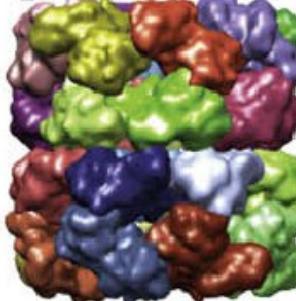
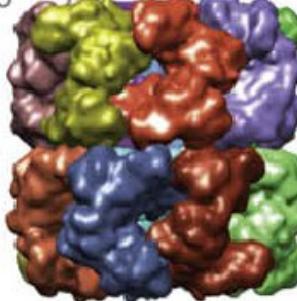
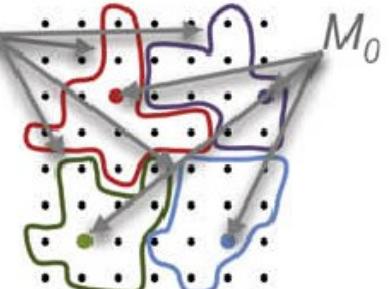
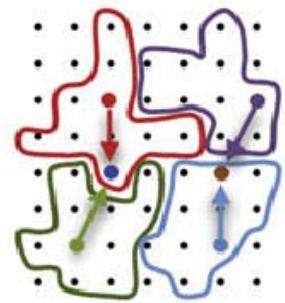
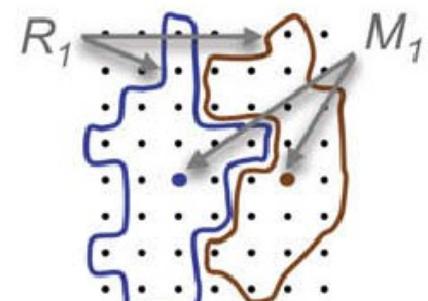


(mol. no: 1) CD /1/M/209 GLU occ: 1.00 bf: 100.09 ele: C pos: (167.80,210.31,219.52)

# Partitioning Maps: Watershed Algorithm

1D-analog



$D_0$  $D_1$  $D_2$  $D_3$  $M_0$  $M_1$  $M_2$  $M_3$  $R_0$  $R_1$  $R_2$  $R_3$  $R_0$  $M_0$  $M_1$ 

 Coot

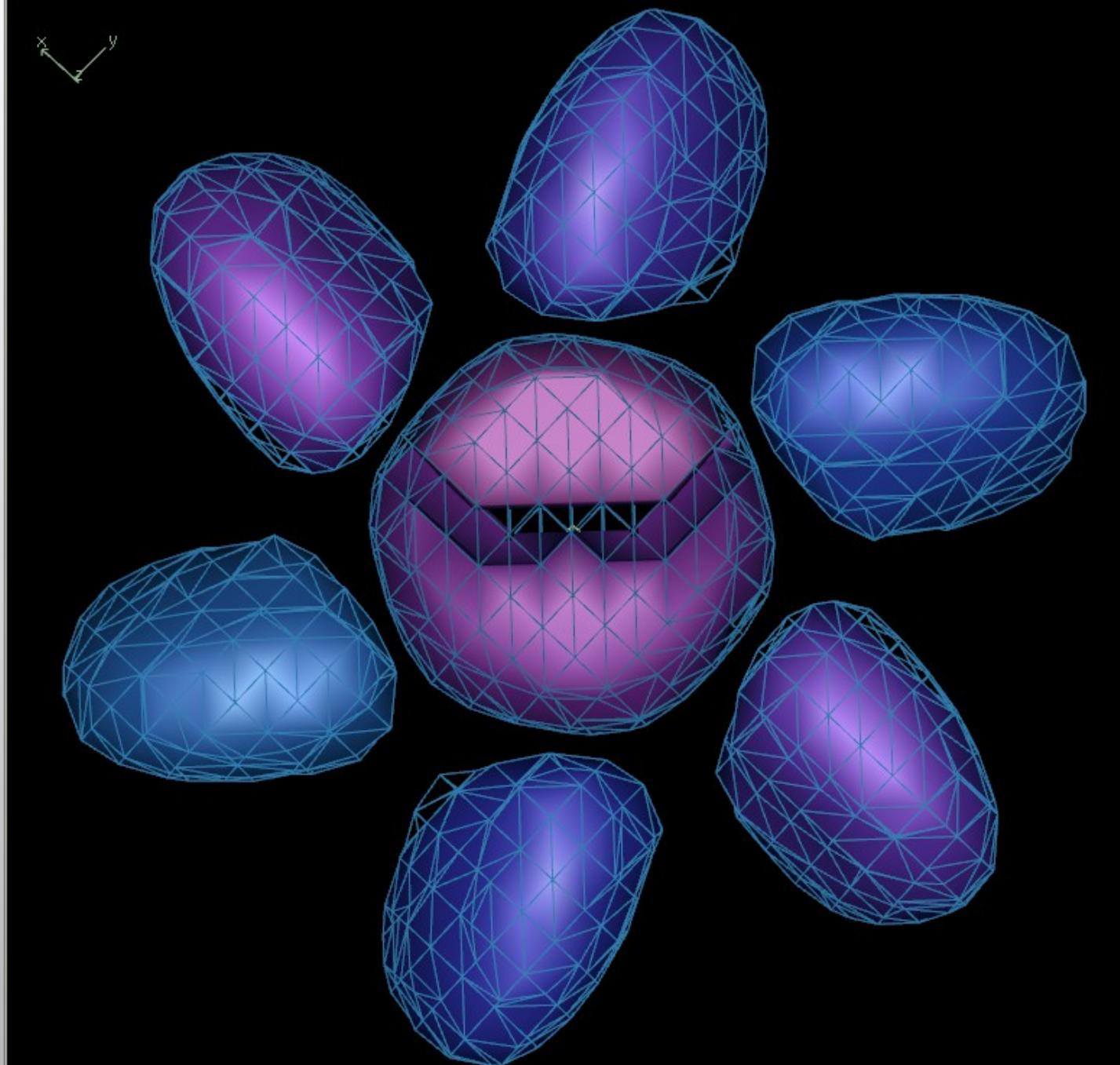
File Edit Calculate Draw Measures Validate HID About Extensions Solid

 Reset View

 Display Manager

 R/RC

 Map



 ●

 ○

 ⌈

 ⌈

 ⌈

 ▶

 ▶

 ▲

 ▾

 ▲

 ▲

 ▲

 ▾

 ▲

 ▾

 ▲

 ▾

 ▲

 ▾

 ▾

 ▾

 ▾

 ▾

Good morning Paul. Welcome to Coot

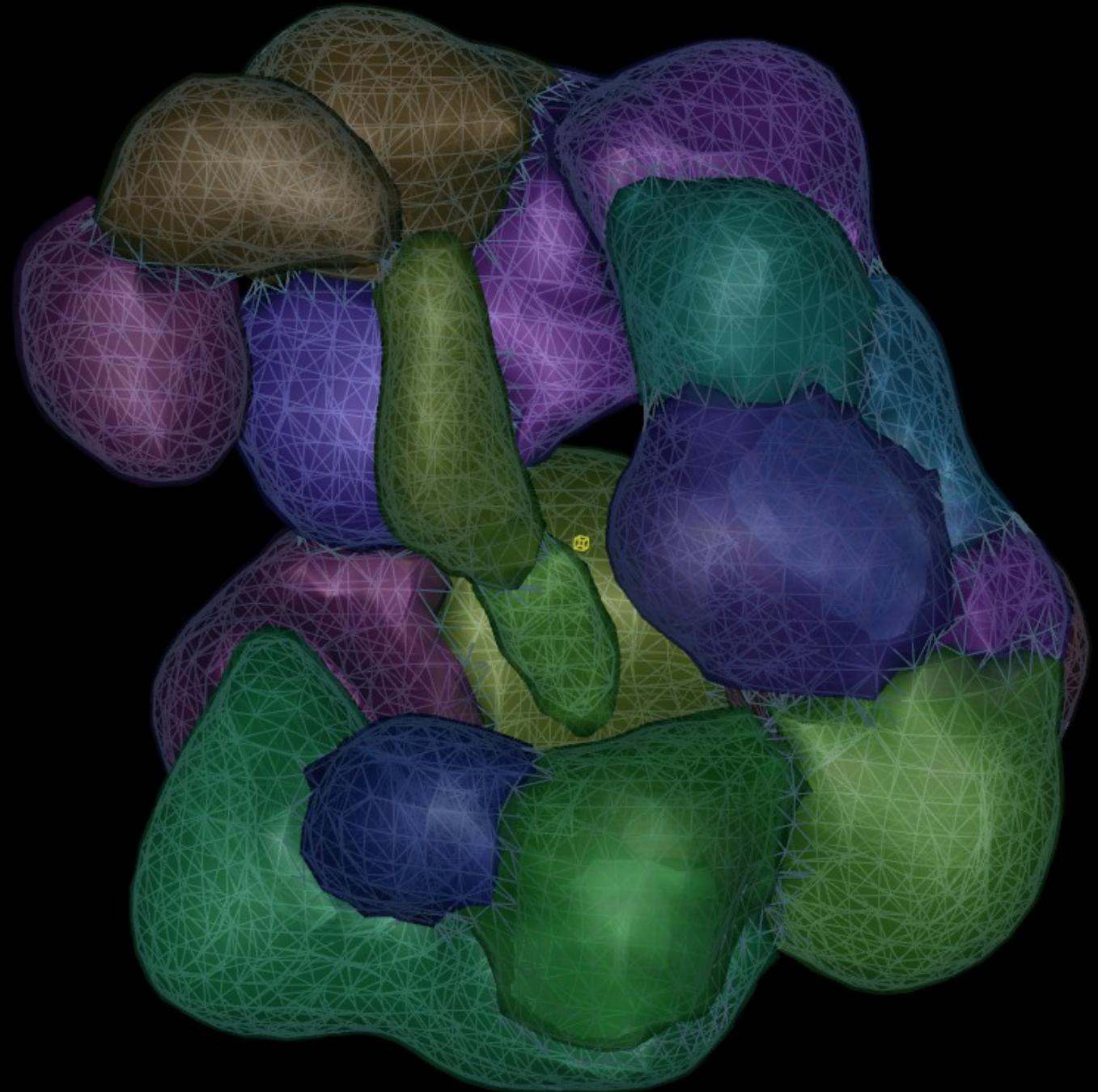
Coot

File Edit Calculate Draw Measures Validate HID About Extensions Solid

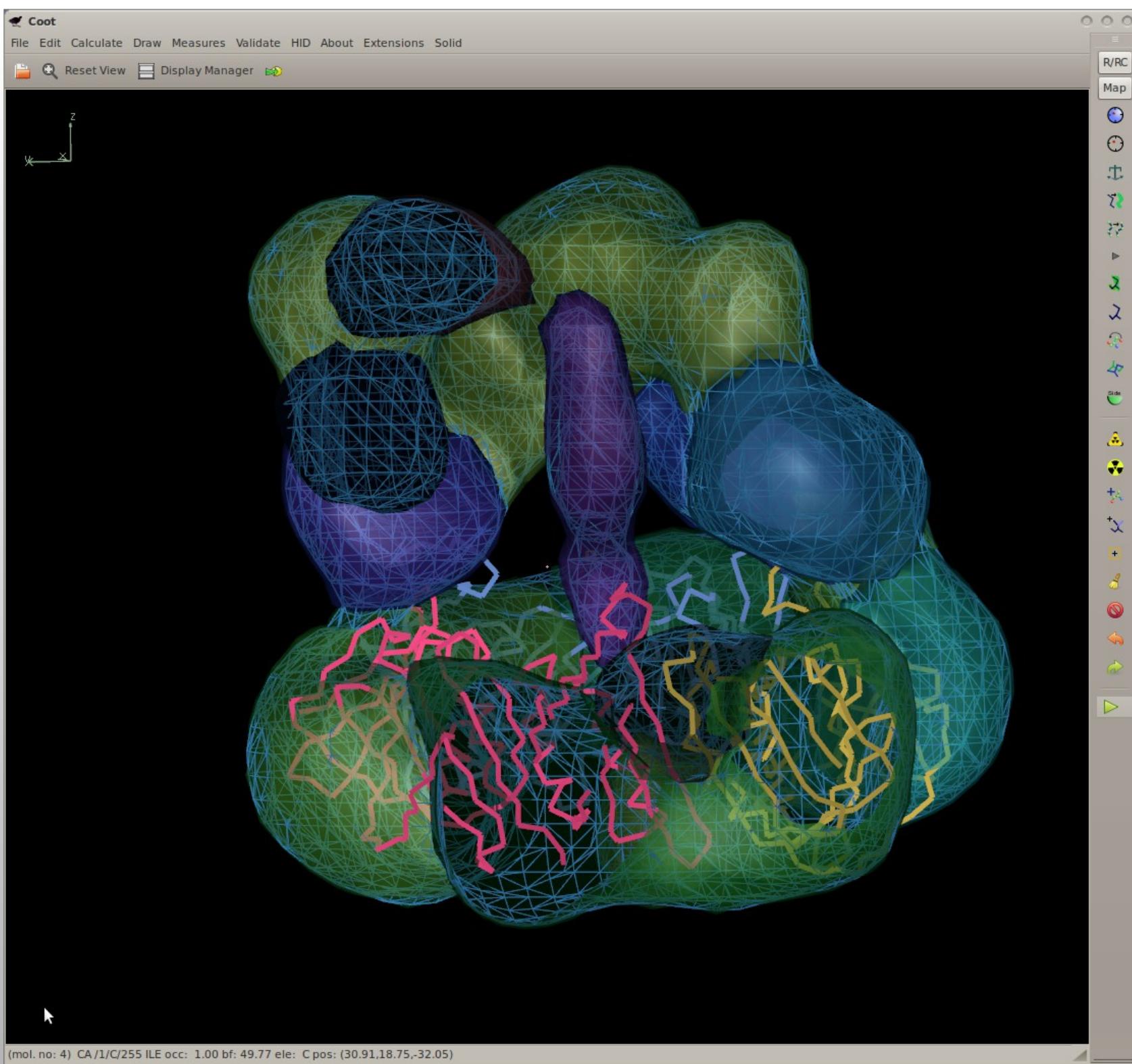
Reset View Display Manager

R/RC

Map



Successfully read coordinates file test.pdb. Molecule number 1 created.



# 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)

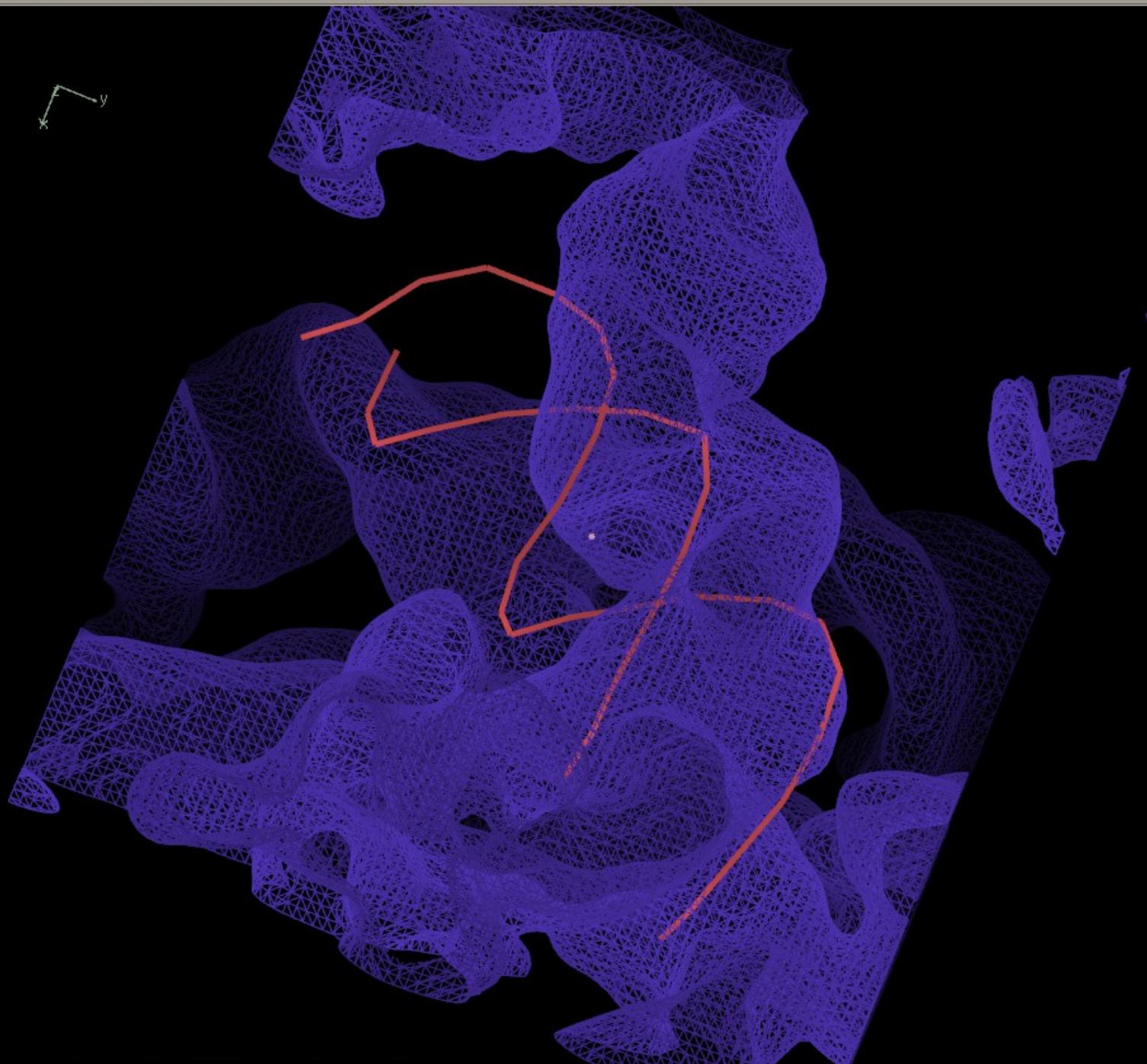
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph

Reset View Display Manager Full screen Sphere Refine

R/RC

Map

x  
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)

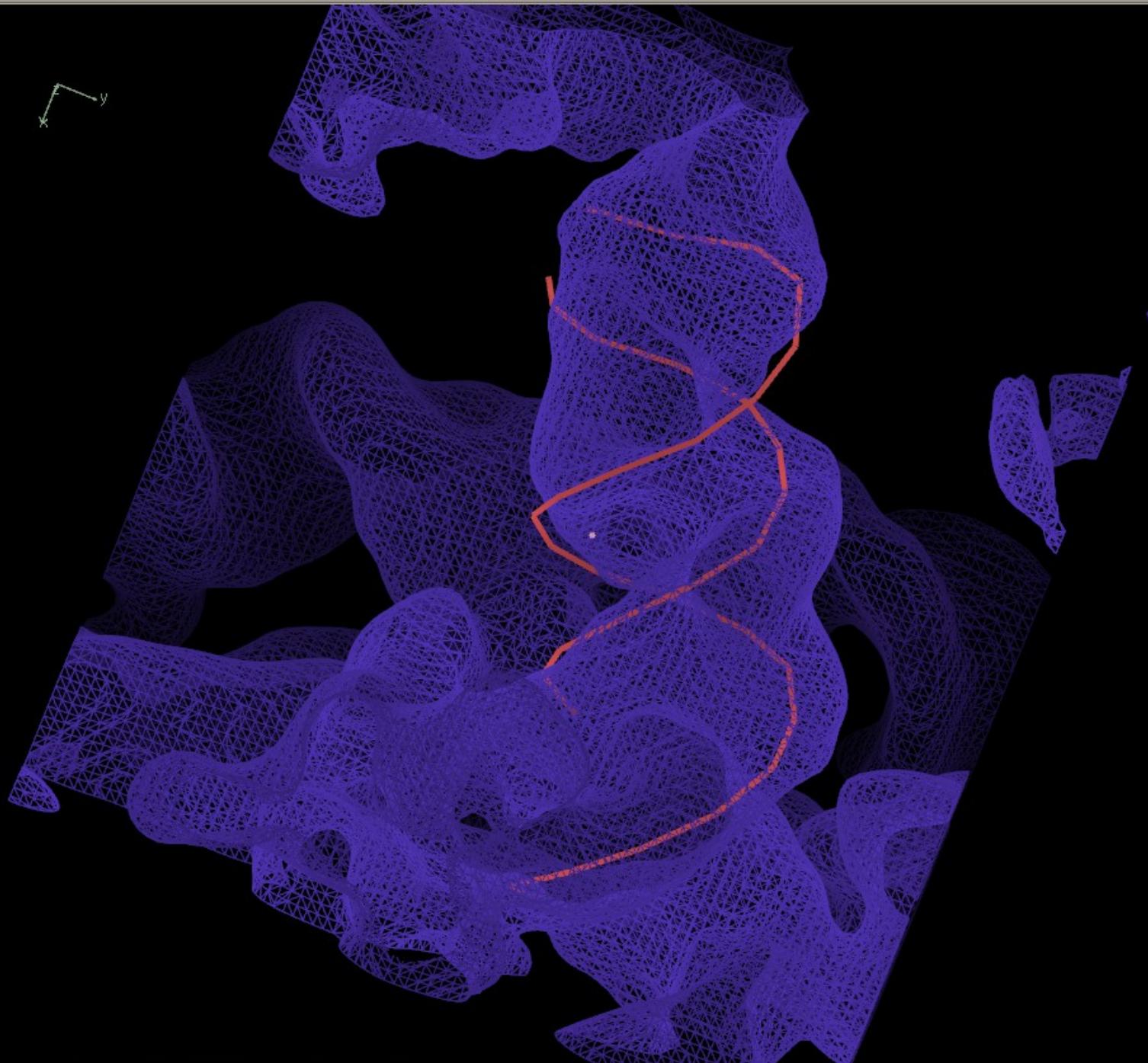
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph

Reset View Display Manager Full screen Sphere Refine

R/RC

Map

x  
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)

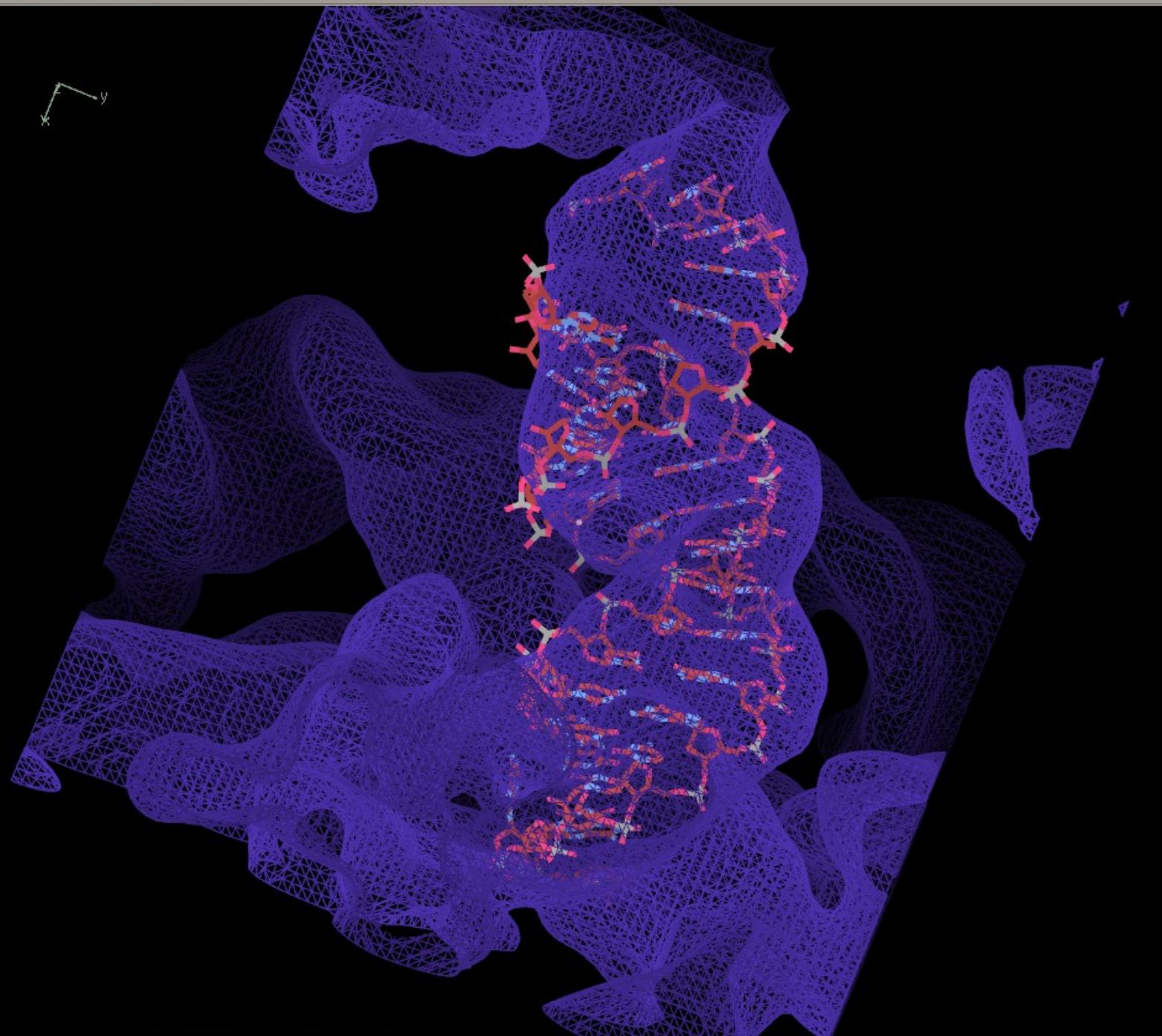
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph

Reset View Display Manager Full screen Sphere Refine

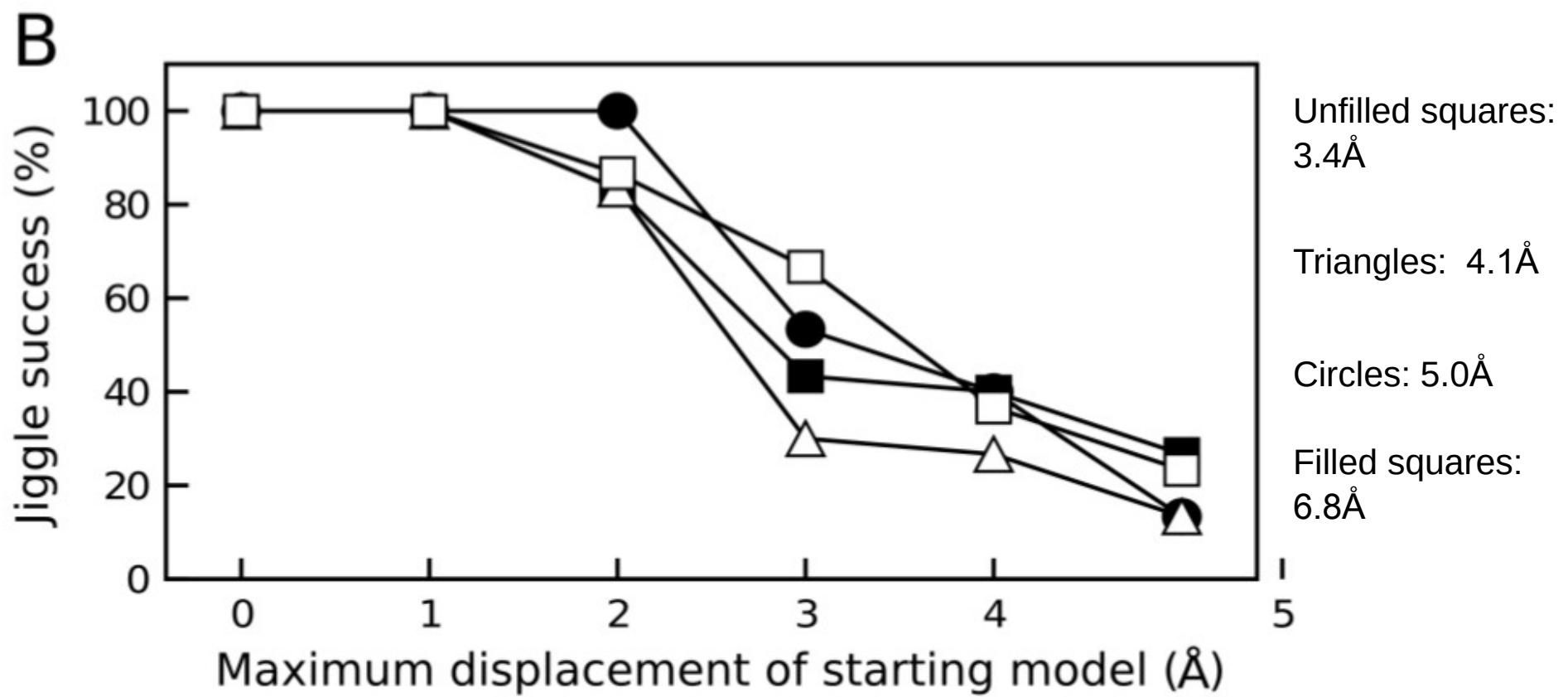
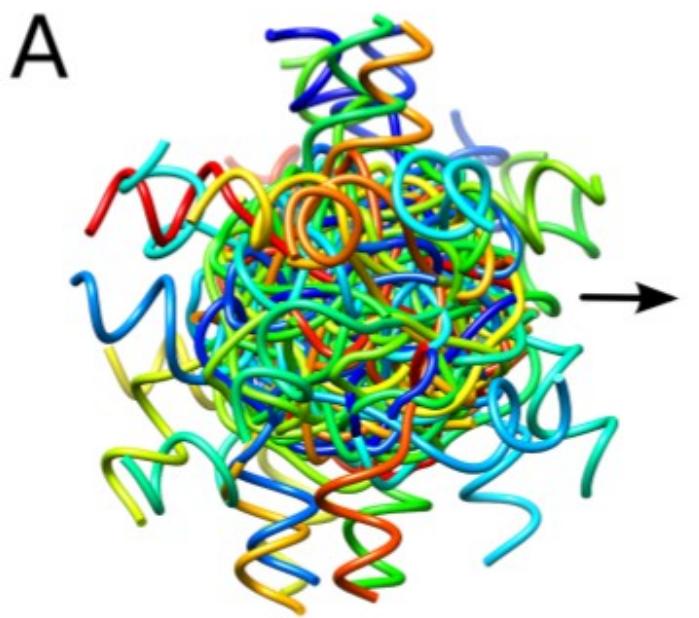
R/RC

Map

x  
y



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



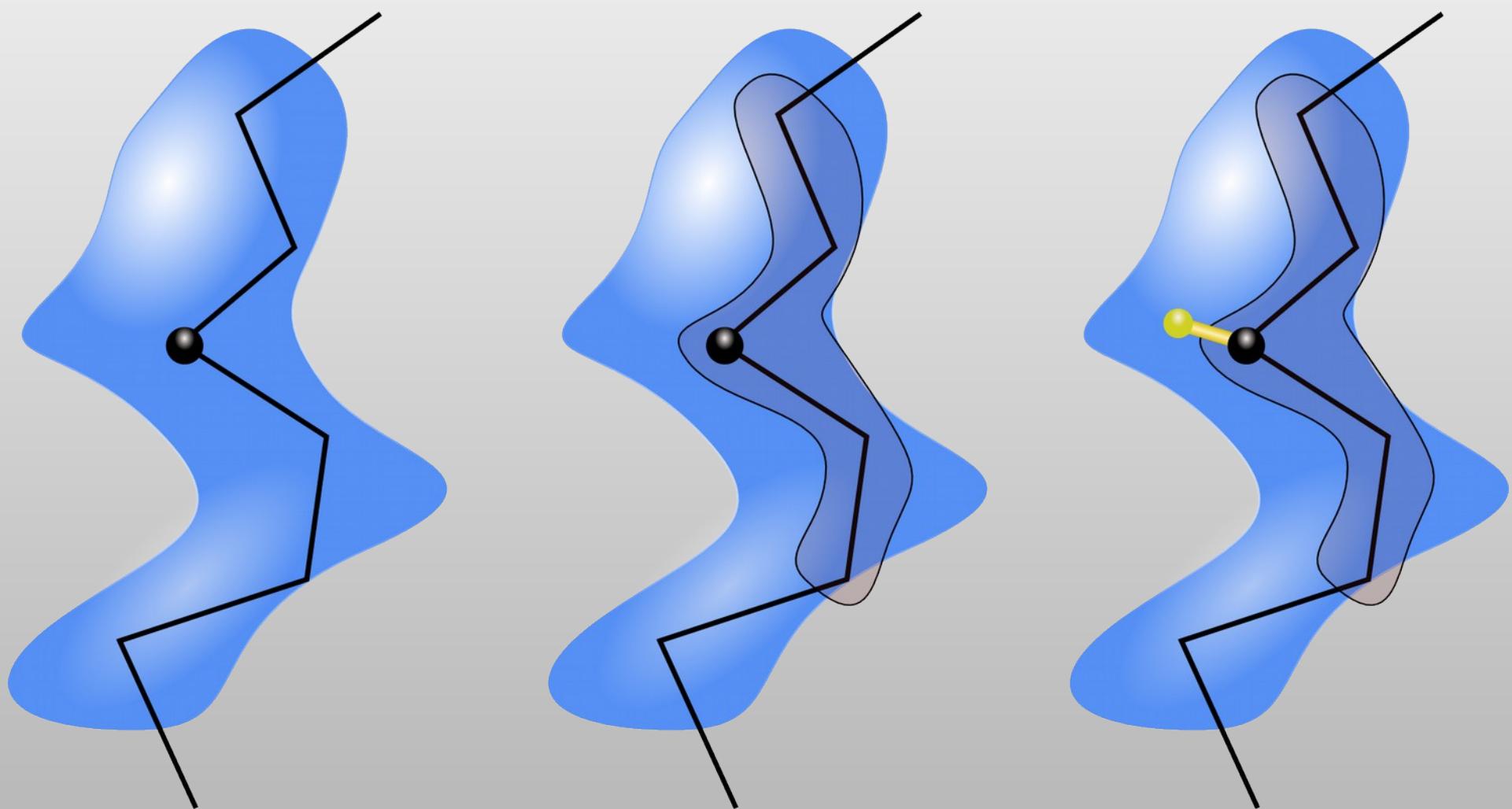
So we have our ideal RNA or homologous protein sitting roughly in the density

(not a great fit)

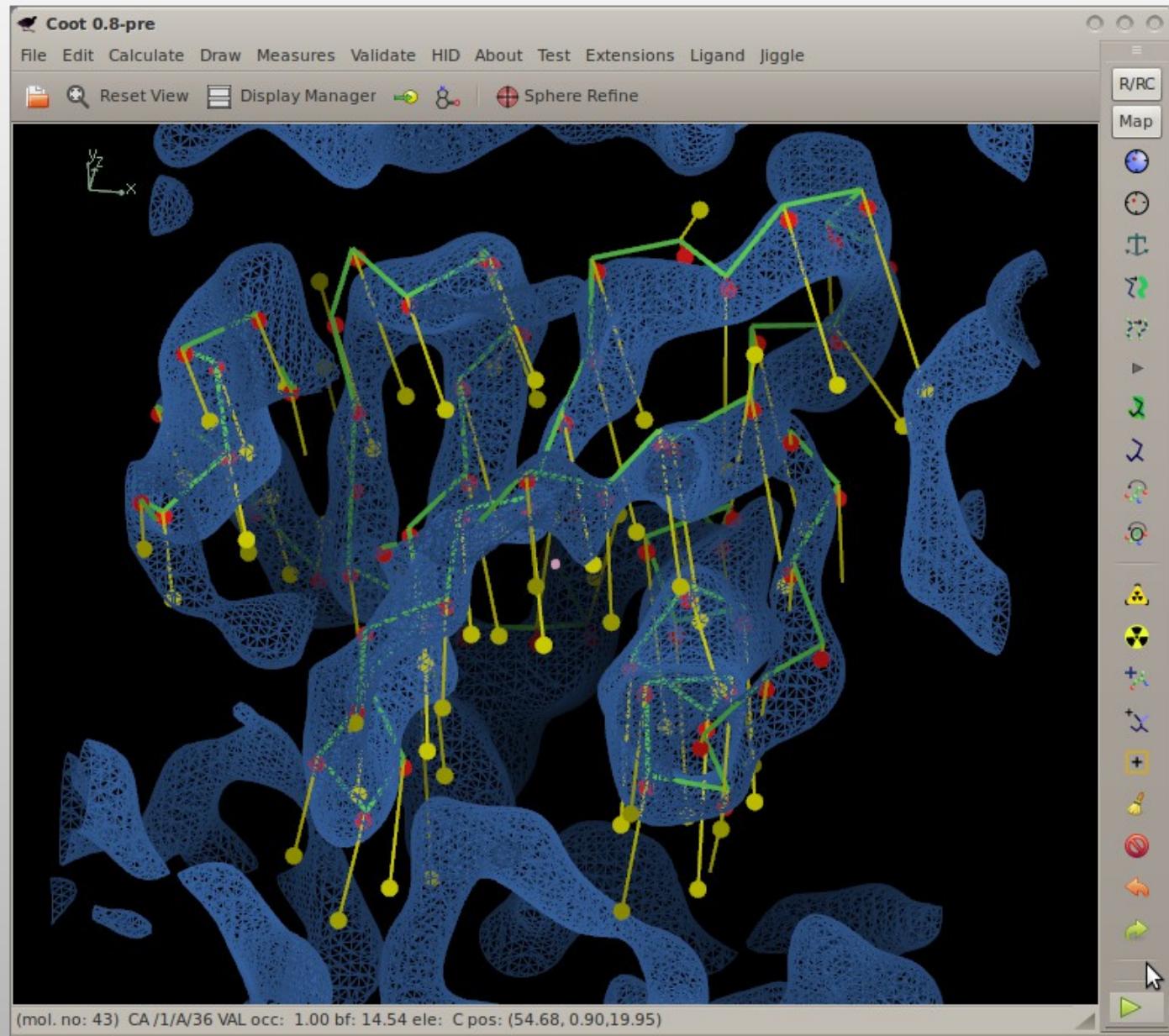
# Model Morphing: How it Works

- 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

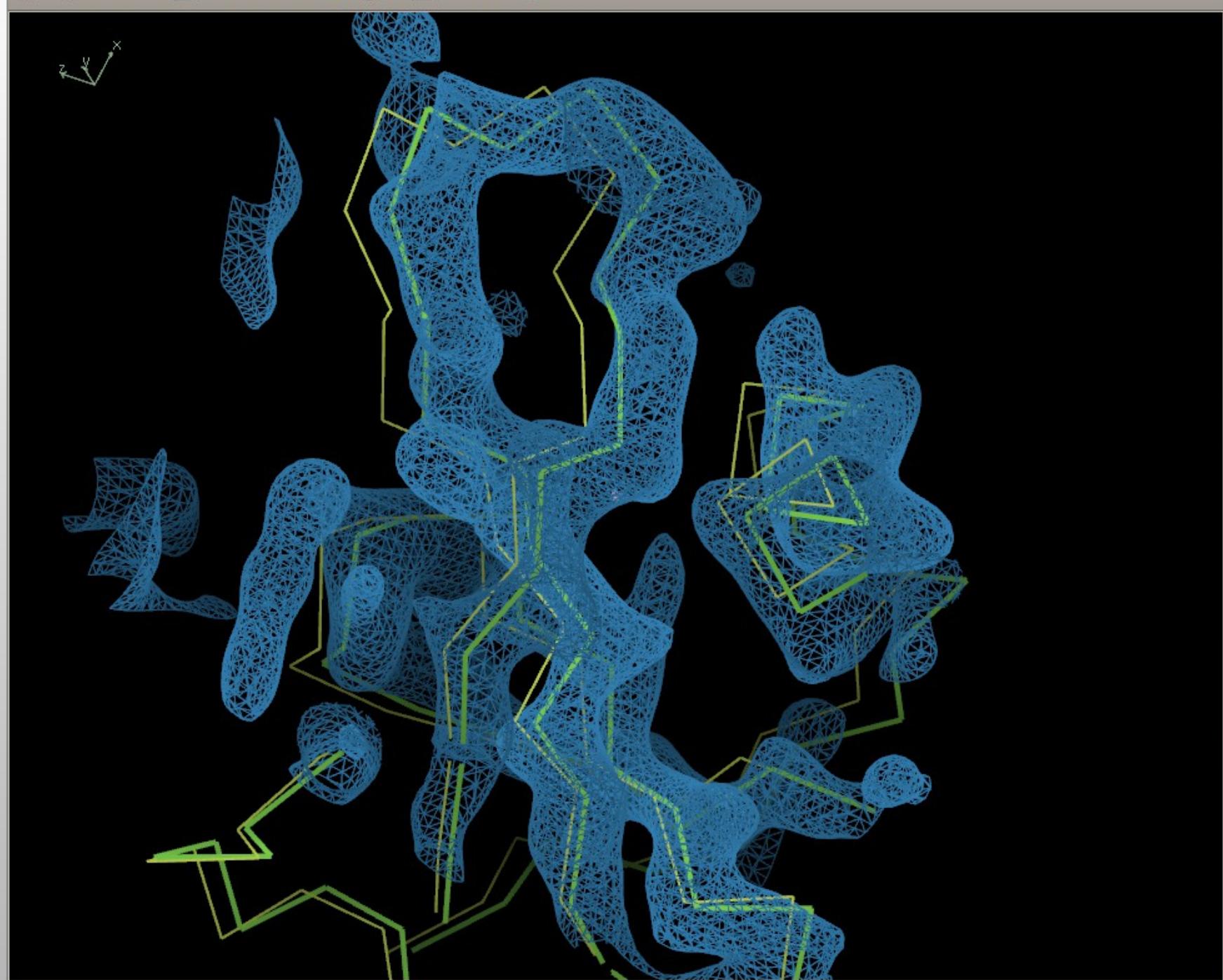
- What are the residues in the environment of a residue?
  - What are their RTs?
  - Create a metric 'distance', sort on that
  - Discard the top and bottom 25%
  - Use remaining RTs to generate average
  - ...which is then applied to central residue
- Repeat for all residues
- Larger environment radii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



Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.

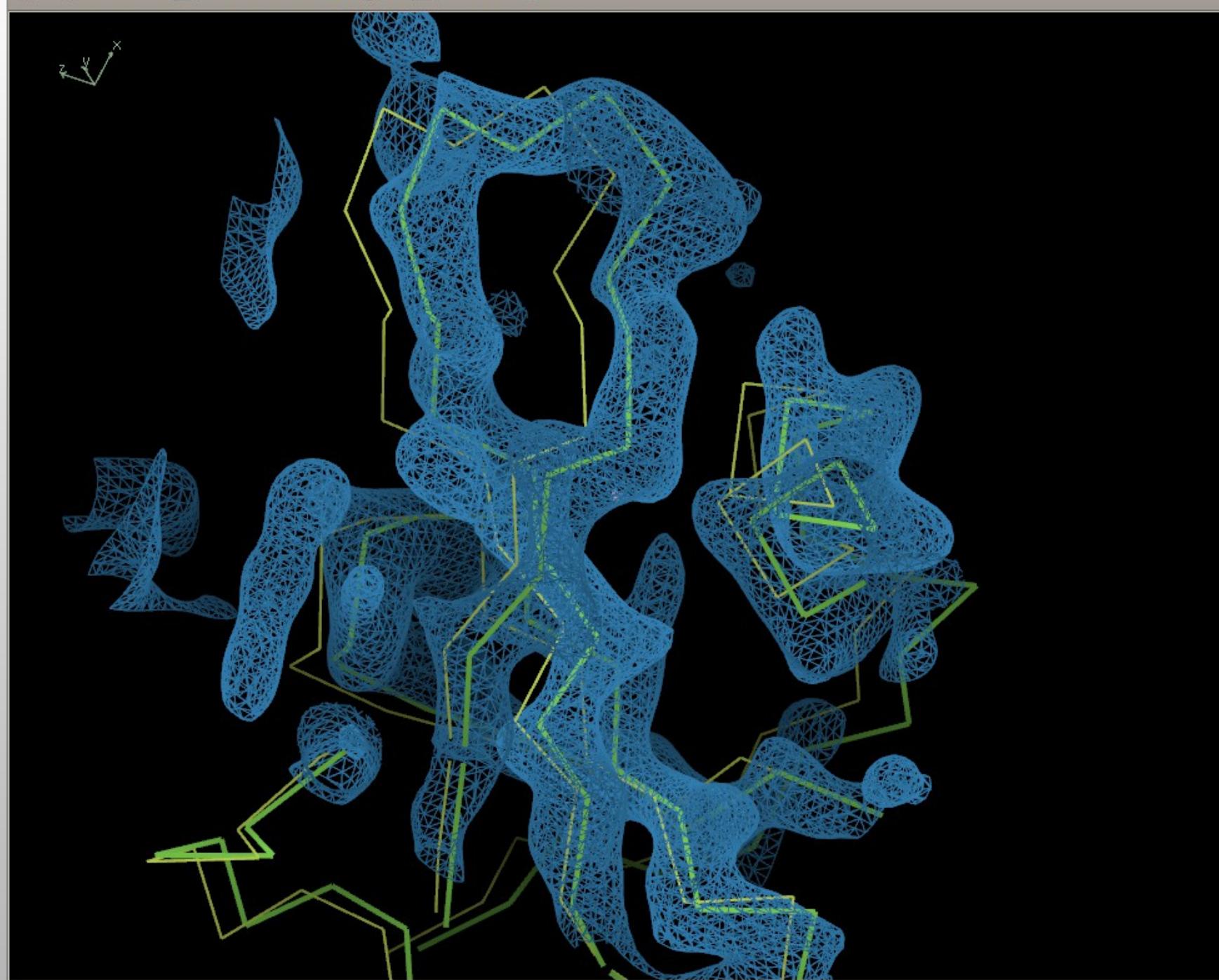
Coot 0.8-pre (revision 4826)

File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph

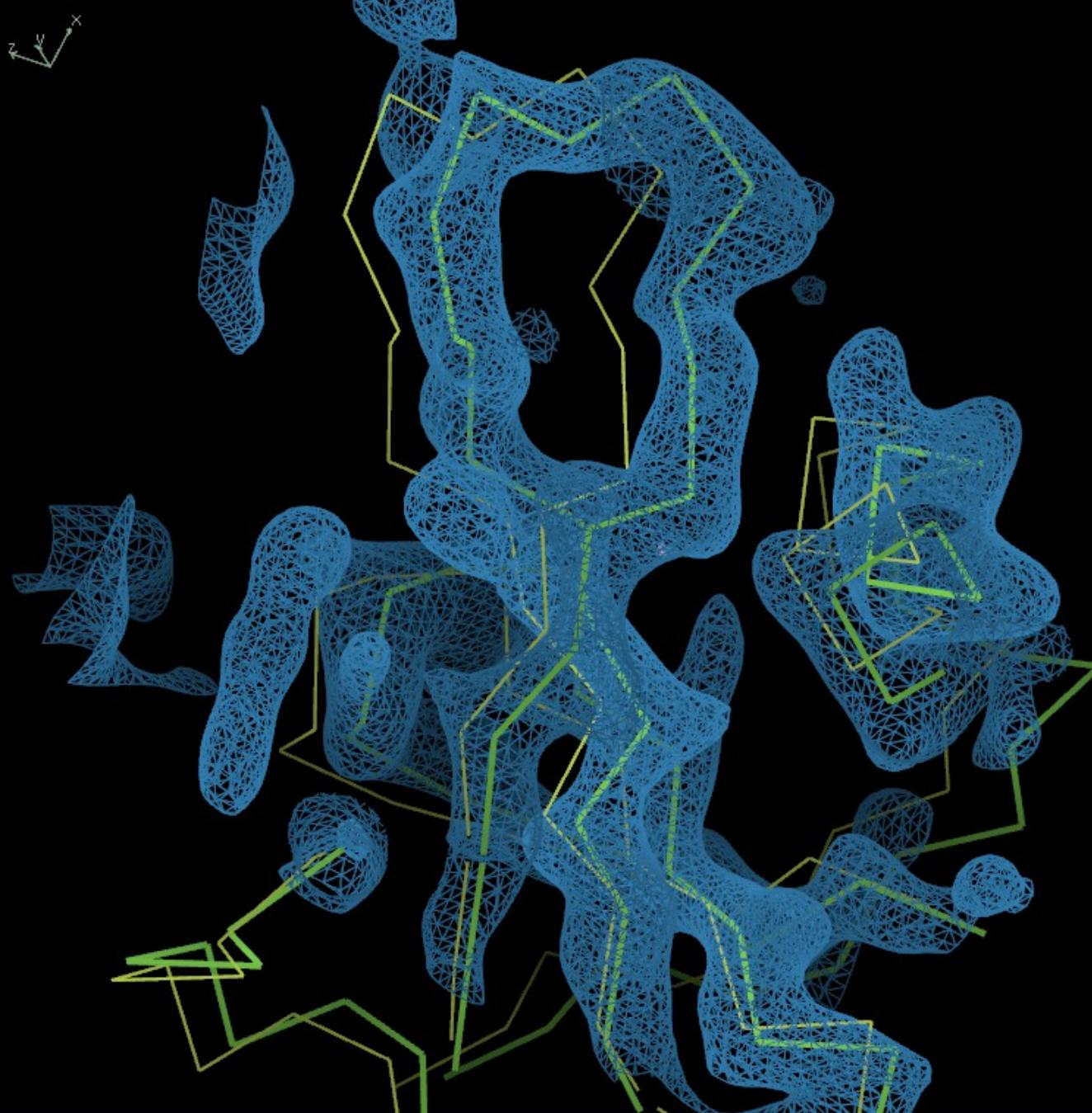
Reset View Display Manager

Full screen

Sphere Refine

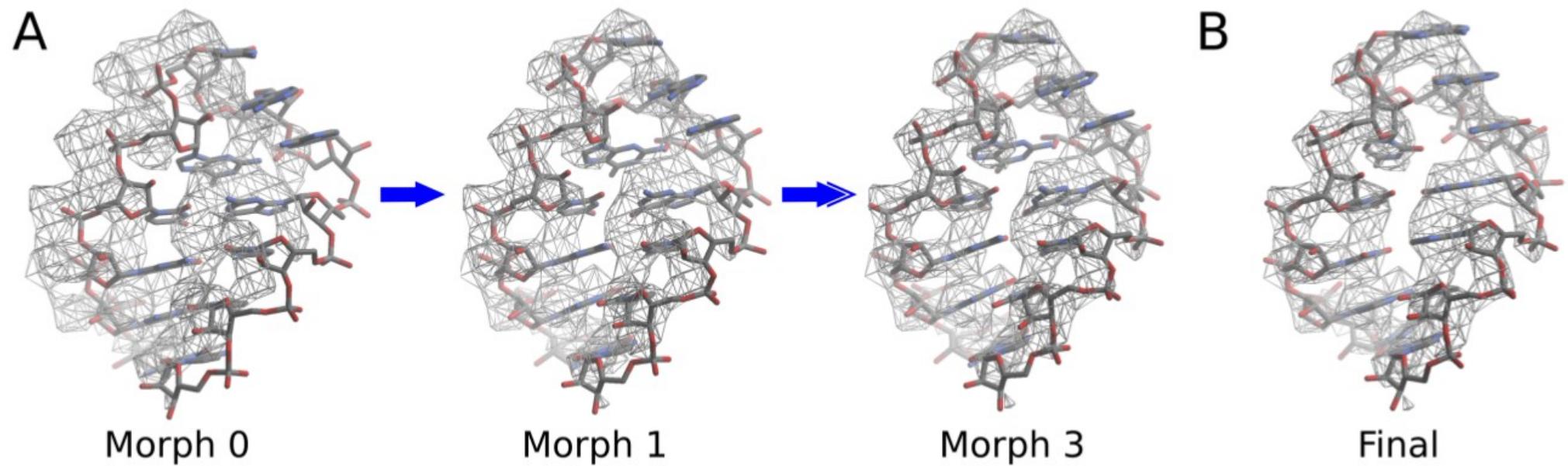


Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb. Molecule number 2 created.





# Model Morphing



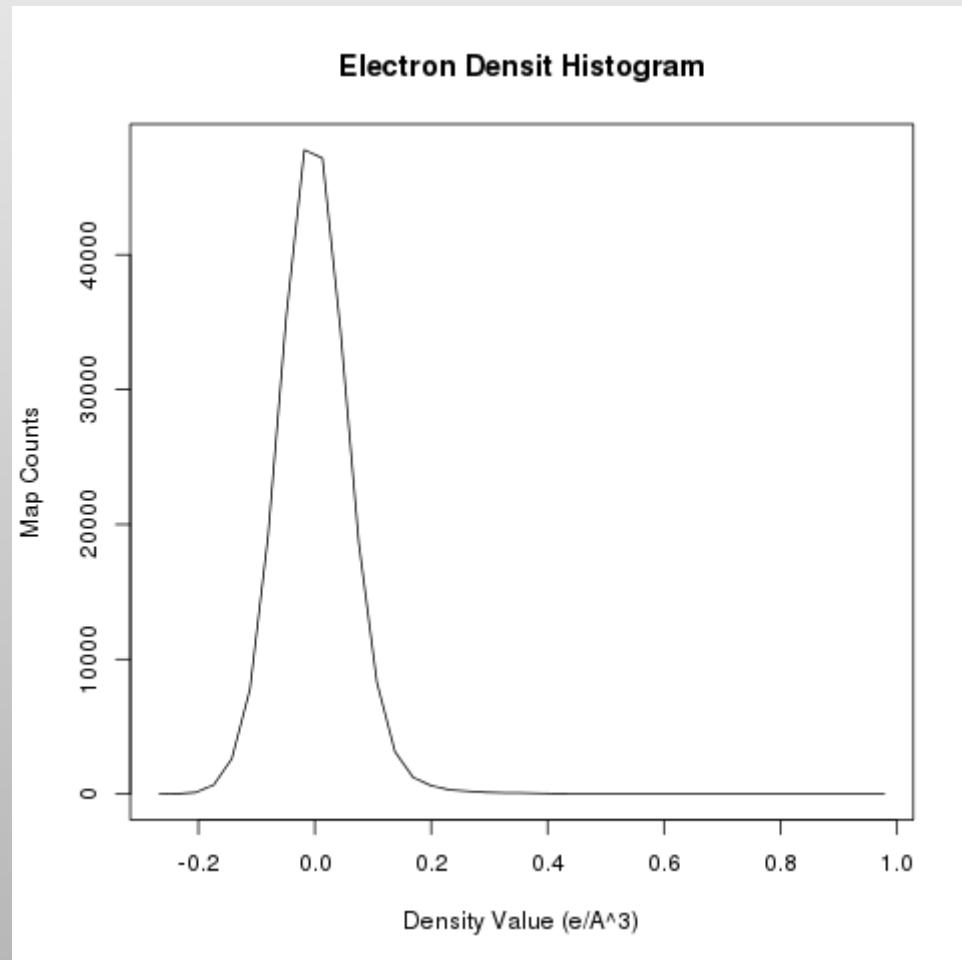
# Helix Fitting

- 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

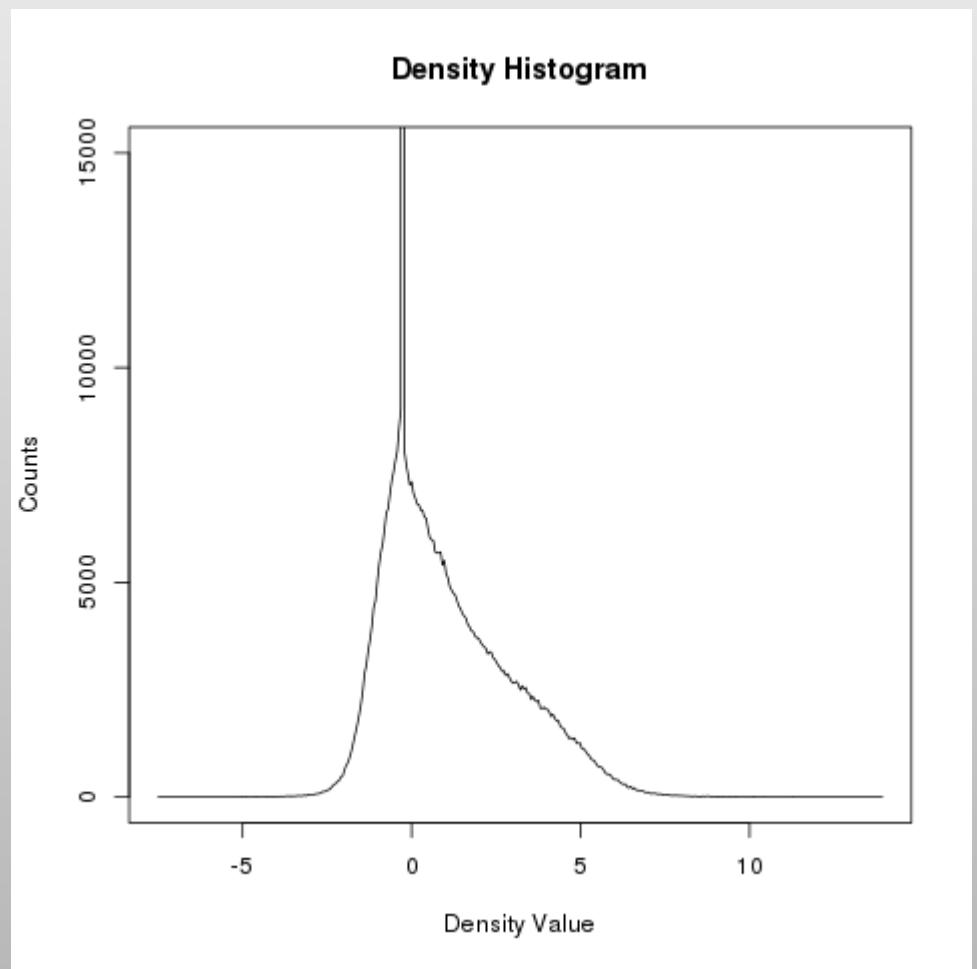
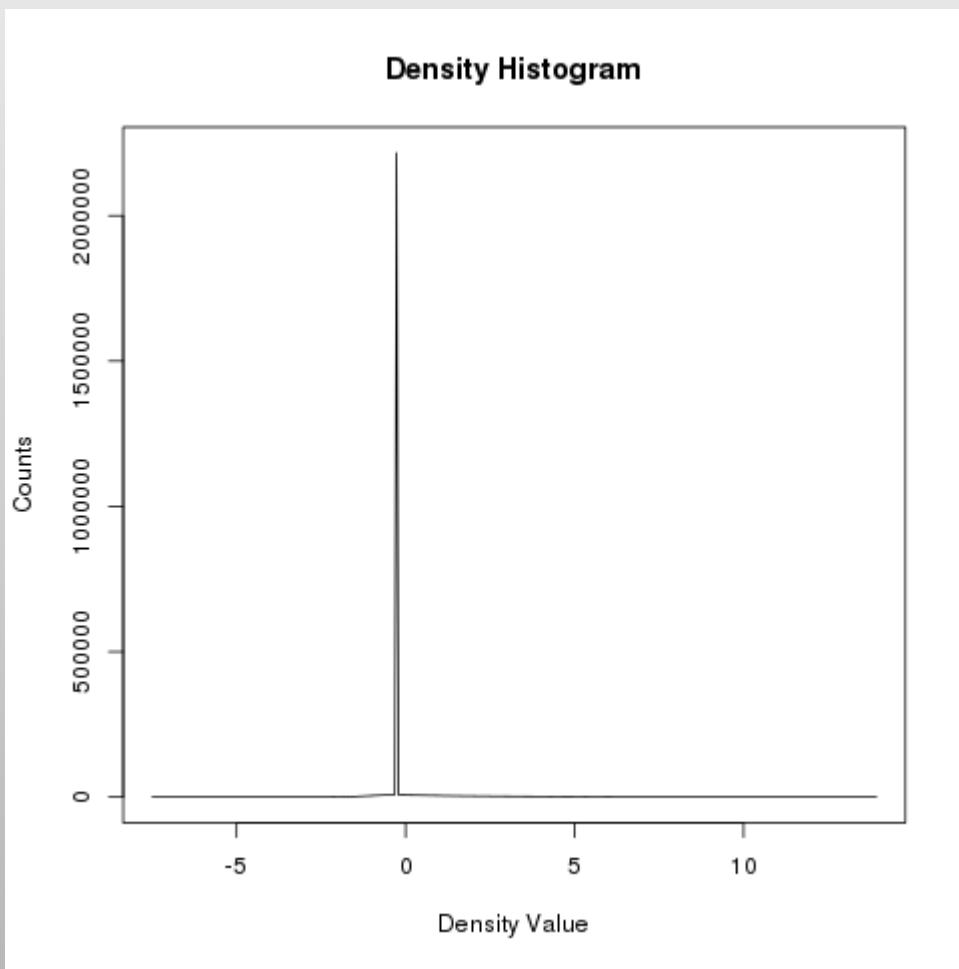
- The distribution of electron density is quite unlike that of x-ray maps

Typical Density Histogram  
from an  
X-ray map



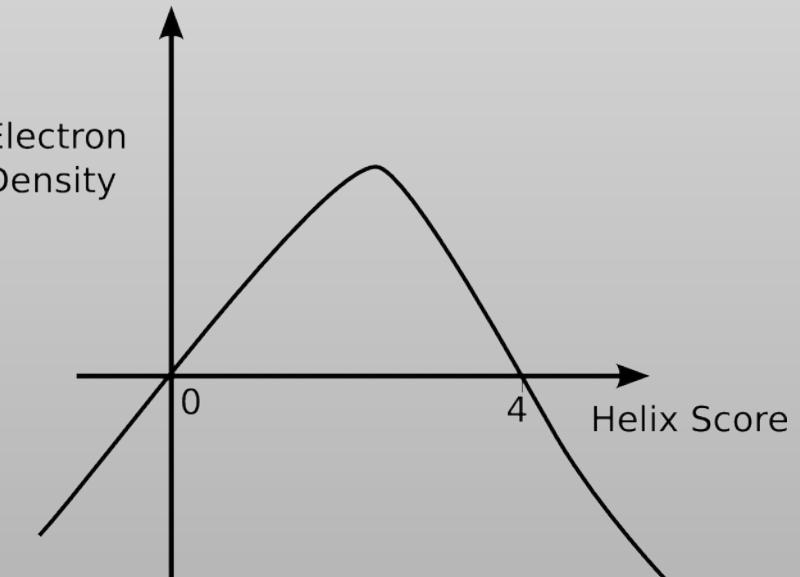
# Helix Fitting

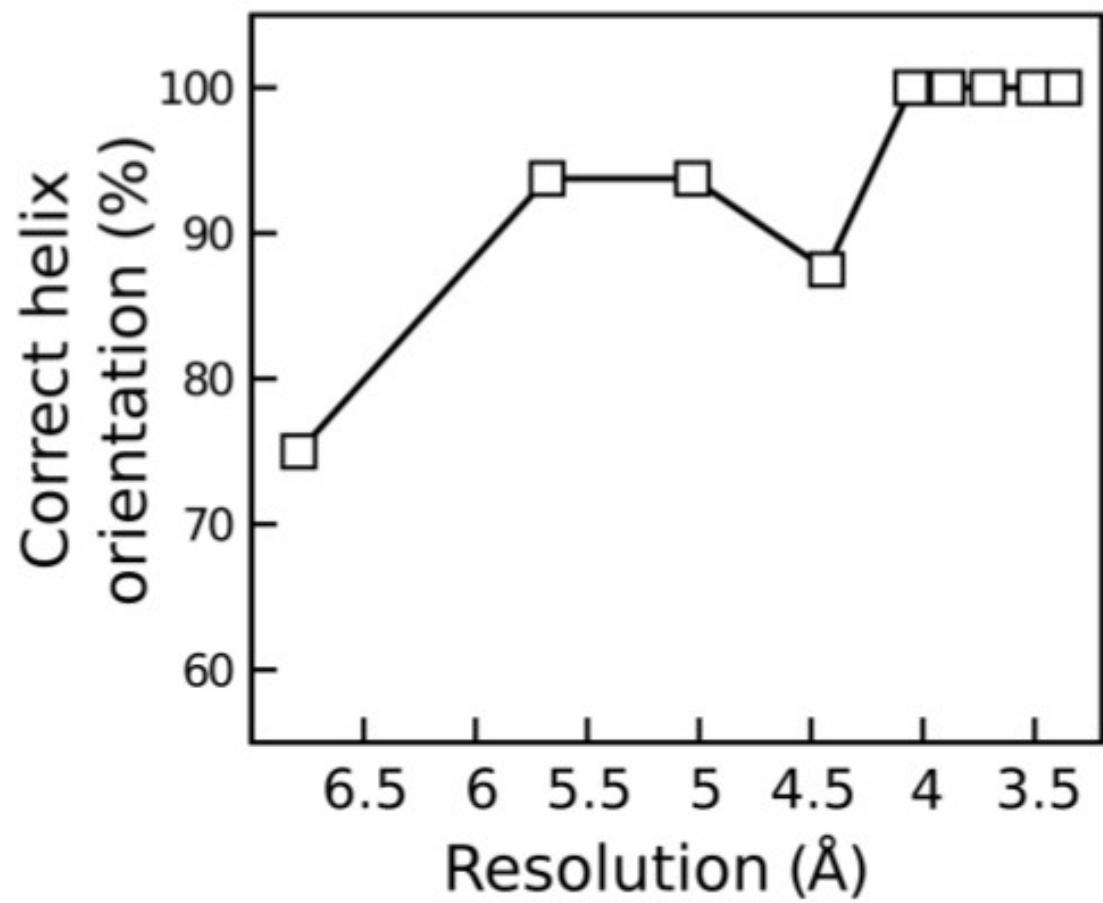
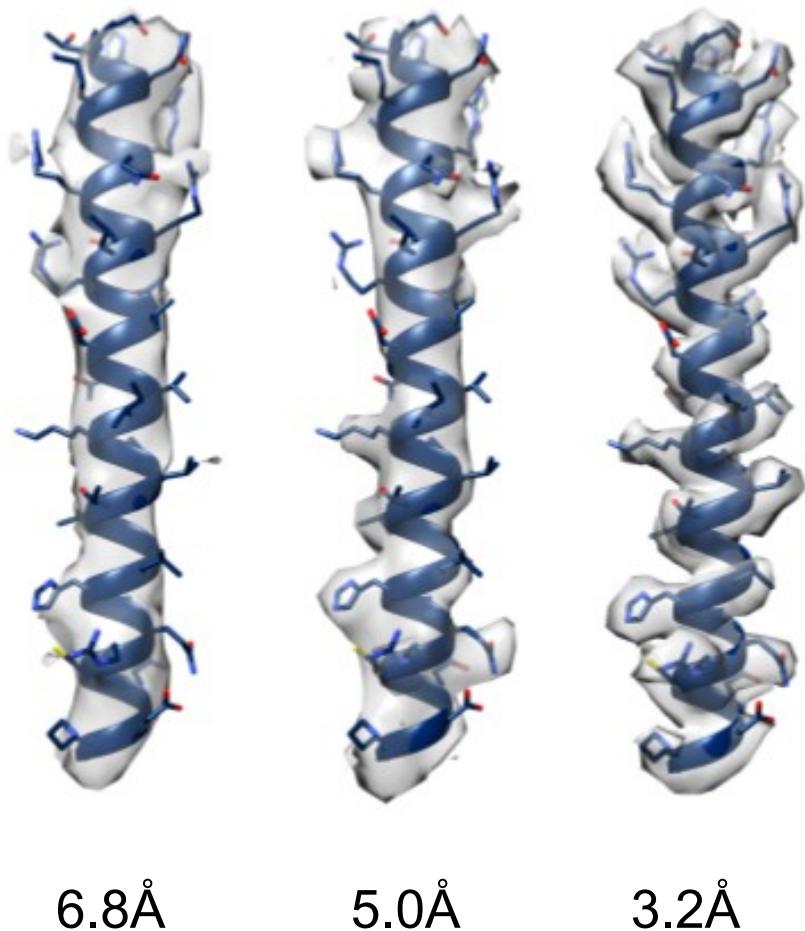
- The distribution of electron density is quite unlike that of x-ray maps



# 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



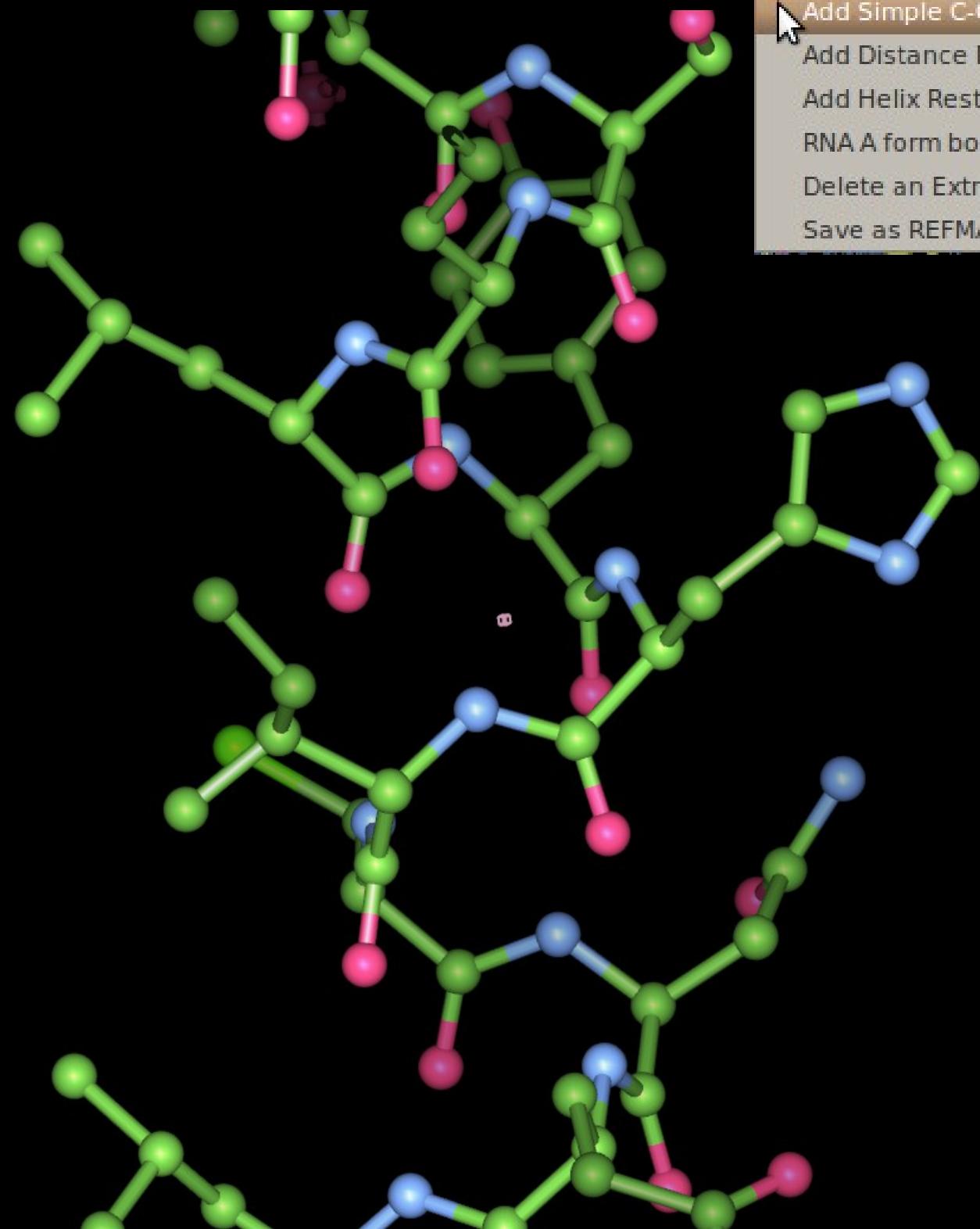


# Additional Restraints

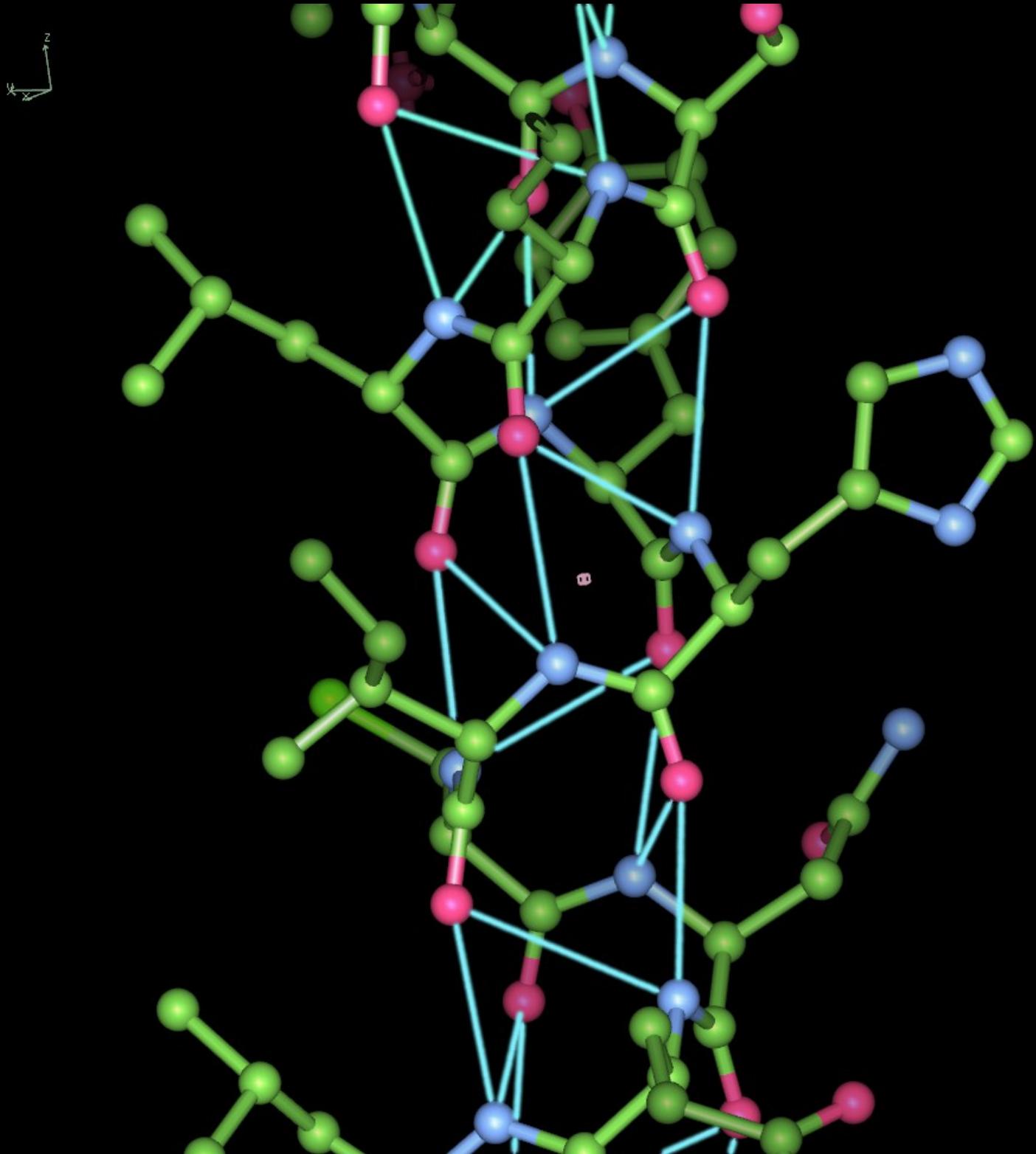
# Restraints Editing in Coot

- 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

*x*  
*y*  
*z*

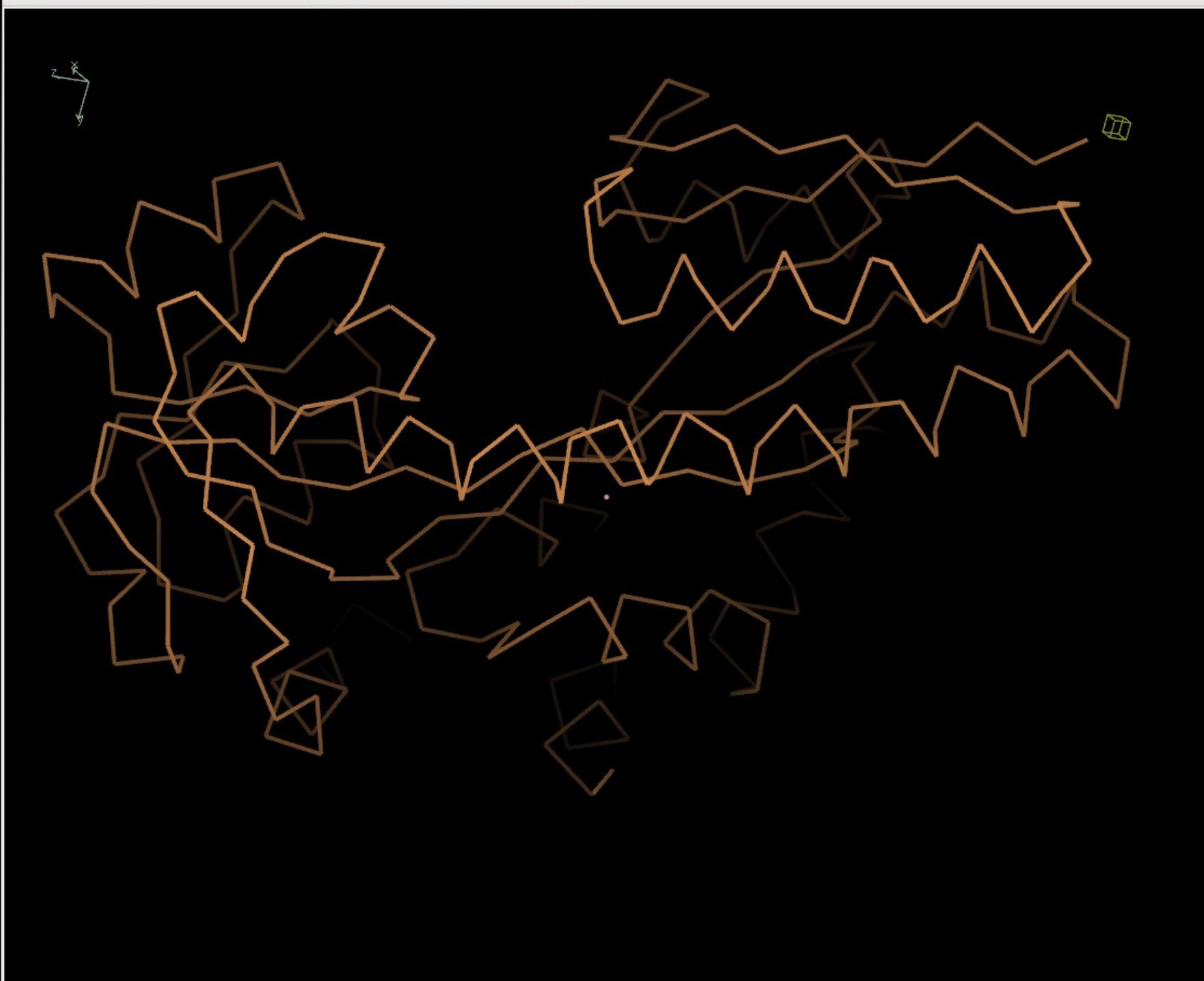


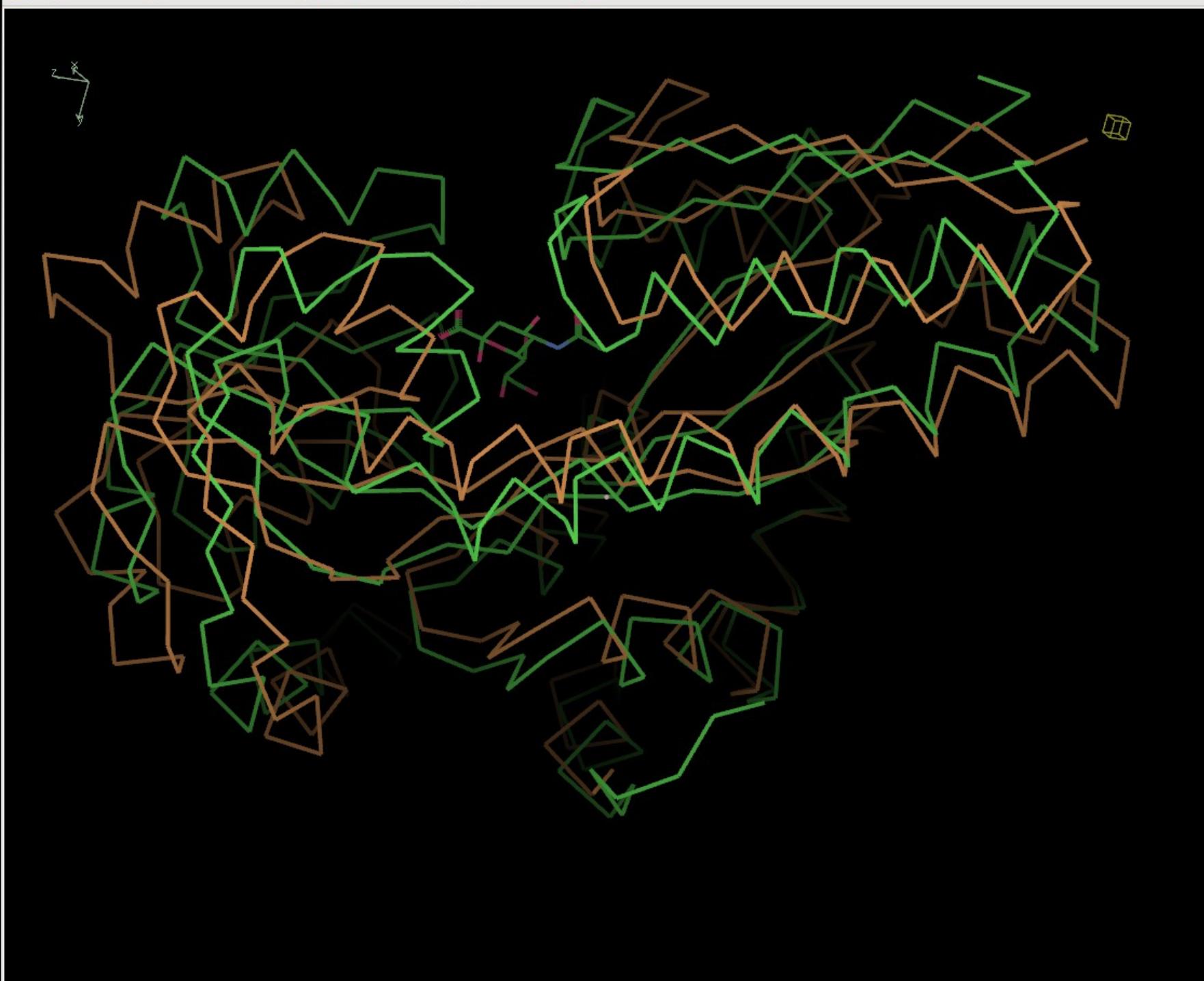
- Add Simple C-C Single Bond Restraint...
- Add Distance Restraint...
- Add Helix Restraints...
- RNA A form bond restraints...
- Delete an Extra Restraint...
- Save as REFMAC restraints...

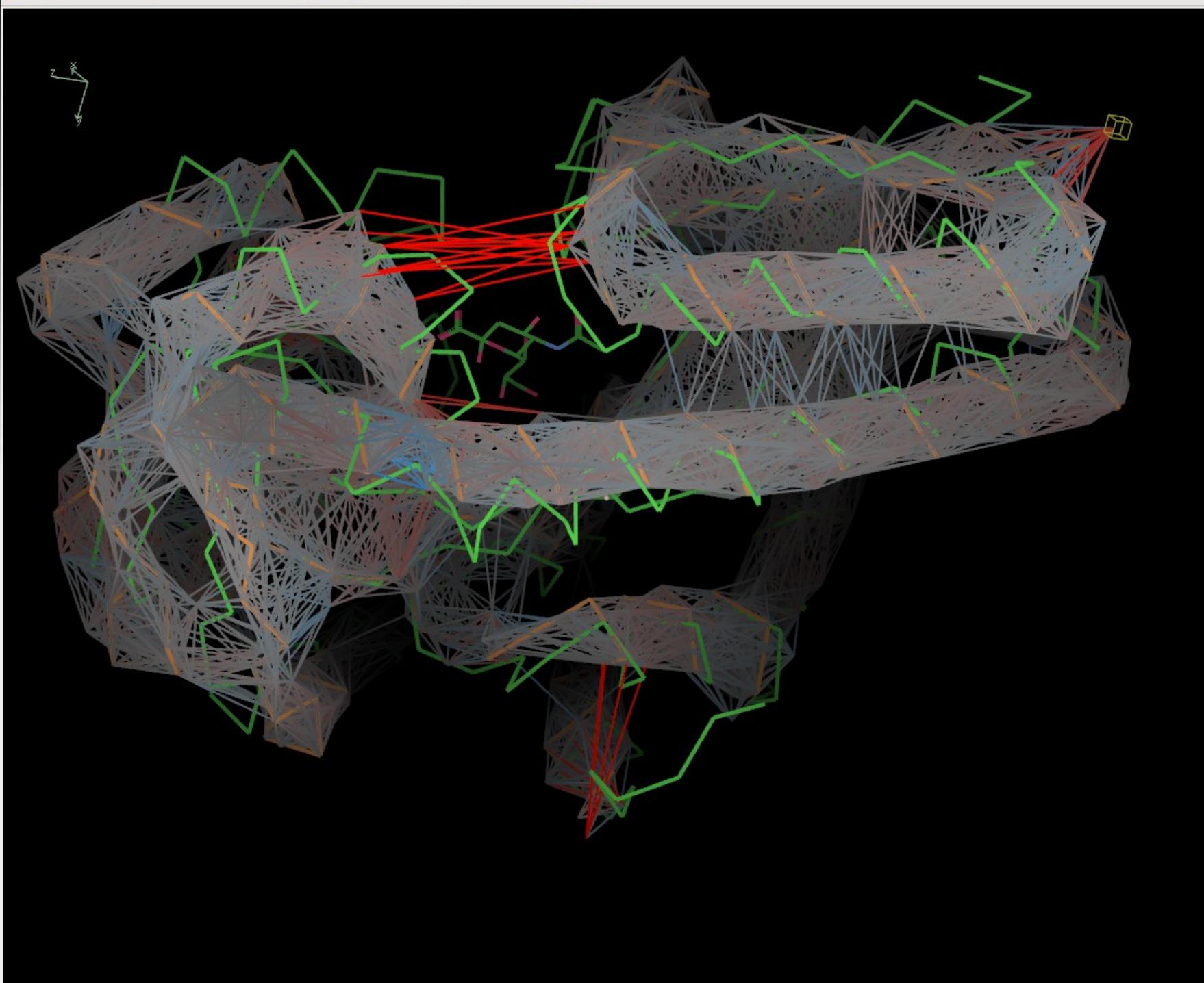


# ProSMART Interface

- 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

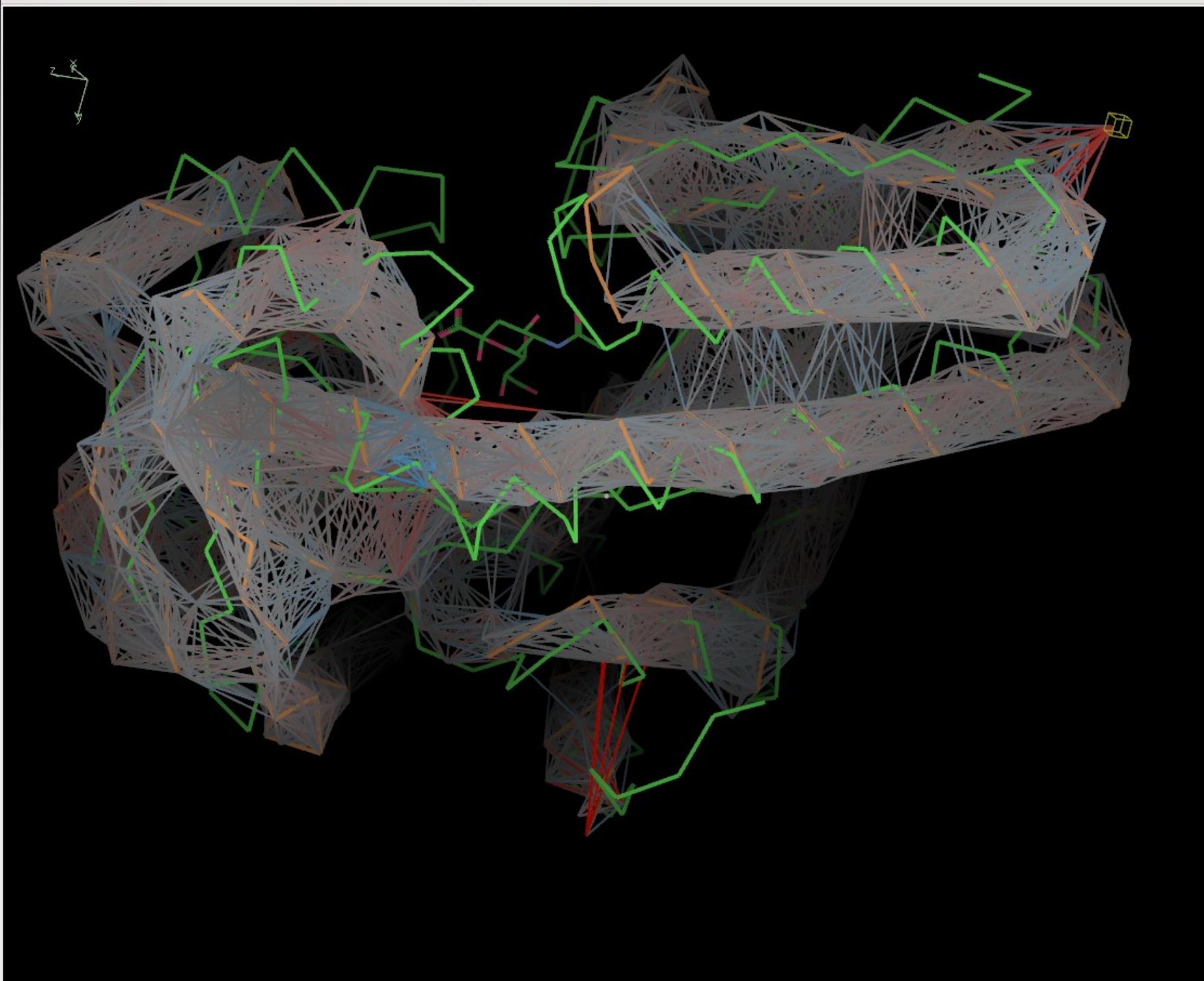
  Reset View  Display Manager

 Sphere Refine  Sphere Refine +  B

Build NA

R/RC

Map

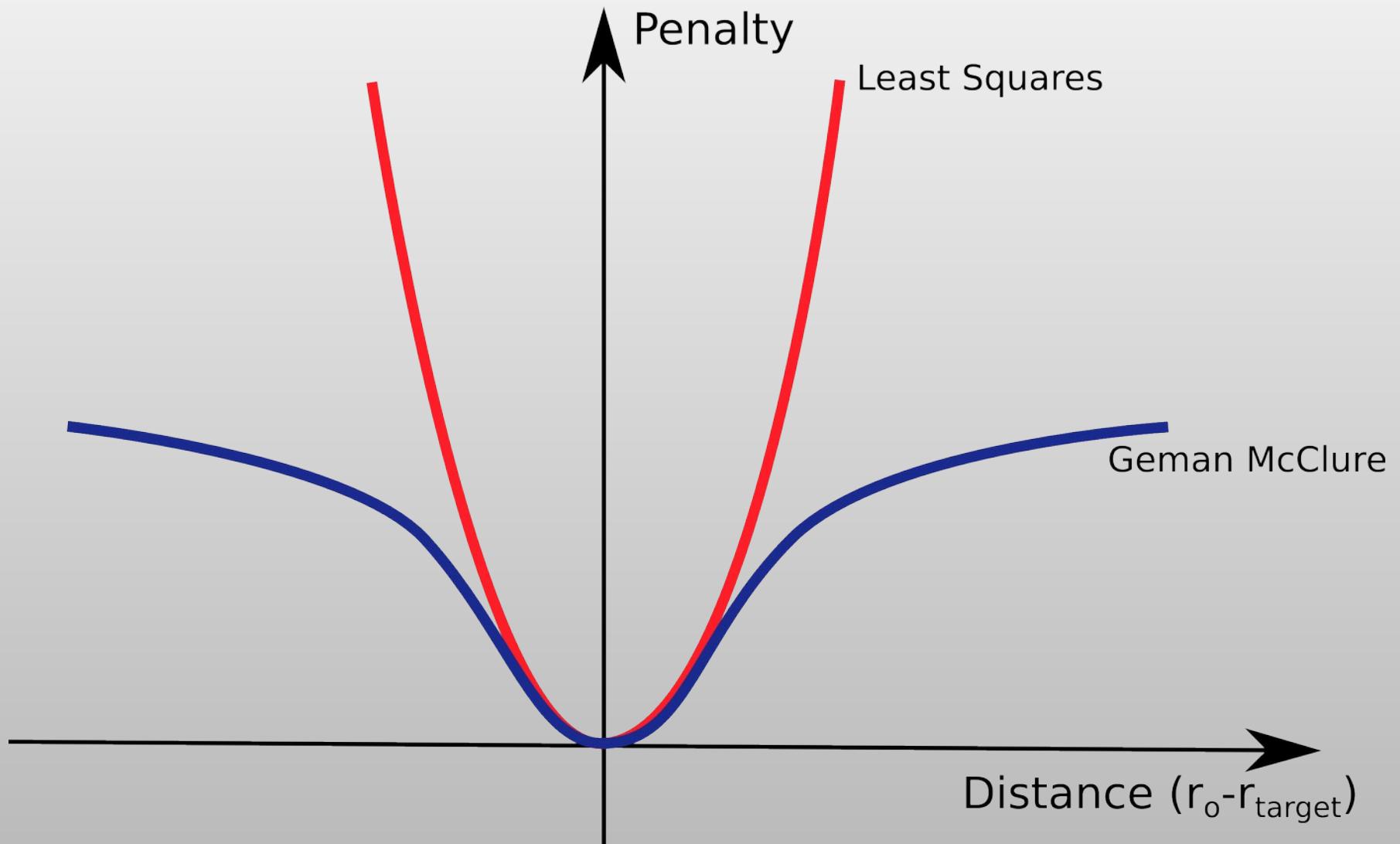


(mol. no: 5) CA /1/A/15 SER occ: 1.00 bf: 34.58 ele: C pos: (18.77,17.11,19.07)

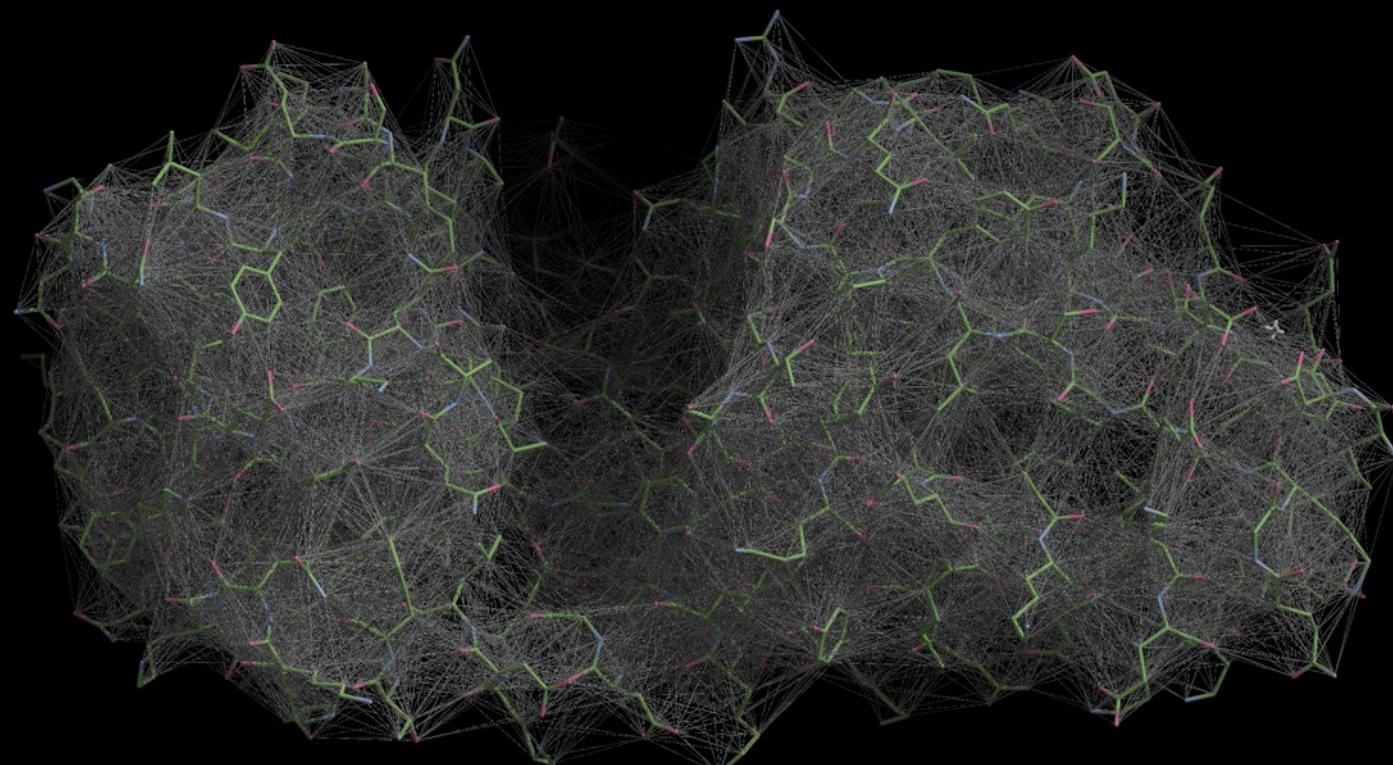
# ProSMART integration

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

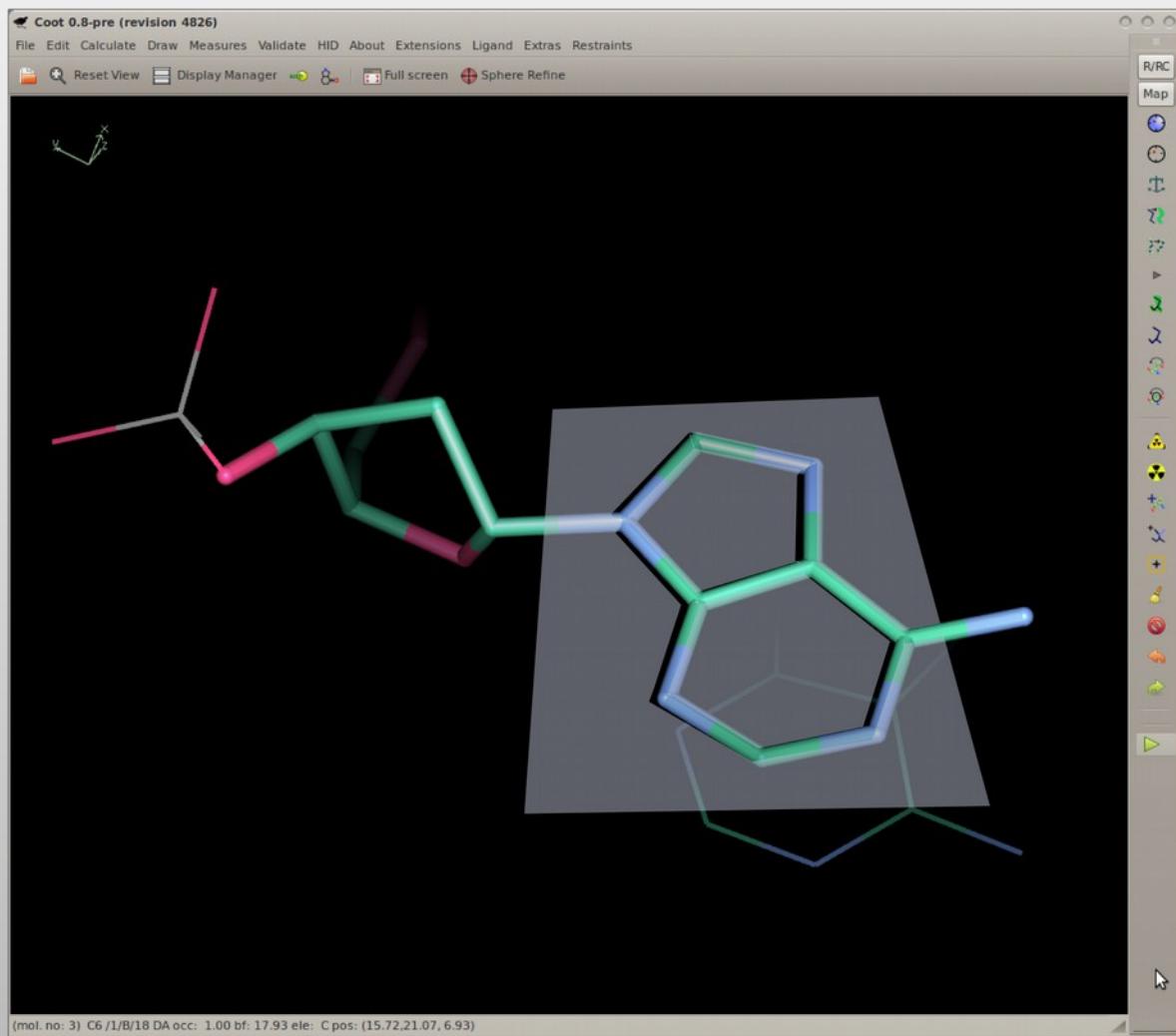
# Modified Target Function



# ProSMART Restraints

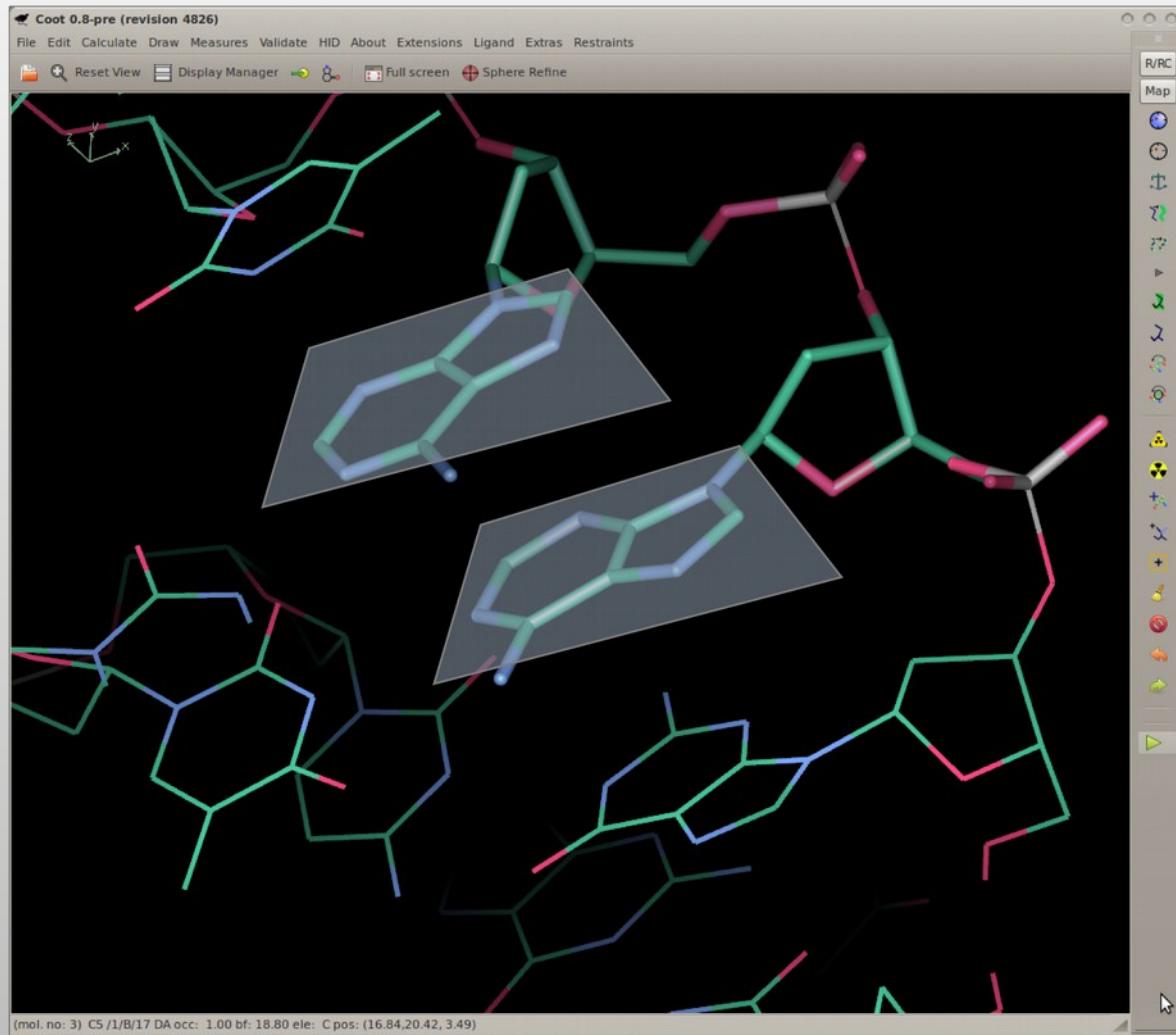


# Plane Restraints



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

# Parallel Planes Restraints



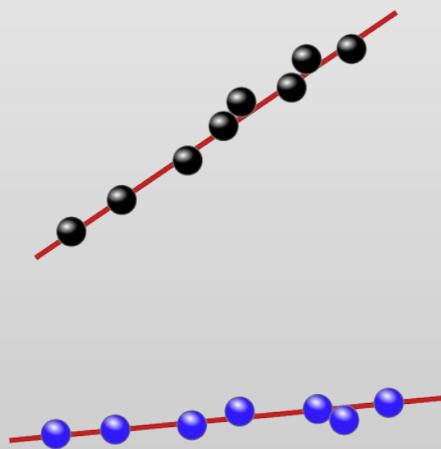
$$S = (a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2$$

Not easy to use in Coot

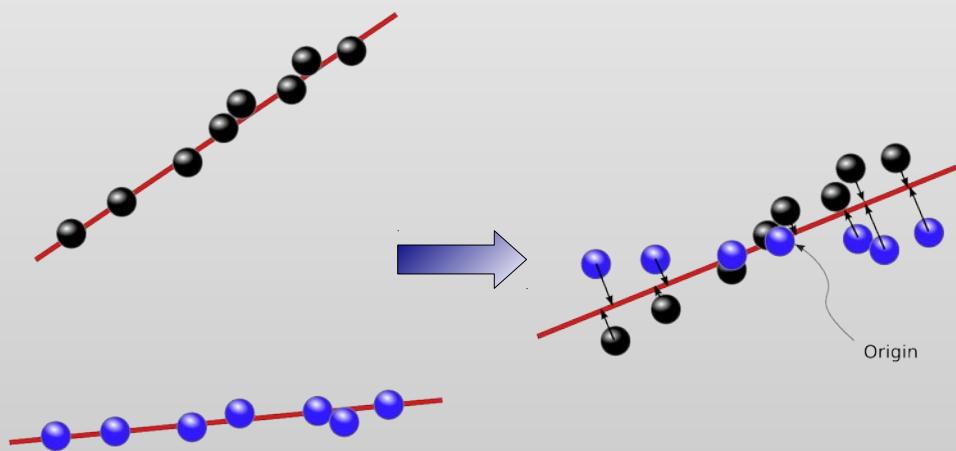
# Parallel Planes Restraints

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

# Parallel Planes Restraints

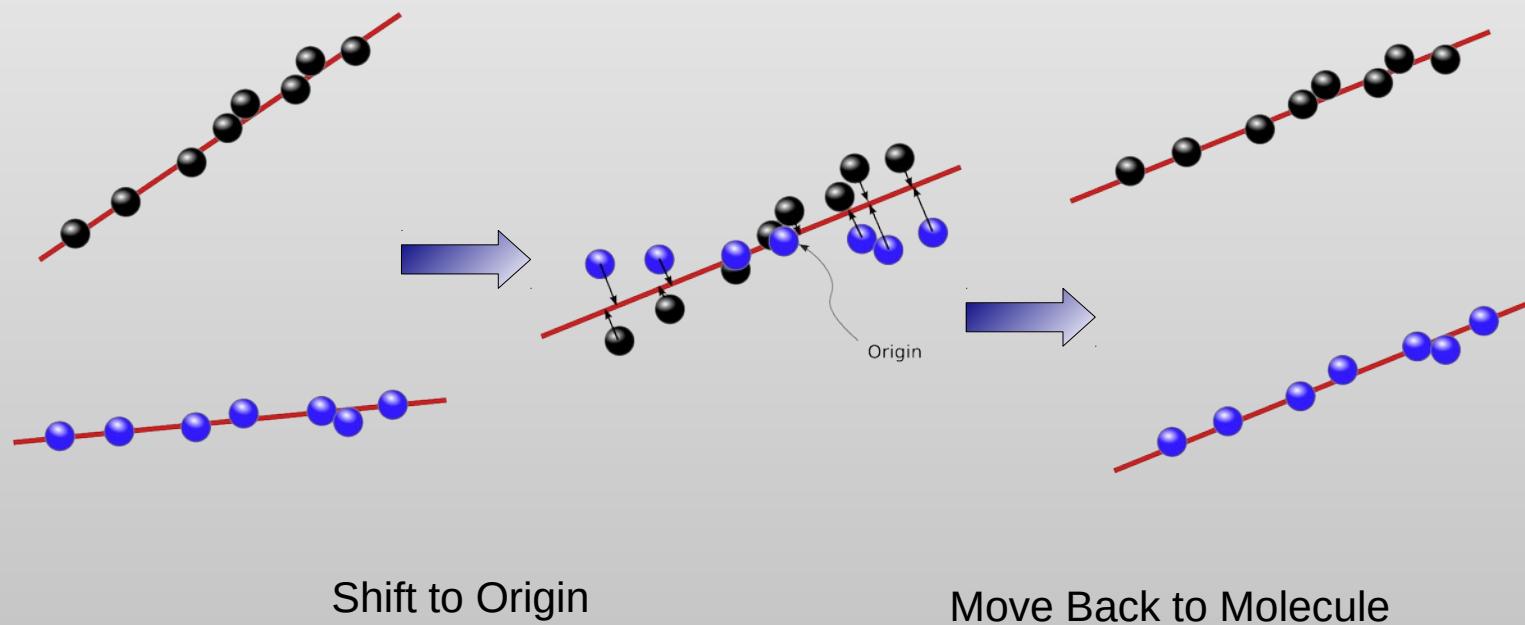


# Parallel Plane Restraints



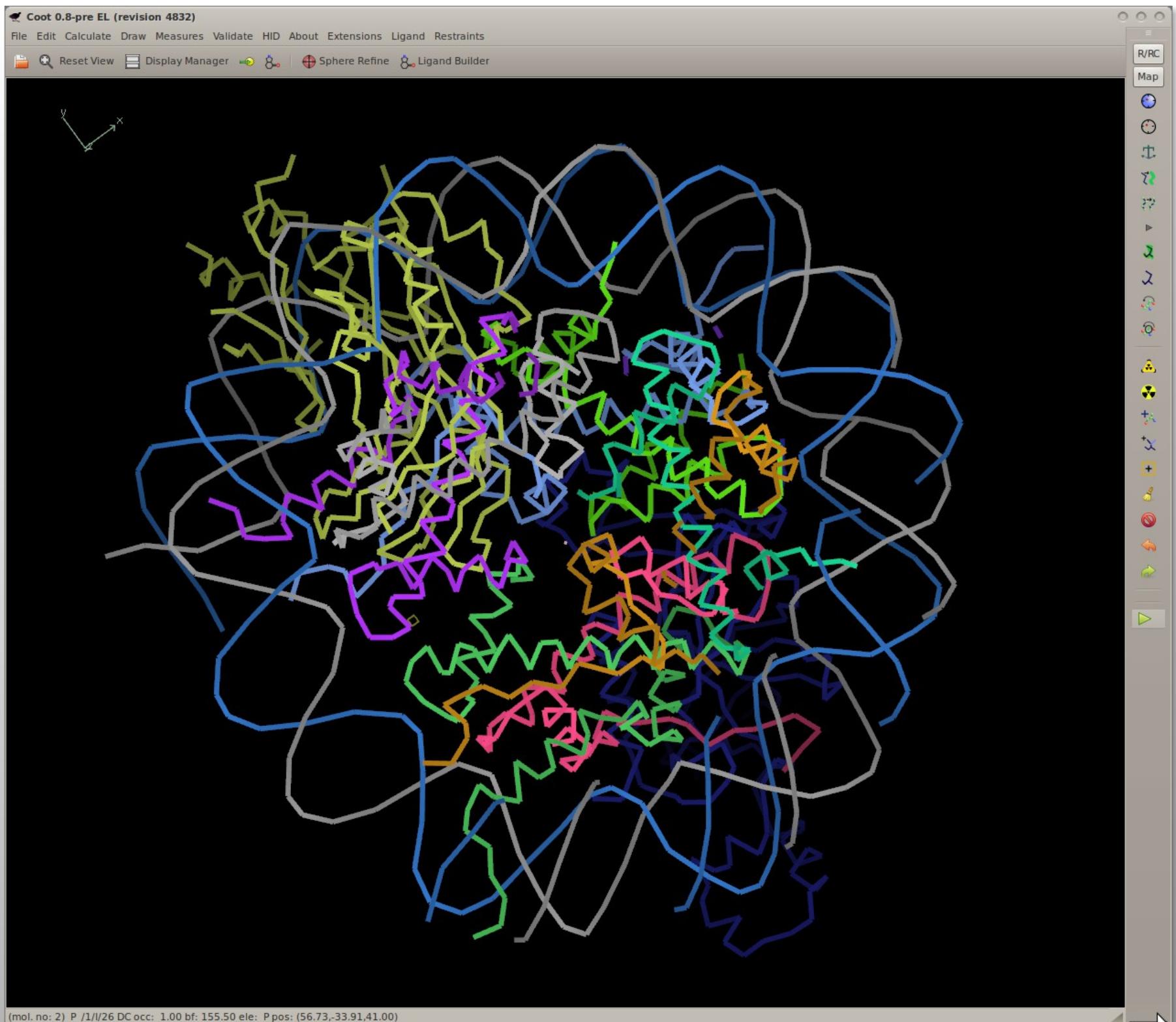
Shift to Origin

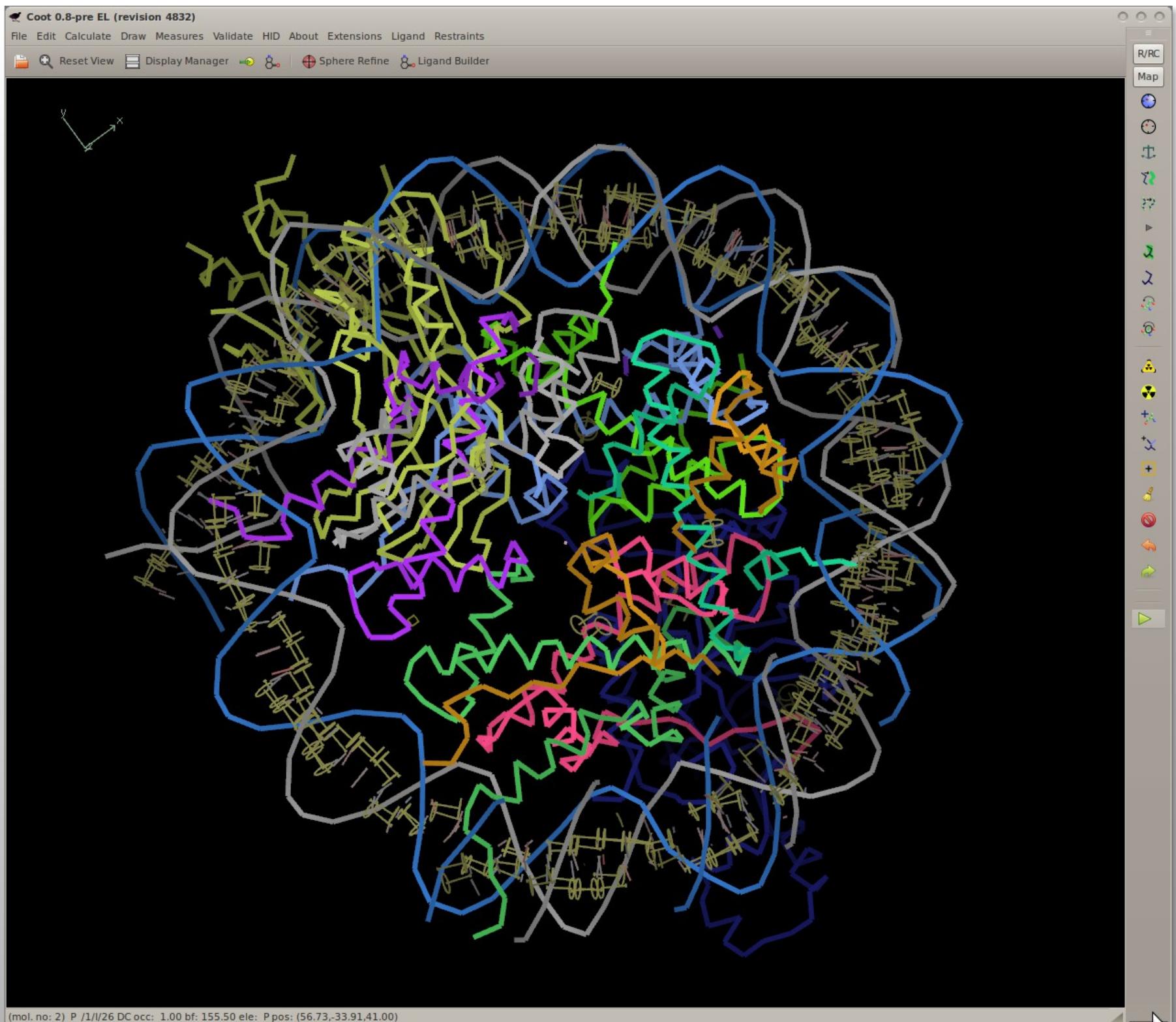
# Parallel Planes Restraints



# Automatic Generation of Base Pairing and Stacking Restraints

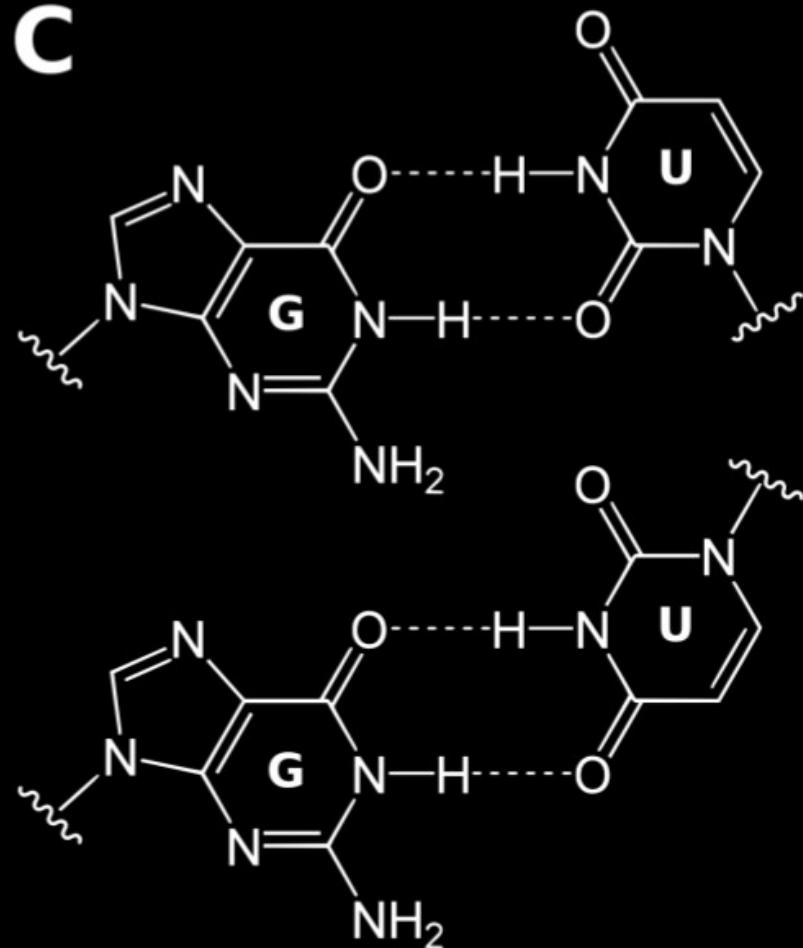
- Fei Long's `libg_d`
  - Provide it with a model and it writes out Refmac restraint descriptions
  - ... which *Coot* can also read
  - *Coot* can also create user-defined base-pairing and stacking restraints



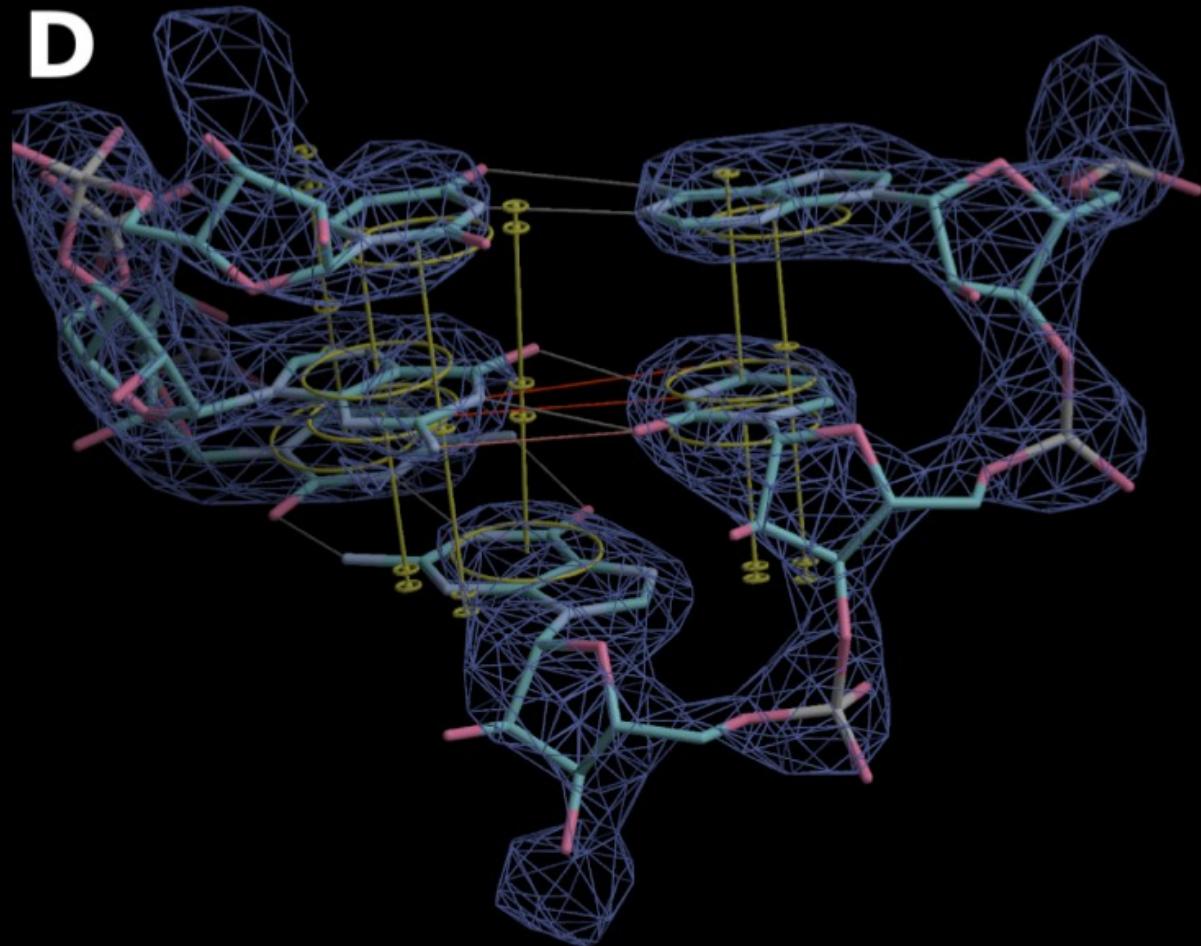


# Libg restraints

C



D



(Watson Crick and)  
Wobble, Reverse Wobble

Representation in Coot

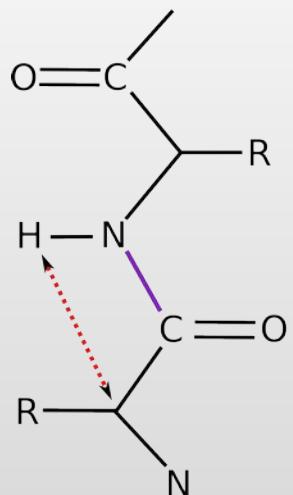
# cis-Peptides

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

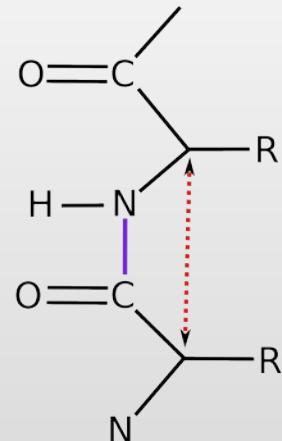
# cis-Peptides

- 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). **D71**, 706-709
  - Touw *et al.*: Detection of trans–cis flips and peptide-plane flips in protein structures *Acta Cryst.* (2015). **D71**, 1604-71614

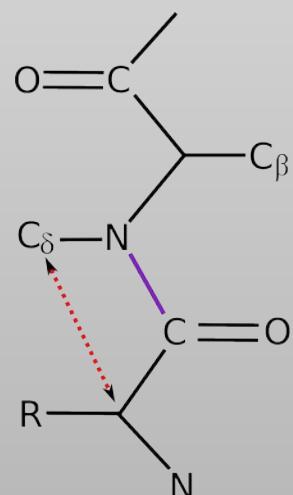
# cis-Peptides



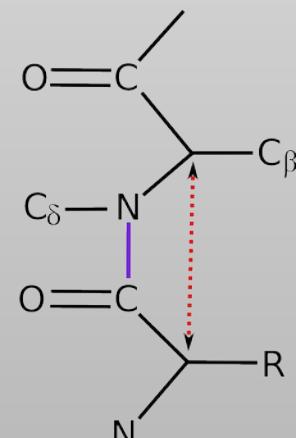
*trans-peptide*



*cis-peptide*

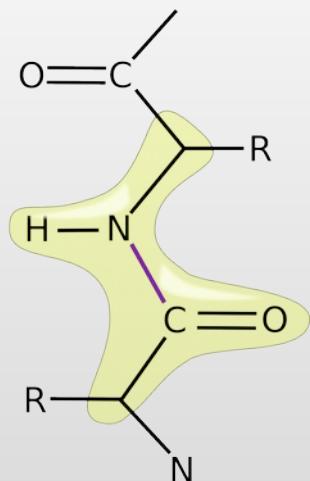


PRO *trans-peptide*

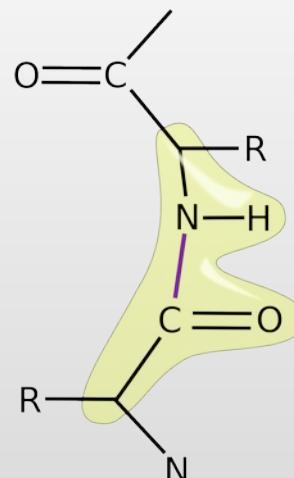


PRO *cis-peptide*

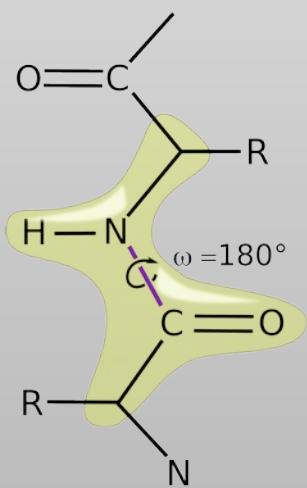
# *cis*-Peptides



*trans-peptide*  
with plane restraints

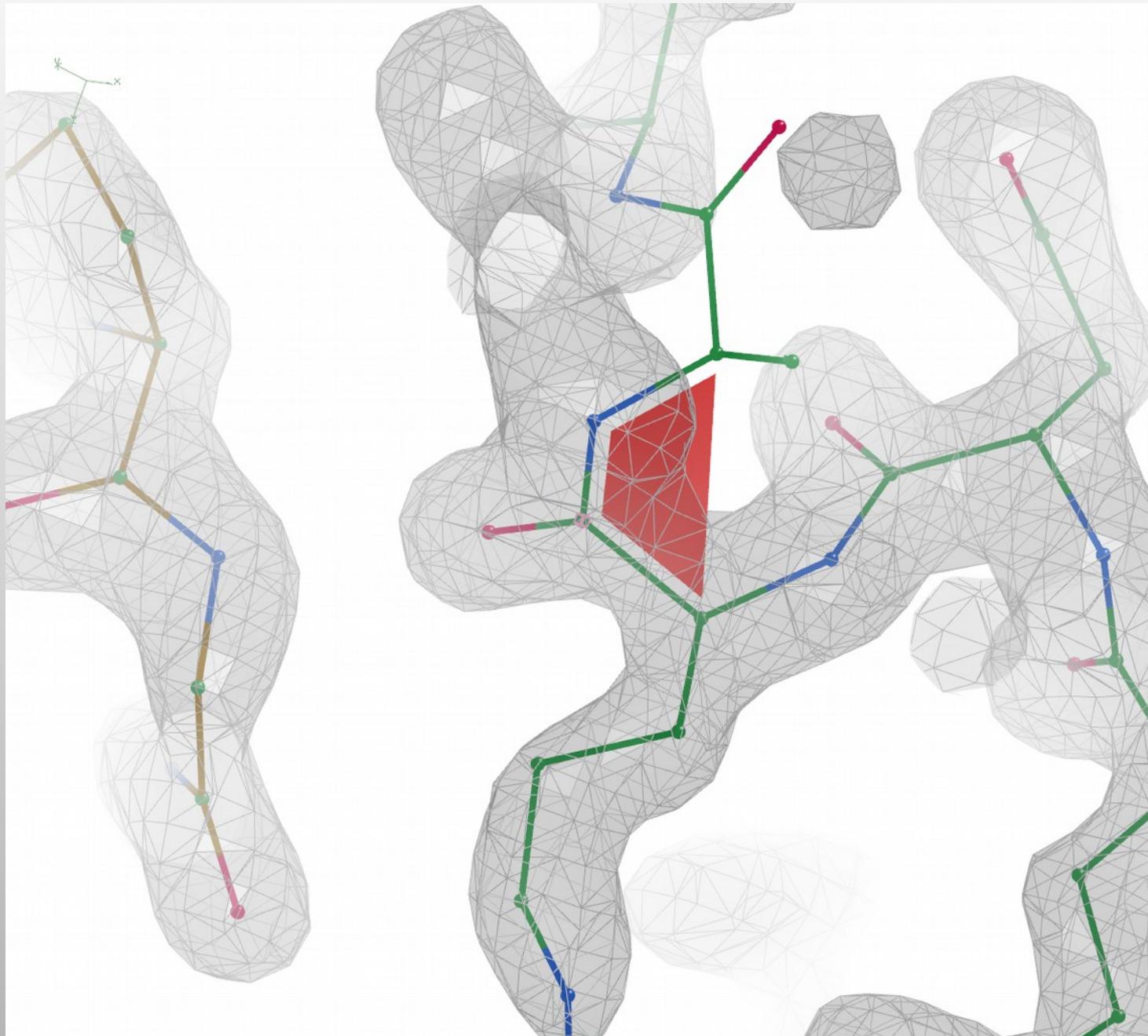


*cis-peptide*  
with plane restraints

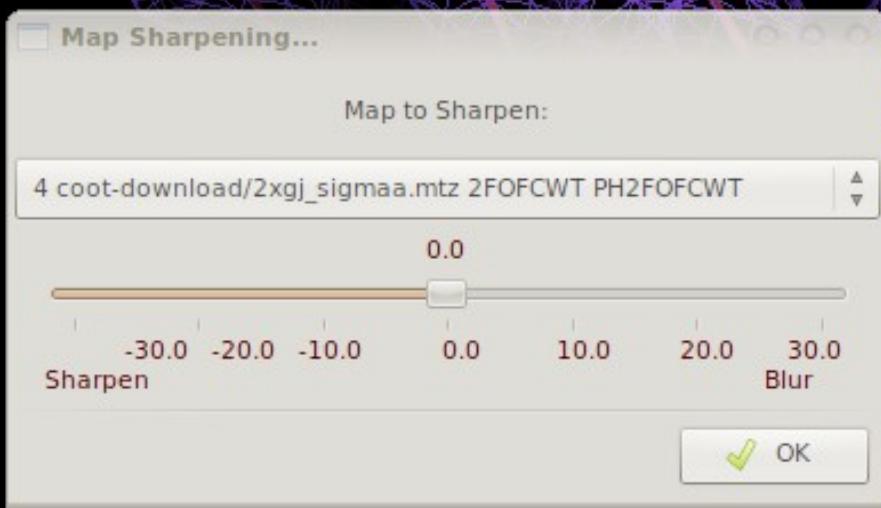
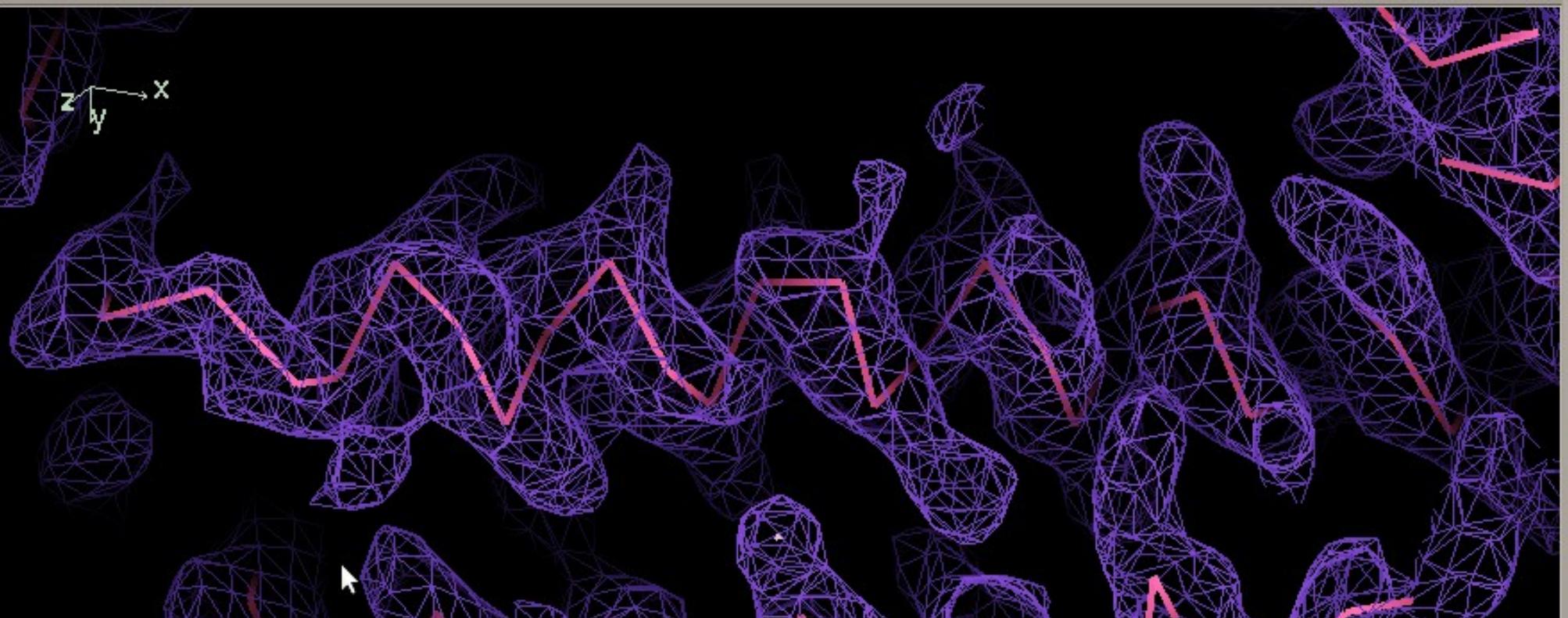


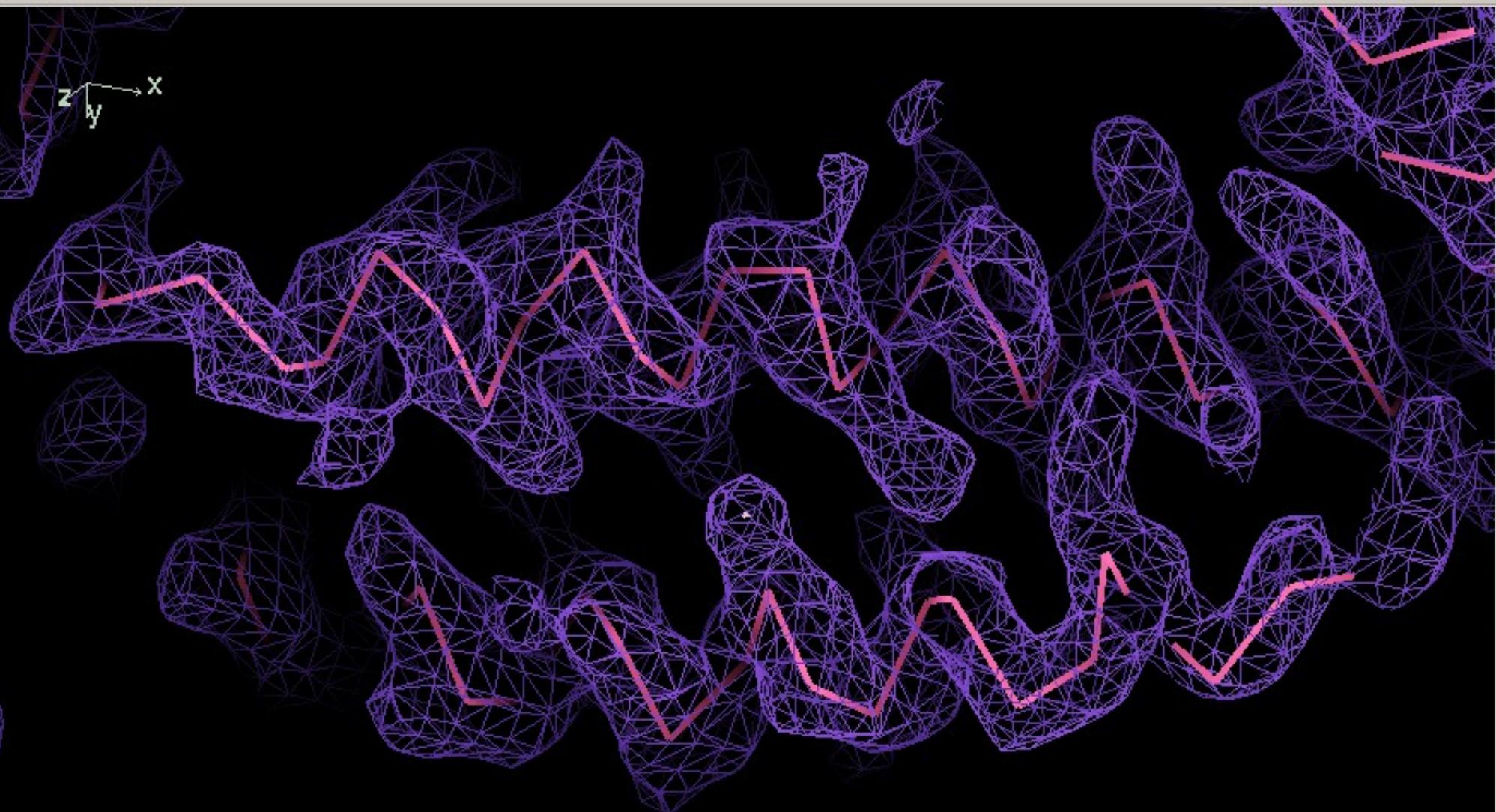
*trans-peptide*  
with plane and trans restraints

# *cis*-peptide Representation



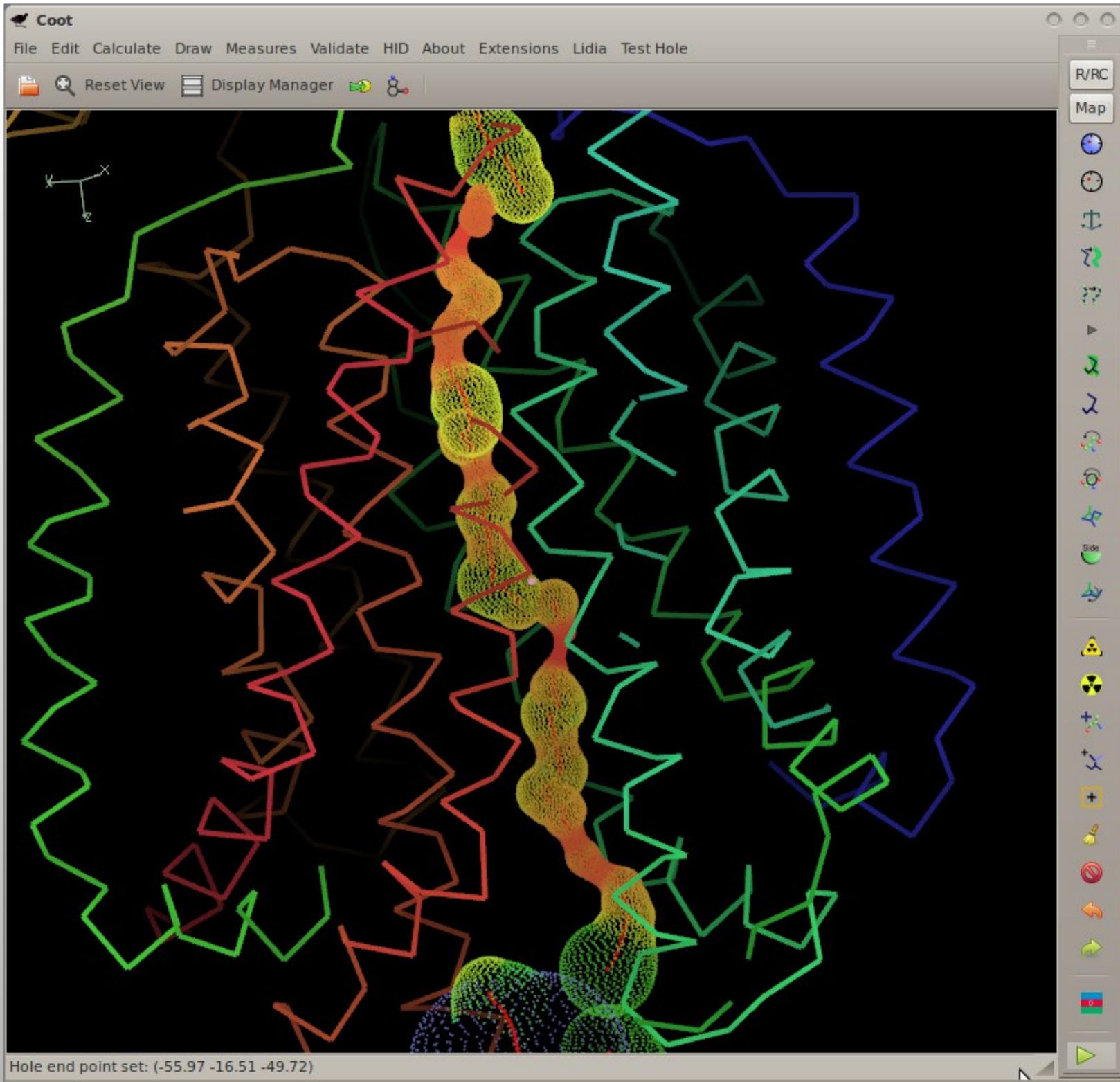
# Interactive Map Sharpening





# Finding Holes

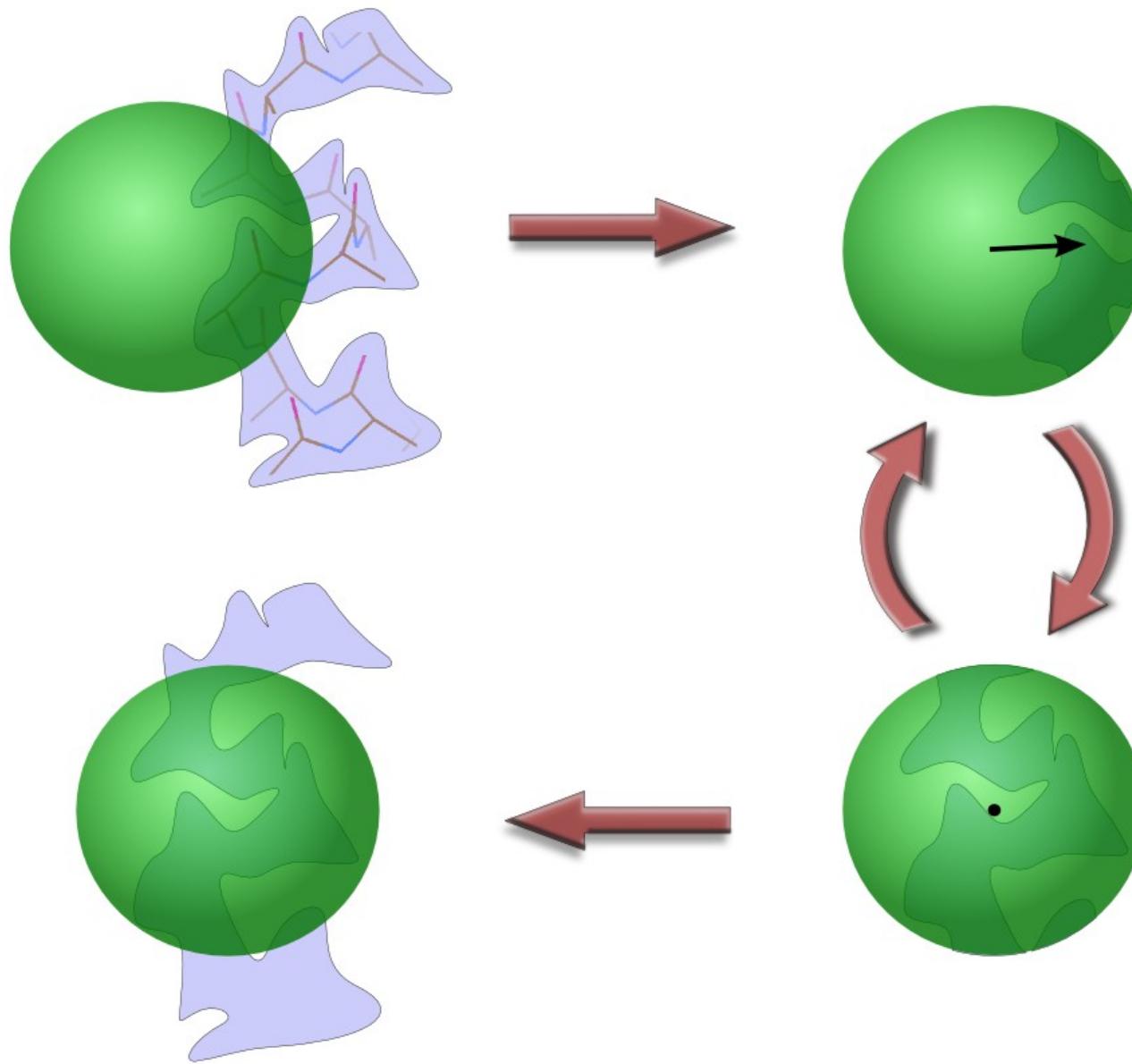
- An implementation of
  - Smart, Goodfellow & Wallace (1993) Biophysics Journal **65**, 2455
  - Atomic radii from AMBER
  - I used
    - radii from CCP4 monomer library
    - sans simulated annealing



# Alpha Helix Placement

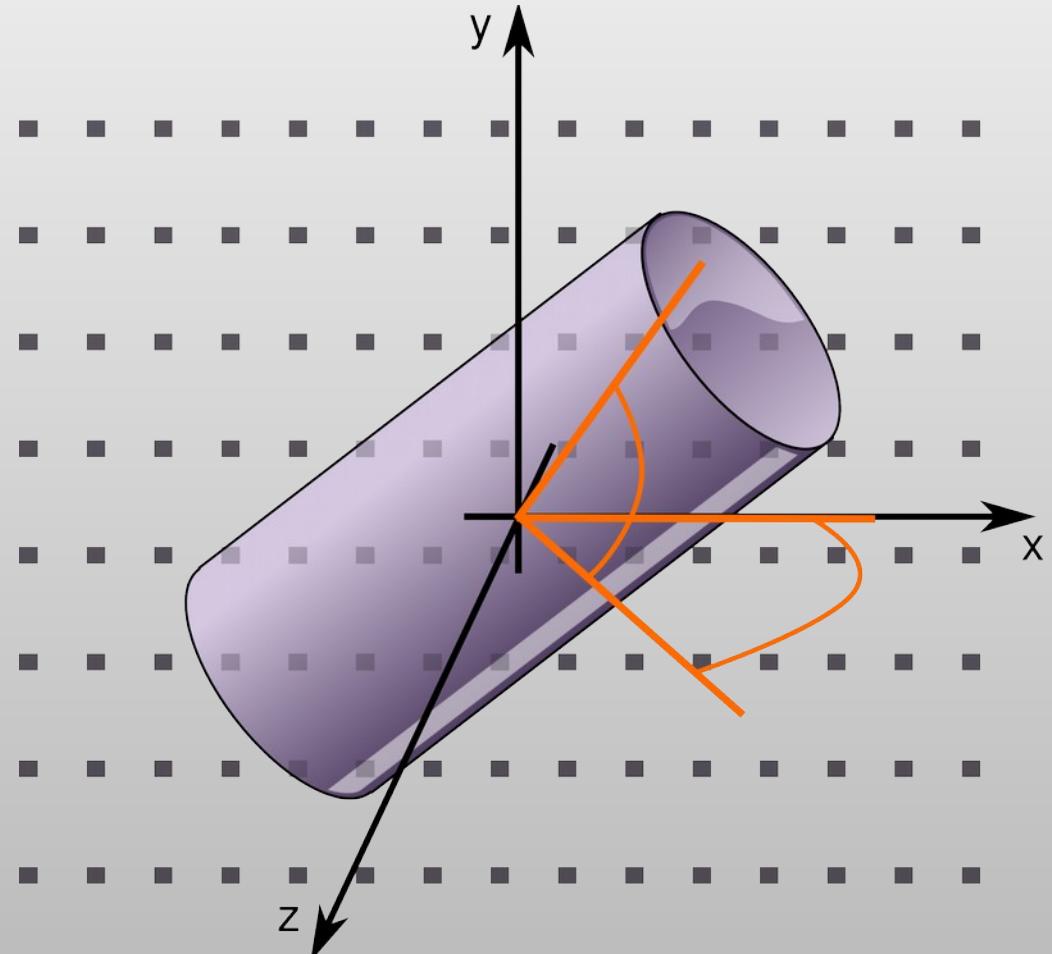
- Scenario: Looking at a new map, not built with automatic tools:
  - “I can see that there’s a helix here - build it for me!”
- From a given point:
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)
  - 1D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

# Centering the Rotation point

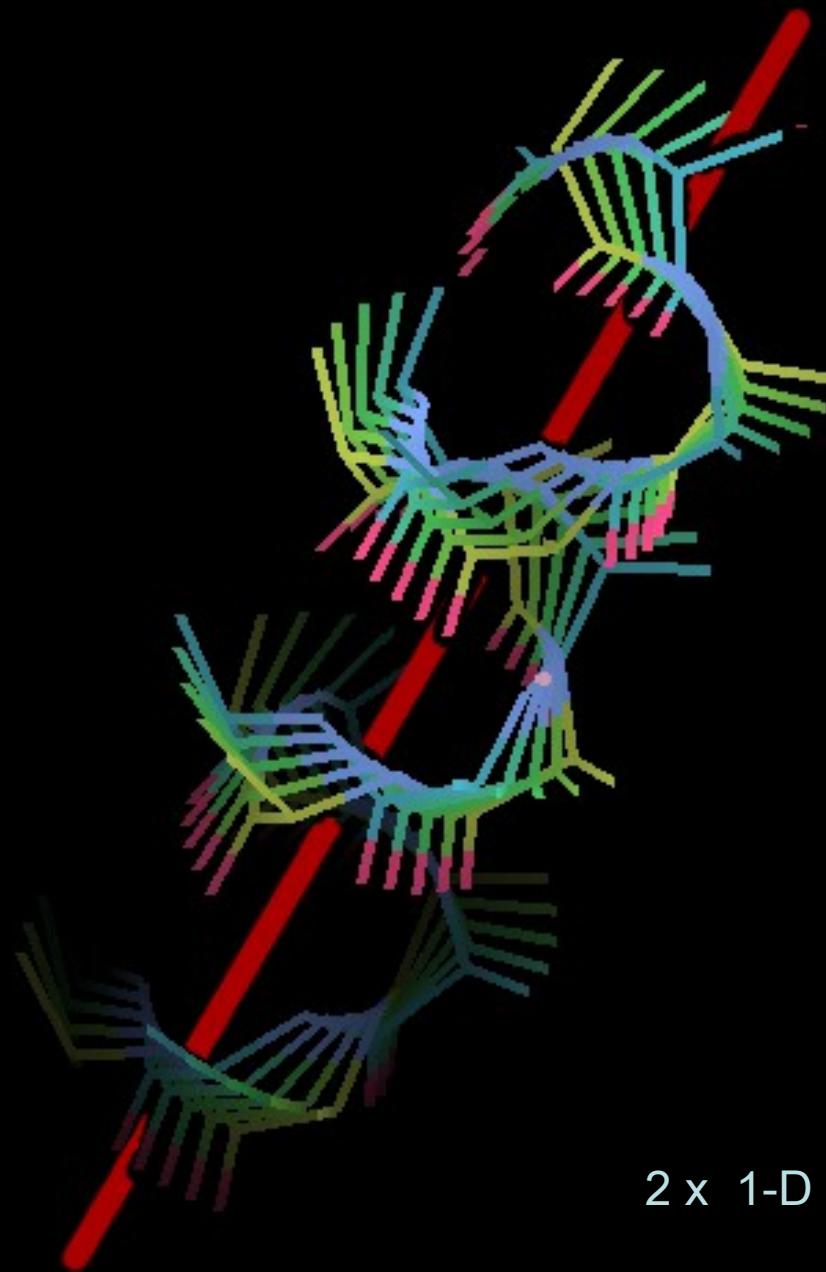


# Cylinder Search

- Pick the orientation that encapsulates the most electron density



Using 2 rotation axes



2 x 1-D Helix orientation searches

Top

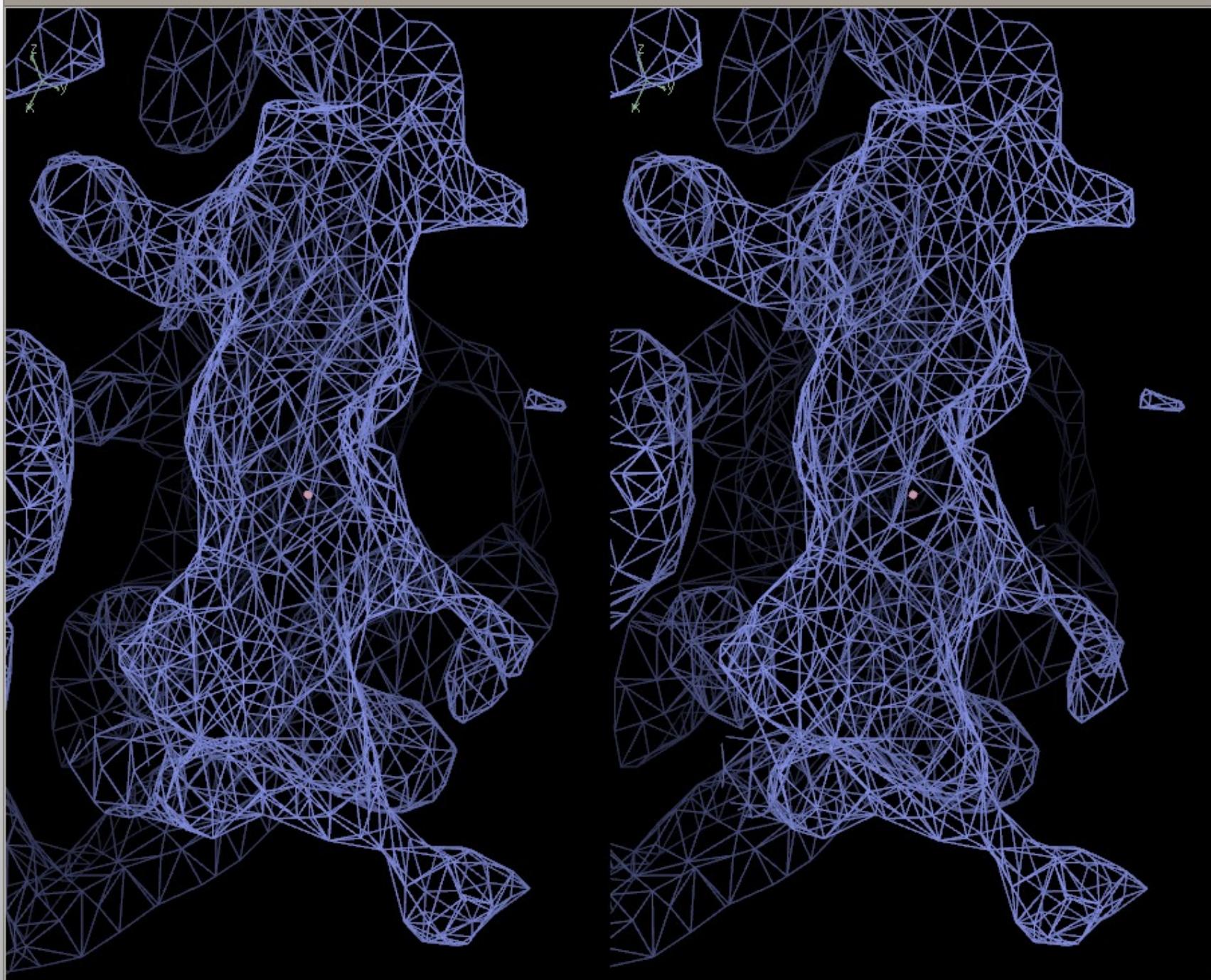


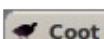
Bottom

Coot

File Edit Calculate Draw Measures Validate HID About Extensions Lidia

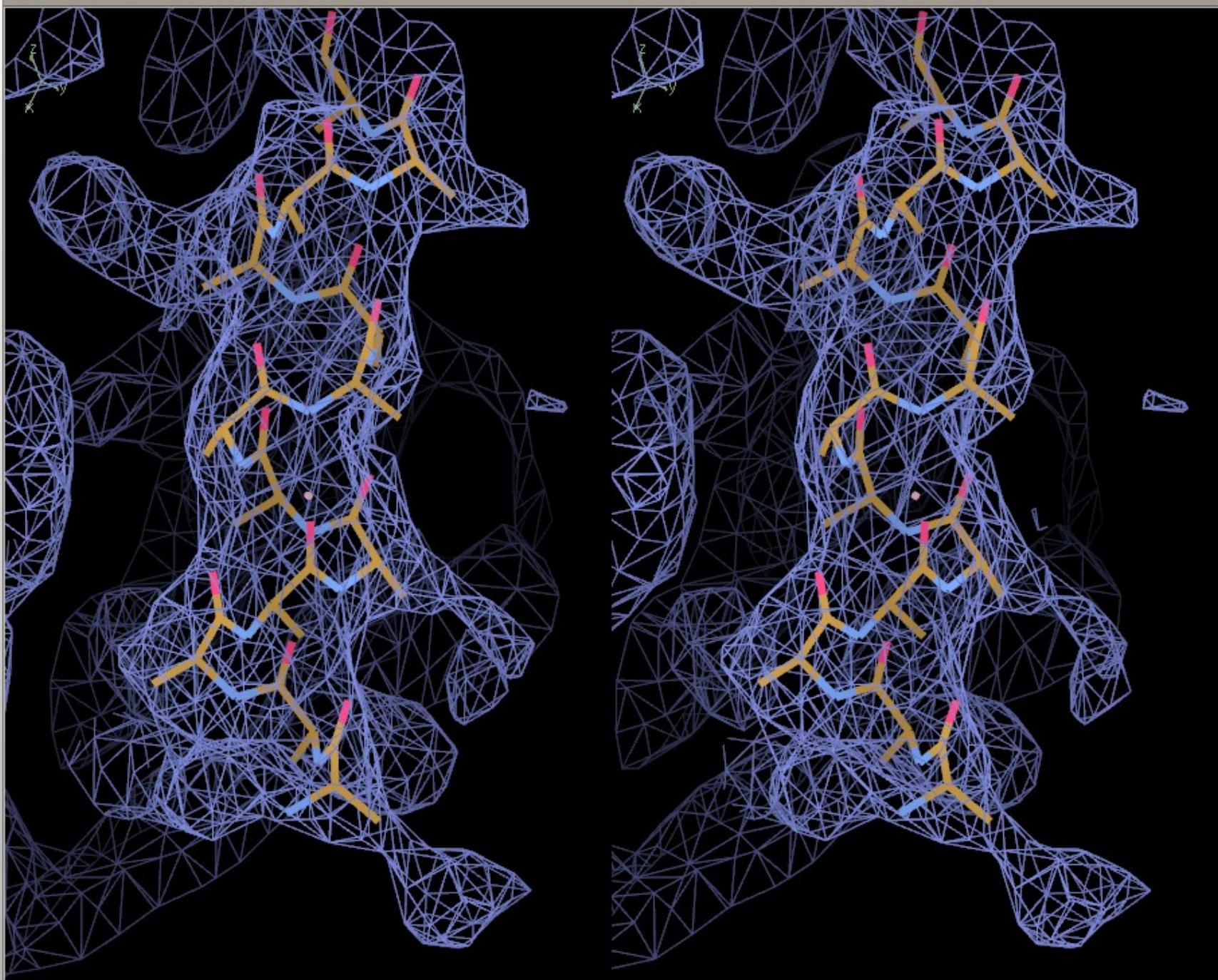
Reset View Display Manager





File Edit Calculate Draw Measures Validate HID About Extensions Lidia

Reset View



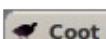
Helix added



R/RC

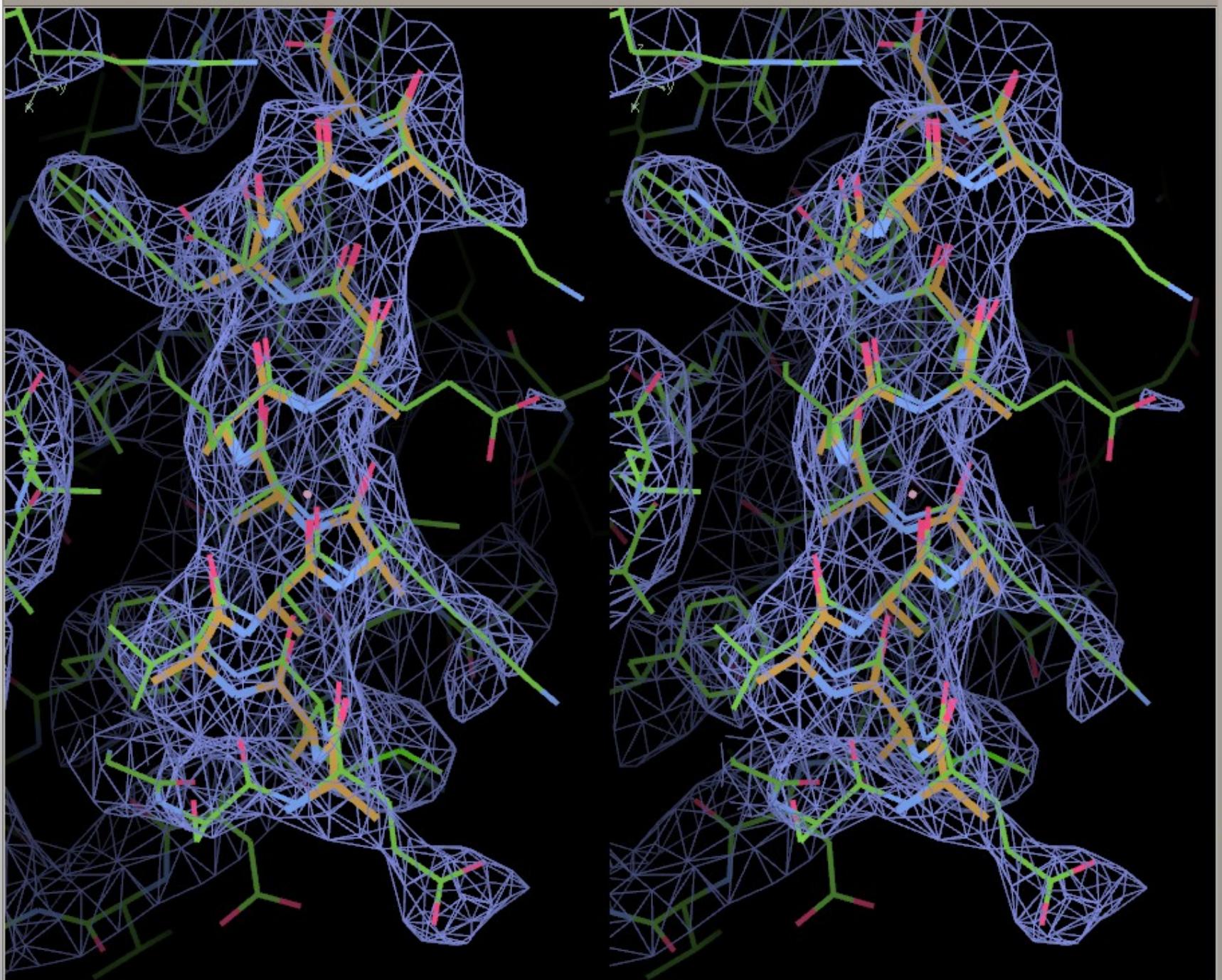
Map





File Edit Calculate Draw Measures Validate HID About Extensions Lidia

Reset View Display Manager



Helix added

...

R/RC

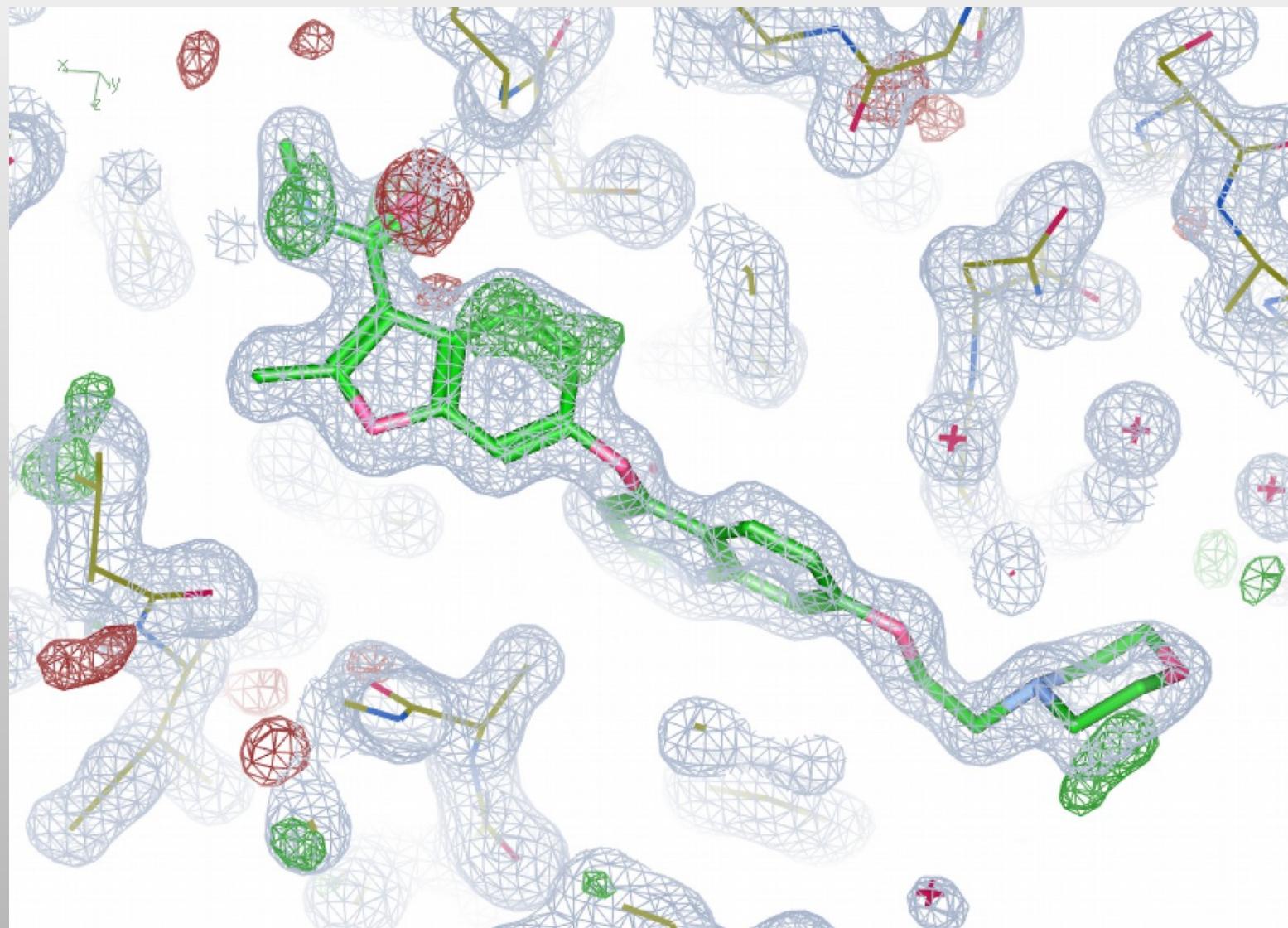
Map



# Making Density Slides with Coot

- White background
- “High” Oversampling (2.3x)
- Pale gray (or very pastel) density colour
- Enable Cut-glass mode 5-10%
- Anti-aliased Coot
  - `$ setenv __GL_FSAA_MODE 5`
  - 0.8.3 will do a better job of anti-aliasing out the box
  - (transfer to CCP4-built binaries)

# Example Density Slide



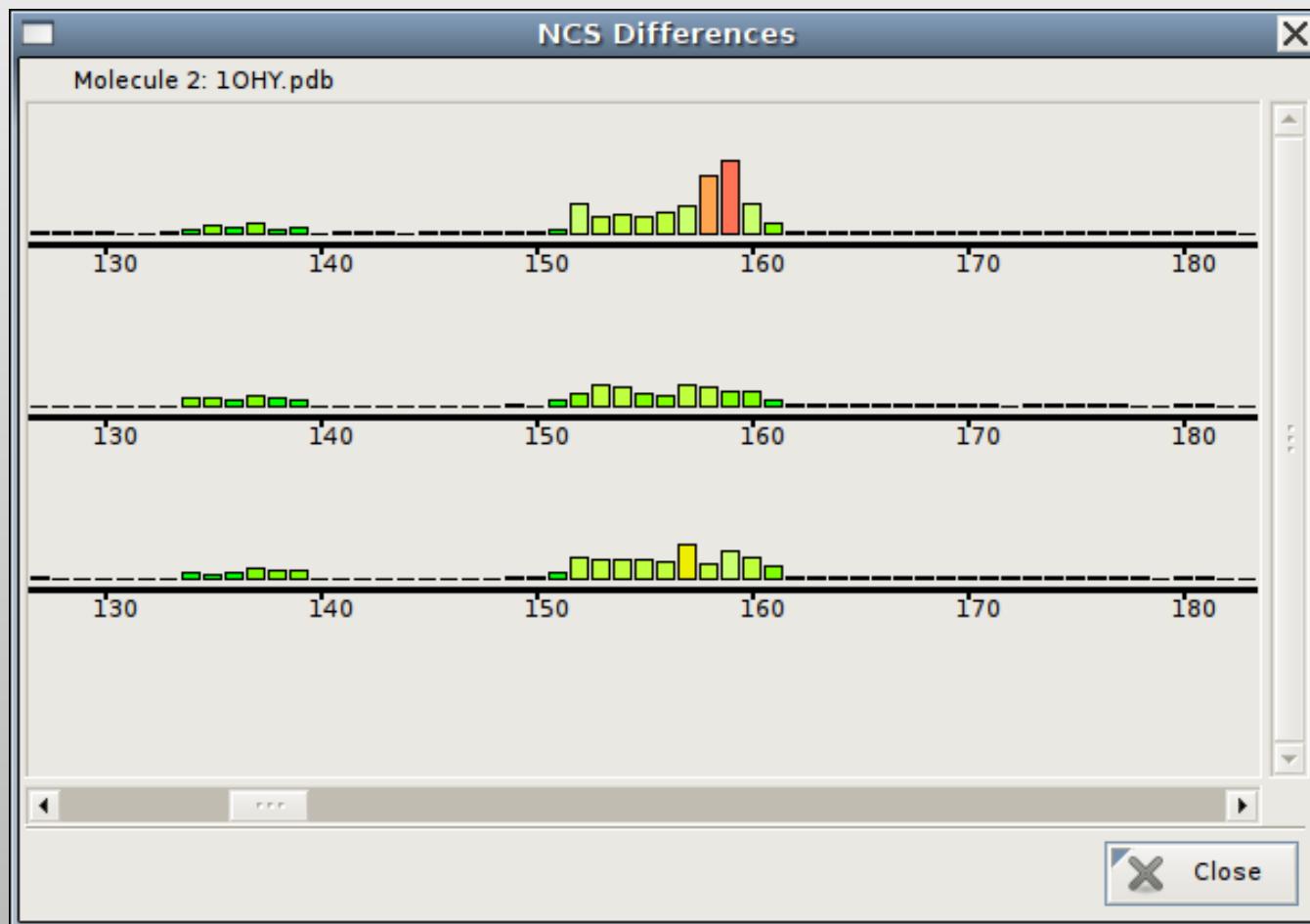
# Handling NCS...

# Handling NCS

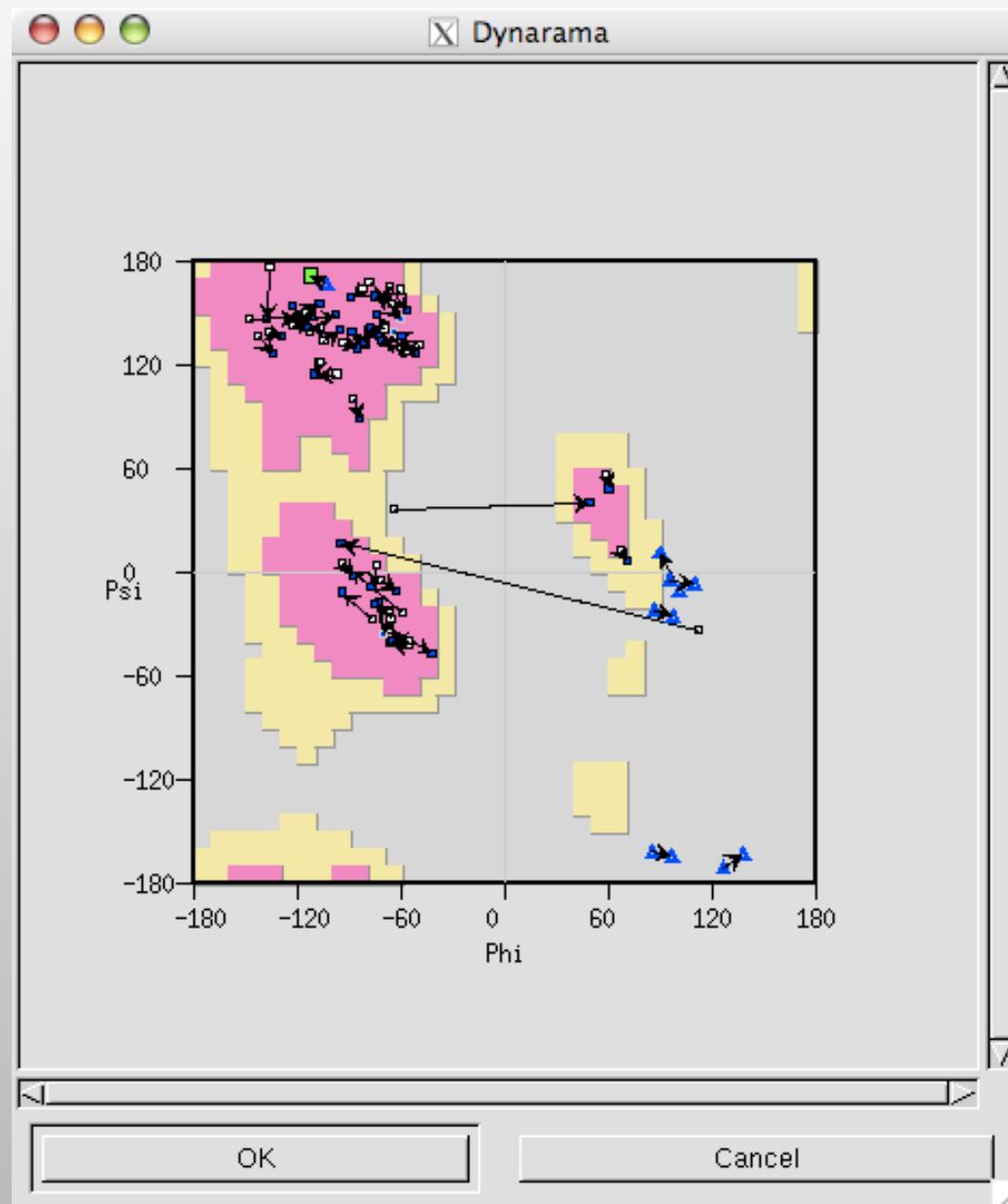
## Typical Scenario:

- 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?

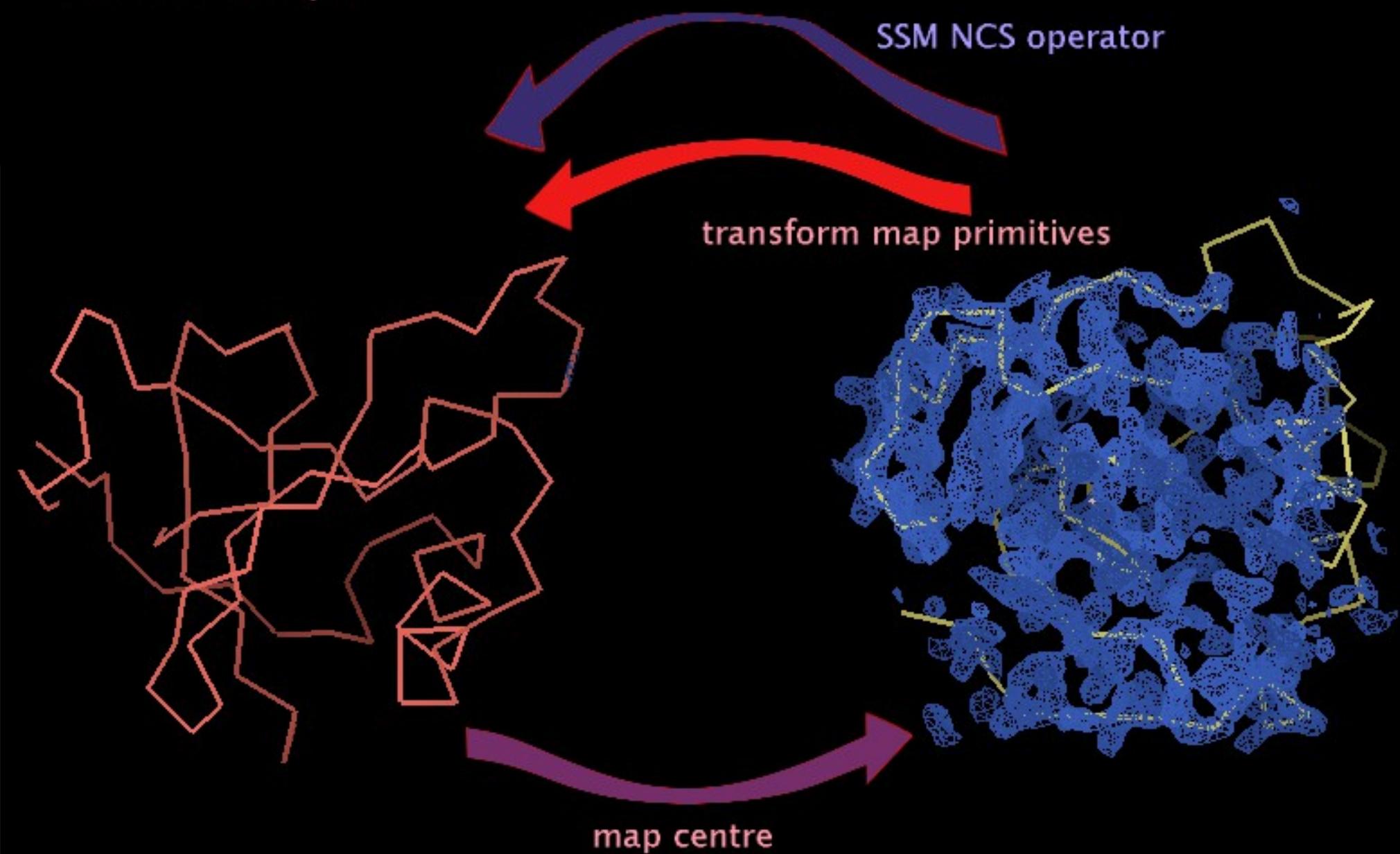
# ... NCS Differences graph



# ...or "Kleywelt" Plots

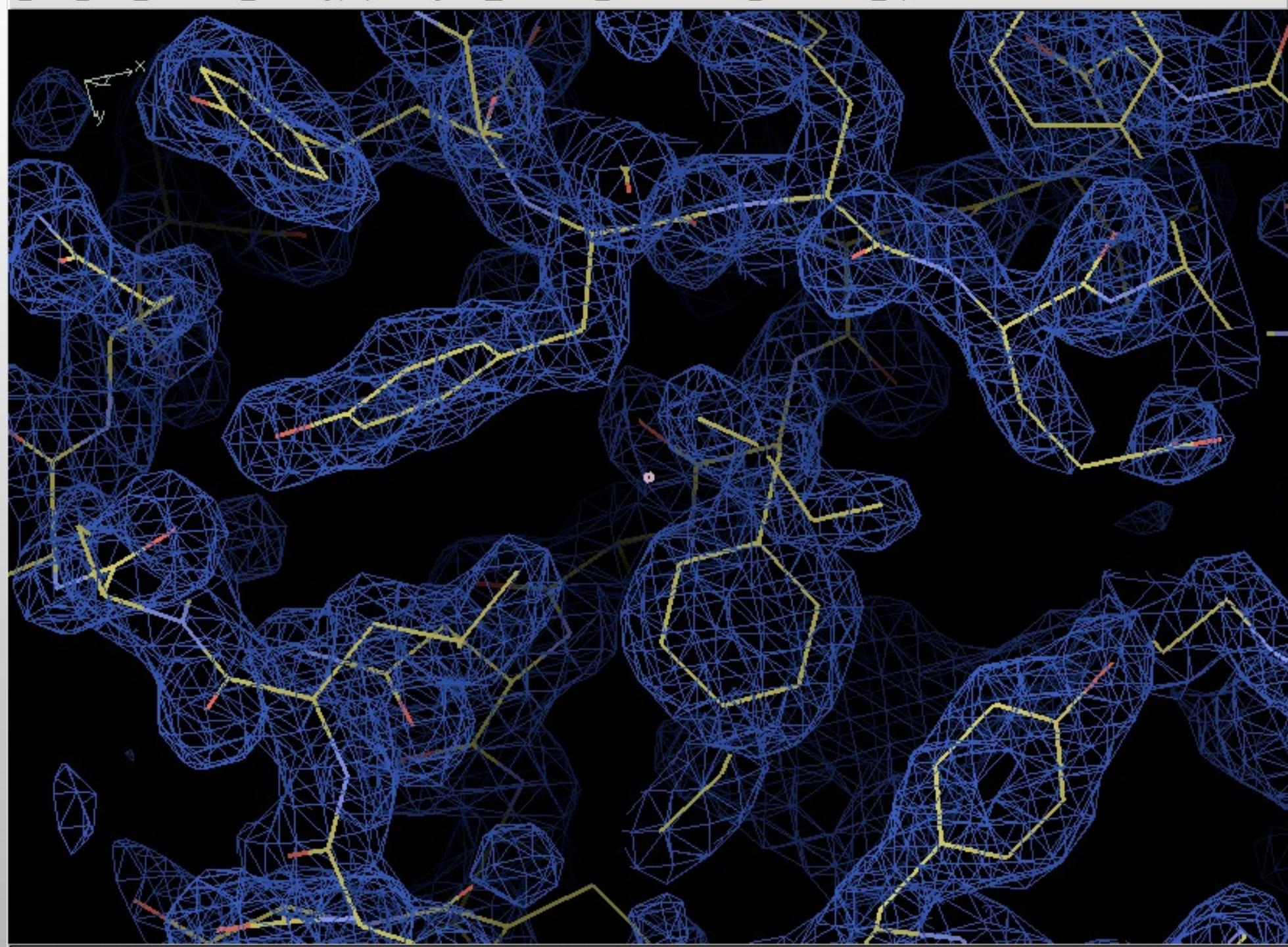


## NCS Overlays



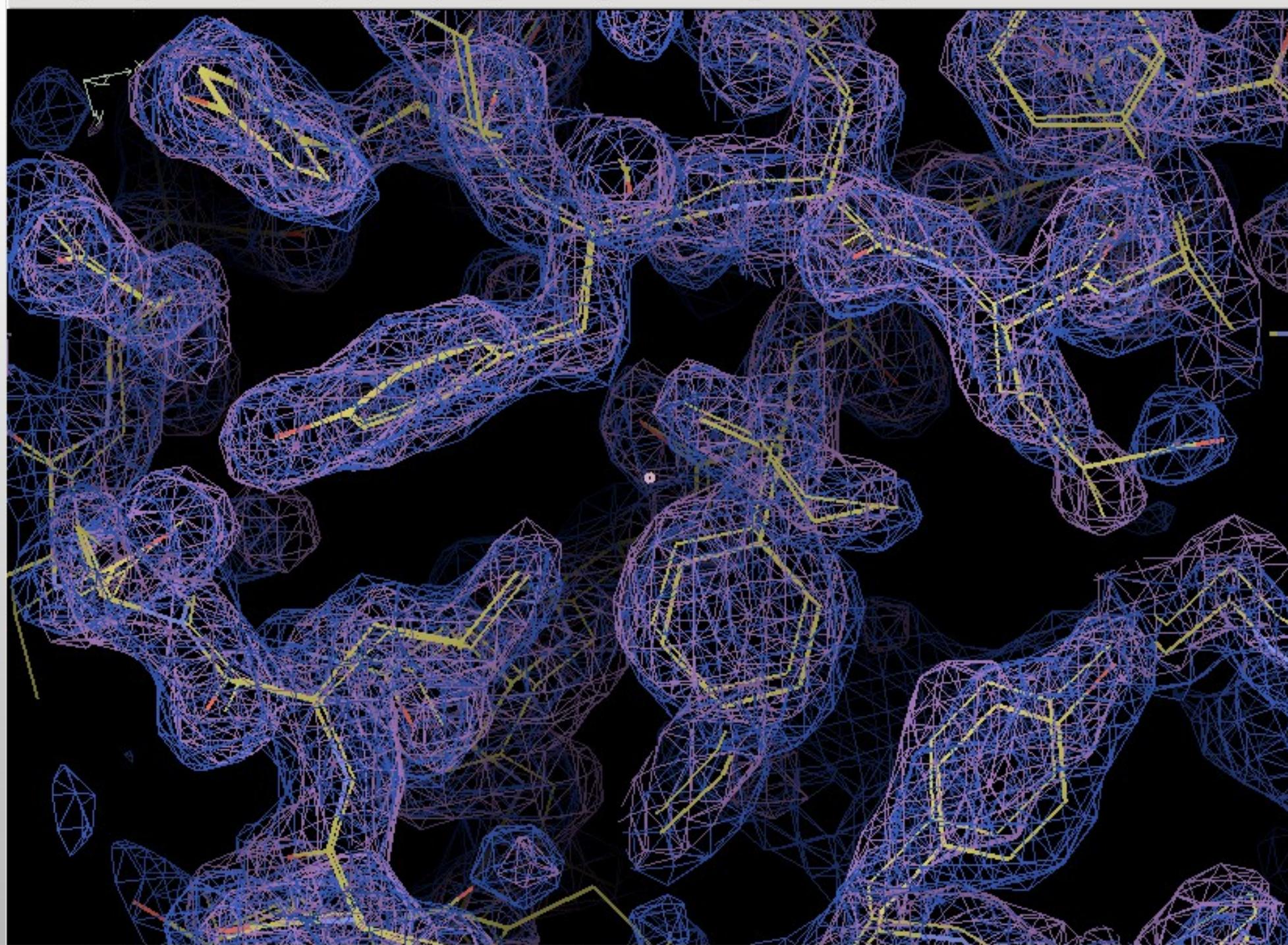
## Coot

File Edit Calculate Draw Display Manager Measures Validate HID Reset View Help



## Coot

File Edit Calculate Draw Display Manager Measures Validate HID Reset View Help



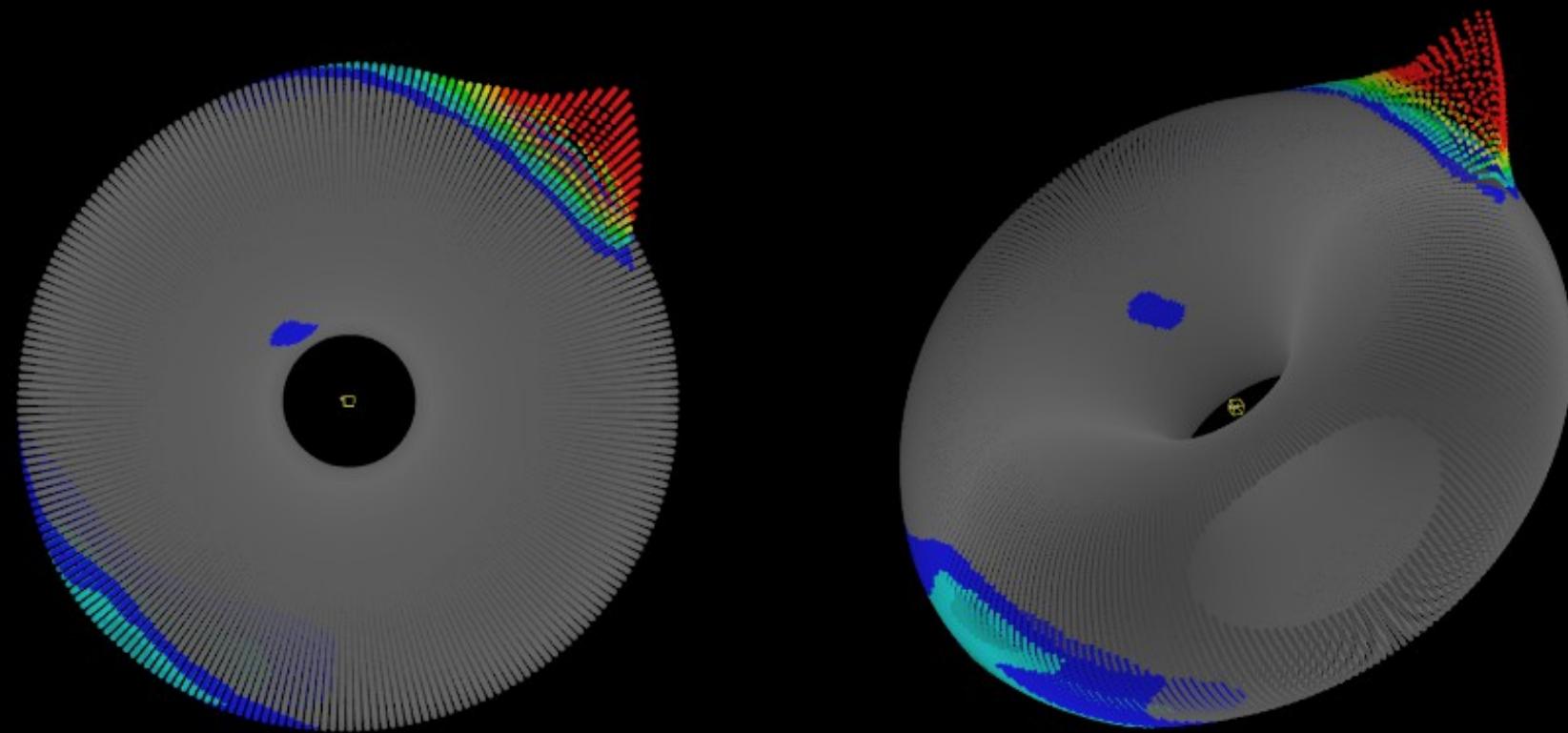
(mol. no: 3) CG /1/A/52 TYR occ: 1.00 bf: 11.63 ele: C pos: (50.36, 2.86, 13.40)

# NCS Model-modification Tools

- Automatic detection of NCS
  - And their operators
- Copy Master NCS molecule to others
  - Applies NCS transformation
- Copy NCS Master residue-range
- Change NCS Master chain
- NCS Skipping
- NCS Ligands

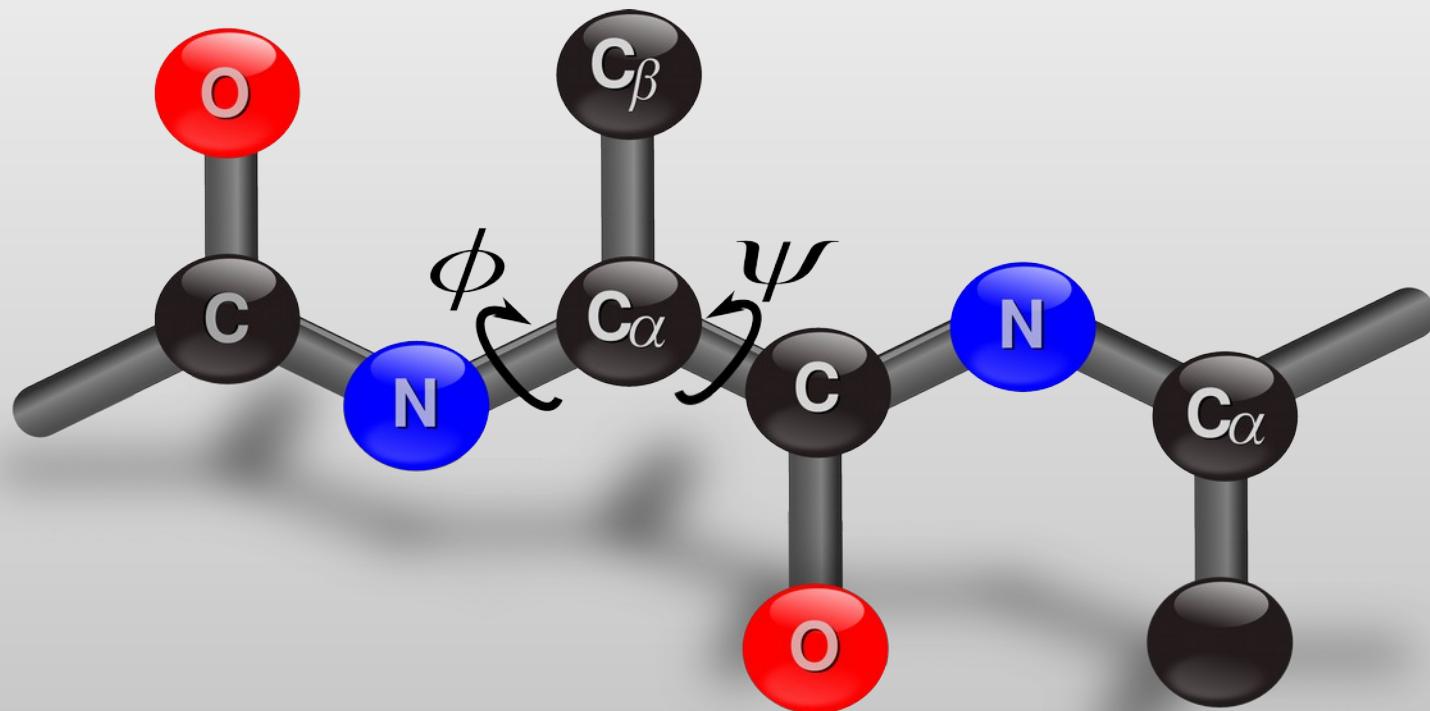
# Acknowledgements

- Keith Wilson
- Eleanor Dodson
- Garib Murshudov
- Libraries, dictionaries
  - Alexei Vagin, Eugene Krissinel
  - Richardsons (Duke)
- Funding
  - BBSRC, CCP4 & MRC



Projection from the surface of a doughnut:  $2 \times 360^\circ$   
(linear scaling)

# Peptide Backbone Geometry



# Typical 2D Projection of Ramachandran Plot

