

# Structural Data: Introduction to X-ray Crystallography & Cryo-EM

**Jesse Rinehart, PhD**

**Biomedical Data Science: Mining & Modeling**  
**CBB 752, Spring 2020**



**Cellular & Molecular Physiology**  
**Yale University School of Medicine**



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

## **Yale Structure Courses:**

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

**MB&B 720a, Macromolecular Structure and Biophysical Analysis**

**MB&B 721b, Macromolecular Interactions and Dynamic Properties**

**MB&B 760b3: Principles of Macromolecular Crystallography**

**MB&B 761b4: X-ray Crystallography Workshop**

**Pharmacology 529b: Structural Pharmacology**

## **Additional Resources:**

“Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models”  
by 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

**“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**

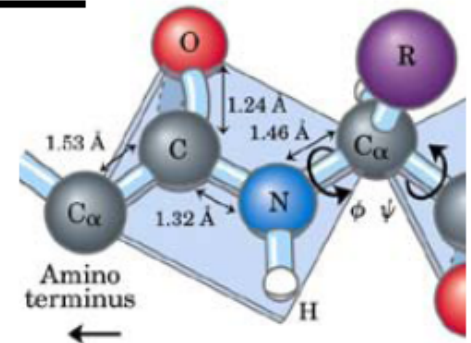
- There's a limit to how small an object can be seen under a light microscope.
- The diffraction limit: you can not image things that are much smaller than the wavelength of the light you are using.
- The wavelength for visible light is measured in hundreds of nanometers, while atoms are separated by 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  $\text{CuK}\alpha$  radiation. Wavelength = 1.54 Å.

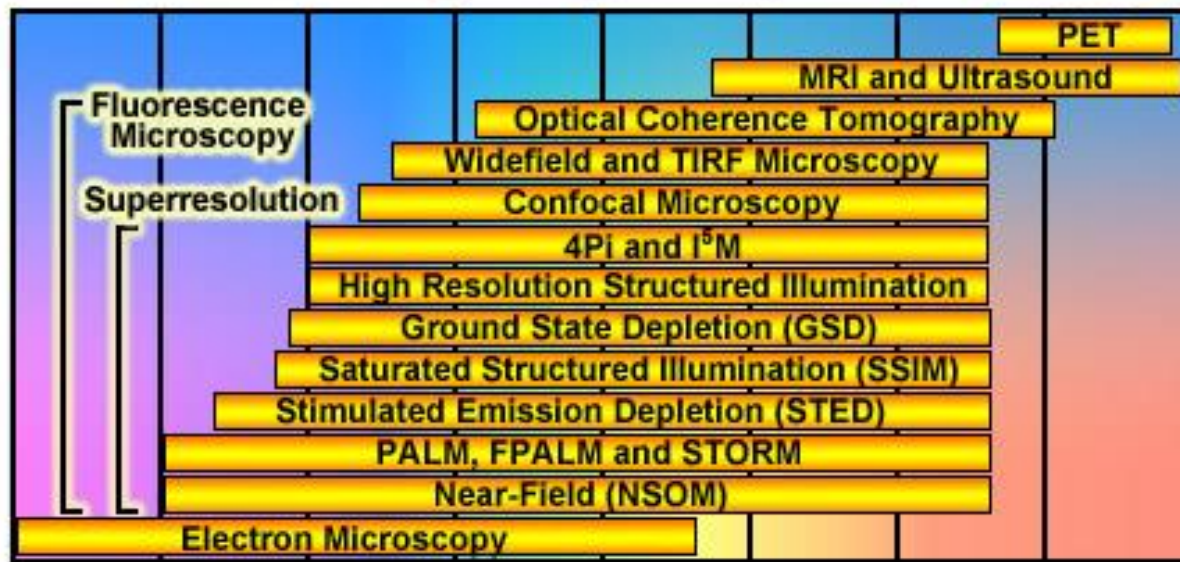
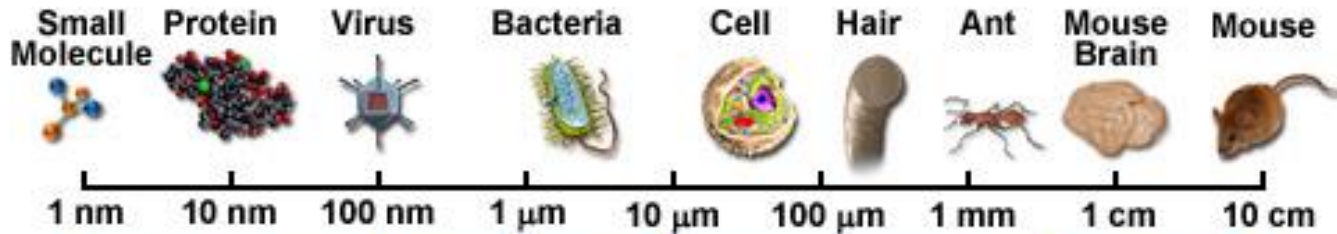
Synchrotron radiation wavelengths in the range 0.5 Å - 2.5 Å.



Yong Xiong

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

## Spatial Resolution of Biological Imaging Techniques



$1 \text{ \AA} = 0.1 \text{ nm}$

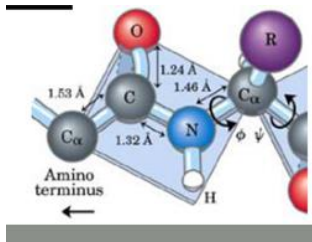
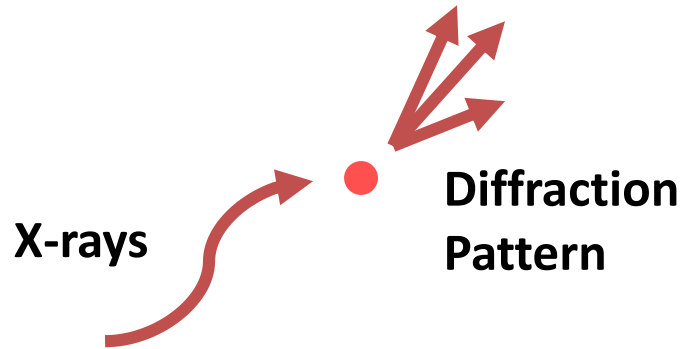


Figure 1

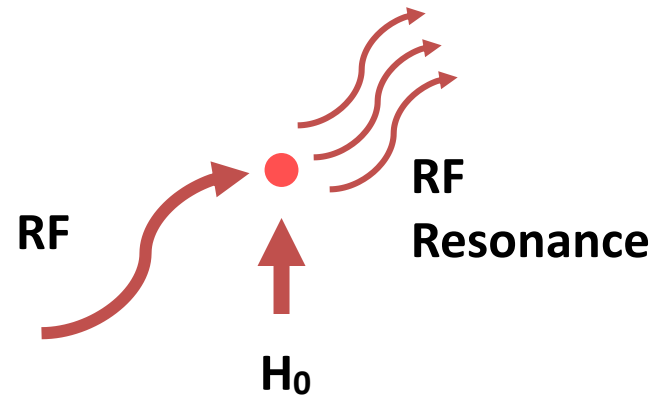
# Experimental Determination of Atomic Resolution Structures

## X-ray



- Direct detection of atom positions
- Crystals

## NMR

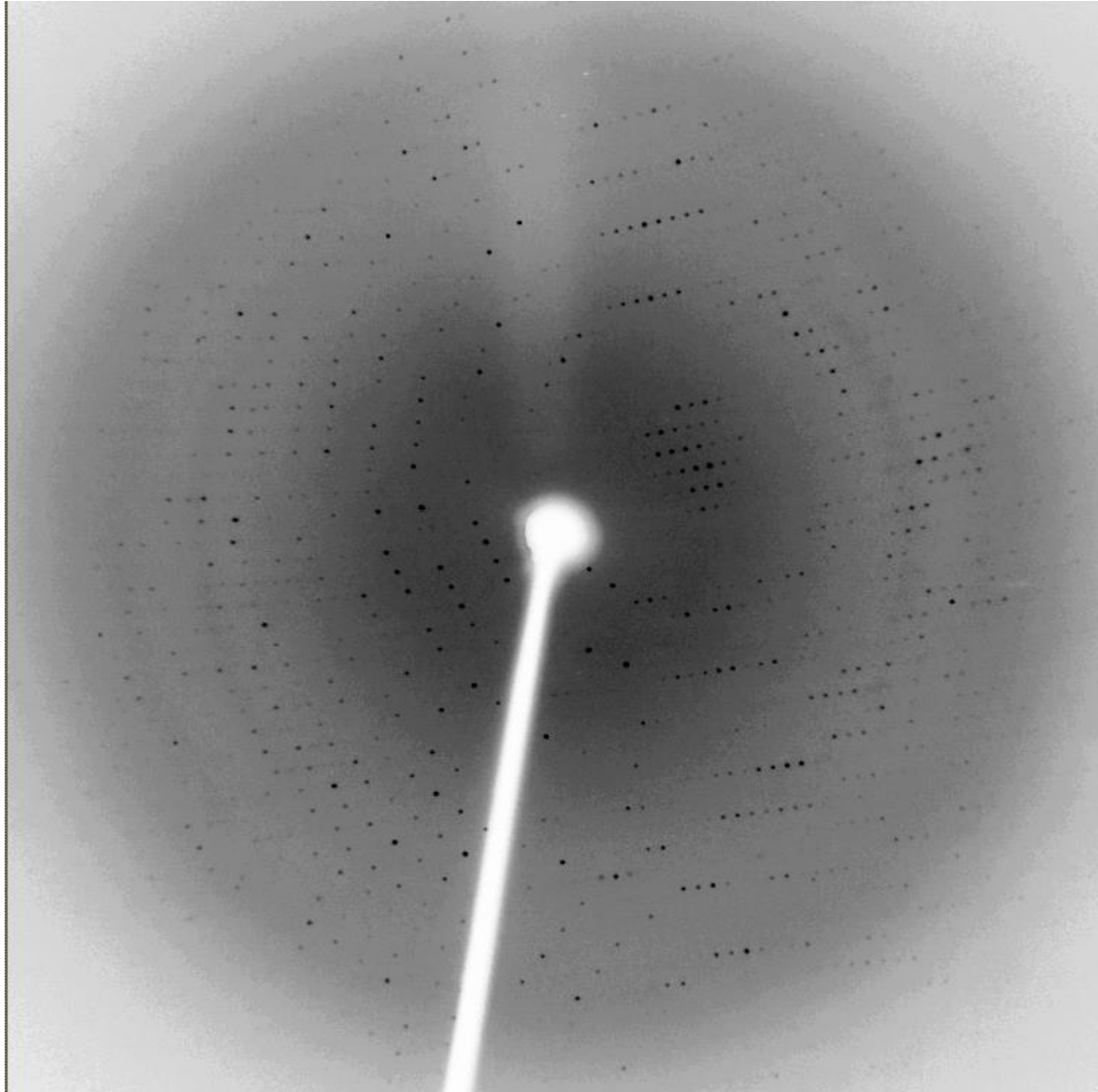


- Indirect detection of H-H distances
- In solution

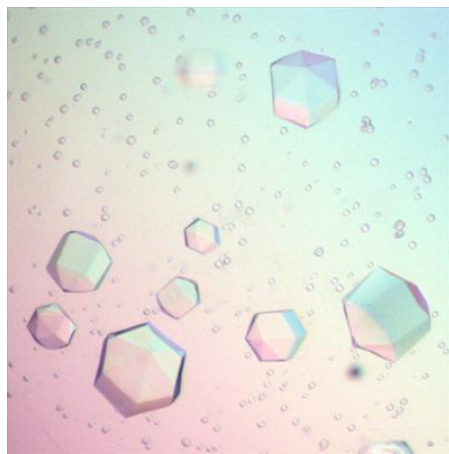
Other methods for determining protein structures:

-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  $\sim N^2$  fold).

# Determination of Protein Crystal Structure

Subcloning



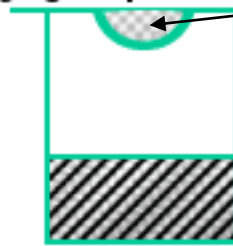
Expression



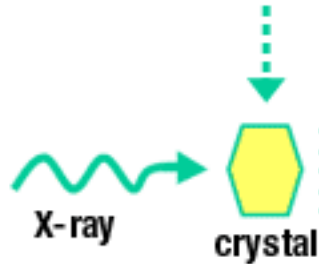
Purification



Crystallization: Hanging Drop with Protein



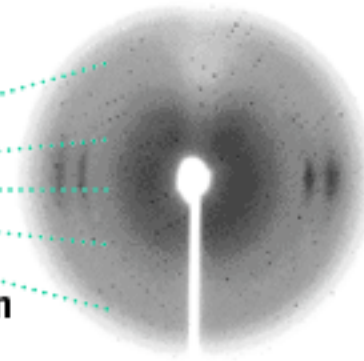
Reservoir with Precipitant



X-ray

crystal

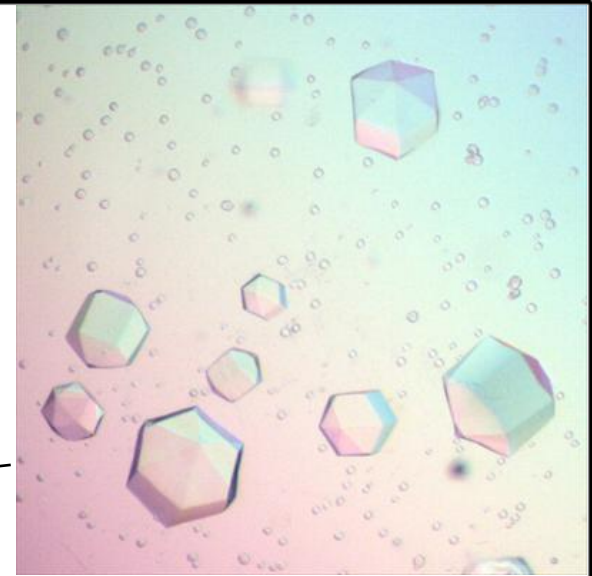
diffraction



detector

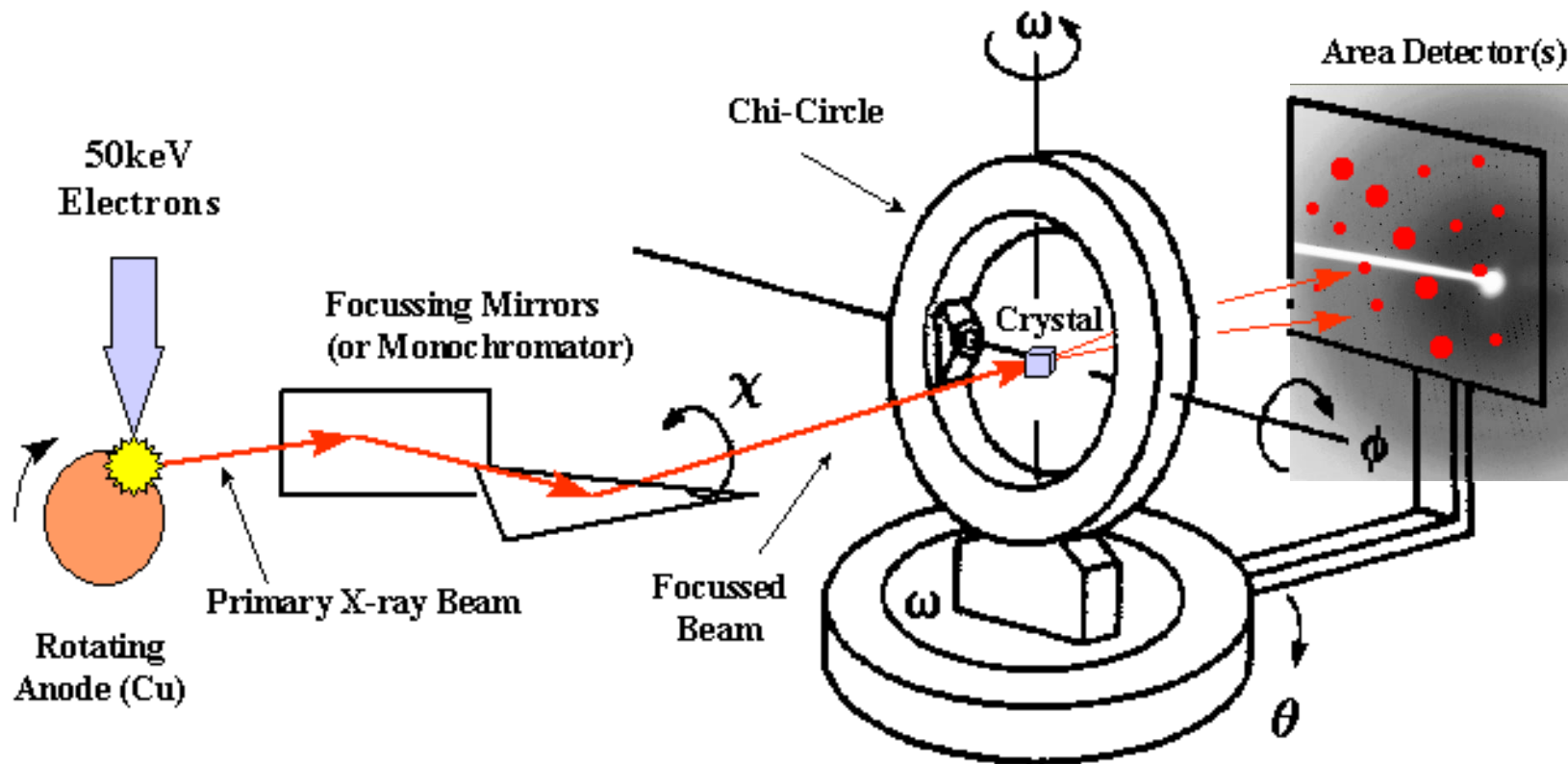


refinement  
modeling





# Data Collection



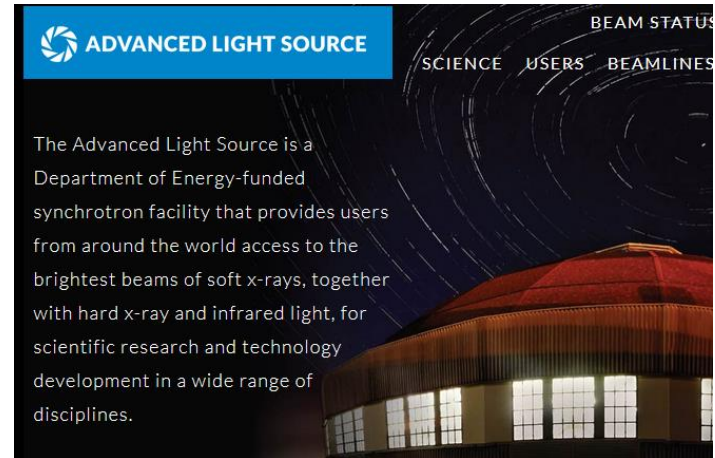
4-Circle Goniometer ( Eulerian or Kappa Geometry)

# Synchrotron X-ray Sources are the method of choice

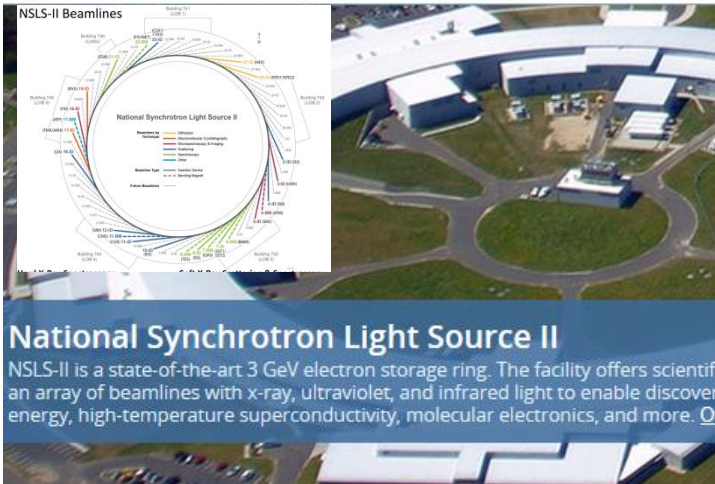
Lab x-ray sources @ 1.54 Å compared to Synchrotron X-ray @ 0.5 Å - 2.5 Å.



APS Chicago



ALS Berkeley

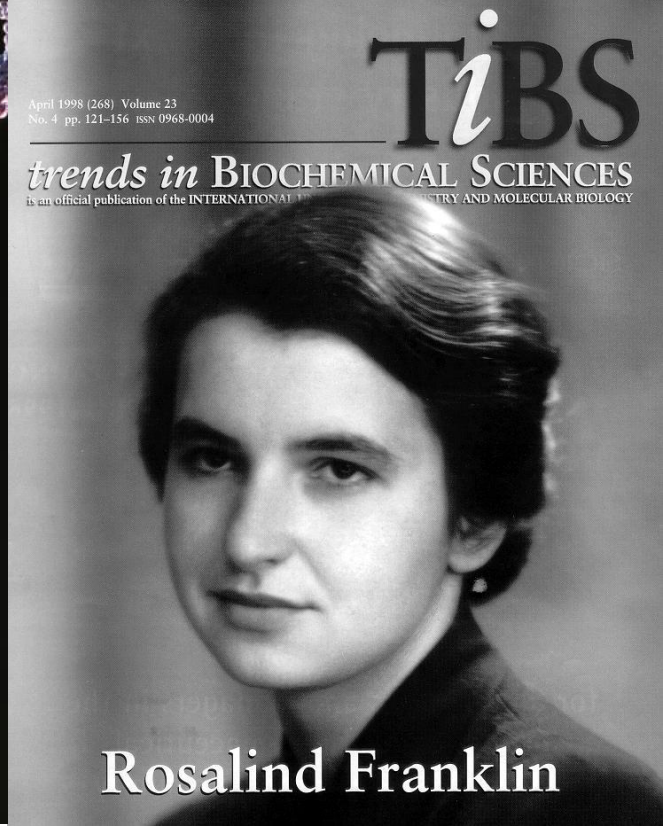
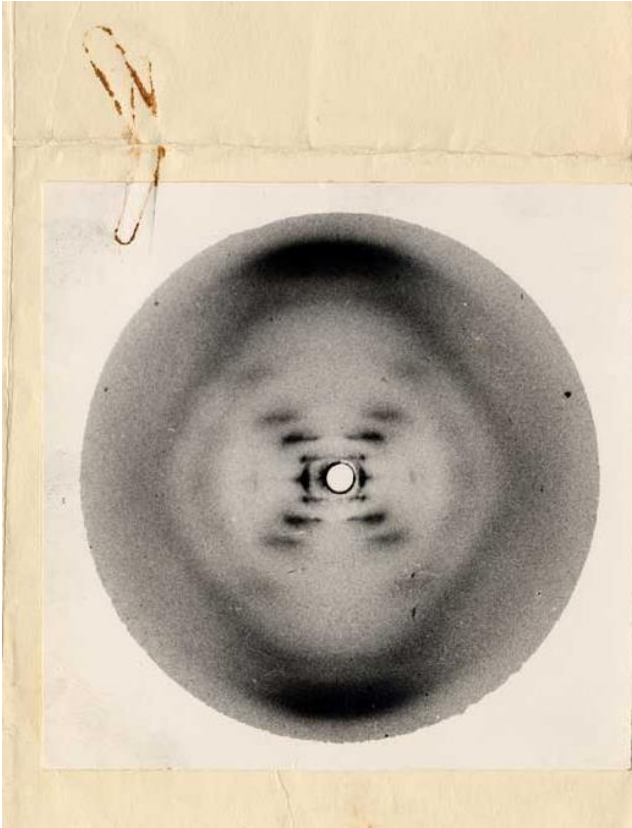


NSLS-II Brookhaven

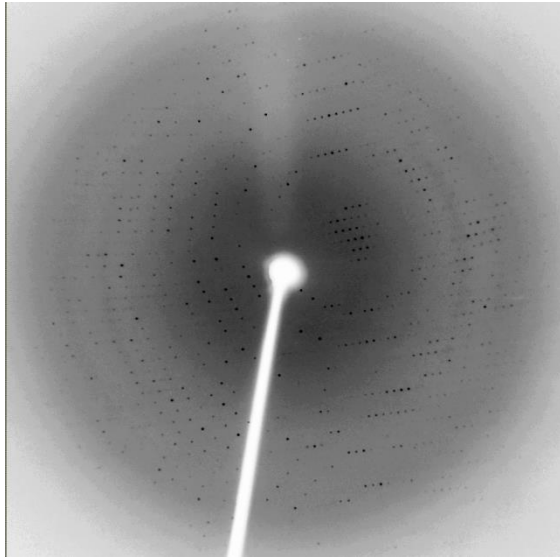


CHESS Ithaca

# Most famous X-ray diffraction pattern



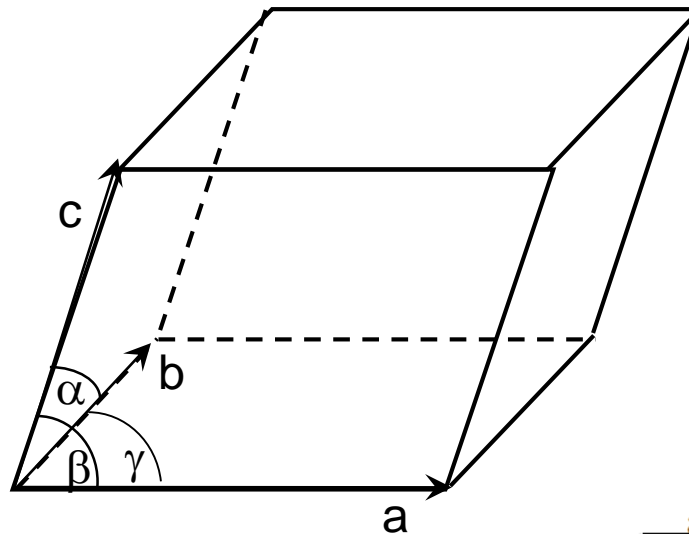
# The information we get from a single diffraction experiment



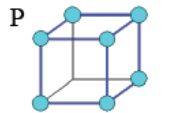
Analyze the pattern  
of the reflections



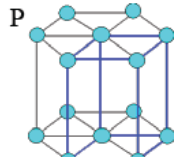
- (a) space group of the crystal
- (b) unit cell dimensions



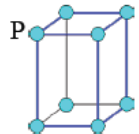
**Cubic**  
 $a = b = c$ ,  
 $\alpha = \beta = \gamma = 90^\circ$



**Hexagonal**  
 $a = b \neq c$ ,  
 $\alpha = \beta = 90^\circ, \gamma = 120^\circ$



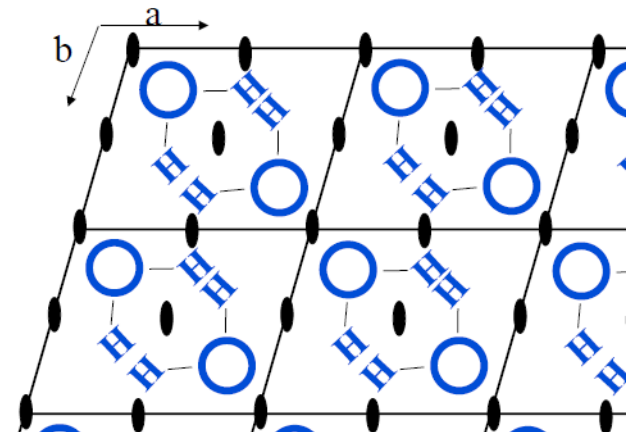
**Trigonal**  
 $a = b \neq c$ ,  
 $\alpha = \beta = 90^\circ, \gamma = 120^\circ$

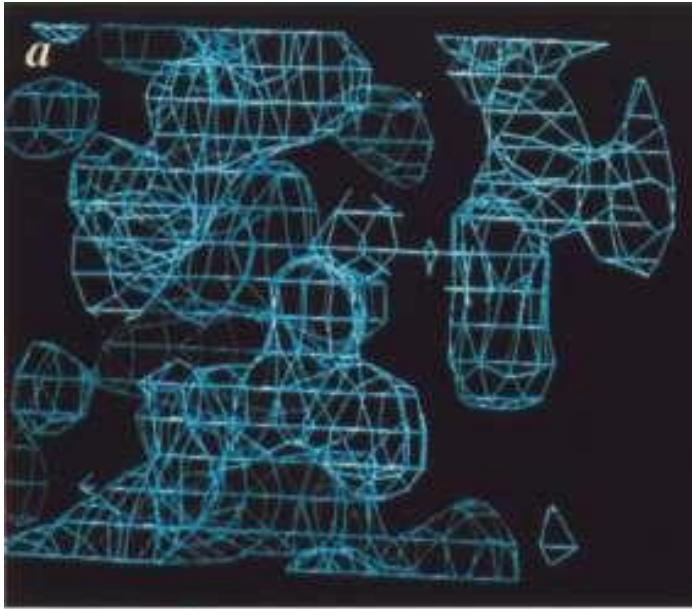


**Tetragonal**  
 $a = b \neq c$ ,  
 $\alpha = \beta = \gamma = 90^\circ$

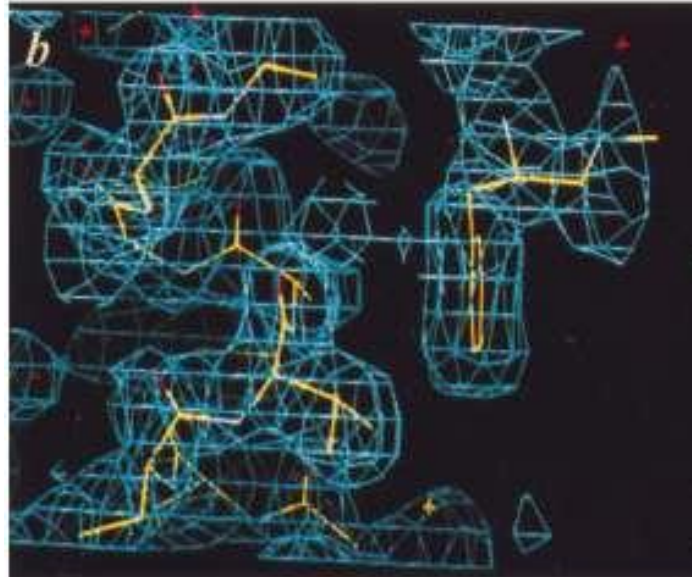
How to understand symmetry?

Crystal = lattice + unit cell content  
 (asymmetric units (asu) content)





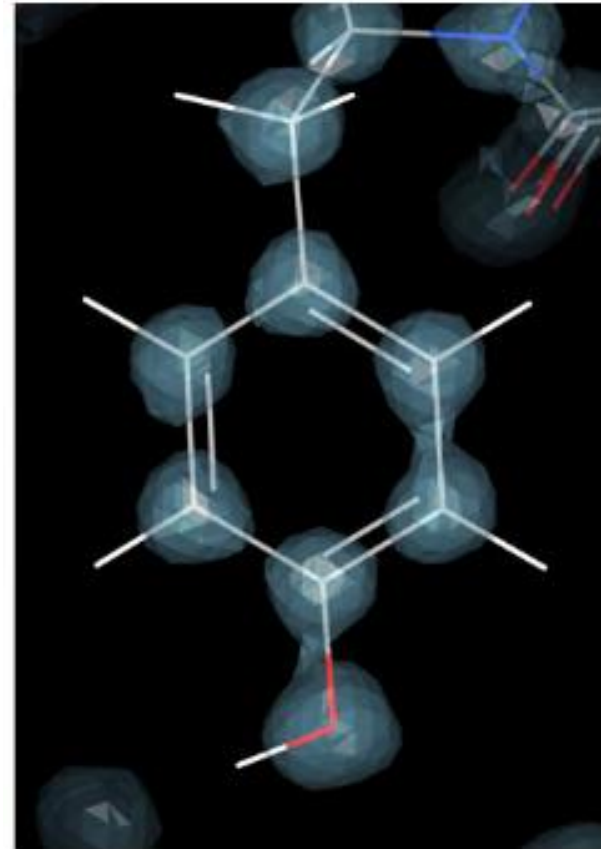
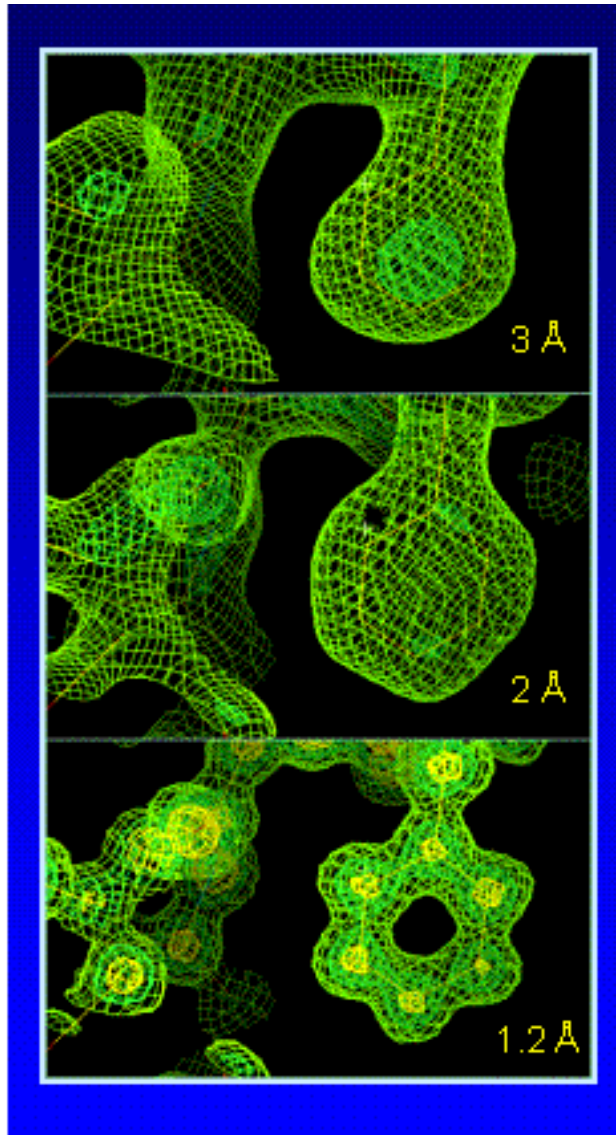
## Electron density map



## Building a structure model

- © 2006
- Academic Press

# The importance of resolution



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

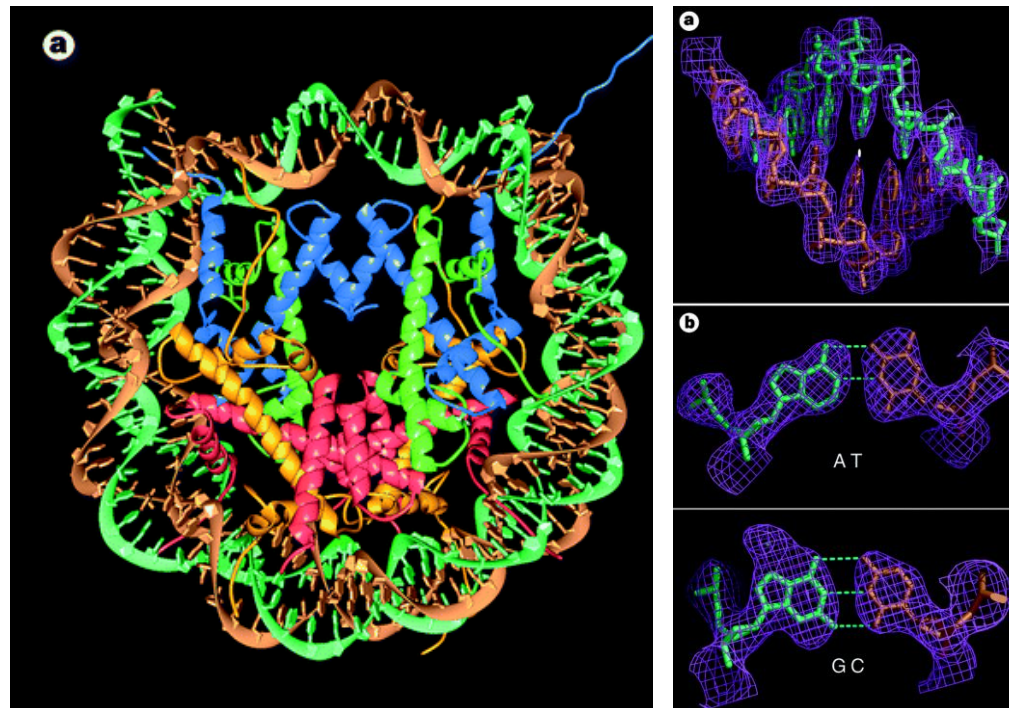
<https://www.rcsb.org/structure/3nir>

# 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

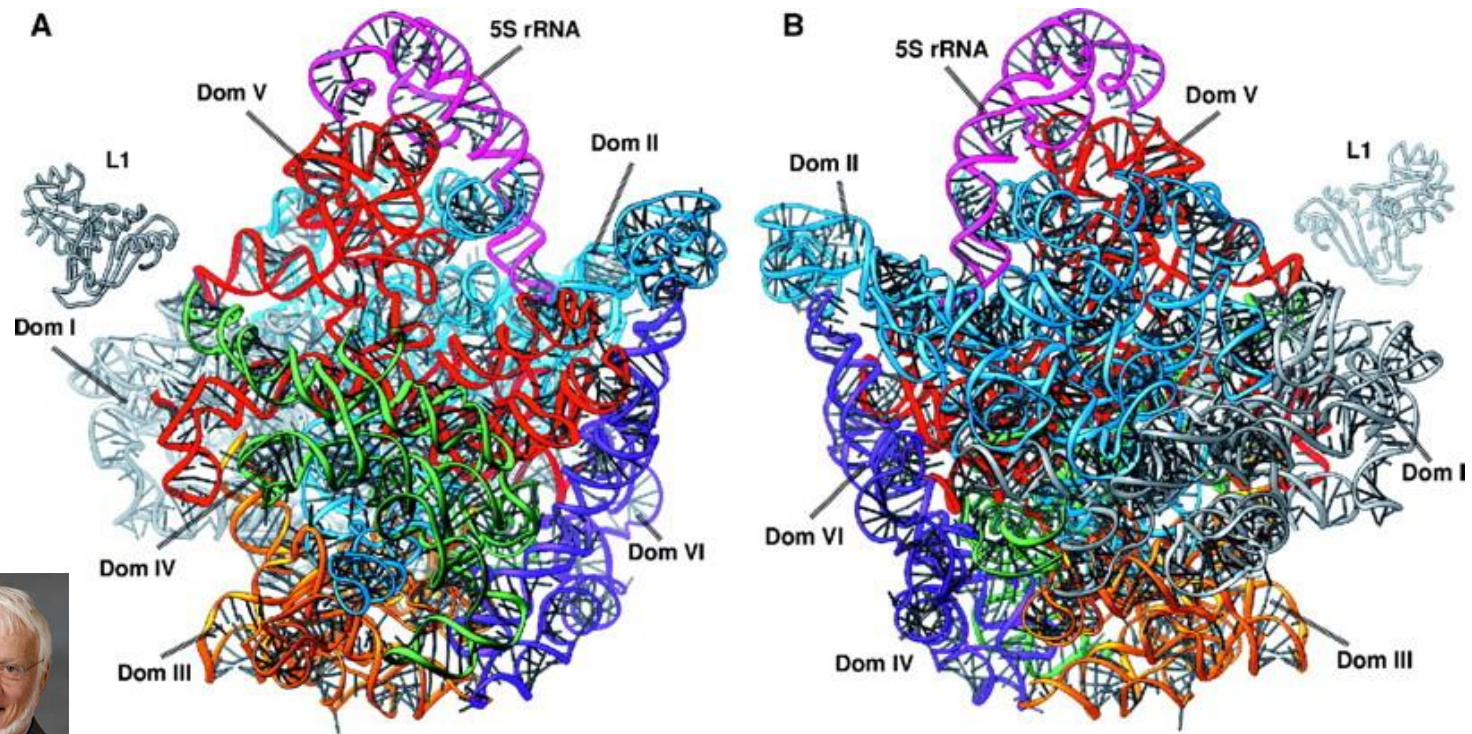
*Institut für Molekularbiologie und Biophysik ETHZ, ETH-Hönggerberg, CH-8093 Zürich, Switzerland*

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.



# The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution

Nenad Ban,<sup>1\*</sup> Poul Nissen,<sup>1\*</sup> Jeffrey Hansen,<sup>1</sup> Peter B. Moore,<sup>1,2</sup>  
Thomas A. Steitz<sup>1,2,3†</sup>



Thomas Steitz shared 2009 Nobel Prize in Chemistry for this structure



# 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/](http://www.biochem.ucl.ac.uk/bsm/cath_new/)

## # of PDB structures

2018: 137,178

2019: 148,268

2020: 159,670

RCSB PDB PROTEIN DATA BANK 159670 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands Go

Advanced Search | Browse by Annotations

PDB-101 Worldwide Protein Data Bank EMDataResource Nucleic Acid Database Worldwide Protein Data Bank Foundation

### Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

## A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### Celebrating 20 YEARS OF Molecule of the Month

### January Molecule of the Month

Twenty Years of Molecules

<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)

# PDB X-ray Structures:

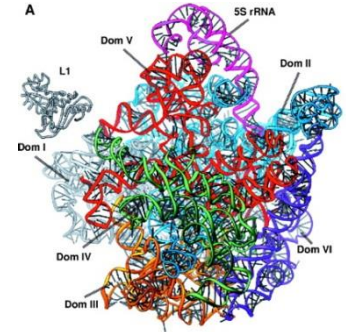
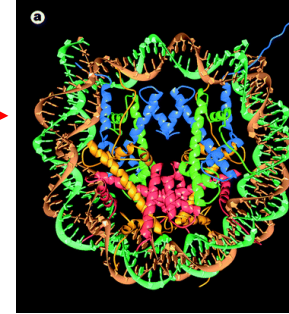
<http://www.rcsb.org/pdb/results/results.do?outformat=&qrid=1B04C26E&tabtoshow=Current>

## 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

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

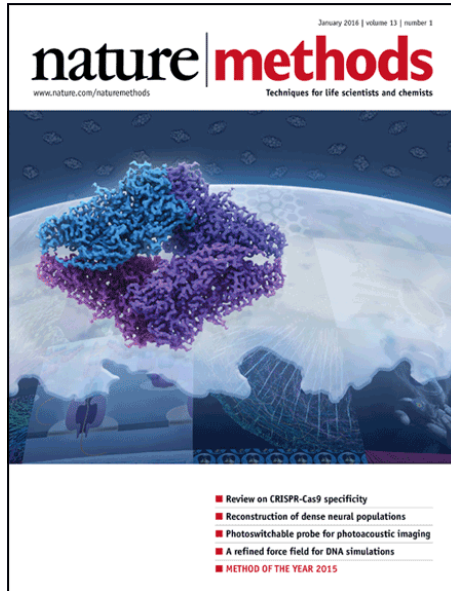
Small % of the total x-ray data

## 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"*

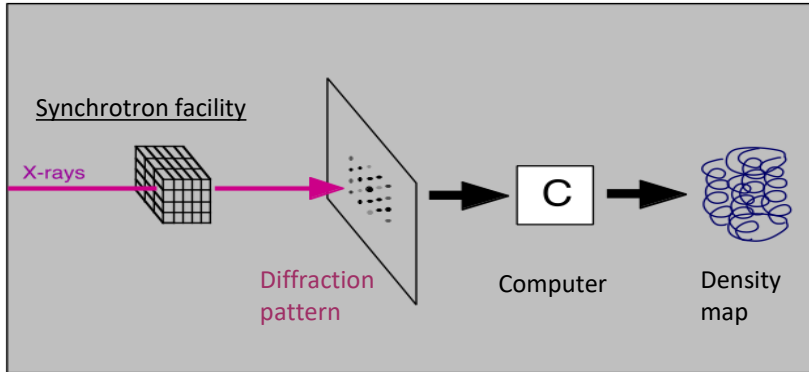


**Jacques Dubochet** (University of Lausanne, Switzerland)

**Joachim Frank** (Columbia University, New York, USA)

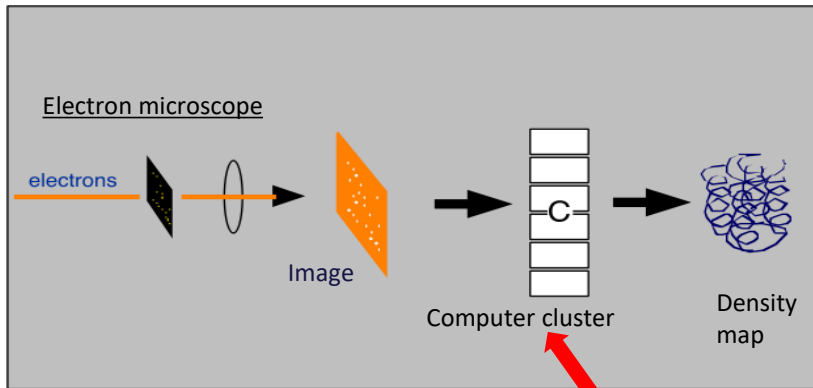
**Richard Henderson** (MRC Laboratory of Molecular Biology, Cambridge, UK)

# Two methods for structure determination



## X-ray crystallography

Well-established (since 1960s)  
Requires well-ordered crystals  
>10<sup>12</sup> copies of protein



## Single-particle cryo-EM

Recent (1990s-present)  
No crystals required!  
~10<sup>5</sup> copies of protein

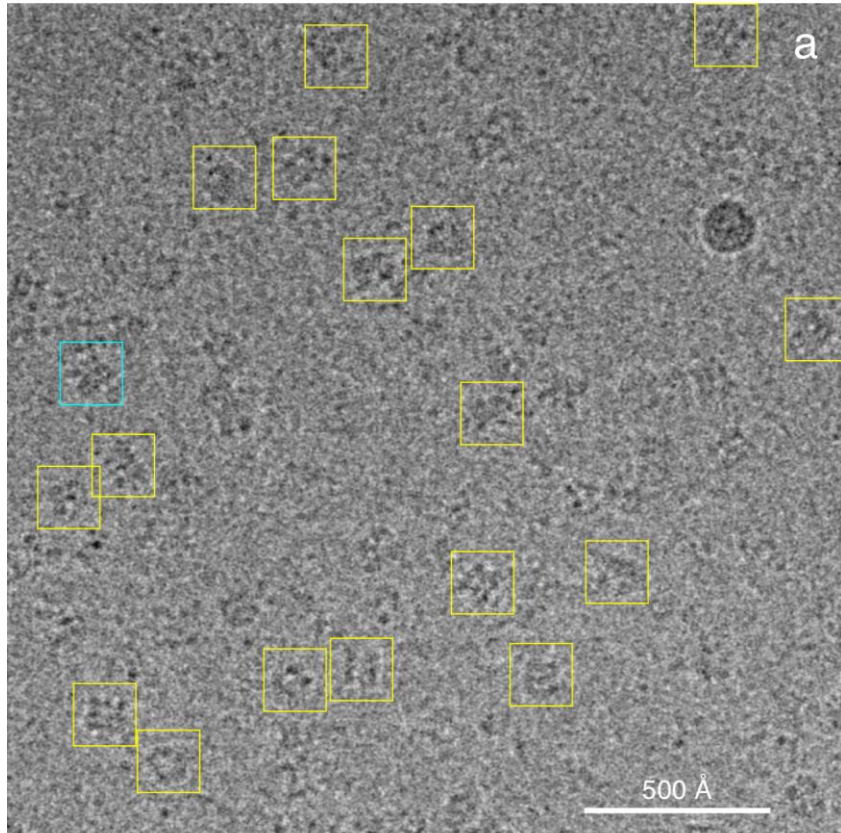
Some gaming PCs  
can now replace  
the cluster 😊



# The Cryo-EM specimen gives only a phase contrast image

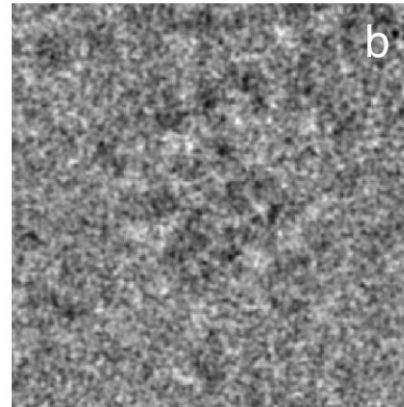
- A constellation of images and data processing are essential.

1/4 of a micrograph, showing some particles

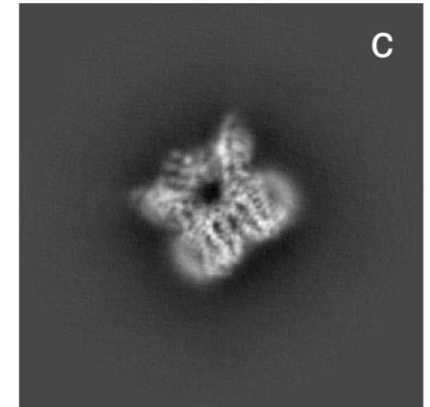


Y. Cheng and D. Julius lab. Nature 2013

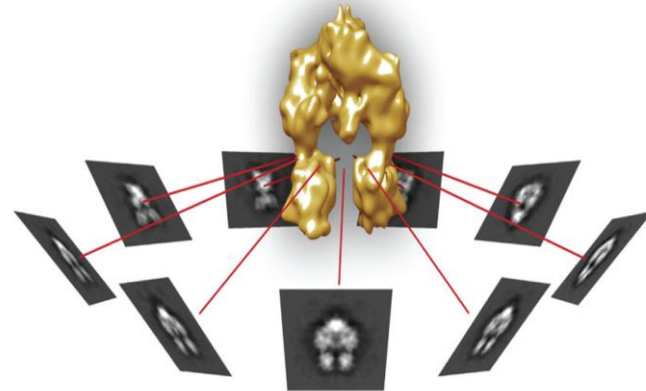
Image



Projection



- orientation assignment and averaging
- 3D reconstruction



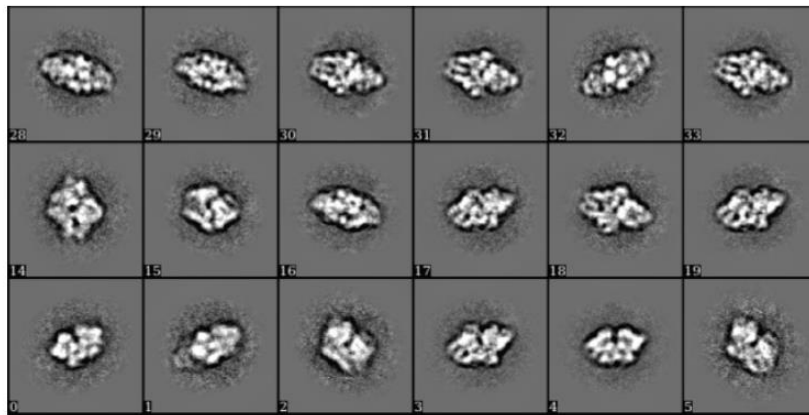
Fred Sigworth

# A landmark study for high-resolution single-particle structures

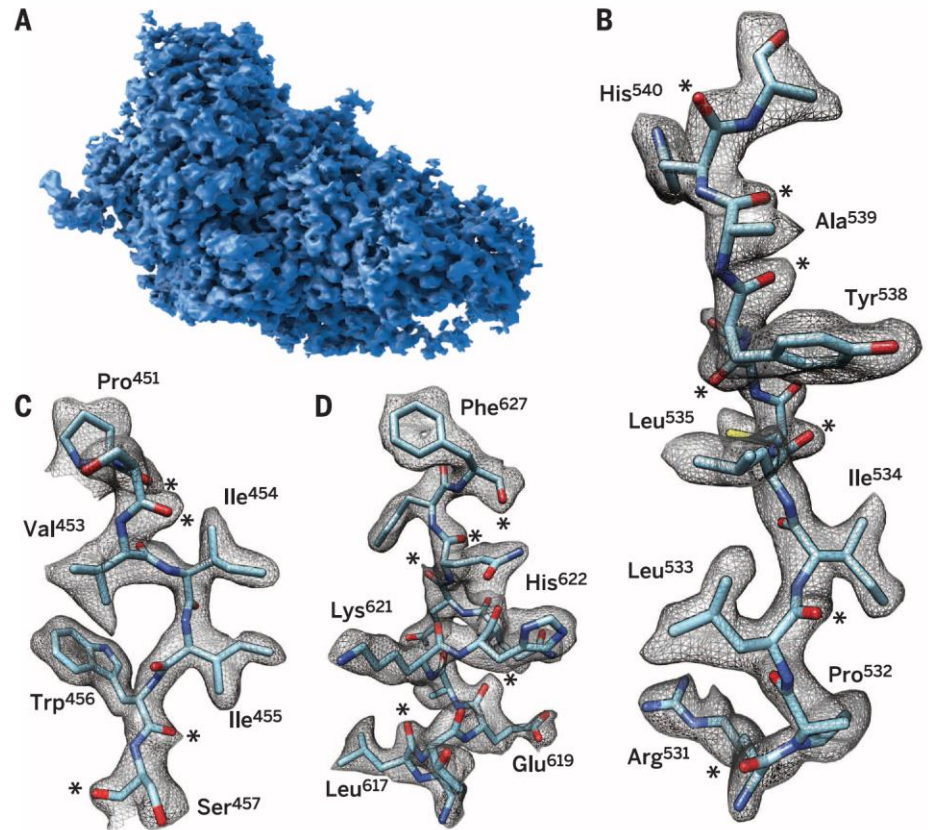
## 2.2 Å resolution cryo-EM structure of $\beta$ -galactosidase in complex with a cell-permeant inhibitor

Alberto Bartesaghi,<sup>1\*</sup> Alan Merk,<sup>1\*</sup> Soojay Banerjee,<sup>1</sup> Doreen Matthies,<sup>1</sup> Xiongwu Wu,<sup>2</sup> Jacqueline L. S. Milne,<sup>1</sup> Sriram Subramaniam<sup>1†</sup>

Science 2015



2D class averages



# New Technologies, Automation, & Computation are accelerating the field



**Krios at National University of Singapore**



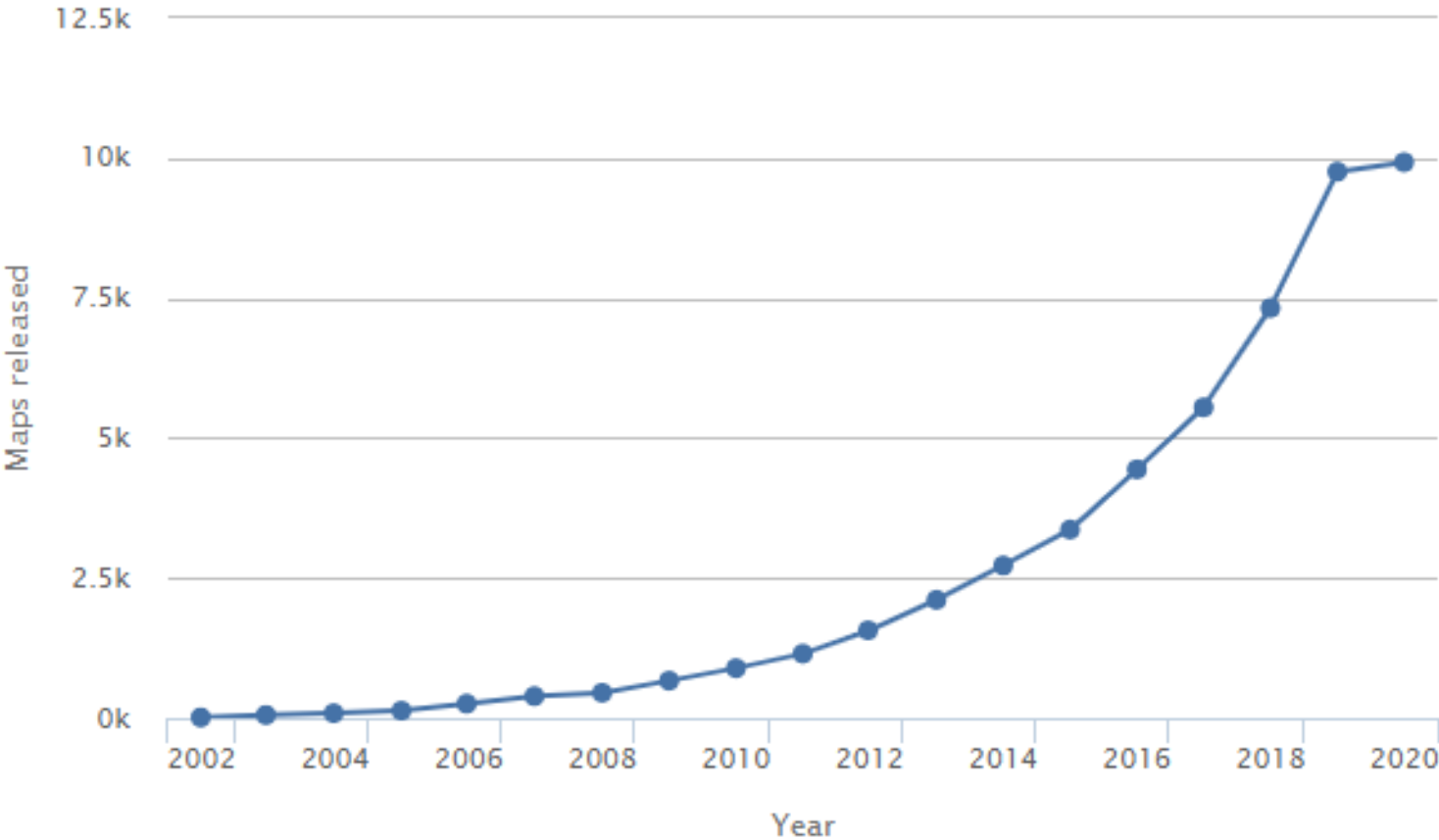
**Control room at Scripps Research Institute, La Jolla**



**Krios TEM installation on Yale's West Campus.**



### Cumulative number of maps released



Source: [https://www.ebi.ac.uk/pdbe/emdb/statistics\\_main.html/](https://www.ebi.ac.uk/pdbe/emdb/statistics_main.html/)

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

Science

RESEARCH ARTICLES

## Recognition of the amyloid precursor protein by human $\gamma$ -secretase

Rui Zhou<sup>1\*</sup>, Guanghui Yang<sup>1\*</sup>, Xuefei Guo<sup>1</sup>, Qiang Zhou<sup>2,3</sup>, Jianlin Lei<sup>1,4</sup>, Yigong Shi<sup>1,2†</sup>

<sup>1</sup>Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. <sup>2</sup>Institute of Biology, Westlake Institute for Advanced Study, Westlake University, 18 Shilongshan Road, Xihu District, Hangzhou 310024, Zhejiang Province, China. <sup>3</sup>School of Life Sciences, Westlake University, 18 Shilongshan Road, Xihu District, Hangzhou 310024, Zhejiang Province, China. <sup>4</sup>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  $\gamma$ -secretase is linked to Alzheimer's disease. We report an atomic structure of human  $\gamma$ -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  $\gamma$ -secretase). A hybrid  $\beta$ -sheet, which is formed by a  $\beta$ -strand from APP and two  $\beta$ -strands from PS1, guides  $\gamma$ -secretase to the scissile peptide bond of APP between its TM and  $\beta$ -strand. Residues at the interface between PS1 and APP are heavily targeted by recurring mutations from AD patients. This structure, together with that of  $\gamma$ -secretase bound to Notch, reveal contrasting features of substrate binding, which may be exploited toward design of substrate-specific inhibitors.

61YC

Recognition of the Amyloid Precursor Protein by Human gamma-secretase

DOI: 10.2210/pdb/6IYC/pdb EMDDataBank: 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

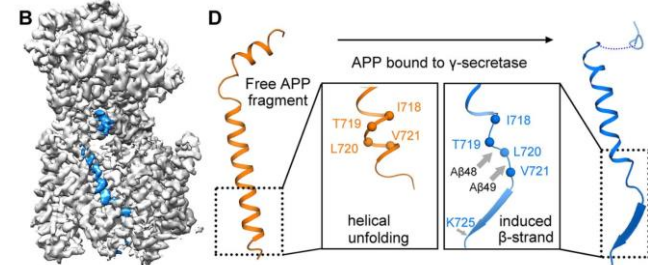
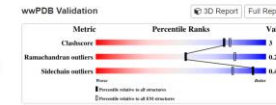
Experimental Data Snapshot

Method: ELECTRON MICROSCOPY

Resolution: 2.6 Å

Aggregation State: PARTICLE

Reconstruction Method: SINGLE PARTICLE



Science

RESEARCH ARTICLES

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

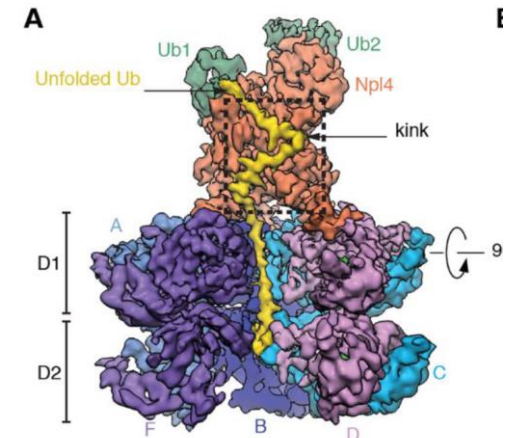
Edward C. Twomey<sup>1\*</sup>, Zhejian Ji<sup>1\*</sup>, Thomas E. Wales<sup>2</sup>, Nicholas O. Bodnar<sup>1</sup>, Scott B. Ficarro<sup>3,4</sup>, Jarrod A. Marto<sup>3,4</sup>, John R. Engen<sup>2</sup>, Tom A. Rapoport<sup>1†</sup>

<sup>1</sup>Department of Cell Biology, Harvard Medical School, and Howard Hughes Medical Institute, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>2</sup>Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA, USA. <sup>3</sup>Department of Cancer Biology, Department of Oncologic Pathology, and Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, MA 02115, USA. <sup>4</sup>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

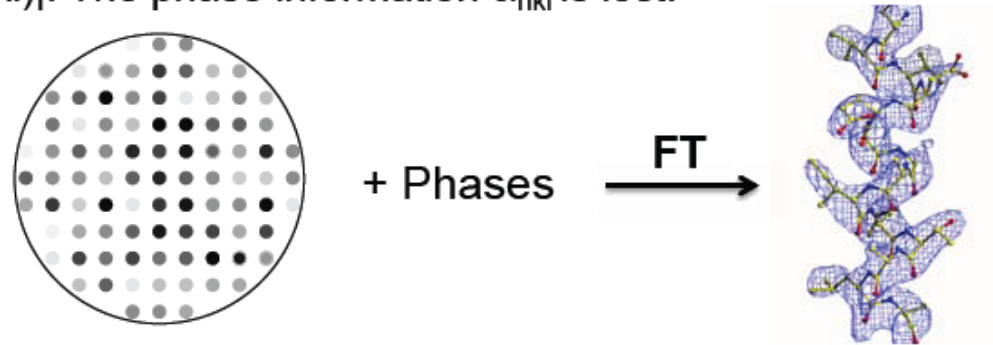
Cite as: E. C. Twomey *et al.*, *Science* 10.1126/science.aax1033 (2019).



PMID: 30630874;30598546;25918421;31249135

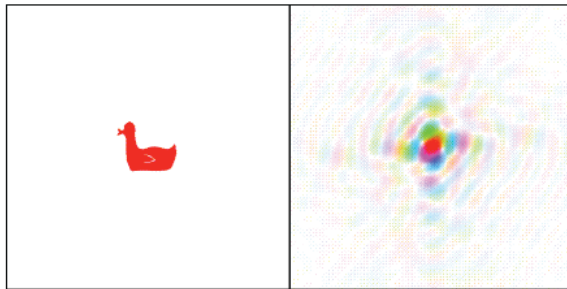
## Appendix

**The phase problem:**  $F(hkl)$  is a complex vector. Measured diffraction data give the amplitude  $|F(hkl)|$ . The phase information  $\alpha_{hkl}$  is lost!

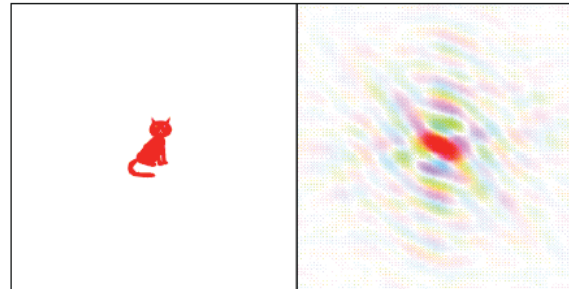


### How important are amplitude and phase?

Fourier Duck and his Fourier transform  
Phase is color coded

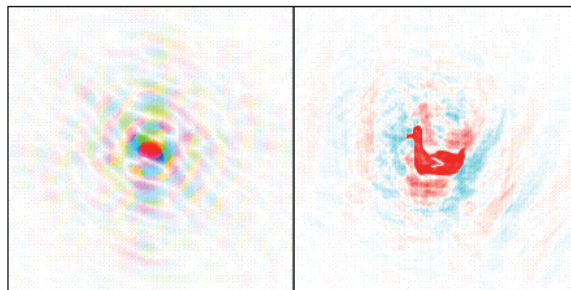


Fourier Cat and his Fourier transform  
Phase is color coded

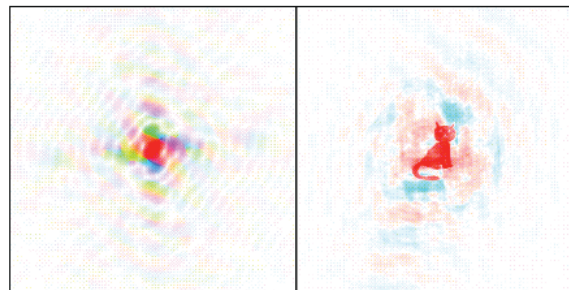


$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F(hkl)| e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}}$$

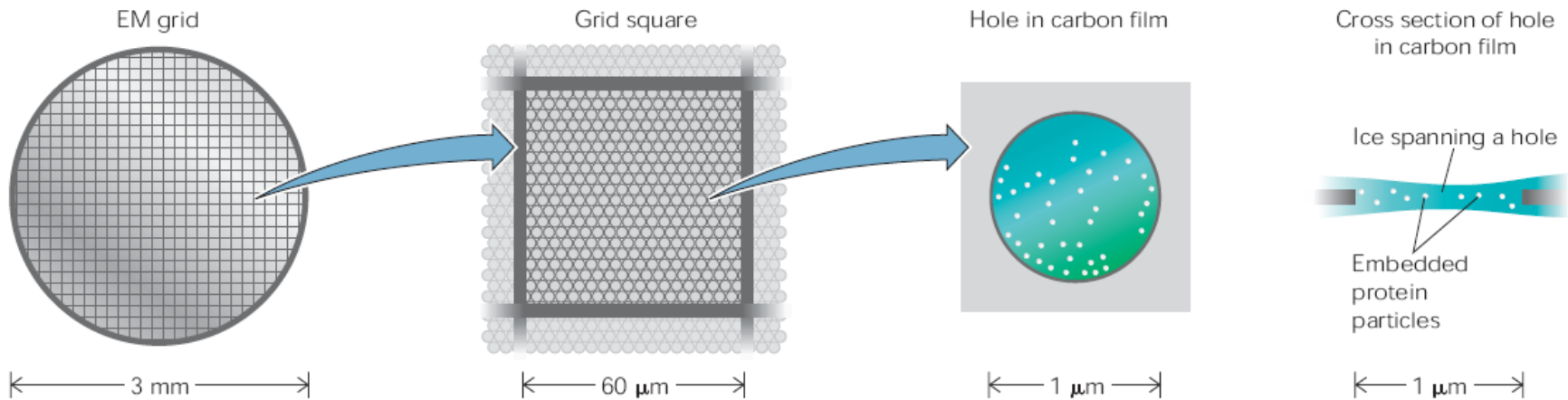
Duck phase and Cat amplitude



Cat phase and Duck amplitude



In a cryo-EM specimen, the fast-frozen sample is supported by a perforated carbon film



Adhering to a standard 3-mm electron microscope grid is a carbon film of  $\sim 500 \text{ \AA}$  thickness perforated with holes 1–2  $\mu\text{m}$  in diameter. The carbon film supports a 1,000- $\text{\AA}$  layer of buffer, in which the particles of interest are embedded. This layer is rapidly frozen in liquid ethane to form vitreous ice. The specimen is maintained continuously below  $-160^\circ\text{C}$  during storage and also during imaging in the electron microscope to prevent the formation of ice crystals.

# PDB explore

RCSB PDB Deposit Search Visualize Analyze Download Learn More

MyPDB Login

PDB-101

WORLDWIDE PDB PROTEIN DATA BANK

EMDataBank Unified Data Resource for 3DHR

NUCLEIC ACID DATABASE

StructuralBiology Knowledgebase



Welcome

Deposit

Search

Visualize

Analyze

Download

## Search Options

Drill Down by Categories

Advanced Search

Sequences

Ligands

Drugs & Drug Targets

Unreleased & New Entries

Browse by Annotation

PDB Statistics

## Explore the PDB Archive



Organism



UniProt Molecule Name



Taxonomy



Experimental Method



X-ray Resolution



Release Date



Polymer Type



Enzyme Classification



SCOP Classification



Protein Symmetry



Protein Stoichiometry



Membrane Proteins

Select a Organism category below:

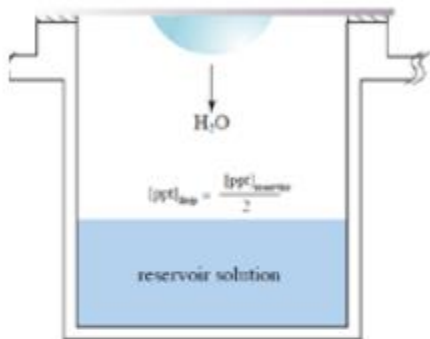


- Homo sapiens (30614)
- Escherichia coli (7911)
- Mus musculus (4945)
- Saccharomyces cerevisiae (3273)
- Bos taurus (2483)
- Rattus norvegicus (2371)
- Mycobacterium tuberculosis (1796)
- Other (60704)

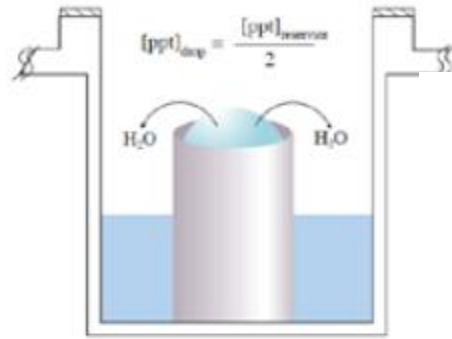
# Some Crystallization Methods:

Vapor diffusion

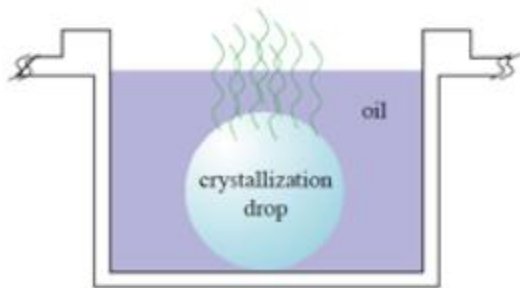
Hanging-drop



Sitting-drop



Batch:  
micro batch under oil



Dialysis

