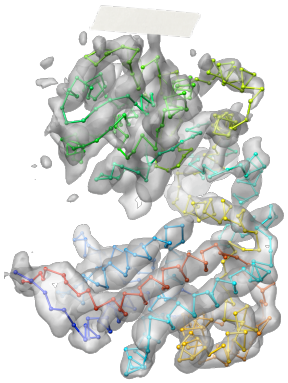
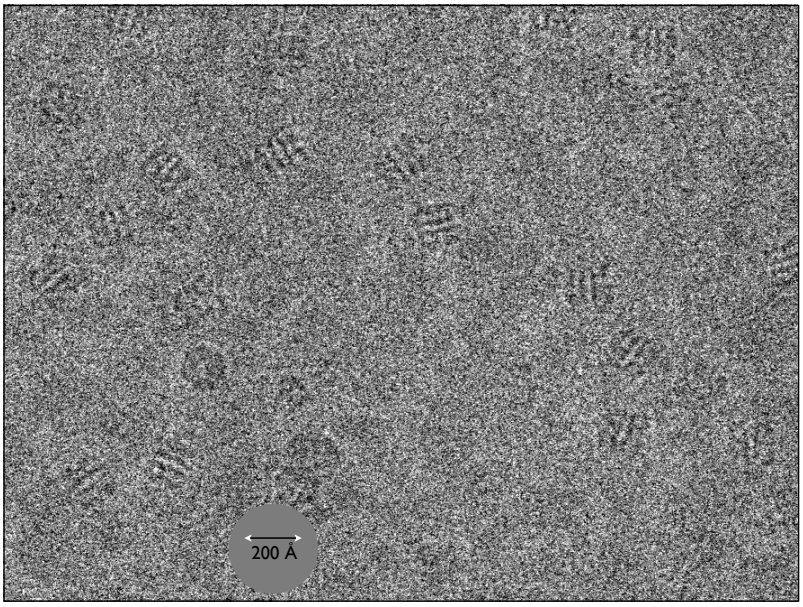
Single Particle Refinement Strategies & Introduction to EMAN2

Steve Ludtke
National Center for Macromolecular Imaging
Baylor College of Medicine

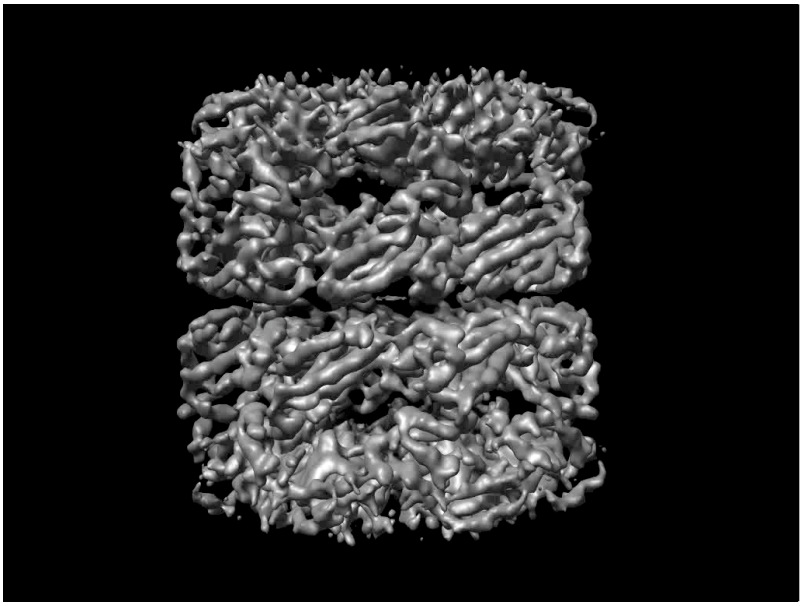


BCM Baylor College of Medicine NCRR NCMI National Center for Macromolecular Imaging NIH Roadmap

1



2



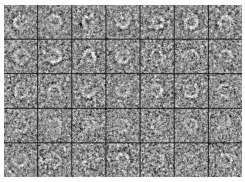
3

2D Analysis

4

2D Refinement

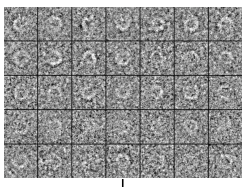
Particles

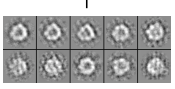


5

2D Refinement

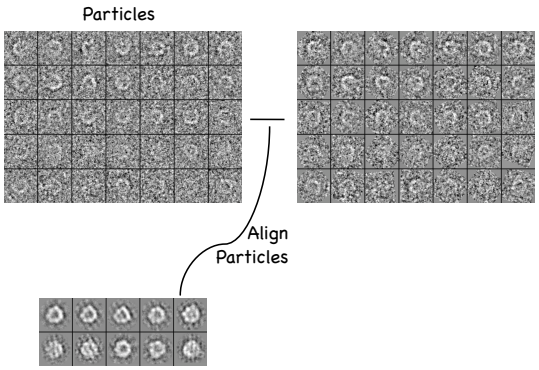
Particles





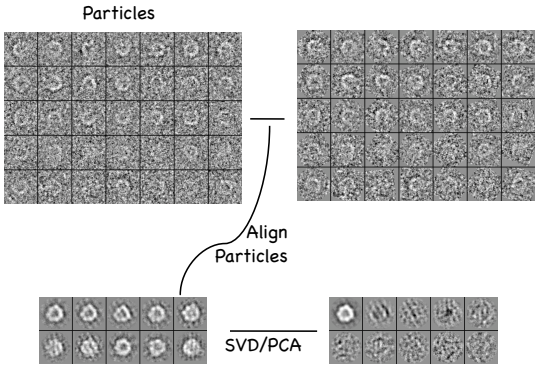
6

2D Refinement



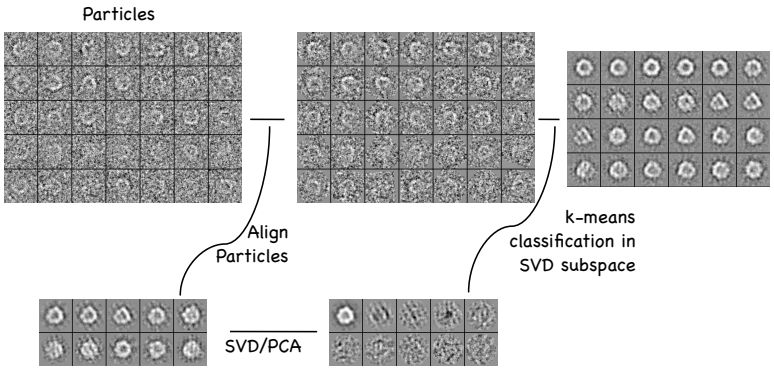
7

2D Refinement



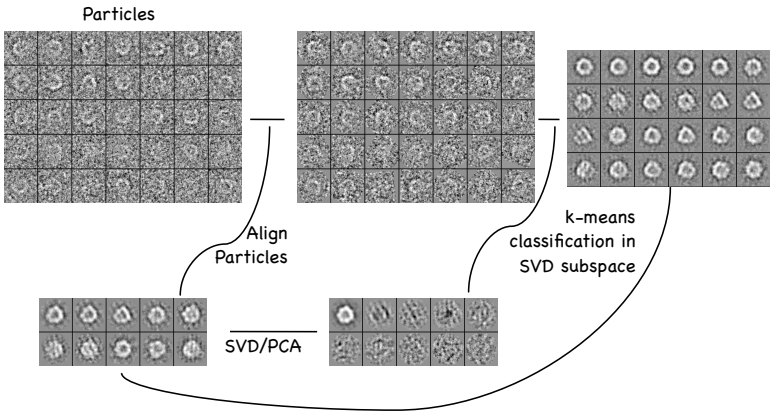
8

2D Refinement



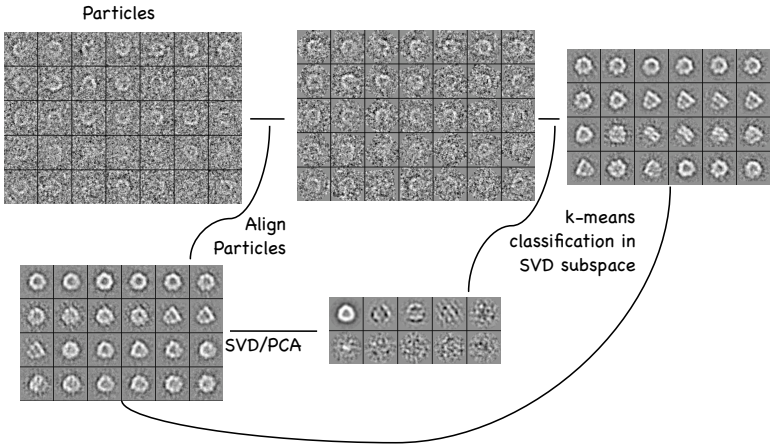
9

2D Refinement



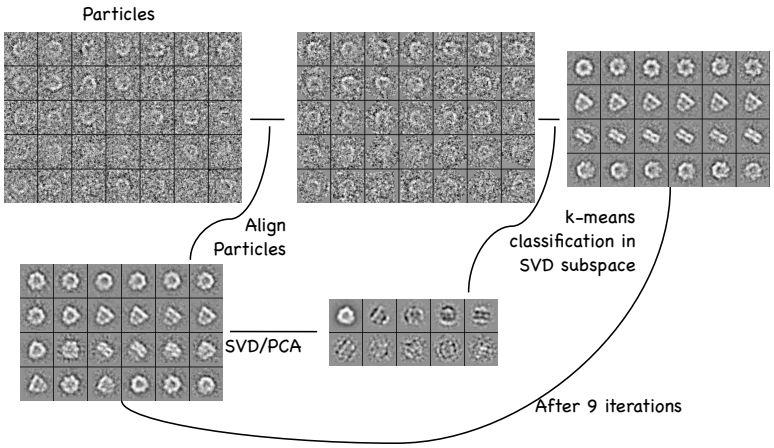
10

2D Refinement

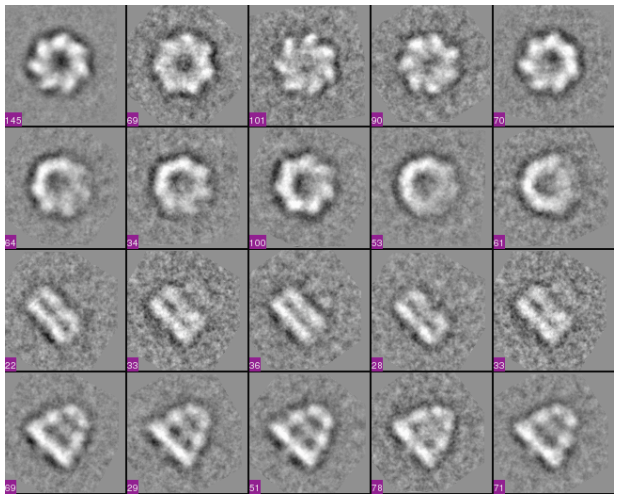


11

2D Refinement



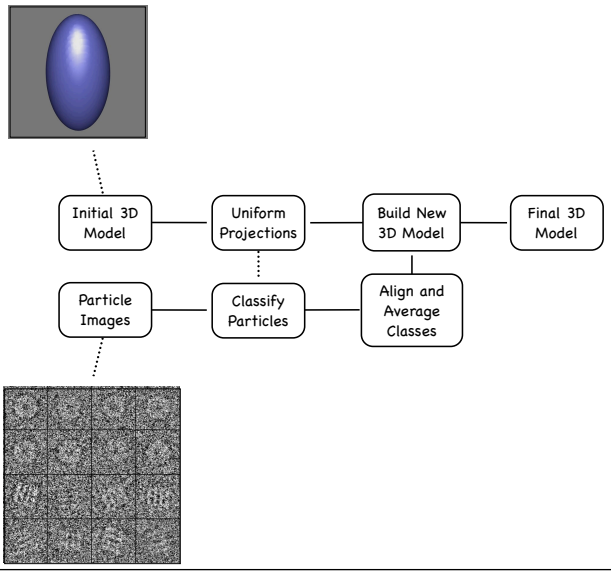
12



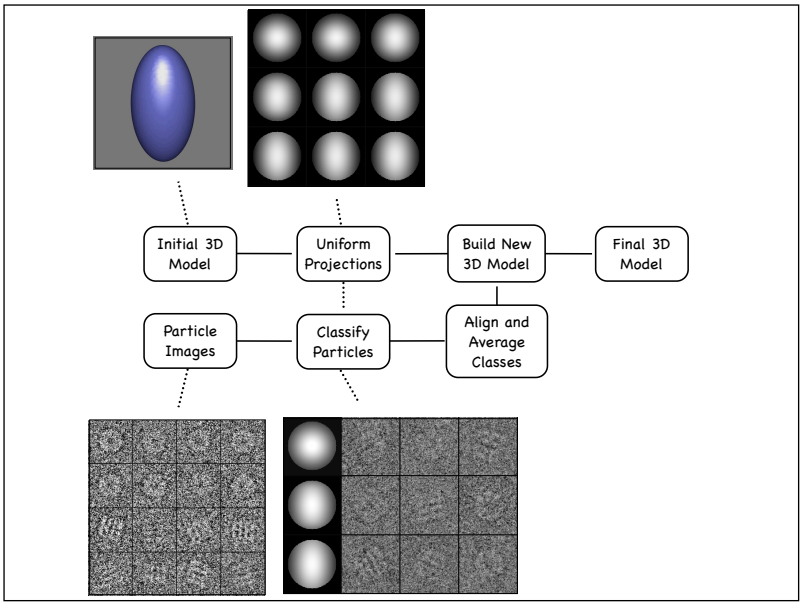
13

3D Reconstruction

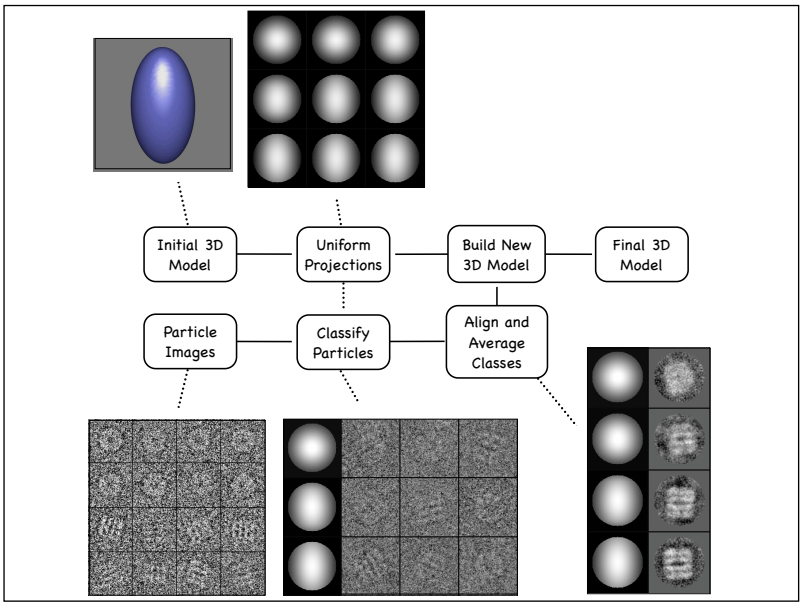
14



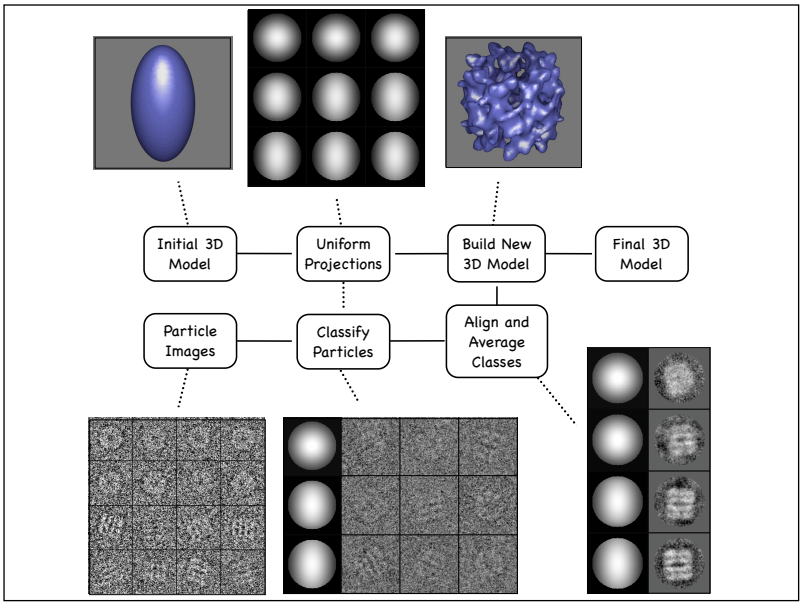
15


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

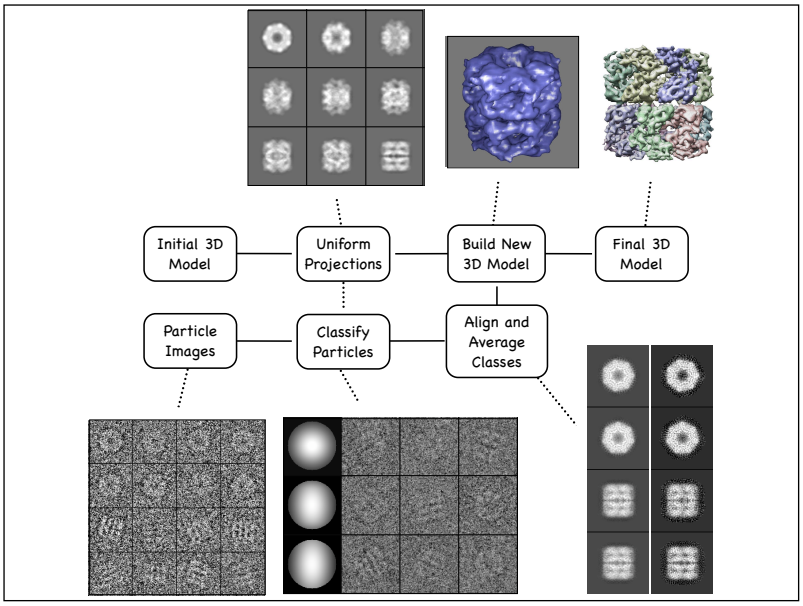
16


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

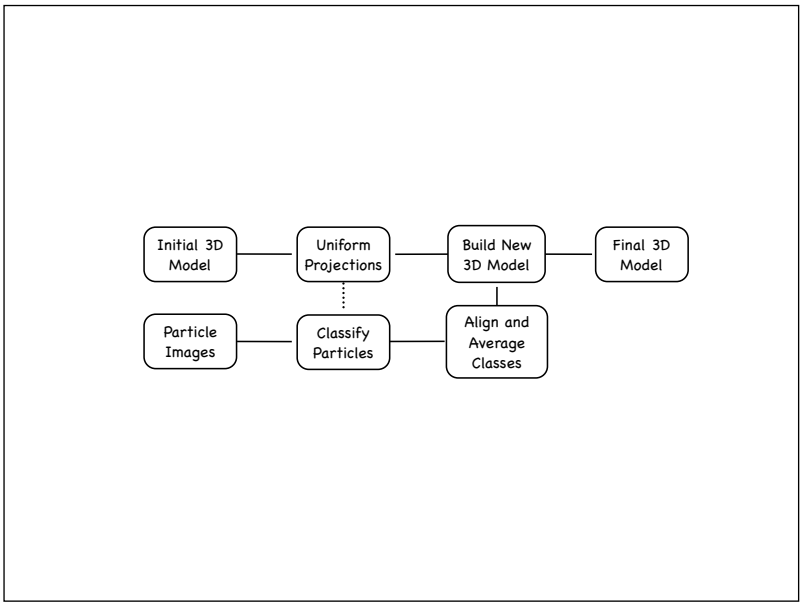
17


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

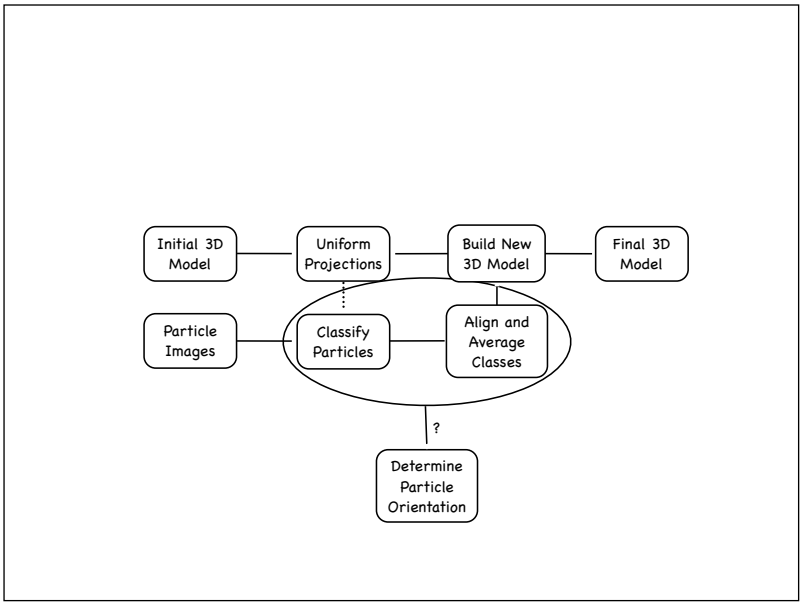
18


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

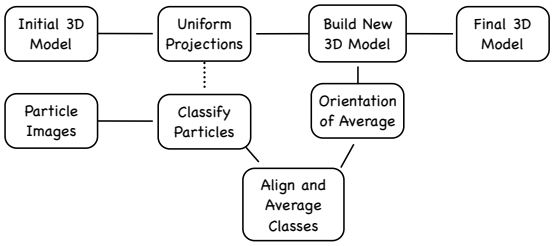
19


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

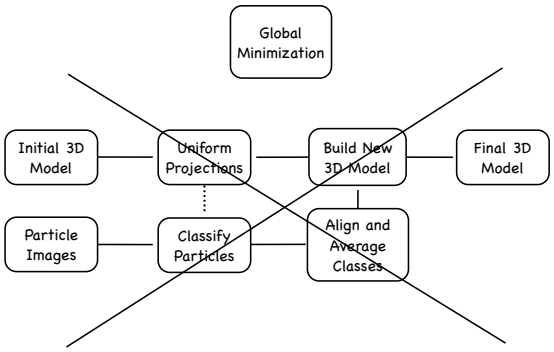
20


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

21



22

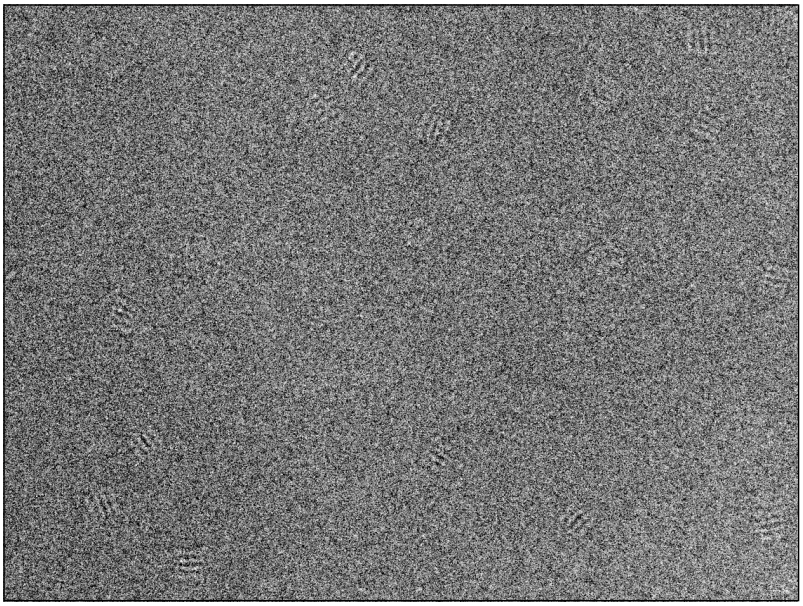


23

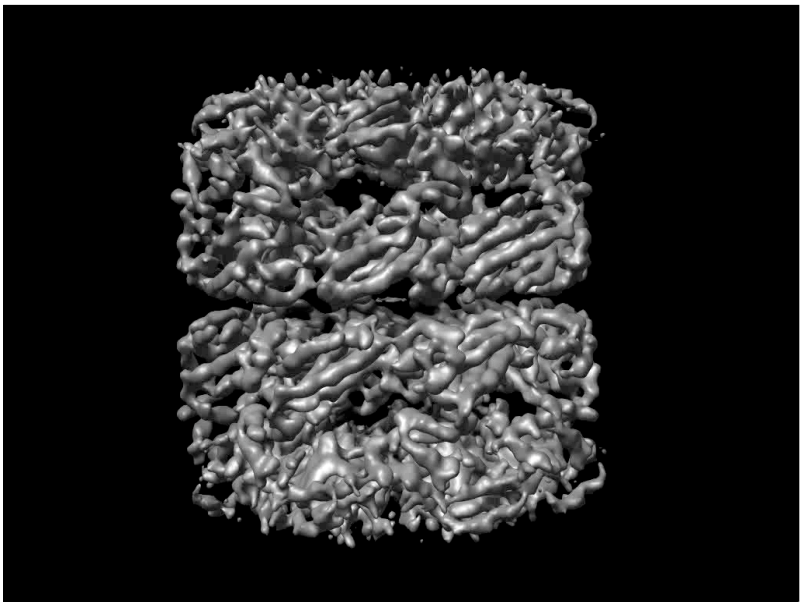
GroEL 2005

- Native, unliganded GroEL, no ATP/ADP (?)
- JEOL 3000SFF (Yoshi-style) at LHe temp
- 6 microscopy sessions, Film
- 825 micrographs, Nikon 9000 @ 6.35 µm scan step
- 60k mag → 1.06 Å/pix
- 135 micrographs used → 20,401 particles
- Defocus 1.2 – 2.2 µm

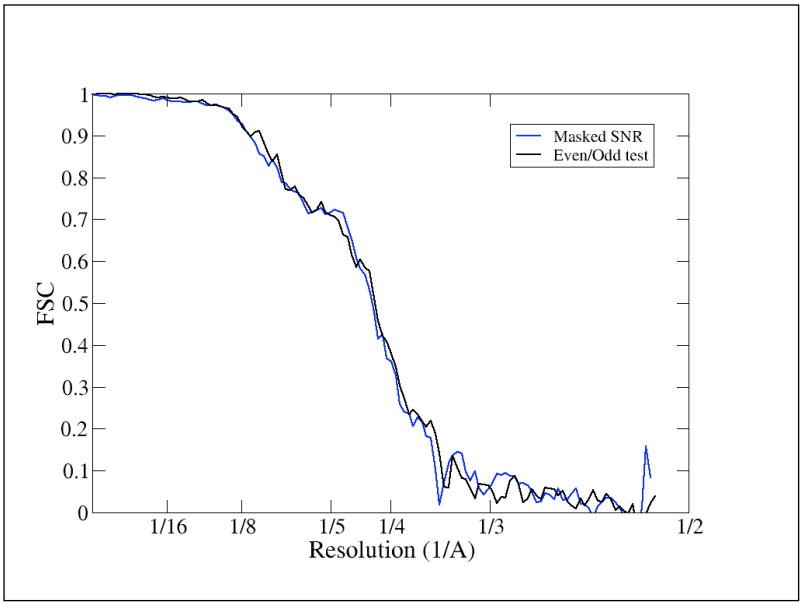
24



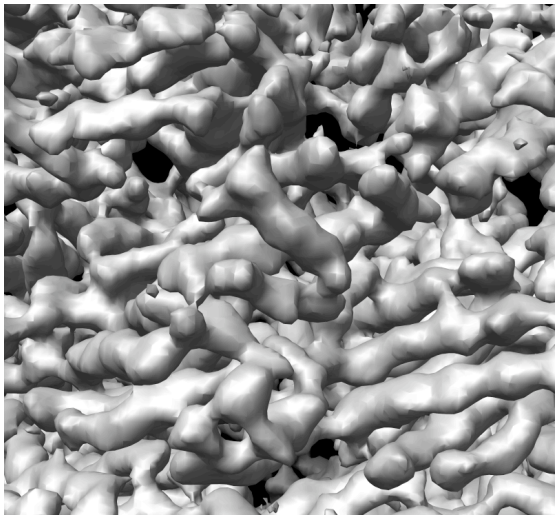
25



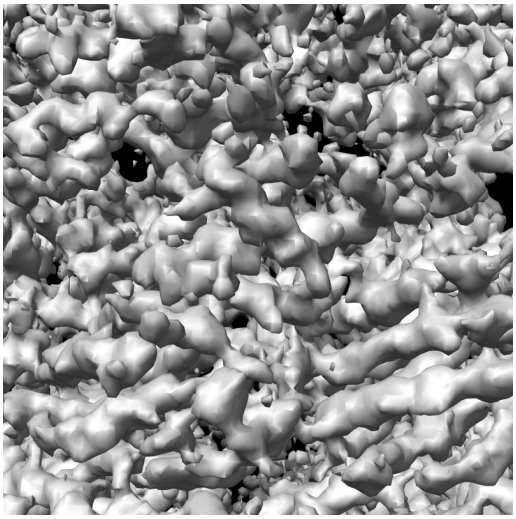
26


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
FSC
1/16
1/8
1/5
1/4
1/3
1/2
Resolution (1/A)
Masked SNR
Even/Odd test

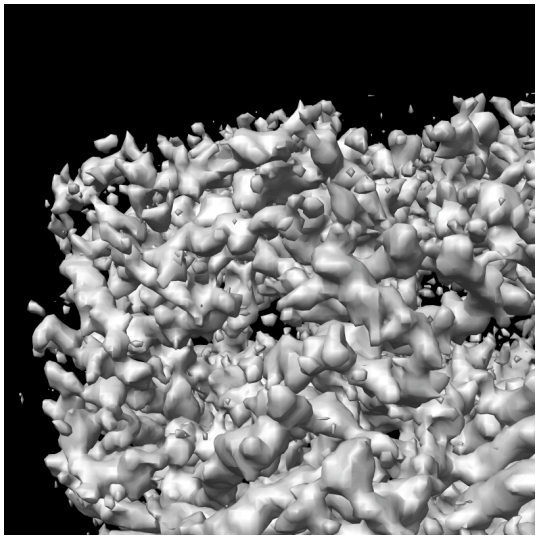
27



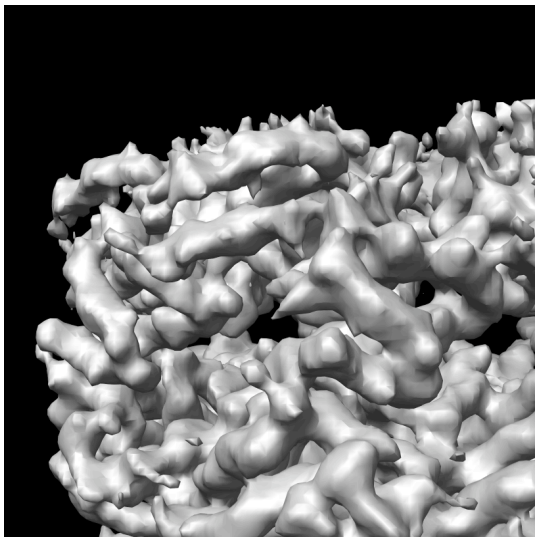
28



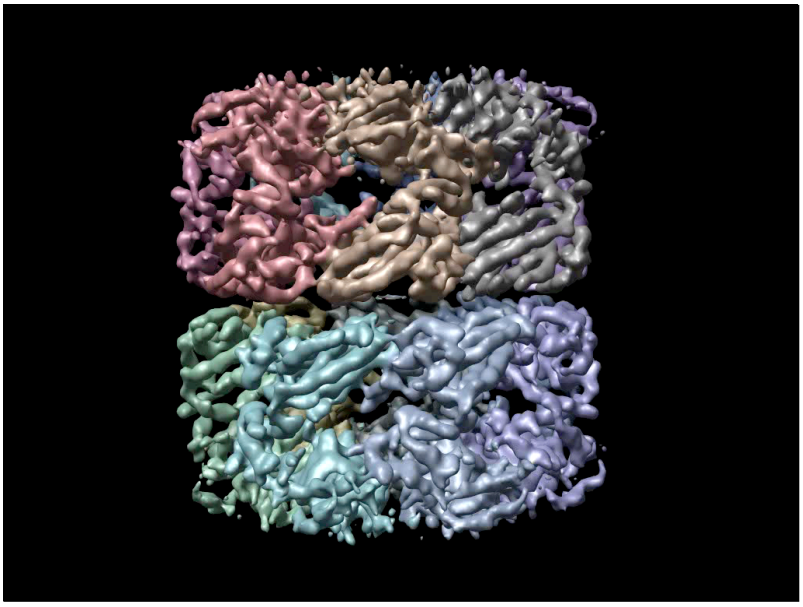
29



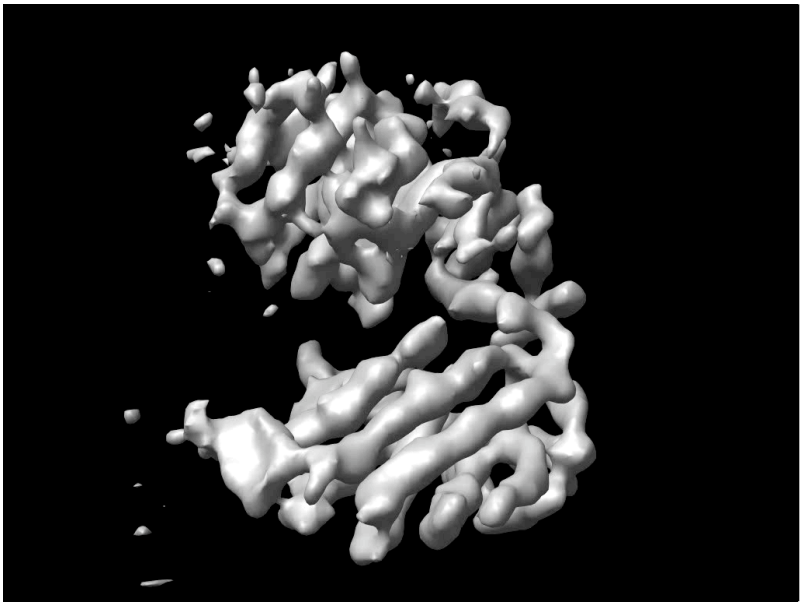
30



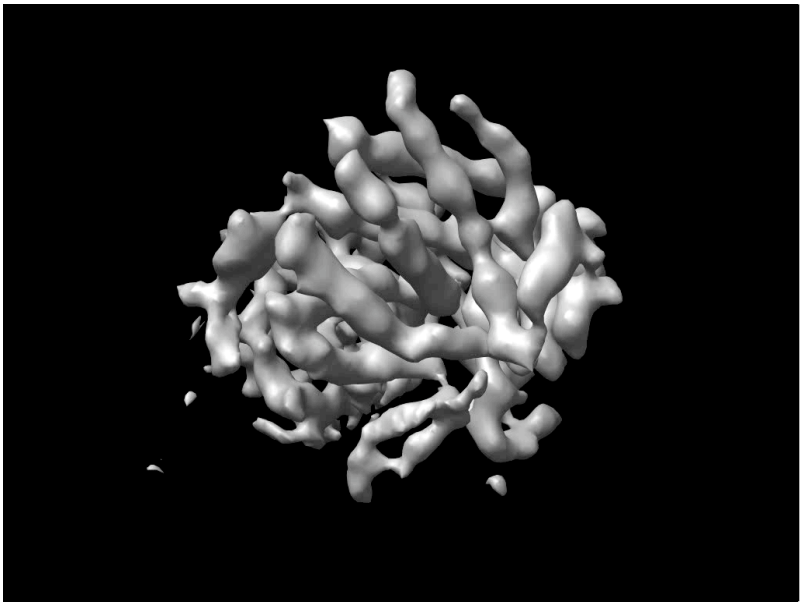
31



32

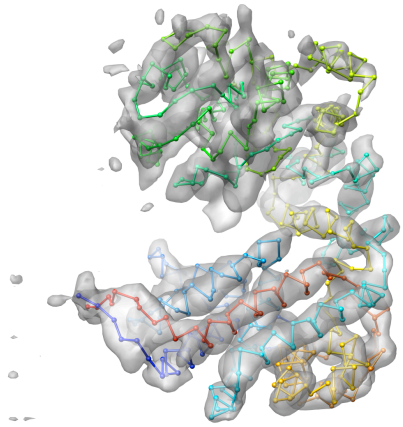


33



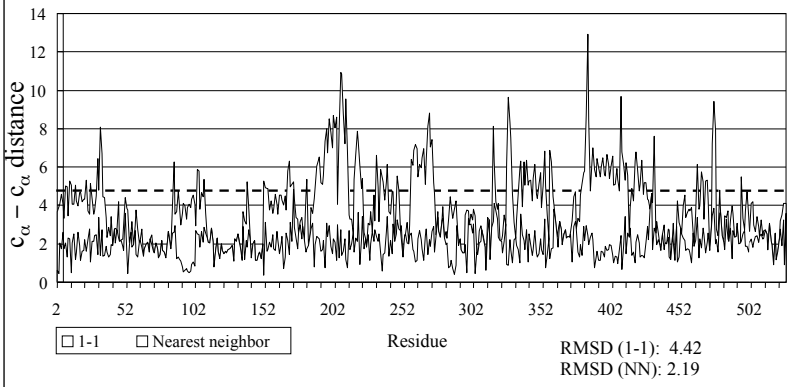
34

GroEL Chain Trace



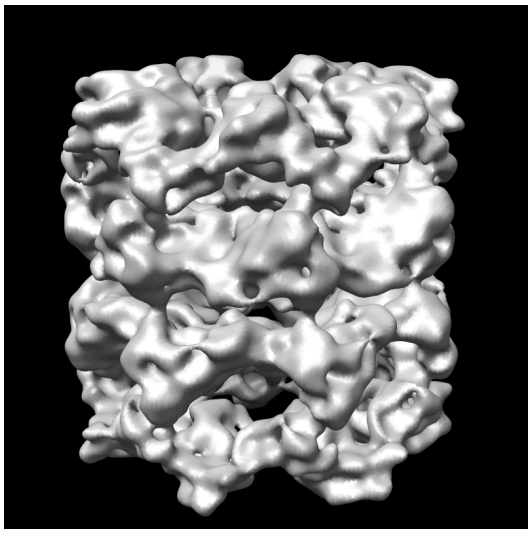
35

RMSD vs. PDB Structure



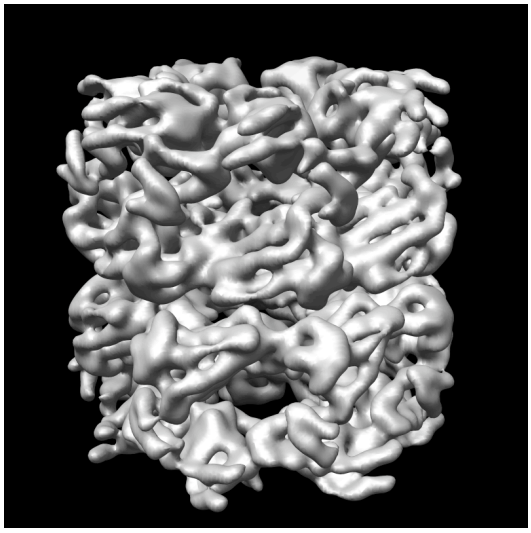
36

11.5 Å GroEL



37

4 Å →11.5 Å GroEL



38

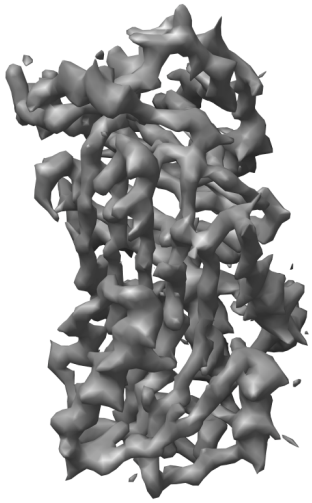
CTF Correction

39

4CAA

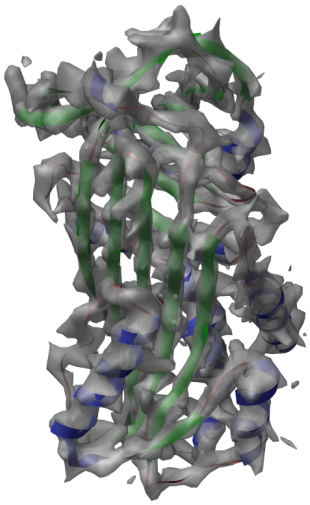
40

4CAA



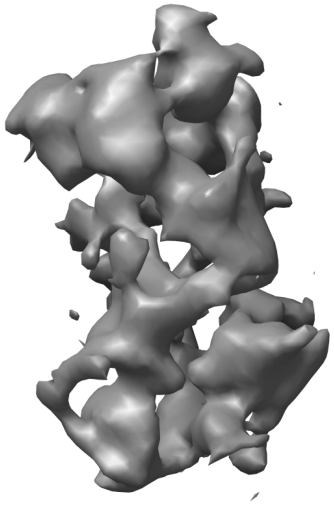
41

4CAA



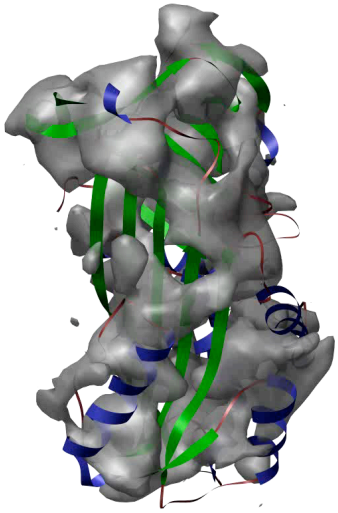
42

No CTF Corr (1 defocus)



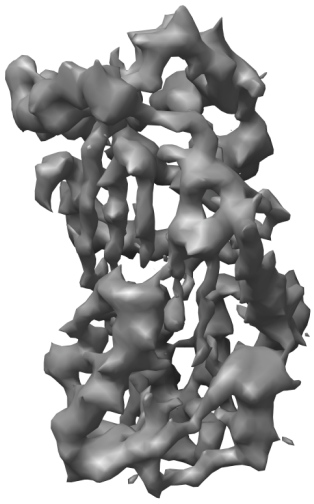
43

No CTF Corr (1 defocus)



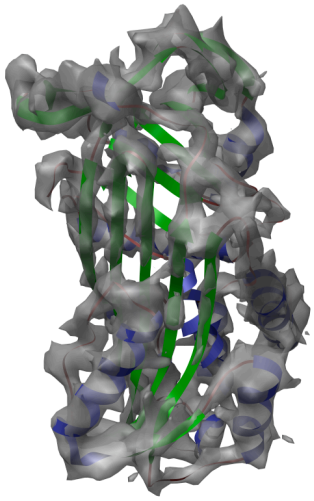
44

Phase Flipped (1 defocus)



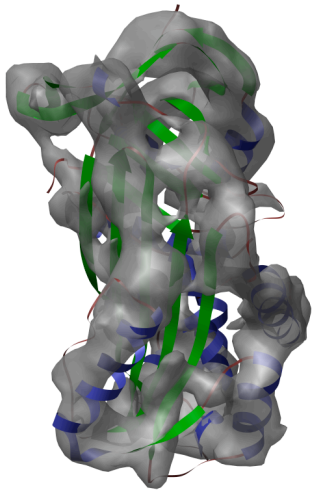
45

Phase Flipped (1 defocus)



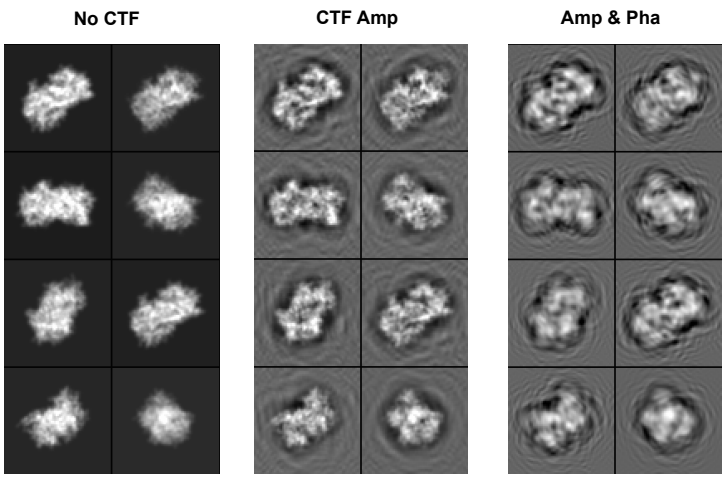
46

Phase Flipped (mult defocus)



47

4CAA in 2D



48

CTF Correction

Measured Image Ideal Particle Random Noise

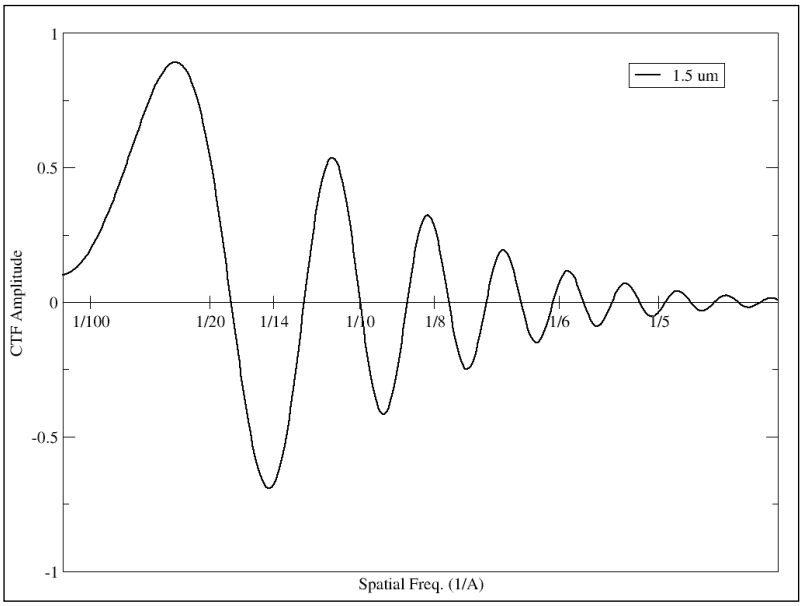
$$\overline{M}(s,\theta) = \overline{F}(s,\theta)C(s)E(s) + \overline{N}(s,\theta)$$

$$C(s) = \sqrt{1-Q^2}\sin\gamma + Q\cos\gamma$$

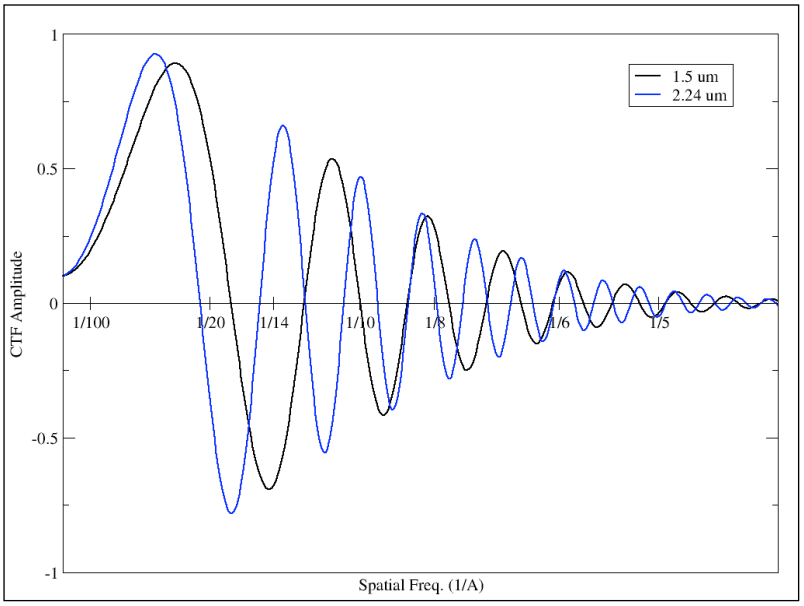
$$\gamma = -\pi\left(\frac{1}{2}C_S\lambda^3 s^4 - \Delta Z\lambda s^2\right)$$

$$E(s) = e^{-Bs^2}$$

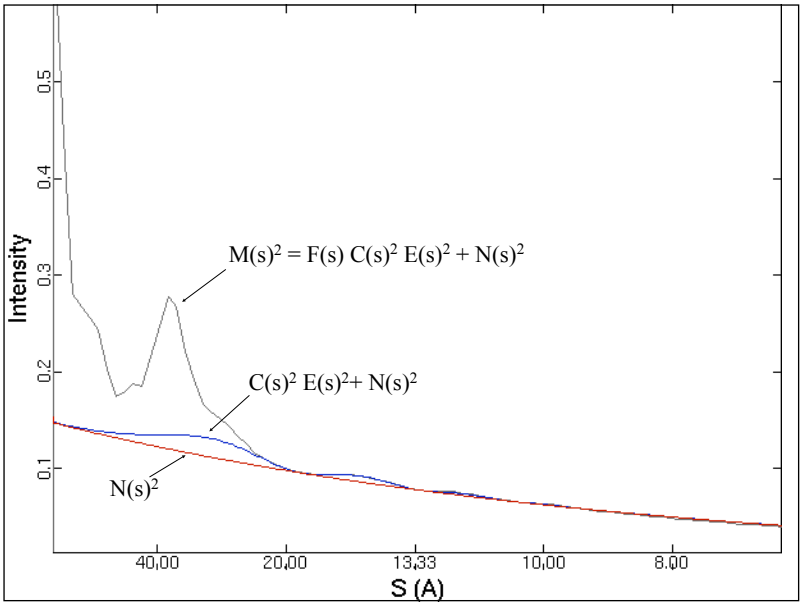
49



50



51



52

CTF Correction

Reconstruction Weight Measured image

$$\overline{T}(s,\theta)=\sum_i k_i \overline{M}_i(s,\theta)$$

$$k_i = ?$$

- Maximize SNR of *T(s,θ)*
- Minimize RMSD between T and F

$$\sqrt{\sum_{x,y}(t(x,y)-f(x,y))^2}$$

53

CTF Correction

Wiener Filter | CTF Correction | SNR Weight

$$\overline{T}(s,\theta)=\frac{F^2(s)R(s)}{1+F^2(s)R(s)}\sum_i \frac{1}{C_i(s)E_i(s)}\frac{R_i(s)}{R(s)}\overline{M}_i(s,\theta)$$

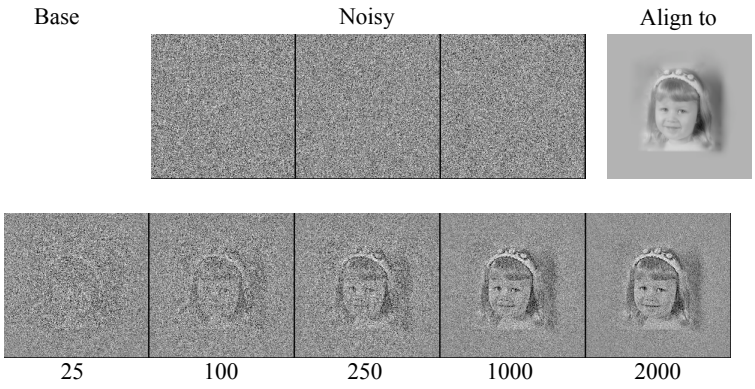
$$R_i(s)=\frac{C_i^2(s)E_i^2(s)}{N_i^2(s)} \qquad R(s)=\sum_i \frac{C_i^2(s)E_i^2(s)}{N_i^2(s)}$$

54

Model Bias ?

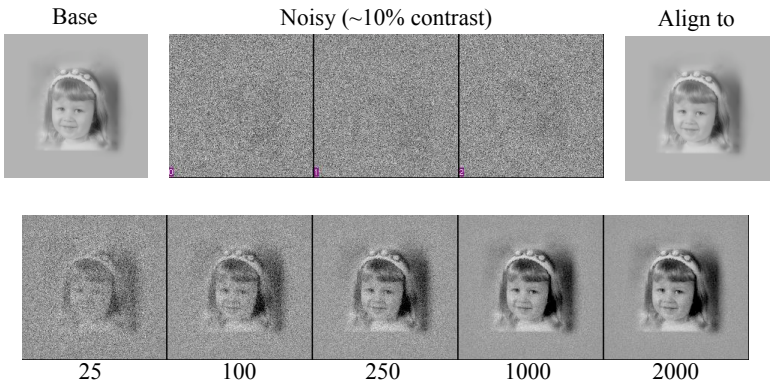
55

Model Bias



56

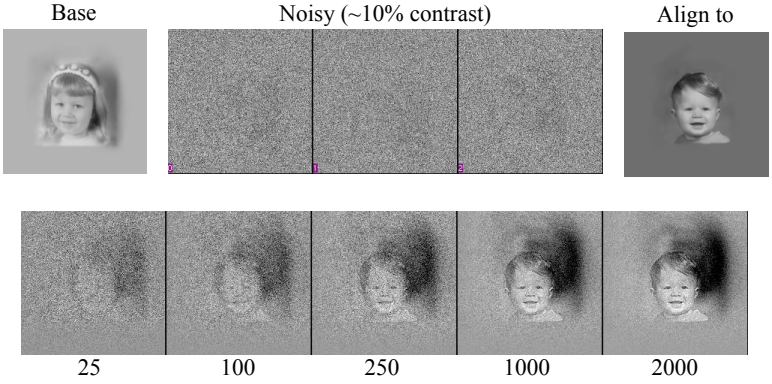
Model Bias



57

Model Bias

Base | Noisy (~10% contrast) | Align to

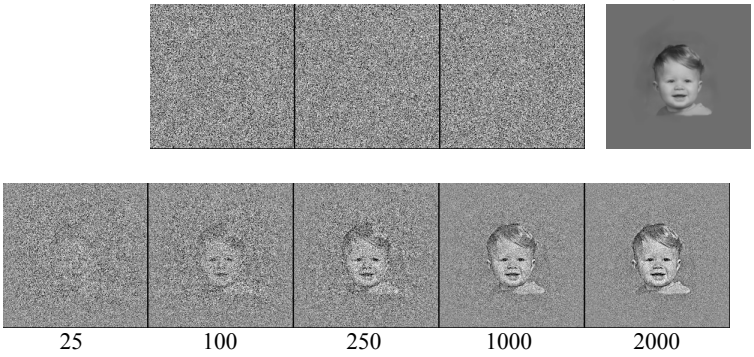


25 | 100 | 250 | 1000 | 2000

58

Model Bias

Base | Noisy | Align to



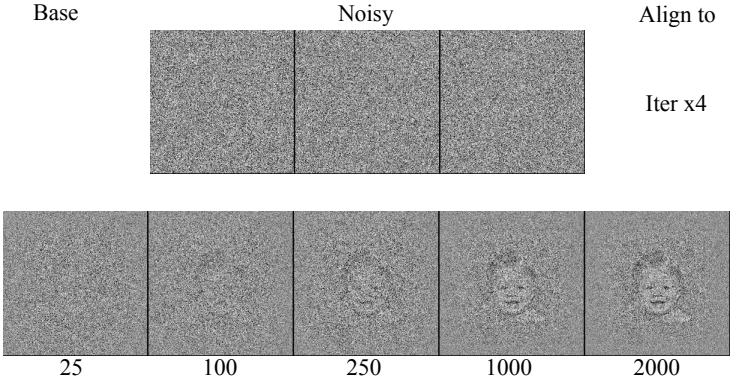
25 | 100 | 250 | 1000 | 2000

59

Model Bias

Base | Noisy | Align to

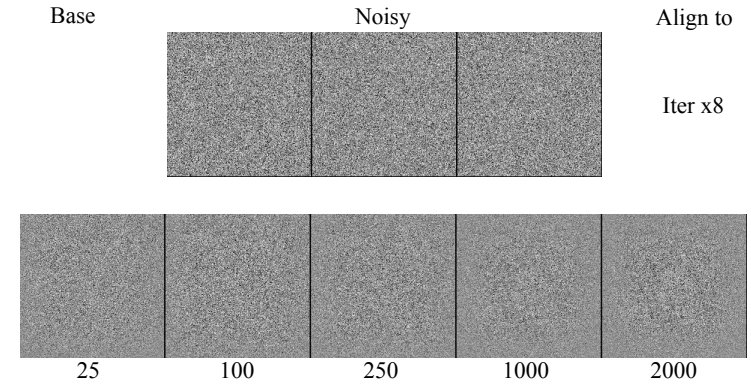
Iter x4



25 | 100 | 250 | 1000 | 2000

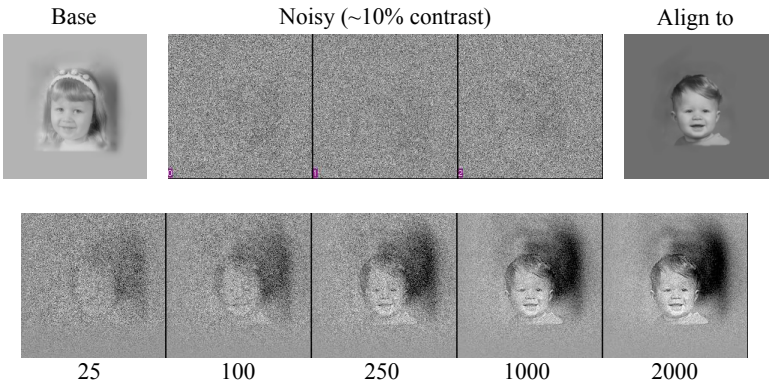
60

Model Bias



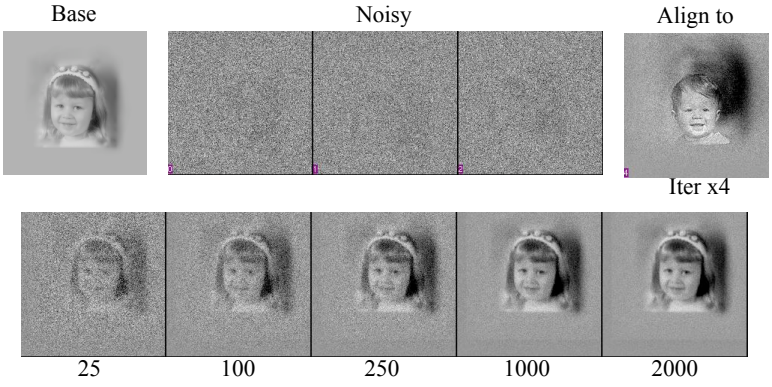
61

Model Bias



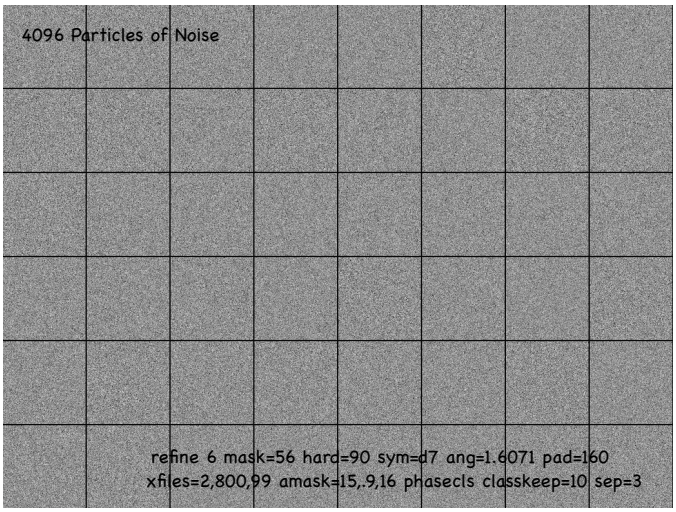
62

Model Bias



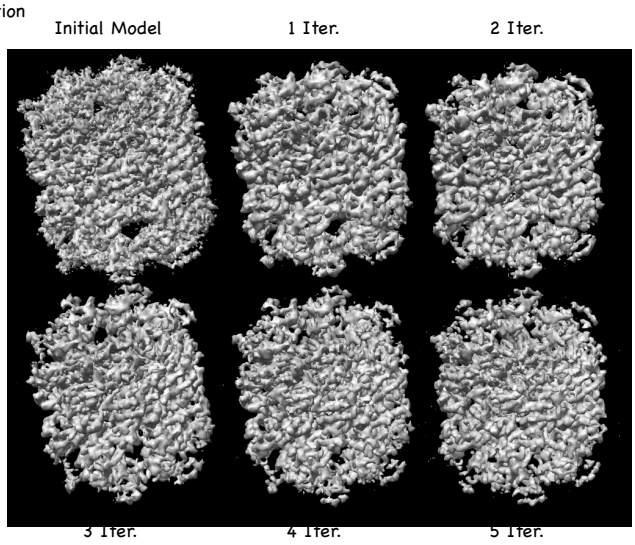
63

How About 3-D ?

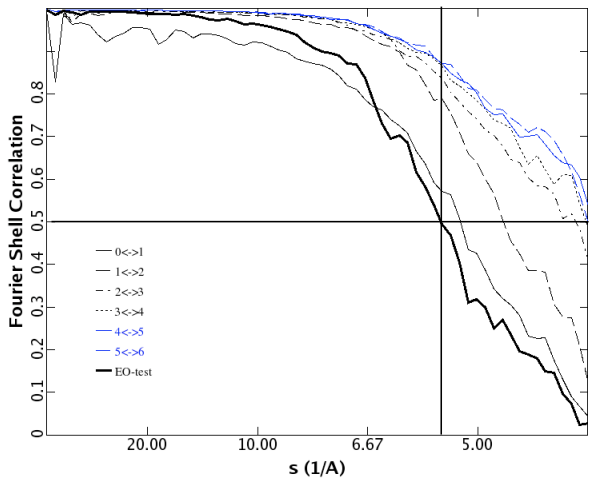


64

no iteration

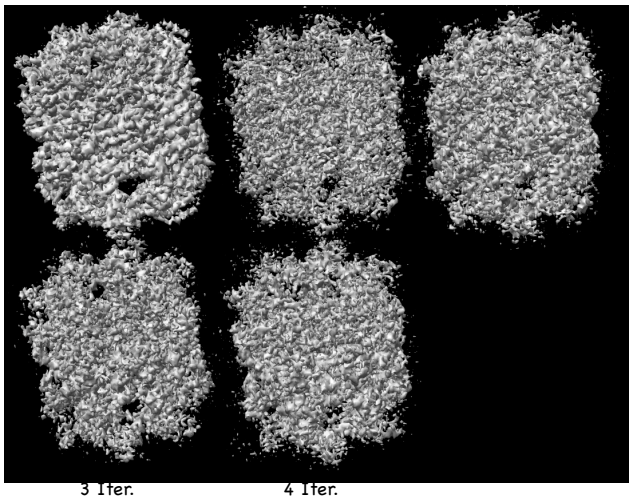


65



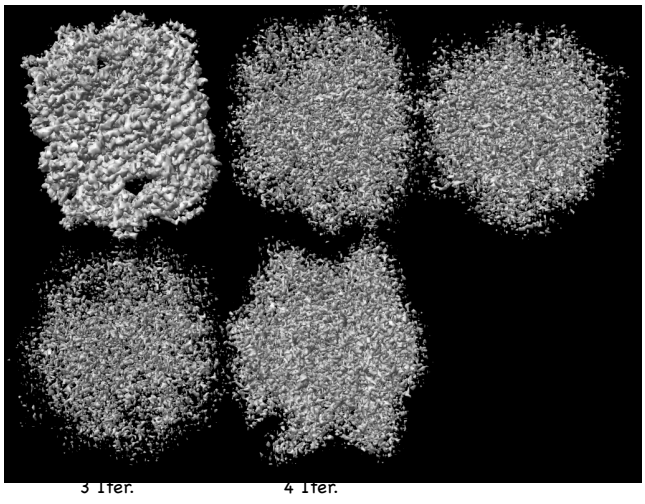
66

1 iteration



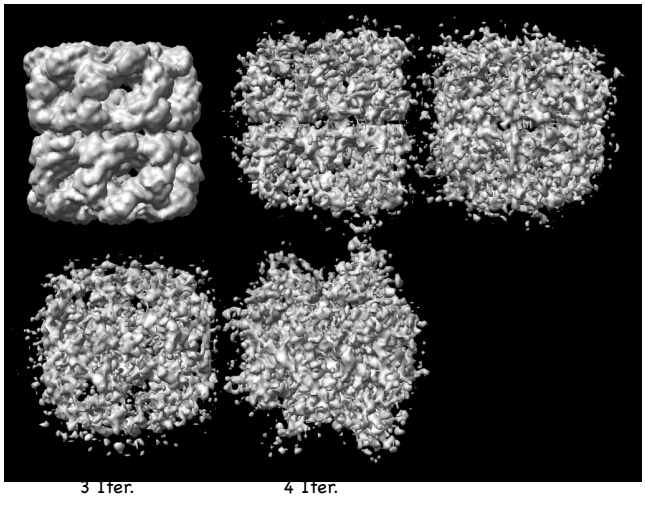
67

6 iterations

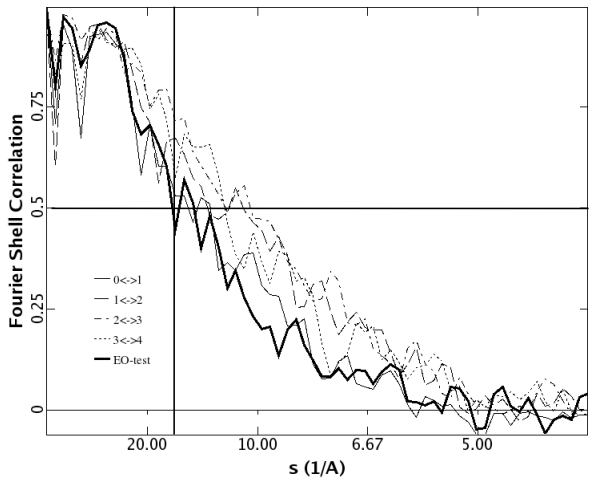


68

6 iterations
(8 A lowpass)



69



70

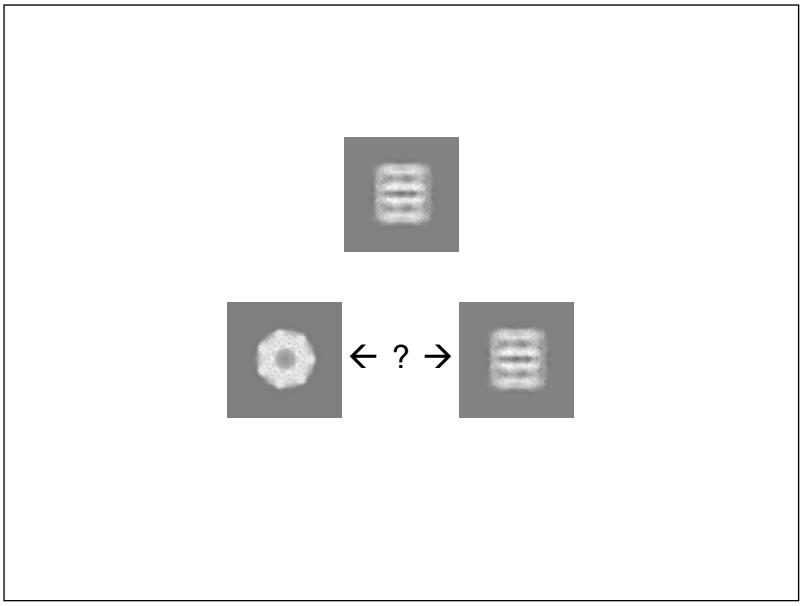
How Do we Stop This ?

- (In EMAN) use classiter>3 for a few rounds
- Use several different (random) starting models and insure that you get a good answer
- Compare 3D models with results of 2D analysis

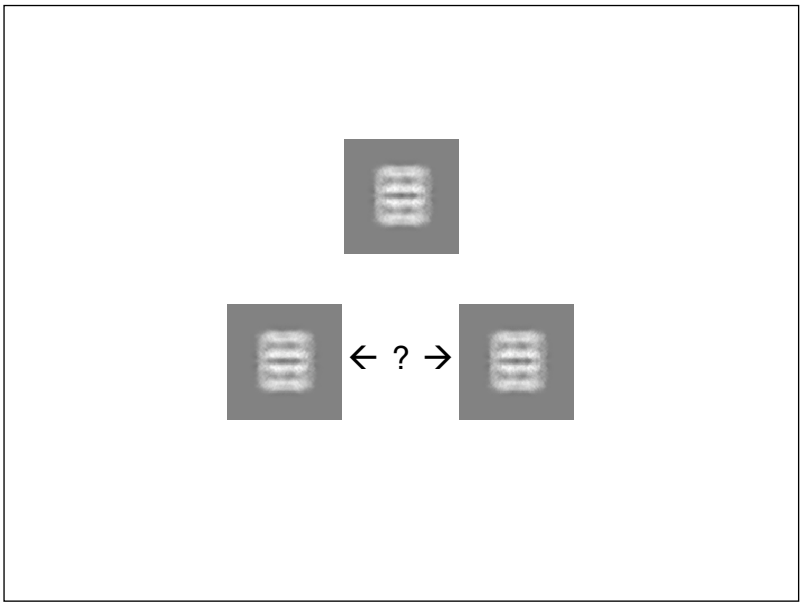
71

Measures of Similarity

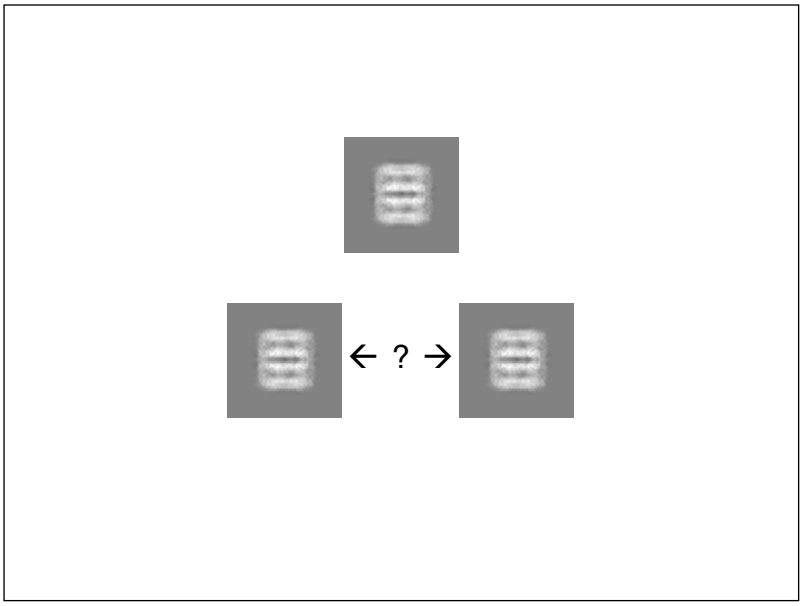
72



73



74

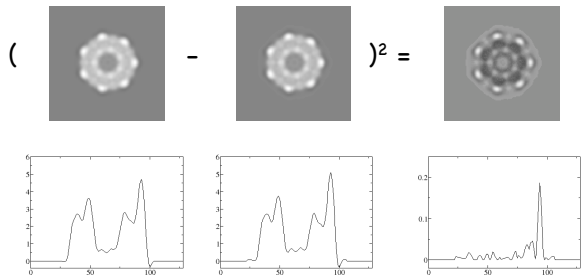


75

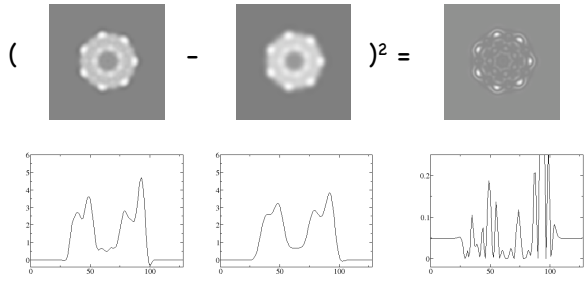
Measures of Similarity

- Correlation coefficient
- Variance (equivalent)
- Phase Residual
- FSC
- Mutual Information
- etc…

76



77



78

EMAN2

79

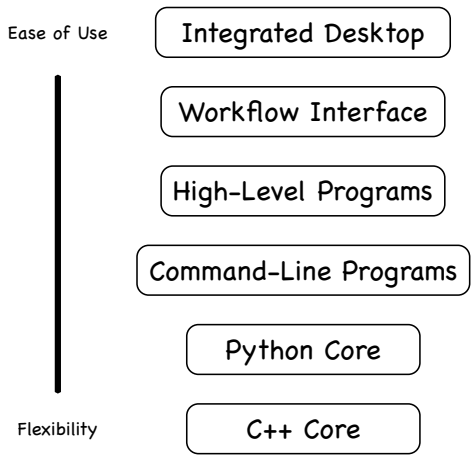
EMAN2 vs. EMAN1

- Improved CTF model
  - Automatic fitting, Astigmatism*, Energy filtered data
- New openGL based GUI
- Workflow infrastructure
- Embedded database for data storage and metadata archival
- EMEN2 Integration *
- Easily extensible image processing infrastructure
- New parallelism strategy ~
- CUDA support *

* - not yet ready for use

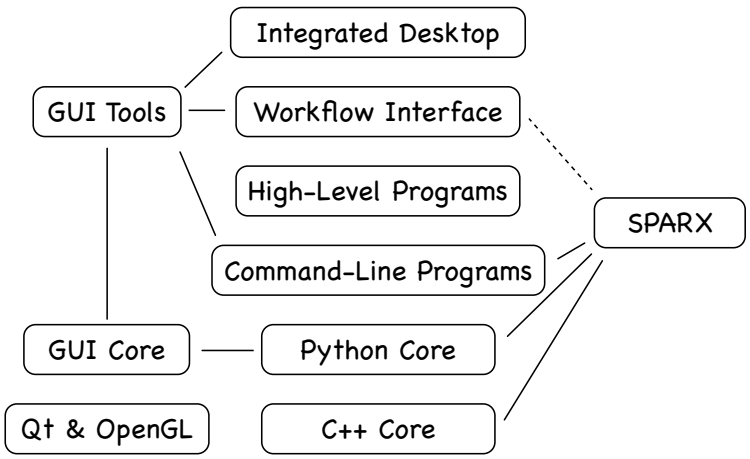
80

EMAN2 Architecture



81

EMAN2 Architecture



Extensible Core

Type	Description	#
Processor	Generic image processing algorithms, filters, masks, thresholds, etc.	157
Aligner	Algorithms used to align 2 images or volumes to each other	11
Projector	Routines to generate 2-D projections of 3-D objects	7
Reconstructor	Routines to reconstruct 3-D objects from 2-D projections	11
Cmp	Similarity metrics used to compare two images or volumes	9
Averager	Average together stacks of images in various ways	9
Analyzer	Perform various operations on sets of images, such as classification or PCA	6
Orientgen	Routines describing how projections cover the asymmetric triangle	6

File Formats

BDB +

MRC	R/W	IMAGIC	R/W
SPIDER	R/W	HDF5	R/W
PIF	R/W	ICOS	R/W
VTK	R/W	PGM	R/W
Amira	R/W	Xplor	W
Gatan DM2	R	Gatan DM3	R
TIFF	R/W	Scans-a-lot	R
LST	R/W	PNG	R/W
Video-4-Linux	R	JPEG	W

Processors

(categories & examples)

- filter
  - filter.lowpass.gauss
  - filter.homomorphic.tophat
- mask
  - mask.sharp
  - mask.gaussian
- math
  - math.sqrt
  - math.laplacian
- misc
  - misc.localnorm
- normalize
  - normalize
  - normalize.edgemean
- testimage
  - testimage.scurve
- threshold
  - threshold.binary
  - threshold.clampminmax
- xform
  - xform.centerofmass
  - xform.fourierorigin.tocenter

85

Similarity Metrics

(cmp)

With Default options, SMALLER -> more similar

- dot - dot product (negative by default)
- frc - Fourier ring correlation (weighted)
- optvariance - 'optimized variance' (EMAN1)
- phase - mean phase error
- quadmindot - Worst of quadrant dot products
- sqeuclidean - sum $(a-b)^2/n$

86

Programs

- 49 Command-Line Programs (EMAN2)

syntax:

e2<name>.py --help

e2<name>.py <file> [--option=value] [--option] [-O]

<> - required parameter

[] - optional parameter

87

GUI

- e2desktop.py (may not be stable yet)
- e2workflow.py
- e2display.py
- and other programs with the --gui option

88

EMEN2

Electron Microscopy Electronic Notebook

89

The Database Problem

- Encourage good recordkeeping (automation)
- Find something after a person leaves the lab
- Reconstruct an experimental protocol refined over years
- Mine experimental data for new information
- Minimize db overhead/administration

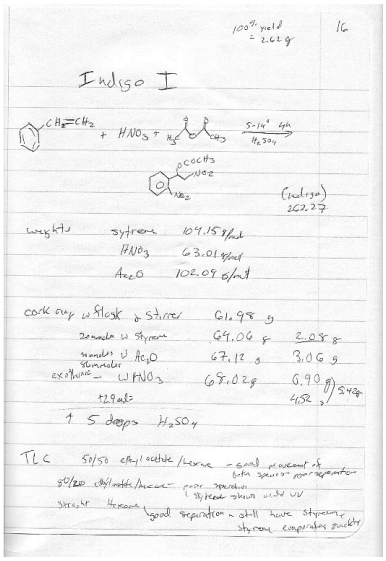
90

Film/Frame Id	Defocus	Mag	Total Dose	ExpTime
ccd_2004031-#1	0.8 um	138.44k	20.0	1.0 s
ccd_2004031-#2	2.5 um	138.44k	20.0	1.0 s
ccd_2004031-#1	0.8 um	138.44k	20.0	1.0 s
ccd_2004031-#2	2.5 um	138.44k	20.0	1.0 s
ccd_2004031-#1	0.8 um	138.44k	20.0	1.0 s
ccd_2004031-#2	2.5 um	138.44k	20.0	1.0 s
ccd_2004031-#1	0.8 um	138.44k	20.0	1.0 s
ccd_2004031-#2	2.5 um	138.44k	20.0	1.0 s
ccd_search1 #1	0.8 um	138.44k	20.0	1.0 s
ccd_search2 #2	2.5 um	138.44k	20.0	1.0 s
ccd_search3 #1	0.8 um	138.44k	20.0	1.0 s
ccd_search4 #2	2.5 um	138.44k	20.0	1.0 s
ccd_search5 #1	0.8 um	138.44k	20.0	1.0 s
ccd_search6 #2	2.5 um	138.44k	20.0	1.0 s

Scientific Database

- Excellent mineability
- Good for centralized databases
- Limited flexibility
- Substantial overhead (design & maintenance)

91

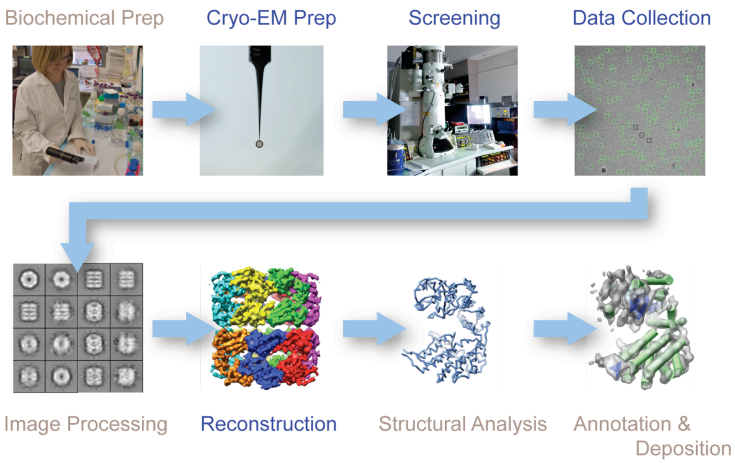


Electronic Lab Notebook

- Excellent flexibility
- Rich information content
- Limited mineability

92

Archive Entire Cryo-EM Pipeline

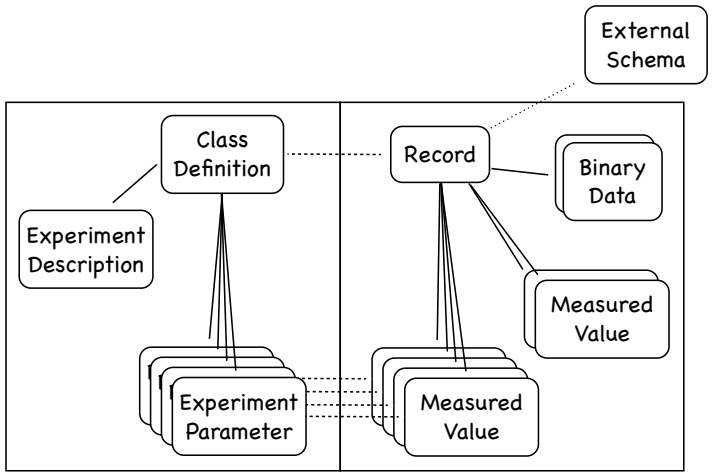


93

Object Oriented Database

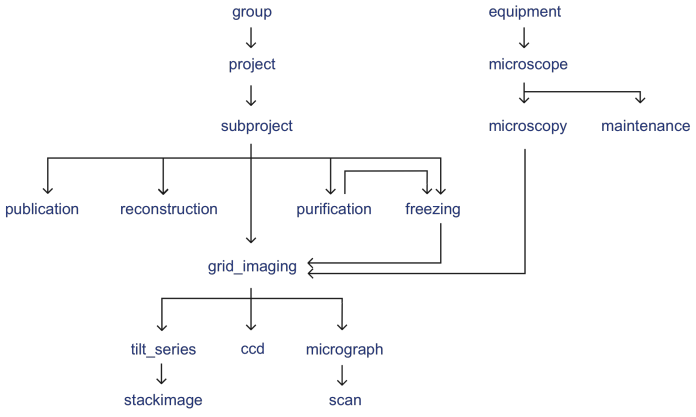
- Flexible records not tied to rigid definitions
- Connectivity between arbitrary records
- Queries not easily posed in a traditional database
- Per-record security
- Security and queries aware of record hierarchy

94



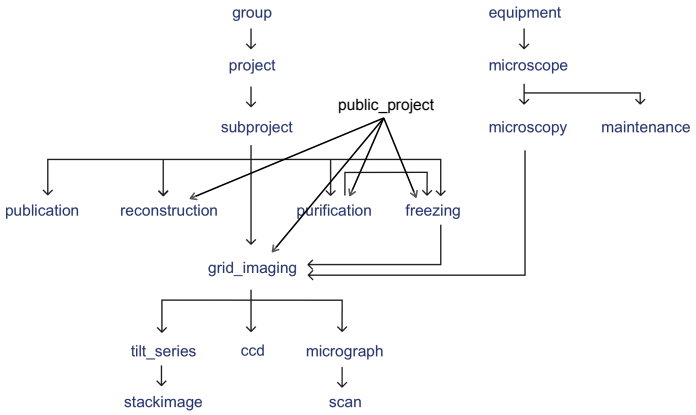
95

Record Organization



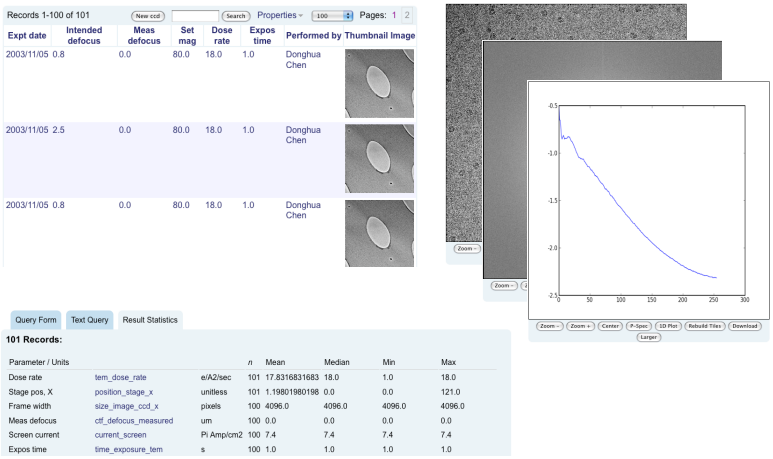
96

Record Organization



97

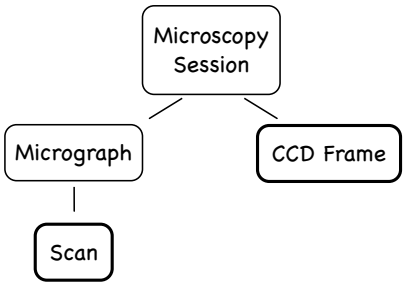
Analysis Tools



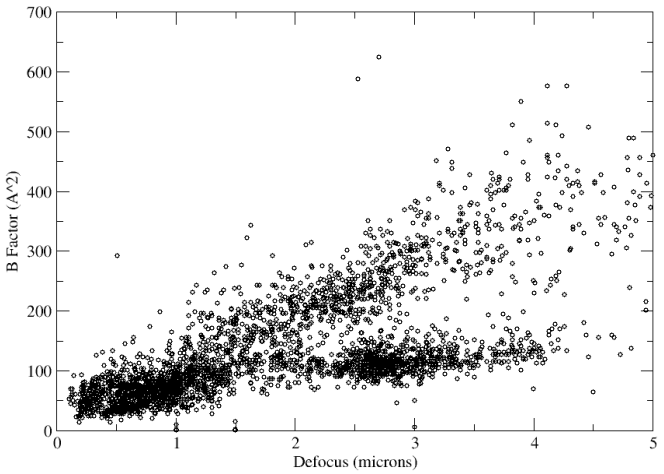
98

plot bfactor vs truedefocus where truedefocus is between 0.1 and 5.0 and bfactor is between 1 and 1000

99



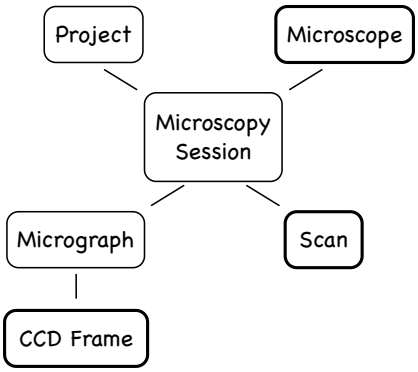
100



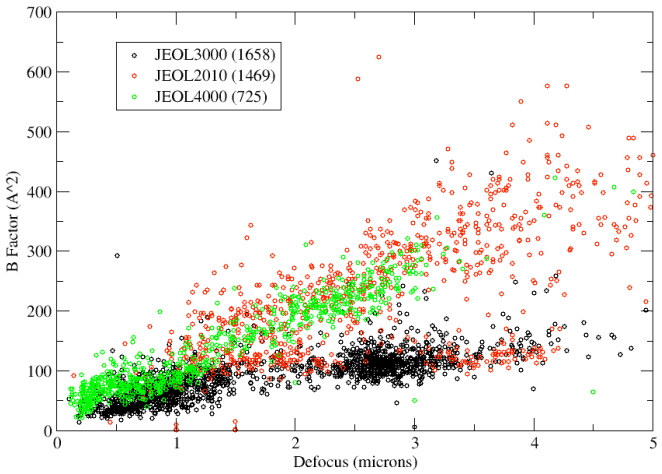
101

plot bfactor vs truedefocus where truedefocus is between 0.1 and 5.0 and bfactor is between 1 and 1000split by microscope

102

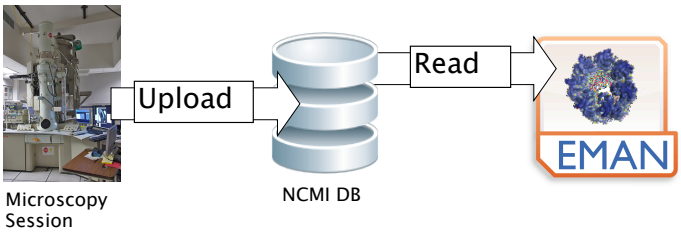


103



104

GOAL: Integrate with EMAN2



105

GOAL: Integrate with EMAN2

Microscopy Session

Upload

NCMI DB

Read

Write

EMAN

106

GOAL: Harvest Data for Export

NCMI DB

XML

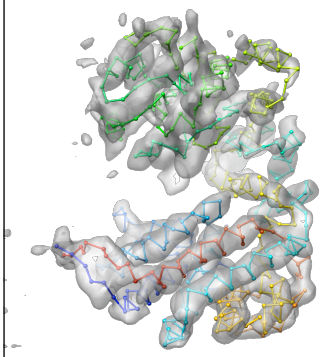
WORLDWIDE PDB PROTEIN DATA BANK

EMDB EM DATA BANK

Value	EMEN2 Parameter	EMDB Name
	microscope: 76	
JEOL 3000SFF	tem_name	microscope
	microscopy: 204173	
kodak SO163 film	film_type	detector
	grid_imaging: 204174	
4.2 K	temperature_specimen	temperature
	micrograph: 204280, 204281, etc.	
25 e/A^2	tem_dose	electronDose
500 nM	ctf_defocus_set	(range) nominalDefocusMin
1200 nM		nominalDefocusMax

107

Acknowledgements



EMEN2

- Ian Rees
- Ed Langley
- Kurt Welgehausen
- Haili Tu
- Deepy Mann

GroEL

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

EMAN2

BCM

- David Woolford
- Guang (Grant) Tang
- Liwei Peng
- Ian Rees
- Phil Baldwin
- Deepy Mann
- Wen Jiang (Purdue)

Via SPARX

- Pawel Penczek (UTH)
- Wei Zhang (UTH)
- Zhengfan Yang (UTH)
- Julien Bert (UTH)
- Stefan Raunser (MPI)
- Christian Spahn (Charité)
- Justus Loerke (Charité)
- Chao Yang (LBNL)

Supported by NIH Roadmap Initiative, NCRR, NIGMS, and the Welch Foundation.

Graphics produced using UCSF Chimera and EMANimator.

108