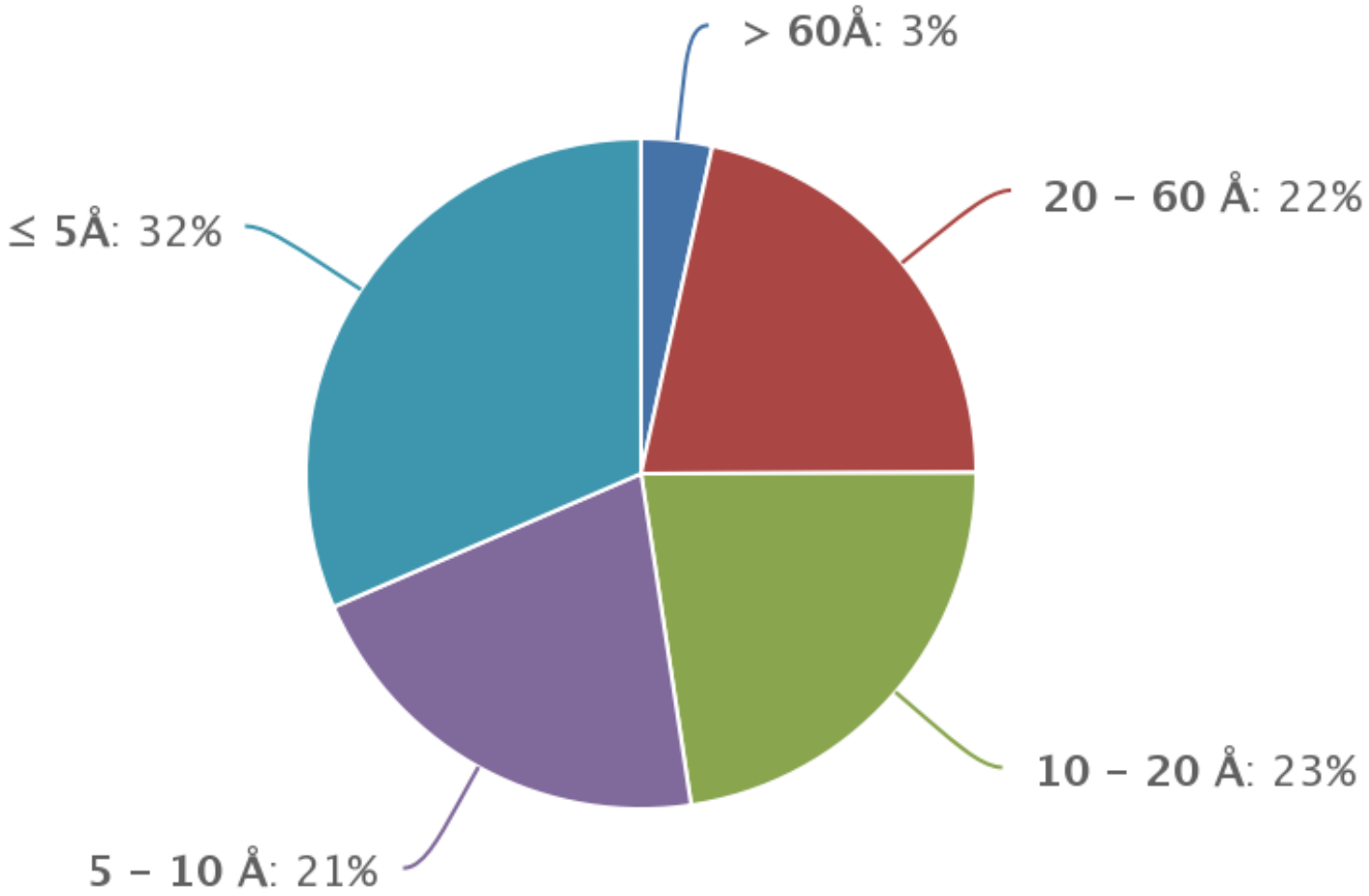
Interpreting “moderate” resolution EM data

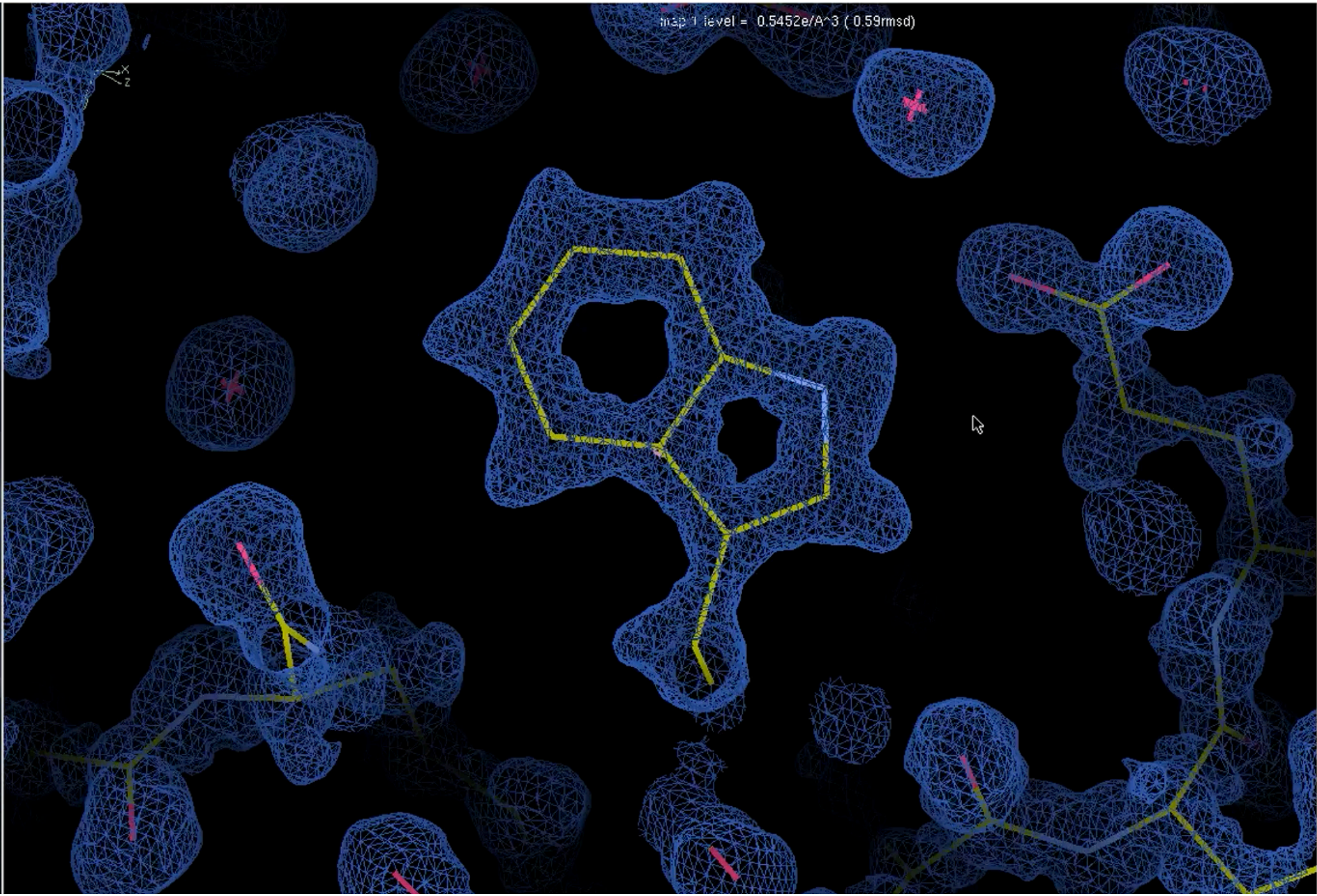
Gira Bhabha & Damian Ekiert
Skirball Institute
NYU School of Medicine

Resolution of cryo EM maps is typically low

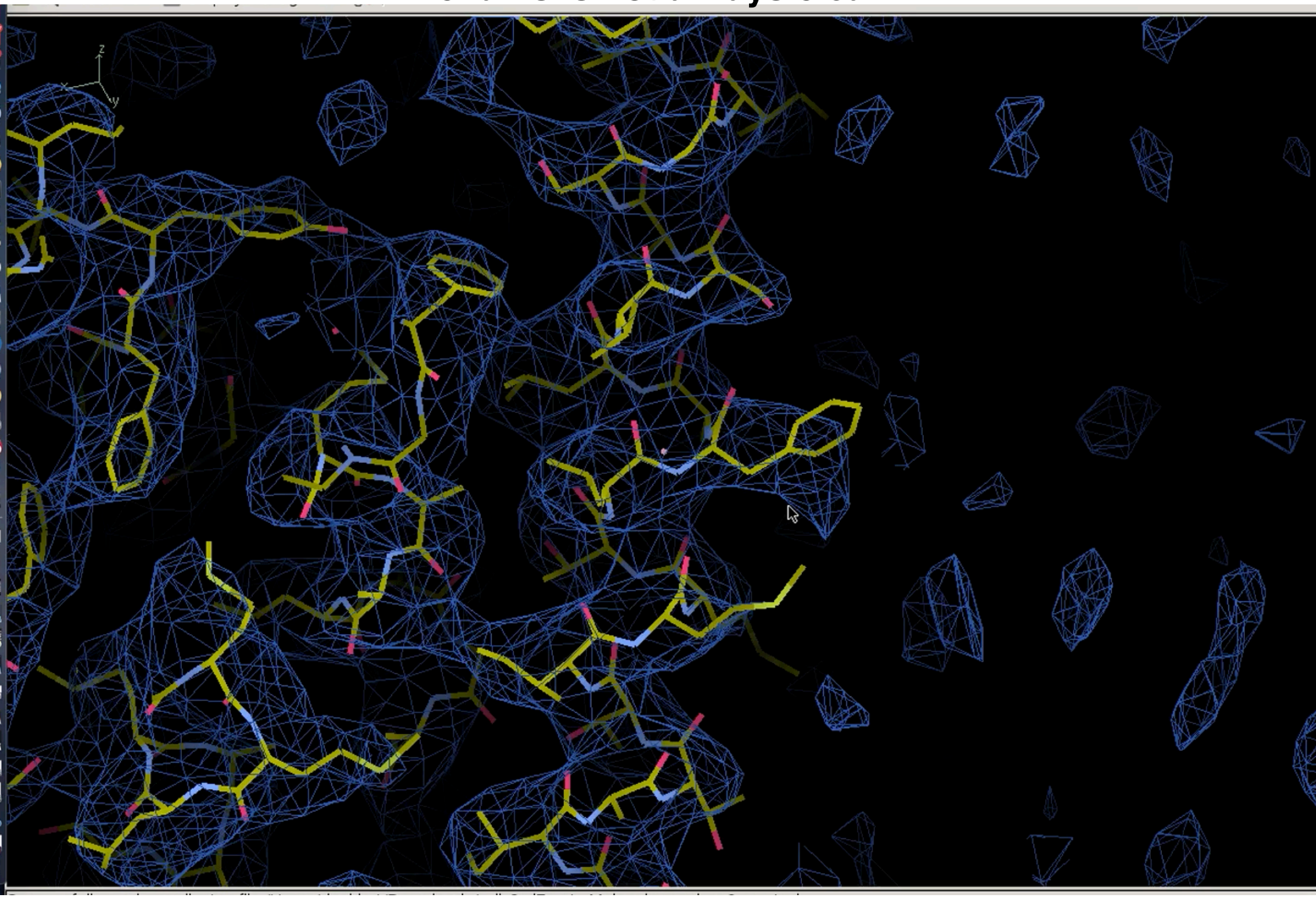


2018: 1771 total maps released, only 98 at 3 Å resolution or higher

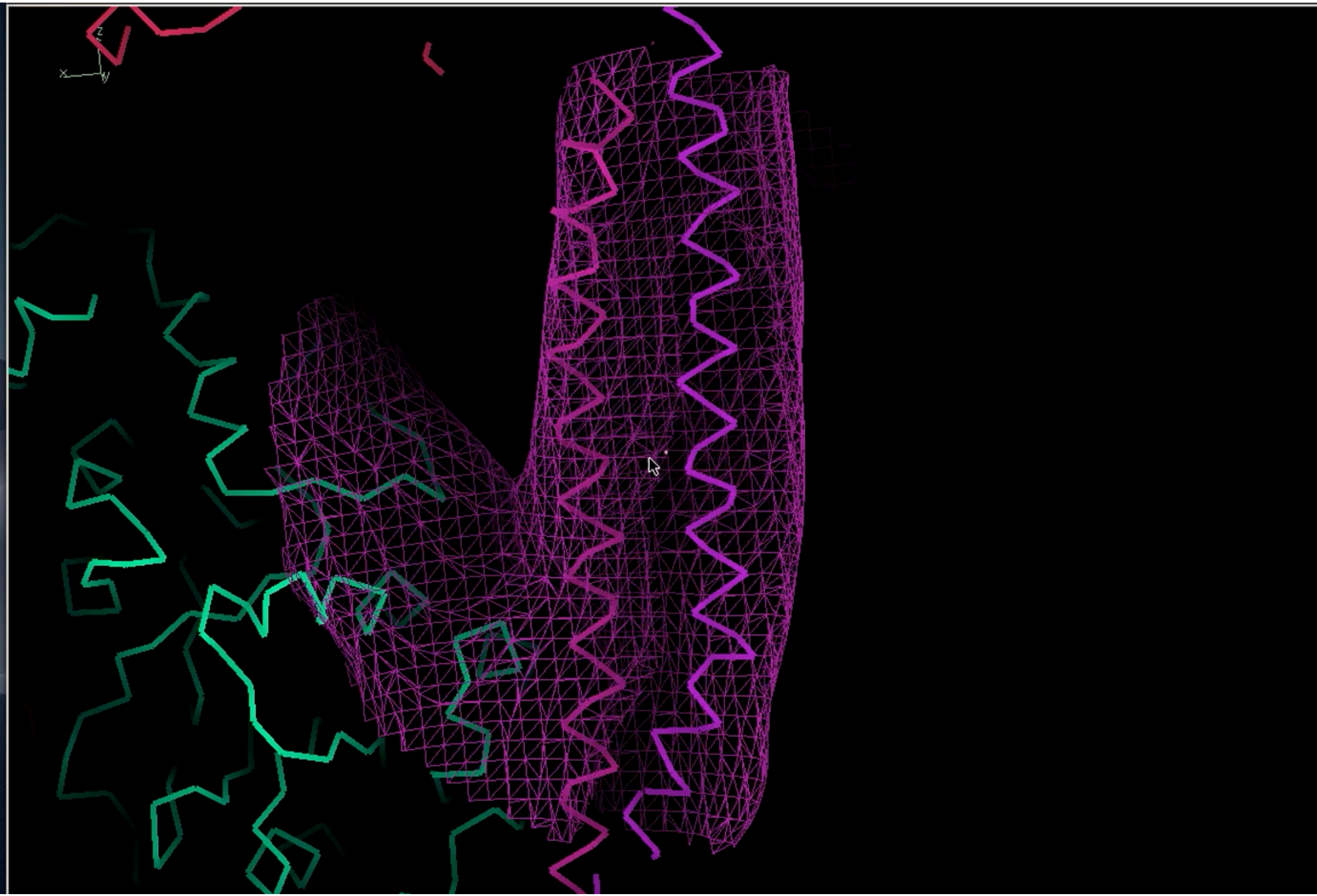
At atomic resolution, position of individual atoms is well-defined



But at “near-atomic” resolution, the position of residues and side chains is not always clear

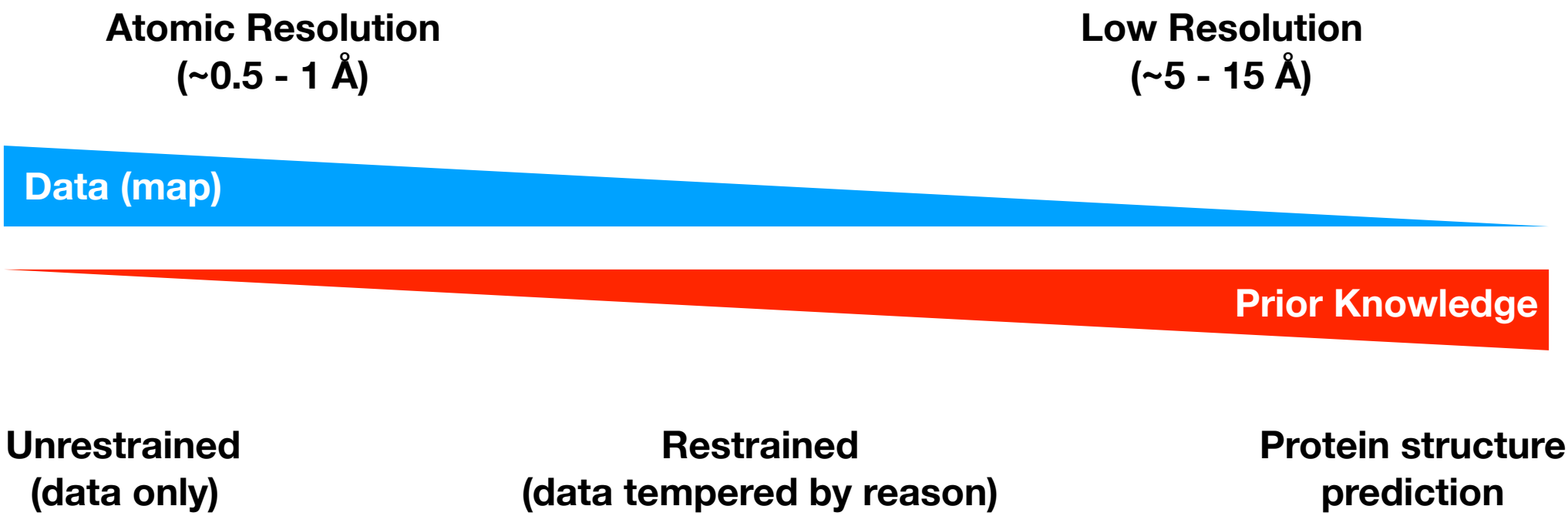


Between ~5-8 Å, we lose our ability to resolve secondary structure



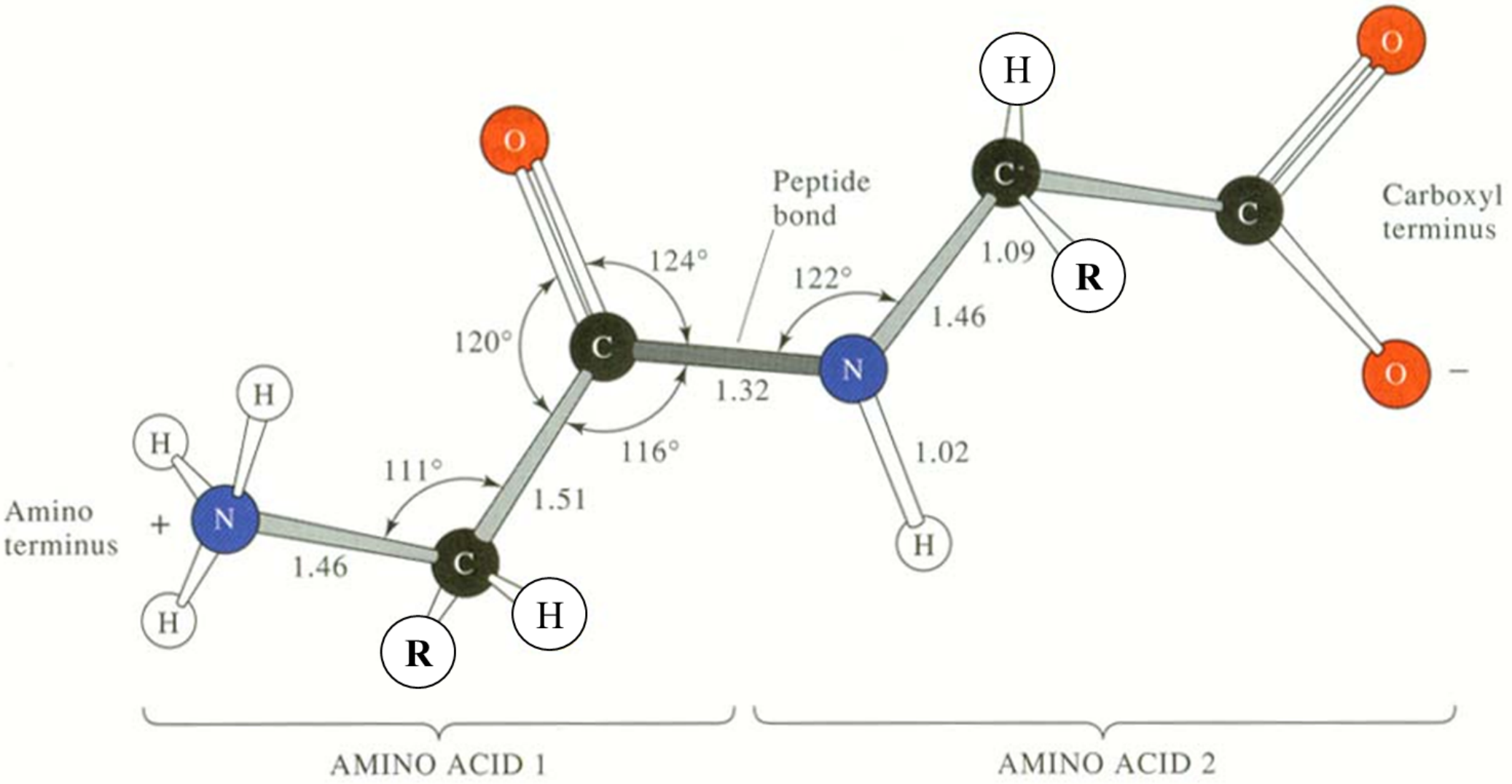
How can we interpret lower resolution cryo EM density maps to gain insights into molecular and mechanistic problems in biology?

Harnessing prior knowledge of protein structure to bridge the gap



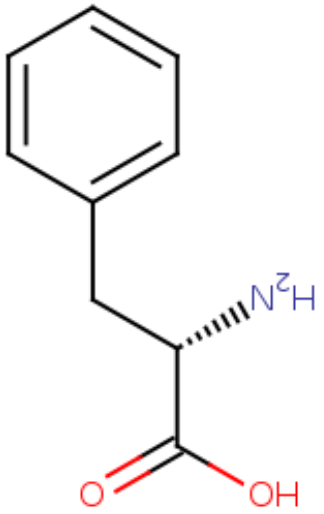
Stereochemistry

Bond lengths, angles, and torsion angles

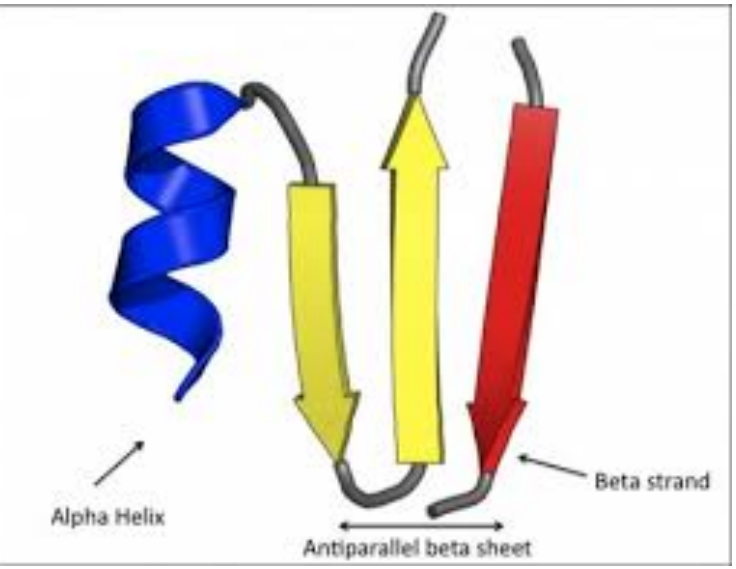


Introduction to protein structure, Branden & Tooze

Stereochemistry



Secondary Structure

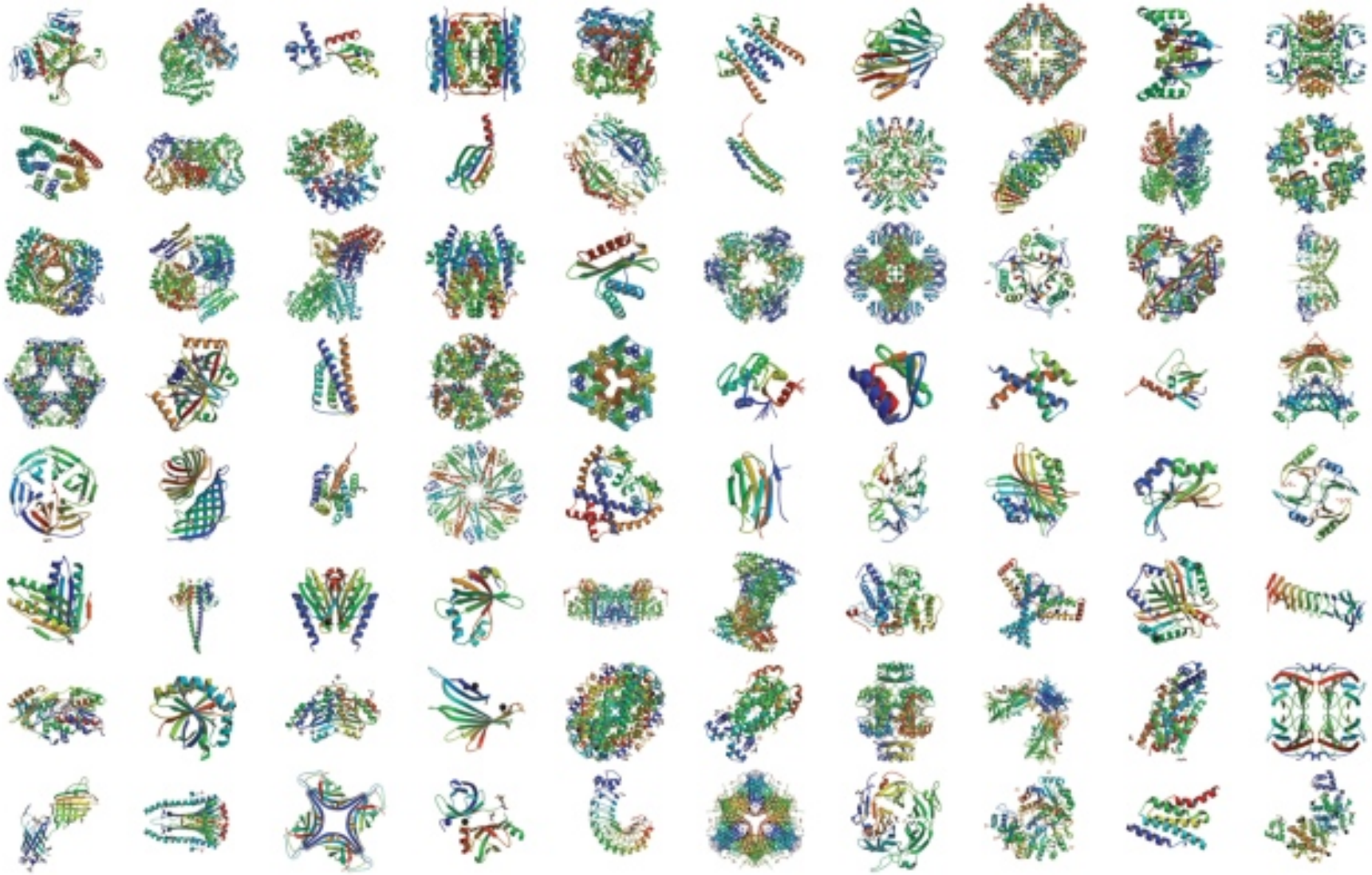


https://www.ebi.ac.uk/training/online/course/biomacromolecular-structures-introduction-ebi-reso/proteins/levels-protein-structure

Folds and domains

Limited number of protein folds (few thousand), and most have been solved

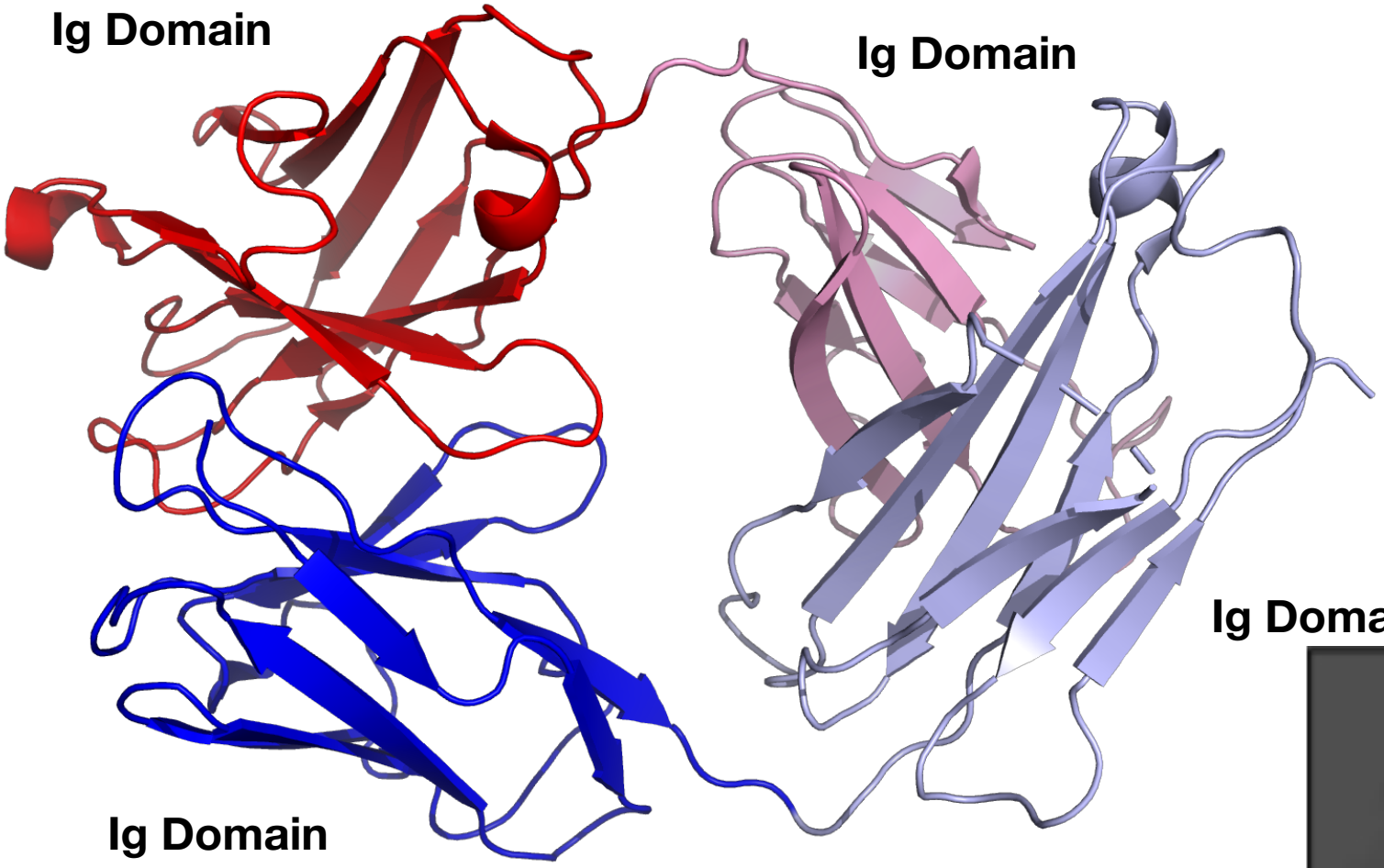
Most likely, *your favorite protein* is composed of one or more known domains

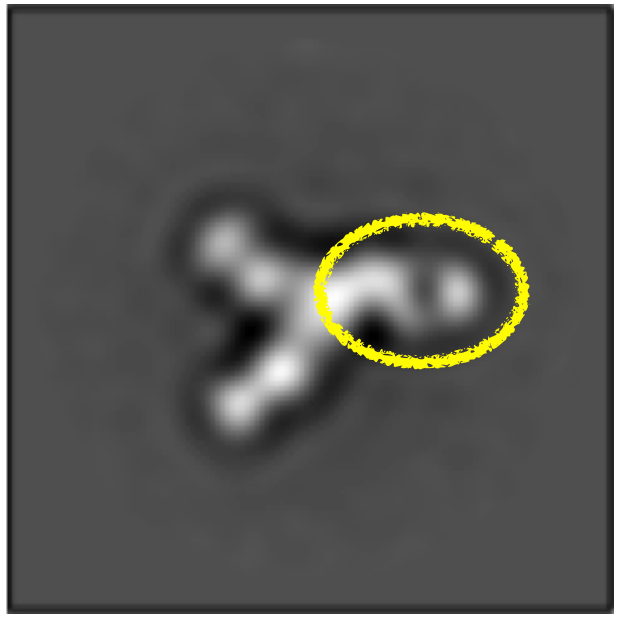


Information from known domains accelerates model building and assigning connectivity at modest resolution, and interpretation of low resolution maps where secondary structure is not apparent

Marc Elsliger

Quaternary Structure

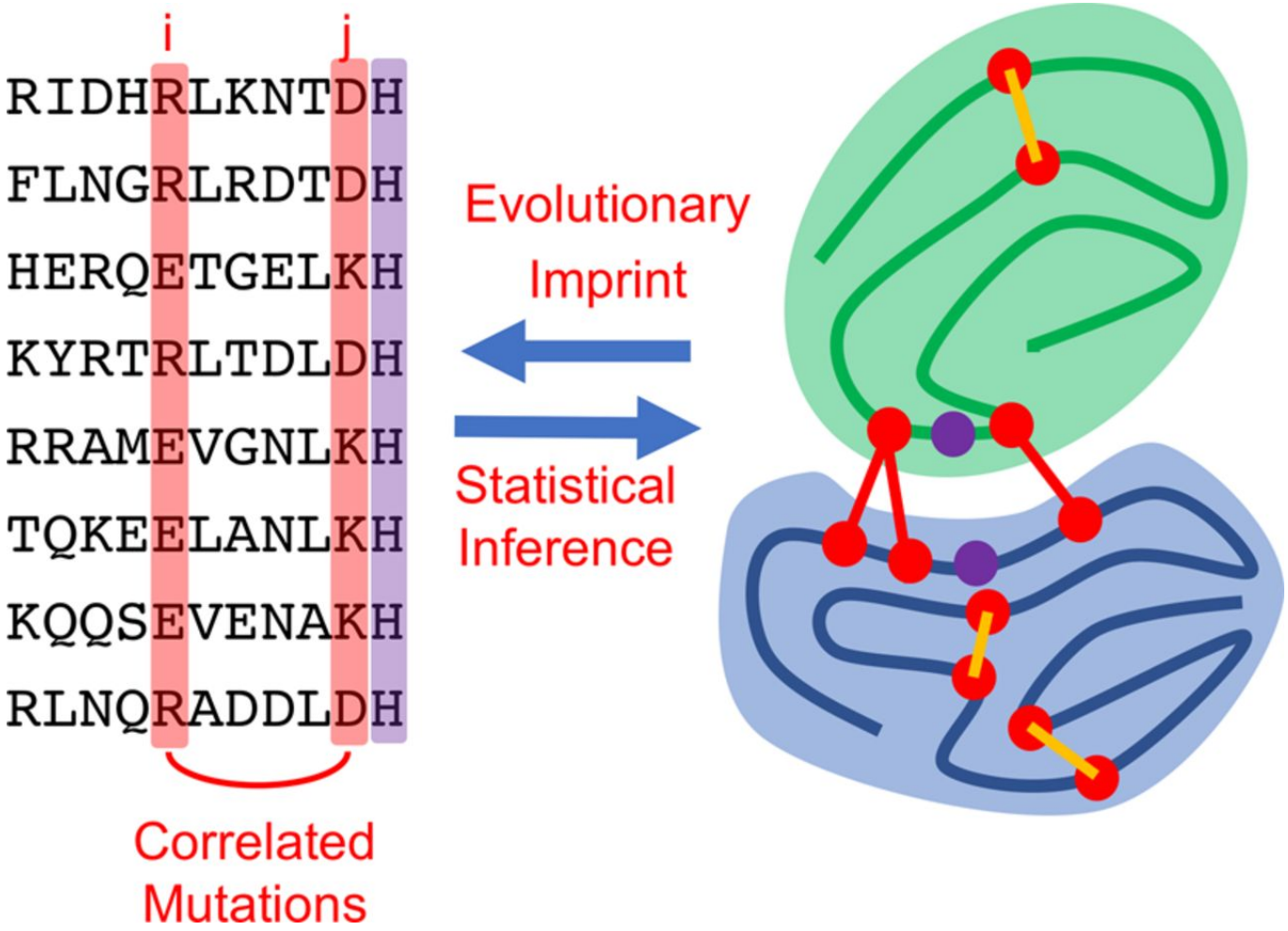




Biochemistry/Bioinformatics

Information about domain-domain interactions and residue-residue proximity can come from a variety of sources:

- **Yeast 2-Hybrid**
- **Co-IP**
- **Cross linking**
- **Co-evolving residue pairs**



Information can be used early to guide modeling, or later to validate the final model

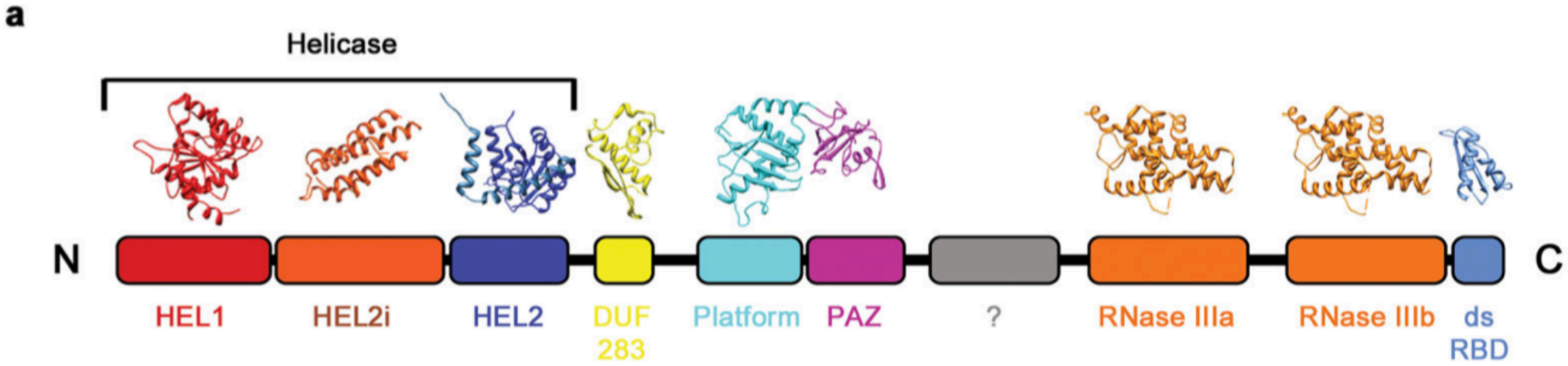
http://www.biochemsoctrans.org/content/45/6/1253.figures-only

1. ~15A - 10A Localizing domains in complexes/large proteins of unknown structure
    1. Example: human Dicer structure (Lau et al, *Nat Struct Mol Biol*. ; 19(4): 436–440. doi:10.1038/nsmb.2268)

2. ~9A - 7A Domain rearrangements in proteins of known structure
    1. Example: conformational changes in dynein motor domain (Niekamp et al, BioRxiv 2019)

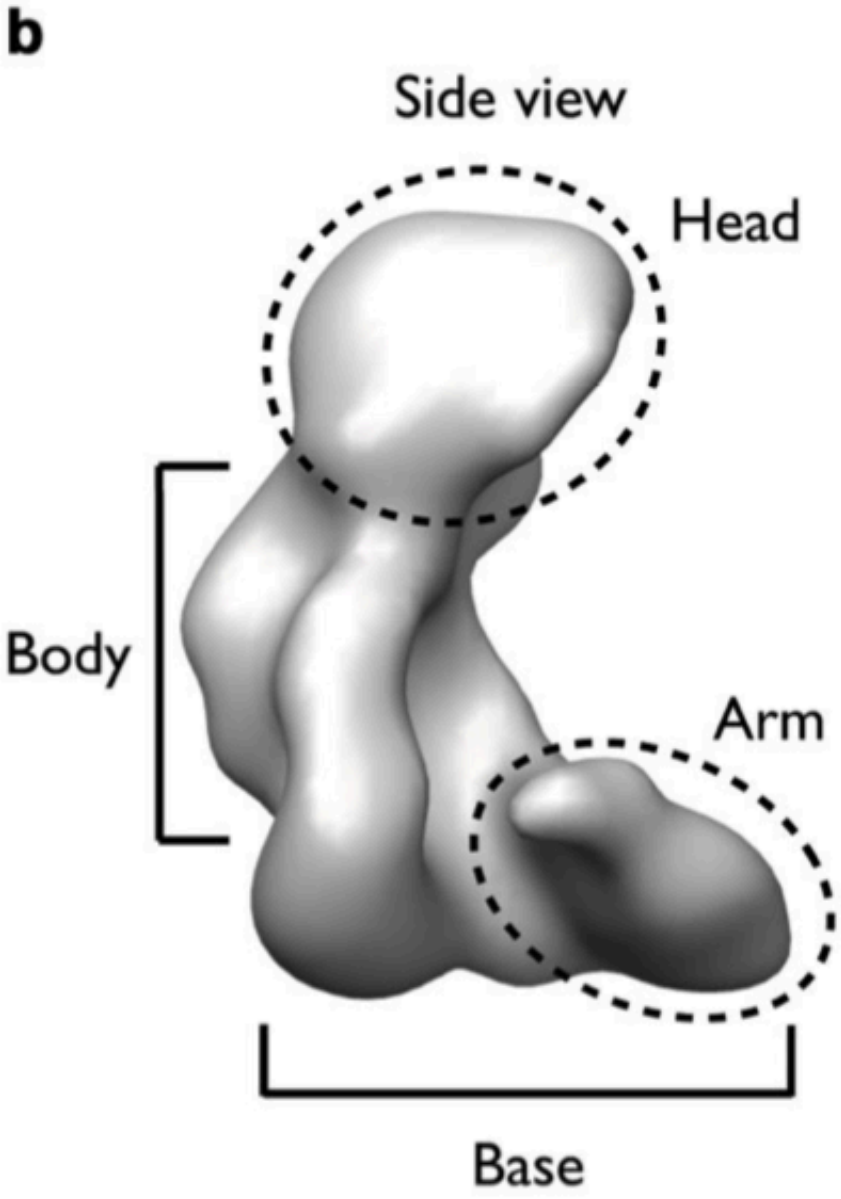
3. ~7A - 4A Interpreting secondary structure and “near-atomic” features

Target protein: human dicer

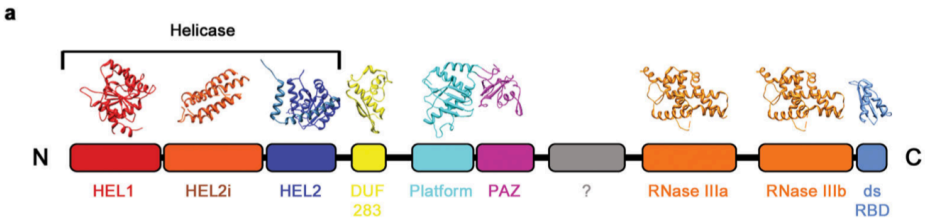


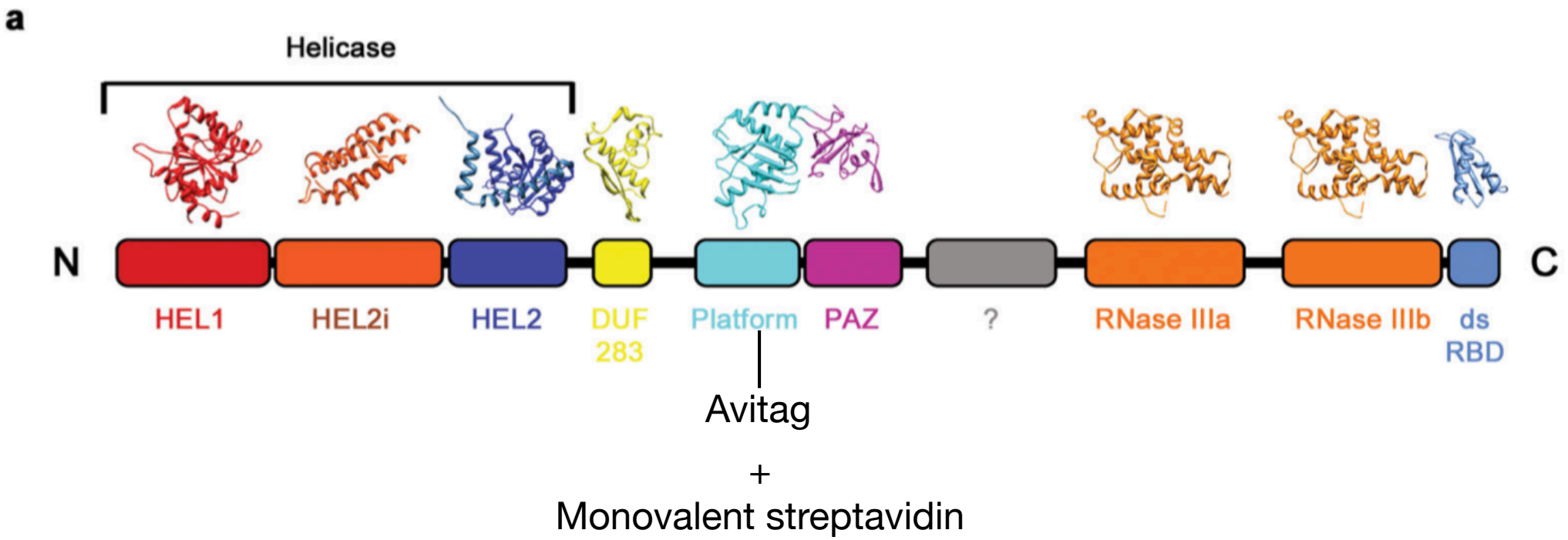
- Discreet domains
- Some crystal structures known
- No information on relative orientations

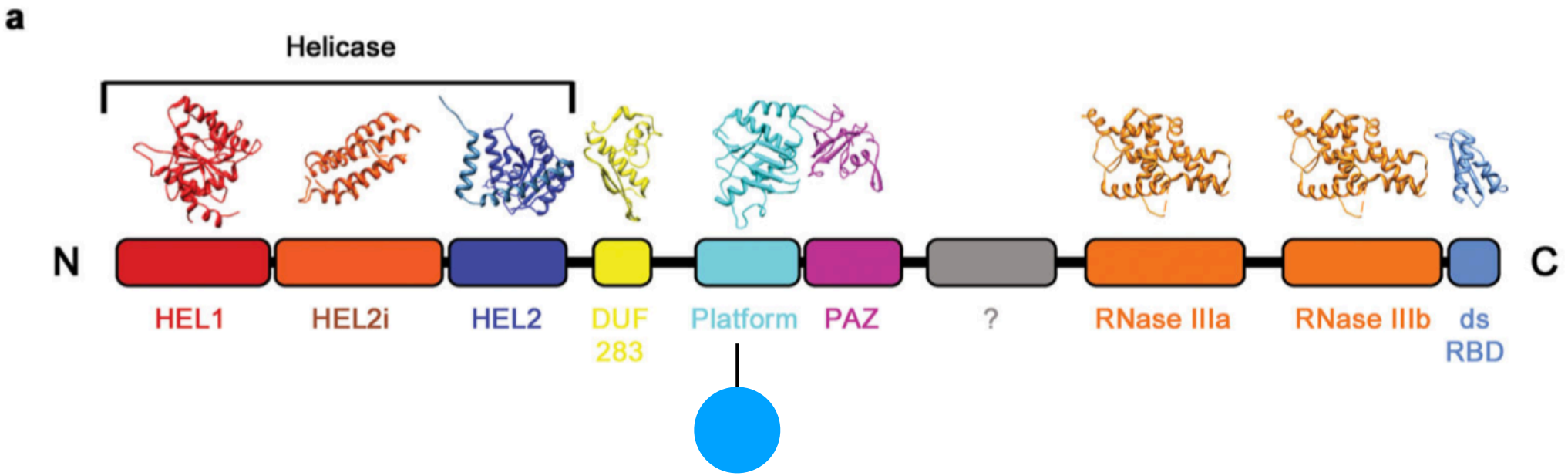
Low resolution map

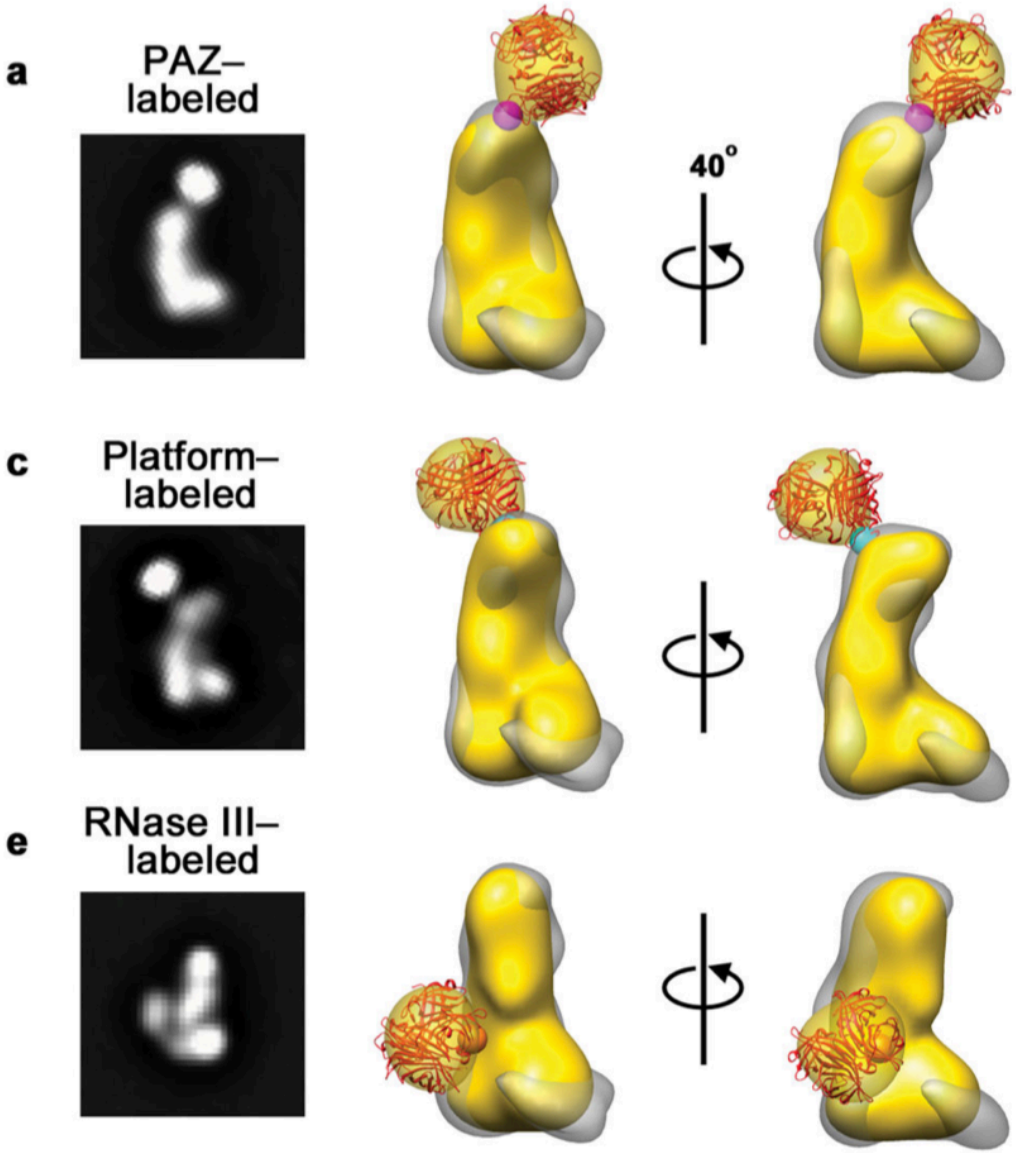


- Features are clearly visible
- What goes where?
- Can domains be fit into this map reliably?

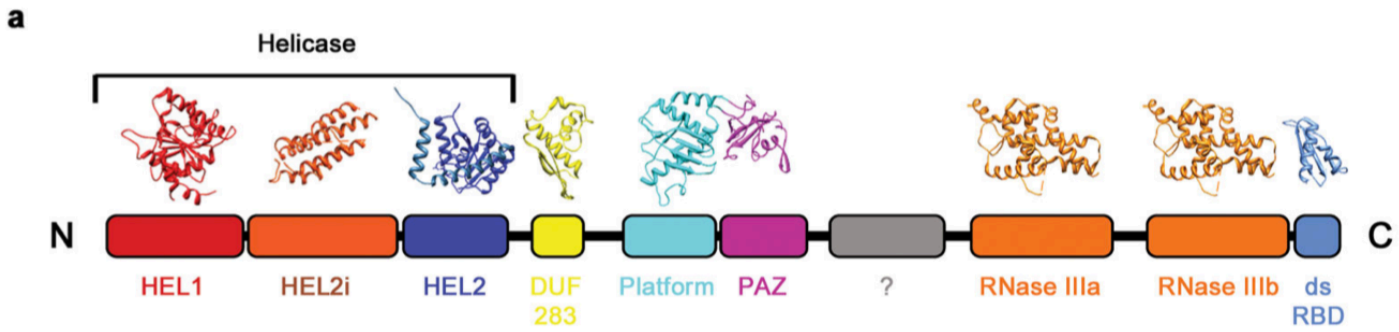


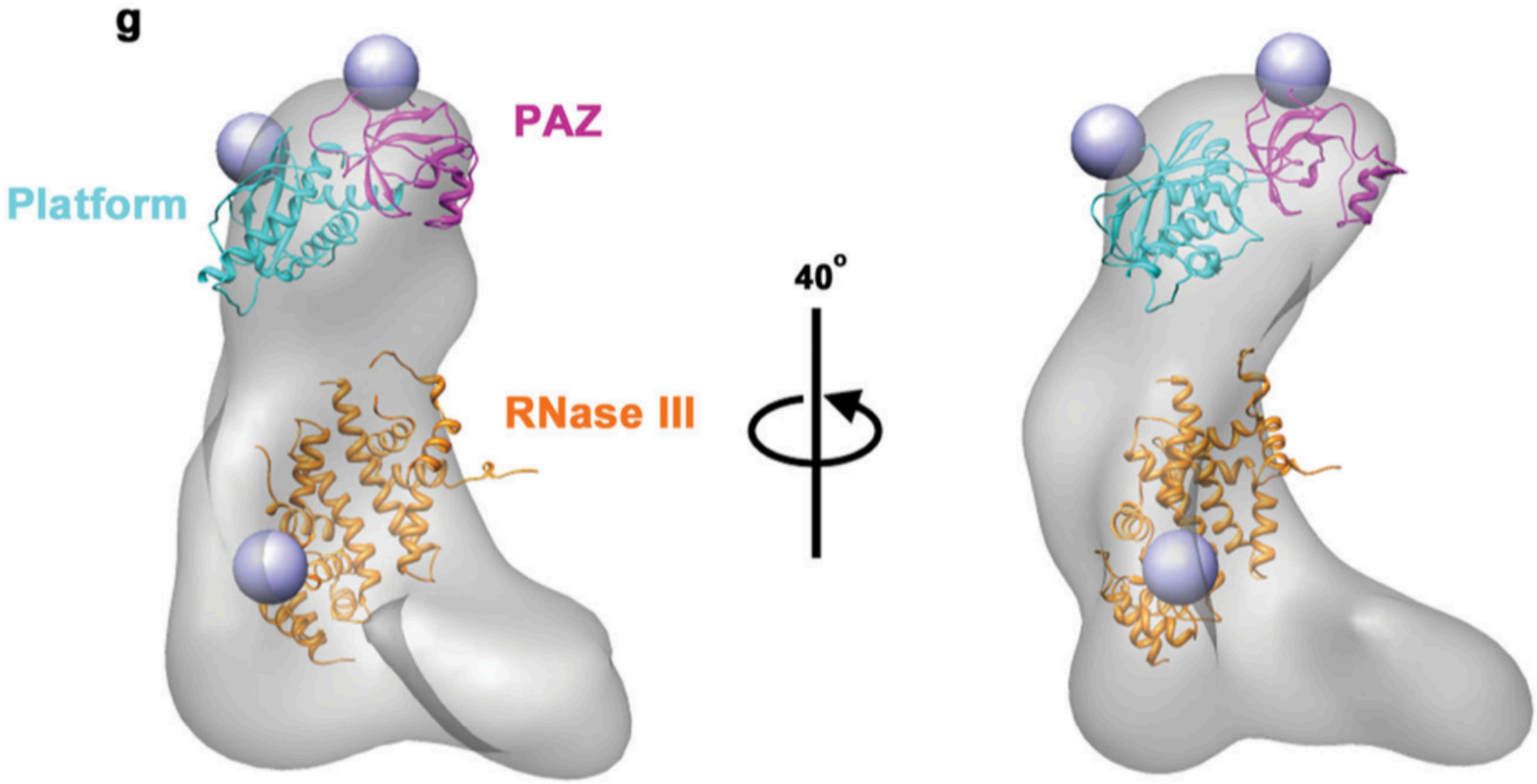

a
Helicase
N
C
HEL1
HEL2i
HEL2
DUF
283
Platform
PAZ
?
RNase IIIa
RNase IIIb
ds
RBD
Avitag
+
Monovalent streptavidin


a
Helicase
N
C
HEL1
HEL2i
HEL2
DUF
283
Platform
PAZ
?
RNase IIIa
RNase IIIb
ds
RBD



- Each domain labeled individually
- Data collected for each construct
- Streptavidin localized on each construct to map domain





- Even at low resolution, insights into relative domain organization are possible

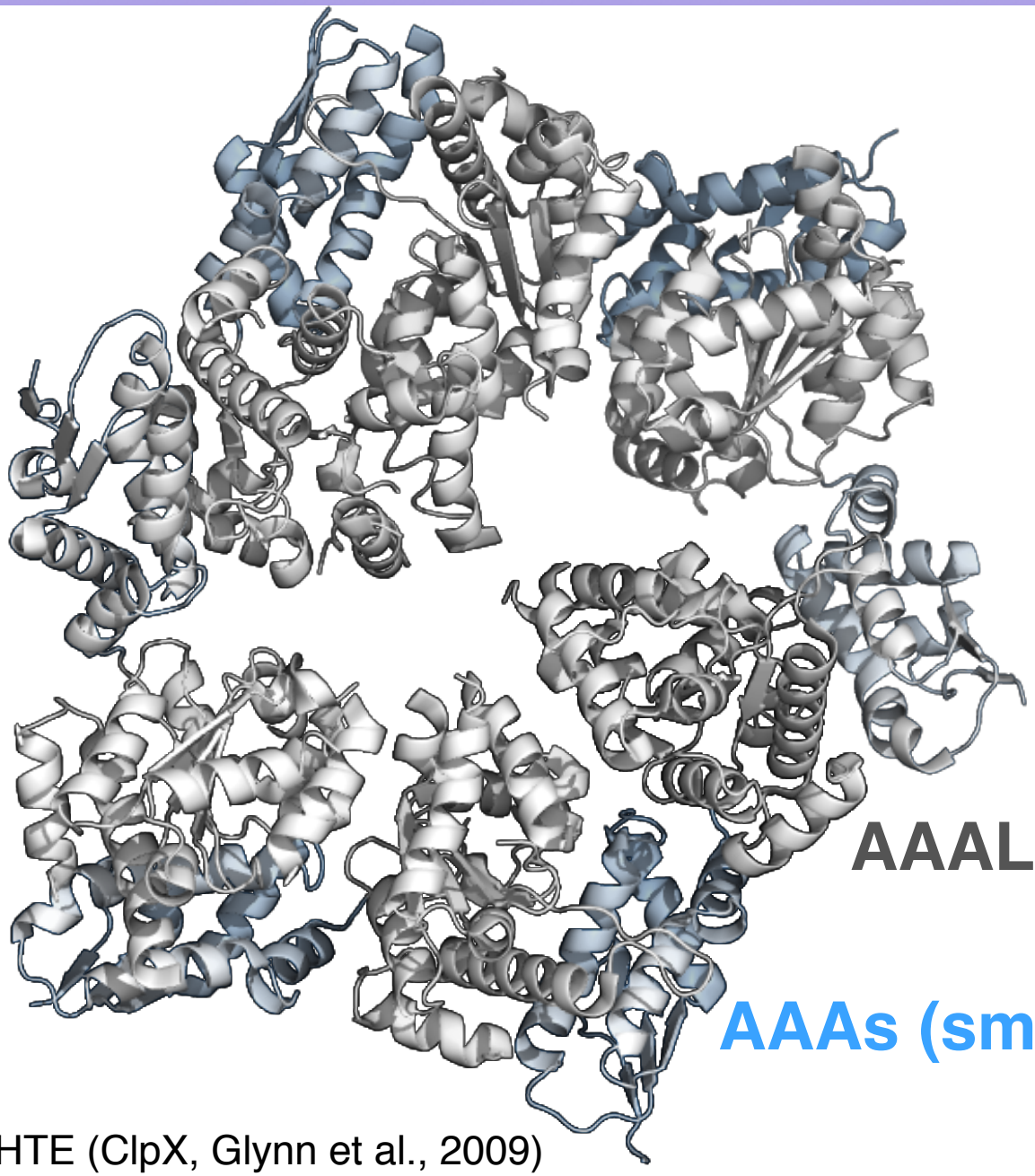
- Depending on your biological question at a given time, resolution where domain-level assignments are possible may be very insightful

1. ~15A - 10A Localizing domains in complexes/large proteins of unknown structure
    1. Example: human Dicer structure (Lau et al, *Nat Struct Mol Biol*. ; 19(4): 436–440. doi:10.1038/nsmb.2268)

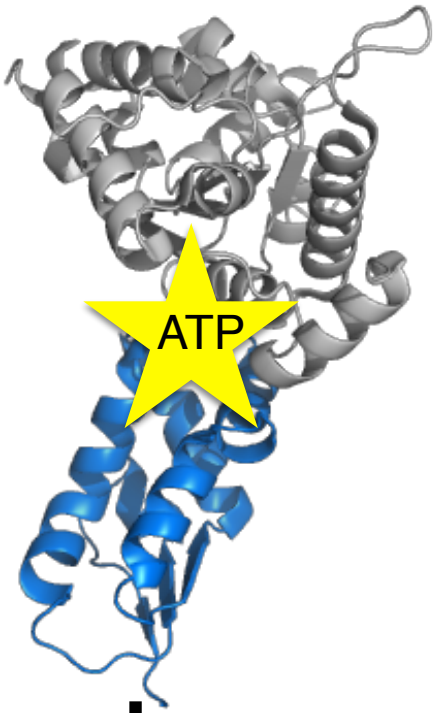
2. ~9A - 7A Domain rearrangements in proteins of known structure
    1. Example: conformational changes in dynein motor domain (Niekamp et al, BioRxiv 2019)

3. ~7A - 4A Interpreting secondary structure and “near-atomic” features

AAA (ATPases Associated with diverse cellular Activities) family of proteins are usually organized as homo-hexametric rings



3HTE (ClpX, Glynn et al., 2009)



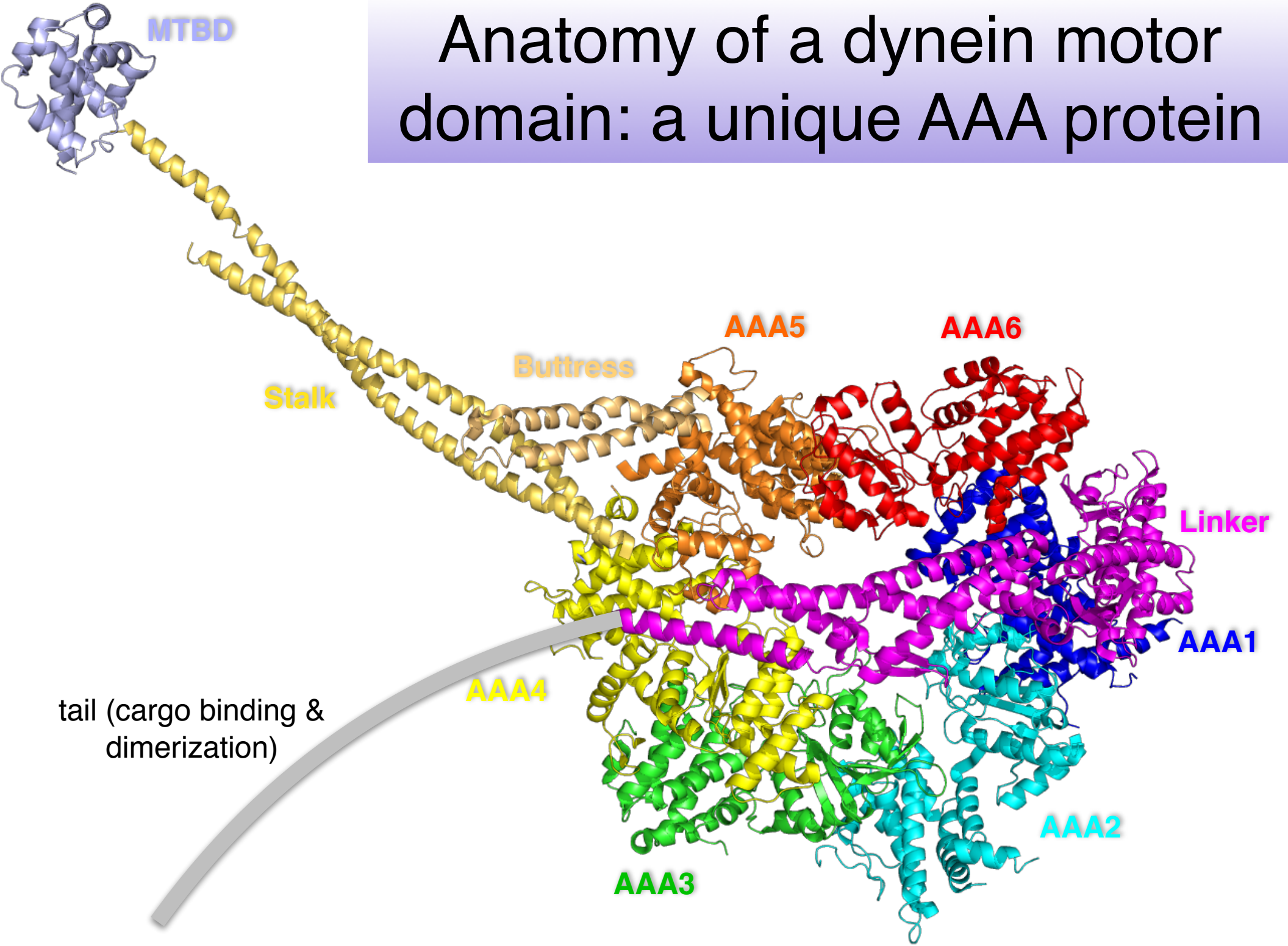
Conformational changes

AAAL (large domain)

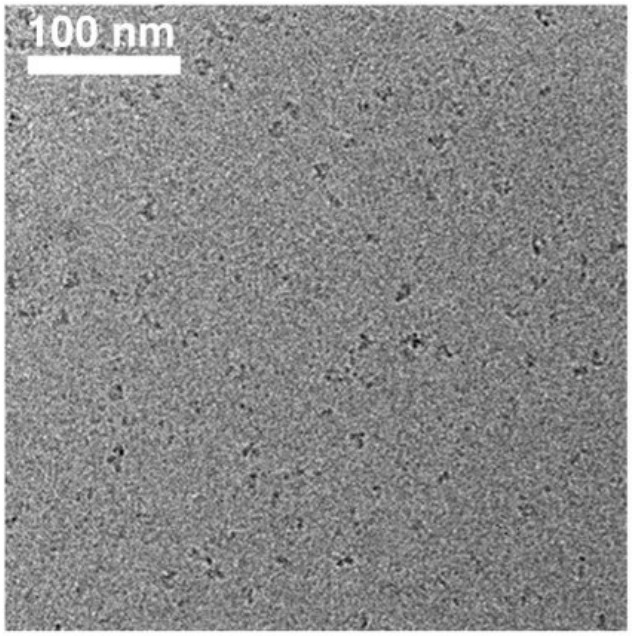
function

AAAs (small domain)

Anatomy of a dynein motor domain: a unique AAA protein

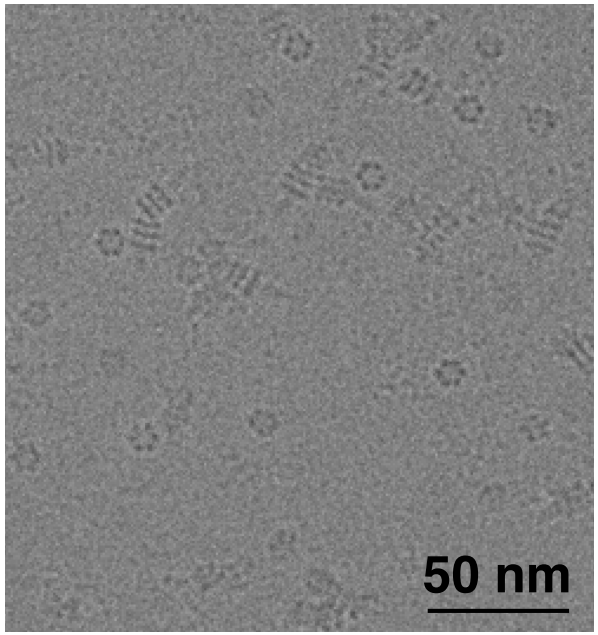


The dynein motor domain is conformationally heterogeneous and not a stellar cryo EM sample*



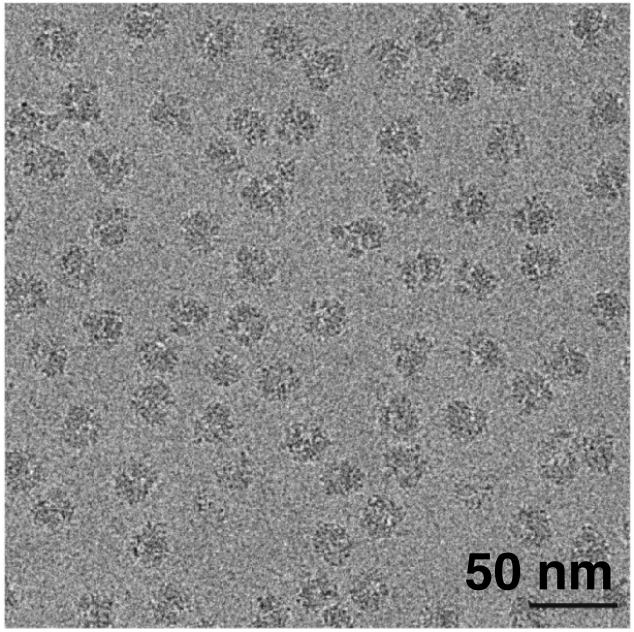
dynein AAA ring

😟



bacterial lipid transporter

😀

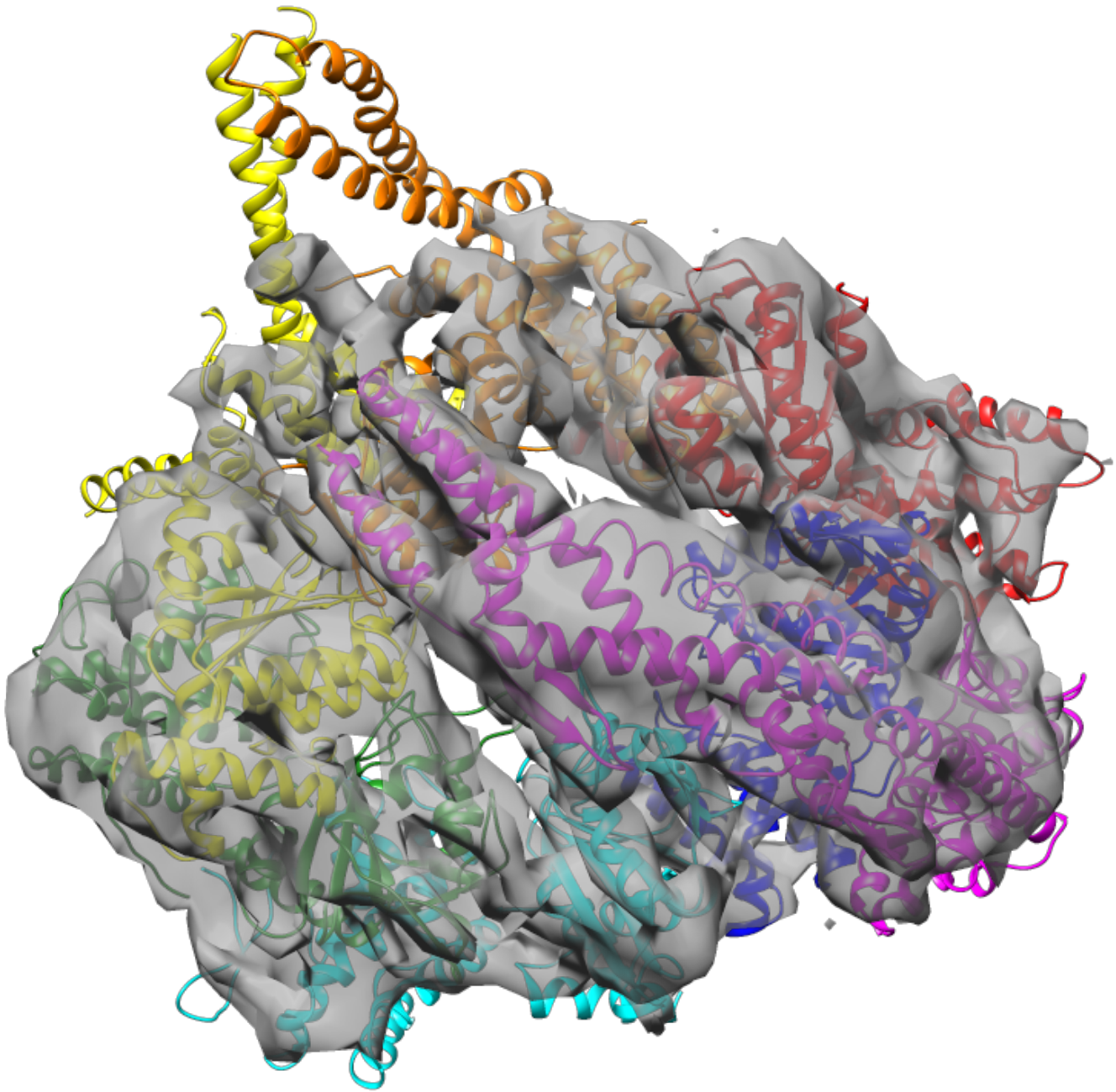


Ribosome, Bai *et al.*, 2013

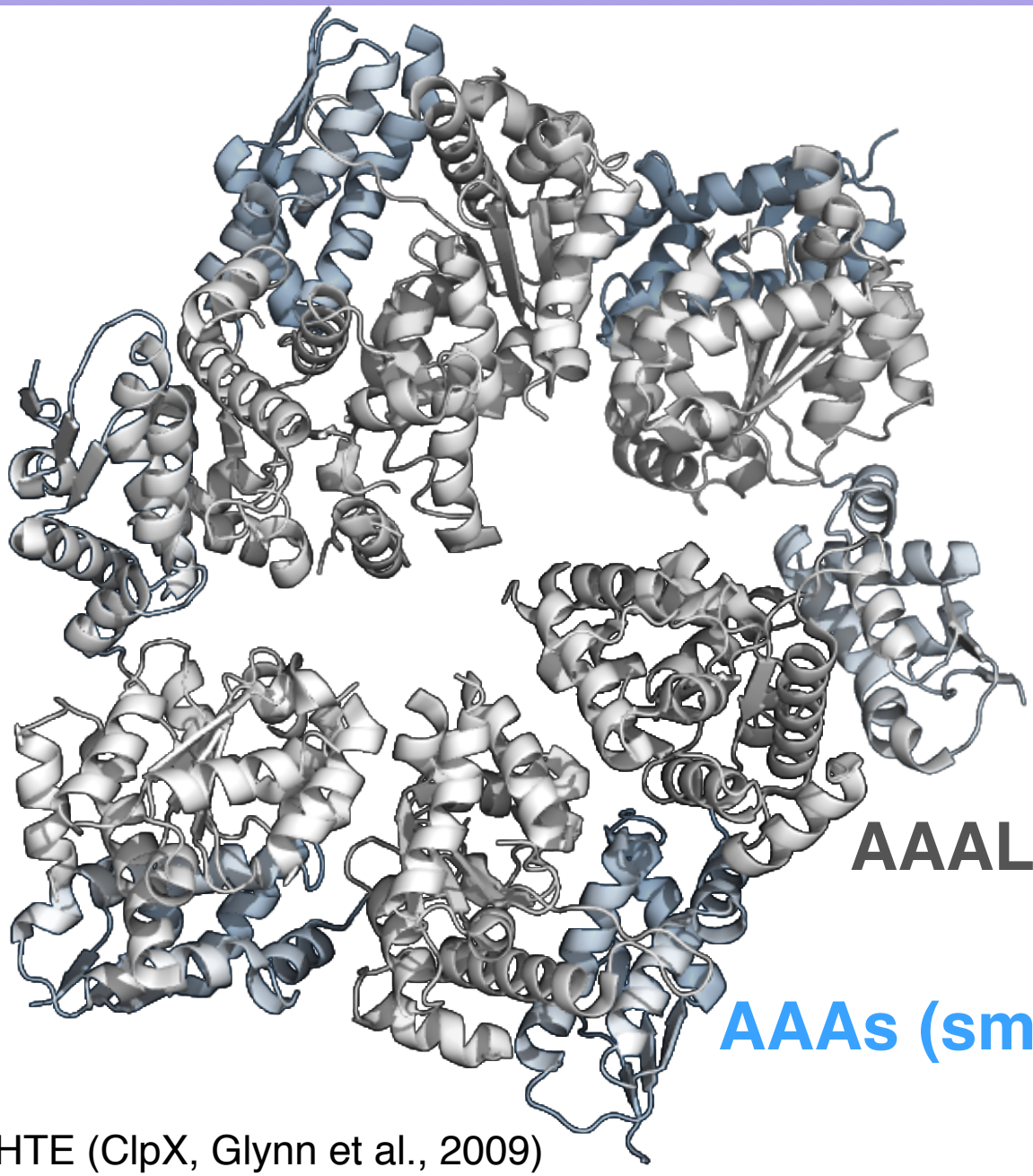
😃

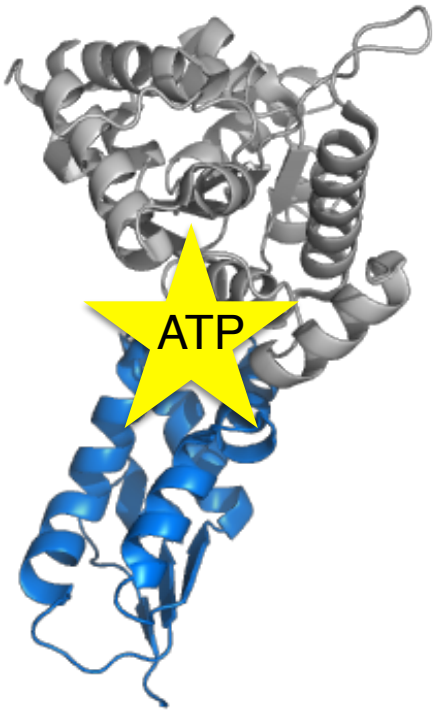
*newer sample preparation strategies may be a game changer…

In “good” areas, secondary structure can be seen, in “average” areas, not so much



But we have enough information to fit each subdomain as a rigid body simultaneously (Chimera)…



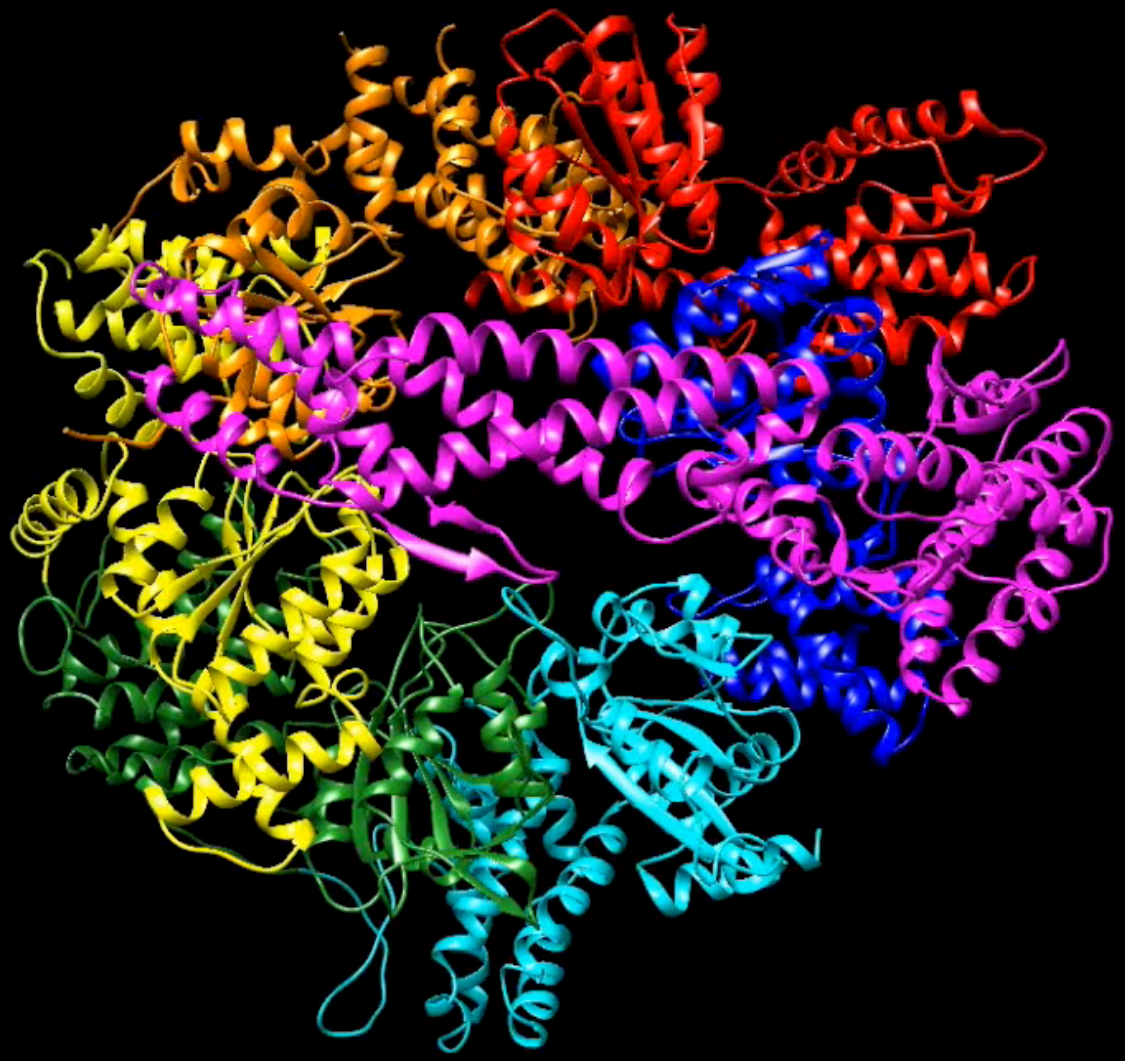


AAAL (large domain)

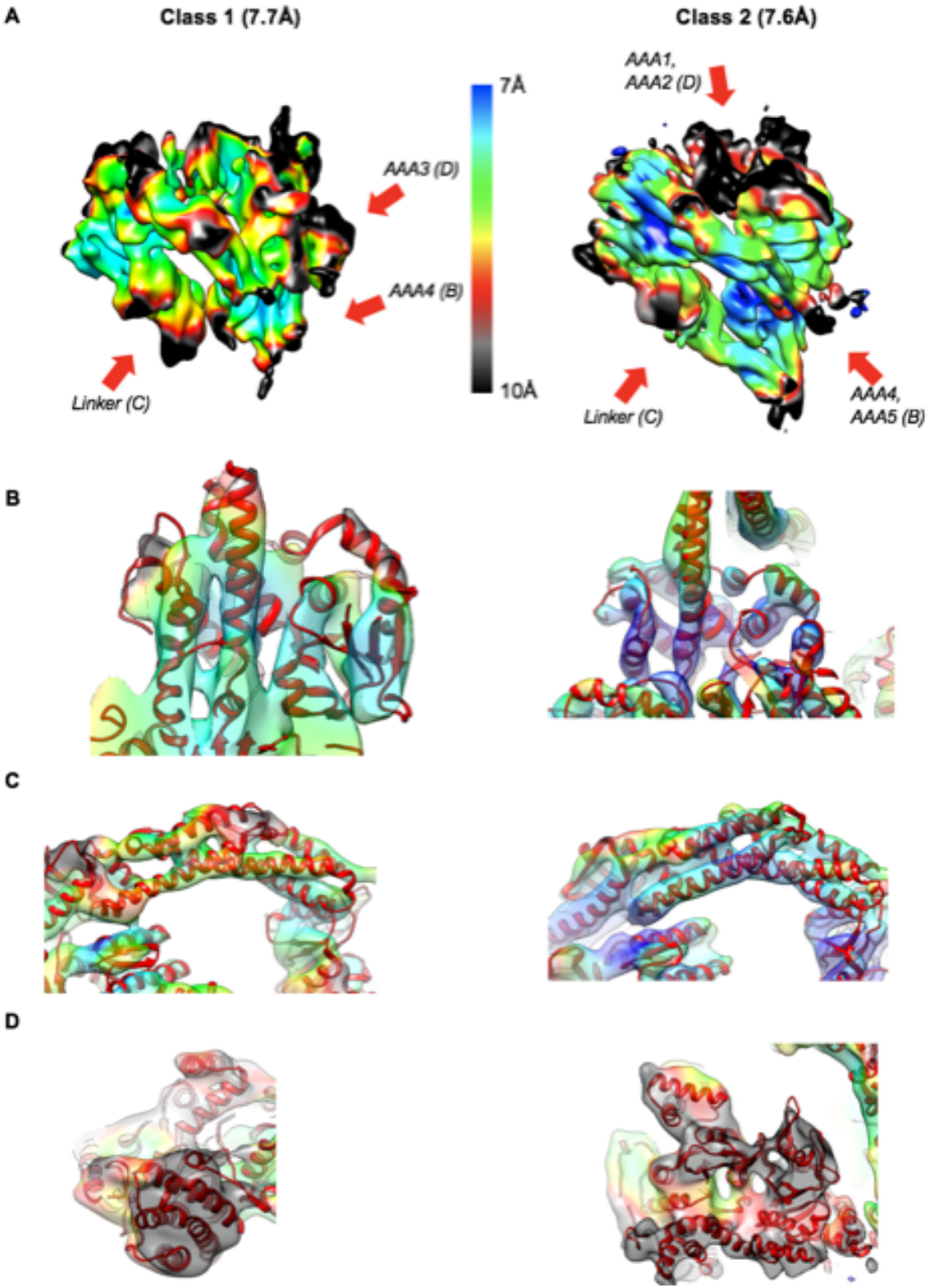
AAAs (small domain)

3HTE (ClpX, Glynn et al., 2009)

Useful info #1: nucleotide-dependent subdomain-level conformational changes*



*snapshots here from X-ray & EM



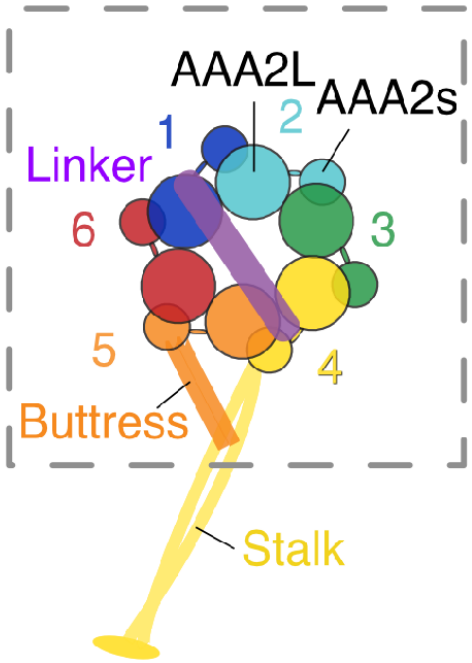
Dynein mutant: two classes distinguished at lower resolution

Key aspects of interpretation:

- Inspect the features of the density in the local area of interpretation
- Confirm ligand binding using **biochemistry/ functional assays** with the same protein prep (ligands are not going to be visible at this resolution)
- Try a few different classification strategies to convince yourself that the classes you find are “real”

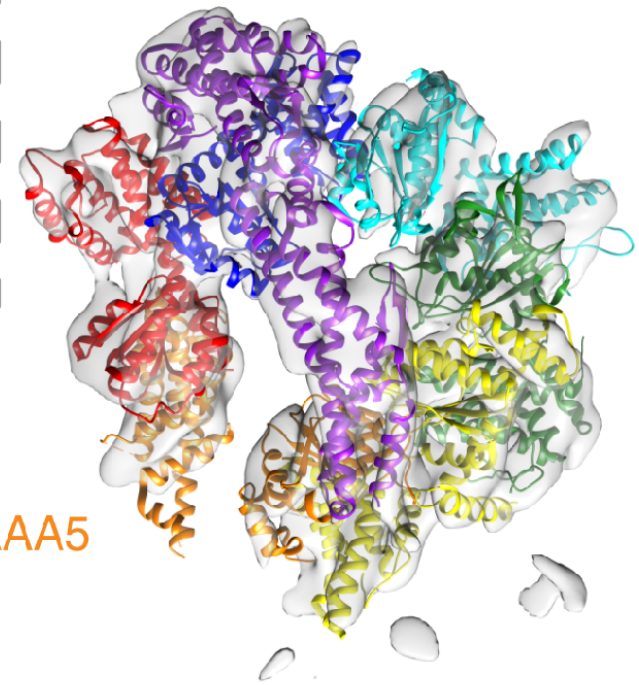
A

Mutant 5 : AMPPNP



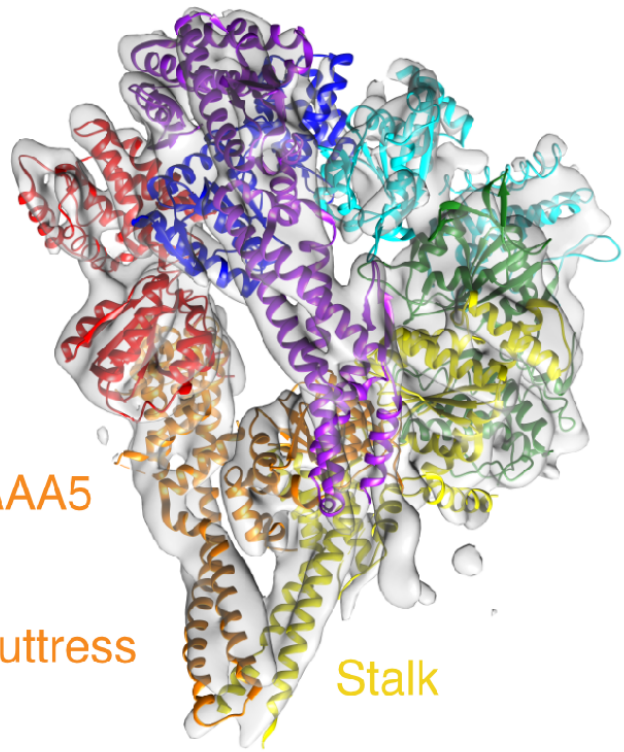
Class 1

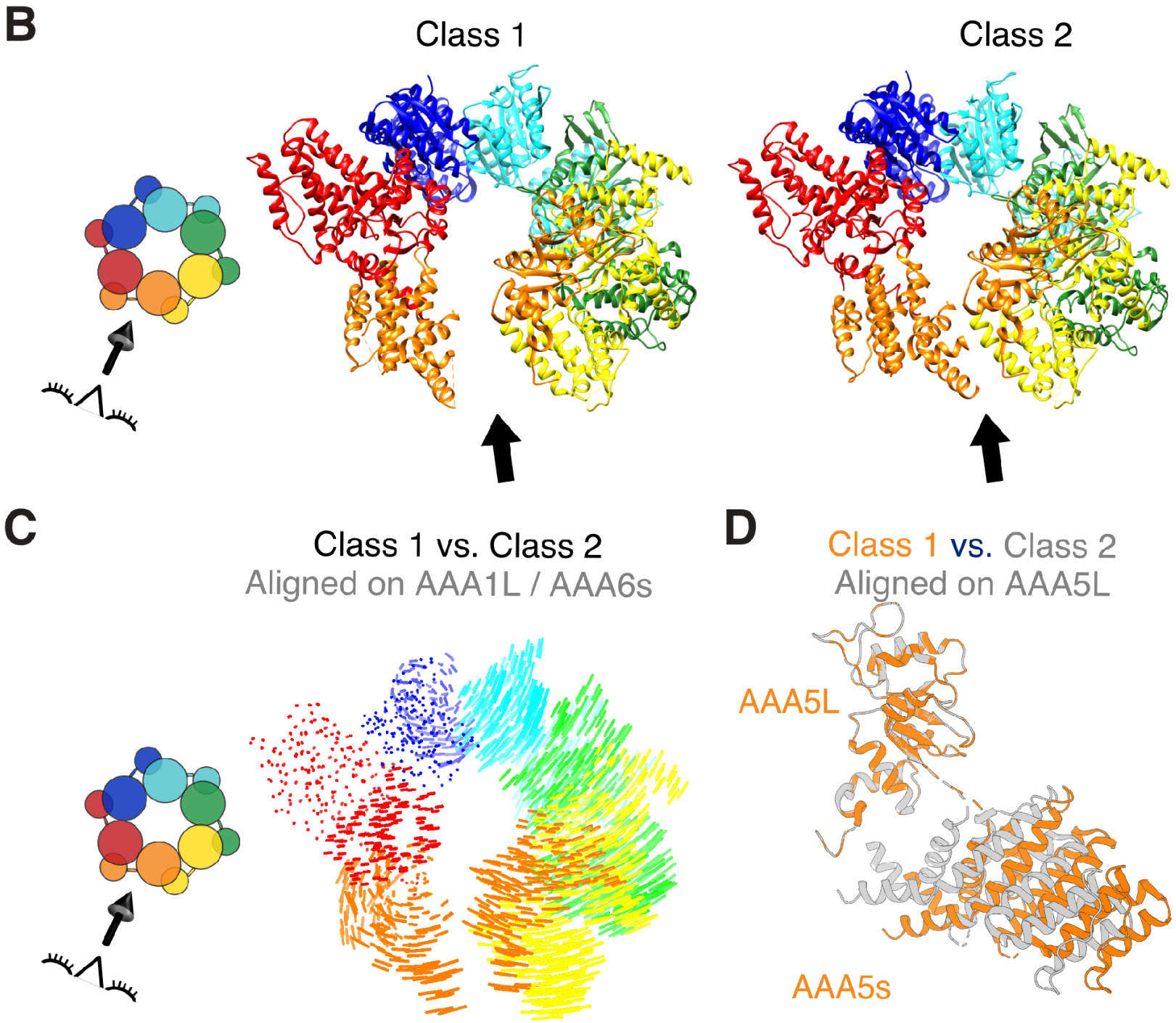
71.3%



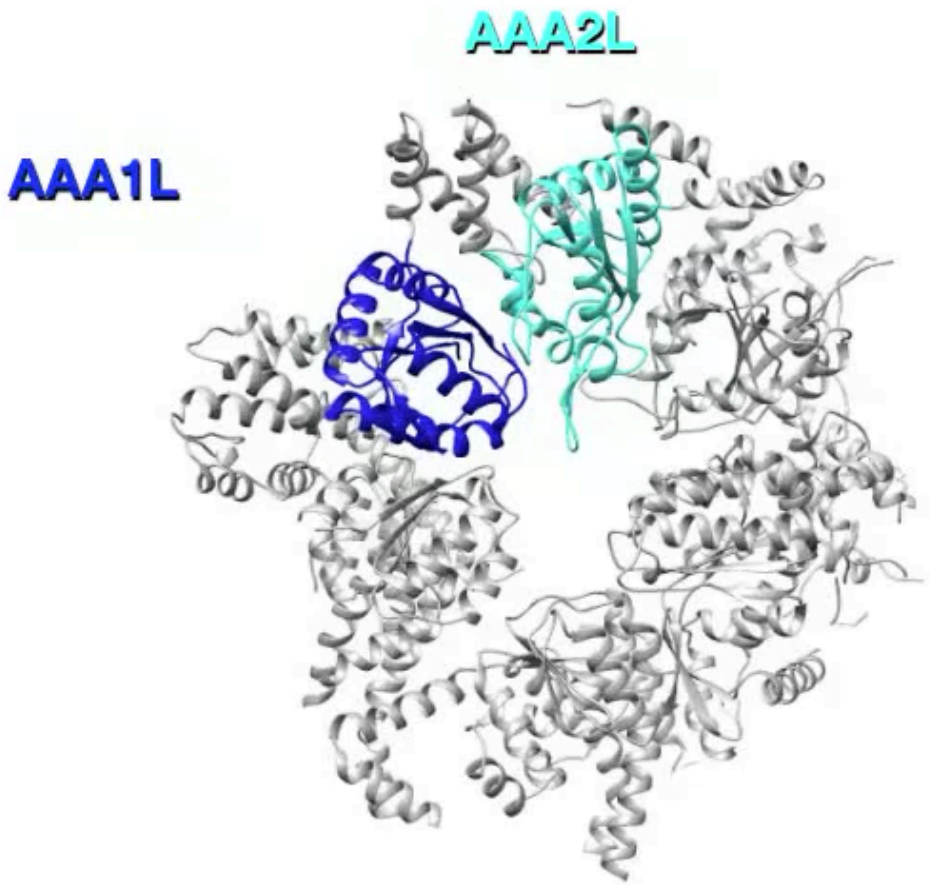
Class 2

28.7%




B
Class 1
Class 2
C
Class 1 vs. Class 2
Aligned on AAA1L / AAA6s
D
Class 1 vs. Class 2
Aligned on AAA5L
AAA5L
AAA5s

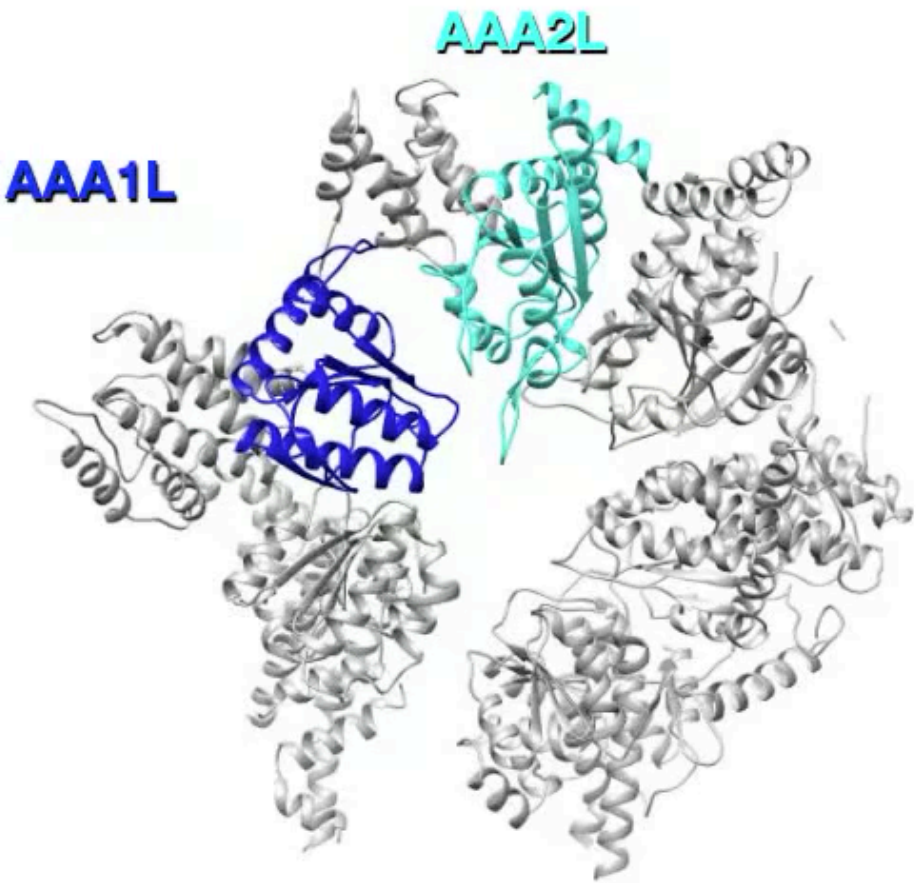
Wild-type



ADP-Vi
Wild-type
PDB: 4RH7

AMPPNP
Wild-type
PDB: 4W8F

Mutant



ADP-Vi
Mutant 5

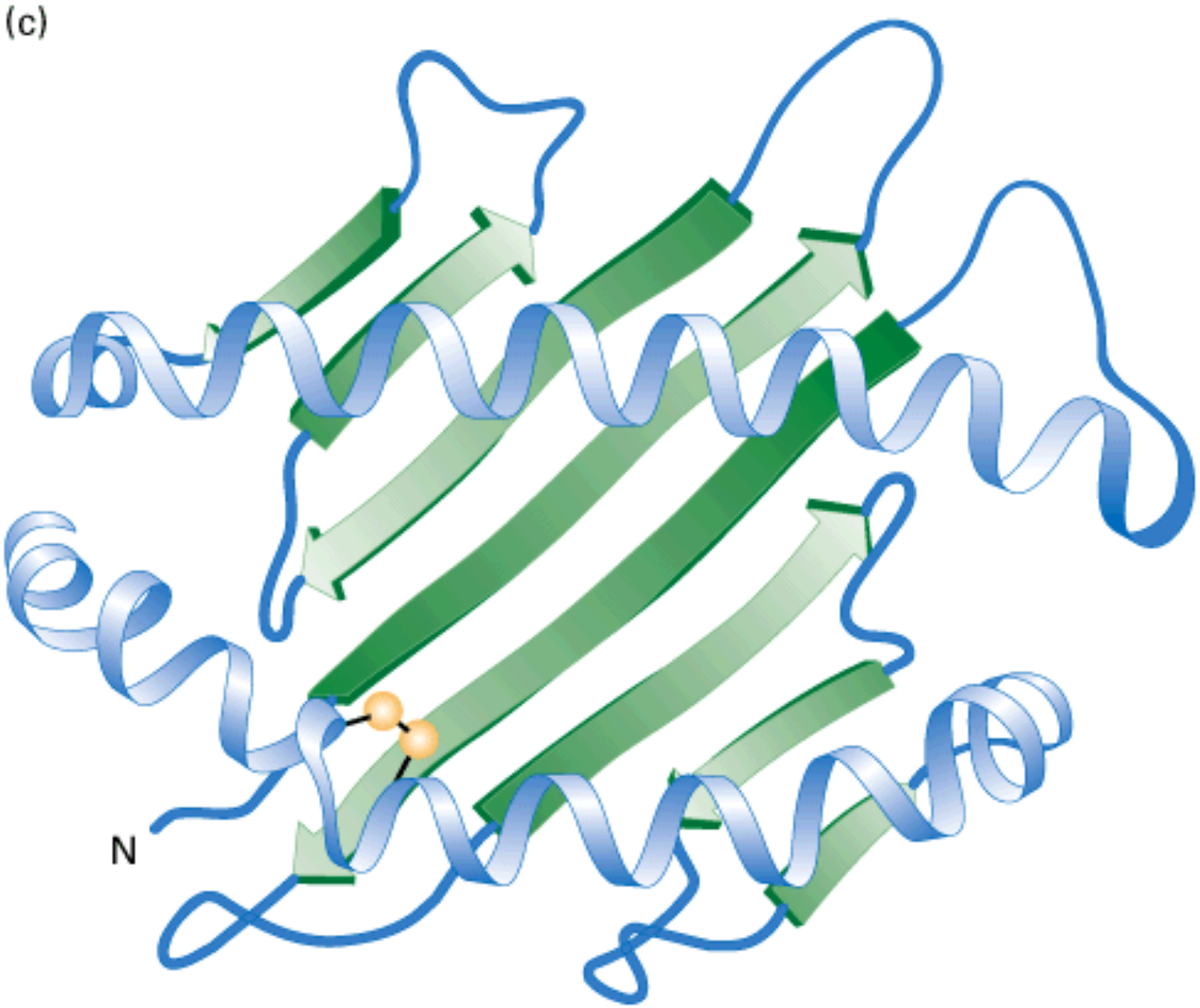
AMPPNP
Mutant 5
Class 1

1. **~15A - 10A Localizing domains in complexes/large proteins of unknown structure**
    1. **Example: human Dicer structure (Lau et al, *Nat Struct Mol Biol*. ; 19(4): 436–440. doi:10.1038/nsmb.2268)**

2. **~9A - 7A Domain rearrangements in proteins of known structure**
    1. **Example: conformational changes in dynein motor domain (Niekamp et al, BioRxiv 2019)**

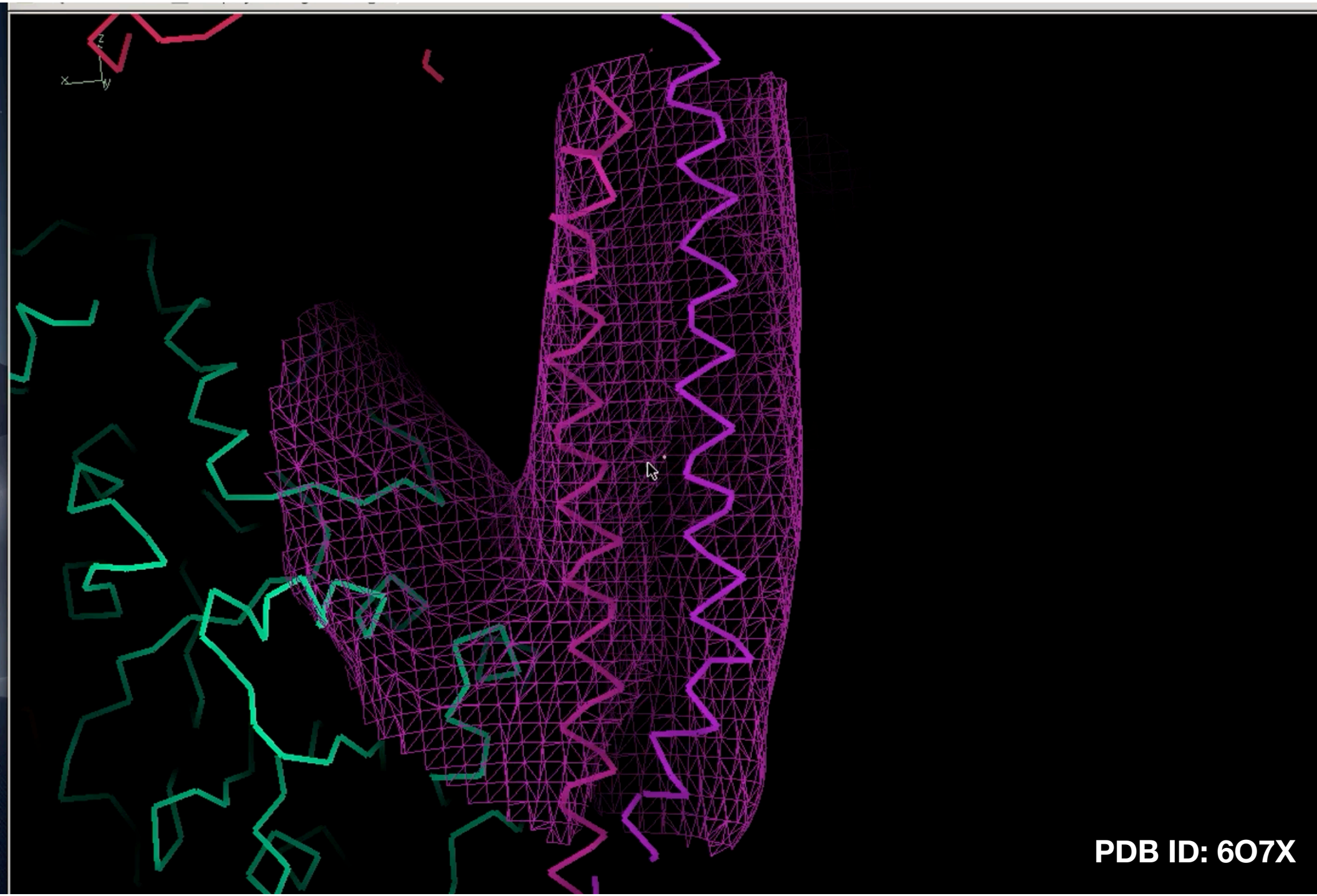
3. **~7A - 4A Interpreting secondary structure and “near-atomic” features**

Finding secondary structure in maps

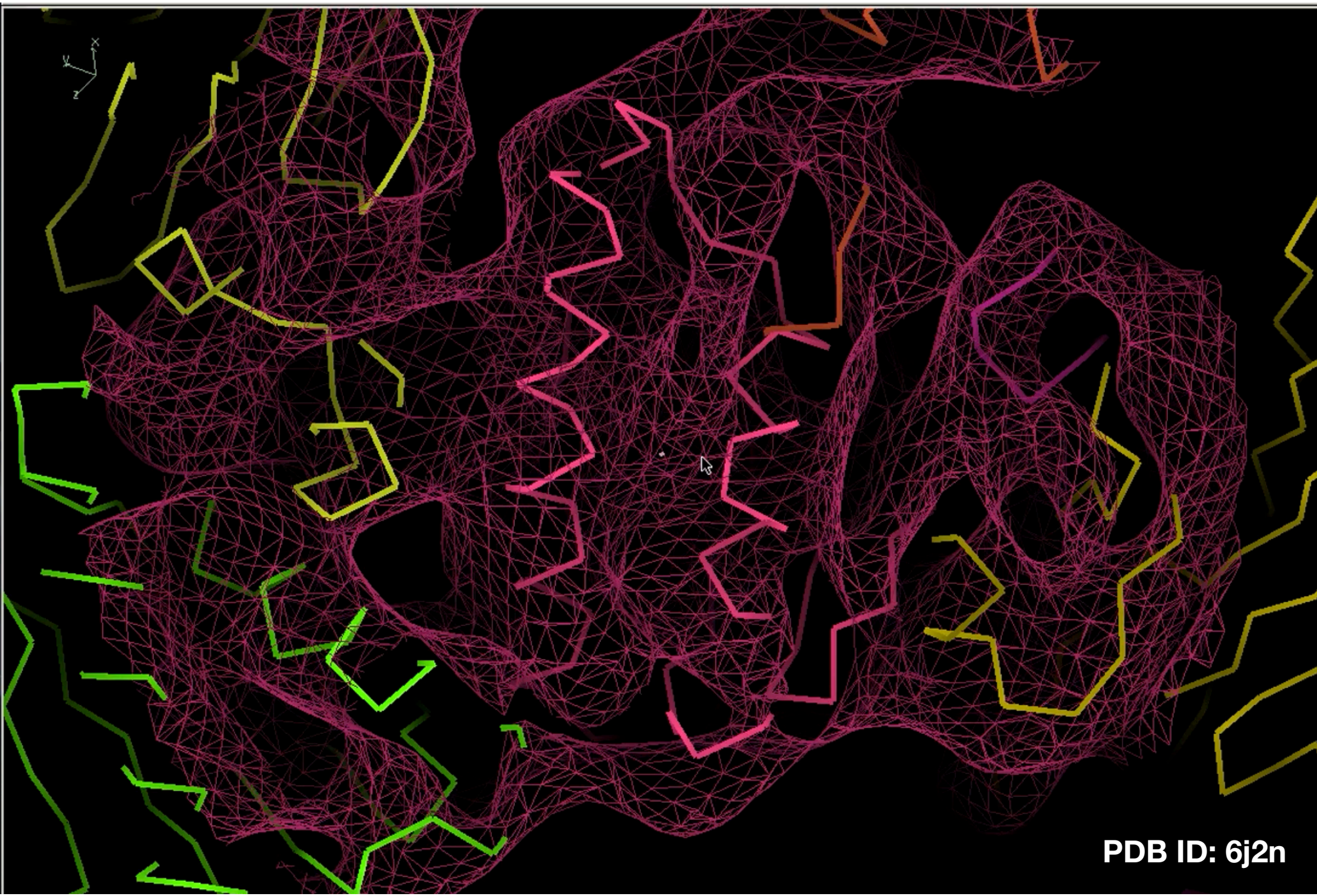


https://owlcation.com/stem/What-Are-Proteins-Part-3-of-3

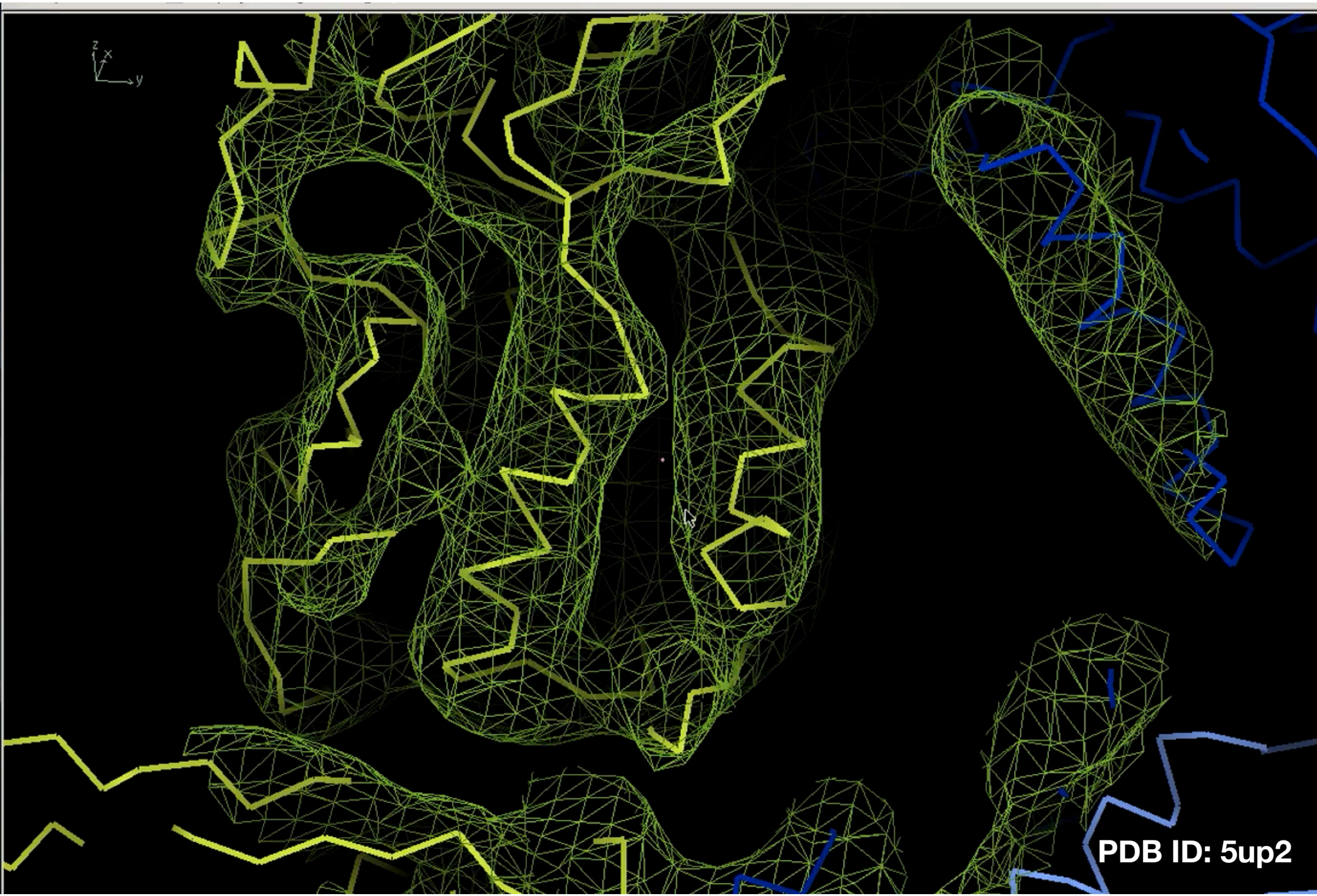
Helices at ~8.5 Å Resolution



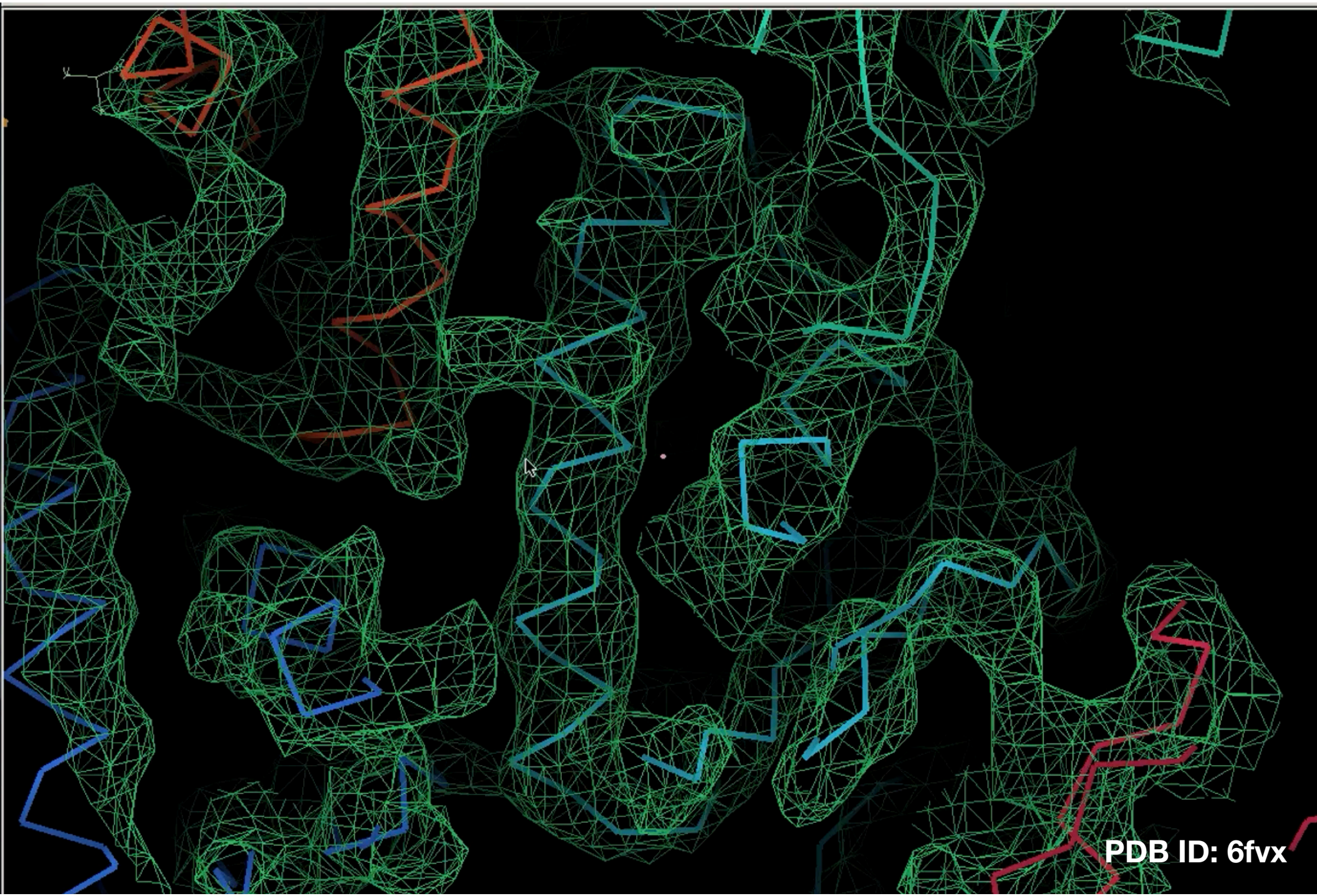
Helices at ~7.5 Å Resolution



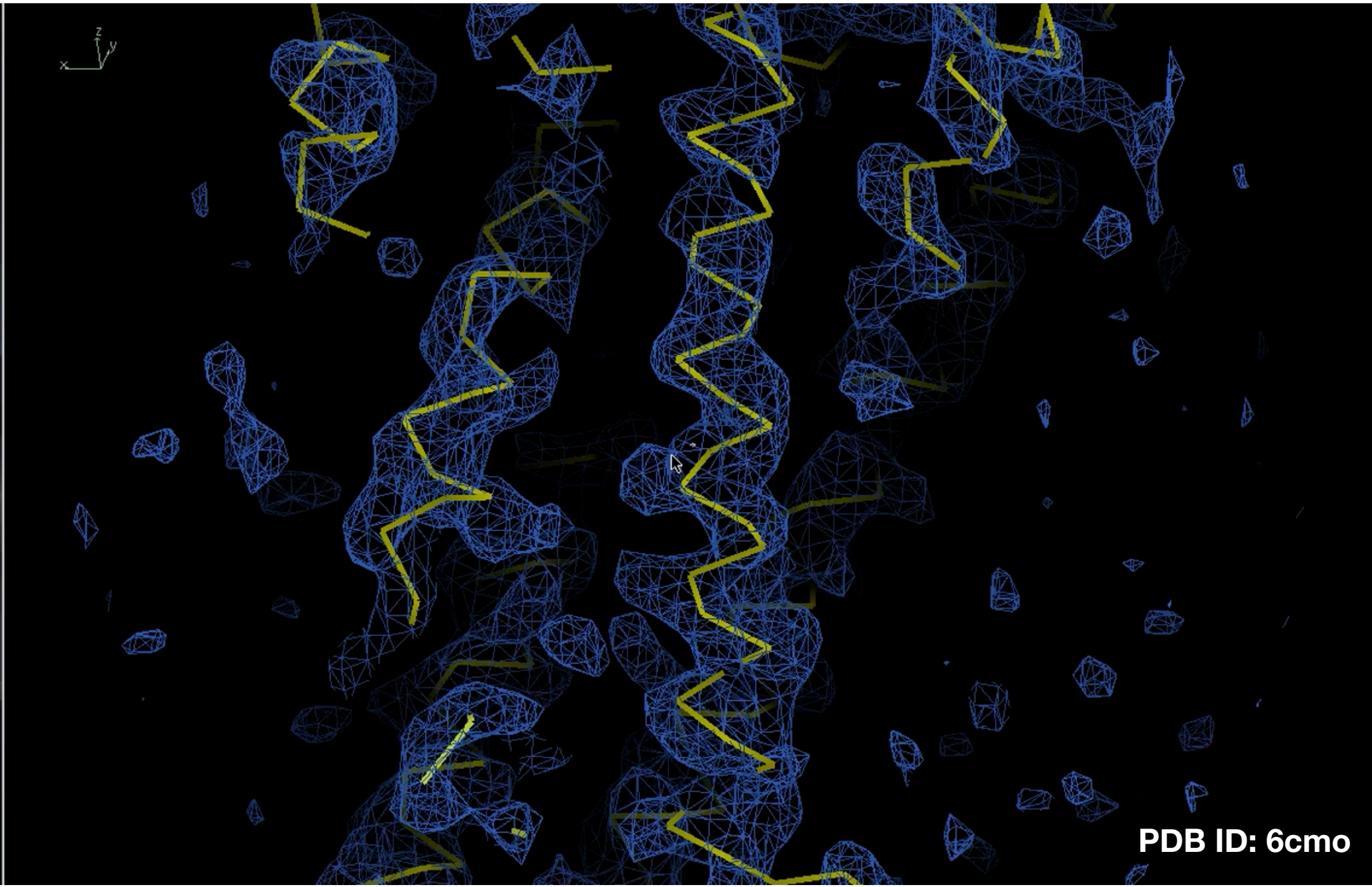
Helices at ~6 Å Resolution



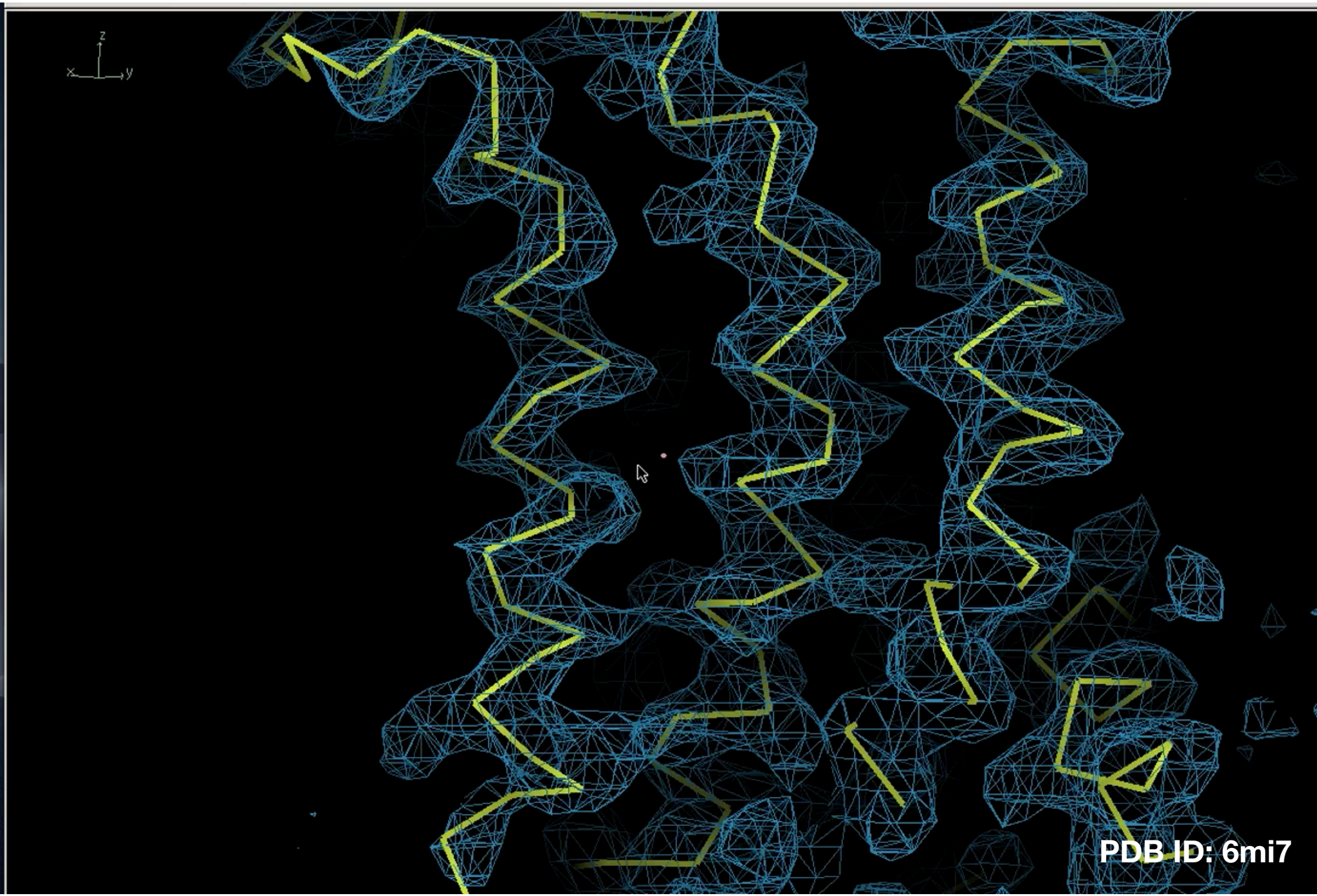
Helices at ~4.9 Å Resolution



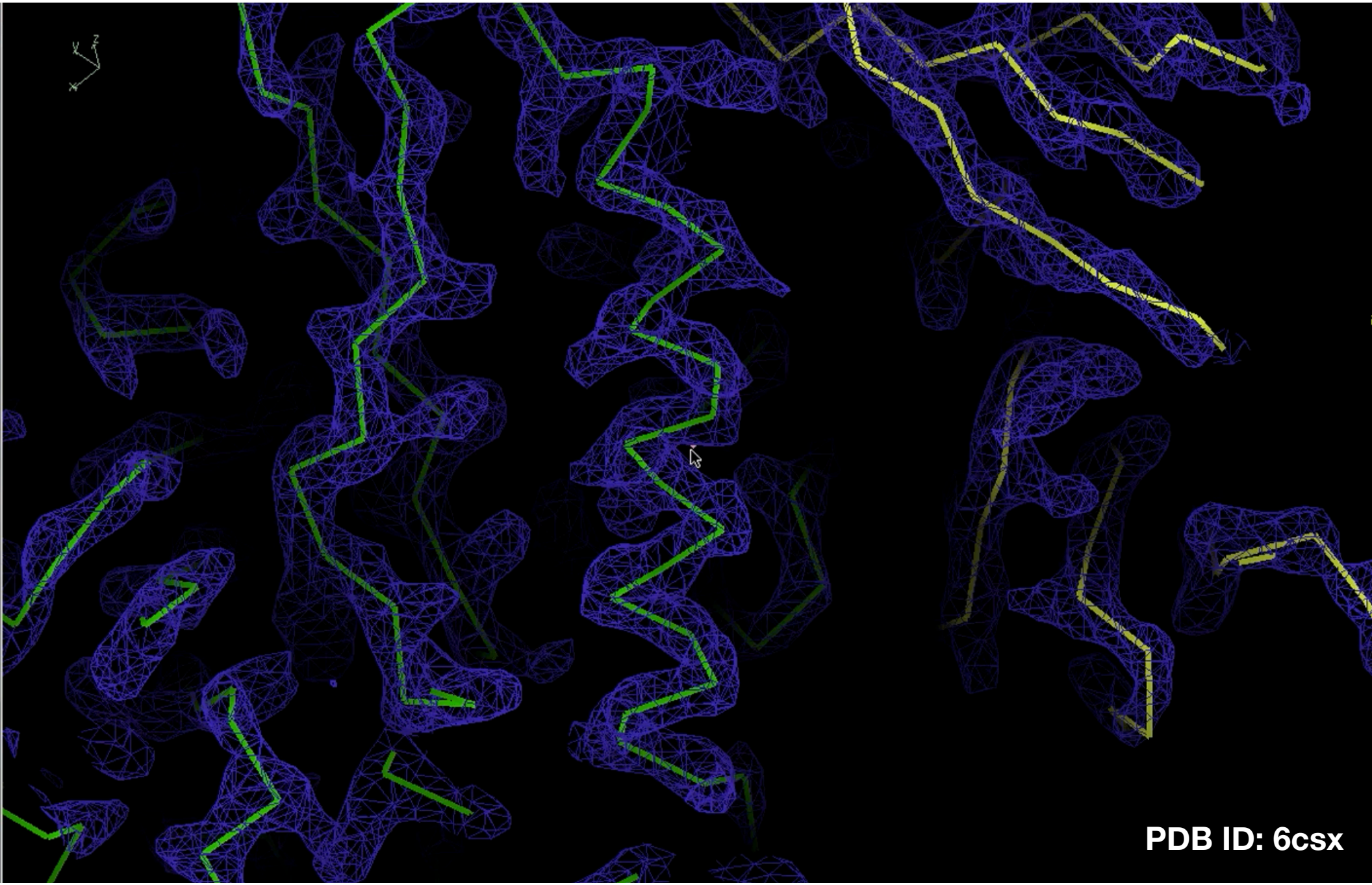
Helices at ~4.5 Å Resolution



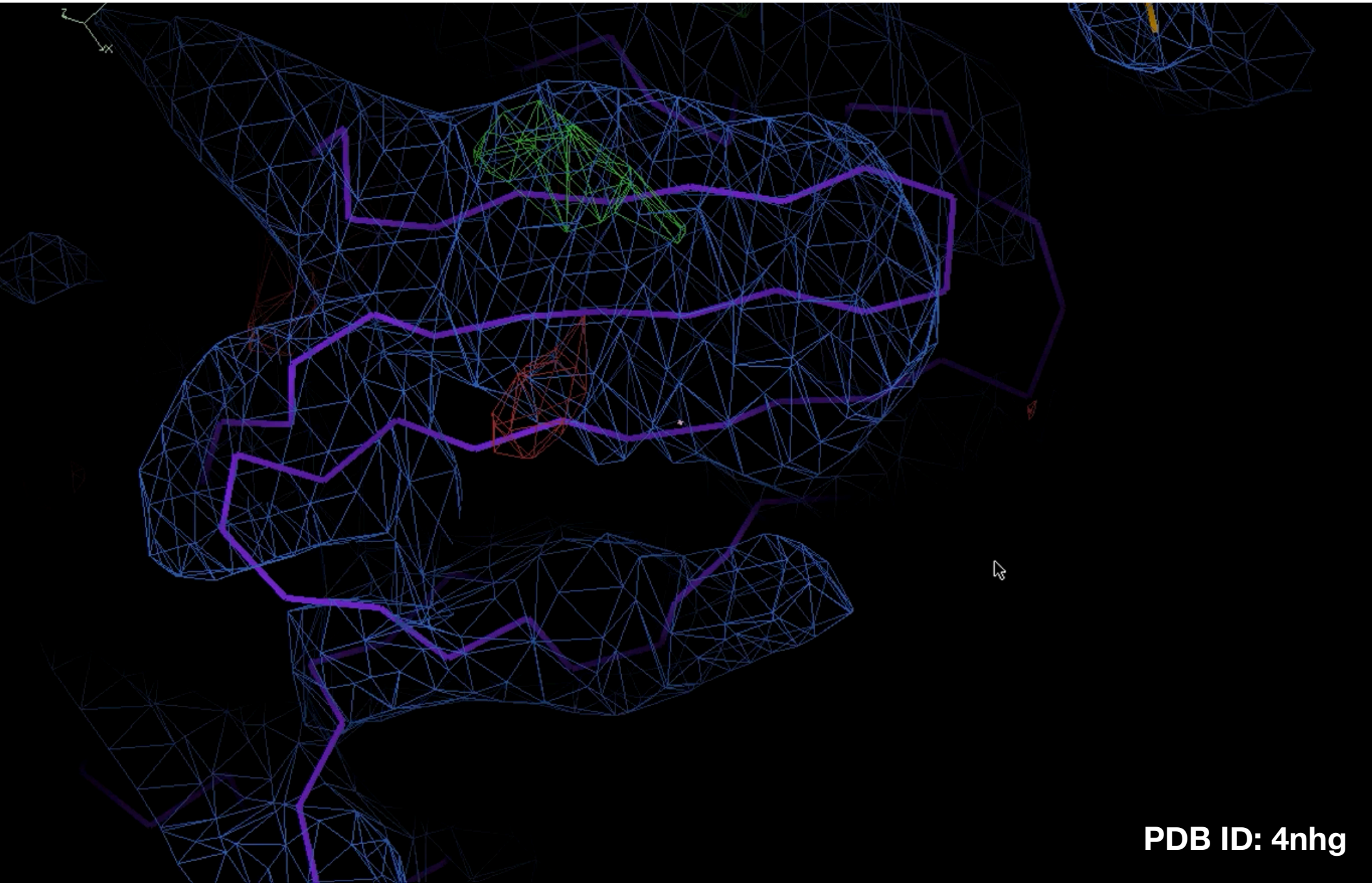
Helices at ~4.2 Å Resolution



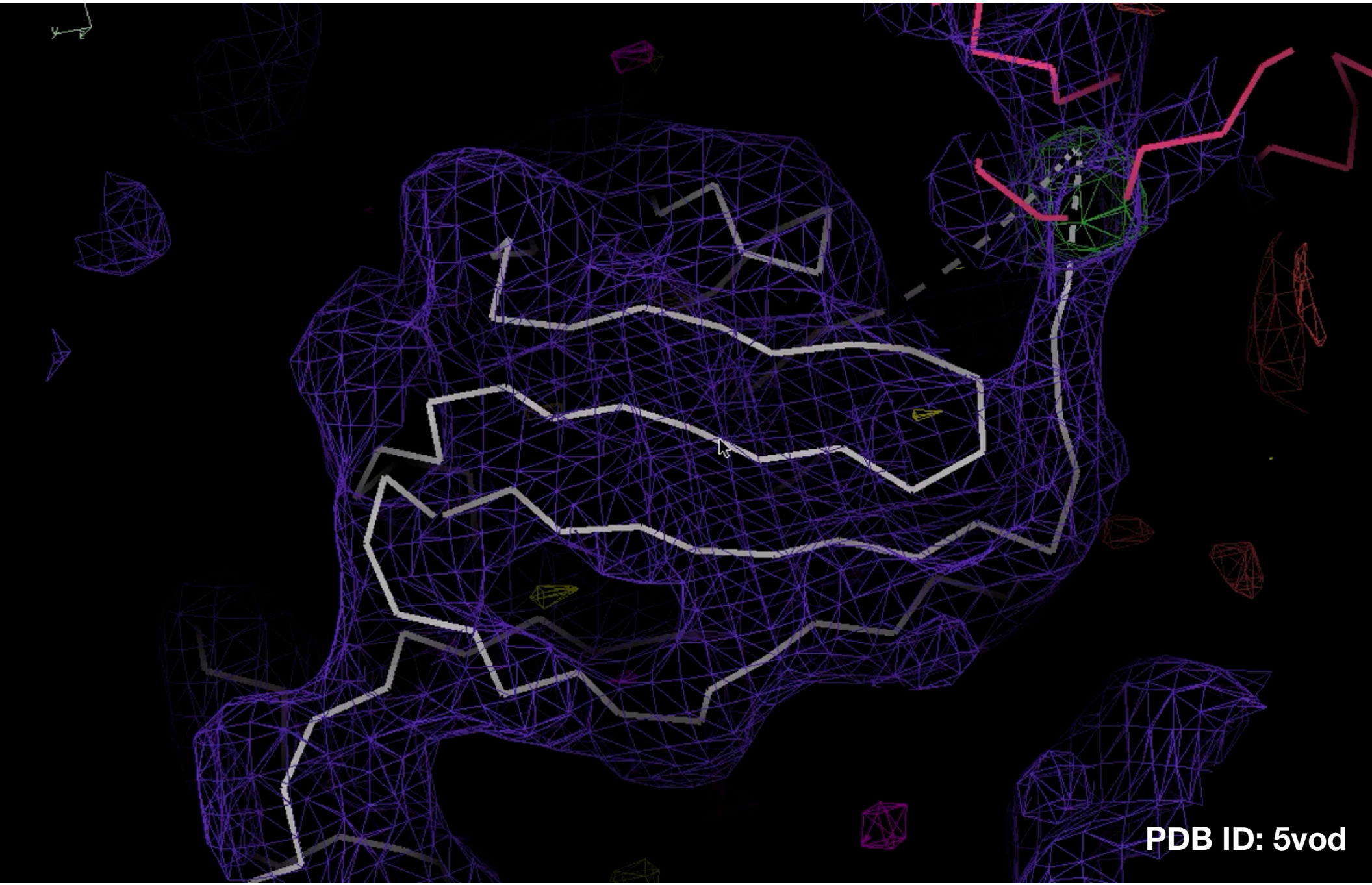
Helices at ~3 Å Resolution



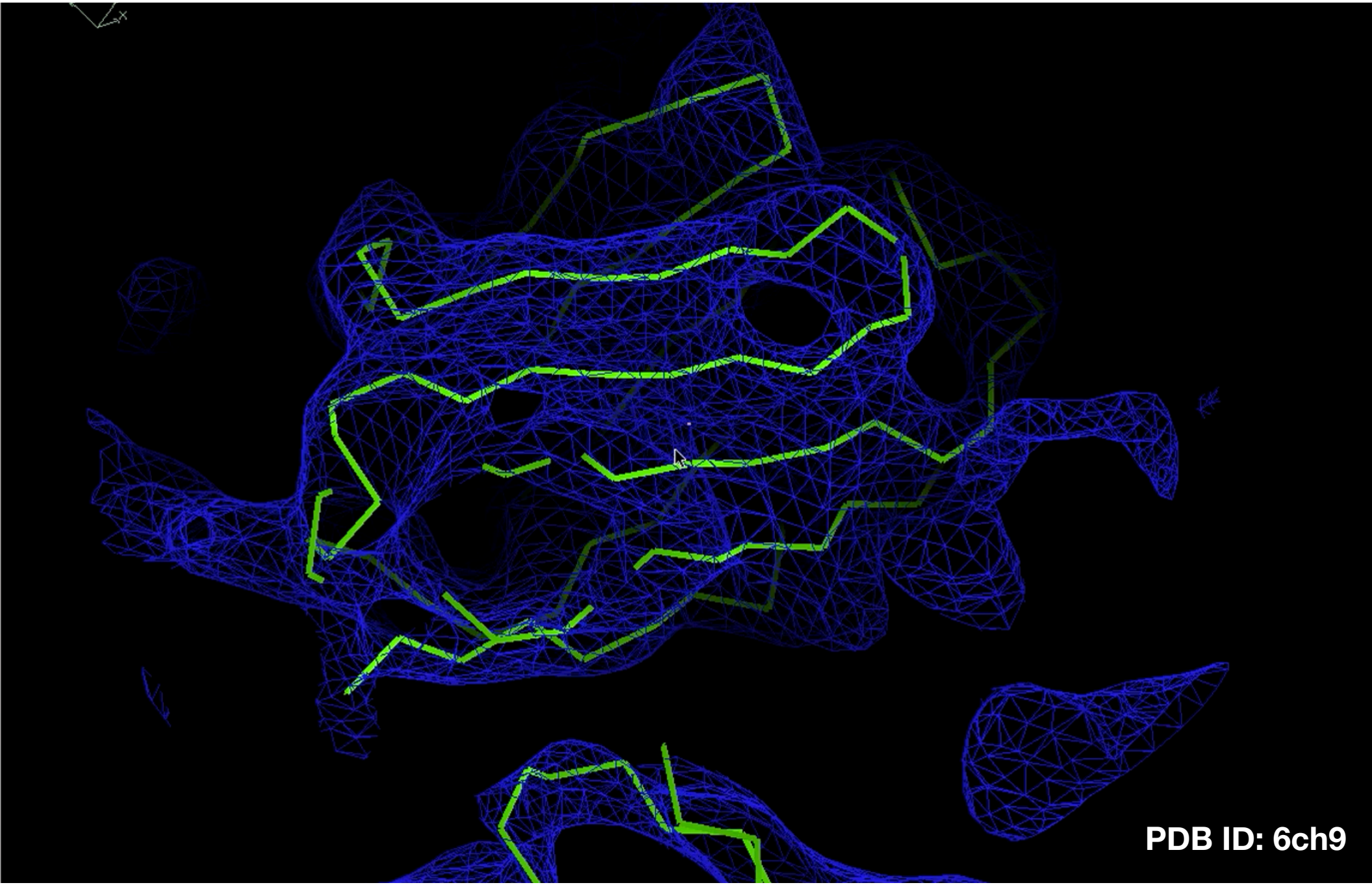
Beta strands at ~8 Å Resolution



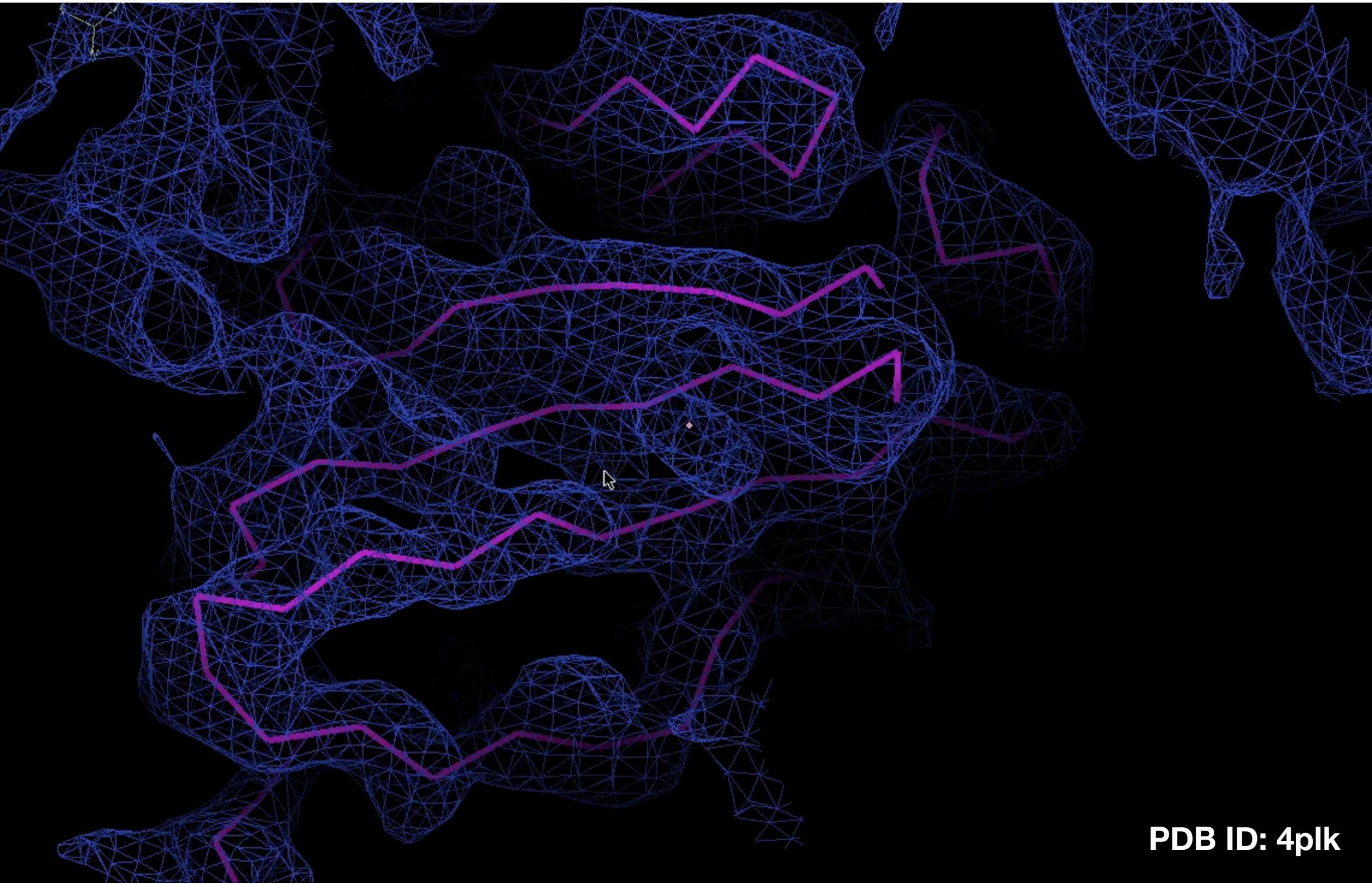
Beta strands at ~5.9 Å Resolution



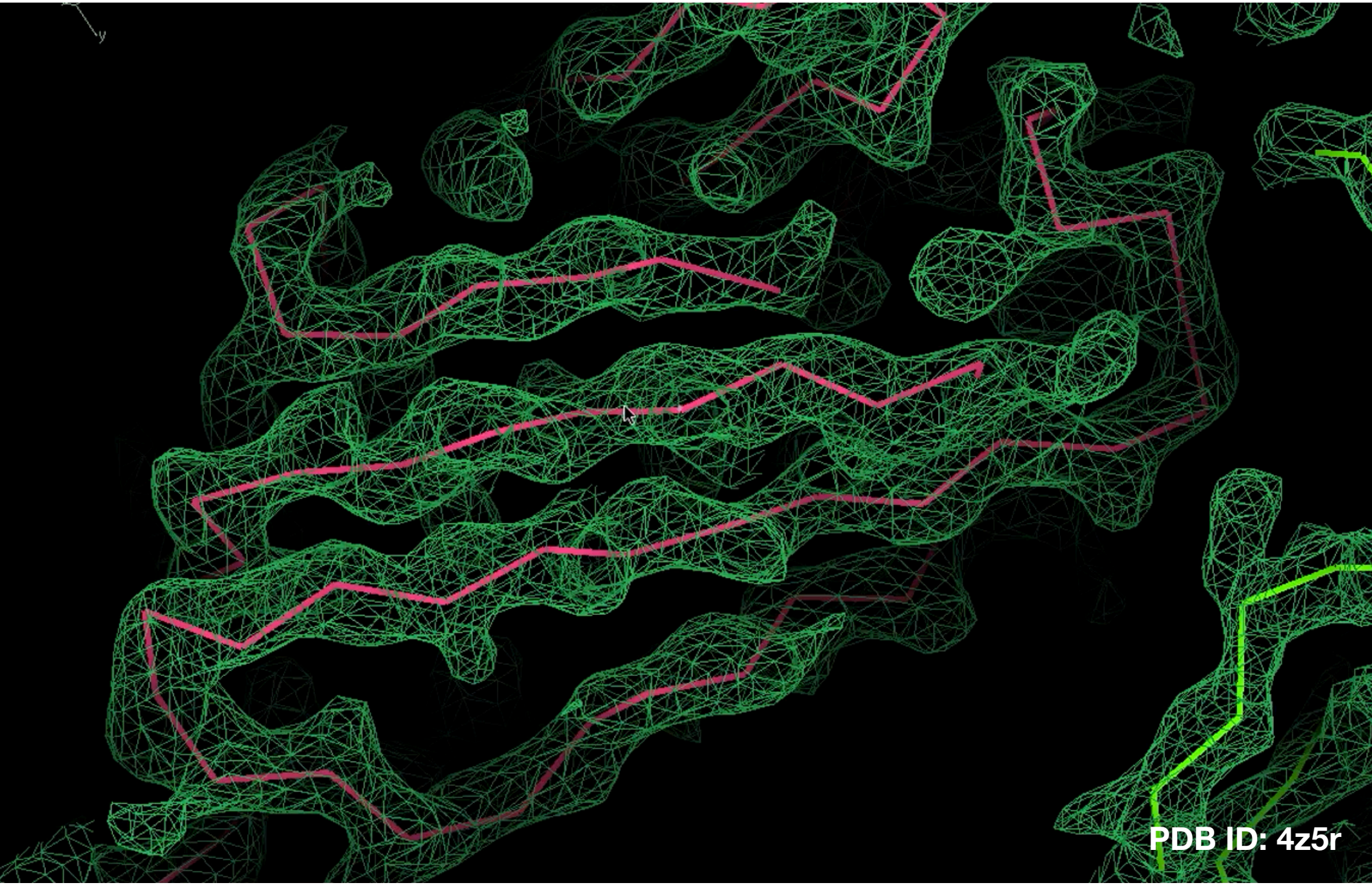
Beta strands at ~4.85 Å Resolution



Beta strands at ~4.5 Å Resolution



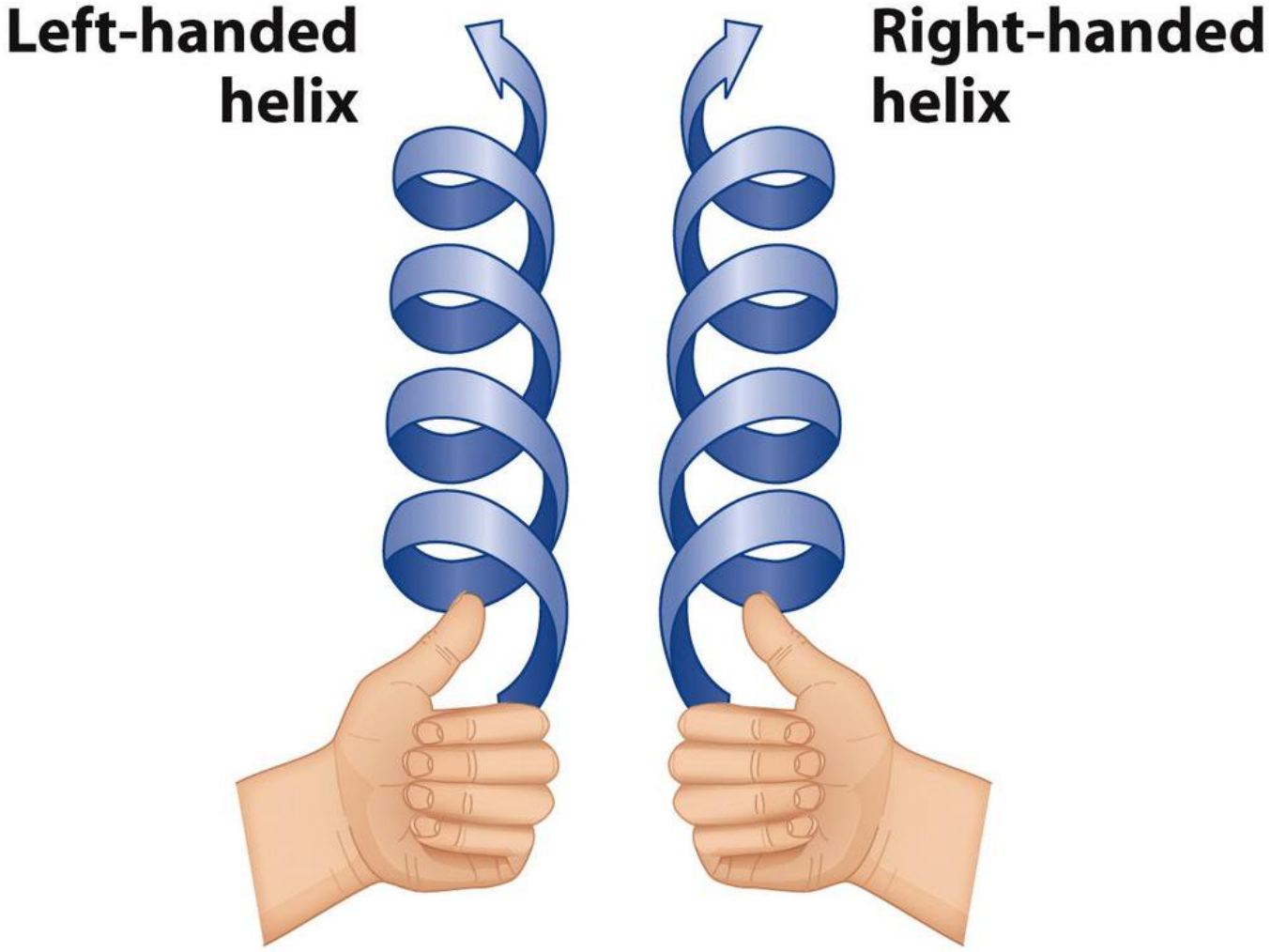
Beta strands at ~3 Å Resolution

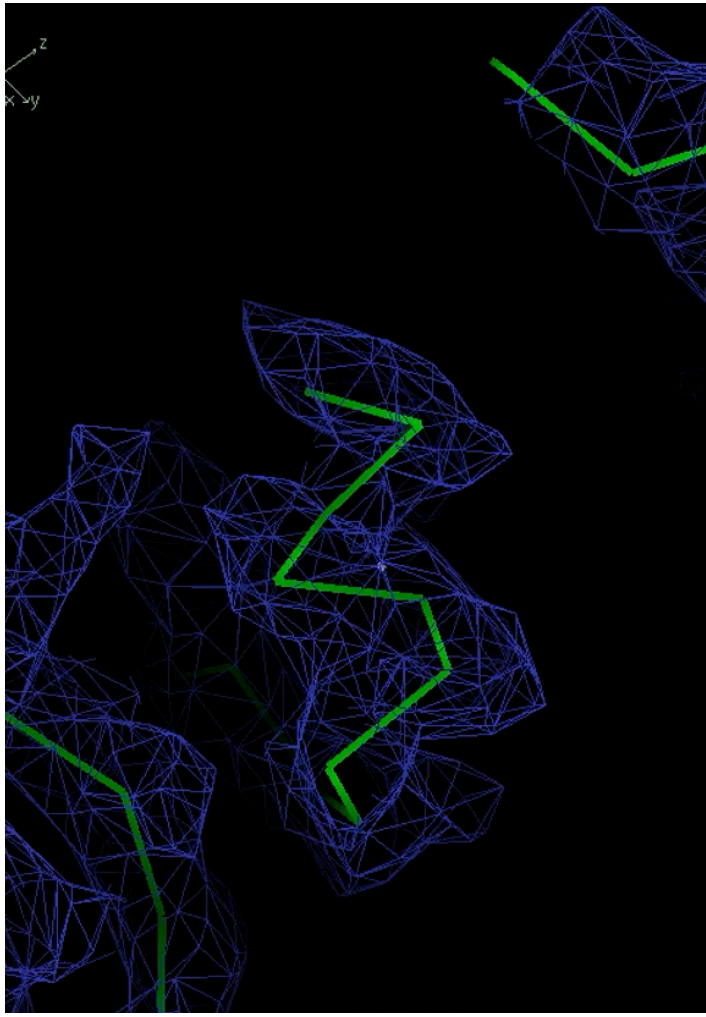


What to expect at

	~15-8 Å Resolution	~8-5 Å Resolution	~5-4 Å Resolution	~4-2 Å Resolution
Helices	Secondary structure absent	Helices apear as tubes	Handedness becoming apparent	Clearly defined handedness
Beta-strands	Secondary structure absent	Sheets are poorly defined slabs of density	Strands begin to resolve	Generally clear

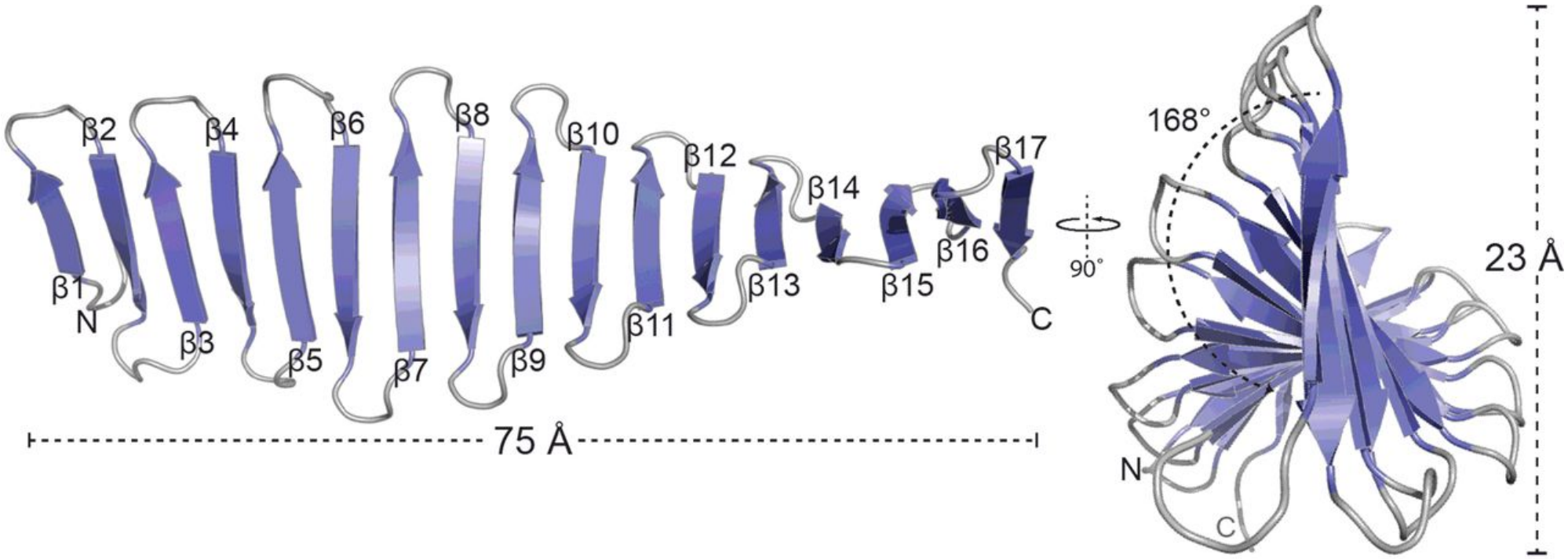
Make sure you get the “hand” of your map right!





Box 4-1
Lehninger Principles of Biochemistry, Fifth Edition

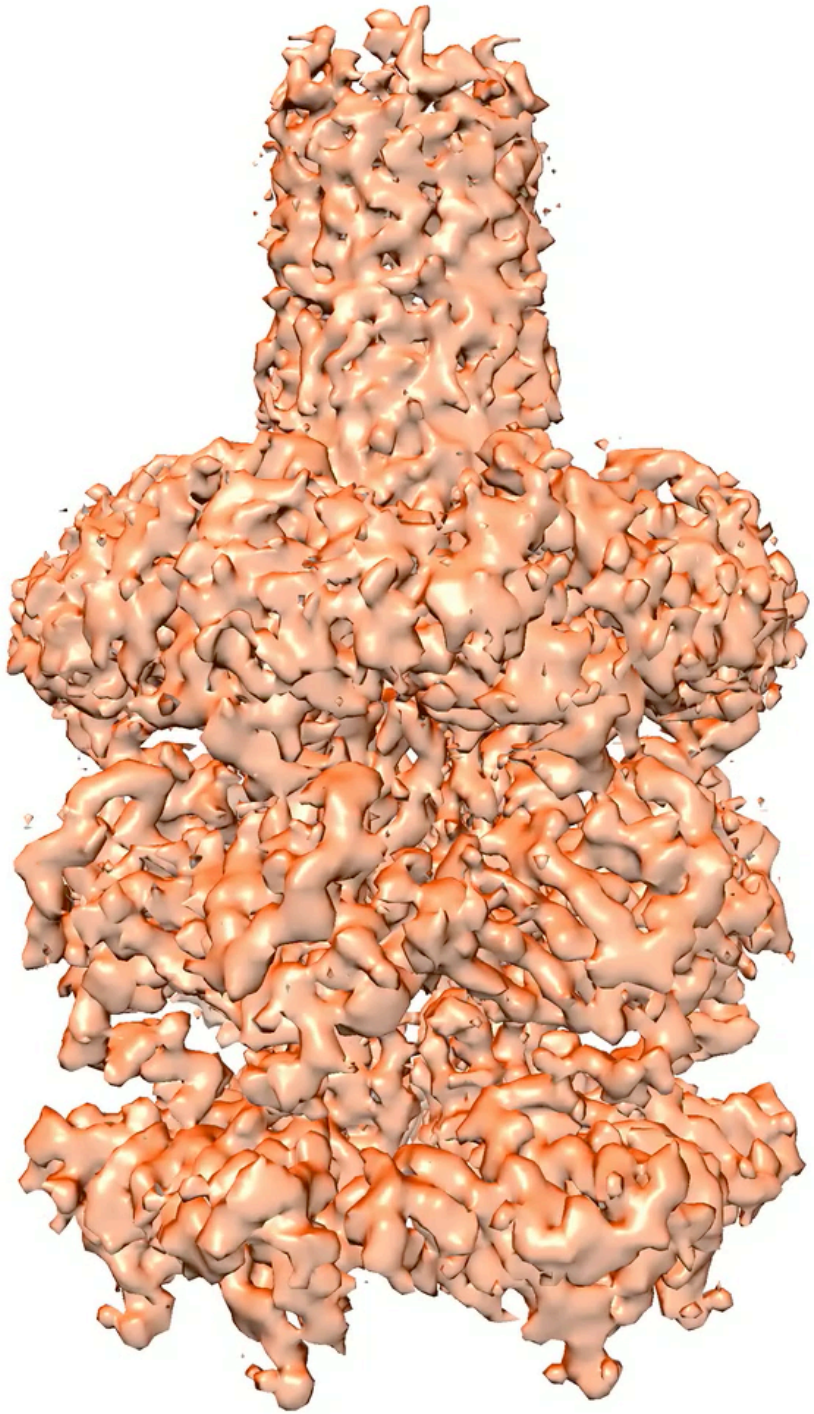
Make sure you get the “hand” of your map right!



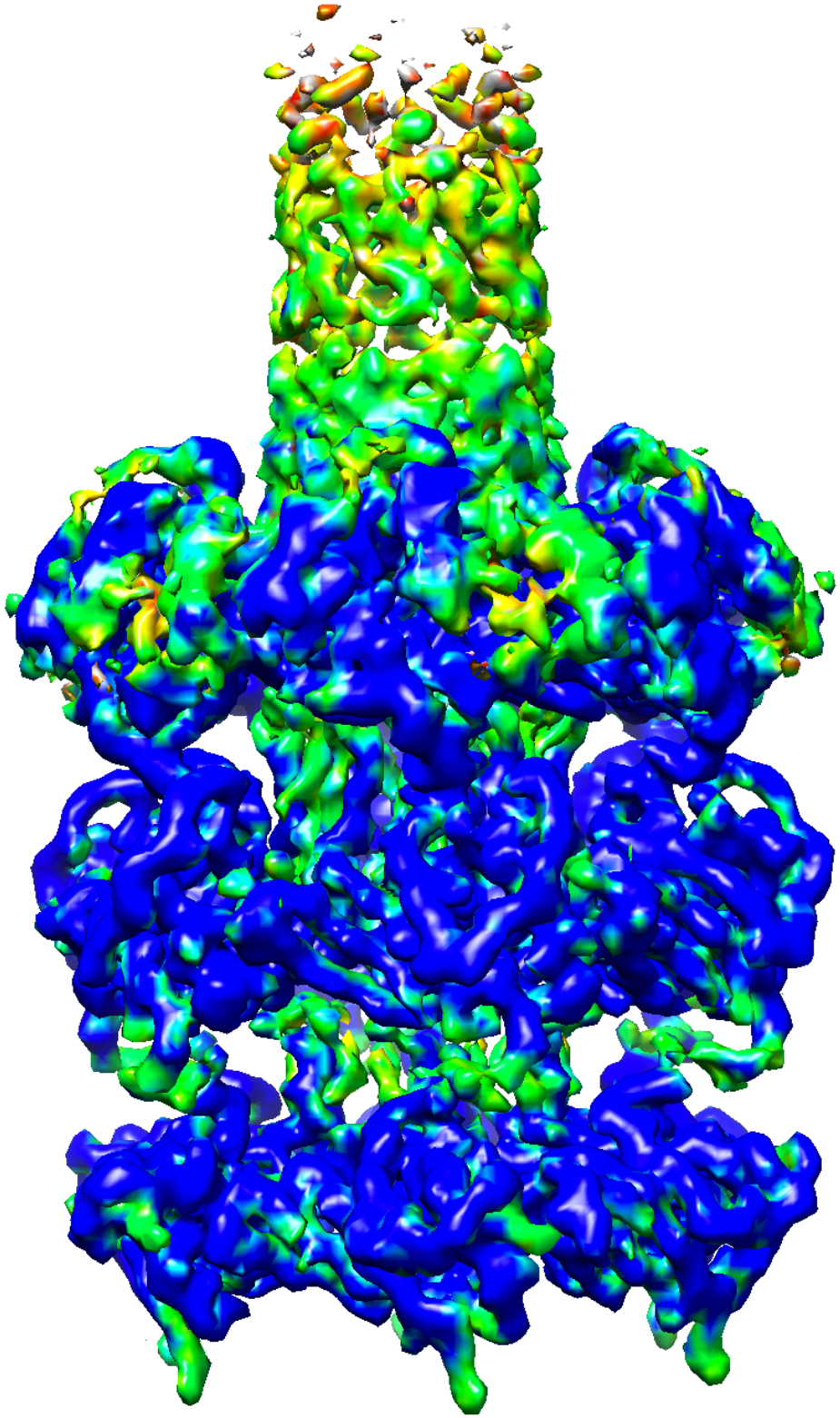
Beta <u>sheets</u> have a left-handed twist

Image credit: Cutts et al, Biochemical Society Transactions, Oct 09, 2015, 43 (5) 838-843

Example case: ~4 Å structure of PqiB

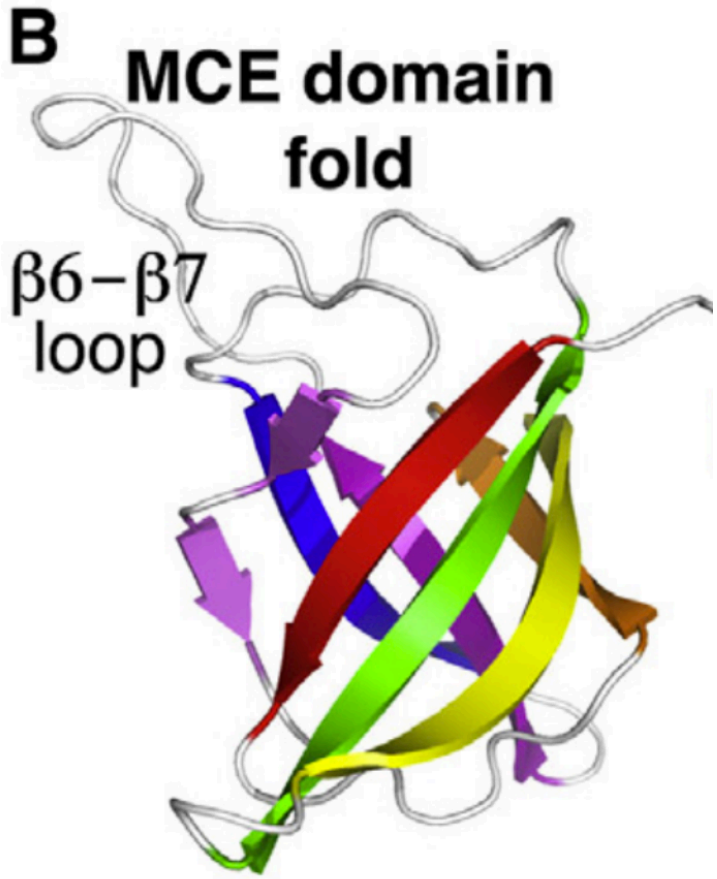


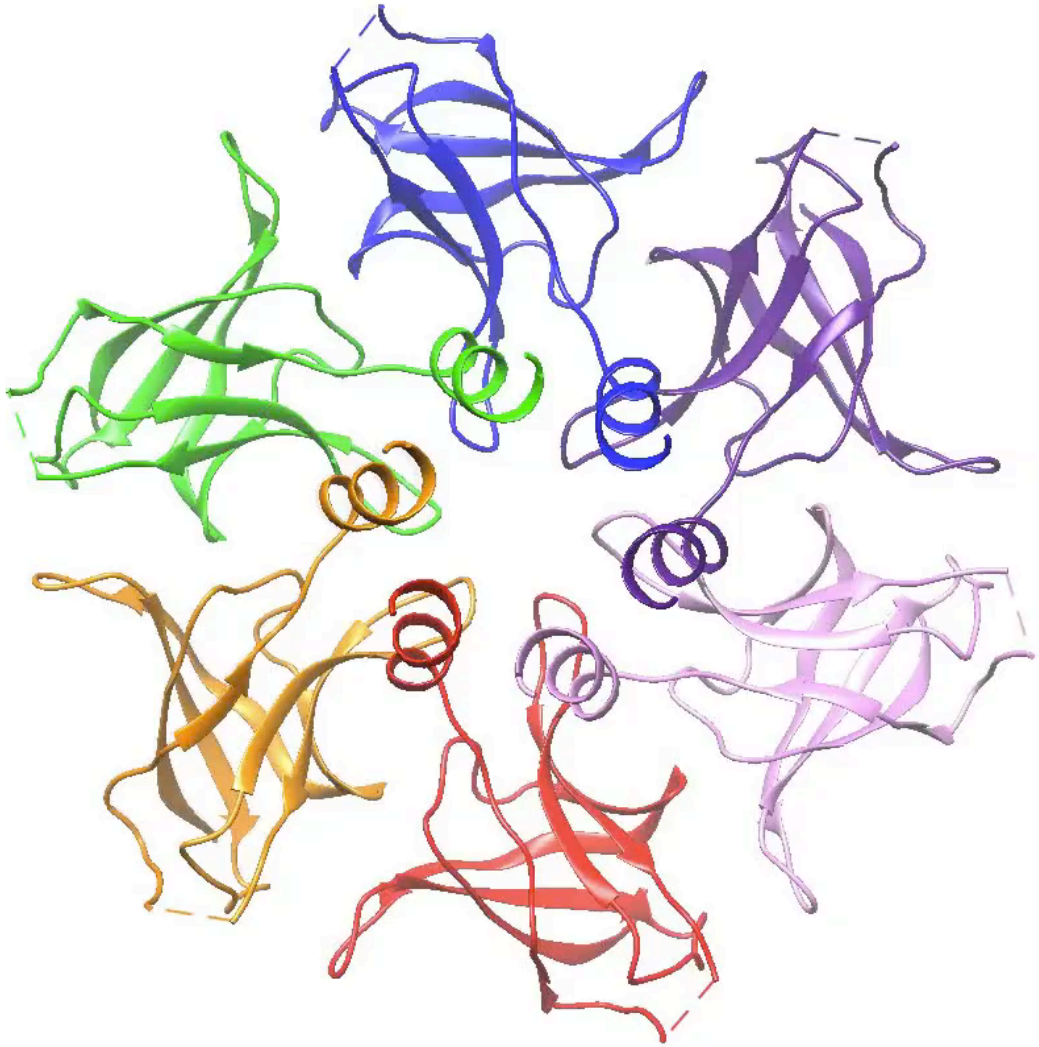
Ekiert & Bhabha et al, Cell 2017



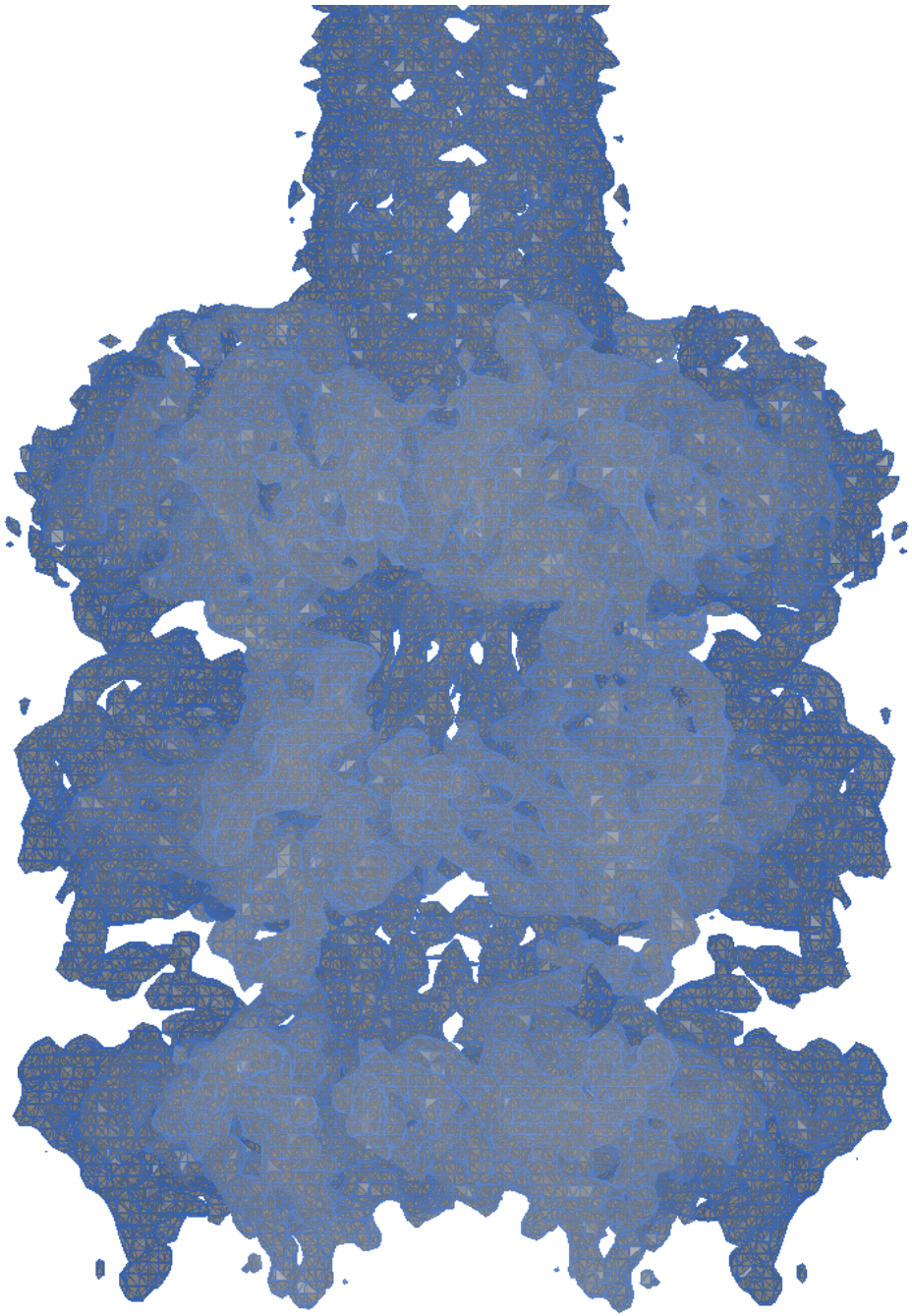
Blue = 3.1A
Green = 4.3A
Red = 6A

Prior Knowledge: Structure of domain and quaternary structure

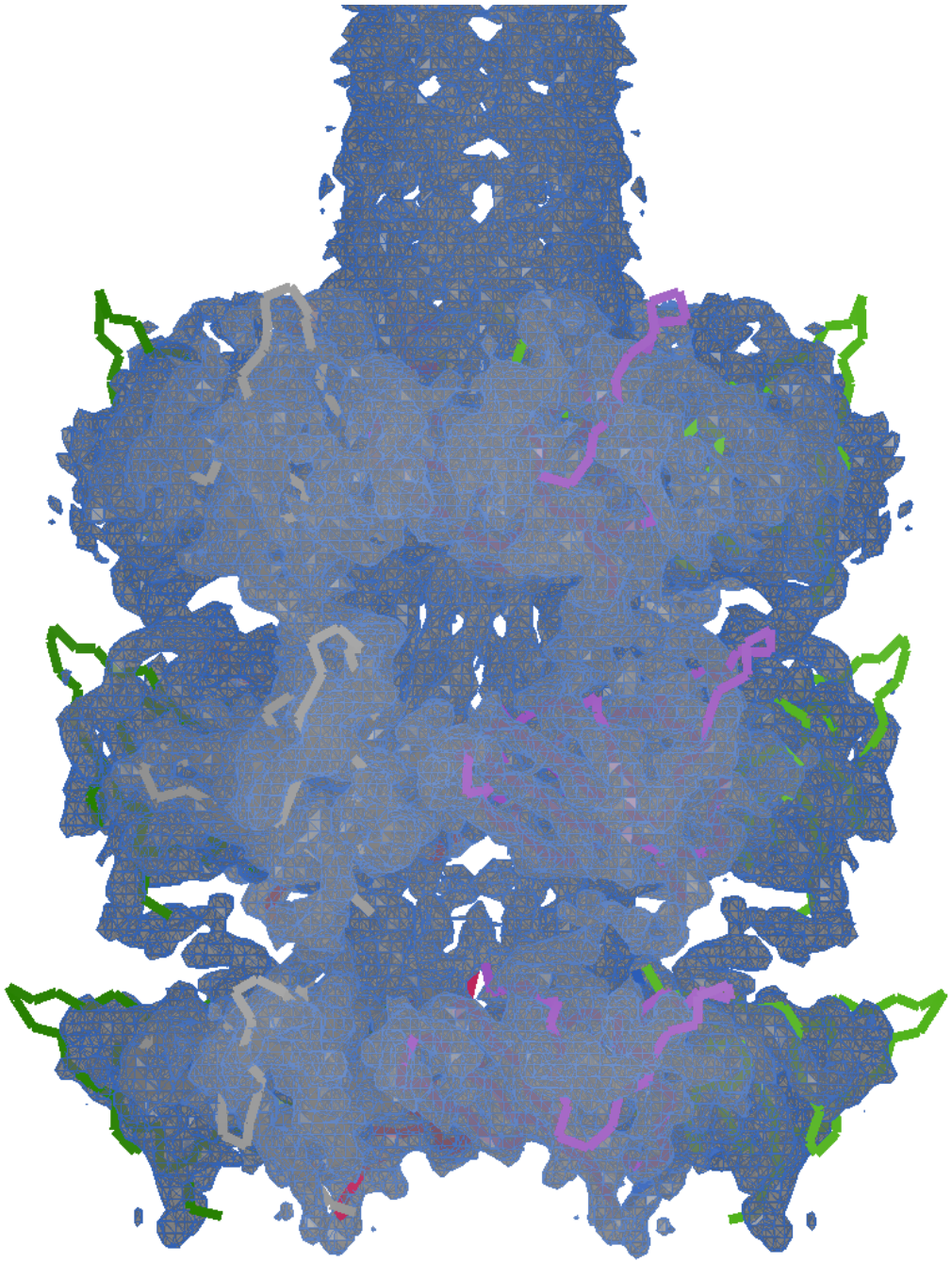


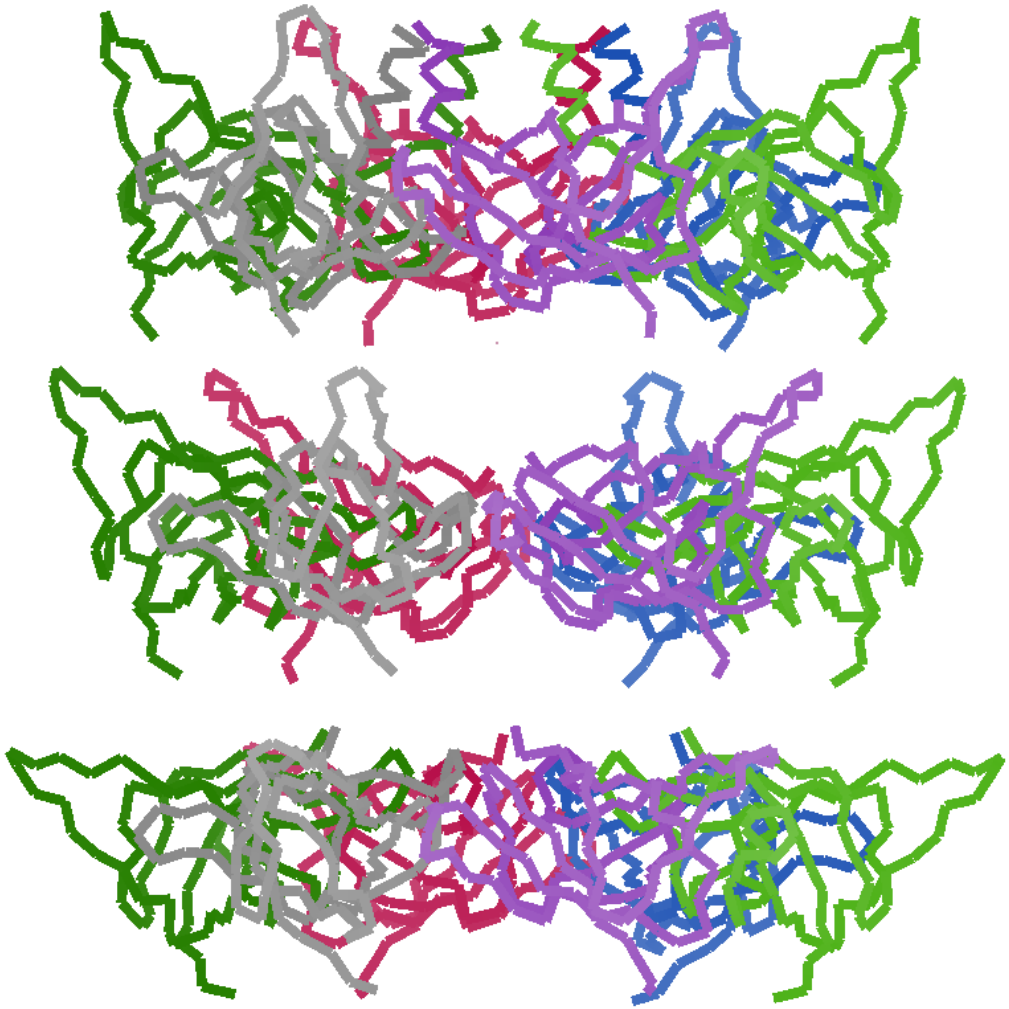


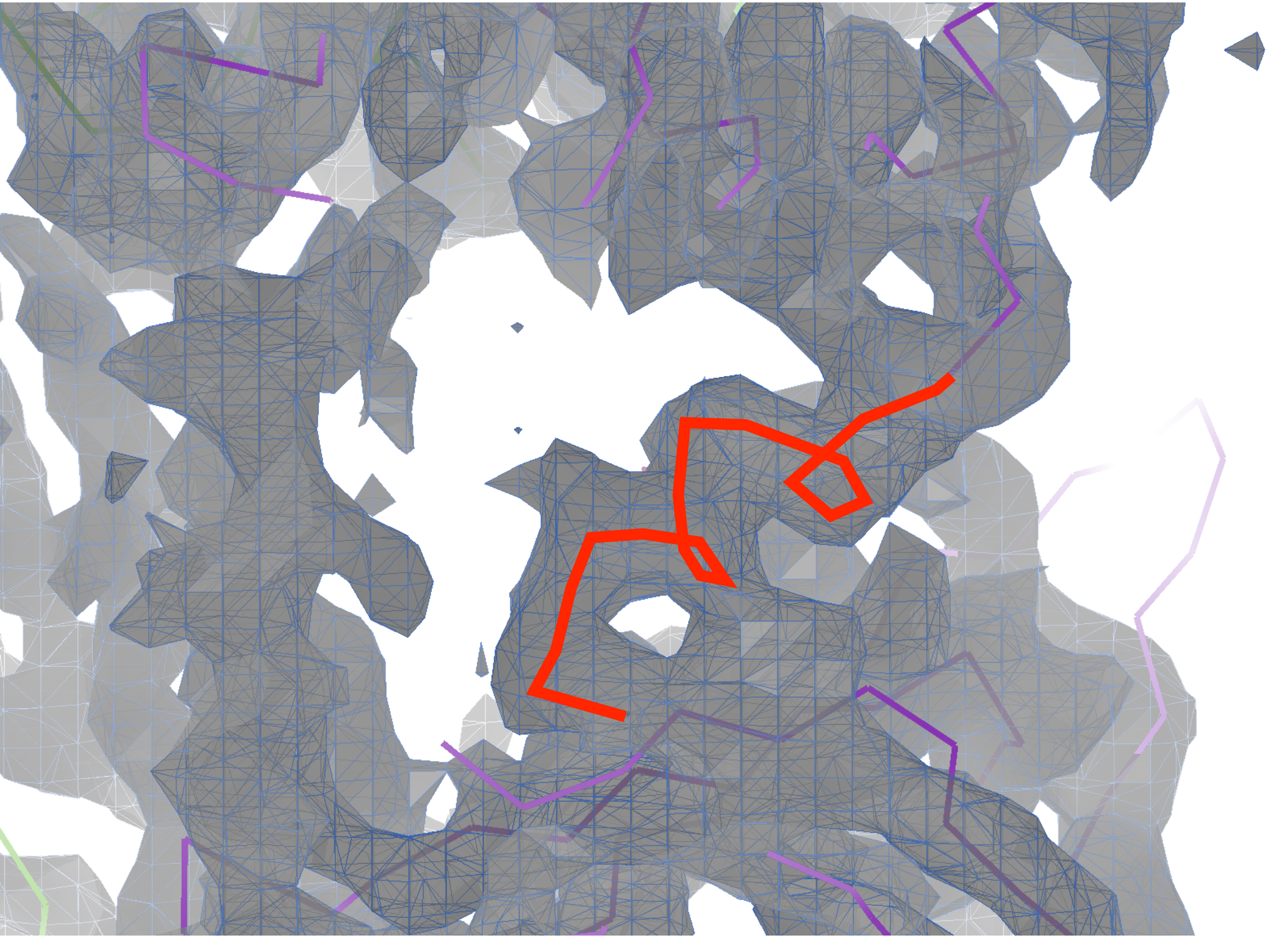
Where to begin: Domain fitting

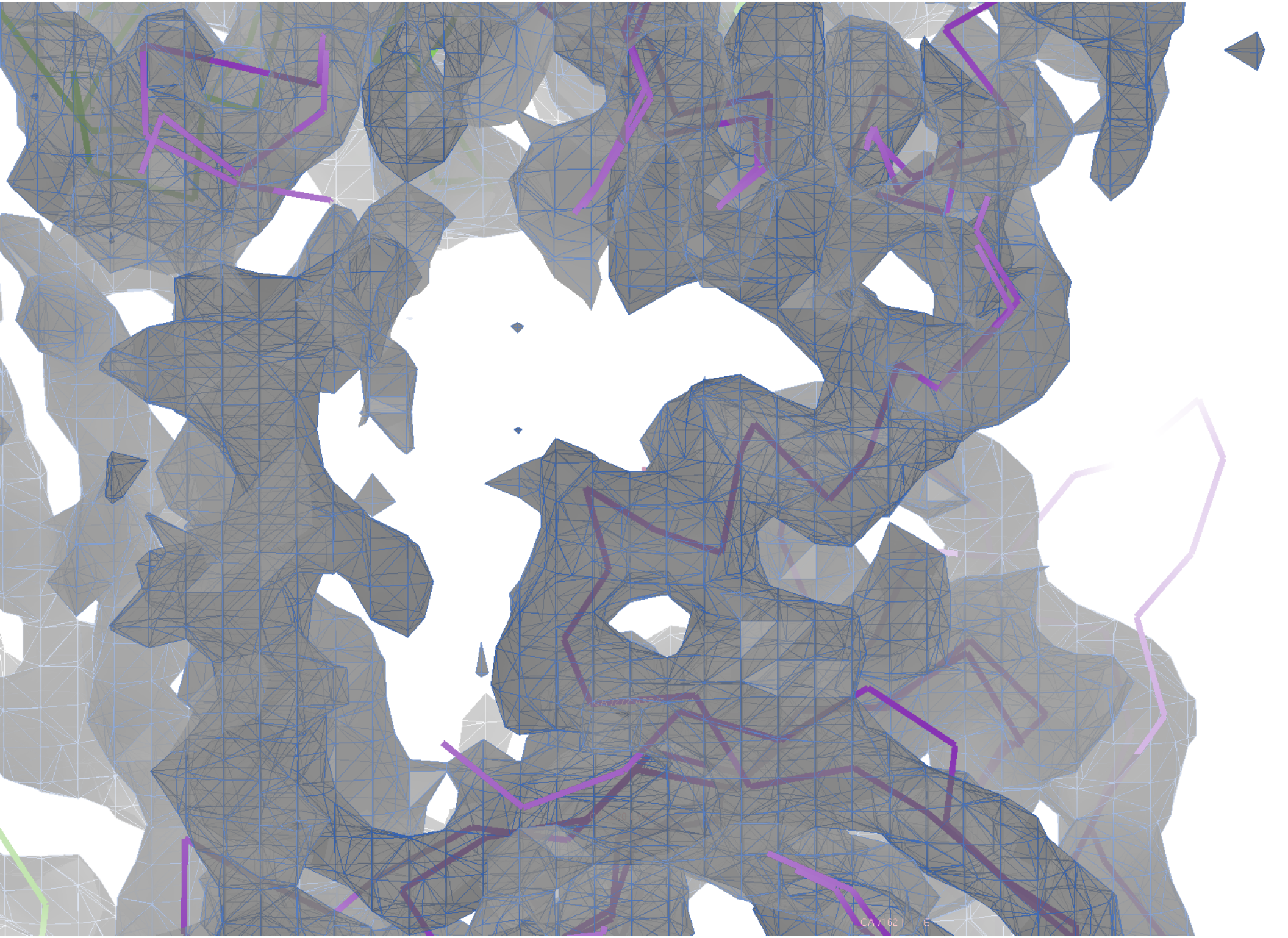


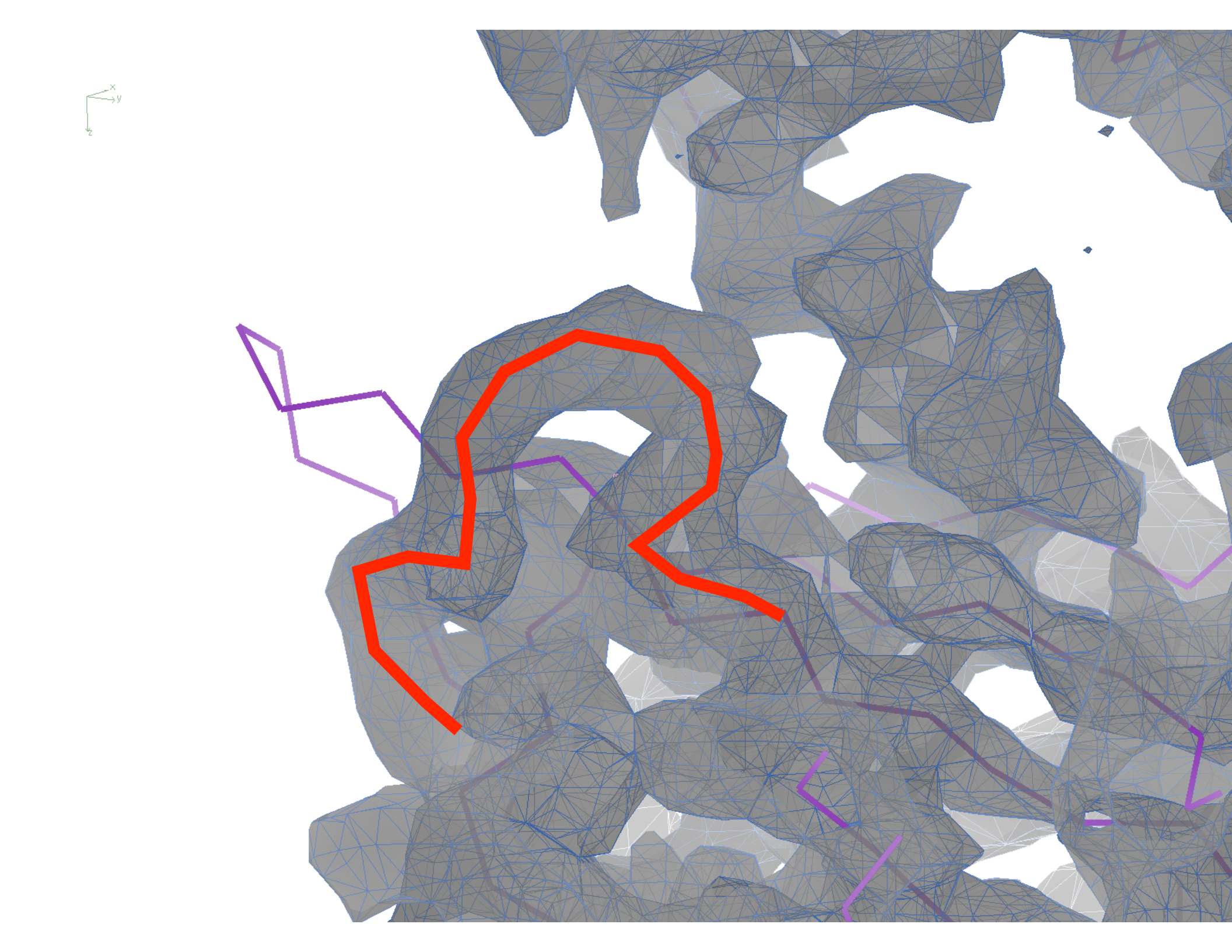
Where to begin: Fitting known domains

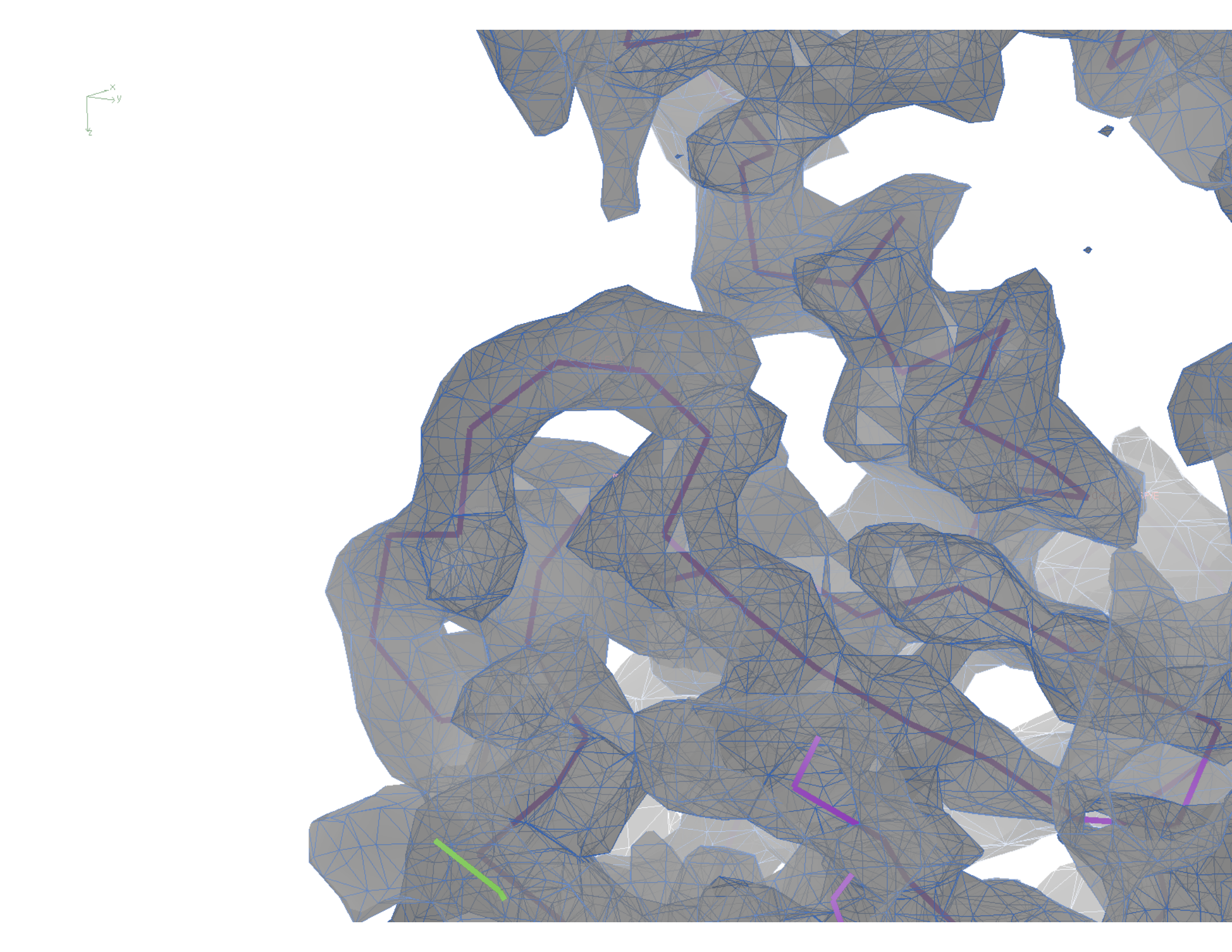




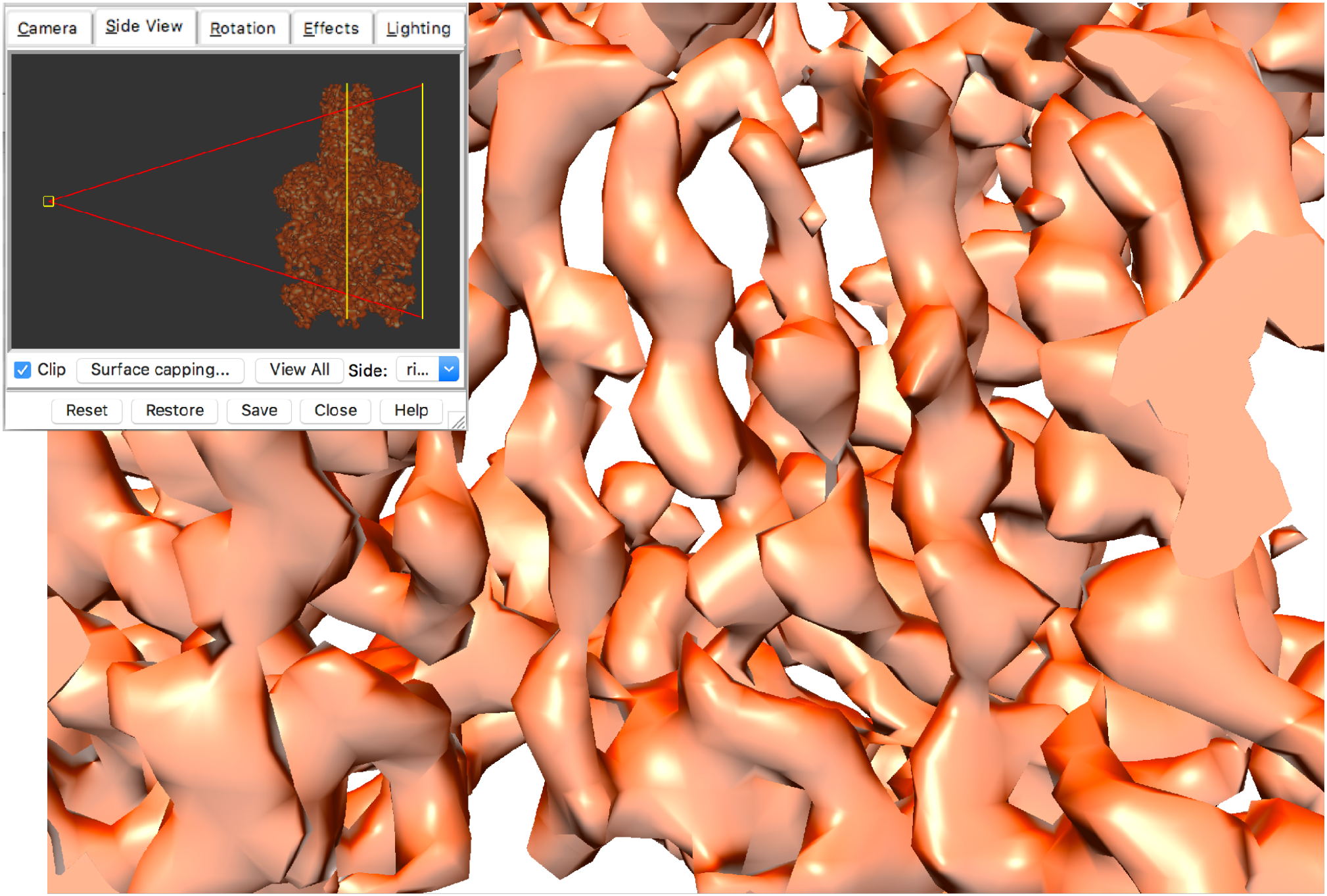


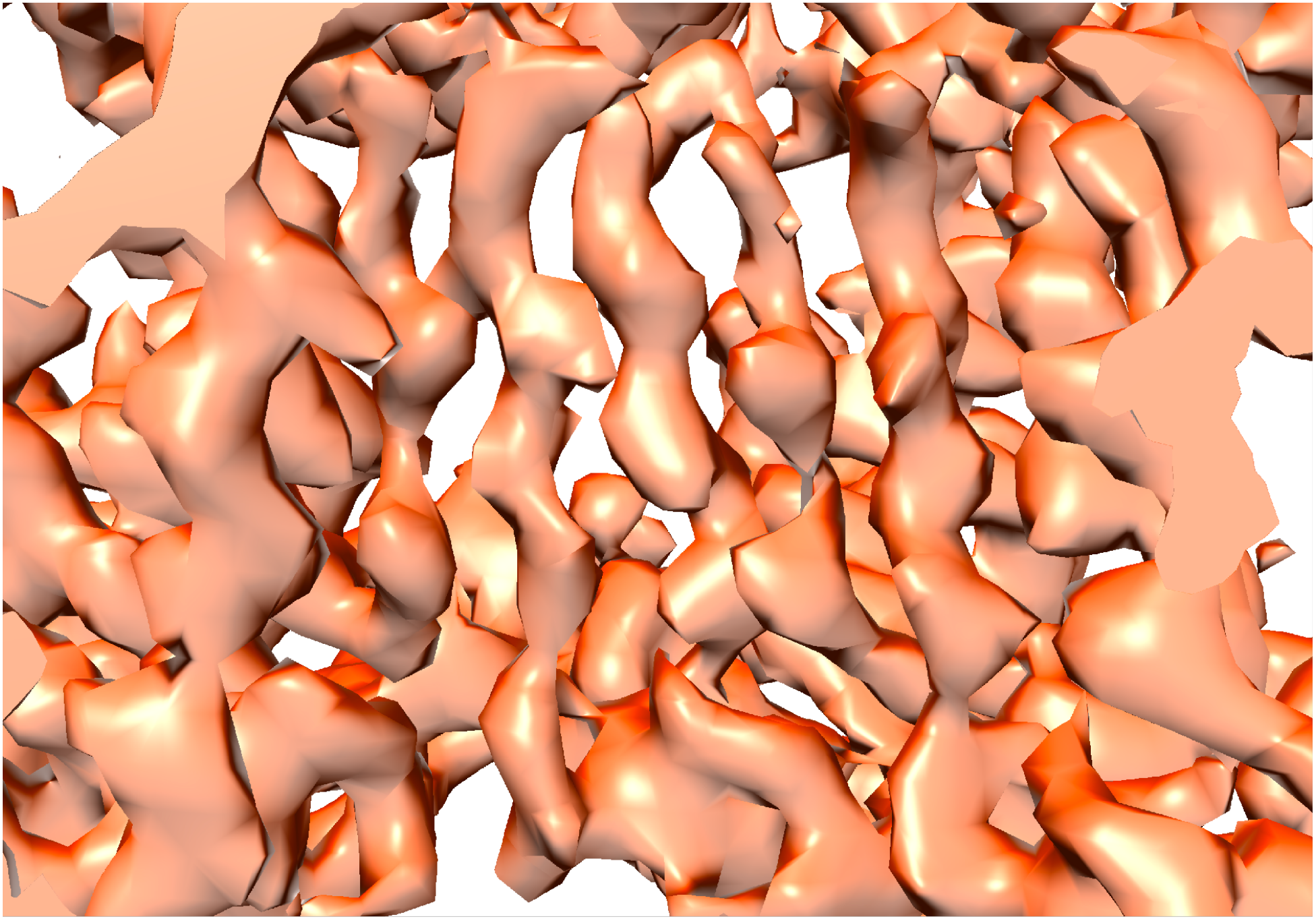


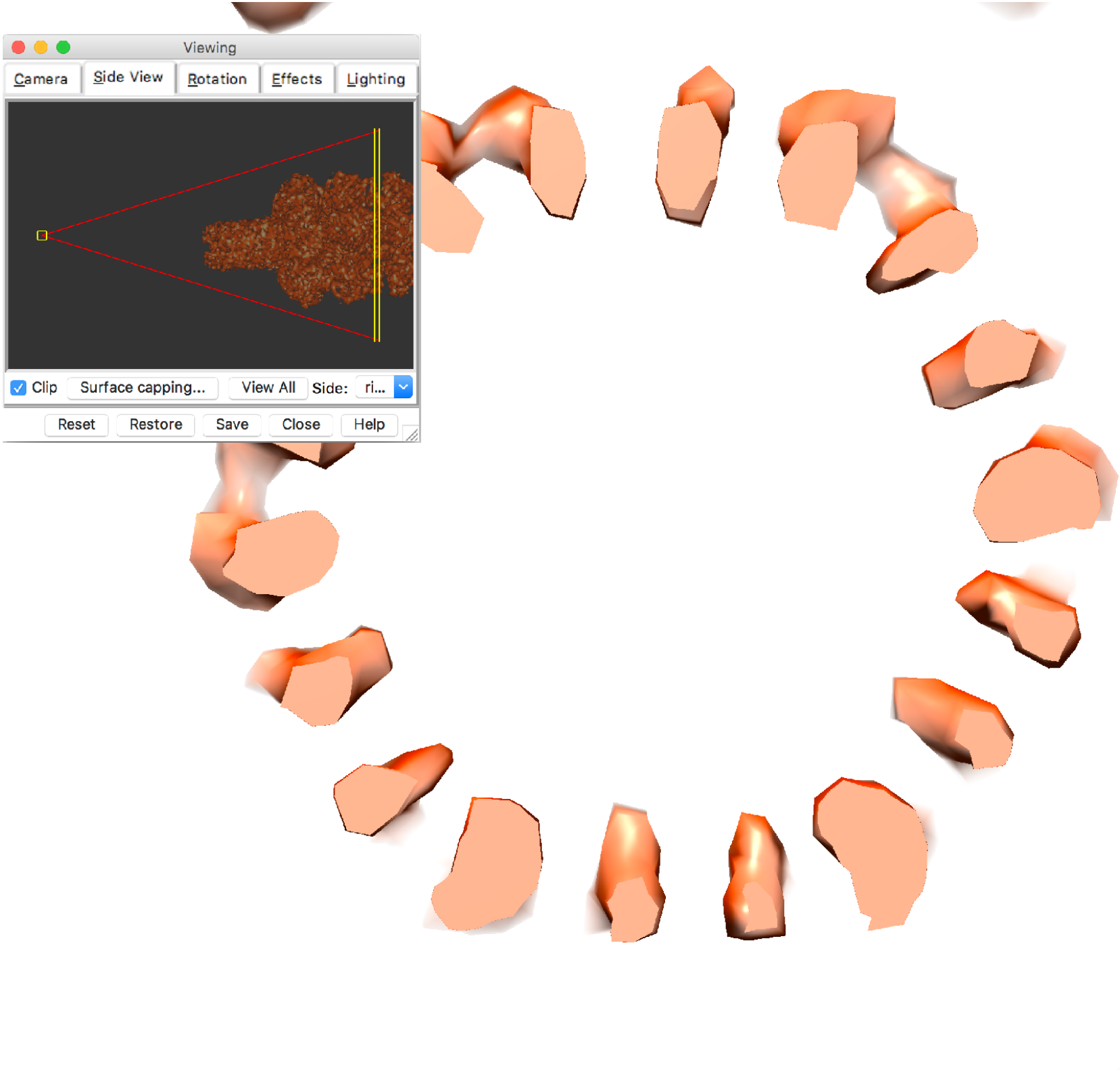

x
y
z

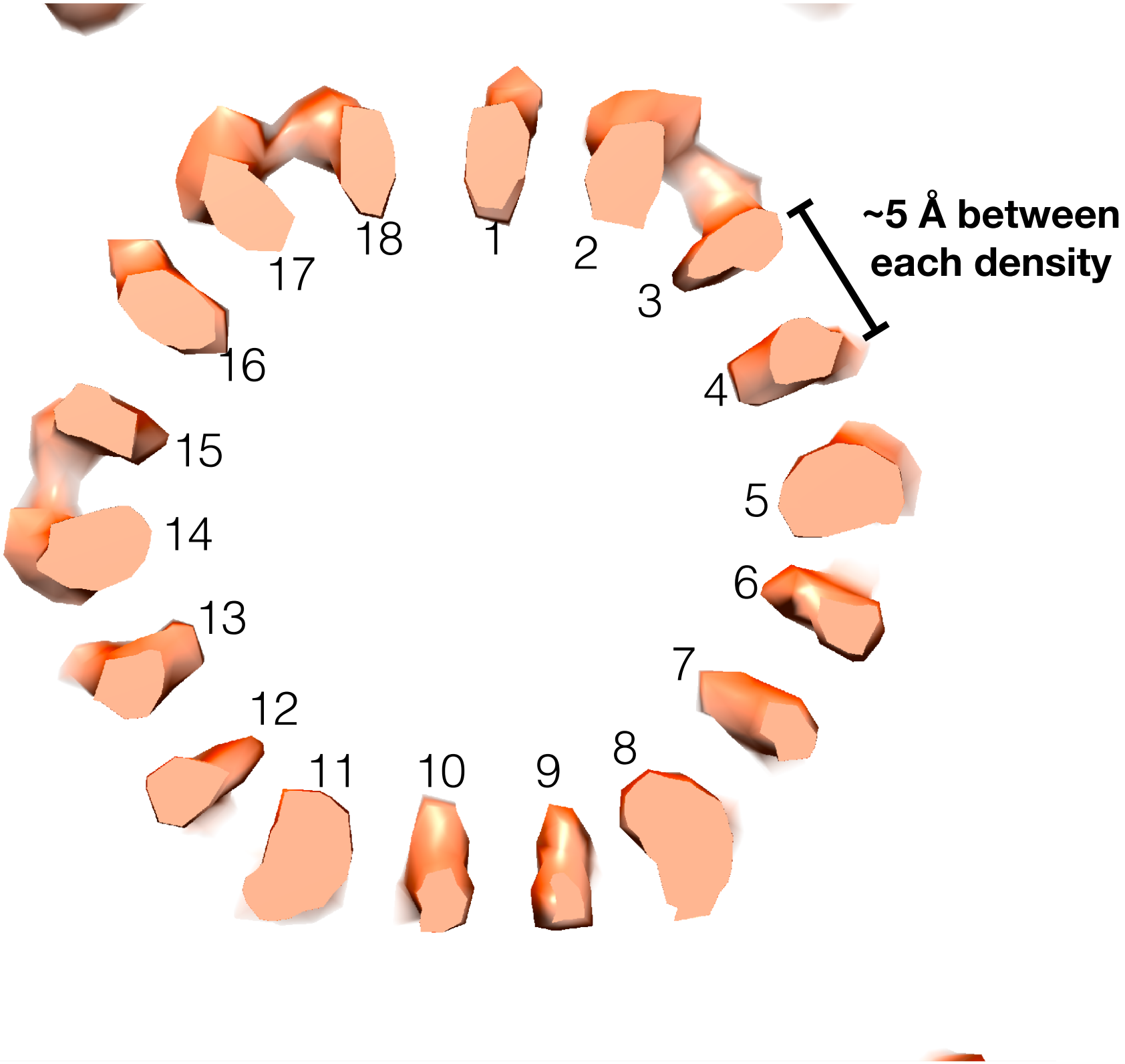

x
y
z

Mystery Densities

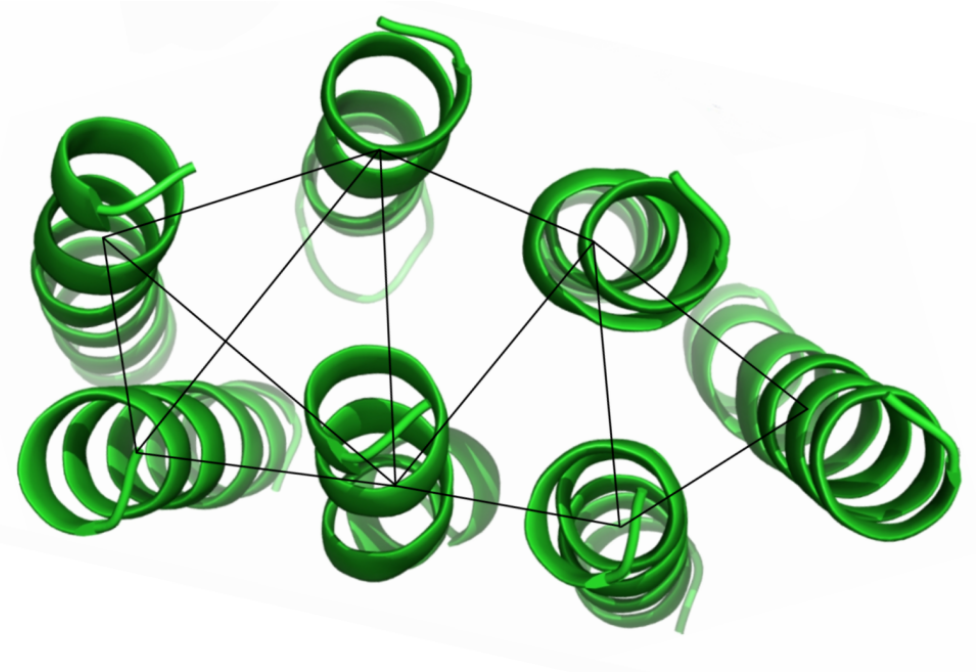




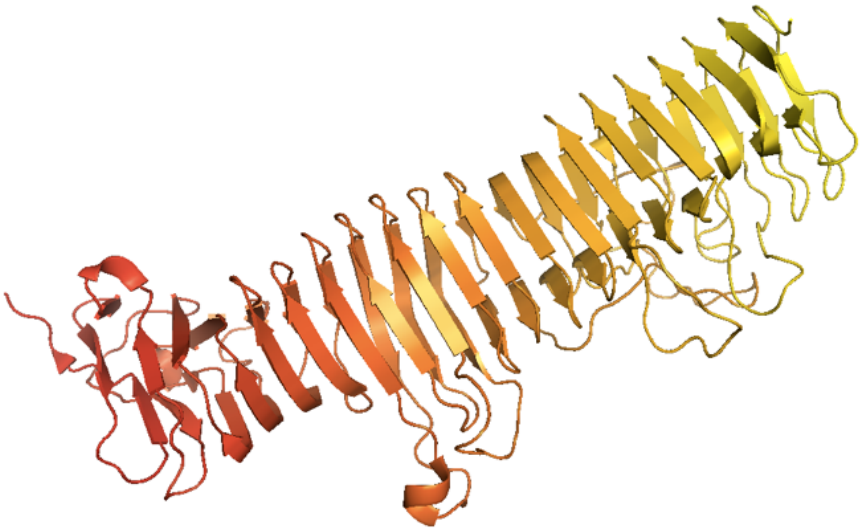

Viewing
Camera
Side View
Rotation
Effects
Lighting
Clip
Surface capping...
View All
Side:
ri...
Reset
Restore
Save
Close
Help


~5 Å between each density
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Helices or Strands?

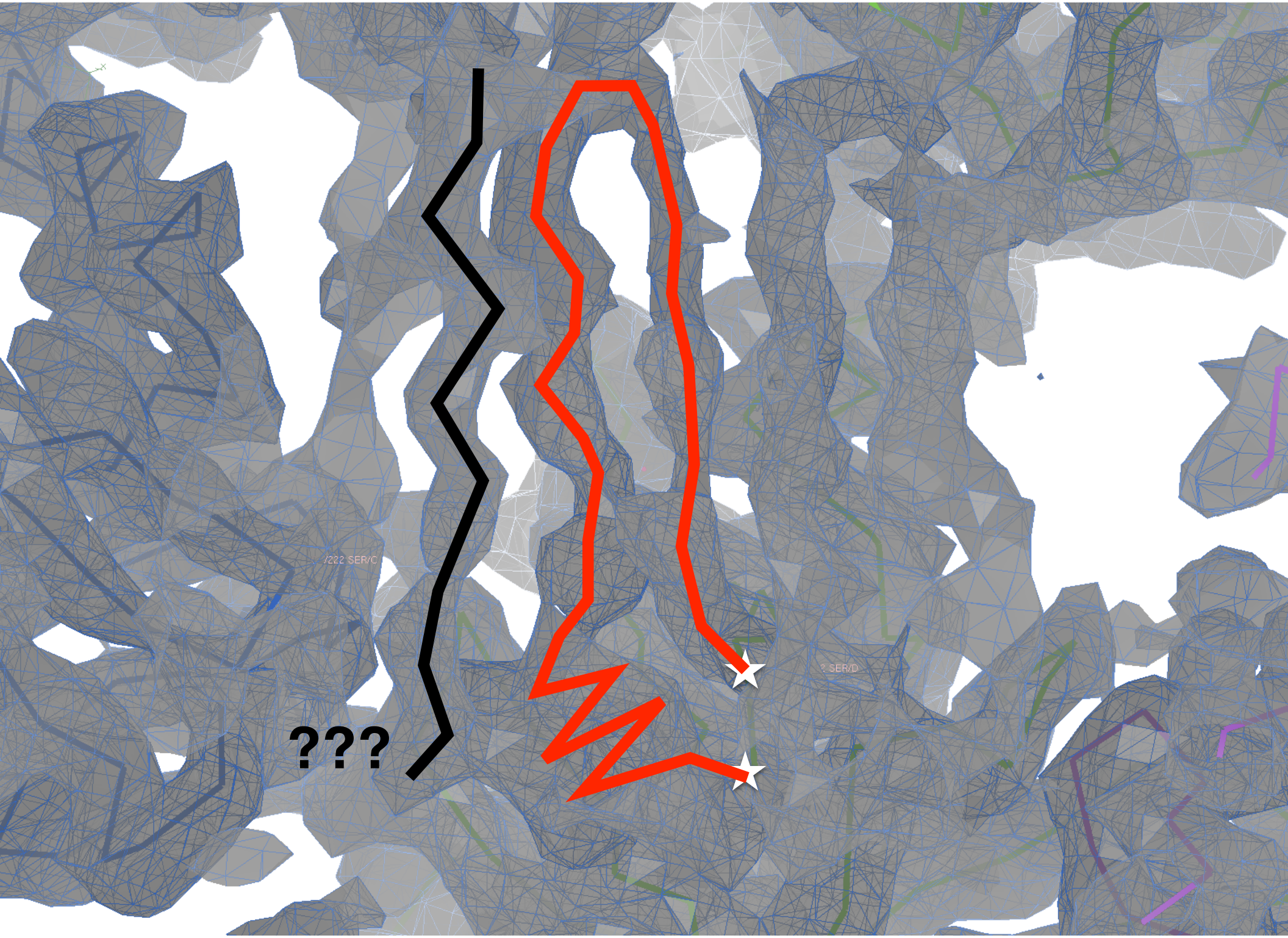


Packing distances between helices: ~7-10 Å



Spacing between beta strands: ~4.5 - 5.0 Å

https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000714
http://www.cureffi.org/2015/02/19/protein-folding-03/



Take-home messages

- Assess the quality of a given area of the map by looking at **features in the density**, not reported resolution

- Don’t over-interpret the map!!

- Be aware of alternate interpretations of the map (if a domain fits equally well in more than one orientation, and your map does not have features to distinguish between the two, then you cannot conclude the right orientation without further experiments)!

- Decide at what level your map allows reliable fitting: rigid body domains? sub-domains? Secondary structure? Amino acid residues? Rotamers? Waters?

- Use all available data:
  - Is your interpretation consistent with biochemical data? Functional data? Mass spectrometry data?