PDBj
Protein Data Bank Japan

蛋白質構造データバンクのデータ検証 高度化と統合化

栗栖源嗣
大阪大学蛋白質研究所

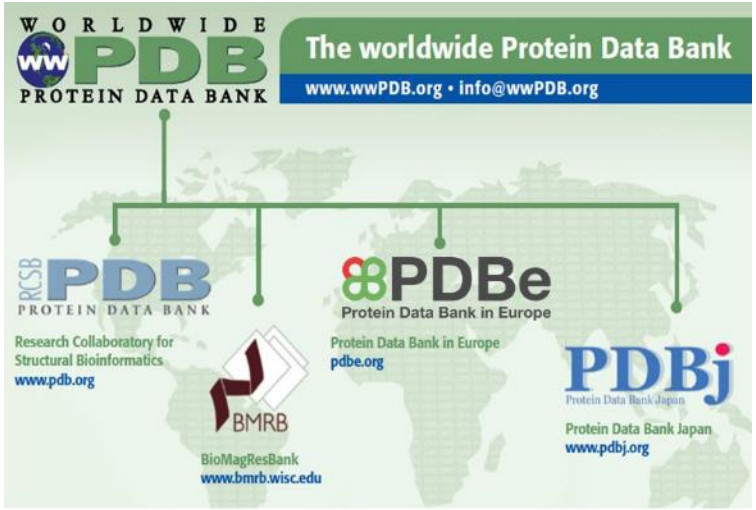
WORLDWIDE PDB
PROTEIN DATA BANK
wwpdb.org

PDBj
Protein Data Bank Japan

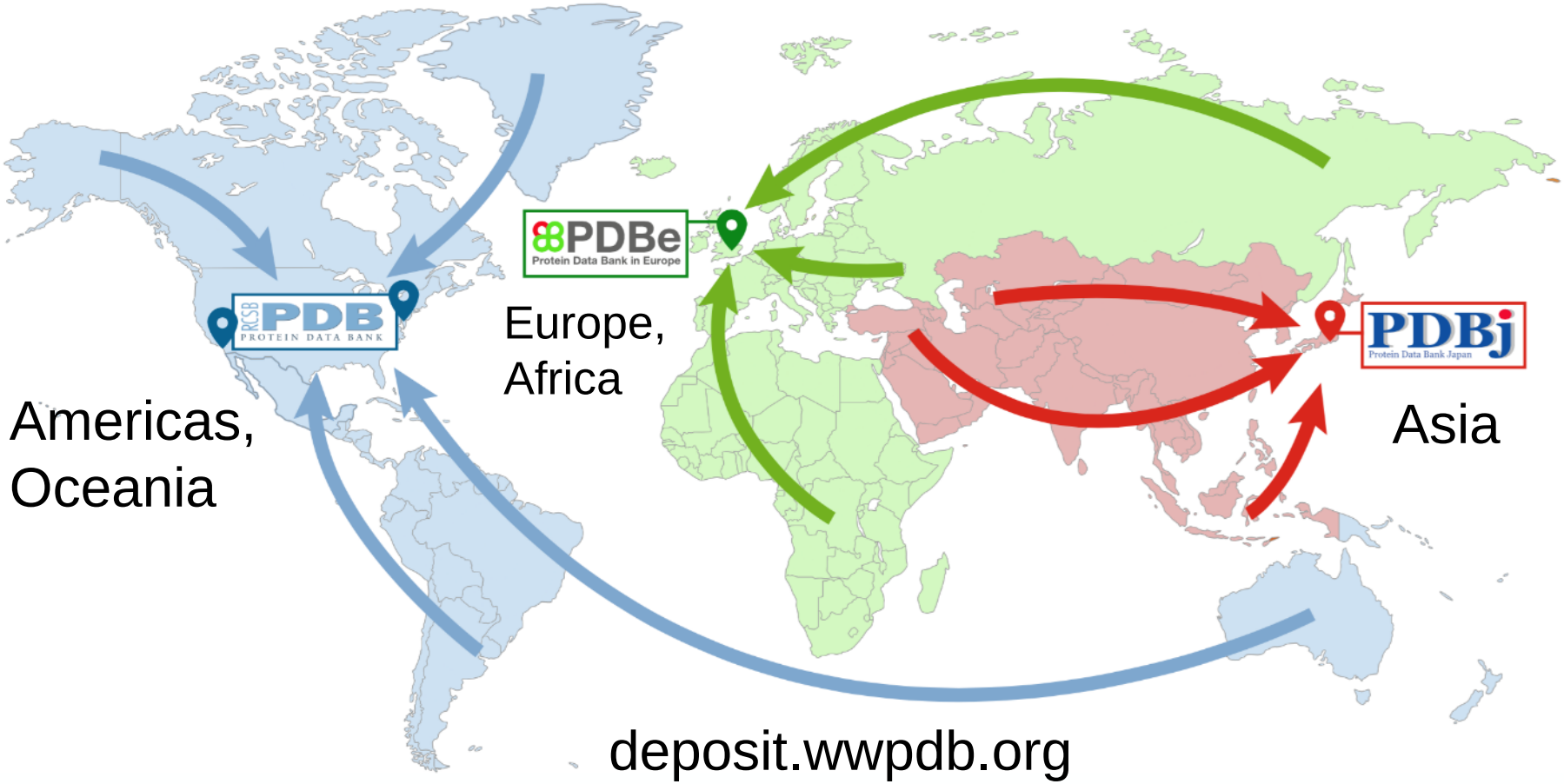
https://pdbj.org

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．

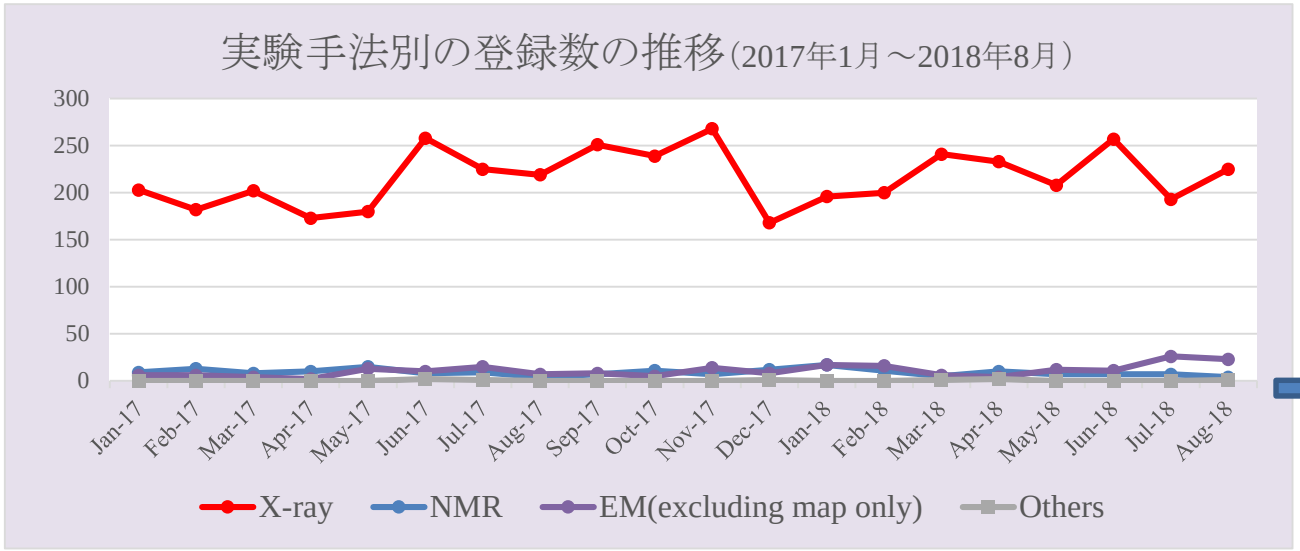
現在のPDBj運営体制（2018.8～）

＊JST-NBDC経費で雇用

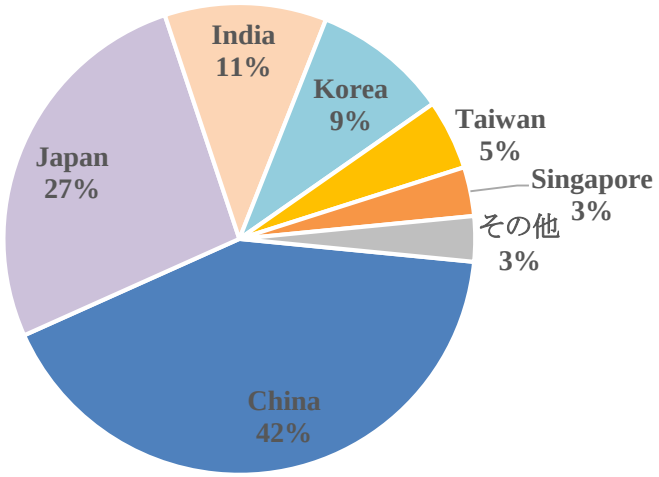
統括責任者
栗栖 源嗣 (大阪大学蛋白質研究所・教授)
PDBjデータベース構築グループ
中川 敦史 (大阪大学蛋白質研究所・教授)
見学 有美子 (大阪大学蛋白質研究所・特任研究員)＊
張 羽澄 (大阪大学蛋白質研究所・特任研究員) ＊
池川 恭代 (大阪大学蛋白質研究所・特任研究員)
佐藤 純子 (大阪大学蛋白質研究所・特任研究員) ＊
金 宙妍 (大阪大学蛋白質研究所・特任研究員)
PDBj国際的な運営高度化グループ
中村 春木 (大阪大学蛋白質研究所・教授)
金城 玲 (大阪大学蛋白質研究所・准教授)
藤 博幸 (関西学院大学理工学部・教授)
山下 鈴子 (大阪大学蛋白質研究所・特任技術専門職員)
工藤 高裕 (大阪大学蛋白質研究所・特任研究員) ＊
岩崎 憲治 (大阪大学蛋白質研究所・准教授)
鈴木 博文 (大阪大学蛋白質研究所・特任助教)
Bekker, Gert-Jan (大阪大学蛋白質研究所・特任助教)
BMRBデータベース管理運営グループ
藤原 敏道 (大阪大学蛋白質研究所・教授)
児嶋 長次郎（横浜国立大学工学部・教授）
宮ノ入 洋平 (大阪大学蛋白質研究所・准教授)
小林 直宏 (大阪大学蛋白質研究所・特任准教授)
岩田 武史 (大阪大学蛋白質研究所・特任研究員) ＊
横地 政志 (大阪大学蛋白質研究所・特任研究員) ＊
研究協力者
輪湖 博 (早稲田大学社会科学総合学術院・教授)
猿渡 茂 (北里大学理学部・准教授) for ProMode
伊藤 暢聡 (東京医科歯科大学大学院・教授)
木下 賢吾 (東北大学大学院情報科学研究科・教授)
Standley, M.Daron (大阪大学微生物病研究所・教授)
加藤 和貴 (大阪大学微生物病研究所・准教授)
事務職員
清水 朋子 (大阪大学蛋白質研究所・特任事務職員)

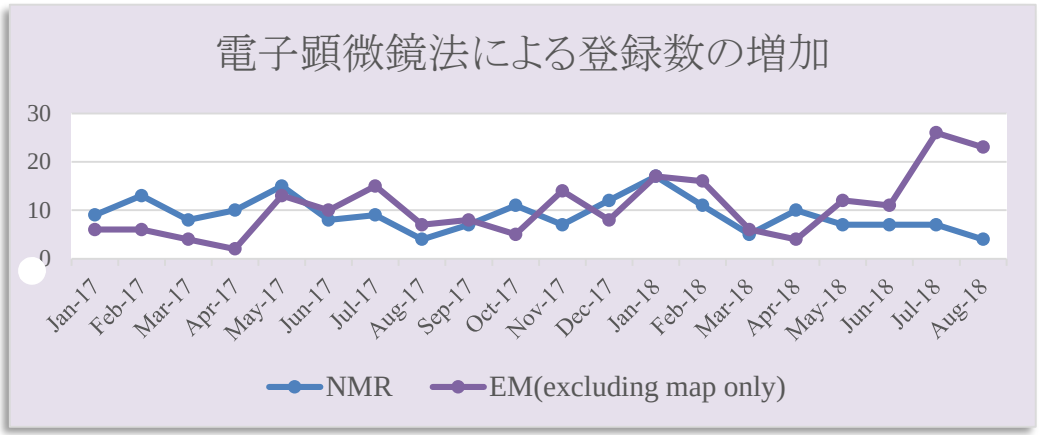
統括責任者
栗栖 源嗣 (大阪大学蛋白質研究所・教授)
PDBjデータベース構築グループ
中川 敦史 (大阪大学蛋白質研究所・教授)
見学 有美子 (大阪大学蛋白質研究所・特任研究員) ＊
張 羽澄 (大阪大学蛋白質研究所・特任研究員) ＊
池川 恭代 (大阪大学蛋白質研究所・特任研究員)
佐藤 純子 (大阪大学蛋白質研究所・特任研究員) ＊
金 宙妍 (大阪大学蛋白質研究所・特任研究員)
PDBj国際的な運営高度化グループ
藤 博幸 (関西学院大学理工学部・教授)
山下 鈴子 (大阪大学蛋白質研究所・特任技術専門職員)
工藤 高裕 (大阪大学蛋白質研究所・特任研究員) ＊
岩崎 憲治 (大阪大学蛋白質研究所・准教授)
鈴木 博文 (大阪大学蛋白質研究所・特任助教)
Bekker, Gert-Jan (大阪大学蛋白質研究所・特任助教)
BMRBデータベース管理運営グループ
藤原 敏道 (大阪大学蛋白質研究所・教授)
児嶋 長次郎（横浜国立大学工学部・教授）
宮ノ入 洋平 (大阪大学蛋白質研究所・准教授)
小林 直宏 (大阪大学蛋白質研究所・特任准教授)
岩田 武史 (大阪大学蛋白質研究所・特任研究員) ＊
横地 政志 (大阪大学蛋白質研究所・特任研究員) ＊
研究協力者
中村 春木 (大阪大学蛋白質研究所・名誉教授)
金城 玲 (ブルネイ・ダルサラーム大学・准教授)
輪湖 博 (早稲田大学社会科学総合学術院・教授)
猿渡 茂 (北里大学理学部・准教授) for ProMode
伊藤 暢聡 (東京医科歯科大学大学院・教授)
木下 賢吾 (東北大学大学院情報科学研究科・教授)
Standley, M.Daron (大阪大学微生物病研究所・教授)
加藤 和貴 (大阪大学微生物病研究所・准教授)
事務職員
清水 朋子 (大阪大学蛋白質研究所・特任事務職員)

PDBj OneDepサイトへのPDB登録数（実験手法別・国別）



同時期における登録者の国分布





PDBj Data-inシステムの特徴・今後

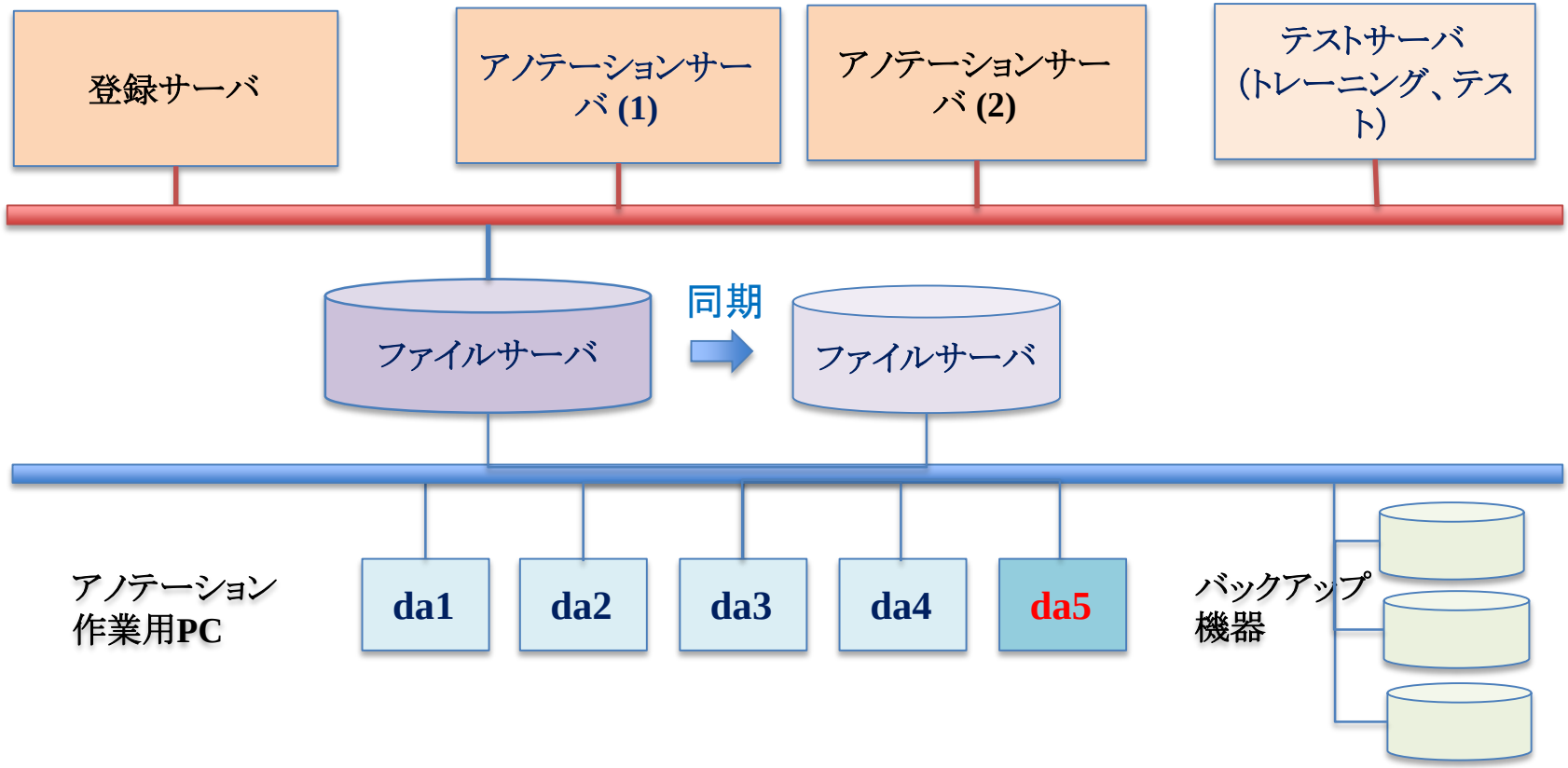
- アノテーションサーバ：高いスペックが要求される（CPU, メモリ）・・・巨大構造などのvalidationのため、外部計算機への投入も検討案
- 登録サーバ（OneDep全体も）：現在は、CentOS 6.x + Python 2.xで運用中だが、ORCiD認証の導入をふまえ、近い将来、CentOS 7.x + Python 3.xへ移行予定
- ファイルサーバ：増大し続けるデータを格納するための容量が必要（今後5年程度）・・・クラウドへの移行の可能性も検討予定
- アノテーション作業用PC：各アノテータに1台ずつ。EMエントリー作業のため、1台は高めのメモリ・グラフィクス性能

Data-In各機器のスペック情報

	機能	スペック	備考
1	登録サーバ	Xeon E5-2680 (8core 2.7GHz), 2CPU, 64GB memory, 200GB SATA SSD x 2 + 600GB SAS HDD x 2	2013- CentOS 6.x
2	アノテーションサーバ1	Xeon E5-4650 (8core 2.7GHz), 4CPU, 128GB memory, 200GB SATA SSD x 2 + 1.2TB SAS HDD x 2	2013- CentOS 6.x
3	アノテーションサーバ2	Xeon E5-2690v4 (14core 2.6GHz), 2CPU, 256GB memory, 480GB SATA SSD x 2 + 1.2TB SAS HDD x 2	2016- CentOS 6.x
4	ファイルサーバ **(Primary)**	Xeon E5-2620v4 (8core 2.1GHz), 2CPU, 128GB memory, 300GB x 2 (for OS) + 6TB x 9 (RAID6)	2017- CentOS 7.x
5	ファイルサーバ **(secondary)**	NAS 6TB x 10 (RAID5), 8GB memory	2017-
6	アノテーションサーバ （次期）	Xeon Gold 6142 (16core 2.6GHz), 2CPU, 768 GB memory, SSD 480GB x 2, HDD 2TB x 3(RAID5)	2018-準備中 CentOS 7.x
7	登録サーバ（次期）	Xeon E5-2630v3 (8core 2.4GHz), 2CPU, 128GB memory, HDD 2TB x 4（RAID10）	2016-準備中（CentOS 6→CentOS 7）
8	アノテーション作業用 **PC**	・Intel Core i5 4460 (4C 3.2GHz) 8GB memory ・Xeon E3-1245 v5 (4C 3.5GHz), 64GB memory, 2TB HDD	CentOS 6.x / CentOS 7.x

PDBj Data-inシステムの機器構成

deposit-
pdbj.wwpdb.org



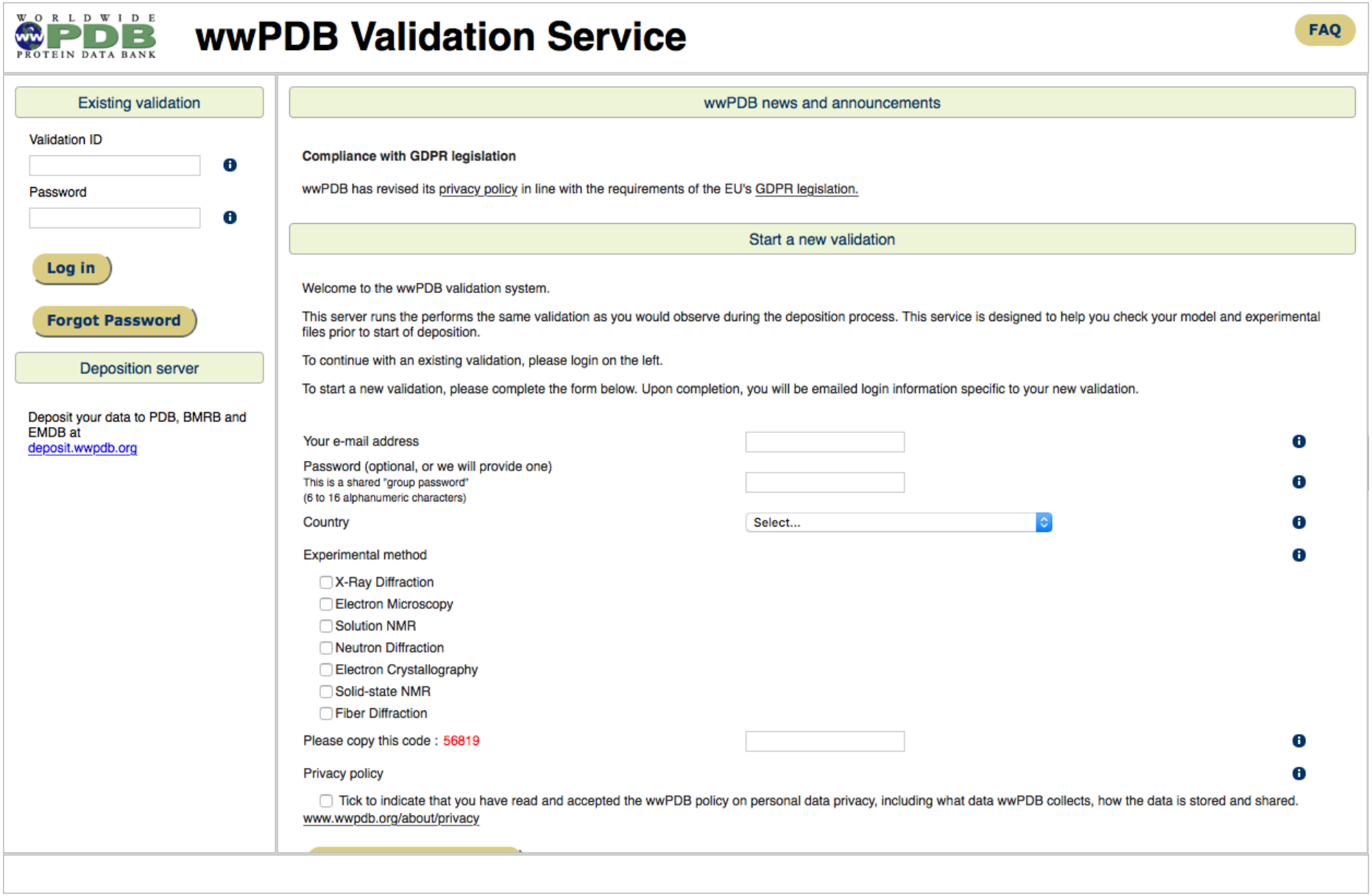
"OneDep" is current PDB deposition and annotation system used since January 2016.

Features:

- A common, web-based deposition interface
- Minimization of manual entry
- Allows submission based on existing depositions
- Enables replacement of coordinate and experimental file prior to submission and after processing
- Preview and download PDB files after submission
- Supports hybrid methods for structure determination
- mmCIF is the master file format, instead of legacy PDB format
- Improved checking for ligand chemistry and polymer sequence consistency
- Communication with PDB annotation staff using web-based interface
- Validation based on recommendations from community Task Forces.

Young, JY et al., Structure, 25(3), 536-545, 2017

Stand-alone wwPDB validation server



https://validate.wwpdb.org

Validation software utilized for generation of wwPDB validation report (2018)

Table 3. Component Software Packages Included in the 2017 Version of the Validation Pipeline

Software Package	Which Section and Metric of the Report the Package Is Used for
MolProbity	model geometry: bond lengths and bond angles of standard protein residues and nucleotides, too-close contacts, Ramachandran outliers, rotamer outliers, RNA suiteness
MAXIT	model geometry: symmetry-related too-close contacts, stereochemistry issues, identification of *cis*-peptides
Mogul *Update (2018, CSD archive)*	model geometry: bond-length and bond-angle outliers in small molecules
Xtriage (Phenix) *Update (Phenix 1.13)*	crystallographic data and refinement statistics: signal-to-noise, twinning
DCC	crystallographic data and refinement statistics: R, R_{free} fit to crystallographic data: R_{free}
EDS *Update (Recmac 7.0v44)*	fit to crystallographic data: real-space R outliers
Cyrange	NMR ensemble composition: identification of well-defined protein cores
RCI	NMR chemical shifts: prediction of protein backbone order parameter from chemical shifts
PANAV	NMR chemical shifts: suggested referencing corrections in chemical shift assignments

Percentile statistics reflecting the state of the archive on December 31st 2017.

Structure 25, 1916–1927, 2017

Validation metrics in wwPDB validation reports

X-ray/EM/NMR

- Geometric & conformational
  - bond, angle, planarity
  - protein backbone conformation
  - protein side-chain conformation
- Atomic & molecular interaction
  - all-atom contacts
  - under packing
  - hydrogen bond quality
- Non-protein
  - nucleic acids (RNA pucker, suite)
  - carbohydrates (N-glycan core)
  - ligands (CSD)
  - ions & other solvent
- Incomplete model (e.g. CA_ONLY)

X-ray

- Structure factor & electron density
  - Wilson plot, outliers, tNCS
  - wrong symmetry
  - twinning
  - agreement (R_{free}, RSR, RSCC)

NMR

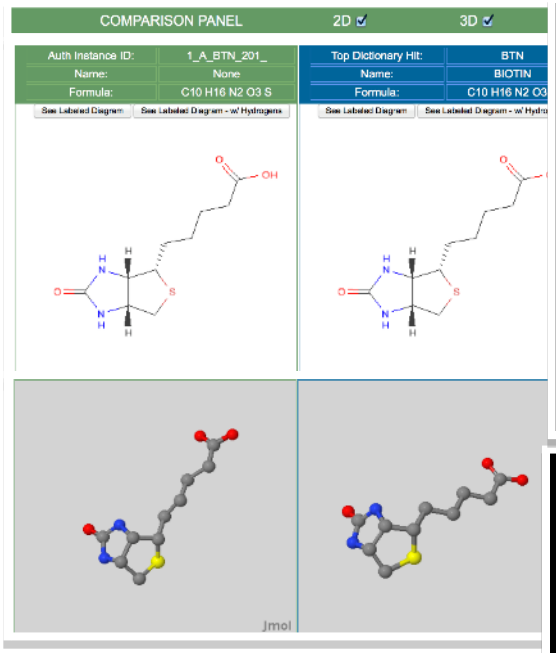
- Chemical shifts
  - completeness
  - outliers
  - estimated reference error
  - random coil index
- Structure ensembles
  - representative model (medoid)
  - domain detection

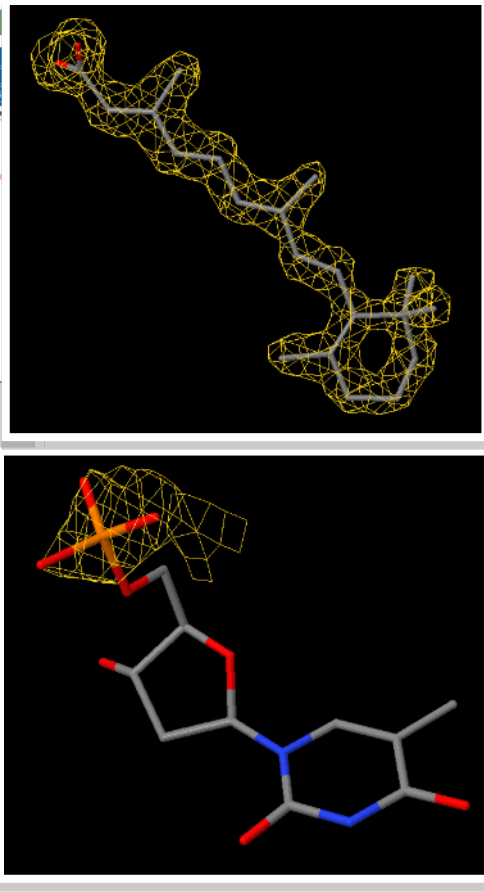
Caveat:

LLDF (Local Ligand Density Fit) has been replaced by a combination of RSR (Real-space R factor) > 0.4 and RSCC (Real-space correlations coefficient) < 0.8 since this March.

Improved Ligand Validation

- Batch search against Chemical Component Dictionary with automated CCD ID assignment
- Captures and displays author-provided chemical information
- Comparison panel
  - 2D and 3D views of ligand for review
  - ID assignment
- Display of local ligand electron density fit



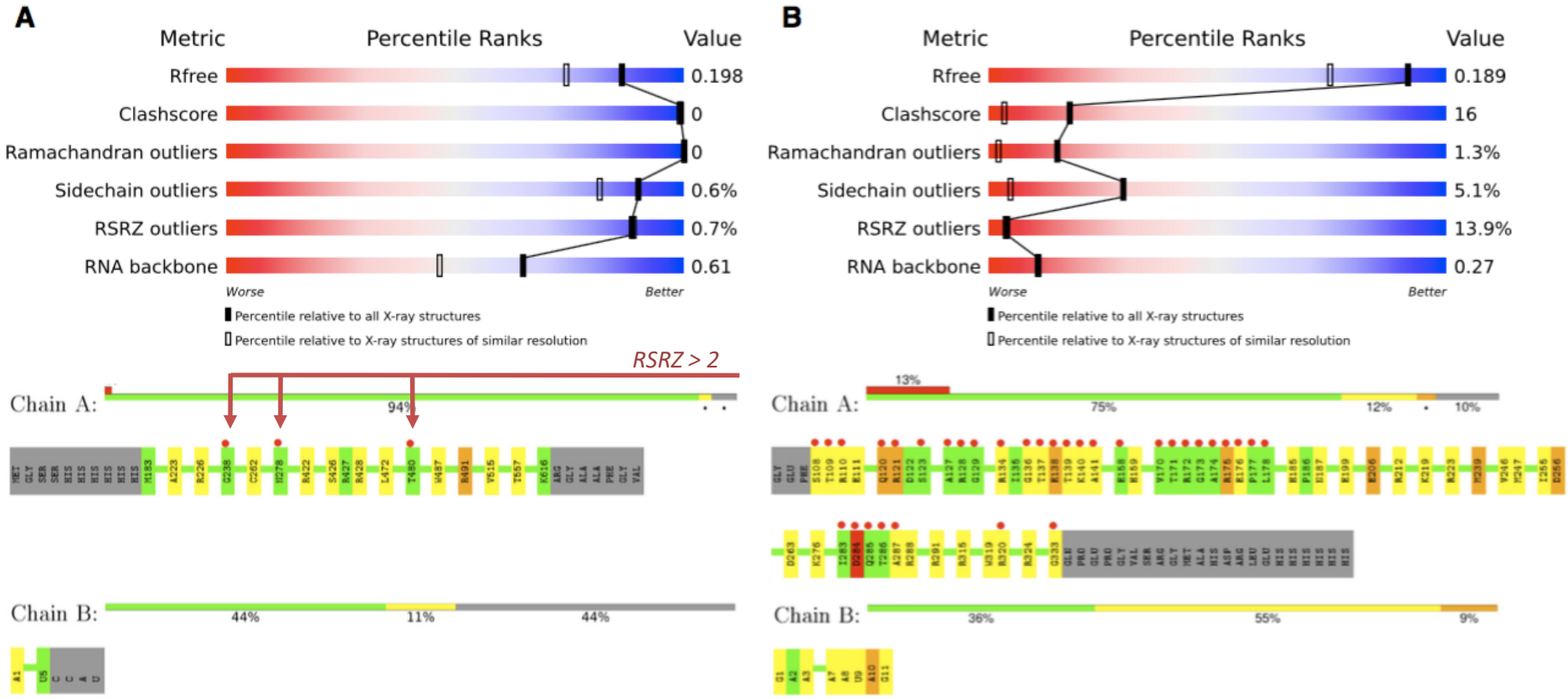


Deposited instance from coordinates (left) and the closest match in the dictionary (right)

Local ligand density display (1.5 sigma omit map)
Top: REA in entry 1CBS with LLDF=1.31 (RSR=0.10, CC=0.95)
Bottom: TMP in entry 3HW4 with LLDF=6.77 (RSR=0.41, CC=0.70)

wwPDB validation report PDFs

Summary quality metrics in wwPDB validation reports



wwPDB validation PDFs are easily reviewed and shared an assessment of structure quality.

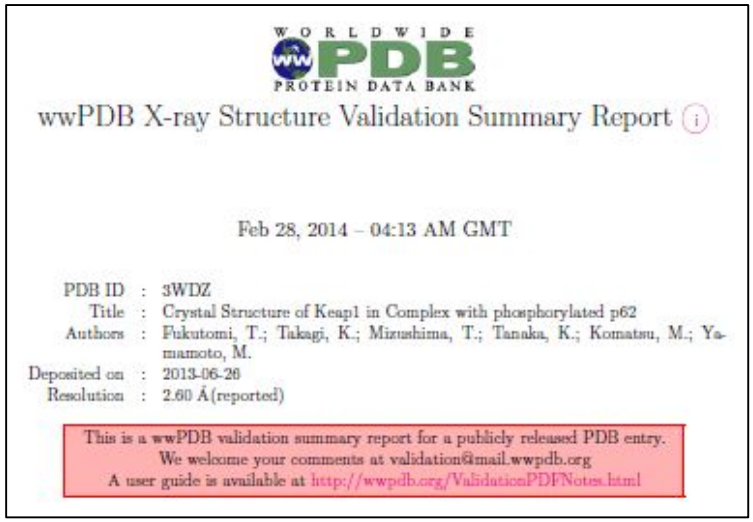
Structure 25, 1916–1927, 2017

セマンティック技術を利用した データ検証の高度化

PDBj
Protein Data Bank Japan

データの信頼性をマシン（AI）が判断できるようになり，
PDBのデータ統合化には極めて重要.

※データ検証ファイルのRDF化



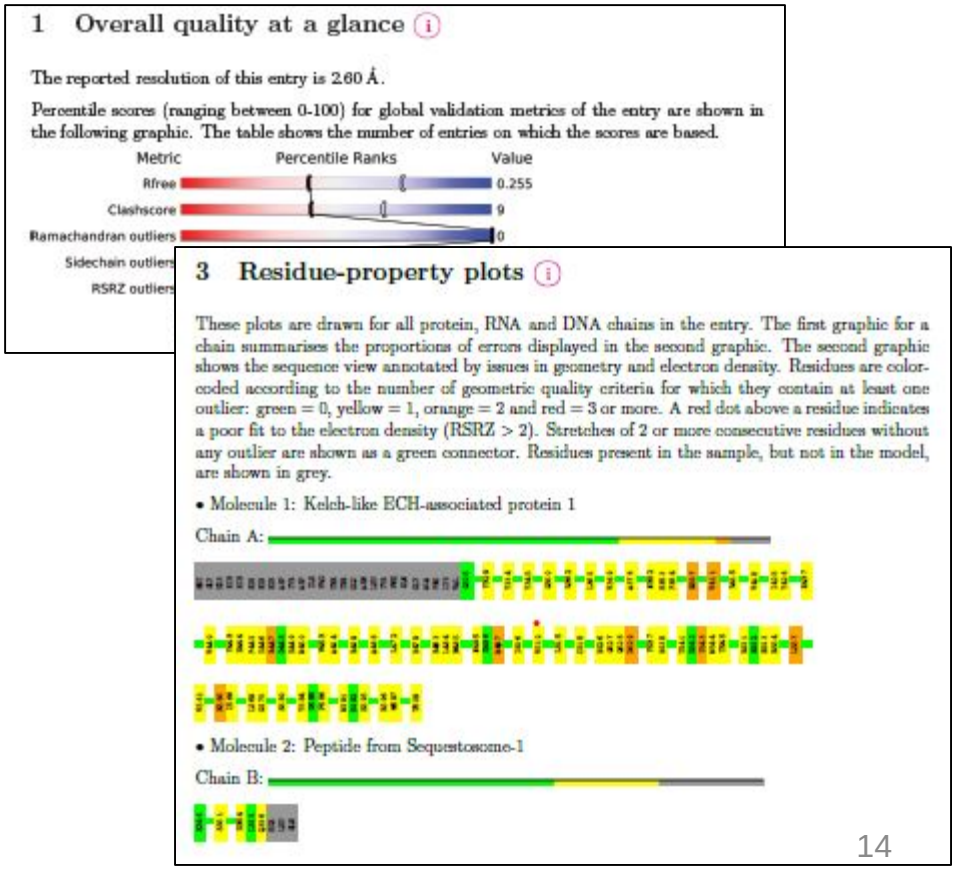
WORLDWIDE PDB
PROTEIN DATA BANK

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



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.

Metric | Percentile Ranks | Value
Rfree 0.255
Clashscore 9
Ramachandran outliers 0
Sidechain outliers
RSRZ outliers

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:

wwPDB validation report PDFs

- **Standard geometry**

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	A	0.47	0/1107	0.71	0/1491

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

- **Too close contacts**

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1091	0	1106	7	0
2	A	22	0	27	2	0
3	A	100	0	0	2	0
All	All	1213	0	1133	9	0

- **Protein backbones**

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	135/137 (98%)	132 (98%)	3 (2%)	0	100	100

- **Protein sidechains**

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	123/123 (100%)	120 (98%)	3 (2%)	52	38

- **Ligand geometry**

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	REA	A	200	-	19,22,22	1.05	1 (5%)	26,30,30	1.02	2 (7%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	200	REA	C1-C6	2.25	1.56	1.53

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	200	REA	C11-C10-C9	-2.40	123.89	127.31
2	A	200	REA	C18-C5-C6	2.08	126.83	124.51

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	200	REA	2	0

Where is the detailed validation data?

https://pdbj.org

Detailed wwPDB validation reports (XML)


```
<?xml version="1.0" ?>
<wwPDB-validation-information xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://wwpdb.org/validation/schema/
    wwpdb_validation_v002.xsd">
    <Entry B_factor_type="FULL" CCP4version="7.0 (Gargrove)" DCC_R="0.18" DCC_Rfree="0.19" DCC_refinement_program="CNS" DataAnisotropy="0.434"
        DataCompleteness="90.54" EDS_R="0.18" EDS_resolution="1.80" EDS_resolution_low="14.93" Fo_Fc_correlation="0.956" IoverSigma="3.77(1.79A)" PDB-R="0.20"
        PDB-Rfree="0.24" PDB-deposition-date="1994-09-28" PDB-resolution="1.80" PDB-resolution-low="8.00" PDB-revision-date="2011-07-13" PDB-revision-number="3"
        RefmacVersion="5.8.0158" RestypesNotcheckedForBondAngleGeometry="REA" TransNCS="The largest off-origin peak in the Patterson function is 9.26% of the
        height of the origin peak. No significant pseudotranslation is detected." TwinFraction="k,h,-l:0.027" TwinL="0.515" TwinL2="0.357"
        WilsonBaniso="[16.802,17.606,11.032,0.000,0.000,0.000]" WilsonBestimate="14.785" XMLcreationDate="Mar 10, 2018 -- 04:41 pm GMT" absolute-percentile-
        DCC_Rfree="90.4" absolute-percentile-clashscore="69.2" absolute-percentile-percent-RSRZ-outliers="100.0" absolute-percentile-percent-rama-
        outliers="100.0" absolute-percentile-percent-rota-outliers="51.8" acentric_outliers="1" angles_rmsz="0.71" attemptedValidationSteps="mogul,molprobity,
        validation-pack,xtriage,eds,percentiles,writexml" babinet_b="141.456" babinet_k="0.156" bonds_rmsz="0.47" bulk_solvent_b="72.956" bulk_solvent_k="0.401"
        centric_outliers="0" clashscore                                                                                                    high-resol-
        relative-percentile-percent-RSR                                                                                                    tile-percent-rota-
        outliers="1.8" low-resol-relati                                                                                                    tile-percent-RSRZ-
        outliers="1.8" low-resol-relati                                                                                                    m-H-reduce="1133"
        num-free-reflections="1496" num                                                                                                    lashscore="122126"
        numPDBids-absolute-percentile-percent-RSRZ-outliers="108989" numPDBids-absolute-percentile-percent-rama-outliers="120053" numPDBids-absolute-percentile-
        percent-rota-outliers="120020" numPDBids-relative-percentile-DCC_Rfree="5253" numPDBids-relative-percentile-clashscore="6077" numPDBids-relative-
        percentile-percent-RSRZ-outliers="5157" numPDBids-relative-percentile-percent-rama-outliers="6011" numPDBids-relative-percentile-percent-rota-
        outliers="6010" num_angles_rmsz="1491" num_bonds_rmsz="1107" pdbid="1CBS" percent-RSRZ-outliers="0.00" percent-free-reflections="10.19" percent-rama-
        outliers="0.00" percent-rota-outliers="2.44" percentilebins="all,1.8,xray" protein-DNA-RNA-entities="1" relative-percentile-DCC_Rfree="92.5" relative-
        percentile-clashscore="62.5" relative-percentile-percent-RSRZ-outliers="100.0" relative-percentile-percent-rama-outliers="100.0" relative-percentile-
        percent-rota-outliers="38.4" xtriage_input_columns="F_meas_au,F_meas_sigma_au"/>
    <ModelledSubgroup NatomsEDS="7" altcode=" " avgoccu="1.000" chain="A" ent="1" icode=" " model="1" num-H-reduce="9" owab="28.790" resname="PRO" resnum="1"
        rota="Cg_endo" rscc="0.908" rsr="0.143" rsrz="0.706" said="A" seq="1">
        <clash atom="HB2" cid="4" clashmag="0.47" dist="1.97"/>
    </ModelledSubgroup>
    <ModelledSubgroup NatomsEDS="8" altcode=" " avgoccu="1.000" chain="A" ent="1" icode=" " model="1" num-H-reduce="6" owab="21.720" phi="-124.5" psi="100.5"
        rama="Favored" resname="ASN" resnum="2" rota="t30" rscc="0.906" rsr="0.144" rsrz="0.485" said="A" seq="2"/>
    <ModelledSubgroup NatomsEDS="11" altcode=" " avgoccu="1.000" chain="A" ent="1" icode=" " model="1" num-H-reduce="9" owab="11.800" phi="-80.0" psi="-9.9"
        rama="Favored" resname="PHE" resnum="3" rota="m-85" rscc="0.966" rsr="0.084" rsrz="-0.490" said="A" seq="3"/>
    <ModelledSubgroup NatomsEDS="6" altcode=" " avgoccu="1.000" chain="A" ent="1" icode=" " model="1" num-H-reduce="5" owab="12.240" phi="-55.6" psi="140.2"
        rama="Favored" resname="SER" re
    <ModelledSubgroup NatomsEDS="4" alt
        rama="Favored" resname="GLY" re
```


Entry-level validation information:
Percentiles, overall validation metrics (e.g. Rfree), statistics.

Residue-level validation information:
Geometric outliers, torsion angles, RSR, RSRZ, clash score, occupancy

validation report of modeled subgroups (residues) repeat…


```
    <ModelledEntityInstance absolute_RSRZ_percentile="100.00" absolute_rama_percentile="100.00" absolute_sidechain_percentile="51.82" angles_rmsz="0.71"
        bonds_rmsz="0.47" chain="A" ent="1" model="1" num_angles_rmsz="1491" num_bonds_rmsz="1107" relative_RSRZ_percentile="100.00"
        relative_rama_percentile="100.00" relative_sidechain_percentile="38.37" said="A"/>
```


Entity-level validation information: Percentiles

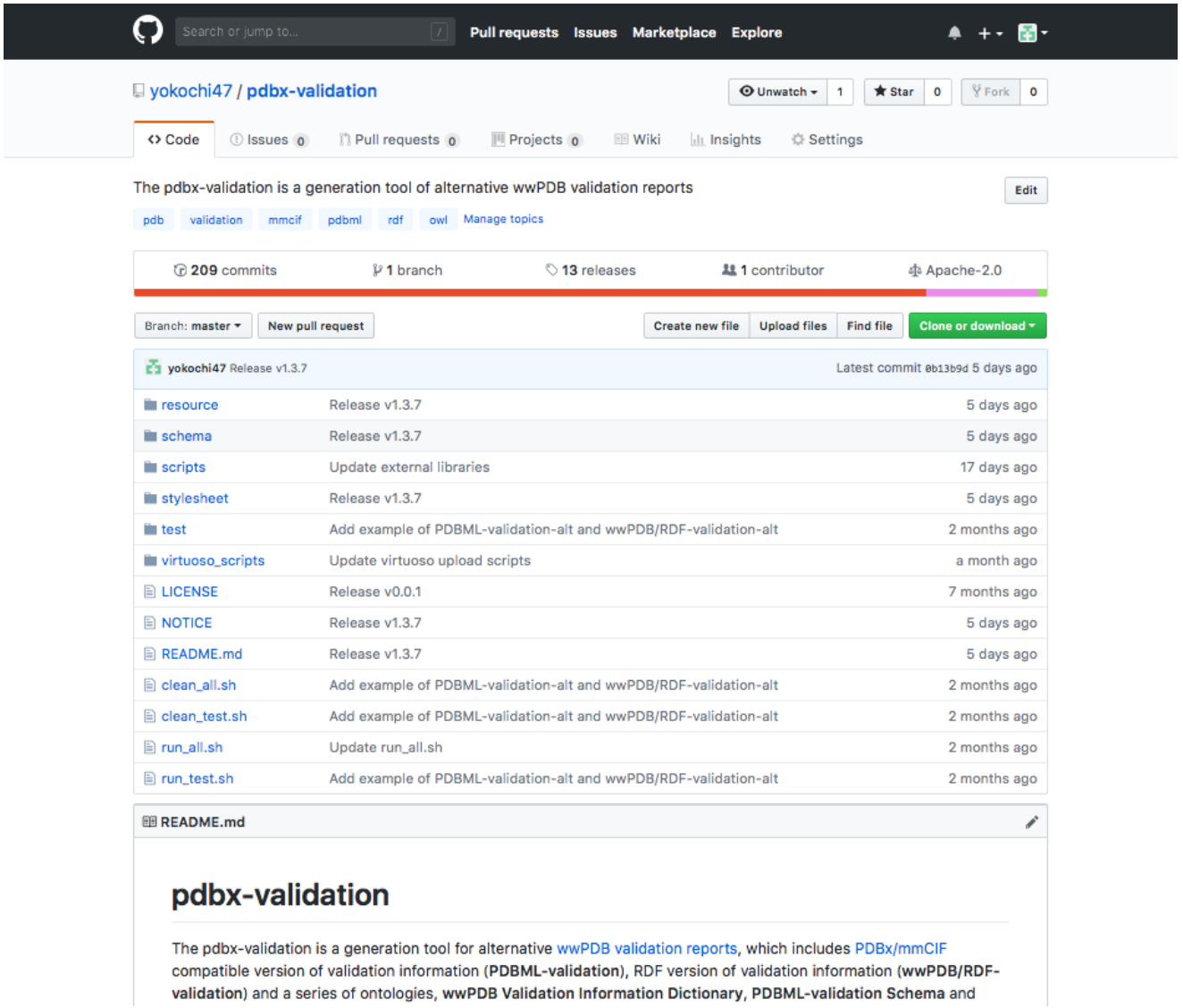
validation report of modeled entities repeat…


```
</wwPDB-validation-information>
```

Known issues on wwPDB validation reports (XML)

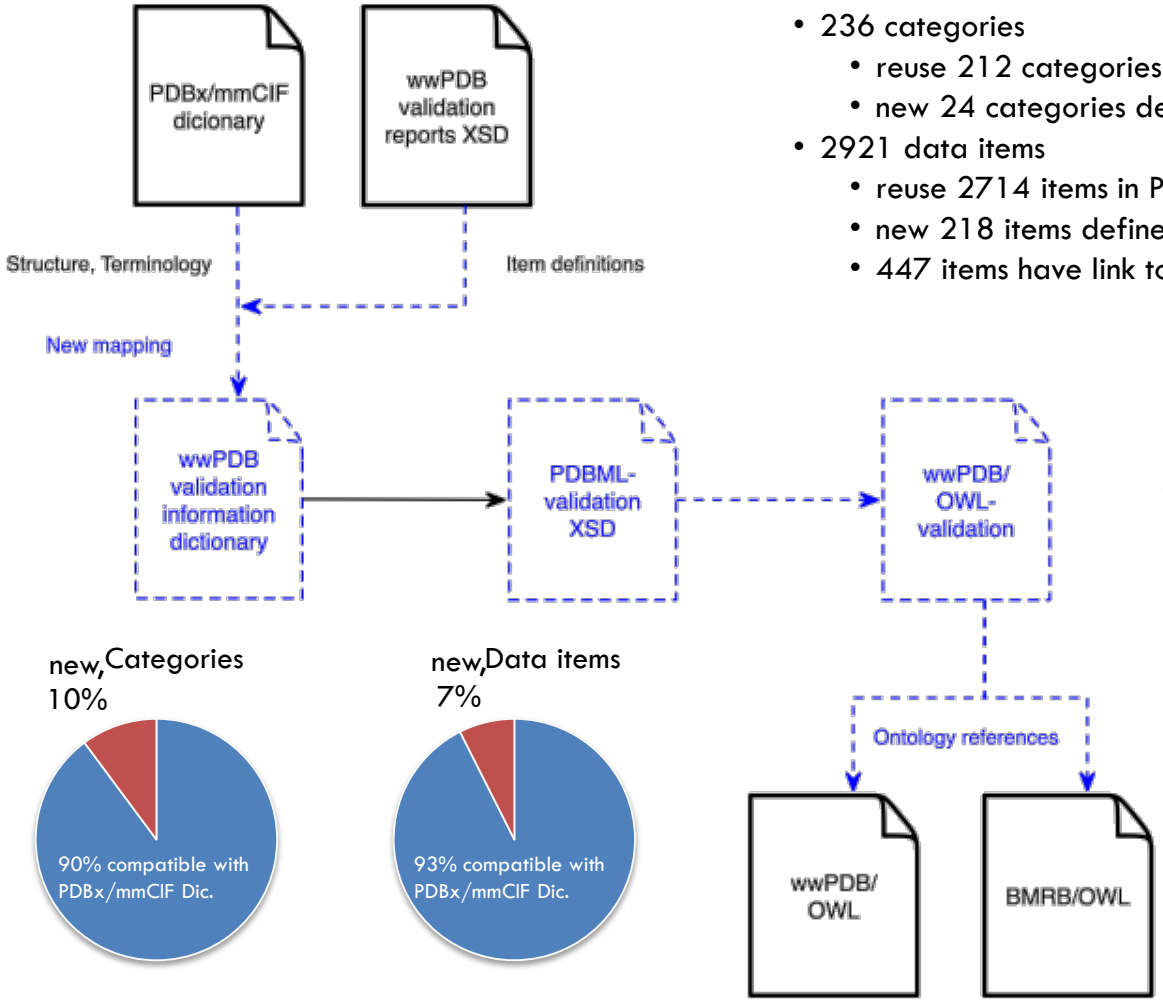
- Incompatible naming with the PDB's manner (PDBx/mmCIF).
  - said -> _atom_site.auth_asym_id
  - cid -> _pdbx_validate_close_contact.id
- Fat attributes
  - It leads to an inefficient search, where program always travels all instances, regardless of whether data exists or not.
- No categories
  - All metrics are tied to entry, entity, and monomer. It is simple, but …
  - It needs tweak for description for
    - angles between adjacent monomers
    - steric collision between entities, or different asymmetric units
    - NMR ensemble structure model
- Name collision
  - For example, 'value' data item indicates one of either chemical shift outlier, random coil index, referencing offset, unparsed chemical shift or unmapped chemical shift.

Reorganization of wwPDB validation reports in manner of the PDBx/mmCIF



https://github.com/yokochi47/pdbx-validation

Reorganization of wwPDB validation reports in manner of the PDBx/mmCIF



wwPDB Validation Information Dictionary, PDBML-validation Schema (XSD)

- 236 categories
  - reuse 212 categories in PDBx/mmCIF Dic.
  - new 24 categories defined
- 2921 data items
  - reuse 2714 items in PDBx/mmCIF Dic.
  - new 218 items defined
  - 447 items have link to validation reports' XSD

wwPDB/OWL-validation

- As for relation with wwPDB/OWL
  - 25 same classes
  - 427 equivalent classes
  - 3795 equivalent properties
- As for relation with BMRB/OWL
  - 134 equivalent properties

https://github.com/yokochi47/pdbx-validation

PDBML-validation and wwPDB/RDF-validation

PDBML-validation:

Note that PDBML-validation is an experimental archive and may be changed or replaced in the future.

```
% rsync -av --delete rsync://bmrbpub.pdbj.org/pdbml-valid .
```

wwPDB/RDF-validation:

Note that wwPDB/RDF-validation is an experimental archive and may be changed or replaced in the future.

```
% rsync -av --delete rsync://bmrbpub.pdbj.org/wwpdb-rdf-valid .
```

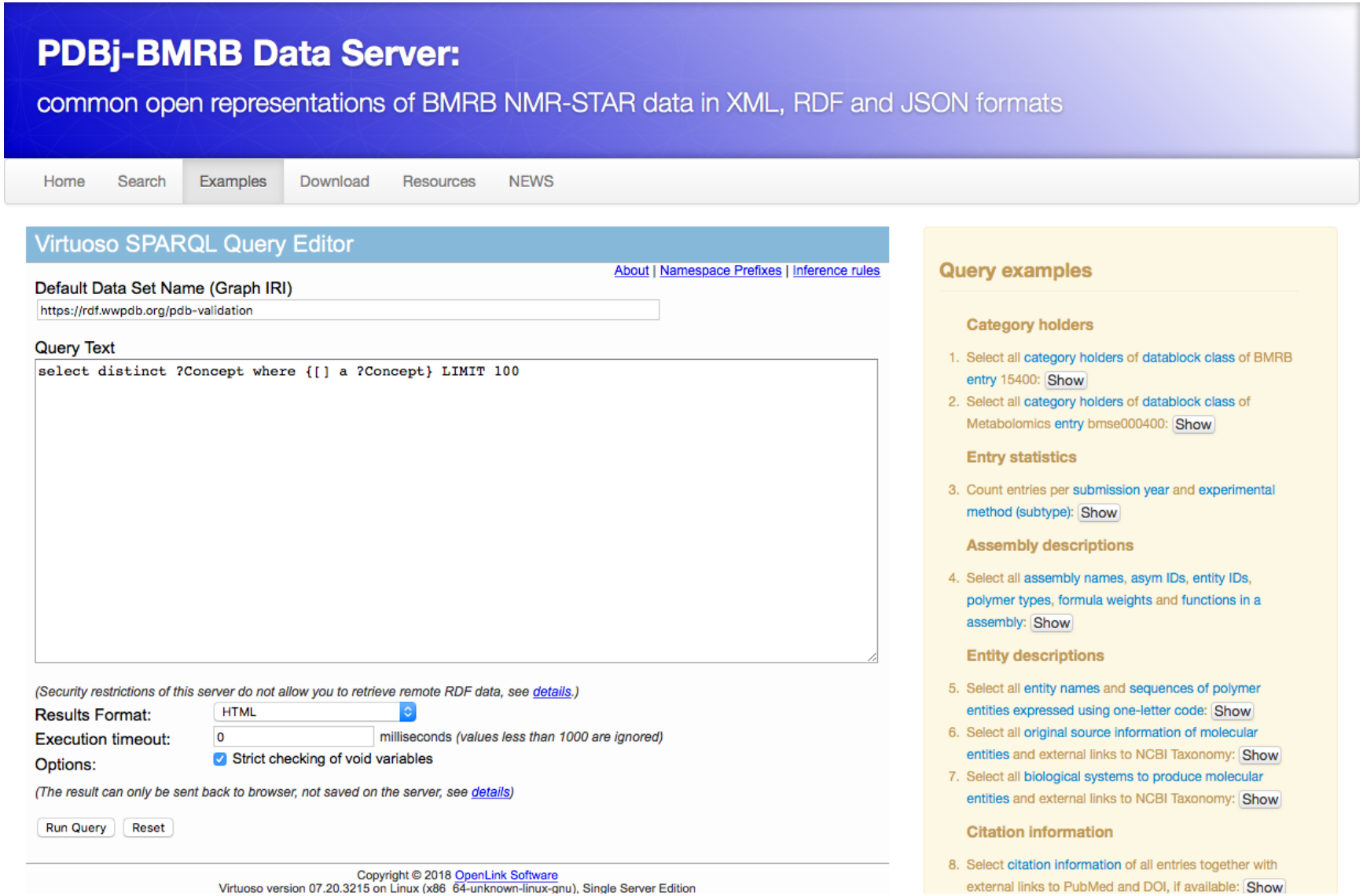
PostgreSQL dump image:

category	description	size (GB)
pdbx_dcc_map	output of MAPMAN used by DCC (RSR, RSCC, LLDF)	10
pdbx_poly_seq_scheme	residue nomenclature mapping for polymer entities	8
struct_mon_prot	structure properties of a protein	6
entity_poly_seq	sequence of monomers in a polymer	2
pdbx_validate_close_contact	close contact with regard to the distance expected	2

https://bmrbpub.pdbj.org

SPARQL endpoint contains wwPDB/RDF-validation graph

https://bmrbpub.pdbj.org



Example #1: Search wwPDB/RDF-validation with SPARQL

Search all enzyme-ligand complexes of which real space R-factor (RSR) of ligand is less than 10%. (showing only essential part of about 30 line-SPARQL query)

```
PREFIX PDBov: <https://rdf.wwpdb.org/schema/pdbx-validation-v1.owl#>
SELECT ?PDB_ID ?enzyme ?ligand ?comp_id MIN(?RSR AS ?minRSR)
FROM <http://rdf.wwpdb.org/pdb-validation>
WHERE {
 ?entity PDBov:link_to_enzyme ?link_to_enzyme ;
        PDBov:entity.pdbx_description ?enzyme ;
        PDBov:of_datablock ?datablock .

 BIND (SUBSTR(STR(?datablock),38,4) AS ?PDB_ID)

…

 FILTER (?ligand != "water" && !STRENDS(?ligand, "ION"))

…

 ?dcc_map PDBov:pdbx_dcc_map.auth_asym_id ?asym_id ;
         PDBov:pdbx_dcc_map.auth_comp_id ?comp_id ;
         PDBov:pdbx_dcc_map.RSR ?RSR .

 FILTER (xsd:float(?RSR) < 0.1)

} GROUP BY ?PDB_ID ?enzyme ?ligand ?comp_id
```

✍ selection of enzyme

✍ ligand selection: non-polymer, not water, not ion

✍ RSR < 0.1

Example #1: Search wwPDB/RDF-validation with SPARQL

Found 15k pairs of enzyme-ligand complexes of which real space R-factor (RSR) of ligand is less than 10%.

PDB ID, enzyme name, ligand name, ligand (3-letters code), minimum RSR value of the ligand

"4CK1","INTEGRASE","(4-CARBOXY-1,3-BENZODIOXOL-5-YL)METHYL-[[2-[(4-METHOXYPHENYL)METHYLCARBAMOYL]PHENYL]METHYL]AZANIUM","OM1","0.081"
"2IOD","Dihydroflavonol 4-reductase","NADP NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE","NAP","0.091"
"5BYR","Iron hydrogenase 1","FE2/S2 (INORGANIC) CLUSTER","FES","0.096"
"2PU0","Enolase","PHOSPHONOACETOHYDROXAMIC ACID","PAH","0.075"
"4LV2","Beta-lactamase","[1-(6-chloropyrimidin-4-yl)-1H-pyrazol-4-yl]boronic acid","N95","0.083"
"3OLE","Pancreatic alpha-amylase","ALPHA-D-GLUCOSE","GLC","0.084"
"4MOR","Pyranose 2-oxidase","DODECAETHYLENE GLYCOL","12P","0.093"
"4FKX","Nucleoside diphosphate kinase","CYTIDINE-5'-DIPHOSPHATE","CDP","0.073"
"4JPU","Cytochrome c peroxidase","PROTOPORPHYRIN IX CONTAINING FE","HEM","0.094"
"1GTV","THYMIDYLATE KINASE","THYMIDINE-5'-DIPHOSPHATE","TYD","0.090"
"5IA2","7-(5-hydroxy-2-methylphenyl)-8-(2-methoxyphenyl)-1-methyl-1H-imidazo[2,1-f]purine-2,4(3H,8H)-dione","7-(5-hydroxy-2-methylphenyl)-8-(2-methoxyphenyl)-1-methyl-1H-imidazo[2,1-f]purine-2,4(3H,8H)-dione","L66","0.077"
"1NFQ","Putative oxidoreductase Rv2002","1,4-DIHYDRONICOTINAMIDE ADENINE DINUCLEOTIDE","NAI","0.090”

Example #2: Search wwPDB/RDF-validation with SPARQL

Search all enzyme-ligand complexes of which percentage of outlier in real space R-factor, defined by Z-score (RSRZ) is larger than 2, of enzyme is less than 1%. (showing only essential part of about 20 line-SPARQL query)

```
PREFIX PDBov: <https://rdf.wwpdb.org/schema/pdbx-validation-v1.owl#>
SELECT ?PDB_ID ?Enzyme (GROUP_CONCAT(?Ligand; SEPARATOR=",") AS ?Ligands)
?RSRZ_outliers_percent
FROM <http://rdf.wwpdb.org/pdb-validation>
WHERE {
 ?map_overall PDBov:pdbx_dcc_map_overall.entry_id ?PDB_ID ;
          PDBov:pdbx_dcc_map_overall.RSRZ_outliers_percent ?RSRZ_outliers_percent .

 FILTER (xsd:float(?RSRZ_outliers_percent) < 0.01)
                                                    ✍ % of outliers in RSRZ < 1%
 BIND (IRI(CONCAT("https://rdf.wwpdb.org/pdb-validation/", ?PDB_ID, "/entityCategory")) AS
?entity_category)

 ?entity_category PDBov:has_entity ?entity .

 ?entity PDBov:link_to_enzyme ?link_to_enzume ;
       PDBov:entity.pdbx_description ?Enzyme .
                                                    ✍selection of enzyme
...

}                                                   ✍ligand selection (omission)
```

Example #2: Search wwPDB/RDF-validation with SPARQL

Found 5k pairs of enzyme-ligand complexes of which percentage of outlier in real space R-factor (RSRZ) of enzyme is less than 1%, 1k pairs for 0%.

PDB ID, enzyme name, ligand name, percentage of outliers in RSR value of the enzyme

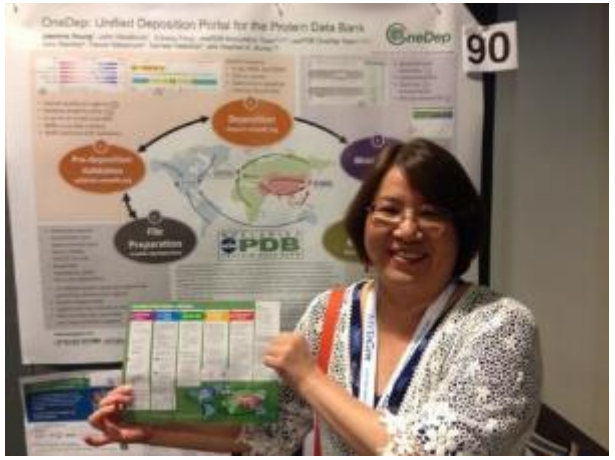
"1BUL","NMC-A BETA-LACTAMASE","2-(1-CARBOXY-2-HYDROXY-2-METHYL-PROPYL)-5,5-DIMETHYL-THIAZOLIDINE-4-CARBOXYLIC ACID,2-(N-MORPHOLINO)-ETHANESULFONIC ACID","0.00"
"5A1G","S-ADENOSYLMETHIONINE SYNTHASE ISOFORM TYPE-2","(DIPHOSPHONO)AMINOPHOSPHONIC ACID,[(3S)-3-amino-3-carboxypropyl]{[(2S,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl]methyl}ethylsulfonium,(4S)-2-METHYL-2,4-PENTANEDIOL,IMIDAZOLE","0.00"
"2DRS","Xylanase Y","GLYCEROL","0.00"
"2AS1","Cytochrome c peroxidase, mitochondrial","PROTOPORPHYRIN IX CONTAINING FE,THIOPHENE-3-CARBOXIMIDAMIDE","0.00"
"1H4W","TRYPSIN IVA","BENZAMIDINE","0.00"
"142L","T4 LYSOZYME","BETA-MERCAPTOETHANOL","0.00"
"4CIK","PLASMINOGEN","5-[(2R,4S)-2-(phenylmethyl)piperidin-4-yl]-1,2-oxazol-3-one","0.00"
"4L4O","Endo-1,4-beta-xylanase","TRIS-HYDROXYMETHYL-METHYL-AMMONIUM","0.00"
"4G5P","Epidermal growth factor receptor","N-{4-[(3-chloro-4-fluorophenyl)amino]-7-[(3S)-tetrahydrofuran-3-yloxy]quinazolin-6-yl}-4-(dimethylamino)butanamide","0.00"
"3GA6","Exodeoxyribonuclease","GLYCEROL","0.00"

Semantic extension of the wwPDB validation reports

	PDF	XML	PDBML-validation	wwPDB/RDF-validation
Human-readability	yes	-	-	-
Validation information	summary	full	full	full
Searchable	no	yes (XQuery)	yes (SQL, XQuery)	yes (SPARQL)
PDBx/mmCIF	-	-	~90% compatible	~90% compatible
URI	-	-	-	supported
Purpose	peer review	data exchange	quick search, data exchange	knowledge sharing

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

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 Core Archive Downloads

Year	Total	Total FTP Archive	Total Website	RCSB PDB FTP Archive	RCSB PDB Website	PDBe FTP Archive	PDBe Website	PDBj FTP Archive	PDBj Website
2017	679,421,200	454,723,083	224,698,117	369,868,918	143,839,718	40,168,072	65,420,002	44,686,093	15,438,397
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
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
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
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
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
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
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

More than 1.8 million/day!

N.B.: Some 2018 data lost due to GDPR.
Hope to be back on track for 2019→.

Geographic Origins of FTP downloads, 2012-2015

PDBで提供しているFile Formats

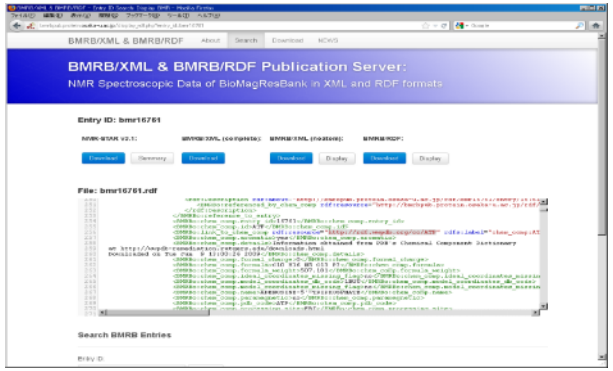
- mmCIF
  - The canonical format of the wwPDB.
  - Ver. 5 released this year!　596のvocabularies
- PDBML
  - “direct translation” of mmCIF into XML.
- (Legacy) PDB format
  - *NOT RECOMMENDED!*
- PDB/RDF
  - Translation of PDBML into RDF/XML (the standard format for the Semantic Web).

wwPDBにおけるRDF化の経緯

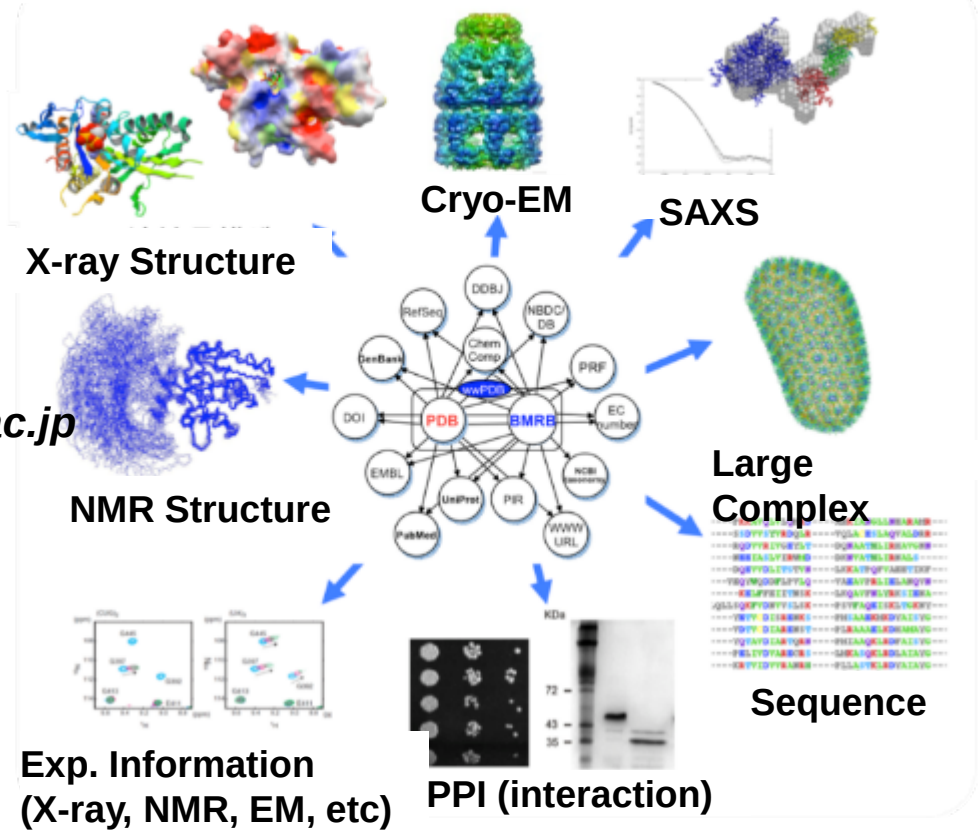
PDBj
Protein Data Bank Japan

wwPDB/RDF
http://rdf.wwpdb.org/

BMRB/RDF
http://bmrbpub.protein.osaka-u.ac.jp

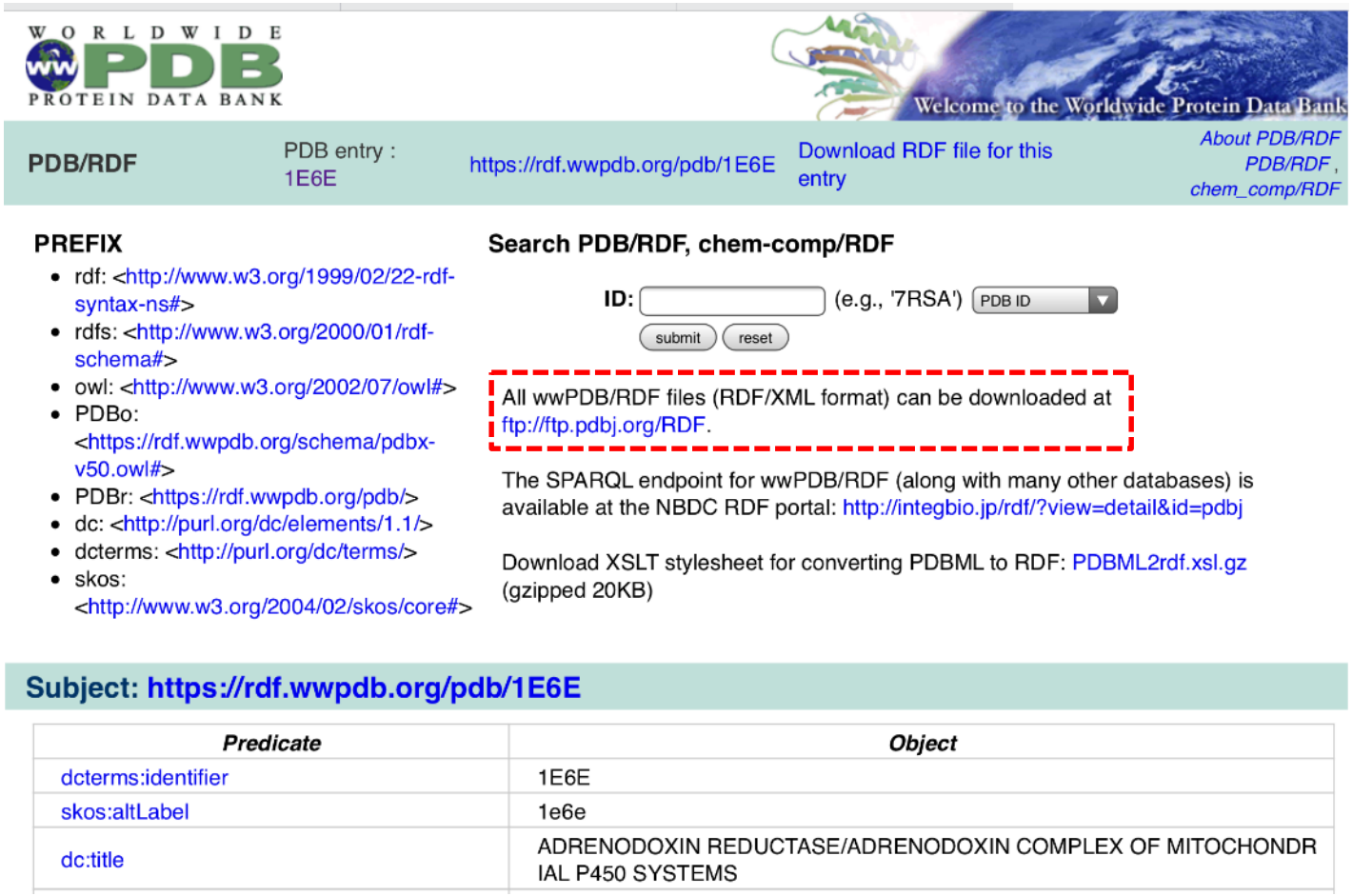


Kinjo et al. (2012) Nucl. Acids Res. 40, D453-D460.
Yokochi et al. (2016) J. Biomed. Semantics, 7:16.




```
<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>
```

wwPDBでPDBjがRDFファイルを公開



WORLDWIDE PDB PROTEIN DATA BANK

Welcome to the Worldwide Protein Data Bank

PDB/RDF PDB entry : 1E6E https://rdf.wwpdb.org/pdb/1E6E Download RDF file for this entry *About PDB/RDF* *PDB/RDF , chem_comp/RDF*

PREFIX

- rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
- rdfs: <http://www.w3.org/2000/01/rdf-schema#>
- owl: <http://www.w3.org/2002/07/owl#>
- PDBo: <https://rdf.wwpdb.org/schema/pdbx-v50.owl#>
- PDBr: <https://rdf.wwpdb.org/pdb/>
- dc: <http://purl.org/dc/elements/1.1/>
- dcterms: <http://purl.org/dc/terms/>
- skos: <http://www.w3.org/2004/02/skos/core#>

Search PDB/RDF, chem-comp/RDF

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

submit reset

All wwPDB/RDF files (RDF/XML format) can be downloaded at ftp://ftp.pdbj.org/RDF.

The SPARQL endpoint for wwPDB/RDF (along with many other databases) is available at the NBDC RDF portal: http://integbio.jp/rdf/?view=detail&id=pdbj

Download XSLT stylesheet for converting PDBML to RDF: PDBML2rdf.xsl.gz (gzipped 20KB)

Subject: https://rdf.wwpdb.org/pdb/1E6E

Predicate	*Object*
dcterms:identifier	1E6E
skos:altLabel	1e6e
dc:title	ADRENODOXIN REDUCTASE/ADRENODOXIN COMPLEX OF MITOCHONDRIAL P450 SYSTEMS

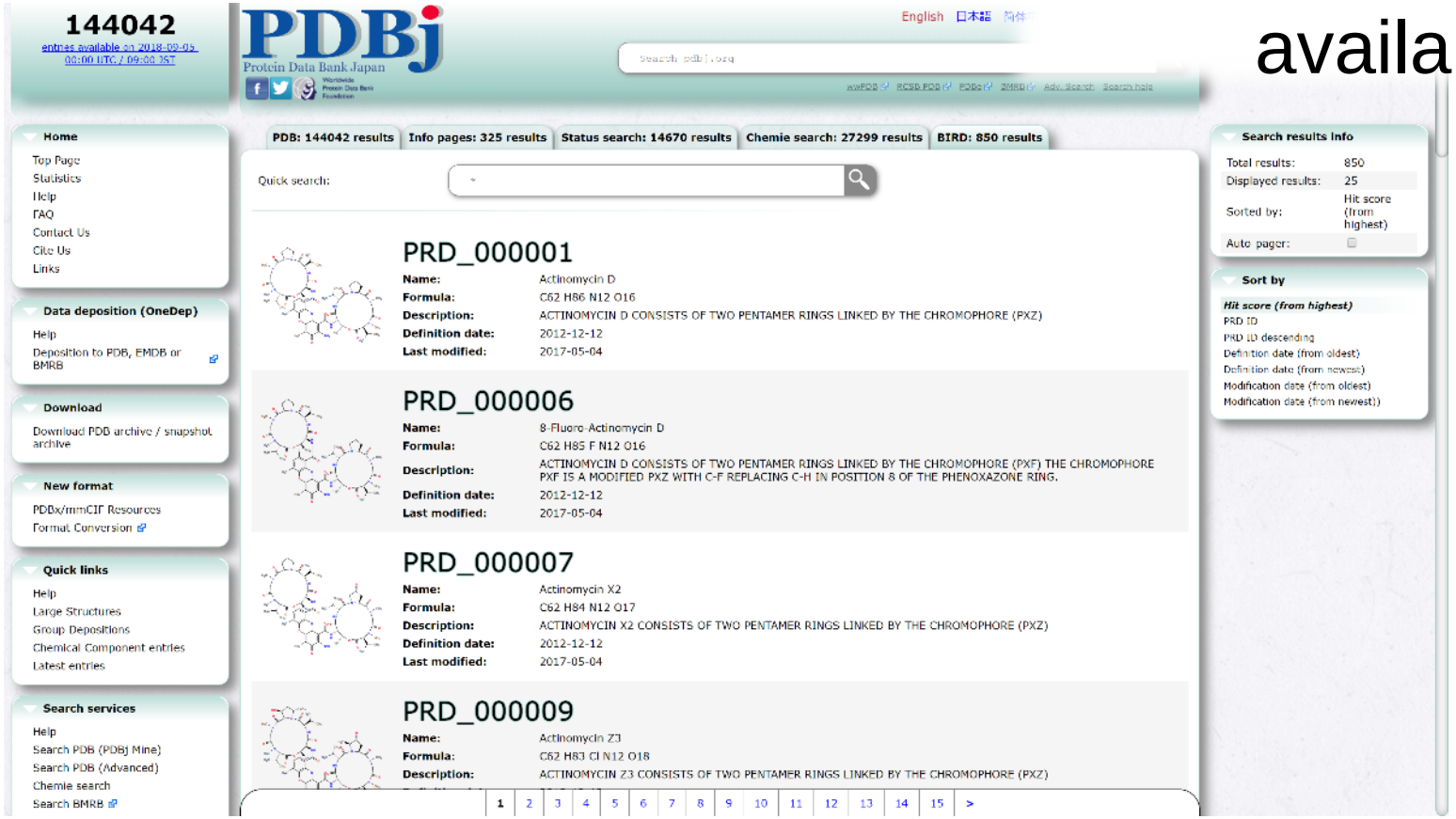
PDBj-Mineの高度化(1)

Sequence navigator now supports large structures

Infrastructure upgraded for speed: weekly releases are now available on Wednesday at 09:00

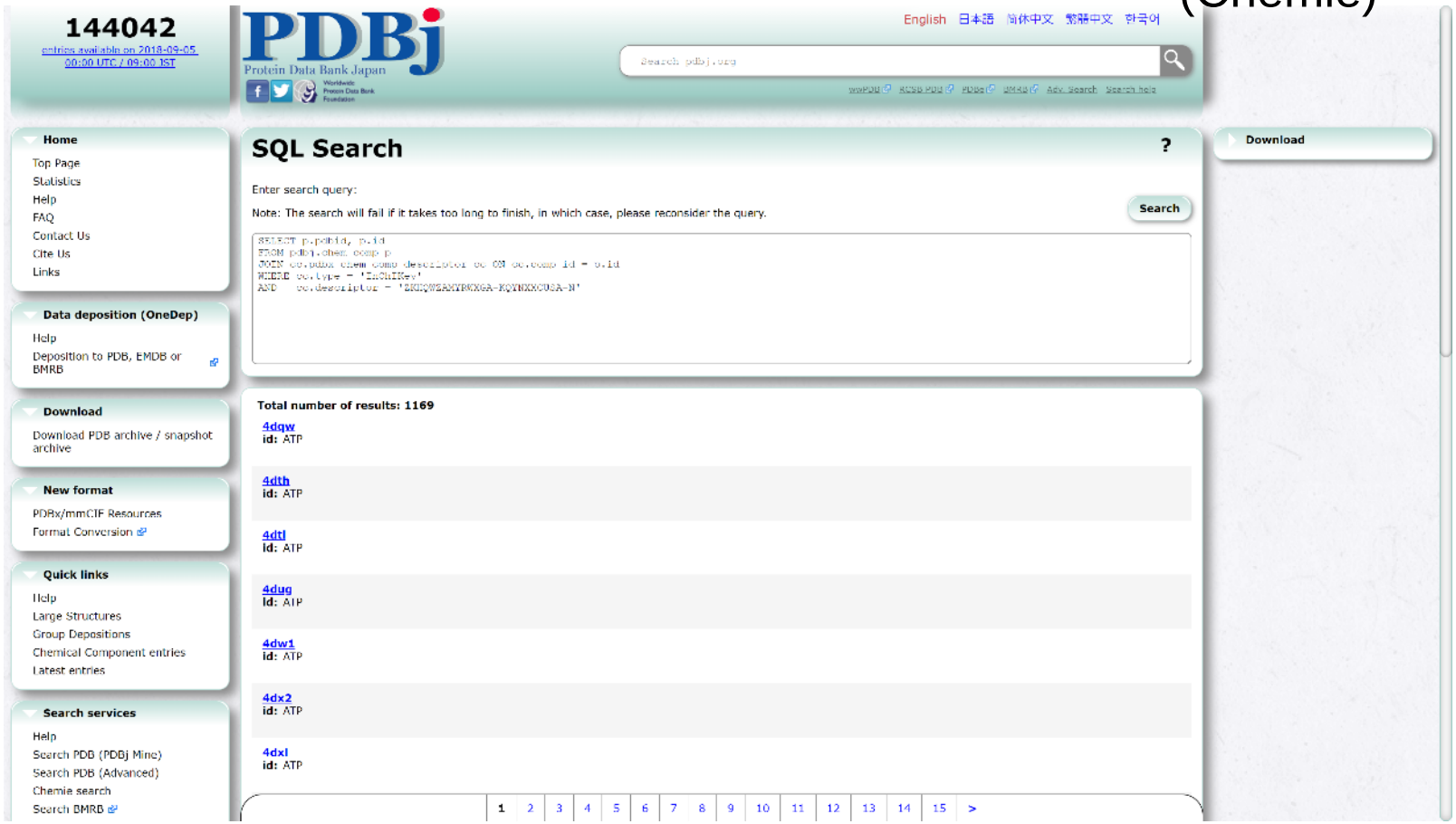
PDBj-Mineの高度化(2)

Mine BIRD/PRD Search is now available



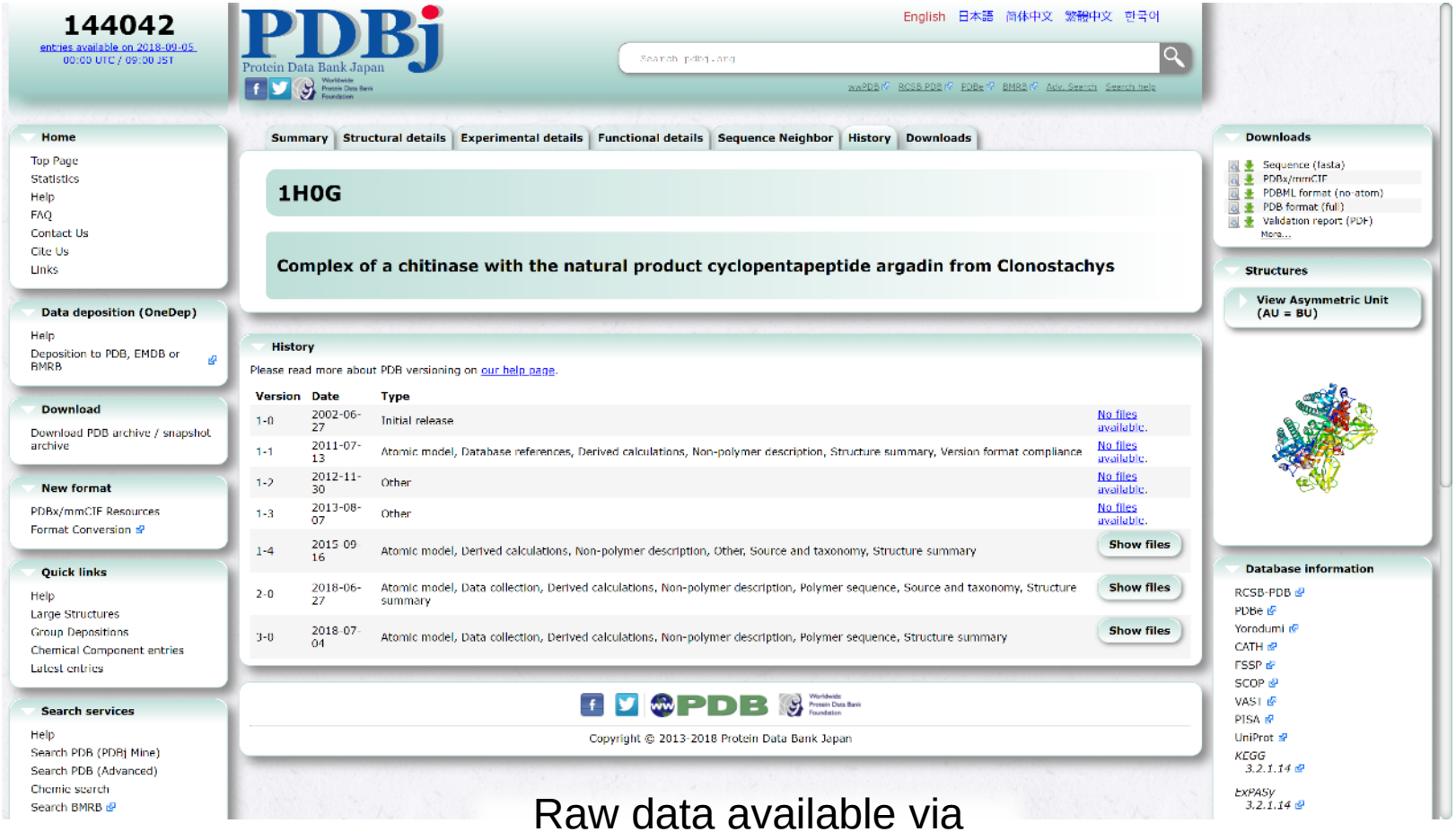
PDBj-Mineの高度化（3）

Added ability to perform SQL search against the chemical_component dictionary (Chemie)



PDBj-Mineの高度化(4)

Added new “History” tab to Mine PDB Entry pages



Raw data available via
http://ftp-versioned.pdbj.org/

PDBj
Protein Data Bank Japan

144042
entries available on 2018-09-05 00:00 UTC / 09:00 JST

PDBj Protein Data Bank Japan

English 日本語 简体中文 繁體中文 한국어

Search pdbj.org

wwPDB RCSB PDB PDBe BMRB Adv. Search Search help

- **Home**
  - Top Page
  - Statistics
  - Help
  - FAQ
  - Contact Us
  - Cite Us
  - Links
- **Data deposition (OneDep)**
  - Help
  - Deposition to PDB, EMDB or BMRB
- **Download**
  - Download PDB archive / snapshot archive
- **New format**
  - PDBx/mmCIF Resources
  - Format Conversion
- **Quick links**
  - Help
  - Large Structures
  - Group Depositions
  - Chemical Component entries
  - Latest entries
- **Search services**
  - Help
  - Search PDB (PDBj Mine)
  - Search PDB (Advanced)
  - Chemle search
  - Search BMRB

PDB Group Deposition list

Group ID	Release date	Title & description
G_1001001 (54)	2018-08-22	Cleavage of RNA by B. Halodurans Ribonuclease H1 *The structures represent different time points in the divalent and monovalent metal-dependent hydrolysis reaction of RNA by Ribonuclease H1, which is bound to an RNA/DNA hybrid in the crystal.*
G_1002034 (26)	2018-08-08	PanDDA analysis group deposition *Nuclease domain of human DCLRE1A screened against the DSPL Fragment Library by X-ray Crystallography at the XChem facility of Diamond Light Source beamline I04-1*
G_1002036 (281)	2018-08-08	PanDDA analysis group deposition *Nuclease domain of human DCLRE1A screened against the DSPL Fragment Library by X-ray Crystallography at the XChem facility of Diamond Light Source beamline I04-1*
G_1002048 (7)	2018-07-18	PanDDA analysis group deposition of models with modelled events (e.g. bound ligands) *human HAO1 screened against the DSPL Fragment Library by X-ray Crystallography at the XChem facility of Diamond Light Source beamline I04-1*
G_1002027 (37)	2018-01-10	A focused fragment library targeting the antibiotic resistance enzyme - oxacillinase-48: synthesis, structural evaluation and inhibitor design,
G_1002040 (26)	2017-12-20	Ligand binding to Cathepsin S
G_1002014 (21)	2017-11-22	Crystal Structure of COMT complex
G_1002041 (4)	2017-11-08	FXIa *Structures deposited in support of D.J.P. Pinto et al., J.Med.C*
G_1002017 (6)	2017-11-01	11betaHSD1 *Structures deposited in support of X.-Y. Ye et al., J.Med.Che*
G_1002010 (10)	2017-08-30	Crystal Structure of COMT complex
G_1002009 (8)	2017-08-30	Crystal Structure of COMT complex
G_1002032 (5)	2017-07-12	FXIa *Structures deposited in support of J.R. Corte et al., Bioorg.Med.Chem.Lett., about to be*
G_1002033 (37)	2017-07-05	Ligand binding to FARNESOID-X-RECEPTOR
G_1002016 (5)	2017-06-21	Crystal Structure of a Factor VIIa complex
G_1002015 (25)	2017-06-21	Crystal Structure of a Factor VIIa complex,
G_1002011 (8)	2017-05-24	Crystal structures of Tyrosine-protein kinase BTK in complex with inhibitors
G_1002026 (6)	2017-05-10	NS5B 1b *Structures deposited in support of K.-S. Yeung et al., J.Med.Chem., about to be submitted*
G_1002023 (257)	2017-03-29	PanDDA analysis group deposition of models of ground state datasets *human BRD1 screened against the 3D-Fragment Consortium Library by X-ray Crystallography at the XChem facility of Diamond Light Source beamline I04-1. Check out the PanDDA event maps at https://zenodo.org/record/290217/files/0_index.html*
G_1002025	2017-03-22	PanDDA analysis group deposition of models of ground state datasets *human SP100 screened against the Maybridge Library by X-ray Crystallography at the XChem facility of Diamond Light Source beamline I04-1. Check out the PanDDA event maps at*

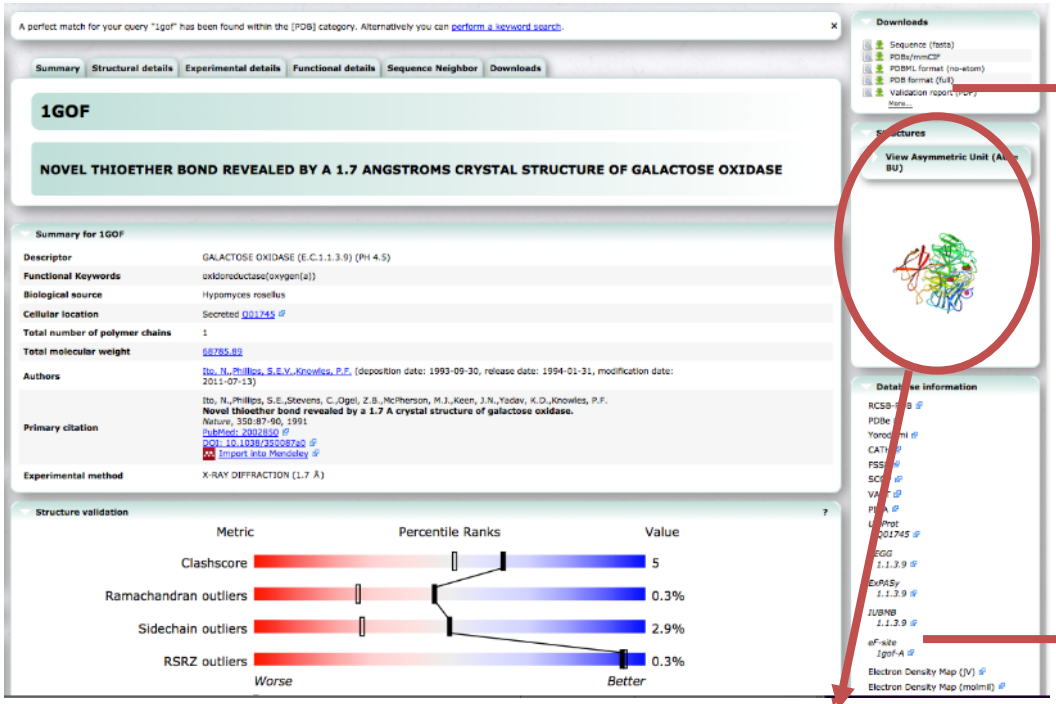
Renewed group deposition list

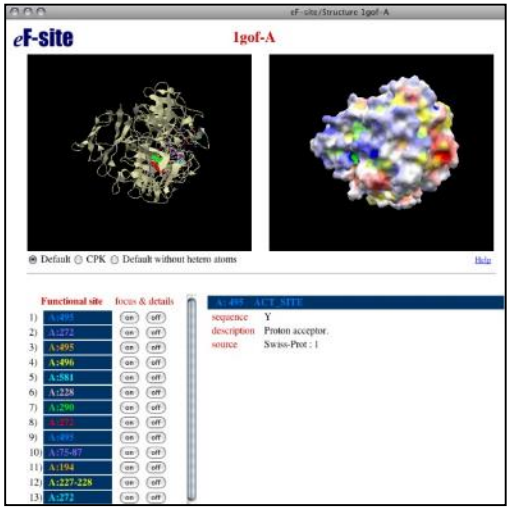
Added release date column + sorted by release date (desc.)

PDBjで開発したData viewer (Molmil)

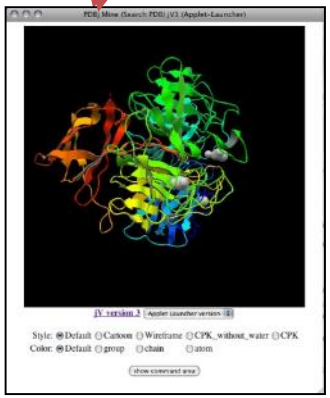
PDBj Protein Data Bank Japan

Amino acid sequence (FASTA)





Data viewer at PDBj



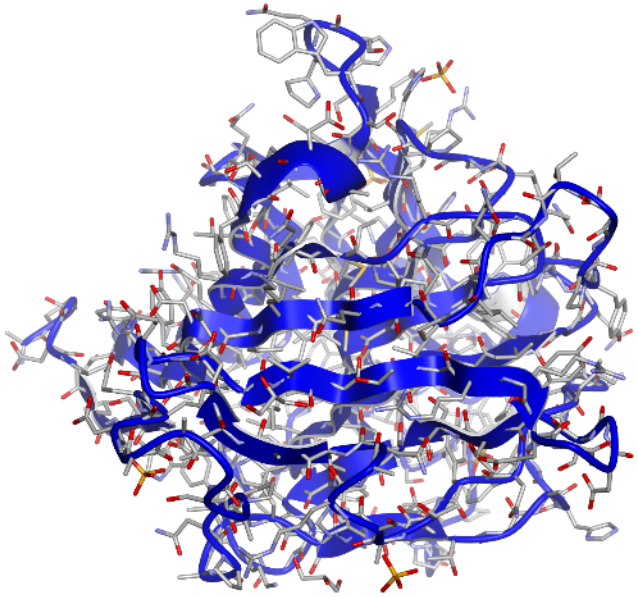
Graphic viewer: *jV* and *Molmil*
http://pdbj.org/molmil/

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

Kinjo et al. NAR 40, D453 (2012)

Data viewerの高度化

Molmil β



Various improvements with respect to the CLI, rendering and the interface

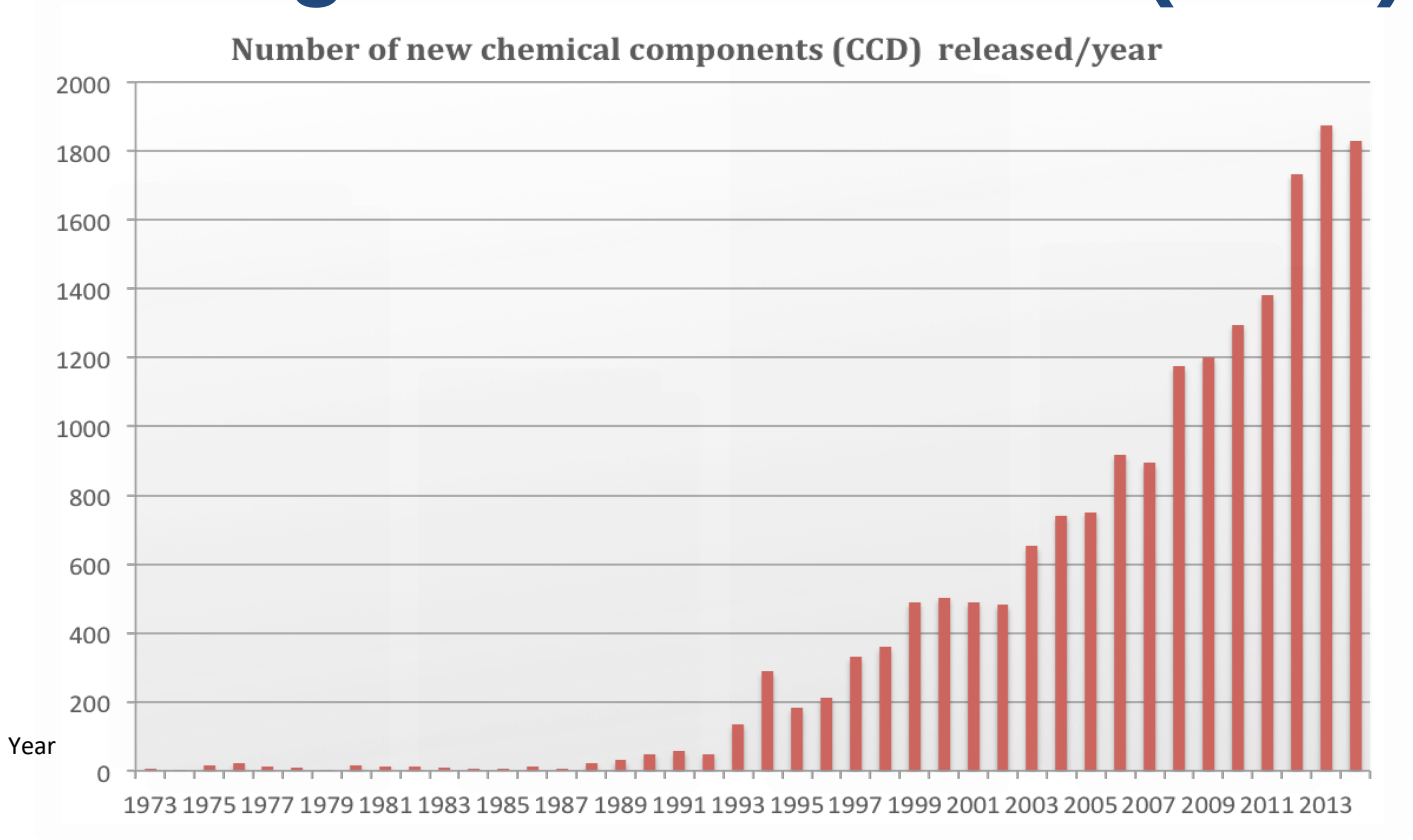
https://pdbj.org/molmil_beta/

```
Pymol-like command interface bound.
fetch 3atg
show sticks, all
cartoon_color blue, model #1
turn y, -40
>
```

Pymol commands

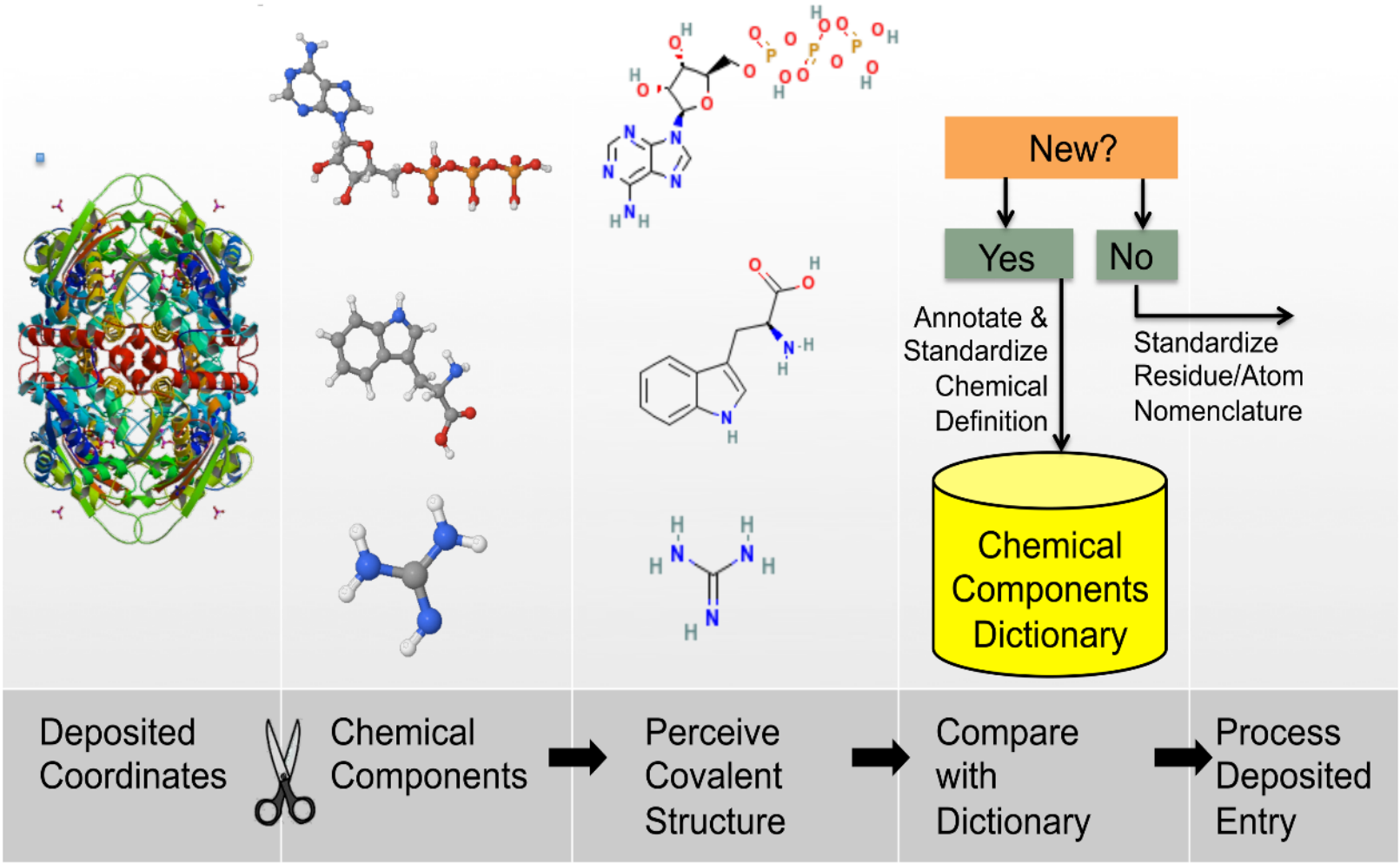
- select (select sc12, resi 12 and sidechain)
- color (color cyan, model #1 and symbol C)
- cartoon_color (cartoon_color cyan, model #1)
- set_color (set_color mycolor 12 12 12)
- show (show sticks, sidechain)
- hide (hide cartoon, model #1)
- turn (turn x, 90)
- move (move x, 90)
- fetch (fetch 1crn)
- load (load https://pdbj.org/rest/displayPromodeEfile?format=anm&id=1ubq_1, format=pdb)
- mplay
- mstop
- origin (origin chain A)
- set (stick_radius f, depth_cue 1/0, orthoscopic on/off, cartoon_smooth_loops 0/2)
- bg_color (bg_color cyan)
- label (label resi 12 and sidechain, Res12)
- save (save filename.pdb, model #1 and name CA, 0, pdb)
- viewport (viewport 500, 500)
- findseq (findseq ACDEF, model #1, my_seq)

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

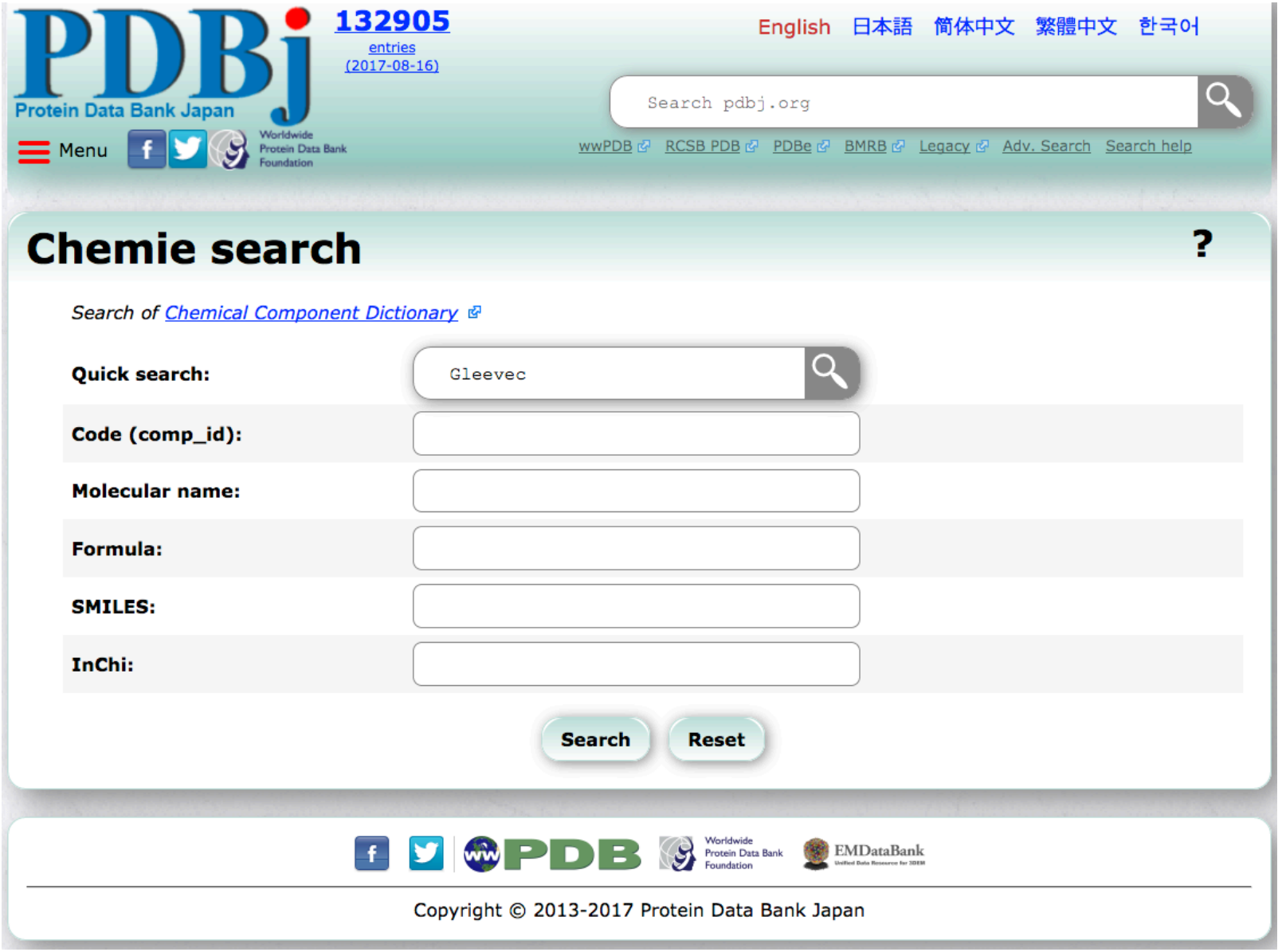


化合物からもPDB情報を検索出来る仕組み作り

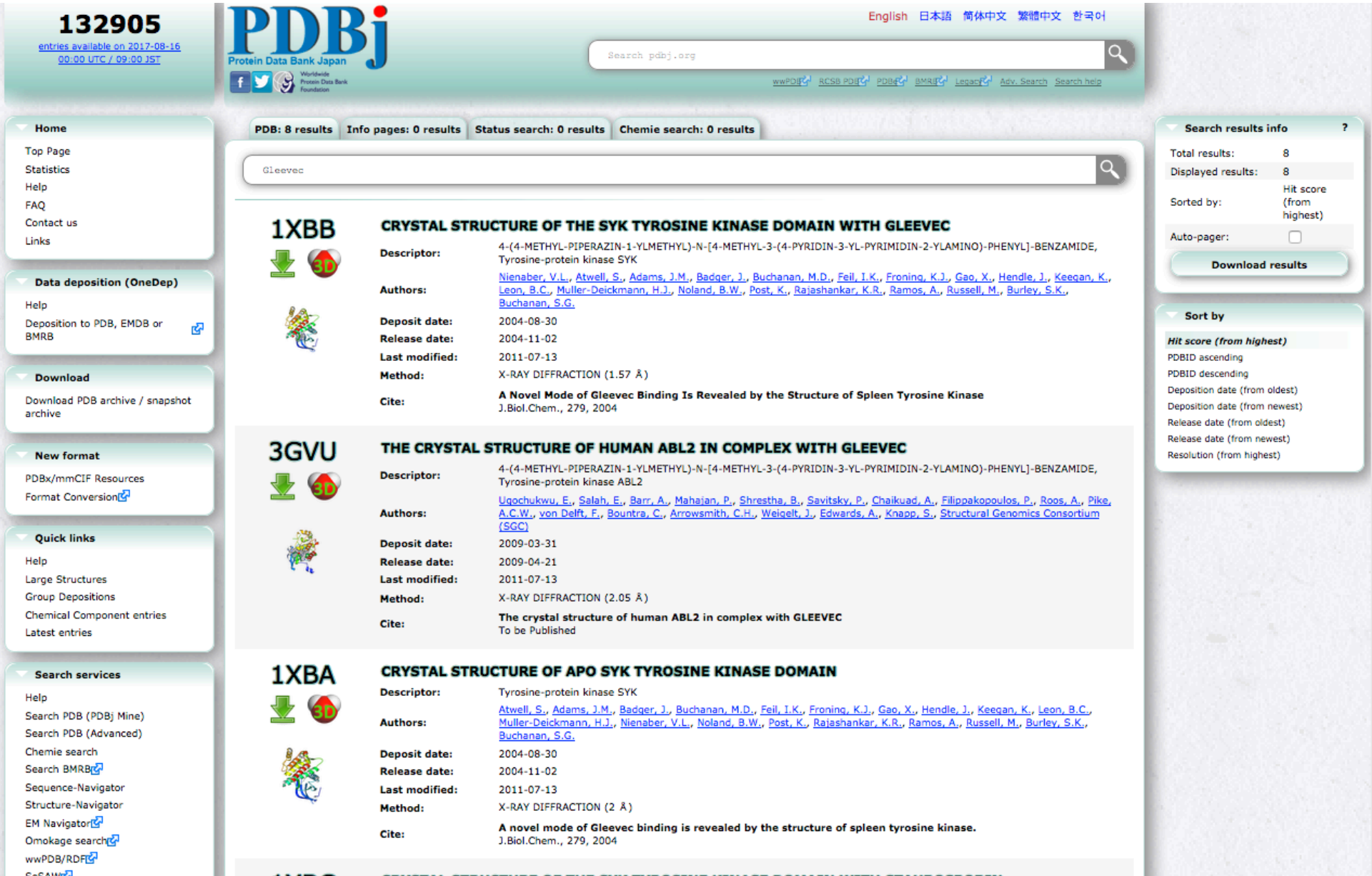
PDBに含まれる化学情報（全データの3/4）



Example1: Gleevec using PDBj-Mine



Explore 1T46: Gleevec using PDBj-Mine (cont.)



132905
entries available on 2017-08-16 00:00 UTC / 09:00 JST

PDBj Protein Data Bank Japan

English 日本語 简体中文 繁體中文 한국어

Search pdbj.org

wwPDB RCSB PDB PDBe BMRB Legacy Adv. Search Search help

Home
Top Page
Statistics
Help
FAQ
Contact us
Links

Data deposition (OneDep)
Help
Deposition to PDB, EMDB or BMRB

Download
Download PDB archive / snapshot archive

New format
PDBx/mmCIF Resources
Format Conversion

Quick links
Help
Large Structures
Group Depositions
Chemical Component entries
Latest entries

Search services
Help
Search PDB (PDBj Mine)
Search PDB (Advanced)
Chemie search
Search BMRB
Sequence-Navigator
Structure-Navigator
EM Navigator
Omokage search
wwPDB/RDF

PDB: 8 results | Info pages: 0 results | Status search: 0 results | Chemie search: 0 results

Gleevec

1XBB — CRYSTAL STRUCTURE OF THE SYK TYROSINE KINASE DOMAIN WITH GLEEVEC

Descriptor: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE, Tyrosine-protein kinase SYK
Authors: Nienaber, V.L., Atwell, S., Adams, J.M., Badger, J., Buchanan, M.D., Feil, I.K., Froning, K.J., Gao, X., Hendle, J., Keegan, K., Leon, B.C., Muller-Deickmann, H.J., Noland, B.W., Post, K., Rajashankar, K.R., Ramos, A., Russell, M., Burley, S.K., Buchanan, S.G.
Deposit date: 2004-08-30
Release date: 2004-11-02
Last modified: 2011-07-13
Method: X-RAY DIFFRACTION (1.57 Å)
Cite: **A Novel Mode of Gleevec Binding Is Revealed by the Structure of Spleen Tyrosine Kinase** J.Biol.Chem., 279, 2004

3GVU — THE CRYSTAL STRUCTURE OF HUMAN ABL2 IN COMPLEX WITH GLEEVEC

Descriptor: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE, Tyrosine-protein kinase ABL2
Authors: Ugochukwu, E., Salah, E., Barr, A., Mahajan, P., Shrestha, B., Savitsky, P., Chaikuad, A., Filippakopoulos, P., Roos, A., Pike, A.C.W., von Delft, F., Bountra, C., Arrowsmith, C.H., Weigelt, J., Edwards, A., Knapp, S., Structural Genomics Consortium (SGC)
Deposit date: 2009-03-31
Release date: 2009-04-21
Last modified: 2011-07-13
Method: X-RAY DIFFRACTION (2.05 Å)
Cite: **The crystal structure of human ABL2 in complex with GLEEVEC** To be Published

1XBA — CRYSTAL STRUCTURE OF APO SYK TYROSINE KINASE DOMAIN

Descriptor: Tyrosine-protein kinase SYK
Authors: Atwell, S., Adams, J.M., Badger, J., Buchanan, M.D., Feil, I.K., Froning, K.J., Gao, X., Hendle, J., Keegan, K., Leon, B.C., Muller-Deickmann, H.J., Nienaber, V.L., Noland, B.W., Post, K., Rajashankar, K.R., Ramos, A., Russell, M., Burley, S.K., Buchanan, S.G.
Deposit date: 2004-08-30
Release date: 2004-11-02
Last modified: 2011-07-13
Method: X-RAY DIFFRACTION (2 Å)
Cite: **A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase.** J.Biol.Chem., 279, 2004

Search results info

Total results:	8
Displayed results:	8
Sorted by:	Hit score (from highest)
Auto-pager:	

Download results

Sort by
Hit score (from highest)
PDBID ascending
PDBID descending
Deposition date (from oldest)
Deposition date (from newest)
Release date (from oldest)
Release date (from newest)
Resolution (from highest)

Home
Top Page
Statistics
Help
FAQ
Contact us
Links

Data deposition (OneDep)
Help
Deposition to PDB, EMDB or BMRB

Download
Download PDB archive / snapshot archive

New format
PDBx/mmCIF Resources
Format Conversion

Quick links
Help
Large Structures
Group Depositions
Chemical Component entries
Latest entries

Search services
Help
Search PDB (PDBj Mine)
Search PDB (Advanced)
Chemie search
Search BMRB
Sequence-Navigator
Structure-Navigator
EM Navigator
Omokage search
wwPDB/RDF
SeSAW
Ligand Binding Sites (GIRAF)
Status search

Molecular Viewers
Help

Summary | Structural details | Experimental details | Functional details | Sequence Neighbor | Downloads

1T46

STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE

Summary for 1T46

Related	1PKG 1T45
Descriptor	Homo sapiens v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog, PHOSPHATE ION, 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-[4-METHYL-3-(4-PYRIDIN-3-YL-PYRIMIDIN-2-YLAMINO)-PHENYL]-BENZAMIDE, ... (4 entities in total)
Functional Keywords	kinase, inhibitor, sti-571, gleevec, transferase activator
Biological source	Homo sapiens (human)
Total number of polymer chains	1
Total molecular weight	36072.48
Authors	Mol, C.D.,Dougan, D.R.,Schneider, T.R.,Skene, R.J.,Kraus, M.L.,Scheibe, D.N.,Snell, G.P.,Zou, H.,Sang, B.C.,Wilson, K.P. (deposition date: 2004-04-28, release date: 2004-06-15, Last modification date: 2011-07-13)
Primary citation	Mol, C.D.,Dougan, D.R.,Schneider, T.R.,Skene, R.J.,Kraus, M.L.,Scheibe, D.N.,Snell, G.P.,Zou, H.,Sang, B.C.,Wilson, K.P. **Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase.** *J.Biol.Chem.*, 279:31655-31663, 2004 PubMed: 15123710 (PDB entries with the same primary citation) DOI: 10.1074/jbc.M403319200 Import into Mendeley
Experimental method	X-RAY DIFFRACTION (1.6 Å)

Structure validation ?

More Asymmetric unit images

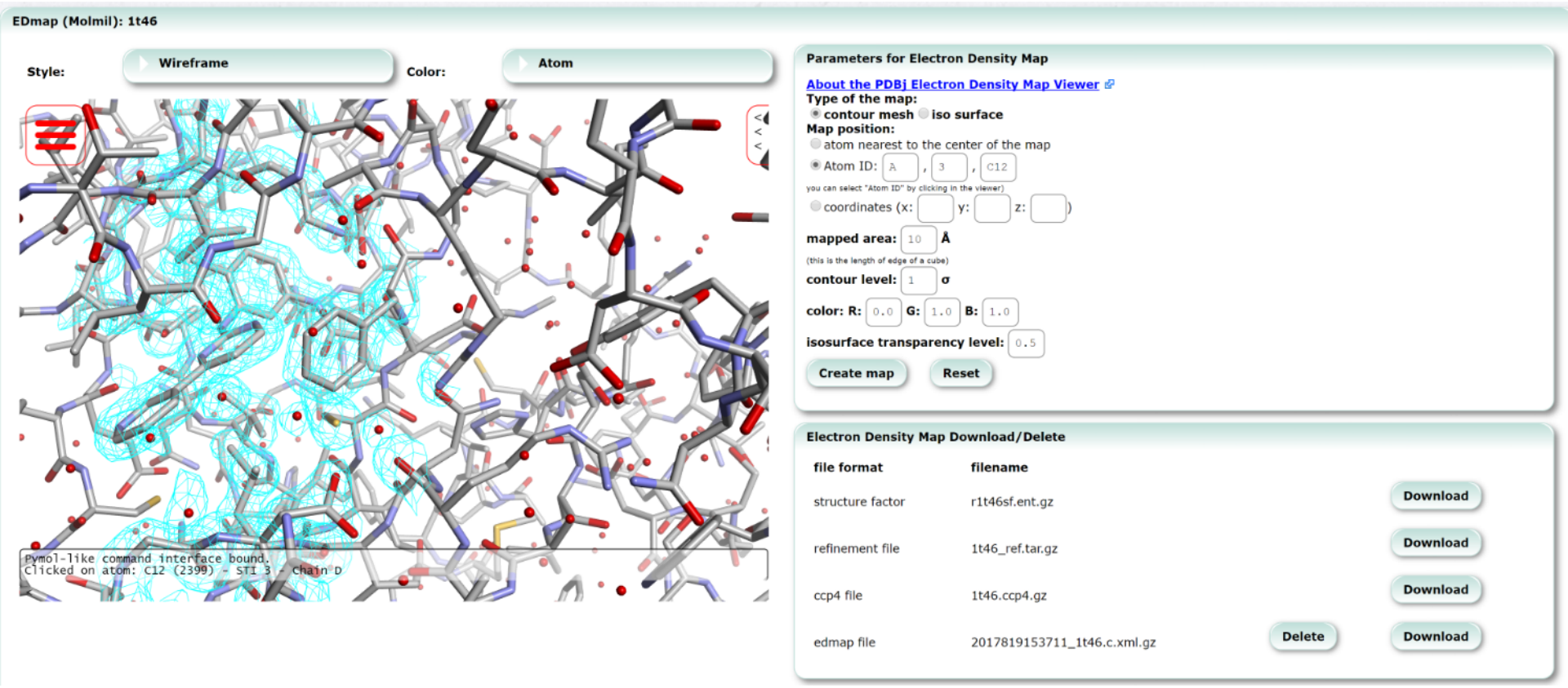
Downloads
Sequence (fasta)
PDBx/mmCIF
PDBML format (no-atom)
PDB format (full)
Validation report (PDF)
More...

Structures
View Asymmetric Unit (AU = BU)

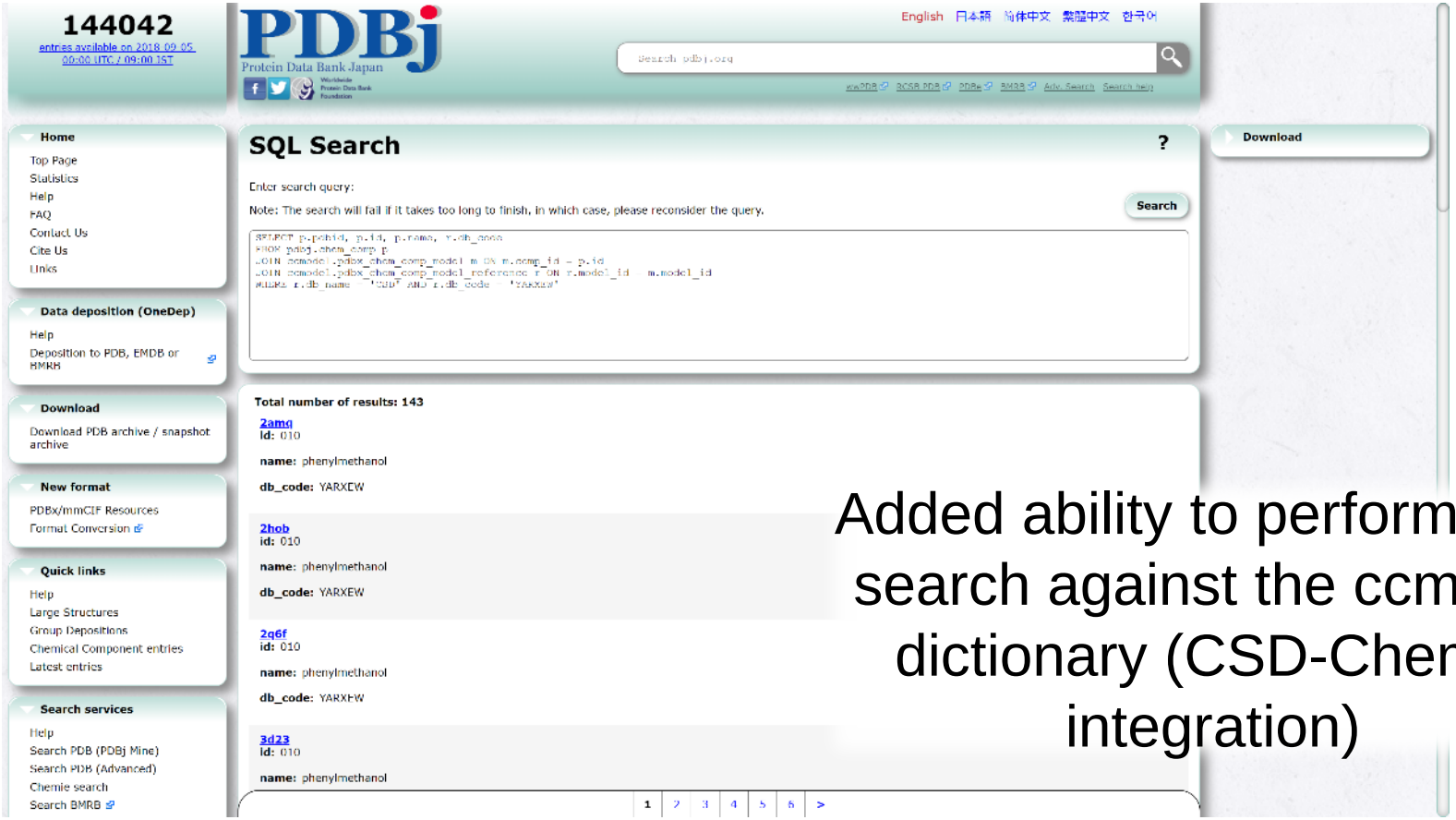
Database information
RCSB-PDB
PDBe
Yorodumi
CATH
FSSP
SCOP
VAST
PISA
UniProt
KEGG 2.7.1.112
ExPASy 2.7.1.112
IUBMB 2.7.1.112
eF-site 1t46.pdb
Electron Density Map (molmil)
wwPDB/RDF
Promode Elastic
EzCatDB (M00323)

Eelectron Density map (Molmil)

Explore 1T46: Gleevec using PDBj-Mine (cont.)



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



Added ability to perform SQL search against the ccmodel dictionary (CSD-Chemie integration)

PDBjのData-outシステム構成について

現在、システム移行の過渡期にあり、以下の通り複数のシステム系が存在する。

1. IPR側システム（2017年3月に大半を2. に移設）
   元から蛋白研に設置していたシステム。（一部のサービスが移植出来ていない）
2. NBDC側システム
   最初は札幌に設置して現在は蛋白研に戻しているシステム
3. 現行メインシステム
   主に蛋白研汎用計算機システム内の機器で構成されたPDBj用システム

1. IPR側システムの機器構成

役割	台数	備考
ウェブフロントエンド pdbjkw1, 2, 3	3	メインのウェブサイト、メールサーバ、DNSサーバなどを提供。3番はJava系ウェブサイト専用。2,3は運用終了。
公開ファイルサーバ pdbjkf1, 2	2	PDB Archiveのftp/rsyncサービスを提供。2台とも運用終了。
内部用ファイルサーバ pdbjkf3～5	2→3→2	当初サーバ2台(3,4)で冗長化してディスクアレイ装置を提供する仕組みを取っていたが、単独のNAS(5)に置き換え、3,4はバックアップ作業に使用。3は故障し4のみテープバックアップに使用。
計算管理サーバ pdbjkh1	1	共通認証(NIS)、ジョブスケジューラ(PBS)
計算実行サーバ pdbjk01～12	12→3	計算管理サーバから割り当てを受けジョブを実行。故障などにより現在は3台まで減少
RDBサーバ　pdbjkr1, 2	2	PDB検索のバックエンドデータベースサーバ。運用終了。
共用ファイルサーバ　filesv	1	eF-site、ProMode、フォーマット変換などの計算を実行

2. NBDC側システム

- 2012年頃、ミラーサイト構築用として新たに導入。
- 2012年7月～2016年2月の間、NBDC札幌データセンターに設置し、主要サービスのみ提供していた。
- 2016年2月、蛋白研内に戻して全サービスを独立実行できるよう再構築。
- 2017年3月21日メイン化。

2. NBDC側システムの機器構成

役割	台数	備考
ウェブフロントエンド nbdcpdbjw1, 3	1	メインのウェブサイト、メールサーバ、DNSサーバなどを提供。3番はJava系ウェブサイト専用。
公開ファイルサーバ nbdcpdbjf1	1	PDB Archiveのftp/rsyncサービスを提供。運用終了。
内部用ファイルサーバ nbdcpdbjf5→nbdcpdbjf6	1	蛋白研に戻した時新たに5番を導入。よりスペックの高い6番に移行し、5番はバックアップ用途に転用。
計算管理サーバ pdbjkh1	1	共通認証（NIS）、ジョブスケジューラ（PBS）
計算実行サーバ pdbjk01～12	12→3	計算管理サーバから割り当てを受けジョブを実行。故障などにより現在は3台まで減少
RDBサーバ　pdbjkr1, 2	2	PDB検索のバックエンドデータベースサーバ。
共用ファイルサーバ　filesv	1	eF-site、ProMode、フォーマット変換などの計算を実行

※細字はIPR側の仕組みを流用

3.現行システム

- 2016年3月、蛋白研汎用計算機システムを更新するタイミングで、合わせて用意して頂いた機器をメインとするシステム。
- 負荷の高い計算は蛋白研汎用計算機システム内の機器（uv）を利用（eF-site、Promode）。
- その他数台の追加購入機器も含む（filesv2、filesv3など）
- 2017年7月19日メイン化。

3.現行システムの機器構成

役割	台数	備考
ウェブフロントエンド pdbjiw1, pdbjiw2	2	メインのウェブサイト、メールサーバ、DNSサーバなどを提供。Java系ウェブサイト専用サーバは設けずこれらのサーバに統合。
公開ファイルサーバ pdbjif1, pdbjif2	2	PDB Archiveのftp/rsyncサービスを提供 50TBの領域をNFSでウェブフロントエンドなどにも提供。
共用ファイルサーバ filesv2、filesv3	2	filesv2はfilesvの後継機、filesv3は計算実行サーバの後継機
内部用ファイルサーバ nbdcpdbjf6	1	NBDC側システムのファイルサーバ（NAS）。前項ファイルサーバのミラー（バックアップ）などの用途で使用。
共用ファイルサーバ　filesv	1	eF-site、ProMode、フォーマット変換などの計算実行
蛋白研計算サーバ　uv		PBSを利用。
計算管理サーバ pdbjkh1	1	共通認証（NIS）、ジョブスケジューラ（PBS）
計算実行サーバ pdbjk01～12	12→3	計算管理サーバから割り当てを受けジョブを実行。故障などにより現在は3台まで減少

※細字は既存の仕組みを流用

4.現行システムの課題

- 現在、すべての機器が大阪大学蛋白質研究所内ネットワーク内にあるため、大阪大学内ネットワークの停止などがあるとサービスが提供できなくなる。
  →Google Cloud Platform上の仮想マシンを使ってメンテナンス中であることだけでも表示するしくみは暫定的に構築した。