Growth, Globalization, and Future of the Protein Data Bank

Stephen K. Burley, Eli Lilly & Co.

October 28th 2011

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Acknowledgements

- Helen Berman - RCSB PDB
- Gerard Kleywegt - PDBe
- John Markley - BMRB
- Haruki Nakamura - PDBj

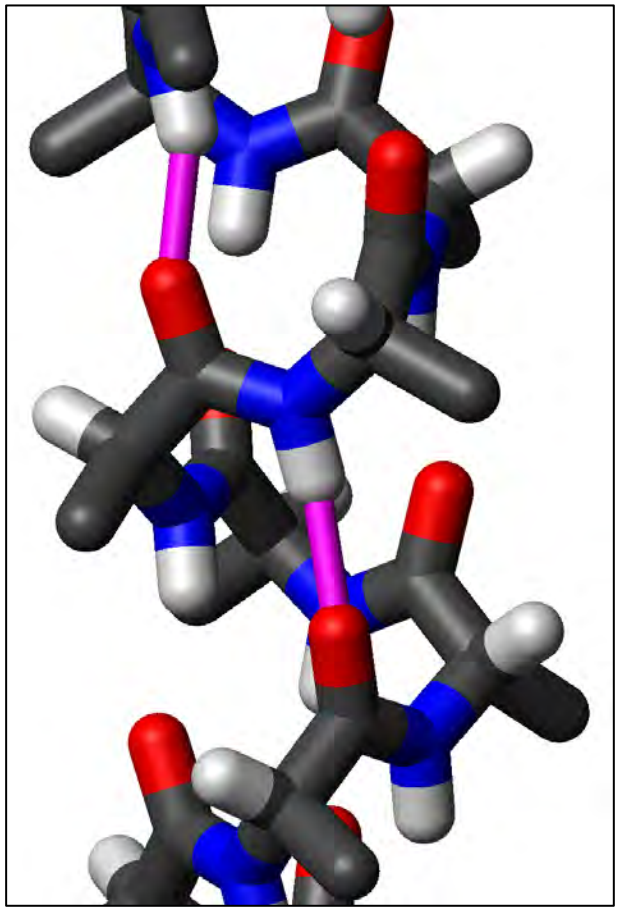
- Phil Bourne - RCSB PDB
- Martha Quesada - RCSB PDB
- Christine Zardecki - RCSB PDB

"We may anticipate that the chemist of the future who is interested in biomolecules will come to rely upon a new structural chemistry, and that great progress will be made, through this technique, in biology and medicine."

Linus Pauling, Nobel Lecture 1954

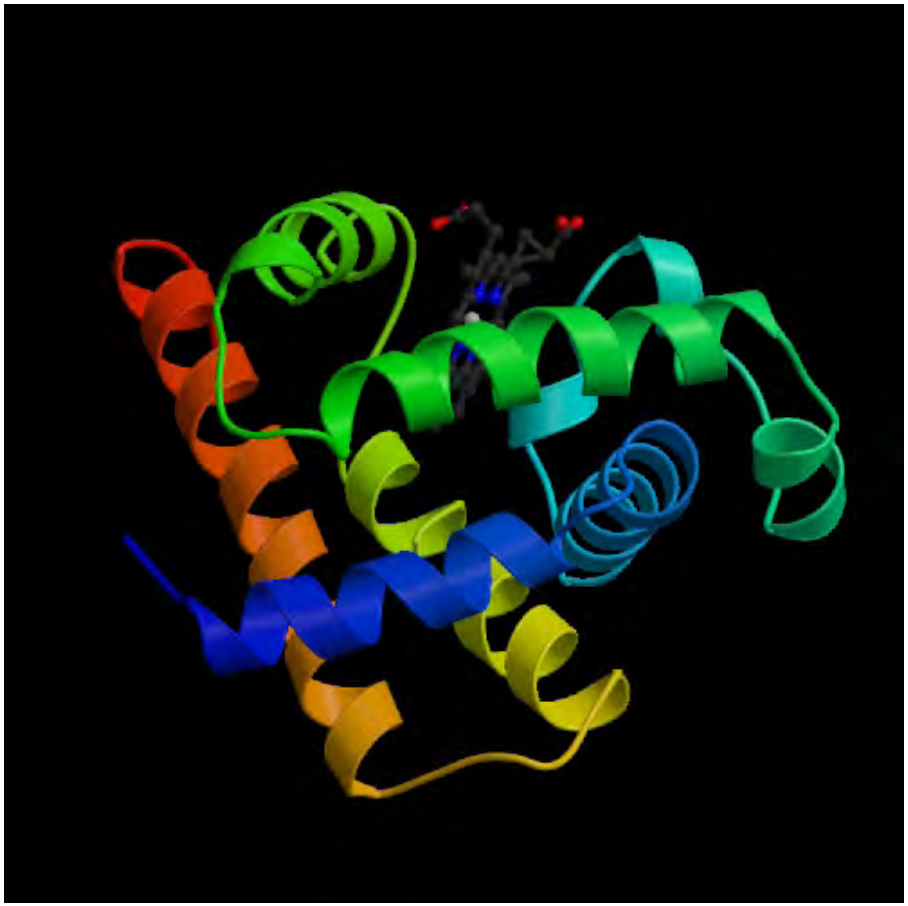
Chemistry→Biological Structure

α-Helix



Pauling and Corey (1951)

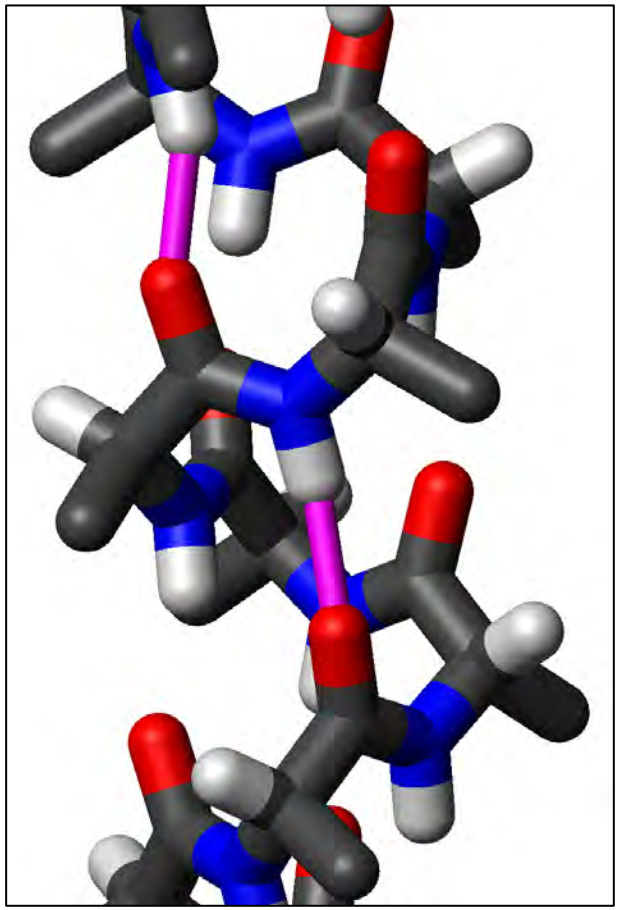
Myoglobin



Kendrew *et al*. (1958)

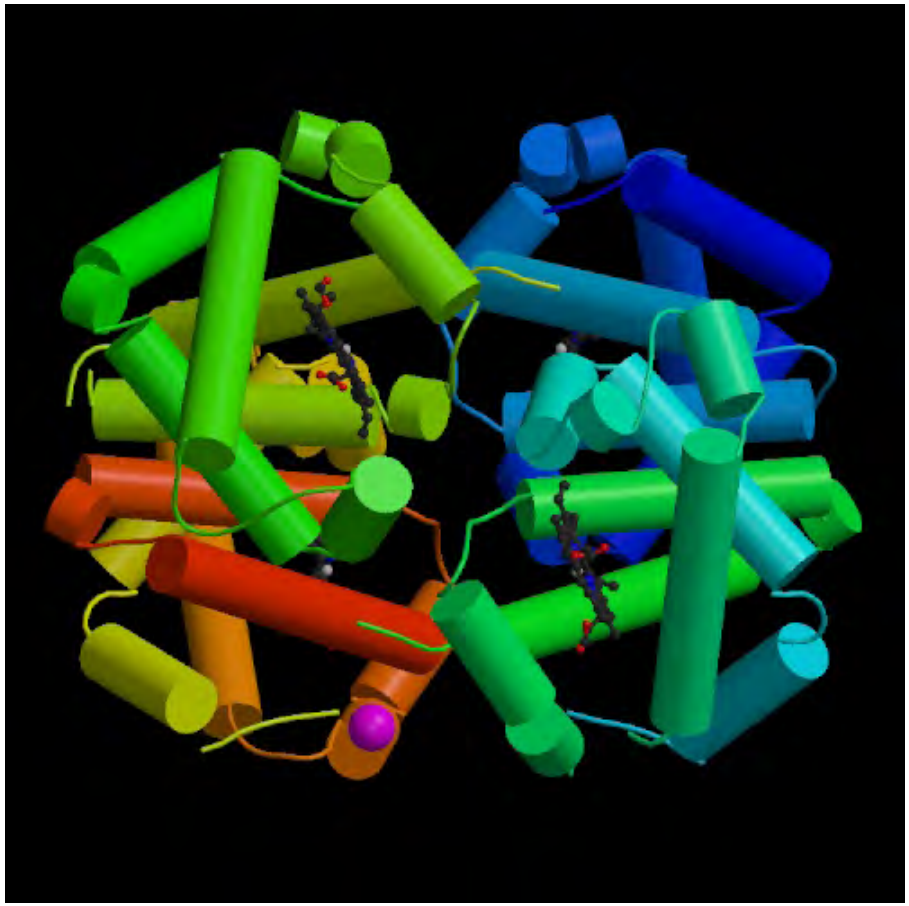
Chemistry→Biological Structure

α-Helix



Pauling and Corey (1951)

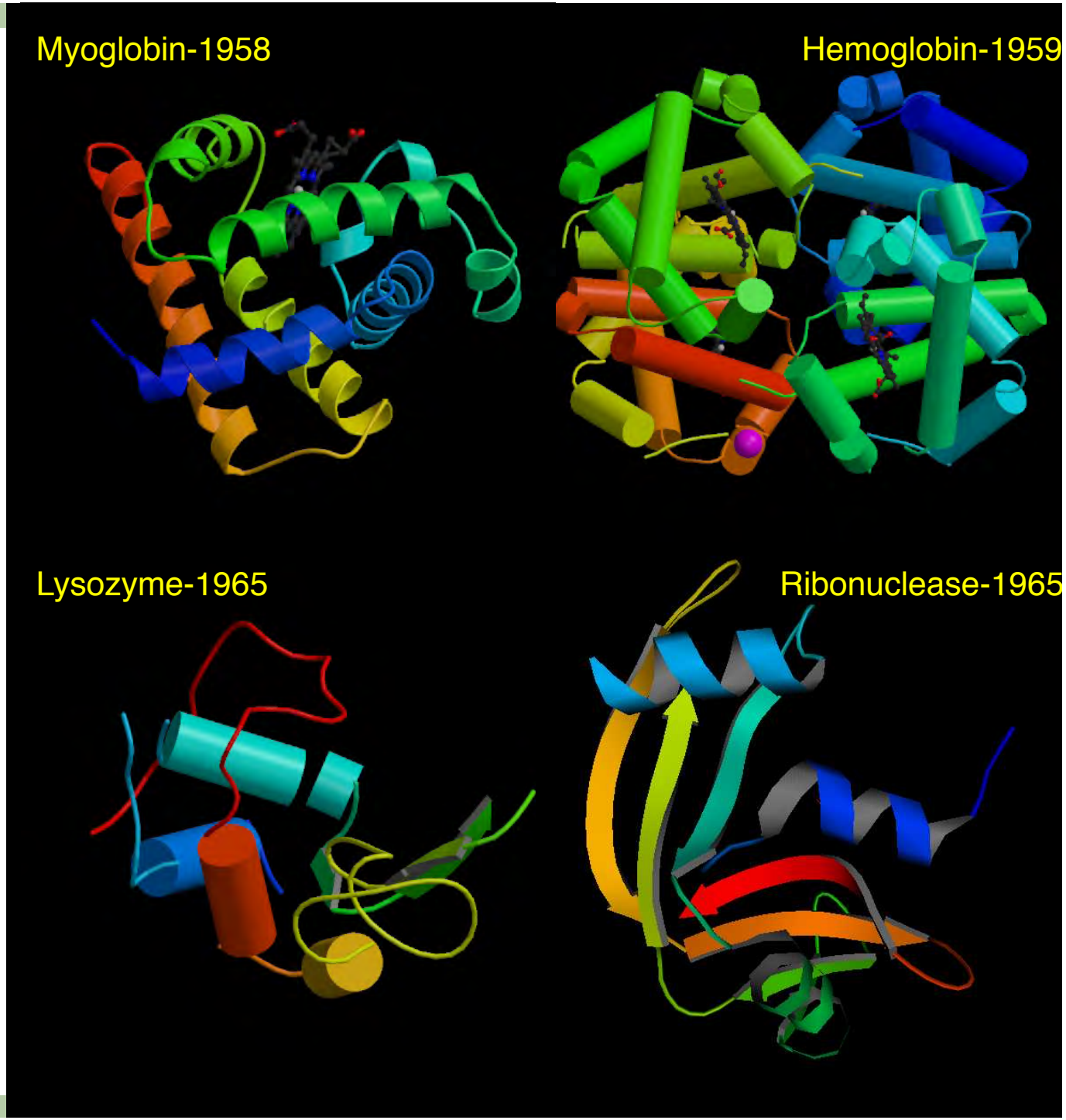
Hemoglobin



Perutz *et al*. (1959)

By the mid-1960s things were getting out of hand!

Something had to be done …



First Electronic, Open Access Resource for the Biological Sciences

CSHL Symposia June 4-11 1971

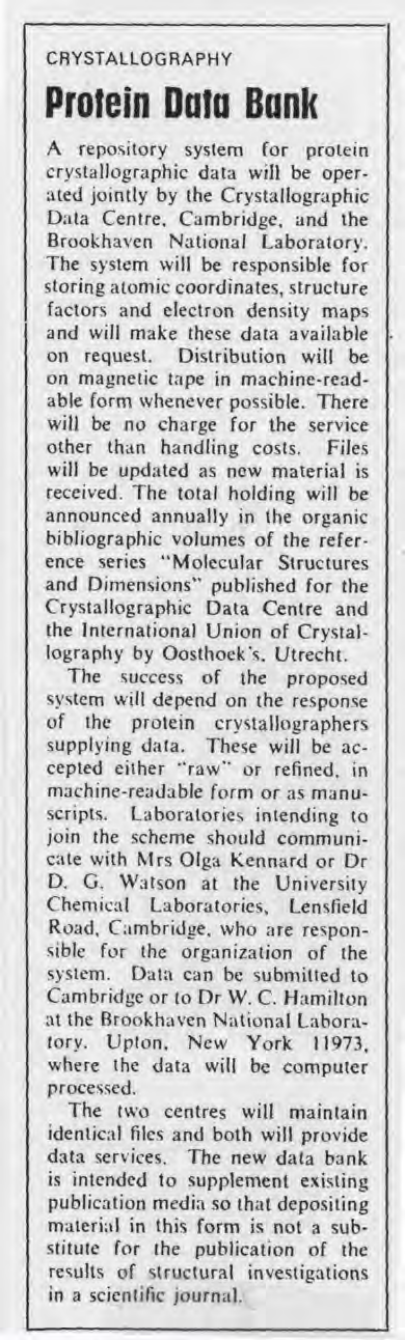
- “Structure and Function of Proteins at the Three-Dimensional Level”
- Organizer: James D. Watson
- Advisors: Aaron Klug, William N. Lipscomb, Max Perutz, David C. Phillips, and Frederic M. Richards

PDB is Born in 1971

- Protein Data Bank found its first home at Brookhaven National Laboratory with only 7 structures
- Founding Director: Walter C. Hamilton

- Announced in *Nature New Biology* with the following caveat:

  "The success of the proposed system will depend on the response of protein crystallographers supplying data."



CRYSTALLOGRAPHY

Protein Data Bank

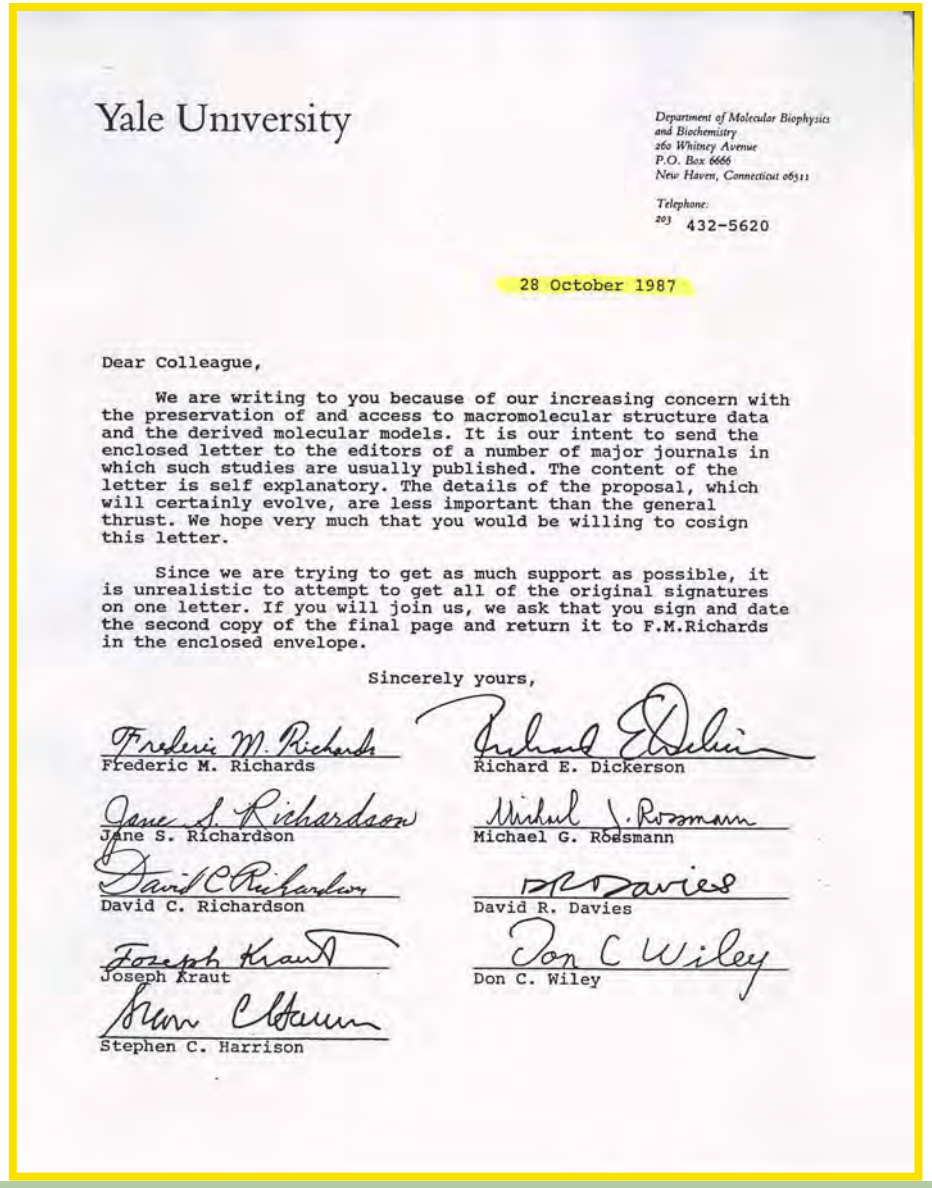
A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available 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.

Growth of the PDB:
The Singular Archive for Macromolecular Structure Data

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

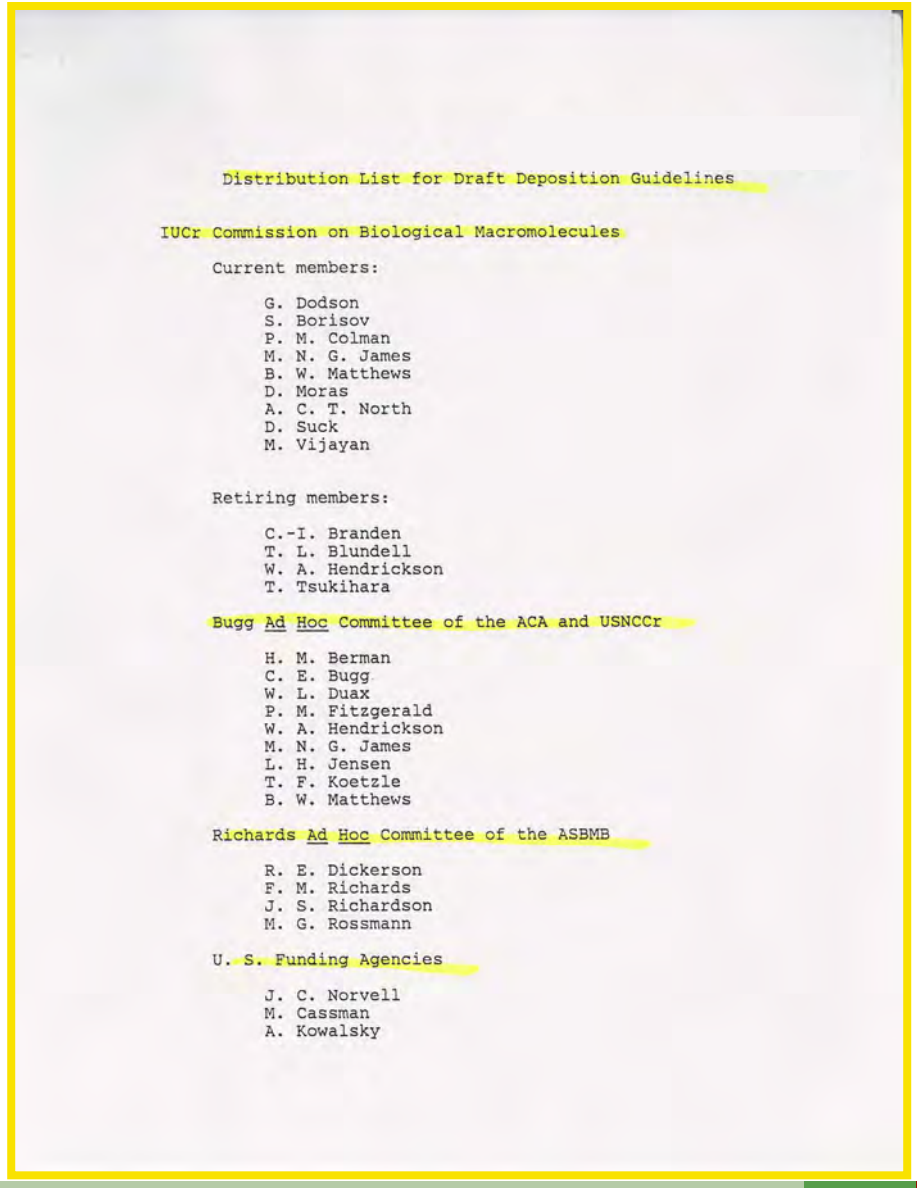
Dear Colleague,

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

Retiring members:

C.-I. Branden
T. L. Blundell
W. A. Hendrickson
T. Tsukihara

Bugg Ad Hoc Committee of the ACA and USNCCr

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

Richards Ad Hoc Committee of the ASBMB

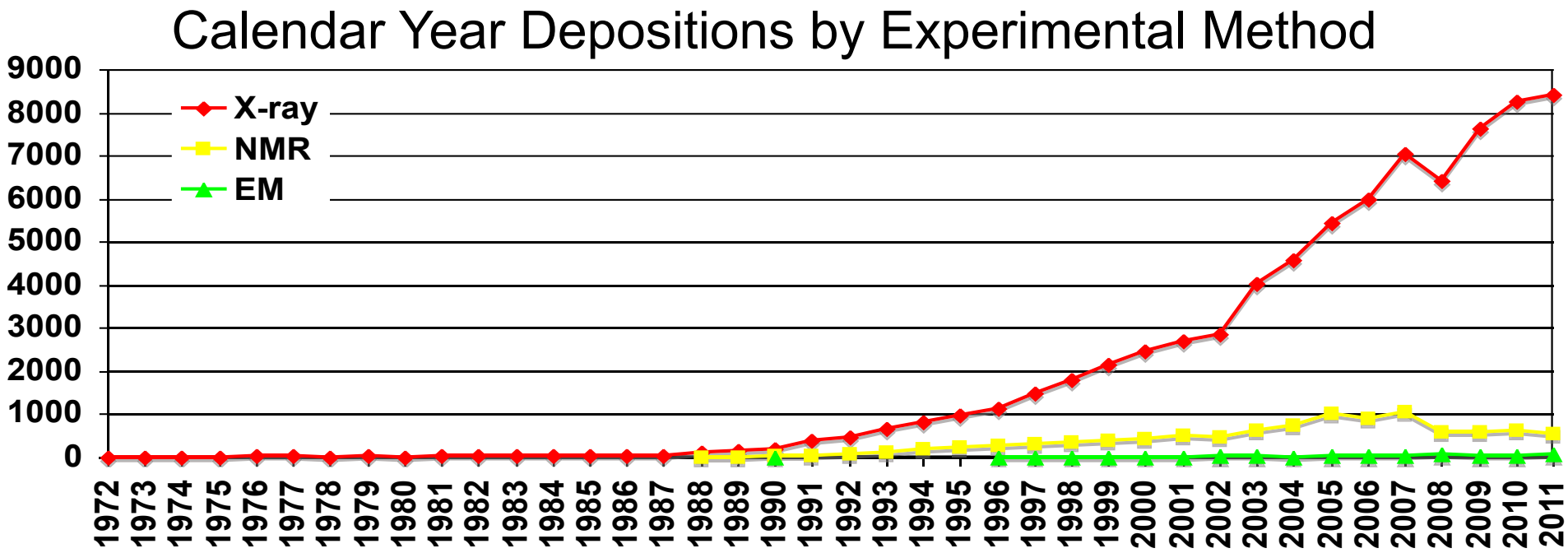
R. E. Dickerson
F. M. Richards
J. S. Richardson
M. G. Rossmann

U. S. Funding Agencies

J. C. Norvell
M. Cassman
A. Kowalsky

10,000-Fold Growth in Four Decades

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

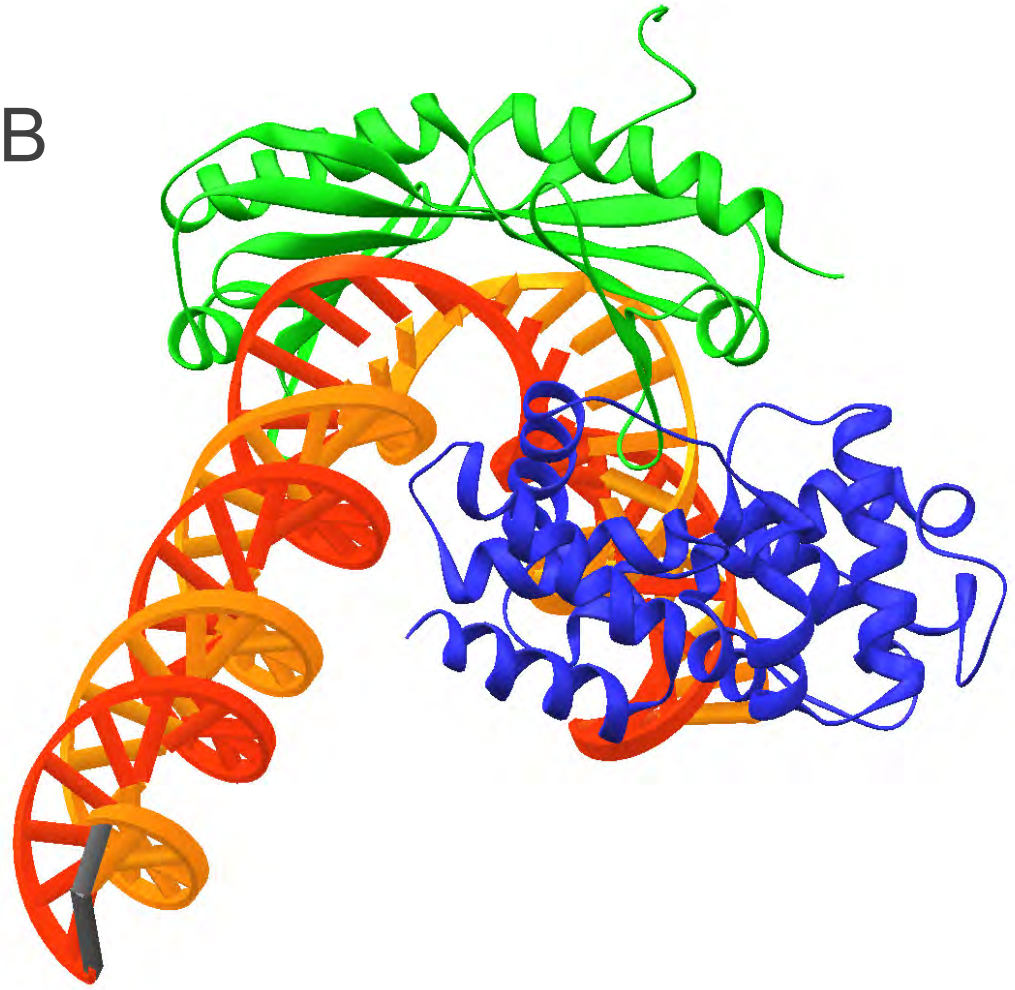


20 Person Years→20 Person Days

- 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

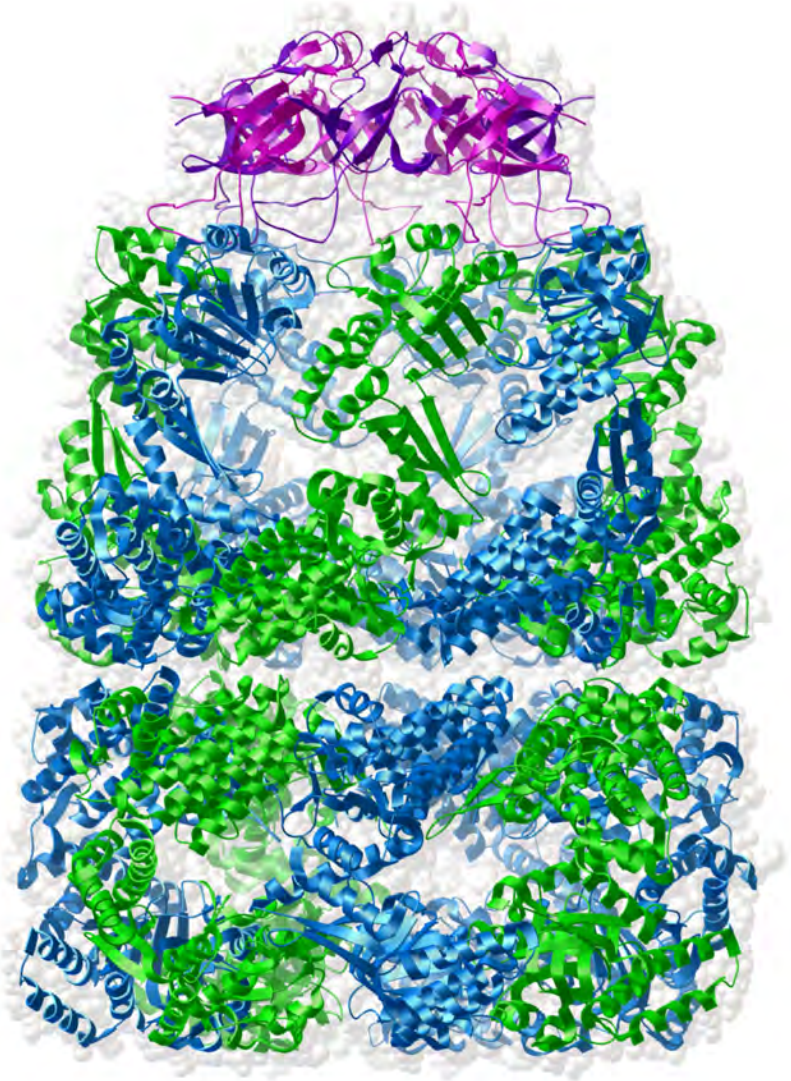
Function Follows Form

- TBP+DNA+TFIIB



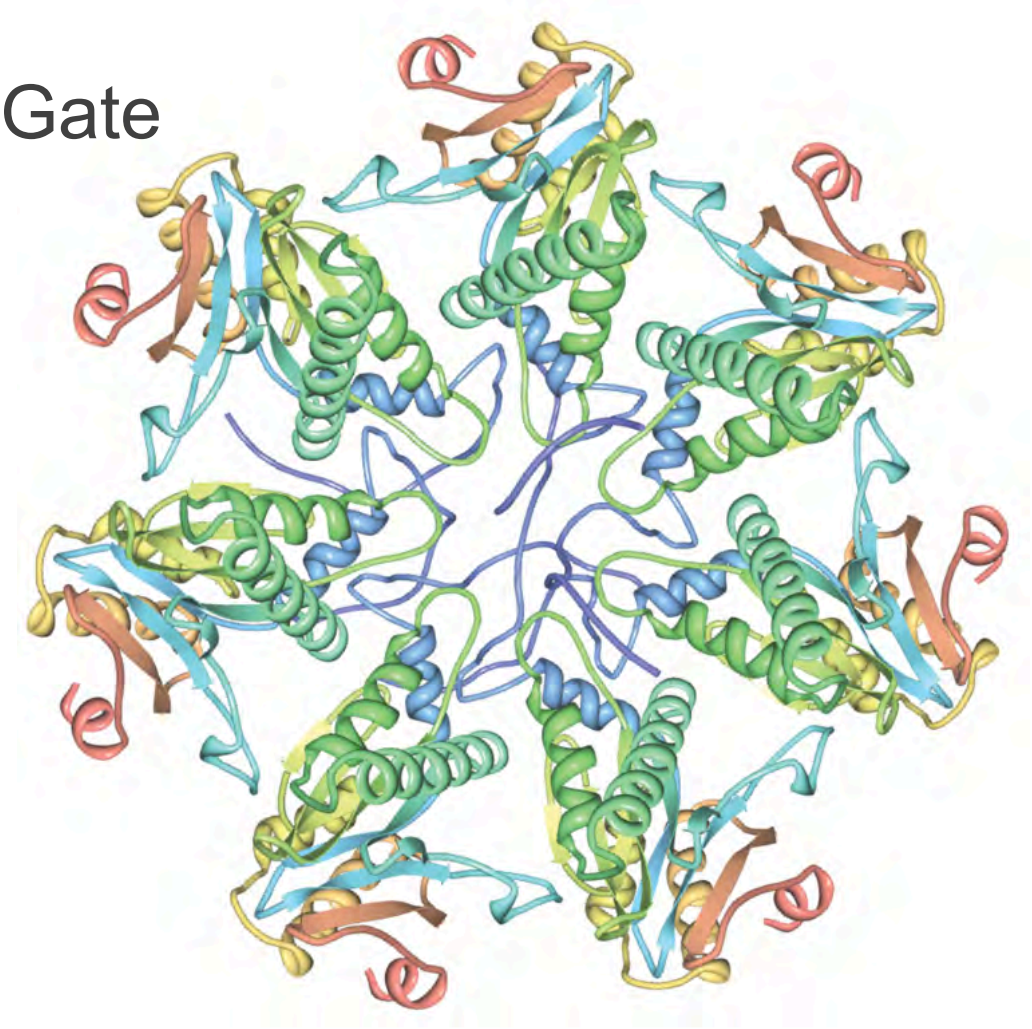
Function Follows Form

- GroEL-GroES



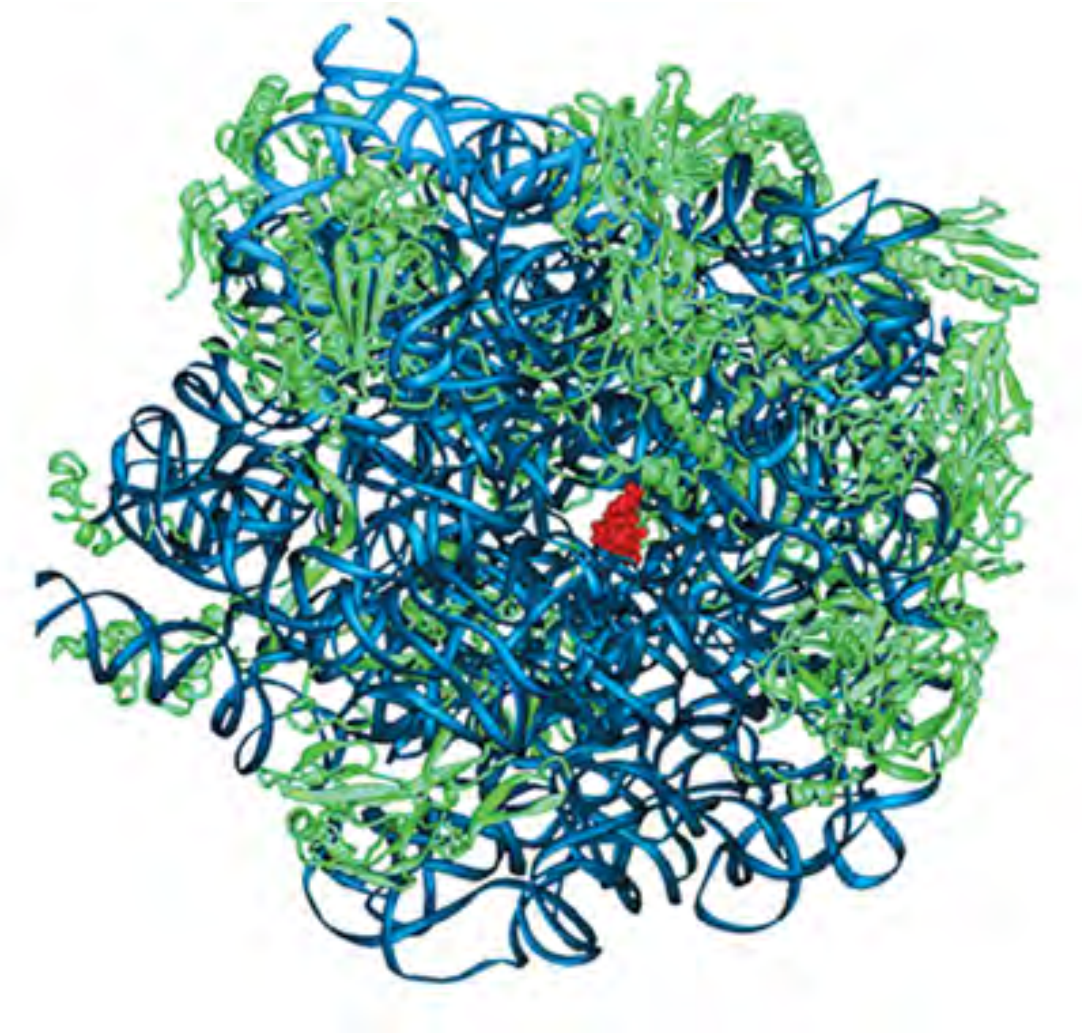
Function Follows Form

- Archael Proteasome Gate



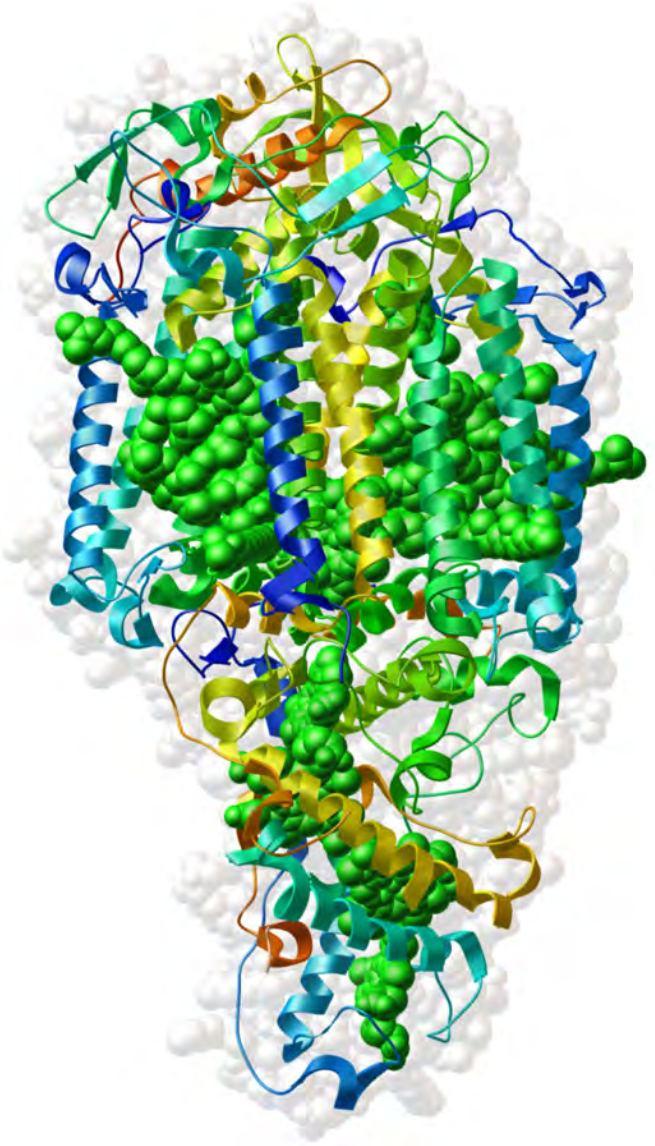
Function Follows Form

- Ribosome



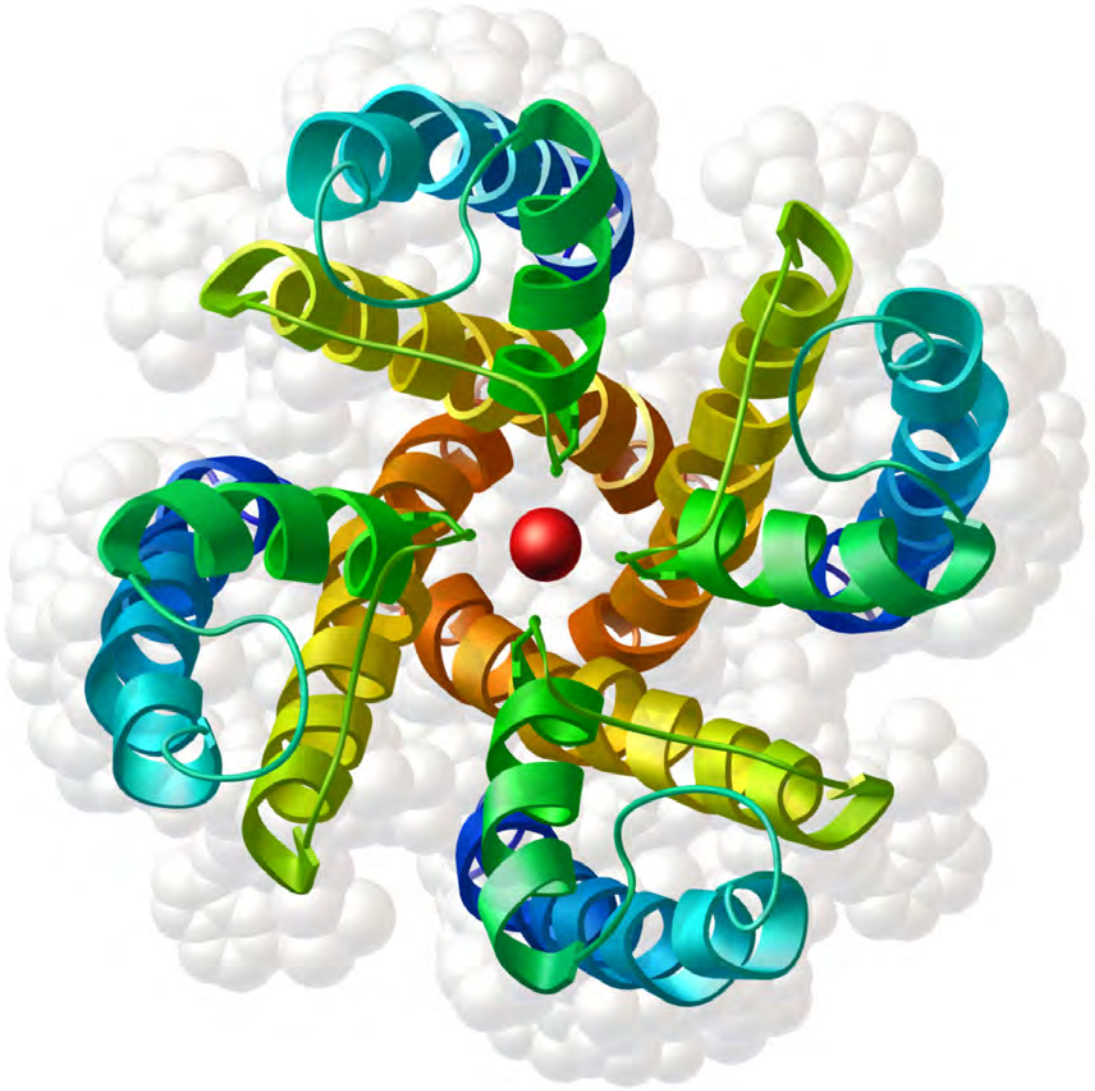
Function Follows Form

- Photosynthetic Reaction Center



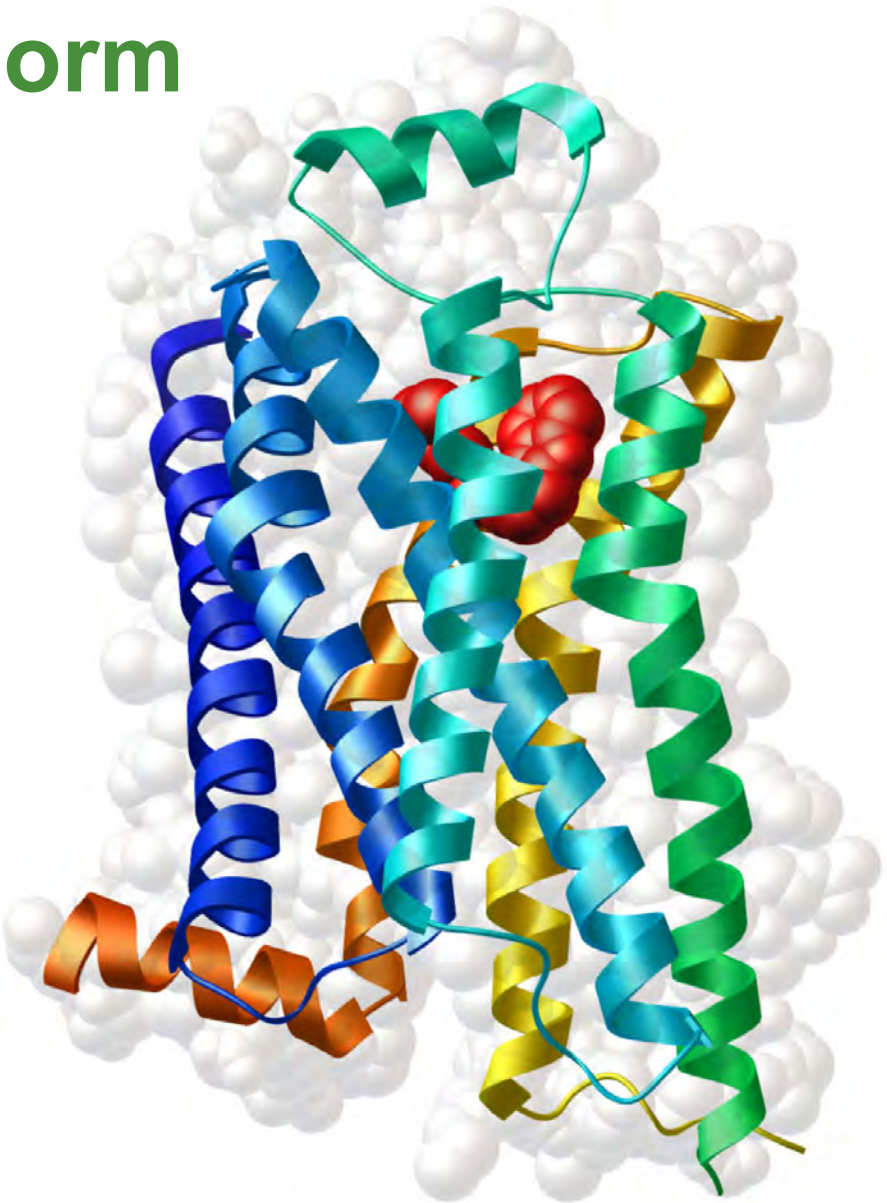
Function Follows Form

- K^+ Channel



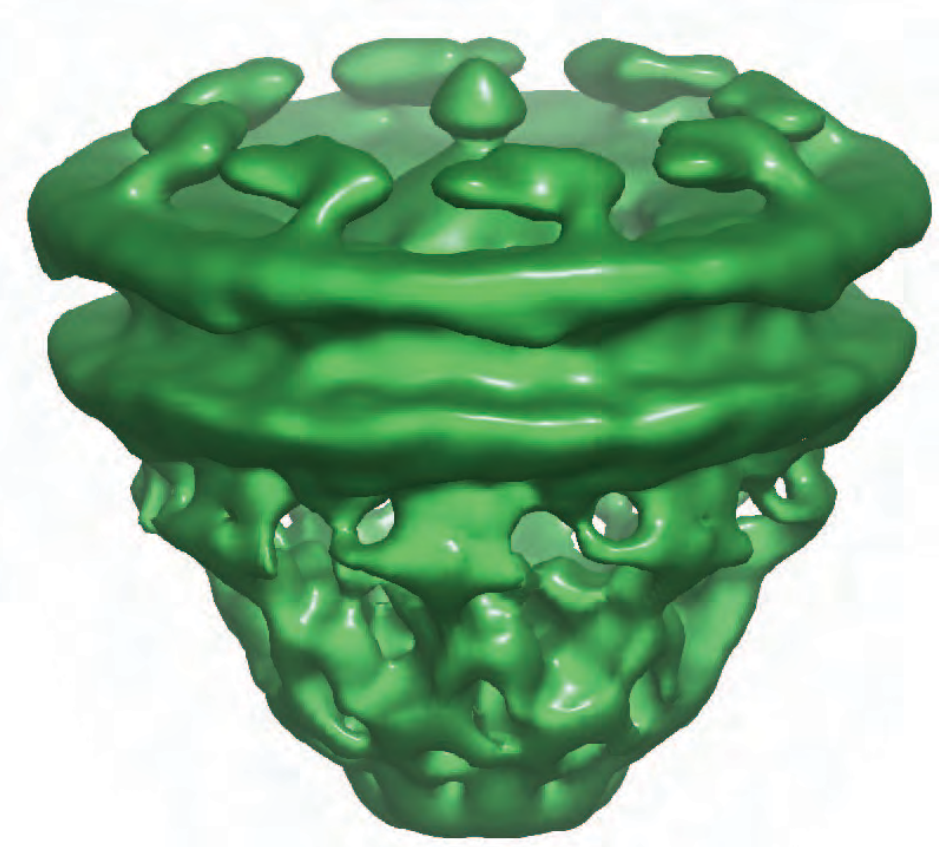
Function Follows Form

- β Adrenergic GPCR



Function Follows Form

- Nuclear Pore Complex



PDB Downloads~210 Million in 2010!

RCSB PDB PDBe PDBj

"Science knows no country, because knowledge belongs to humanity and is the torch that illuminates the world."

Louis Pasteur

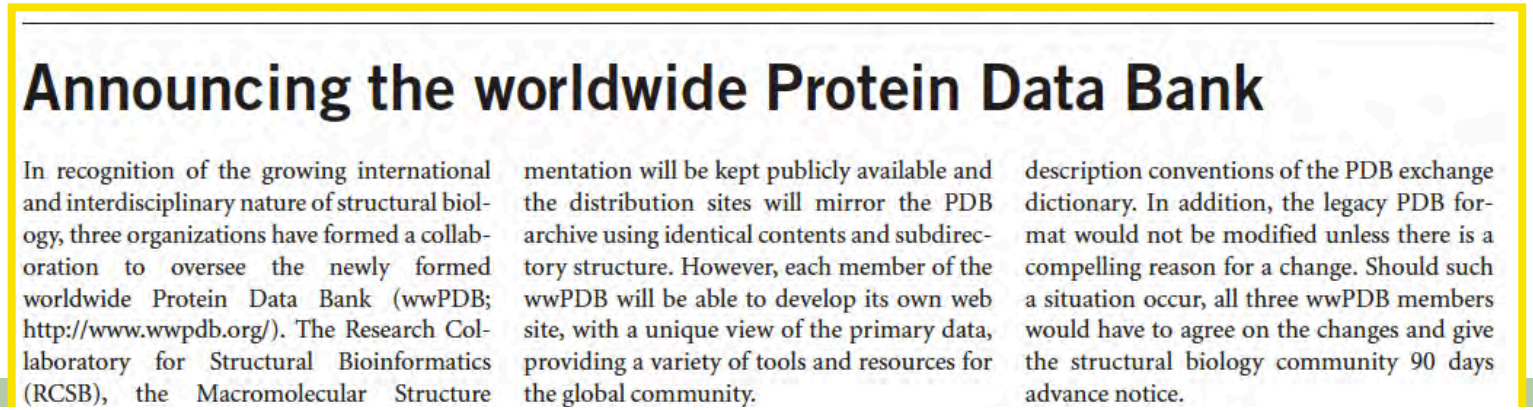
Globalization of the PDB

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

wwPDB Established in 2003

- Membership
  - RCSB PDB (Research Collaboratory for Structural Bioinformatics - Rutgers University/UC San Diego)
  - PDBj (Osaka University)
  - PDBe (EMBL EBI)
  - BMRB (University of Wisconsin)*
- MOU signed July 1st 2003; Amended in 2007*
- Announced in *Nature Structural Biology*



Announcing the worldwide Protein Data Bank

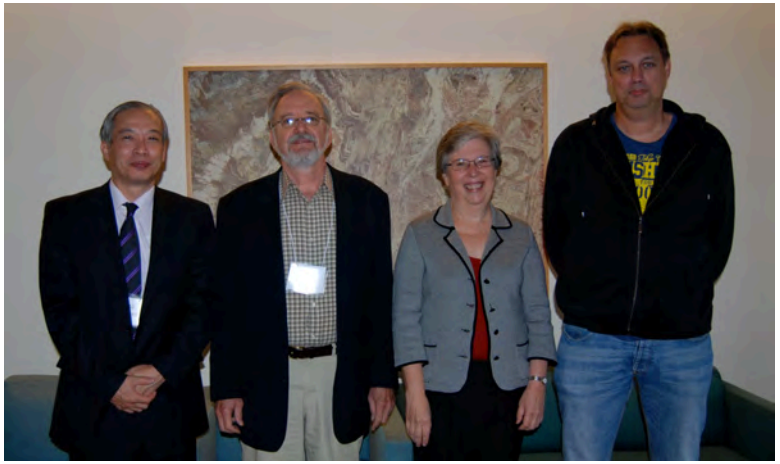
In recognition of the growing international and interdisciplinary nature of structural biology, three organizations have formed a collaboration to oversee the newly formed worldwide Protein Data Bank (wwPDB; http://www.wwpdb.org/). The Research Collaboratory for Structural Bioinformatics (RCSB), the Macromolecular Structure mentation will be kept publicly available and the distribution sites will mirror the PDB archive using identical contents and subdirectory structure. However, each member of the wwPDB will be able to develop its own web site, with a unique view of the primary data, providing a variety of tools and resources for the global community.

description conventions of the PDB exchange dictionary. In addition, the legacy PDB format would not be modified unless there is a compelling reason for a change. Should such a situation occur, all three wwPDB members would have to agree on the changes and give the structural biology community 90 days advance notice.

wwPDB Today

Advisory Committee Meeting 2011

Leadership



PDBj Protein Data Bank Japan

BMRB

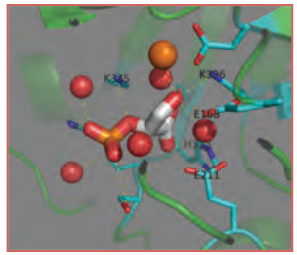
RCSB PDB PROTEIN DATA BANK

EMBL-EBI PDBe PROTEIN DATA BANK EUROPE

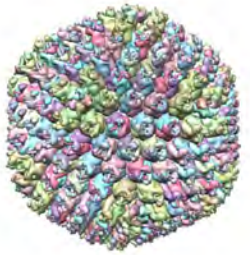
wwPDB Member Responsibilities

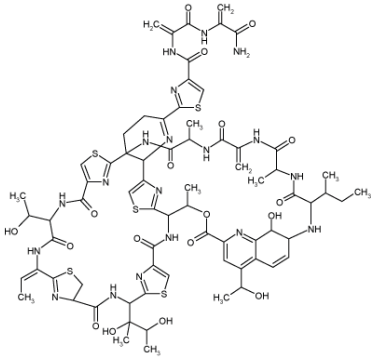
- Members collaborate on “Data In”
  - Issue PDB IDs
- RCSB PDB serves as the Archive Keeper
  - Manage PDB IDs
  - Sole write access
- Members compete on “Data Out”
  - Distribute Identical Data
  - “Market Place of Ideas” concept

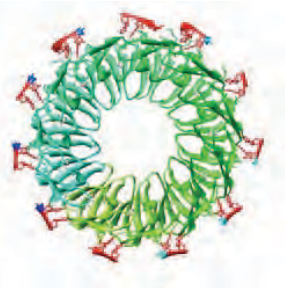
wwPDB Archive Remediation



- 2007: Sequences/Chemistry, Viruses …
- 2009: New Record Types, Enhanced Annotations, Database References, Various Corrections …
- 2011: Peptide Antibiotics, Biological Assemblies, Tagging Entries …
- 2012: Next!

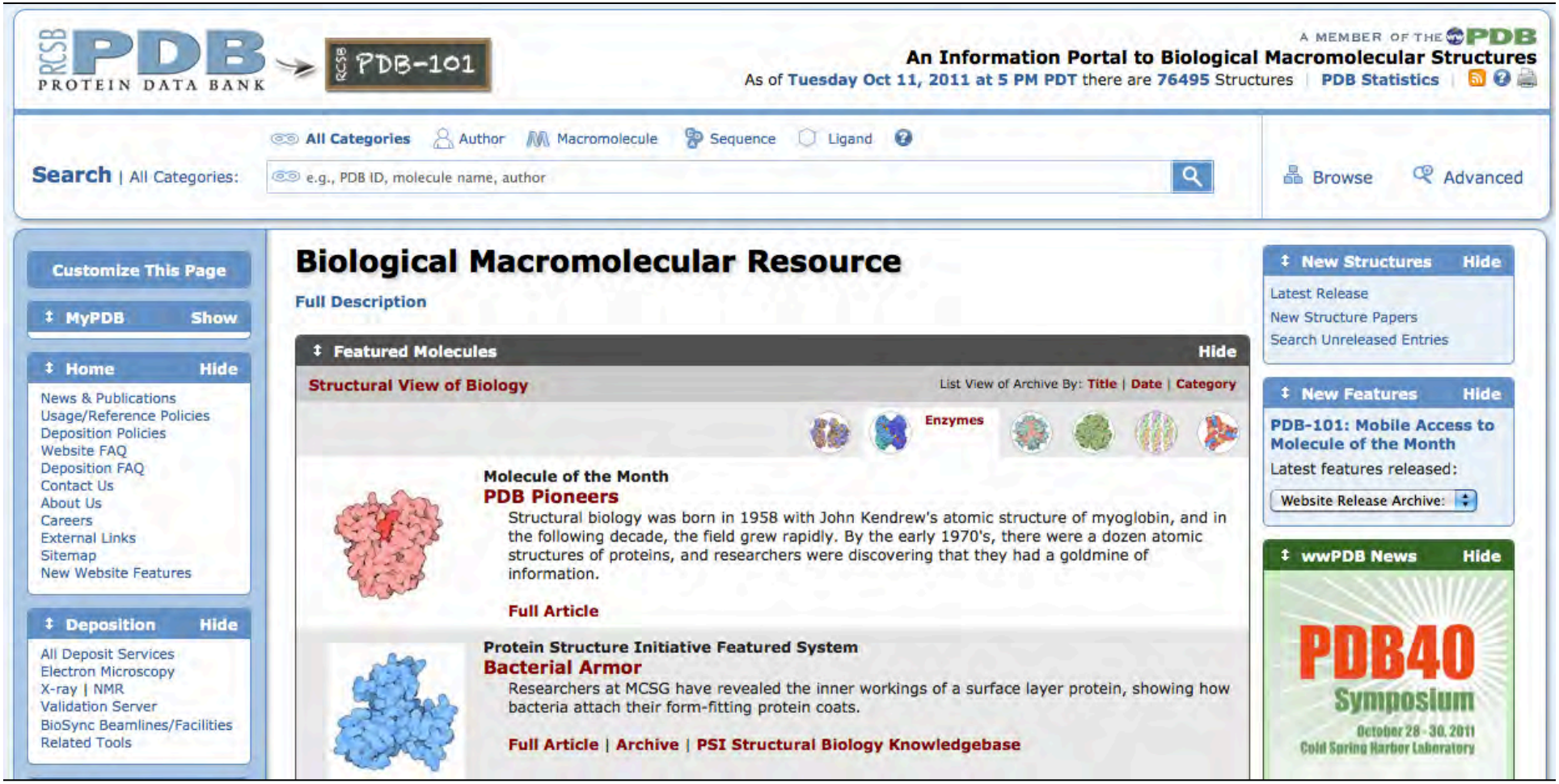




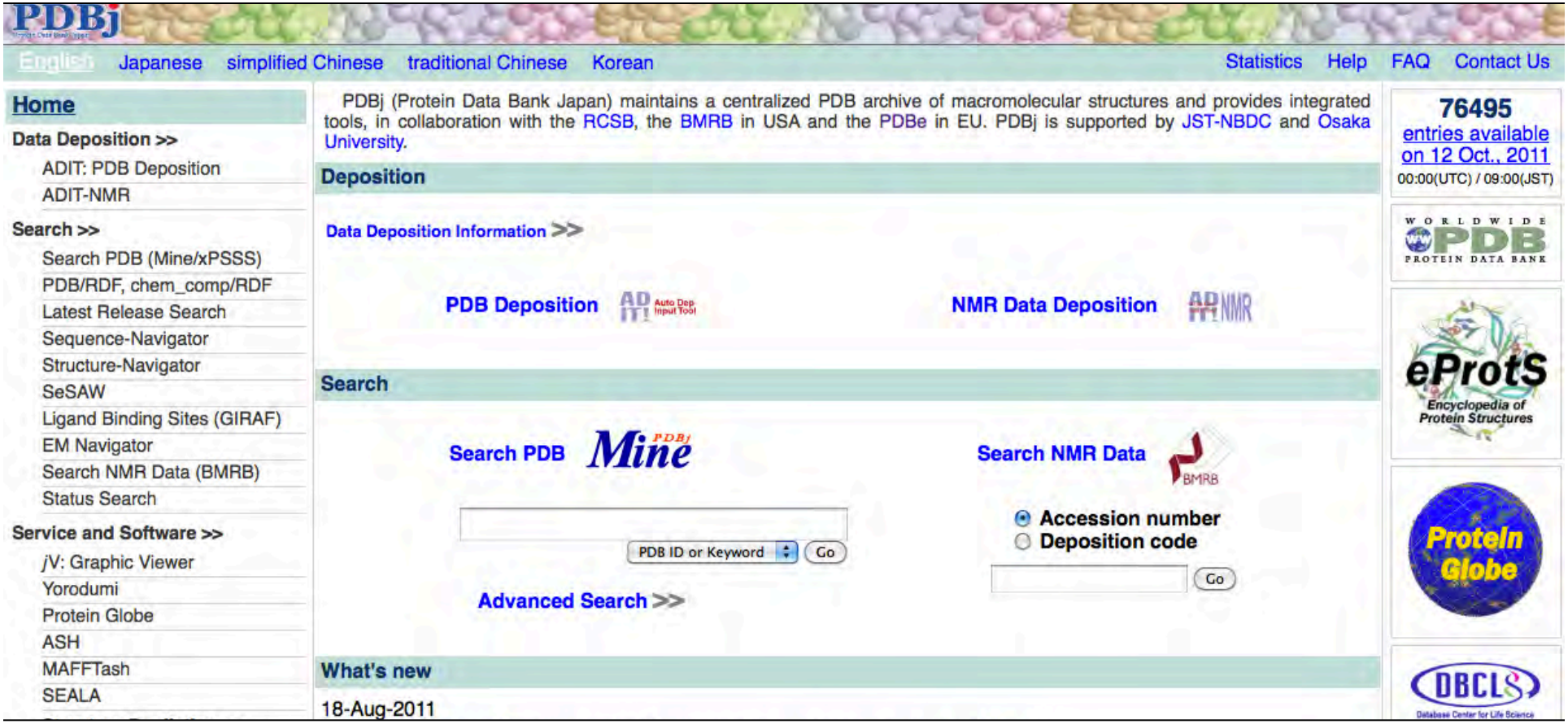


Please visit http://www.wwpdb.org

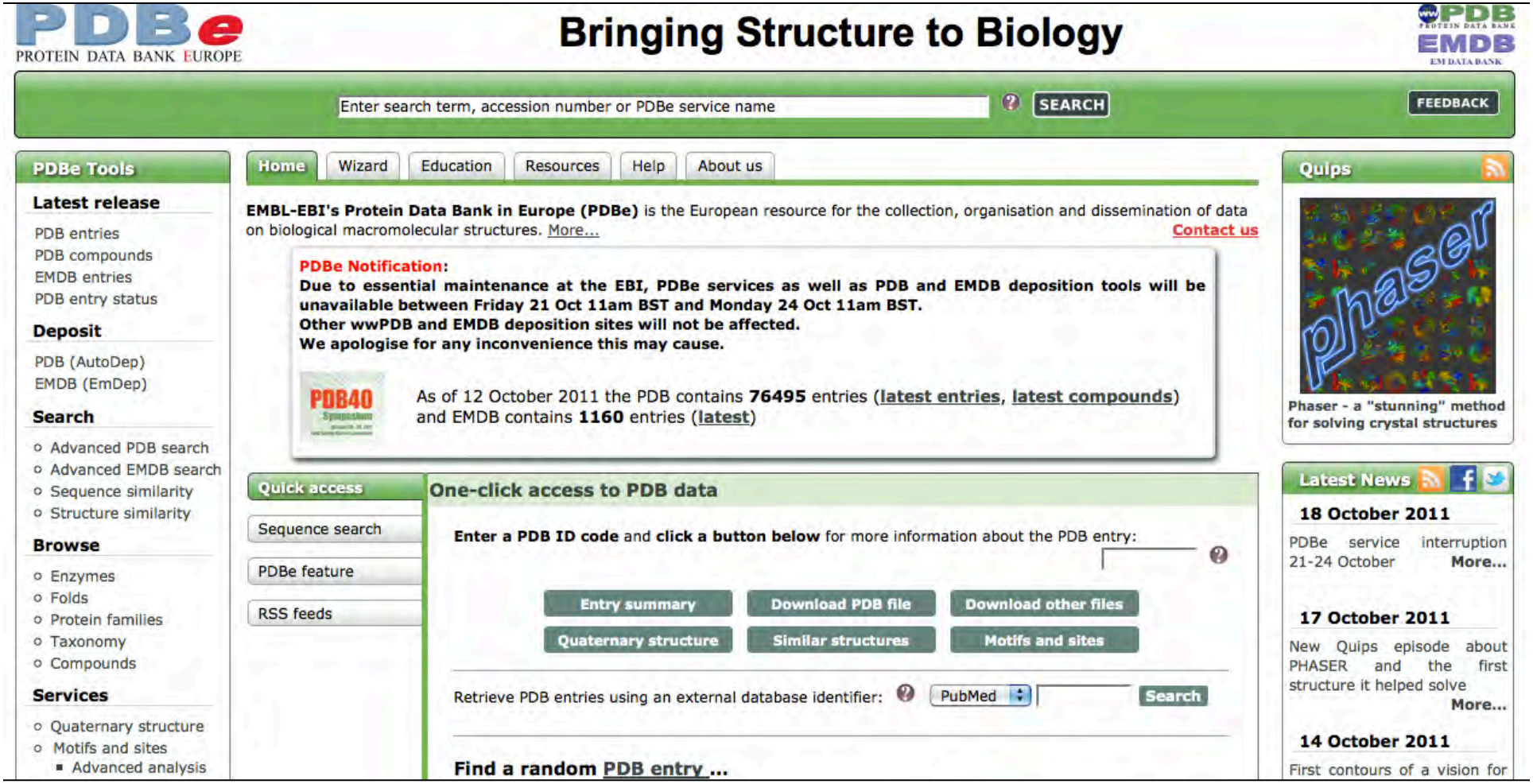
RCSB PDB


RCSB PDB
PROTEIN DATA BANK
PDB-101
A MEMBER OF THE wwPDB
An Information Portal to Biological Macromolecular Structures
As of Tuesday Oct 11, 2011 at 5 PM PDT there are 76495 Structures | PDB Statistics
All Categories | Author | Macromolecule | Sequence | Ligand
Search | All Categories:
e.g., PDB ID, molecule name, author
Browse | Advanced
Customize This Page
MyPDB Show
Home Hide
News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features
Deposition Hide
All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamlines/Facilities
Related Tools
Biological Macromolecular Resource
Full Description
Featured Molecules Hide
Structural View of Biology
List View of Archive By: Title | Date | Category
Enzymes
Molecule of the Month
PDB Pioneers
Structural biology was born in 1958 with John Kendrew's atomic structure of myoglobin, and in the following decade, the field grew rapidly. By the early 1970's, there were a dozen atomic structures of proteins, and researchers were discovering that they had a goldmine of information.
Full Article
Protein Structure Initiative Featured System
Bacterial Armor
Researchers at MCSG have revealed the inner workings of a surface layer protein, showing how bacteria attach their form-fitting protein coats.
Full Article | Archive | PSI Structural Biology Knowledgebase
New Structures Hide
Latest Release
New Structure Papers
Search Unreleased Entries
New Features Hide
PDB-101: Mobile Access to Molecule of the Month
Latest features released:
Website Release Archive:
wwPDB News Hide
PDB40
Symposium
October 28 - 30, 2011
Cold Spring Harbor Laboratory

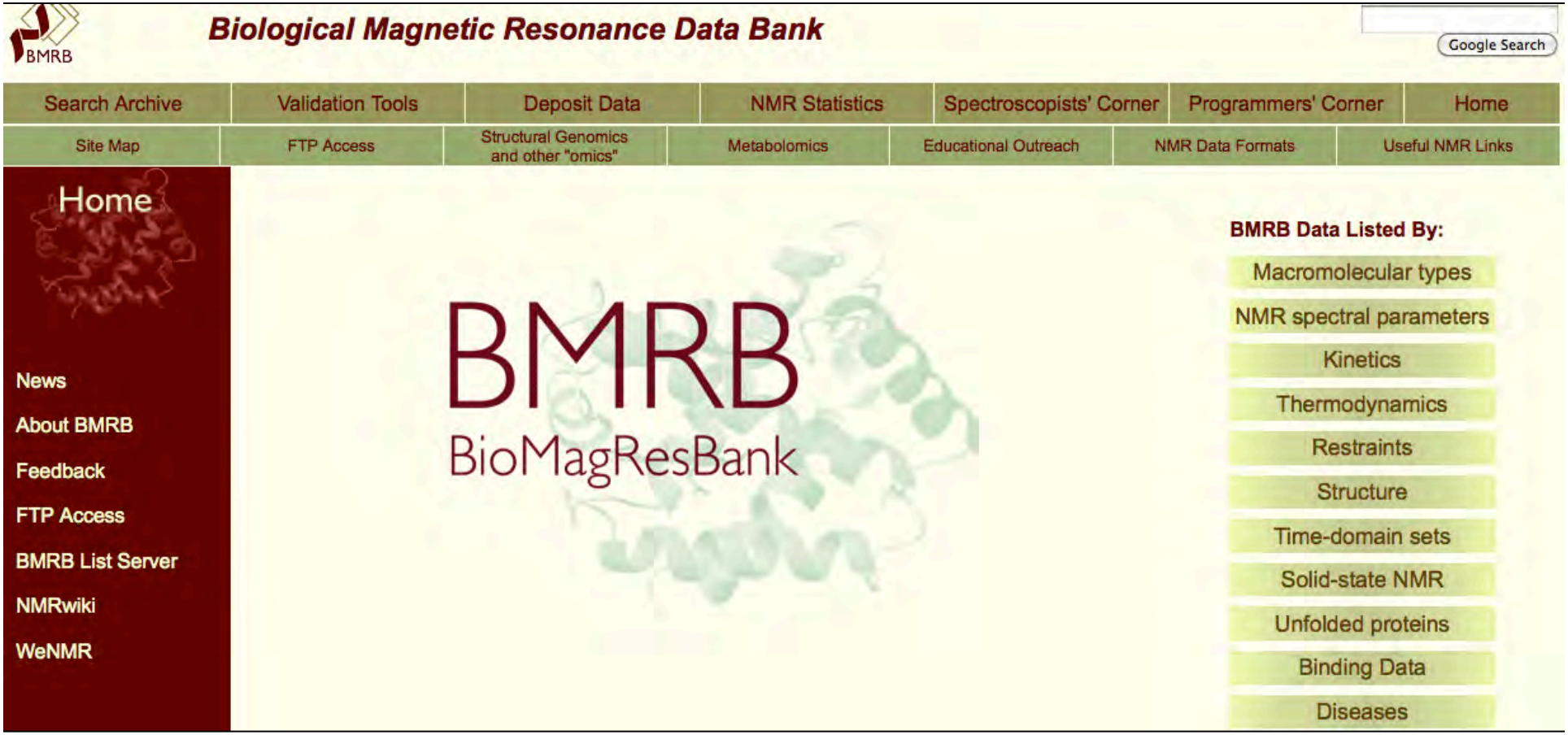
PDBj


PDBj
English Japanese simplified Chinese traditional Chinese Korean
Statistics Help FAQ Contact Us
Home
Data Deposition >>
ADIT: PDB Deposition
ADIT-NMR
Search >>
Search PDB (Mine/xPSSS)
PDB/RDF, chem_comp/RDF
Latest Release Search
Sequence-Navigator
Structure-Navigator
SeSAW
Ligand Binding Sites (GIRAF)
EM Navigator
Search NMR Data (BMRB)
Status Search
Service and Software >>
jV: Graphic Viewer
Yorodumi
Protein Globe
ASH
MAFFTash
SEALA
PDBj (Protein Data Bank Japan) maintains a centralized PDB archive of macromolecular structures and provides integrated tools, in collaboration with the RCSB, the BMRB in USA and the PDBe in EU. PDBj is supported by JST-NBDC and Osaka University.
Deposition
Data Deposition Information >>
PDB Deposition
NMR Data Deposition
Search
Search PDB
Search NMR Data
PDB ID or Keyword
Go
Accession number
Deposition code
Go
Advanced Search >>
What's new
18-Aug-2011
76495
entries available on 12 Oct., 2011
00:00(UTC) / 09:00(JST)
WORLDWIDE PDB PROTEIN DATA BANK
eProtS
Encyclopedia of Protein Structures
Protein Globe
DBCLS
Database Center for Life Science

PDBe


PDBe
PROTEIN DATA BANK EUROPE
Bringing Structure to Biology
PDB
PROTEIN DATA BANK
EMDB
EM DATA BANK
Enter search term, accession number or PDBe service name
SEARCH
FEEDBACK
PDBe Tools
Latest release
PDB entries
PDB compounds
EMDB entries
PDB entry status
Deposit
PDB (AutoDep)
EMDB (EmDep)
Search
Advanced PDB search
Advanced EMDB search
Sequence similarity
Structure similarity
Browse
Enzymes
Folds
Protein families
Taxonomy
Compounds
Services
Quaternary structure
Motifs and sites
Advanced analysis
Home
Wizard
Education
Resources
Help
About us
EMBL-EBI's Protein Data Bank in Europe (PDBe) is the European resource for the collection, organisation and dissemination of data on biological macromolecular structures. More...
Contact us
PDBe Notification:
Due to essential maintenance at the EBI, PDBe services as well as PDB and EMDB deposition tools will be unavailable between Friday 21 Oct 11am BST and Monday 24 Oct 11am BST.
Other wwPDB and EMDB deposition sites will not be affected.
We apologise for any inconvenience this may cause.
PDB40
As of 12 October 2011 the PDB contains 76495 entries (latest entries, latest compounds) and EMDB contains 1160 entries (latest)
Quick access
Sequence search
PDBe feature
RSS feeds
One-click access to PDB data
Enter a PDB ID code and click a button below for more information about the PDB entry:
Entry summary
Download PDB file
Download other files
Quaternary structure
Similar structures
Motifs and sites
Retrieve PDB entries using an external database identifier:
PubMed
Search
Find a random PDB entry ...
Quips
phaser
Phaser - a "stunning" method for solving crystal structures
Latest News
18 October 2011
PDBe service interruption 21-24 October More...
17 October 2011
New Quips episode about PHASER and the first structure it helped solve More...
14 October 2011
First contours of a vision for

BMRB


BMRB
Biological Magnetic Resonance Data Bank
Google Search
Search Archive
Validation Tools
Deposit Data
NMR Statistics
Spectroscopists' Corner
Programmers' Corner
Home
Site Map
FTP Access
Structural Genomics and other "omics"
Metabolomics
Educational Outreach
NMR Data Formats
Useful NMR Links
Home
News
About BMRB
Feedback
FTP Access
BMRB List Server
NMRwiki
WeNMR
BMRB
BioMagResBank
BMRB Data Listed By:
Macromolecular types
NMR spectral parameters
Kinetics
Thermodynamics
Restraints
Structure
Time-domain sets
Solid-state NMR
Unfolded proteins
Binding Data
Diseases

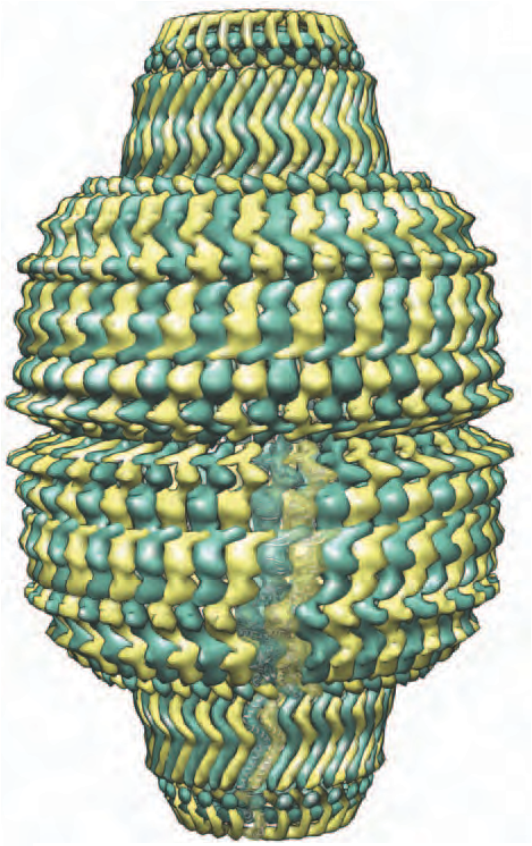
PDB of the Future

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

What: Responding to User Needs

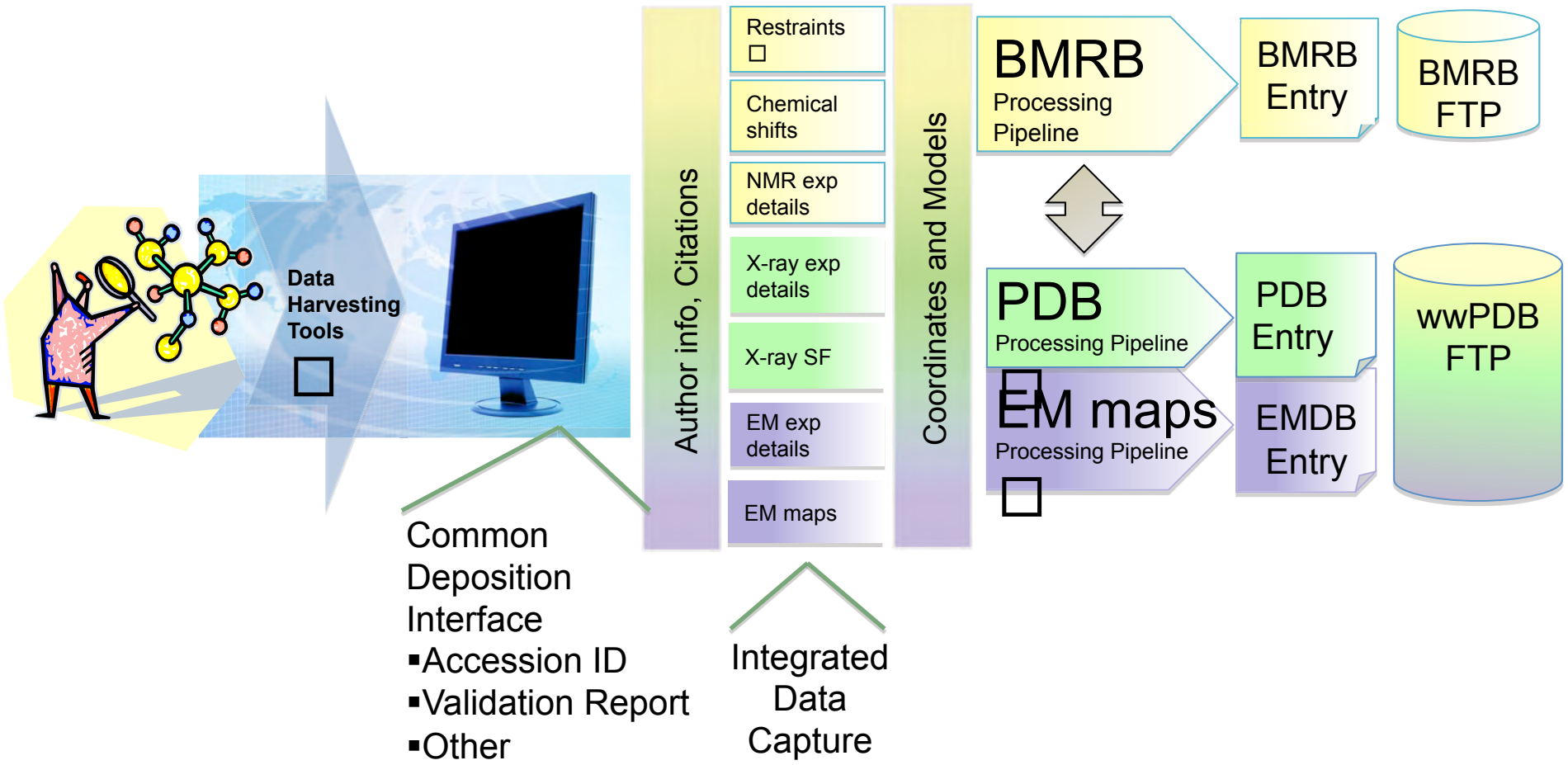
- Higher deposition rates
- Increasingly complex structures
- Enhanced validation
- Expanded annotation
- Hybrid methods



Why: What's In It For …

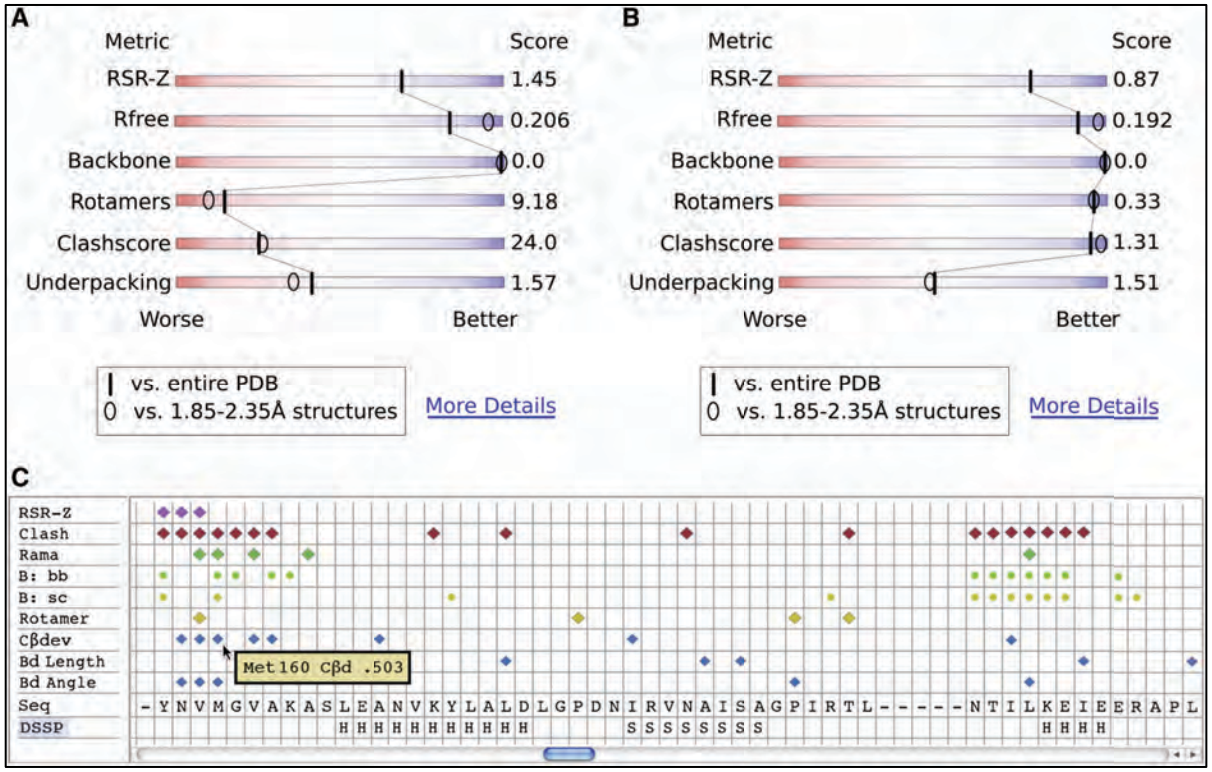
- Depositors
  - Interactive Deposition Interface
  - Validation/Annotation
  - Increased Efficiency
  - Support New/Hybrid Methods
- Annotators
  - Increased Throughput
  - Advanced Annotation
- Researchers/Educators
  - Highest Quality Archive

How: Common Deposition Tool



How: Enhanced Validation Systems

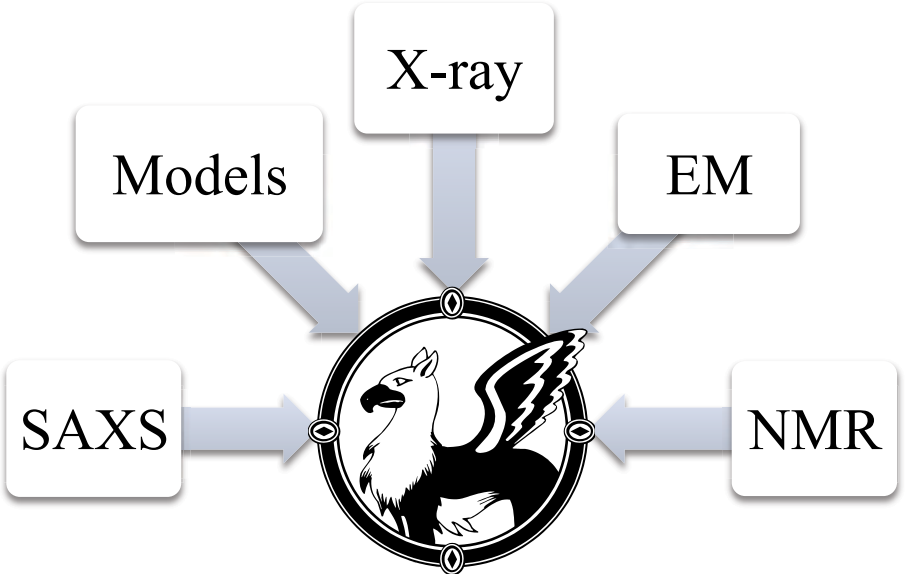
- User Driven
- Transparent
- Interpretable



From *A new generation of crystallographic validation tools for the Protein Data Bank* Read *et al*. (2011) *Structure 19*, 1395-1412.

How: Hybrid Methods→New Biology

- Today: Combined Neutron/X-ray Structures
- Tomorrow: Scientist's Choice



- Extensible Dictionary
- Modular System

Plus ça change
(Plus c'est la même chose)

The more things change
(The more they remain the same)

WORLDWIDE PDB
PROTEIN DATA BANK

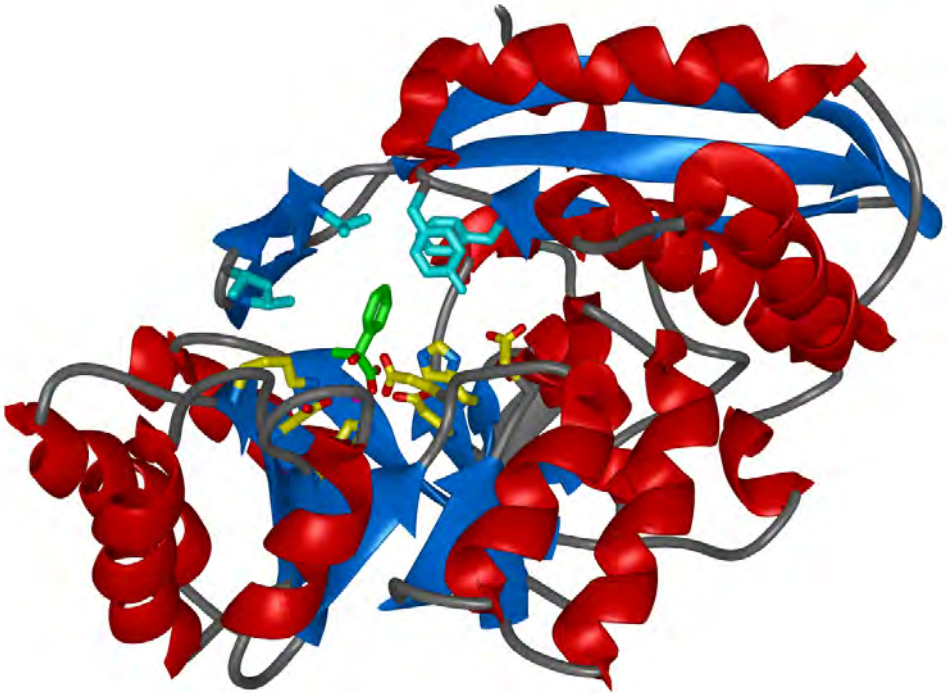
wwpdb.org

1971: What does it Look Like?

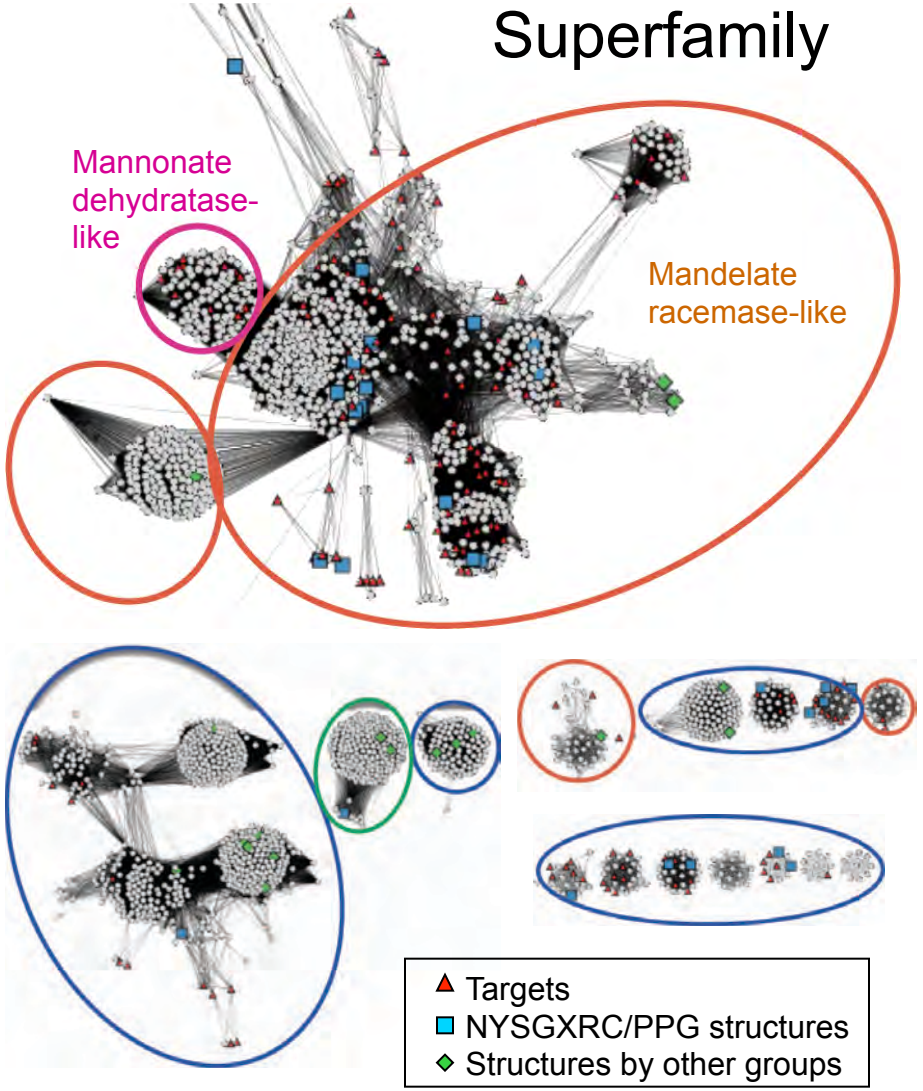
Sperm Whale Myoglobin—Kendrew *et al*. (Irving Geis)

2011: What does it Look Like?

Triosphosphate Isomerase $(\alpha\beta)_8$ Barrel Superfold from Phillips *et al*. occurs in ~5% of all proteins



Enolase Superfamily



WORLDWIDE PDB
PROTEIN DATA BANK

PDB40 Symposium

October 28 - 30, 2011
Cold Spring Harbor Laboratory

Come celebrate four decades of innovation in structural biology

Speakers

- **Cheryl Arrowsmith**, University of Toronto, Canada
- **David Baker**, University of Washington
- **Ad Bax**, NIH/DHHS/NIDDK/LCP
- **Axel Brunger**, Stanford University/HHMI
- **Stephen K. Burley**, Eli Lilly & Co.
- **Wah Chiu**, Baylor College of Medicine
- **Johann Deisenhofer,** UT Southwestern Medical Center
- **Angela Gronenborn**, University of Pittsburgh
- **Richard Henderson**, MRC Lab. of Molecular Biology
- **Wayne Hendrickson**, Columbia University
- **Mei Hong**, Iowa State University
- **Brian Matthews**, University of Oregon
- **Jane Richardson**, Duke University Medical Center
- **Michael Rossmann**, Purdue University
- **Andrej Sali**, University of California, San Francisco
- **David Searls**, Independent Consultant
- **Susan Taylor**, University of California, San Diego
- **Janet Thornton**, EMBL EBI, Hinxton
- **Soichi Wakatsuki**, IMMS-KEK
- **Kurt Wüthrich**, The Scripps Research Institute, ETH Zürich

meetings.cshl.edu/meetings/pdb40.shtml