

Protein Data Bank

Wintersemester 2011/12

Peter Güntert

RCSB PDB PROTEIN DATA BANK **PDB-101** A MEMBER OF THE PDBs

As of Tuesday Jan 31, 2012 at 4 PM PST there are 78992 Structures **PDB Statistics**

Search | All Categories:

Biological Macromolecular Resource

Full Description

Featured Molecules

Structural View of Biology

Molecule of the Month
Messenger RNA capping
 In our cells, transcription is not just a simple process of reading DNA and building a complimentary RNA strand. Almost immediately after RNA polymerase begins, the cell is making changes. When the mRNA is only about 30 nucleotides long, the cell makes its first change: it connects a guanosine nucleotide to the end.
Full Article

Protein Structure Initiative Featured System
Disordered Proteins
 Looking through the thousands of structures in the PDB, we get the impression that proteins must have a stable, folded structure to be functional. Recently, however, it has become clear that many proteins use disorder when performing their jobs.
Full Article | Archive | PSI Structural Biology Knowledgebase

Explore Archive

| | | |
|------------------------------|----------------------------|---|
| Organism | Taxonomy | Organism |
| Exp. Method | X-Ray Resolution | <ul style="list-style-type: none"> Homo sapiens (19280) Escherichia coli (4223) Mus musculus (2423) Saccharomyces cerevisiae (2117) Bos taurus (1993) Rattus norvegicus (1681) Escherichia coli K-12 (1278) Other (42166) |
| Release Date | Polymer Type | <input type="button" value="Show all"/> |
| Enzyme Classification | SCOP Classification | |

New user? Try the browser [Compatibility Check](#) and information on [Getting Started](#), or the [Narrated Tutorial](#).


The RCSB PDB is managed by two members of the RCSB: Rutgers and UCSD, and is funded by NSF, NIGMS, DOE, NLM, NCI, NINDS, and NIDDK.

www.pdb.org

RCSB PDB News

Weekly | Quarterly | Yearly

2012-01-31
Announcement: Meeting on Hybrid Methods



A symposium on the Structural Analysis of Supramolecular Assemblies by Hybrid Methods will be held March 14-18, 2012 in Lake Tahoe, CA. [more](#)

Free Webinar on How to use the RCSB PDB
 Join OpenHelix on Wed Feb 15 to learn how to search, visualize, and explore structures using the RCSB PDB. [more](#)

- Newsletter Published
- NI Science Olympiad Protein Modeling Results
- Web Services for Accessing PDB Data

PDB Current Holdings Breakdown

| Exp.Method | Proteins | Nucleic Acids | Protein/NA Complexes | Other | Total |
|---------------------|----------|---------------|----------------------|-------|-------|
| X-RAY | 64591 | 1337 | 3187 | 2 | 69117 |
| NMR | 8108 | 966 | 186 | 7 | 9267 |
| ELECTRON MICROSCOPY | 277 | 22 | 101 | 0 | 400 |
| HYBRID | 42 | 3 | 2 | 1 | 48 |
| other | 138 | 4 | 5 | 13 | 160 |
| Total | 73156 | 2332 | 3481 | 23 | 78992 |

(Click on any number to retrieve the results from that category.)

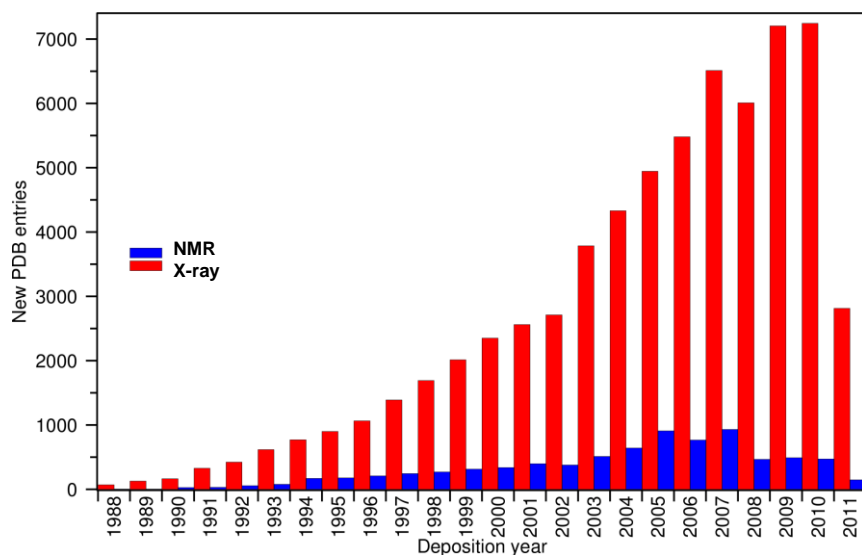
58512 structures in the PDB have a structure factor file.

6572 structures in the PDB have an NMR restraint file.

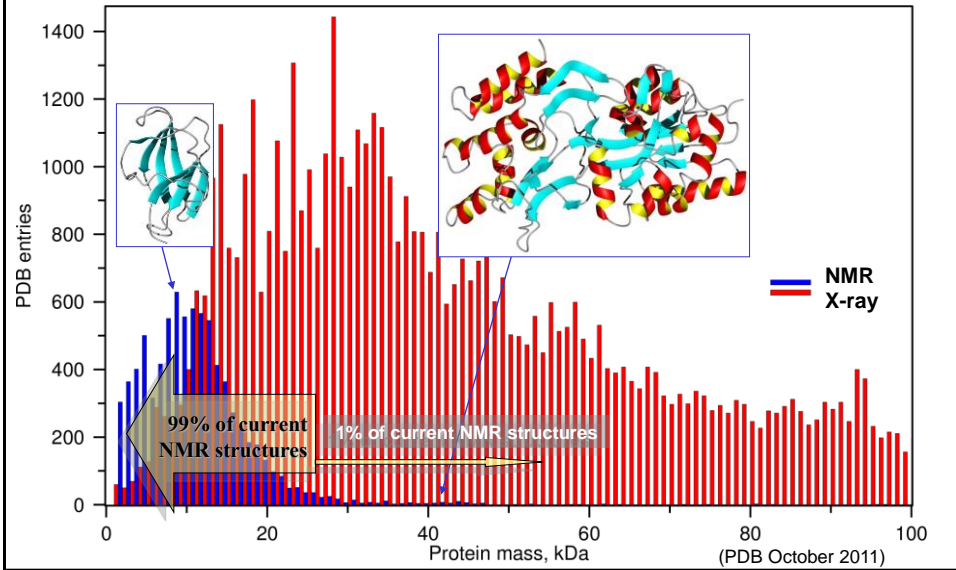
343 structures in the PDB have a chemical shifts file.

31.01.2012

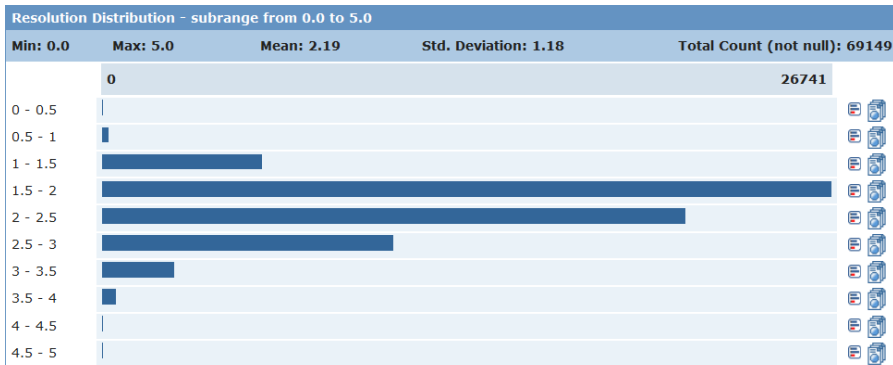
Number of released X-ray and NMR structures in the PDB (October 2011)



Protein structures in the PDB

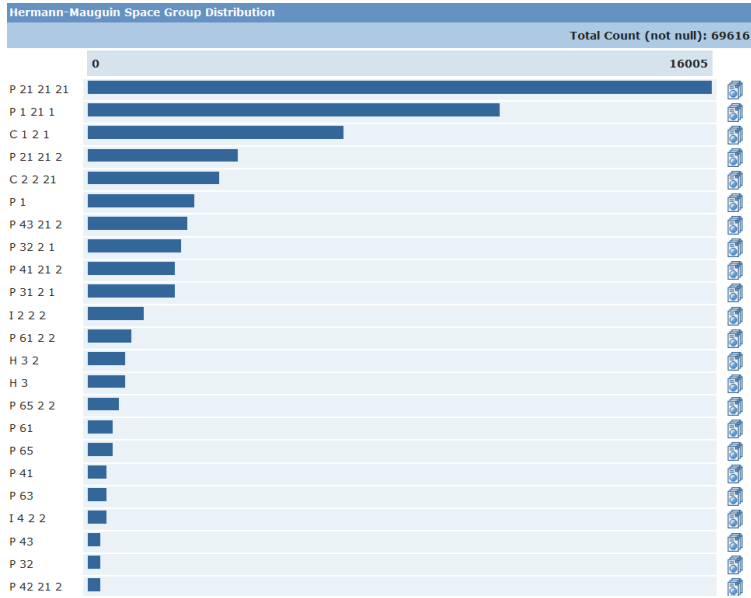


PDB: Resolution distribution

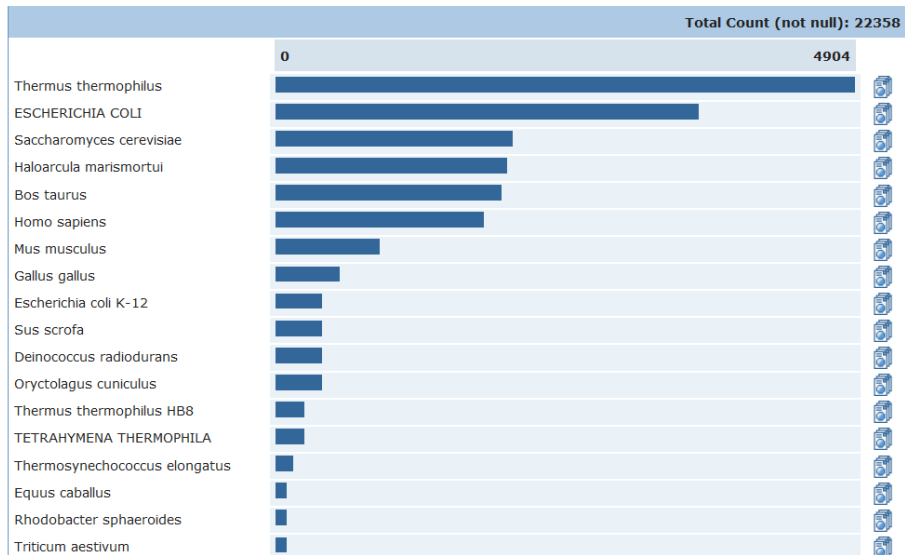


There are 1551 values outside the current range - for a complete histogram [click here](#).

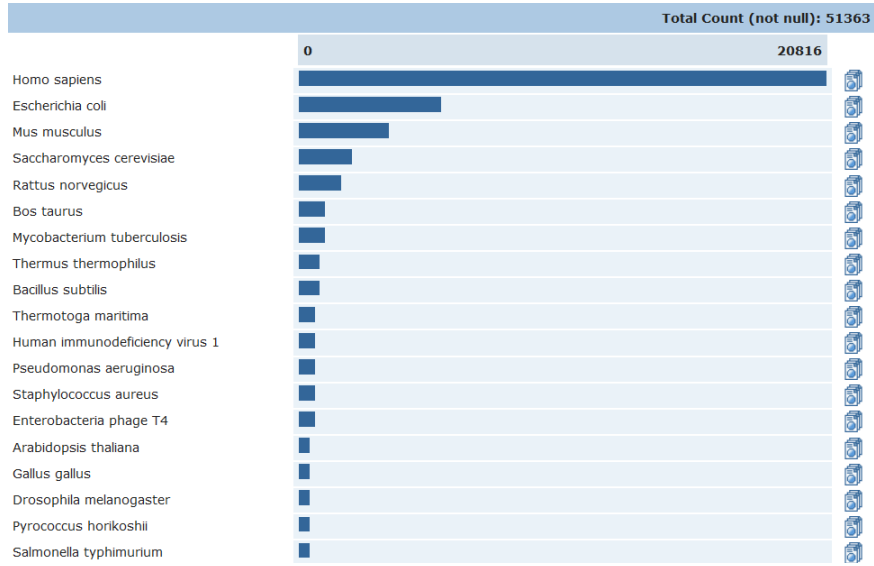
PDB: Distribution by space group



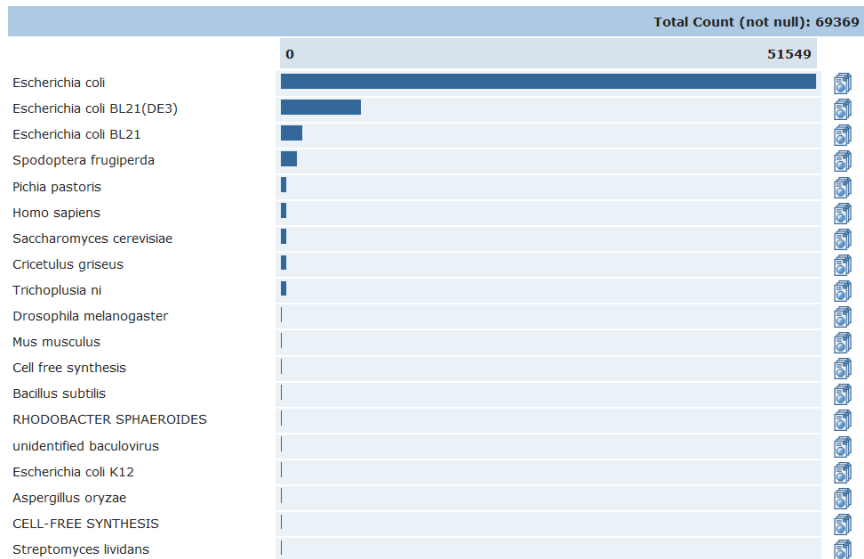
Distribution by natural source organism



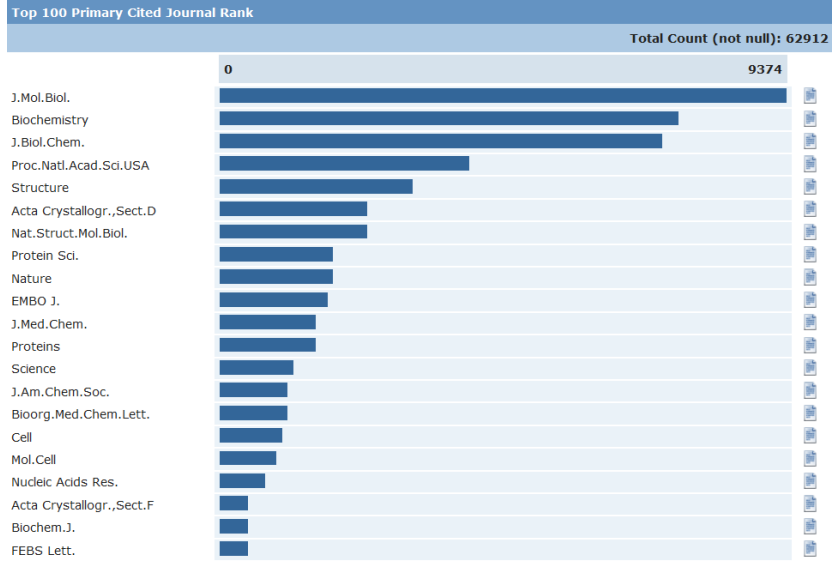
Distribution by gene source organism



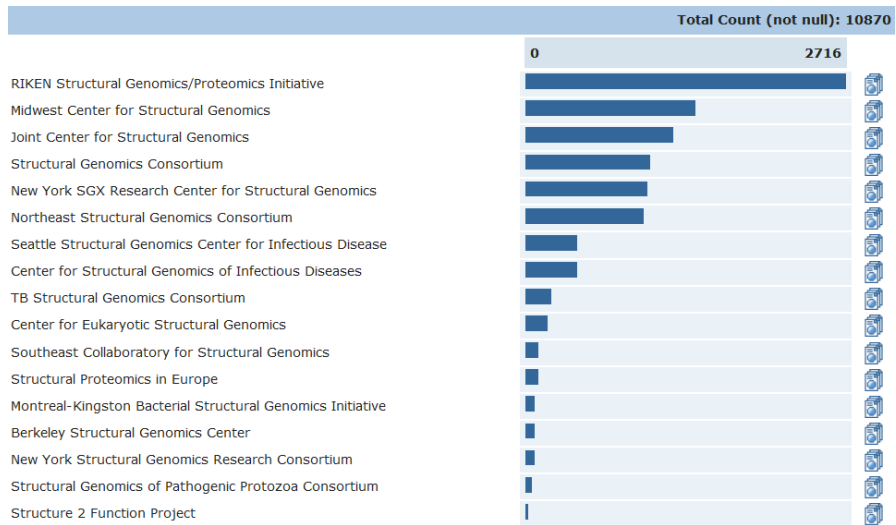
Distribution by expression organism



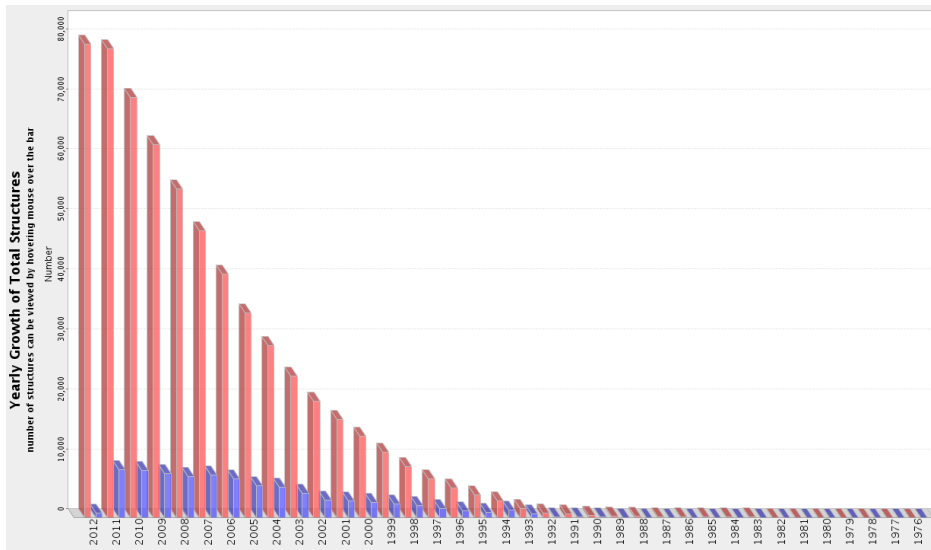
Distribution by primary cited journal



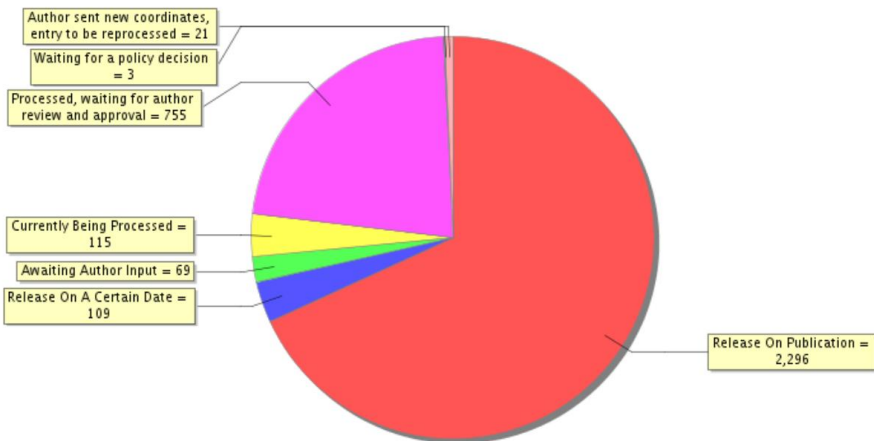
Distribution by structural genomics center



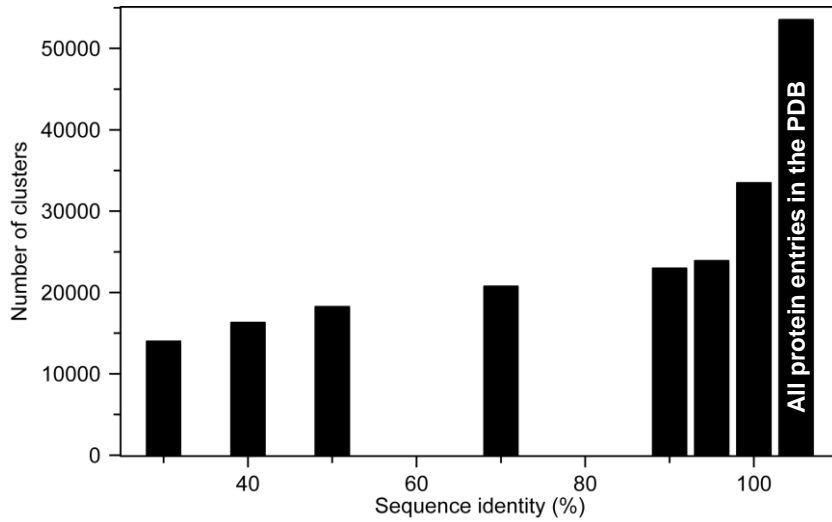
PDB: Total structures



PDB: Status of unreleased entries



PDB redundancy



PDB entry 1U2P

Summary | Derived Data | Sequence | Seq. Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Crystal structure of Mycobacterium tuberculosis Low Molecular Protein Tyrosine Phosphatase (MPTpA) at 1.9A resolution

DOI:10.2210/pdb1u2p/pdb

1u2p [Display Files](#) [Download Files](#) [Print this Page](#)

Biological Molecule 9

Primary Citation

Crystal Structure of Low-Molecular-Weight Protein Tyrosine Phosphatase from Mycobacterium tuberculosis at 1.9-A Resolution
 Madhurantakam, C., Rajakumara, E., Mazumdar, P.A., Saha, B., Mitra, D., Wiker, H.G., Sankaranarayanan, R., Das, A.K. (2005) J.Bacteriol. 187: 2173-2181
 PubMed: 15743966 [Search Related Articles in PubMed](#)

PubMed Abstract: The low-molecular-weight protein tyrosine phosphatase (LMWPTpase) belongs to a distinctive class of phosphotyrosine phosphatases widely distributed among prokaryotes and eukaryotes. We report here the crystal structure of LMWPTpase of microbial origin, the first of its kind from Mycobacterium tuberculosis. The ... [Read More & Search PubMed Abstracts](#)

Molecular Description

Classification: **Hydrolase**
 Structure Weight: 17954.65
 Molecule: low molecular weight protein-tyrosine-phosphatase
 Polymer: 1 Type: polypeptide(L) Length: 163
 Chains: A
 EC#: 3.1.3.48

Source

Polymer: 1 Scientific Name: **Mycobacterium tuberculosis** Expression System: **Escherichia coli**

Related PDB Entries

| ID | Details |
|------|--|
| 1U2Q | Crystal structure of Mycobacterium tuberculosis low molecular weight protein tyrosine phosphatase (MPTpA) at 2.5 A resolution with glycerol in the active site |

Ligand Chemical Component

| Identifier | Name | Formula | Links |
|------------|--------------|---------|---|
| CL | CHLORIDE ION | Cl | B S I |

Derived Data

- CATH Classification v3.2.0 - (1 Domains)
- PFAM Classification - (1 Domains)
- GO Terms - (4 Terms)

Deposition Summary

Authors: Madhurantakam, C., Rajakumara, E., Mazumdar, P.A., Saha, B., Mitra, D., Wiker, H.G., Sankaranarayanan, R., Das, A.K.
 Deposition: 2004-07-20
 Release: 2005-03-22
 Last Modified (REVDAT): 2009-02-24

Experimental Details

Method: X-RAY DIFFRACTION
 Experimental Data: [1](#) [1](#) [EDS](#) [2](#)
 Resolution(Å): 1.90
 R-Value: 0.202 (obs.)
 R-Free: 0.227
 Space Group: **P 2₁ 2₁ 2₁**
 Unit Cell:
 Length(Å) Angles(°)
 a = 40.82 a = 90.00
 b = 53.61 b = 90.00
 c = 68.49 c = 90.00

3-D Viewers

Jmol SimpleViewer Protein Workshop Other Viewers

Oligomeric State: MONOMERIC

Summary **Derived Data** Sequence Seq. Similarity Literature Biol. & Chem. Methods Geometry Links

Derived Data **1u2p** [Display Files](#) [Download Files](#) [Print this Page](#)

Derived Data: CATH Classification (version v3.2.0) [Hide](#)

| Domain | Class | Architecture | Topology | Homology |
|---------|------------|-----------------------|---------------|----------|
| 1u2pA00 | Alpha Beta | 3-Layer(aba) Sandwich | Rossmann fold | |

Derived Data: PFAM Classification [Hide](#)

| Chain | PFAM Accession | PFAM ID | Description | Type | Clan ID |
|-------|-------------------------|---------|--|--------|---------|
| A | PF01451 | LMWpC | Low molecular weight phosphotyrosine protein phosphatase | Domain | |

Derived Data: GO Terms [Hide](#)

| Polymer | Molecular Function | Biological Process | Cellular Component |
|--|--|--|--|
| low molecular weight protein-tyrosine-phosphatase (1U2P:A) | <ul style="list-style-type: none"> phosphoprotein phosphatase activity protein tyrosine phosphatase activity hydrolase activity | <ul style="list-style-type: none"> protein amino acid dephosphorylation | <ul style="list-style-type: none"> none |

Summary **Derived Data** **Sequence / Structure Details** Sequence Seq. Similarity Literature Biol. & Chem. Methods Geometry Links

1u2p [Display Files](#) [Download Files](#) [Print this Page](#)

Redundancy Reduction and Sequence Clustering
View the clustering results for 1U2P.

Sequence Display
The structure **1U2P** has in total **3** chains. Out of these **3** are sequence-unique. Currently viewing **unique chains** only. [[show all chains](#)] [[show 3D in 3mol](#)]

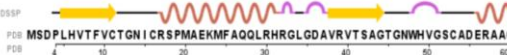
Chain Display


Chain A [polymer 1] [[help](#)] [[fasta](#)] [[text/markup](#)]

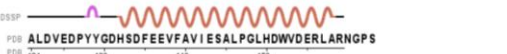
Description: low molecular weight protein-tyrosine-phosphatase
 Chain Type: polypeptide(L)
 UniProt reference: [P65716](#)
 Length: 163 residues
 CATH domain assignment: [1u2pA00](#) : 156 residues [?](#)
 [hide] [reference]
 DSSP secondary structure: 43% helical (7 helices; 71 residues)
 [hide] [reference]
 13% beta sheet (5 strands; 22 residues)
 More annotations:
 Select:

Currently displayed: SEQRES sequence. [[display external \(UniProt/PIR\) sequence](#)]

Sequence Details

DSSP 
 PDB **MSDPLHVTFCYCTGNICRSPMAEKMFAOQLRHRGLGDAVRVTSAGTGNHHVGVSCADERAAG**
 PDB 4 10 20 30 40 50 60

DSSP 
 PDB **VLR AHGYPTDHRAAQVGT EHLAADLLVALDRNHARLLRQLGVEAARVMLRSFDP RSGTH**
 PDB 61 70 80 90 100 110 120

DSSP 
 PDB **ALDVEDPYGDHSDFEVFAVIESALPGLHDWDERLARNGPS**
 PDB 121 130 140 150

PDB entry 1U2P: Header, source

```

HEADER      HYDROLASE                               20-JUL-04  1U2P
TITLE       CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS LOW
TITLE       2 MOLECULAR PROTEIN TYROSINE PHOSPHATASE (MPTPA) AT 1.9A
TITLE       3 RESOLUTION
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: LOW MOLECULAR WEIGHT PROTEIN-TYROSINE-
COMPND      3 PHOSPHATASE;
COMPND      4 CHAIN: A;
COMPND      5 SYNONYM: PTPASE;
COMPND      6 EC: 3.1.3.48;
COMPND      7 ENGINEERED: YES
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: MYCOBACTERIUM TUBERCULOSIS;
SOURCE      3 ORGANISM_TAXID: 1773;
SOURCE      4 GENE: MPTPA;
SOURCE      5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE      6 EXPRESSION_SYSTEM_TAXID: 562;
SOURCE      7 EXPRESSION_SYSTEM_STRAIN: SG13009;
SOURCE      8 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE      9 EXPRESSION_SYSTEM_PLASMID: PQE30
KEYWDS      HYDROLASE, TYROSINE PHOSPHATASE, MYCOBACTERIUM

```

PDB entry 1U2P: Authors

```

EXPDTA      X-RAY DIFFRACTION
AUTHOR      C. MADHURANTAKAM, E. RAJAKUMARA, P. A. MAZUMDAR, B. SAHA, D. MITRA,
AUTHOR      2 H. G. WIKER, R. SANKARANARAYANAN, A. K. DAS
REVDAT      2 24-FEB-09 1U2P 1 VERSN
REVDAT      1 22-MAR-05 1U2P 0
JRNL        AUTH C. MADHURANTAKAM, E. RAJAKUMARA, P. A. MAZUMDAR, B. SAHA,
JRNL        AUTH 2 D. MITRA, H. G. WIKER, R. SANKARANARAYANAN, A. K. DAS
JRNL        TITL CRYSTAL STRUCTURE OF LOW-MOLECULAR-WEIGHT PROTEIN
JRNL        TITL 2 TYROSINE PHOSPHATASE FROM MYCOBACTERIUM
JRNL        TITL 3 TUBERCULOSIS AT 1.9-A RESOLUTION
JRNL        REF J. BACTERIOL. V. 187 2175 2005
JRNL        REFN ISSN 0021-9193
JRNL        PMID 15743966
JRNL        DOI 10.1128/JB.187.6.2175-2181.2005
REMARK      1
REMARK      2
REMARK      2 RESOLUTION. 1.90 ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT.
REMARK      3 PROGRAM : CNS 1.1
REMARK      3 AUTHORS : BRUNGER, ADAMS, CLORE, DELANO, GROS, GROSSE-
REMARK      3 : KUNSTLEVE, JIANG, KUSZEWSKI, NILGES, PANNU,
REMARK      3 : READ, RICE, SIMONSON, WARREN

```

PDB entry 1U2P: Refinement

```

REMARK 2 RESOLUTION.      1.90  ANGSTROMS.
REMARK 3
REMARK 3 REFINEMENT.
REMARK 3   PROGRAM       : CNS 1.1
REMARK 3   AUTHORS      : BRUNGER, ADAMS, CLORE, DELANO, GROS, GROSSE-
REMARK 3                   : KUNSTLEVE, JIANG, KUSZEWSKI, NILGES, PANNU,
REMARK 3                   : READ, RICE, SIMONSON, WARREN
REMARK 3
REMARK 3 REFINEMENT TARGET : ENGH & HUBER
REMARK 3
REMARK 3 DATA USED IN REFINEMENT.
REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 1.90
REMARK 3 RESOLUTION RANGE LOW  (ANGSTROMS) : 24.96
REMARK 3 DATA CUTOFF              (SIGMA(F)) : 0.000
REMARK 3 DATA CUTOFF HIGH          (ABS(F))  : 1161871.740
REMARK 3 DATA CUTOFF LOW          (ABS(F))  : 0.0000
REMARK 3 COMPLETENESS (WORKING+TEST) (%)   : 99.6
REMARK 3 NUMBER OF REFLECTIONS       : 12309
REMARK 3
REMARK 3 FIT TO DATA USED IN REFINEMENT.
REMARK 3 CROSS-VALIDATION METHOD          : THROUGHOUT
REMARK 3 FREE R VALUE TEST SET SELECTION : RANDOM
REMARK 3 R VALUE                          (WORKING SET) : 0.202
REMARK 3 FREE R VALUE                      : 0.227
REMARK 3 FREE R VALUE TEST SET SIZE      (%) : 5.000
REMARK 3 FREE R VALUE TEST SET COUNT     : 616
REMARK 3 ESTIMATED ERROR OF FREE R VALUE : 0.009

```

PDB entry 1U2P: Missing residues

```

REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465  M RES C  SSSEQI
REMARK 465    MET A    1
REMARK 465    SER A    2
REMARK 465    ASP A    3
REMARK 465    ASN A   160
REMARK 465    GLY A   161
REMARK 465    PRO A   162
REMARK 465    SER A   163

```

PDB entry 1U2P: Ramachandran plot outliers

```

REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: TORSION ANGLES
REMARK 500
REMARK 500 TORSION ANGLES OUTSIDE THE EXPECTED RAMACHANDRAN REGIONS:
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE) .
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,4X,F7.2,3X,F7.2)
REMARK 500
REMARK 500 EXPECTED VALUES: GJ KLEYWEGT AND TA JONES (1996) . PHI/PSI-
REMARK 500 CHOLOGY: RAMACHANDRAN REVISITED. STRUCTURE 4, 1395 - 1400
REMARK 500
REMARK 500 M RES CSSEQI      PSI      PHI
REMARK 500   CYS A 16      -83.04   -122.74

```

PDB entry 1U2P: Sequence

```

DBREF 1U2P A 1 163 UNP P65716 PTPA_MYCTU 1 163
SEQRES 1 A 163 MET SER ASP PRO LEU HIS VAL THR PHE VAL CYS THR GLY
SEQRES 2 A 163 ASN ILE CYS ARG SER PRO MET ALA GLU LYS MET PHE ALA
SEQRES 3 A 163 GLN GLN LEU ARG HIS ARG GLY LEU GLY ASP ALA VAL ARG
SEQRES 4 A 163 VAL THR SER ALA GLY THR GLY ASN TRP HIS VAL GLY SER
SEQRES 5 A 163 CYS ALA ASP GLU ARG ALA ALA GLY VAL LEU ARG ALA HIS
SEQRES 6 A 163 GLY TYR PRO THR ASP HIS ARG ALA ALA GLN VAL GLY THR
SEQRES 7 A 163 GLU HIS LEU ALA ALA ASP LEU LEU VAL ALA LEU ASP ARG
SEQRES 8 A 163 ASN HIS ALA ARG LEU LEU ARG GLN LEU GLY VAL GLU ALA
SEQRES 9 A 163 ALA ARG VAL ARG MET LEU ARG SER PHE ASP PRO ARG SER
SEQRES 10 A 163 GLY THR HIS ALA LEU ASP VAL GLU ASP PRO TYR TYR GLY
SEQRES 11 A 163 ASP HIS SER ASP PHE GLU GLU VAL PHE ALA VAL ILE GLU
SEQRES 12 A 163 SER ALA LEU PRO GLY LEU HIS ASP TRP VAL ASP GLU ARG
SEQRES 13 A 163 LEU ALA ARG ASN GLY PRO SER
HET CL A 164 1
HETNAM CL CHLORIDE ION
FORMUL 2 CL CL 1-
FORMUL 3 HOH *152 (H2 O)

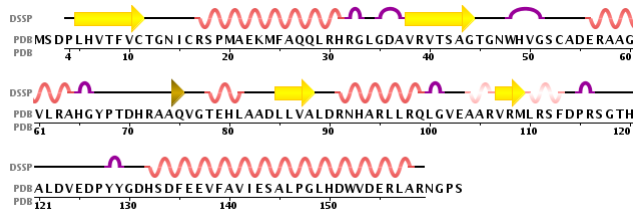
```

PDB entry 1U2P: Secondary structure

```

HELIX  1  1  CYS A   16  ARG A   32  1  17
HELIX  2  2  ASP A   55  HIS A   65  1  11
HELIX  3  3  GLY A   77  ALA A   82  1   6
HELIX  4  4  ASP A   90  LEU A  100  1  11
HELIX  5  5  GLU A  103  ALA A  105  5   3
HELIX  6  6  ARG A  111  ASP A  114  5   4
HELIX  7  7  ASP A  131  ARG A  159  1  29
SHEET  1  A 4 VAL A   38  GLY A   44  0
SHEET  2  A 4 LEU A   5   CYS A  11  1  N LEU A   5   O ARG A  39
SHEET  3  A 4 LEU A  85  ALA A  88  1  O VAL A  87  N THR A   8
SHEET  4  A 4 VAL A 107  MET A 109  1  O ARG A 108  N LEU A  86
SITE   1 AC1  4 THR A  12  GLY A  13  ARG A  17  HOH A 171

```



Secondary structure assignment: DSSP algorithm

Kabsch, W., Sander, C. *Biopolymers* 22, 2577–2637 (1983)

The definitions of H-bonded features form a hierarchy:

1. H-bonds are defined.
2. Based on them, turns and bridges.
3. Based on them, α -helices and β -ladders, including common imperfections such as helical kinks and β -bulges.

Each structural feature is defined independently of the others and structural overlaps are resolved by defining a secondary structure summary that assigns a single state to each residue.

DSSP-Algorithm: H-bonds

Hydrogen bonds in proteins have little wave-function overlap and are well described by an electrostatic model. We calculate the electrostatic interaction energy between two H-bonding groups by placing partial charges on the C,O (+ q_1 , - q_1) and N,H (- q_2 , + q_2) atoms, i.e.,

$$E = q_1 q_2 (1/r(\text{ON}) + 1/r(\text{CH}) - 1/r(\text{OH}) - 1/r(\text{CN})) * f$$

with $q_1 = 0.42e$ and $q_2 = 0.20e$, e being the unit electron charge and $r(\text{AB})$ the interatomic distance from A to B. In chemical units, r is in Å, the dimensional factor $f = 332$, and E is in kcal/mol. A good H bond has about -3 kcal/mol binding energy. We choose a generous cutoff to allow for bifurcated H bonds and errors in coordinates and assign an H bond between C=O of residue i and N-H of residue j if E is less than the cutoff, i.e.,

“Hbond(ij)=: [E < -0.5kcal/mole].”

PDB entry 1U2P: Crystal data

```

CRYST1    40.816   53.610   68.486   90.00   90.00   90.00 P 21 21 21    4
ORIGX1    1.000000   0.000000   0.000000           0.000000
ORIGX2    0.000000   1.000000   0.000000           0.000000
ORIGX3    0.000000   0.000000   1.000000           0.000000
SCALE1    0.024500   0.000000   0.000000           0.000000
SCALE2    0.000000   0.018653   0.000000           0.000000
SCALE3    0.000000   0.000000   0.014602           0.000000

```

| Experimental Details | | Hide |
|-----------------------------------|-------------------------|------|
| Method: X-RAY DIFFRACTION | | |
| Exp. Data: | | |
| Structure Factors | | |
| EDS | | |
| Resolution[Å]: | 1.90 | |
| R-Value: | 0.202 (obs.) | |
| R-Free: | 0.227 | |
| Space Group: | P 21 21 21 ^p | |
| Unit Cell: | | |
| Length [Å] | Angles [°] | |
| a = 40.82 | α = 90.00 | |
| b = 53.61 | β = 90.00 | |
| c = 68.49 | γ = 90.00 | |

PDB entry 1U2P: Coordinates

| | | | | | | | | | | | |
|------|----|-----|-----|---|---|-------|---------|--------|------|-------|---|
| ATOM | 1 | N | PRO | A | 4 | 6.719 | -12.134 | 26.603 | 1.00 | 18.91 | N |
| ATOM | 2 | CA | PRO | A | 4 | 6.735 | -10.746 | 27.122 | 1.00 | 18.45 | C |
| ATOM | 3 | C | PRO | A | 4 | 6.209 | -9.735 | 26.108 | 1.00 | 16.72 | C |
| ATOM | 4 | O | PRO | A | 4 | 6.701 | -9.658 | 24.983 | 1.00 | 16.64 | O |
| ATOM | 5 | CB | PRO | A | 4 | 8.174 | -10.427 | 27.495 | 1.00 | 20.82 | C |
| ATOM | 6 | CG | PRO | A | 4 | 8.942 | -11.387 | 26.584 | 1.00 | 20.17 | C |
| ATOM | 7 | CD | PRO | A | 4 | 8.093 | -12.664 | 26.557 | 1.00 | 22.00 | C |
| ATOM | 8 | N | LEU | A | 5 | 5.207 | -8.963 | 26.521 | 1.00 | 16.15 | N |
| ATOM | 9 | CA | LEU | A | 5 | 4.605 | -7.937 | 25.674 | 1.00 | 14.51 | C |
| ATOM | 10 | C | LEU | A | 5 | 5.700 | -6.960 | 25.244 | 1.00 | 14.38 | C |
| ATOM | 11 | O | LEU | A | 5 | 6.564 | -6.600 | 26.042 | 1.00 | 15.34 | O |
| ATOM | 12 | CB | LEU | A | 5 | 3.513 | -7.204 | 26.458 | 1.00 | 13.81 | C |
| ATOM | 13 | CG | LEU | A | 5 | 2.639 | -6.180 | 25.737 | 1.00 | 14.69 | C |
| ATOM | 14 | CD1 | LEU | A | 5 | 1.815 | -6.864 | 24.656 | 1.00 | 15.29 | C |
| ATOM | 15 | CD2 | LEU | A | 5 | 1.725 | -5.506 | 26.754 | 1.00 | 15.24 | C |

Keyword

Atom index

Atom name

Residue name
Chain identifier

Residue number

x coordinate

y coordinate

z coordinate

Occupancy

B-factor

Element



PROTEIN STRUCTURE CLASSIFICATION

[Home](#) | [Search](#) | [Documentation](#) | [Tools](#) | [Download](#)

🏠 Home

Welcome to CATH

CATH is a manually curated classification of protein domain structures. Each protein has been chopped into structural domains and assigned into homologous superfamilies (groups of domains that are related by evolution). This classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.

[Search the CATH database >>](#)
[Find out more about CATH >>](#)

New in CATH v3.4

CATH v3.4 is built from 104,238 PDB chains. We have added the following data since v3.3:

- 49 folds (total 1,282)
- 163 superfamilies (total 2,549)
- 1,311 sequence families (total 11,330)
- 24,232 domains (total 152,920)

[Download CATH data >>](#)

New in Gene3D 10.2

Gene3D 10.2 (released Sep 2011) uses CATH domains to provide 16,118,154 structural annotations for 14,963,305 protein sequences. The latest release also offers a number of new features:

- Interaction Network Data
- Interactive Graphical Representations
- Genome Comparisons

[Goto Gene3D >>](#)

Using CATH

- [Search](#)
- [Browse](#)
- [Download](#)
- [Tutorials](#)
- [Introduction to CATH](#)
- [FAQ](#)

CATH Tools

- [Find My Sequence](#)
- [Find My Structure](#)
- [CATHEDRAL Structure Scan](#)
- [Structure Comparison Server](#)
- [Linking to CATH](#)

About CATH

- [Release Statistics](#)
- [Glossary](#)
- [CATH Team](#)
- [References](#)

www.cathdb.info

CATH Protein Structure Classification

- CATH is a manually curated classification of protein domain structures.
- Each protein has been chopped into structural domains and assigned to homologous superfamilies (groups of domains related by evolution).
- This classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.
- CATH is a tree-like organization of nodes that begins with the class node (i.e. the first branch-point of the tree) and ends with the domain nodes (i.e. the leaves of the tree).
- Each node below the class have a parent that they belong to, e.g. the parent of the **H** level (Homologous superfamily) is the **T** (Topology).
- Additionally, each node above the domains have child nodes that belong to them, e.g. the child nodes of a given **H** (Homologous superfamily) level are the S35 families (domains clustered at > 35% sequence identity).

CATH Hierarchical classification

C Class: Class is determined according to the secondary structure composition and packing within the structure. Four major classes are recognized; mainly-alpha, mainly-beta, alpha-beta, and low secondary structure content.

A Architecture: This describes the overall shape of the domain structure as determined by the orientations of the secondary structures but ignores the connectivity between the secondary structures. It is assigned manually using a simple description of the secondary structure arrangement e.g. barrel or 3-layer sandwich. Reference is made to the literature for well-known architectures (e.g the β -propeller or α four helix bundle).

T Topology (Fold family): Structures are grouped according to whether they share the same topology or fold in the core of the domain, that is, if they share the same overall shape and connectivity of the secondary structures in the domain core. Domains in the same fold group may have different structural decorations to the common core.

H Homologous Superfamily, H-level: This level groups together protein domains which are thought to share a common ancestor and can therefore be described as homologous. Similarities are identified either by high sequence identity or structure comparison using SSAP.

S, O, L, I, D Sequence Family Levels: Domains within each H-level are subclustered into sequence families using multi-linkage clustering at S = 35, O = 60, L = 95, and I = 100% sequence identity levels. The D-level acts as a counter within each S100 family.

CATH
PROTEIN STRUCTURE CLASSIFICATION

[Home](#) | [Search](#) | [Documentation](#) | [Tools](#) | [Download](#)

Home > Search > 1 CathDB: V3_4_0 (change)

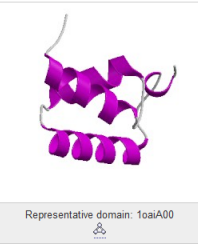
Class: 1
Mainly Alpha

Classification Lineage (1)

| CATH Code | Level Description | Links |
|-----------|-------------------|-------|
| 1 | Mainly Alpha | |

Summary of Non-Redundant Representatives

| | 5 | 376 | 839 | 2763 | 3571 | 4679 | 9217 | 32396 |
|---|---|-----|-----|------|------|------|------|-------|
| - | 5 | 376 | 839 | 2763 | 3571 | 4679 | 9217 | 32396 |






Representative domain: 1oaiA00

Non-Redundant Representatives (5) [Help](#)

What is a Non-Redundant Representative? [Expand](#)

Architecture Entries in Class 1 (5)

| CATH Level | CATH code | Name | Representative Domain | Representative Keywords | Representative Thumbnail | Number of Domains |
|------------|-----------|-------------------|-----------------------|---|---|-------------------|
| 1.10 | 1oaiA00 | Orthogonal Bundle | 1oaiA00 | Nuclear RNA export factor 1 Homo sapiens Protein binding |  | 22755 |
| 1.20 | 1mz9A00 | Up-down Bundle | 1mz9A00 | Mus musculus Cartilage oligomeric matrix protein |  | 8531 |
| 1.25 | 1wa5B00 | Alpha Horseshoe | 1wa5B00 | Protein import into nucleus Saccharomyces cerevisiae Protein binding Protein transporter activity Protein targeting to membrane |  | 667 |

SCOP Structural Classification of Proteins


Structural Classification of Proteins

Root: scop

Classes:

1. [All alpha proteins](#) [46456] (284)
2. [All beta proteins](#) [48724] (174)
3. [Alpha and beta proteins \(a/b\)](#) [51349] (147)
4. [Alpha and beta proteins \(a+b\)](#) [53931] (376)
5. [Multi-domain proteins \(alpha and beta\)](#) [56572] (66)
6. [Membrane and cell surface proteins and peptides](#) [56835] (58)
7. [Small proteins](#) [56992] (90)
8. [Coiled coil proteins](#) [57942] (7)
9. [Low resolution protein structures](#) [58117] (26)
10. [Peptides](#) [58231] (121)
11. [Designed proteins](#) [58788] (44)

Experimental structures of proteins with essentially non-natural sequences. Not a true class



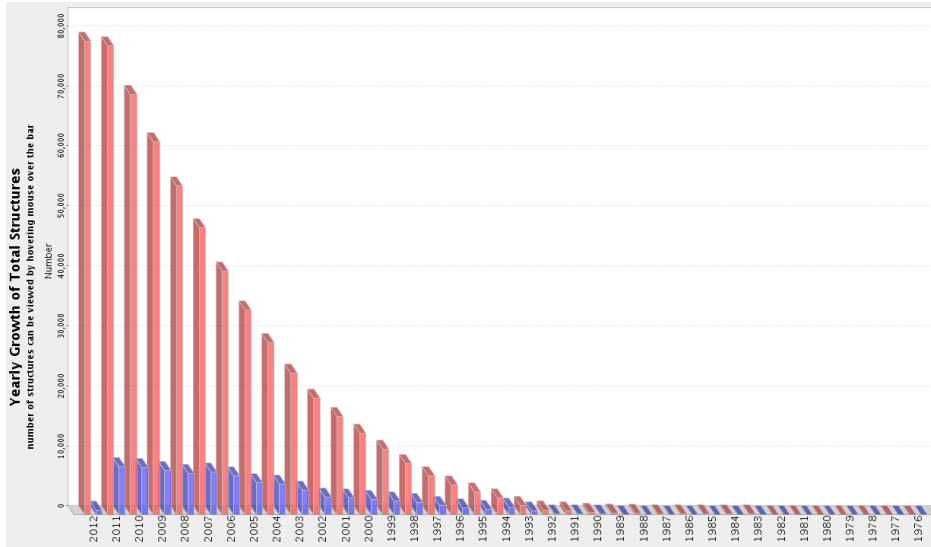
Scop Classification Statistics

SCOP: Structural Classification of Proteins. 1.75 release
38221 PDB Entries (23 Feb 2009). 110800 Domains. 1 Literature Reference (excluding nucleic acids and theoretical models)

| Class | Number of folds | Number of superfamilies | Number of families |
|------------------------------------|-----------------|-------------------------|--------------------|
| All alpha proteins | 284 | 507 | 871 |
| All beta proteins | 174 | 354 | 742 |
| Alpha and beta proteins (a/b) | 147 | 244 | 803 |
| Alpha and beta proteins (a+b) | 376 | 552 | 1055 |
| Multi-domain proteins | 66 | 66 | 89 |
| Membrane and cell surface proteins | 58 | 110 | 123 |
| Small proteins | 90 | 129 | 219 |
| Total | 1195 | 1962 | 3902 |

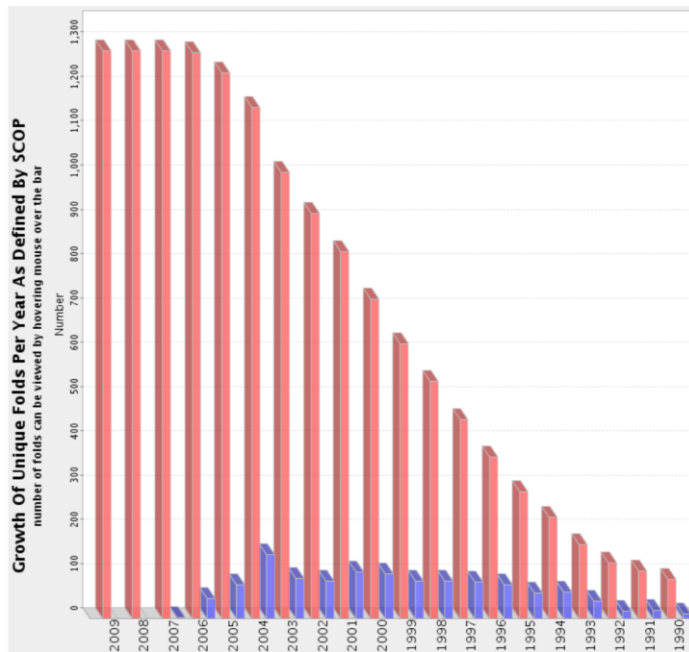
<http://scop.mrc-lmb.cam.ac.uk/scop>

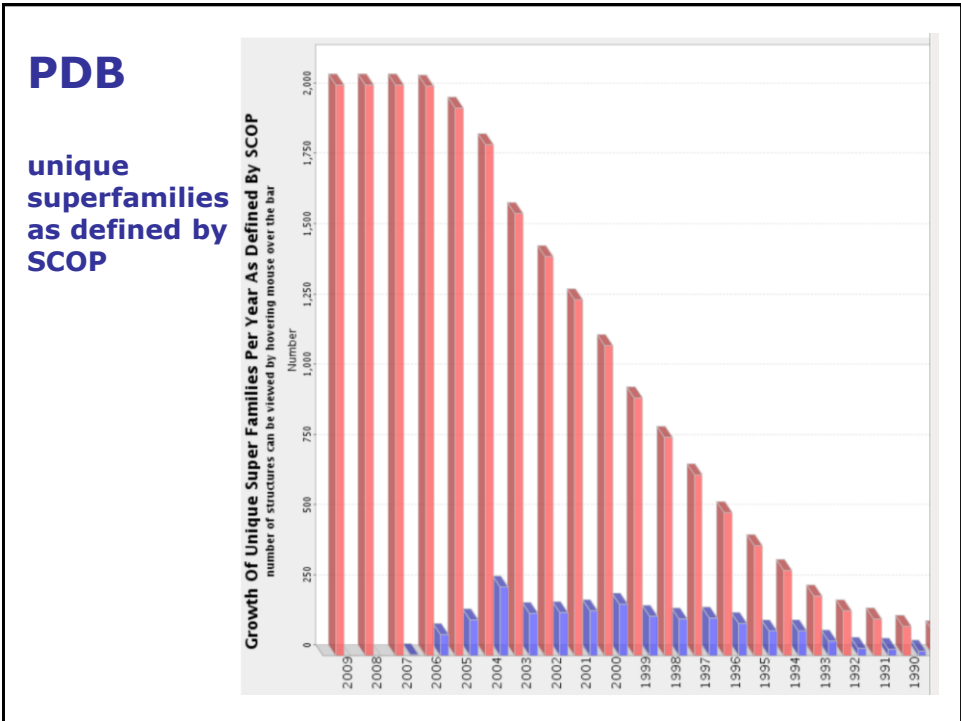
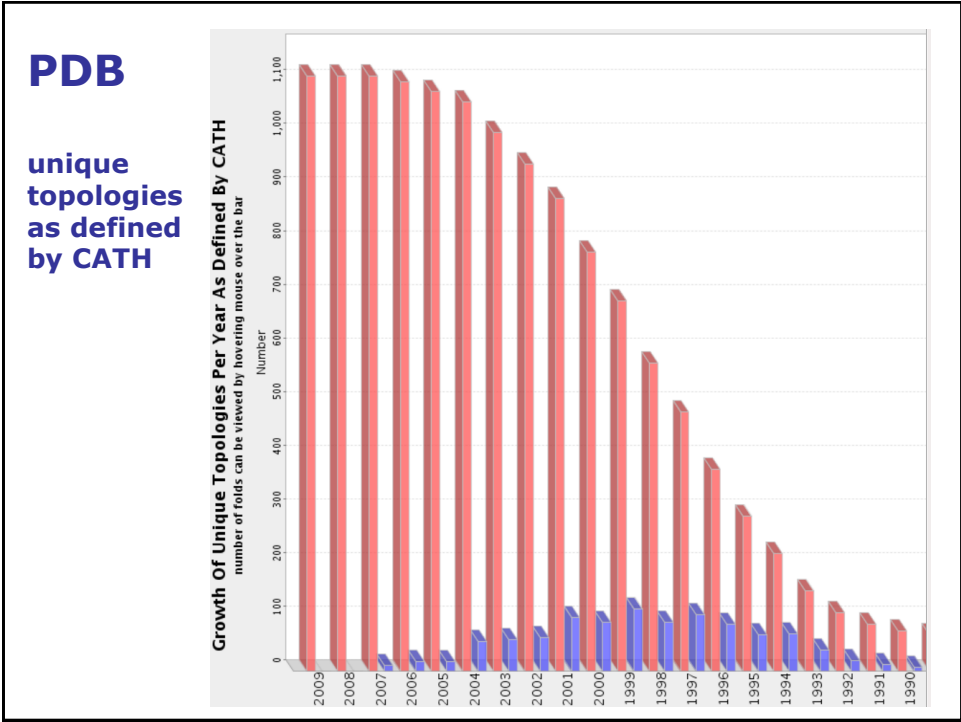
PDB: Total structures



PDB

unique folds as defined by SCOP





PDB

unique superfamilies as defined by CATH

