

(Aspekte der Thermodynamik in der Strukturbiologie)

Einführung in die Bioinformatik

Wintersemester 2012/13

Peter Güntert

Protein Data Bank (PDB)

Outline

- PDB content overview
- Protein structure classification
- Format and content of PDB entries
- Structure comparison

The screenshot displays the RCSB PDB website interface. At the top, it features the RCSB PDB logo and the text 'An Information Portal to Biological Macromolecular Structures'. Below the logo, there is a search bar with the text 'Search | All Categories: e.g., PDB ID, molecule name, author'. The main content area is titled 'Biological Macromolecular Resource' and includes a 'Full Description' section. This section highlights 'Molecule of the Month: Transfer-Messenger RNA' and 'Designer Proteins'. The 'Designer Proteins' section describes the engineering of new proteins with novel structures and functions. The interface also includes a sidebar with navigation options like 'Home', 'Deposition', 'Tools', and 'Help'. At the bottom, there is a footer with the text '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.' and the website URL 'www.pdb.org'.

Protein Data Bank (PDB)

- Contains all publicly available experimentally determined three-dimensional protein structures
- One entry for each structure with an accession code consisting of 1 digit (1–9) and 3 characters (A–Z, 1–9), e.g. 1ABC

PDB Current Holdings Breakdown 22.01.2013

Exo.Method	Proteins	Nucleic Acids	Protein/NA Complexes	Other	Total
X-RAY	72060	1432	3686	3	77181
NMR	8564	1017	191	7	9779
ELECTRON MICROSCOPY	339	39	123	0	501
HYBRID	45	3	2	1	51
other	147	4	5	13	169
Total	81155	2495	4007	24	87681

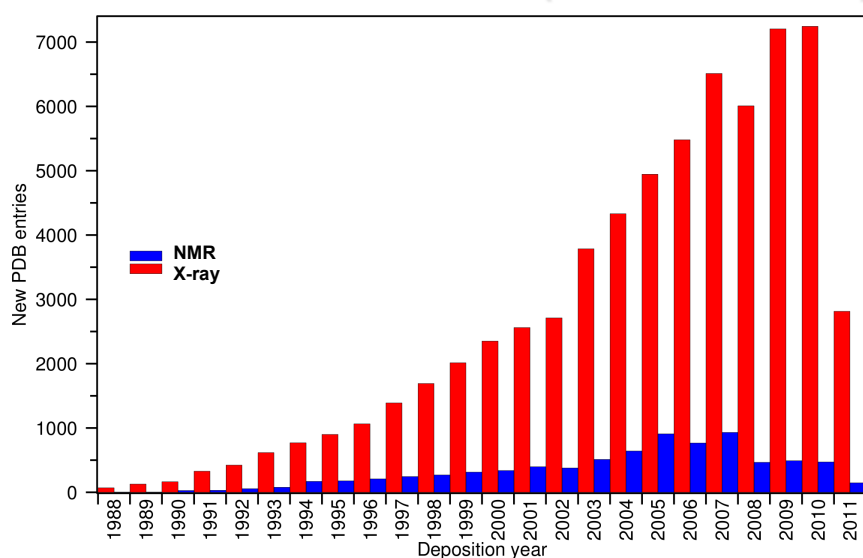
(Click on any number to retrieve the results from that category.)

[66634](#) structures in the PDB have a structure factor file.

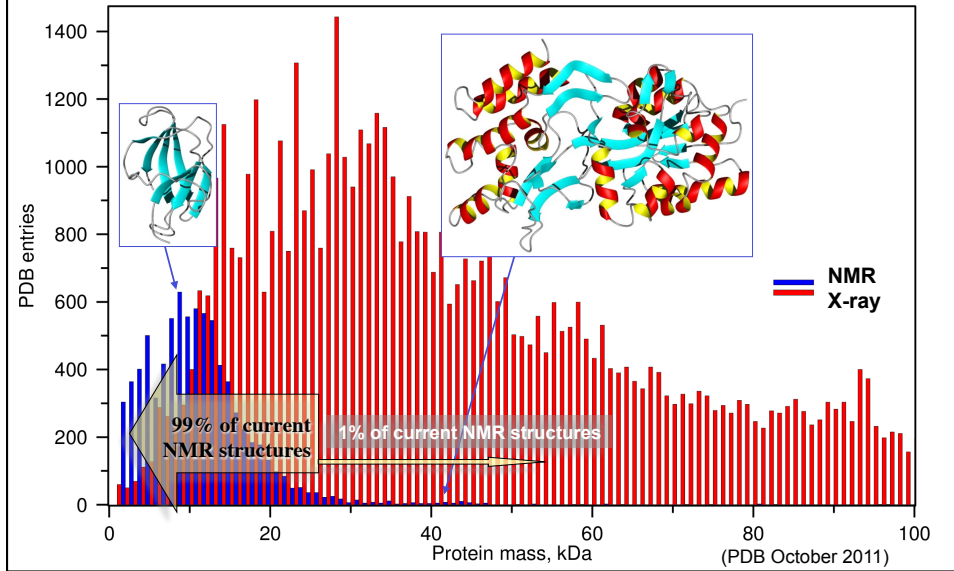
[7086](#) structures in the PDB have an NMR restraint file.

[844](#) structures in the PDB have a chemical shifts file.

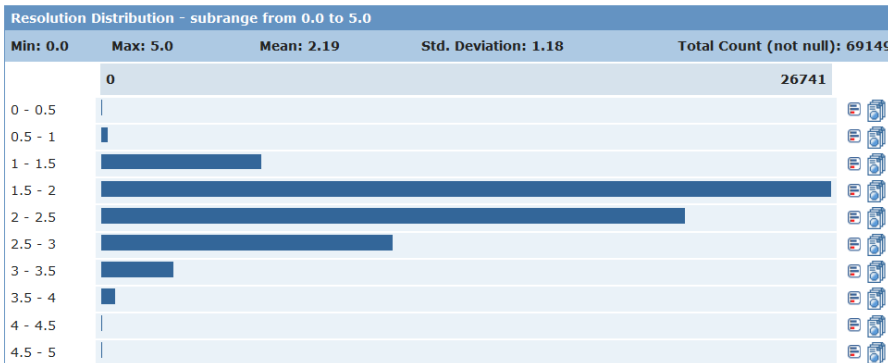
Number of released X-ray and NMR structures in the PDB (October 2011)



Protein structures in the PDB

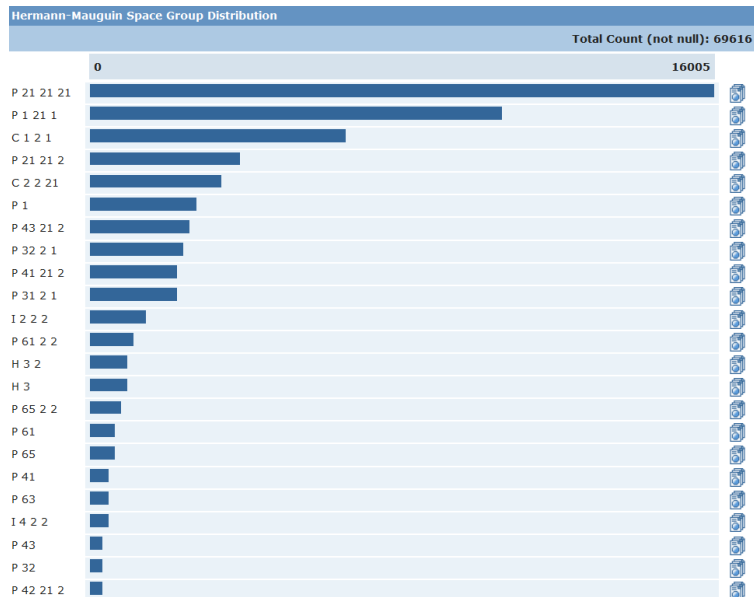


PDB: Resolution distribution

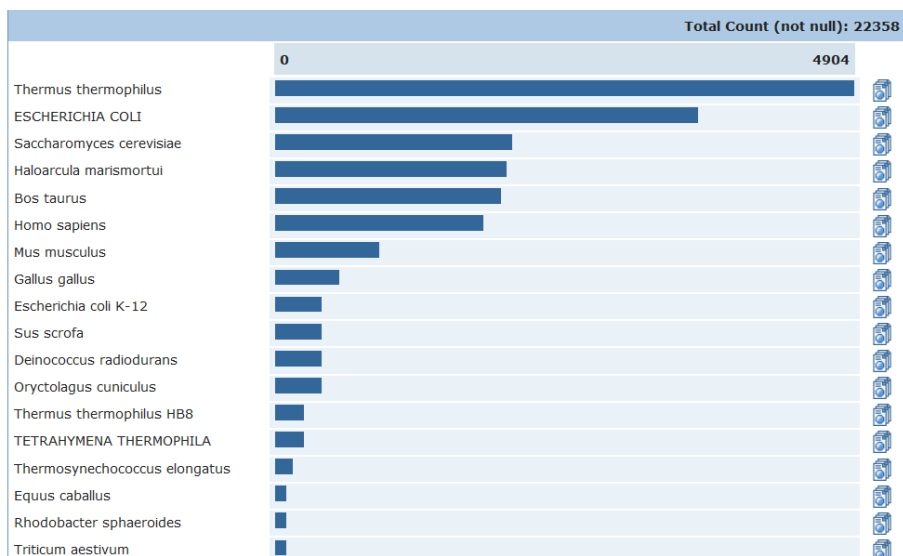


There are 1551 values outside the current range - for a complete histogram [click here](#).

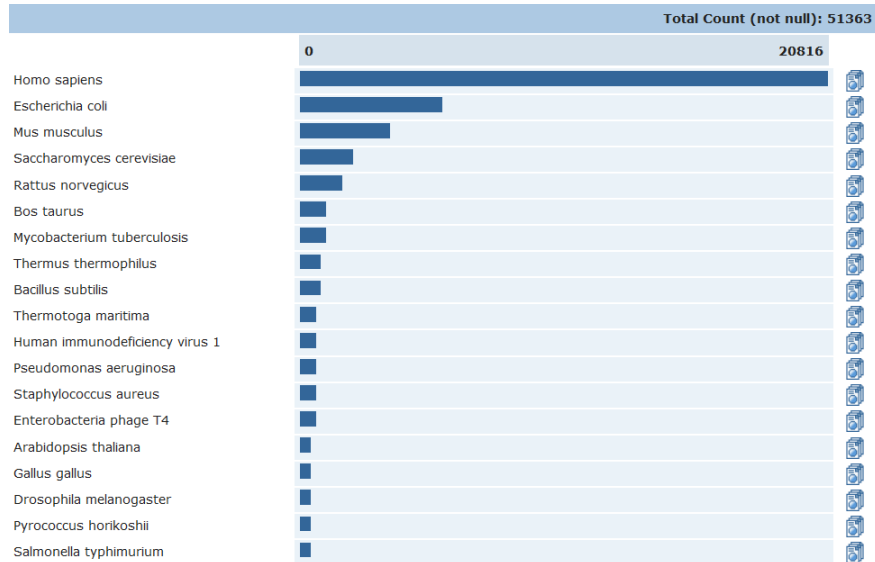
PDB: Distribution by space group



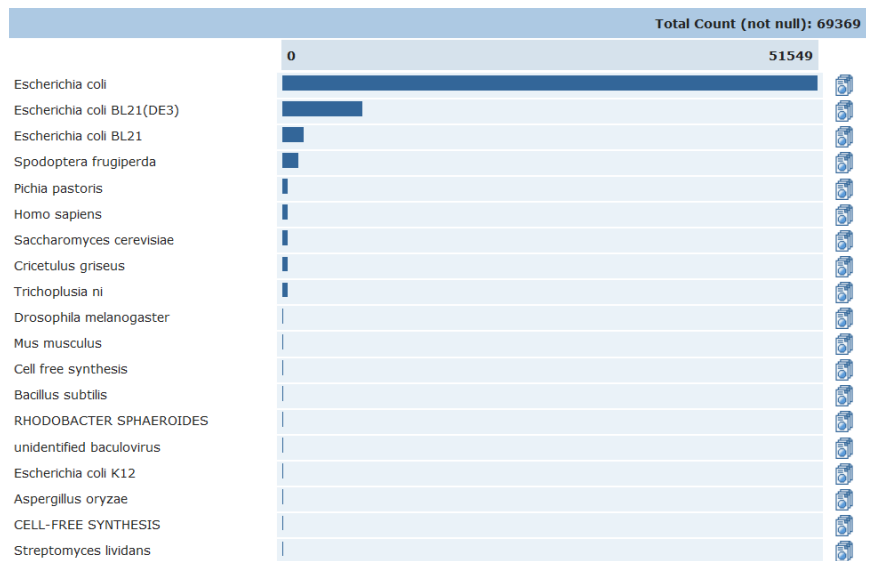
Distribution by natural source organism



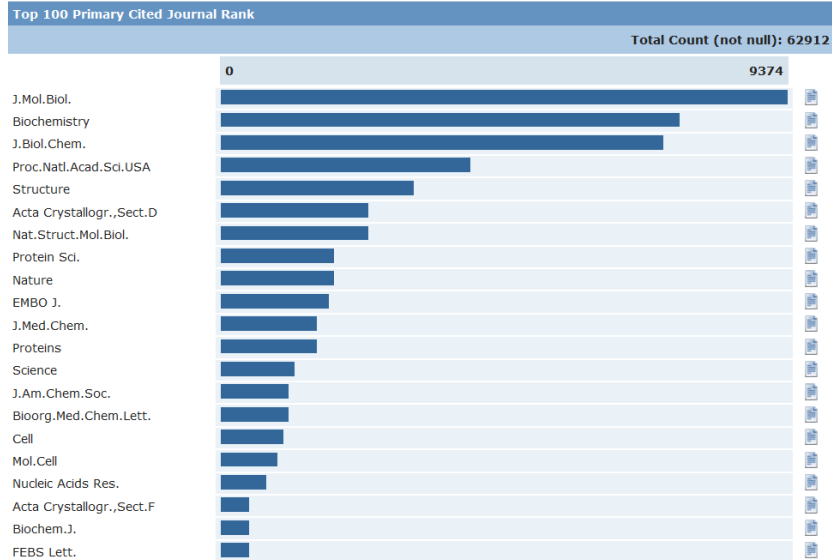
Distribution by gene source organism



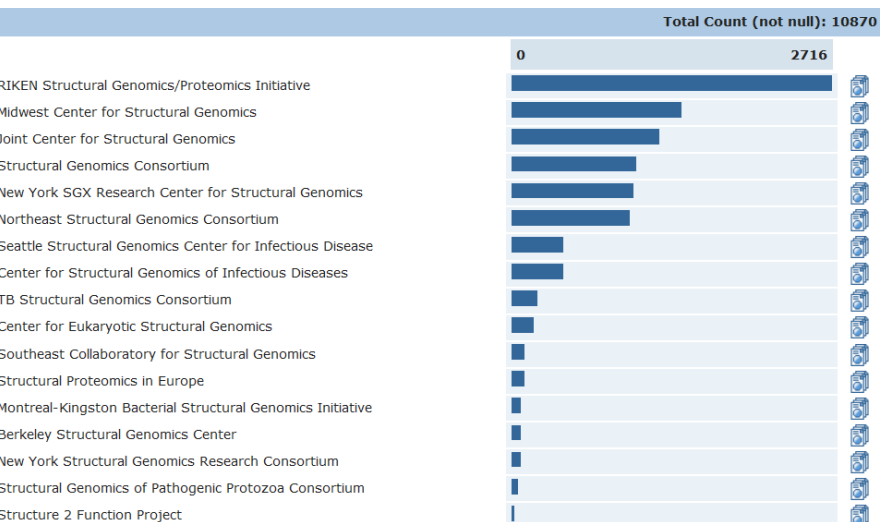
Distribution by expression organism



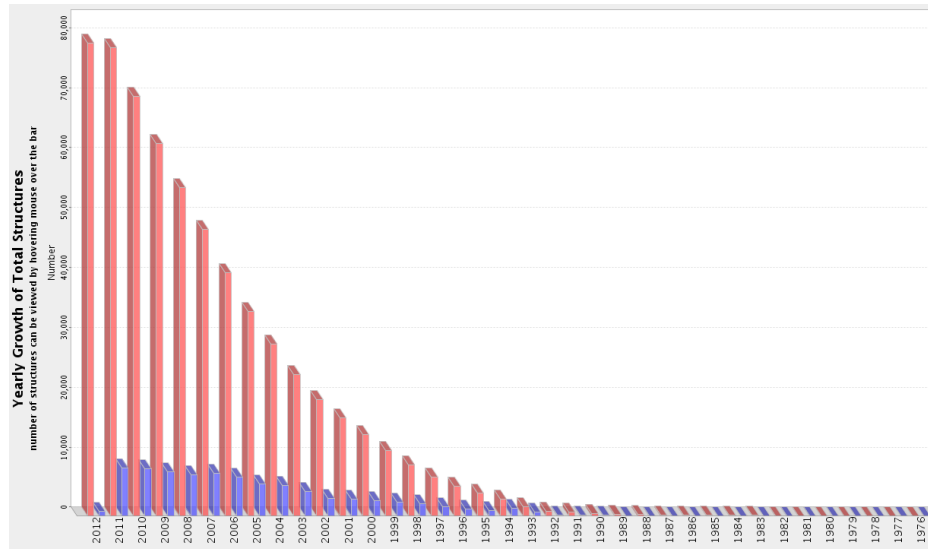
Distribution by primary cited journal



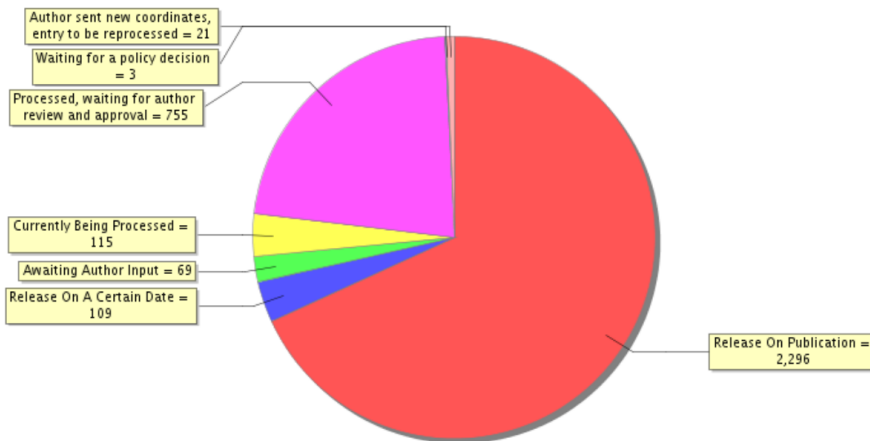
Distribution by structural genomics center



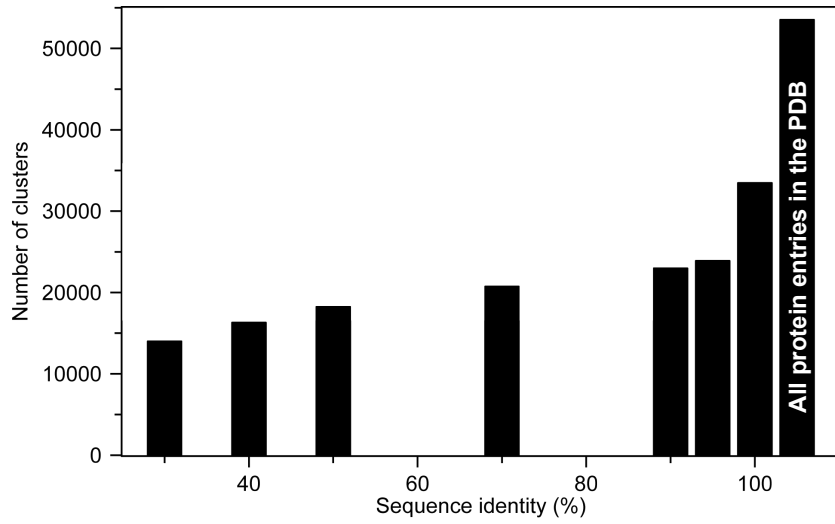
PDB: Total structures



PDB: Status of unreleased entries



PDB redundancy



Structure classification

CATH Home Search * Browse Download About Support Search CATH by keywords or ID

CATH / Gene3D

16 million protein domains classified into 2,626 superfamilies

Get Started » Search » Download » Take the Tour »

What's New?

The CATH website has recently undergone a big overhaul. We really hope you find the new pages more useful, easier to use and quicker to load. Please [get in touch](#) and let us know what you think.

Latest News

CATH @ ECCB 2012 September 9, 2012

"Using CATH-Gene3D to study the evolution of your protein and find its function" - Prof Orengo presents the new CATH website at ECCB

Latest Release

CATH v3.5	
based on PDB dated September 20, 2011	
173,536	CATH Domains
2,626	CATH Superfamilies
51,334	PDBs

Gene3D v11	
released March 18, 2012	
1,639	Cellular Genomes
1,016	Viral Genomes
14,963,305	Protein Sequences
16,297,076	CATH Domain Predictions

www.cathdb.info

CATH News	Get Started	Download	About
Support	Documentation	WebServices	Orengo Group
Jobs	Tutorials	Software	Web accessibility

© CATH, 1997-2012

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.



PROTEIN STRUCTURE CLASSIFICATION

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

Search
Clear

Home > Search > 1 CathDB: V3_4_0 (change)

Class: 1

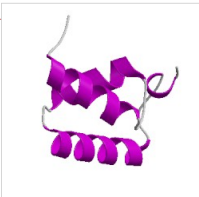
Mainly Alpha

Classification Lineage (1)

CATH Code	Level Description	Links
1	Mainly Alpha	

Summary of Non-Redundant Representatives

	S	O	L	I	D			
	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	1.10	Orthogonal Bundle	1oaiA00	Nuclear RNA export factor 1 Homo sapiens Protein binding		22755
1.20	1.20	Up-down Bundle	1mz9A00	Mus musculus Cartilage oligomeric matrix protein		8531
1.25	1.25	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

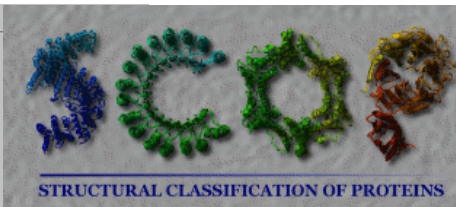


Root: scop

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)

Enter [search](#) key:



STRUCTURAL CLASSIFICATION OF PROTEINS

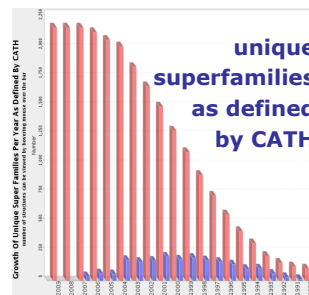
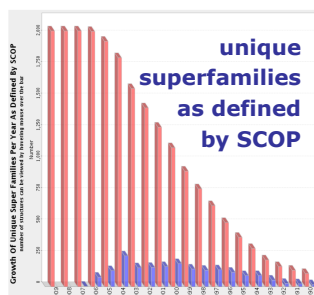
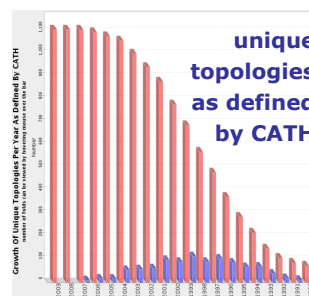
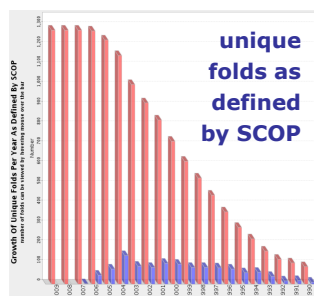
Scop Classification Statistics

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>

New folds in the PDB



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

DOI:10.2210/pdb1u2p/pdb

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: 2175-2181
 PubMed: 15743966 [Search Related Articles in PubMed](#)

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

Molecular Description [Hide](#)
 Classification: **Hydrolase**
 Structure Weight: 17554.65
 Molecule: low molecular weight protein-tyrosine-phosphatase
 Polymer: 1 Type: polypeptide(L) Length: 163
 Chains: A
 EC#: 3.1.3.48 [GO](#)

Source [Hide](#)
 Polymer: 1 Scientific Name: **Mycobacterium tuberculosis** Expression System: **Escherichia coli**

Related PDB Entries [Hide](#)

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 [Hide](#)

Identifier	Name	Formula	Links
CL	CHLORIDE ION	Cl	CS DOI PubMed

Derived Data [Hide](#)

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

Biological Molecule [More Images...](#)
 3-D Viewers: [Jmol](#) [SimpleViewer](#) [Protein Workshop](#) [Other Viewers](#)

Deposition Summary [Hide](#)
 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 (NEVDAT): 2009-02-24

Experimental Details [Hide](#)
 Method: X-RAY DIFFRACTION
 Experimental Data: [EDS](#)
 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

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 GO	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** | Seq. Similarity | Literature | Biol. & Chem. | Methods | Geometry | Links

Sequence / Structure Details **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 1 chains. Out of these 1 are sequence-unique.
Currently viewing **unique chains** only. [[show all chains](#)] [[show 3D in Jmol](#)]

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 **1u2pAB** : 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 **MSDPLHVTFCVCTGNI CRSPMAEKMFAQQLRHRGLGDAVRTSAGTGNWHVGS CADERAAG**
PDB 4 10 20 30 40 50 60

DSSP
PDB **VLR AHGYPTDHR AAGVGEHLAADLLVALDRNHARLLRQLGVEAARVMLRSFDP RSGTH**
PDB 61 70 80 90 100 110 120

DSSP
PDB **ALDVEDPYYGDSDFEEVFAVIESALPGLHWDWDERLARNGPS**
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)

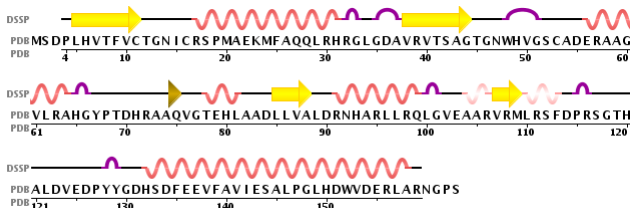
```

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  0.000000
ORIGX2  0.000000  1.000000  0.000000  0.000000  0.000000
ORIGX3  0.000000  0.000000  1.000000  0.000000  0.000000
SCALE1  0.024500  0.000000  0.000000  0.000000  0.000000
SCALE2  0.000000  0.018653  0.000000  0.000000  0.000000
SCALE3  0.000000  0.000000  0.014602  0.000000  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

Unterlagen zur Vorlesung

<http://www.bpc.uni-frankfurt.de/guentert/wiki/index.php/Teaching>