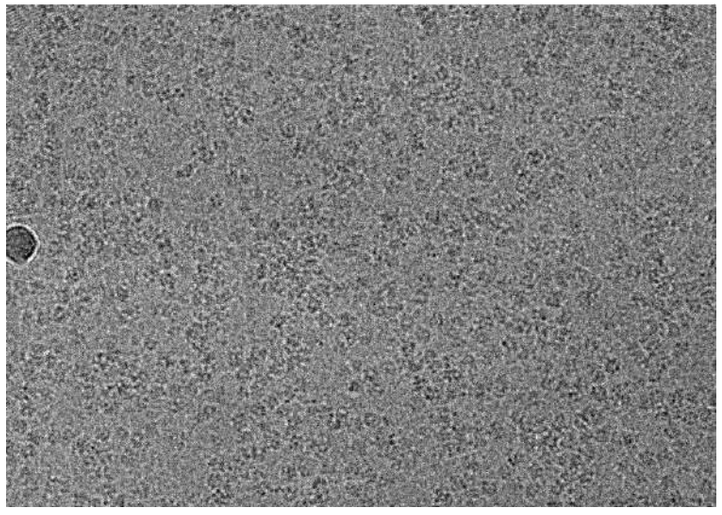
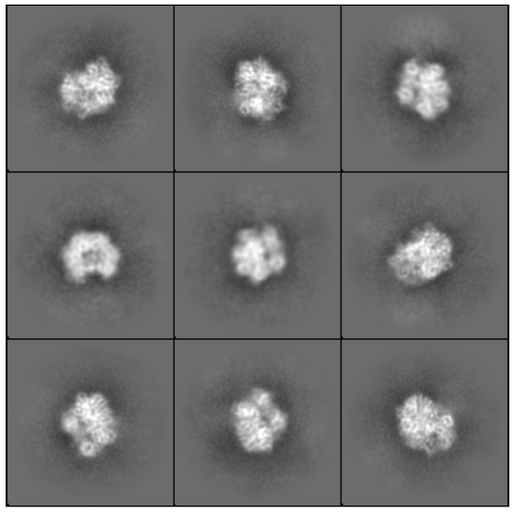
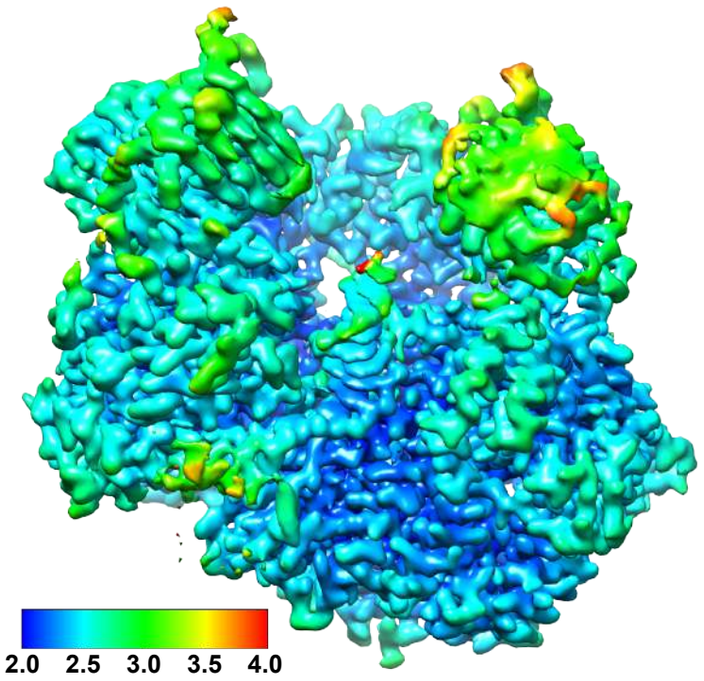
Interpretation and Limitations of Single-Particle Cryo-EM Analysis

Rich Hite
Memorial Sloan Kettering Cancer Center

Cryo-EM structure determination



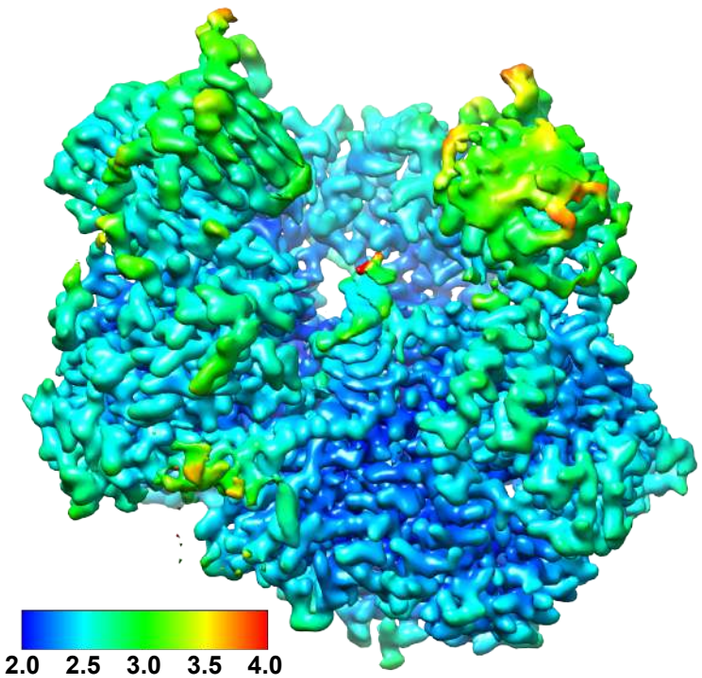




Casteneda*, Schrecker* *et al*, NSMB. 2022

You have a density map, now what?

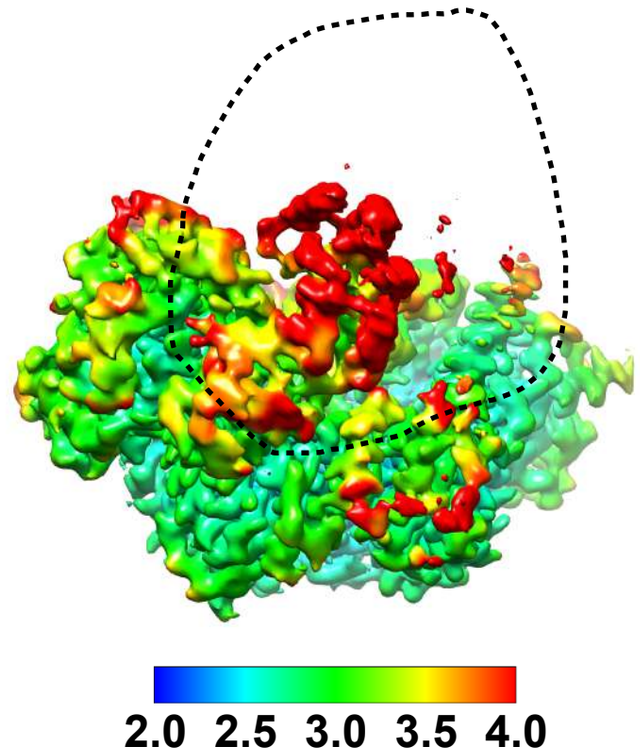
- **First important question:**
  - What is the question that you are using cryo-EM structure determination to answer?
- **Second question:**
  - What evidence do you have that supports the hypothesis that your structure faithfully represents the protein or complex that wish to study?
- **Third question:**
  - Is the domain that you are interested in studying well-resolved in all, some or none of your maps?
  - If so, at what resolution?
  - Can you see different conformations the domain?



Casteneda*, Schrecker* *et al*, NSMB. 2022

You have a density map, now what?

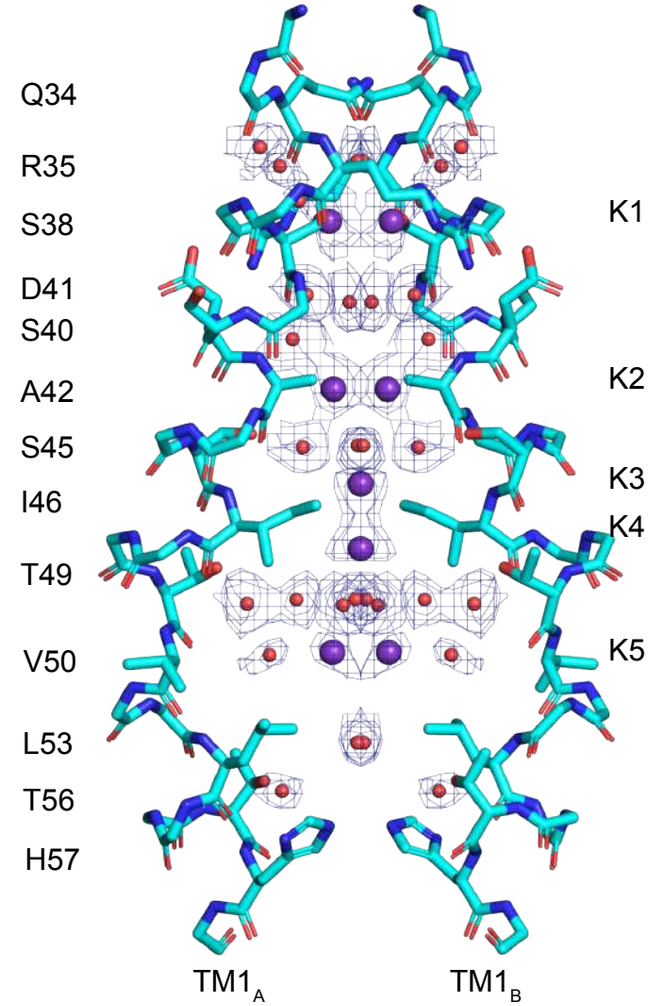
- **First important question:**
  - What is the question that you are using cryo-EM structure determination to answer?
- **Second question:**
  - What evidence do you have that supports the hypothesis that your structure faithfully represents the protein or complex that wish to study?
- **Third question:**
  - Is the domain that you are interested in studying well-resolved in all, some or none of your maps?
  - If so, at what resolution?
  - Can you see different conformations the domain?

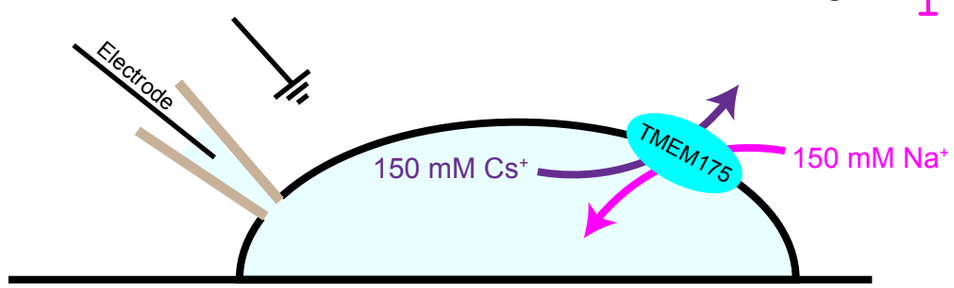


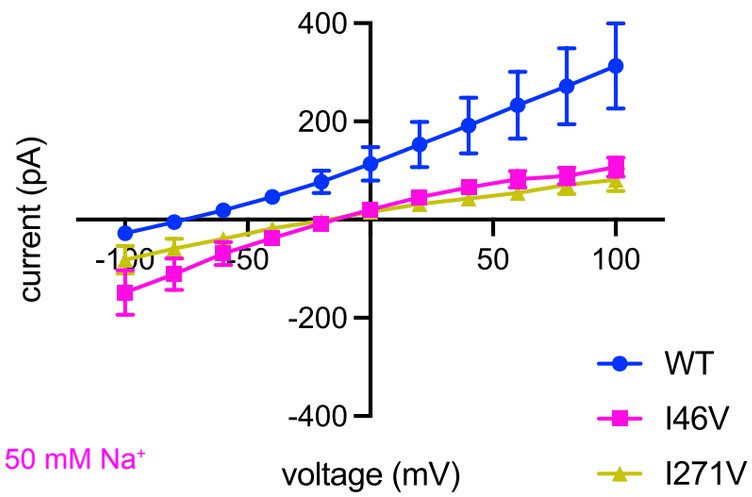
Casteneda*, Schrecker* *et al*, NSMB. 2022

You have a density map, now what?

- Functional assessment
- Second question - What orthogonal approaches can you employ to support the hypothesis that your density map faithfully represents your protein of interest?
  - Can you reconstitute your protein or complex in an *in vitro* functional assay with the same purified components that are used for structure determination?
- Does the domain architecture agree with previous data?



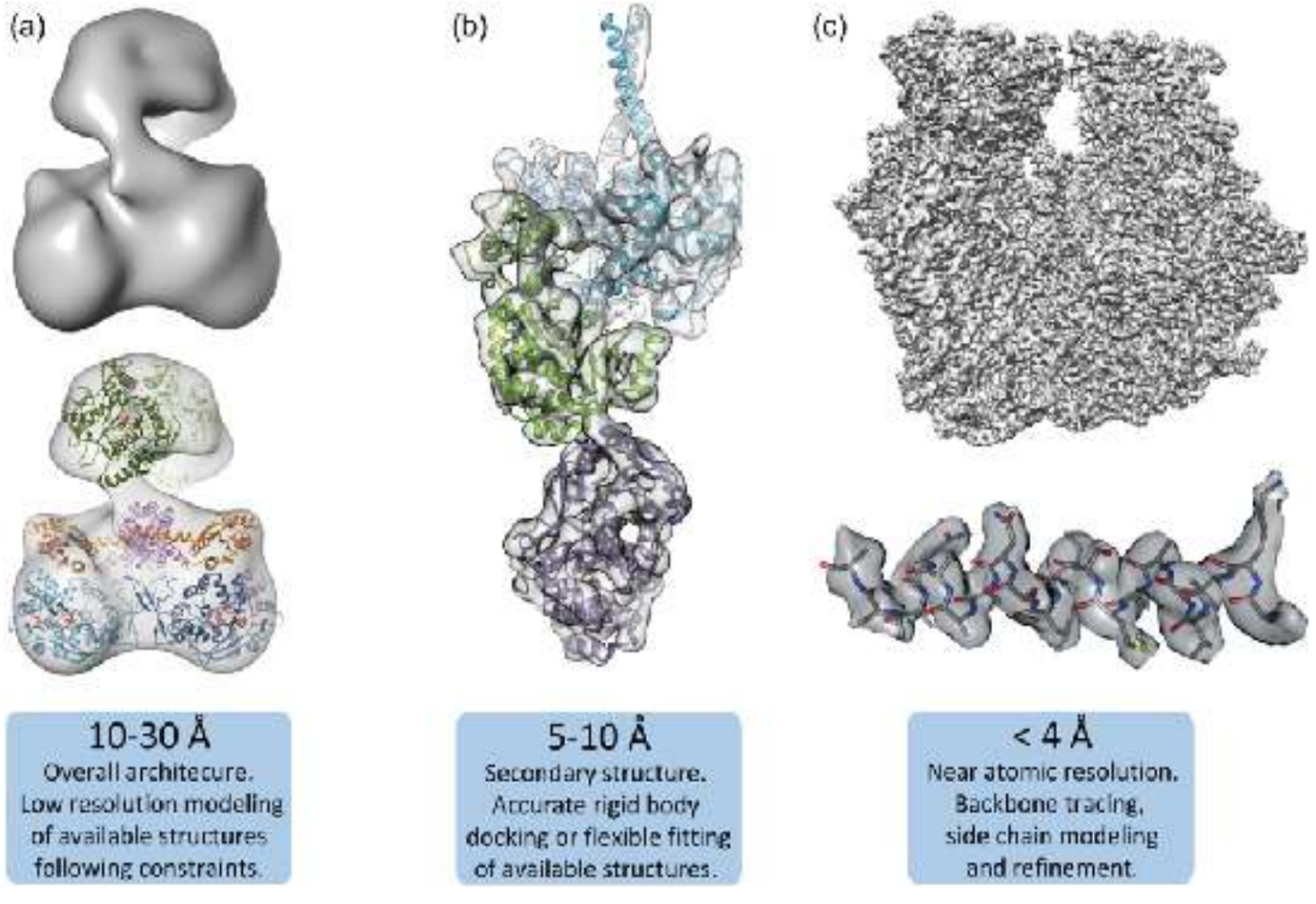




Oh *et al*, biorxiv. (2021)

Resolution - what does it mean?

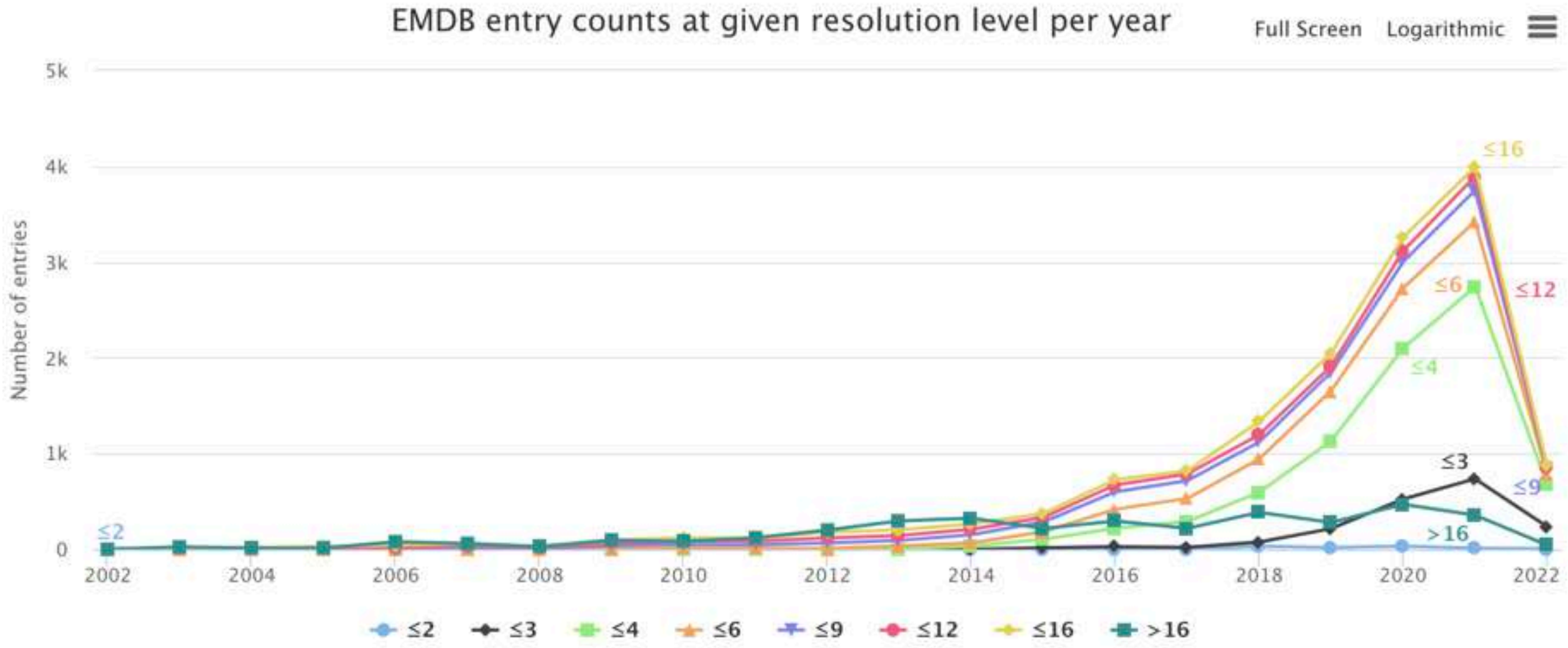
- Resolution corresponds to the distance between two objects that can be reliably separated in the map
- In EM, resolution is estimated by the Fourier shell correlation (FSC)
- FSC is determined by comparing the cross correlation between two independently calculated maps in Fourier space resolution shells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

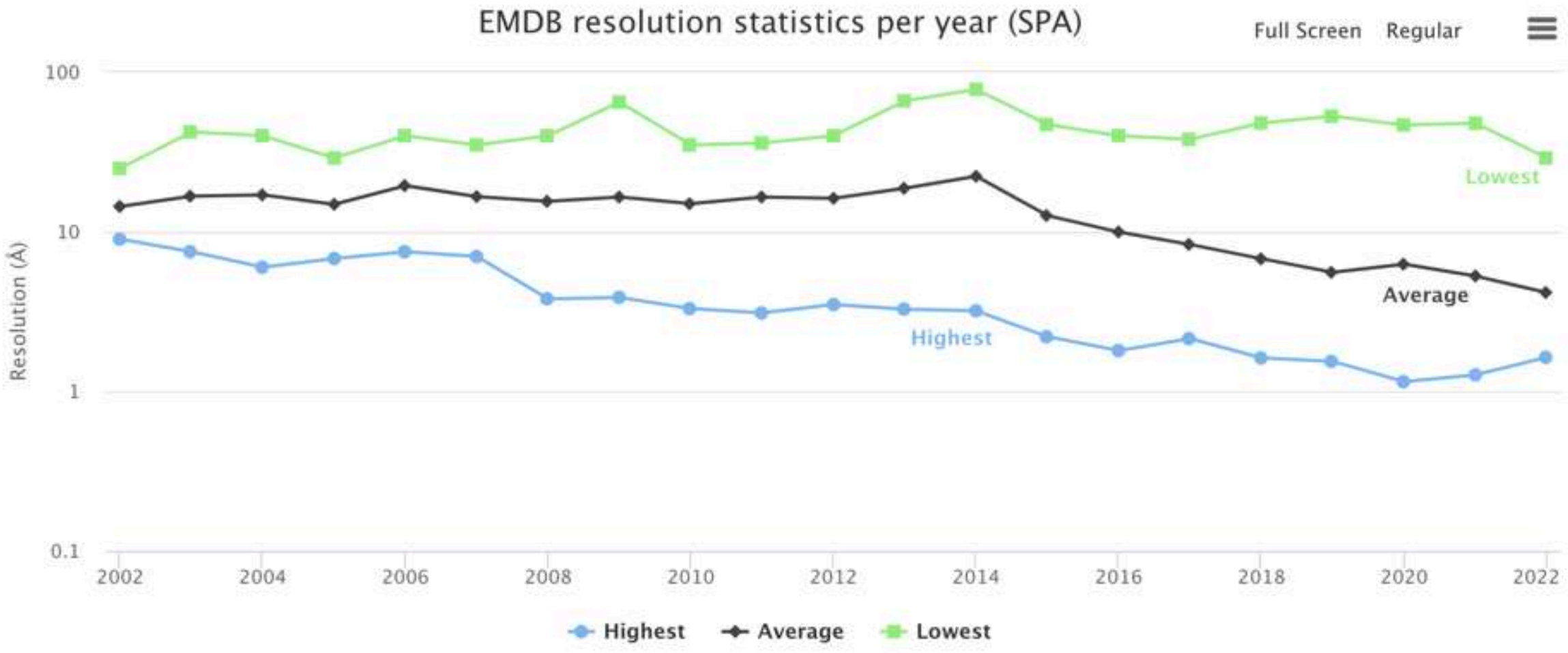
- What is the meaning of 4 Å map? A 3 Å map?
- How does the resolution influence our interpretation?
- What do we gain from higher resolution reconstructions?
- How can we improve the resolution of reconstruction?



EMDB 2022.03.15

Resolution - what does it mean?

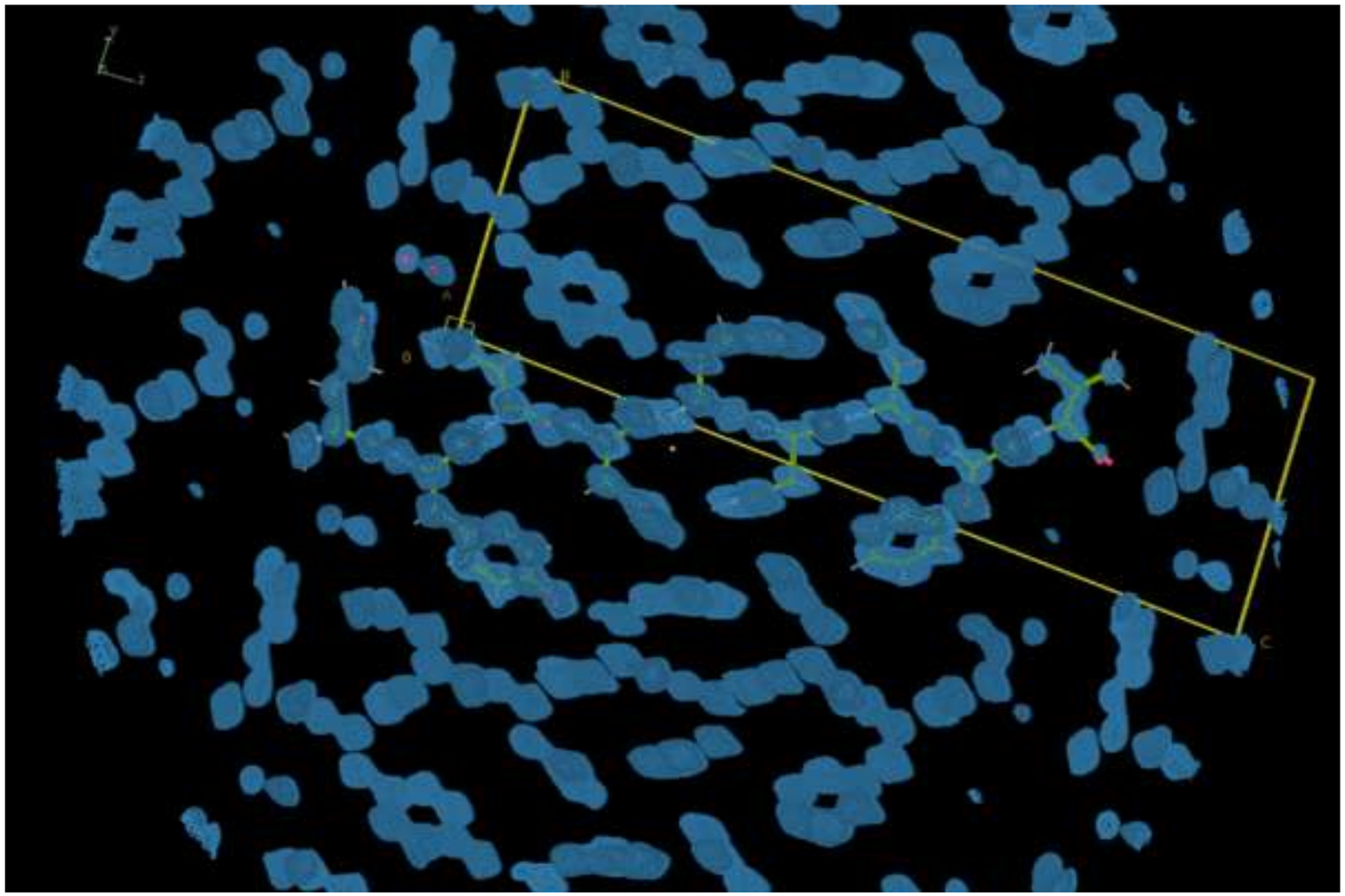
- What is the meaning of 4 Å map? A 3 Å map?
- How does the resolution influence our interpretation?
- What do we gain from higher resolution reconstructions?
- How can we improve the resolution of reconstruction?



EMDB 2022.03.15

Resolution - what does it mean?

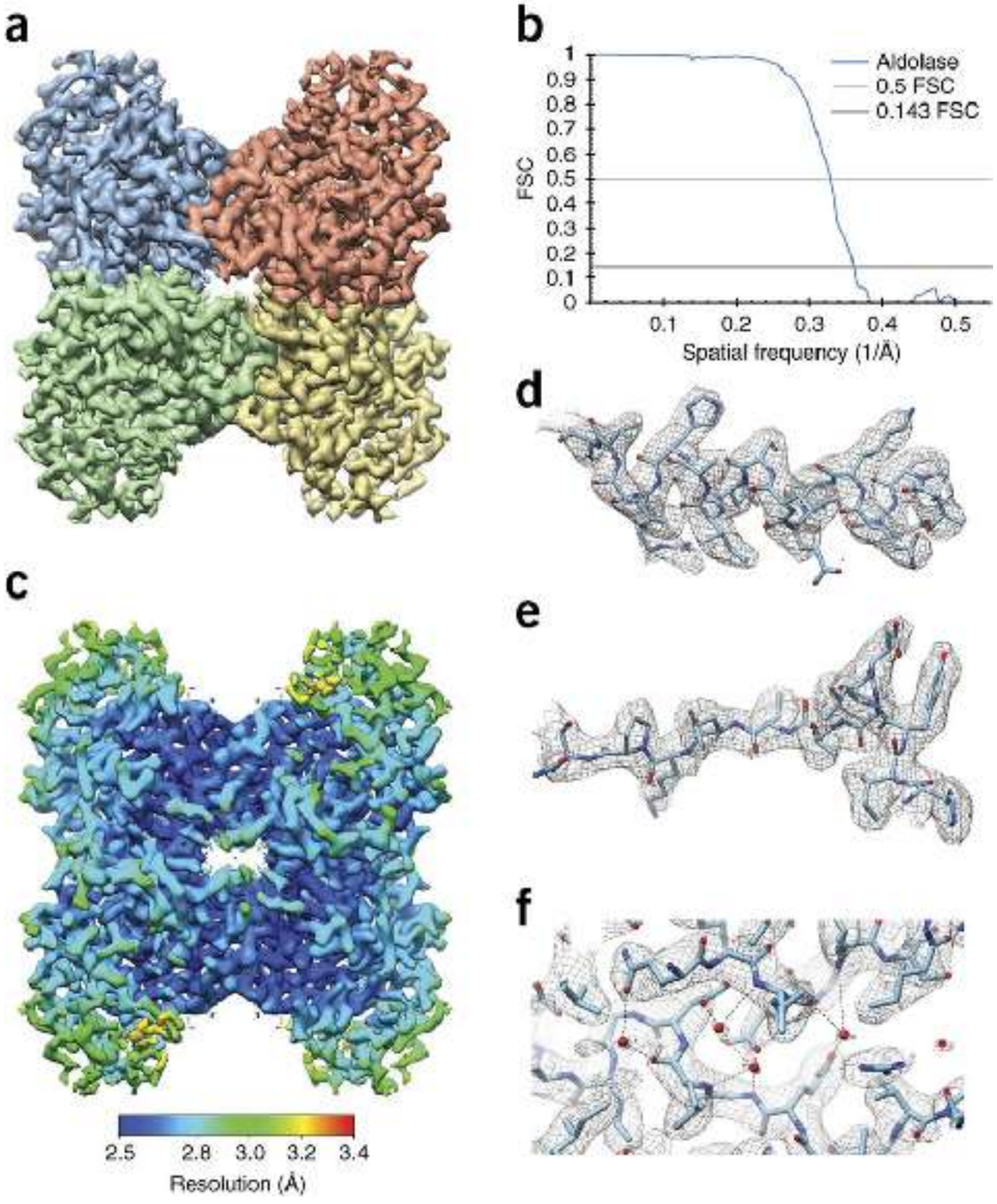
Sub-1 Å microED of bank vole prion protein 168-176 QYNNQNNFV



EMDB 7017
Gallagher-Jones M et al, NSMB **25** 131-134 (2018)

Resolution - what does it mean?

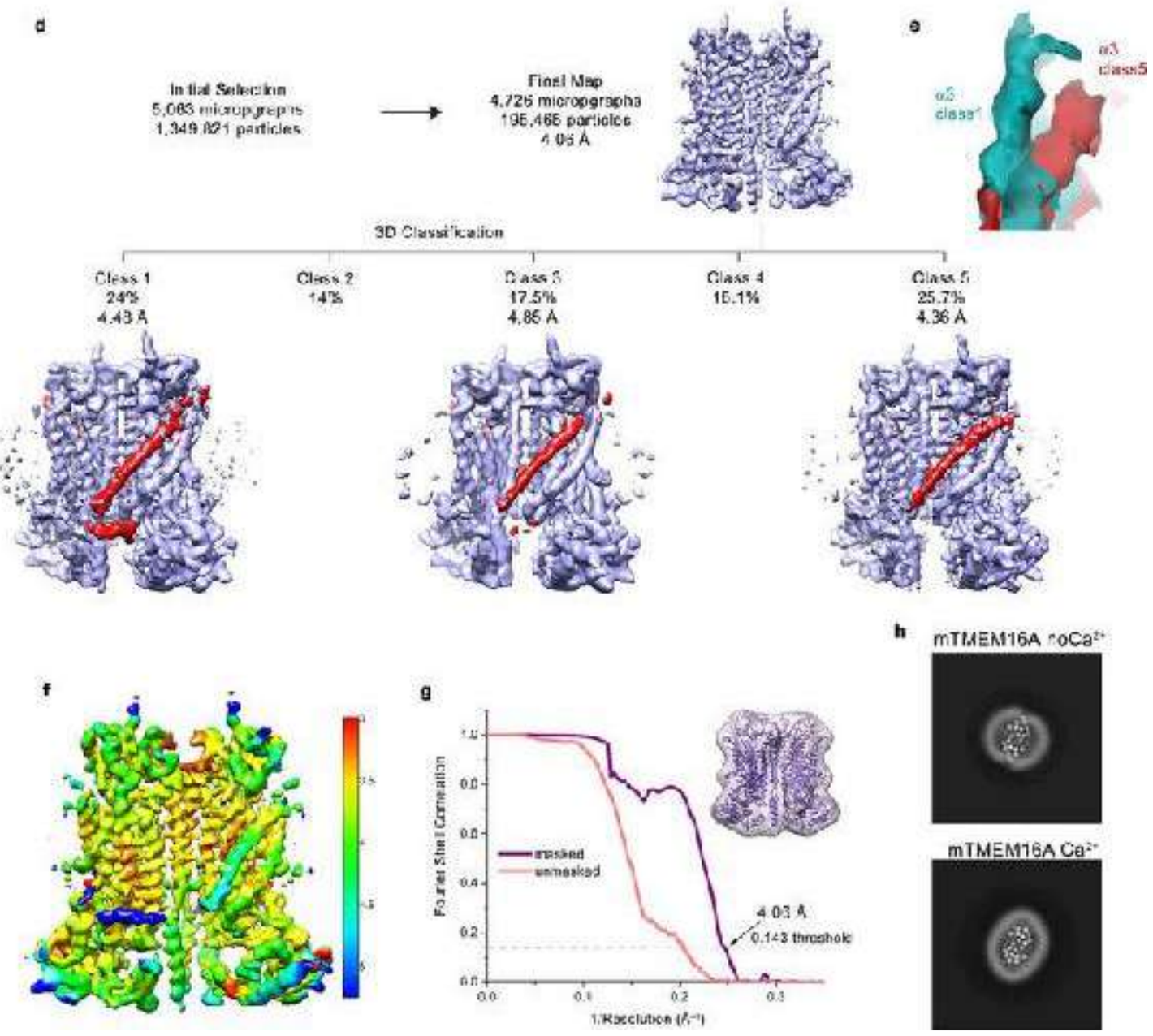
2.6 Å single particle reconstruction of aldolase from 200 keV



EMDB 8743
Herzik et al., *Nat. Methods* **14** 1075-1078 (2017)

Resolution - what does it mean?

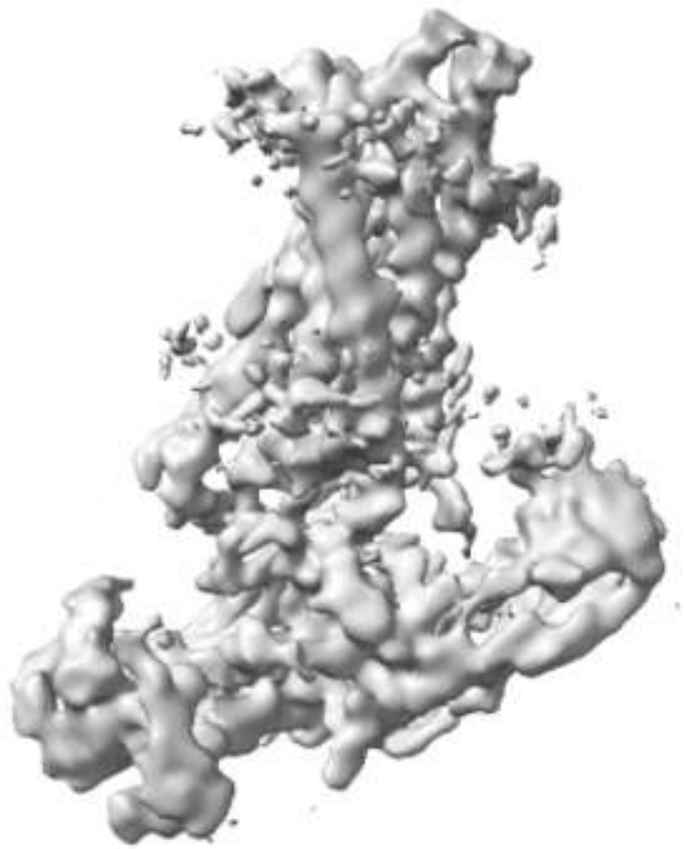
4.0 Å single particle reconstruction of TMEM16A (ion channel)



EMDB 3860
Paulino et al., *Nature* **552** 421-425 (2017)

Resolution - what does it mean?

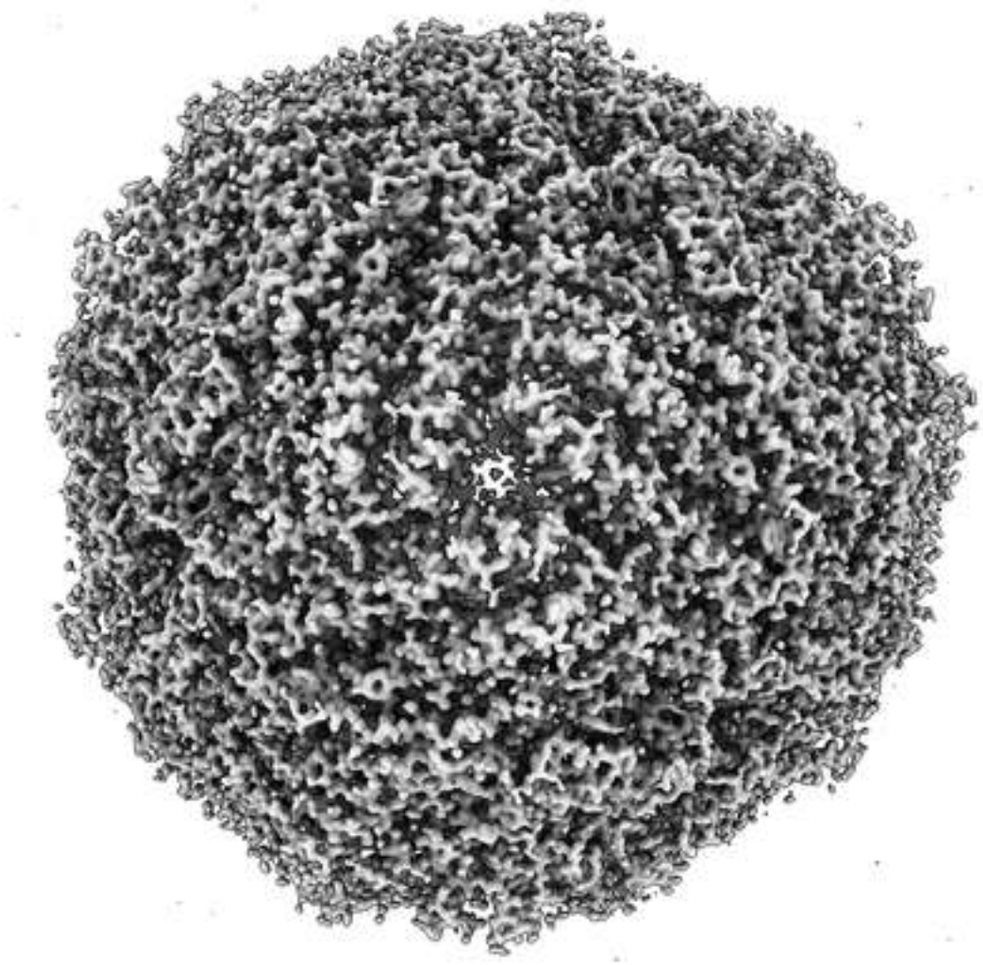
1. What are common features of samples that yield high-resolution cryo-EM reconstructions?
2. What are common limitations that can degrade the resolution of cryo-EM reconstructions?
3. At what stages in the cryo-EM structure determination process can improvements be made to improve the resolution and interpretability?



EMDB - 9865

Features of high-quality samples for SPA

- Highest resolution single-particle cryo-EM structure in the EMDB is apoferritin at 1.15 Å
- Apoferritin accounts for 19 of the 20 deposited structures at resolutions greater than 1.6 Å (other AAV-2)
- What makes these samples particularly good for cryo-EM analysis?
- Why don’t other samples achieve resolutions better than 2 Å?

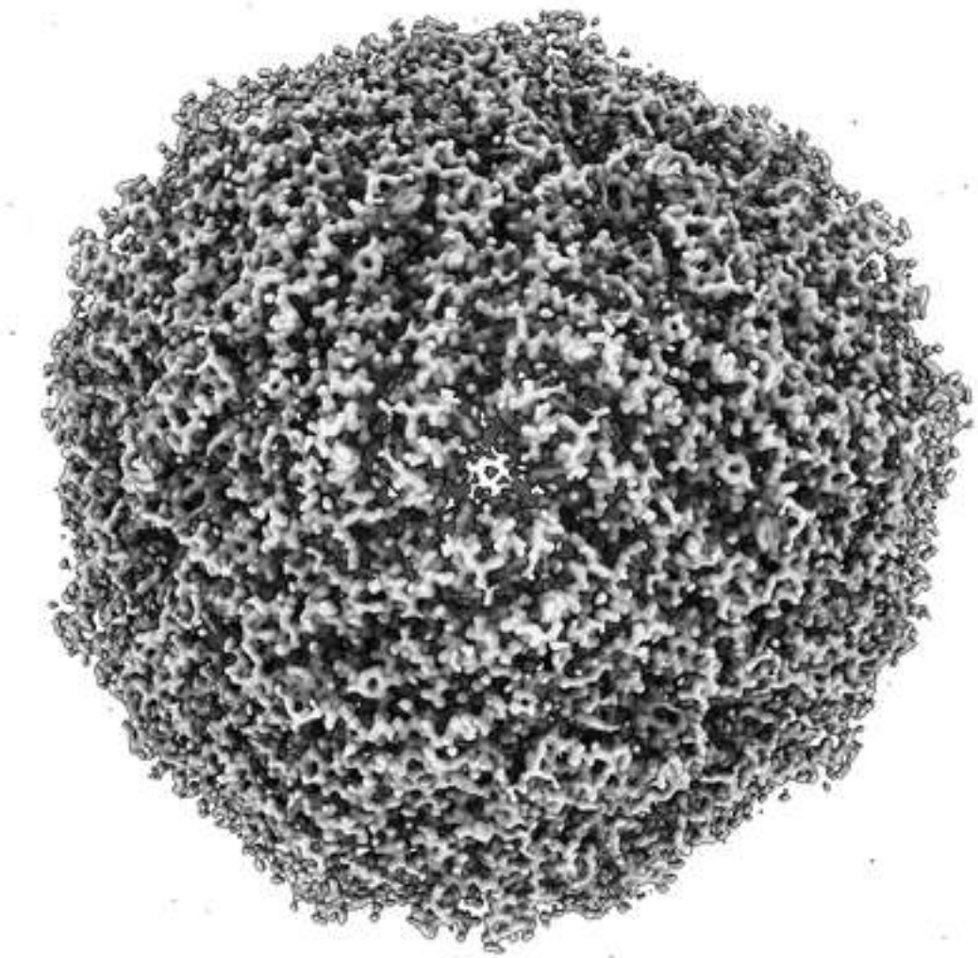


EMDB - 9865

What if you don’t work on apoferritin?

Features that can limit resolution:

- Low or no symmetry
- Difficult to produce
  - Low concentration
  - Sticks to the carbon support / nothing in the holes
- Heterogeneous samples
  - Compositional heterogeneity
  - Conformational heterogeneity
- Particle instability
  - Weak interactions
  - Damage by the air-water interface
- Preferred orientation

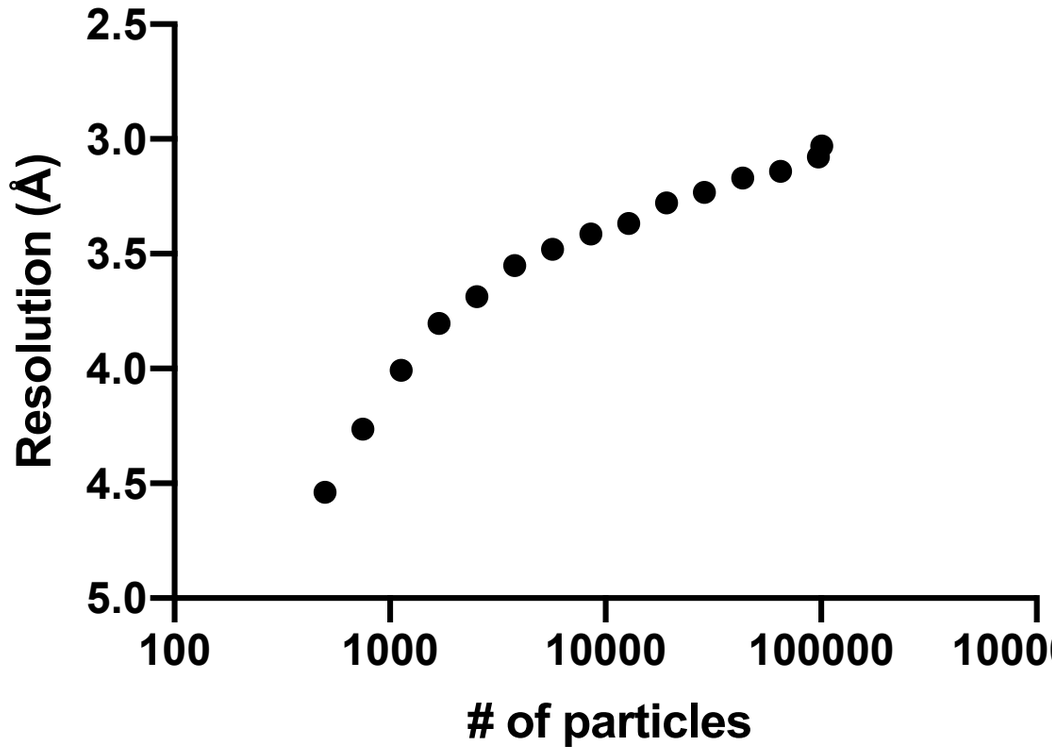


EMDB - 9865

Approaches to improve resolution

Ways to improve resolution:

- Brute-force data collection - millions of particles
- Structure-determination chaperones
  - Antibody fragments
  - Nanobodies
- Identify small-molecule modulators of protein activity
- Identify protein-cofactors that modulate activity
- Molecular biology to truncate away flexible domains

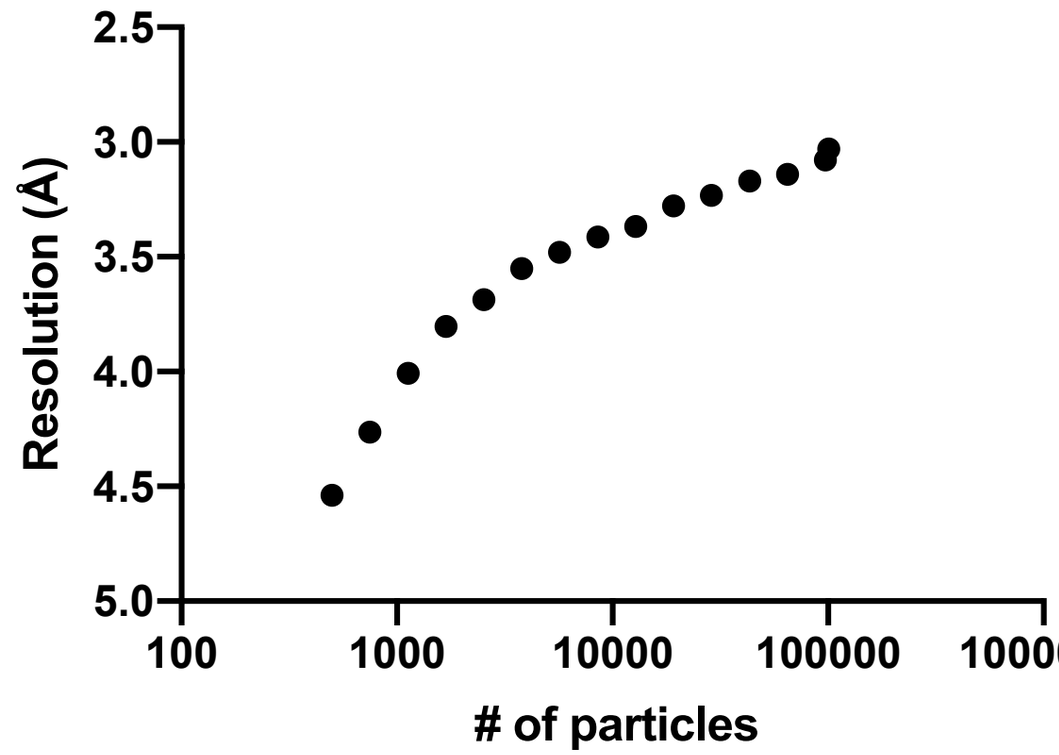


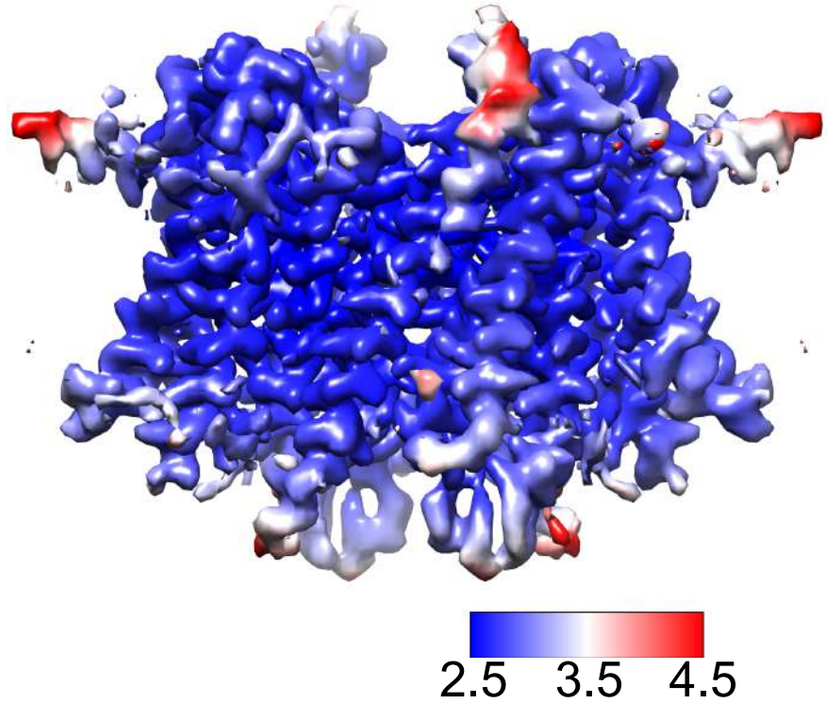
2014 - 400 images a day
2017 - 800 images a day
2018 - 2000 images a day
2019 - 3500 K3 (1.5x area) images a day
2020 - 7000 K3 images per day

Approaches to improve resolution

Ways to improve resolution:

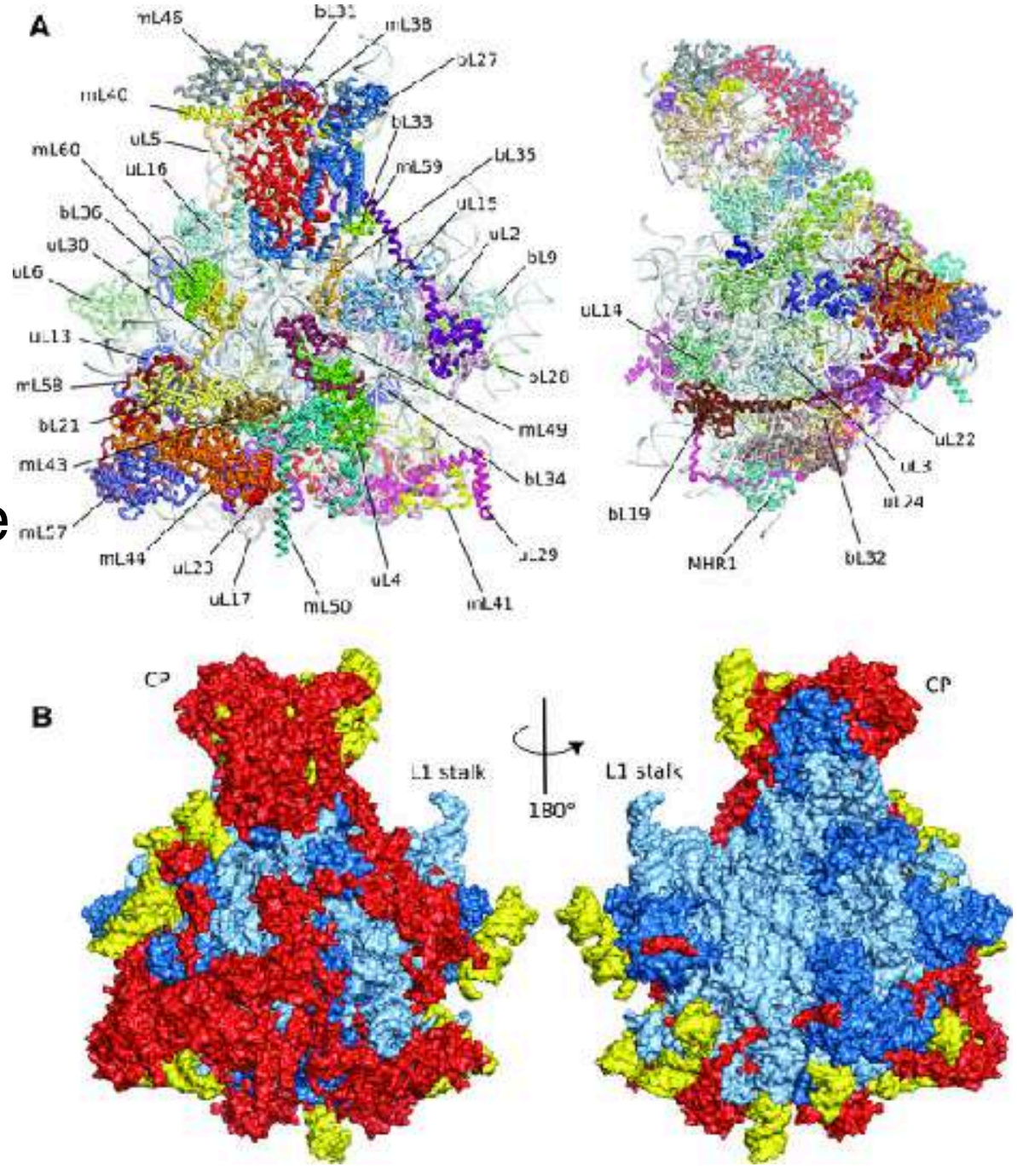
- Brute-force data collection - millions of particles
- Structure-determination chaperones
  - Antibody fragments
  - Nanobodies
- Identify small-molecule modulators of protein activity
- Identify protein-cofactors that modulate activity
- Molecular biology to truncate away flexible domains





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



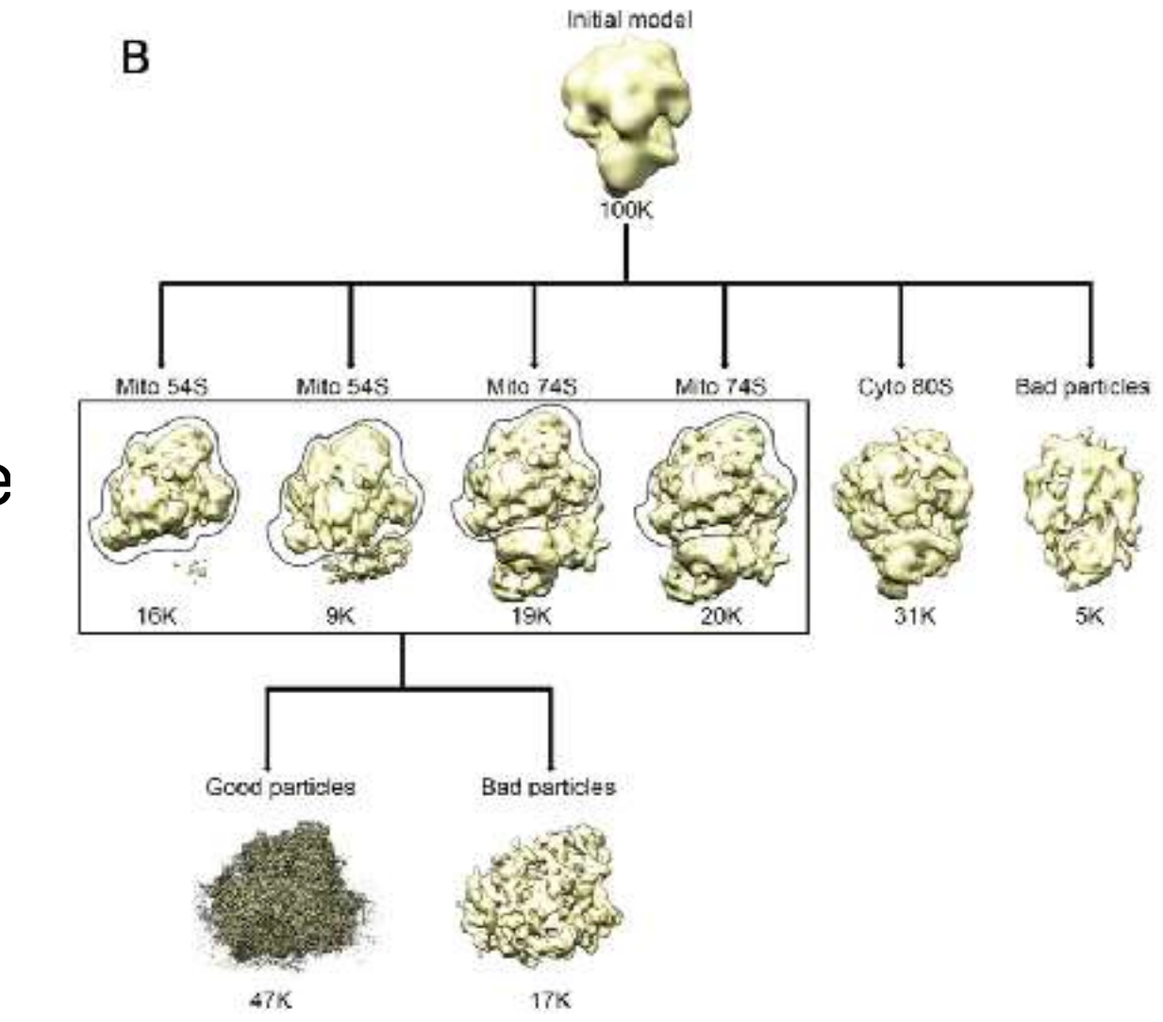
Amunts *et al*, Science. (2014) 343:1485-1489

Heterogeneity - biochemical approaches

- Optimizing biochemistry can often help to minimize sample heterogeneity and is generally the best place start to improve 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)
  - Domain deletions can also reduce sample heterogeneity by removing flexible domains or linkers
- 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

Heterogeneity - computational approaches

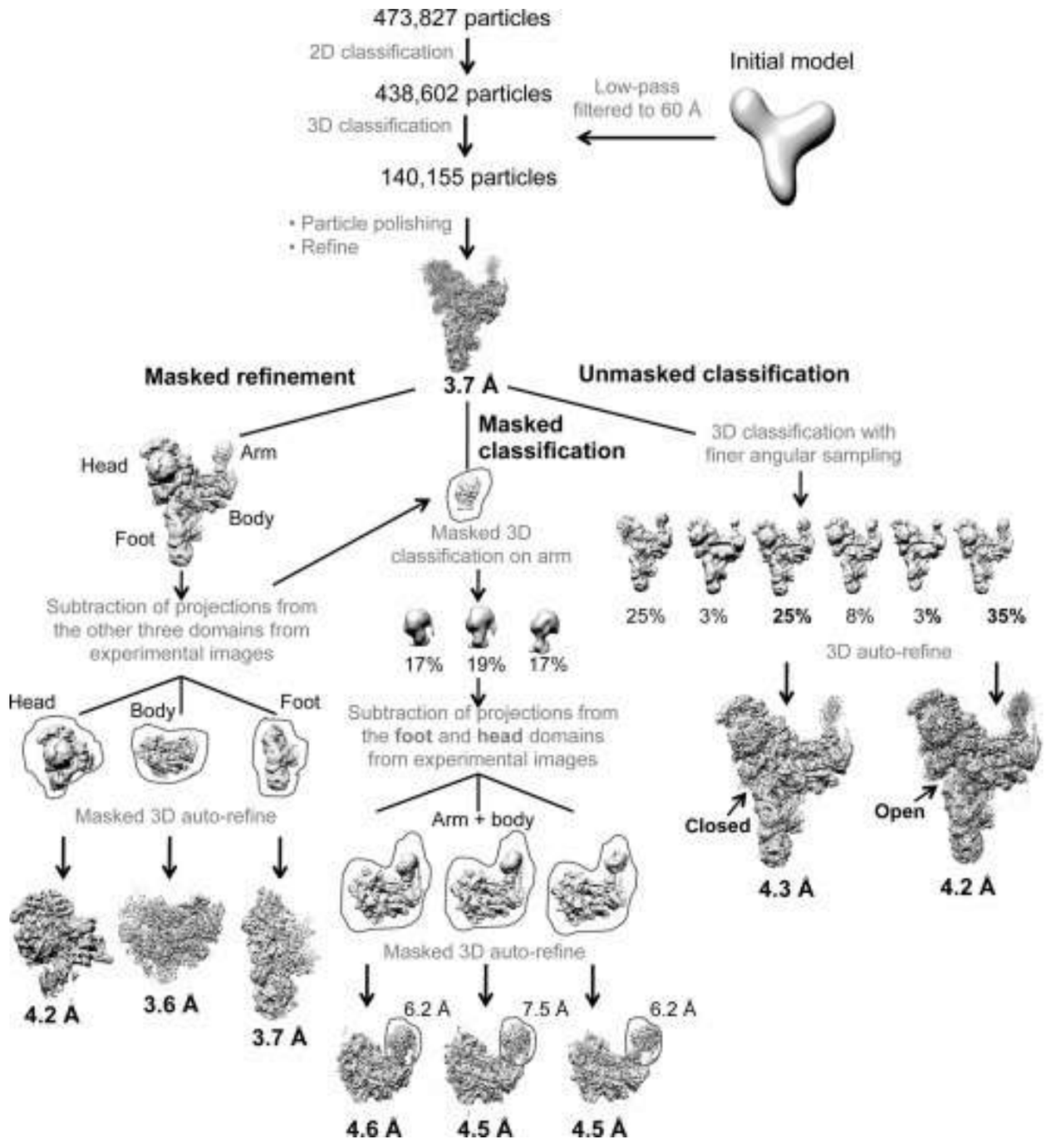
- There are multiple sources of heterogeneity in cryo-EM samples
  - Compositional heterogeneity - mixture of different components or mixtures with varying subunit stoichiometry
  - Structural heterogeneity - domains of the specimen can adopt multiple conformations
  - In some cases, both types of heterogeneity 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

Heterogeneity - computational approaches

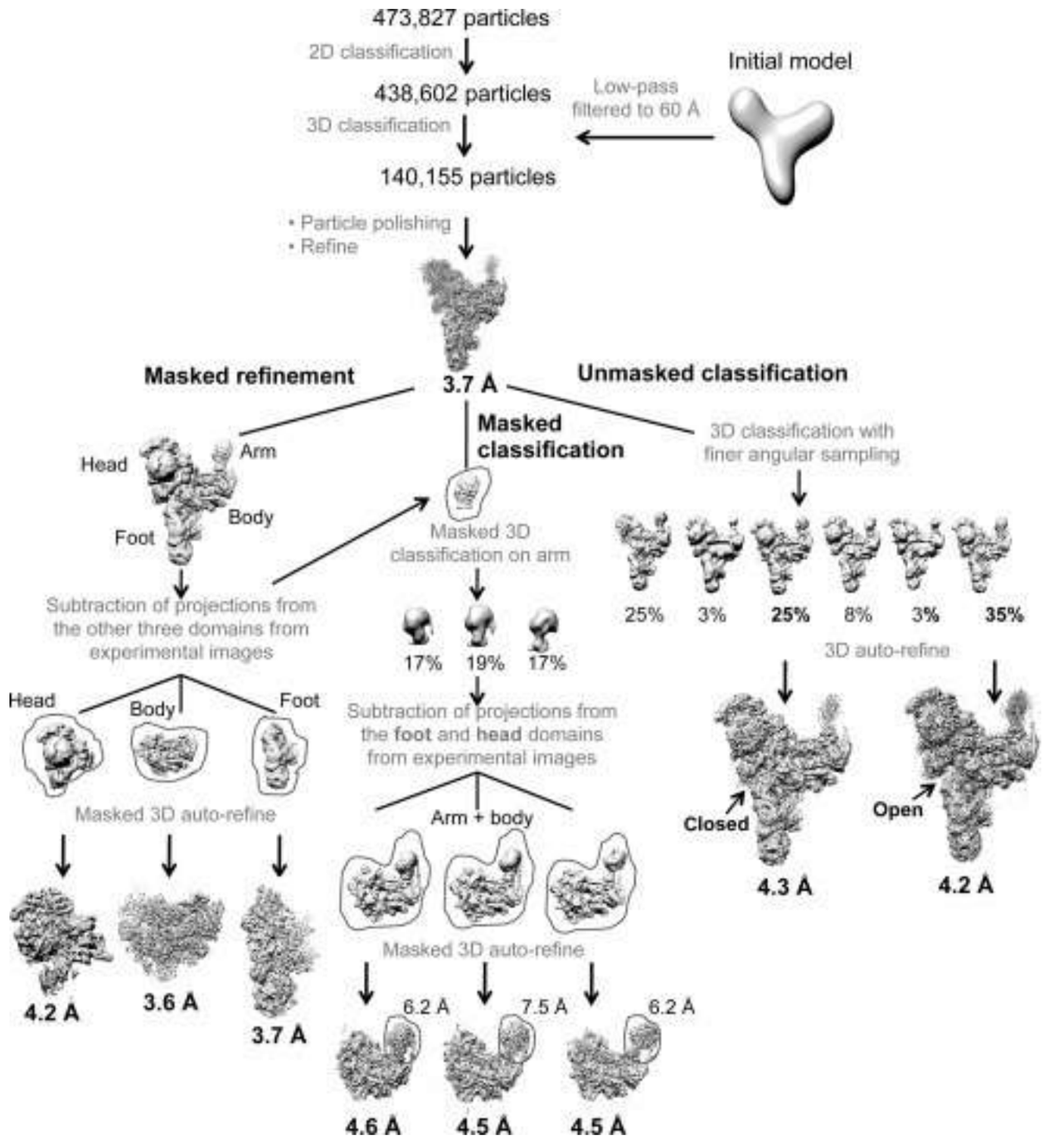
- Heterogeneity may be unavoidable for some samples, particularly large dynamic complexes and must be dealt with computationally after image acquisition
- There are now several different software packages that sort and classify particles, allowing one to create “pure” subsets of the particles images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



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

Heterogeneity - computational approaches

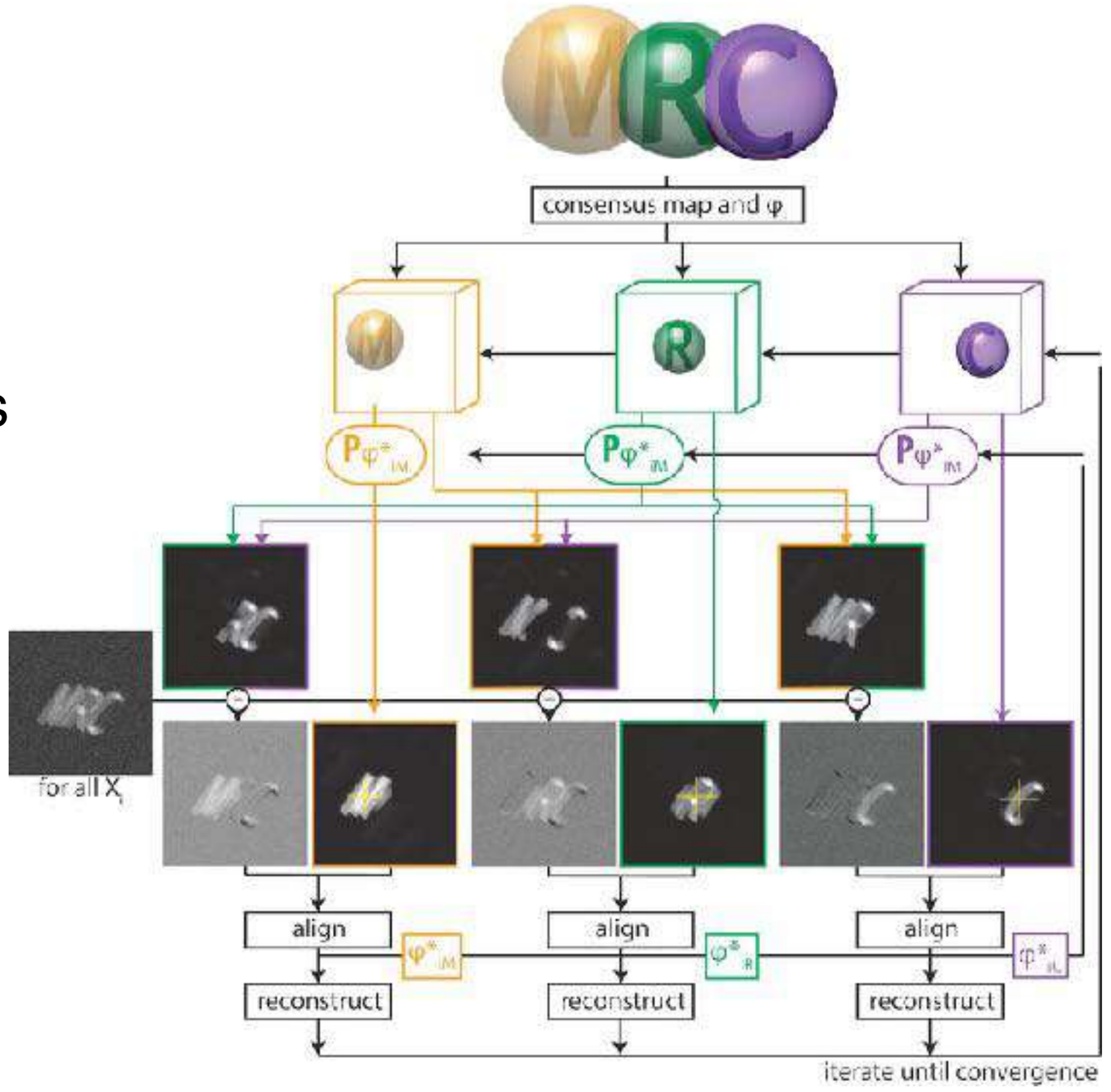
- Classification can be enhanced through the use of masks
- A mask can be placed around the region of interest - allowing independent sorting of different domains
- This multi-classification approach is particularly powerful for samples that have multiple different types of movements
- Another modification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

Heterogeneity - computational approaches

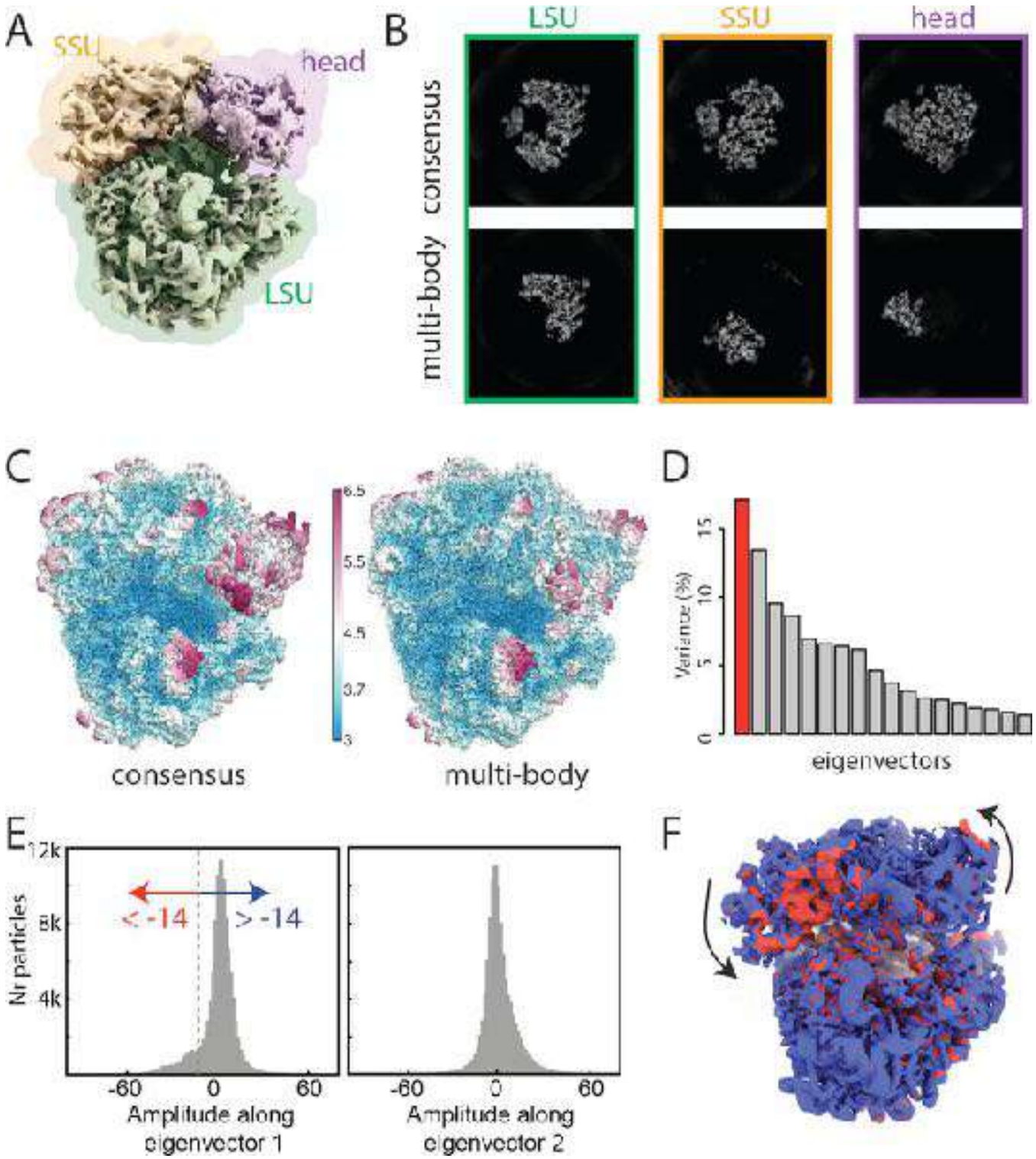
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Heterogeneity - computational approaches

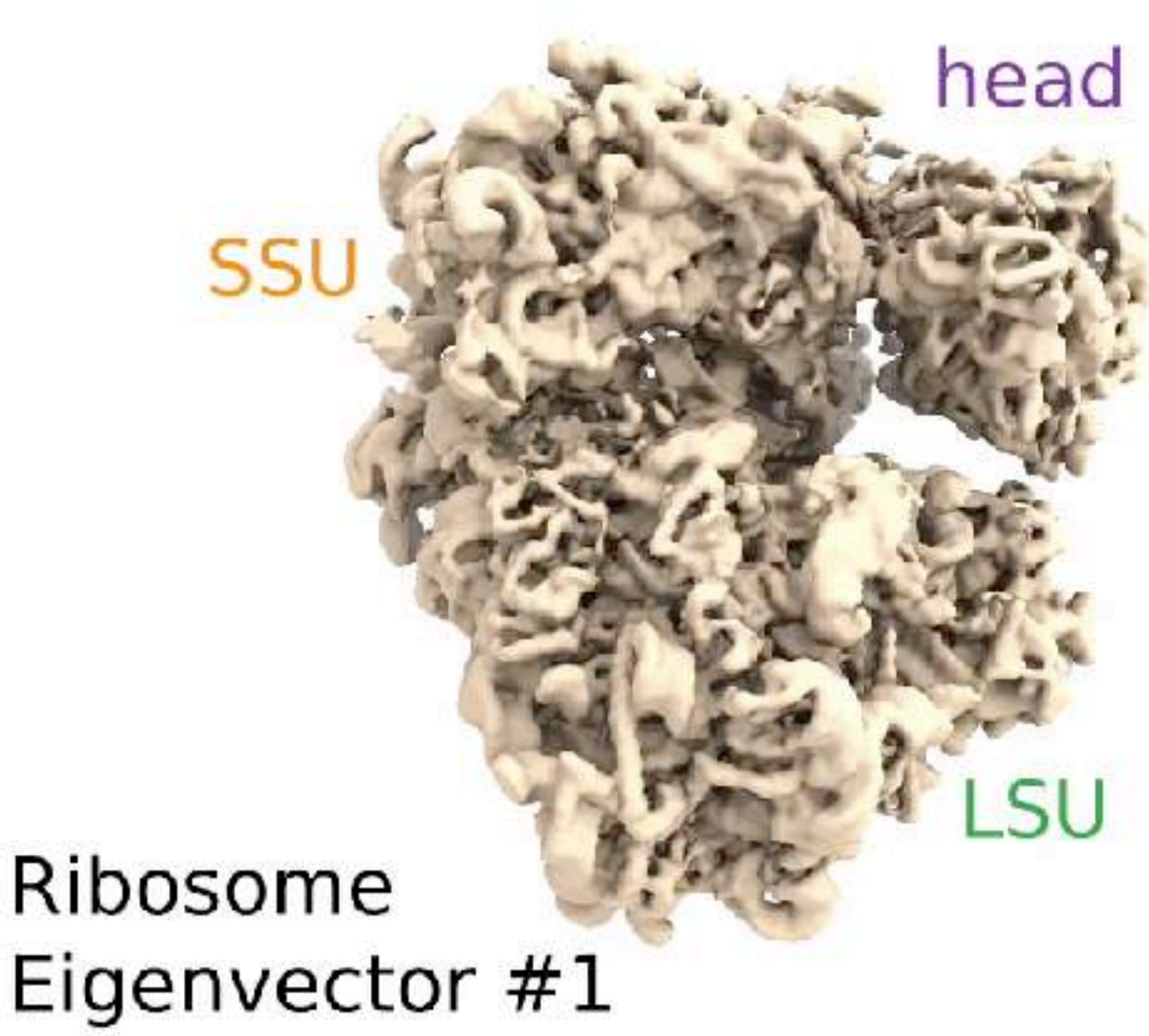
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Heterogeneity - computational approaches

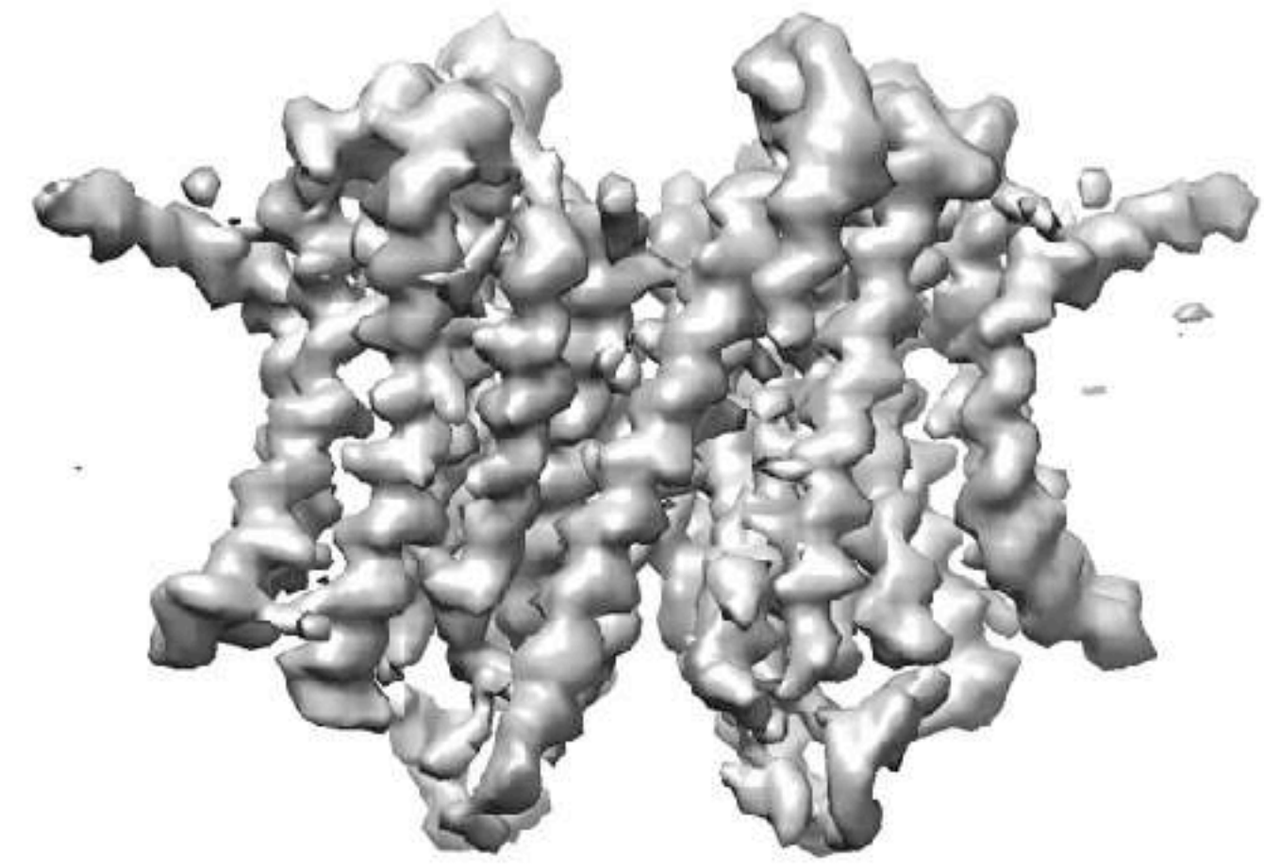
- Relion has an automated procedure to apply masks based upon distinct flexible domains
- This approach is known as multi-body refinement
- It also determines the vectors of movement allowing an understanding of conformational dynamics across protein complexes



Nakane et al, eLife. (2018) e36861

Heterogeneity - computational approaches

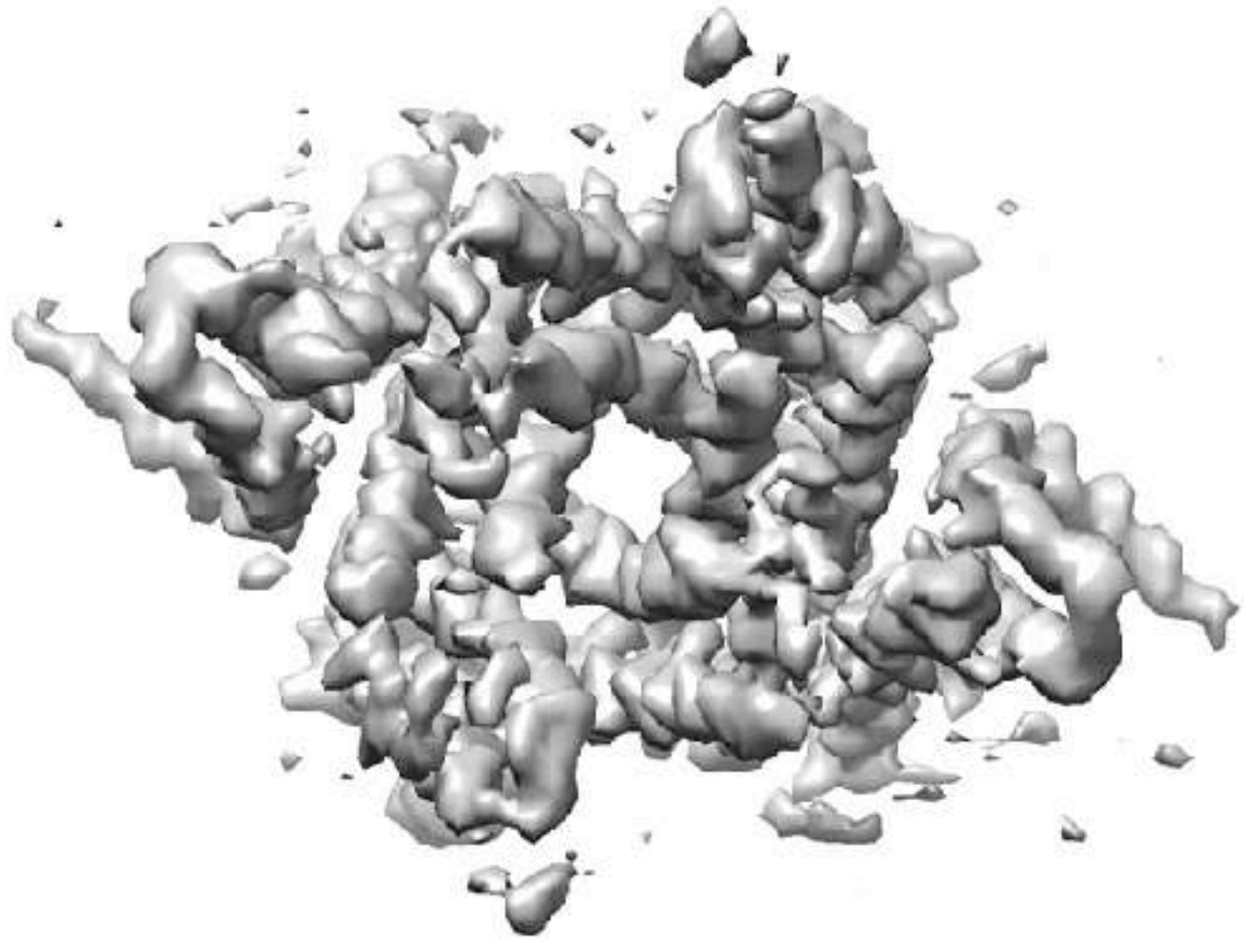
- CryoSPARC has an alternative approach that does not require *a priori* information about flexible interfaces known as 3D variability analysis
- It employs eigenvector analysis to identify the principal components present between particles in the data set
- These analyses can identify compositional and conformational heterogeneity within cryo-EM data



https://cryosparc.com/docs/tutorials/3d-variability-analysis/

Heterogeneity - computational approaches

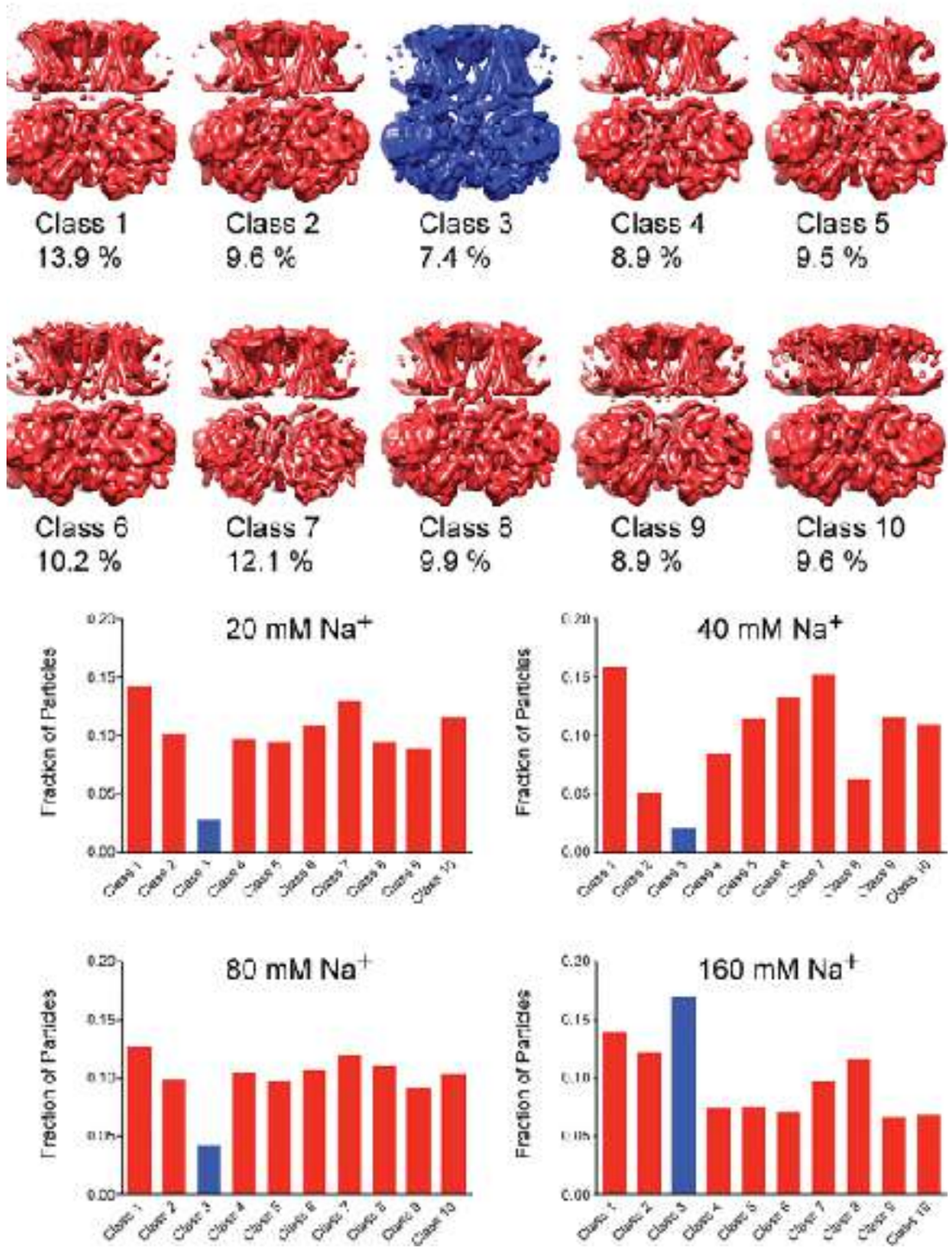
- CryoSPARC has an alternative approach that does not require *a priori* information about flexible interfaces known as 3D variability analysis
- It employs eigenvector analysis to identify the principal components present between particles in the data set
- These analyses can identify compositional and conformational heterogeneity within cryo-EM data



https://cryosparc.com/docs/tutorials/3d-variability-analysis/

Benefits of heterogeneity?

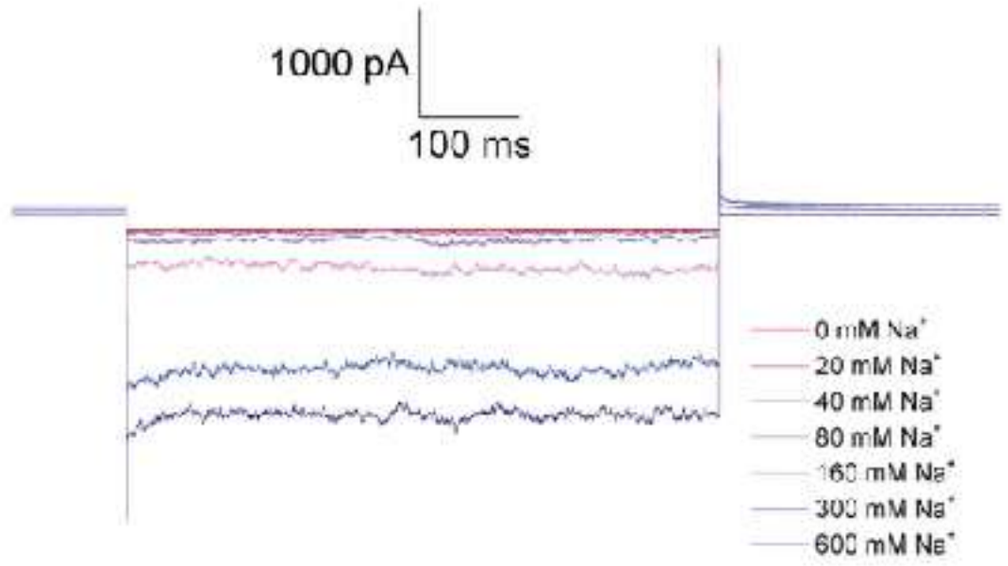
- How can you use heterogeneity to better understand the biology of your samples?
- Does your heterogeneity correlate with functional changes?
- Always test to ensure that your representative density map is actually representative of your sample, and not merely some small portion of the particles that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

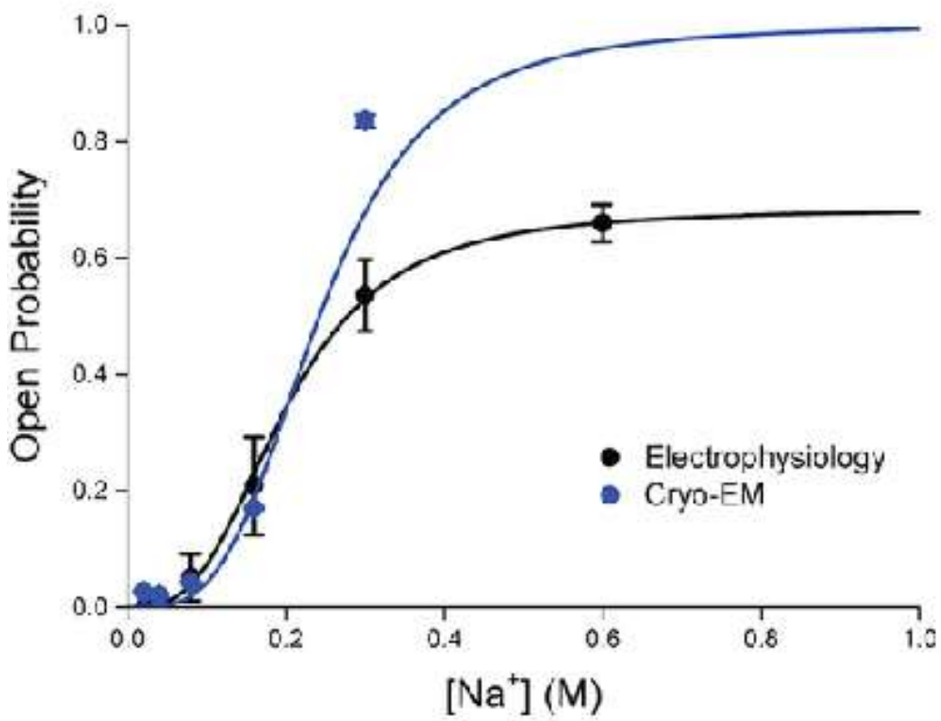


Hite and MacKinnon, Cell. (2017) 168:390-399

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?





Hite and MacKinnon, Cell. (2017) 168:390-399