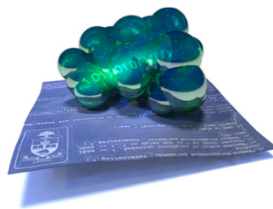


A
BIOINFORMATICS
COURSE

PDB



BORIS STEIPE

*DEPARTMENT OF BIOCHEMISTRY – DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO*

PDB http://rcsb.org/

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to **123456** Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligand Go

Advanced Search | Browse by Annotations

PDB-101 PDB EMDatabank Nucleic Acid Database Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

A Structural View of Biology

This resource is powered by the Protein Data Bank archive—information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Events and Activities

October Molecule of the Month

Dipeptidyl Peptidase 4

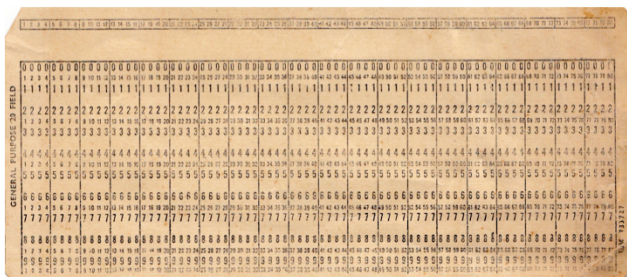
pd101.rcsb.org/motm Contact Us

...the **PDB** (Protein structure DataBase) is **the** central repository for 3D structural data of proteins and nucleic acids. RCSB is the Research Collaboratory for Structural Bioinformatics.

If you look closely at the screenshot, you will see that the PDB held 123,456 coordinate files (as of October 16, 2016). That's the actual number, not a placeholder. How cool is that! (Apologies for this completely useless bit of trivia. It's 134,091 in October 2017.)

PDB FILE FORMAT

- Flat file, strictly column oriented, 80 character records
- Human readable
- **Human editable**



80-character punch card. (From Wikipedia)

Flat File: A datafile without indexing structure or hierarchy. In contrast, to *relational database*, or *data grammar*.

The “PDB” file format is the most widely used format to exchange structure data. It is a keyword controlled, column-oriented, fixed-width flat file format, derived from the information processing constraints of computer punch cards.

For decades, some researchers have been advocating for a more rigorous and flexible XML based data grammar, but the existing solutions have never been widely adopted.

PDB FILE FORMAT COMPONENTS: HEADER

```
HEADER      IMMUNOGLOBULIN                      01-MAR-93   2IMM   2IMM   2
COMPND      IMMUNOGLOBULIN VL DOMAIN (VARIABLE DOMAIN OF KAPPA LIGHT
COMPND      2 CHAIN) OF MCPC603                      2IMM   3
SOURCE      HUMAN (HOMO $SAPIENS) RECOMBINANT SYNTHETIC M603 GENE 2IMM   4
AUTHOR      B.STEIPE,R.HUBER                          2IMM   5
REVDAT      1 15-JUL-93 2IMM 0                          2IMM   6
REMARK      1                                          2IMM   7
REMARK      1 REFERENCE 1                               2IMM   8
REMARK      1 AUTH  B.STEIPE,A.PLUCKTHUN,R.HUBER      2IMM   9
REMARK      1 TITL 2 REFINED CRYSTAL STRUCTURE OF A RECOMBINANT 2IMM 10
REMARK      1 TITL 3 IMMUNOGLOBULIN DOMAIN AND A      2IMM 11
REMARK      1 TITL 3 COMPLEMENTARITY-DETERMINING REGION 1-GRAFTED MUTANT 2IMM 12
REMARK      1 REF  J.MOL.BIOL. V. 225 739 1992        2IMM 13
REMARK      1 REFN ASTM JMOBAK UK ISSN 0022-2836      070 2IMM 14
REMARK      1 REFN ASTM JMOBAK UK ISSN 0022-2836      070 2IMM 15

[...]
```

```
REMARK      2                                          2IMM 23
REMARK      2 RESOLUTION. 2.00  ANGSTROMS.             2IMM 24
REMARK      3                                          2IMM 25

[...]
```

The header section of a PDB file.

PDB FILE FORMAT COMPONENTS: SEQUENCE

```
[...]  
SEQRES  1  114  ASP ILE VAL MET THR GLN SER PRO SER SER LEU SER VAL  2IMM  35  
SEQRES  2  114  SER ALA GLY GLU ARG VAL THR MET SER CYS LYS SER SER  2IMM  36  
SEQRES  3  114  GLN SER LEU LEU ASN SER GLY ASN GLN LYS ASN PHE LEU  2IMM  37  
SEQRES  4  114  ALA TRP TYR GLN GLN LYS PRO GLY GLN PRO PRO LYS LEU  2IMM  38  
SEQRES  5  114  LEU ILE TYR GLY ALA SER THR ARG GLU SER GLY VAL PRO  2IMM  39  
SEQRES  6  114  ASP ARG PHE THR GLY SER GLY SER GLY THR ASP PHE THR  2IMM  40  
SEQRES  7  114  LEU THR ILE SER SER VAL GLN ALA GLU ASP LEU ALA VAL  2IMM  41  
SEQRES  8  114  TYR TYR CYS GLN ASN ASP HIS SER TYR PRO LEU THR PHE  2IMM  42  
SEQRES  9  114  GLY ALA GLY THR LYS LEU GLU LEU LYS ARG  2IMM  43
```

Explicit (*above*) and implicit (*from coordinates*) sequence may differ !

Always be extremely cautious with sequence numbers! There is no well defined starting point for numbering sequence!! The numbers you count in a FASTA file and the numbers you read in a PDB record may be totally different!!!

The *explicit* sequence is found in the SEQRES section of the PDB file. This corresponds to what the crystallographer put into the experiment and it is in general the sequence that is recorded as the PDB-sequence for the purpose of database searches.

The *implicit* sequence can be derived from the actual coordinates. It is frequently not the same because flexible, disordered termini and loops are not recorded.

Additional complications arise from “insertion codes”. These are letters that allow the insertion of residues for a common numbering scheme for families of homologous sequences. In principle this is a good idea, since this makes comparison of residues much easier. But strings such as “23A” can no longer be treated as sequence *numbers* - they are sequence *labels* and using them correctly can be a challenge. The existence of insertion codes breaks common assumptions that novice programmers make about the structure of PDB files.

PDB FILE FORMAT COMPONENTS: COORDINATE RECORDS

```
ATOM      119  CA  ARG A  18           8.386  51.105  35.847  1.00  7.30           C
```

```
HETATM    877  O   HOH A  321        -4.169  60.050  40.145  1.00  3.00           O
```

**Record
type**

The PDB format is **strictly** column oriented !

The coordinate records contain the actual 3D-data. Their structured is the same for ATOM (protein and DNA components) and HETATM records (ligands, cofactors, solvent etc.).

Read the Coordinate section of the PDB format specification.
(<http://www.wwpdb.org/documentation/file-format>)

COORDINATE RECORDS

Pitfalls:

Atomname is a mix of Chemical element and bond topology. "CA.." ≠ ".CA."

Sequence number is actually a string - Chain and insertion code are required to make it unique (e.g **B 123A**).

ATOM	119	CA	ARG	A	18	8.386	51.105	35.847	1.00	7.30	C
------	-----	----	-----	---	----	-------	--------	--------	------	------	---

Atom number

Amino acid type

Sequence number

Chain

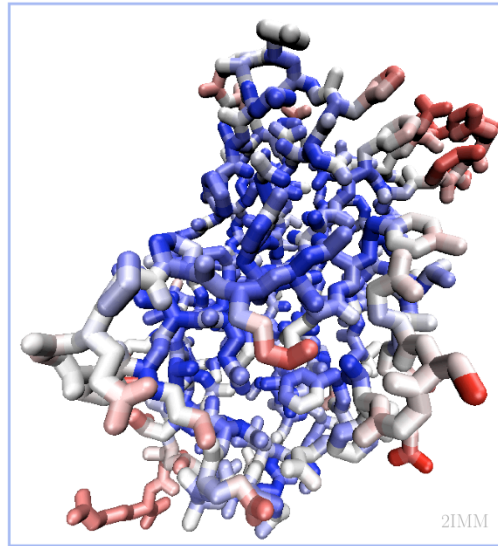
Atom name

The PDB format is **strictly** column oriented !

Potential pitfalls (very incomplete list):

- Record type: changes not consistently applied for modifications
- Atom number: rarely used and a nuisance to update when changing.
- Atom name: be careful about columns " CA " is C^α, but "CA " is calcium. Usually this is obvious from context though.
- Amino acid type: careful about non-natural amino-acids e.g. selenocysteine, selenomethionine etc. Some very old files use TRY for TRP.
- Chain: may be blank (" ") in older files. Recently was changed to ("A") even in files that contain only a single chain.
- Alternate location: only sometimes given in very high resolution structures.
- Sequence number might not be unique if insertion codes are used.
- X, Y, and Z are given in Å; (10⁻¹⁰m = 0.1 nm) values in a cartesian (i.e. orthogonal) coordinate system; but origin and orientation is arbitrary!
- Occupancy can describe: special locations, partially bound ligands, unobserved fragments of structure...
- B-values, (also called temperature factors) are a measure of the volume of space around into which a given electron density is distributed. Usually considered to be isotropic!

COORDINATE RECORDS



ATOM	119	CA	ARG	A	18	8.386	51.105	35.847	1.00	7.30	C
------	-----	----	-----	---	----	-------	--------	--------	------	------	---

$$B = 8\pi^2 \left\langle \vec{x}^2 \right\rangle$$

$$B = 80; \bar{x} \approx 1.0 \text{ \AA}$$

Occupancy

B

Occupancy describes the fraction of sites that are occupied by that atom in the crystal. Occupancy can be less than 1.0 for ligands of partially occupied sites, or for atoms in special crystallographic positions – e.g. on a two-fold axis of symmetry. Modelled coordinates that are not supported by electron density from the experiment and have only been added for cosmetic reasons are usually given an occupancy of 0.0

B factors describe the thermal fluctuation of atoms and can be interpreted as a mean coordinate displacement IF the thermal fluctuation were isotropic (the same in each direction). A *B* factor of 80 is approximately what you would expect of bulk, disordered solvent, it corresponds to a mean displacement of approximately one Å. *B* factors are commonly very low in the well-ordered and conformationally constrained core of a protein, they are high at the surface, especially for side-chains and flexible loops.

The interpretation of *B* factors is not entirely straightforward, because poor quality crystals can give rise to higher values, actual movement in proteins is anything but isotropic, and static disorder / alternate conformations are modelled through increased *B* factors as well.

A PDB FILE CONTAINS

Contents of a PDB file:

- Polypeptide Chain(s)
- Nucleic Acid(s)
- Heteroatoms e.g. prosthetic groups, cofactors, ligands ...
- Typically, H-Atoms are invisible ...

Structures are in a crystal lattice, have crystal contacts.

Asymmetric units are not guaranteed to be biological units.

Origin of the coordinate system and rotation is arbitrary.

Coordinates are time- and population- averaged experiments.

PDB QUERY ... by PDB ID

rcsb.org

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures
 PROTEIN DATA BANK

Search by PDB ID, author, macromolecule, sequence, or ligand Go

Advanced Search | Browse by Annotations | Search History (1) | Previous Results (123456)

PDB-101 PDB EMDataBank Nucleic Acid Database Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

Biological Assembly 1 Display Files Download Files

2IMM

Refined crystal structure of a recombinant immunoglobulin domain and a complementarity-determining region 1-grafted mutant

DOI: 10.2210/pdb2imm/pdb

Classification: [IMMUNOGLOBULIN](#)

Deposited: 1993-03-01 Released: 1993-07-15

Deposition author(s): [Steipe, B.](#), [Huber, R.](#)

Organism: [Mus musculus](#)

Structural Biology Knowledgebase: 2IMM [SBKB.org](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
 Resolution: 2.0 Å
 R-Value Work: 0.149

wwPDB Validation 3D Report Full Report

Metric	Percentile Ranks	Value
Clashscore		2
Ramachandran outliers		0.9%
Sidechain outliers		4.1%

View In 3D: NGL or JSmol or PV (in Browser)

Standalone Viewers

Display a menu for "www.rcsb.org/pdb/explore/explore.do?structureid=2IMM"

[Contact Us](#)

The most straightforward access to a coordinate set in the PDB is to enter the four-character PDB ID into the search field. PDB IDs have a digit as the first character, then three (letter/digit)s. Thus the keyspace is $10 \times 36 \times 36 \times 36 = 466,560$. At the time the PDB was created with seven structures in 1971, each of which took years to complete in a herculean effort, no one could foresee that we would run out of IDs! Lesson learned.

PDB QUERY ... by keyword

rcsb.org

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligand

Advanced Search | Browse by Annotations | Search History (2) | Previous Results (1669)

PDB-101 PDB EMDatabank Nucleic Acid Database Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

1,869 Structures 132 Unreleased Structures 912 Citations 403 Ligands 21 News & PDB-101 Articles

Search Parameter:

Refinements Currently showing 1 - 25 of 1869 Page: 1 of 75

ORGANISM

- Mus musculus (957)
- Homo sapiens (915)
- Human immunodeficiency vi ... (167)
- Gallus gallus (70)
- Influenza A virus (66)
- Lama glama (60)
- Streptomyces lividans (44)
- Other (94)

UNIPROT MOLECULE NAME

- Ig kappa chain C region (120)
- Epilope albumin alpha-1 (100)

View: Detailed | Reports: Select a Report | Displaying 25 Results

Sort: Release Date: Newest to Oldest

5E08

Specific Recognition of a Single-stranded RNA Sequence by an Engineered Synthetic Antibody Fragment

[Shao, Y.](#), [Huang, H.](#), [Qin, D.](#), [Li, N.S.](#), [Koide, A.](#), [Staley, J.P.](#), [Koide, S.](#), [Kossiakoff, A.A.](#), [Piccirilli, J.A.](#)

(2016) J Mol Biol

Keyword searches are useful, but unreliable. Antibody? Immunoglobulin? Fab fragment? Bence-Jones protein?

PDB QUERY ... by keyword

rcsb.org

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures
PROTEIN DATA BANK

Search by PDB ID, author, macromolecule, sequence, or ligands Go

[Advanced Search](#) | [Browse by Annotations](#) | [Search History \(1\)](#) | [Previous Results \(123456\)](#)

Advanced Search Interface

Text Search ?

Search the full text of the mmCIF coordinate file

Text Result Count

1869
PDB Entries
(Structures)

OR

Text Search ?

Search the full text of the mmCIF coordinate file

Text Result Count

717
PDB Entries
(Structures)

[Add Search](#) [Contact Us](#)

[Display a menu](#)

There are 1,869 structures annotated with they keyword “antibody” and 717 structures annotated with “immunoglobulin”.

PDB QUERY ... by keyword

rcsb.org

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligand Go

Advanced Search | Browse by Annotations | Search History (1) | Previous Results (2399)

PDB-101 PDB EMDatabank Nucleic Acid Database Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

2,399 Structures 1214 Citations 461 Ligands

Search Parameter: Refine Search Save Search to MyPDB

Text Search for: antibody or Text Search for: immunoglobulin

Refinements Currently showing 1 - 25 of 2399 Page: 1 of 96 Previous Next

ORGANISM

- Homo sapiens (1224)
- Mus musculus (1034)
- Human immunodeficiency vi ... (167)
- Gallus gallus (74)
- Influenza A virus (68)
- Lama glama (61)
- Streptomyces lividans (44)
- Other (238)

UNIPROT MOLECULE NAME

View: Detailed **Reports:** Select a Report Displaying 25 Results

Sort: Release Date: Newest to Oldest Download

5K6U Download File View File

Sidekick-1 immunoglobulin domains 1-4, crystal form 1

Goodman, K.M., Yamagata, M., Jin, X., Mannepilli, S., Katsamba, P.S., Ahlsen, G., Sergeeva, A.P., Honig, B., Sanes, J.R., Shapiro, L.

Contact Us

Note that the number of results – 2,399 structures – is not the same as the sum of the two individual keyword searches. 1,869 “antibody” structures plus 717 “immunoglobulin” structures would have given 2,586 structures.

Keyword searches are good to find *some* structures, but they will virtually never find *all* structures of a desired type.

PDB QUERY ... by properties

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

Advanced Search Interface

Structure Description ?

Search by the description of the structure (PDB 'COMPND' record or mmCIF_entity.pdbx_description value)

Contains: Result Count

724
PDB Entries
(Structures)
284
Ligands

AND

Macromolecule Type ?

Search based on whether the structure contains chains of certain molecule types (e.g. protein vs. DNA)

Contains Protein Yes Result Count

Contains DNA Yes 3942

Contains RNA Ignore PDB Entries

Contains DNA/RNA Hybrid Ignore (Structures)

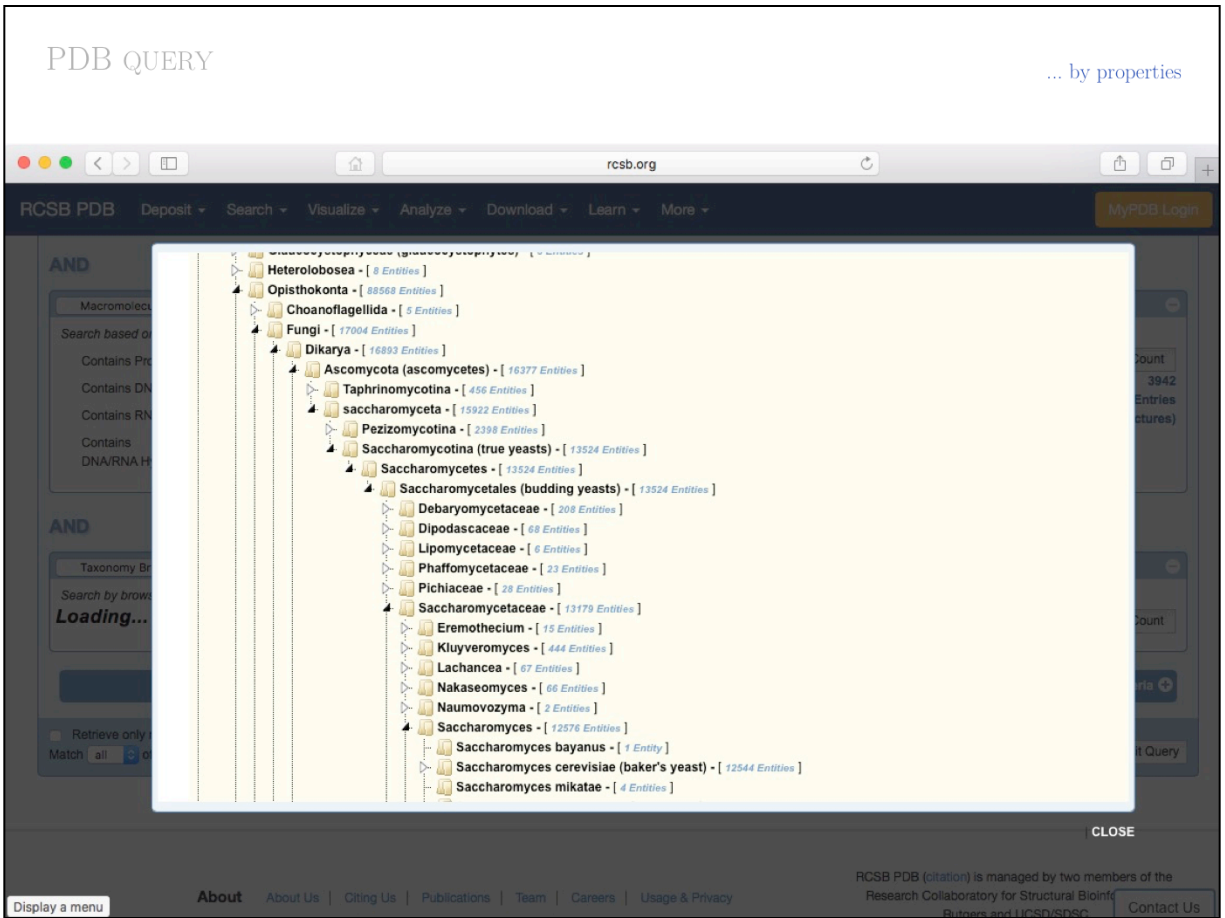
[Add Search Criteria](#)

Retrieve only representatives at 90% sequence identity ?

Match all of the above conditions. Results Structures Clear All Parameters Submit Query

Display a menu Contact Us

The Advanced Query interface is a flexible and powerful search tool for more sophisticated queries. Here a search for potential transcription factor/DNA complexes.



You can search for structures from specific species via the taxonomy browser.

PDB QUERY RESULTS ... as list

rcsb.org

RCSB PDB - Advanced Search RCSB PDB - Search Results

RCSB PDB Deposit Search Visualize Analyze Download Learn More [MyPDB Login](#)

Displaying 25 Results

ORGANISM

Saccharomyces cerevisiae (263)

Mus musculus (1)

Saccharomyces cerevisiae (1)

UNIPROT MOLECULE NAME

Meiosis-specific transcri ... (11)

DNA polymerase eta (9)

TATA-box-binding protein (9)

Transcription initiation ... (9)

Transcription initiation ... (7)

Intron-encoded endonuclea ... (7)

DNA repair protein RAD14 (6)

[Refine Query](#)

TAXONOMY

Eukaryota (264)

EXPERIMENTAL METHOD

X-ray (154)

Electron Microscopy (108)

Solution NMR (2)

X-RAY RESOLUTION

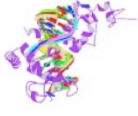
less than 1.5 Å (1)

1.5 - 2.0 Å (19)

2.0 - 2.5 Å (71)

View: Detailed **Reports:** Select a Report

Sort: Release Date: Newest to Oldest [Download](#)



5A39: Entity 4 containing Chain G, H [Download File](#) [View File](#)

Structure of Rad14 in complex with cisplatin containing DNA


[Koch, S.C., Kuper, J., Gasteliger, K.L., Simon, N., Strasser, R., Eisen, D., Geiger, S., Schneider, S., Kisker, C., Carell, T.](#)

(2015) Proc Natl Acad Sci U S A 112 8272

Released: 6/24/2015 **Macromolecule:** DNA REPAIR PROTEIN RAD14 (protein)

Method: X-ray Diffraction **Resolution:** 2.8 Å **Unique Ligands:** CPT, ZN

Residue Count: 312



4YIR: Entity 1 containing Chain A [Download File](#) [View File](#)

Crystal structure of Rad4-Rad23 crosslinked to an undamaged DNA

[Chen, X., Velmurugu, Y., Zheng, G., Park, B., Shim, Y., Kim, Y., Liu, L., Van Houten, B., He, C., Ansari, A., Min, J.H.](#)

(2015) Nat Commun 6 5849

Display a menu [Contact Us](#)

Output can be in list format ...

PDB QUERY RESULTS ... as report table

RCSB PDB - Advanced Search RCSB PDB - Grid Report

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

Custom Report Total of 104 results.

Click on column headers to sort up/down. Click again to reverse order. Download options: [EXCEL](#) | [EXCEL 2007 or later](#) | [CSV](#)

Page 1 of 6 View 1 - 20 of 104

	PDB ID	Structure Title	Resolution	R All	R Free
	<input type="text"/> x	<input type="text"/> x	<input type="text"/> x	<input type="text"/> x	<input type="text"/> x
1	1A0A	PHOSPHATE SYSTEM POSITIVE REGULATORY PROTEIN PHO4/DNA COMPLEX	2.80		0.284
2	1AKH	MAT A1/ALPHA2/DNA TERNARY COMPLEX	2.50		0.302
3	1APL	CRYSTAL STRUCTURE OF A MAT-ALPHA2 HOMEODOMAIN-OPERATOR COMPLEX SUGGESTS A GENERAL MODEL FOR HOMEODOMAIN-DNA INTERACTIONS	2.70		
4	1D66	DNA RECOGNITION BY GAL4: STRUCTURE OF A PROTEIN/DNA COMPLEX	2.70		
5	1DGC	THE X-RAY STRUCTURE OF THE GCN4-BZIP BOUND TO ATF/CREB	3.00		

Display a menu Contact Us

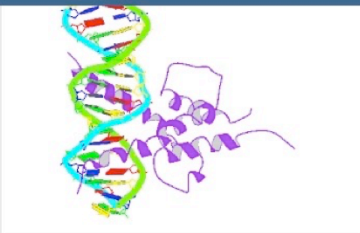

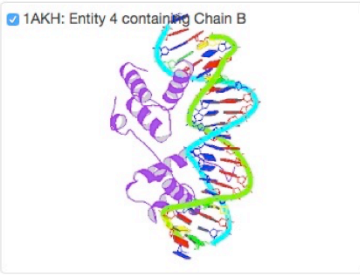



... customizable tables ...

PDB QUERY RESULTS ... as collage of images

rcsb.org

RCSB PDB - Advanced Search RCSB PDB - Search Results

RCSB PDB Deposit Search Visualize Analyze Download Learn More [MyPDB Login](#)

<p>UNIPROT MOLECULE NAME</p> <p>Meiosis-specific transcri ... (11) DNA polymerase eta (9) TATA-box-binding protein (9) Transcription initiation ... (9) Transcription initiation ... (7) Intron-encoded endonuclea ... (7) DNA repair protein RAD14 (6) Refine Query</p> <p>TAXONOMY</p> <p>Eukaryota (264)</p> <p>EXPERIMENTAL METHOD</p> <p>X-ray (154) Electron Microscopy (108) Solution NMR (2)</p> <p>X-RAY RESOLUTION</p> <p>less than 1.5 Å (1) 1.5 - 2.0 Å (19) 2.0 - 2.5 Å (31) 2.5 - 3.0 Å (53) 3.0 and more Å (50) Refine Query</p> <p>RELEASE DATE</p>		
	<input checked="" type="checkbox"/> 1AKH: Entity 4 containing Chain B 	<input checked="" type="checkbox"/> 1APL: Entity 3 containing Chain C, D 
	<input checked="" type="checkbox"/> 1D66: Entity 3 containing Chain A, B 	<input checked="" type="checkbox"/> 1DGC: Entity 2 containing Chain A 

Display a menu for www.rcsb.org/pdb/explore/explore.do?structureid=1AKH [Contact Us](#)

... or a gallery overview that allows you to quickly scan for the type of structure you want to consider.

LIMITATIONS

In principle, structures can be obtained at **atomic resolution**.

This means we can identify the location of individual water molecules!

In practice, structures are **time-averaged and population averaged**.

We see atomic resolution only for well-ordered atoms.

Experimental artefacts are possible, in particular crystal packing artefacts.

Experimental errors are possible, in particular when modeling disordered structure (B-factors!); *fraud is uncommon but has occurred at times.*

All (refined) structures combine experimental information with idealized stereochemistry.

Fortunately, experience shows that structures obtained with different experimental methods are very similar.

Quality metrics exist, but are not trivial to interpret.

Biological relevance has to be taken into account.

BIOLOGICAL ASSEMBLY

rcsb.org

RCSB PDB - Advanced Search

1QPI: CRYSTAL STRUCTURE OF TETRACYCLINE REPRESSOR/OPERATOR COMPLEX Structure Sum...

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures

Search by PDB

Advanced Search

PDB-101 PDB-EMBL EMDatabank

Structure Summary 3D View Annotations Sequence Sequence Similarity Literature

Asymmetric Unit

1QPI

CRYSTAL STRUCTURE OF TETRACYCLINE REPRESSOR/OPERATOR COMPLEX

DOI: 10.2210/pdb

Classification: TR

Deposited: 1999-0

Deposition autho

Organism: *Esche*

Expression Syste

Structural Biolog

Experimental Dat

Method: X-RAY D

Resolution: 2.5 Å

REMARK 290 CRYSTALLOGRAPHIC SYMMETRY

REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: C 2 2 2 1

REMARK 290

REMARK 290 SYMOP SYMMETRY

REMARK 290 NNNMM OPERATOR

REMARK 290 1555 X, Y, Z

REMARK 290 2555 -X, -Y, 1/2+Z

REMARK 290 3555 -X, Y, 1/2-Z

REMARK 290 4555 X, -Y, -Z

REMARK 290 5555 1/2+X, 1/2+Y, Z

REMARK 290 6555 1/2-X, 1/2-Y, 1/2+Z

REMARK 290 7555 1/2-X, 1/2+Y, 1/2-Z

REMARK 290 8555 1/2+X, 1/2-Y, -Z

Display Files Download Files

View in 3D: NGL or JSmol or PV (in Browser)

Contact Us

The crystallographic asymmetric unit does not necessarily contain a functional molecule

In the example above, only one chain of the tet-repressor dimer is seen bound to only one strand of B-DNA. The second chain and strand can be generated through a symmetry operation (180° rotation and translation), thus it contains the same coordinate information and does not need to be separately stored.

However, in order to study the functional molecule, the redundant coordinates have to be combined to a homodimer. The term **biological unit** describes a coordinate set that (presumably) depicts a homooligomer in its *functional* state.

All molecules in the crystal lattice can be generated from the crystallographic symmetry operations specified in a PDB file, for the space group of the crystal. But it may not be obvious which of the symmetry replicates might actually be involved in a physiological interaction and which ones have only been induced by the crystallization process. In the tet-repressor example, the crystallographic space group has eight(!) symmetry related monomers in the unit cell of the crystal lattice.

BIOLOGICAL ASSEMBLY

RCSB PDB - Advanced Search

1QPI: CRYSTAL STRUCTURE OF TETRACYCLINE REPRESSOR/OPERATOR COMPLEX Structure Sum...

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB Login

RCSB PDB An Information Portal to 123456 Biological Macromolecular Structures

Search by PDB Advanced Search

PDB-101 PDB-EMBL EMDatabank Nucleic Acid Database Structural Biology Knowledgebase Worldwide Protein Data Bank Foundation

Structure Summary 3D View Annotations Sequence Sequence Similarity Literature

Biological Assembly 1

1QPI

CRYSTAL STRUCTURE OF TETRACYCLINE REPRESSOR/OPERATOR COMPLEX

DOI: 10.2210/pdb

Classification: TR

Deposited: 1999-0

Deposition autho

Organism: *Esche*

Expression Syste

Structural Biolog

Experimental Dat

Method: X-RAY D

Resolution: 2.5 Å

REMARK 290 CRYSTALLOGRAPHIC SYMMETRY

REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: C 2 2 2 1

REMARK 290

REMARK 290 SYMOP SYMMETRY

REMARK 290 NNNMMM OPERATOR

REMARK 290 1555 X, Y, Z

REMARK 290 2555 -X, -Y, 1/2+Z

REMARK 290 3555 -X, Y, 1/2-Z

REMARK 290 4555 X, -Y, -Z

REMARK 290 5555 1/2+X, 1/2+Y, Z

REMARK 290 6555 1/2-X, 1/2-Y, 1/2+Z

REMARK 290 7555 1/2-X, 1/2+Y, 1/2-Z

REMARK 290 8555 1/2+X, 1/2-Y, -Z

Display Files Download Files

View in 3D: NGL or JSmol or PV (in Browser)

Contact Us

The crystallographic asymmetric unit does not necessarily contain a functional molecule

The "Biological Assembly" complements the single chain to a complete repressor homodimer/double-stranded DNA complex.

STRUCTURE AT THE NCBI

https://www.ncbi.nlm.nih.gov/Structure/MMDB/docs/mmdb_how_to.html

NCBI Structure

Structure Group ▾ 3D Macromolecular Structures ▾ Conserved Domains ▾ PubChem ▾ BioSystems ▾

3D Macromolecular Structures

OVERVIEW SEARCH HOW TO HELP NEWS FTP PUBLICATIONS DISCOVER

How to use the Molecular Modeling Database (MMDB) and related resources: examples

This page provides quick start guides for some common types of searches.
The MMDB Help document provides additional search tips. Once records of interest are retrieved, follow Entrez's "Links" to discover associations among previously disparate data.

Retrieve specific subsets of resolved structures

- Find structures for a **gene/protein product of interest**.
- Find 3D structures bound to a **specific chemical** (e.g., aspirin).
- Find 3D structures bound to a **specific small biopolymer** (e.g., a peptide).
- Retrieve 3D structures for a **specific type of molecule**, such as protein, RNA, DNA, protein+chemical, etc.
- Retrieve all structures from a **particular source organism or taxon**.

Find structural templates for proteins

- Align a query protein to a similar sequence from a 3D structure and interactively view sequence/structure relationships.

Find structures that are similar in 3D shape (using VAST)

- Find structures with **similarly shaped proteins or 3D domains**, regardless of their degree of sequence similarity.
- Find structures with **similarly shaped biological assemblies**, regardless of their degree of sequence similarity. [VAST+]
- Compare a **newly resolved 3D structure in PDB format** against the 3D coordinates of known structures already in the public domain. [VAST Search]

View 3D structures

- Interactively view a 3D structure and its corresponding sequence data to **examine sequence-structure relationships**.
- Identify putative **active site residues**.

MMDB - very well integrated with Entrez databases but not with external databases ...
... and no support for stereo-viewing!

Display a menu

Revised 23 September 2016

Structure services exist at the NCBI and the EBI with distinct services. Explore.

NDB

ndbserver.rutgers.edu

About NDB Standards Education To

ndb NUCLEIC ACID DATABASE

A Portal for Three-dimensional Structural Information
As of 12-Oct-2016 num

Search DNA Search RNA Advanced Search

Welcome to the NDB

The NDB contains information about experimentally-determined nucleic acids and complex assemblies. Use the NDB to perform searches based on annotations relating to sequence, structure and function, and learn about nucleic acids.

Search Structures

Search DNA
Search DNA and its complexes

Search RNA
Search for RNA structures in the NDB archive or in the Non-Redundant list

Advanced Search
Search for structures based on structural features, chemical features, binding modes, citation and experimental information

Featured Tools

RNA 3D Motif Atlas, a representative collection of RNA 3D internal and hairpin motifs

Non-redundant Lists of RNA-containing 3D structures

RNA Base Triple Atlas, a collection of motifs consisting of two RNA basepairs

WebFR3D, a webserver for symbolic and geometric searching of RNA 3D structures

R3D Align, an application for detailed nucleotide to nucleotide alignments of RNA 3D structures

urx035.pdb
(Hammerhead Ribozyme)

Inconveniently, neither the PDB nor the NDB contain a complete set of public RNA structure coordinate files.

The Nucleic Acid Structure Database

<http://ndbserver.rutgers.edu/>

PDBsum

EMBL-EBI Services Research Training About us

PDBsum Pictorial database of 3D structures in the Protein Data Bank

Databases > Structure Databases > PDBsum

PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB). It shows the molecule(s) that make up the structure (ie protein chains, DNA, ligands and metal ions) and schematic diagrams of the interactions between them. [Read more...](#)

PDB code (4 chars) **Find** Example: "1klv"

Text search **Search**
Scans all TITLE, HEADER, COMPND, SOURCE and AUTHOR records in the PDB (eg to find a given protein by name).

Search by sequence

Search
Perform FASTA search vs all sequences in the PDB to get a list of the closest matches.

Search by

UniProt id: <input type="text"/> <small>(eg P03023, LACI, ECOLI, etc)</small>	Pfam id: <input type="text"/> <small>(eg PF07992)</small>	Ensembl id: <input type="text"/> <small>(eg ENSG00000086205, ENST00000256999)</small>
Search	Search	Search

Notes

Display a menu

Contents
PDBsum contains 126,889 entries, including 2,433 superseded
Last update: 15 October, 2016

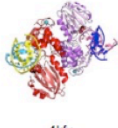
In-house version
PDBsum proprietary
In-house version for companies to process own structures (see below left)

Related databases

EC-PDB
Enzyme 3D structures organized by the E.C. numbering hierarchy.

DrugPort
Structures of drugs and their target proteins in the PDB.

SAS
Searches sequence against all PDB sequences



PDBsum is a secondary database that stores analysis and interpretation information for PDB coordinate sets.

<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>

PDBSUM

ebi.ac.uk


Go to PDB code: 2imm go

Top page Protein Ligands Clefts Links

Protein chain A PDB id 2imm

Chain A (114 residues)

CATH structural classification (1 domain):
 Domain Links CATH no. Class Architecture
 1 CATH d2 2.60.40.10 = **Mainly Beta Sandwich**



Protein chain A highlighted (click to view)

Jmol **Strap**

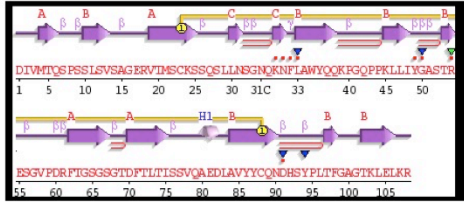
Motifs

Secondary structure

- Wiring diagram
- Residue conservation

ProMotif

- 3 sheets
- 5 beta hairpins
- 3 beta bulges
- 13 strands
- 1 helix
- 17 beta turns
- 2 gamma turns
- 1 disulphide



FASTA file

Domain 1

Key:

Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...

Helix Strand

Motifs: β beta turn γ gamma turn β beta hairpin

Disulphides: S-S disulphide bond

Residue contacts: \bullet to ligand

PDB SITE records: ∇ AC1 ∇ AC2

Analysis of sequence's residue conservation

Related protein sequences in the PDB

Display a menu

Schematics like this “wiring diagram” help to get an overview of protein topology and sequence organization. I find the high-level descriptions in PDBsum very useful to begin the study of a particular protein.

<http://steipe.biochemistry.utoronto.ca/abc>

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA