Week 4

Structural Biology

4.1 Tools of Structural Biology

- Today's lecture, for time's sake, will not focus on structural biology so much as it will focus on the *tools* of structural biology.
- **Structural biology**: The determination of the 3D structures of biomolecules and the study of their structure-function relationships, aimed at understanding the molecular mechanisms of biomolecules' functions and interactions.
- Today, we will focus on X-ray crystallography, NMR, and electron microscopy.
  - These are direct methods for the determination of 3D structure.
- Different length scales in biology.
  - E.g., small molecules through eukaryotic cells.
  - Different techniques have different ranges over which they can be useful for determining structure.
  - NMR, then X-ray crystallography, then single particle cryo-EM, then cryo-electron tomography (for organelles), then light microscopy (from small to large).
- X-ray diffraction played an essential role in the early days of molecular biology.
  - Before crystallography, X-rays were still used to detect and describe compounds such as keratin from hair.
  - Based on these diffraction patterns, Pauling was able to describe the basic elements of protein structure (e.g., α-helices and β-pleated sheets).
    - Pauling was able to guess the structures from very rudimentary, blurry data, such as...
      - ➢ 9.6 Å: Radius of an α-helix;
      - ➢ 4.6 Å: The distance between hydrogen-bonded strands in a β-pleated sheet.
    - This is what made Pauling a great chemist.
  - Rosalind Franklin's photograph 51. 3.4 Å corresponds to the stacking between the bases.
    - UChicago has a graduate course dedicated to interpreting X-ray diffraction patterns.
    - Zhao recommends we read Watson and Crick's original 1953 paper: "Molecular Structure of Nucleic Acids: A structure for Deoxyribose Nucleic Acid."
    - Dickerson in 1980: Crystal structure analysis of a complete turn of B-DNA.
    - Review by Eisenberg in 2003: The discovery of the α-helix and β-sheet, the principal structural features of proteins.
      - ➢ Eisenberg was Zhao's grad mentor.

- Modern X-ray crystallography.
  - Dickerson crystallized DNA, took an X-ray diffraction pattern, and thus was able to determine the position of every atom in DNA.
  - In-house X-ray sources vs. synchrotron X-ray sources, like the over 1 km loop building at Argonne, used for biomolecule characterization.
- Examples of high-quality protein single crystals shown.
- Workflow for macromolecular crystallography.
  - Grow a crystal, take it to a synchrotron, do an exposure to X-rays, rotate the crystal a degree or two, and do another exposure. Zhao had to fly from LA to Chicago to use Argonne's synchrotron while in grad school!
  - From this "movie" of exposures, you can get the electron structure and, with practice, resolve that into amino acids and other atoms.
  - You then fold the atom/amino acid sequence into a protein.
  - This whole workflow takes about 1 day today.
  - In the 80s-90s, this would be a graduate student's 4-5 year project and, if successful, would likely result in a *Nature* publication.
- Bottlenecks of macromolecular chemistry.
  - You need to clone the gene and express it to get the protein (most proteins, save a few such as RuBisCO, cannot be purified in sufficient quantities from natural sources). Cloning success rate: 100%. Expression success rate: 66%.
  - Then you need to purify it. Success rate: 35%.
  - Then you need to grow diffraction-quality crystals or take an NMR spectrum. Success rate: 29.5%.
  - The latter two are the biggest bottle necks; you lose a lot of your starting material in each one.
  - Crystallization is difficult because (most) proteins did not evolve to be crystallized, they evolved to function.
- The future of X-ray crystallography.
  - Argonne is the best synchrotron in the United States.
  - X-ray free electron laser (xFEL). LCLS (Linear Coherent Light Source) close to Stanford. 2-mile tunnel under the 280. You shoot electrons down a tube, vibrate them with magnets along the length of the tunnel so that they emit X-rays, trap the electrons at the end, and then just make use of the directed light. You get a super powerful beam (a billion times stronger than third generation synchrotrons).
  - Zhou and Zhao compared resolution from APS and LCLS on the same (type of) crystal; LCLS has higher resolution.
    - The beam is so strong that you damage the crystal though; at every point, you can only take one shot.
  - Radiation damage free-diffraction: Super fast exposure; take your diffraction before you cause damage.
- Future of macromolecular crystallography: microED.
  - You don't have to grow diffraction-quality crystals here (which are about 5 µm in size).
  - You can use nanometer-scale crystals instead.
  - microED: Micro-electron diffraction.

- NMR.
  - You still need to purify protein, but then you do NMR sample prep, acquire data, process the spectrum, and then do structural analysis.
  - NMR does not give you a map; it gives you the distance between different atoms. If you have enough of these constraints, you can calculate a structure.
  - In chemistry, you typically collect one-dimensional spectra; in biochemistry, you typically collect three- or four-dimensional spectra.
    - Most OChem spectra are 1-dimensional.
    - 2D spectra: You have to define a specific sequence to get resonance of both types of molecules together.
    - 3D includes one more type of atom that you want to resonate.
    - 4D is if you introduce some radiofrequency change over time, providing another dimension of information.
    - Physically, this is the most difficult technology.
    - The physics behind NMR is the toughest, as it dips into quantum mechanics.
  - Advantages:
    - Protein in native environment (crystal packing in XRD might introduce artifacts).
    - Information on dynamic structure (more on this later).
    - Information on protein interaction with other biomolecules.
      - ➢ For example, you can run an NMR of the protein and then of the protein mixed with some ligand.
      - ➢ This tells you how the protein interacts with the ligand.
      - ➢ This is a very hard experiment to run with XRD because you have to soak the crystal in ligand or ligate the protein and then crystallize it. In the words of Zhao, this is a "pain in the ass" to do.
  - Limitations:
    - Isotopic labeling required (^{13}C and ^{15}N at least).
    - Difficult for large proteins or complicated folding structures (if it's a large protein, you'll be stuck into a local minimum).
    - You need amino acid sequence information (including PTM) in advance (and the protein usually has to be 10-20 kDa).
  - Reason to run an NMR experiment: Not for a *de novo* 3D structure, but for that ligand-protein interaction.
- NMR is good for probing interactions.
  - Every peak in a 2D N-H spectrum corresponds to an amino acid (amide bond).
  - You can see a shift in proteins as they're ligated.
- The future of NMR.
  - Feed a cell nutrients and take an NMR of a whole cell.
- Recent revolution in structural biology.
  - 2017 Nobel prize for cryoEM[1].
- Two major techniques: SPA and CryoET.
  - SPA is single-particle analysis. Embed viri in a tray, take a 2D projection, do FTs, and reconstruct the 3D structure. Multiple particles averaged.

[1] What dad mentioned years ago?

  - Cryo-electron tomography/STA (single tilt analysis; there used to be double tilt analysis but it was too hard to keep the stage stable): Focus on one single particle, but tilt the stage within the microscope. Usually used to look at subcellular organelles.
  - SPA is already comparable to XRD in its ability to generate atomic-level resolution.
  - Workflow: Aqueous solution → sample vitrification (rapid cooling) → low-dose image collection by cryoEM → SPA or cryoEM → structure.

- Methods for SPA have been developed for decades.
  - Began in the 1970s with reconstruction of an icosahedral structure.
  - Modern cryoEM began around 2011.

- Problems that prevented high-resolution cryoEM reconstruction.
  - Radiation damage: Only limited amount of electron dose can be used since light atoms, e.g., hydrogens can be damaged.
  - Bad detection: Only limited amount of signal is recorded.
  - Beam-induced motion: Hard to avoid.
  - Sample heterogeneity: No crystal lattice as constraints.
  - Problems 1-3 were solved with a good camera in 2012. Multiple exposures aligned and averaged (take multiple shots [a movie] over 4-5 seconds and then align the relative positions and average).

- Direct electron direction camera.
  - Used to use a CCD camera. The screen of an electron microscope shows a green fluorescent image[2]; to digitize the image, we use a CCD camera; detects photons only, so we convert electrons to photons with a scintillator. This causes a lot of signal loss, though, because of the conversion.
  - The invention of DDC gets rid of the scintillator and fiber-optic coupling, allowing literal direct detection of electrons.
  - DQE: Detective Quantum Efficiency goes up.

- SPA revolution.
  - The structures that SPA focused on were ones that were very difficult or impossible to crystallize.
  - People have tried to crystallize membrane proteins for decades, but in 2010, SPA cryoEM gave another way.

- Workflow for single-particle cryoEM.
  - Prepare the sample and then plunge it into liquid ethane.
  - Ethane has a very large heat capacity (not liquid nitrogen). We don't want crystalline ice; we want vitrous ice, so that the ice crystal doesn't affect the experiment.

- Advantages of cryoEM analysis:
  - Removes the crystallization bottleneck.
  - Dynamics can show you different states of a molecule.
  - For example, this is how we figured out the different conformations/rotation of ATP synthase.

- Future of cryoEM: Cryo-electron tomography.
  - Currently mainly used to look at larger structures, e.g., organelles.
  - Used to study how SARS-CoV-2 infects cells.

[2] Think of the TEM machine in the GCIS sub-basement.

    - Allows for a better understanding of its S-proteins, as well.

- CryoEM's limitations.
  - The sample cannot be too thick.
  - A eukaryotic cell is typically too thick.
  - Circumventing this: FIB milling (focused ion beam). Takes off part of the cell.
  - Put everything in an SEM (to guide your progress), do the milling, and then transfer to a TEM.

- Comparison of structural biology techniques.

		single-particle cryoEM	**X-ray crystallography**	**xFEL**	**microED**	**NMR**
Setup	imaging source	electron	X-ray	X-ray	electron	Magnetic field and RF pulses
	lens system	yes	no	no	yes	no
Sample	form	solution	crystal	micro-xtal / xtal in cell	micro-xtal	solution
	quantity	low	high	high	low	high
Throughput	sample screen	low	high	high	low	low
	data collection	days	minutes	hours to days	hours	days
	data processing	weeks	days	days-weeks	days	weeks
Limit	resolution	up to 1.2 Å	better than 1.0 Å	better than 1.0 Å	better than 1.0 Å	N/A
	MW	> 60 kDa	no	no	no	< 100 kDa
Pain Point		screen freezing conditions	growing crystals	growing a ton of micro-xtals	growing micro-xtals	isotopic labeling
Unique Benefit		multiple states	anomalous scattering	radiation damage free	few micro-xtals	dynamic information

Table 4.1: Comparison of structural biology techniques.

- Graduate course (though 1-2 undergrads take it every year): BCMB 32600 Methods in Structural Biology.
  - Did SARS-CoV-2 last time.

- Quick survey of Cross-β diffraction pattern and amyloid.
  - Zhao's research topic as a graduate student.
  - Difference between β-pleated sheets (XRD points expand out linearly) and cross-β patterns (XRD points expand out perpendicularly), the latter of which are generated by amyloids.
  - Eisenberg published seven peptides whose structure they determined with XRD and which ran perpendicular to the fibers.
  - Fibers are like 1D crystals which are very hard to crystallize. Breakthrough in 2015: used microED to determine the structures of the very small "invisible" (under light microscope) crystals.
  - Solid-state NMR helps, too.
  - cryoEM helps more.

- Nobel prizes in 1962.
  - Crick was a grad student of Perutz; Perutz and Kendrew determined the first crystal structures of protein.

  - Wilkins was the mentor of Franklin; she had already passed away by 1962. She was not recognized; women are still not recognized enough.

- Future Nobel prize in Zhao's evaluation.
  - John Jumper (former grad student at UChi) develops AlphaFold2.
    - Combines multiple sequence alignment (MSA, genetic information) with pair representation (distance matrix, analogous to NOE spectrum, structural geometrical information) as the input.
    - Introducing attention-based neural-network architecture...
- Think of a neural network as a large machine with a lot of knobs.
  - Once the knobs are tuned with existing data, the machine is capable of predicting, decoding, and analyzing unknown data.
- AlphaFold2 data flow.
  - The key step is MSA + pair as input.
- Prediction of a human methyltransferase that has not been crystallized with a high confidence.
  - AlphaFold2 predicted the structure, however! Collaboration between Zhao and Chuan He.
- Try it with your own protein using ColabFold via Google.
- Current limitations.
  - Not implemented for nucleic acids.
  - Static structures.
  - Sequence length limit.
  - Insensitive to point mutations.
  - Poor performance for antibody recognition.
- One critical reason why AlphaFold is so successful.
  - The database has gotten huge in the last several decades. High quality experimental training sets are available.
  - Another advantage is high quality sequence technology.
- Remaining challenges: Complex structures, dynamic structures, and intrinsically disordered proteins.
  - Still difficult due to issues computing the energy landscape of large biological complexes.
  - Nuclear pore: Complex with over 1000 proteins.
  - Science: Volume 376, issue 6598, 10 June 2022 reviews research surrounding the nuclear pore.
  - How do we extract the dynamic information, regardless?
- Is structural biology still cool?
  - Various reasons it's still needed.
- Last class, we talked about Ramachandran plots.
  - Ramachandran statistics is used as a validation method in X-ray crystallography and single-particle cryoEM. Not used in AlphaFold.
- Keep my eye on Nick Korn; seems to really know what's going on and asks good questions.
- More info will be provided later on how this info will be incorporated into future exams.
- Reach out to Zhao to talk more about his work! Seems very related to analytical chemistry. What is his view of the field and how my math background can help?