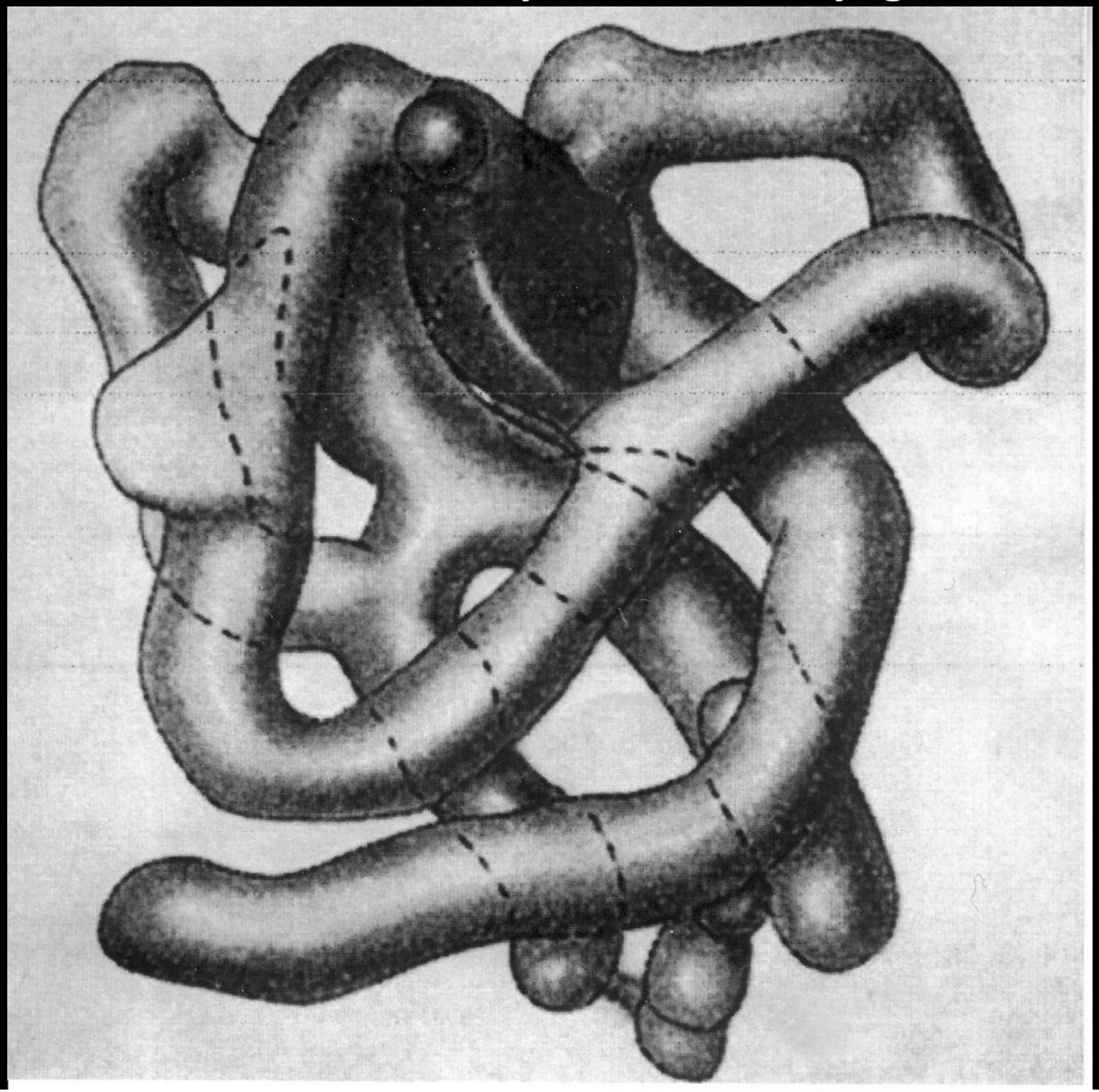
**Michael Rossmann**

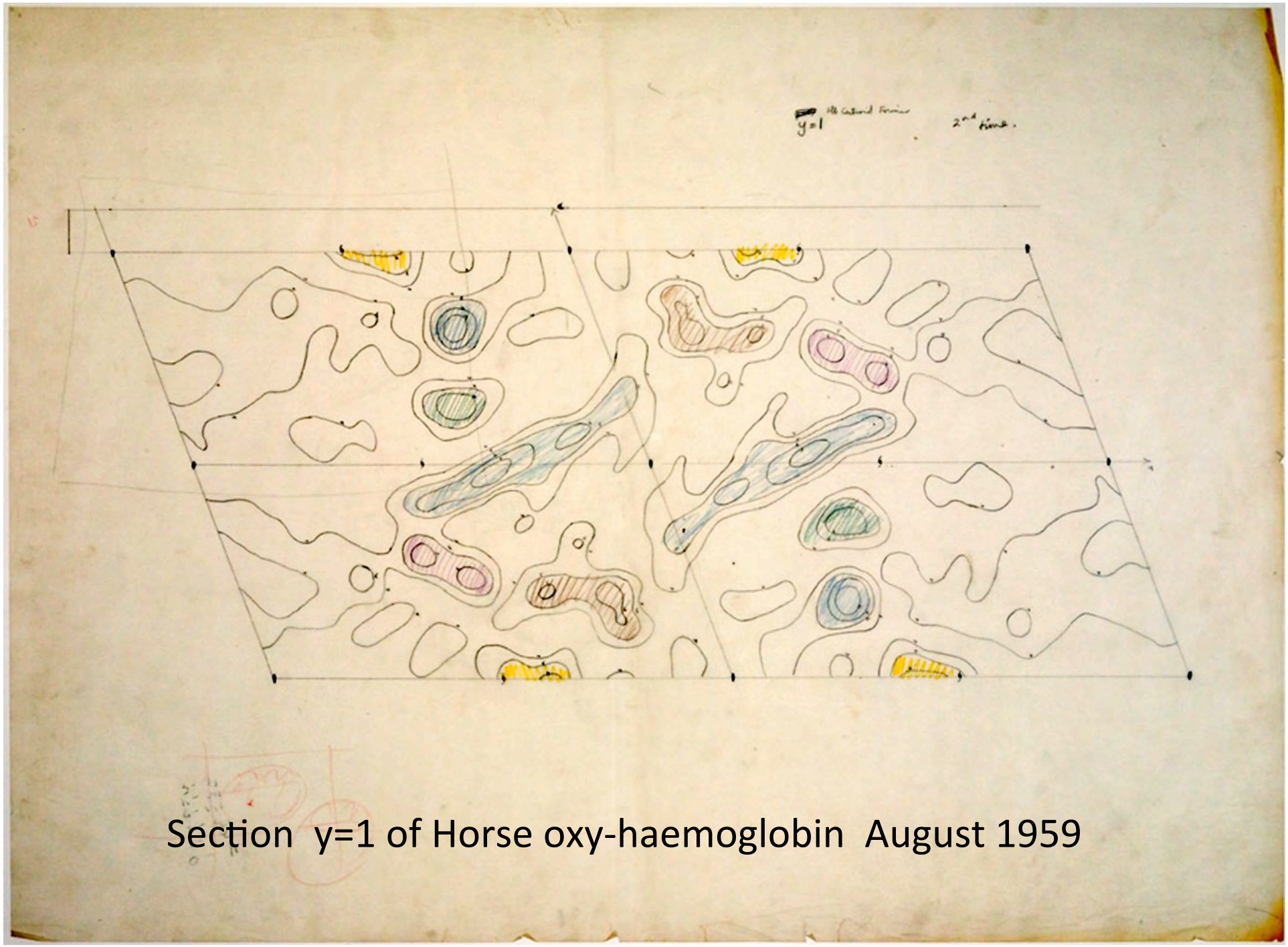
mr@purdue.edu

# PDB40

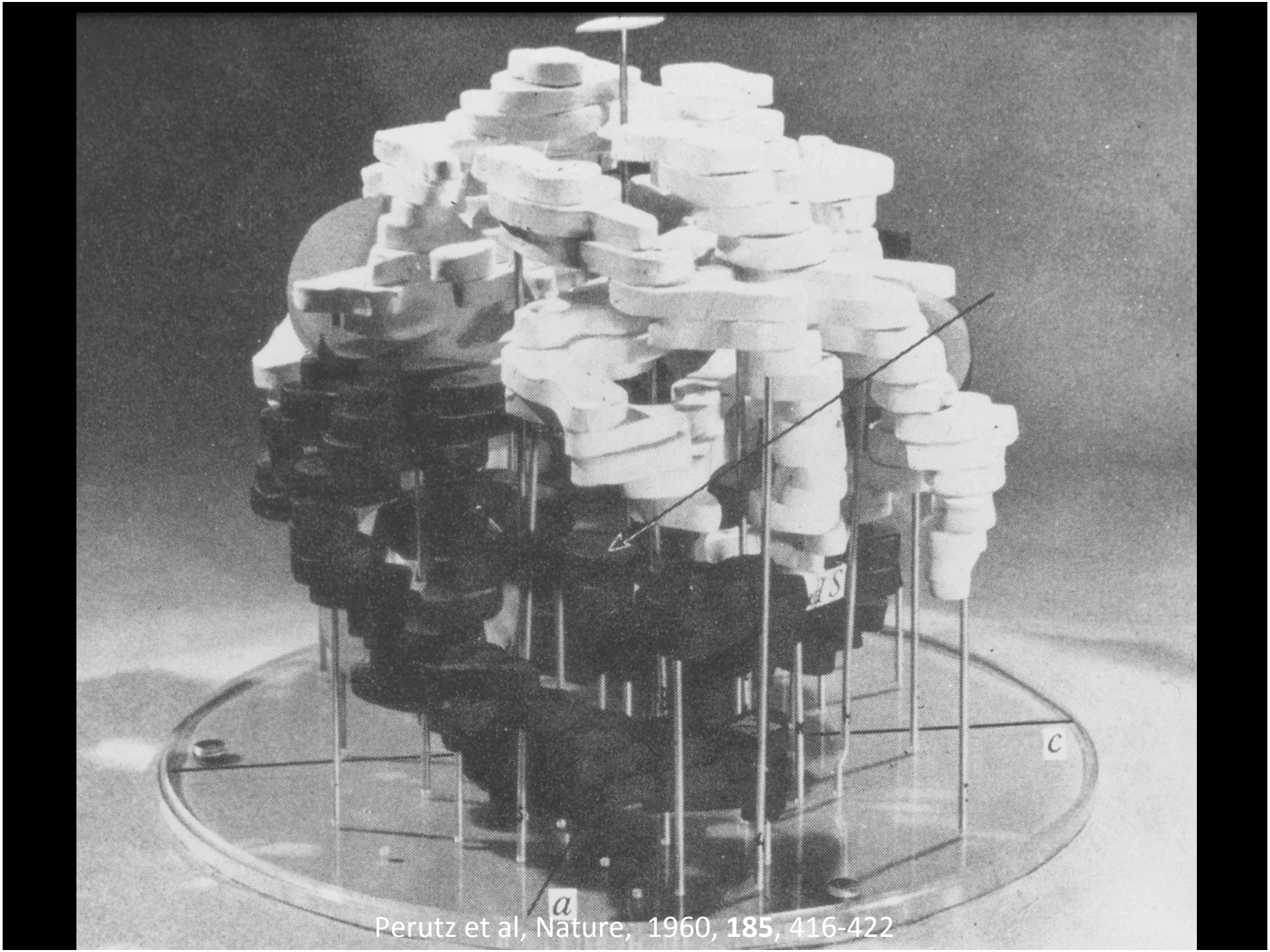
## A historical perspective

# The 6Å resolution model of sperm whale myoglobin. 1957



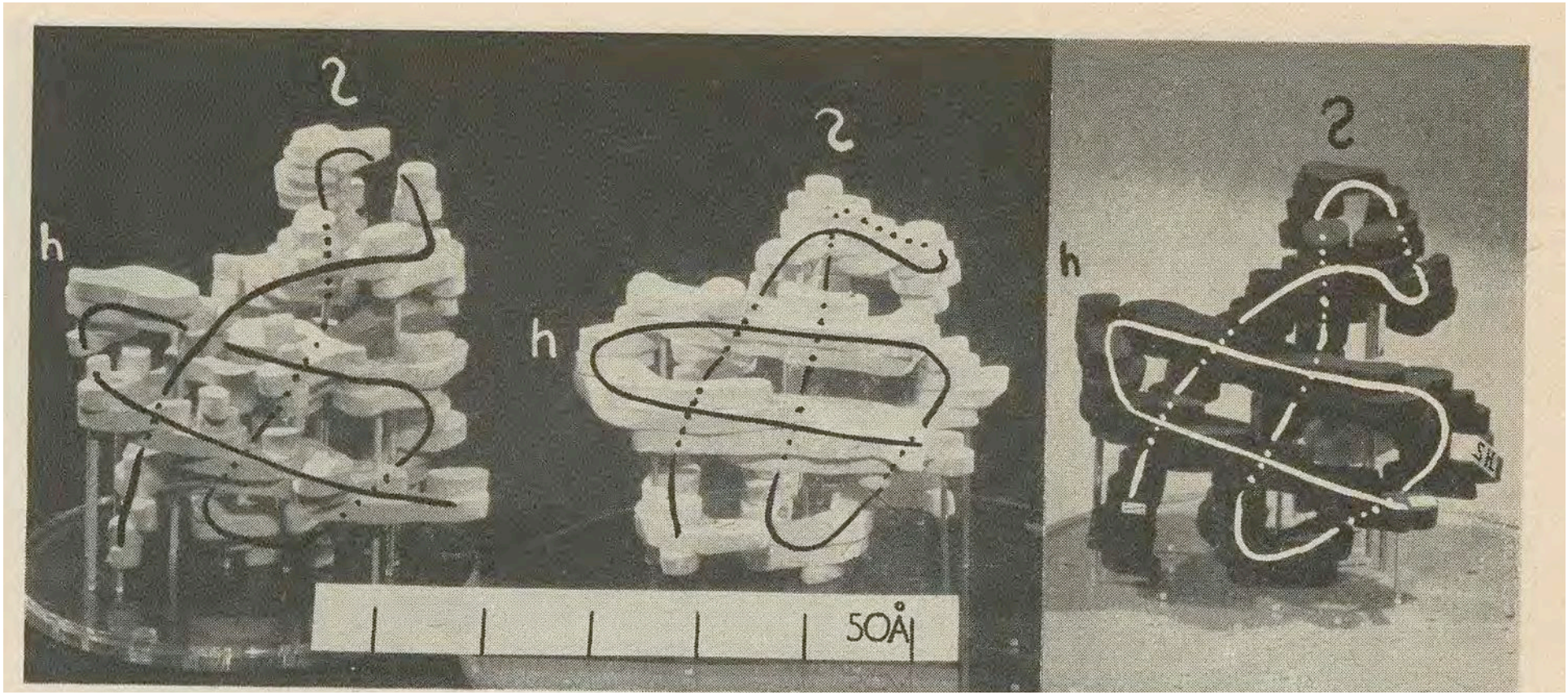



Section y=1 of Horse oxy-haemoglobin August 1959



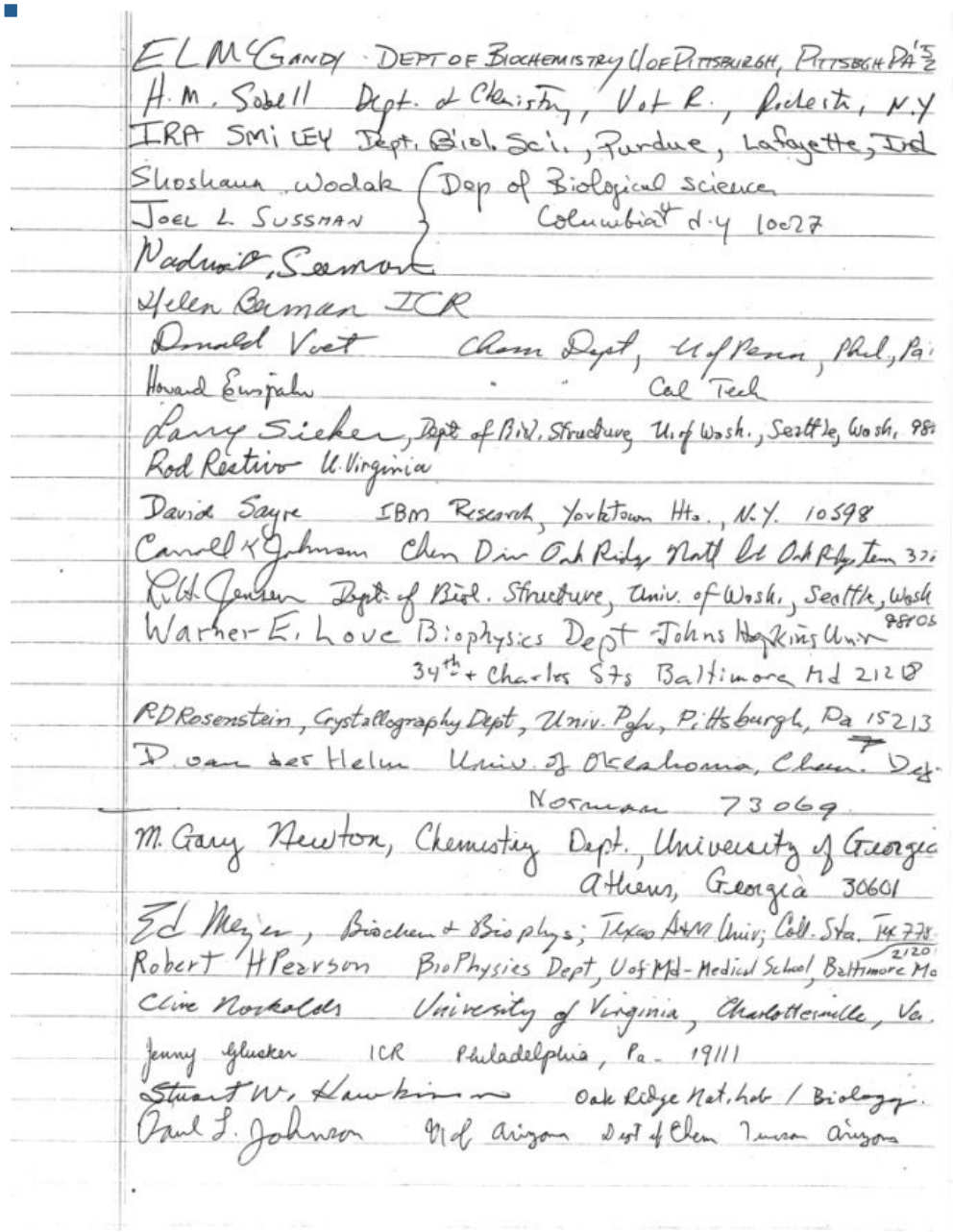

Perutz et al, Nature, 1960, **185**, 416-422

**Myoglobin** **Haemoglobin alpha** **Haemoglobin beta**




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

**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**



ELMcGandy Dept of Biochemistry U of Pittsburgh, Pittsbgh Pa 15213
H.M. Sobell Dept. of Chemistry, Vet R., Rochester, N.Y
IRA Smiley Dept. Biol. Sci., Purdue, Lafayette, Ind
Shoshana Wodak } Dep of Biological Science
Joel L. Sussman } Columbia N.Y 10027
Nadrian Seeman
Helen Berman ICR
Donald Voet Chem Dept, U of Penn, Phil, Pa
Howard Einspahr " " Cal Tech
Larry Sieker, Dept of Biol. Structure, U. of Wash., Seattle, Wash, 98
Rod Restivo U. Virginia
David Sayre IBM Research, Yorktown Hts., N.Y. 10598
Carroll K Johnson Chem Div Oak Ridge Natl Lab Oak Ridge, Tenn 37
R.H. Jensen Dept. of Biol. Structure, Univ. of Wash., Seattle, Wash 98105
Warner E. Love Biophysics Dept Johns Hopkins Univ
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. Tex 778
Robert H Pearson BioPhysics Dept, U of Md - Medical School, Baltimore Md 2120
Clive Nockolds University of Virginia, Charlottesville, Va.
Jenny Glucker ICR Philadelphia, Pa. 19111
Stuart W. Hawkinson Oak Ridge Nat. Lab / Biology
Paul L. Johnson U of Arizona Dept of Chem Tucson Arizona

# 1971 Cold Spring Harbor Conference

## Informal meeting called by Max Perutz

Participants:

Jan Drenth
Walter Hamilton
Bill Lipscomb
Max Perutz
David Phillips
Fred Richards
Michael Rossmann

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 results of structural investigations in a scientific journal.

**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

# Frances Bernstein

- 1974 began as programmer for CRYSNET
- Worked on annotation of more than 8000 PDB entries
- 1998 retired from BNL prior to RCSB PDB award

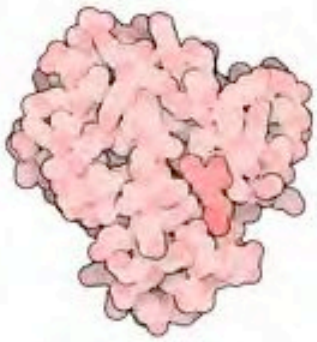
The CRYSNET consortium was funded by the NSF. Its goal was to allow distributed computing. In this case, Brookhaven provided high speed computing for calculating Fouriers and doing refinements. Remote sites could pick up the results via the (very young) internet.

# 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**

# Protein Data Bank in 1973



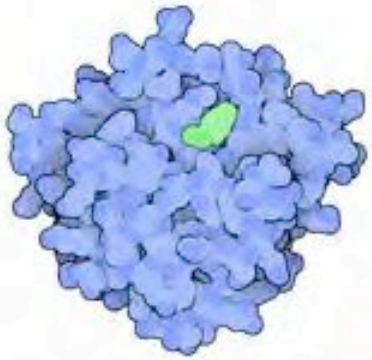
myoglobin

lamprey
hemoglobin

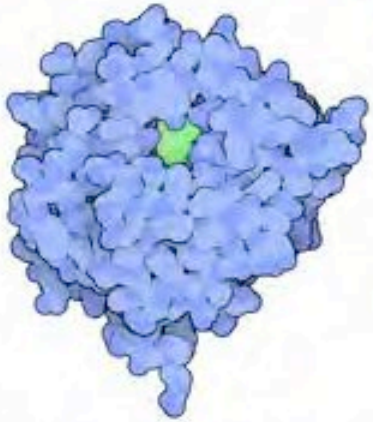
rubredoxin

cytochrome b5

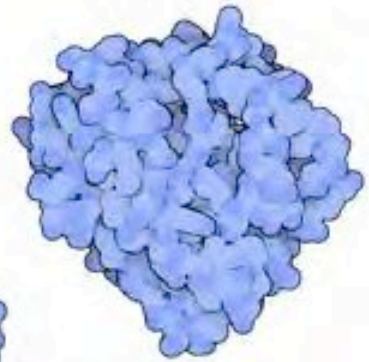
trypsin inhibitor



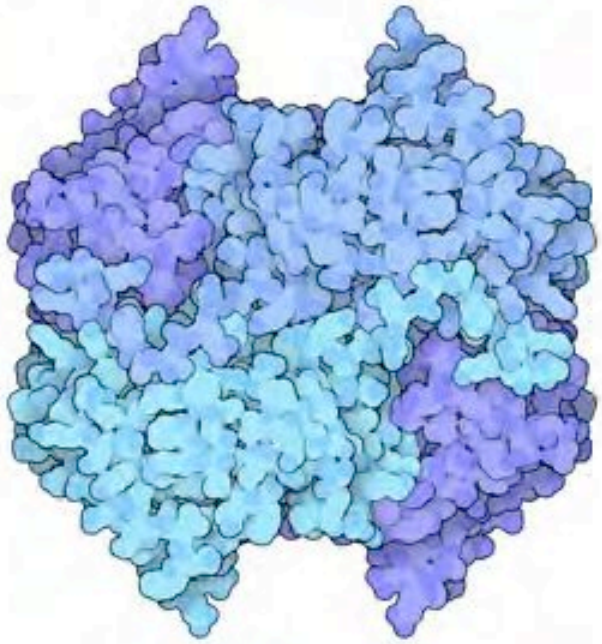
chymotrypsin



carboxypeptidase



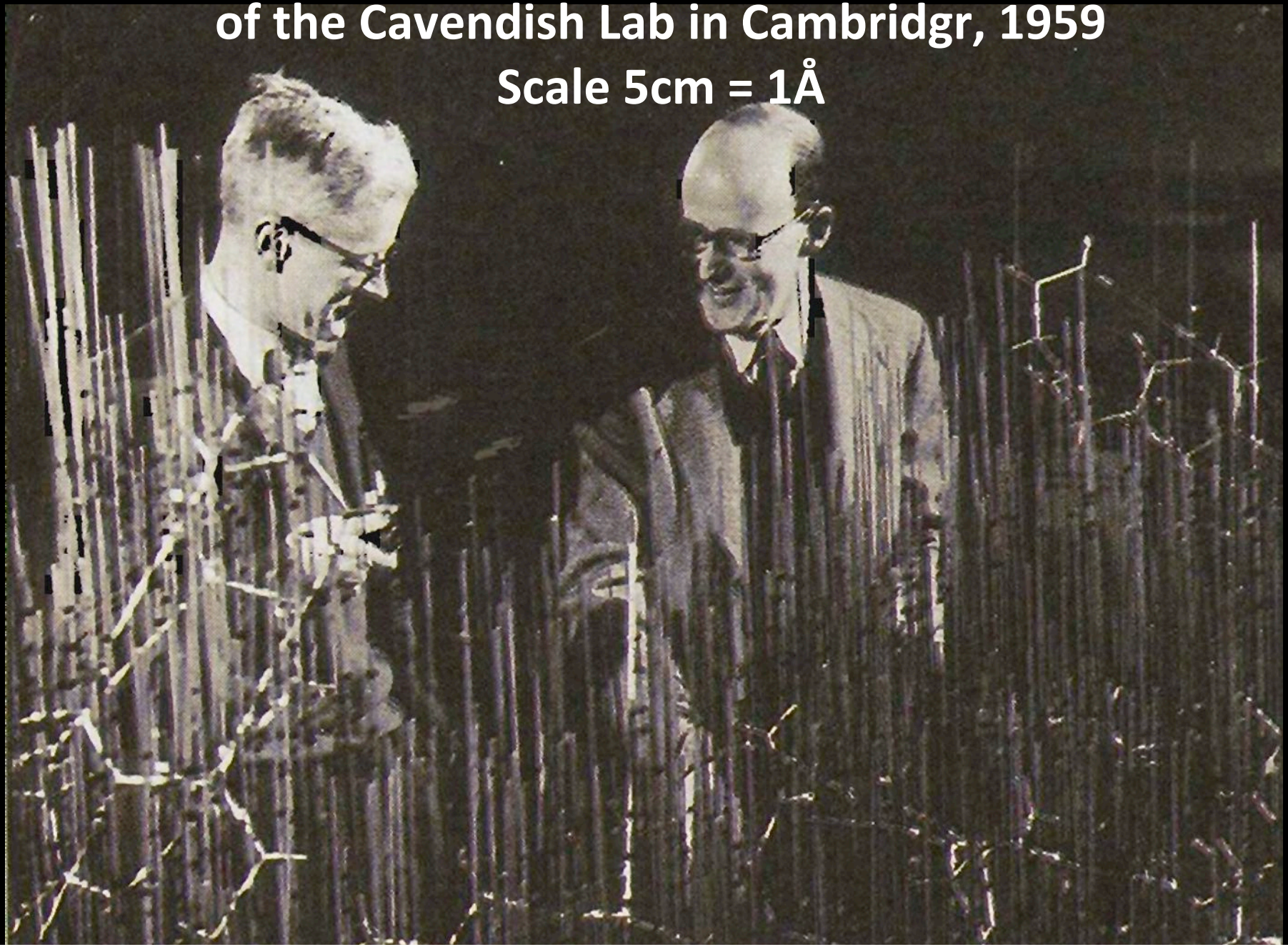
subtilisin

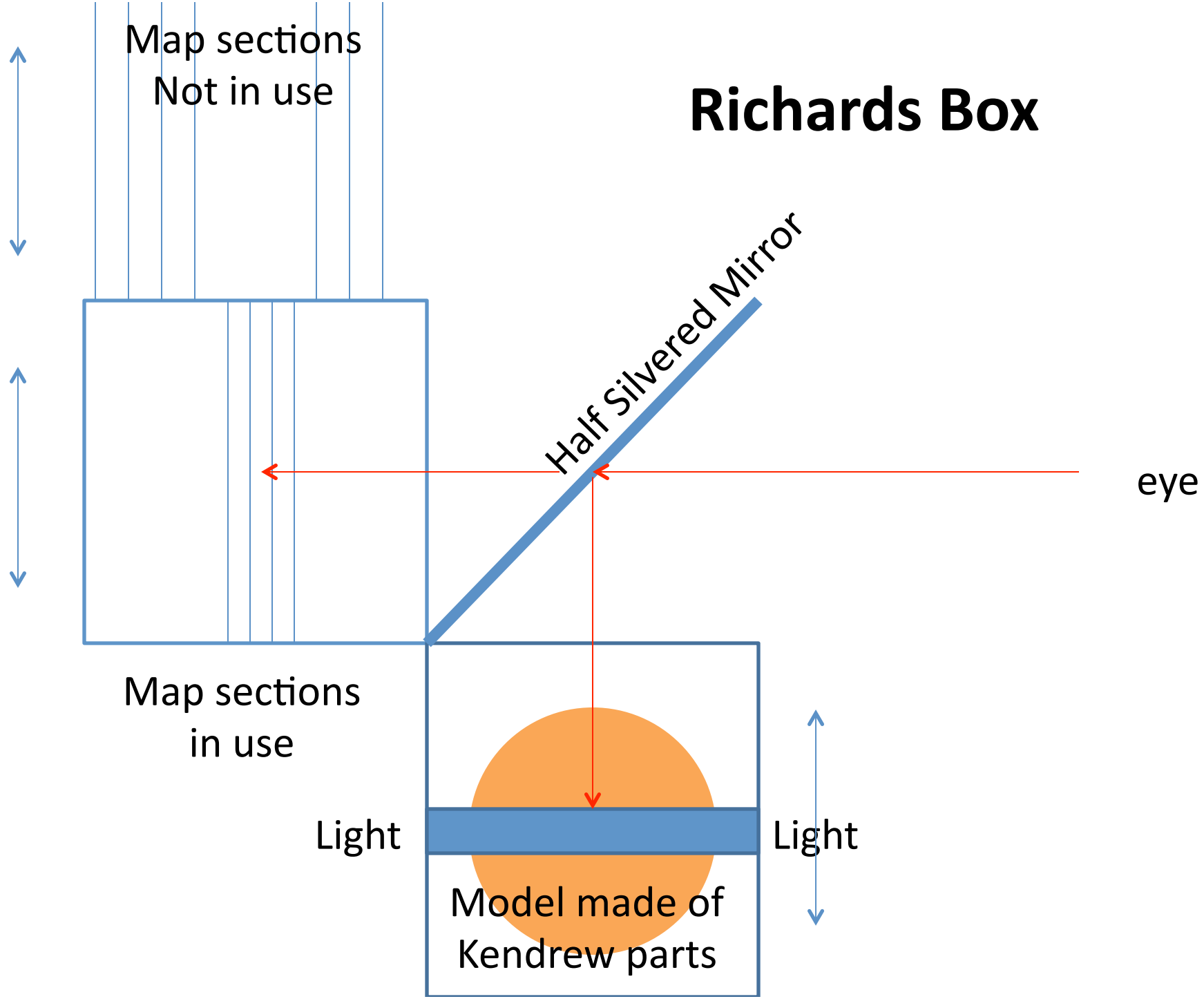


lactate dehydrogenase

# Building a model of myglobin in the old cycletron room of the Cavendish Lab in Cambridgr, 1959

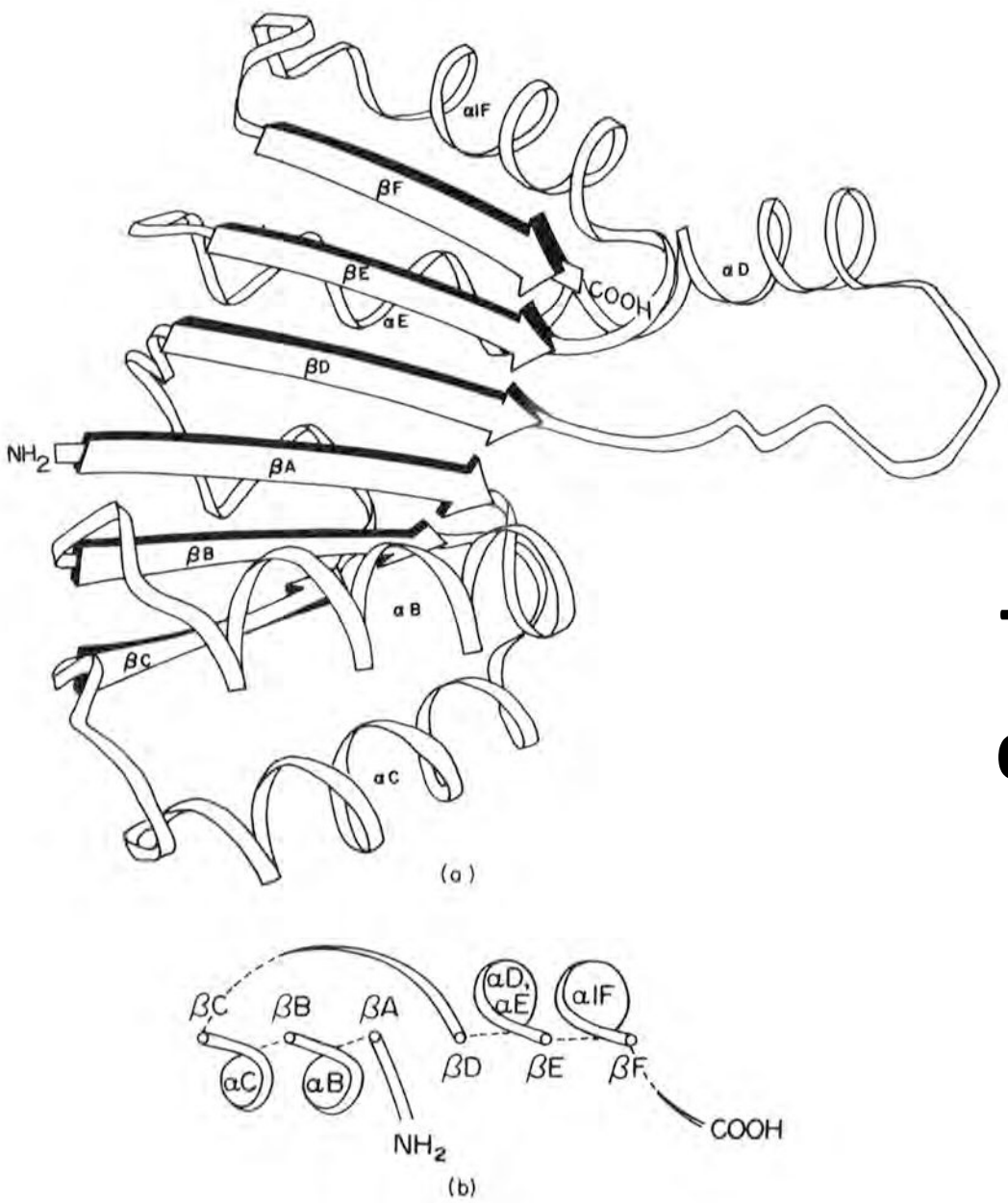
## Scale 5cm = 1Å




Map sections
Not in use
Richards Box
Half Silvered Mirror
eye
Map sections
in use
Light
Light
Model made of
Kendrew parts

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

Scale: 2cm = 1 Å




# The NAD binding domain of lactate dehydrogenase 1970

Number 1

September '74

# Protein Data Bank

## Newsletter

We thought it would be a good idea to mail out a report on the status of the Protein Data Bank and to establish at this time a regular newsletter. Some of the material in this first edition of the newsletter may be familiar to you. We promise that the next edition will be much smaller!

## Deposition of Coordinates

Data may be deposited by filling out the form in Appendix 1. Tape or cards rather than a listing are appreciated. Mail these to

T.F. Koetzle
Department of Chemistry
Brookhaven National Laboratory
Upton, New York 11973
Telephone: 516-345-4384

## Coordinate Directory

The coordinate sets in final distributable form are listed below along with coordinate sets soon to be available (marked *):

carboxypeptidase A
carp muscle calcium binding parvalbumin
α-chymotrypsin
cytochrome $b_5$
flavodoxin *
D-glyceraldehyde-3-phosphate dehydrogenase *
horse hemoglobin (deoxy and met)
lactate dehydrogenase
lamprey hemoglobin
lysozyme *
myoglobin
pancreatic trypsin inhibitor
papain
rubredoxin
staphylococcal nuclease
subtilisin
thermolysin*

## Format

The format of the coordinates is given in Appendix 2. Torsion angles, structure factors and phases are also available for some proteins, as indicated in Appendix 3.

# Early PDB deposition procedures

November 30, 1981

Dr. Georg Schulz
Abteilung Biophysik
Max Planck Institut fuer
Medizinische Forschung
Jahnstrasse 29
6900 Heidelberg
West Germany

Dear Dr. Schulz:

Thank you for depositing the full coordinate set for glutathione reductase with the Protein Data Bank. We have extracted the data from the tape with no difficulty and will be processing it as soon as possible.

In the last three months we have received twenty coordinate sets (we usually get about thirty per year) and so we have fallen behind in our processing efforts. We are delighted to see this growth in the field and in our data base and hope that we will be able to get up-to-date soon.

With best wishes,

Sincerely,

Frances C. Bernstein

FCB:sk

# Early PDB deposition procedures

November 30, 1981

Dr. Georg Schulz
Abteilung Biophysik
Max Planck Institut fuer
Medizinische Forschung
Jahnstrasse 29
6900 Heidelberg
West Germany

Dear Dr. Schulz:

Thank you for depositing the full coordinate set for glutathione reductase with the Protein Data Bank. We have extracted the data from the tape with no difficulty and will be processing it as soon as possible.

In the last three months we have received twenty coordinate sets (we usually get about thirty per year) and so we have fallen behind in our processing efforts. We are delighted to see this growth in the field and in our data base and hope that we will be able to get up-to-date soon.

With best wishes,

Sincerely,

Frances C. Bernstein

FCB:sk

*Acta Cryst.* (1980). B36, 819–823

# Three-Dimensional Coordinates from Stereodiagrams of Molecular Structures

BY MICHAEL G. ROSSMANN AND PATRICK ARGOS

*Department of Biological Sciences, Purdue University, West Lafayette, Indiana* 47907, *USA*

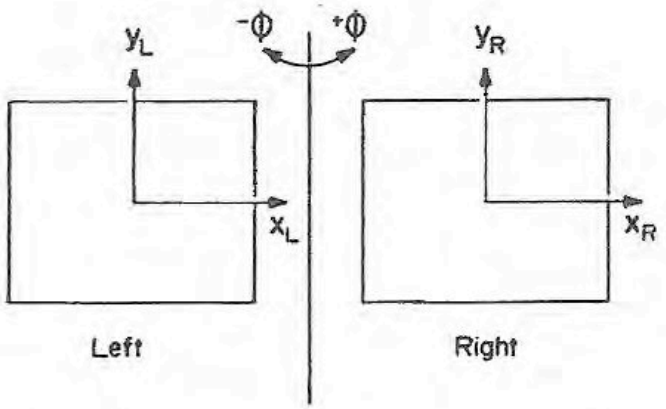



Fig. 1. Definition of coordinates in viewing a stereodiagram. The $z$ axis is perpendicular to the page.

$$\left.\begin{aligned} x &= \frac{x_L + x_R}{2\cos\varphi} \times q, \\ y &= \frac{y_L + y_R}{2} \times q, \\ z &= \frac{x_L - x_R}{2\sin\varphi} \times q, \end{aligned}\right\}$$

where

$$q = \frac{1}{1 + \left(\dfrac{x_L - x_R}{2v\sin\varphi}\right)}.$$

## The ethics question

The tradition of science is to gather and publish facts. Others may wish either to verify the facts by repeating the observations or to use these results to obtain a fundamental understanding of Nature in terms of a unifying concept or correlation. The accepted practice is to extract information from the literature, acknowledge its source, and to build upon it. The trend to withhold coordinates appears to be at odds with this long-standing tradition of scientific endeavor and exchange. Furthermore, coordinates are sometimes given only to close associates thus stifling a healthy public debate. Nevertheless, the present authors foresee that the technique published here may be considered a 'sharp' practice by some, although it is only extracting information from publications. This is evidenced by resistance to suggestions that coordinates be deposited with the Brookhaven Data Bank upon publication of high-resolution structures (*cf. Instructions to Authors* of the *Journal of Biological Chemistry*, 1979; *Crystallography of Molecular Biology*, 1976).

## Publication and deposition of structural data

1976. Erice meeting recommendation by Michael Rossmann in book of abstracts: “A prerequisite for publication should be the deposition of coordinates at the Brookhaven Protein Data Bank”

1976(?) Martha Ludwig and Michael Rossmann, assistant editors of JBC, create compulsory deposition policy for JBC

1980. Extraction of co-ordinates from stereo diagrams.

1988. Janet Smith, chair of Gordon Conference, organizes public discussion

1988. More and more journals requiring deposition of coordinates

1990. NIH requires coordinate deposition for funding.

2008. PDB requires deposition of structure amplitudes with coordinates

## Diversification of PDB archive

1989. First NMR data

1991. First EM data

Two thirds of all structures deposited with the PDB are based on Molecular Replacement in recent years, some based on *ab-initio* predicted structures.

Deposition of structural predictions?

How much experimental data should be required for inclusion in the PDB?

Problems of validation

## PDB archive distribution

1971 Magnetic tapes, floppy discs **$100**

1992 CD-ROM discs **$100**

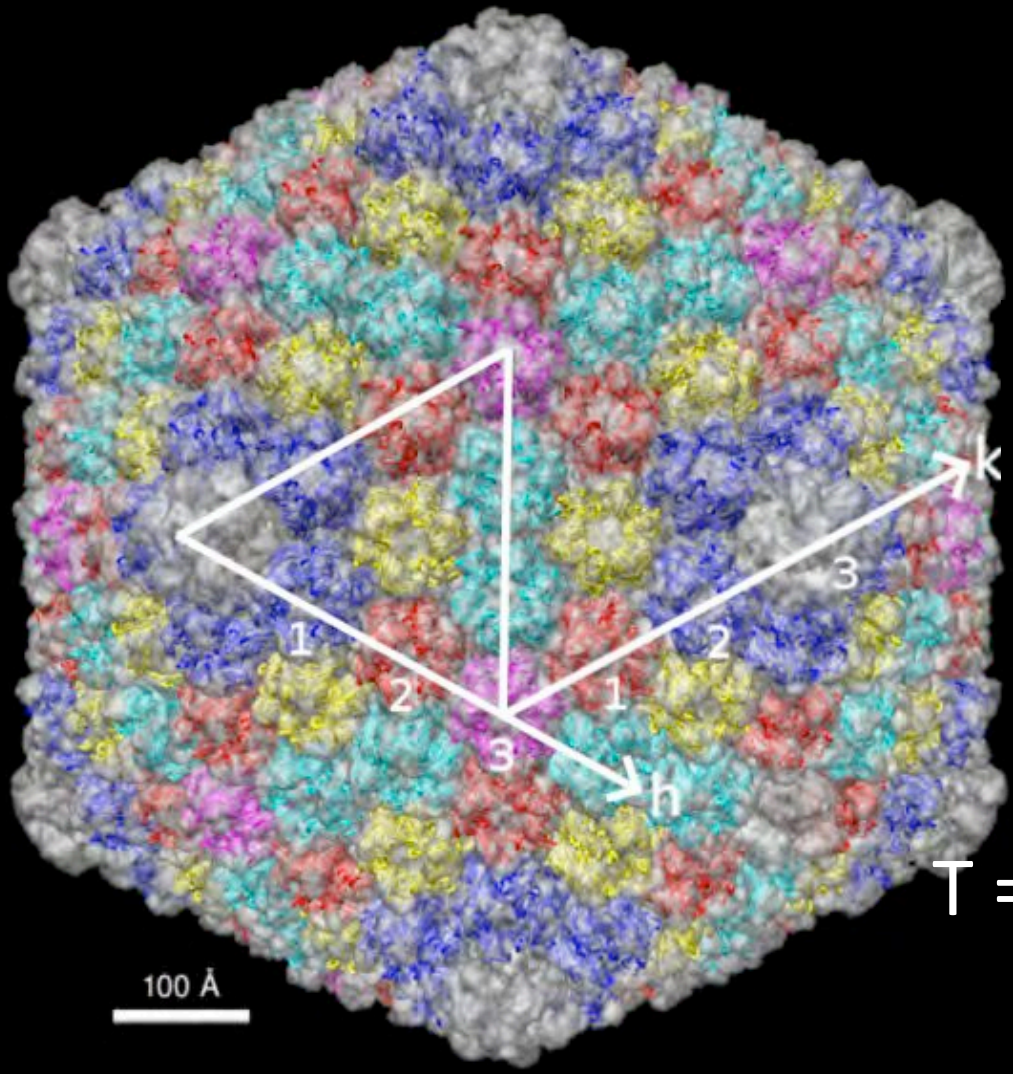
1994 PDB Browser could read records in the header sections

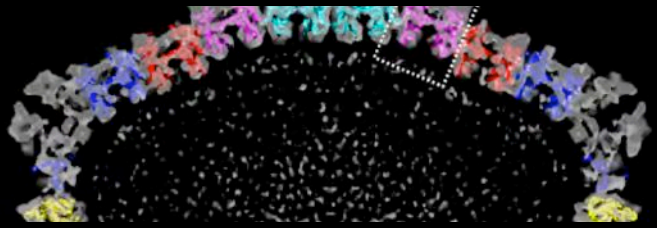
1998 Distributions became **free** under the Research Collaboratory for Structural Biology (RCSB) PDB by giving yearly time-stamped snapshots of the archive, available at ftp://snapshots.wwpdb.org/ .

2005 Last physical distribution
(In 2005 it took 8 DVDs for one distribution)

# **Sputnik** Virus, a “virophage” .

## It uses Mimivirus as it’s host



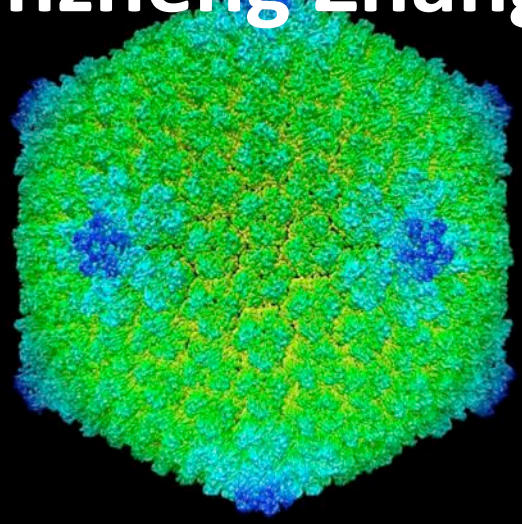
T = 27;  Diameter = 747 Å

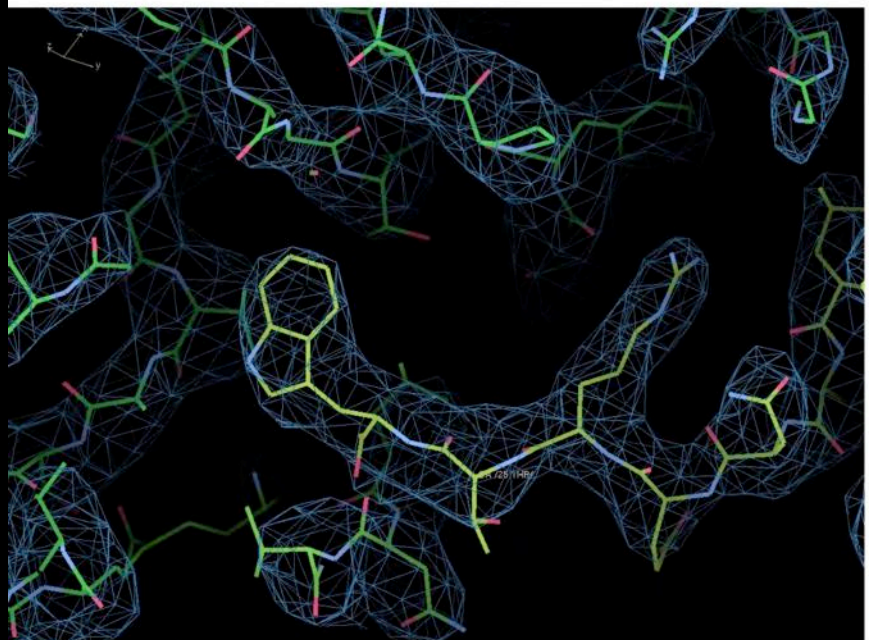
Sun et al 2010

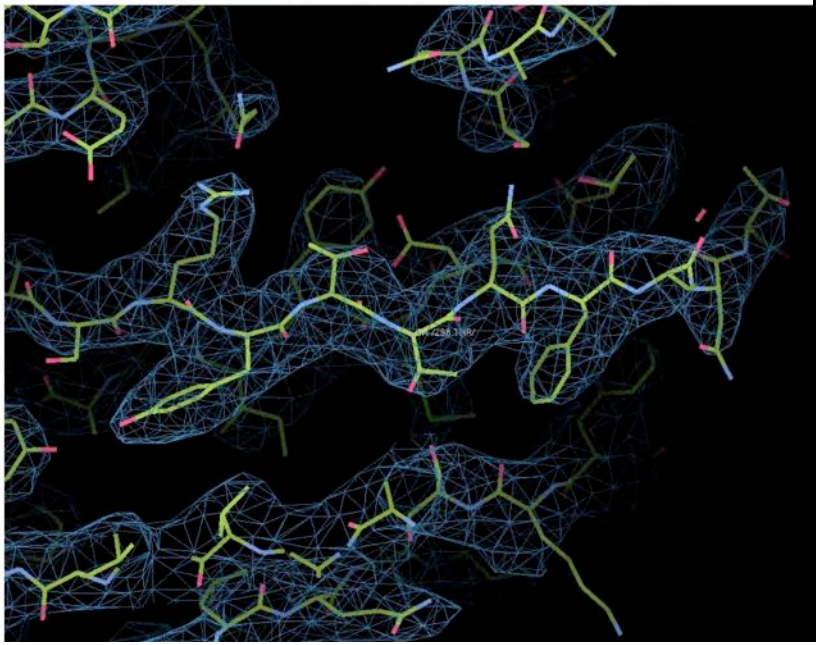
*J. Virol.* **84**:894-897

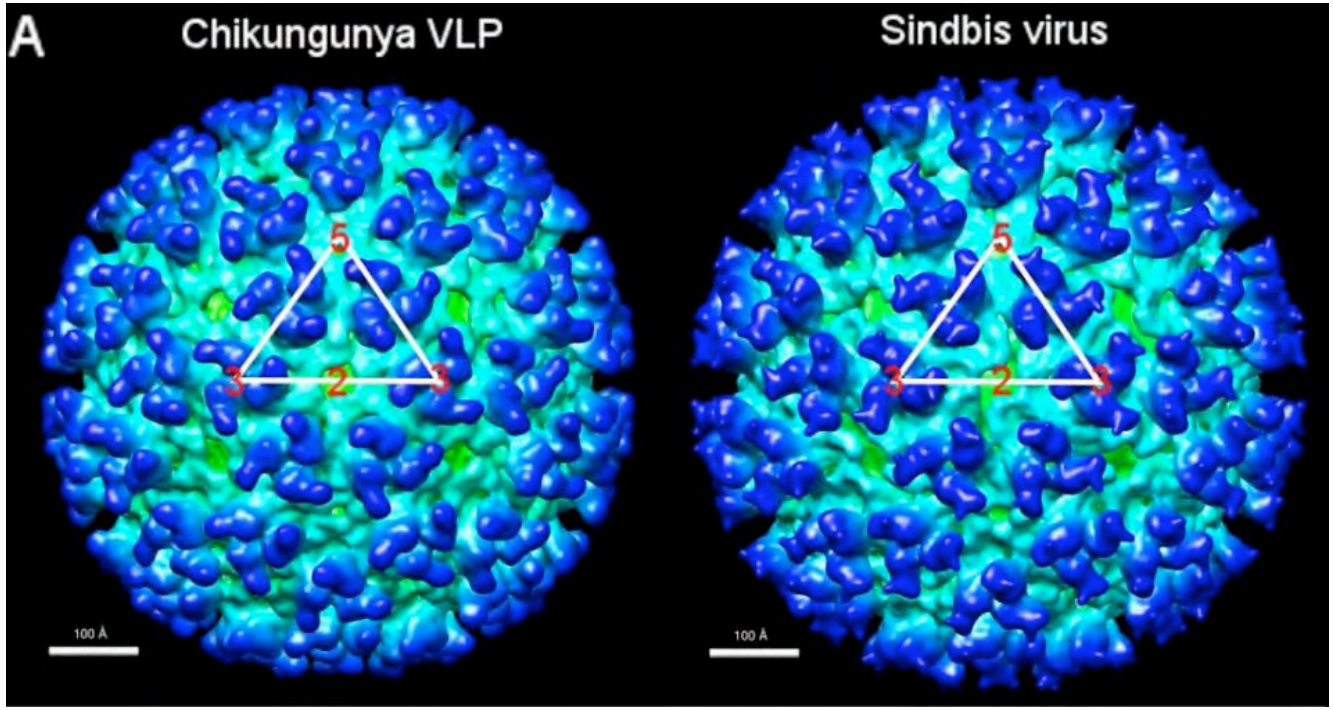
# Sputnik virus at 3.5 Å resolution by cryo EM

**Xinzheng Zhang, Siyang Sun, Ye Xiang**



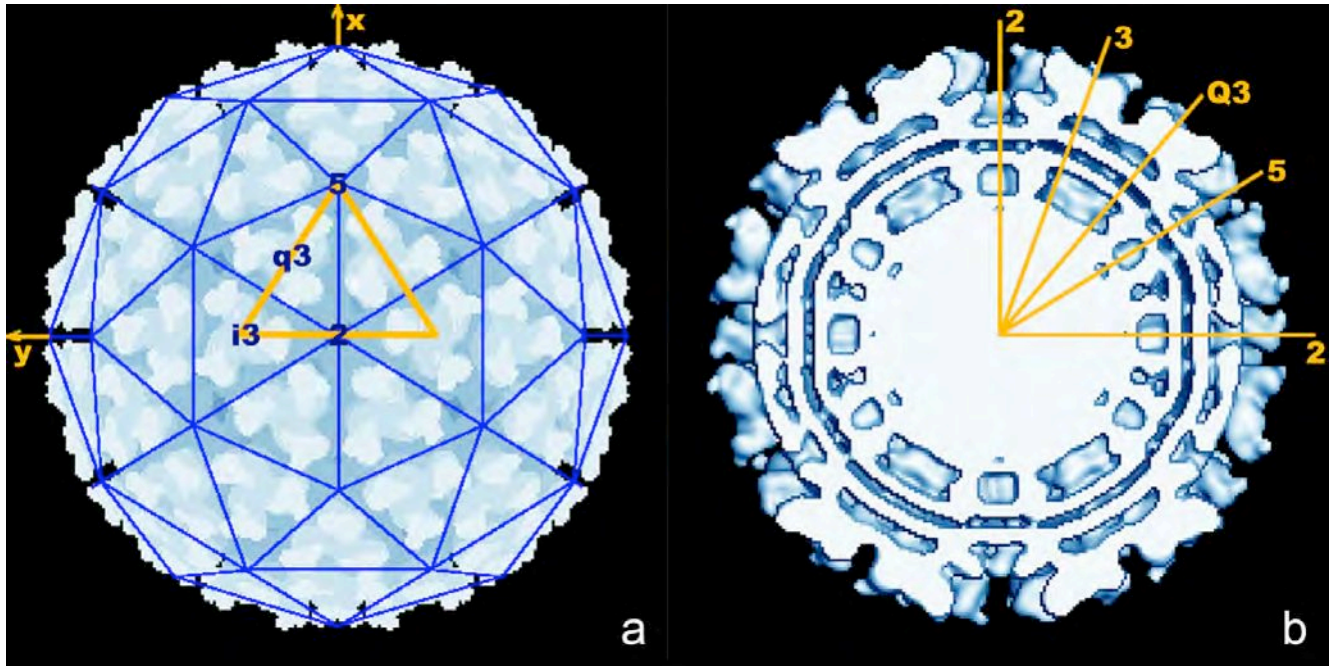







# Alpha viruses

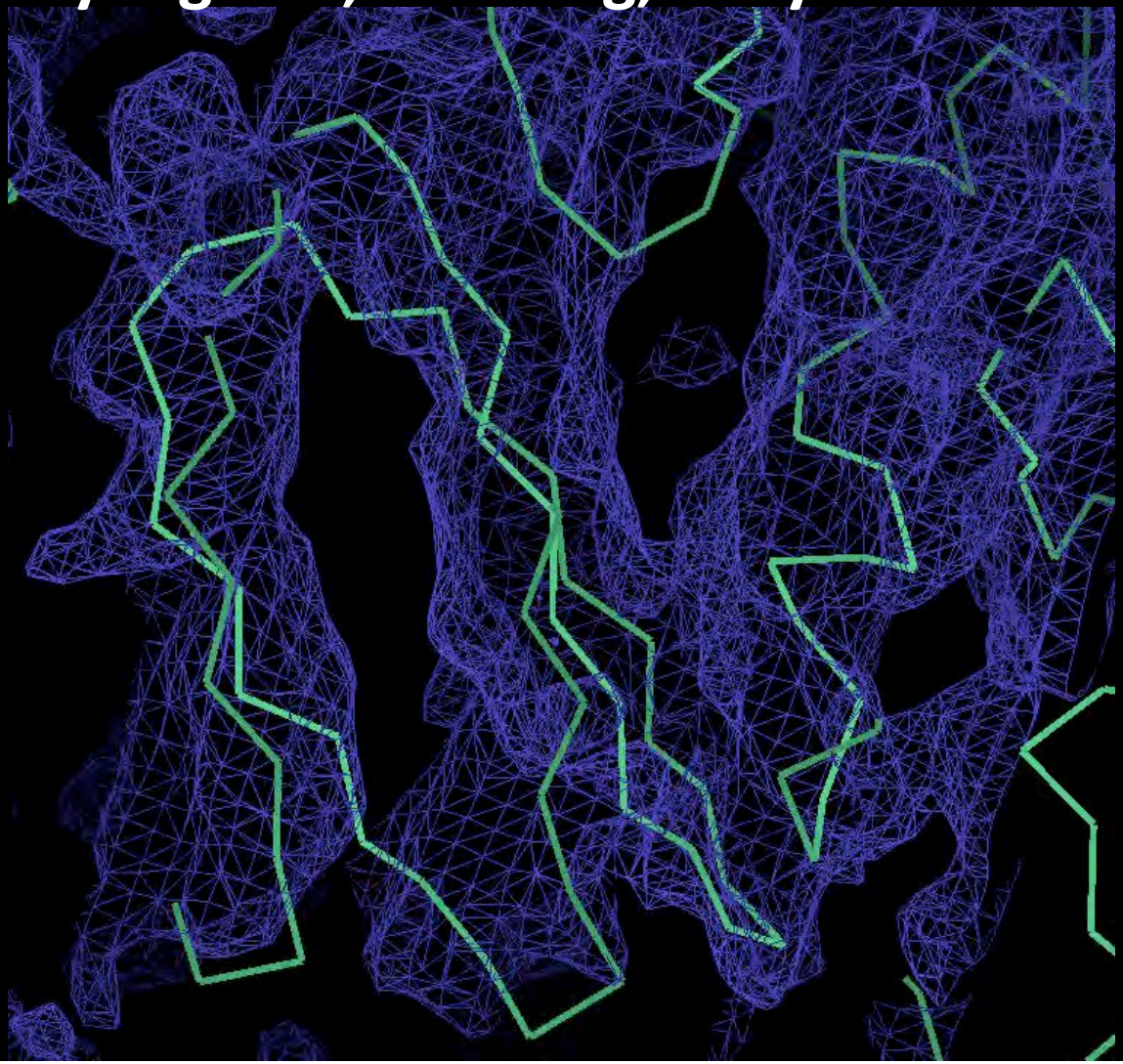
The Chikungunya virus like particles are closely similar to Sindbis virus




Sindbis virus surface view (left) and cross section (right)
A lipid membrane separates the inner nuclear capsid from the outer glycoproteins

The 4.2 Å resolution CHIKV VLP cryoEM map (dark blue)
View of the Cα backbone of the EM fit determined
Domain C β-barrel (baby blue)

**Siyang Sun, Ye Xiang, Gary Nabel**



# The T4 structure

**Purdue University**

Anthoni Batisti

Paul Chipman

Andrei Fokine

Victor Kostyuchenko

Pertr Leiman

Siyang Sun

**Tokyo Institute of Technology**

Fumio Arisaka

Shuji Kanamaru

**University of Maryland**

Lindsay Black

**Moscow**

Vadim Mesyanzhinov

Mikhail Schneider

**Catholic University of America**

Venigella Rao

**NIH Bethesda**

Bijan Ahvazi

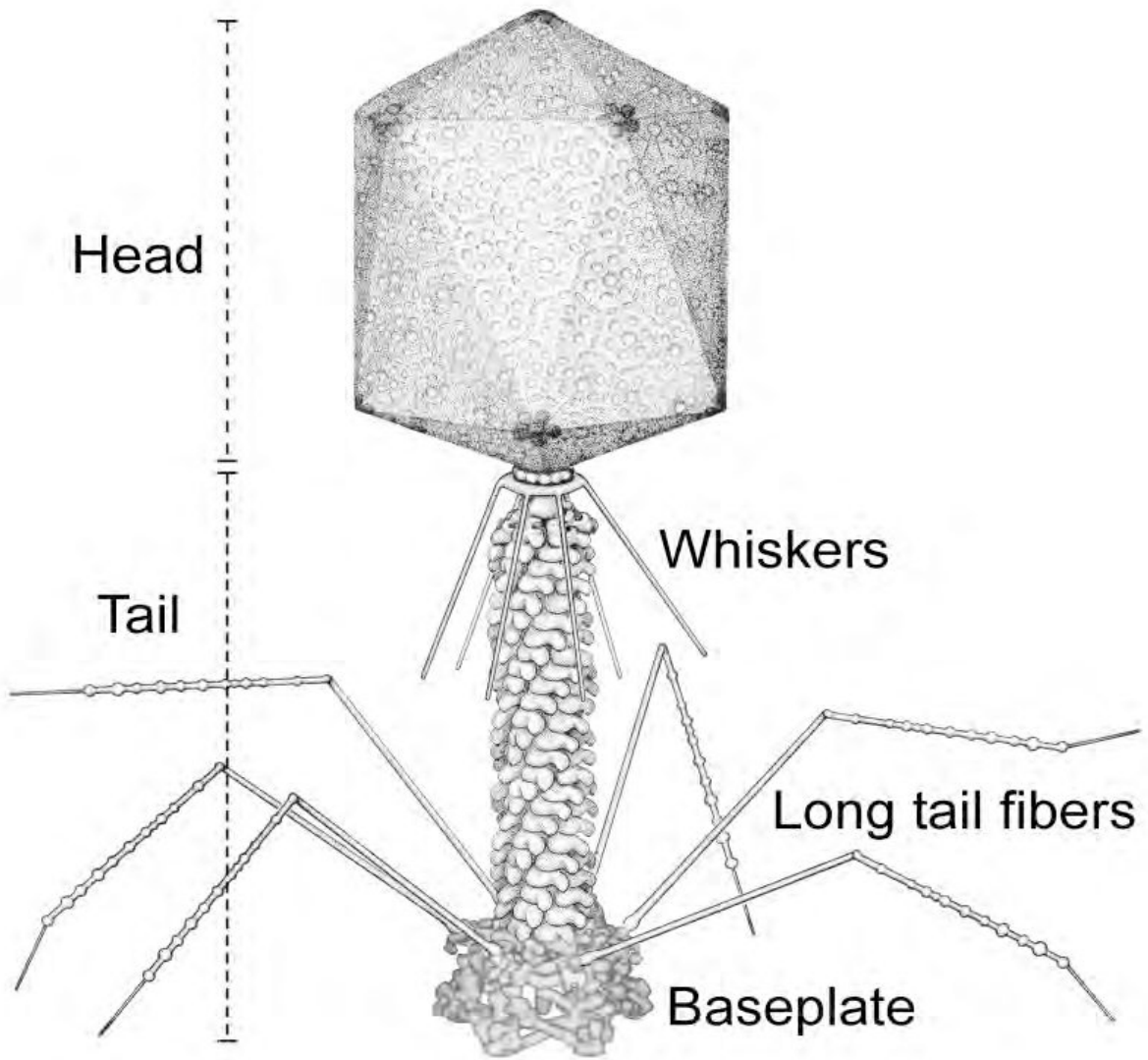
Karen Boeshans

Alasdair Steven

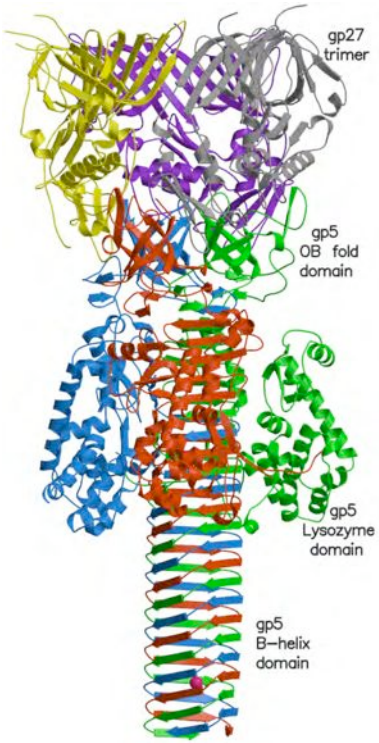
**Santiago**

Mark van Raaij

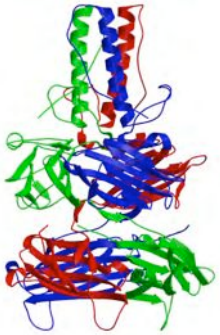
# Too complex to crystallize: Bacteriophage T4



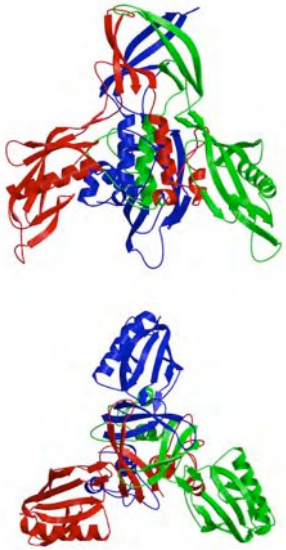
# Crystal structures of the baseplate proteins



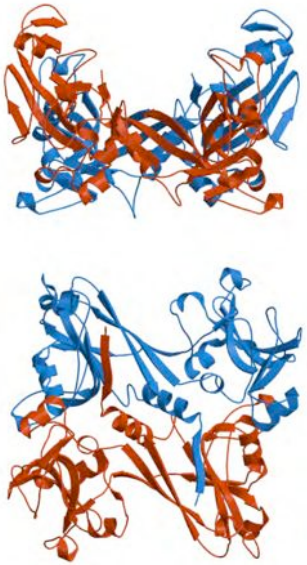

gp5-gp27



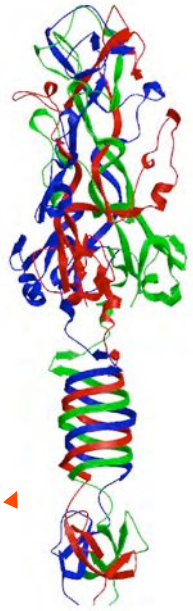
gp9,
LTF
attachment



gp11,
STF attachment

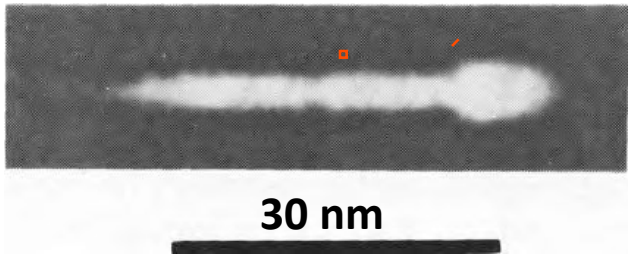


gp8

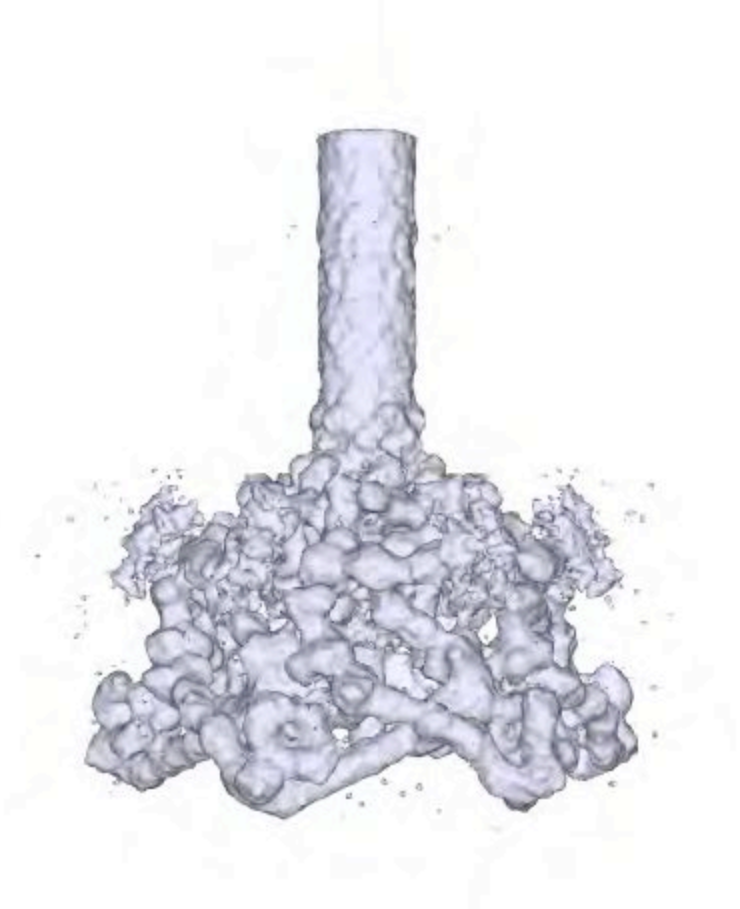


gp12
C-term

gp12, STF

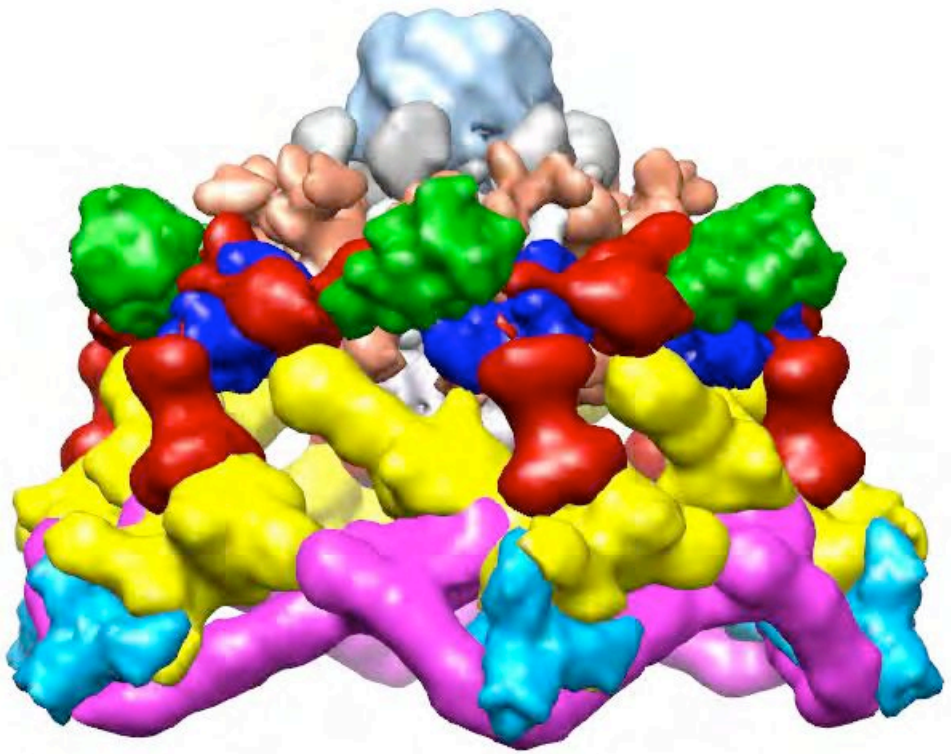


# Hexagonal conformation (tube-baseplates)

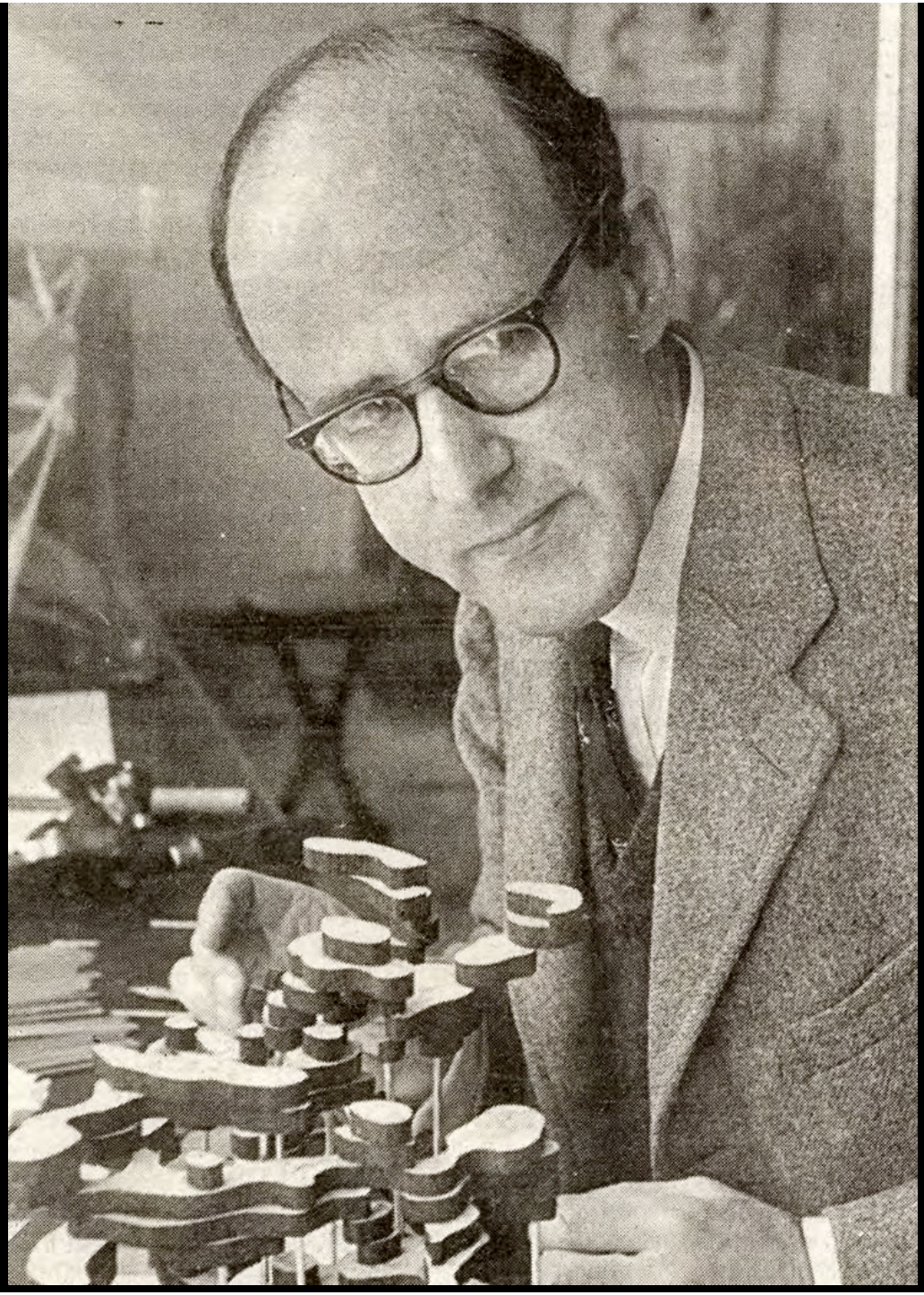


- Initial model – hexagonal prism connected to a tube
- Sixfold symmetry
- 945 particles used in the reconstruction
- Defoci 1.5 – 3.5 μm
- 12 Å resolution

# Hexagonal conformation (tube-baseplates)



- Initial model – hexagonal prism connected to a tube
- Sixfold symmetry
- 945 particles used in the reconstruction
- Defoci 1.5 – 3.5 μm
- 12 Å resolution

**Helen**
**2011**

# Max circa 1959