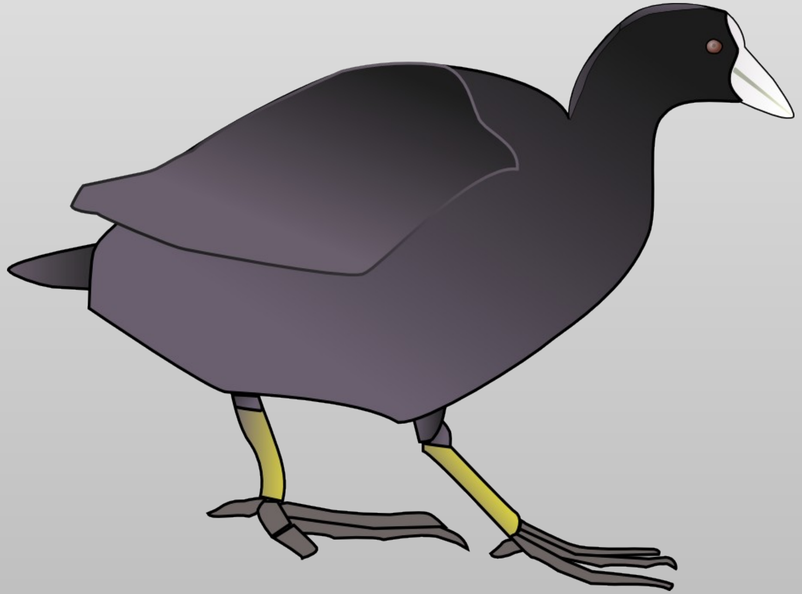
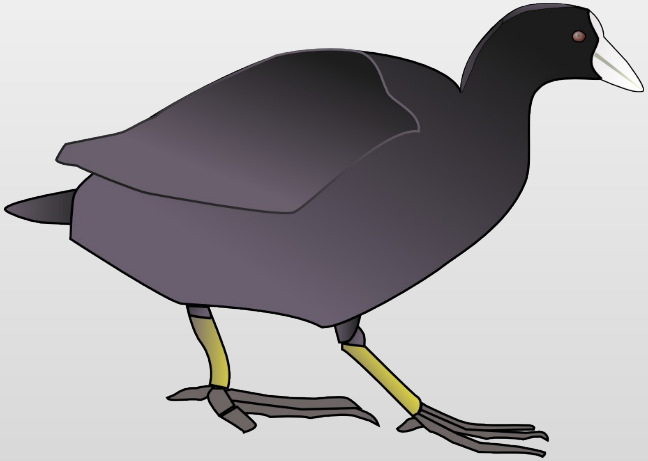
Model-Building of Proteins Using X-ray Data With *Coot*



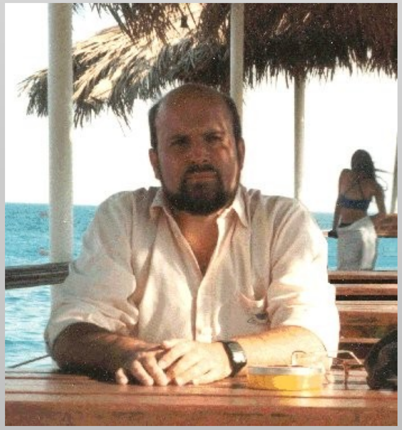
Paul Emsley
May 2013

Acknowldegments, Collaborators



Bernhard
Lohkamp

Kevin
Cowtan



Eugene
Krissinel

Stuart
McNicholas

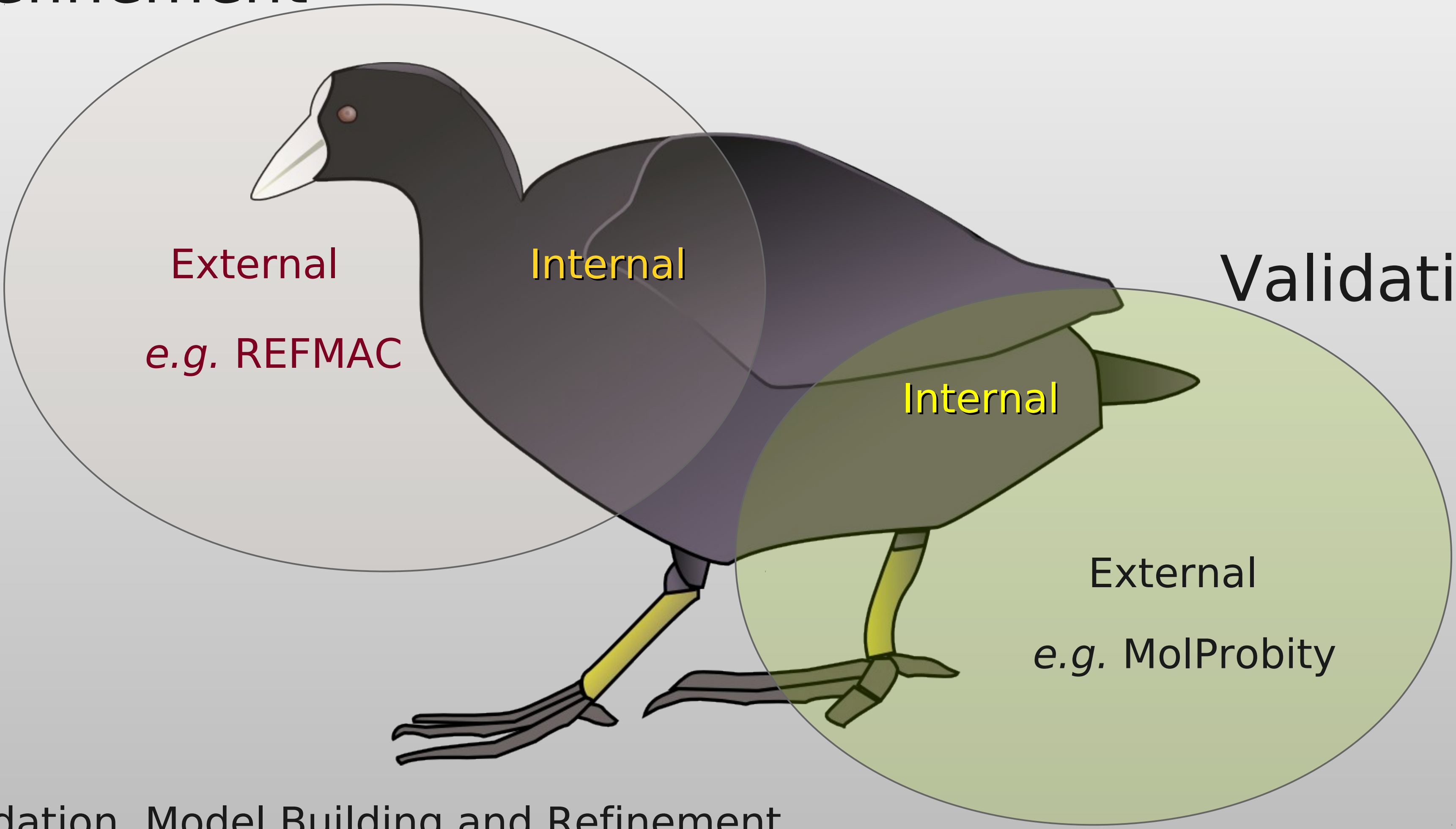
Martin
Noble

Alexei
Vagin

But Why..?

- In these days of automation...
    - why should I use something interactive?
    - why are you forcing me to think?
- It's the Nature of the Problem
    - partial/refined model to deposition quality
        - Problematic due to the variety of ways in which a model can be wrong
        - Hypothesis generation and evaluation
- It takes a brain to validate and correct
    - (for at least the next 5 years...)

Feature Integration



Refinement

External

e.g. REFMAC

Internal

Validation

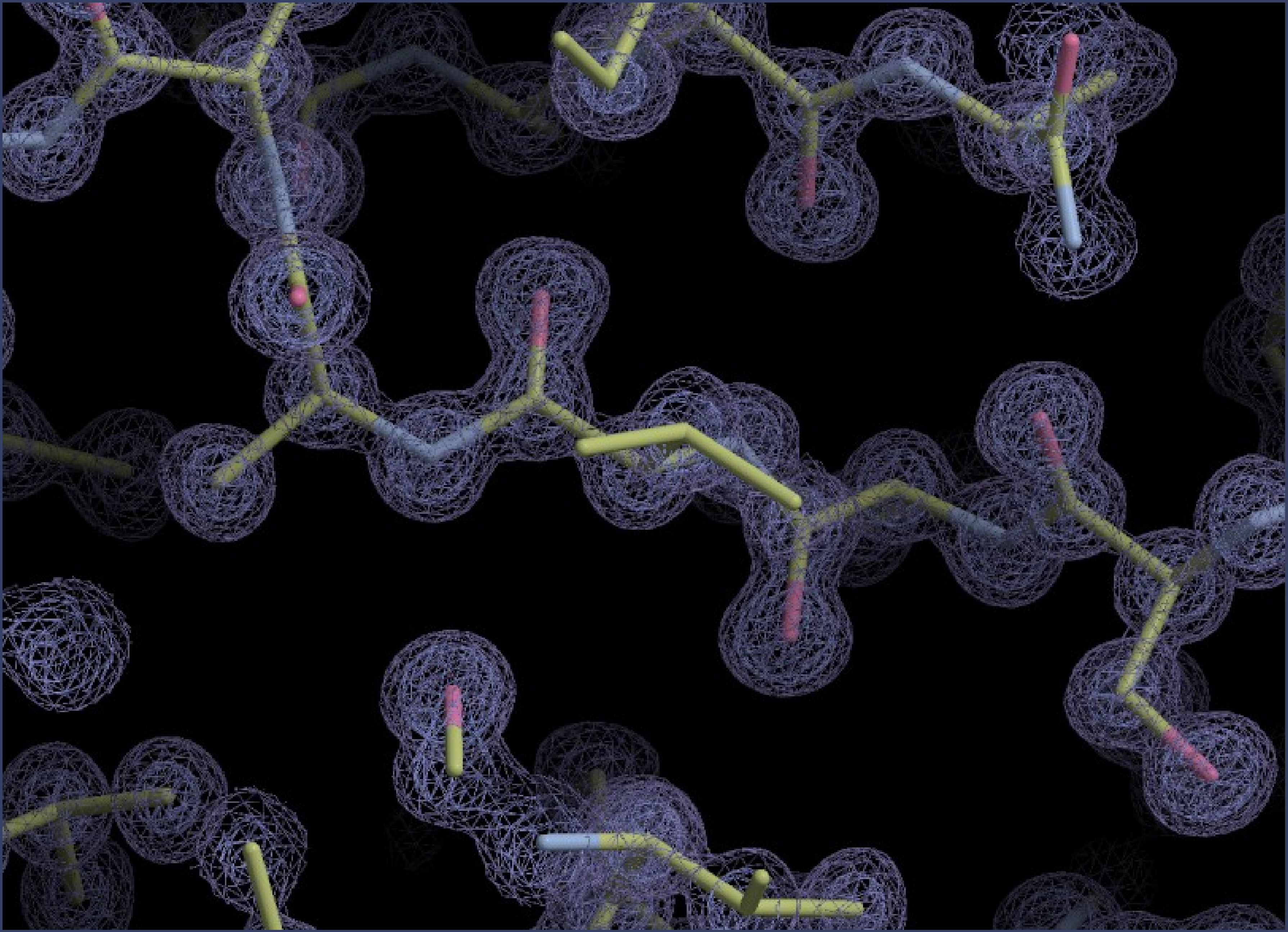
Internal

External

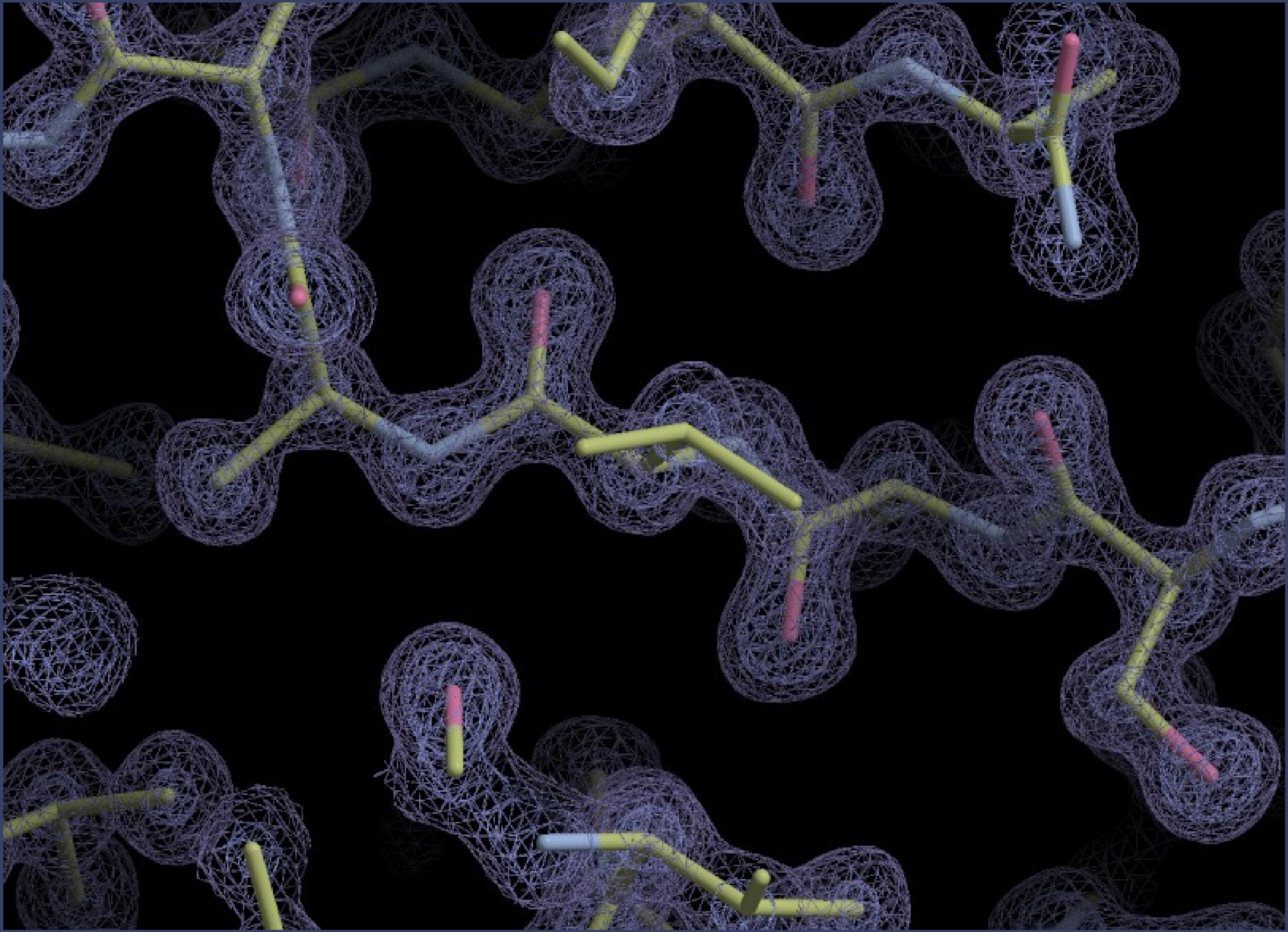
e.g. MolProbity

Validation, Model Building and Refinement should be used together

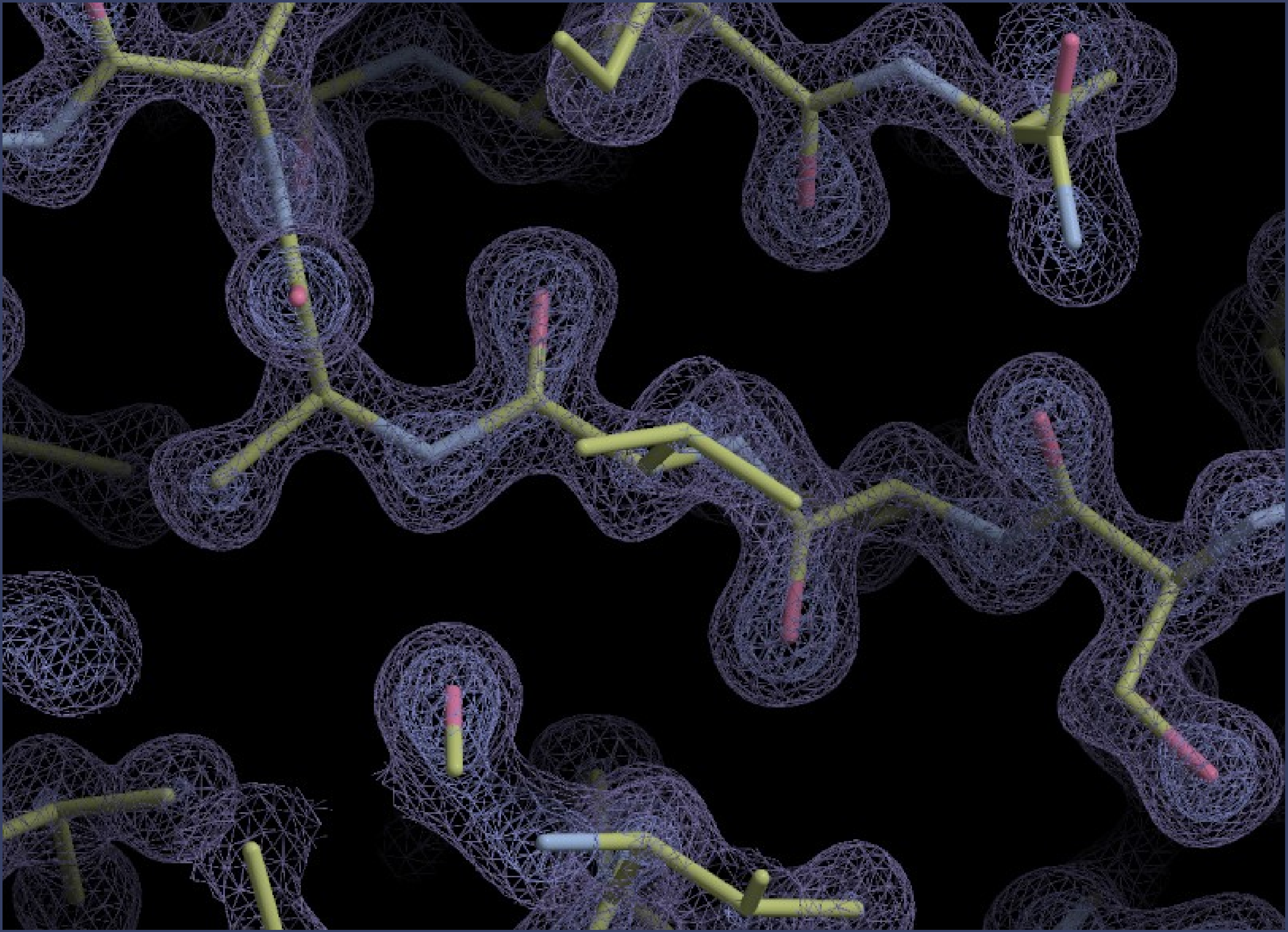
1.0Å



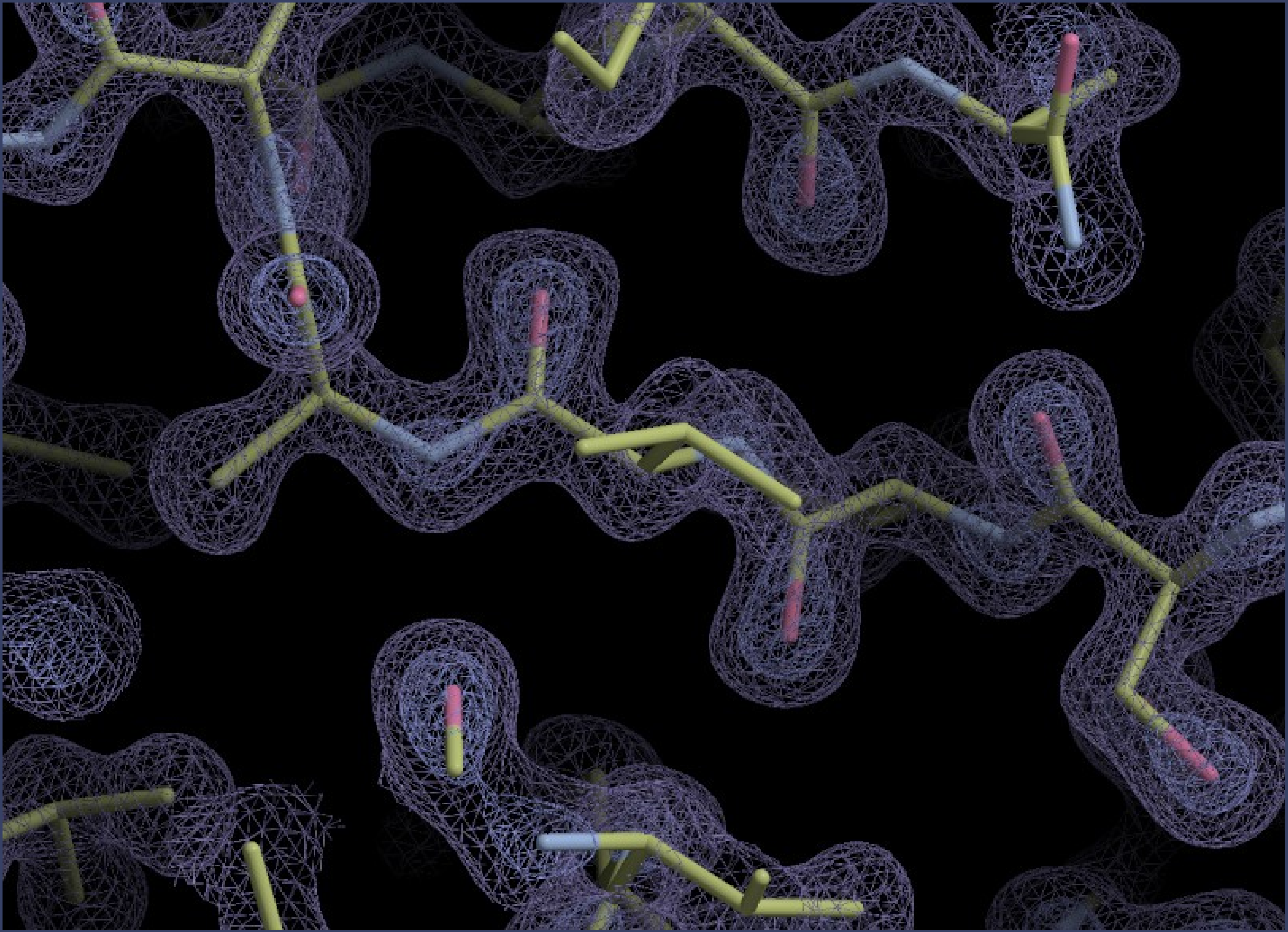
1.2Å



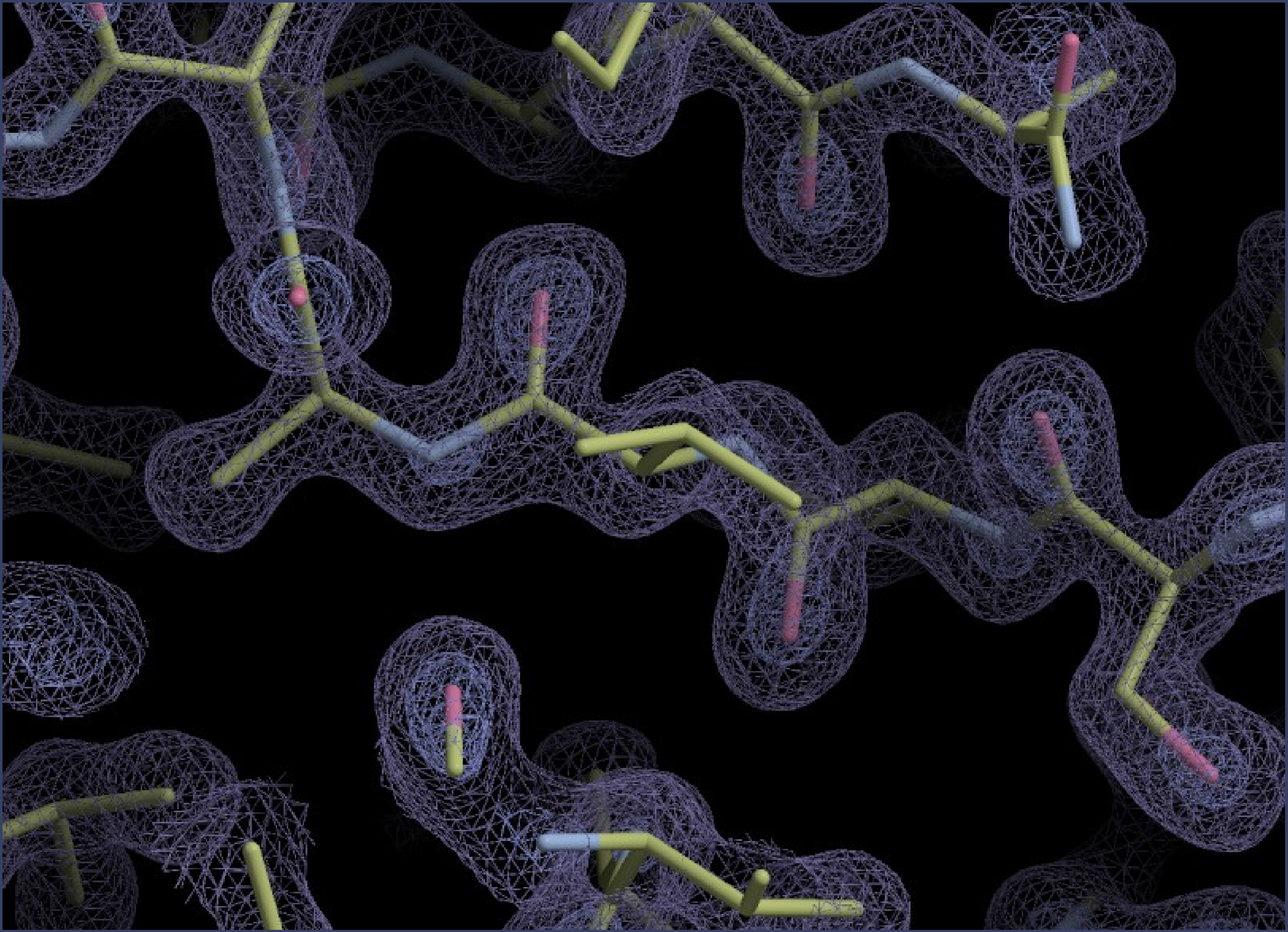
1.4Å



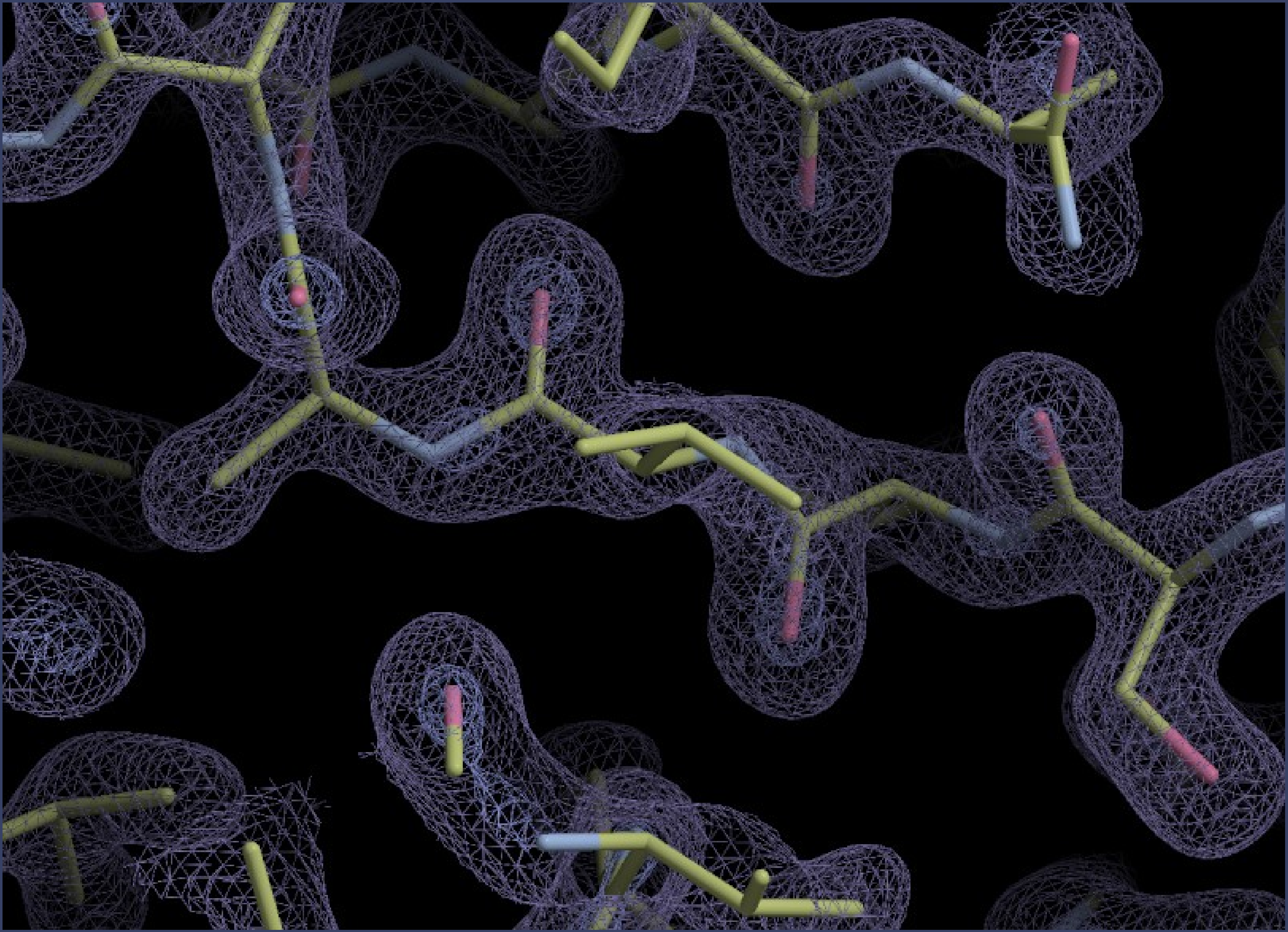
1.6Å



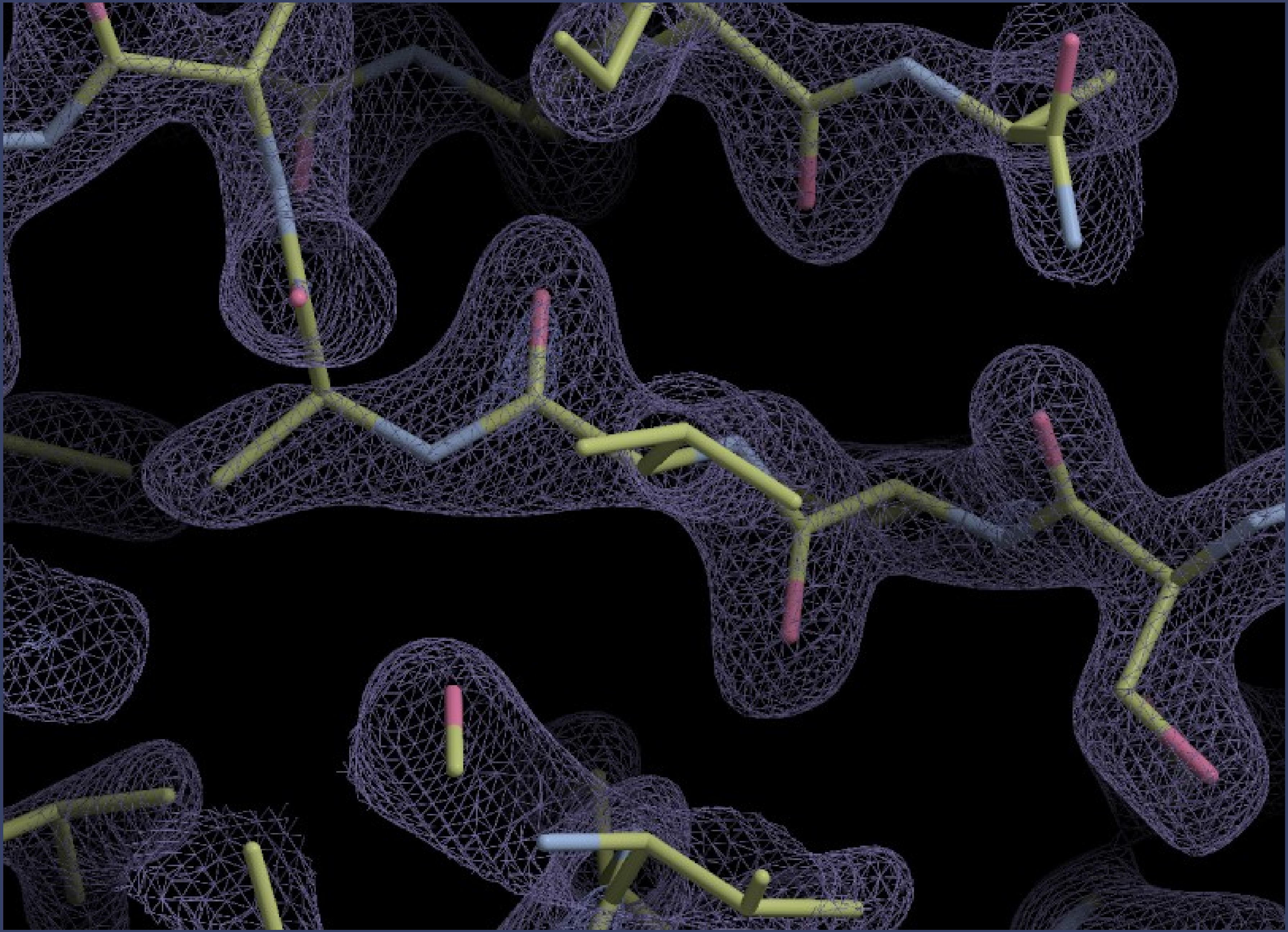
1.8Å



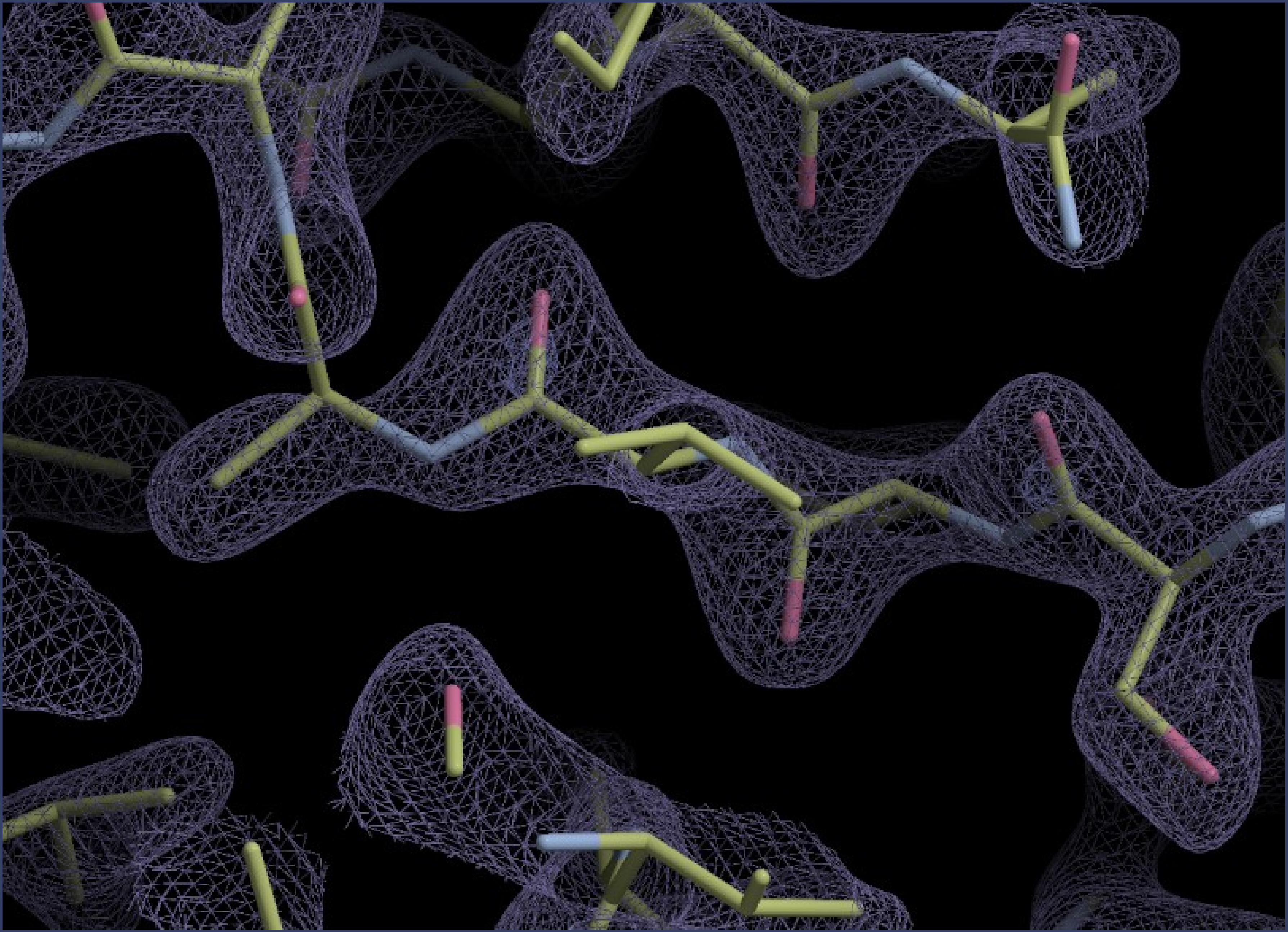
2.0Å



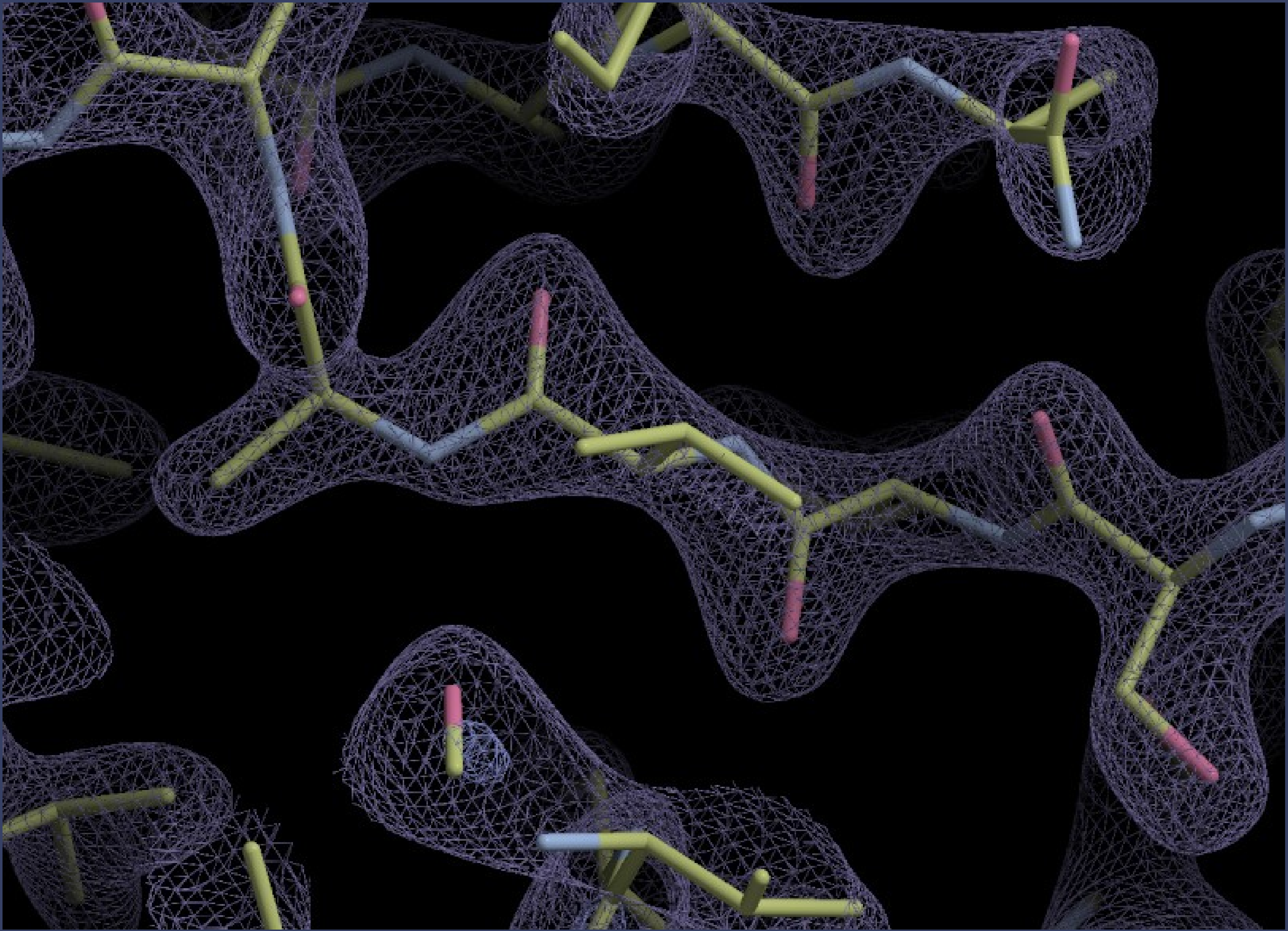
2.2Å



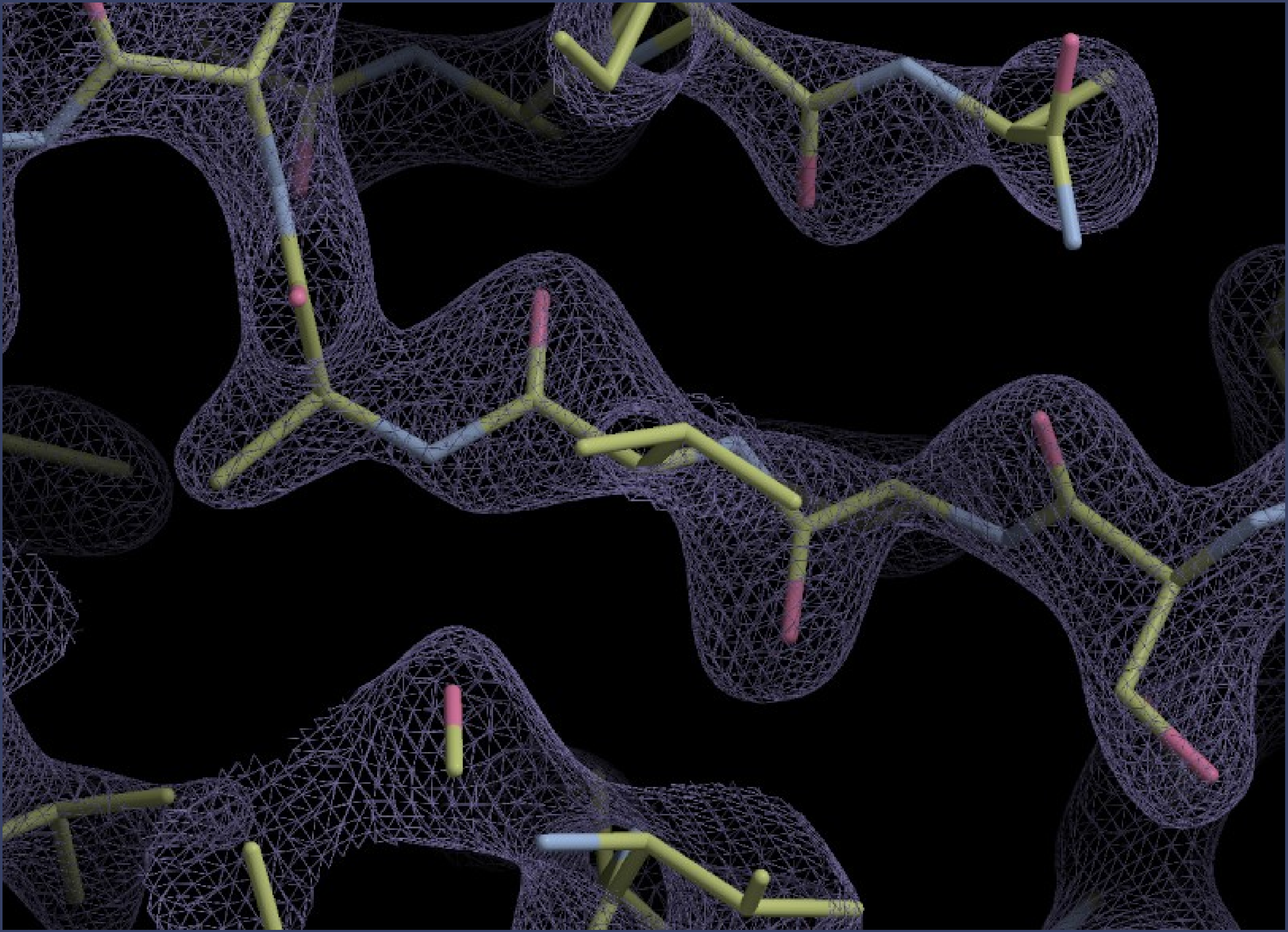
2.4Å



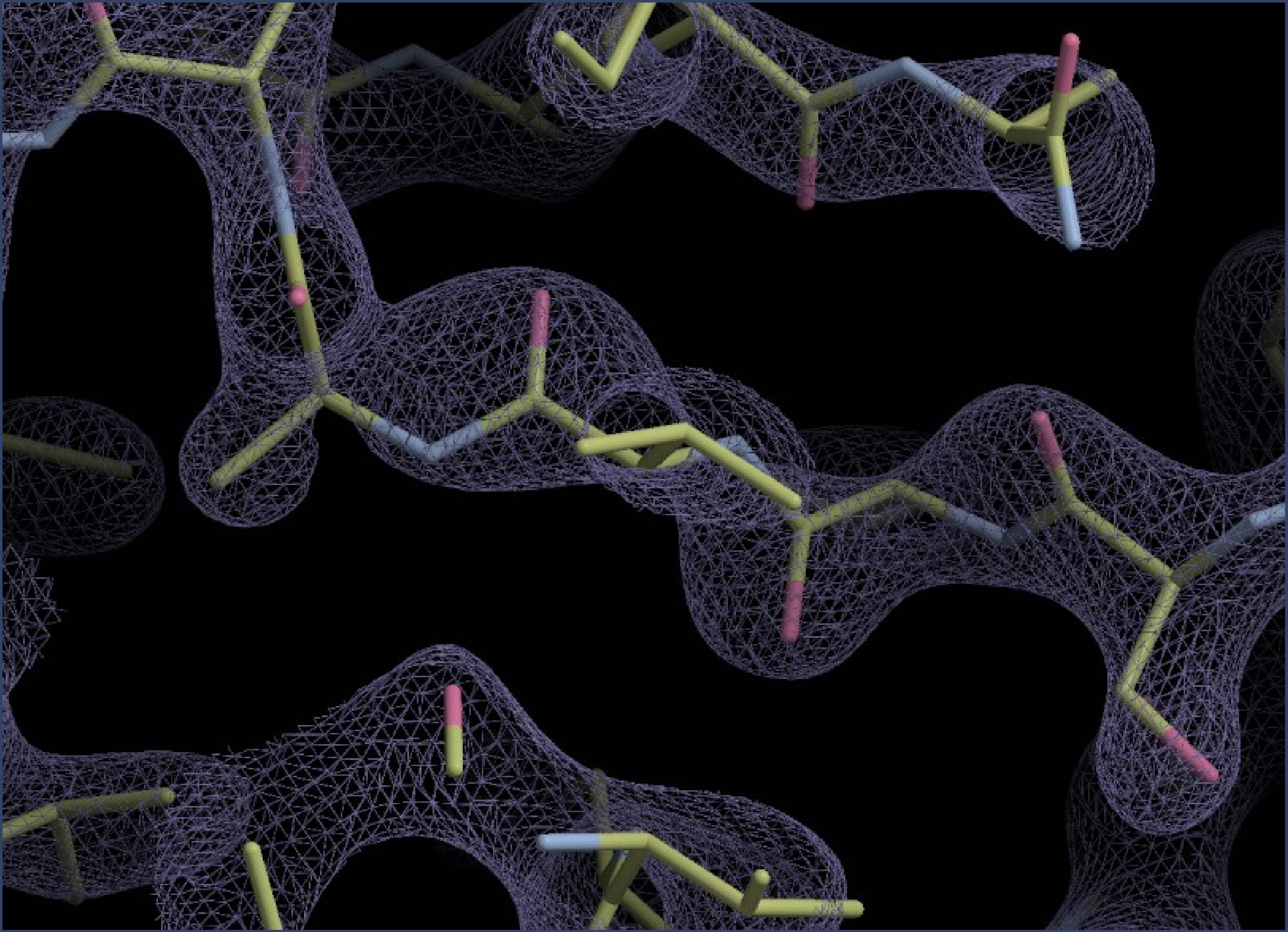
2.6Å



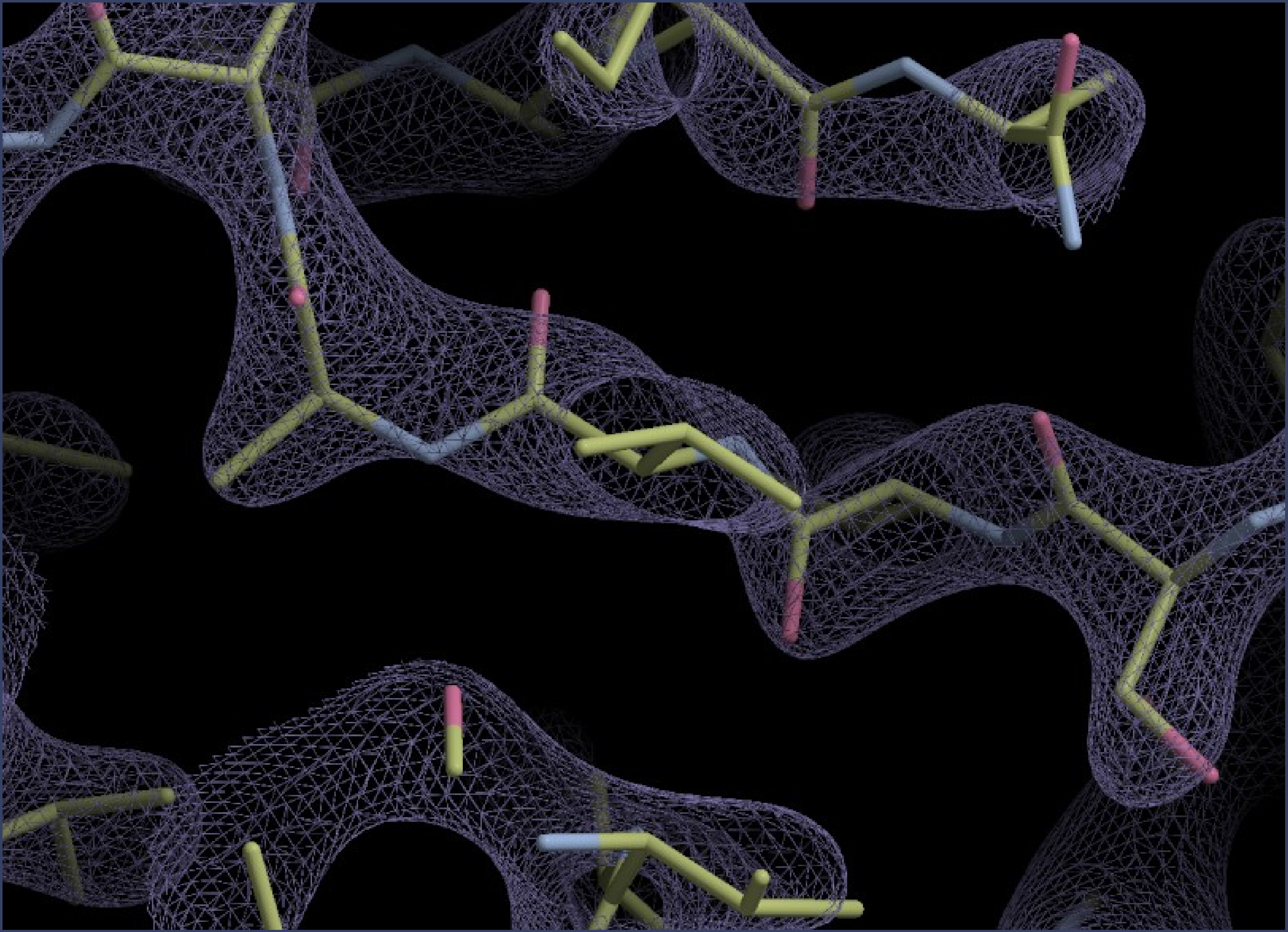
2.8Å



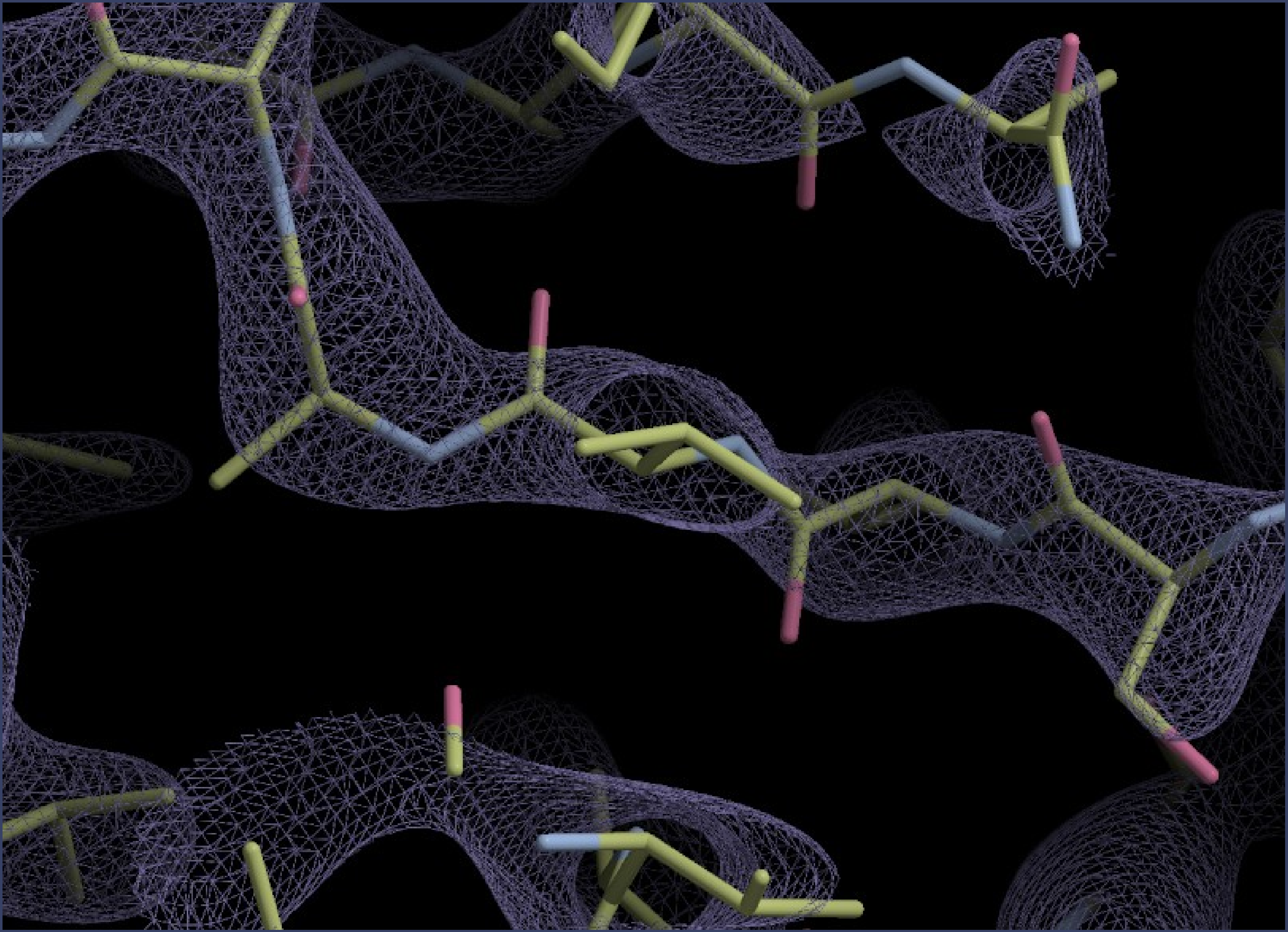
3.0Å



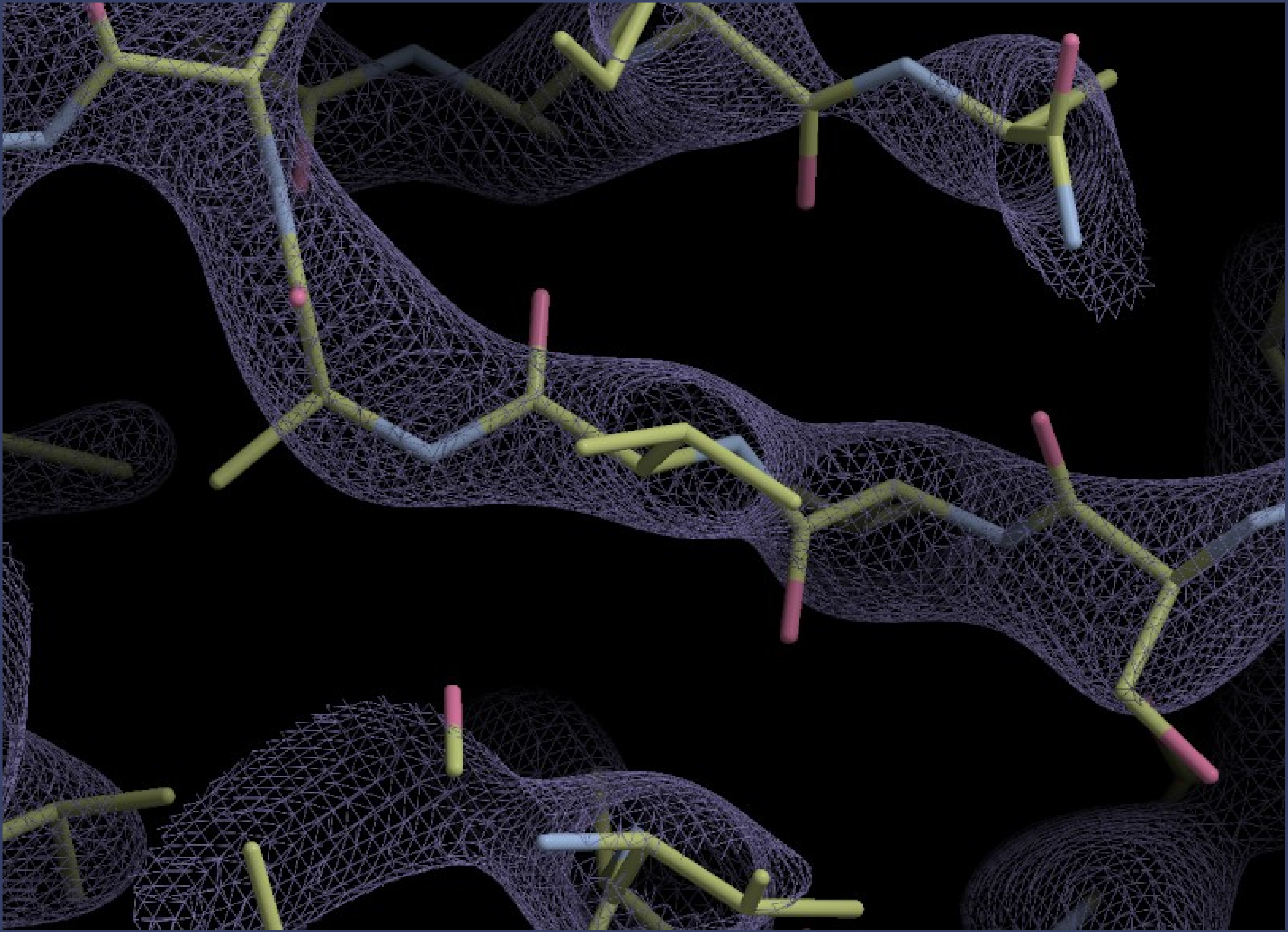
3.2Å



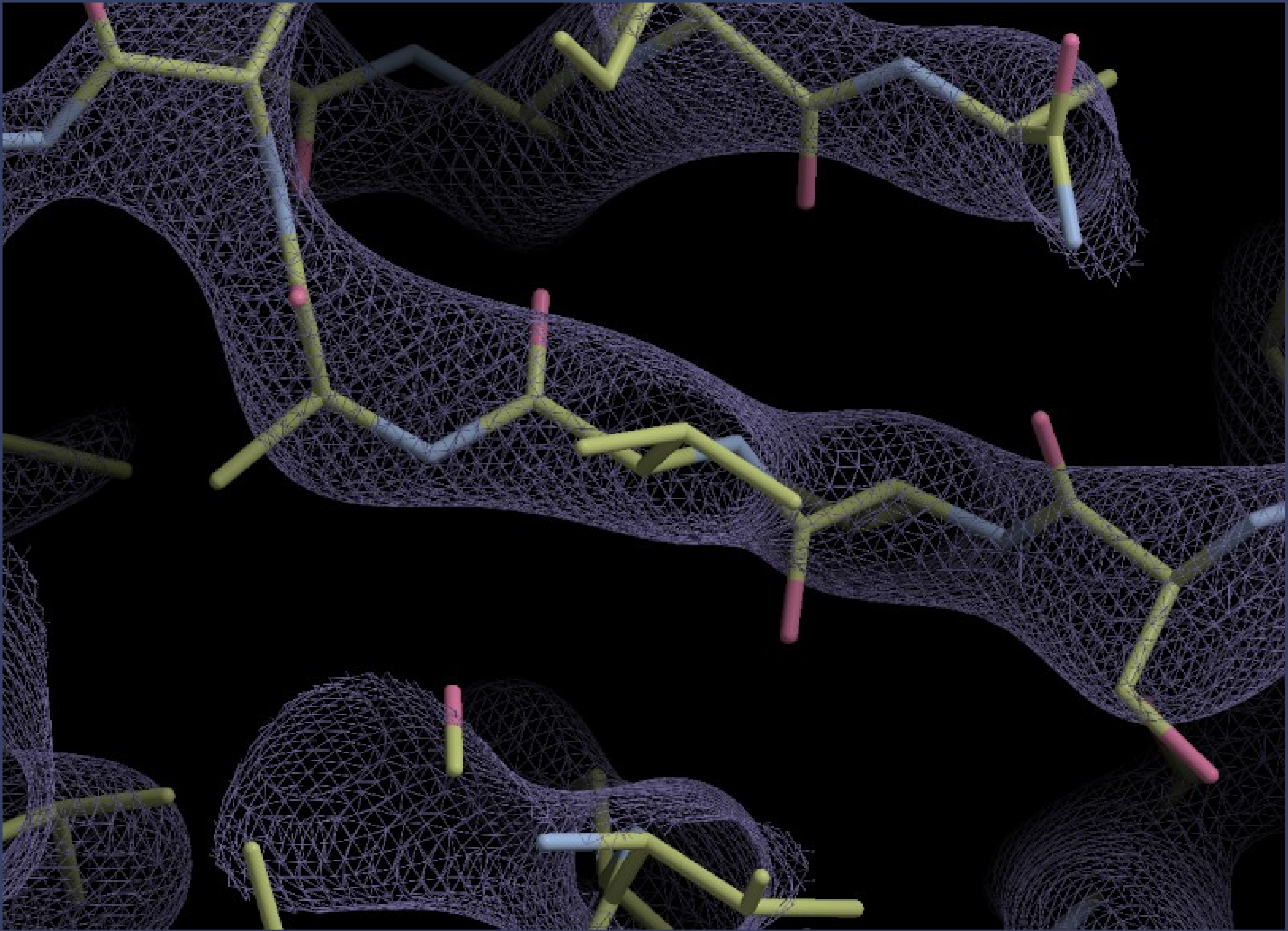
3.4Å



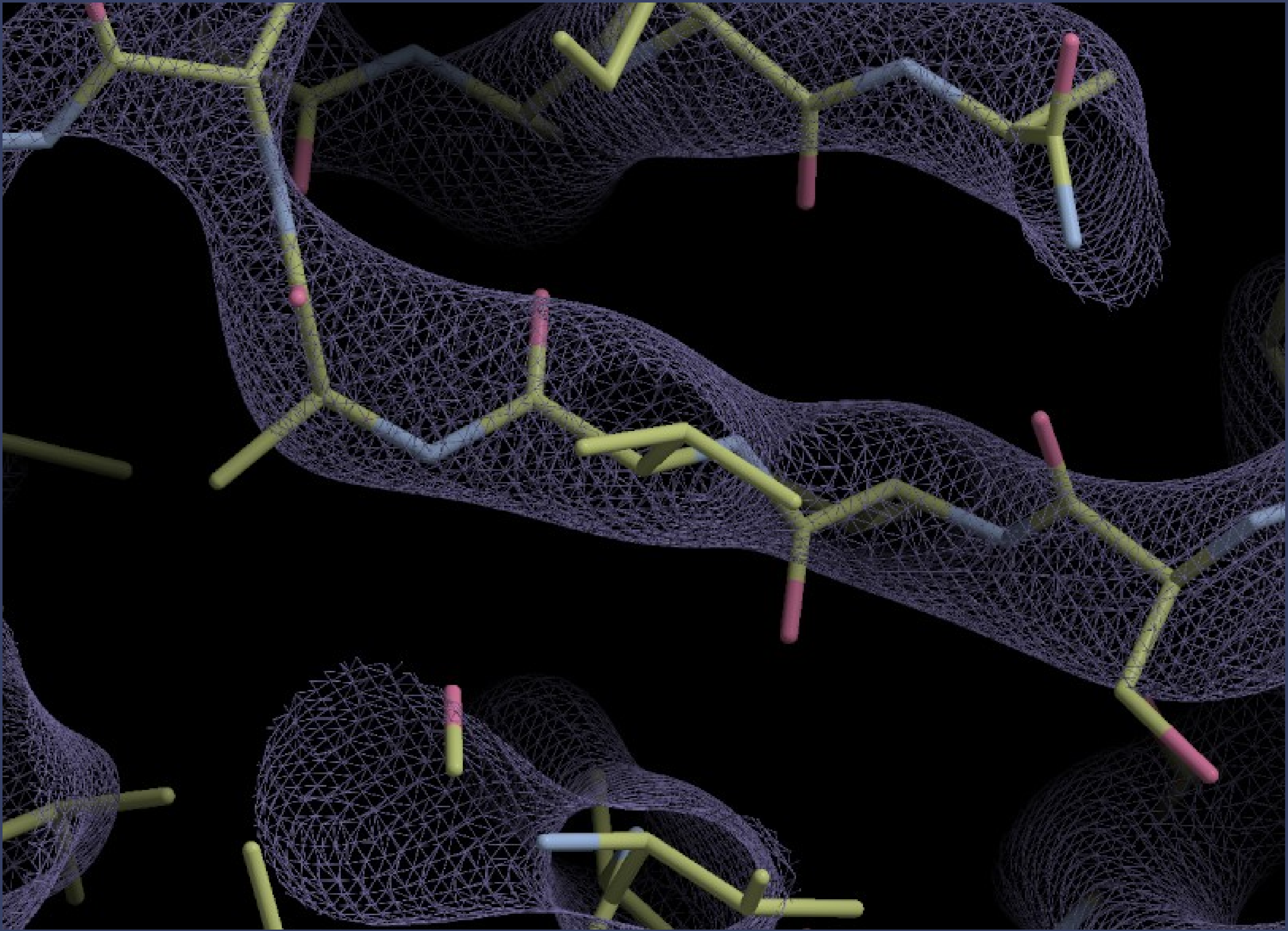
3.6Å



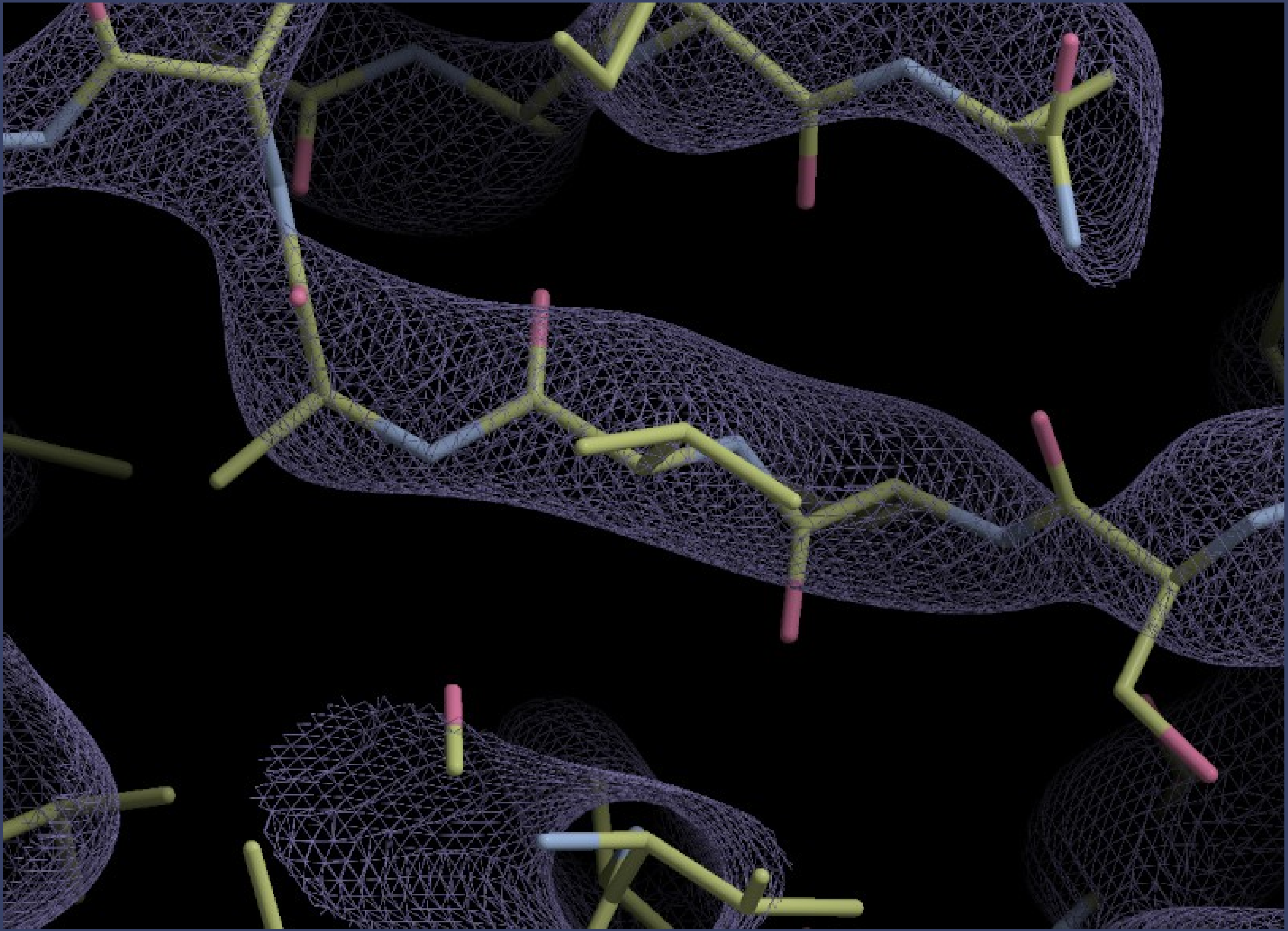
3.8Å



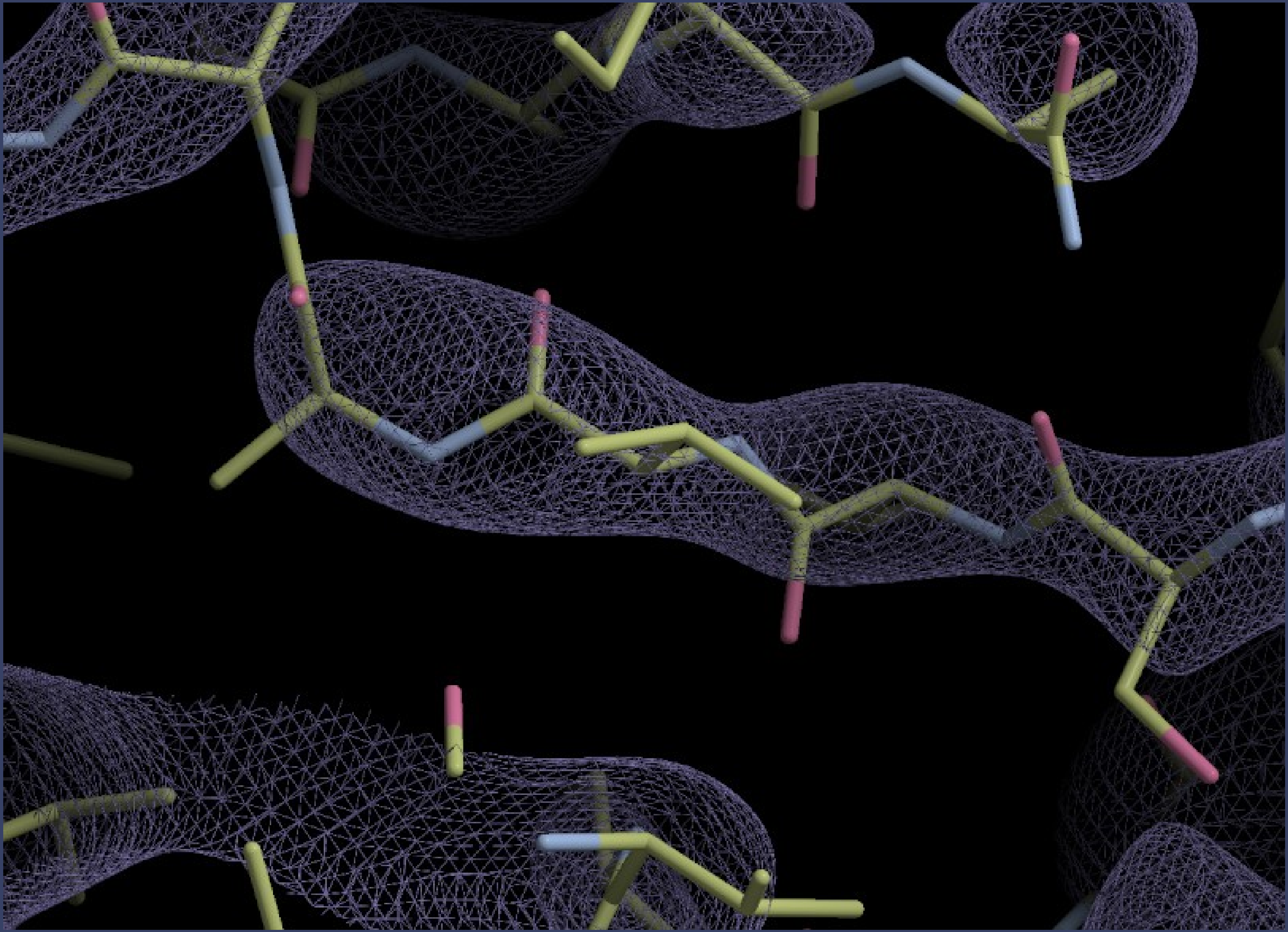
4.0Å



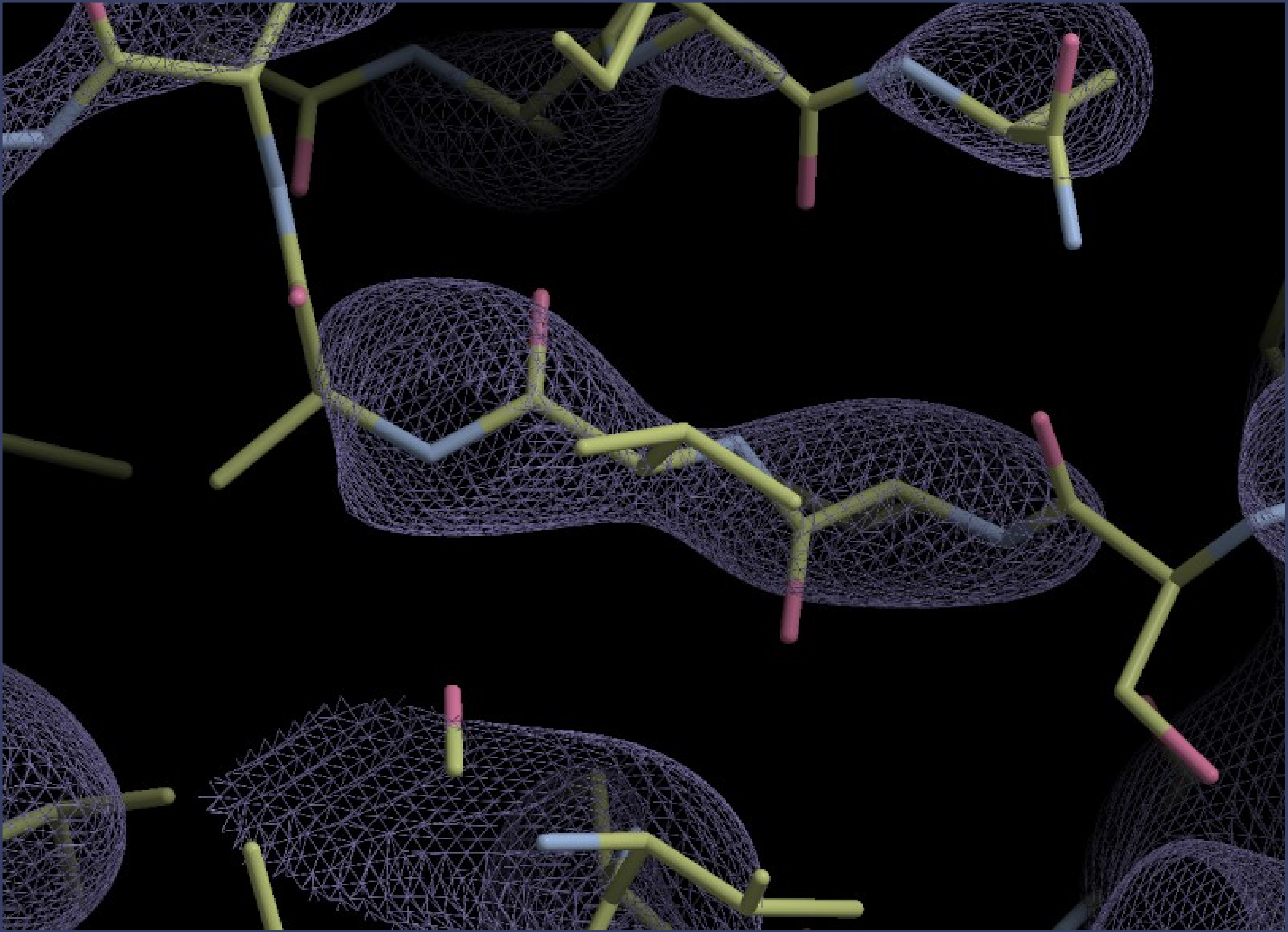
4.2Å



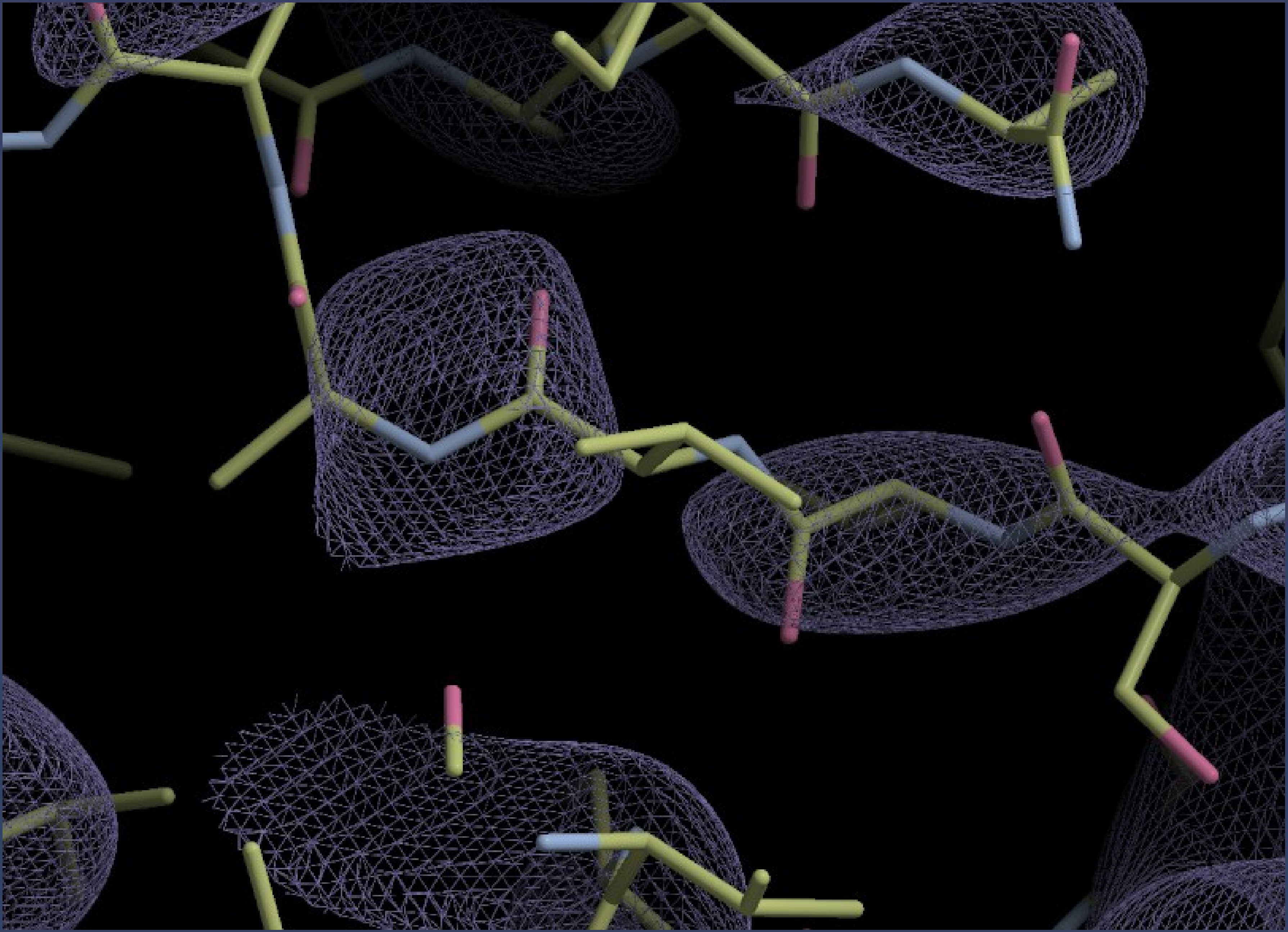
4.4Å



4.6Å



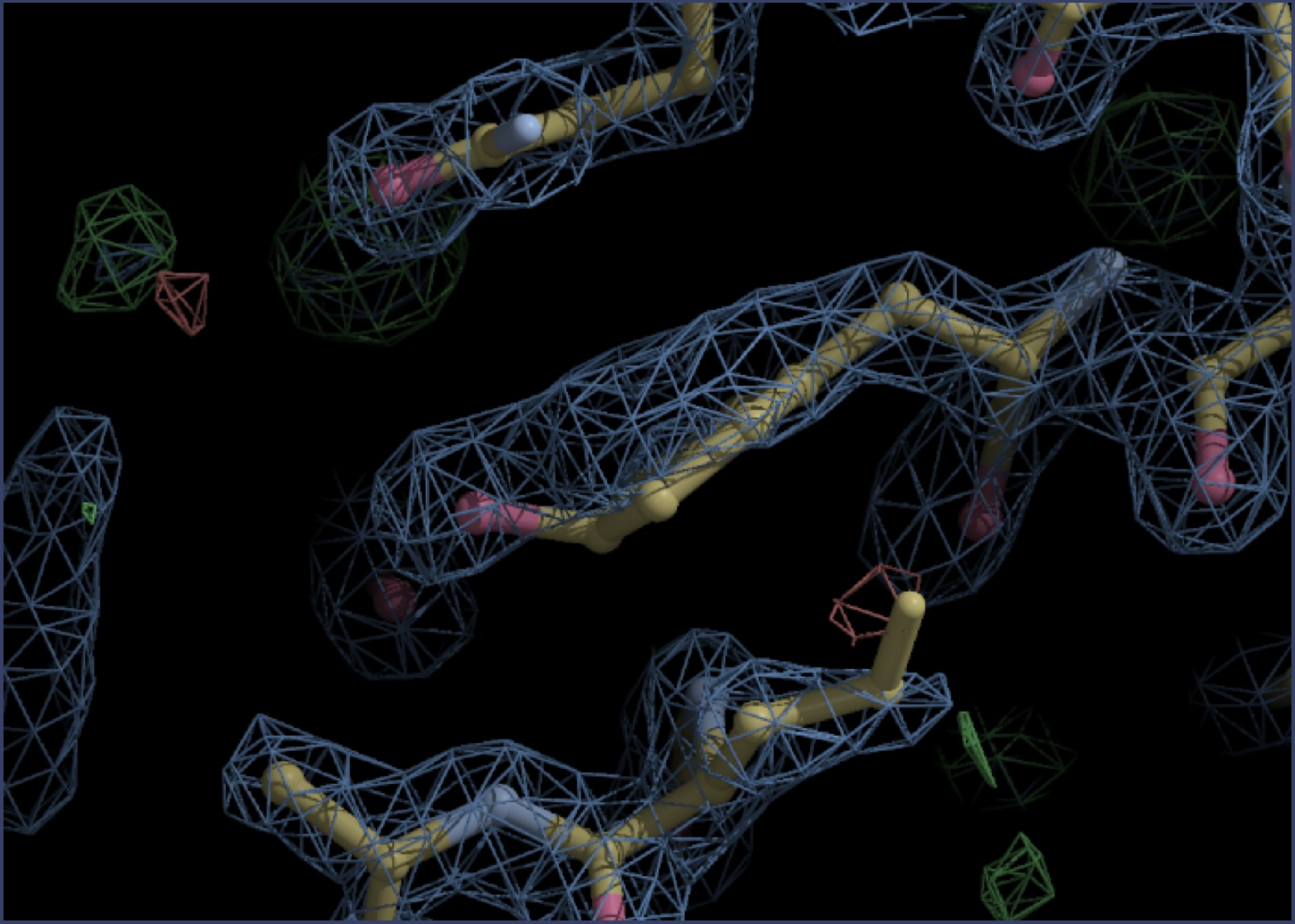
4.8Å



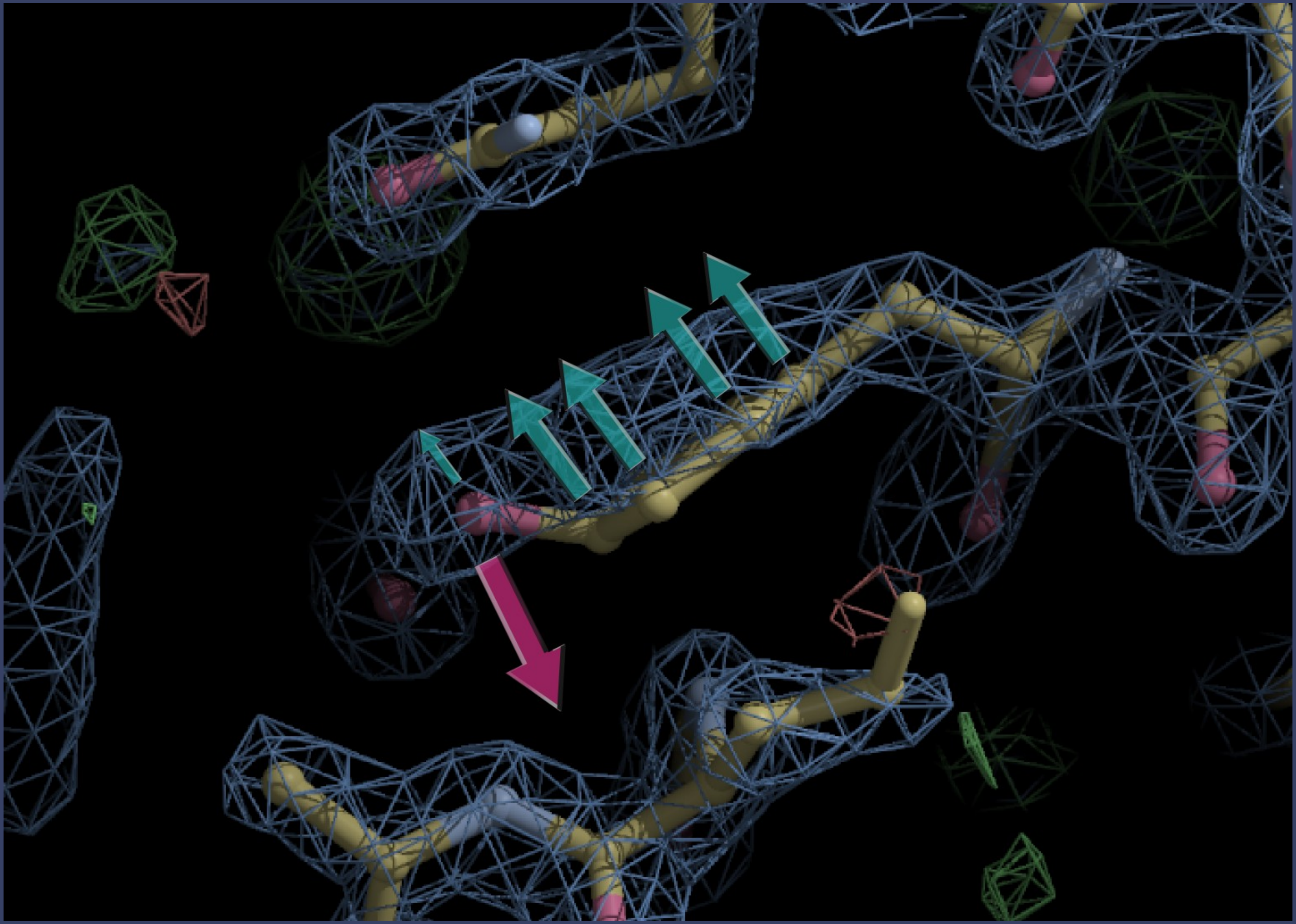
What is “Refinement”?

- 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

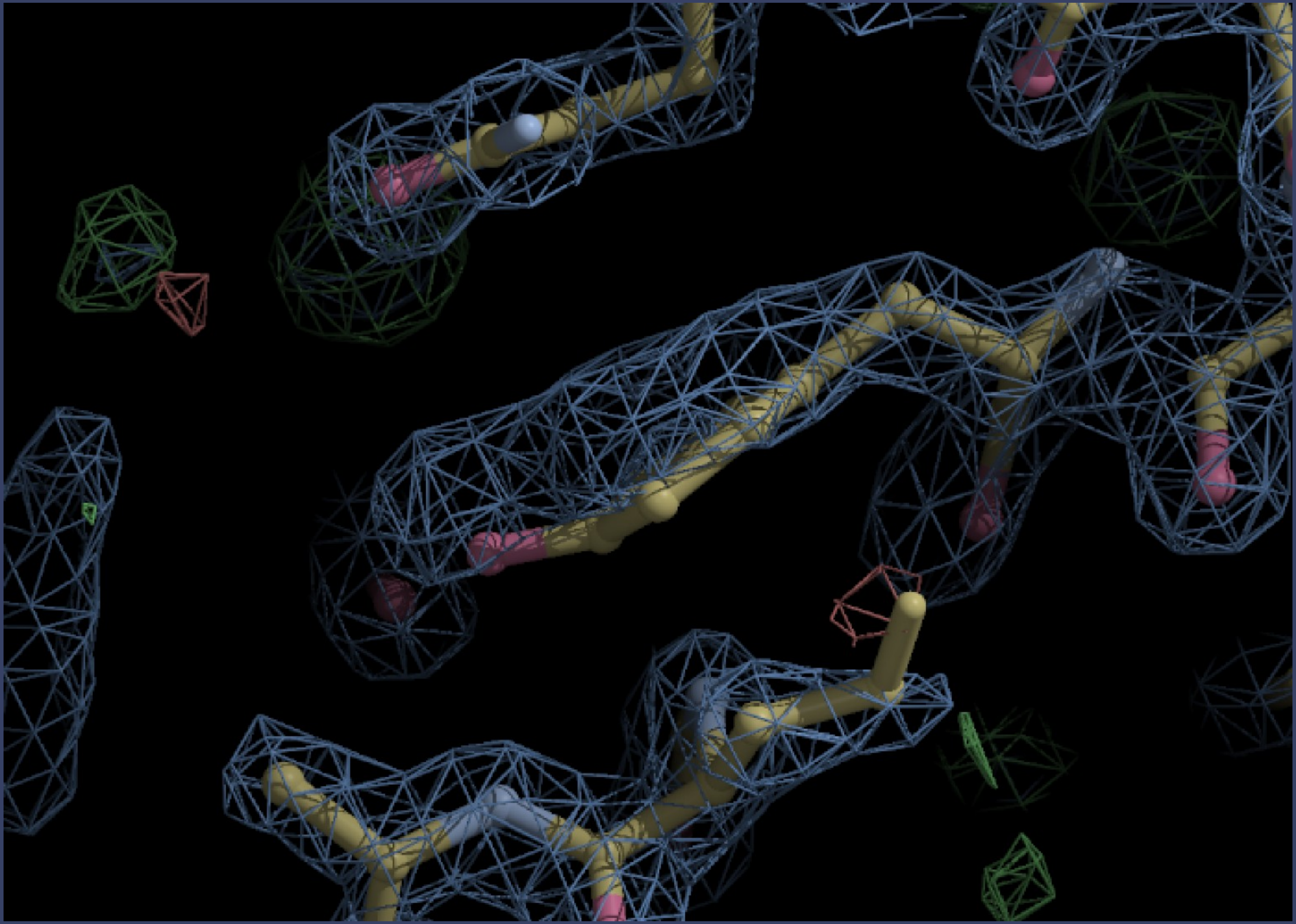
Distorted Geometry Pre-Refinement



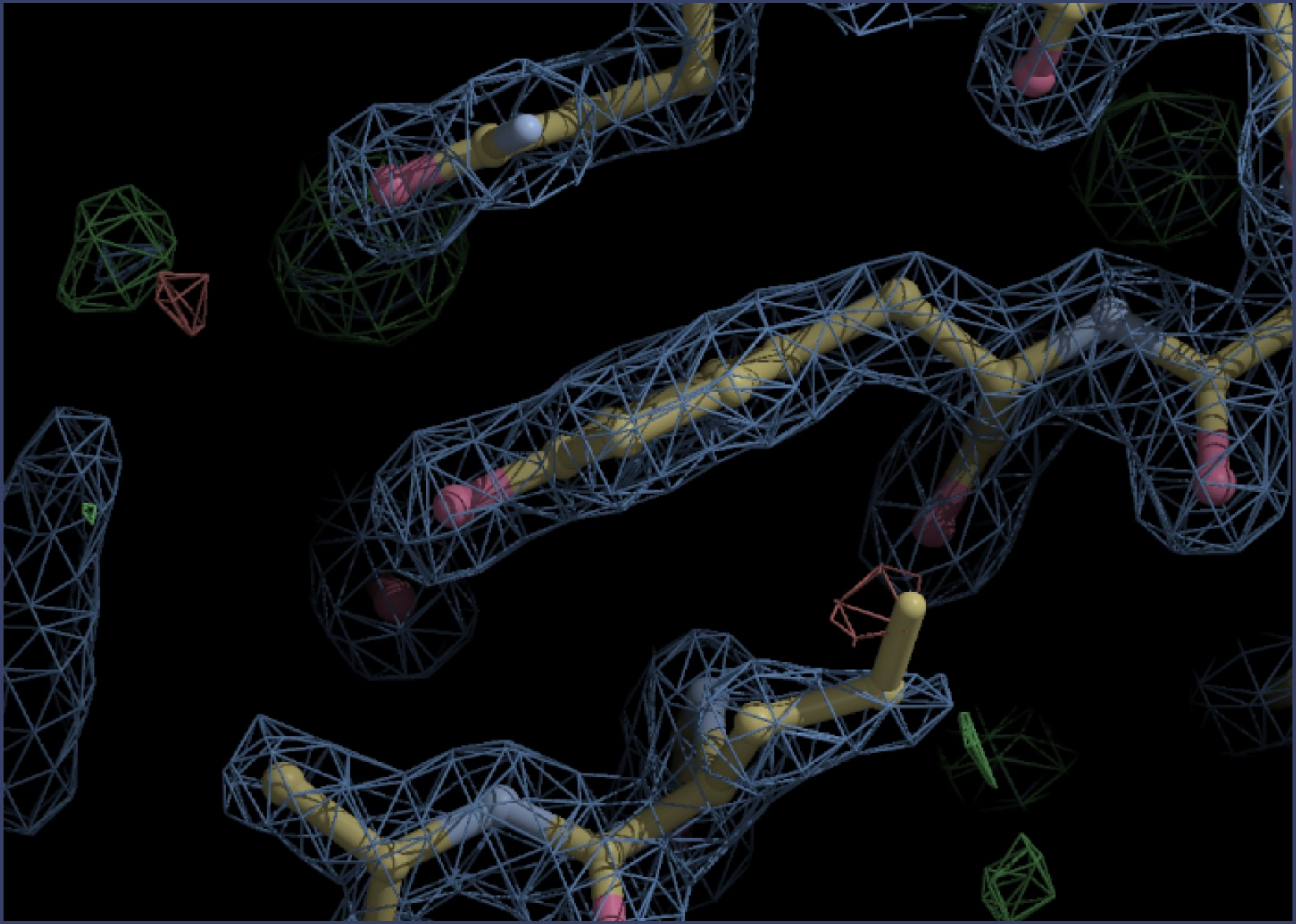
Refinement Gradients



Refinement: Cycle 3



Refinement Cycle 200: Minimized

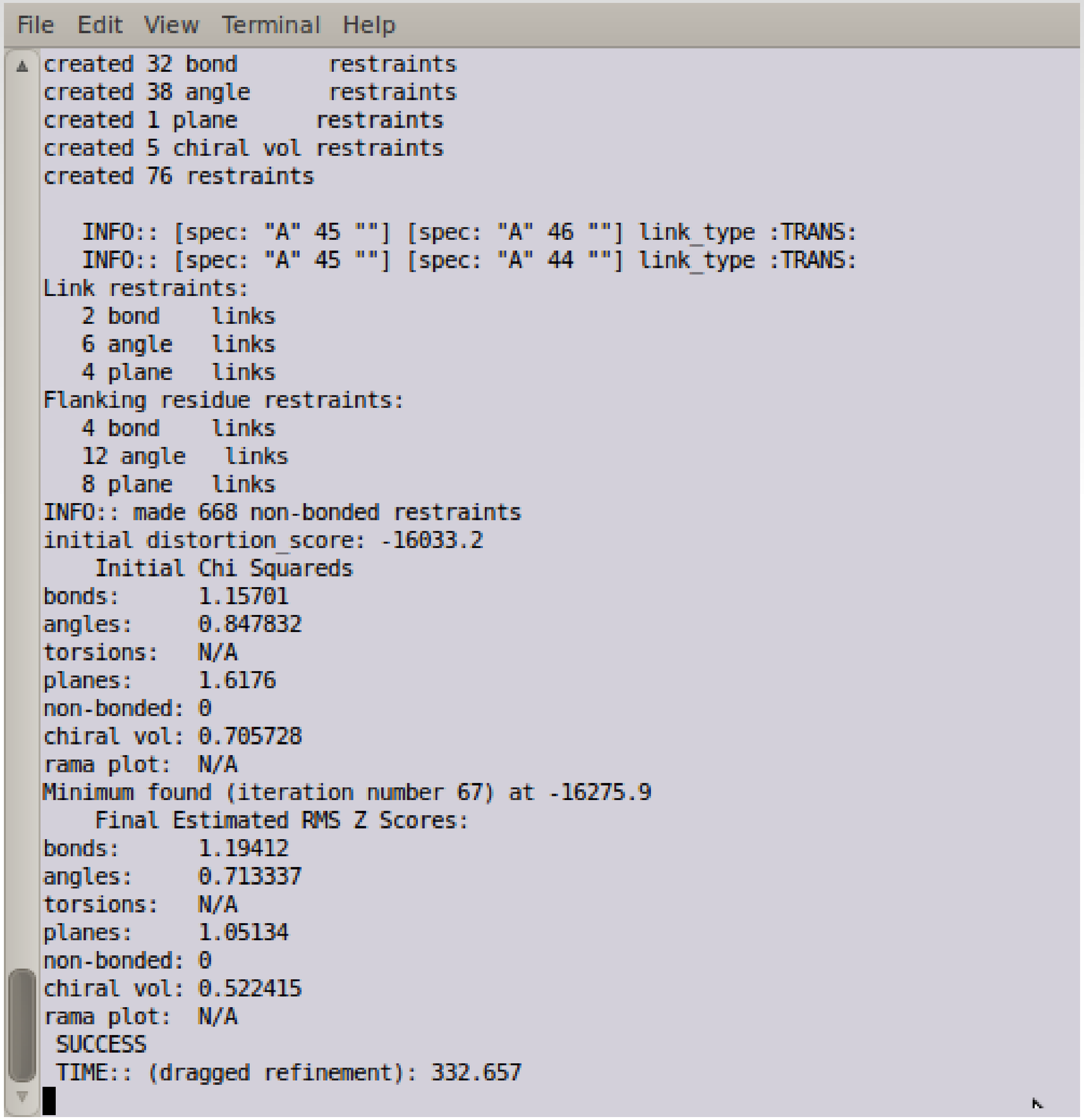


Real Space Refinement

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:



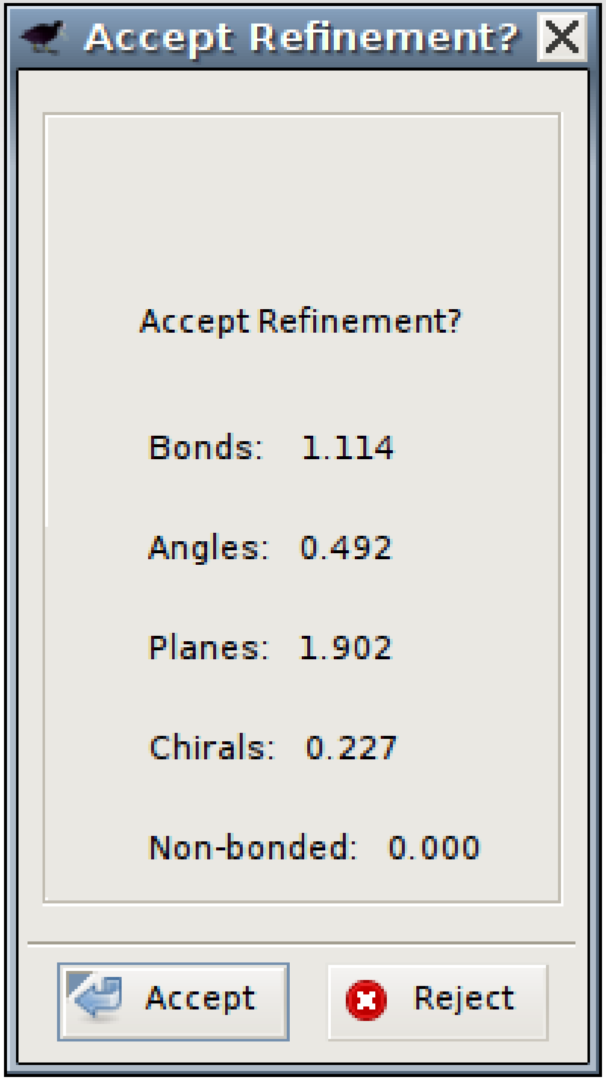
```
created 32 bond       restraints
created 38 angle      restraints
created 1 plane      restraints
created 5 chiral vol restraints
created 76 restraints

   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 main window immediately)

Representation of Results:



Accept Refinement?

Accept Refinement?

Bonds: 1.114

Angles: 0.492

Planes: 1.902

Chirals: 0.227

Non-bonded: 0.000

Second attempt...

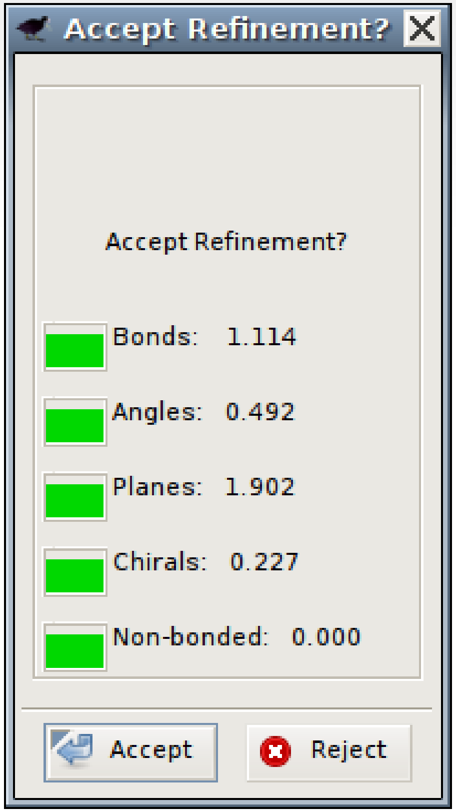
Student Reaction:

“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



Accept Refinement?

🟩 Bonds: 1.114

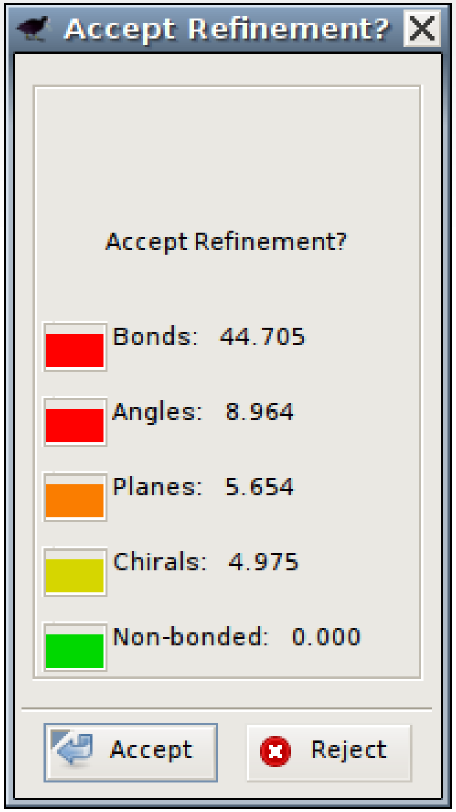
🟩 Angles: 0.492

🟩 Planes: 1.902

🟩 Chirals: 0.227

🟩 Non-bonded: 0.000

Good refinement



Accept Refinement?

🟥 Bonds: 44.705

🟥 Angles: 8.964

🟧 Planes: 5.654

🟨 Chirals: 4.975

🟩 Non-bonded: 0.000

Bad refinement

Refinement Techniques

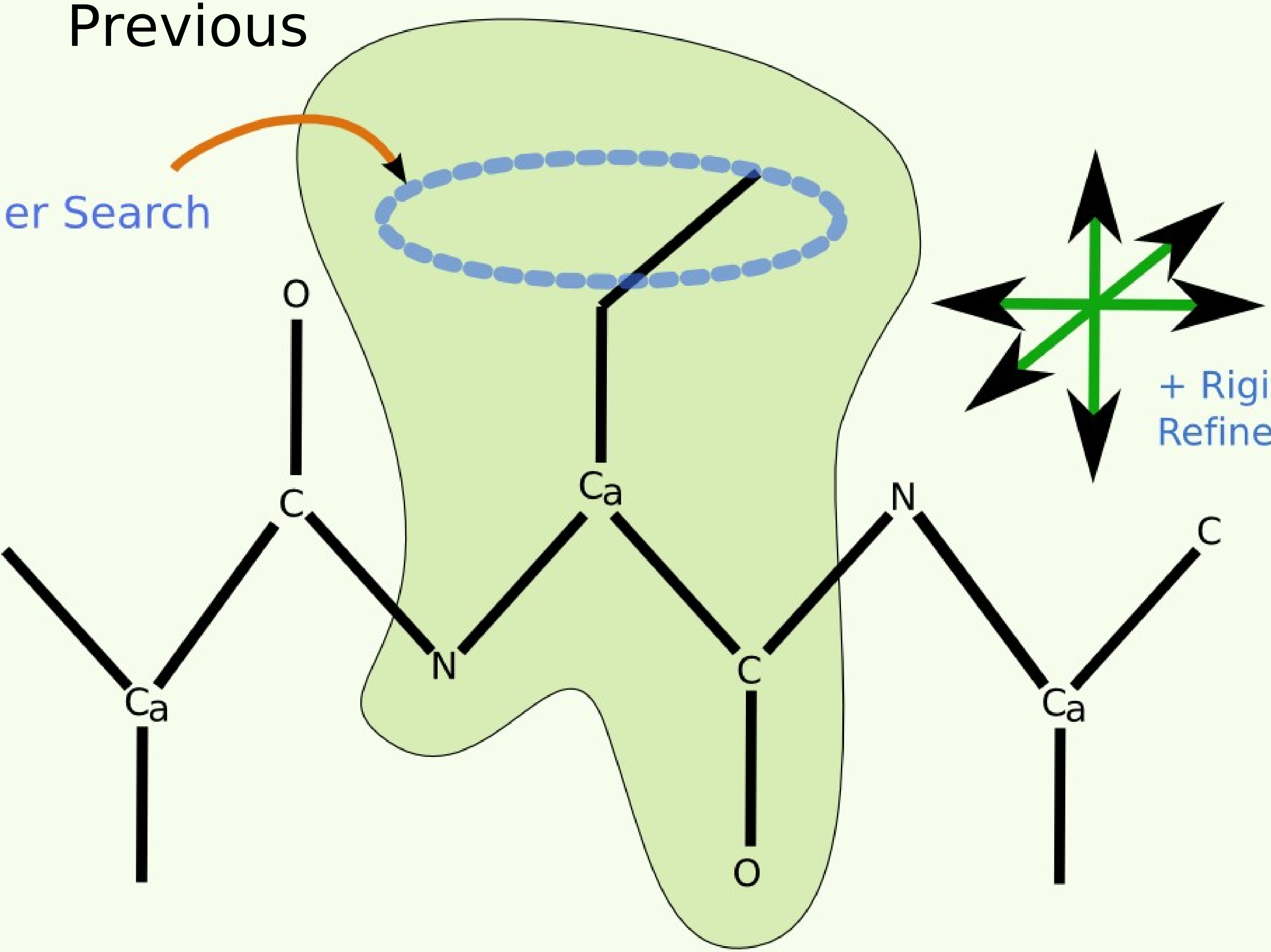
- Single-Atom Drag
  - Over-dragging
- Key-bindings:
  - Triple Refine
  - Single Residue Refine with Auto-accept

Low Resolution Model-Building

- Helix fitting
- “Backrub” rotamers

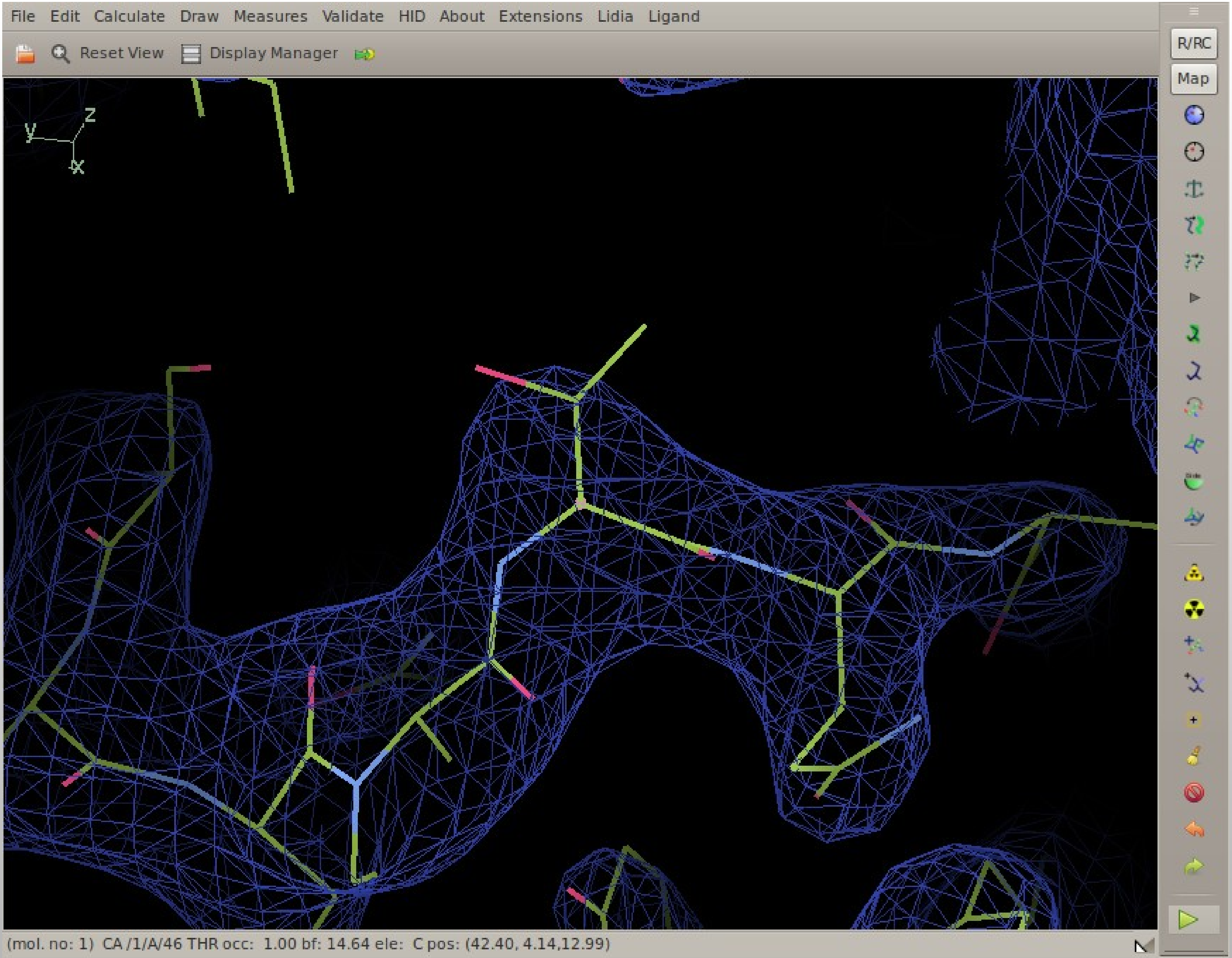
~~Current~~ Low Resolution Rotamer Search

Previous

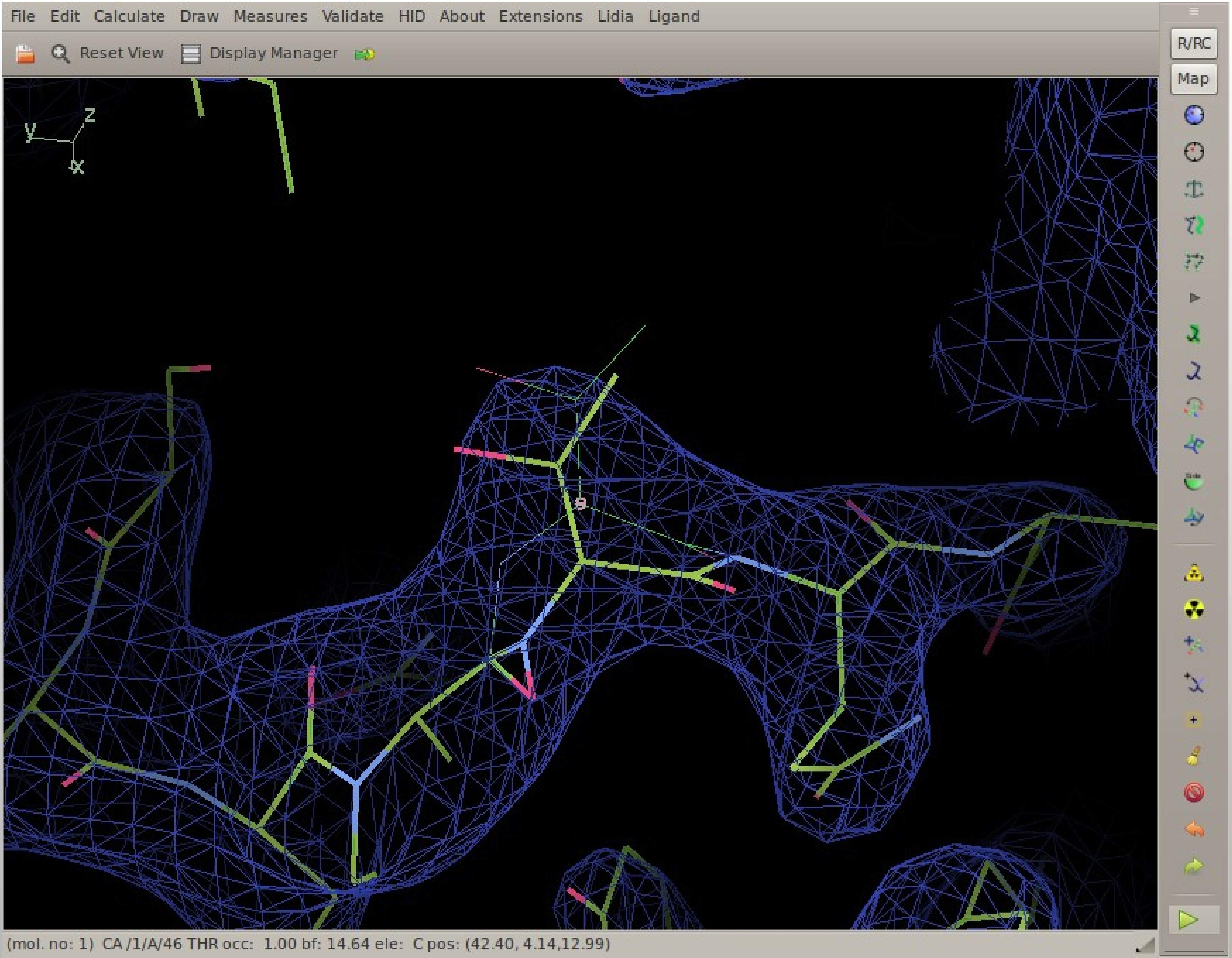


Rotamer Search

+ Rigid Body Refinement



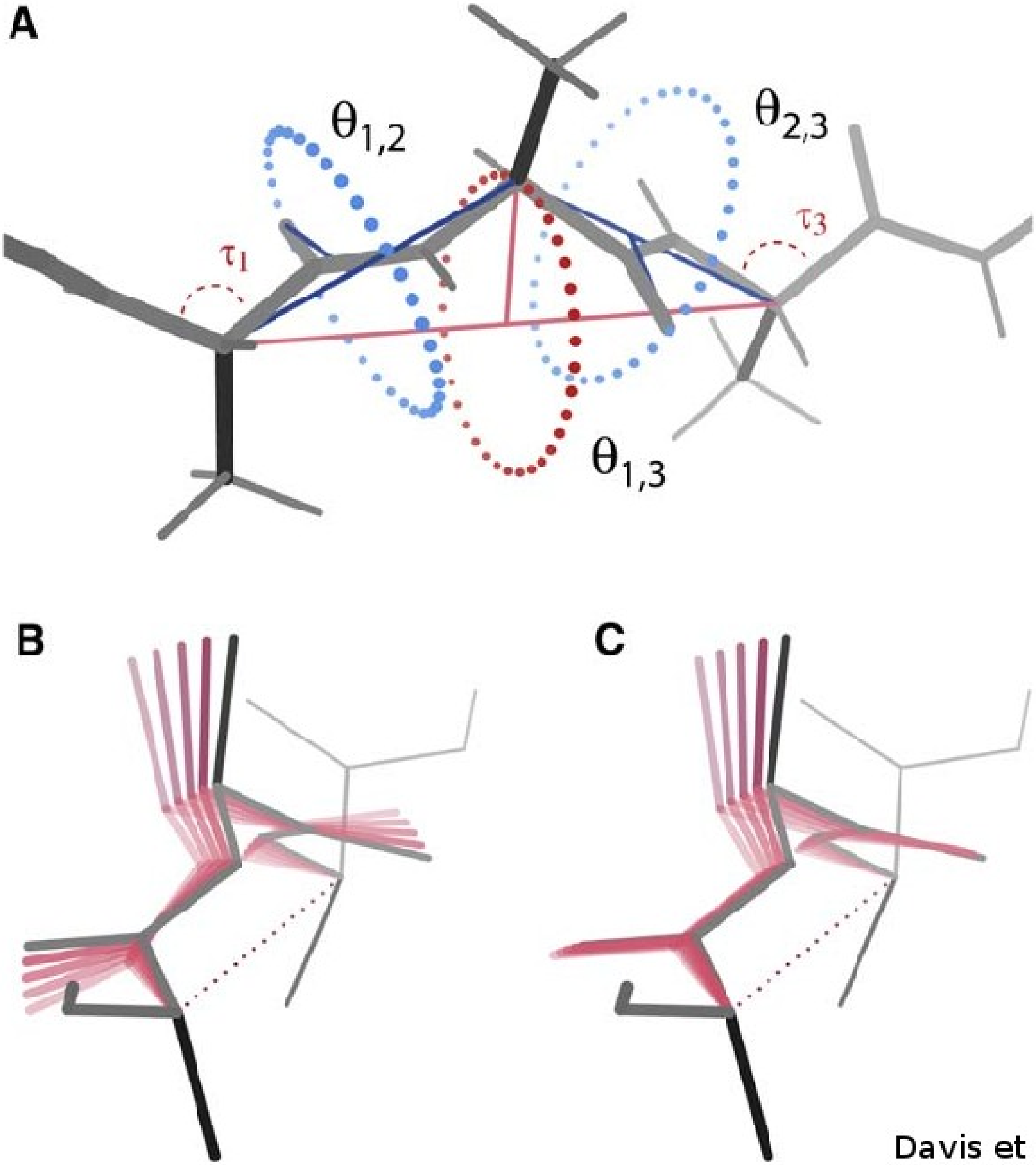
```
(mol. no: 1)  CA /1/A/46 THR occ:  1.00 bf: 14.64 ele:  C pos: (42.40, 4.14,12.99)
```



(mol. no: 1) CA /1/A/46 THR occ: 1.00 bf: 14.64 ele: C pos: (42.40, 4.14,12.99)

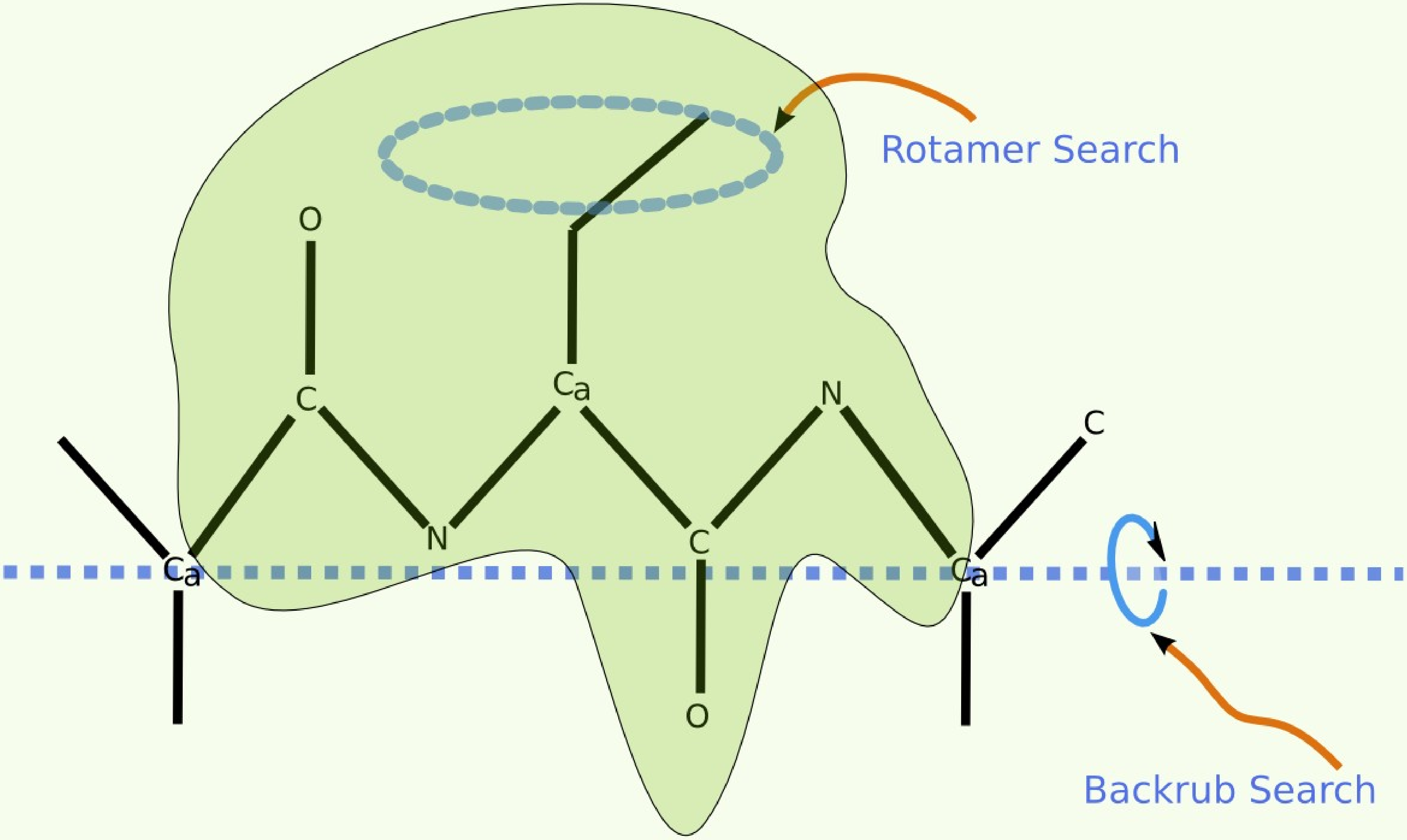
Using Backrubs

- for rotamer fitting

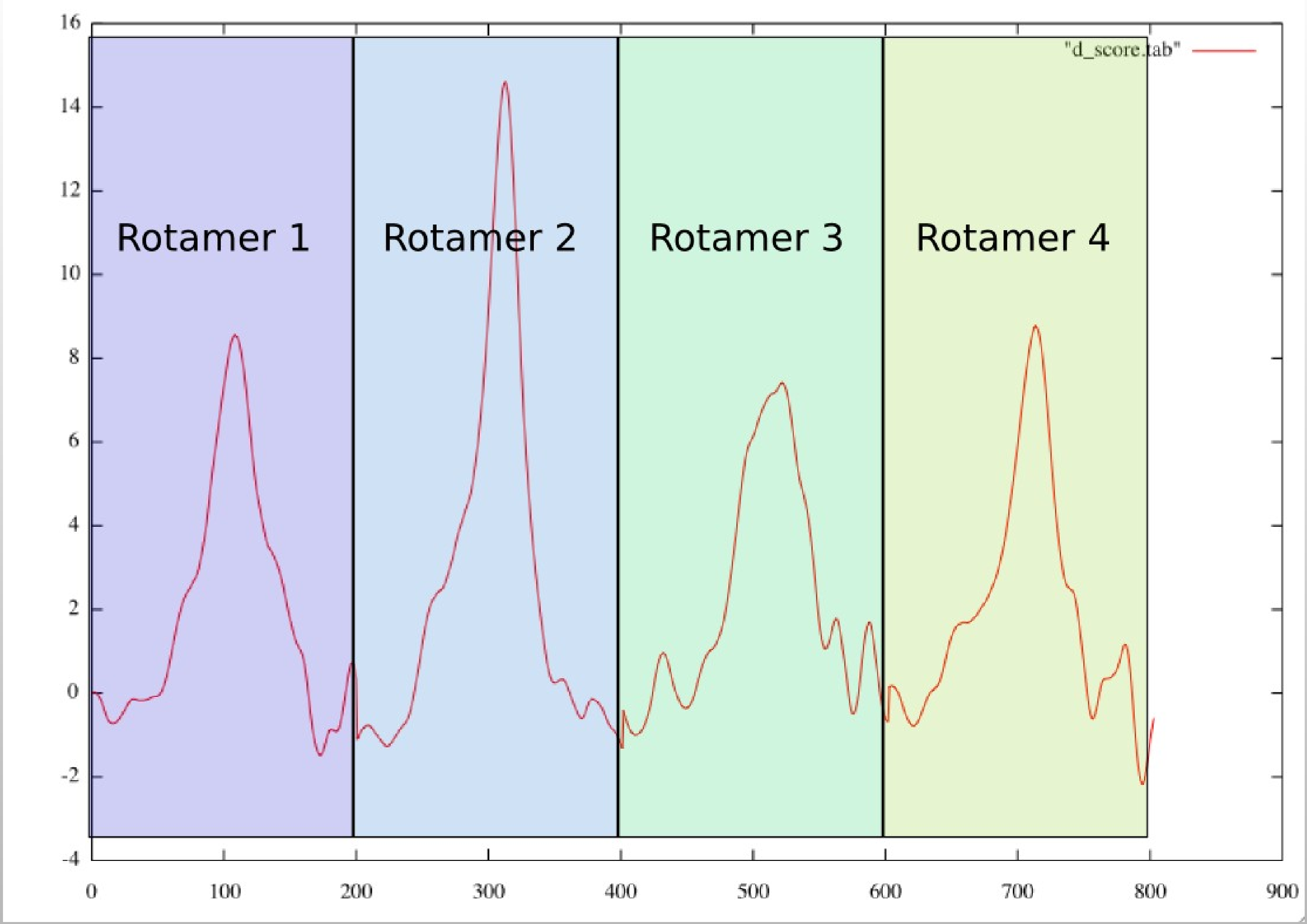


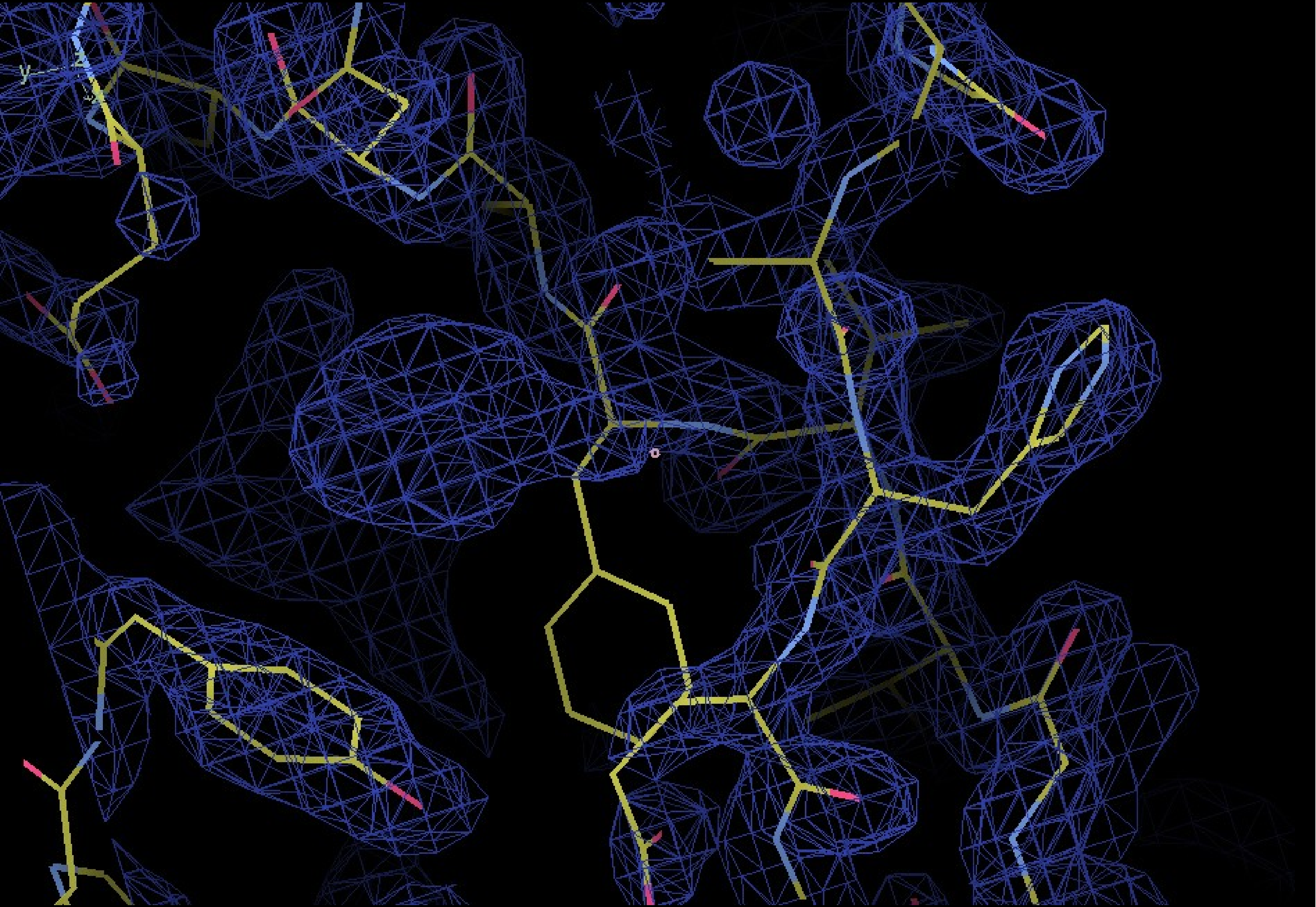
Davis et al. (2006) Structure

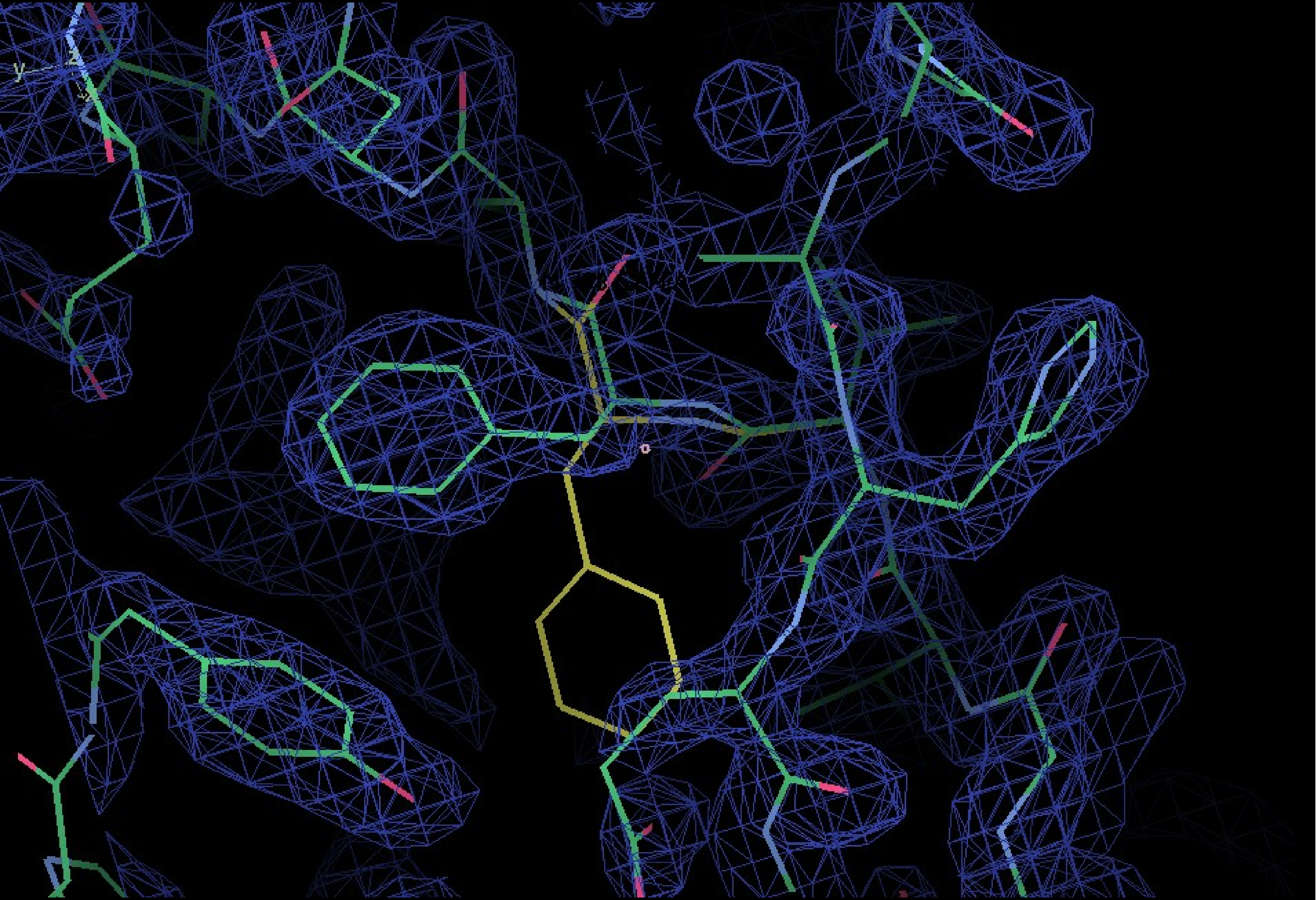
New Low Resolution Rotamer Search



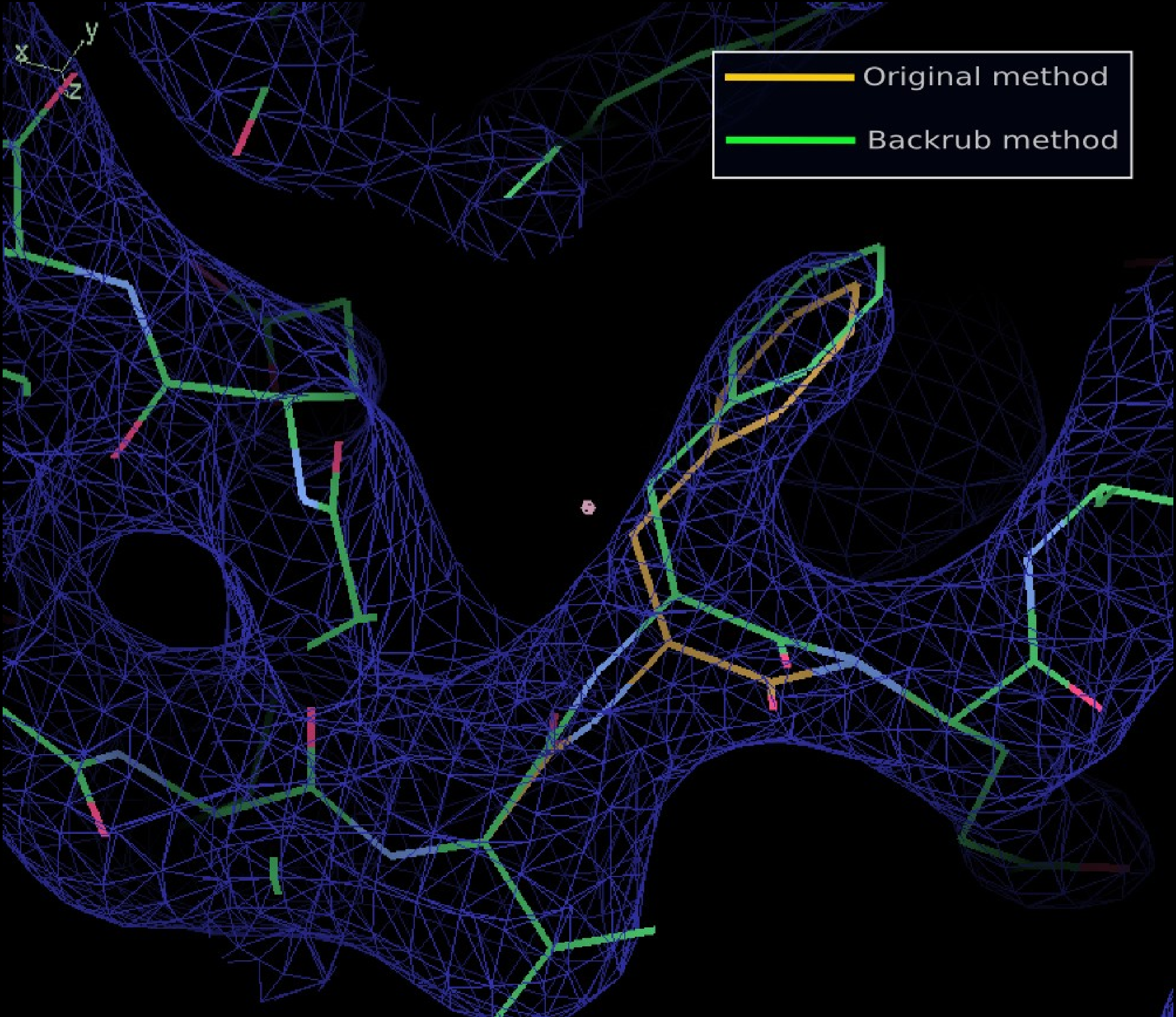
After Fitting Tools in KING/Molprobity


"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
z
x

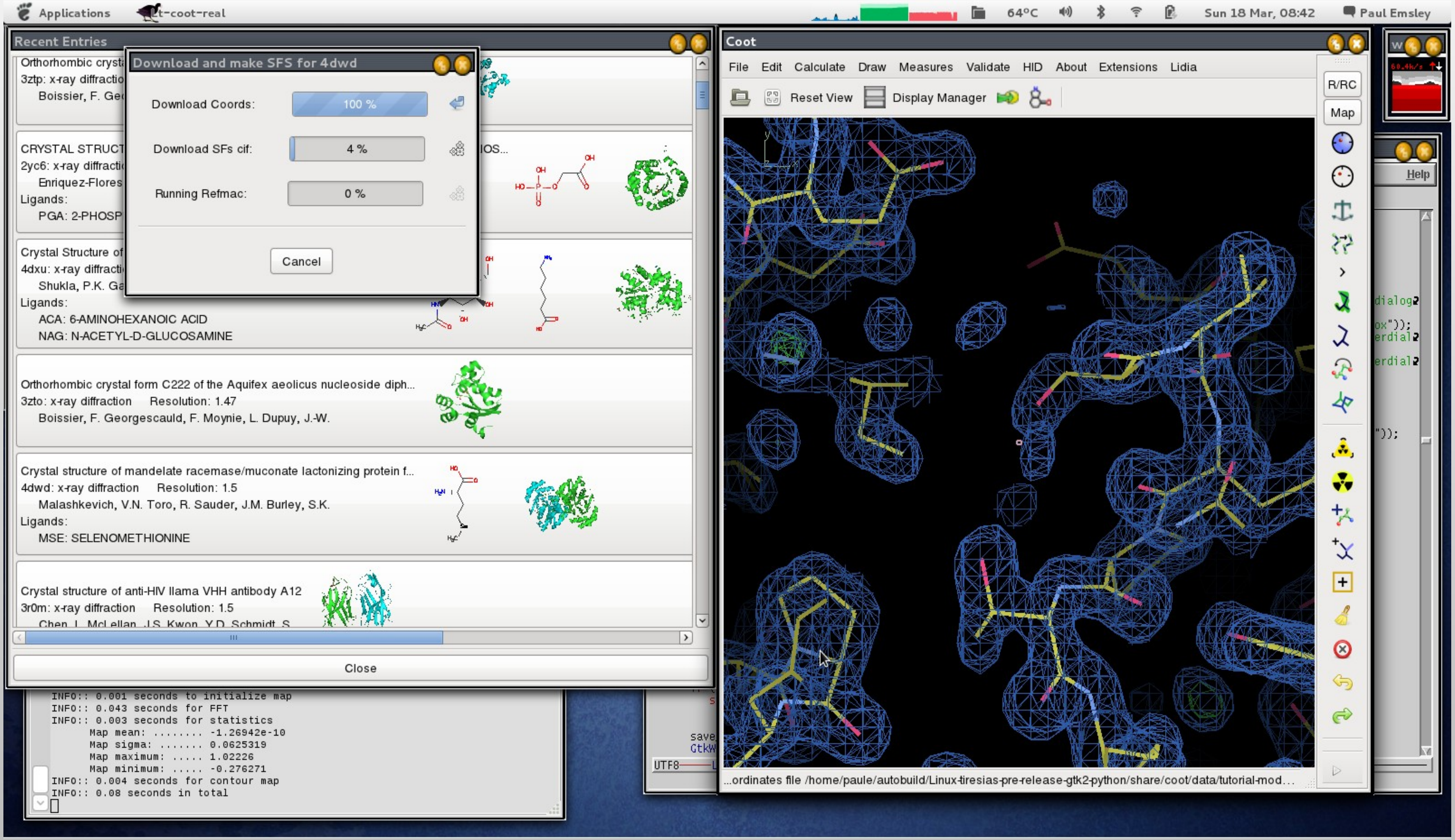

x
y
z
Original method
Backrub method

Networking...

- PDBe interface...
- Drag and drop
    - Also with drugbank

PDBe Recent Structures

JSON parser, network threaded code



Download Coords: 100 %

Download SFs cif: 4 %

Running Refmac: 0 %

Orthorhombic crysta
3ztp: x-ray diffractio
Boissier, F. Ge

CRYSTAL STRUCT … IOS...
2yc6: x-ray diffracti
Enriquez-Flores
Ligands:
PGA: 2-PHOSP

Crystal Structure of
4dxu: x-ray diffracti
Shukla, P.K. Ga
Ligands:
ACA: 6-AMINOHEXANOIC ACID
NAG: N-ACETYL-D-GLUCOSAMINE

Orthorhombic crystal form C222 of the Aquifex aeolicus nucleoside diph...
3zto: x-ray diffraction Resolution: 1.47
Boissier, F. Georgescauld, F. Moynie, L. Dupuy, J.-W.

Crystal structure of mandelate racemase/muconate lactonizing protein f...
4dwd: x-ray diffraction Resolution: 1.5
Malashkevich, V.N. Toro, R. Sauder, J.M. Burley, S.K.
Ligands:
MSE: SELENOMETHIONINE

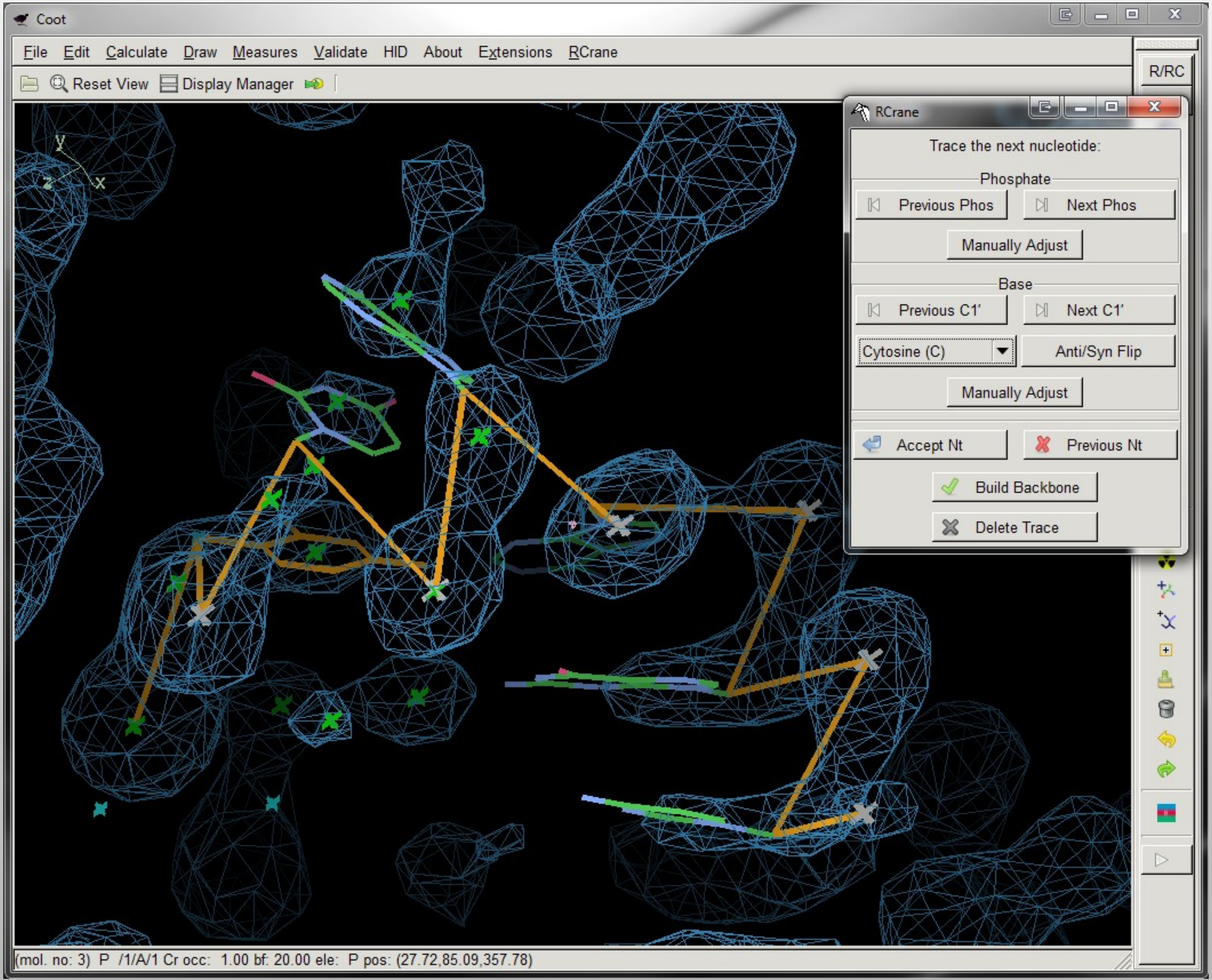
Crystal structure of anti-HIV llama VHH antibody A12
3r0m: x-ray diffraction Resolution: 1.5
Chen, L. McLellan, J.S. Kwon, Y.D. Schmidt, S.

```
INFO:: 0.001 seconds to initialize map
INFO:: 0.043 seconds for FFT
INFO:: 0.003 seconds for statistics
      Map mean: ........ -1.26942e-10
      Map sigma: ....... 0.0625319
      Map maximum: ..... 1.02226
      Map minimum: ..... -0.276271
INFO:: 0.004 seconds for contour map
INFO:: 0.08 seconds in total
```

...ordinates file /home/paule/autobuild/Linux-tiresias-pre-release-gtk2-python/share/coot/data/tutorial-mod...

Using the API

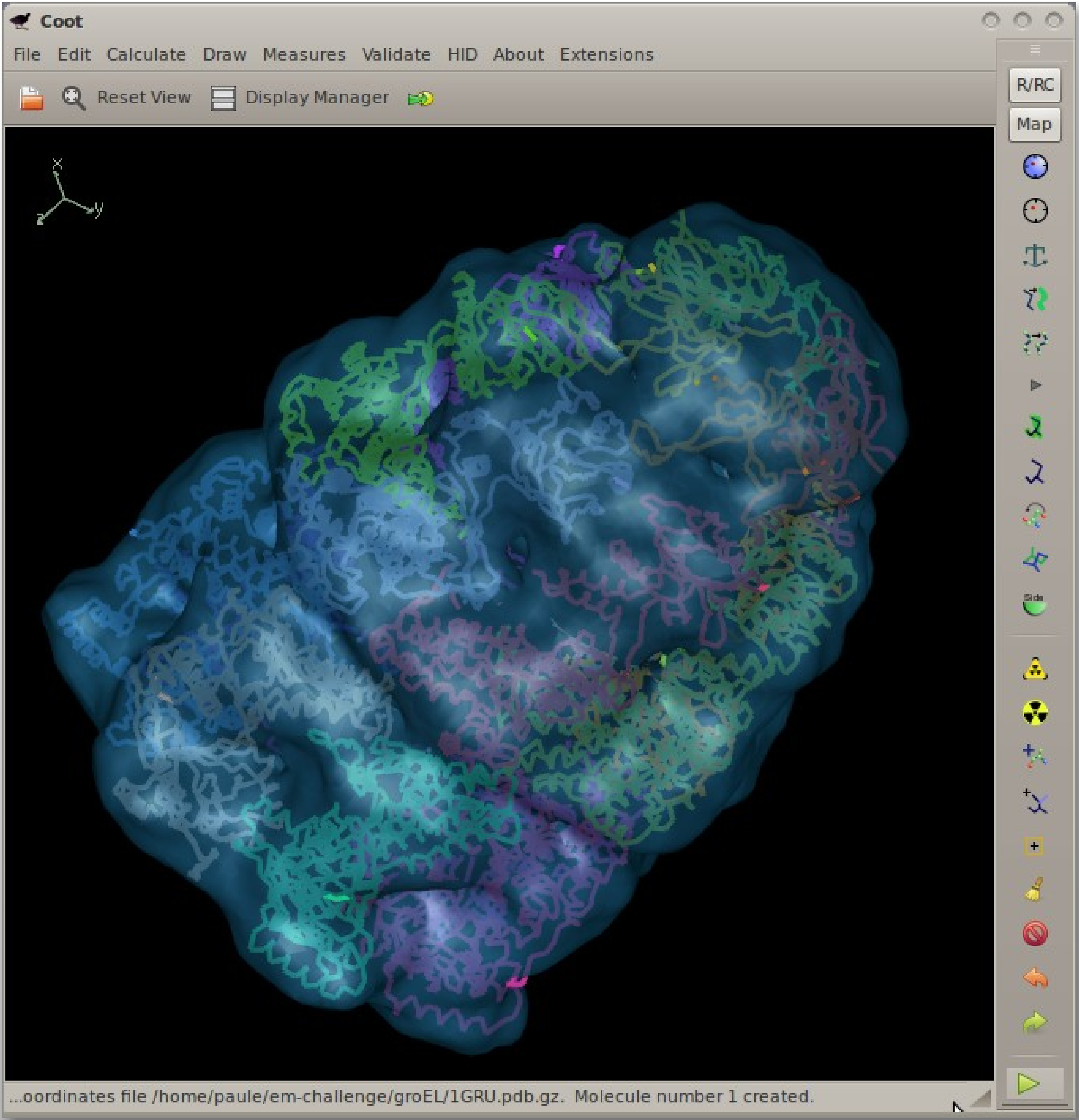
RCrane: Semi-Automated Building of RNA



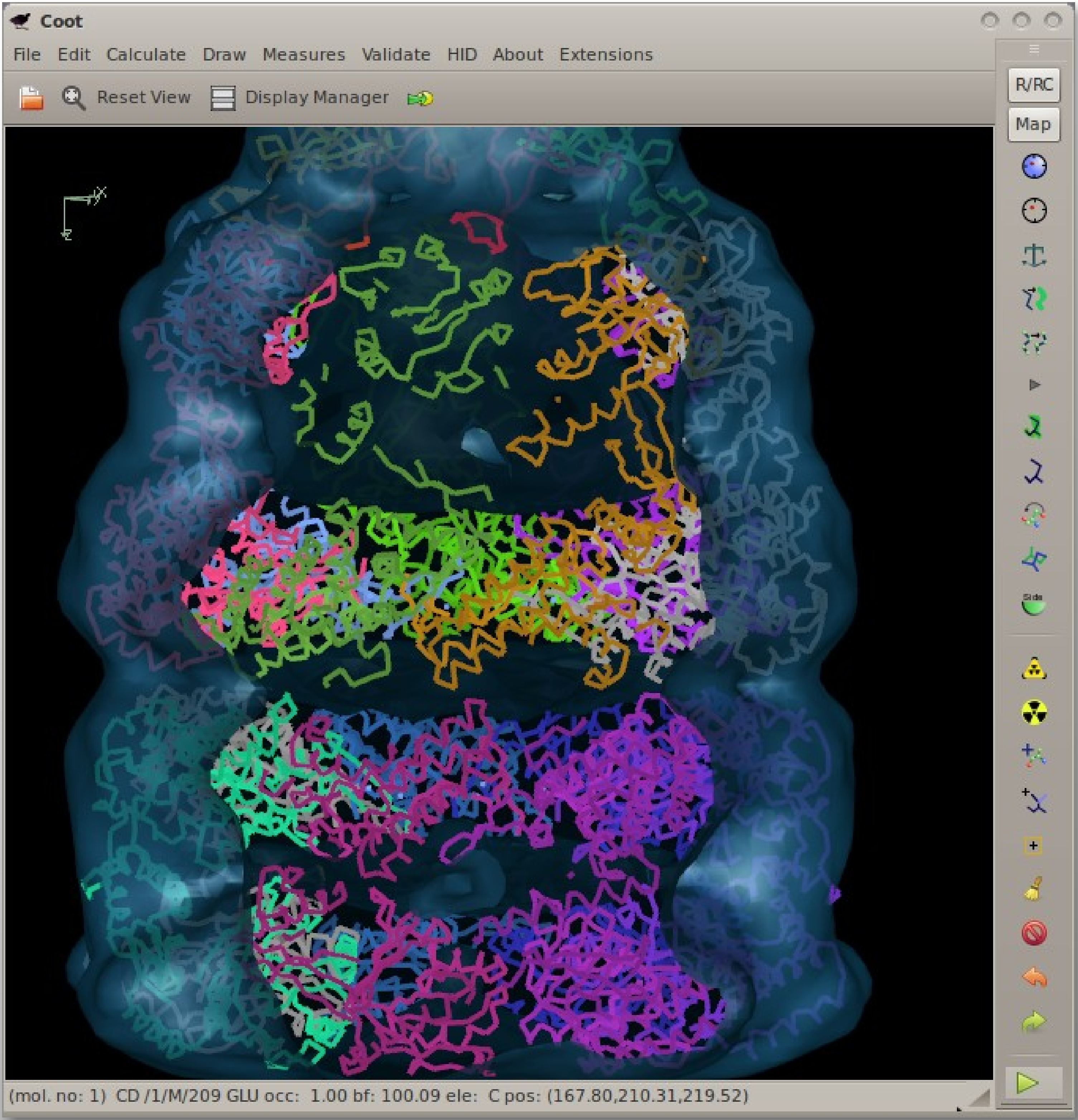
Trace the next nucleotide:

```
(mol. no: 3) P /1/A/1 Cr occ:  1.00 bf: 20.00 ele:  P pos: (27.72,85.09,357.78)
```

Handling EM maps



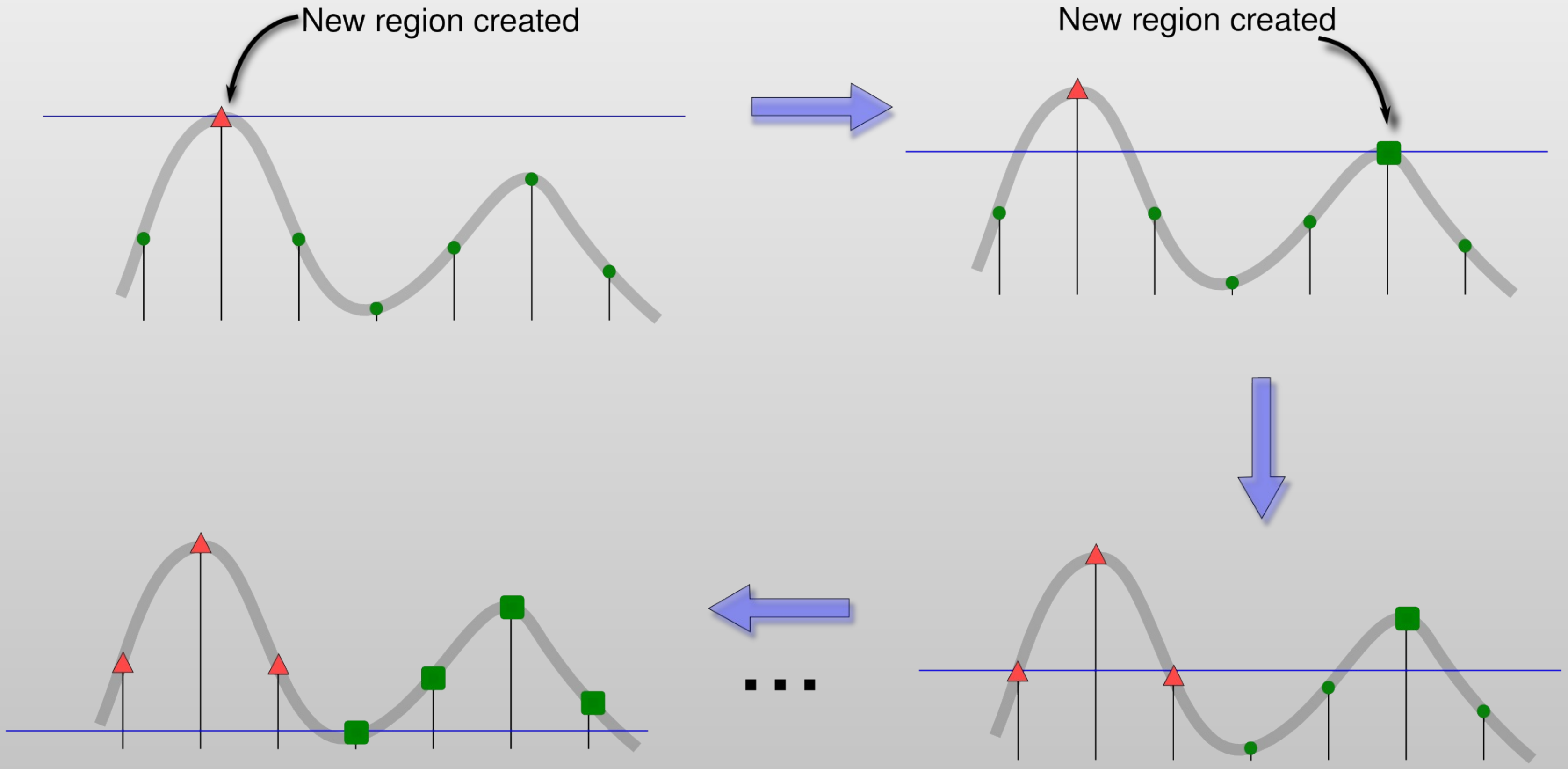
...oordinates file /home/paule/em-challenge/groEL/1GRU.pdb.gz. Molecule number 1 created.



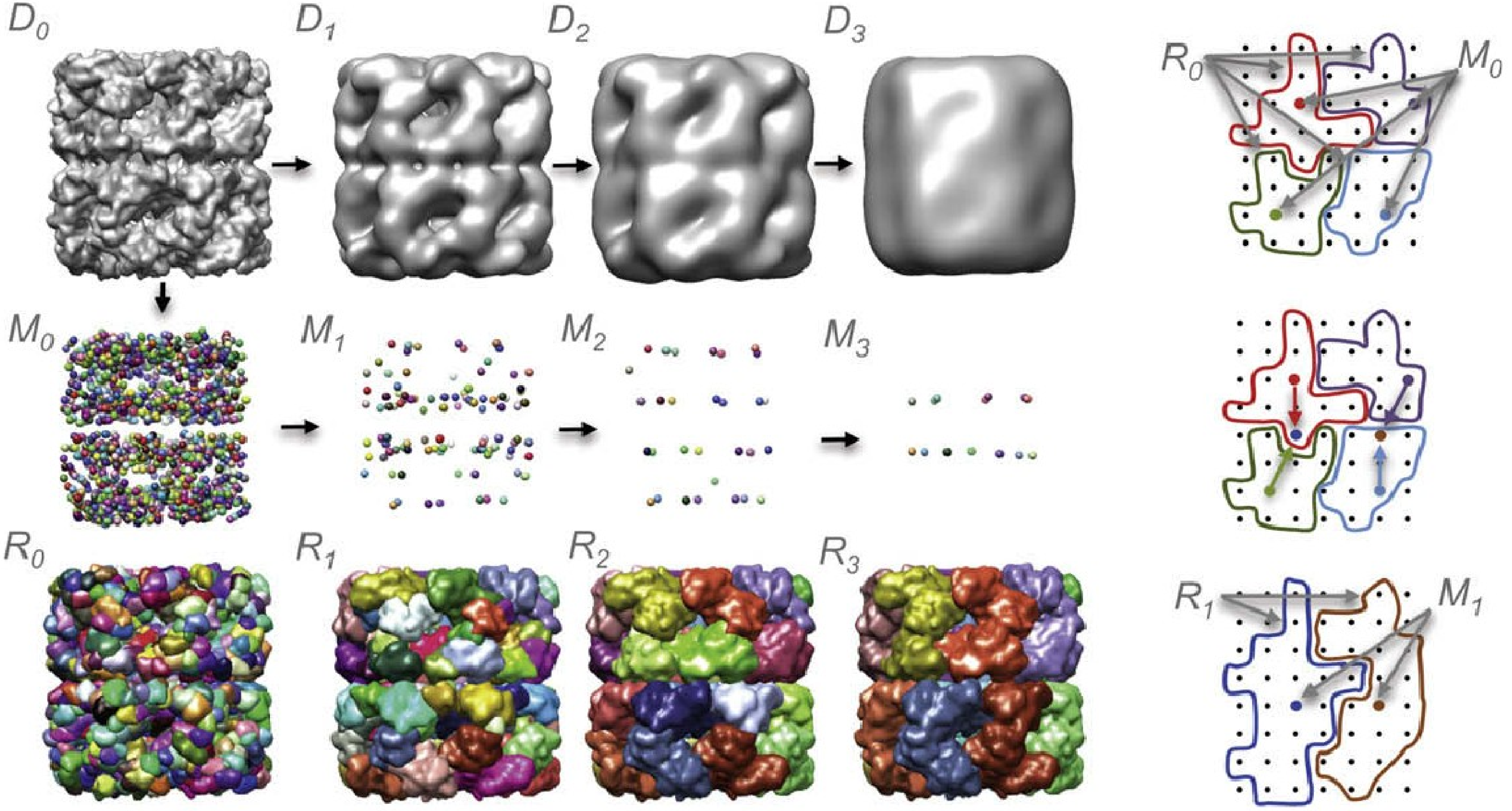
```
(mol. no: 1)  CD /1/M/209 GLU occ:  1.00 bf: 100.09 ele:  C pos: (167.80,210.31,219.52)
```

Partioning Maps: Watershed Algoritm

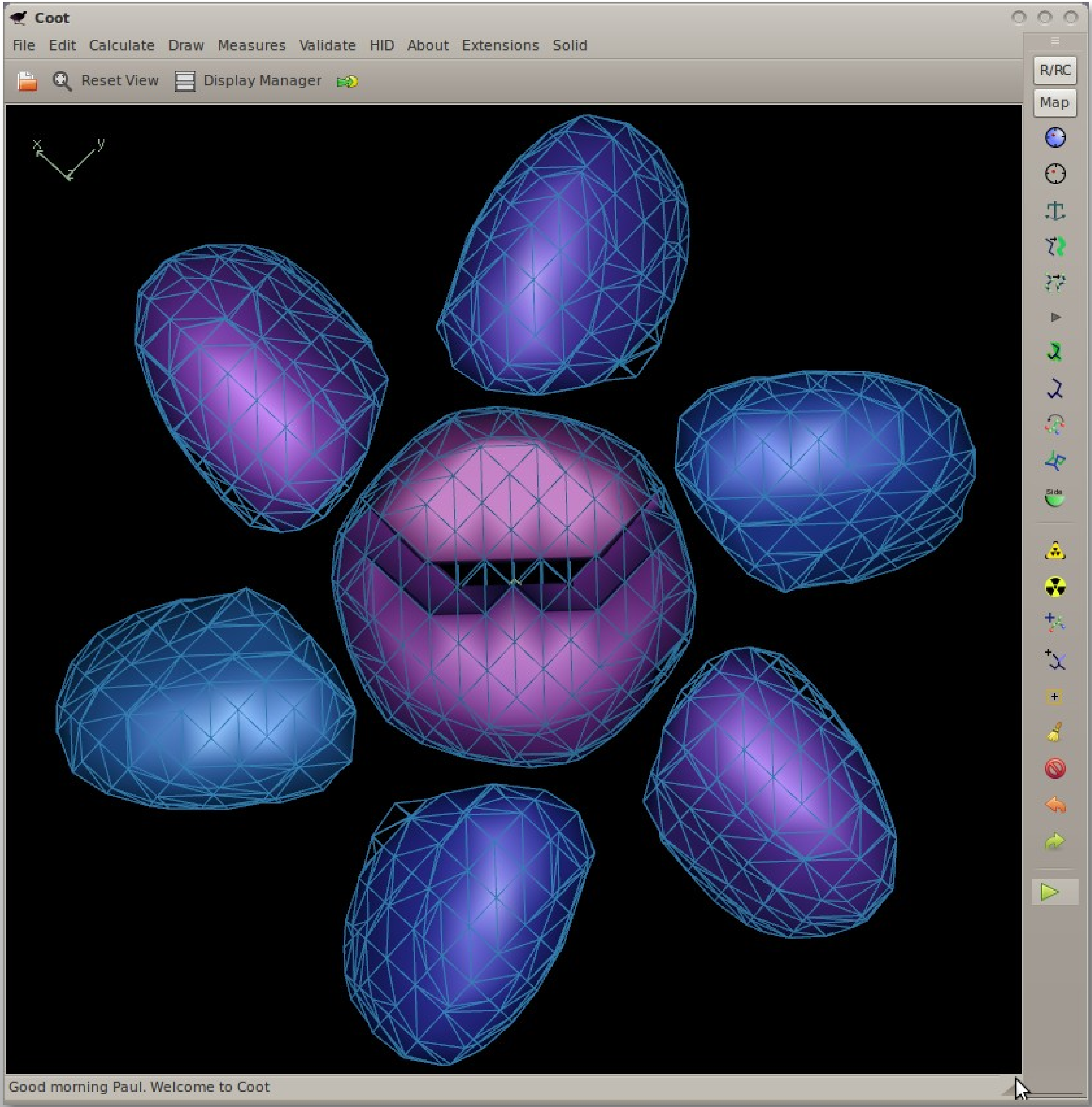
1D-analog



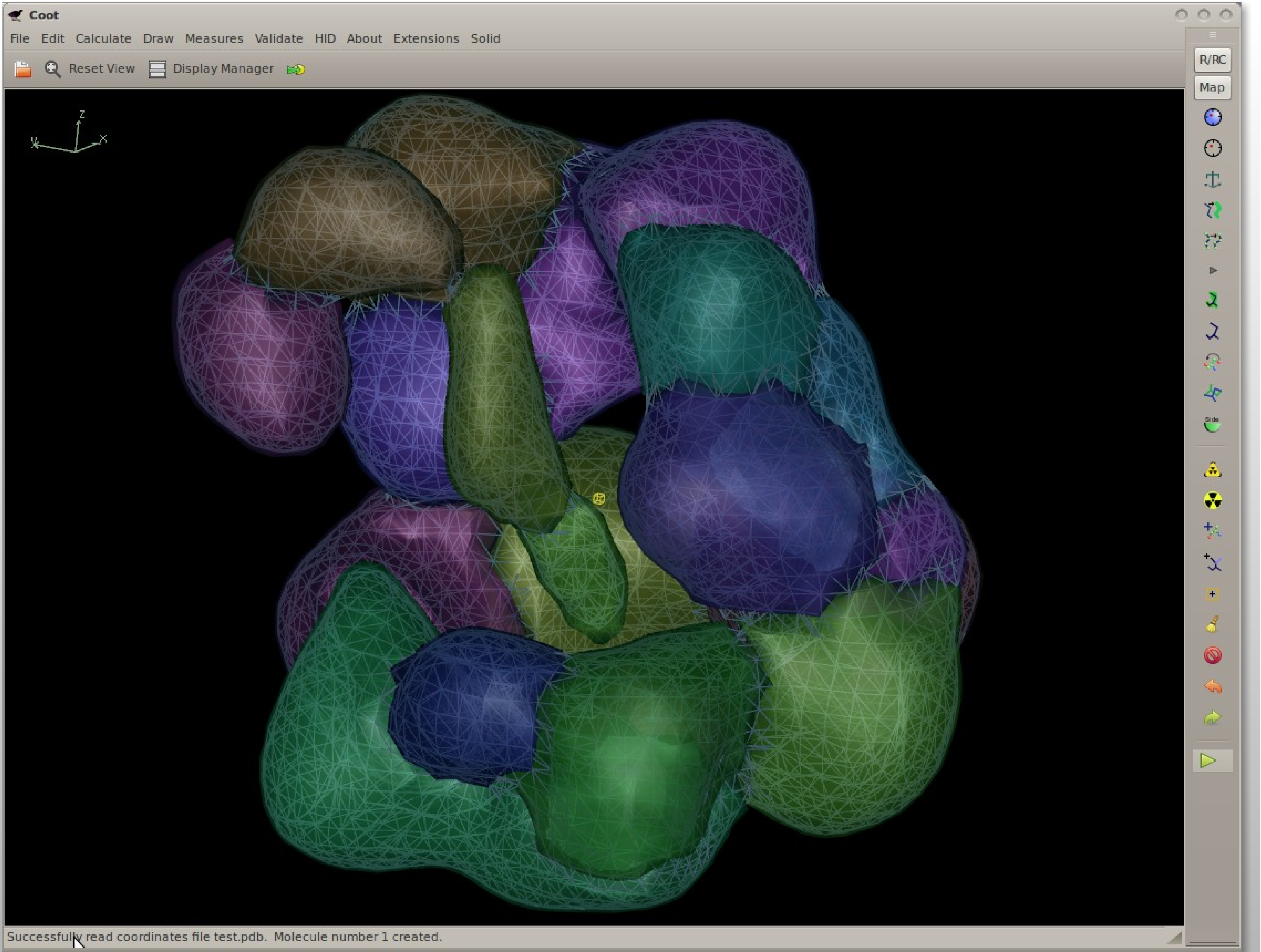
▲ 🟩 } Different segments



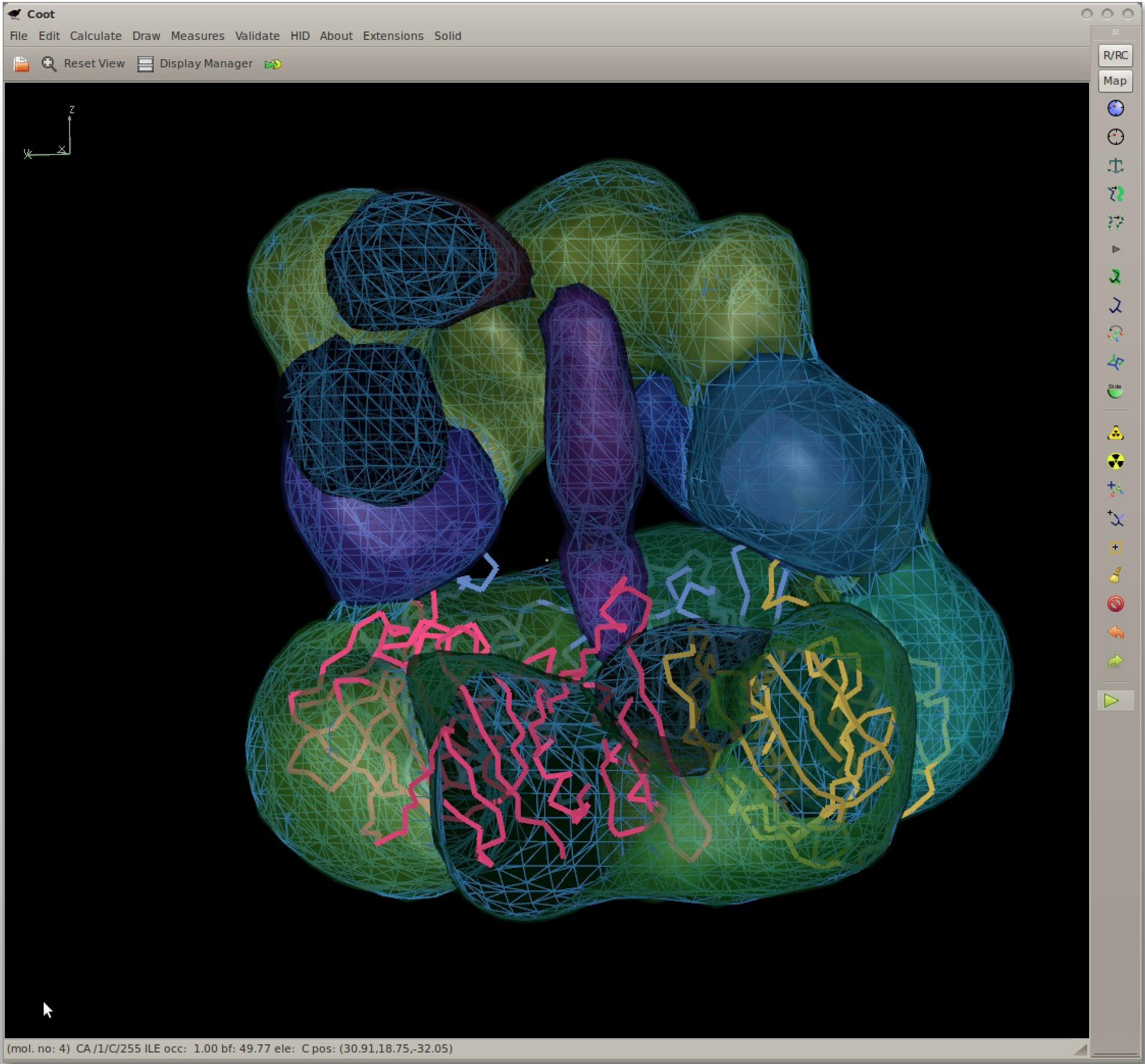
Pintilie *et al.* (2010)
J.Struct.Biol.



Good morning Paul. Welcome to Coot



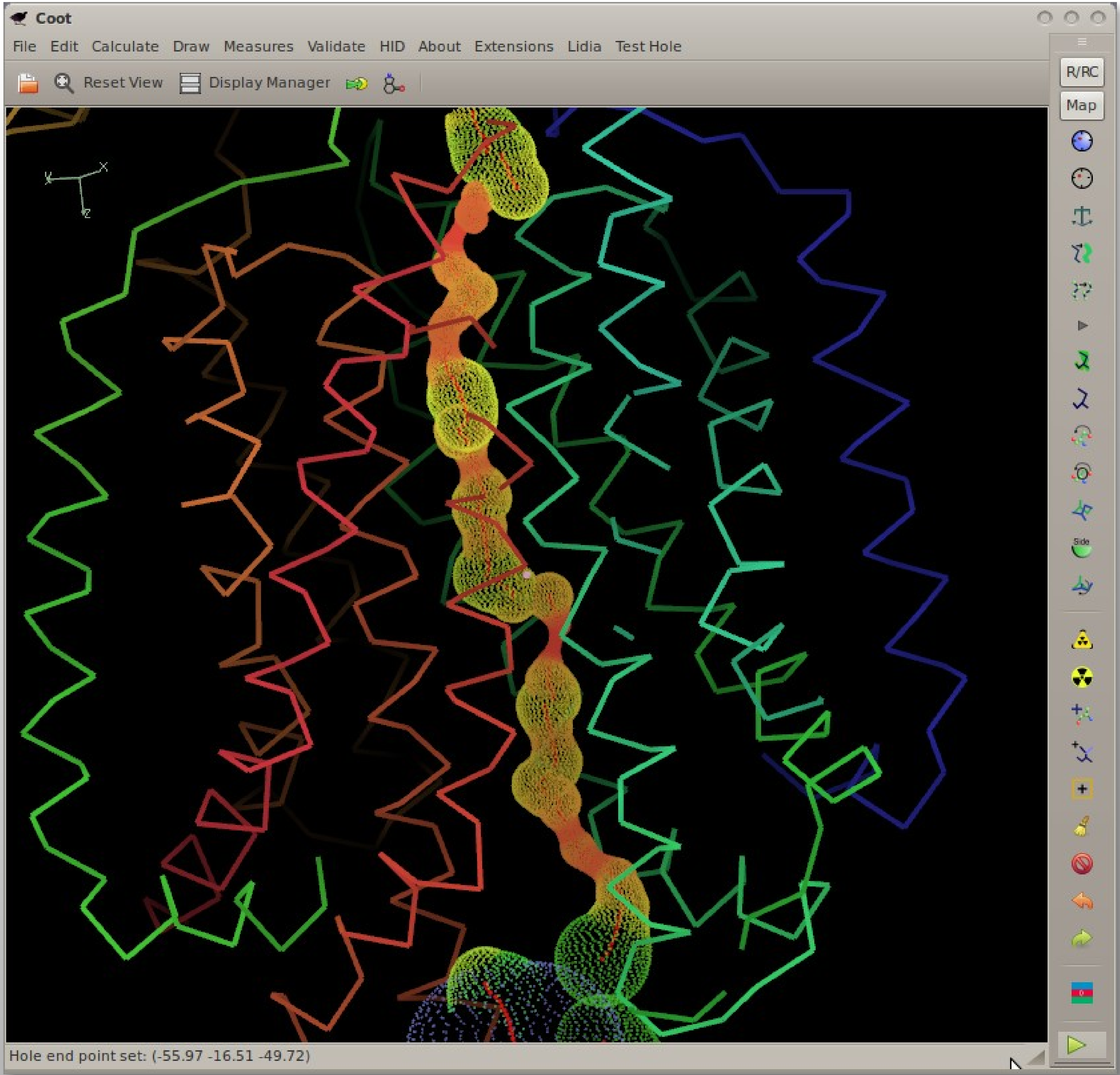
Successfully read coordinates file test.pdb. Molecule number 1 created.



```
(mol. no: 4)  CA /1/C/255 ILE occ:  1.00 bf: 49.77 ele:  C pos: (30.91,18.75,-32.05)
```

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)

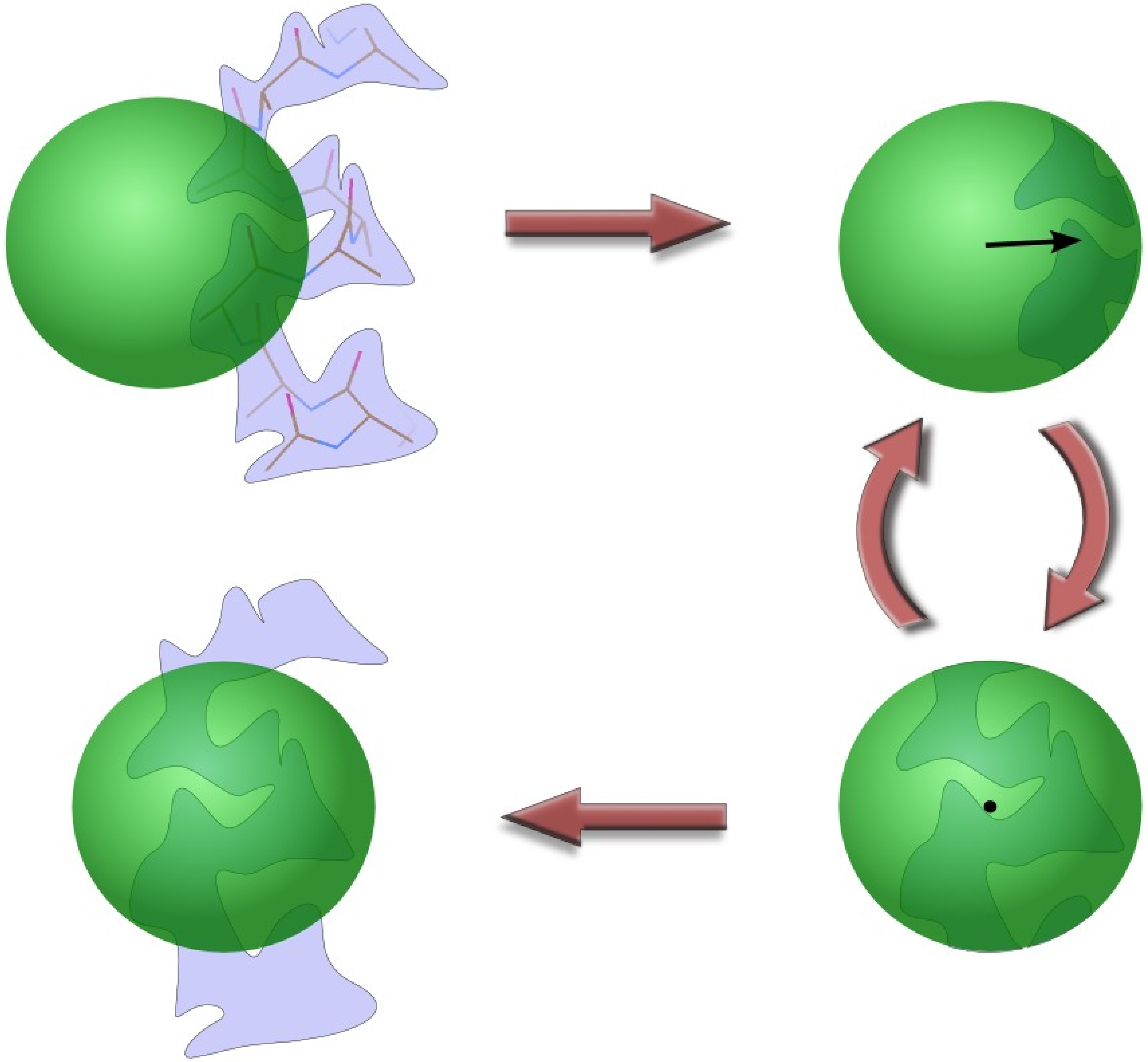


Hole end point set: (-55.97 -16.51 -49.72)

Alpha Helix Placement

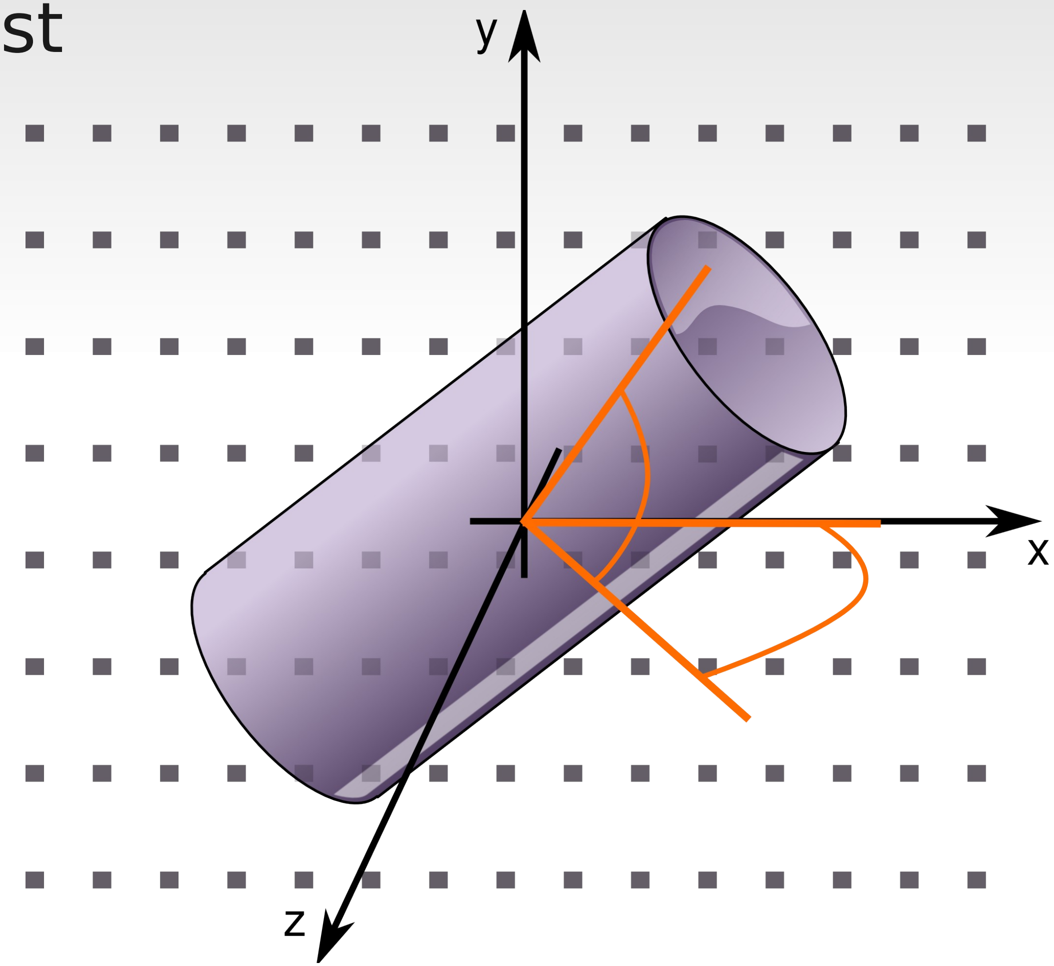
- Scenario: Looking at a new map, not built with automatic tools:
  - “I can see that there’s a helix here - build it for me!”
- From a given point:
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)
  - 1D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

Centering the Rotation point

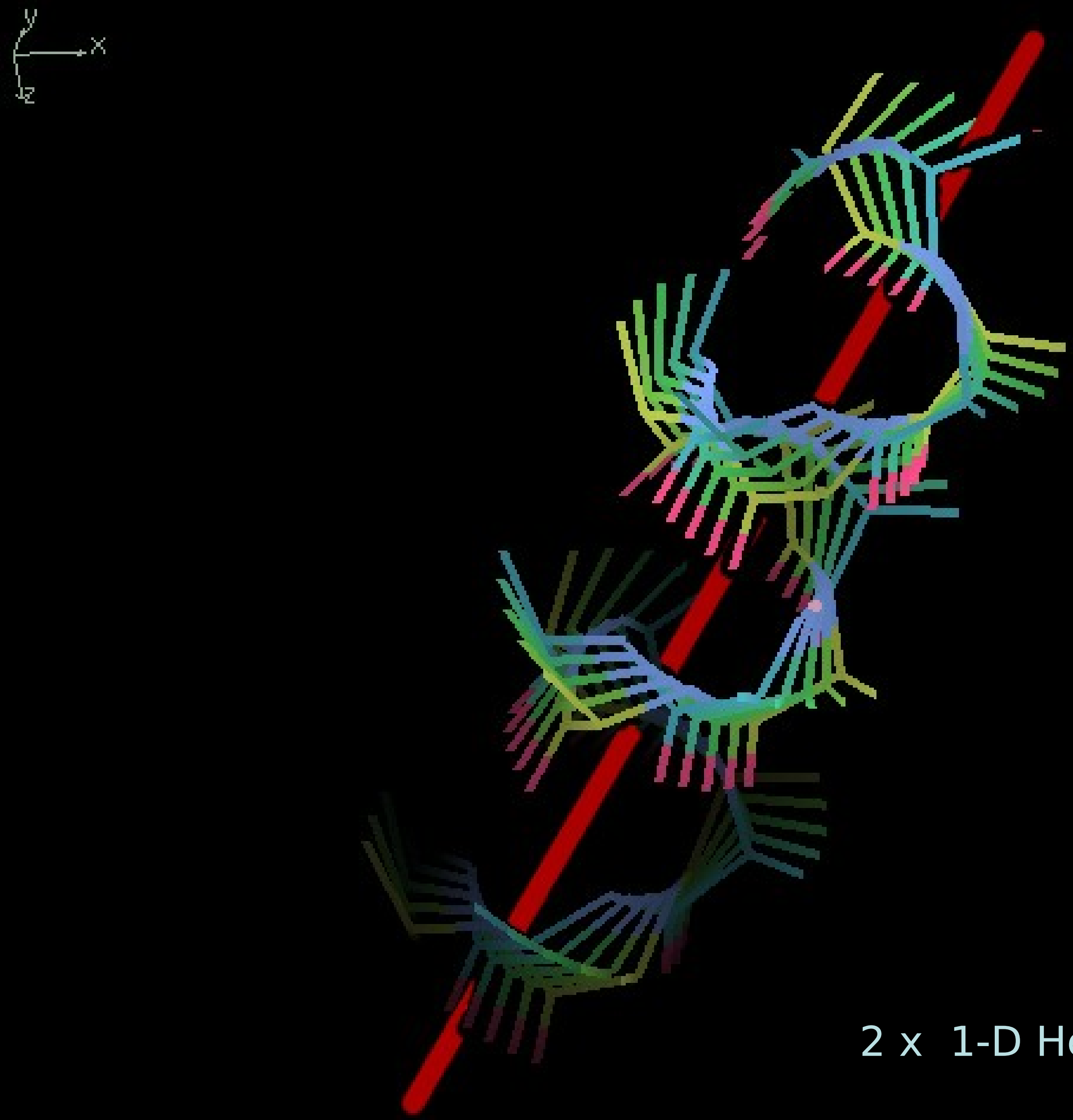


Cylinder Search

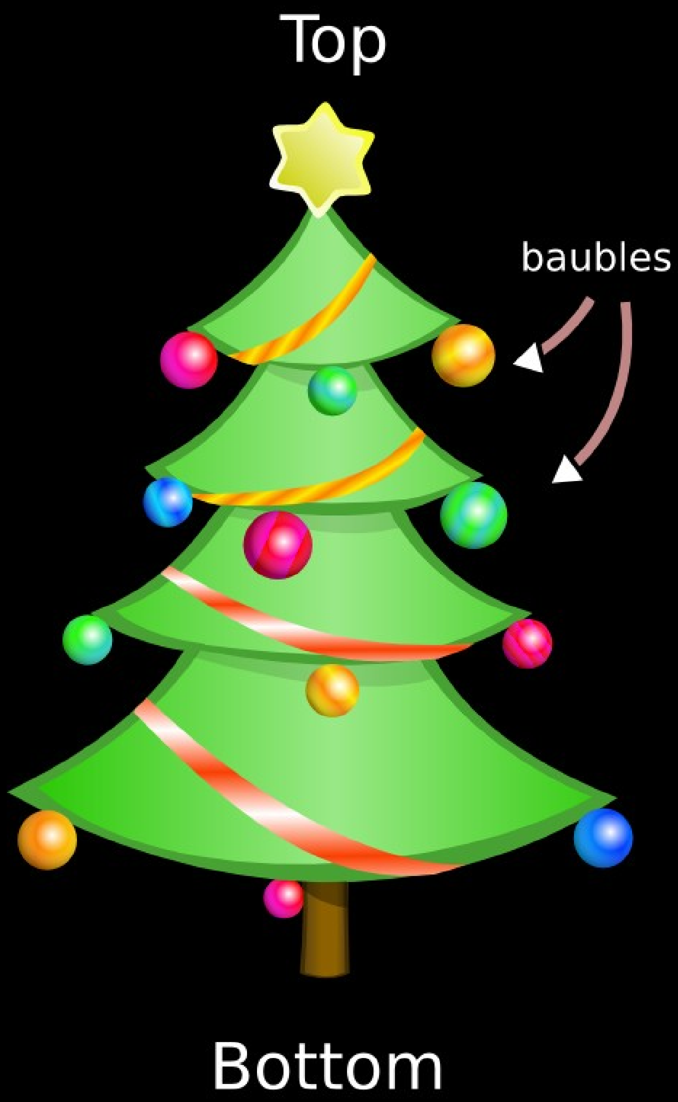
- Pick the orientation that encapsulates the most electron density

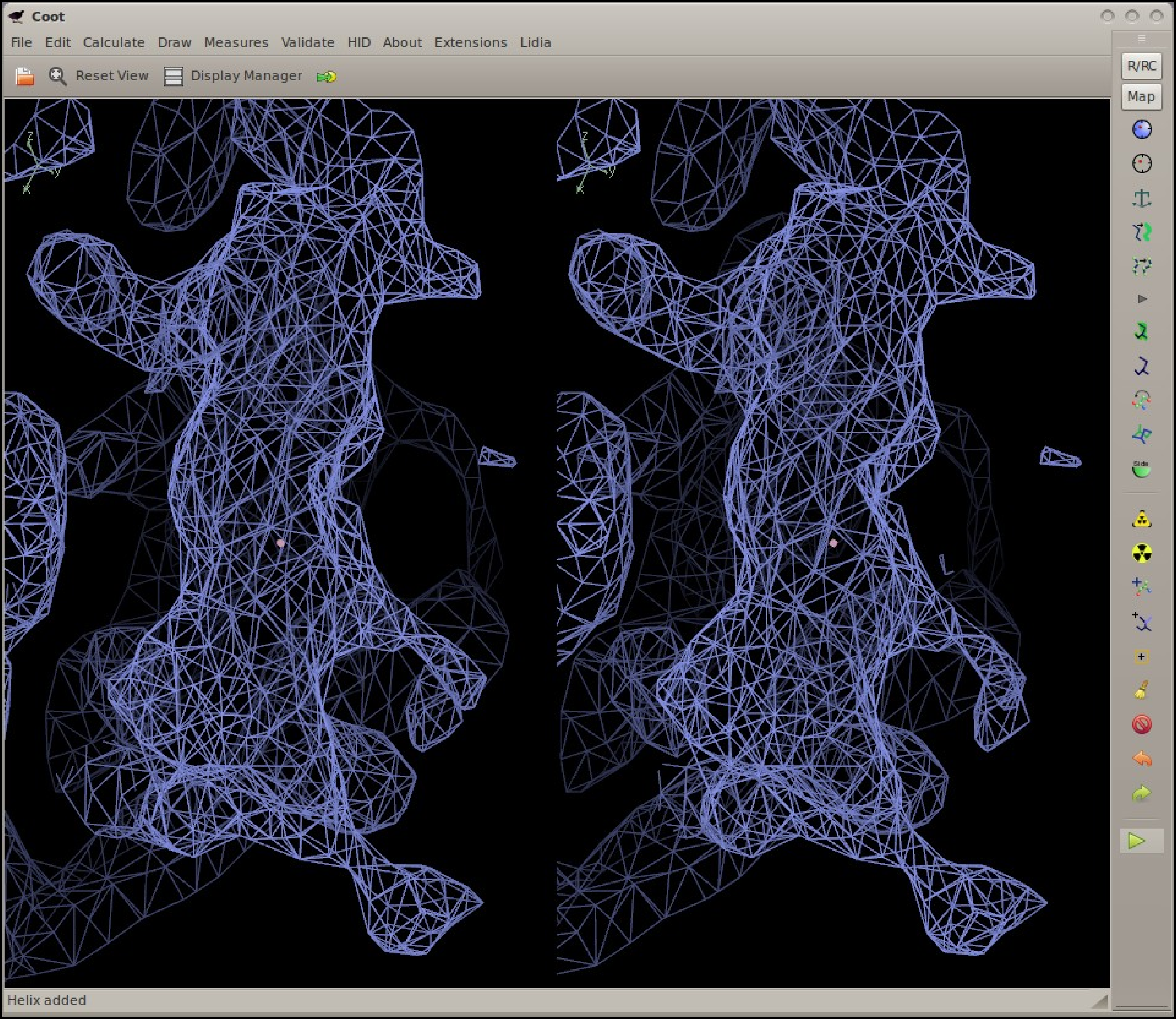


Using 2 rotation axes

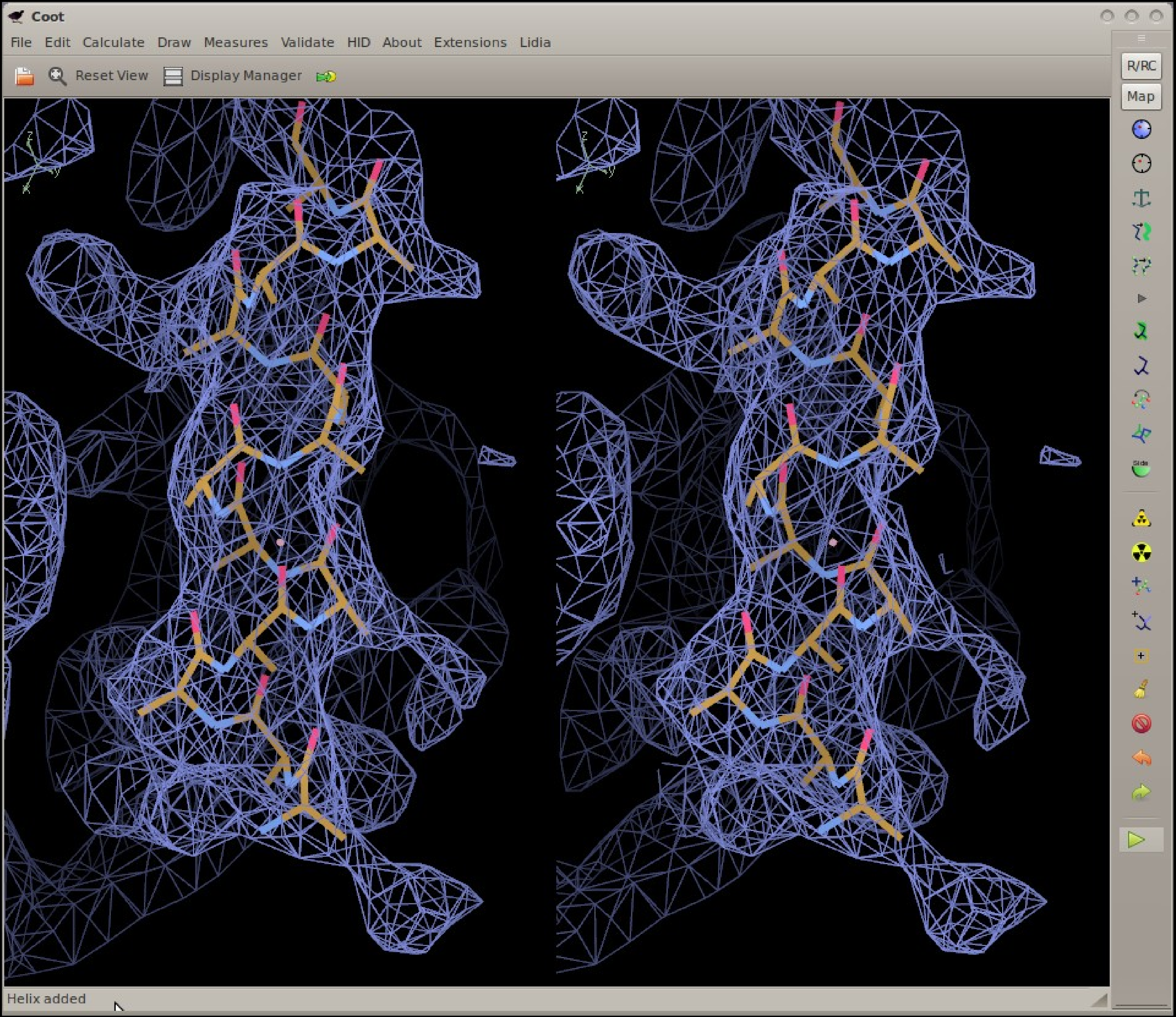


2 x 1-D Helix orientation searches

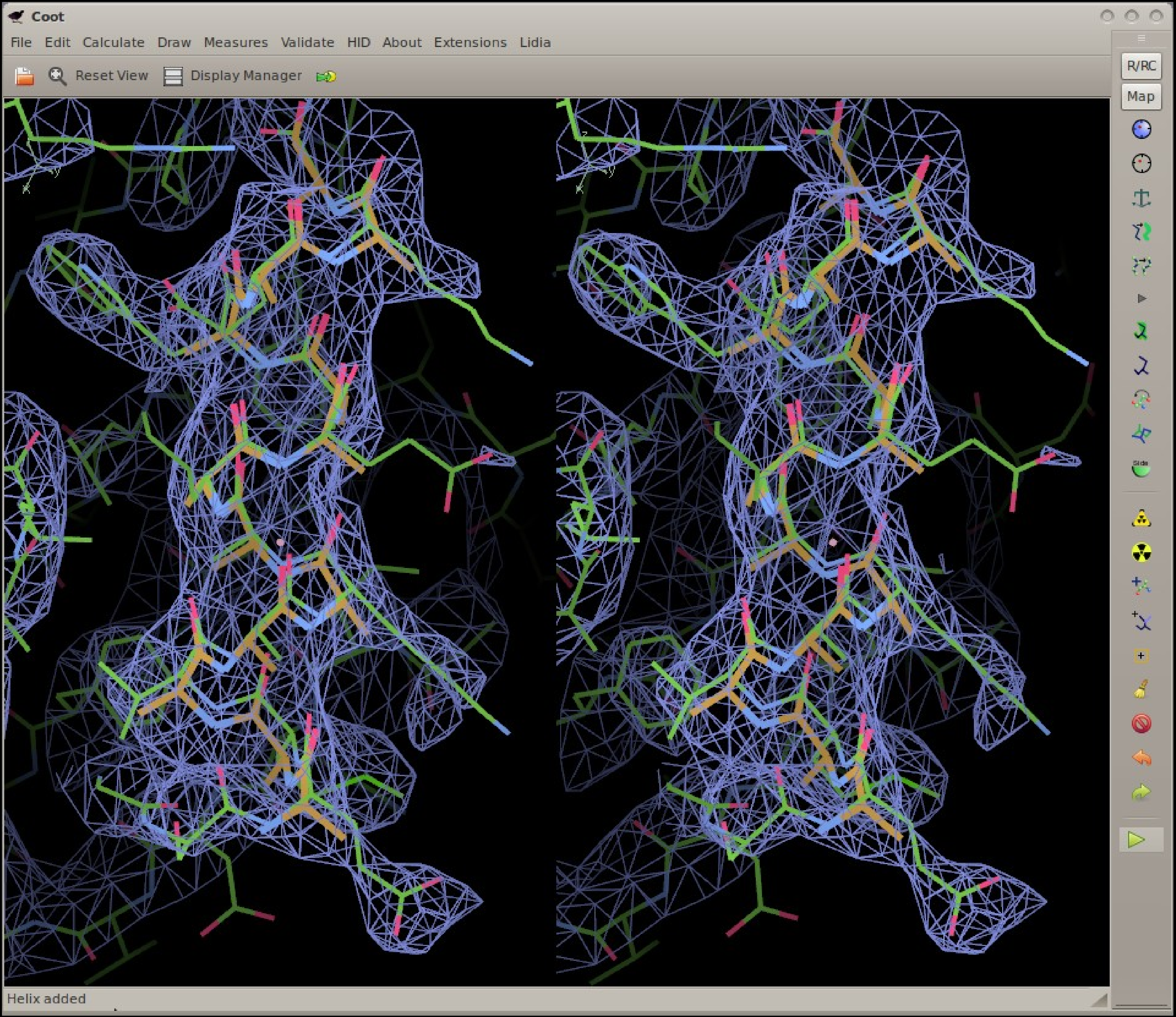

Top
baubles
Bottom


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

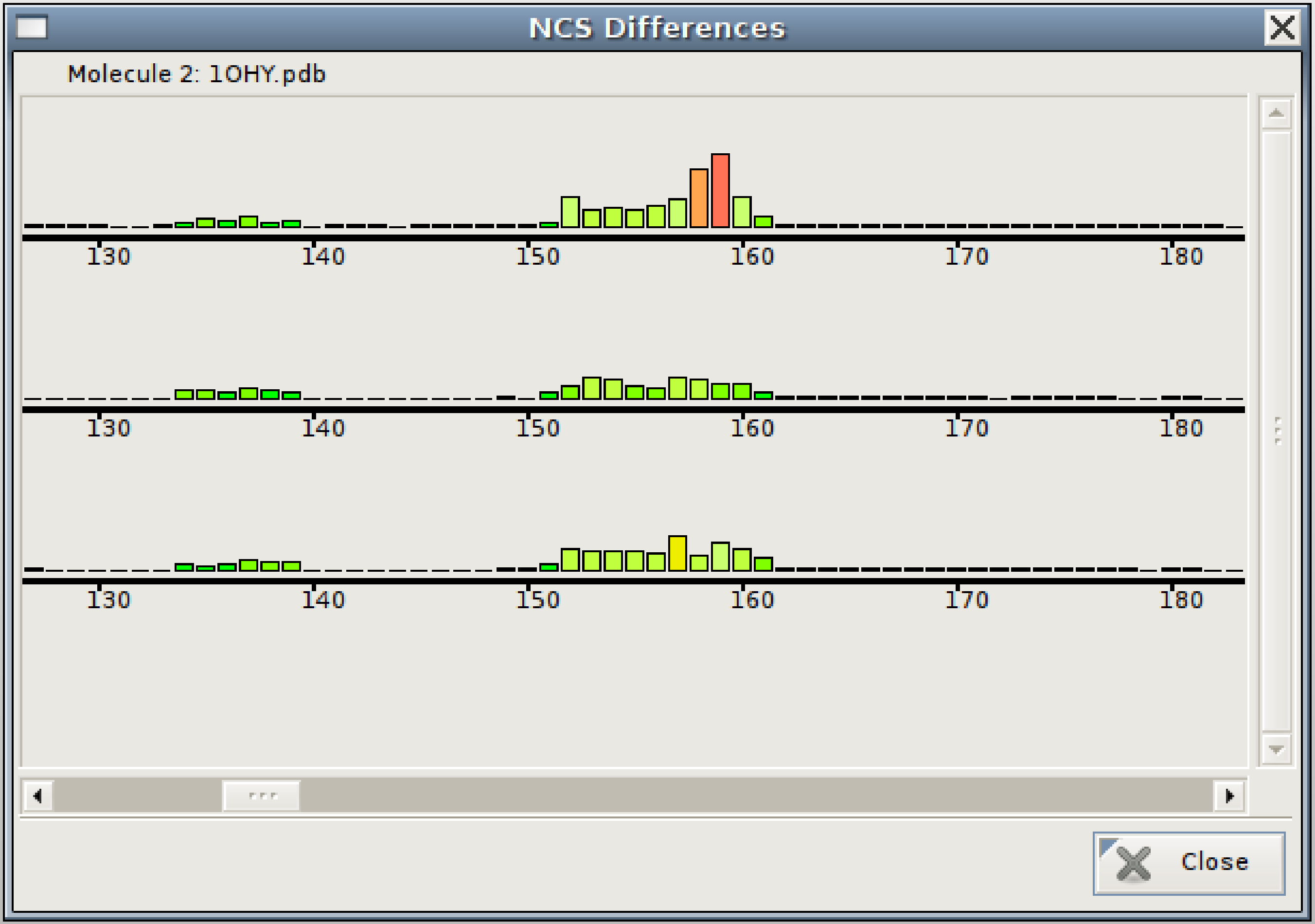
Handling NCS…

Handling NCS

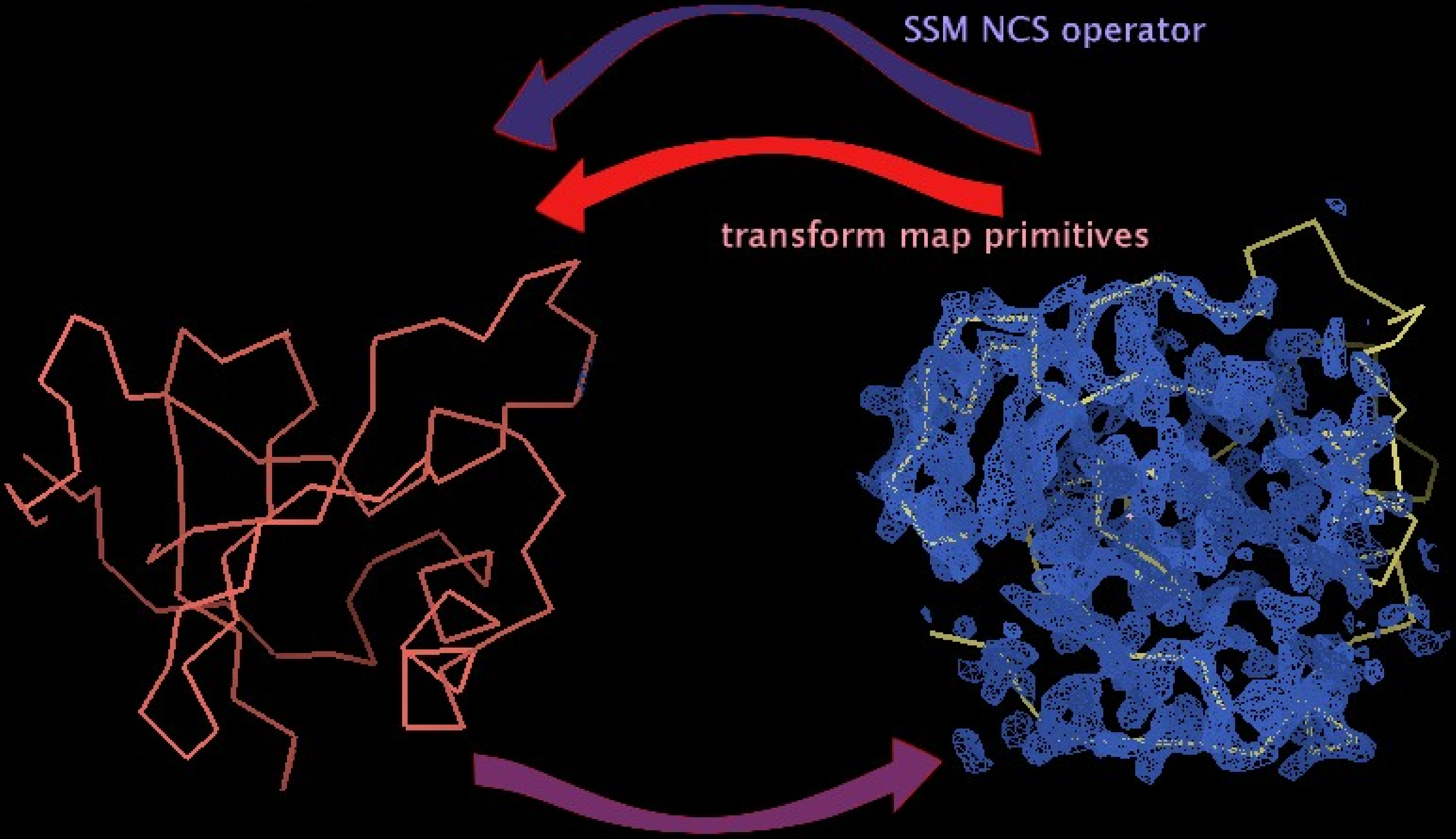
Typical Scenario:

- 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



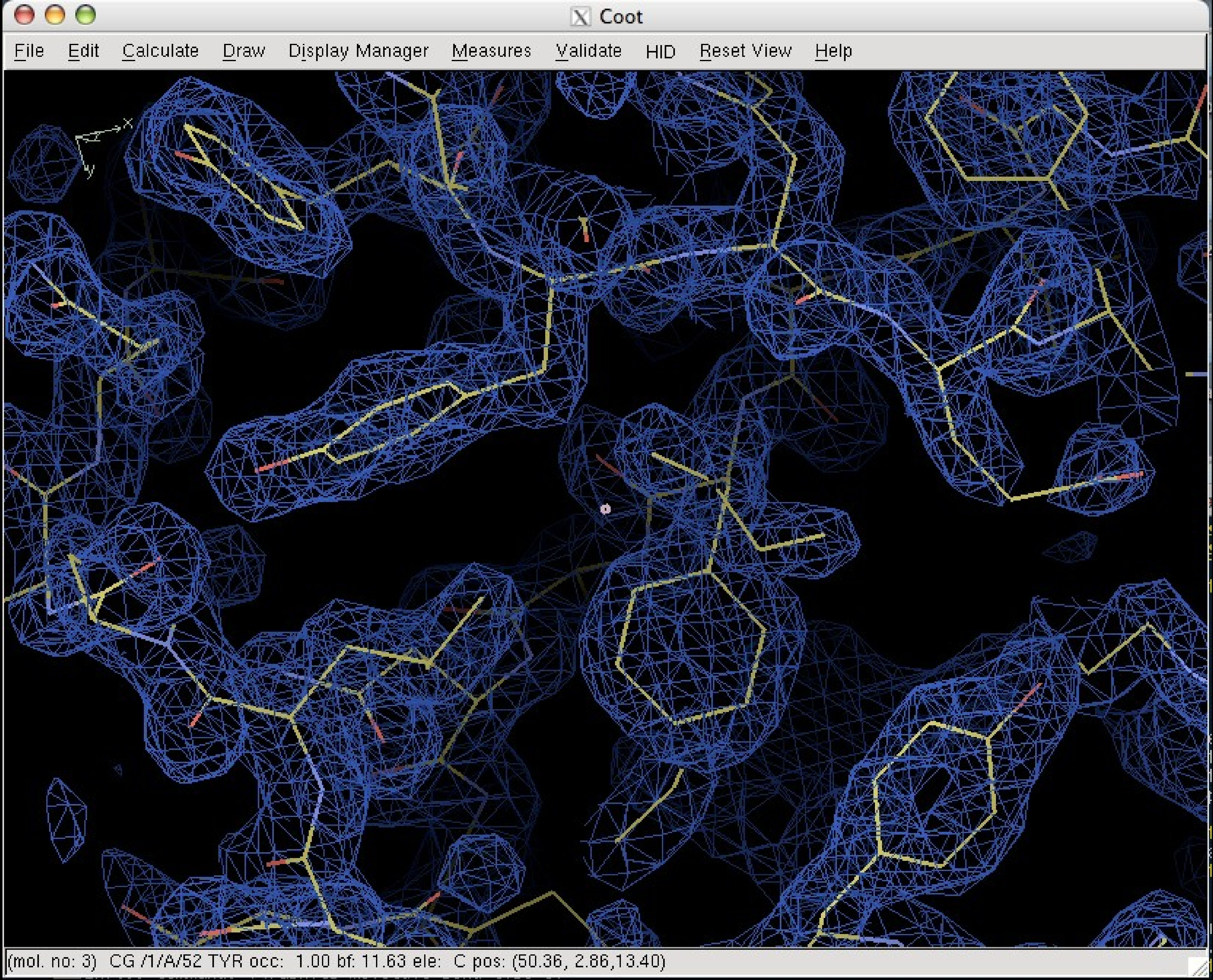
NCS Overlays



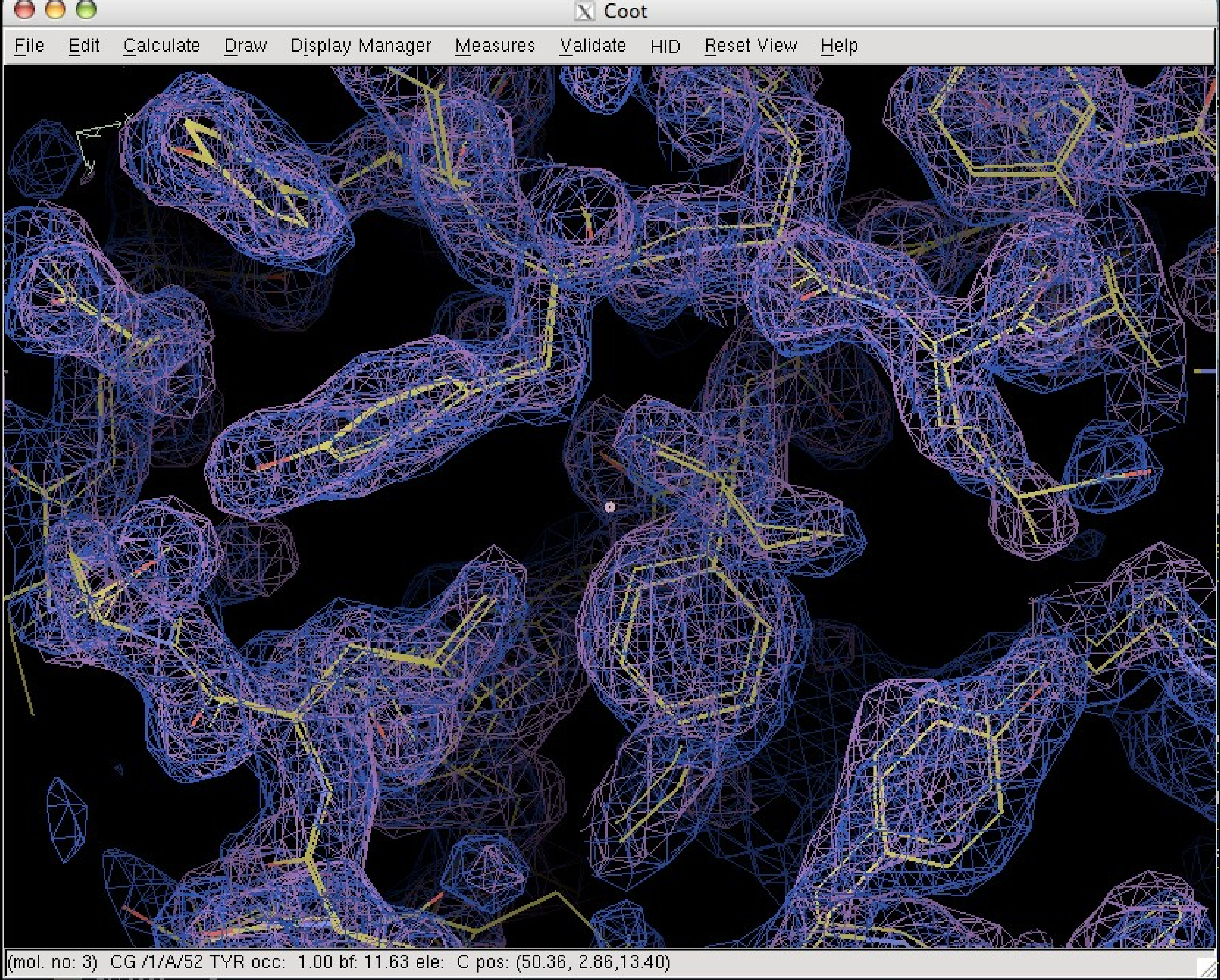
SSM NCS operator

transform map primitives

map centre



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

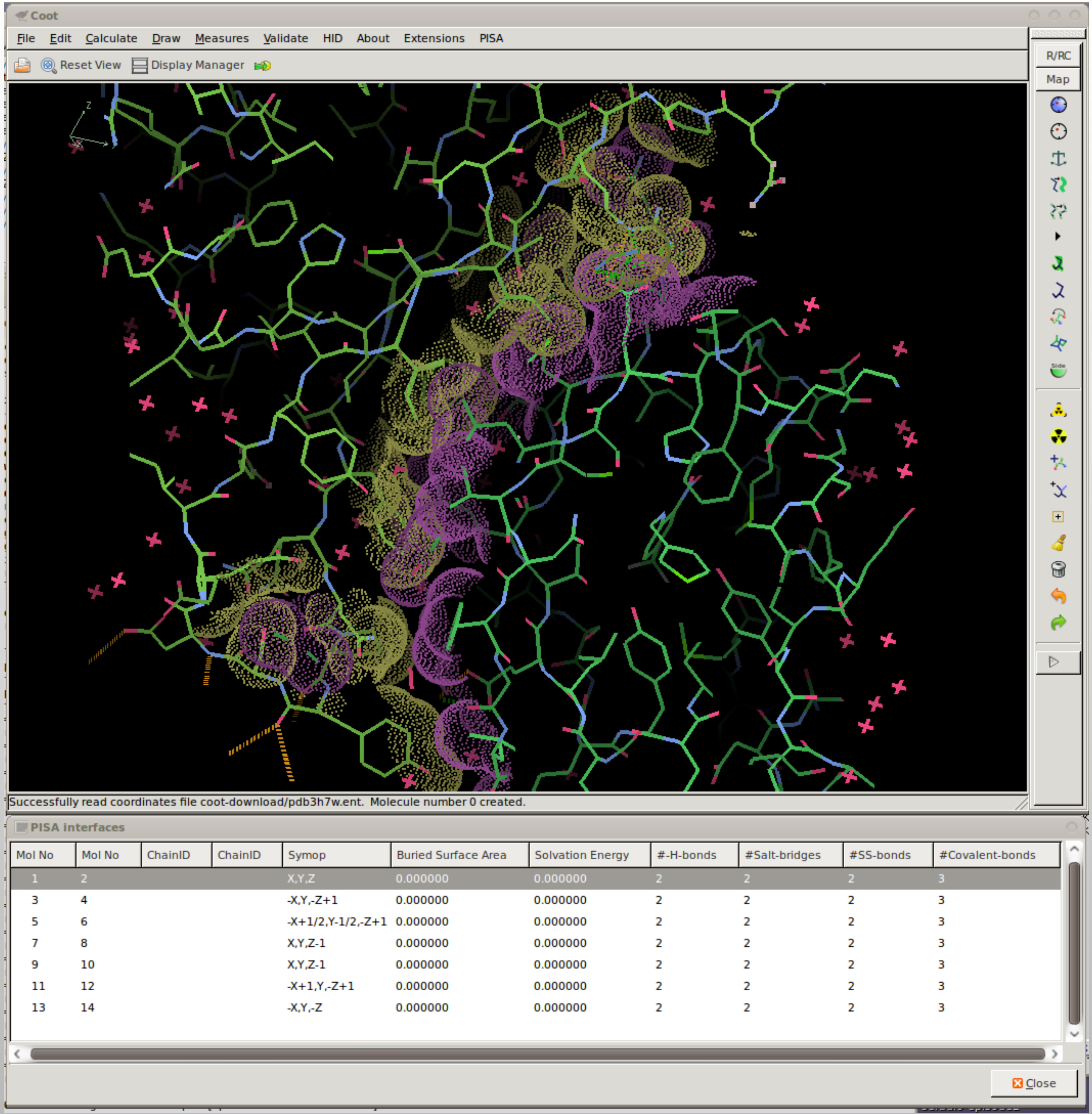


```
(mol. no: 3)  CG /1/A/52 TYR occ:  1.00 bf: 11.63 ele:  C pos: (50.36, 2.86,13.40)
```

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

Interfaces and Assemblies: Interface to PISA

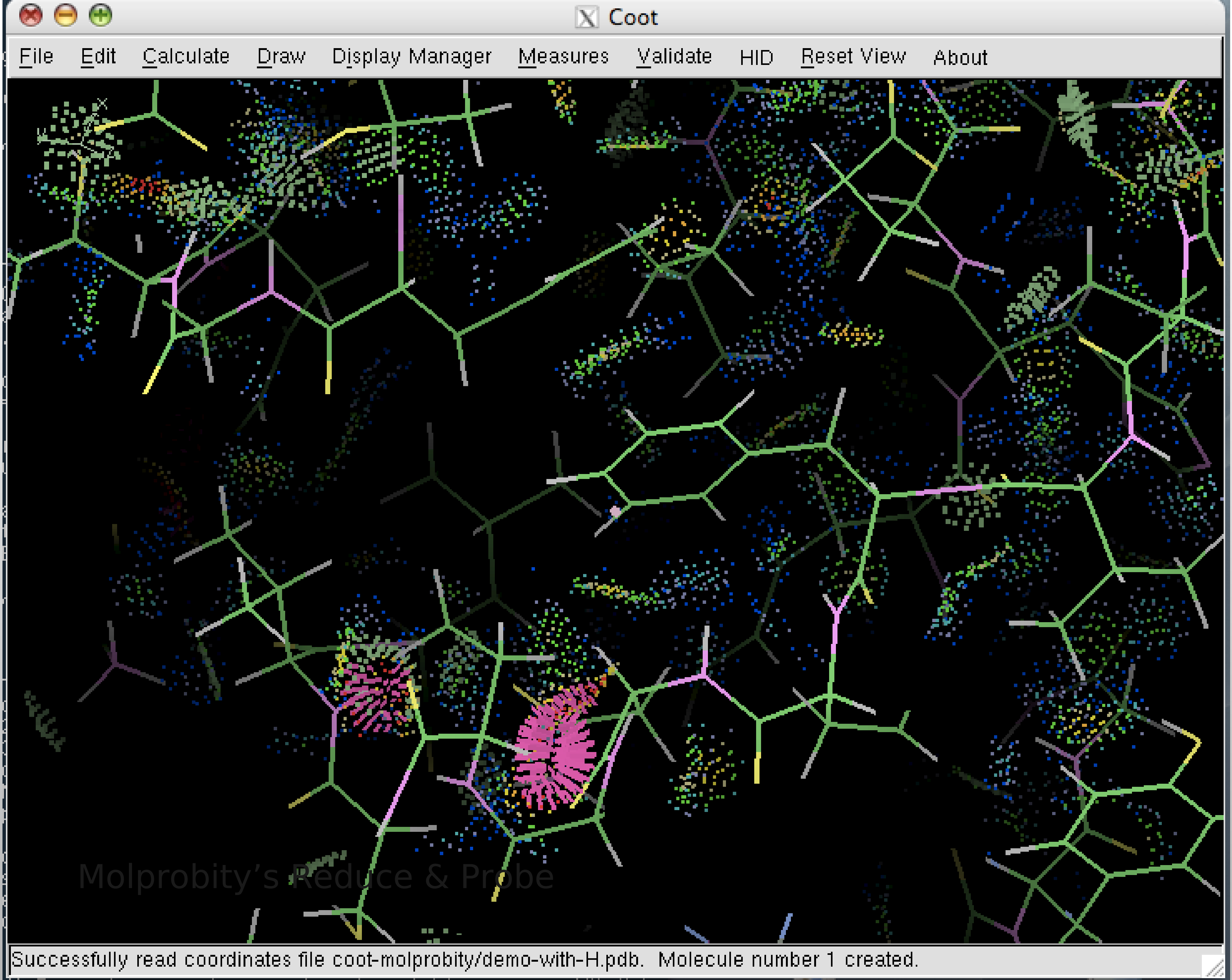


Successfully read coordinates file coot-download/pdb3h7w.ent. Molecule number 0 created.

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

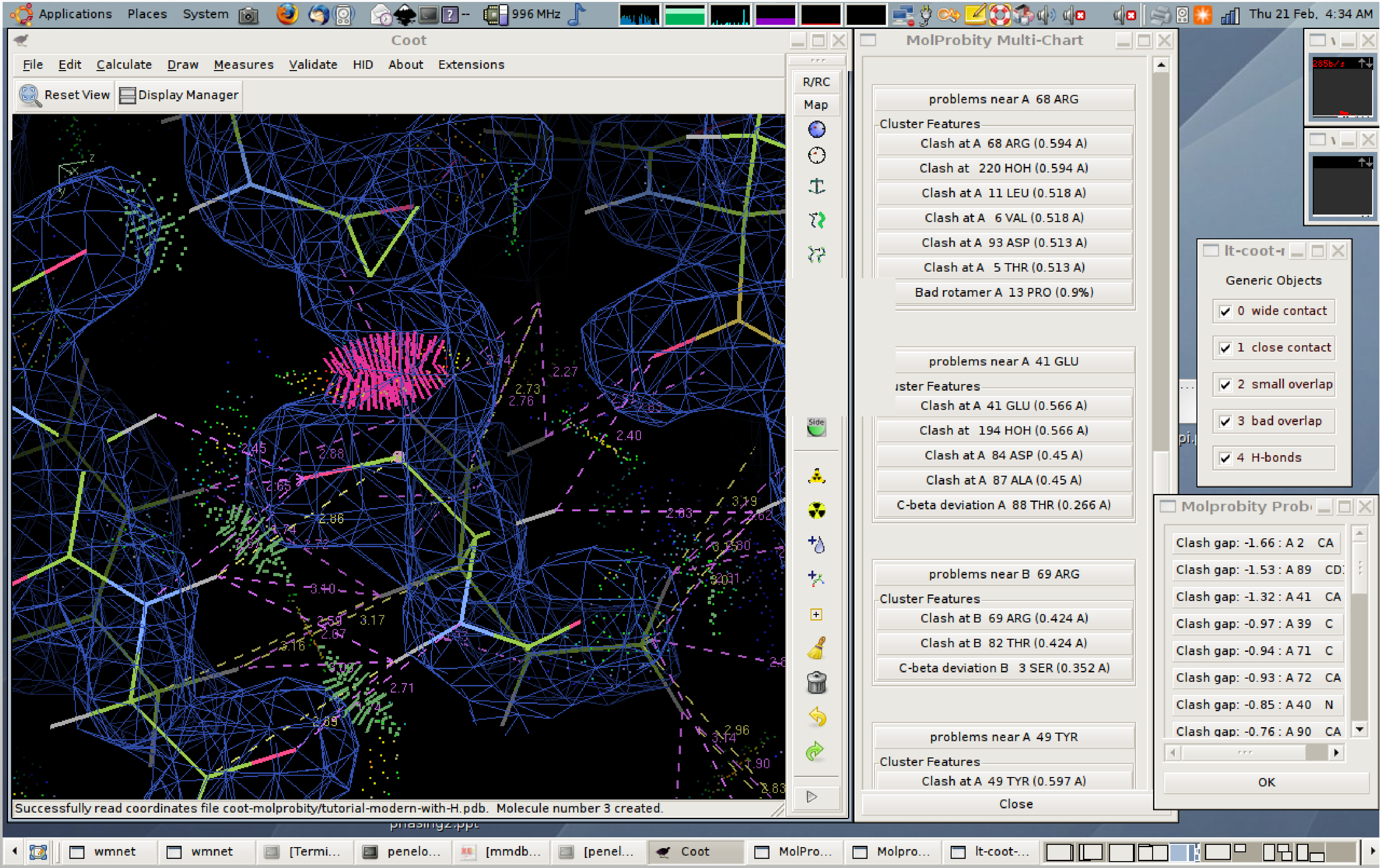
Validation

- My preferred tools:
    - Difference Map Peaks
    - Probe clashes



Molprobity's Reduce & Probe

Successfully read coordinates file coot-molprobity/demo-with-H.pdb. Molecule number 1 created.



problems near A 68 ARG

Cluster Features

Clash at A 68 ARG (0.594 A)

Clash at 220 HOH (0.594 A)

Clash at A 11 LEU (0.518 A)

Clash at A 6 VAL (0.518 A)

Clash at A 93 ASP (0.513 A)

Clash at A 5 THR (0.513 A)

Bad rotamer A 13 PRO (0.9%)

problems near A 41 GLU

ıster Features

Clash at A 41 GLU (0.566 A)

Clash at 194 HOH (0.566 A)

Clash at A 84 ASP (0.45 A)

Clash at A 87 ALA (0.45 A)

C-beta deviation A 88 THR (0.266 A)

problems near B 69 ARG

Cluster Features

Clash at B 69 ARG (0.424 A)

Clash at B 82 THR (0.424 A)

C-beta deviation B 3 SER (0.352 A)

problems near A 49 TYR

Cluster Features

Clash at A 49 TYR (0.597 A)

Generic Objects

- ☑ 0 wide contact
- ☑ 1 close contact
- ☑ 2 small overlap
- ☑ 3 bad overlap
- ☑ 4 H-bonds

Clash gap: -1.66 : A 2 CA

Clash gap: -1.53 : A 89 CD

Clash gap: -1.32 : A 41 CA

Clash gap: -0.97 : A 39 C

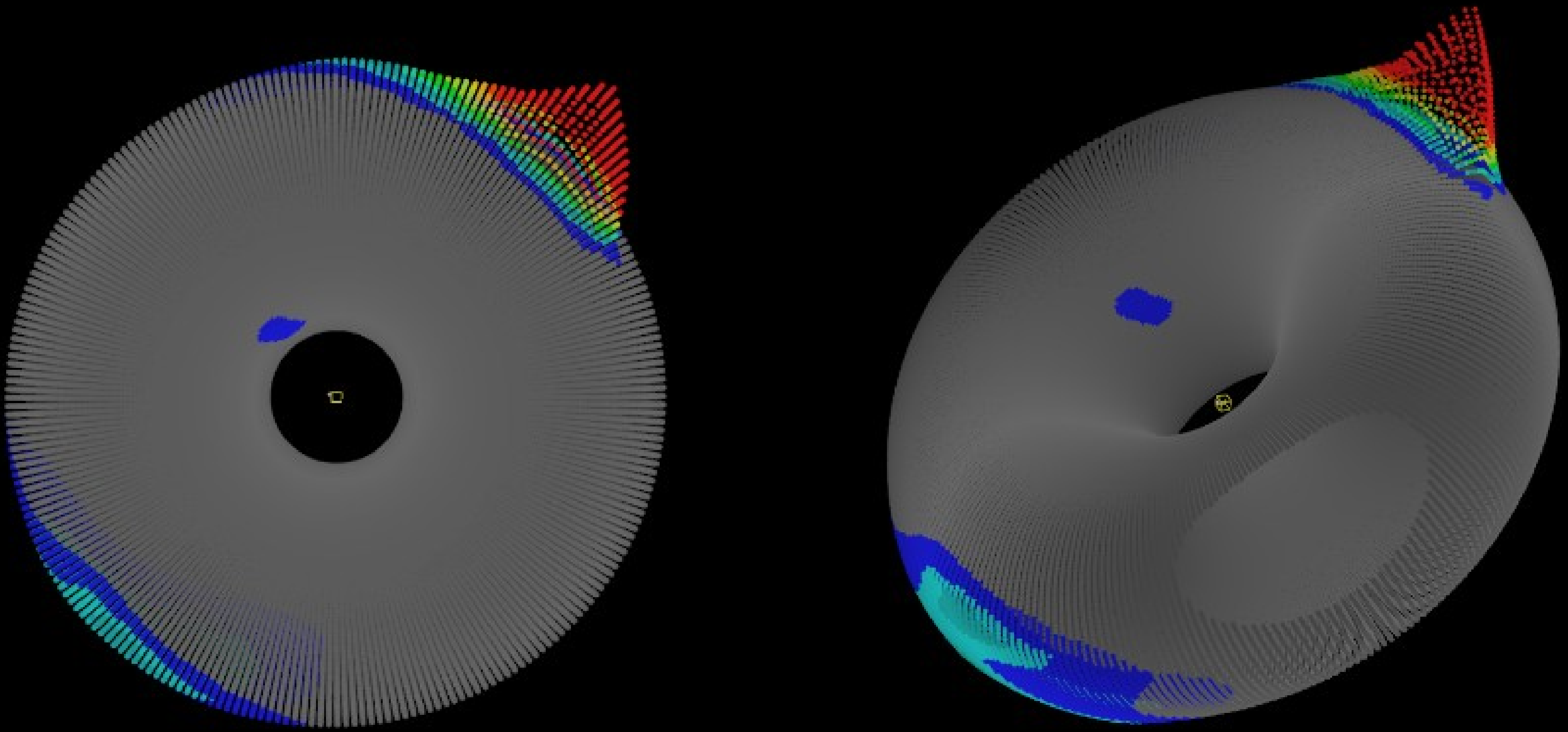
Clash gap: -0.94 : A 71 C

Clash gap: -0.93 : A 72 CA

Clash gap: -0.85 : A 40 N

Clash gap: -0.76 : A 90 CA

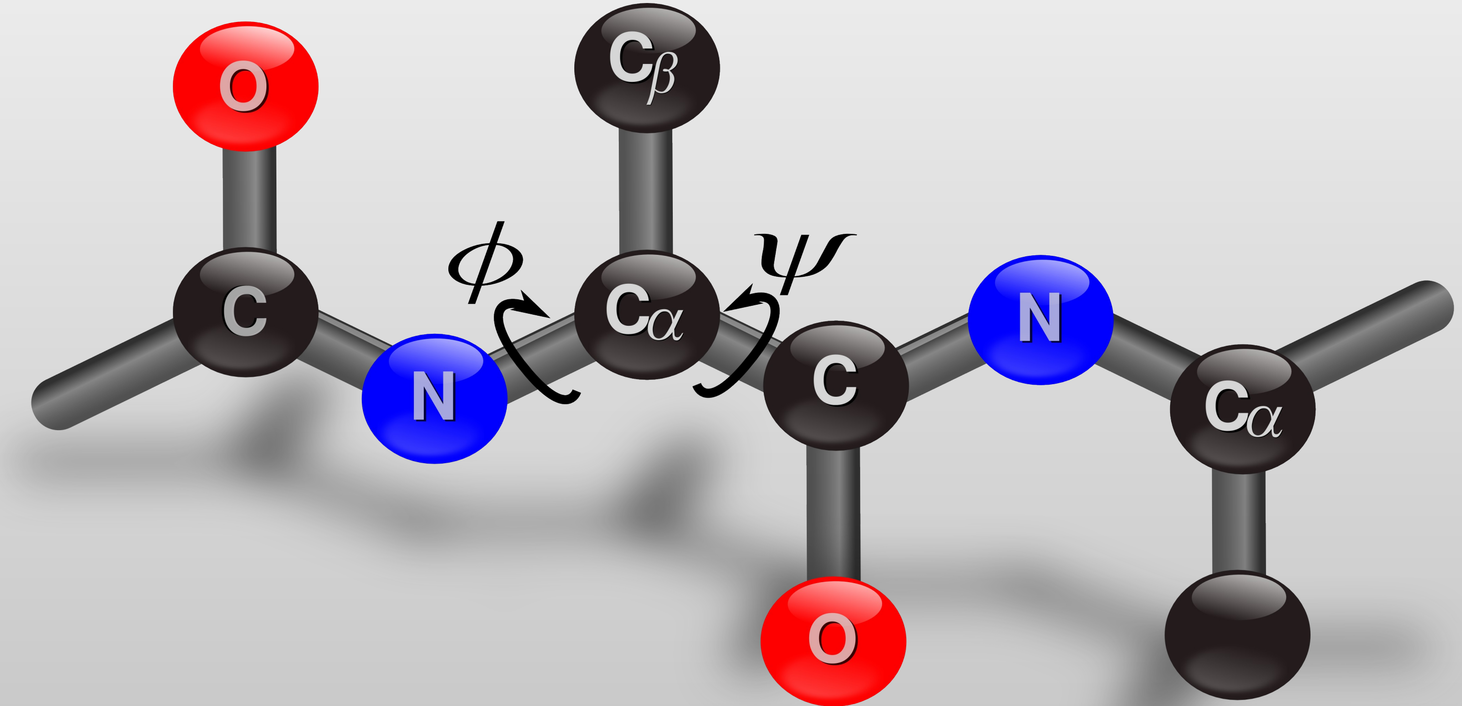
Successfully read coordinates file coot-molprobity/tutorial-modern-with-H.pdb. Molecule number 3 created.



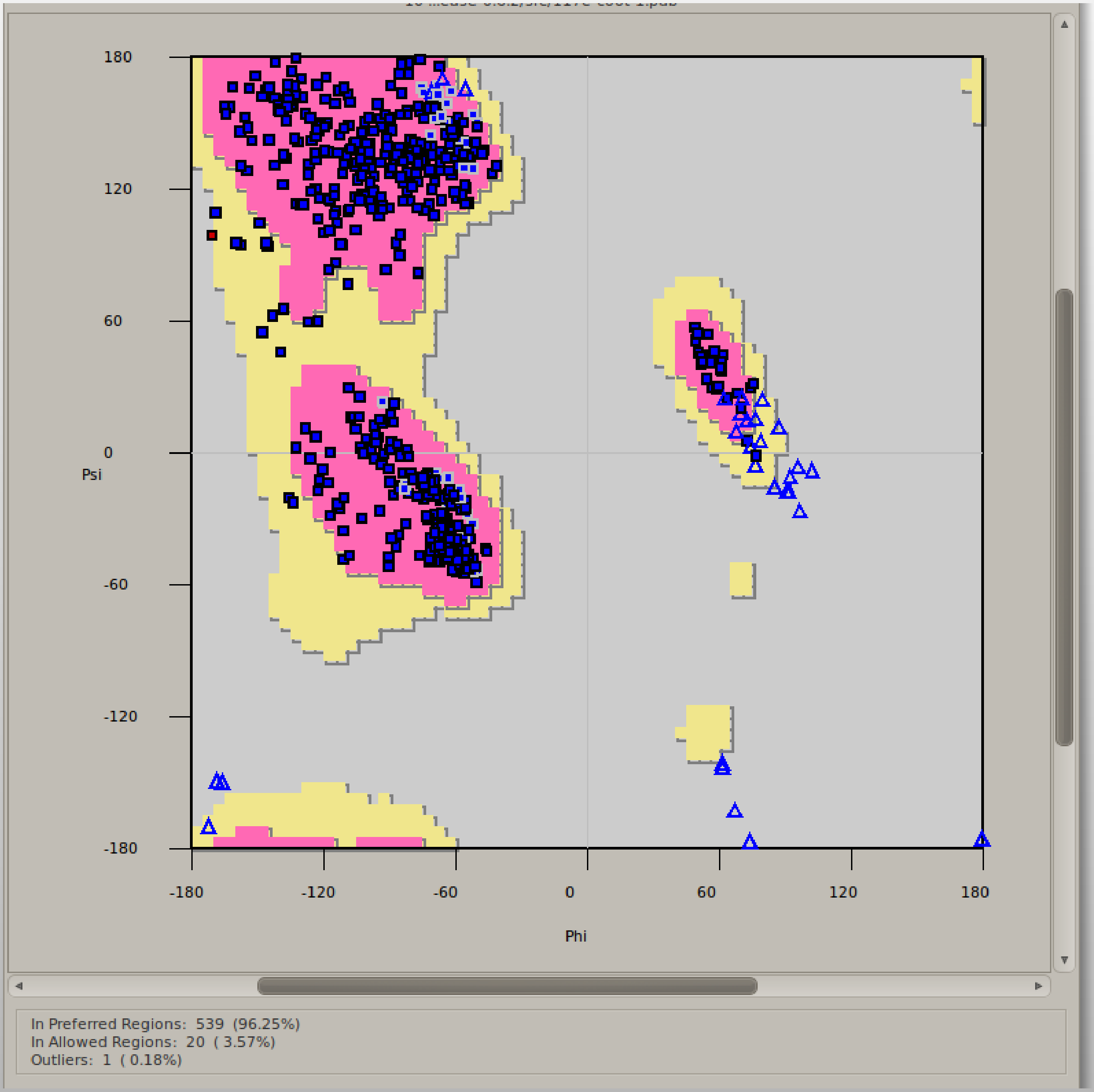
Projection from the surface of a doughnut: 2 x 360°

(linear scaling)

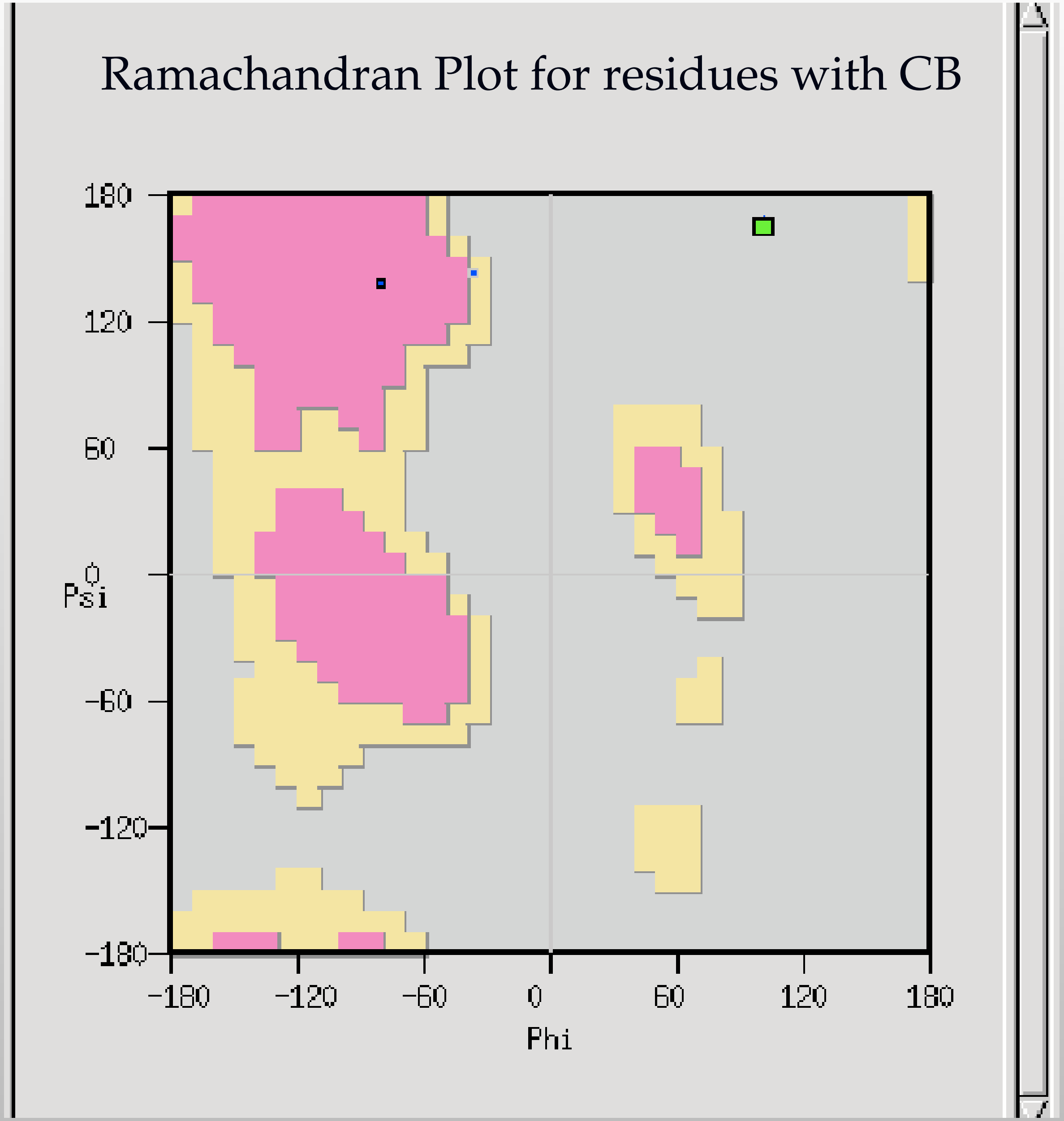
Peptide Backbone Geometry



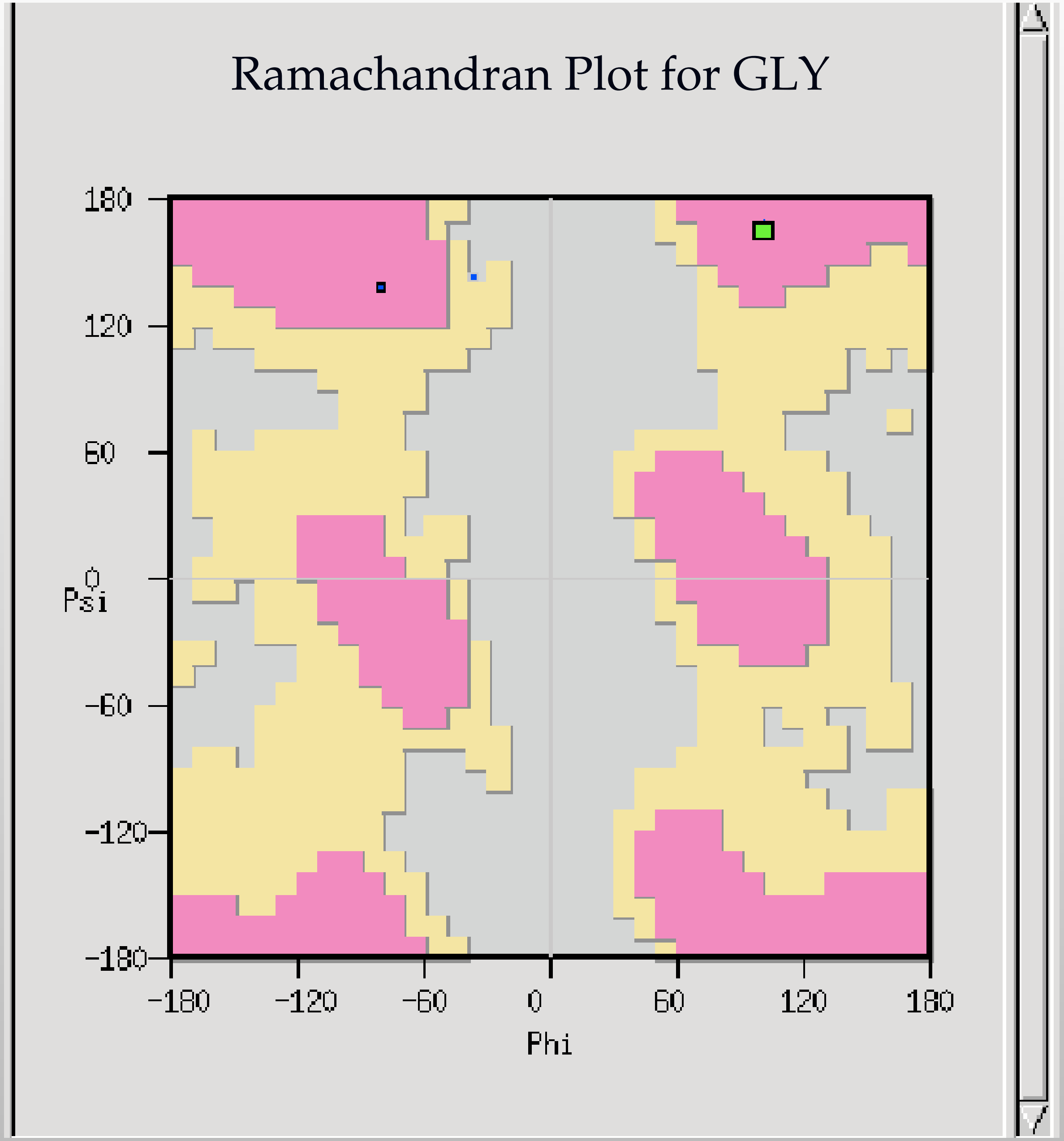
Typical 2D Projection of Ramachandran Plot



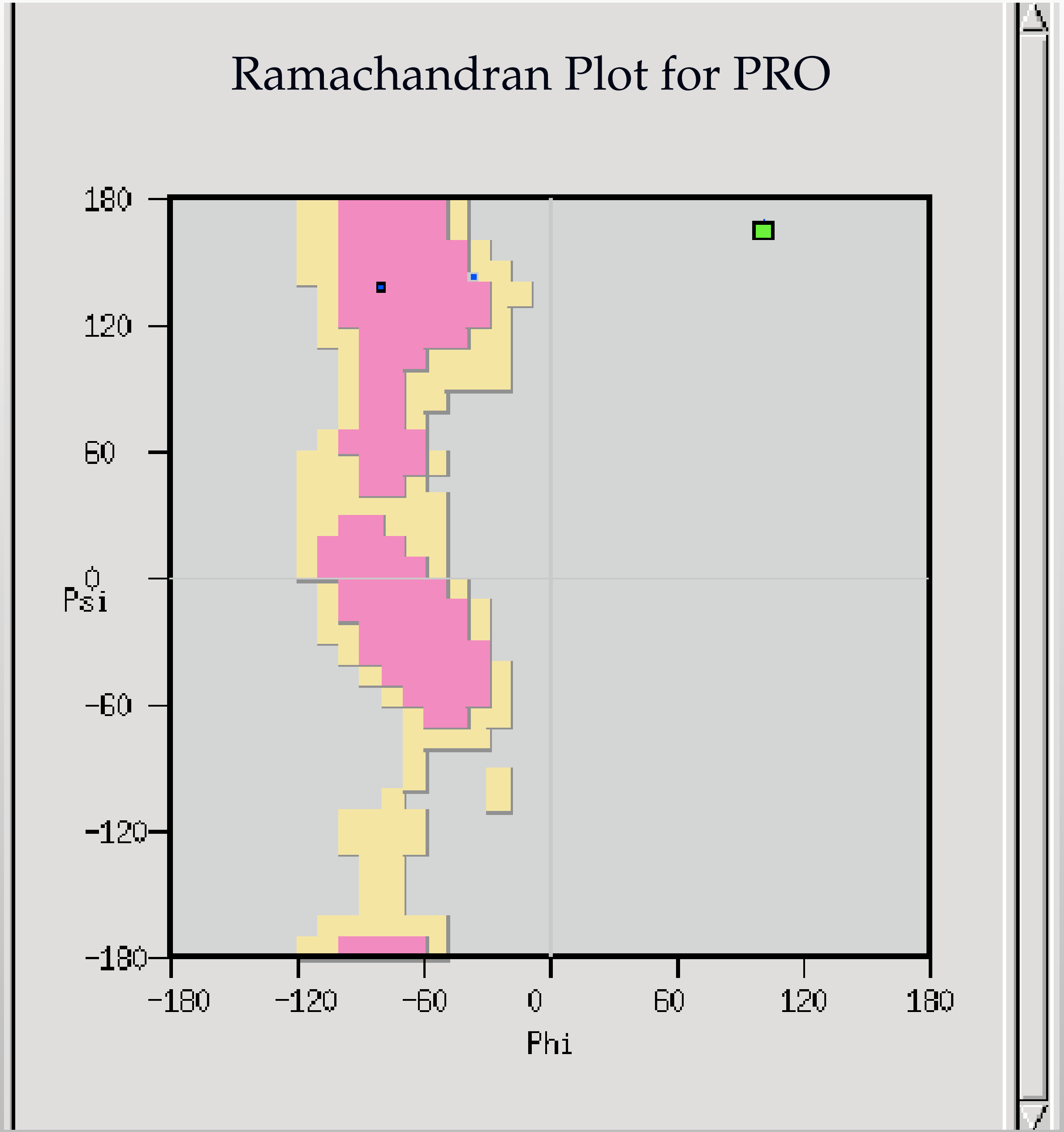
In Preferred Regions: 539 (96.25%)
In Allowed Regions: 20 (3.57%)
Outliers: 1 (0.18%)



Ramachandran Plot for residues with CB



Ramachandran Plot for GLY
180
120
60
0
Psi
-60
-120
-180
-180
-120
-60
0
60
120
180
Phi



Note: PRE-Pro is another distinct distribution, but not yet encoded in Coot

Acknowledgements

- Keith Wilson
- Eleanor Dodson
- Max Crispin

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

- Funding
  - BBSRC, CCP4 & MRC