

Biological Chemistry Laboratory
Biology 3515/Chemistry 3515
Spring 2023

Lecture 29:

Cryo-electron Microscopy
and
Protein Structure Prediction

25 April 2023

©David P. Goldenberg

University of Utah

goldenberg@biology.utah.edu

Special Thanks to the 2023 TAs and Lab Instructor!

- Calder Lake
- Leon Guerra
- Erik Smith
- Juli Kim
- Adam Rupper

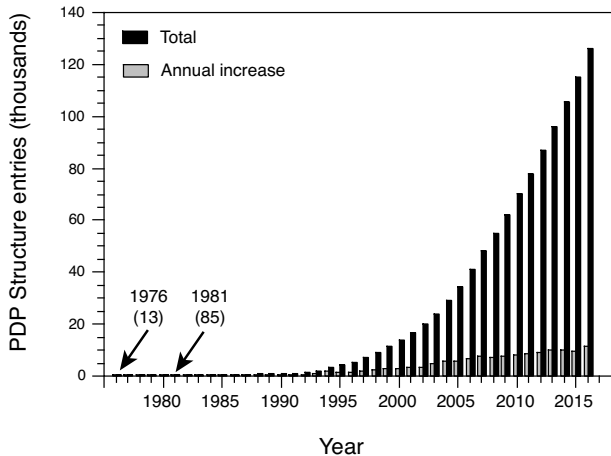
Announcements

- Final Exam:
Monday, 1 May
10:30–11:30 AM
CSC 208
- Review Session with the TAs
Thursday, 27 April
5:15 PM
ASB 210

Free (almost) Clicker Points!

If $\geq 80\%$ of the class submits student feedback, everyone will receive 10 bonus clicker points.

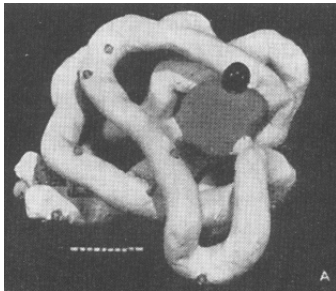
The Protein Data Bank Since 1976



What happened in the 1990s?

- Genetic engineering:
Ability to make large amounts of many proteins.
- Synchrotron X-ray sources:
Much faster data collection.
- Bigger and faster computers.

The First Myoglobin Structure and Myoglobin Diffraction Data



- Calculation required 70 min on EDSAC computer, with \approx 1 KByte memory and 700 Hz clock speed.
- 1980s: VAX 11/780 computer, with up to 4 Mbyte memory and 5 MHz clock speed. (\$160,000)
- 2023 iPhone (budget model): 64 GByte memory and 3 GHz processor.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. & Phillips, D.C. (1958). *Nature*, 181, 662–666.

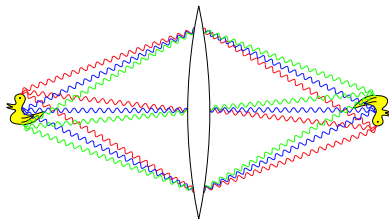
Dickerson, R. (1992). A little ancient history. *Protein Sci.*, 1, 182–186.

<http://dx.doi.org/10.1002/pro.5560010118>

Resolution: Not All Crystal Structures are Equal!

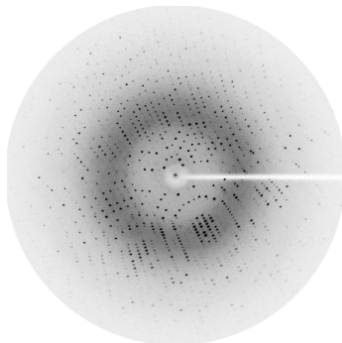
Resolution in optical images depends on:

- Wavelength of light.
- Quality of lens.
- Aperture of lens.



Resolution of crystal structures depends on:

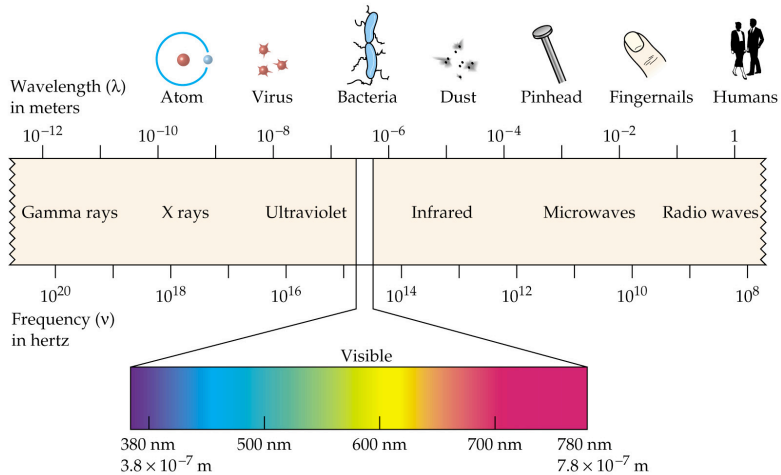
- Wavelength of X-rays (usually not a limitation).
- Quality of crystal, *i.e.*, how well-ordered the molecules are.
- Data with largest diffraction angle used in calculation.
- Accurate phase determination.



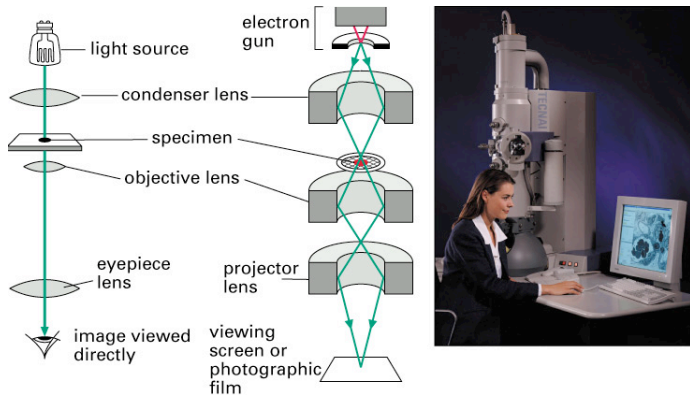
Challenges in Protein Crystallography

- Requires crystals!
 - Relatively large amounts of protein, mg or greater.
 - Search for crystallization conditions.
 - Especially difficult for large and dynamic structures, such as membrane proteins and molecular motors.
- Solving the phase problem:
 - Finding suitable heavy-atom derivatives.
 - Enhanced computational methods.
 - Becomes more difficult with larger structures.

The Electromagnetic Spectrum



The Transmission Electron Microscope



- First electron microscope built by Ernst Ruska in 1930
- 1986 Nobel Prize in Physics

Strengths and Limitations of Electron Microscopy

■ Strengths

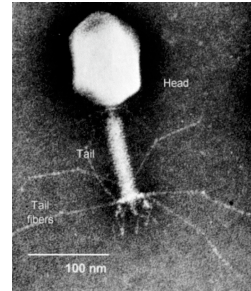
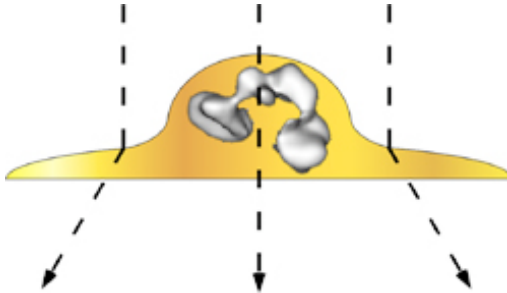
- Short wavelengths: $\lambda = 0.025 \text{ \AA}$ for 200 kV electron.
- Electromagnetic lenses can focus electrons and produce images directly.
- Theoretical resolution $< 1 \text{ \AA}$.

■ Limitations

- Vacuum required.
- Sample damage.
- Weak signal from light elements.

Negative stain

- Most useful for macromolecular structures such as viruses or protein complexes.
- Particles are embedded in a salt of a heavy metal, e.g., uranyl acetate.

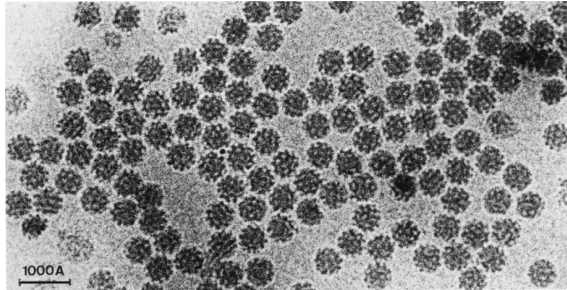
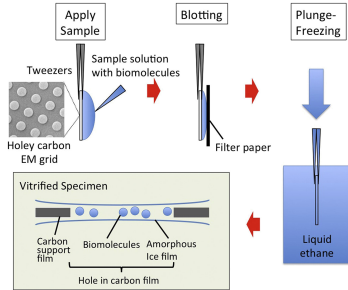


- Image represents a projection of places that are *not* occupied by the protein.
- Relatively high resolution, $\approx 10 \text{ \AA}$.
- Only shows surfaces, and structure can be distorted.

Diagram from http://www.snaggledworks.com/em_for_dummies

Electron micrograph of virus particle by Hans Ackermann.

A Breakthrough: Cryogenic Freezing of Unstained Biological Samples



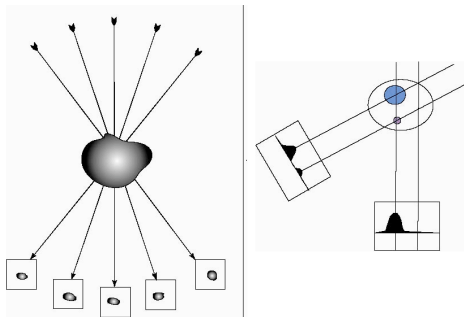
S

- Very rapid freezing - “vitreous ice”
- Low temperature minimizes radiation damage
- Very low contrast

Adrian, M., Dubochet, J., Lepault, J. & McDowell, A. (1984). Cryo-electron microscopy of viruses. *Nature*, 308, 32–36.

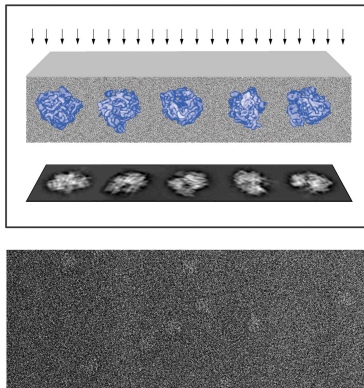
<http://dx.doi.org/10.1038/308032a0>

Electron Tomography: 3-Dimensional Reconstruction from a Tilt Series



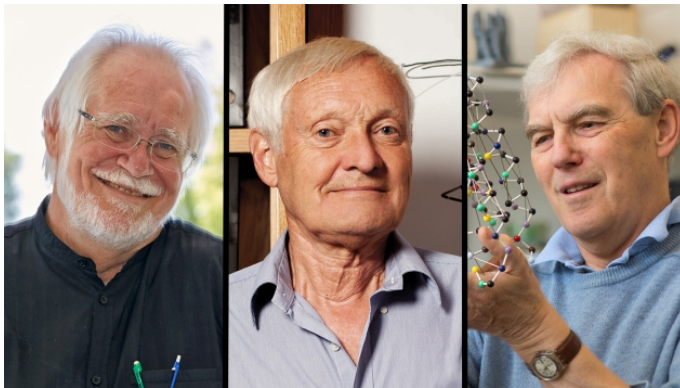
- Each image is like a shadow, *i.e.*, a projection onto two-dimensions.
- 3-dimensional structure is reconstructed from multiple projection views.
- Different views are generated tilting the sample in the microscope.
- Commonly used for cellular structures.
- Similar to an x-ray CAT scan (Computed Axial Tomography)

Reconstruction from Randomly-Oriented Single Particles



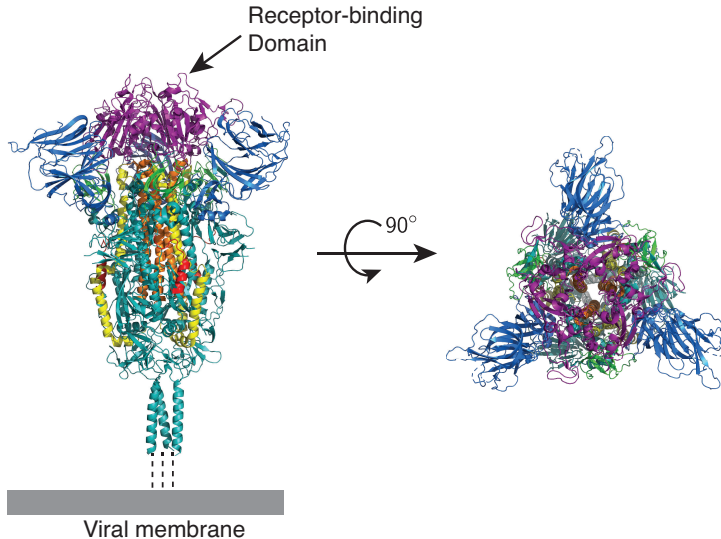
- Low-dose images of 1000s of individual particles are oriented and averaged.
- Computer program replaces crystallization!

2017 Nobel Prize in Chemistry



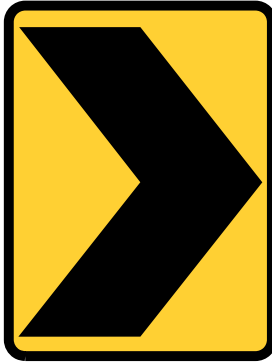
- Jacques Dubochet
- Joachim Frank
- Richard Henderson

Cryo-EM Structure of SARS-CoV-2 Spike



Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh Jr., R. M., Rawson, S., Rits-Volloch, S. & Chen, B. (2020). *Science*, 369, 1586–1592. <http://doi.org/10.1126/science.abd4251>, PDB entry 6XR8

Warning!

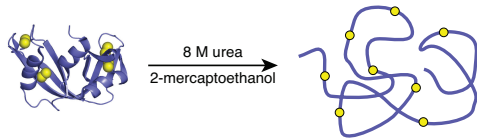


Direction Change

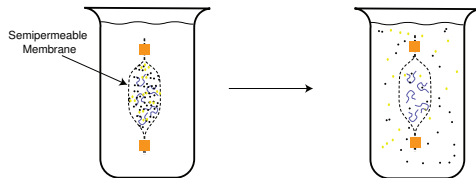
Protein Structure Prediction

The Anfinsen Experiment

- Unfolding and reduction of RNase A:



- Removal of urea and 2-mercaptoethanol by dialysis in the presence of O_2 :



- Recovery of active RNase A, with properly formed disulfides!
- Demonstrated that information required to specify three-dimensional structure is contained in the amino-acid sequence.
- Implied that structure could be predicted from amino-acid sequence.

Somme Approaches to Predicting Protein Structures

■ Hierarchical approach:

- Determine propensities of different amino acids to form α -helices and β -strands.
- Use propensities to predict segments of polypeptide chain that will form α -helices and β -strands.
- Assemble secondary-structure elements into overall fold.
- Doesn't really work!

■ Template-based modeling:

- Identify a protein with a sequence very similar to the protein of interest, and with a known three-dimensional structure.
- Adjust the known structure to accommodate the sequence of the protein of interest.
- Works pretty well when the template structure is 50% or more identical to the unknown structure, but accuracy is limited.

■ Physics-based modeling:

- Build a computer model of the polypeptide chain, in arbitrary conformation.
- Apply mathematical functions that describe all of the forces acting on individual atoms.
- Simulate process of sampling conformations to find those with minimum energies.
- Now feasible with very small proteins, but with high computational cost.

Inferring Residue-Residue Contacts from Co-evolution

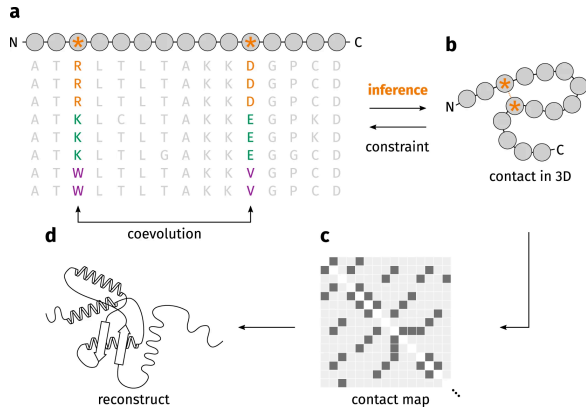


Figure from: Bittrich, S., Schroeder, M. & Labudde, D. (2019). StructureDistiller: Structural relevance scoring identifies the most informative entries of a contact map. *Scientific Reports*, 9, 18517.

<https://www.nature.com/articles/s41598-019-55047-4>

Göbel, U., Sander, C., Schneider, R. & Valencia, A. (1994). Correlated mutations and residue contacts in proteins. *Proteins: Struct. Funct. Bioinf.*, 18, 309–317.

<https://doi.org/10.1002/prot.340180402>

A Deep-learning Neural Network

Figure 1.6. Deep representations learned by a digit-classification model

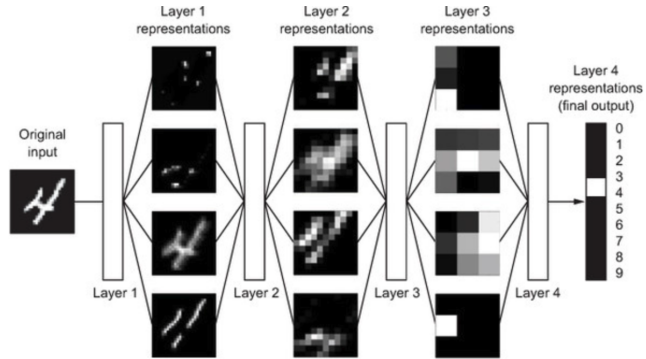
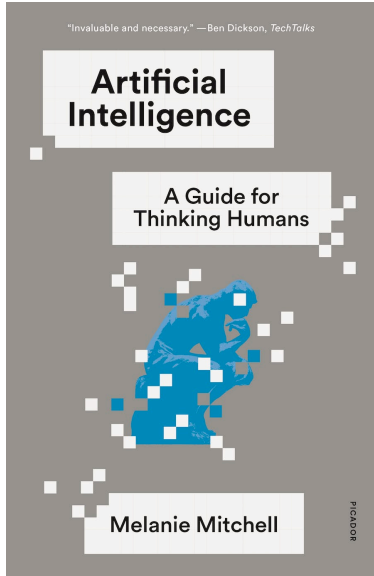
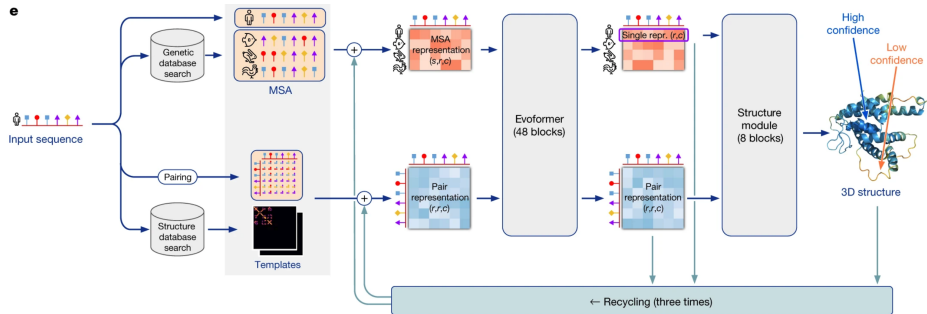


Figure from: Chollet, F. (2018). *Deep Learning with Python*. Manning.

A Recommended Book

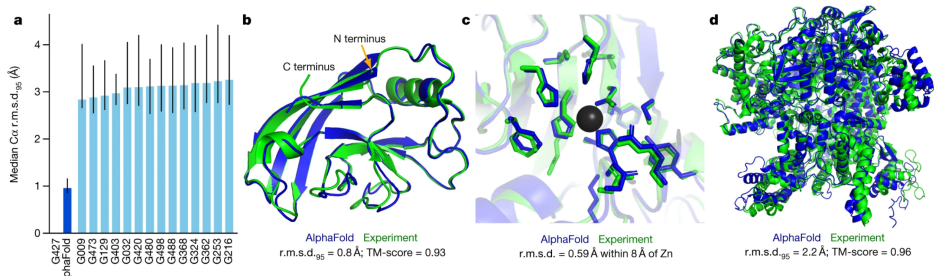


AlphaFold: Protein Structure Prediction Using Deep Learning



Jumper, J., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.
<https://doi.org/10.1038/s41586-021-03819-2>

AlphaFold Results



Jumper, J., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.
<https://doi.org/10.1038/s41586-021-03819-2>

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)–Round XIII. *Proteins: Struct. Funct. Bioinf.*, 87, 1011–1020.
<https://doi.org/10.1002/prot.25823>

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) Help: [AlphaFold DB search help](#)

Feedback on structure: [Contact DeepMind](#)

AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.