Heterogeneity in Single Particles

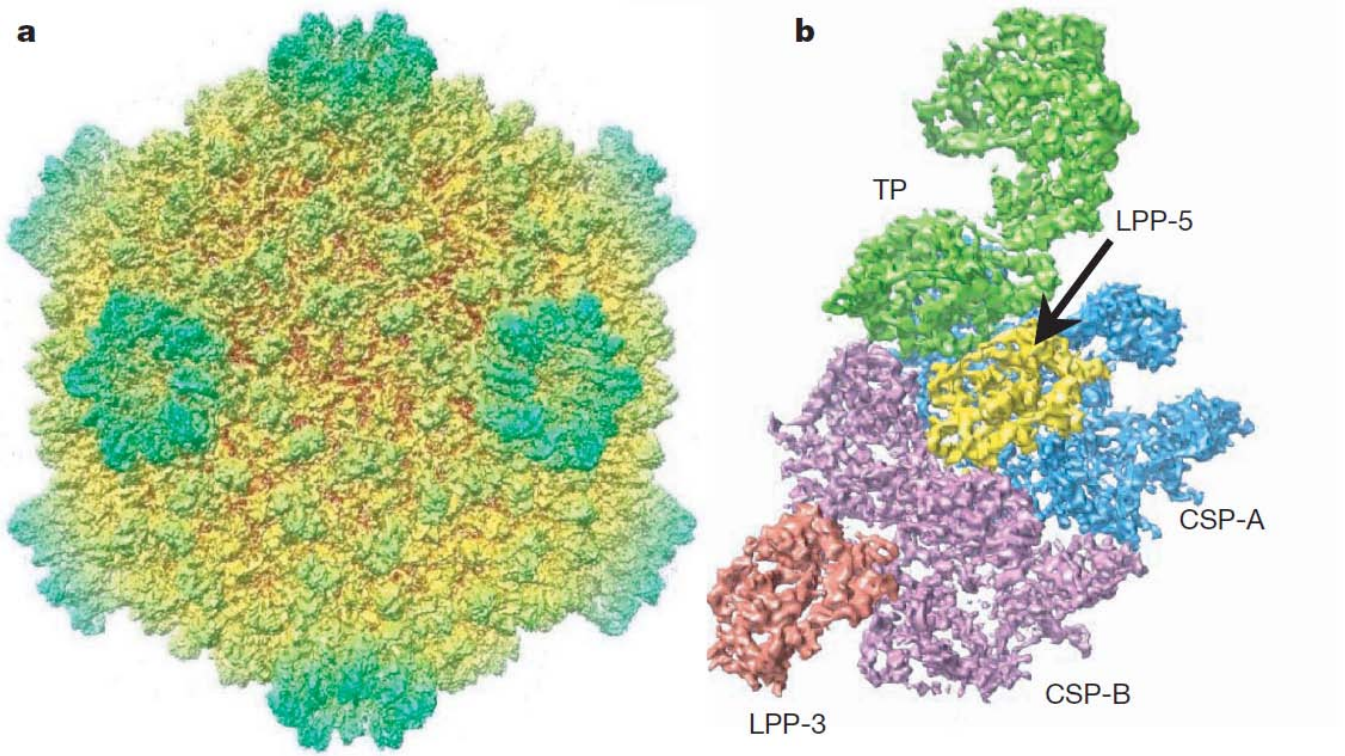
- Degrees of right and wrong
- Ways to increase reliability
- Detecting problems
- Different types of heterogeneity
- Overview of classification methods (Sigworth)
  - Classification as a problem of clustering in factor space
  - Brief intro to supervised classification
  - ML and the EM algorithm
  - ML with a prior probability (MAP estimation)

Heterogeneity in Single Particles…

- ML classification (Sigworth)
- ML-like restraints & classification
- Continuous deformation models (Sigworth)
  - Continuous vs. discrete models
  - Reconstructing continuous models using morphings--2D results.

Degrees of Right and Wrong

Cytoplasmic Polyhedrosis Virus

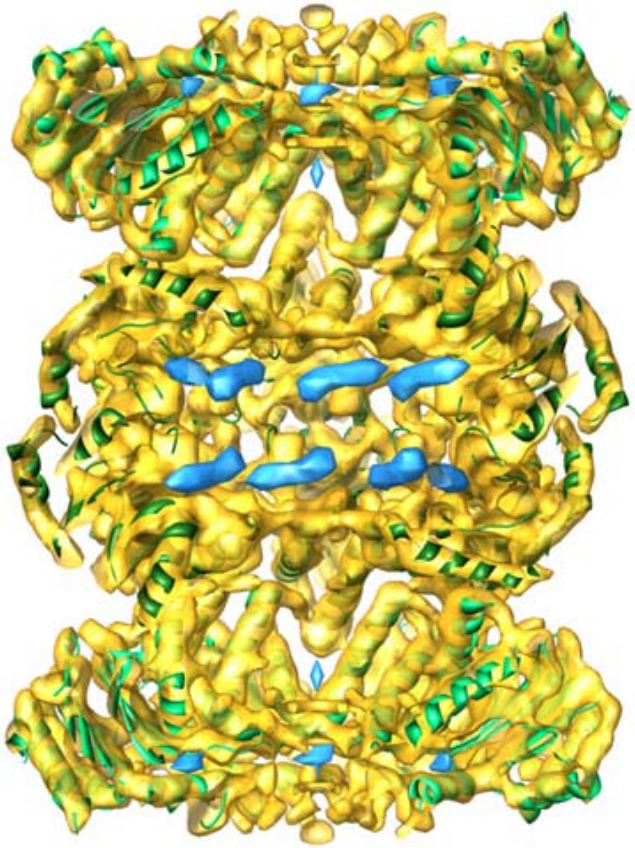


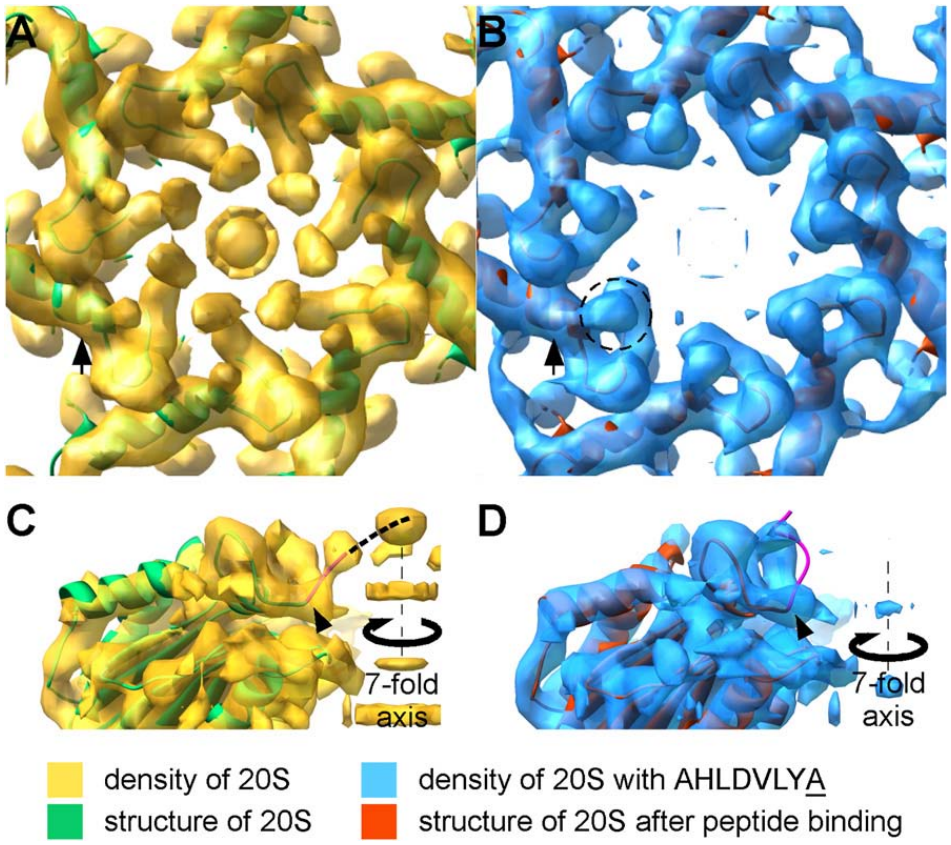
3.88 Å resolution

Atomic structure visible

Yu et al. 2008

20S Proteasome



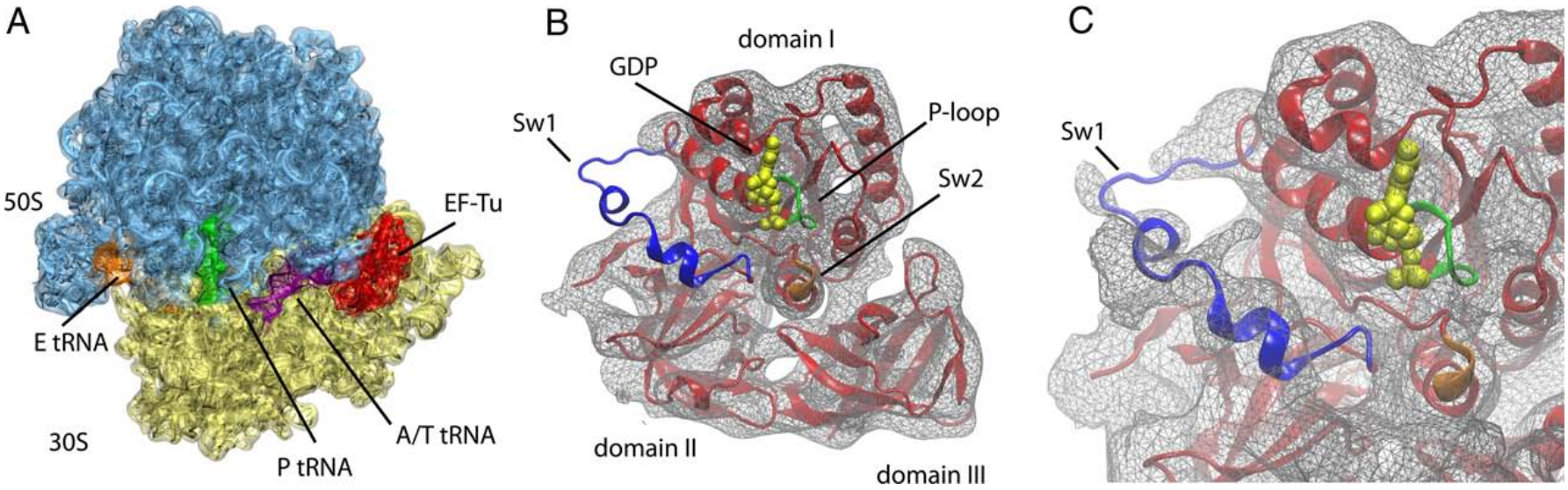


Resolution between 6 and 8 Å

Secondary structure visible
Correlation with existing atomic models

Rabl et al. 2008

80S Ribosome



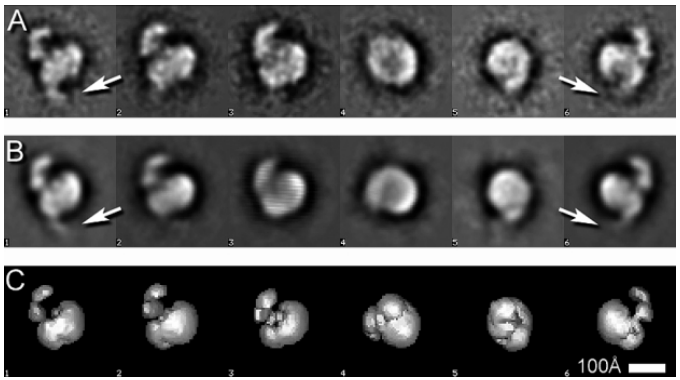
6.7 Å Resolution

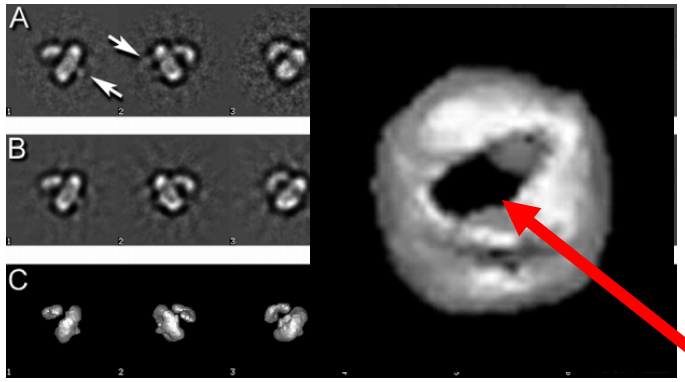
Secondary structure visible
Correlation with existing atomic models

Villa et al. 2009

L-Type Ca^{2+} Channel

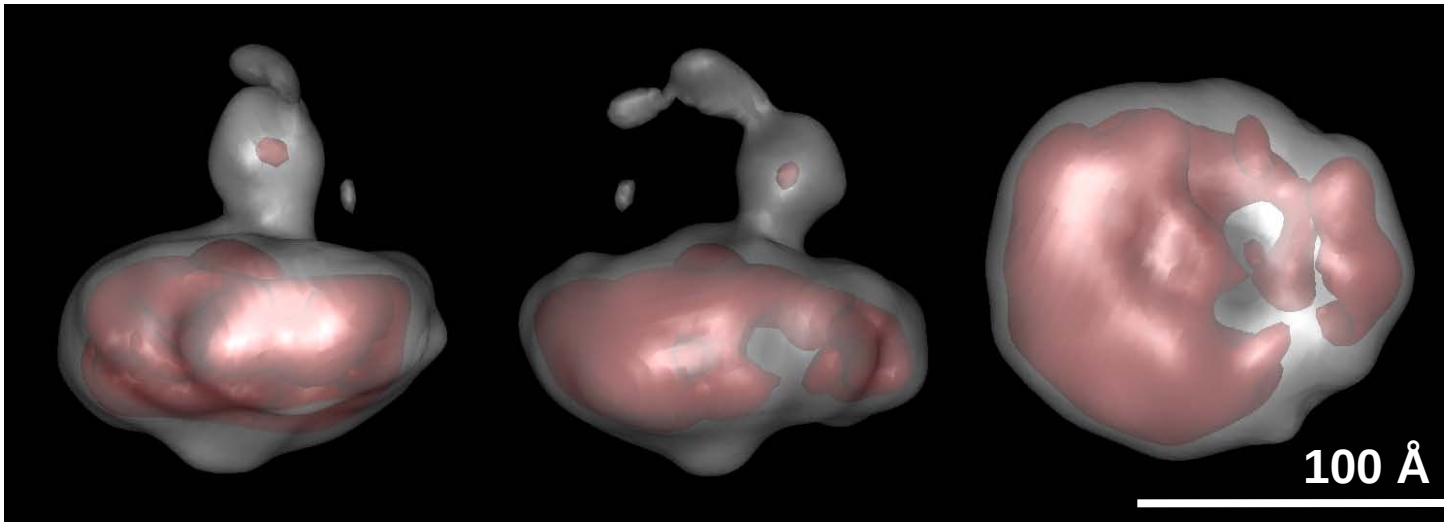
Neg. stain





Cryo

hollow

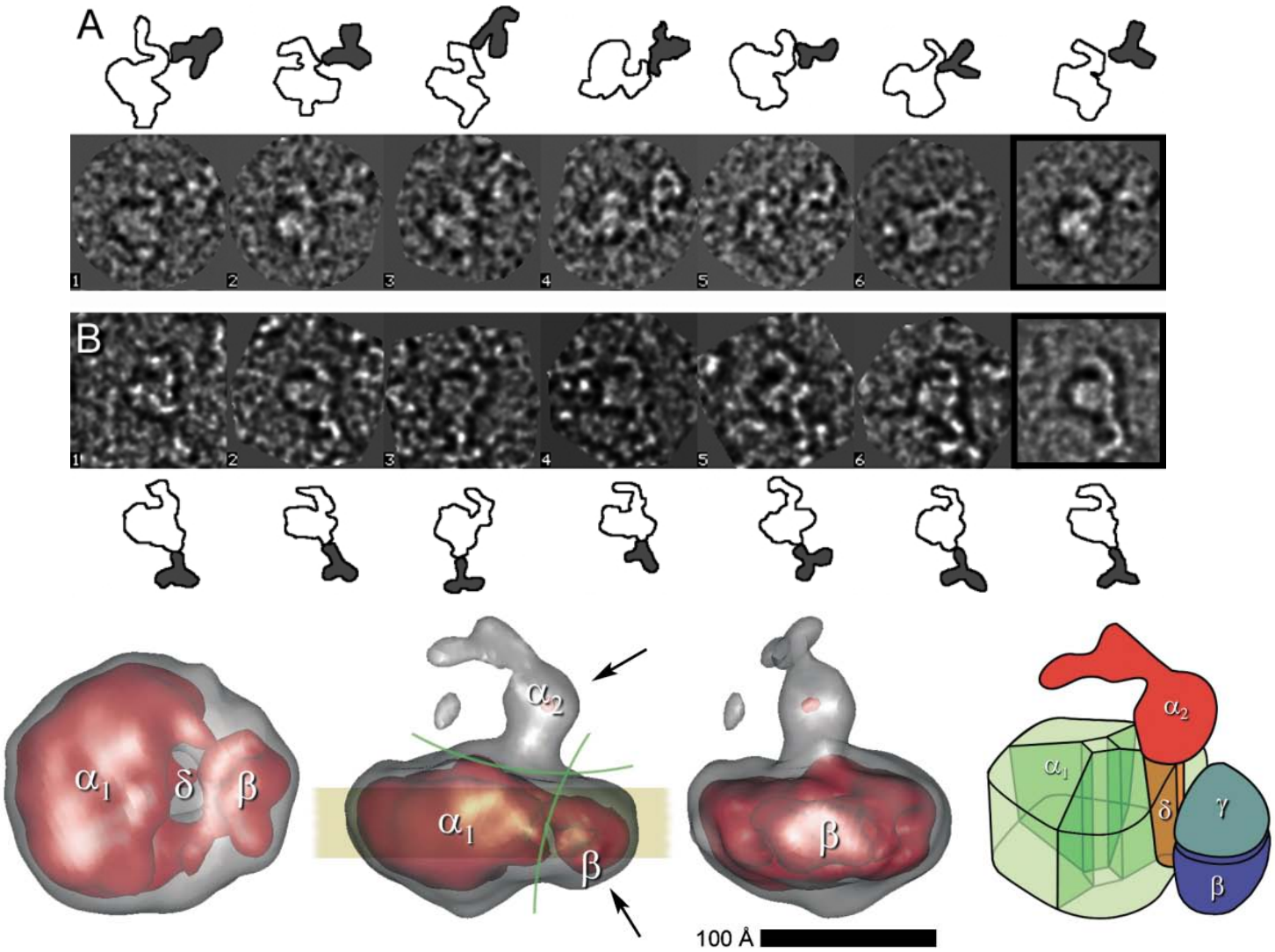


23 Å Resolution

Secondary structure NOT visible
No existing atomic models available

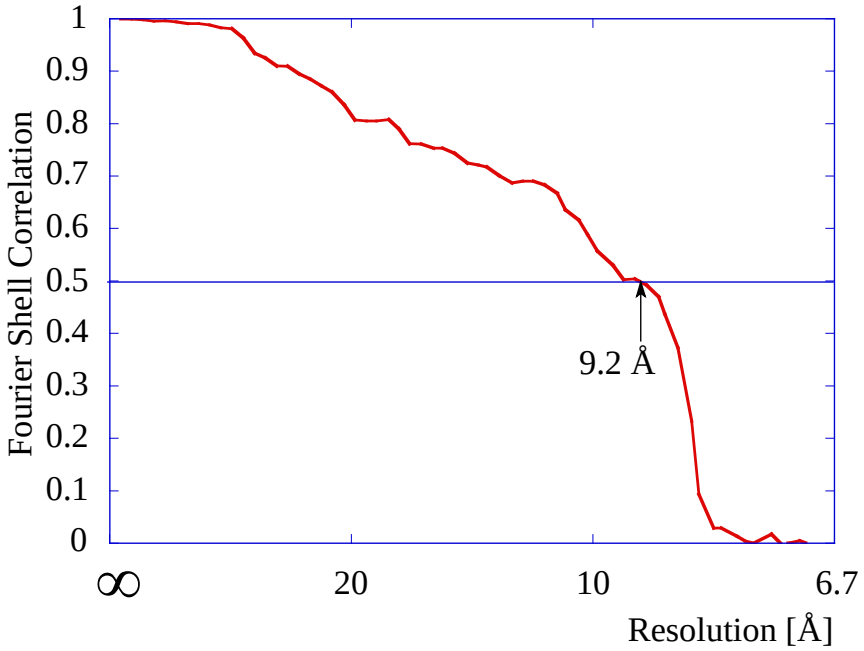
Wolf et al. 2003

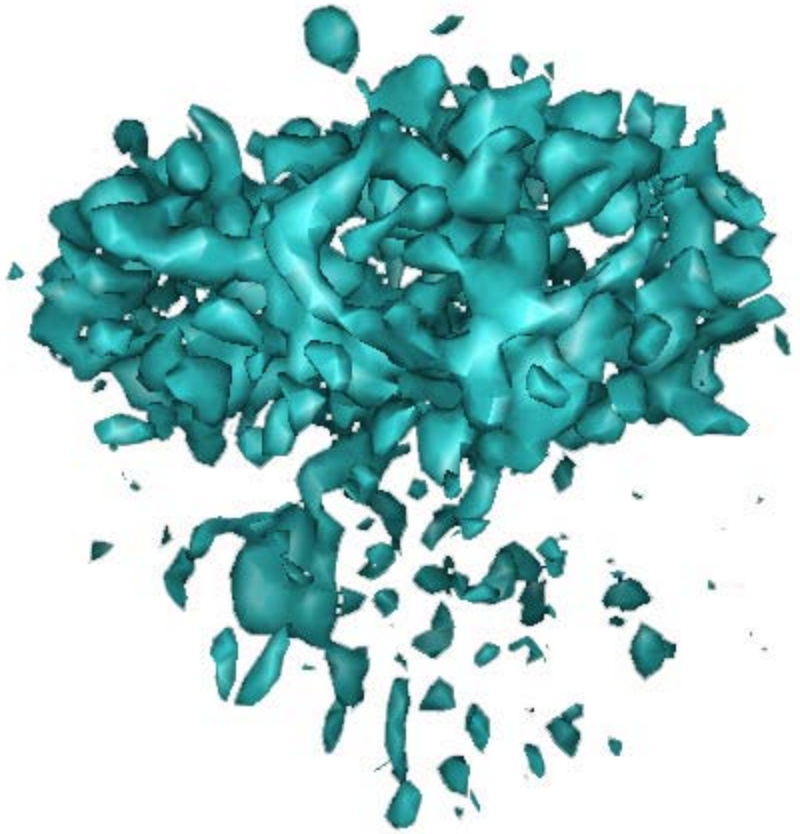
Interpretation



Wolf et al. 2003

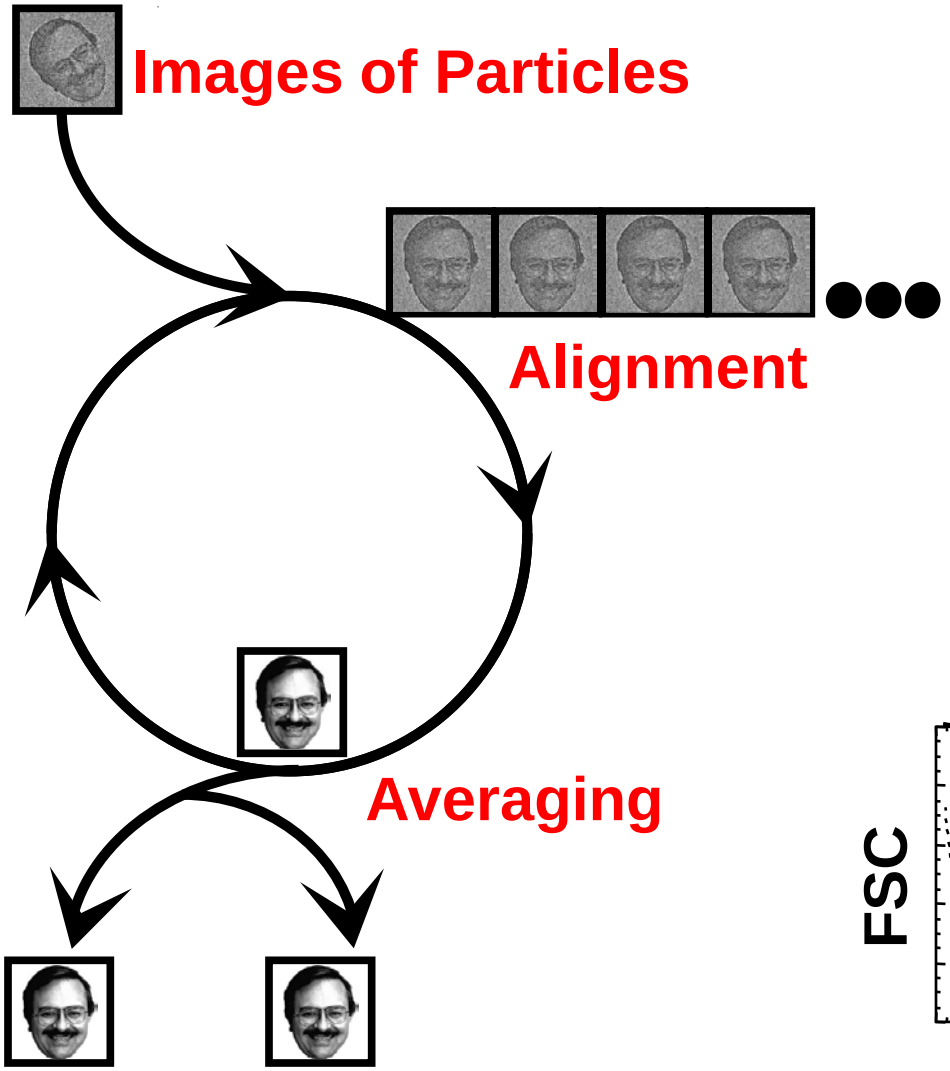
Over-Refinement

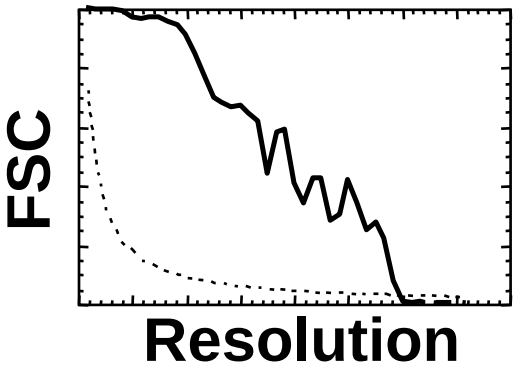




Wolf et al. 2002, unpublished

Resolution Measurement



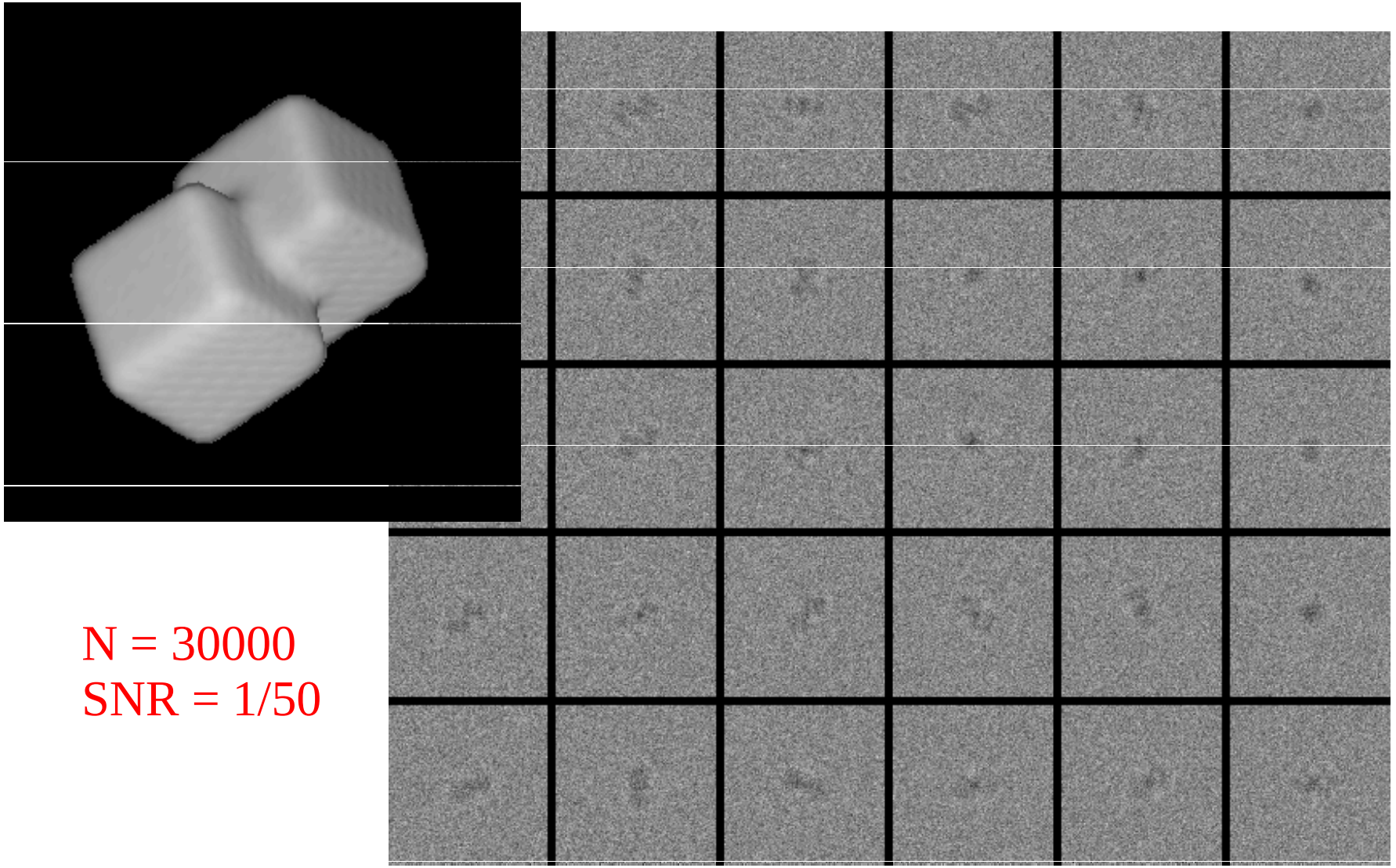


Seeing is NOT Always Believing

100 Images 1000 Images Reference

Ways to Increase Reliability

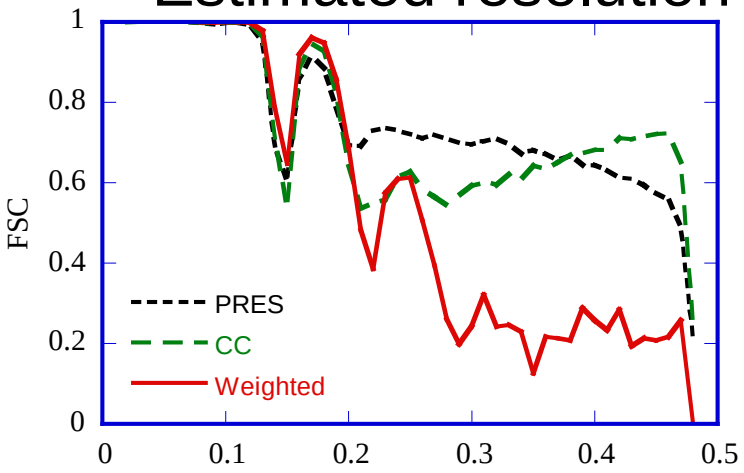
Computer Simulation



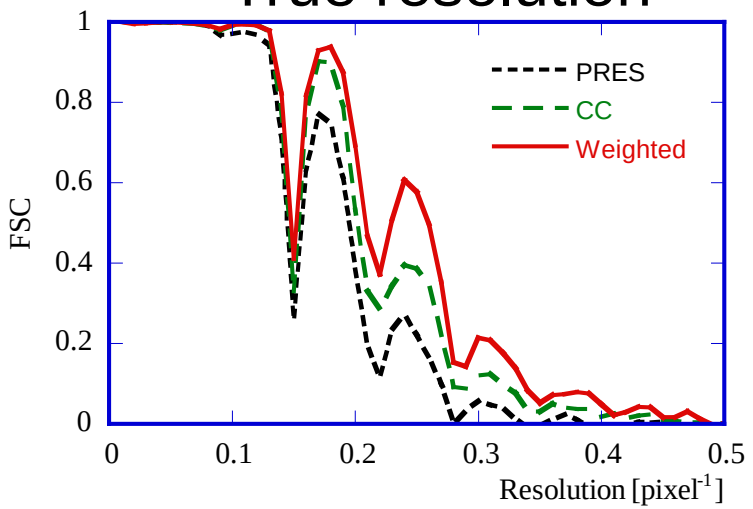
N = 30000
SNR = 1/50

Different Refinement Targets

Estimated resolution



True resolution

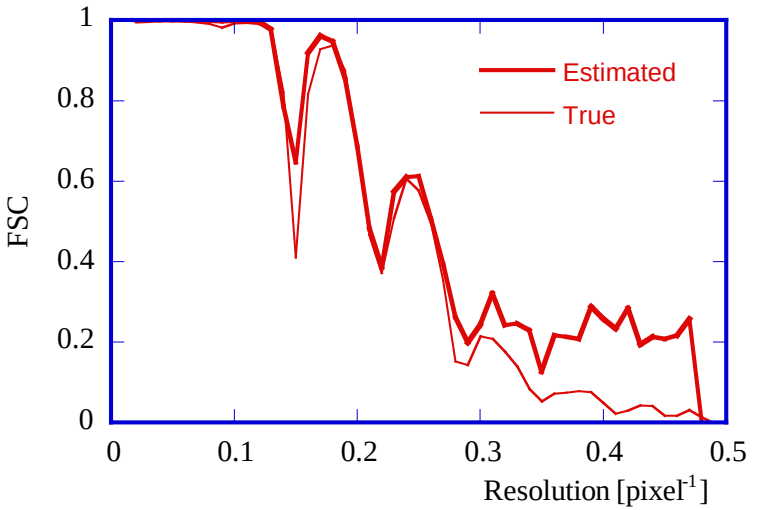


Target functions:

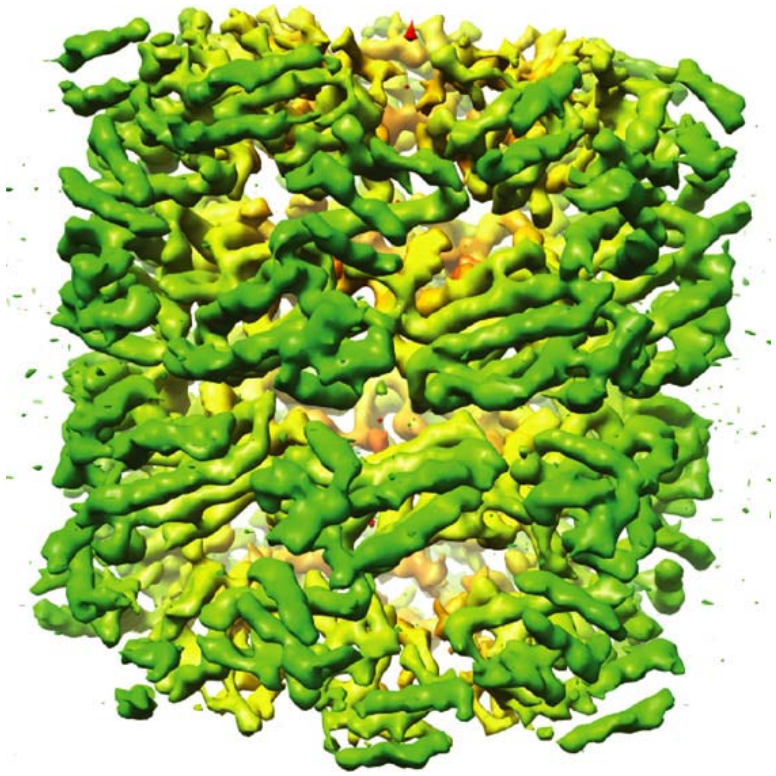
Phase residual

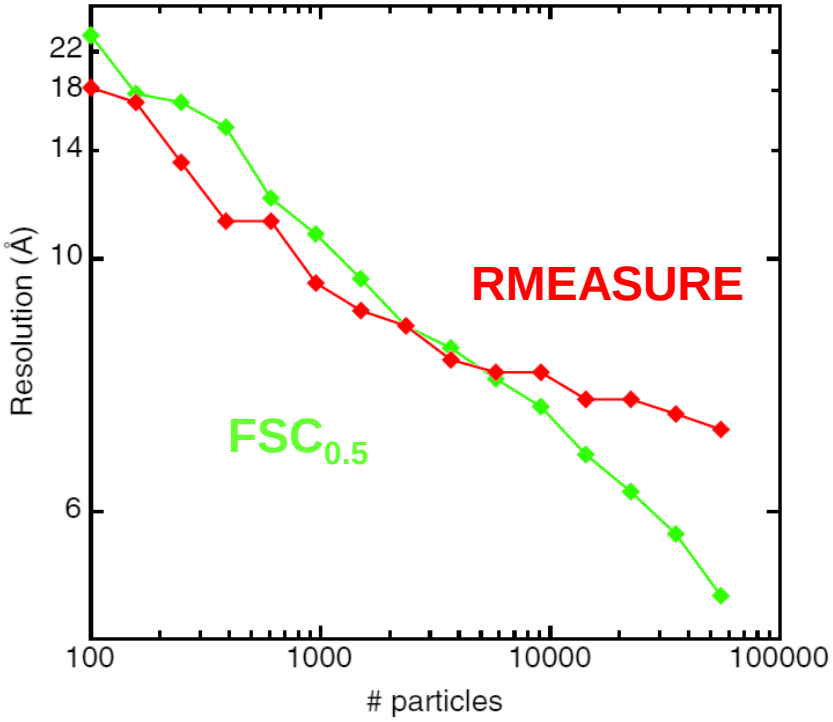
Linear correlation coefficient

Weighted correlation coefficient (signal-to-noise weighting)



Resolution Measurement

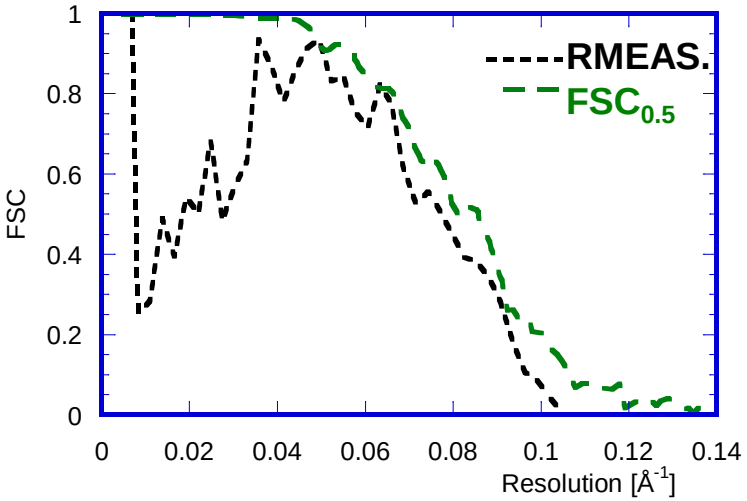


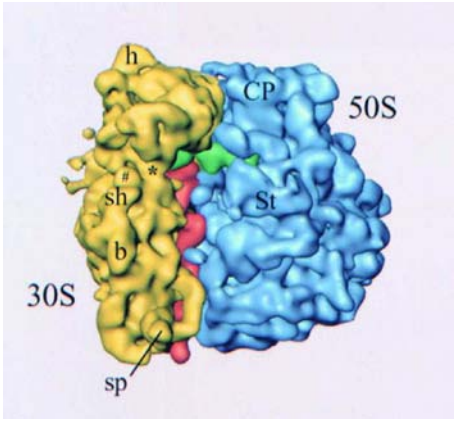


"The resolution reported by RMEASURE […] was more consistent with the details observed in the reconstructions."

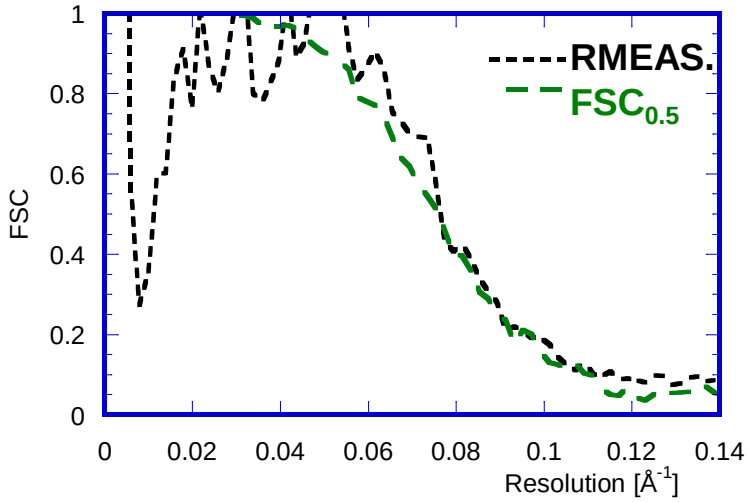
Stagg et al. 2008

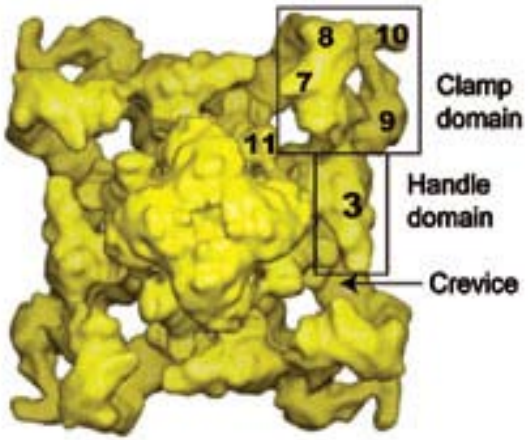
More RMEASURE Tests



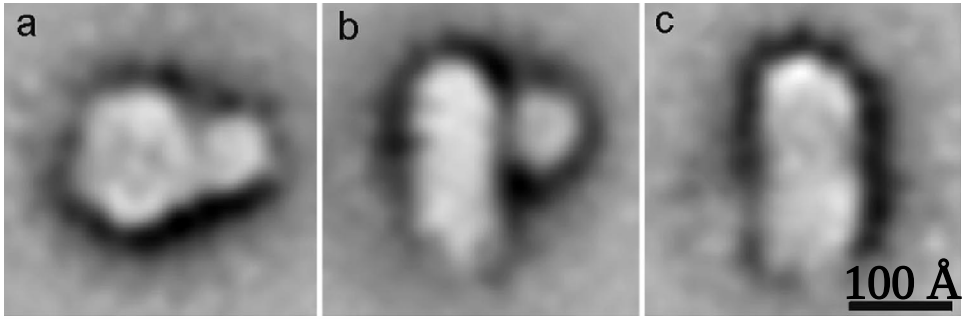


Gabashvili et al. 2000

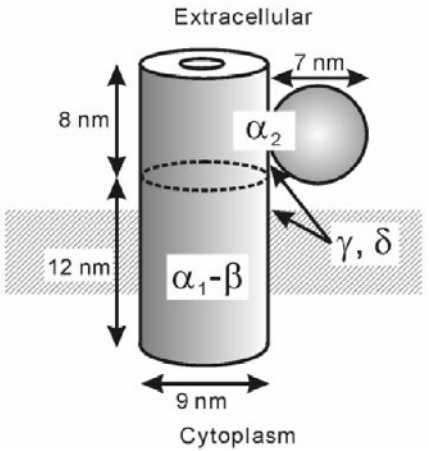




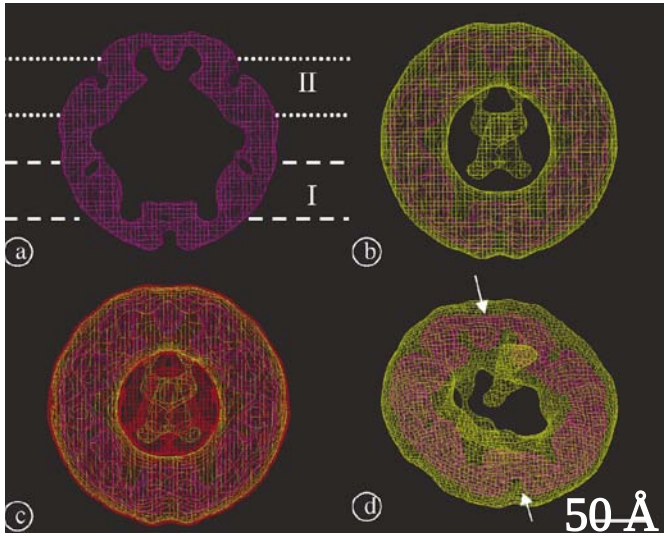
Samso et al. 2005



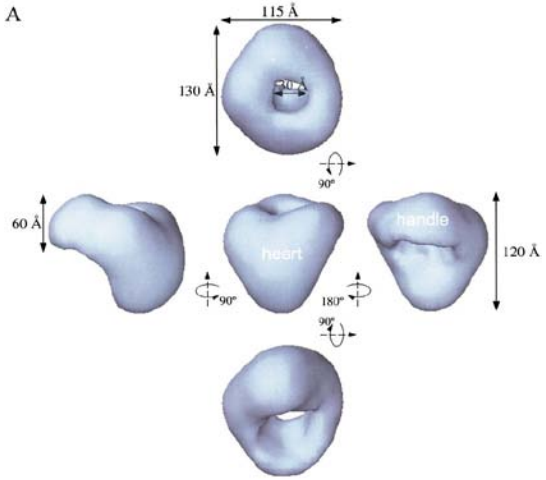
Murata et al. 2001

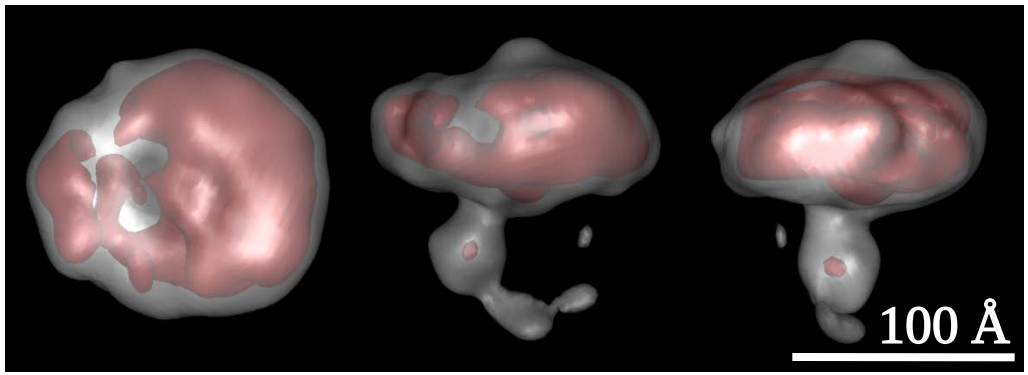


Wang et al. 2002



Serysheva et al. 2002

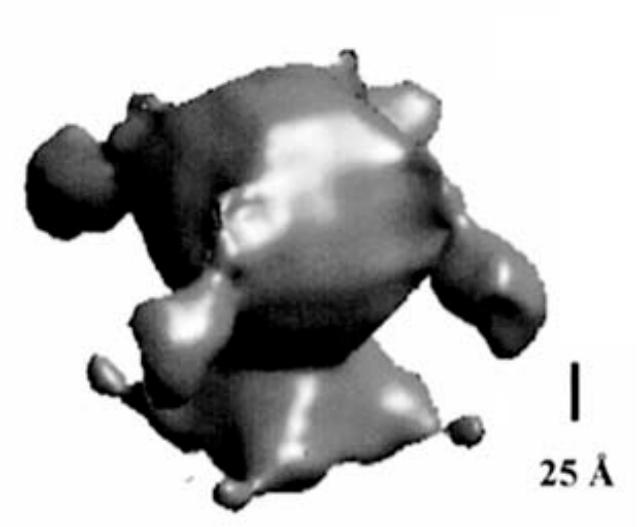




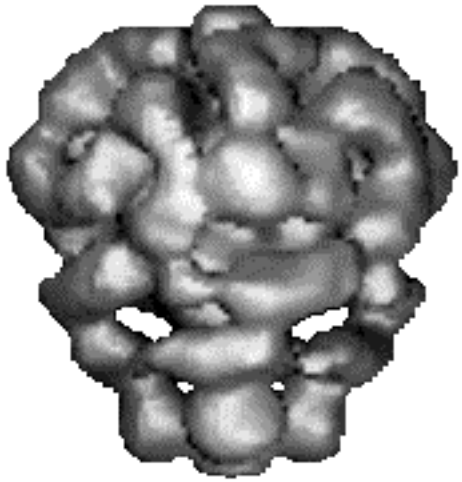
Wolf et al. 2003

The Many Faces of a Channel

IP_3 Receptor

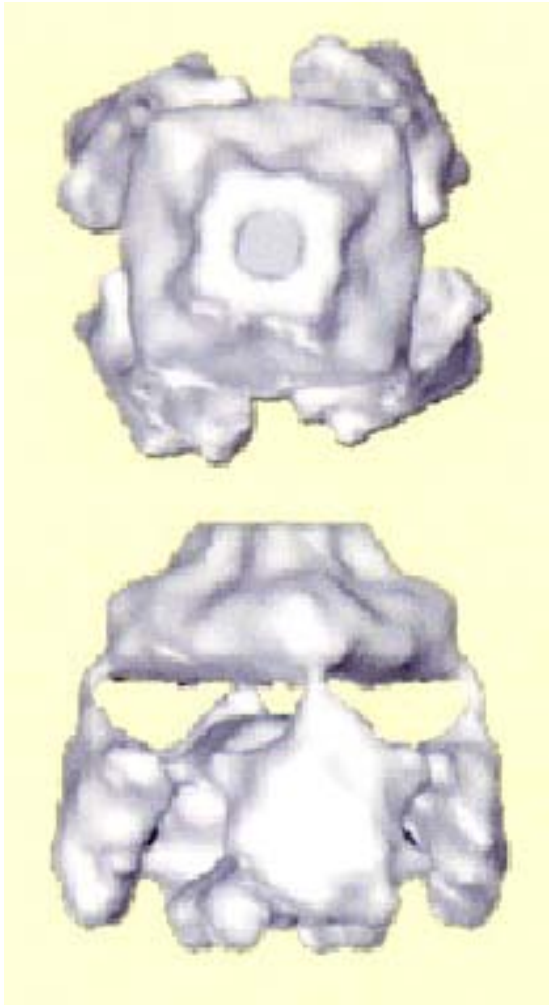


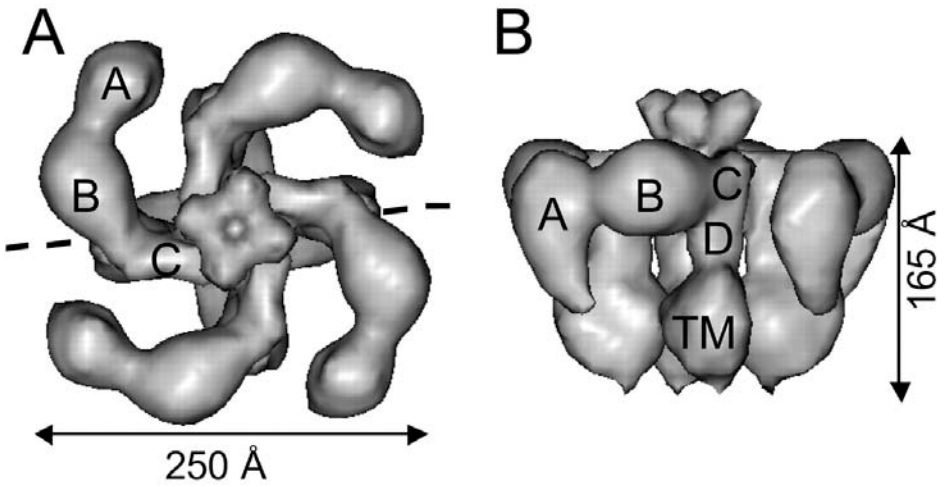
Jiang et al. 2002



Sato et al. 2004

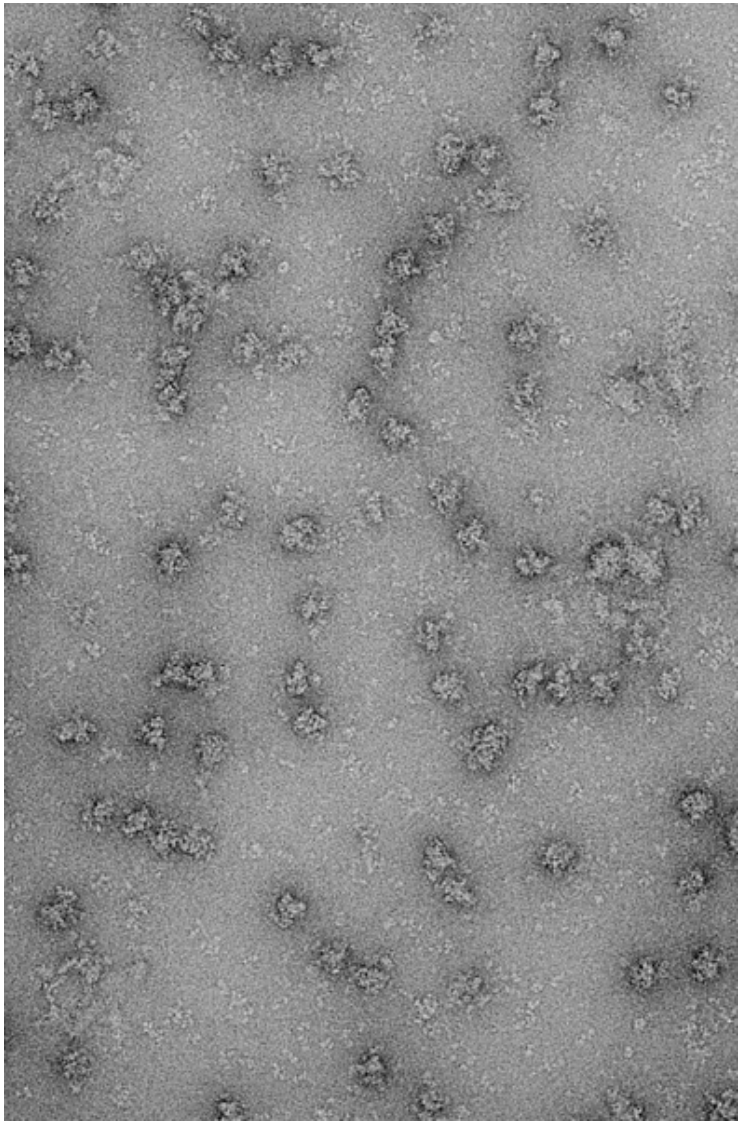
da Fonseca et al. 2003

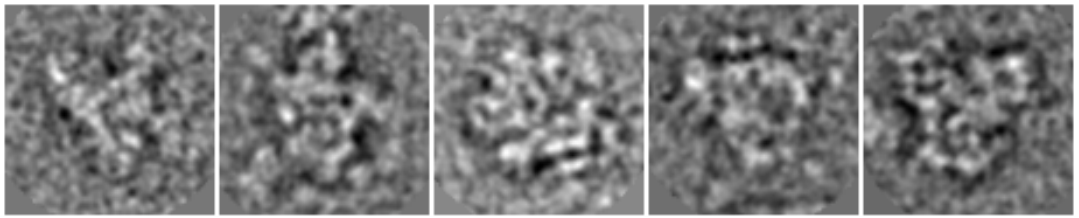


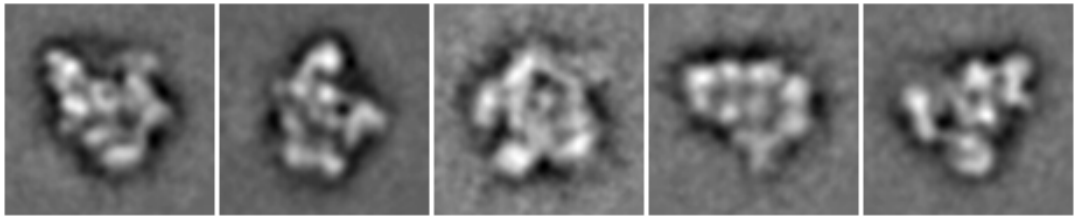


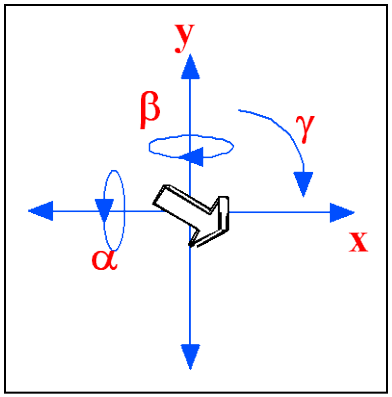
Serysheva et al. 2003

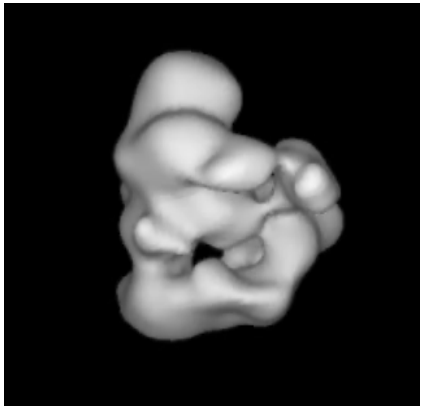
Spliceosome

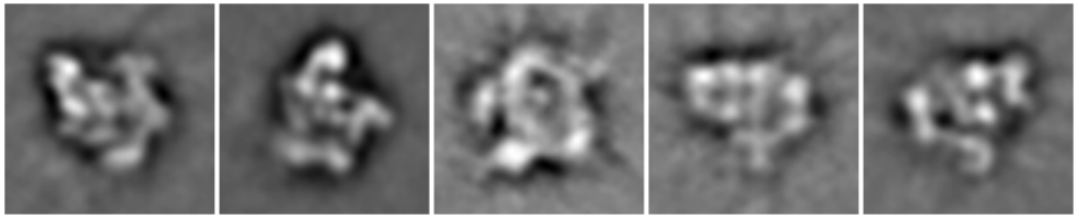






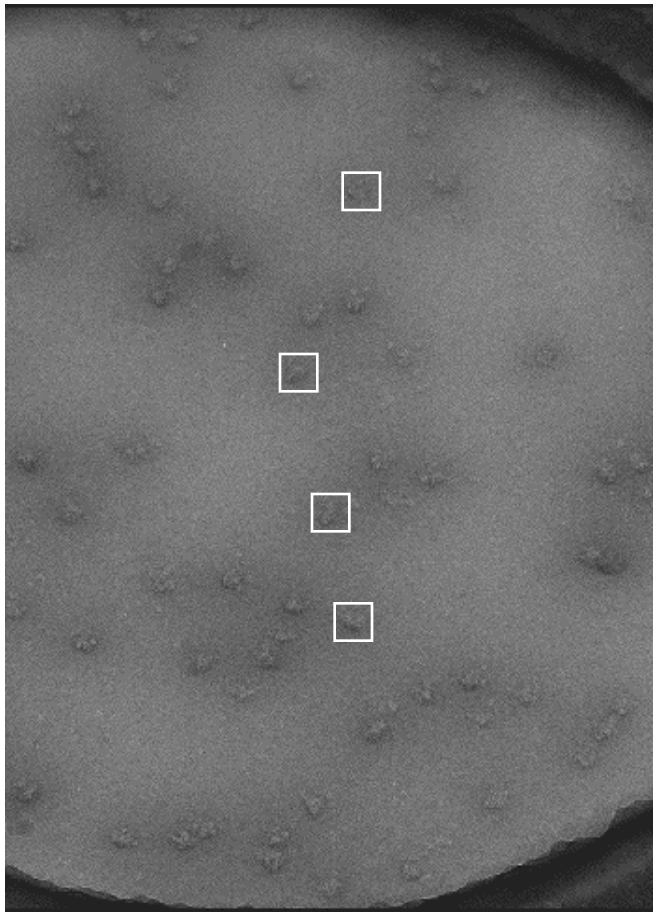




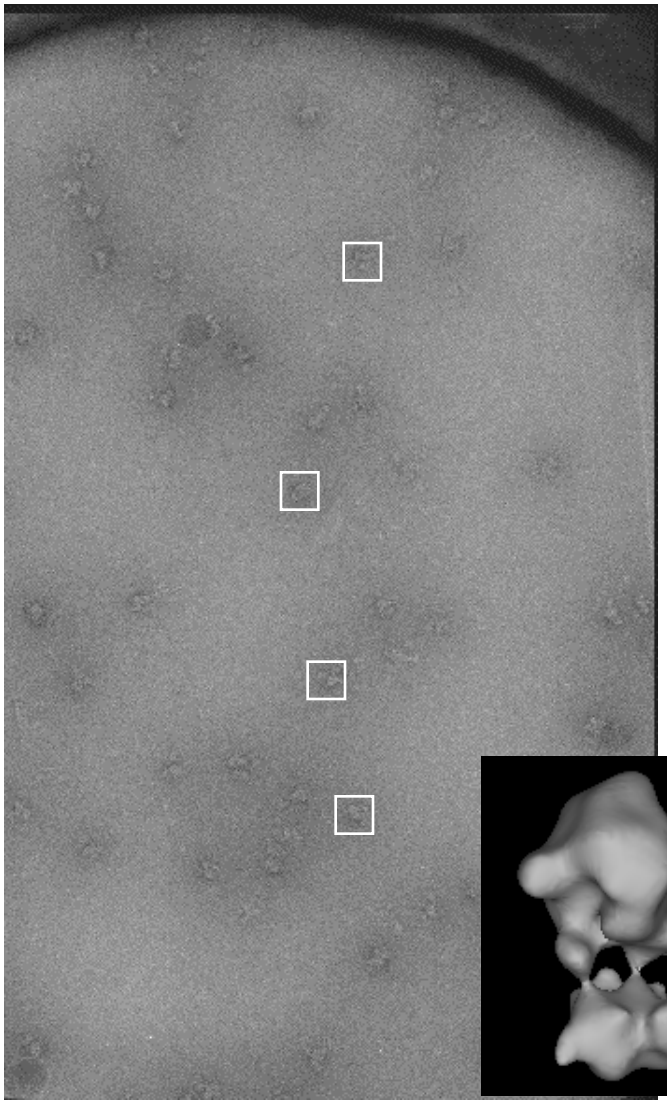


Jurica et al. 2002, unpublished

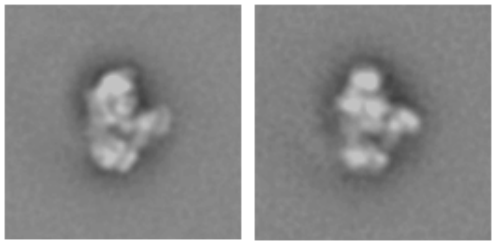
Random Conical Tilt



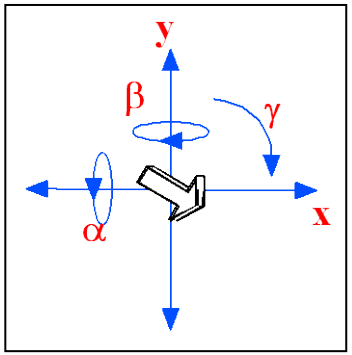
40° tilt

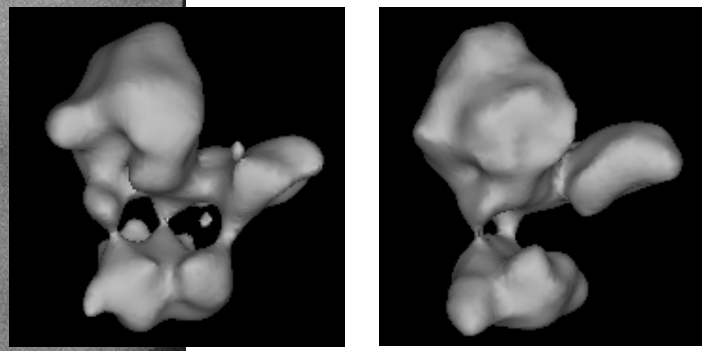


Untilted



Class Averages

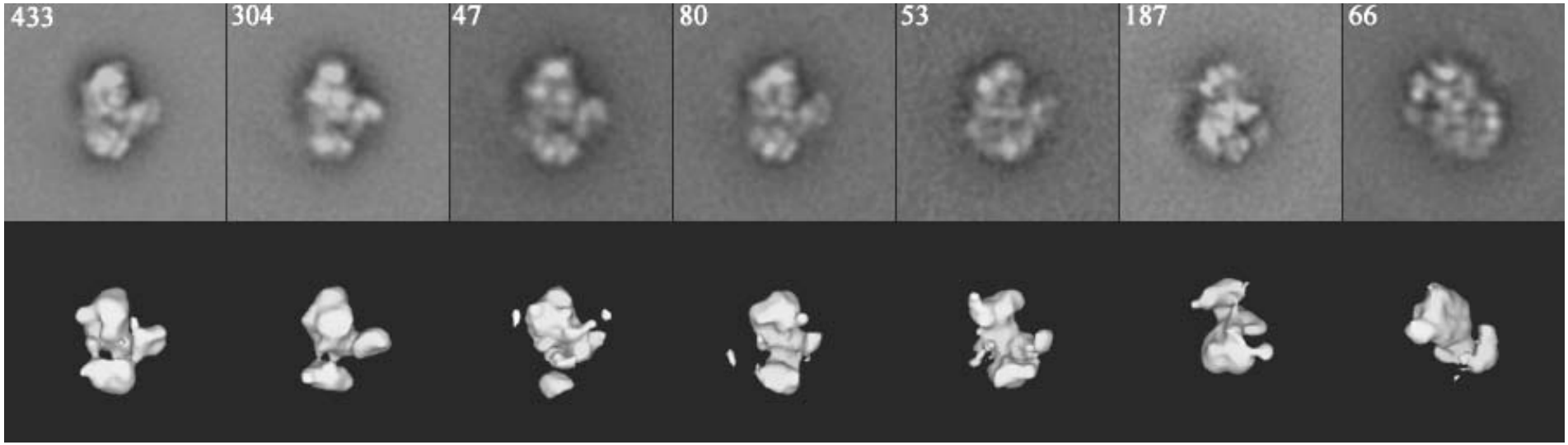




Structures

Jurica et al. 2003

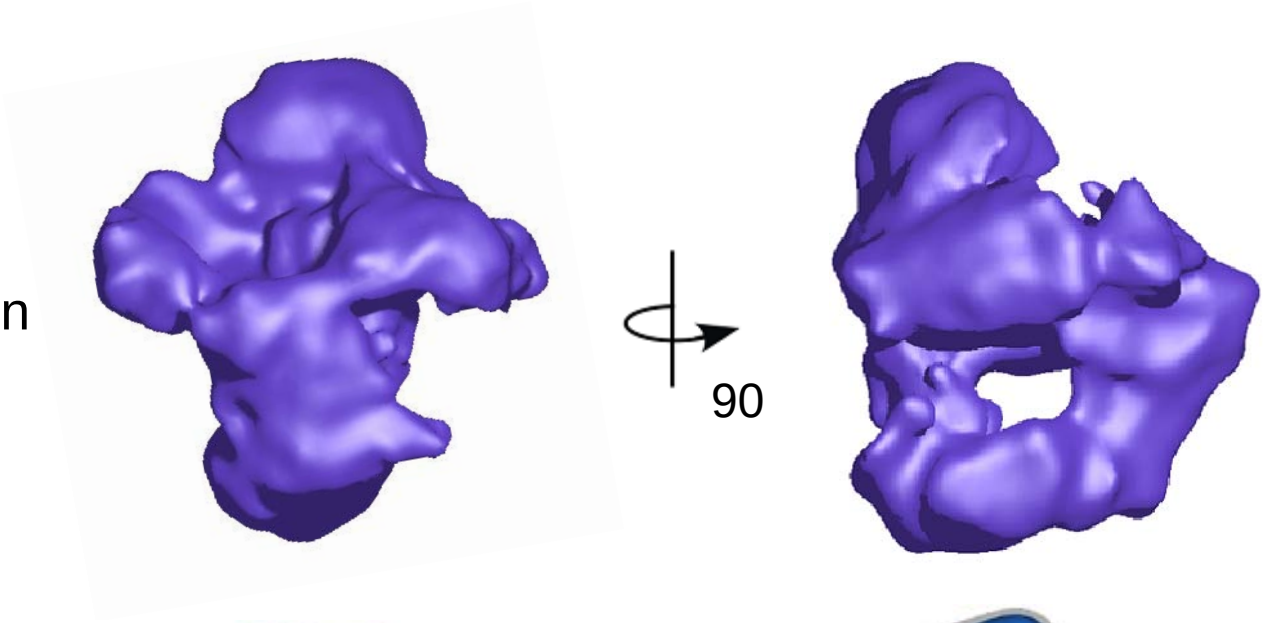
Classification



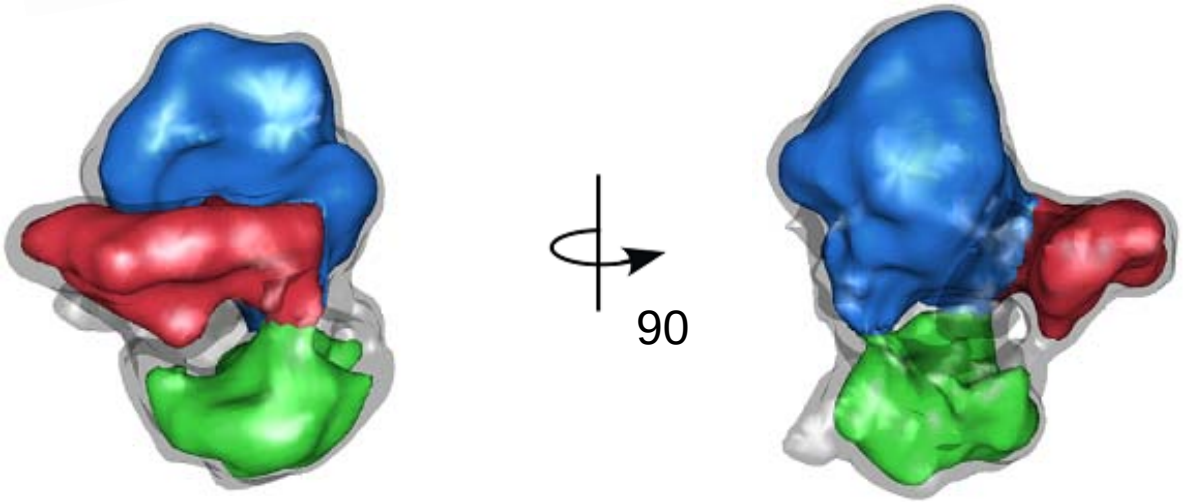
Jurica et al. 2003

Two Methods - Two Structures

Angular reconstitution (no tilts)

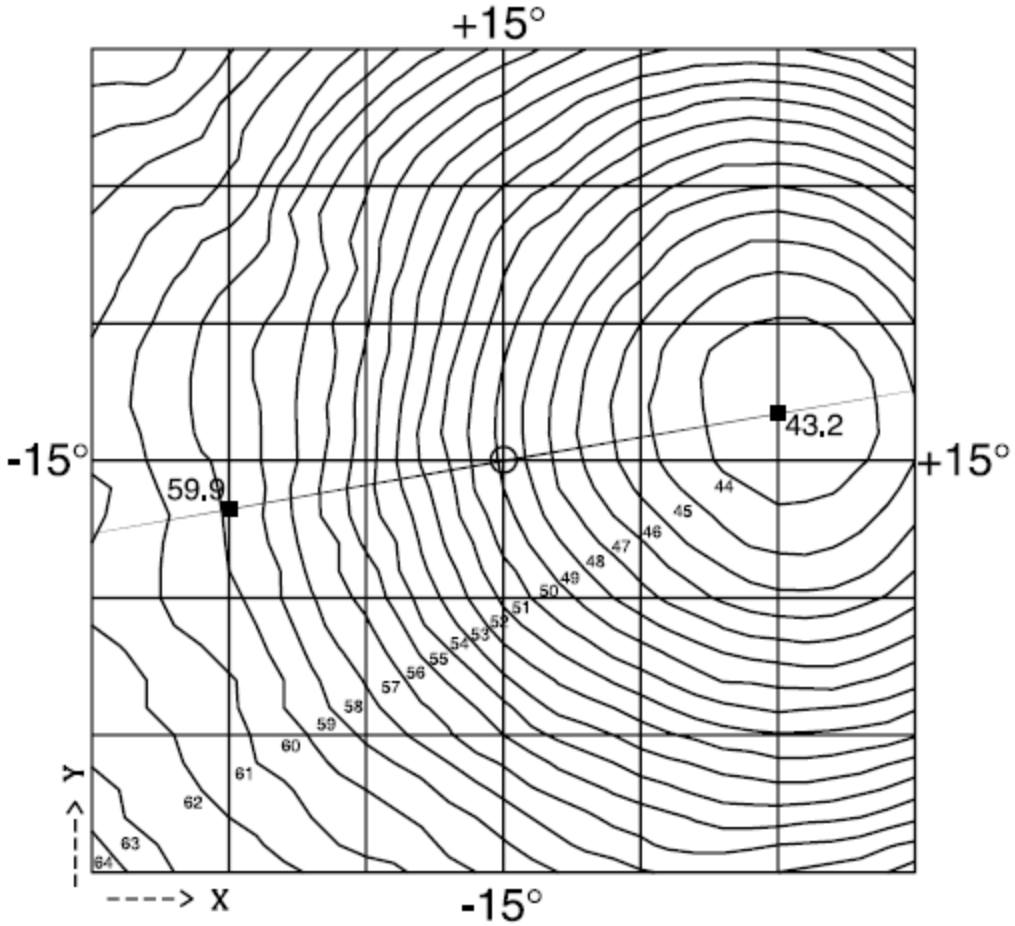


Random conical tilt



Jurica et al. 2003

Tilt Experiments

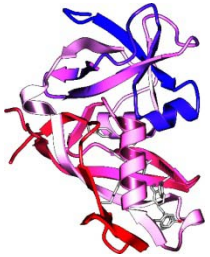


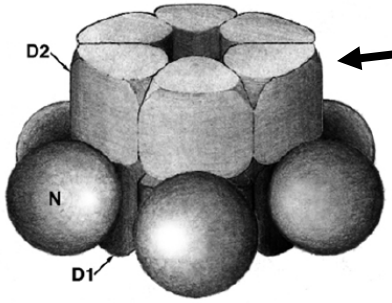
Rosenthal & Henderson 2003

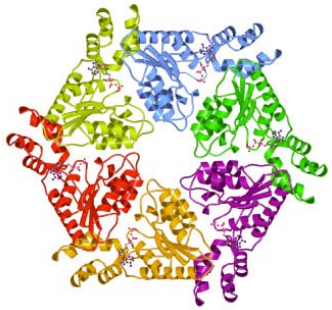
N-ethylmaleimide Sensitive Factor

Yu et al. (1999)
May et al. (1999)

N





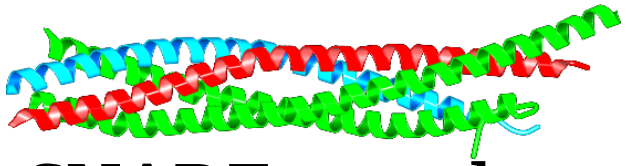


D2

Yu et al. (1998)
Lenzen et al. (1998)

α-SNAP

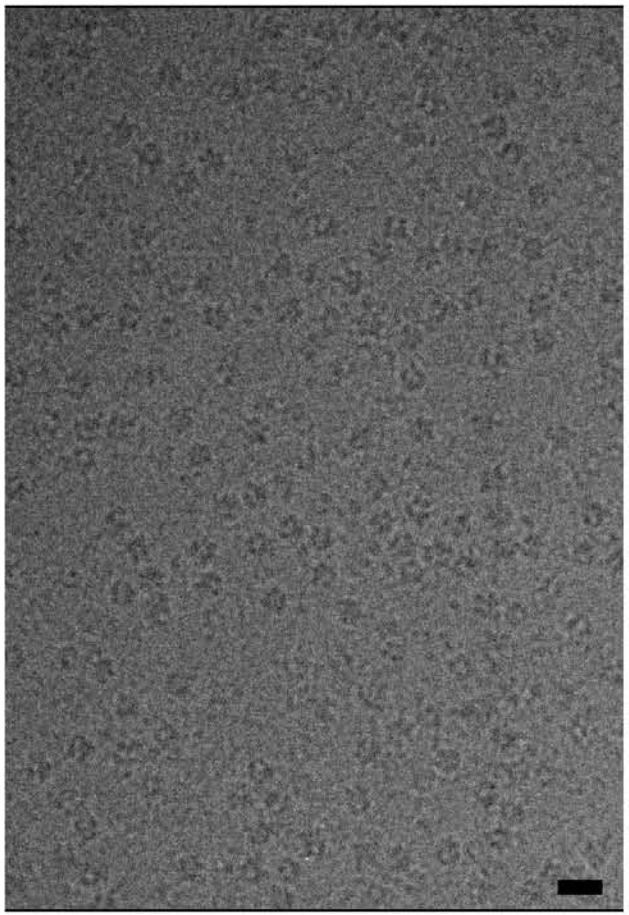
Rice& Brünger (1999)



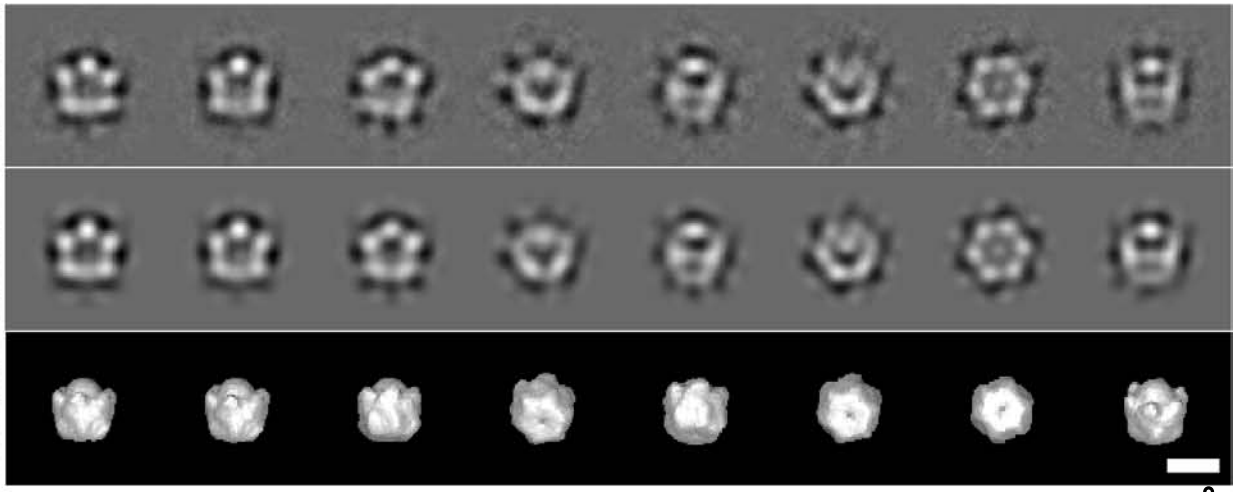
SNARE complex

Sutton et al. (1998)

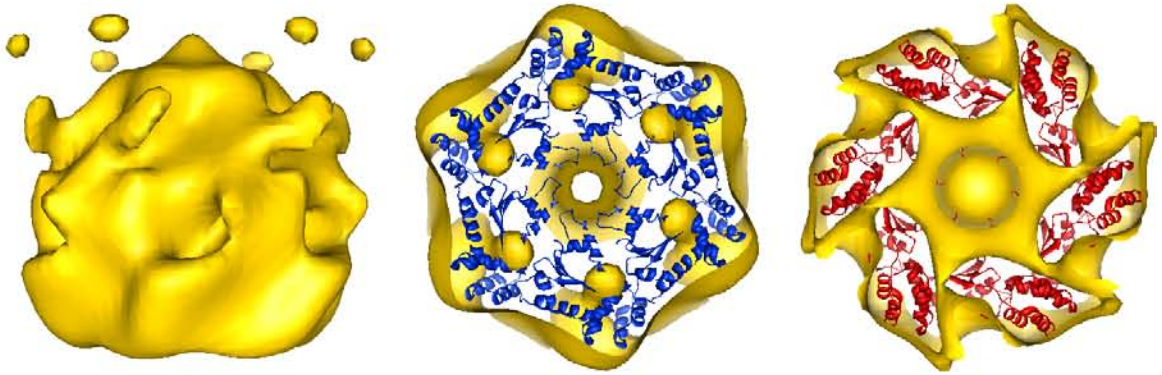
Reconstruction



200 Å

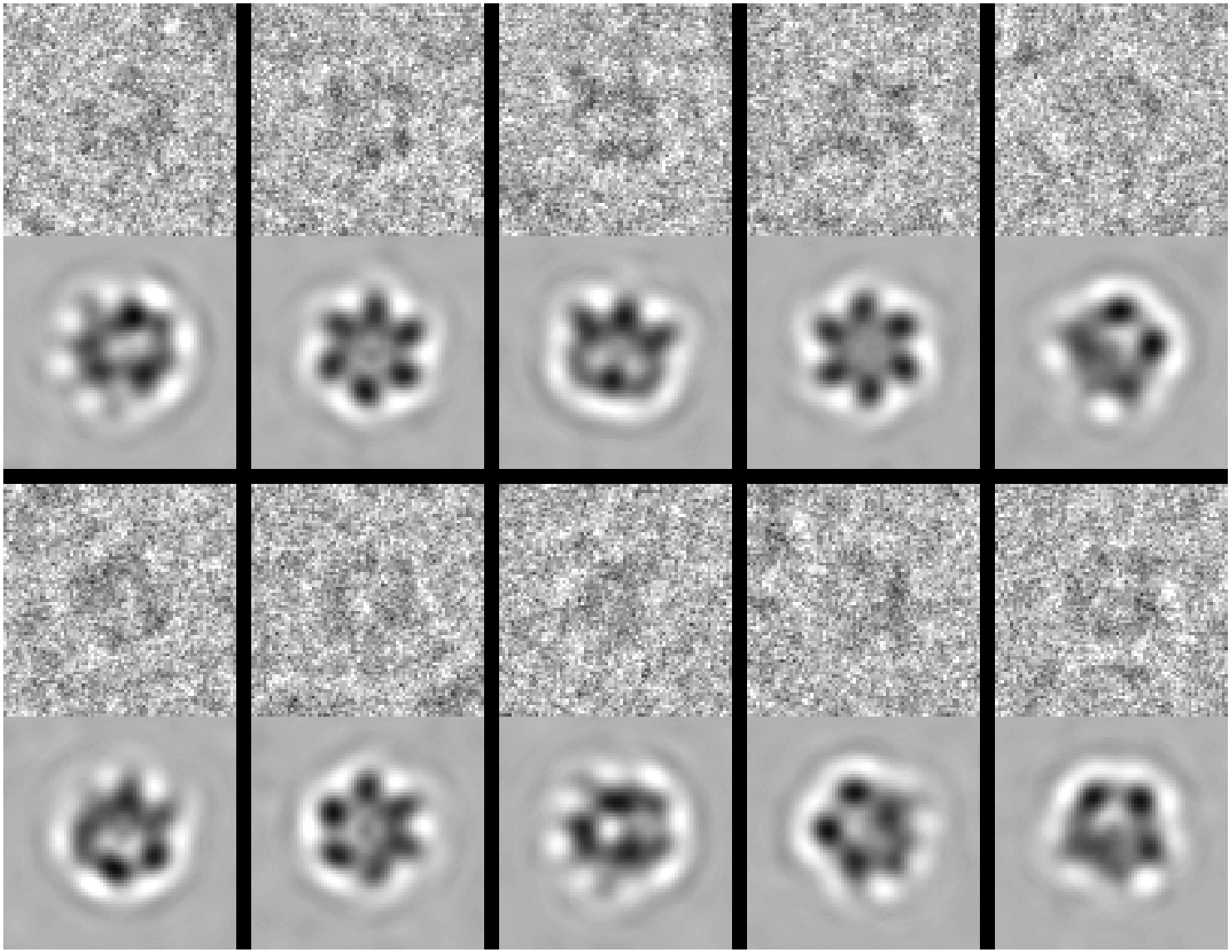


100 Å

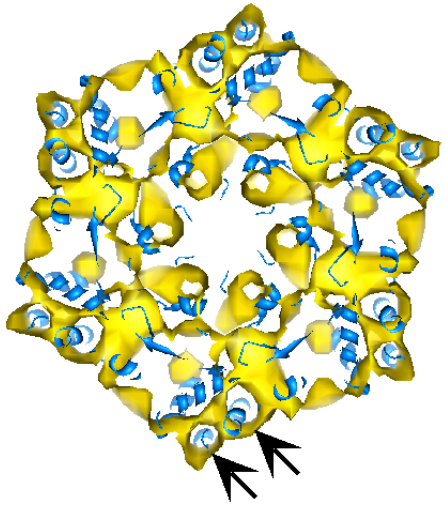


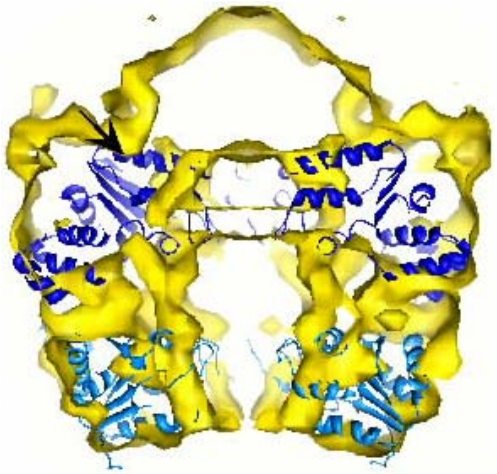
Fürst et al. 2003

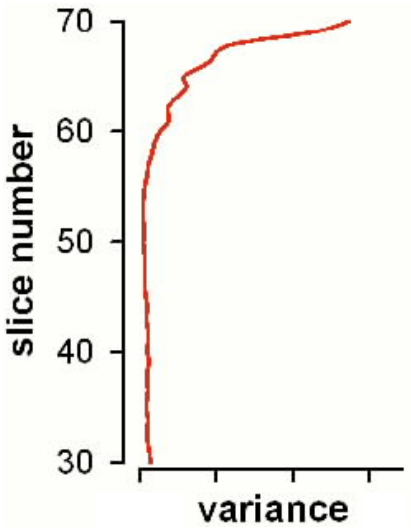
Matching References

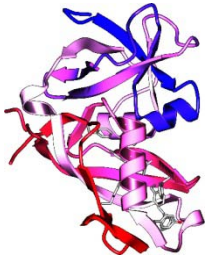


Interpretation?



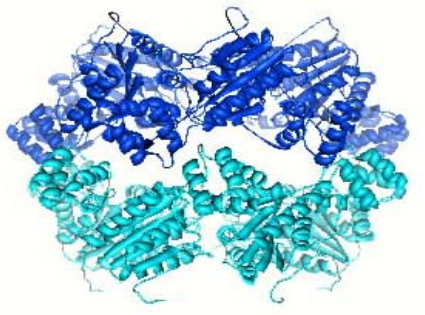




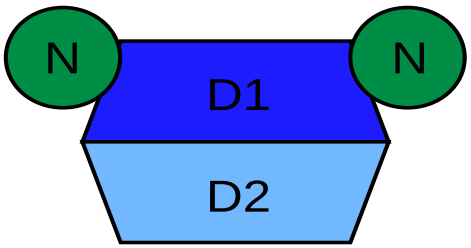


Yu et al. 1999
May et al. 1999

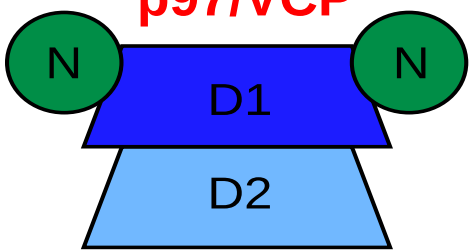
Sutton et al. 1998



Fürst et al. 2003



p97/VCP

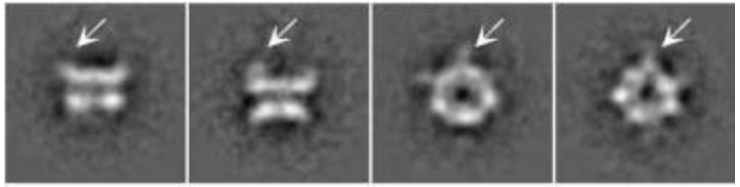


Detecting Problems

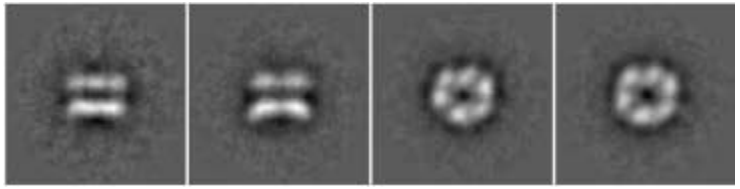
- Often not straight forward!
- Does it look like a ball?
- Is it hollow?
- Does the reference match the particles?
- Does is correlate with known structures?
- Can the high-resolution details be verified?
- Does it make sense (biology, molecular mass)?
- How does the structure refine?
- Is there heterogeneity (variance, classification)?

Different Types of Heterogeneity

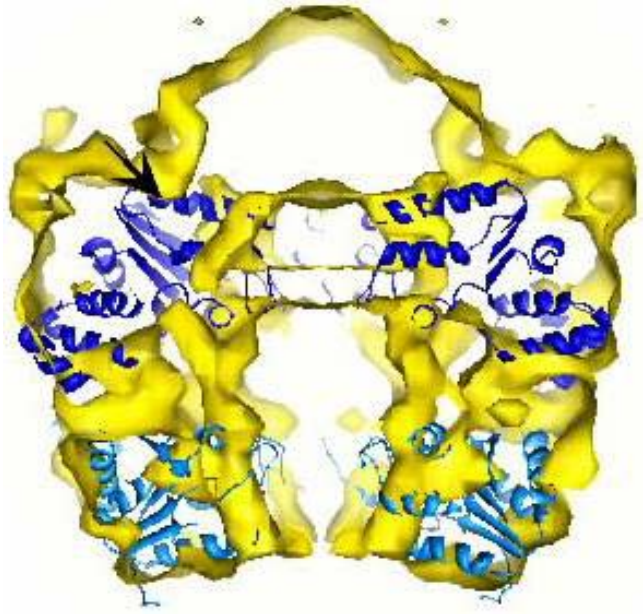
Conformational Heterogeneity

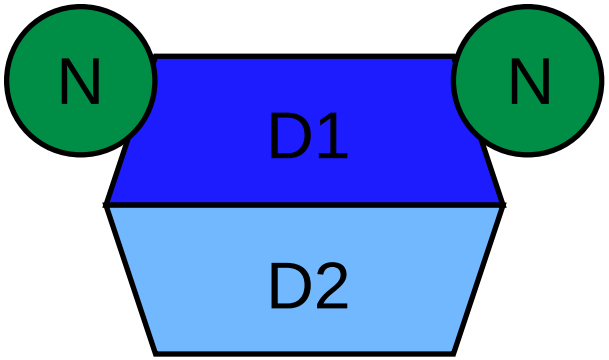


NSF



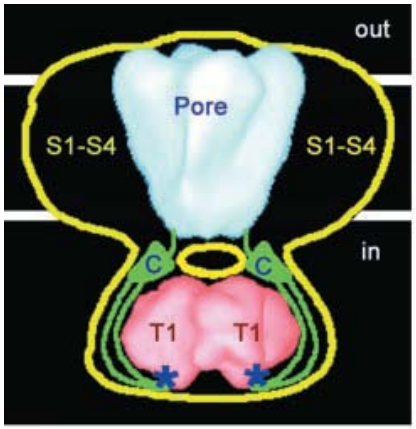
NSFΔN



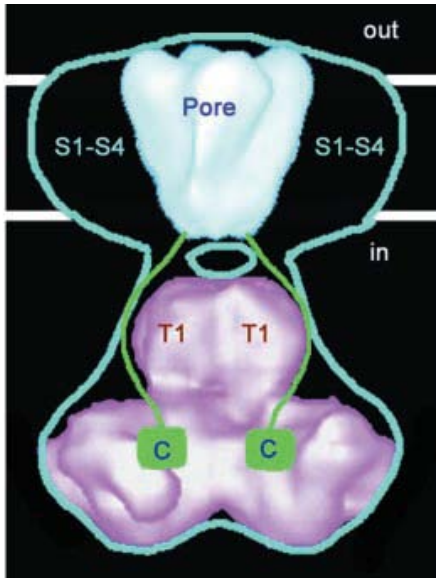


Fürst et al. 2003

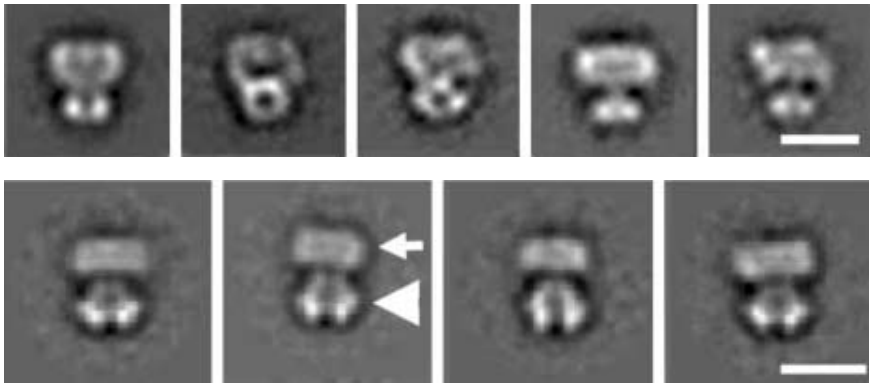
Compositional Heterogeneity



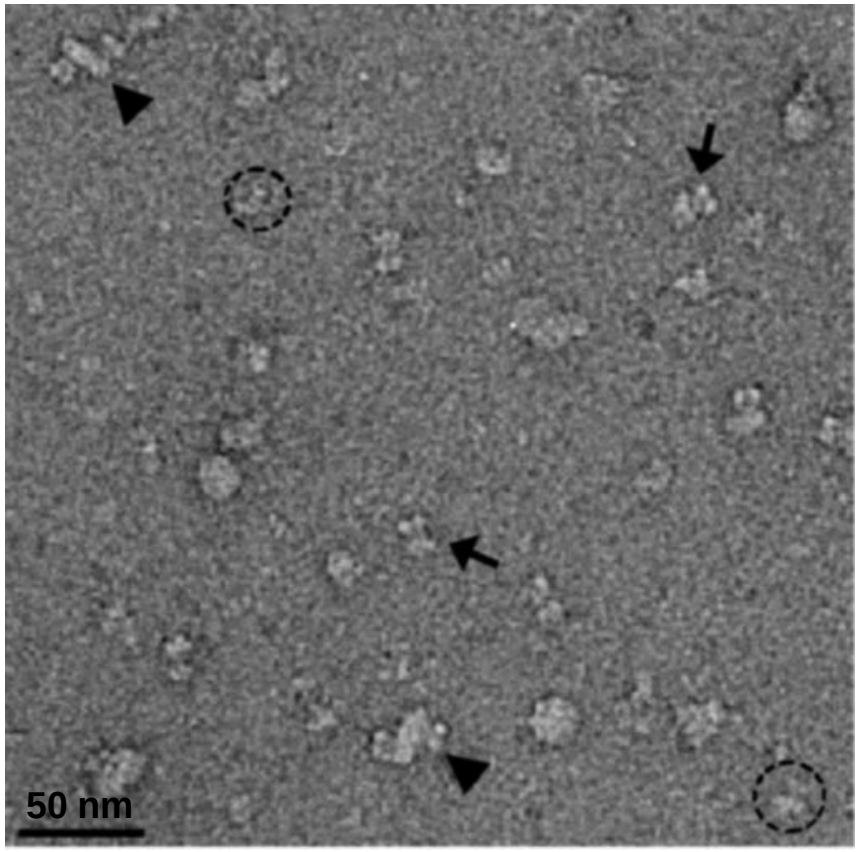
Shaker α_4



Shaker $\alpha_4\beta_4$



10 nm



Sokolova et al. 2003

Classification Methods…

Fred Sigworth

ML-Like Restraints & Classification

Poor Man's Maximum Likelihood

$$L(\Theta) = \sum_{i=1}^{N} \ln \int p_i(\phi, X_i \mid \Theta) d\phi$$

p_i = joint probability

$$\Theta = (A, \sigma, x_0, y_0, \sigma_{xy})$$

If SNR high, then p_i essentially zero everywhere except when particle aligned with reference (p_i similar to delta function):

$$\phi_i = \text{argmax}\left[p_i(\phi, X_i \mid \Theta)\right] = \text{argmax}\left[\ln p_i(\phi, X_i \mid \Theta)\right]$$
$$= \text{argmax}\left[X_i \bullet A + \sigma^2 \ln f(\phi \mid \Theta)\right]$$

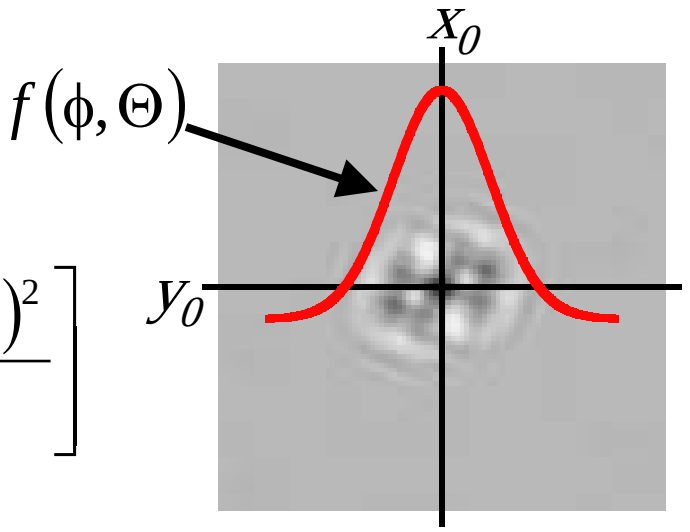
X_i : image I $\quad$ σ : standard deviation of noise in image

A : reference image $\quad$ ϕ : particle params

Sigworth 1998

Parameter Restraints

$$\phi_i = \text{argmax}\left[X_i \bullet A + \sigma^2 \ln f\left(\phi \mid \Theta\right)\right]$$



$$f\left(\phi,\Theta\right) = \frac{1}{4\pi^2\sigma_{xy}^2}\exp\left[-\frac{\left(x-x_0\right)^2+\left(y-y_0\right)^2}{2\sigma_{xy}^2}\right]$$

X_i : image i

A : reference image

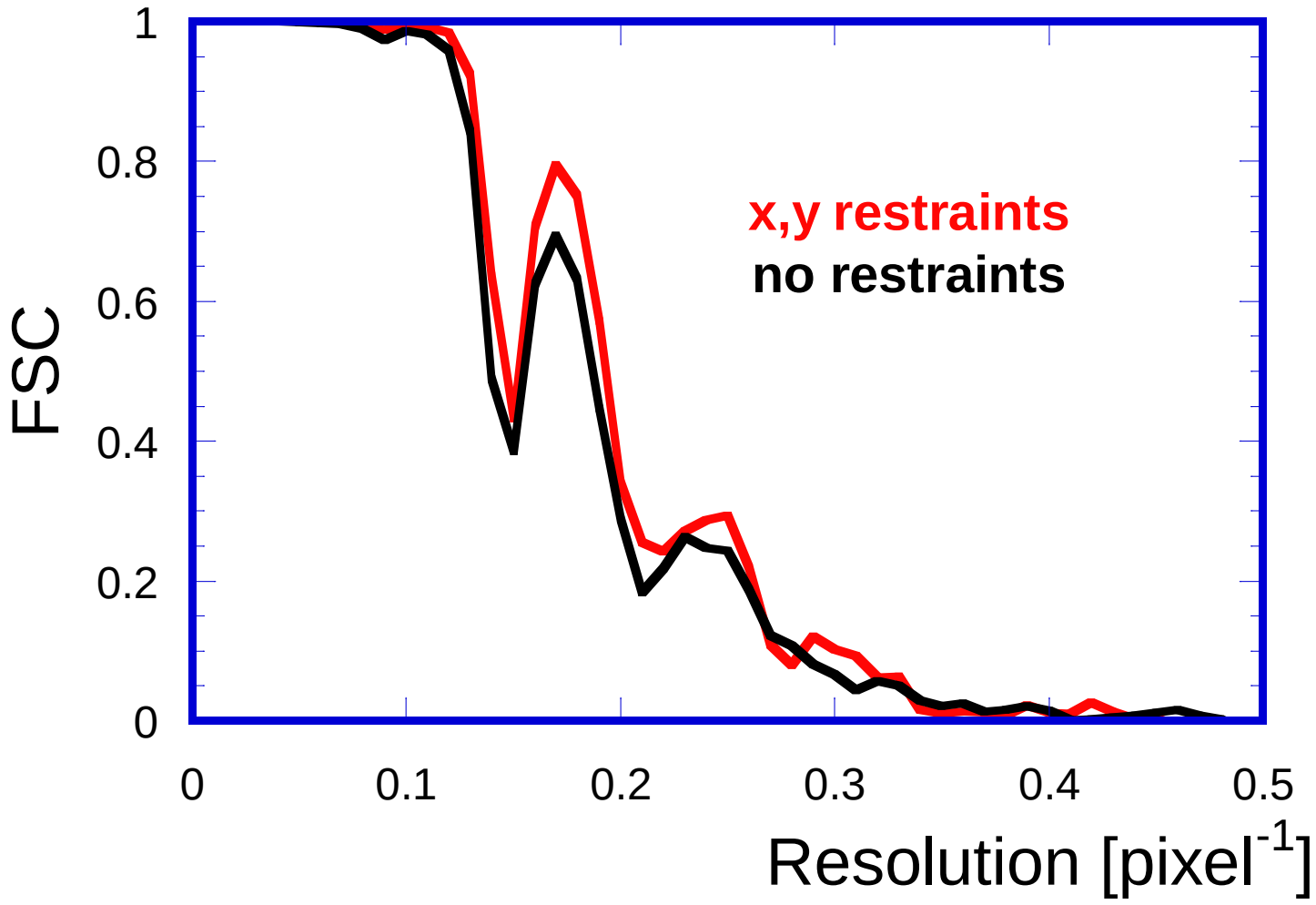
ϕ : particle params

σ : standard deviation of noise in image

x_0, y_0: average x,y coords in data set

σ_{xy} : std. deviation of x,y coords

Computer Simulation



Poor Man's ML Classification

Assume K classes with class averages A_k:

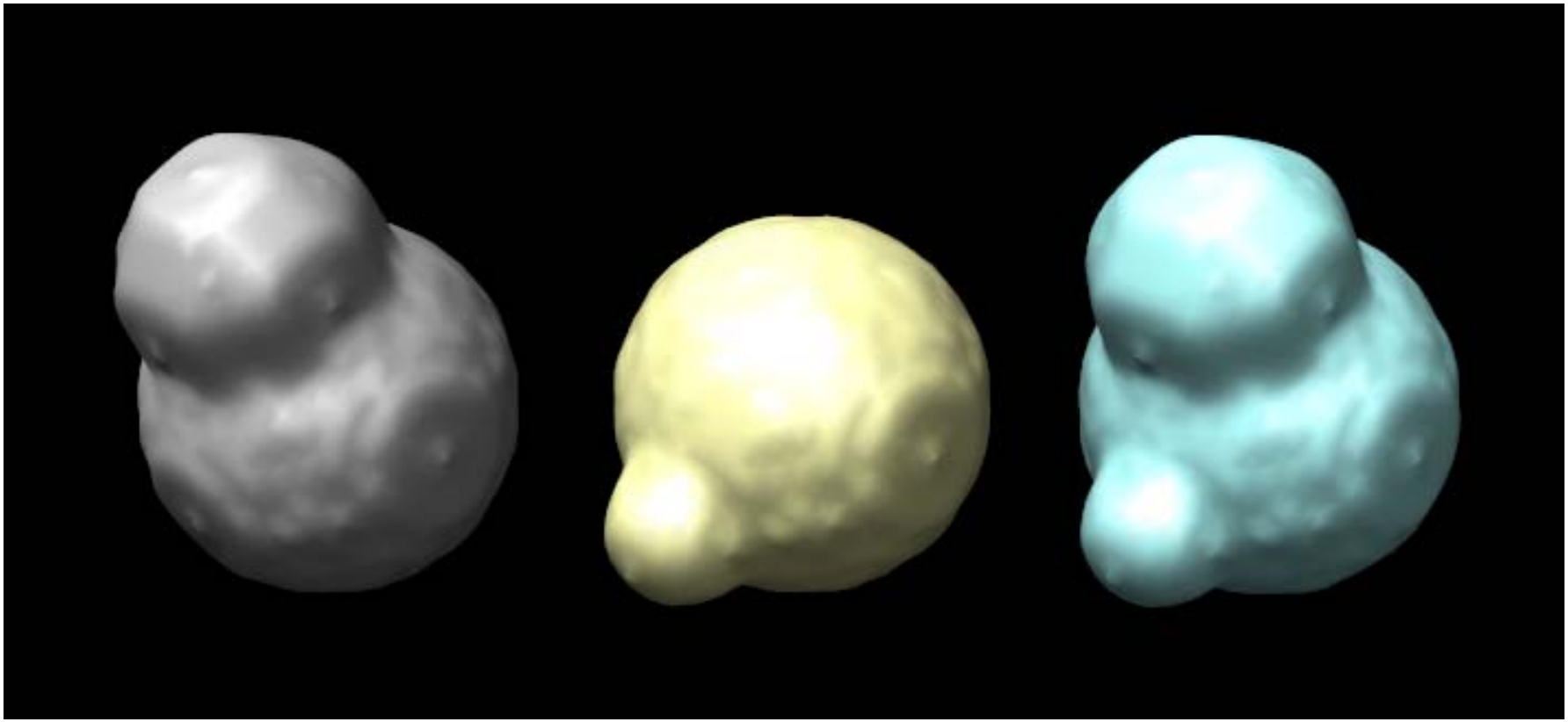
$$A_k^{(n+1)} = \frac{\sum_{i=1}^{N} X_i\left(\phi_{i,k}\right) q_{i,k}^{(n+1)}}{\sum_i q_{i,k}^{(n+1)}}$$

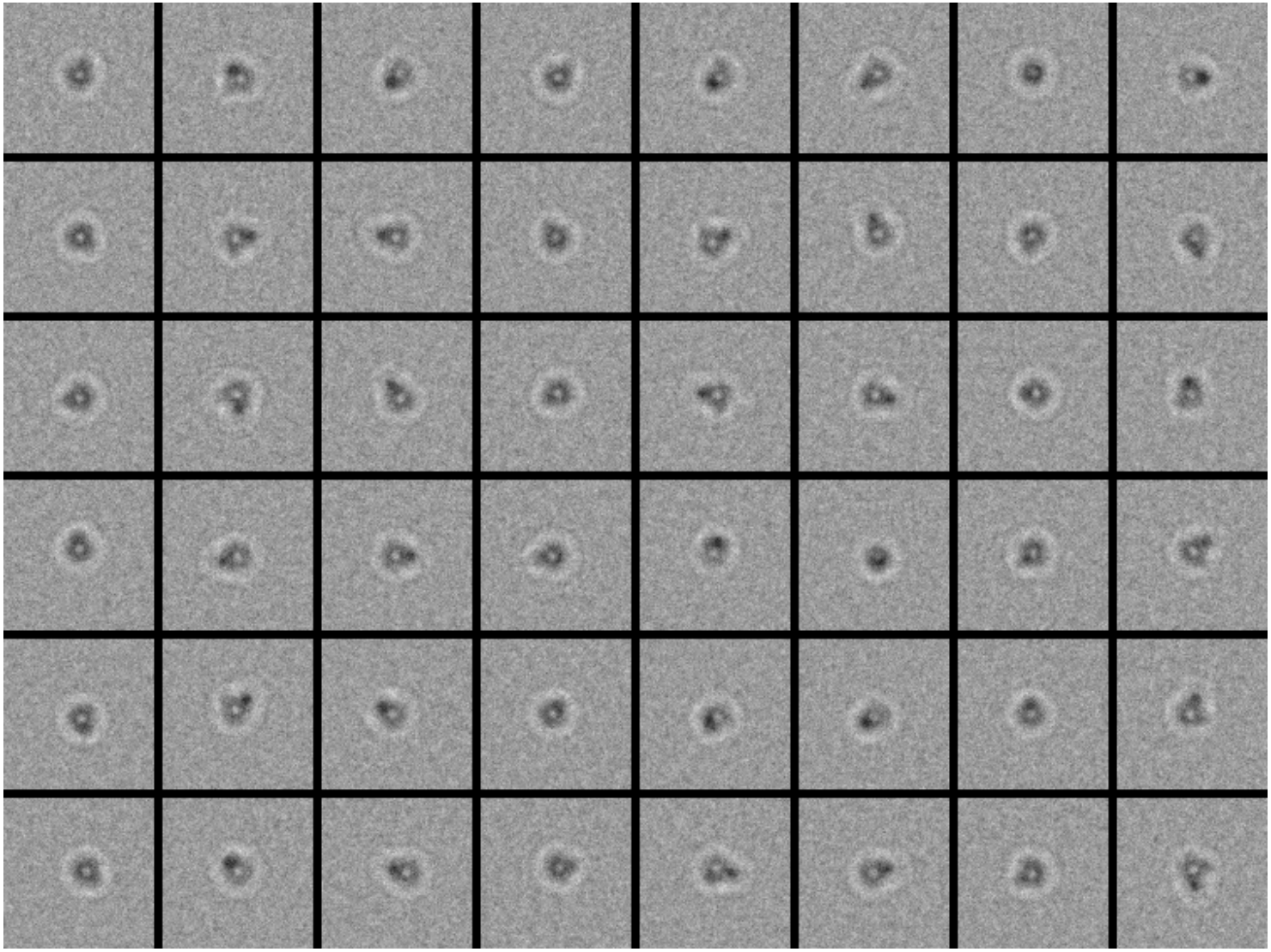
$$q_{i,k}^{(n+1)} = p_{i,k}^{(n)}\left(\phi_{i,k}, X_i \middle| \Theta_k^{(n)}\right) a_k^{(n)} \Big/ \sum_{k=1}^{K} p_{i,k}^{(n)}\left(\phi_{i,k}, X_i \middle| \Theta_k^{(n)}\right) a_k^{(n)}$$

$$a_k^{(n)} = \frac{\sum_{i=1}^{N} q_{i,k}^{(n)}}{N}$$

$$\phi_{i,k} = \operatorname{argmax}\left[X_i \bullet A_k + \sigma^2 \ln f\left(\phi \middle| \Theta\right)\right]$$

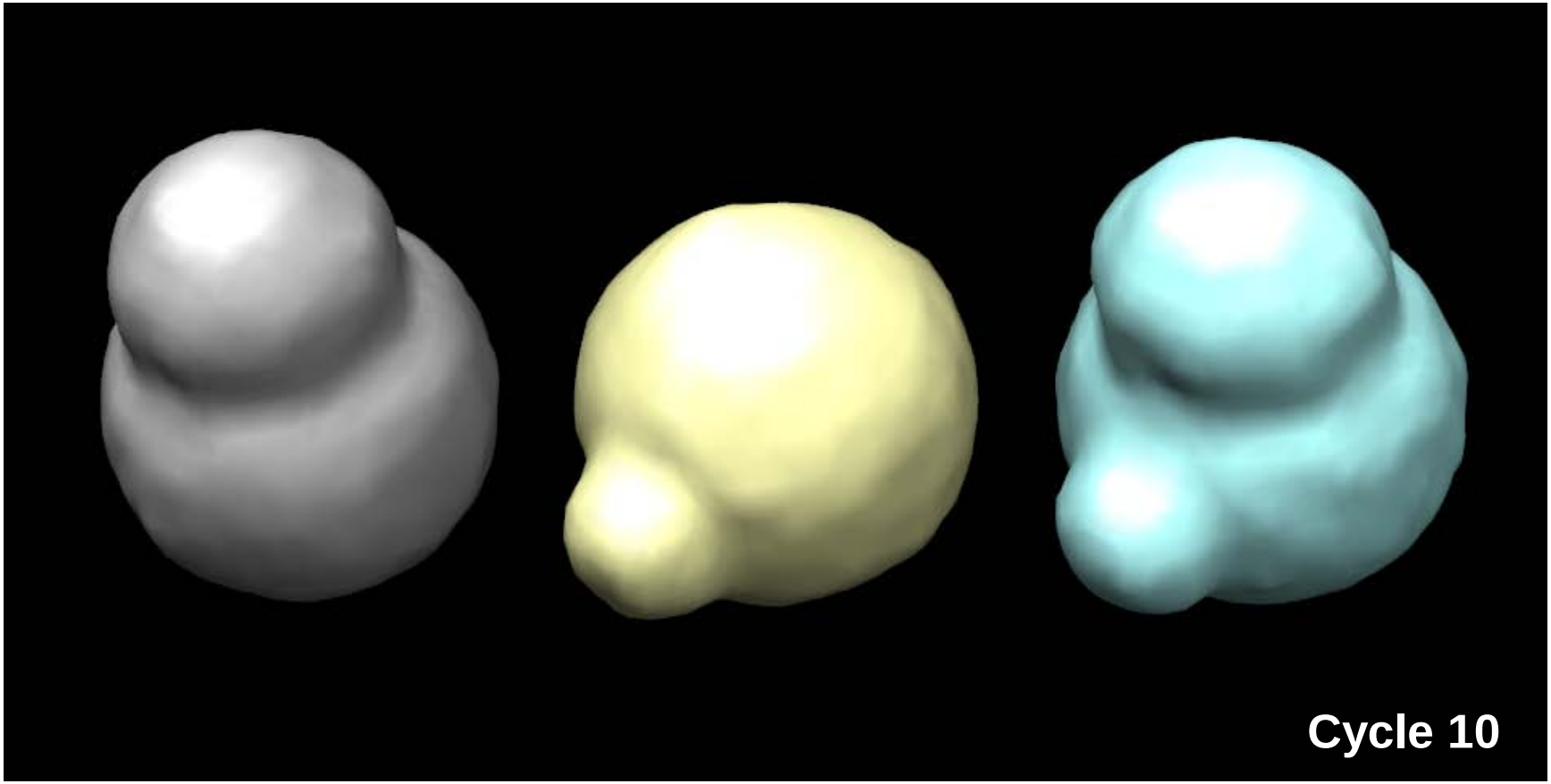
Test Structures





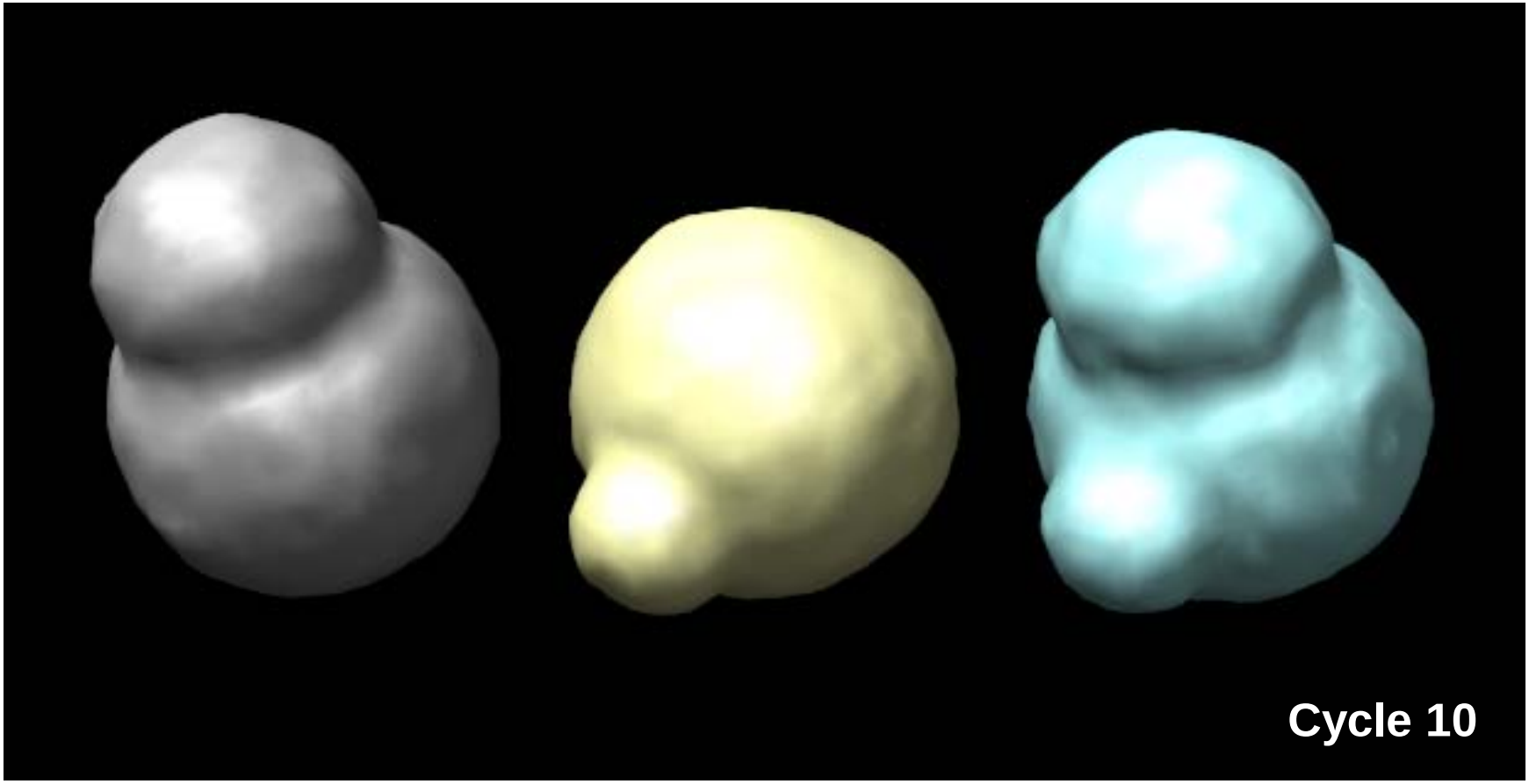
10000 images of each structure in random orientations SNR ~ 1

Correlation Classification

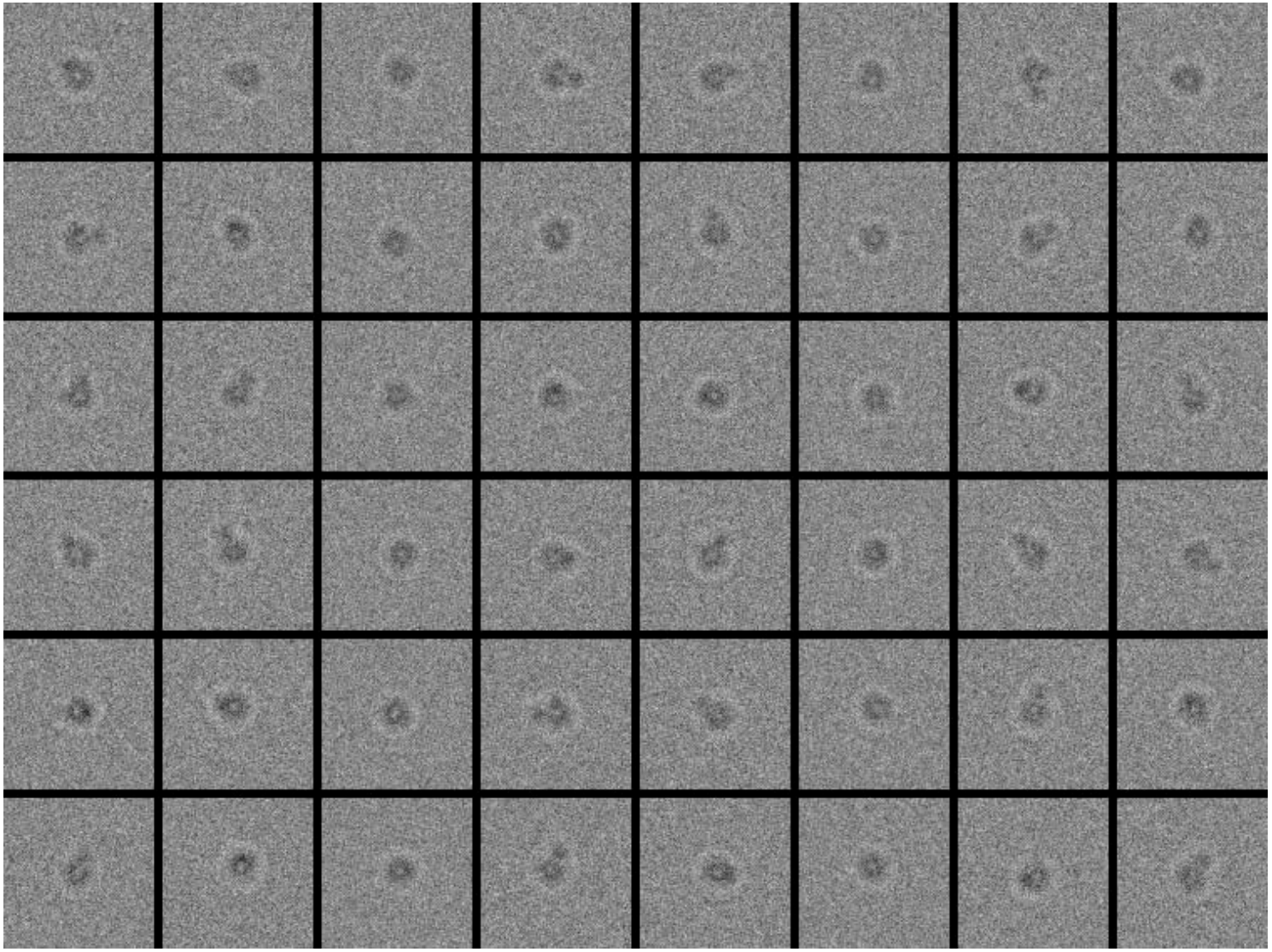


SNR ~ 1 **Correct: 99.2%**

ML-Like Classification

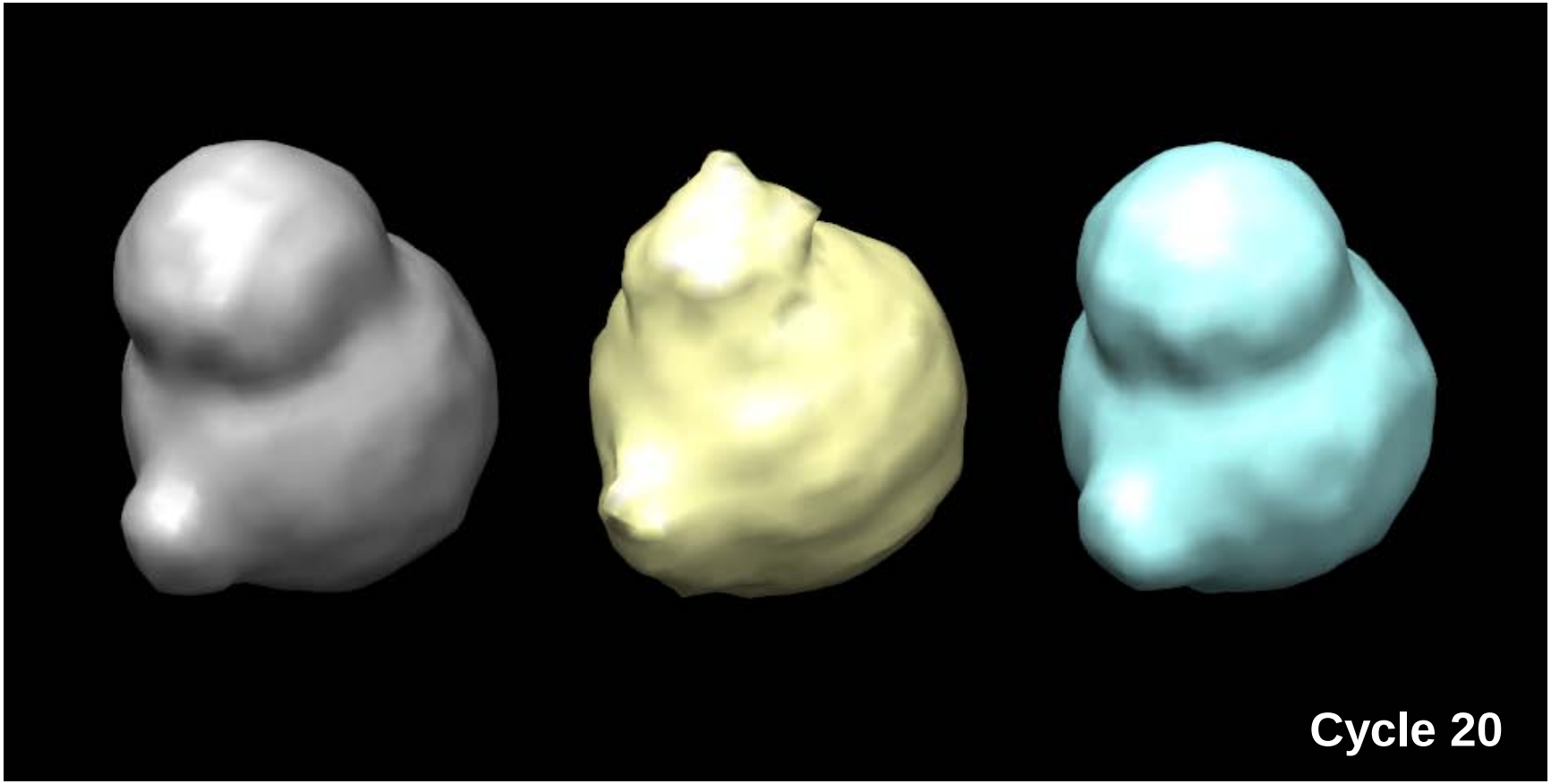


SNR ~ 1 **Correct: 94.3%**



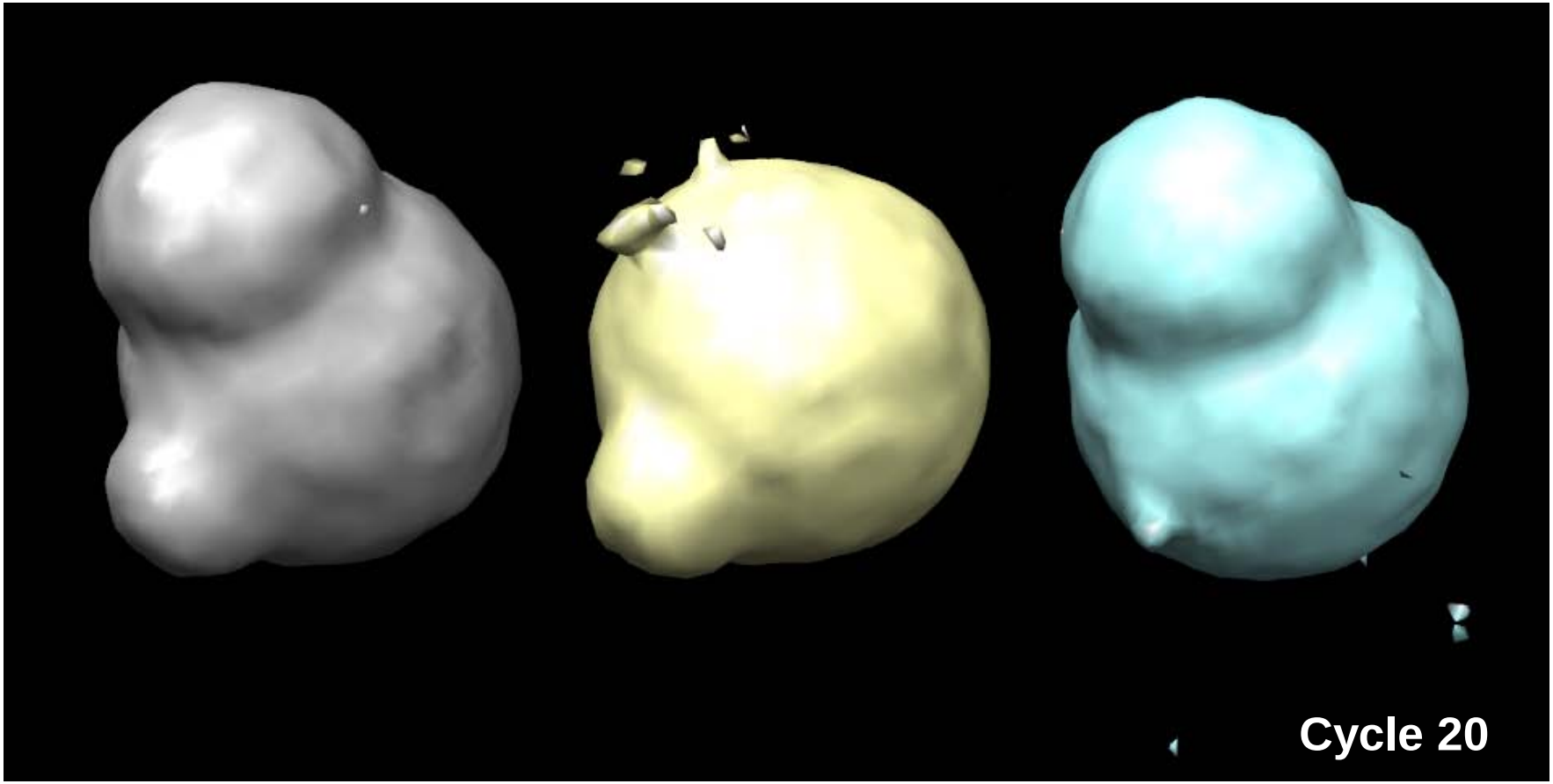
10000 images of each structure in random orientations SNR ~ 0.1

Correlation Classification



SNR ~ 0.1 **Correct: 62.6%**

ML-Like Classification



SNR ~ 0.1 **Correct: 86.5%**

Continuous Deformation Models…

Fred Sigworth