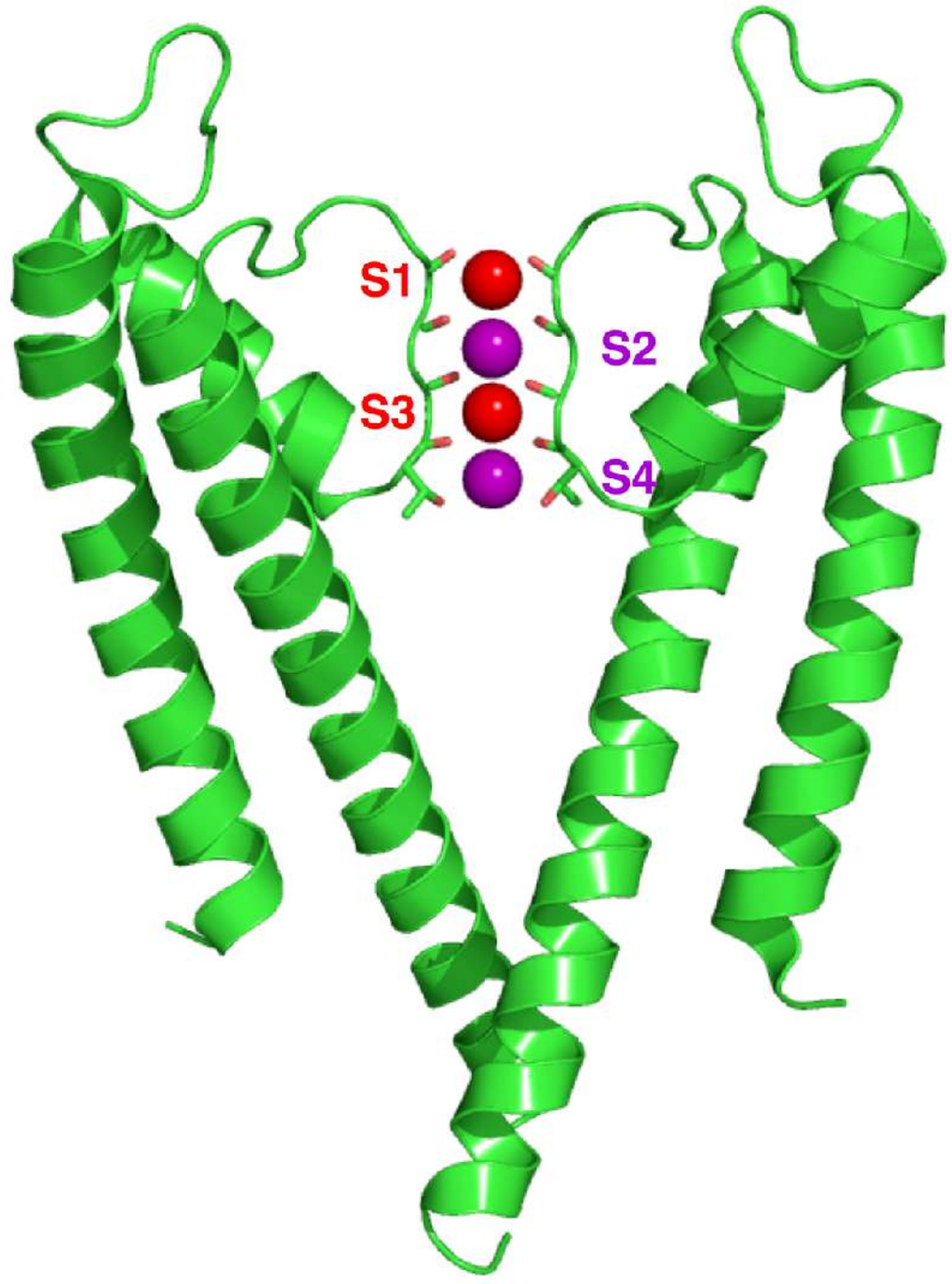
Interpretation and limitation of Single Particle Cryo-EM analysis

Rich Hite
Memorial Sloan Kettering Cancer Center

Overview

1. What is structural biology?
2. Why is cryo-electron microscopy currently so popular among structural biologists?
3. What advances have allowed cryo-electron microscopy to become so widely adopted?
4. What are the limitations of Cryo-EM?



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

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

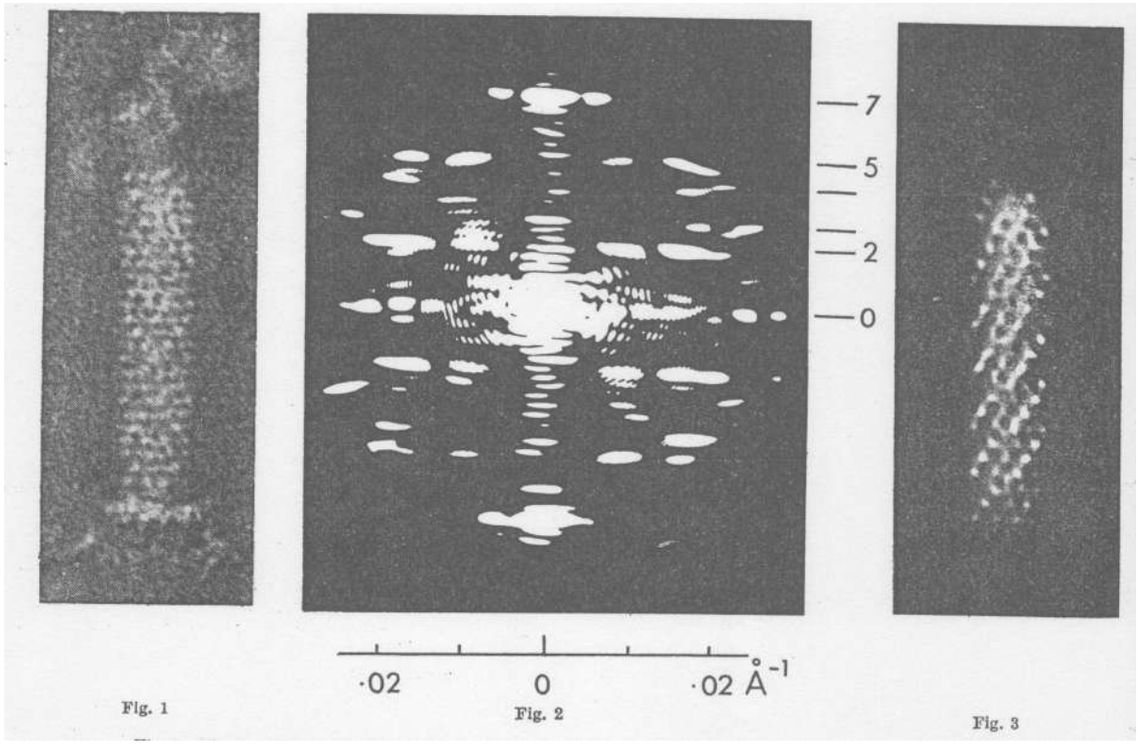


Fig. 1

Fig. 2

Fig. 3

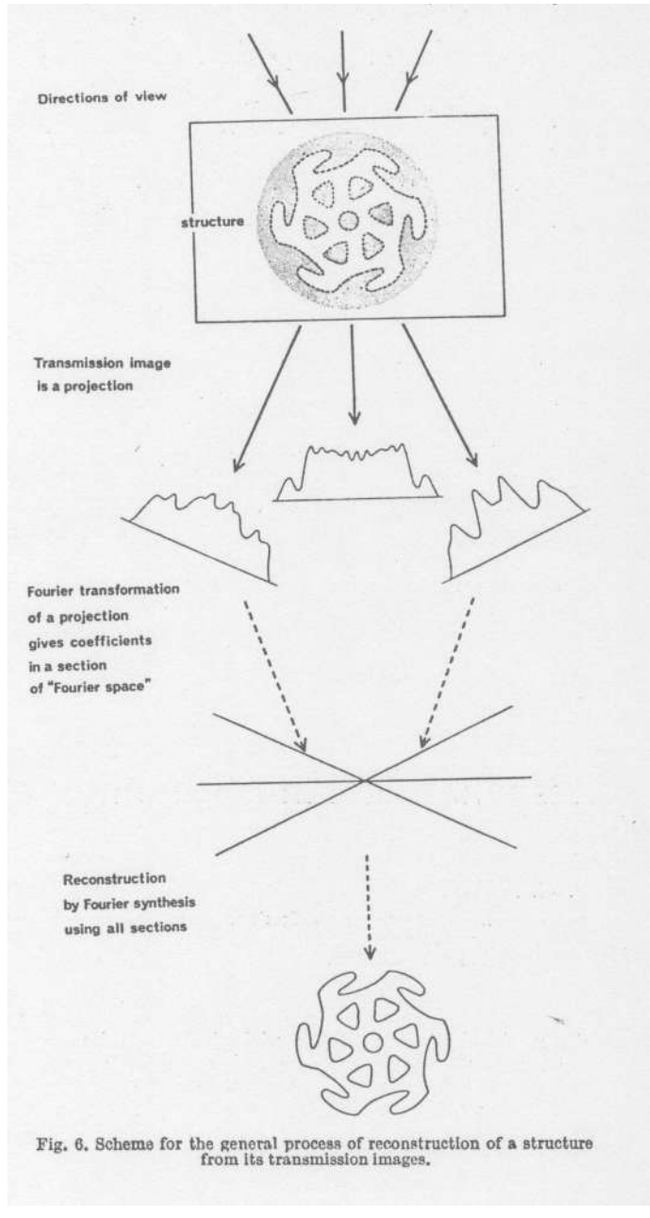
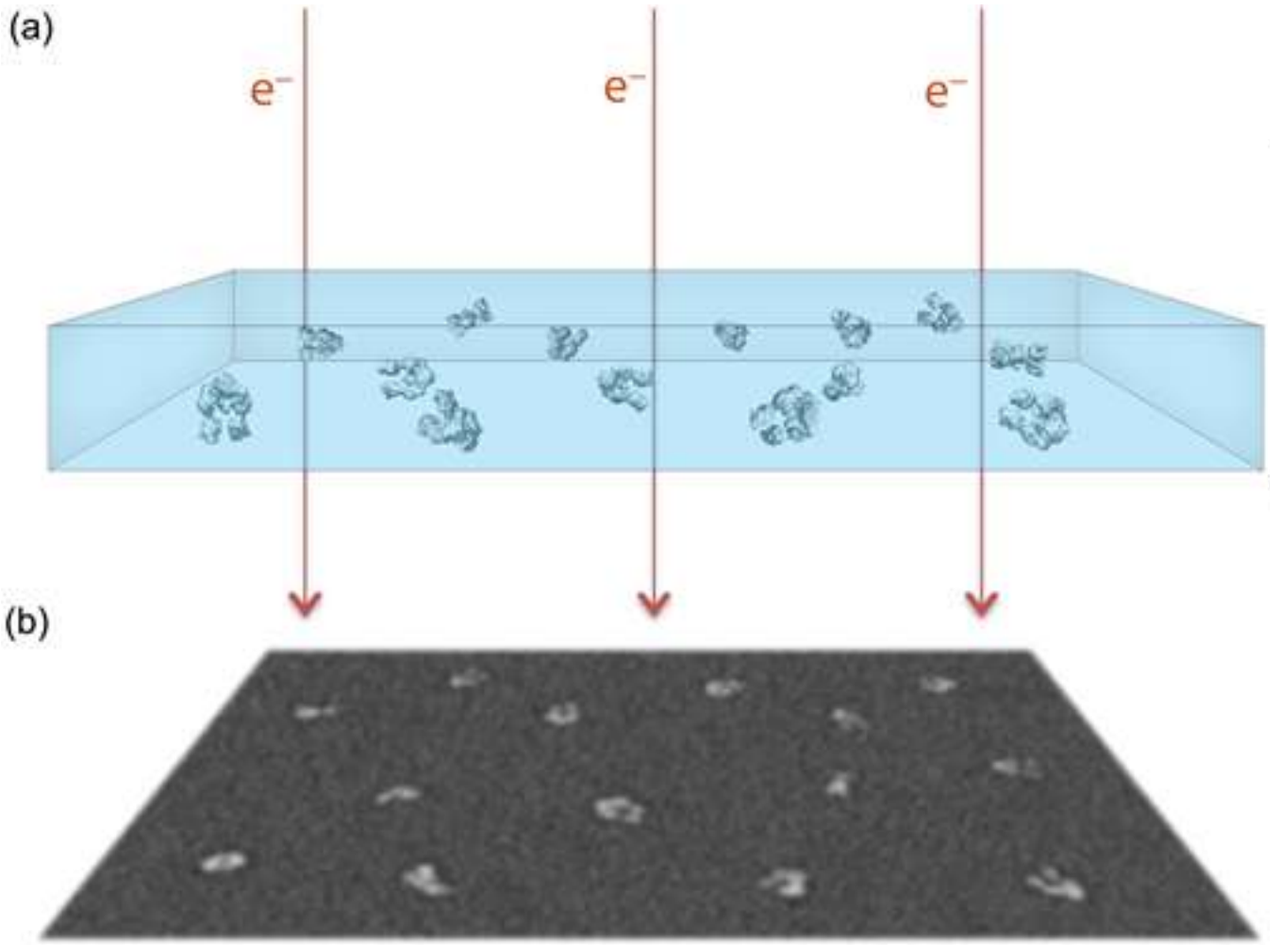


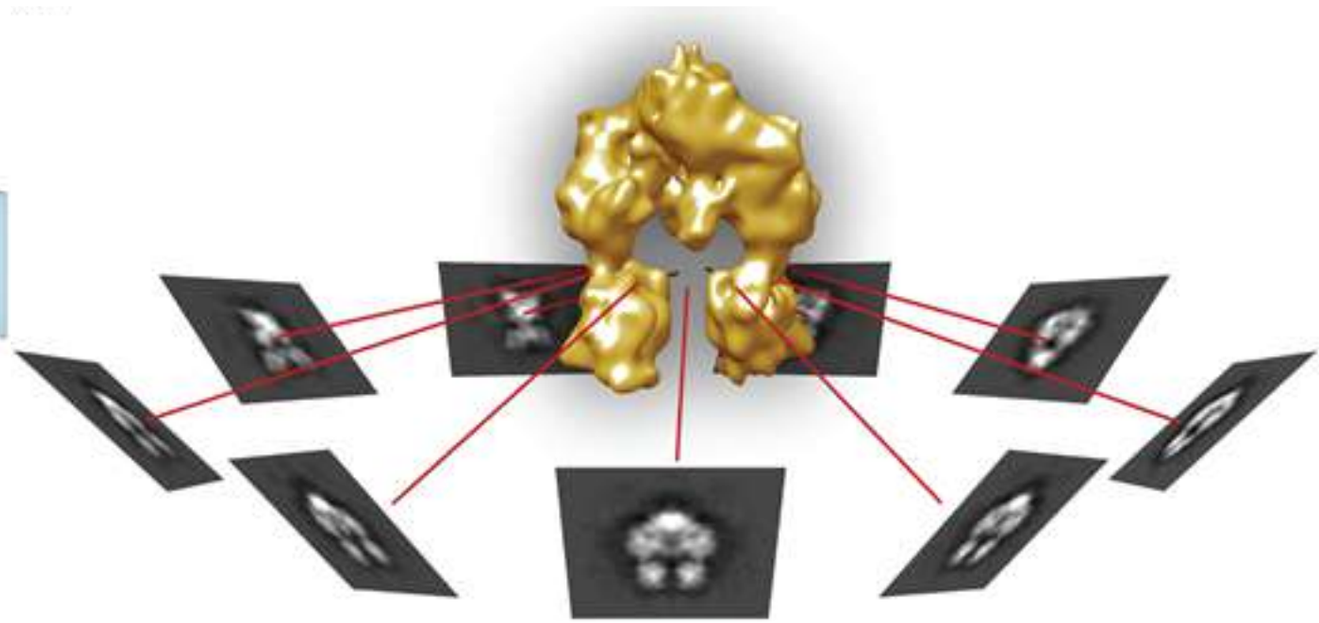
Fig. 6. Scheme for the general process of reconstruction of a structure from its transmission images.

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

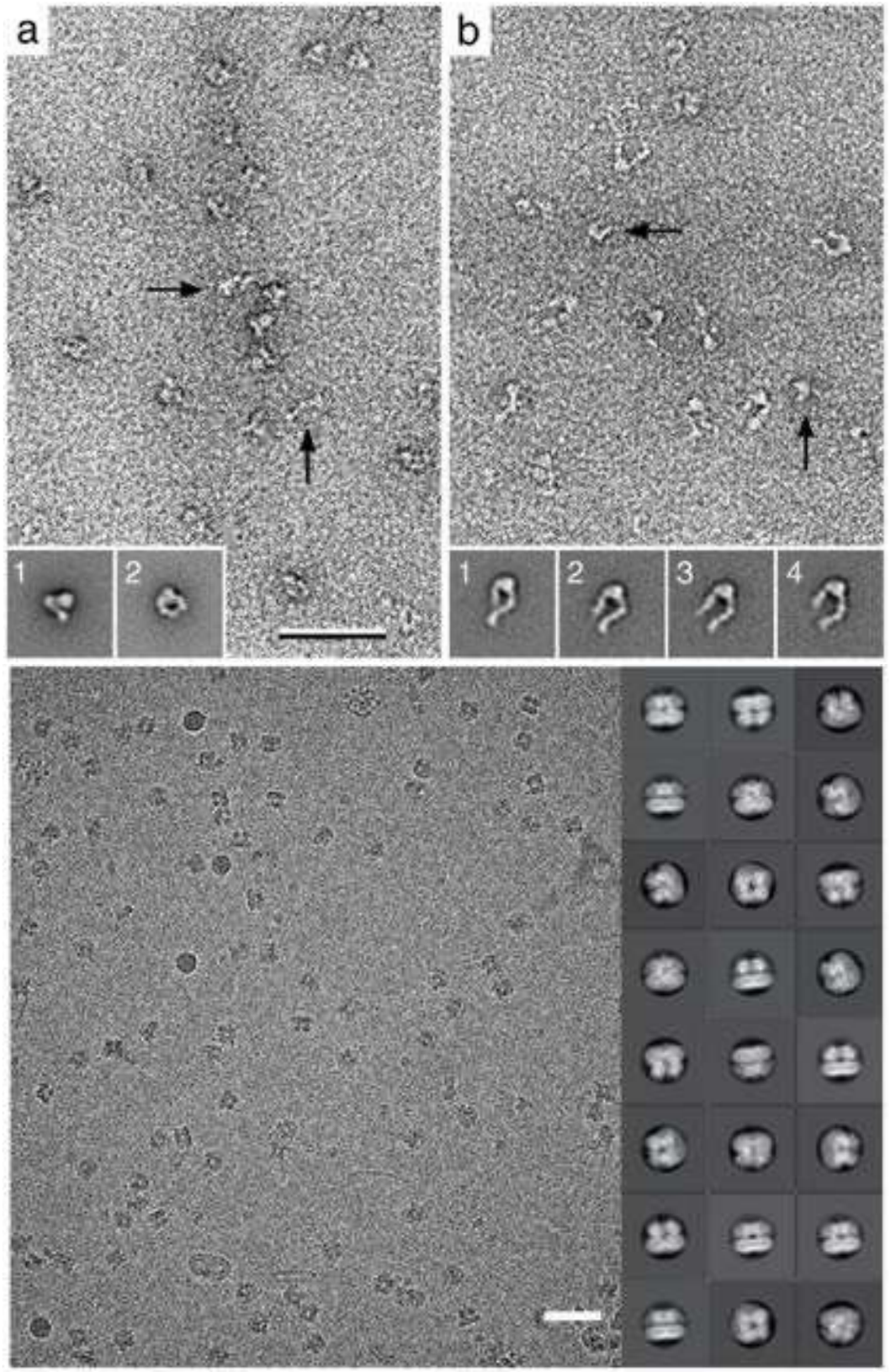




Skiniotis and Southworth. Microscopy. Vol. 65, (2016) pp. 9-22

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.

Development of high-resolution cryo-EM

- **Vitrification** of viruses and proteins (early 1980's)
  - Enabled the first cryo-EM images to be recorded of individual protein complexes
  - Proteins no longer needed to be crystallized in order to be visualized
- Proteins were much less sensitive to beam-induced radiation damage
- Resolution was drastically improved

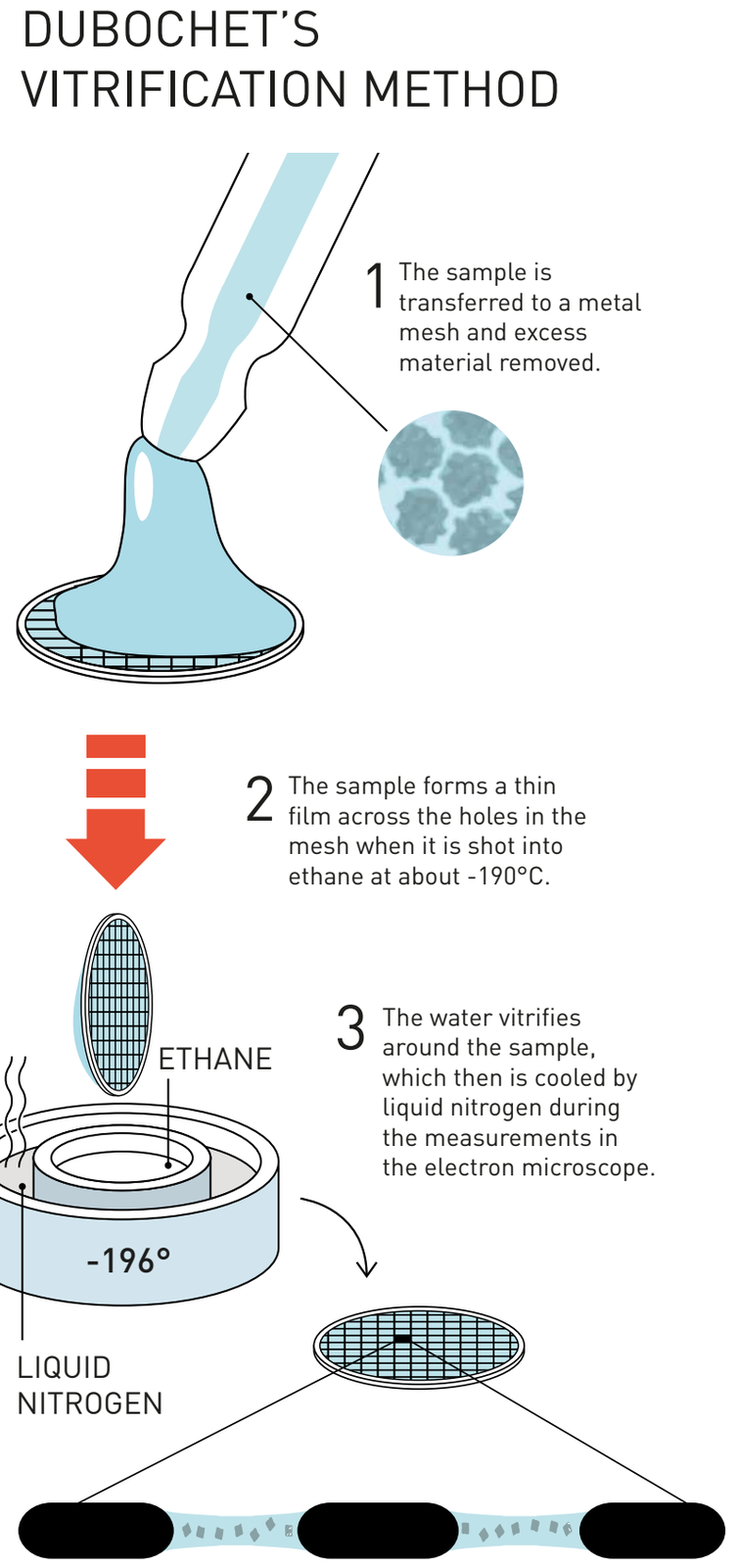
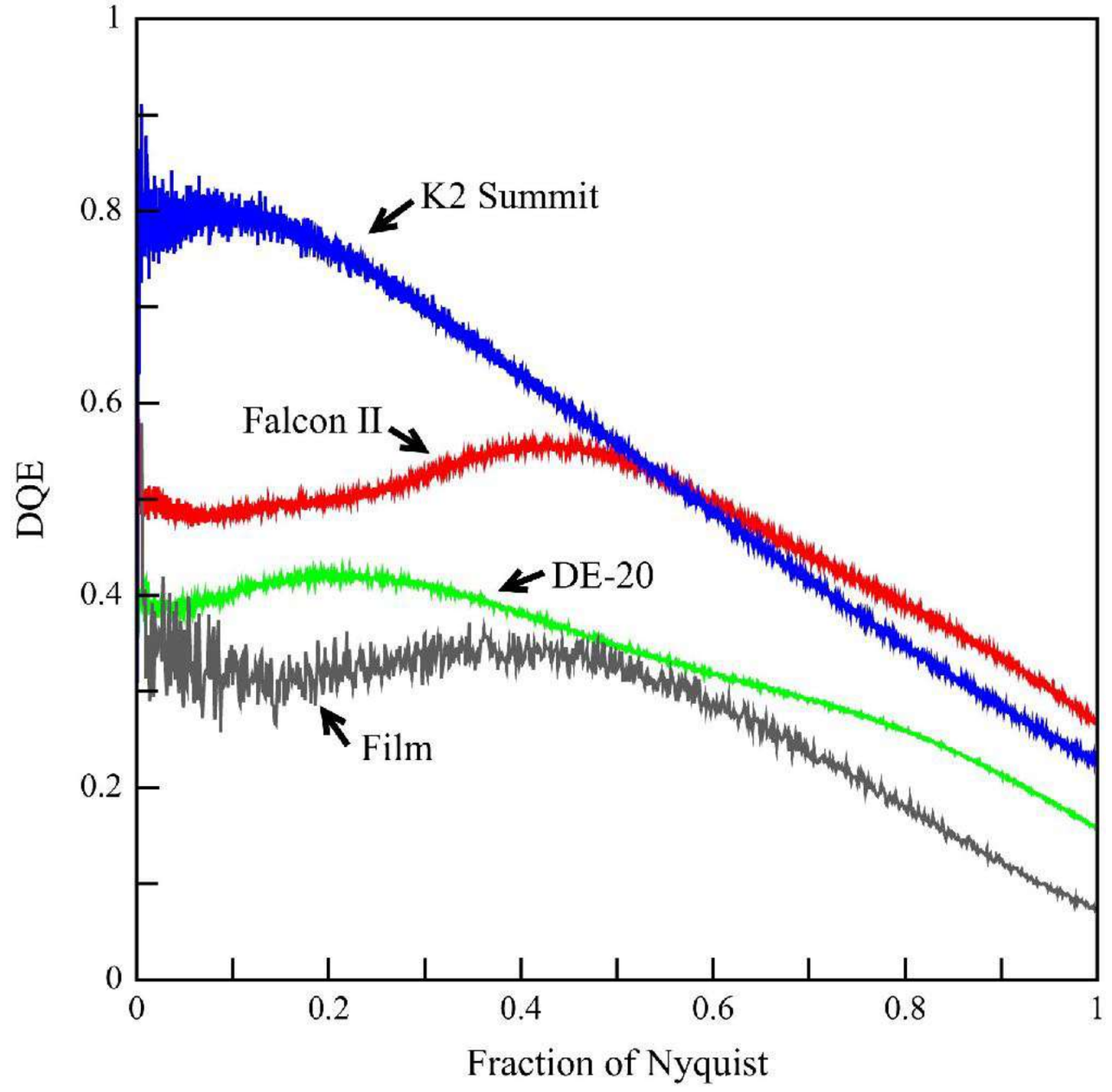


Illustration: ©Johan Jarnestad/The Royal Swedish Academy of Sciences

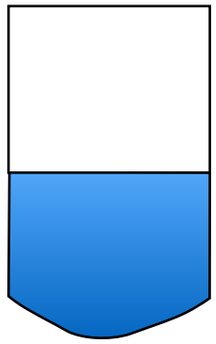
Direct electron detectors

- 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 rapid cryo-EM structure determination

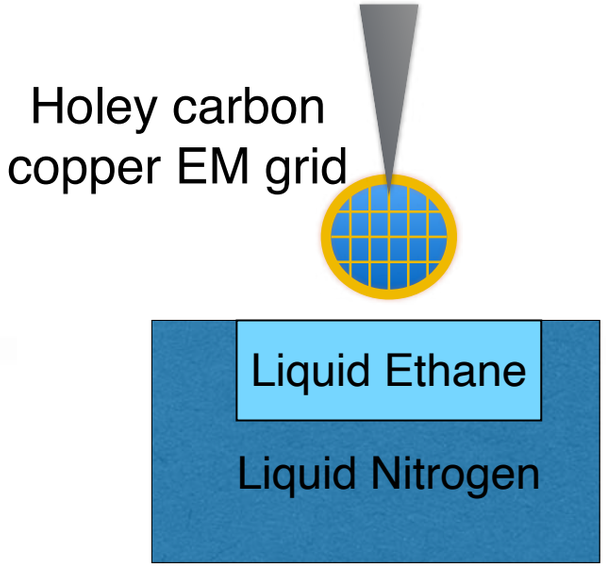


McMullan et al, Ultramicroscopy, Vol 147 (2014) pp.156-163.

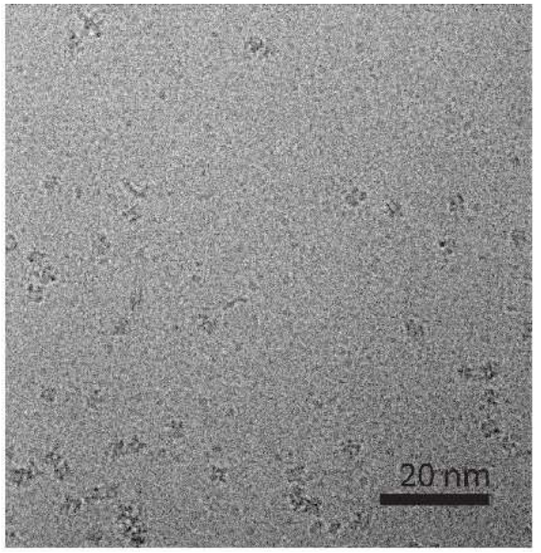
Cryo-EM single particle analysis



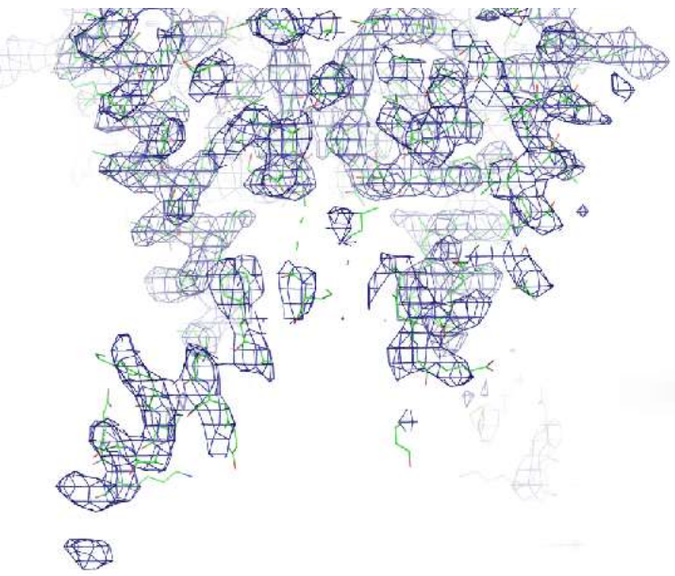
Purify sample



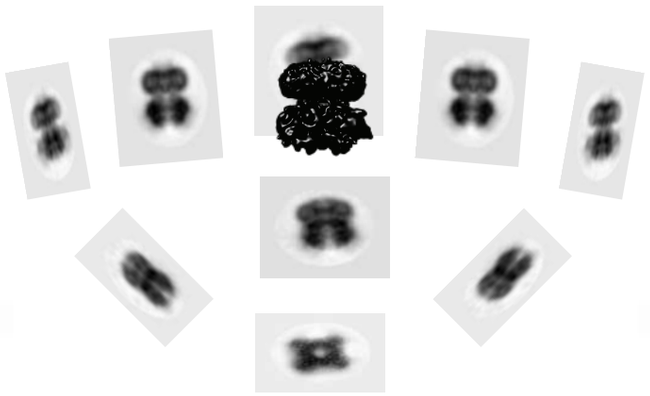
Freeze sample



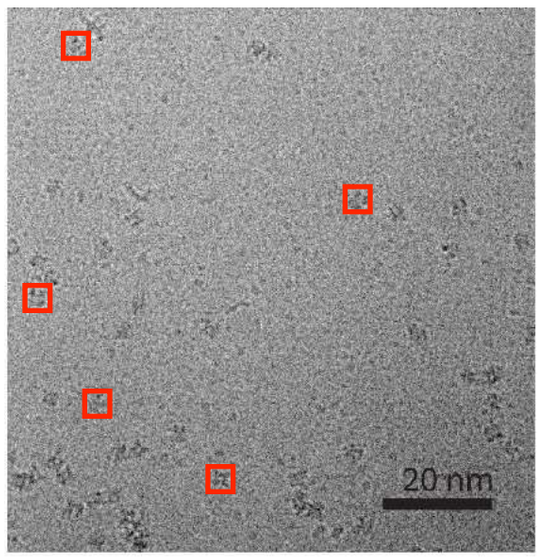
Collect images



Build and refine
atomic model



Reconstruct 3D volume
by projection matching



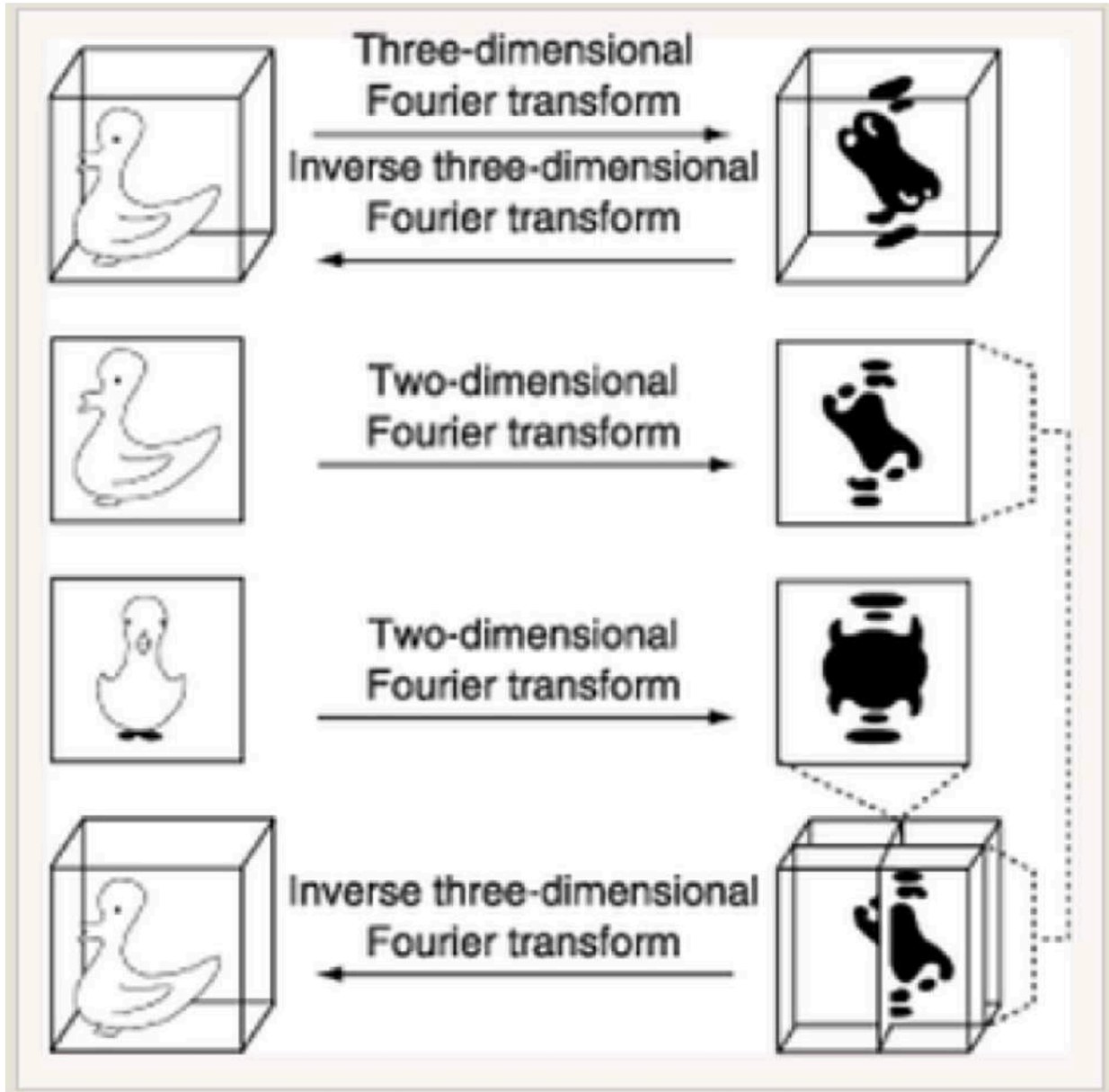
Select particles

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

Cryo-EM image reconstruction

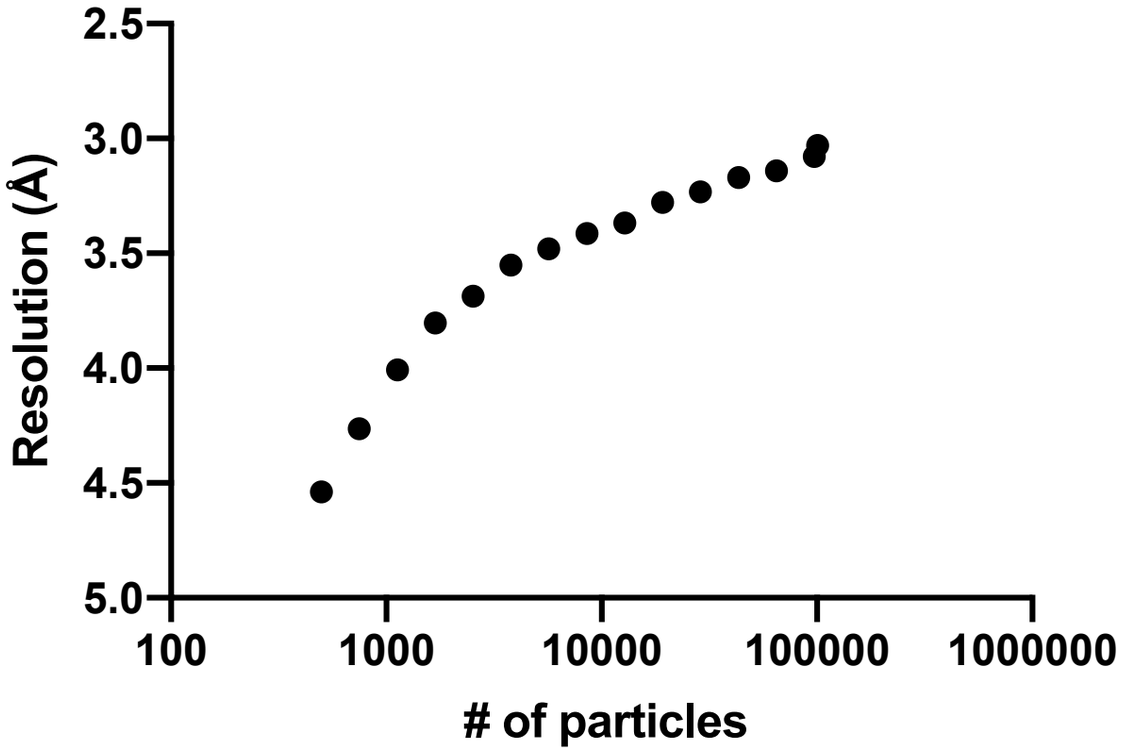
- How to obtain 3D dimensional information from 2D images?
- We rely on the Central section theorem which states that a 2D Fourier transform of the 2D projection of a 3D density is equal to the central section of the 3D Fourier transform of the density perpendicular to the direction of the projection
- By imaging many different projection of the sample, we can sample the entire 3D Fourier transform of the density
- Using an inverse Fourier transform, we can then visualize the 3D volume of the object



Frank, 2006

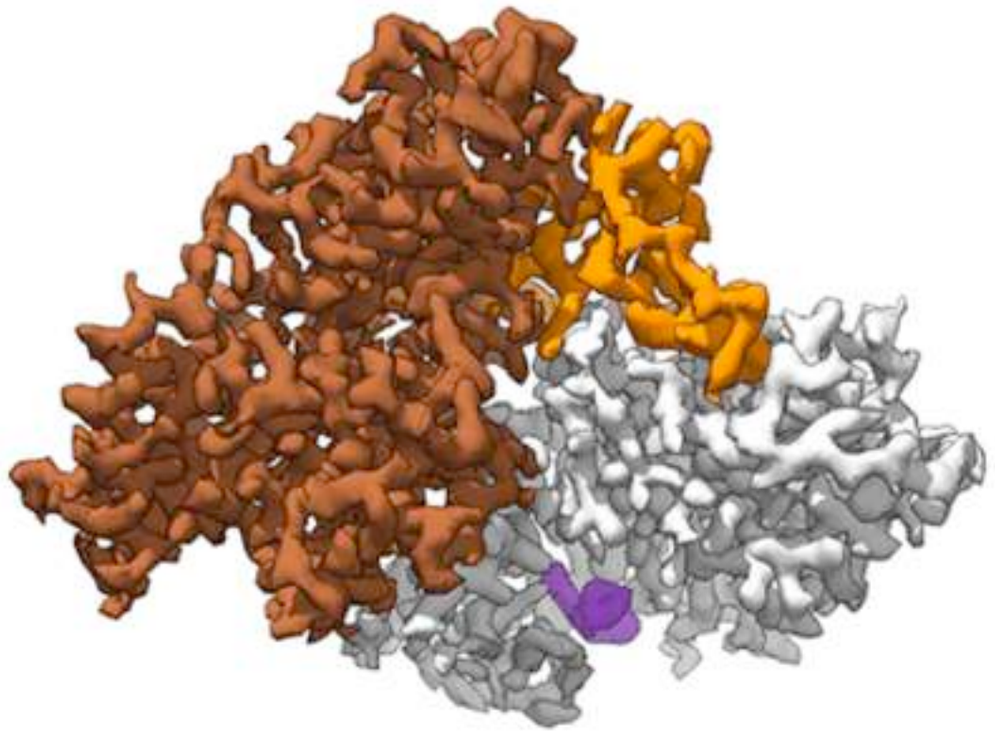
Why millions of particles?

- Biological specimen are extremely sensitive to radiation damage - electron beam can break bonds
- To minimize the effects of radiation damage, a very low number of electrons is used to image the specimen
- Signal-to-noise ratio of cryo-EM data is low, especially at high resolution
- Averaging techniques can alleviate the low signal-to-noise issue because the signal is invariant from particle to particle while the noise is random and will vary between particles



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?

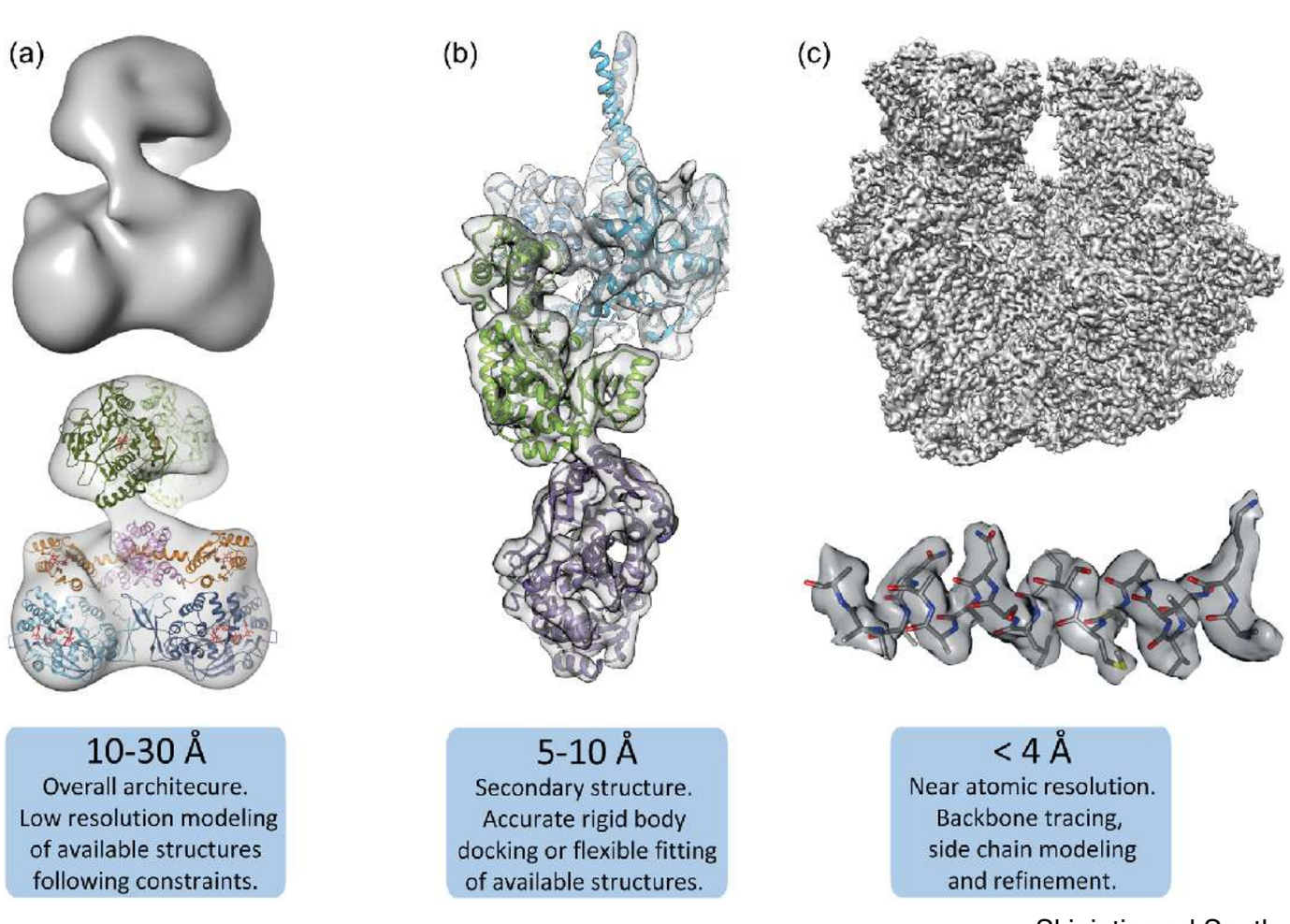


EMD-12042
2.5 Å

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

Resolution - what does it mean?

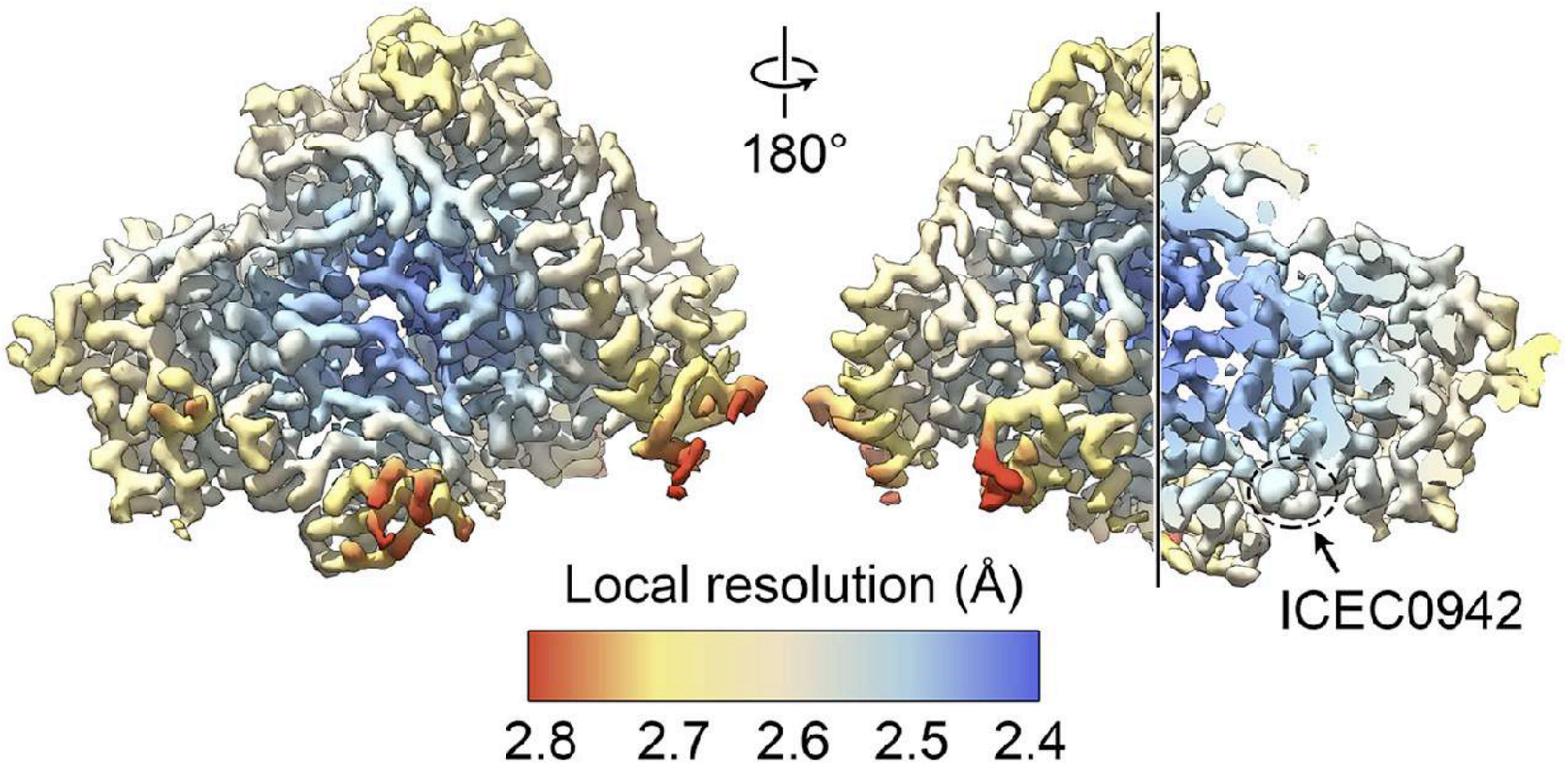
- Resolution corresponds to the distance between two objects that can be reliably separated in the map
- In EM, resolution is estimated by the Fourier shell correlation (FSC)
- FSC is determined by comparing the cross correlation between two independently calculated maps in Fourier space resolution shells
- 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



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

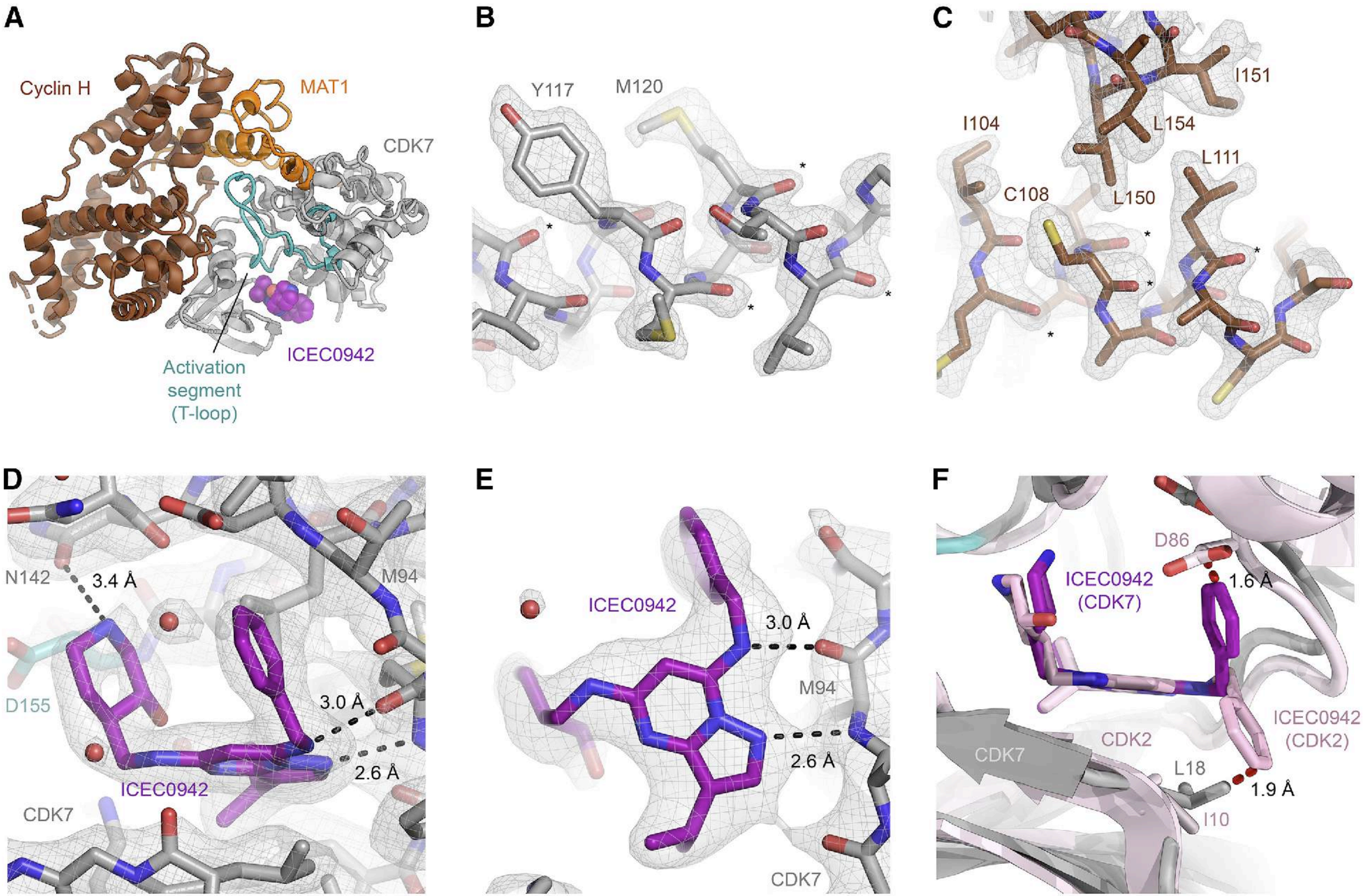
Resolution - what does it mean?

- 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.

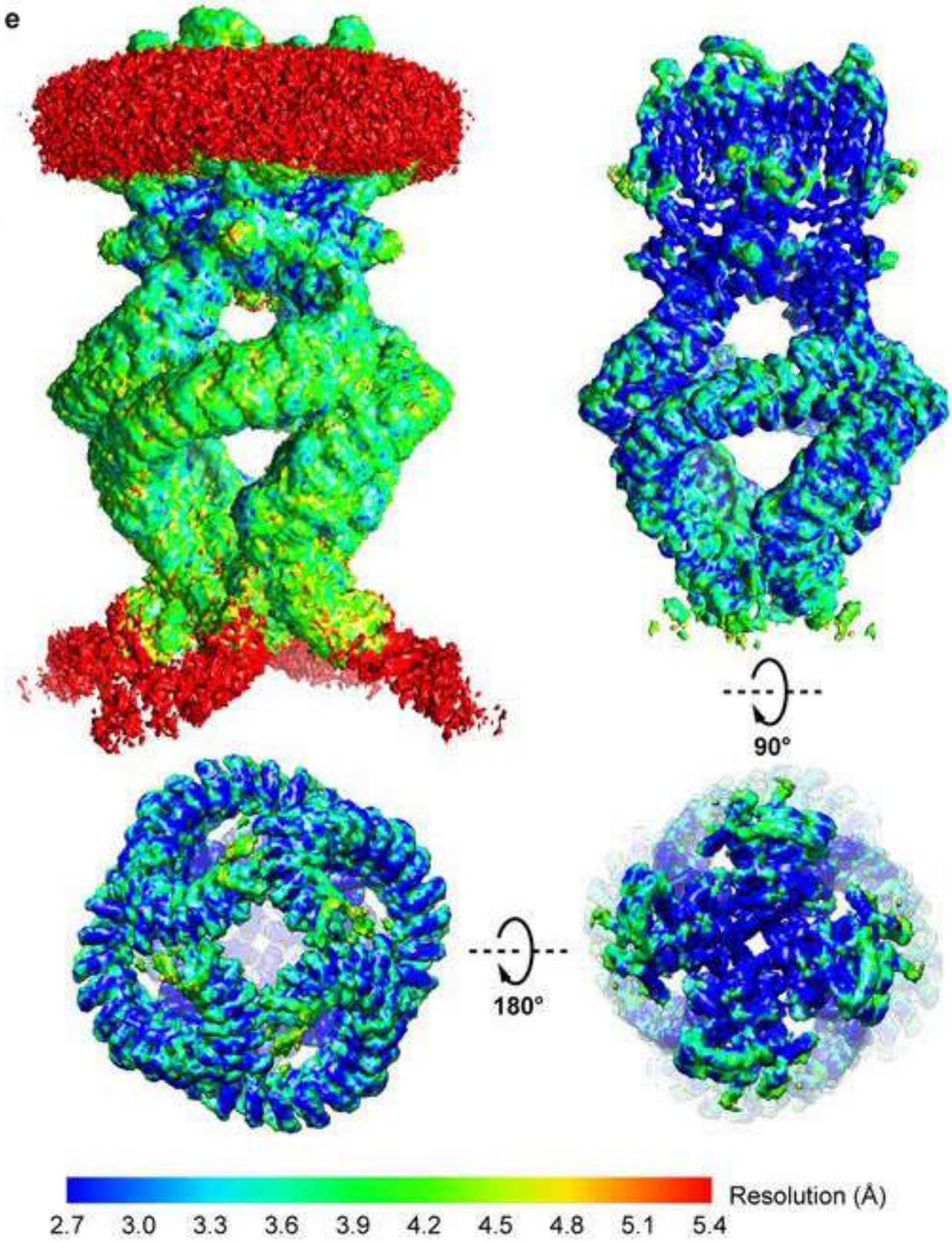
Resolution - what does it mean?



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

Resolution - what does it mean?

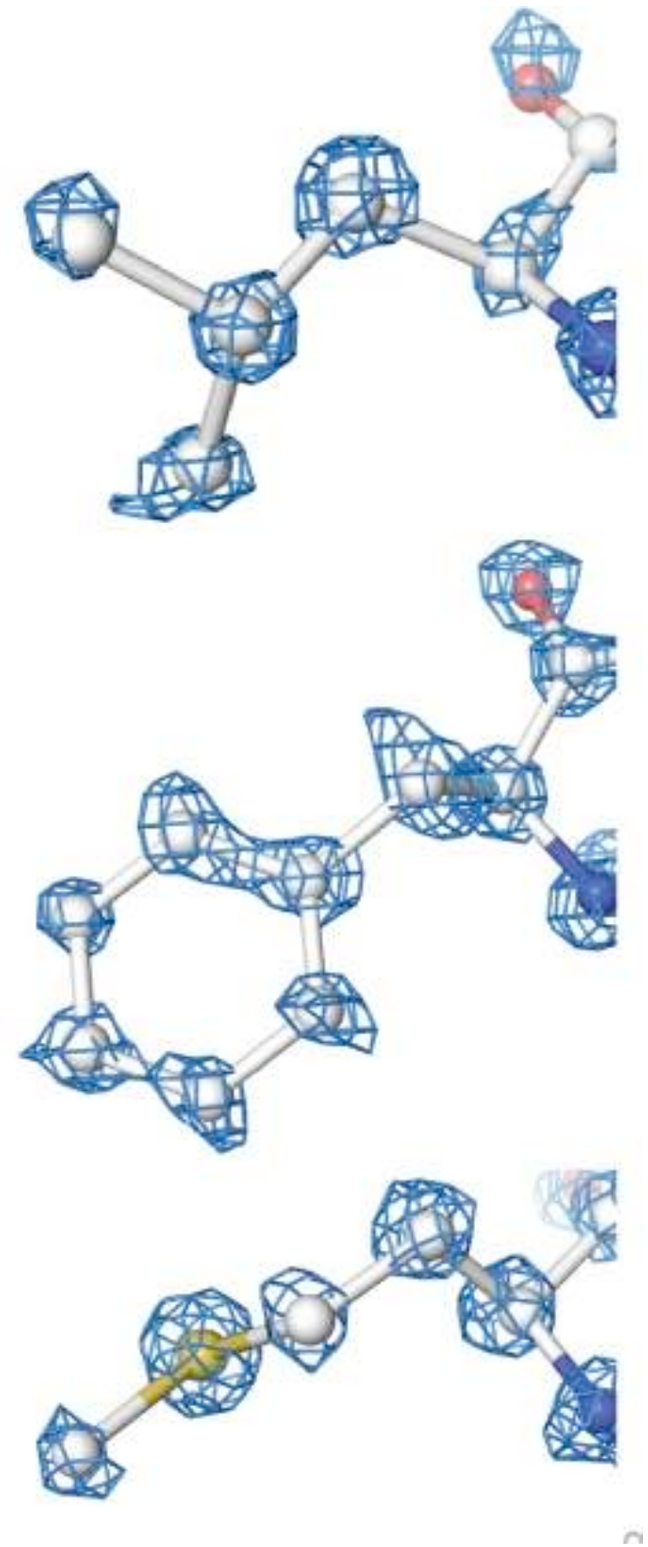
- Overall resolution does not equal local details
- Quality can vary greatly within a map and care should be taken to not over interpret poorly ordered domains
- Different resolutions have different types of features that can be visualized in the density map
  - 7 Å - alpha-helices
  - 4.5 Å - beta-strands
  - 4 Å - large side chains
  - > 3 Å water molecules
- Local resolution calculations can be performed to estimate local resolution
- Density slices can also be extremely informative for evaluating local map quality



Jin *et al*, Nature. (2017) 547. pp 118-122

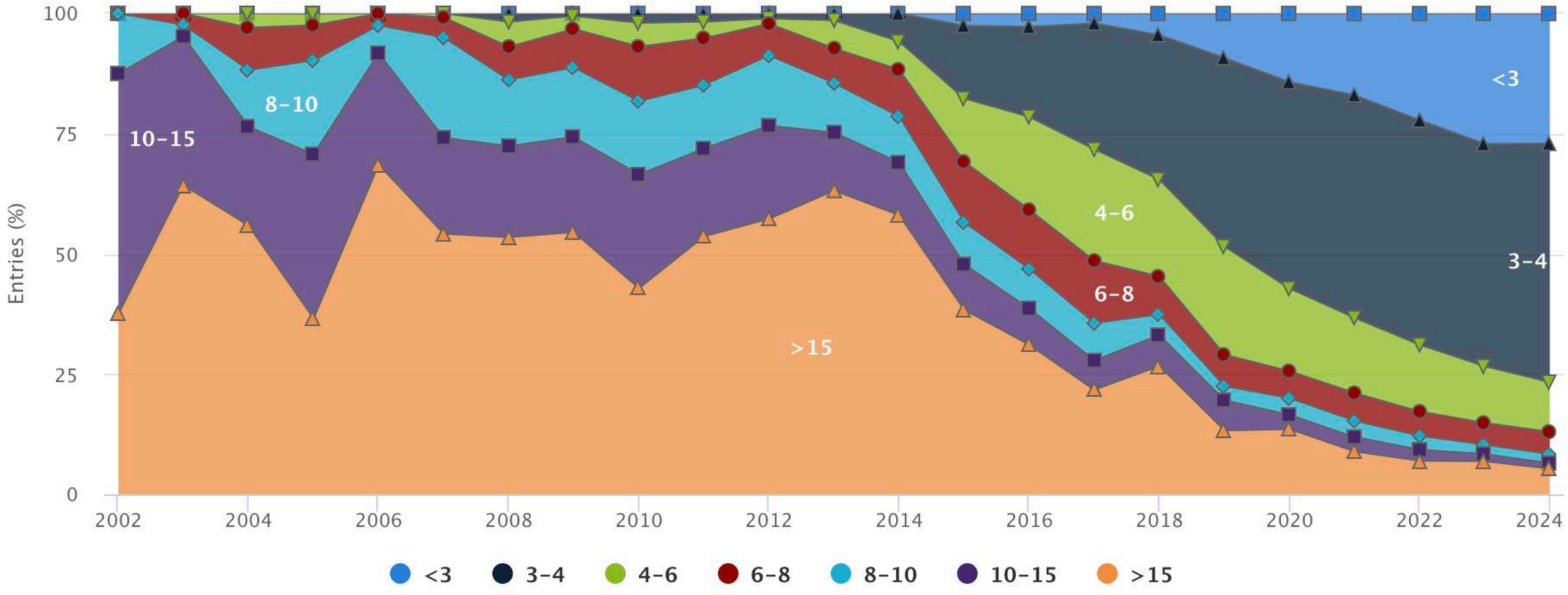
Resolution - what does it mean?

- Overall resolution does not equal local details
- Quality can vary greatly within a map and care should be taken to not over interpret poorly ordered domains
- Different resolutions have different types of features that can be visualized in the density map
  - 7 Å - alpha-helices
  - 4.5 Å - beta-strands
  - 4 Å - large side chains
  - > 3 Å water molecules
  - 1.2 Å allows resolving individual atoms



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

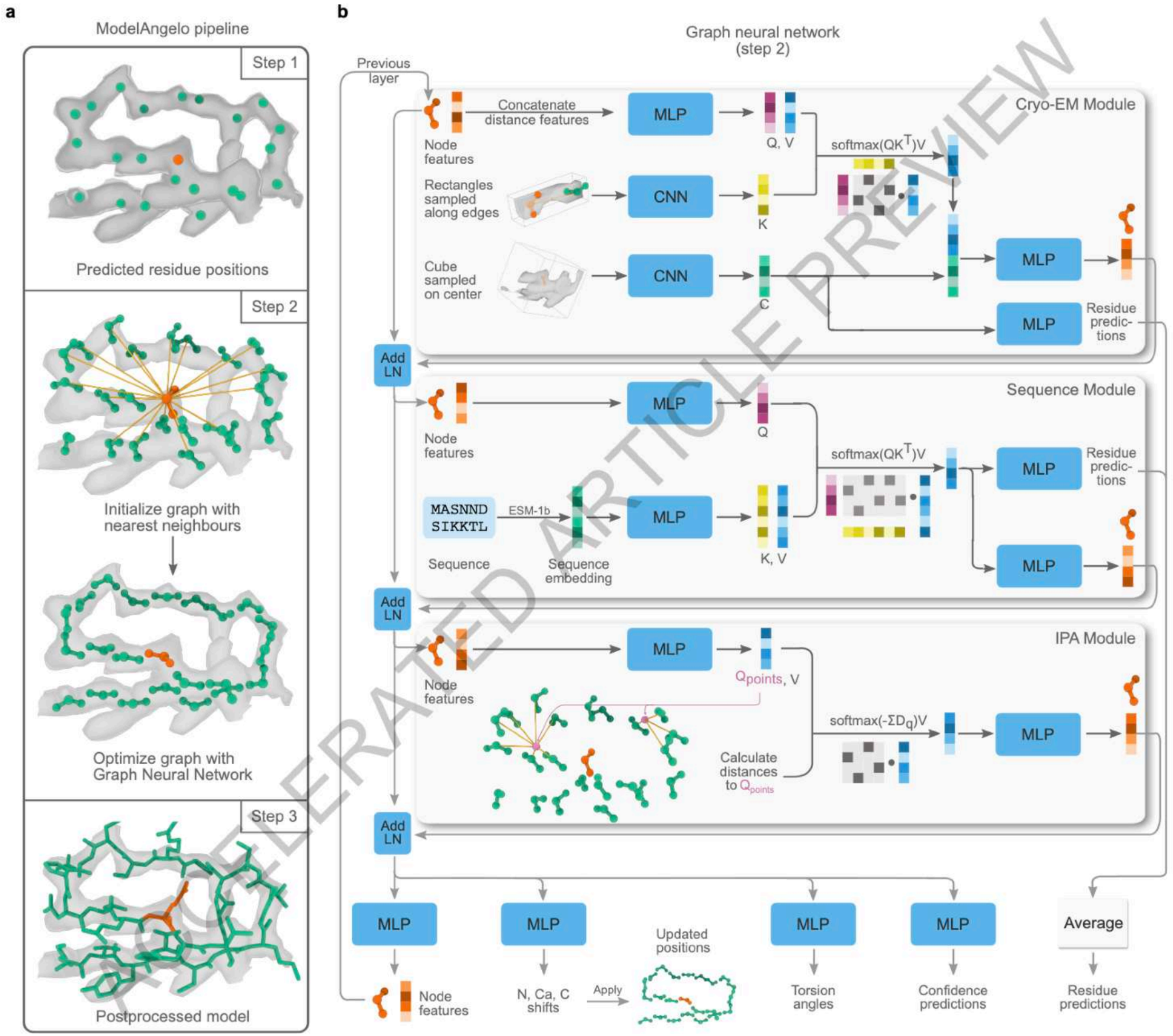
Resolution - what does it mean?



EMDB 2024.3.14

Map to Model

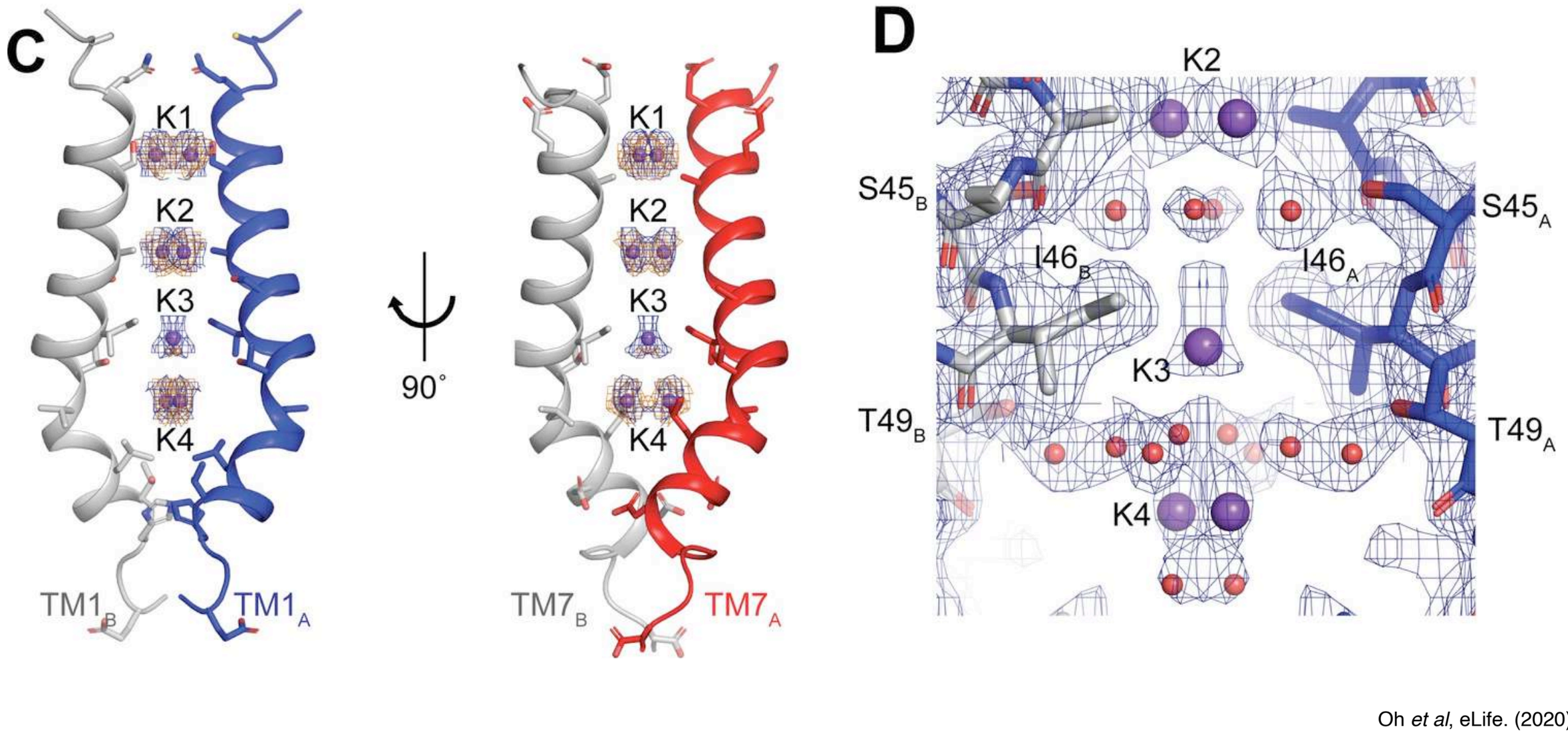
- Model building has been rapidly accelerated
- Hand building of high-resolution structures has been replaced with automated approaches
- Modelangelo is one of them that can rapidly build into maps at better than 3.5 Å with high accuracy



Jamali *et al*, Nature. (2024)

Is all density in a map is the same?

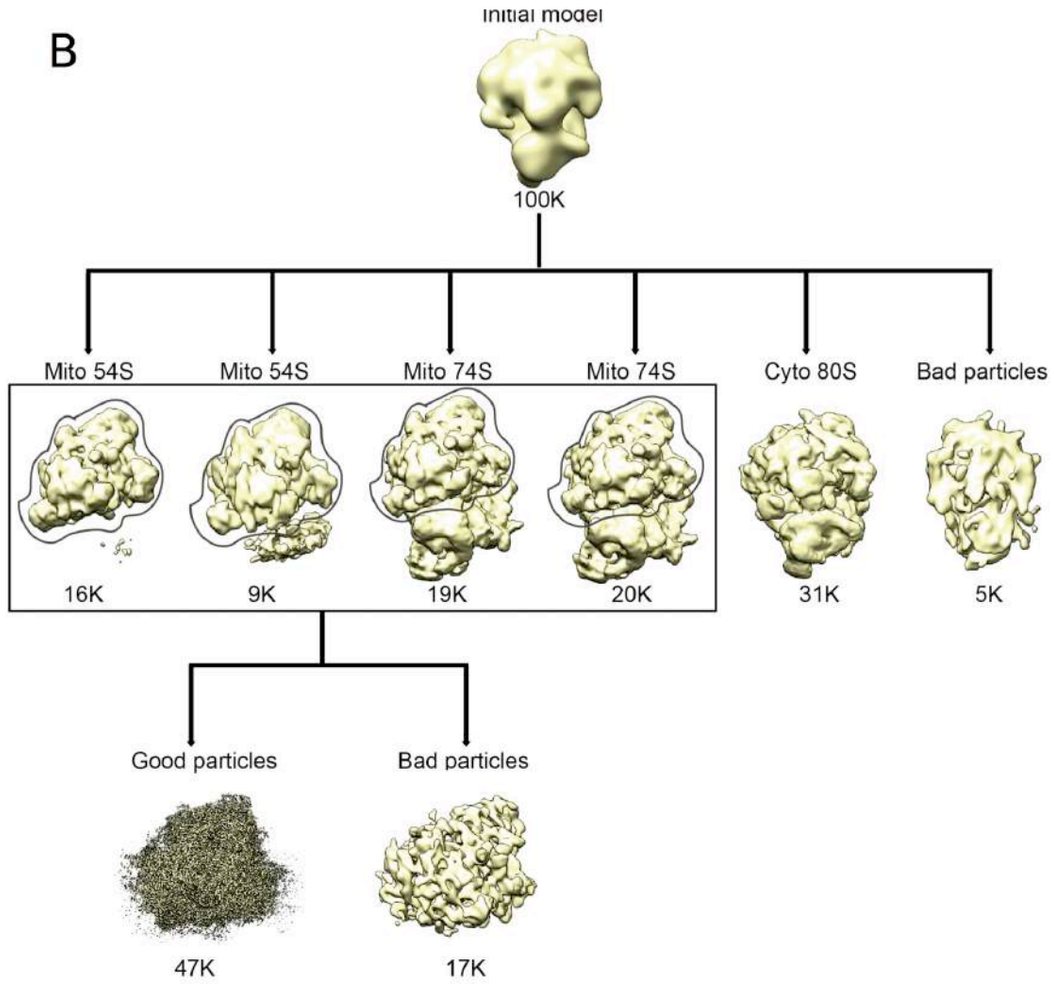
- Identifying non-protein densities in cryo-EM is quite challenging as there few tools that allow accurate discrimination between atoms in density maps.
- In some circumstances one can replace an atom with a similarly structured atom to distinguish ion binding sites



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



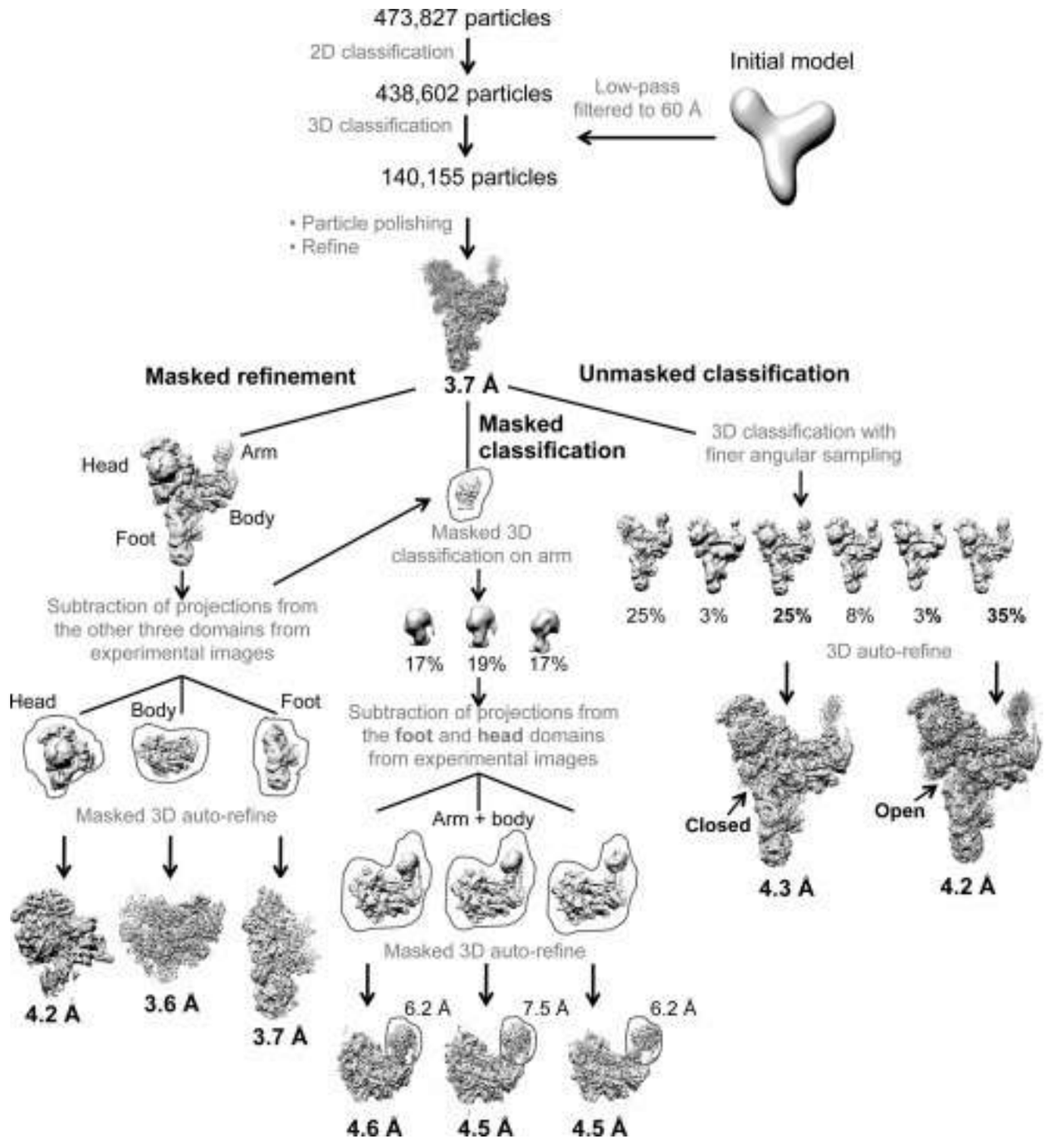
Amunts *et al*, Science. (2014) 343:1485-1489

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

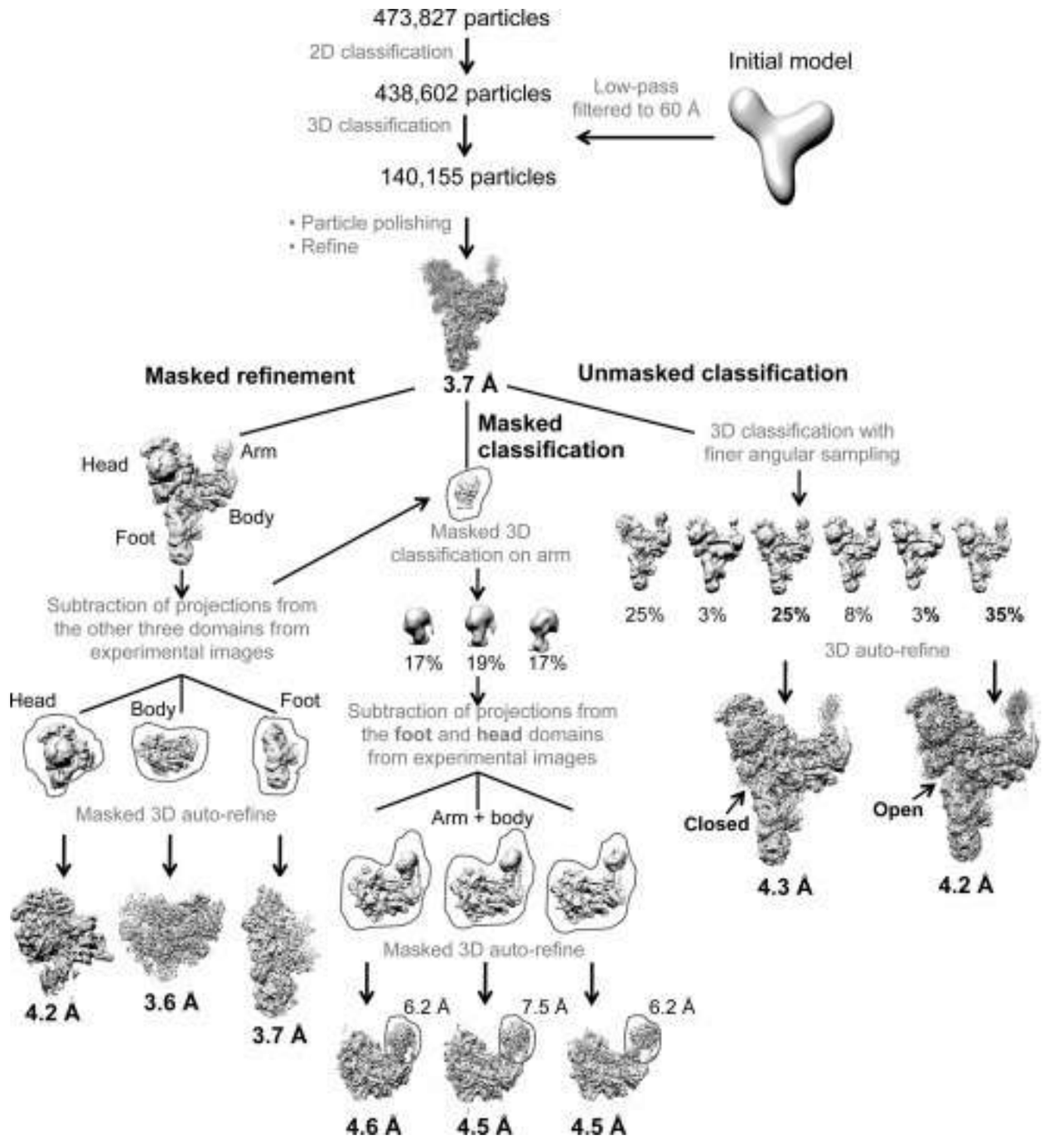
- Heterogeneity may be unavoidable for some samples and must be dealt with computationally after image acquisition
- There are now several different software packages that sort and classify particles, allowing one to create “pure” subsets of the particles images
- The simplest approach is classify based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



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

Overcoming heterogeneity - computation

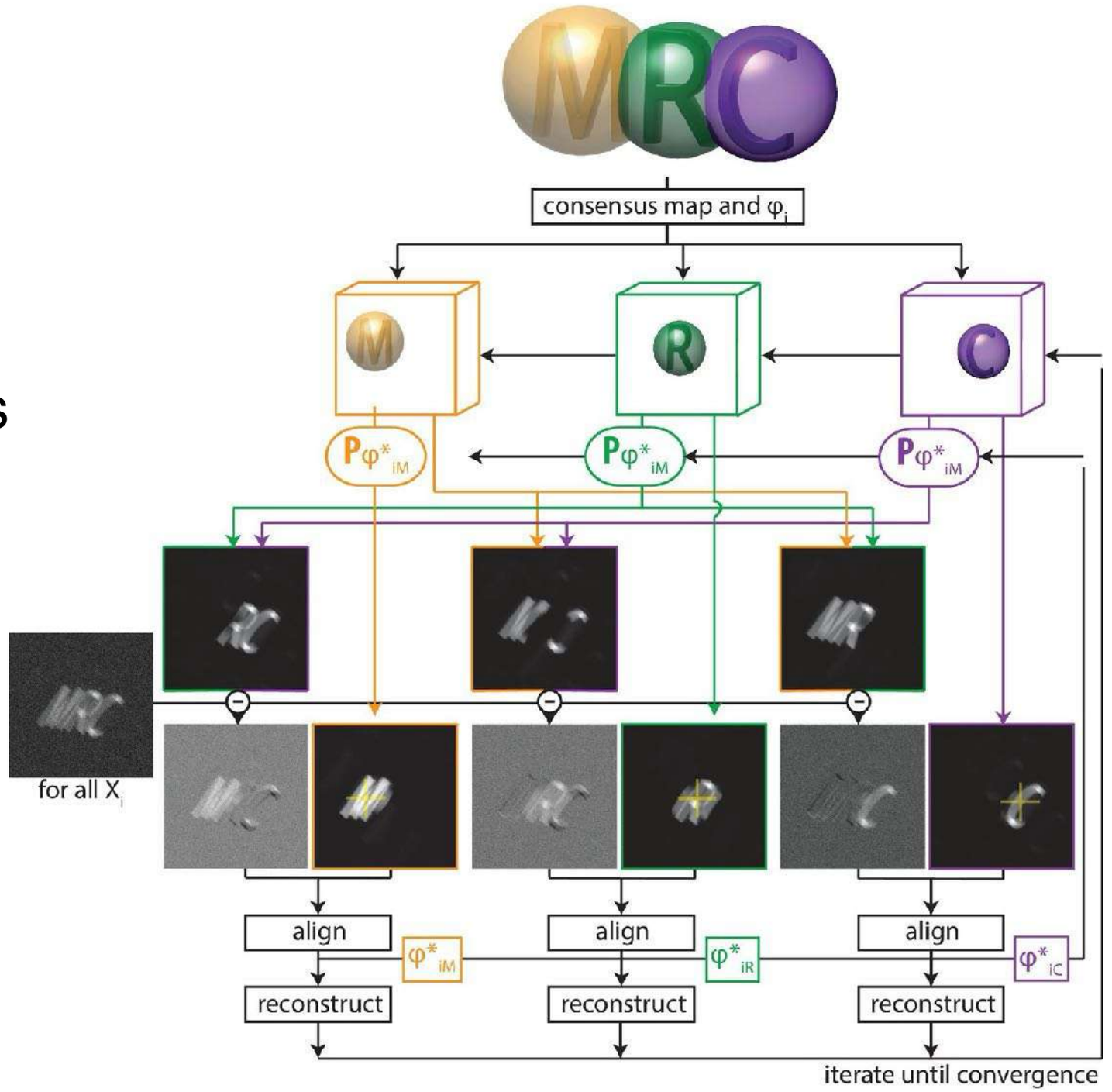
- Classification can be enhanced through the use of masks
- A mask can be placed around the region of interest - allowing independent sorting of different domains
- This multi-classification approach is particularly powerful for samples that have multiple different types of movements
- Another modification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

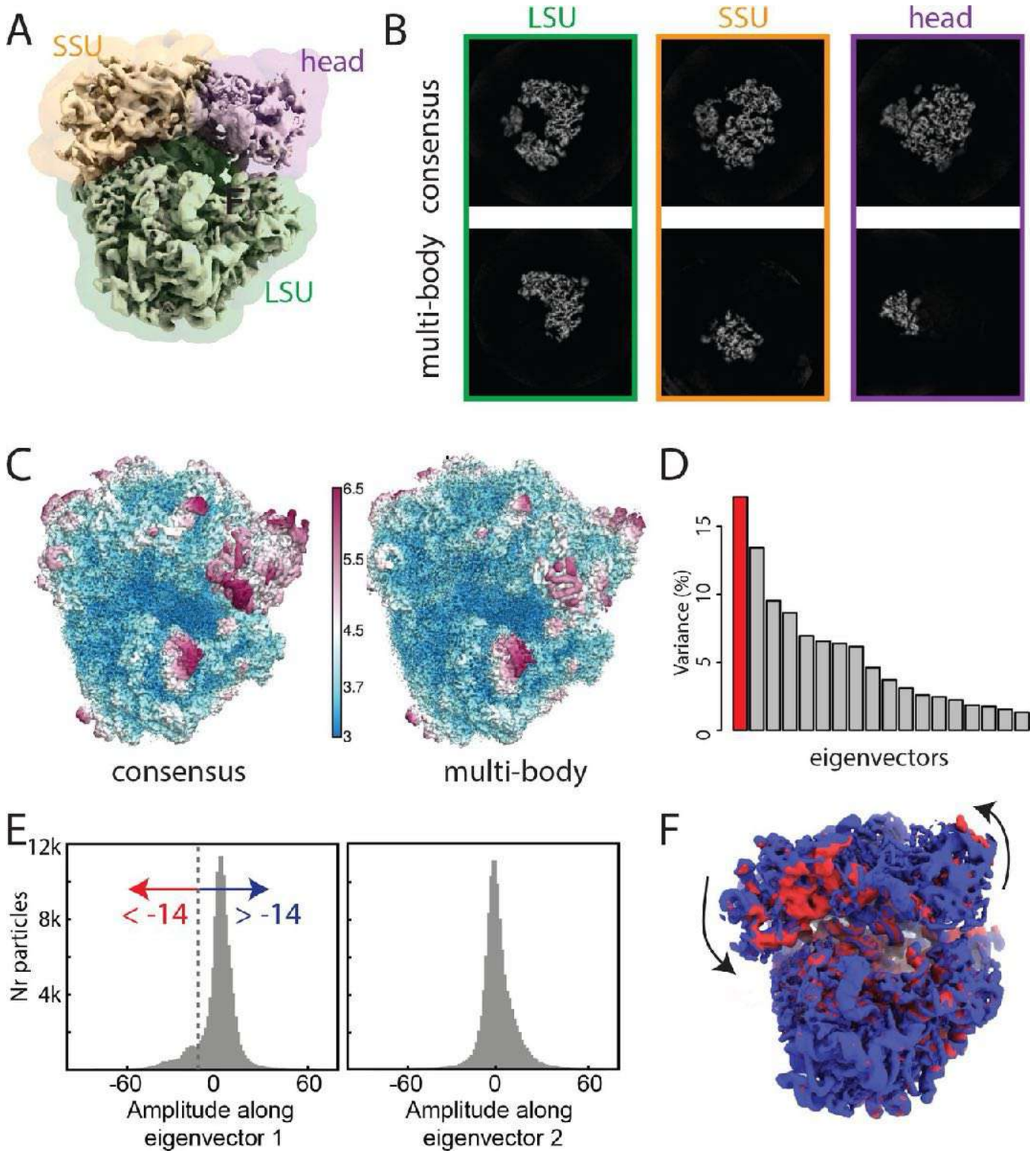
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Overcoming heterogeneity - computation

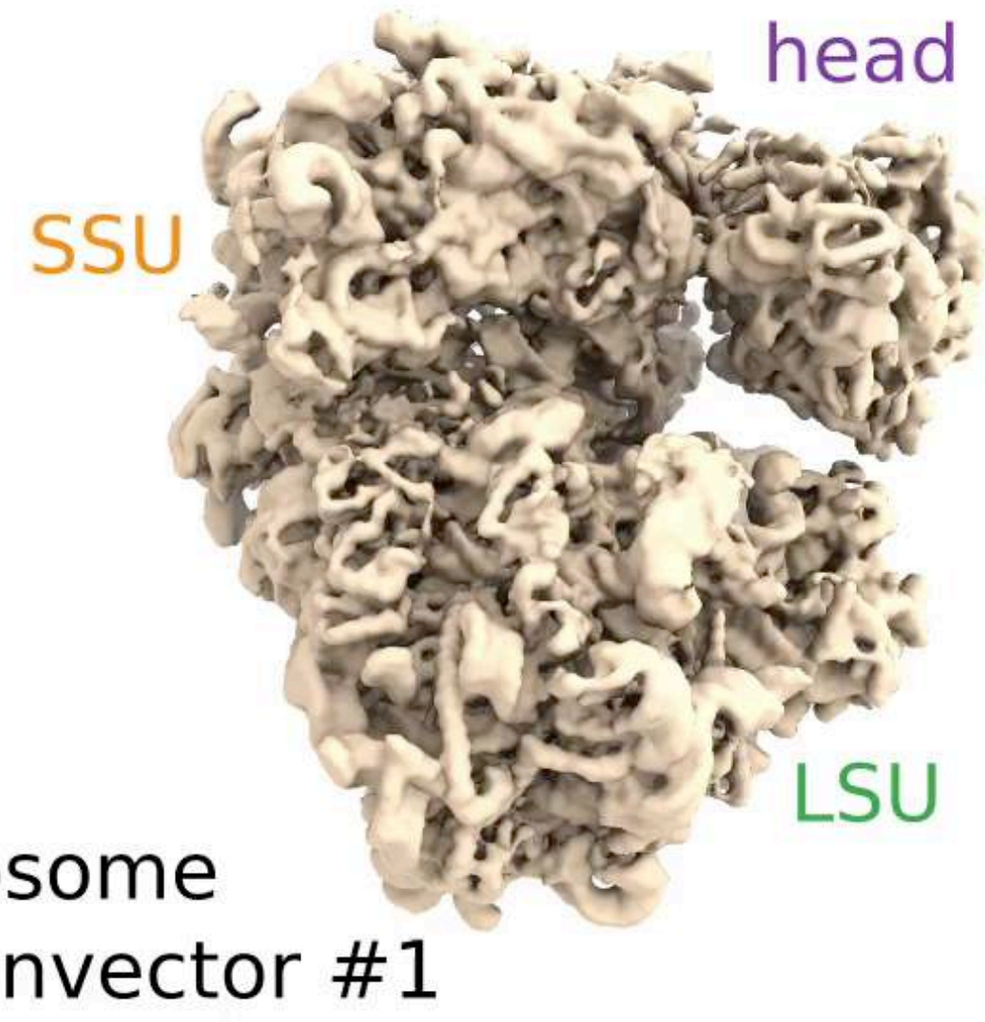
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Overcoming heterogeneity - computation

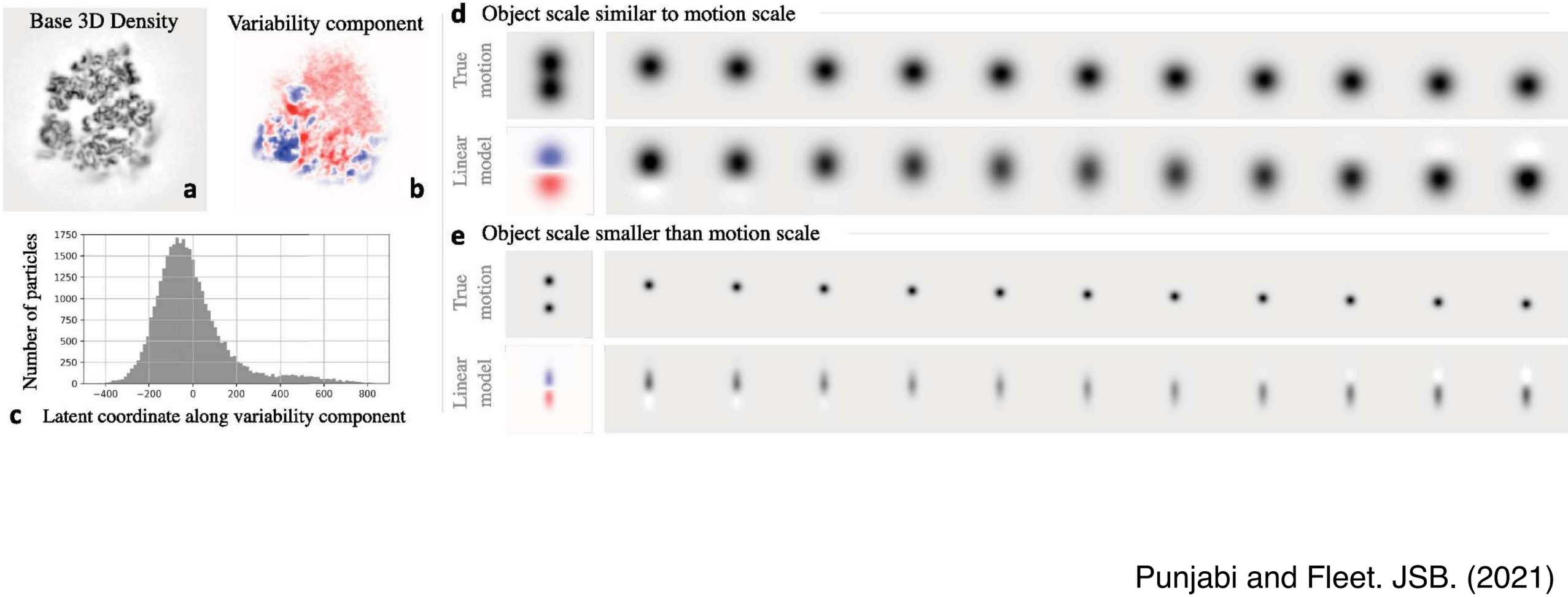
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes
- Requires knowledge of the major domains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



Nakane et al, eLife. (2018) e36861

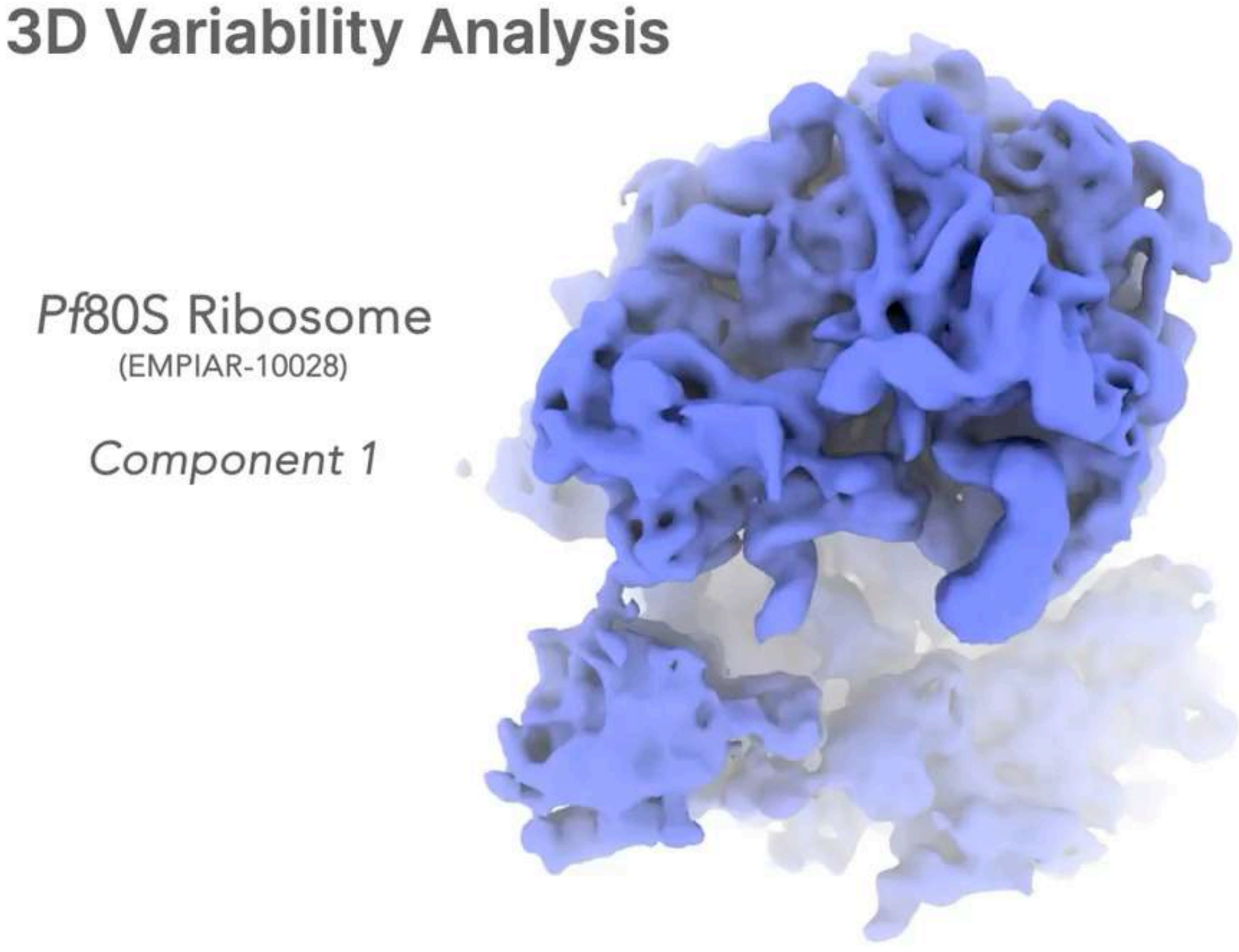
Overcoming heterogeneity - computation

- Cryosparc has an alternative approach that they call 3D variability analysis
- Generates frames of a movie that reveal conformational dynamics along N principal components
- Requires classified particles (to some extent)



Punjabi and Fleet. JSB. (2021)

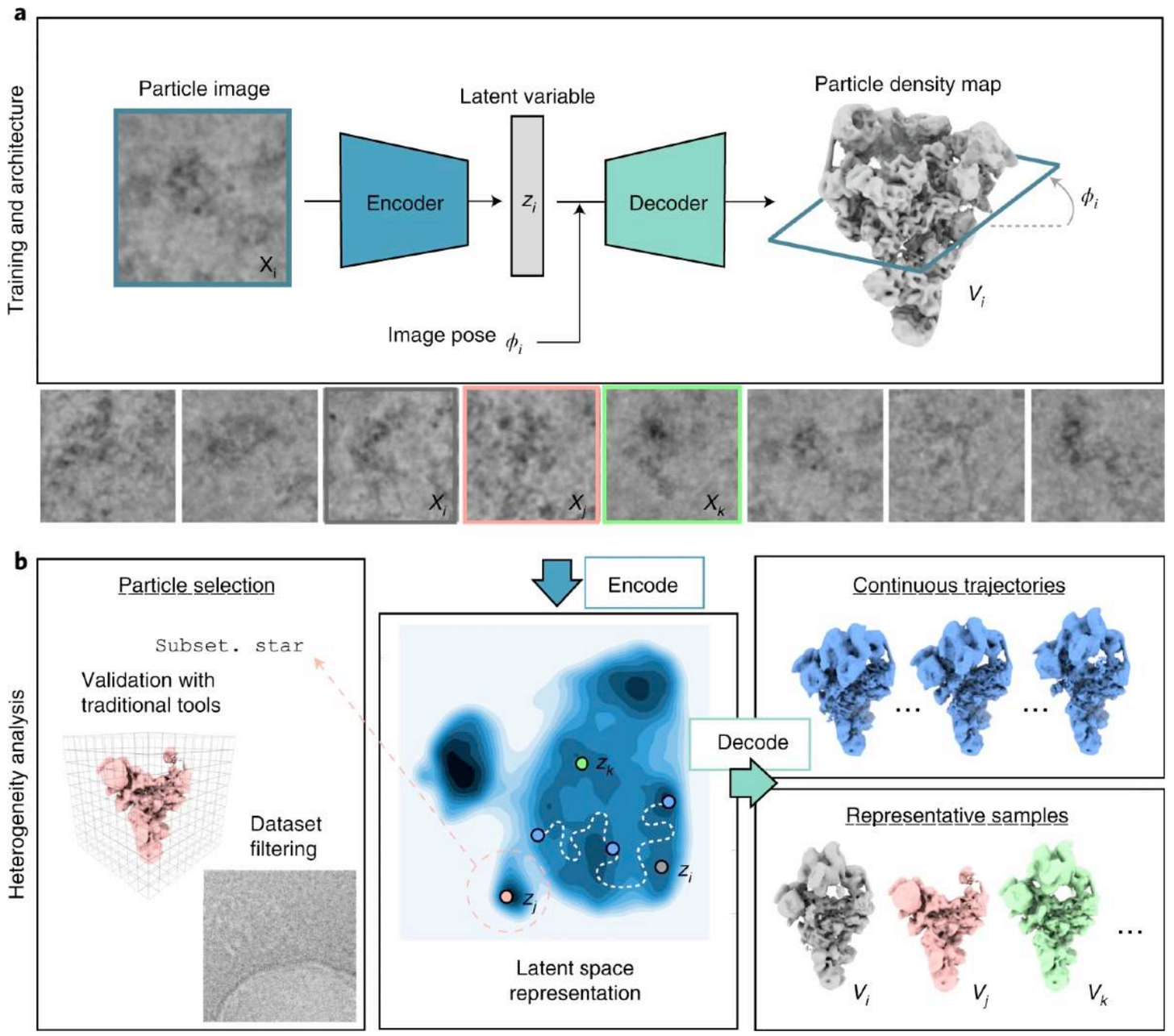
Overcoming heterogeneity - computation



Punjabi and Fleet. JSB. (2021)

Overcoming heterogeneity - computation

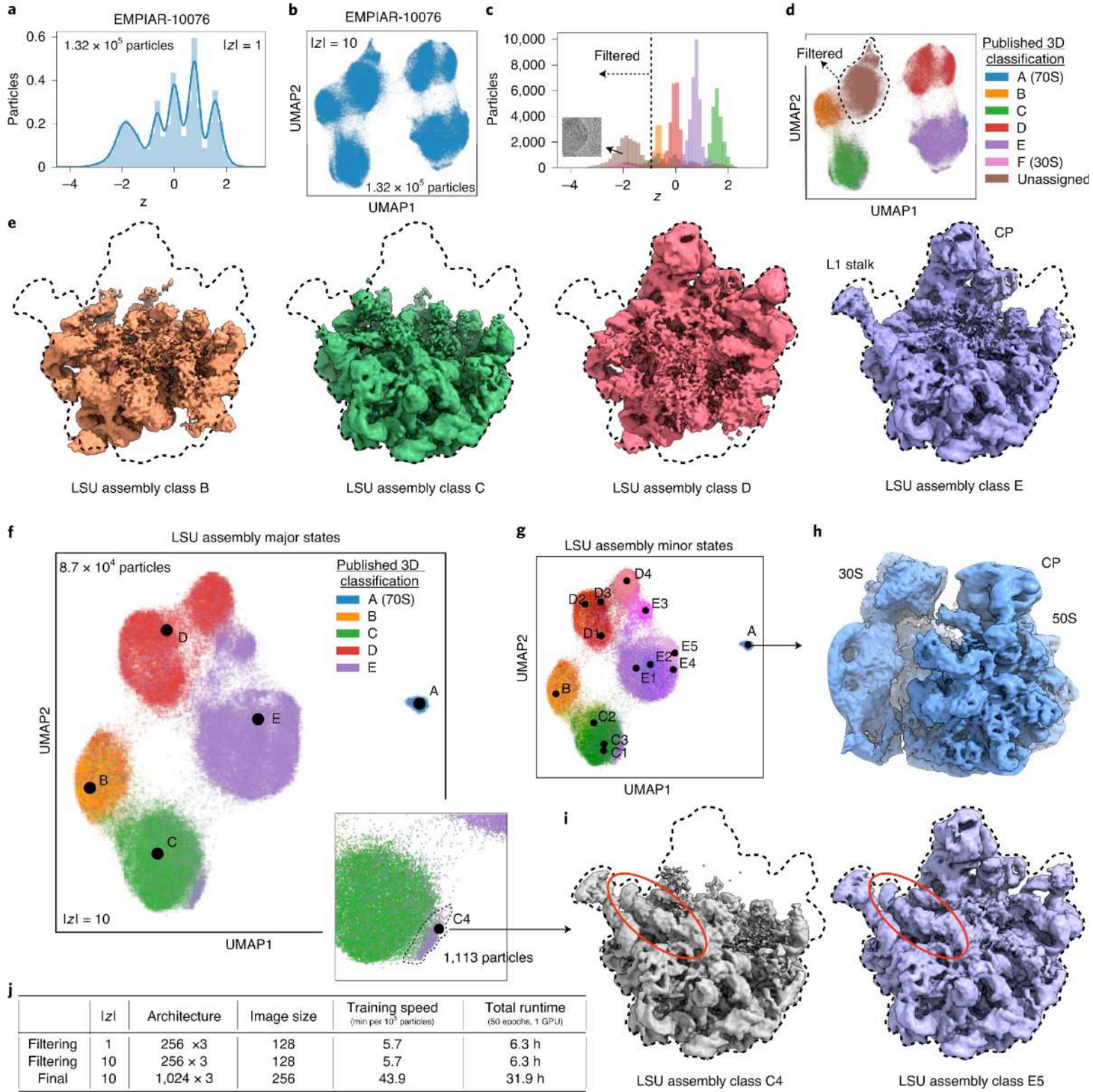
- CryoDRGN - deep learning approach to visualize conformational dynamics
- Does not require classification, uses particles directly and sorts into a latent space representation
- Allows analysis of highly complex data sets



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



j

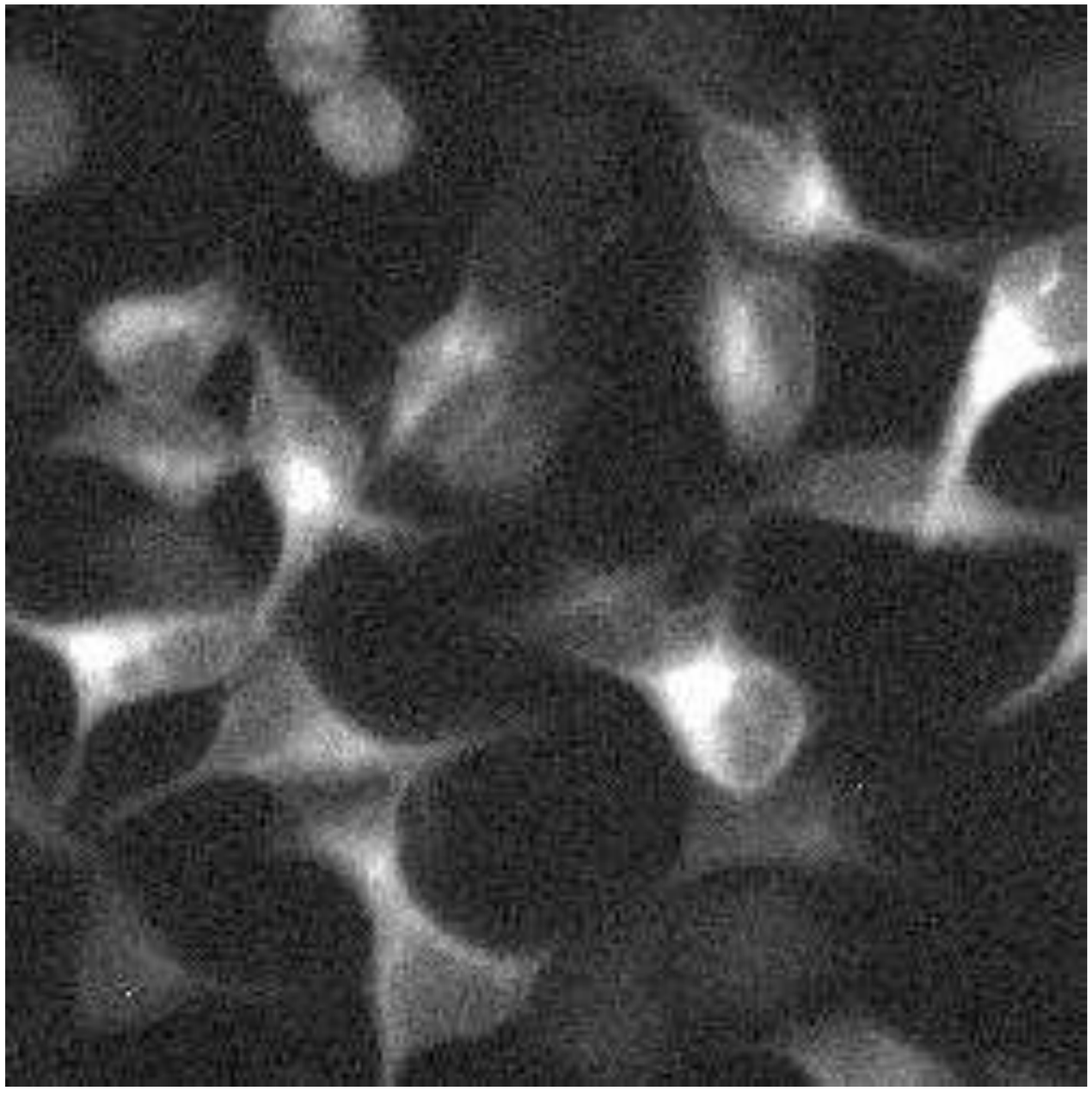
	|z|	Architecture	Image size	Training speed (min per 10^3 particles)	Total runtime (50 epochs, 1 GPU)
Filtering	1	256 ×3	128	5.7	6.3 h
Filtering	10	256 × 3	128	5.7	6.3 h
Final	10	1,024 × 3	256	43.9	31.9 h

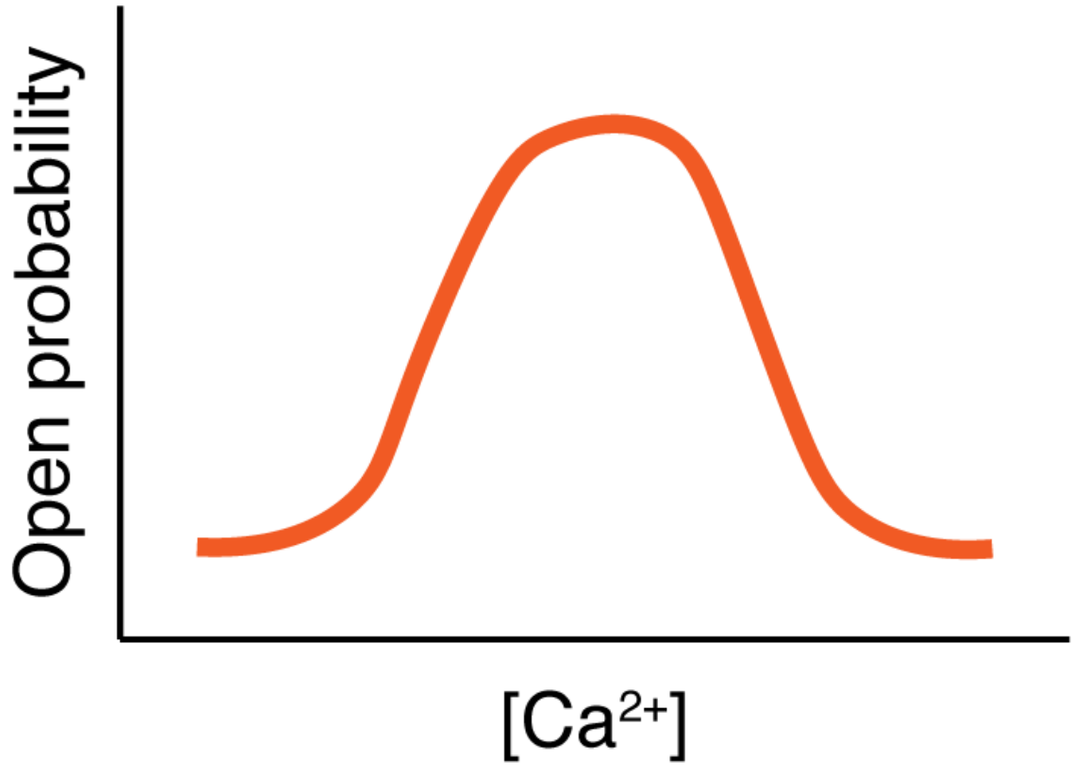
Zhong et al., Nat Methods. (2021)

Benefits of heterogeneity?

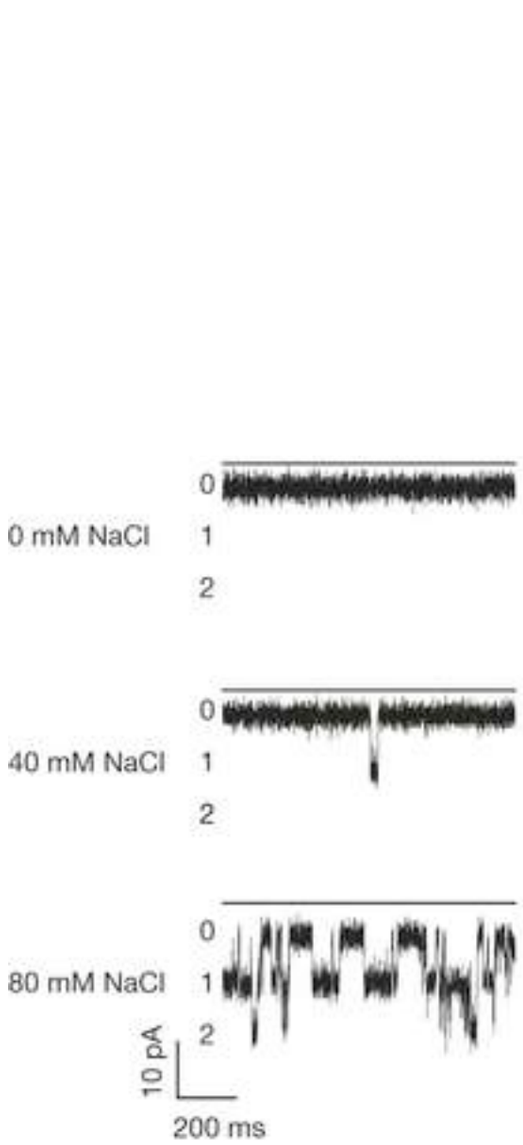
- How can you use heterogeneity to better understand the biology of your samples?
- Does your heterogeneity correlate with functional changes?
- Always test to ensure that your representative density map is actually representative of your sample, and not merely some small portion of the particles that 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?

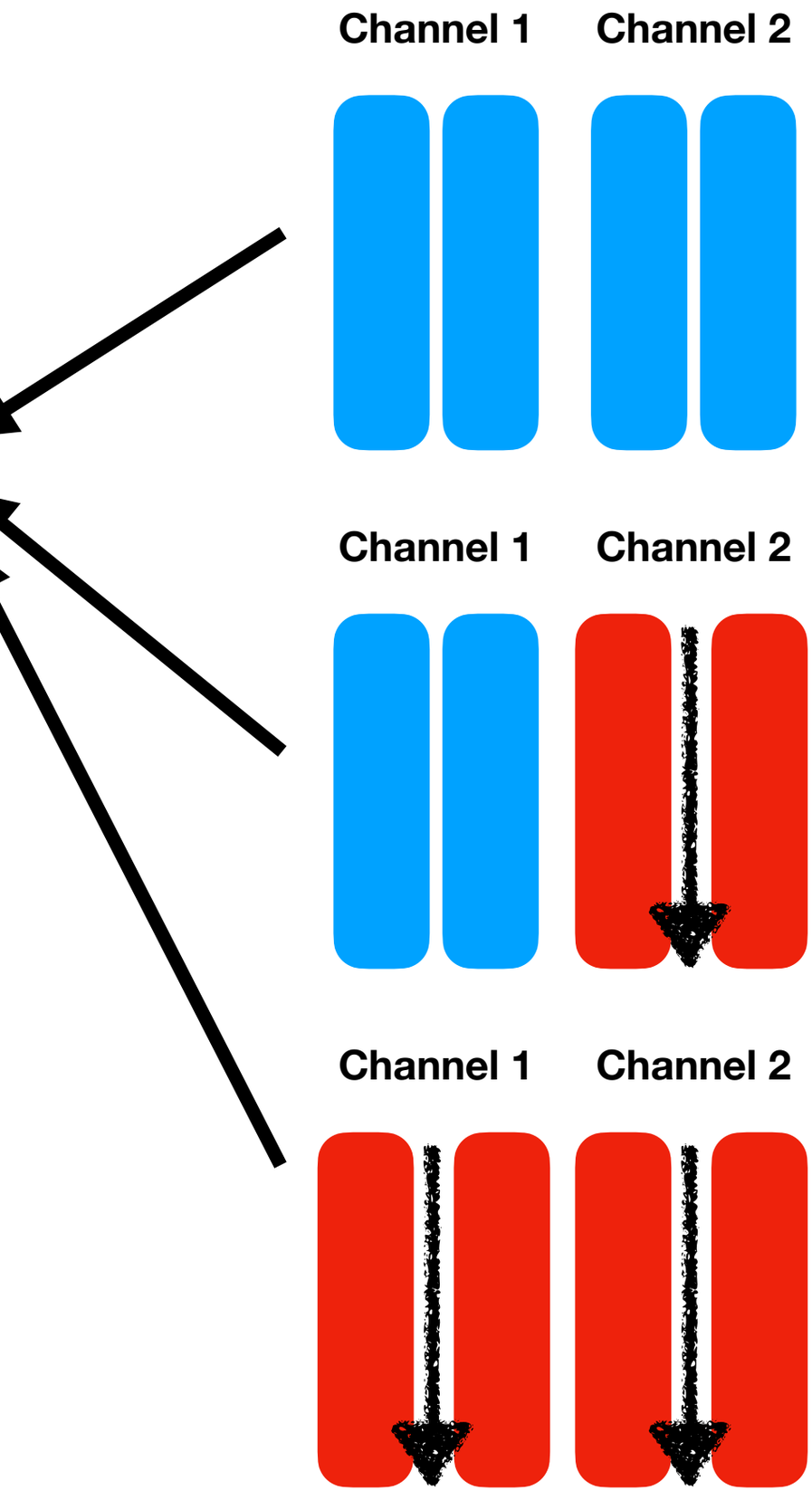
Cytosolic Ca^{2+} signaling



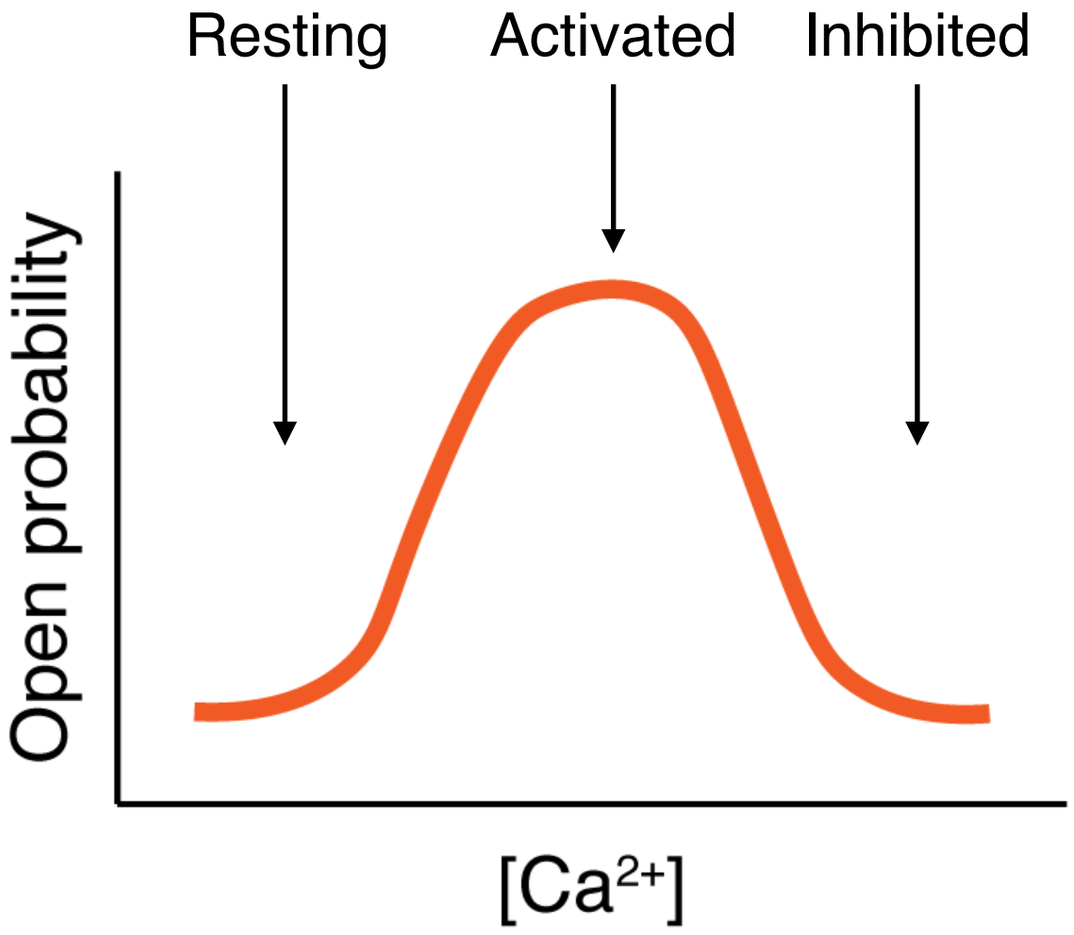


Ion channel gating

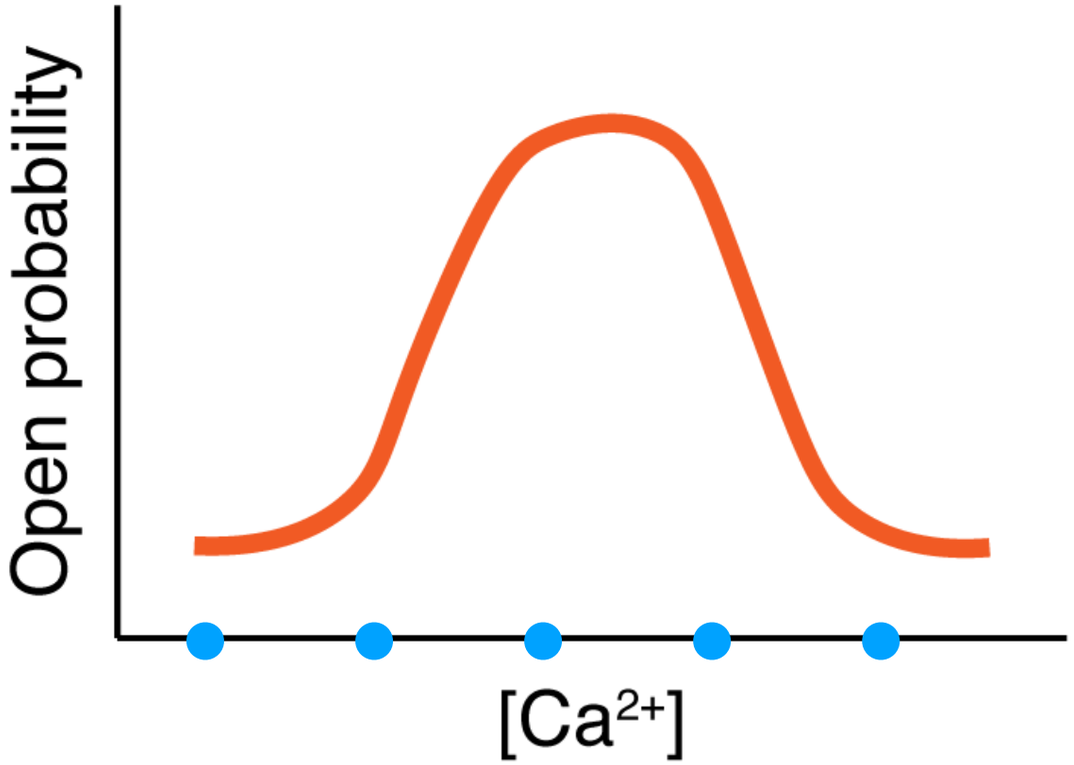


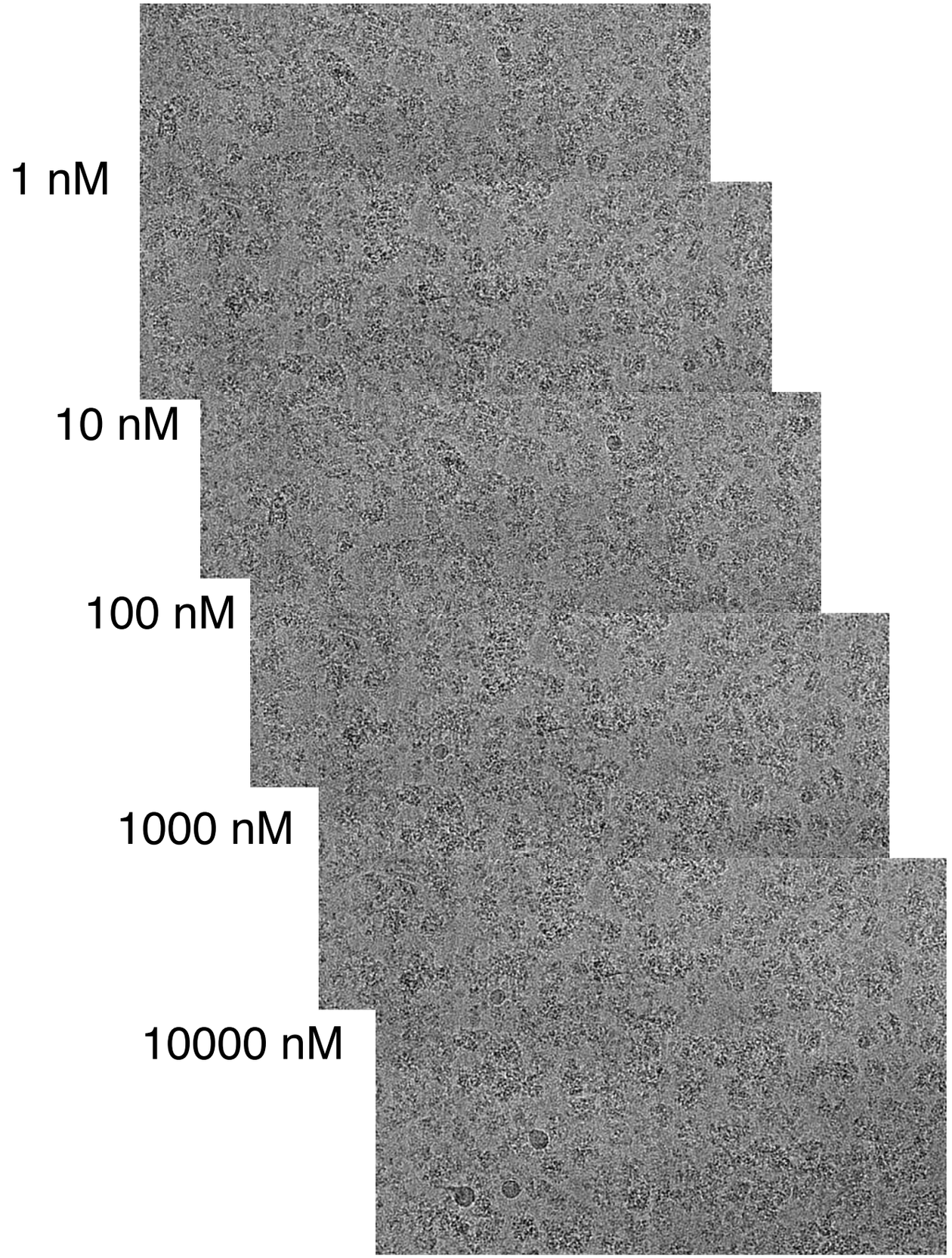


Visualizing ligand-dependent activation and inhibition

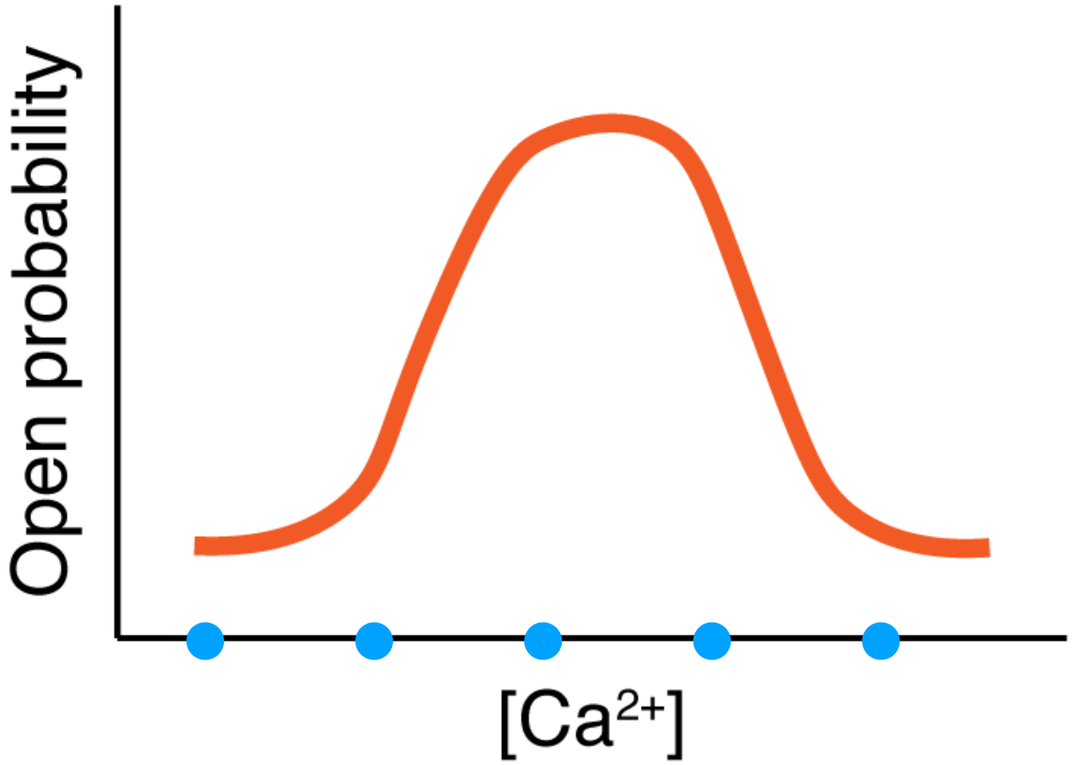


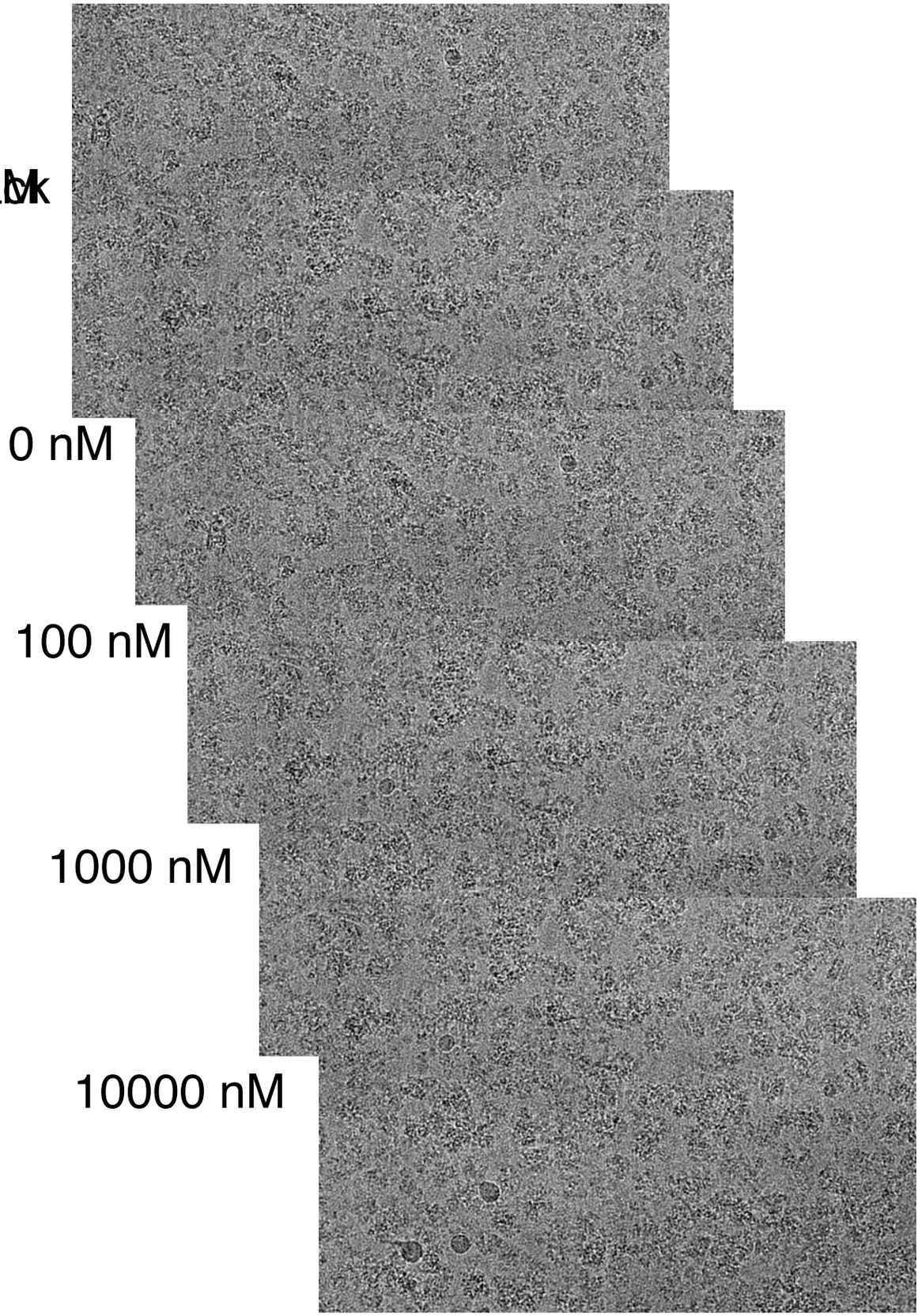
Visualizing ligand-dependent activation and inhibition





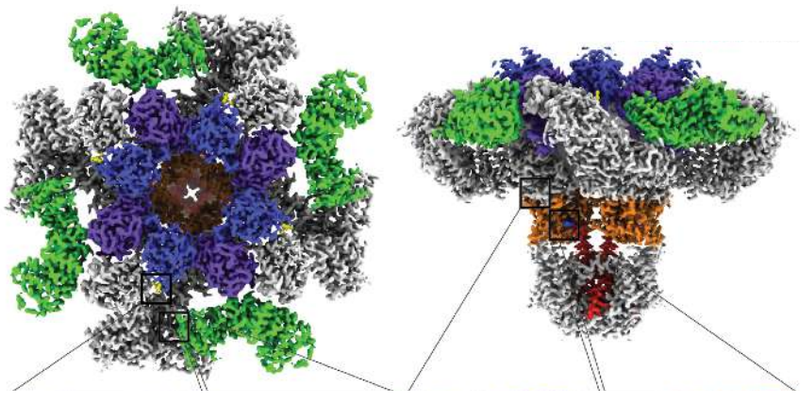
Visualizing ligand-dependent activation and inhibition



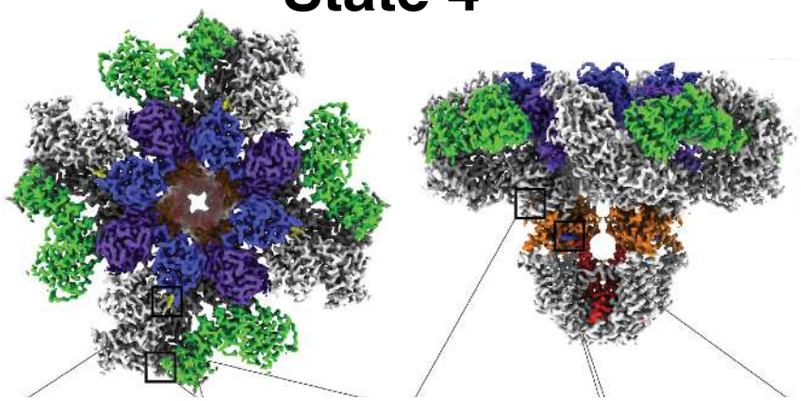


Interpreting the classes from the titration

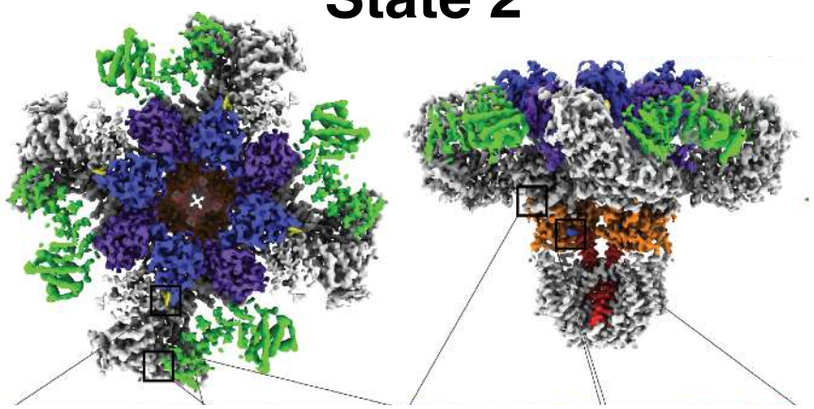
State 1



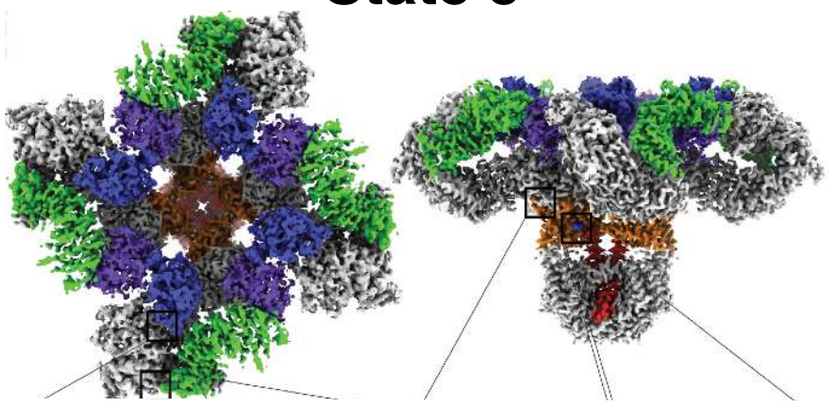
State 4



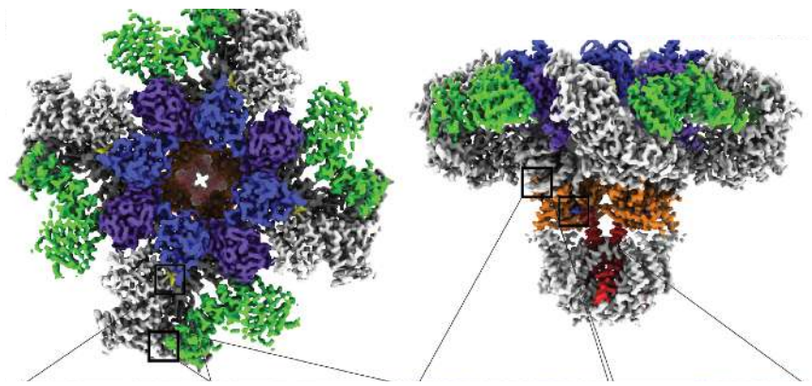
State 2



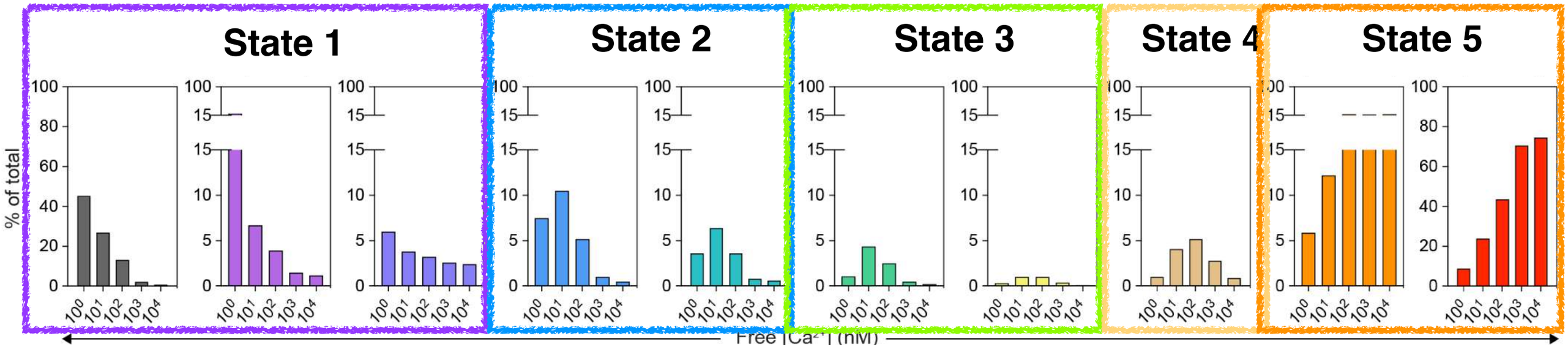
State 5



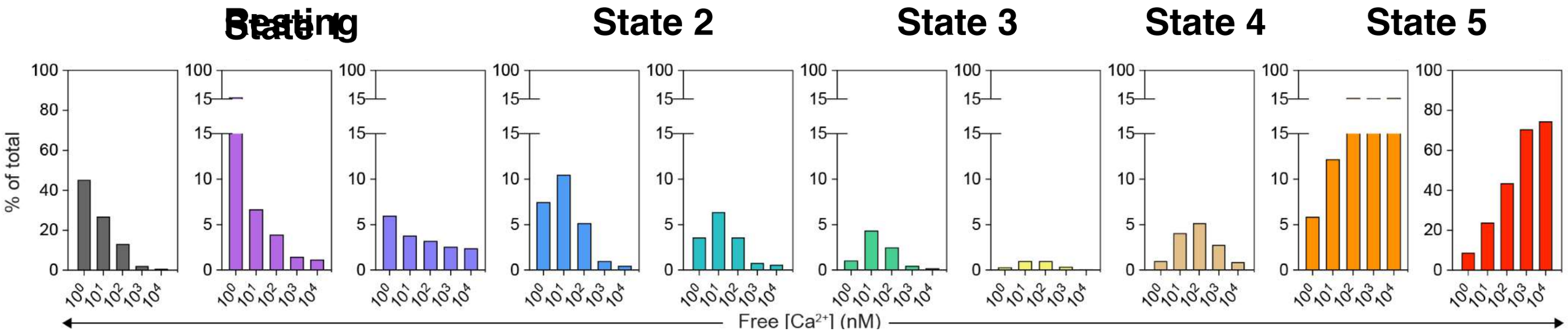
State 3

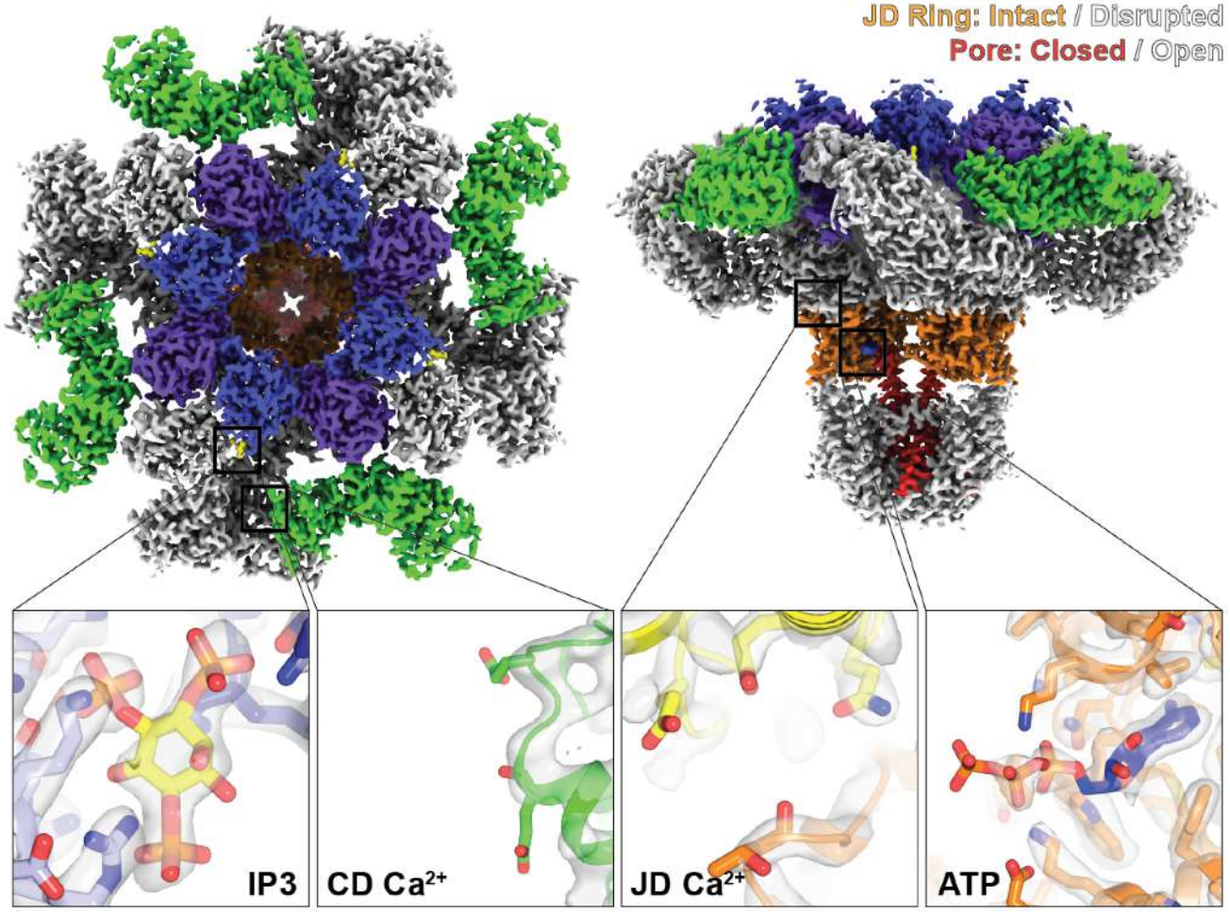


Ca^{2+} dependence of conformational state

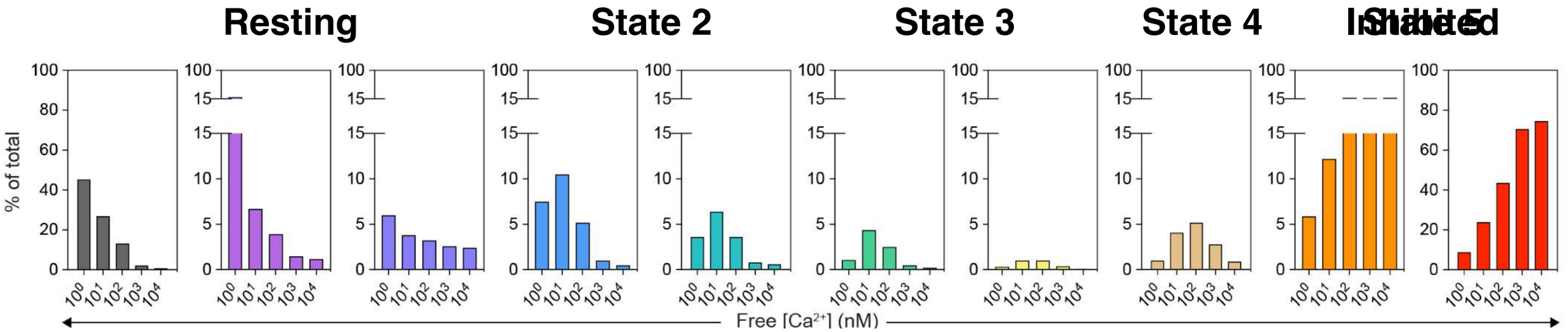


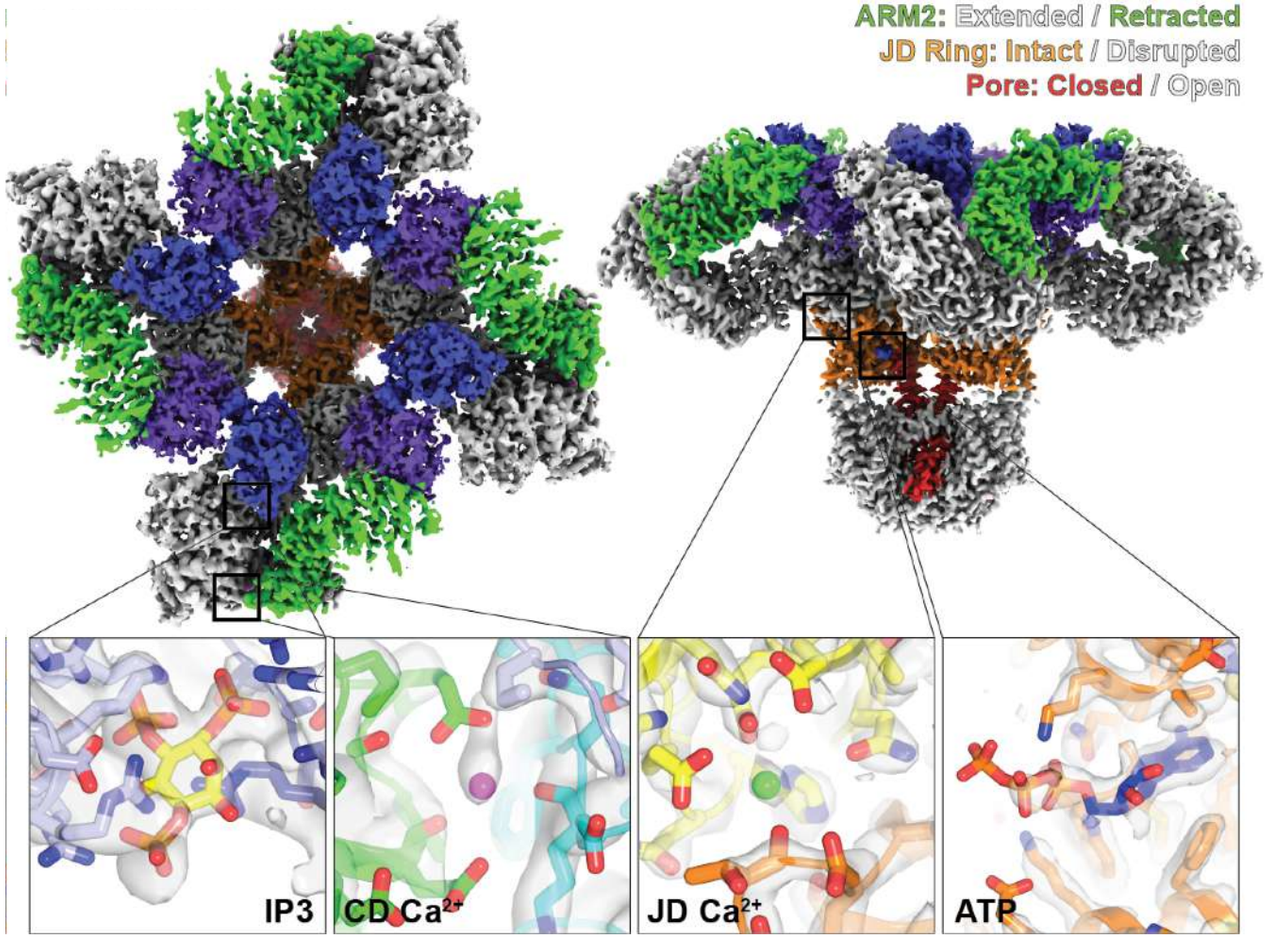
Ca^{2+} dependence of conformational state



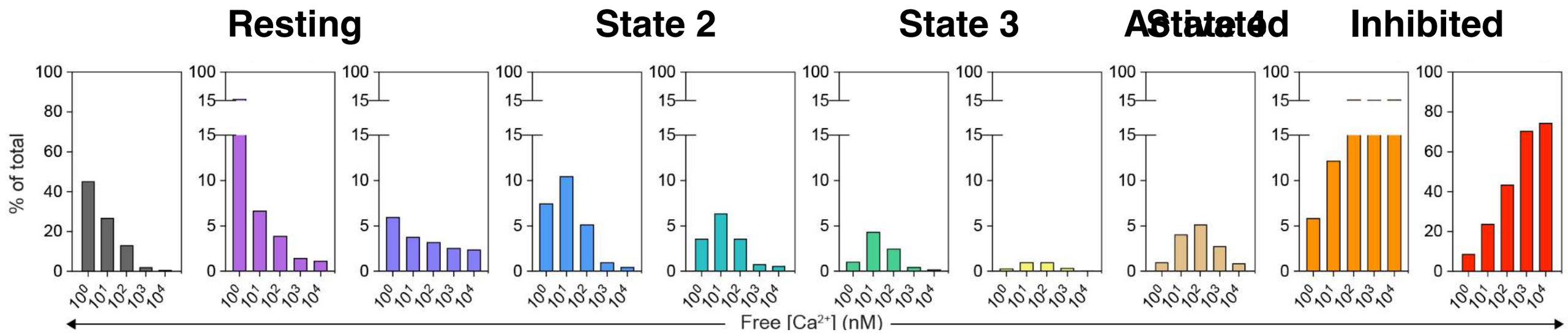


Ca^{2+} dependence of conformational state

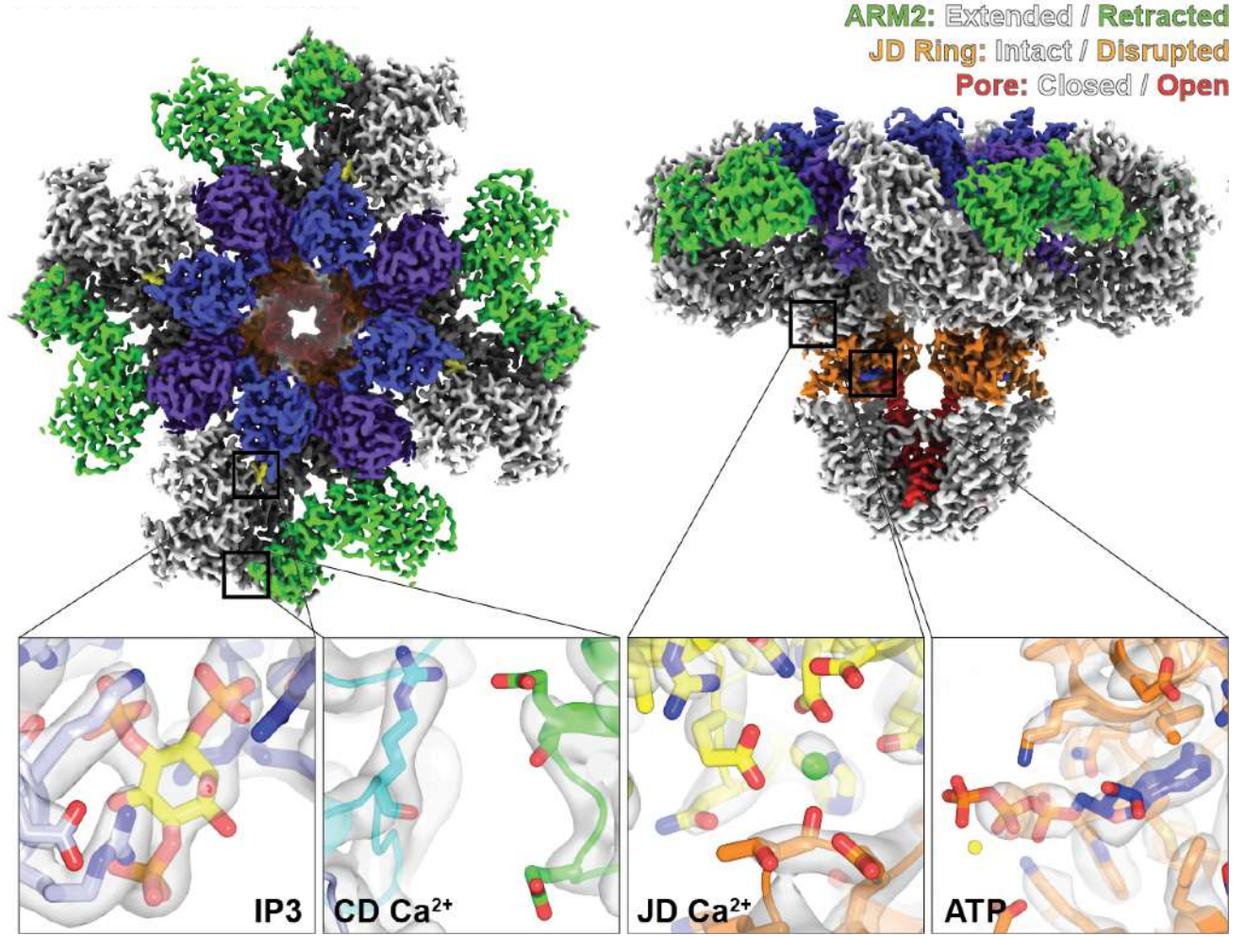


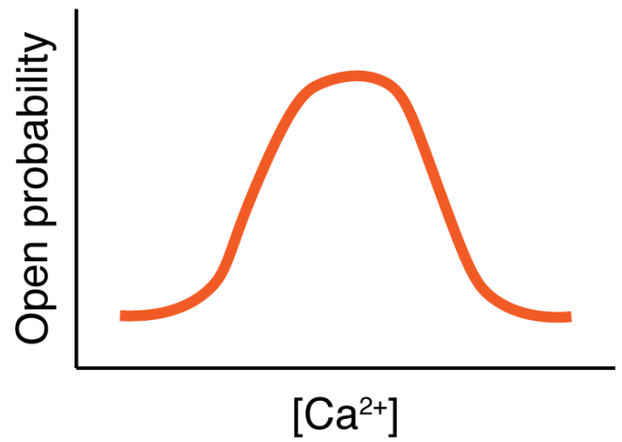


Ca^{2+} dependence of conformational state

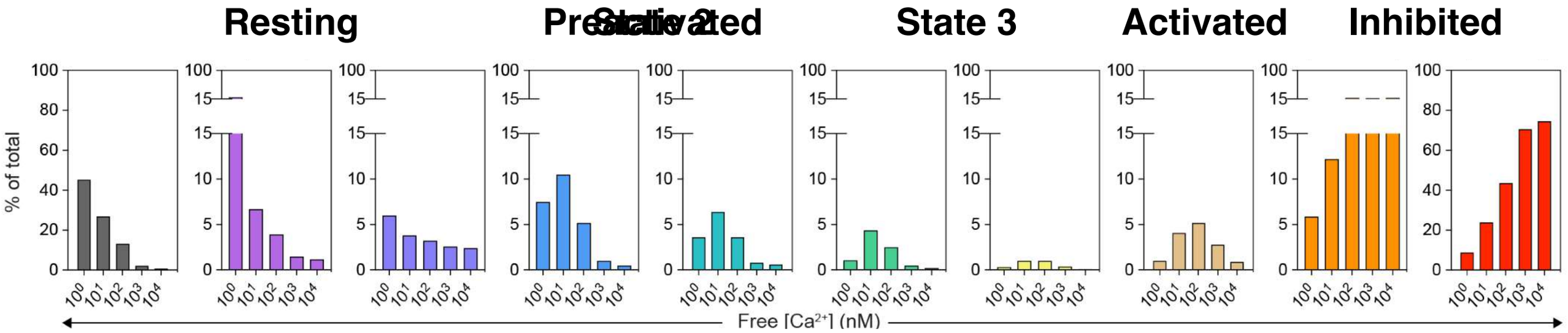


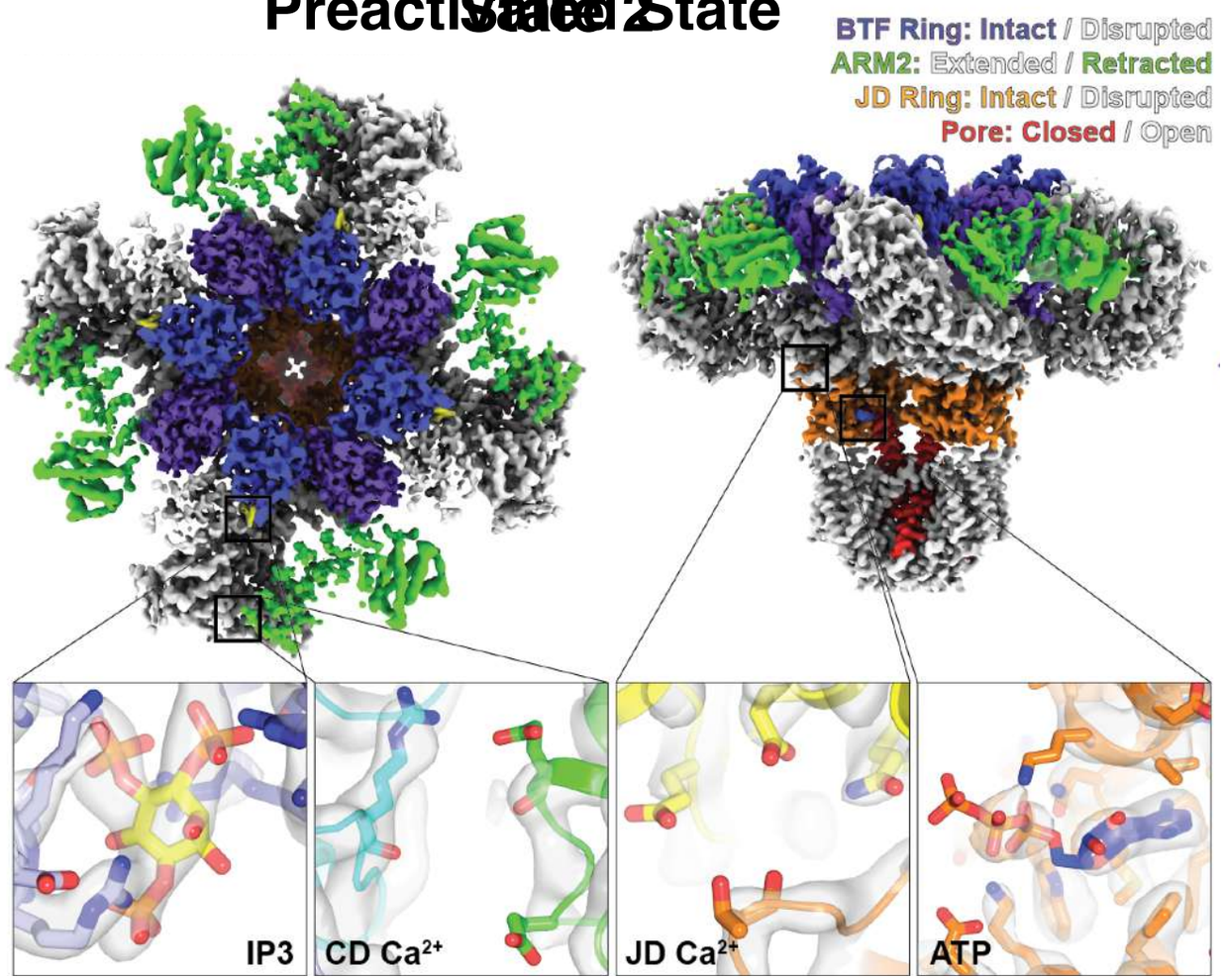
Activated State



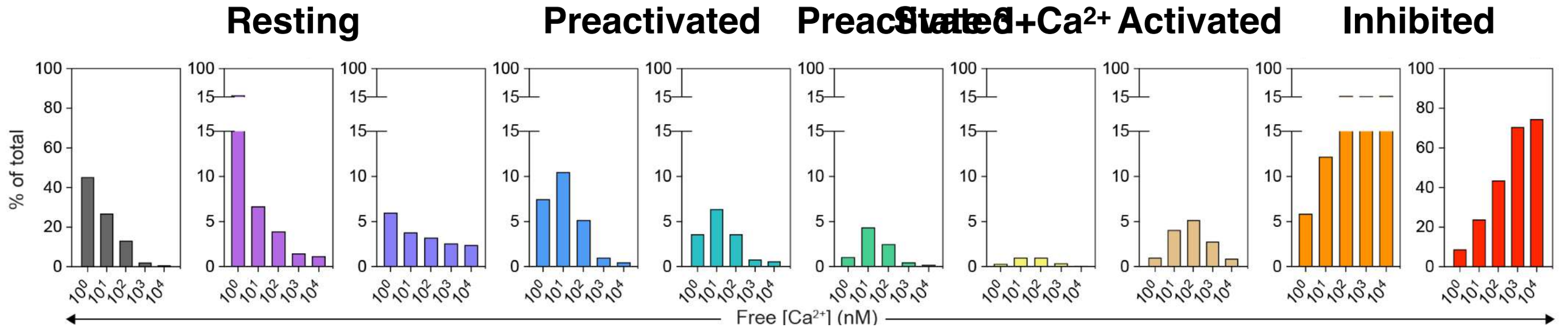


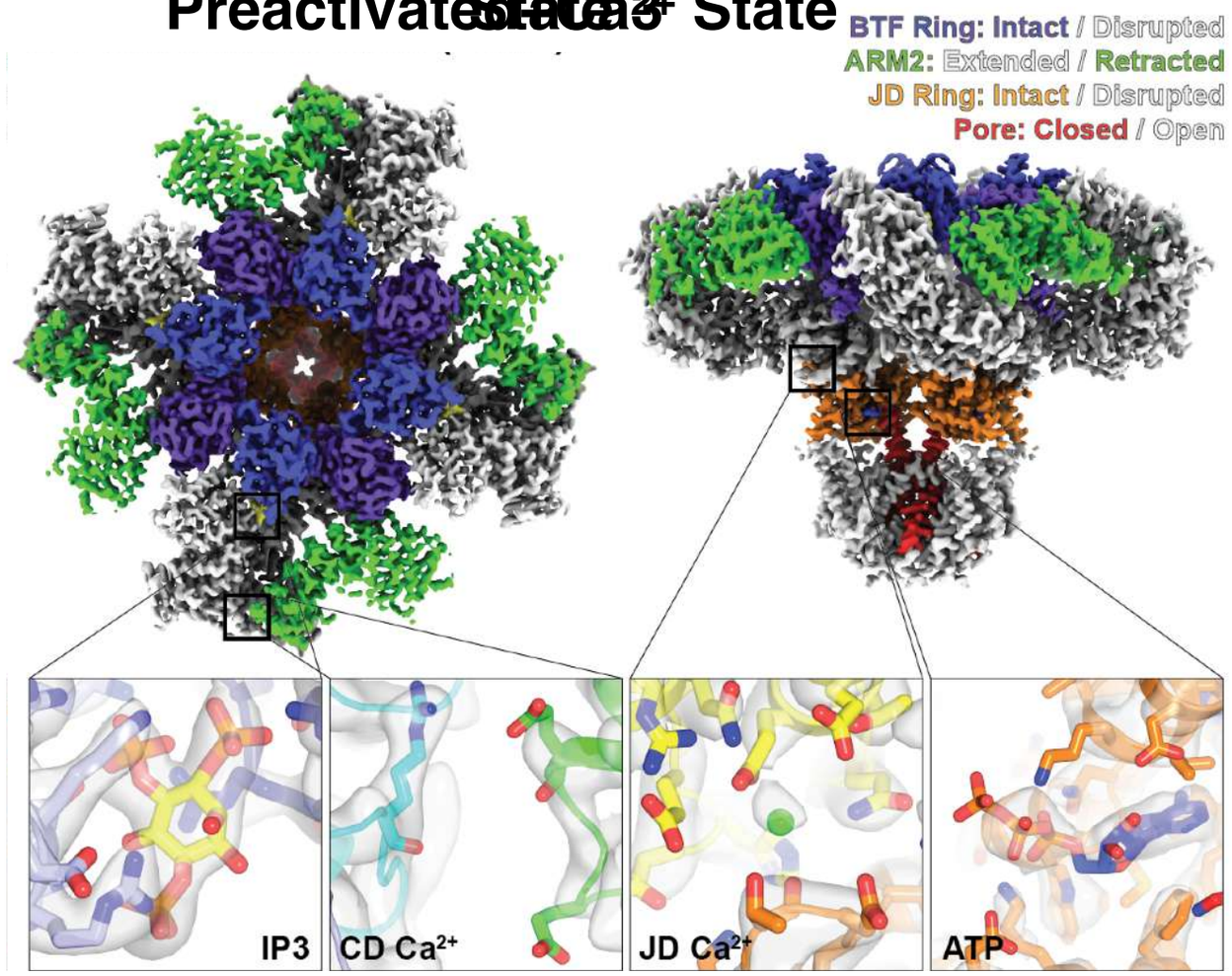
Ca^{2+} dependence of conformational state





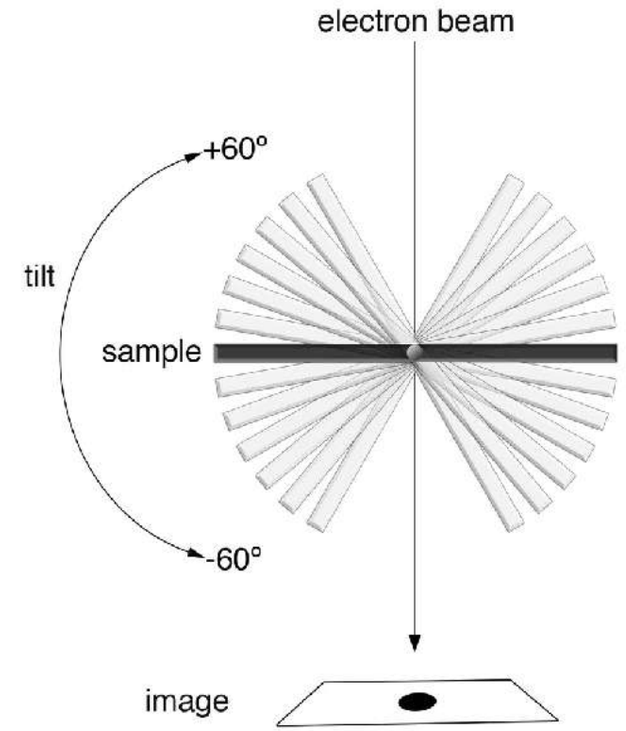
Ca^{2+} dependence of conformational state

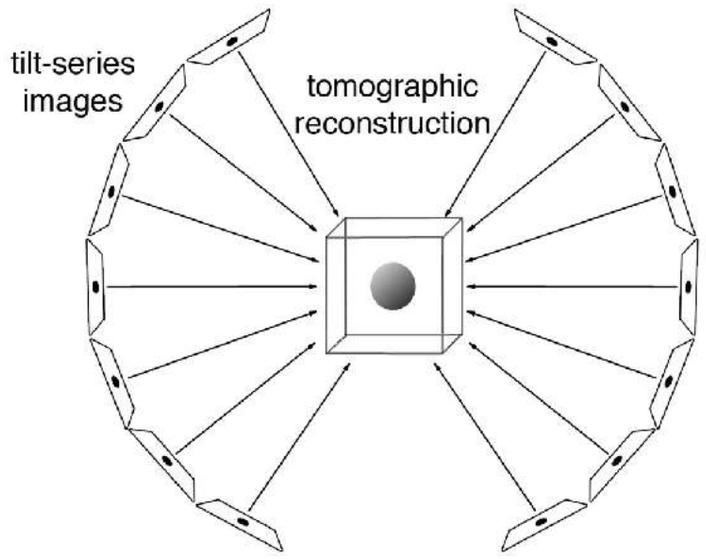




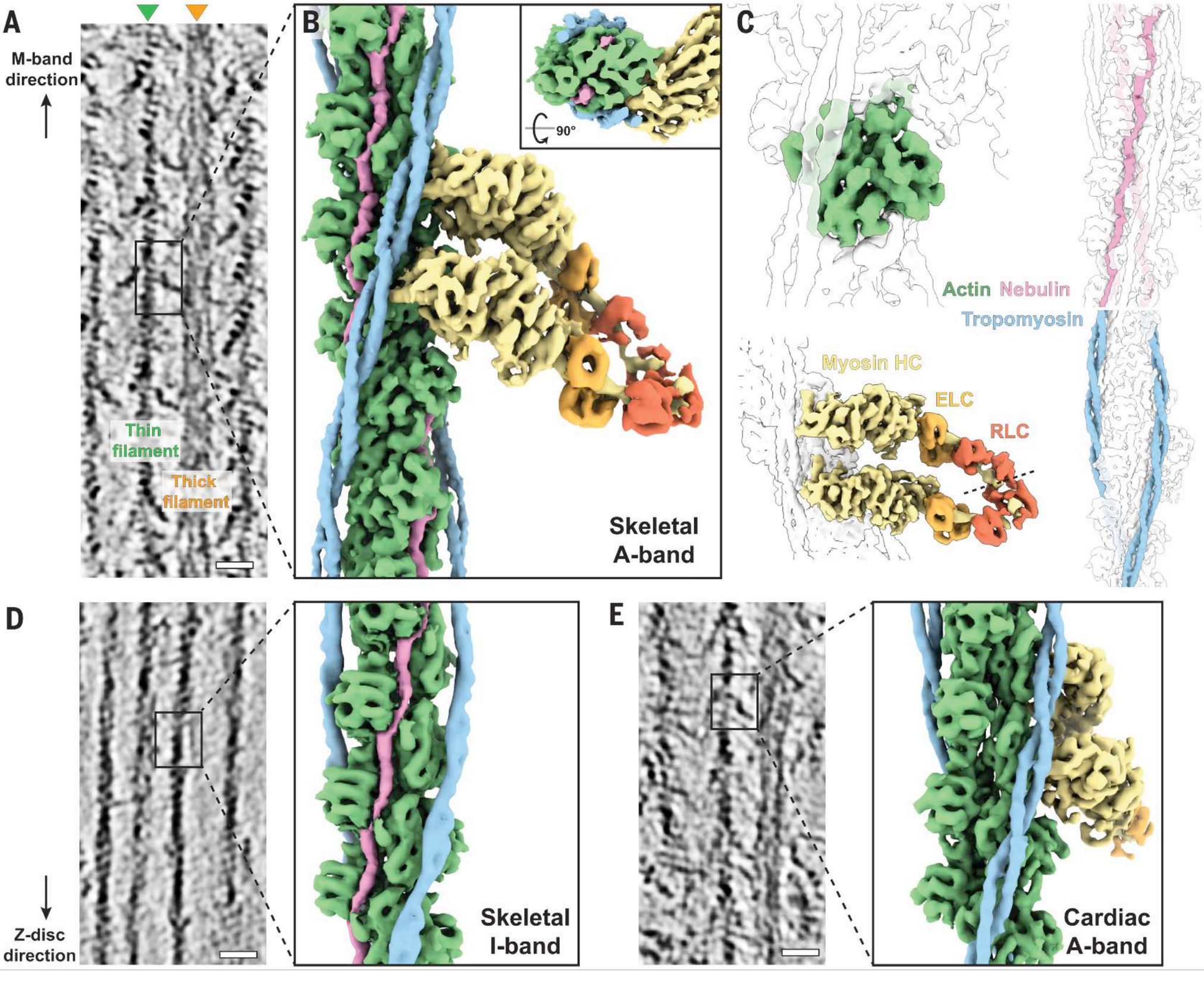
In cellulo heterogeneity?

- Cryo-electron tomography
- Many images of the same specimen
- Images are collected at different tilt angles to allow different views of the sample
- Images can then be reconstructed into volumes
- Samples can be purified, crude or recorded from samples in vitrified cells



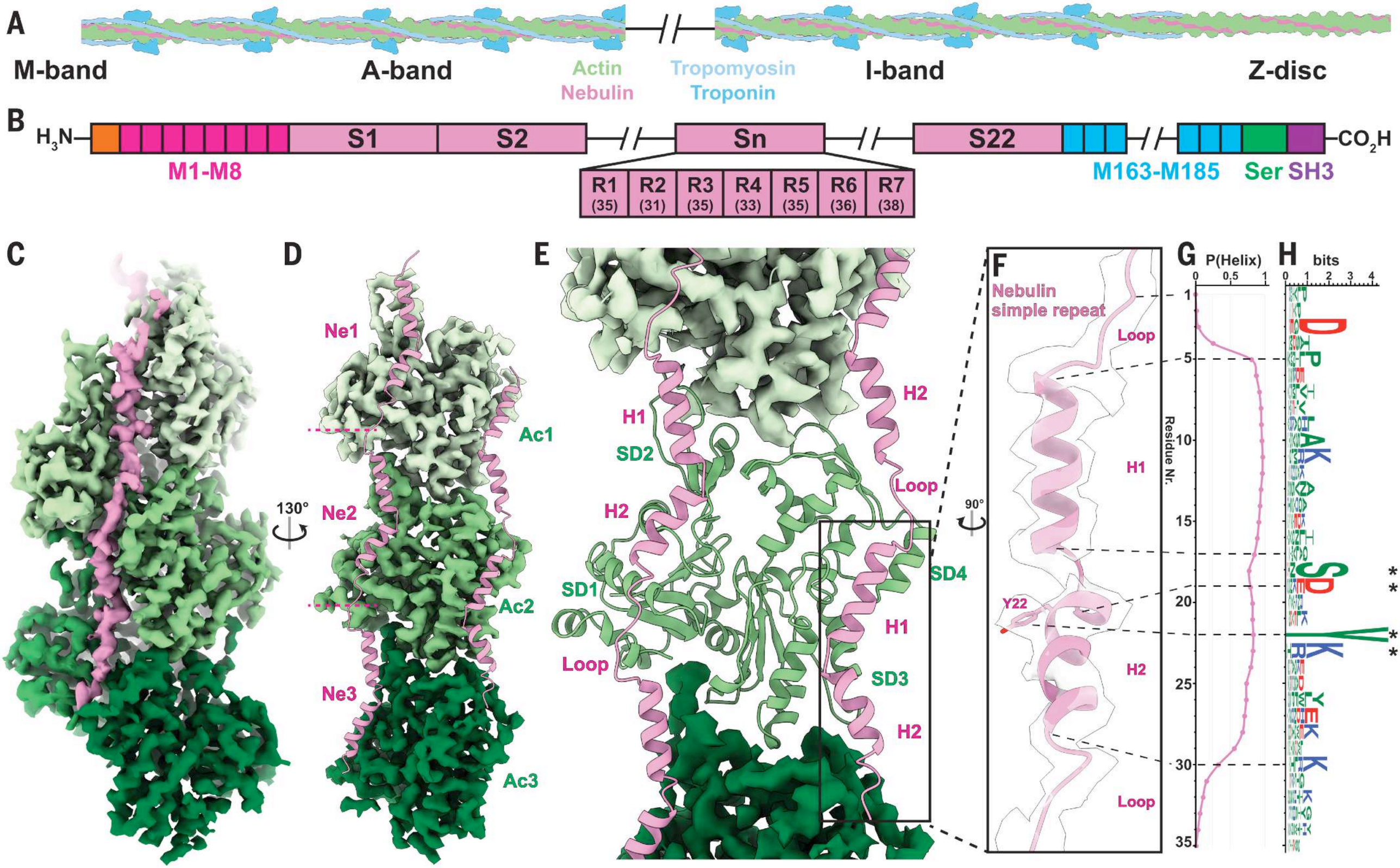


Sub-tomogram averaging



Wang et al, Science. (2022)

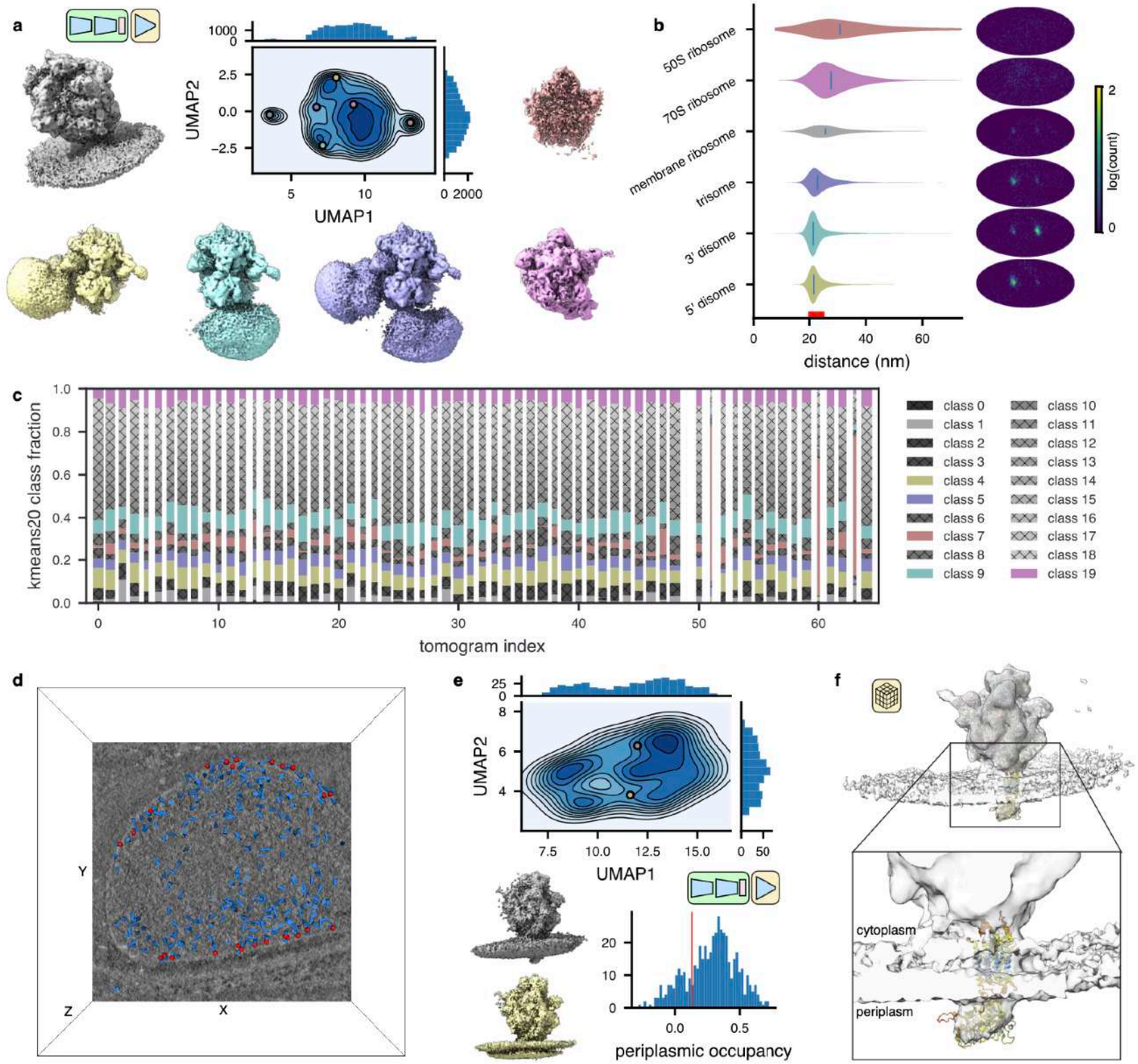
Sub-tomogram averaging



Wang et al, Science. (2022)

Heterogeneity in sub-tomogram averages

- TomoDRGN is an application of cryoDRGN to tomographic data, allowing one to visualize complex landscapes of complexes in cells



Powel and Davis, bioRxiv. (2023)

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/