



Published in final edited form as:

Structure. 2013 September 3; 21(9): 1485–1491. doi:10.1016/j.str.2013.07.010.

How Community Has Shaped the Protein Data Bank

Helen M. Berman^{1,*}, Gerard J. Kleywegt², Haruki Nakamura³, and John L. Markley⁴

¹RCSB PDB, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ USA 08854

²PDBe, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

³PDBj, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

⁴BioMagResBank, Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706 USA

Abstract

Following several years of community discussion, the Protein Data Bank (PDB) was established in 1971 as a public repository for the coordinates of three-dimensional models of biological macromolecules. Since then, the number, size, and complexity of structural models have continued to grow, reflecting the productivity of structural biology. Managed by the Worldwide PDB organization, the PDB has been able to meet increasing demands for the quantity of structural information and of quality. In addition to providing unrestricted access to structural information, the PDB also works to promote data standards and to raise the profile of structural biology with broader audiences. In this perspective, we describe the history of PDB and the many ways in which the community continues to shape the archive.

Over the last 40 years, the Protein Data Bank (PDB) has grown from a small data repository promoted by the structural biology community to a critical international resource used around the world in a broad range of scientific research endeavors. Today, more than 90,000 structures are publicly available, and more than 300 million coordinate sets were downloaded in the past year (Figure 1). Many factors have contributed to the growth of the PDB. The scientific drivers for structural biology are numerous, including the desire to understand protein function, to design new pharmaceutical agents, and to understand how molecular machines work. The rapid development and uptake of the latest technologies for protein production, data collection, data analysis and visualization have played a vital part in the ability to determine the three-dimensional (3D) structures of increasingly complex systems. The role of community in shaping the PDB and helping it evolve into a vital resource for biological research cannot be underestimated. In this article we trace the evolution of PDB-community interactions from rather informal town meetings to formally constituted Advisory Committees and Task Forces that work with the Worldwide Protein

Data Bank (wwPDB; <http://wwpdb.org>) (Berman et al., 2003) to establish policy and to improve the contents and quality of the data.

Establishment of the PDB

The first 3D structures of proteins were determined in the 1950s by X-ray crystallography (Kendrew et al., 1958, 1960; Perutz et al., 1960). Around the same time, Cyrus Levinthal and others were carrying out pioneering research on protein folding (Levinthal, 1968), structure prediction, and visualization. The structural biologists recognized the need to share their data with people who wanted to analyze them. A series of ad hoc meetings at American Crystallographic Association (ACA) conferences and at a symposium at Cold Spring Harbor Laboratory (CSHL) in 1971 (Cold Spring Laboratory Press, 1972) involving both producers and potential users of these data culminated in the establishment of the PDB at Brookhaven National Laboratory (BNL). A paper in *Nature New Biology* announced the PDB as an international collaboration with sites in the US and the UK (Protein Data Bank, 1971). A symposium celebrating the 40th anniversary of the PDB was held at CSHL in 2011 (Berman et al., 2012; Burley, 2013) and was attended by past and present data curators and scientists who have contributed to the PDB, including some who were at the original CSHL meeting (Figure 2).

PDB Organization

When the PDB began in 1971, it was a small organization headquartered at BNL. Early on, distribution sites were established in Cambridge, UK, and in Osaka, Japan. In 1998, the Research Collaboratory for Structural Bioinformatics (RCSB) took over the management of the PDB from BNL (Berman et al., 2000). The European Bioinformatics Institute in the UK created the Macromolecular Structure Database (now Protein Data Bank in Europe, PDBe) (Velankar et al., 2012) for both deposition and distribution of PDB data. The Protein Data Bank Japan was established at Osaka University Protein Data Bank Japan (PDBj) (Kinjo et al., 2012). In 2003, in recognition of the global nature of the PDB, the wwPDB organization was established with RCSB PDB, PDBe, and PDBj as its founding members (Berman et al., 2003). In 2006, BioMagResBank (BMRB) joined wwPDB (Berman et al., 2007). The mission of the wwPDB is to ensure that the PDB archive will remain a single, global, uniform, and freely available archive. A Memorandum of Understanding was signed with guidelines for representing and processing the data using the same algorithms and procedures. To achieve this, the wwPDB collaborates on a variety of projects as described and made available at wwpdb.org (Table 1). In the interest of science, member organizations each continue to develop their own websites competitively with advanced tools and services (Table 1).

In addition to the Advisory Committees that report to each individual wwPDB member site, a wwPDB Advisory Committee (wwPDB AC) was established with representatives of each of the partner sites, community representatives for X-ray crystallography, NMR and electron microscopy as well as regional representatives from China and India. The committee meets yearly to review progress and policies related to the representation, deposition, processing, and validation of PDB entries.

Development of Guidelines for Structural Data Deposition and Release

Data deposition was optional when the PDB was established in the early 1970s. Even as the number of published articles describing structure determinations increased, some authors were hesitant to submit data. By the 1980s, when only a few hundred structures were available, some members of the community became very vocal about the need for mandatory deposition (Barinaga, 1989). Fred Richards at Yale circulated a petition (Figure 3) signed by many scientists demanding that the data created with public funds be made broadly and freely available. At the same time, the International Union for Crystallography (IUCr) and the ACA created committees to formulate guidelines for structure deposition. Among the matters discussed were what data should be deposited, when they should be deposited, and when they should be released. The people involved were experts in structural biology who were mindful of the needs of both the depositors and the users of the data. After several years of careful deliberation, a set of guidelines was published recommending data deposition in the PDB archive, and the release of data upon publication (with options of a 1-year hold for coordinates, and up to 4 years for structure factors) (International Union of Crystallography, 1989). That seminal publication set the stage for requirements adopted by funding agencies, and, most important, one by one the journals adopted the IUCr guidelines. Nowadays, a PDB identifier (and, increasingly, a wwPDB validation report) is required by virtually all journals that publish 3D structures and those structures (and the supporting experimental data) must be released to the public at the time of publication. In 2003, the PDB instituted a maximum of a 1-year hold limit after deposition. Following recommendations of the wwPDB Advisory Committee, deposition of X-ray structure factor data and NMR restraints was made mandatory in February 2008. In December 2010, deposition of NMR chemical shifts also became mandatory.

Methods for releasing data to the public have evolved along with technology (Berman, 2008). Coordinate data were originally deposited and distributed via postal mail using either magnetic tape or punched cards. In time, the entire archive was distributed on CD-ROM (and later DVD) to a list of subscribers. Early Internet servers (Gopher and FTP) could be used to access data entries, but users needed to know the IDs of the entries of interest. In the 1990s, websites made it possible for users to access data and, more important, to search for data using a variety of characteristics. The easy access of websites and search tools (Table 1) has opened up the usage of PDB data to a diverse community of researchers, students, and the general public. The discoverability and utility of PDB data are further enhanced through, usually reciprocal, cross-referencing with a large variety of chemical and biological information resources (Velankar et al., 2013).

PDB File Format

The original PDB file format was simple, human readable, and designed to utilize the IBM punched card format used to store data (Bernstein et al., 1977). The format outlived the devices that were used to read the physical punched cards. Over time, data items were added to better describe the structure determination, but the atom-site records (that contain the 3D coordinates) remained the same. Although the format was very popular and used by many people, it did present problems. Not all the data items were unambiguously described, and,

more important, the punched card format limited each line to 80 characters, thus restricting the number and names of atoms, residues, and chains that could be accommodated for a single entry. For example, a PDB format file can support a maximum of 99,999 atoms and 62 chains. Large and complex structures such as ribosomes require multiple PDB files to represent their numerous atoms and chains.

In 1990, the IUCr charged a committee headed by Paula Fitzgerald to come up with a better representation for macromolecular structures. A series of workshops was held, and in 1996 a new format called the Macromolecular Crystallographic Information File (mmCIF) was announced. The first mmCIF data dictionary contained more than 3,000 terms that could be used to describe a biological macromolecular structure and related experimental details (Fitzgerald et al., 2005). The format was machine-readable, the definitions were unambiguous, and there were no restrictions on macromolecule size. Although adoption of the new format by the community was limited, the wwPDB has continued to expand the mmCIF dictionary to include terms for NMR and 3DEM; the extended dictionary is named PDBx. With large structures becoming a common reality, it became clear to more and more members of the community that the PDB format was woefully inadequate, and there were calls for a change.

In 2011, the wwPDB convened a workshop (Figure 4A) at the EBI with developers of X-ray and NMR structure-determination software and charged this group to come up with a format that would be produced and output from their programs and used to submit and retrieve data from the PDB archive. Their choice was mmCIF/PDBx, in which the atom-site records follow a style that is similar to the more traditional PDB format (Figures 4B–4C). A working group was established and its members are implementing support for the mmCIF/PDBx format into their structure-determination packages, including CCP4 (Winn et al., 2011), Phenix (Adams et al., 2002), and Refmac (Murshudov et al., 2011). To facilitate the integration of structure data with other life sciences data resources, the wwPDB also provides data in formats compatible with semantic web technologies (Chen et al., 2013), such as XML (Westbrook et al., 2005) and Resource Description Framework (Kinjo et al., 2012) formats.

PDB Content

When the PDB was founded, the only structure-determination method used was X-ray crystallography. In the 1980s, NMR began to be used to determine structures, and in the 1990s the first electron-microscopy structures were deposited into the PDB. The mmCIF format was expanded to accommodate the terms necessary to describe all structure-determination methods and renamed PDBx as noted above. For a long time, the PDB also included theoretical models. In 2005, a workshop attended by both theorists and experimentalists was held to discuss whether this was a good practice. The outcome of that workshop was a recommendation that the PDB should contain only structures that were determined by experimental methods and that a separate portal should be developed to store *in silico* models. These recommendations were published in *Structure* and adopted by the wwPDB (Berman et al., 2006). Several years later, the Protein Modeling Portal was

established as a collaboration between the Protein Structure Initiative Structural Biology Knowledgebase and the Swiss Institute of Bioinformatics (Arnold et al., 2009).

Validation

An area of major concern has long been how best to validate the structures deposited and archived in the PDB (Brändén and Jones, 1990). Over the years, a set of tools developed by various members of the community was incorporated into the PDB processing pipelines. A few structures had to be obsoleted by their authors because of errors, and, of much greater concern, a few structures in the PDB appeared to be fraudulent. In 2008, the wwPDB convened an X-Ray Validation Task Force (VTF) chaired by Randy Read, whose members are acknowledged experts in the field of X-ray structure validation. The charge to the committee was to advise the wwPDB on the approaches it should implement so as to validate new depositions of structures based on X-ray crystallographic data. Over a period of about 2 years, the group investigated and discussed the best ways to validate crystal structures. As part of this process, the X-Ray VTF used a variety of methods to check the entire PDB archive and produced archive-wide statistics for a large number of quality indicators. An important outcome was the design of a graphic to illustrate the results of validation checks. Key metrics are represented by bars colored red (for worse) to blue (for better), and the values of the entry are positioned relative to entries in the entire PDB and to entries of similar resolution. This graphic makes it much easier to spot anomalies in a particular entry. The recommendations and detailed analyses were published in *Structure* (Read et al., 2011) and form the basis of a new validation suite being implemented by the wwPDB (Gore et al., 2012).

NMR is a much younger field, and the rules for validation are not as well established. A wwPDB NMR VTF chaired by Guy Montelione and Michael Nilges and consisting of experts in NMR as well as representatives from the X-Ray VTF was established in 2009. This NMR VTF has recently published its recommendations for the creation of validation standards for NMR-derived structures (Montelione et al., 2013; in this issue of *Structure*). Similarly, structures derived by 3D electron-microscopy methods are difficult to assess. The EMDatabank partners (PDBe, RCSB PDB, and National Center for Macromolecule Imaging) convened an EM VTF, chaired by Richard Henderson and Andrej Sali. The EM VTF has published a paper in *Structure* concluding that there is an urgent need to develop validation standards for both maps and models (Henderson et al., 2012). In recent years, structures have been submitted to the PDB that are either fully or partially derived from small-angle scattering data. A wwPDB Small Angle Scattering Task Force, chaired by Jill Trehwella, has begun its deliberations on how best to handle such models and recently published a set of recommendations (Trehwella et al., 2013).

The wwPDB Does Not Stand Alone

As this brief overview shows, the definition of the “PDB community” has evolved over time. In the beginning, it consisted mostly of X-ray crystallographers who needed a reliable place to store and retrieve data. Since then, the depositor community has expanded to include NMR spectroscopists and electron microscopists as well as scientists who use

combinations of methods to determine structures. Over time, the user community has expanded from the small group of activists who saw the need for an archive to do their research to an ever-larger group of users who far outnumber the depositors. Indeed, in a 2012 survey conducted by RCSB PDB, 75% of the ~1,000 respondents had never deposited a structure. These users included computational biologists and statisticians, pharmaceutical and medical scientists, and ecologists. The data-curation community has likewise expanded from a single expert annotator at BNL to groups of trained annotators and computer programmers with expertise in various aspects of structural biology and chemistry working at sites around the world. In the early days, curation policies were decided by a series of ad hoc committees that drew up guidelines that are the basis for today's policies. Nowadays, policies are decided by formally charged advisory committees drawn from the community of PDB users at each site, as well as the wwPDB AC, which has representation from the global community. Finally, to assemble the expertise needed to validate structures and to ensure high data quality, the wwPDB has constituted validation task forces consisting of members with deep knowledge of each of the structure-determination methods represented in the PDB archive. Members of the wwPDB also are involved in “community challenges” such as CASP (<http://predictioncenter.org/>), CAPRI (<http://pdbe.org/capri/>), and CASD-NMR (<http://www.wenmr.eu/wenmr/casd-nmr>). The partners of the wwPDB maintain electronic mailing lists, host events, and participate in professional society meetings in which lively discussions and debates are carried out. The tradition of community collaboration that gave birth to the PDB will no doubt continue and will ensure that the archive is responsive to the science it contains as well as the new science it enables.

ACKNOWLEDGMENTS

The wwPDB thanks staff past and present and all members of the PDB community.

RCSB PDB is supported by NSF DBI 0829586, NIGMS, DOE, NLM, NCI, NINDS, NIDDK; PDBe by EMBL-EBI, Wellcome Trust (088944), BBSRC (BB/J007471/1, BB/I02576X/1, and BB/K016970/1), NIGMS (1R01 GM079429-01A1), and EU (284209); PDBj by NBDC-JST; and BMRB by NLM P41 LM05799.

REFERENCES

- Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 2002; 58:1948–1954. [PubMed: 12393927]
- Arnold K, Kiefer F, Kopp J, Battey JN, Podvinec M, Westbrook JD, Berman HM, Bordoli L, Schwede T. The Protein Model Portal. *J. Struct. Funct. Genomics.* 2009; 10:1–8. [PubMed: 19037750]
- Barinaga M. The missing crystallography data. *Science.* 1989; 245:1179–1181. [PubMed: 2781276]
- Berman HM. The Protein Data Bank: a historical perspective. *Acta Crystallogr. A.* 2008; 64:88–95. [PubMed: 18156675]
- Berman HM, Westbrook JD, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
- Berman HM, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* 2003; 10:980. [PubMed: 14634627]
- Berman HM, Burley SK, Chiu W, Sali A, Adzhubei A, Bourne PE, Bryant SH, Dunbrack RL Jr. Fidelis K, Frank J, et al. Outcome of a workshop on archiving structural models of biological macromolecules. *Structure.* 2006; 14:1211–1217. [PubMed: 16955948]

- Berman HM, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* 2007; 35(Database issue):D301–D303. [PubMed: 17142228]
- Berman HM, Kleywegt GJ, Nakamura H, Markley JL. The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure.* 2012; 20:391–396. [PubMed: 22404998]
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977; 112:535–542. [PubMed: 875032]
- Braändén C-I, Jones TA. Between objectivity and subjectivity. *Nature.* 1990; 343:687–689.
- Burley SK. PDB40: The Protein Data Bank celebrates its 40th birthday. *Biopolymers.* 2013; 99:165–169. [PubMed: 23280389]
- Chen H, Yu T, Chen JY. Semantic Web meets Integrative Biology: a survey. *Brief. Bioinform.* 2013; 14:109–125. [PubMed: 22492191]
- Cold Spring Laboratory Press. *Cold Spring Harb. Symp. Quant. Biol.* 1972; 36
- Fitzgerald, PMD.; Westbrook, JD.; Bourne, PE.; McMahon, B.; Watenpugh, KD.; Berman, HM. 4.5 Macromolecular dictionary (mmCIF). In: Hall, SR.; McMahon, B., editors. *International Tables for Crystallography G Definition and exchange of crystallographic data.* Springer; Dordrecht, The Netherlands: 2005. p. 295–443.
- Gore S, Velankar S, Kleywegt GJ. Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* 2012; 68:478–483. [PubMed: 22505268]
- Gutmanas A, Oldfield TJ, Patwardhan A, Sen S, Velankar S, Kleywegt GJ. The role of structural bioinformatics resources in the era of integrative structural biology. *Acta Crystallogr. D Biol. Crystallogr.* 2013; 69:710–721. [PubMed: 23633580]
- Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, et al. Outcome of the first electron microscopy validation task force meeting. *Structure.* 2012; 20:205–214. [PubMed: 22325770]
- Hopper P, Harrison SC, Sauer RT. Structure of tomato bushy stunt virus. V. Coat protein sequence determination and its structural implications. *J. Mol. Biol.* 1984; 177:701–713. [PubMed: 6481803]
- International Union of Crystallography Commission on Biological Macromolecules. Policy on publication and the deposition of data from crystallographic studies of biological macromolecules. *Acta Crystallogr. A.* 1989; 45:658.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* 1958; 181:662–666. [PubMed: 13517261]
- Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution. *Nature.* 1960; 185:422–427. [PubMed: 18990802]
- Kinjo AR, Suzuki H, Yamashita R, Ikegawa Y, Kudou T, Igarashi R, Kengaku Y, Cho H, Standley DM, Nakagawa A, Nakamura H. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res.* 2012; 40(Database issue):D453–D460. [PubMed: 21976737]
- Levinthal C. Are there pathways for protein folding? *Extrait du Journal de Chimie Physique.* 1968; 65:44–45.
- Montelione GT, Nilges M, Bax A, Güntert P, Herrmann T, Markley JL, Richardson J, Schwieters C, Vuister GW, Vranken W, et al. Recommendations of the wwPDB NMR Structure Validation Task Force. *Structure.* 2013; 21(this issue):1563–1570. [PubMed: 24010715]
- Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 2011; 67:355–367. [PubMed: 21460454]
- Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North ACT. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature.* 1960; 185:416–422. [PubMed: 18990801]
- Protein Data Bank. Escape from lysosomes. *Nat. New Biol.* 1971; 233:223. [PubMed: 20480989]

- Read RJ, Adams PD, Arendall WB 3rd, Brunger AT, Emsley P, Joosten RP, Kleywegt GJ, Krissinel EB, Lütke T, Otwinowski Z, et al. A new generation of crystallographic validation tools for the protein data bank. *Structure*. 2011; 19:1395–1412. [PubMed: 22000512]
- Trehwella J, Hendrickson WA, Sato M, Schwede T, Svergun D, Tainer JA, Westbrook J, Kleywegt GJ, Berman HM. Report of the wwPDB Small-Angle Scattering Task Force. Data Requirements for Biomolecular Modeling and the PDB. *Structure*. 2013; 21:875–881. [PubMed: 23747111]
- Velankar S, Alhroub Y, Best C, Caboche S, Conroy MJ, Dana JM, Fernandez Montecelo MA, van Ginkel G, Golovin A, Gore SP, et al. PDBe: Protein Data Bank in Europe. *Nucleic Acids Res*. 2012; 40(Database issue):D445–D452. [PubMed: 22110033]
- Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin MJ, Kleywegt GJ. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res*. 2013; 41(Database issue):D483–D489. [PubMed: 23203869]
- Westbrook JD, Ito N, Nakamura H, Henrick K, Berman HM. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*. 2005; 21:988–992. [PubMed: 15509603]
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr*. 2011; 67:235–242. [PubMed: 21460441]

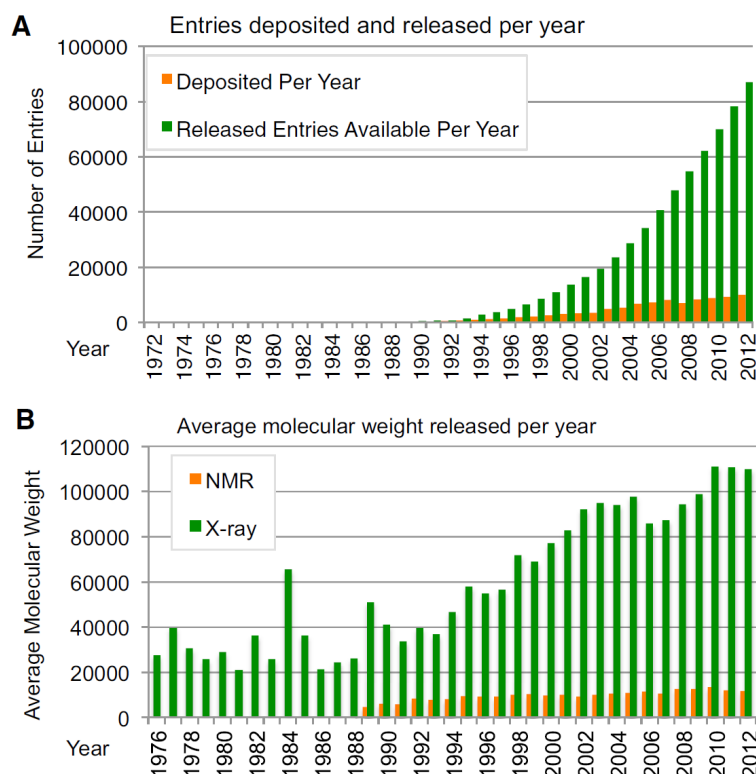


Figure 1. Growth of the PDB Archive

(A) The number of entries deposited (orange) and the total number of entries available in the archive (green) each year.

(B) Average molecular weight of entries in the archive from NMR (orange) and X-ray (for the asymmetric unit, green) experiments, by year of entry release. The spike seen in 1984 is the result of the release of a tomato bushy stunt virus structure (2TBV) (Hopper et al., 1984).



Figure 2. Past and Present PDB People, in Attendance at PDB40

Front row, from left to right: Francis Bernstein, Martha Quesada, Gerard J. Kleywegt, Tom Koetzle, Helen M. Berman, Haruki Nakamura, John Markley, Miri Hirshberg, and Joel Sussmann. Second row: Judith L. Flippen-Anderson, Peter Rose, Gary L. Gilliland, T.N. Bhat, Jasmine Young, Buvanewari Coimbatore Narayanan, Monica Sekharan, Irina Persikova, and Sutapa Ghosh. Third row: Spencer Bliven, Shuchismita Dutta, Guanghua Gao, Zukang Feng, Tom Oldfield, David Micallef, Luigi Di Costanzo, Catherine L. Lawson, Sanchayita Sen, Christine Zardecki, and Chisa Kamada. Fourth row: Wolfgang F. Bluhm, Chunxiao Bi, Chenghua Shao, Dimitris Dimitropoulos, Andreas Prli , Geoffrey Barton, Sameer Velankar, Brian Hudson, Vladimir Guranovi , John Westbrook, Philip E. Bourne. Back row: Margaret Gabanyi, Eldon Ulrich, Genji Kurisu, Atsushi Nakagawa, Nomi Ron, Lihua Tan, Maria Voigt, Huanwang Yang, Rachel Kramer Green, Greg Quinn, and David S. Goodsell.

Photo by Constance Brukin.

Yale University



Department of Molecular Biophysics
and Biochemistry
260 Whitney Avenue
P.O. Box 6668
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
Frederic M. Richards

Richard E. Dickerson
Richard E. Dickerson

Jane S. Richardson
Jane S. Richardson

Michael G. Rossmann
Michael G. Rossmann

David C. Richardson
David C. Richardson

David R. Davies
David R. Davies

Joseph Kraut
Joseph Kraut

Don C. Wiley
Don C. Wiley

Stephen C. Harrison
Stephen C. Harrison

Figure 3. Frederic M. Richards Initiated This Petition to Develop Deposition Policies for PDB Data

The full petition was signed by hundreds of scientists and sent to various journals.

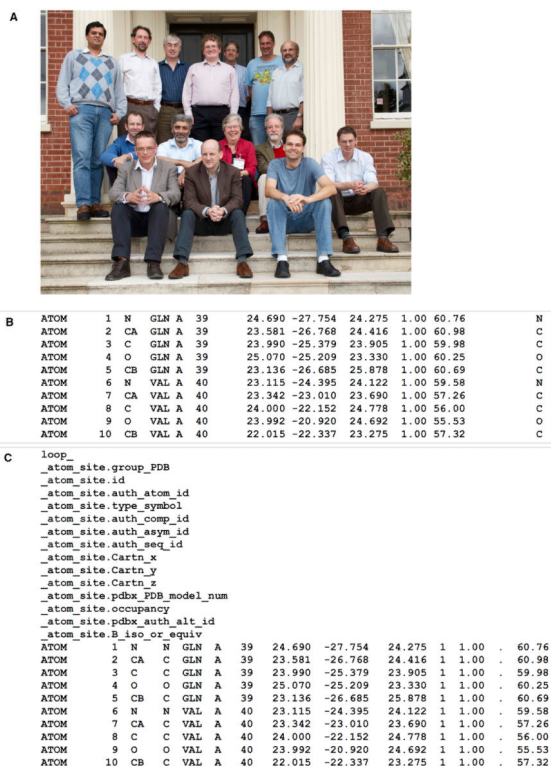


Figure 4. Participants in the 2011 wwPDB Format Workshop in Cambridge, UK and Comparison of File Formats

(A) Workshop participants: Top row, from left to right: Sameer Velankar, Paul Adams, Randy Read, Thomas Womack, John Westbrook, Gerard Kleywegt, and Eugene Krissinel. Middle row: John Ionides, Garib Murshudov, Helen Berman, Gérard Bricogne, and Tom Oldfield. Bottom row: Oliver Smart, Paul Emsley, and Ralf Grosse-Kunstleve.

(B and C) PDB and PDBx/mmCIF file formats for representing atomic coordinate data.

Table 1

The Worldwide Protein Data Bank Areas of Collaboration between the wwPDB Partners

Area of Collaboration	Examples
Policy issues	definition of mandatory items and data for deposition, e.g., structure-factor amplitudes
Archive releases	weekly updates exchanged between all sites; simultaneous release of identical copies of the archive
Validation standards	validation task forces provide recommendations that the wwPDB partners implement
Format specifications	PDBx/mmCIF, PDBML, and PDB
Chemical component descriptions	detailed description of new chemical components, including ideal coordinates
Deposition and annotation procedures	agreement on common procedures to describe, for example, quaternary structure or Ramachandran outliers, or the use of reference resources, e.g., for sequence annotation
Archive quality and remediation	regular review of quality, consistency, and integrity issues leading to large-scale archive-wide remediation
Journal interactions	recommendations with respect to wording of deposition and release requirements; coordination of publications and data release

Reprinted from Gutmanas et al. (2013).

Each wwPDB partner develops separate websites hosting advanced tools and services that utilize PDB data: RCSB PDB (<http://rcsb.org>), PDBe (<http://pdbe.org>), PDBj (<http://pdbj.org>), and BMRB (<http://www.bmrwisc.edu>).