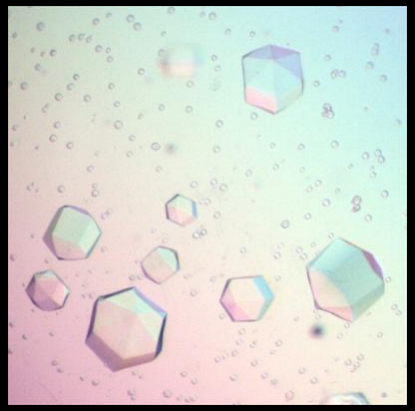
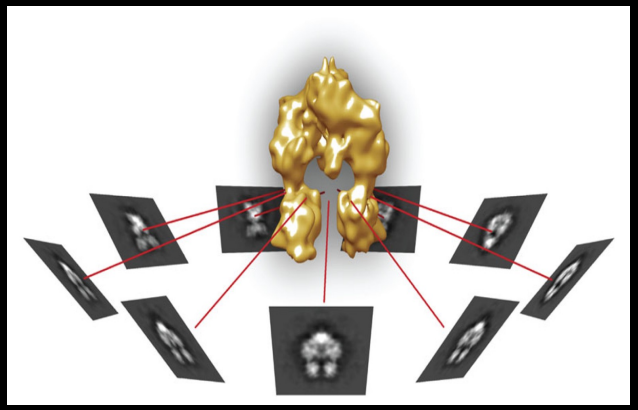
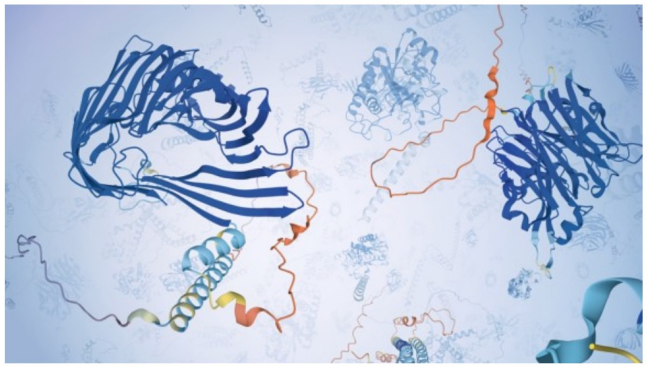




Structural Data:
X-ray Crystallography & Cryo-EM & AI

Jesse Rinehart, PhD

Biomedical Data Science: Mining & Modeling

CBB 752, Spring 2022

Cellular & Molecular Physiology
Yale University School of Medicine

YALE SYSTEMS BIOLOGY INSTITUTE

Yale Structure Courses:

MB&B529b / PHAR529b, Structural Biology and Drug Discovery

MB&B711b / C&MP711b, Practical cryo-EM Workshop

MB&B720a, Macromolecular Structure and Biophysical Analysis

C&MP 710b/MB&B 710b4, Electron Cryo-Microscopy for Protein Structure Determination

MB&B635a / ENAS518a, Quantitative Approaches in Biophysics and Biochemistry

Additional Resources:

"Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models"
Gale Rhodes (Third Edition, 2006 Elsevier/Academic Press)

"Crystallography 101" http://www.ruppweb.org/Xray/101index.html

"Single particle electron cryomicroscopy: trends, issues and future perspective."
Vinothkumar KR, Henderson R. Q Rev Biophys. 2016 pubmed:27658821

"Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity"
Eva Nogales & Sjors HW Scheres, Mol. Cell 015 May PMID: 26000851

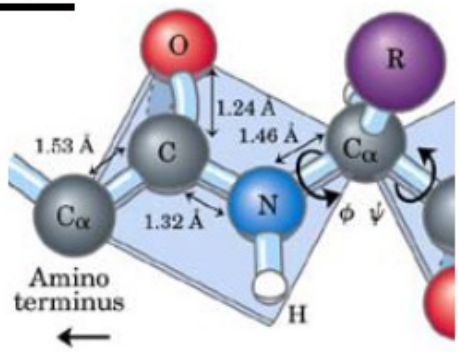
Thank you to **Yong Xiong** and **Fred Sigworth** for contributions to this lecture

"Just as we see objects around us by interpreting the light reflected from them, x-ray crystallographers "see" molecules by interpreting x-rays diffracted from them."
- Gale Rhodes

- There's a limit to how small an object can be seen under a light microscope.
- The diffraction limit: you can not image things that are much smaller than the wavelength of the light you are using.
- The wavelength for visible light is measured in hundreds of nanometers, while atoms are separated by distances of the order of 0.1nm, or 1Å.

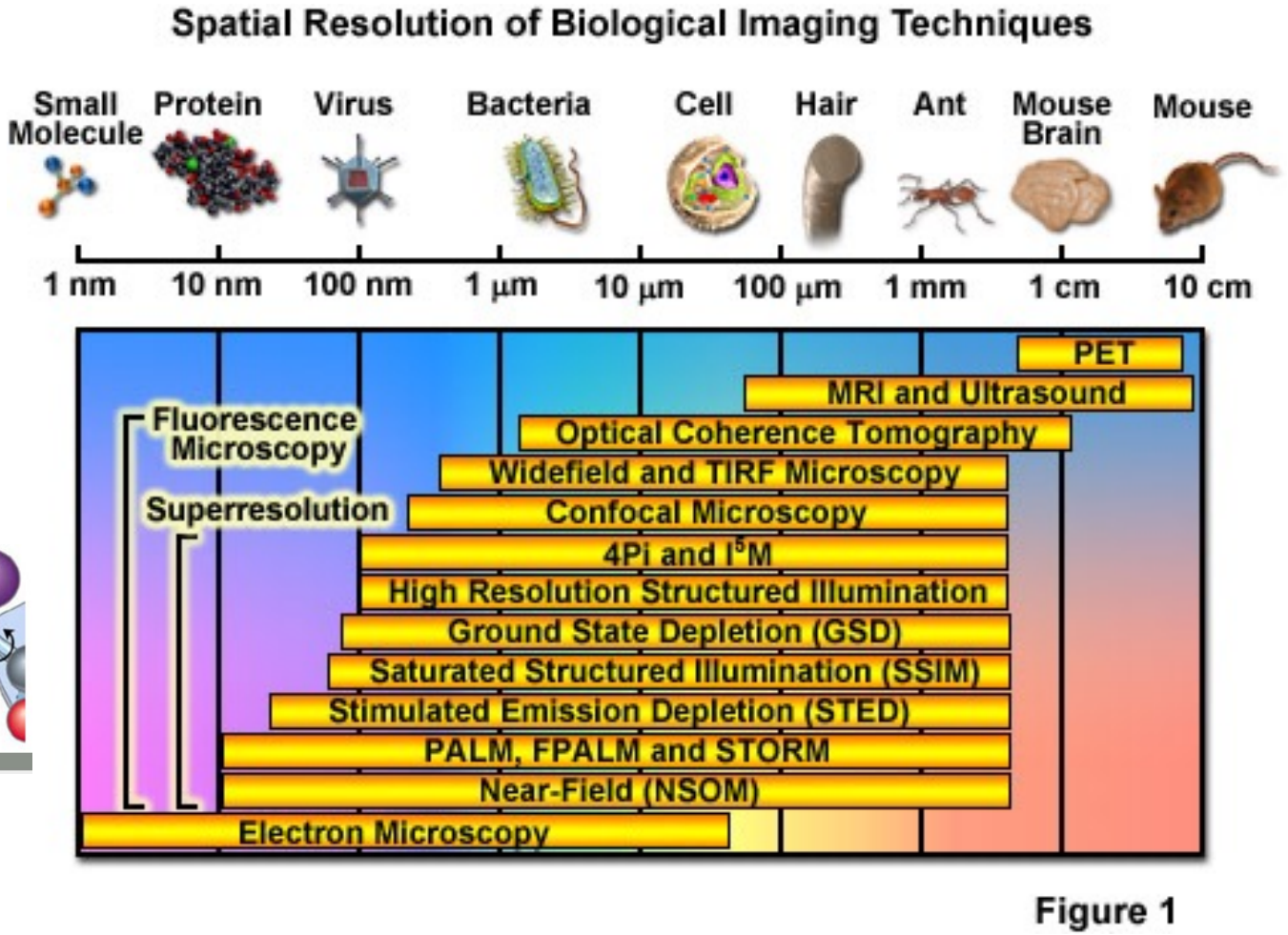
We need to use x-rays to resolve atomic features.

Distances between atoms are small:
Lab x-ray sources use CuKα radiation. Wavelength = 1.54 Å.
Synchrotron radiation wavelengths in the range 0.5 Å - 2.5 Å.



Yong Xiong

The 2014 Nobel Prize in Chemistry: Eric Betzig, W.E. Moerner, and Stefan Hell
"The development of super-resolved fluorescence microscopy"



1Å = 0.1nm

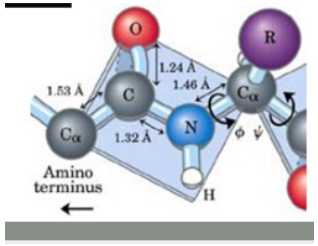
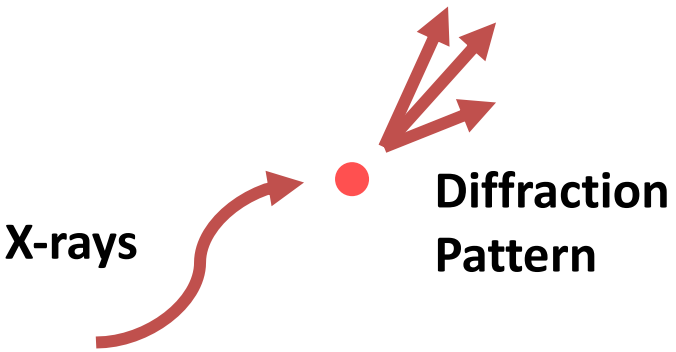


Figure 1

http://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction.html

Experimental Determination of Atomic Resolution Structures

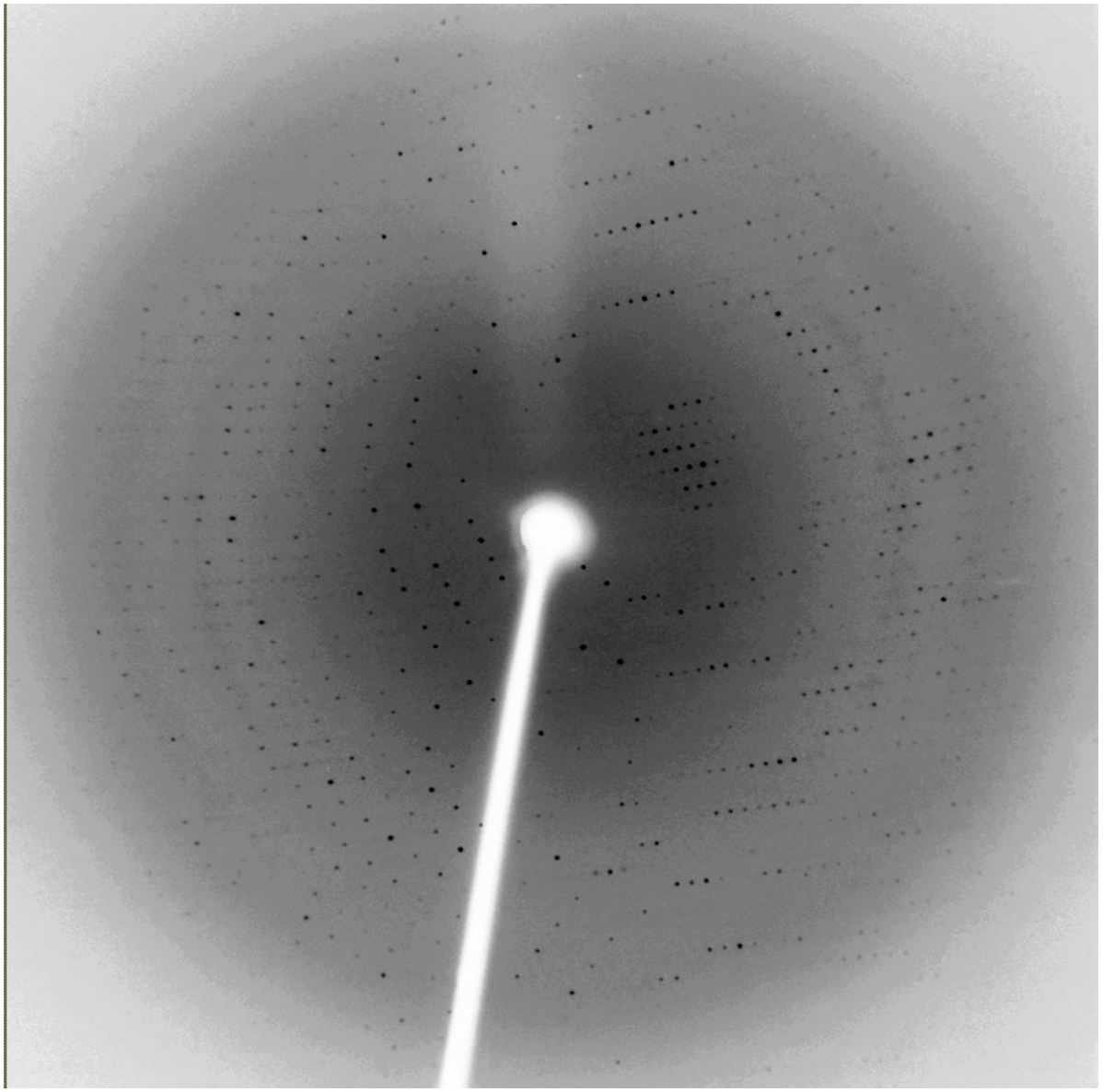
X-ray



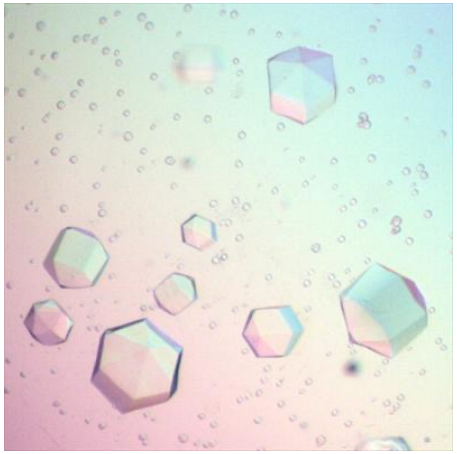
- Direct detection of atom positions
- Crystals required

Other methods for determining protein structures:
-EM (Electron Microscopy), Cryo-EM, ESR/Fluorescence

Image of X-ray diffraction of a protein crystal



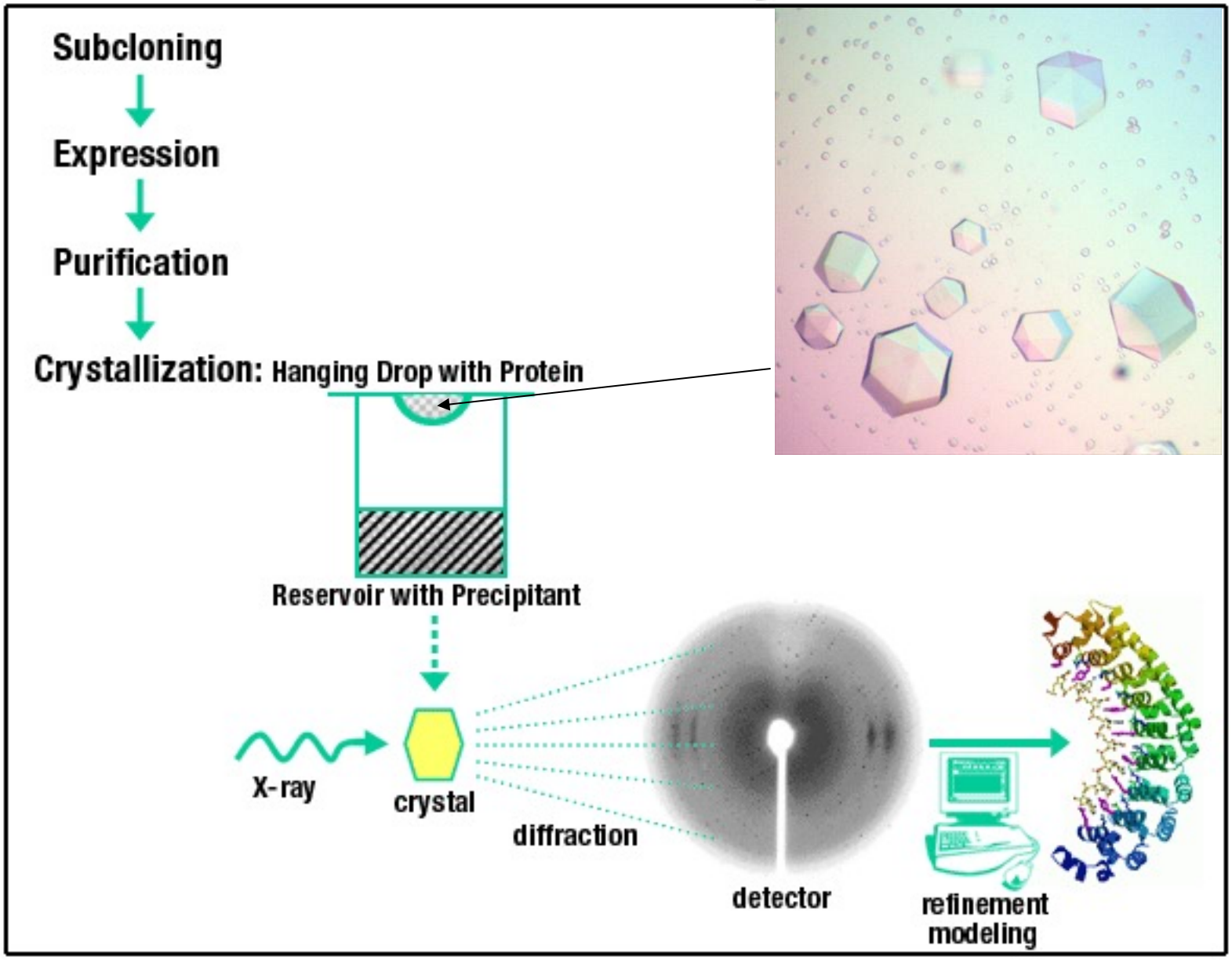
Why Crystals?



X-rays are scattered by electrons, too weak to record scattering from a single molecule. Crystals are therefore used because they present many molecules (N) in exactly the same orientation. The scattering from each of the N molecules interferes constructively to give a measurable diffraction pattern (enhanced ~N^2 fold).

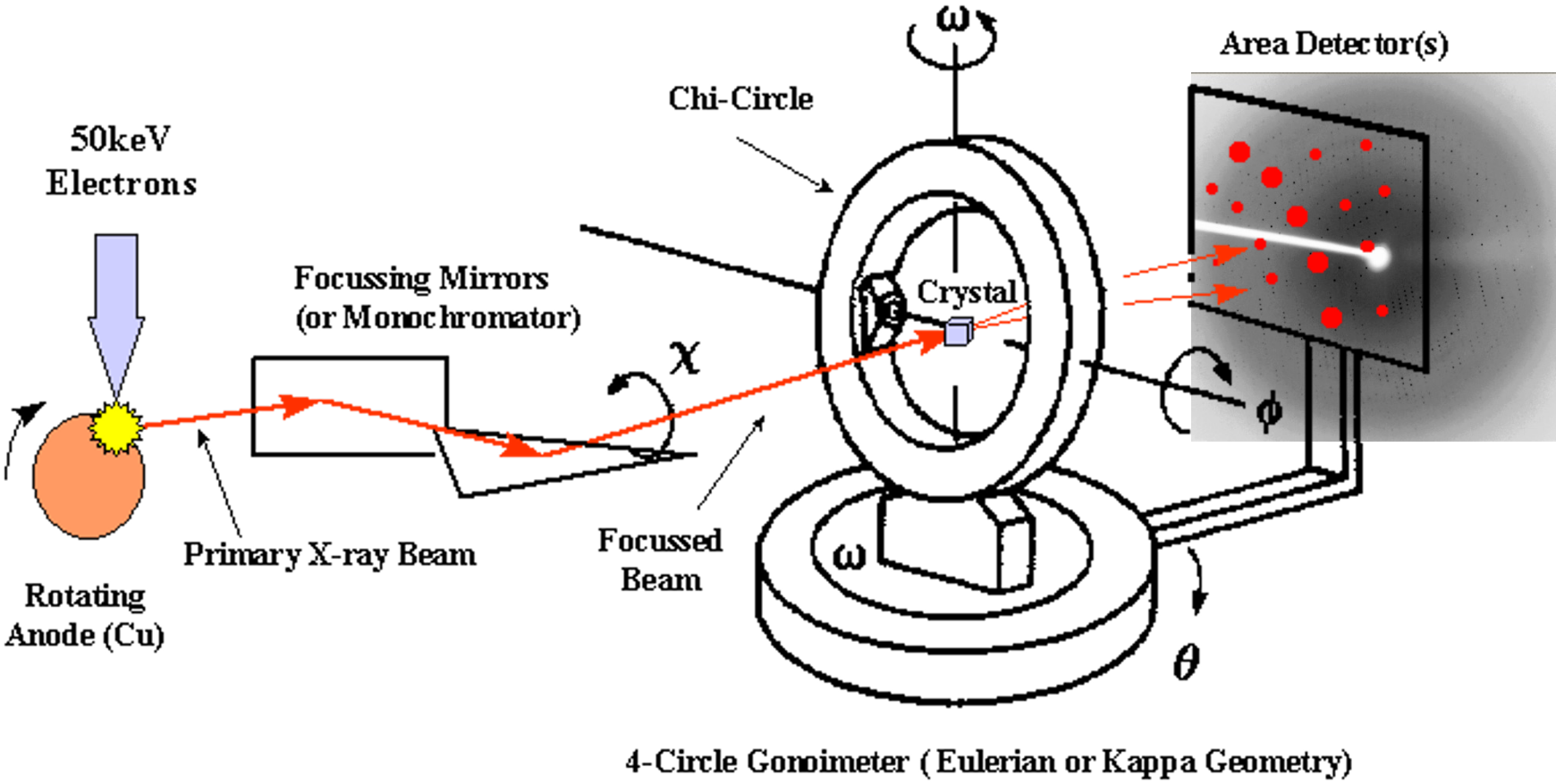
Yong Xiong

Determination of Protein Crystal Structure



http://www.noble.org/PlantBio/Wang/crystallography.html

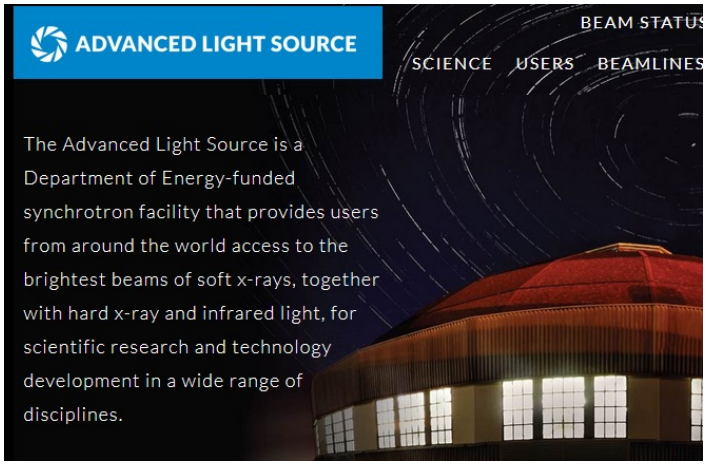
Data Collection



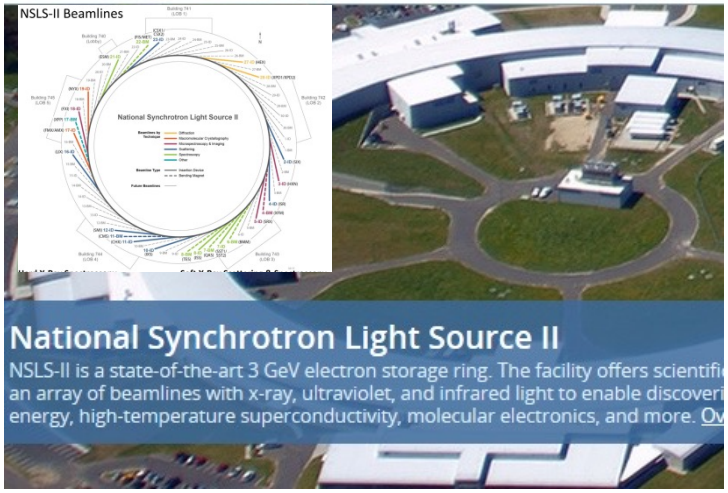
Synchrotron X-ray Sources are the method of choice

Lab x-ray sources @ 1.54 Å compared to Synchrotron X-ray @ 0.5 Å - 2.5 Å.

APS Chicago



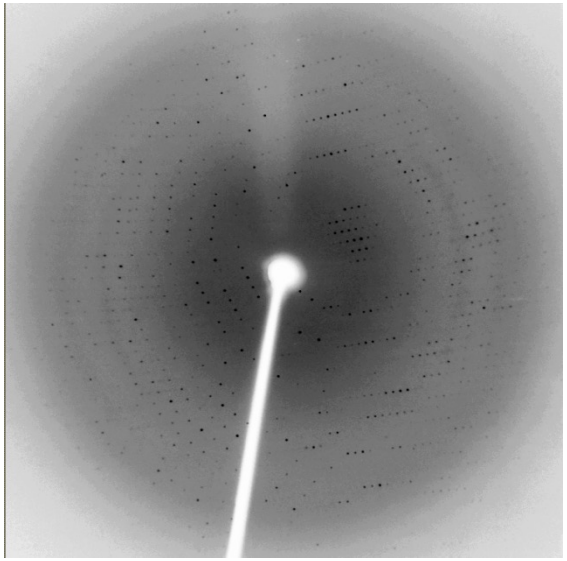
ALS Berkeley



NSLS-II Brookhaven

CHESS Ithaca

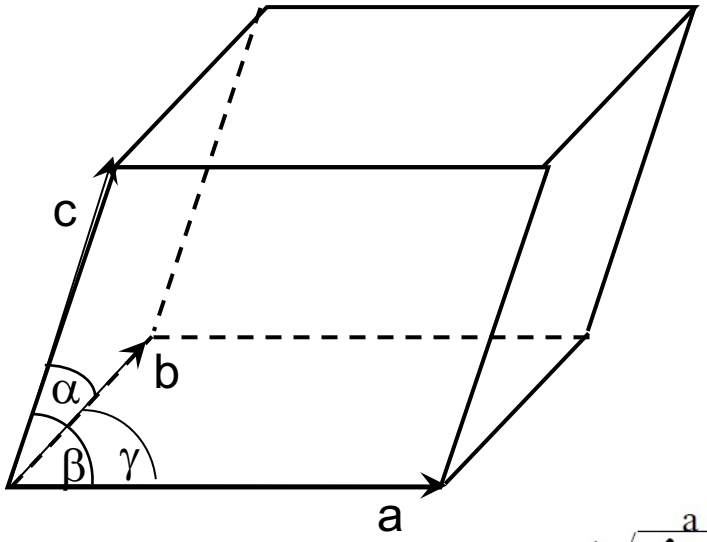
The information we get from a single diffraction experiment



Analyze the pattern of the reflections

(a) space group of the crystal

(b) unit cell dimensions

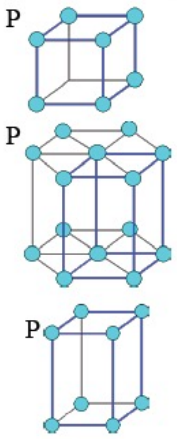


Cubic
a = b = c,
α = β = γ = 90°

Hexagonal
a = b ≠ c,
α = β = 90°, γ = 120°

Trigonal
a = b ≠ c,
α = β = 90°, γ = 120°

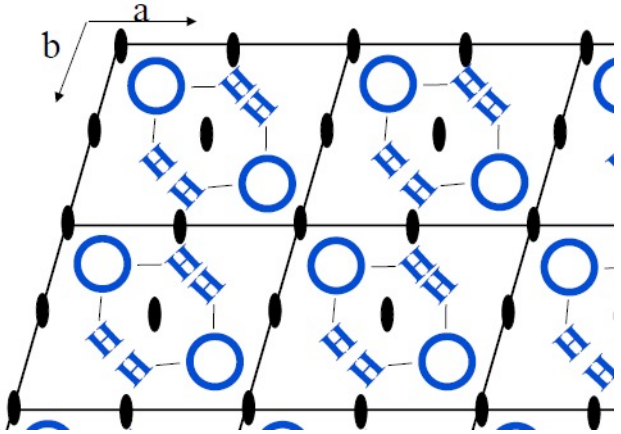
Tetragonal
a = b ≠ c,
α = β = γ = 90°

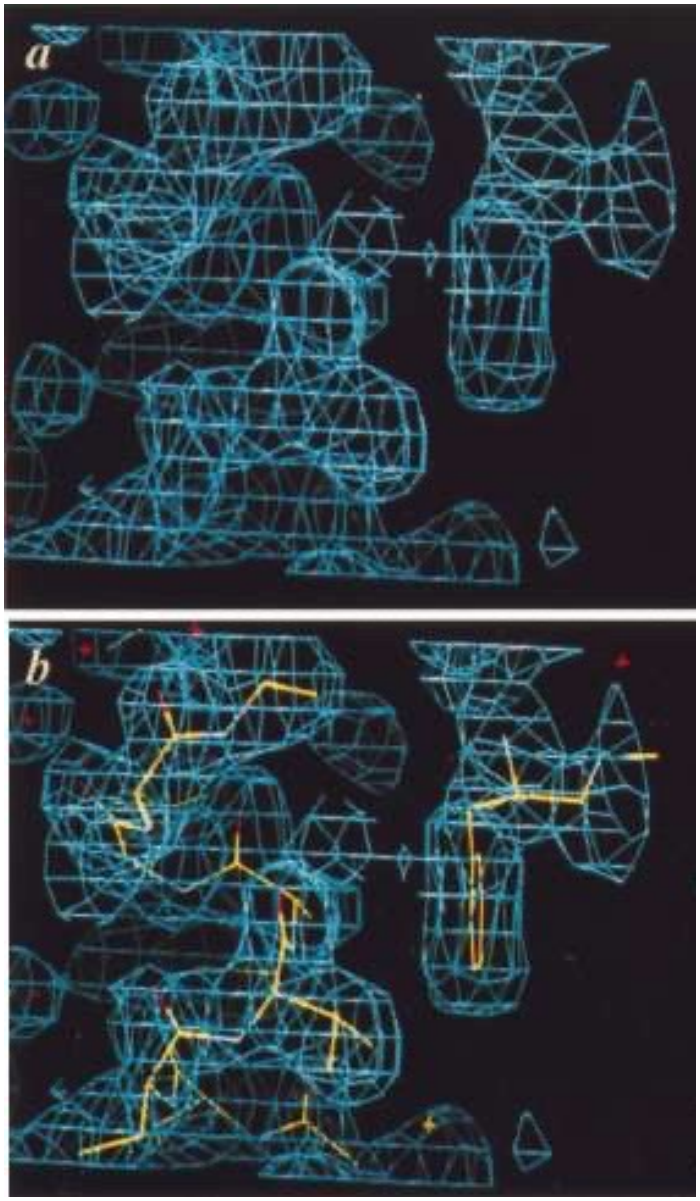


How to understand symmetry?

Crystal = lattice + unit cell content

(asymmetric units (asu) content)

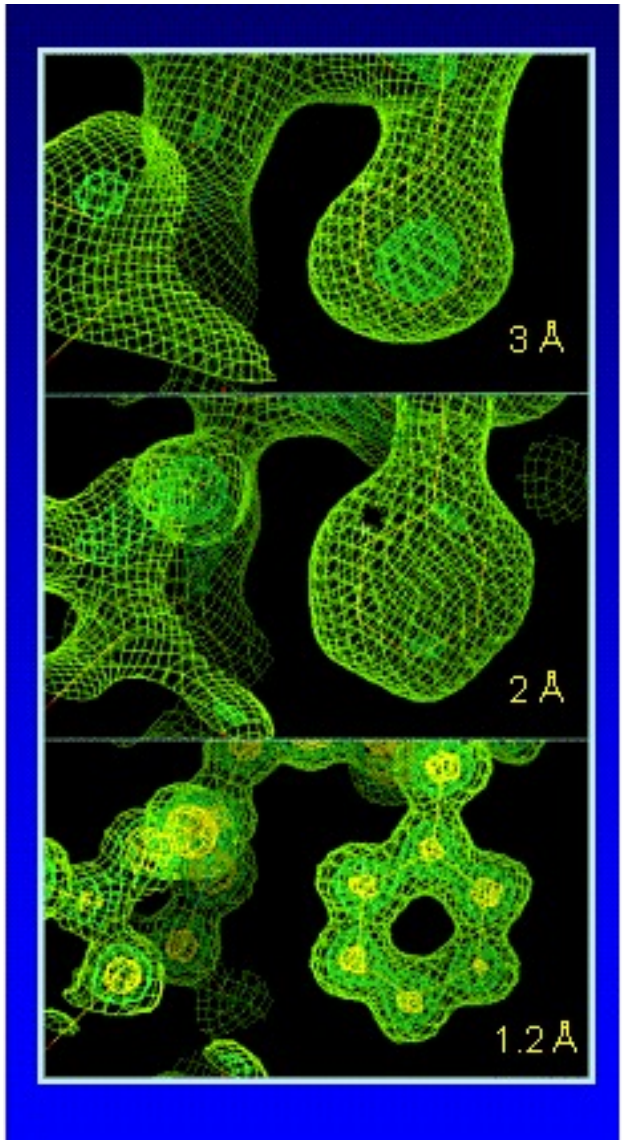




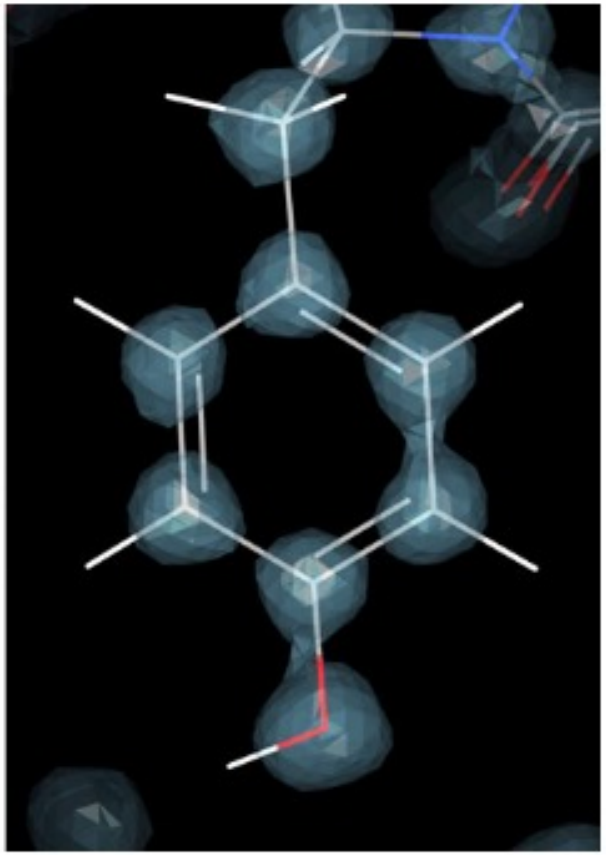
Electron density map

Building a structure model

The importance of resolution



http://www.ruppweb.org/Xray/101index.html



Crystal structure of small protein crambin at 0.48 A resolution
Schmidt, A., et al (2011) Acta Crystallography 67: 424-429

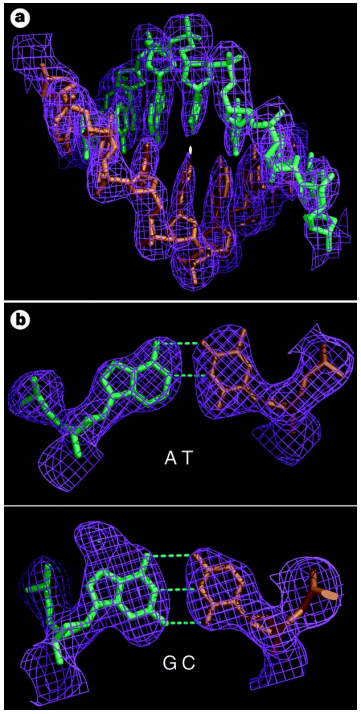
https://www.rcsb.org/structure/3nir

Crystal structure of the nucleosome core particle at 2.8 Å resolution

Karolin Luger, Armin W. Mäder, Robin K. Richmond, David F. Sargent & Timothy J. Richmond

Institut für Molekularbiologie und Biophysik ETHZ, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

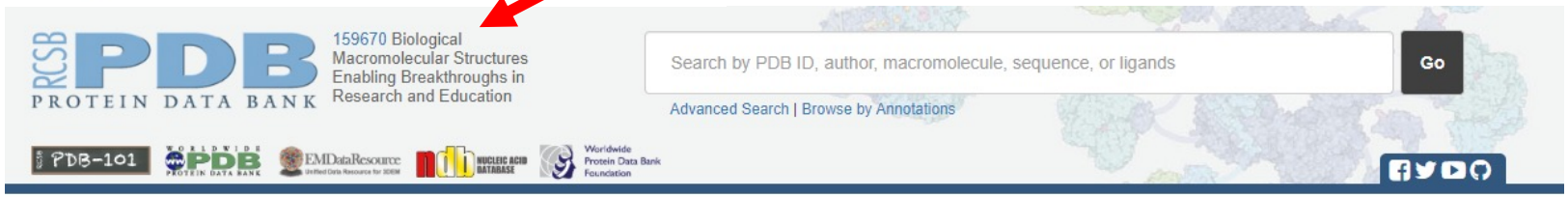
The X-ray crystal structure of the nucleosome core particle of chromatin shows in atomic detail how the histone protein octamer is assembled and how 146 base pairs of DNA are organized into a superhelix around it. Both histone/histone and histone/DNA interactions depend on the histone fold domains and additional, well ordered structure elements extending from this motif. Histone amino-terminal tails pass over and between the gyres of the DNA superhelix to contact neighbouring particles. The lack of uniformity between multiple histone/DNA-binding sites causes the DNA to deviate from ideal superhelix geometry.

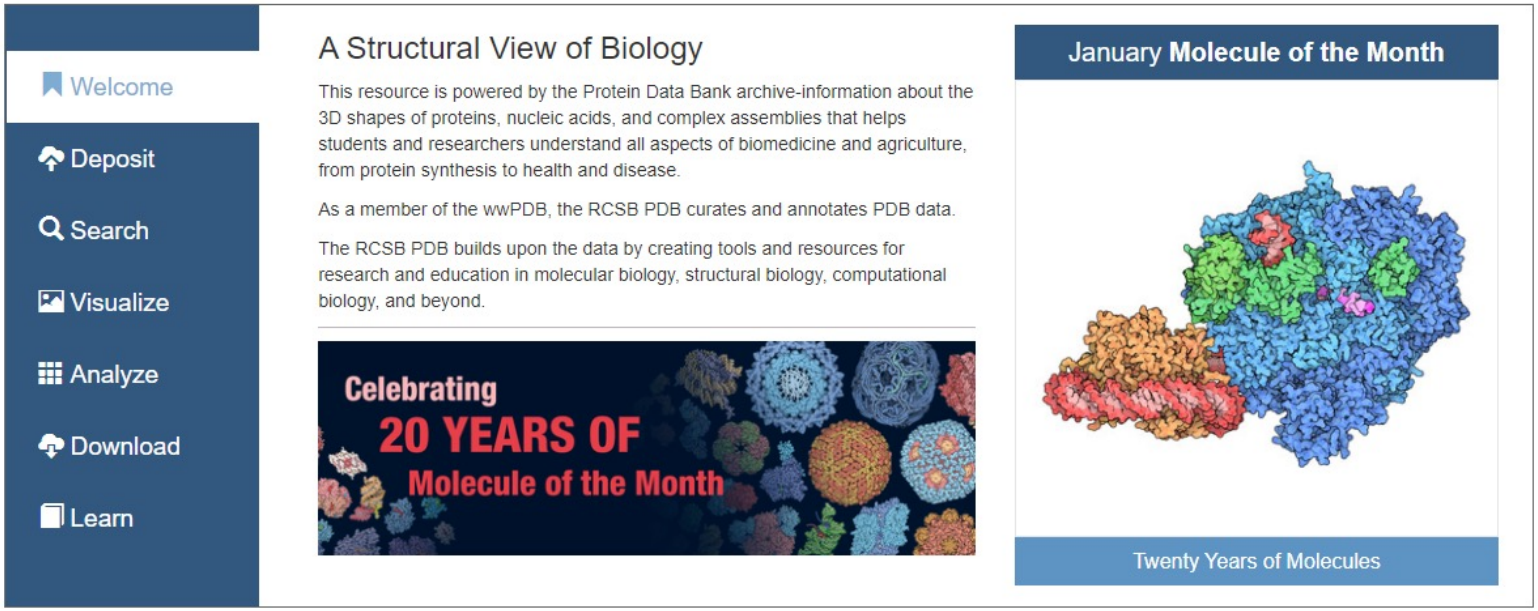


Structure Databases

- **Where does protein structural information reside?**
  - **PDB:**
    - http://www.rcsb.org/pdb/
  - **MMDB:**
    - http://www.ncbi.nlm.nih.gov/Structure/
  - **FSSP:**
    - http://www.ebi.ac.uk/dali/fssp/
  - **SCOP:**
    - http://scop.mrc-lmb.cam.ac.uk/scop/
  - **CATH:**
    - http://www.biochem.ucl.ac.uk/bsm/cath_new/

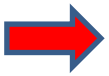
```
# of PDB structures
  2018:   137,178
  2019:   148,268
  2020:   159,670
```





https://pdb101.rcsb.org/learn/videos/what-is-a-protein-video

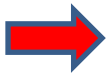
PDB: What species are the structures from?

human 

ORGANISM

Homo sapiens (42668)
Escherichia coli (9294)
Mus musculus (6313)
Saccharomyces cerevisiae (4133)
synthetic construct (3707)
Rattus norvegicus (2988)
Bos taurus (2852)
Other (77188)

Which methods?

X-ray 

EXPERIMENTAL METHOD

X-ray (132583) Resolution range 15 - 0.48 Å
Solution NMR (12391)
Electron Microscopy (2783) Resolution range 70 - 1.8 Å
Hybrid (138)
Electron Crystallography (112)
Solid-State NMR (101)
Neutron Diffraction (66)
Fiber Diffraction (38)
Solution Scattering (32)
Other (24)

http://www.rcsb.org/pdb/home/home.do

PDB X-ray Structures:

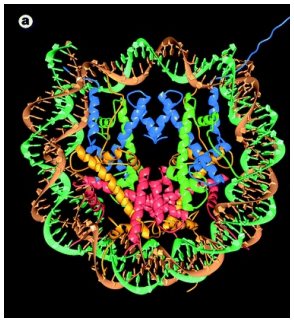
http://www.rcsb.org/pdb/results/results.do?outformat=&qrid=1B04C26E&tabtoshow=Current

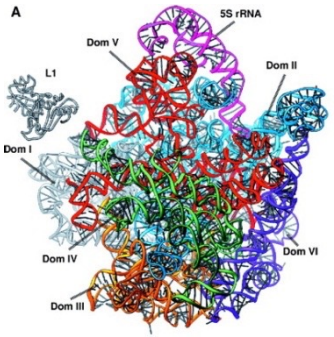
ORGANISM

Homo sapiens (37692)
Escherichia coli (8330)
Mus musculus (5352)
Saccharomyces cerevisiae (3437)
synthetic construct (3305)
Rattus norvegicus (2623)
Bos taurus (2570)
Other (reached drill-down ... (71122)

POLYMER TYPE

Protein (124178)
Mixed (6508)
DNA (1074)
RNA (819)





MEMBRANE PROTEINS → Small % of the total x-ray data

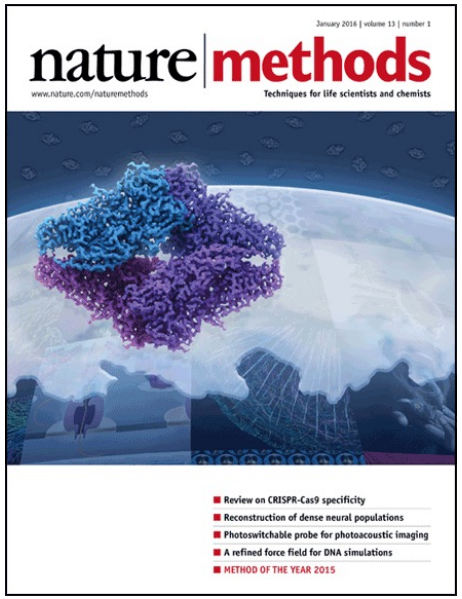
ALPHA-HELICAL (3071)
BETA-BARREL (914)
MONOTOPIC MEMBRANE PROTEINS (486)

Tools for Viewing Structures

- **Jmol**
  - **http://jmol.sourceforge.net**
- **PyMOL**
  - **http://pymol.sourceforge.net**
- **Swiss PDB viewer**
  - **http://www.expasy.ch/spdbv**
- **Mage/KiNG**
  - **http://kinemage.biochem.duke.edu/software/mage.php**
  - **http://kinemage.biochem.duke.edu/software/king.php**
- **Rasmol**
  - **http://www.umass.edu/microbio/rasmol/**

Cryo-EM for biomolecular structures

2015 Method of the Year: Single-particle Cryo-EM



METHOD OF THE YEAR 2015

At *Nature Methods* we are ringing in a new year with our celebration of single-particle cryo-electron microscopy (cryo-EM) as our Method of the Year 2015. Cryo-EM has its roots in work first performed in the 1960s. It has steadily progressed over the past few decades as a medium-resolution structural technique for obtaining information about macromolecular samples that resist analysis by X-ray crystallography. But very recent technical advances, especially the development of direct-detection cameras, have enabled the field to achieve impressive leaps in resolution—even reaching the near-atomic realm of X-ray crystallography—and, by extension, biological applicability. An Editorial, News Feature, Primer, Historical Commentary and Commentary discuss how cryo-EM works, what it is used for, how the field began, why now is such an exhilarating time, and where the field is going in the future. We also cast our predictions about methods with exciting potential in our Methods to Watch section.

Special feature starts on p19

2017 Nobel Prize in Chemistry

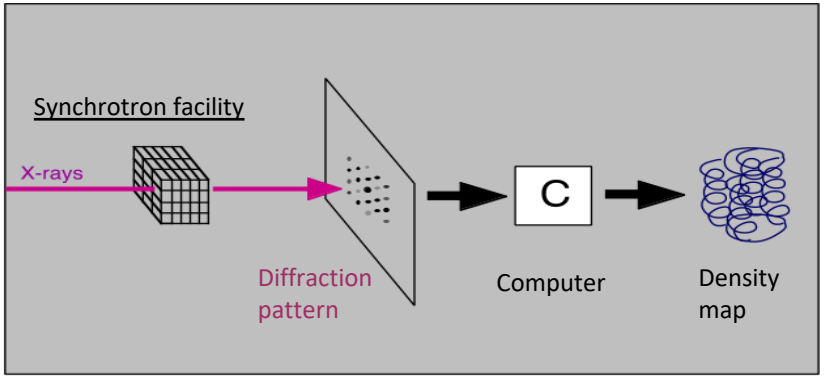
"for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution"

Jacques Dubochet (University of Lausanne, Switzerland)

Joachim Frank (Columbia University, New York, USA)

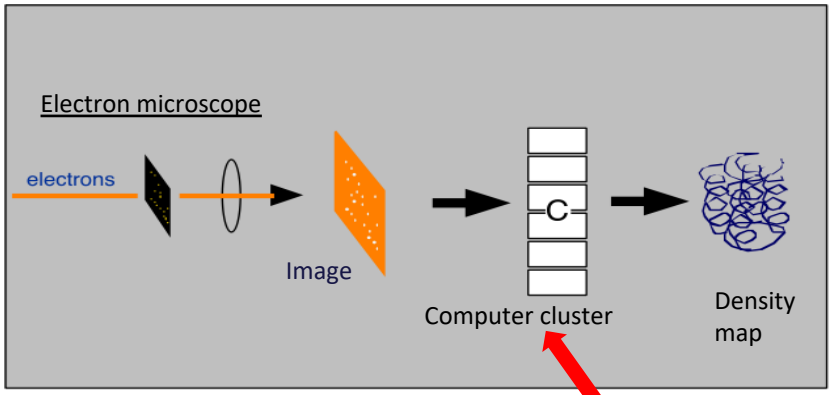
Richard Henderson (MRC Laboratory of Molecular Biology, Cambridge, UK)

Two methods for structure determination



X-ray crystallography

Well-established (since 1960s)
Requires well-ordered crystals
>10^{12} copies of protein



Single-particle cryo-EM

Recent (1990s-present)
No crystals required!
~10^{5} copies of protein

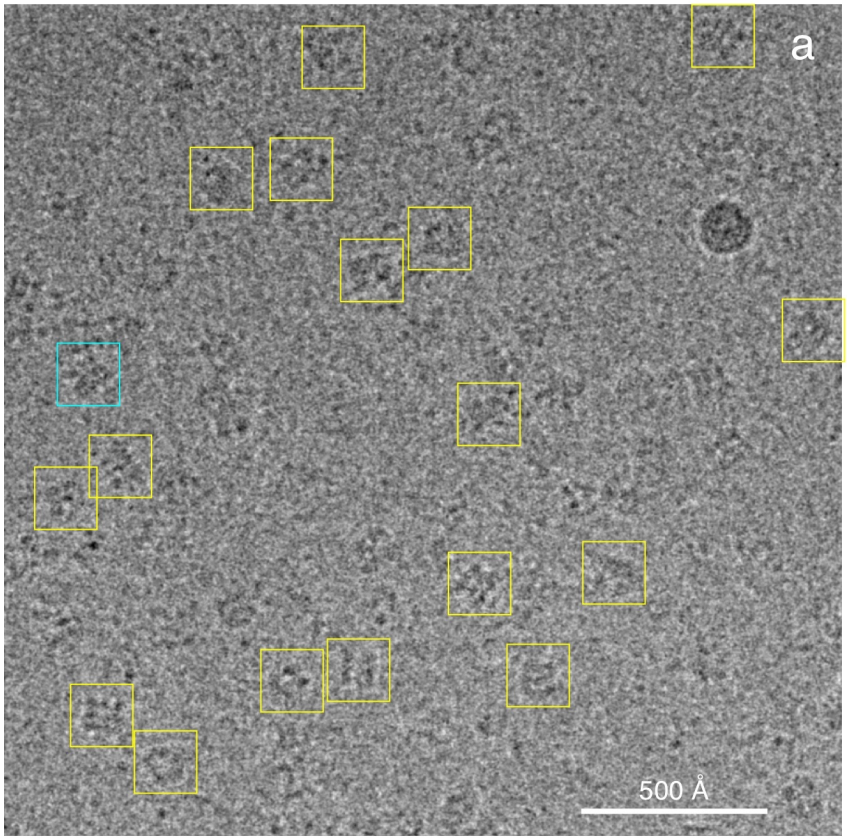
Some gaming PCs can now replace the cluster ☺

Fred Sigworth

The Cryo-EM specimen gives only a phase contrast image

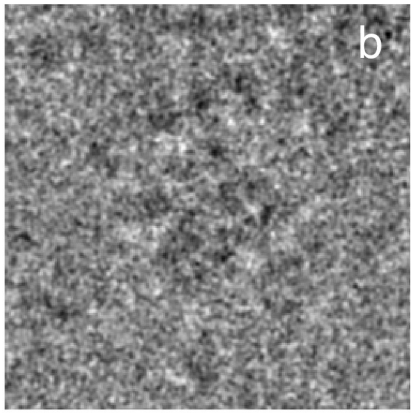
- A constellation of images and data processing are essential.

1/4 of a micrograph, showing some particles

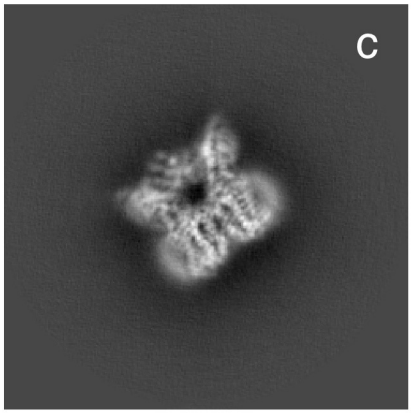


Y. Cheng and D. Julius lab. Nature 2013

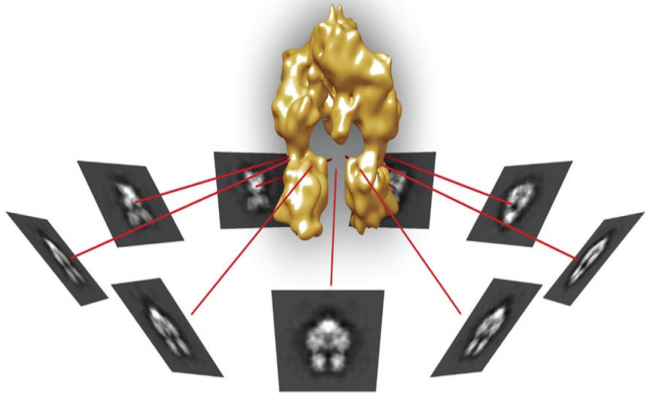
Image



Projection



- orientation assignment and averaging
- 3D reconstruction



Fred Sigworth

Cryo-EM: membrane proteins, protein complexes, proteins difficult to crystalize

Science

RESEARCH ARTICLES

Cite as: R. Zhou *et al.*, *Science* 10.1126/science.aaw0930 (2019).

Recognition of the amyloid precursor protein by human γ-secretase

Rui Zhou[1]*, Guanghui Yang[1]*, Xuefei Guo[1], Qiang Zhou[2,3], Jianlin Lei[1,4], Yigong Shi[1,2]†

[1]Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]Institute of Biology, Westlake Institute for Advanced Study, Westlake University, 18 Shilongshan Road, Xihu District, Hangzhou 310024, Zhejiang Province, China. [3]School of Life Sciences, Westlake University, 18 Shilongshan Road, Xihu District, Hangzhou 310024, Zhejiang Province, China. [4]Technology Center for Protein Sciences, Ministry of Education Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China.

*These authors contributed equally to this work.

†Corresponding author. Email: shi-lab@tsinghua.edu.cn

Cleavage of amyloid precursor protein (APP) by the intramembrane protease γ-secretase is linked to Alzheimer's disease. We report an atomic structure of human γ-secretase in complex with a transmembrane APP fragment at 2.6-Å resolution. The transmembrane helix (TM) of APP closely interacts with five surrounding TMs of PS1 (the catalytic subunit of γ-secretase). A hybrid β-sheet, which is formed by a β-strand from APP and two β-strands from PS1, guides γ-secretase to the scissile peptide bond of APP between its TM and β-strand. Residues at the interface between PS1 and APP are heavily targeted by recurring mutations from AD patients. This structure, together with that of γ-secretase bound to Notch, reveal contrasting features of substrate binding, which may be exploited toward design of substrate-specific inhibitors.

6IYC

Recognition of the Amyloid Precursor Protein by Human gamma-secretase

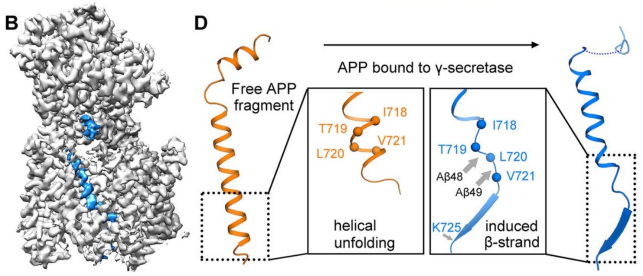
DOI: 10.2210/pdb6IYC/pdb EMDataBank: EMD-9751

Classification: MEMBRANE PROTEIN
Organism(s): Homo sapiens
Expression System: Homo sapiens
Mutation(s): 2

Deposited: 2018-12-14 Released: 2019-01-23
Deposition Author(s): Zhou, R., Yang, G., Guo, X., Zhou, Q., Lei, J., Shi, Y.
Funding Organization(s): National Natural Science Foundation of China

Experimental Data Snapshot
Method: ELECTRON MICROSCOPY
Resolution: 2.6 Å
Aggregation State: PARTICLE
Reconstruction Method: SINGLE PARTICLE

wwPDB Validation



Science

RESESARCH ARTICLES

Cite as: E. C. Twomey *et al.*, *Science* 10.1126/science.aax1033 (2019).

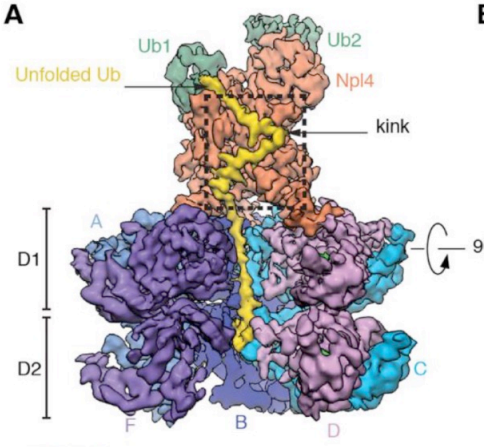
Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding

Edward C. Twomey[1]*, Zhejian Ji[1]*, Thomas E. Wales[2], Nicholas O. Bodnar[1], Scott B. Ficarro[3,4], Jarrod A. Marto[3,4], John R. Engen[2], Tom A. Rapoport[1]†

[1]Department of Cell Biology, Harvard Medical School, and Howard Hughes Medical Institute, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. [2]Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA. [3]Department of Cancer Biology, Department of Oncologic Pathology, and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115, USA. [4]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.

***These authors contributed equally to this work.**

†Corresponding author. Email: tom_rapoport@hms.harvard.edu

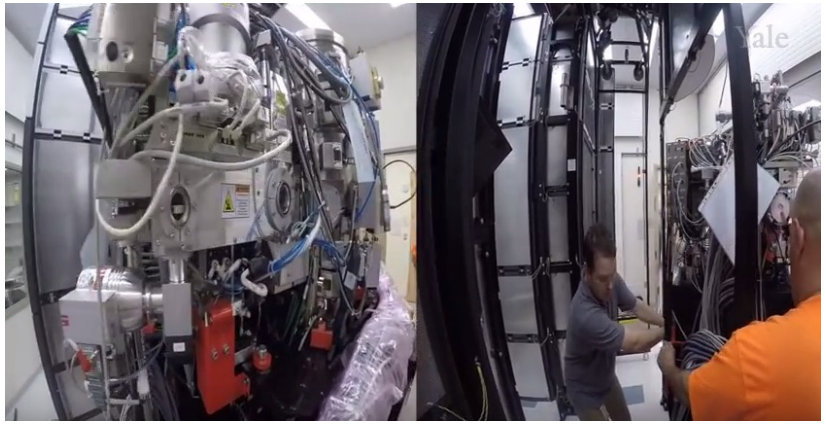


PMID: 30630874;30598546;25918421;31249135

New Technologies, Automation, & Computation are accelerating the field

Krios at National University of Singapore

Control room at Scripps Research Institute, La Jolla



Krios TEM installation on Yale's West Campus.

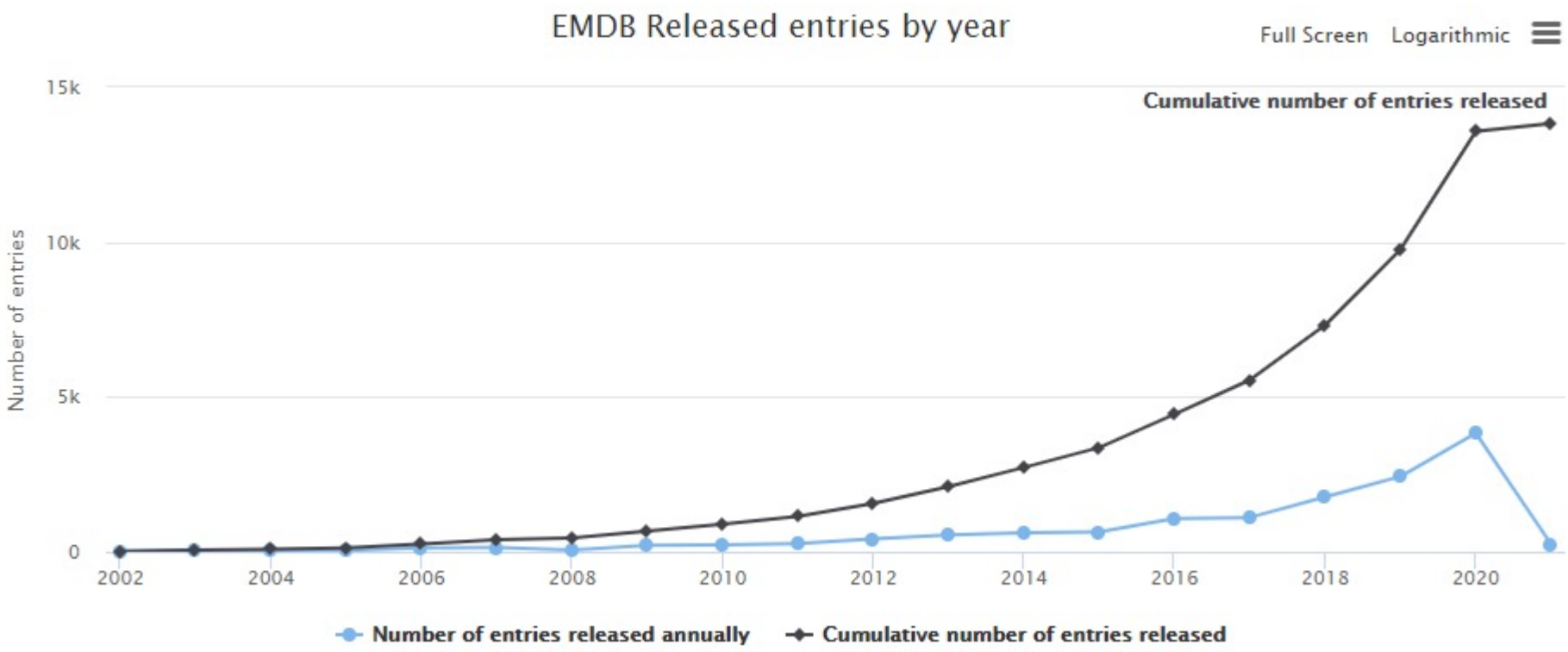
Fred Sigworth

EMBL-EBI | Services | Research | Training | About us | EMBL-EBI

EMDB
Electron Microscopy Data Bank

Search EMDB...
Examples: 1001, Apoferritin, Tomography, Rossmann MG, 5A1A
advanced search

Home | Deposition | Documentation | Resources | FTP Archive | REST API | About | Feedback | Share



https://wwwdev.ebi.ac.uk/emdb/statistics

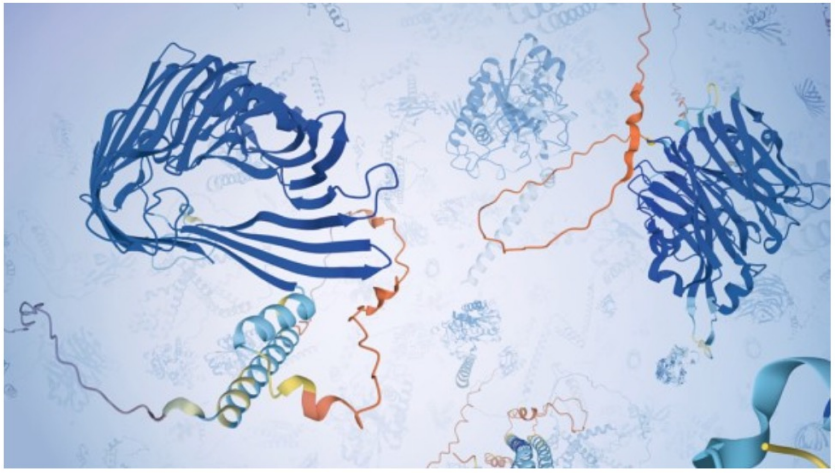
02/15/2021
PDB, X-RAY = 154,039 entries
EMDB, EM = 13,827 entries

AI deep-learning-based methods solved the protein folding problem

FOCUS | 11 JANUARY 2022

Method of the Year 2021: Protein structure prediction

Protein structure prediction is our Method of the Year 2021, for the remarkable levels of accuracy achieved by deep learning-based methods in predicting the 3D structures of proteins and protein complexes, essentially solving this long-standing challenge.



Excellent perspective & overview:

"The impact of AlphaFold2 one year on." Jones, D.T., Thornton, J.M. *Nature Methods* **19,** 15–20 (2022). PMID: 35017725

Key literature:

(AlphaFold)

Senior, A. W. et al. *Nature* **577**, 706–710 (2020). PMID: 34293799.

Jumper, J. et al. *Nature* **596**, 583–589 (2021). PMID: 34265844.

Tunyasuvunakool, K. et al. *Nature* **596**, 590–596 (2021) PMID: 34293799.

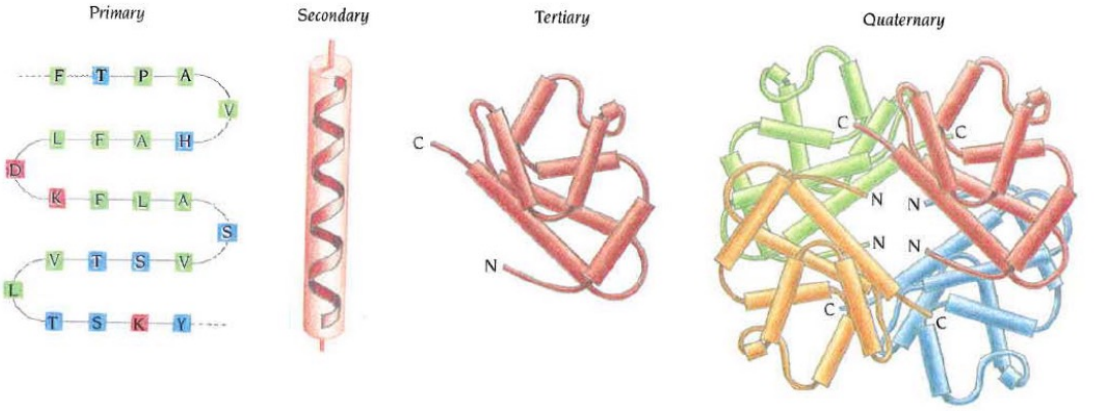
(RoseTTA)

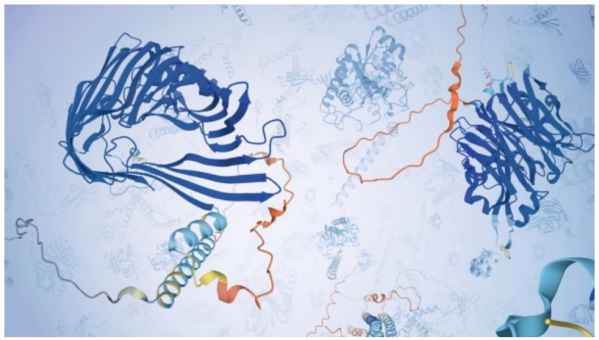
Baek, M. et al. *Science* **373,** (2021) PMID: 34282049

The protein-folding problem was first posed over 50 years ago:

What is the physical code by which an amino acid sequence dictates fold?

Can we devise a computer algorithm to predict protein structures from their sequences?



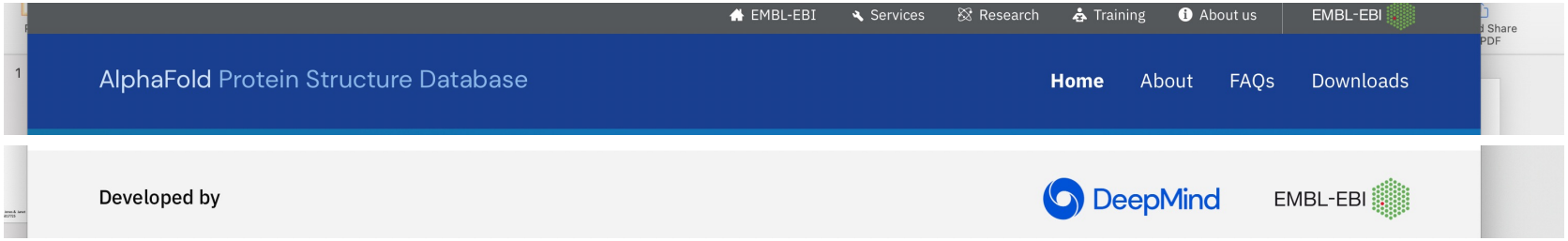


The Protein-Folding Problem, 50 Years On, Dill K and Maccallum, J.L. Science, 2012, PMID: 23180855
Proteins and Protein Structure (Branden, C. and Tooze, J. *Introduction to Protein Structure*)

AI: deep-learning-based methods AlphaFold2 and RoseTTA solved the protein folding problem

- In the past year, the deep-learning-based (AI) methods AlphaFold2 and RoseTTA fold have managed to solve the protein folding problem

https://alphafold.ebi.ac.uk



- European Molecular Biology Laboratory and DeepMind predicted structures for over **350,000 proteins for 21 model organisms** and made them freely available at the AlphaFold Protein Structure Database

- plans for expanding predictions to millions of structures in 2022.
  - Structure of UniRef904 ~ over 130 million protein models by the middle of 2022 and **ultimately providing more than 700 times as many models as there are experimentally derived structures currently available in the PDB.**

Jones, D.T., Thornton, J.M. *Nat Methods* **19,** 15–20 (2022). PMID: 35017725