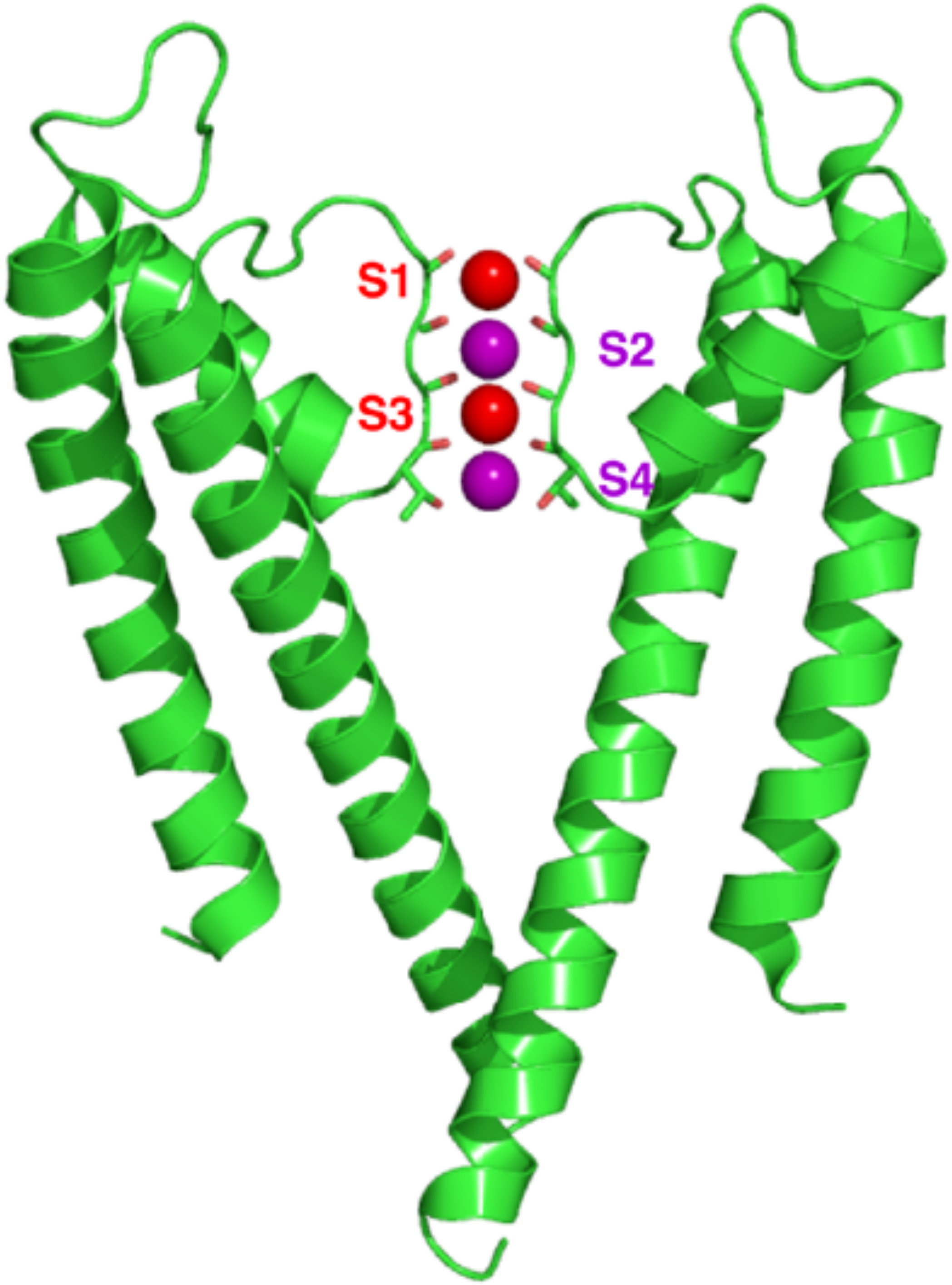
Interpretation and Limitation of Single Particle Cryo-EM analysis

Rich Hite
Memorial Sloan Kettering Cancer Center

Structural biology in 2025?

1. What is the value of an experimentally determined structure in the era of alphafold structure prediction?
2. How can we design experiments to maximize the benefits of determining cryo-EM structures?



Doyle et al. Science, Vol. 280, (1998) pp 69-77.

Advantages of cryo-EM

- Crystallization is not necessary
  - Growing well-ordered protein crystals can require years of work
  - Large protein complexes are particularly difficult to crystallize
- Immediate feedback regarding sample quality
  - Particles can be directly viewed in the images
  - Reconstructions can be calculated on the fly
- Compositional and conformational heterogeneities can be overcome by particle classification
  - Proteins rarely adopt a single conformation in physiological conditions
  - Classification can resolve protein dynamics

Molecular electron microscopy

- First EM structure -bacteriophage T4 tail - De Rosier and Klug (1968)
  - Applied helical averaging techniques to resolve the structure at ~35 Å
  - Described the general principles for Fourier synthesis of EM images

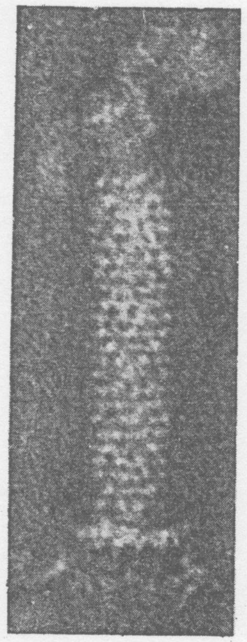


Fig. 1

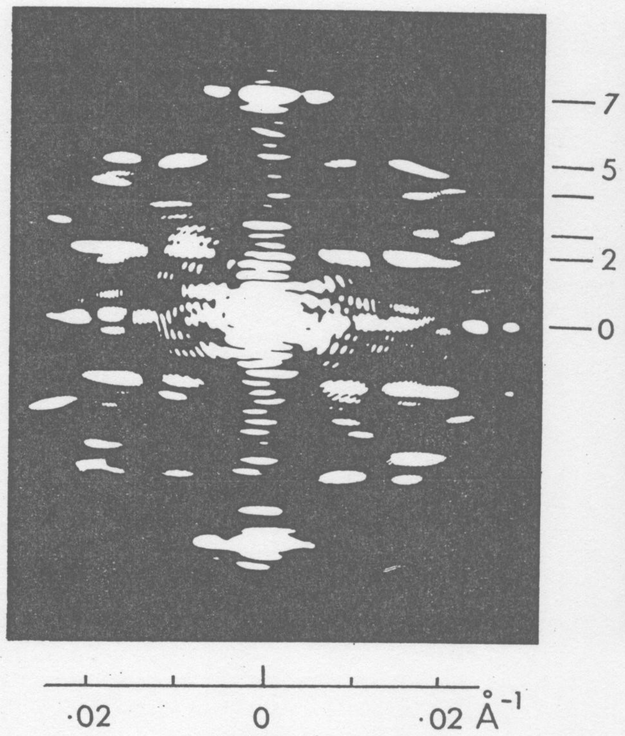


Fig. 2

Fig. 3

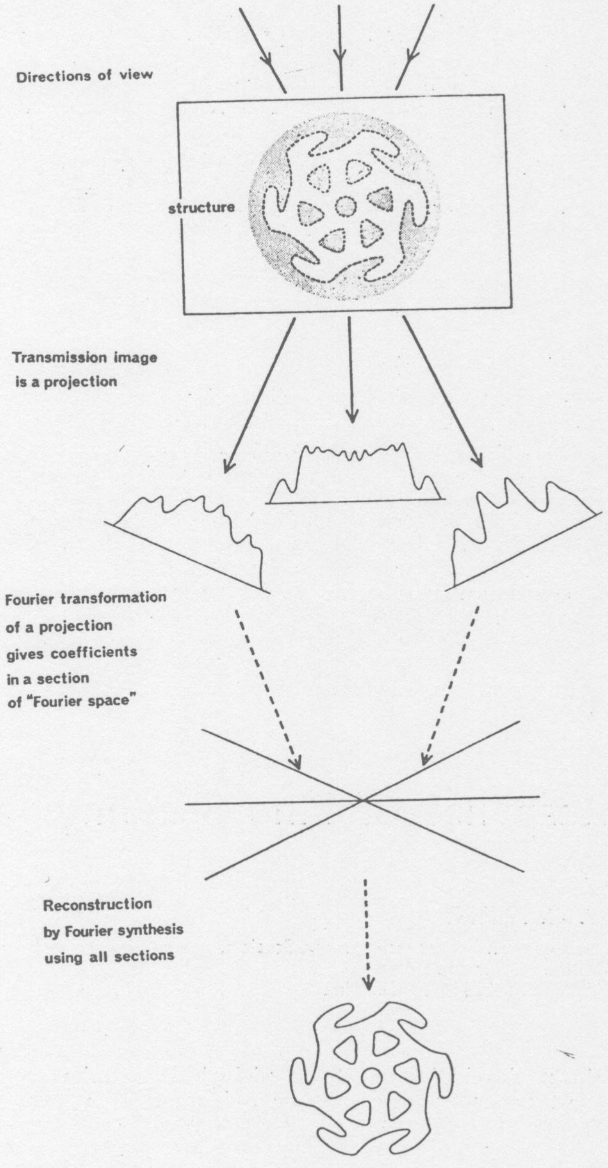
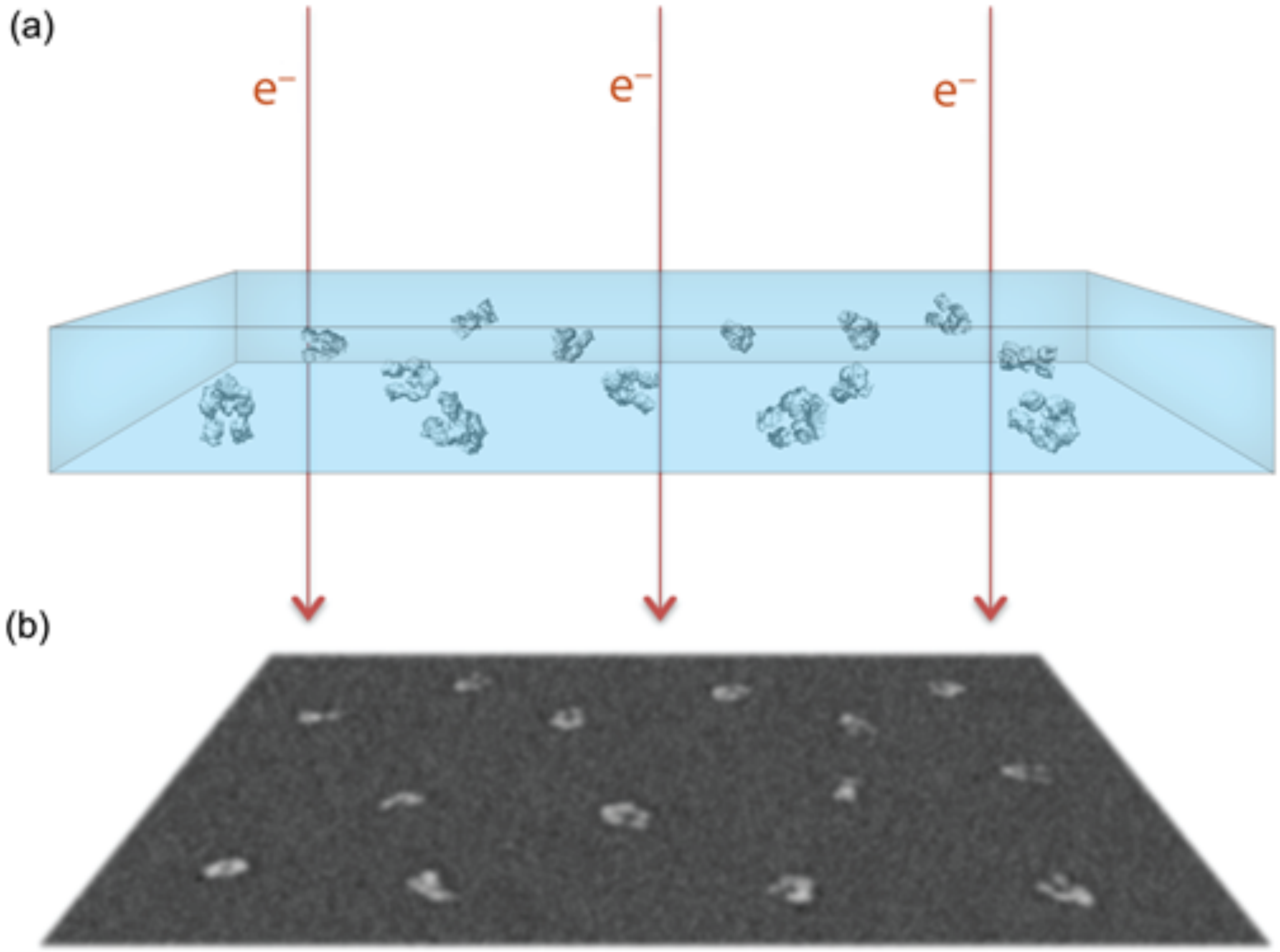


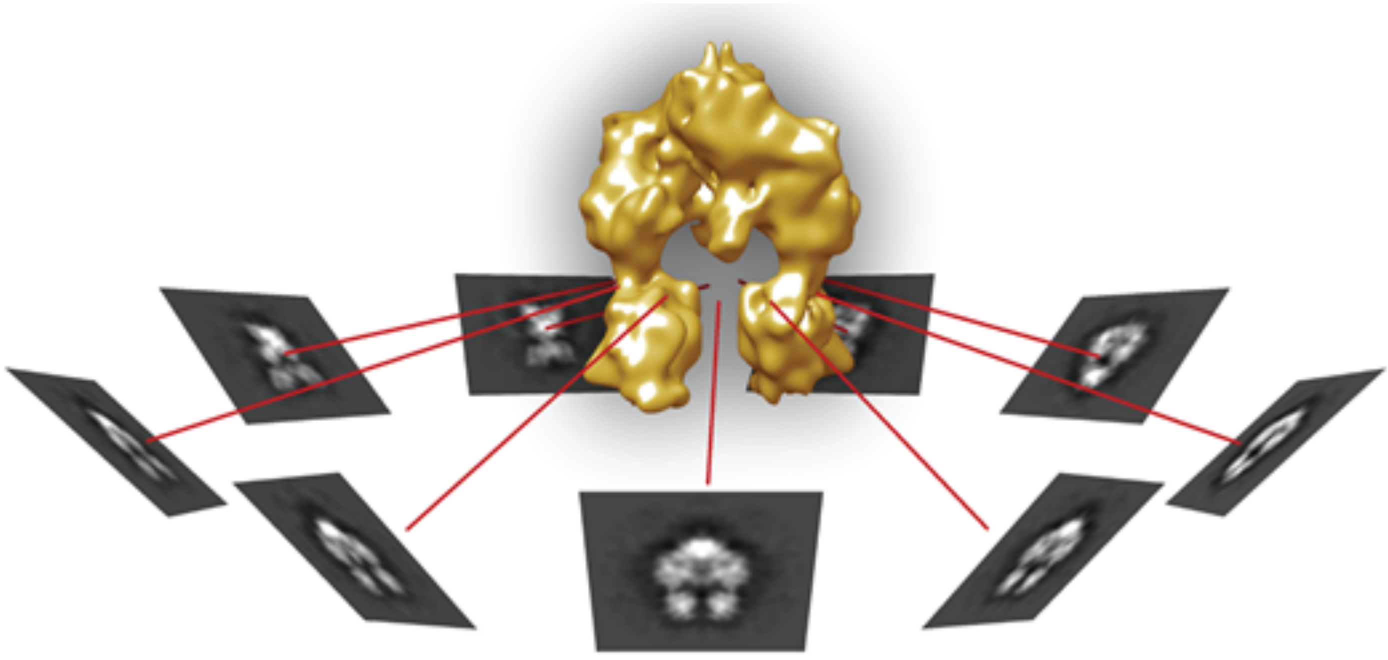
Fig. 6. Scheme for the general process of reconstruction of a structure from its transmission images.

De Rosier and Klug. Nature. Vol. 217, (1968) pp. 130 -134

Single-particle electron microscopy

- Cryo-EM images are 2D projections of 3D objects
- Each image therefore can fully describe a single view of the object
- 3D reconstructions are generated from combining many views together

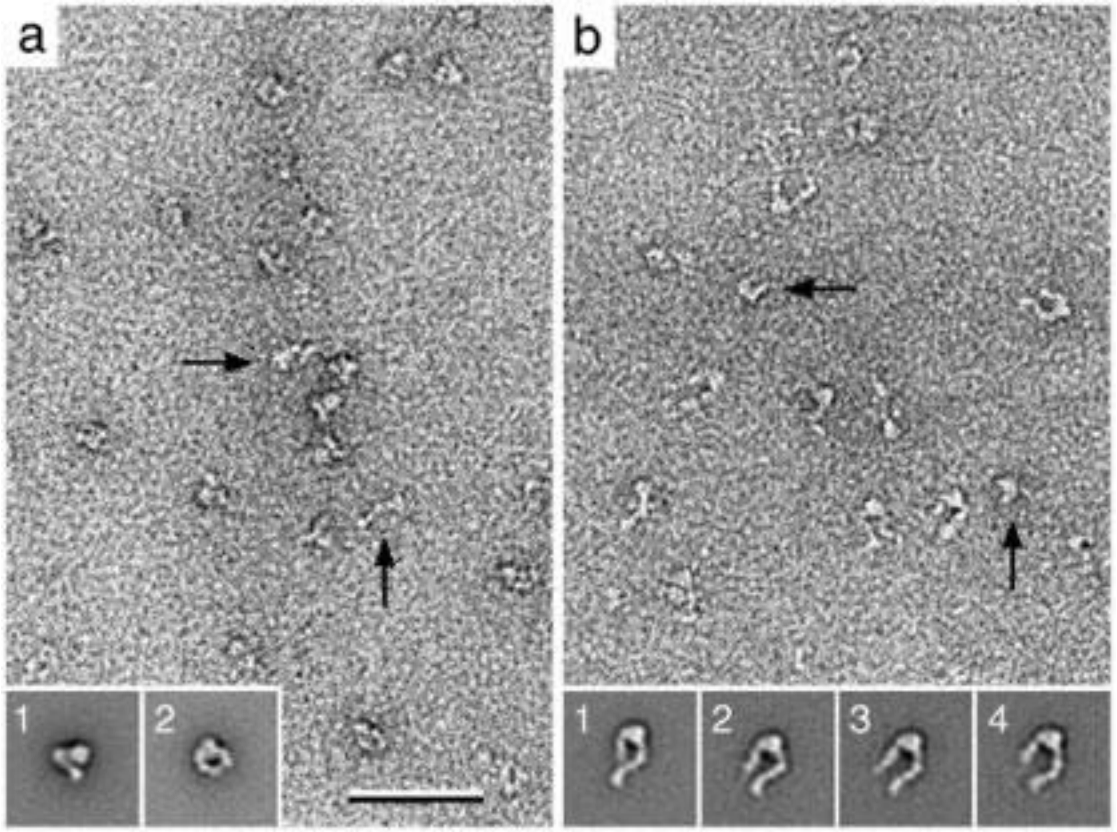


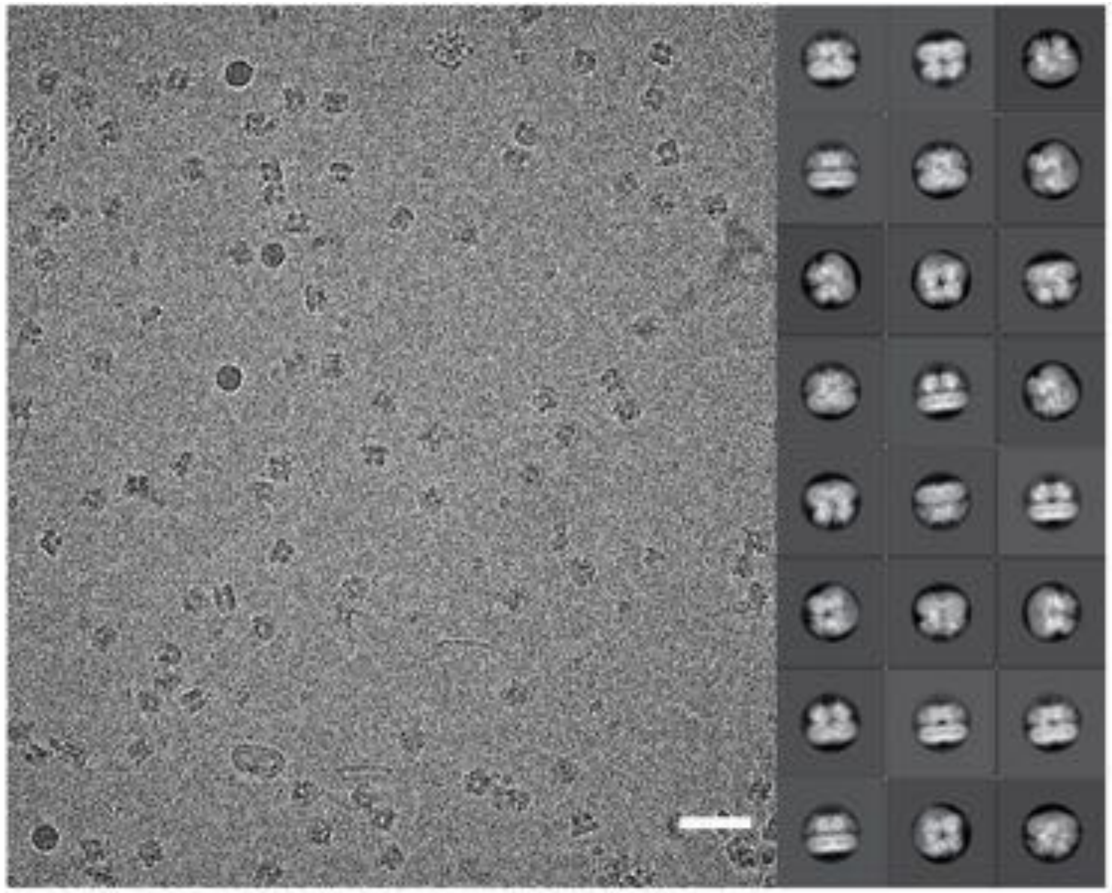


Skiniotis and Southworth. Microscopy. Vol. 65, (2016) pp. 9-22

Specimen preparation techniques

- Negative staining - embed the protein in a thin layer of heavy metal
  - Heavy metal scatters electron strongly
  - Visualize areas without stain (i.e. areas with protein)
  - Resolve the envelope of a protein complex
  - Limited to ~20 Å (grain size of the staining metal)
- Cryogenic vitrification - embed the protein in a thin layer of vitreous ice
  - Protein scatters more strongly than vitreous ice
  - Resolve complete protein structure (when ordered)
  - Resolution is limited by protein conformational stability

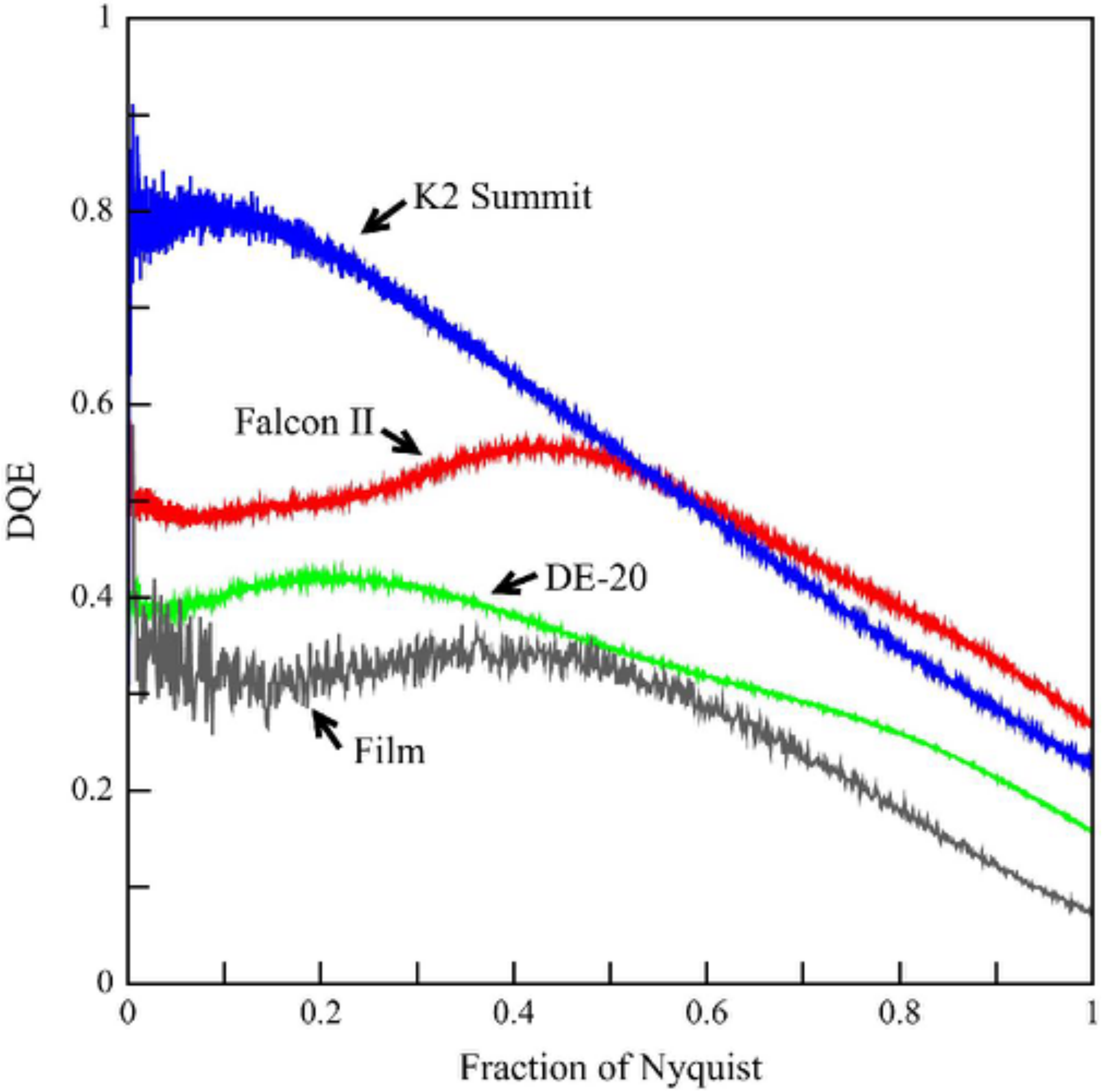




Ohi et al. Biol Proced Online. Vol 6, (2004) pp. 23–34.
Hite et al. Nature. Vol 541, (2017) pp. 46–51.

Direct electron detectors

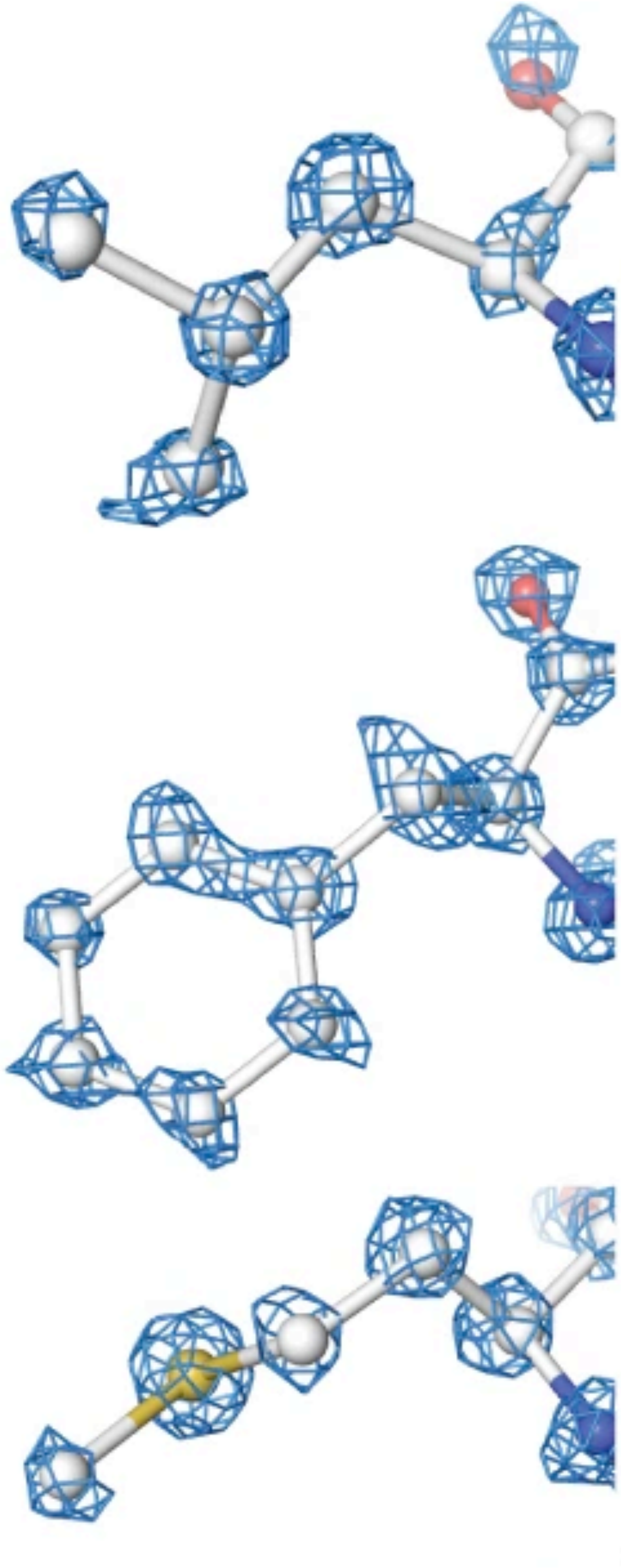
- Traditionally, EM images were recorded on film negatives and then digitized with a scanner
- In 2013, **direct electron detectors** were commercialized
  - Massively increased signal-to-noise ratio
  - Enabled data collection to be automated, increasing the number of images that can be obtained
- DEDs sparked the current era of high-resolution cryo-EM structure determination



McMullan et al, Ultramicroscopy, Vol 147 (2014) pp.156-163.

Overview - challenging specimen

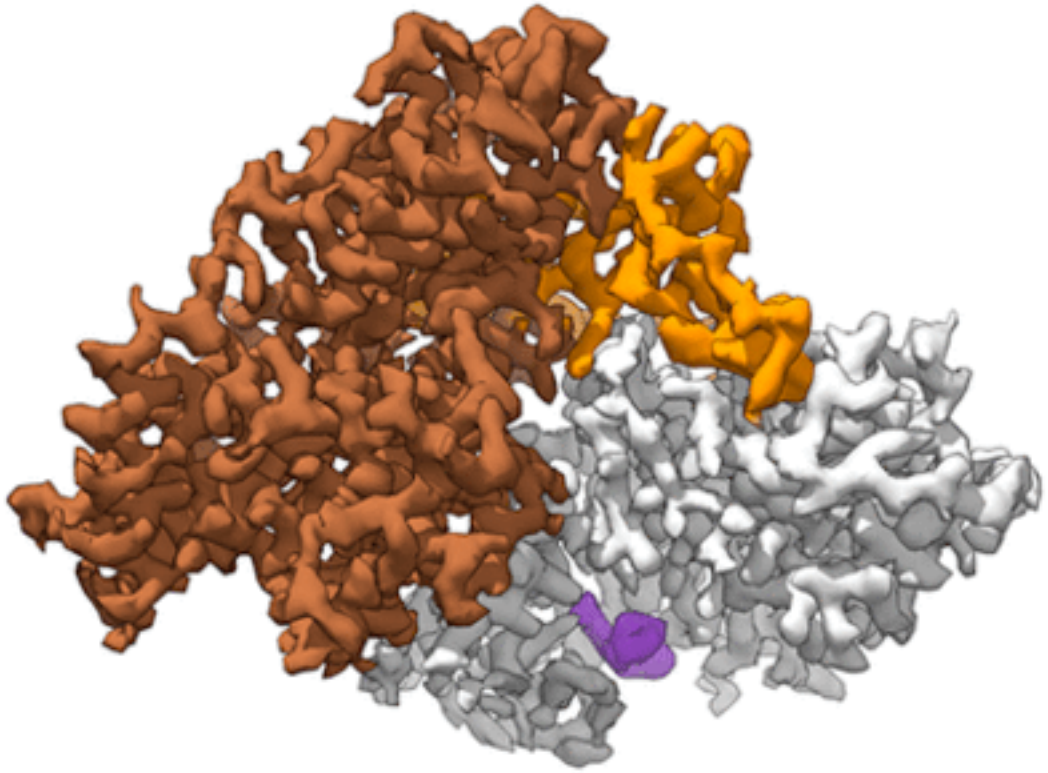
1. What features allow cryo-EM structures to achieve high-resolution?
2. What are the benefits of resolving high-resolution structures?
3. What can be learned from low-resolution structures?
4. What limitations persist that can degrade the resolution?
5. What approaches exist to help improve cryo-EM structure determination?



Nakane *et al*, Nature. (2020) 587. pp 156–162.

How to evaluate a cryo-EM map

- How was the map validated?
  - What evidence supports the hypothesis that the map faithfully describes the structure of the sample?
  - What methods can you use to provide additional evidence?
- Map resolution and interpretation
  - What is the resolution of the map?
  - How uniform is the resolution?
  - Do the features of the map (i.e. secondary structure, side chains) correspond to the resolution?
  - Was the model properly positioned in the structure and how was it validated?
- Sample heterogeneity
  - Does the sample contain a mixture of different proteins and/or conformations?
  - What can you learn from the heterogeneity of the sample?

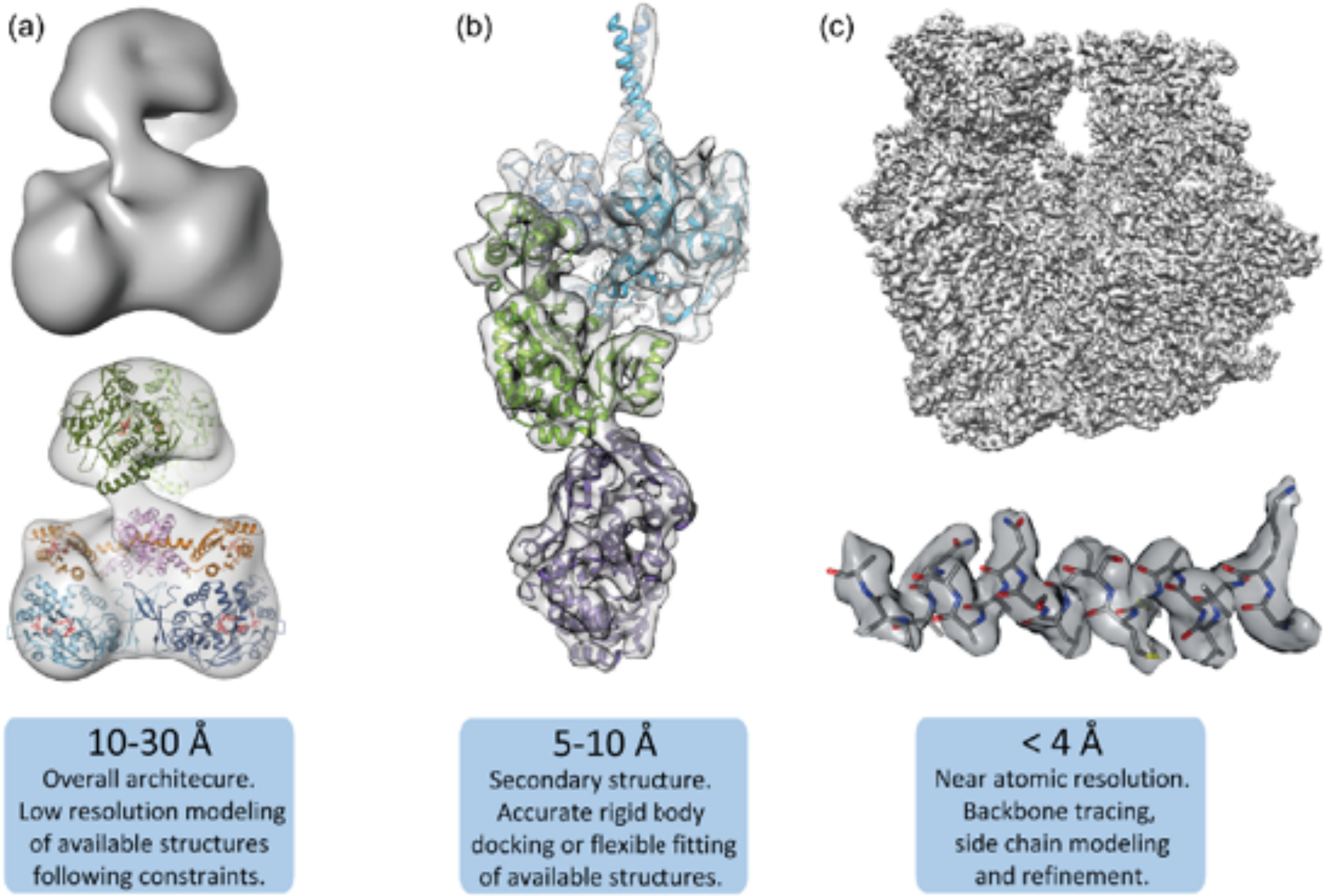


EMD-12042
2.5 Å

Greber BJ *et al*, Biophys. J (2021) pp 667-686.

Resolution - what does it mean?

- Resolution corresponds to the distance between two objects that can be reliably separated in the map
- In EM, resolution is estimated by the Fourier shell correlation (FSC)
- FSC is determined by comparing the cross correlation between two independently calculated maps in Fourier space resolution shells
- Practically, this is done by separating a data set (10,000 - 1,000,000 particle images) into two halves and determining independent reconstructions of each half
- Different resolutions have different types of features that can be visualized in the density map



10-30 Å
Overall architecure.
Low resolution modeling
of available structures
following constraints.

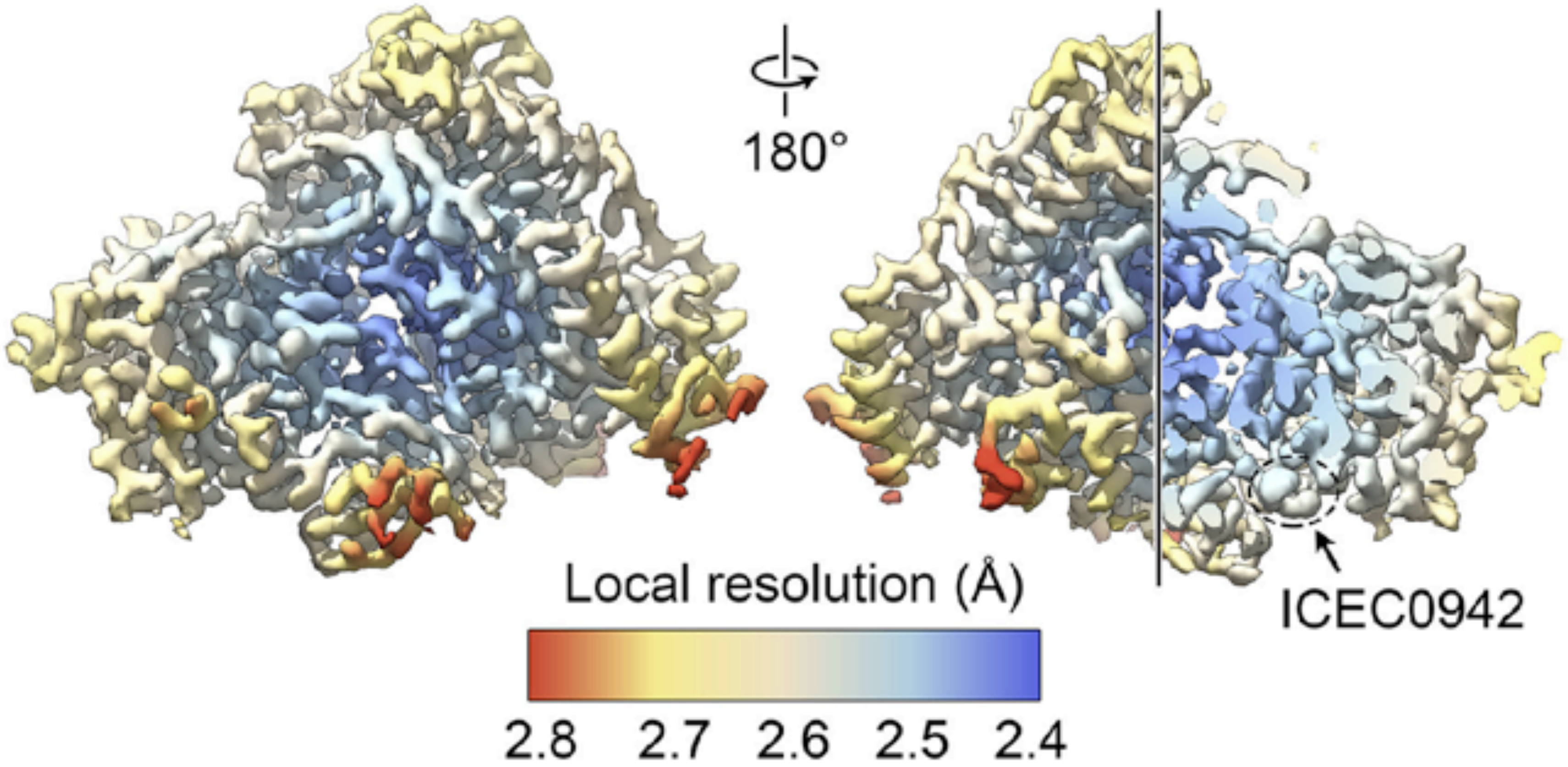
5-10 Å
Secondary structure.
Accurate rigid body
docking or flexible fitting
of available structures.

< 4 Å
Near atomic resolution.
Backbone tracing,
side chain modeling
and refinement.

Skiniotis and Southworth. Microscopy. Vol. 65, (2016) pp. 9-22

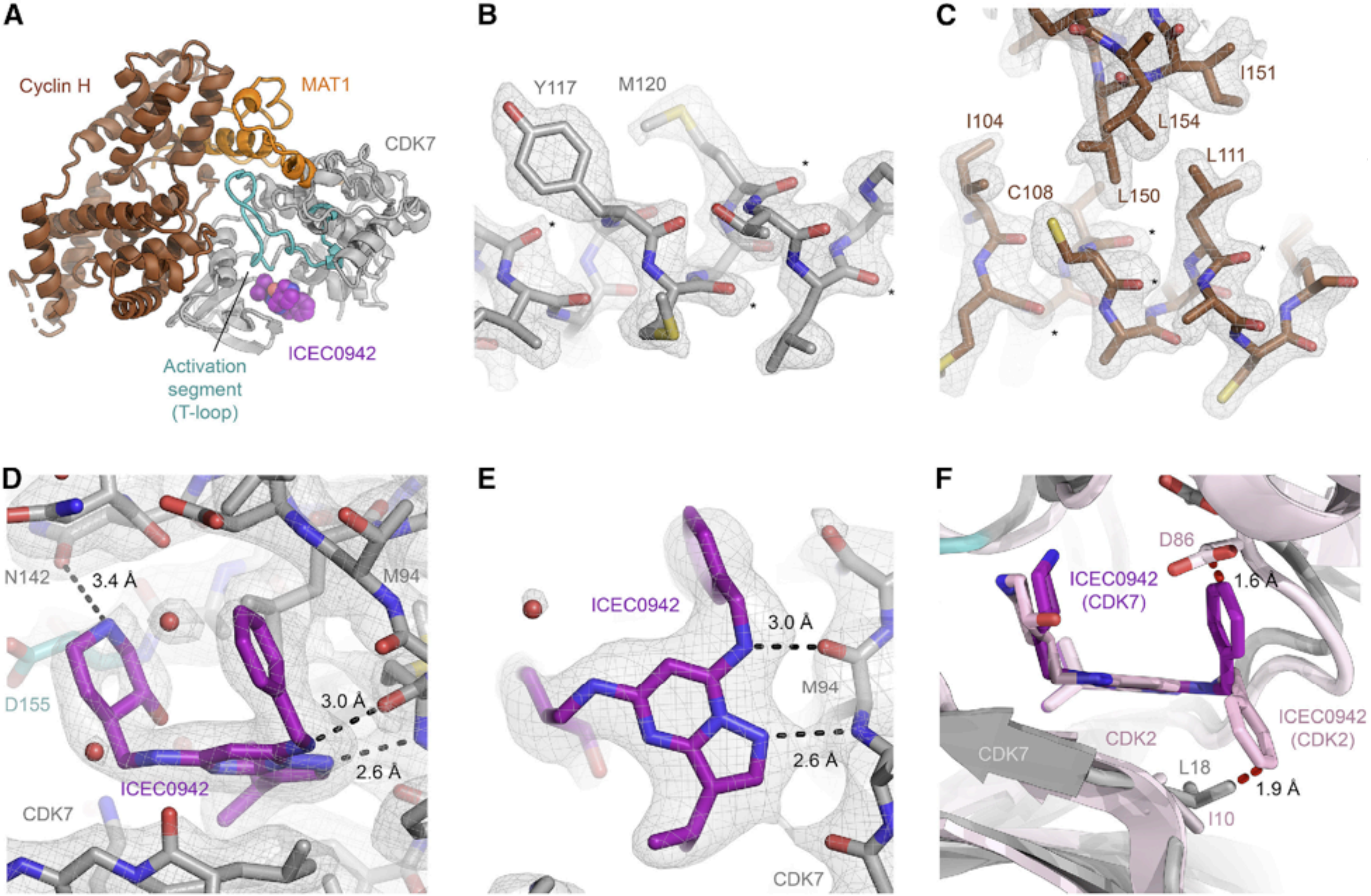
Resolution - what does it mean?

- Resolution corresponds to the distance between two objects that can be reliably separated in the map
- In EM, resolution is estimated by the Fourier shell correlation (FSC)
- FSC is determined by comparing the cross correlation between two independently calculated maps in Fourier space resolution shells
- Practically, this is done by separating a data set (10,000 - 1,000,000 particle images) into two halves and determining independent reconstructions of each half
- Different resolutions have different types of features that can be visualized in the density map



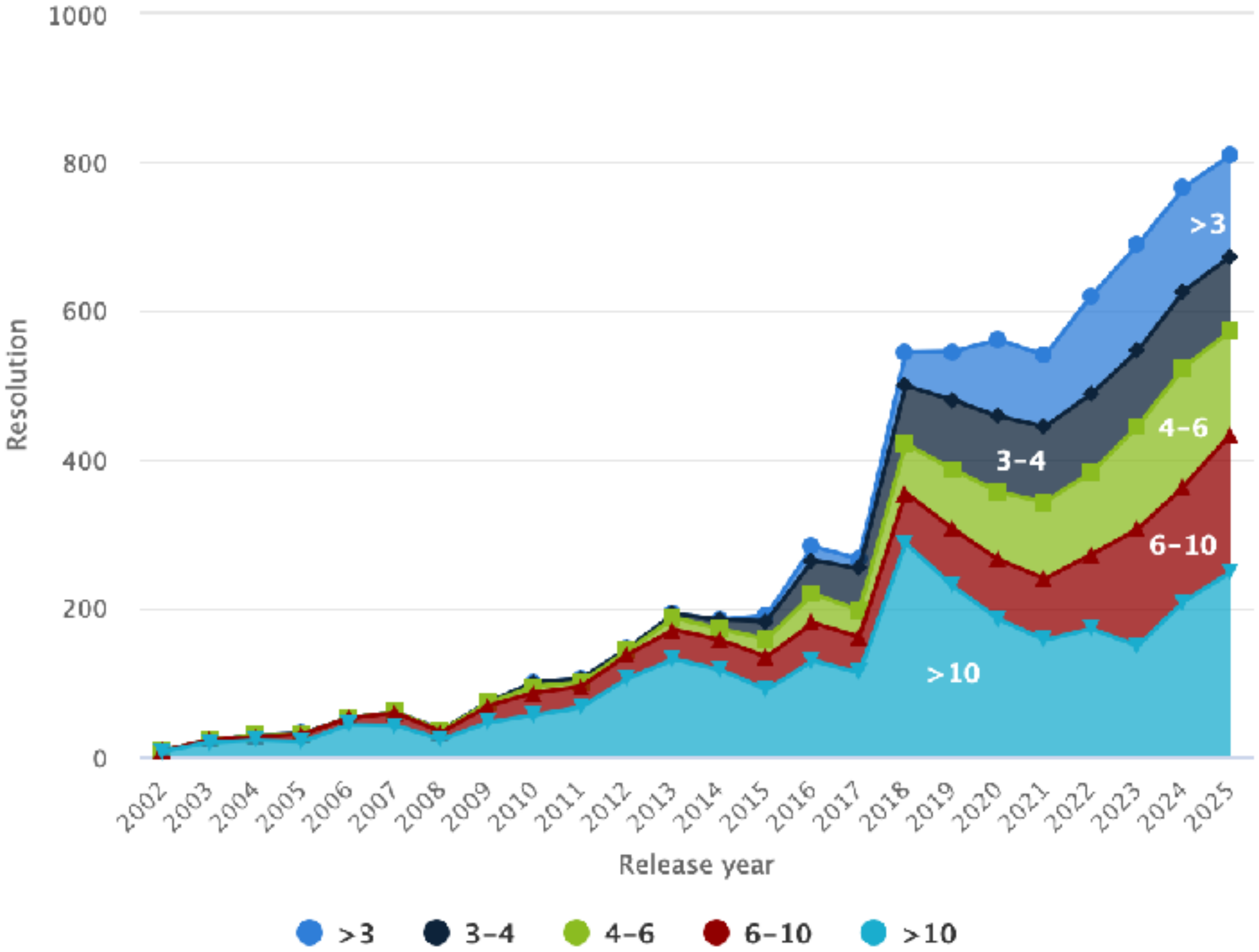
Greber BJ *et al*, Biophys. J (2021) pp 667-686.

Resolution - what does it mean?



Greber BJ *et al*, Biophys. J (2021) pp 667-686.

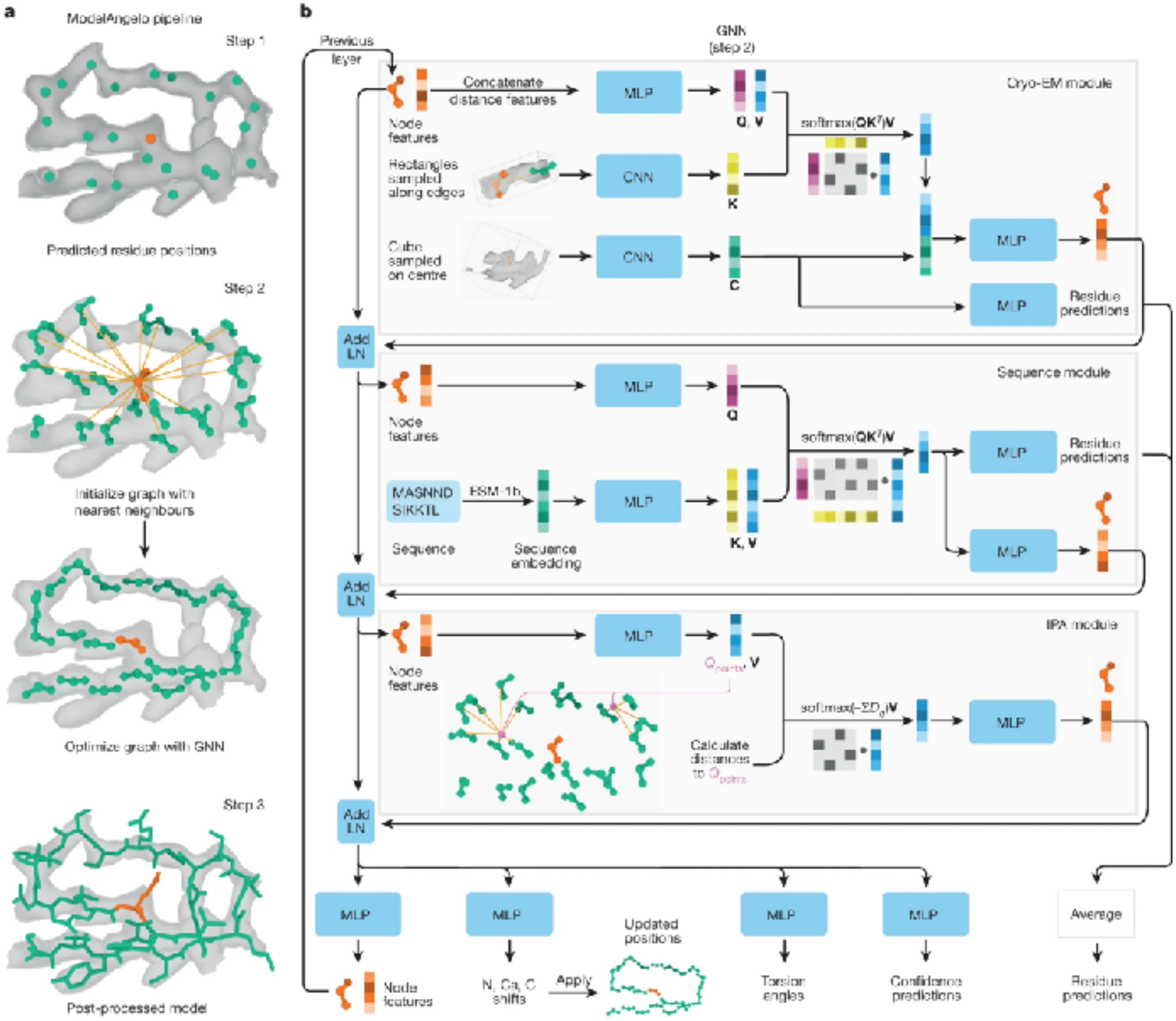
Resolution - what does it mean?



EMDB 2026.3.17

Map to Model

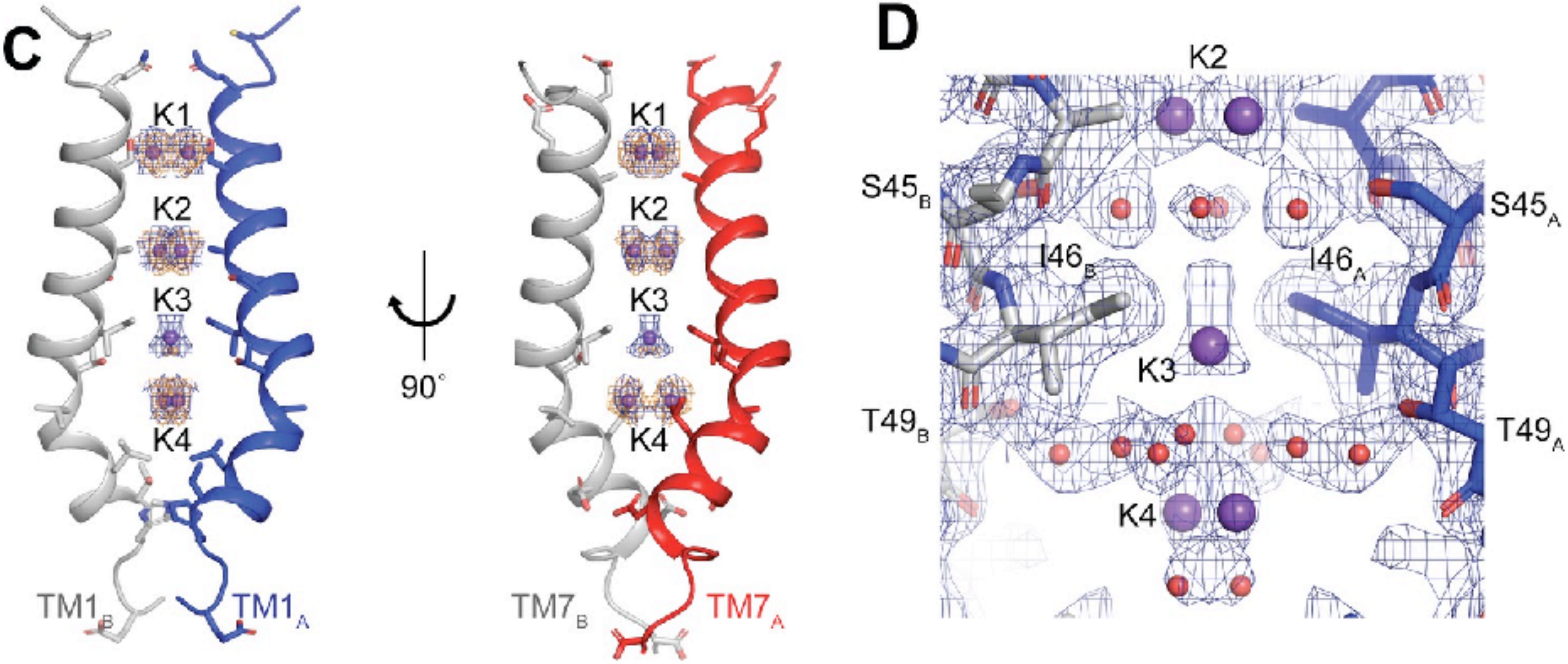
- Model building has been rapidly accelerated
- Hand building of high-resolution structures has been replaced with automated approaches
- Modelangelo is one of them that can rapidly build into maps at better than 3.5 Å with high accuracy



Jamali *et al*, Nature. (2024)

Is all density in a map is the same?

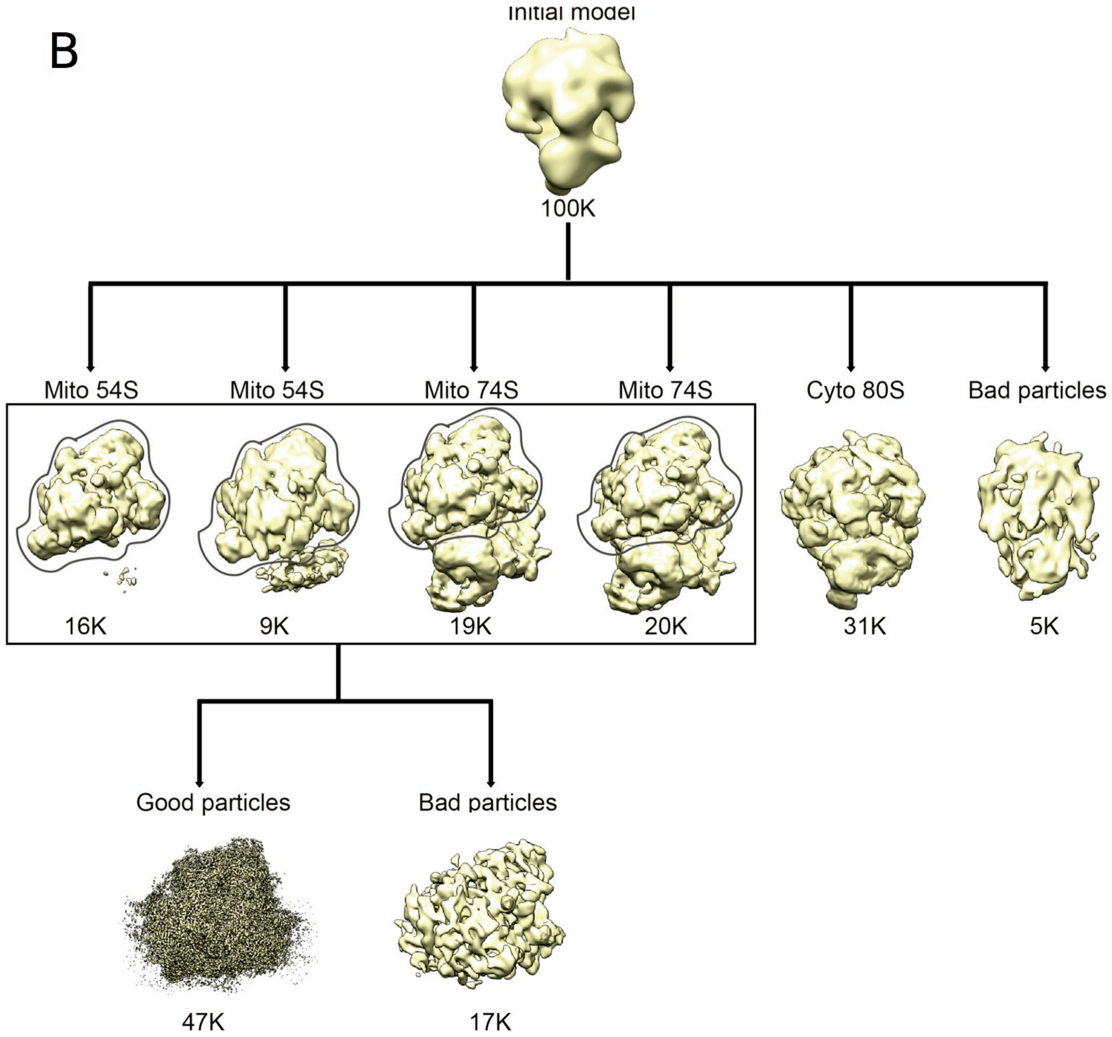
- Identifying non-protein densities in cryo-EM is quite challenging as there few tools that allow accurate discrimination between atoms in density maps.
- In some circumstances one can replace an atom with a similarly structured atom to distinguish ion binding sites



Oh *et al*, eLife. (2020)

Sample heterogeneity

- There are multiple sources of heterogeneity in sample preparation
  - Compositional heterogeneity - mixture of different components or mixtures with varying subunit stoichiometry
  - Structural heterogeneity - domains of the specimen can adopt multiple conformations
  - In some cases, both types of heterogeneity exist within a single sample
- These will degrade the resolution of reconstructions if not sorted computationally, but can provide insights into function of the specimen



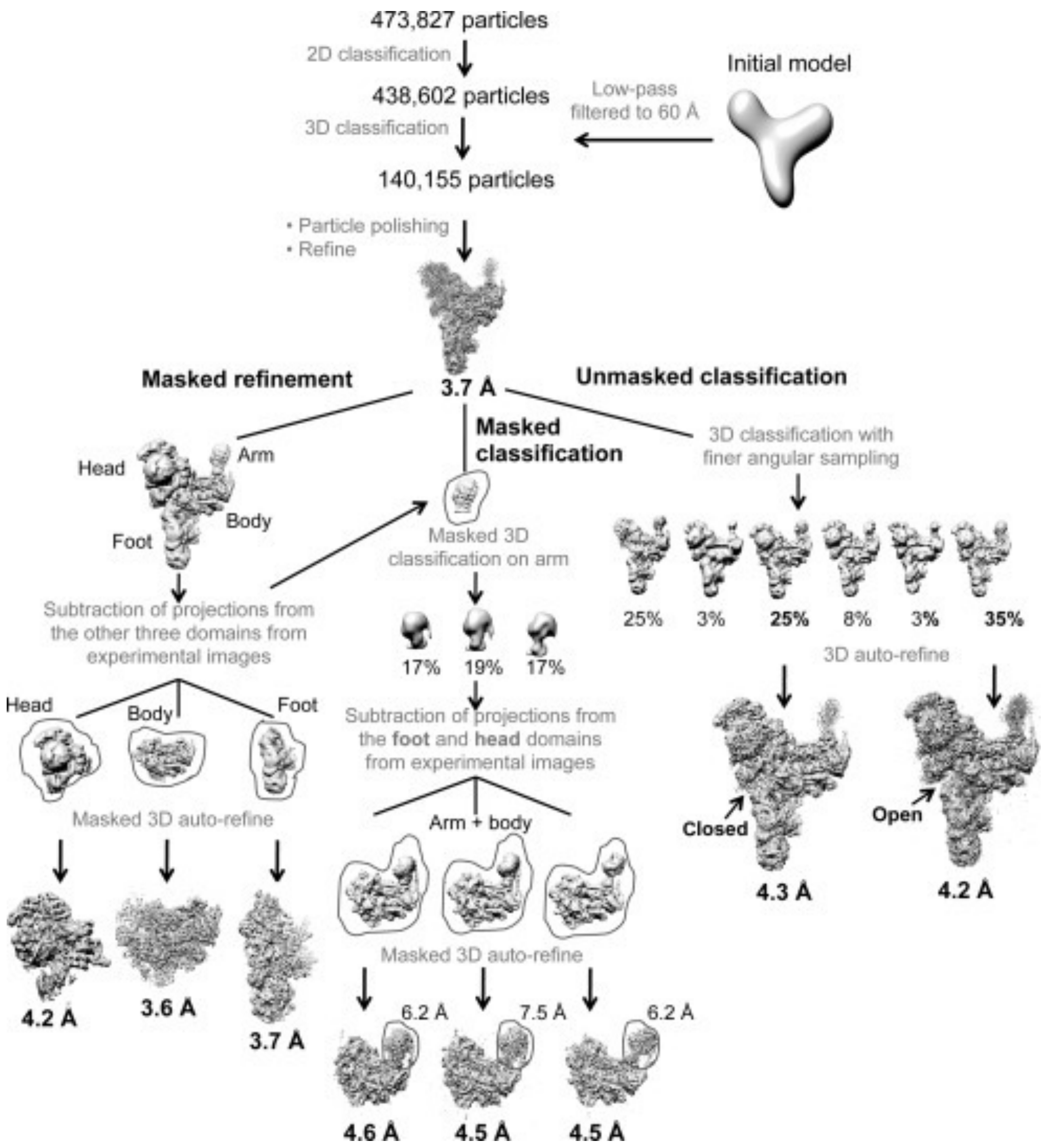
Amunts *et al*, Science. (2014) 343:1485-1489

Overcoming heterogeneity - biochemistry

- Optimizing biochemistry can often help to alleviate heterogeneity and is generally the best place start to improve sample quality
  - Improvements in sample purification can reduce compositional heterogeneity by obtaining a more uniform starting sample
  - Structural heterogeneity can be minimized by altering purification conditions (i.e. presence of activating or inhibiting ligands, different pH or salt conditions)
  - Construct alterations can also reduce sample heterogeneity by removing flexible domains
- In some cases chemical cross-linking can helpful to reduce flexibility
  - Testing cross-linking reagents with different lengths and varying the concentration can be helpful to optimize conditions
  - However, it is essential that the chemically cross-linked structure be validated with a non-cross-structure to demonstrate the the cross-linking does not introduce artifactual protein-protein interactions

Overcoming heterogeneity - computation

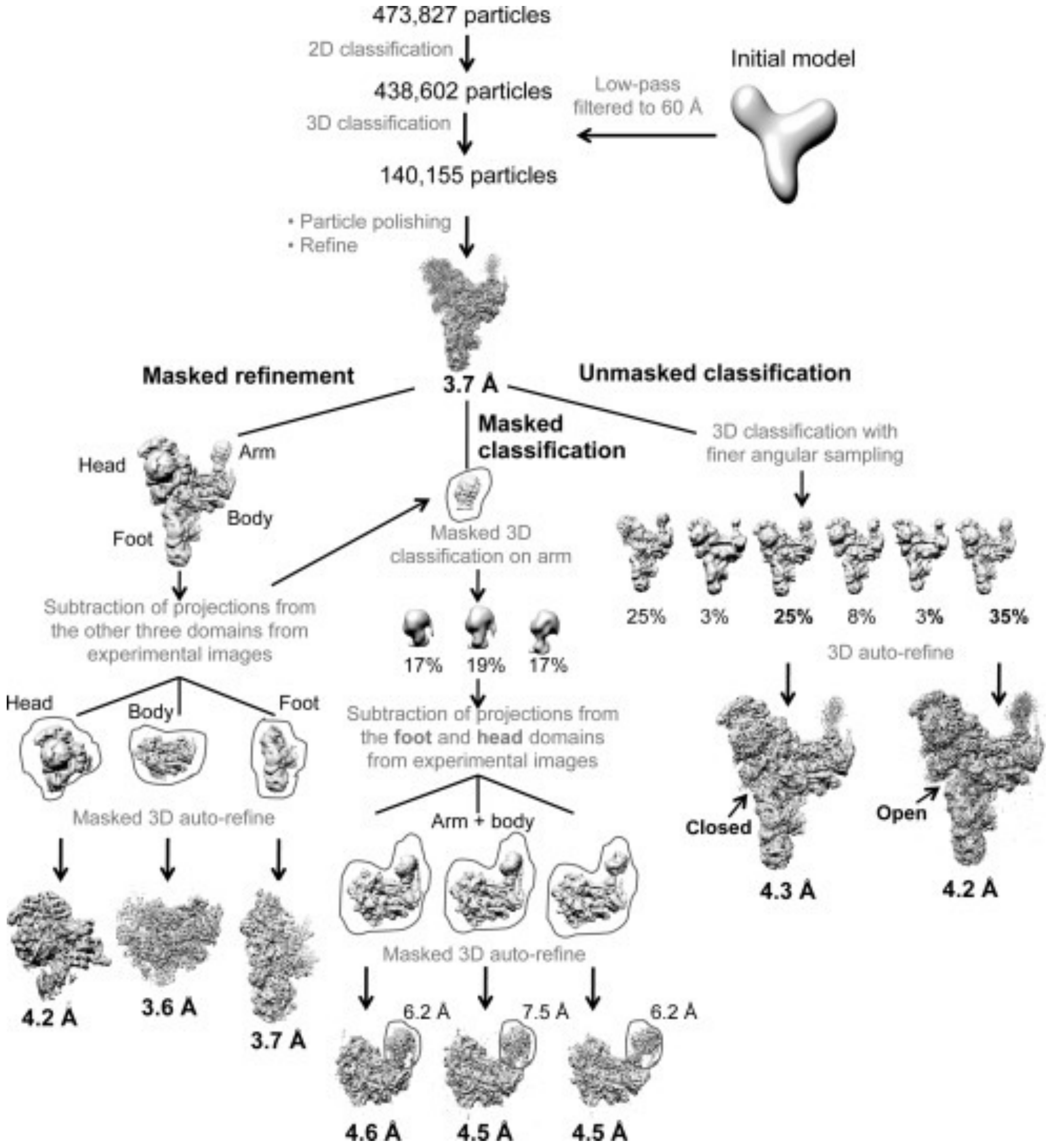
- Heterogeneity may be unavoidable for some samples and must be dealt with computationally after image acquisition
- There are now several different software packages that sort and classify particles, allowing one to create “pure” subsets of the particles images
- The simplest approach is classify based upon the entire molecule, which works well with large conformational differences



Scheres, Meth Enz. (2016) 579:125-157

Overcoming heterogeneity - computation

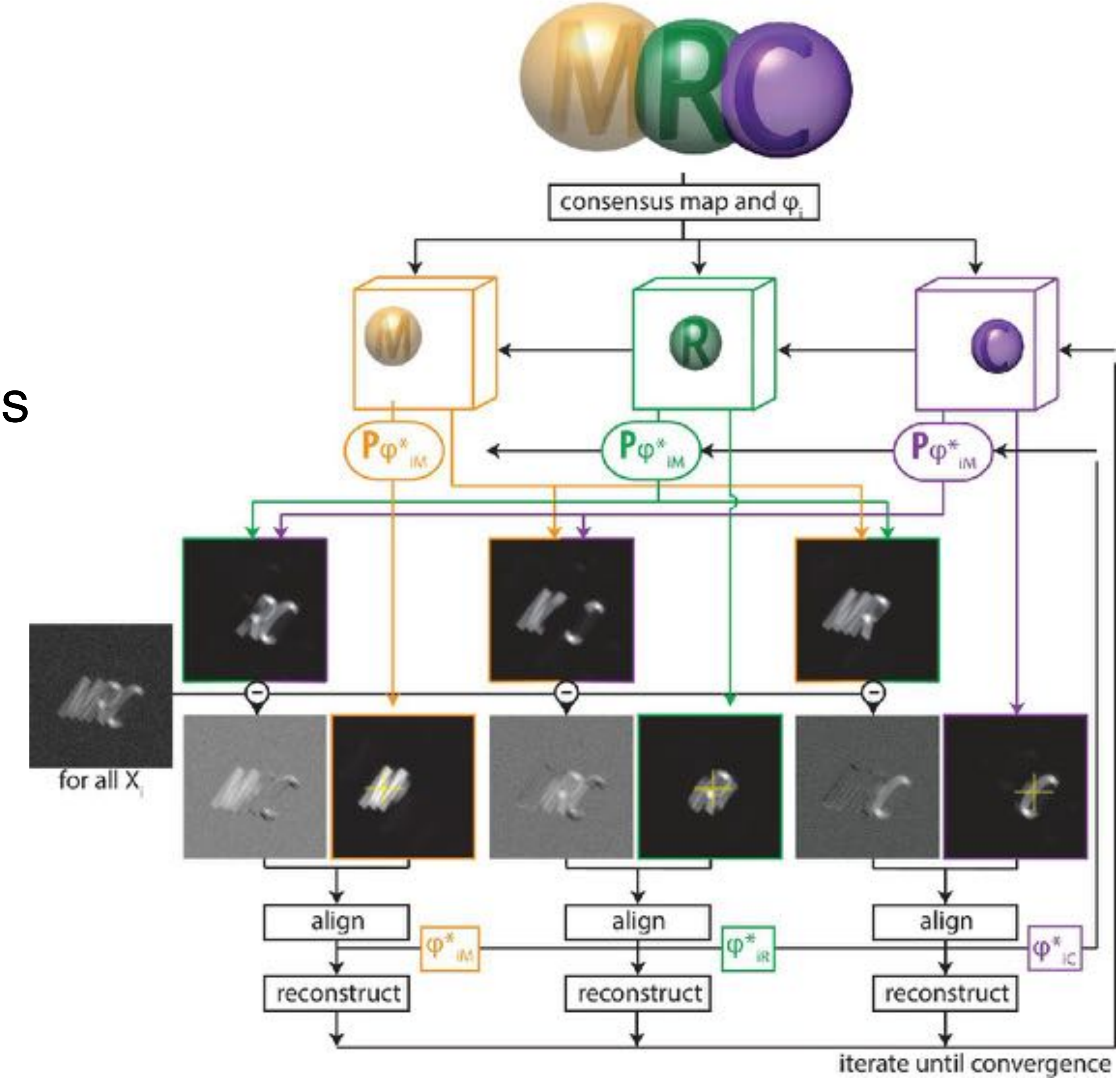
- Classification can be enhanced through the use of masks
- A mask can be placed around the region of interest - allowing independent sorting of different domains
- This multi-classification approach is particularly powerful for samples that have multiple different types of movements
- Another modification to classification is the use of background subtraction prior to classification to reduce the signal of constant domains during classification



Scheres, Meth Enz. (2016) 579:125-157

Overcoming heterogeneity - computation

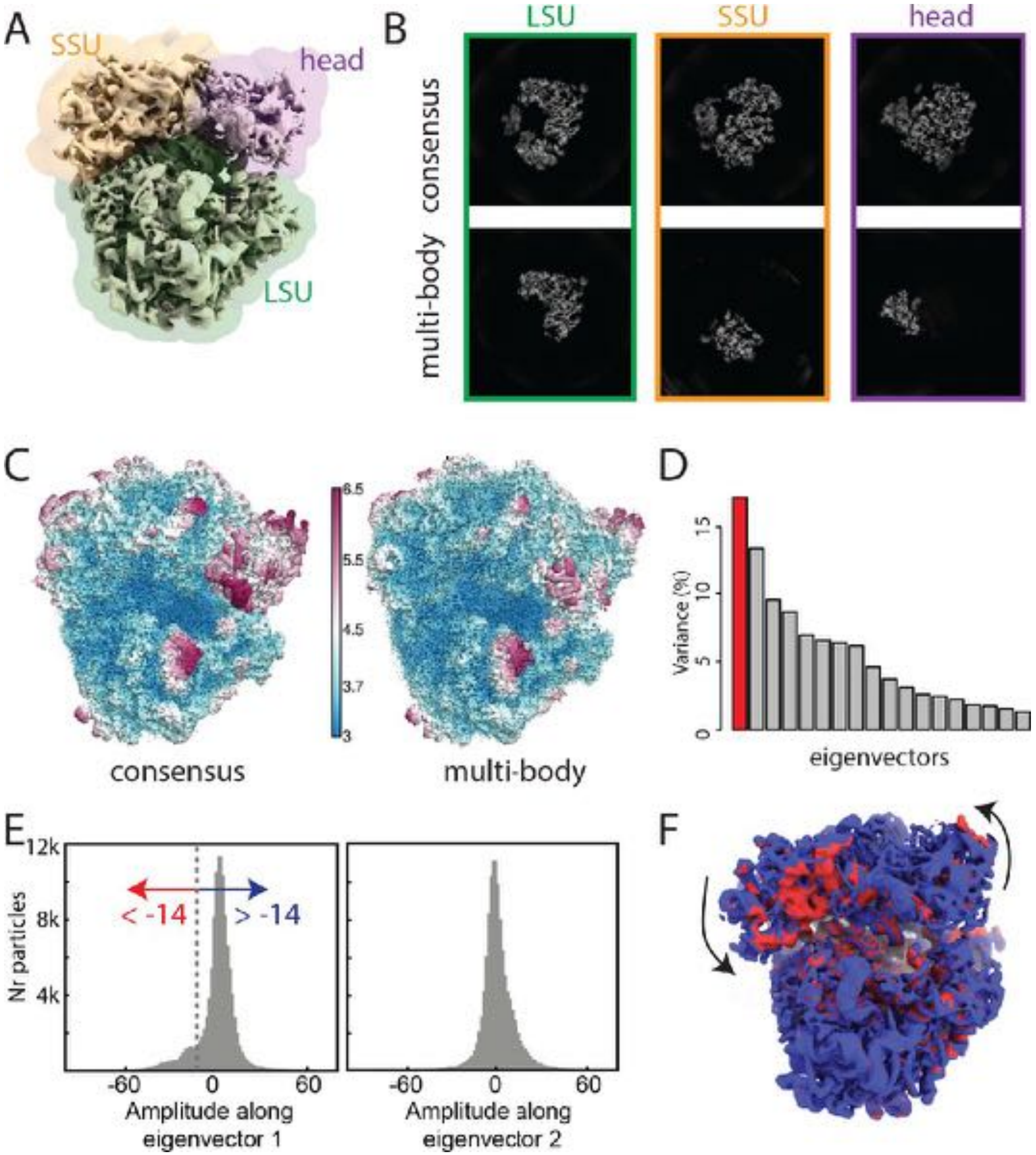
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Overcoming heterogeneity - computation

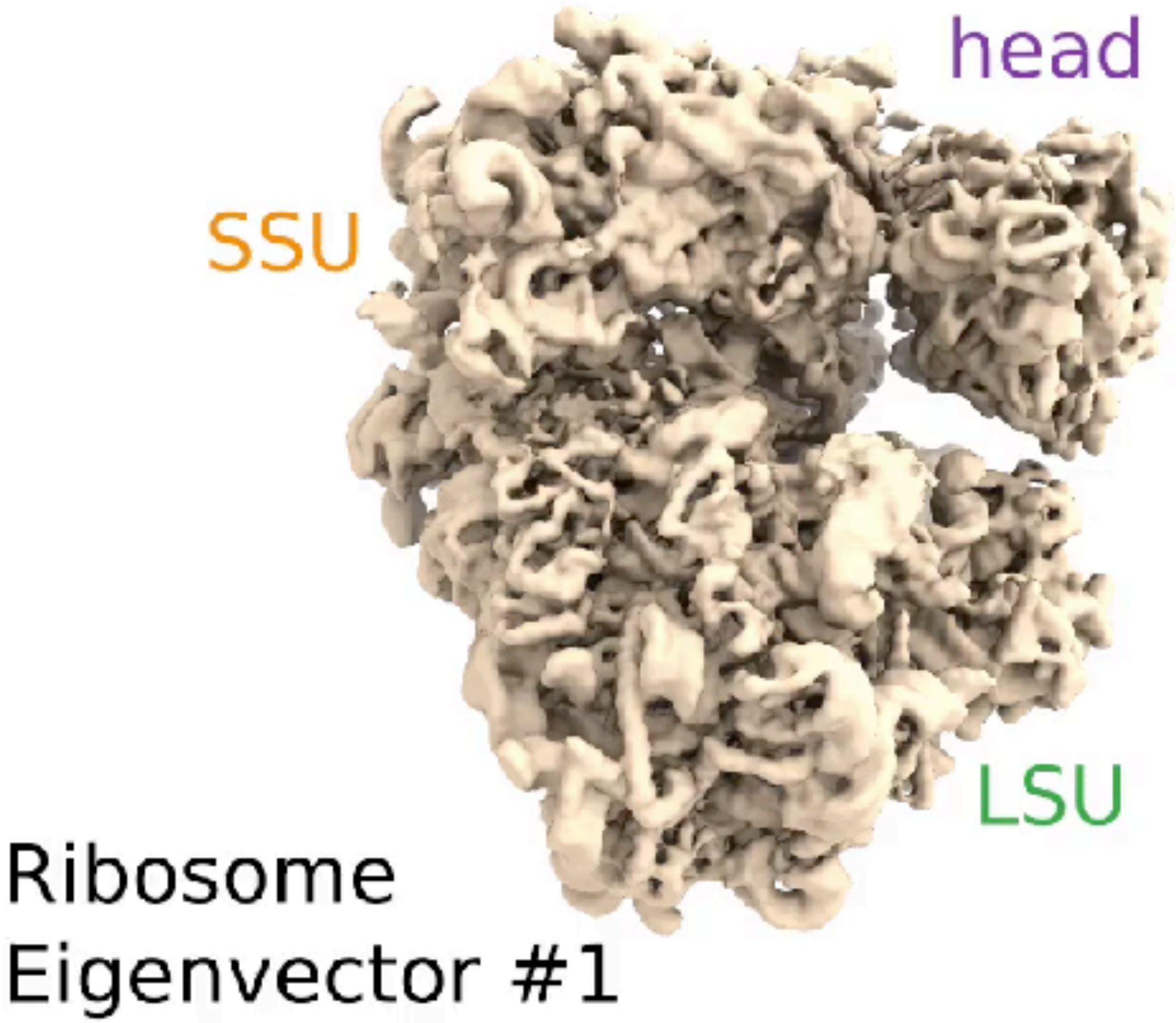
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

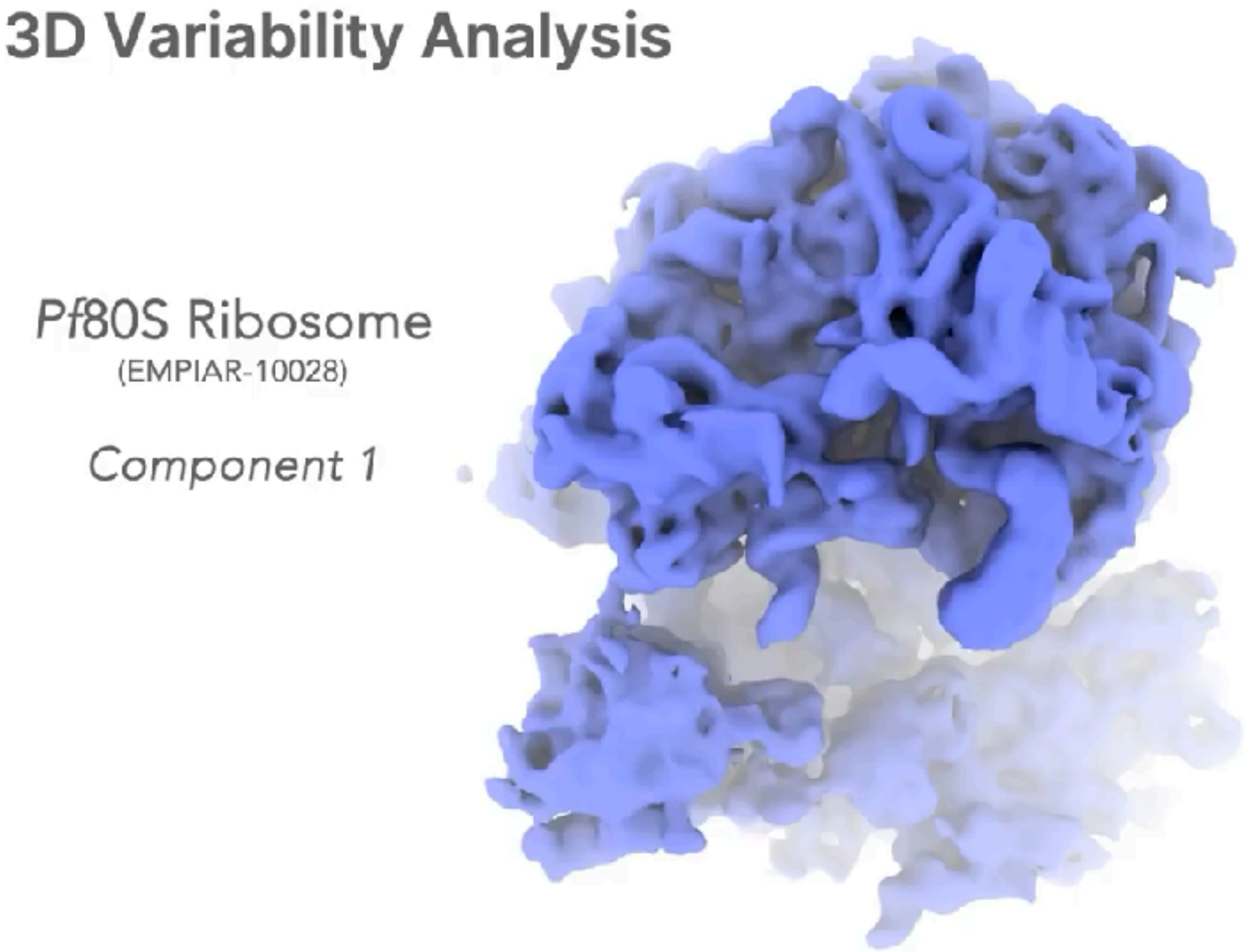
Overcoming heterogeneity - computation

- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes
- Requires knowledge of the major domains of movement



Nakane et al, eLife. (2018) e36861

Overcoming heterogeneity - computation



3D Variability Analysis

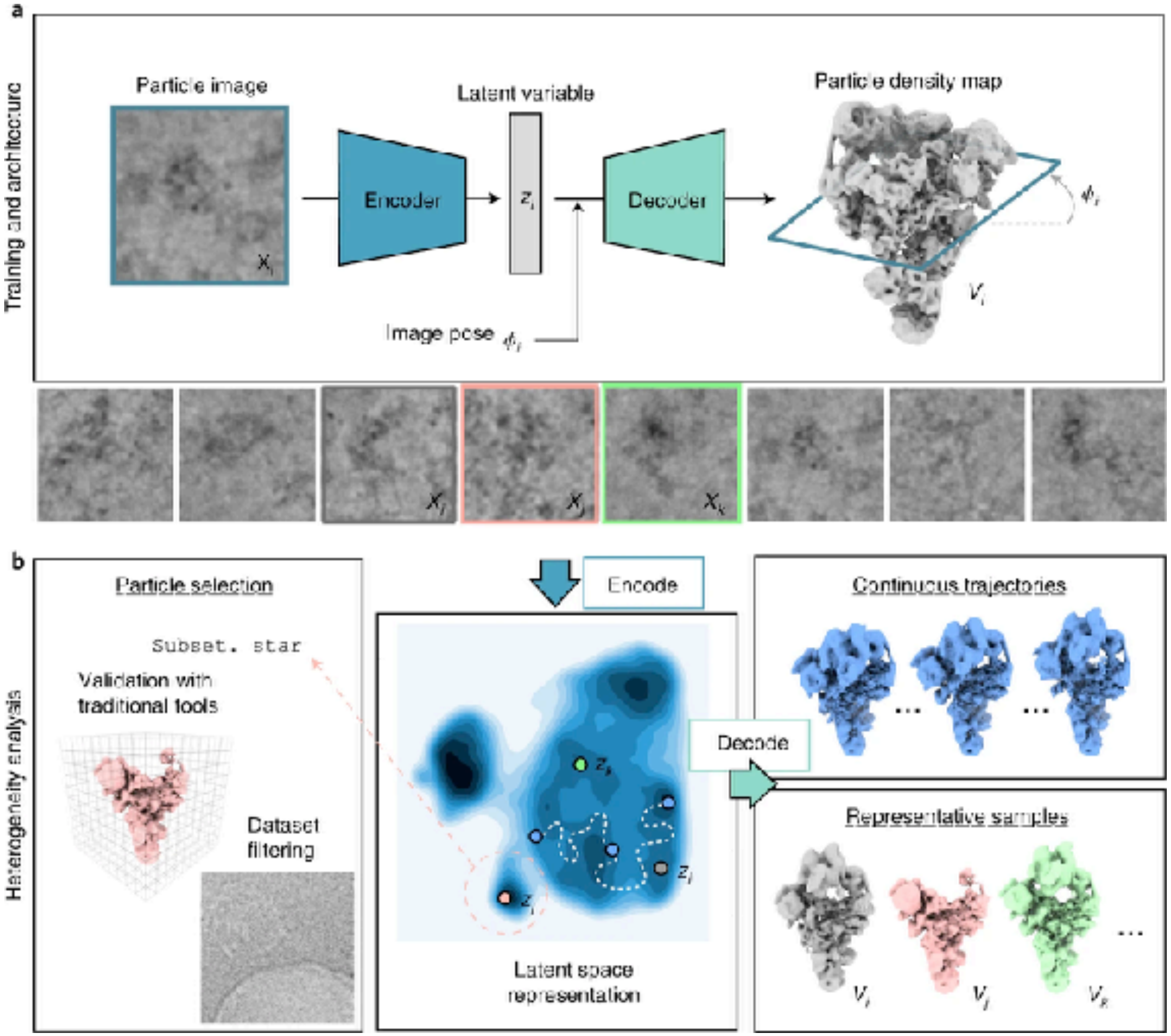
*Pf*80S Ribosome
(EMPIAR-10028)

Component 1

Punjabi and Fleet. JSB. (2021)

Overcoming heterogeneity - computation

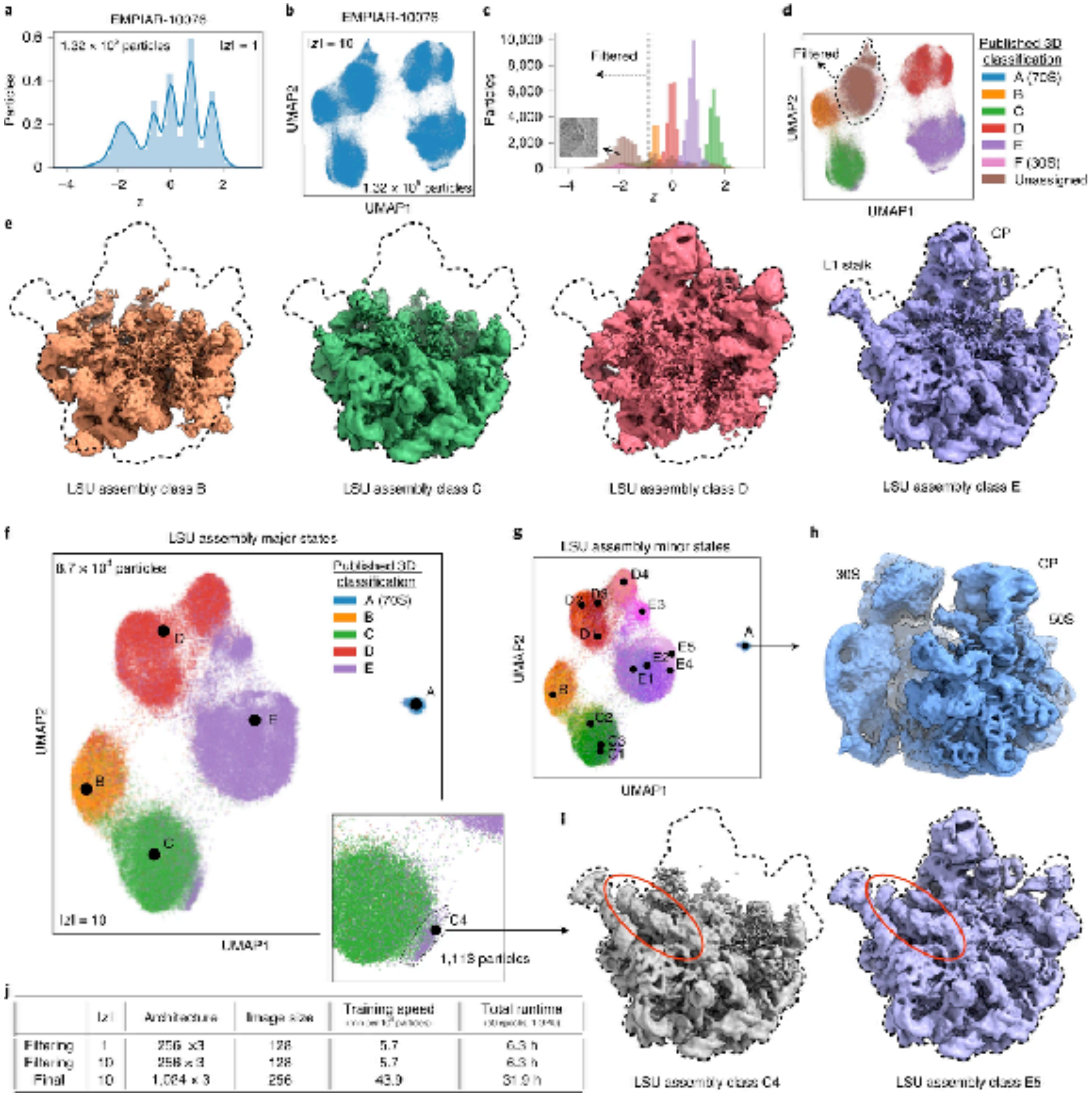
- CryoDRGN - deep learning approach to visualize conformational dynamics
- Does not require classification, uses particles directly and sorts into a latent space representation
- Allows analysis of highly complex data sets



Zhong et al., Nat Methods. (2021)

Overcoming heterogeneity - computation

- CryoDRGN - deep learning approach to visualize conformational dynamics
- Does not require classification, uses particles directly and sorts into a latent space representation
- Allows analysis of highly complex data sets

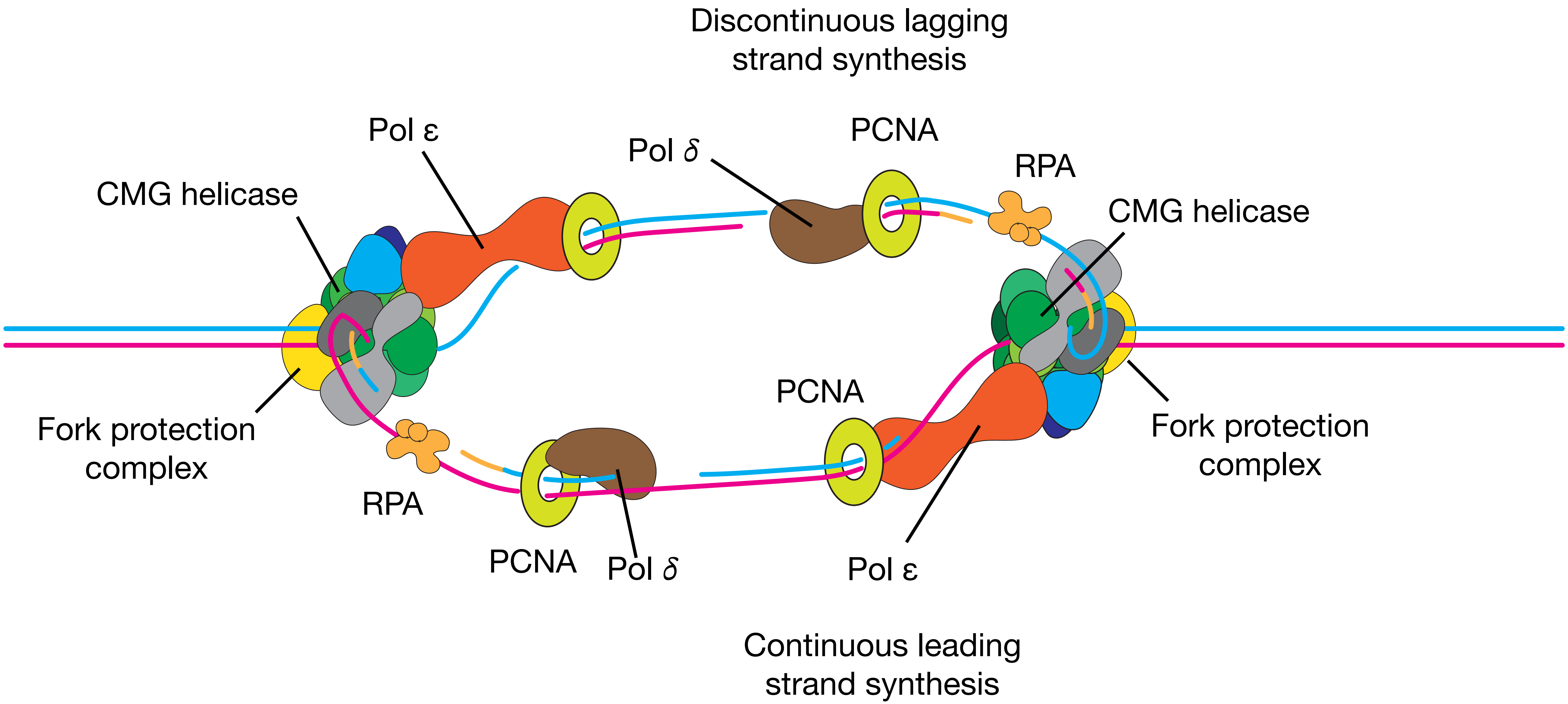


Zhong et al., Nat Methods. (2021)

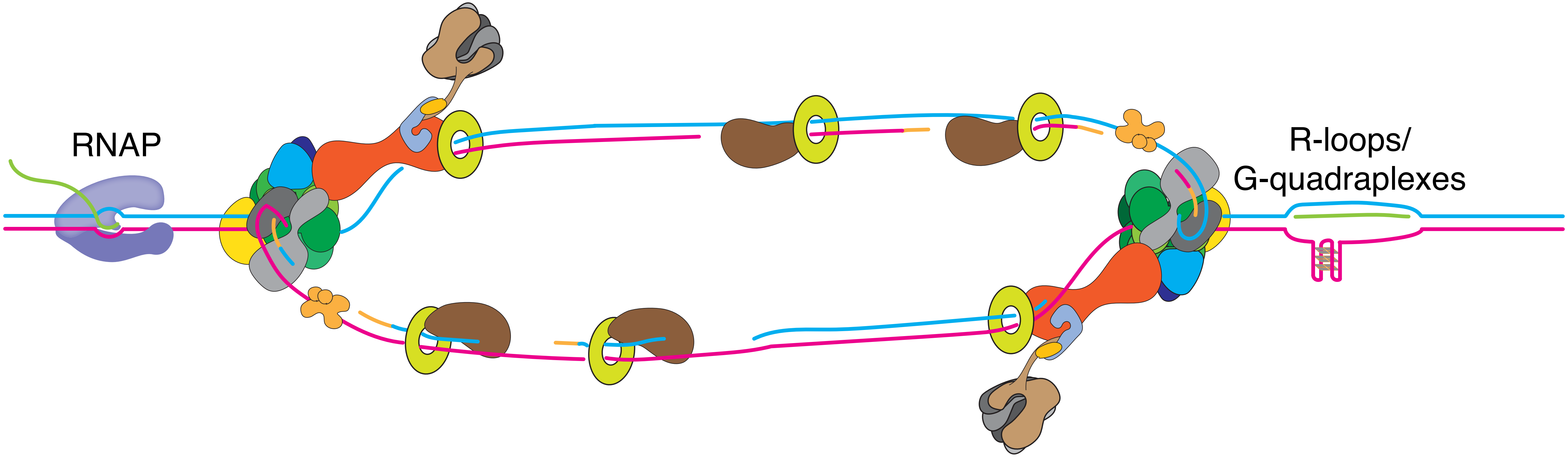
Benefits of heterogeneity?

- How can you use heterogeneity to better understand the biology of your samples?
- Does your heterogeneity correlate with functional changes?
- Always test to ensure that your representative density map is actually representative of your sample, and not merely some small portion of the particles that generate a high-resolution structure?
  - If the map does result from a very small fraction of particles, try to understand why?
  - Can you test activity to see if that makes sense biologically?

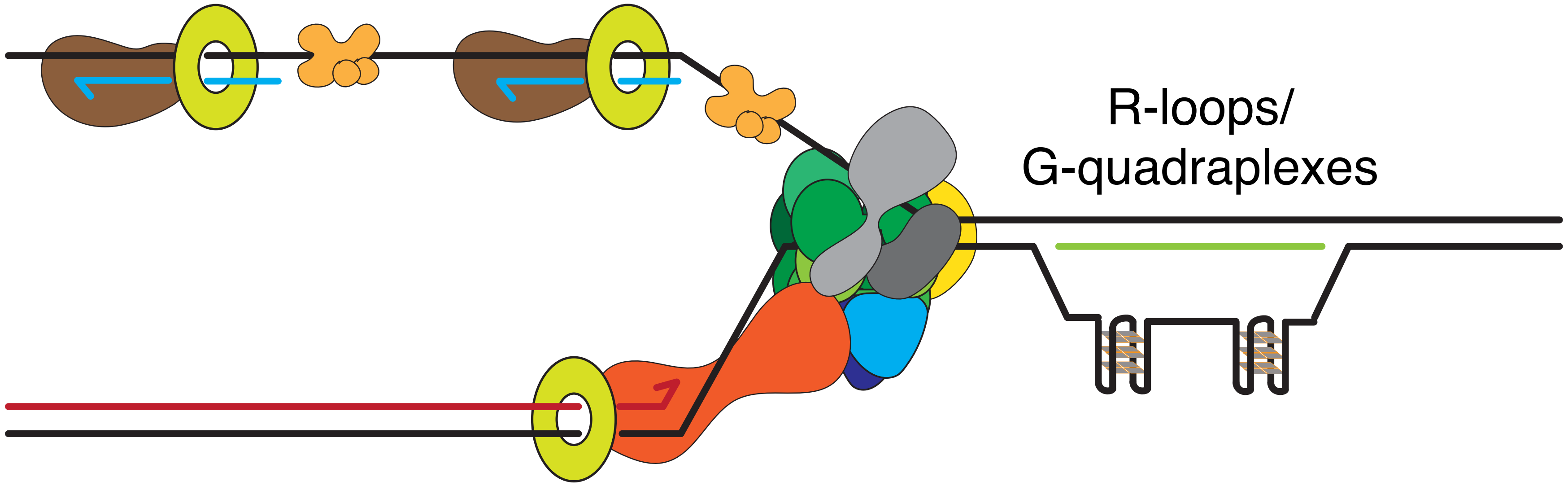
Case Study 1: DNA replication



Impediments to replisome function

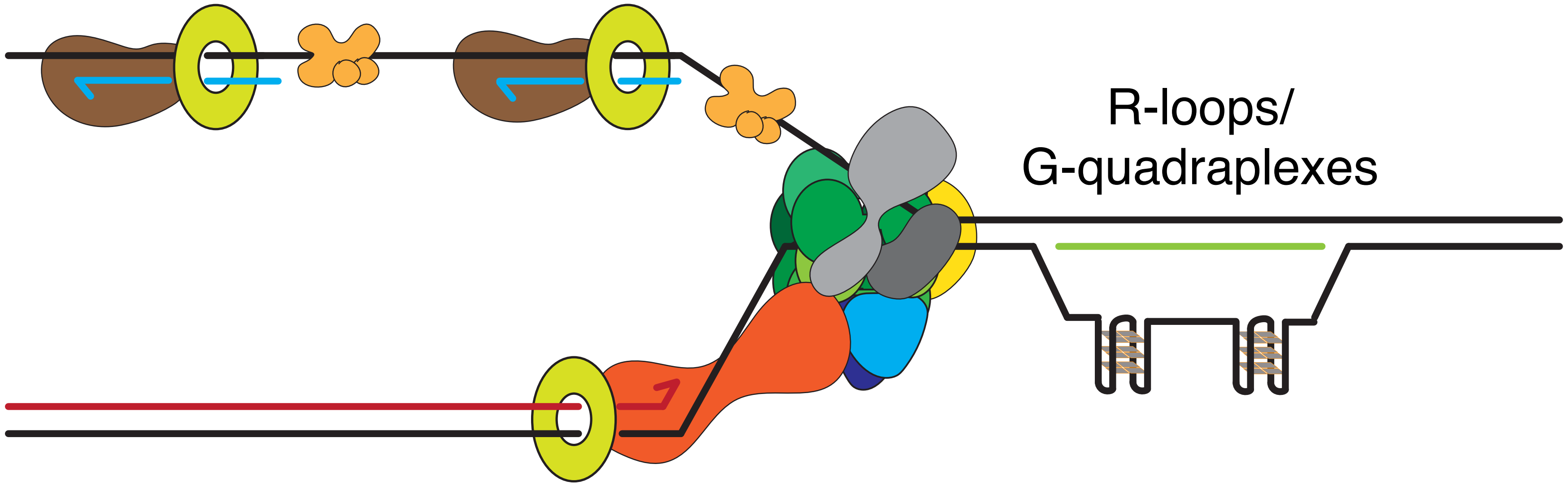


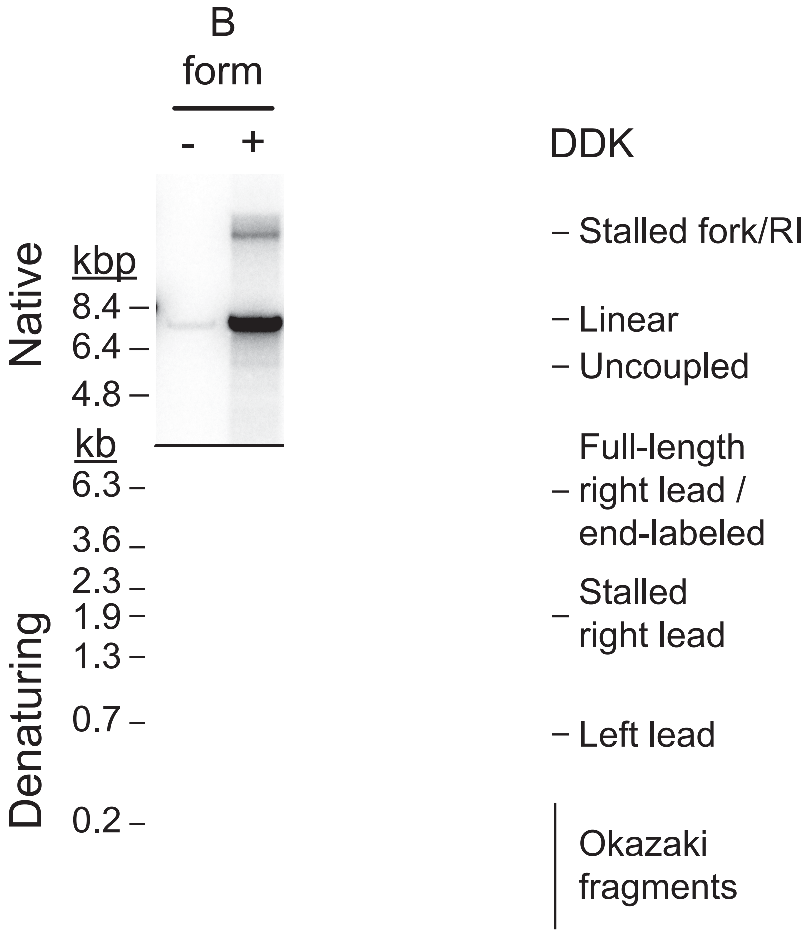
G4 quadraplexes are impediments to replisome function



Kumar et al., 2021, PMID 34494544

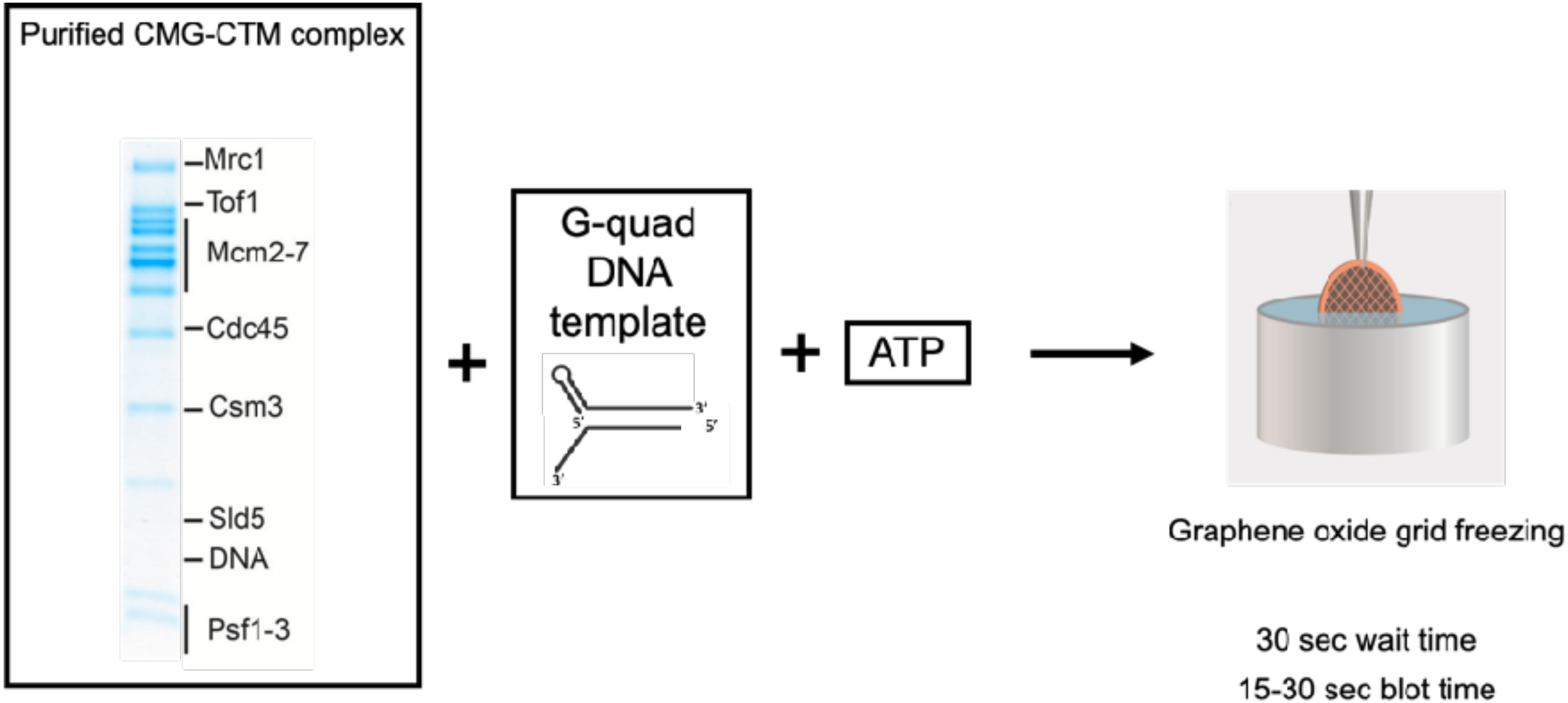
G4 quadraplexes are impediments to replisome function





Kumar et al., 2021, PMID 34494544

Grid preparation for structural analysis



Purified CMG-CTM complex

\+

G-quad DNA template

\+

ATP

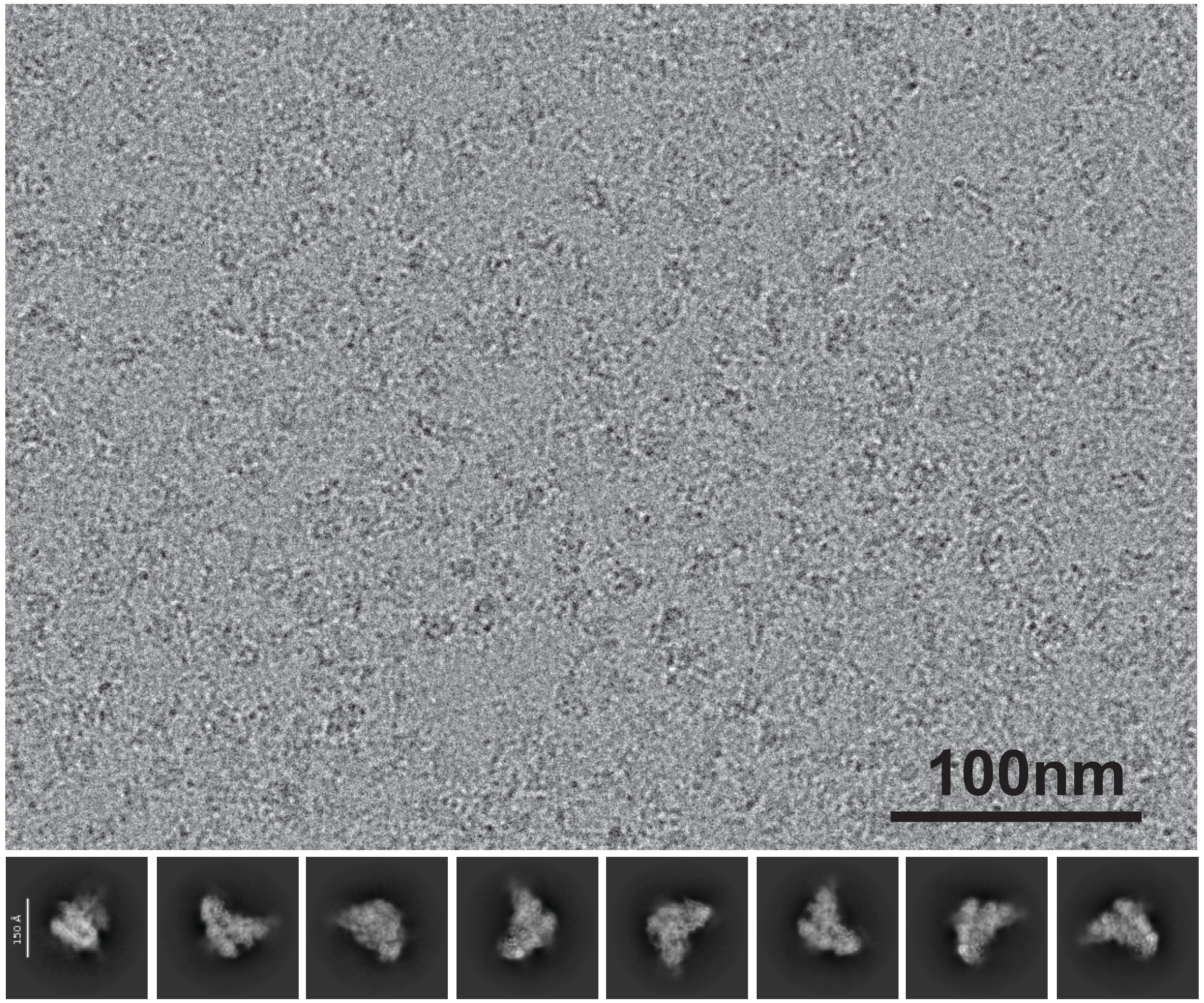
→

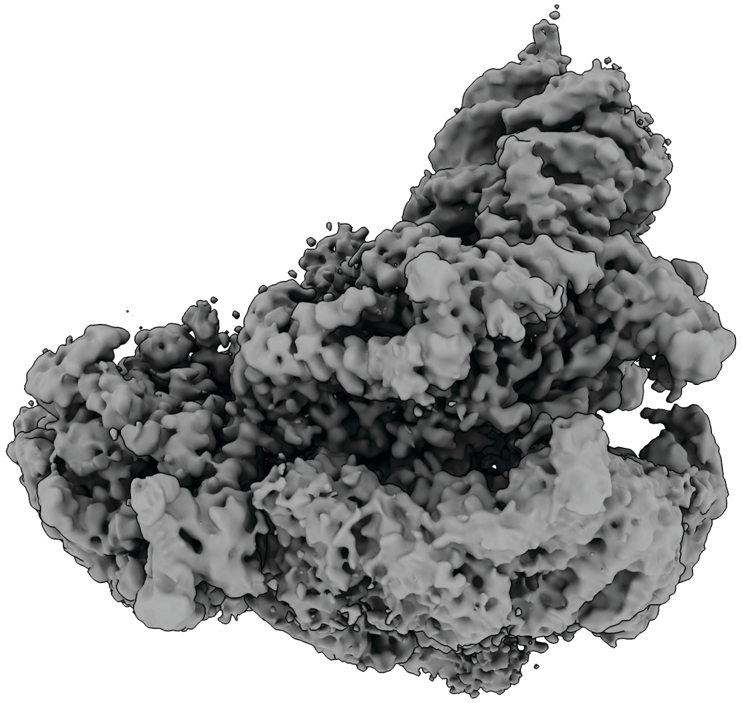
Graphene oxide grid freezing

30 sec wait time

15-30 sec blot time

Cryo-EM analysis

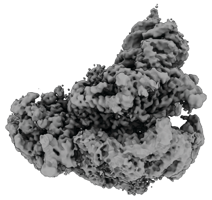




294,032 particles, 2.99 Å

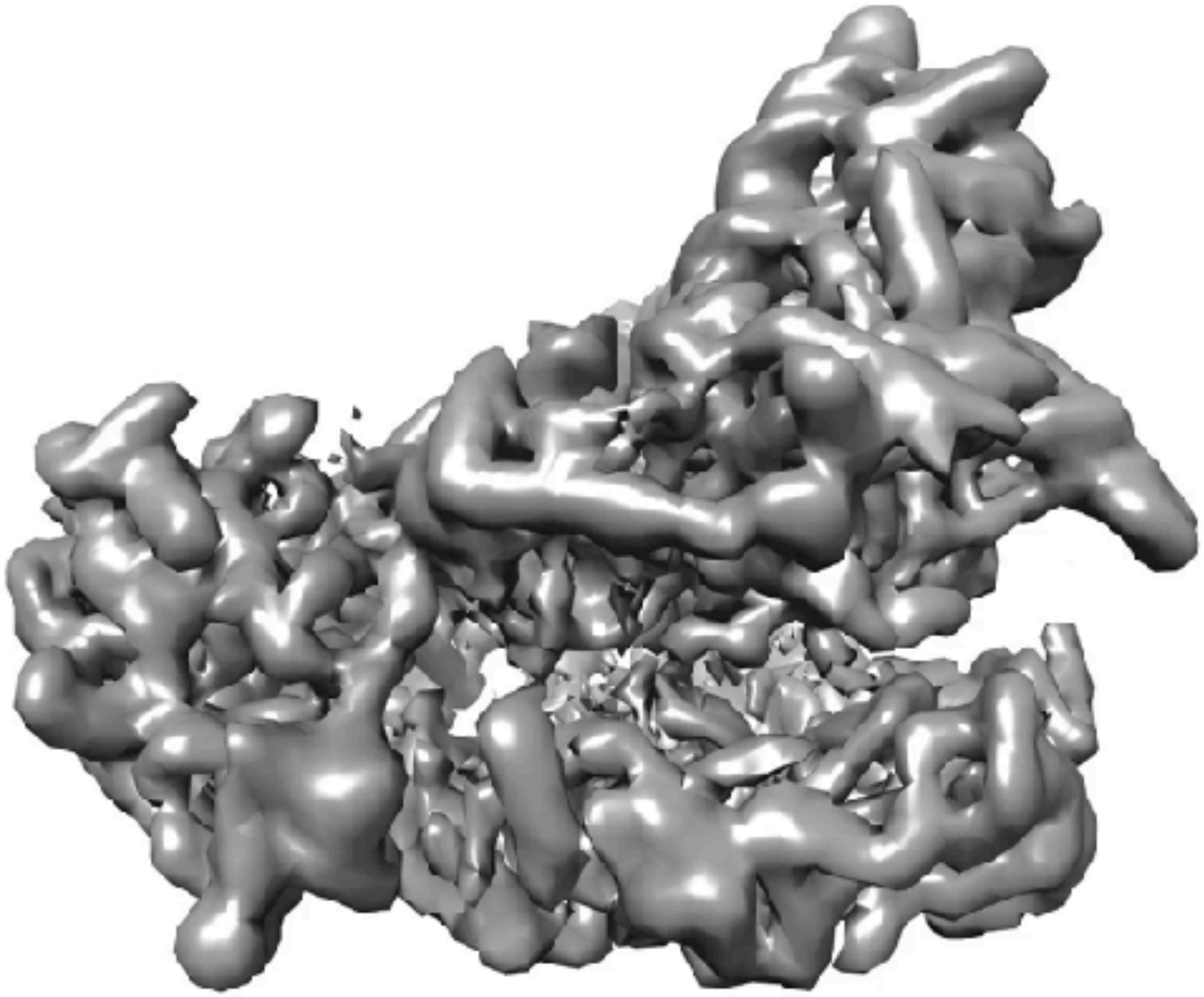
Kumar*, Allwein* et al., 2025, *Science*

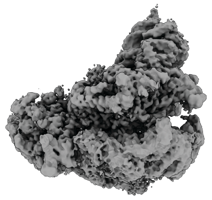
Cryo-EM analysis



294,032 particles, 2.99 Å

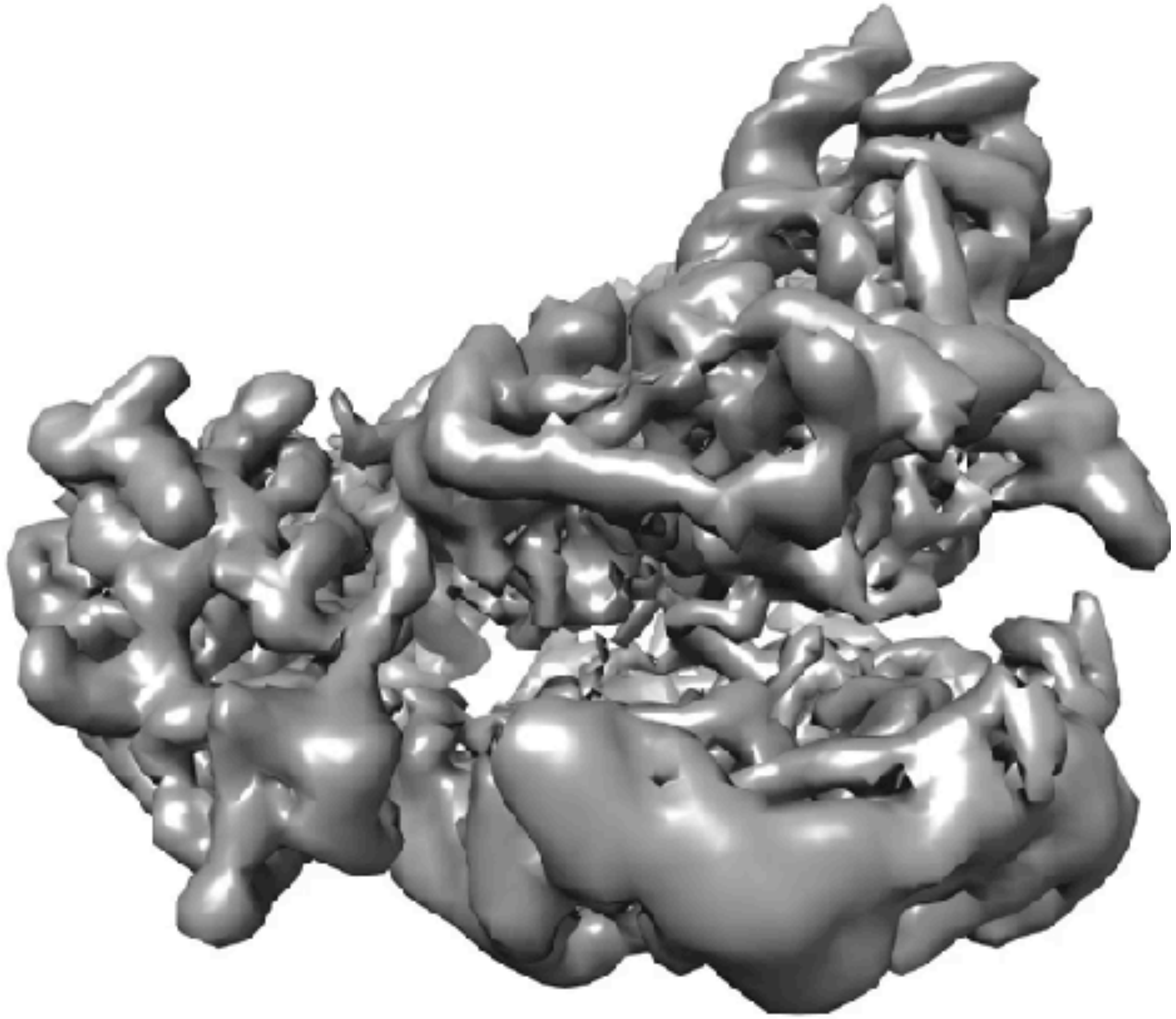
Cryo-EM analysis

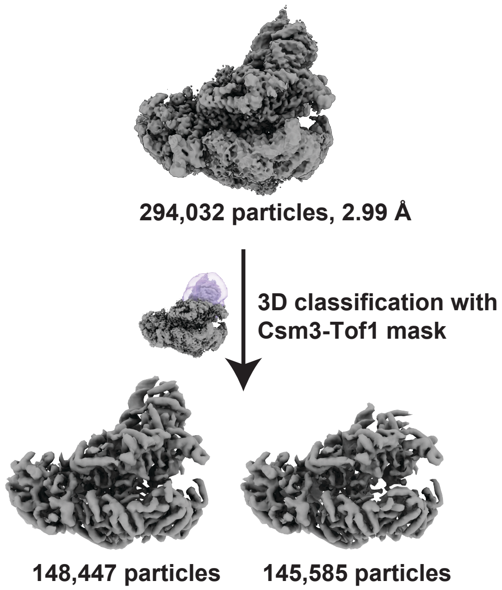




294,032 particles, 2.99 Å

Cryo-EM analysis





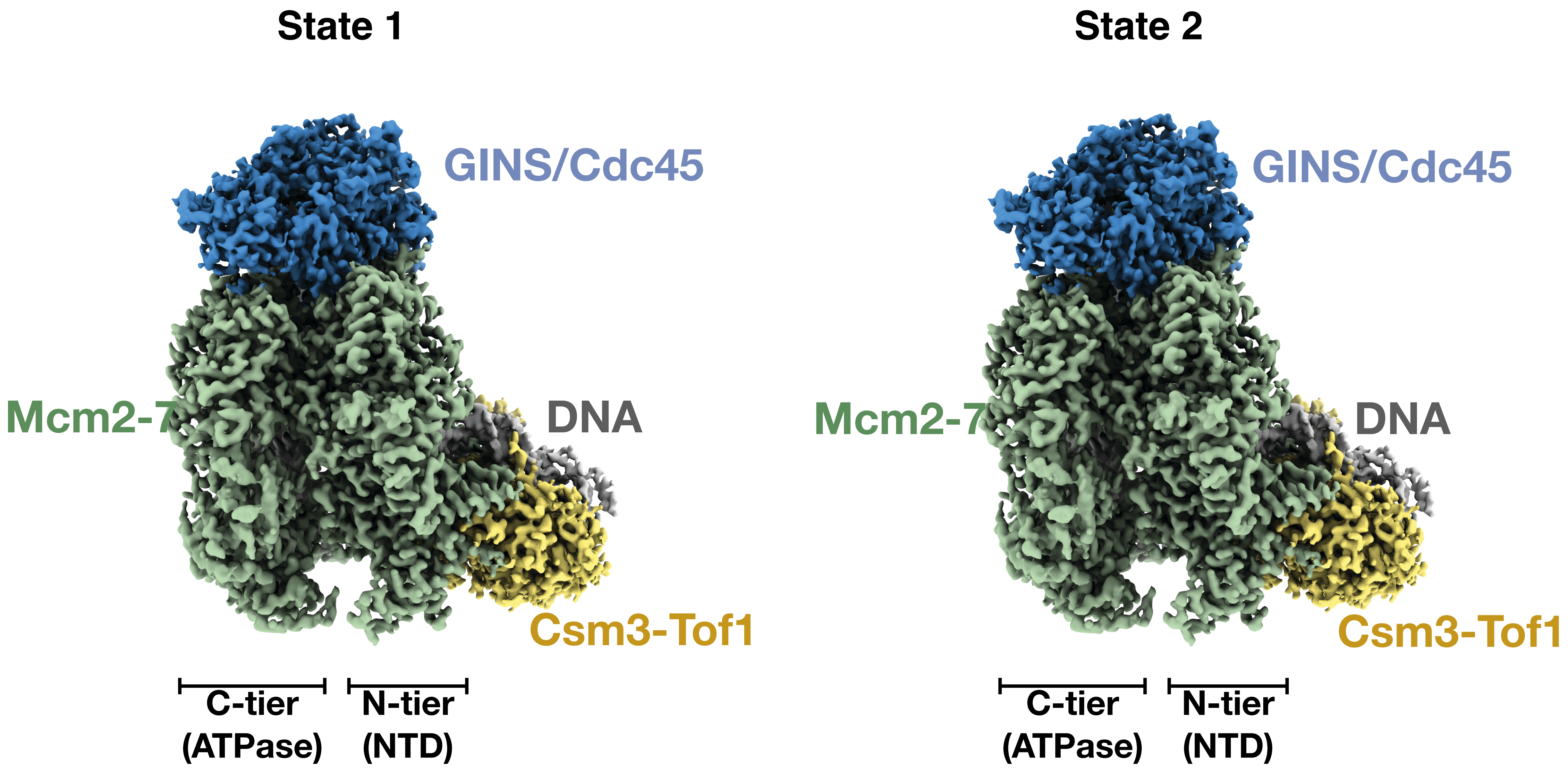
294,032 particles, 2.99 Å

3D classification with
Csm3-Tof1 mask

148,447 particles

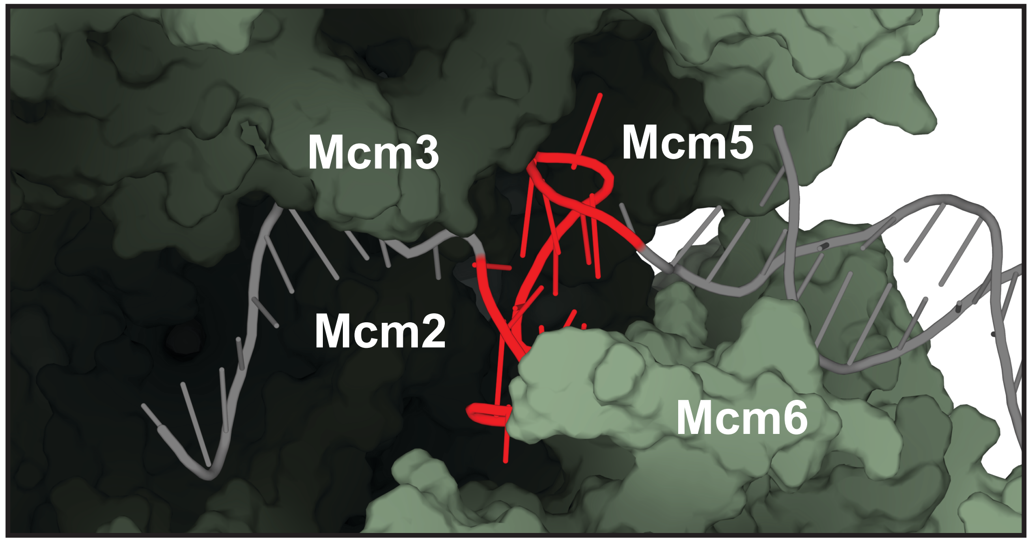
145,585 particles

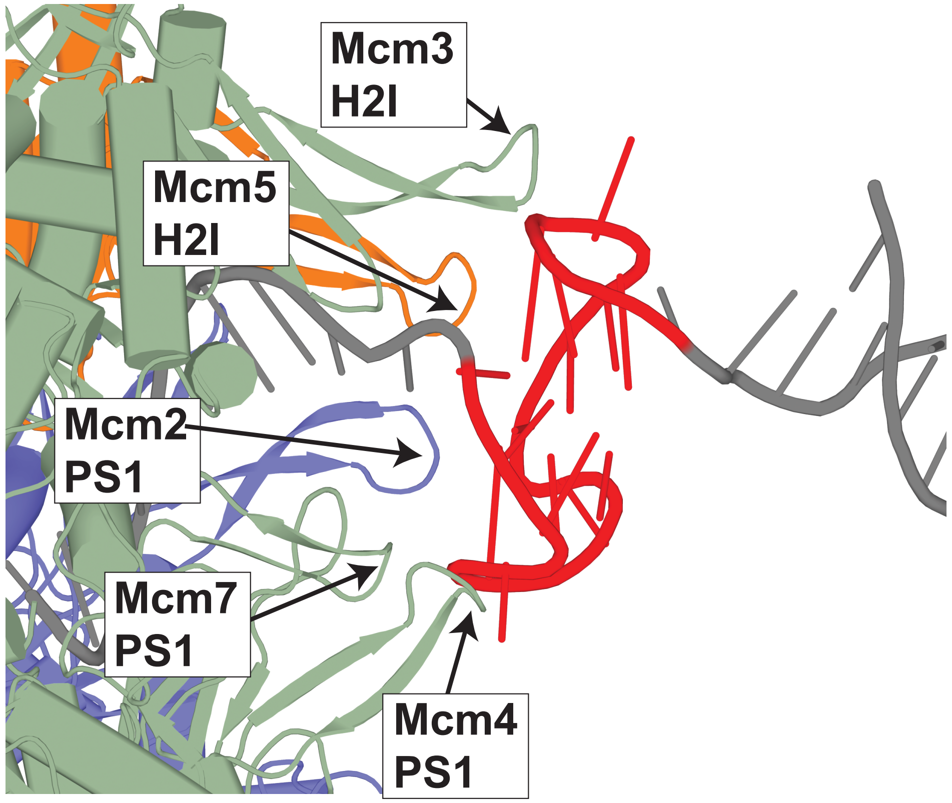
State 1 vs state 2



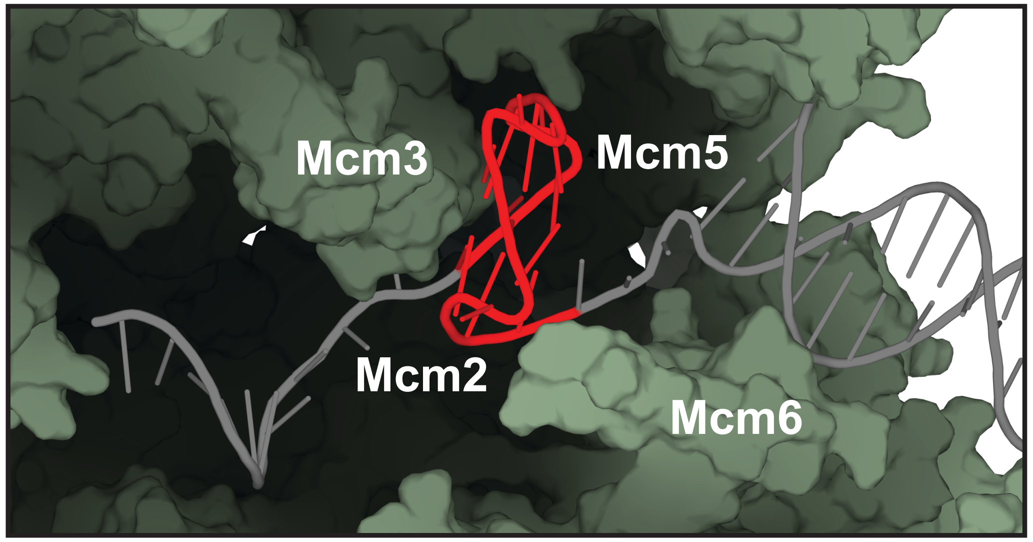
State 1 vs state 2

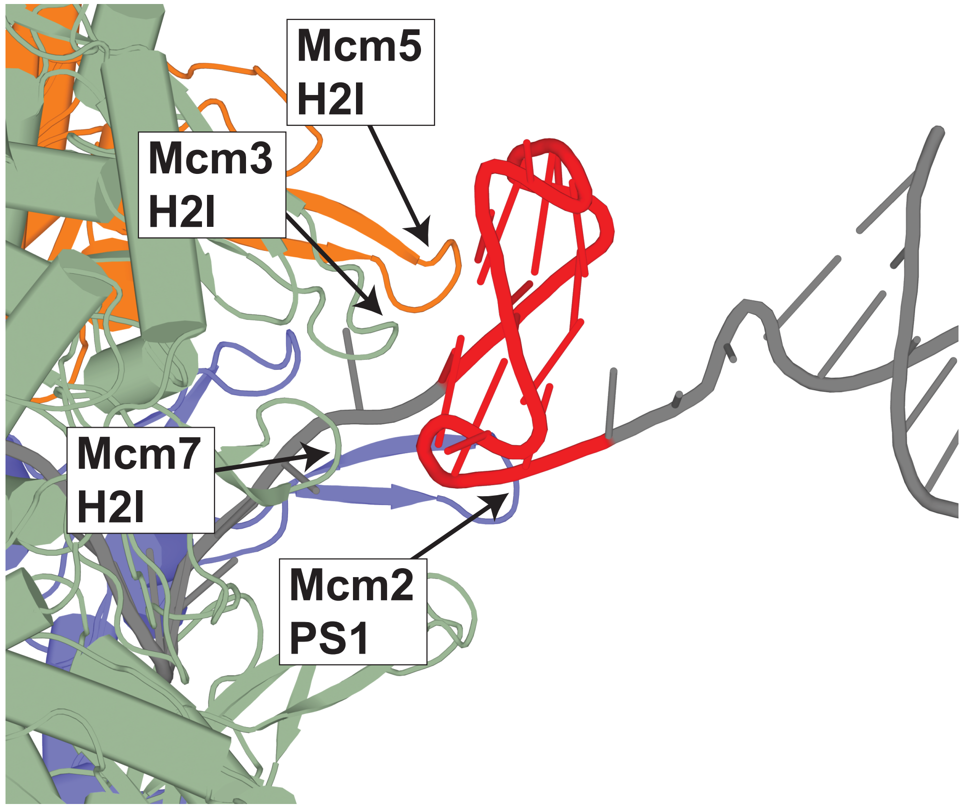
Stall state 1



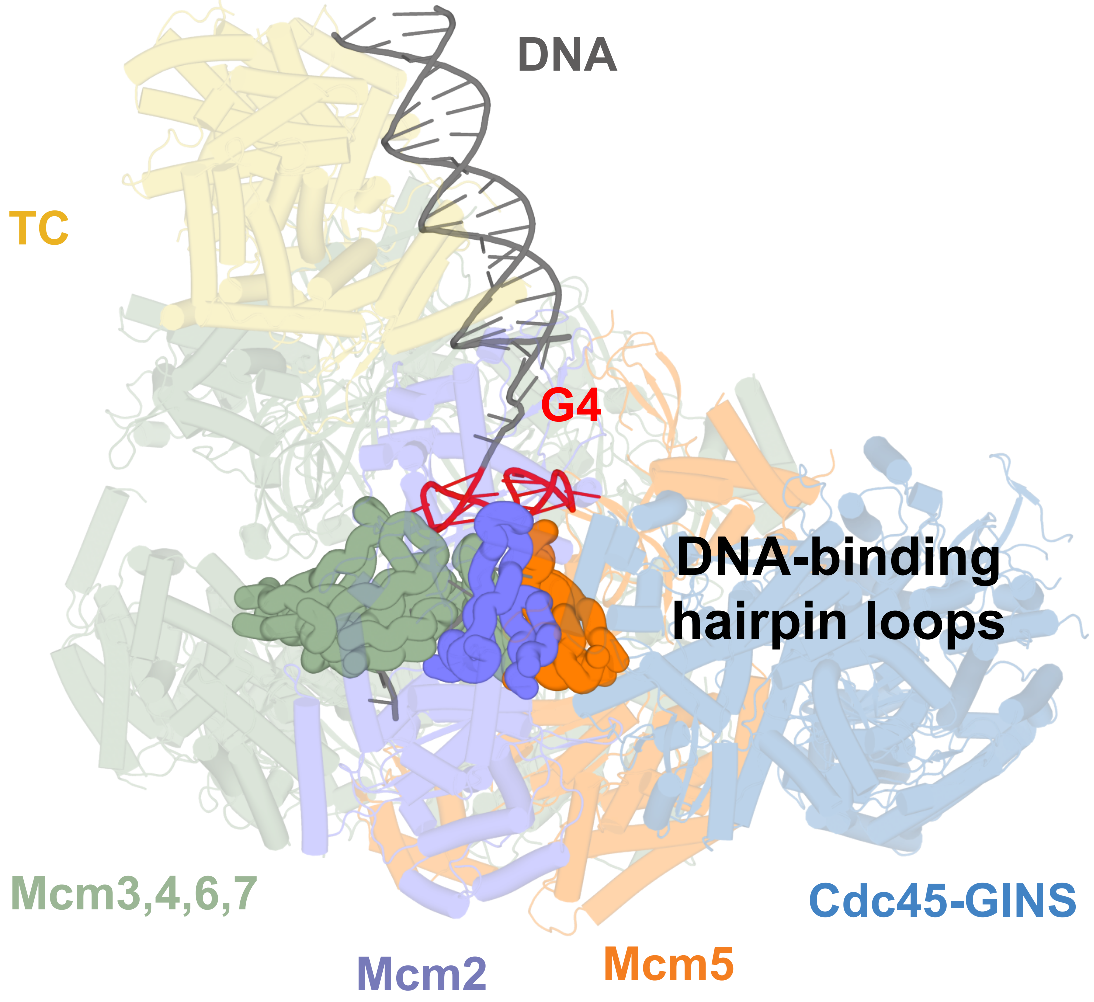


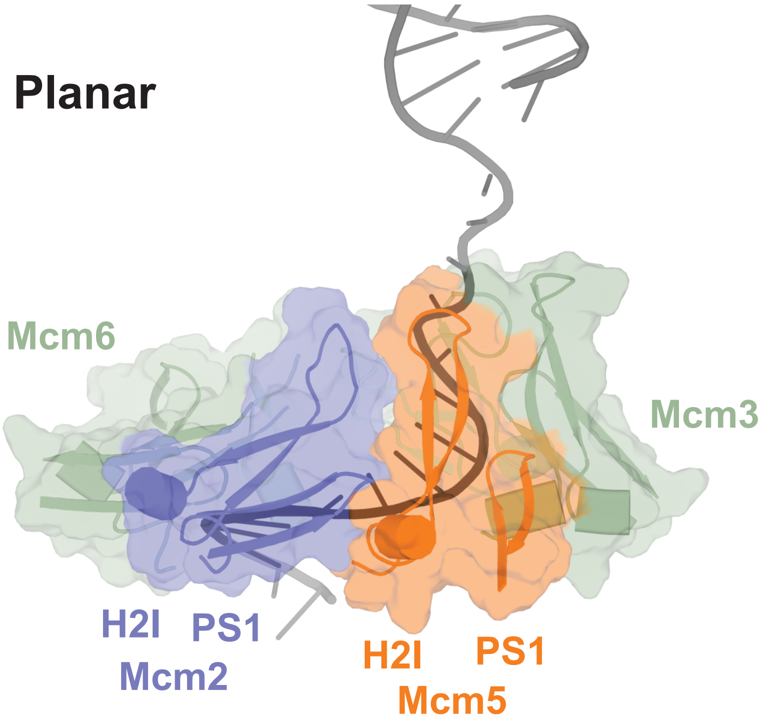
Stall state 2

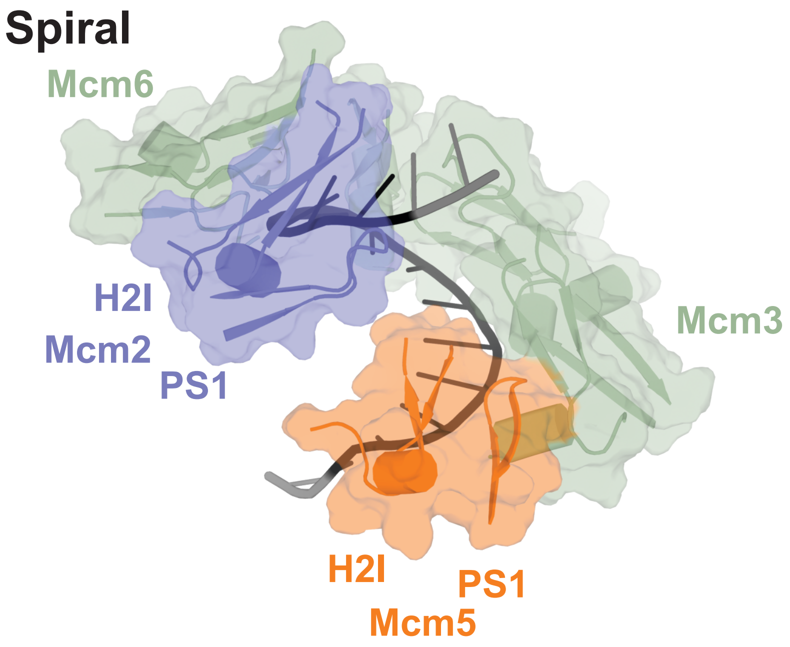




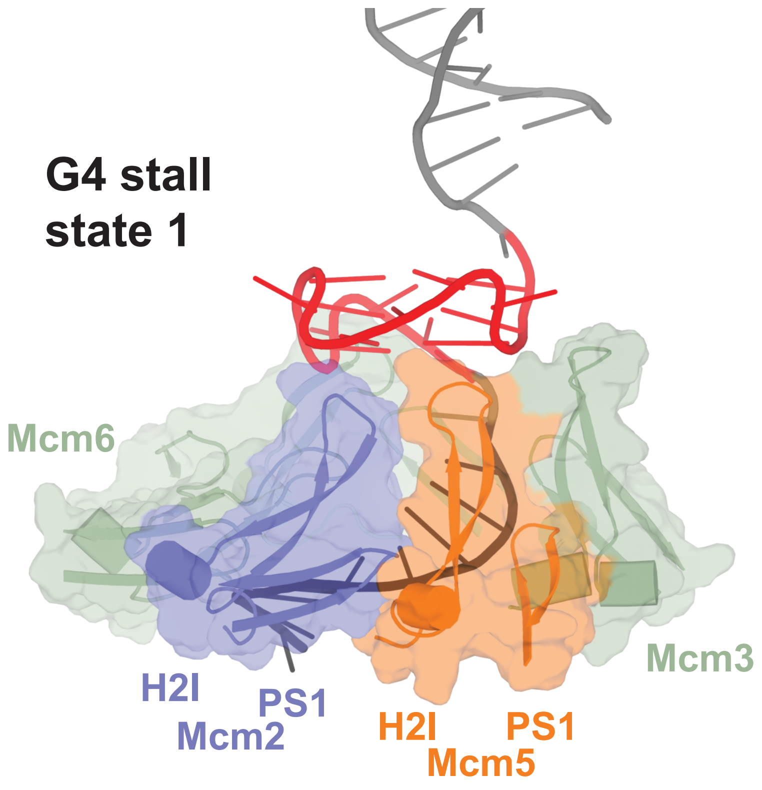
DNA translocation by CMG

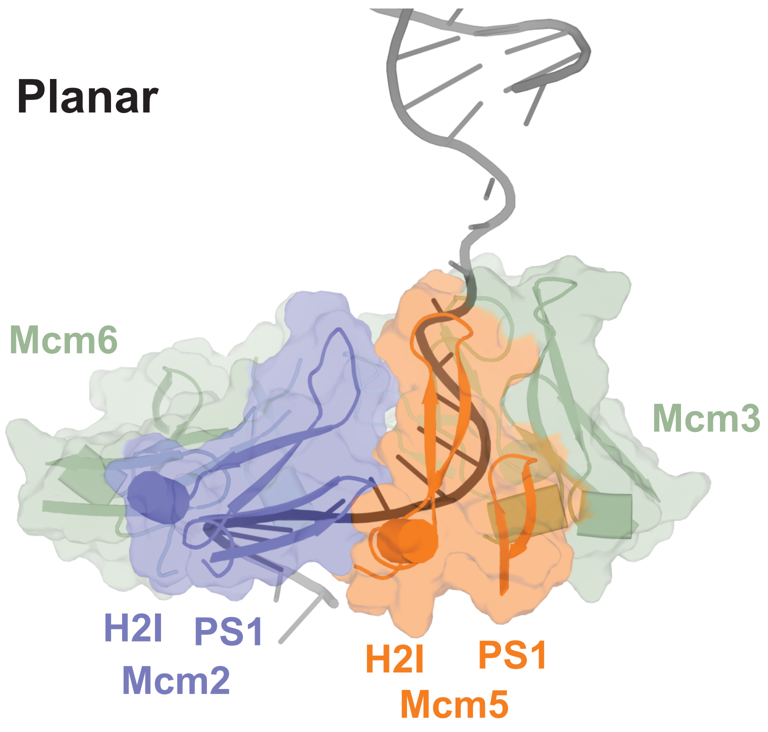


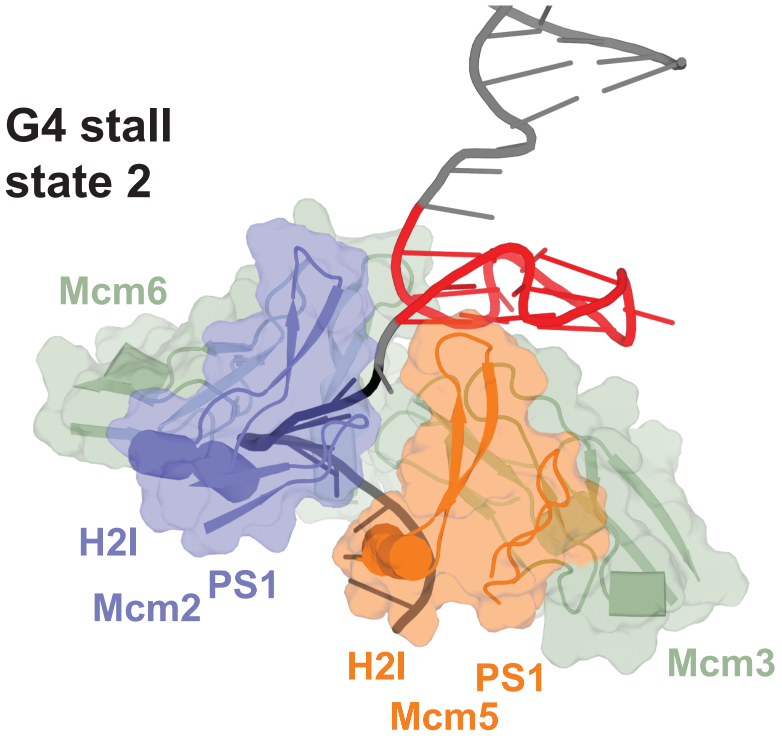


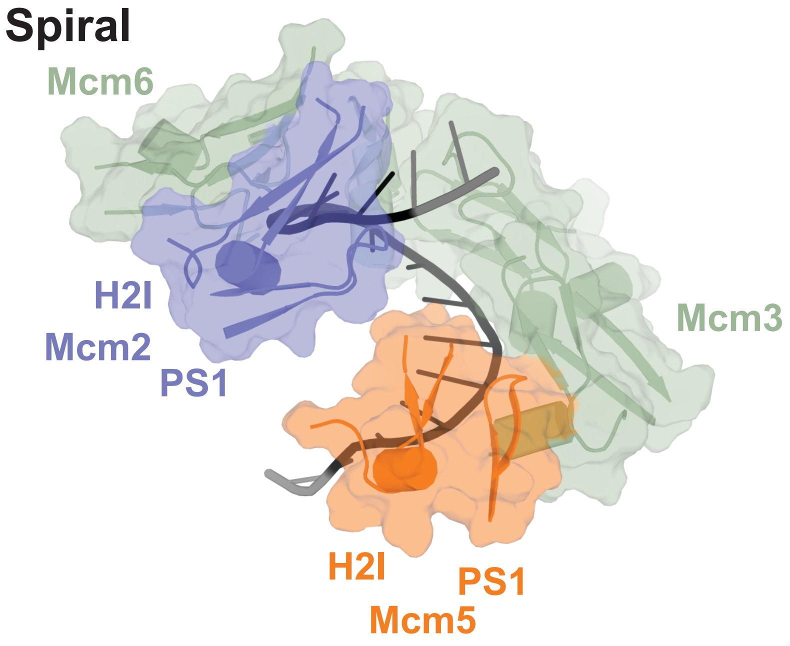


DNA translocation by CMG

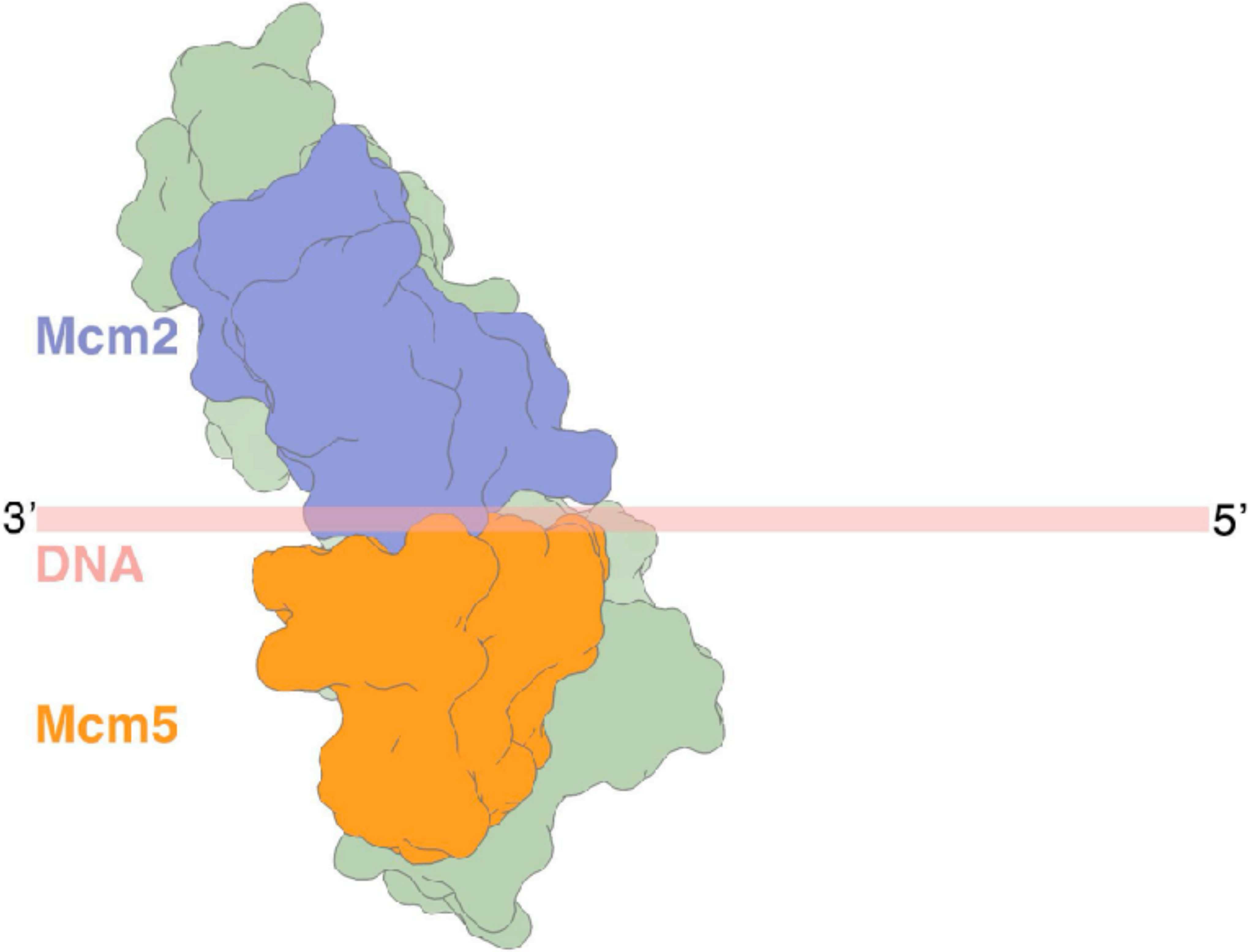

G4 stall state 1


Planar

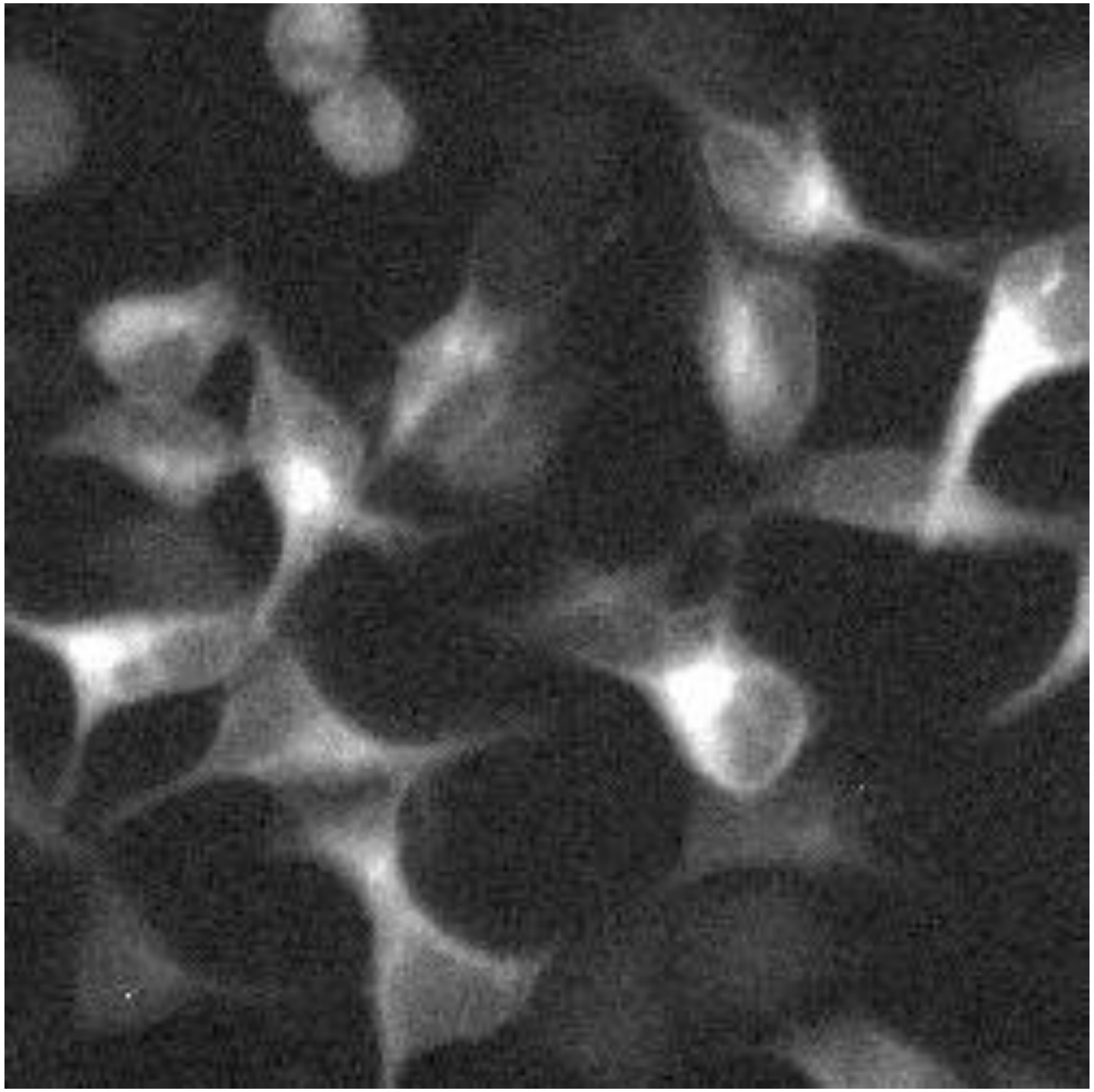

G4 stall state 2

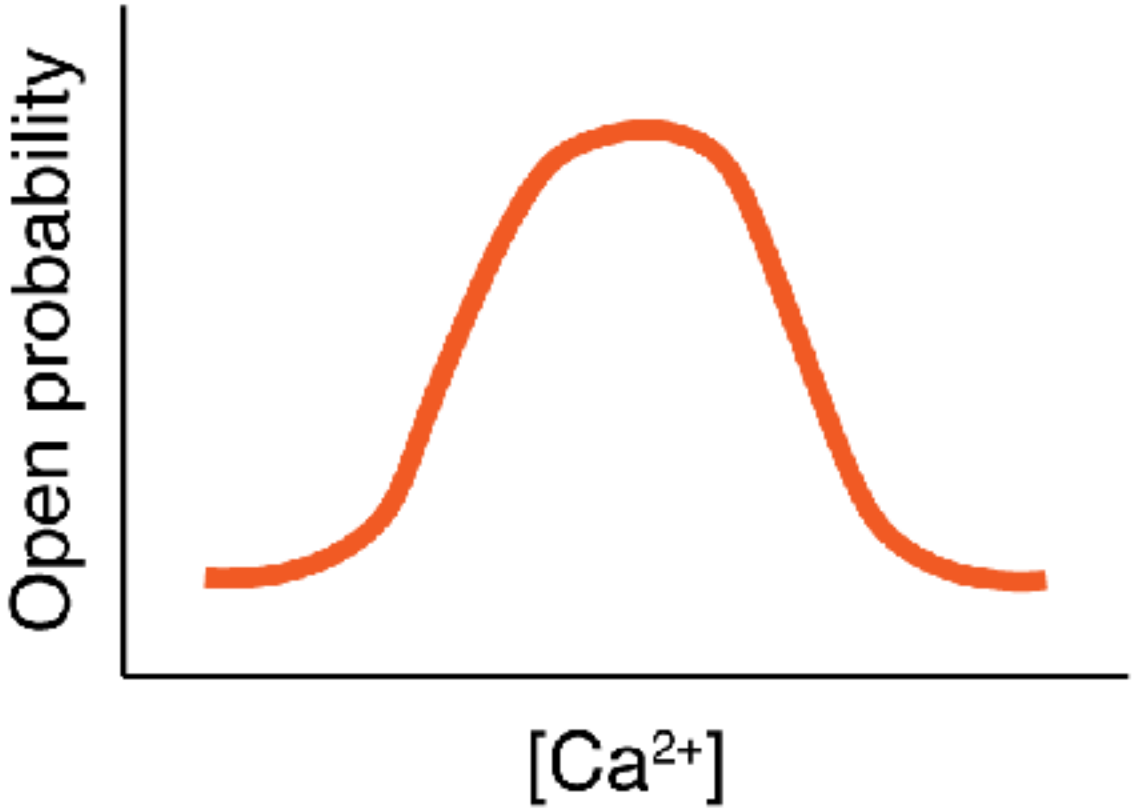

Spiral

Model for DNA translocation

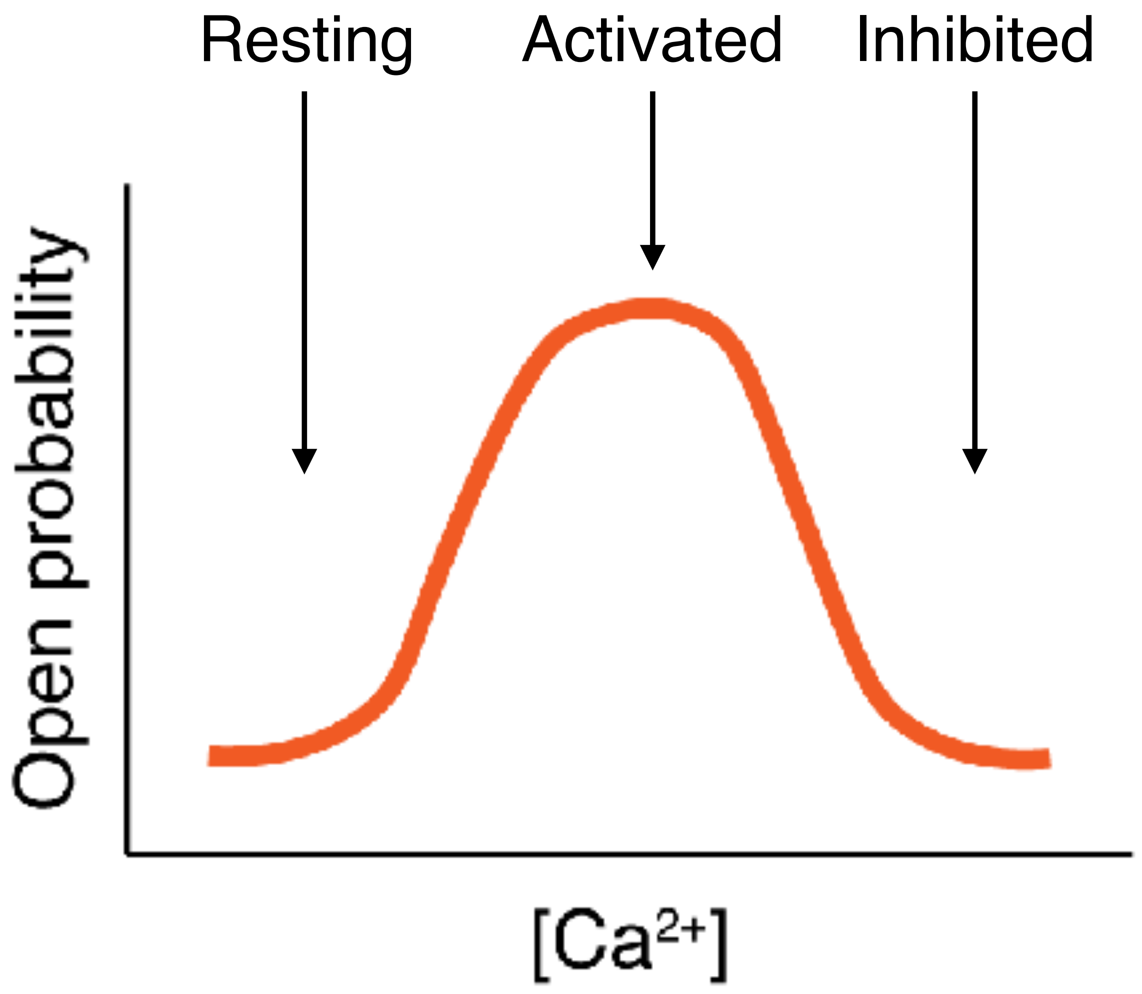


Case Study 2: Cytosolic Ca^{2+} signaling

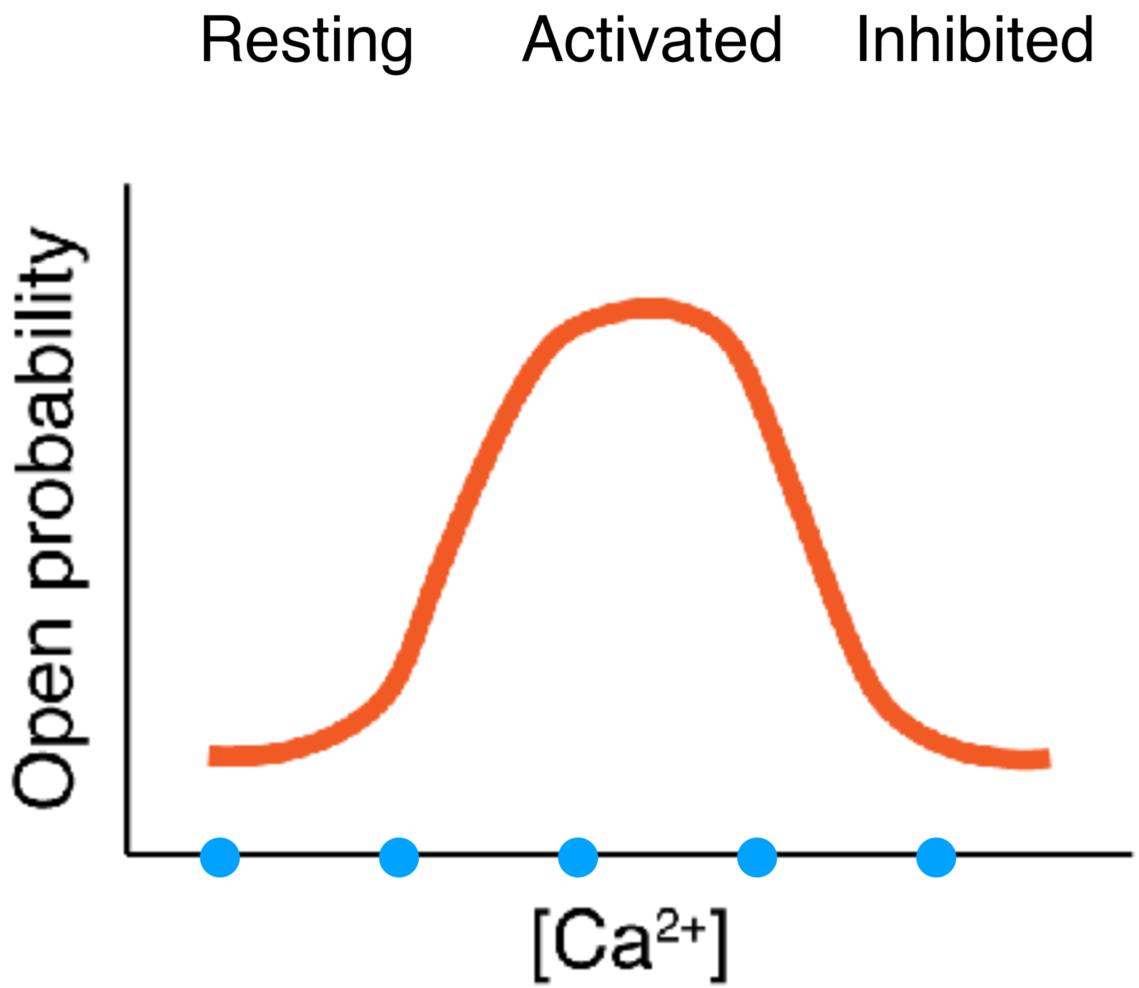


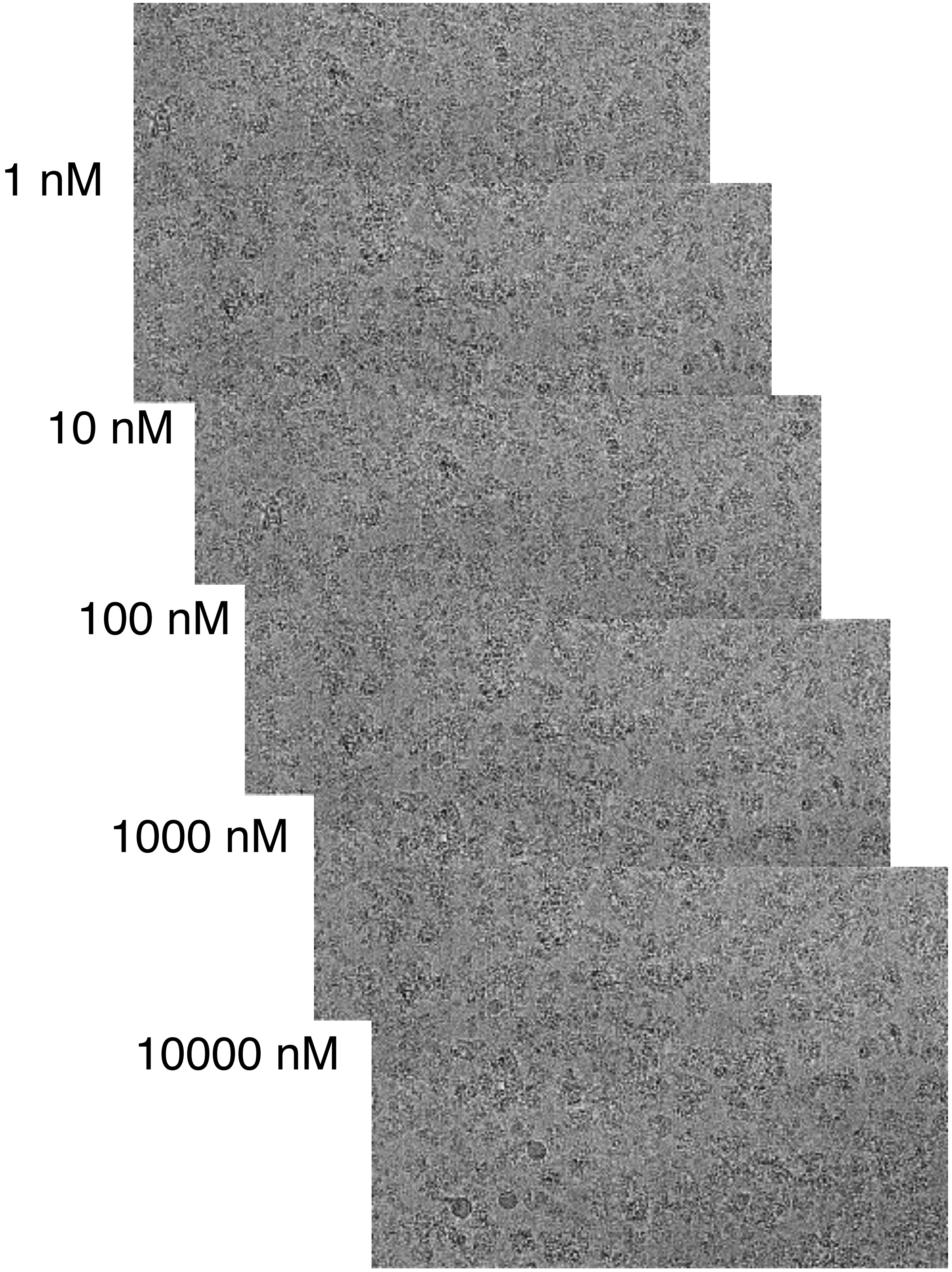


Visualizing ligand-dependent activation and inhibition

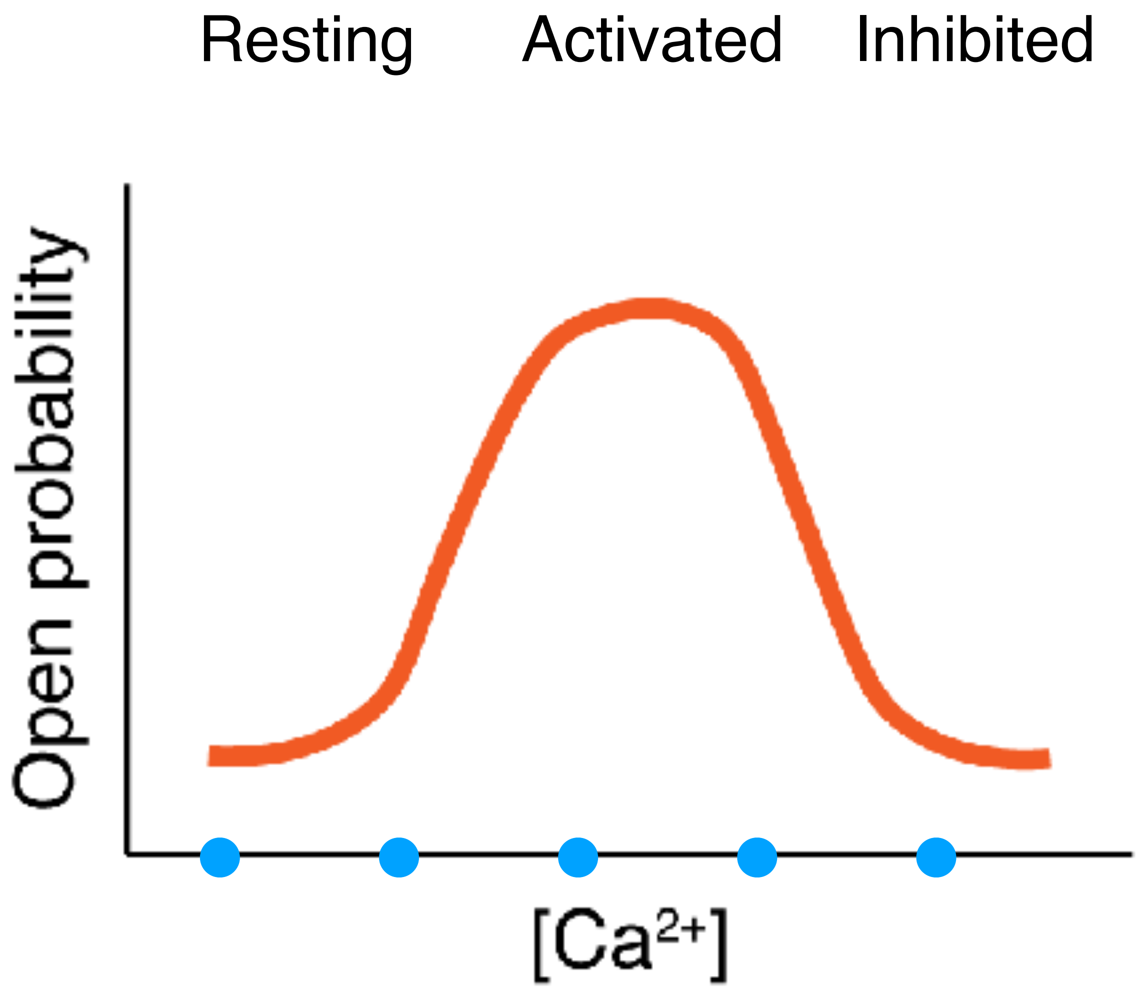


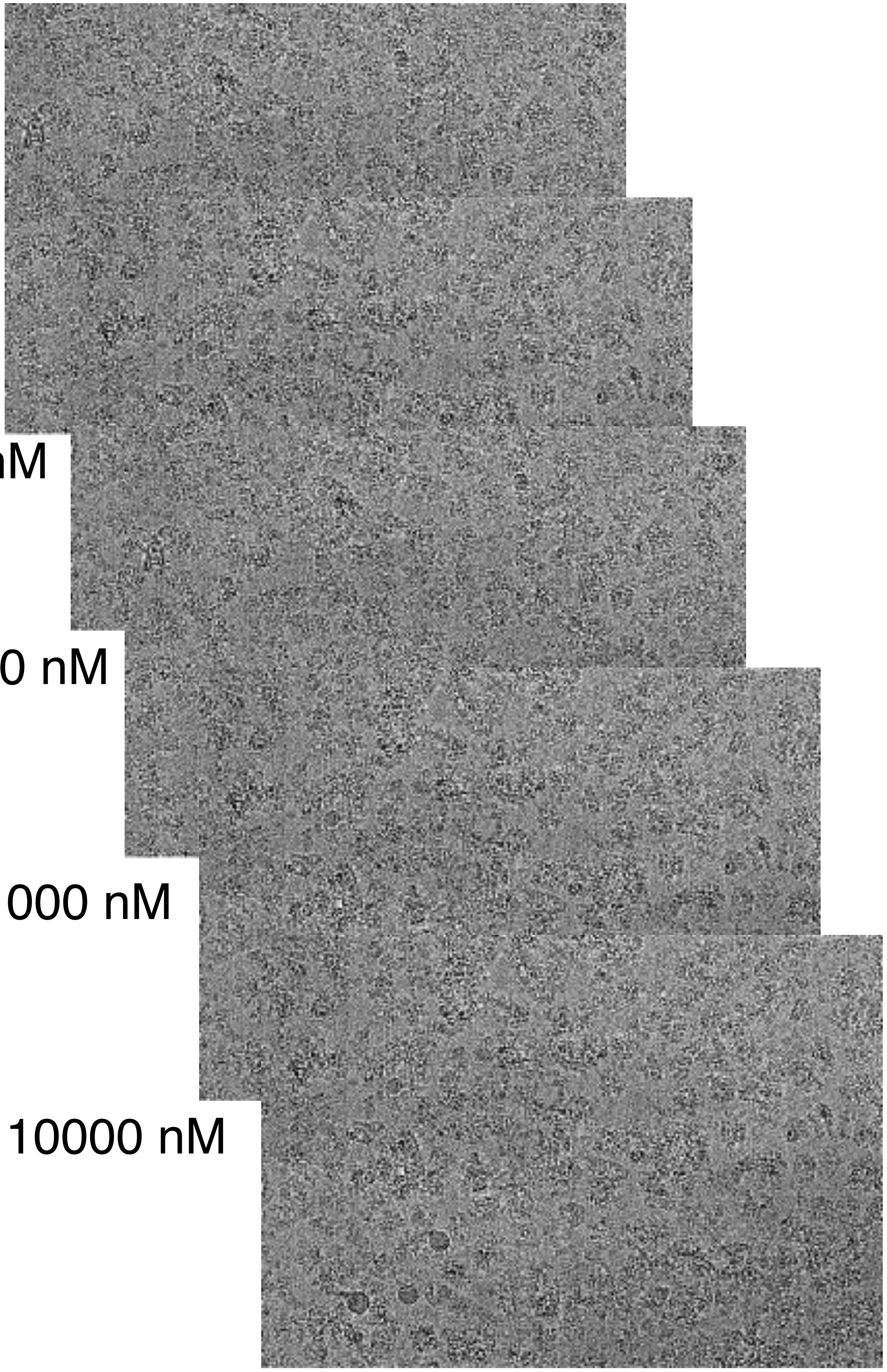
Visualizing ligand-dependent activation and inhibition





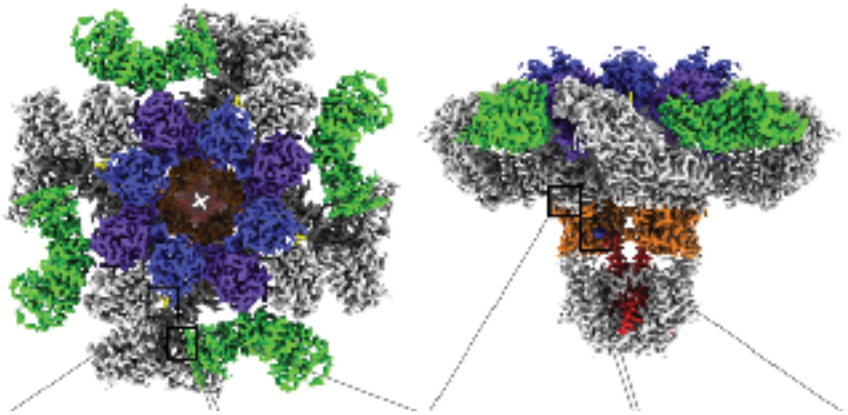
Visualizing ligand-dependent activation and inhibition



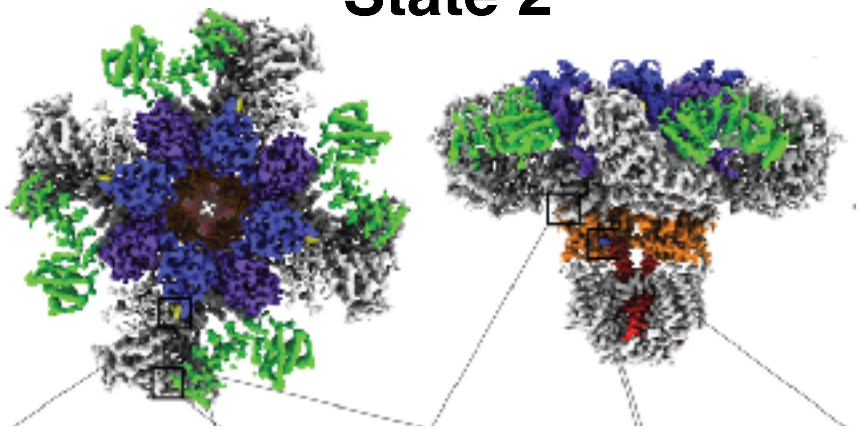


Interpreting the classes from the titration

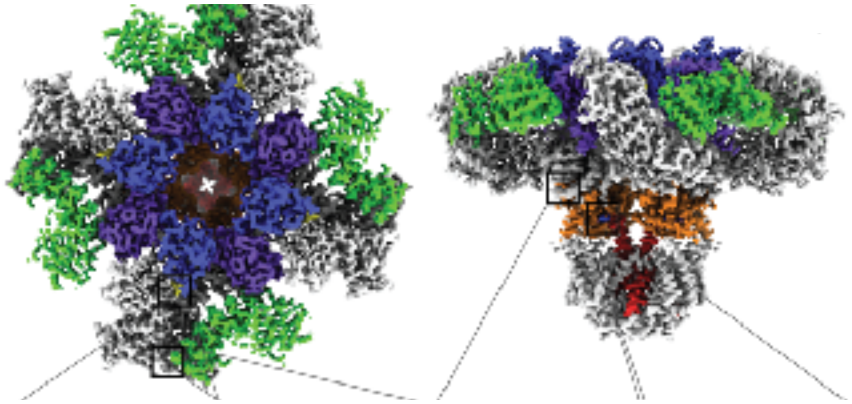
State 1



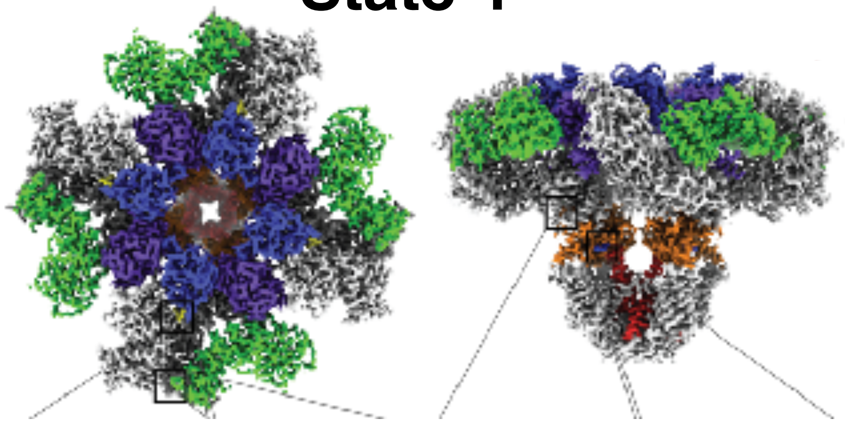
State 2



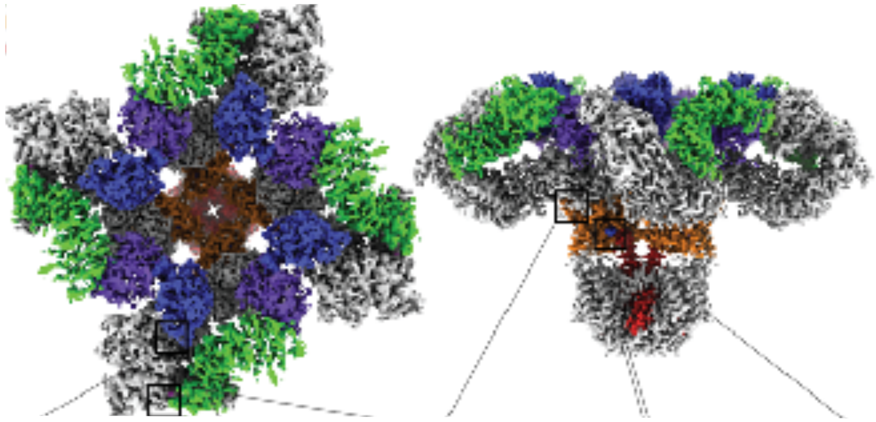
State 3



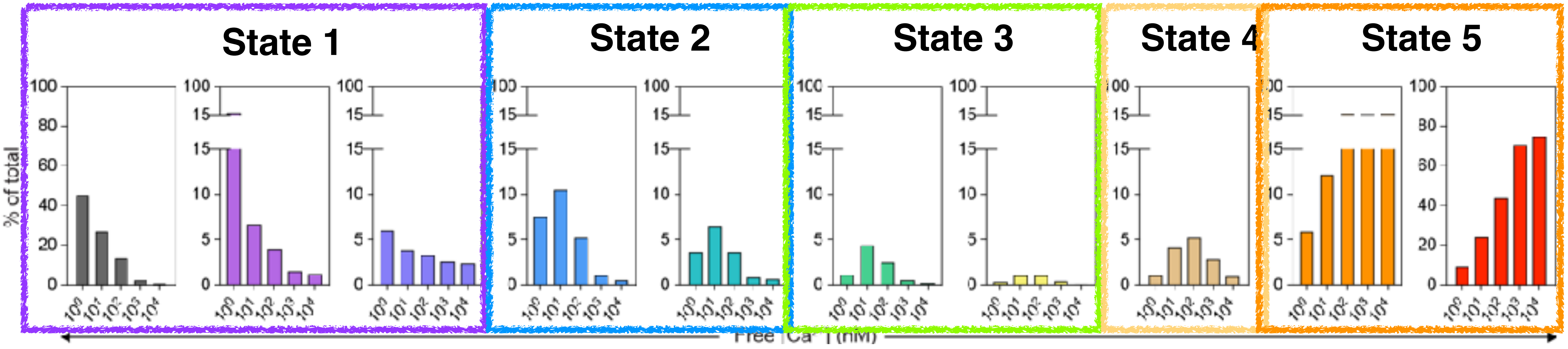
State 4



State 5



Ca^{2+} dependence of conformational state



Ca2+ dependence of conformational state

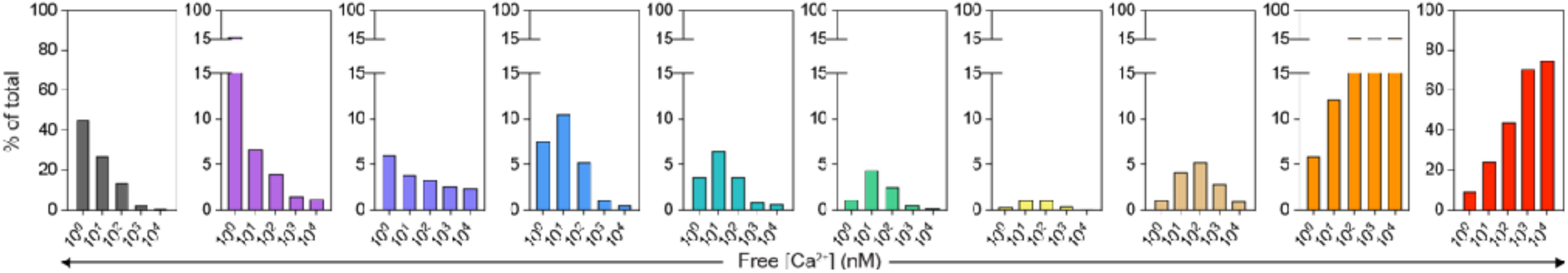
Resting **State 1**

State 2

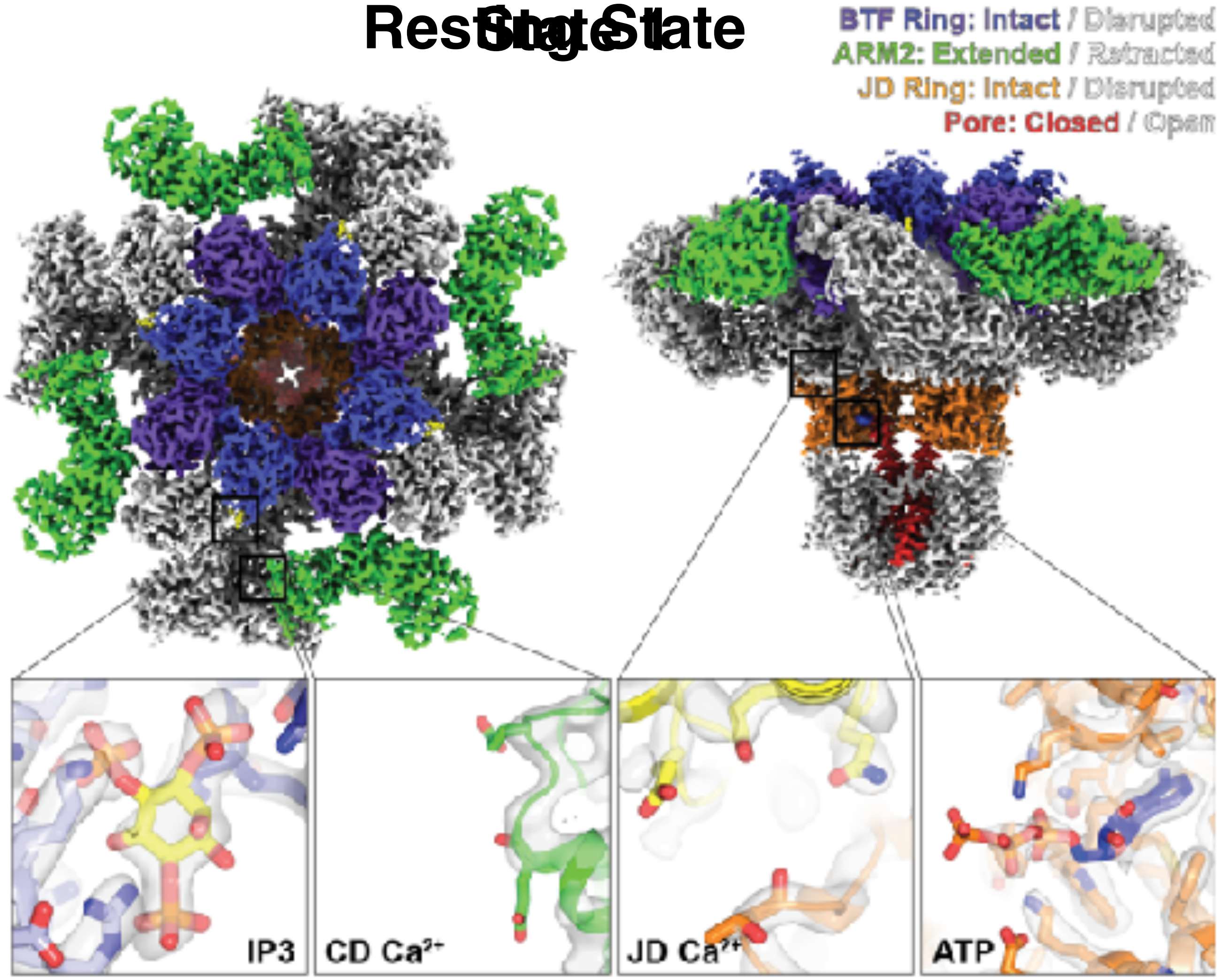
State 3

State 4

State 5

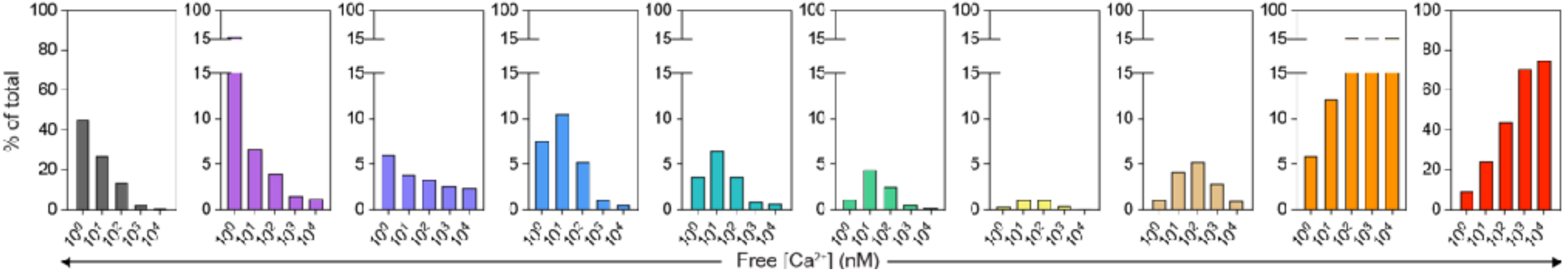


Resting State **State 1**

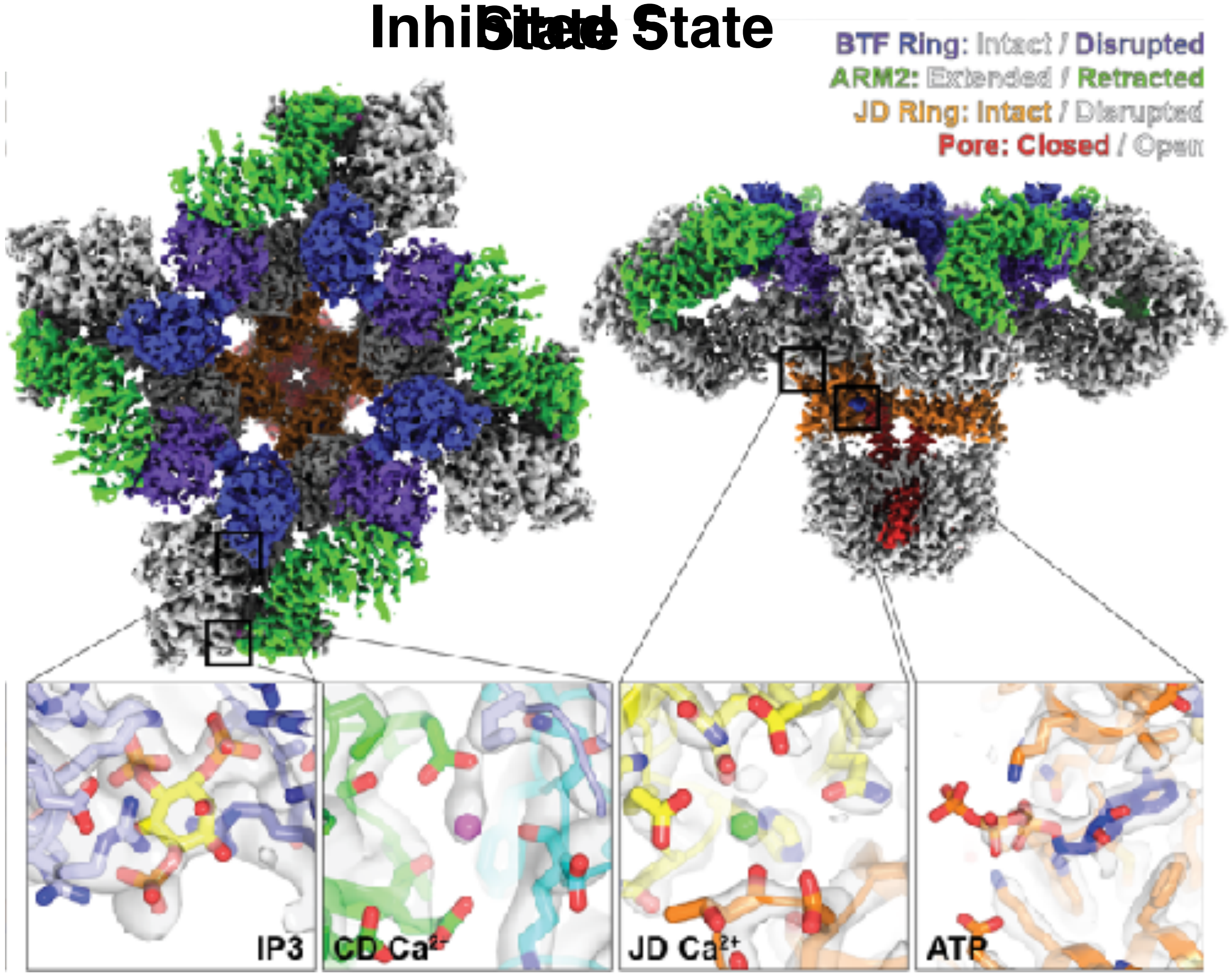


Ca^{2+} dependence of conformational state

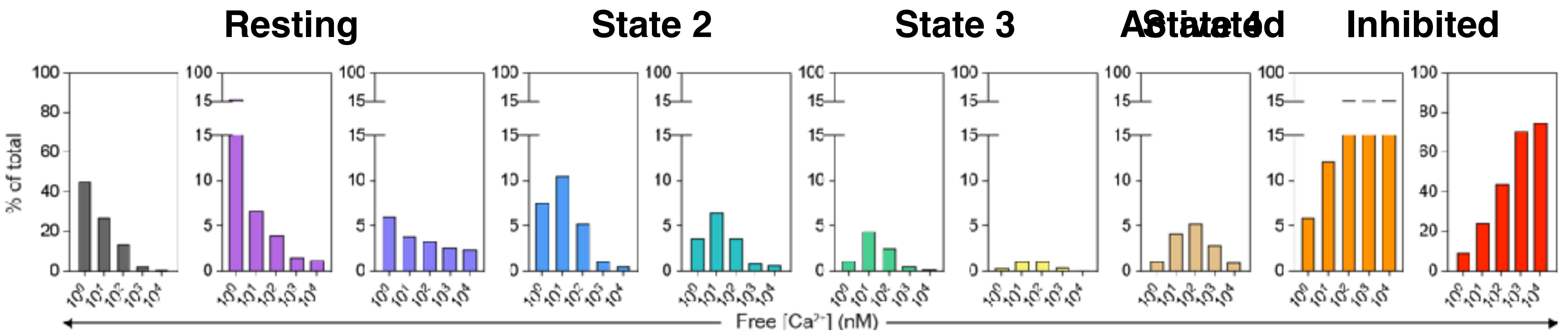
Resting **State 2** **State 3** **State 4** **Inhibited** **State 5**



Inhibited State **State 5**

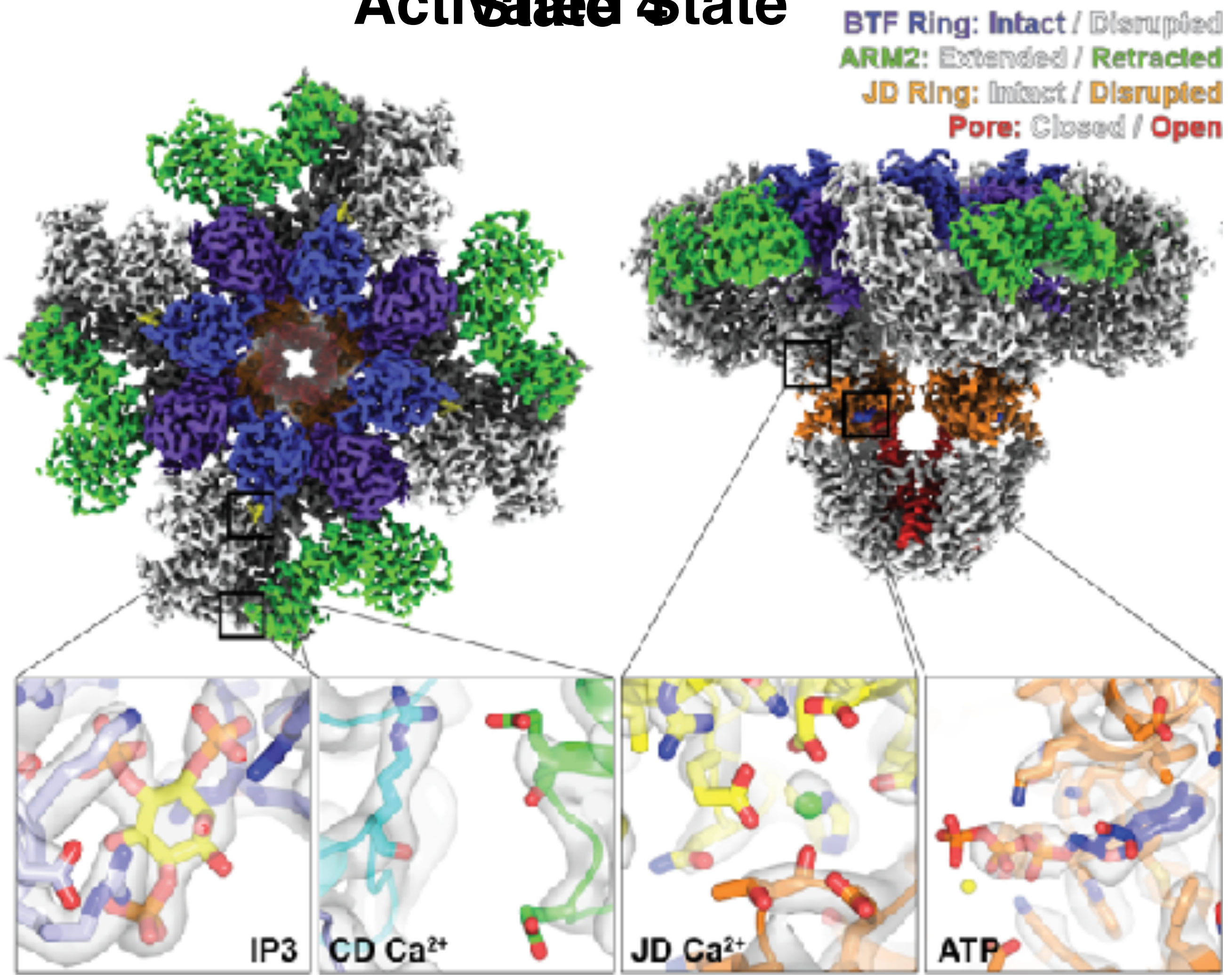


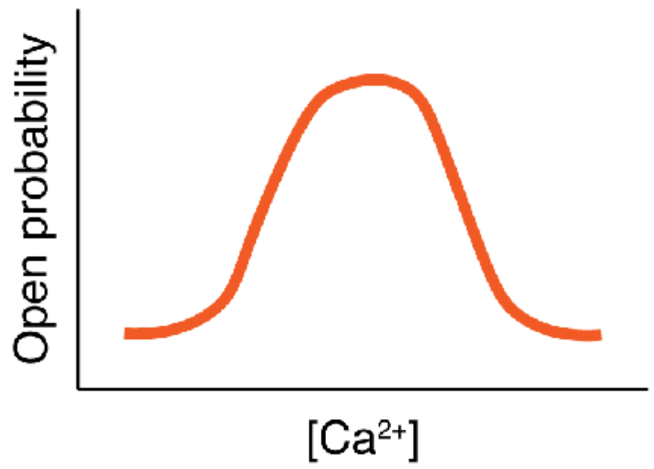
Ca2+ dependence of conformational state



Activated State

State 4





Ca^{2+} dependence of conformational state

Resting

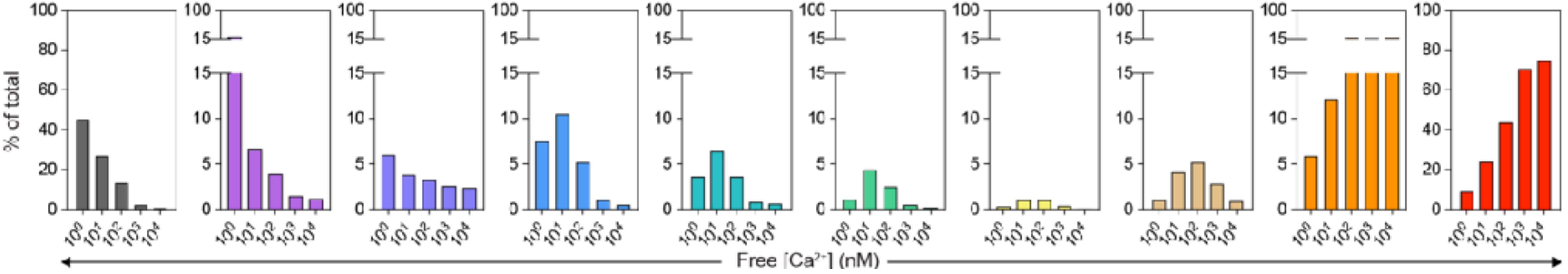
Preactivated

State 2

State 3

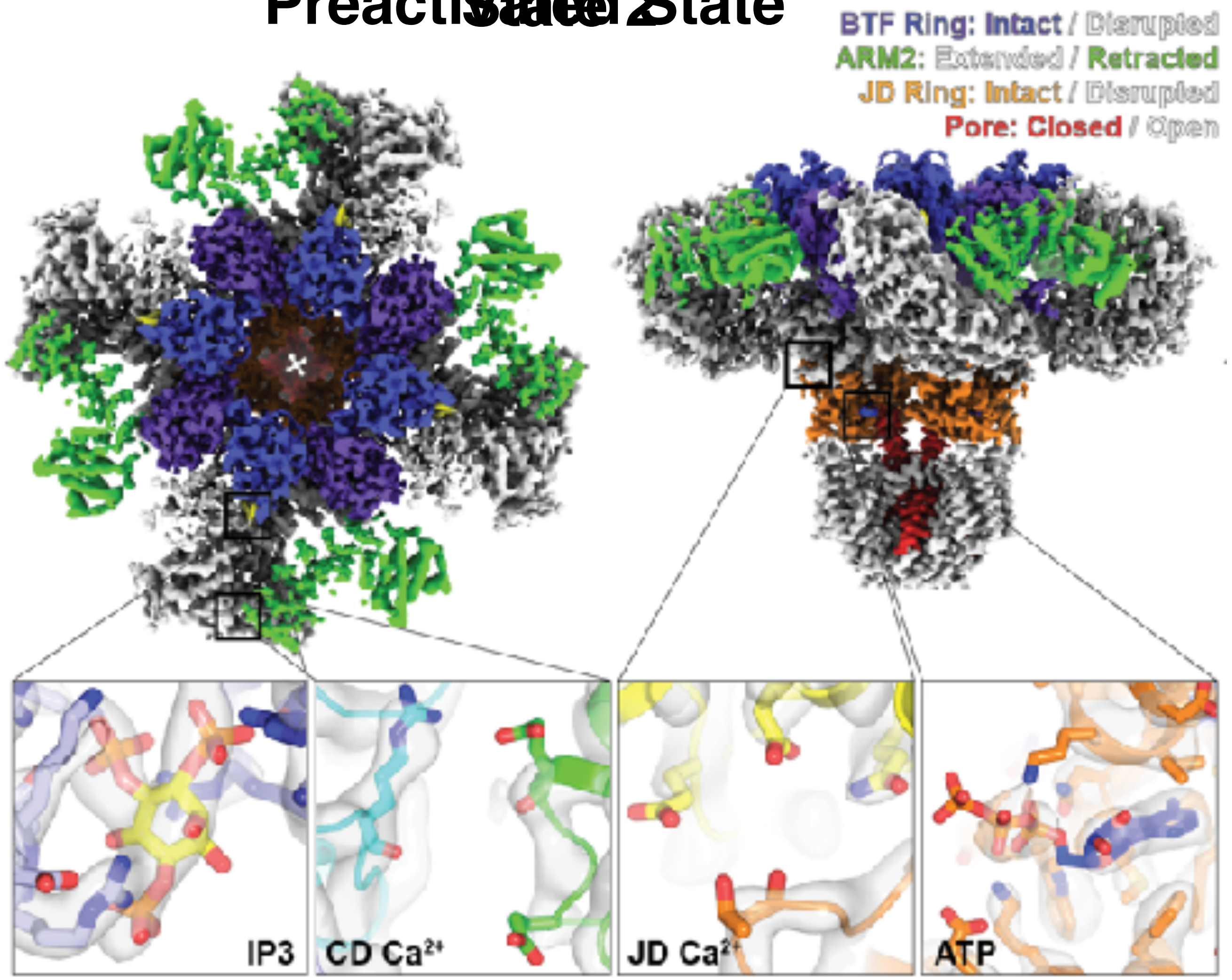
Activated

Inhibited



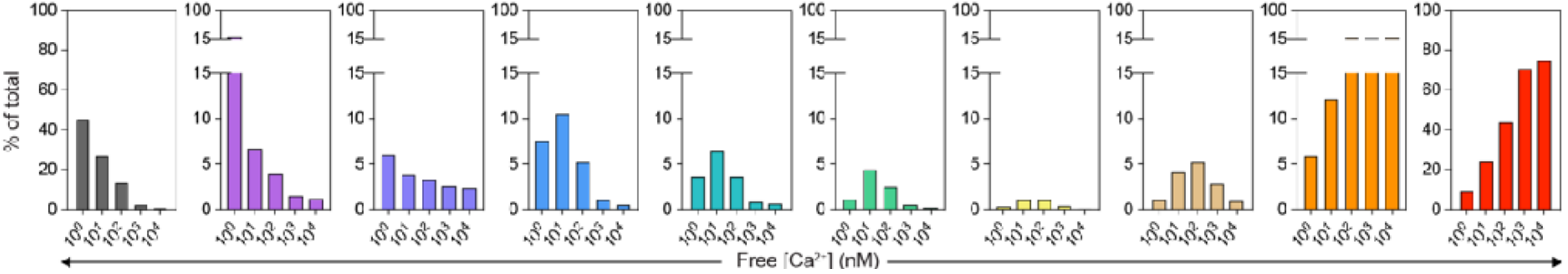
Preactivated State

State 2



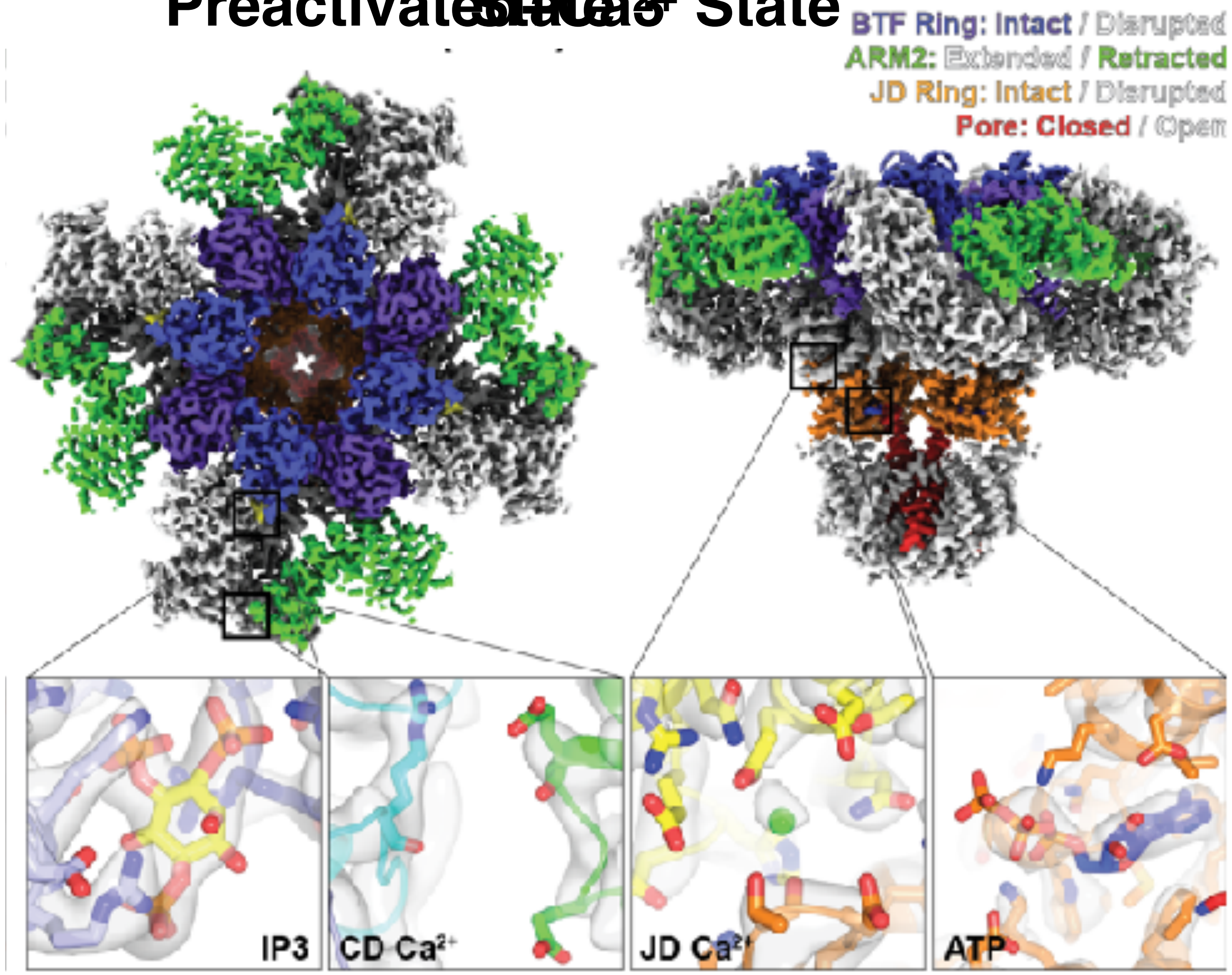
Ca^{2+} dependence of conformational state

Resting **Preactivated** **Preactivated+Ca^{2+}** **State 1** **Activated** **Inhibited**



Preactivated+Ca^{2+} State

State 3



Resources

- Cryo-EM databases
  - Electron microscopy data bank - all published EM structures are deposited and released upon publication
  - Protein data bank - all atomic structures of proteins (crystallography, NMR and EM) are deposited and released upon publication
- Cryo-EM online courses
  - CalTech Getting Started in Cryo-EM http://cryo-em-course.caltech.edu/videos
  - LMB Cryo-EM course 2017 https://www2.mrc-lmb.cam.ac.uk/research/scientific-training/electron-microscopy/
- NYC resources
  - NYSBC Simons Electron Microscopy winter-spring course - offered annually with distinguished lecturers from New York City and beyond http://semc.nysbc.org/the-winter-spring-2020-em-course/