



PDBjとwwPDBの最近の活動について (Recent activities of PDBj and wwPDB)

栗栖源嗣 (Genji Kurisu)

大阪大学蛋白質研究所

(Institute for Protein Research, Osaka Univ.)

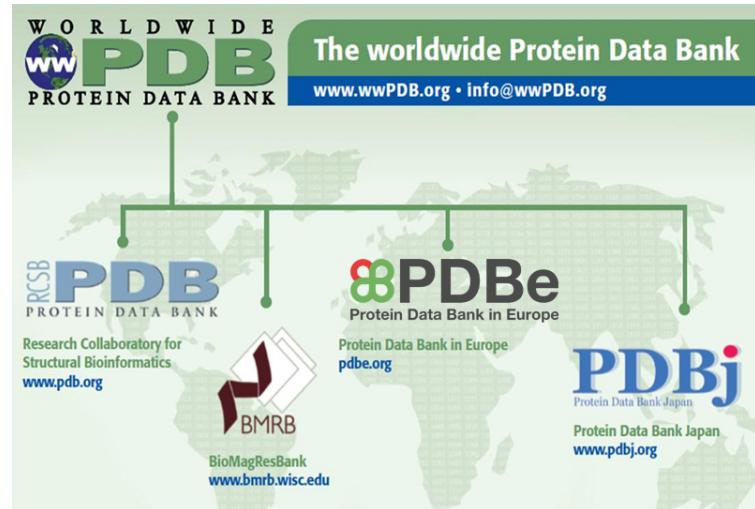


wwpdb.org



生体高分子の3次元構造(形)に関する情報を集めた国際 的データベース

1971年からのデータが集積され、情報は無償で利用できる。運営は各国(米国、欧州、日本)の政府機関による研究費用でまかなわれている。2003年からは国際組織wwPDBとして活動している(PDBjは創立メンバー)。



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.

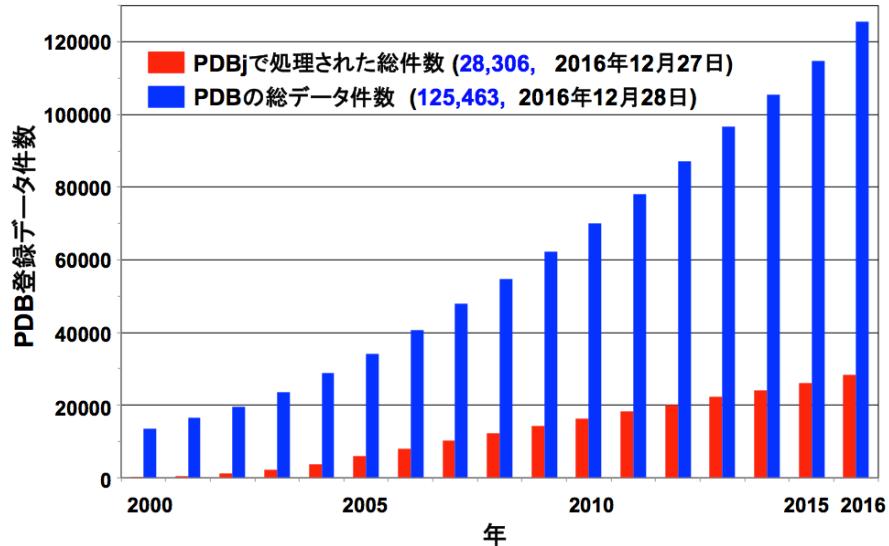
Web interface

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

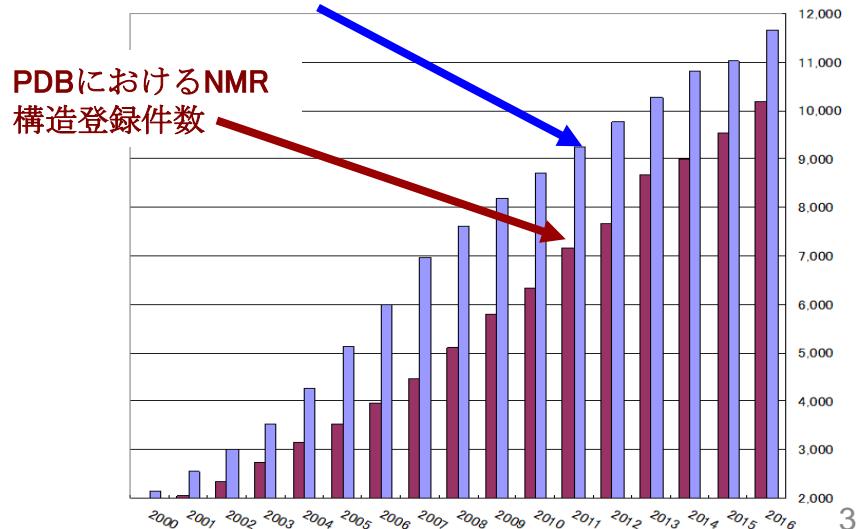
PDBjの活動

- **Data-inの活動:**
wPDBの一員として品質管理をしつつ登録作業を実施
新たな標準フォーマット等の開発(PDB/RDF, BMRB/RDF)
- **Data-out の活動:**
共通データのダウンロードサイト(毎週アプデート)の運営
関連DBとの統合化や二次データベース・ツールの開発



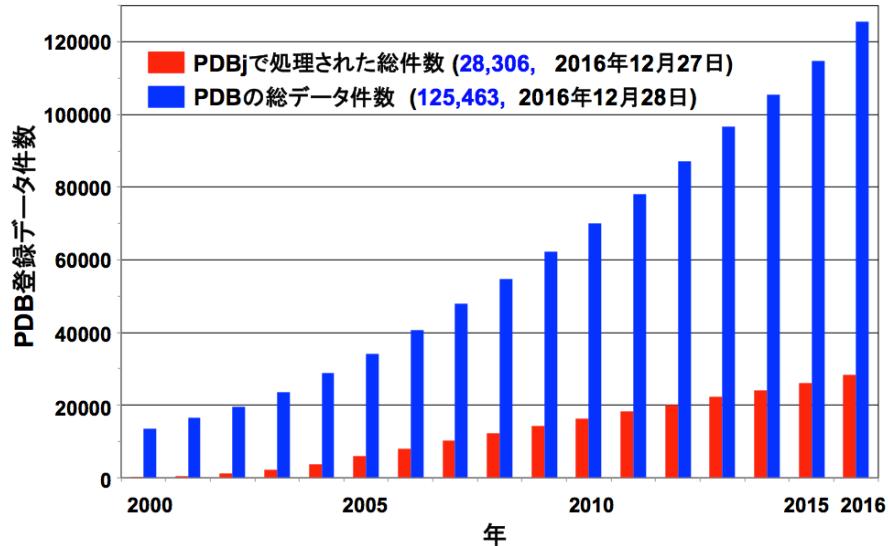
BMRB(化学シフト), PDB(原子座標)の登録総数の推移

BMRB化学シフト登録件数 (10,250, 2016年12月30日)



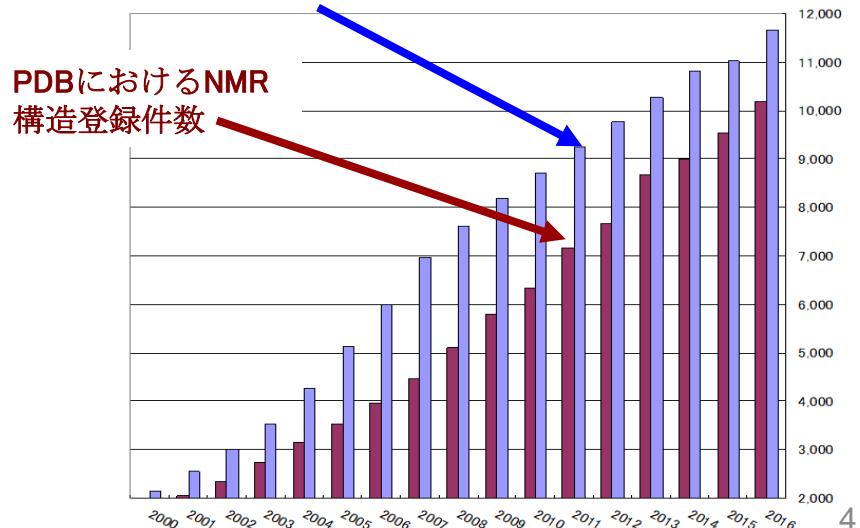
PDBjの活動

- **Data-inの活動:**
wwPDBの一員として品質管理をしつつ登録作業を実施
新たな標準フォーマット等の開発(PDB/RDF, BMRB/RDF)
- **Data-out の活動:**
共通データのダウンロードサイト(毎週アプデート)の運営
関連DBとの統合化や二次データベース・ツールの開発

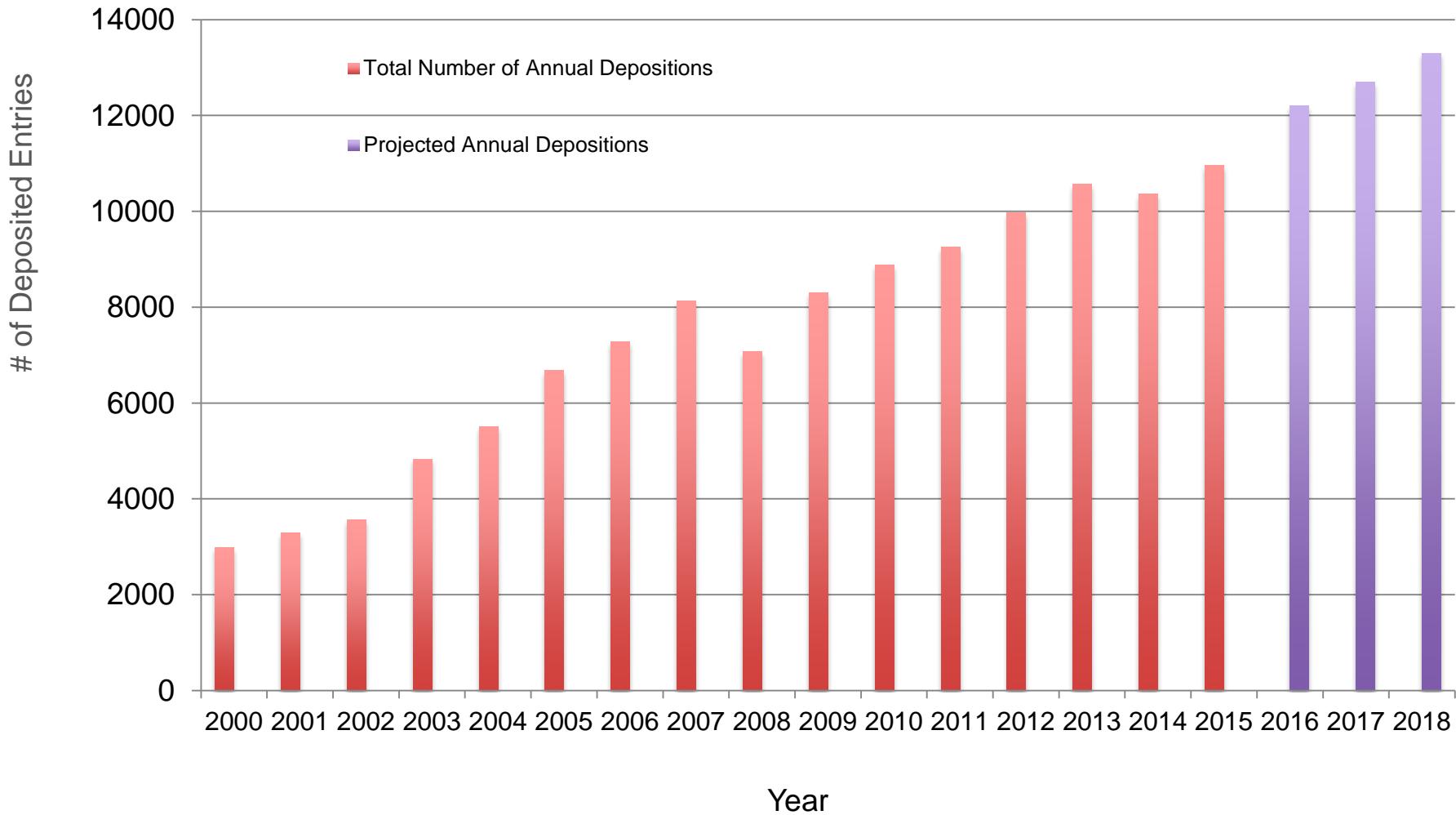


BMRB(化学シフト), PDB(原子座標)の登録総数の推移

BMRB化学シフト登録件数 (10,250, 2016年12月30日)



増え続けるPDB登録数

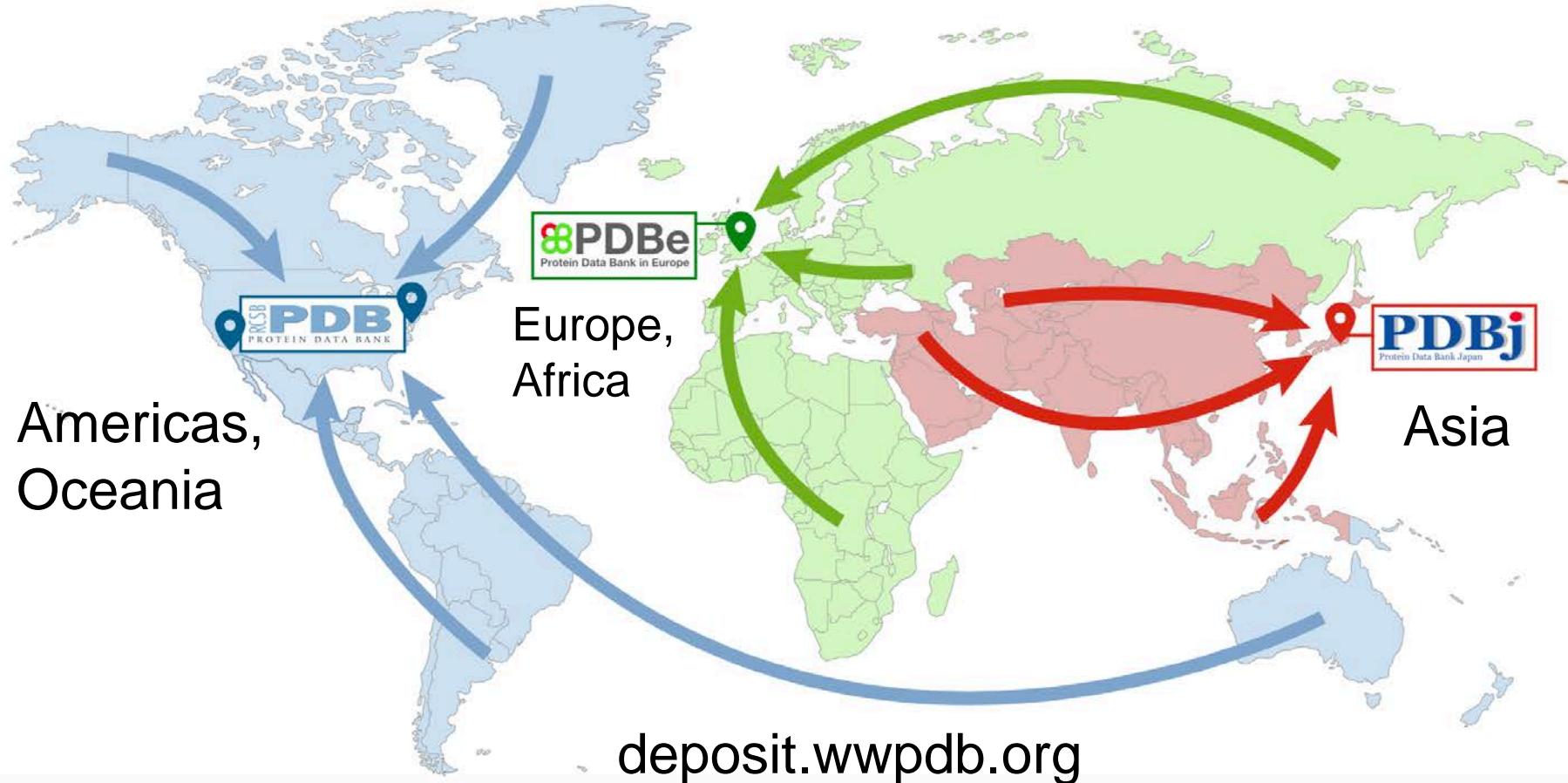


共通登録サイト導入による効率化

- wwPDB datacenterの共通登録サイト
 - X線, NMR, 電子顕微鏡の各手法で決定した座標を全て取り扱う
 - 実験データ(構造因子, 化学シフトと距離束縛情報, EMマップ)の登録も行う
- OneDep logo



PDBjは日本を中心にアジア地区からのデータ登録に責任を持つ



引き続き、アジア地区のデータ登録とデータ検証に責任を持つ。

ORCID IDの収集

- 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)

データ検証(Validation)レポートの発行と実験データの公開



wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 04:13 AM GMT

PDB ID : 3WDZ
 Title : Crystal Structure of Keap1 in Complex with phosphorylated p62
 Authors : Fukutomi, T.; Takagi, K.; Mizushima, T.; Tanaka, K.; Komatsu, M.; Yamamoto, M.
 Deposited on : 2013-06-20
 Resolution : 2.60 Å (reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
 We welcome your comments at validation@mail.wwpdb.org
 A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

Validation report のweb公開

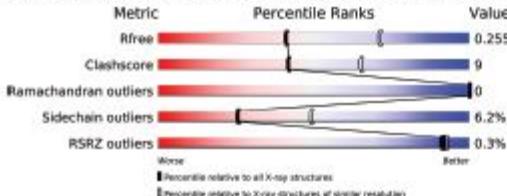
登録時に必須とされている実験情報

- X-ray: Structure Factor(構造因子)
- NMR: 化学シフトと距離拘束情報
- EM: 3DEM volume マップ

1 Overall quality at a glance i

The reported resolution of this entry is 2.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

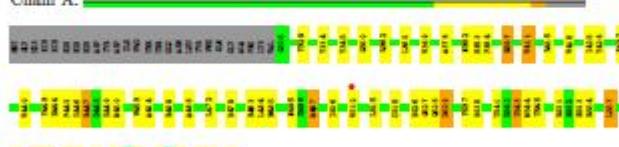


3 Residue-property plots i

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarizes the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Kelch-like ECH-associated protein 1

Chain A:



- Molecule 2: Peptide from Sequestosome-1

Chain B:



wwPDBによるValidation reportが 論文peer reviewのデフォルトに

PDBj
Protein Data Bank Japan

EDITORIAL

nature
structural &
molecular biology

Nature Struct. Mol. Biology, 23 (10), 871, 2016

Where are the data?

Here, we announce two policy changes across Nature journals: data-availability statements in all published papers and official Worldwide Protein Data Bank (wwPDB) validation reports for peer review.

“We are now taking a further step and are requesting official wwPDB validation reports for peer review.

These reports are made available by the wwPDB after data deposition

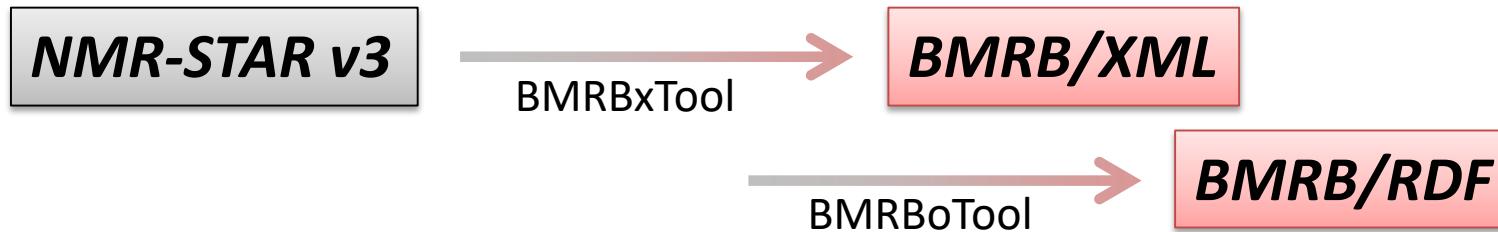
(<http://www.wwpdb.org/validation/validation-reports>). Other Nature journals will soon follow suit.”

BMRBは構造以外の情報も含む

Year	NMRによって 決定された 溶液構造の数	構造情報を伴わな い化学シフトの登録 (揺らぎや相互作用 を示すデータ)	Total
2013	486	308	714
2014	506	240	746
2015	427	333	760
2016	507	234	741

揺らぎや相互作用などの情報も新フォーマットへ

Yokochi et al., *J. Biomed. Sem.* (2016)



多機能検索エンジン

Search

Everything Sequence

e.g., Entry ID, Macromolecule, Gene Ontology or Author

Found 467 Documents (0.735 - 1.365 seconds) Order by Export as

BMRB 66 PDB 104 EMDB 3 Swiss-Prot 262 Metabolomics Ligand Expo

Showing 1 - 10 of 98 Documents, order by Relevance

1 BMRB: 17103

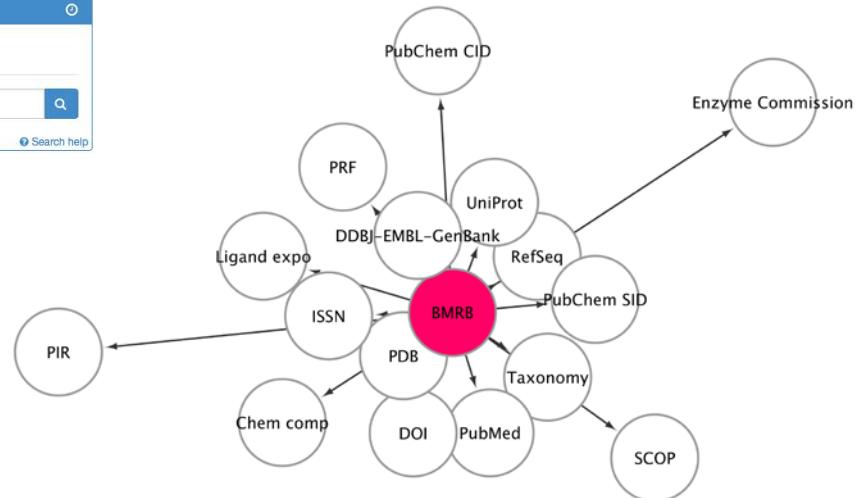
The structure of the calcium-sensitizer, dfbp-o, in complex with the N-domain of troponin C and the switch region of troponin I

2L1R score: 1750.000

Authors: Robertson, I.M., Sun, Y., Li, M.X., Sykes, B.D.
Assembly: calcium-sensitizer, dfbp-o, in complex with troponin C and the switch region of troponin I
Entity: 1. cNTnC (polymer), 89 monomers, 10062.23 Da [Detail](#) 2. cNTn(144-163) (polymer), 20 monomers, 2214.660 Da [Detail](#) 3. CA (non-polymer), 40.078 Da 4. dfbp-o (non-polymer), 197.239 Da

Total weight: 12514.208 Da
Max. entity weight: 10062.23 Da
Source organism: Homo sapiens [Detail](#)
Exptl. method: SOLUTION NMR
Refine. method: simulated annealing
Data set: assigned_chemical_shifts
Chem. Shift Compl.: Sequence coverage: 97.2 %, Completeness: 72.7 %, Completeness (bb): 73.4 % [Detail](#)
Release date: 2009-12-08

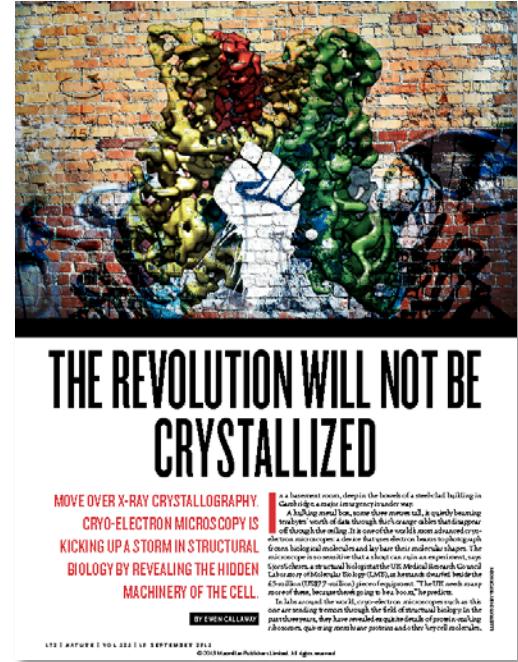
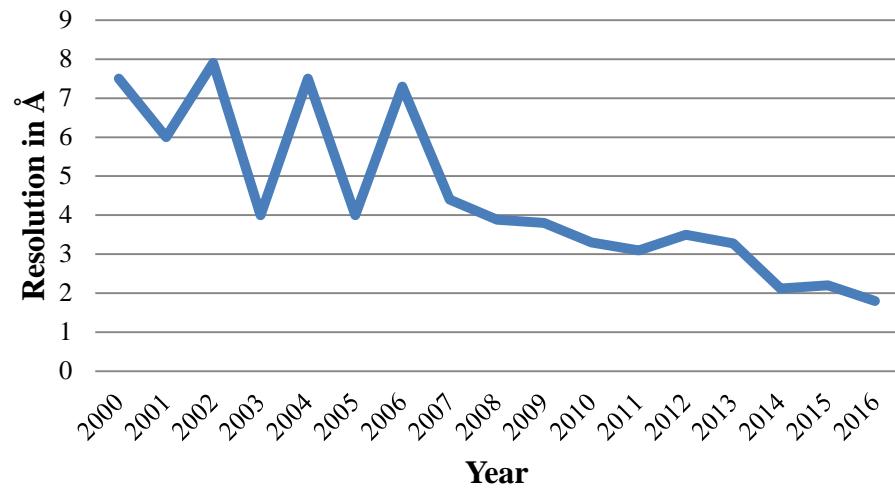
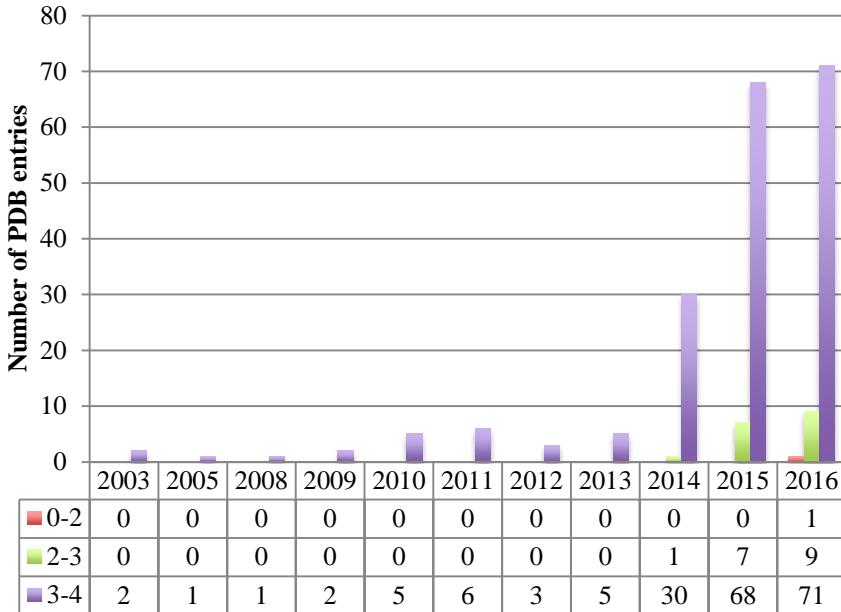
更に、機械が理解出来るフォーマットへ



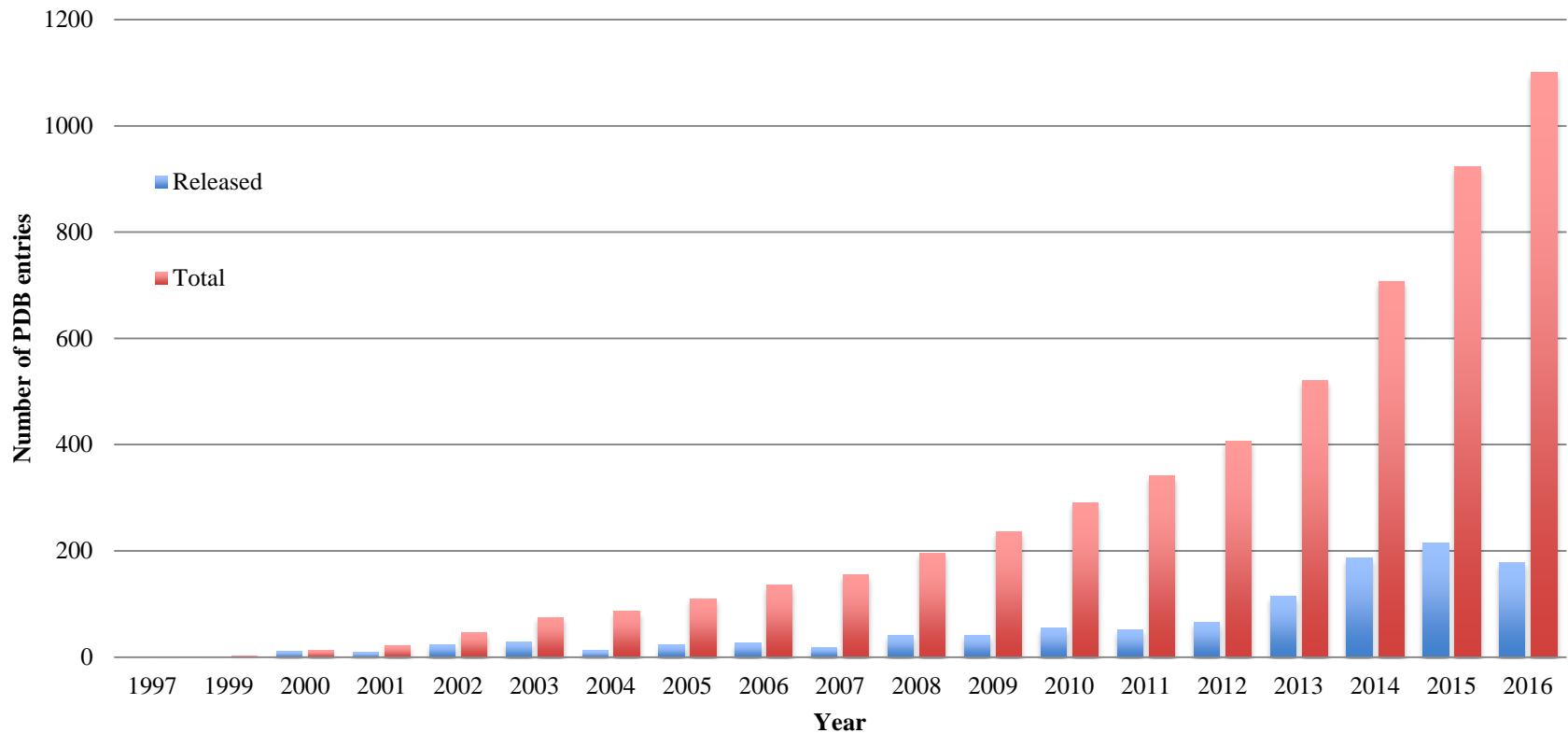
多岐に及ぶデータベースリンク

“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)



3DEM Entriesの増加



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

178 new entries released Jan 1 - Aug 1, 2016

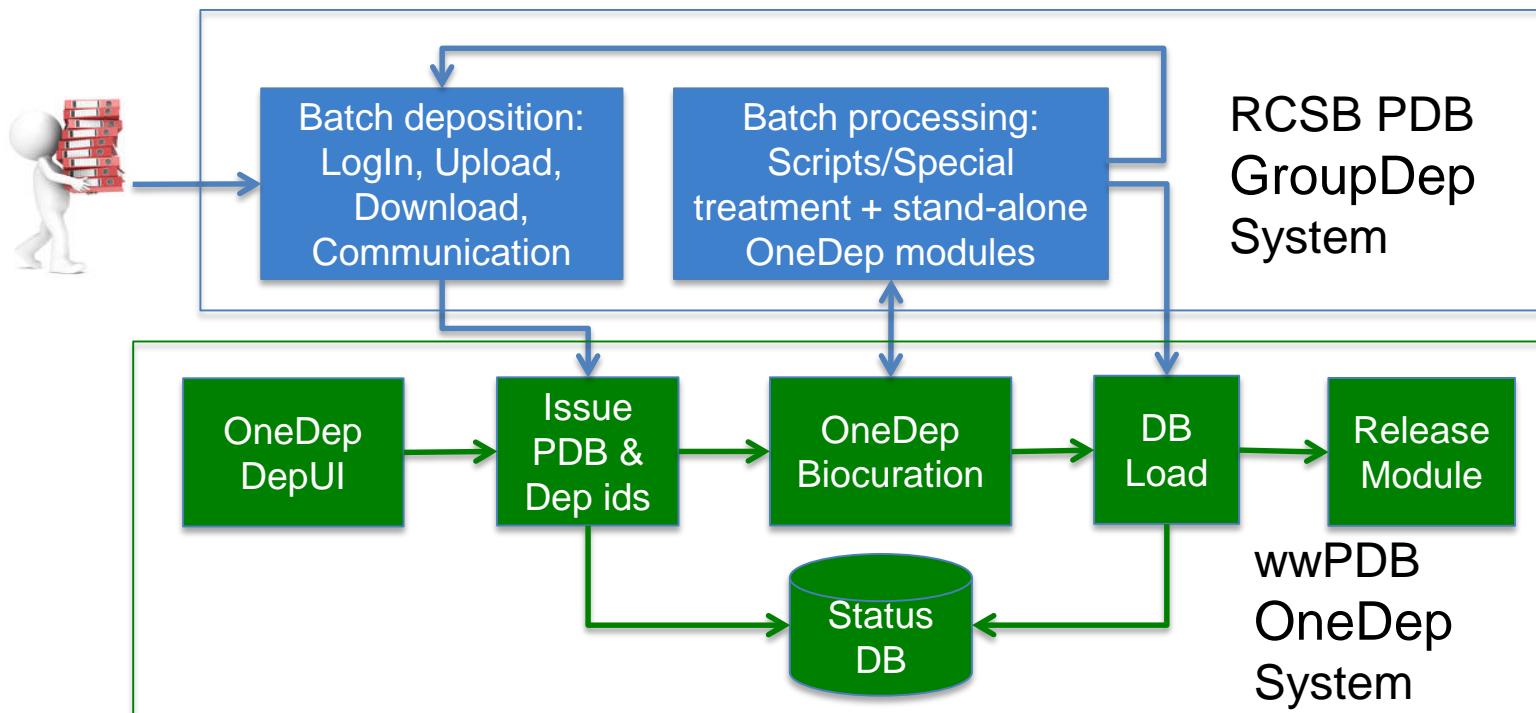
wwPDBの3DEMデータ登録ポリシー

- 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)

企業ユーザーからのバッチ登録にも対応

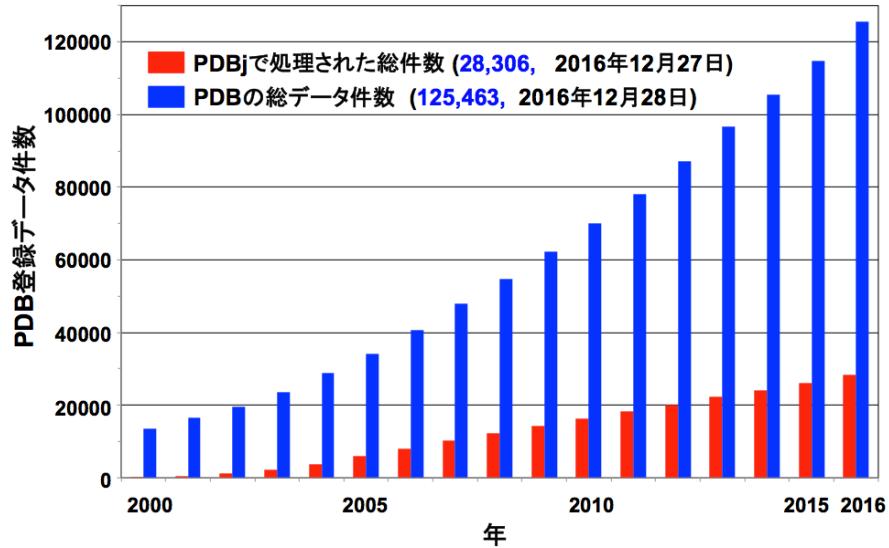


- Group Deposition processing
 - Requirements set by wwPDB OneDep Team
 - Provided support for D3R Blind Challenges



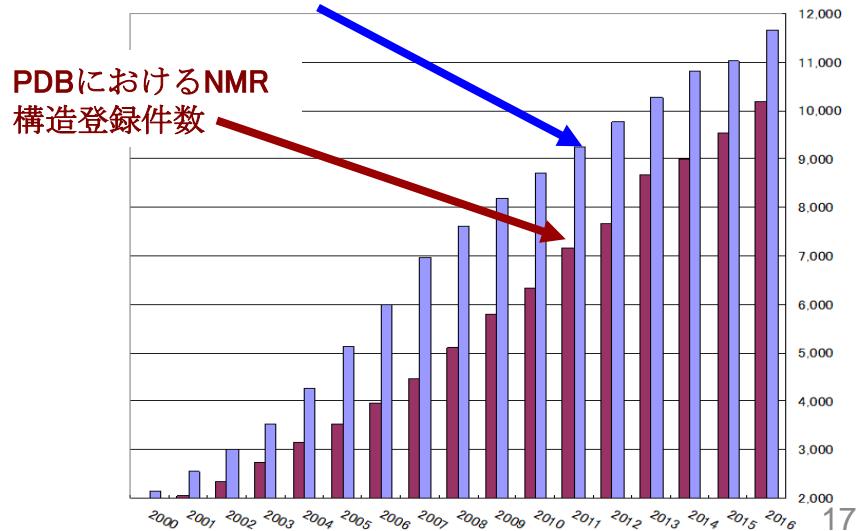
PDBjの活動

- **Data-inの活動:**
wwPDBの一員として品質管理をしつつ登録作業を実施
新たな標準フォーマット等の開発(PDB/RDF, BMRB/RDF)
- **Data-out の活動:**
共通データのダウンロードサイト(毎週アプデート)の運営
関連DBとの統合化や二次データベース・ツールの開発



BMRB(化学シフト), PDB(原子座標)の登録総数の推移

BMRB化学シフト登録件数 (10,250, 2016年12月30日)



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



PDB IDのバージョン化(1)

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

PDB IDのバージョン化(2)

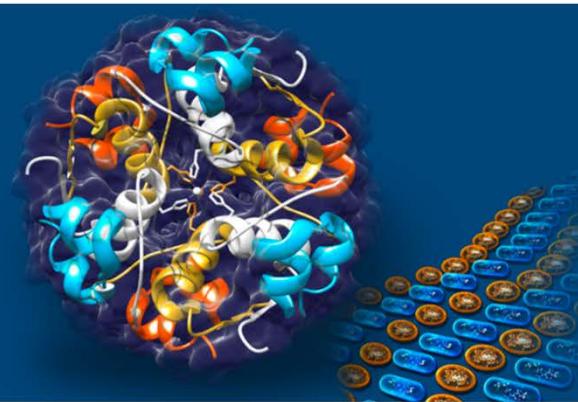
- Enable revisions to entries updated by the Depositor of Record (e.g., Version 1-0 → 1-1; 1-0 → 2-0)
 - wwPDB will NOT assign a new PDB ID going forward (for Depositor of Record revision only)
- Introduce new PDB ID code format
 - Allow more informative and transparent delivery of revised data files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

wwPDB Foundationに寄附を！



Worldwide
Protein Data Bank
Foundation

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.

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

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
- 2017 Events
 - OneDep Summi
 - wwPDB AC