

# Activities of PDBj and wwPDB: A new PDB format, Data Deposition, Validation, and Data Integration

Haruki Nakamura



Institute for Protein Research

*<http://pdbj.org/>*

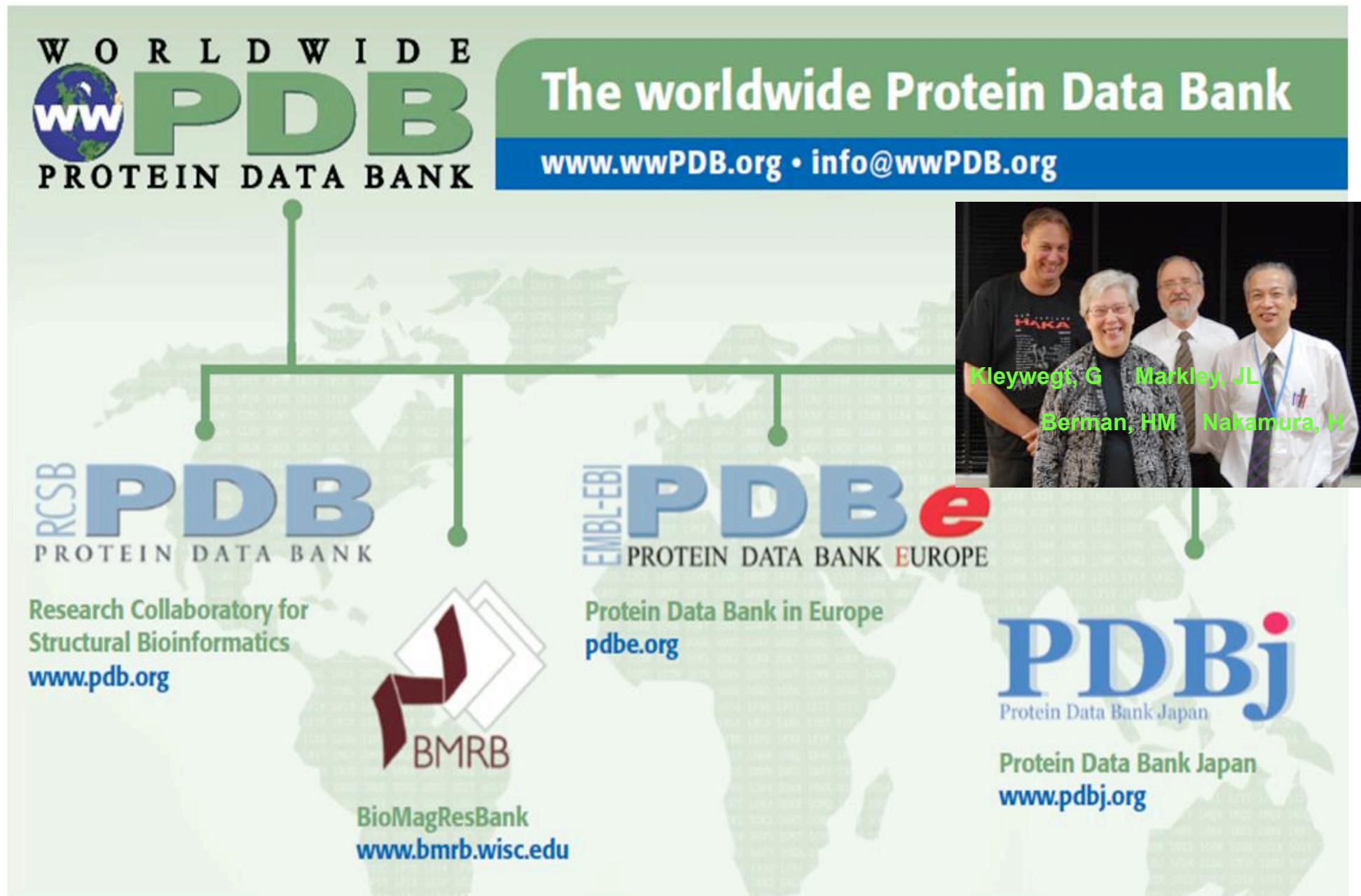
*<http://wpdb.org/>*

# wwPDB

**WORLDWIDE PDB PROTEIN DATA BANK**

The worldwide Protein Data Bank

[www.wwPDB.org](http://www.wwPDB.org) • [info@wwPDB.org](mailto:info@wwPDB.org)



**RCSB PDB**  
PROTEIN DATA BANK

Research Collaboratory for  
Structural Bioinformatics  
[www.pdb.org](http://www.pdb.org)

**BMRB**

BioMagResBank  
[www.bmrb.wisc.edu](http://www.bmrb.wisc.edu)

**EIBL-EBI PDBe**  
PROTEIN DATA BANK EUROPE

Protein Data Bank in Europe  
[pdbe.org](http://pdbe.org)

**PDBj**  
Protein Data Bank Japan

Protein Data Bank Japan  
[www.pdbj.org](http://www.pdbj.org)



Kleywegt, G    Markley, JL  
Berman, HM    Nakamura, H

# **wwPDB Advisory Committee on Sept 27, 2013 Rutgers Univ.**





# Protein Data Bank Japan

*<http://pdbj.org/>*

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

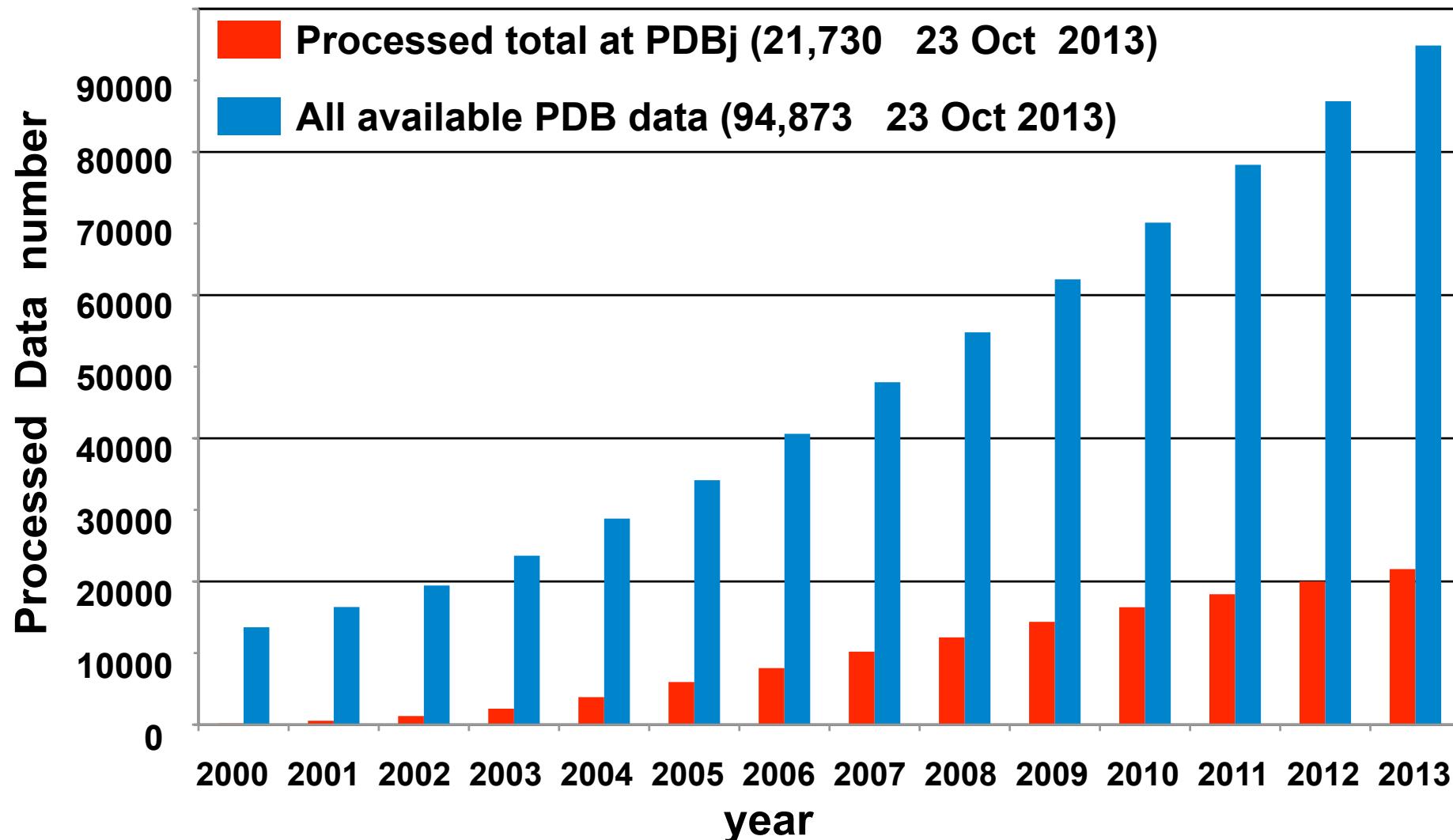


PDBj staffs (April 2013)

# 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, characteristic at each wwPDB member site except common ftp-site data.

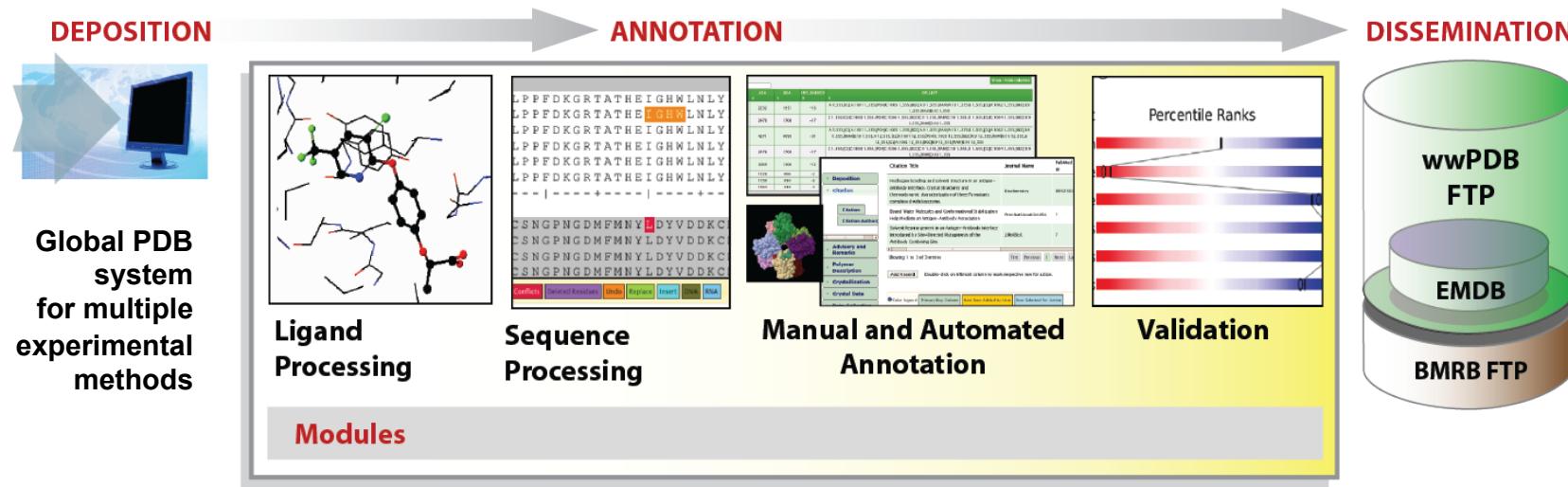
# Data-in at PDBj and wwPDB



PDBj curates and processes about a Quarter of the deposited data, mainly from Asian and Oceania regions

# New Annotation System

## For data increase and high quality data management



- Enables workload balancing and increased productivity
- Better quality assurance of **ligand chemistry and polymer sequences**
- **PDBx/mmCIF** is the master file format
- Validation suites based on recommendations from expert task forces; **X-ray validation pipeline** is available as a stand-alone server
- System will support all accepted experimental methods

# Validation Report

- **Version 1.0** in production use since August 2013
  - <http://www.wwpdb.org/validation.html>
  - Fixing occasional bugs
  - Collecting feedback to inform possible changes

[validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)
- **January 2014** – validation data for all X-ray structures will be made publicly available through the wwPDB ftp sites

wwPDB X-ray Structure Validation Report [\(i\)](#)

Aug 13, 2013 – 09:45 AM BST

PDB ID : 1CBS  
Title : CRYSTAL STRUCTURE OF CELLULAR RETINOIC-A CID-BINDINGPROTEINS I AND II IN COMPLEX WITH ALL-TRANS-RETINOIC ACID AND A SYNTHETIC RETINOID  
Authors : Kleywegt, G.J.; Bergfors, T.; Jones, T.A.  
Deposited on : 1994-09-28  
Resolution : 1.80 Å(reported)

**DISCLAIMER**  
This is a preliminary version of a new style of wwPDB validation report.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	trunk21216
Percentile statistics	:	20591
Refmac	:	5.8.0043
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk21216

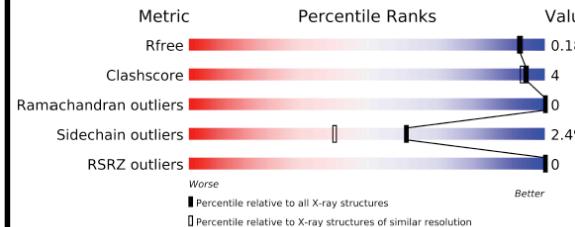
# Validation Report

- **Summary**
  - Quality vs. all PDB X-ray
  - Quality vs. entries at similar resolution
  - Overview of residue-based quality for every polymer
  - Table of ligands that may need attention

## 1 Overall quality at a glance (i)

The reported resolution of this entry is 1.80 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.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R <sub>free</sub>	65580	5522 (1.84-1.76)
Clashscore	76988	5040 (1.82-1.78)
Ramachandran outliers	75395	6528 (1.84-1.76)
Sidechain outliers	75377	6529 (1.84-1.76)
RSRZ outliers	65576	5522 (1.84-1.76)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.



The following table lists non-polymeric compounds that contain outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	NAG	A	401	-	X
2	NAG	C	401	-	X

# Validation Report

- **Residue quality**

**Red dots: poor density (**RSR-Z > 2**, as in EDS)**

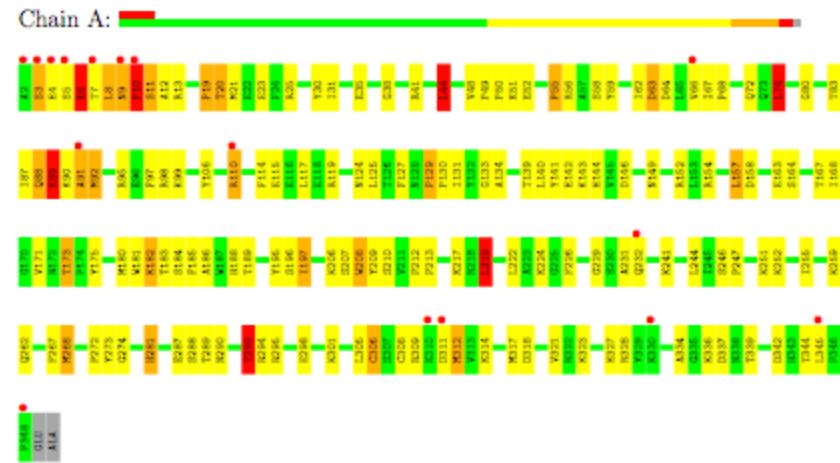
- **Model/data fit ligands etc.**

– “**LLDF**” –*Local Ligand Density Fit* = Z-score of ligand RSR relative to nearby polymeric residues

## 3 Residue-property plots

The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence annotated by issues in geometry and electron density. Residues are color-coded according to the number of different types of geometric errors found. Green signifies no errors, yellow, orange and red 1, 2, and 3 or more errors respectively. A red dot above a residue indicates a problem with electron density. Regions of sequence for which no errors are detected are indicated by a green connector.

- Molecule 1: Jumonji domain-containing protein 2A



# Data-out from PDBj

<http://pdbj.org/> and <http://legacy.pdbj.org/>

93970  
件が利用できます。(2013-09-18  
00:00 UTC / 09:00 JST)

ホーム  
トップページ  
統計情報  
ヘルプ  
FAQ  
お問い合わせ  
リンク集  
PDBアーカイブ

データ登録  
ADIT: PDBへの登録  
ADIT-NMR  
データ登録について

検索  
Search PDB (PDBj Mine)  
Search PDB (詳細検索)  
Large Structures  
Search BMRB  
Sequence-Navigator  
Structure-Navigator  
EM Navigator  
wwwPDB/RDF  
SesAW  
Ligand Binding Sites (GIRAF)  
最新の公開エントリー  
未公開エントリーのステータス

サービス&ソフトウェア  
二次データベース  
教材  
PDBjについて  
カスタマイズ



English 日本語 简体中文 繁體中文 한국어

Search pdbj.org  
wwwPDB RCSB PDB BMRB PDBe IJST-NBDC IPR 豊来版 Share



今月の分子



**PDBj-BMRB**

日本蛋白質構造データバンク (PDBj: Protein Data Bank Japan) は、IJST-NBDC と大阪大学 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。

## 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。この説明は、[対話型チュートリアル](#)をご覧ください。

新しいPDBj ウェブサイトが、**5月23日より、正式なPDBjウェブサイトになりました。** 更なる改良のために、このPDBj ウェブサイトは、下記URLから引き続きお使い頂けます。<http://legacy.pdbj.org/>

## メンテナスのお知らせ

9月25日(水)の午前11時から午前11時半の間、システム保守作業のため、[ProMode\\_Elastic](#)サービスがご利用頂ける間帯を避けてご利用頂きますようお願い致します。

## 最新情報

### ニュース (2013年9月6日)

wwPDBのNMR検証作業チームの提言をまとめた論文が公開されました

### ニュース (2013年9月6日)

論文が公開されました「How Community Has Shaped the Protein Data Bank」



English

PDBj BMRB

PDBj-BMRBでは生体高分子のNMR実験データの登録受付・公開をPDBjグループの一員として行っています。

**BMRB Search**  
Retrieve entry by BMRB accession No.  
 search  
 Accession number  Deposition code  
登録後に該当するもAccNoで検索すると直進で貢献です。

Retrive by keywords (Google)  
 search  
 BMRB Database www

**Deposit Data**  
ADIT-NMR AP NMR  
生体高分子のNMR実験データの登録はこちら。▶

SMS-Dep SMSDep  
低分子量生体分子のNMR実験データと希少溶解データ登録▶

**BMRB Database**  
Osaka site BMRB  
BMRBウェブサイトはこちら。▶

**Contents**

Over view of the Submission Process	Submitting data	ADIT-NMR or SMS-Dep	FAQ
BMRB(データバンク)へ データ登録の流れ	登録サイトADIT-NMRの 使い方、登録項目	どちらに登録するべきか	よく寄せられるご質問を まとめてあります
NMR Tool BOX	BMRBxTool	Tool List	Contact us
実験解析ツール	データベース活用ツール	Validation Tool Deposit Tool PDBj-BMRB Tool	お問い合わせ

**Information**  
Feb. 04-2013 2013年2月6日以降にADIT-NMRにてPDBに登録される座標と実験データは、同じ公開条件(ステータス)に従います。[\(リンク\)](#)

**Logos:**  
PDBj Protein Data Bank Japan  
BMRB  
NBDC  
NBDC  
INSTITUTE OF PROTEIN RESEARCH

# Summary page

<http://pdbj.org/>

91761  
Entries available on 2012-06-26  
00:00 UTC / 09:00 JST

**PDBj**  
Protein Data Bank Japan

English 日本語 简体中文 繁體中文 한국어 Advanced search

Search pdbj.org

wwwPDB RCSB BMRB PDBe ISTM-NBDC IPR Legacy Share

**1GOF**

NOVEL THIOETHER BOND REVEALED BY A 1.7 ANGSTROMS CRYSTAL STRUCTURE OF GALACTOSE OXIDASE

**Summary for 1GOF**

**Descriptor:** GALACTOSE OXIDASE (E.C.1.1.3.9) (PH 4.5)

**Functional Keywords:** OXIDOREDUCTASE(OXYGEN(A))

**Biological source:** Hypomyces rosellus

**Cellular location:** Secreted

**Total number of polymer chains:** 1

**Total molecular weight:** 59785.88

**Authors:** Ito, N., Phillips, S.E.V., Knowles, P.F. (deposition date: 1993-09-30, release date: 1994-01-31, modification date: 2011-07-13)

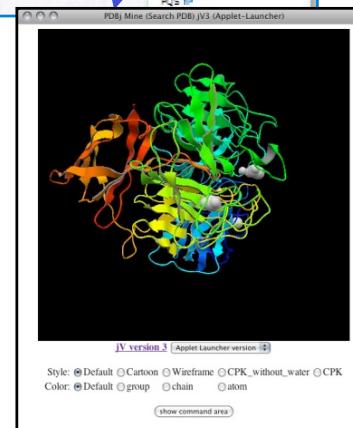
**Primary citation:** Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., Knowles, P.F. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature*, 350:87-90, 1991  
DOI: 10.1038/350087a0

**Experimental method:** X-RAY DIFFRACTION (1.7 Å)

**More Asymmetric unit images**

Copyright © 2013 PDBj. All Rights Reserved.

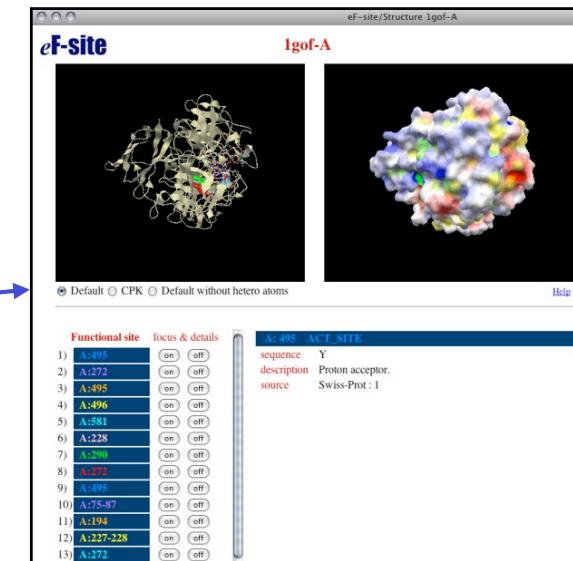
## Data viewer at PDBj



Graphic viewer: **jV**  
<http://pdbj.org/jV/>

## Amino acid sequence (FASTA)

```
ftp://ftp.pdbj.org/XML/pdbmlplus/fasta_seq_data/1gof_seq.txt
>1-GOFA:GALACTOSE_OXIDASE
ASAPIGSAISRNNNAWVTCDSAGSGNECNKAIDGNKDTFWHTFYGANGDPKPPHTYIDMK
TTQNVNGLSMLPFDQDGNGQWIGHREVVLSDGTNGNSGVASGSWFADSTTXYSNPETRP
ARTVRVLVAITEANGOPWTSIAEINVQASSYTAPQFLGRNGFTIDLPIVPAAAIEPTS
GRVLMWSYRNDAFGCGSPGICITLSSWDPTGIVSDRTVTVTHKDMCPGICSMODNGQIV
VTGQNDAKRTSLYDSSDWIPEGPDNMQVARGYQSAATMDEGRVFTIGGSNSGGVKEKNGE
WPSFSRSTWVTRKQVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY
EGSDFVSKVSEKQGRRVQVAFVQVAFVQVAFVQVAFVQVAFVQVAFVQVAFVQVAFVQVAFV
DTTYKQNPNS1VVRVHS1SLLPDPGRVFNGGGCLCDDCTTNEFDAQ1FTPNLYNSNGNL
ATRPKIRTRTSTSQVKVGRIT1STDSS1KASLRLRVGTAHTWNTDQRRIPLTLTNNGN
SYSFQVPSDSCVALPGYWHLFVMNSAGVPVASVIRVTPQ
```



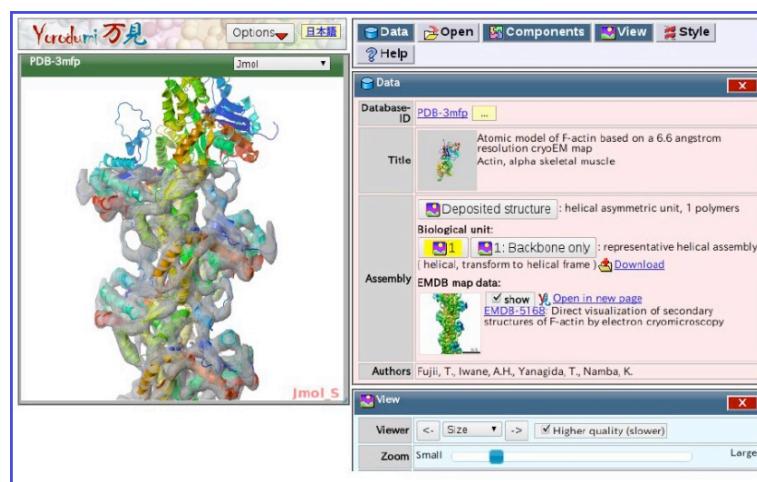
Molecular surface DB: **eF-site**  
<http://ef-site.hgc.jp/eF-site/>

Kinjo et al. NAR 40, D453 (2012)

# PDBj services



**EM Navi:** EMDB browser



**Yorodumi:** PDB & EMDB

**GIRAF query upload**

Interface type

- @nonpolymer binding
- @DNA and RNA binding
- @All types of ligands (nonpolymer, DNA, RNA, peptides, and others).
- @PPI (protein-protein interfaces).
- @All (ligands + PPI).

Input PDB ID:

or upload a PDB file:  ファイルを選択 ファイルが選択していません

Chain IDs (optional):  (comma-separated multiple IDs [e.g., "A,B"] or "all" are allowed.)

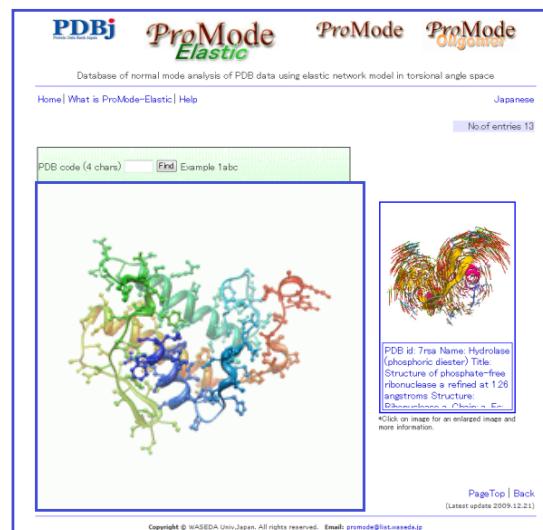
Limit target PDB entries (optional):  (comma-separated multiple IDs [e.g., "101m,1a00"] or "all" are allowed.)

Number of displayed results (optional):  100

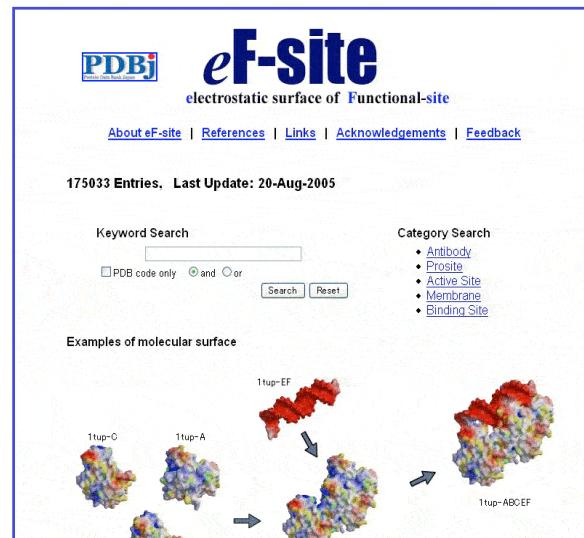
Your email address (optional):

DB version: 2013-09-21:713107 interfaces

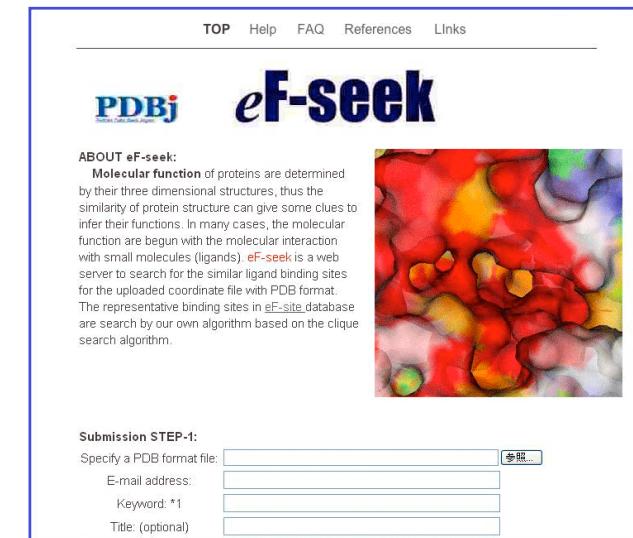
**GIRAF:** Similar binding site



**ProMode:** Normal Mode



**eF-site:** Mol. SurfaceDB



**eF-seek:** Similar surface

# EM Navigator

3次元電子顕微鏡データナビゲーター [ [English](#) / [日本語](#) ]

■ トップ ■ ギャラリー ■ リスト ■ 分布図 ■ 統計情報 ■ ビューア ■ ? 説明

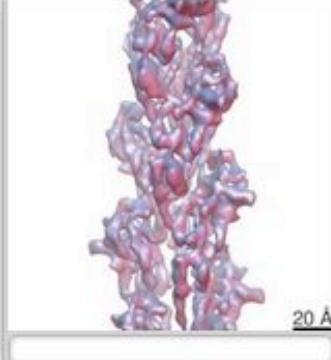
■ PDBj > ■ EM Navigator

データを見る 詳しく

• さがす : (キーワード / EMDB ID / PDB ID)  EMDBがPDBのID、あるいはキーワードを入力

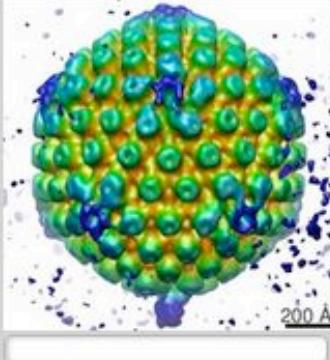
• 眺める : ■ ギャラリー ■ リスト ■ 表

ムービースロット ■ 再生 方向 ■ シャッフル



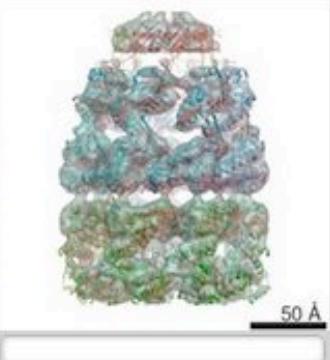
20 Å

■ EMDB-1980



200 Å

■ EMDB-5452



50 Å

■ EMDB-2325

## 情報

**EM Navigatorとは？**

- 生体分子や生体組織の3次元電子顕微鏡データを、気軽にわかりやすく眺めるためのウェブサイトです。
- [EMDB](#) と [PDB](#) のデータを利用しています ([統計情報](#))
- 分子・構造生物学の専門家にも、初心者や専門外のかたにも利用していただけるサイトを目指しています。
- PDBが運営しています。

[? 詳しくはこちら](#)

**お知らせ**

- 2013-09-11: 公開データ






その他の最新データ: [EMDB付随情報](#), [EMDB更新](#), [PDB更新](#)

- 2013-09-04: 公開データ







## New format PDB: PDBx/mmCIF

- PDB format is almost **40 years old** and does not support today's science.
  - PDB Record format limitations
    - **Max. 62 chains**
    - **Max. 99,999 atoms**
    - **No bond orders or chirality specified for ligands**
    - **No support for NMR, EM, hybrid methods, ...**
    - **Meta-data specification cumbersome and inflexible**
- 
- **Preserve backward compatibility where possible**
  - **web service to create PDB format data files**
  - **Start in early 2014**

# Current Supported Archival Formats

## *protein structure format universe*

PDB (ca. 1974)

PDBx/mmCIF (ca. 1997)

PDBML (ca. 2005)

RDF (ca. 2011)



In managing the formats, PDBx is the master format.

# PDB Format Example

```
REMARK 3 DATA USED IN REFINEMENT.  
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.57  
REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 23.00  
REMARK 3 DATA CUTOFF (SIGMA(F)) : 0.000  
REMARK 3 COMPLETENESS FOR RANGE (%) : NULL  
REMARK 3 NUMBER OF REFLECTIONS : 43316  
REMARK 3  
REMARK 3 FIT TO DATA USED IN REFINEMENT.  
REMARK 3 CROSS-VALIDATION METHOD : NULL  
REMARK 3 FREE  
REMARK 3 R VALUE : 0.0000  
REMARK 3 R VALUE FREE : 0.0000  
REMARK 3 FREE  
REMARK 3 FREE  
REMARK 3 FREE  
REMARK 3 FREE
```

- Record-oriented with fixed column format
- Metadata in semi-structured remarks
- Documentation by example
- Most widely used and supported archival format

ATOM	1	N	VAL A 363	22.711	-1.937	11.725	1.00	33.52	A
ATOM	2	CA	VAL A 363	21.557	-0.831	11.024	1.00	32.13	C
ATOM	3	C	VAL A 363	20.954	-1.757	9.943	1.00	31.73	C
ATOM	4	O	VAL A 363	19.737	-1.906	9.845	1.00	30.94	O
ATOM	5	CB	VAL A 363	21.883	0.552	10.391	1.00	33.45	C

# PDBx/mmCIF Format Example

- Name – value pairs

```
_exptl.entry_id          1XBB
_exptl.method            'X-RAY DIFFRACTION'
_exptl.crystals_number   1
```

- Tables

```
loop
  _data
  _data
  _data
  _data
  _data
  _data
    _database_PDB_rev.replaces
    _database_PDB_rev.status
 1 2004-11-02 2004-08-30 0 1XBB ?
 2 2005-03-22 ?           1 1XBB ?
 3 2009-02-24 ?           1 1XBB ?
```

- Simple syntax
- Named data items
- Data semantics defined in the PDBx data dictionary
- Software support in most popular languages

# PDBML Example

```
<PDBx:entity_polyCategory>
  <PDBx:entity_poly entity_id="1">
    <PDBx:type>polypeptide (L) </PDBx:type>
    <PDBx:nstd_linkage>no</PDBx:nstd_linkage>
    <PDBx:nstd_monomer>no</PDBx:nstd_monomer>
    <PDBx:wpdbx_seq_one_letter_code>
      D T V I . T O S D A S T . S A S V C E T V V T I T C R A S C N T H N V T . A W Y Q Q K O C K S D O T . T . V V V T T T I . A D G
    </PDBx:wpdbx_seq_one_letter_code>
  </PDBx:entity_poly>
</PDBx:entity_polyCategory>
```

- Three flavors of XML files:
  - fully marked-up files
  - files without atom records
  - files with a more space efficient encoding of atom records
- Follows naming and semantics of the PDBx data dictionary



# RDF Example

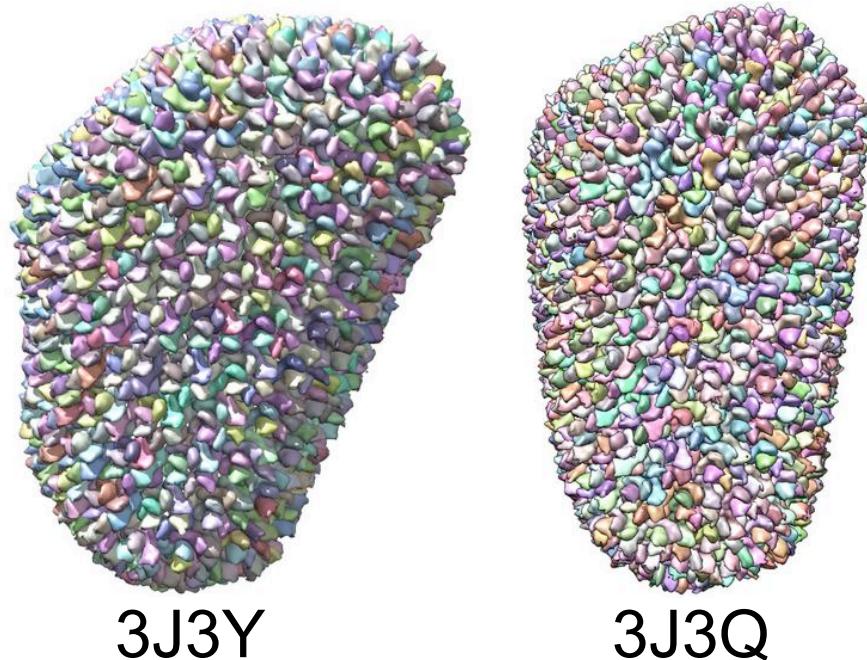
- Entry point for semantic web and reasoning systems
  - Translates data items in PDBx/mmCIF schema into triples with URL identifiers
  - Follows naming and semantics of the PDBx data dictionary
  - `http://pdbj.org/rdf/<pdbID>/<categoryName>/<pkey1>,...`
  - For example, `http://pdbj.org/rdf/1GOF/entity/1`

# Transitional Home for Large Structures

Large single entries are now stored separately on the wwPDB ftp site, and PDB internally produces *divided/split* PDB format files.

[ftp://ftp.wwpdb.org/pub/pdb/data/large\\_structures/mmCIF/](ftp://ftp.wwpdb.org/pub/pdb/data/large_structures/mmCIF/)  
[ftp://ftp.wwpdb.org/pub/pdb/data/large\\_structures/XML/](ftp://ftp.wwpdb.org/pub/pdb/data/large_structures/XML/)

HIV-1 Capsid 3J3Q –  
• 1356 chains  
• >2M atoms  
• 25 – PDB format entries



# Providing Format Compatibility

- Adopt a *PDB friendly* mmCIF/PDBx style -
  - All records on a single text line
  - Columns presented in standard column order.
  - Tabular presentation with leading record names
    - (e.g. ATOM, CELL, REFINEREFINE)
  - Method independent features in left-most column
    - S (e.g. identifiers & coordinates)
  - Method specific features in the right-most columns
    - (e.g. ADPs, NMR order/disorder parameters)
  - Continue to support PDB nomenclature semantics
    - (e.g. PDB style chains, residue numbering, and insertion codes)
- Large entries will be internally converted to divided/split PDB format files.

ATOM	1	N	GLN	A	39	24.690	-27.754	24.275	1.00	60.76		N
ATOM	2	CA	GLN	A	39	23.581	-26.768	24.416	1.00	60.98		C
ATOM	3	C	GLN	A	39	23.990	-25.379	23.905	1.00	59.98		C
ATOM	4	O	GLN	A	39	25.070	-25.209	23.330	1.00	60.25		O
ATOM	5	CB	GLN	A	39	23.136	-26.685	25.878	1.00	60.69		C
ATOM	6	N	VAL	A	40	23.115	-24.395	24.122	1.00	59.58		N
ATOM	7	CA	VAL	A	40	23.342	-23.010	23.690	1.00	57.26		C
ATOM	8	C	VAL	A	40	24.000	-22.152	24.778	1.00	56.00		C
ATOM	9	O	VAL	A	40	23.992	-20.920	24.692	1.00	55.53		O
ATOM	10	CB	VAL	A	40	22.015	-22.337	23.275	1.00	57.32		C

PDB

```

loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.auth_atom_id
_atom_site.type_symbol
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_seq_id
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.pdbx_PDB_model_num
_atom_site.occupancy
_atom_site.pdbx_auth_alt_id
_atom_site.B_iso_or_equiv

```

ATOM	1	N	N	GLN	A	39	24.690	-27.754	24.275	1	1.000	.	60.760
ATOM	2	CA	C	GLN	A	39	23.581	-26.768	24.416	1	1.000	.	60.980
ATOM	3	C	C	GLN	A	39	23.990	-25.379	23.905	1	1.000	.	59.980
ATOM	4	O	O	GLN	A	39	25.070	-25.209	23.330	1	1.000	.	60.250
ATOM	5	CB	C	GLN	A	39	23.136	-26.685	25.878	1	1.000	.	60.690
ATOM	6	N	N	VAL	A	40	23.115	-24.395	24.122	1	1.000	.	59.580
ATOM	7	CA	C	VAL	A	40	23.342	-23.010	23.690	1	1.000	.	57.260
ATOM	8	C	C	VAL	A	40	24.000	-22.152	24.778	1	1.000	.	56.000
ATOM	9	O	O	VAL	A	40	23.992	-20.920	24.692	1	1.000	.	55.530
ATOM	10	CB	C	VAL	A	40	22.015	-22.337	23.275	1	1.000	.	57.320
ATOM	11	N	N	ALA	A	41	24.560	-22.804	25.797	1	1.000	.	54.570

PDBx/mmCIF

# PDBx/mmCIF Software Support

- **Phenix and Refmac** – produce native PDBx files for deposition
- **MMDB** - macromolecular object library in CCP4
- **iotbx.cif/ucif** - CCTBx C++/Python IO library with dictionary validation
- **CCIF** – CCP4 C++ library with FORTRAN support and dictionary validation
- **CBFLib** - ANSI-C library for CIF & imgCIF files
- **mmLIB** - Python toolkit supporting CIF & mmCIF
- **BioPython** - Python toolkit for computational biology
- **PyCifRW** - Python CIF/mmCIF parsing tools
- **BioJava** - Java mmCIF IO package
- **STAR::Parser** – Perl mmCIF parser and molecular object library
- **RCSBTools** - C++/Python parsing and dictionary validation tools plus many other supporting format conversion and data management applications
- **Visualization** - **Chimera, Jmol, OpenRasMol**  
PDB actively working with community developers to help fill in missing functionalities. Two workshops scheduled in Fall 2013 ...

# Converter Program: CIFTr

<http://mmcif.pdb.org>

The screenshot shows the RCSB PDB website with a blue header bar. The header includes the RCSB PDB logo, a search bar with "PDB Exchange Dictionary" and a "Go" button, and a link to "Dictionary Home". Below the header, the page title "Dictionary Resources" is displayed in bold blue text. A detailed text block explains the use of mmCIF dictionaries for PDB entries, mentioning IUCr Core, mmCIF, Image and symmetry dictionaries, and extensions for NMR, Cryo-EM, and protein production data. A bulleted list provides links to various resources, including a red circle highlighting the "PDB" and "PDBML" links in the "Converting mmCIF to PDB and PDBML" item. Another red circle highlights the "here" link at the end of the list. At the bottom, a section titled "Dictionary Content and Representation" lists background information and a specific chapter from the International Tables for Crystallography.

**A MEMBER OF THE OPDB**  
**An Information Portal to Biological Macromolecular Structures**

PDB Home | Contact Us

Dictionary Home | PDBML Home | Software Tools Home    Search: PDB Exchange Dictionary

## Dictionary Resources

The Protein Data Bank (PDB) uses macromolecular Crystallographic Information File (mmCIF) data dictionaries to describe the information content of PDB entries. The PDB Exchange data dictionary consolidates content from a variety of crystallographic dictionaries including: the IUCr Core, mmCIF, Image and symmetry dictionaries. The PDB Exchange Dictionary also includes extensions describing NMR, Cryo-EM, and protein production data. PDB data processing, data exchange, annotation, and database management operations all make heavy use of the data format and the content of the PDB Exchange Dictionary. Software tools are used to convert mmCIF data files to the older PDB format and to PDBML/XML.

- [mmCIF/PDBx example files for PDB entries containing large structures](#)
- [PDBx/mmCIF tutorial](#) for the protein structure modules of BioJava
- Data files in mmCIF format can be downloaded from the [RCSB PDB website](#) or by [ftp](#).
- Software tools are available for [preparing](#) and [editing](#) depositions.
- Software tools are available for converting mmCIF data files to [PDB](#) and [PDBML](#) formats
- A complete list of PDB software tools for managing PDB data in mmCIF format can be found [here](#).

## Dictionary Content and Representation

- [Background and Introduction](#) about mmCIF
- [Chapter 3.6. Classification and use of macromolecular data.](#) (PDF) in *International Tables for Crystallography G. Definition and exchange of crystallographic data*, S.R. Hall and B. McMahon, Editors. 2005, Springer: Dordrecht, The Netherlands. p. 144-198.
  - [Appendix 3.6.2 The Protein Data Bank exchange dictionary](#) (PDF) in *International Tables for Crystallography G. Definition and exchange of crystallographic data*, S.R. Hall and B. McMahon, Editors. 2005, Springer: Dordrecht, The Netherlands. p. 195-198.

# Converter Program: CIFTr

<http://mmcif.pdb.org>

  
A MEMBER OF THE   
**An Information Portal to Biological Macromolecular Structures**

PDB Home | Contact Us 

Software Tools Home | Dictionary Home | PDBML Home

## CIFTr

CIFTr is an application program for translating files in mmCIF format into files in PDB format. CIFTr is currently available in source form and binary form for several UNIX platforms.

CIFTr can be used to translate the files released that are in mmCIF format at <ftp://ftp.rcsb.org/pub/pdb/data/structures/>.

**Installation and usage notes for the source distribution of CIFTr**

Click here to [download source distribution of CIFTr](#).

**Installation and usage notes for the binary distribution of CIFTr**

Click here to [download binary](#) (Linux, SGI, OSF, MACOSX, or SUN) distribution of CIFTr.

© RCSB PDB

# PDB/RDF format for Semantic Web

Service from wwPDB by Akira R. Kinjo (PDBj) & Tom Oldfield (PDBe)

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

*Kinjo et al. (2012) Nucl. Acids Res. 40, D453-D460.*

WORLDWIDE  
**wwPDB**  
PROTEIN DATA BANK

Welcome to the Worldwide Protein Data Bank

**PDB/RDF**

*About PDB/RDF*  
*PDB/RDF , chem\_comp/RDF*

**PDB ID:**  (e.g., '7RSA') **PDB ID**

**property:**  (e.g., 'PDBo:entity.pdbx\_description')

**keywords:**  (e.g., 'alcohol')

Download XSLT stylesheet for converting PDBML to RDF: [PDBML2rdf.xsl.gz](#) (gzipped 22KB)

## In UniProt RDF:

```
<rdf:Description rdf:about="http://rdf.wwpdb.org/pdb/1BY4">
  <rdf:type rdf:resource="http://purl.uniprot.org/core/Structure_Resource"/>
  <database rdf:resource="http://purl.uniprot.org/database/PDB"/>
  <method rdf:resource="http://purl.uniprot.org/core/X-Ray_Crystallography"/>
  <resolution rdf:datatype="http://www.w3.org/2001/XMLSchema#float">2.10</resolution>
</rdf:Description>
```

# BMRB/XML and /RDF for Semantic Web

(developed by PDBj and BMRB)

## NMR-STAR v3

```
#####
# Entry information #
#####

save_entry_information
    _Entry.SF_category          entry_information
    _Entry.SF_framecode          entry_information
    _Entry.ID                    15400
    _Entry.Title                 ;
Backbone and side chain chemical shift assignments of the F153-to-5-fluorotryptophan mutant of human cardiac troponin C
;
    _Entry.Version_type          new
    _Entry.Submission_date       2007-07-20
    _Entry.Accession_date        2007-07-20
    _Entry.Last_release_date     .
    _Entry.Original_release_date .
    _Entry.Origination           author
    _Entry.NMR_STAR_version      3.0.8.100
    _Entry.Original_NMR_STAR_version 3.0.8.100
    _Entry.Experimental_method    NMR
    _Entry.Experimental_method_subtype solution
    _Entry.Details
    _Entry.BMRB_internal_directory_name bmr15400.str
```

BMRBxTool

## BMRB/XML

```
<BMRBx:entryCategory>
<BMRBx:entry id="15400">
<BMRBx:accession_date>2006-12-07+09:00</BMRBx:accession_date>
<BMRBx:bmrbl_internal_directory_name xsi:nil="true"/>
<BMRBx:details xsi:nil="true"/>
<BMRBx:experimental_method>NMR</BMRBx:experimental_method>
<BMRBx:experimental_method_subtype>SOLUTION</BMRBx:experimental_method_subtype>
<BMRBx:last_release_date xsi:nil="true"/>
<BMRBx:nmr_star_version>3.0.8.100</BMRBx:nmr_star_version>
<BMRBx:original_nmr_star_version>3.0.8.100</BMRBx:original_nmr_star_version>
<BMRBx:original_release_date xsi:nil="true"/>
<BMRBx:origination>author</BMRBx:origination>
<BMRBx:sf_category>entry_information</BMRBx:sf_category>
<BMRBx:sf_framecode>entry_information</BMRBx:sf_framecode>
<BMRBx:submission_date>2006-12-07+09:00</BMRBx:submission_date>
<BMRBx:title>Backbone and side chain chemical shift assignments of the F153-to-5-fluorotryptophan mutant of human cardiac troponin C</BMRBx:title>
<BMRBx:version_type>original</BMRBx:version_type>
</BMRBx:entry>
</BMRBx:entryCategory>
```

bmr15400.xml

## BMRB/RDF

```
<BMRBo:has_entryCategory>
<BMRBo:entryCategory rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400/entryCategory">
    <BMRBo:has_entry>
        <BMRBo:entry rdf:about="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400/entry/15400">
            <BMRBo:of_datablock rdf:resource="http://bmrpub.protein.osaka-u.ac.jp/rdf/bmr15400"/>
            <BMRBo:entry_id>15400</BMRBo:entry_id>
            <BMRBo:entry.accession_date>2006-12-07+09:00</BMRBo:entry.accession_date>
            <BMRBo:entry.experimental_method>NMR</BMRBo:entry.experimental_method>
            <BMRBo:entry.experimental_method_subtype>SOLUTION</BMRBo:entry.experimental_method_subtype>
            <BMRBo:entry.nmr_star_version>3.0.8.100</BMRBo:entry.nmr_star_version>
            <BMRBo:entry.original_nmr_star_version>3.0.8.100</BMRBo:entry.original_nmr_star_version>
            <BMRBo:entry.origination>author</BMRBo:entry.origination>
            <BMRBo:entry.sf_category>entry_information</BMRBo:entry.sf_category>
            <BMRBo:entry.sf_framecode>entry_information</BMRBo:entry.sf_framecode>
            <BMRBo:entry.submission_date>2006-12-07+09:00</BMRBo:entry.submission_date>
            <BMRBo:entry.title>Backbone and side chain chemical shift assignments of the F153-to-5-fluorotryptophan mutant of human cardiac troponin C</BMRBo:entry.title>
            <BMRBo:entry.version_type>original</BMRBo:entry.version_type>
        </BMRBo:entry>
        </BMRBo:has_entry>
    </BMRBo:entryCategory>
</BMRBo:has_entryCategory>
```

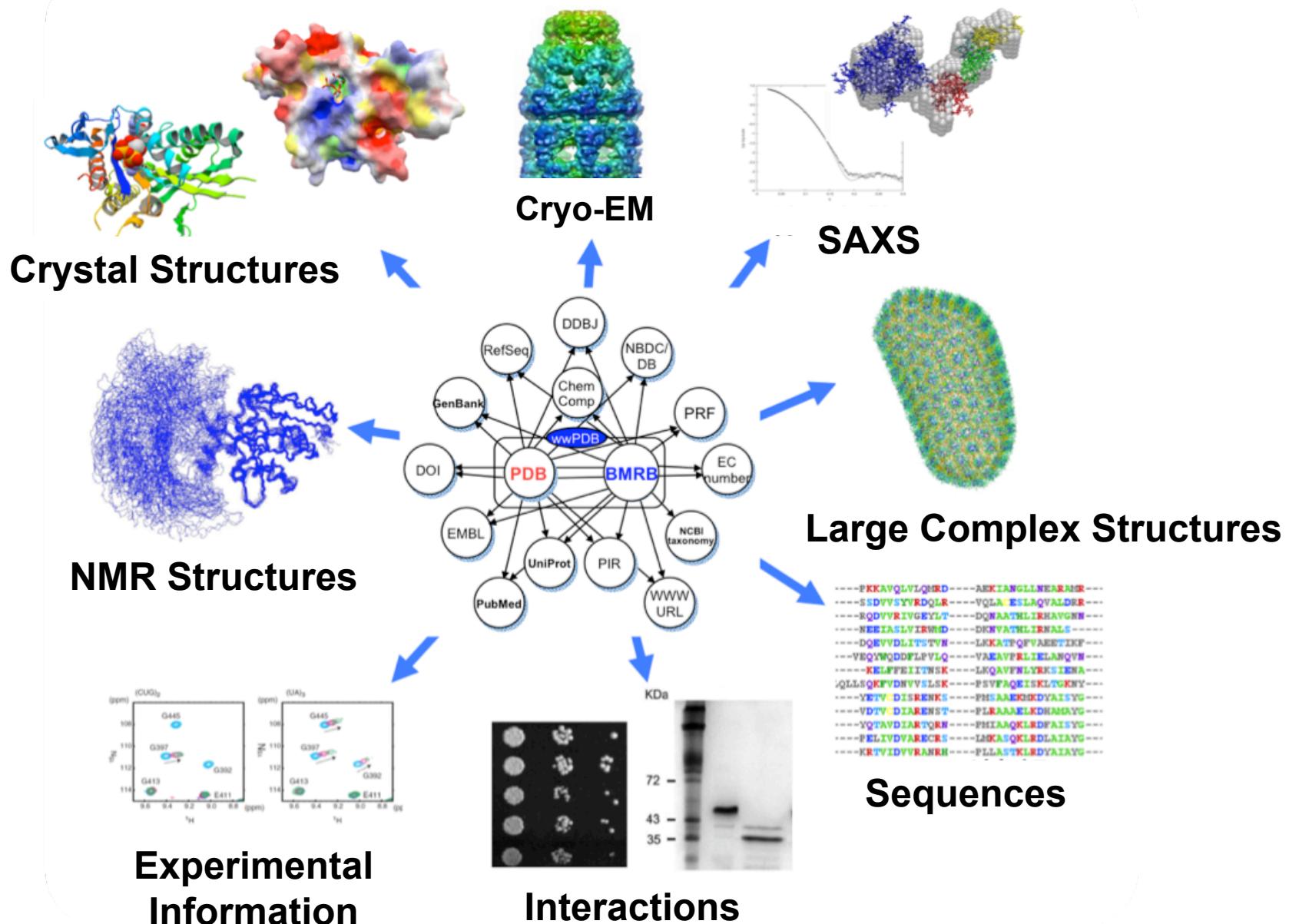
bmr15400.rdf

BMRBoTool

Format	NMR-STAR v3	BMRB/XML	BMRB/RDF
Ontology	NMR-STAR v3 dictionary	BMRB/XML schema	BMRB/OWL
Validation	ADIT-NMR & manual	Schema validation	RDF validator
Data type	Text file (.str)	XML file (.xml)	RDF file (.rdf)

Yokochi, M. et al., in preparation

# Links among PDB/RDF, BMRB/RDF and many other databases



# PDBj: Protein Data Bank Japan (2013)

- **Head**

- [Nakamura, Haruki, Ph. D.](#) (Prof., IPR, Osaka Univ.)

- **Group for PDB Database Curation**

- Nakagawa, Atsushi, Ph. D. (Group Leader, Prof., IPR, Osaka Univ.)
- Matsuda, Makoto, Ph. D. (IPR, Osaka Univ.)
- Igarashi, Reiko (IPR, Osaka Univ.)
- Kengaku, Yumiko (IPR, Osaka Univ.)
- Cho, Hasumi, Ph. D. (IPR, Osaka Univ.)
- Ikegawa, Yasuyo (IPR, Osaka Univ.)
- Sato, Junko (IPR, Osaka Univ.)

- **Group for Development of new tools and services**

- [Kinjo, Akira R., Ph. D.](#) (IPR, Osaka Univ.)
- Iwasaki, Kenji, Ph. D. (IPR, Osaka Univ.)
- Suzuki, Hirofumi, Ph. D. (IPR, Osaka Univ.)
- Yamashita, Reiko (IPR, Osaka Univ.)
- Kudou, Takahiro (IPR, Osaka Univ.)
- Nishikawa, Ken, Ph. D. (Guest Prof., IPR, Osaka Univ.)
- Bekker, Gert-Jan (IPR, Osaka Univ.)

- **Group for BMRB**

- Fujiwara, Toshimichi, Ph. D. (Group Leader, Prof. Osaka Univ.)
- Akutsu, Hideo, Ph. D. (Prof., IPR, Osaka Univ.)
- Kojima, Chojiro, Ph. D. (IPR, Osaka Univ.)
- Kobayashi, Naohiro, Ph. D. (IPR, Osaka Univ.)
- Iwata, Takeshi (IPR, Osaka Univ.)
- Takahashi, Ami (IPR, Osaka Univ.)
- Yokochi, Masashi (IPR, Osaka Univ.)

- **Collaboratory Researchers**

- [Wako, Hiroshi, Ph. D.](#) (Prof., Waseda Univ.) (for Pro Mode)
- Ito, Nobutoshi, Ph. D. (Prof., Tokyo Medical and Dental Univ.)
- [Kinoshita, Kengo, Ph.D.](#) (Prof., Tohoku Univ.) (for e F-site)
- [Standley, Daron, Ph. D.](#) (IFReC, Osaka Univ.) (for SeqNavi, StructNavi, SeSAW, and ASH)
- Katoh, Kazutaka, Ph. D. (IFReC, Osaka Univ.) (for MAFFTash)

- **Secretary**

- Haruki, Nahoko (IPR, Osaka Univ.)