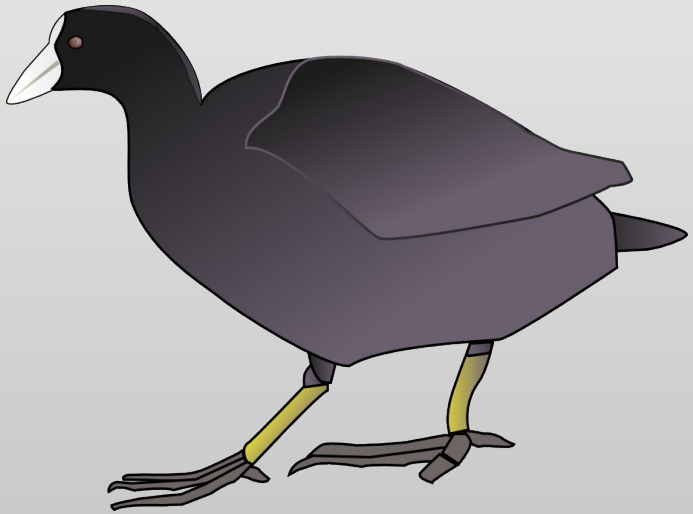
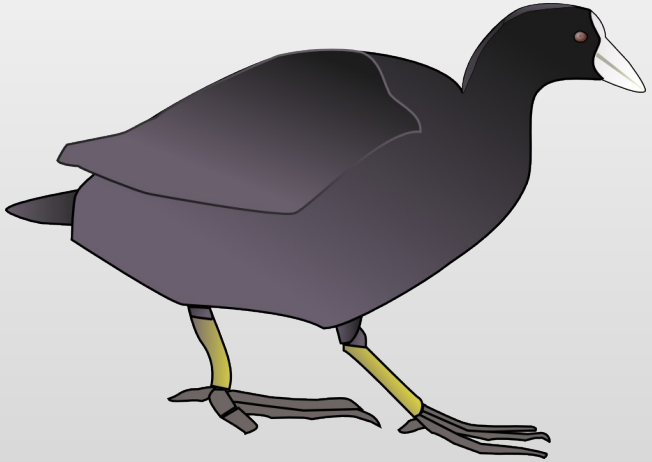
Modelling Macromolecules with *Coot*

- Overview
  - Real Space Refinement
  - A Sample of Tools
  - Tools for Cryo-EM
  - Tools for Ligands
  - [Carbohydrates]



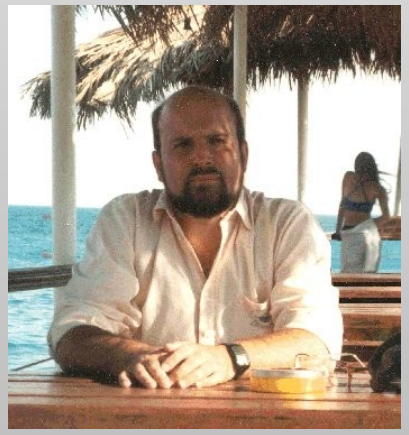
Paul Emsley
MRC Laboratory of Molecular Biology

Acknowldegments, Collaborators



Bernhard Lohkamp

Kevin Cowtan



Eugene Krissinel

Stuart McNicholas

Martin Noble

Alexei Vagin

Coot

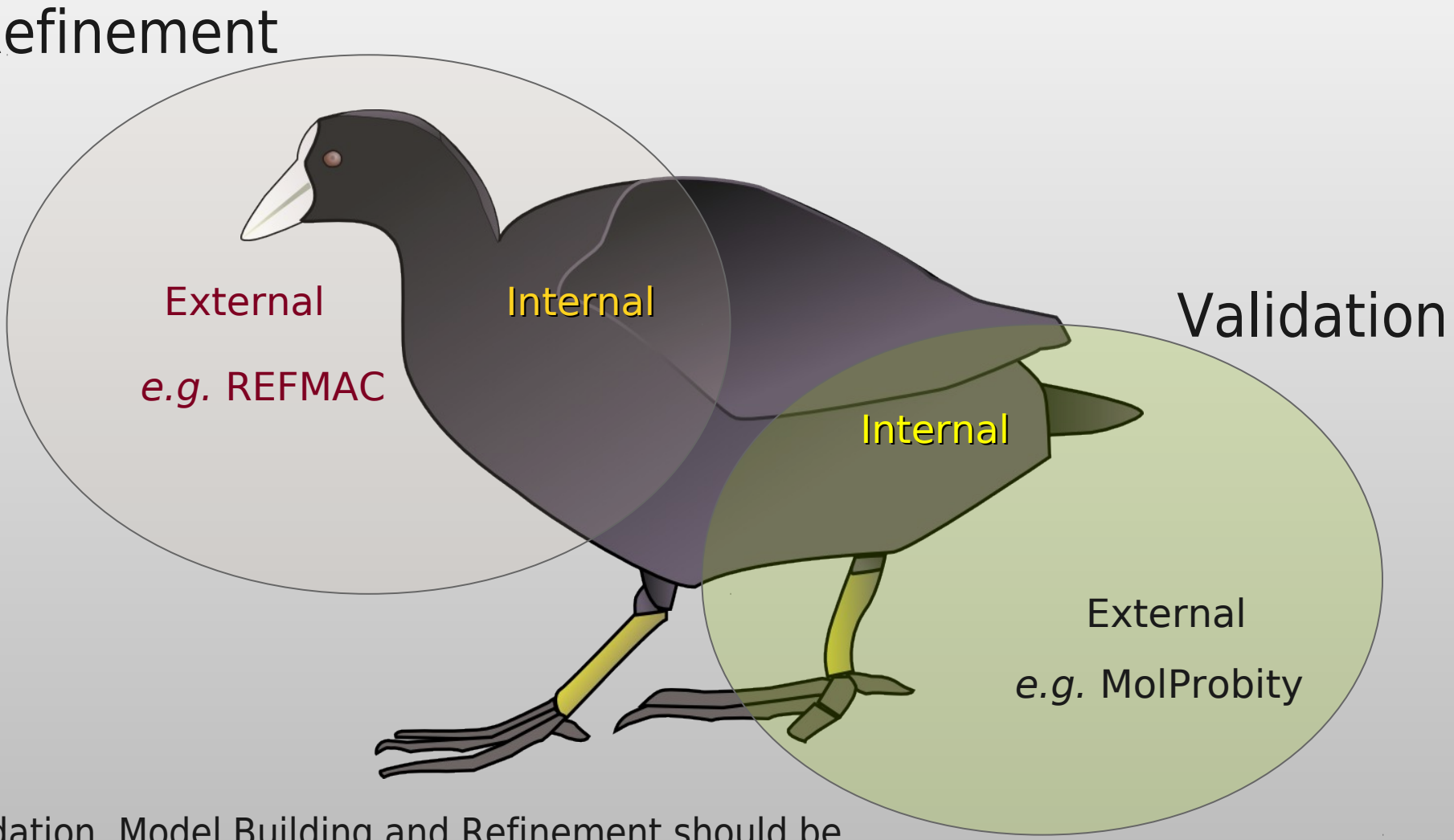
- Crystallographic Object-Oriented Tool-kit
- Primarily a tool for the interpretation of electron density generated from X-ray data
  - with tools for modelling:
    - rotate/translate, rotamers,
    - refinement & regularization
    - add, delete
    - ligand fitting and analysis
  - to be used post-automation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

Why bother?

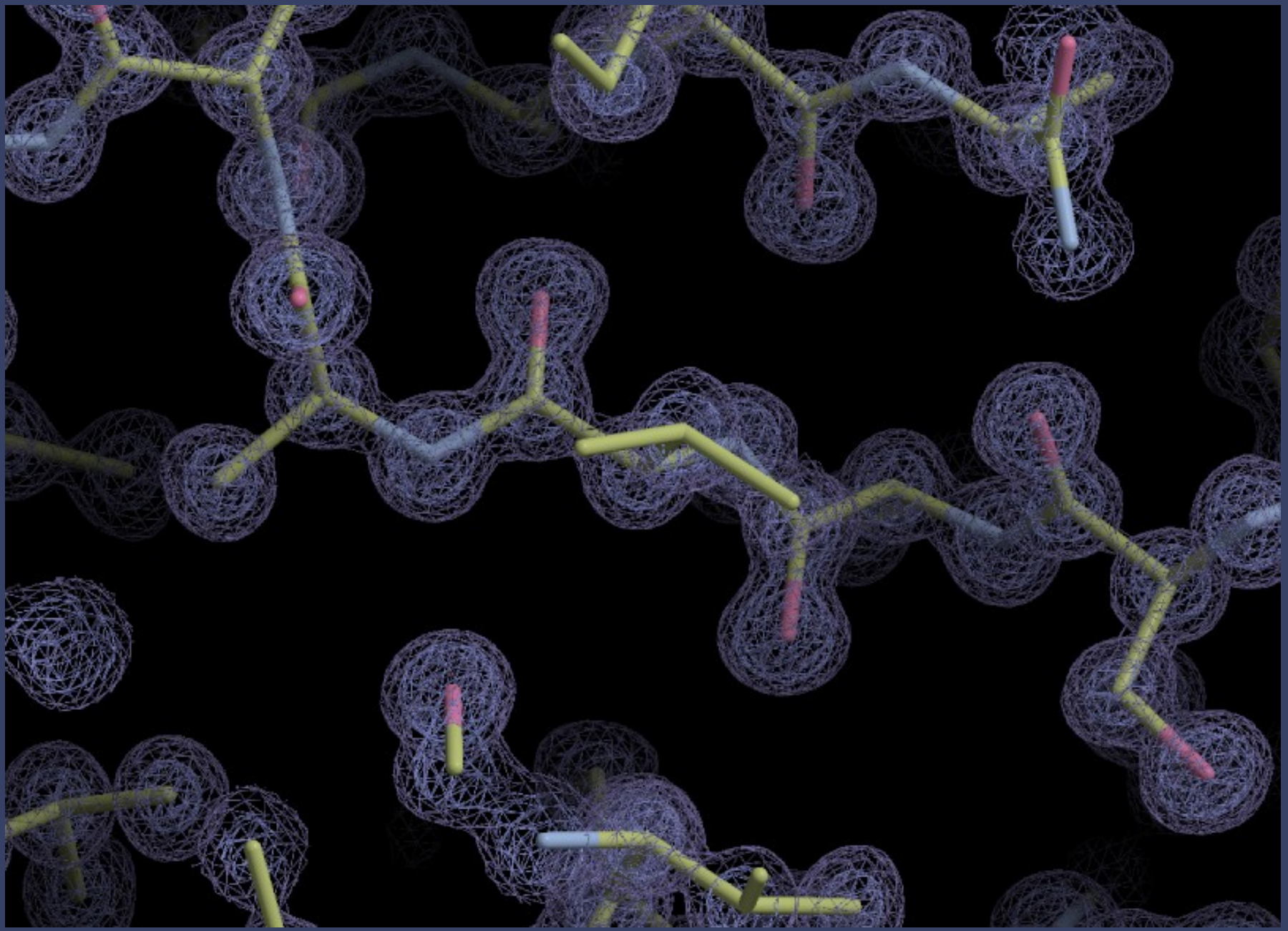
- Automated (complete) model-building still impractical
  - Extremely demanding
  - It takes a brain to validate
- Concerted motion of atoms connected by geometric restraints is difficult
- Coot is built with Novice users in mind
  - (but not exclusively)
  - because using the **key-bindings** will turn you from

    Noob → Pro

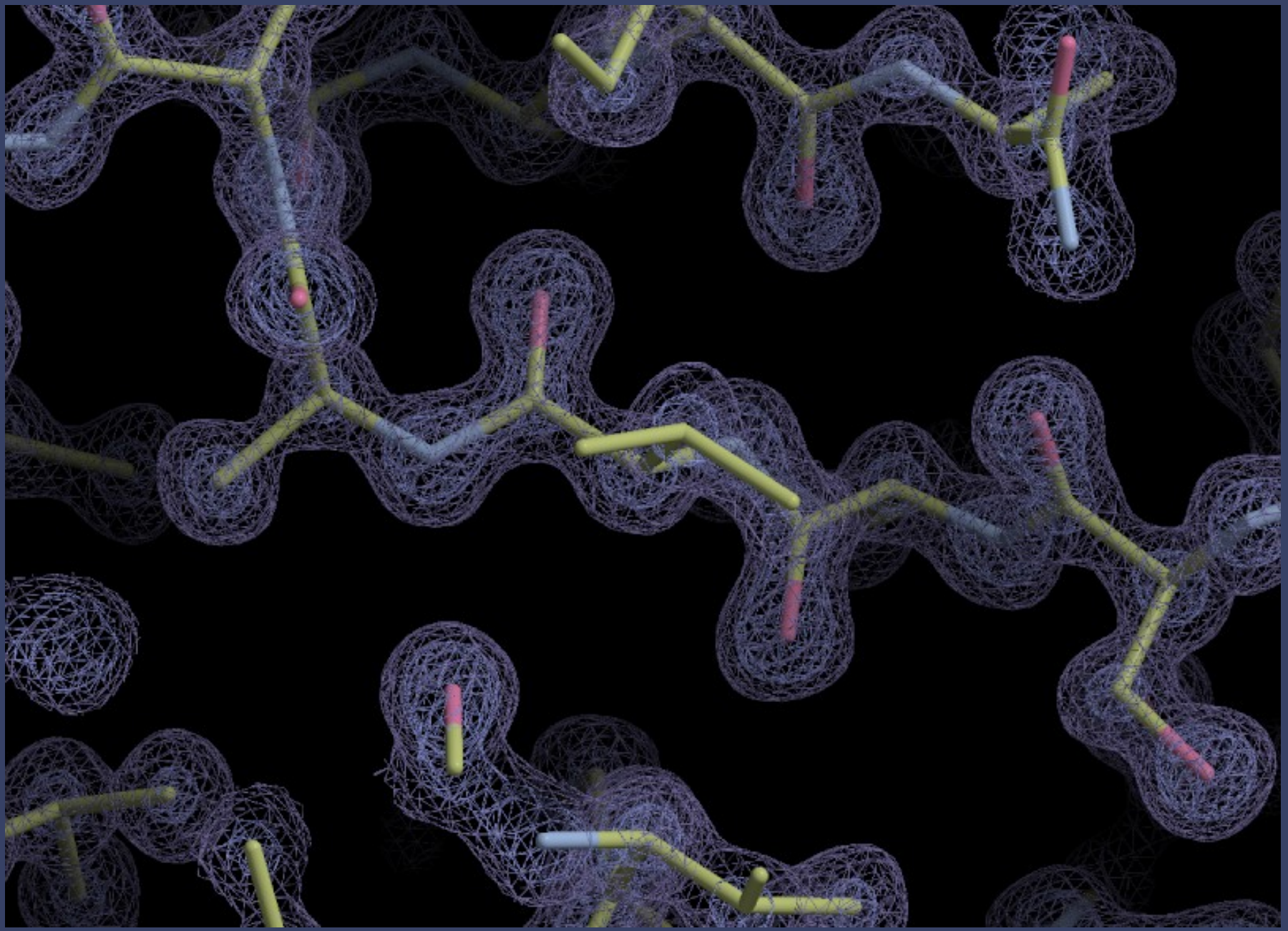
Feature Integration



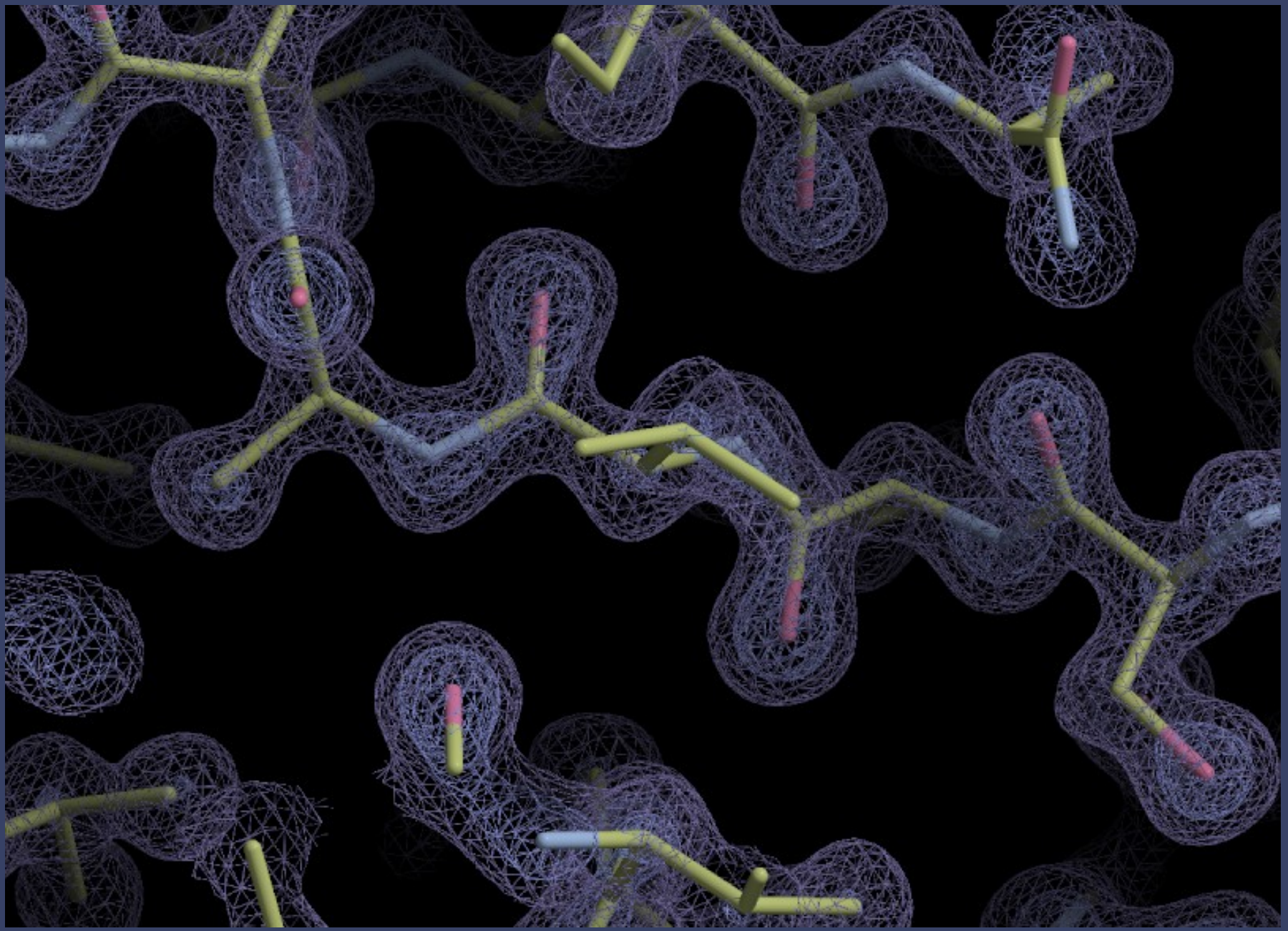
Validation, Model Building and Refinement should be used together


1.0Å

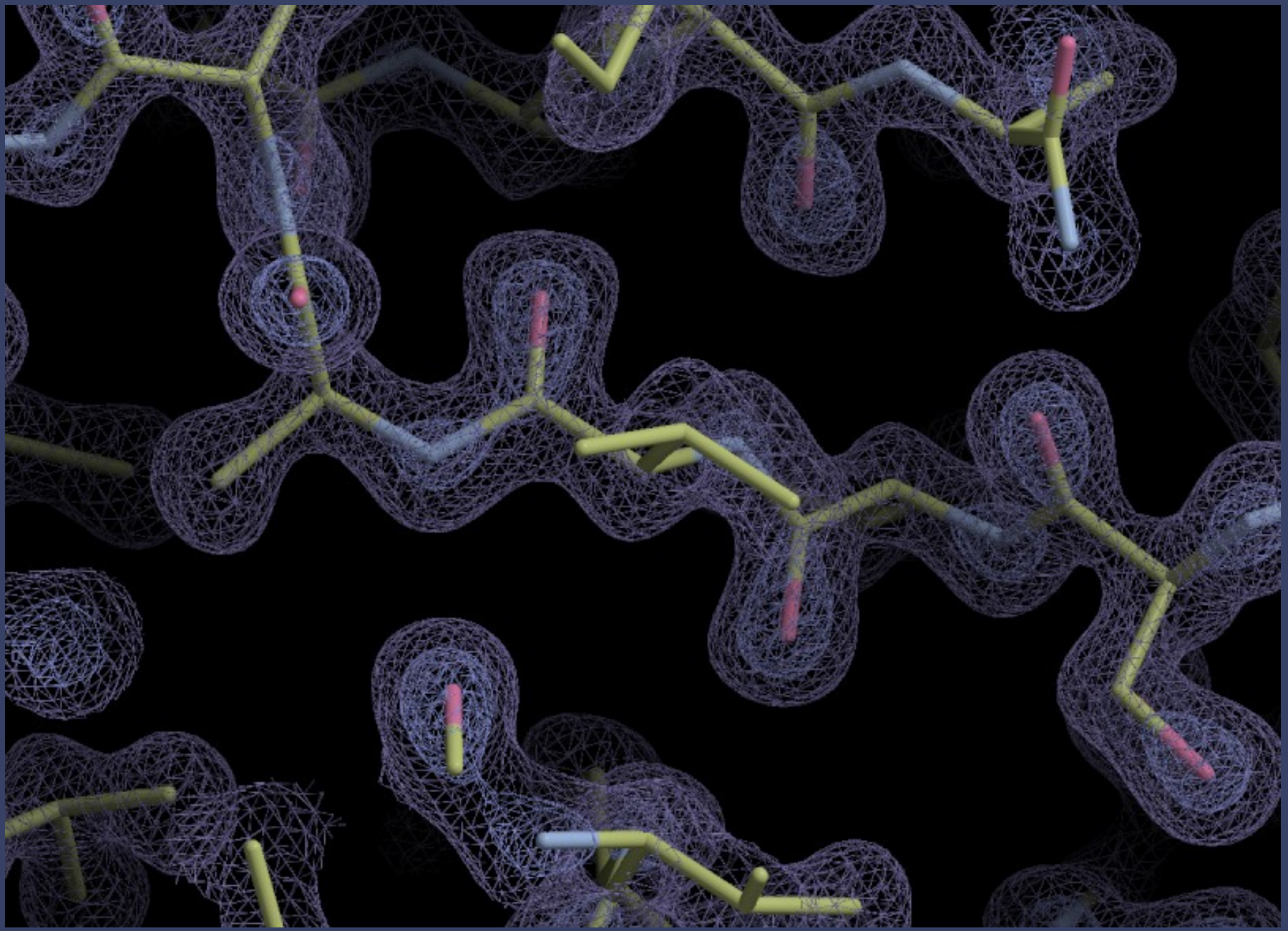
1.2Å



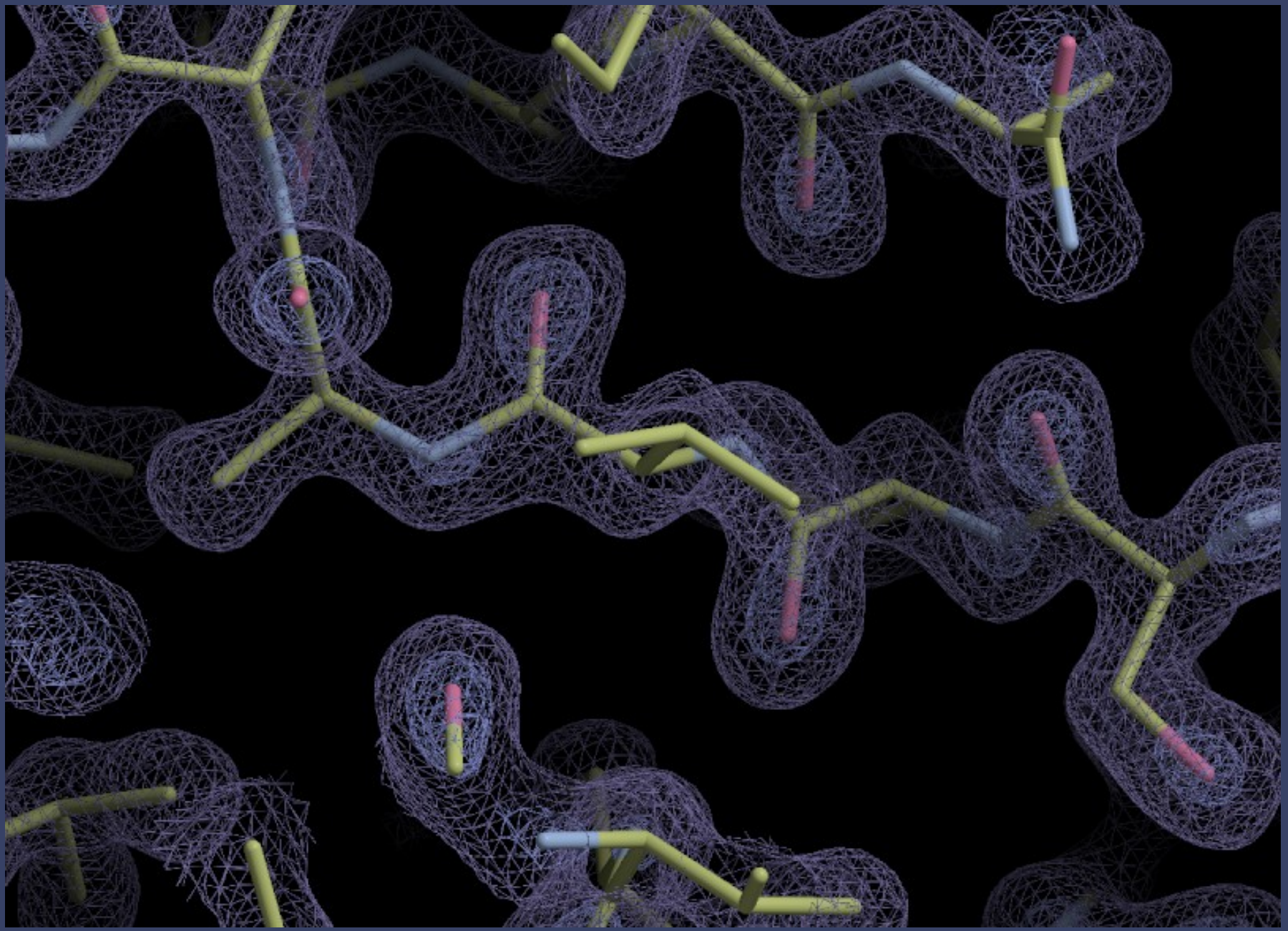
1.4Å



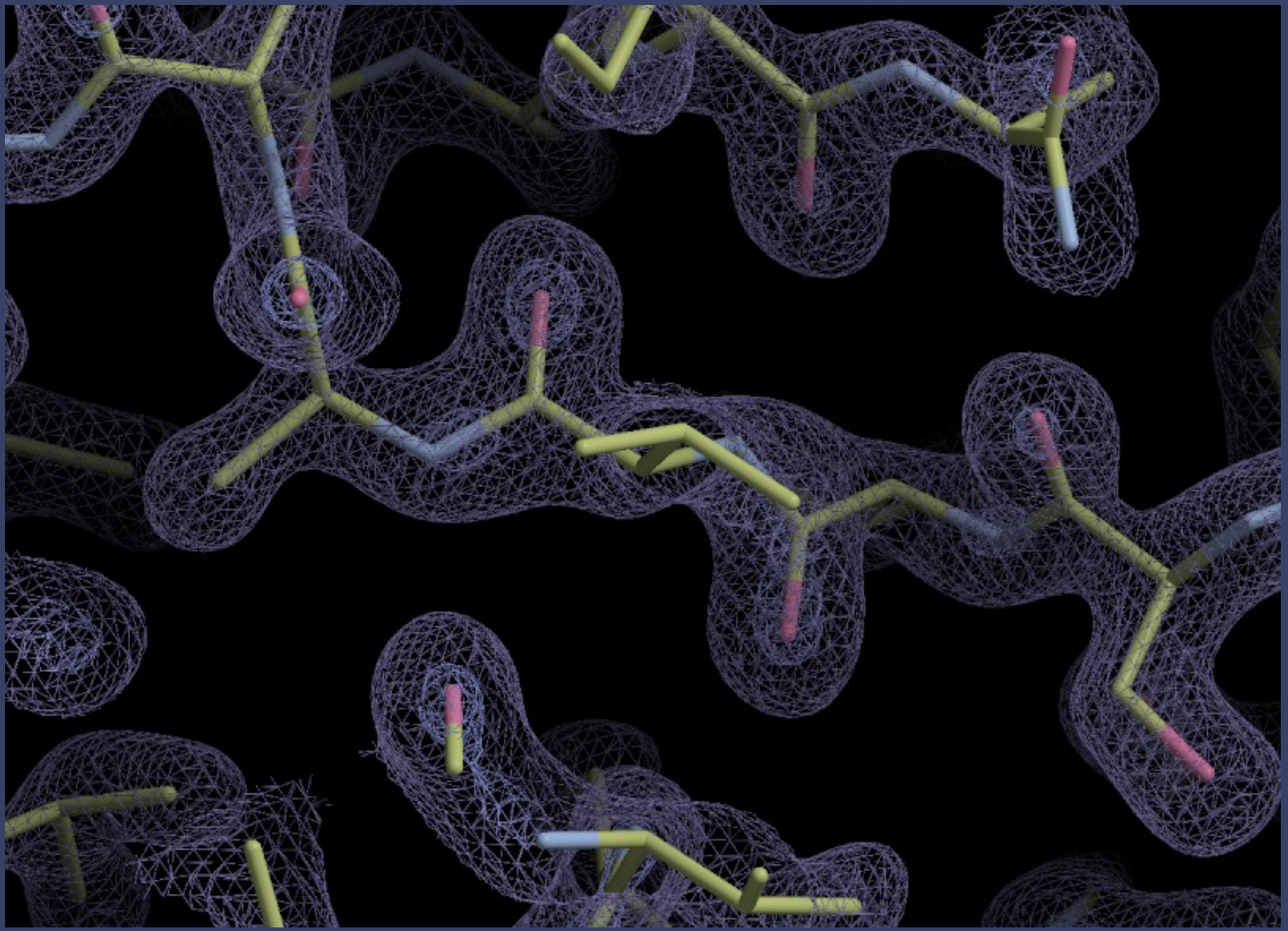
1.6Å



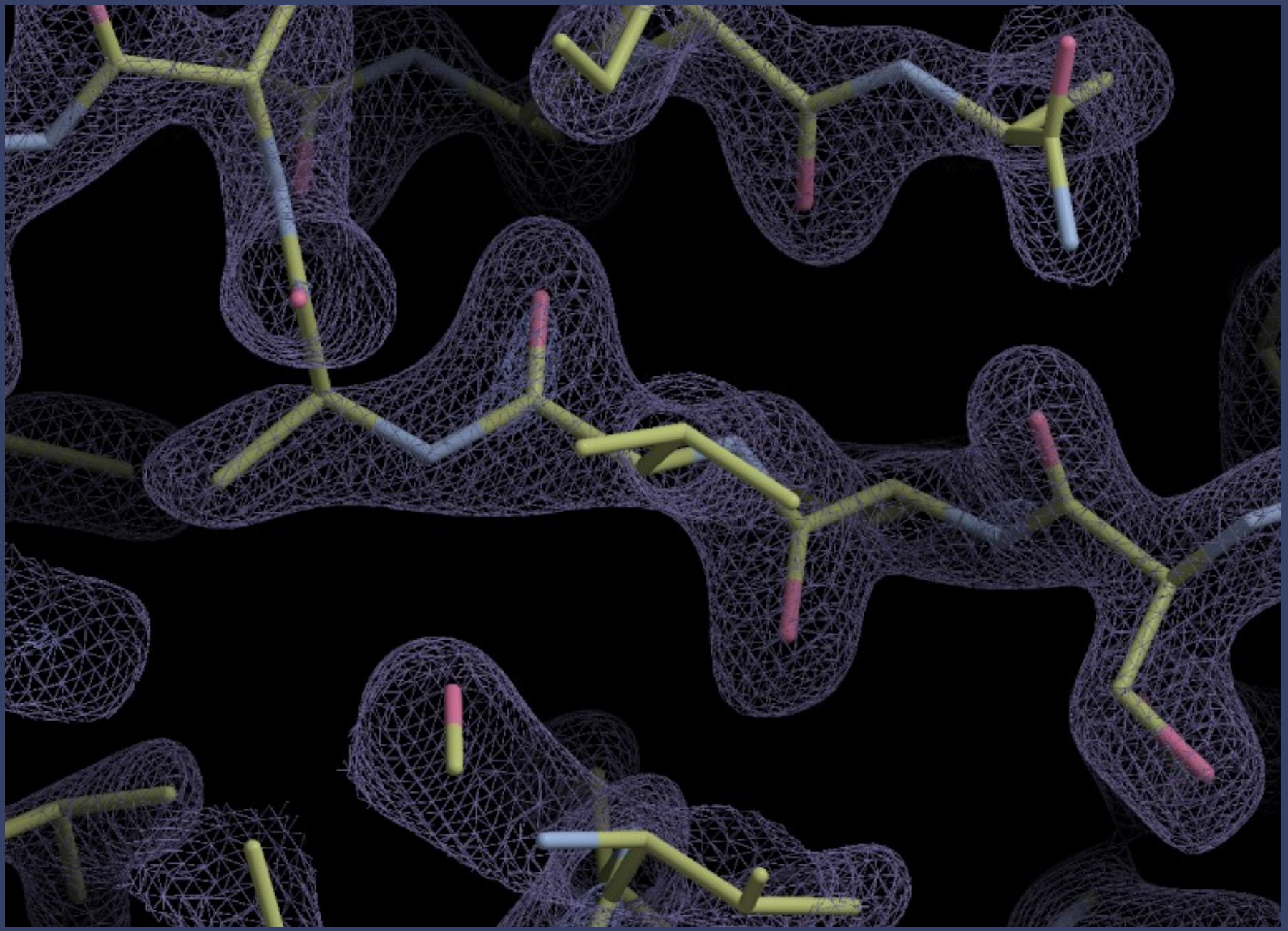
1.8Å



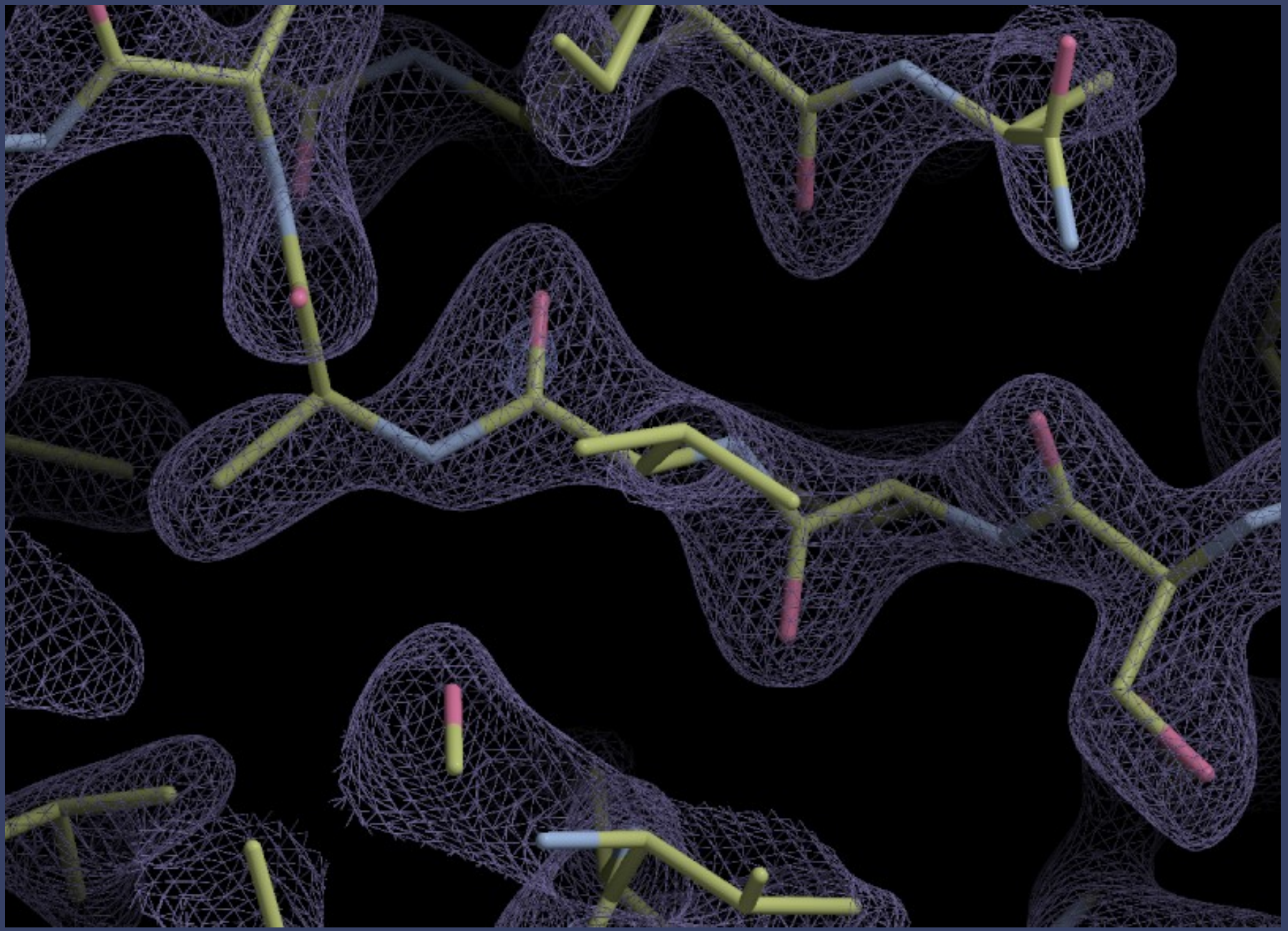
2.0Å



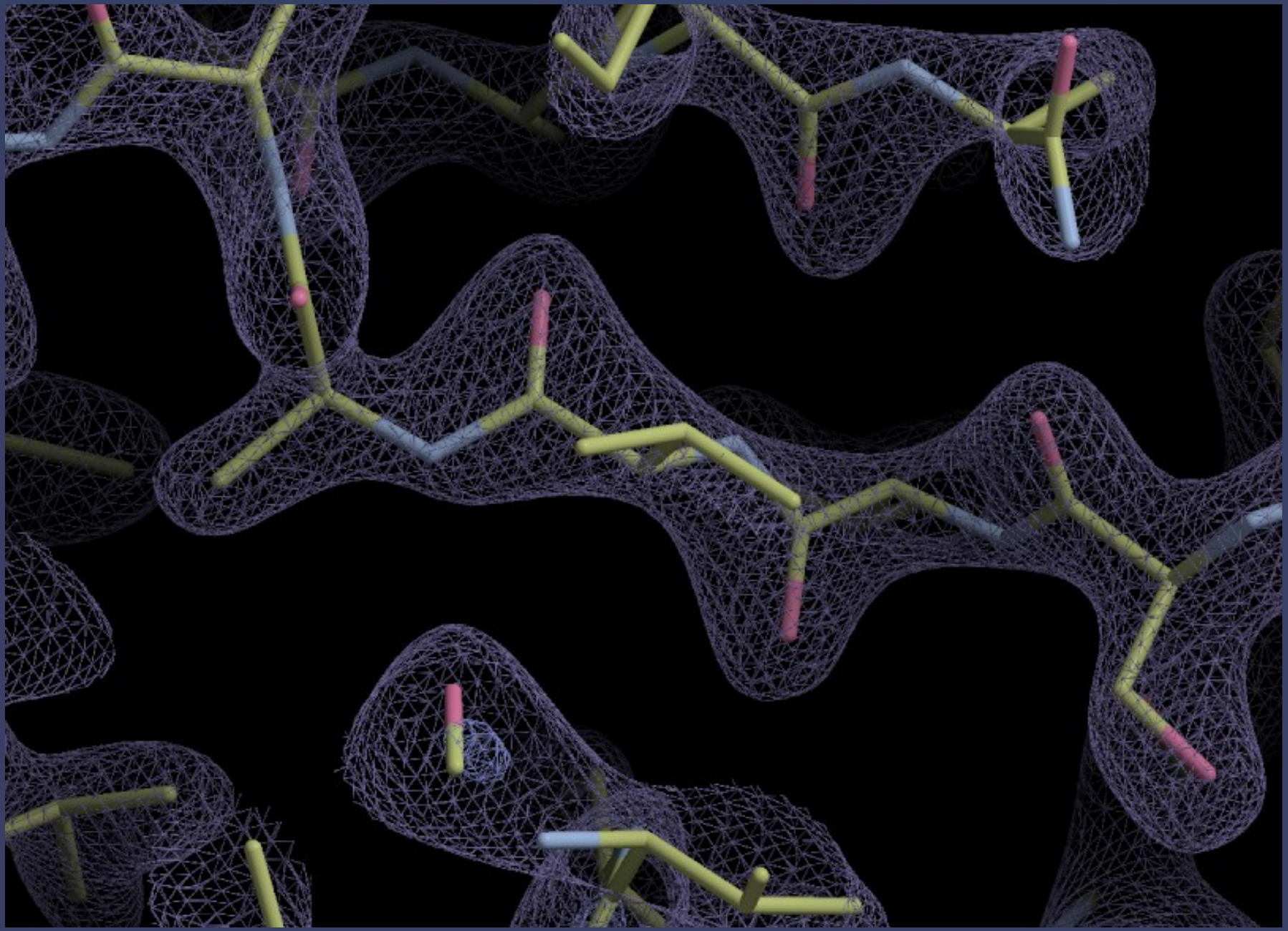
2.2Å



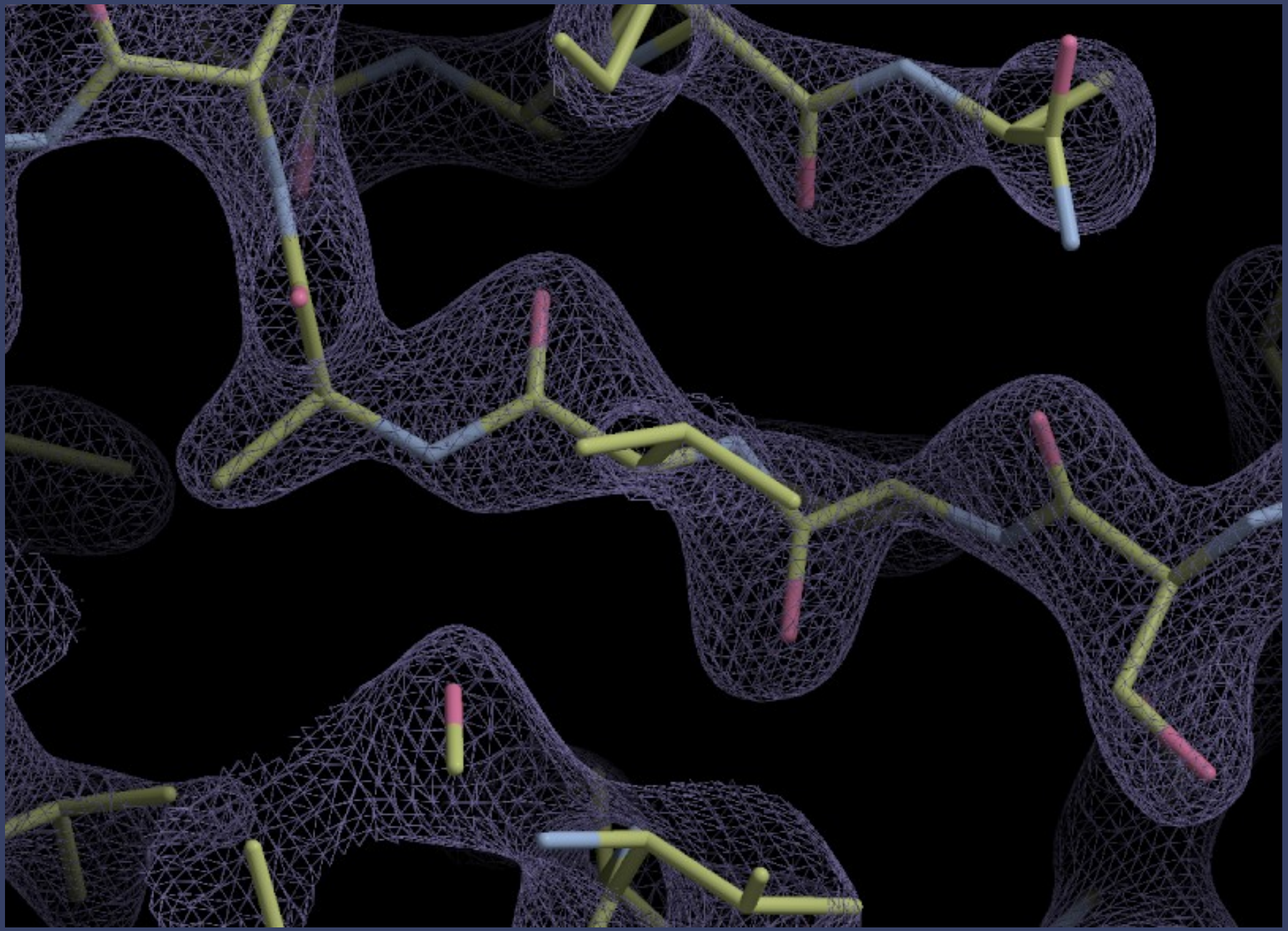
2.4Å

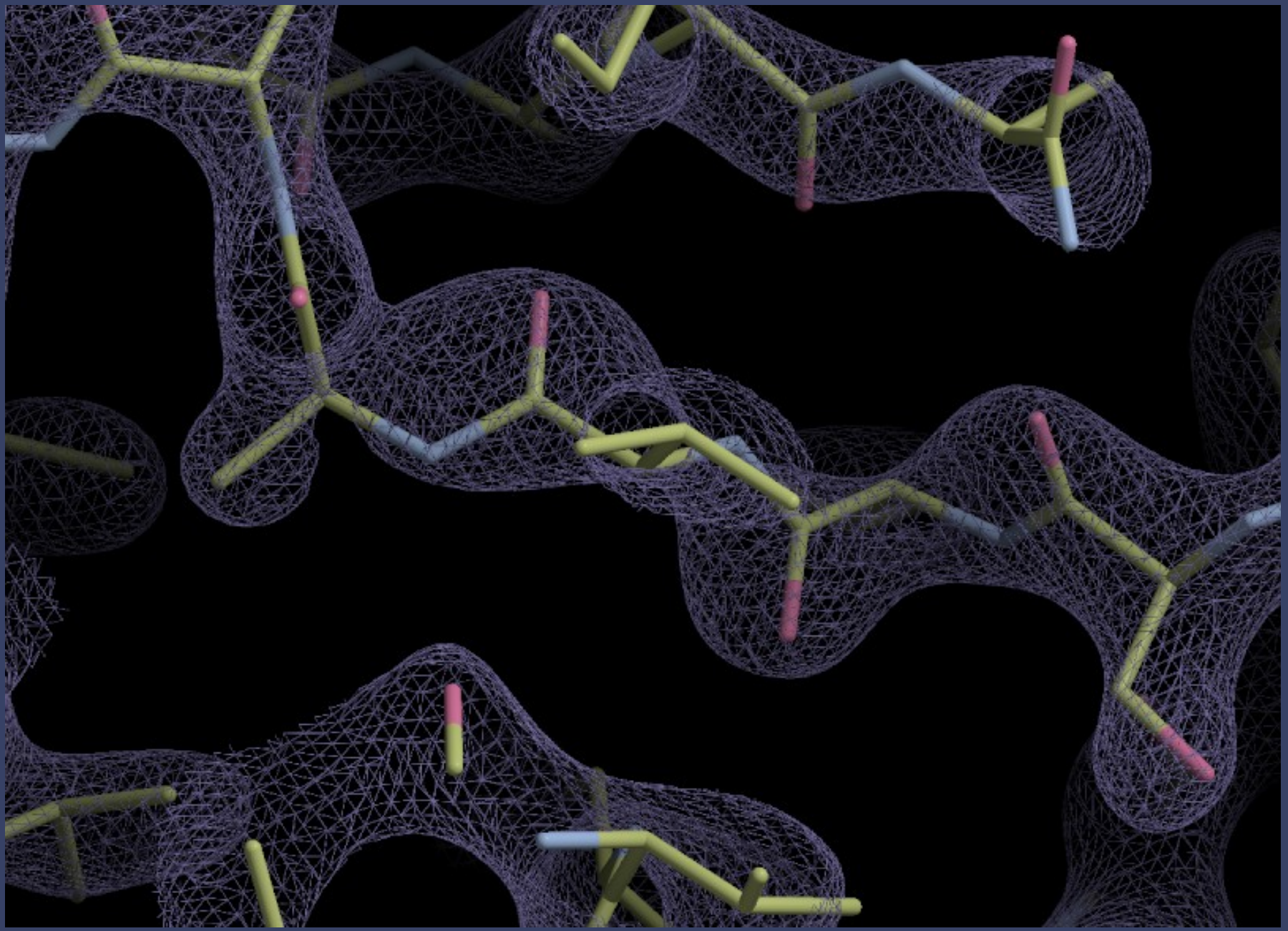


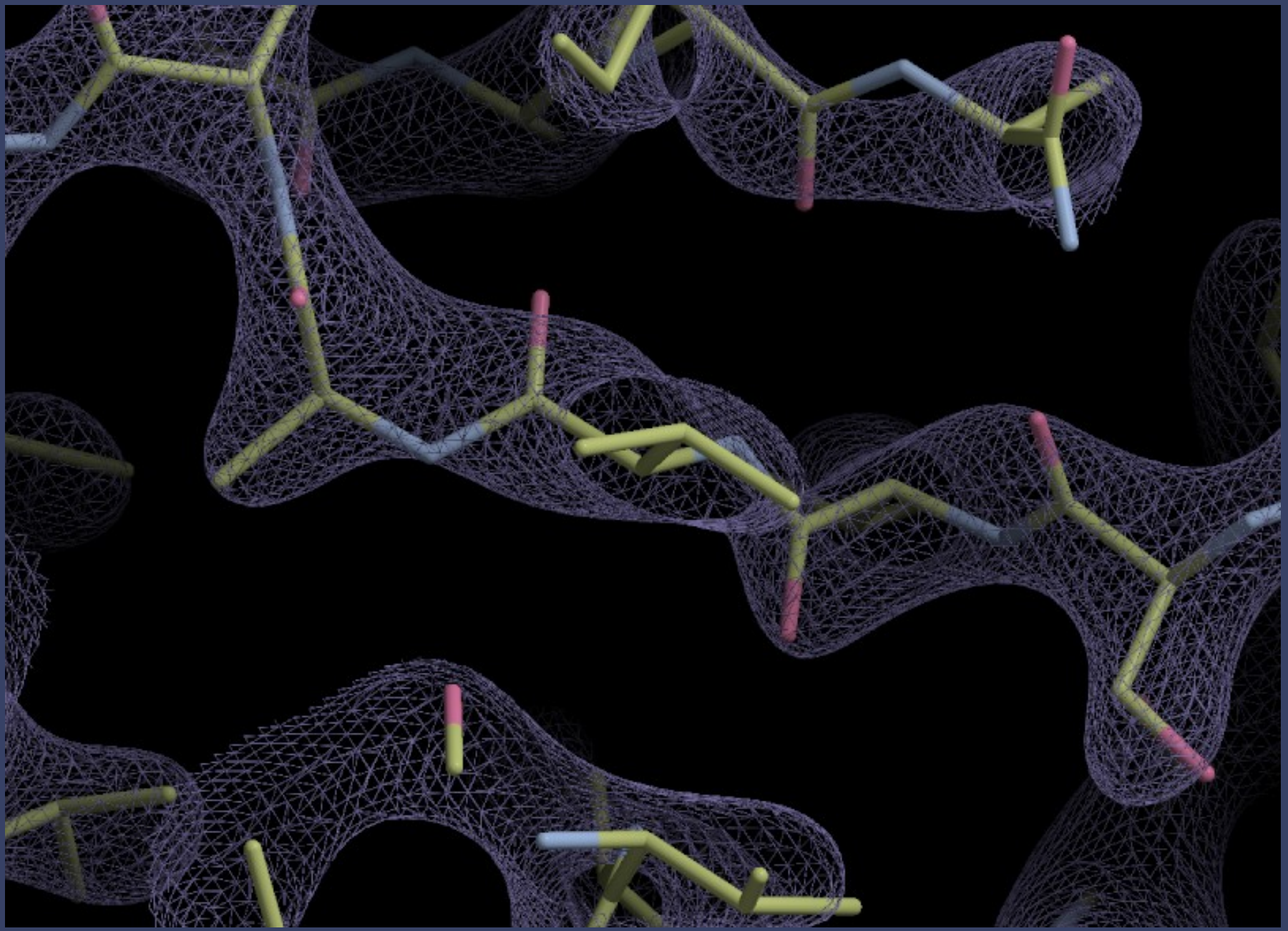
2.6Å



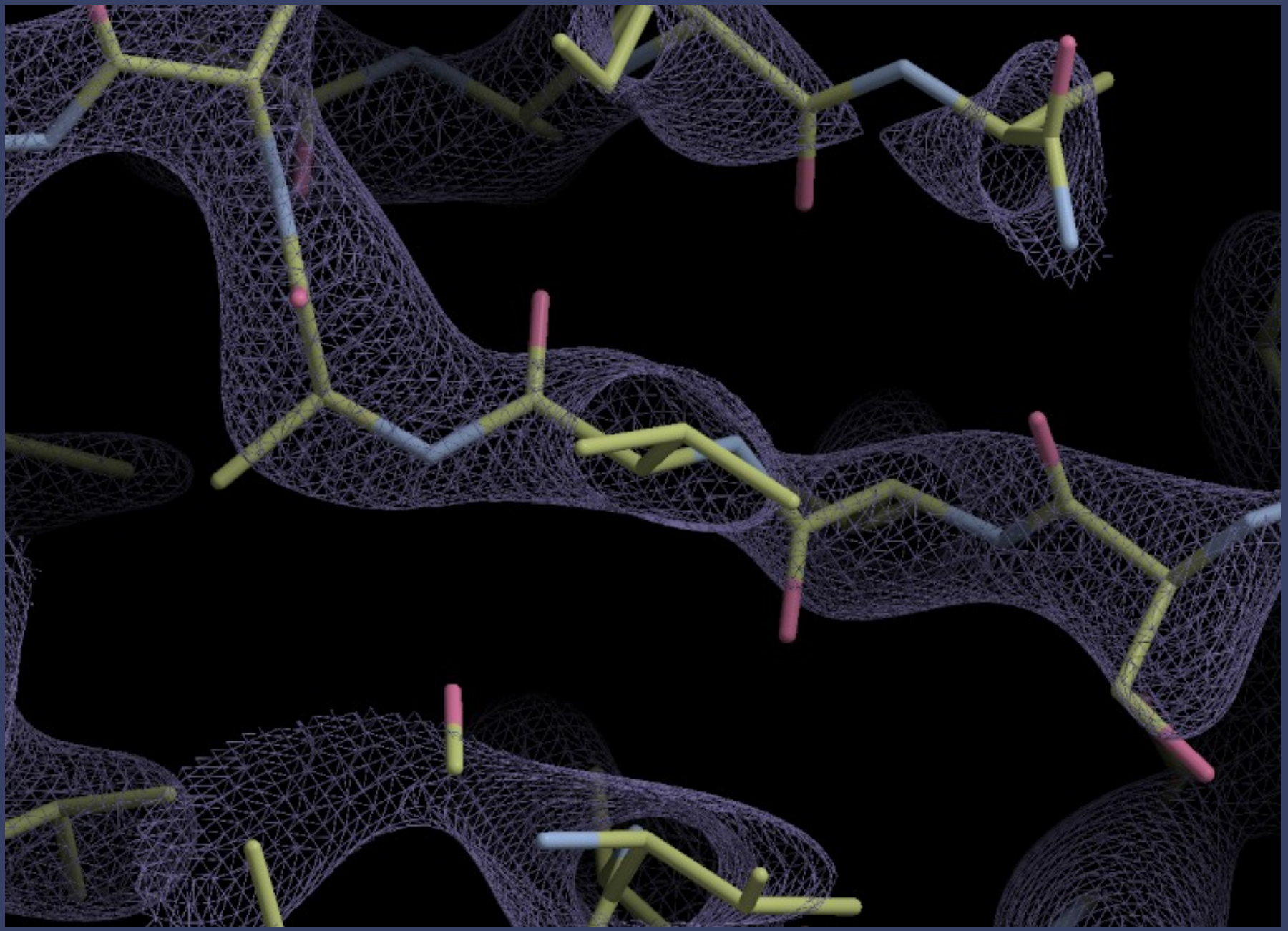
2.8Å



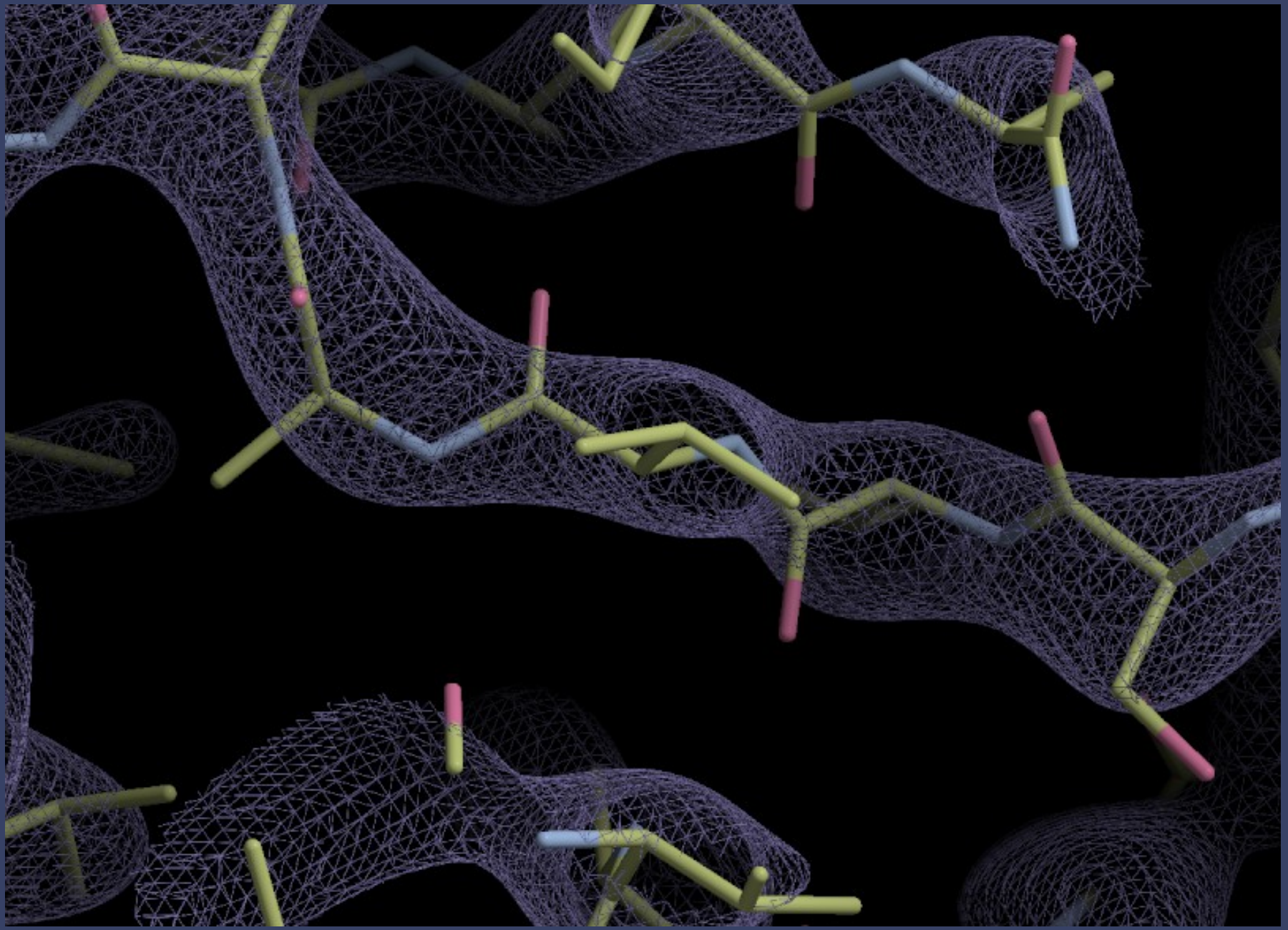

3.0Å


3.2Å

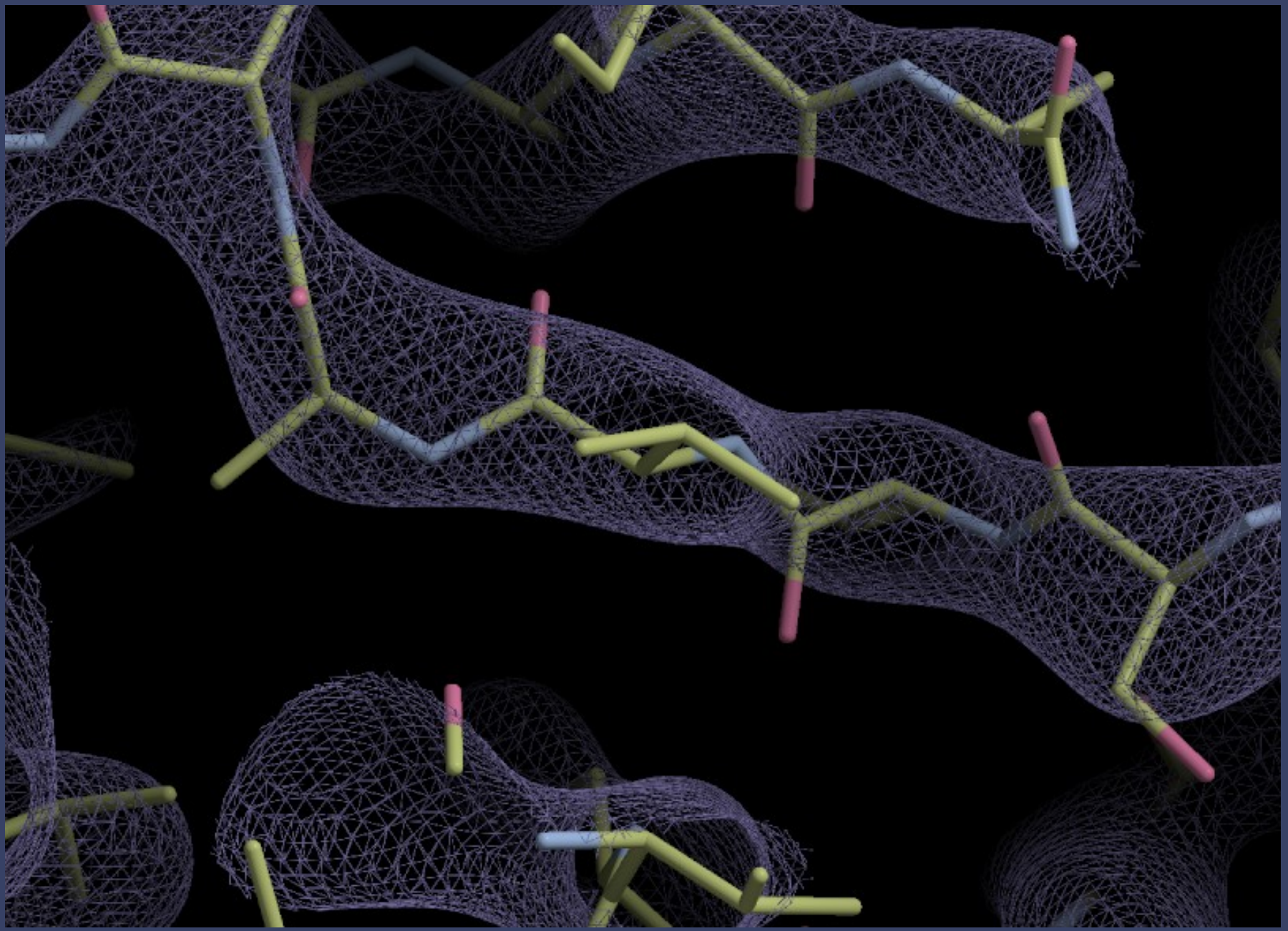
3.4Å

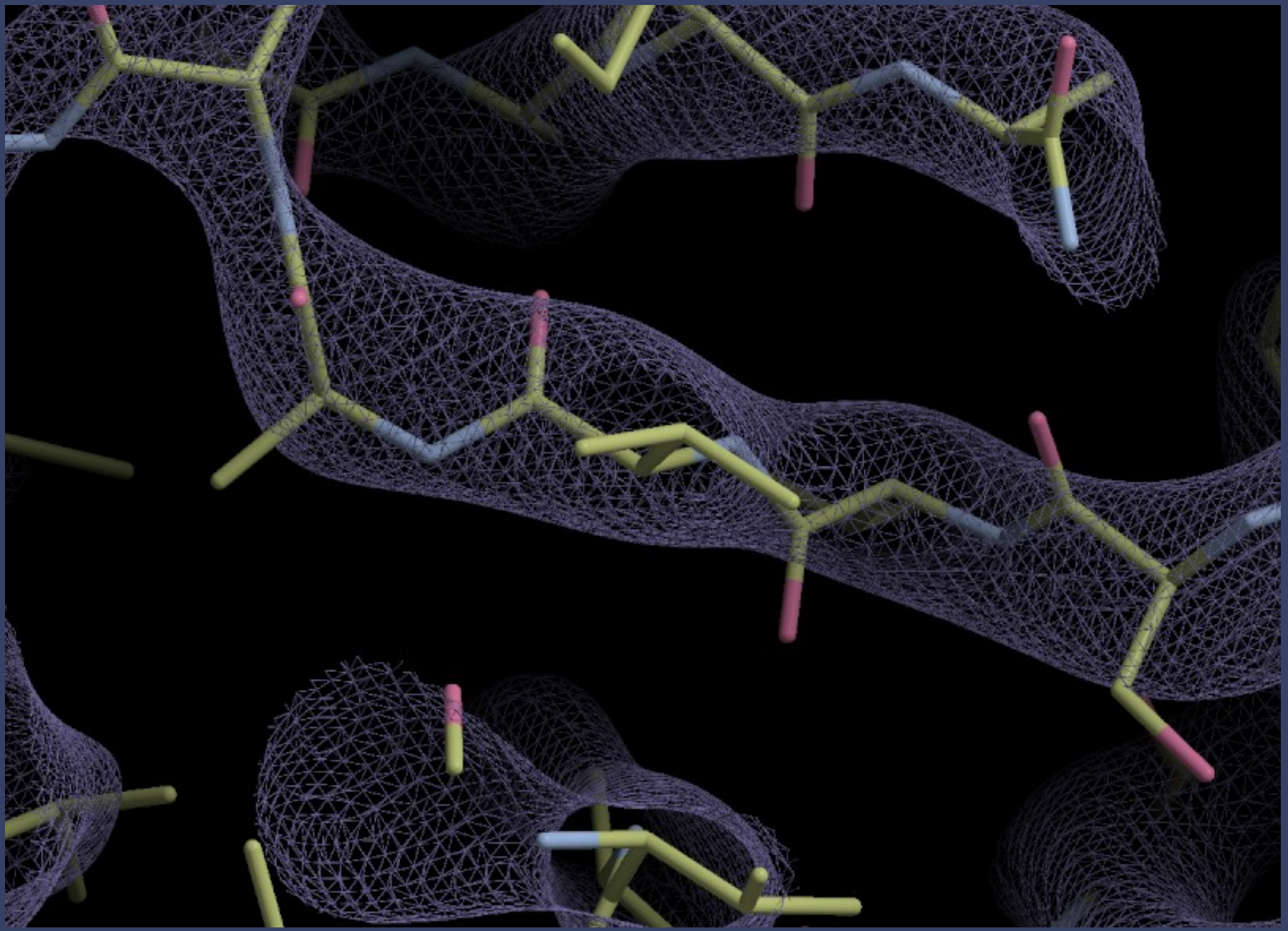


3.6Å

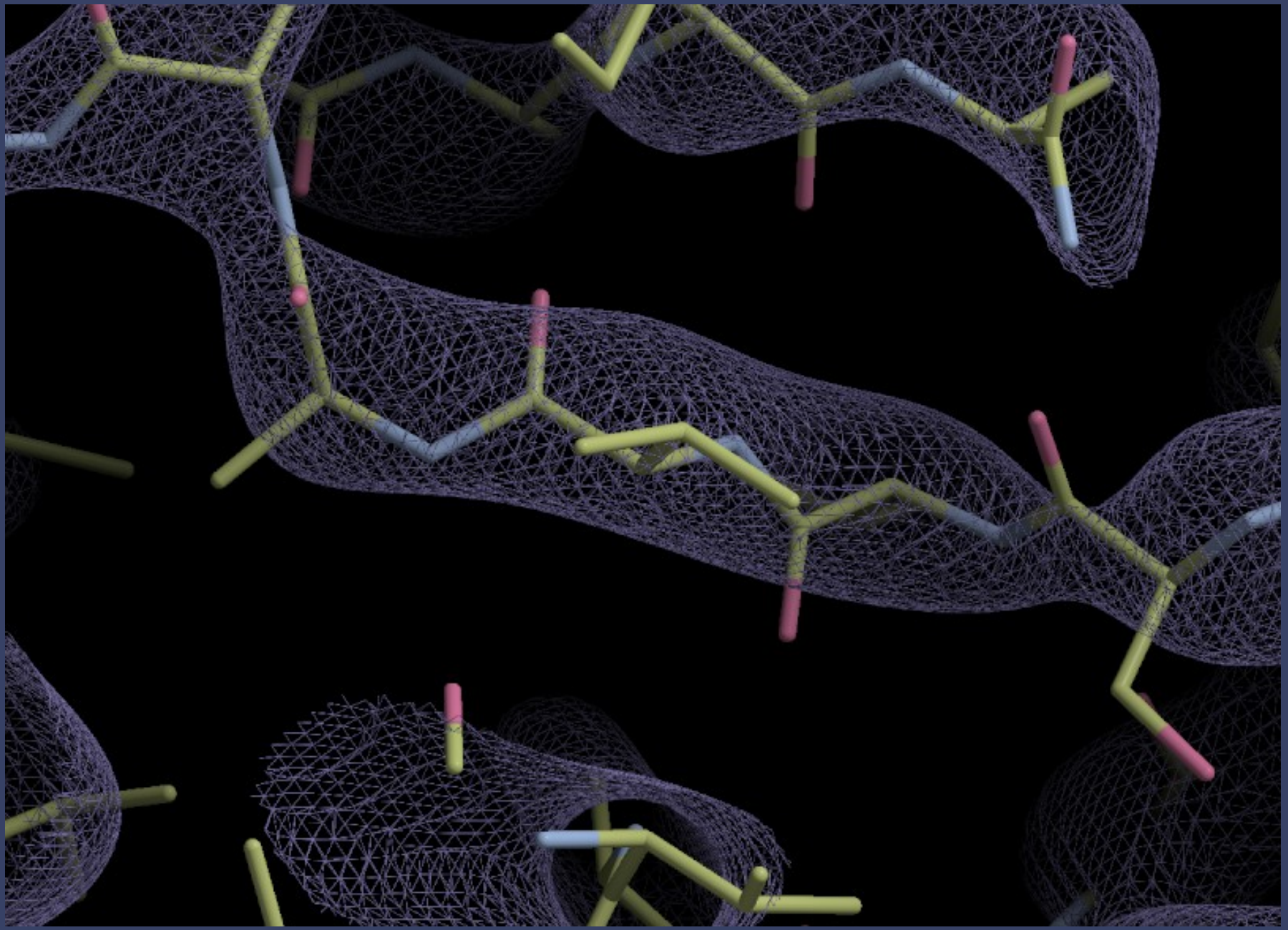


3.8Å

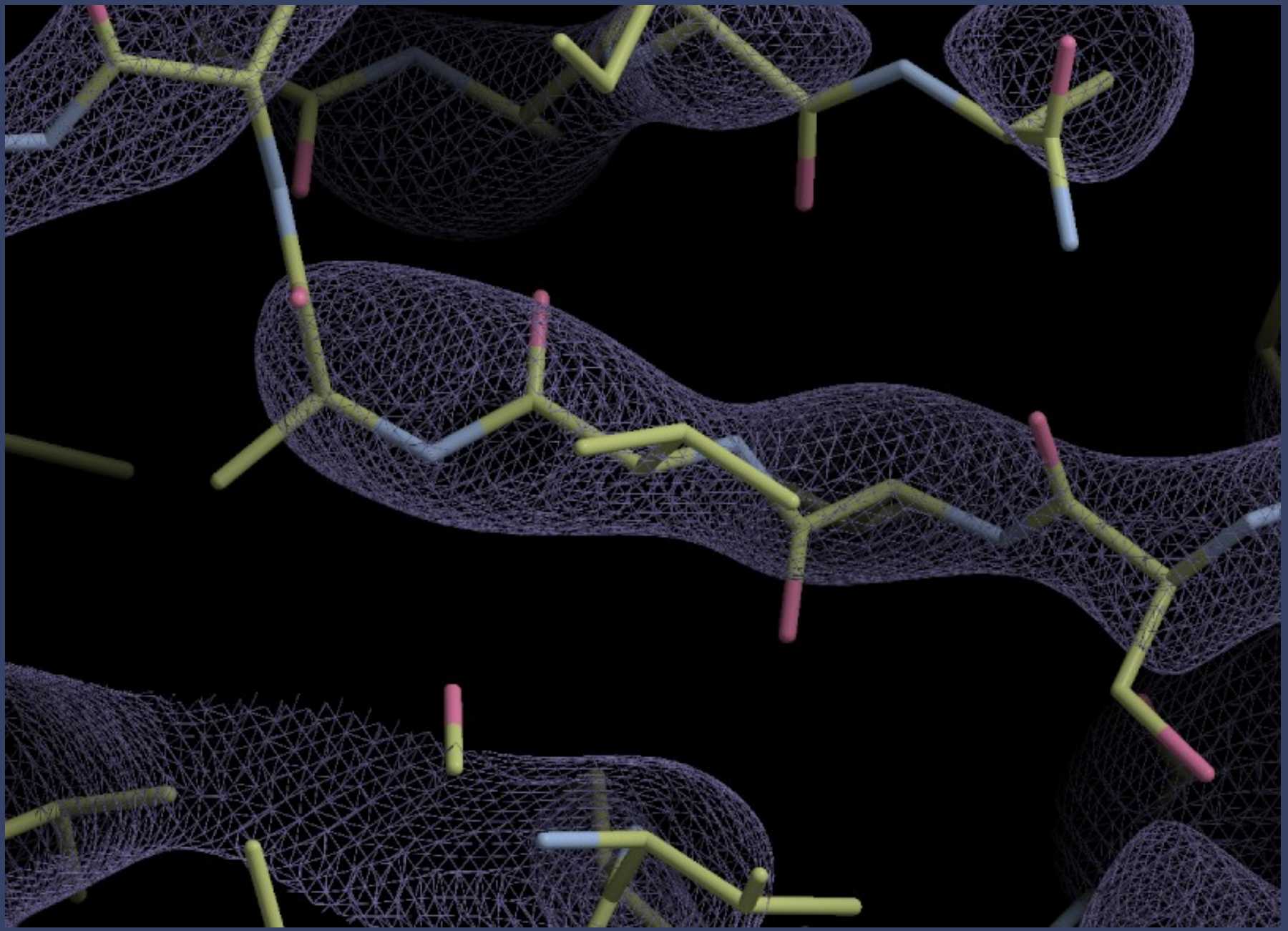



4.0Å

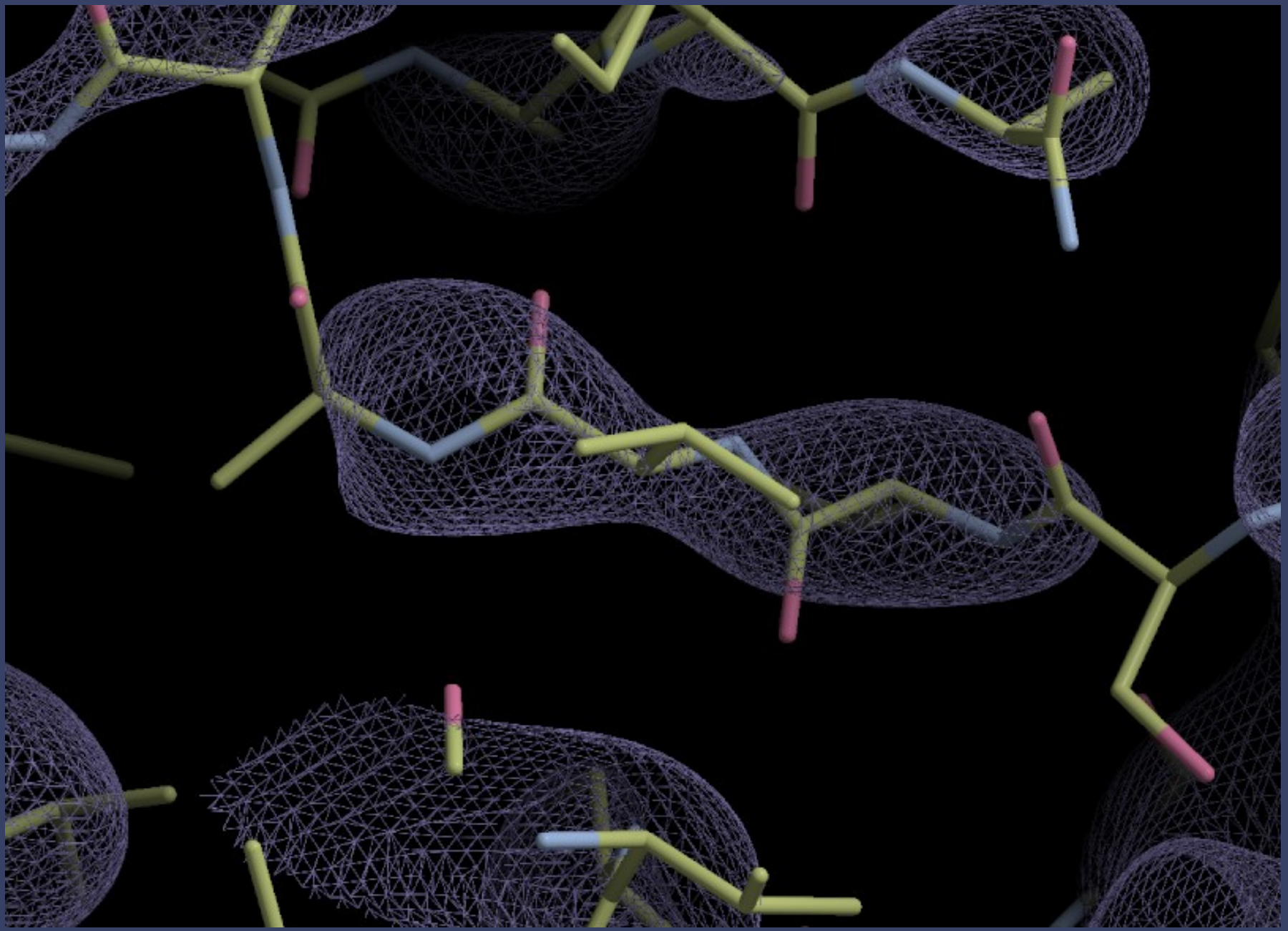
4.2Å



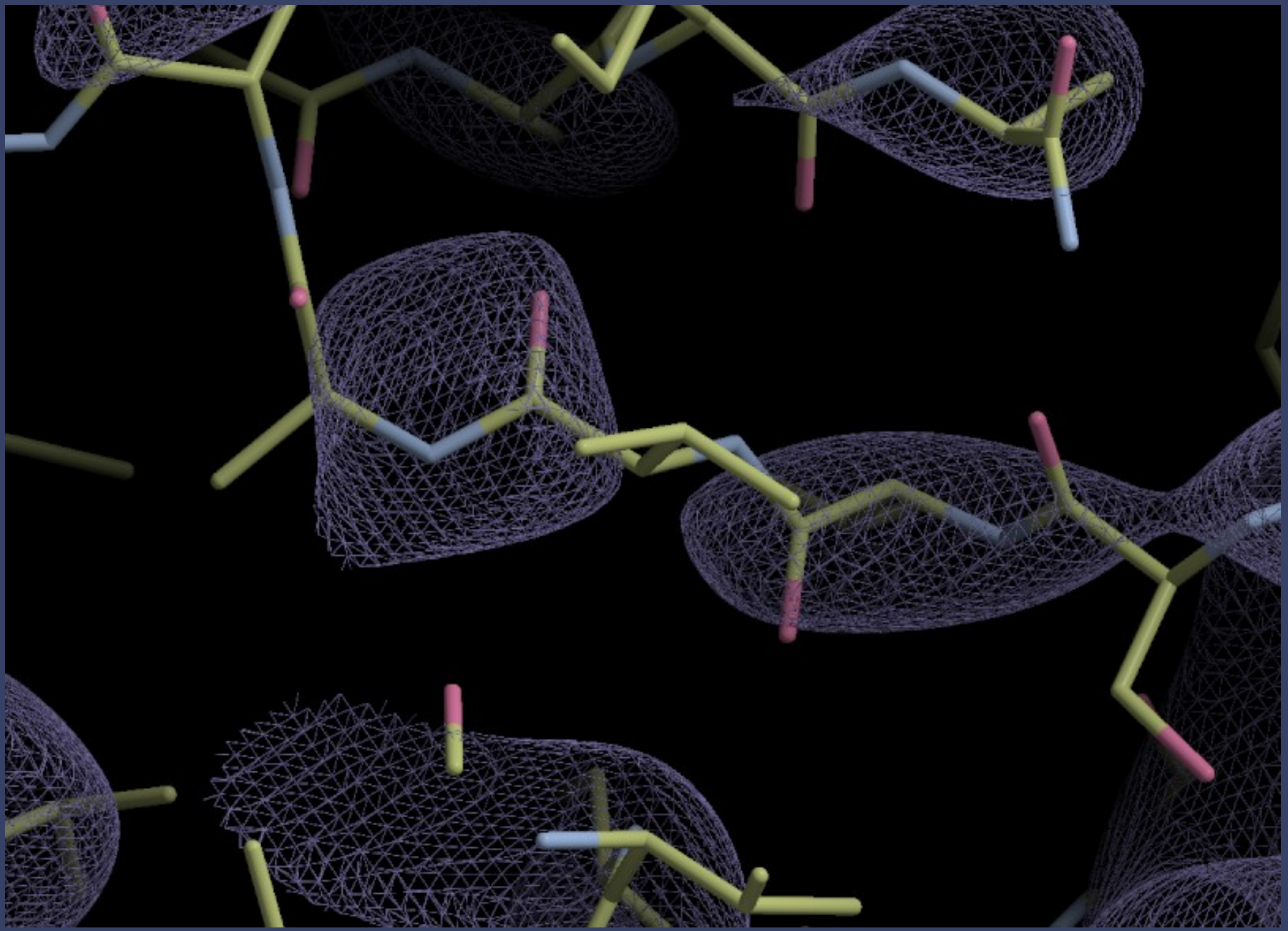
4.4Å



4.6Å



4.8Å



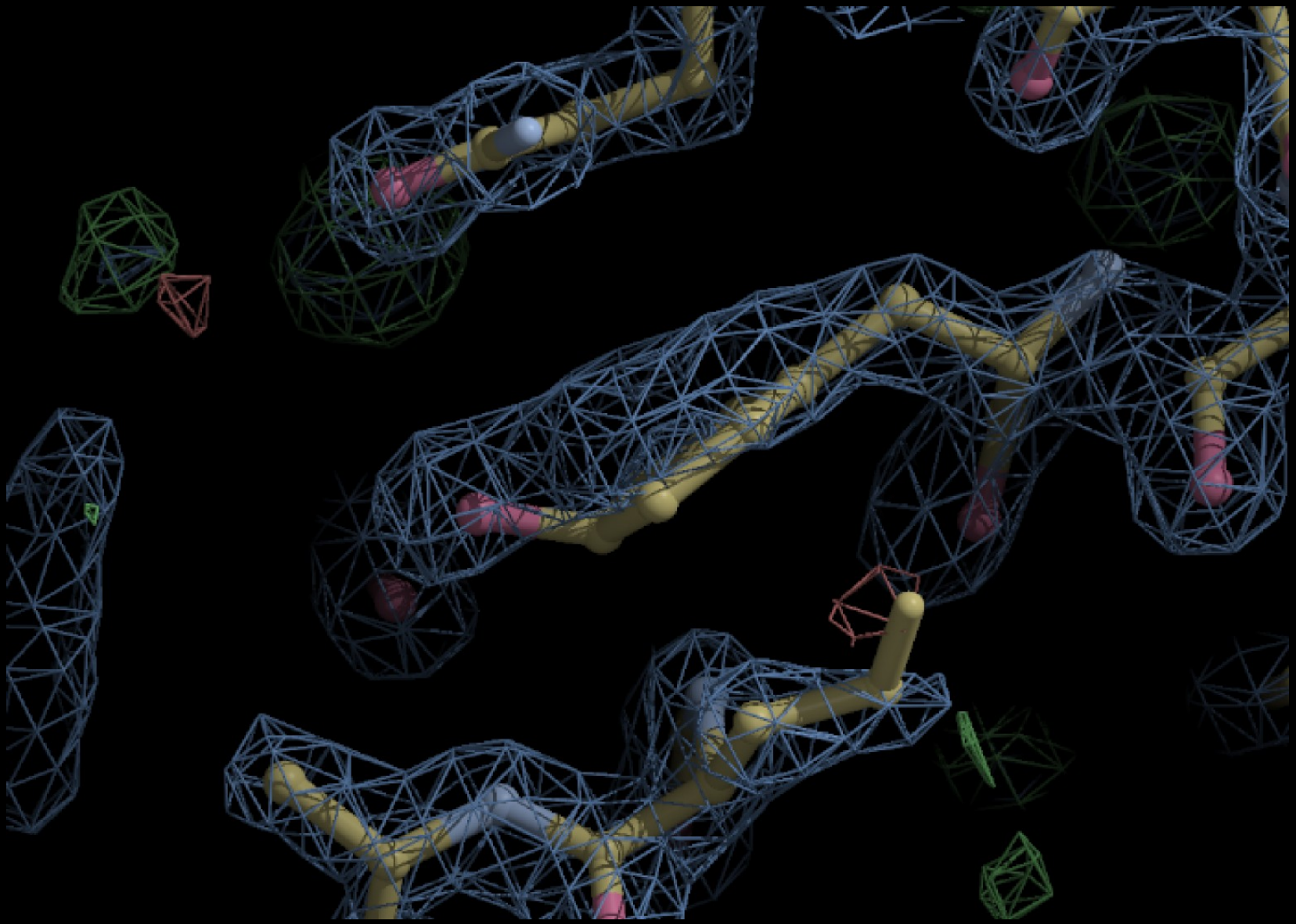
What is “Refinement”?

- 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

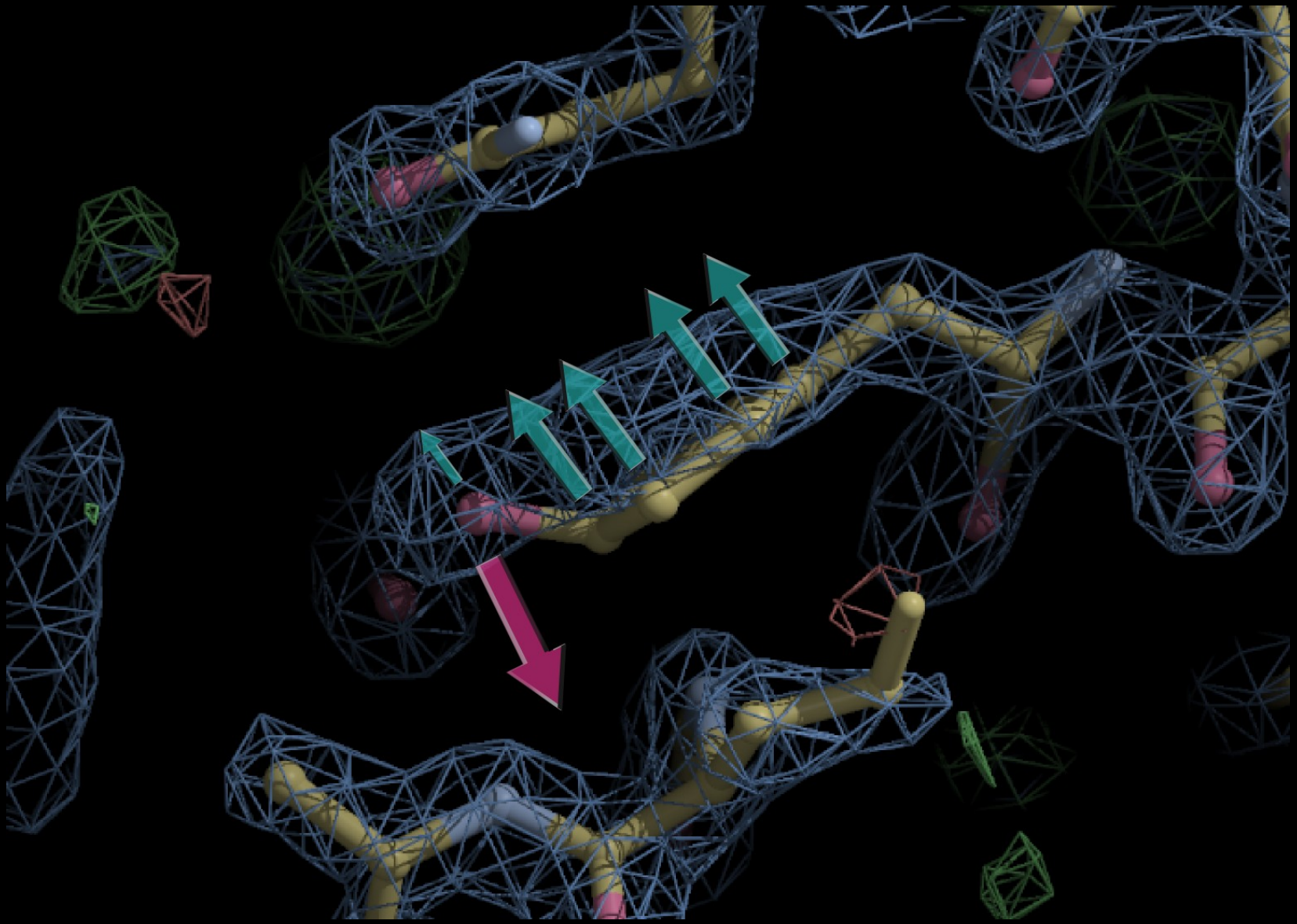
Real Space Refinement

- Major Feature of Coot
  - Gradient-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
  - e.g. Sphere Refine

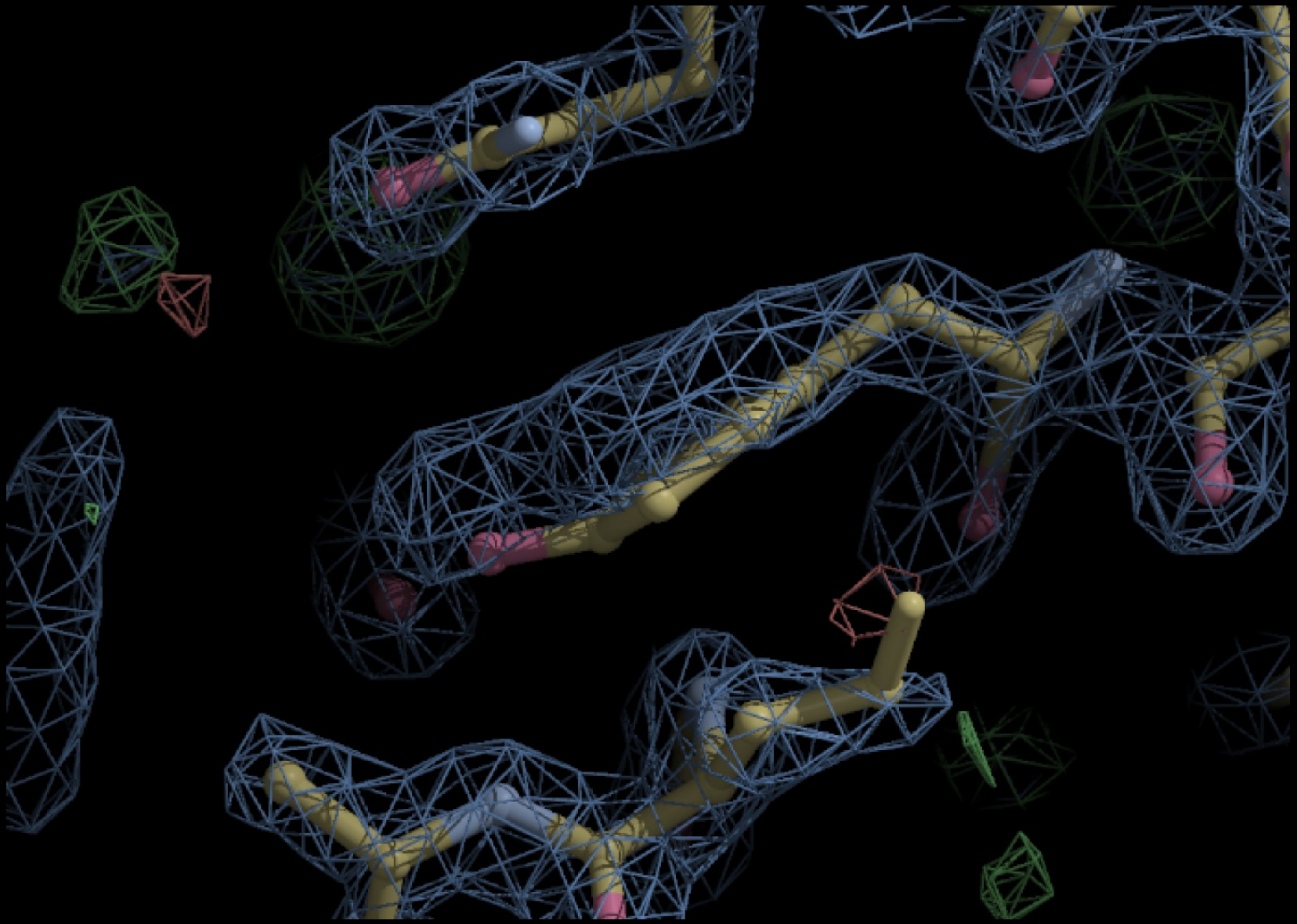
Distorted Geometry Pre-Refinement



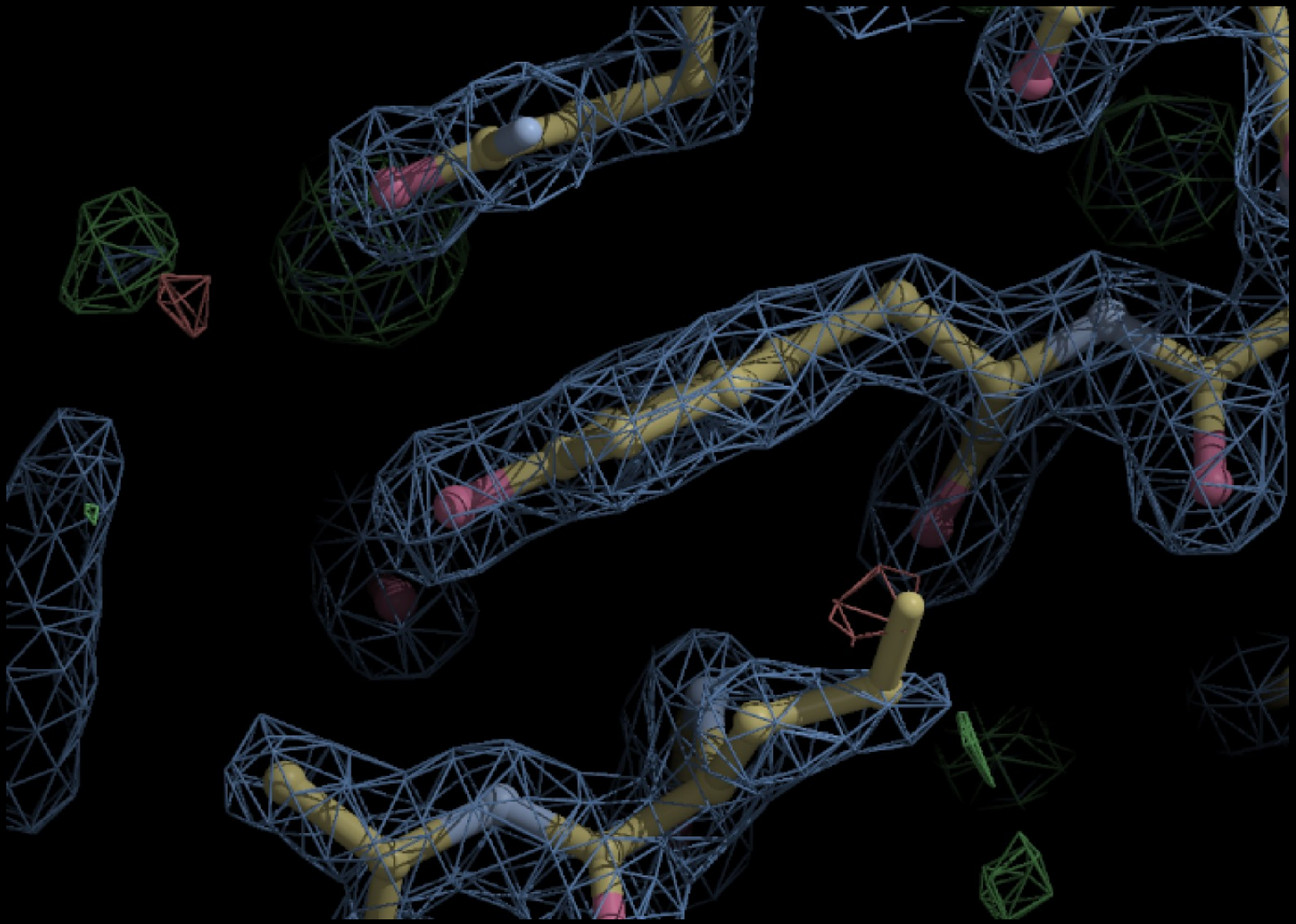
Refinement Gradients



Refinement: Cycle 3



Refinement Cycle 200: Minimized



Representation of Results:

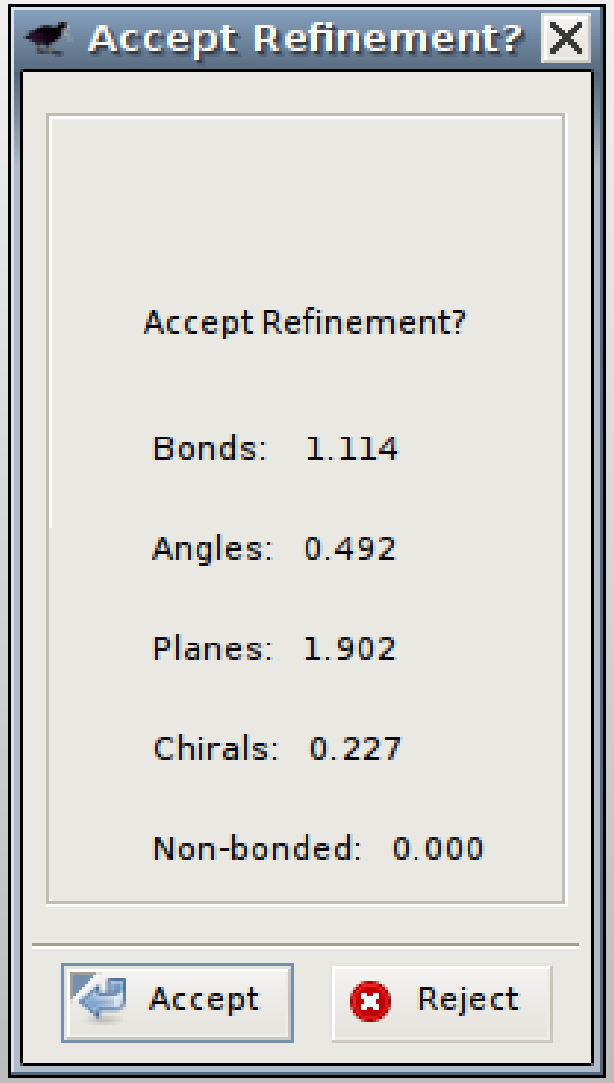
```
File  Edit  View  Terminal  Help
created 32 bond       restraints
created 38 angle      restraints
created 1 plane      restraints
created 5 chiral vol restraints
created 76 restraints

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

The first attempt

Student Reaction:

“Oh, I don't look at that window…”

Representation of Results:



Second attempt...

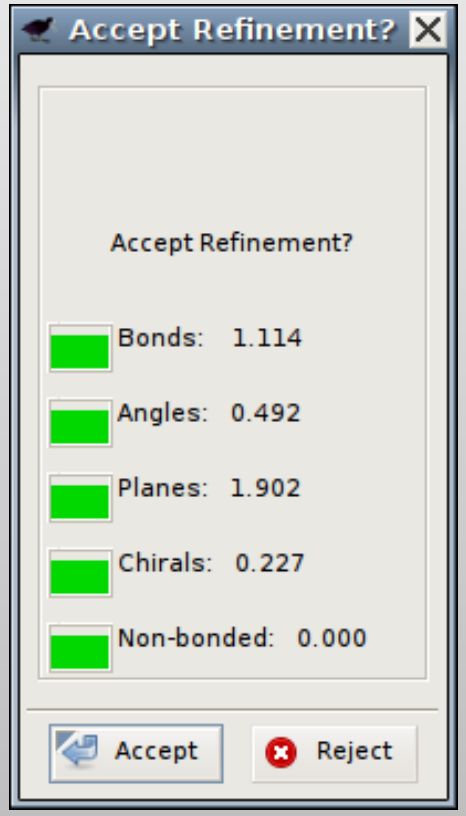
Student Reaction:

“Oh, box of meaningless numbers.

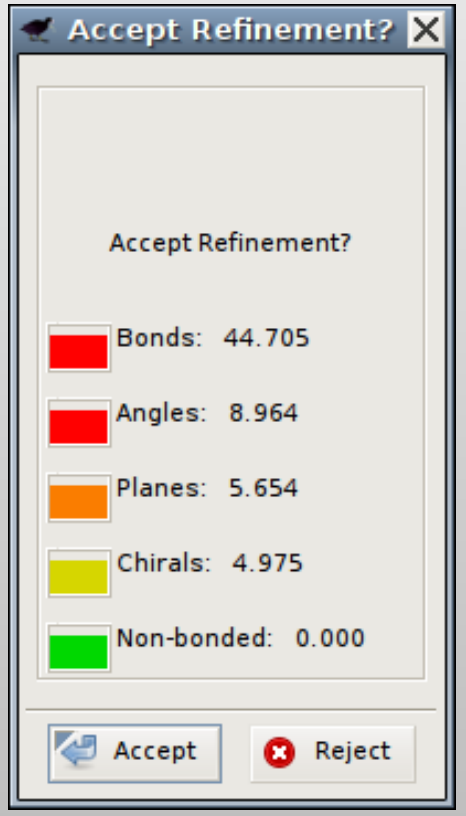
Go away”

Representation of Results: “Traffic Lights”

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



Good refinement



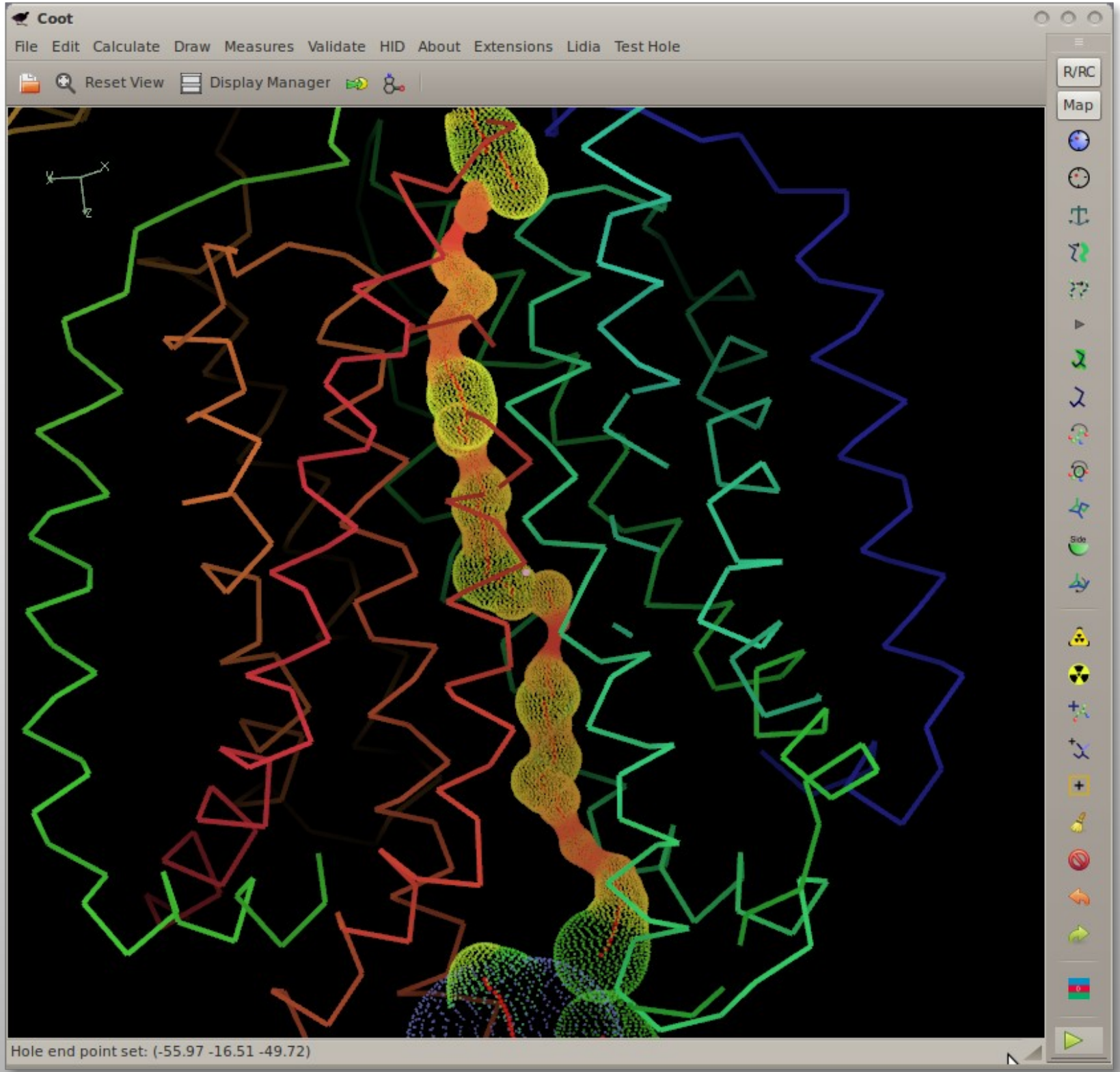
Bad refinement

Refinement Techniques

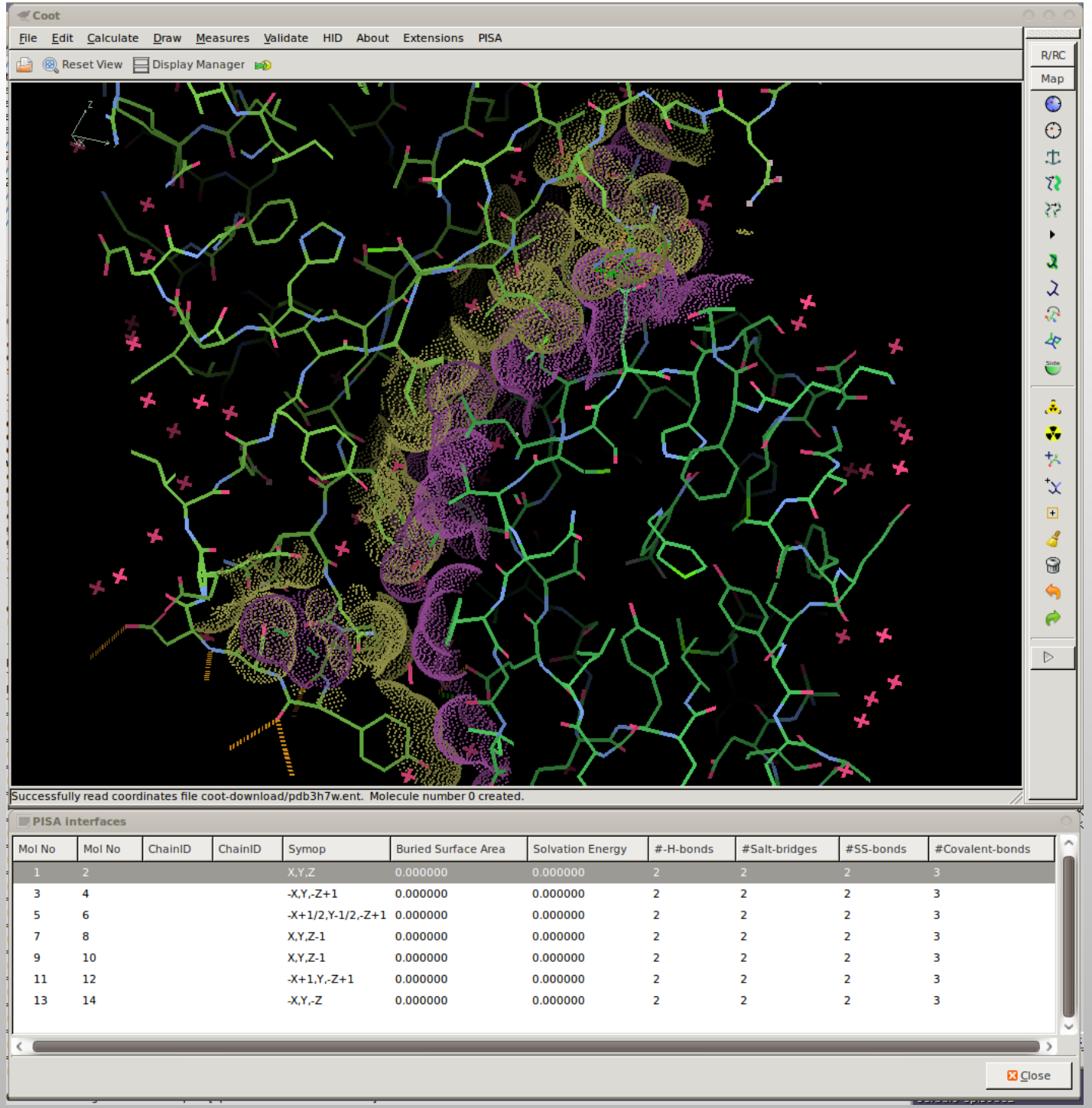
- Dragging an atom with Sphere Refine...
  - too much moves, so use:
- Single-Atom Drag
  - Over-dragging
- Key-bindings:
  - Triple Refine
  - Single Residue Refine with Auto-accept
- Ramachandran Refinement
  - Best done with hydrogens
- Parallel Plane Restraints

Finding Holes

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

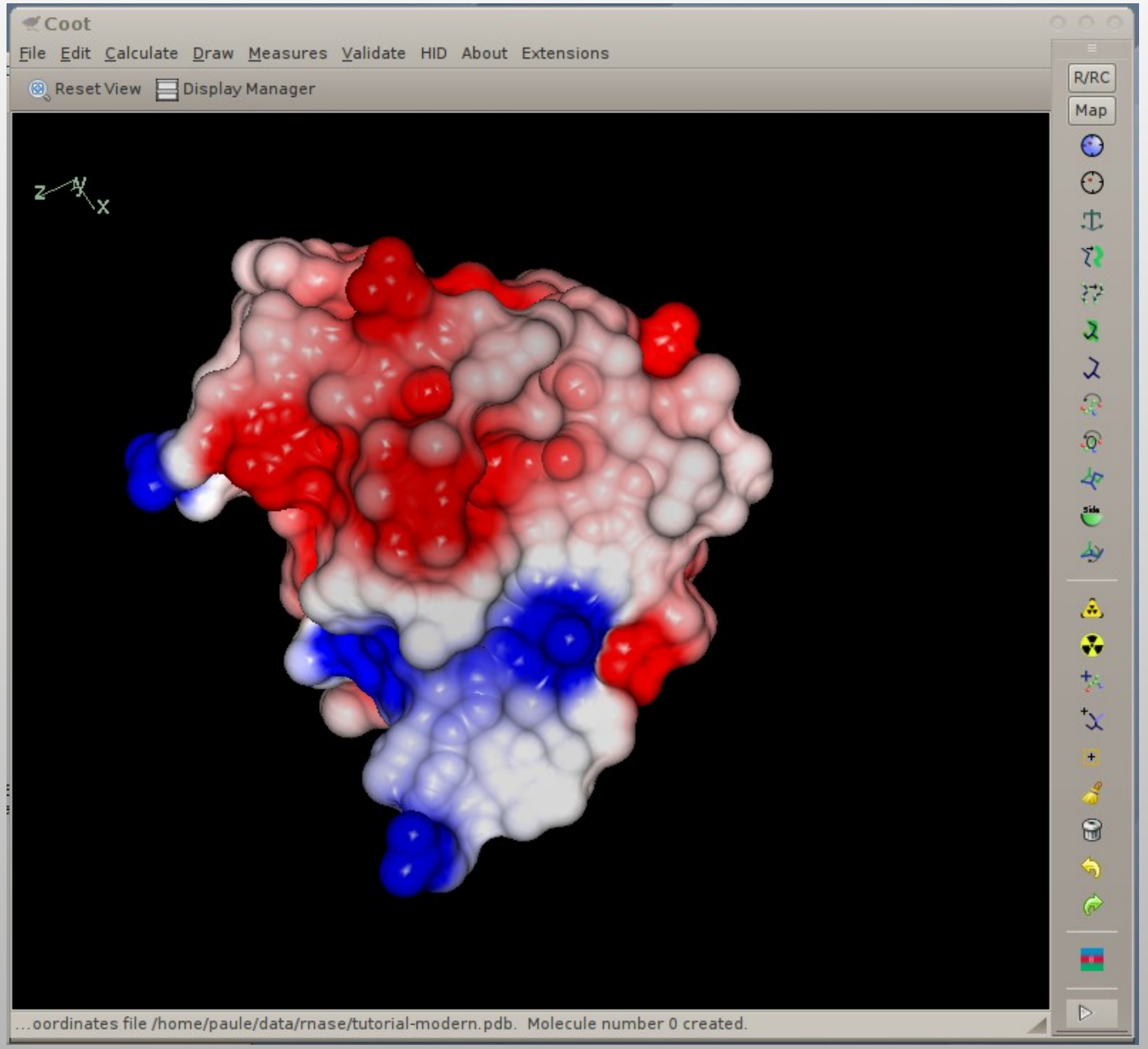

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia Test Hole
Reset View
Display Manager
R/RC
Map
Side
Hole end point set: (-55.97 -16.51 -49.72)

Interfaces and Assemblies: Interface to PISA

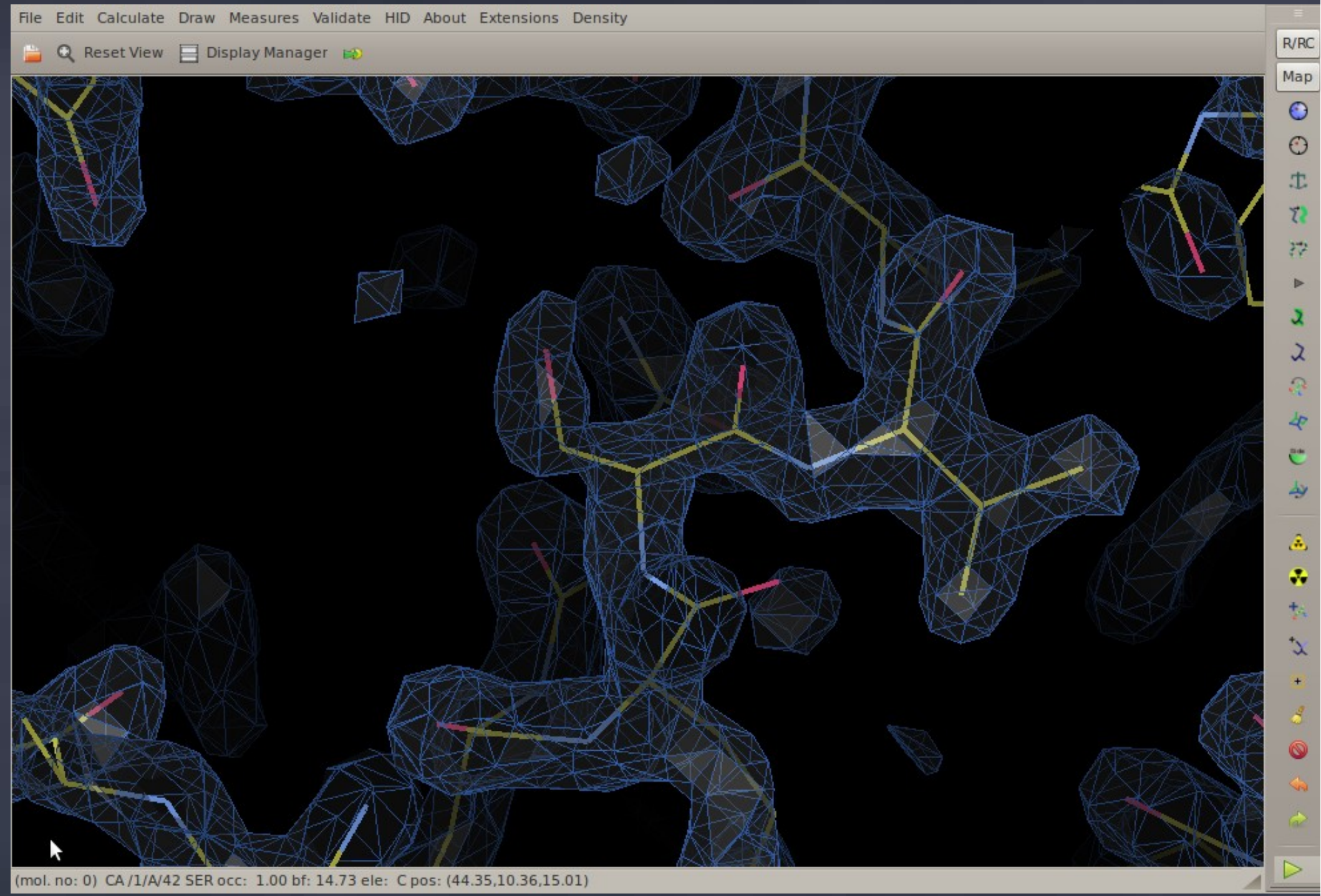


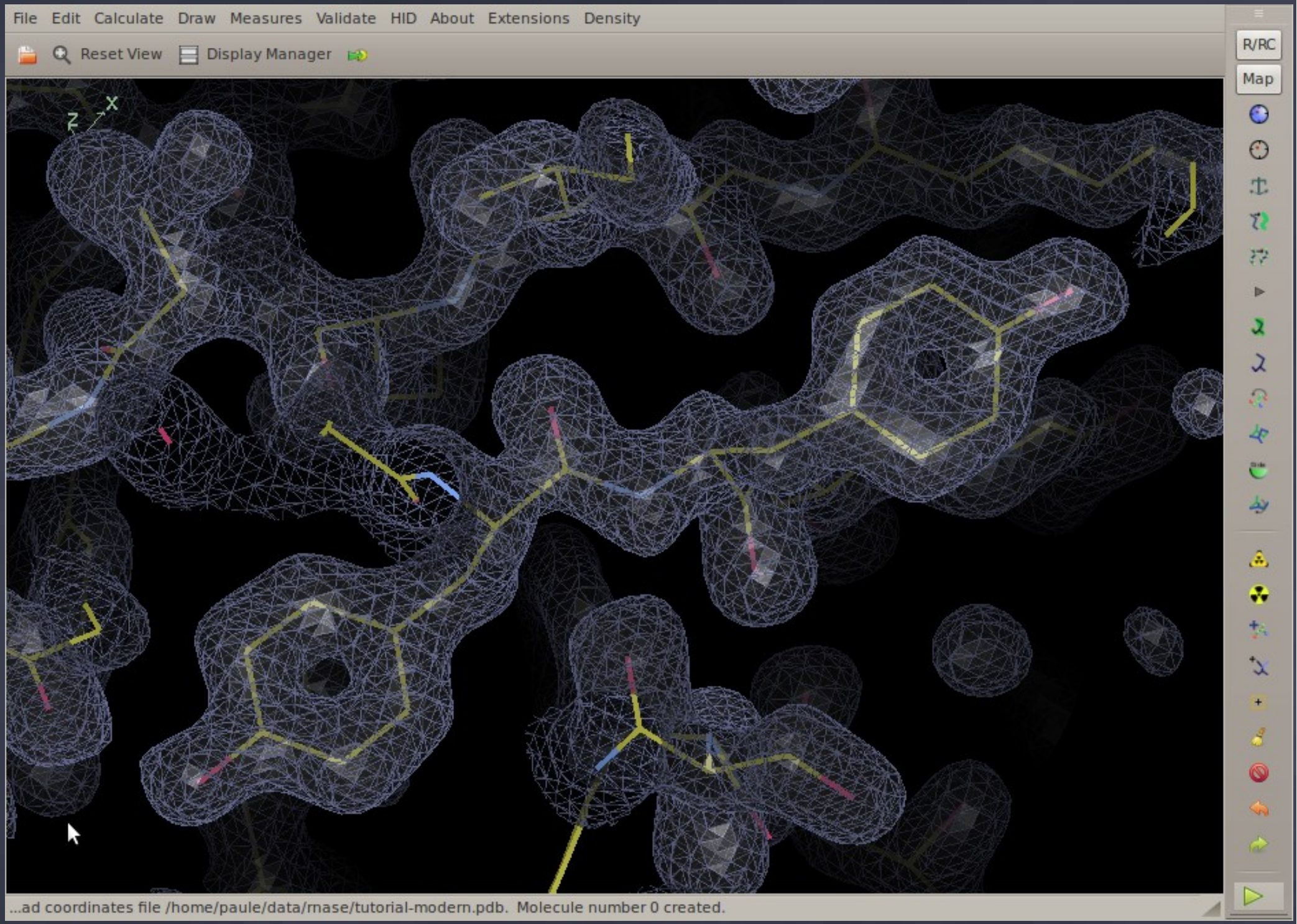
Mol No	Mol No	ChainID	ChainID	Symop	Buried Surface Area	Solvation Energy	#-H-bonds	#Salt-bridges	#SS-bonds	#Covalent-bonds
1	2			X,Y,Z	0.000000	0.000000	2	2	2	3
3	4			-X,Y,-Z+1	0.000000	0.000000	2	2	2	3
5	6			-X+1/2,Y-1/2,-Z+1	0.000000	0.000000	2	2	2	3
7	8			X,Y,Z-1	0.000000	0.000000	2	2	2	3
9	10			X,Y,Z-1	0.000000	0.000000	2	2	2	3
11	12			-X+1,Y,-Z+1	0.000000	0.000000	2	2	2	3
13	14			-X,Y,-Z	0.000000	0.000000	2	2	2	3

Some Representation Tools



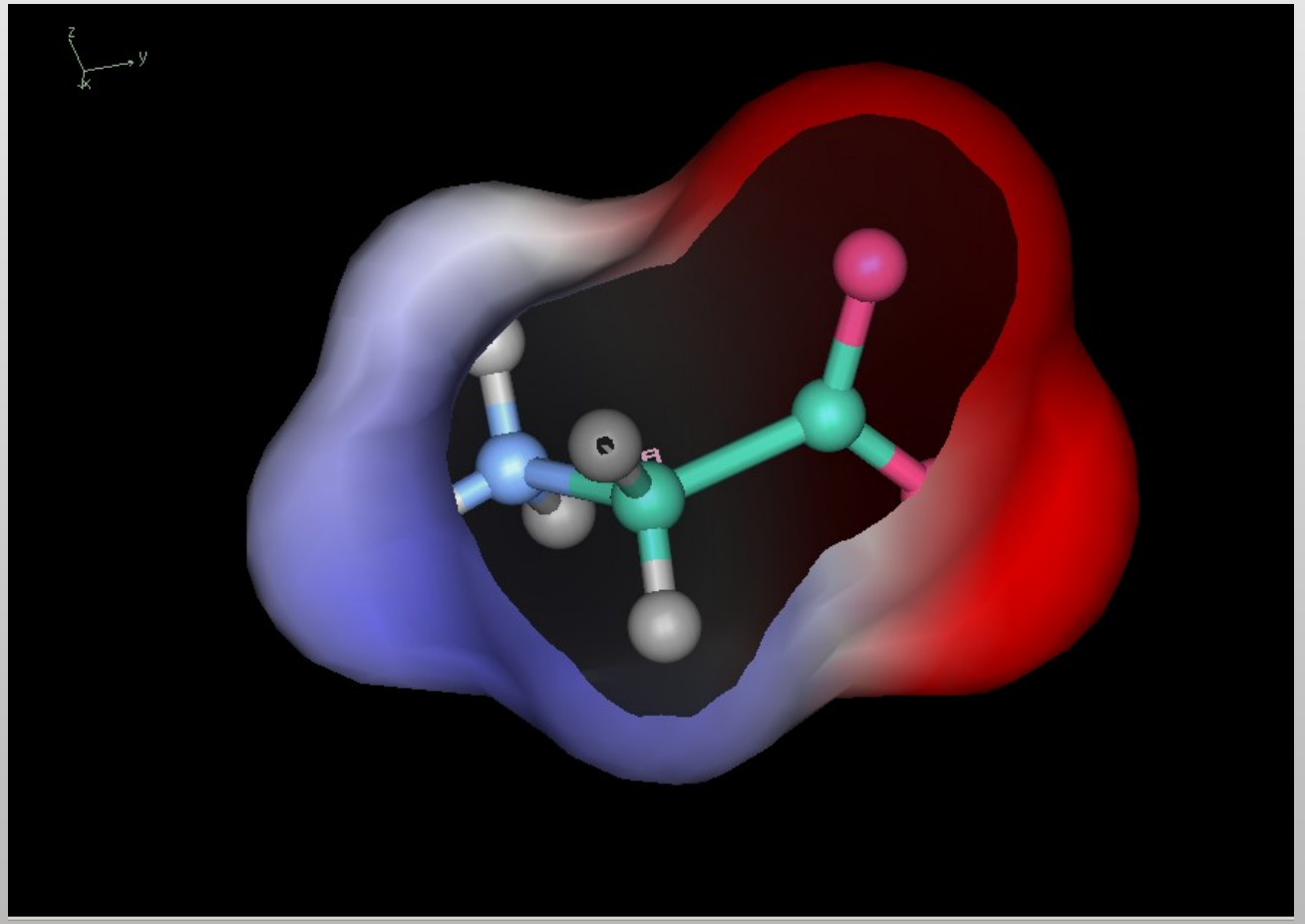
Gruber & Noble (2007)


File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
(mol. no: 0) CA /1/A/42 SER occ: 1.00 bf: 14.73 ele: C pos: (44.35,10.36,15.01)

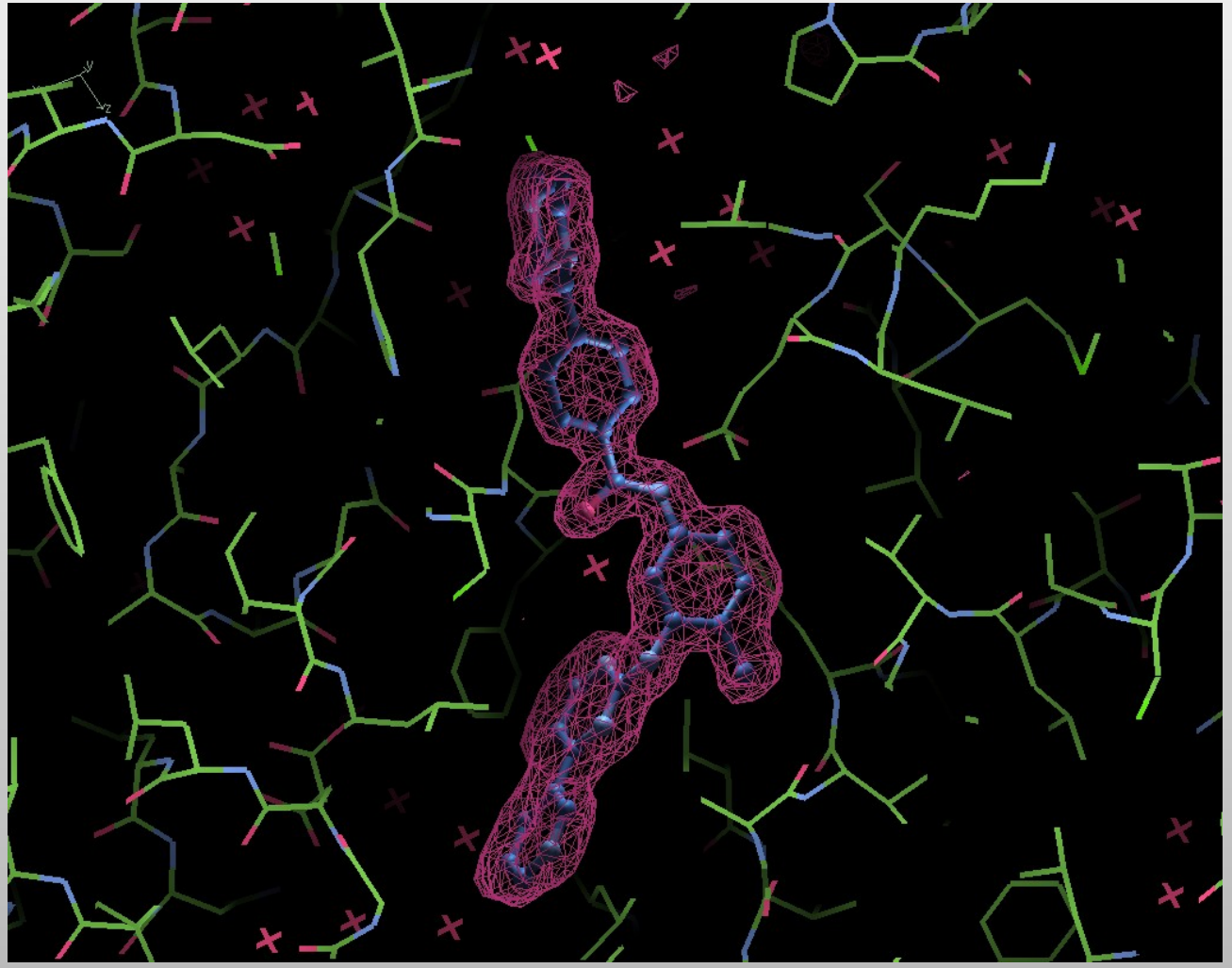

File Edit Calculate Draw Measures Validate HID About Extensions Density
Reset View
Display Manager
R/RC
Map
Z
X
...ad coordinates file /home/paule/data/rnase/tutorial-modern.pdb. Molecule number 0 created.

Other Things

- Surfaces that use dictionary partial charges

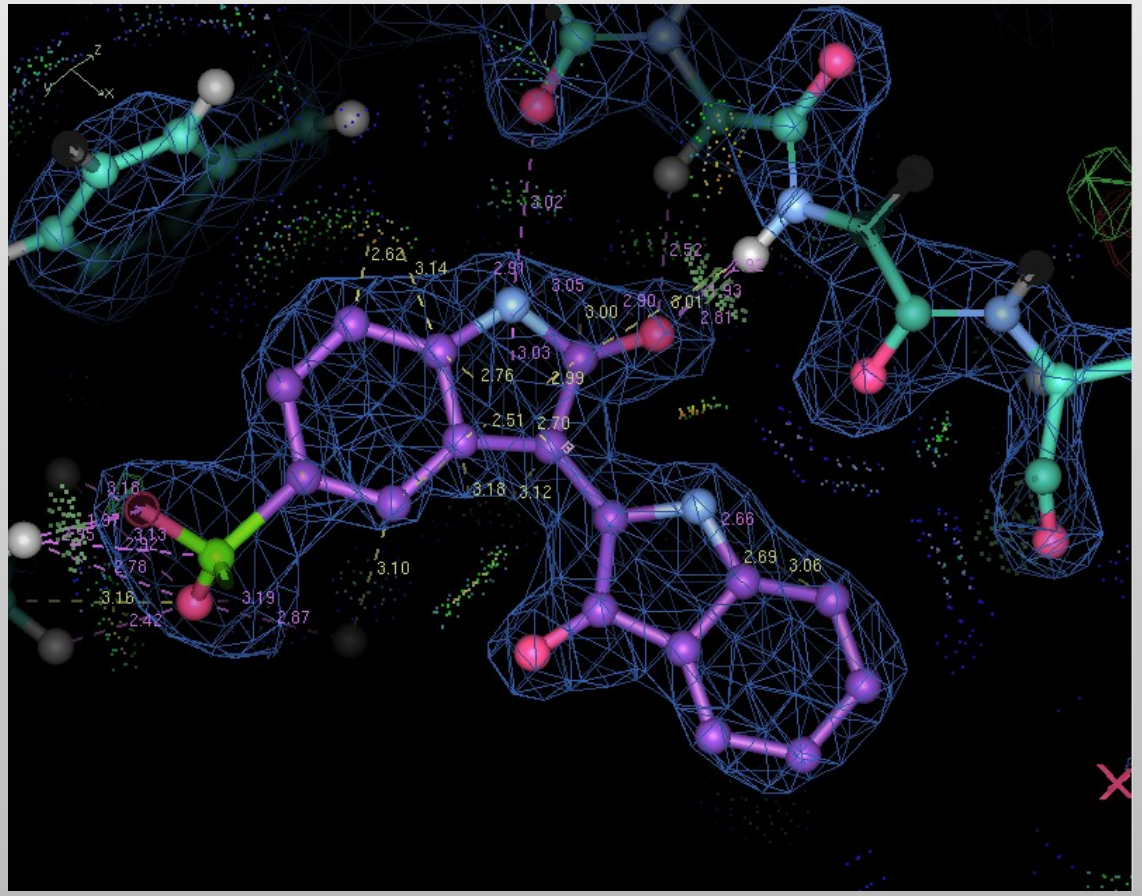


Some Representation Tools



Other Things

- Molprobity dots for ligands
  - Highlight interesting site



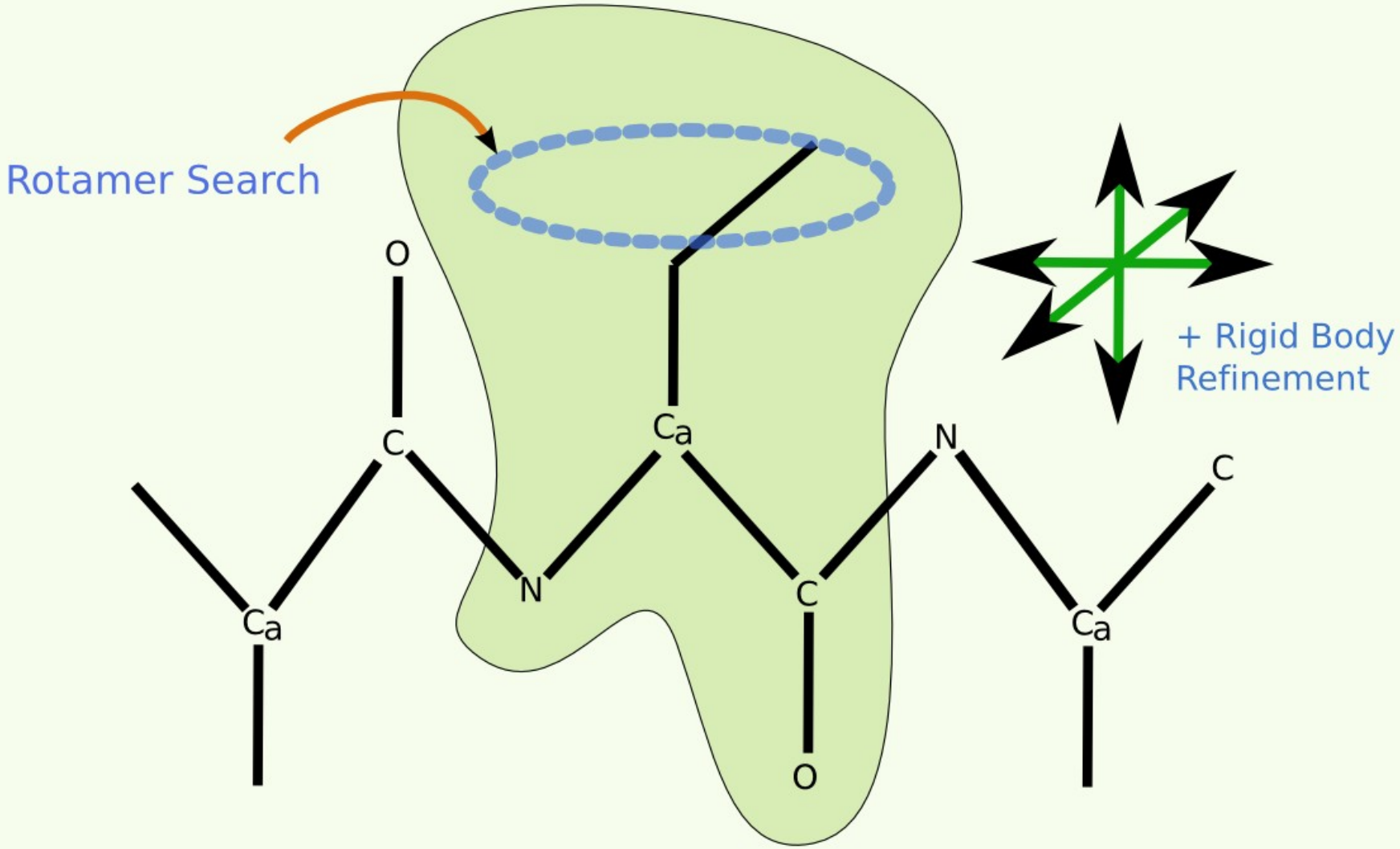
Tools for Cryo-EM and Low-Resolution

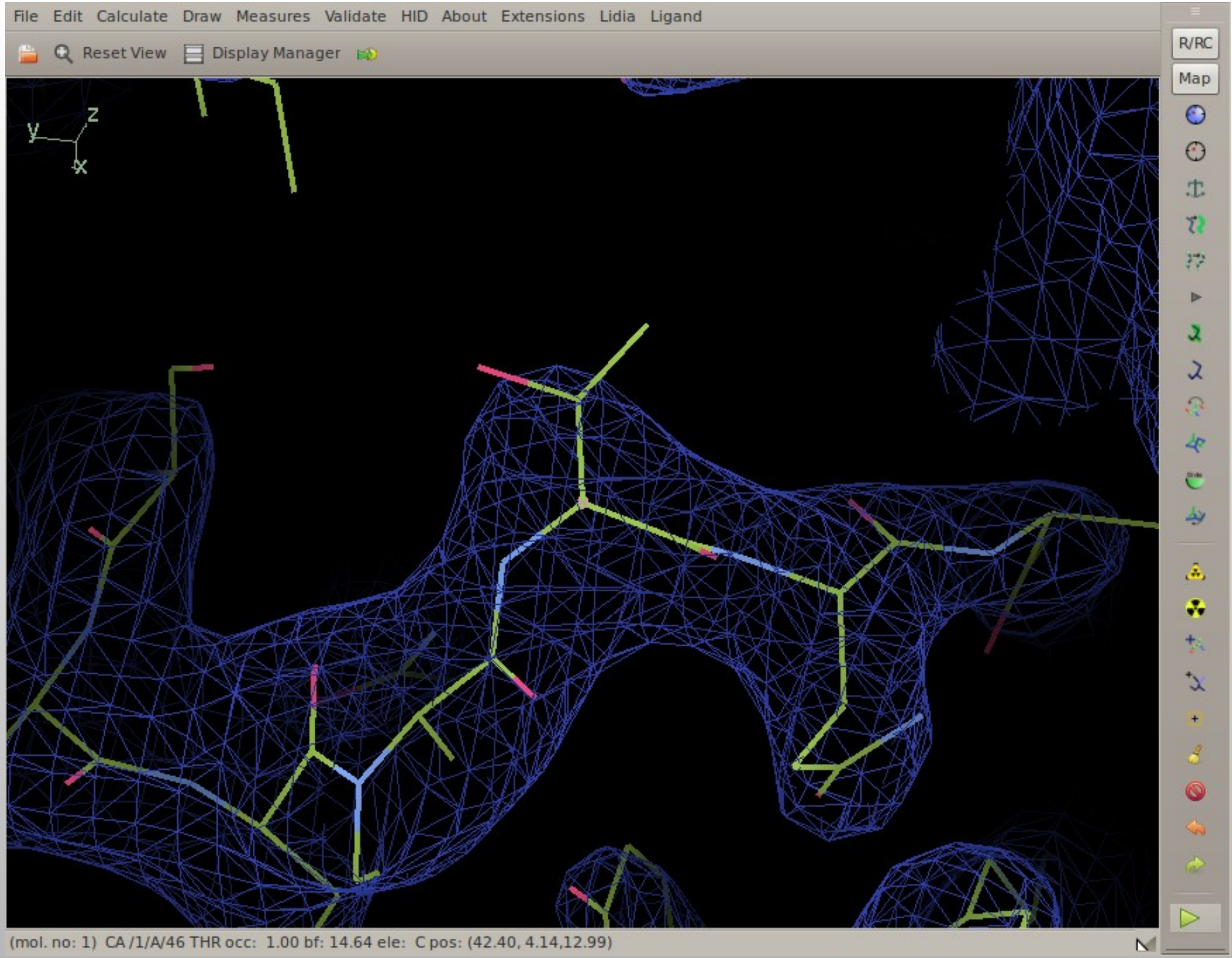
”Backrub Rotamers”

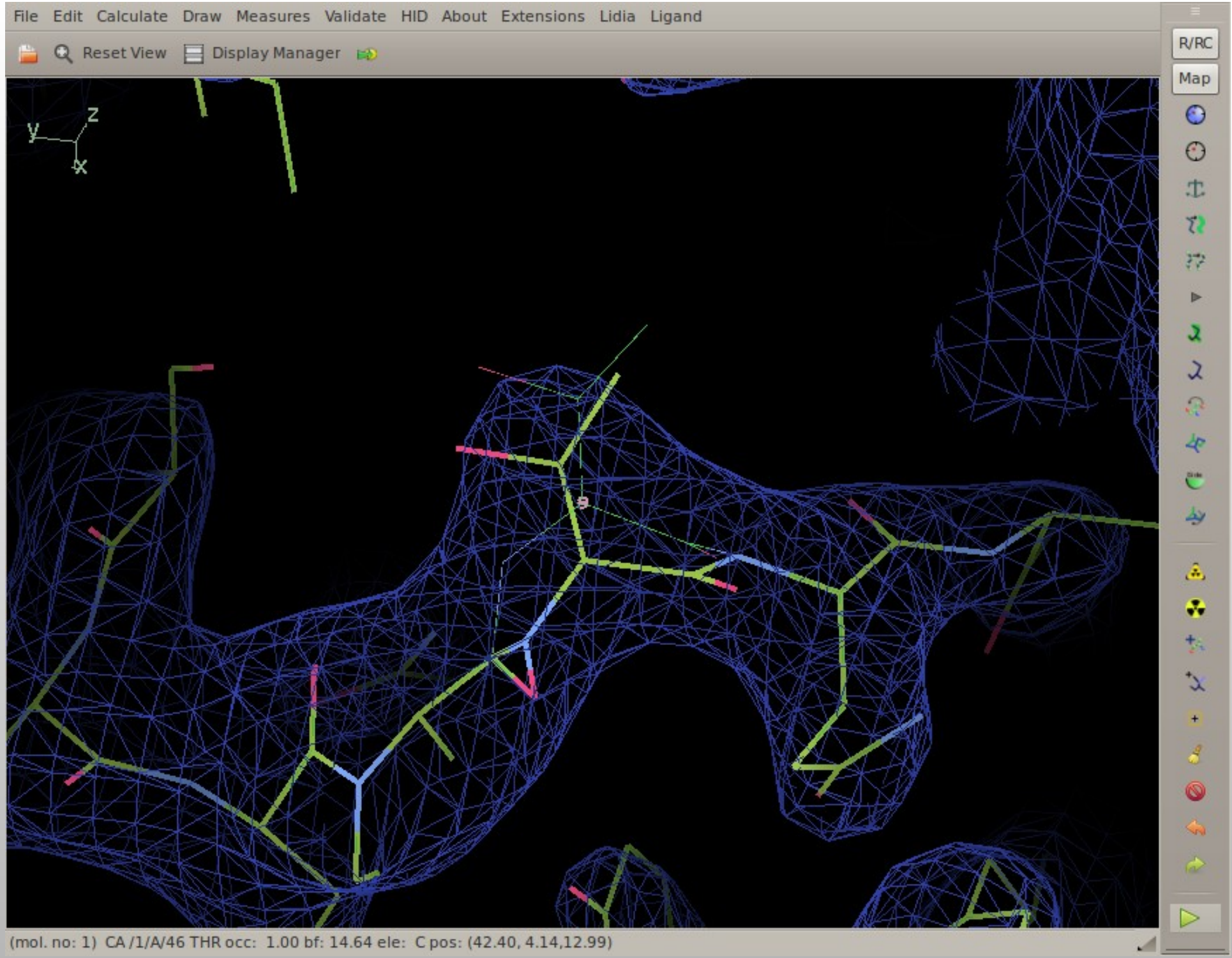
- High probability models with low resolution data

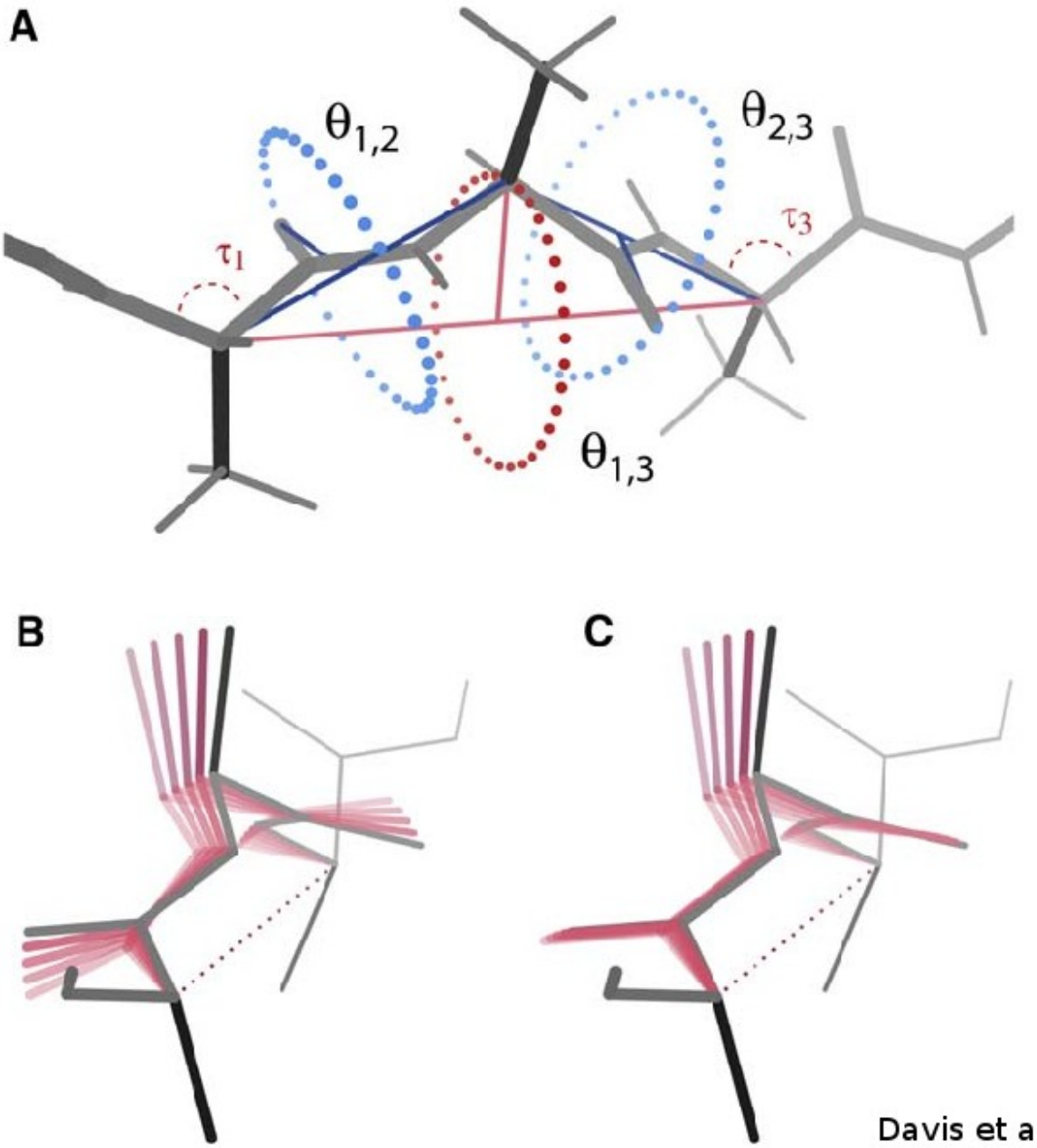
Previous

~~Current~~ Low Resolution Rotamer Search



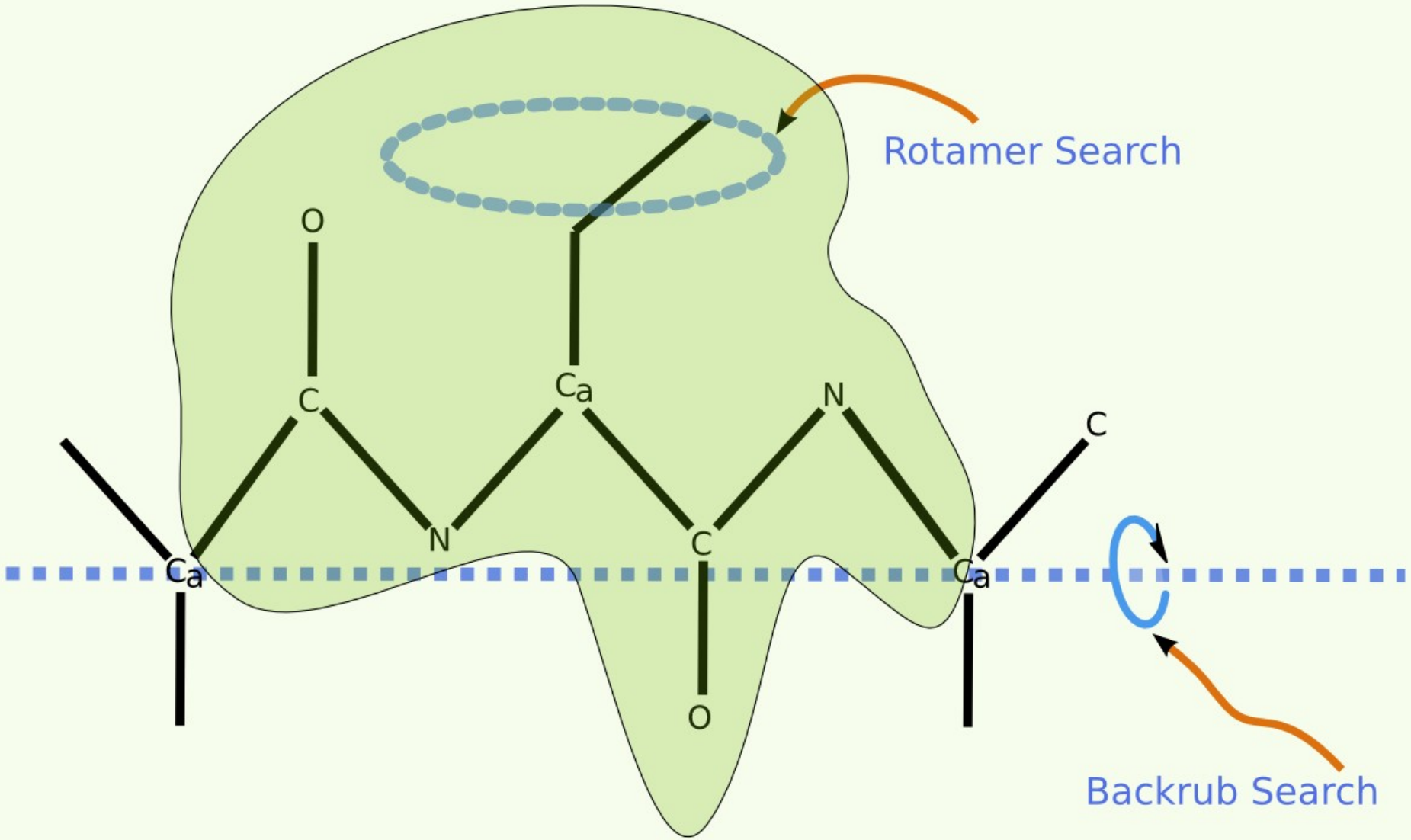

File Edit Calculate Draw Measures Validate HID About Extensions Lidia Ligand
Reset View
Display Manager
R/RC
Map
Z
Y
X
(mol. no: 1) CA /1/A/46 THR occ: 1.00 bf: 14.64 ele: C pos: (42.40, 4.14,12.99)


File Edit Calculate Draw Measures Validate HID About Extensions Lidia Ligand
Reset View
Display Manager
R/RC
Map
Z
Y
X
(mol. no: 1) CA /1/A/46 THR occ: 1.00 bf: 14.64 ele: C pos: (42.40, 4.14,12.99)

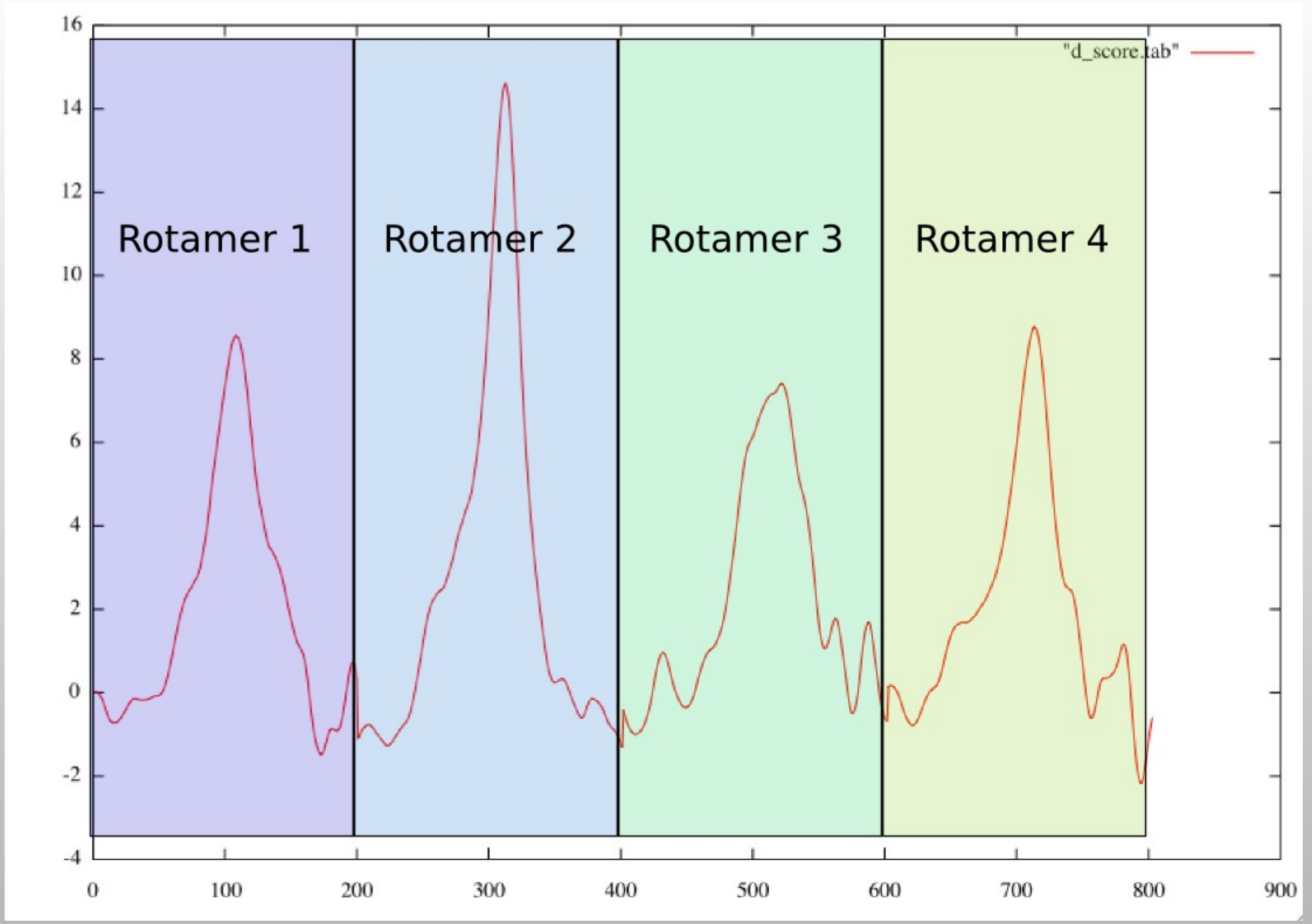


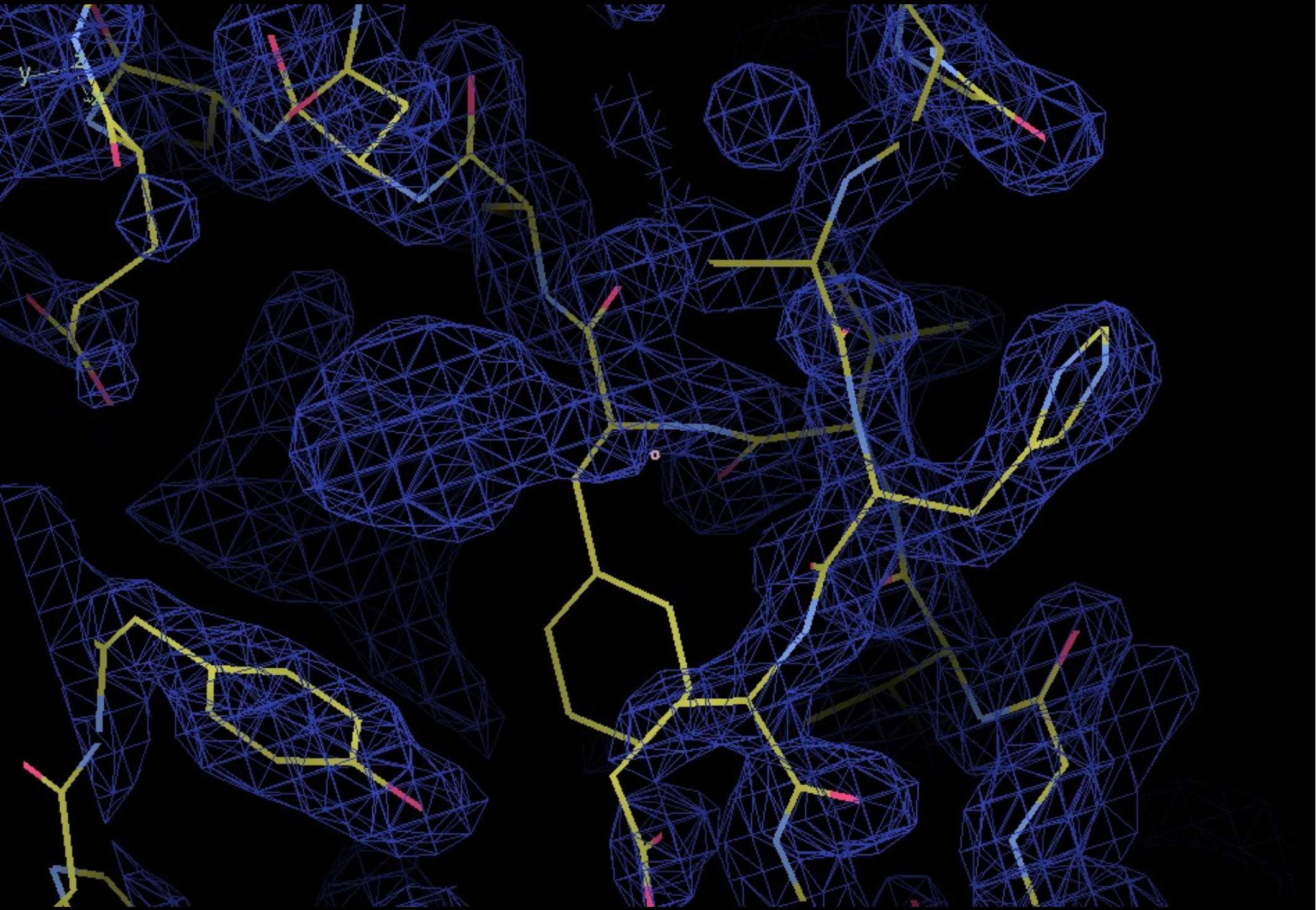
Davis et al. (2006) Structure

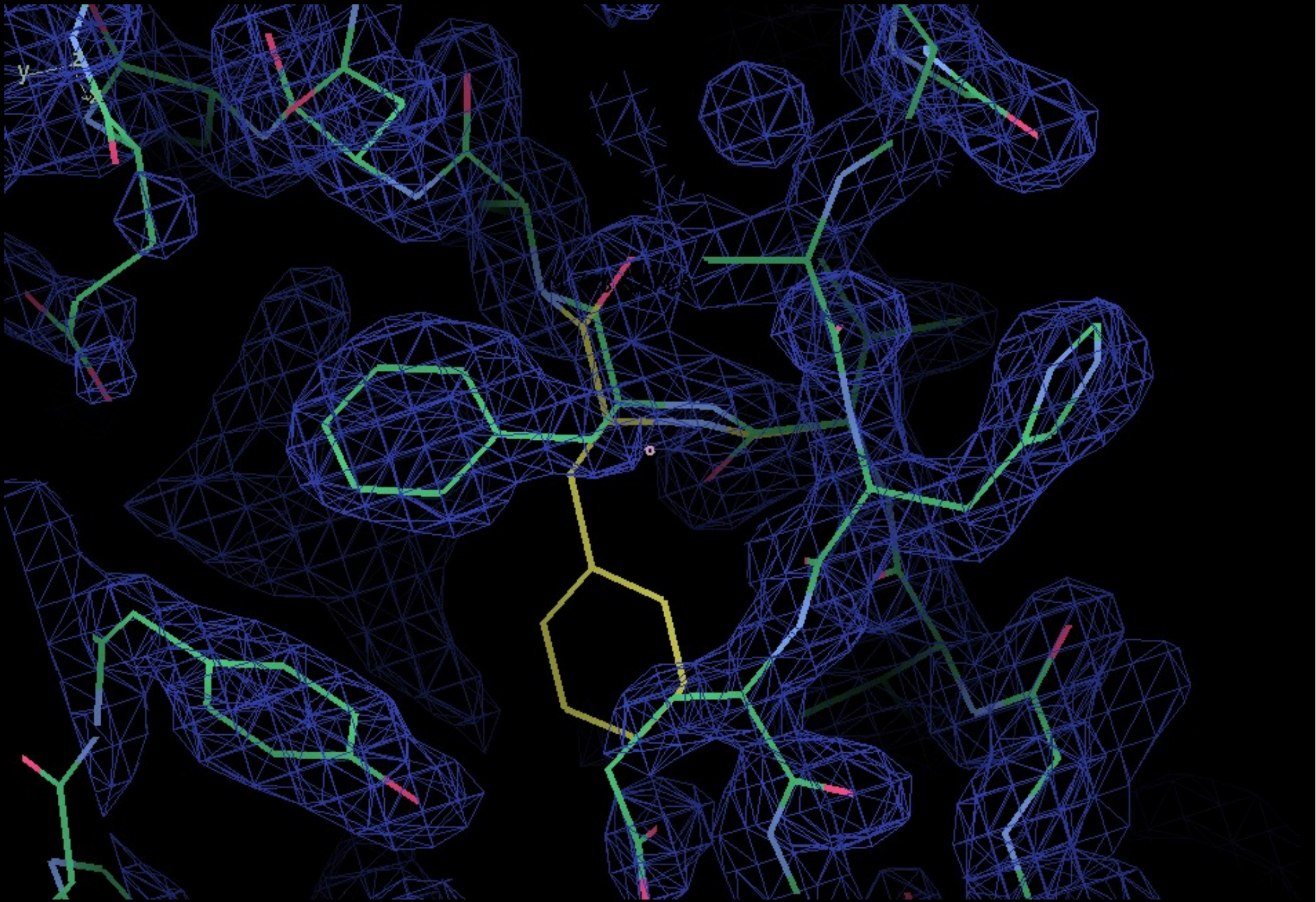
New Low Resolution Rotamer Search

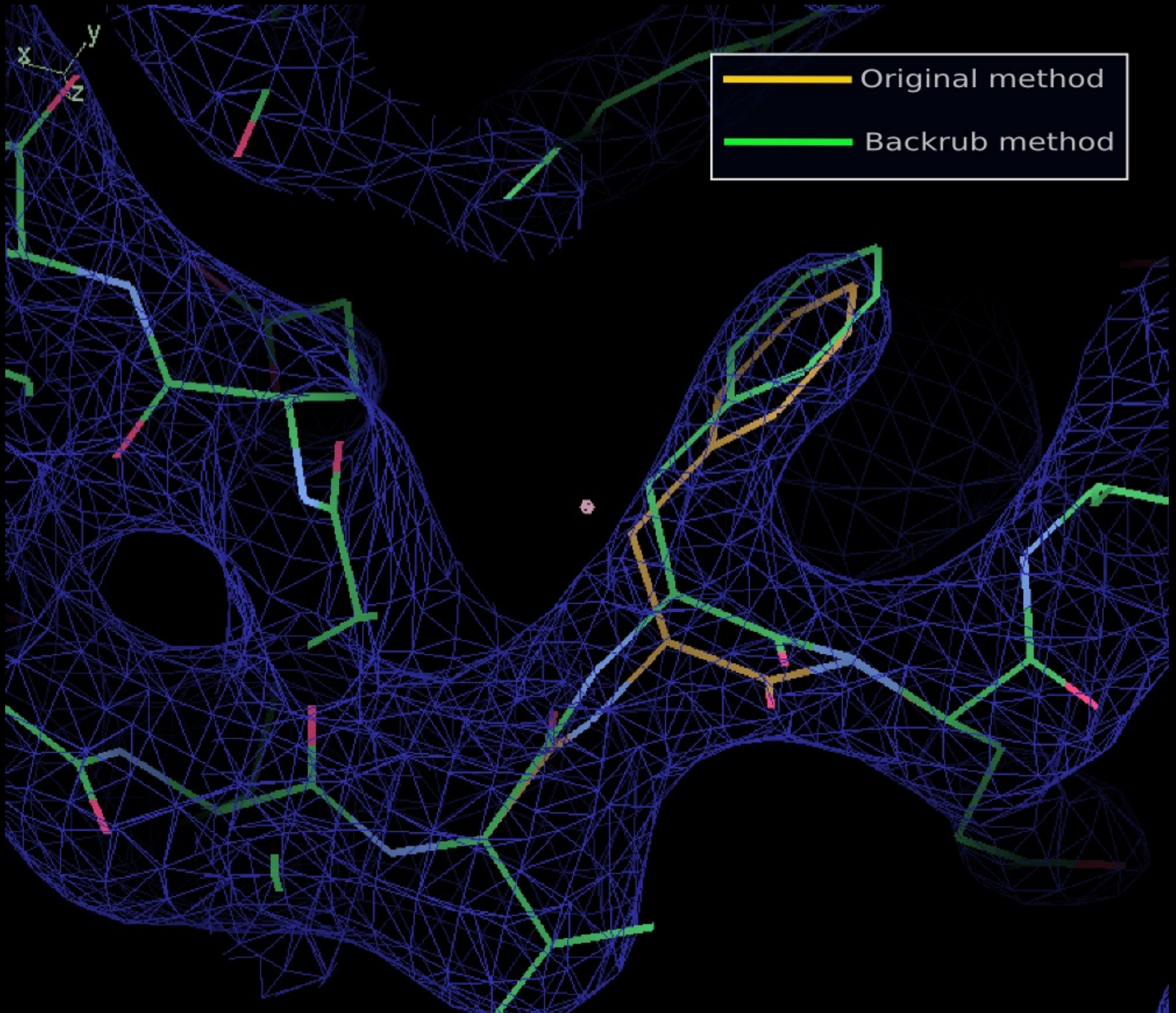


After Fitting Tools in KING/Molprobity


"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






y
x
z
Original method
Backrub method

To turn it on...

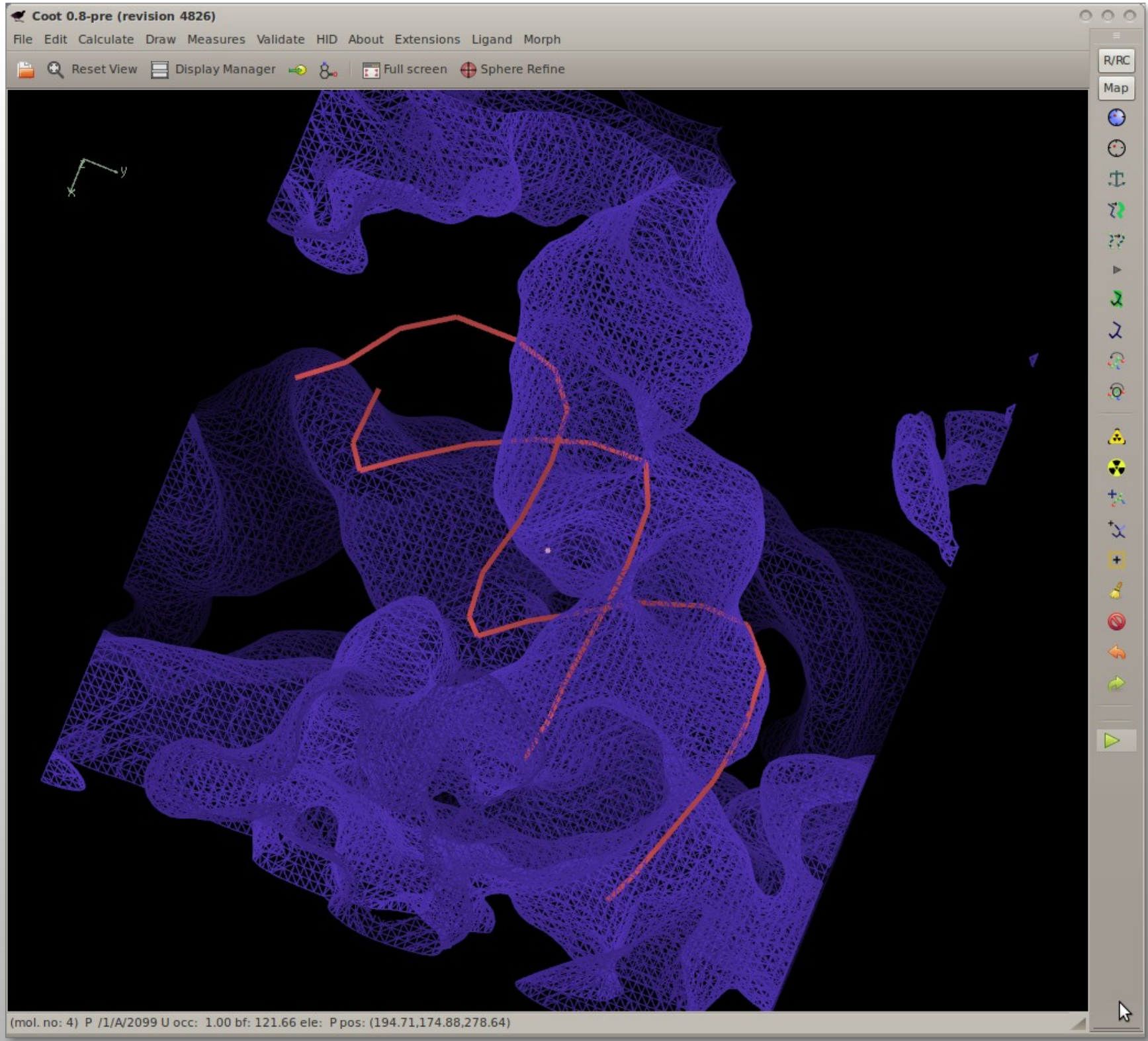
- (ROTAMERSEARCHLOWRES)

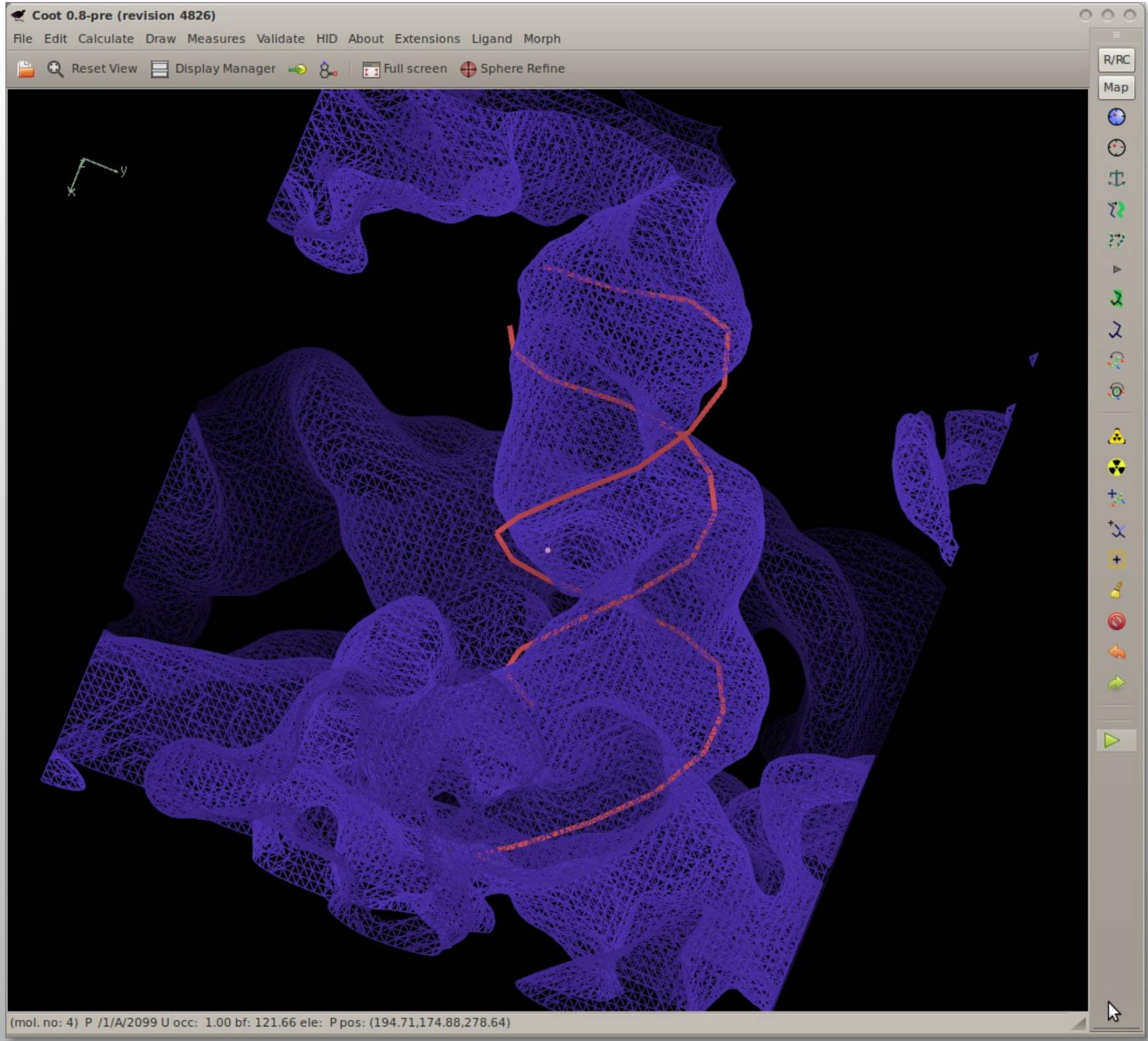
Jiggle Fit

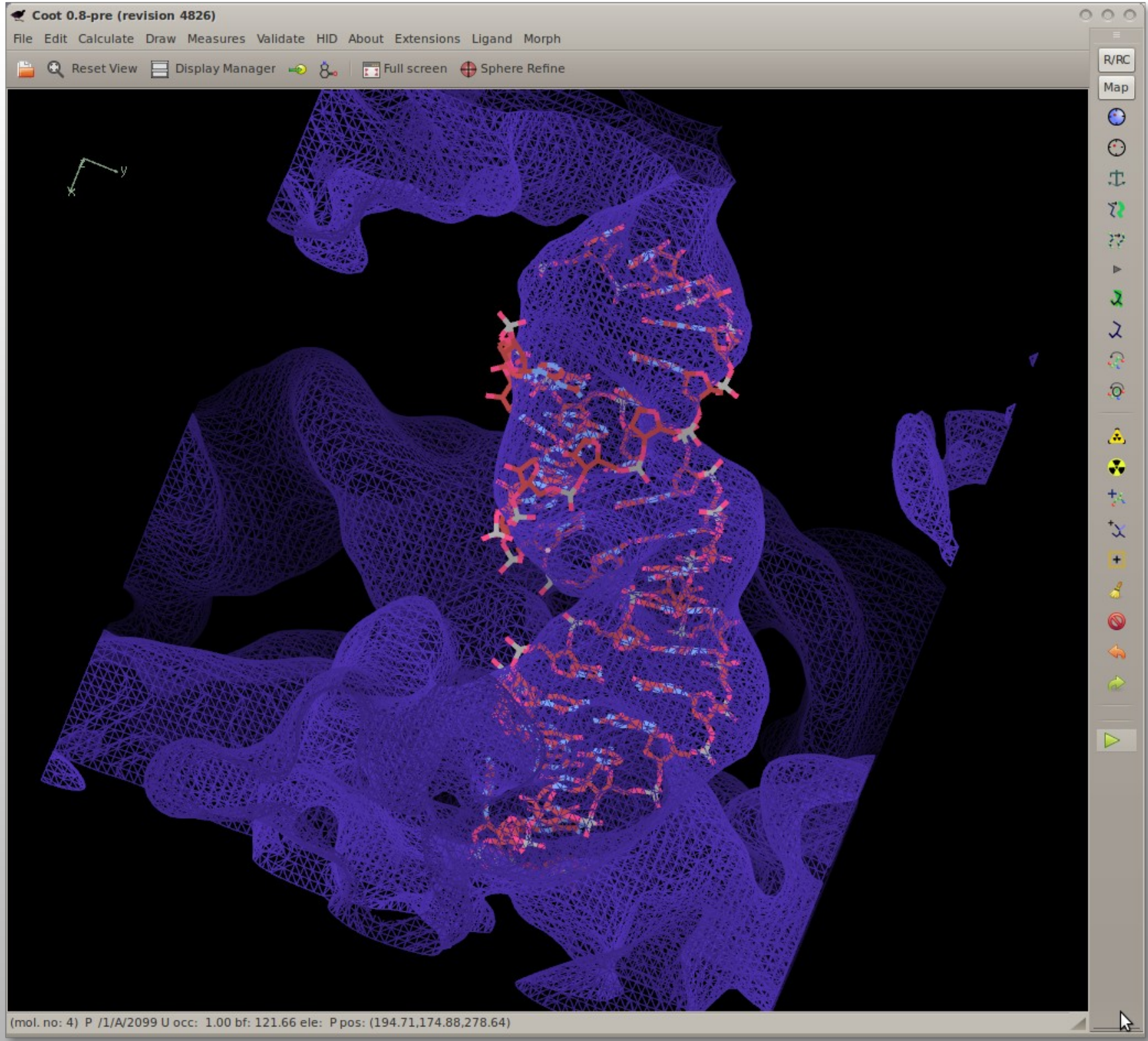
- How do I rotate and translate these atoms to fit the density?
  - 6-dimensional problem
- Originally used to fit simple ligands/solvent molecules to blobs of density
- Now extended to fit arbitrary atom selections
  - *e.g.* by Chain

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)

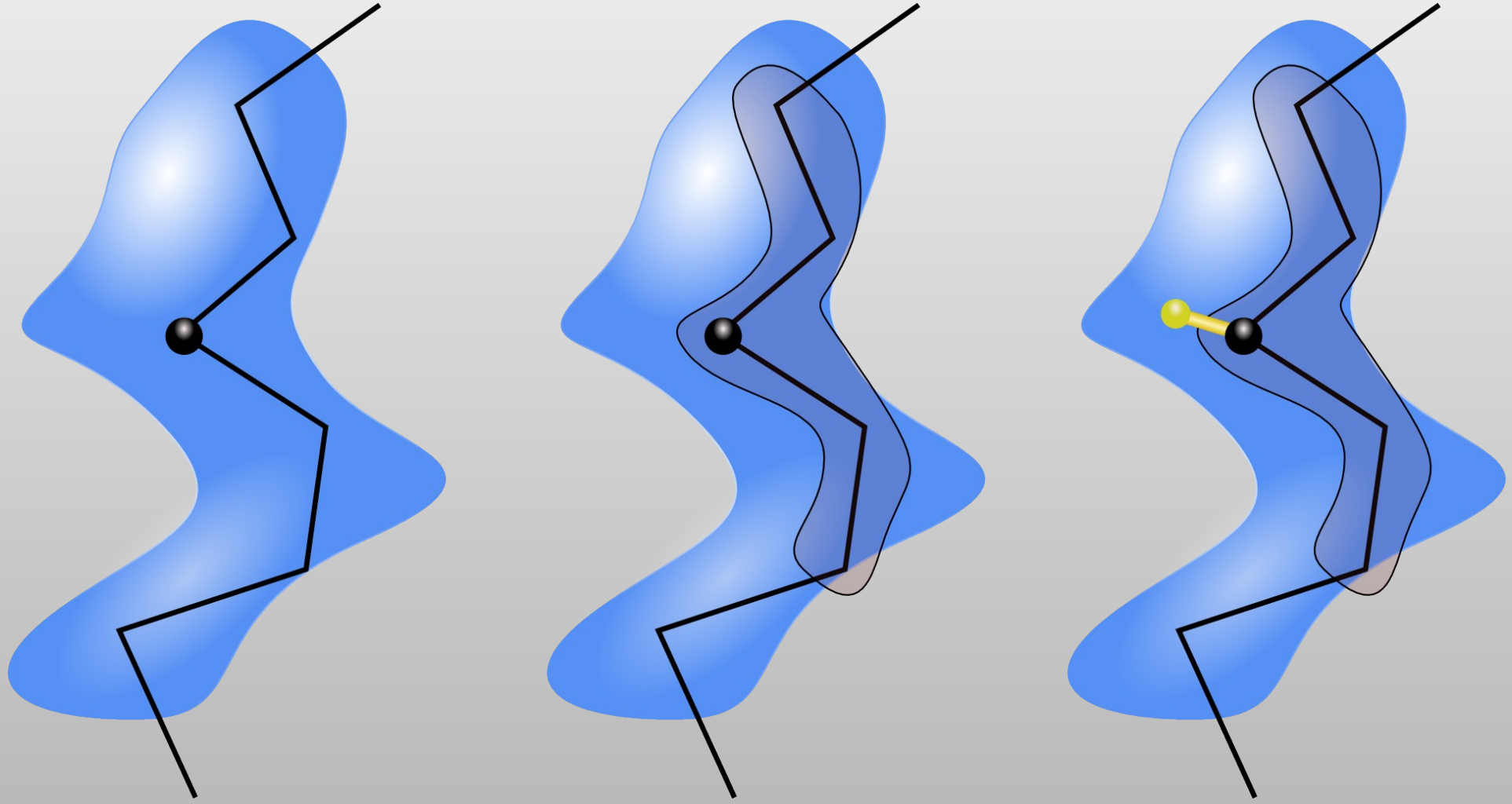

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
y
(mol. no: 4) P /1/A/2099 U occ: 1.00 bf: 121.66 ele: P pos: (194.71,174.88,278.64)


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)

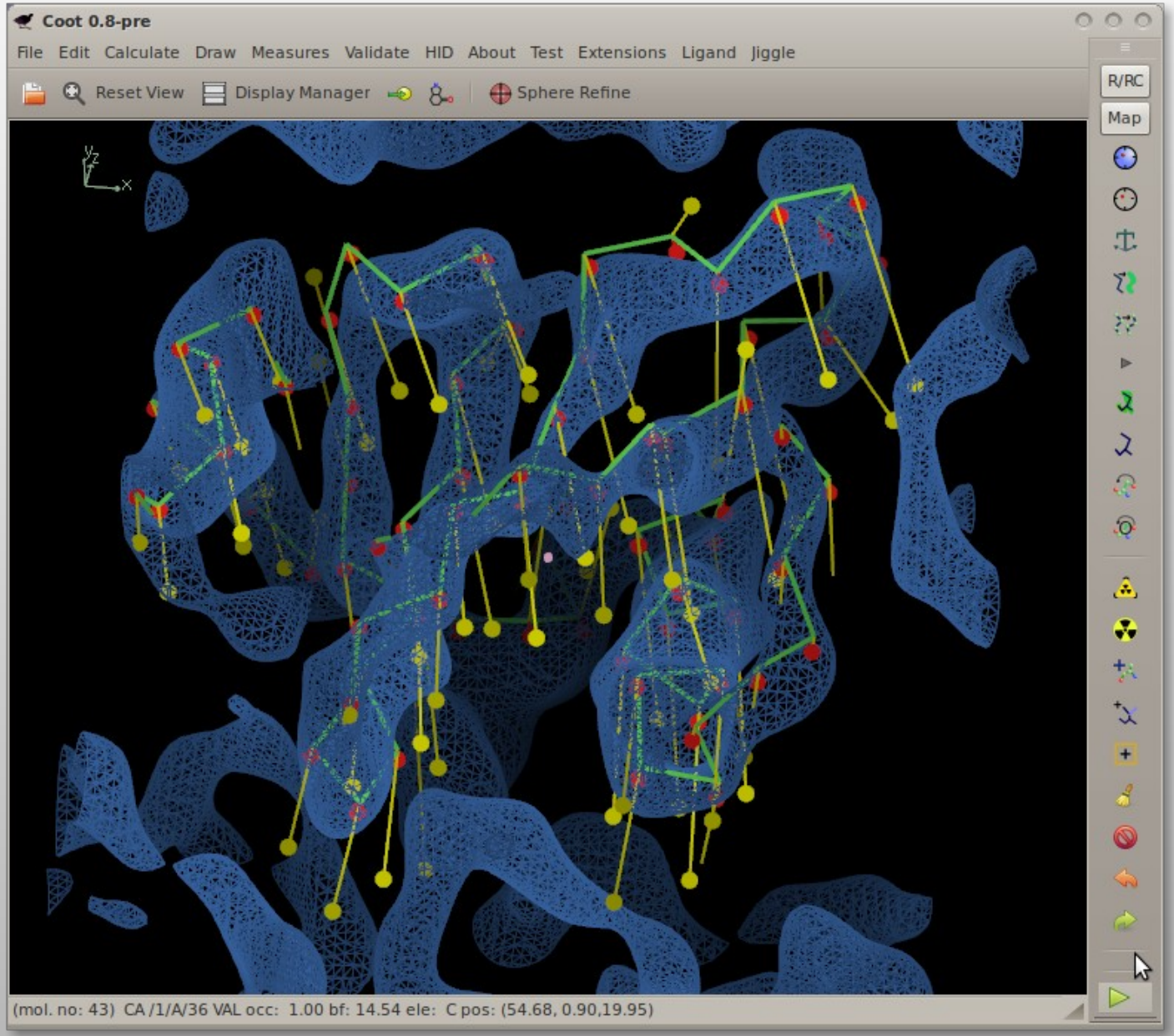
Model Morphing: How it Works

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

Model Morphing: Generating the Raw RTs

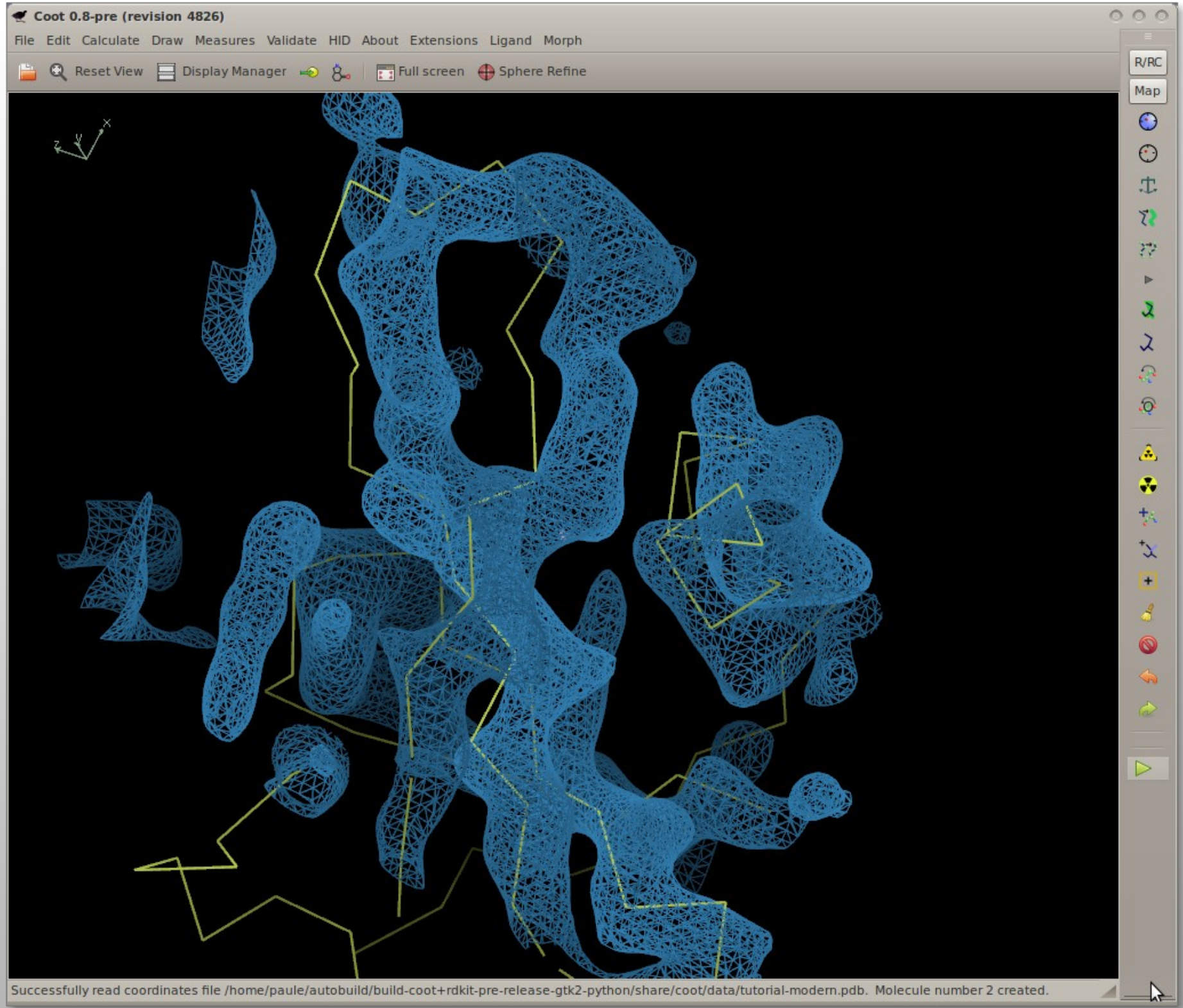


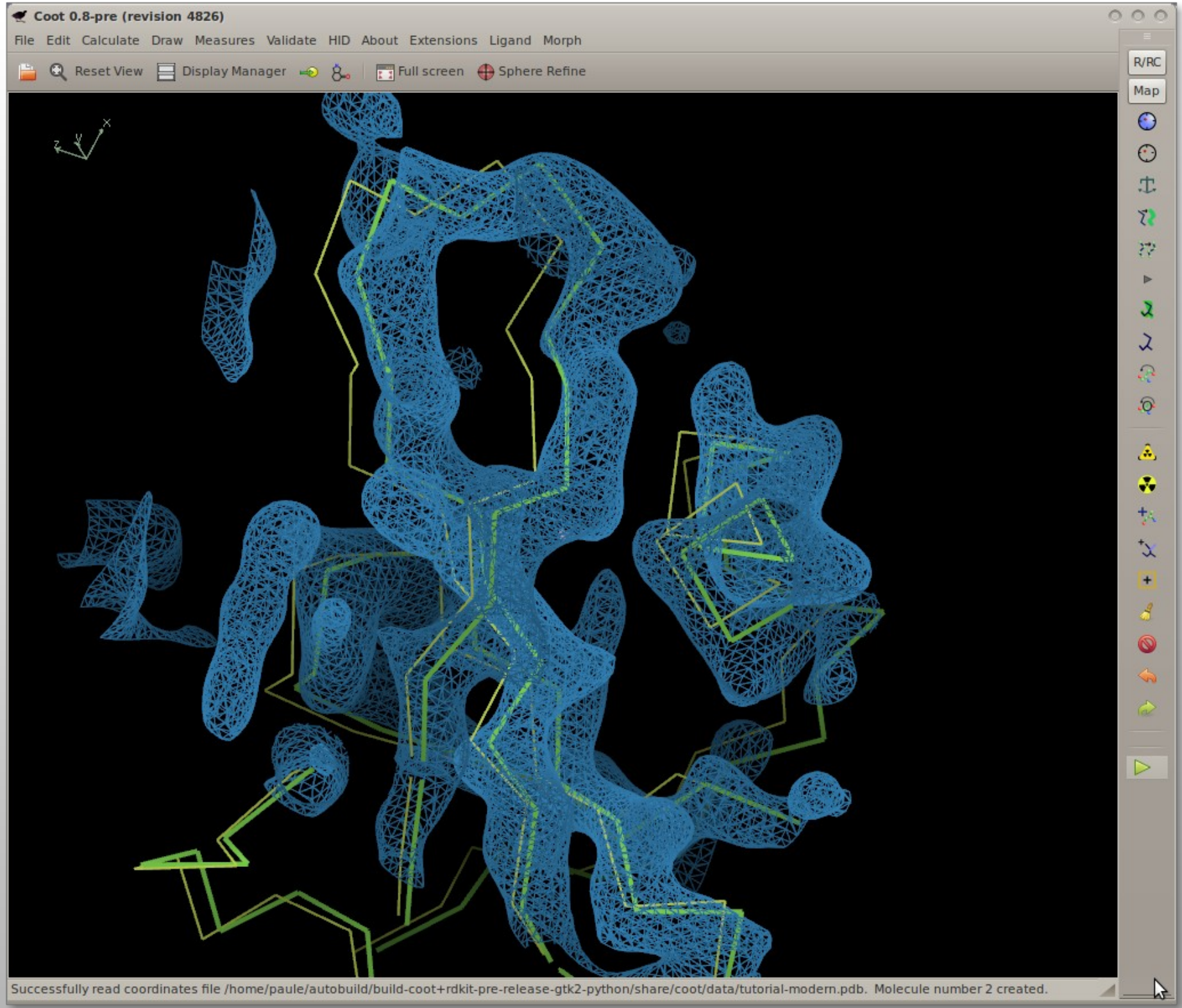
Model Morphing: Example

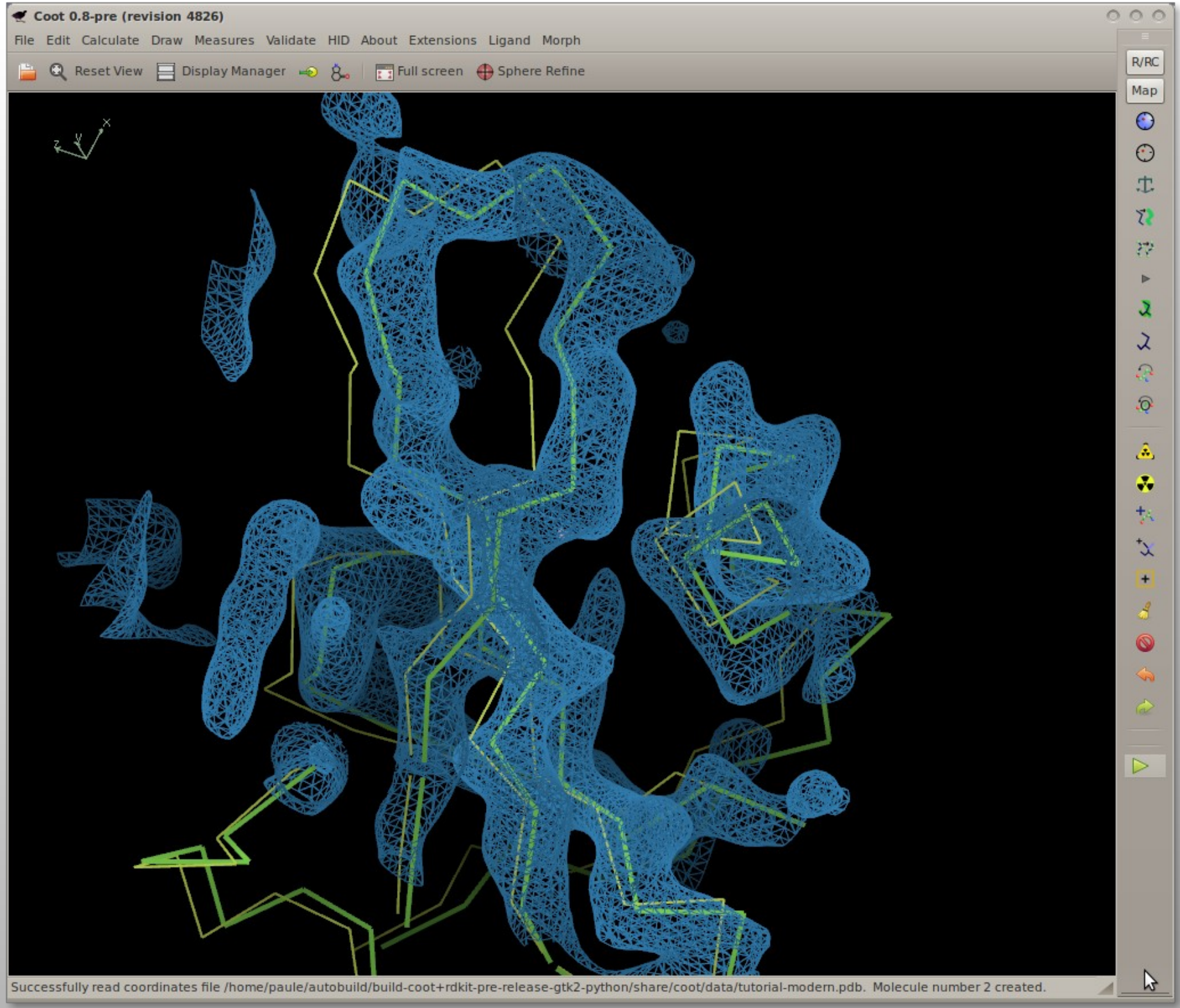


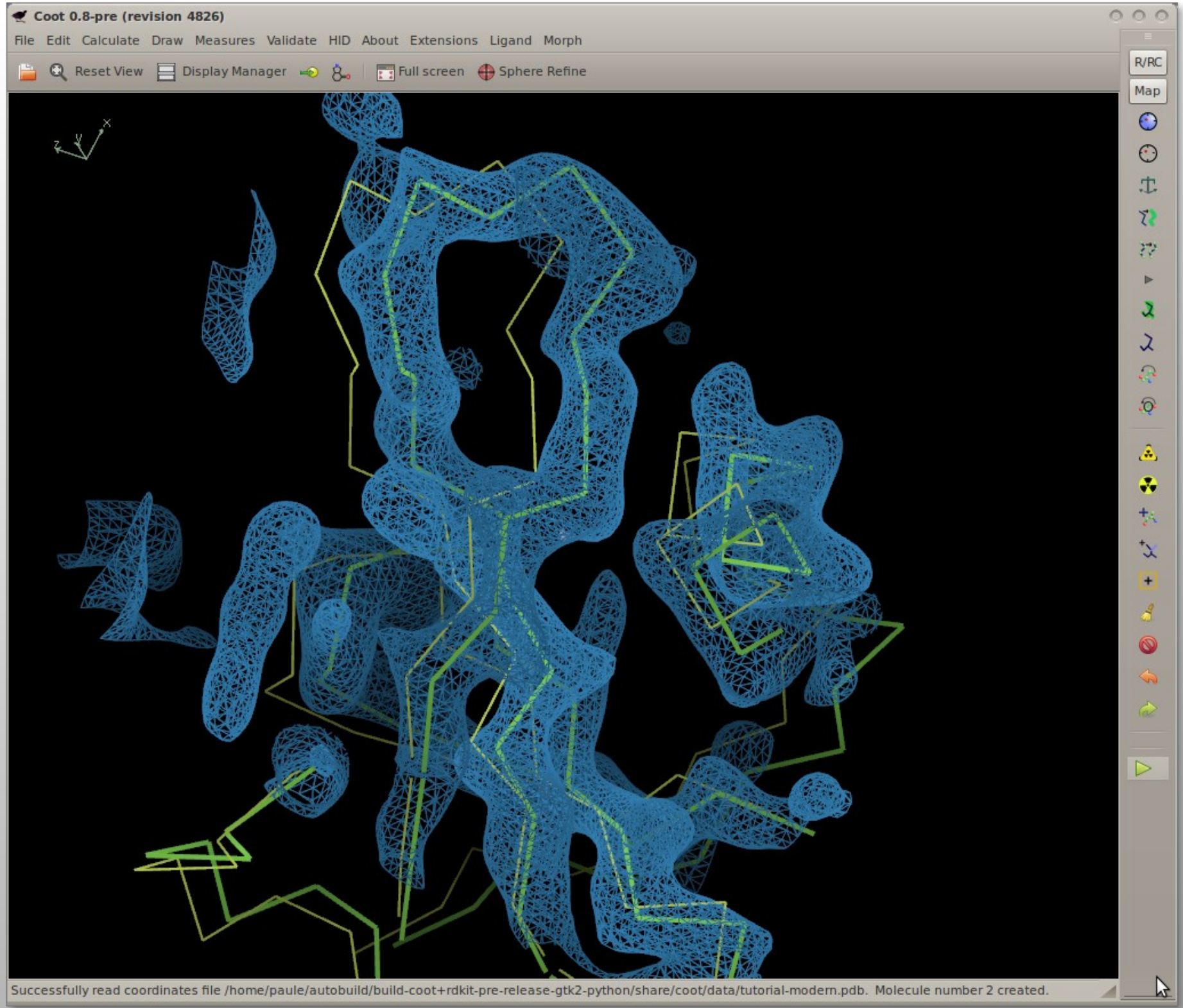
Model Morphing: Robust Averaging

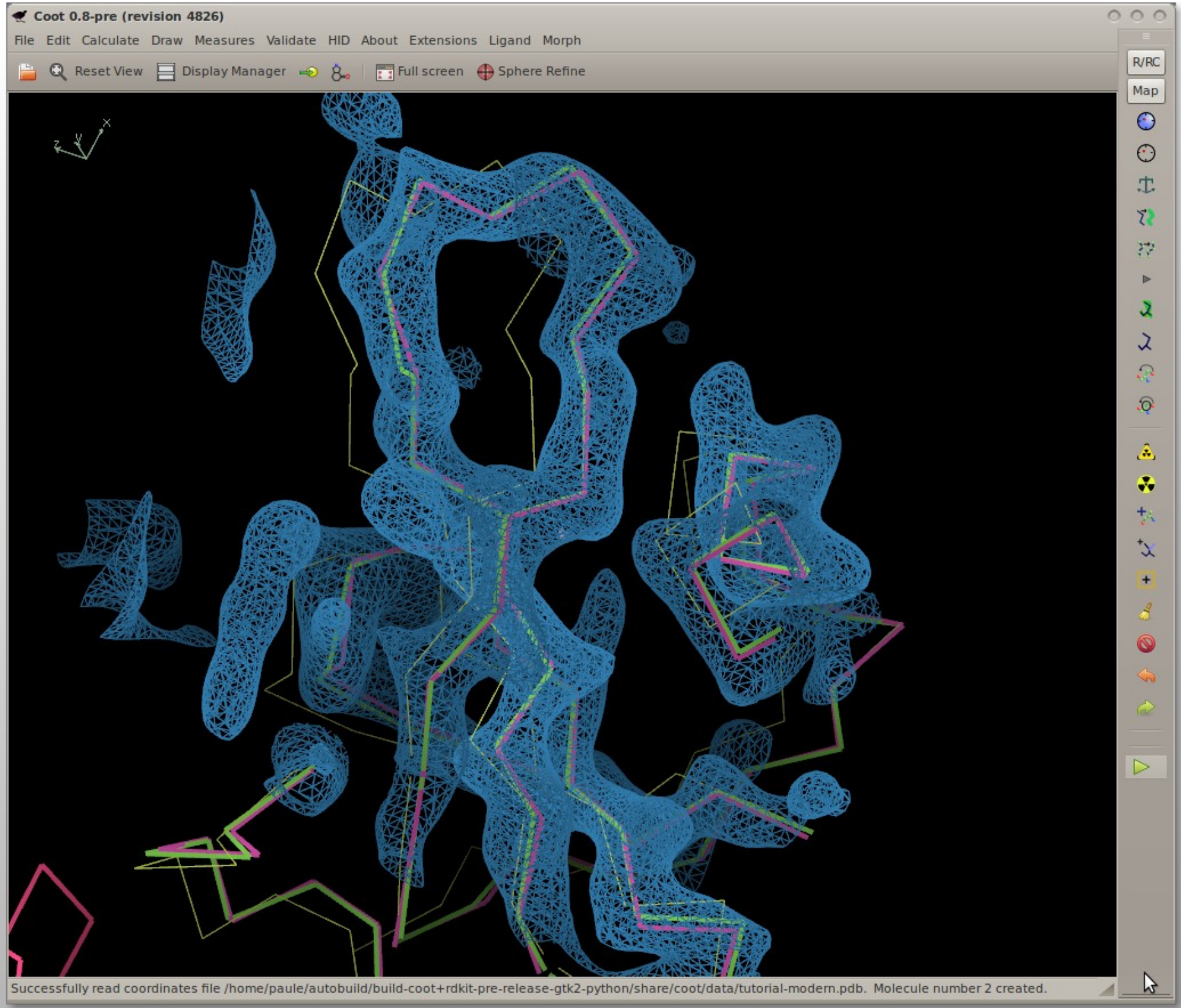
- What are the residues in the environment of a residue?
  - What are their RTs?
  - Create a metric 'distance', sort on that
  - Discard the top and bottom 25%
  - Use remaining RTs to generate average
  - ...which is then applied to central residue
- Repeat for all residues
- Larger environment radii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


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


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


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

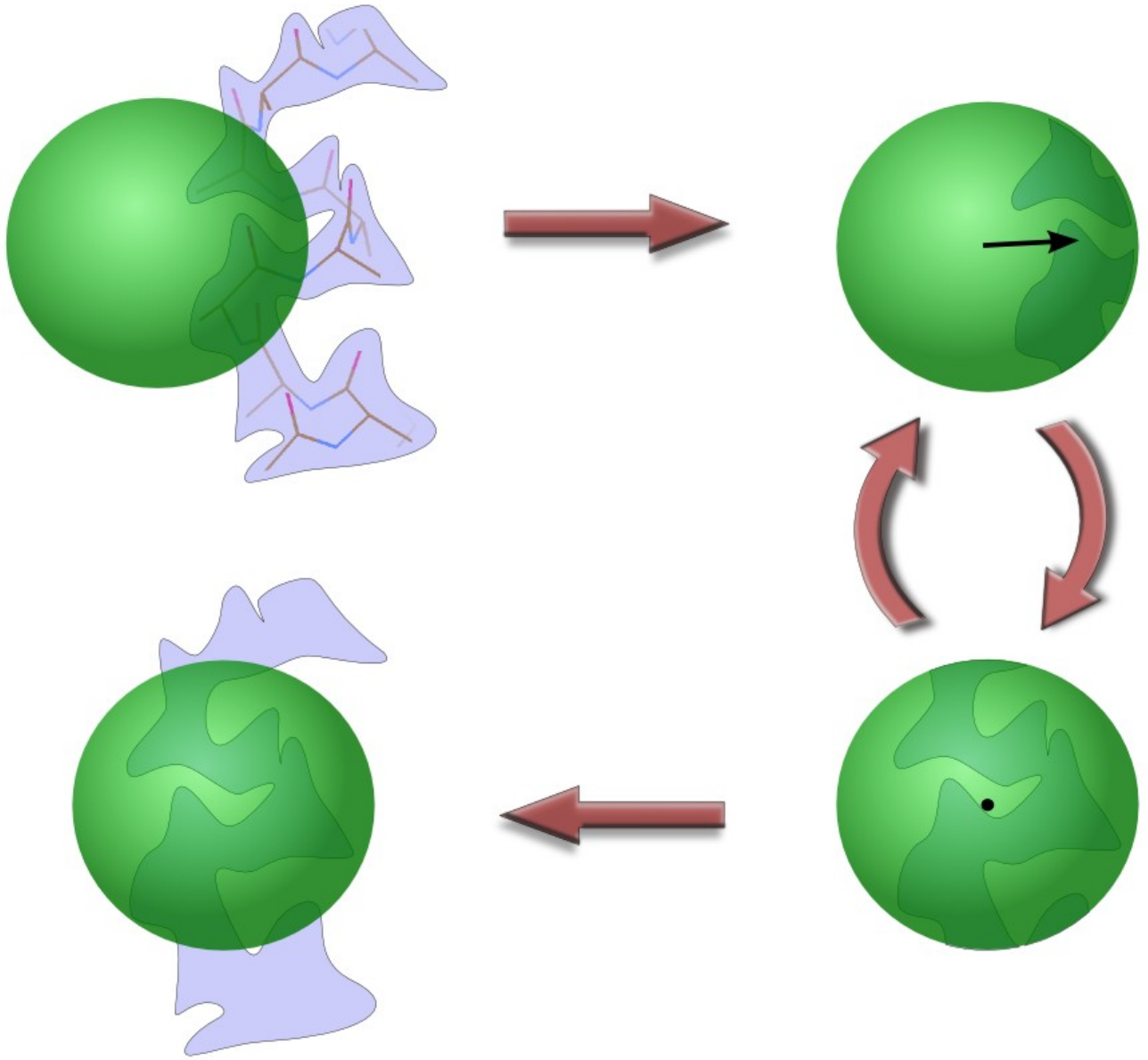

Coot 0.8-pre (revision 4826)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Morph
Reset View
Display Manager
Full screen
Sphere Refine
R/RC
Map
Successfully read coordinates file /home/paule/autobuild/build-coot+rdkit-pre-release-gtk2-python/share/coot/data/tutorial-modern.pdb.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


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

Alpha Helix Placement

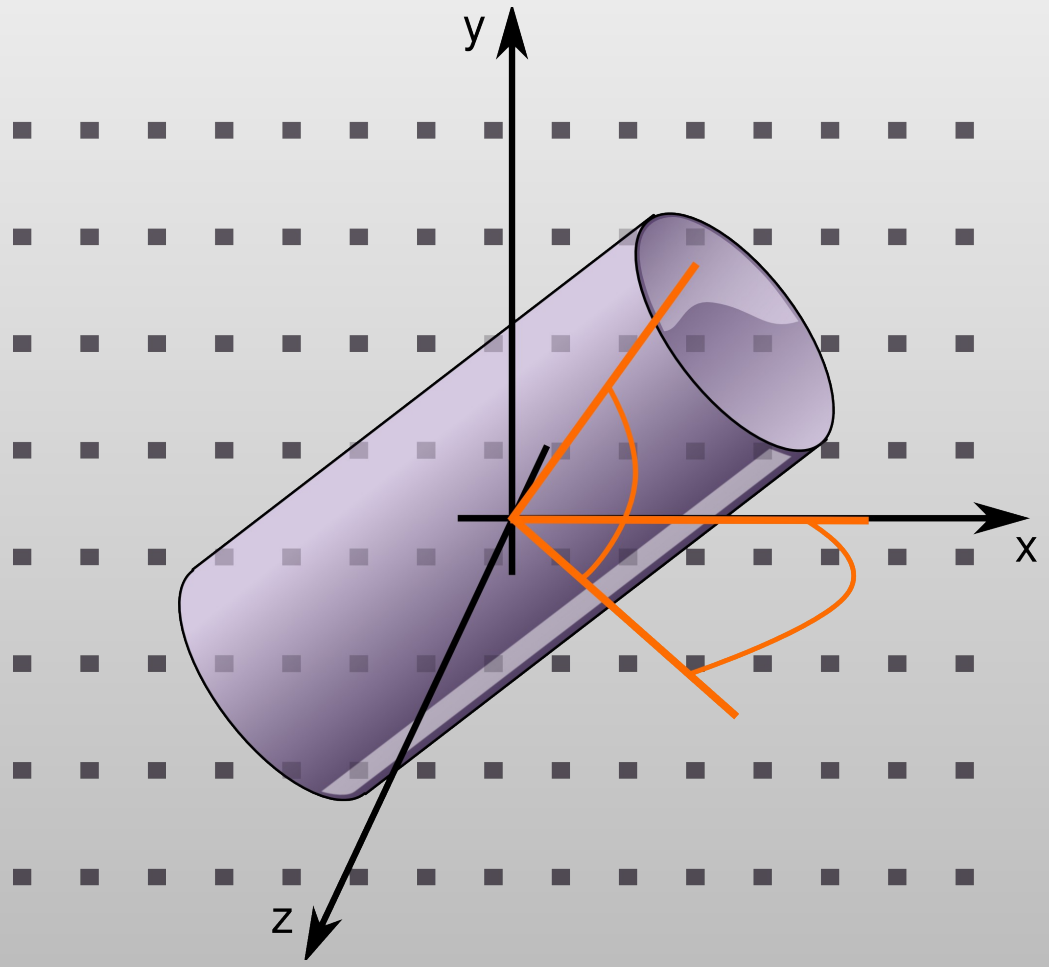
- Scenario: Looking at a new map, not built with automatic tools:
  - “I can see that there’s a helix here - build it for me!”
- From a given point:
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)
  - 1D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

Centering the Rotation point

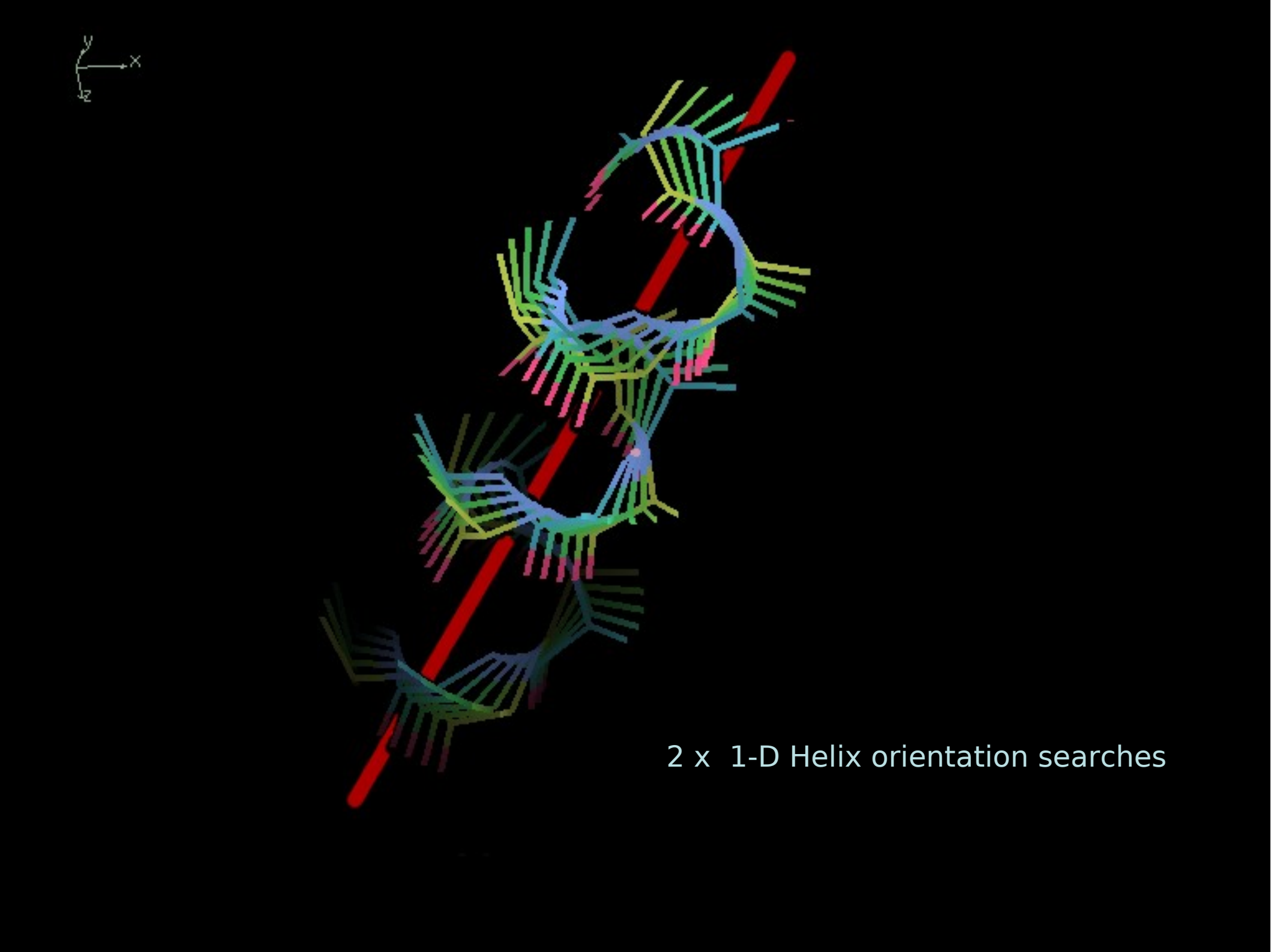


Cylinder Search

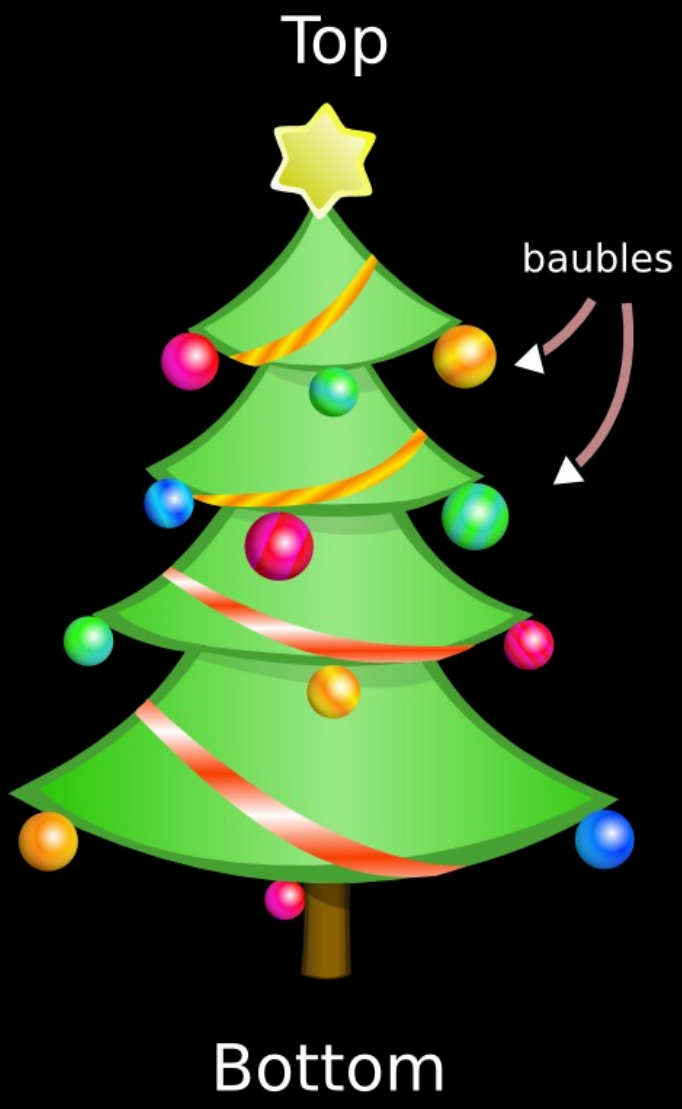
- Pick the orientation that encapsulates the most electron density

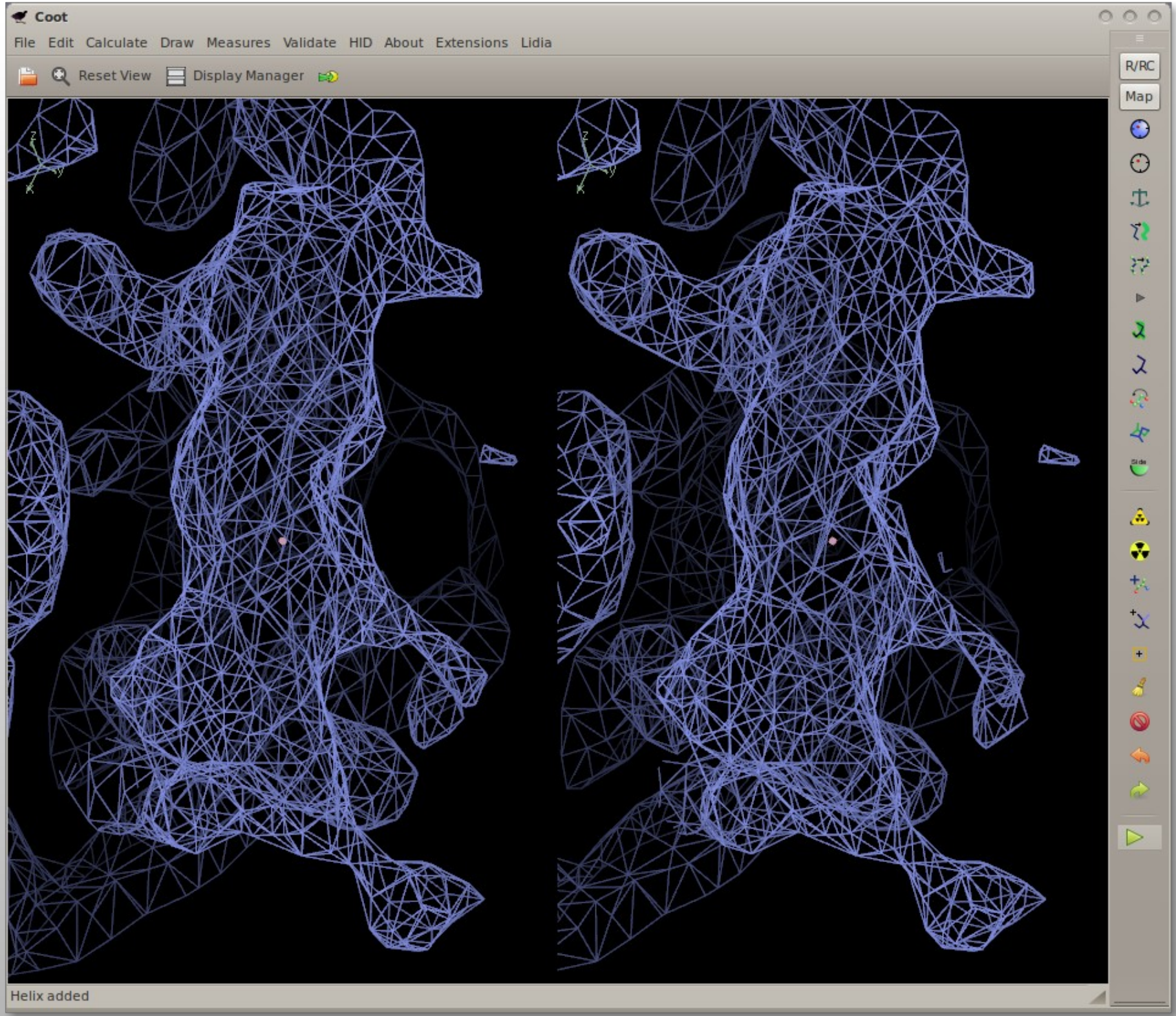


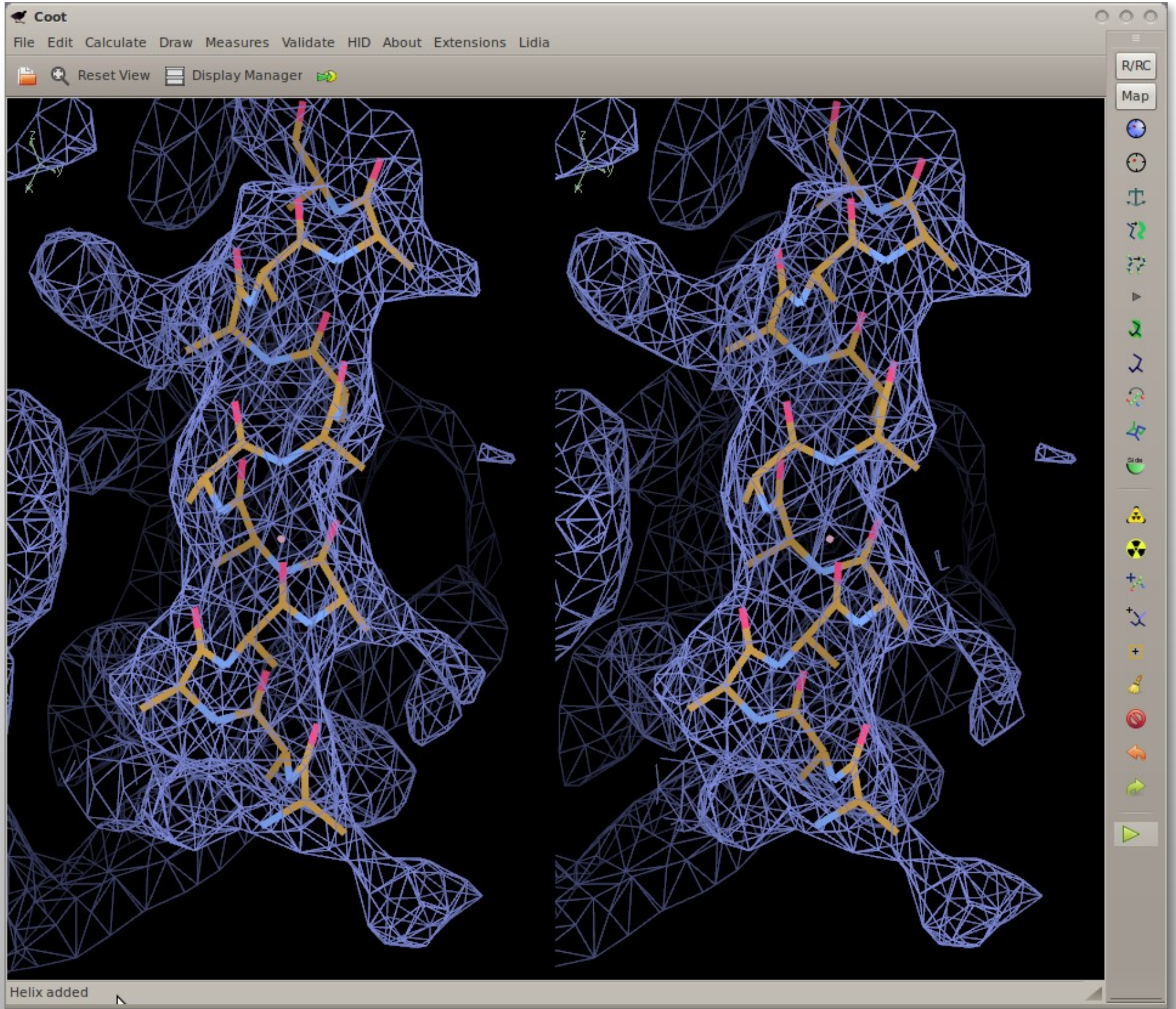
Using 2 rotation axes

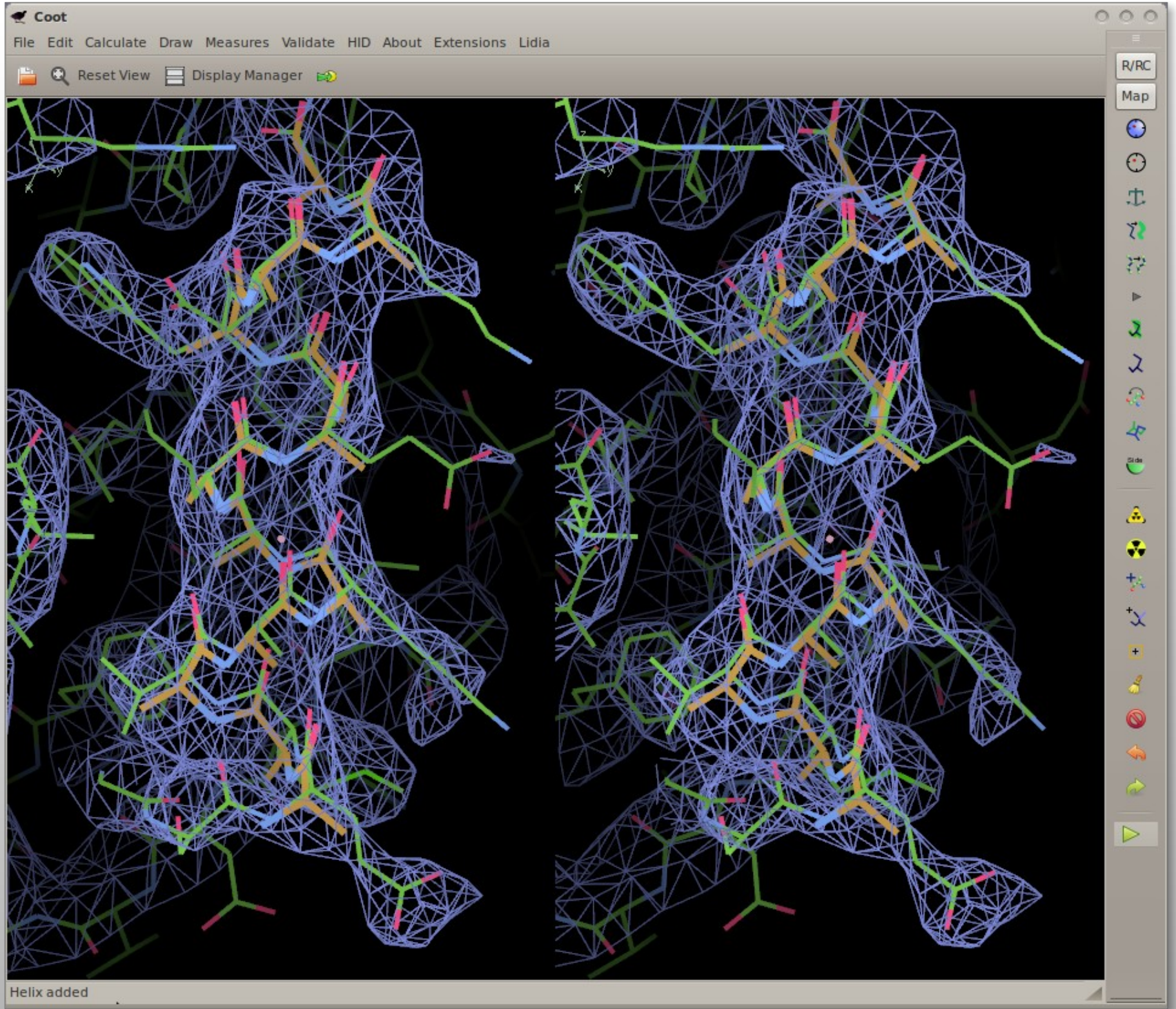


2 x 1-D Helix orientation searches


Top
baubles
Bottom


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added

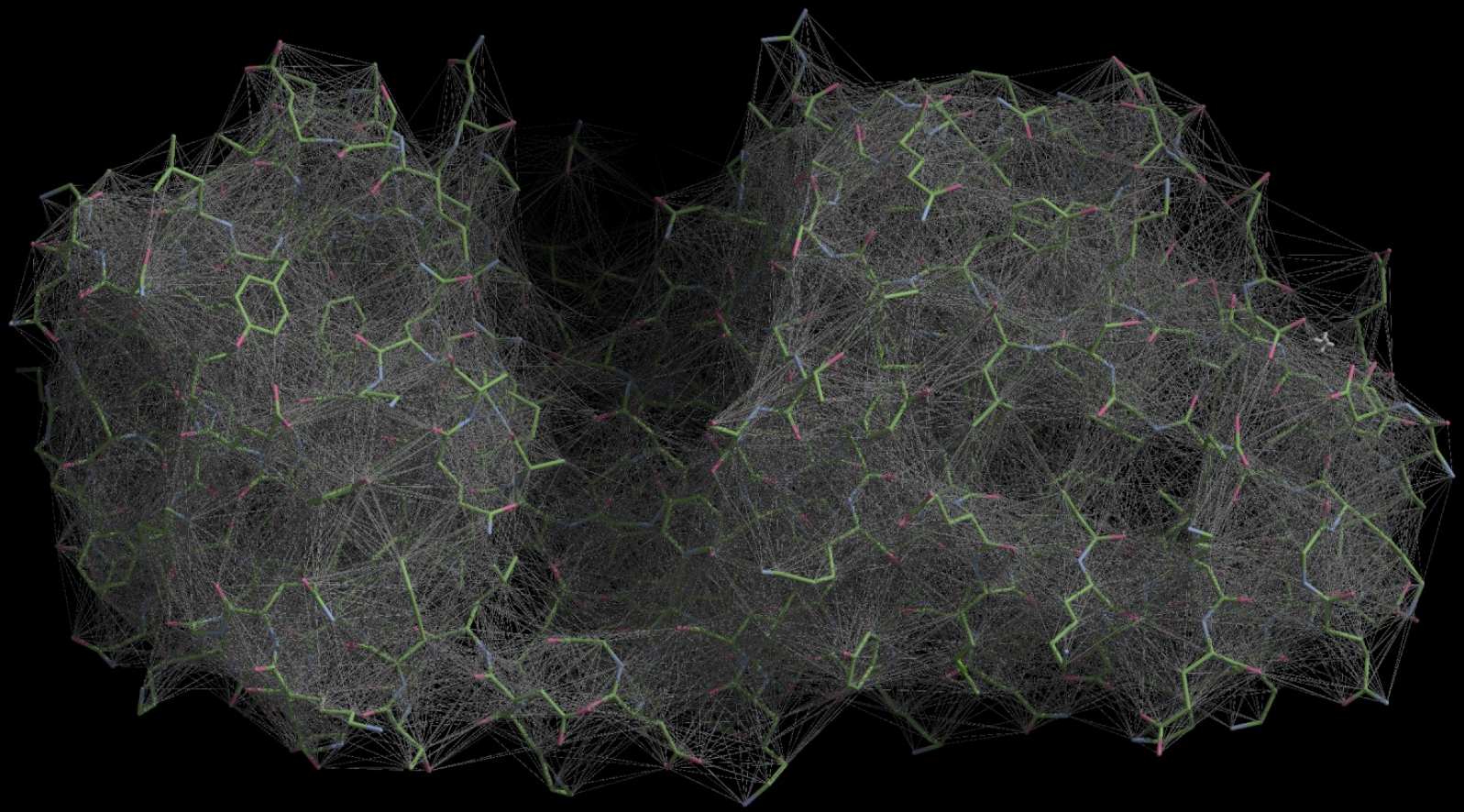

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Helix added

Additional Restraints

ProSMART integration

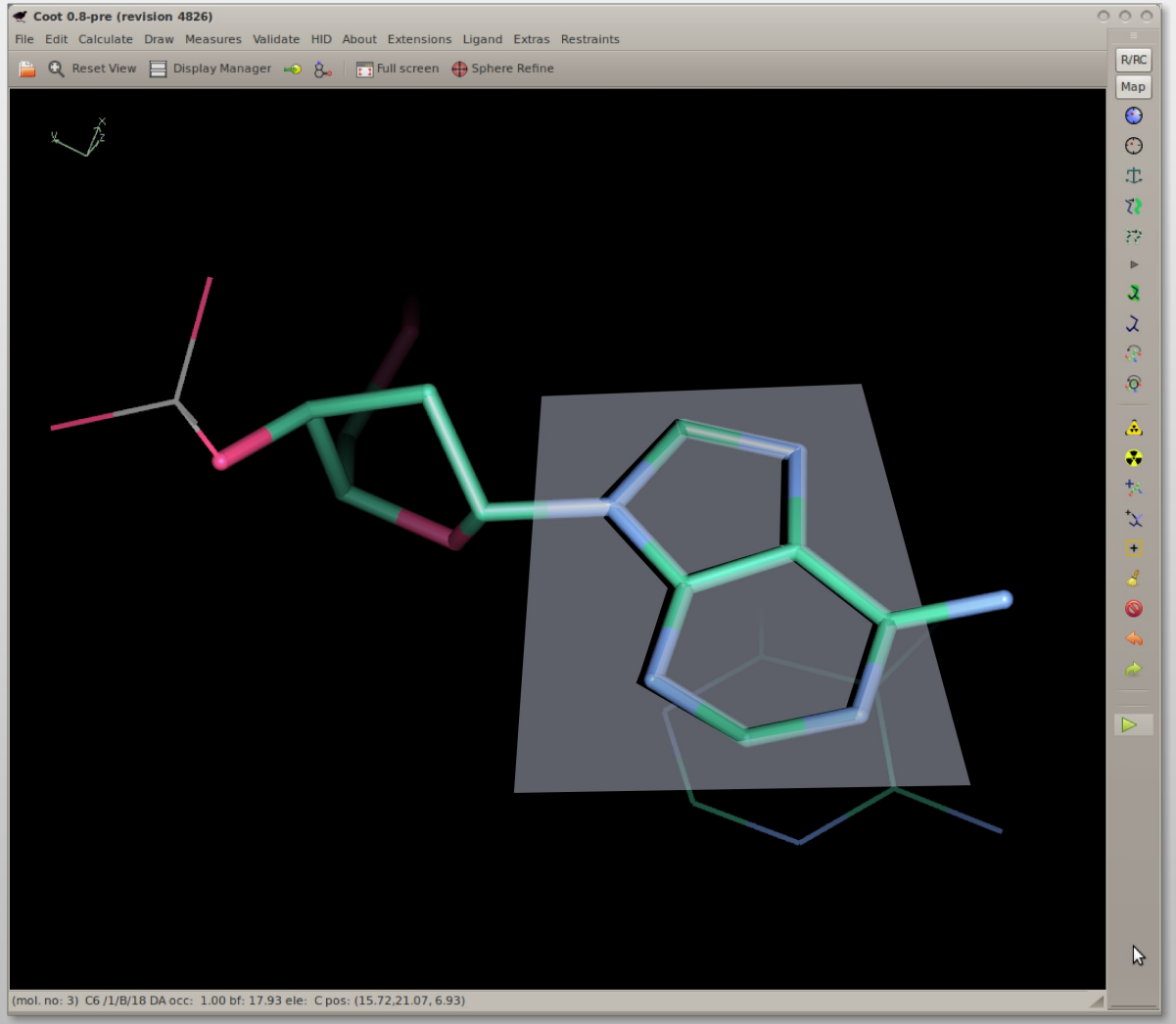
- ProSMART generates distance restraints from homologous structures
  - to be applied to current model for refinement
  - now available in *Coot*

ProSMART Restraints



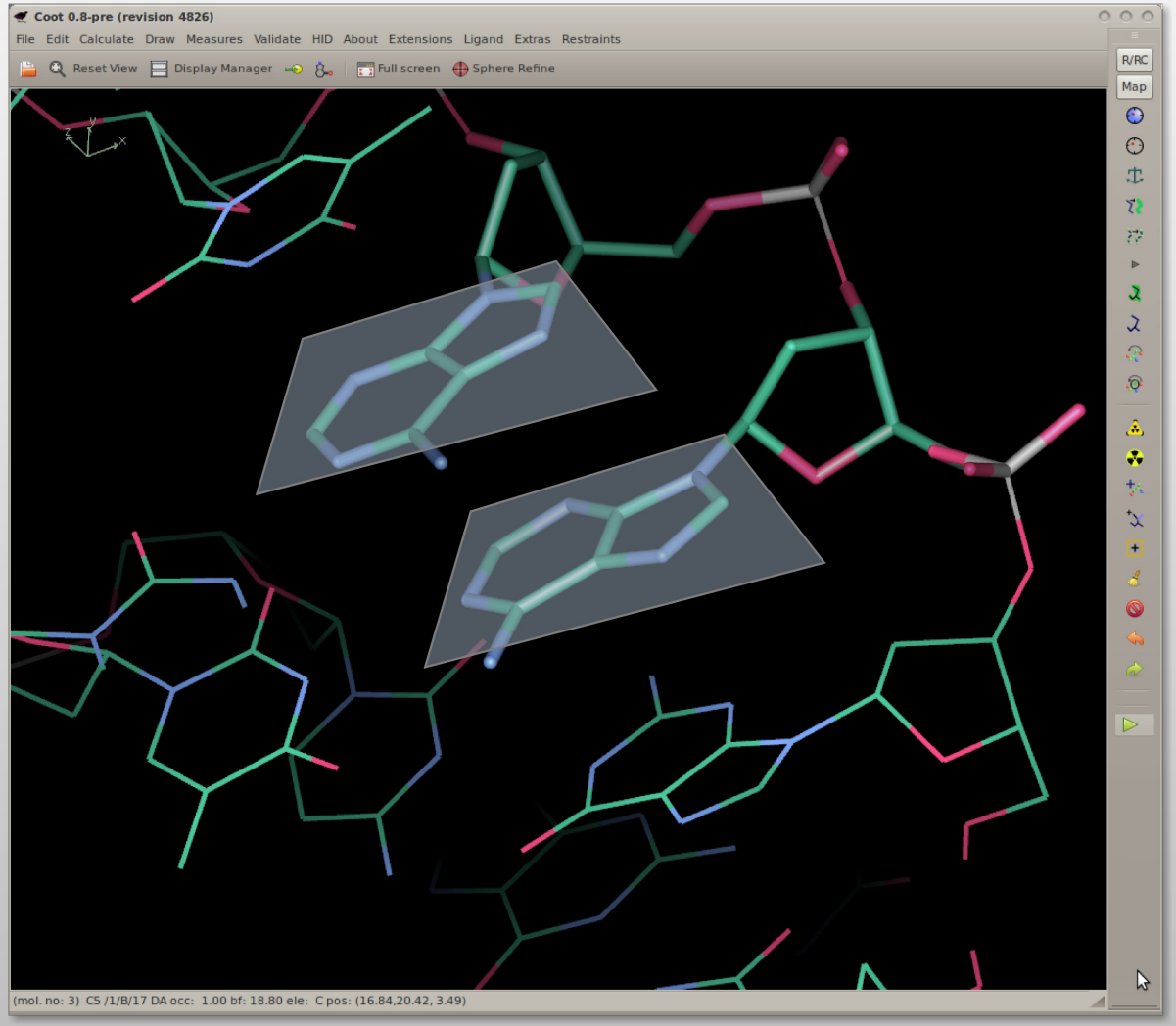
Tools for RNA

Plane Restraints



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

Parallel Planes Restraints



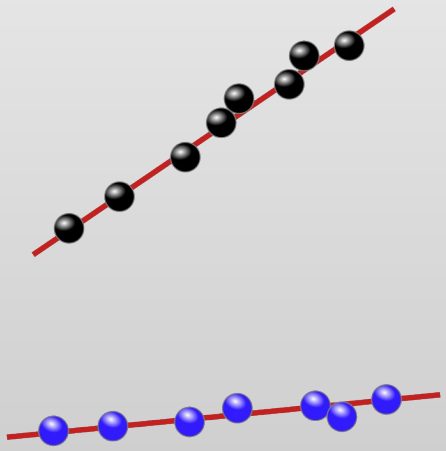
$$S = (a_1 - a_2)^2 + (b_1 - b_2)^2 + (c_1 - c_2)^2$$

Not easy to use in Coot

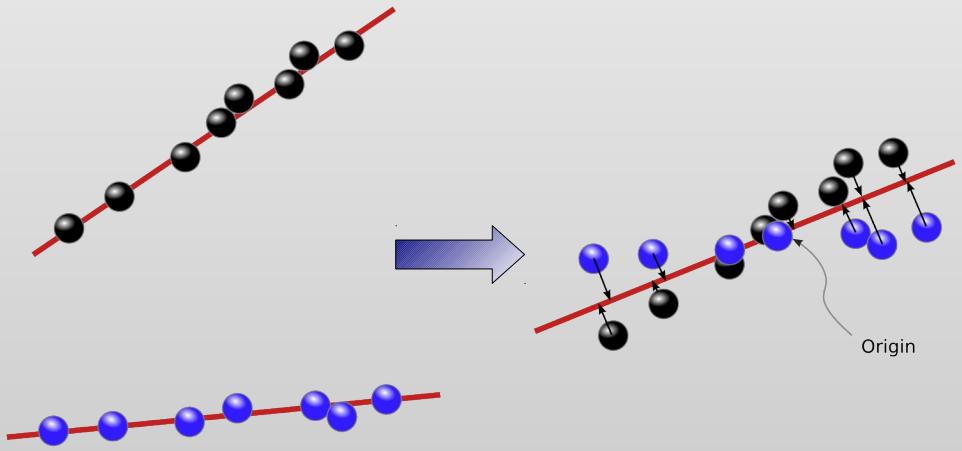
Parallel Planes Restraints

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

Parallel Planes Restraints

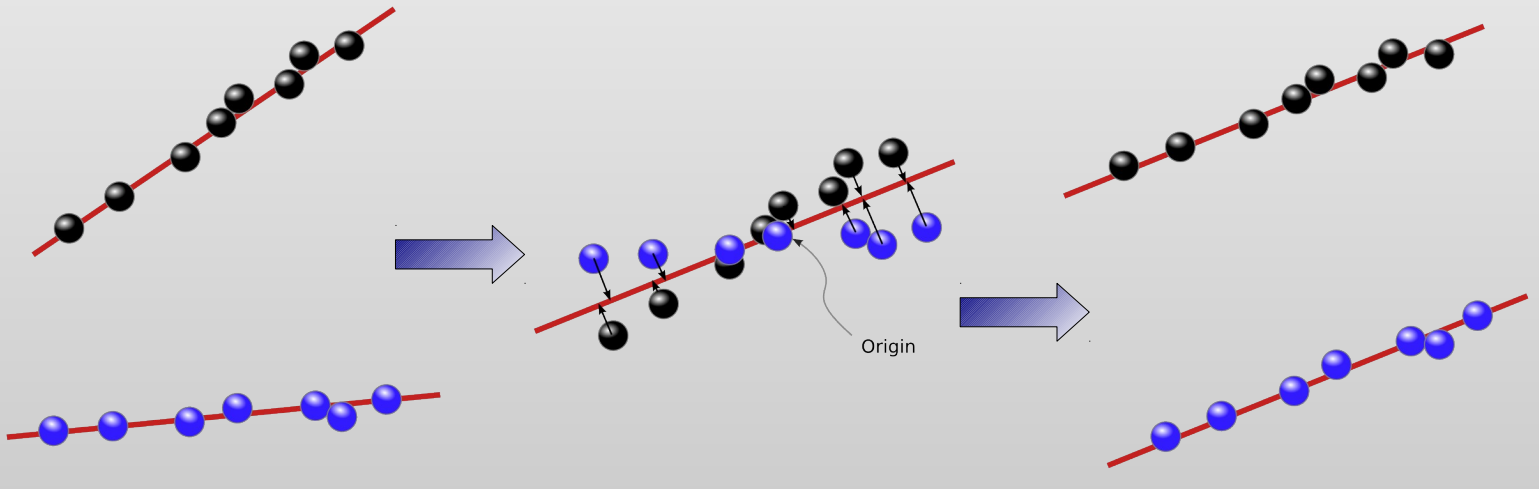


Parallel Plane Restraints



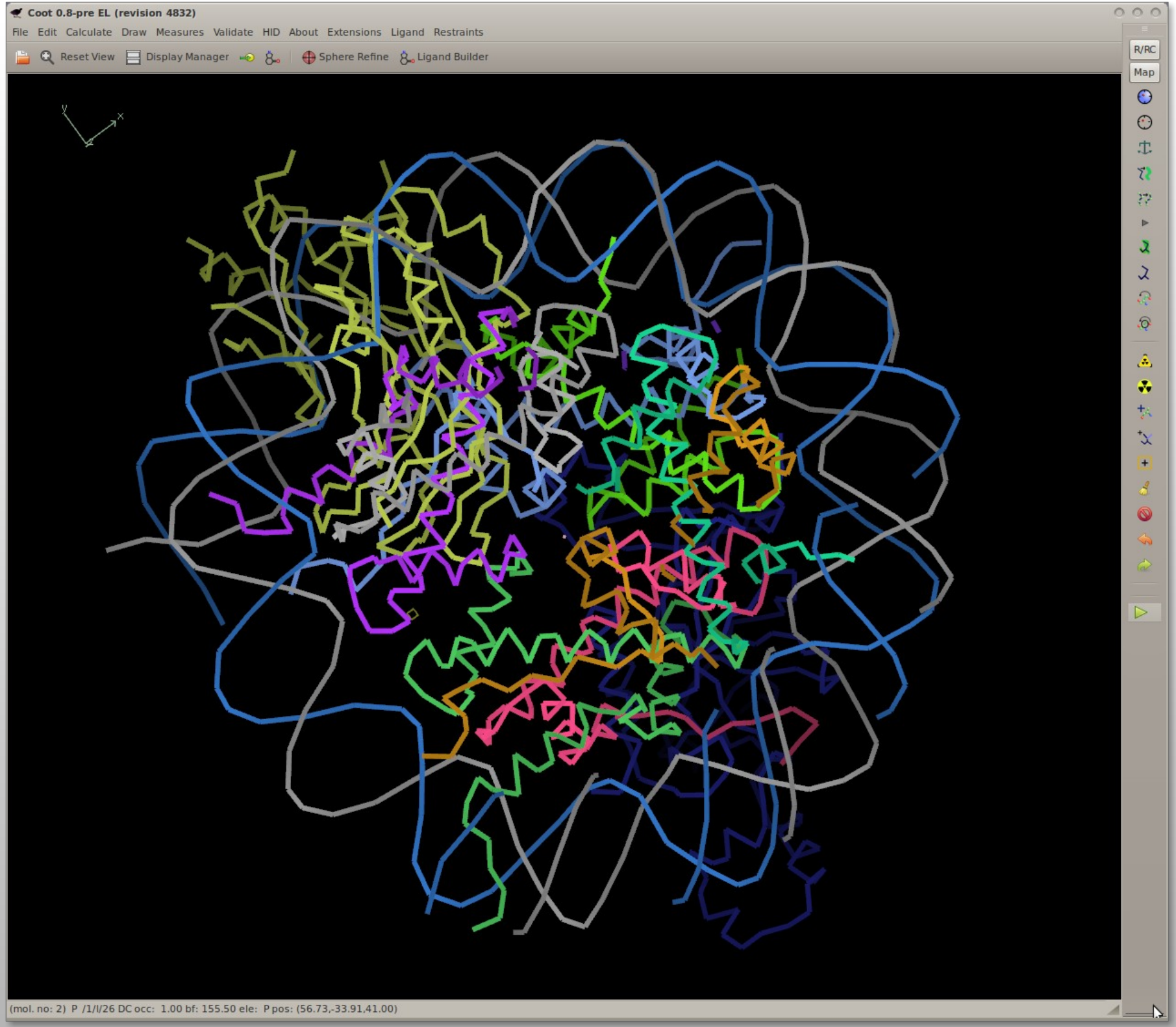
Shift to Origin

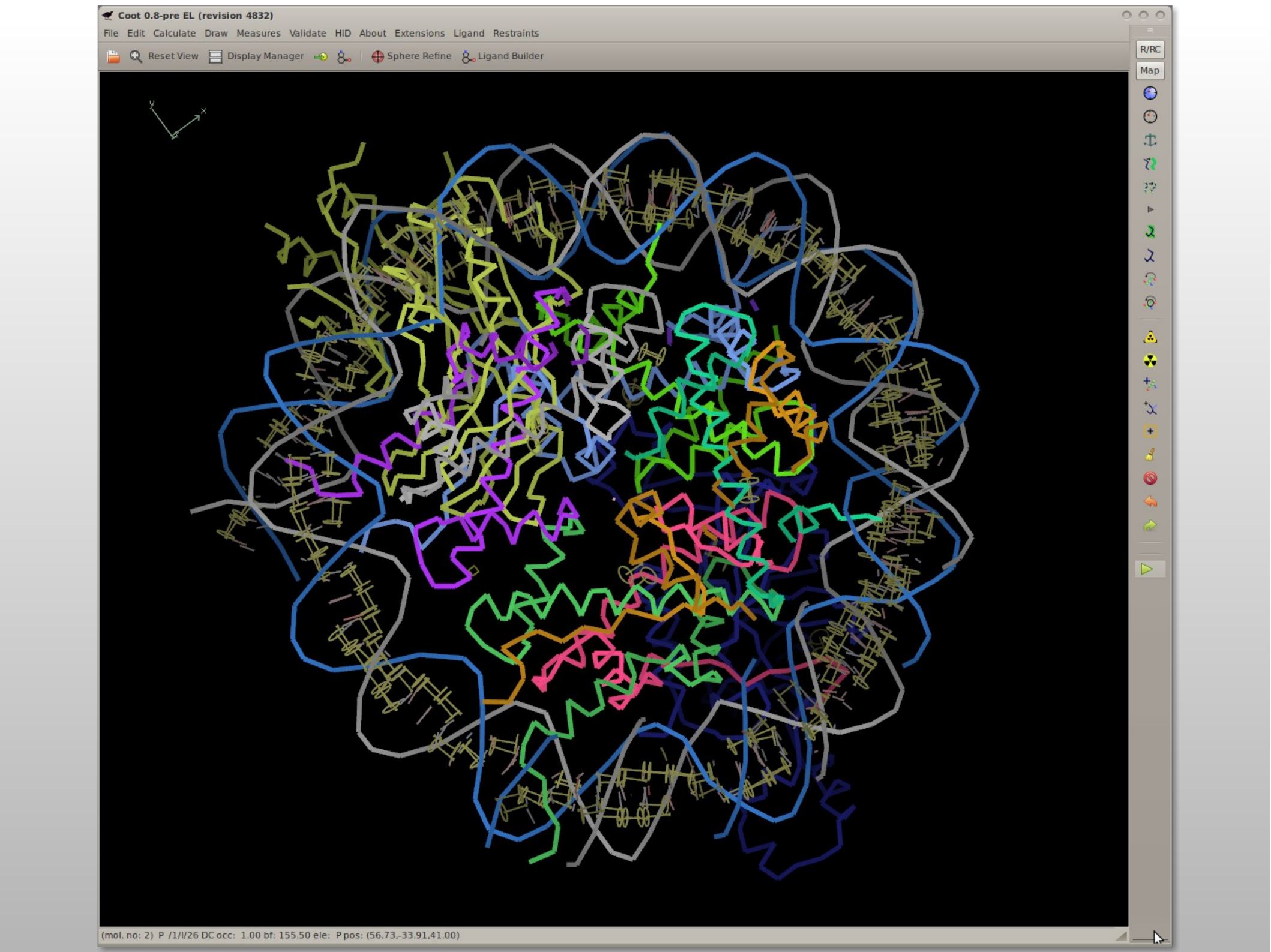
Parallel Planes Restraints

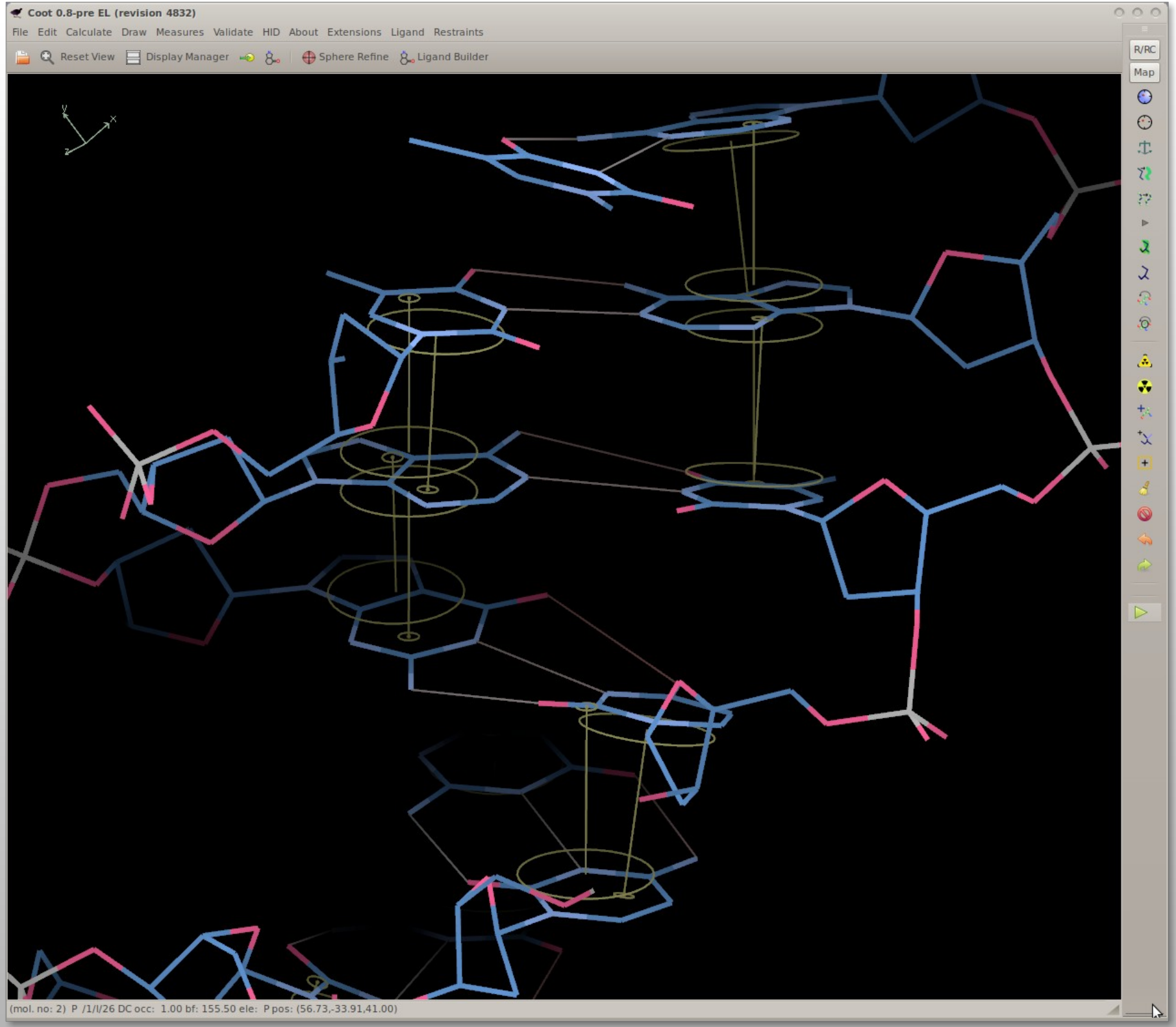


Shift to Origin

Move Back to Molecule


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

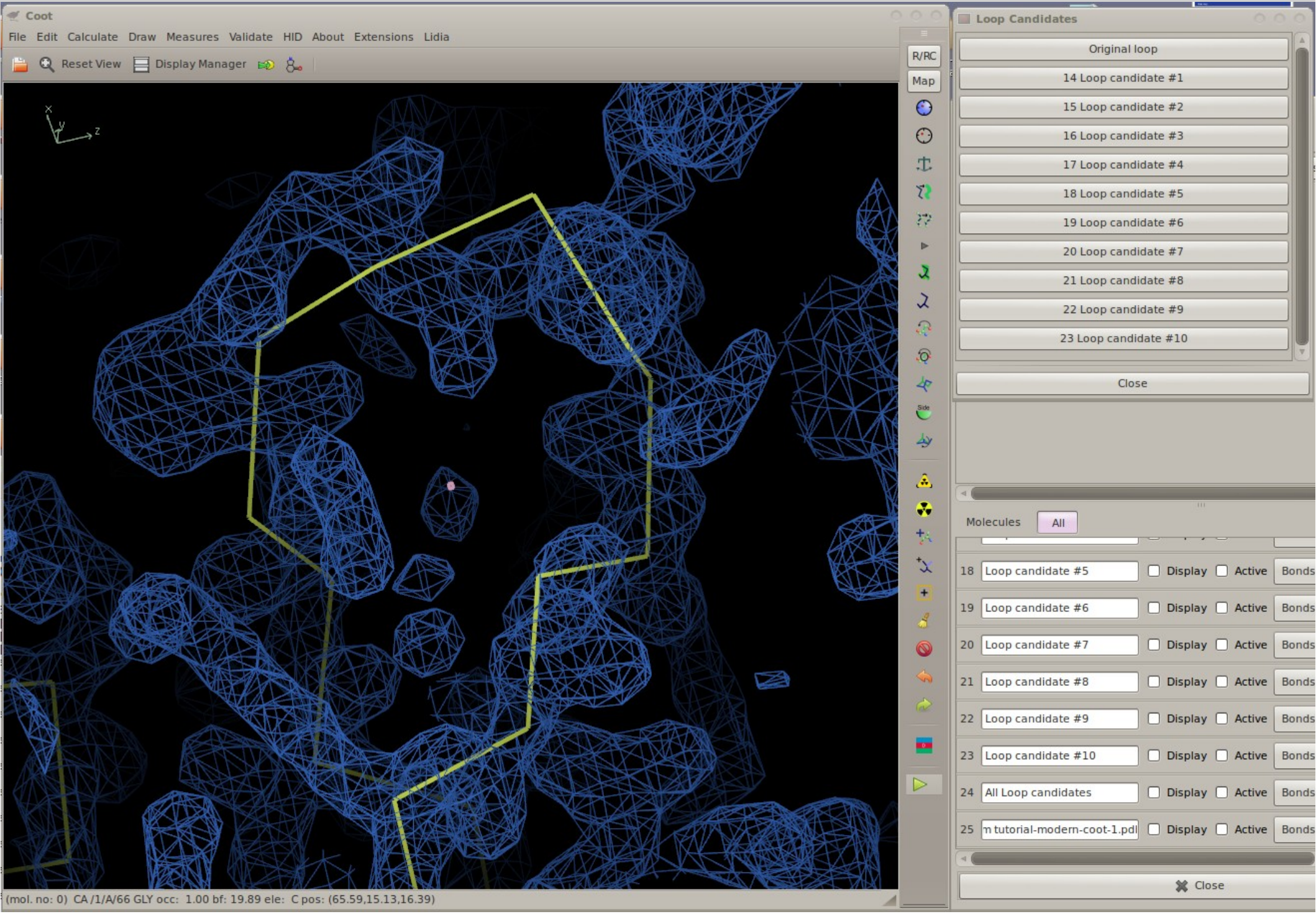

Coot 0.8-pre EL (revision 4832)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Restraints
Reset View
Display Manager
Sphere Refine
Ligand Builder
R/RC
Map
(mol. no: 2) P /1/I/26 DC occ: 1.00 bf: 155.50 ele: P pos: (56.73,-33.91,41.00)

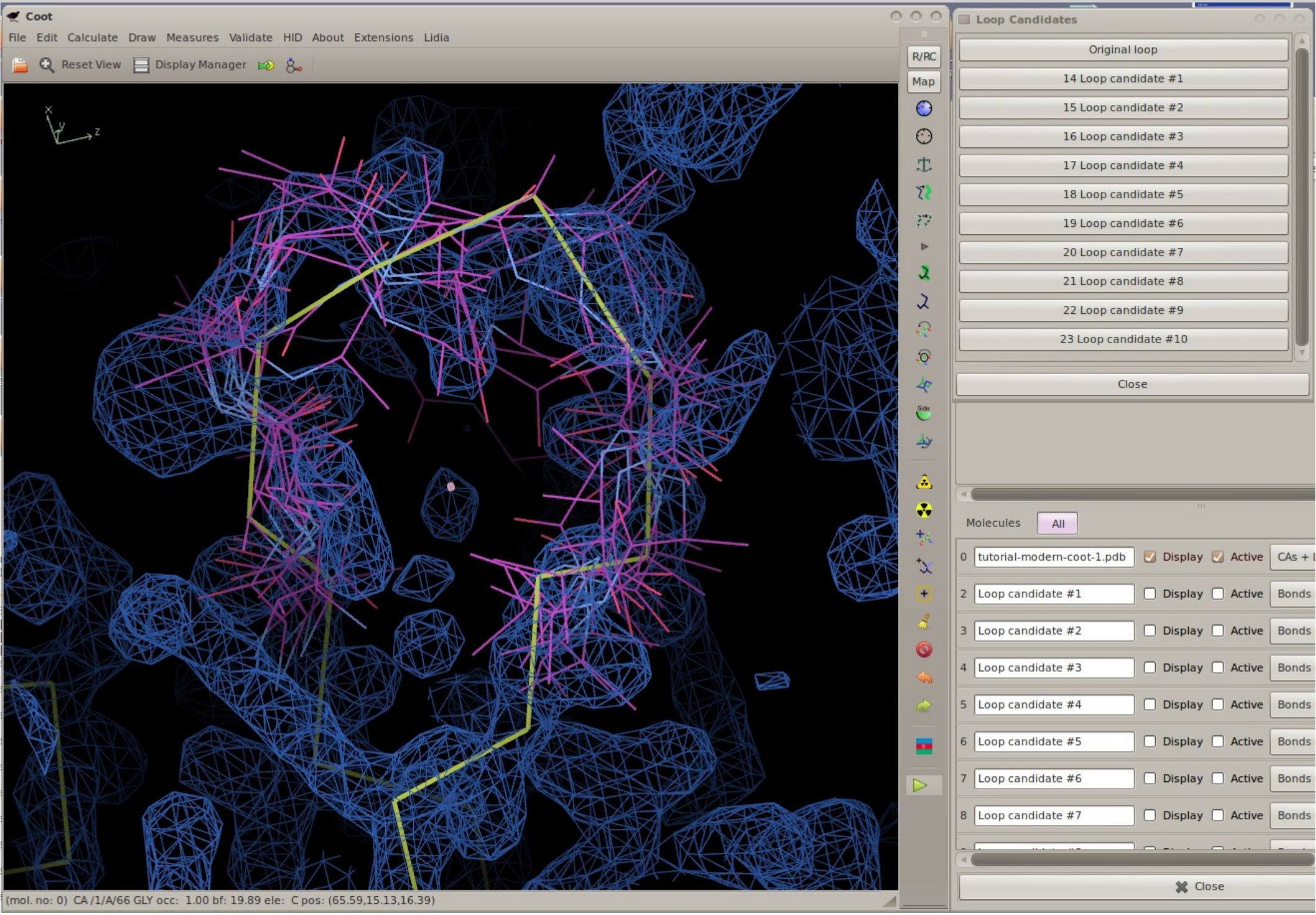

Coot 0.8-pre EL (revision 4832)
File Edit Calculate Draw Measures Validate HID About Extensions Ligand Restraints
Reset View
Display Manager
Sphere Refine
Ligand Builder
R/RC
Map
(mol. no: 2) P /1/I/26 DC occ: 1.00 bf: 155.50 ele: P pos: (56.73,-33.91,41.00)

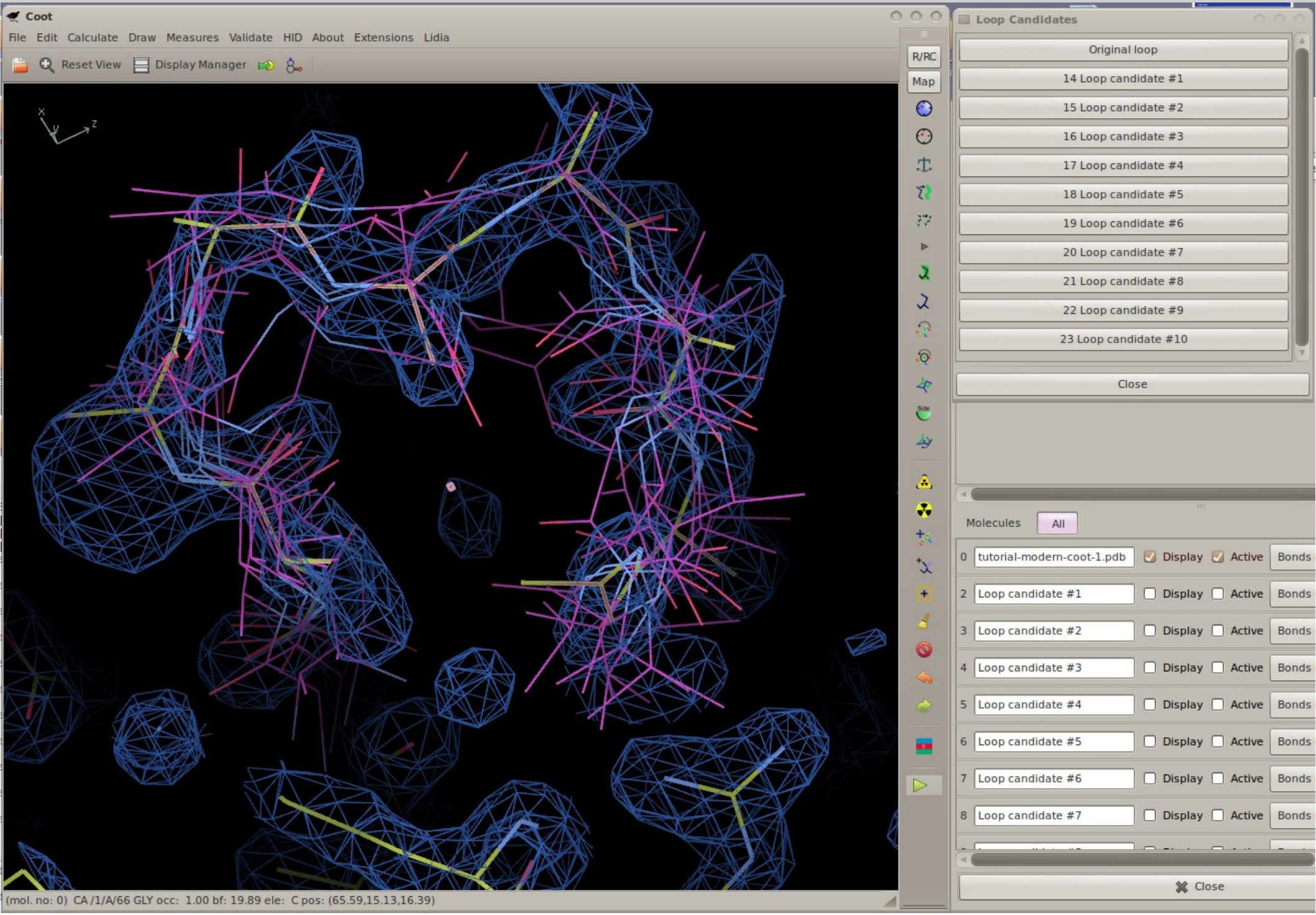
Loop Fitting Tool:

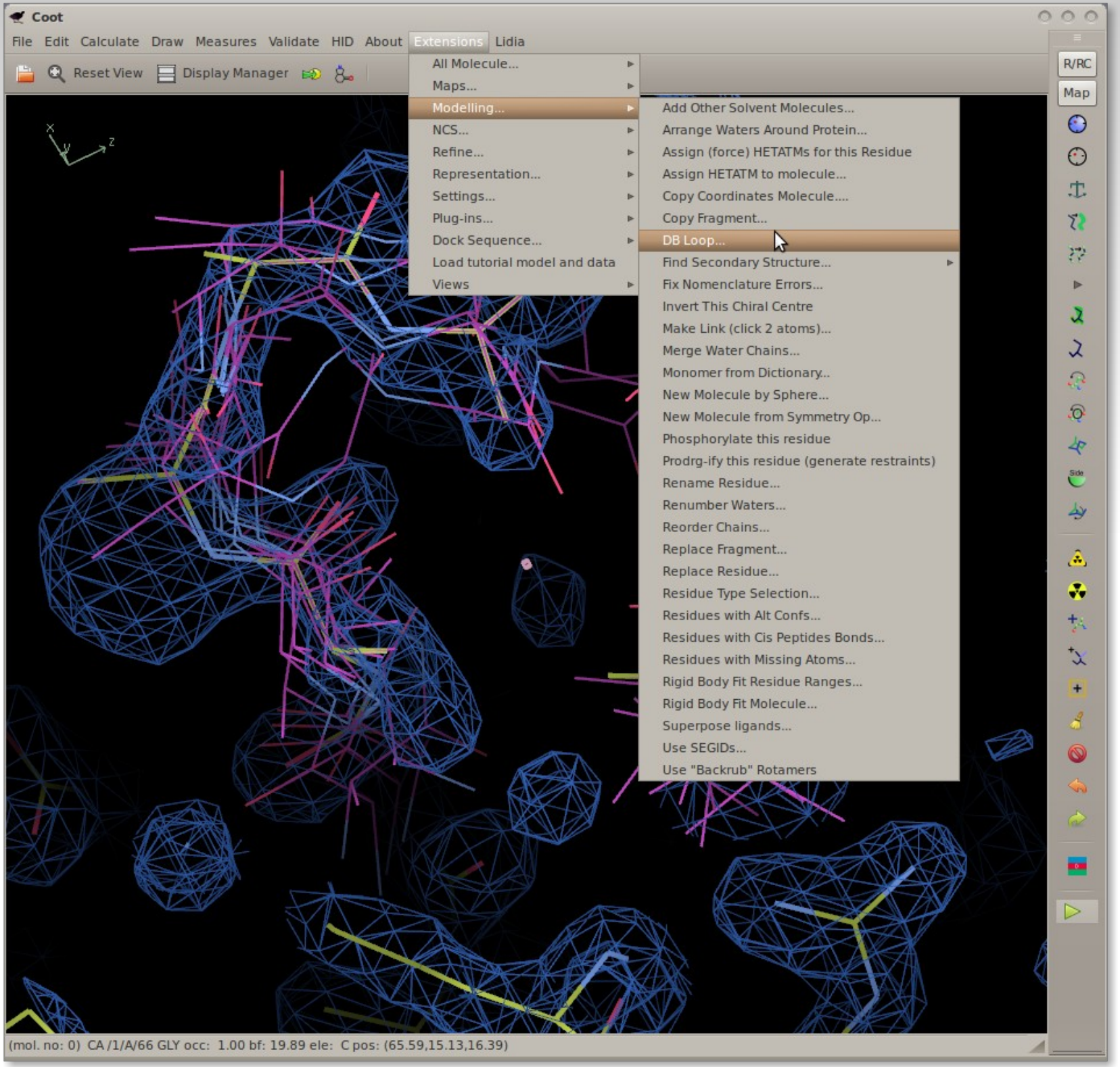
(sloop)

Kevin Cowtan


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Loop Candidates
Original loop
14 Loop candidate #1
15 Loop candidate #2
16 Loop candidate #3
17 Loop candidate #4
18 Loop candidate #5
19 Loop candidate #6
20 Loop candidate #7
21 Loop candidate #8
22 Loop candidate #9
23 Loop candidate #10
Close
Molecules
All
18 Loop candidate #5 Display Active Bonds
19 Loop candidate #6 Display Active Bonds
20 Loop candidate #7 Display Active Bonds
21 Loop candidate #8 Display Active Bonds
22 Loop candidate #9 Display Active Bonds
23 Loop candidate #10 Display Active Bonds
24 All Loop candidates Display Active Bonds
25 m tutorial-modern-coot-1.pdl Display Active Bonds
Close
(mol. no: 0) CA /1/A/66 GLY occ: 1.00 bf: 19.89 ele: C pos: (65.59,15.13,16.39)

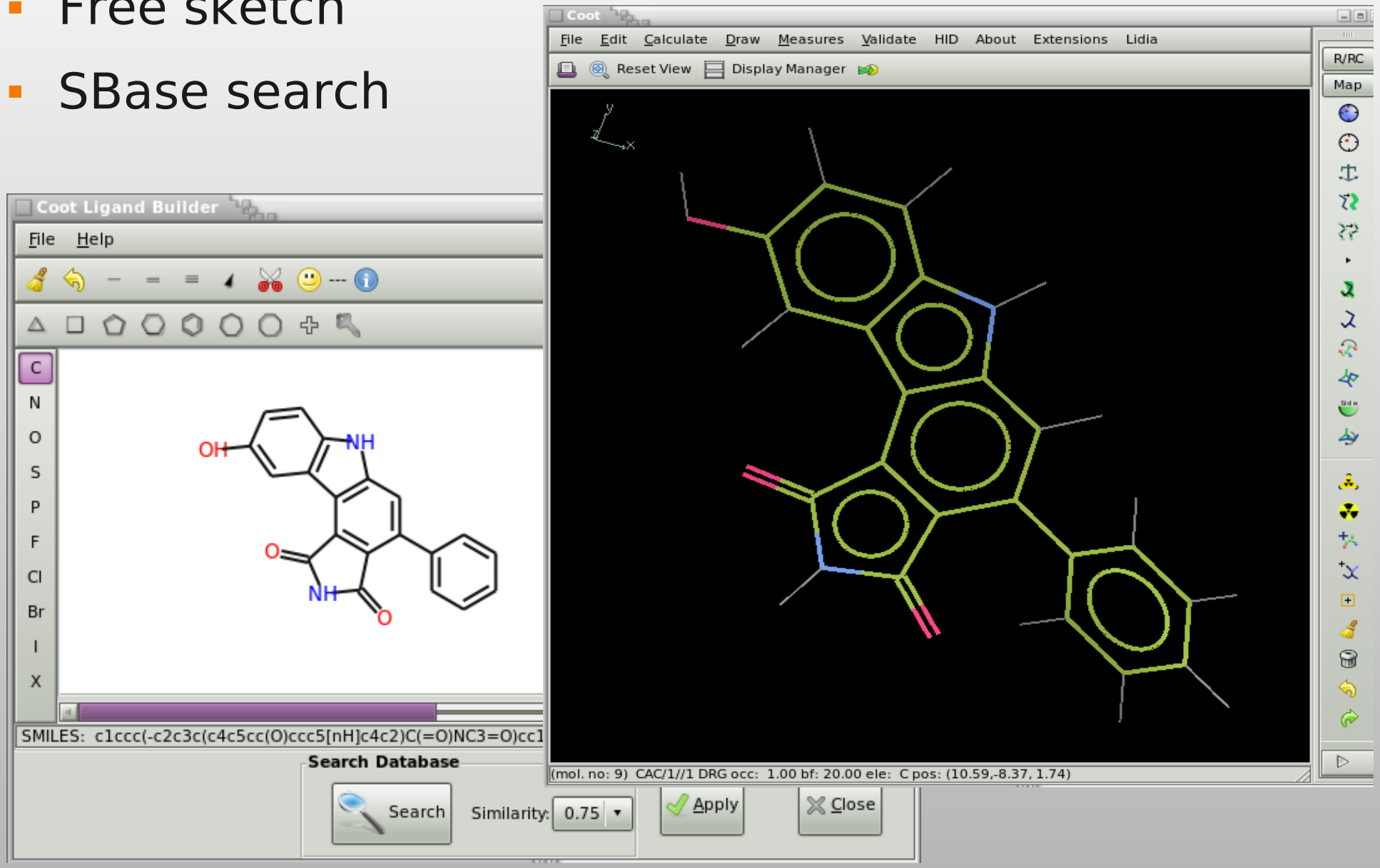

Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Loop Candidates
Original loop
14 Loop candidate #1
15 Loop candidate #2
16 Loop candidate #3
17 Loop candidate #4
18 Loop candidate #5
19 Loop candidate #6
20 Loop candidate #7
21 Loop candidate #8
22 Loop candidate #9
23 Loop candidate #10
Close
Molecules
All
0 tutorial-modern-coot-1.pdb Display Active CAs + I
2 Loop candidate #1 Display Active Bonds
3 Loop candidate #2 Display Active Bonds
4 Loop candidate #3 Display Active Bonds
5 Loop candidate #4 Display Active Bonds
6 Loop candidate #5 Display Active Bonds
7 Loop candidate #6 Display Active Bonds
8 Loop candidate #7 Display Active Bonds
Close
(mol. no: 0) CA /1/A/66 GLY occ: 1.00 bf: 19.89 ele: C pos: (65.59,15.13,16.39)


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
R/RC
Map
Loop Candidates
Original loop
14 Loop candidate #1
15 Loop candidate #2
16 Loop candidate #3
17 Loop candidate #4
18 Loop candidate #5
19 Loop candidate #6
20 Loop candidate #7
21 Loop candidate #8
22 Loop candidate #9
23 Loop candidate #10
Close
Molecules
All
0 tutorial-modern-coot-1.pdb Display Active Bonds
2 Loop candidate #1 Display Active Bonds
3 Loop candidate #2 Display Active Bonds
4 Loop candidate #3 Display Active Bonds
5 Loop candidate #4 Display Active Bonds
6 Loop candidate #5 Display Active Bonds
7 Loop candidate #6 Display Active Bonds
8 Loop candidate #7 Display Active Bonds
Close
(mol. no: 0) CA /1/A/66 GLY occ: 1.00 bf: 19.89 ele: C pos: (65.59,15.13,16.39)


Coot
File Edit Calculate Draw Measures Validate HID About Extensions Lidia
Reset View
Display Manager
All Molecule...
Maps...
Modelling...
NCS...
Refine...
Representation...
Settings...
Plug-ins...
Dock Sequence...
Load tutorial model and data
Views
Add Other Solvent Molecules...
Arrange Waters Around Protein...
Assign (force) HETATMs for this Residue
Assign HETATM to molecule...
Copy Coordinates Molecule....
Copy Fragment...
DB Loop...
Find Secondary Structure...
Fix Nomenclature Errors...
Invert This Chiral Centre
Make Link (click 2 atoms)...
Merge Water Chains...
Monomer from Dictionary...
New Molecule by Sphere...
New Molecule from Symmetry Op...
Phosphorylate this residue
Prodrg-ify this residue (generate restraints)
Rename Residue...
Renumber Waters...
Reorder Chains...
Replace Fragment...
Replace Residue...
Residue Type Selection...
Residues with Alt Confs...
Residues with Cis Peptides Bonds...
Residues with Missing Atoms...
Rigid Body Fit Residue Ranges...
Rigid Body Fit Molecule...
Superpose ligands...
Use SEGIDs...
Use "Backrub" Rotamers
R/RC
Map
(mol. no: 0) CA /1/A/66 GLY occ: 1.00 bf: 19.89 ele: C pos: (65.59,15.13,16.39)

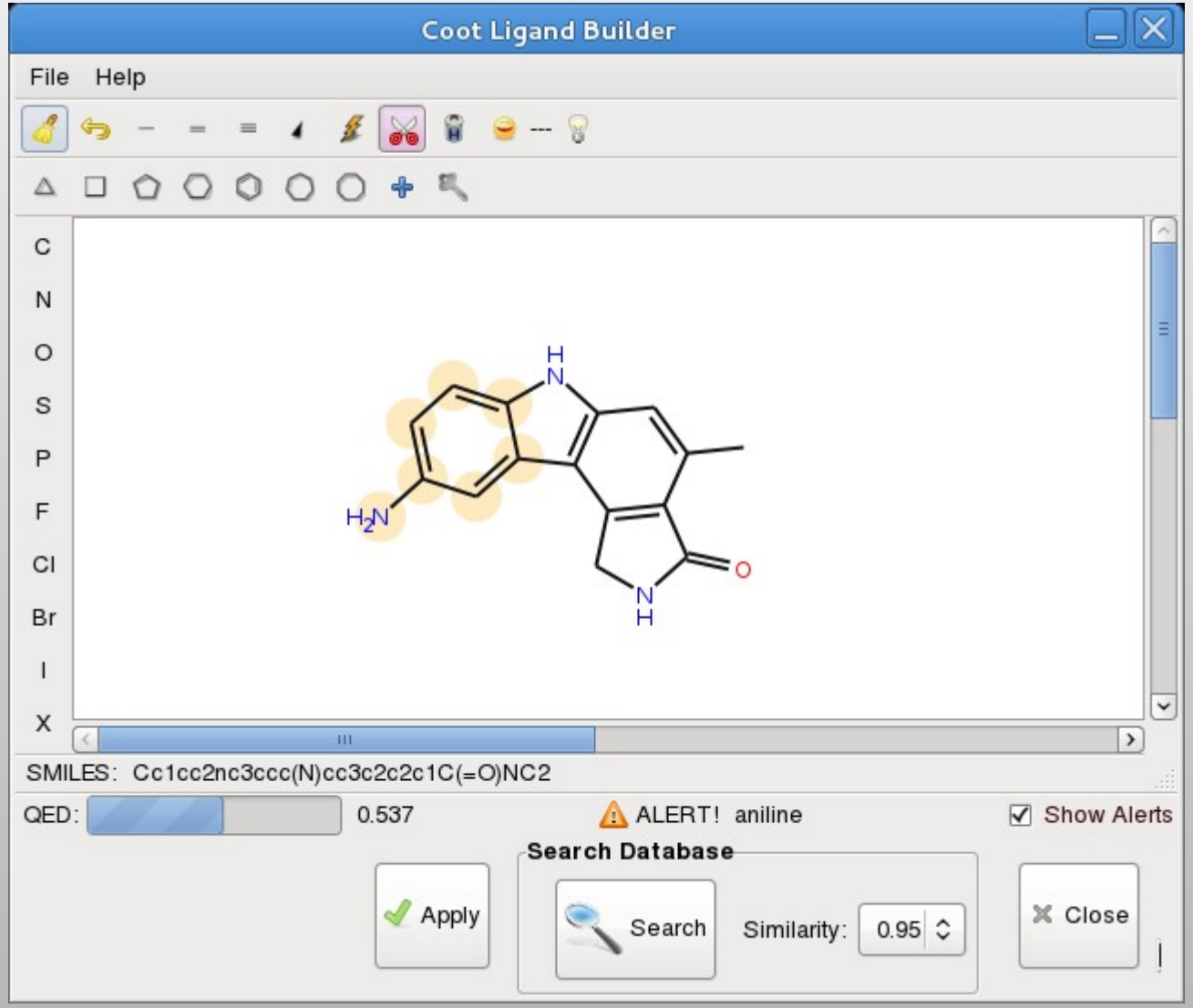
2D Ligand Builder

- Free sketch
- SBase search



2D Sketcher

- Structural Alerts



•Check *vs*. vector of SMARTS
  - (from Biscu-it)

QED Score

Quantitative Evaluation of Drug-likeness

ARTICLES
PUBLISHED ONLINE: 24 JANUARY 2012 | DOI: 10.1038/NCHEM.1243
nature chemistry

Quantifying the chemical beauty of drugs

G. Richard Bickerton[1], Gaia V. Paolini[2], Jérémy Besnard[1], Sorel Muresan[3] and Andrew L. Hopkins[1]*

Drug-likeness is a key consideration when selecting compounds during the early stages of drug discovery. However, evaluation of drug-likeness in absolute terms does not reflect adequately the whole spectrum of compound quality. More worryingly, widely used rules may inadvertently foster undesirable molecular property inflation as they permit the encroachment of rule-compliant compounds towards their boundaries. We propose a measure of drug-likeness based on the concept of desirability called the quantitative estimate of drug-likeness (QED). The empirical rationale of QED reflects the underlying distribution of molecular properties. QED is intuitive, transparent, straightforward to implement in many practical settings and allows compounds to be ranked by their relative merit. We extended the utility of QED by applying it to the problem of molecular target druggability assessment by prioritizing a large set of published bioactive compounds. The measure may also capture the abstract notion of aesthetics in medicinal chemistry.

The concept of drug-likeness provides useful guidelines for early-stage drug discovery[1,2]. Analysis of the observed distribution of some key physicochemical properties of approved drugs, including molecular mass (M_r), hydrophobicity and polarity, reveals that they occupy preferentially a relatively narrow range of possible values[3]. Compounds that fall within this range are described as 'drug-like'. This definition holds in the absence of any obvious structural similarity to an approved drug. It has been shown that the preferential selection of drug-like compounds increases the likelihood of surviving the well-documented high rates of attrition in drug discovery[4].

Drug-likeness can be rationalized by considering how simple physicochemical properties impact molecular behaviour *in vivo*, with particular respect to solubility, permeability, metabolic stability and transporter effects. Indeed, drug-likeness is often used as a proxy for oral bioavailability. However, drug-likeness provides a broad composite descriptor that implicitly captures several criteria,

Paradoxically, since the publication of the seminal paper by Lipinski *et al.*[5] there appears to be a growing epidemic, which Hann has termed 'molecular obesity'[6], among new pharmacological compounds (Supplementary Fig. S1). Compounds with higher relative M_r and lipophilicity have a higher probability of attrition at each stage of clinical development[4,9–11]. Thus, the inflation of physicochemical properties that increases the risks associated with clinical development may explain, in part, the decline in productivity of small-molecule drug discovery over the past two decades[4]. However, the mean molecular properties of new pharmacological compounds are still considered Lipinski compliant, even though their property distributions are far from historical norms.

Although the Ro5 is predictive of oral bioavailability, 16% of oral drugs violate at least one of the criteria and 6% fail two or more (although this does include natural products and substrates of transporters) (Supplementary Fig. S2a and Supplementary Table S1). High-profile drugs, such as atorvastatin (Lipitor) and montelukast

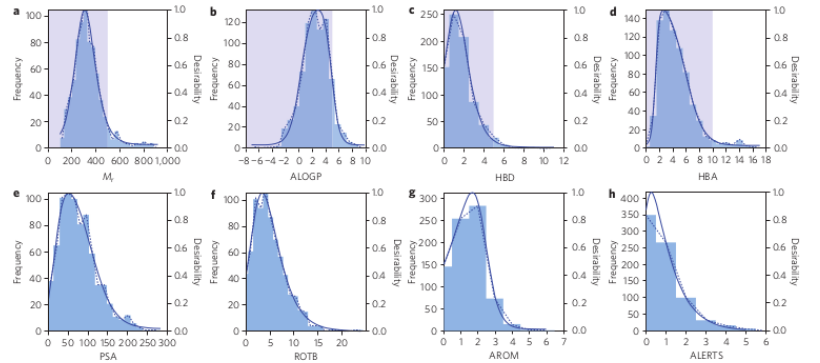


Figure 1 | Histograms of eight selected molecular properties for a set of 771 orally absorbed small molecule drugs. a–h. Molecular properties M_r (**a**), lipophilicity estimated by atom-based prediction of ALOGP (**b**), number of HBDs (**c**), number of HBAs (**d**), PSA (**e**), number of ROTBs (**f**), number of AROMs (**g**) and number of ALERTS (**h**). The Lipinski-compliant areas are shown in pale blue in (**a**), (**b**), (**c**) and (**d**). The solid blue lines describe the ADS functions (equation (2)) used to model the histograms. The parameters for each function are given in Supplementary Table S1.

design[17,18], prioritization of molecular targets, penetration of the central nervous system[19] and estimating the reliability of screening data[20]. The concept was introduced originally by Harrington[15] in the area of process engineering and further refined by Derringer and Suich[21]. Desirability takes multiple numerical or categorical parameters measured on different scales and describes each by an individual desirability function. These are then integrated into a single dimensionless score. In the case of compounds, a series of desirability functions (d) are derived, each of which corresponds to a different molecular descriptor. Combining the individual desirability functions into the QED is achieved by taking the geometric

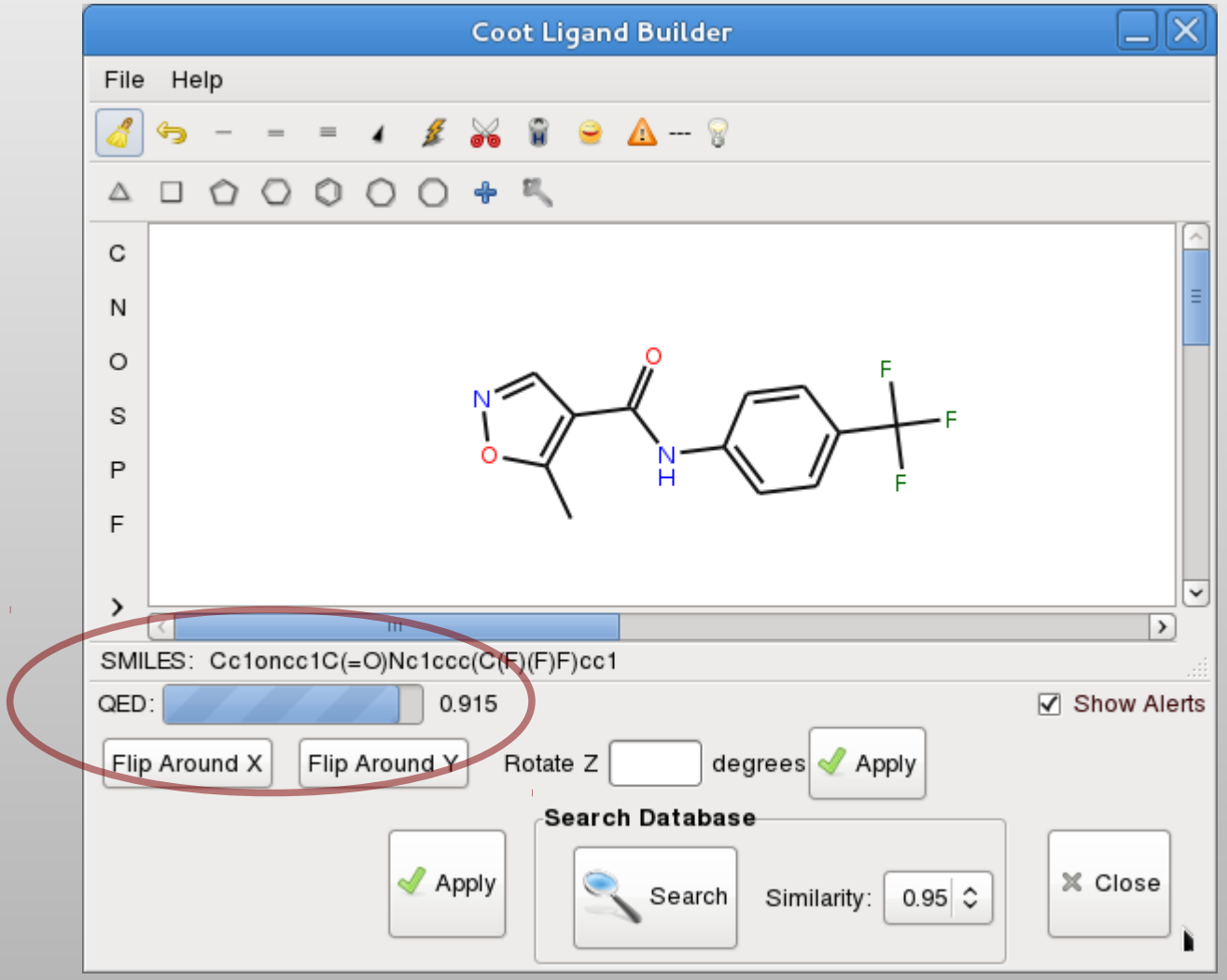
asymmetric double sigmoidal (ADS) functions, which are also shown in Fig. 1 over the same range. The general ADS function is shown in equation (2), where $d(x)$ is the desirability function for molecular descriptor x:

$$d(x) = a + \frac{b}{\left[1+\exp\left(-\frac{x-c+\frac{d}{2}}{\quad}\right)\right]}\left[1-\frac{1}{\left[1+\exp\left(-\frac{x-c-\frac{d}{2}}{\quad}\right)\right]}\right]$$

Bickerton et al (2012) *Nature Chemistry*

2D Sketcher

- QED score



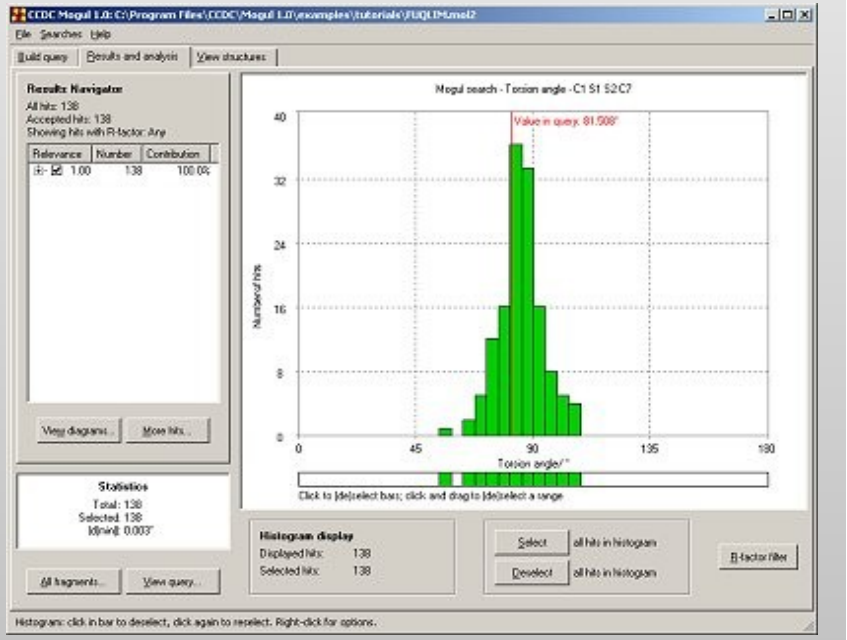
Silicos-it's
Biscu-it™

Ligand Utils

- “Get Molcule”
  - Uses network connection to Wikipedia
- Get *comp-id* ligand-description from PDBe
  - downloads and reads (*e.g.*) AAA.cif
    - (extracted from chemical component library)
- Drag and drop
  - Uses network connection to get URLs
  - or file-system files
- pyrogen, acedrg
  - restraints generation

Parmatisation issues... (what if they are wrong?)

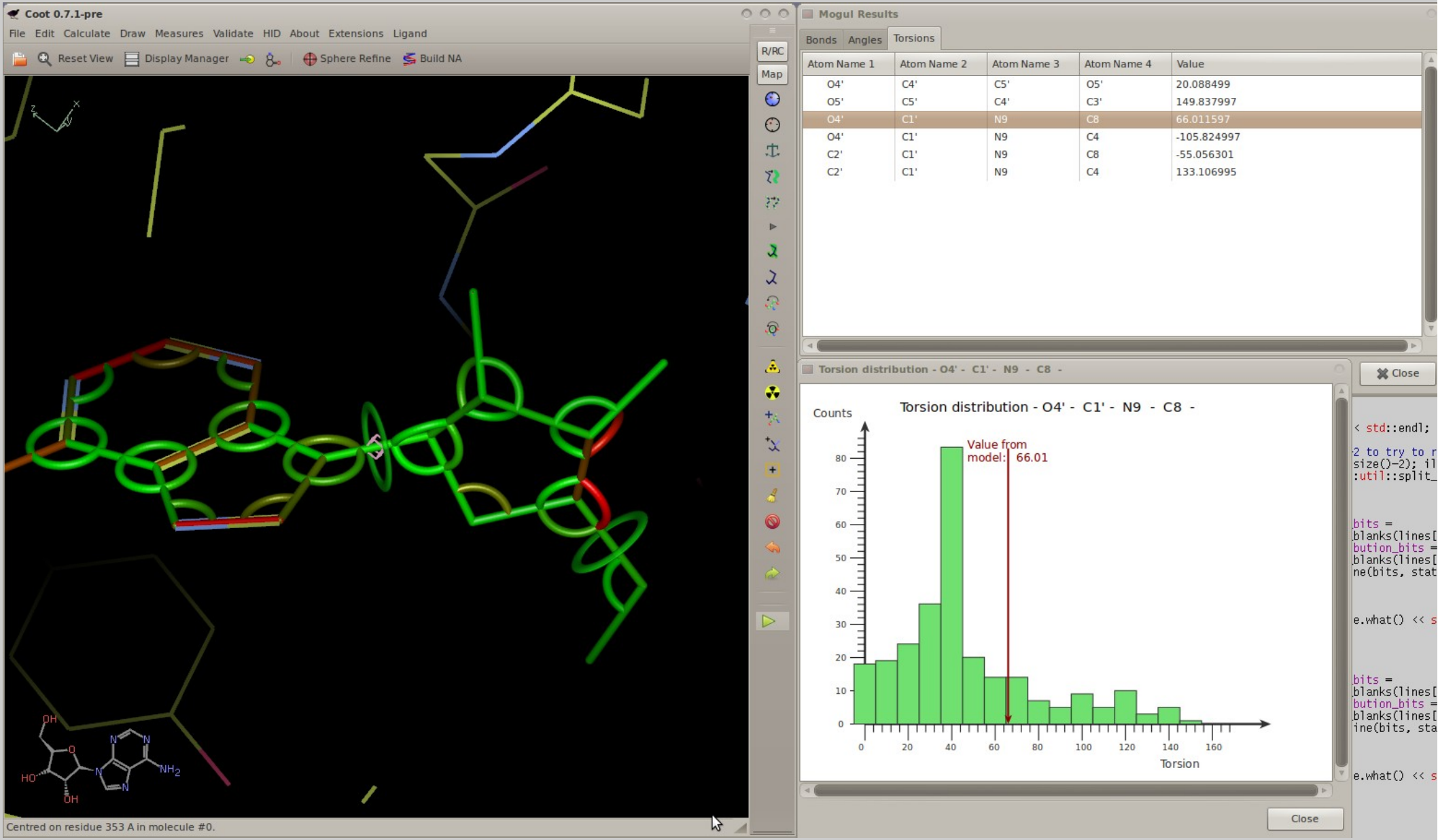
- Perfect refinement with incorrect parameters
  → distorted structure
- CSD's Mogul
  - Knowledge-base of geometric parameters based on the CSD
  - Can be run as a “batch job”
  - Mean, median, mode, quartiles, Z-scores.



Ligand Validation

- Mogul plugin in Coot
  - Run mogul, graphical display of results
  - Update restraints (target and esds for bonds and angles)
  - CSD data not so great for plane, chiral and torsion restraints
    - (not by me, anyway)

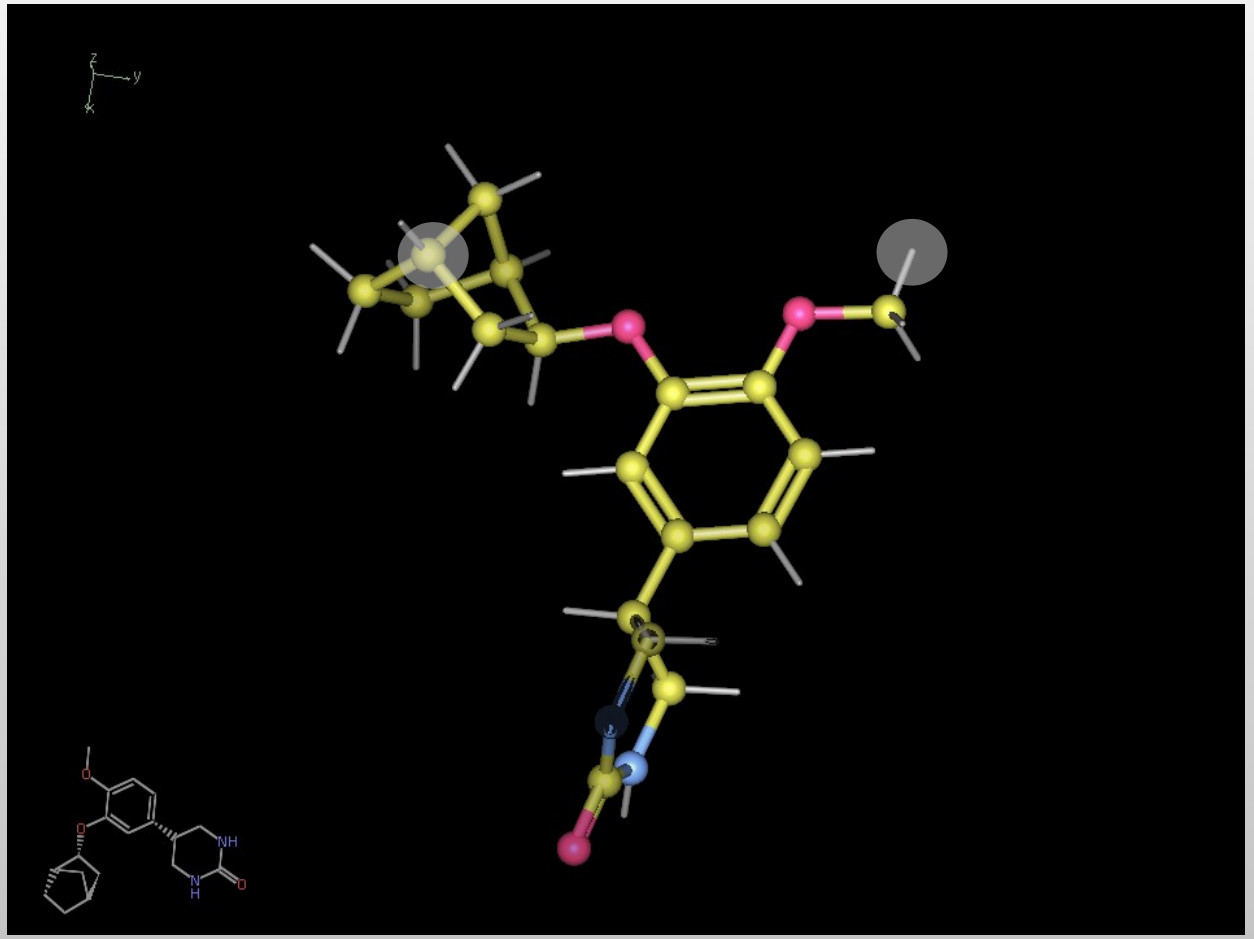
Mogul Results Representation



New Software for Restraints Generation

COD Atom Types

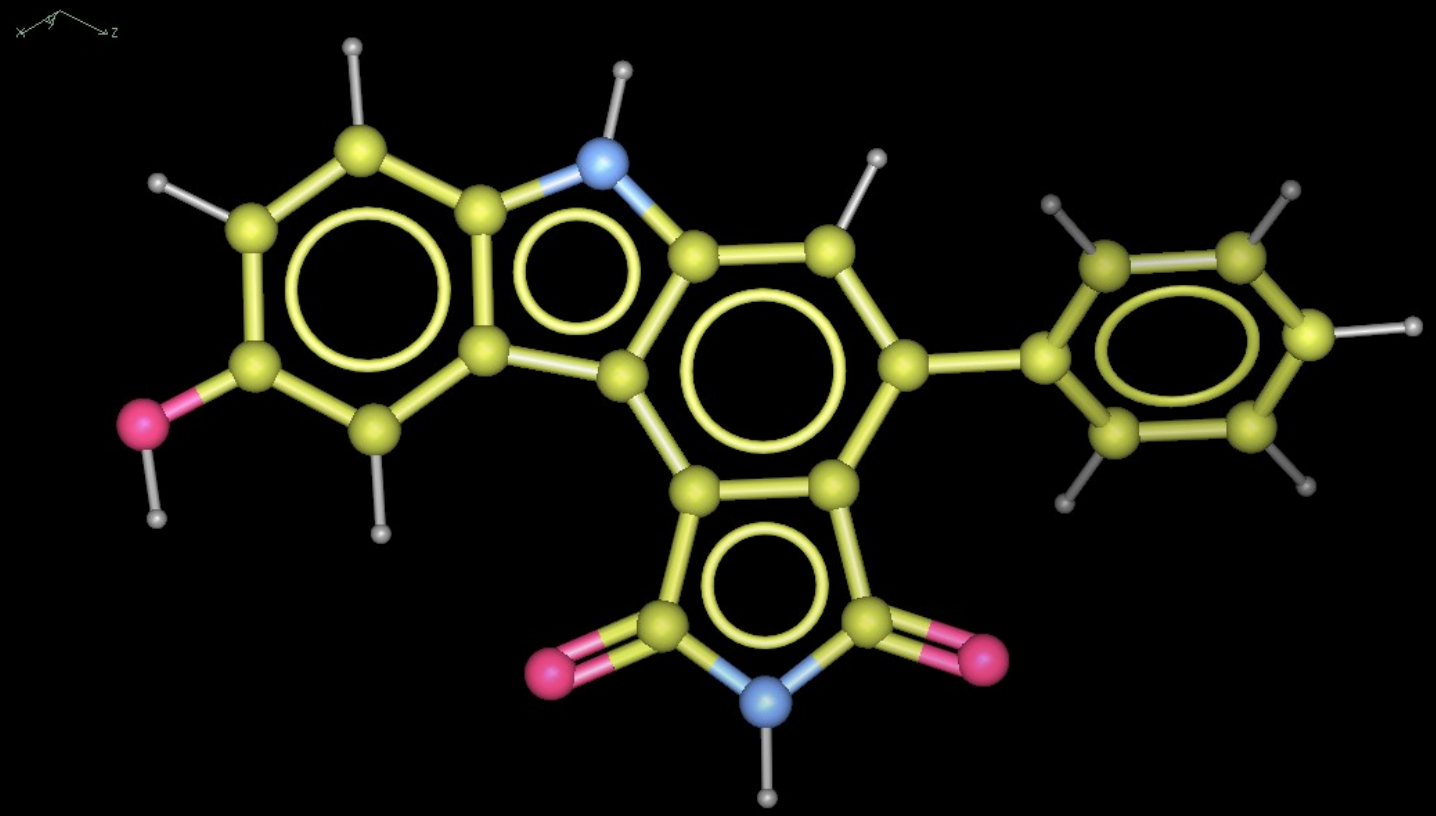
- COD
  - 2nd order-based



H1B: H(CHHO)
C9: C[5,5,6](C[5,5]CHH)(C[5,6]CHH)(C[5,6]CHO)(H)

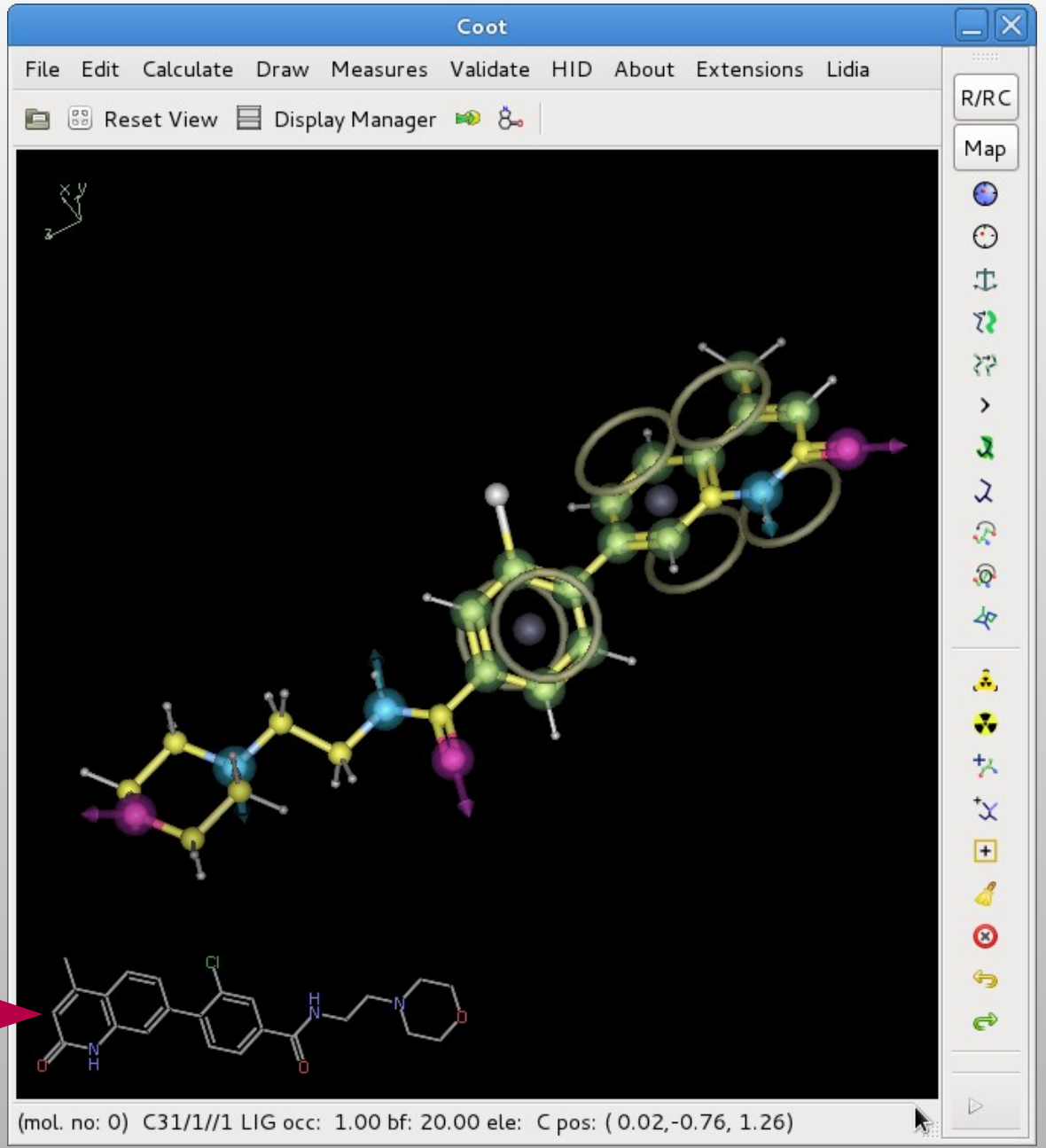
Ligand Represenation

- Bond orders (from dictionary restraints)



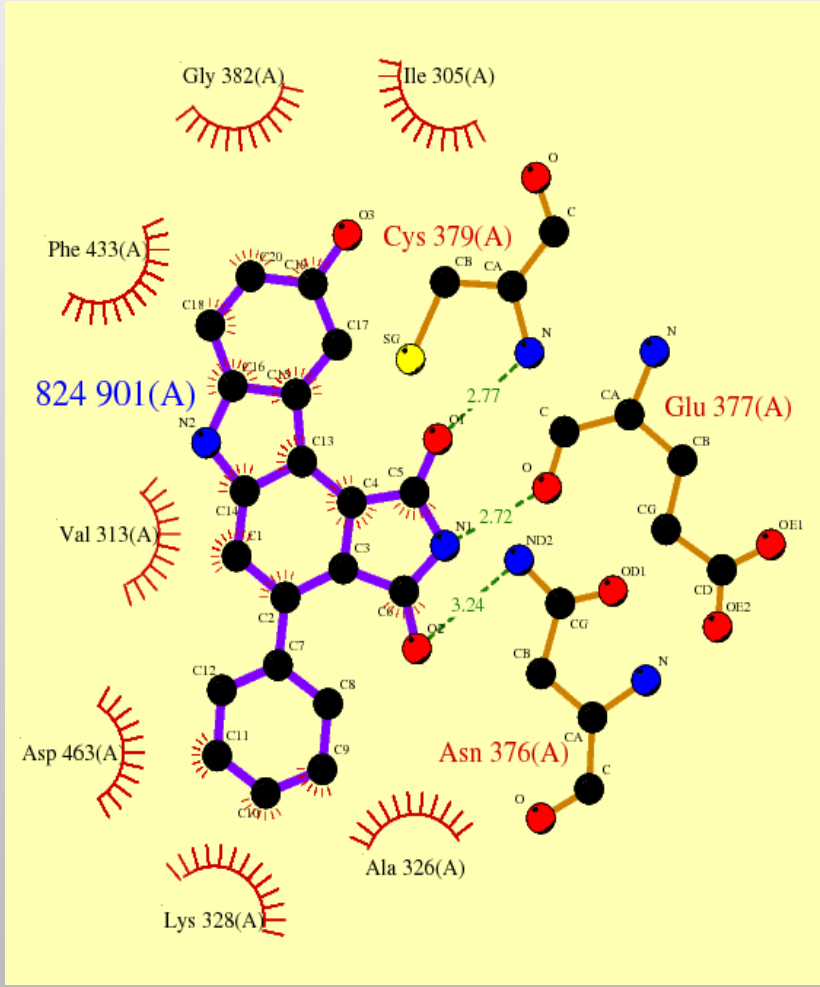
Chemical Features

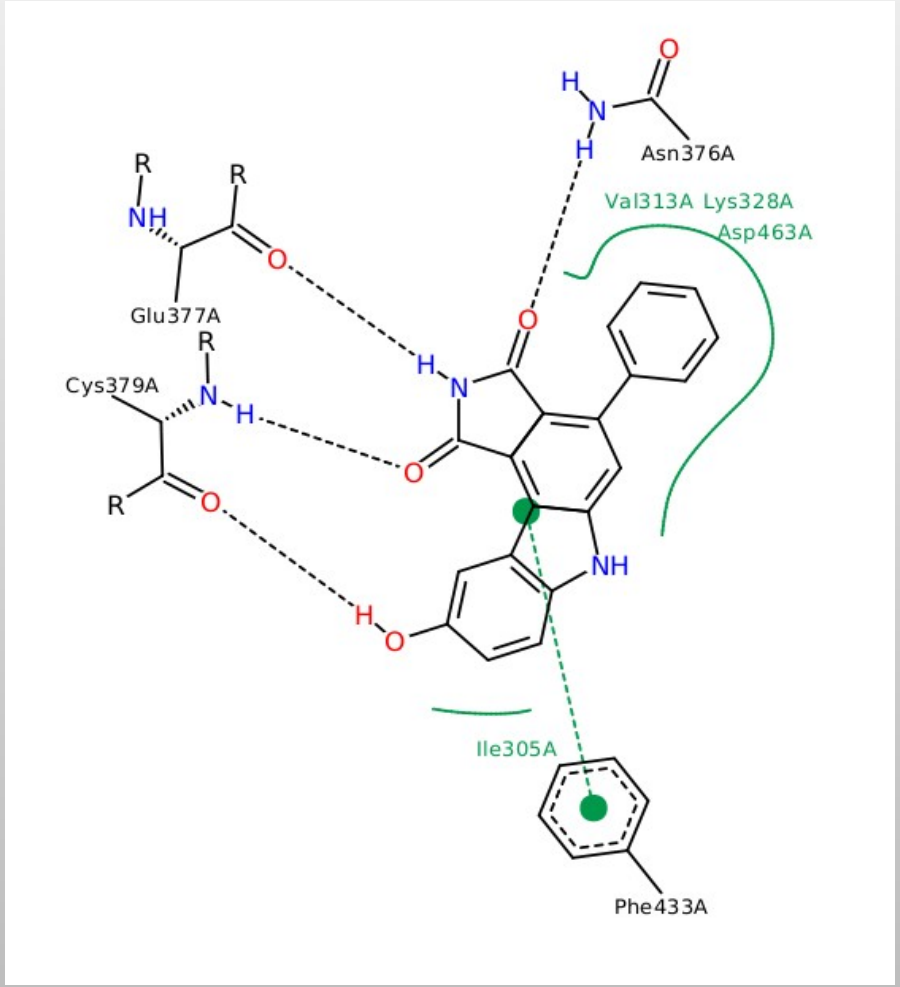
...and on the fly thumbnailing



Ligand Environment Layout

- 2d Ligand pocket layout (ligplot, poseview)





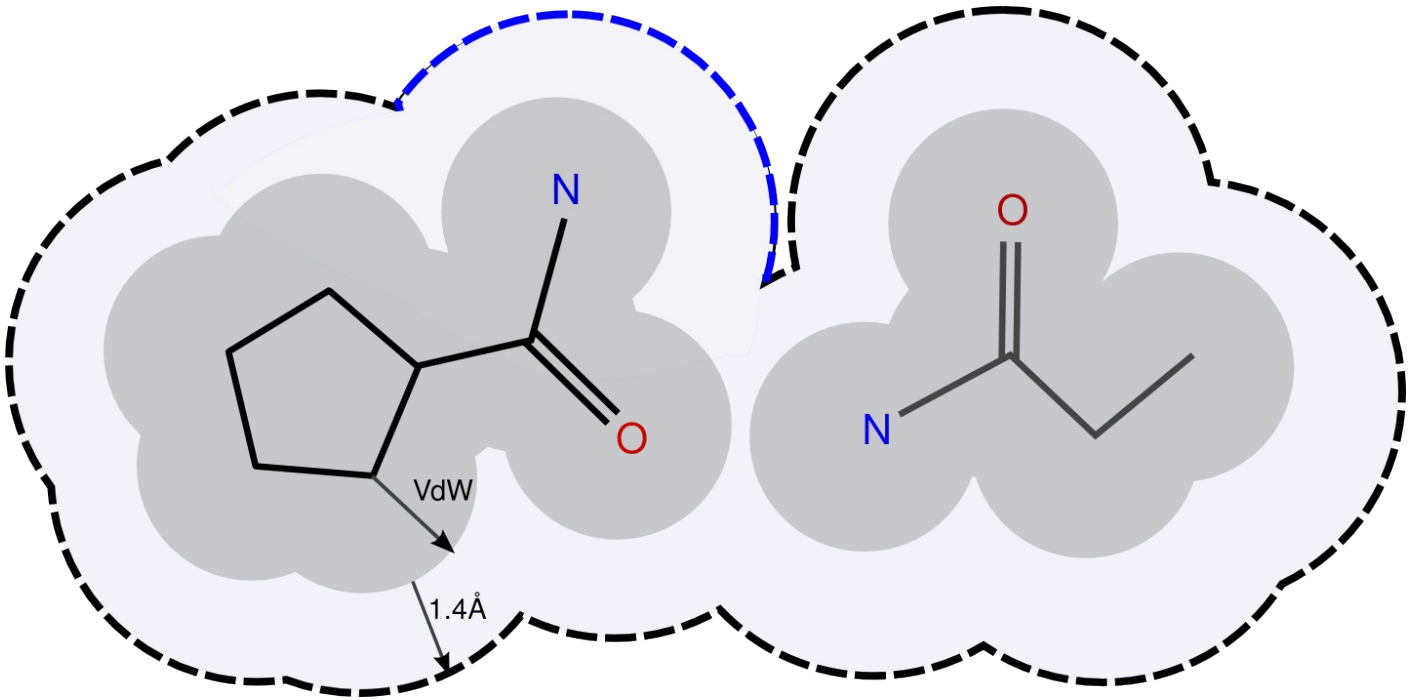
Can we do better? - Interactivity?

Ligand Environment Layout

- Binding pocket residues
- Interactions
- Substitution contour
- Solvent accessibility halos
- Solvent exclusion by ligand

Solvent Exposure

- Identification of solvent accessible atoms



Ligand Enviroment Layout

- Considerations
  - 2D placement and distances should reflect 3D metrics (as much as possible)
    - H-bonded residues should be close the atoms to which they are bonded
  - Residues should not overlap the ligand
  - Residues should not overlap each other
  - *c.f.* Clark & Labute (2007)

Layout Energy Terms

$$E = \sum\sum w_{ij}(d_{ij}^2 - D_{ij}^2) +$$

Residues match 3D Distances

$$\sum\sum \exp(-\frac{1}{2}d_{ij}^2) +$$

Residues don't overlay each other

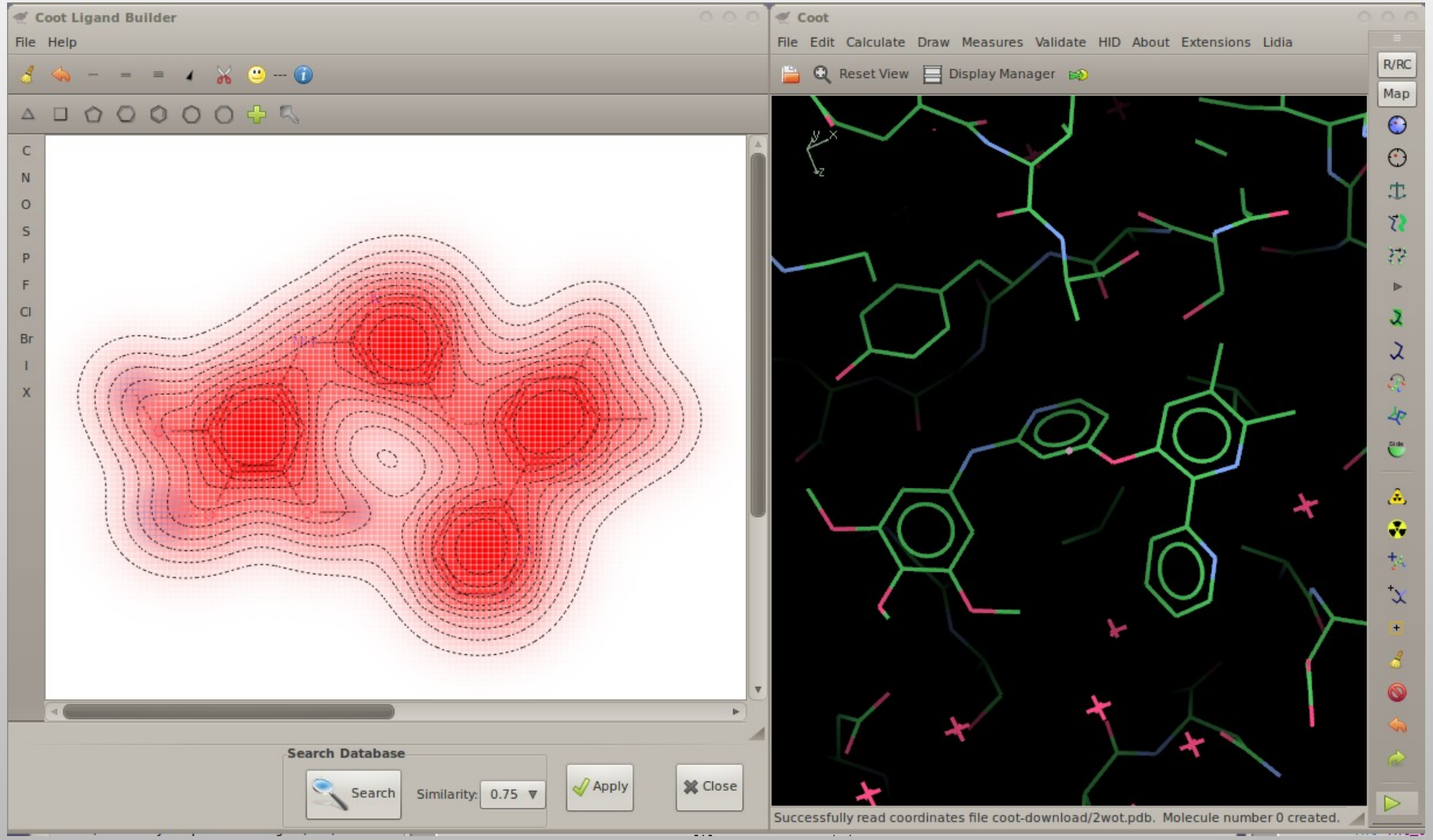
$$\sum\sum (d_{ik}^2 - D_{ik}^2) +$$

Residues are close to H-bonding ligand atoms

$$\sum\sum \exp(-\frac{1}{2}d_{ik}^2)$$

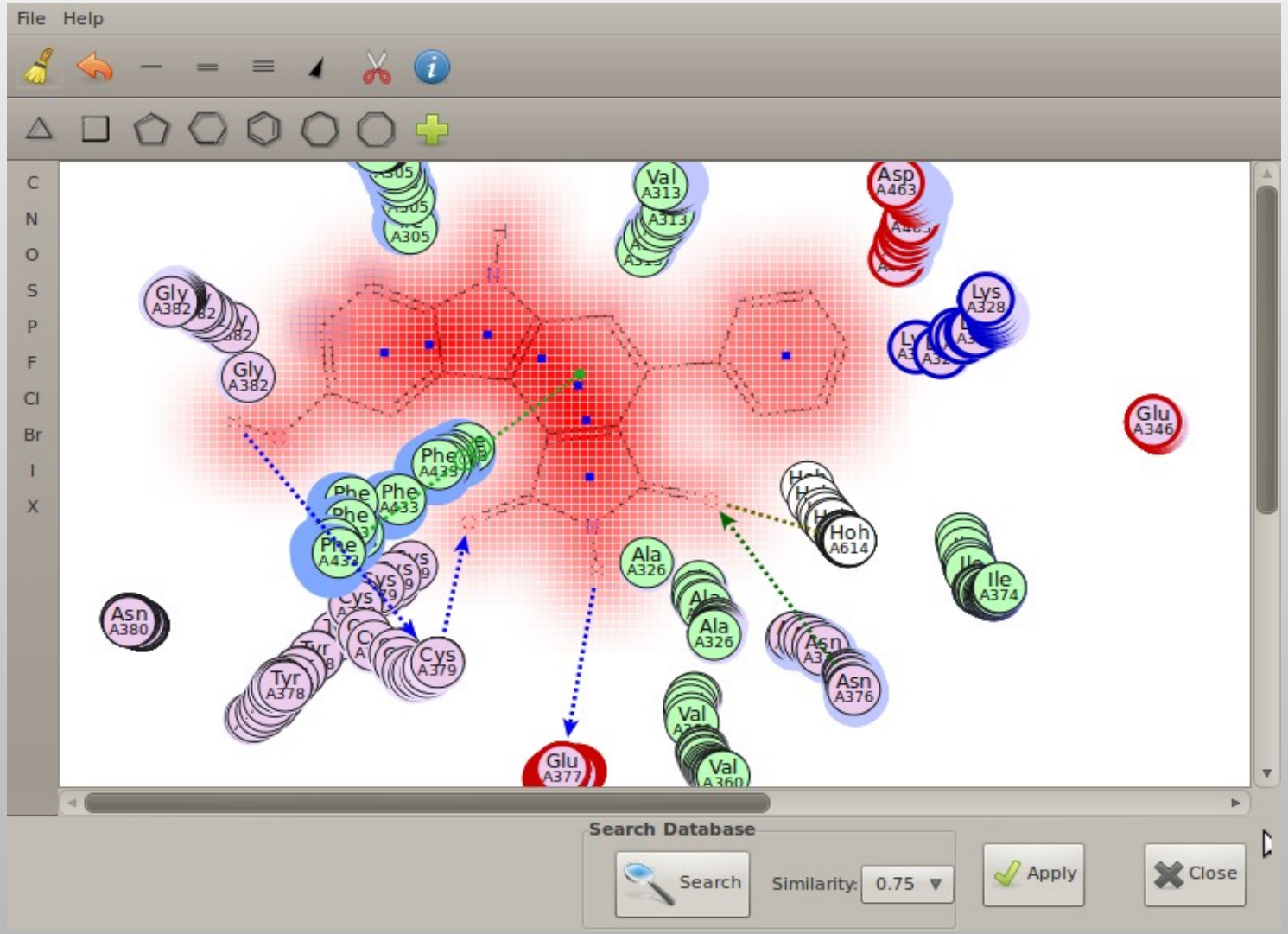
Residues don't overlap ligand

”Don't overlap the ligand”



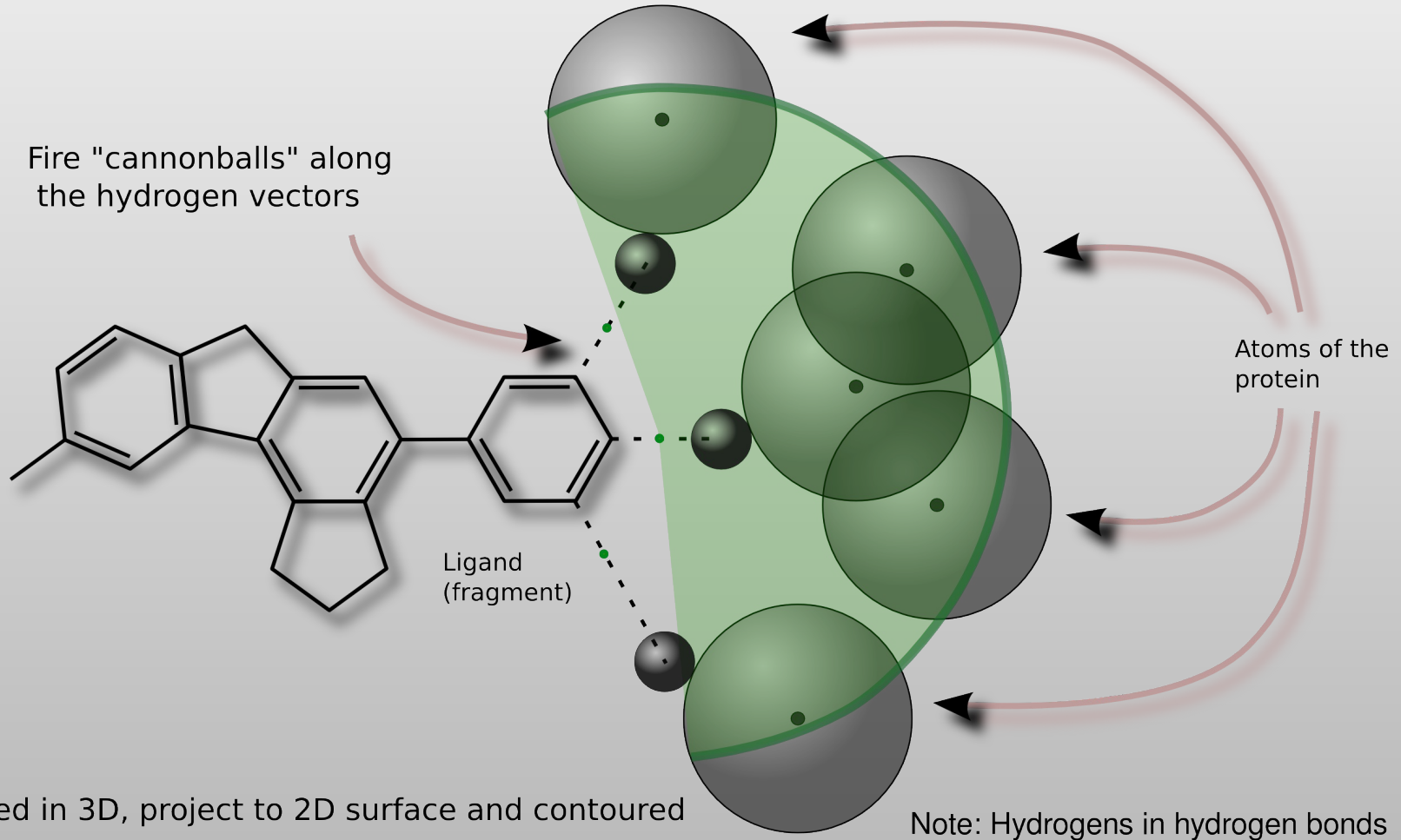
Ligand Environment Layout

- Residue position minimisation



Determination of the Substitution Contour

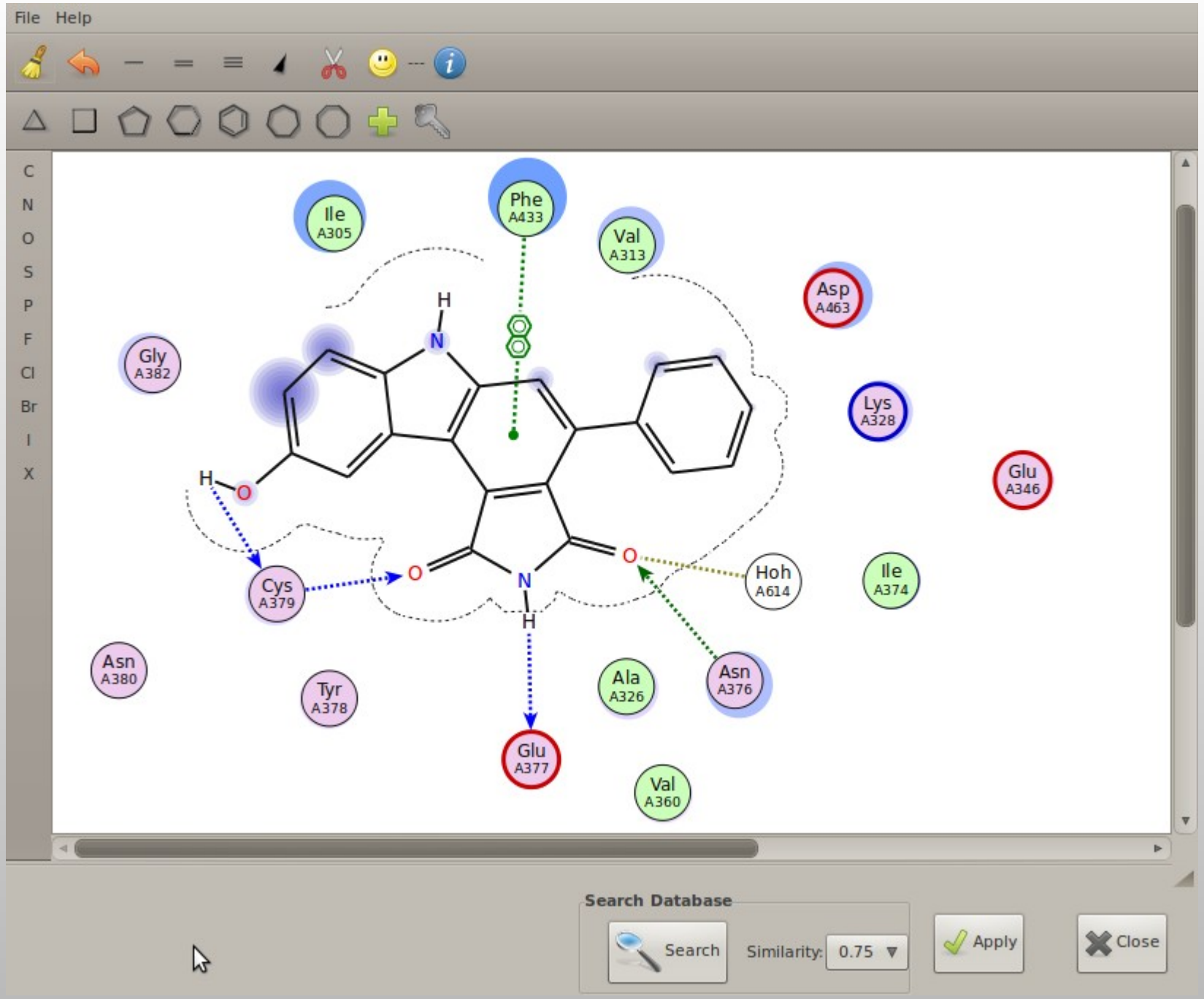
How far can we go (in the direction of the hydrogens) before hitting atoms of the protein?

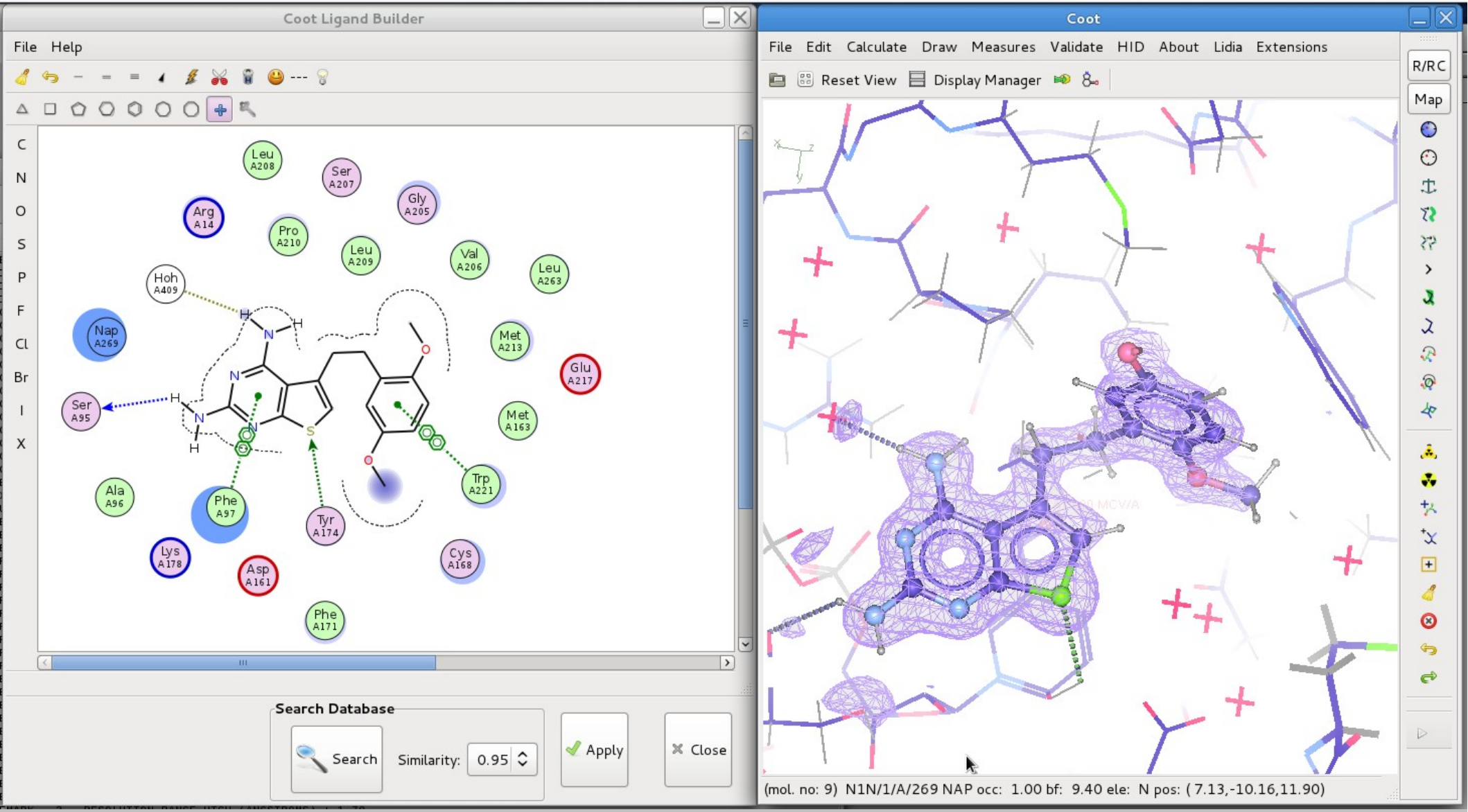


Determined in 3D, project to 2D surface and contoured

c.f. Clarke & Labute (2007)

Note: Hydrogens in hydrogen bonds are a confounding factor


File Help
C
N
O
S
P
F
Cl
Br
I
X
Ile A305
Phe A433
Val A313
Asp A463
Gly A382
Lys A328
Glu A346
Cys A379
Hoh A614
Ile A374
Asn A380
Tyr A378
Ala A326
Asn A376
Glu A377
Val A360
Search Database
Search
Similarity: 0.75
Apply
Close


Coot Ligand Builder
File Help
Leu A208
Ser A207
Gly A205
Arg A14
Pro A210
Leu A209
Val A206
Leu A263
Hoh A409
Nap A269
Met A213
Glu A217
Ser A95
Met A163
Trp A221
Ala A96
Phe A97
Tyr A174
Lys A178
Asp A161
Cys A168
Phe A171
Search Database
Search
Similarity: 0.95
Apply
Close
Coot
File Edit Calculate Draw Measures Validate HID About Lidia Extensions
Reset View
Display Manager
R/RC
Map
(mol. no: 9) N1N/1/A/269 NAP occ: 1.00 bf: 9.40 ele: N pos: (7.13,-10.16,11.90)

Acknowledgements

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

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

Modelling Carbohydrates

- Validation,
- Model-building,
- Refinement

Problematic Glycoproteins

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

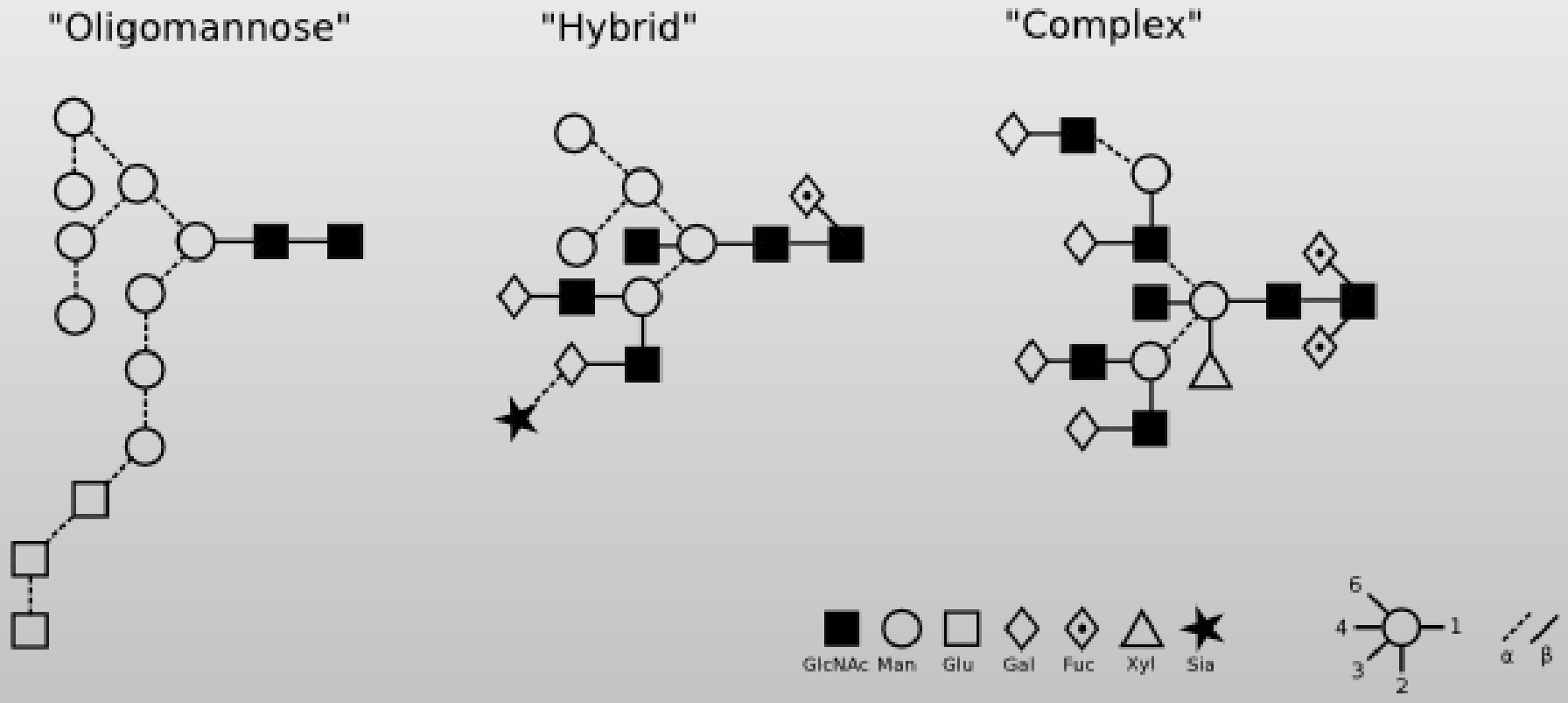
Modelling Carbohydrates

- Validation,
- Model-building,
- Refinement

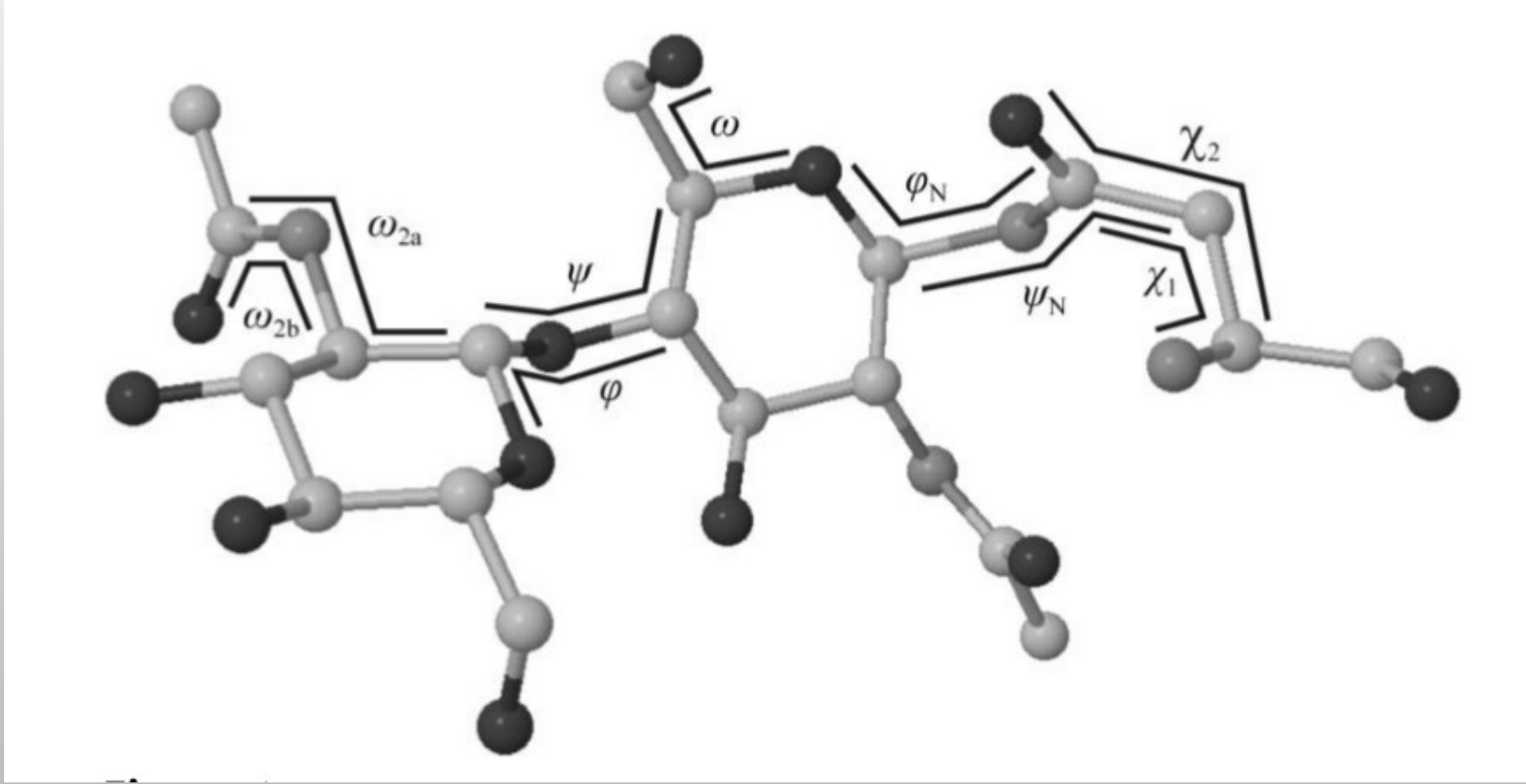
Problematic Glycoproteins

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

Validate the Tree: N-linked carbohydrates



Carbohydrate Links

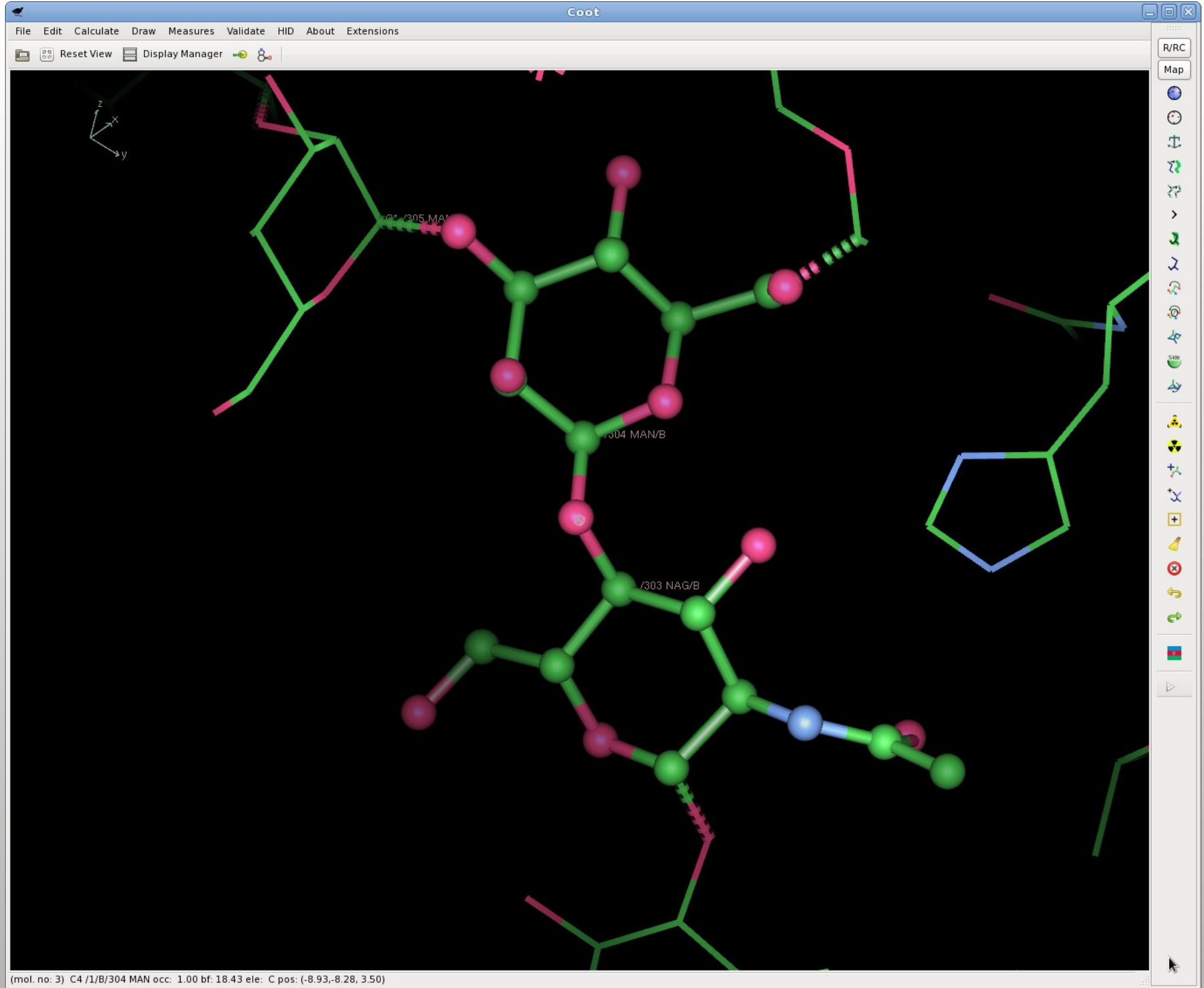


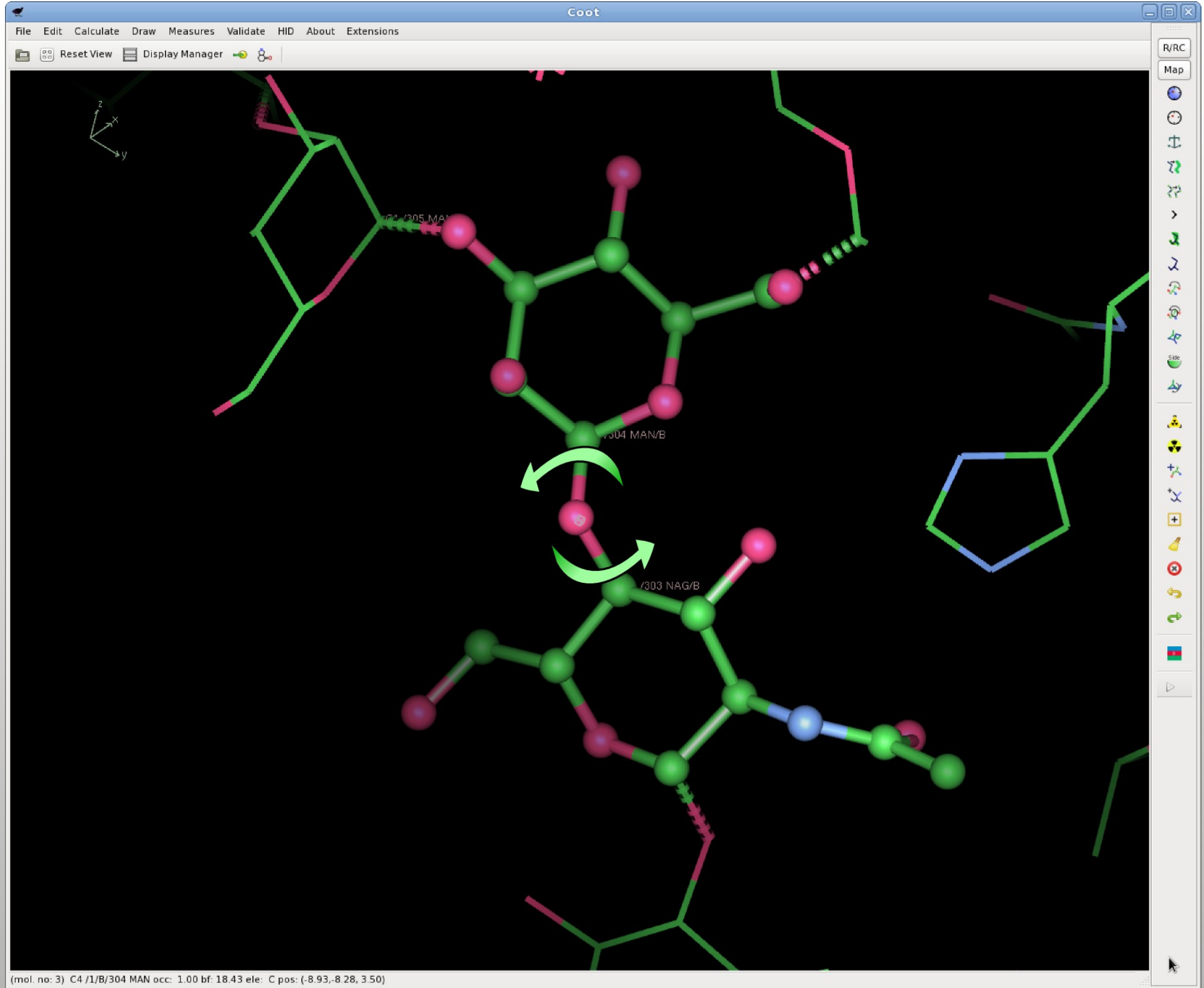
Thomas Lütteke (2007)

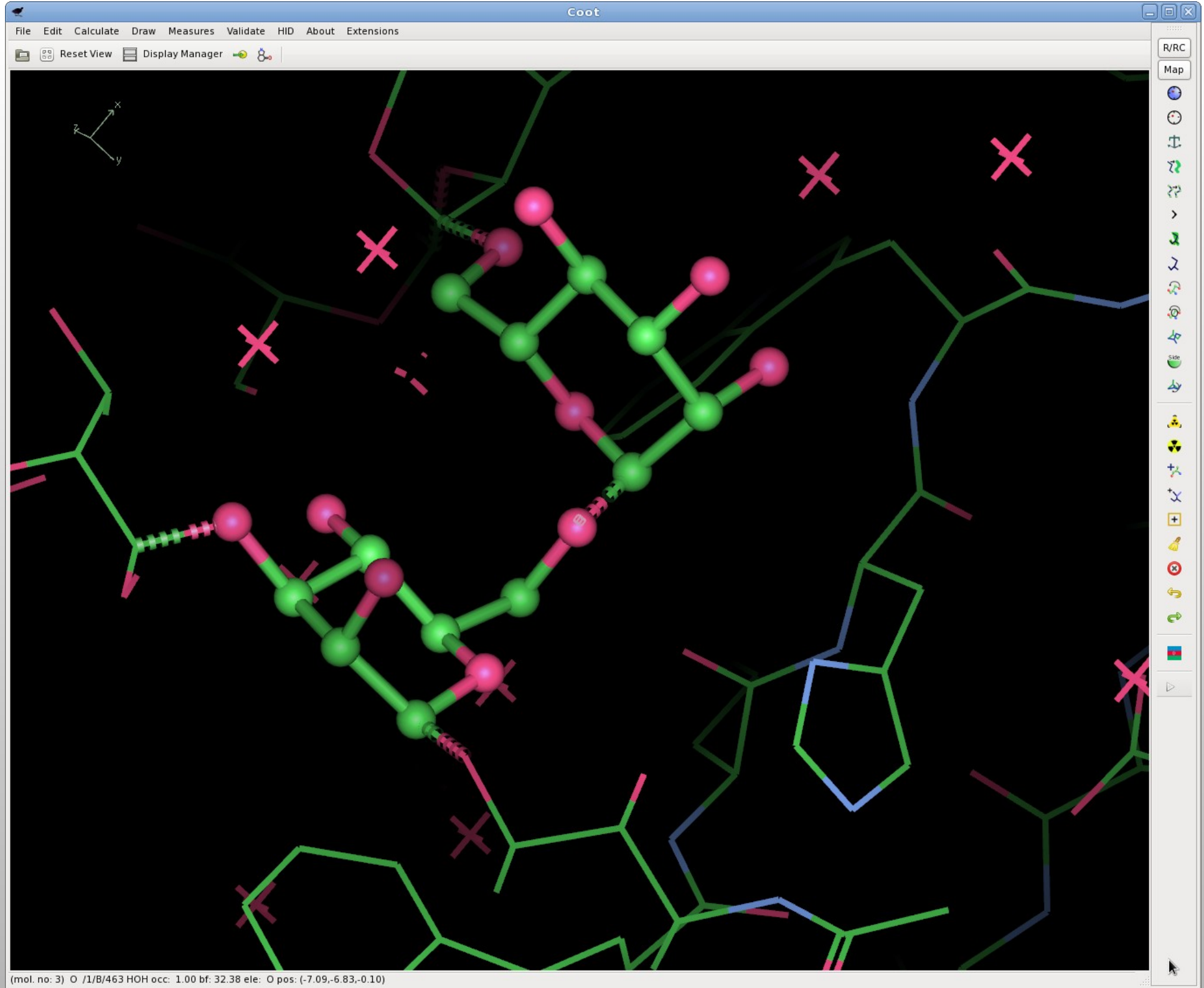
Linking Oligsaccharides/Carbohydrates:

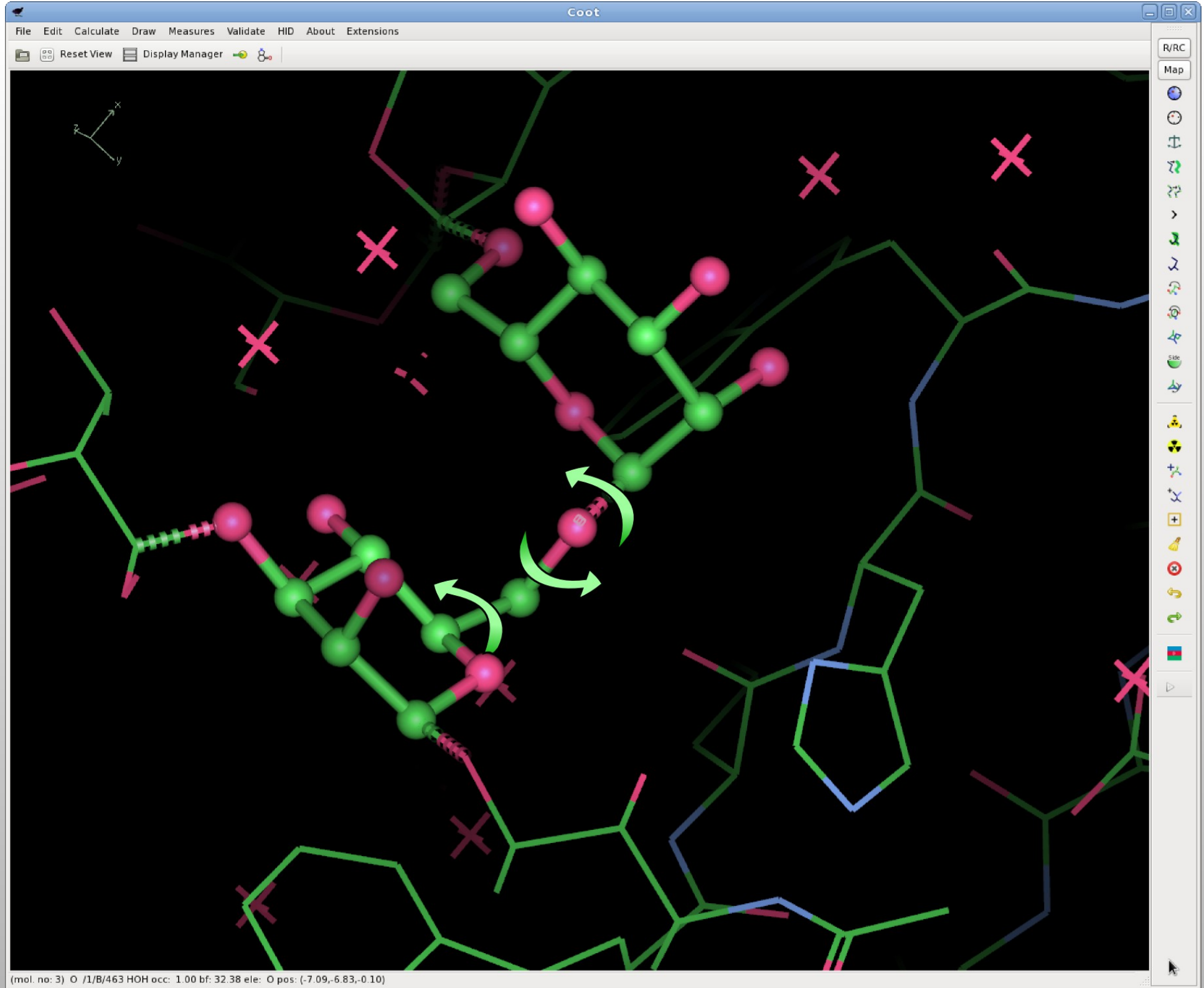
LO/Carb

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

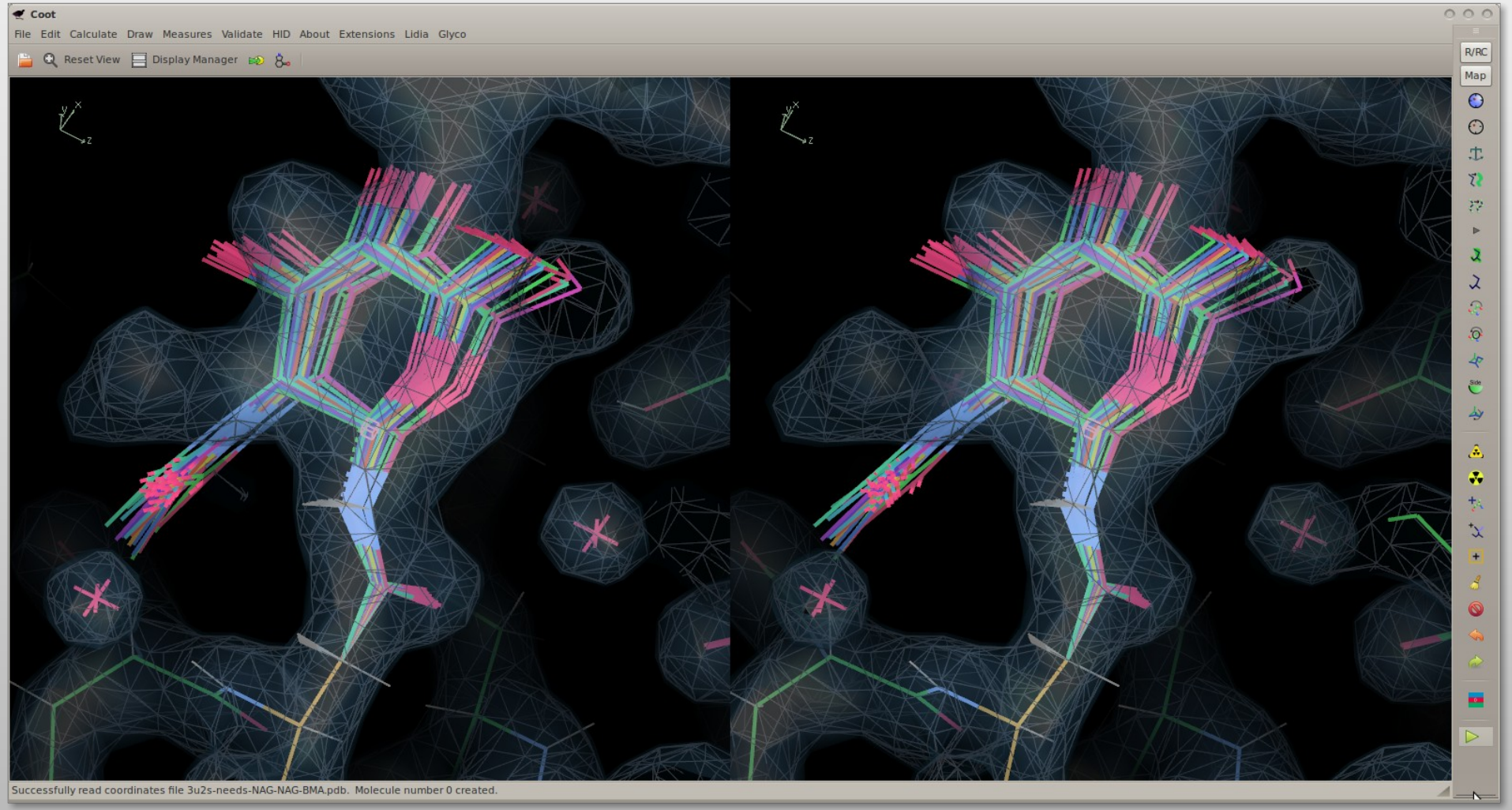

Coot
File
Edit
Calculate
Draw
Measures
Validate
HID
About
Extensions
Reset View
Display Manager
R/RC
Map
/304 MAN/B
/303 NAG/B
(mol. no: 3) C4 /1/B/304 MAN occ: 1.00 bf: 18.43 ele: C pos: (-8.93,-8.28, 3.50)


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


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


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

Refinement Trials (NAG-ASN example)



Acknowledgements

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

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