

PDBjとwwPDBの活動について

栗栖源嗣

大阪大学蛋白質研究所



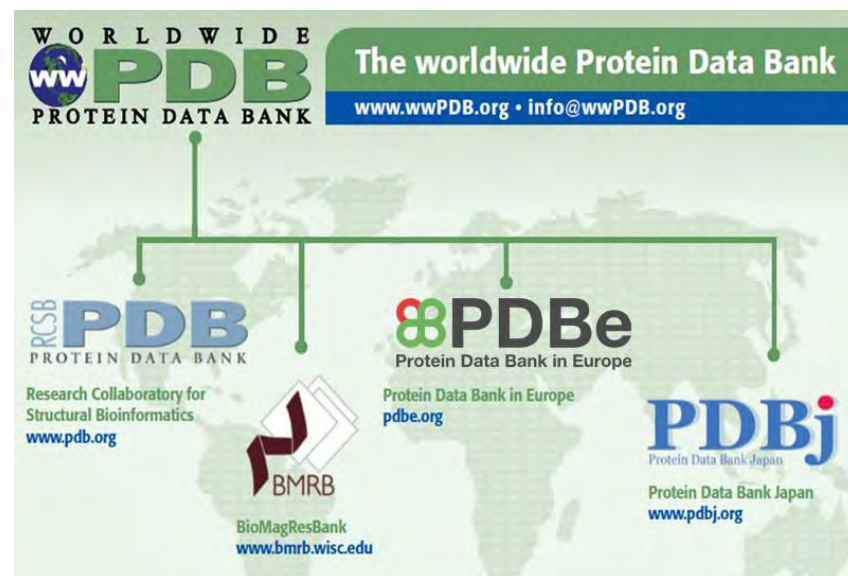
wwpdb.org



Protein Data Bank Japan

<http://pd bj.org/>

Since 2001, PDBj has been managed at **Institute for Protein Research, Osaka University** as a member of the **wwPDB**, to curate and process the deposited data for an open and single archive.



PDBj Publication

Nucleic Acids Research Advance Access published October 26, 2016

Nucleic Acids Research, 2016 **1**
doi: 10.1093/nar/gkw962

Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures

Akira R. Kinjo*, Gert-Jan Bekker, Hirofumi Suzuki, Yuko Tsuchiya, Takeshi Kawabata, Yasuyo Ikegawa and Haruki Nakamura

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

Received September 15, 2016; Revised October 06, 2016; Editorial Decision October 07, 2016; Accepted October 11, 2016

ABSTRACT

The Protein Data Bank Japan (PDBj, <http://pdbj.org>), a member of the worldwide Protein Data Bank (wwPDB), accepts and processes the deposited data of experimentally determined macromolecular structures. While maintaining the archive in collaboration with other wwPDB partners, PDBj also provides a wide range of services and tools for analyzing structures and functions of proteins. We herein outline the updated web user interfaces together with RESTful web services and the backend relational database that support the former. To enhance the interoperability of the PDB data, we have previously developed PDB/RDF, PDB data in the Resource Description Framework (RDF) format, which is now a wwPDB standard called wwPDB/RDF. We have enhanced the connectivity of the wwPDB/RDF data by incorporating various external data resources. Services for searching, comparing and analyzing the ever-increasing large structures determined by hybrid methods are also described.

major changes regarding user interfaces and analysis tools as well as additional data provided. The previously described Resource Description Framework (RDF) format, PDB/RDF, is now one of the wwPDB standard formats called wwPDB/RDF and is enhanced with supplementary information in order to connect PDB data with other biological data resources.

USER INTERFACES

User interfaces include interactive (and graphical) web interfaces for humans and RESTful web services for computer programs. We also expose our backend database in the forms of web services or dump files for enabling very complex queries. These are described in turn.

Web interface

The web interface of PDBj was updated to provide a uniform integrated interface for the available services as well as to provide a scalable interface for devices ranging from smartphones to workstations. This update incorporates several innovative/renovative features as described below.

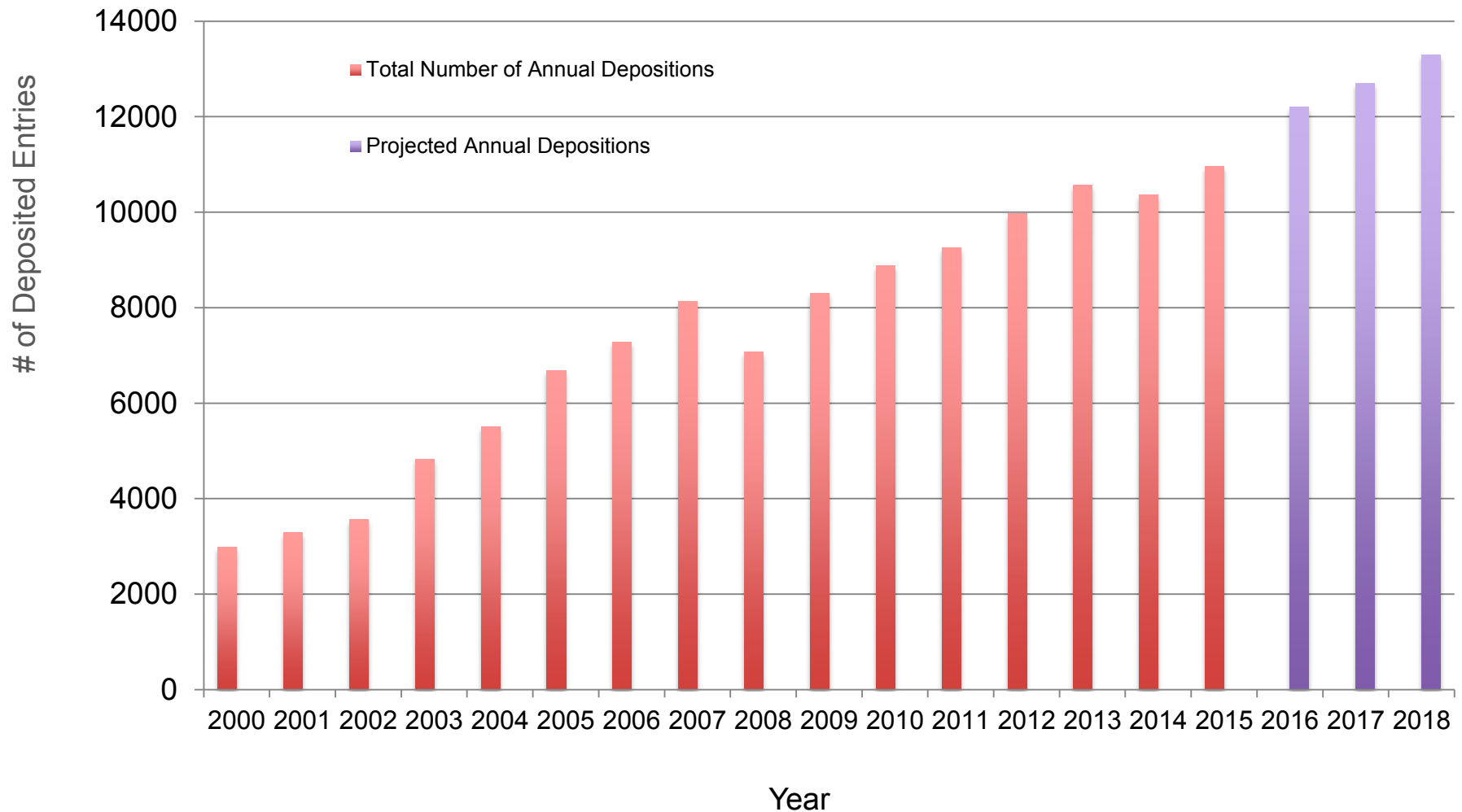
We have implemented various functionalities to ease

Downloaded from <http://nar.oxfordjournals.org/> at Osaka Daigaku Ningensho on October 26, 2016

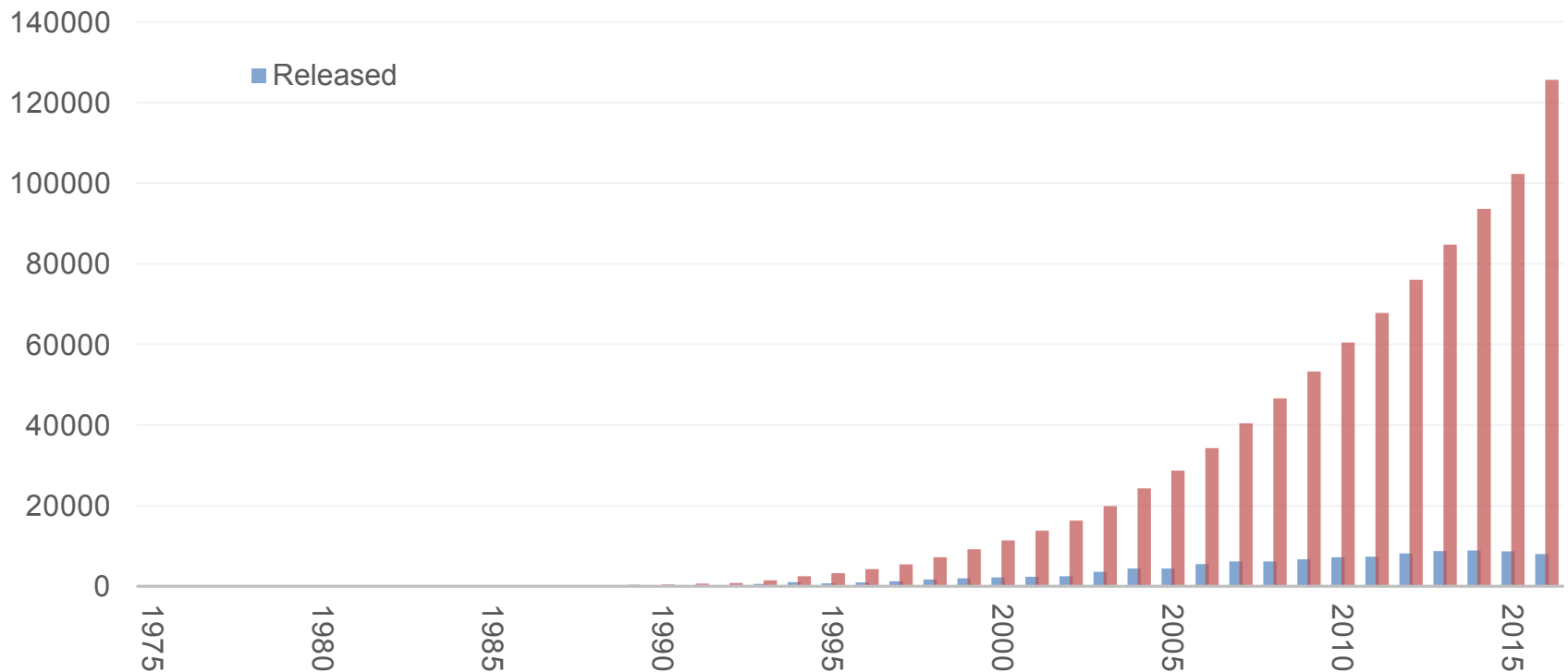
Activities/Services of each member of the wwPDB

- **“Data-in”** activity, common in all the wwPDB members with high quality control. For that purpose, new format, data deposition, and validation system are developed
- “Data-out” services, common archive as the ftp site and the characteristic services by each wwPDB member

Growing Number of Depositions



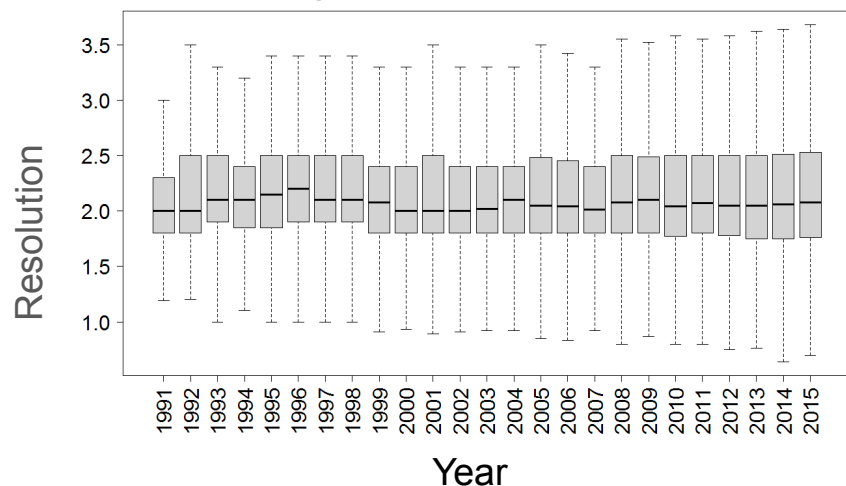
Growth of PX entries



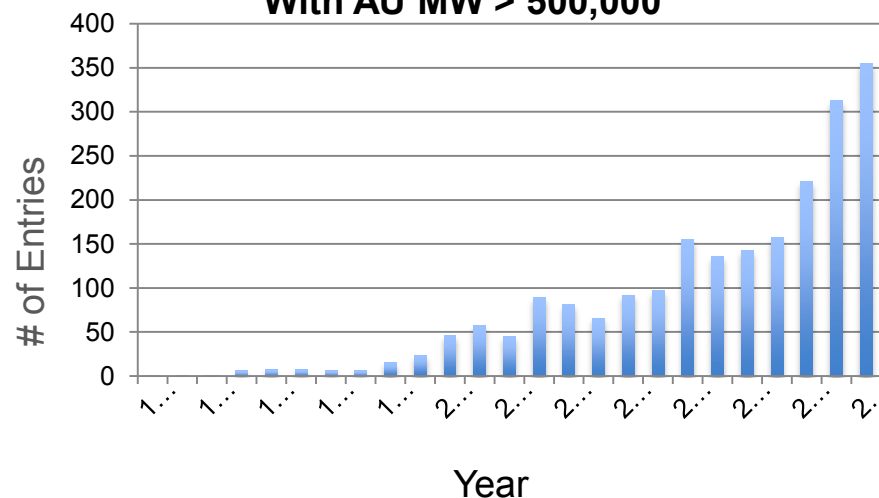
>9,000 New PX Entries Projected for Calendar 2016

Growing Complexity in PX Deposits

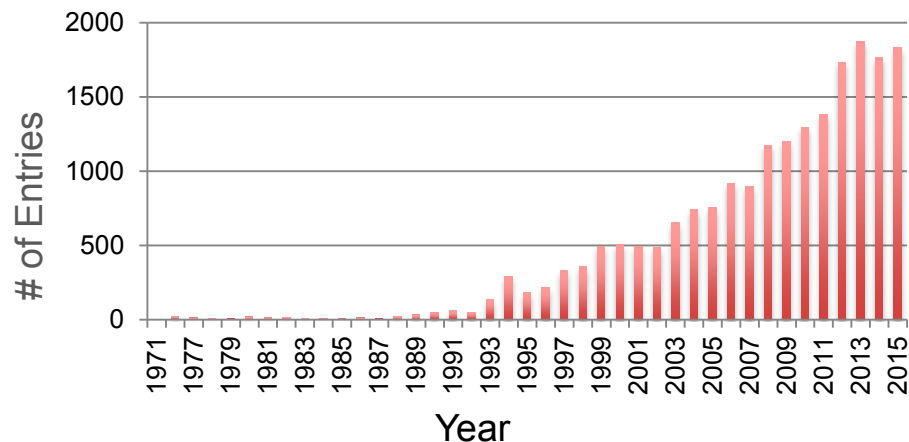
Annual Distribution for High Resolution Limit



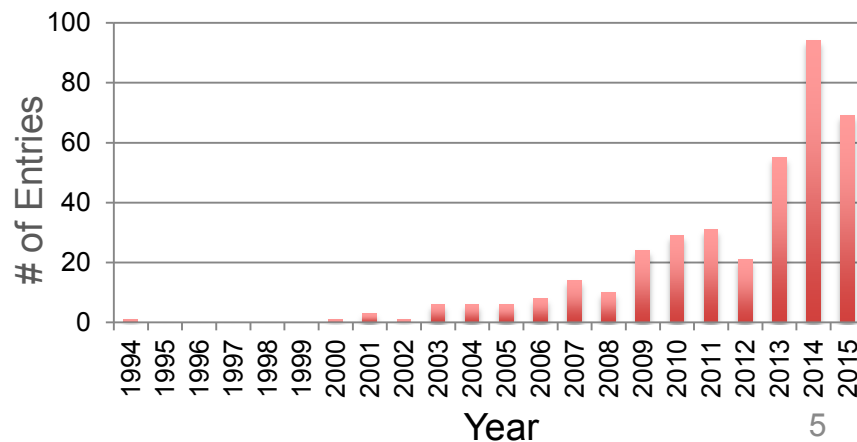
Annual Released Structures With AU MW > 500,000



Total Number of New CCD Entries



Annual Released Large Structures (chains > 62 & atoms > 99999)



Rebranding of D&A → OneDep

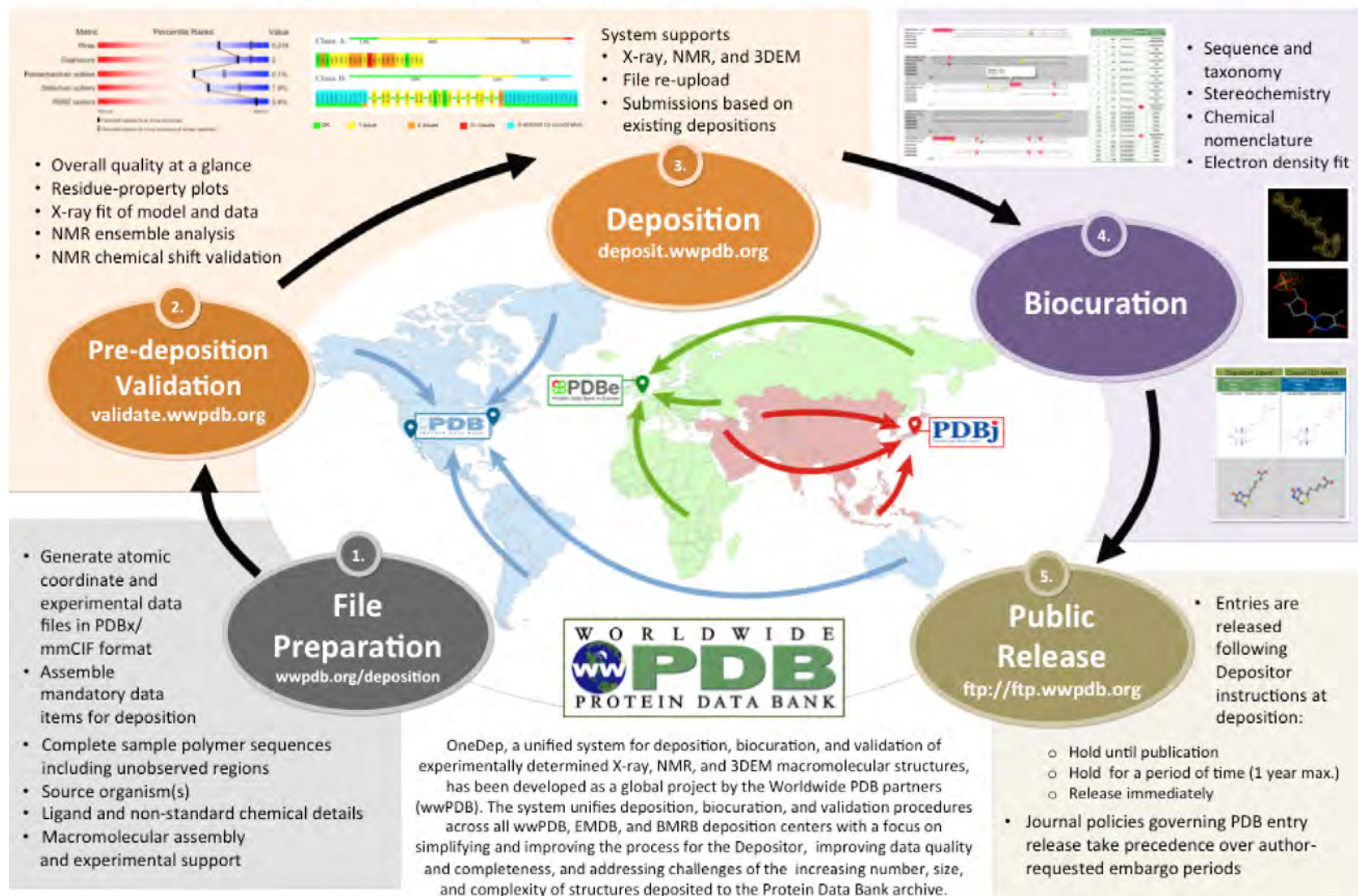
- Rebranding Process
 - Project Team nominated names/logos
 - Project Leadership recommended
 - wwPDB Pls made final selection
- OneDep logo



→ 詳細はこの後の中川敦史先生のご講演
で

OneDep: Unified Deposition Portal for the Protein Data Bank

wwPDB Partners - RCSB PDB, PDBe, PDBj, and BMRB



Acknowledgements



wwPDB X-ray VTF 2.0 Meeting

Nov 16-17, 2015 at EMBL-EBI

- Ligands (wwPDB Ligand Validation Workshop, feedback, density-fit analysis and display, Buster Report and ligands, Radiation damage effects, Metal validation, Carbohydrate issues)
- Proteins (MolProbity, Cis-peptides, HNQ flips, Clashes and false positives)
- X-ray-specific (Xtrriage update, serial crystallography, NCS)
- wwPDB issues (pipeline, reports, metadata, annotation, prioritization)



VTF Members: Paul Adams, Gérard Bricogne, Dave Brown, Paul Emsley, Richard Henderson, Nobutoshi Ito, Robbie Joosten, Thomas Lütkeke, Michael Nilges, Arwen Pearson, Tassos Perrakis, Randy Read (Chair), Jane Richardson, Janet Smith, Tom Terwilliger, Ian Tickle, Gert Vriend

wwPDB Attendees: Burley, Feng, Gutmanas, Velankar, Westbrook

Ligand Validation Workshop White Paper Published in 2016

- Adams *et al.* (2016) *Structure* 24, 502-508.
- 57 co-authors from 42 institutions/organizations
- Recommendations endorsed unanimously by wwPDB X-ray VTF 2.0

Cell^{press}

Structure
Meeting Report

Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop

Paul D. Adams,¹ Kathleen Aertgeerts,² Cary Bauer,³ Jeffrey A. Bell,⁴ Helen M. Berman,^{5,6} Talapady N. Bhat,⁷ Jeff M. Blaney,⁸ Evan Bolton,⁹ Gerard Bricogne,¹⁰ David Brown,^{11,12} Stephen K. Burley,^{5,6,13,*} David A. Case,⁶ Kirk L. Clark,¹⁴ Tom Darden,¹⁵ Paul Emsley,¹⁶ Victoria A. Feher,^{17,*} Zukang Feng,^{5,6} Colin R. Groom,^{18,*} Seth F. Harris,⁸ Jorg Hendle,¹⁹ Thomas Holder,⁴ Andrzej Joachimiak,²⁰ Gerard J. Kleywegt,²¹

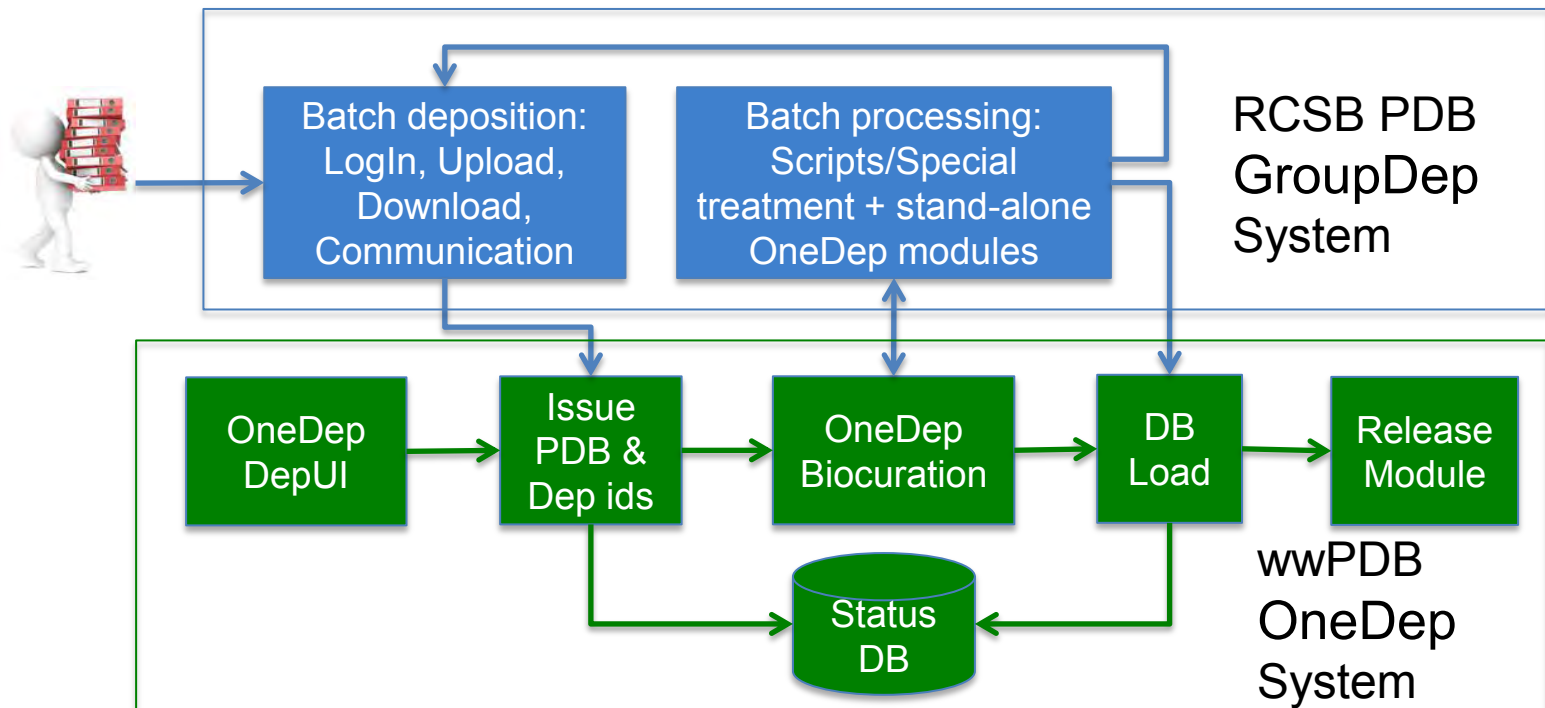
(Author list continued on next page)

Impact of 2-Stage PDB Data Release

- Every Saturday by 3:00 UTC the wwPDB website provides the following for every new entry stage for Wednesday release:
 - Sequence/s (amino acid or nucleotide) for each distinct polymer
 - Where appropriate, InChI string(s) for each distinct ligand and crystallization pH value(s)
- Support/Statistics for CAMEO
 - 4066 targets: 26 predictors for protein structure prediction
 - 14200 targets: 5 predictors for ligand binding
- Support/Statistics for CAPRI, CASP, and D3R
 - CAPRI: 11 targets: 41 teams
 - CASP: 134 targets: 221 groups registered, 16099 models
 - D3R: Blinded Challenges predicting docking pose/binding affinity for 2 targets/211 compounds; Weekly CELPP challenge coming

Enabling Depositions from Industry

- Group Deposition processing
 - Requirements set by wwPDB OneDep Team
 - Provided support for D3R Blind Challenges



Activities/Services of each member of the wwPDB

- “Data-in” activity, common in all the wwPDB members with high quality control. For that purpose, new format, data deposition, and validation system are developed
- “**Data-out**” services, common archive as the ftp site and the characteristic services by each wwPDB member

Download Statistics

Year	Total	Total FTP Archive	Total Website	RCSB PDB FTP Archive	RCSB PDB Website	PDBe FTP Archive	PDBe Website	PDBj FTP Archive	PDBj Website
2009	328,362,536	271,116,934	57,245,602	222,984,760	53,507,785	30,141,339	1,475,116	17,990,835	2,262,701
2010	294,326,976	213,180,966	81,146,010	159,248,214	64,569,658	34,383,219	14,017,349	19,549,533	2,559,003
2011	383,131,048	276,952,286	106,178,762	204,939,406	81,560,098	40,960,368	18,515,245	31,052,512	6,103,419
2012	376,944,070	255,837,735	121,106,335	213,510,347	90,438,501	21,601,103	23,982,801	20,726,285	6,685,033
2013	441,262,210	296,176,290	145,085,920	215,331,908	97,549,580	43,684,850	37,762,496	37,159,532	9,773,844
2014	512,227,251	339,193,721	173,033,530	237,168,615	110,115,316	52,362,370	48,031,414	49,662,736	14,886,800
2015	534,339,871	368,244,766	166,095,105	255,346,630	111,802,897	48,544,330	41,127,219	64,353,806	13,164,989

More than 1.5 million / day



Geographic origins of FTP downloads, 2012-2015

Collection of ORCID IDs

- Successfully Implemented Apr 11, 2016
- Metrics (Apr 11 – Aug 31, 2016):
 - ~8% of Depositions have ORCID ID (374/4713)
 - 170 unique ORCID IDs (92 identified as PIs)
- Plans to Increase ORCID Adoption
 - Expand to all entry authors to provide ORCID (2017)
 - Distribute collected ORCID IDs at ftp archive (2017)
 - Mandatory going forward (2018)

File Versioning: Objectives

Current Issues:

- Loss of connection between PDB ID and Publication under current wwPDB Obsolete/Supersede Policy
- Current wwPDB Policy represents a non-trivial barrier to revisions by the Depositor of Record

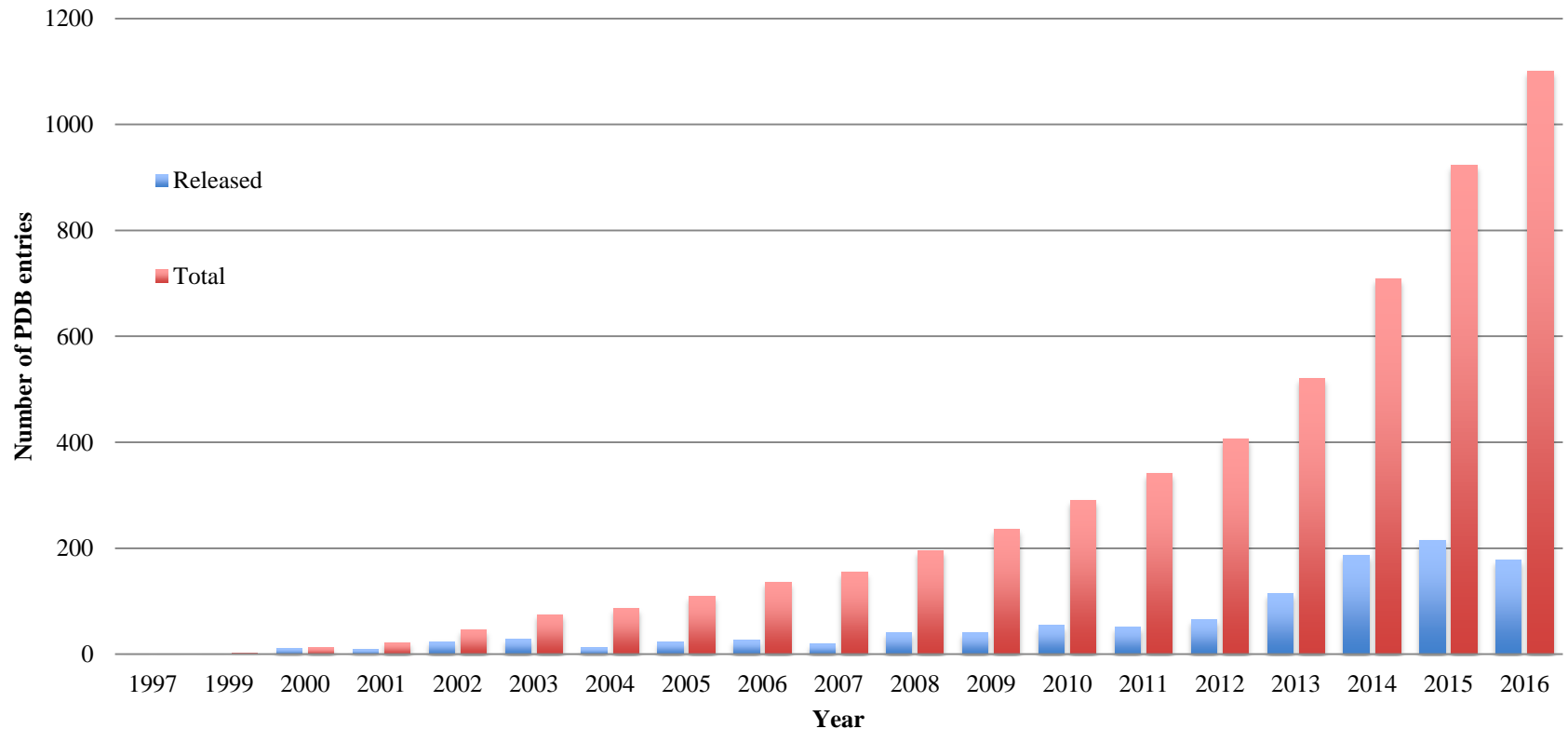
Objectives:

- Introduce new procedure to manage revision of atomic coordinates by the Depositor of Record
- Establish a robust extensible framework for versioning of all archival data

File Versioning: Planning Process

- User feedback solicited
- Enable revisions to entries updated by the Depositor of Record (e.g., Version 1-0 → 1-1; 1-0 → 2-0)
 - wwPDB will NOT assign a new PDB ID going forward (for Depositor of Record revision only)
- Introduce new PDB ID code format
 - Allow more informative and transparent delivery of revised data files
 - With PDB prefix and extension of 4 characters (e.g., from “1ABC” to “PDB_00001ABC”)
- Example: PDB_00001ABC_XYZ_V2-2.cif.gz

Growth of PDB EM Entries

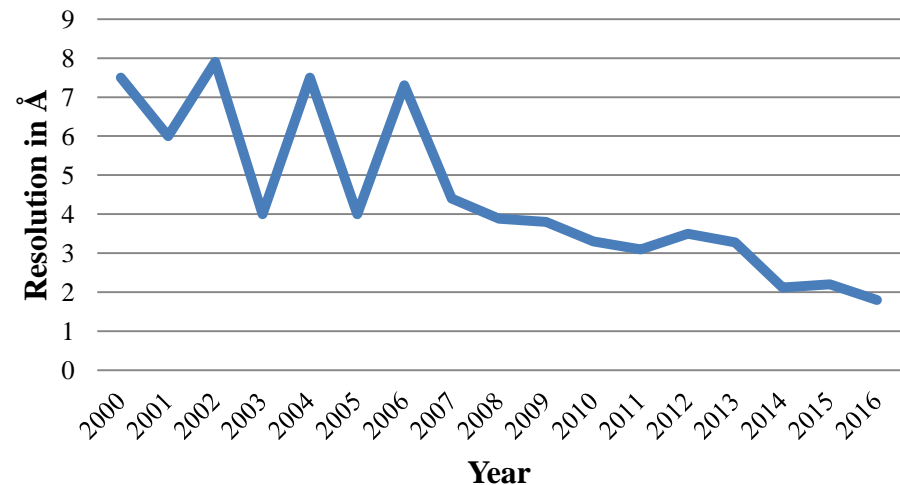
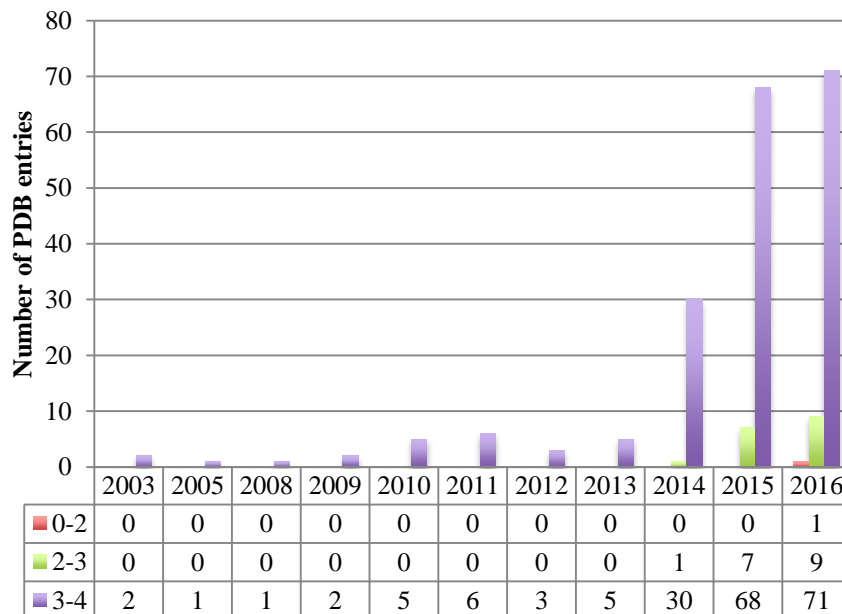


As of August 1, 2016, >1100 EM entries in the PDB archive

178 new entries released Jan 1 - Aug 1, 2016

“Resolution Revolution”

- 1.8Å structure in 2016 (PDB ID 5K12; EMD-8194)
- Increasing number of 3DEM structures at 2-4Å resolution (75 in calendar 2015 and 80 in first 7 months of 2016)



New wwPDB Policy for 3DEM Data

- Effective Sep 6, 2016, deposition of atomic models determined by 3DEM to the PDB requires prior or simultaneous deposition of the associated 3DEM mass density maps to EMDB
- For joint PDB/EMDB depositions, the hold period is the same for both map(s) and model(s)

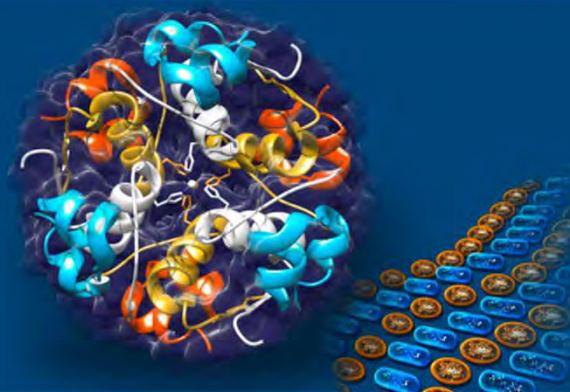
wwPDB Foundation Progress



HOME EVENTS SPONSORS AND DONATIONS BOARD

The Worldwide Protein Data Bank Foundation supports the **outreach activities of the wwPDB** that are crucial to the future of the PDB archive, including workshops, symposia, and advisory meetings.

SUPPORT US



About Us

The wwPDB Foundation was established in 2010 to raise funds in support of the outreach activities of the wwPDB. The Foundation has raised funds to help support PDB40, a symposium celebrating the 40th anniversary of the archive; workshops; and educational publications.

The Foundation is chartered as a 501(c)(3) entity exclusively for scientific, literary, charitable, and educational purposes.

Individual and institutional donations to the wwPDB are critical to the future of the PDB archive.

The Protein Data Bank Archive



Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The worldwide Protein Data Bank



The **Worldwide PDB (wwPDB)** organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

wwPDB data centers serve as deposition, annotation, and distribution sites of the PDB archive. Each site offers tools for searching, visualizing, and analyzing PDB data.

- Website released
- Fundraising on-going
- 2016 Events
 - OneDep Summit
 - Economics and Impact of the Protein Data Bank (PDB) Archive
- HFSP Meeting on Sustainability



SciDataCon 2016

<http://foundation.wwpdb.org/>