

Ligand representation and validation in the Protein Data Bank

Akira R. Kinjo, Haruki Nakamura, Genji Kurisu
PDBj, Institute for Protein Research, Osaka University



wwpdb.org



Protein Data Bank

- PDB: 1st Open Access digital resource in biology (est. in 1971 with 7 entries)
- Initially, managed jointly by data centers in US and UK
- Today, single global PDB macromolecular structure archive (>138,000 entries)

Nature New Biology **233**, page 223 (1971)

(>26,000 entries with sugars)

CRYSTALLOGRAPHY

Protein Data Bank

A repository system for protein crystallographic data will be operated jointly by the Crystallographic Data Centre, Cambridge, and the Brookhaven National Laboratory. The system will be responsible for storing atomic coordinates, structure factors and electron density maps and will make these data available on request. Distribution will be on magnetic tape in machine-readable form whenever possible. There will be no charge for the service other than handling costs. Files will be updated as new material is received. The total holding will be announced annually in the organic bibliographic volumes of the reference series "Molecular Structures and Dimensions" published for the Crystallographic Data Centre and the International Union of Crystallography by Oosthoek's, Utrecht.

The success of the proposed system will depend on the response of the protein crystallographers supplying data. These will be

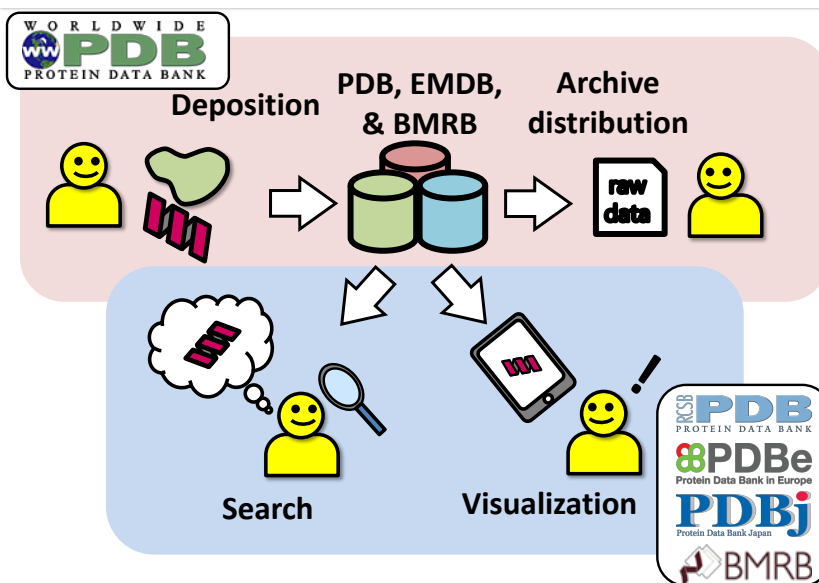
Worldwide Protein Data Bank (wwPDB)

- Ensures data are freely and globally available
 - Members
 - RCSB PDB (US)*Archive Keeper
 - PDBj (Osaka University, Japan)
 - PDBe (EMBL-EBI)
 - BioMagResBank (University Wisconsin, Madison, US)
- } Founding Members
- Collaborate on data processing and annotation
 - Each site provides different websites that offer different services and views of the data

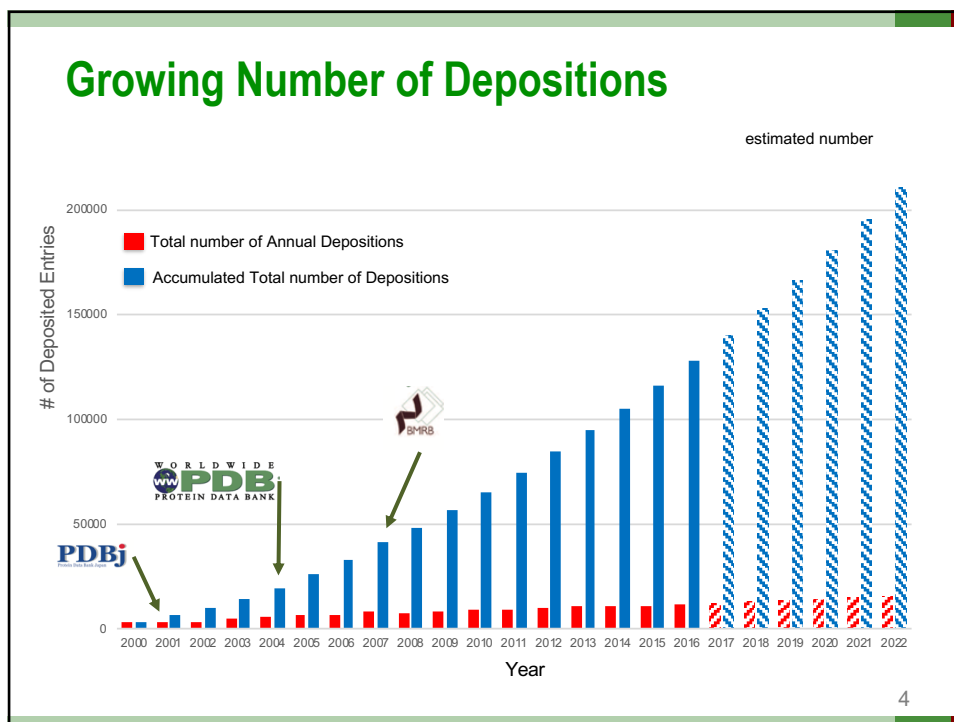


2

wwPDB collaboration



3



PDB File formats from wwPDB

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

Other data provided from wwPDB

- Validation Report
Translation to the RDF format is on going.
- PDB archive of Structure factors (for crystal structures)
- BMRB archive of NMR distance restraints
- EMDB archive of Cryo-TEM maps

6

mmCIF: an example

The screenshot shows the wwPDB website interface. The main content area is titled 'Resources' and lists various data formats available for download. The 'mmCIF' entry is highlighted with a pink oval. A pink arrow points from this entry to the 'ダウンロード' (Download) section on the right side of the page.

| フォーマット | ファイル名 (ファイルサイズ) | 操作 | |
|----------------|--|-----------------------------------|--------|
| 全ての情報 | pdb1qpf.ent.gz (109.89 KB) | ダウンロード | |
| 全ての情報 (非圧縮) | pdb1qpf.ent (456.73 KB) | ダウンロード | |
| ヘッダのみ | pdb1qpf.ent.gz (7.93 KB) | ダウンロード | |
| mmCIF | 1qpf.cif.gz (140.02 KB) | ダウンロード | |
| 全ての情報 | 1qpf.xml.gz (214.1 KB) | ダウンロード | |
| PDBML | ヘッダのみ | 1qpf-noatom.xml.gz (36.35 KB) | ダウンロード |
| | 最終情報のみ | 1qpf-exactom.xml.gz (120.21 KB) | ダウンロード |
| | 全ての情報 | 1qpf-plus.xml.gz (217.34 KB) | ダウンロード |
| PDBMLplus | ヘッダのみ | 1qpf-plus-noatom.xml.gz (39.6 KB) | ダウンロード |
| | 付録情報のみ | 1qpf-add.xml.gz (3.24 KB) | ダウンロード |
| RDF | 1qpf.rdf.gz (26.03 KB) | ダウンロード | |
| 構造因子 | 1qpf.ent.gz (558.08 KB) | ダウンロード | |
| 生物学的単位 (PDB形式) | 1qpf.ent.gz (105.21 KB) (A) *author_defined_assembly: 1 molecule(s) (monomeric) | ダウンロード | |
| PDF | 1qpf_validation.pdf.gz (231.36 KB) | ダウンロード | |
| PDF-full | 1qpf_full_validation.pdf.gz (296.43 KB) | ダウンロード | |
| 構造化レポート | XML | 1qpf_validation.xml.gz (32.32 KB) | ダウンロード |

The 'ダウンロード' (Download) section on the right lists various data formats available for download, including Sequence (fasta), PDBML (圧縮済み), PDBML (非圧縮), mmCIF, 構造レポート (PDF), and others. The 'mmCIF' entry is highlighted with a pink oval and a pink arrow pointing to it.

7

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.

8

More about mmCIF

- Context-free grammar (STAR [Self-defining Text Archive and Retrieval])
- All the categories and items are defined in the PDBx/mmCIF dictionary.
- For details, see <http://mmcif.wwpdb.org/>

9

A closer look at mmCIF

```

data_1GOF ← This tag starts a data ("datablock" is the unit of data [entry])
#
entry.id 1GOF ← entry ID (PDB ID) is "1GOF".
#
_audit_conform.dict_name    mmcif_pdbx.dic
_audit_conform.dict_version 5.287
_audit_conform.dict_location http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
_database_2.database_id    PDB
_database_2.database_code  1GOF
#
....

```

Provenance information is also included.

10

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.

11

Example of “loop” structure

```

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 ? ? ? ?
#

```

Start of a loop

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

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

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

12

Atomic coordinates in mmCIF

```

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

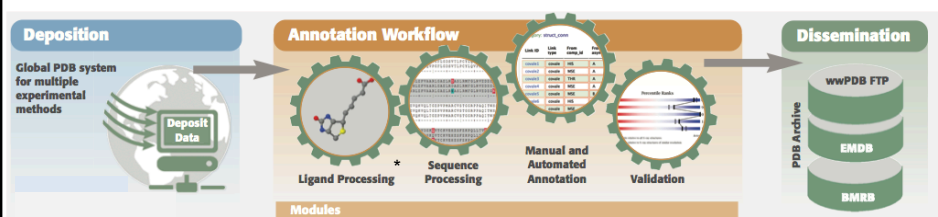
```

Main category groups

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

14

wwPDB Common Deposition & Annotation



- Enables workload balancing and has increased productivity
- Better quality assurance of polymer sequences and ligand chemistry
- PDBx/mmCIF is now the master file format
- Validation based on recommendations from expert task forces
- Federation with other Data Resources (e.g., EMDB, SASBDB, ...)

15

wwPDB Common Deposition & Annotation

<https://wwpdb.org>

The screenshot shows the wwPDB website homepage. At the top, there is a navigation bar with links for 'Validate Structure', 'Deposit Structure', 'All Deposition Resources', and 'Download Archive'. The 'Deposit Structure' link is circled in red. Below the navigation bar, there is a main content area with a large image of a protein structure and a 'OneDep' logo. The page is divided into several sections: 'wwPDB Members' (listing BMRB, PDBe, PDBj, and RCSB PDB), 'wwPDB Resources' (listing Data Dictionaries, Annotation, and Community Input), and 'News & Announcements' (listing recent updates and announcements).

16

wwPDB Common Deposition & Annotation

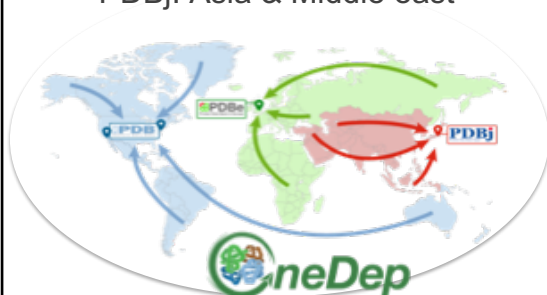
The screenshot shows the wwPDB OneDep System login and deposition page. The page is titled 'wwPDB OneDep System' and includes a 'Start a new deposition' section with the URL <http://deposit.wwpdb.org/deposition/>. There is a login form with fields for 'Deposition ID' and 'Password', and buttons for 'Log in' and 'Forgot Password'. Below the login form, there is a 'wwPDB regions' section with a world map and a 'wwPDB news' section. The 'Please select the location of the institute of your PI' instruction is circled in red. Below this instruction, there is a 'Country' dropdown menu with a list of countries including United Kingdom, United States, Japan, Afghanistan, Alder Islands, Albania, Algeria, American Samoa, Andorra, Angola, Anguilla, Antarctica, Antigua And Barbuda, Argentina, Armenia, Aruba, Australia, Austria, Azerbaijan, Bahamas, and Bahrain.

17

wwPDB Common Deposition & Annotation

- As of 2016 region-based processing of D&A-deposited entries:

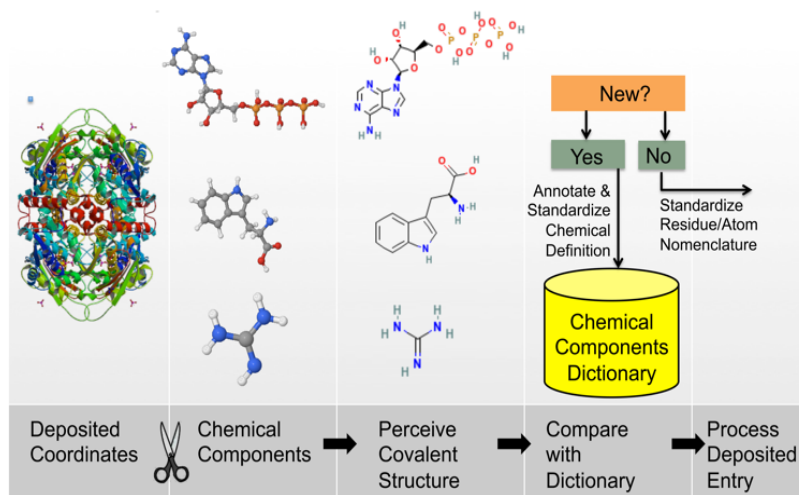
- RCSB PDB: Americas & Oceania
- PDBe: Europe & Africa
- PDBj: Asia & Middle east



| 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 | 6676 | 3563 | 2110 | 1005 |
| 2006 | 7282 | 4252 | 1945 | 1085 |
| 2007 | 8130 | 4703 | 2299 | 1128 |
| 2008 | 7073 | 4106 | 1994 | 973 |
| 2009 | 8300 | 5069 | 2173 | 1058 |
| 2010 | 8876 | 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 |

18

Chemical Component Deposition Pipeline



19

Chemical Component Dictionary (CCD)

- Complete descriptions of constituent small molecules in experimentally-determined 3D macromolecular structures in the PDB
- Data items include
 - Atom Nomenclature
 - Connectivity/Chirality
 - Chemical Formula, InChI/SMILES, etc.
 - Molecular Names
 - Idealized 3D Structure
 - 3D Structure Exemplar from PDB Archive

20

Chemical Component Dictionary Entry

a)

```

data_HTP
#
  _chem_comp_id          HTP
  _chem_comp_name       4-HYDROXYPROLINE
  _chem_comp_type       "L-PEPTIDE LINKING"
  _chem_comp_pdbx_type  ATOMP
  _chem_comp_formula    "C5 H9 N O3"
  _chem_comp_mol_wt     133.130
  _chem_comp_mol_weight 131.130
  _chem_comp_charge      0
  _chem_comp_release_date 1999-07-08
  _chem_comp_modified_date 2008-04-23
  _chem_comp_pdbx_release_status
  _chem_comp_pdbx_replaced_by
  _chem_comp_pdbx_replaces
  _chem_comp_formula_weight
  _chem_comp_chem_letter_code
  _chem_comp_chem_letter_code
  _chem_comp_pdbx_model_coordinates_details
  _chem_comp_pdbx_model_coordinates_missing_flag
  _chem_comp_pdbx_ideal_coordinates_details
  _chem_comp_pdbx_ideal_coordinates_missing_flag
  _chem_comp_pdbx_model_coordinates_ob_code
  _chem_comp_pdbx_processing_site
#
  
```

b)

```

loop
  _chem_comp_atom_comp_id
  _chem_comp_atom_atom_id
  _chem_comp_atom_alt_atom_id
  _chem_comp_atom_type_symbol
  _chem_comp_atom_charge
  _chem_comp_atom_pdbx_allen
  _chem_comp_atom_pdbx_aromatic_flag
  _chem_comp_atom_pdbx_leaving_atom_flag
  _chem_comp_atom_pdbx_stereo_config
  _chem_comp_atom_model_Cartn_x
  _chem_comp_atom_model_Cartn_y
  _chem_comp_atom_model_Cartn_z
  _chem_comp_atom_pdbx_model_Cartn_x_ideal
  _chem_comp_atom_pdbx_model_Cartn_y_ideal
  _chem_comp_atom_pdbx_model_Cartn_z_ideal
  _chem_comp_atom_pdbx_ordinal
HTP N  H  0  1  W  N  -3.266 14.585 44.188  0.169  1.369 -0.292 1
HTP CA  C  0  1  W  N  -2.955 15.769 43.044 -0.384 -0.603 -0.493 2
HTP O  O  0  1  W  N  -1.447 15.609 43.039 -0.813 -0.473 -0.013 3
HTP O  O  0  1  W  N  -0.722 14.484 43.563 -0.233  0.764  0.750 4
HTP CB  C  0  1  W  N  -2.468 16.576 43.629  0.315 -0.854 -0.259 5
HTP CC  C  0  1  W  N  -4.437 17.482 42.339  1.447 -0.139  0.595 6
HTP CD  C  0  1  W  N  -4.948 17.483 43.753  1.840  1.159 -0.271 7
HTP OD1 O  0  1  W  N  -5.693 16.815 42.294  2.817 -0.911 -0.071 8
HTP OXT O  0  1  W  N  -0.976 14.502 42.469 -2.614 -1.683 -0.423 9
HTP H  H  0  1  W  N  -3.480 16.847 44.765 -0.107  1.981 -0.020 10
HTP HA  HA  0  1  W  N  -3.285 14.794 43.068 -0.225 -0.278 -1.540 11
HTP HB2 HB  0  1  W  N  -2.247 17.141 42.398  0.864 -1.092 -1.375 12
HTP HB3 ZB  0  1  W  N  -3.790 15.930 41.026  0.678 -1.873 -0.153 13
HTP HB2 HB  0  1  W  N  -4.508 16.399 42.726  2.052  0.048 -1.505 14
HTP HB2 HB  0  1  W  N  -4.958 18.005 44.370  3.018  1.665 -1.289 15
HTP HB2 HB  0  1  W  N  -3.457 16.733 43.448  1.332  1.965 -0.243 16
HTP BD1 BD  0  1  W  N  -5.999 14.664 42.181  3.769 -0.479 -0.009 17
HTP HXT HXT  0  1  W  N  -0.027 14.511 42.499 -3.320 -1.944 -0.698 18
#
  
```

c)

```

loop
  _chem_comp_bond_comp_id
  _chem_comp_bond_atom_id_1
  _chem_comp_bond_atom_id_2
  _chem_comp_bond_valence_order
  _chem_comp_bond_pdbx_aromatic_flag
  _chem_comp_bond_pdbx_stereo_config
  _chem_comp_bond_pdbx_ordinal
HTP N  CA  S  S  S  N  1
HTP N  CD  S  S  S  N  2
HTP N  B  S  S  S  N  3
HTP CA  CS  S  S  S  N  4
HTP CA  CS  S  S  S  N  5
HTP CA  BA  S  S  S  N  6
HTP C  O  S  S  S  N  7
HTP C  OXT S  S  S  N  8
HTP C  O  S  S  S  N  9
HTP CB  HB2 S  S  S  N  10
HTP CB  HB3 S  S  S  N  11
  
```

d)

```

#
  _pdbx_chem_comp_descriptor_comp_id
  _pdbx_chem_comp_descriptor_type
  _pdbx_chem_comp_descriptor_program
  _pdbx_chem_comp_descriptor_program_version
  _pdbx_chem_comp_descriptor_descriptor
HTP SMILES  ACGLABS  12.01 O=C(O)C(=O)O=C1
HTP SMILES  CACTVS  3.895 O=C(O)C(=O)O=C1C(C)O=C1
HTP SMILES  CANONICAL  1.2-5 C1C(=O)C(=O)C(=O)O=C1O
HTP SMILES  "OpenEye OEToolkits"  1.7-5 C1C(=O)C(=O)O=C1O
HTP InChI  InChI  1.03 "InChI=1S/C5H9NO3/c7-3-1-(5(8)9)6-2-3/h3-4,6-7H,1
HTP InChIKey  InChIKey  1.03 PHTTEVYHMQDQ-INTCNVIGSA-N
#
loop
  _pdbx_chem_comp_descriptor_comp_id
  _pdbx_chem_comp_descriptor_type
  _pdbx_chem_comp_descriptor_program
  _pdbx_chem_comp_descriptor_program_version
  _pdbx_chem_comp_descriptor_descriptor
  
```

Atom names

Stereochemistry & aromaticity

Model coordinates

Ideal coordinates

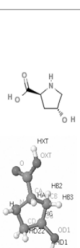
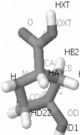
Connected atoms

Bond type

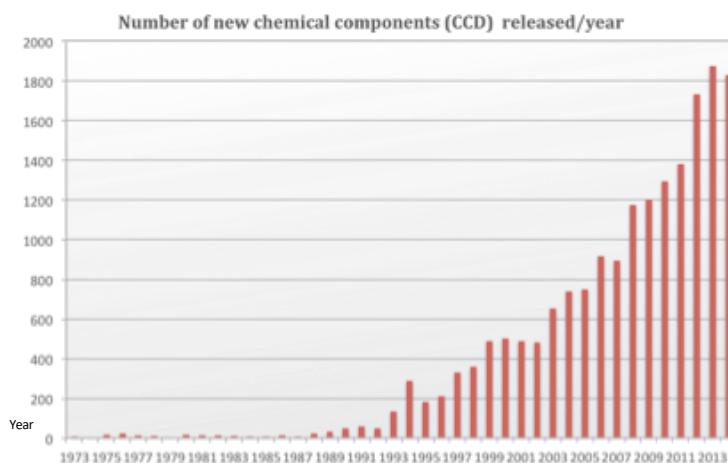
Stereochemistry & aromaticity

SMILES

InChI 21

Growth of Chemical Components in PDB

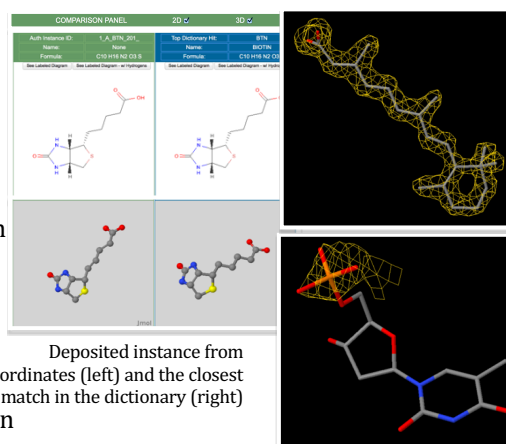


Growth trends new Chemical Components in the PDB

22

Improved Ligand Annotation

- 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



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**)

23

wwPDB/CCDC/D3R Ligand Validation Workshop

Meeting Objectives: To bring together co-crystal structure determination experts from Academe and Industry with Crystallography and Computational Chemistry Software Developers to discuss, develop, and recommend:

- Best practices PDB archive deposition/validation of co-crystal structures
- Editorial/Refereeing/Publication standards for co-crystal structures
- Improvements in ligand representation across the PDB Archive



24

Workshop White Paper

- White Paper describing recommendations re deposition/validation and editorial/refereeing/publication standards is published in *Structure* 24, 502-508 (2016)

CellPress

Structure
Meeting Report

Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop

Paul D. Adams,¹ Kathleen Aertgeerts,² Cary Bauer,³ Jeffrey A. Bell,⁴ Helen M. Berman,^{5,6} Talapady N. Bhat,⁷ Jeff M. Blaney,⁸ Evan Bolton,⁹ Gerard Bricogne,¹⁰ David Brown,^{11,12} Stephen K. Burley,^{5,6,13,*} David A. Case,⁶ Kirk L. Clark,¹⁴ Tom Darden,¹⁵ Paul Emsley,¹⁶ Victoria A. Feher,^{17,*} Zukang Feng,^{5,6} Colin R. Groom,^{18,*} Seth F. Harris,⁸ Jorg Hendle,¹⁹ Thomas Holder,⁴ Andrzej Joachimiak,²⁰ Gerard J. Kleywegt,²¹

(Author list continued on next page)

25

Improved Validation

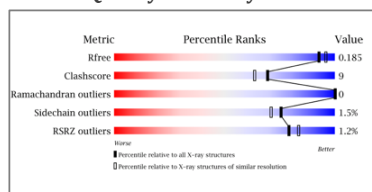
- wwPDB Validation Task Forces X-ray, NMR, SAS
- wwPDB/EMDataBank VTF for EM
- Recommendations about validating new and existing structures
 - Implemented in software pipeline
 - Produces summary report (PDF) and XML file with detailed statistics
- Validation at different stages
 - While determining/depositing the structure
 - After annotation (official; should be sent to journals)
 - Upon release (publicly available; updated annually)

26

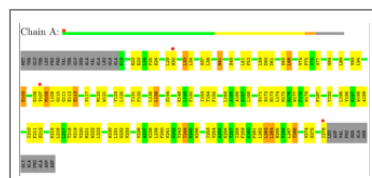
X-ray Validation Report

- Model Quality
 - Bond lengths and angles (outlier info, RMS-Z)
 - Chirality, planarity
 - Close contacts (including worst clashes, MolProbity clash score)
 - Torsion angles (Ramachandran statistics, protein rotamers)
 - Ligand geometry (Mogul analysis)
- Residue Plots
 - Residues with model-quality outliers (0, 1, 2, >2)
 - Residues with RSR-Z > 5 are highlighted
 - Residues not observed

Overall Quality Summary



Residue Plots



27

Validation Report is requested for peer review

EDITORIAL

nature
structural &
molecular biology

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

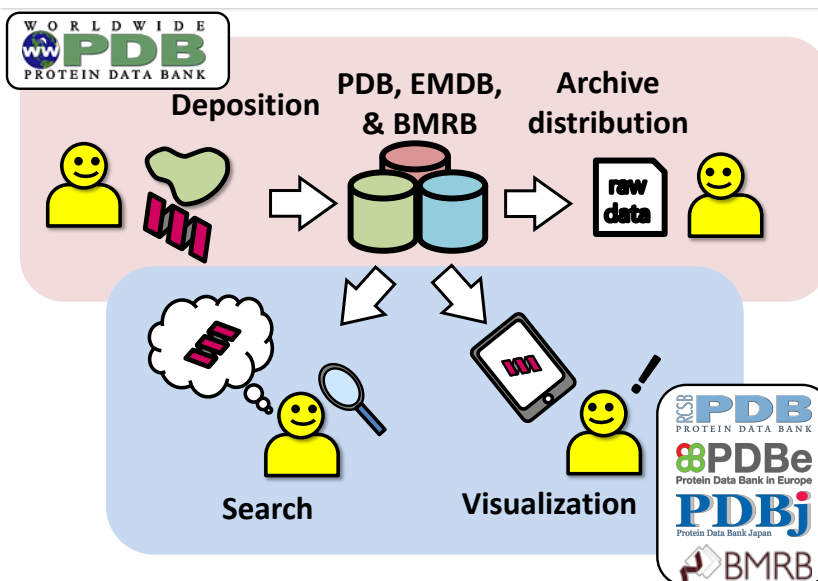
Where are the data?

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

We are now taking a further step and are **requesting official wwPDB validation reports for peer review**. These reports are made available by the wwPDB after data deposition (<http://www.wwpdb.org/validation/validation-reports>). Other Nature journals will soon follow suit.

28

wwPDB collaboration



29

Worldwide Protein Data Bank (wwPDB)

The collage shows various parts of the PDB website: a structural view of a protein, a news article about carbohydrates, a guide for new visitors, and a search interface for the Biological Magnetic Resonance Data Bank.

30

Wh

The screenshot displays the PDB entry page for 5BVU, titled 'Crystal structure of Thermoanaerobacterium xyloxyticum GH116 beta-glucosidase'. It includes a 3D structure view, classification as a hydrolase, and a wwPDB Validation plot showing metrics like R-free, Clashscore, Ramachandran outliers, and Sidechain outliers.

31

DOI Landing Page Layout (Planned)

WORLDWIDE PDB PROTEIN DATA BANK

VALIDATION • DEPOSITION • DATA DICTIONARIES • DOCUMENTATION • TASK FORCES • STATISTICS • ABOUT • wwPDB Foundation

About

- Contact Us
- Publications
- Advisory Committee
- Agreement
- Outreach
- FAQ

Data Access Options for 10.2210/pdb1kip/pdb

This page presents data access options for PDB entry 10.2210/pdb1kip/pdb. Questions about this page may be sent to info@wwpdb.org.

Data download options:

- Structure coordinates (PDBx/mmCIF)
- Structure coordinates (PDBML)
- Structure coordinates (PDB)
- Structure coordinates (PDF)
- X-ray diffraction data (PDBx/mmCIF)

Further resources for entry 10.2210/pdb1kip/pdb at: PDBx PDB RCSB PDB

Download Archive

RCSB PDB ftp | PDBx ftp | PDBj ftp

Instructions

Archive Snapshots

RCSB PDB | PDBj

Cite wwPDB:

Nature Structural Biology 10, 980 (2003)
doi: 10.1038/nsb1203-980

More publications

News & Announcements

Members:

PDBj BMRB PDBe PDB

wwPDB Foundation

© wwPDB

32

PDBj Mine2 RDB (<https://pdbj.org>)

PDBj 132905

English 日本語 简体中文 繁體中文 বাংলা

Search

132905

Guide for first time visitors

For an introduction to the new web interface, please read [Using PDBj](#) and [PDBj](#). An introduction to the customization features offered by the new PDBj web interface can be found [here](#). To get a more in-depth explanation on the various features of the PDBj website, please take a look at the [Interactive Tutorial](#).

The [PDBj website](#) will no longer be updated after July 12, 2017 and will be closed at the end of August 2017.

- Relational database working behind the PDBj.
- Docs: <https://pdbj.org/mine-rdb-docs>
 - Complete database schema with diagrams.
- Web SQL interface: <https://pdbj.org/mine>
- REST API: https://pdbj.org/rest/mine2_sql
- SQL dump: <ftp://ftp.pdbj.org/mine2/>
 - Requires **PostgreSQL** >= 9.3
 - See <https://pdbj.org/help/mine2-rdb-local-install>
- Many examples: <https://pdbj.org/help/mine2-sql>

33

Schema diagrams
<https://pdbj.org/mine-rdb-docs>

The screenshot shows the PDBj website interface. On the left, there is a sidebar with navigation options like 'Home', 'SQL Search', 'Download', 'New Format', 'Quick links', and 'Search services'. The main content area displays the search results for '133093', including a table with columns like 'Name', 'Chemical ID', 'PDB ID', and 'Header'. A diagram on the right side of the page illustrates the relationships between various database tables. The tables shown include 'chem_comp_bond', 'chem_comp_atom', 'chem_comp_plane_atom', and 'chem_comp_atom: partial_charge'. Arrows indicate foreign key relationships between columns in different tables. For example, 'chem_comp_bond' has columns 'atom_id_1' and 'atom_id_2' that reference 'chem_comp_atom'. 'chem_comp_atom' has columns like 'atom_id', 'atm_name', 'model_cartn_x', 'model_cartn_y', 'model_cartn_z', 'partial_charge', 'pdbev_all_atom_id', 'pdbev_all_comp_id', 'pdbev_all_entity_id', 'pdbev_all_entity_type', and 'pdbev_all_id'. 'chem_comp_atom: partial_charge' has columns 'atom_id', 'number of non-null entries', and 'The partial charge assigned to this atom'.

A running example on PDBj web site.

34

Integration with SIFTS

- “Structure Integration with Function, Taxonomy and Sequence” developed by PDBe & UniProt.
 - <https://www.ebi.ac.uk/pdbe/docs/sifts/>
- Integrates *UniProt*, *NCBI Taxonomy*, *Gene Ontology*, *Pfam*, *EC code*, *PubMed*, *SCOP*, *CATH* with PDB.
- The “Quick access” data of SIFTS are integrated into the PDBj Mine RDB.
 - c.f., <https://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>

New integration

- Chemical Component Dictionary (cc)
 - PDB's (3-letter) chemical components
 - Includes InChi keys, SMILES, etc.
- Chemical Component Model Data (ccmodel)
 - Xref to *Cambridge Structure Database (CSD)*
- BIRD (prd)
 - “Biologically Interesting Molecule Reference Dictionary”
 - Peptide-like antibiotic and inhibitor molecules.

36

Example1: Gleevec using PDBj-Mine

The screenshot shows the PDBj website interface. At the top, there is a header with the PDBj logo (Protein Data Bank Japan) and the number 132905. There are language options: English, 日本語, 繁体中文, 繁體中文, and 한국어. A search bar is present with the text 'Search pdbj.org'. Below the header is a 'Chemie search' section with a question mark icon. Underneath, it says 'Search of Chemical Component Dictionary'. There are several input fields: 'Quick search:' with 'Gleevec' entered, 'Code (comp_id):', 'Molecular name:', 'Formula:', 'SMILES:', and 'InChi:'. At the bottom of the search form are 'Search' and 'Reset' buttons. The footer contains social media icons, the PDB logo, and copyright information: 'Copyright © 2013-2017 Protein Data Bank Japan'.

37

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

The screenshot shows the PDBj search results page for the query 'Gleevec'. The search returned 8 results. The top three results are highlighted:

- 1XBB**: CRYSTAL STRUCTURE OF THE SYK TYROSINE KINASE DOMAIN WITH GLEEVEC. Description: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-(4-METHYL-3-(4-PYRIDIN-3-YL-PYRIDIN-2-YLMINO)-PHENYL) BENZAMIDE, Tyrosine-kinase kinase SYK. Authors: Saitoh, S., Kuroki, S., Sakai, I.H., Banno, J., Kubota, M.D., Ishi, I., Fujita, K.I., Gu, X., Yoshii, J., Kamei, A., Imai, K., Nishikawa, K., Nakai, S.S., Oka, S., Sakagami, A.S., Sakita, A., Sasaki, S., Sato, S.S., Kubota, S.S.
- 3GVU**: THE CRYSTAL STRUCTURE OF HUMAN ABL3 IN COMPLEX WITH GLEEVEC. Description: 4-(4-METHYL-PIPERAZIN-1-YLMETHYL)-N-(4-METHYL-3-(4-PYRIDIN-3-YL-PYRIDIN-2-YLMINO)-PHENYL) BENZAMIDE, Tyrosine-kinase kinase ABL3. Authors: Uchibayashi, S., Sato, E., Sato, A., Nakano, T., Shirota, S., Saitoh, T., Chikuda, A., Mizukami, T., Imai, A., Ito, A.C.P., van Driel, R., Boucher, C., Akamatsu, C., Nakai, S., Uehara, A., Imai, S., Structural Genomics Consortium (SGC).
- 1XBA**: CRYSTAL STRUCTURE OF APO SYK TYROSINE KINASE DOMAIN. Description: Tyrosine-kinase kinase SYK. Authors: Saitoh, S., Kuroki, S., Sakai, I.H., Banno, J., Kubota, M.D., Ishi, I., Fujita, K.I., Gu, X., Yoshii, J., Kamei, A., Imai, K., Nishikawa, K., Nakai, S.S., Oka, S., Sakagami, A.S., Sakita, A., Sasaki, S., Sato, S.S., Kubota, S.S.

The screenshot shows the structural details page for PDB entry 1T46. The title is 'STRUCTURAL BASIS FOR THE AUTOINHIBITION AND STI-571 INHIBITION OF C-KIT TYROSINE KINASE'. The entry is a monomer (1 chain) with a total molecular weight of 36273.8. The primary citation is: 'Mori, C.S., Ouyang, D.R., Schneider, T.R., Stone, R.J., Kitah, M., Schiele, D.R., Small, C.P., Zhou, H., Song, B.C., Wilson, K.P. Structural basis for the autoinhibition and STI-571 inhibition of c-KIT tyrosine kinase. J Biol Chem. 279:31659-31663, 2004. doi:10.1074/jbc.M411220200. [Open in PubMed] [Open in CAS] [Open in NCBI] [Open in Europe PMC]'. The experimental method is X-RAY DIFFRACTION (2.6 Å). A red dashed circle highlights the 'Electron Density map (Molmil)' link in the 'Database Information' section.

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

The screenshot shows the PDBj-Mine Electron Density Map Viewer for PDB entry 1T46. The main view displays a 3D molecular model with electron density maps overlaid. The interface includes a 'Style' section with 'Wireframe' and 'Color' options, and a 'Parameters for Electron Density Map' section with various settings like 'Type of the map', 'Map position', 'mapped area', 'contour level', 'color', and 'isosurface transparency level'. Below this is a table for 'Electron Density Map Download/Delete' with columns for 'file format', 'filename', and 'Download/Delete' buttons.

| file format | filename | Download/Delete |
|------------------|----------------------------|-----------------|
| structure factor | 1T46f.ent.gz | Download |
| refinement file | 1T46_ref.tot.gz | Download |
| log4 file | 1T46.log4.gz | Download |
| edmap file | 201701151711_1T46_e.ent.gz | Delete Download |

40

Example 2. PDB entries containing "HEM"

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

The screenshot shows the PDBj website for PDB entry 1O1M. The 'mmCIF tree view' is displayed, listing various data categories. The 'pdbx_entity_nonpoly' category is highlighted, and a table below it shows the contents of this category.

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

"pdbx_entity_nonpoly" category

PDB entries containing “HEM” sorted 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 search query is entered in the 'SQL Search' field. The results are displayed in a table format, showing the PDB ID and the count of HEM residues in the asymmetric unit. The top results are:

| PDB ID | cnt |
|--------|-----|
| 5k9k | 96 |
| 2I7a | 84 |
| 4u8u | 36 |
| 4rkn | 32 |
| 2yr0 | 28 |

Example 3. BIRD: Biologically Interesting Molecule Reference Dictionary

- See <https://www.wwpdb.org/data/bird>
- Antibiotics, inhibitors, etc.

In 1KQE:

```
_pdbx_molecule_features.prd_id      PRD_000154
_pdbx_molecule_features.name      'MINI-GRAMICIDIN A DIMER'
_pdbx_molecule_features.type      Polypeptide
_pdbx_molecule_features.class      Antibiotic
_pdbx_molecule_features.details
;THE N-TERMINI OF THE TWO IDENTICAL PEPTIDES, EACH
A TRUNCATED GRAMICIDIN A WERE LINKED BY A SUCCINIC
ACID IN A HEAD-TO-HEAD MANNER.
;
#
```

Combining with BIRD

Find PDB entries containing antibiotics of molecular weight less than 1000 Da.

```
SELECT mf.pdbid, rm.name
FROM pdbj.pdbx_molecule_features mf
JOIN prd.pdbx_reference_molecule rm
      ON rm.prd_id = mf.prd_id
WHERE rm.class = 'Antibiotic'
AND   rm.formula_weight < 1000.0
```

The "prd" schema (for some historical reasons...)

<https://pdbj.org/mine-rdb-docs?schema=prd>

44

List BIRD entries or their types according to popularity

```
SELECT prd_id, name, COUNT(pdbid)
FROM pdbx_molecule_features
GROUP BY prd_id,name
ORDER BY COUNT DESC
```

Total number of results: 835

| | | |
|--------------------|--|-----------|
| prd_id: PRD_000020 | name: D-Phe-Pro-Arg-CH2Cl | count: 51 |
| prd_id: PRD_000238 | name: Ac-Asp-Glu-Val-Asp-CHK | count: 46 |
| prd_id: PRD_000142 | name: Cyclosporin A | count: 30 |
| prd_id: PRD_000398 | name: N-((2S)-2-[[N-acetyl-L-threonyl-L-isoleucyl]amino]hexyl)-L-norleucyl-L-glutamyl-N-5-[[amino(mino)methyl]-L-ornithinamide | count: 29 |
| prd_id: PRD_001243 | name: CARFILZOMIB, bound form | count: 28 |
| prd_id: PRD_000454 | name: Saquinavir | count: 27 |
| prd_id: PRD_000557 | name: Pepstatin | count: 24 |

```
SELECT type, COUNT(pdbid)
FROM pdbx_molecule_features
GROUP BY type
ORDER BY COUNT DESC
```

Total number of results: 17

| | |
|--------------------------|-------------|
| type: Peptide-like | count: 1032 |
| type: Oligopeptide | count: 146 |
| type: Cyclic peptide | count: 130 |
| type: Polypeptide | count: 117 |
| type: Glycopeptide | count: 38 |
| type: Cyclic desipeptide | count: 23 |
| type: Thiopeptide | count: 18 |
| type: Peptaibol | count: 15 |
| type: Non-polymer | count: 8 |
| type: Cyclic lipopeptide | count: 3 |
| type: Lipopeptide | count: 3 |

45

Example 4. Combining with CC model

Find PDB entries containing a compound corresponding to a Cambridge Structure Database (CSD) entry.

```
SELECT p.pdbid, p.id, p.name, r.db_code
FROM pdbj.chem_comp p
JOIN ccmode1.pdbx_chem_comp_model m
      ON m.comp_id = p.id
JOIN ccmode1.pdbx_chem_comp_model_reference r
      ON r.model_id = m.model_id
WHERE r.db_name = 'CSD' AND r.db_code = 'YARXEW'
```

The "ccmodel" schema.

<https://pdbj.org/mine-rdb-docs?schema=ccmodel>

46

Combining with CC

Find PDB entries containing monomers with the given InChIKey.

```
SELECT p.pdbid, p.id
FROM pdbj.chem_comp p
JOIN cc.pdbx_chem_comp_descriptor cc
      ON cc.comp_id = p.id
WHERE cc.type = 'InChIKey'
AND cc.descriptor = 'ZKHQWZAMYRWXGA-KQYNXXCUSA-N'
```

The "pdbj" schema is the default and can be omitted.

Chemical Component Dictionary entries are under the "cc" schema.

For complete information:

<https://pdbj.org/mine-rdb-docs?schema=cc>

47

Integration with GlyTouCan

Listing PDB ID's that contain GlyTouCan entries.

SQL Search

```
SELECT cc.pdbid, g.acc
FROM glytoucan.chem_comp g
JOIN chem_comp cc
  ON cc.id = g.chem_comp_id
```

Total number of results: 26005

1308

acc: G50720WY

1309

acc: G50720WY

131f

acc: G50720WY

133y

acc: G03144EF

134e

acc: G50720WY

136e

acc: G50720WY

138f

acc: G50720WY

13ca

acc: G50720WY

1365

acc: G50720WY

136f

acc: G50720WY

138h

acc: G50720WY

13af

acc: G50720WY

The "glytoucan" schema contains only 1 table: chem_comp kindly provided by the GlyTouCan team & Dr. I. Yamada.



48

Another CC example (sugars!)

SQL Search

Enter search query:

```
SELECT comp_id, type, name
FROM cc.chem_comp
WHERE type ILIKE '%saccharide%'
```

Total number of results: 536

comp_id: 045

type: D-SACCHARIDE

name: beta-D-fructofuranosyl-(2->6)-beta-D-fructofuranosyl alpha-D-glucopyranoside

comp_id: 0AT

type: D-SACCHARIDE

name: 2-(deoxyamino)-2-deoxy-6-O-phosphono-alpha-D-glucopyranose

comp_id: 0BD

type: D-saccharide

name: 3-methyl-1-(2-methylpropyl)butyl 4-O-beta-L-galactopyranosyl-beta-D-glucopyranoside

comp_id: 0MK

type: L-SACCHARIDE

name: L-ribose

comp_id: 0NZ

type: SACCHARIDE

name: 2-deoxy-6-O-phosphono-beta-D-arabino-hexopyranose

comp_id: 0TS

type: D-saccharide 1,4 and 1,4 linking

name: beta-D-glucopyranosyl-(1->4)-4-thio-beta-D-glucopyranosyl-(1->4)-beta-D-glucopyranosyl-(1->4)-4-thio-beta-D-

Find all "saccharides" from the Chemical Component Dictionary.

There are too many varieties in annotation: D-SACCHARIDE, D-saccharide, etc. :(

49

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

Feel free to ask any questions at:
<https://pd bj.org/contact?tab=PDBjmaster>

50

Acknowledgements



51