PDBj
Protein Data Bank Japan

PDBjの活動の紹介

栗栖源嗣
大阪大学蛋白質研究所

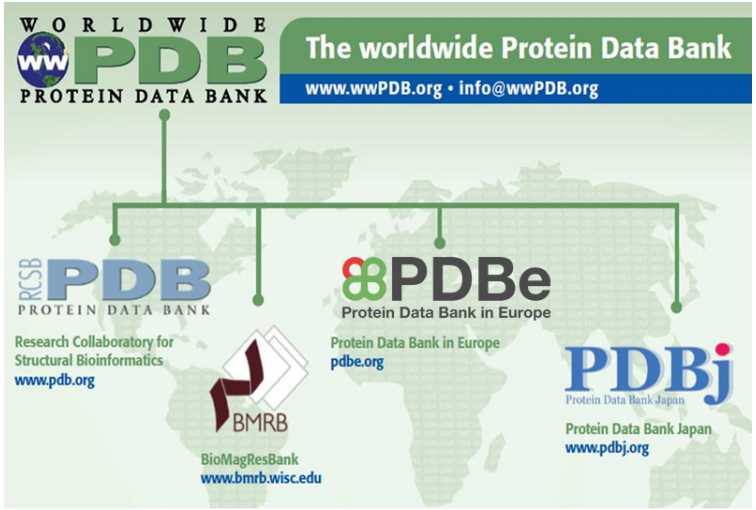
WORLDWIDE PDB
PROTEIN DATA BANK
wwpdb.org

PDBj
Protein Data Bank Japan

https://pdbj.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）．



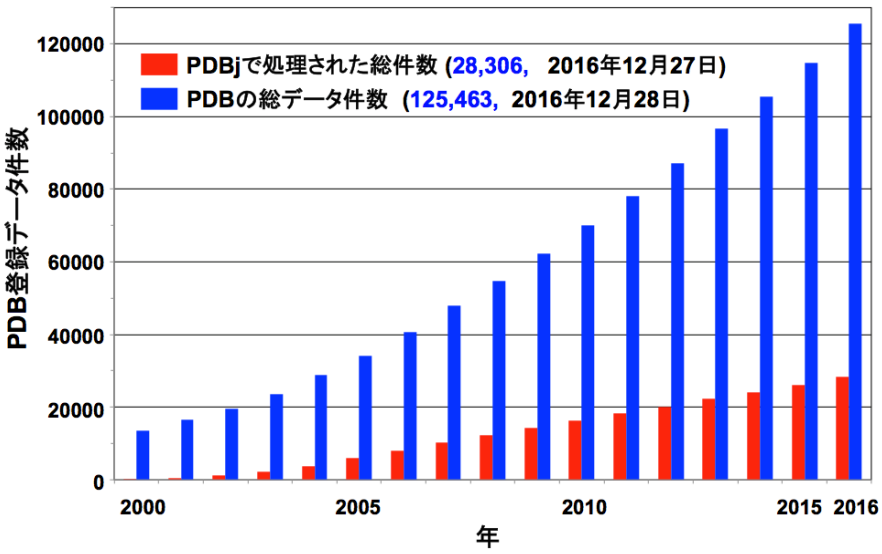
PDBjの活動

- 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

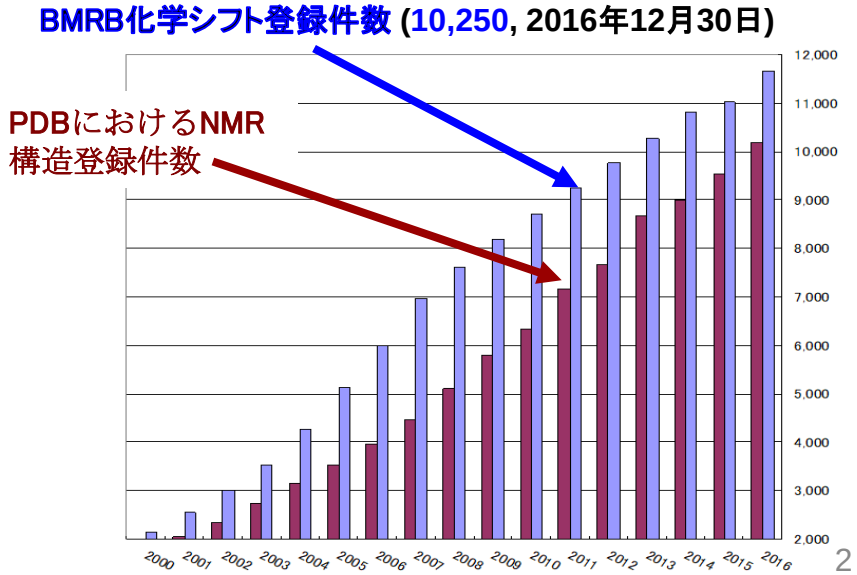
  新たな標準フォーマット等の開発(PDB/RDF, BMRB/RDF)

- 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



構造データのData-inの現状

1. 結晶解析

回折イメージ → 構造因子（hkl） → 座標データ（xyz）

配列やヘテロ化合物の情報，実験条件や生物種の情報などmetadataを付与して，PDBアーカイブとしてwwPDBが一元管理

アジア・オセアニア地区では，公的アーカイブは存在せず．
論文発表したデータ保管は研究者個人にゆだねられている．

構造データのData-inの現状

2. NMR構造解析

Spectral データ
- Chemical shifts
- Coupling constants
- Time-domain data
- Spectral peak lists

Relaxationデータ
- R1/T1, R2/T2,
- R1rho/T1rho, NOEs

→ 座標データ

配列やヘテロ化合物の情報，実験条件や生物種の情報などmetadataを付与して，PDBアーカイブへ

日本蛋白質構造データバンク（PDBj: Protein Data Bank Japan）は、JST-NBDCと大阪大学蛋白質研究所の支援を受け、米国 RCSB、 BMRB、および欧州 PDBeと協力して、生体高分子の立体構造データベースを国際的に統一化されたPDBアーカイブとして運営するとともに、様々な解析ツールを提供しております。

PDBjのOneDep登録サービス一時停止のお知らせ

12月17日（日）の午前9時より、システムのメンテナンス作業のため、**PDBjサイトのOneDep登録サービス**が一時的にご利用頂けなくなります。停止時間は3～4時間程度の予定ですが、作業進捗により、前後する場合があります。ご不便をおかけしますが、この時間帯を避けてご利用下さい。

初めての利用者のためのガイド

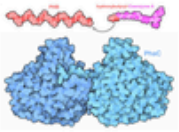
新しいウェブインターフェースを初めてご利用の方は、PDBjのウェブインターフェースについてをご覧ください。 PDBjウェブサイトの様々な機能についての説明は、対話型チュートリアルをご覧ください。

必要なサービスを探す

最新情報

今月の分子

216: 生分解性プラスチック (Biodegradable Plastic)



今月の分子のリスト

BMRB検索

PDBj-BMRB

Accession number / Deposition code

Try the omni-search bar!

Search

WORLDWIDE PDB PROTEIN DATA BANK

パートナー

Worldwide Protein Data Bank Foundation
EMDataBank
Copyright © 2013-2017 日本蛋白質構造データバンク

構造データのData-inの現状

3.　電子顕微鏡構造解析

画像データ → 密度マップ → 座標データ(xyz)

EMPIAR

欧州EBIがパイロット的にデータベースを運用中.

EMDB
Electron Microscopy Data Bank

配列やヘテロ化合物の情報，実験条件や生物種の情報などmetadataを付与して，PDBおよびEMDBアーカイブへ

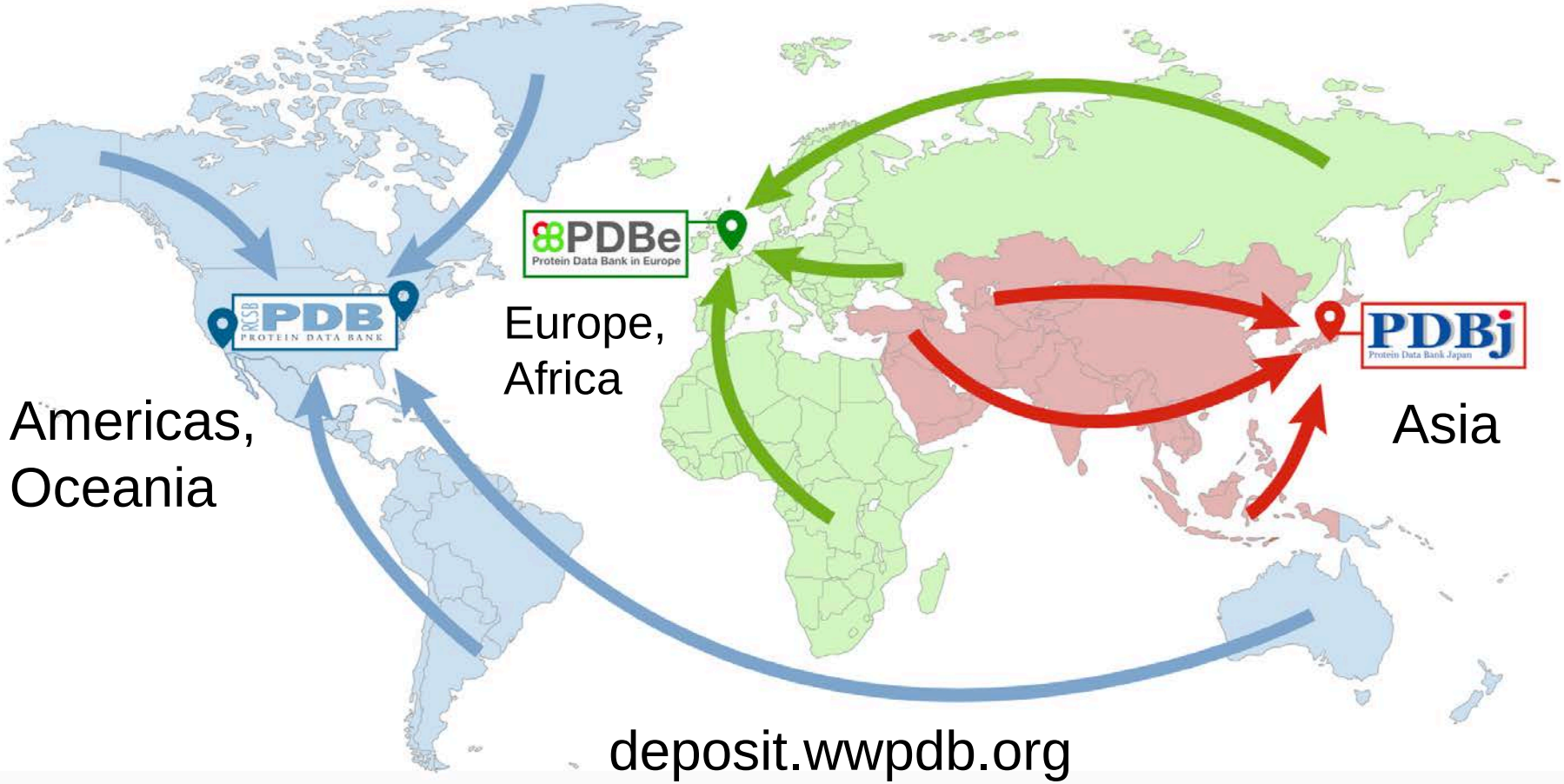
PDBj
Protein Data Bank Japan

BINDSから参画

PDBのデータは，どのように作られているのでしょうか？

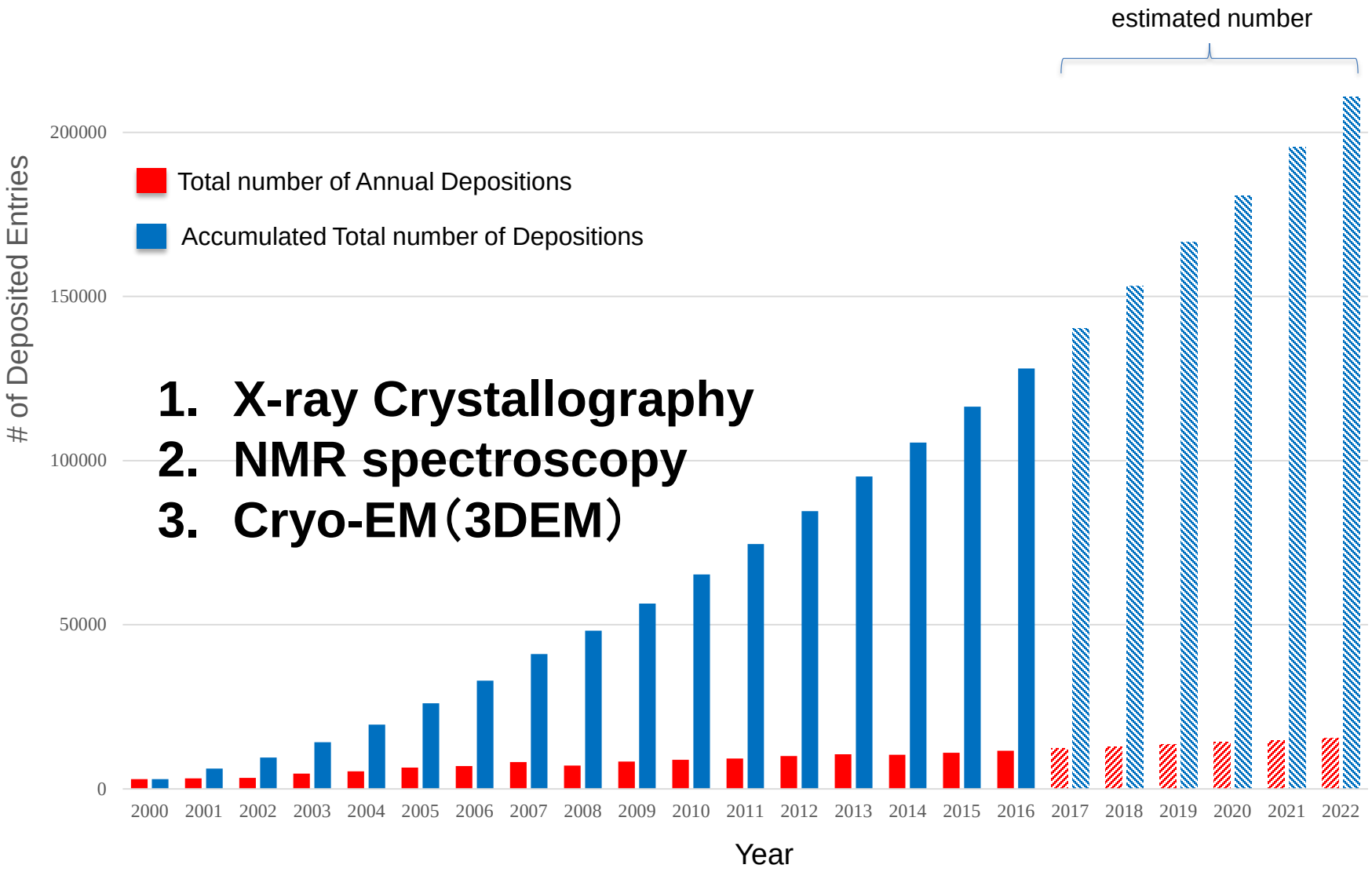
https://wwpdb.org

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.

増え続けるPDB登録数



1. X-ray Crystallography
2. NMR spectroscopy
3. Cryo-EM（3DEM）

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

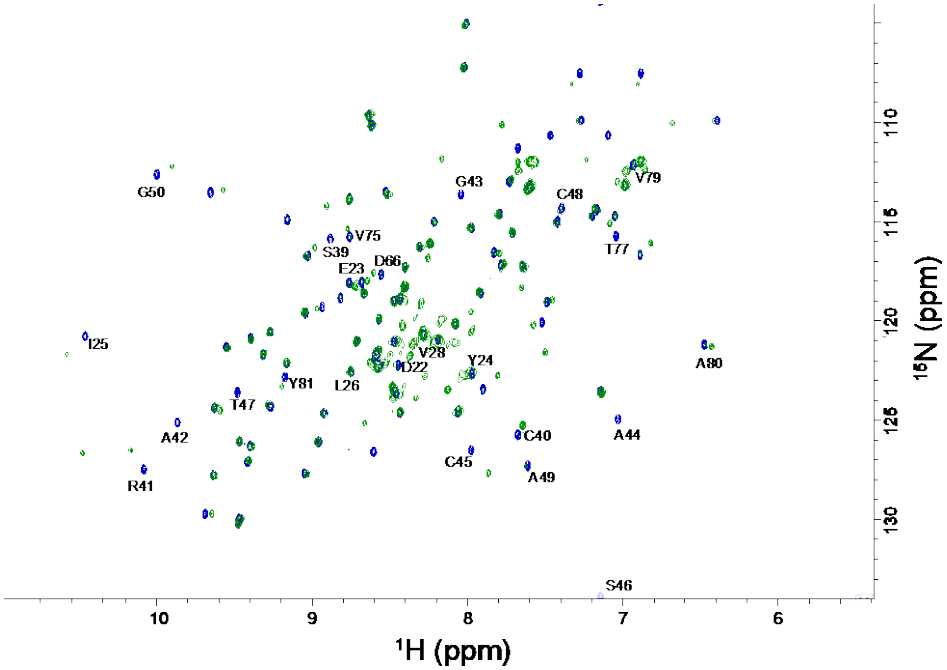
化学シフトのみの登録例

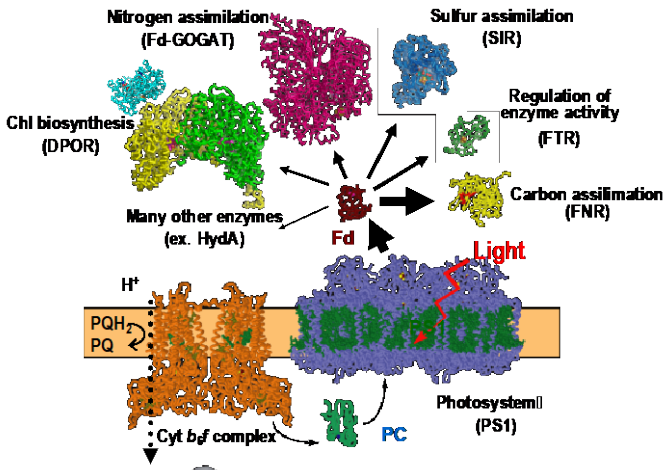
BMRB ID:11596

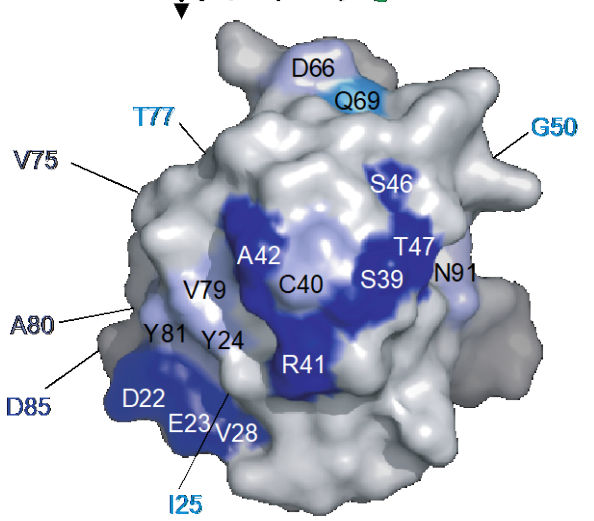
Mutoh, *et al*, *Biochemistry* (2015)

Green: with PSI

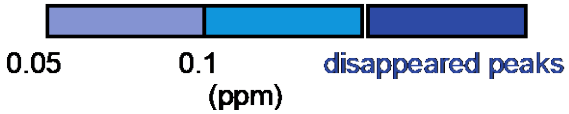
Blue: without PSI





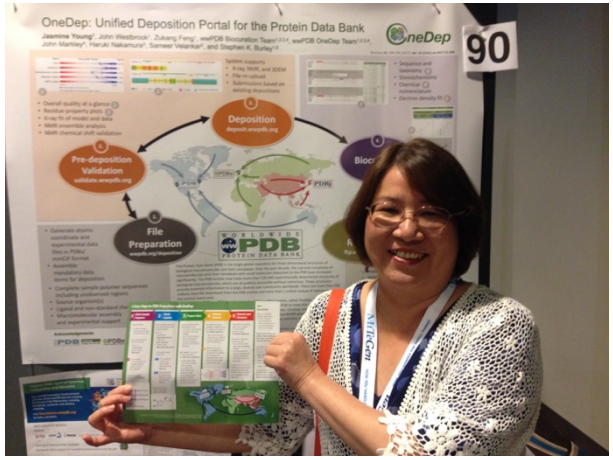


T. elongatus Fd with PSI



wwPDBメンバーとしての国際連携

wwPDB Summit: May EMBL-EBI



OneDep Posters: ACA, ECM, AsCA, APPA, Biocuration Society Meetings

wwPDB PDBx/mmCIF Meeting: July EMBL-EBI

wwPDB Booth: IUCr India

PDBで提供しているFile Formats

- mmCIF
  - The canonical format of the wwPDB.
  - Ver. 5 released this year! 596のvocabularies
- PDBML
  - “direct translation” of mmCIF into XML.
- (Legacy) PDB format
  - *NOT RECOMMENDED!*
- PDB/RDF
  - Translation of PDBML into RDF/XML (the standard format for the Semantic Web).

PDB IDのバージョン化 & 8文字化

現行の問題点

- 座標更新時には、同一の実験データに由来するものでもPDB IDが変更される
  → 論文中で使われるPDB ID と実際に公開されるPDBIDとが一致しない

新規 PDB ID コードの導入（案）

- 分かり易く透明性のあるversion管理による更新データファイル名称
- “PDB”を先頭につけ、4文字コードを拡張
  (例：現行の“1ABC” から “PDB_00001ABC”のように変更・拡張)
- PDB IDコードの例: PDB_00001ABC_XYZ_V2-2.cif.gz

実装の計画（案）

- version管理がなされた新規のPDB IDコードによるftp サイトを公開
  （ツリー構造として構造情報だけでなく関連する実験情報、メタ情報を含む）
- 現行のPDBIDを付けた ftp サイトを継続して公開
  （最新versionの構造情報のみを公開）

データ検証（Validation）レポートの発行と実験データの公開

PDBj
Protein Data Bank Japan

WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report

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 : 2013-06-26
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

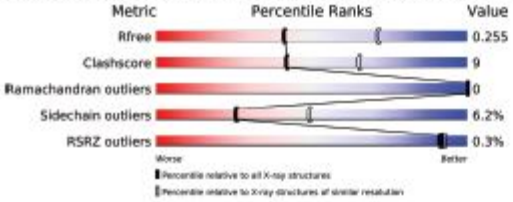
登録時に必須とされている実験情報

- **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

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.

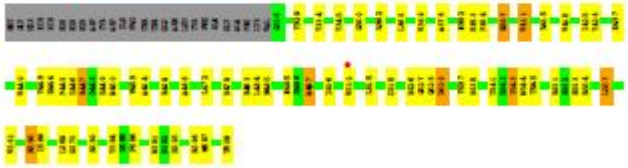


3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises 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:

Validation reportの例

PDBとBMRBのデータをどうやって使うのか？（本講習会の趣旨）

https://pdbj.org

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

PDBデータのダウンロード数

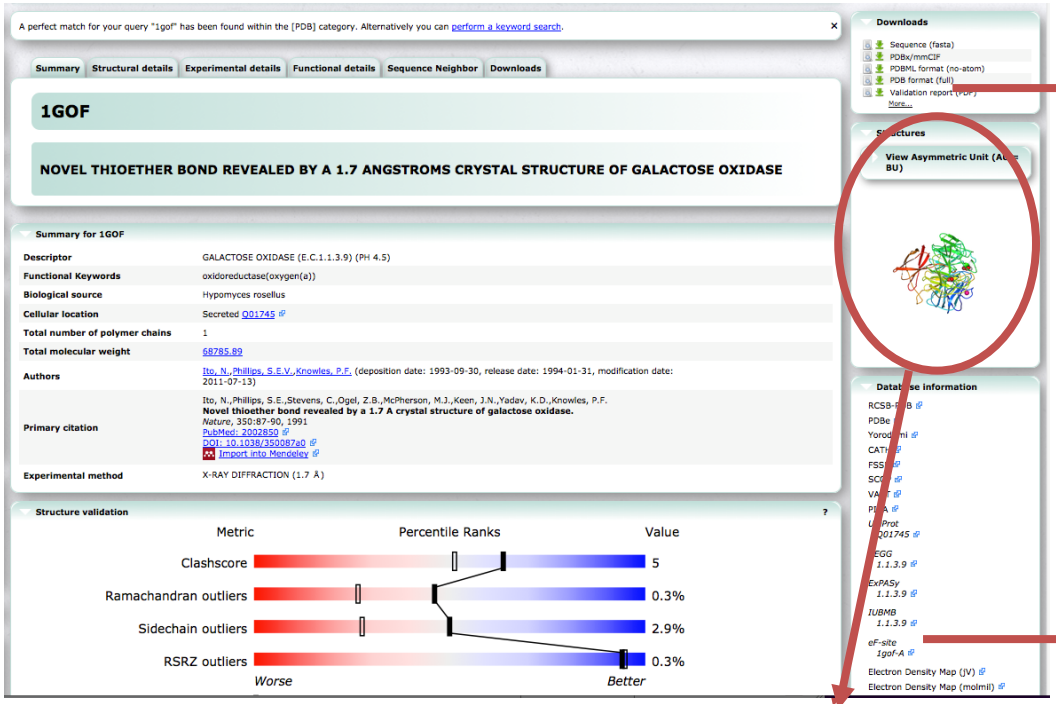
Year	Total	Total FTP Archive	Total Website	RCSB PDB FTP Archive	RCSB PDB Website	PDBe FTP Archive	PDBe Website	PDBj FTP Archive	PDBj Website
2009	328,362,536	271,116,934	57,245,602	222,984,760	53,507,785	30,141,339	1,475,116	17,990,835	2,262,701
2010	294,326,976	213,180,966	81,146,010	159,248,214	64,569,658	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
2016	591,876,087	366,677,897	225,198,190	293,648,366	161,208,456	30,274,284	44,432,830	42,755,247	19,556,904

More than 1.5 million/day

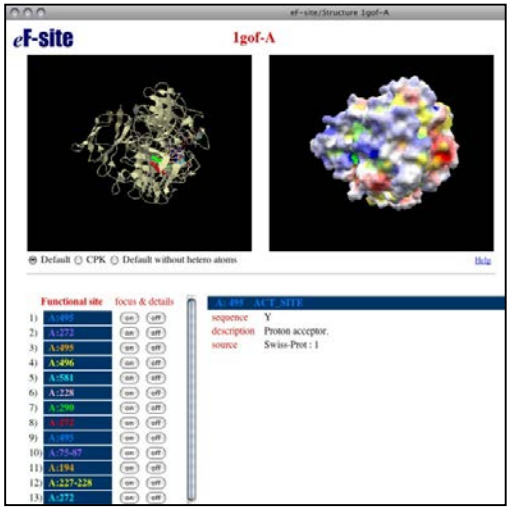
Geographic origins of FTP downloads, 2012-2015

Data-out の活動例

https://pdbj.org/



Amino acid sequence (FASTA)



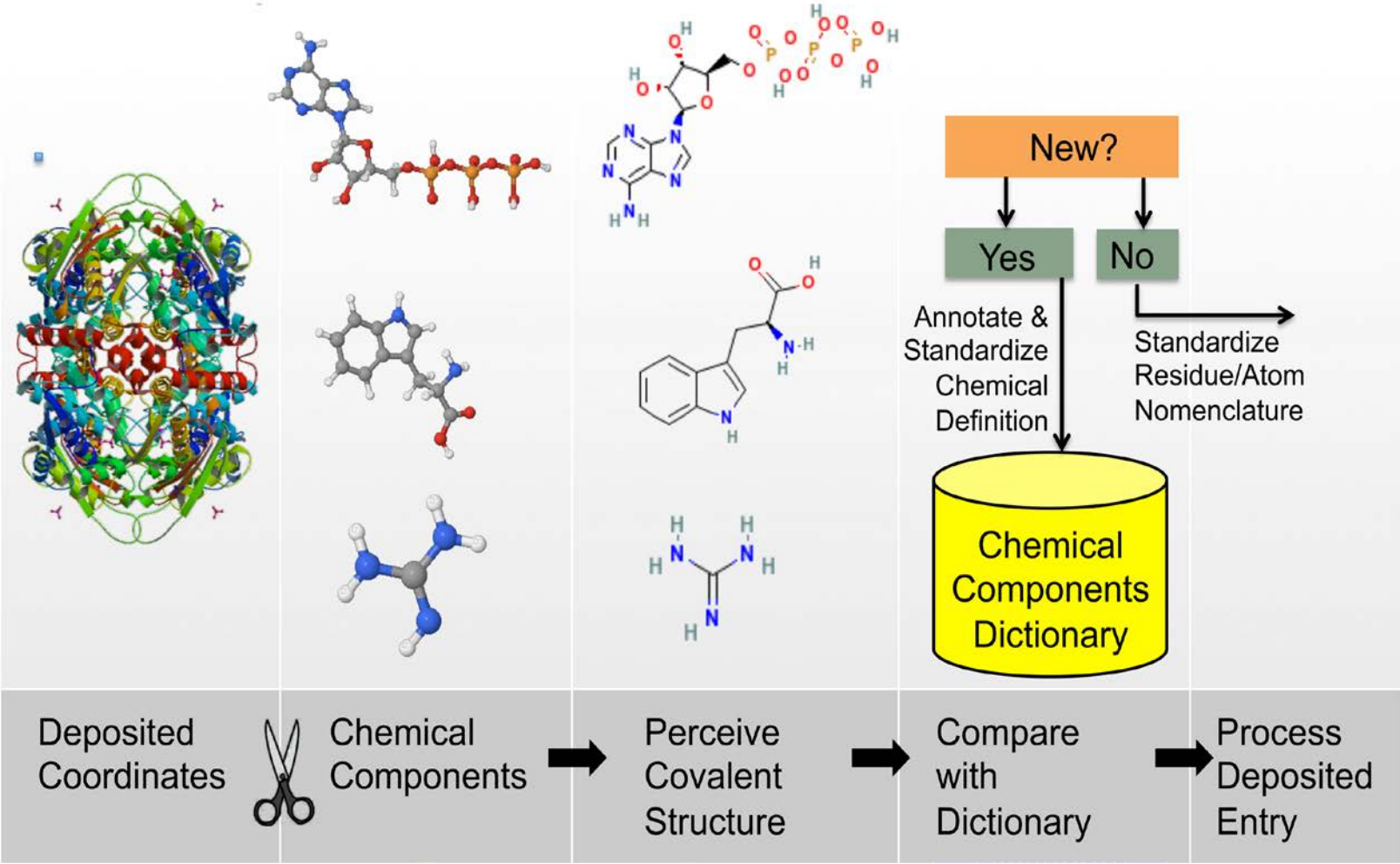
Molecular surface DB: *eF-site*
http://ef-site.hgc.jp/eF-site/

Data viewer at PDBj

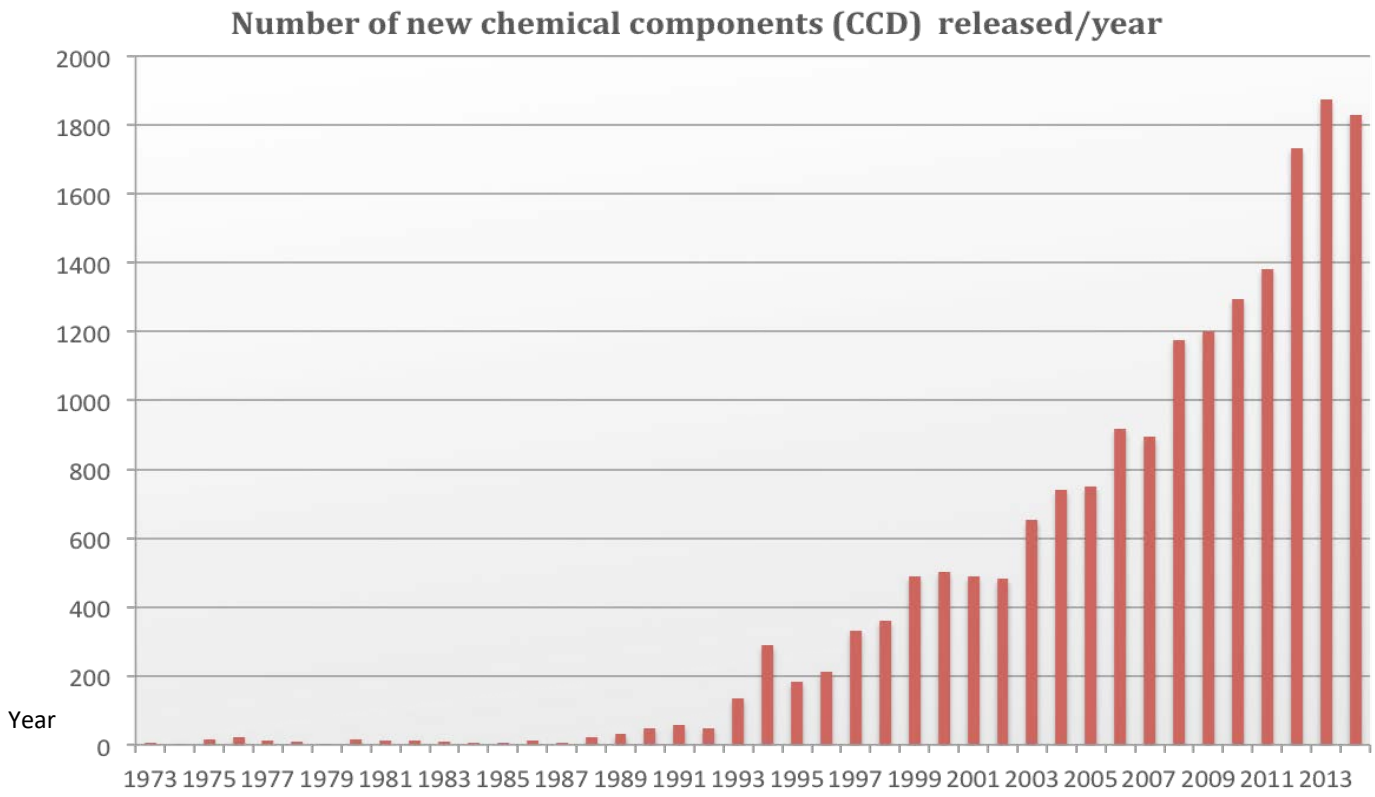
Graphic viewer: *jV* and *Molmil*
http://pdbj.org/jV/

Kinjo et al. NAR 40, D453 (2012)

PDBに含まれる化学情報（全データの3/4）



化合物データの活用と統合化
- Cambridge Structural Data (CSD) -



PDBに含まれる化合物データ(CCD)を，これから
もっと使いやすく提供していきます．

終わりに‥

質問はお気軽にPDBjまで！

Feel free to ask any questions at:
https://pdbj.org/contact?tab=PDBjmaster