X-Ray Crystallography

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Topics:

1. **Image Formation** *(optical illusions)*

   Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

2. **Protein Data Bank (PDB)**

   Data mining and Protein Structure Analysis Tools

---

3. **X-Ray Crystallography**

   a) 100 years of X-ray Crystallography

   b) Crystal Growth – Materials / Methods

   c) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit

   d) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron

   e)Theory of Diffraction – Bragg's Law / Reciprocal Space

   f) Data Collection – Methods / Detectors / Structure Factors

   g) Structure Solution – Phase Problem: MIR / MR / MAD

   h) Refinements and Models / Analysis and presentation of results

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Image Formation

Abbe (~1873):

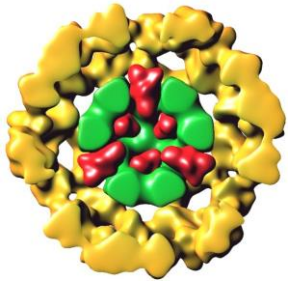
Limit Res. ~ λ/2

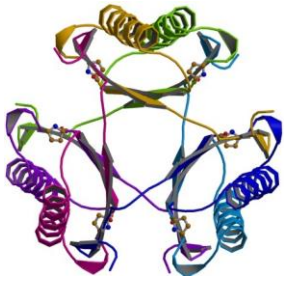
- Light Photography

  λ ~ 400 - 700 nm

- Electron Microscopy

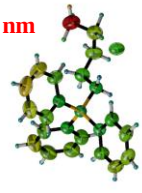
  λ ~ 0.001 - 0.1 nm

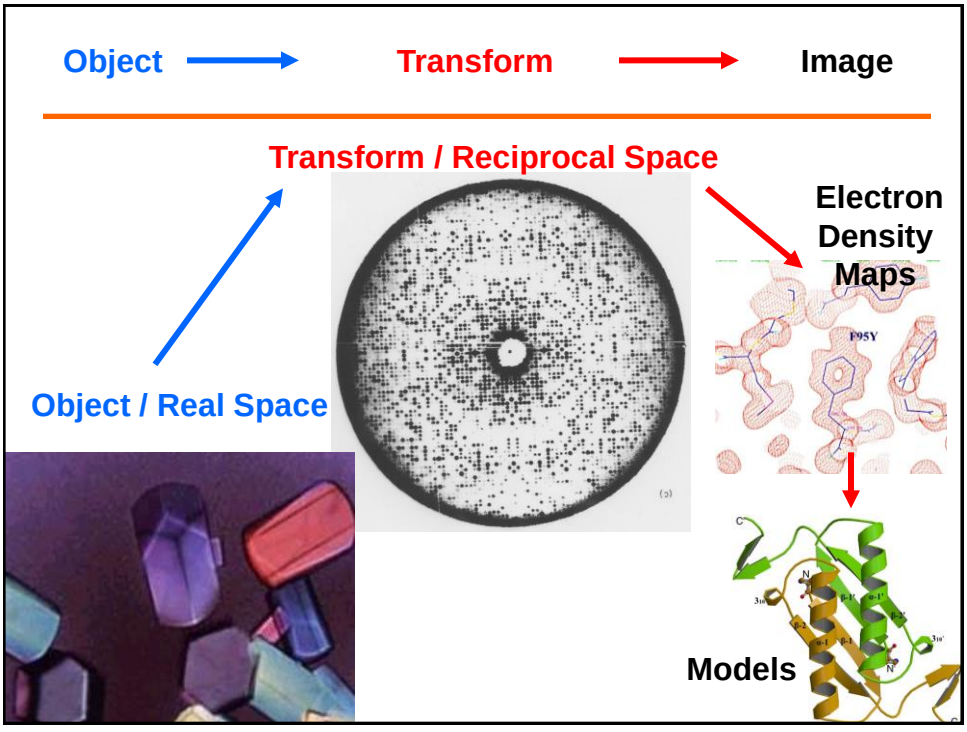




- X-Ray or NMR

  λ ~ 0.1 nm




Object
Transform
Image
Transform / Reciprocal Space
Electron Density Maps
Object / Real Space
F95Y
Models

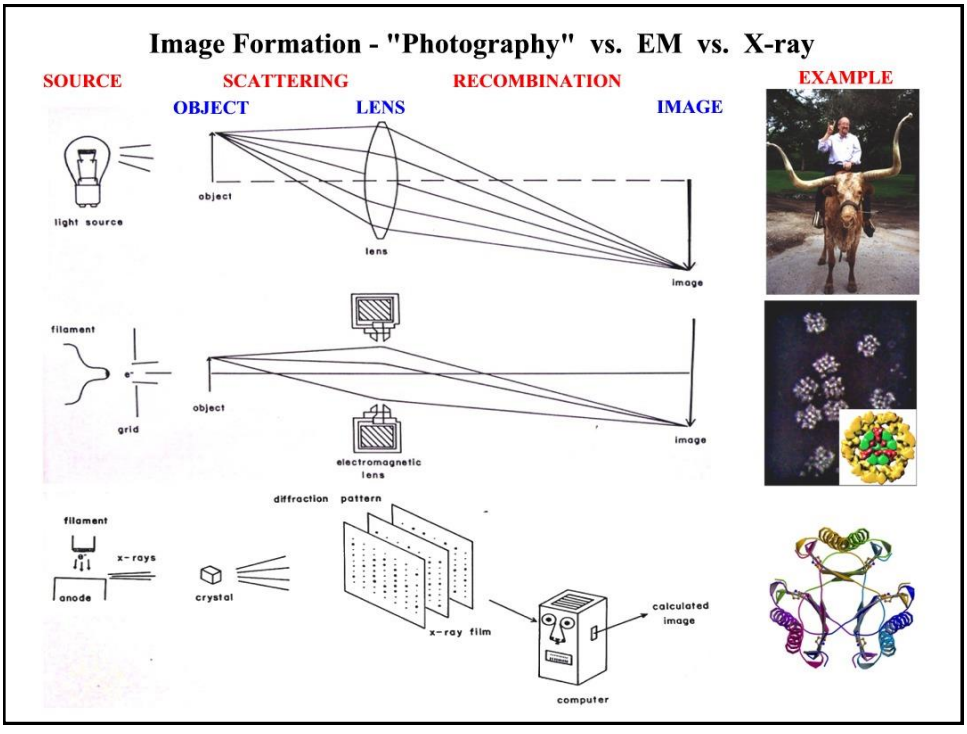
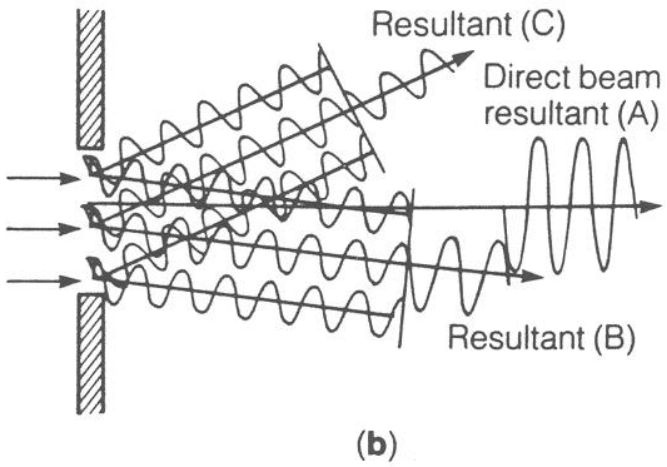


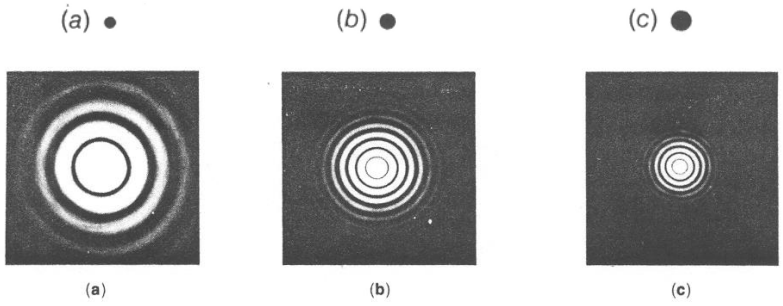
Image Formation - "Photography" vs. EM vs. X-ray
SOURCE
SCATTERING
RECOMBINATION
EXAMPLE
OBJECT
LENS
IMAGE
light source
object
lens
image
filament
grid
object
electromagnetic lens
image
filament
x-rays
anode
crystal
diffraction pattern
x-ray film
computer
calculated image

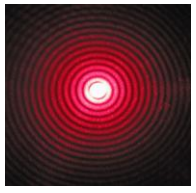
Transforms / Reciprocal Space



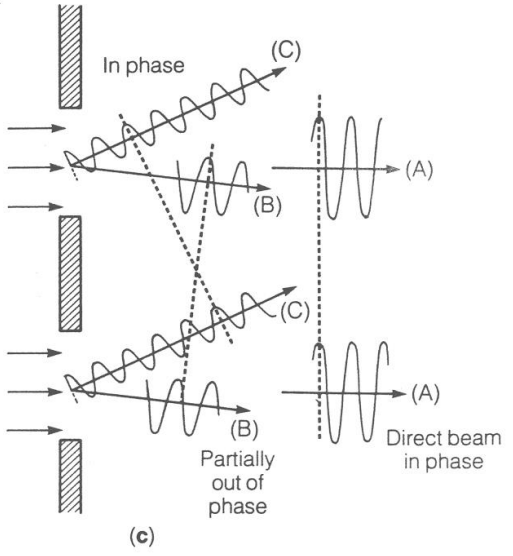
Transforms / Reciprocal Space

Different size holes

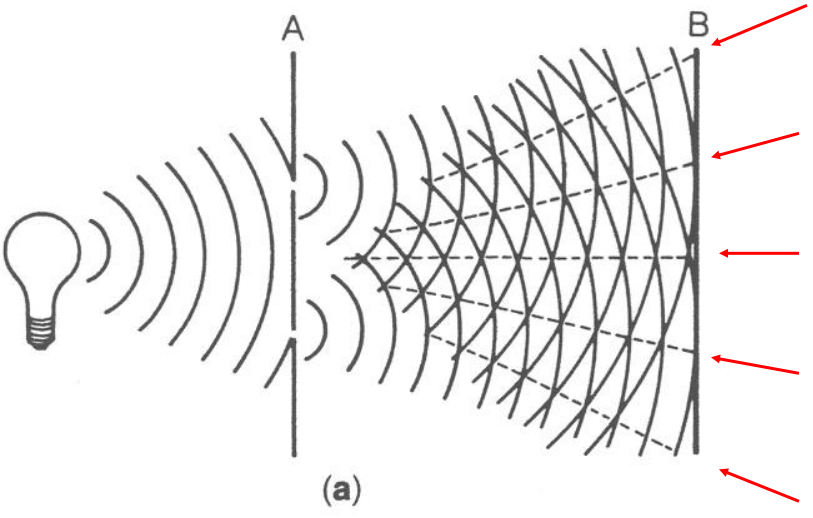




Transforms / Reciprocal Space

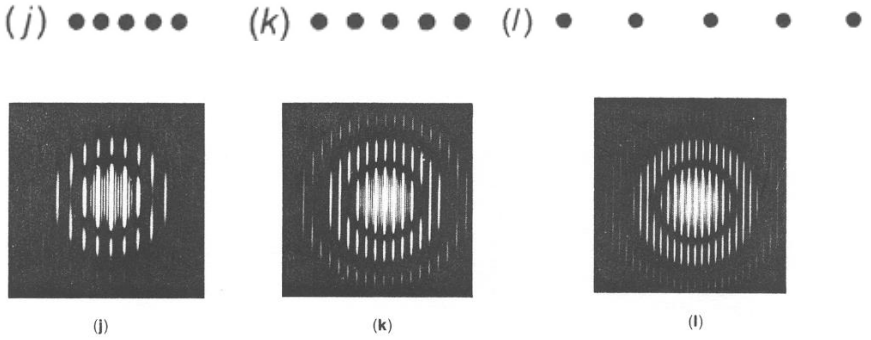


Transforms / Reciprocal Space



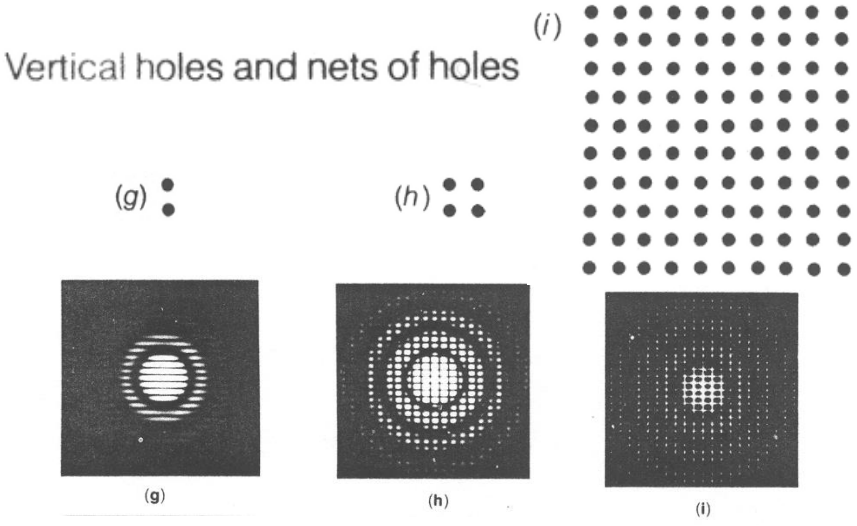
Transforms / Reciprocal Space

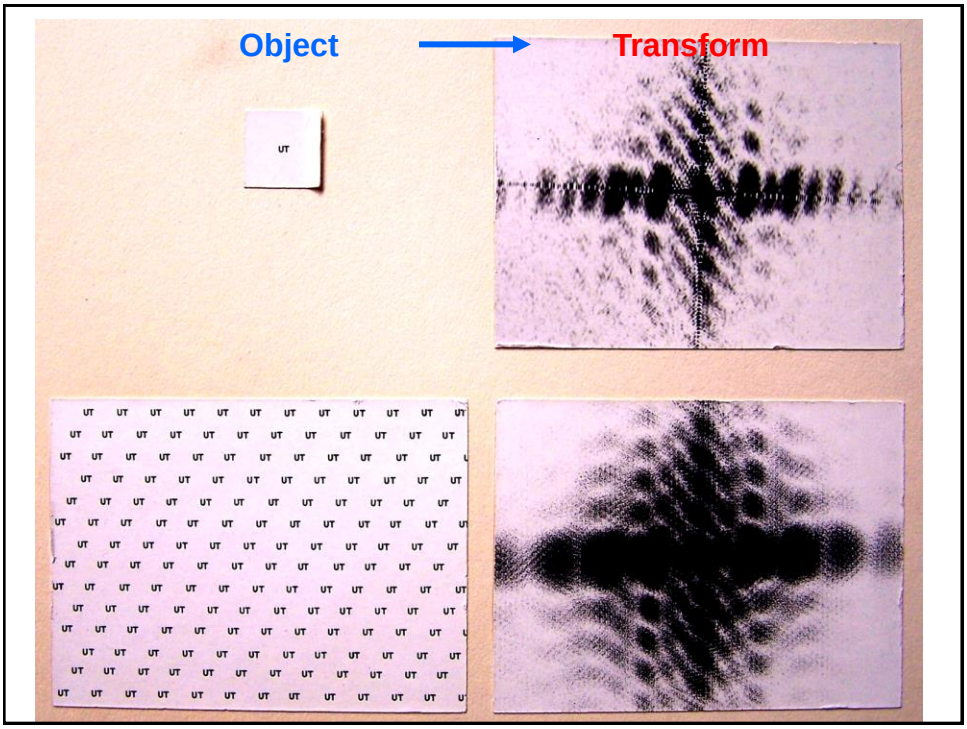
Five horizontal holes
with various spacings



Transforms / Reciprocal Space

Vertical holes and nets of holes




Object
Transform

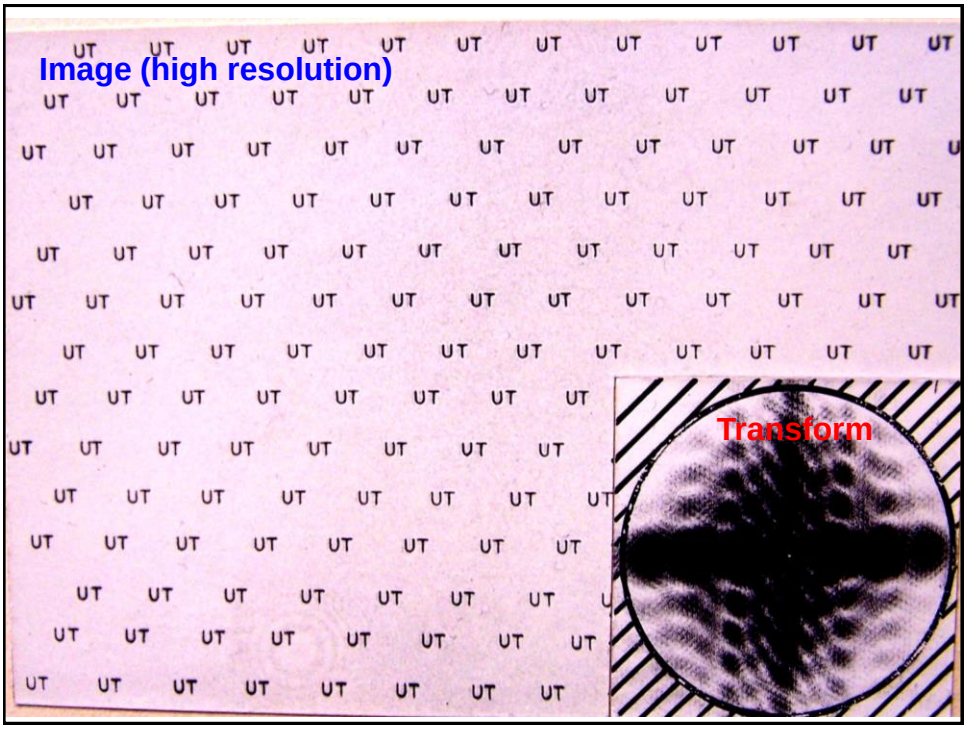
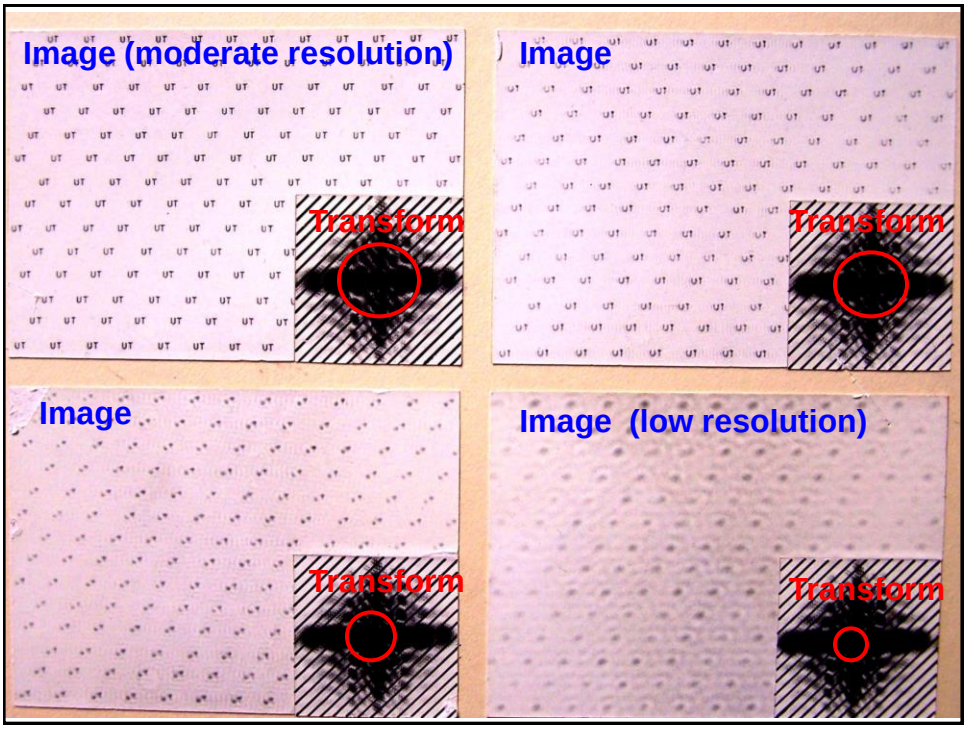
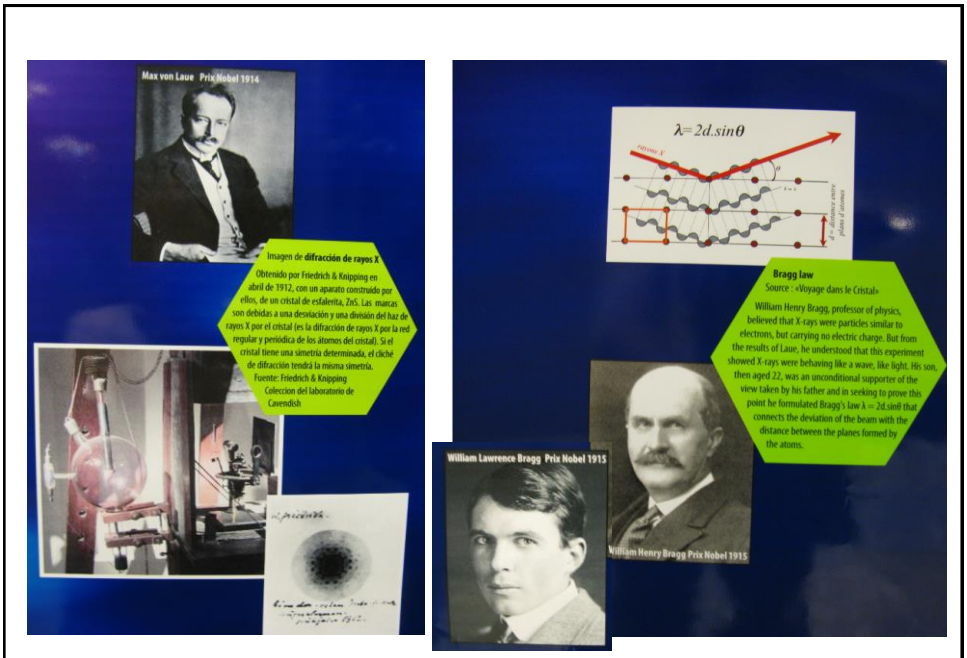


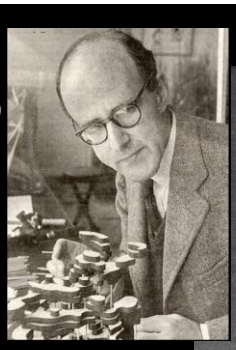
Image (high resolution)
Transform

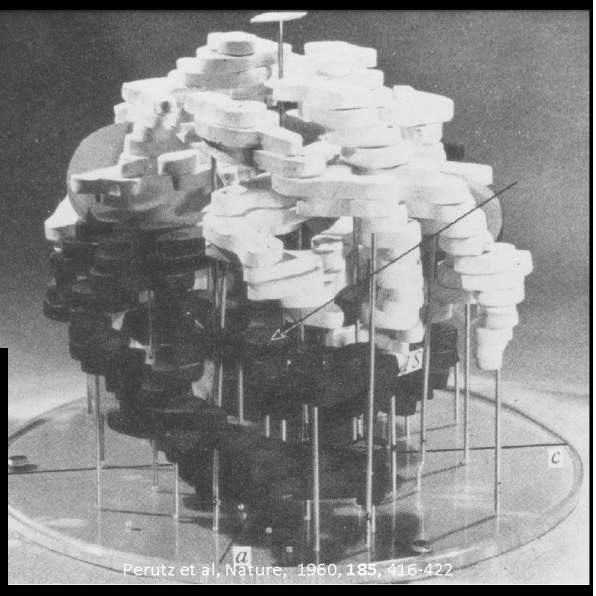




Max von Laue (1912/1914) **William & Lawrence Bragg (1913/1915)**

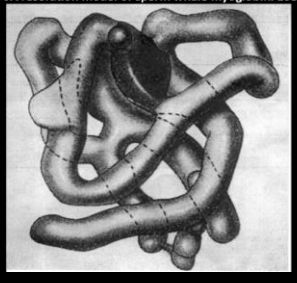
Max
circa 1959

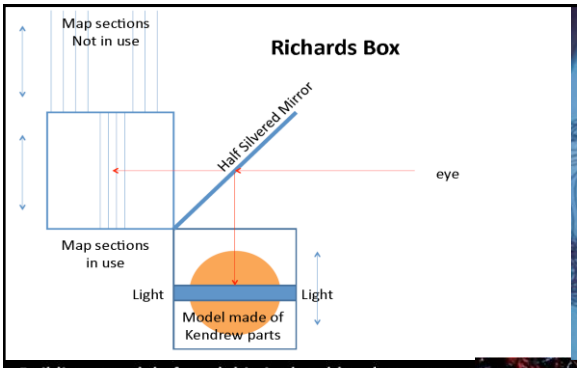


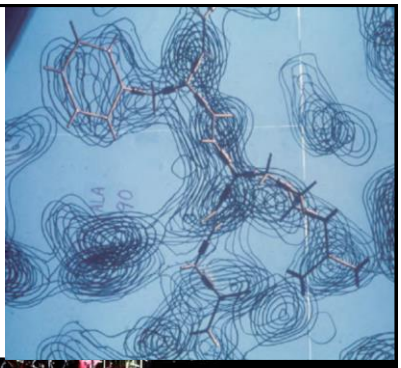


Perutz et al, Nature, 1960, 185, 416-422

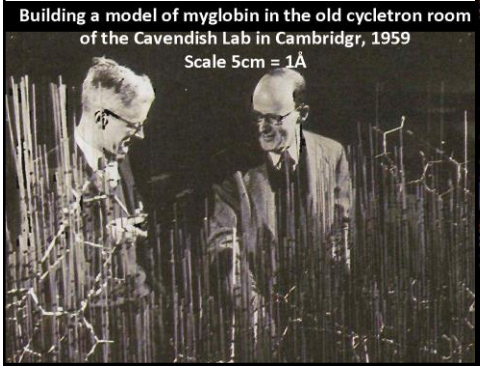
The 6Å resolution model of sperm whale myoglobin. 1957







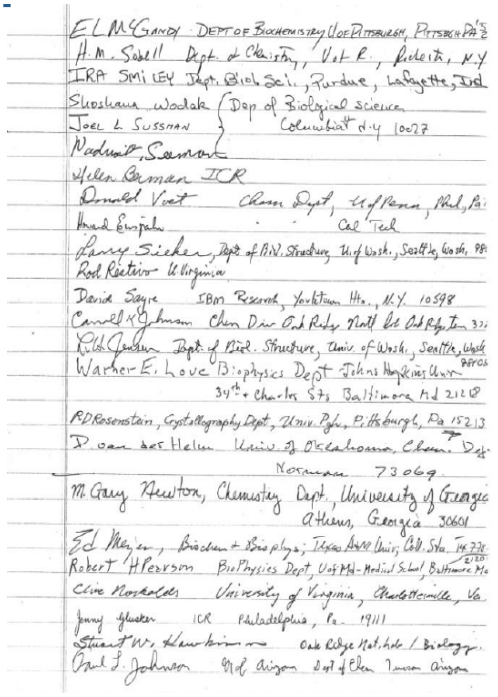
Building a model of myglobin in the old cycletron room of the Cavendish Lab in Cambridgr, 1959
Scale 5cm = 1Å



Discovering the nucleotide binding fold while building the Lactate dehydrodenase model
1970

Scale: 2cm = 1 Å

A petition to establish a central repository for atomic coordinate data of protein structures was written at the American Crystallographic Association Winter Meeting, Columbia, SC, February 1971


E L McGandy Dept of Biochemistry Univ Pittsburgh, Pittsburgh Pa
H. M. Sobell Dept. of Chemistry, Vet R., Roberts, N.Y.
IRA Smiley Dept. Biol. Sci., Purdue, Lafayette, Ind.
Shoshana Wodak Dep of Biological sciences
Joel L Sussman Columbia U. N.Y. 10027
Rodewald, Samson
Helen Berman ICR
Donald Voet Chem Dept, U of Penn, Phil, Pa
Brad Sutton Cal Tech
Larry Sieker, Dept of B.M. Structure, U of Wash., Seattle, Wash. 98
Rod Rietiov U Virginia
David Sayre IBM Research, Yorktown Hts., N.Y. 10598
Carroll K Johnson Chem Div Oak Ridge Natl Lab Oak Ridge Tenn 37
Lyle Jensen Dept of Biol. Structure, Univ of Wash., Seattle, Wash
Walter E. Love Biophysics Dept Johns Hopkins Univ 98105
34th + Charles Sts Baltimore Md 21218
R D Rosenstein, Crystallography Dept, Univ. Pgh., Pittsburgh, Pa 15213
D. van der Helm Univ. of Oklahoma, Chem. Dept.
Norman 73069
M. Gary Newton, Chemistry Dept., University of Georgia
Athens, Georgia 30601
Ed Meyer, Biochem + Biophys; Texas A&M Univ, Coll. Sta. Tx 778
Robert H Pierson BioPhysics Dept, Univ Md - Medical School, Baltimore Md 21201
Clive Maskalas University of Virginia, Charlottesville, Va.
Jenny Glusker ICR Philadelphia, Pa. 19111
Stuart W. Hawkinson Oak Ridge Natl Lab / Biology
Carl I. Johnson Mel Anson Dept of Chem Tucson Arizon

Michael Rossmann – PDB40 address

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is received. The total holding will be announced annually in the organic bibliographic volumes of the reference series "Molecular Structures and Dimensions" published for the Crystallographic Data Centre and the International Union of Crystallography by Oosthoek's, Utrecht.

The success of the proposed system will depend on the response of the protein crystallographers supplying data. These will be accepted either "raw" or refined, in machine-readable form or as manuscripts. Laboratories intending to join the scheme should communicate with Mrs Olga Kennard or Dr D. G. Watson at the University Chemical Laboratories, Lensfield Road, Cambridge, who are responsible for the organization of the system. Data can be submitted to Cambridge or to Dr W. C. Hamilton at the Brookhaven National Laboratory, Upton, New York 11973, where the data will be computer processed.

The two centres will maintain identical files and both will provide data services. The new data bank is intended to supplement existing publication media so that depositing material in this form is not a substitute for the publication of the

Creation of PDB announced in 1971 (*Nature New Biology* 1971, 233, 223)


Walter Hamilton, Helen Berman, Tom Koetzle in 1972

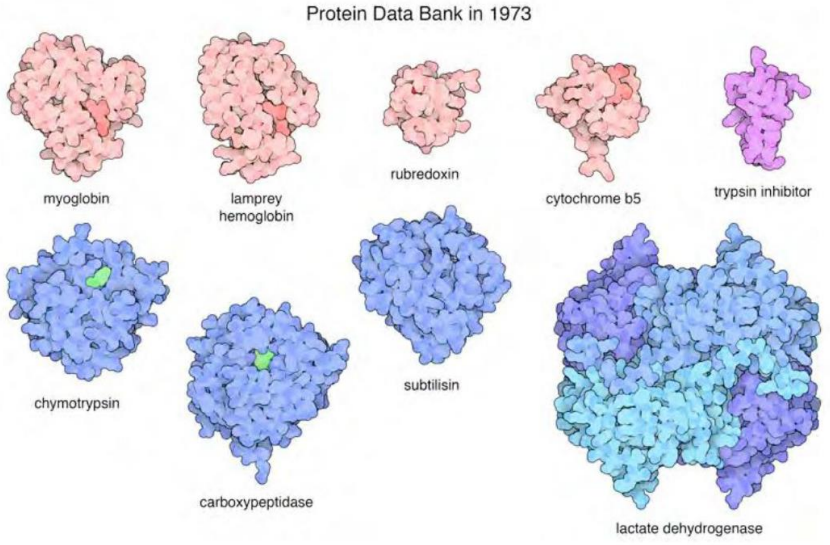

Walter Hamilton and Harold W. Wyckoff at the CSHL meeting in 1971

The First protein structures

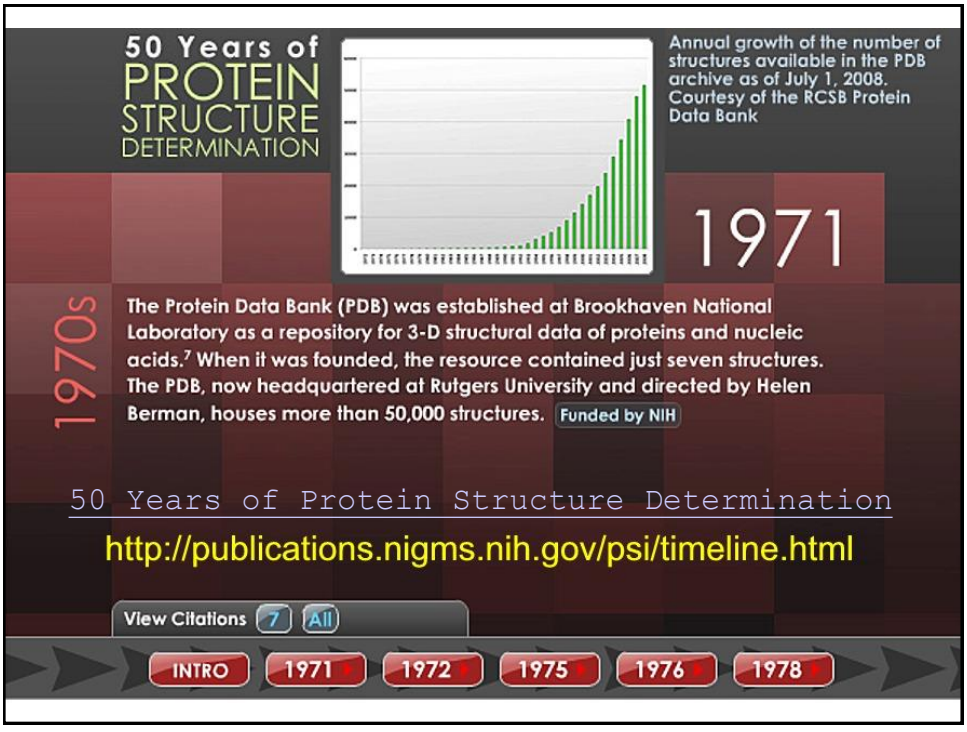
1958	6.0 Å Myoglobin	Cambridge	John Kendrew
1959	5.5 Å oxy-Haemogbin	Cambridge	Max Perutz
1959	2.0 Å Myglobin	Cambridge	John Kendrew
1965	HEW lysozyme	RI London	David Phillips
1967	**Carboxypeptidase**	**Harvard**	**Bill Lipscomb**
1968	**Ribonuclease**	**Yale**	**Fred Richards**
1968	**Chymotrypsin**	**Cambridge**	**David Blow**
1968	Papain	Groningen	Jan Drenth
1970	2.8 Å oxy Haemogbin	Cambridge	Max Perutz
1970	De-oxy Haemoglobin	Cambridge	Max Perutz
1970	**Lactate dehydrognase**	**Purdue**	**Michael Rossmann**
1971	Staphylococcal nuclease	MIT	Al Cotton
1971	Carbonic anhydrase	Uppsala	Anders Liljas
1972	**Subtilisin**	**Groningen**	**Wim Hol**
1972	**Lamprey Haemoglobin**	**Johns Hopkins**	**Werner Love**
1972	**Rubridoxin**	**U of Washington**	**Lyle Jensen**
1972	**Trypsin inhibitor**	**Max Plank**	**Robert Huber**
1973	**Cytochrome b5**	**Washington U**	**Scott Matthews**

1973 PDB holdings in red

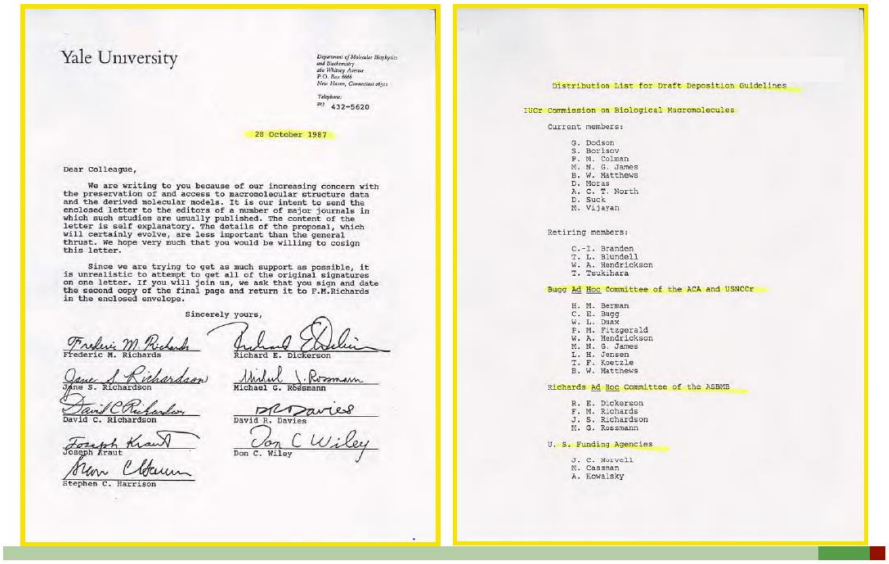
Michael Rossmann – PDB40 address



Michael Rossmann – PDB40 address


50 Years of
PROTEIN
STRUCTURE
DETERMINATION
Annual growth of the number of structures available in the PDB archive as of July 1, 2008. Courtesy of the RCSB Protein Data Bank
1971
1970s
The Protein Data Bank (PDB) was established at Brookhaven National Laboratory as a repository for 3-D structural data of proteins and nucleic acids.[7] When it was founded, the resource contained just seven structures. The PDB, now headquartered at Rutgers University and directed by Helen Berman, houses more than 50,000 structures. Funded by NIH
50 Years of Protein Structure Determination
http://publications.nigms.nih.gov/psi/timeline.html
View Citations 7 All
INTRO
1971
1972
1975
1976
1978

1987: Users Compel Deposition



Yale University

Department of Molecular Biophysics and Biochemistry
260 Whitney Avenue
P.O. Box 6666
New Haven, Connecticut 06511

Telephone:
203 432-5620

28 October 1987

Dear Colleague,

We are writing to you because of our increasing concern with the preservation of and access to macromolecular structure data and the derived molecular models. It is our intent to send the enclosed letter to the editors of a number of major journals in which such studies are usually published. The content of the letter is self explanatory. The details of the proposal, which will certainly evolve, are less important than the general thrust. We hope very much that you would be willing to cosign this letter.

Since we are trying to get as much support as possible, it is unrealistic to attempt to get all of the original signatures on one letter. If you will join us, we ask that you sign and date the second copy of the final page and return it to F.M.Richards in the enclosed envelope.

Sincerely yours,

Frederic M. Richards
Richard E. Dickerson
Jane S. Richardson
Michael G. Rossmann
David C. Richardson
David R. Davies
Joseph Kraut
Don C. Wiley
Stephen C. Harrison

Distribution List for Draft Deposition Guidelines

IUCr Commission on Biological Macromolecules

Current members:

G. Dodson
S. Hovlov
P. M. Colman
M. N. G. James
B. W. Matthews
D. Moras
A. C. T. North
D. Suck
N. Vijayan

Retiring members:

C.-I. Branden
T. L. Blundell
W. A. Hendrickson
Z. Iwakhara

Buge Ad Hoc Committee of the ACA and USNCCr

H. M. Berman
C. E. Bugg
W. L. Duax
F. H. Fitzgerald
W. A. Hendrickson
M. N. G. James
L. H. Jensen
Z. F. Koetzle
B. W. Matthews

Richards Ad Hoc Committee of the ASBMB

R. E. Dickerson
F. M. Richards
J. S. Richardson
M. G. Rossmann

U. S. Funding Agencies

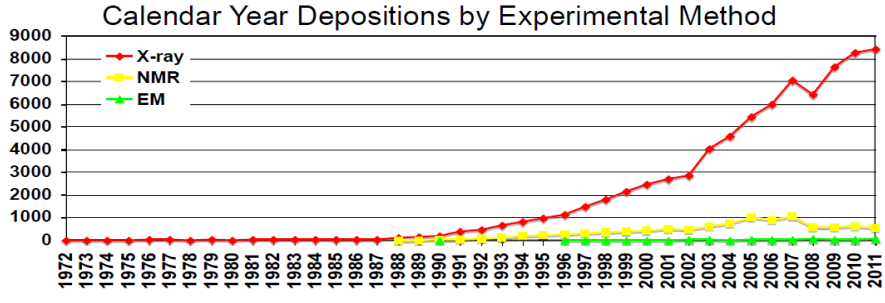
J. C. Norvell
N. Cassman
A. Kowalsky

Stephen K. Burley – PDB40 address

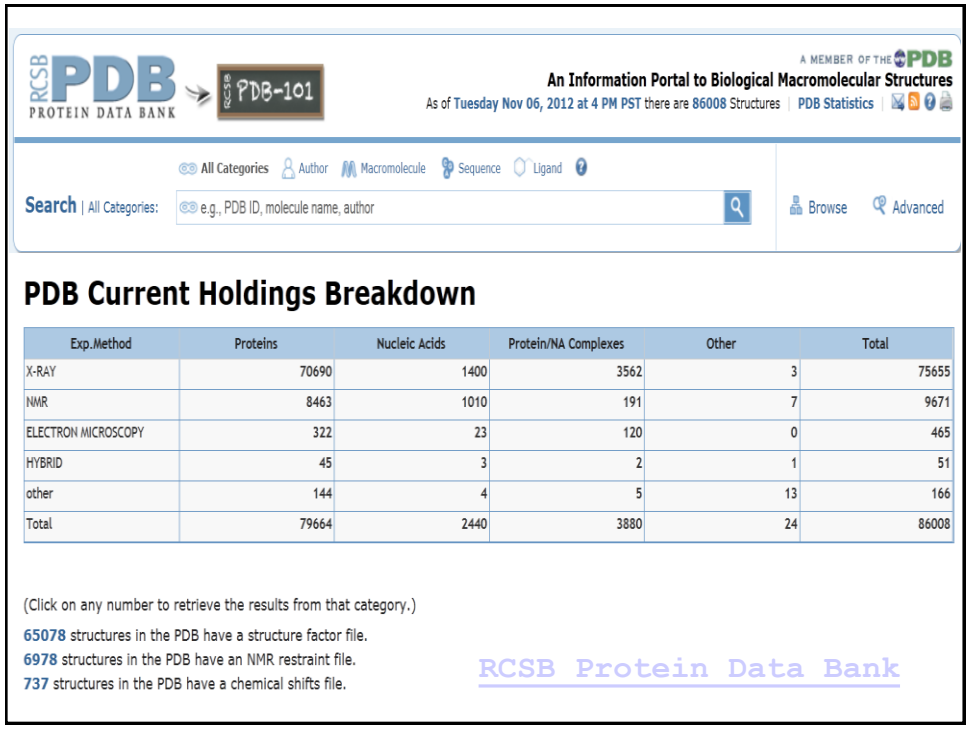
10,000-Fold Growth in Four Decades

http://www.wwpdb.org/PDB40.html

- 7→>76,000 entries
- 2011 will see ~9,000 depositions
- Electron Microscopy is beginning to hit its stride



Stephen K. Burley – PDB40 address



PDB Current Holdings Breakdown

Exp.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	70690	1400	3562	3	75655
NMR	8463	1010	191	7	9671
ELECTRON MICROSCOPY	322	23	120	0	465
HYBRID	45	3	2	1	51
other	144	4	5	13	166
Total	79664	2440	3880	24	86008

(Click on any number to retrieve the results from that category.)

65078 structures in the PDB have a structure factor file.
6978 structures in the PDB have an NMR restraint file.
737 structures in the PDB have a chemical shifts file.

RCSB Protein Data Bank

Statistics For PDB Structures That Are Deposited And Processed By Year And Site

Year	Total Depositions	Deposited To			Processed By		
		RCSB PDB	PDBj	PDBe	RCSB PDB	PDBj	PDBe
2000	2983	2445	10	528	2297	158	528
2001	3287	2673	118	496	2408	383	496
2002	3565	2769	289	507	2401	657	507
2003	4830	3488	673	669	3135	1026	669
2004	5508	3796	900	812	3082	1614	812
2005	6678	4507	1166	1005	3563	2110	1005
2006	7282	5145	1052	1085	4252	1945	1085
2007	8130	5399	1603	1128	4703	2299	1128
2008	7073	5452	648	973	4106	1994	973
2009	8300	6715	527	1058	5069	2173	1058
2010	8878	6912	593	1373	5464	2041	1373
2011	9250	7172	582	1496	5938	1816	1496
2012	9972	7695	601	1676	6409	1887	1676
2013	9010	6856	607	1547	5774	1689	1547
TOTAL	94746	71024	9369	14353	58601	21792	14353

Note: Includes theoretical models and entries later withdrawn or obsoleted
Last Updated: 6 Nov 2013

PDB

20 Person Years→20 Person Days

- Faster and Faster Computing
- Graphical Display (Geis→Frodo→O→COOT→…)
- Simulated Annealing Refinement
- Gene Cloning/Protein Expression Systems
- Protein Purification/Engineering
- Crystallization Strategies (Factorial, LCP, …)
- Data Collection: Cryogenics/Area Detectors
- Synchrotron Beamlines→MAD/SAD Phasing
- Automated Map Interpretation/Model Building
- Micro Focus X-ray Beamlines

Stephen K. Burley – PDB40 address

Analyze – structure (Ramachandran Plot) and biochemistry

Publish in leading biochemical or structural biology journal

Contribute results (coordinates, etc.) to PDB

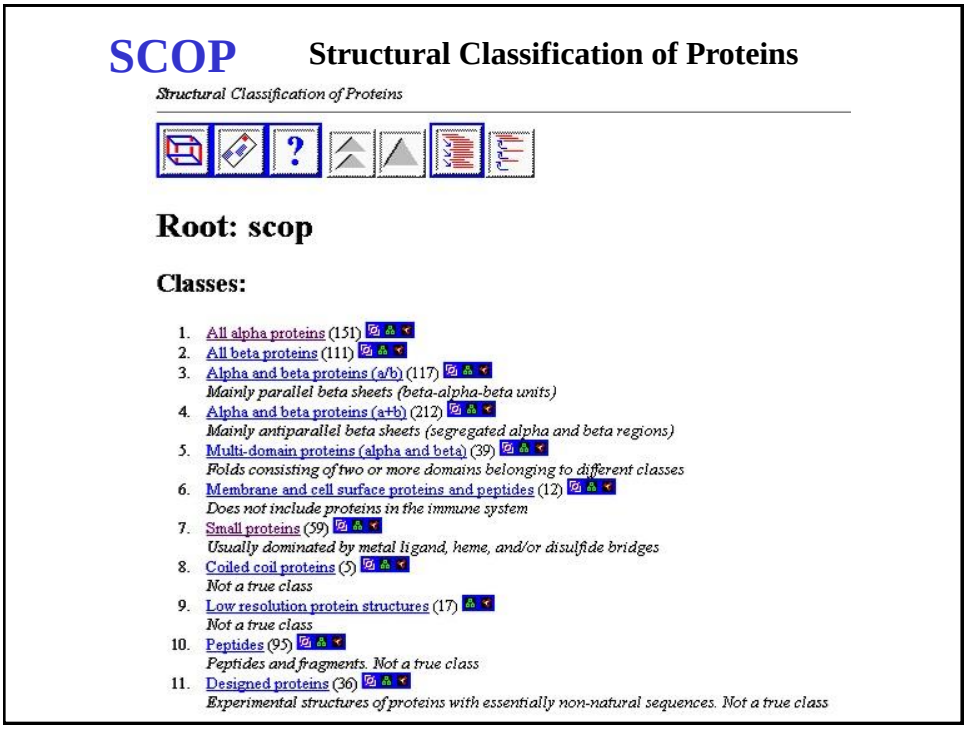
****

Data Mining

Visualization programs (Cn3D / RasMol / SwissPDBV / etc)

SCOP – Structural Classification of Proteins

CATH – Classification / Arch / Topology

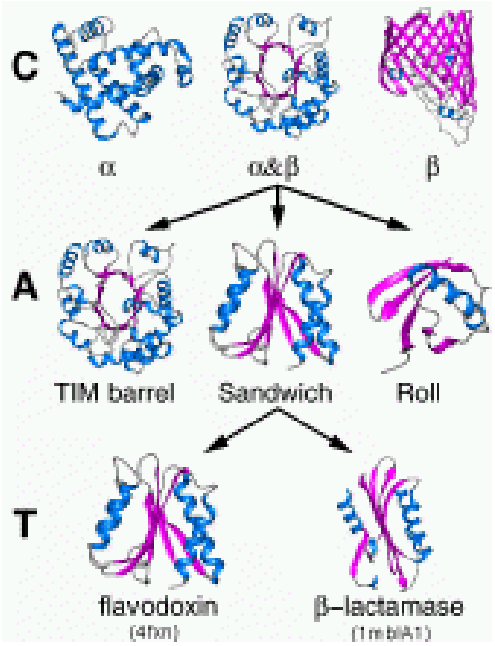


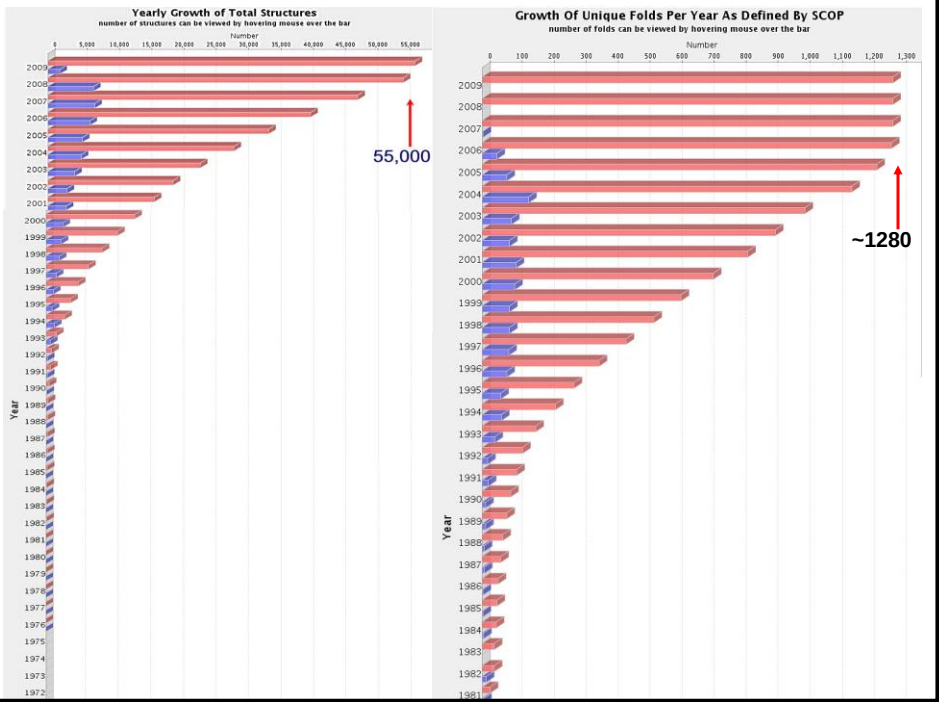
CATH - Protein Structure Classification

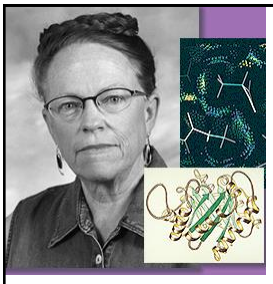
CATH is a novel hierarchical classification of protein domain structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T), and Homologous (H) Superfamily

Class, derived from secondary structure content, is assigned for more than 90% of protein structures automatically. **Architecture**, which describes the gross orientation of secondary structures, independent of connectivities, is currently assigned manually. The **topology** level clusters structures according to their topological connections and numbers of secondary structures. The **homologous superfamilies** cluster proteins with highly similar structures and functions. The assignments of structures to toplogy families and homologous superfamilies are made by sequence and structure comparisons.

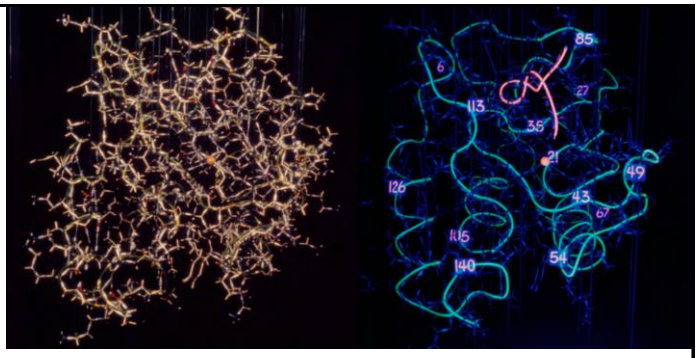
CATH


C
α
α&β
β
A
TIM barrel
Sandwich
Roll
T
flavodoxin
(4fxn)
β–lactamase
(1mbiA1)

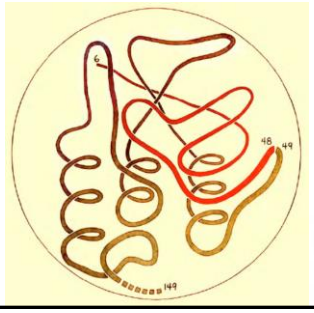

Yearly Growth of Total Structures
number of structures can be viewed by hovering mouse over the bar
Number
Year
55,000
Growth Of Unique Folds Per Year As Defined By SCOP
number of folds can be viewed by hovering mouse over the bar
Number
Year
~1280

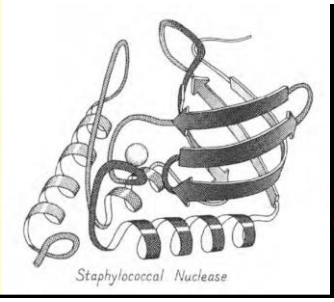


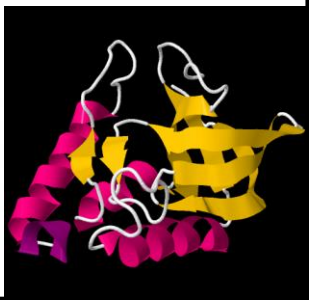
Jane Richardson –
Duke Univ.

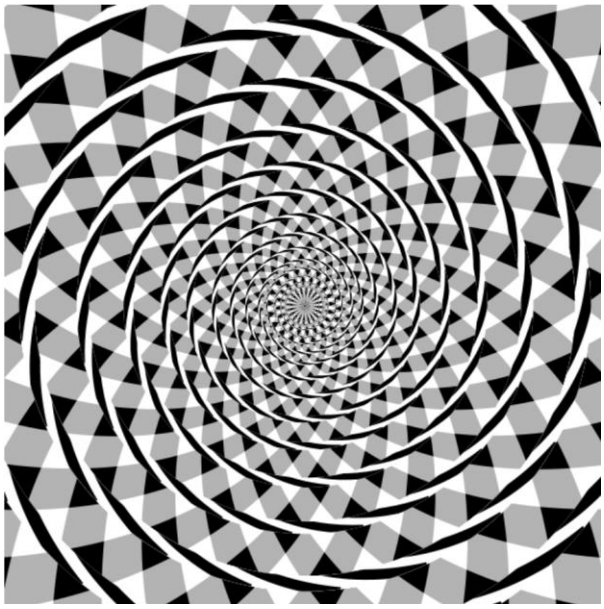


Models of Staphylococcal nuclease





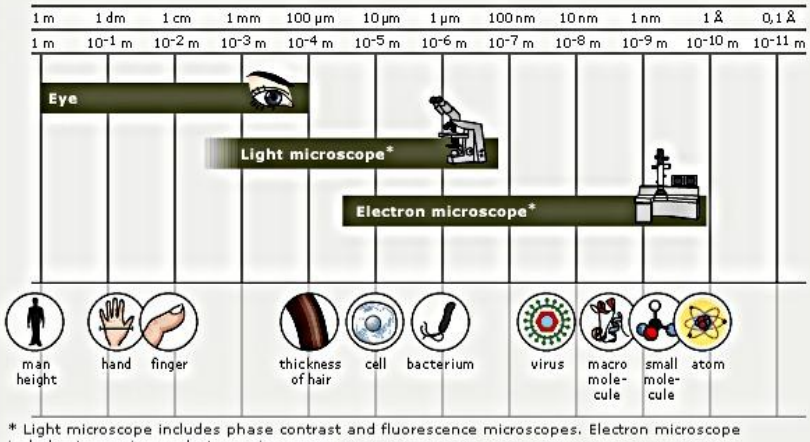




MICROSCOPES BACK

Resolving Power Line

What can you see with the different types of microscopes? The human eye is capable of distinguishing objects down to a fraction of a millimeter. With the use of light and electron microscopes it is possible to see down to an angstrom and study everything from different cells and bacteria to single molecules or even atoms.



* Light microscope includes phase contrast and fluorescence microscopes. Electron microscope includes transmisson electron microscope.

MICROSCOPES BACK

Time Line

Z. Janssen

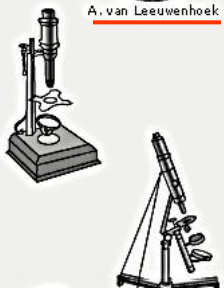
14th century – The art of grinding lenses is developed in Italy and spectacles are made to improve eyesight.

1590 – Dutch lens grinders Hans and Zacharias Janssen make the first microscope by placing two lenses in a tube.

A. van Leeuwenhoek

1667 – Robert Hooke studies various object with his microscope and publishes his results in Micrographia. Among his work were a description of cork and its ability to float in water.

1675 – Anton van Leeuwenhoek uses a simple microscope with only one lens to look at blood, insects and many other objects. He was first to describe cells and bacteria, seen through his very small microscopes with, for his time, extremely good lenses.



18th century – Several technical innovations make microscopes better and easier to handle, which leads to microscopy becoming more and more popular among scientists. An important discovery is that lenses combining two types of glass could reduce the chromatic effect, with its disturbing halos resulting from differences in refraction of light.

E. Abbe

R. Zsigmondy

F. Zernike

E. Ruska

G. Binnig and H. Rohrer

1830 – Joseph Jackson Lister reduces the problem with spherical aberration by showing that several weak lenses used together at certain distances gave good magnification without blurring the image.

1878 – Ernst Abbe formulates a mathematical theory correlating resolution to the wavelength of light. Abbes formula make calculations of maximum resolution in microscopes possible.

1903 – Richard Zsigmondy develops the ultramicroscope and is able to study objects below the wavelength of light.
The Nobel Prize in Chemistry 1925 »

1932 – Frits Zernike invents the phase-contrast microscope that allows the study of colorless and transparent biological materials.
The Nobel Prize in Physics 1953 »

1938 – Ernst Ruska develops the electron microscope. The ability to use electrons in microscopy greatly improves the resolution and greatly expands the borders of exploration.
The Nobel Prize in Physics 1986 »

1981 – Gerd Binnig and Heinrich Rohrer invent the scanning tunneling microscope that gives three-dimensional images of objects down to the atomic level.
The Nobel Prize in Physics 1986 »

Resolution Limit

The diffraction limit of a microscope is,

$$d = \frac{\lambda}{2n\sin\theta}$$

where d is the resolvable feature size, λ is the wavelength of light, n is the index of refraction of the medium being imaged in, and θ (depicted as α in the inscription) is the half-angle subtended by the optical objective lens.

Memorial to Ernst Abbe
University of Jena

Secondary school named Ernst Abbe Gymnasium

X-Ray Crystallography

"If a picture is worth a thousand words, then a macromolecular structure is priceless to a physical biochemist." – van Holde

Topics:

1. **Image Formation** ***(optical illusions)***

   Resolution / Wavelength (Amplitude, Phase) / Light Microscopy / EM / X-ray / (NMR)

2. **Protein Data Bank (PDB)**

   Data mining and Protein Structure Analysis Tools

---

3. **X-Ray Crystallography**

   a) 100 years of X-ray Crystallography

   b) Crystal Growth – Materials / Methods

   c) Crystal Lattices - Lattice Constants / Space Groups / Asymmetric Unit

   d) X-ray Sources – Sealed Tube / Rotation Anode / Synchrotron

   e)Theory of Diffraction – Bragg's Law / Reciprocal Space

   f) Data Collection – Methods / Detectors / Structure Factors

   g) Structure Solution – Phase Problem: MIR / MR / MAD

   h) Refinements and Models / Analysis and presentation of results