NCMI Workshop on Single Particle Processing and Visualization - 2007

Thanks to our Sponsors:
Agouron Institute
TLC², University of Houston
National Center for Research Resources
National Institute of General Medical Sciences
JEOL, Inc.

Steve Ludtke
sludtke@bcm.edu

NCRR

NCMI
National Center for Macromolecular Imaging

Announcements

- If you have a dot, you have a spot in the teaching classroom
- For cab rides back to the airport, there is a signup sheet at the registration desk
- Daily Agenda

LAPTOP USERS

- Dot on your badge if you will have a desktop in the afternoon

- Copy the entire DVD to your linux/OSX partition
- After copying, make sure permissions are set:
- chmod –R u+w *demo
- Check software installation:
- cd eman-demo/samples
- v2 imagic.hed

Workshop Materials

- Note that all workshop materials, including final versions of presentations will be archived at:

http://ncmi.bcm.edu/ncmi/events/workshops/workshops_66

- Use the new EMAN/EMAN2 Wiki for questions:

http://blake.bcm.edu/emanwiki

- make an account for yourself
- subscribe to the page to be notified when there is an answer

Topics

- Single Particle Reconstruction
  - EMAN, EMAN2 (Steve Ludtke), SPARX(Pawel Penczek)
- Visualization
  - Chimera (Tom Ferrin & Tom Goddard), MGLTools/Vision (Yong Zhao)
- Map Interpretation
  - AIRS (Matt Baker)
- Homology Modeling
  - Modeller (Ben Webb & Maya Topf)
- Theory & Background (Wah Chiu, Pawel Penczek, Kim Henrick, Mike Schmid, Mike Marsh, Ian Rees)

Recommended Computer
(Jan 2007)

- Core 2 duo
- 2-4 gigs memory
- 250+ gig SATA hard drive
- Good NVidia or ATI graphics card
  - Most expensive models don't help isosurface rendering much
  - 'industrial' cards, like quadro, support stereo in a window, but otherwise the same
- Dual monitor solutions can be very useful

Processor Comparison

Relative Performance (speedtest):

- 180 - SGI Octane R10k 0.25 Ghz
- 400 - PIII 0.8 Ghz
- 1000 - P4/Xeon(old) 3.2 Ghz
- 1800 - Athlon64 3500 2.2 Ghz
- 3400 - Core 2 duo 2.6 Ghz

All numbers are per-core
Athlon64 ~ Opteron at the same speed
Core 2 duo ~ New Xeon at same speed

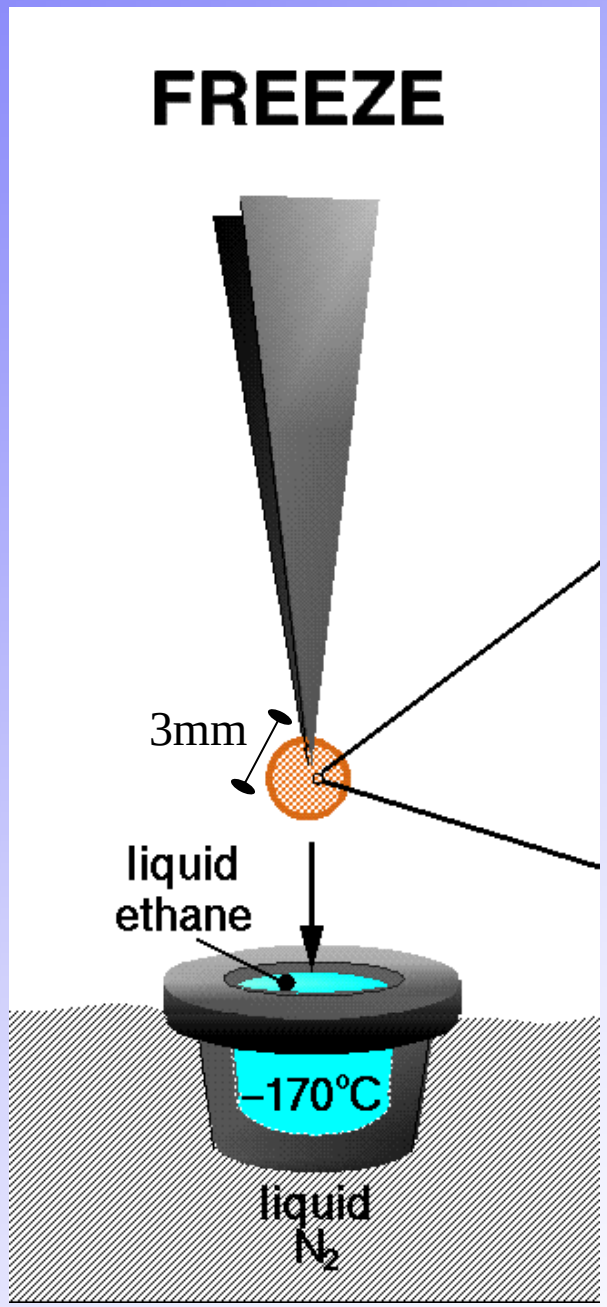
Unit Refresher

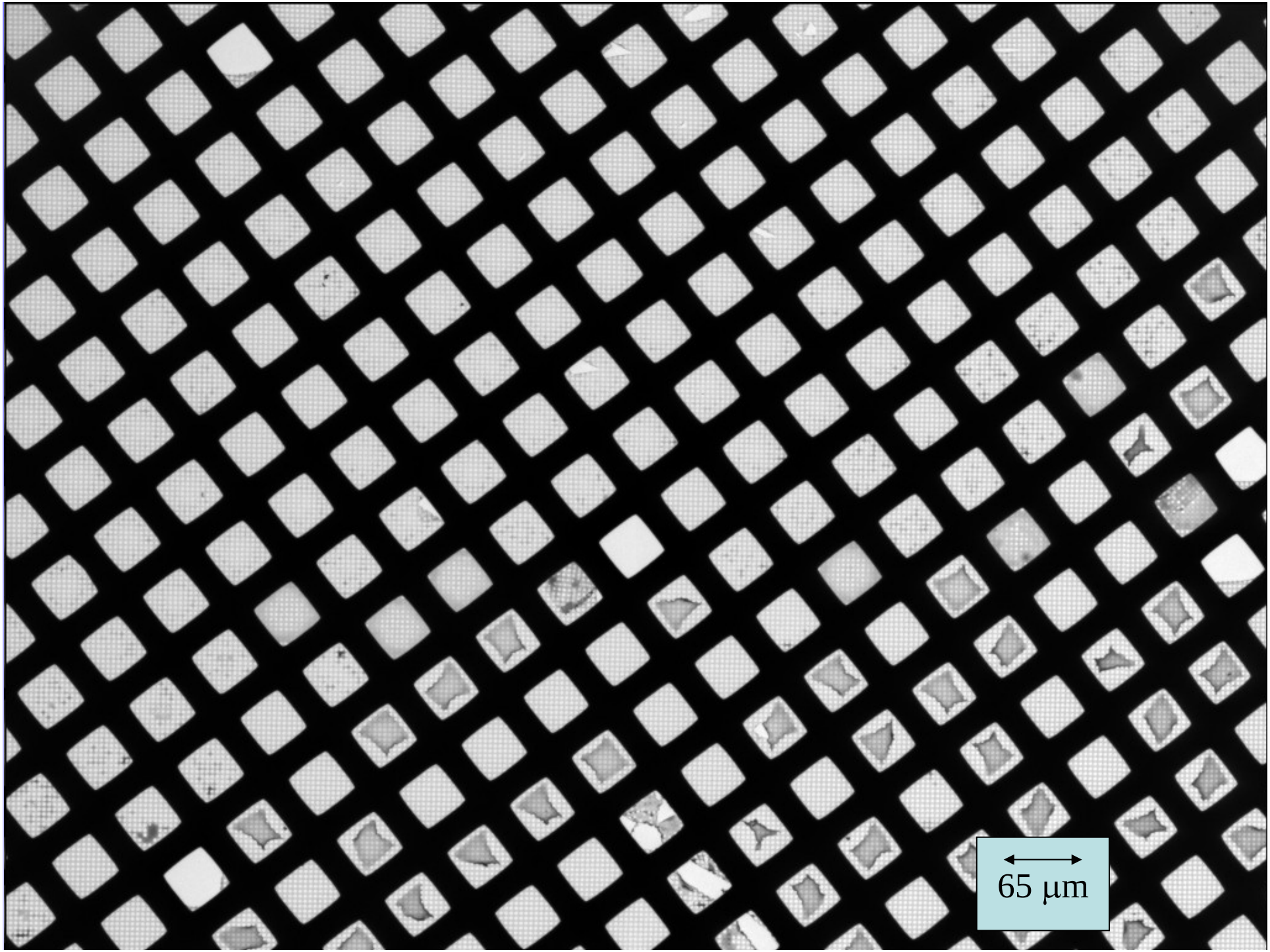
Trivial unit definitions:

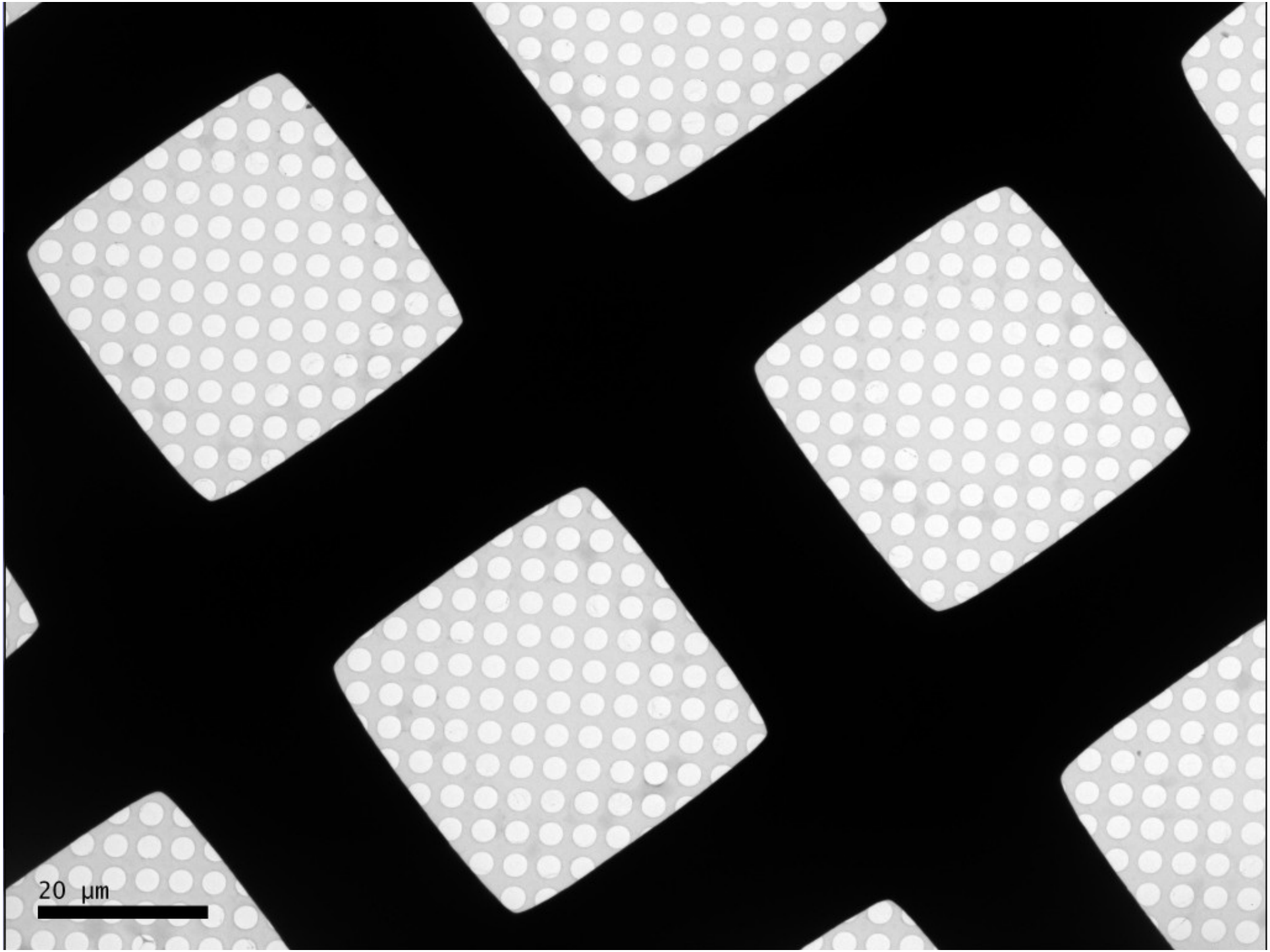
- 1 Angstrom = 1 Å = 0.1 nm ~ size of an atom
- <3 Å resolution required in x-ray crystallography for a backbone trace
- In SPA, defocus usually specified in μm, 1 μm = 10,000 Å
- Rule of thumb for SPA → 3x oversampling, ie – to achieve 3 Å resolution, you need 1 Å/pixel

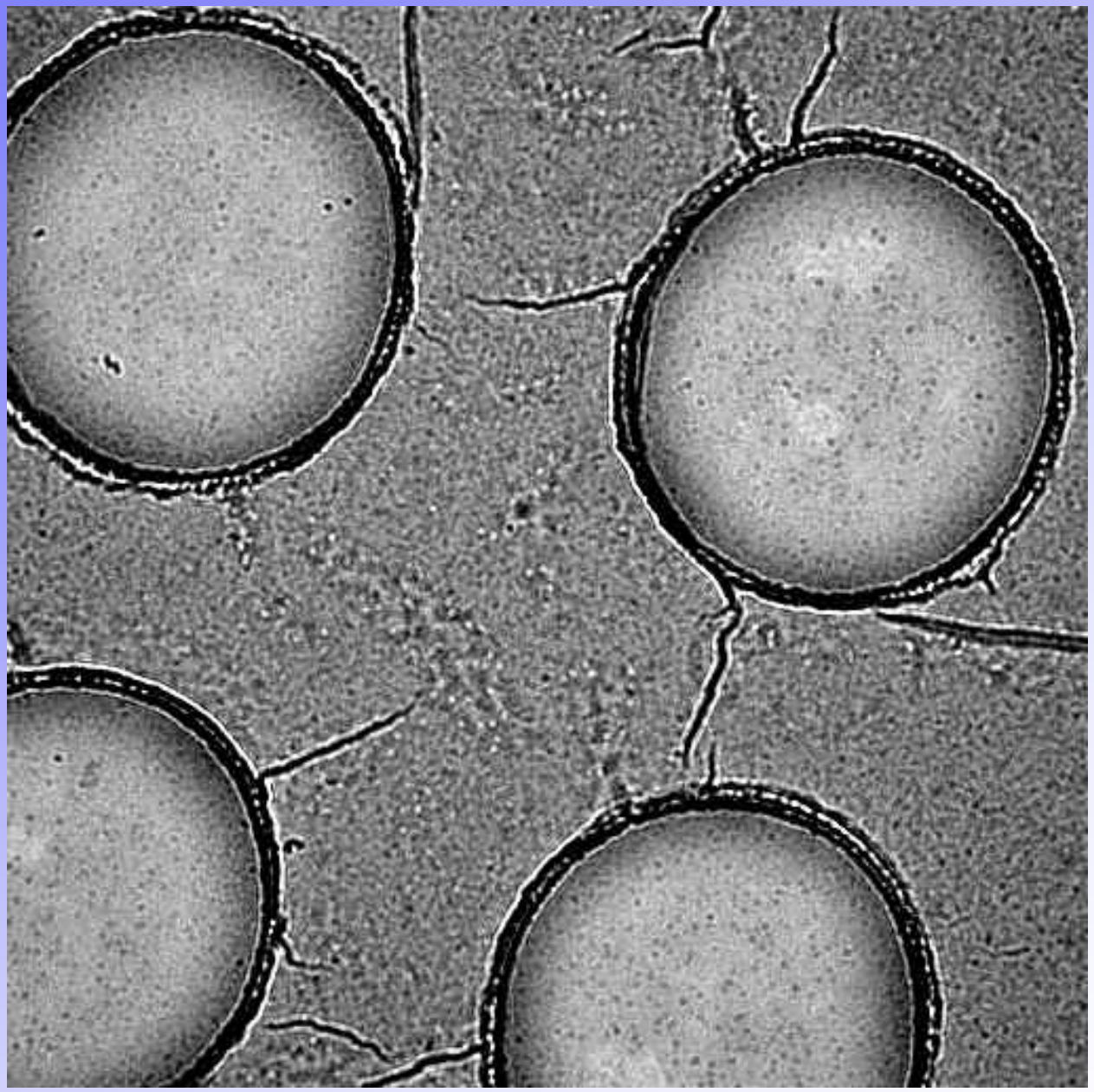
Requirements for Single Particle Analysis

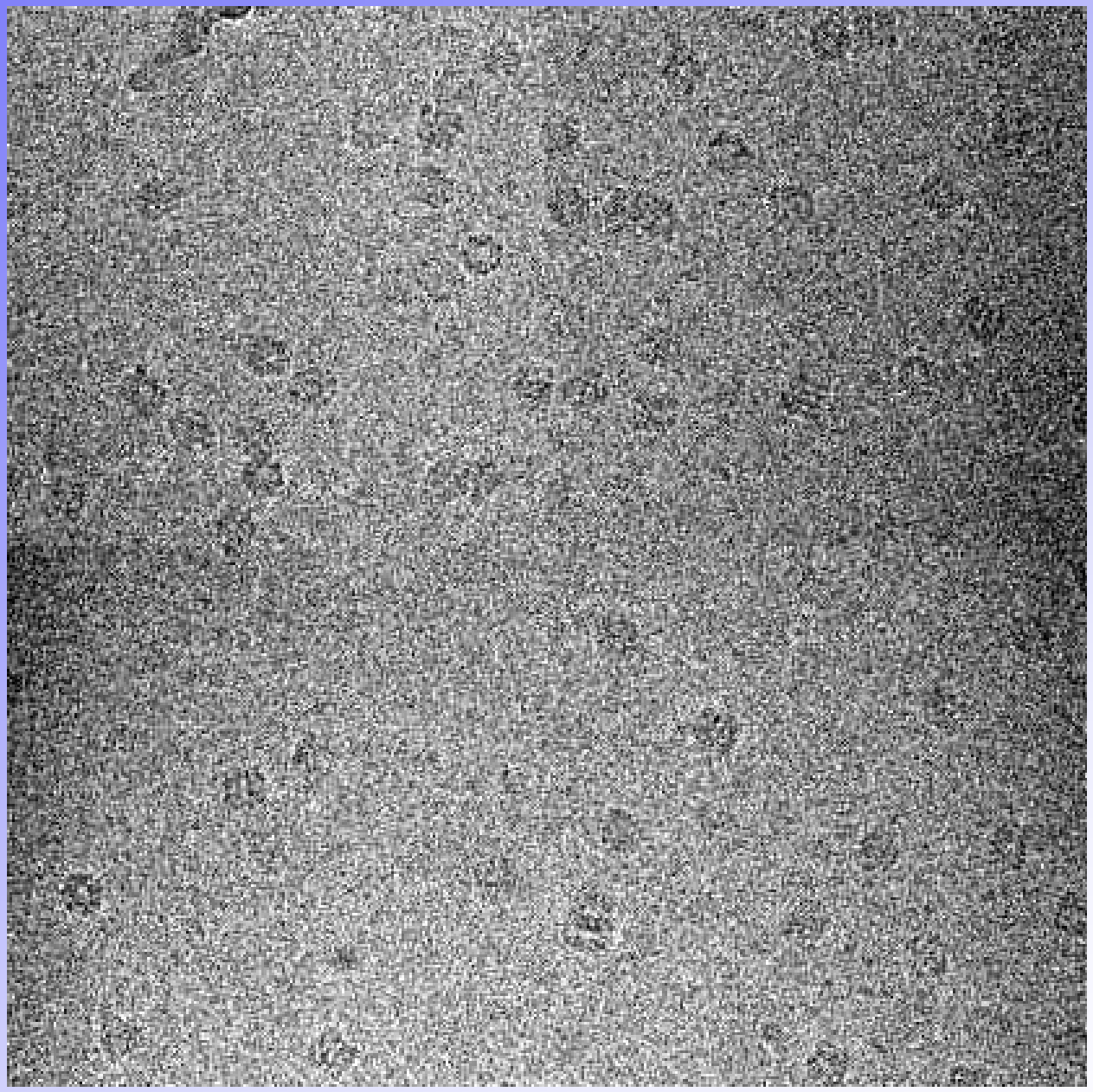
- Soluble, monodisperse
- ~20 μl @ 0.1 – 1 mg/ml (makes ~4 grids)
- Bigger is better (>150 kDa)
- High purity 95%+, 99% is better
- Buffer is important (eliminate glycerol!)
- In theory, 1 grid+1 day → <1nm resolution
  - In practice …

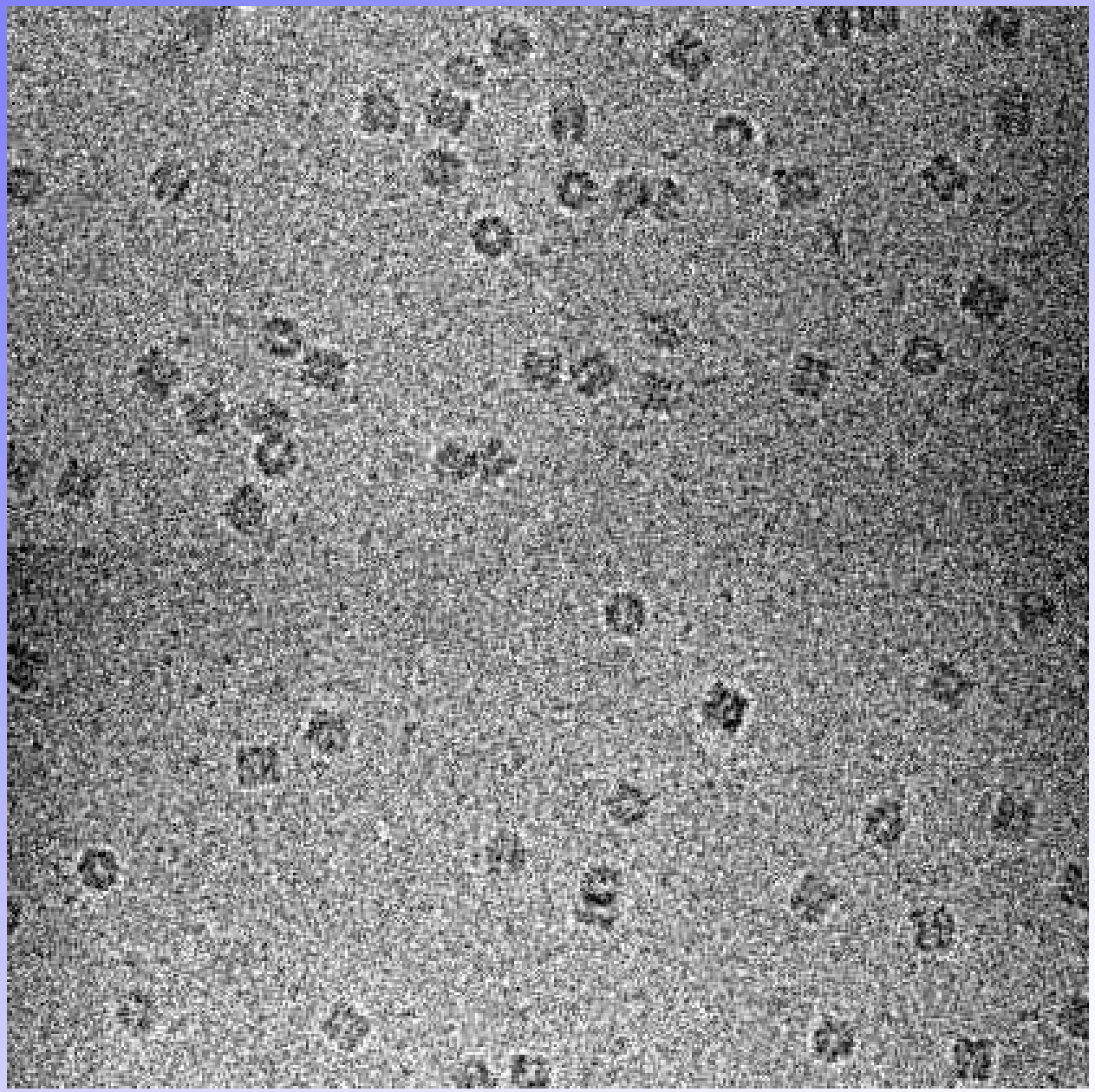

FREEZE
3mm
liquid
ethane
–170°C
liquid
N_2

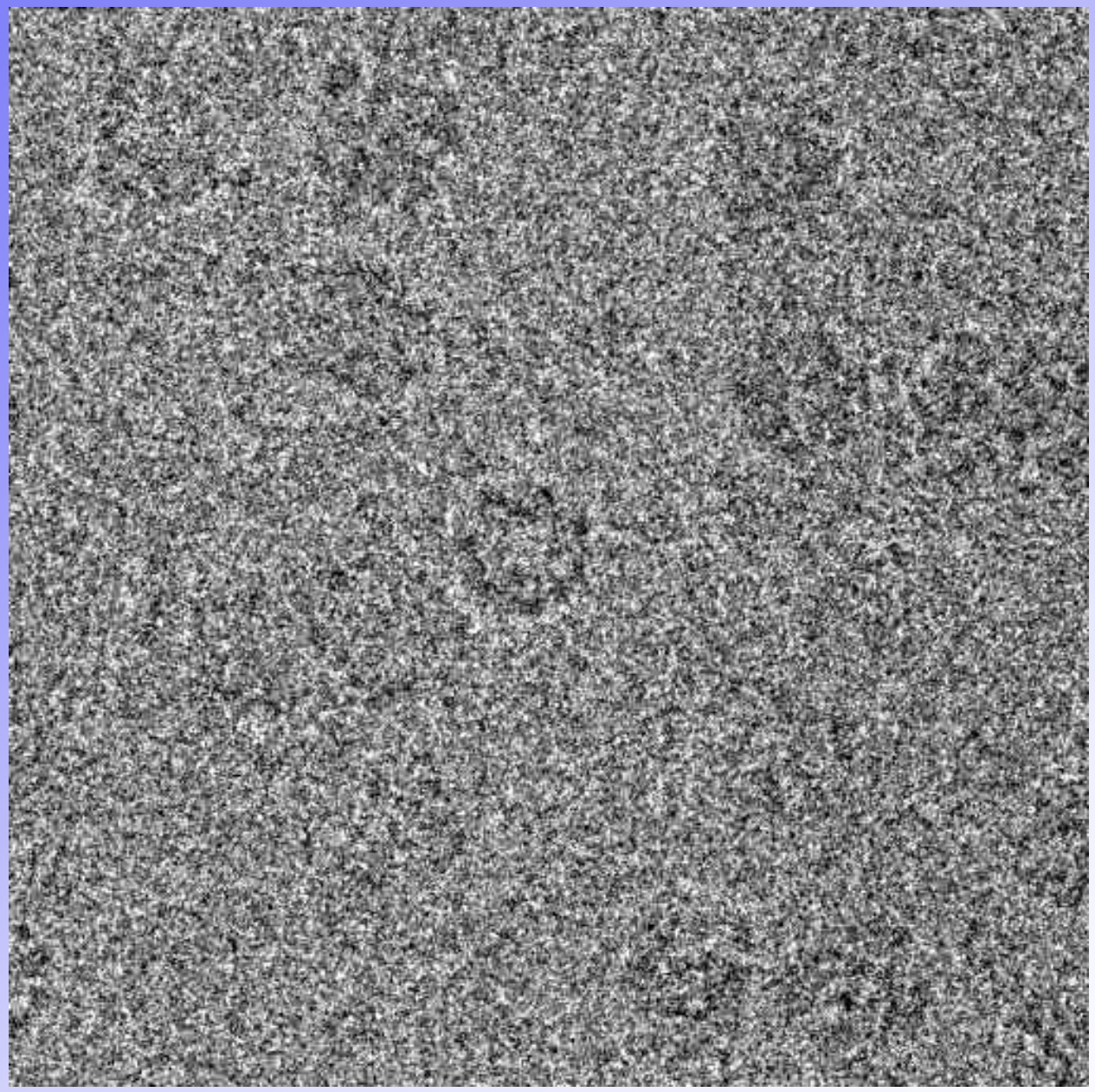

65 μm


20 µm


~1 μm


1000 Å


1000 Å


200 Å

Canonical SPA in EMAN

- Image Acquisition
- Particle Picking
- 2-D Analysis
- Symmetry/Low Resolution Model
- Determine CTF Parameters
- High Resolution Refinement
- Post-processing
- Dynamics Analysis

Canonical SPA in EMAN

- Image Acquisition
- Particle Picking
- 2-D Analysis
- Symmetry/Low Resolution Model
- Determine CTF Parameters
- High Resolution Refinement
- Post-processing
- Dynamics Analysis

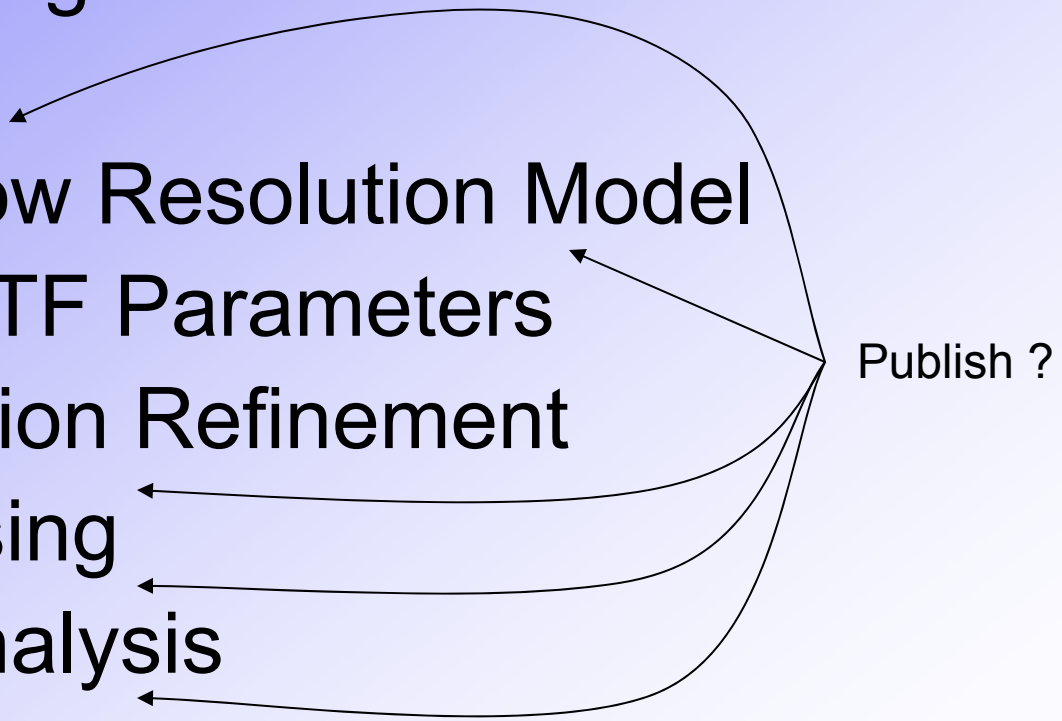


Image Aquisition

- Pick your defocus range
  - Envelope function related to defocus even on modern FEG scopes
  - As close to focus as possible while still able to locate the particles → +1 μm, random values (for EMAN)
  - Focal Pairs ?
- CCD
  - Adequate sampling. For low symmetry, 2/5 Nyquist is a good rule of thumb (res=5*A/pix)
- Film
  - Scanner is important. Bit depth less important than scanner envelope function (e2scannereval.py)

Image Aquisition

- Avoid continuous carbon substrate
  - Decreases contrast
  - Difficult to separate c-film CTF from specimen
  - Makes CTF correction difficult

Particle Picking

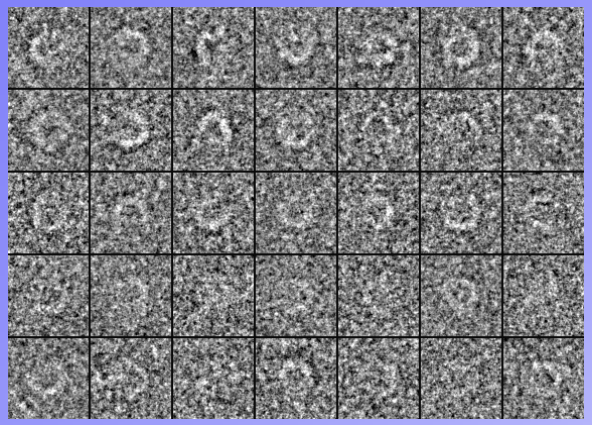
- boxer, batchboxer, helixboxer, e2boxer.py
- Outside software ?
- manual or semi-automated process
- False positives are very dangerous, but also beware of excluding views you weren't expecting
- Mixing microscopes possible, but rarely worthwhile (tomorrow)

2-D Analysis

- Even if you know the quaternary structure, still worthwhile. May be surprises.
- At least 1000-2000 particles with uniform orientation distribution, perhaps fewer if symmetry or preferred orientation
- Look for dynamics or degradation
- ‘shrink’ particles for speed
- # classes ≤ # particles/20
- refine2d.py **not** startnrclasses

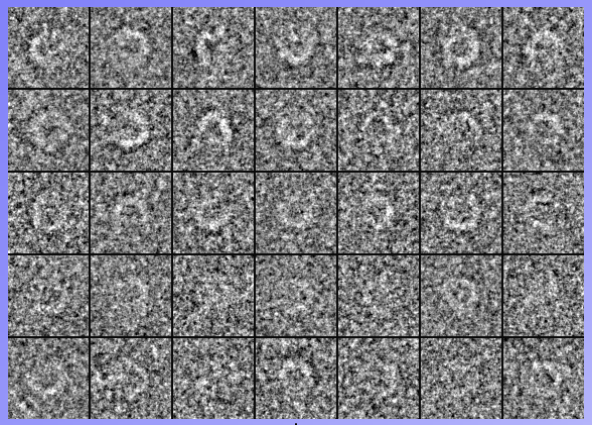
refine2d.py

Particles

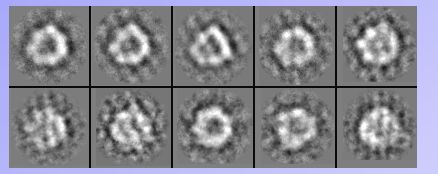


refine2d.py

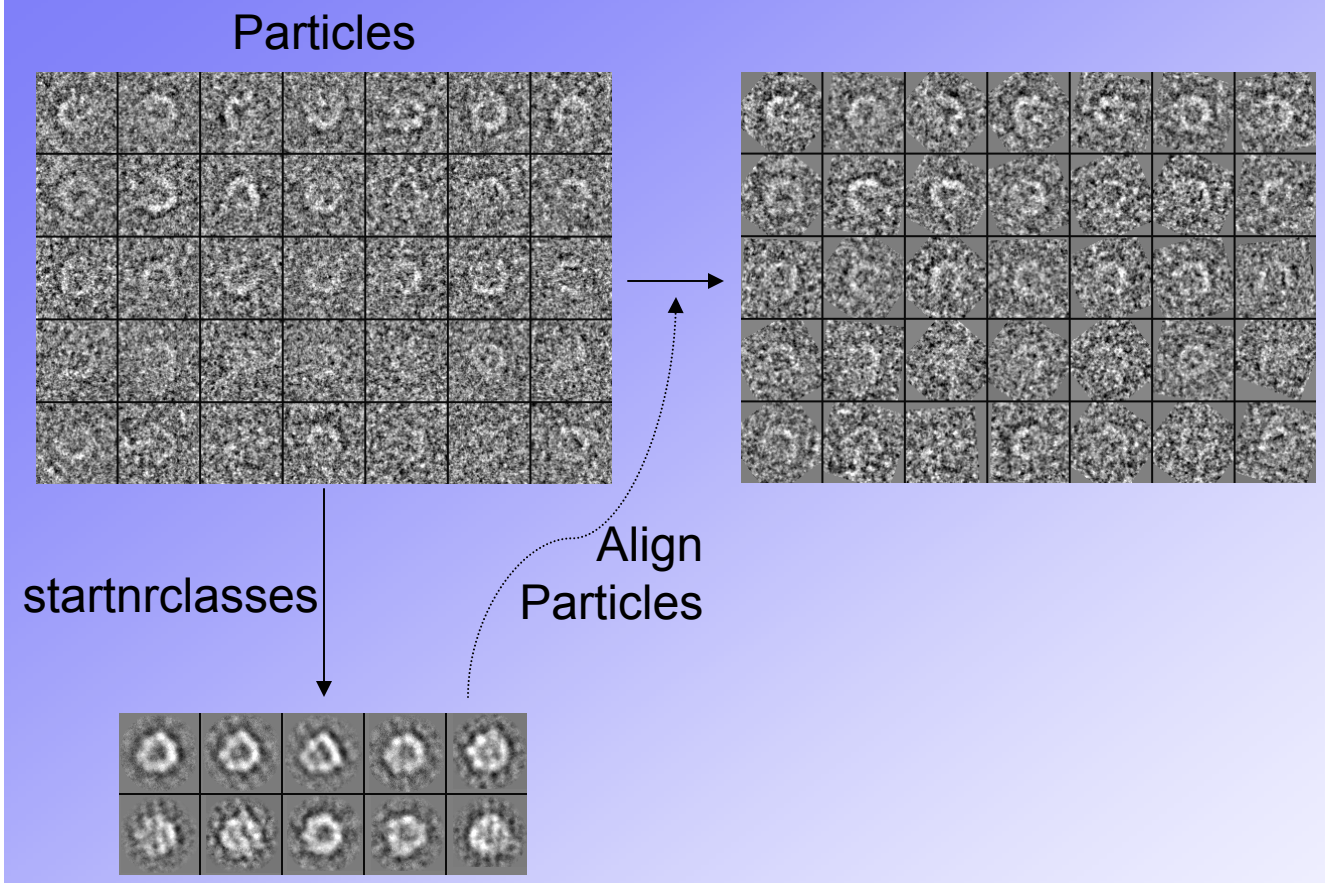
Particles



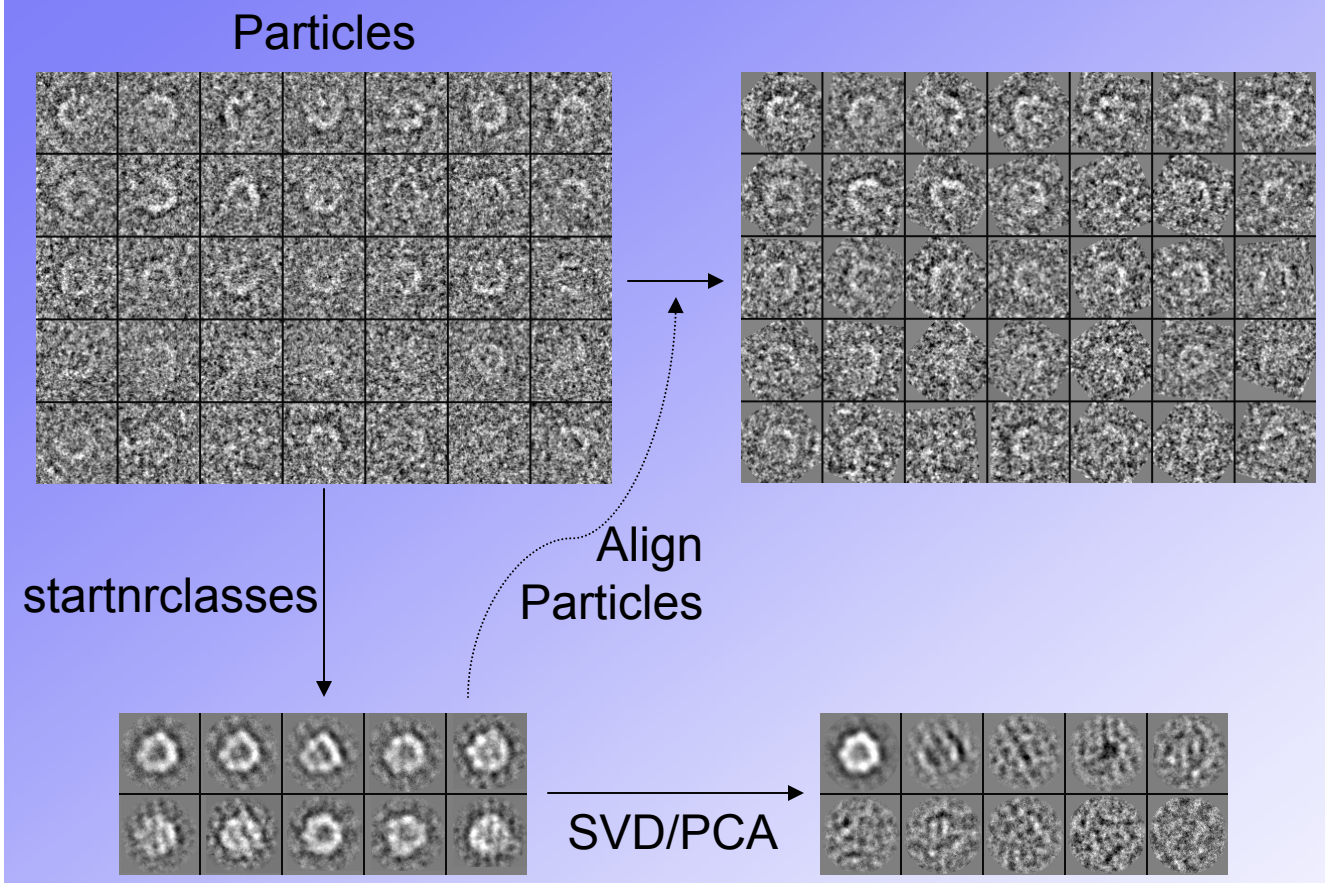
startnrclasses



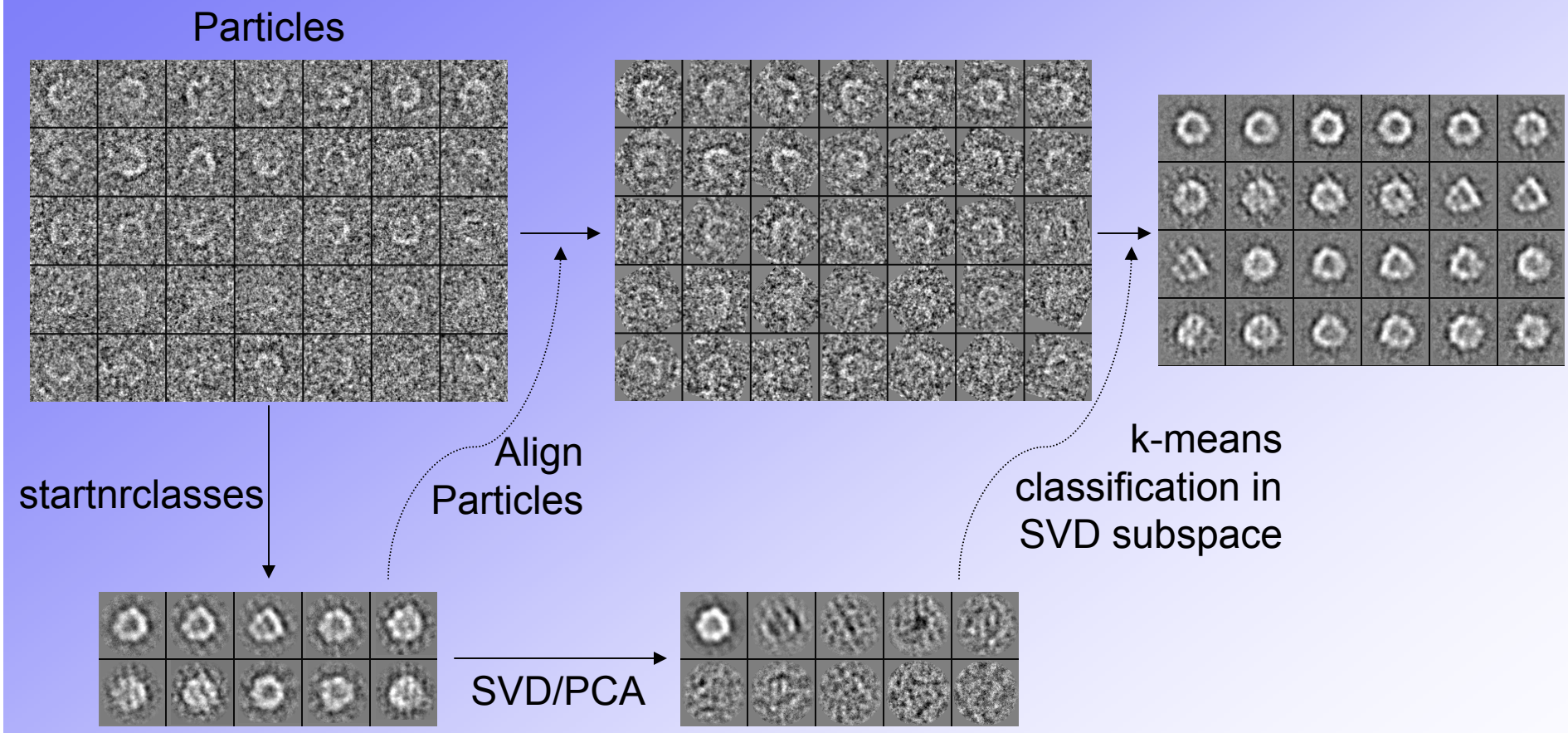
refine2d.py



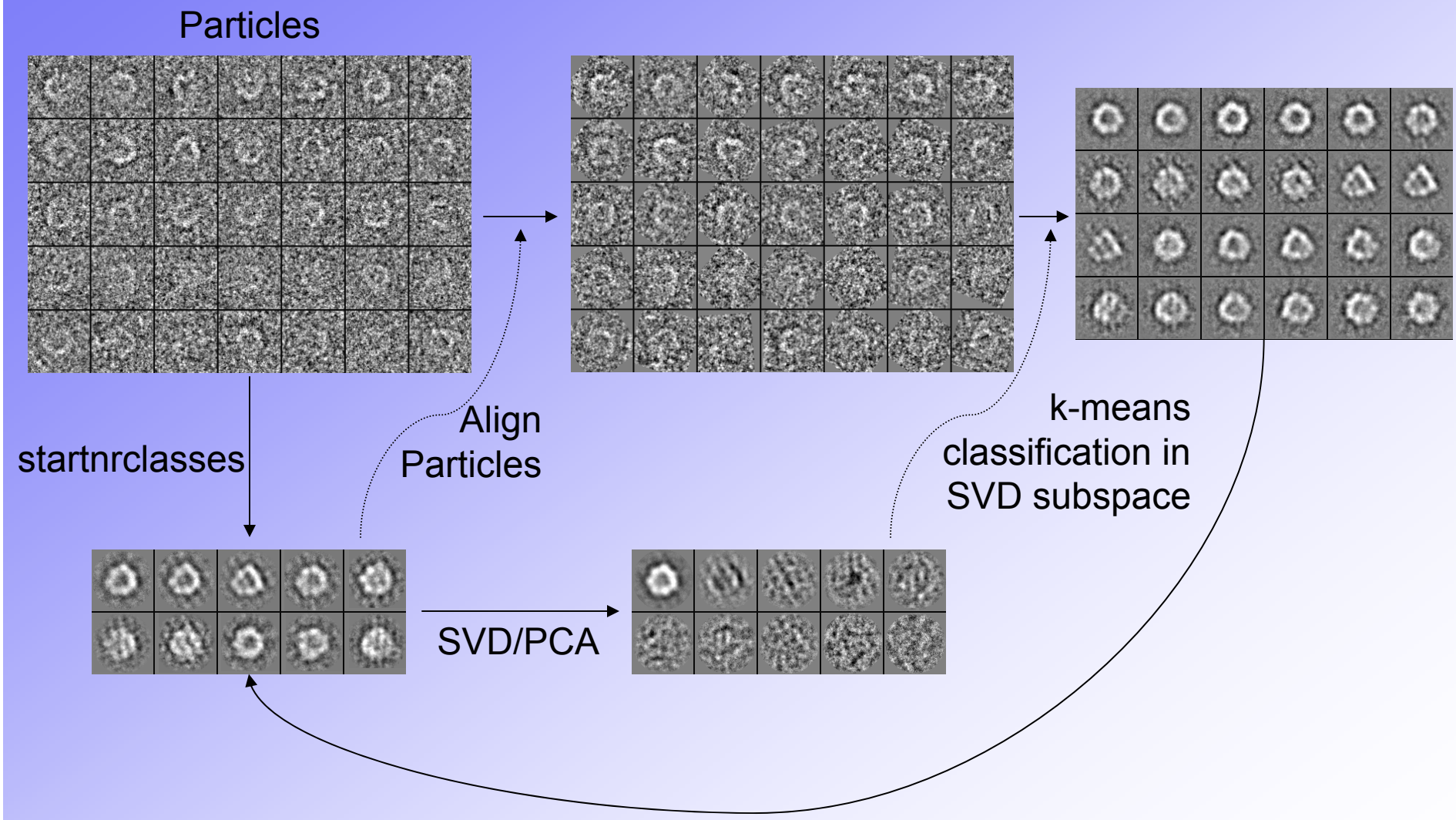
refine2d.py



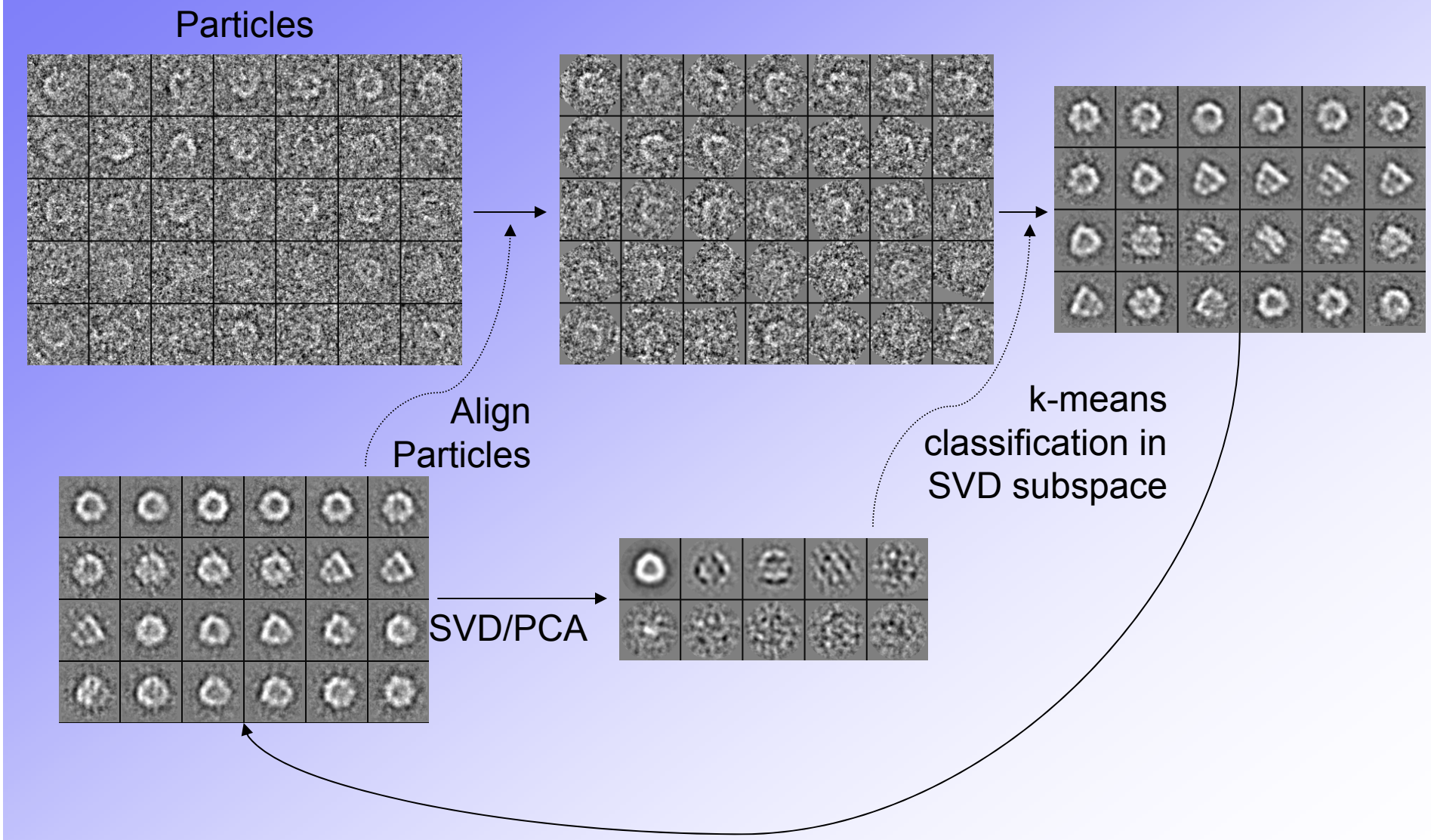
refine2d.py


Particles
startnrclasses
Align
Particles
SVD/PCA
k-means
classification in
SVD subspace

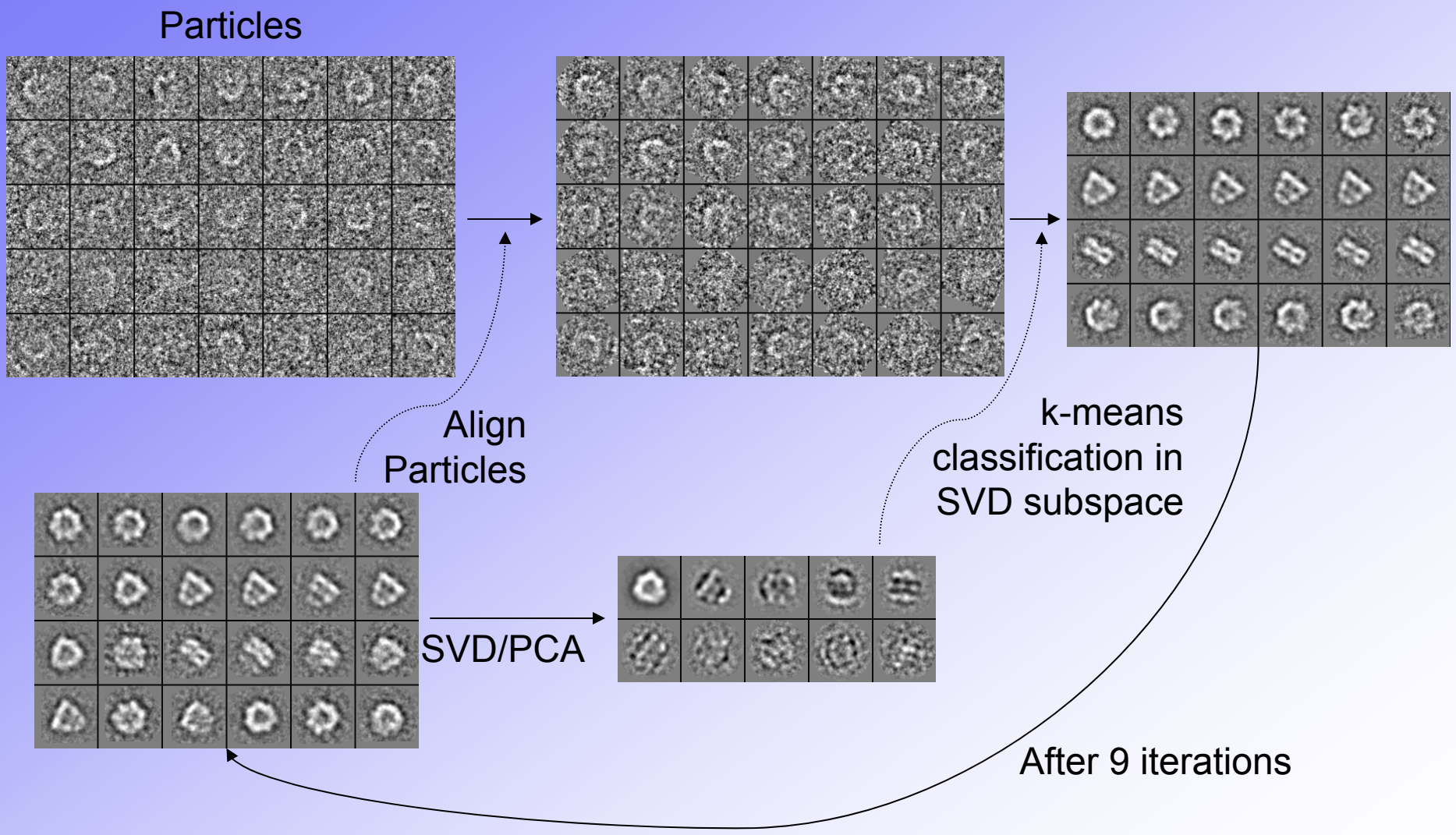
refine2d.py


Particles
startnrclasses
Align
Particles
SVD/PCA
k-means
classification in
SVD subspace

refine2d.py


Particles
Align
Particles
SVD/PCA
k-means
classification in
SVD subspace

refine2d.py


Particles
Align
Particles
SVD/PCA
k-means
classification in
SVD subspace
After 9 iterations

Symmetry / Initial Model

- May be obvious from 2-D refinement
- Double-check with quick 3-D refinements from class-averages → also gives starting model
- If still ambiguous, may need better data or tomography
- Starting models need not be very good. General shape/size is sufficient.
- Compare projections and class-averages. Must agree !

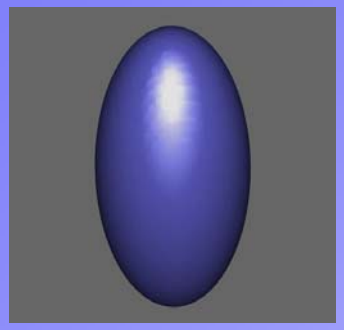
CTF Determination

(current EMAN correction scheme)

- (tomorrow!)
- Focus on high resolution more than low resolution. Significant mismatches can be tolerated at low resolution.
- B-factors and amp-contrast should be fairly constant, unless the data clearly justifies a difference.
- Fewer particles → noisier power spectra → overestimating B-factor
- Use ctfcw=, not ctfc=

High Resolution Refinement

- Most efficient to move stepwise, not jump straight from 20 Å → 5 Å.
- Beware of over-refinement / model bias
- classiter=0 and setsf= dangerous, especially together
- Resolution ≠ Resolvability !
- Masking ~ solvent flattening, helps a lot



Initial 3D Model

Uniform Projections

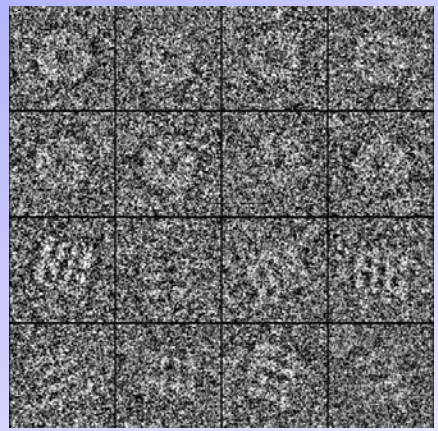
Build New 3-D Model

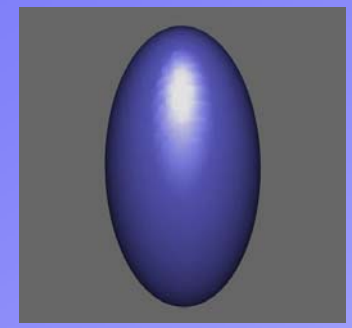
Final 3D Model

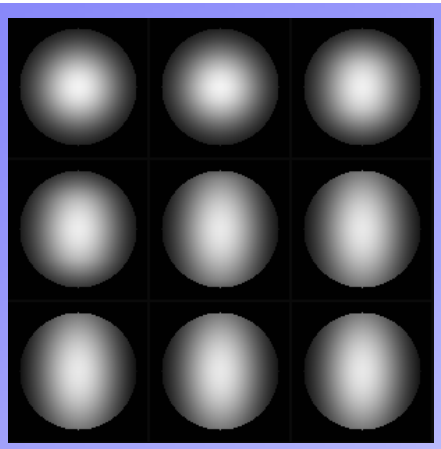
Particle Images

Classify Particles

Align and Average Classes







Initial 3D Model

Uniform Projections

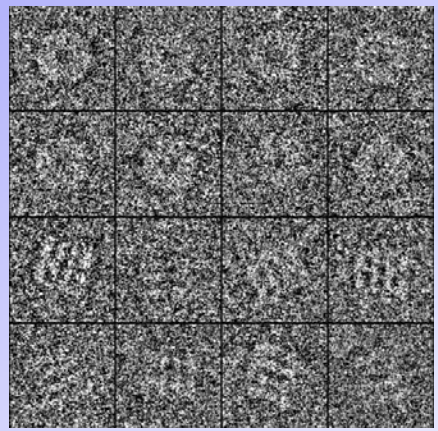
Build New 3-D Model

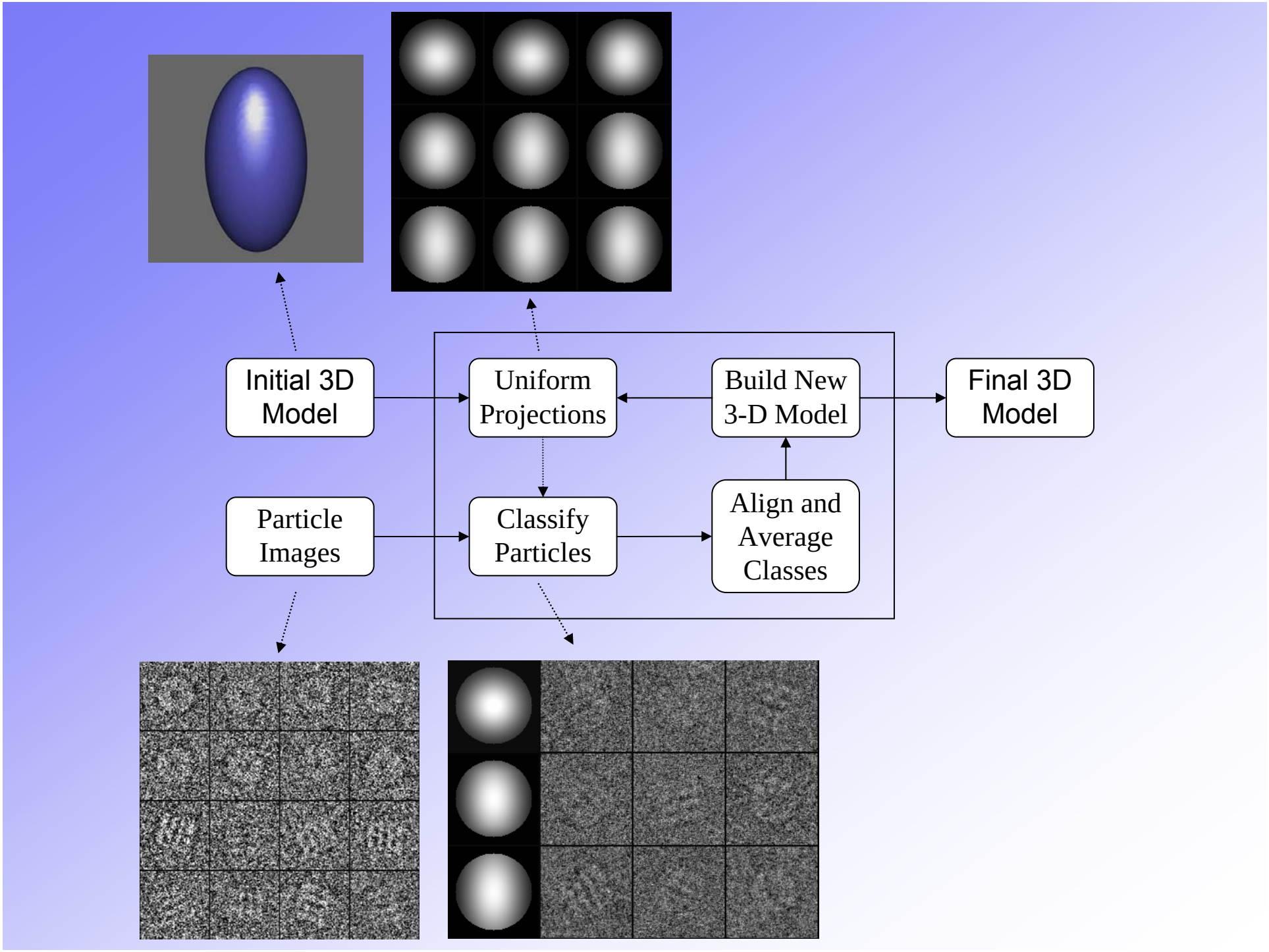
Final 3D Model

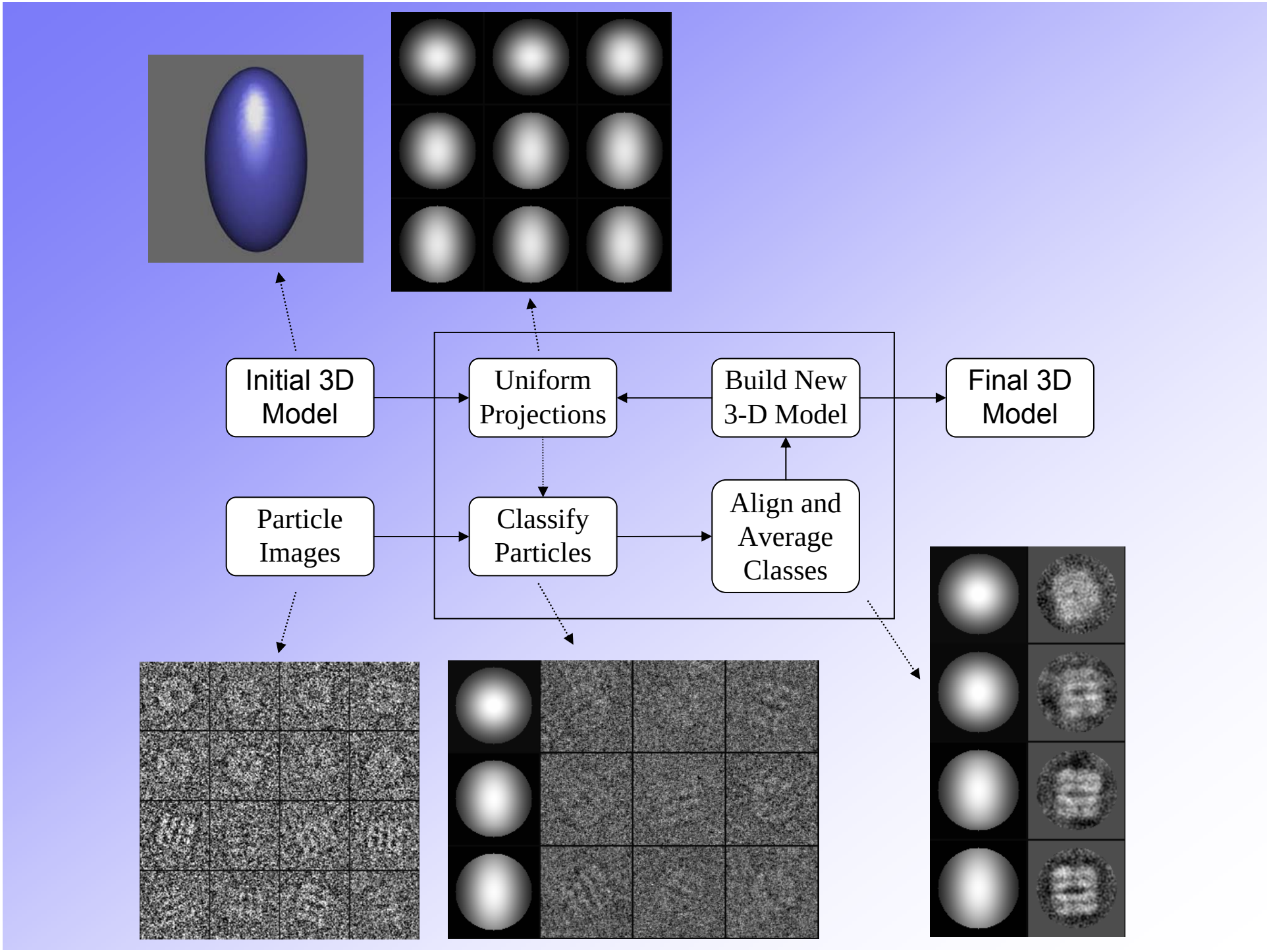
Particle Images

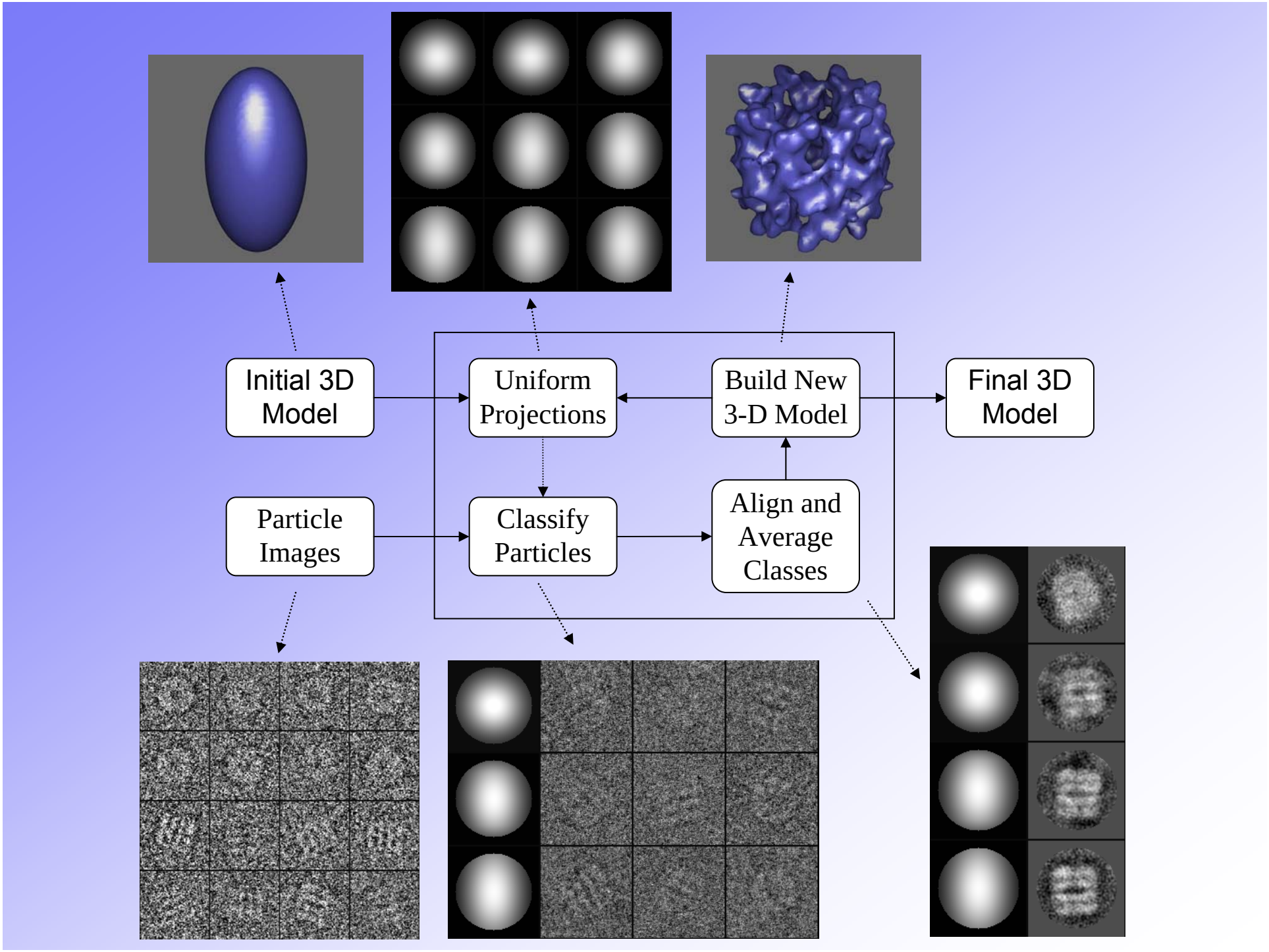
Classify Particles

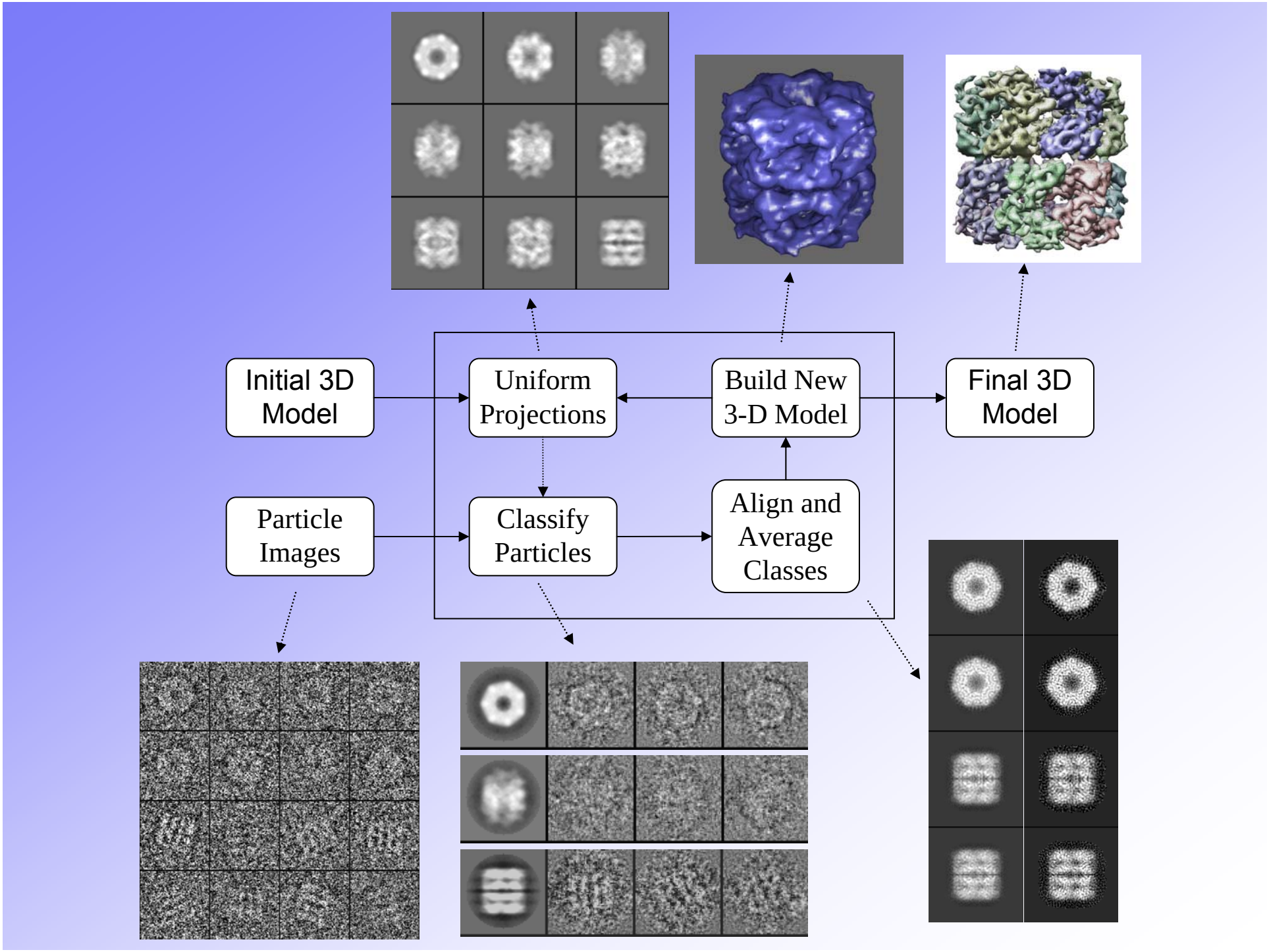
Align and Average Classes



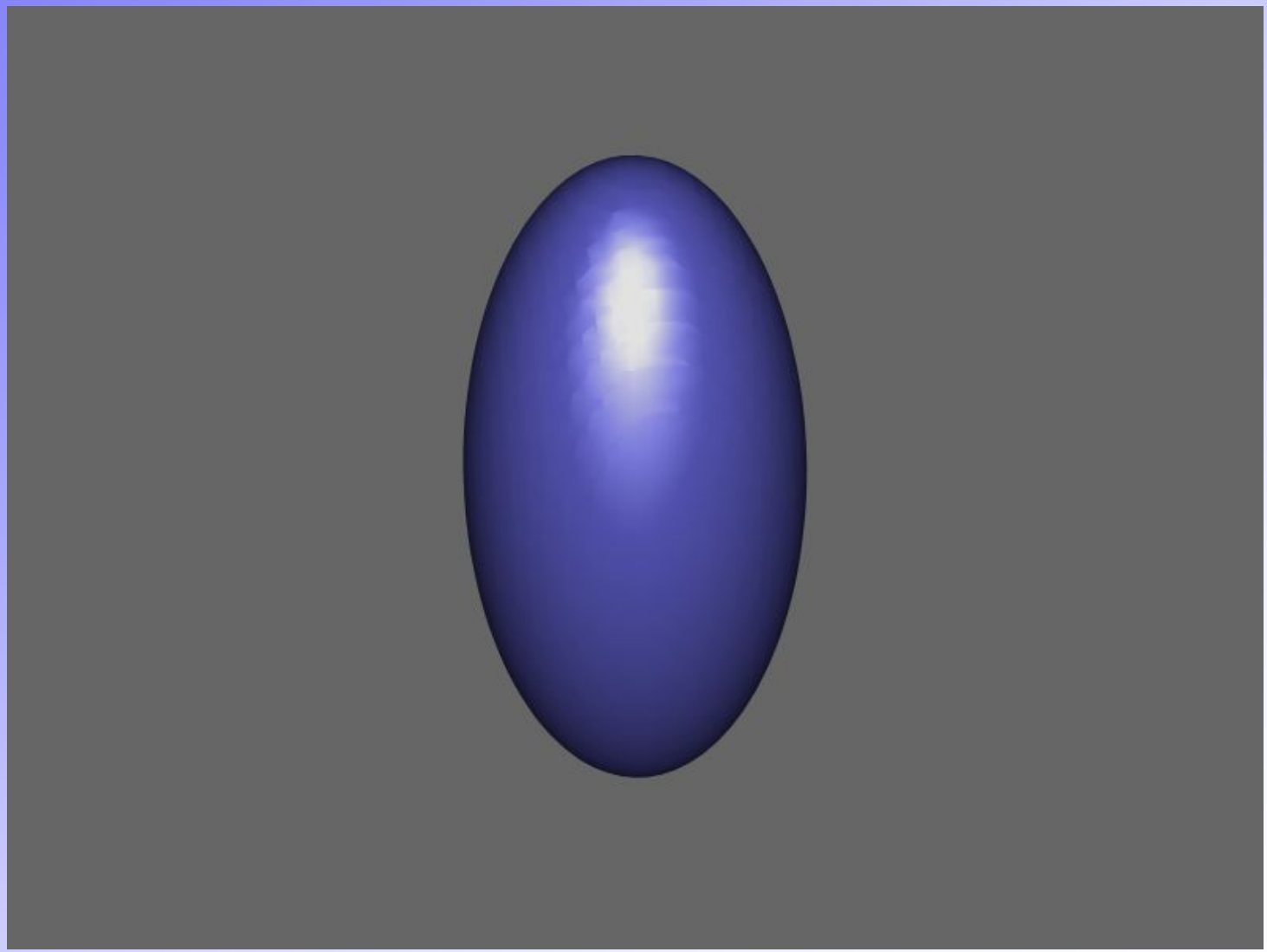

Initial 3D Model
Uniform Projections
Build New 3-D Model
Final 3D Model
Particle Images
Classify Particles
Align and Average Classes


Initial 3D Model
Uniform Projections
Build New 3-D Model
Final 3D Model
Particle Images
Classify Particles
Align and Average Classes

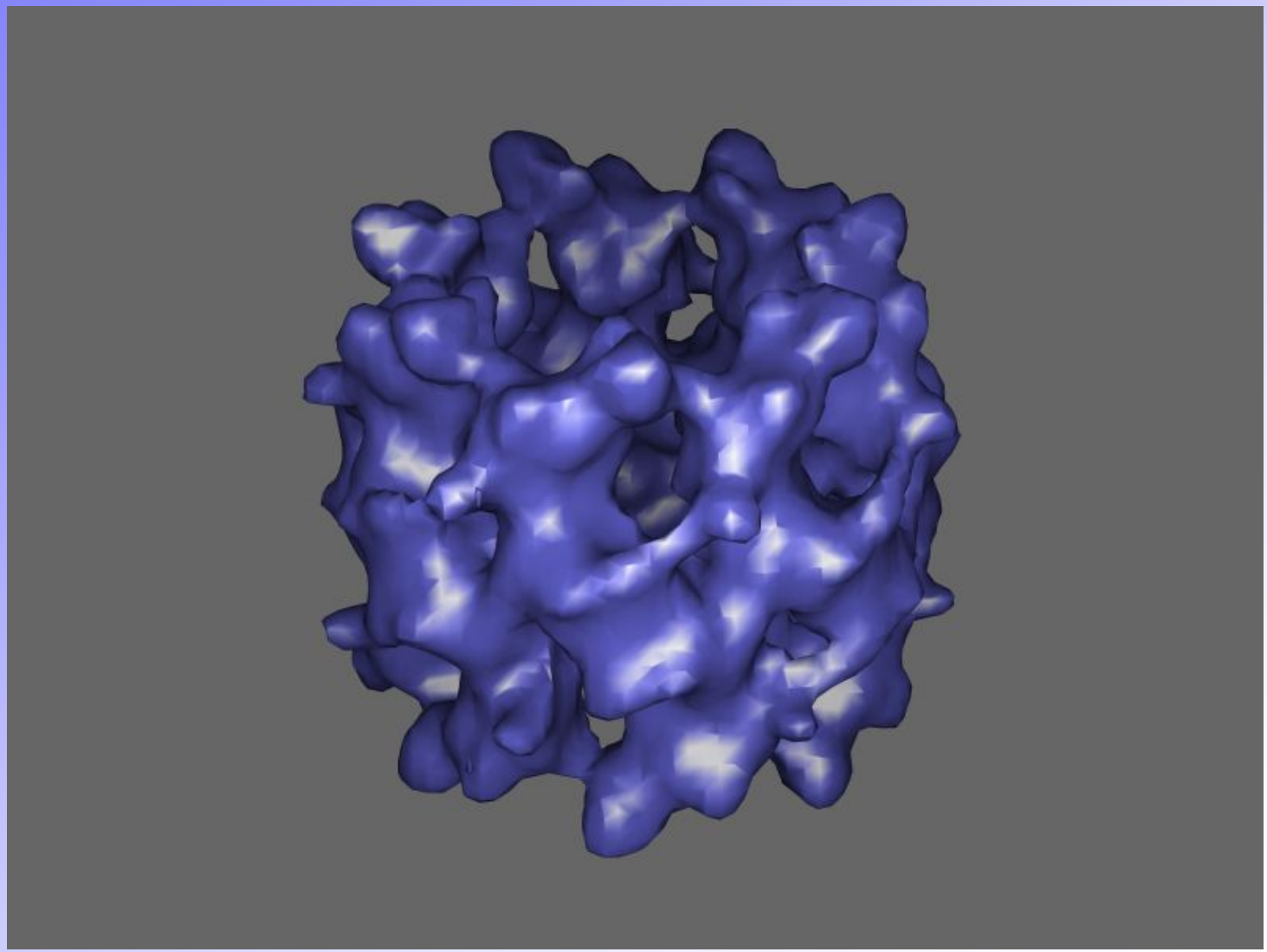

Initial 3D Model
Uniform Projections
Build New 3-D Model
Final 3D Model
Particle Images
Classify Particles
Align and Average Classes


Initial 3D Model
Uniform Projections
Build New 3-D Model
Final 3D Model
Particle Images
Classify Particles
Align and Average Classes

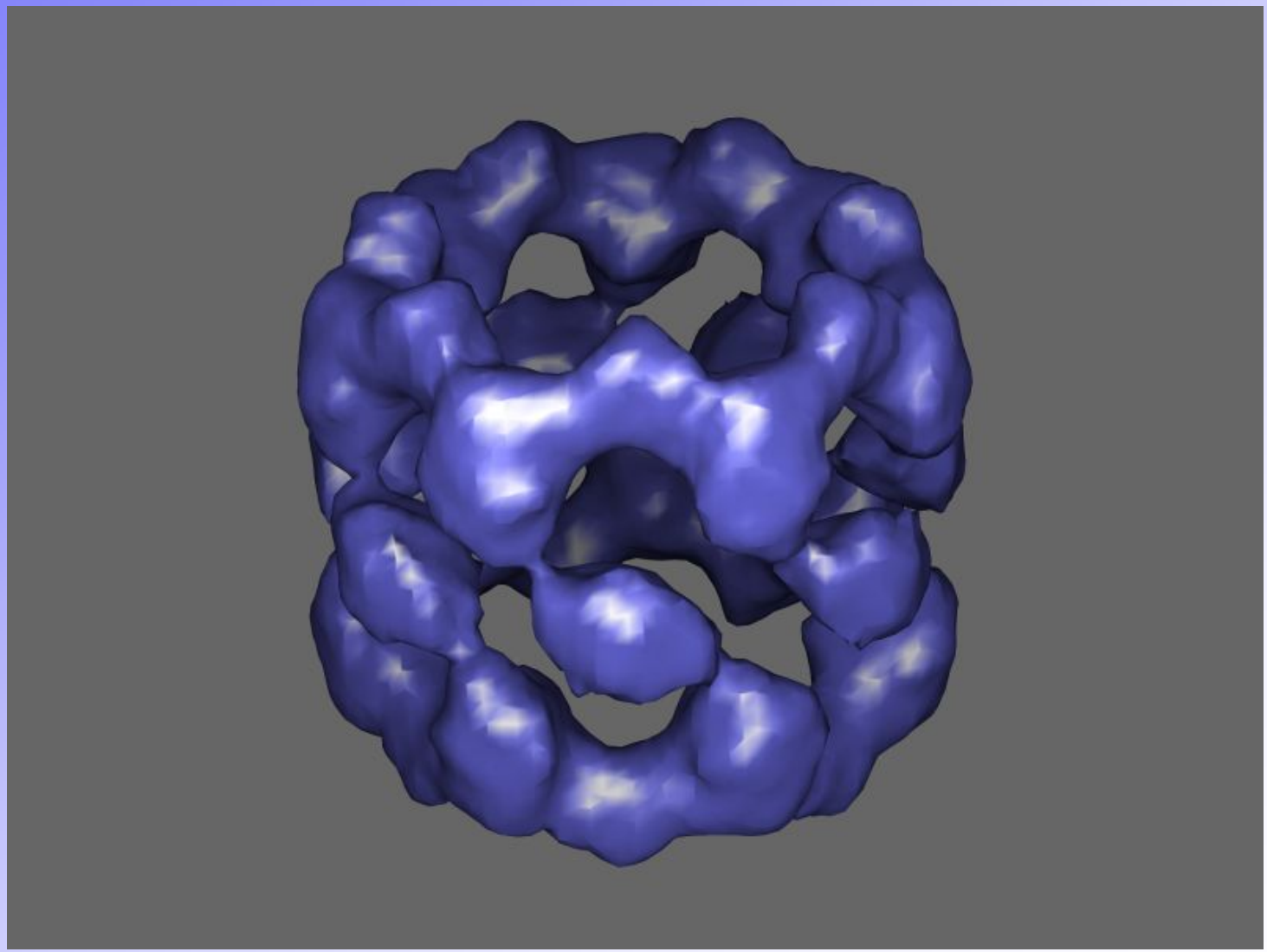
Refine from Gaussian Ellipsoid



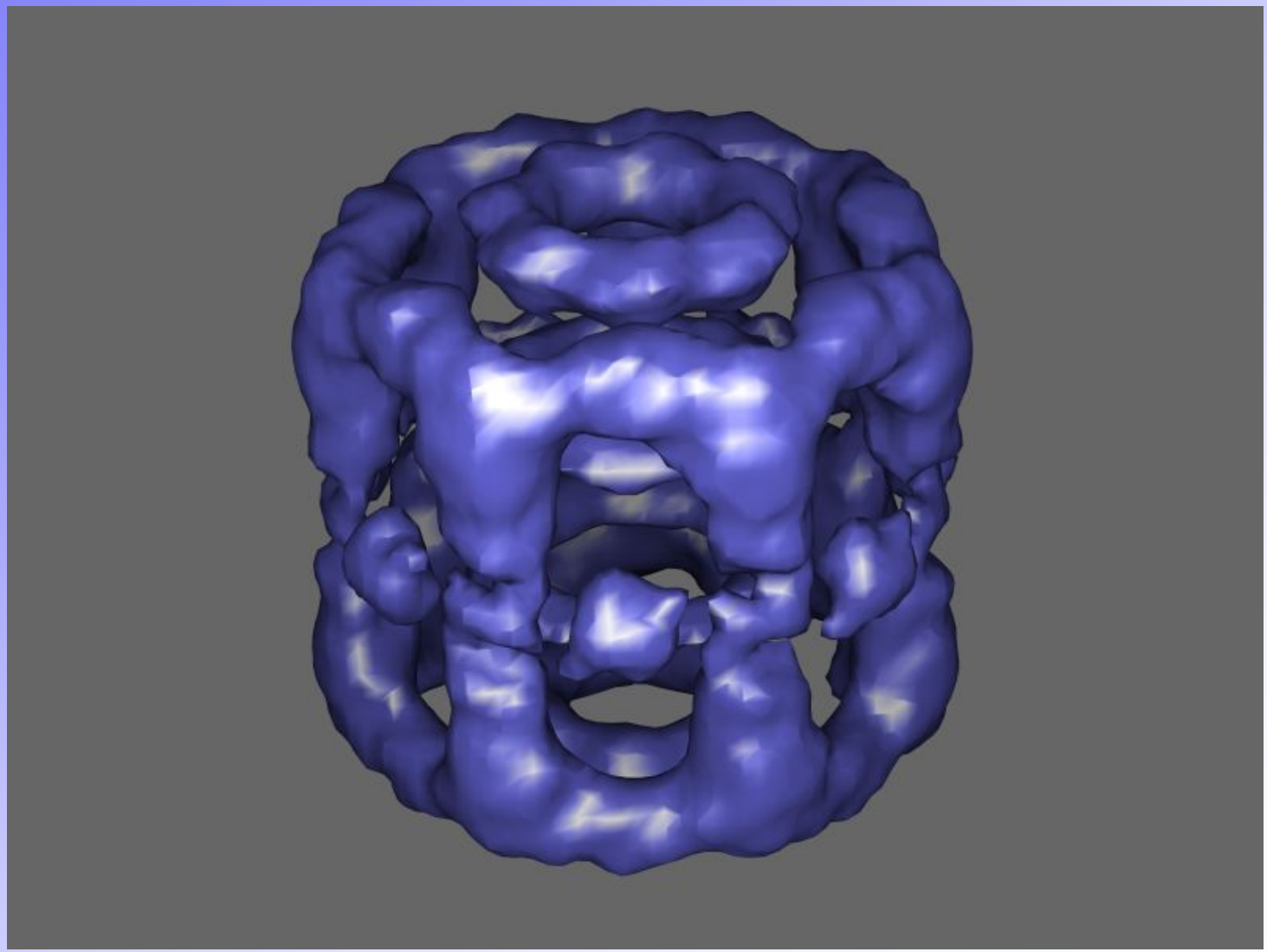
Iteration 1



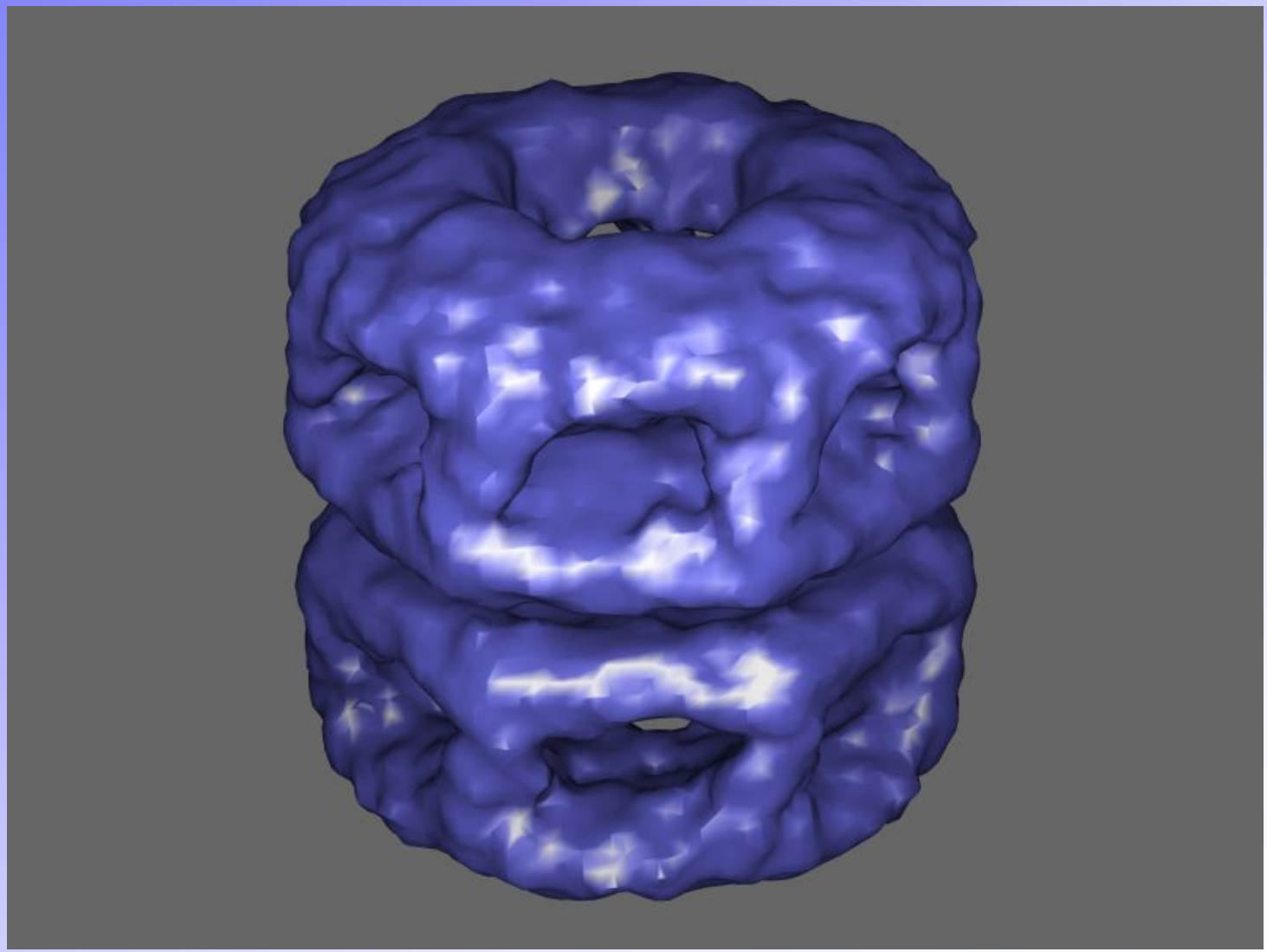
Iteration 2



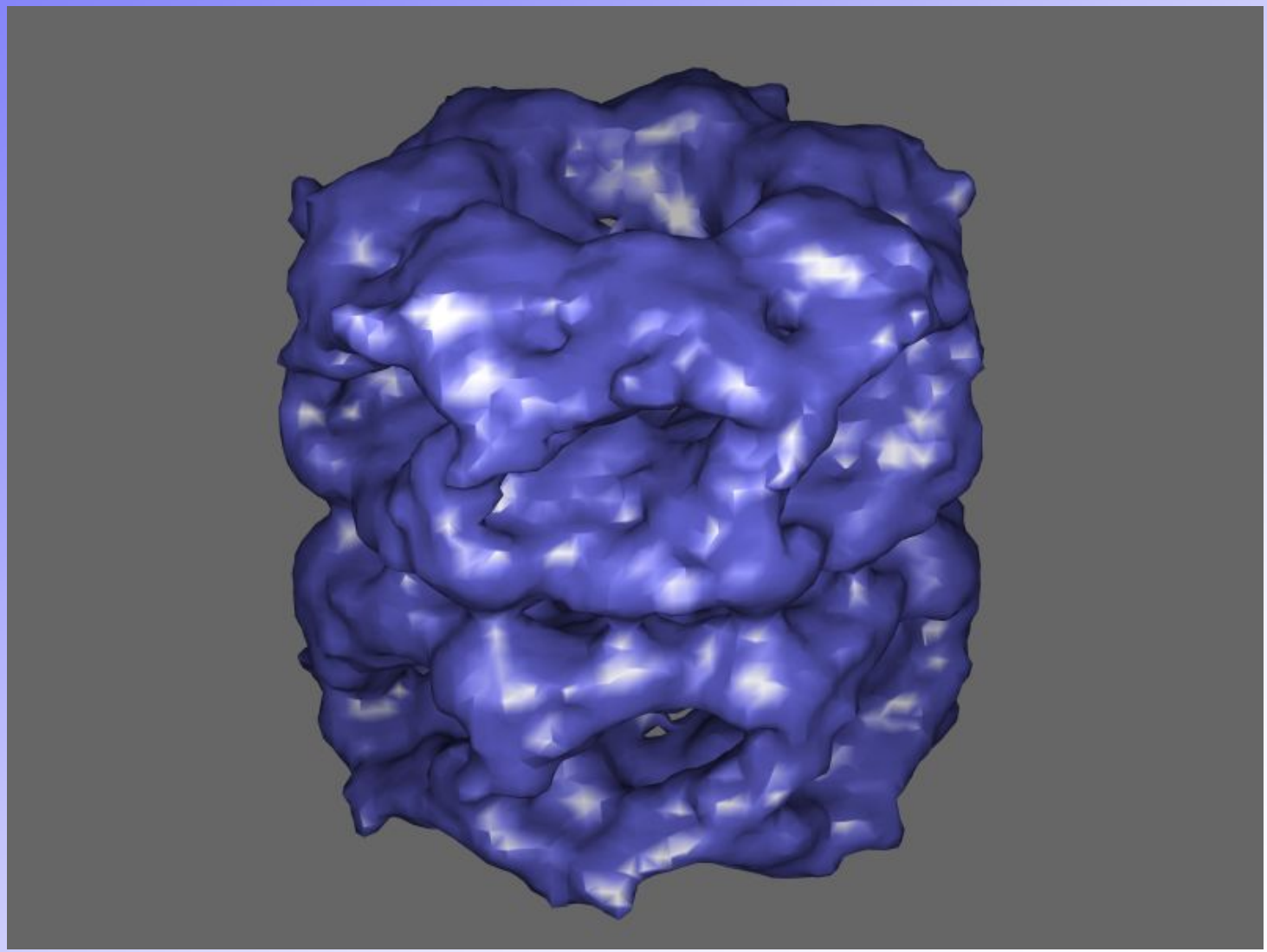
Iteration 3



Iteration 4

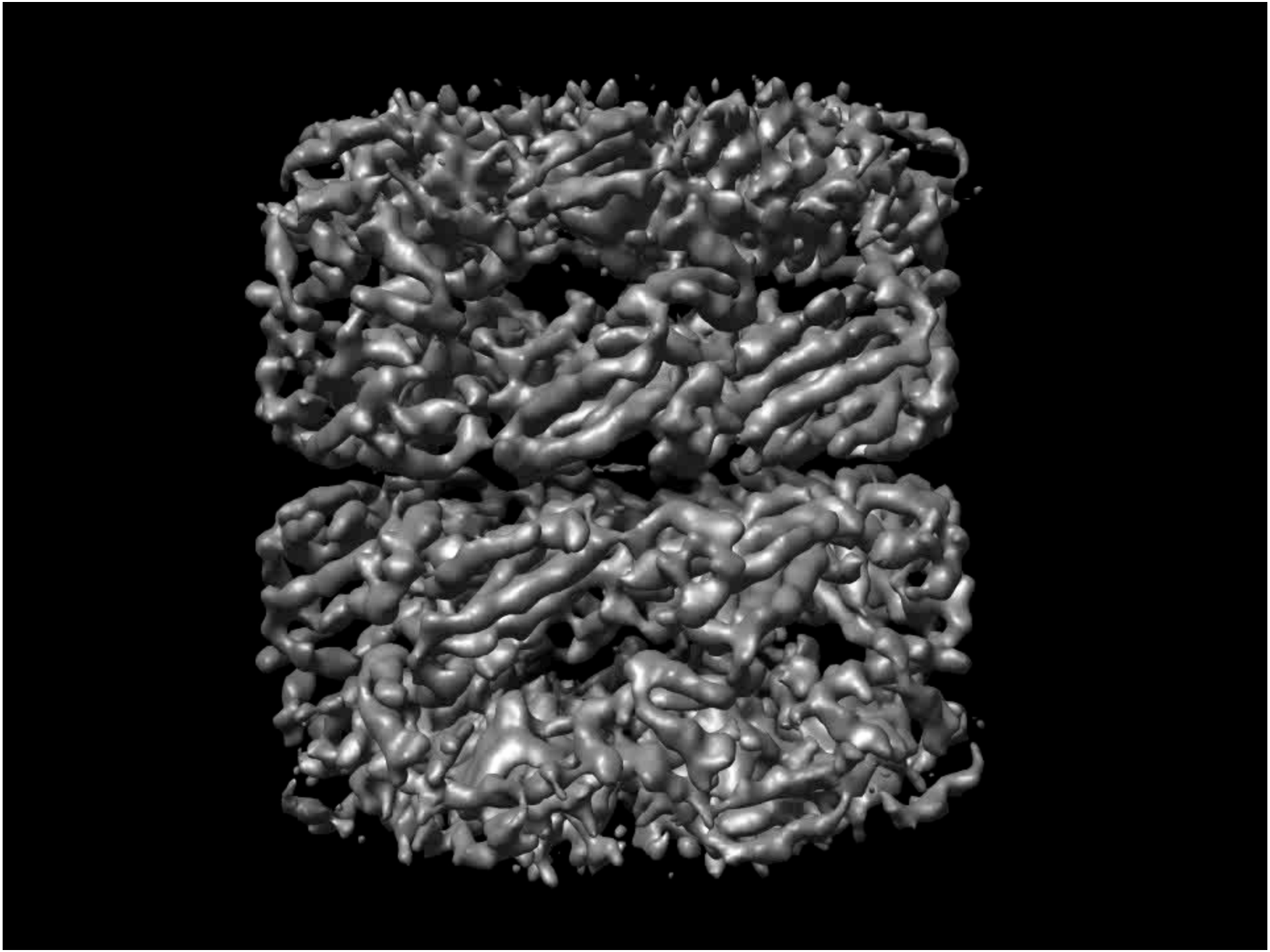


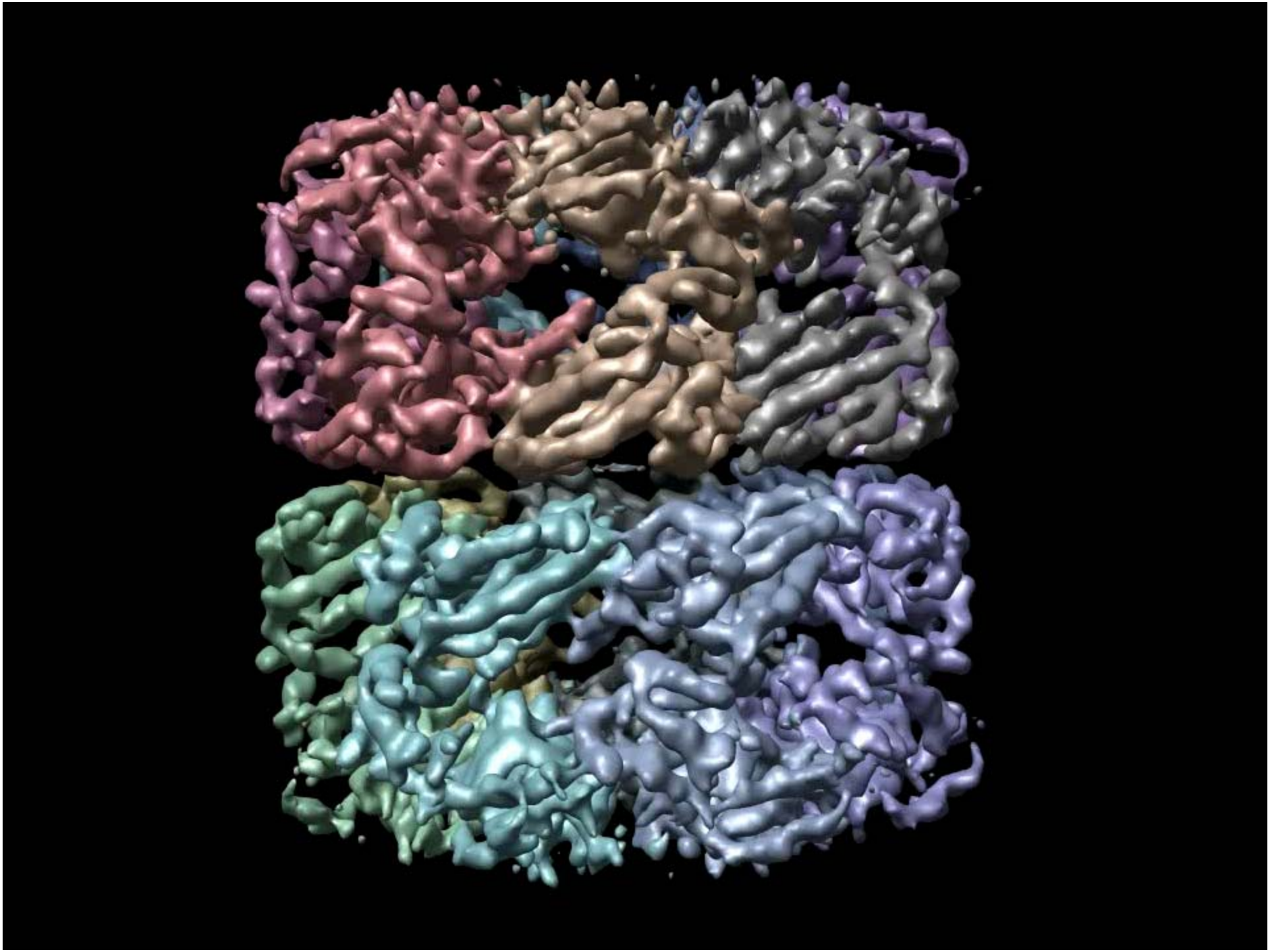
Iteration 5

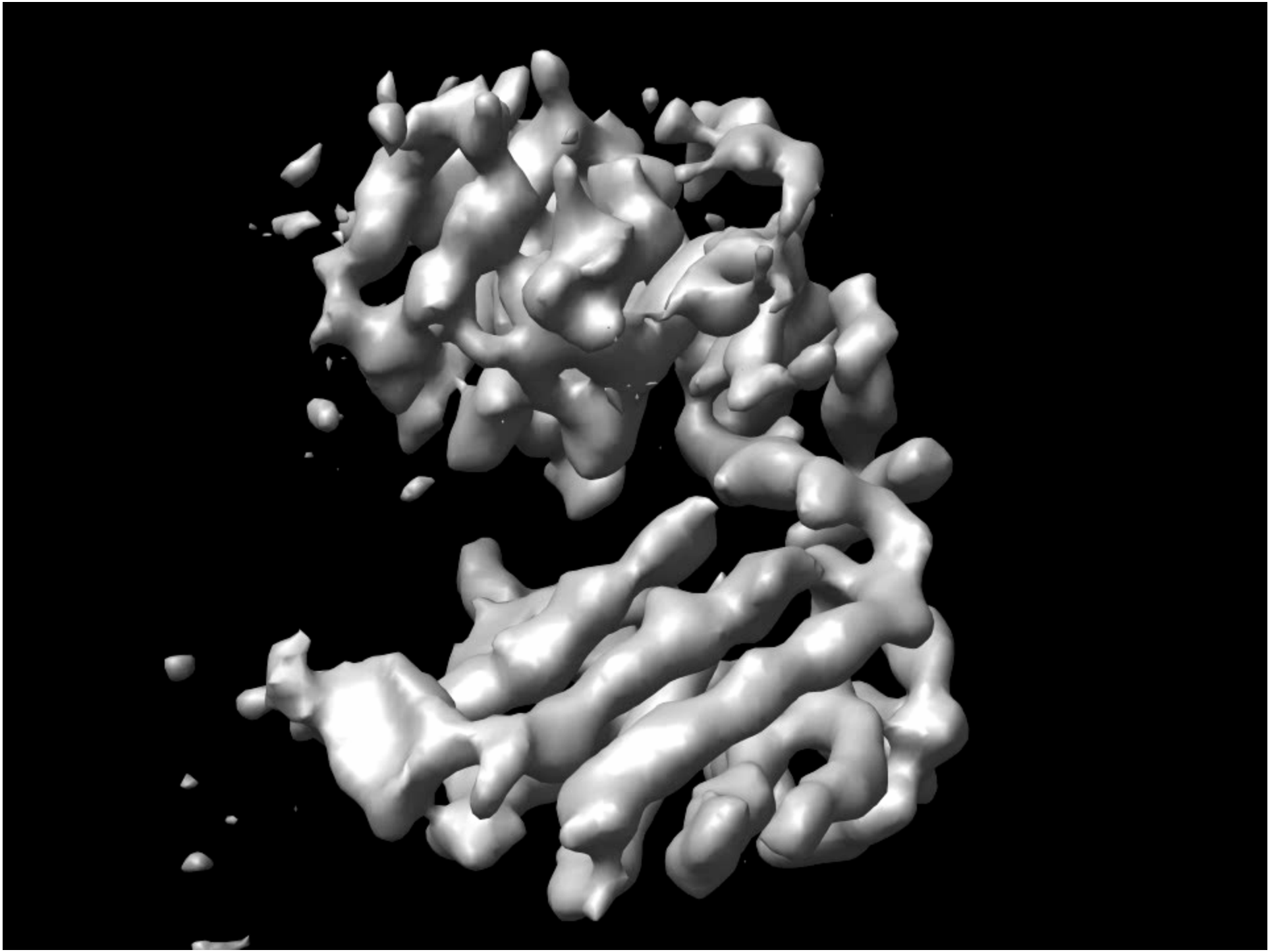


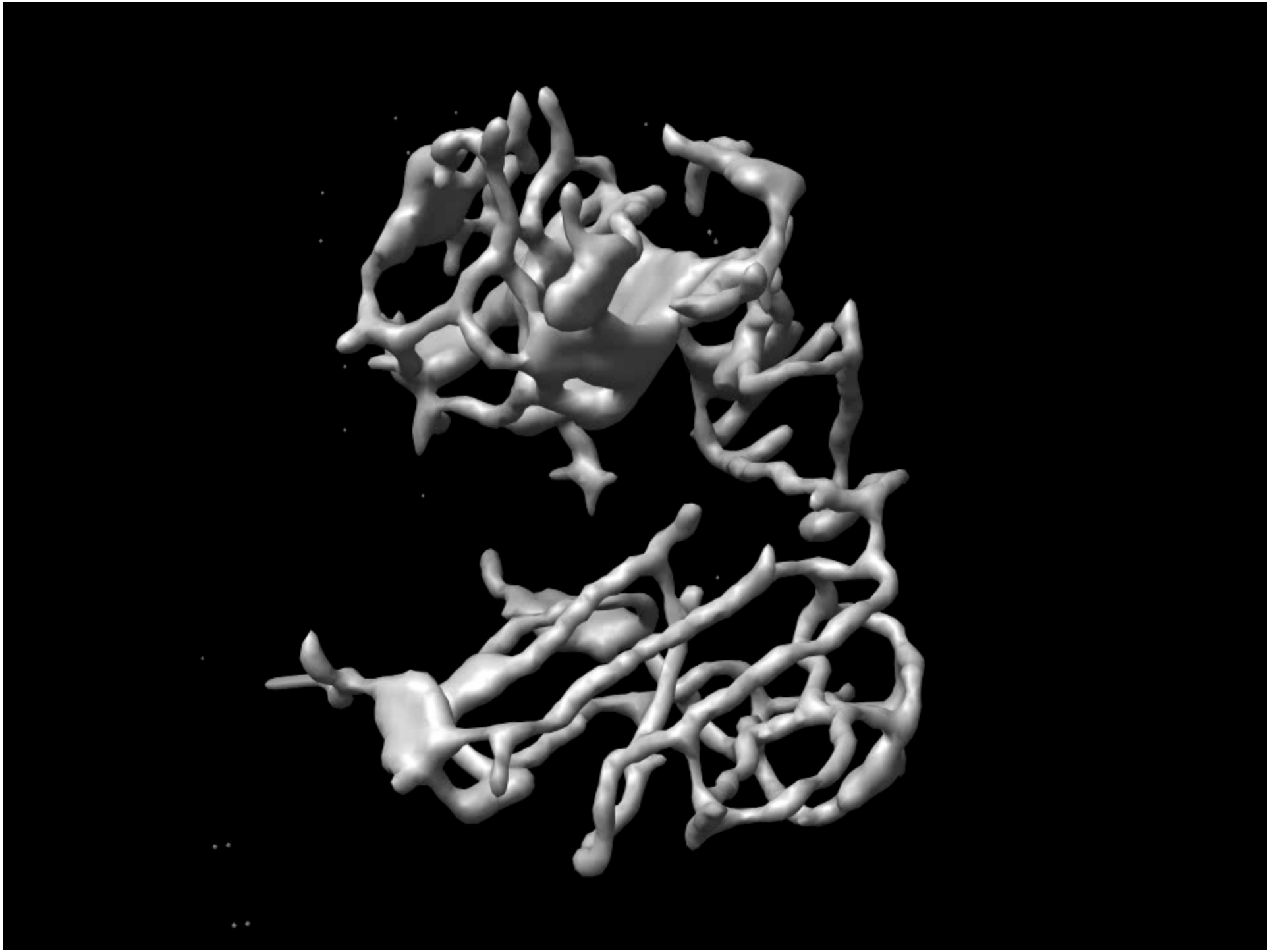
Postprocessing

- 10 – 20 Å
  - Xtal structures / homology modeling of components (foldhunter)
- 5 – 10 Å
  - Secondary structure analysis (ssehunter)
- < 5 Å
  - Backbone tracing, atomistic models (multiple tools)





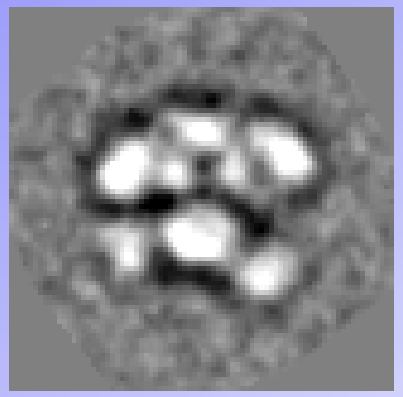


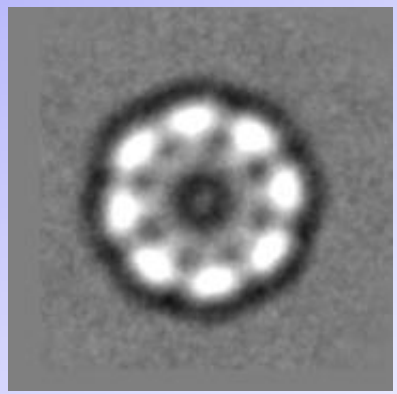


Dynamics

- Look at refine2d.py results in the context of your high resolution model
- ‘bootstrapping’ method to identify flexibility/heterogeneity
- Use multirefine to generate multiple structures from 1 heterogeneous data set

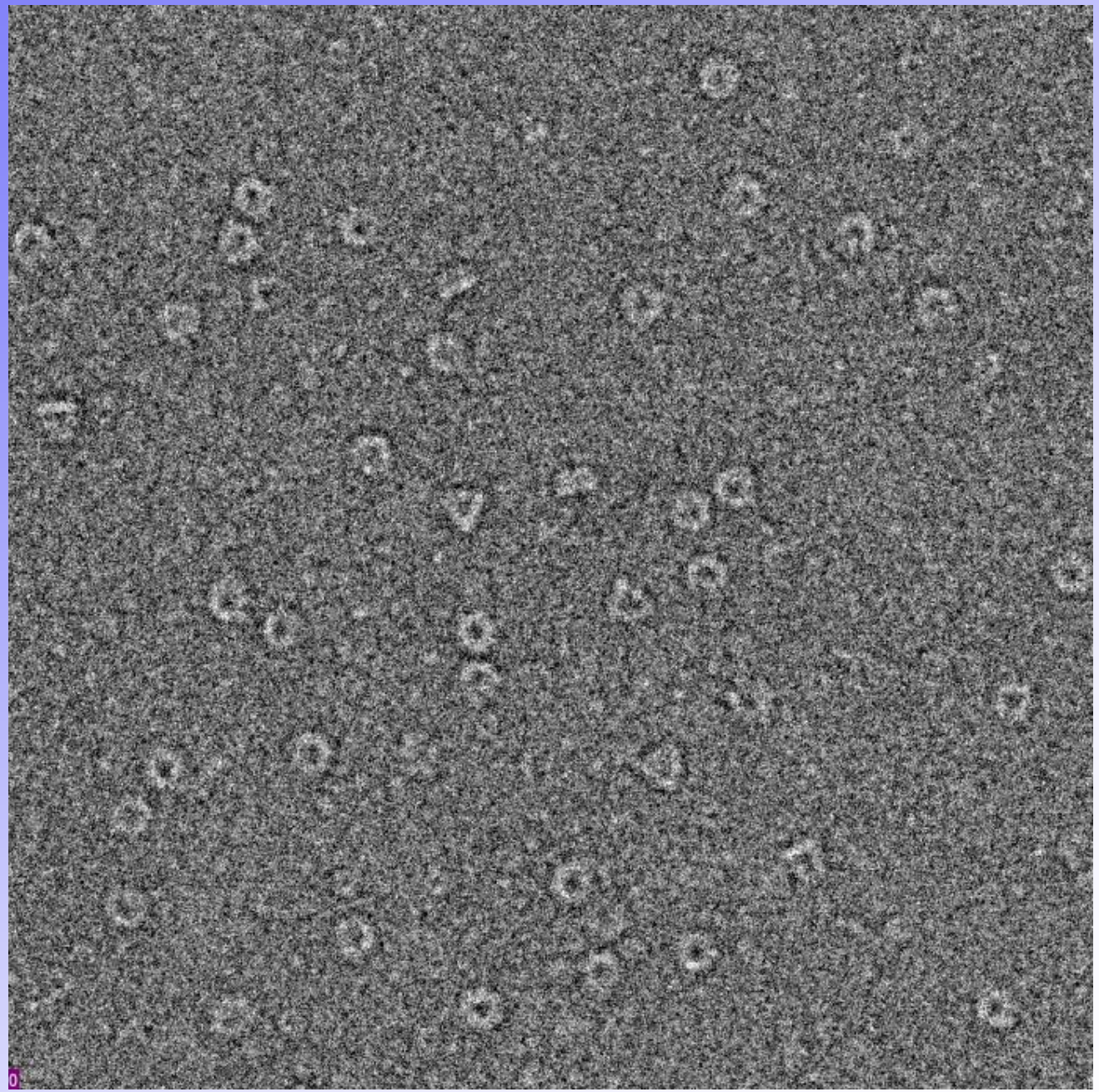
Dynamics
(refine2d.py)

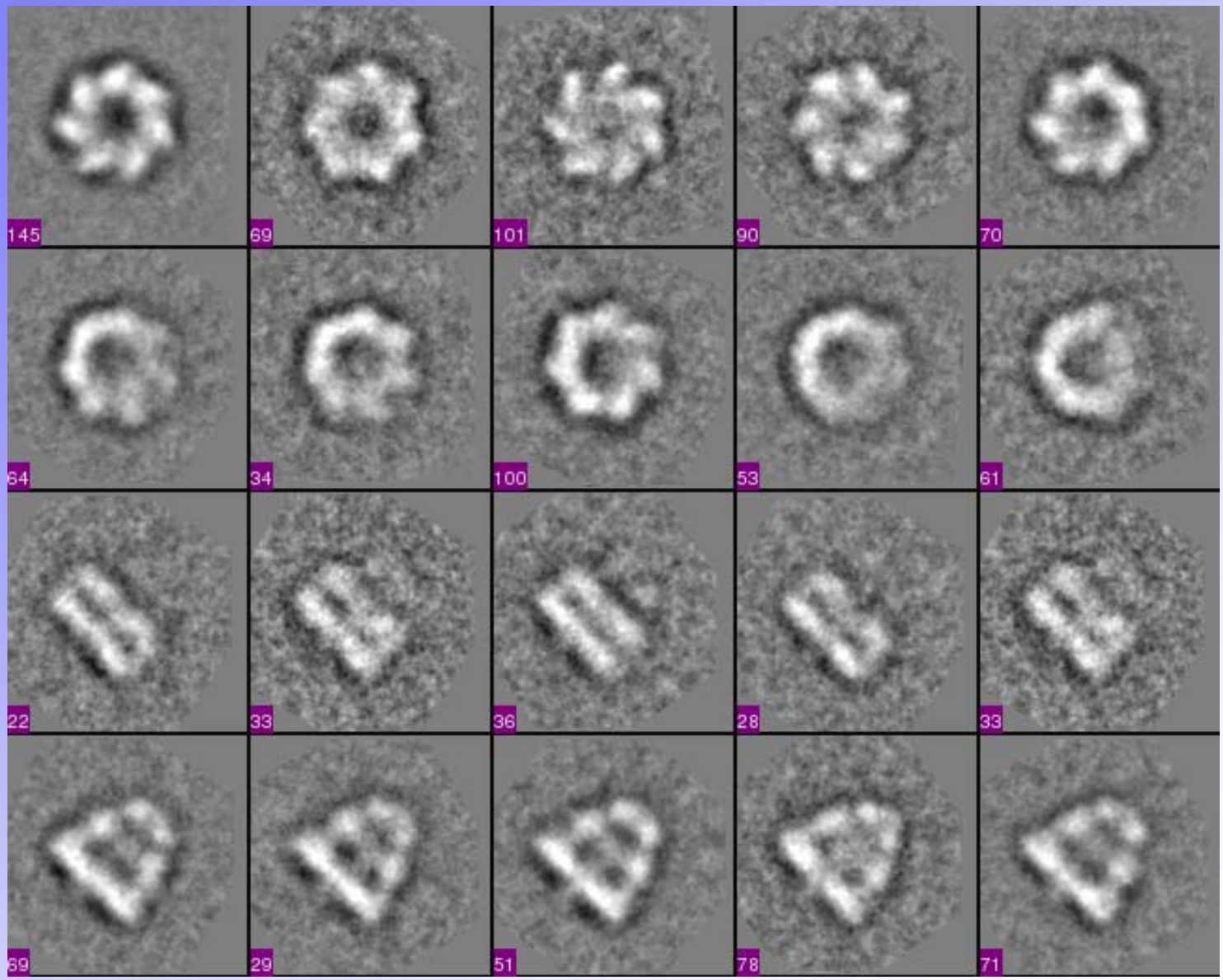




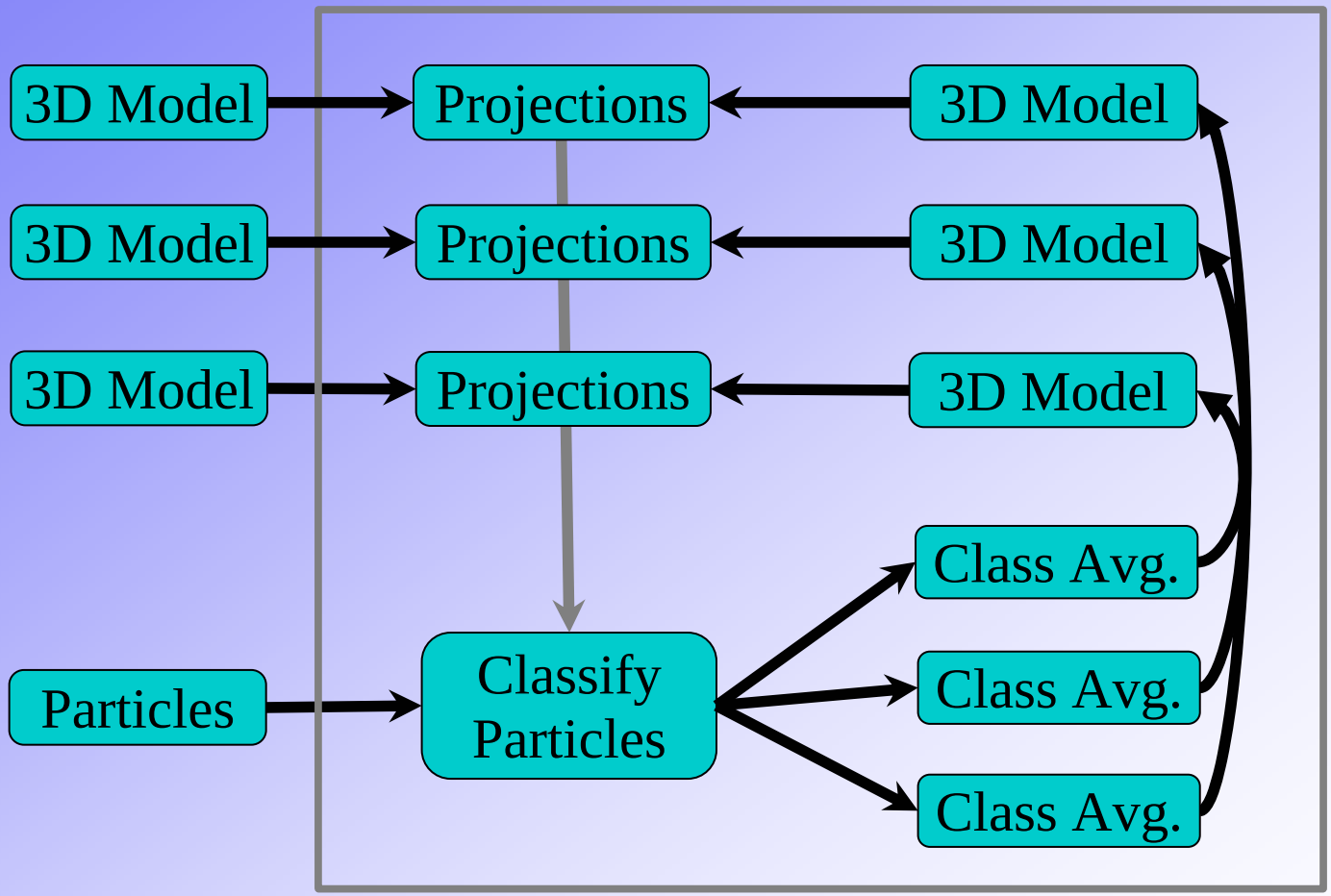
- Refine2d.py with a small number of classes
- Refine2d.py again on particles from a single view
- e2stacksort.py
- e2stackanim.py

SR398+GroES+ATP

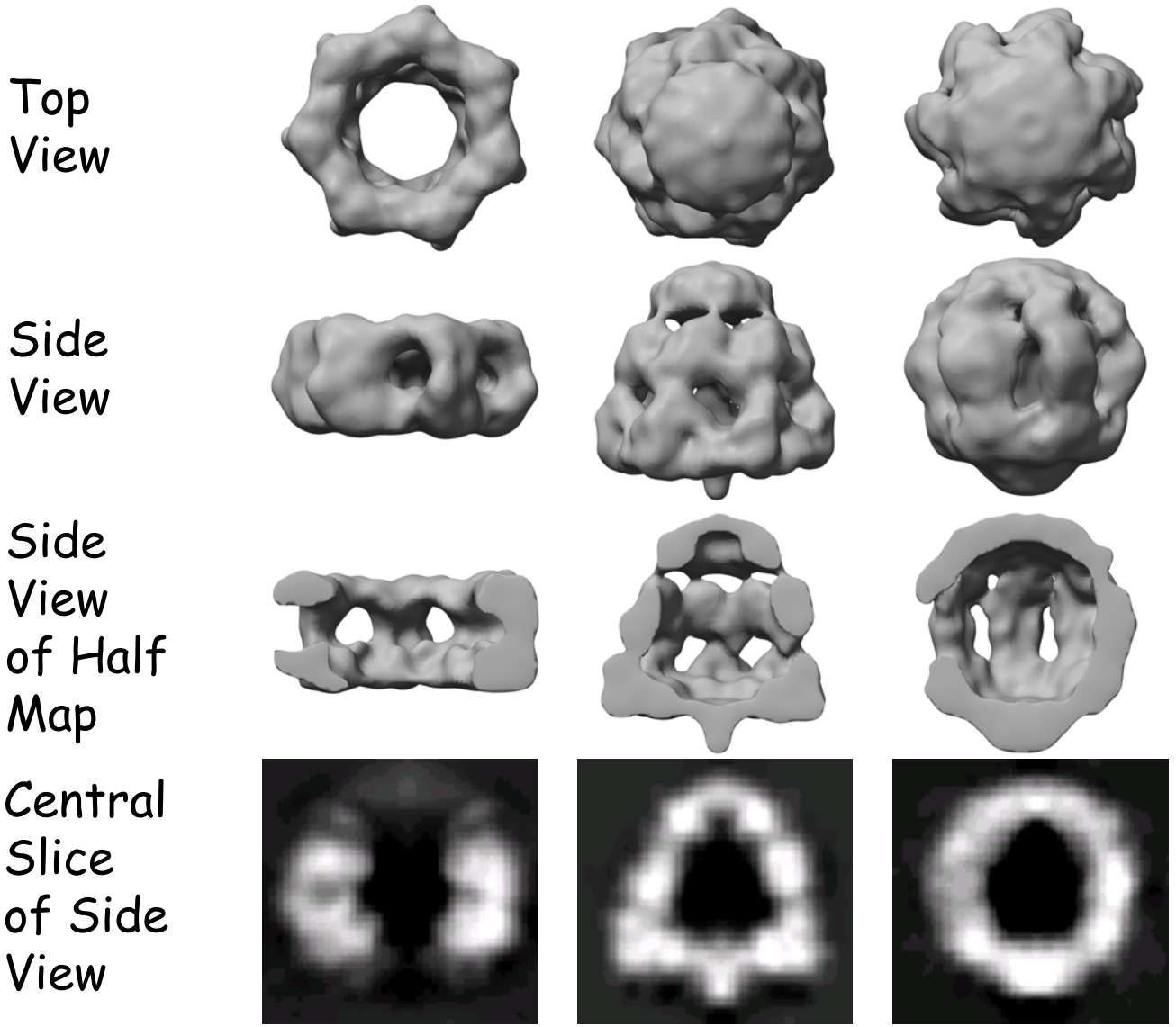



145
69
101
90
70
64
34
100
53
61
22
33
36
28
33
69
29
51
78
71

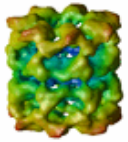
EMAN Multireference Refinement

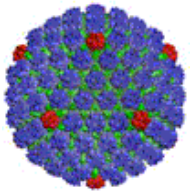


SR398+GroES+Mg-ATP



Acknowledgements





GroEL
- **Donghua Chen**
- **Matt Baker**
- **David Chuang (UTSW)**
- **Jiu-Li Song (UTSW)**
- **Wah Chiu**

Research supported by the NCRR and NIGMS

Graphics produced using Chimera (CGL at UCSF) (supported by NIH P41 RR-01081) and EMANimator.