Growth, Globalization, and Future of the Protein Data Bank

Stephen K. Burley, Eli Lilly & Co.

October 28th 2011
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

Hemoglobin

Pauling and Corey (1951)  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.”
Growth of the PDB: The Singular Archive for Macromolecular Structure Data
1987: Users Compel Deposition

Yale University

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 sign 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
Don C. Wiley

Yale University

29 October 1987

Distribution List for Draft Deposition Guidelines

IUCr Commission on Biological Macromolecules

Current members:
G. Dodson
S. Berisso
F. H. Colman
M. N. G. James
B. W. Matthews
D. Moras
A. C. T. North
D. Philips
M. Vijayan

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

Bugg Ad Hoc Committee of the ACA and USNC
H. M. Berman
C. F. Rugg
W. L. Dux
P. M. Fitzgerald
W. A. Hendrickson
M. N. G. James
L. H. Jensen
T. F. Koetzle
E. W. Matthews

Richards Ad Hoc Committee of the ASBMB
R. E. Dickerson
F. K. Richards
J. S. Richardson
M. G. Rossmann

U. S. Funding Agencies
J. H. Norvell
M. Cassman
A. Kowalsky
10,000-Fold Growth in Four Decades

- 7$\rightarrow$>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

- \( \beta \) Adrenergic GPCR
Function Follows Form

- Nuclear Pore Complex
PDB Downloads~210 Million in 2010!

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

Louis Pasteur
Globalization of the PDB
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  

...
wwPDB Today

Advisory Committee Meeting 2011

Leadership
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!
The Worldwide Protein Data Bank (wwPDB) consists of organizations that act as deposition, data processing and distribution centers for PDB data. The founding members are RCSB PDB (USA), PDB (Europe) and PDB (Japan). The BMRB (USA) group joined the wwPDB in 2006. The mission of the wwPDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community.

This site provides information about services provided by the individual member organizations and about projects undertaken by the wwPDB.

wwPDB Statement on Retraction of PDB Entries

21-October-2011

PDB40 Symposium Update

Celebrate four decades of innovation in structural biology with the wwPDB October 28-30, 2011 at CSHL. View the preliminary program (PDF) and poster abstracts (PDF). Limited space is available—register today!

14-October-2011

First contours of a vision for the future of validation at the PDB

The Worldwide Protein Data Bank (wwPDB; [wwpdb.org](http://www.wwpdb.org)) is pleased to direct PDB depositors and users to the recommendations of the wwPDB X-ray Validation Task Force (VTF) that were published in the journal Structure this week (2011, vol. 19: 1395-1412).
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 PDBa in EU. PDBj is supported by JST-NBDC and Osaka University.

**Deposition**

**Data Deposition Information**

**Search**

**Search PDB**

*Mine*

**Search NMR Data**

- Accession number
- Deposition code

**Advanced Search**

**What's new**

18-Aug-2011
PDBe

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...

PDB 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.

As of 12 October 2011 the PDB contains 76495 entries (latest entries, latest compounds) and EMDB contains 1160 entries (latest).

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...
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

Common Deposition Interface
- Accession ID
- Validation Report
- Other

Author info, Citations
- Restraints
- Chemical shifts
- NMR exp details
- X-ray exp details
- X-ray SF
- EM exp details
- EM maps

Integrated Data Capture

BMRB
Processing Pipeline
- BMRB Entry
- BMRB FTP

PDB
Processing Pipeline
- PDB Entry
- wwPDB FTP
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)
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

Mannionate dehydratase-like

Mandelate racemase-like

| ▲ | Targets |
| --- |
| □ | NYSGXRC/PPG structures |
| ♦ | Structures by other groups |
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