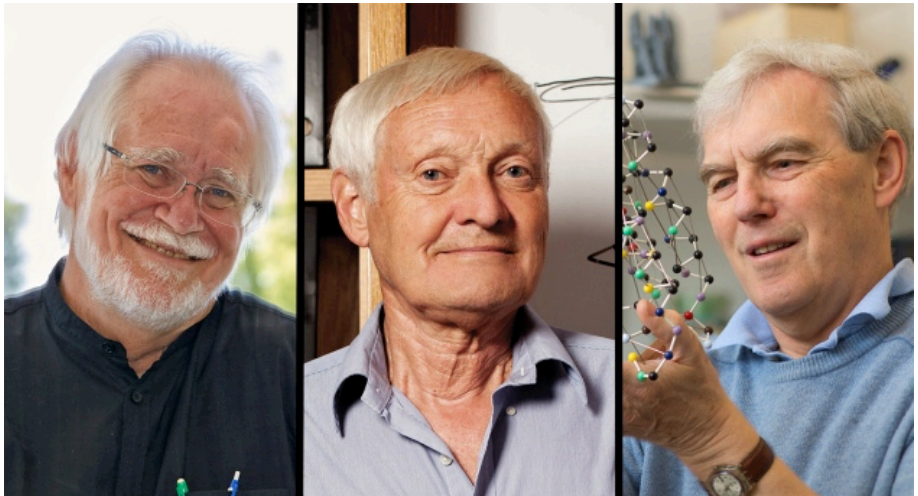
Single-Particle Cryo-Electron Microscopy

Robbie Ostrow, Trevor Tsue and Shalom Rottman-Yang

What is Cryo-EM?

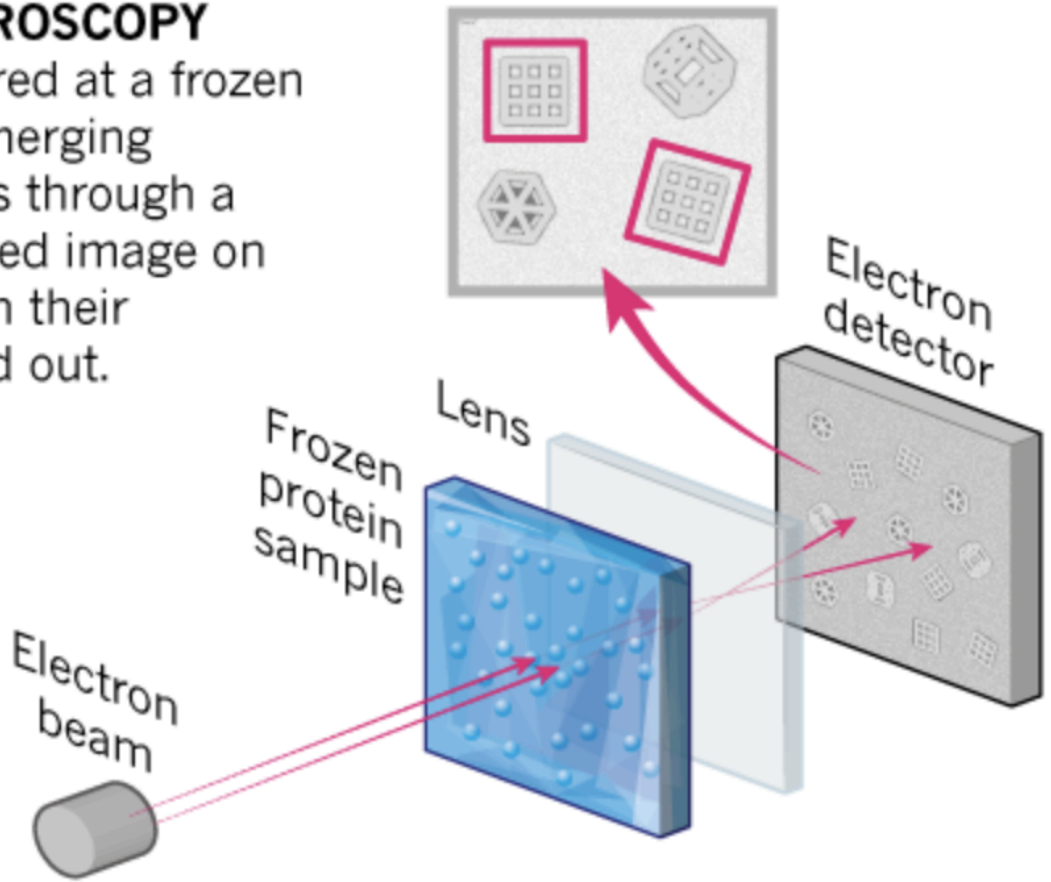
- Finds 3-D structure of molecules
- Developed in the 1970s, but massive development in the last few years
  - Better cameras and more processing power
- 2017 Nobel Prize!



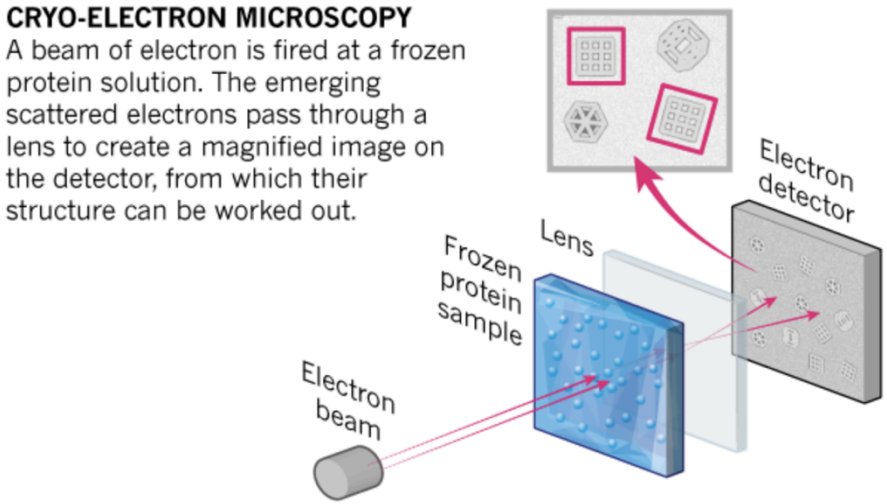
nature.com

CRYO-ELECTRON MICROSCOPY

A beam of electron is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, from which their structure can be worked out.



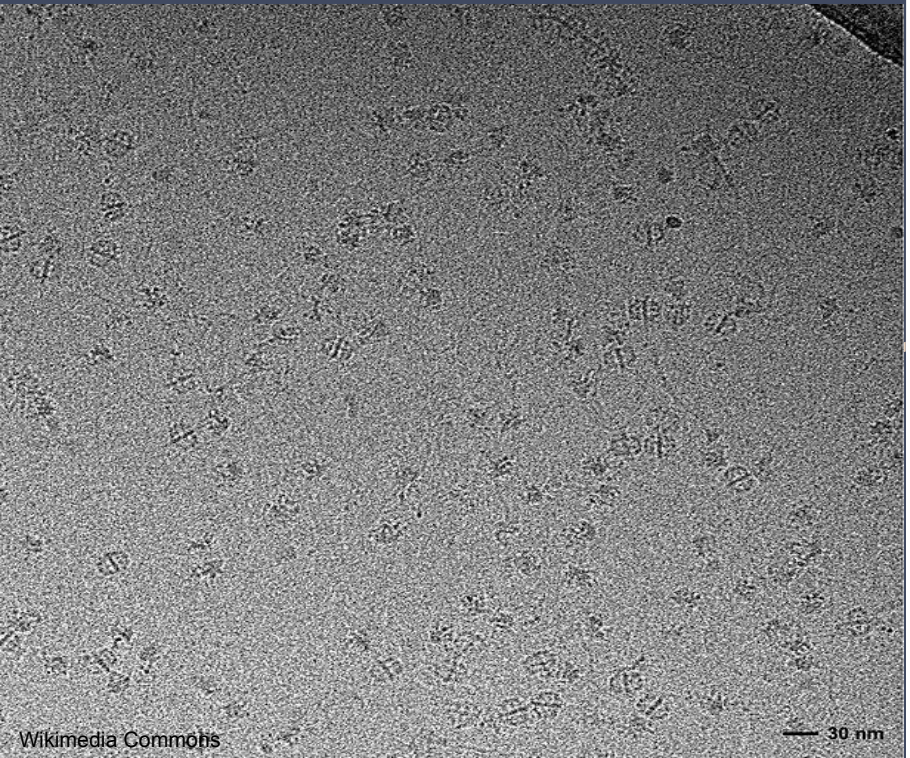
How does Cryo-EM work?



1. Protein Purification and specimen preparation
2. Take a 2-D image (or a series of images) with an electron microscope
3. Pick out particles
4. Classify and align
5. 3-D reconstruction
6. Refine and validate
7. Done! (Maybe?)

Difficulties of Cryo-EM

- Every one of these steps has its own set of challenges.



Wikimedia Commons

Image of C. thermophilum lysate

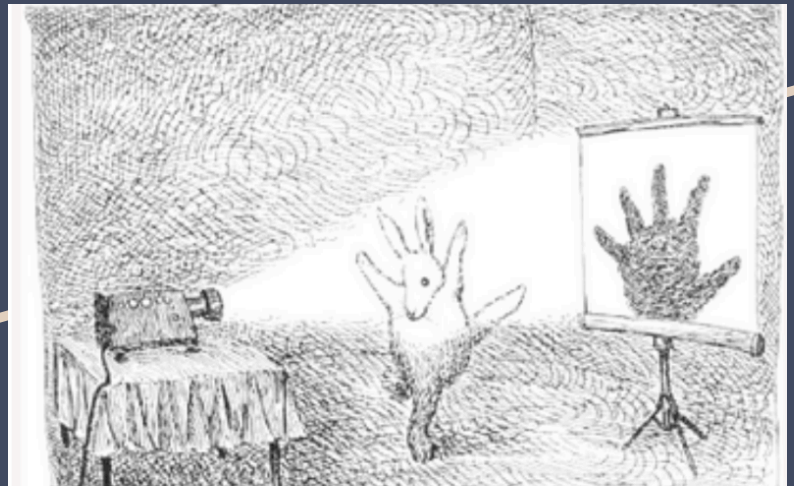
Difficulties of Cryo-EM

1. Protein Purification and specimen preparation

- Can be tricky, depending on the particle.
- Vacuum dries out particles
- Electrons damage unprotected particles
- Straining/vitrification are the most common techniques.

Difficulties of Cryo-EM

2. Take a 2-D image (or a series of images) with an electron microscope

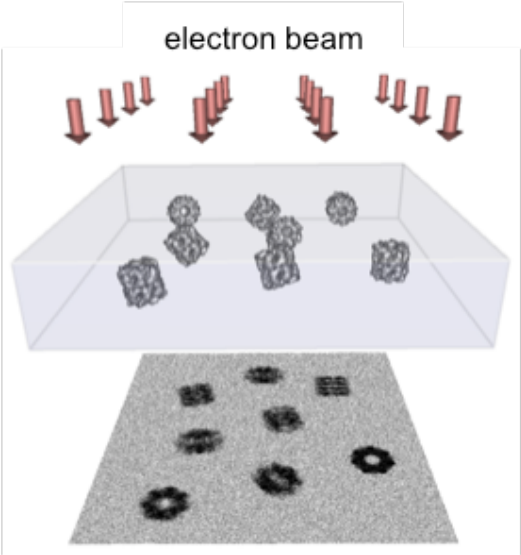


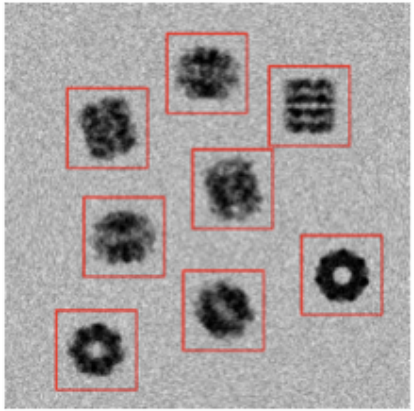
- Contrast very low
  - Images need to be taken out of focus!
- Microscopes have to be calibrated extraordinarily well
- Lots of noise
- Each image is essentially a noisy 2-D shadow at a random angle

Difficulties of Cryo-EM

3. Pick out particles

- Too many to do by hand, but often lacking a good model
- RELION



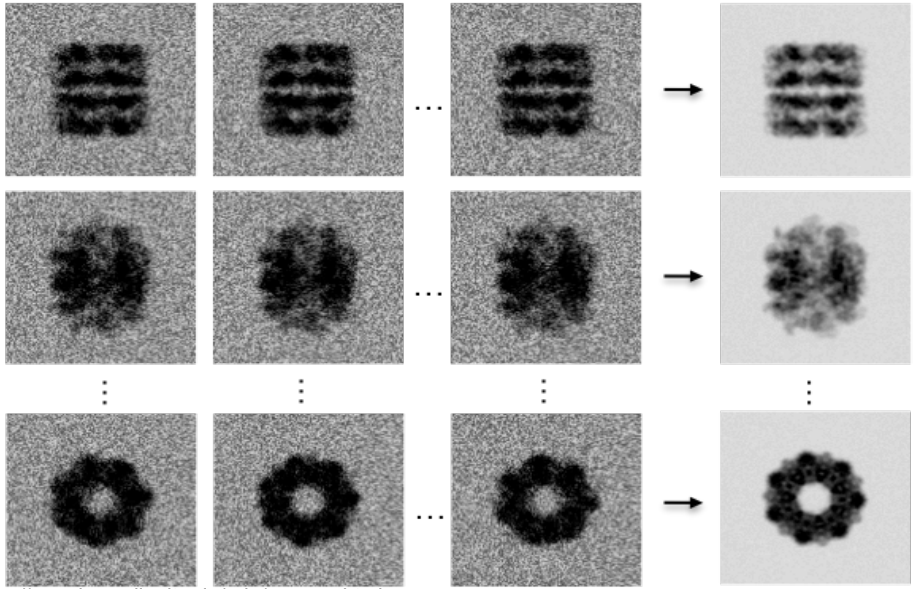


http://people.csail.mit.edu/gdp/cryoem.html

Difficulties of Cryo-EM

4. Classify and align

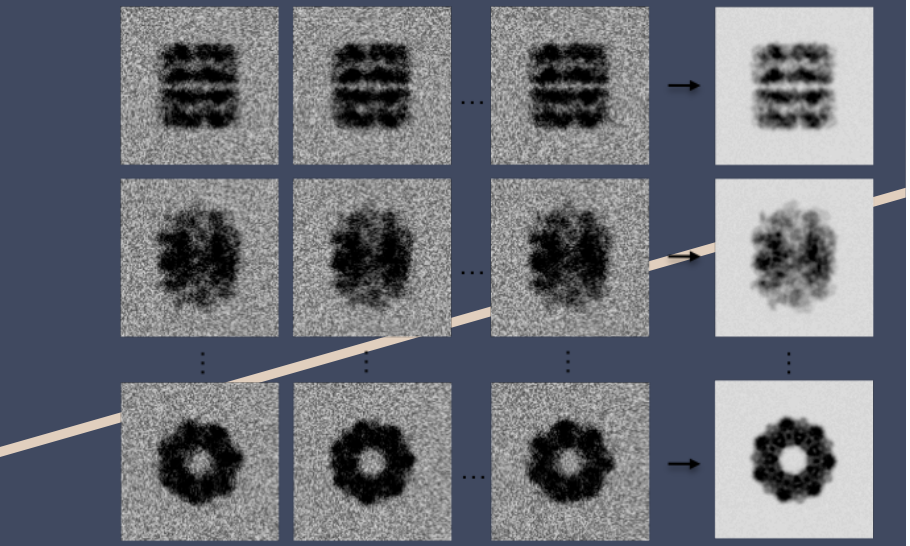
- 2-D images need to be clustered to create 3-D reconstruction
  - But 3-D reconstruction needed to get the right clustering!



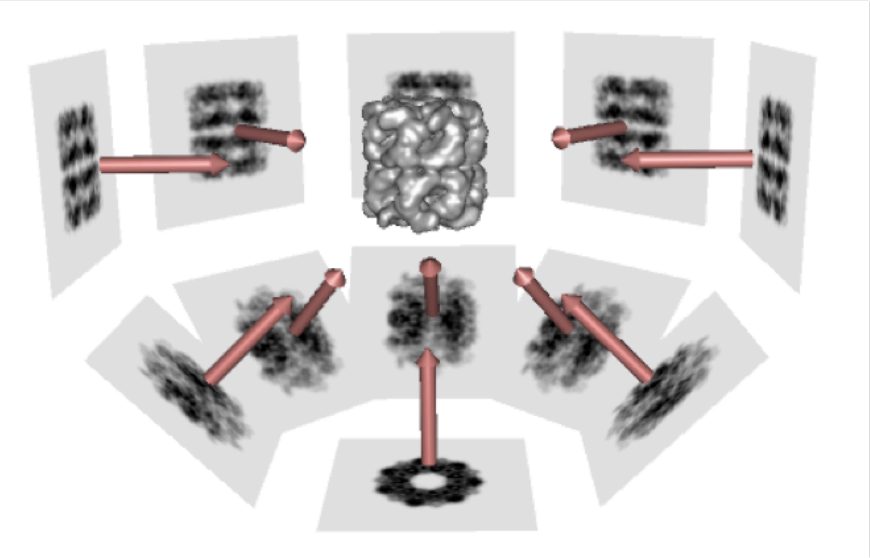
http://people.csail.mit.edu/gdp/cryoem.html

Difficulties of Cryo-EM

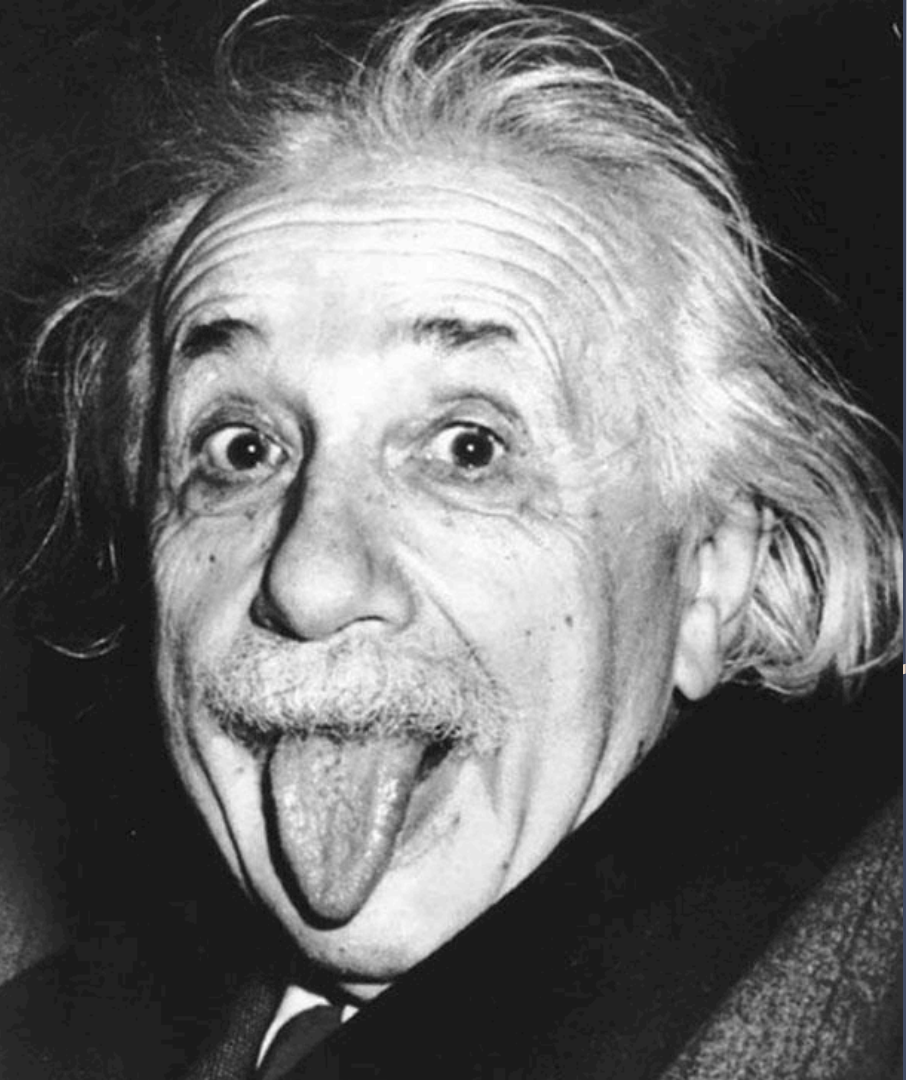
4. 3-D Reconstruction

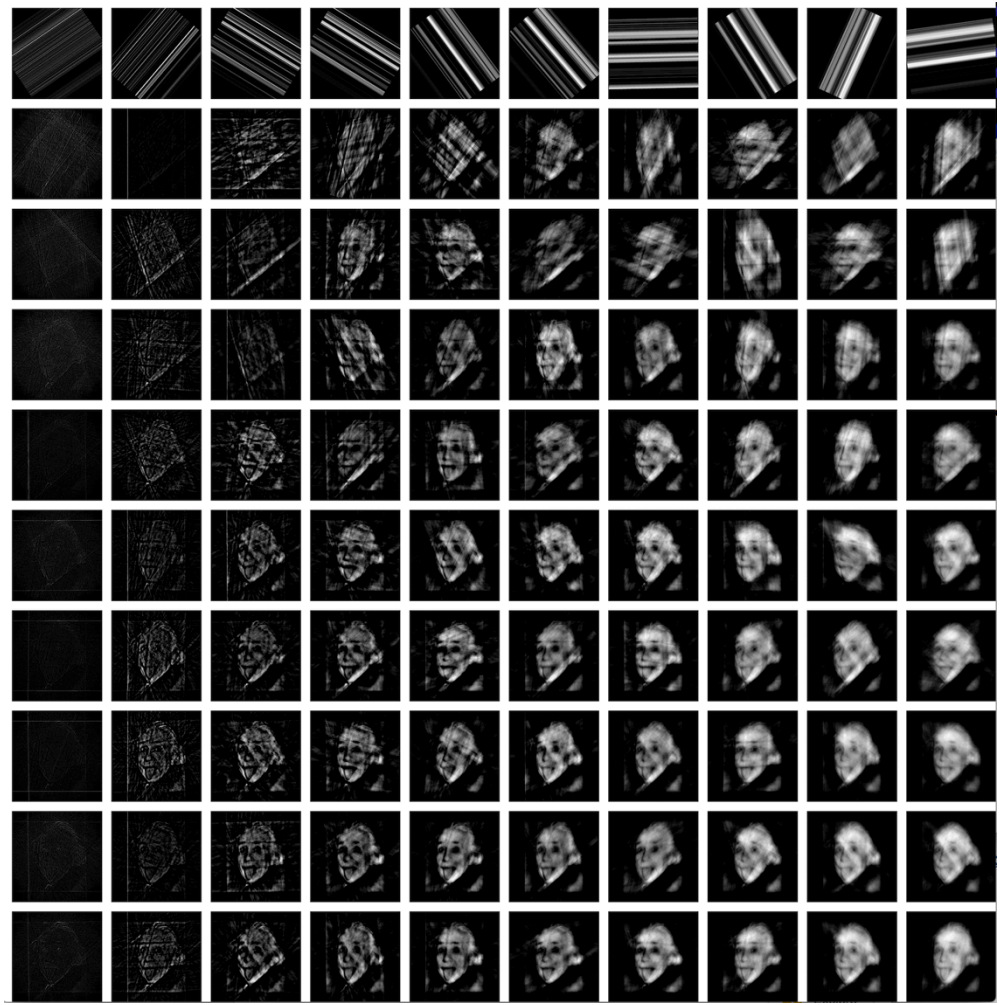


- Combine 2-D projections
- Filtered back-projection



http://people.csail.mit.edu/gdp/cryoem.html





Advantages of Cryo-EM

- Despite all of the difficulties, still often easier (and much cheaper) then crystallography
  - Especially for large particles
- Crystallography can change the conformation of particles

Improvements being made on all of these steps.

1. Protein Purification and specimen preparation
    a. Minimize heterogeneity (and classify)
2. Take a 2-D image (or a series of images) with an electron microscope
3. Pick out particles
4. Classify and align
5. 3-D reconstruction
6. Refine and validate
7. Done! (Maybe?)

Trajectories of the ribosome as a Brownian nanomachine

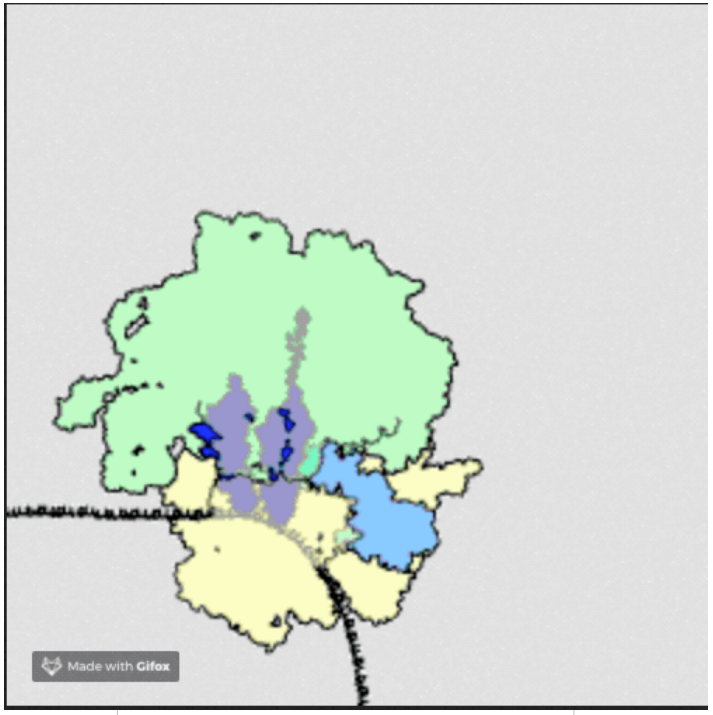
Dashti et. al.

How does Cryo-EM work?

1. Protein Purification and specimen preparation
    a. Minimize heterogeneity (and classify)
2. Take a 2-D image (or a series of images) with an electron microscope
3. Pick out particles
4. Classify and align
5. 3-D reconstruction
6. Refine and validate
7. Done! (Maybe not?)

The ribosome as a Brownian machine

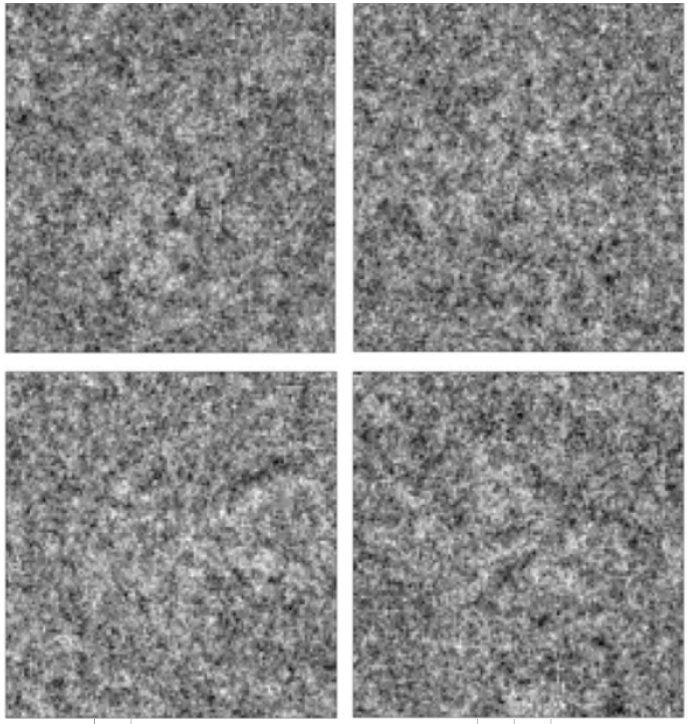
- Exploits random motions of the molecules in its environment to do work.
- Ribosome is widely regarded as a prototypical machine.
- But almost every protein acts as a Brownian nanomachine.



Wikimedia Commons

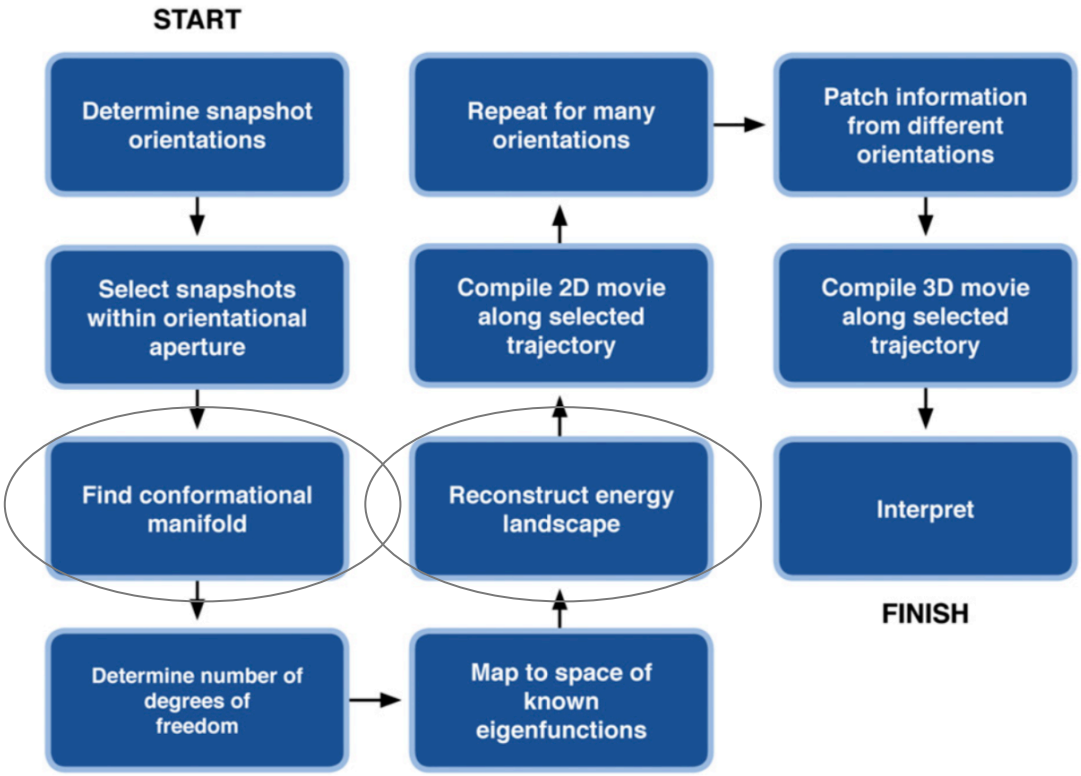
Sample preparation

- Yeast 80S Ribosome
- 849,914 images from ~4700 micrographs
- No (or very little) mRNA or tRNA



Dashti et. al.

Algorithm

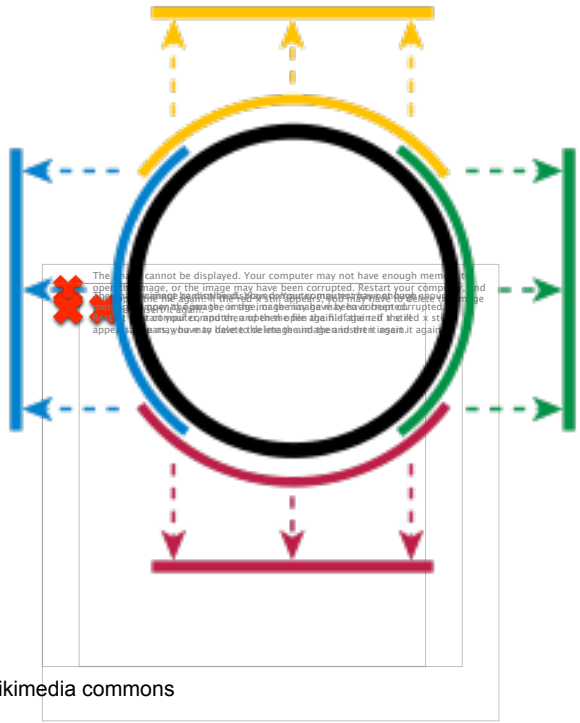

START
Determine snapshot orientations
Select snapshots within orientational aperture
Find conformational manifold
Determine number of degrees of freedom
Map to space of known eigenfunctions
Reconstruct energy landscape
Compile 2D movie along selected trajectory
Repeat for many orientations
Patch information from different orientations
Compile 3D movie along selected trajectory
Interpret
FINISH

Manifolds

- Space that is locally homeomorphic to Euclidean space.

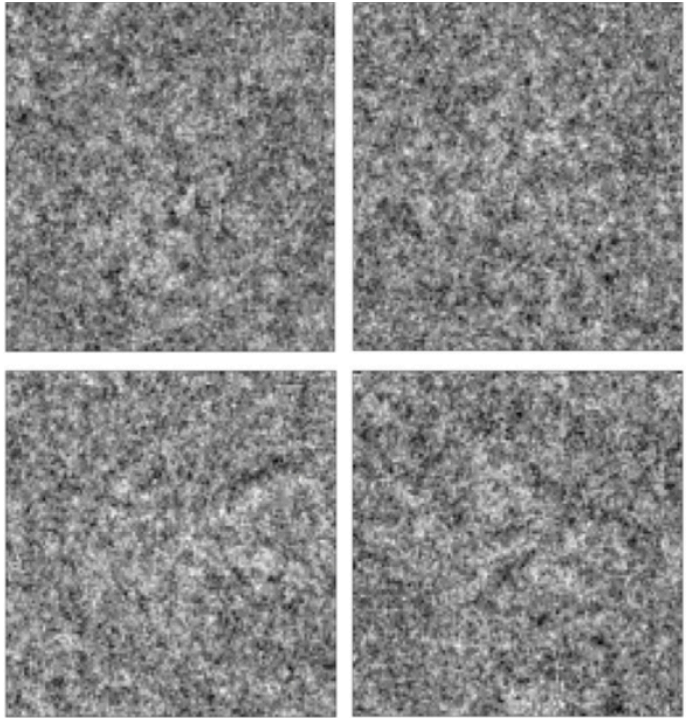
Formally,

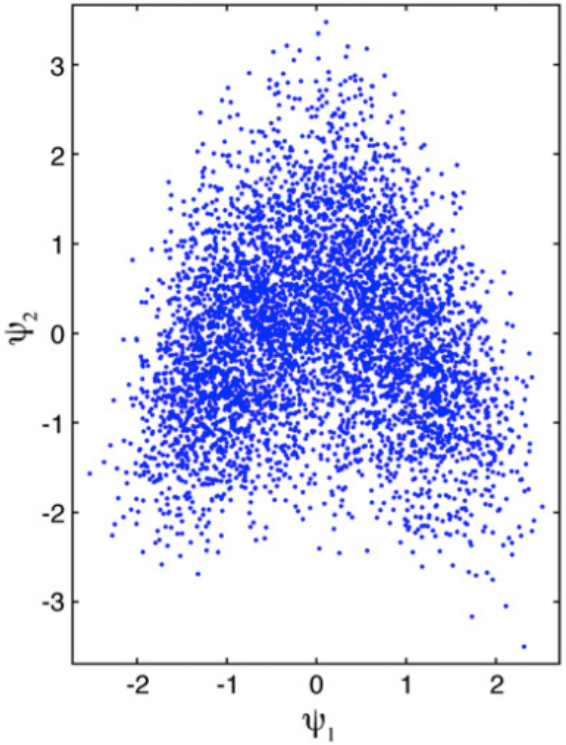
$$\mathbf{B}^n = \left\{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + x_2^2 + \cdots + x_n^2 < 1\right\}$$



Wikimedia commons

Conformational manifolds

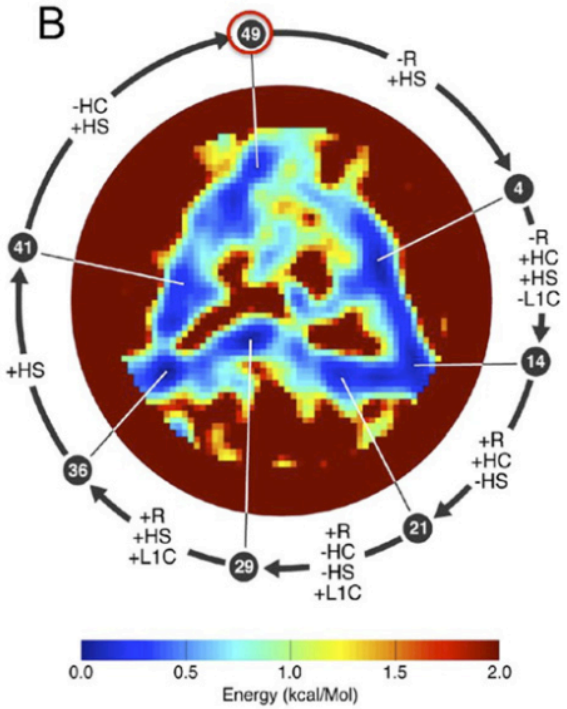




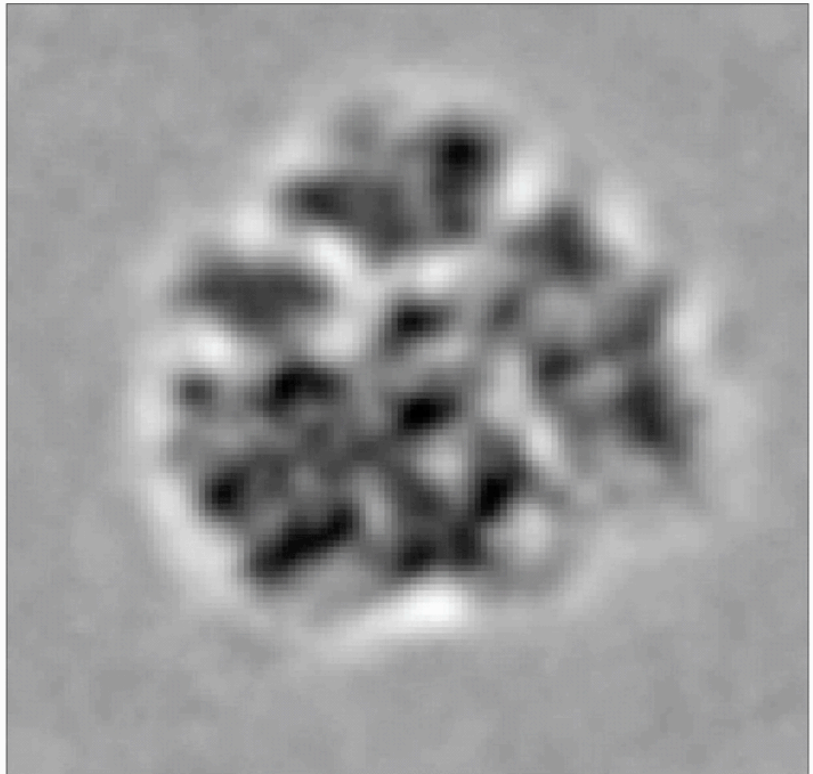
Dashti et. al.

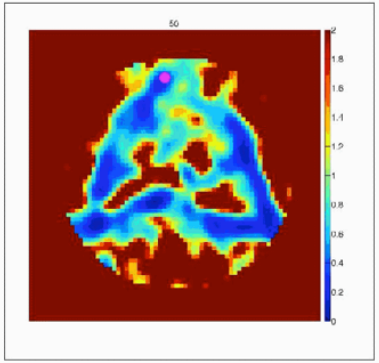
Energy landscape

- Reconstructed from relative proportions of samples in micrographs
- Susceptible to bias?

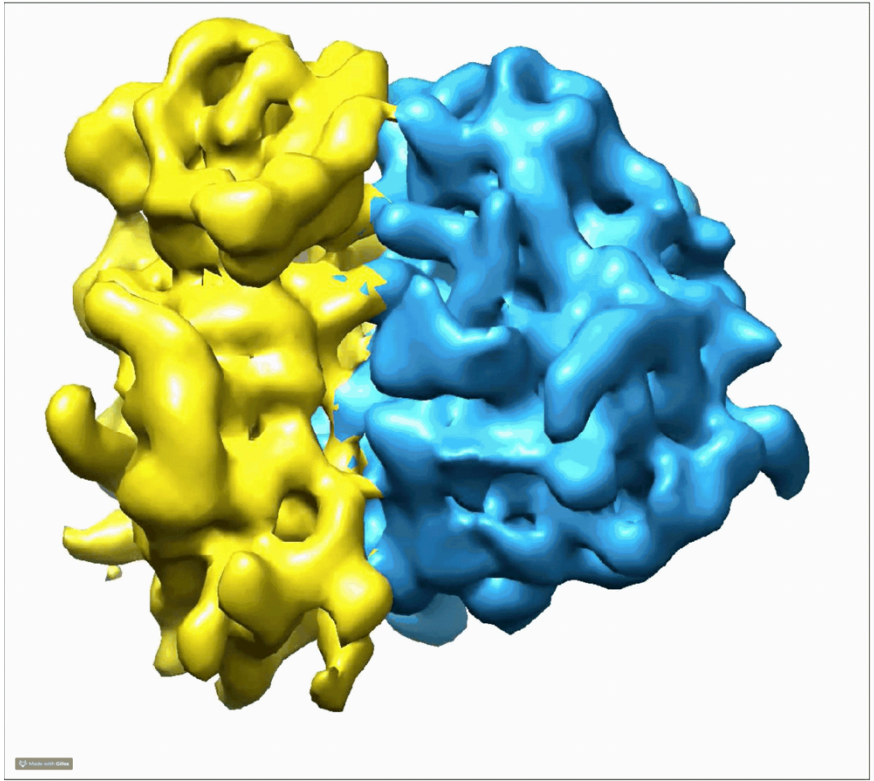


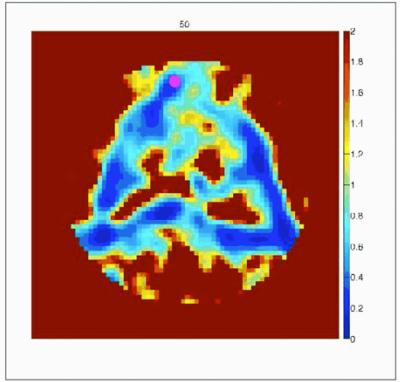
2D Movie



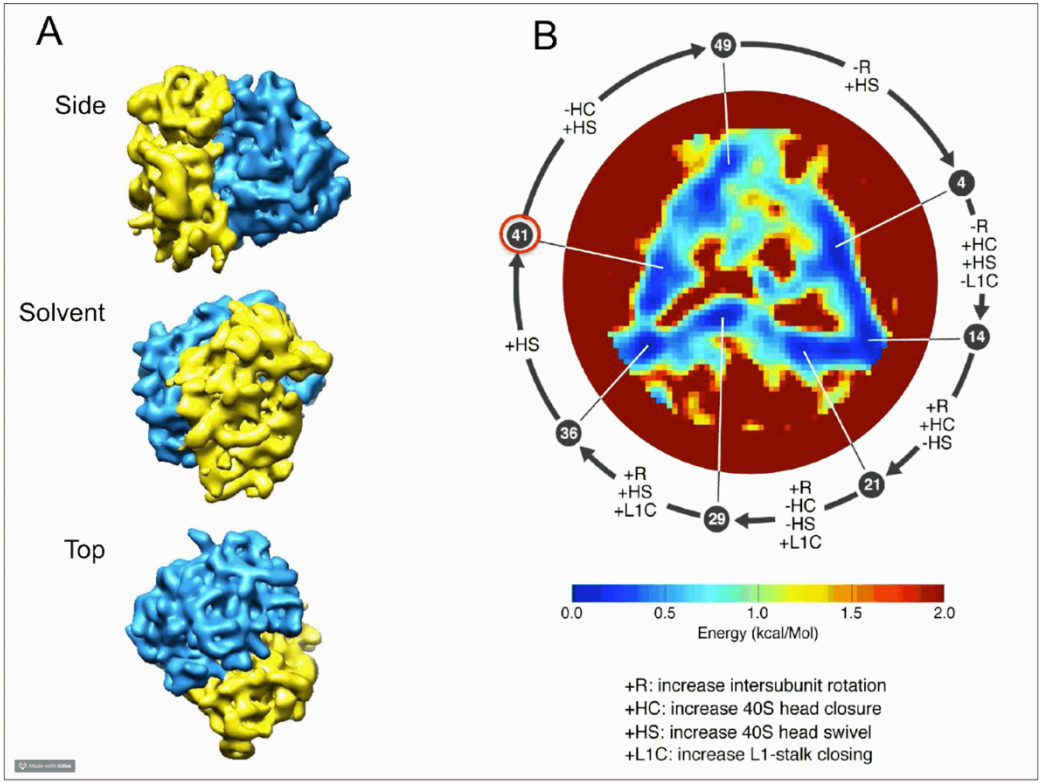


3D Movie





Results


A
Side
Solvent
Top
B
49
-R
+HS
4
-R
+HC
+HS
-L1C
14
+R
+HC
-HS
21
+R
-HC
-HS
+L1C
29
+R
+HS
+L1C
36
+HS
41
-HC
+HS
0.0
0.5
1.0
1.5
2.0
Energy (kcal/Mol)
+R: increase intersubunit rotation
+HC: increase 40S head closure
+HS: increase 40S head swivel
+L1C: increase L1-stalk closing

Strengths

- Split cryo-EM structures into 50 classes, rather than the usual 5.
- Allows a movie to be made based on inferred free-energy
- Easily extensible to other types of particles

Limitations

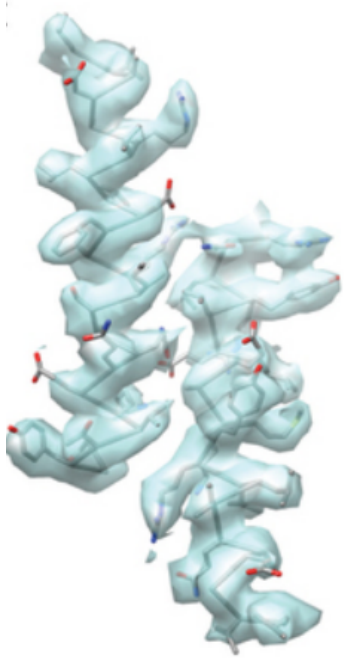
- Non-translating ribosomes might not traverse the same paths.
- No way to confirm correctness except “looks like it makes sense”
- How useful is a composite movie?
- Movie is based on close-ness, not time. (can’t distinguish between forwards and backwards time)
- Some ribosomes were selected by hand

Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta

Wang R.Y., Song Y., Barad B.A., Cheng Y., Fraser J.S., DiMaio F.

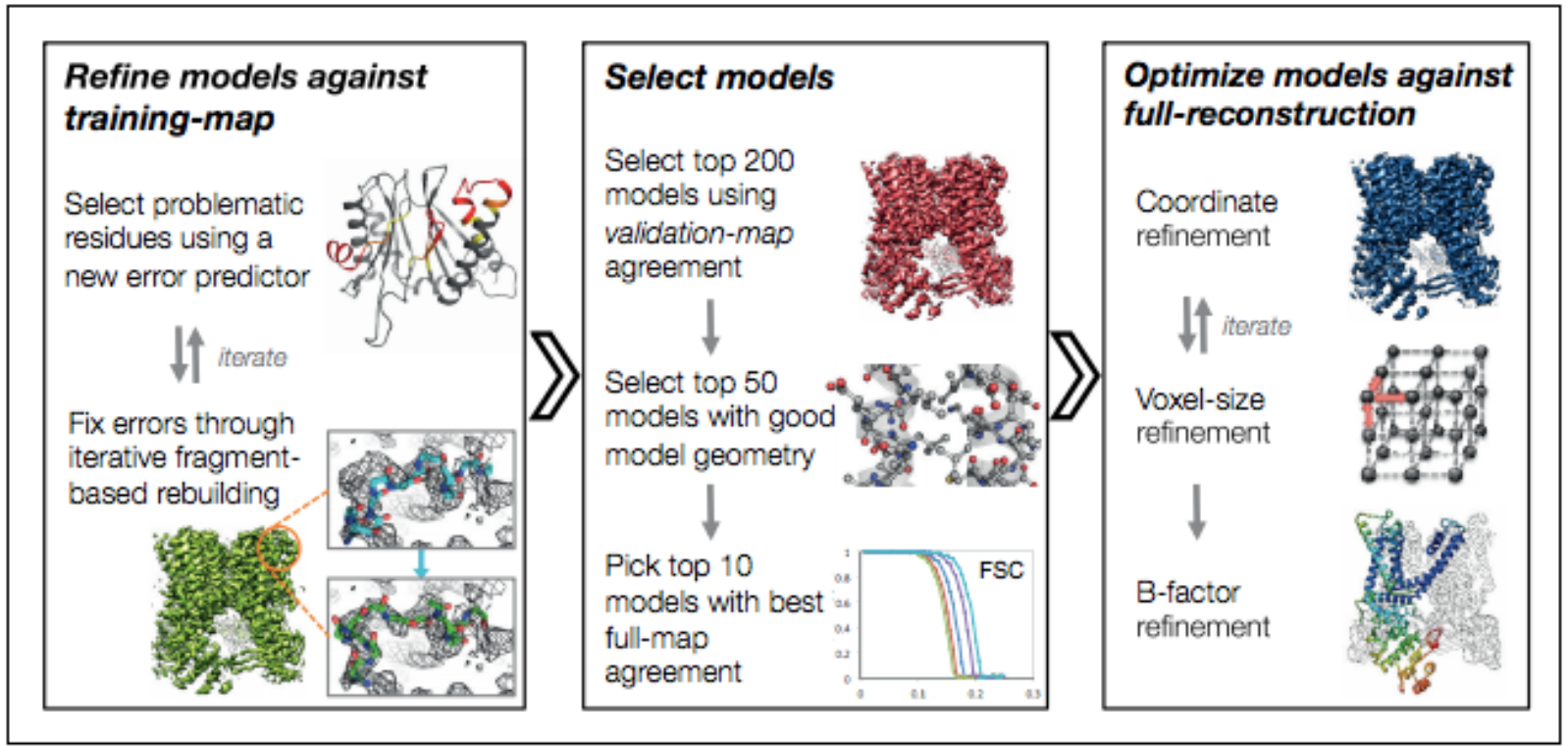
Background

- Cryo-EM can provide near-atomic resolution
- All-atom models can be built from density maps given by cryo-EM
  - Atom coordinates cannot be assigned precisely
  - Some molecular interactions may not be captured
- Currently, usually build model into the density map manually



Automatic refinement in this paper

- 3-stage approach to automatically refine manually-traced cryo-EM models

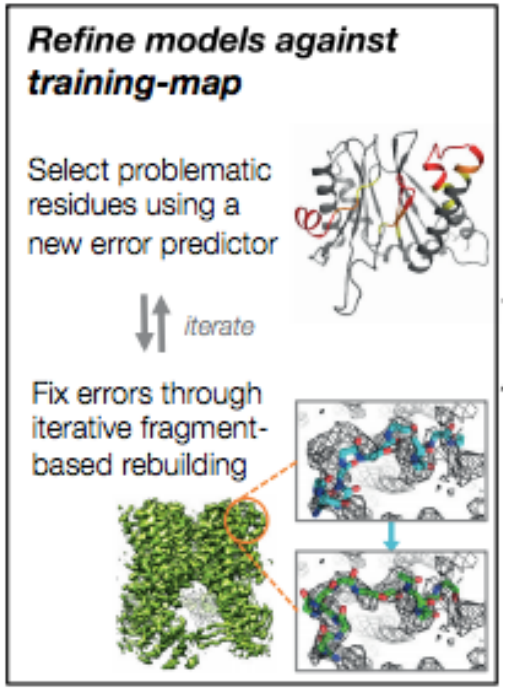


Versus previous work

- Had previously created a tool for local rebuilding for refining homology models
- Improvements in this approach allow for correcting significant backbone errors

Stage 1a: Model refinement using training map

- **Takeaway**: Relax the structure and then choose the worst fitting residues

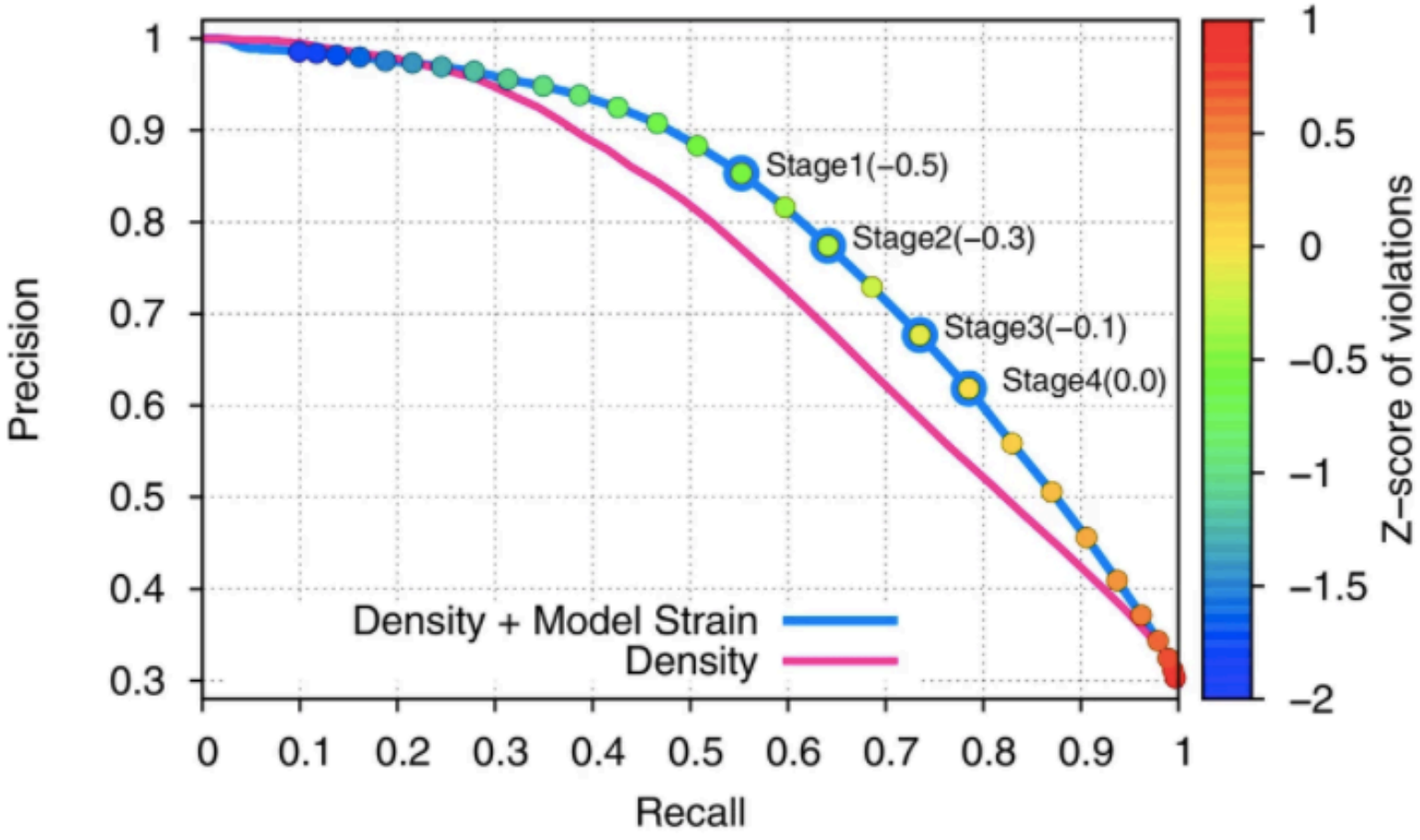


- Training map is a 'half-map' – a full 3d density map created using only half of the cryo-EM data
- Run Rosetta relax (wiggles sidechains to trigger local strain)
- Choose the worst residues:

$$Z_{error}^{(i)} = w_{dens} \cdot Z_{dens}^{(i)} + w_{lcldens} \cdot Z_{lcldens}^{(i)} + w_{bonded} \cdot Z_{bonded}^{(i)} + w_{rama} \cdot Z_{rama}^{(i)}$$

  - learn weights from known structure dataset
  - fit-to-density is measured by real-space correlation coefficient

1a: Why include model strain?



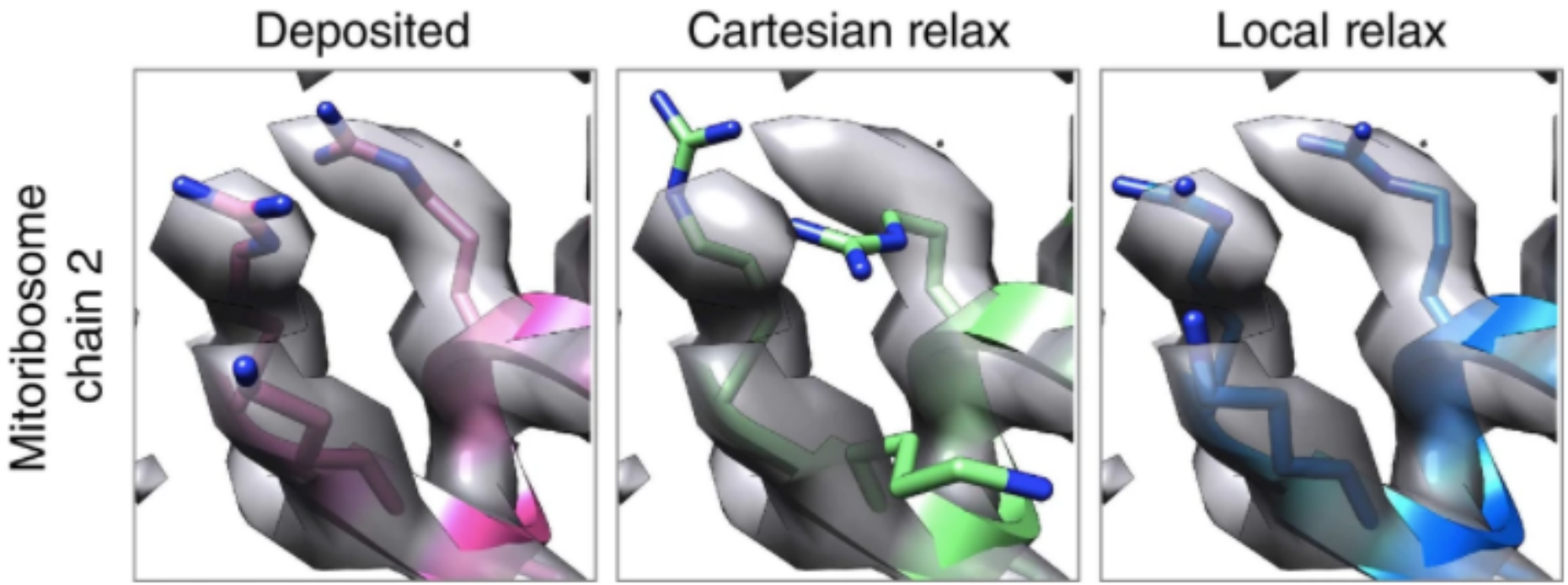
Stage 1b: Iterative fragment-based rebuilding

- **Takeaway**: Rebuild fragments of the model with Monte Carlo sampling

- Choose a 'bad' residue
- Choose a set of known backbone conformations based on local sequence
- Run Monte Carlo (randomized, small-step) optimizations using energy functions and fit-to-density
- Take the best result using fit-to-density

- After iteration with 1a, run LocalRelax (repeatedly choose a residue with many nearby residues, and run relax on that neighborhood)

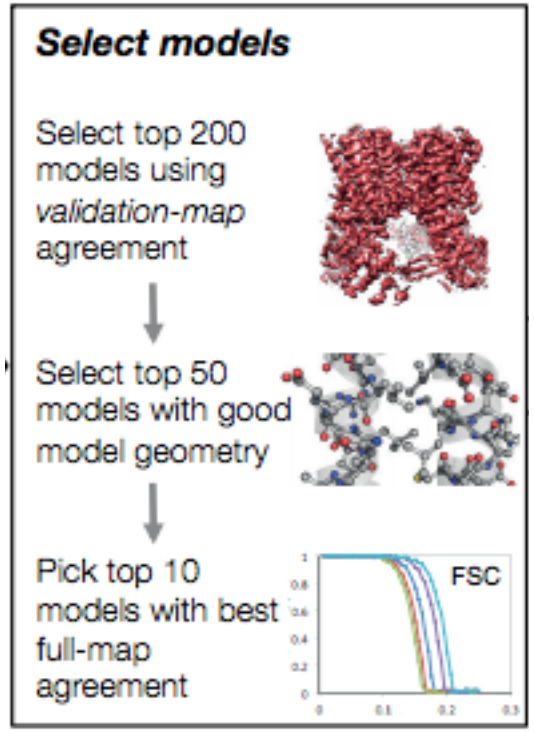
1b: Why optimize locally and not globally?



Stage 2: Model selection

- **Takeaway**: Select the best models using validation map and then full map

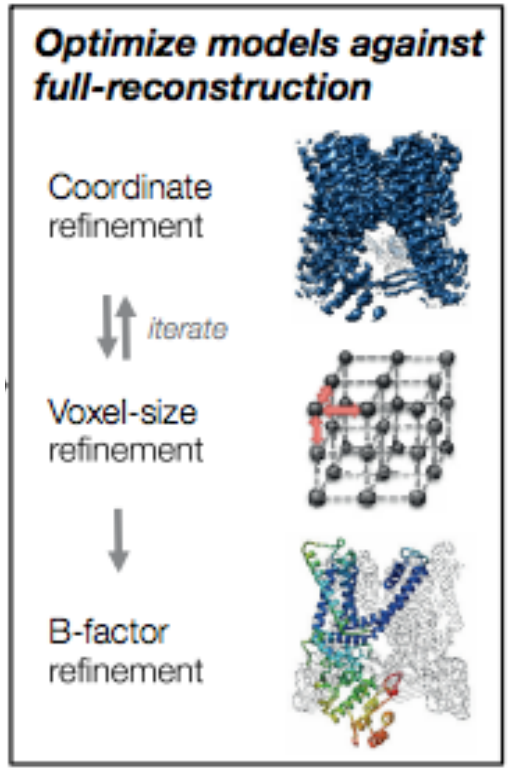
- Choose best models from stage 1 as according to:
  - Fit to validation 'half-map'
  - Model geometry (MolProbity score)
  - Fit to full map



Stage 3: Model optimization

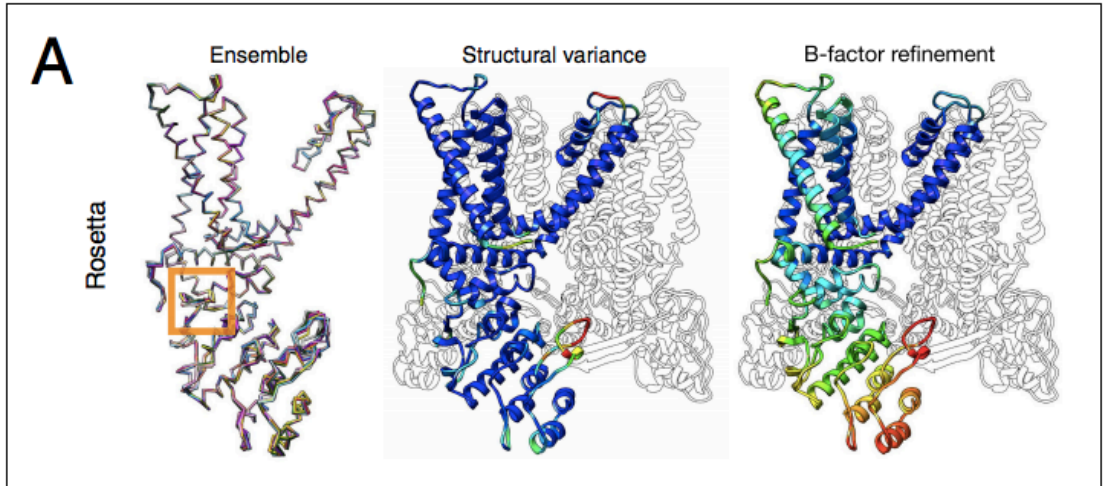
- **Takeaway**: Further refine selected models against full map without overfitting

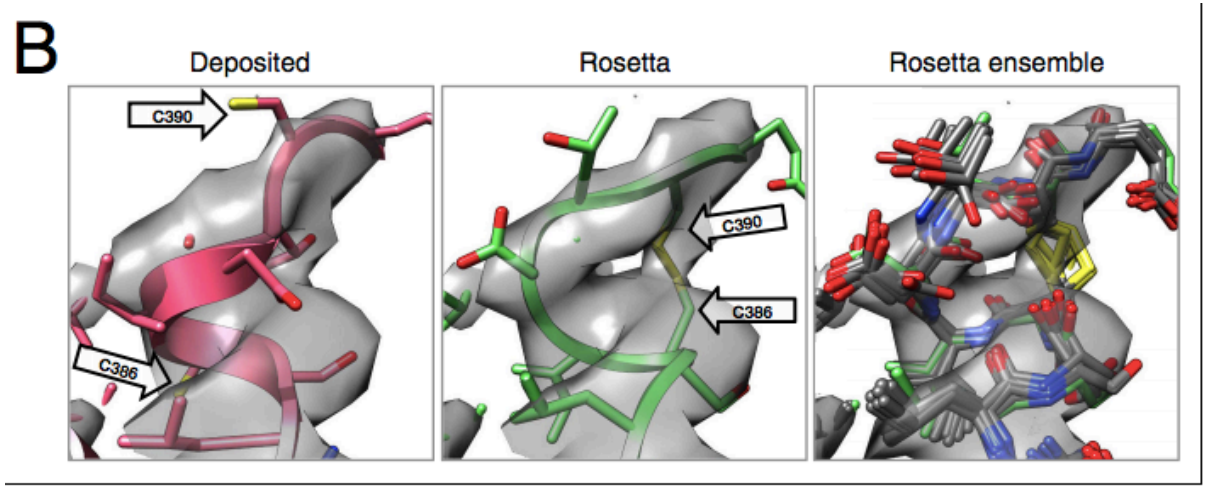
- Voxel-size refinement: optimize voxel size and origin of map density based on RSCC with experimental map
- Coordinate refinement: Rosetta LocalRelax with the full (not half-) map



Applying to 3 solved cryo-EM reconstructions: TRPV1

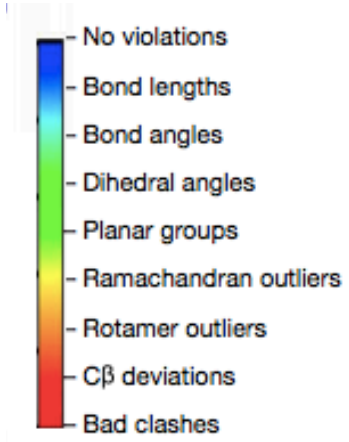
- Capsaicin receptor / vanilloid receptor 1
- Better MolProbity score (model geometry), slightly worse fit-to-density, and better EMRinger score (model-to-map backbone agreement)
- Found disulfide link not built in manual model

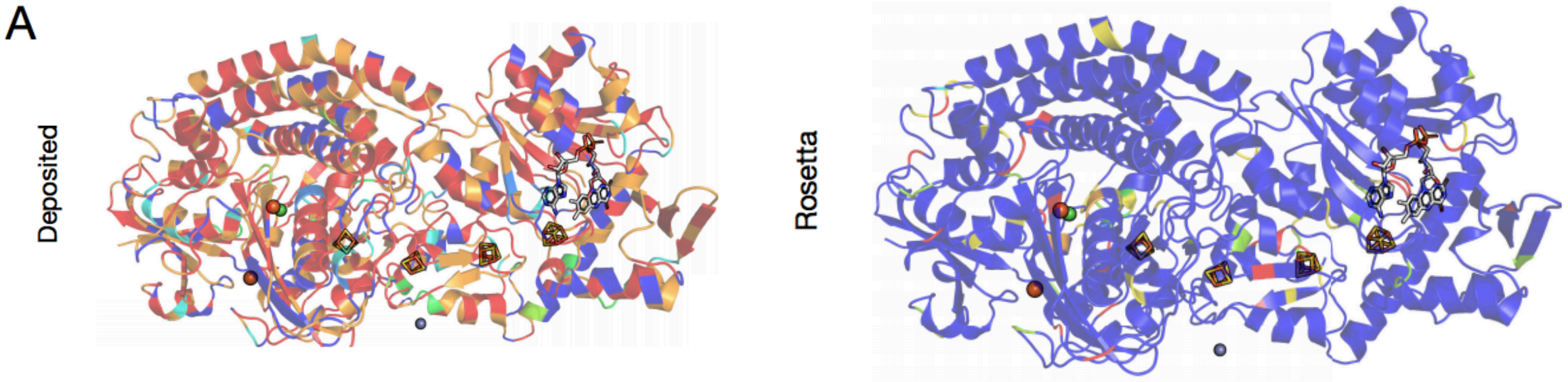




Applying to 3 solved cryo-EM reconstructions: F_{420}-reducing [NiFe] hydrogenase complex

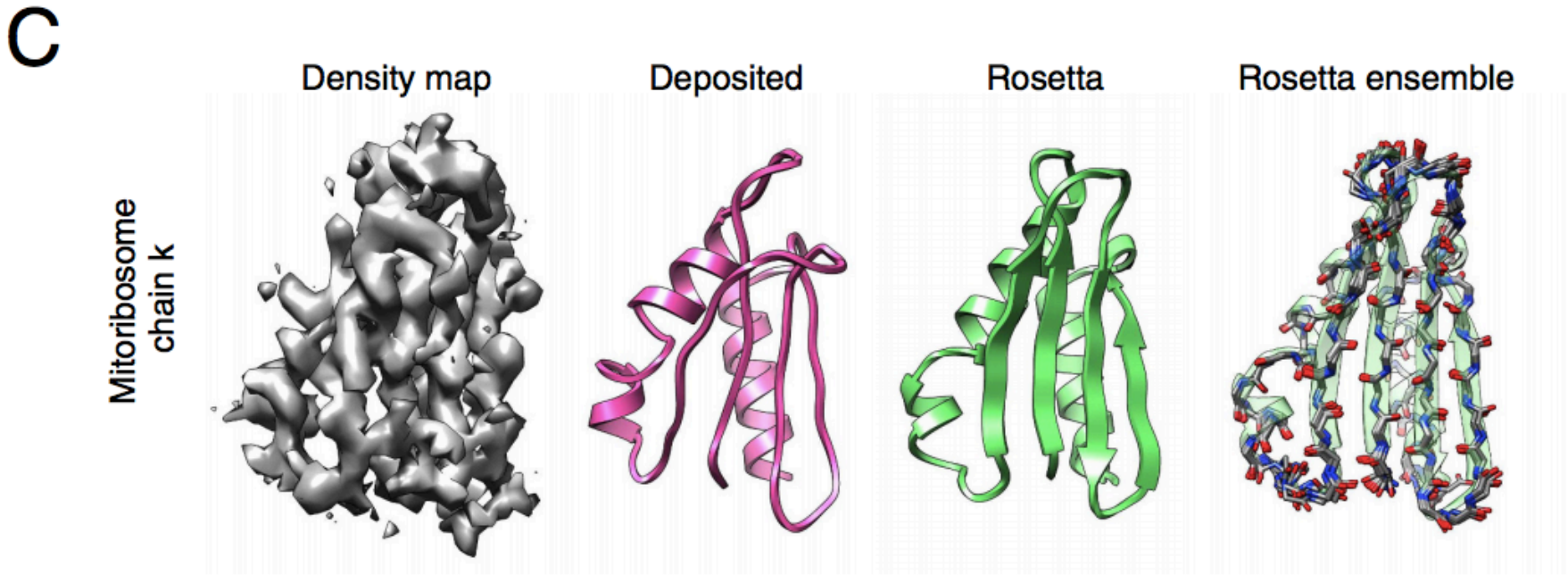
- Assembly of proteins with many covalently-bound ligands
- Better MolProbity and EMRinger scores, but worse fit-to-density





Applying to 3 solved cryo-EM reconstructions: mitochondrial ribosome large subunit

- 48 protein chains and 2 RNA chains
- Better MolProbity score on all protein chains due to better backbone geometry



Strengths

- Can handle backbone errors
- Uses physically based forcefield and known structures to ‘fill in’ information missing due to resolution
- Can avoid overfitting (lower fit-to-density but better model geometry)
- Not manual!

Limitations

Table 2. Comparison of structure refinement results between Rosetta and phenix.real_space_refine*.

	RSCC*,†,‡ validation map	iFSC*,†,§ validation map	EMRinger Score*,† validation map	MolProbity†				Number of residues with better RSCC†,¶
				Score	Clash score	Rotamer outliers [%]	Ramachandran favored [%]	
TRPV1	0.785 / 0.790	0.546 / 0.566	1.84 / 1.90	1.59 / 1.48	4.30 / 2.14	0.00 / 0.00	94.41 / 91.72	86 / 250
Frh	0.835 / 0.835	0.504 / 0.517	1.36 / 1.27	1.68 / 1.62	7.99 / 3.66	0.68 / 0.13	96.31 / 92.67	677 / 1328
Mitoribosome	0.832 / 0.832	0.476 / 0.478	2.05 / 1.98	1.88 / 1.62	6.17 / 4.08	0.38 / 0.00	90.19 / 93.49	415 / 564

- phenix used 0.24 CPU hours; Rosetta used 5000 CPU hours (5 hrs per trajectory)

Limitations

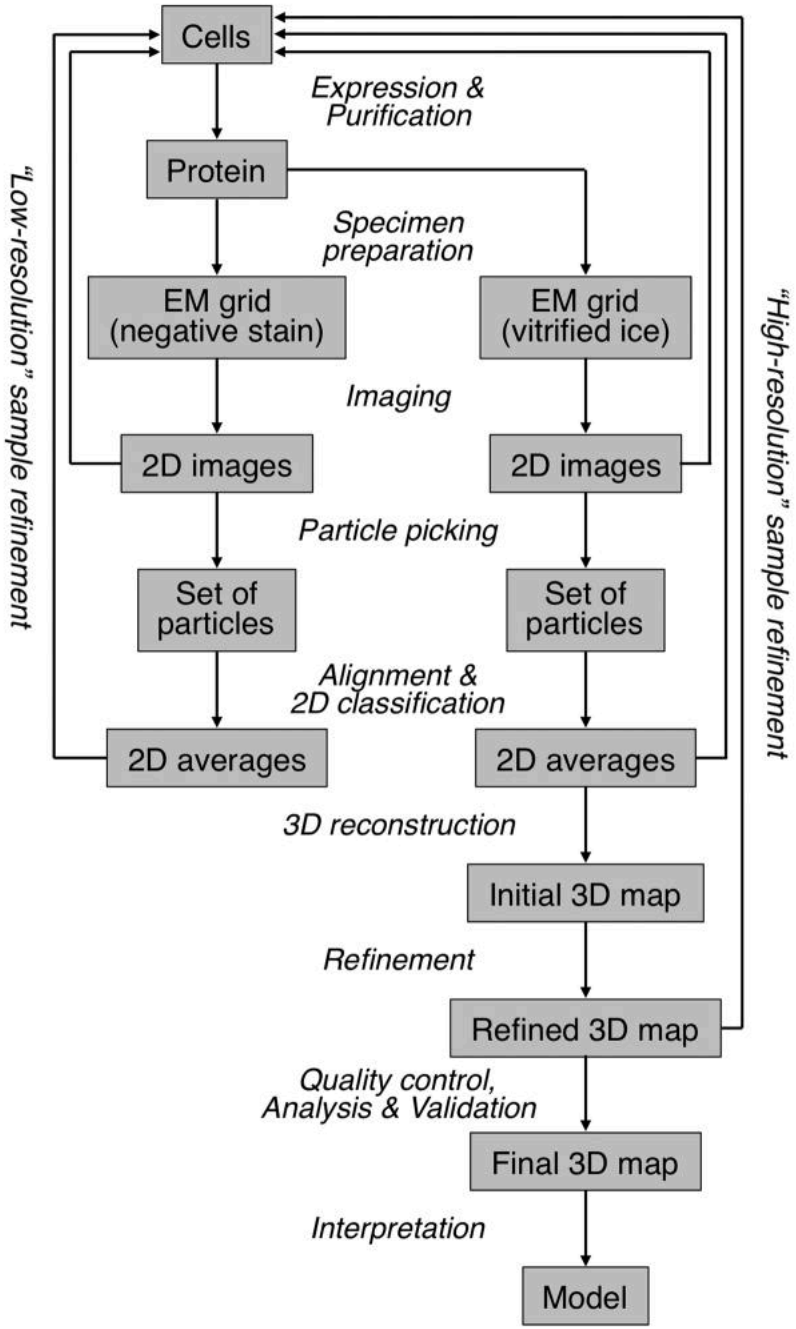
- Only looks at applicability to three structures; no wider-scale evaluation of performance
- Not especially elegant
- Still requires a manually traced model to start from

A Bayesian View on Cryo-EM Structure Determination

Sjors H. W. Scheres

2D to 3D

- 2D Reconstruction
  - Particle Alignment
  - Particle Picking
  - Clustering
- 3D Reconstruction
  - Combine 2D Images
  - Back Projection
  - Filtering

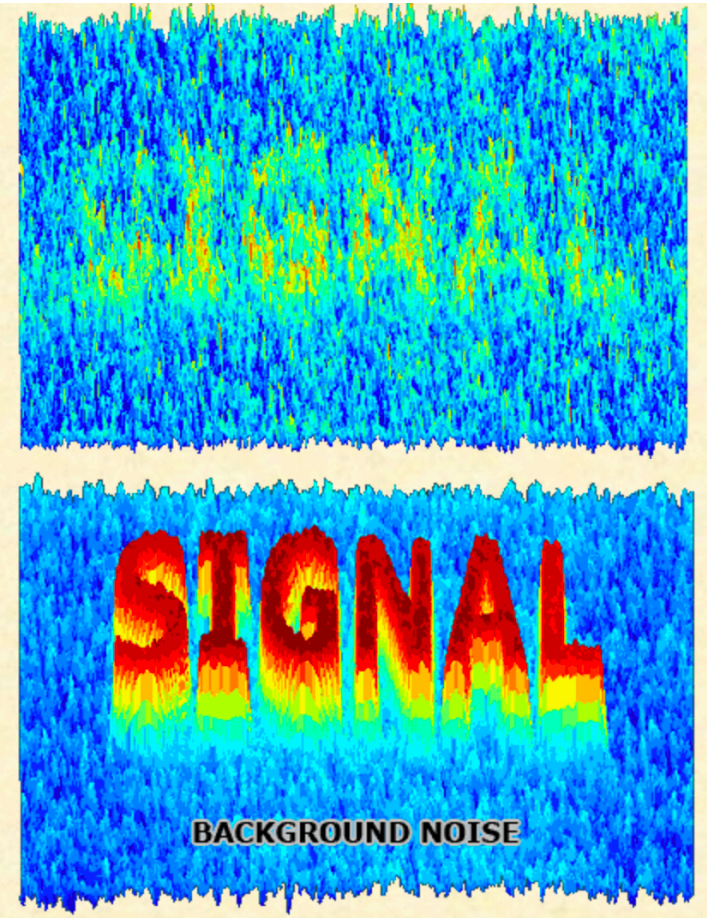


Cheng, Y., et al. (2015). A primer to single-particle cry—electron microscopy. Cell, 161(3), 438-449.

Difficulties

- Noise
- Random Orientations
- Potential Bias in Clustering (Chicken and Egg Problem)
- Overfitting

Signal to Noise Ratio (SNR)



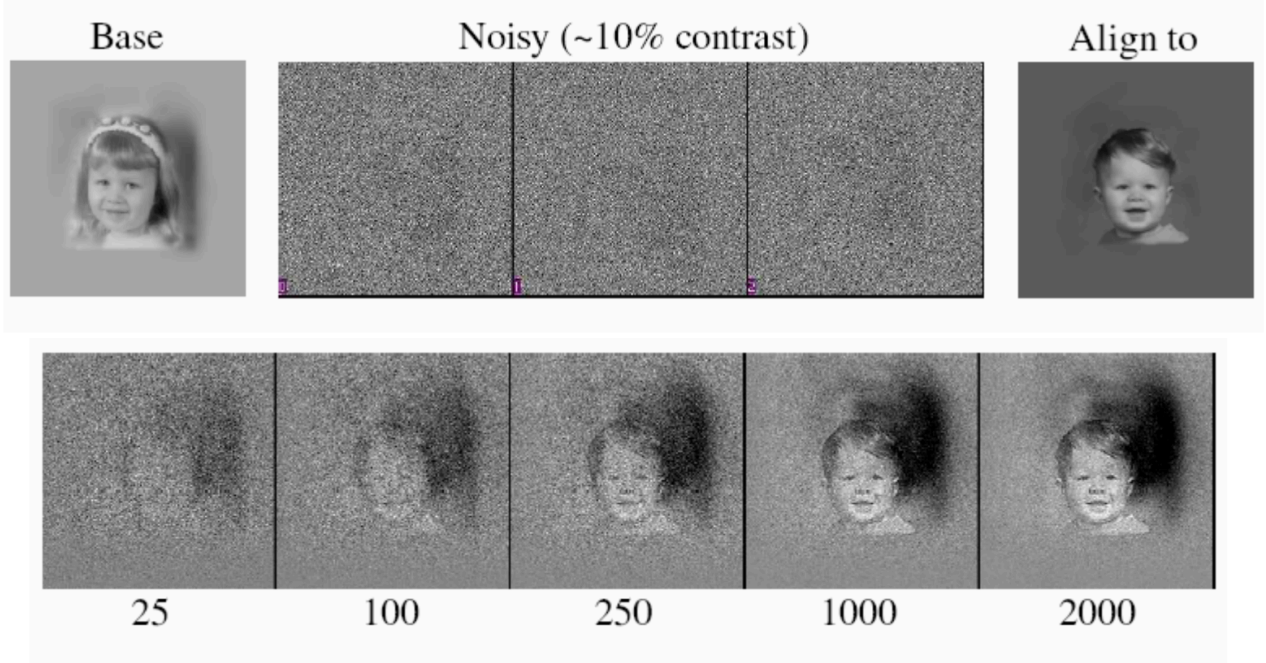
Low SNR

High SNR

https://textureoftime.wordpress.com/2016/03/17/dopamine-adhd-and-signal-to-noise-ratio/

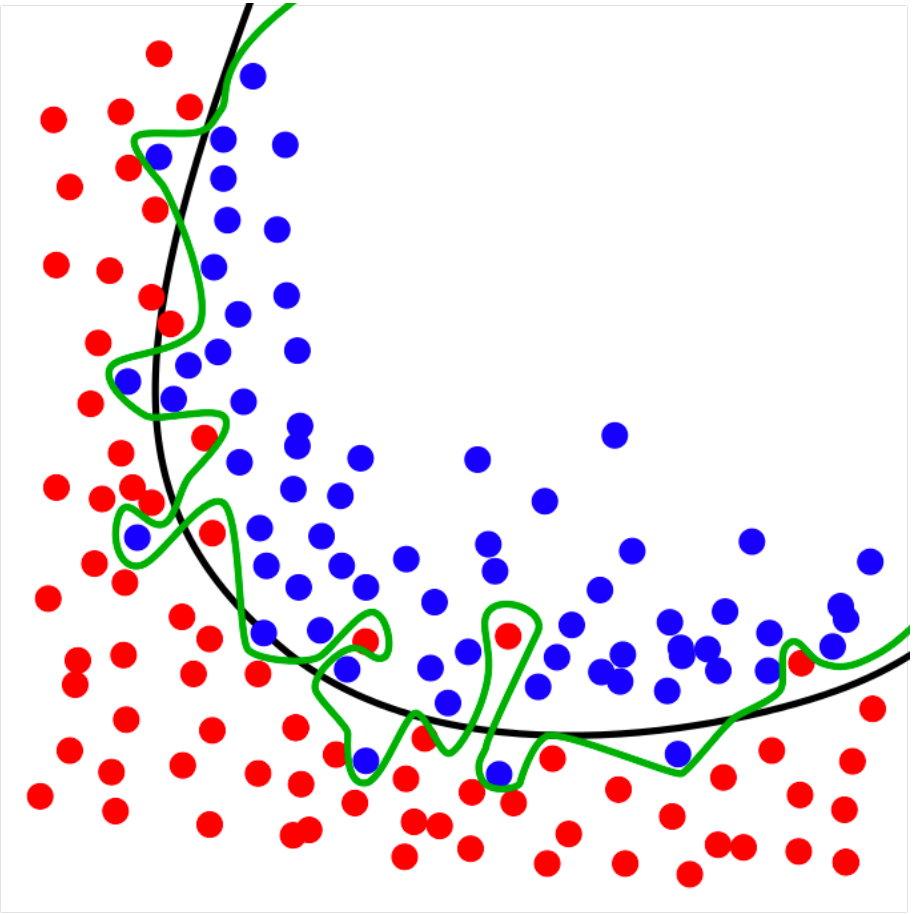
Chicken and Egg

Caveat: Model Bias



Ludtke, S. http://biomachina.org/courses/structures/091.pdf

Overfitting



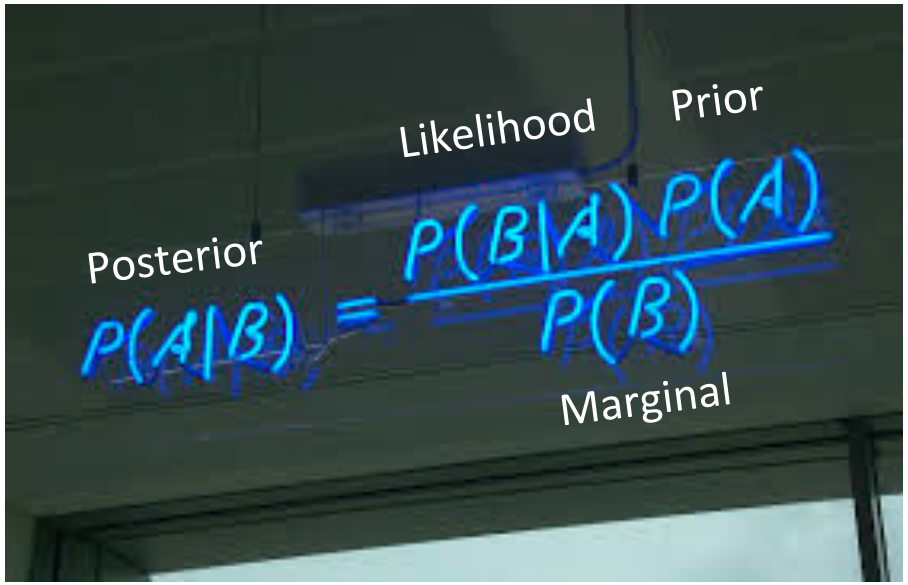
https://en.wikipedia.org/wiki/Overfitting

Smoothness

- Prevent overfitting of noise
- Limits reconstruction at frequencies where SNR is low
- Implemented through *ad hoc* filtering procedures

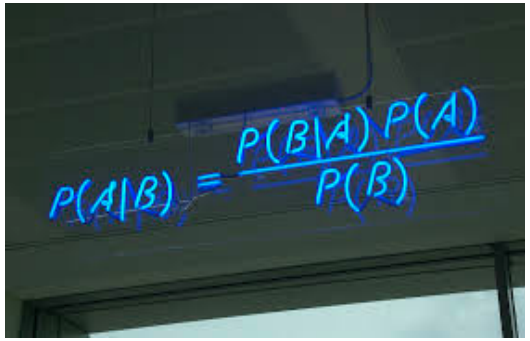
A Statistical Approach

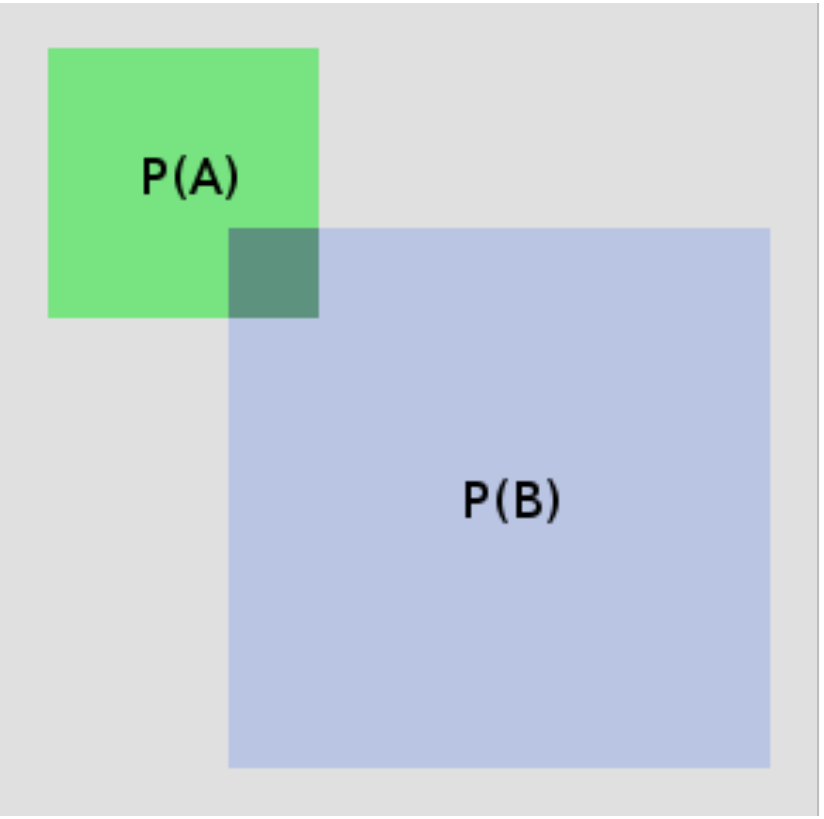
- Old Approach: particle alignment, class averaging, filtering, and 3D reconstruction
- New Approach: maximize a single probability function



https://en.wikipedia.org/wiki/Bayes%27_theorem

Bayes' Theorem





In: P(A) =
P(B) =
P(B|A) = /

P(B|A) x P(A) =

Out: P(A|B) = /

https://wiki.lesswrong.com/wiki/Bayes'_theorem

https://en.wikipedia.org/wiki/Bayes%27_theorem

$$P(\Theta|X,Y) \propto P(X|\Theta,Y)P(\Theta|Y)$$

$$P(X_i|\phi,\Theta,Y) = \prod_{j=1}^{J} \frac{1}{2\pi\sigma_{ij}^2}\exp \times\left(\frac{|X_{ij}-\mathrm{CTF}_{ij}\sum_{l=1}^{L}\mathbf{P}_{jl}^{\phi}V_l|^2}{-2\sigma_{ij}^2}\right)$$

$$P(X|\Theta,Y) = \prod_{i=1}^{N}\int_{\phi} P(X_i|\phi,\Theta,Y)P(\phi|\Theta,Y)d\phi$$

$$P(\Theta|Y) = \prod_{l=1}^{L}\frac{1}{2\pi\tau_l^2}\exp\left(\frac{|V_l|^2}{-2\tau_l^2}\right)$$

$$\Gamma_{i\phi}^{(n)} = \frac{P\left(X_i|\phi,\Theta^{(n)},Y\right)P\left(\phi|\Theta^{(n)},Y\right)}{\int_{\phi'}P\left(X_i|\phi',\Theta^{(n)},Y\right)P\left(\phi'|\Theta^{(n)},Y\right)d\phi'}$$

$$V_l^{(n+1)} = \frac{\sum_{i=1}^{N}\int_{\phi}\Gamma_{i\phi}^{(n)}\sum_{j=1}^{J}\mathbf{P}^{\phi_{lj}^T}\frac{\mathrm{CTF}_{ij}X_{ij}}{\sigma_{ij}^{2(n)}}d\phi}{\sum_{i=1}^{N}\int_{\phi}\Gamma_{i\phi}^{(n)}\sum_{j=1}^{J}\mathbf{P}^{\phi_{lj}^T}\frac{\mathrm{CTF}_{ij}^2}{\sigma_{ij}^{2(n)}}d\phi + \frac{1}{\tau_l^{2(n)}}}$$

$$\tau_l^{2(n+1)} = \frac{1}{2}|V_l^{(n+1)}|^2$$

$$\sigma_{ij}^{2(n+1)} = \frac{1}{2}\int_{\phi}\Gamma_{i\phi}^{(n)}|X_{ij}-\mathrm{CTF}_{ij}\sum_{l=1}^{L}\mathbf{P}_{jl}^{\phi}V_l^{(n)}|^2 d\phi$$

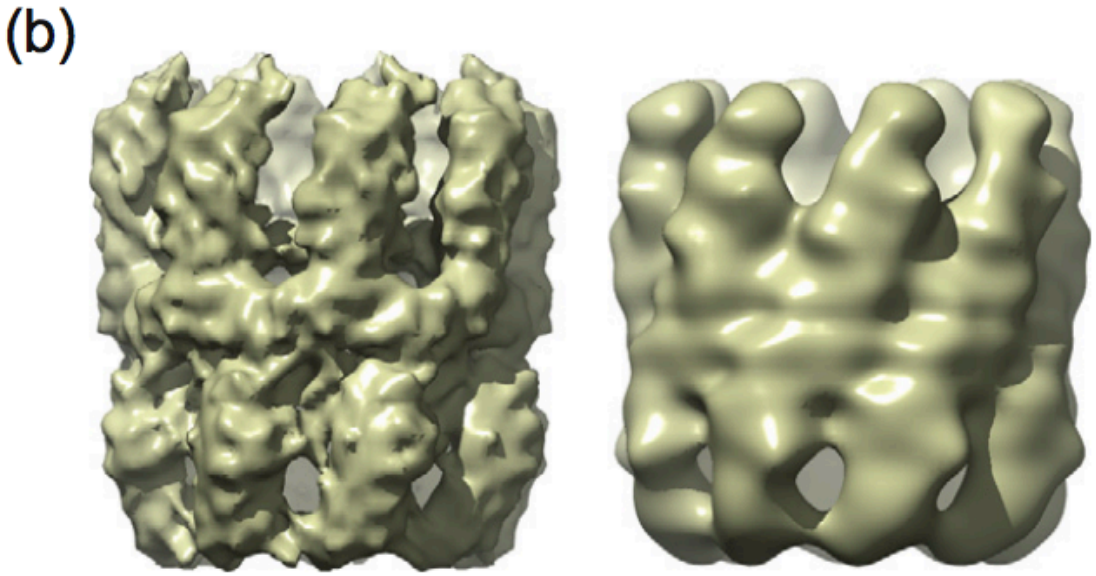
NOTE: There is a parameter T

Scheres, S. (2012). A Bayesian View on Cryo-EM Structure Determination. Journal of Molecular Biology, 415, 406–418.

Intuition

- Assume noise and signals are both independent and Gaussian distributed
- Same assumptions as old filters
- Smoothness: limits power at high frequency components
- Prevent overfitting
- Maximize a single probability function

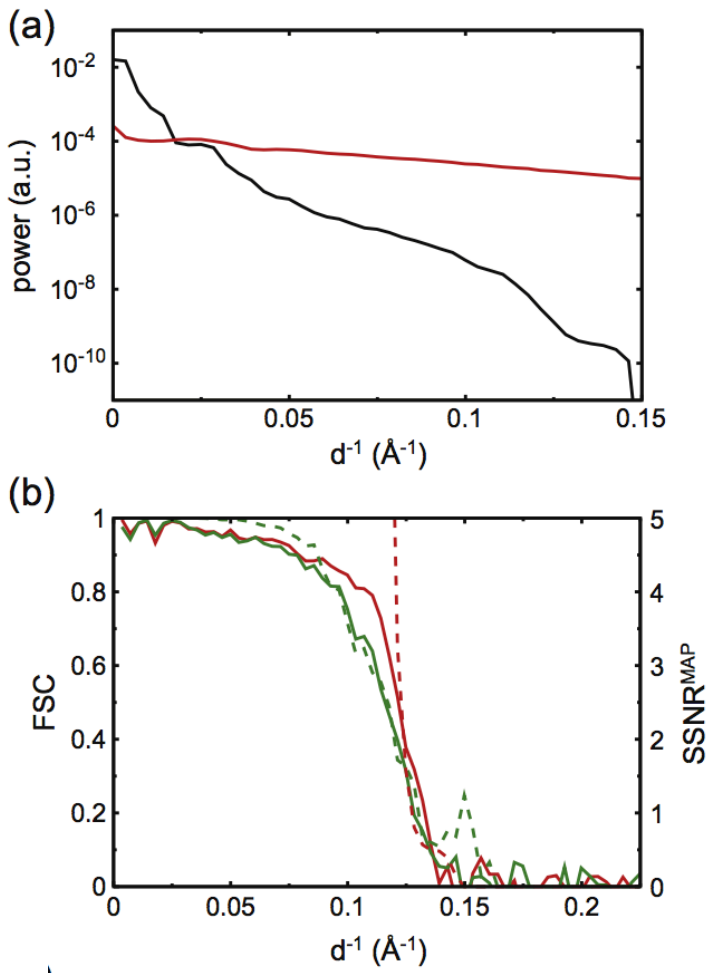
Noise Reduction



Old Method: XMIPP New Method: MAP

Scheres, S. (2012). A Bayesian View on Cryo-EM Structure Determination. Journal of Molecular Biology, 415, 406–418.

Resolution Increased

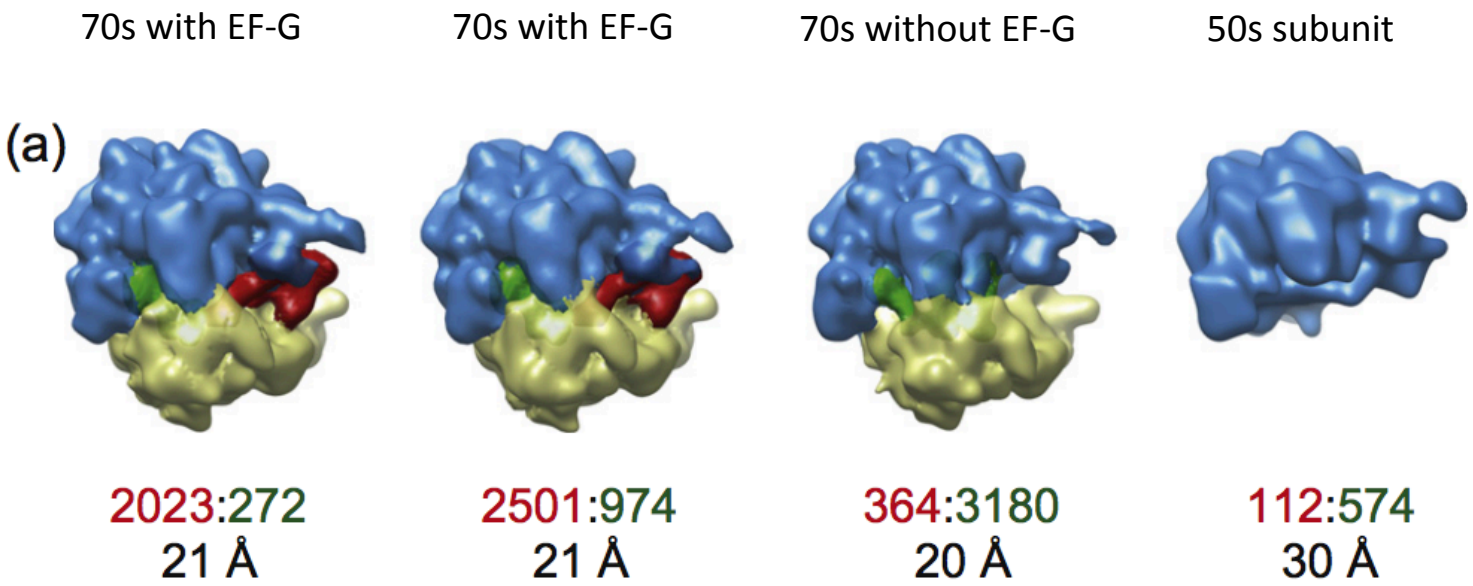


Noise
MAP

XMIPP
MAP

Scheres, S. (2012). A Bayesian View on Cryo-EM Structure Determination. Journal of Molecular Biology, 415, 406–418.

Minority Classes Discovered (K=4)



Scheres, S. (2012). A Bayesian View on Cryo-EM Structure Determination. Journal of Molecular Biology, 415, 406–418.

Strengths

- Standardizes reconstruction (more objective)
- Takes out most arbitrary decisions
- Focus on one task (probability function) instead of multiple steps
- Allows use of more powerful prior knowledge

Limitations

- Doesn’t completely remove parameters
  - K classes and T
- Assumes independence of Fourier components and noise
- Didn’t tune parameters for XMIPP