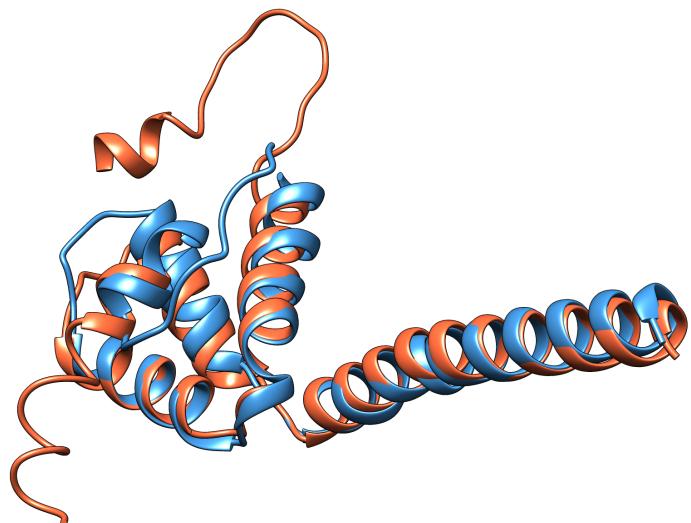


Understanding of cryo-EM: Image formation and 3D reconstruction

written by LATEX

Yining Wang

August 16, 2021



Contents

1 cryo-EM versus Negative-staining EM	5
2 Evaluation Criteria during structure analysis using cryo-EM	5
2.1 Protein preparation	5
2.2 Negative staining	5
2.3 Diagnostic cryo-EM	5
2.4 Initial cryo-EM data collection	6
2.5 High-resolution cryo-EM data collection	6
3 Image processing and 3-D reconstruction	6
3.1 Fourier Transform	6
3.1.1 Fourier Transform	6
3.1.2 Nyquist Frequency	7
3.1.3 Convolution	7
3.2 Image filtering	7
3.3 Central Section theorem and 3-D reconstruction	8
3.4 Common model for information restoration	8
3.5 Orientation Determination	8
4 Single Particle Analysis: Workflow	8
4.1 Micrographs evaluations	9
4.2 CTF estimation of each micrograph	10
4.2.1 Key factors	10
4.2.2 CTF estimation	10
4.2.3 CTF correction	10
4.3 Particle picking	11
4.4 2D Image Alignment and classification	11
4.4.1 Aims	11
4.4.2 Methods	11
4.5 Initial model problem	11
4.6 3D refinement	12
4.7 Resolution assessment and map validation	12
4.8 3D Classification	14
4.8.1 Discrete-state methods	14
4.8.2 Continuous-state methods	14

Abstract

This one is a brief summary of my two-month internship at IGDB of Chinese Academy of Science. It is a great experience and I really appreciate the opportunity to learn some cutting-edge technology like cryo electron microscope and other bioinformatic techniques.

Other techniques when applying cryo-EM are relatively easier to understand so in this note I laid more importance on those parts which are more obscure.

In other words, this note states mostly about the theoretical part about image formation and the basic operation of cryo-EM.

After all, the whole process of this study will last for a long period of time, hopefully we can get a good result.

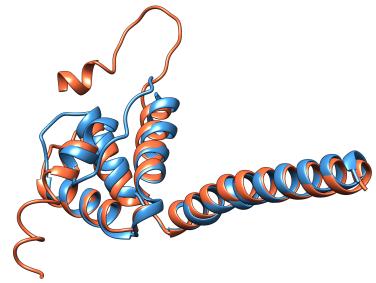


Figure 1: HMW-GS predicted by AlphaFold(Pipeline)

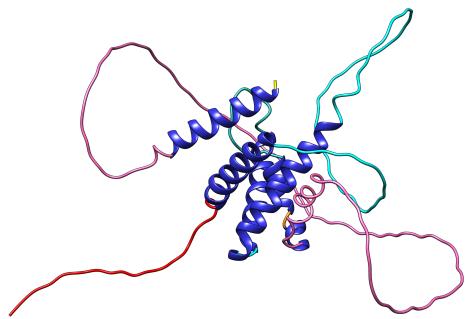


Figure 2: LMW-GS predicted by AlphaFold(Colab version)

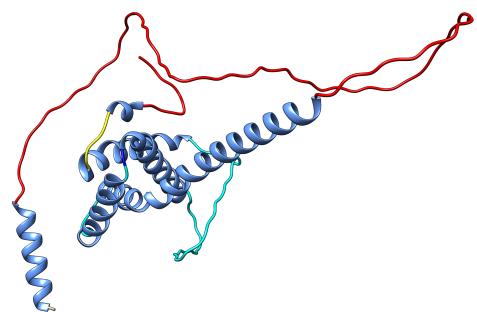


Figure 3: Gliadin predicted by AlphaFold(Colab version)

1 cryo-EM versus Negative-staining EM

- Using Negative-staining EM to get a initial model and then using cryo-EM to REFINE the model
- Not all structures need initial model!(symmetric or high-quality enough or in a nice shape!)
- The sample preparation process is much easier using ns-EM

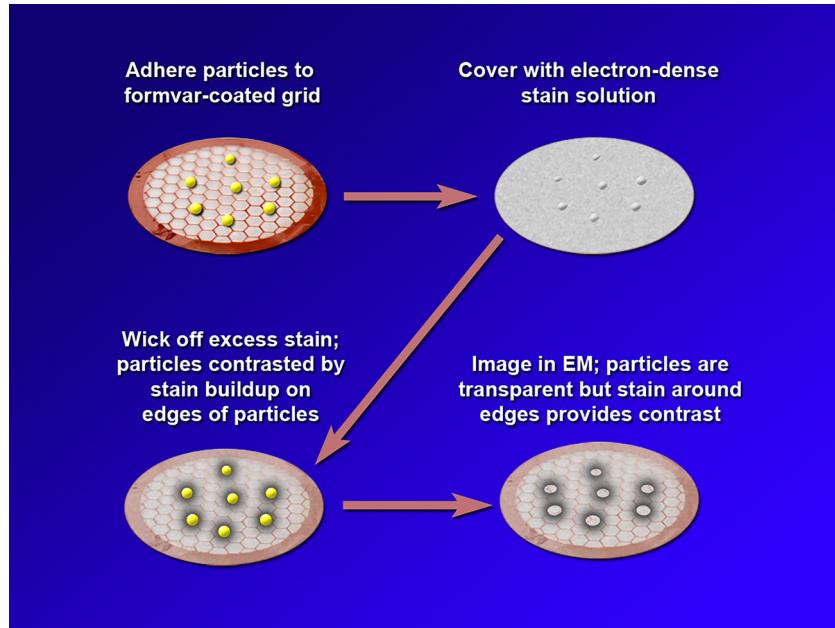


Figure 4: Negative-staining EM

2 Evaluation Criteria during structure analysis using cryo-EM

2.1 Protein preparation

- Composition
- Purity and homogeneity
- Stability (buffer composition)
- Biochemical activity

2.2 Negative staining

- Discrete particle
- Stability
- Particle size and particle shape

2.3 Diagnostic cryo-EM

- Stability
- Particle size and shape
- concentration and particle distribution

2.4 Initial cryo-EM data collection

- High-resolution 2D classes
- Initial 3D model
- Orientation and distribution
- Particle yield

2.5 High-resolution cryo-EM data collection

- Tilt pairs
- Motion statistics
- Angular accuracy
- Conformational states

3 Image processing and 3-D reconstruction

3.1 Fourier Transform

3.1.1 Fourier Transform

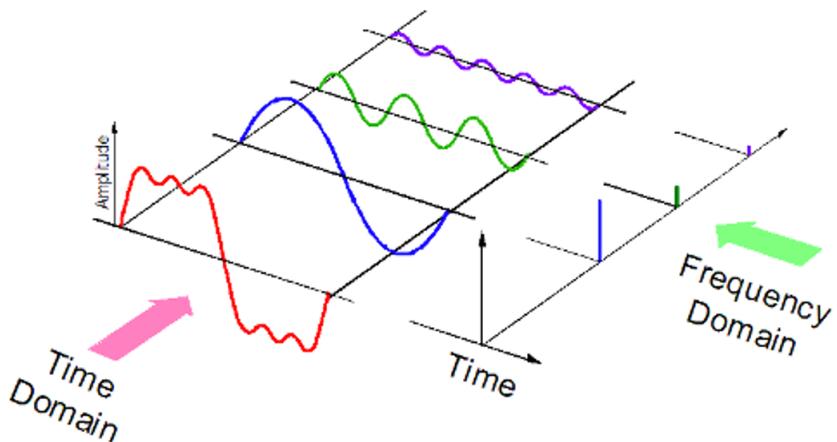


Figure 5: Fourier transform

Fourier Transform allows us to describe a specific real-space matter in the frequency domain. It answers the question that how do the density values of each point in space overlay. Hence it gives us a new angle to understand familiar substance. Composition of different spatial frequencies forms a image in real space \Leftrightarrow reciprocal space. There are two main information conveyed by a WAVE : Amplitude(A) and Phase shift(φ), and through a Fourier Transform these two elements could be linked.

$$F = |A| \cdot e^{i\varphi_F}$$

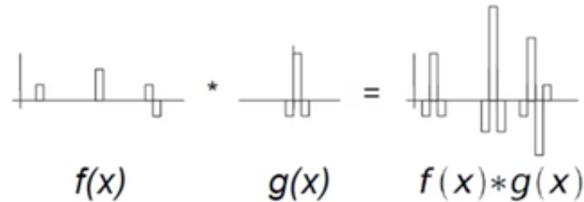
The process of image formation is a process of FOURIER TRANSFORM !

3.1.2 Nyquist Frequency

The highest frequency information in the image, The magnitude of the change cannot exceed the size of one pixel. If the size of a pixel is 2\AA , as a result the smallest wavelength can be described is 4\AA . For example, if we want to get a resolution of 1 *angstrom* the pixel size needed is 0.5 *angstrom*

3.1.3 Convolution

Every spot on the image becomes a Convolution kernel(CTF in this case) and then stacked together.



Continuous Real Space Convolution:

$$f(x)*g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

$$h(x) = f(x)*g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

Figure 6: Convolution theorem 01

$$\begin{aligned} h(x) &= f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(t)g(x-t) dx \\ V(x) \otimes P(x) &= Im(x) \end{aligned}$$

$P(x)$:Point scatter function – This related only to the imaging system, it has nothing to do with the sample.

The Fourier transform of the convolution of two functions is the product of their respective Fourier transforms

3.2 Image filtering

Take FFT \Rightarrow Multiplied by filter \Rightarrow IFT

- High-pass filter: Good for high resolution details
- Low-pass filter: Increase contrast and SNR
- Bond-pass filter: Contrast Transform Funtion

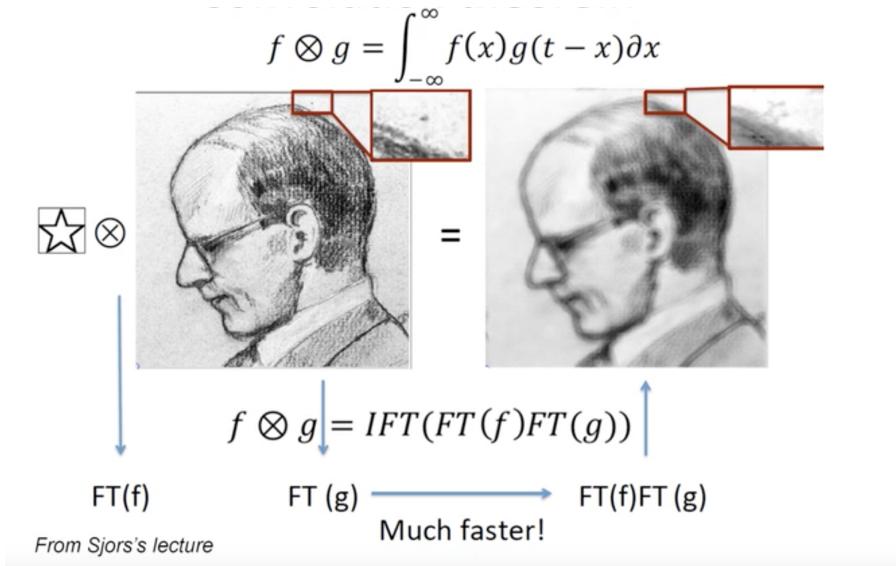


Figure 7: Convolution theorem 02

3.3 Centeral Section theorem and 3-D reconstruction

There is an exact relationship between Fourier transform and its mathematical projection

3.4 Common model for information restoration

$$X = PSF \otimes P_\varphi V \otimes MTF + Noise$$

As it is shown above, the main problem for us is to solve the parameter V form X. Under our specific circumstance the function is below

- real space: $X = PSF \otimes P_\varphi V \otimes MTF + N$
- reciprocal space: $X = CTF \cdot \hat{P}_\varphi \hat{V} \cdot MTF + \hat{N}$
- $CTF = \sin[\Pi \cdot \lambda \cdot \Delta f \cdot s^2 + \frac{\pi \cdot \lambda^3 \cdot C_s \cdot s^4}{2}]$

\hat{P}_φ can be solved only when the ORIENTATION can be determined!

3.5 Orientation Determination

Maximum likelihood & Bayesian method The common line technique and Projetcion matching

4 Single Particle Analysis: Workflow

In this section, each part of the SPA will be noted carefully. Main subsections range from evaluation of the micrograph to the final 3D refinement of the model!

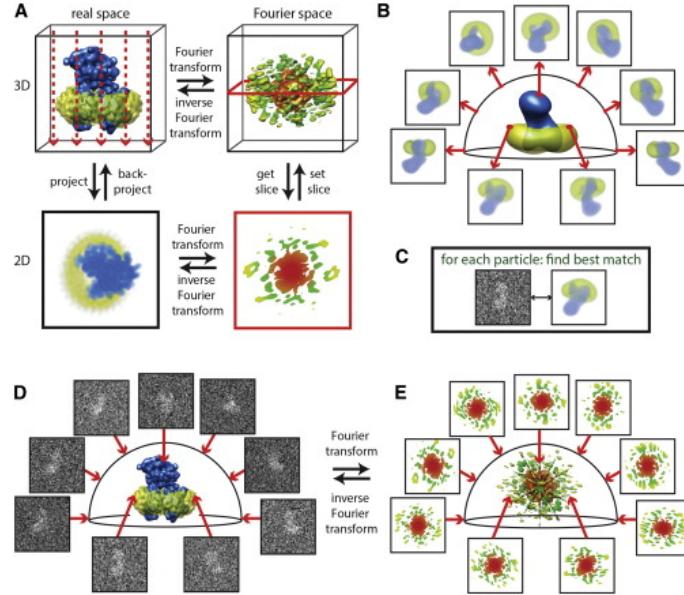


Figure 8: Central Section theorem

4.1 Micrographs evaluations

There are several common phenomena in this part. To make evaluation about the micrograph the very first thing to get is the POWER SPECTRUM of the photo, any slight drift, astigmatism or loss of resolution can be discovered clearly. Also, if there are many THON rings, it means the ice layer is way too thin, on the opposite, if the number of rings are too low, it means you got a thick layer of ice. More rings usually means that the high-resolution details of the image are great, but CAUTION , the thin layer of ice may also bring damage to the specimen.

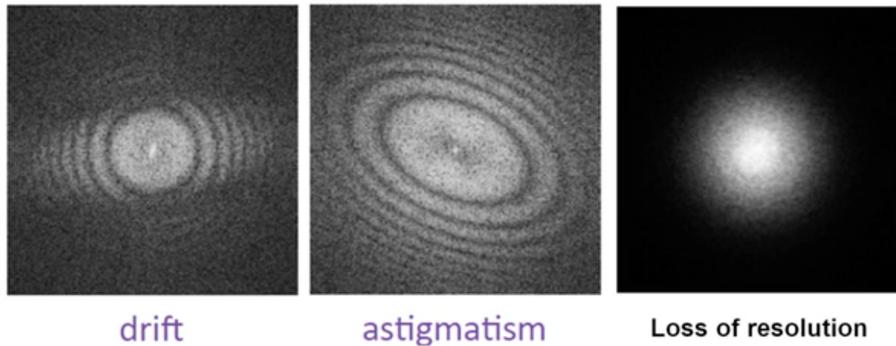


Figure 9: Drift, astigmatism and loss of resolution

4.2 CTF estimation of each micrograph

$$CTF = \sin[\Pi \cdot \lambda \cdot \Delta f \cdot s^2 + \frac{\pi \cdot \lambda^3 \cdot C_s \cdot s^4}{2}]$$

4.2.1 Key factors

There are few key factors which determine the Contrast Transfer Function.

- Power spectrum
- Microscope parameter (Kv, Cs, pixel size...)
- Defocus(Δf) —> Delocalization

4.2.2 CTF estimation

As the function shown above the main parameter in solving CTF is Δf and Astigmatism, to solve these two items, the method used is to compare the CTF image to the Thon Ring Pattern (Exhaustively), until a perfect match is complete!

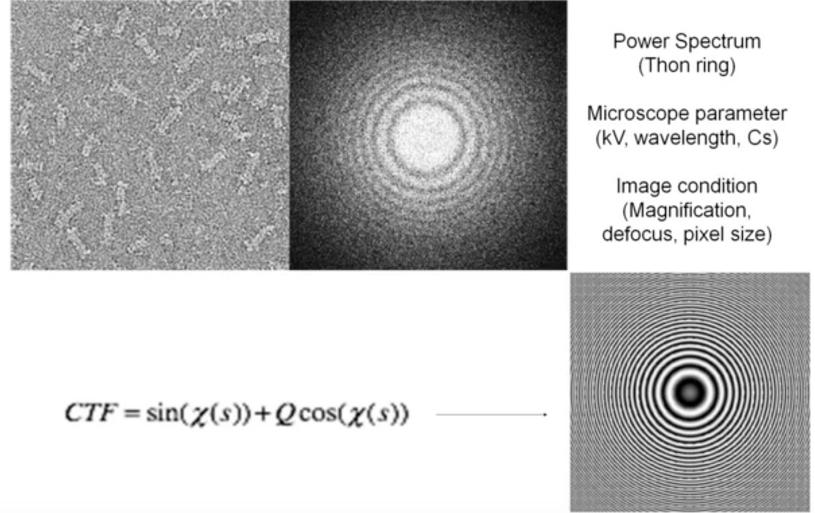


Figure 10: CTF estimation

4.2.3 CTF correction

$$-1 = e^{i\pi}(180^\circ \text{phase shift})$$

CTF can be corrected by phase flipping (adding 180° to the phases) after measurement of Δf . But a dataset should comprise images within an adequate range of defocus in order to recover informations in zones of zero-crossing of CTF.

► PHASE FLIPPING

$$\begin{aligned} X(s) &= CTF(s) \cdot F(S) + N(S) \\ M(s) &= CTF(s)^2 \cdot F(s) + CTF(s) \cdot N(s) \end{aligned}$$

►WEINER FILTER

$$M(s) = X(s) \cdot \frac{CTF(s)}{\overline{CTF(s)^2} + 1/SNR}$$

$$SNR = \sigma^2(F(s))/\sigma^2(N(s)) [SSNR(s) = \frac{M(s)^2 - N(s)^2}{N(s)^2}]$$

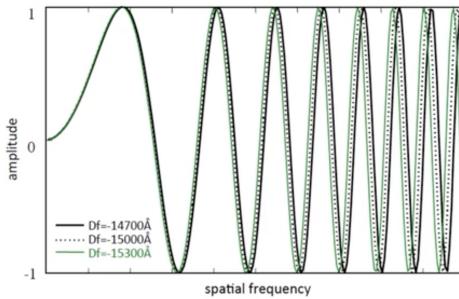


Figure 11: Small variation in defocus have significant effect in High-resolution details

4.3 Particle picking

- Manual particle selection & Autopick
- Box Size (1.5–2 time → good for FT)
- cross-correlation

4.4 2D Image Alignment and classification

4.4.1 Aims

- clean the dataset
- increase SNR to generate averaged micrographs
- classification to purify particles

4.4.2 Methods

1. 2D Image Rotation Alignment (rotation alignment)
2. Maximum likelihood

4.5 Initial model problem

Initial model problem plays a fundamental part in the image formation, because the following part (Projection Mapping) needs an appropriate initial model to make 3D refinement, the main methods are below:

- Random Conical Tilt (Missing cone probably)
- Class average and Common line method (Good SNR needed)
- Electron Tomography and sub-tomo average

Refinement

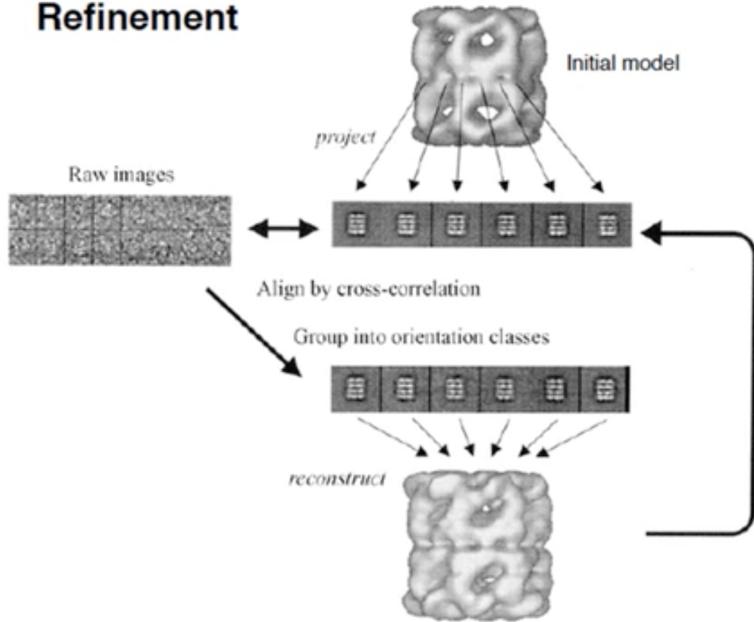


Figure 12: Iterative Refinement–Projection Mapping

4.6 3D refinement

Major methods used in 3D refinement are Projection mapping and Bayesian Method

4.7 Resolution assessment and map validation

Once a concentration map is built, the job to do is to make assessment of the resolution and furthermore, apply the map validation, if the resolution of the image is high enough and the side chains can be determined, the validation of the map can be done easily. But we can not always get hi-resolution images so the proper methods and algorithms to do resolution assessment and map validation is needed

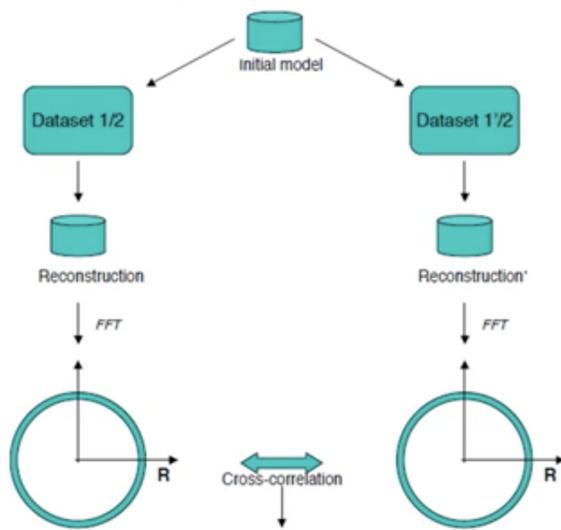
$$F_L \cdot F_R^* = |F_L| |F_R| - e^{i \cdot (\varphi_L - \varphi_R)}$$

The function above is a simplified version of Fourier Shell coefficient, by using this method the resolution of the initial model can be calculated. If the resolution isn't estimated correctly, a phenomenon called over fitting will come up and affect further section. PARTICLE NUMBER DOES MATTER!!! Because the number of the particles will determine Signal-to-Noise Rate and in advance, determine the resolution, and there is a simple way to estimate the number of particle needed to reach a certain resolution by using the diameter of the particle

$$\text{Num} = \frac{\pi \cdot D}{(m) \cdot \text{Res}} \times 100 \times 5$$

Once the resolution assessment is done the next part is to apply map validation, one of the mainstream method is called Tilt pair Validation.

Fourier Shell Coefficient (FSC)



$$FSC(k, \Delta k) = \frac{Re \left| \sum_{[k, \Delta k]} F_1(k) F_2^*(k) \right|}{\left[\sum_{[k, \Delta k]} |F_1(k)|^2 |F_2(k)|^2 \right]^{1/2}}$$

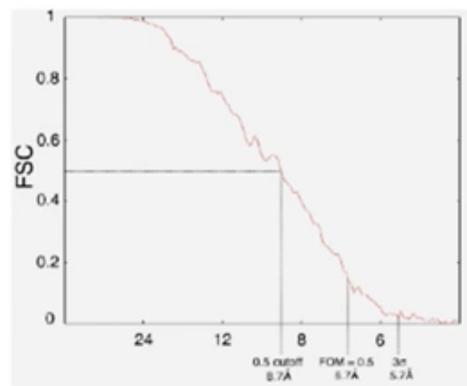
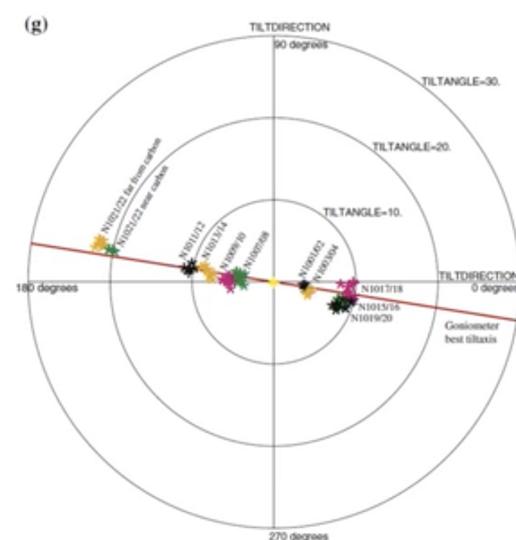
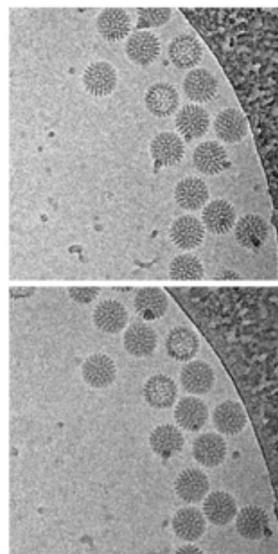


Figure 13: Fourier shell coefficient



Tilt-Pair Analysis of Images from a Range of Different Specimens in Single-Particle Electron Cryomicroscopy

Figure 14: Tilt Pair Validation

4.8 3D Classification

Although the process stated above can be done by computer programs like RELION, it is not that easy because the exsistance of two types of heterogeneity: Composition heterogeneity and Conformational heterogeneity. Because the SPA is valide under the hypothesis of all the particles are exactly SAME, so special computational strategies are needed to solve these two problems, MDA(Multivariate Date Analysis) is a prevailing one. Discrete-state heterogeneity(apple/pear) is often easier to deal with than Continuous-state (green apple/light-green apple)

4.8.1 Discrete-state methods

- Multireference classification
- Maximum likelihood methods
- Statistical Analysis methods

4.8.2 Continuous-state methods

There is a clear relation between the possibility of a electron's apperance on certain spot and it's energy

$$P \propto e^{-\frac{E}{kT}}$$

There are two methods to deal with the problem

1. Hybrid EM NMA
2. Manifold Embedding Method

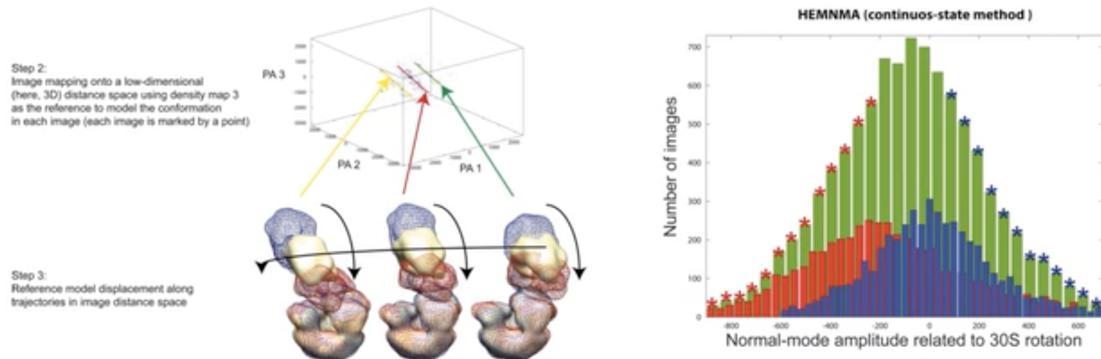


Figure 15: HEMNMA

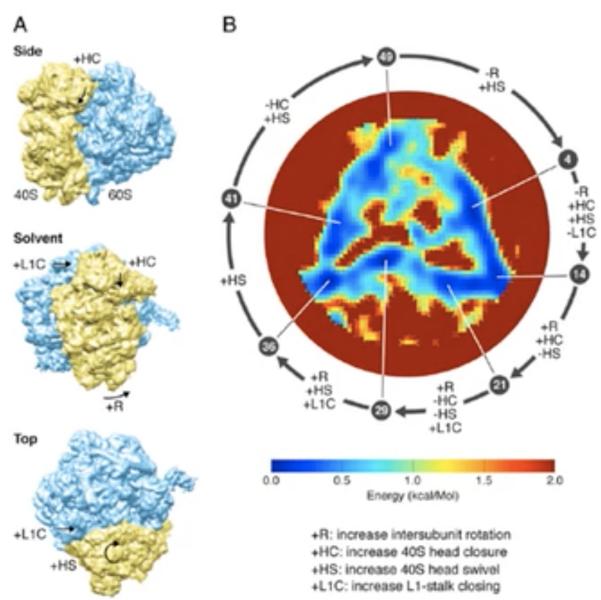


Figure 16: Manifold Embedding Method