TMM 3102: Protein Structure, Function and Disease

- Structural Biology Methods: Cryo-Electron Microscopy (Cryo-EM) (October 5th, 2021)

Jyh-Yeuan (Eric) Lee, Assistant Professor, BMI
How “Tiny” Can We See?

From humans eyes to analytical instruments, we are all limited to how small objects we can see.

For cells, we can easily observed under a light microscope and with more detailed information using electron microscopes.

To see objects at atomic resolution, so far, we know X-ray crystallography, transmission electron microscopy, and NMR spectroscopy can enable such high resolution imaging.

This course will selectively focuses on these three methodologies that enable vast protein structure-function studies so far.
Electron Microscopy in Biology

Transmission electron microscopy (TEM)
1) section of specimen
2) 2-D projection images

Scanning electron microscopy (SEM)
1) usually entire 3-D objects
2) surface tomographic pictures

(UMass Med, EM Facility)
Travel passage of electron beams

Figure 4.13: Ray diagram of a transmission electron microscope and a scanning electron microscope
Dimensions of biological objects. Using the thumb for example, when we keep zooming-in, the dimension to see things becomes smaller and smaller:

- 20 mm: thumb
- 2 mm: fingerprint
- 0.2 mm: skin tissues
- 20 μm: skin cells
- 2 μm: organelles
- 0.2 μm: protein networks
- 20 nm: protein tertiary/quaternary structure
- 2 nm: single protein molecule
- 0.2 nm (or 2 Å): covalent bonds

Cryo-EM
Milestones: Timeline

- **1968**: Birth of cryo-EM
  - Taylor and Glaeser

- **1974**: First 3D-reconstruction
  - DeRosier and Klug

- **1982**: Vitrification technology
  - Dubochet et al.

- **1987**: Atomic structure of bacteriorhodopsin by electron crystallography
  - Henderson et al.

- **1990**: Single particle: random conical tilt
  - Radermacher, Frank et al.

- **1994**: Projection matching
  - Penczek, Frank et al.

- **1997**: 2D maximum-likelihood
  - Sigworth

- **1998**: 10 Å virus structure
  - Böttcher, Crowther et al.; Conway, McDowell et al.

- **2008**: Direct electron-detection cameras used for motion correction
  - Grigorieff et al.

- **2012**: De novo tracing <200 kDa
  - Scheres, Shi et al.

- **2013**: <3 Å structures
  - Carragher, Potter et al.; Subramaniam and colleagues

- **2015**: Near-atomic maps of 30,000 ribosomes
  - Scheres et al.

(Eva Nogales, Nat Methods, 2016)
Making Structural Biology Possible

2017: Cryo-Electron Microscopy (cryo-EM)

From left: Jacques Dubochet, Joachim Frank and Richard Henderson developed cryo-electron microscopy.

Molecular-imaging pioneers scoop Nobel

Chemistry prize hails work on cryo-electron microscopy.

Only 2-D projections are recorded \((x,y)!!\)
Milestones: 3-D Reconstruction

At that stage, having come into structural biology through X-ray diffraction in which all the phases of the Fourier components, also observed through Bragg diffraction from the crystal lattice, had to be determined indirectly, I also thought that electron diffraction was intrinsically more promising than electron microscopy because the elegant simplicity of recording electron diffraction patterns compared favourably with the multiple difficulties of recording good images. Therefore spent several years trying to extend the resolution of the bacteriorhodopsin structure using anumber of diffraction-based approaches. Figure 3summarises the different ideas we tried. Tom Cieska tried to make heavy atom derivatives.[10] Joyce Baldwin and Michael Rossmann tried molecular replacement.[11,12] David Agard tried to extend the phases using a multi-parameter model building approach (unpublished). Although all of these approaches gave hints of success that were encouraging at times, none of them were powerful enough to give phases that resulted in convincing maps that were interpretable much beyond the resolution obtained in 1975. It was not until Tzyy-Wen Jeng and Wah Chiudemonstrated, in a collaboration with Fritz Zemlin,[13] that images showing clearly visible diffraction spots at 3.9 Å resolution could be obtained from thin 3D crystals of rattlesnake venom crotoxin using an electron microscope in Berlin with a liquid-helium superconducting objective lens, that I became convinced electron cryomicroscopy could produce high quality images. We therefore embarked, as last resort, on using electron cryomicroscopy for high-resolution phase determination (see Figure 4). In earlier years, Bob Glaeser's group had shown that freezing thin 3D crystals of catalase could produce good electron diffraction patterns and images[14,15] and that there was a benefit in terms of reduced radiation damage,[16] but I had been unconvinced by earlier attempts to show that electron cryomicroscope images of purple membrane contained high-resolution information.[17]

The change of emphasis from diffraction to imaging proved to be very challenging. I began with a visit to Jacques Dubochet's laboratory at the European Molecular Biology Laboratory (EMBL) in Heidelberg in 1984 working with Jean Lepault to record images on their hybrid Zeiss/Siemens microscope with the same design of superconducting liquid-helium objective lens as on the Berlin microscope. We spent a week with that home-constructed microscope, which turned out to be very unreliable. Fortunately, we managed to obtain just one image that showed diffraction beyond 4 Å resolution, although because of the difficulty of alignment and the short mean time between failures, that image had over 5000 Å of astigmatism. We did not pursue further imaging at EMBL. Nevertheless, that was the first image that allowed us to begin developing procedures for the computer-based processing of

---

Figure 1. The first projection structure at 7 Å resolution of the purple membrane calculated in October 1974 using 36 reflections obtained by room-temperature electron diffraction and imaging of glucose-embedded 2D crystals of bacteriorhodopsin.[7]

Figure 2. The structure of bacteriorhodopsin at 7 Å resolution in 3D from 18 images and 15 diffraction patterns. The collage shows (a) freeze-fracture picture from Walther Stoeckenius, (b) electron diffraction pattern obtained much later using a phosphor/fibre-optics/CCD camera, (c) the 1975 balsawood model of a single bacteriorhodopsin molecule.[8]

(Anwer & Henderson, J Mol Biol, 1975)
Milestones: 3-D Reconstruction

Higher Resolution (~2Å) and Bigger Molecules/Complexes

Aquaporin (1.9 Å)
(Gonen et al, Nature, 2005)

β-Galactosidase (1.5 Å)
(Bartesaghi et al, Structure, 2018)

Mitochondrial Complex I (4.2 Å)
(Zhu et al, Nature, 2016)
October surprise: 1.22/1.25 Å (apoferritin)

**Fig. 3** Apoferritin reconstruction. **a**, B-factor plots for reconstructions using: high-order aberration (+aberr) and Ewald sphere correction (+Ewald; orange); high-order aberration correction only (blue); and no correction (grey). B-factors estimated from the slopes of fitted straight lines are shown in the same colours. Numbers in parentheses and error bars represent estimated and sample s.d.s from sevenfold random resampling, respectively. **b–d**, Densities from the 1.22 Å map (blue) for M100 (b), F51 (c) and L175 (d). **e**, Hydrogen-bonding network around Y32 and water-302 is visible in the difference map (green, positive; orange, negative). **f**, The α-helix hydrogen-bonding network involving residues 20NRQIN27, shown as in **e**.

*(Nakane et al, 2020, Nature)*
Cryo-EM applications in protein structural biology:

1) Single particle analysis:
(no need to generate protein crystals)

a. Prepare purified protein samples and freeze them.

b. Take electron micrographs of isolated protein particles.

c. Collect several of “identical” images and add altogether to enhance the signal of the objects (darker area).

d. Then place amino acid models in the electron density.

(Bai et al, eLife, 2013)
Microelectron Diffraction (MicroED) & Electron Crystallography

Cryo-EM applications in protein structural biology:

2) Microelectron diffraction:
   (need to generate protein crystals, but much smaller than that for X-ray crystallography)
   
   a. Prepare purified protein samples, crystallize and freeze them.
   
   b. Take electron micrographs of protein crystals and diffraction images.
   
   c. Use diffractions to establish the amplitude information and images to find the phase information, then calculate the structural factors to generate the electron clouds for the target proteins.
   
   d. Model building like that in X-ray crystallography.

(Martynowycz & Gonen, Curr Opin Colloid Interf Sci, 2018)
Cryo-EM applications in protein structural biology:

3) Electron tomography: (suitable for large protein-protein network, organelles, subcellular structures)

a. Prepare purified/ homogenous samples and freeze them.

b. Take electron micrographs of isolated protein particles at different angles almost at the same time.

c. Reconstruct the 3-D images instantaneously.
Optical & X-ray Diffractions

Schematic diagrams to illustrate the optical difference between X-ray crystallography (bottom) and electron/light microscopy (top):

Microscopy (top):
When a periodic object is subjected to laser or electron beams, the diffracted beam (information in reciprocal space, i.e., FT product) is processed by a physical lens which then converts the information back to image in real space, i.e., IFT.

X-ray (bottom):
The information of the periodic object is only collected on the diffraction plane, because of lack of X-ray lens. To generate the real image back, it requires algorithms to process such IFT operation as would be done in microscopy.

If \( F(X) = \text{FT}[f(x)] \), then \( f(x) = \text{IFT}[F(X)] \), where \( \text{FT} = \text{Fourier transform} \) & \( \text{IFT} = \text{Inverse Fourier transform} \). Note: important in X-ray crystallography and 3D reconstruction algorisms.
Optical & X-ray Diffractions

Schematic diagrams to illustrate the instrumental difference between X-ray crystallography (left) and cryo-electron microscopy (right).

**X-RAY CRYSTALLOGRAPHY**
X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.

**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.
Anatomy of light and electron microscopes

Light microscope:
1) bright-field
2) fluorescence
3) UV

Usually made of glass/quartz

Electron microscope:
1) transmission (TEM)
2) scanning (SEM)
3) scanning/transmission (STEM)

Made of electric magnets
Why do we need to know optics?

Ideal instrument

Real instrument
**Electron Gun**

- **Thermionic Emission Gun**
  - Electrons are emitted from the heated filament.
  - Lanthanum Hexaboride (LaB6) Single crystal: \(~1900\text{K}\)
  - Hairpin Tungsten (W) filament: \(\sim 2600\text{K}\)

- **Field Emission Gun** *(high coherency)*
  - Electrons are extracted from the W emitter tip surface by the high electric field \(\sim 10\text{V/nm.}\
  - Cold cathode type: room temperature
  - Schottky type: \(\sim 1800\text{K}\)
Why Do You Need to Know Electron Optics?

Ideal instrument Real instrument

Phosphor Screen or Digital Detectors

Camera and Viewing System

Low-Dose Imaging

- **Search Mode:** Low mag, low dose rate $10^{-3} \text{e/Å}^2/\text{sec}$
- **Focus Mode:** High dose in an adjacent area at chosen magnification
- **Photo Mode:** Defocus at the desirable value for an anticipated resolution. Use a total dose of $\sim 20-50 \text{e/Å}^2$

Defocus will be an important factor to determine the image quality.
Imaging: Parameters to Consider

- Dose rate: “low-dose” mode

(Grant & Grigorieff, eLife, 2015)
Cryo-EM Workflow

Green: Protein biochemistry
Sample preparation

Light blue
Imaging
Data collection

Blue
Image processing
Model building

Purple
Model validation

(Costa et al, Meth Mol Biol, 2017)
(Dörr, Nat Meth, 2016)
How do we start? 4 aspects.

- Core knowledge
- Biochemistry & Sample preparation
- Data collection
- Processing & Data analysis
Bottlenecks

- Core knowledge
- Biochemistry & Sample preparation
- Data collection
- Processing & Data analysis
Protein Sample Quality

So what kind of sample quality do we need to do cryo-EM?

Use single-particle analysis as example:

a. It needs to be as pure as that used for crystallization.

b. No need of mg quantity of protein yields. In general, when one can see a clean and sharp protein bands by silver staining, it's sufficient enough.

Note: After running an SDS-PAGE gels, proteins can be stained in many ways. Common staining is done with Coomassie blue dye, but its sensitivity is limited to $\mu g$-$mg$ proteins. Silver staining however can detect protein quantity at as low as ng.

Silver staining

(But no need to purify several mg proteins!)

(Abeyrathne & Grigorieff, PLOS ONE, 2017)
Protein Sample Quality

(Functionally characterized)

(Qu et al, Cell, 2018)
EM Samples: Negative Staining

- Principle:
  - Embedding objects in a layer of heavy-metal salts that surround the proteins like a shell.
  - Shape of objects are visible in contrast to the optically opaque stains.
- Benefits:
  - Small amount of proteins (0.01 mg/mL)
  - Easy and quick (preparation and imaging)
  - No need of high-end microscope; diagnostic
- Downsides:
  - Low resolution (e.g., high noise from stains)
  - Artifacts (lack of hydration)

(Brenner & Horne, BBA, 1959)
The protein solution is spread on a very thin carbon film perforated with small 1 µm diameter holes. The sample is rapidly frozen in liquid ethane to form vitreous ice (water molecules frozen randomly, not in an ice lattice). This fixes the molecule, prevents water evaporation, and helps protect the protein from radiation damage - even if bonds are broken, the atoms are fixed in place.

http://faculty.washington.edu/lw32/cryoem_home.php
When preparing for cryo-EM samples, one important factor is to avoid formation of crystalline ice (right panel). Formation and size of crystal ices is directly associated with the cooling speed (middle panel). The process to avoid big crystal ice and freeze samples with <1nm ice crystals is call **vitrification**, and the ice is called vitreous ice. To achieve in cryo-EM sample preparation, we use a quick-freezing device (plunger) and freeze the sample in **liquid ethane** at -183°C.

Note: Liquid nitrogen, while colder than ethane, is not good, because it boils samples first and causes easy formation of big crystalline ice.
Cryo-EM Workflow

Green:
Protein biochemistry
Sample preparation

Light blue
Imaging
Data collection

Blue
Image processing
Model building

Purple
Model validation

(Costa et al, Meth Mol Biol, 2017)
(Dörr, Nat Meth, 2016)
The Fourier Transform is a tool that breaks a waveform (a function or signal) into an alternate representation, characterized by sine and cosines. The Fourier Transform shows that any waveform can be re-written as the sum of sinusoidal functions.
2-D images to 3-D Reconstruction

General concept to generate 3-D images from 2-D images:

An electron micrograph shows what it looks like a sheet-like picture when the light sources penetrate the samples. This generate 2-D projections.

Taking a duck image as an example:

a. One takes pictures of the duck from various angles, where each picture is a 2-D projected image.

b. In order to mathematically combine all 2-D information together, one needs to operate such combination using reciprocal information. Hence converting all 2-D images by FT.

c. Combination of all 2-D reciprocal information results a convoluted image in 3-D.

d. An inverted FT transforms the reciprocal image into a real 3-D image of a duck.
A. Raw image data typically seen from a cryo-EM experiments. Protein particles are usually shown as the dark objects. This image shows particles of a proteasome complex.

B. However, often time, each particle looks “blurred”, largely because of image drifting during the picture taking, as well as the microscopic movement of protein molecules in the ice. "Motion correction" is thus necessary to help enhance the image quality, i.e., making the images sharper.

C. We use power spectrum to evaluate the quality of an image. Power spectrum can be seen as a theoretical diffraction pattern of the image in A.

D. Corrected and sharpened images from A. As indicated in C, the corrected picture clearly reveals potential information that can be resolved as better as 3Å.

(Cheng et al, Cell, 2015)
Computed Diffraction Pattern

\[ F^2(s) \cdot CTF^2(s) \cdot Env^2(s) + N^2(s) \]

- Structural Factor
- Envelope Function
- Contrast Transfer Function
- Background Noise
CTF Assessment (Power Spectra)

Good:
• Isotropic
• Thon rings at high resolution

Bad:
Thon rings only at low resolution

Bad:
Missing Thon rings at certain direction due to drift (can be corrected if movies are recorded)

Bad:
Elliptic Thon rings due to astigmatism (can be useful if properly processed)
Motion Correction

- Stage drift
- Beam-induced sample motion

(Shilov et al, J Struct Biol, 2012)

(Zivanov et al, IUCrJ, 2019)
Particle Selection & 2-D Classes

Picking particles is the first step of cryo-EM image processing. It allows researchers to generate a library of various shape of 2-D projected images. In general, one can pick by naked eyes (left), but these days, for millions of such particle images, we use some computer algorithm to help pick particle images (right).
Based on the automated particle selection in the last slide, 50 2-D classes were generated. This process is called "2-D classification". It shows us what types of 2-D pictures are and whether the automated picking gives us redundant, useful, or useless 2-D images. In the case here, reading from left to right in rows, we see the first 16 classes look reasonable and can be used for further data processing. The rest 34 images either do not have sufficient information or simply garbage.
From 2-D images to a 3-D model, there are generally two ways to construct the initial model and refine the final model. Random conical tilt (left) uses the imaginary angles rendered by each different 2-D projections to rebuild the 3-D picture. Since 2010s, people also started to use machine learning (right) to generate initial model and refine the final models.
This makes people suspect that there are more shape changes (i.e., conformations) in this fuzzy area.

Voilà! There are many conformations.

(Scheres, Meth Enzymol, 2016)
Resolution Assessment

- Split particles into even and odd halves, reconstruct and compare models

- Early literature (<2005): differential phase residue

- Now: Fourier Shell Correlation (FSC)
  - Easy to compute. Invariant to filtering/sharpening level
  - Threshold? (0.5, 0.33, 0.14, 3σ)
  - Masking?
  - Split data? When?

$$FSC = \frac{\sum (F_1 \cdot F_2^*)}{\sqrt{(\sum |F_1|^2)(\sum |F_2|^2)}}$$

$$FSC = \frac{SNR}{SNR + 1}$$

Similar concept to CC$_{1/2}$ as discussed in X-ray crystallography.

Think like % error.
Resolution Assessment: example

FSC = 0.143, Å⁻¹ ~ 0.28, ~3.6 Å

FSC = 0.5, Å⁻¹ ~ 0.26, ~3.8 Å

Difference criteria lead to different resolution estimates.

(Cheng et al, Cell, 2015)
Model Validation:

- Average map value at atom positions
- **Map-model Cross-correlation** (in real space)
- Map-model FSC (in reciprocal space)
- EMRinger
- Q-score
- ...

Molmap 2Å  Molmap 3.5Å

\(S^2C^2\) Workshop, Stanford
Model Validation: Model-Map Correlation Coefficient

- Typically referred as CC, map CC, map correlation or real-space correlation
  A metric to show how well the model fits the map.
  (Equivalent to R-factor for crystallography)

Table 3
Summary of map correlation coefficients used in this work.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Region of the map used in calculation</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC-box</td>
<td>Whole map</td>
<td>Similarity of maps</td>
</tr>
<tr>
<td>CC-mask</td>
<td>Jiang &amp; Brünger (1994) mask with a fixed radius</td>
<td>Fit of the atomic centers</td>
</tr>
<tr>
<td>CC-volume</td>
<td>Mask of points with the highest values in the model map</td>
<td>Fit of the molecular envelope defined by the model map</td>
</tr>
<tr>
<td>CC-peaks</td>
<td>Mask of points with the highest values in the model and in the target maps</td>
<td>Fit of the strongest peaks in the model and target maps</td>
</tr>
<tr>
<td>CC_vr_mask</td>
<td>Same as CC_mask but atomic radii are variable and function of resolution, atom type and ADP</td>
<td>Fit of the atomic images in the given map</td>
</tr>
</tbody>
</table>

(Afonine et al, Acta Cryst D, 2018)
**Table 1** – data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Titan Krios 300kV, FEI Falcon II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel size (Å)</td>
<td>1.39</td>
</tr>
<tr>
<td>Defocus range (µm)</td>
<td>-0.5 to -3.5</td>
</tr>
</tbody>
</table>

**Reconstruction (RELION)**

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Membrane Domain</th>
<th>Peripheral Arm</th>
<th>64k class (for 42kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of rotations (°)</td>
<td>0.573</td>
<td>0.711</td>
<td>0.728</td>
<td>0.591</td>
</tr>
<tr>
<td>Accuracy of translations (pixel)</td>
<td>0.308</td>
<td>0.400</td>
<td>0.400</td>
<td>0.325</td>
</tr>
<tr>
<td>B-factor from post-processing</td>
<td>-88</td>
<td>-85</td>
<td>-83</td>
<td>-89</td>
</tr>
<tr>
<td>B-factor for map visualisation</td>
<td>-100</td>
<td>-150</td>
<td>-100</td>
<td>-120</td>
</tr>
<tr>
<td>Final resolution (Å)</td>
<td>3.9</td>
<td>4.1</td>
<td>3.9</td>
<td>4.0</td>
</tr>
</tbody>
</table>

**Model refinement (PHENIX)**

<table>
<thead>
<tr>
<th></th>
<th>Complete model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution limit (Å)</td>
<td>3.9</td>
</tr>
<tr>
<td>Number of residues</td>
<td>8037</td>
</tr>
<tr>
<td>Map CC (whole unit cell)</td>
<td>0.758</td>
</tr>
<tr>
<td>Map CC (around atoms)</td>
<td>0.782</td>
</tr>
<tr>
<td>Rmsd (bonds)</td>
<td>0.009</td>
</tr>
<tr>
<td>Rmsd (angles)</td>
<td>1.04</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>86.0</td>
</tr>
</tbody>
</table>

**Validation**

<table>
<thead>
<tr>
<th></th>
<th>24.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-atom clashscore</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0.5</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>12.5</td>
</tr>
<tr>
<td>Favoured (%)</td>
<td>87.0</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>0.1</td>
</tr>
<tr>
<td>Molprobity score</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Pros and Cons (X-ray v.s. EM)

• Pros:
  • Except MicroED/2D crystals, no need of crystallization
  • Smaller demand of purified proteins than X-ray crystallography
• Cons:
  • Molecular size:
    • >200 kD (100-200 kD, pushing the limit)
  • Resolution:
    • Mostly 3-5 Å
    • Overfitting
  • Conformational variability (same issue for X-ray crystallography)
    • Only a small number of functional states are solved.
    • Preferred orientation (single particles in particular)
  • Limited validation criteria
Cryo-EM Friendly Samples

**Repeating assemblies**

- **2D crystals**
  - (high tilts needed for 3D)

- **Helical filaments or tubes**
  - (no tilts needed)

**Single particle**
- no tilts needed, if randomly oriented

**Icosahedral viruses**
- (or other symmetrical particles)

**Asymmetric**
- (>200kDa)
  - (e.g. ribosomes)

Saibil, Acta Cryst. 2000, D56:1215
Challenges & Opportunities: Dynamic States

- Direct capture of macromolecular complexes from small volume cell culture
- Determine atomic structure of all states (N >> 1) in the mixture
- Order the states and elucidate the complete functional process

Only small number of discrete states are solvable now

Conformational variability

One conformation per image; perfect and simple.

Still OK, if three conformational possibilities per image, then can still sort out the best possible one.

In reality, several conformations mix together for one image, making it hard to tell what specific conformation the image represents.
It’s all about S/N ratio (SNR).

Resolution Revolution?
Started 40 years ago.
More to come.
Cryo-EM v.s. X-ray Crystallography

Friends and Rivals

a. A short loop (Met22 – Asn24) in the β subunit from the crystal structure does not fit well into the final 3D density map, but is easily corrected. While not in a crystal contact, these residues do have much higher temperature factors in the 3.4 Å crystal structure than the average. b. We remodeled this loop to fit better into our 3.3 Å density map.

(Li et al, Nat Methods, 2012)