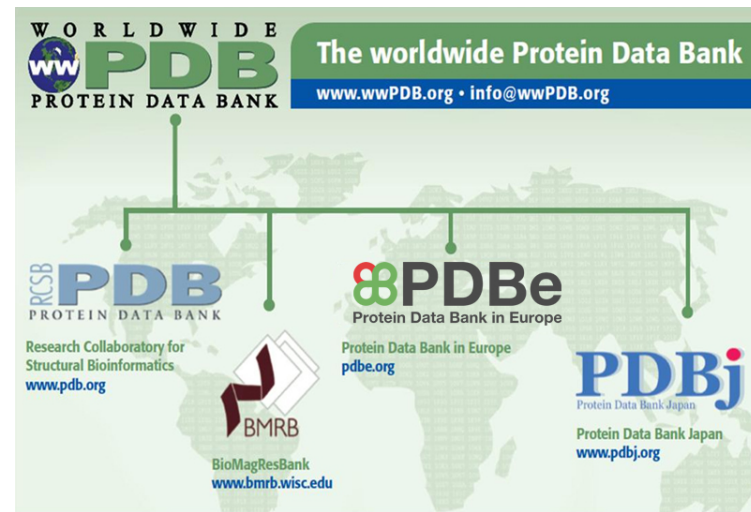




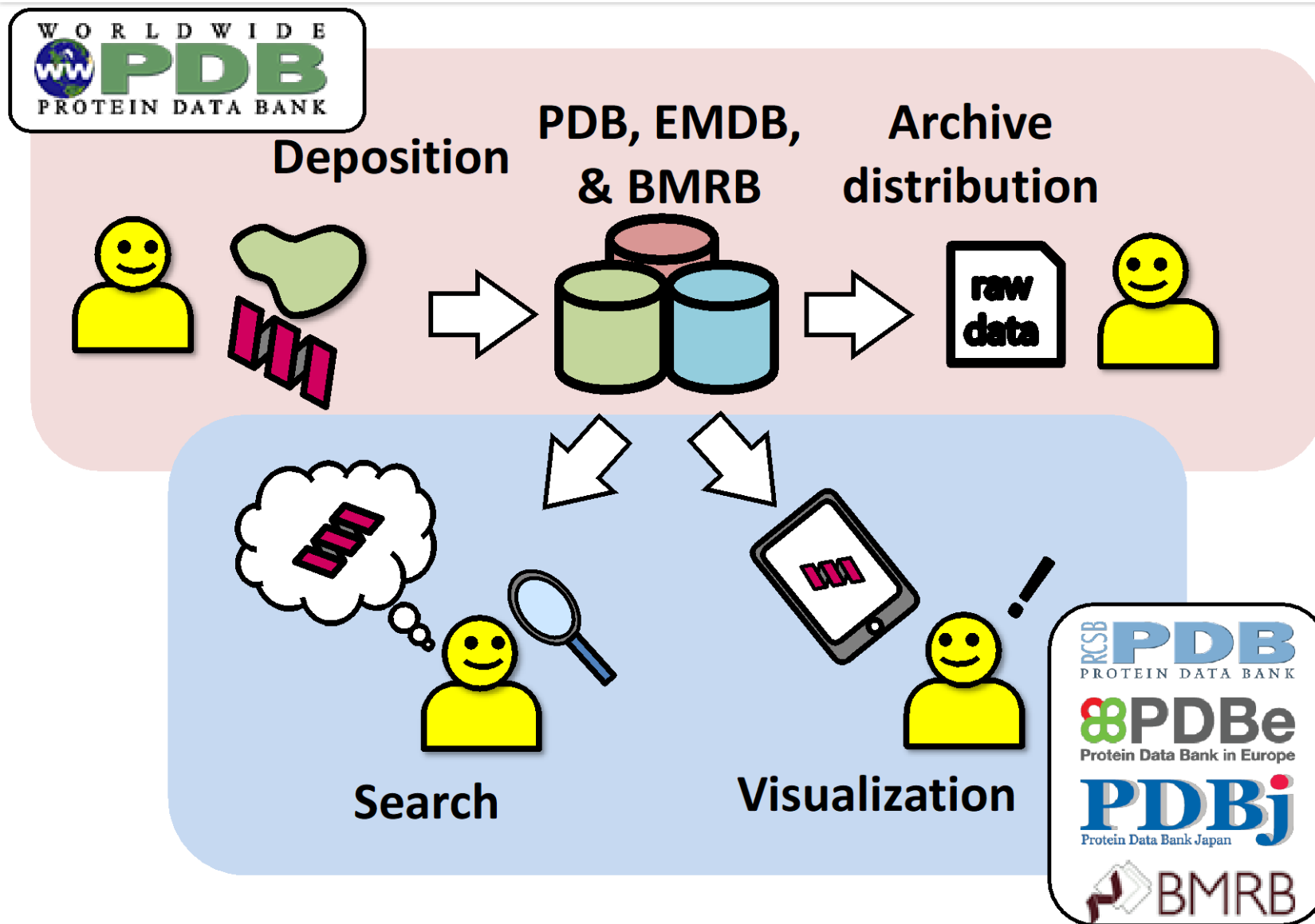
Protein Data Bank Japan

<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.



wwPDB collaboration

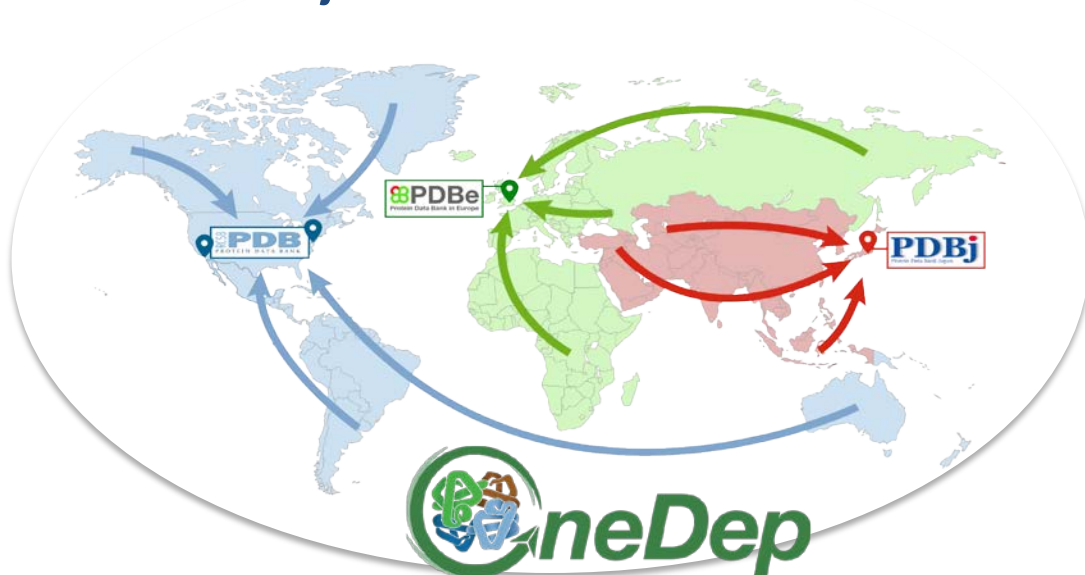


Activities/Services of each member of the wwPDB

- **“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**

PDBj is the Data Center in Asia

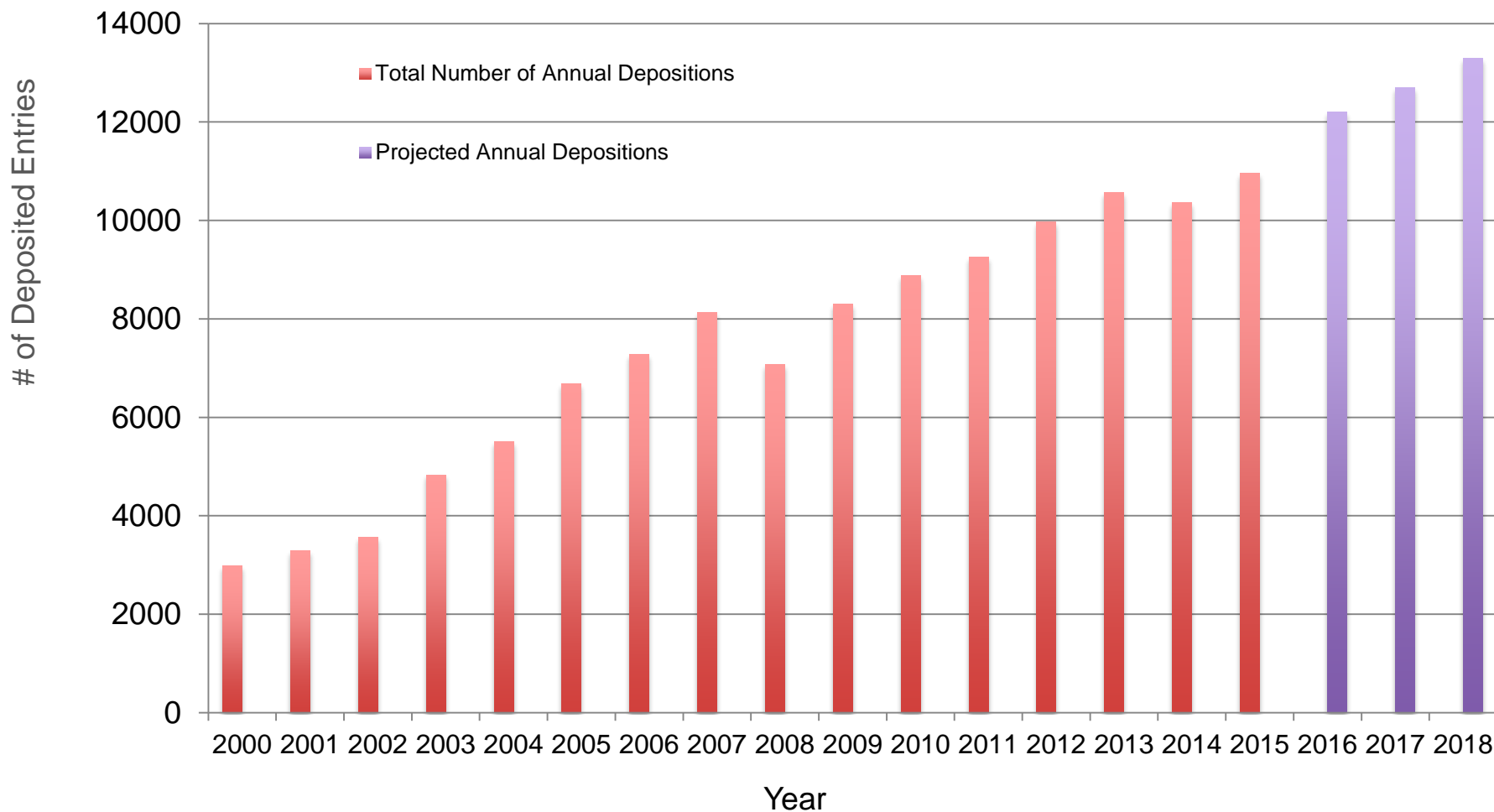
- As of 2016 region-based processing of D&A-deposited entries:
 - RCSB PDB: Americas & Oceania
 - PDBe: Europe & Africa
 - PDBj: Asia & Middle east



Year	Total Depositions	Processed By		
		RCSB PDB	PDBj	PDBe
2000	2983	2297	158	528
2001	3287	2408	383	496
2002	3565	2401	657	507
2003	4830	3135	1026	669
2004	5508	3082	1614	812
2005	6678	3563	2110	1005
2006	7282	4252	1945	1085
2007	8130	4703	2299	1128
2008	7073	4106	1994	973
2009	8300	5069	2173	1058
2010	8878	5464	2041	1373
2011	9250	5938	1816	1496
2012	9972	6408	1888	1676
2013	10566	6652	2128	1786
2014	10364	6038	1781	2545
2015	10958	4845	2100	4013
2016	11614	5326	2238	4050
2017	2577	1579	394	604
TOTAL	131815	77266	28745	25804

132K 77K 29K 26K 4

Growing Number of Depositions



Official wwPDB Validation Report is requested for peer review



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 : 2013-06-26

Resolution : 2.60 Å (reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

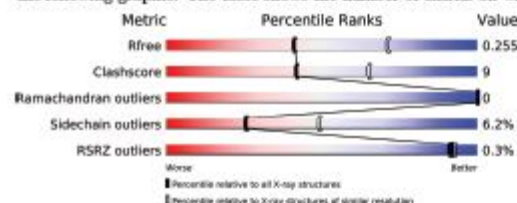
We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://www.pdb.org/ValidationPDFNotes.html>

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 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



- Molecule 2: Peptide from Sequestosome-1



Example of Validation report

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**

File Versioning: Planning Process

Current Issues:

- Loss of connection between PDB ID and Publication under current wwPDB Obsolete/Supersede Policy
- 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**
 - With PDB prefix and extension of 4 characters (e.g., from “1ABC” to “**PDB_00001ABC**”)
- Example: PDB_00001ABC_XYZ_V2-2.cif.gz

Data-out from PDBj

<http://pd bj.org/>

Amino acid sequence (FASTA)

A perfect match for your query "1gof" has been found within the [PDB] category. Alternatively you can [perform a keyword search](#).

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(s))

Biological source Hypomyces rosellus

Cellular location Secreted [Q01745](#)

Total number of polymer chains 1

Total molecular weight 68785.82

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.V., Knowles, P.F., 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. PubMed: 2002850 DOI: 10.1038/350087a0 Import into Mendeley](#)

Experimental method X-RAY DIFFRACTION (1.7 Å)

Structure validation

Metric	Percentile Ranks	Value
Clashscore		5
Ramachandran outliers		0.3%
Sidechain outliers		2.9%
RSRZ outliers		0.3%

Worse Better

Downloads

- Sequence (fasta)
- PDBx/mmCIF
- PDBML format (no atom)
- PDB format (full)
- Validation report (pdf)
- MacK...

Structures

View Asymmetric Unit (ASU)

Data base information

RCSB-PDB

PDBe

Yorodumi

CATH

PSS

SCOP

VAST

PIR-A

UniProt

201745

EGG

1.1.3.9

Expasy

1.1.3.9

IUBMB

1.1.3.9

eF-site

1gof-A

Electron Density Map (JV)

Electron Density Map (molmil)

```
>1GOF:GALACTOSE OXIDASE
ASAP1GSA1SRNNNAVTCDEAGQNECHRAIDGNKPTWTFYGCAGDPEFPPTTIDMK
TQNVNGLSHLPKQDQGNHICABEYVLSQGTWGFVADGSWADSTYKIDNETHP
ARVYKLVATTEANGQWFTSTAEINVFQASVYAPQCLGKACPTDLSFYAAALFPTF
GVLNMSYRNDAFGSGFDGTLTSSWDPSTGIVEDATVTVTKEDHFCFGISNDSGQIV
VYGNDAKRTLLTSSSDSWIPCPQWQVANGVQSRATNSDGRVTFIGSGSGGVKFERGE
VTEFEKRTKTLNPAVYRMLTARQQLFEGHNAVLYQKSGSVVQAGFSTANNTTE
CSGDVKSAGKQENRGVAFDAMCNNAVYDAVKGILTFQGSFYQSDATTHANITLG
EPQTEPNTVFANGLYFARTFTSTVVLPGSTFIFQGGKGFIFEDSTPVPFTIIVPEQ
DTPYKQFHSIVNVNLSILLPGQVFMGGGLGDCCTHNPDAQIFPFTLYENHGL
ATPRKITRTSTQGVKGRITISTOSEISKALINGYATNTVTNTQNRPLTLNNGH
SYEFQVPEDSGVALPCTNMLFYHDSAGVFVASTIRVTO
```

eF-site

1gof-A

Default CPK Default without hetero atoms

Functional site

Residue	Focus	Details
1) A:495	Yes	No
2) A:572	Yes	No
3) A:495	Yes	No
4) A:496	Yes	No
5) A:594	Yes	No
6) A:228	Yes	No
7) A:209	Yes	No
8) A:495	Yes	No
9) A:495	Yes	No
10) A:75-87	Yes	No
11) A:194	Yes	No
12) A:227-228	Yes	No
13) A:572	Yes	No

A:495

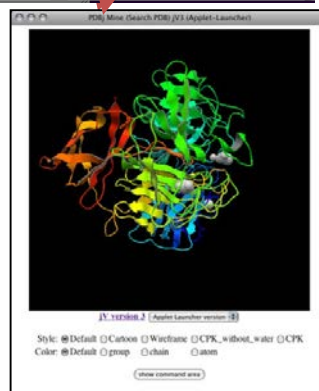
sequence Y

description Proton acceptor

source Swiss-Prot: 1

Data viewer at PDBj

Graphic viewer: **jV**
and **Molmil**
<http://pd bj.org/jV/>



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

Kinjo et al. NAR 40, D453 (2012)

ED map viewer

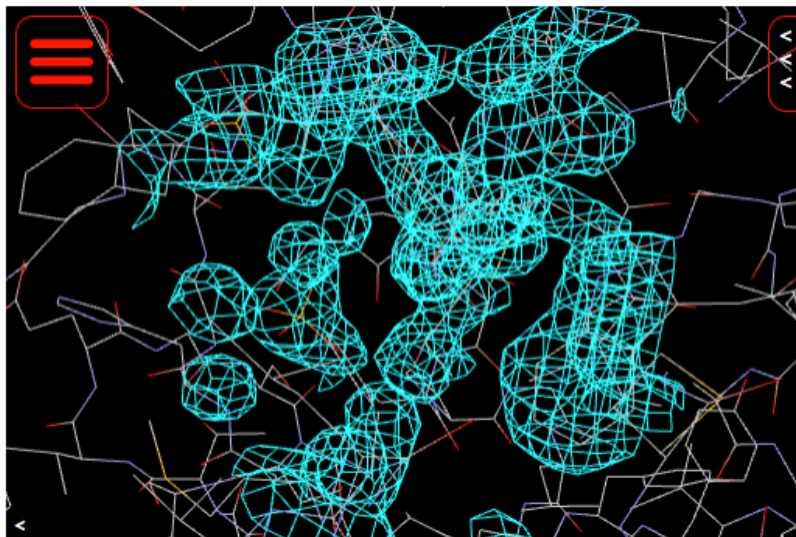
EDmap (Molmil): 2xvy

Style:

Wireframe

Color:

Atom



Parameters for Electron Density Map

[About the PDBj Electron Density Map Viewer](#)

Type of the map:

☒ contour mesh ☐ iso surface

Map position:

☒ atom nearest to the center of the map

☐ Atom ID: , ,

you can select "Atom ID" by clicking in the viewer)

☐ coordinates (x: y: z:)

mapped area: Å

(this is the length of edge of a cube)

contour level: σ

color: R: G: B:

isosurface transparency level:

Create map

Reset

Electron Density Map Download/Delete

file format

filename

structure factor r2xvysf.ent.gz

Download

refinement file 2xvy_ref.tar.gz

Download

ccp4 file 2xvy.ccp4.gz

Download

edmap file 2016518184431_2xvy.c.xml.gz

Delete

Download

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



Tools for Protein Science

New tools and functions in Data-out activities at Protein Data Bank Japan (PDBj)

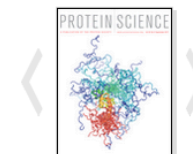
Akira R. Kinjo, Gert-Jan Bekker, Hiroshi Wako, Shigeru Endo, Yuko Tsuchiya, Hiromu Sato, Hafumi Nishi, Kengo Kinoshita, Hirofumi Suzuki, Takeshi Kawabata, Masashi Yokochi, Takeshi Iwata, Naohiro Kobayashi, Toshimichi Fujiwara, Genji Kurisu, Haruki Nakamura ✉

Accepted manuscript online: 17 August 2017 Full publication history

DOI: 10.1002/pro.3273 View/save citation

Cited by (CrossRef): 0 articles Check for updates Citation tools ▼

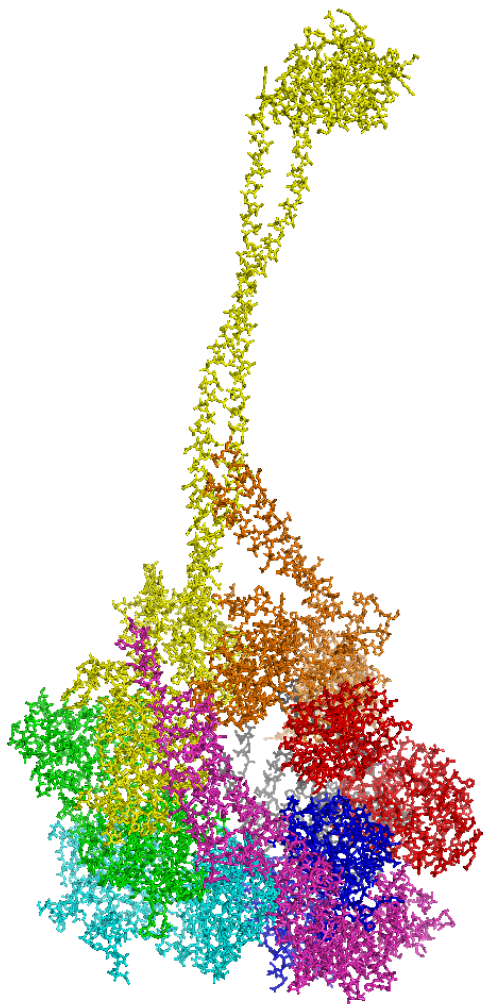
Accepted Articles



Browse Accepted Articles
Accepted, unedited articles published online and citable. The final edited and typeset version of record will appear in future.

www.pdbj.org
www.pdbj.org/advance/doi/10.1093/nar/gkw962 at Osaka Daigaku Ningen on Oct 26

X-ray crystal structure



Experimental Data

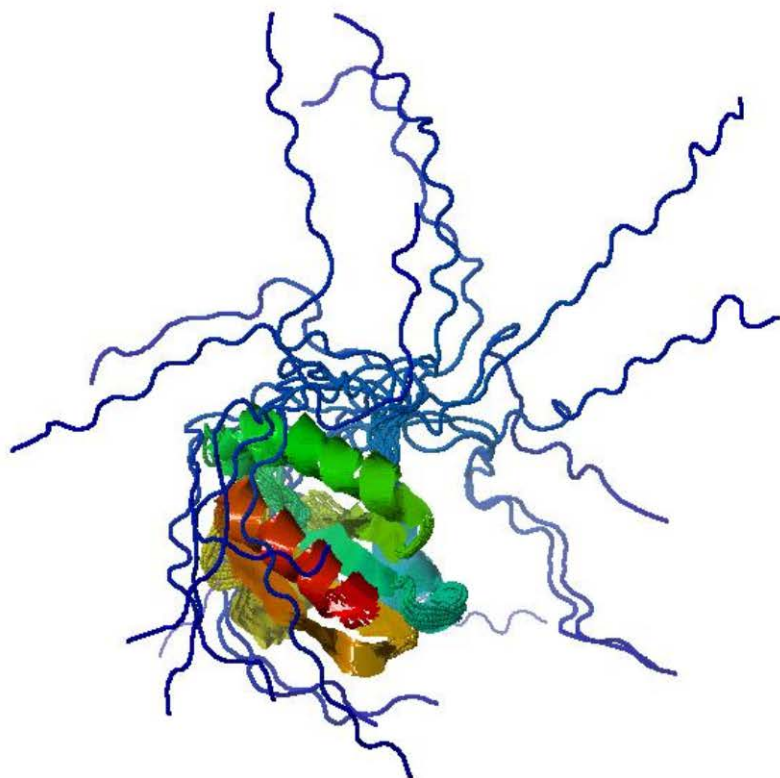
One set of Structure factors

Deposition

One coordinates file

PDB-3vkh
Cytoplasmic Dynein

Structure Ensembles by NMR



Experiment

1 sample condition
100 conformers

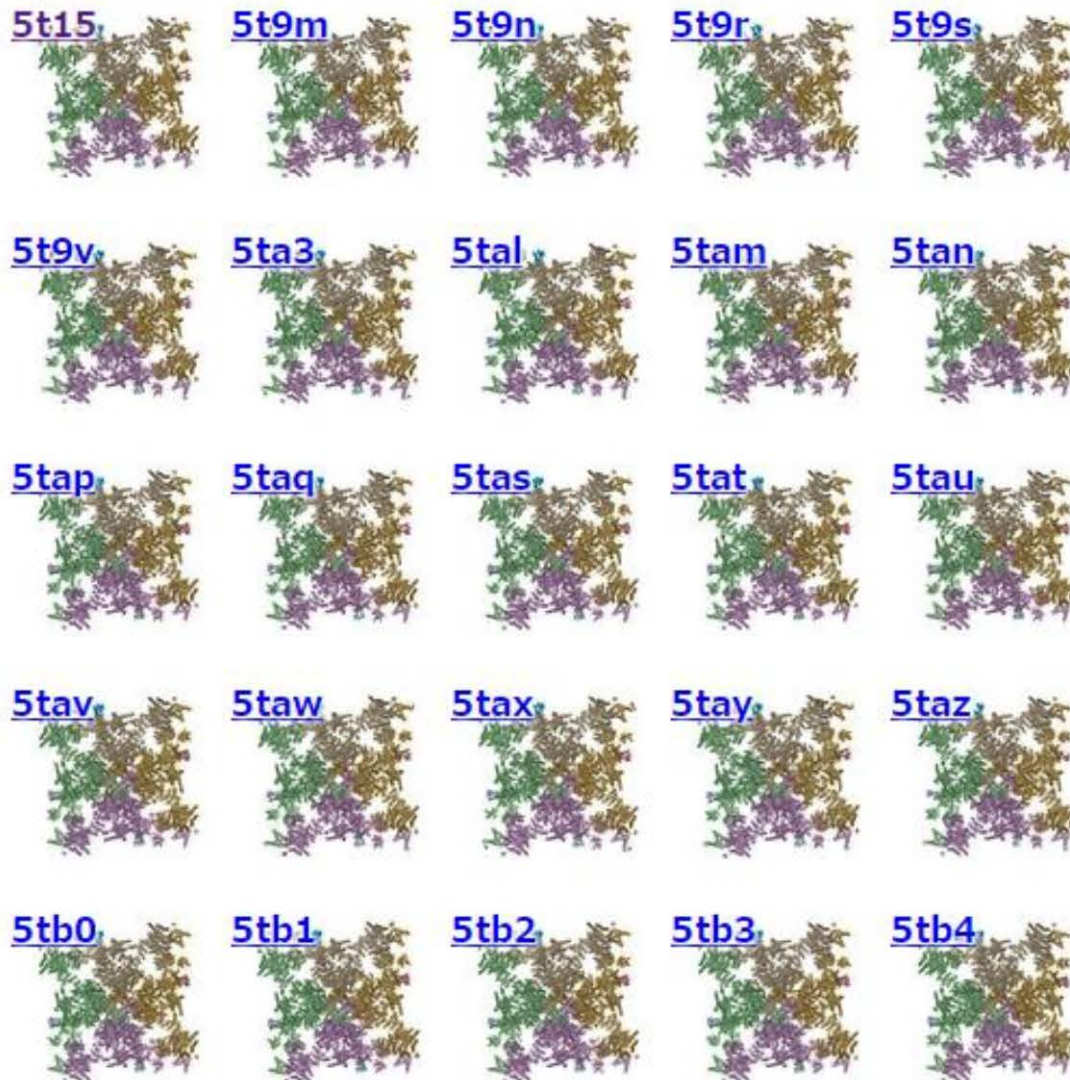
Deposition

20 conformers
1 entry

PDB-6ali

putative thioredoxin

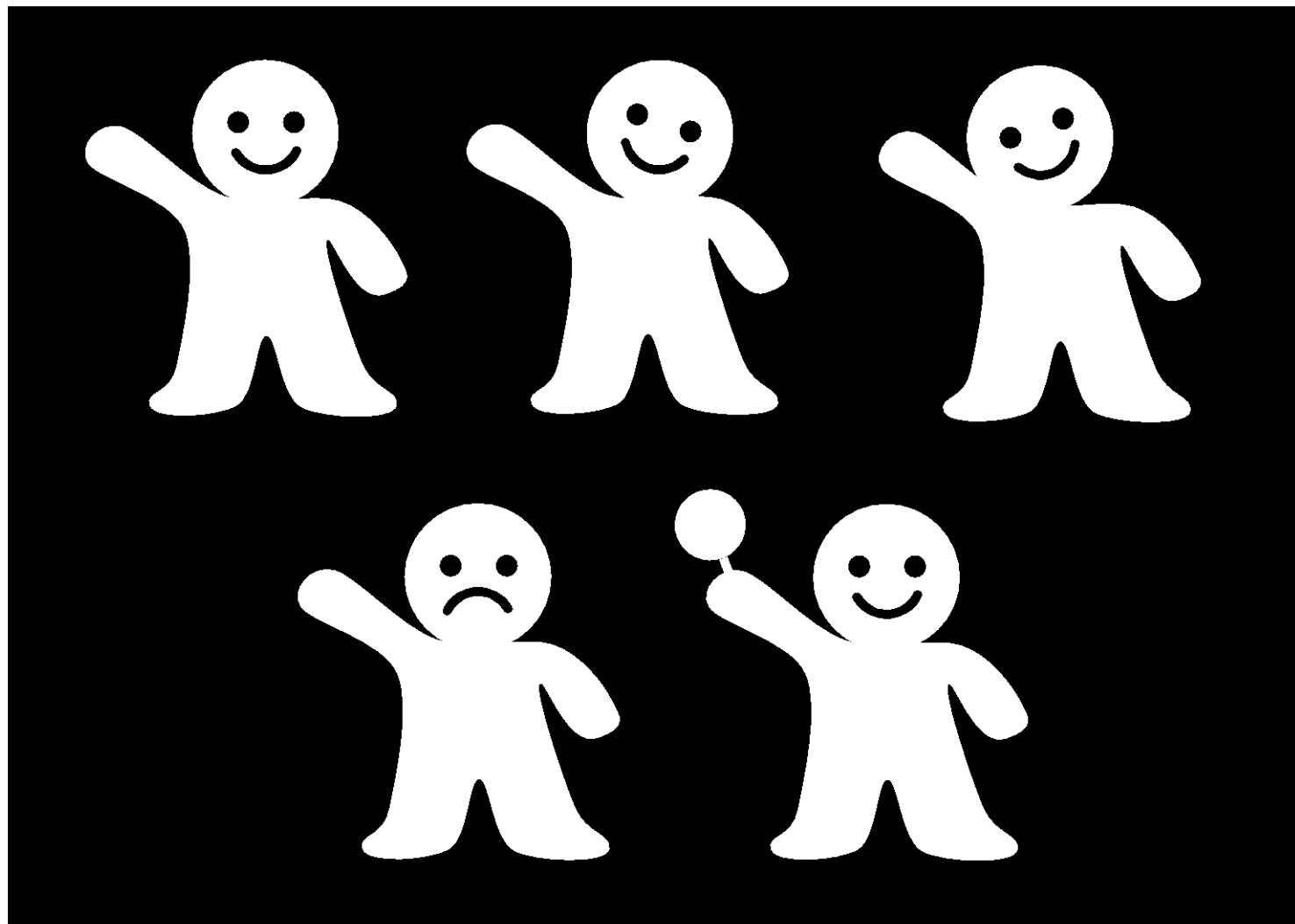
Structure Ensembles by 3DEM



5 samples
x (1 all + 4 classes)
= 25 PDB entries

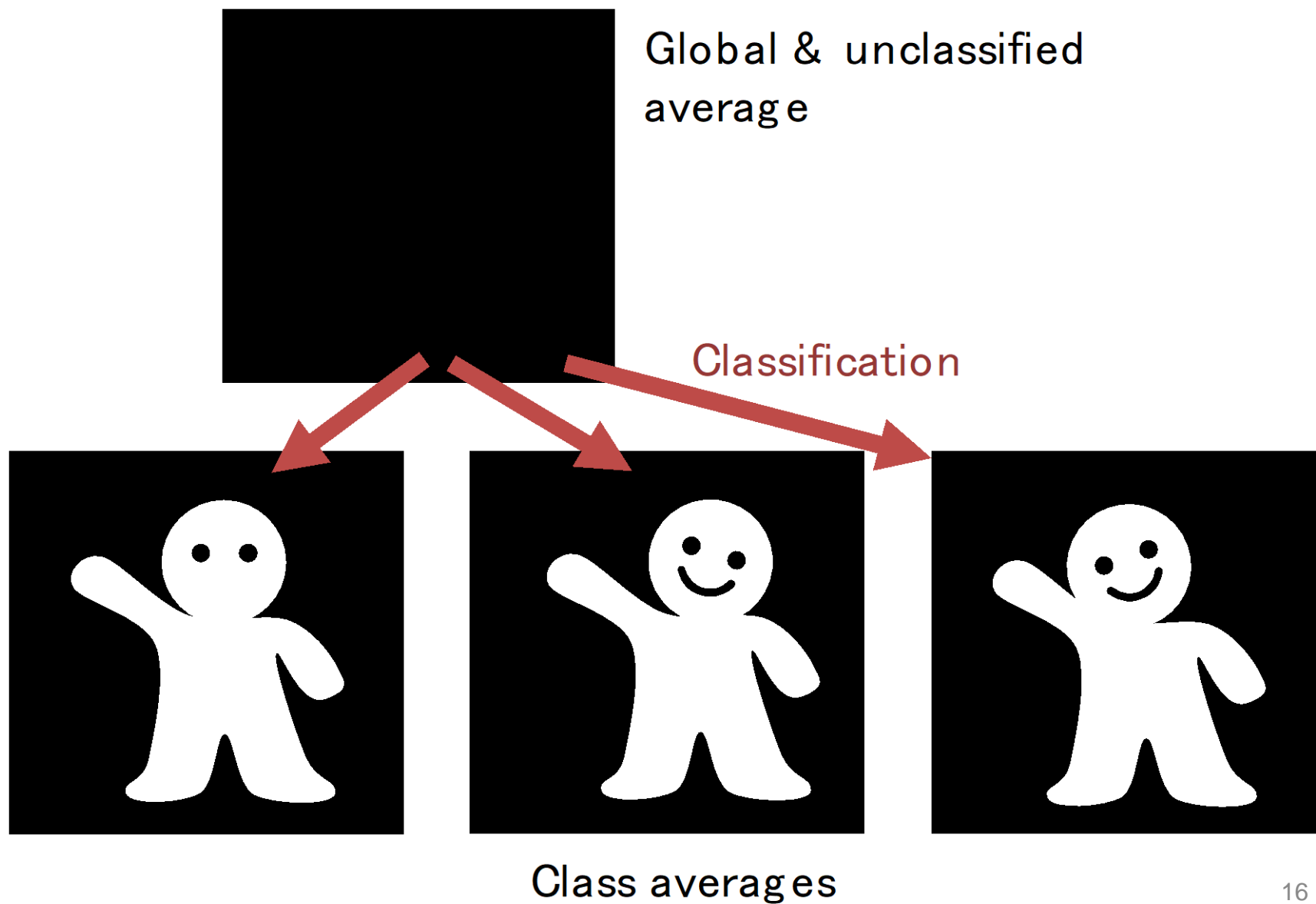
Ryanodine receptor
(calcium channel)
des Georges *et al.*,
Cell, 2016

Flexibility in Biomolecule

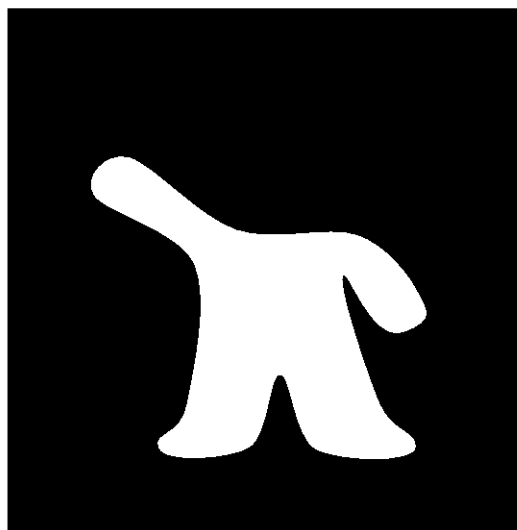


flexibility, heterogeneity & inoccupancy

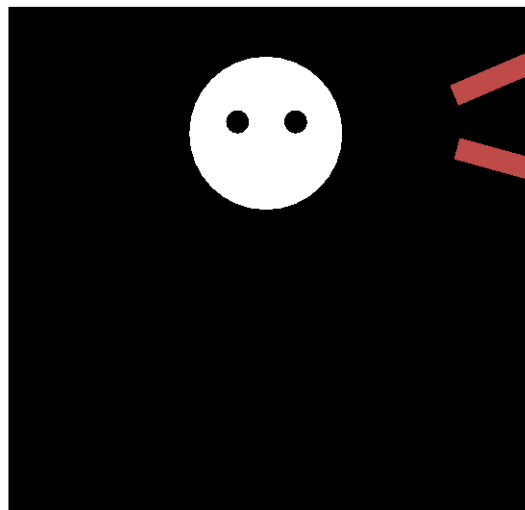
Classification



Local averaging & classification



Local (masked)
average



Classification



Local (masked)
classification

File Formats

- mmCIF
 - The canonical format of the wwPDB.
 - Ver. 5 released this year!
- 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).

wwPDB/RDF for Semantic Web

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

Yokochi et al. (2016) *J. Biomed. Semantics*, 7:16.

wwPDB/RDF

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

WORLDWIDE PDB PROTEIN DATA BANK

Welcome to the Worldwide Protein Data Bank

PDB/RDF About PDB/RDF PDB/RDF_chem_comp/RDF

PDB ID: (e.g., 'TRSA') PDB ID

property: (e.g., 'PDB entity.pdbx_description')

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

submit reset

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

BMRB/RDF

<http://bmrbpub.protein.osaka-u.ac.jp>

BMRB/XML & BMRB/RDF Publication Server

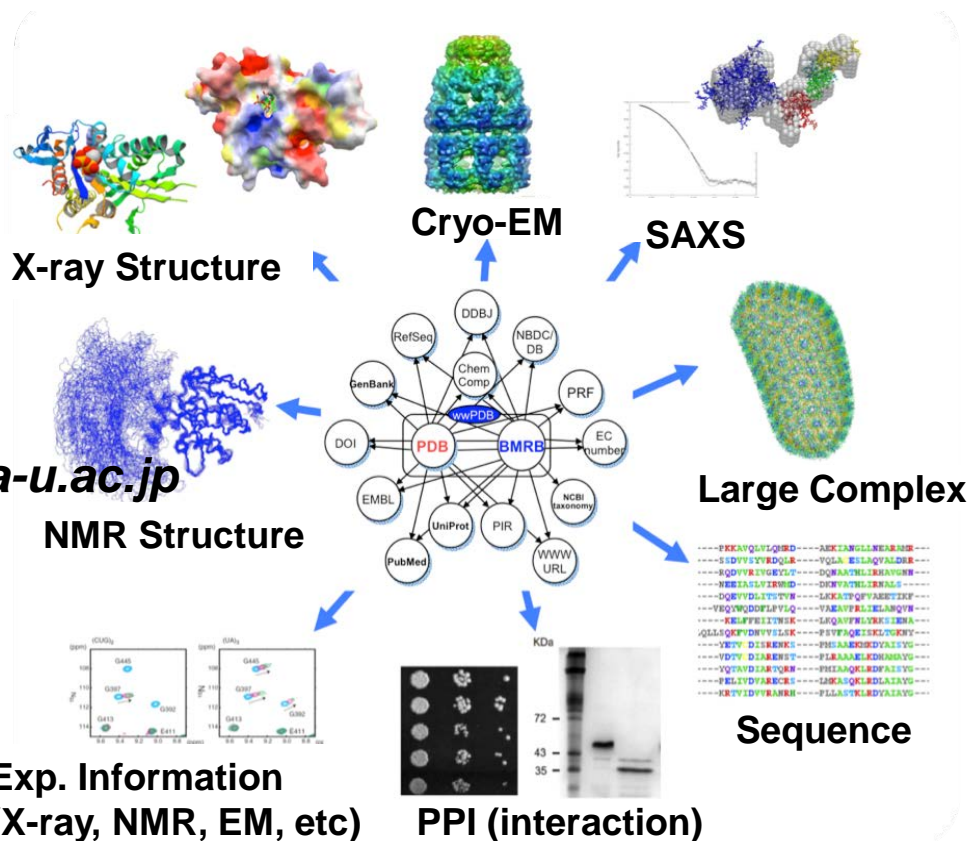
NMR Spectroscopic Data of BioMagResBank in XML and RDF formats

Entry ID: bmrb16761

BMRB/XML (compsets) BMRB/XML (sequences) BMRB/RDF

File: bmrb16761.rdf

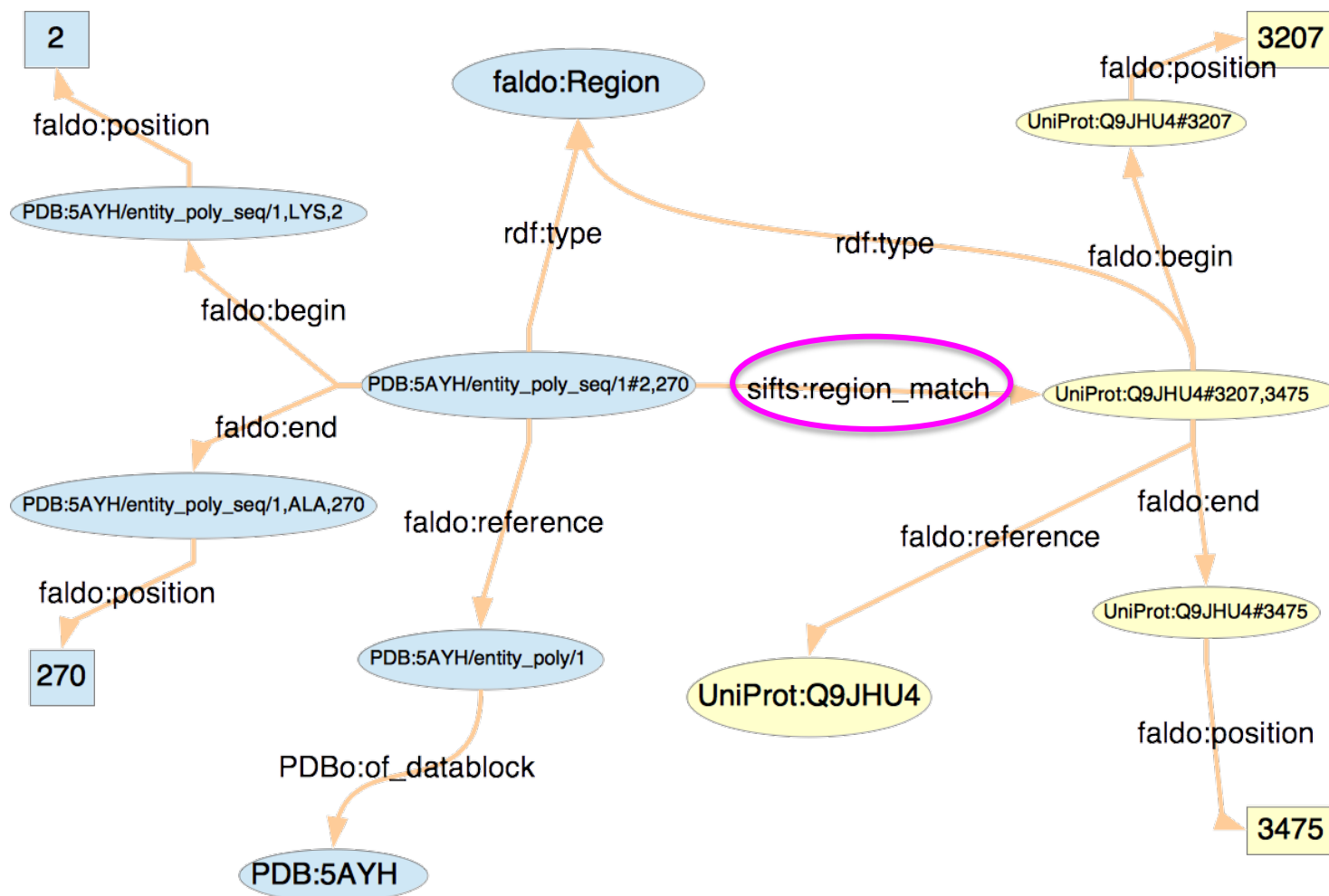
Search BMRB Entries



```
<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>
```

Usage of SIFTS information

(Structure integration with function, taxonomy and sequence)



SIFTS/RDF has been developed with EBI group: FALDO (Feature Annotation Location Description Ontology: Bolleman et al, 2016) is used at BioHackathon

mmCIF: an example

The screenshot shows the PDBj website interface for the entry 1gof. The main content area is titled 'Resources' and lists various file formats and their corresponding file names and sizes. The 'mmCIF' format is highlighted with a pink oval, and a pink arrow points to it from the right sidebar.

フォーマット	ファイル名 (ファイルサイズ)	操作
PDB	全ての情報: pdb1gof.ent.gz (109.89 KB)	画面表示
	全ての情報 (非圧縮): pdb1gof.ent (456.73 KB)	画面表示
	ヘッダのみ: pdb1gof.ent.gz (7.93 KB)	画面表示
mmCIF	1gof.cif.gz (140.02 KB)	画面表示
PDBML	全ての情報: 1gof.xml.gz (214.1 KB)	画面表示
	ヘッダのみ: 1gof-noatom.xml.gz (36.35 KB)	画面表示
	座標情報のみ: 1gof-extatom.xml.gz (120.21 KB)	画面表示
PDBMLplus	全ての情報: 1gof-plus.xml.gz (217.34 KB)	画面表示
	ヘッダのみ: 1gof-plus-noatom.xml.gz (39.6 KB)	画面表示
	付加情報のみ: 1gof-add.xml.gz (3.24 KB)	画面表示
RDF	1gof.rdf.gz (26.03 KB)	画面表示
構造因子	r1gofsf.ent.gz (558.08 KB)	画面表示
生物学的単位 (PDB形式)	1gof.pdb1.gz (105.21 KB) (A) *author_defined_assembly, 1 molecule(s) (monomeric)	画面表示
PDF	1gof_validation.pdf.gz (231.36 KB)	画面表示
PDF-full	1gof_full_validation.pdf.gz (296.43 KB)	画面表示
検証レポート	1gof_validation.xml.gz (32.37 KB)	画面表示

The right sidebar contains a 'ダウンロード' (Download) section with links to various file formats: Sequence (fasta), PDB形式 (全ての情報), PDBML (ヘッダのみ (no-atom)), mmCIF, and 検証レポート (PDF). A pink arrow points from the 'mmCIF' link in this sidebar to the 'mmCIF' entry in the main table.

mmCIF basics

- Data are divided into “categories”.
 - *_category.item*
 - e.g., _entry.id “entry” is the category name, “id” is an item of the “entry” category.
 - “_entry.id 1GOF” The value of “id” item of “entry” category is “1GOF”.
- Two ways of presenting data.
 - key-value: if only one value exists for an item.
 - loop: if more than one item exists for an item.

Example of key-value pairs

_cell.entry_id	1GOF
_cell.length_a	98.000
_cell.length_b	89.400
_cell.length_c	86.700
_cell.angle_alpha	90.00
_cell.angle_beta	117.80
_cell.angle_gamma	90.00
_cell.Z_PDB	4
_cell.pdbx_unique_axis	?
#	

The last “#” is a convention to indicate the end of a category.

Example of “loop” structure

```

loop_ ← Start of a loop
_entity.id
_entity.type
_entity.src_method
_entity.pdbx_description
_entity.formula_weight
_entity.pdbx_number_of_molecules
_entity.details
_entity.pdbx_mutation
_entity.pdbx_fragment
_entity.pdbx_ec
1 polymer      man  'GALACTOSE OXIDASE'  68579.250  1    ? ? ? 1.1.3.9
2 non-polymer  syn  'COPPER (II) ION'      63.546    1    ? ? ? ?
3 non-polymer  syn  'SODIUM ION'           22.990    1    ? ? ? ?
4 non-polymer  syn  'ACETIC ACID'          60.052    2    ? ? ? ?
5 water       nat  water                18.015    316  ? ? ? ?
#

```

A list of items.
“One item per line” is just a convention.

The last “#” is a convention to indicate the end of a loop.

- Each item is whitespace-delimited.
- In the same order as the item list (above).
- Use quotes (') for data with whitespace.

Atomic coordinates

```

loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.Cartn_x_esd
_atom_site.Cartn_y_esd
_atom_site.Cartn_z_esd
_atom_site.occupancy_esd
_atom_site.B_iso_or_equiv_esd
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
ATOM      1      N   N   . ALA  A  1  1  ? 38.840  0.236  1.012  1.00 34.65 ? ? ? ? ? ? 1  ALA  A  N   1
ATOM      2      C   CA  . ALA  A  1  1  ? 38.356 -0.999  0.357  1.00 42.26 ? ? ? ? ? ? 1  ALA  A  CA  1
ATOM      3      C   C   . ALA  A  1  1  ? 37.098 -1.547  1.056  1.00 41.25 ? ? ? ? ? ? 1  ALA  A  C   1
ATOM      4      O   O   . ALA  A  1  1  ? 36.619 -0.946  2.028  1.00 29.44 ? ? ? ? ? ? 1  ALA  A  O   1
ATOM      5      C   CB  . ALA  A  1  1  ? 39.398 -2.114  0.379  1.00 40.70 ? ? ? ? ? ? 1  ALA  A  CB  1
ATOM      6      N   N   . SER  A  1  2  ? 36.610 -2.666  0.495  1.00 32.67 ? ? ? ? ? ? 2  SER  A  N   1
ATOM      7      C   CA  . SER  A  1  2  ? 35.411 -3.244  1.202  1.00 34.90 ? ? ? ? ? ? 2  SER  A  CA  1
ATOM      8      C   C   . SER  A  1  2  ? 35.683 -4.740  1.081  1.00 38.30 ? ? ? ? ? ? 2  SER  A  C   1
ATOM      9      O   O   . SER  A  1  2  ? 36.827 -5.147  0.747  1.00 28.59 ? ? ? ? ? ? 2  SER  A  O   1
ATOM     10      C   CB  . SER  A  1  2  ? 34.063 -2.660  0.823  1.00 24.49 ? ? ? ? ? ? 2  SER  A  CB  1
ATOM     11      O   OG  . SER  A  1  2  ? 33.031 -3.308  1.686  1.00 20.37 ? ? ? ? ? ? 2  SER  A  OG  1

```

Main category groups

- `_entity` (Info about molecules)
 - `entity`, `entity_poly`, `pdbox_entity_nonpoly`, ...
- `_atom` (Info about atoms)
 - `atom_site`
- `_struct` (structural info)
 - `struct`, `struct_conf`, `struct_sheet`, `struct_conn`, `pdbox_struct_assembly`, ...
- `_chem_comp` (chemical components)
 - `chem_comp`
- `_citation` (literature info)
 - `citation`, `citation_author`, ...

Ex. PDB entries containing “HEM”

```
SELECT pdbid
FROM pdbx_entity_nonpoly
WHERE comp_id = 'HEM'
```

101M

SPERM WHALE MYOGLOBIN F46V N-BUTYL ISOCYANIDE AT PH 9.0

Summary for 101M

Descriptor MYOGLOBIN, SULFATE ION, PROTOPORPHYRIN IX CONTAINING FE, ... (5 entities in total)

Functional Keywords ligand binding, oxygen storage, oxygen binding, heme, oxygen transport

Biological source Physeter catodon (sperm whale)

Total number of polymer chains 1

Total molecular weight 18112.8

Authors Smith, R.D., Olson, J.S., Phillips Jr., G.N. (deposition date: 1997-12-13, release date: 1998-04-08, Last modification date: 2011-07-13)

Primary citation Smith, R.D. Correlations between Bound N-Alkyl Isocyanide Orientations and Pathways for Ligand Binding in Recombinant Myoglobins Thesis, Rice, 1999

Experimental method X-RAY DIFFRACTION (2.97 Å)

NMR Information

Downloads

- Sequence (FASTA)
- PDBx/mmCIF
- PDBx/mmCIF (no-atom)
- PDB format (full)
- Validation report (PDF)
- Max...

Structures

View Asymmetric Unit (AU = BU)

Database information

- RCSB-PDB
- PDBE
- Yorodumi
- CATH
- FSPP
- SCOP
- VAST
- PISA
- UniProt
- PFam
- PD0042
- CP-Info

PDBx/mmCIF: 101m - Protein Data Bank Japan

Secure | <https://pdbj.org/displayMMCIFTree?pdbid=101m>

“mmCIF tree view”

- + pdbx_xplor_file
- + struct
- + struct_keywords
- + struct_asym
- + struct_biol
- + struct_conf
- + struct_conf_type
- + struct_conn
- + struct_conn_type
- + struct_site
- + struct_site_gen
- + database_PDB_matrix
- + atom_sites
- + atom_type
- + atom_site
- + pdbx_poly_seq_scheme
- + pdbx_nonpoly_scheme
- + pdbx_struct_assembly
- + pdbx_struct_assembly_gen
- + pdbx_struct_oper_list
- + pdbx_struct_special_symmetry
- + pdbx_struct_conn_angle
- + pdbx_audit_revision_history
- + pdbx_audit_revision_details
- + pdbx_audit_revision_group
- + software
- + pdbx_validate_torsion
- pdbx_entity_nonpoly

entity_id	name	comp_id
2	SULFATE ION	SO4
3	PROTOPORPHYRIN IX CONTAINING FE	HEM
4	N-BUTYL ISOCYANIDE	NBN
5	water	HOH

“pdx_entity_nonpoly” category

Ex. PDB entries containing “HEM” by the number of HEM’s in asymmetric unit

```
SELECT a.pdbid, count(DISTINCT a.id) AS cnt
FROM pdbx_entity_nonpoly e
JOIN struct_asym a ON a.pdbid = e.pdbid
                    AND a.entity_id = e.entity_id
WHERE e.comp_id = 'HEM'
GROUP BY a.pdbid
ORDER BY cnt DESC
```

The screenshot shows the PDBj website interface. The top navigation bar includes the PDBj logo, the text 'Protein Data Bank Japan', and a search bar. Below the navigation bar, the 'SQL Search' section is active, displaying the search query entered by the user. The results section shows a total of 4198 results, with a list of entries and their corresponding counts.

SQL Search

Enter search query:

```
SELECT a.pdbid, count(DISTINCT a.id) AS cnt
FROM pdbx_entity_nonpoly e
JOIN struct_asym a ON a.pdbid = e.pdbid
                    AND a.entity_id = e.entity_id
WHERE e.comp_id = 'HEM'
GROUP BY a.pdbid
ORDER BY cnt DESC
```

Total number of results: 4198

PDB ID	cnt
4rkm	96
2j7a	84
4u8u	36
4rkn	32

Retrieve PDB chain sequence matching to the Pfam accession “PF00046” (Homeobox) and having resolution better than 2.0 Angstrom and sequence length greater than or equal to 58 (residues)

```
SELECT s.*, r.ls_d_res_high as reso,
       LENGTH(p.pdbx_seq_one_letter_code_can) as len,
       ('>' || s.pdbid || s.chain) as header,
       p.pdbx_seq_one_letter_code_can as aaseq
FROM sifts.pdb_chain_pfam s
JOIN refine r on r.pdbid = s.pdbid
JOIN entity_poly p on p.pdbid = s.pdbid
      AND s.chain = ANY(regexp_split_to_array(p.pdbx_strand_id, ','))
WHERE pfam_id = 'PF00046'
AND r.ls_d_res_high < 2.0
AND LENGTH(p.pdbx_seq_one_letter_code_can) >= 58
ORDER BY reso, len, s.chain
```

* “entity_poly.pdbx_seq_one_letter_code_can” contains AA seq in one-letter codes.

If you want to do complicated queries,
we may be able to help!

(質問はお気軽にPDBjまで)

Feel free to ask any questions at:

<https://pdbj.org/contact?tab=PDBjmaster>