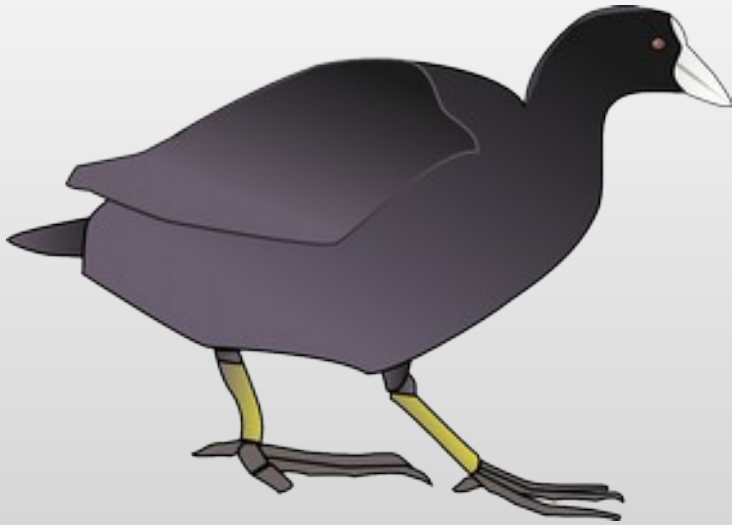


# Model-Building using X-ray data (with *Coot*)

**Paul Emsley,**  
MRC Laboratory of Molecular Biology  
Cambridge, UK

(don't print this out)

# Coot Collaborators



Bernhard  
Lohkamp



Kevin  
Cowtan



Eugene  
Krissinel



Stuart  
McNicholas



Martin  
Noble



Alexei  
Vagin

# A bit of context

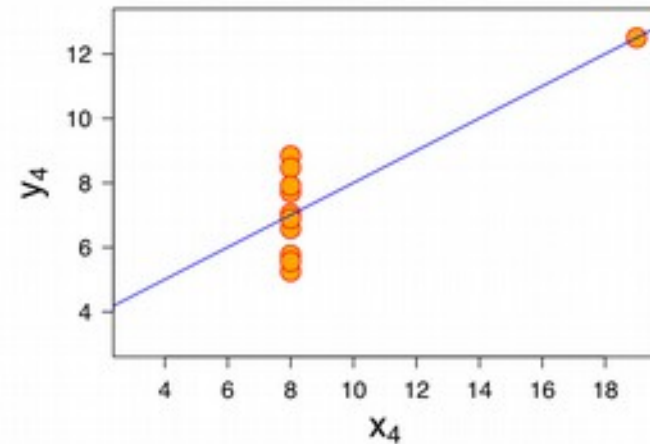
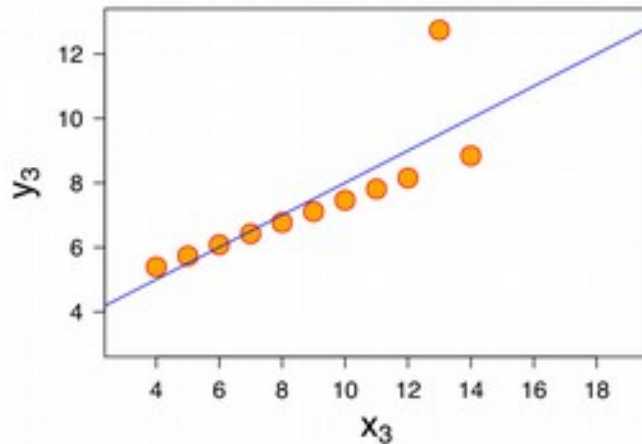
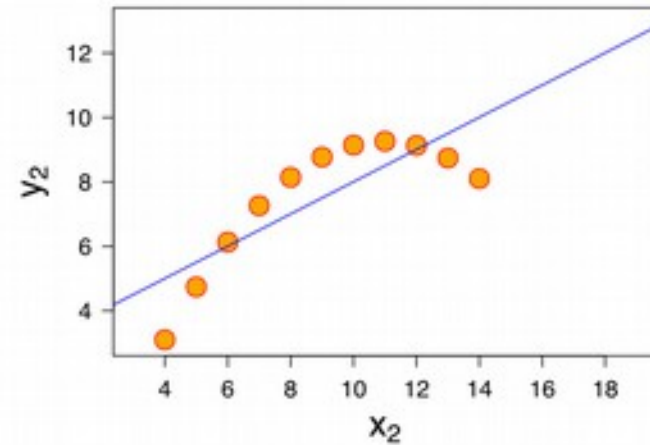
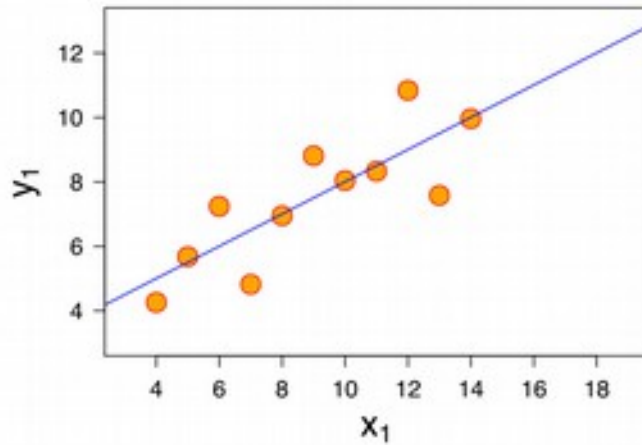
- Why use 3D graphics?

**do we have carbohydrate slides?  
or in other presentation?**

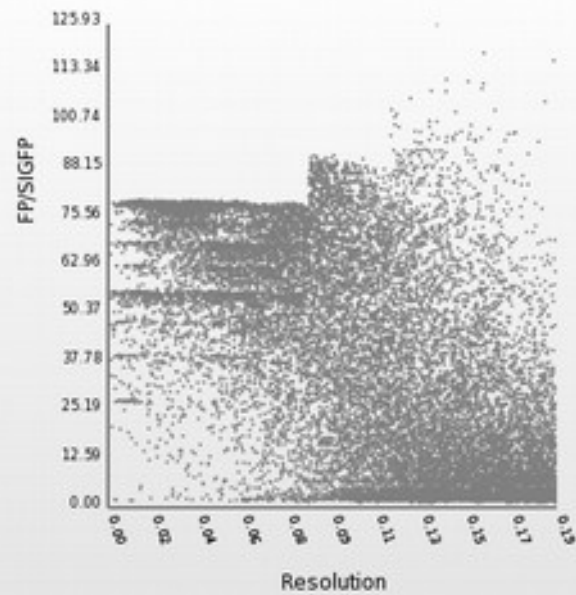
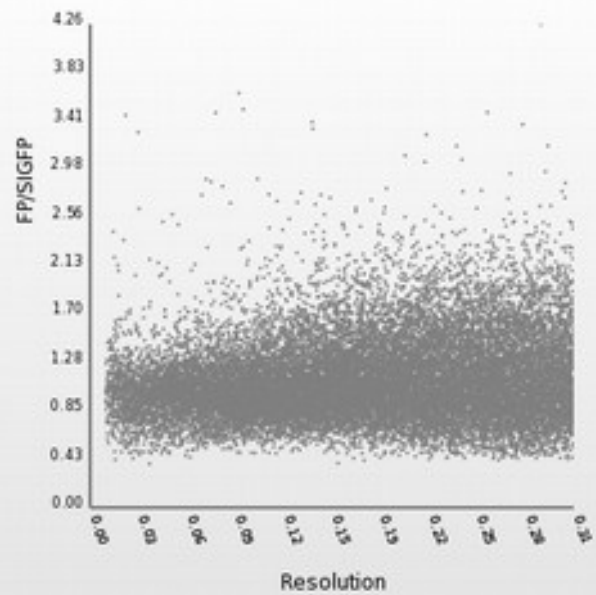
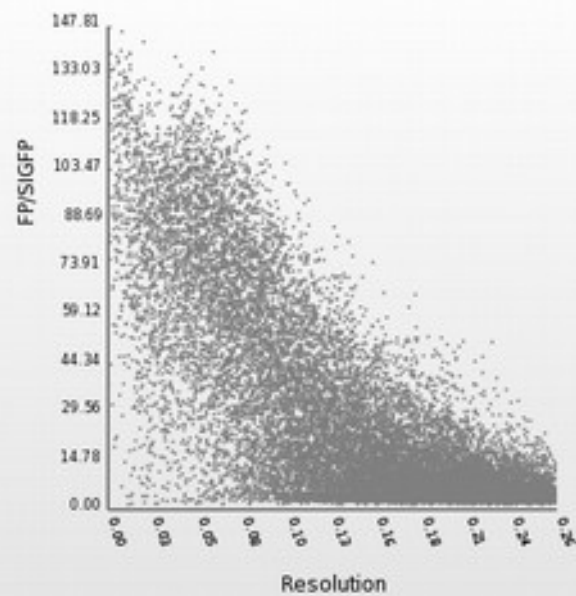
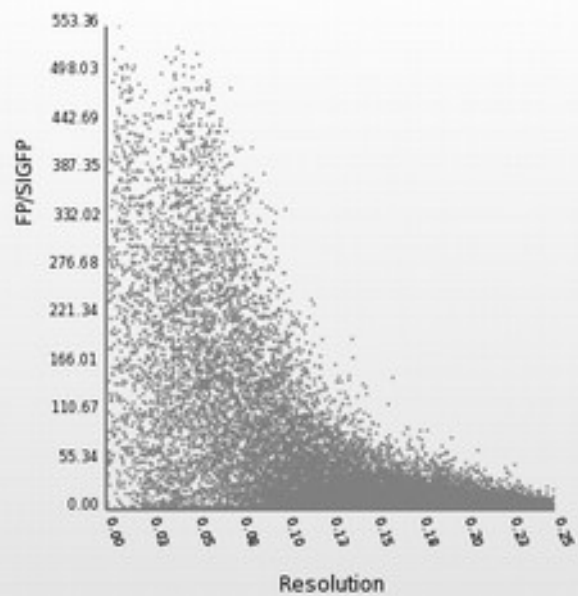
# Summary Statistics

- Are useful, but don't tell the whole story
- Let's say we have 10 data points
  - X mean            9
  - Y mean            7.5
  - correlation        0.816
  - regression         $y = 3 + \frac{1}{2} x$

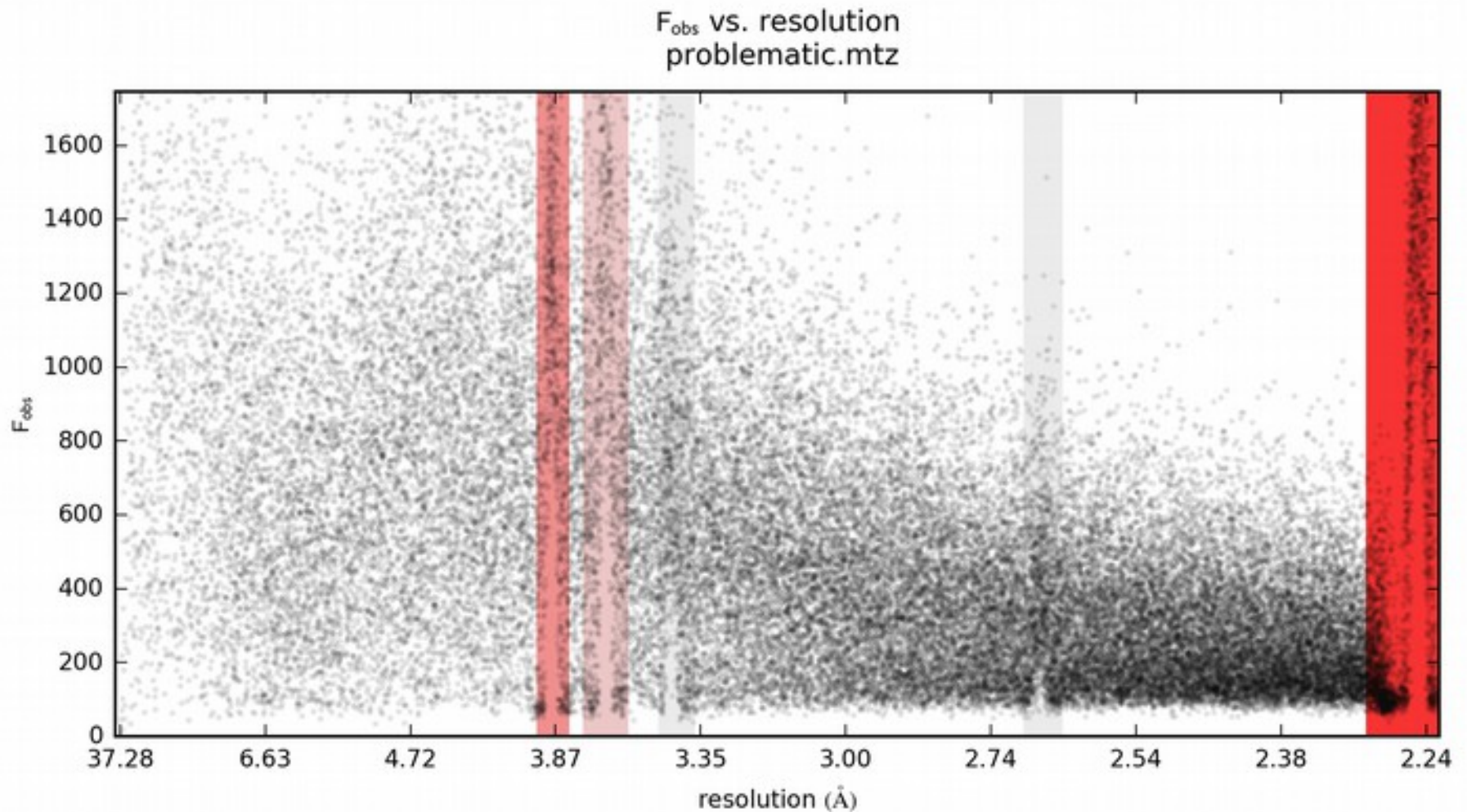
# View Your Data and Model



Anscombe's Quartet



# Auspex: Icefinder





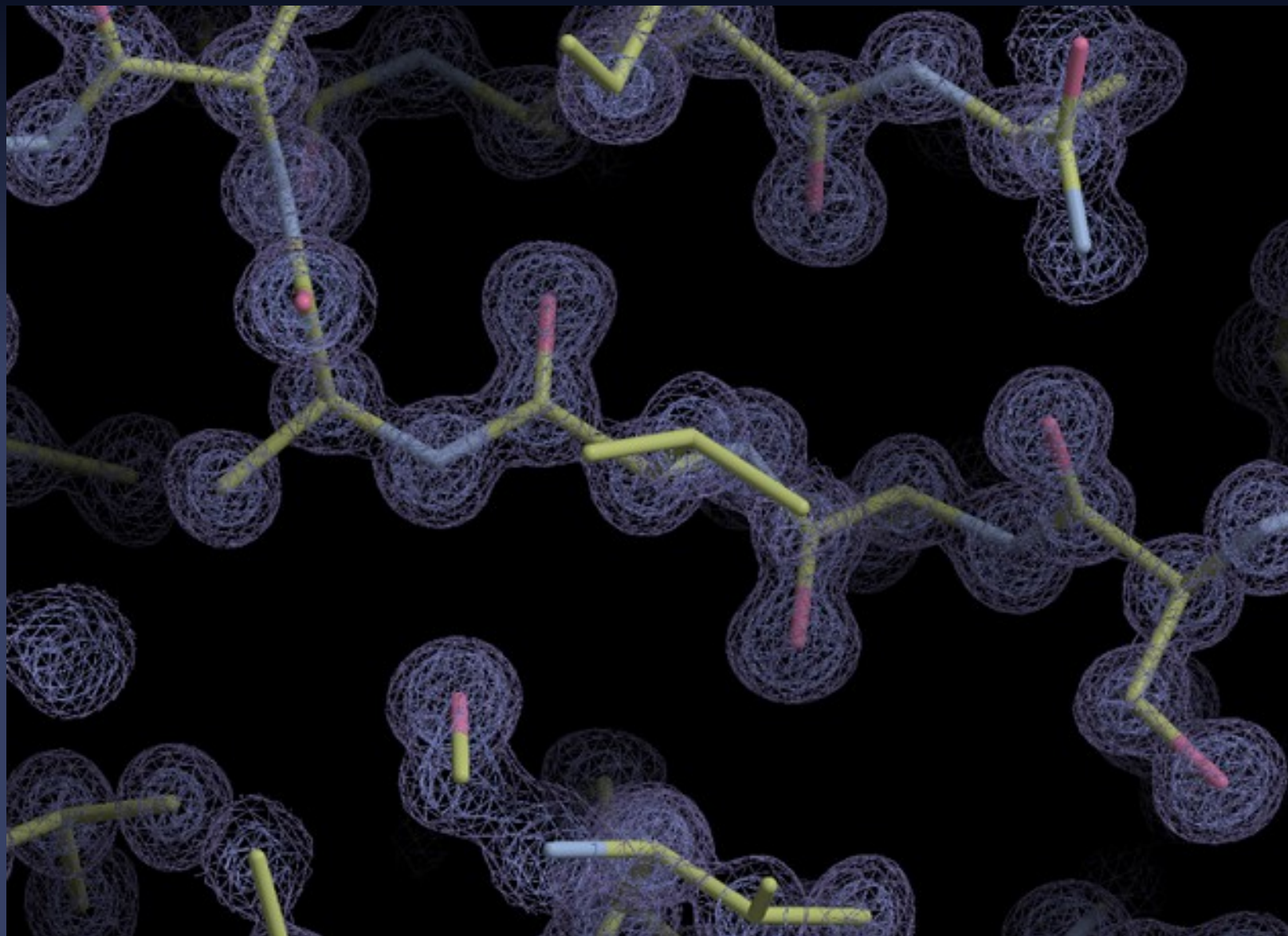
# Coot

- Molecular Graphics application
  - Protein Crystallographic model-building tools
  - Designed to “fill the gap” where automatic methods fail
    - (generally, we don't use molecular graphics programs to do what automatic methods can do)
- Interface to other programs: SHELXL, Refmac, Libcheck, Probe&Reduce (Molprobit), EBI, EDS, Povray... and others

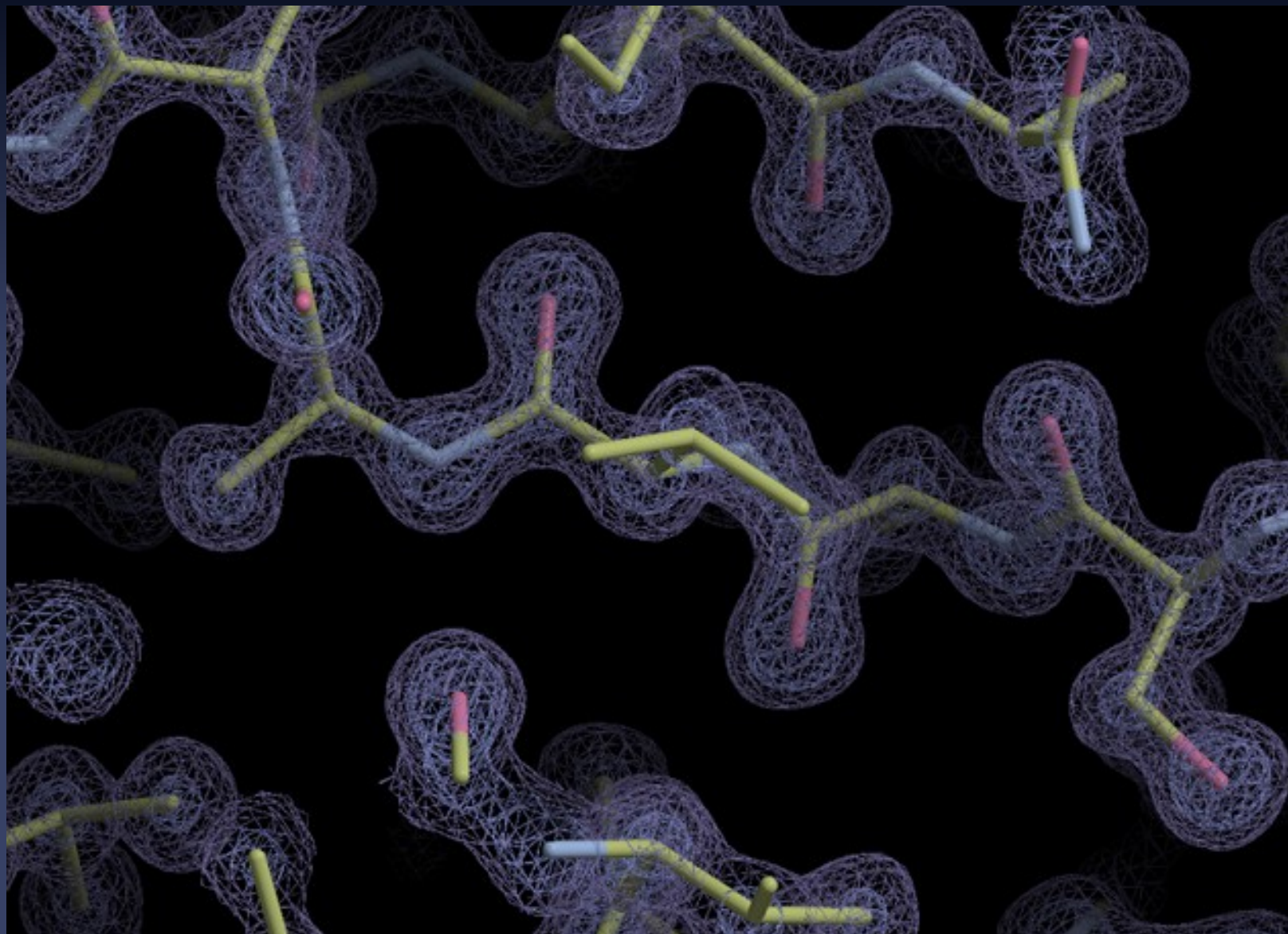
# But Why Bother?

- Automated model-building for complete models is still impossible
  - It takes a brain to validate
- Concerted correction/improvement of a model is difficult on the larger scale

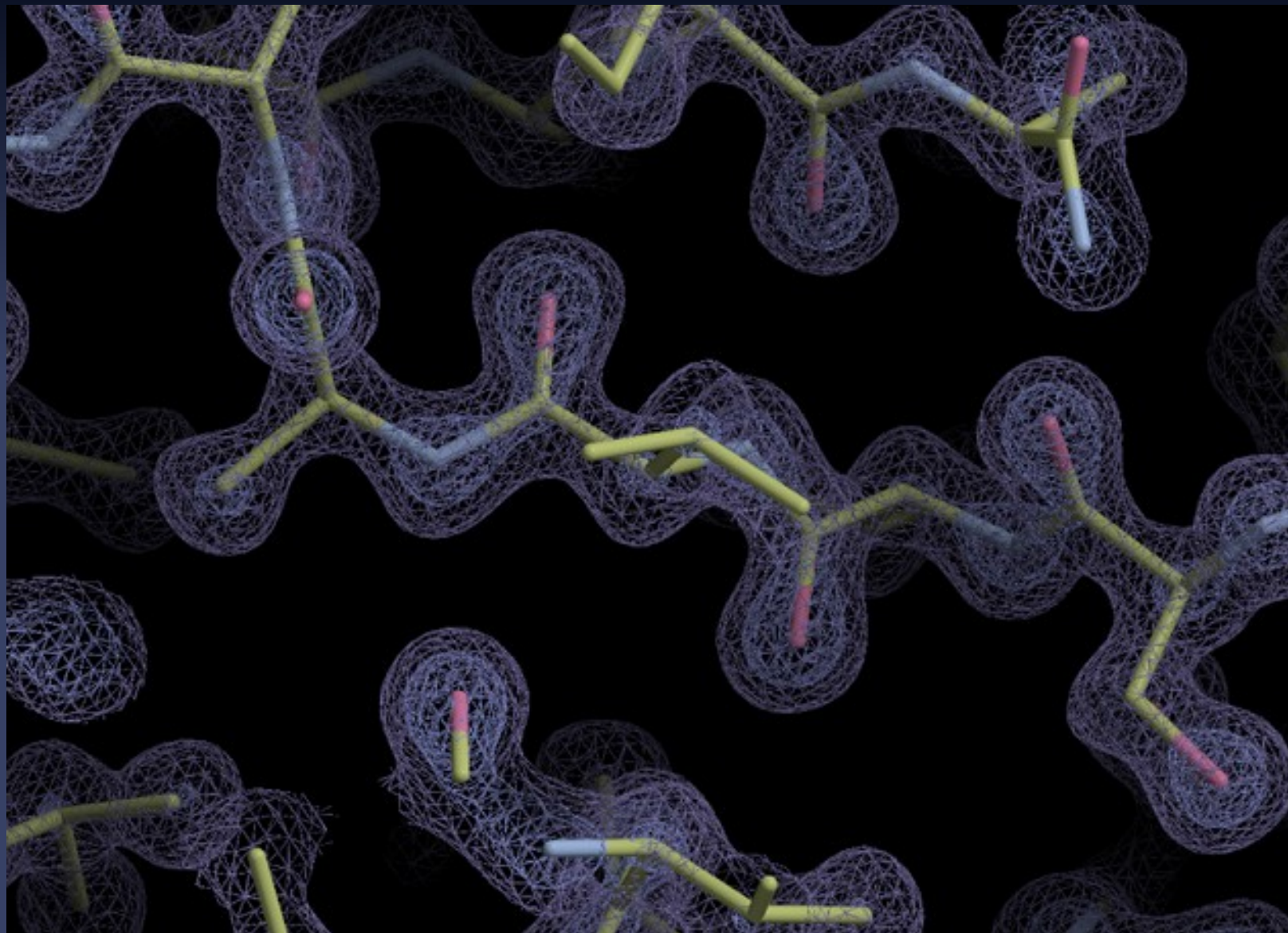
1.0Å



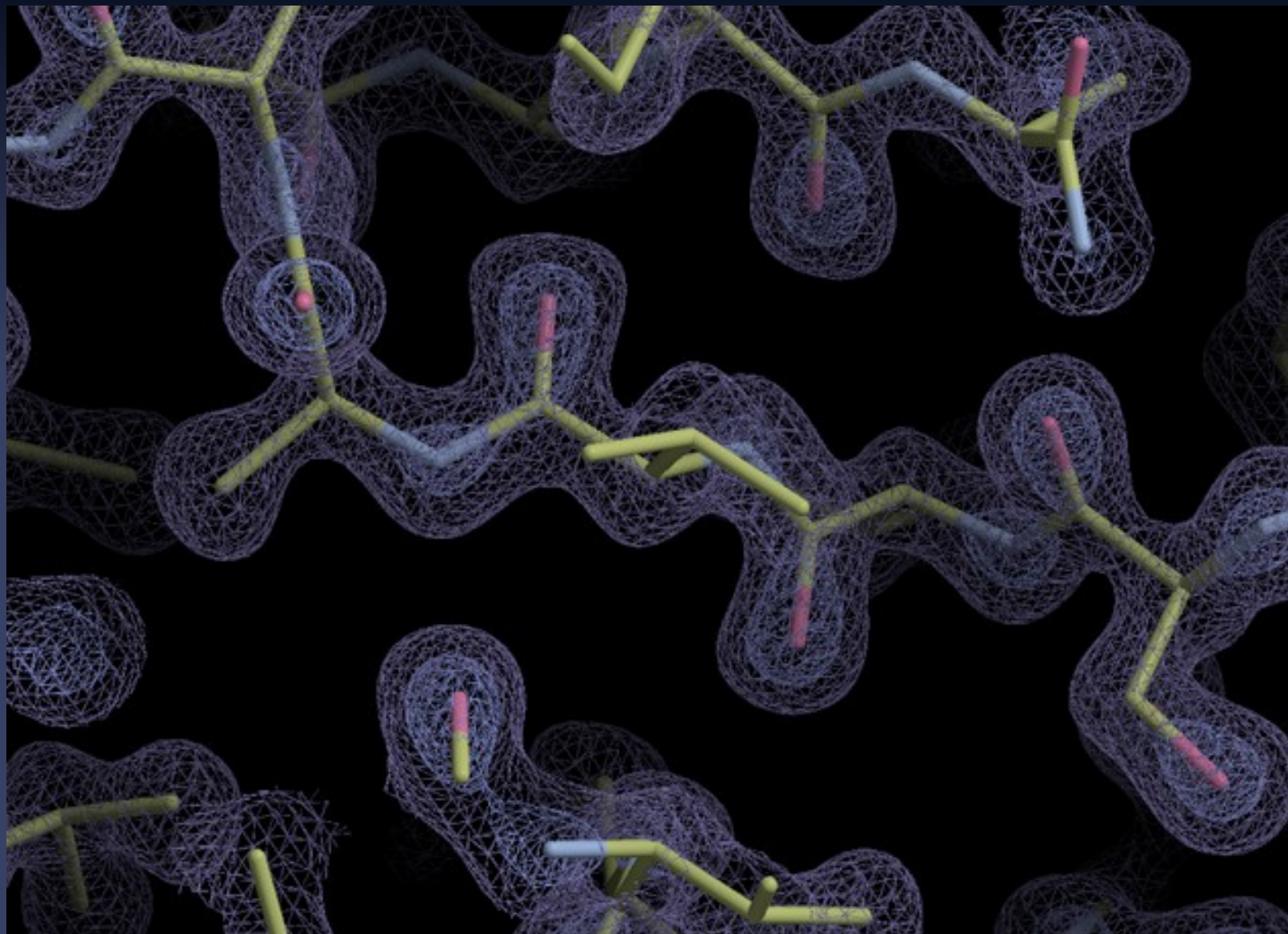
1.2Å



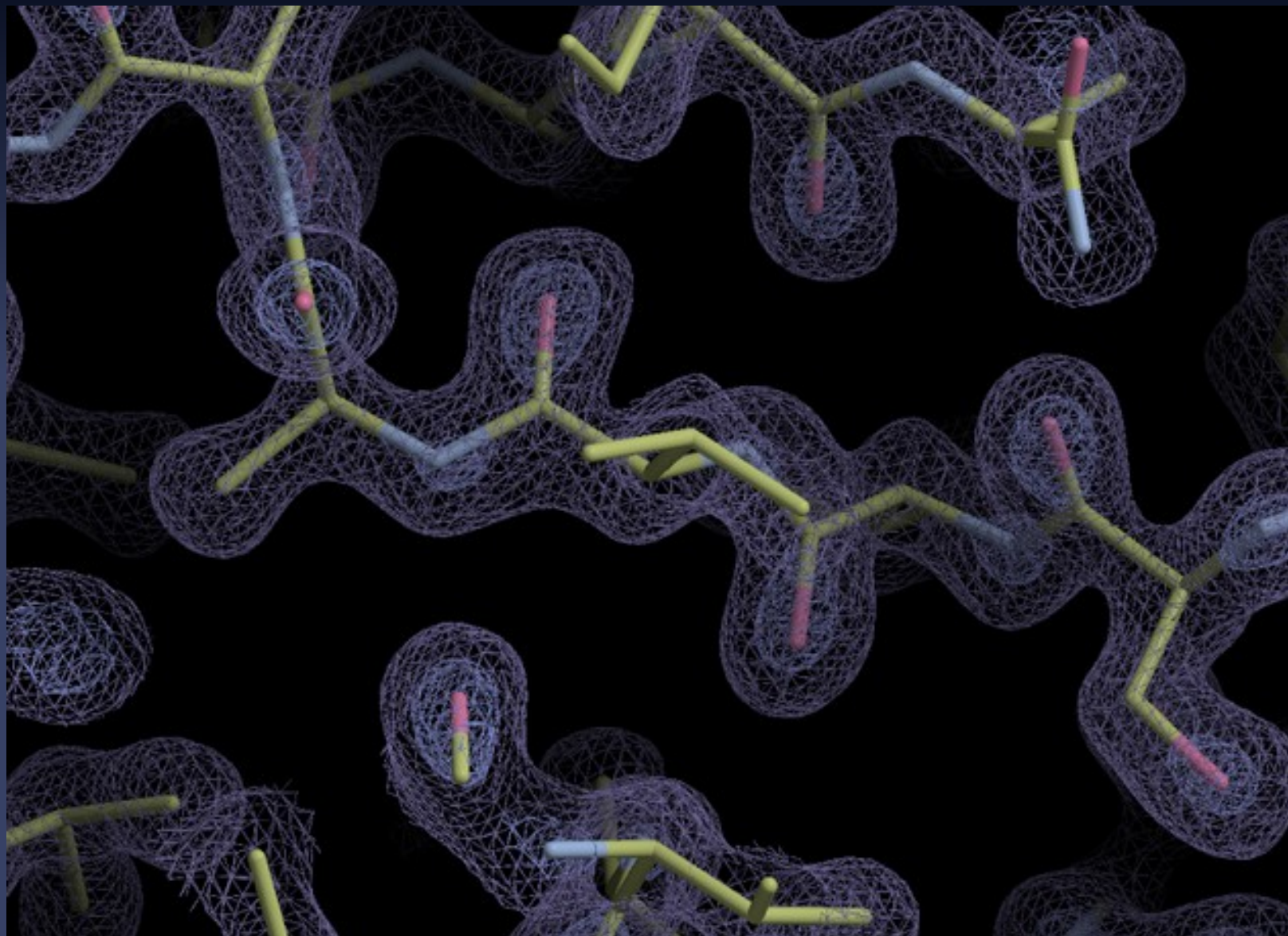
1.4Å



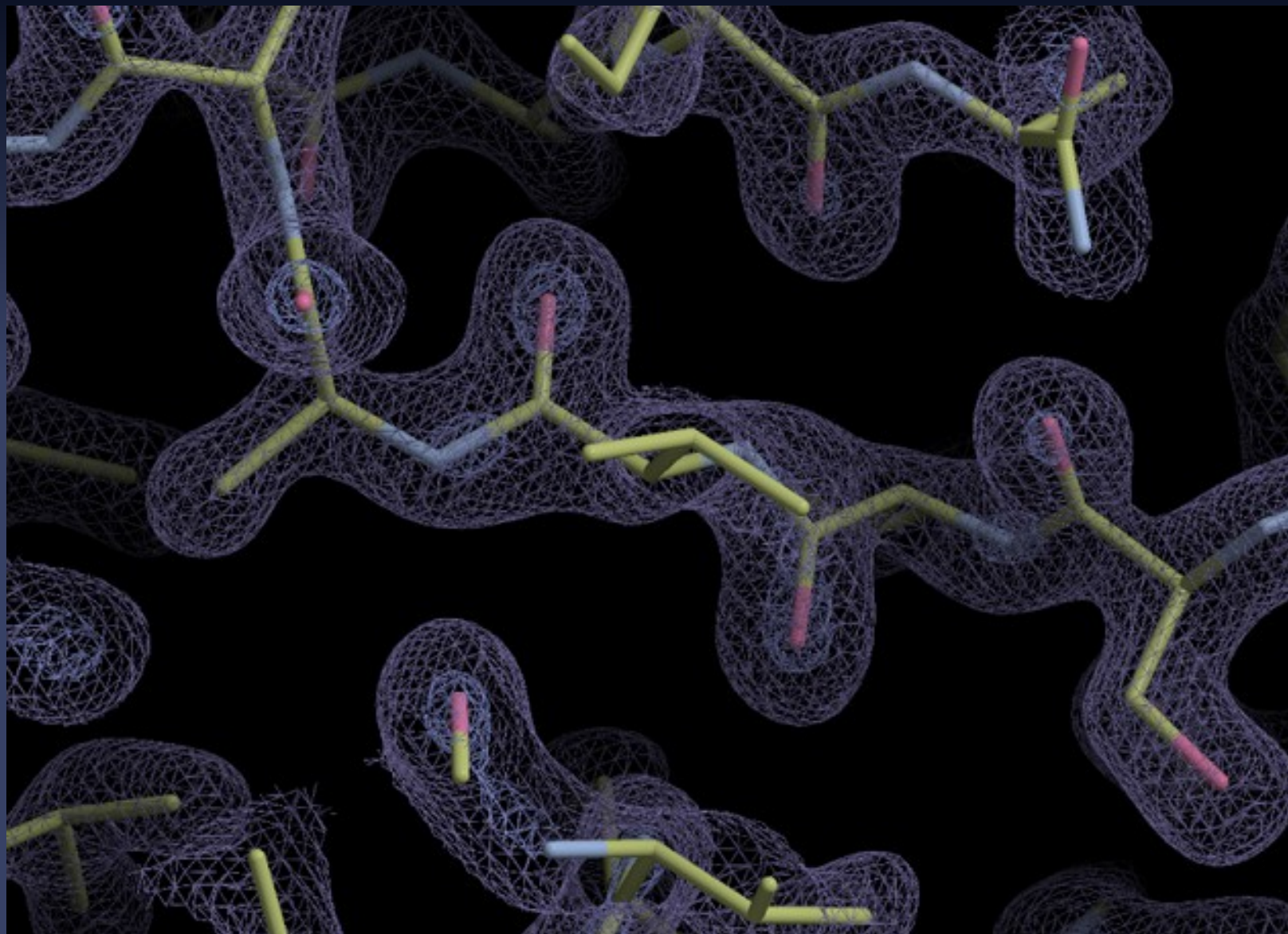
1.6Å



1.8Å

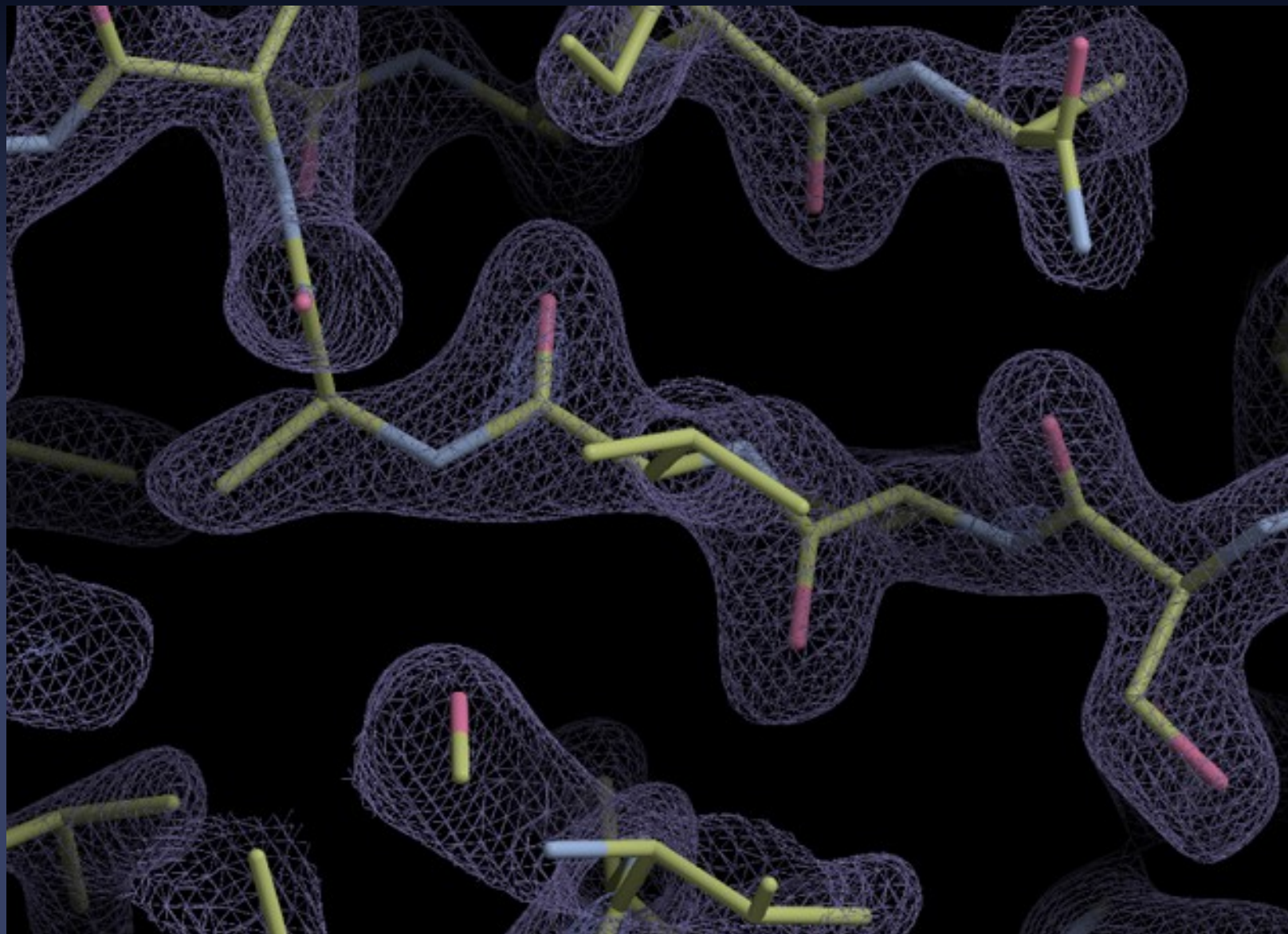


2.0Å

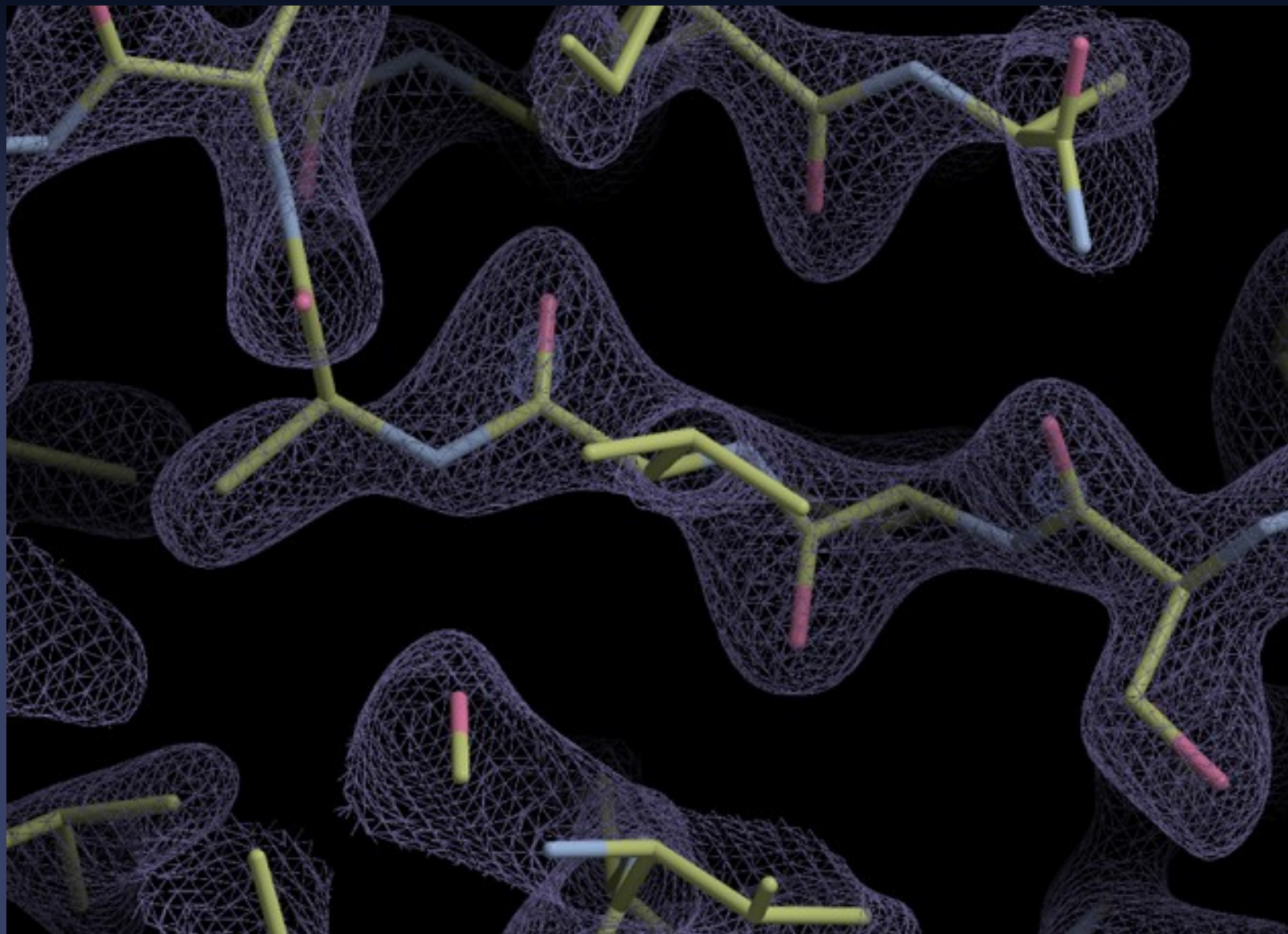




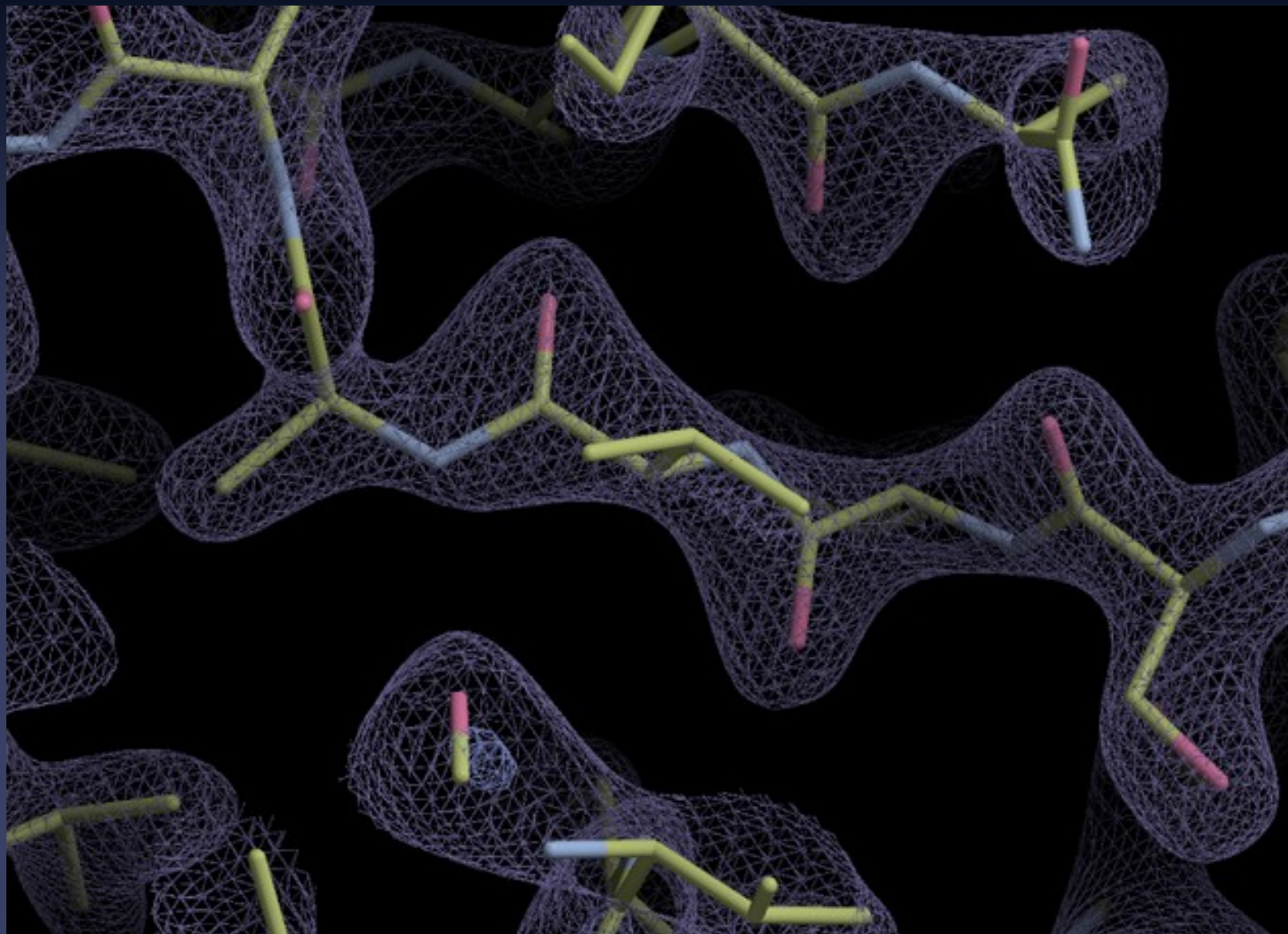
2.2Å



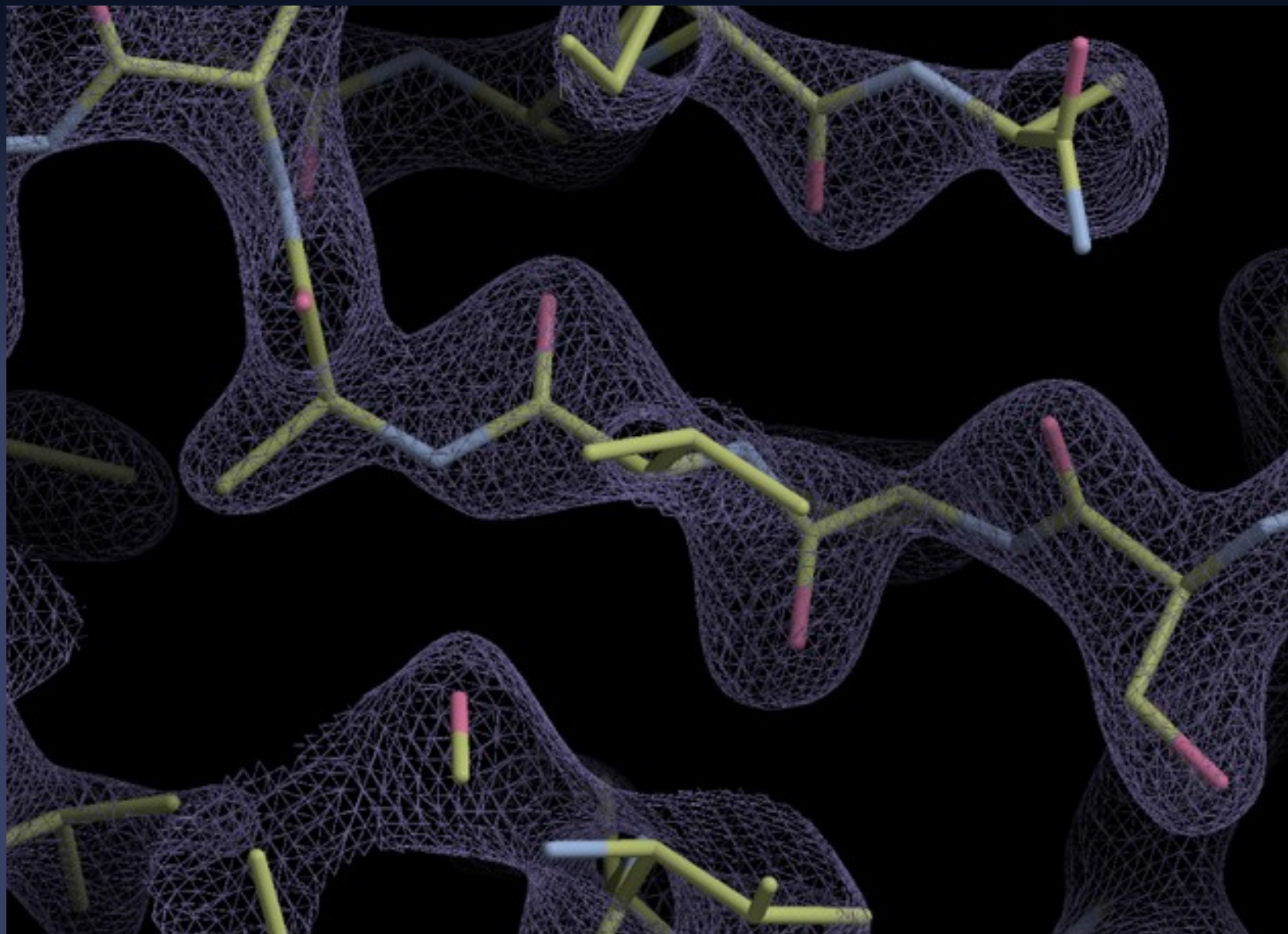
2.4Å



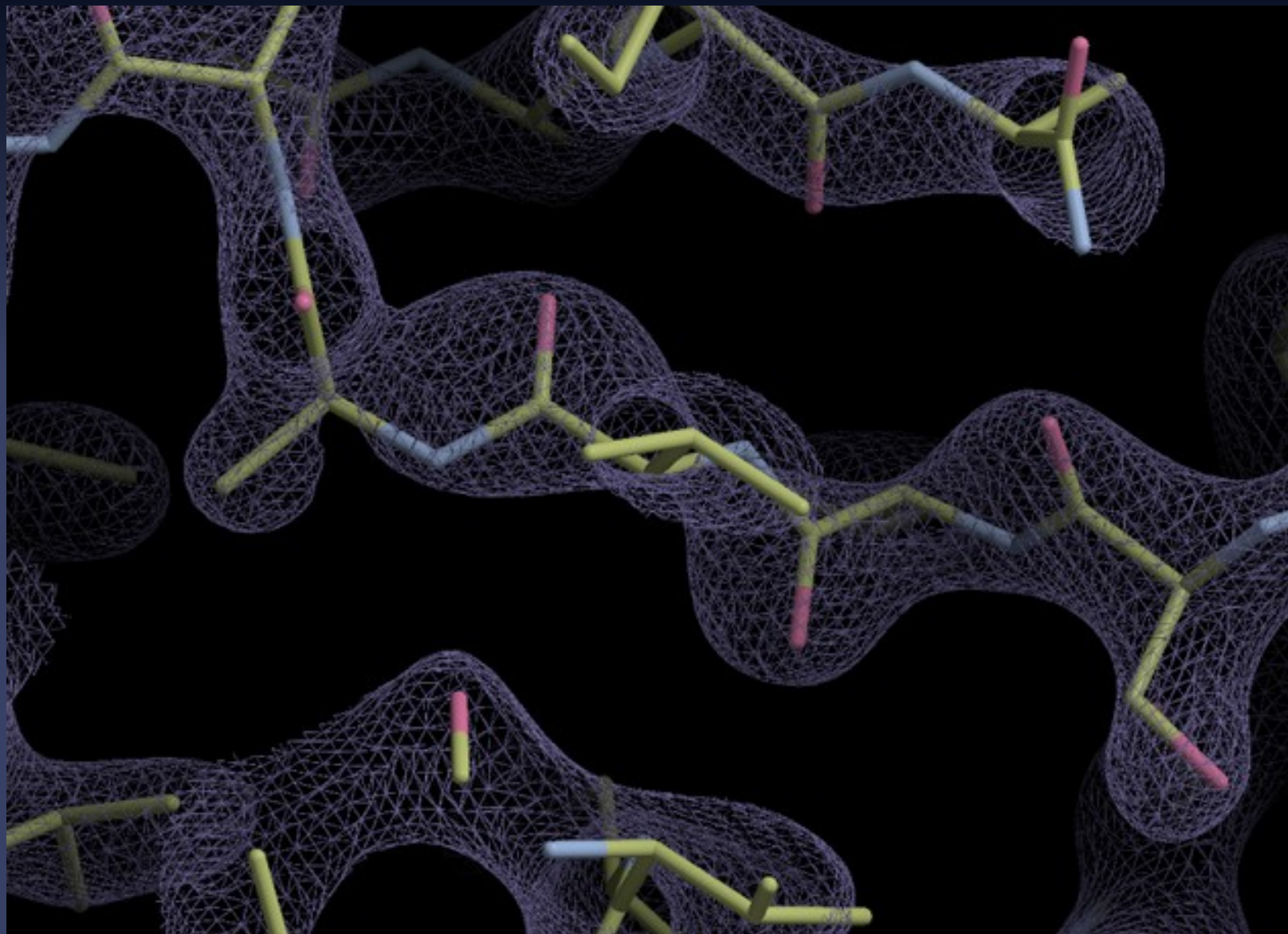
2.6Å



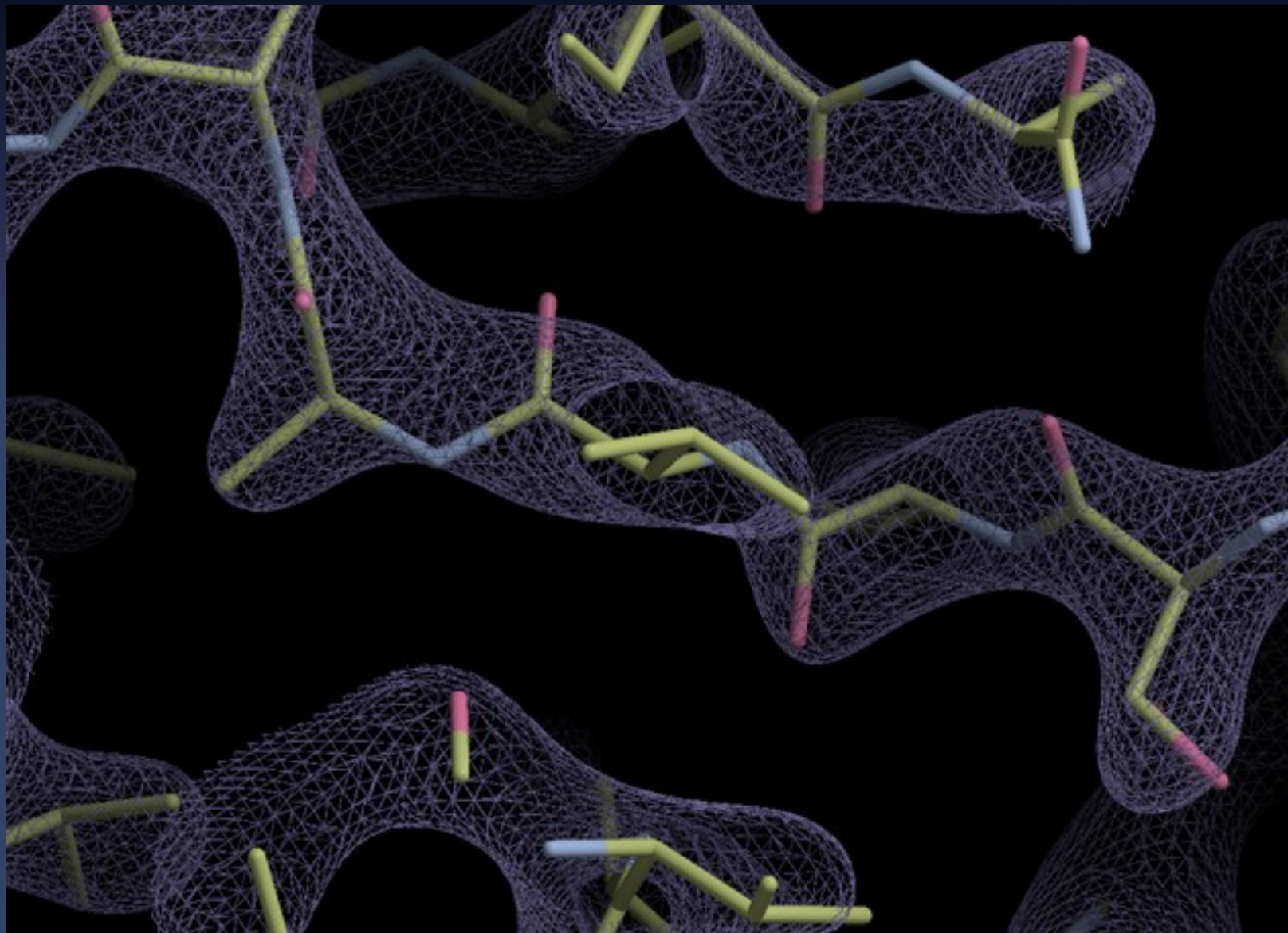
2.8Å



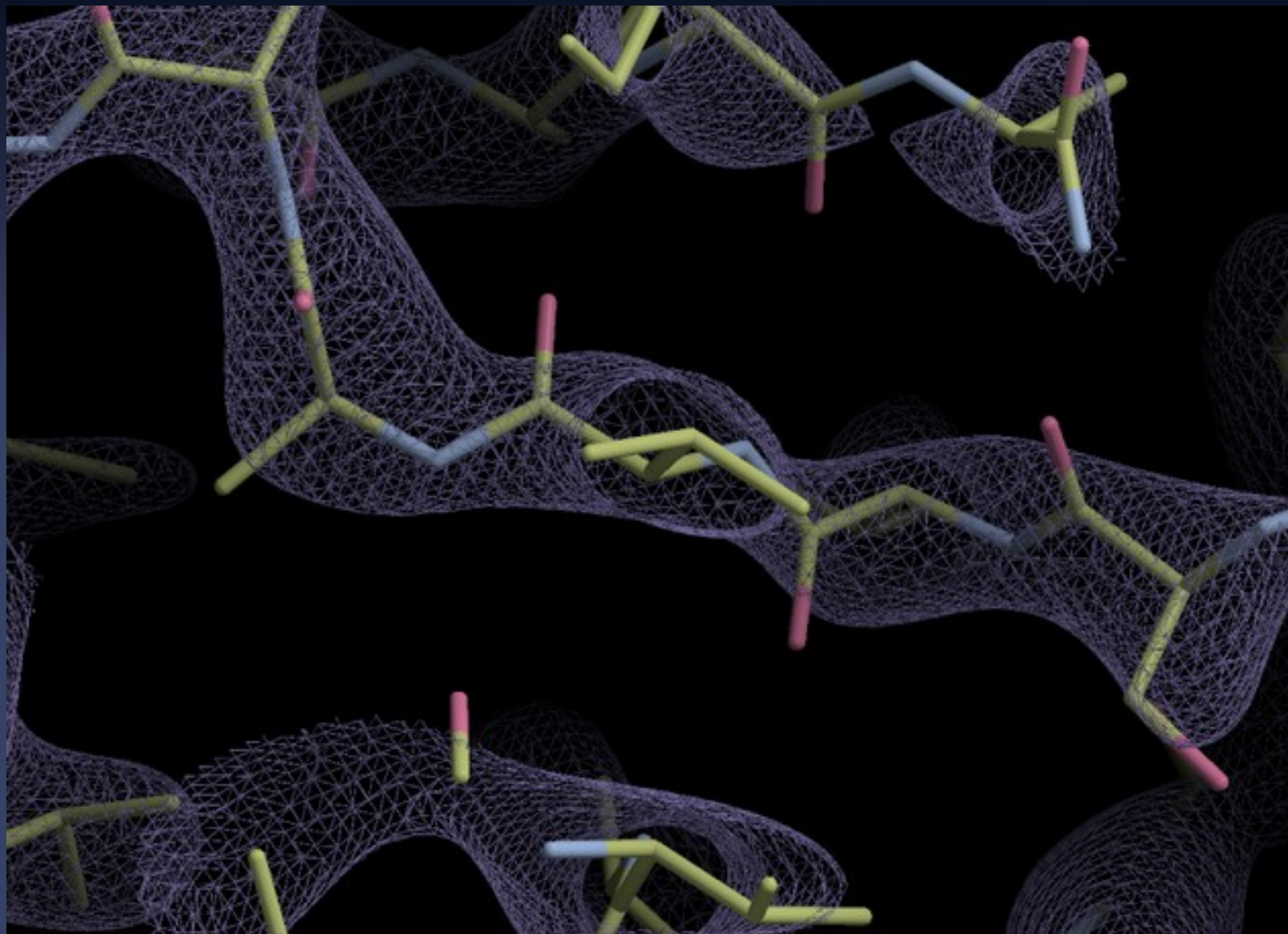
3.0Å



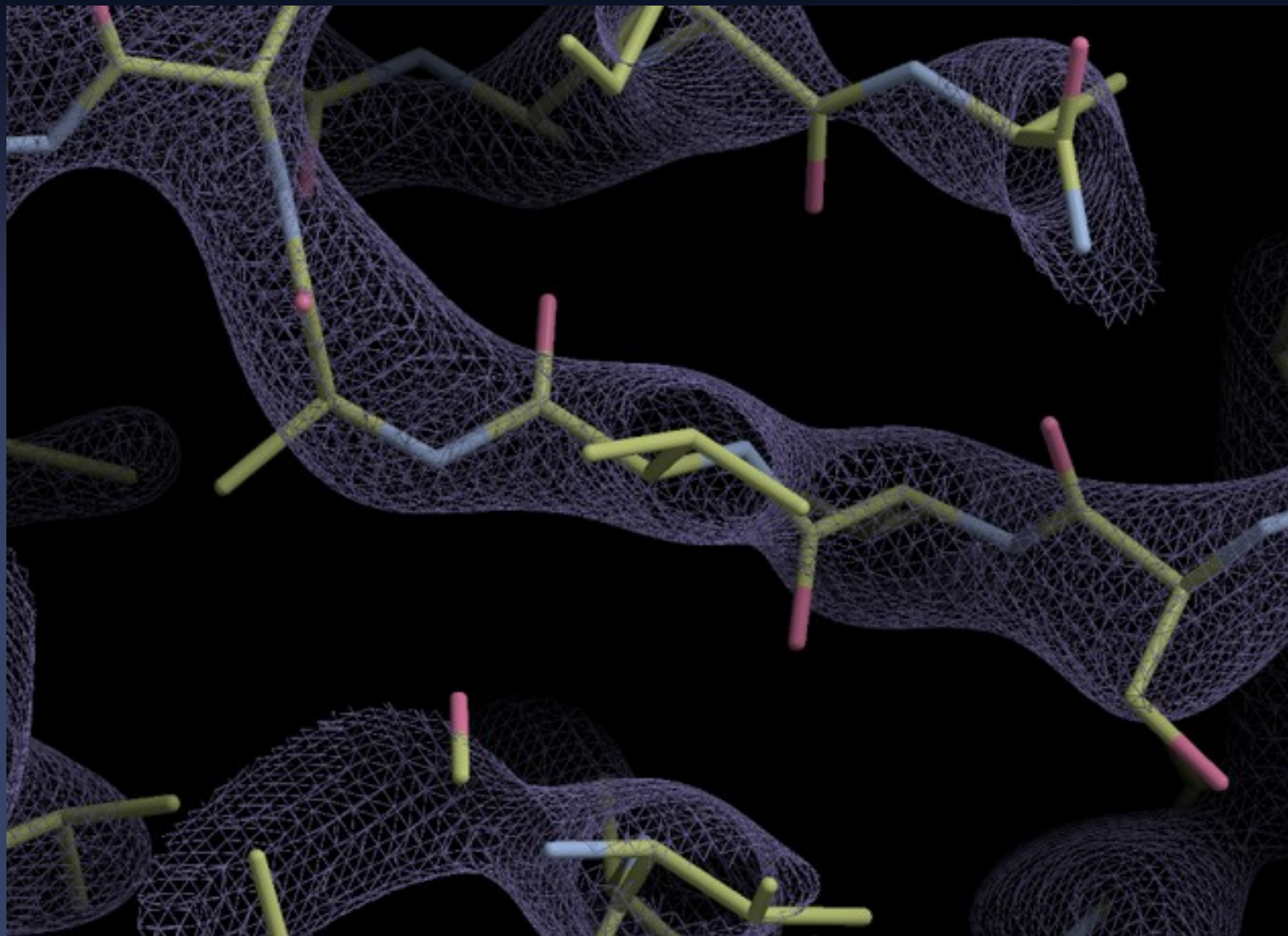
3.2Å



3.4Å

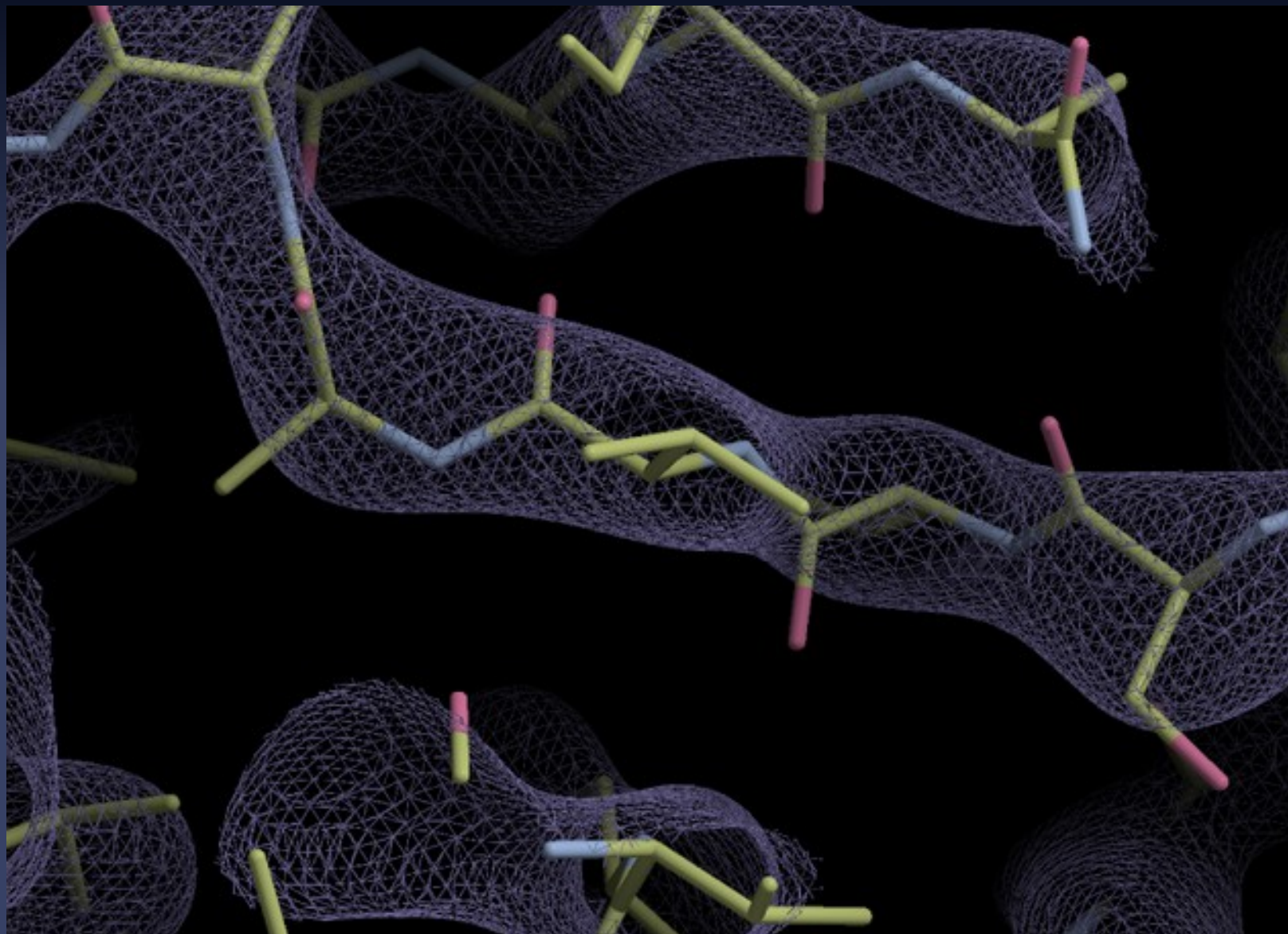


3.6Å

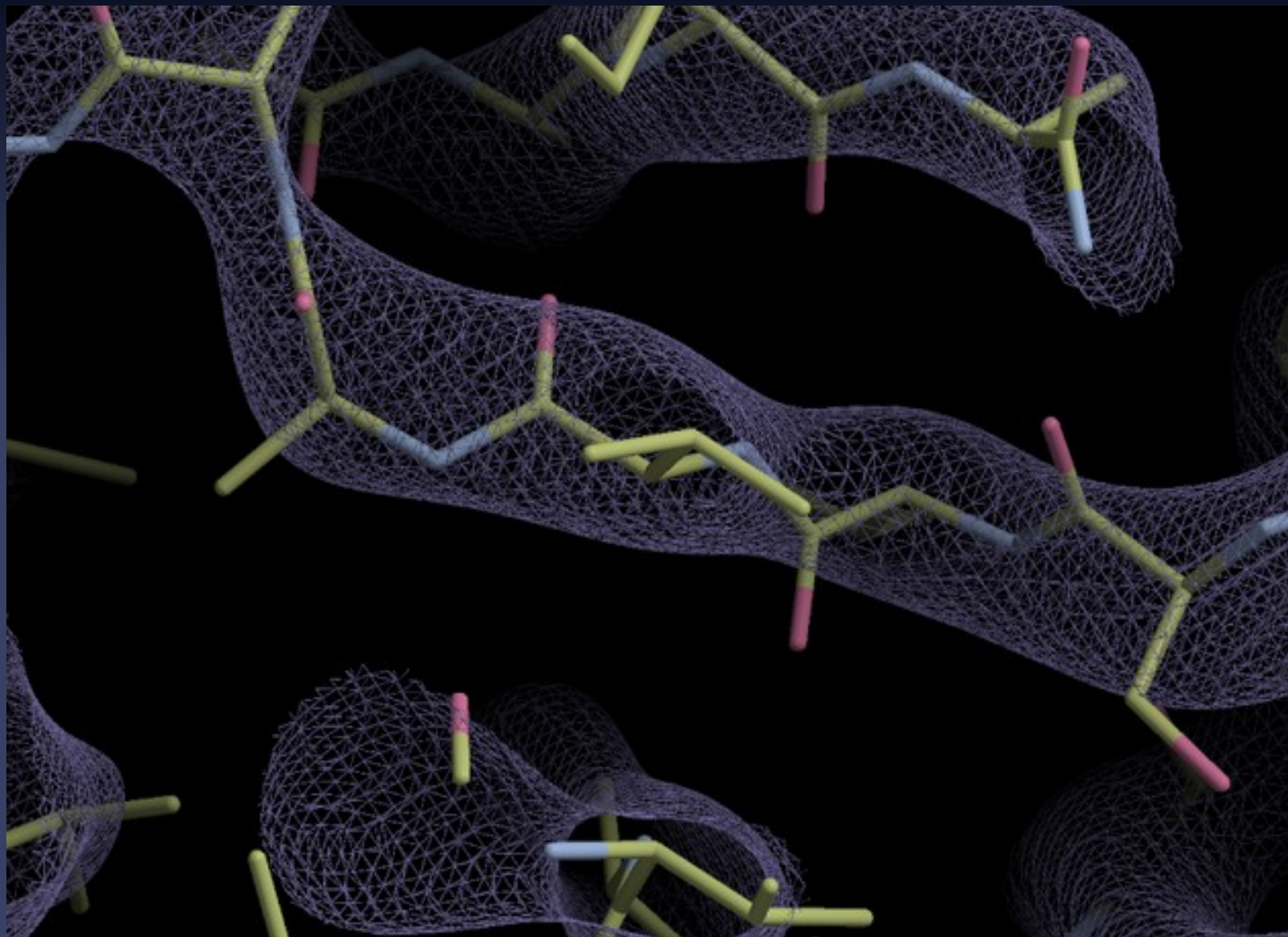




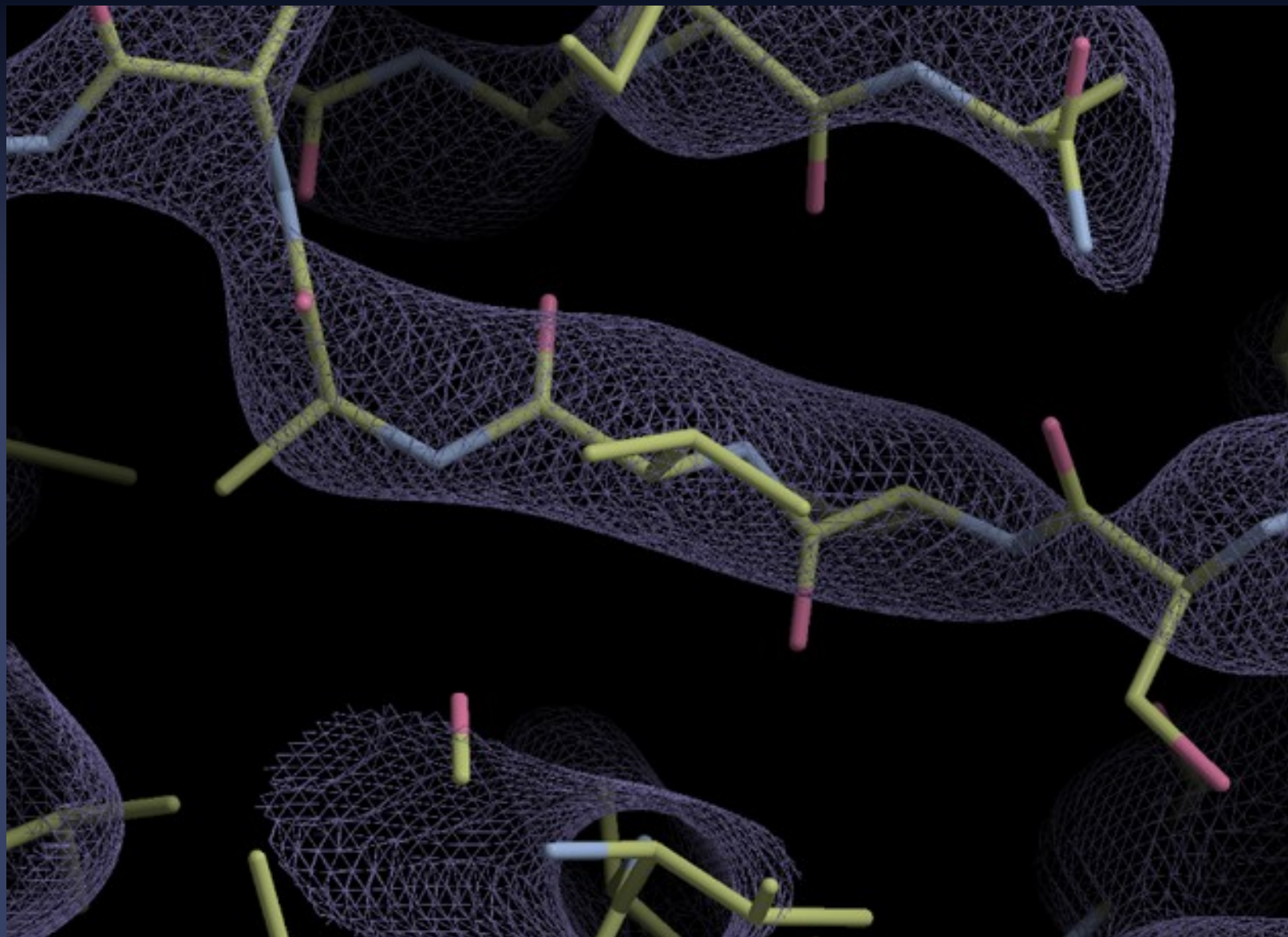
3.8Å



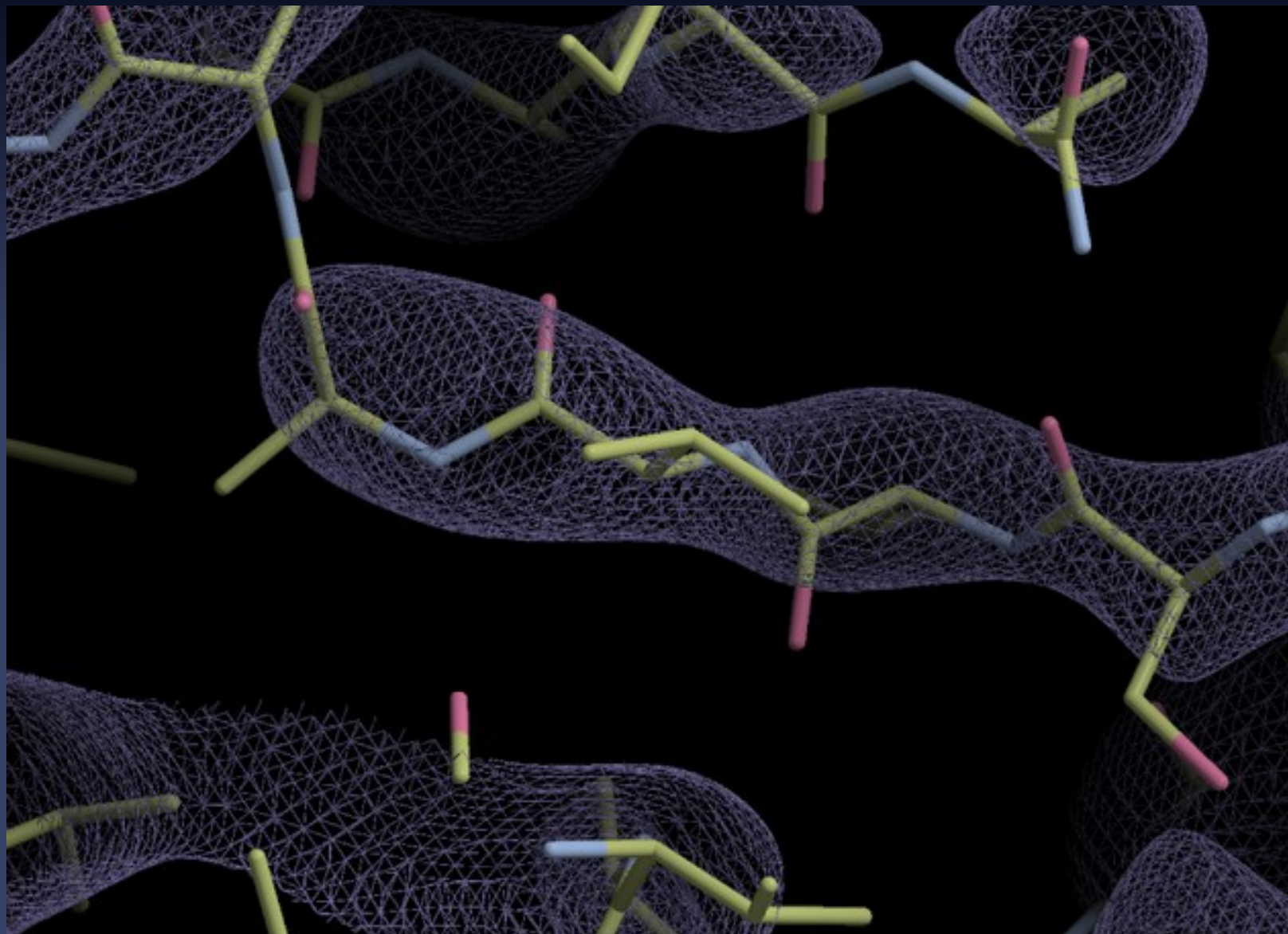
4.0Å



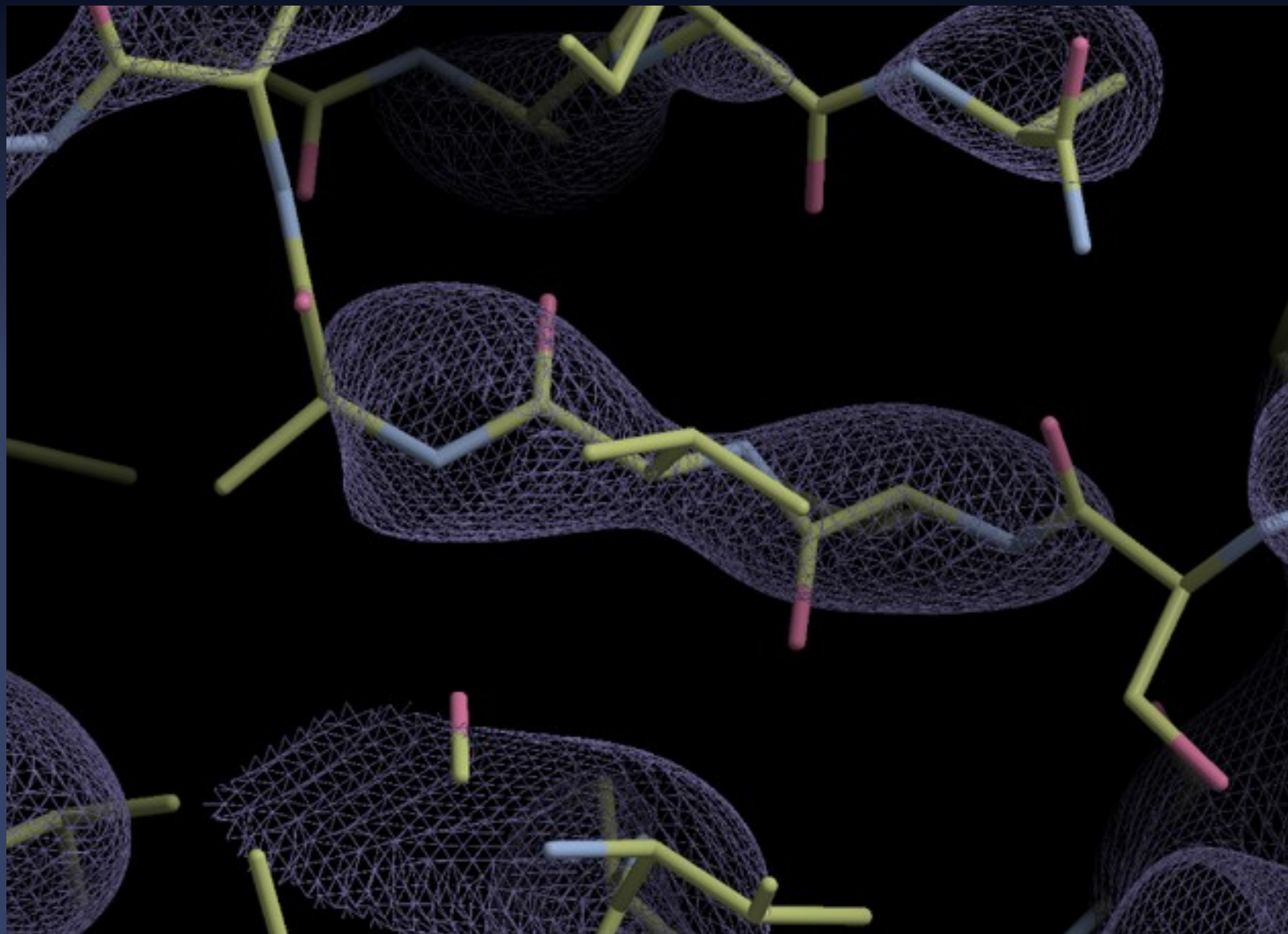
4.2Å



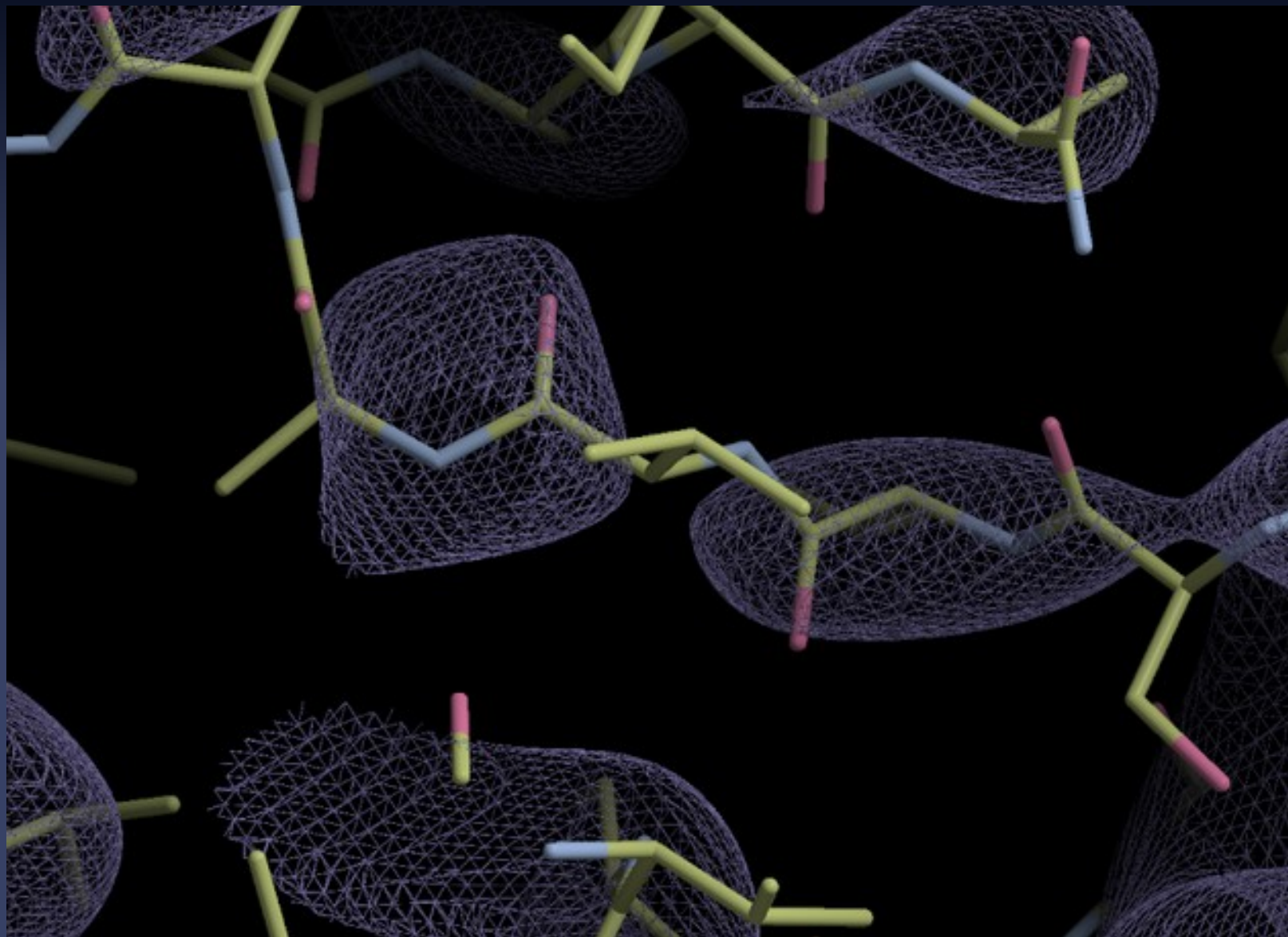
4.4Å



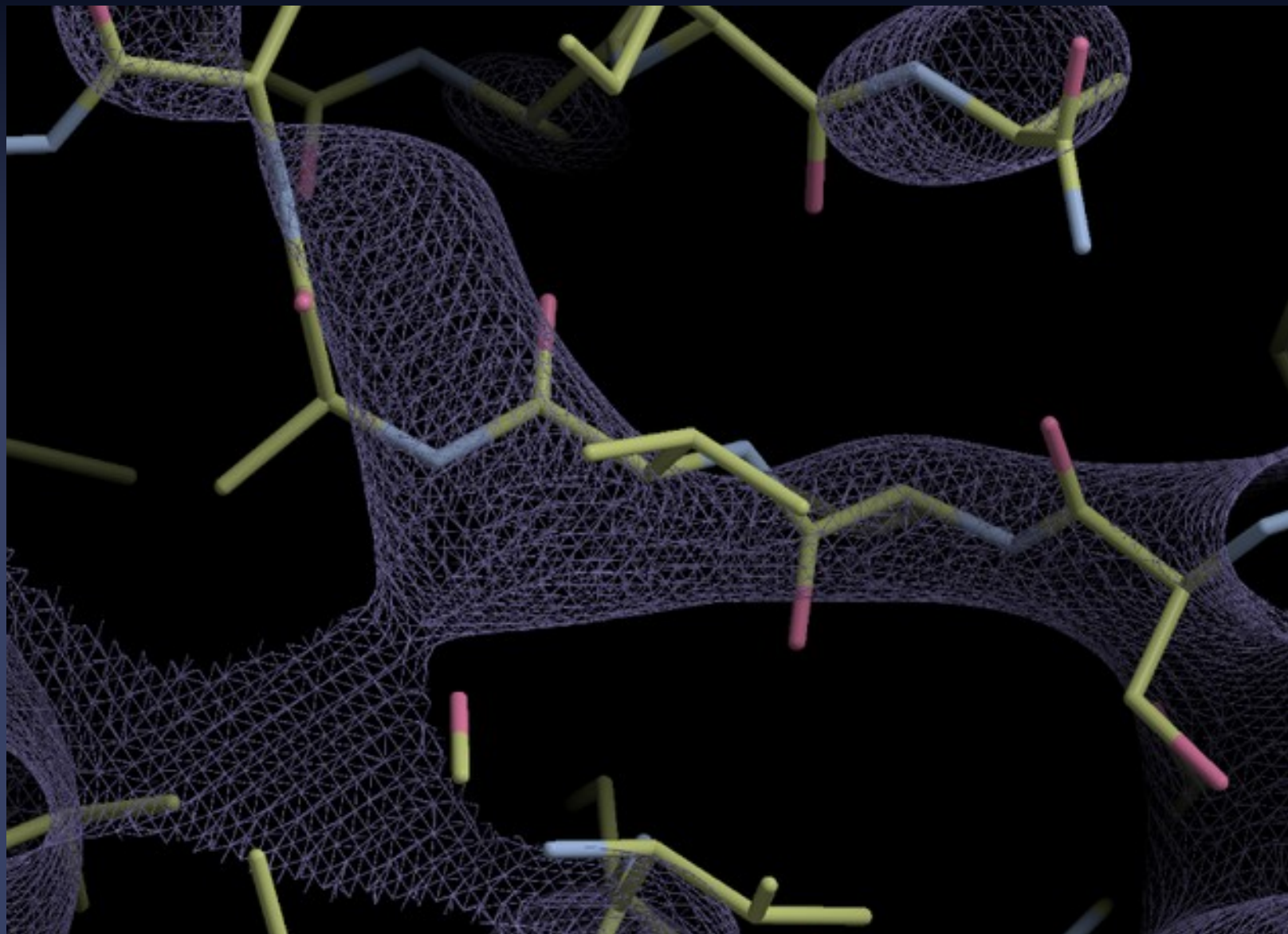
4.6Å



4.8Å

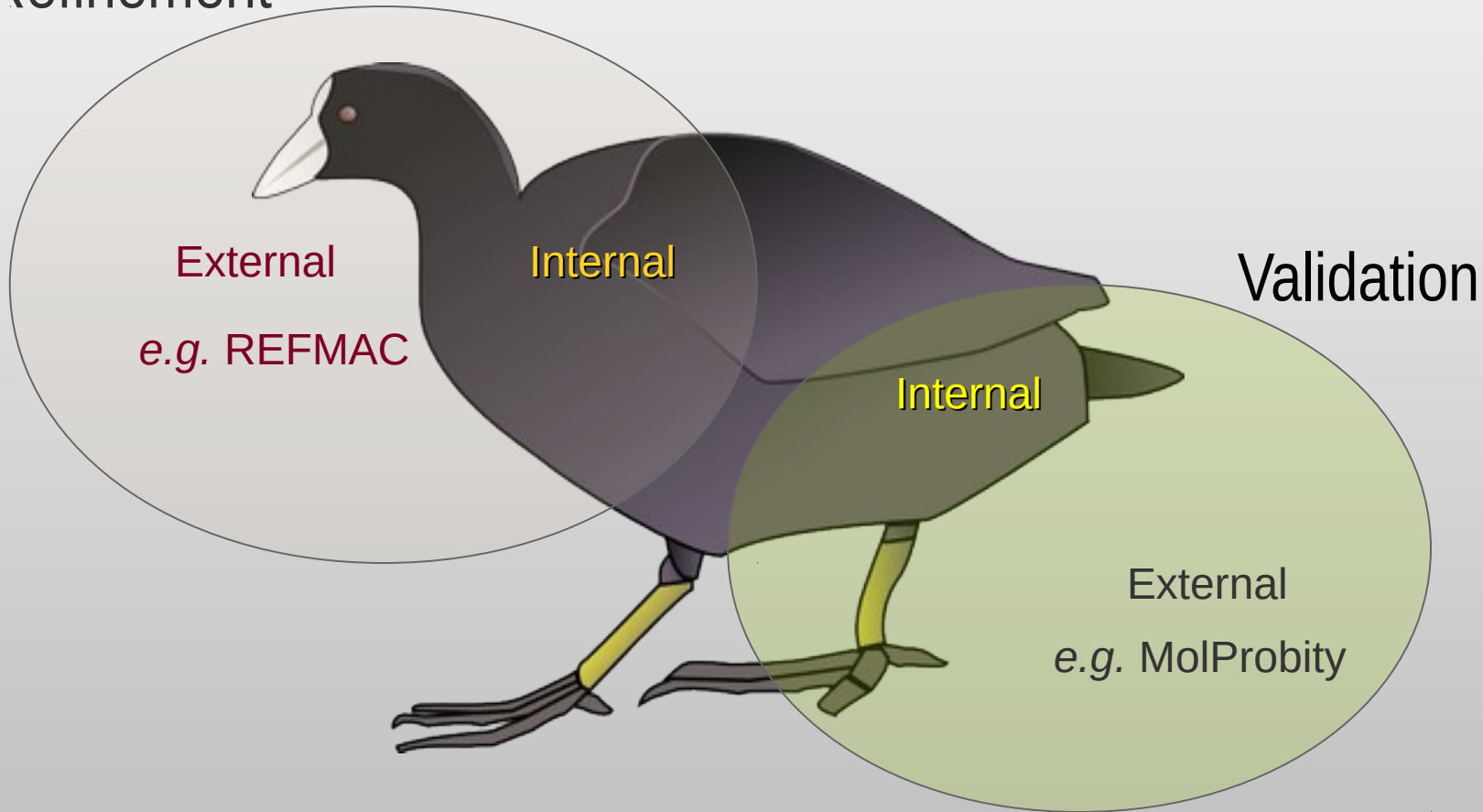


5.0Å



# Feature Integration

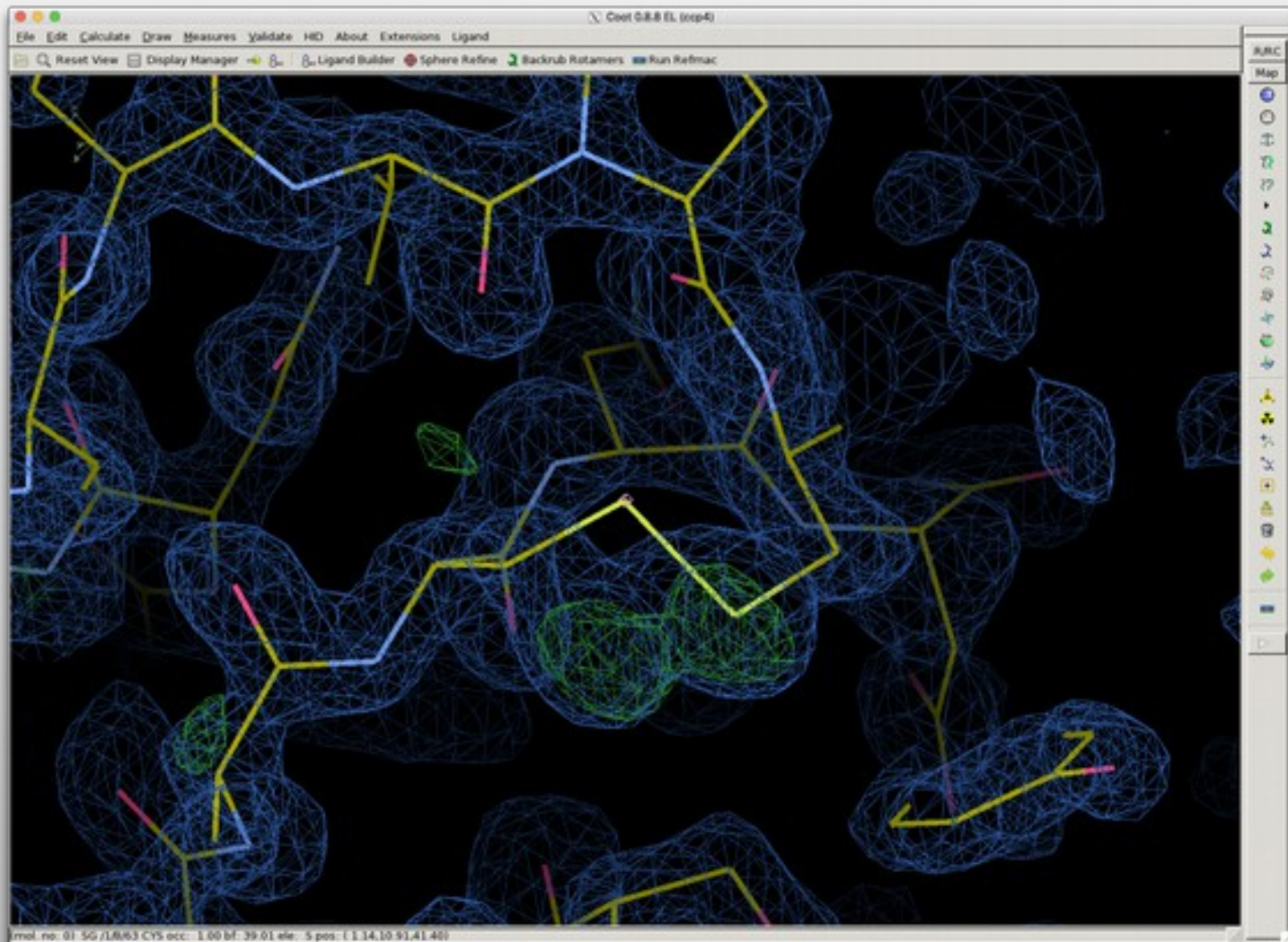
Refinement



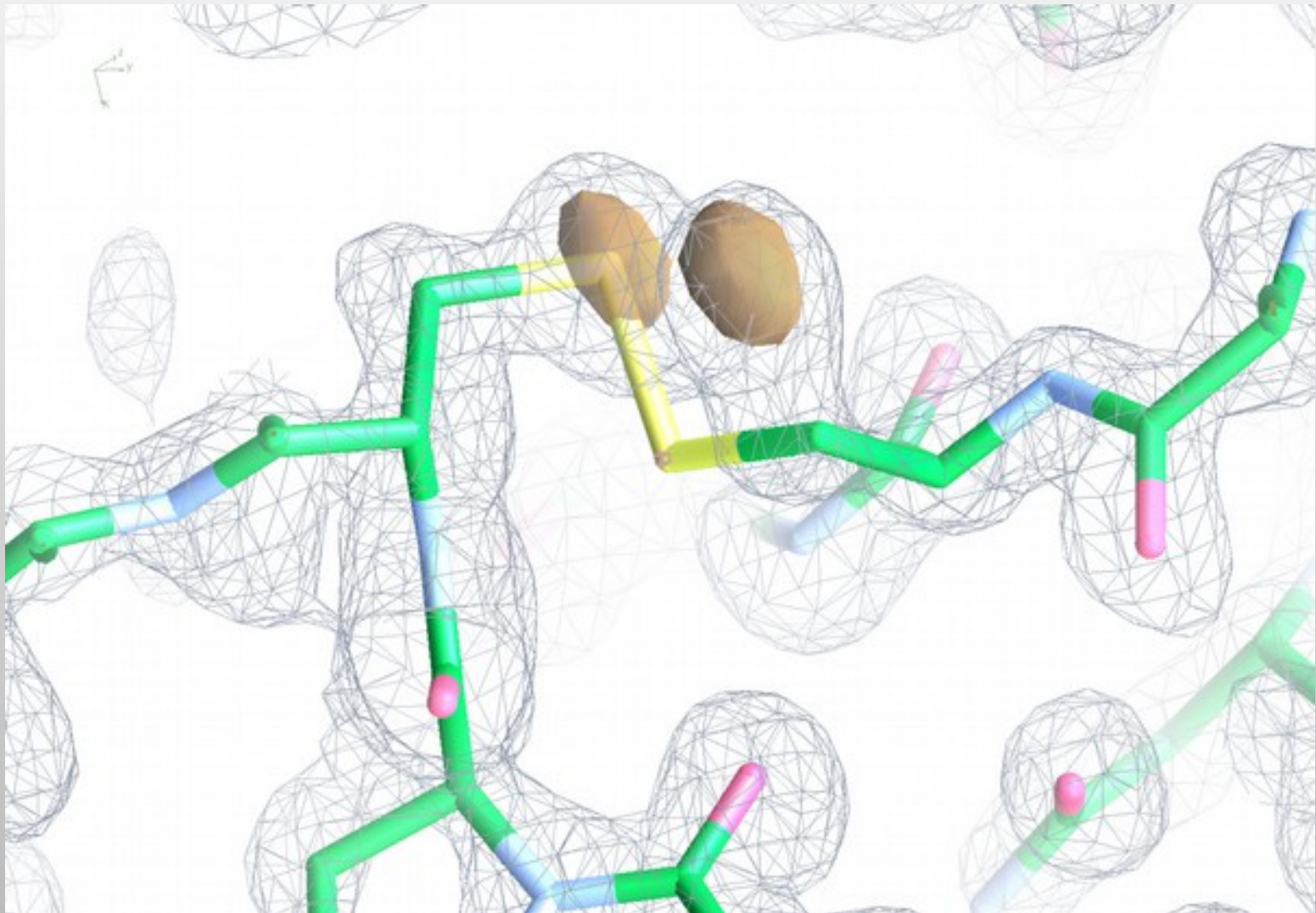
Validation, Model Building and Refinement should be used together



# Fixing what auto-building doesn't get right



# Fixing what auto-building doesn't get right



# What is “Refinement”?

- The adjustment of model parameters (co-ordinates) so that the calculated structure factors match the observations as nearly as possible
  - In “one-shot” real-space refinement, such as in Coot, this translates to:
    - move the atoms into as high density as possible while minimizing geometrical distortions

# Real Space Refinement

- Major feature of Coot
  - Gradient minimizer (BFGS derivative)
  - Based on mmCIF standard dictionary
  - Minimizing bonds, angles, planes, non-bonded contacts, torsions, chiral volumes
  - Additional user-defined restraints,
    - secondary structure restraints
    - homologous protein local environment restraints
- Provides “interactive refinement”

Refinement in *Coot* has been extended in several ways...

# What prior geometric information do we have?

- We know chemistry....
  - We know bond lengths and uncertainties
  - We know bond angles and uncertainties
  - We know the chiral centres
  - We know which atoms should lie in a plane
  - We know (more or less) about torsions
- We combine the gradients from the data with those from molecular mechanics in the minimisation

# REFMAC Monomer Library

## chem\_comp\_bond

```
loop_  
_chem_comp_bond.comp_id  
_chem_comp_bond.atom_id_1  
_chem_comp_bond.atom_id_2  
_chem_comp_bond.type  
_chem_comp_bond.value_dist  
_chem_comp_bond.value_dist_esd  
ALA      N      H      single      0.860      0.020  
ALA      N      CA     single      1.458      0.019  
ALA      CA     HA     single      0.980      0.020  
ALA      CA     CB     single      1.521      0.033  
ALA      CA     C      single      1.525      0.021  
ALA      C      O      double      1.231      0.020
```

## APPENDIX A

### Regularization and refinement derivatives

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

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

## A1. Bonds

$$S_{\text{bond}} = \sum_{i=1}^{N_{\text{bonds}}} (b_i - b_{0_i})^2,$$

where  $b_{0_i}$  is the ideal length (from the dictionary) of the  $i$ th bond,  $\mathbf{b}_i$  is the bond vector and  $b_i$  is its length.

$$\frac{\partial S_i}{\partial x_m} = \frac{\partial S_i}{\partial b_i} \frac{\partial b_i}{\partial x_m} = [2(b_i - b_{0_i})] \frac{\partial b_i}{\partial x_m},$$

$$b_i = [(x_m - x_k)^2 + (y_m - y_k)^2 + (z_m - z_k)^2]^{1/2}.$$

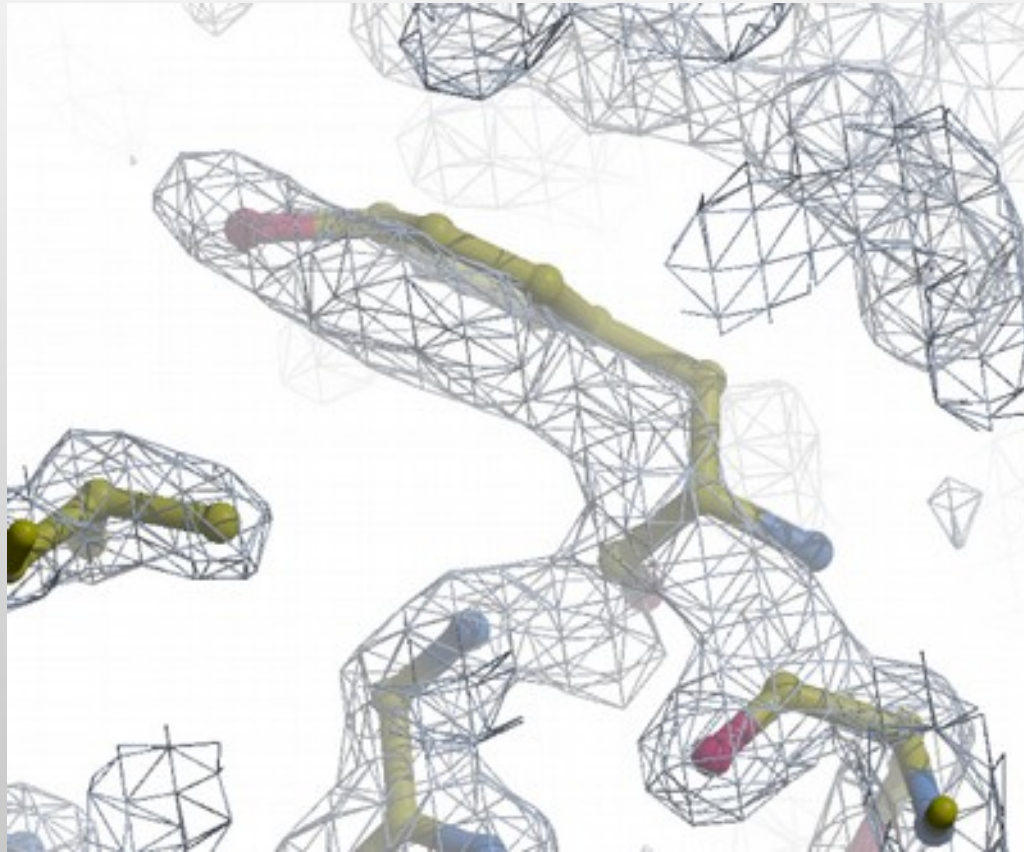
Therefore

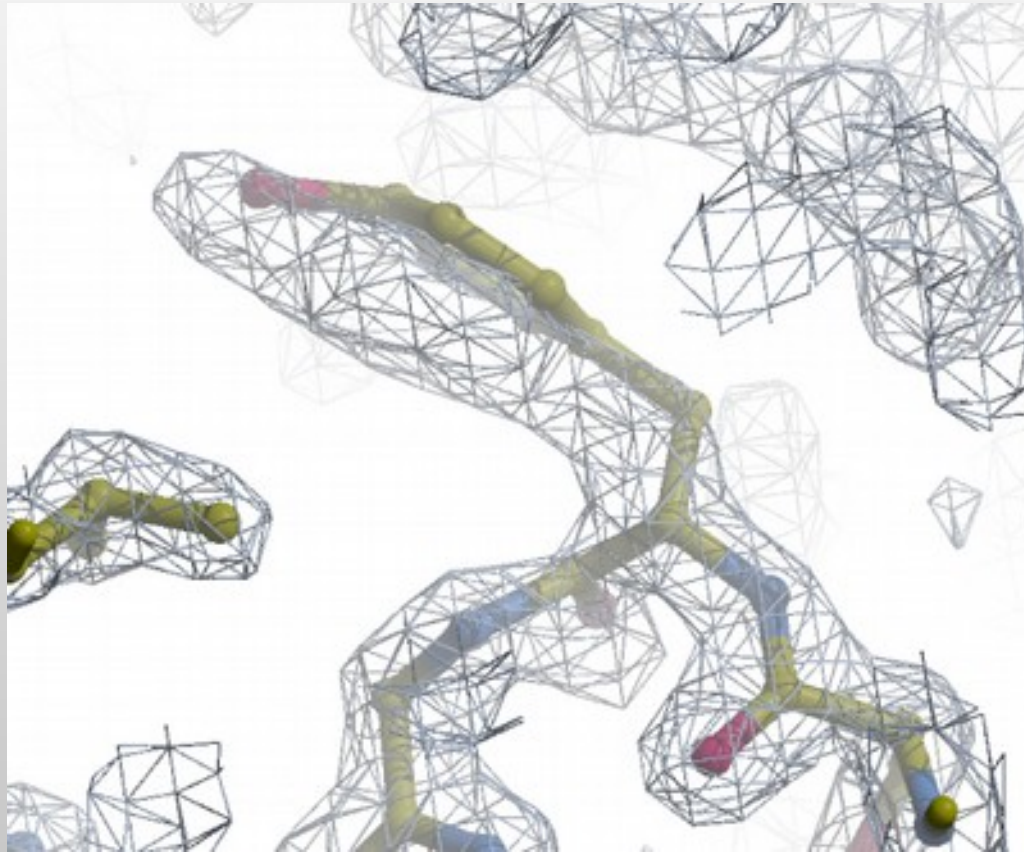
$$\frac{\partial b_i}{\partial x_m} = \left( \frac{1}{2} \frac{1}{b_i} \right) 2(x_m - x_k) = \frac{(x_m - x_k)}{b_i}$$

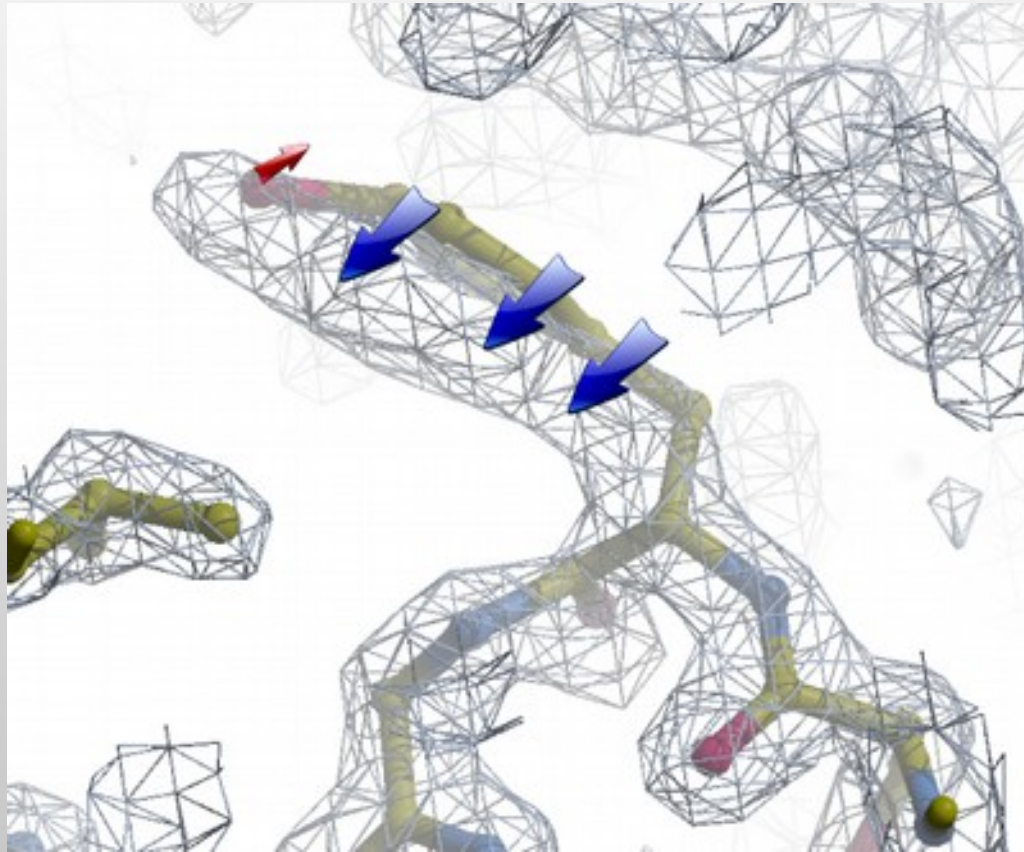
and

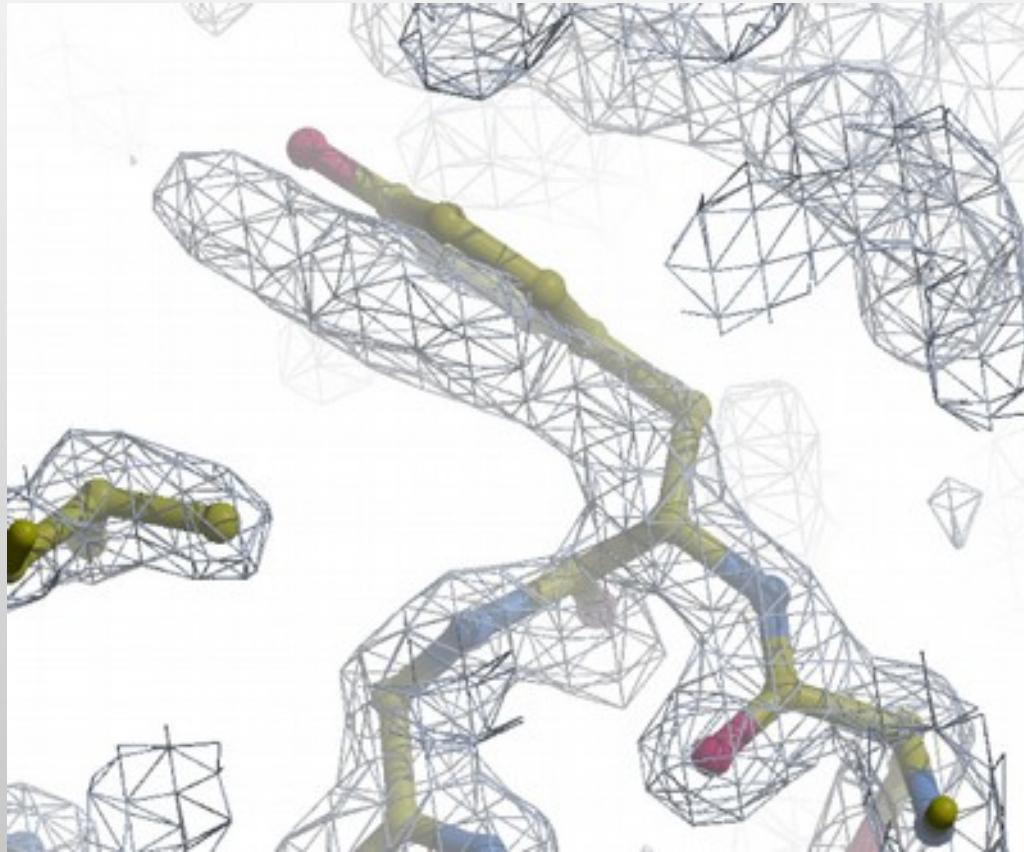
$$\frac{\partial S_i}{\partial x_m} = 2[b_i - b_{0_i}] \frac{(x_m - x_k)}{b_i}.$$

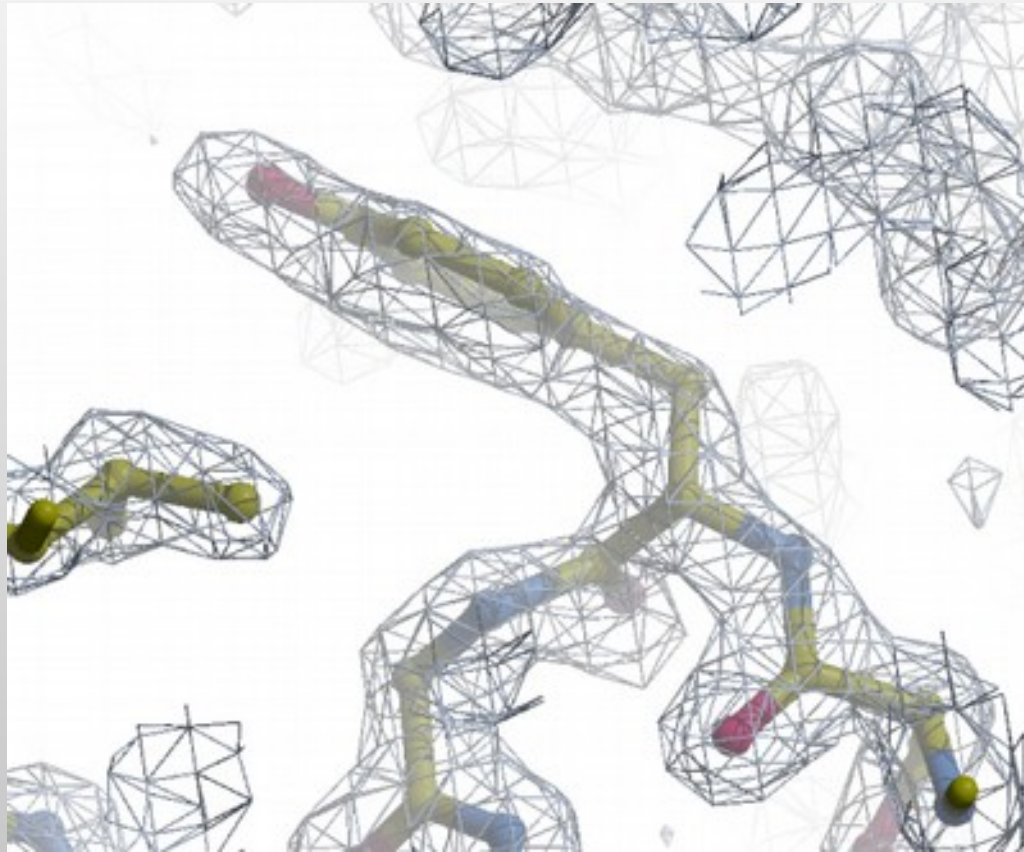


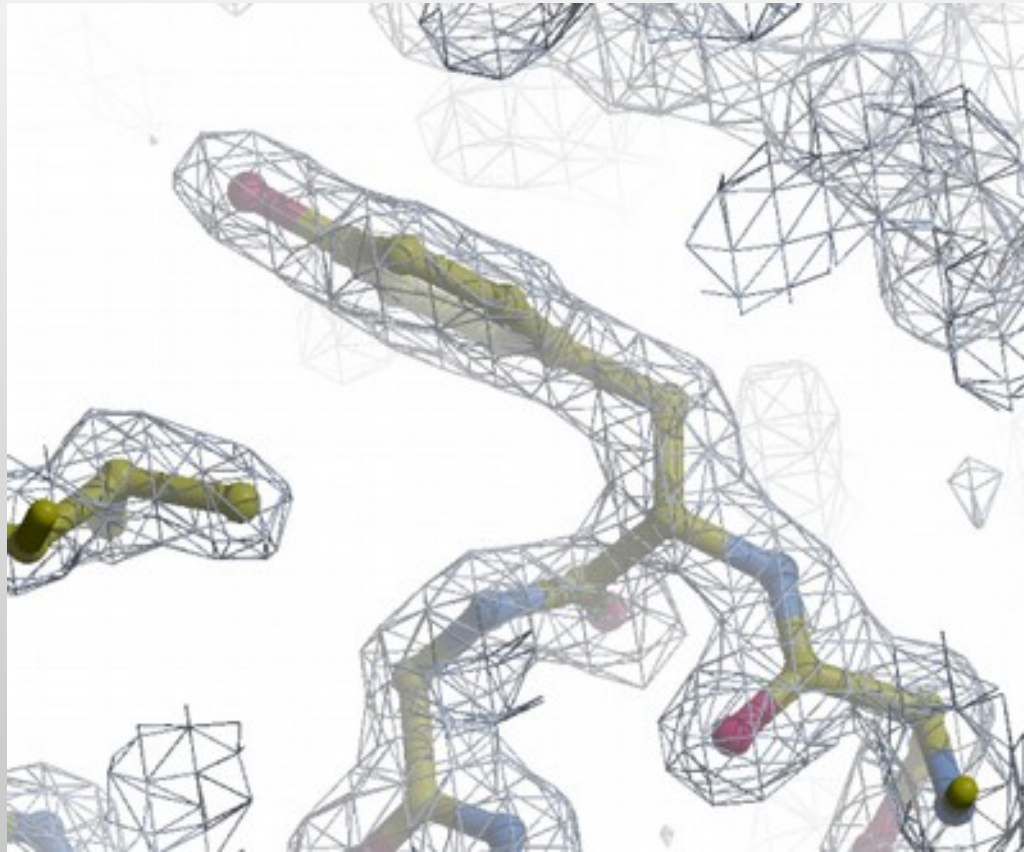






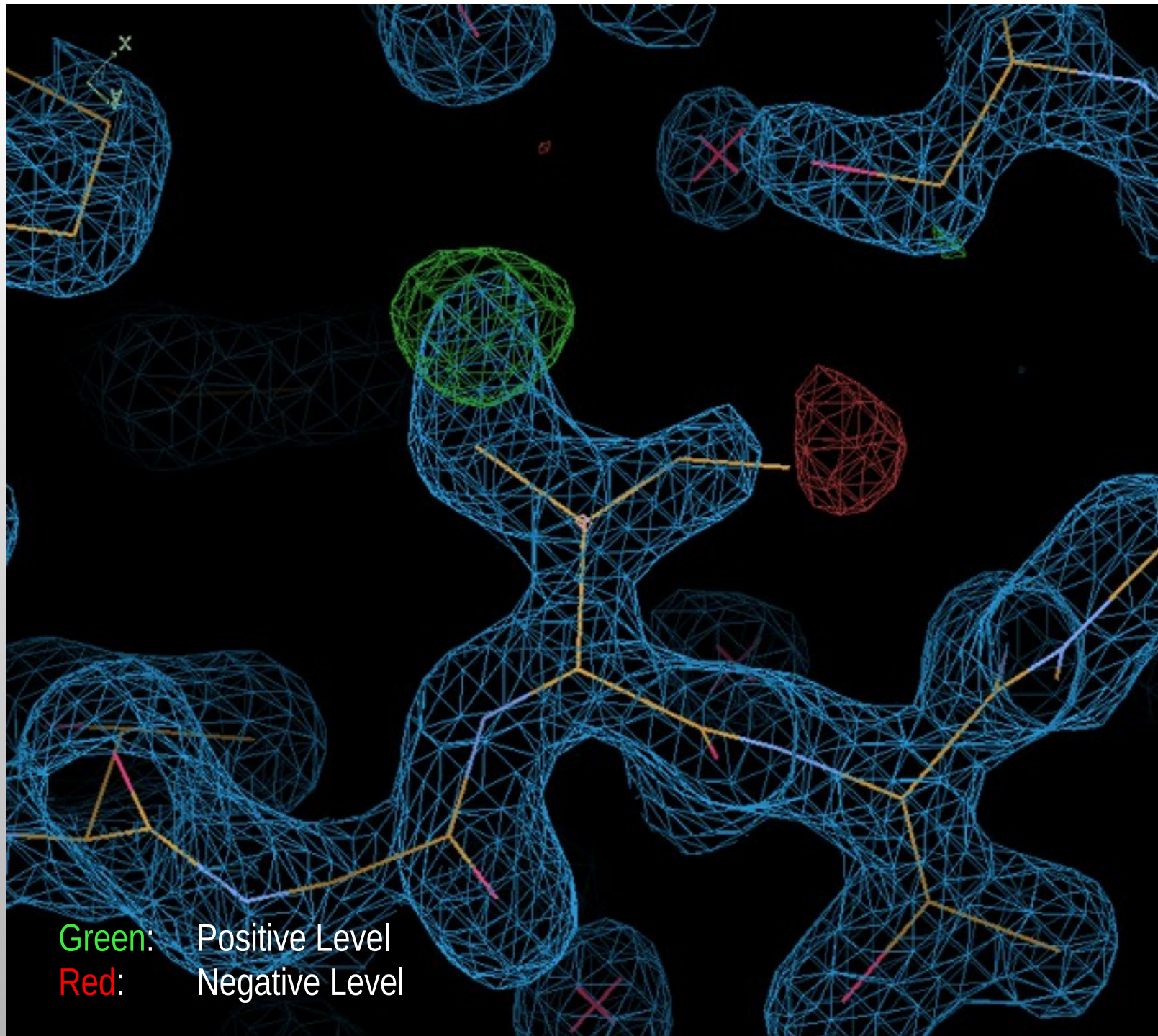






# Different types of electron density maps

- “Experimental” maps
  - maps that result directly from the crystallographic data analysis: MIR, MAD, SAD
- Direct Maps:
  - where the atoms are
- Coefficients  $F_o - F_c$  (“difference map”)
  - Identifies errors in the model. Locations in space where there should be atoms show positive peaks, while locations where the model contains atoms that should not be there show negative peaks.



Green: Positive Level  
Red: Negative Level



# 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

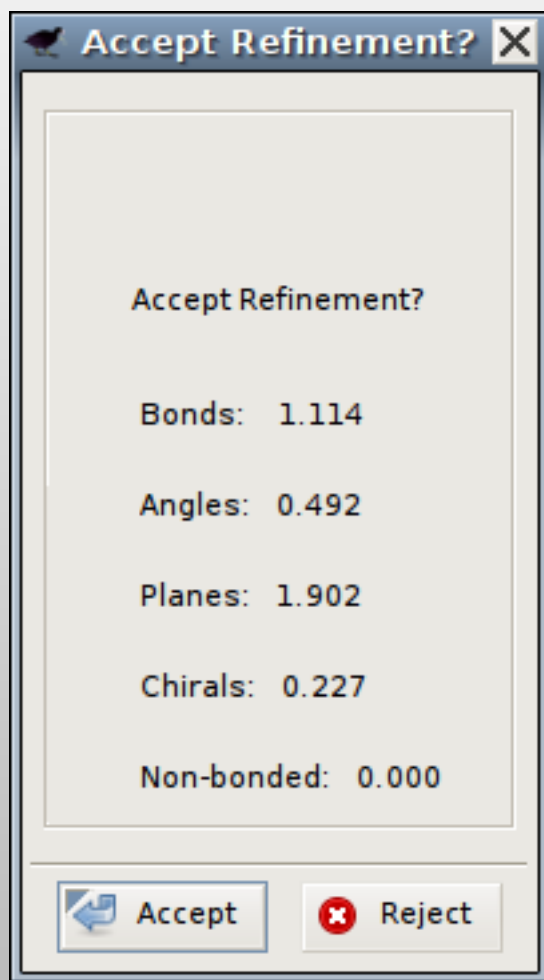
INFO:: [spec: "A" 45 "" ] [spec: "A" 46 "" ] link_type :TRANS:
INFO:: [spec: "A" 45 "" ] [spec: "A" 44 "" ] link_type :TRANS:
Link restraints:
  2 bond      links
  6 angle     links
  4 plane     links
Flanking residue restraints:
  4 bond      links
 12 angle     links
  8 plane     links
INFO:: made 668 non-bonded restraints
initial distortion score: -16033.2
  Initial Chi Squares
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

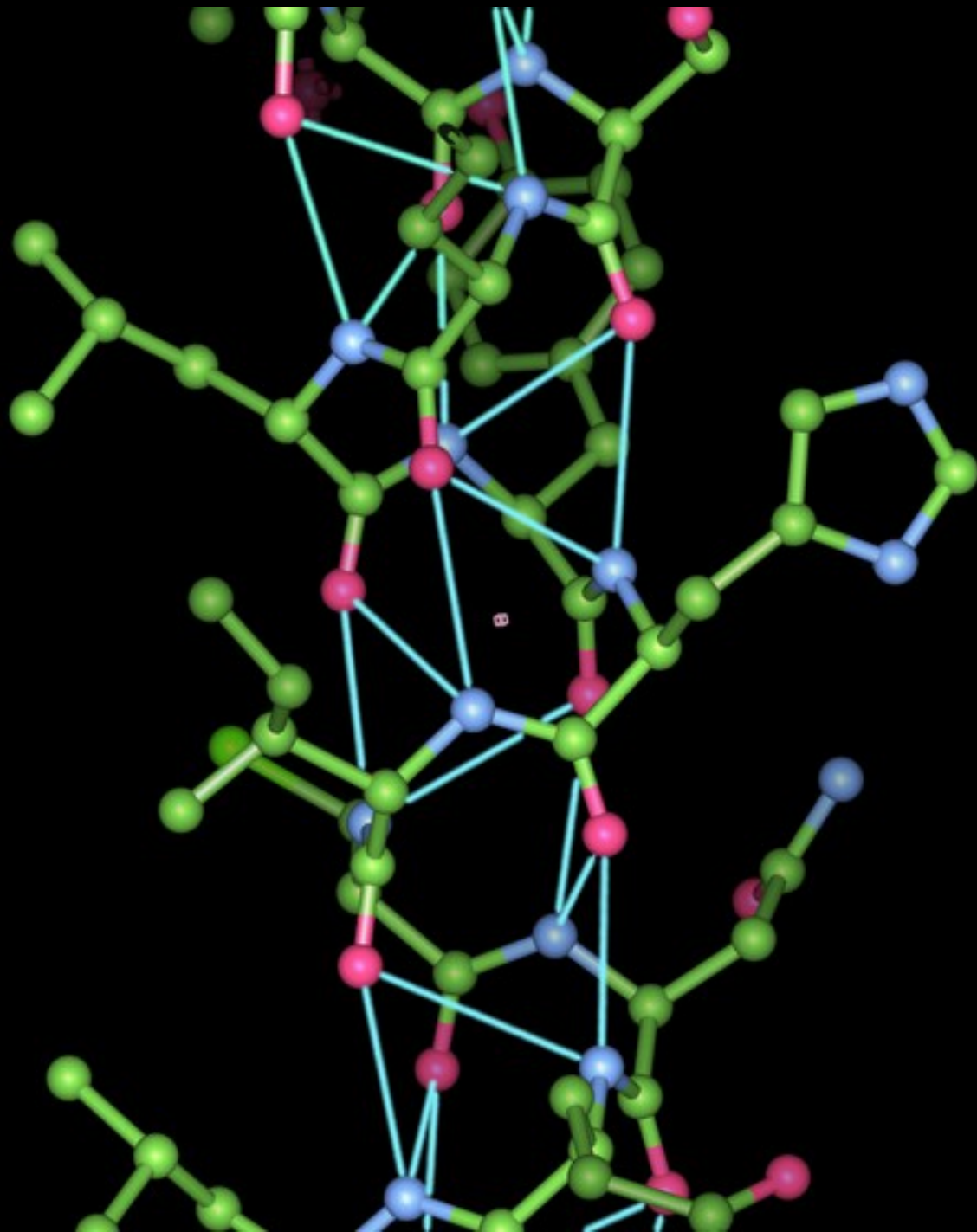


Bad refinement

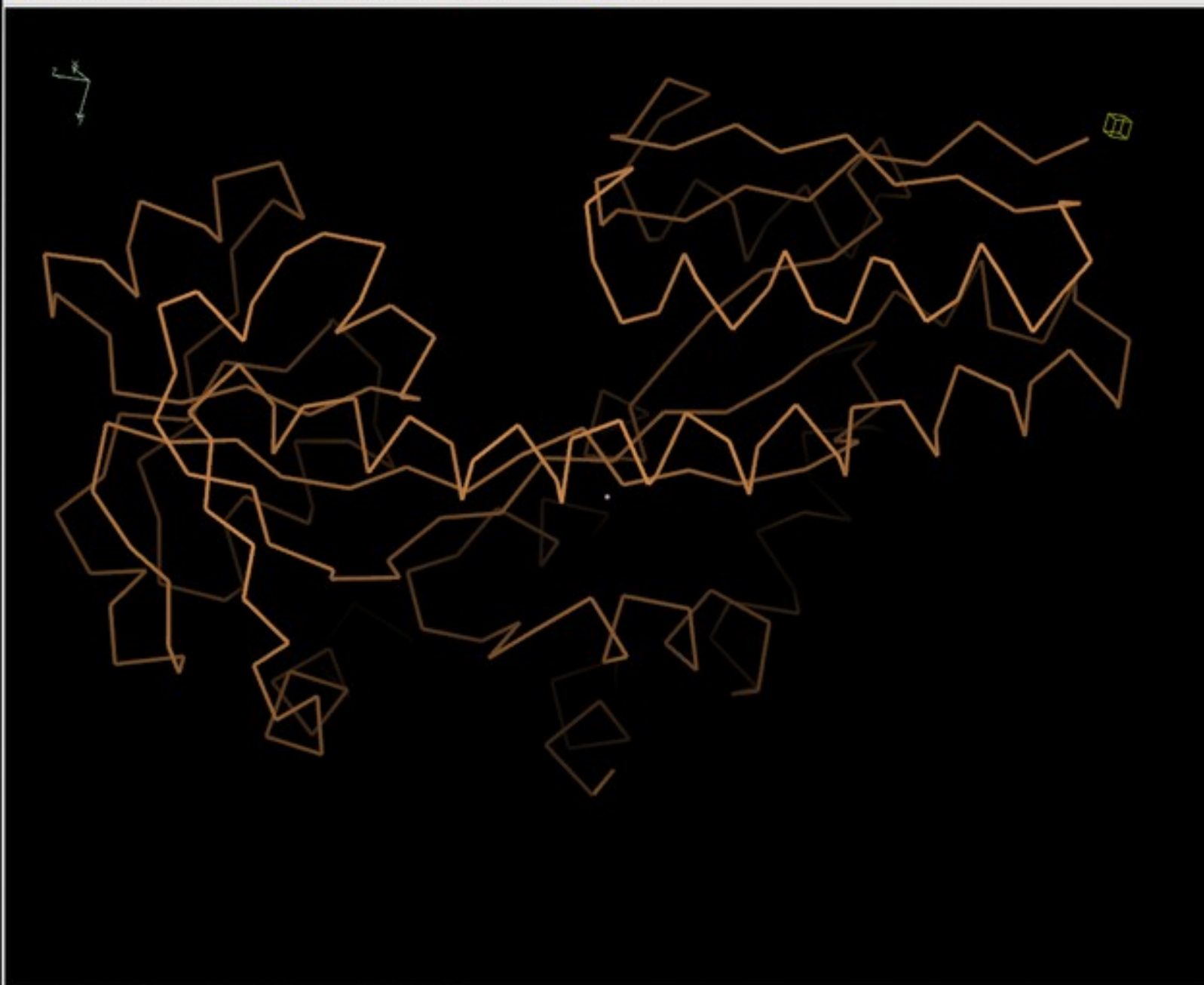
# ProSMART Interface

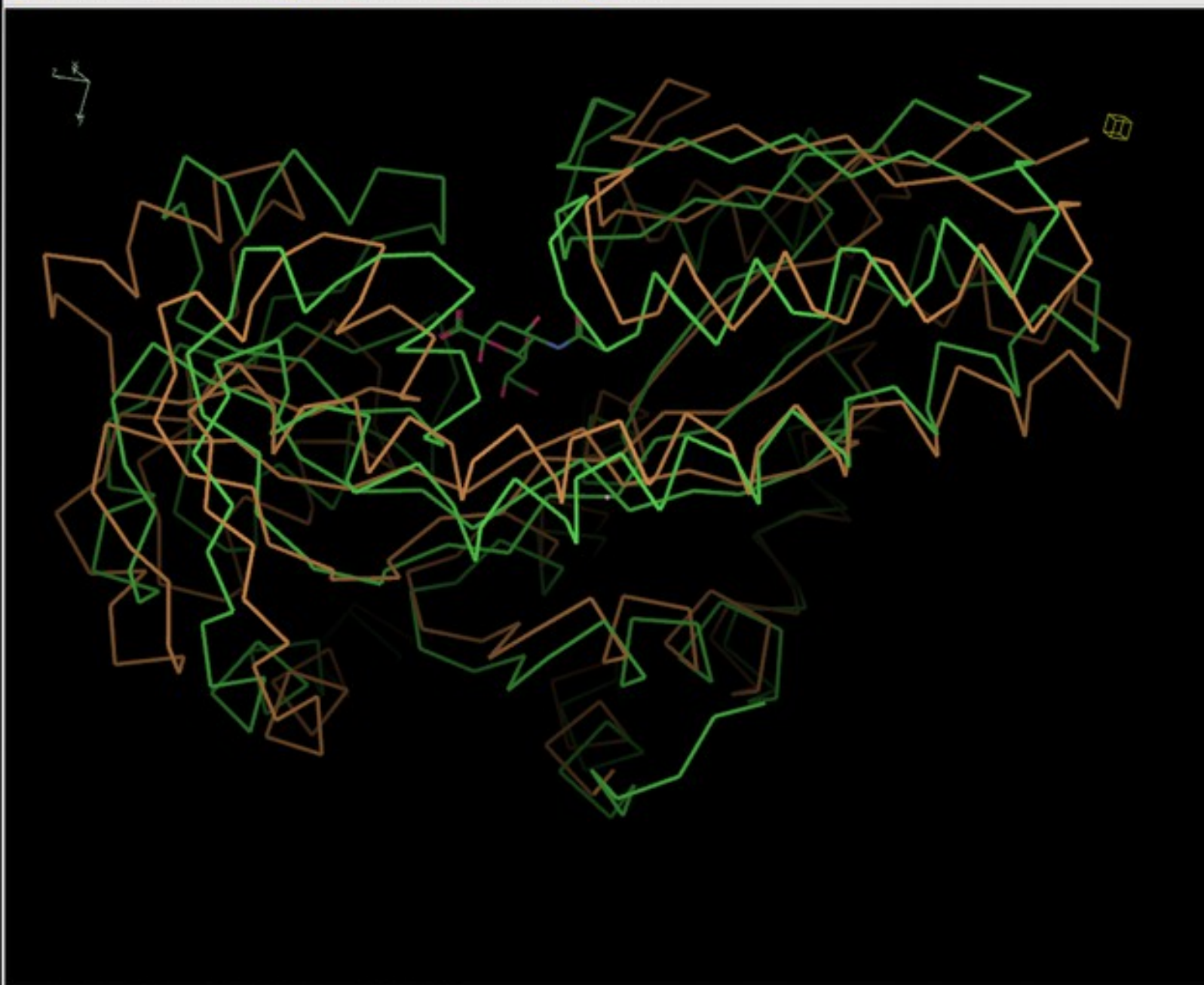
- Use previous-solved “template” structures to inform the refinement of the (low resolution) target protein
- Conformation-independent structural comparison/superposition
- and restraint generation

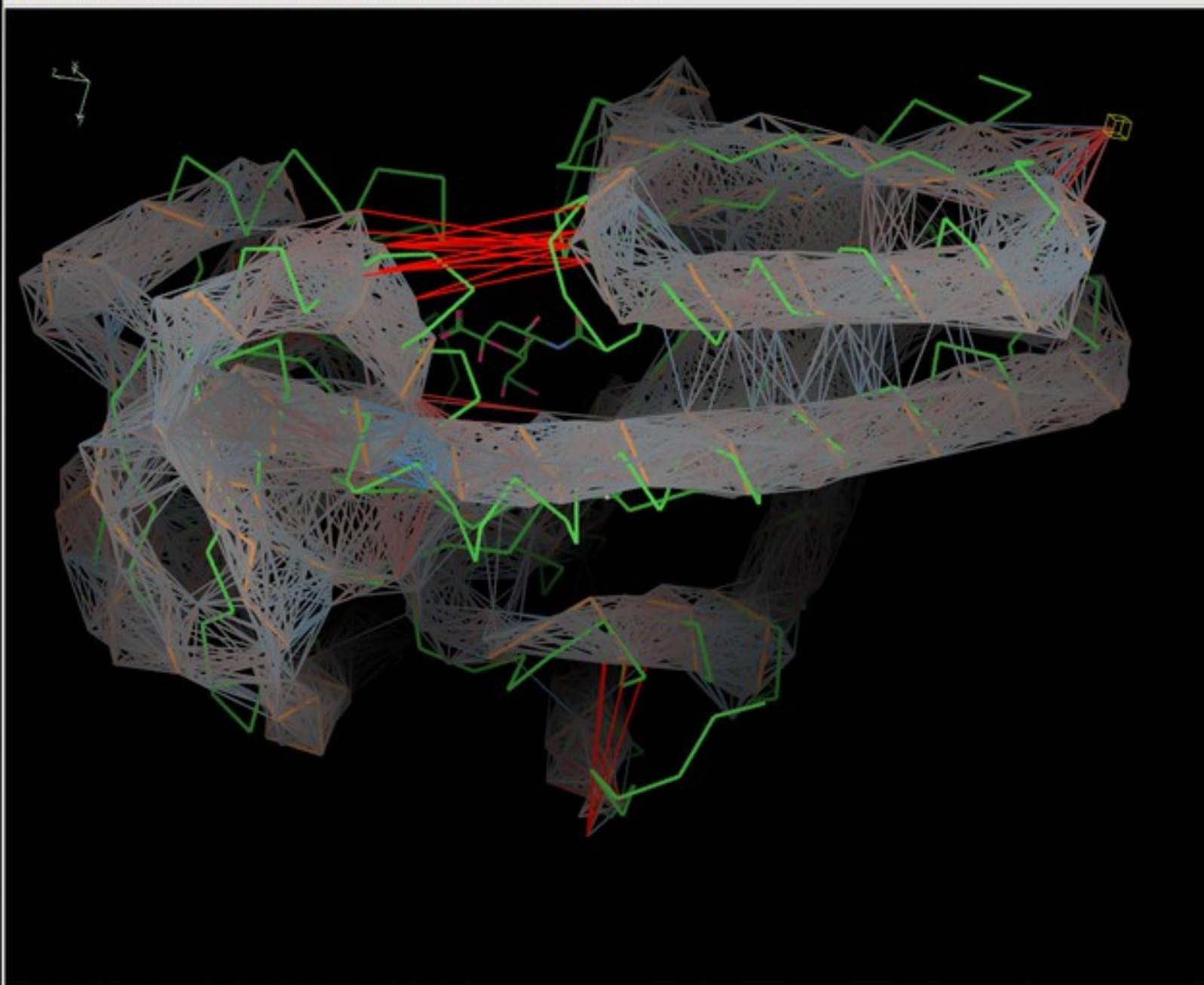
1



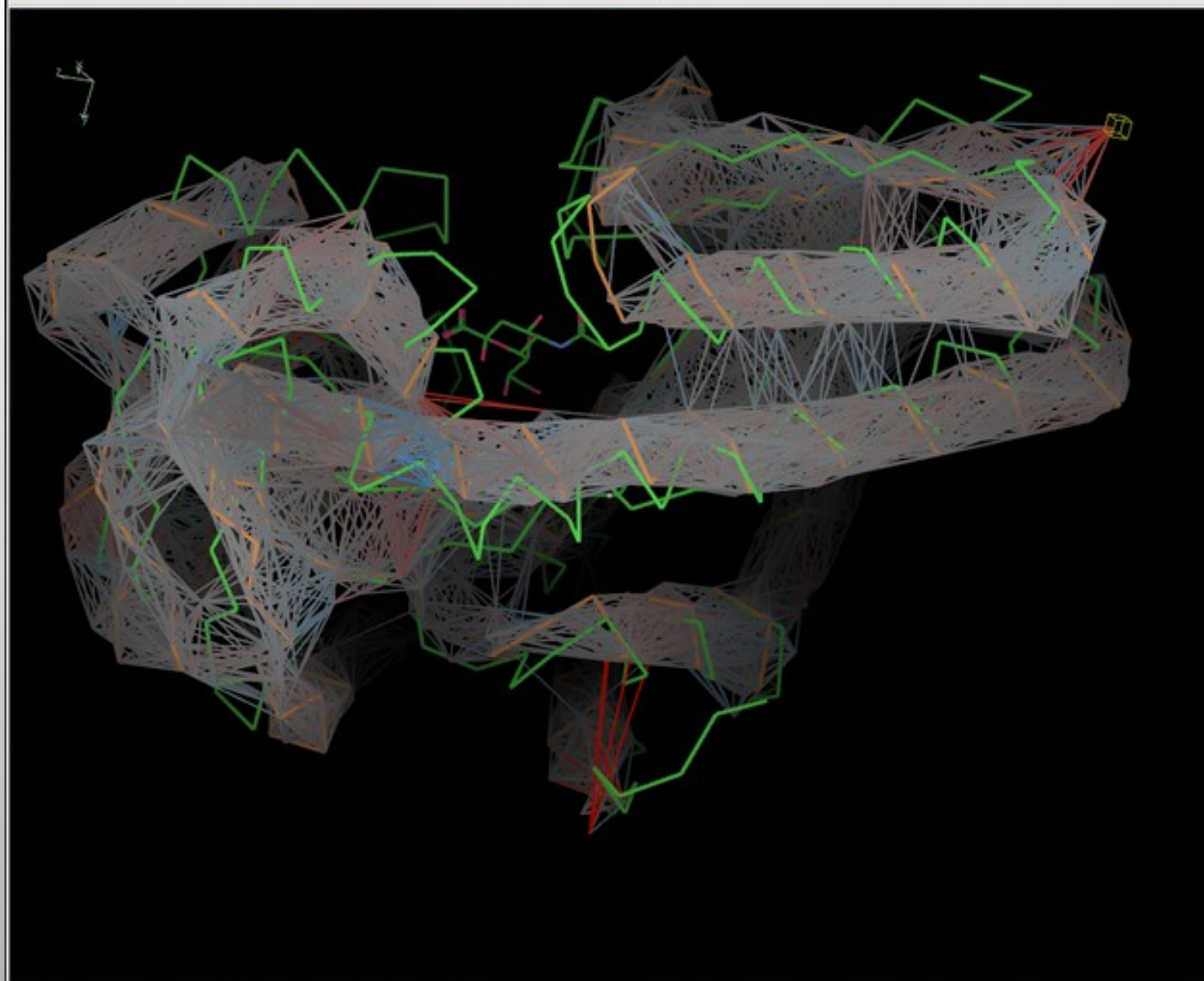
8







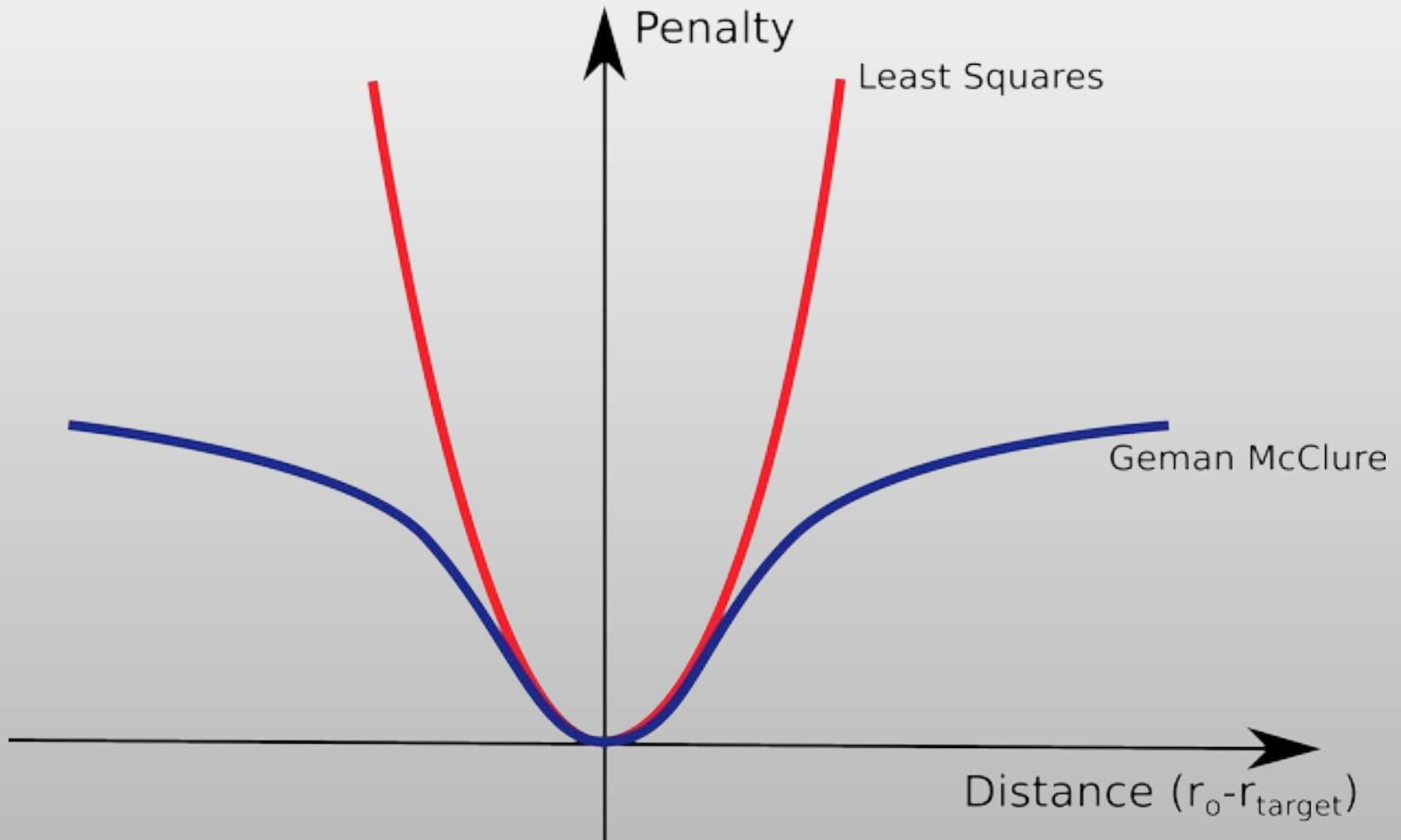




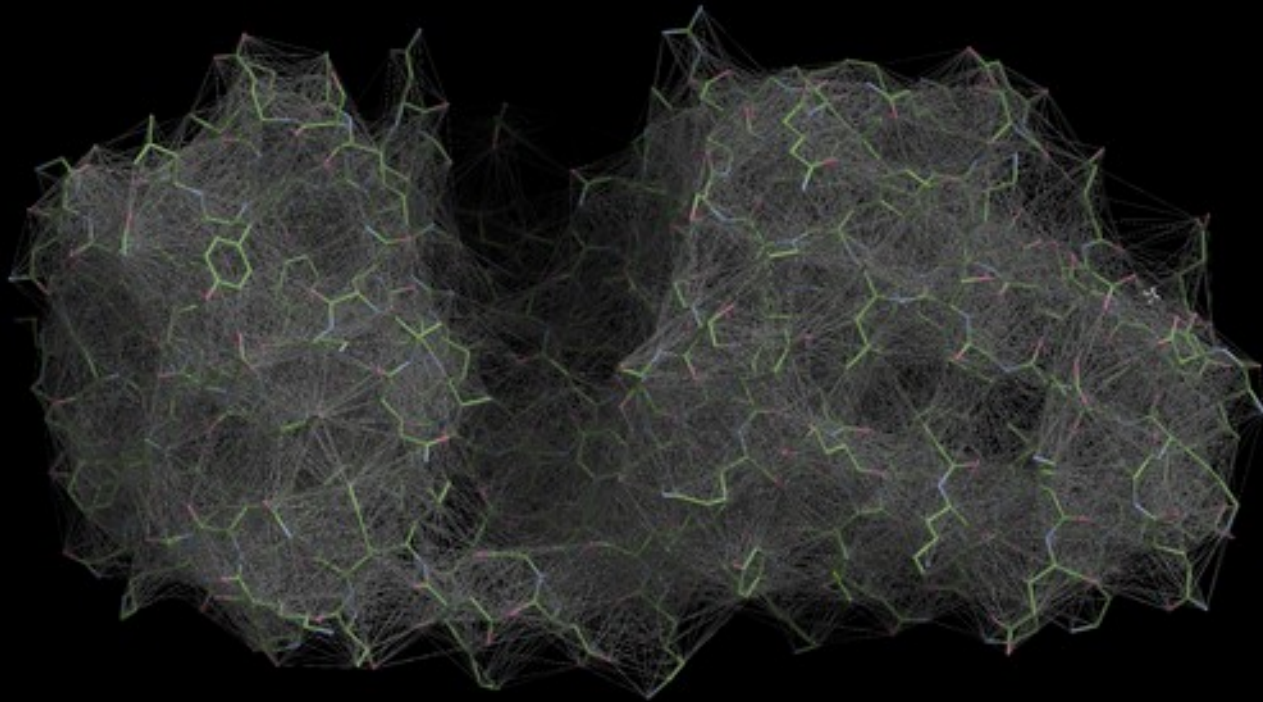
# ProSMART integration

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

# Modified Target Function



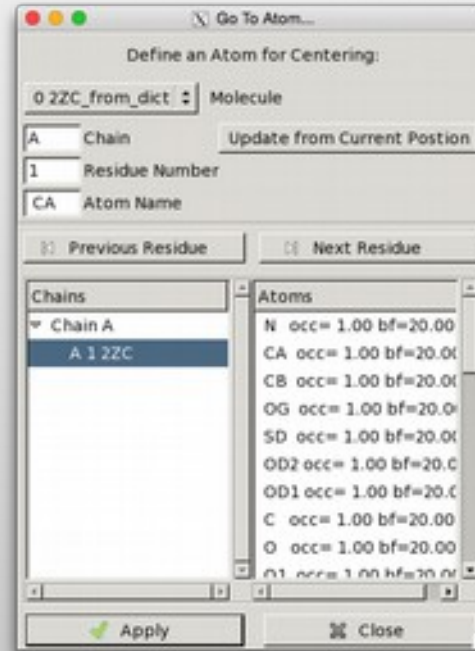
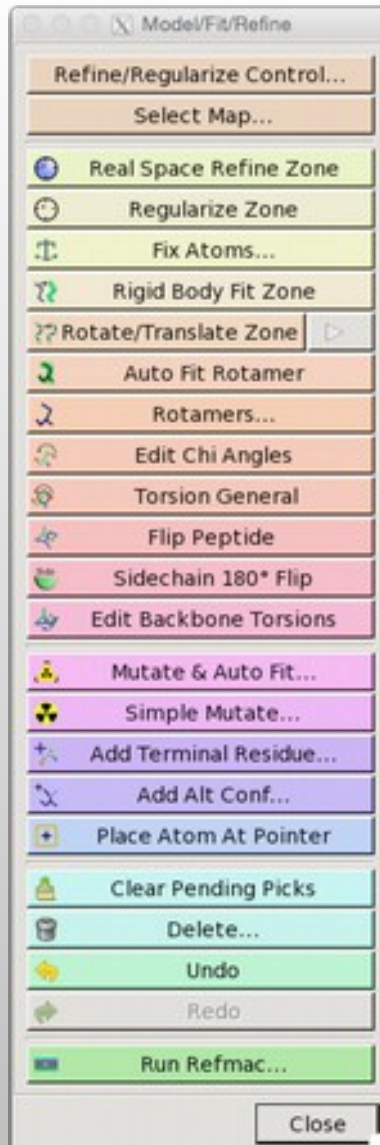
# ProSMART Restraints



# A note on *Coot's* GUI

- It used to be clean
- Now lots of features have been added without much thought
- “Somewhat difficult to navigate”
- “Hidden” hot-keys

# IISTDTIDIW



- If I See This Dialog Then I'm Doing It Wrong

# Refinement Techniques

- Single-Atom Drag
  - Over-dragging
- Key-bindings:
  - Triple Refine “T”, with auto-accept: “H”
  - Single Residue Refine: “R” with Auto-accept: “X”
  - Add Residue: “Y”
  - Autofit rotamer” “J”
    - Residue Flip: E, Shift: Opt-Alt- → Rotate: Ctl Shft - →
  - Hybridization-aware residue fragment rotation: “Shift F”

# Cool Key-binding Crib-Sheet

Cool Crib Sheet

October 16, 2014

## 1 Keyboard

### 1.1 Chasing Mousetrails

⌘-Track On To Active Window  
⌘-Track Display-Content Window

### 1.2 Previous/Next Window

⌘-Space? Next Window  
⌘-Shift? ⌘-Space? Previous Window

### 1.3 Command Modifier

⌘-? go to an item of the command palette (the "C" is what if the window has one)

### 1.4 Go To Modifier

⌘-G (Number number) <Enter>  
<Enter> to the given window (you can provide a window title)

### 1.5 Next/Next Chain

⌘-? <other ACS chain>

### 1.6 "Undo" More

⌘-? to undo the more recent actions sequentially (eg. move back after reentering after reading a new PDF file)

\*The chain of your undo/redo actions is limited to 100 (eg. when the window contains a text)

### 1.7 Previous/Next Window

⌘-Space? Next Window  
⌘-Shift? ⌘-Space? Previous Window

### 1.8 Keyboard CSI Angles

Instead of pressing the buttons in the CSI Angles button list, you can use keyboard "C" for CSI, "O" for CSI on.

### 1.9 Keyboard Copying

⌘-C? or ⌘-V? to change the content text

### 1.10 Keyboard Labelling

⌘-? to label chosen item

### 1.11 Quick Save As

⌘-S? to save the state and any connected molecules (to default file name)

### 1.12 Keyboard Modifier Info

⌘-Info then click on window to open Modifier Info dialog

\*Click "Info" and ⌘-? in the pop-up window, not the button dialog.

### 1.13 Keyboard Translation

⌘-Space? Push View (J? translation)  
⌘-Space? Pull View (J? translation)

### 1.14 Keyboard Undo/Redo

⌘-Z Undo last modification  
⌘-Y Redo last modification  
⌘-X Undo last mouse navigation

### 1.15 Editing

⌘-D Copy active molecule  
⌘-V Paste active molecule

### 1.16 Keyboard Zoom and Clip

⌘-Z Zoom out  
⌘-N Zoom in  
⌘-J Show clip  
⌘-F Toggle clip

### 1.17 Command

⌘-Command

### 1.18 Skeleton

⌘-Command skeleton around current point?

### 1.19 Continuous Rotate

⌘-Toggle continuous spin

### 1.20 Bates Mode

⌘-Toggle into Bates mode mode?

## 2 Mouse

Mouse actions are occasionally augmented with keyboard modifiers

⌘-Left Mouse Drag	Rotate view
⌘-Right Mouse Drag	Translate view
⌘-Left Mouse Click	Label atom
⌘-Right Mouse Drag	Zoom in and out
⌘-Right Mouse Drag	Change clipping and front screen Z

The movement is along an axis

⌘-Right/Down-left click	up-left/down-right click
⌘-Left/Down-right click	up-right/down-left click

⌘-Right Mouse Drag Rotate View about Screen  
Middle Mouse Click Center on atom  
Scroll wheel Forward Increase map rotation time  
Scroll wheel Backward Decrease map rotation time

Intermediate (rotated) atoms can be dragged around by clicking on them

⌘-Left Mouse Drag	Move all intermediate atoms by same value as shown with
⌘-Left Mouse Drag with "A" key	move front atom
⌘-Left Mouse Drag with "V" key	Move a single atom

## 3 Refinement Extras

Use "A" to define a rotation angle? with a single click. Useful in Refinement and Regeneration.

- Click "Red/Space Before Zero"
- Click on an atom
- Press the "V" key

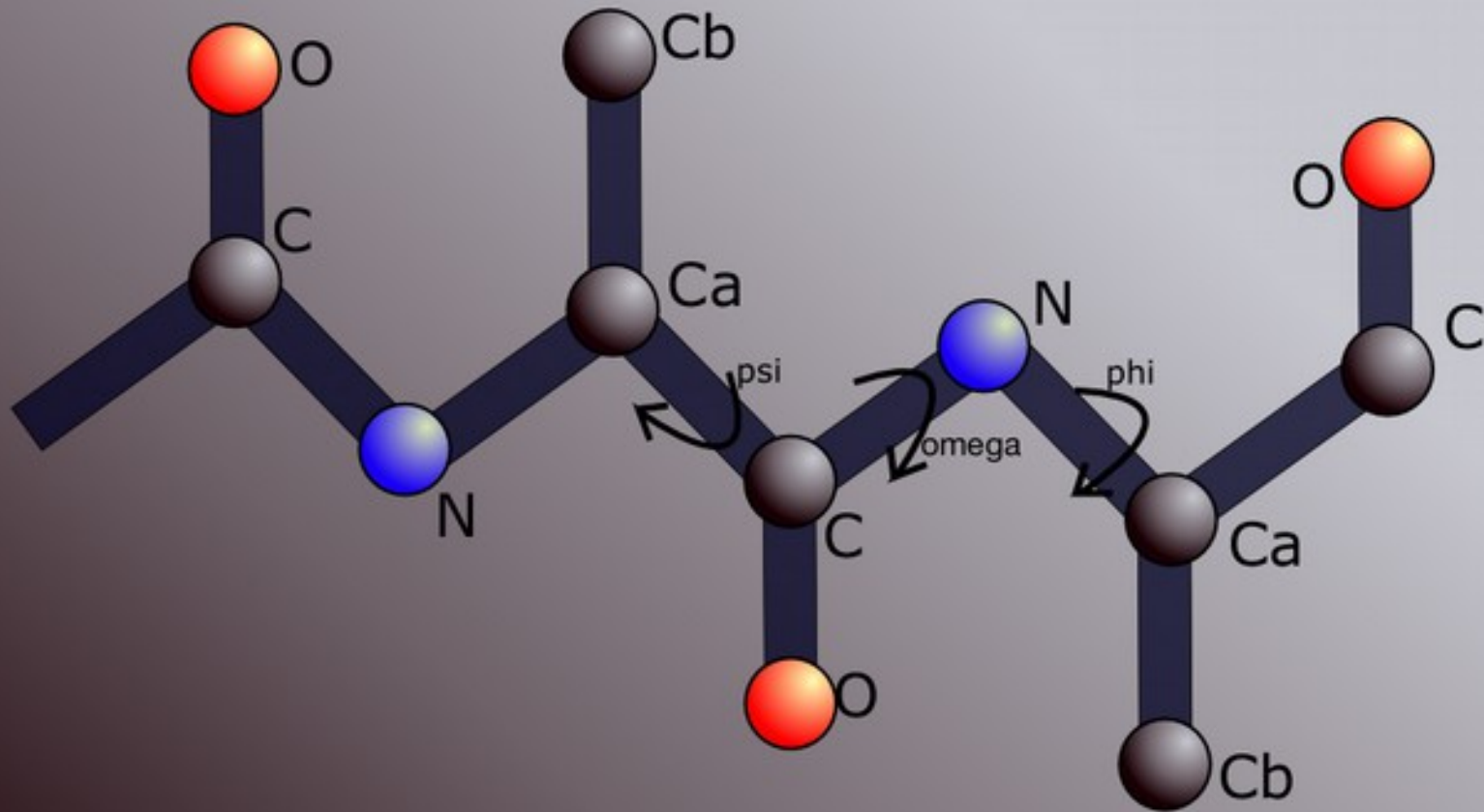
\*? is available in being displayed before the other atoms mode

## 4 Template Key-Bindings

E	File Legend
C	On To Back (under context)
H	Toggle/Show editor
I	Single File This Window
R	PDF Format Selection
B	Reduce Active Window
T	Toggle Window Before
X	Reduce Active Window and Auto-escape
A	Auto-escape
V	Auto-Resized Window
⌘-Q	Remove Dialog for Window
⌘-B	System Before
⌘-S	Accept State Position
⌘-R	System Regeneration
⌘-F	Delete Window/Dialogs
⌘-V	Undo Symmetry View
⌘-N	Save CSI Angles
⌘-M	Auto-Name to File
⌘-A	Red and Back to Legend



# Peptide Torsion Angles



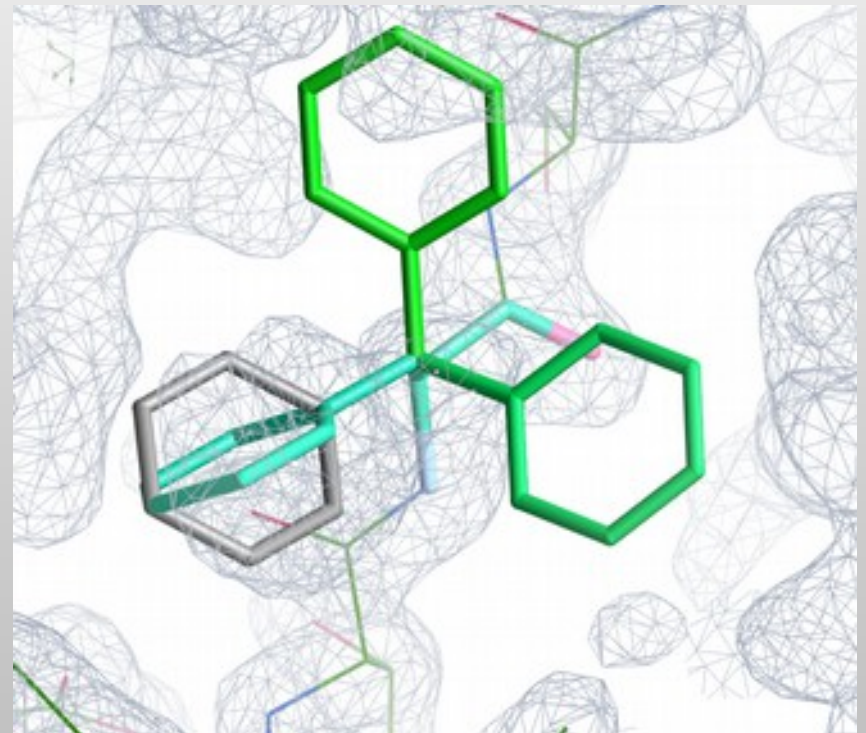
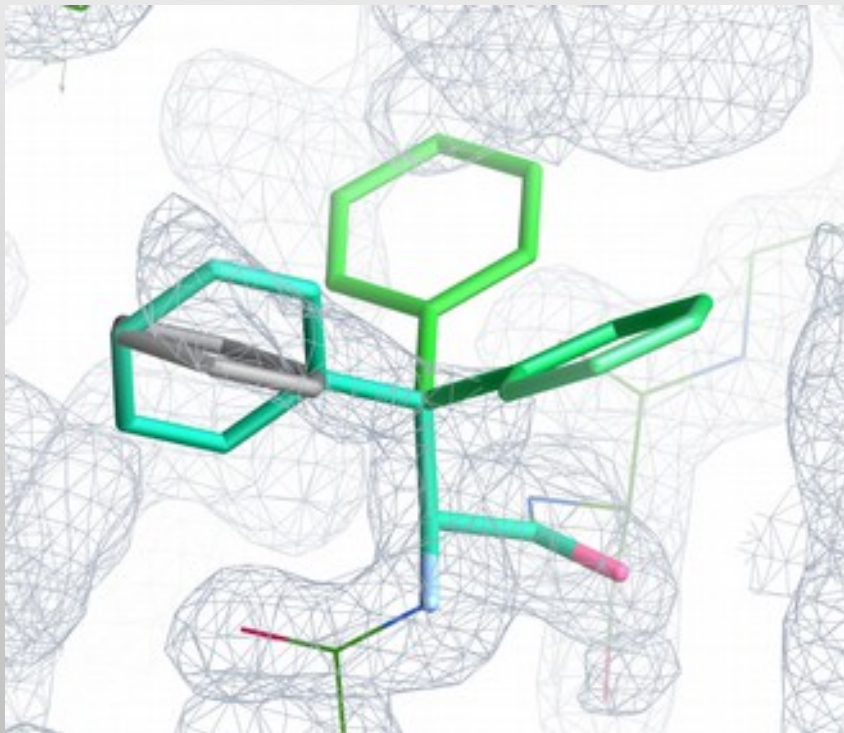
# Rotamer Searching

- Two methods
  - Traditional
  - Backrub

# Rotamers

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

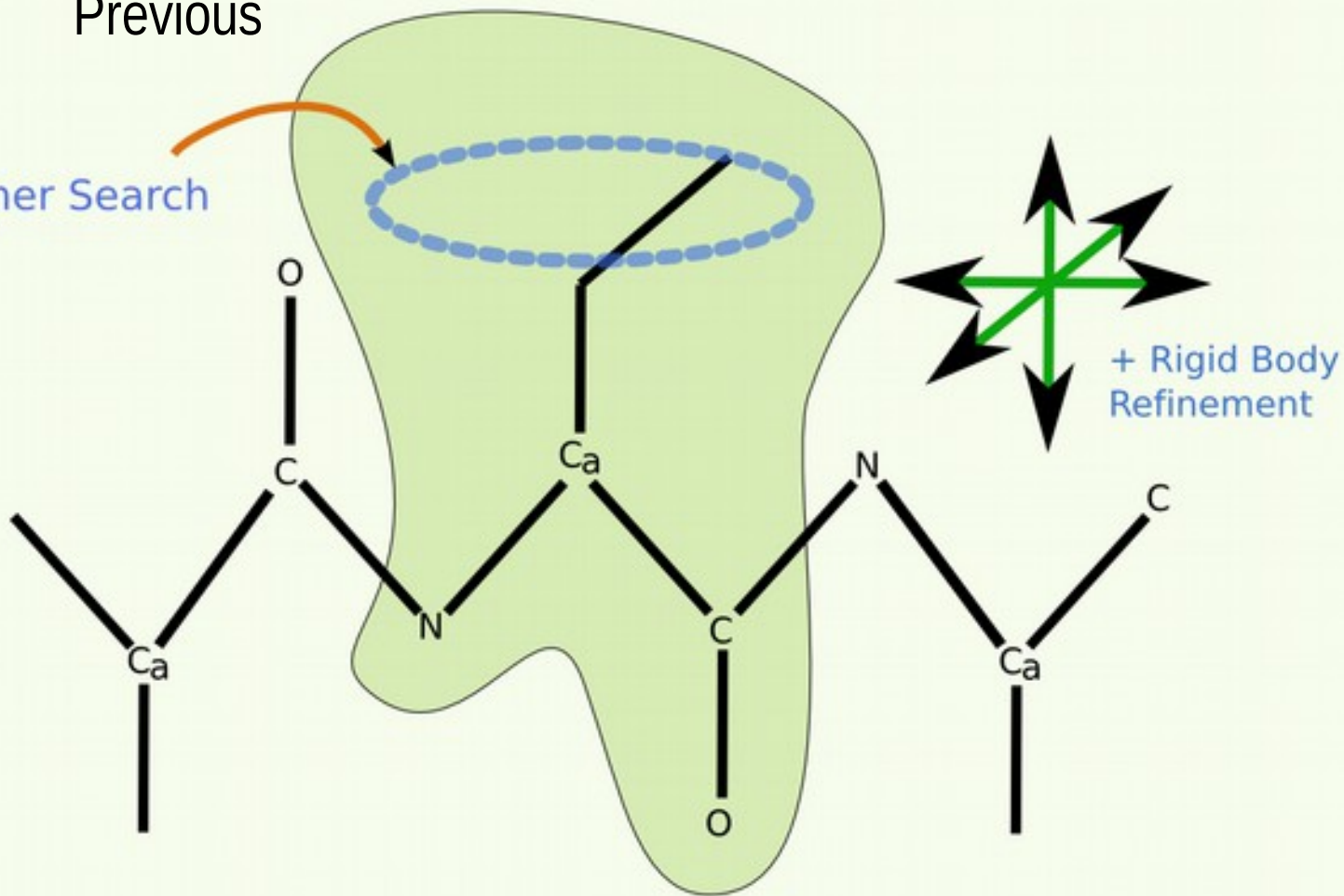
# 4 PHE Rotamers

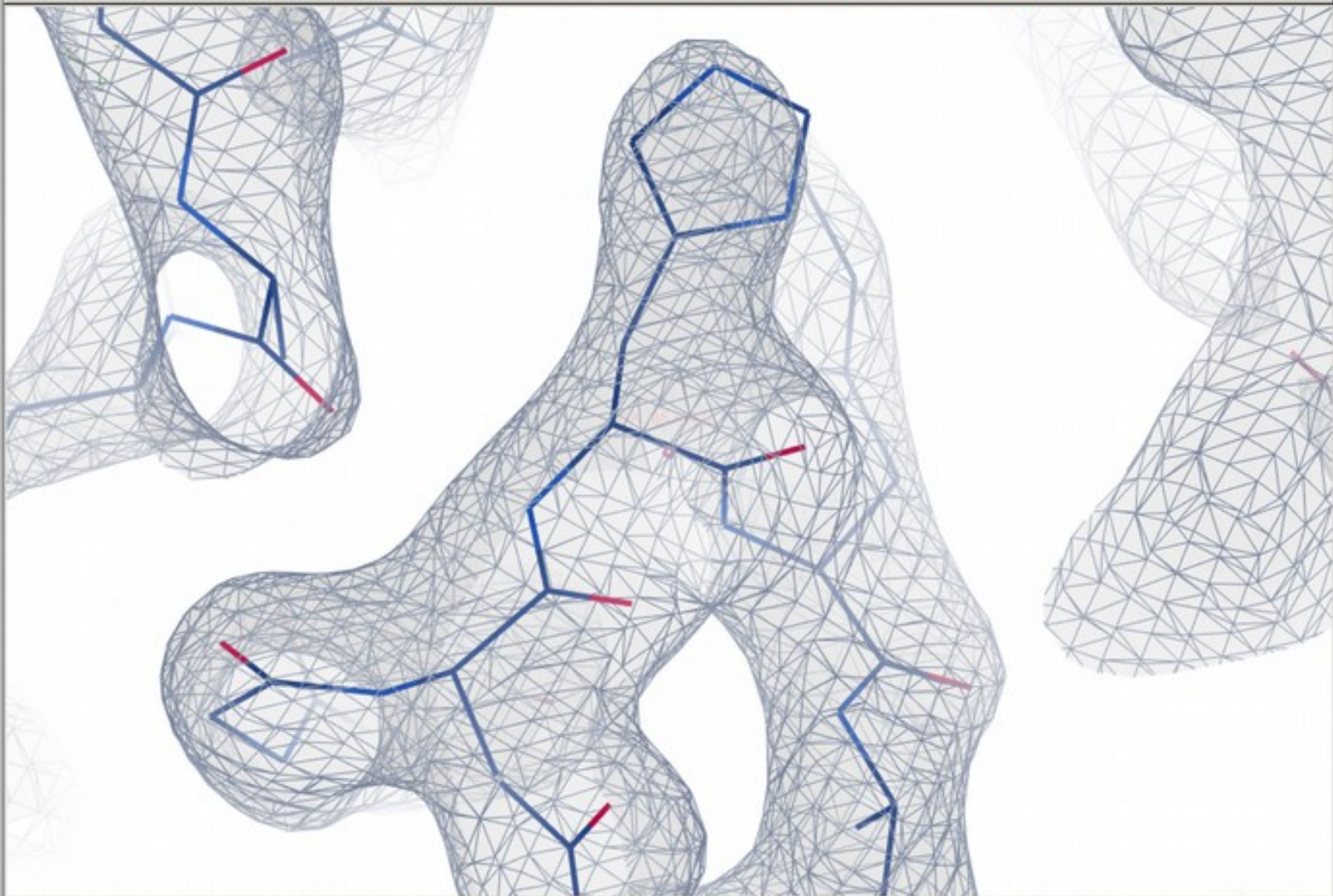


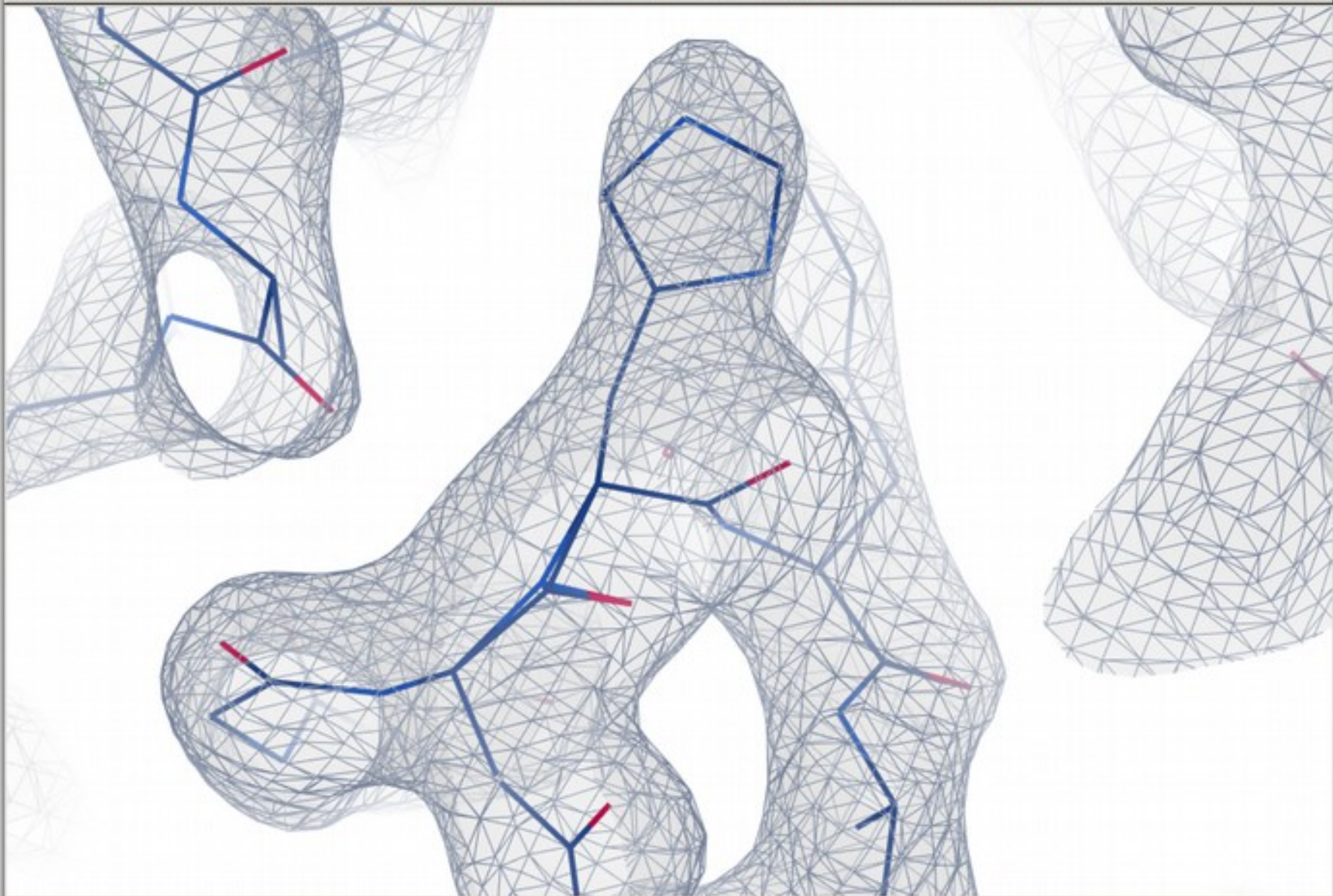
# ~~Current~~ Low Resolution Rotamer Search

Previous

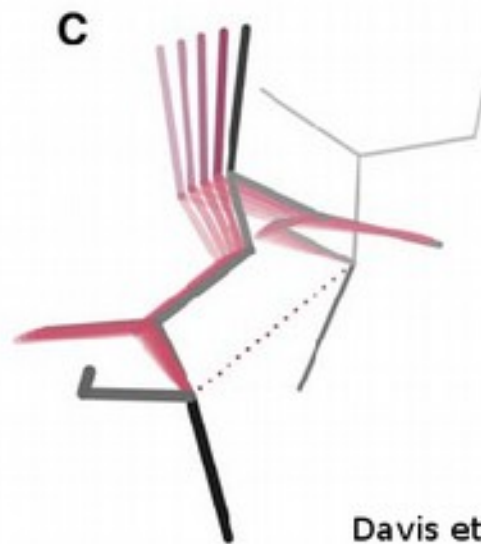
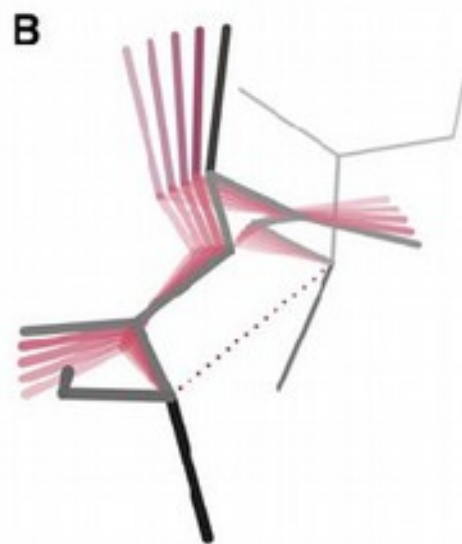
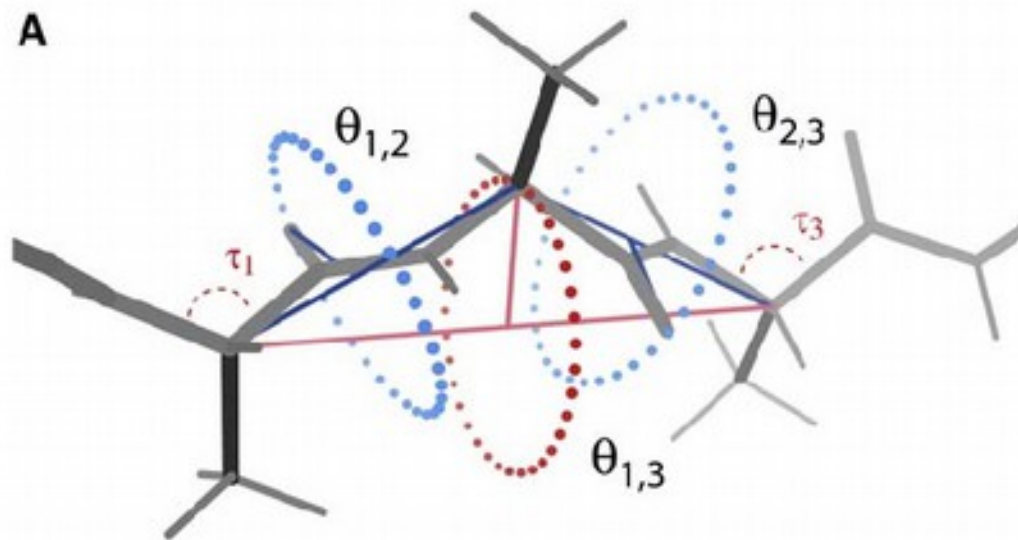
Rotamer Search







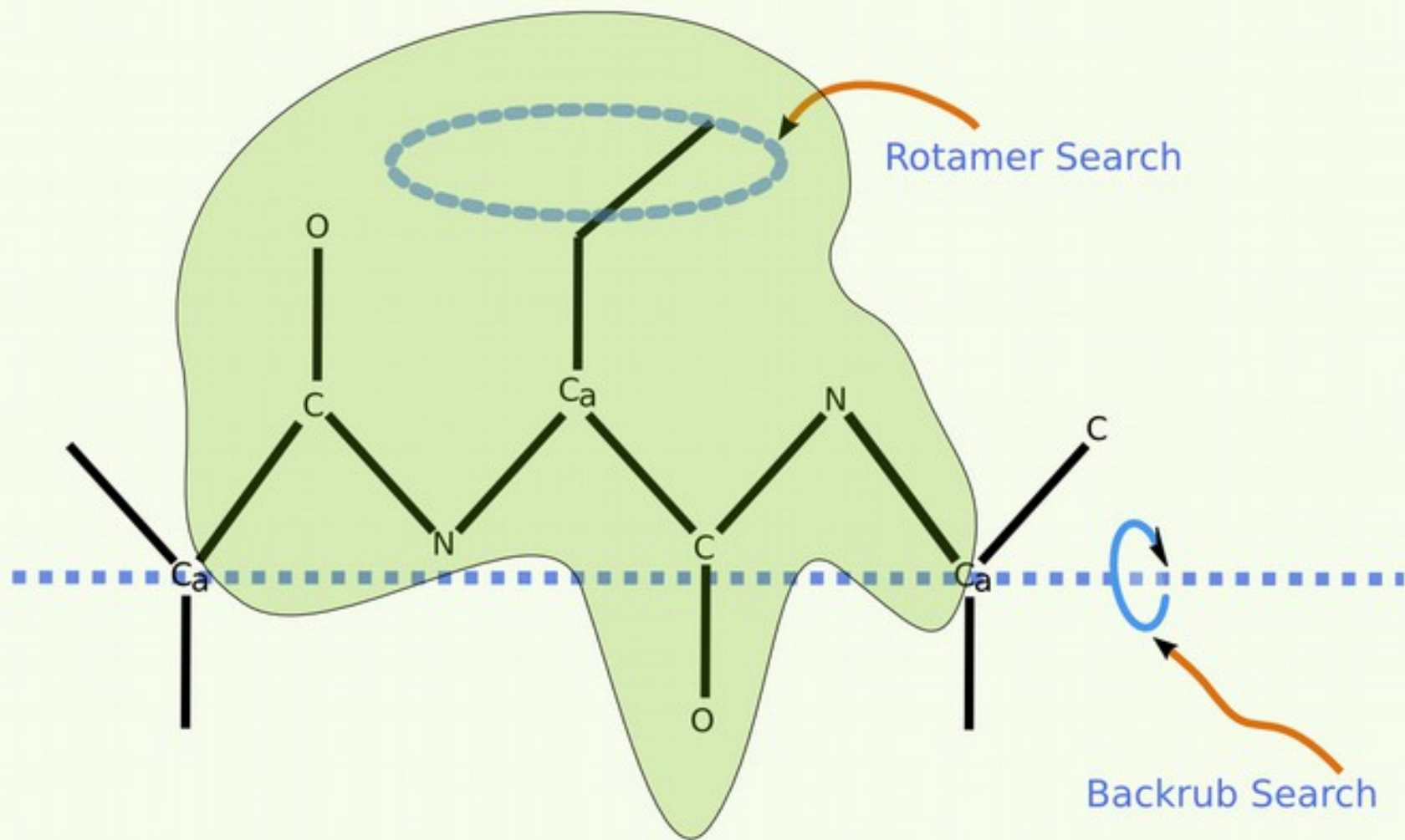
Vertical toolbar containing various icons for map display, rotation, translation, and other molecular modeling functions.



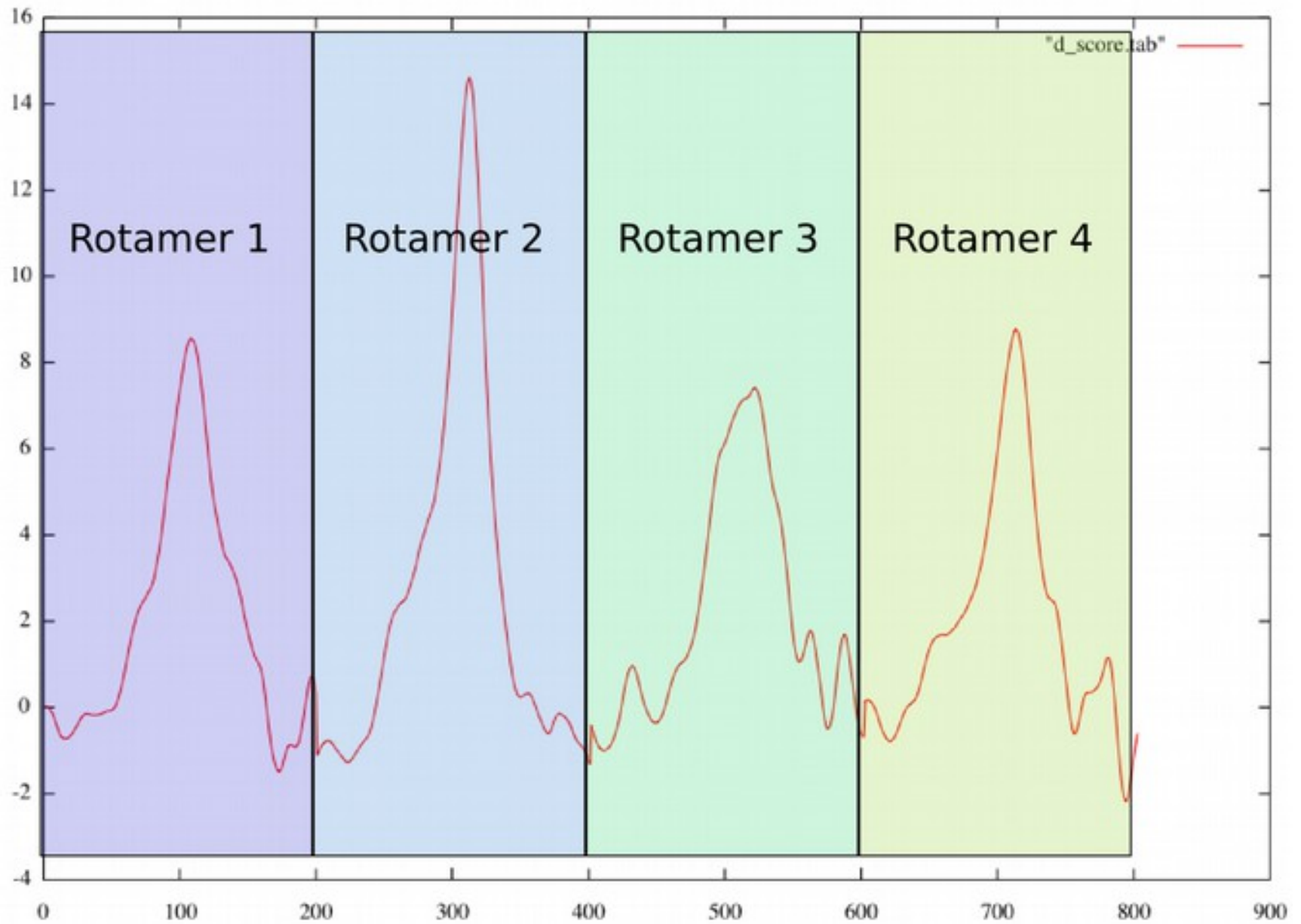
Davis et al. (2006) Structure



# New Low Resolution Rotamer Search



After Fitting Tools in KING/Molprobity





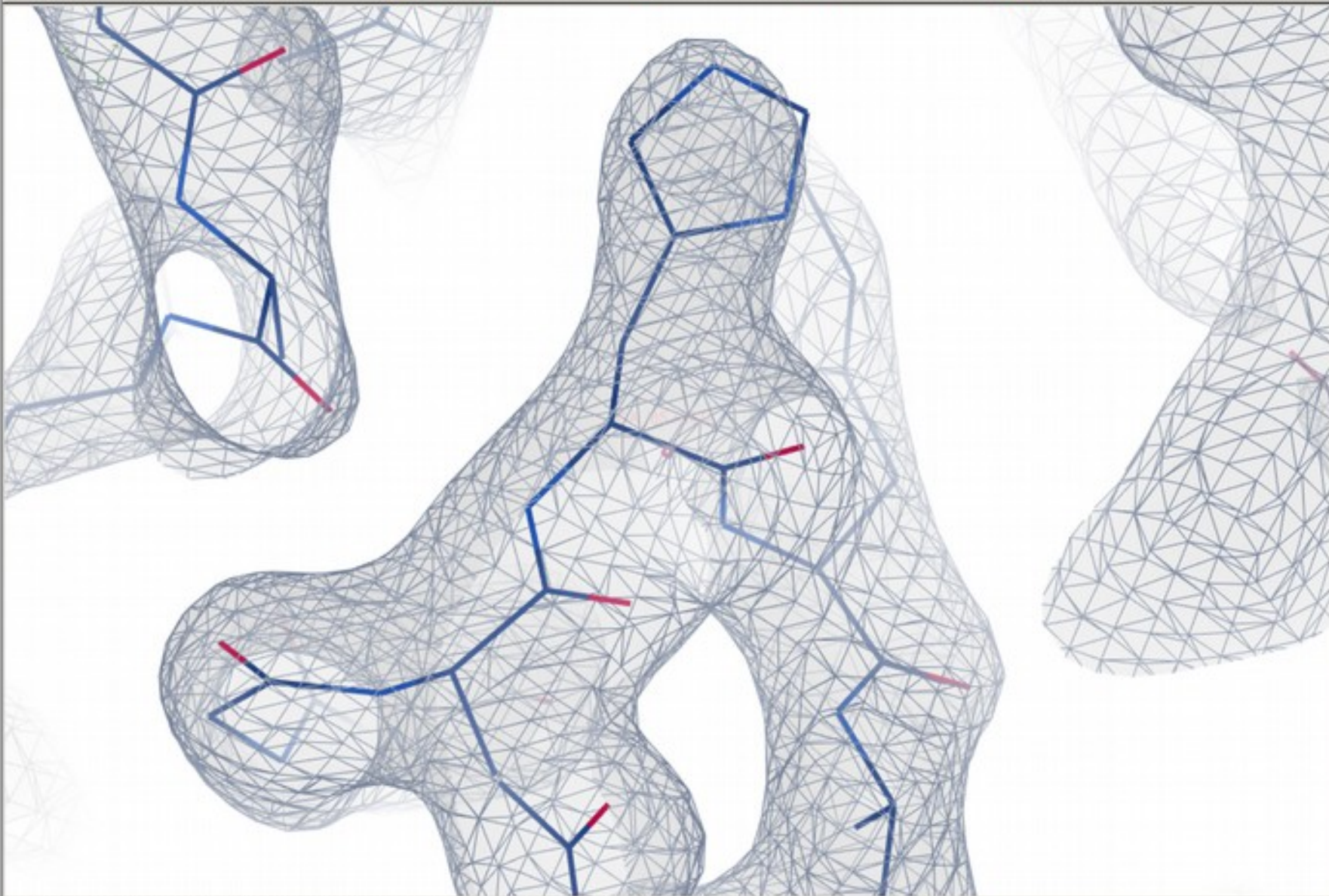
Coot 0.8.7-pre EL (revision count 6456)

File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug

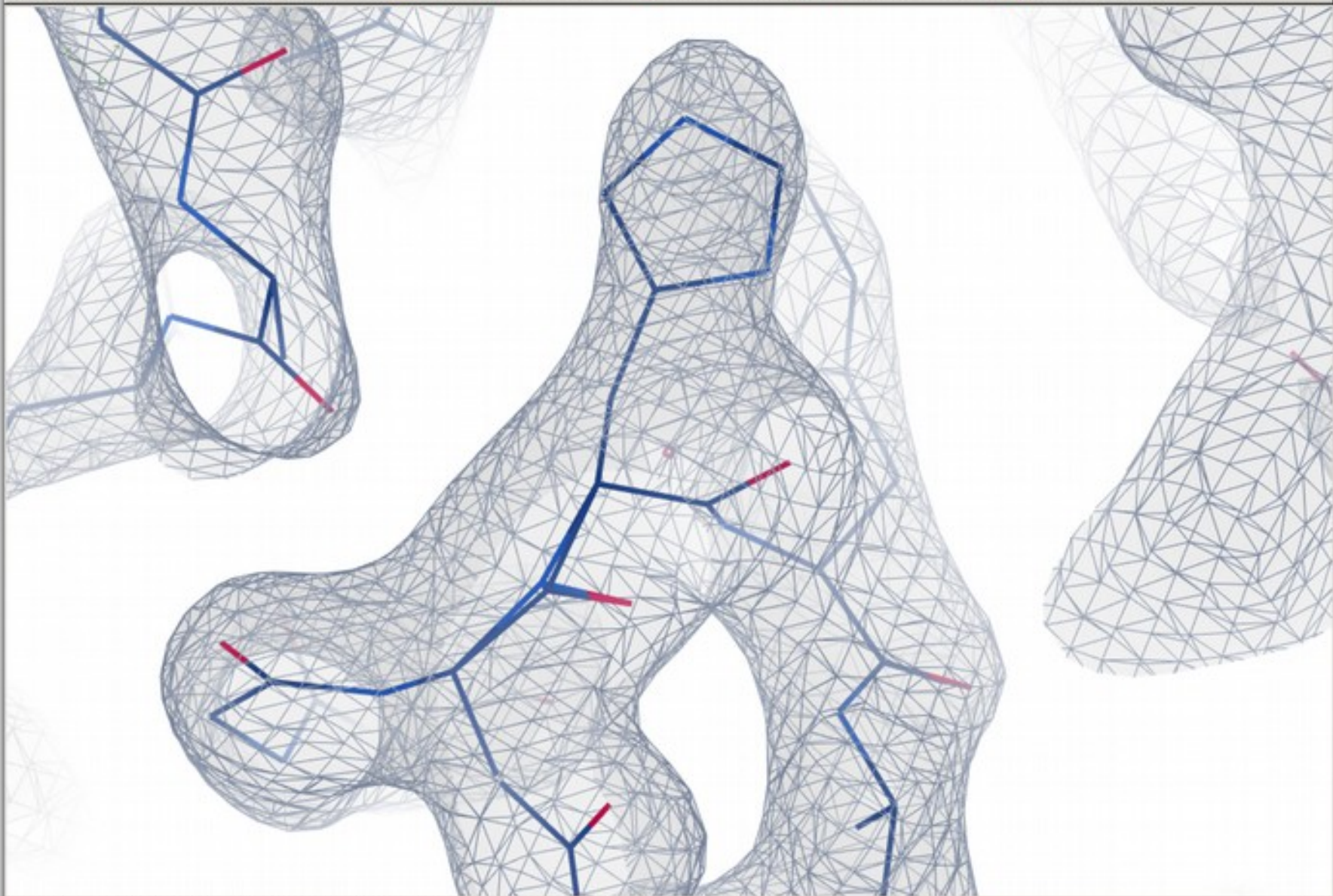
Reset View Display Manager Ligand Builder Sphere Refine Sphere Refine + Backrub Rotamers

R/RC

Map



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





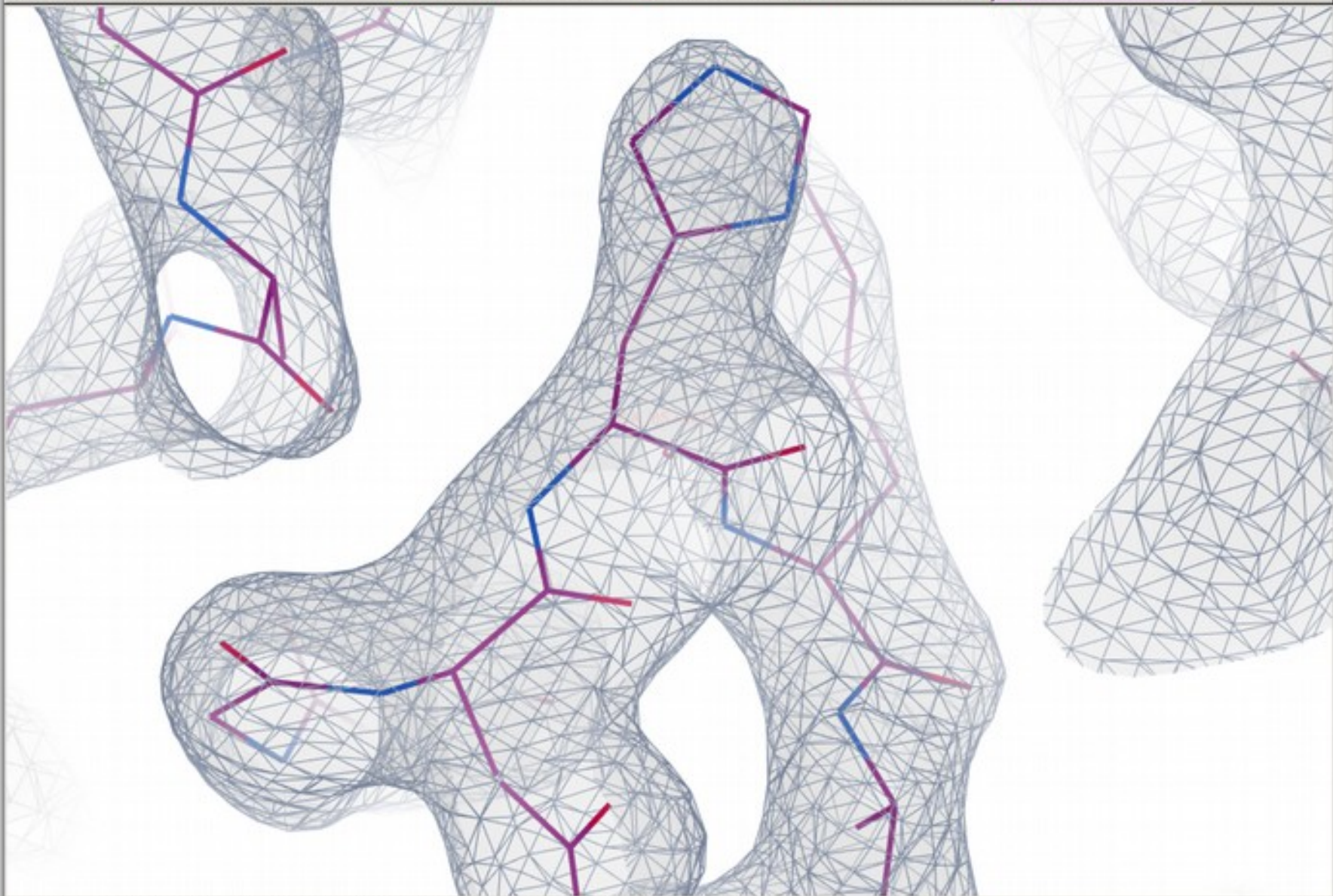
Coot 0.8.7-pre EL (revision count 6456)

File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug

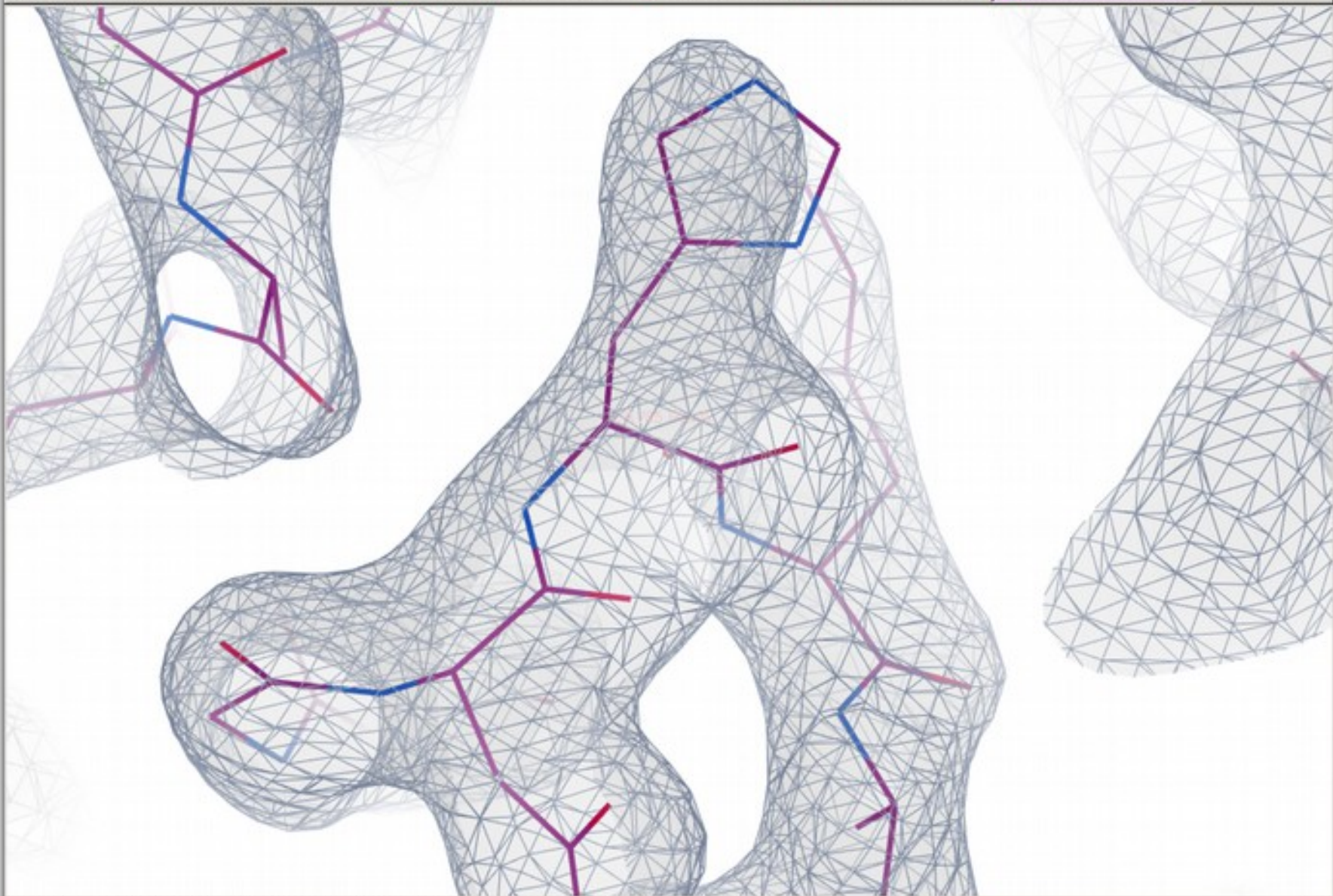
Reset View Display Manager Ligand Builder Sphere Refine Sphere Refine + Backrub Rotamers

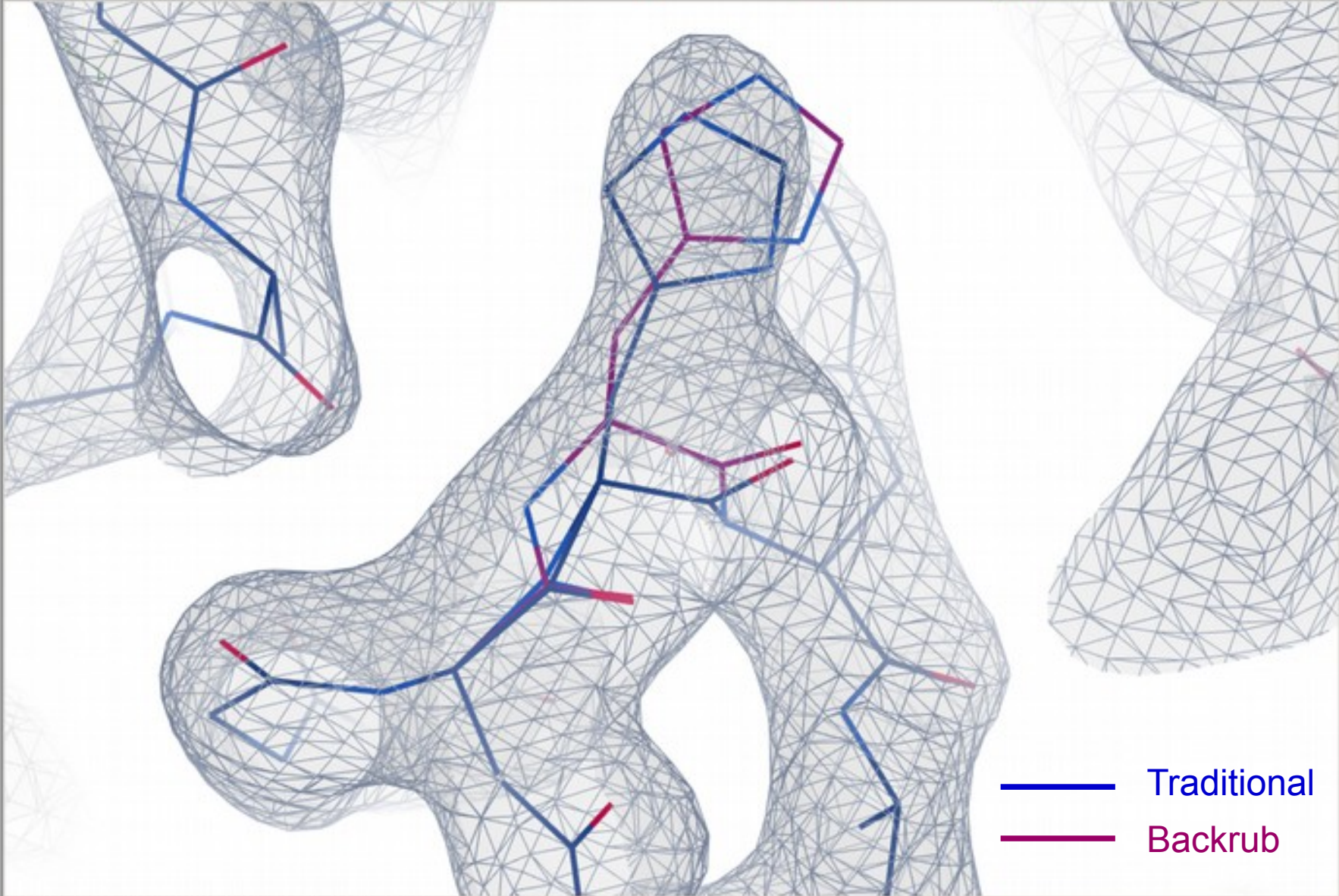
R/RC

Map



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

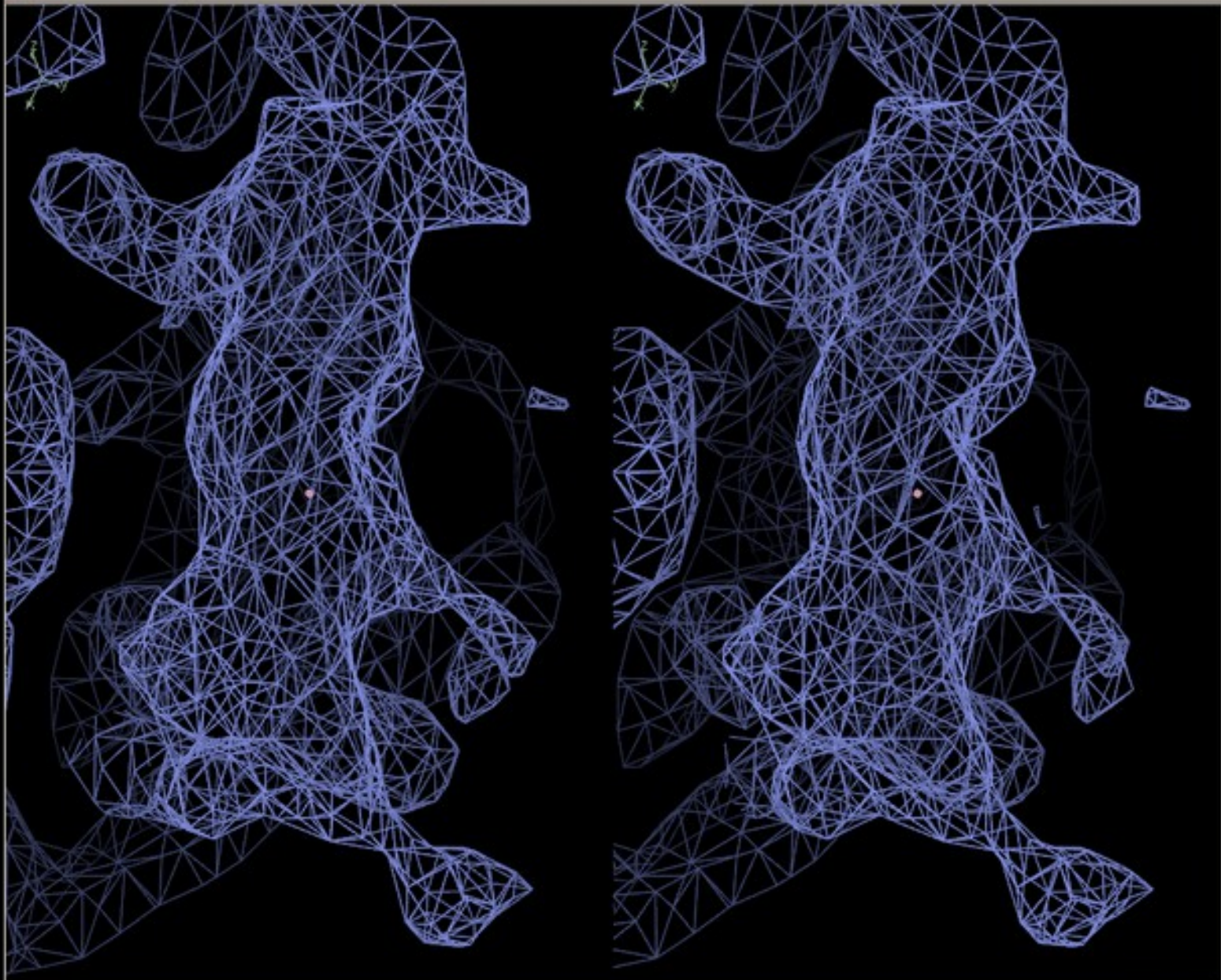




— Traditional  
— Backrub

# Helix-Building

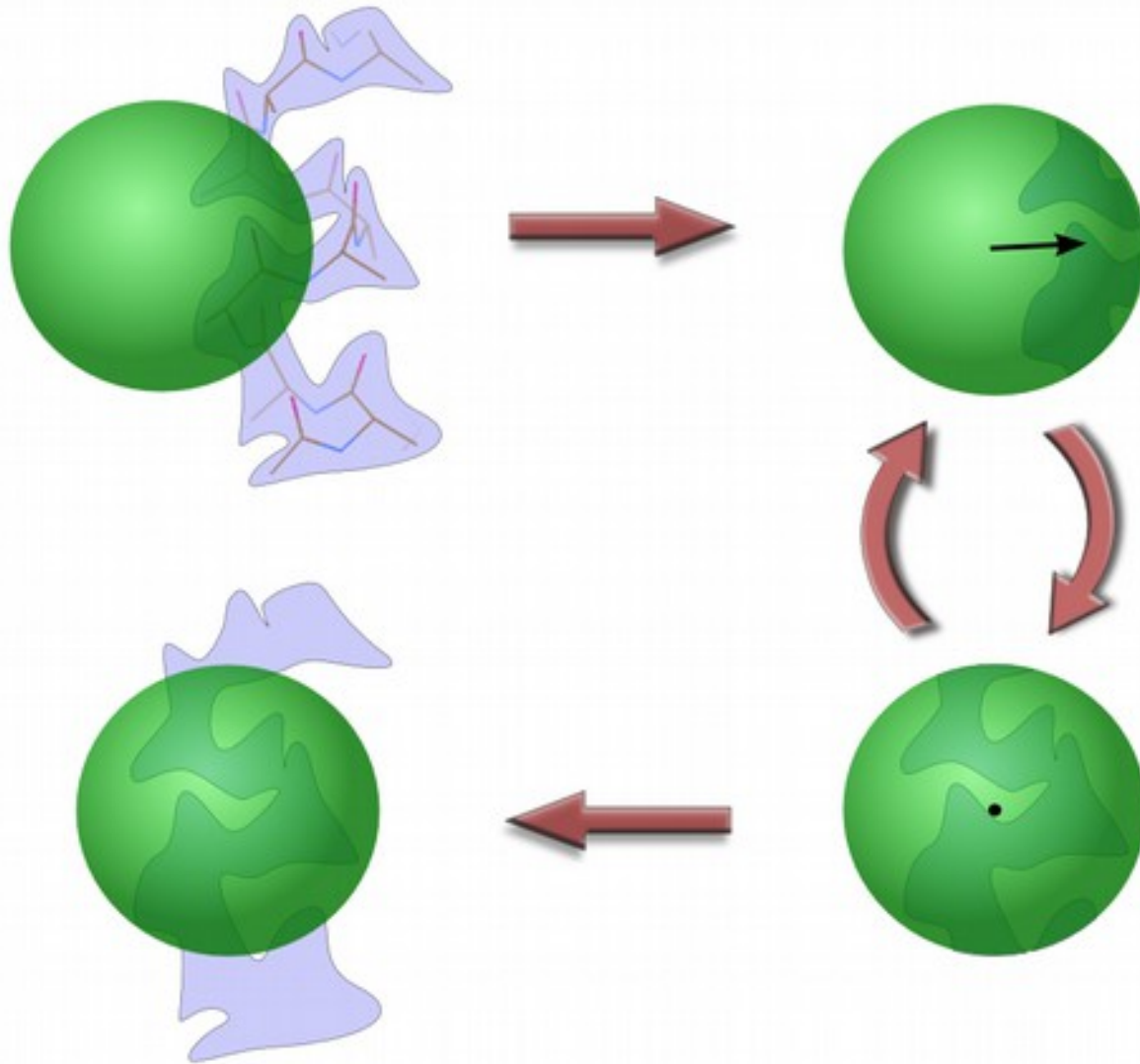




# Alpha Helix Placement

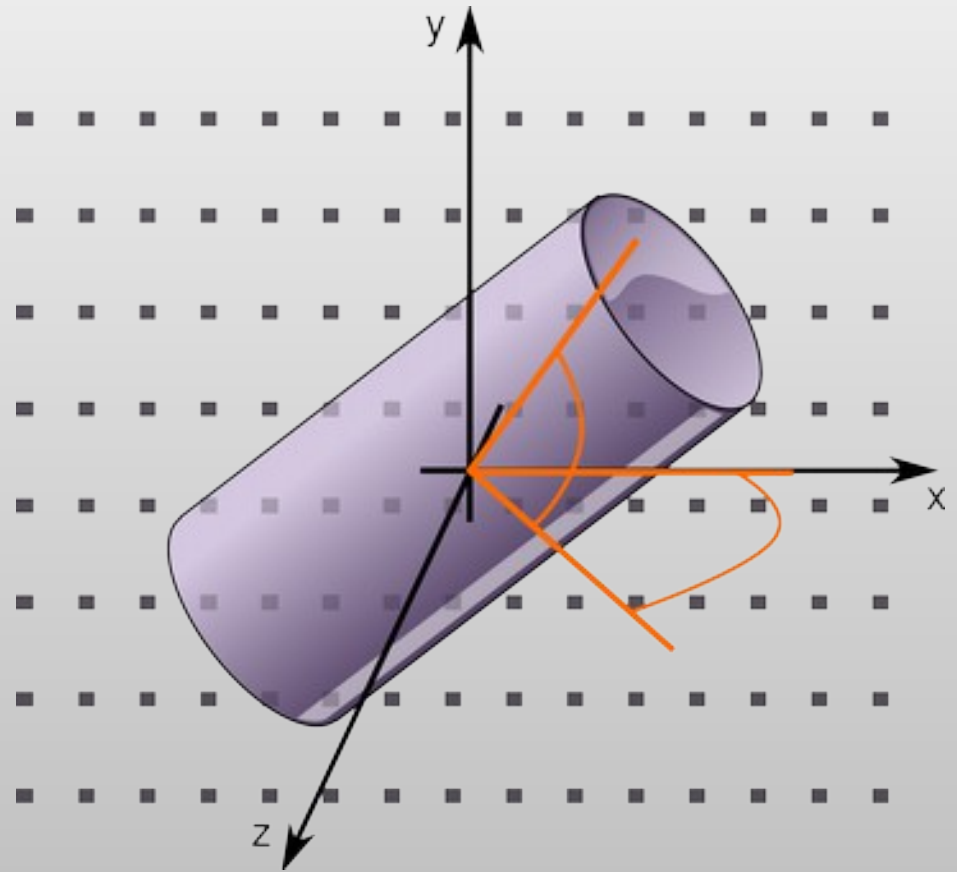
- **Scenario: Looking at a new map, not built with automatic tools:**
  - “I can see that there’s a helix here - build it for me!”
- **From a given point:**
  - Move to local averaged maximum
  - Do a 2D MR-style orientation search on a cylinder of electron density
  - Build a helix (both directions)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

# Centering the Rotation point

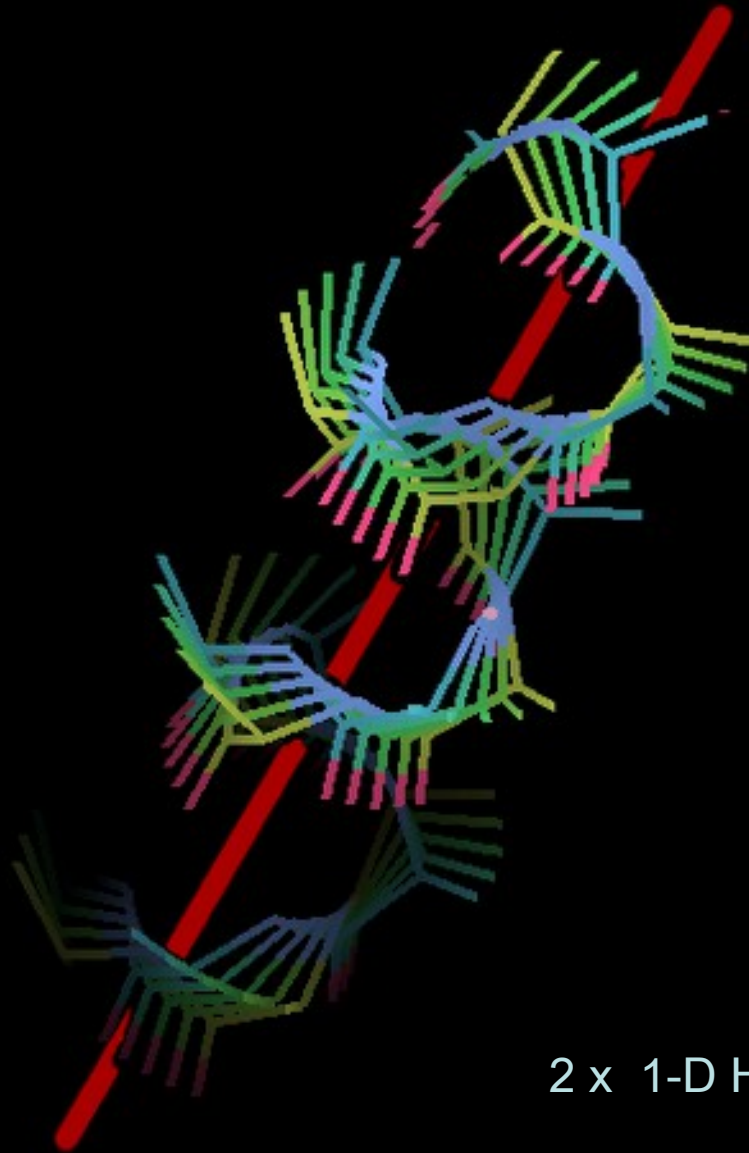


# Helix Fitting: Cylinder Search

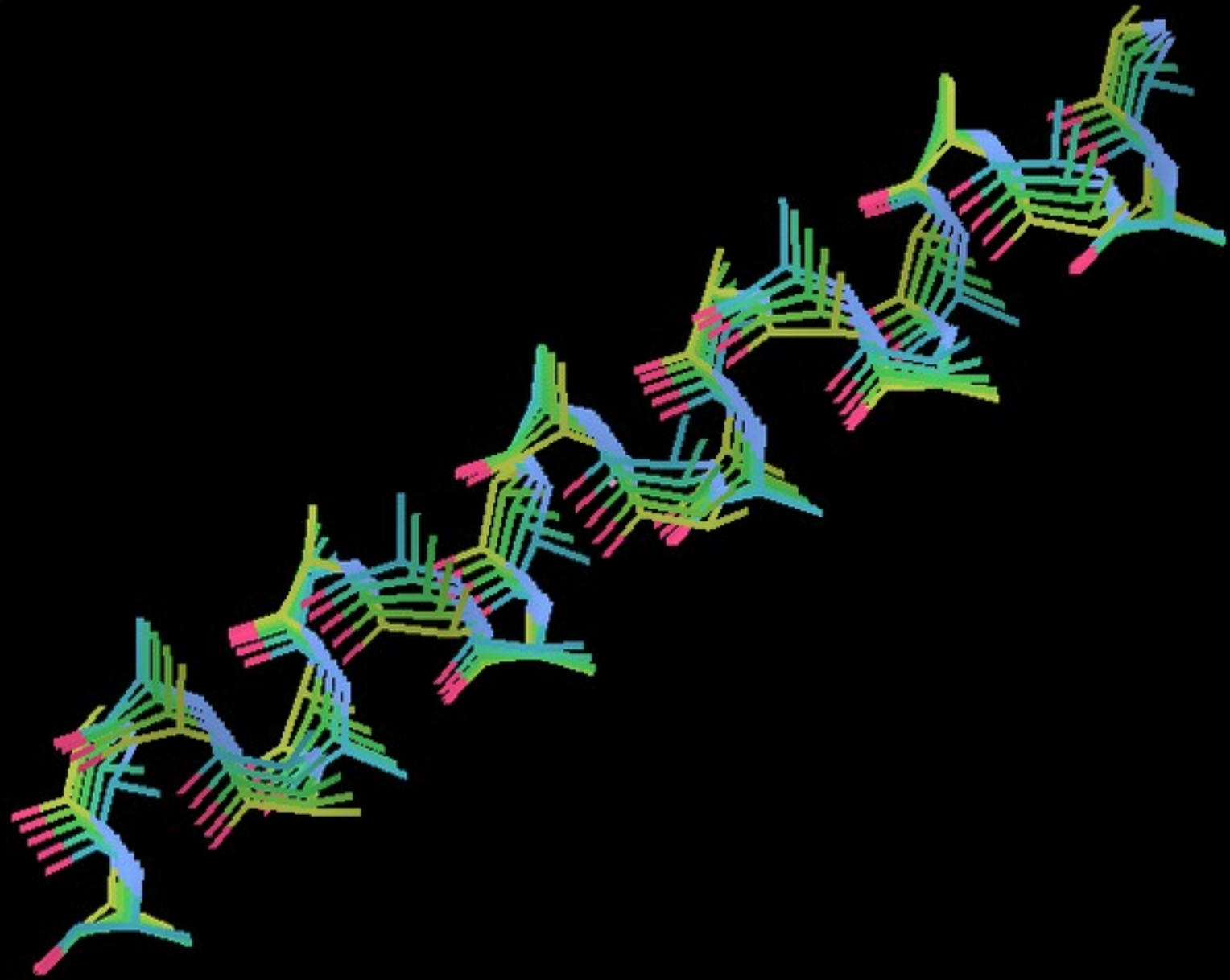
- Pick the orientation that encapsulates the most electron density

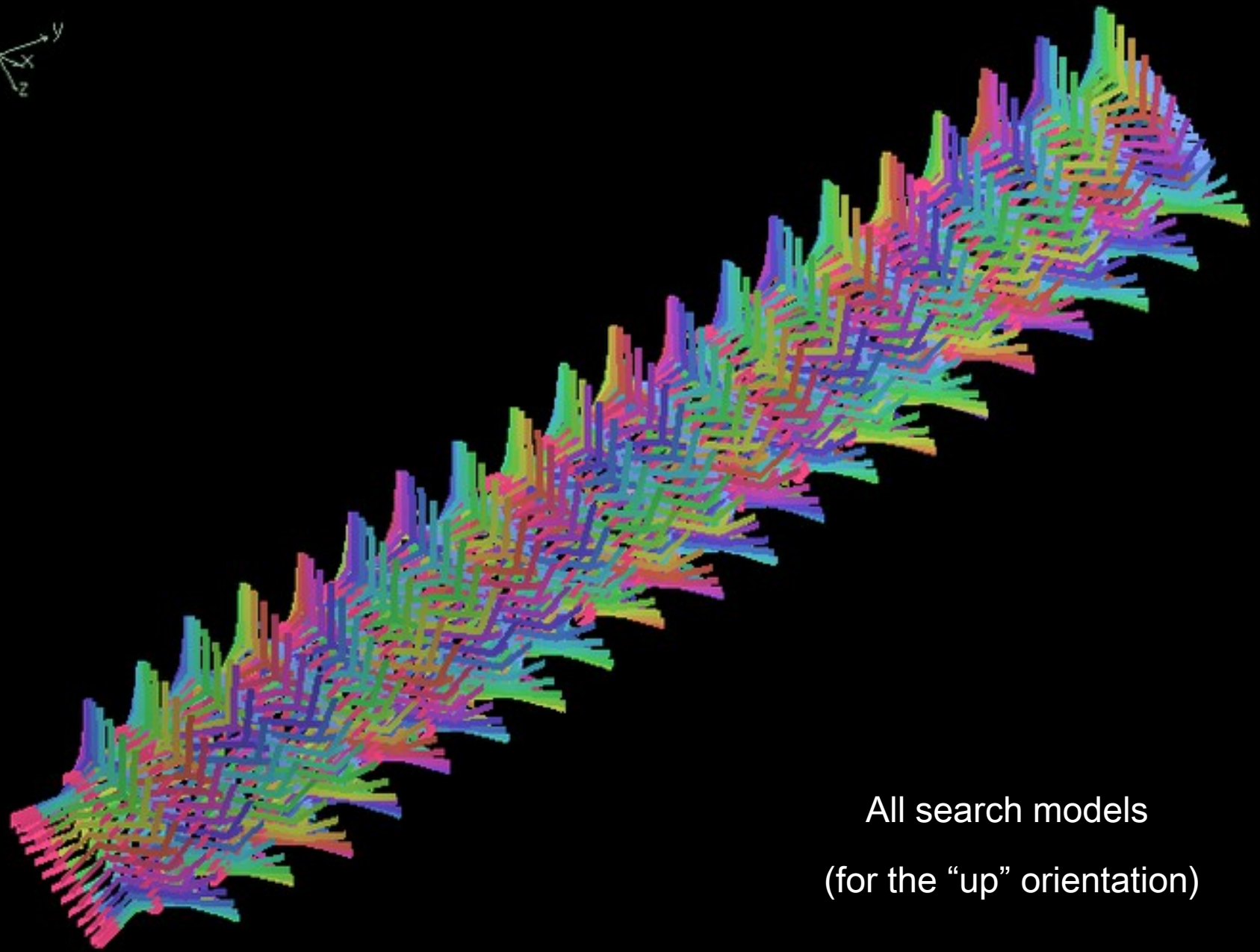


Using 2 rotation axes

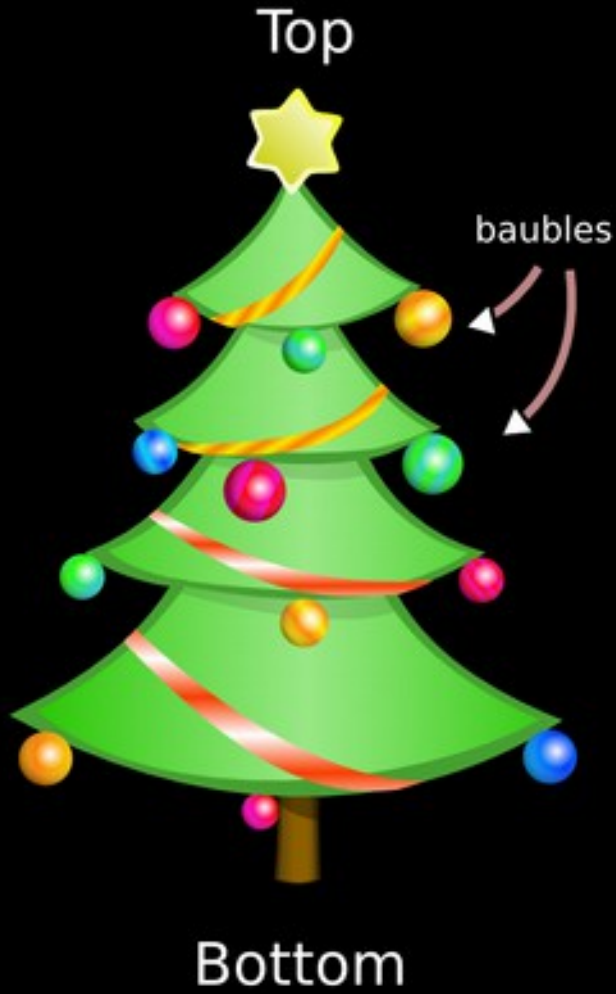


2 x 1-D Helix orientation searches





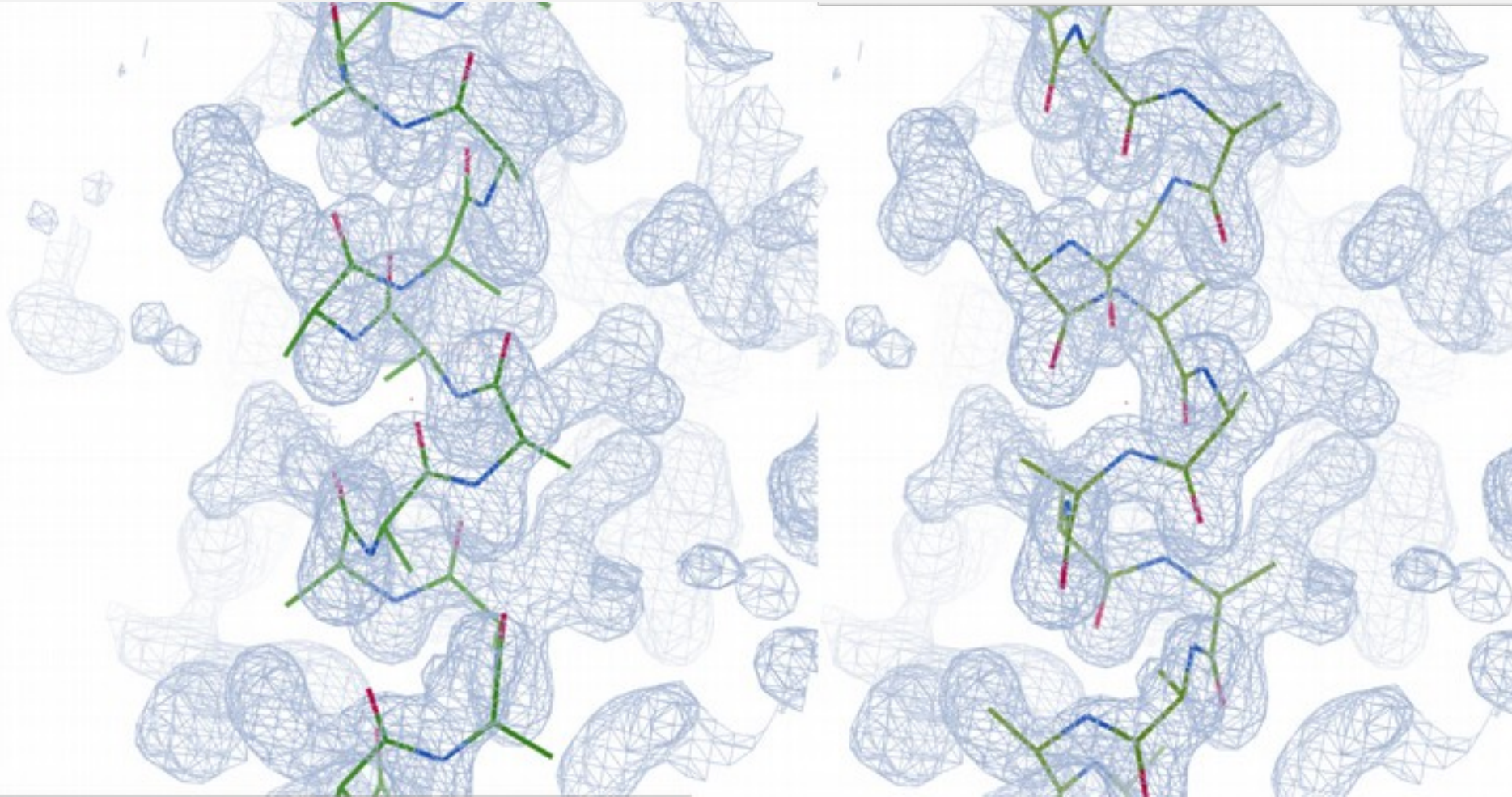
All search models  
(for the “up” orientation)





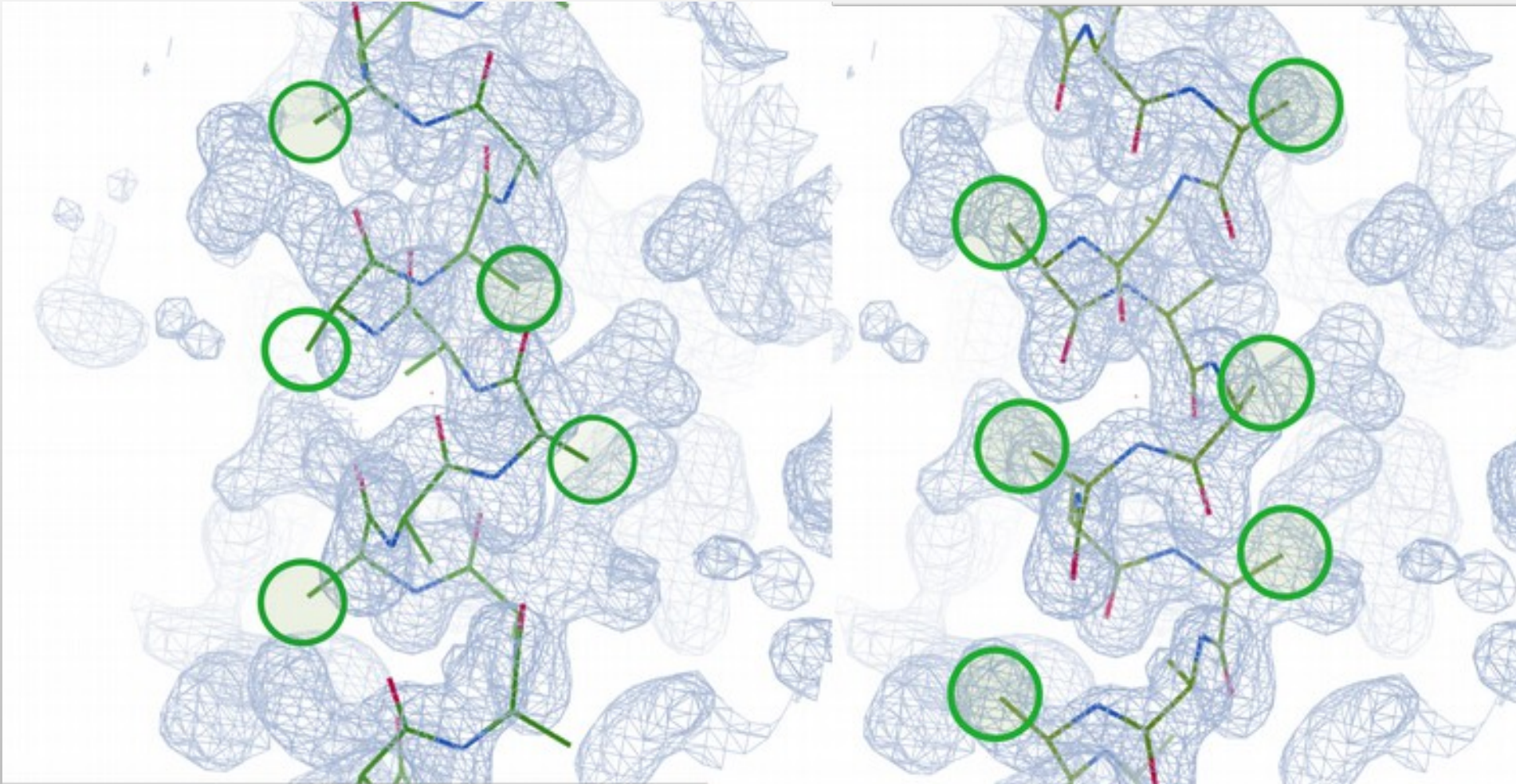
# Helix Fitting

## Comparing orientation hypotheses

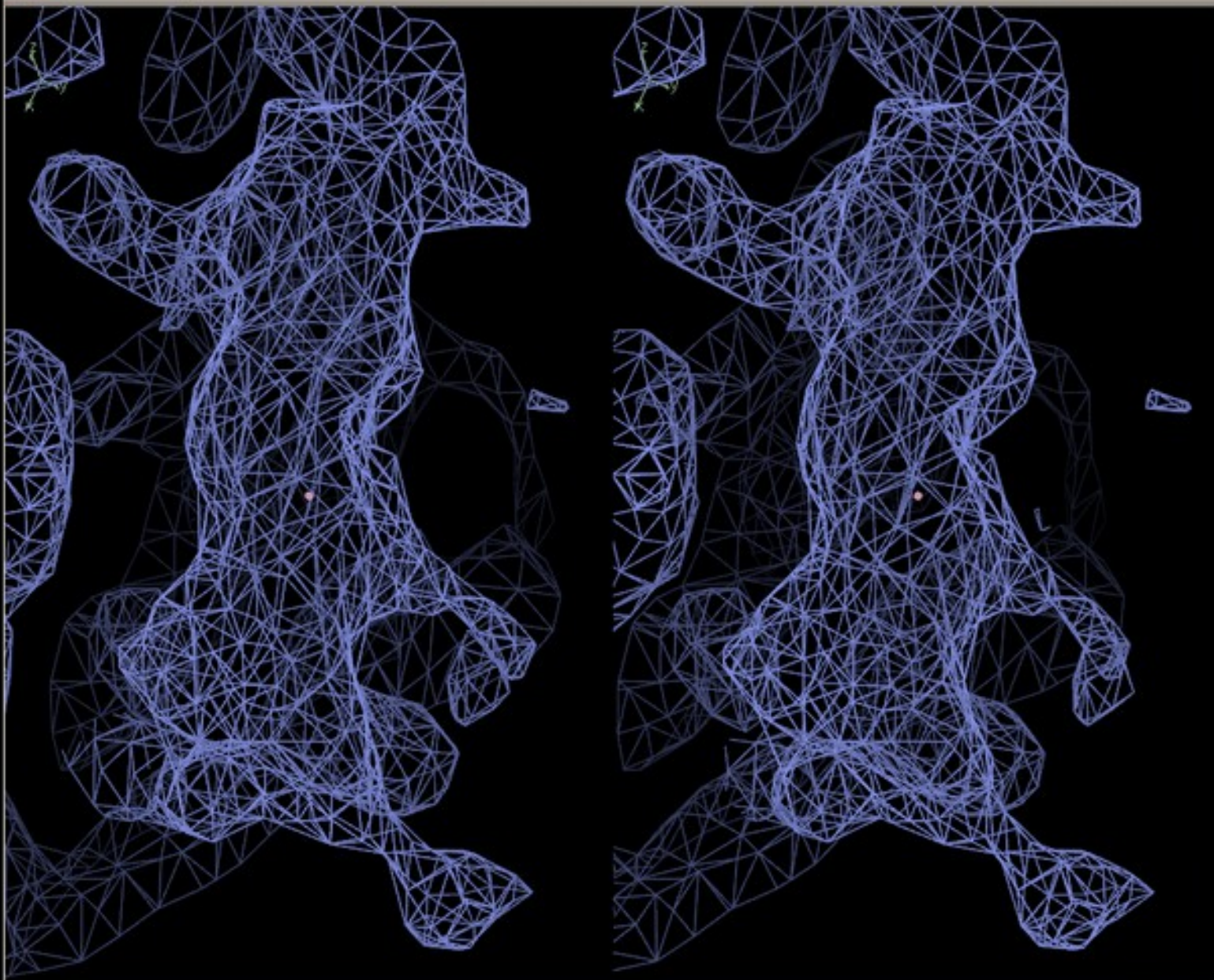


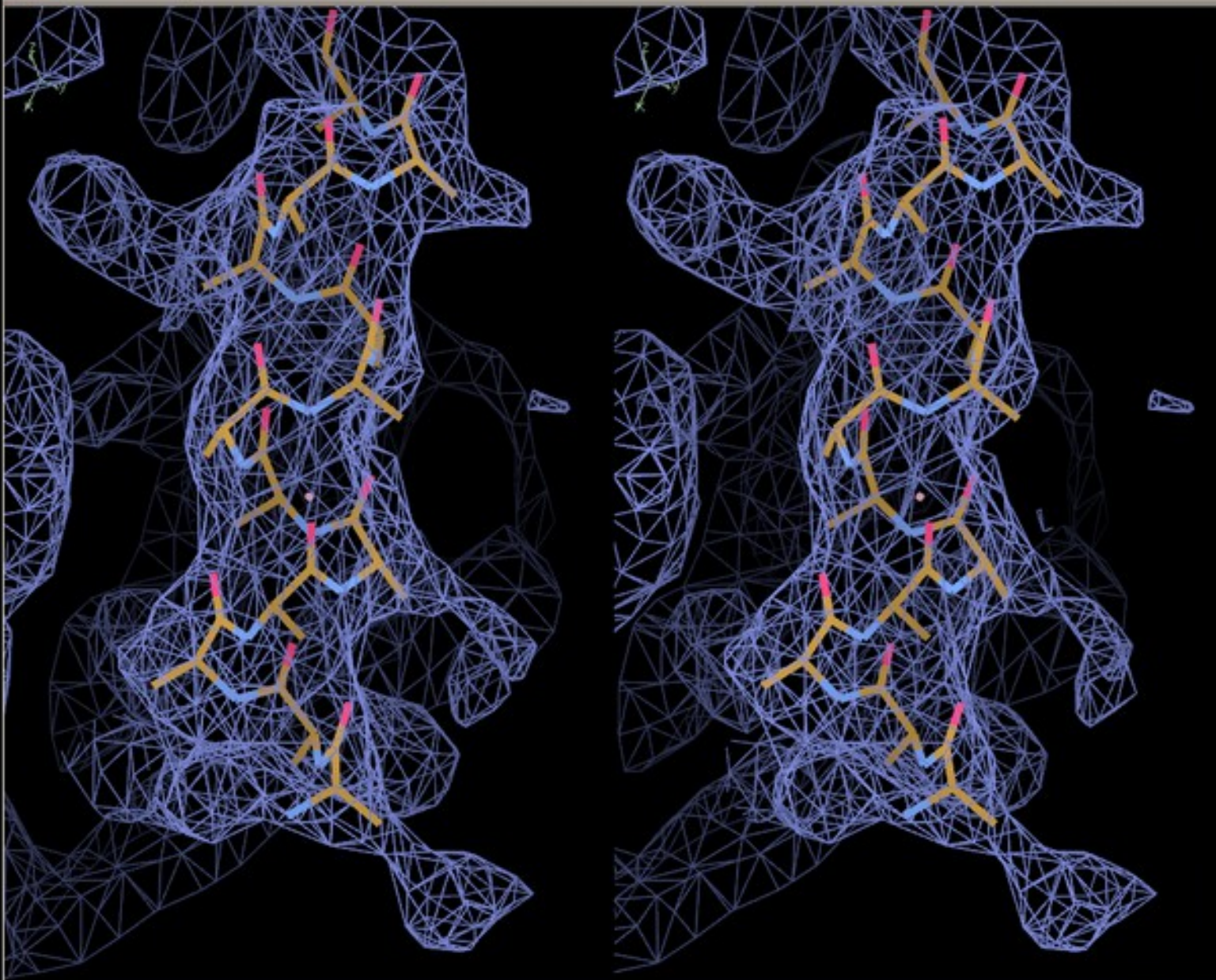
# Helix Fitting

## Comparing orientation hypotheses

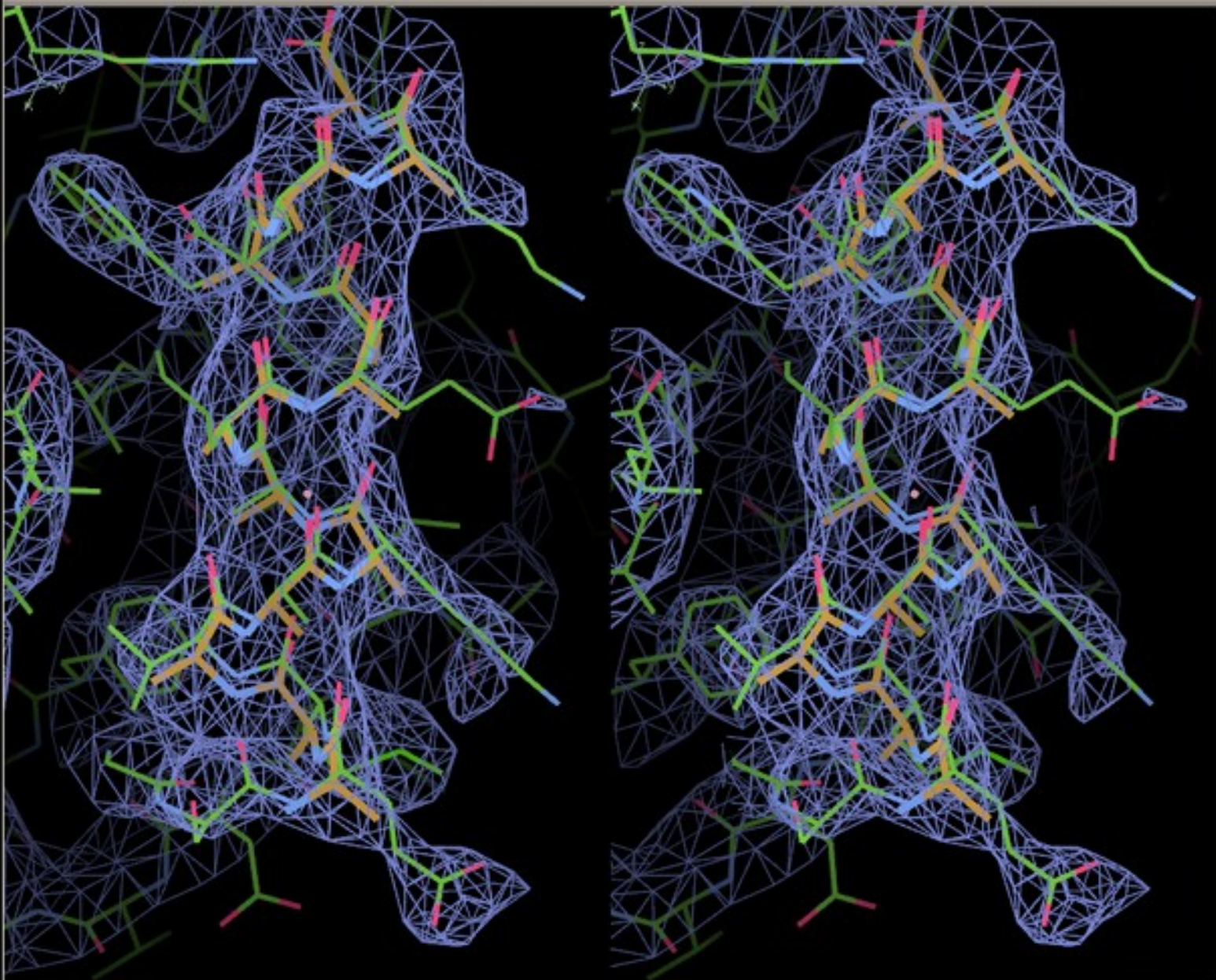


c-betas are not fitted and are used for scoring





- Navigation icons: Home, Back, Forward, Search, etc.
- Display Manager icons: Toggle visibility of various components like the map, model, and helix.
- Tools: Rotate, Translate, Scale, etc.
- Validation icons: Radiation symbol, etc.
- Other utility icons: Plus sign, etc.



Helix added

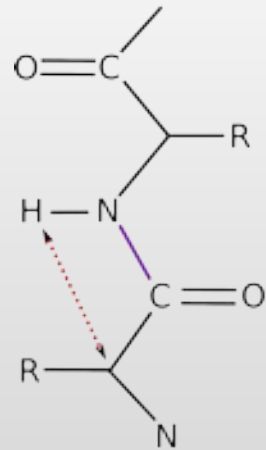
# cis-Peptides

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

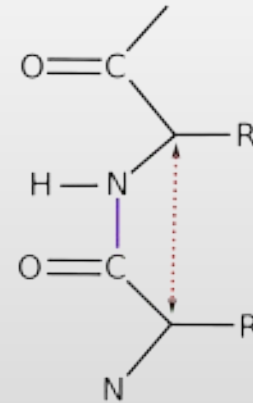
# cis-Peptides

- A number of papers 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

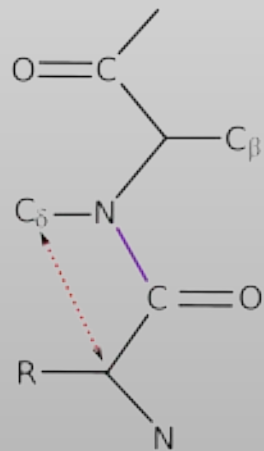
# cis-Peptides



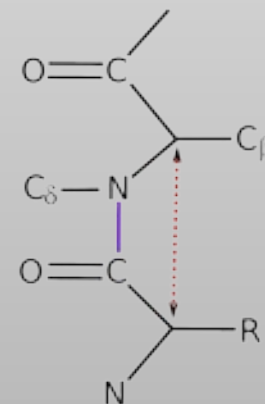
*trans-peptide*



*cis-peptide*



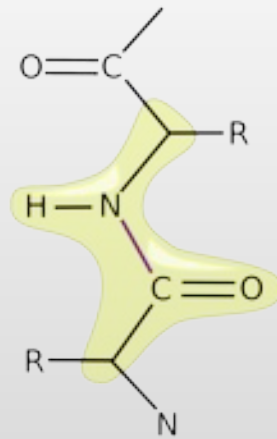
*PRO trans-peptide*



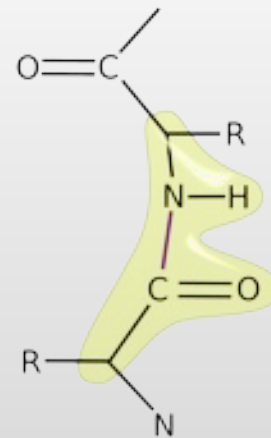
*PRO cis-peptide*



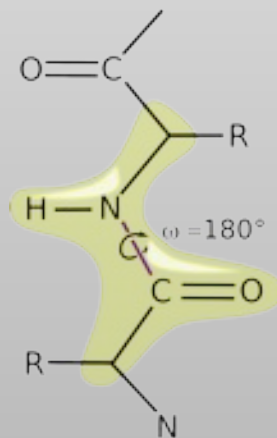
# *cis*-Peptides



*trans*-peptide  
with plane restraints

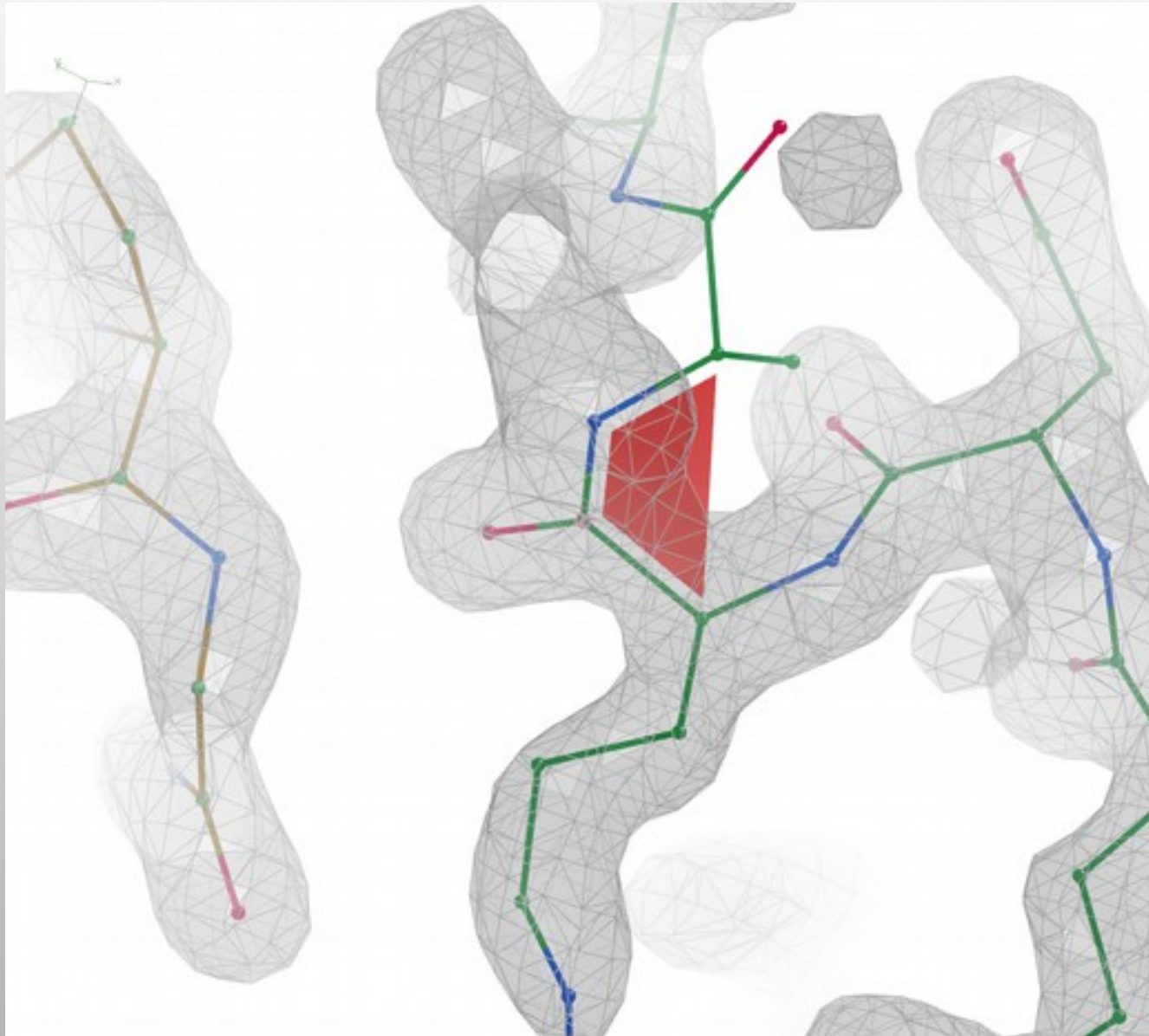


*cis*-peptide  
with plane restraints



*trans*-peptide  
with plane and trans restraints

# *cis*-peptide Representation



Pre-PRO



Twisted-trans



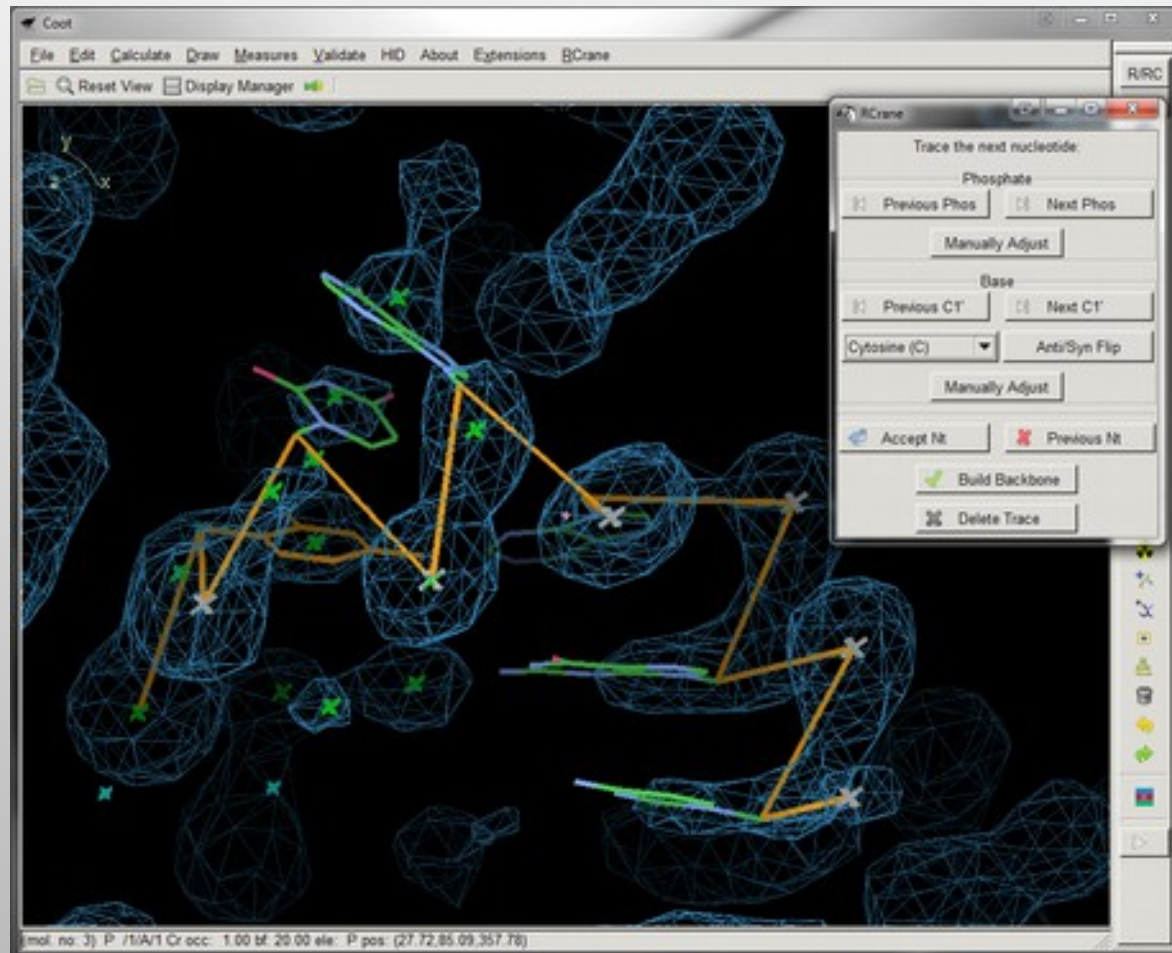
Non-pre-PRO



# A Sample of Tools

- A few extra tools...

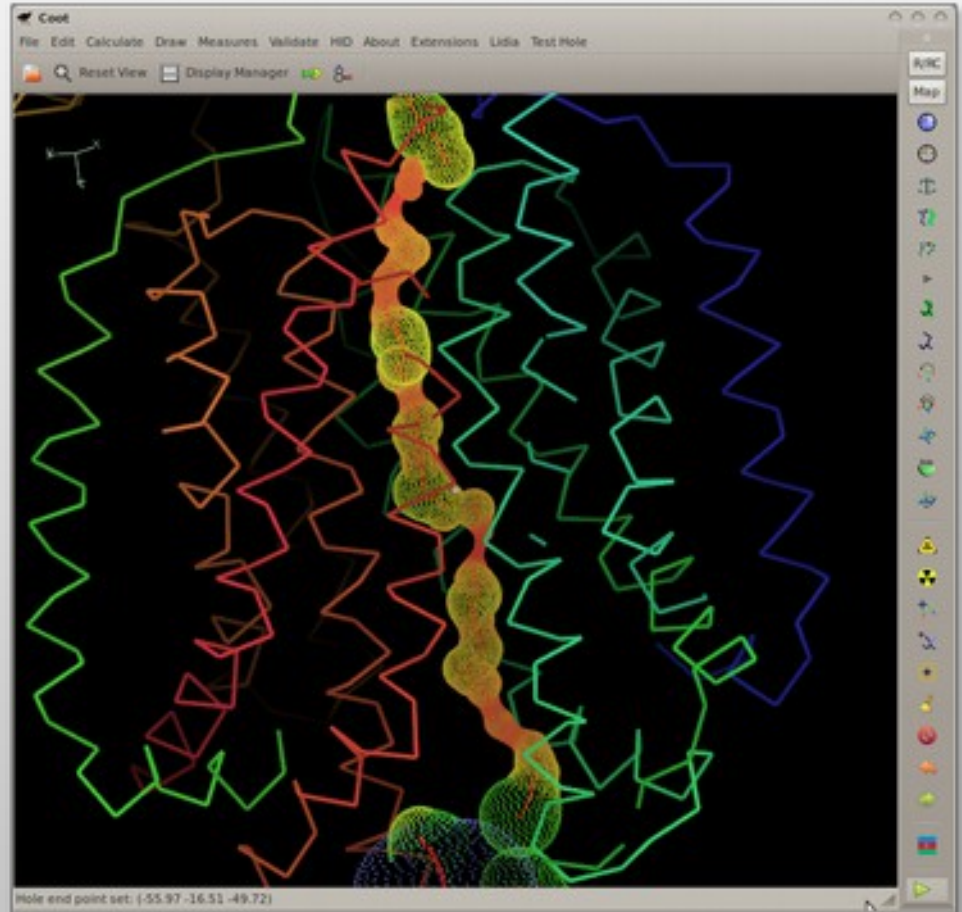
# RCrane: Semi-automated RNA building

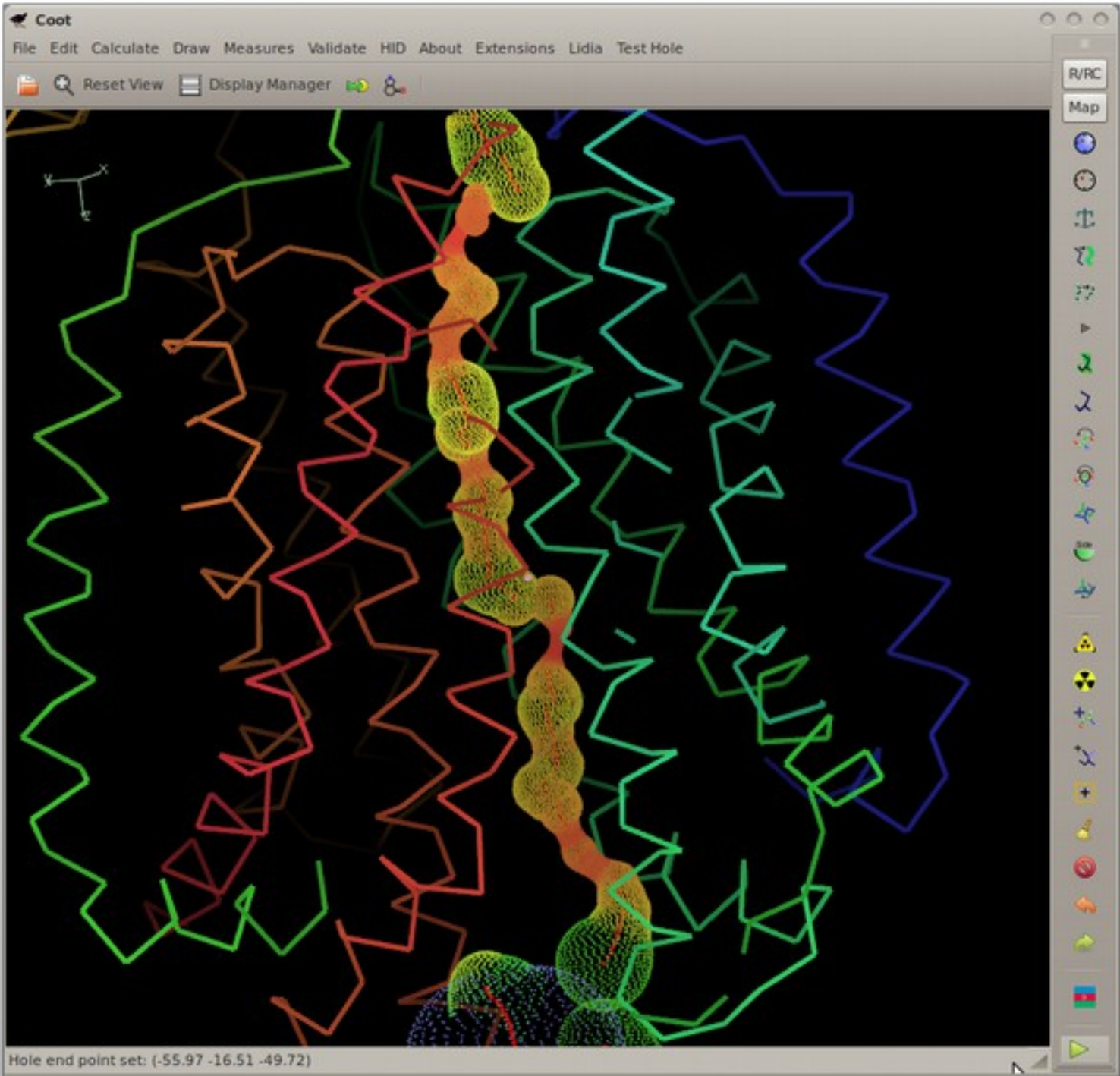


Kevin Keating

# Coot: 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



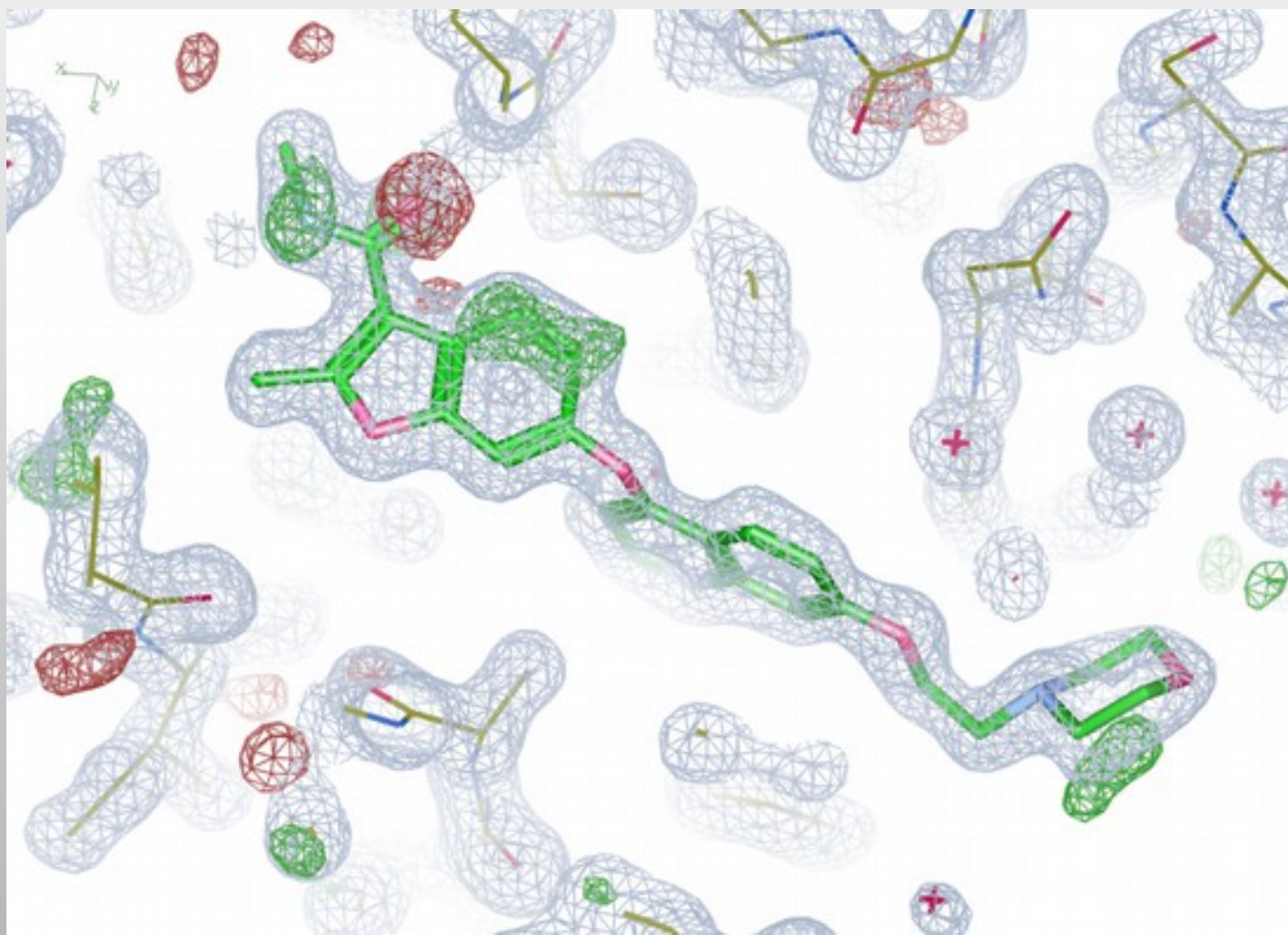


Hole end point set: (-55.97 -16.51 -49.72)

# Making Density Slides with Coot

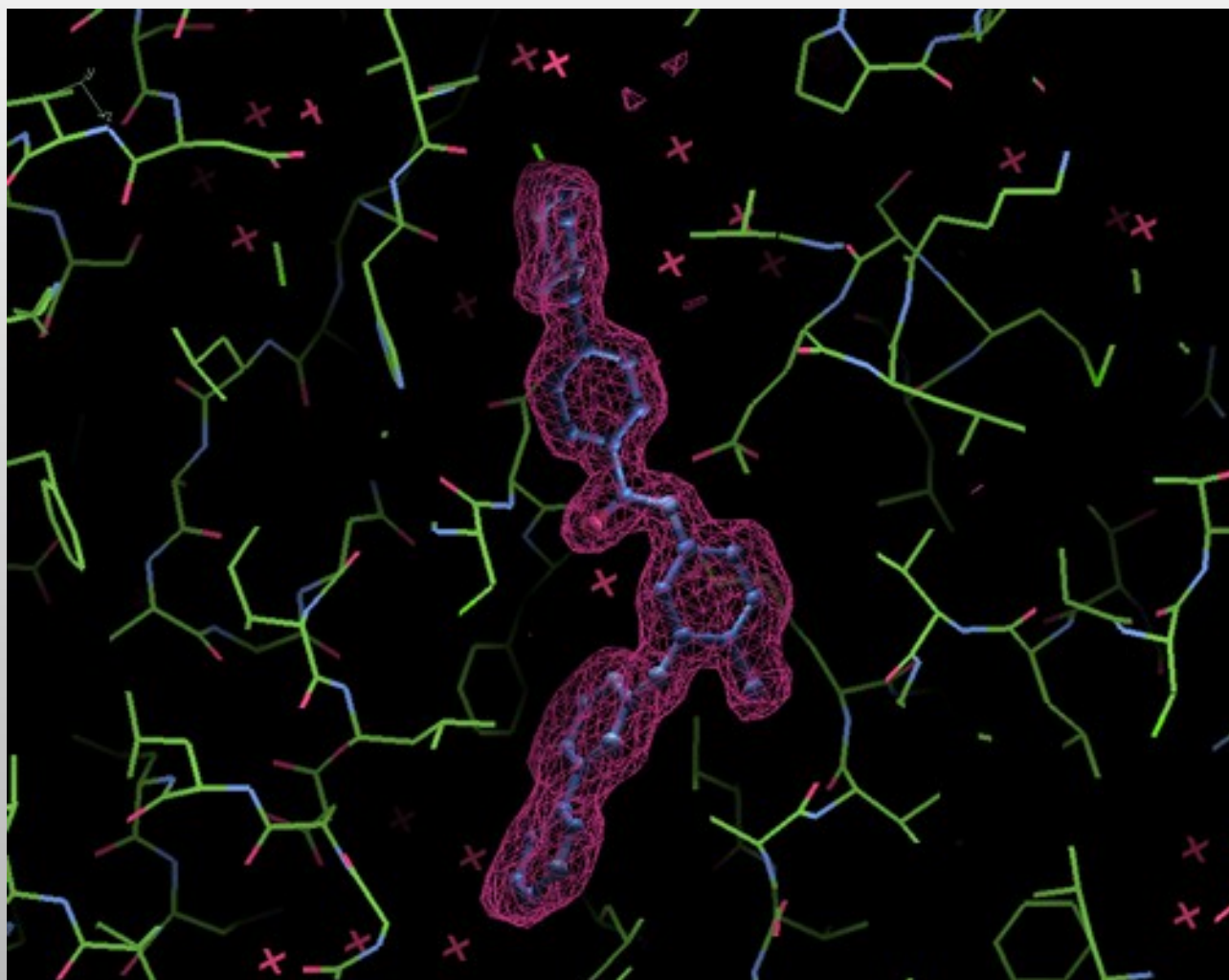
- White background
- “High” Oversampling (2.3x)
- Pale gray (or very pastel) density colour
- Enable Cut-glass mode 5-10%
- Anti-aliased Coot
  - `$ setenv __GL_FSAA_MODE 5`
  - 0.8.3 will do a better job of anti-aliasing out the box
  - (transfer to CCP4-built binaries)

# Example Density Slide

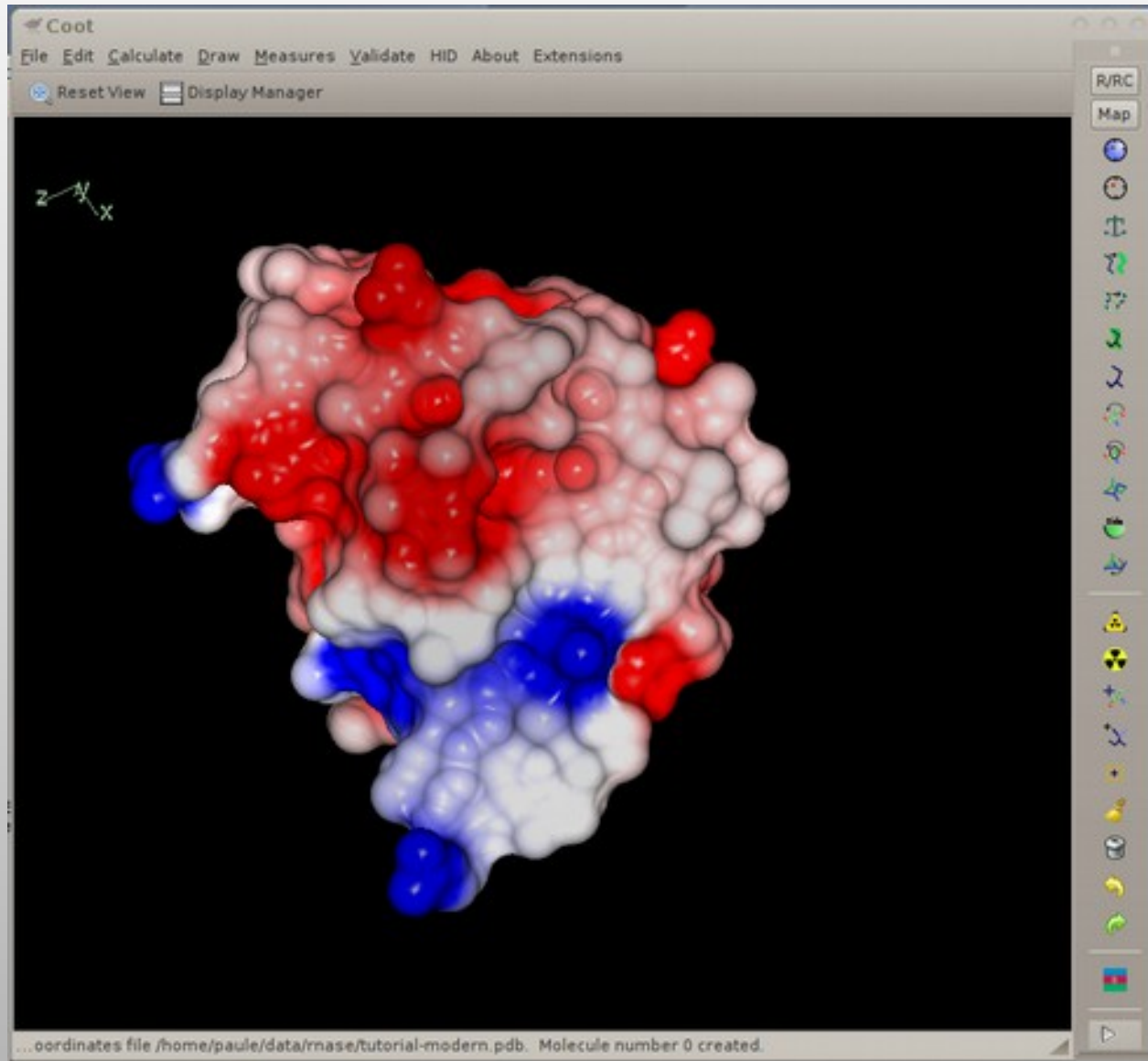




# Some Representation Tools



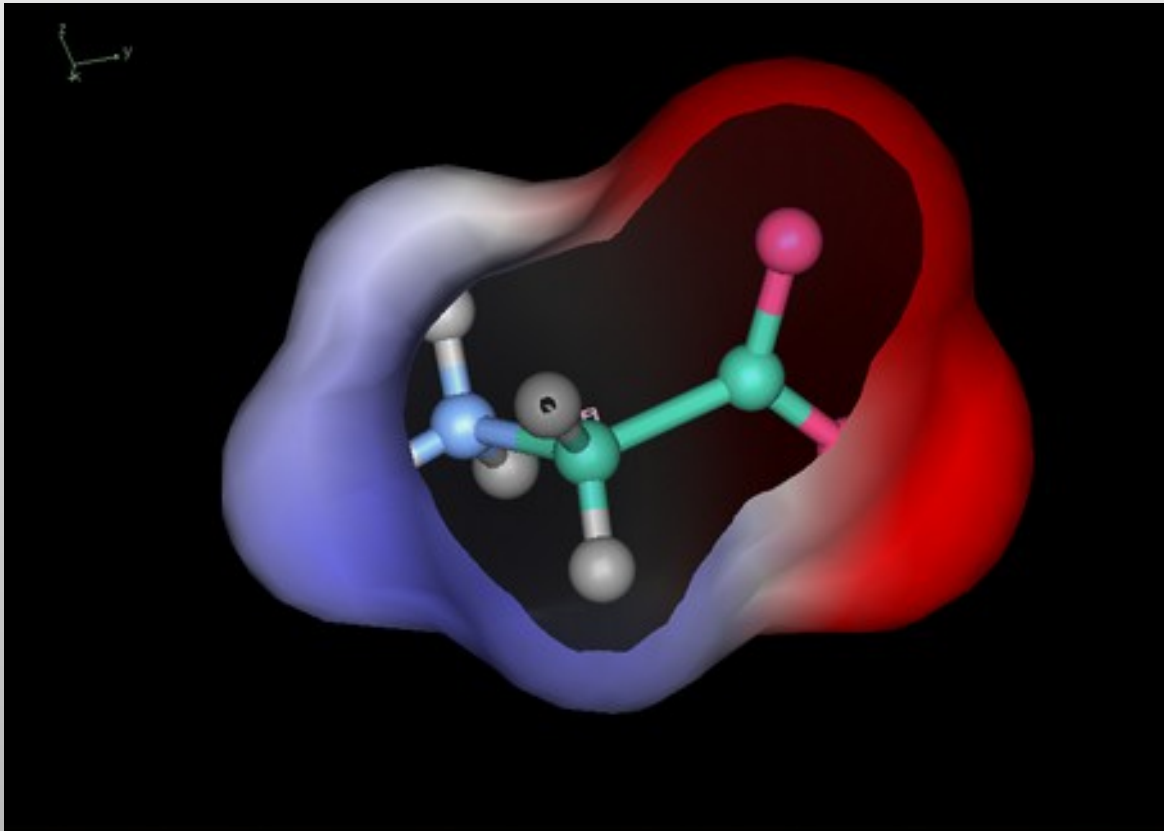
# Some Representation Tools



Gruber & Noble  
(2007)

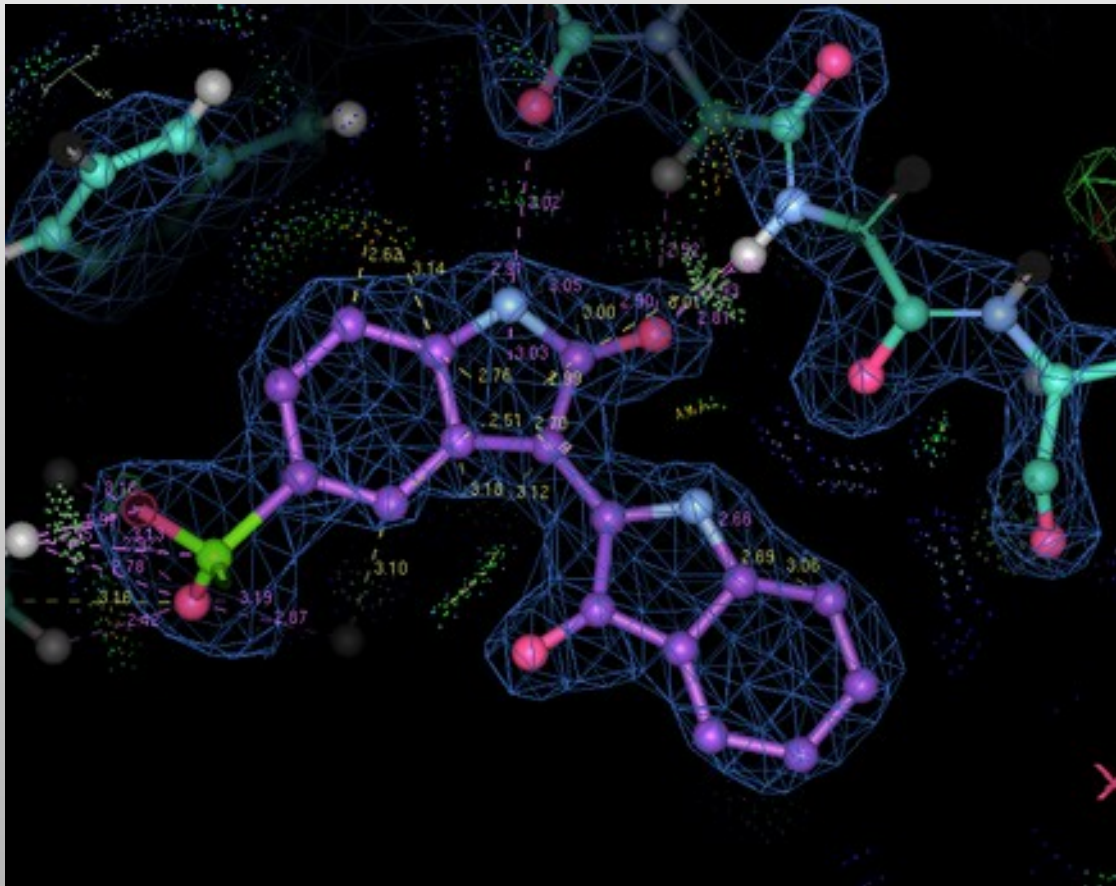
# Other Things

- Surfaces that use dictionary partial charges

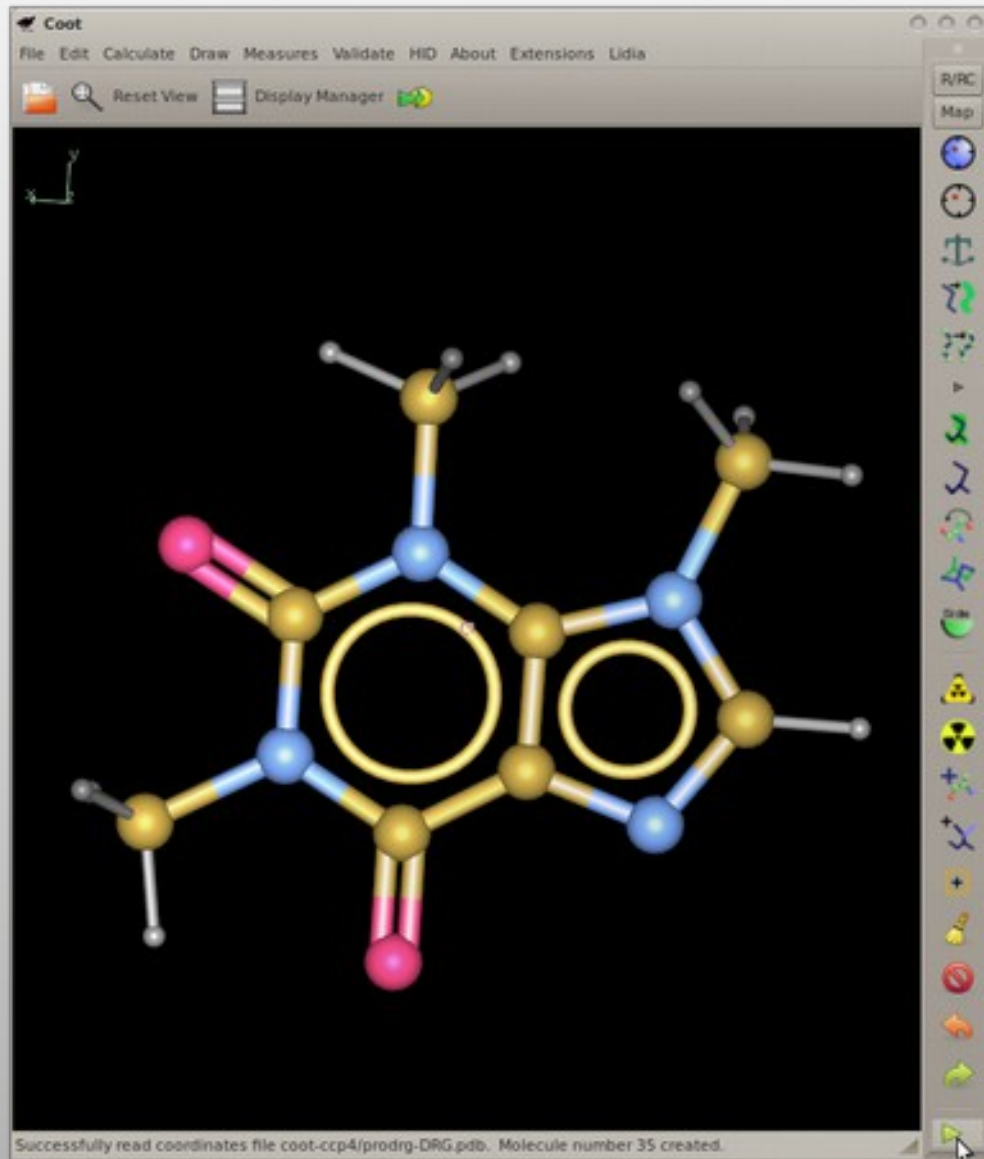


# Other Tools

- Molprobity dots for ligands
- Highlight interesting site



# Representing Bond Orders

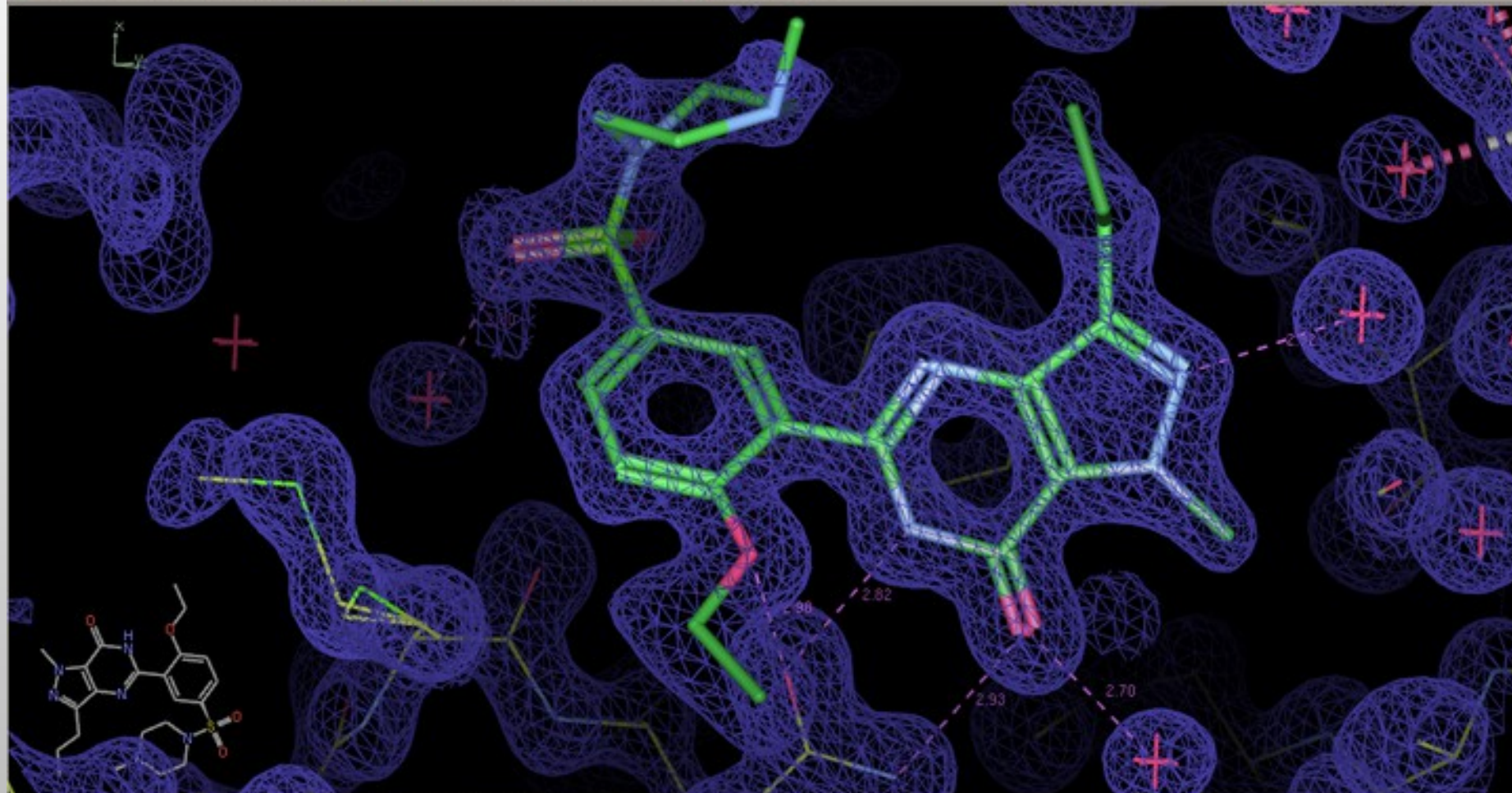


File Edit Calculate Draw Measures Validate HID About Extensions Ligand

Reset View Display Manager Ligand Builder Sphere Refine Backrub Rotamers

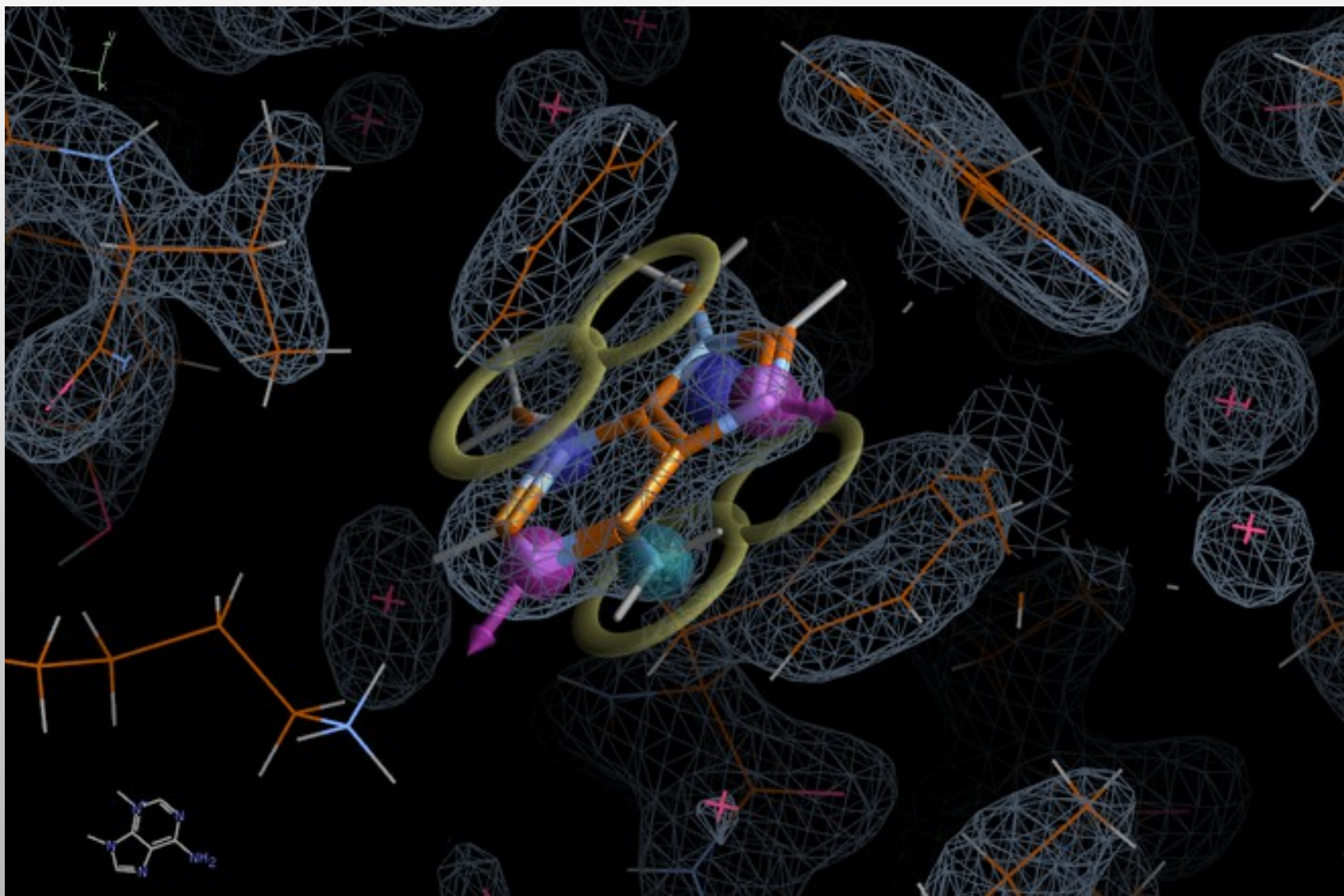
R/R/C

Map



(mol. no: 6) C9 /1/A/501 VIA occ: 1.00 bf: 14.44 ele: C pos: (27.49,29.50,63.65)

# Chemical Features



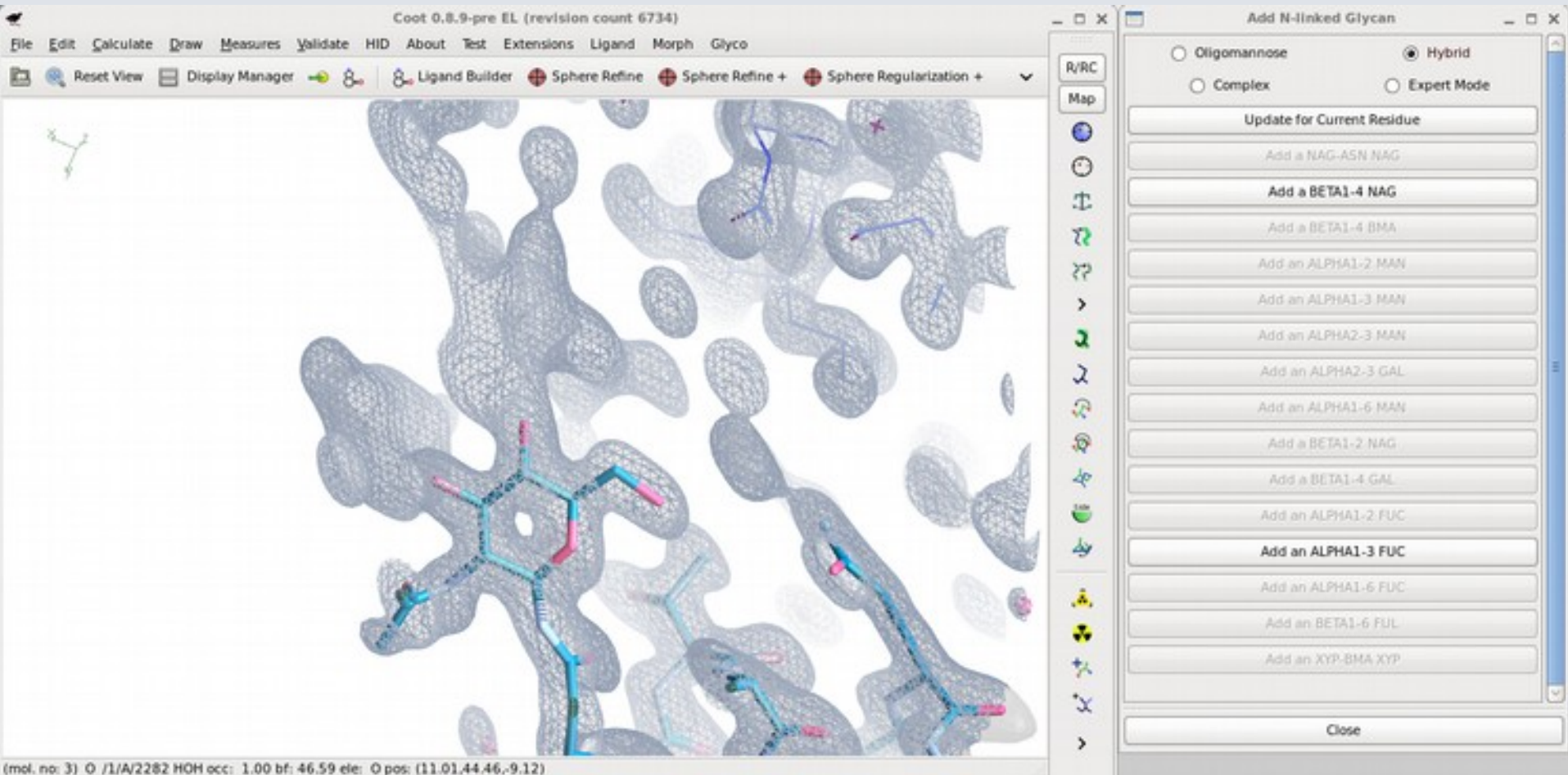
# A Few Tools More...

- Fitting Low-Resolution/EM maps
- Ligands:
  - dictionaries
  - ligand-fitting
  - analysis
- Carbohydrate-fitting
  - N-linked glycosylation



# N-linked Carbohydrates

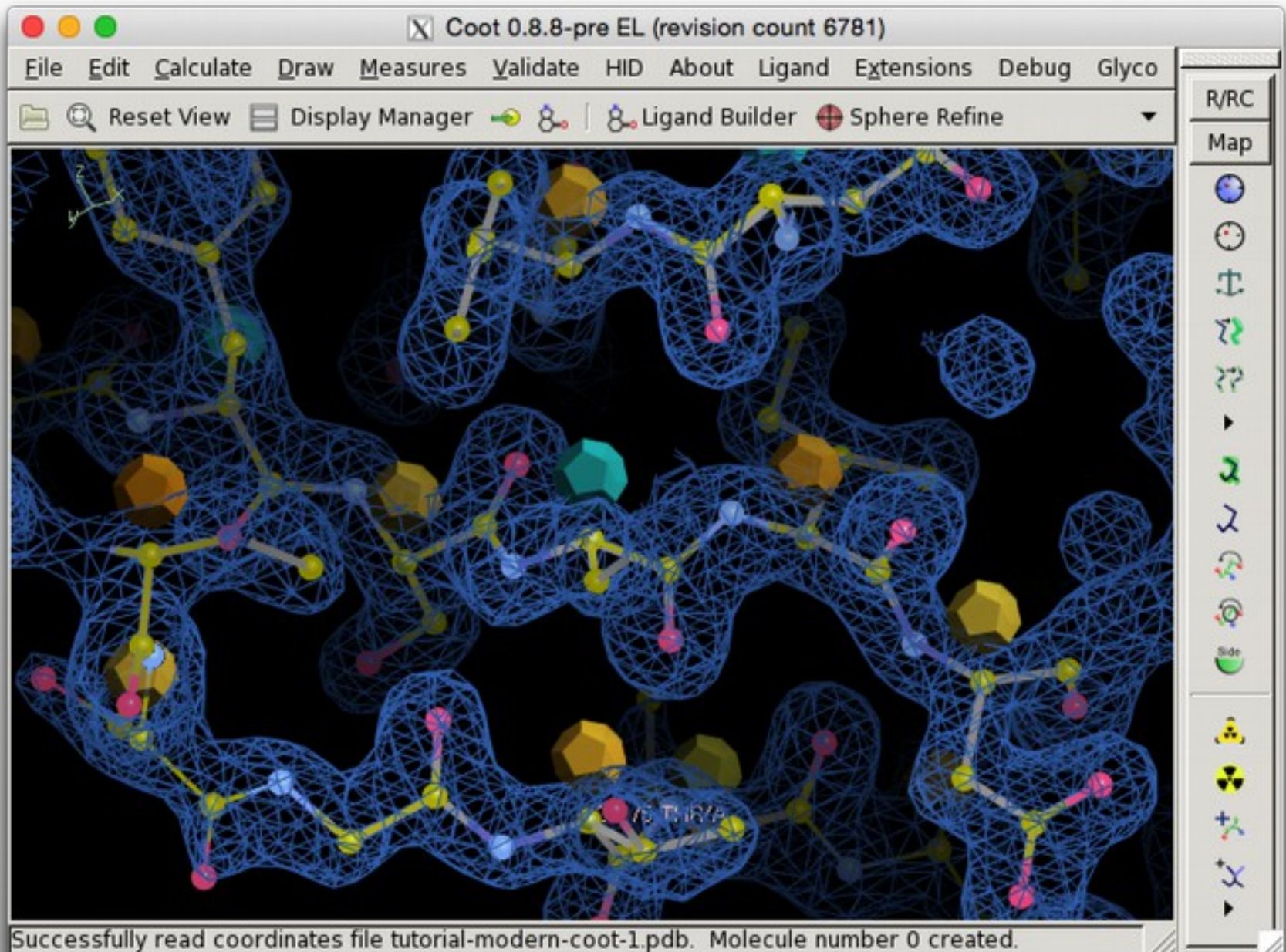
- Improved algorithm and re-worked GUI



# *Coot* Futures

- Routine Geman-McClure distance restraints
  - Multi-threading/parallel processing
- GPU usage:
  - Refinement
  - Contouring
  - Representation
- Interactive Ramachandran, rotamer and clash markup

# Interactive Rotamer Goodness



Coot 0.8.8-pre EL (revision count 6781)

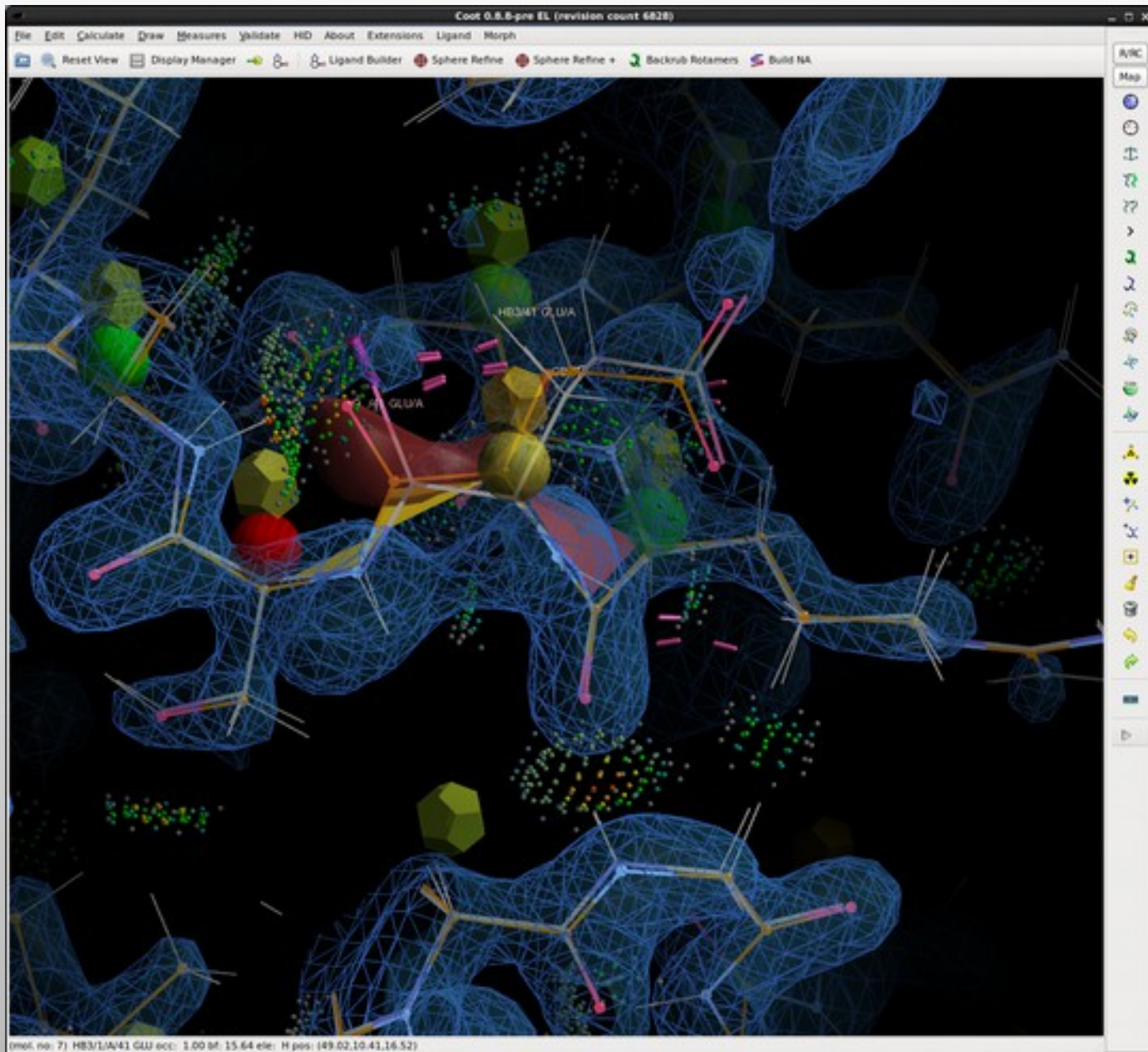
File Edit Calculate Draw Measures Validate HID About Ligand Extensions Debug Glyco

Reset View Display Manager Ligand Builder Sphere Refine

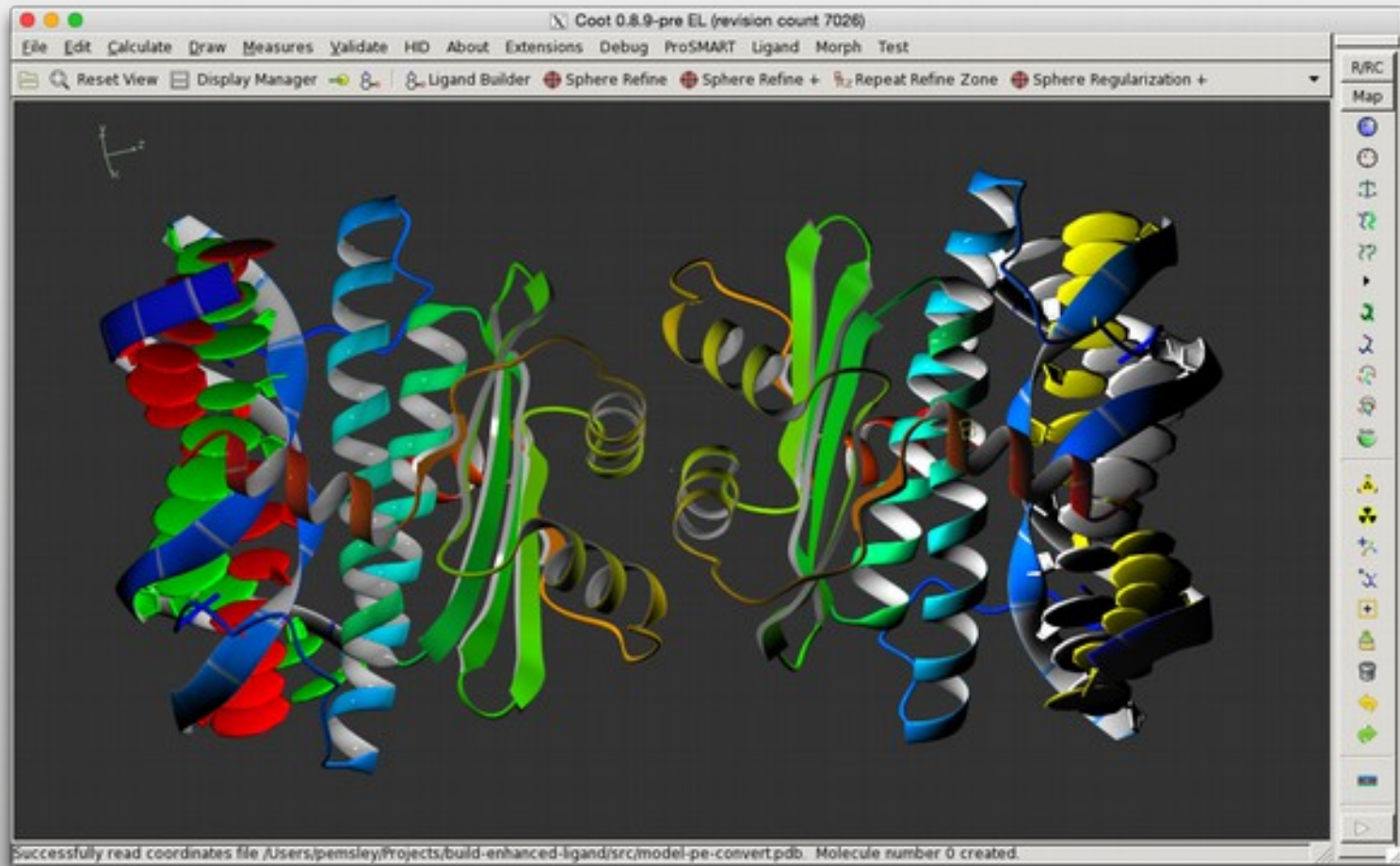
R/RC  
Map

Successfully read coordinates file tutorial-modern-coot-1.pdb. Molecule number 0 created.

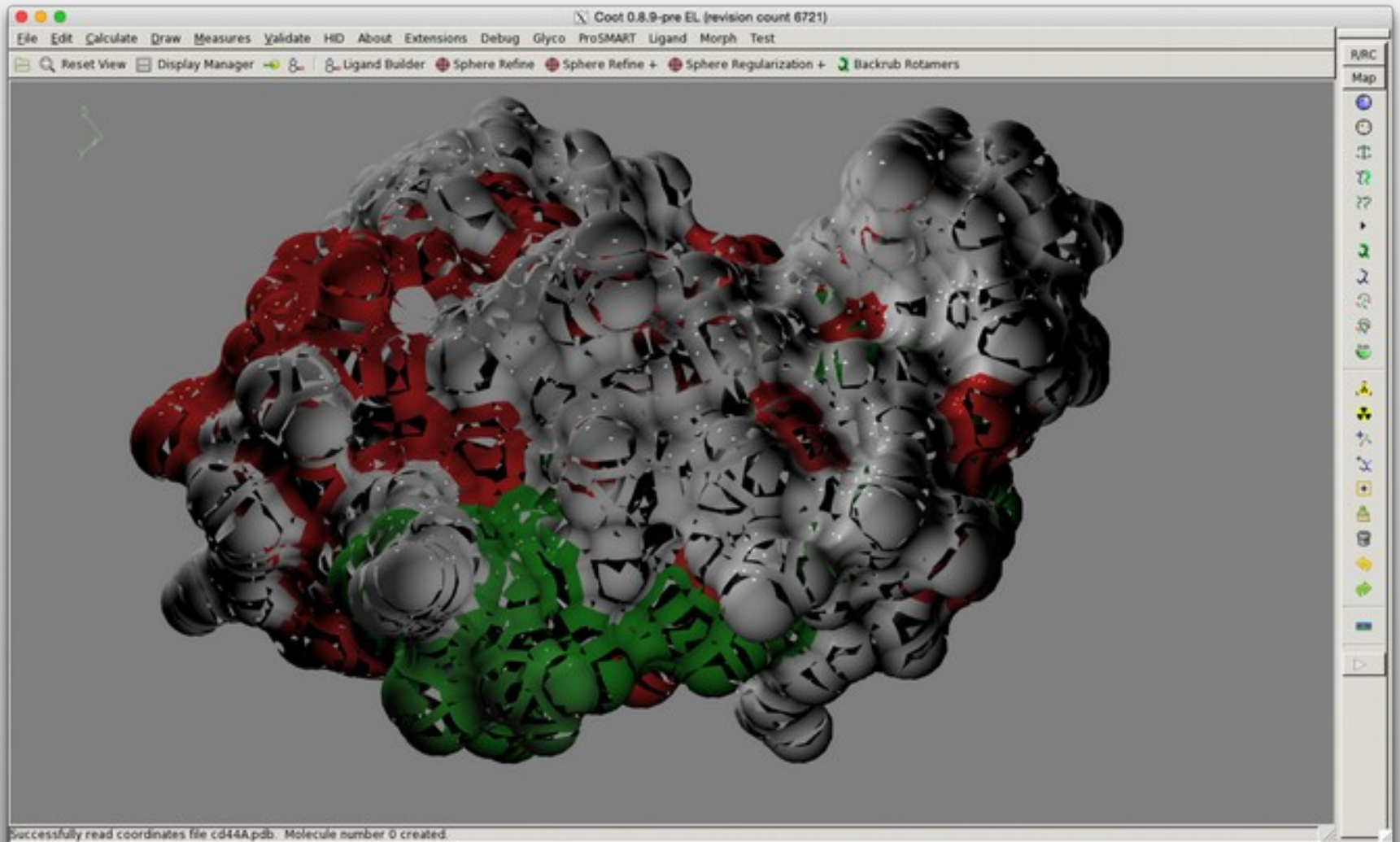
# Coot Futures: Multi-Criteria Markup



# Coot Futures: GPU Ribbons

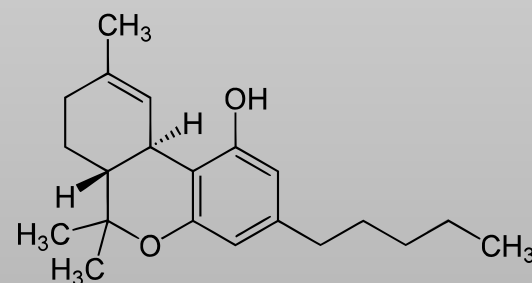
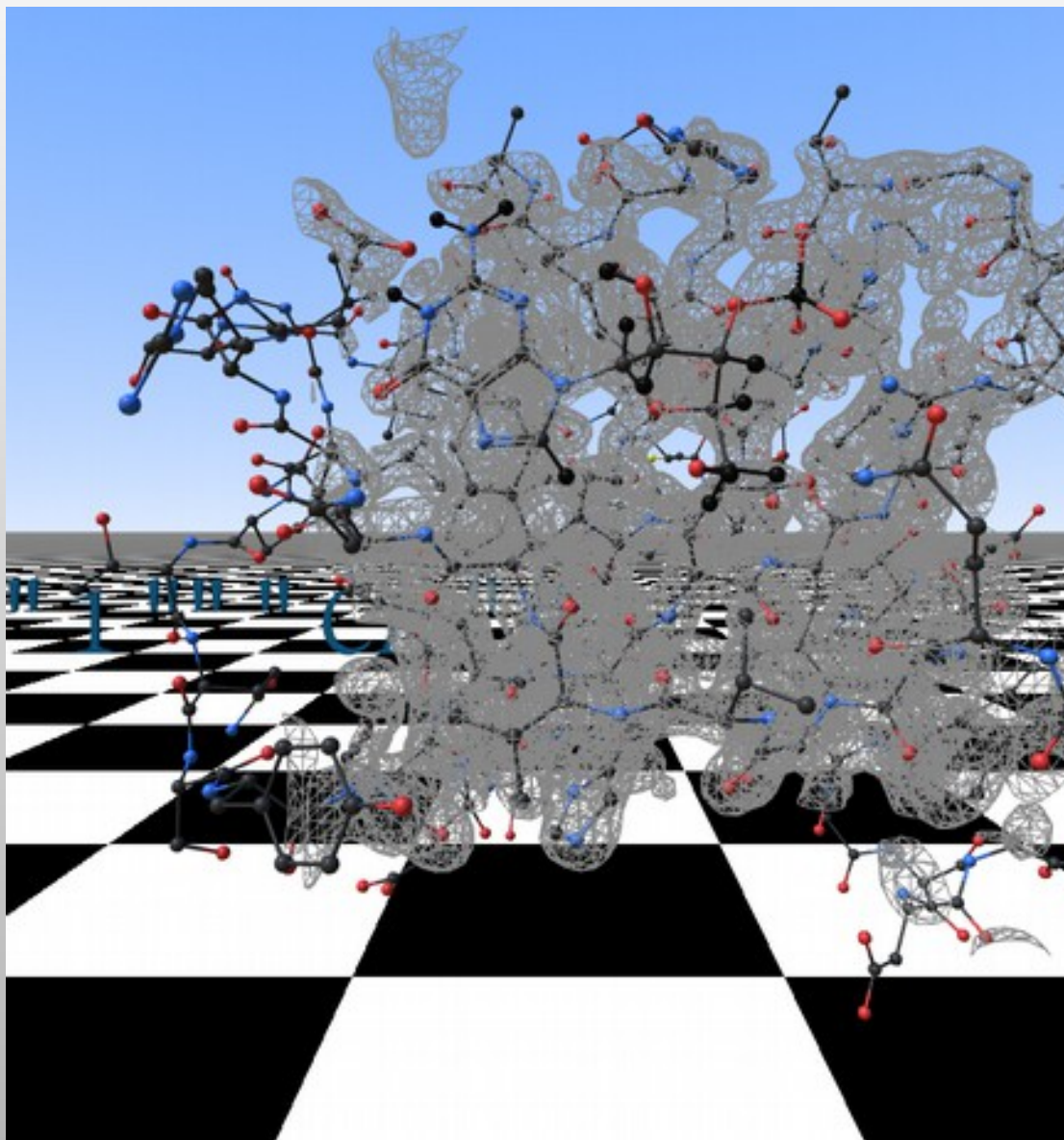


# Coot Futures: GPU Surfaces



# Cool Futures: Virtual Reality

Hamish Todd



Don't fear the mushroom

- AR Coot





# Acknowledgements

- Martin Noble
- Kevin Cowtan
- Bernhard Lohkamp
- Colleagues at LMB MRC
  
- Libraries, dictionaries
  - Alexei Vagin, Eugene Krissinel
  - Richardsons (Duke)
  
- Funding
  - BBSRC, CCP4 & MRC

# **Non-Crystallographic Symmetry**

# What is Non-Crystallographic Symmetry?

- 2 or more copies of a molecule in the unit cell not related by crystallographic symmetry
- Crystallographic copies of molecules are (of course) treated as if they were exactly the same across the unit cell – and indeed across the whole crystal
- Non-crystallographically related molecules provide different representations of the same molecule
  - This can be useful for model-building
  - But difficult to use in practice

# Handling NCS

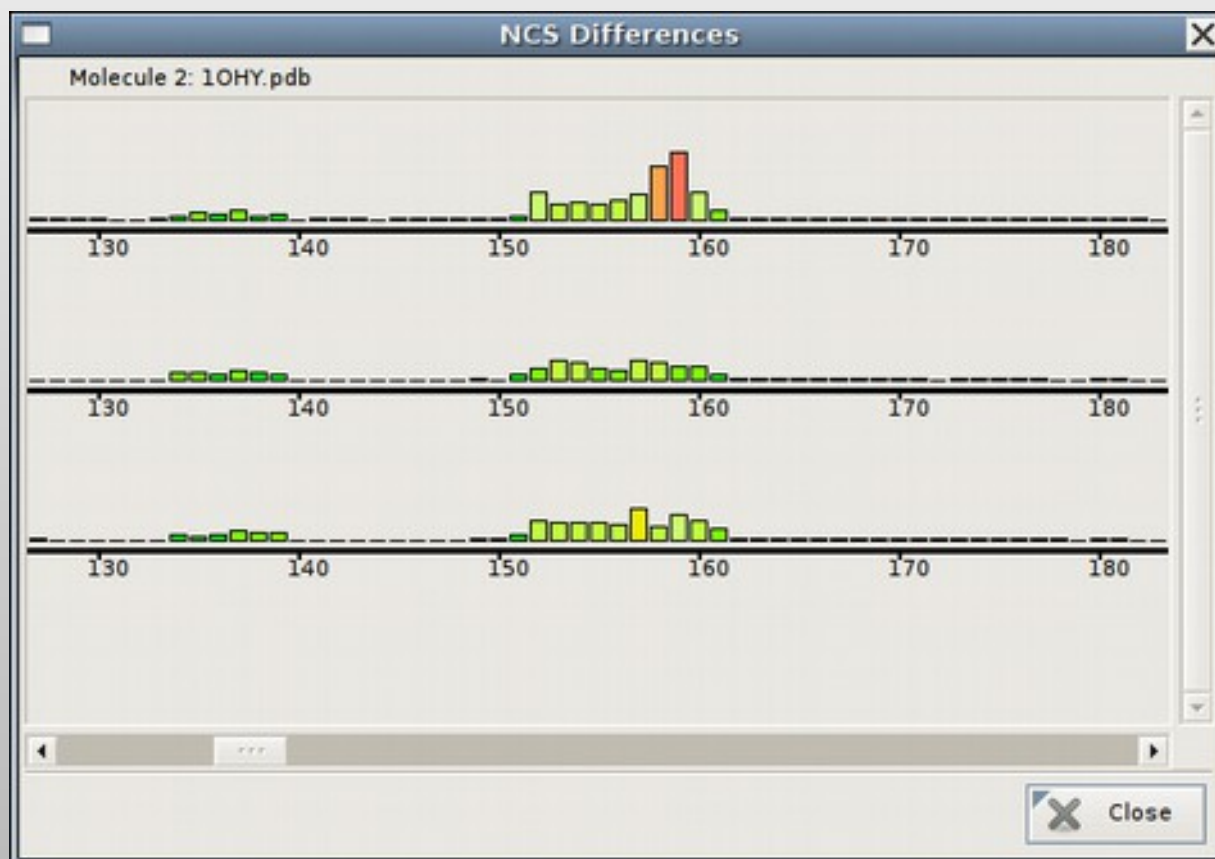
- What are the Problems?
- Strict NCS:
  - NCS should appear like crystallographic symmetry does [exact copies]
- Non-Strict NCS:
  - Molecules are different
  - How to cope with differences, but minimize unnecessary rebuilding?

# Handling NCS

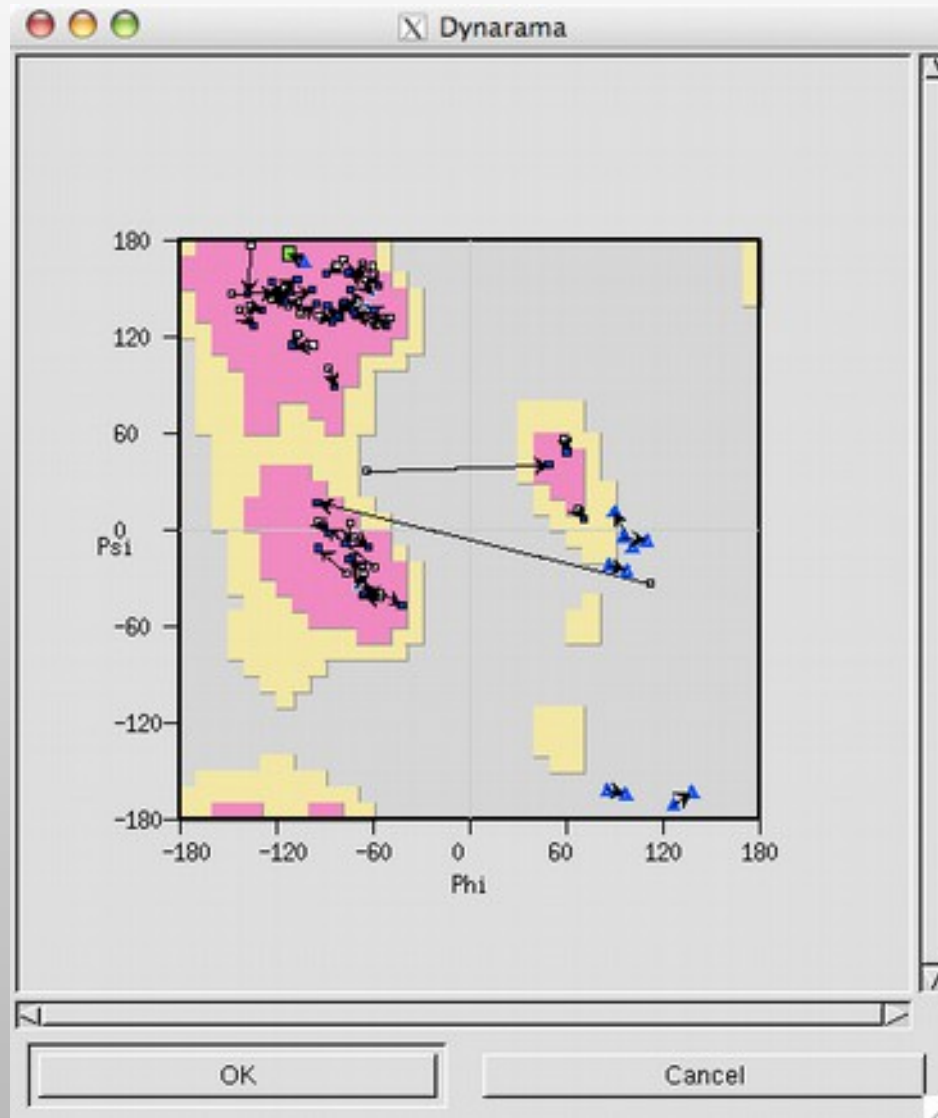
## Typical Scenario:

- 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?

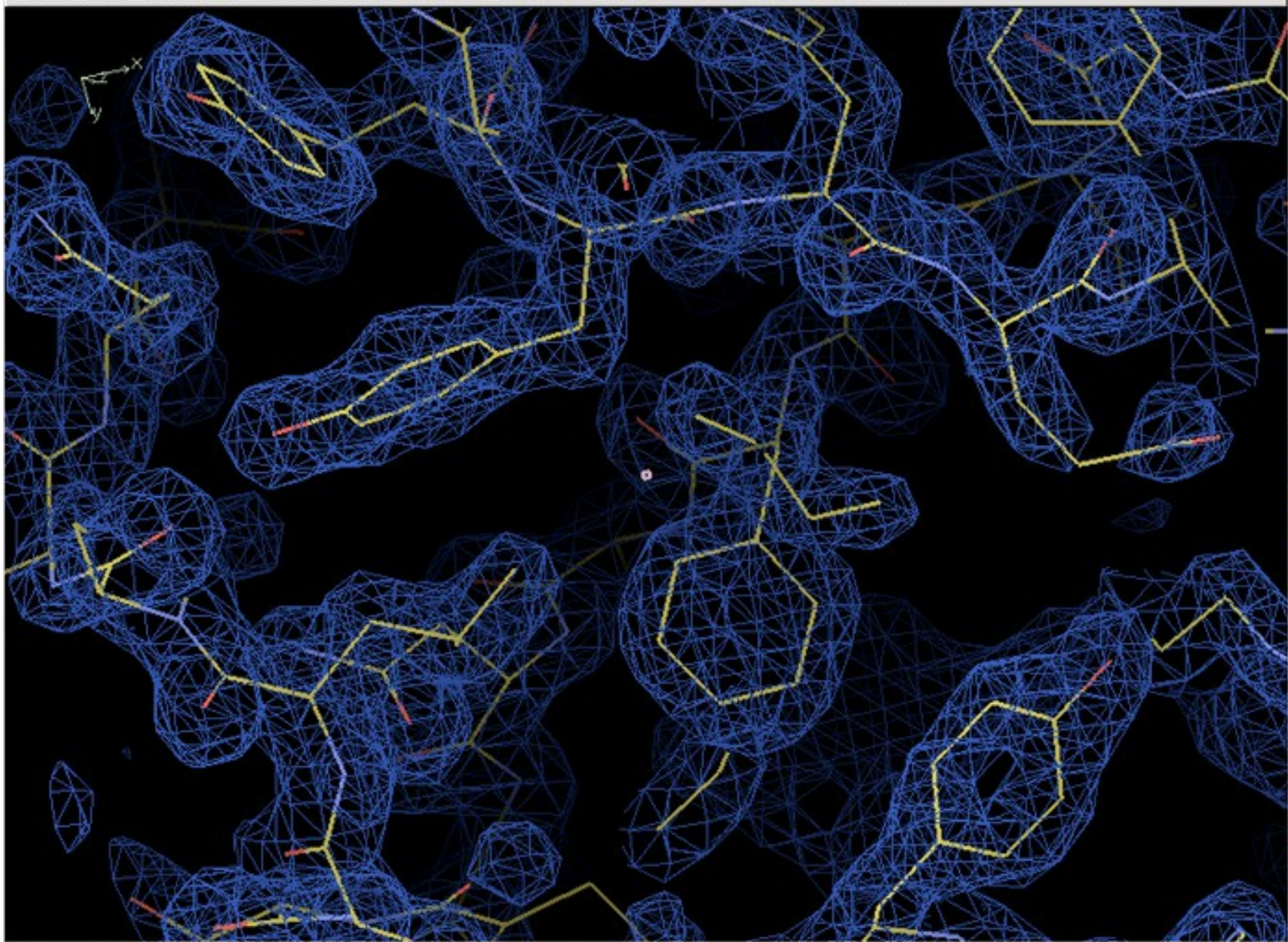
# ...or new NCS Differences graph



# ...or Kleywegt Plots[\*]



[\*] Named by George Sheldrick





# NCS Overlays

SSM NCS operator

transform map primitives

map centre

