Applying the Automated EM Pipeline: One quarter of a million particles of GroEL per day

Or what do I do with all these data?

Outline

- What are the steps one takes to use automation in practice?
- What are the obstacles one encounters along the way to a reconstruction?

Reconstruction pipeline

- Data Acquisition
  - Leginon
- Particle picking
  - Selexon
- CTF estimation
  - ACE
- Selecting "good" data
  - Database queries
  - ???
- Reconstruction

Background

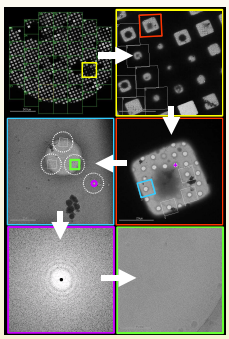
- GroEL has been our driver for developing both automated data collection and automated data analysis
- 150,000 particles/24 hours a year ago
- Over the last year, led to the development of
  - Environment monitoring
  - Database reports
  - Training data for ACE
  - Optimize protocols for single particle reconstruction with EMAN and Frealign
  - Creation of JAHCs grids

Data Acquisition

Automated data acquisition with Leginon

Automated microscope control

Multiscale Imaging



Suloway *et al.* (2005) J. Struct. Biol., In press.

How long does it take?

- Setup
  - 1 h on a good day - 5 h on a bad day
    - Stability of microscope/problems with specimen
- Acquisition
  - Creating the atlas
    - 15 min
  - Finding holes
    - ~30s for square image
    - < 1s for hole image
  - Focusing
    - 10s for algorithm + 5-30s for melting ice
  - Reading and correcting the high-resolution exposures
    - ~30s / exposure

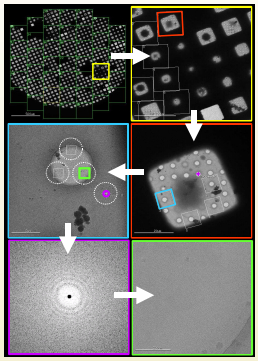


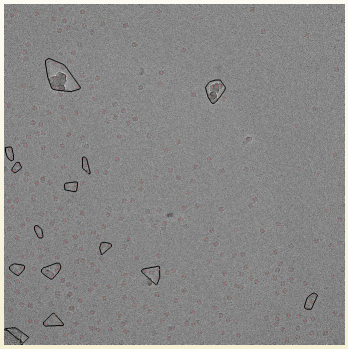
Image collection statistics

- Defocus pairs: 552
  - 50,000X, 2.263 Å/pix, -0.8 to -2.0 µm defocus
  - Hundreds of particles per image
- Focus images: 273
  - 50,000X
- Holes visited: 318
  - 5000X, 179 Å/pix, -150 µm defocus
- Squares visited: 32
  - 800X, 558 Å/pix, -2mm defocus
- Total time: 25h

Picking particles

Automated particle picking

Selexon

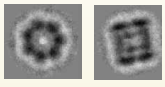


~95% accurate

280,000 particles picked

Roseman (2004) JSB, 145
Zhu *et al.* (2004) IEEE ISBI04 conference

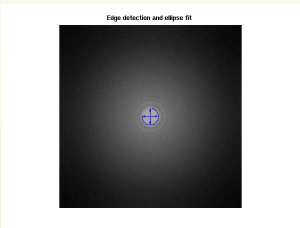
How long does it take?

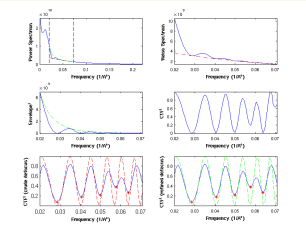
- Setup
  - Creating templates ⟷ 
    - 1-2 hours
  - Setting parameters
    - 30 min
- Automated particle picking
  - ~2 min/micrograph

CTF Estimation

Automated CTF estimation

ACE



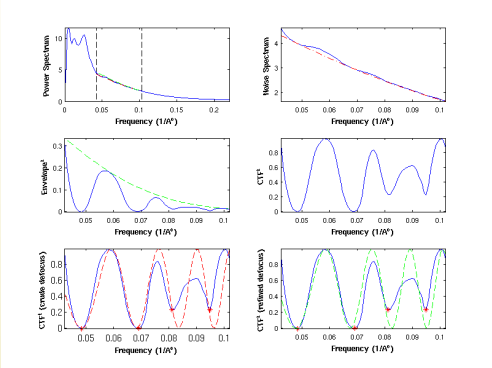


Mallick *et al.* (2005) Ultramicroscopy,104

ACEMAN

- Reads Imagic stacks instead of entire micrographs
- Uses EMAN formulation for noise and envelope
- So far does not include structure factors
  - Structure factors should be implemented w/i a month

ACEMAN



How long does it take?

- Setup
  - 1 minute
- Automated CTF estimation
  - ~1 minute/micrograph
  - Slightly faster with ACEMAN

Database reports

http://cronus3.scripps.edu/dbem/summary.php?expId=1933

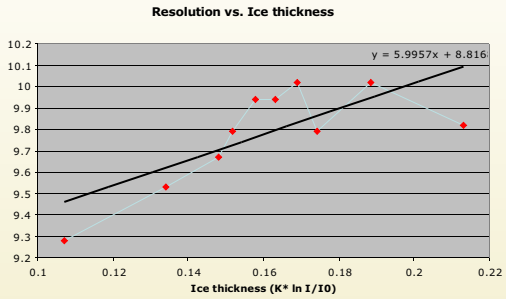
The bottom line: How do these parameters affect the reconstruction?

- Can we sort the data in such a way that we focus only on “good” particles?
  - Sort by ice thickness
  - Sort by ACE data
  - Sort by drift
  - Sort by temperature
  - ???

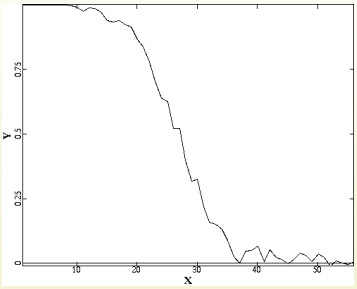
Sorting particles by ice thickness

- Sorting scheme
  - Throw away any micrograph with ACE confidence value < 0.8 (manually verified that all fits > 0.8 are correct)
  - Take defocus measurements from ACE and sort micrographs into small (0.5-1.0), medium (1.0-1.5), and large (1.5-3.0) defocus sets
  - Sort defocus sets and split into 10 subsets by increasing ice thickness
  - Find set with least ptcls and randomly remove ptcls from other sets until all have same # ptcls (~15,800)
- Result is 10 sets of particles with equivalent range of defoci
- Reconstruct each set using EMAN

Resolution decreases with increasing ice thickness

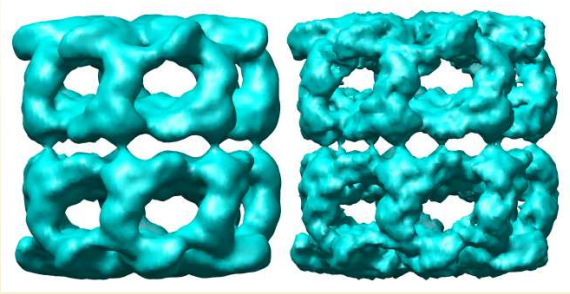


FSC of highest resolution structure



Resolution = 9.3Å

The structure of GroEL



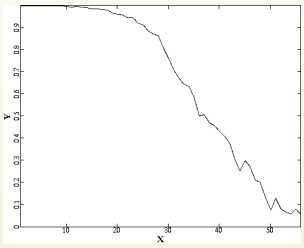
Thinnest ice structure

Amplitude corrected via Spider

Sorting particles by ice thickness - amp. corrected

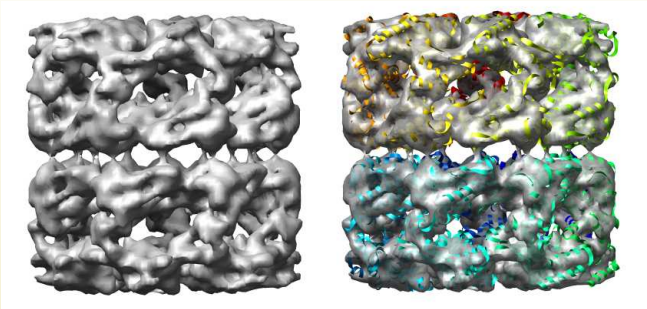
- Sorting scheme
  - Use ACEMAN to estimate noise and envelope, but use original ACE estimation for defocus
  - Throw away any micrograph with ACE confidence value < 0.8
  - Take defocus measurements from ACE and sort micrographs into small (0.5-1.0), medium (1.0-1.5), and large (1.5-3.0) defocus sets
  - Sort defocus sets and split into 10 subsets by increasing ice thickness
  - Find set with least ptcls and randomly remove ptcls from other sets until all have same # ptcls (~15,800)
- Result is 10 sets of particles with equivalent range of defoci
- Reconstruct each set using EMAN
  - Apply envelope correction to class averages towards the end of the refinement

FSC of thinnest ice



Resolution = 6.5Å
Nyquist = 4.526Å

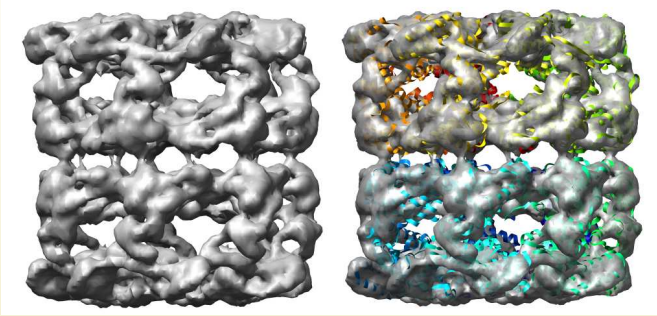
GroEL at 6.5Å?



Can we get even higher resolution?

- Refine with all 280,000 ptcls
- Average volumes from multiple reconstructions
- What do we do about amplitudes?
- What is the resolution?!!!

Average of all volumes

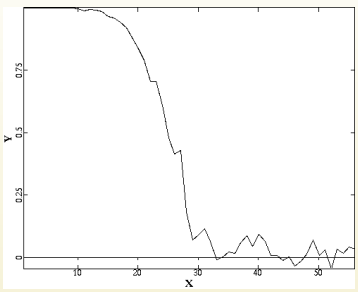


Volume was amplitude corrected via Spider

Average of all volumes

QuickTime™ and a
H.264 decompressor
are needed to see this picture.

What is the resolution?

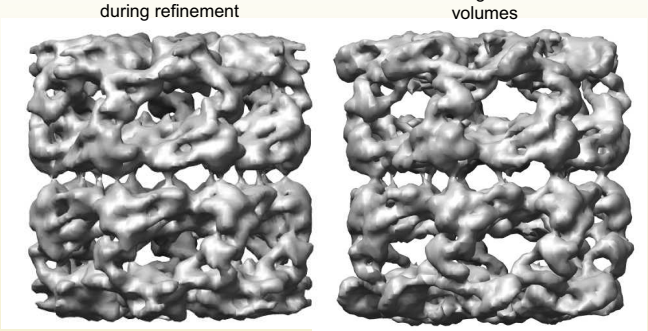


Resolution ($FSC_{0.5}$) = 10.2Å

Comparison with 6.5Å

Amplitude corrected during refinement

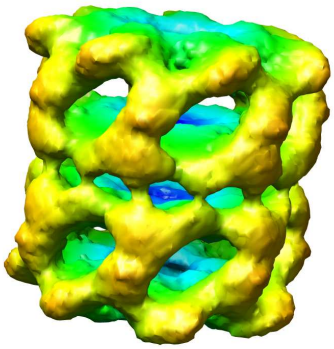
Average of 10 volumes



6.5Å?

10.2Å?

The pipeline in action



Acknowledgments

- Leginon
  - Denis Fellman
  - Jim Pulokas
  - Christian Suloway
  - Joel Quispe
  - Anchi Cheng
- ACE
  - Satya Mallick
- Selexon
  - Yuanxin Zhu
  - Alan Roseman