

KFC/STBI

Structural bioinformatics

03 structure storage
and structural DBs - PDB

Karel Berka

Jak ukládat struktury

- formáty – PDB (mmCIF)
 - xyz souřadnice
 - tagy
 - atomové typy
- databáze – www.pdb.org

PDB formát: vzor

```

HEADER      LYASE (OXO-ACID)                                01-OCT-91   12CA      12CA    2
COMPND      CARBONIC ANHYDRASE /II (CARBONATE DEHYDRATASE) (/HCA II) 12CA    3
COMPND      2 (E.C.4.2.1.1) MUTANT WITH VAL 121 REPLACED BY ALA (/V121A) 12CA    4
SOURCE      HUMAN (HOMO SAPIENS) RECOMBINANT PROTEIN                12CA    5
AUTHOR      S.K.NAIR,D.W.CHRISTIANSON                          12CA    6
REVDAT      1   15-OCT-92 12CA      0                            12CA    7
JRNL        AUTH   S.K.NAIR,T.L.CALDERONE,D.W.CHRISTIANSON,C.A.FIERKE 12CA    8
JRNL        TITL   ALTERING THE MOUTH OF A HYDROPHOBIC POCKET.        12CA    9
JRNL        TITL 2  STRUCTURE AND KINETICS OF HUMAN CARBONIC ANHYDRASE 12CA   10
JRNL        TITL 3  /II$ MUTANTS AT RESIDUE VAL-121                  12CA   11
JRNL        REF    J.BIOL.CHEM.                                V. 266 17320 1991 12CA   12
JRNL        REFN   ASTM JBCHA3   US ISSN 0021-9258                071 12CA   13
REMARK      1                                             12CA   14
REMARK      2                                             12CA   15
REMARK      2 RESOLUTION. 2.4  ANGSTROMS.                      12CA   16
REMARK      3                                             12CA   17
REMARK      3 REFINEMENT.                                       12CA   18
REMARK      3  PROGRAM                                PROLSQ        12CA   19
REMARK      3  AUTHORS                                HENDRICKSON,KONNERT        12CA   20
REMARK      3  R VALUE                                0.170          12CA   21
REMARK      3  RMSD BOND DISTANCES                    0.011  ANGSTROMS 12CA   22
REMARK      3  RMSD BOND ANGLES                      1.3    DEGREES    12CA   23
REMARK      4                                             12CA   24
REMARK      4 N-TERMINAL RESIDUES SER 2, HIS 3, HIS 4 AND C-TERMINAL 12CA   25
REMARK      4 RESIDUE LYS 260 WERE NOT LOCATED IN THE DENSITY MAPS AND, 12CA   26
REMARK      4 THEREFORE, NO COORDINATES ARE INCLUDED FOR THESE RESIDUES. 12CA   27

```

.....

PDB (cont.)

SHEET	3	S10 PHE	66	PHE	70	-1	O	ASN	67	N	LEU	60	12CA	68
SHEET	4	S10 TYR	88	TRP	97	-1	O	PHE	93	N	VAL	68	12CA	69
SHEET	5	S10 ALA	116	ASN	124	-1	O	HIS	119	N	HIS	94	12CA	70
SHEET	6	S10 LEU	141	VAL	150	-1	O	LEU	144	N	LEU	120	12CA	71
SHEET	7	S10 VAL	207	LEU	212	1	O	ILE	210	N	GLY	145	12CA	72
SHEET	8	S10 TYR	191	GLY	196	-1	O	TRP	192	N	VAL	211	12CA	73
SHEET	9	S10 LYS	257	ALA	258	-1	O	LYS	257	N	THR	193	12CA	74
SHEET	10	S10 LYS	39	TYR	40	1	O	LYS	39	N	ALA	258	12CA	75
TURN	1	T1 GLN	28	VAL	31			TYPE VIB (CIS-PRO 30)					12CA	76
TURN	2	T2 GLY	81	LEU	84			TYPE II (PRIME) (GLY 82)					12CA	77
TURN	3	T3 ALA	134	GLN	137			TYPE I (GLN 136)					12CA	78
TURN	4	T4 GLN	137	GLY	140			TYPE I (ASP 139)					12CA	79
TURN	5	T5 THR	200	LEU	203			TYPE VIA (CIS-PRO 202)					12CA	80
TURN	6	T6 GLY	233	GLU	236			TYPE II (GLY 235)					12CA	81
CRYST1	42.700	41.700	73.000	90.00	104.60	90.00	P	21				2	12CA	82
ORIGX1	1.000000	0.000000	0.000000			0.000000							12CA	83
ORIGX2	0.000000	1.000000	0.000000			0.000000							12CA	84
ORIGX3	0.000000	0.000000	1.000000			0.000000							12CA	85
SCALE1	0.023419	0.000000	0.006100			0.000000							12CA	86
SCALE2	0.000000	0.023981	0.000000			0.000000							12CA	87
SCALE3	0.000000	0.000000	0.014156			0.000000							12CA	88
ATOM	1	N	TRP	5		8.519	-0.751	10.738	1.00	13.37			12CA	89
ATOM	2	CA	TRP	5		7.743	-1.668	11.585	1.00	13.42			12CA	90
ATOM	3	C	TRP	5		6.786	-2.502	10.667	1.00	13.47			12CA	91
ATOM	4	O	TRP	5		6.422	-2.085	9.607	1.00	13.57			12CA	92
ATOM	5	CB	TRP	5		6.997	-0.917	12.645	1.00	13.34			12CA	93
ATOM	6	CG	TRP	5		5.784	-0.209	12.221	1.00	13.40			12CA	94
ATOM	7	CD1	TRP	5		5.681	1.084	11.797	1.00	13.29			12CA	95
ATOM	8	CD2	TRP	5		4.417	-0.667	12.221	1.00	13.34			12CA	96
ATOM	9	NE1	TRP	5		4.388	1.418	11.515	1.00	13.30			12CA	97
ATOM	10	CE2	TRP	5		3.588	0.375	11.797	1.00	13.35			12CA	98
ATOM	11	CE3	TRP	5		3.837	-1.877	12.645	1.00	13.39			12CA	99
ATOM	12	CZ2	TRP	5		2.216	0.208	11.656	1.00	13.39			12CA	100
ATOM	13	CZ3	TRP	5		2.465	-2.043	12.504	1.00	13.33			12CA	101
ATOM	14	CH2	TRP	5		1.654	-1.001	12.009	1.00	13.34			12CA	102

.....

mmCIF

Výrazně popisnější krystalografický slovníkový formát – podobný XML, používá kategorie:

- atom group
 - Categories that describe the properties of atoms.
- audit group
 - Categories that describe dictionary maintenance and identification.
- cell group
 - Categories that describe the unit cell.
- chemical group
 - Categories that describe chemical properties and nomenclature.
- chem comp group
 - Categories that describe components of chemical structure.
- chem link group
 - Categories that describe links between components of chemical structure.
- citation group
 - Categories that provide bibliographic references.
- computing group
 - Categories that describe the computational details of the experiment.
- compliance group
 - Categories that are included in this dictionary specifically to comply with previous dictionaries.
- database group
 - Categories that hold references to entries in databases that contain related information.
- diffn group
 - Categories that describe details of the diffraction experiment.
- entity group
 - Categories that describe chemical entities.
- entry group
 - Categories that pertain to the entire data block.
- exptl group
 - Categories that hold details of the experimental conditions.
- geom group
 - Categories that hold details of molecular and crystal geometry.
- iucr group
 - Categories that are used for manuscript submission and internal processing by the Intl Union of Crystallography.
- pdb group
 - Categories that pertain to the file-format or data-processing codes used by the Protein Data Bank.

International Tables for Crystallography Volume G: Definition and exchange of crystallographic data

First online edition (2006)

[ISBN: 978-1-4020-3138-0](#)

[doi:](#)

[10.1107/9780955360206000107](#)

<http://it.iucr.org/Ga/contents/>

21th Century – “Century of life science data”

Nucleic acid sequences



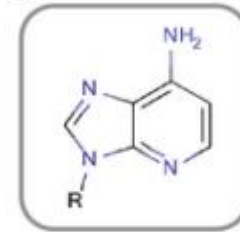
European Nucleotide Archive
ArrayExpress
Expression Atlas
...

Protein (biomacromolecule) structures



Protein Data Bank

Drug-like molecules



PubChem
Zinc
DrugBank
PDB CCD
ChEBI
ChEMBL
...

Databáze – není jich málo...

Structure and sequence/structure databases

SCOP

CATH

FSSP

Molecular Modeling Database

CAMPASS

ISSD

Library of Protein Family Cores (LPFC)

3D_ALI (a database of aligned protein structures
and related sequences)

IDITIS (relational database and query tool for proteins)

HSSP

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

<http://www.biochem.ucl.ac.uk/bsm/cath/>

<http://www2.ebi.ac.uk/dali/fssp/>

<http://www.ncbi.nlm.nih.gov/Structure/>

<http://www-cryst.bioc.cam.ac.uk/~campass/>

<http://www.protein.bio.msu.su/issd/>

<http://WWW-SMI.Stanford.EDU/projects/helix/LPFC/>

http://www.embl-heidelberg.de/argos/ali/ali_info.html

<http://www.oxmol.co.uk/prods/iditis/>

<http://www.sander.embl-heidelberg.de/hssp/>

Speciality databases

HIV Protease Database

Nucleic Acid Database

Prolysis (protease and protease inhibitor Web server)

International Immunogenetics Database (IMGT)

Enzyme Structures Database

<http://www-lbsc.ncifcrf.gov/HIVdb/>

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

<http://delphi.phys.univ-tours.fr/Prolysis/>

<http://imgt.cnusc.fr:8104/>

<http://www.biochem.ucl.ac.uk/bsm/enzymes/>

Features databases

Molecular Movements Database

OLDERADO

PROCAT

Protein Quaternary Structures (PQS)

ReLiBase (receptor-ligand complexes database)

PROMISE

PDBSum

Biological Macromolecule Crystallization Database (BMCD)

<http://bioinfo.mbb.yale.edu/MolMovDB/>

<http://neon.chem.le.ac.uk/olderado/>

<http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html>

<http://pqs.ebi.ac.uk/>

<http://www2.ebi.ac.uk:8081/home.html>

<http://bioinf.leeds.ac.uk/promise/>

<http://www.biochem.ucl.ac.uk/bsm/pdbsum/>

<http://h178133.nist.gov:4400/bmcd/bmcd.html>

Resources

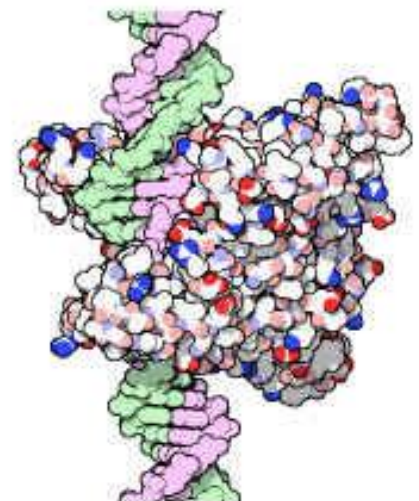
Protein Data Bank

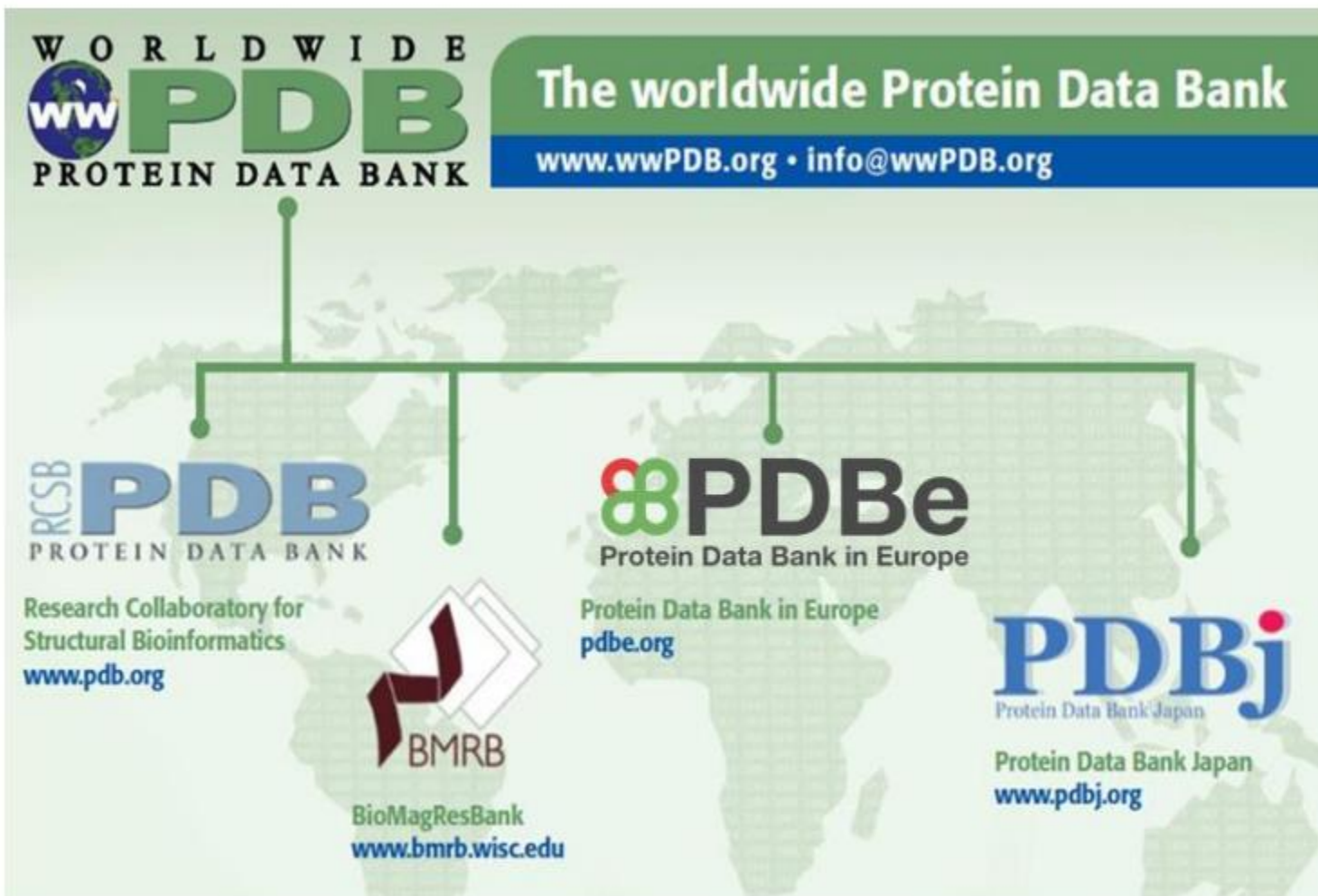
<http://www.rcsb.org/pdb/>

Protein DataBank (PDB)

- Hlavní a klíčová databáze pro řešení molekulárně biologických otázek
- Protein Databank
 - PDB od roku 1972 Brookhaven National Laboratory (BNL)
 - Jedinečné mezinárodní úložiště makromolekulárních strukturních dat
 - dnes pod Research Collaboratory for Structural Bioinformatics (RCSB)

<http://www.rcsb.org/>



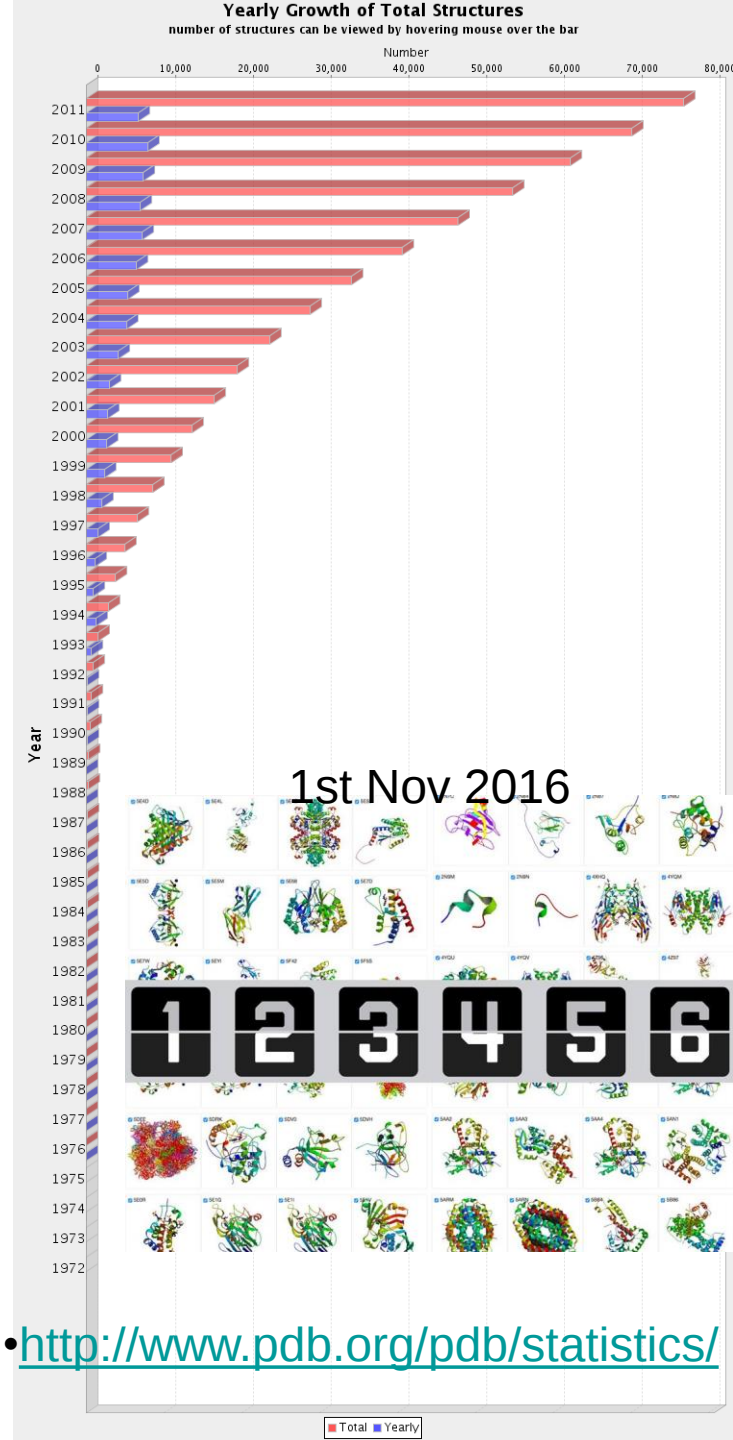


> 120k biomacromolecular structures

PDB statistics

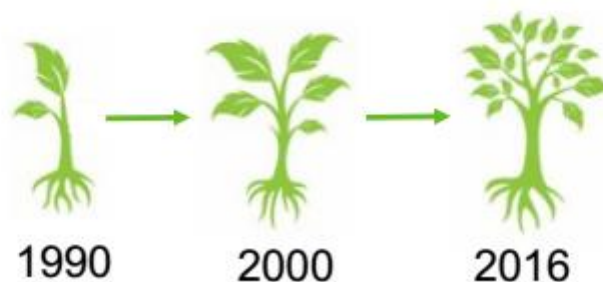
říjen 2011

- Struktury 76814
 - X-ray 67125
 - NMR 9108
 - EM 380
 - Proteiny 71148
 - NA 2312
 - Protein/NA 3331
- zajímavosti, nejvíc struktur:
 - organism-Thermus thermophilus (4751)
 - gene source Homo Sapiens (20127)
 - R 1.5-2.0 Å (26000)
 - Software – CNS (25276)
 - EC – 3.2.1.17 Lysozyme (1150)
 - Residues - 100-300 (30000)
 - Journal J.Mol.Biol. (9247)



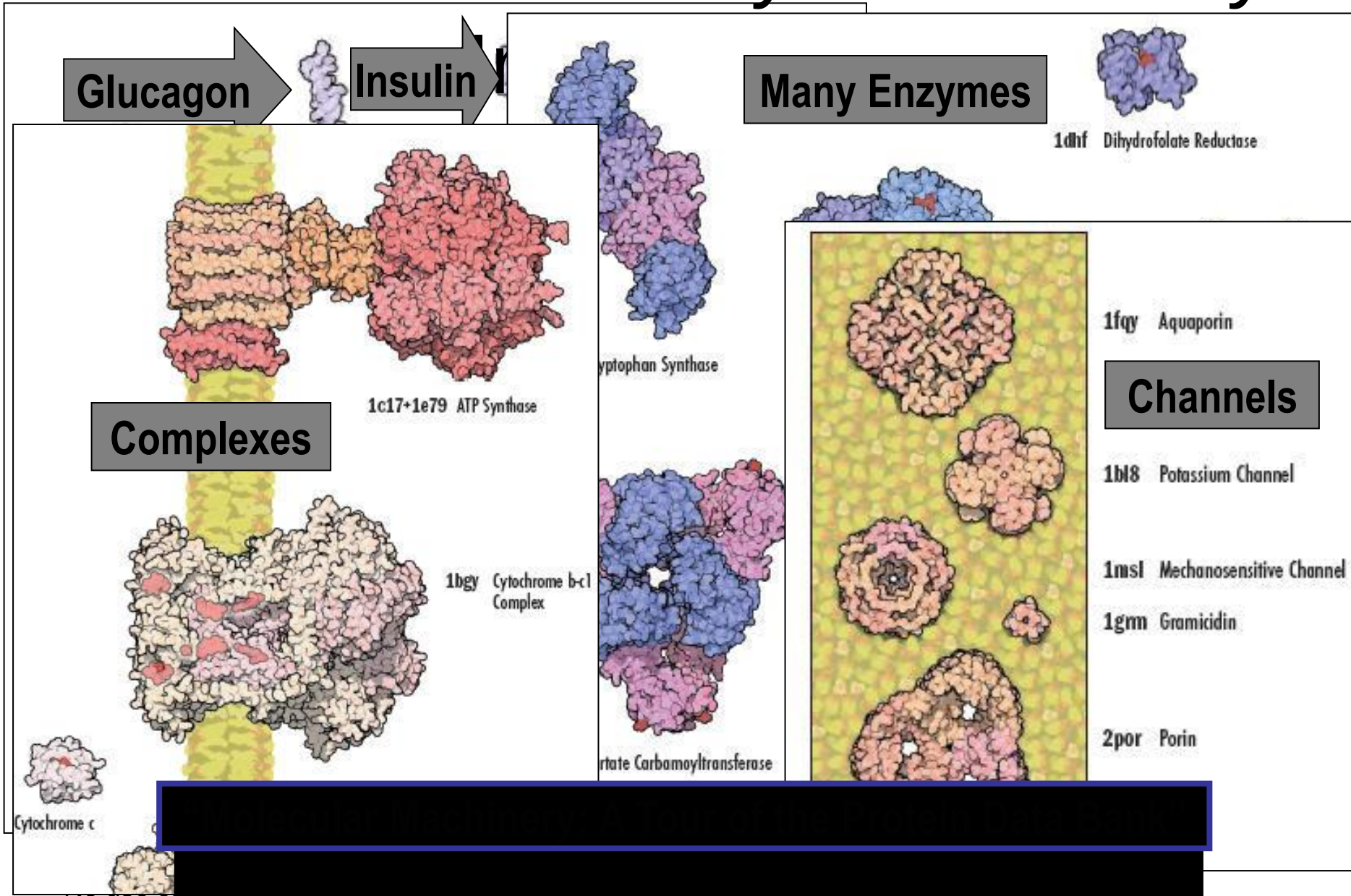
Growth of PDB

Year	1990	2000	2016
Testosterone binding proteins	0	4	72
Cytochrome P450	1	22	121
Lectine LecB	0	9	43
Apoptosis regulators BAK a BAX	0	3	190
Methemoglobine	4	11	37
Can we perform a comparison / analysis?	No	?	Yes



**A richness of data about
individual protein families / folds.**

Structural Diversity - Constantly



An Example of a Structure

Summary Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

ESCHERICHIA COLI PROTEIN ARAC

2ARA

Display Files ▾
Download Files ▾
Share this Page ▾

Explore next

Primary Citation

Structural basis

Soisson, S.M.
Wolberger, C.

Journal: (1997)

PubMed: 9103

Search Related

PubMed Abstract

The crystal structure of the domain of the determined in angstrom structure protein adopt

β-strands

α-helices

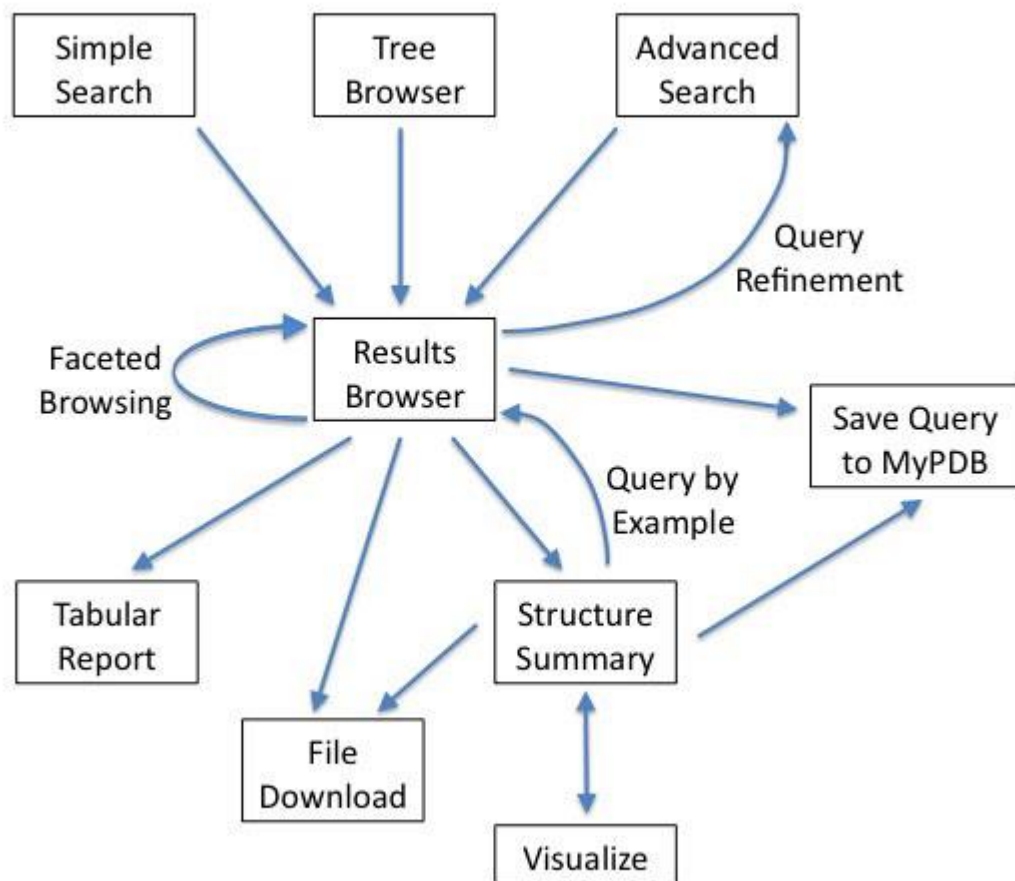
sugar-binding domains

Classified by Structure

Molecule: A
Polymer: 1
Chains: A
Fragment: S

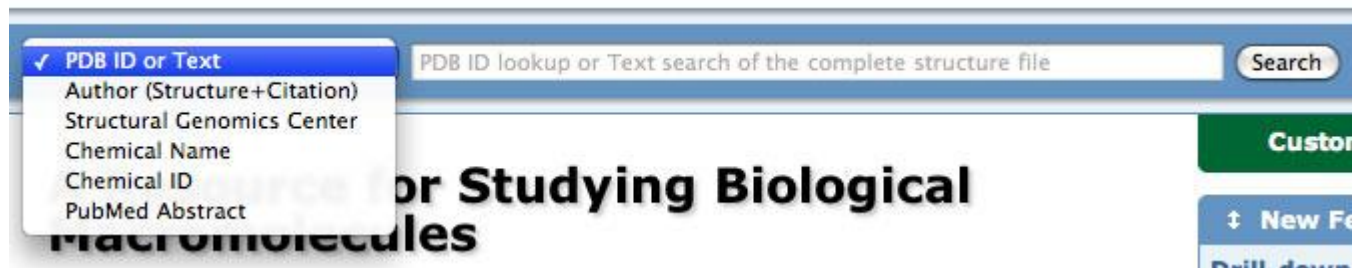
Copyright OpenHelix.

Hlavní složky RCSB PDB website



Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases
Nature Interaction Database doi:10.1038/pid.2011.1

Základní vyhledávání



- top navigation bar
- PDB ID (e.g., 4HHB)
- full-text search
- Výběr z
 - “Author”
 - “Makromolecule”
 - “Sequence”
 - “Ligand”

Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases
Nature Interaction Database doi:10.1038/pid.2011.1

Results Browser

- Nahoře:
 - unreleased entries,
 - citations, ligands
 - web pages.
- Rafinace vyhledávání
 - výběry dle typů
 - faceted browsing
 - advanced search
- Pracovní menu
 - file downloads,
 - tabular reports
 - sort
- List PDB IDs
 - (e.g., 2X51),
 - structure titles
 - thumbnail images

4207 Structure Hits218 Unreleased Structures1860 Citations1586 Ligand Hits96 Web Page Hits

Query Parameters:
Text Search for: KINASE

Query Refinements ? Show

Organism

Taxonomy

Exp. Method

X-Ray Resolution

Release Date

Polymer Type

Enzyme Classification

SCOP Classification

Refine Query

Remove Similar:

☒ Display/Download:

Generate Reports:

Sort by:

Results per Page:

Displaying results 1 - 25 of 4207 total | Page 1 of 169 | Jump to page: GO

☒ **2X51**

M6 DELTA INSERT1
Authors: [Squires, G.](#), [Houdusse, A.](#)
Release Date: 2011-01-26
Classification: [Motor Protein/signaling Protein](#)
Experiment: X-RAY DIFFRACTION with resolution of 2.20 Å
Compound: 2 Polymers [[Display Full Polymer Details](#) | [Display for All Results](#)]
3 Ligands [[Display Full Ligand Details](#) | [Display for All Results](#)]
Citation: Not Available.
Molecule of the Month: [Nitric Oxide Synthase](#), [Myosin](#), [Anthrax Toxin](#), [Calmodulin](#)

☒ **3KF9**

Crystal structure of the SdCen/skMLCK complex
Authors: [Radu, L.](#), [Assairi, L.](#), [Blouquit, Y.](#), [Durand, D.](#), [Miron, S.](#), [Charbonnier, J.B.](#), [Craescu, C.T.](#)
Release Date: 2011-01-26
Classification: [Cell Cycle/calcium Binding Protein](#)
Experiment: X-RAY DIFFRACTION with resolution of 2.60 Å
Compound: 2 Polymers [[Display Full Polymer Details](#) | [Display for All Results](#)]
1 Ligand [[Display Full Ligand Details](#) | [Display for All Results](#)]
Citation: Not Available.
Molecule of the Month: [Nitric Oxide Synthase](#), [Myosin](#), [Anthrax Toxin](#), [Calmodulin](#)

Refine search

- zpřesnění výběrových kritérií

4207 Structure Hits 218 Unreleased Structures 1860 Citations 1586 Ligand Hits 96 Web Page Hits

Query Parameters:
Text Search for: KINASE

Query Refinements ? Hide

 **Organism**

- Homo sapiens (man) (2230)
- Mus musculus (mouse) (202)
- Escherichia coli (184)
- Bos taurus (domestic cow) (125)
- Saccharomyces cerevisiae (yeast) (118)
- Rattus norvegicus (rats) (104)
- Mycobacterium tuberculosis (92)
- Other (1197)

 **Taxonomy**

- Eukaryota (3151)
- Bacteria (852)
- Archaea (115)
- Viruses (102)
- Other (11)
- Unassigned (7)

 **Experimental Method**

- X-RAY (3840)
- Solution NMR (366)
- Electron Microscopy (1)

 **X-Ray Resolution**

- less than 1.5 Å (130)
- 1.5 - 2.0 Å (1006)
- 2.0 - 2.5 Å (1543)
- 2.5 - 3.0 Å (915)
- 3.0 and more Å (246)
- more choices...

 **Release Date**

- before 2000 (370)
- 2000 - 2005 (802)
- 2005 - 2010 (2362)
- 2010 - today (673)
- this year (59)
- this month (59)
- more choices...

 **Polymer Type**

- Protein (4194)
- Mixed (12)
- RNA (1)

 **Enzyme Classification**

- 2: Transferases (3317)
- 3: Hydrolases (97)
- 4: Lyases (35)
- 5: Isomerases (27)
- 6: Ligases (15)
- 1: Oxidoreductases (4)

 **SCOP Classification**

- Alpha and beta proteins (a+b) (992)
- Alpha and beta proteins (a/b) (536)
- All beta proteins (260)
- All alpha proteins (251)
- Small proteins (75)
- Peptides (13)
- Multi-domain proteins (alpha an ... (12)
- Other (13)

 **Refine Query**

Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases
Nature Interaction Database doi:10.1038/pid.2011.1

Structure Summary

- Tabs – další odkazy na:
 - *Sequence* (+ secondary-structure annotations);
 - *Annotations* (external data from SCOP, CATH, GO, Pfam, Structural Biology Knowledgebase);
 - *Seq. Similarity* (sequence clusters);
 - *3D Similarity* (entries with similar 3D structures);
 - *Literature* (primary citation and related articles);
 - *Biol. & Chem.* (biological and chemical details);
 - *Methods* (experimental details on the structure determination method);
 - *Geometry* (bond lengths and angles, Ramachandran plot);
 - *Links* (hyperlinks to external resources with data related to the specific PDB entry).

[Summary](#)
[Sequence](#)
[Annotations](#)
[Seq. Similarity](#)
[3D Similarity](#)
[Literature](#)
[Biol. & Chem.](#)
[Methods](#)
[Geometry](#)
[Links](#)

CRYSTAL STRUCTURE OF THE CATALYTIC SUBUNIT OF CYCLIC ADENOSINE MONOPHOSPHATE-DEPENDENT PROTEIN KINASE

DOI:10.2210/pdb2cpk/pdb

2CPK

[Display Files](#)
[Download Files](#)
[Share this Page](#)

ENTRY 2CPK SUPERSEDES 1CPK

Primary Citation

Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.

Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., Sowadski, J.M.

Journal: (1991) *Science* **253**: 407-414

PubMed: [1862342](#)

Search Related Articles in PubMed

PubMed Abstract:

The crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase complexed with a 20-amino acid substrate analog inhibitor has been solved and partially refined at 2.7 Å resolution to an R factor of 0.212. The magnesium adenosine...

[\[Read More & Search PubMed Abstracts \]](#)

Related Citations in PDB Entry (REMARK 1)

Hide

Structure of a Peptide Inhibitor Bound to the Catalytic Subunit of Cyclic Adenosine Monophosphate-Dependent Protein Kinase

Knighton, D.R., Zheng, J., Teneyck, L.F., Xuong, N.-H., Taylor, S.S., Sowadski, J.M.

(1991) *Science* **253**: 414

Expression of the Catalytic Subunit of C/AMP-Dependent Protein Kinase in Escherichia Coli

Slice, L.W., Taylor, S.S.

(1989) *J.Biol.Chem.* **264**: 20940

Molecular Description

Hide

Classification: [Transferase\(phosphotransferase\)](#)

Structure Weight: 42964.13

Molecule: cAMP-DEPENDENT PROTEIN KINASE, CATALYTIC SUBUNIT

Polymer: 1 **Type:** polypeptide(L) **Length:** 350

Chains: E

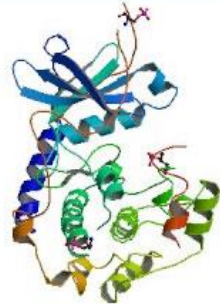
EC#: [2.7.1.37](#)

Molecule: PEPTIDE INHIBITOR 20-MER

Polymer: 2 **Type:** polypeptide(L) **Length:** 20

Chains: I

Biological Assembly



[More Images...](#)

[View in Jmol](#) [SimpleViewer](#)

[Other Viewers](#) [Protein Workshop](#)

Biological assembly assigned by authors and generated by PISA (software)

MyPDB Personal Annotations

Hide

To save personal annotations, please [login](#) to your MyPDB account.

Deposition Summary

Hide

Authors: Knighton, D.R., Zheng, J., Teneyck, L.F., Ashford, V.A., Xuong, N.-H., Taylor, S.S., Sowadski, J.M.

Deposition: 1992-10-21

Release: 1993-01-15

Last Modified (REVDAT): 2009-02-24

Previous versions: [1CPK](#)

Experimental Details

Hide

Advanced search

- Vyhledávání dle:
 - text search
 - structure type
 - sequence motifs
 - ligand
 - biologická data (EC, TC, GO)
 - experimentální metoda
 - publikace
- Result count
- Kombinování kritérií

The screenshot displays the 'Advanced Search Interface' with two search criteria sections. The first section, 'Text Search', has a dropdown menu set to 'Text Search' and a text input field containing 'KINASE'. To the right of the input field, a 'Result Count' button shows '4207 PDB Entries'. The second section, 'Citation', has a dropdown menu set to 'Citation'. Below it, a note reads: 'Search by citation records. Uncheck 'Primary citation only' to include secondary citations.' The 'Citation' section includes several input fields: 'Author Name' (empty), 'Title' (with a 'Contains:' dropdown and an empty input field), 'Primary citation only' (checked), 'Journal' (with a 'Nature' dropdown), and 'Year' (with an '=' dropdown and an empty input field). To the right of these fields, a 'Result Count' button shows '1711 PDB Entries'. At the bottom of the interface, there is a 'Remove Similar Sequences at' dropdown set to '90%' and an 'Identity' button with a question mark. Below this, a 'Match' dropdown is set to 'all' with the text 'of the above conditions.' To the right of the 'Match' dropdown are two buttons: 'Clear All Parameters' and 'Submit Query'.

Report

- Tabulka s výsledky

Citation Report Total of 80 results.

Click on column headers to sort up/down. Click again to reverse order. Download options:  Type value in text boxes under column headers to filter the data set. 

PDB ID	Title	Primary Citation Author	Pub. Year	Journal Name	Volume	First Page	PubMed ID
13PK	Synergistic effects of su	Bernstein, B.E., Michels, P.A., Hol,	1997	Nature	385	275	9000079
1AD5	Crystal structure of the	Sicheri, F., Moarefi, I., Kuriyan, J.	1997	Nature	385	602	9024658
1AM0	Structural Basis of RNA	Jiang, F., Kumar, R.A., Jones, R.A.	1996	Nature	382	183	8700212
1AP7	Structure of the cyclin	Luh, F.Y., Archer, S.J., Domaille, P.	1997	Nature	389	999	9353127
1AWJ	Regulatory intramolecu	Andreotti, A.H., Bunnell, S.C., Fen	1997	Nature	385	93	8985255
1B47	Structure of the amino	Meng, W., Sawadkiosol, S., Burak	1999	Nature	398	84	10078535
1B89	Clathrin self-assembly i	Ybe, J.A., Brodsky, F.M., Hofmann	1999	Nature	399	371	10360576
1B17	Structural basis for inh	Russo, A.A., Tong, L., Lee, J.O., Je	1998	Nature	395	237	9751050
1B18	Structural basis for inh	Russo, A.A., Tong, L., Lee, J.O., Je	1998	Nature	395	237	9751050
1BLX	Crystal structure of the	Brotherton, D.H., Dhanaraj, V., Wi	1998	Nature	395	244	9751051
1BXD	NMR structure of the h	Tanaka, T., Saha, S.K., Tomomori,	1998	Nature	396	88	9817206
1C1Y	The 2.2 Å crystal struct	Nassar, N.	1995	Nature	375	554	7791872
1CF4	Structure of the small G	Mott, H.R., Owen, D., Nietispach,	1999	Nature	399	384	10360579
1CRK	Structure of mitochond	Fritz-Wolf, K., Schnyder, T., Wallin	1996	Nature	381	341	8692275
1E00	Crystal Structure of Fib	Pellegrini, L., Burke, D.F., Von De	2000	Nature	407	1029	11069186
1ERK	Atomic structure of the	Zhang, F., Strand, A., Robbins, D.,	1994	Nature	367	704	8107865
1FIN	Mechanism of CDK acti	Jeffrey, P.D., Russo, A.A., Polyak,	1995	Nature	376	313	7630397
1FMK	Three-dimensional stru	Xu, W., Harrison, S.C., Eck, M.J.	1997	Nature	385	595	9024657
1IRK	Crystal structure of the	Hubbard, S.R., Wei, L., Ellis, L., H	1994	Nature	372	746	7997262
1JKY	Crystal structure of the	Pannifer, A.D., Wong, T.Y., Schwa	2001	Nature	414	229	11700563

Filter Results Reload Results Customize Columns Page 1 of 5 View 1 - 20 of 80

☒ Display/Download:  ☒ Generate Reports:

- Custom Reports
 - Image Collage
 - Customizable Table
- Summary Reports
 - Structure
 - Sequence
 - Ligands
 - Structural Genomics Center
 - Primary Citation
 - Biological Details
- Experimental Reports
 - X-ray
 - Crystallization
 - Data Collection
 - Refinement
 - Refinement Parameters
 - Unit Cell
 - NMR
 - Representative Model
 - Spectrometer
 - Sample Conditions
 - Software
 - NMR Refinement
 - NMR Ensemble

3PE4 Structure substrate
 Authors:
 Release Date:
 Experiment:
 Compound:
 Citation:

3PCR Structure
 Authors:
 Release Date:
 Experiment: X-RAY DIFFRACTION with res

Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases
Nature Interaction Database doi:10.1038/pid.2011.1

myPDB



Welcome Primer!

- [Saved Query Manager](#)
- [Personal Annotations](#)
- [Display Settings](#)
- [User Account](#)
- [Logout](#)

Query Manager

Your saved queries are listed below. You can assign queries to be executed during each weekly update of the RCSB PDB. Click on email notification to turn automatic notification on or off. Notifications about new structures matching your queries are sent on either a weekly or monthly schedule based on your account settings. Email notifications are only sent if new structures match your queries.

The query name and query description are generated automatically. Click on the name or description to edit the text. Click the save button that appears after you have made your changes.

Name	Query Description	Email Notification	Last Run	Next Scheduled Run	Run Query (No Email)	Delete
User Query	Text Search for: KINASE	Yes	01/27/2011	02/01/2011		

Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases
Nature Interaction Database doi:10.1038/pid.2011.1

Access to More

**Authors,
Release Date,
Classification,
Experiment,
Compound &
Citation**

Authors: Consani Textor, L. , Holton, S. , Wilmanns, M. 
Release Date: 2010-12-08  **Classification:** Transcription 
Experiment: X-RAY DIFFRACTION with resolution of 2.90 Å
Compound: 3 Polymer [\[Display Full Polymer Details | Display for All Results \]](#)
 1 Ligand [\[Display Full Ligand Details | Display for All Results \]](#)
Citation: Not Available.

polymerase-Gfh1 complex (Crystal type 1)

Structure Sum

- One for every structure
- Organized similarly
- It provides access to a wealth of data
- Customizable with a widget framework
- Re-arranged sections in red
- Use Reset Layout to return to default

Report tabs
Title & ID

Primary Citation

Molecular Description

Ligand Component

Source

Related Entries

External Data (orange)

Experimental Details

MyPDB

Deposition Summary

Images & Viewing Options

Menus

Reset Layout

Sequence Report

- Graphic of secondary structures
- Choose domain assignments or annotations from dropdown menu
- Specify how you want to view graphic

Set viewing preferences

References below

The screenshot shows the RCSB PDB sequence report for entry 2ARC. The 'Sequence' tab is selected in the top navigation bar. The page title is 'ESCHERICHIA COLI REGULATORY PROTEIN ARABINOSIDE OPERON REGULATORY PROTEIN 2ARC'. The 'Results Tabs' section includes 'Sequence / Structure Details', 'Redundancy Reduction and Sequence Clustering', and 'Sequence Display'. The 'Sequence Display' section states: 'The structure 2ARC has in total 2 chains. Out of these 1 are sequence-unique. Currently viewing unique chains only. [show all chains] [Show 3D in Jmol]'. The 'Chain Display' section shows 'Chain A (polymer 1) [help] [fasta] [sequence & DSSP]'. The 'Description' is 'ARABINOSIDE OPERON REGULATORY PROTEIN'. The 'Identical chains' are 'B'. The 'Chain Type' is 'polypeptide(L)'. The 'UniProtKB reference' is 'P0A9E0'. The 'Select' dropdown menu is open, showing options: 'domain assignment from sequence', 'pfam', 'interpro', and 'secondary structure'. The 'Preferences' section includes 'Jmol - 3D parameters' (show Jmol: No, Jmol width: 450, Jmol height: 450) and 'Page parameters' (Sequence to view: PDB SEQRES sequence, Label residue ids by: ATOM record only, Font size: 12, Residues per row: 60, Chains to view: Unique chains (1 chains)). A 'Submit' button is at the bottom.

Sequence

ESCHERICHIA COLI REGULATORY PROTEIN ARABINOSIDE OPERON REGULATORY PROTEIN 2ARC

Display Files
Download Files
Share this Page

Sequence / Structure Details

Redundancy Reduction and Sequence Clustering

View the clustering results for 2ARC.

Sequence Display ⓘ

The structure 2ARC has in total 2 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] [sequence & DSSP]

Description ARABINOSIDE OPERON REGULATORY PROTEIN

Identical chains B

[show all chains]

Chain Type polypeptide(L)

UniProtKB reference P0A9E0

Select

domain assignment from sequence

pfam

interpro

secondary structure

Preferences

Jmol - 3D parameters

show Jmol No

Jmol width 450 Jmol height 450

Page parameters

Sequence to view PDB SEQRES sequence Label residue ids by ATOM record only

Font size 12 Residues per row 60

Chains to view Unique chains (1 chains)

Submit

C

No use or reproduction
without express written

Annotations Report

ESCHERICHIA COLI REGULATORY COMPLEXED WITH L-ARABINOSE

2ARC

Database links

Display Files
Download Files
Share this Page

Annotation data related to this entry.

SCOP

Domain Annotation: SCOP Classification (version 1.75)						
Domain Info	Class	Fold	Superfamily	Family	Domain	Species
d2arca_	All beta proteins	Double-stranded beta-helix	Regulatory protein AraC	Regulatory protein AraC	Regulatory protein AraC	Escherichia coli [TaxId: 562]
d2arcb_	All beta proteins	Double-stranded beta-helix	Regulatory protein AraC	Regulatory protein AraC	Regulatory protein AraC	Escherichia coli [TaxId: 562]

CATH

Domain Annotation: CATH Classification (version 3.3.0)				
Domain	Class	Architecture	Topology	Homology
2arcA00	Mainly Beta	Sandwich	Jelly Rolls	
2arcB00	Mainly Beta	Sandwich		

PFAM

Protein Family Annotation: PFAM Classification			
Chain	PFAM Accession	PFAM ID	Description
A	PF02311	AraC_binding	AraC-like ligand binding domain
B	PF02311	AraC_binding	AraC-like ligand binding domain

SBKB

Structural Biology Knowledgebase Data

Information from the Structural Biology Knowledgebase

- Models from the Protein Model Portal: 299 models
- Protein Targets from TargetDB: 0 targets
- Related Biological Annotations: >14 annotations
- Related Clones from PSI:Biological Materials Repository: 0 clones
- Related Protocols from PepcDB: 0 protocols

Customize

25

1169 Structure Hits 588 Citations 313 Ligand Hits GO Hits SCOP Hits CATH Hits

Query Parameters:
CathTree Search for Jelly Rolls (2.60.120)

Query Refinements ?

Organism Taxonomy Exp. Method X-Ray Resolution

Enzyme Classification SCOP Classification

Refine Query

Display/Download: Generate Reports: Sort by

Displaying results 1 - 25 of 1169 total | Page 1

3D4K Concanavalin A Complexed to a Synthetic Analog of the

Sequence Similarity Report

[Summary](#) [Sequence](#) [Annotations](#) [Seq. Similarity](#) [3D Similarity](#) [Literature](#) [Bion. & Chem.](#) [Methods](#) [Geometry](#) [Links](#)

ESCHERICHIA COLI REGULATORY PROTEIN ARAC COMPLEXED WITH L-ARABINOSE

2ARC

[Display Files](#) [Download Files](#) [Share this Page](#)

Sequence Clustering and Redundancy Reduction Results

Cluster # 3319 Members. Sequence cutoff: 95% Cluster size: 4 sequences

Sequence Clusters	Rank ?	PDB id	Entity id	Chains	Description	Details	EC Number	Taxonomy
Entity #1: Chains: 1								
Cluster Sequence Similarity								
100%	☐	Download	Details			— Select Comparison Method — Submit ?		
95%	☐ 1	2ARC	1	A, B	ARABINOSE OPERON REGULATORY PROTEIN	SUGAR-BINDING/DIMERIZATION DOMAIN		562
90%								
70%	☐ 2	2AAC	1	A, B	ARAC	SUGAR-		562
50%								
40%								
30%								

Documentation [Click here for more detailed d](#)

Redundancy in the Protein Data Bank

As the single worldwide repository for macromolecular structures, the Protein Data Bank holds a body of data that contains considerable redundancy in regard to both sequence and structure. We have incorporated into the query interface the ability to select a subset of structures from which similar sequences have been largely removed. In most cases, the selected subset will contain far fewer structures than the complete result set. However, the following caveats should be kept in mind:

- Sequence similarity is defined on a chain basis, but results are returned on a structure basis.

Statistics

The following table shows the number of non-redundant sequences as determined by blastclust at several levels of sequence identity.

Method	Description	# of Clusters
blast	100% identity	40231
blast	95% identity	28822
blast	90% identity	27708

Biology & Chemistry Report

Summary Sequence Information Sequence Similarity 3D Similarity Literature **Biology & Chemistry** Methods Geometry Links

ESCHERICHIA COLI REGULATORY PROTEIN ARAC COMPLEXED WITH L-ARABINOSE **2ARC**

Display Files ▾
Download Files ▾
Share this Page ▾

Biology and Chemistry Report

↑ Structure Details ? **Help** Hide

Structure Keywords

Keywords TRANS
Text TRANS

Polymeric Molecule

Chain A,B

Description
Fragment
Nonstandard Linkage
Nonstandard Monomers
Polymer Type
Formula Weight
Source Method
Entity Name

↑ Protein Details ? **Help** Hide

UniProtKB Information

Chain	SWS/UNP ID	SWS/UNP Accession(s)
A,B	ARAC_ECOLI	P0A9E0

Keywords and Names

Chain(s)	RCSB Name	UniProtKB Name	UniProtKB Keywords
A,B	ARABINOSE OPERON REGULATORY PROTEIN	Arabinose operon regulatory protein	3D-structure , Activator , Arabinose catabolism , Carbohydrate metabolism , Complete proteome , Cytoplasm , DNA-binding , Repressor , Transcription , Transcription regulation

↑ Gene Details ? **Help** Hide

Genetic Source

Chain A,B	Scientific Name	Genus	Host Scientific Name	Host Genus
A,B	Scientific name of the organism.	ESCHERICHIA COLI	Escherichia coli	Escherichia
		HOMO SAPIENS	Escherichia coli	Escherichia
		SACCHAROMYCES CEREVISIAE		

Ligands and Prosthetic Groups

ID	Name
ARA	ALPHA-L-ARABINOSE

Mouse over for details

Methods Report

Summary | Sequence | Annotation | Secondary Structure | Literature | Bibliography | **Methods** | Geometry | Links

X-RAY DIFFRACTION **2ARC** [Display Files](#) [Download Files](#) [Share this Page](#)

Materials and Methods page

Crystallization Hide

Crystallization Experiments

pH 7.25

Details PROTEIN WAS CRYSTALLIZED BY MICROSEEDING FROM 18-20% PEG 8000, 100 MM TRIS-HCL, PH 7.25, 40 MM MAGNESIUM ACETATE AND 0.2% (W/V) L-ARABINOSE

Crystal Data Hide

Unit Cell

Length	(Å)	Angle	(°)
a =	39.75	α =	90
b =	93.84	β =	95.62
c =	50.33	γ =	90

Space Group

Space Group Name: Hermann-Mauguin space-group symbol. Note that the Hermann-Mauguin symbol does not necessarily contain complete information about the symmetry and the space-group origin. If used, always supply the FULL symbol from International Tables for Crystallography Vol. A (2002) and indicate the origin and the setting if it is not implicit. If there is any doubt that the equivalent positions can be uniquely deduced from this symbol, specify the attribute pos_as_xyz in category symmetry_equiv or attribute space_group_name_Hall in category symmetry data items as well. Leave spaces between symbols referring to different axes.
P 1 21/m 1
P 2/n 2/n 2/n (origin at -1)
R -3 2/m

Diffraction Hide

Diffraction Experiment

Diffn ID

Data Collection Temperature

Diffraction Detector

Detector IMAGE PLATE

Type FUJI

Details SPHERICAL RH COATED MIRROR

Collection Date 1995-09

- Precise experimental details
- Organized similarly
- Mouse over to read definitions
- Sections depend on methodology

Educational Resources

[Contact Us](#) | [Print](#)

PDB ID or T

Understanding PDB Data: Looking at Structures

The PDB archive is a repository of atomic coordinates and other information describing proteins and other important biological macromolecules. Structural biologists use methods such as **X-ray crystallography**, **NMR spectroscopy**, and **cryo-electron microscopy** to determine the location of each atom relative to each other in the molecule. They then deposit this information, which is then annotated and publicly released into the archive by

Looking at Structures

- Introduction
- Biological Assemblies
- Dealing with Coordinates
- Methods for Determining Structure
- Missing Coordinates and Biological Assemblies
- Molecular Graphics Programs
- Resolution
- R-value and R-free

Educational Resources

[Kiosk Viewer](#) | [Posters](#) | [Website Tutorials](#) | [Animations](#) | [Activities/Lessons](#) | [Exhibit Materials](#) | [Events](#) | [Poster Prize](#)

These Educational Resources are aimed at teachers and students interested in learning about protein and nucleic acid structures, and how to use the RCSB PDB site ([PDF Overview](#)). For researchers and educators interested in how to understand PDB data, how to visualize structures, how to read coordinate files, and potential challenges in exploring the archive, please see the [Looking at Structures](#) feature.

Molecules in Motion Kiosk Viewer



The **Molecules in Motion Kiosk Viewer** is a full-screen animation program that displays structures from different angles and perspectives, and focuses on chemical components within the structure (**PDF flyer**).

The Kiosk program runs on Mac, Windows and some versions of Linux (i.e. CentOS).

**Click on
Molecule
of the
Month**

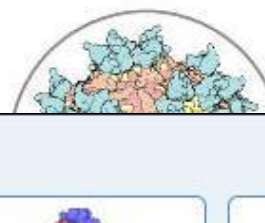
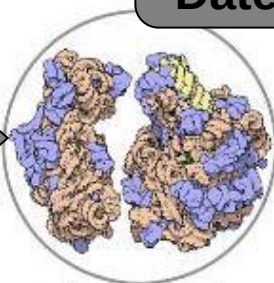
Copyright OpenHelix
No use or reproduction
without express written

Molecule of the Month: Structural View of Biology

View by Title,
Date or Category

List View of Archive By: [Title](#) | [Date](#) | [Category](#)

Click on



Transcription



Enhanceosome



Oct and Sox
Transcription



Poly(A)
Polymerase

Protein Synthesis

The major molecules of protein synthesis, ribosomes to folded proteins, are available

Structure of Nucleic Acids

DNA Maintenance

Replication

Transcription

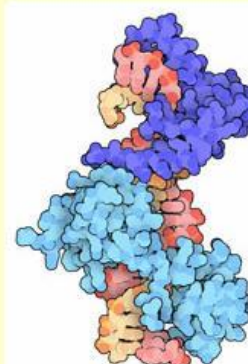
Translation

Protein Folding - Modification - D

The RCSB PDB's Molecule of the Month series by OpenHelix. The model is for all types of users, but especially for those who are interested in the structure and function of the molecule. a discussion of the structure and function of the molecule. a discussion of the structure and function of the molecule.

Copyright OpenHelix.
No use or reproduction
without express written

Oct and Sox Transcription Factors



The development of a complete human being from a single cell is one of the great miracles of life. A human egg cell contains about 30,000 genes that encode proteins, and of these, about 3,000 of these genes encode transcription factors. Transcription factors determine when genes will be turned on and turned off, orchestrating the many processes involved in the development of an embryo and the many tasks performed by each cell after a child is born. Amazingly, there is only about 1

transcription factor for every 10 genes, posing a puzzle: how does this limited set of proteins control the many genes and processes that must be regulated?

[Click to Read More.](#)

A Protein Information Goldmine

RCSB PDB
PROTEIN DATA BANK

A MEMBER OF THE **PDDB**
Portal to Biological Macromolecular Structures
PST there are 69967 Structures | PDB Statistics

Contact Us | Print

structure file Search Advanced Search

MyPDB Hide

Summary Sequence Annotations Seq. Similarity 3D Similarity Literature Biol. & Chem. Me

Chain A (polymer 1) [help] [fasta] [sequence & DSSP]

Description: ARABINOSE OPERON REGULATORY PROTEIN

Identical chains: B

[show all chains]

Chain Type: polypeptide(L)

UniProtKB reference: P0A9E0

Length: 164 residues

scop domain assignment: d2arca Regulatory protein AraC: 161 residues

[hide] [reference]

dssp secondary structure: 25% helical (3 helices; 41 residues)
33% beta sheet (10 strands; 55 residues)

More annotations: Select

Currently displayed: SEQRES sequence. [display external (UniProtKB) sequence]

Sequence Details [View image]

scop: Regulatory protein AraC (d2arca)

dssp: [Diagram showing secondary structure elements]

PDB: D P L L P G Y S F N A H L V A G L T P I E A N G Y L D F F I D R P L G M K G Y I L N L T I R G Q G V V K N Q G R E F V C

PDB: 7 10 20 30 40 50 60 66

scop: Regulatory protein AraC (d2arca)

dssp: [Diagram showing secondary structure elements]

PDB: R P G D I L L F P P G E I H H Y G R H P E A R E W Y H Q W V Y F R P R A Y W H E W L N W P S I F A N T G F F R P D E A H

PDB: 67 70 80 90 100 110 120 126

scop: Regulatory protein AraC (d2arca)

dssp: [Diagram showing secondary structure elements]

PDB: Q P H F S D L F G Q I I N A G Q G E G R Y S E L L A I N L L E Q L L R R M E A I N E S

PDB: 127 140 150 160

2ARC



Quality of data in PDB – why validate?

Structural biology community found that some published structures contained serious errors

Nightmare before Christmas



Retraction

WE WISH TO RETRACT OUR RESEARCH ARTICLE "STRUCTURE OF MdhA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporter" and both of our Reports "Structure of the ABC transporter MdhA in complex with ADP vanadate and lipopolysaccharide" and "X-ray structure of the EmrE multidrug transporter in complex with a substrate" (1–3).

The recently reported structure of Sav1866 (4) indicated that our MdhA structures (1, 2, 3) were incorrect in both the hand of the structure and the topology. Thus, our biological interpretations based on these inverted models for MdhA are invalid.

As in-house data reduction programs introduced a change in sign for anomalous differences. This program, which was not part of a conventional data processing package, converted the anomalous pairs (h⁺ and l[−]) to (h[−] and l⁺), thereby introducing a sign change. As the diffraction data collected for each set of MdhA crystals and for the EmrE crystals were processed with the same program, the structures reported in (1–3, 5, 6) had the wrong hand.

The error in the topology of the original MdhA structure was a consequence of the low resolution of the data as well as breaks in the elec-

tron density for the connecting loop regions. Unfortunately, the use of the multiplicity refinement procedure still allowed us to obtain reasonable refinement values for the wrong structures.

The Protein Data Bank (PDB) files 1HSQ, 1PF4, and 1ZZR for MdhA and 1S7H and 2F2M for EmrE have been moved to the archive of obsolete PDB entries. The MdhA and EmrE structures will be recalculated from the original data using the proper sign for the anomalous differences, and the new Cα coordinates and structure factors will be deposited.

We very sincerely regret the confusion that these papers have caused and, in particular, subsequent research efforts that were unproductive as a result of our original findings.

GEORFREY CHENG, CHRISTOPHER B. ROSE,
CHRISTOPHER L. REYES, OWEN FORNAGIS,
YIN-JU CHEN, ANDY K. CHEN

Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

References

1. G. Cheng, C. B. Rose, *Science* **299**, 1770 (2002).
2. C. L. Reyes, G. Cheng, *Science* **300**, 3028 (2002).
3. D. Poudyal, Y. J. Chen, A. P. Chen, G. Cheng, *Science* **310**, 2940 (2005).
4. W. J. Deane, K. P. Locher, *Nature* **443**, 188 (2004).
5. G. Cheng, *J. Mol. Biol.* **309**, 409 (2001).
6. C. Ma, G. Cheng, *Proc. Natl. Acad. Sci. USA* **101**, 2652 (2004).

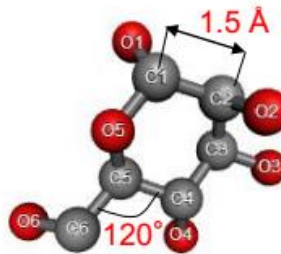
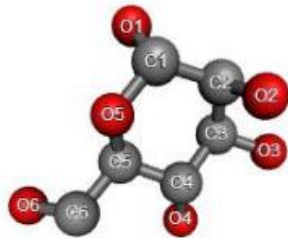
What to validate?

How?

Software

Geometry (3D)

Against a template

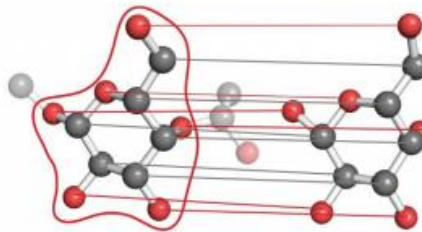
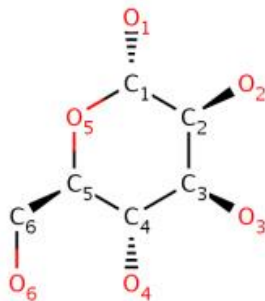


Proteins and nucl. acids:
WHAT_CHECK, PROCHECK,
PROCHECK-NMR, AQUA,
MolProbity, OOPS

Ligands:
ValLigURL, Mogul, Coot, PHENIX

Topology (2D)

Against tabular values



Proteins and nucl. acids:
-||-

Ligands:
PDB Care, MotiveValidator*,
ValidatorDB**

*Sehna, Svobodová et al., **Nucl. Acids Res.** (2014)

** Svobodová. Jaiswal et al.. **Nucl. Acids Res.** (2015)

Quality of data in PDB

- Validation of ligands



Estimations:

- 20-30% of ligands contain errors**

Methodology:

- Topology validation
 - All ligands in PDB
 - Important ligand classes
- Template molecules from wwPDB CCD

Sustainability of results:

- Weekly updated database ValidatorDB***

* Lütke et al., Carbohydr. Res. (2004)

** Liebeschuetz et al., J. Comput. Aided Mol. Des. (2012)

*** Svobodová, Jaiswal et al., Nucl. Acids Res. (2015)



not all ligands are equally wrong

- 83% of ligands are OK, 9% incomplete, 8% with chirality errors
- Better quality: Drug molecules
- Problematic quality: Carbohydrates, organometals

	All molecules	Polycyclic	Carbo- hydrates	Mannose derivates	Organo- metals	Experi- mental drugs	Approved drugs
Number of validated molecules	238153	6804	57302	6341	22600	37450	1934
Incomplete	8.9%	6.7%	5.9%	3.5%	18.0%	6.1%	3.2%
Missing only atoms	5.9%	3.1%	4.2%	3.0%	6.0%	5.0%	0.9%
Missing rings	2.6%	3.0%	1.5%	0.1%	10.7%	0.6%	2.0%
Degenerate	0.5%	0.6%	0.2%	0.4%	1.4%	0.5%	0.3%
Wrong chirality	7.9%	5.5%	4.0%	7.6%	16.5%	2.1%	2.8%
Wrong C chirality	2.4%	3.5%	4.0%	7.4%	2.5%	1.7%	2.8%
Wrong Metal chirality	1.4%	0.0%	0.0%	0.0%	14.3%	0.0%	0.0%
Wrong High order chirality	4.3%	1.9%	0.0%	0.2%	0.0%	0.4%	0.0%
Wrong Planar chirality	1.1%	0.0%	0.0%	0.0%	10.5%	0.1%	0.8%

> 2 times better

> 30% better

> 30% worse

> 2 times worse

Další strukturní databáze

- **MSD:** The Macromolecular Structure Database
 - A relational database representation of clean Protein Data Bank (PDB)
- **3DSeq:** 3D sequence alignment server
 - Annotation of the alignments between sequence database and the PDB
- **FSSP:**
 - Based on exhaustive all-against-all 3D structure comparison of protein structures currently in the Protein Data Bank (PDB)
- **DALI:**
 - Fold Classification based on Structure-Structure Assignments
- **3Dee:**
 - Database of protein domain definitions wherein the domains have been clustered on sequence and structural similarity
- **NDB:** Nucleic Acid Structure Database
 - database of NA structures

by

Jaroslav Koča, Radka Svobodová Vařeková, Lukáš Pravda, Karel Berka, David Sehnal, Stanislav Geidl and Michal Otyepka

Structural Bioinformatics Databases of General Use	21
Karel Berka	
3.1 How a Biomacromolecule Looks Codes What It Does	21
3.2 Worldwide Protein Data Bank (PDB) – essential structure repository	23
3.2.1 Protein Data Bank in Europe (PDBe)	24
3.2.2 RCSB PDB	26
3.3 Other notable databases	27
3.3.1 PDBsum – Pictorial view on PDB database	27
3.3.2 PDB_REDO and WHY_NOT databases for curated structures	27
3.3.3 CATH and Pfam databases for classification of protein folds and sequences	27
3.3.4 PDBFlex, Pocketome and PED3 databases to analyze protein flexibility and disorder	28
3.3.5 OPM and MemProtMD databases for membrane protein	29
3.3.6 NDB and GFDB databases for other macromolecules	29
3.3.7 UniProt and ChEMBL databases – power of connection	30
3.4 Conclusion	30
3.5 Exercises	31
3.5.1 Use of PDBe	31
3.5.2 Use of RCSB and ChEMBL	31
3.5.3 Use of PDBsum	32
3.5.4 Use of CATH	32
References	33

PDBe-KB

The screenshot shows a web browser window with the PDBe-KB website. The browser's address bar shows the URL `ebi.ac.uk/pdbe/pdbe-kb/protein`. The website's header includes navigation links for EMBL-EBI, Services, Research, Training, About us, and a search bar. Below the header, the PDBe-KB logo is displayed, followed by a search bar with the example text "Ex. - hemoglobin, BRCA1_HUMAN". A navigation bar contains links for Home, Aggregated Views of Proteins, PDBe Graph API, PDBe Graph Database, Partners, Guidelines, Contact, and Feedback. The main content area features the heading "What are the Aggregated Views of Proteins?" in green. Below this heading are four icons representing different data types: Structures, Small-molecules, Macromolecular Interactions, and Functional Annotations. A paragraph explains that the aggregated view provides a comprehensive overview of structural data available in PDB for a full-length protein sequence, including UniProt identifiers, PDB entries, small-molecules, macromolecular interaction sites, and functional annotations. A link "Learn more about PDBe-KB..." is provided. Below the paragraph is a section titled "Explore PDBe-KB Aggregated Views of Proteins" with a search bar and examples: "2etx" and "Q14676". The browser's taskbar at the bottom shows the Windows logo, a search bar, and various application icons.

AlphaFold Protein Structure Data x PDBe - Knowledge Base x PDBe-KB Protein Pages x +

ebi.ac.uk/pdbe/pdbe-kb/protein

EMBL-EBI Services Research Training About us EMBL-EBI

PDBe-KB
Protein Data Bank in Europe Knowledge Base

Ex. - hemoglobin, BRCA1_HUMAN

Home Aggregated Views of Proteins PDBe Graph API PDBe Graph Database Partners Guidelines Contact Feedback

What are the Aggregated Views of Proteins?

Structures Small-molecules Macromolecular Interactions Functional Annotations

The PDBe-KB aggregated view of proteins provides a comprehensive overview of structural data available in PDB for a full-length protein sequence. Either PDB or UniProt identifiers can be used to display all the available data, and in particular all the PDB entries related to the protein sequence, all the observed small-molecules interacting with the protein, all the macromolecular interaction sites and partners, and additional functional annotations, such as sequence conservation and druggable sites. [Learn more about PDBe-KB...](#)

Explore PDBe-KB Aggregated Views of Proteins

Search by PDB accession or UniProt accession...

Examples: [2etx](#) [Q14676](#)

Sem zadejte hledaný výraz

14:55 04.10.2021

PDBe-KB

- Focused on individual proteins
 - UniprotID is key
- Gathers data across individual PDBID and individual resources (e.g. ChannelsDB)

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism

BETA

Search

Examples: [Free fatty acid receptor 2](#) [At1g58602](#) [Q5VSL9](#) [E. coli](#) [Help: AlphaFold DB search help](#)

AlphaFold DB provides open access to protein structure predictions for the human proteome and 20 other key organisms to accelerate scientific research.

<https://www.alphafold.ebi.ac.uk/>

Complete structures of 20 model organisms

Species	Common Name	Reference Proteome	Predicted Structures	Download
<i>Arabidopsis thaliana</i>	<i>Arabidopsis</i>	UP000006548 	27,434	Download (3642 MB)
<i>Caenorhabditis elegans</i>	Nematode worm	UP000001940 	19,694	Download (2601 MB)
<i>Candida albicans</i>	<i>C. albicans</i>	UP000000559 	5,974	Download (965 MB)
<i>Danio rerio</i>	Zebrafish	UP000000437 	24,664	Download (4141 MB)
<i>Dictyostelium discoideum</i>	<i>Dictyostelium</i>	UP000002195 	12,622	Download (2150 MB)
<i>Drosophila melanogaster</i>	Fruit fly	UP000000803 	13,458	Download (2174 MB)
<i>Escherichia coli</i>	<i>E. coli</i>	UP000000625 	4,363	Download (448 MB)
<i>Glycine max</i>	Soybean	UP000008827 	55,799	Download (7142 MB)
<i>Homo sapiens</i>	Human	UP000005640 	23,391	Download (4784 MB)
<i>Leishmania infantum</i>	<i>L. infantum</i>	UP000008153 	7,924	Download (1481 MB)
<i>Methanocaldococcus jannaschii</i>	<i>M. jannaschii</i>	UP000000805 	1,773	Download (171 MB)
<i>Mus musculus</i>	Mouse	UP000000589 	21,615	Download (3547 MB)
<i>Mycobacterium tuberculosis</i>	<i>M. tuberculosis</i>	UP000001584 	3,988	Download (421 MB)
<i>Oryza sativa</i>	Asian rice	UP000059680 	43,649	Download (4416 MB)
<i>Plasmodium falciparum</i>	<i>P. falciparum</i>	UP000001450 	5,187	Download (1132 MB)
<i>Rattus norvegicus</i>	Rat	UP000002494 	21,272	Download (3404 MB)
<i>Saccharomyces cerevisiae</i>	Budding yeast	UP000002311 	6,040	Download (960 MB)
<i>Schizosaccharomyces pombe</i>	Fission yeast	UP000002485 	5,128	Download (776 MB)
<i>Staphylococcus aureus</i>	<i>S. aureus</i>	UP000008816 	2,888	Download (268 MB)
<i>Trypanosoma cruzi</i>	<i>T. cruzi</i>	UP000002296 	19,036	Download (2905 MB)

elixir
CZECH
REPUBLIC

Download

mmCIF file

Predicted aligned error

4

Biological function (Microbial infection) Proposed to be involved in transcriptional activation by EBV EBNA2 of CBF-1/RBPJ-repressed promoters. [go to UniProt](#)

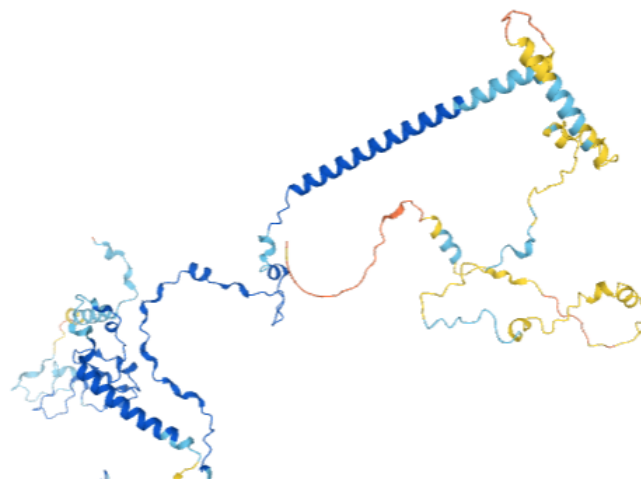
Model Confidence:

Very low (pLDDT < 50)

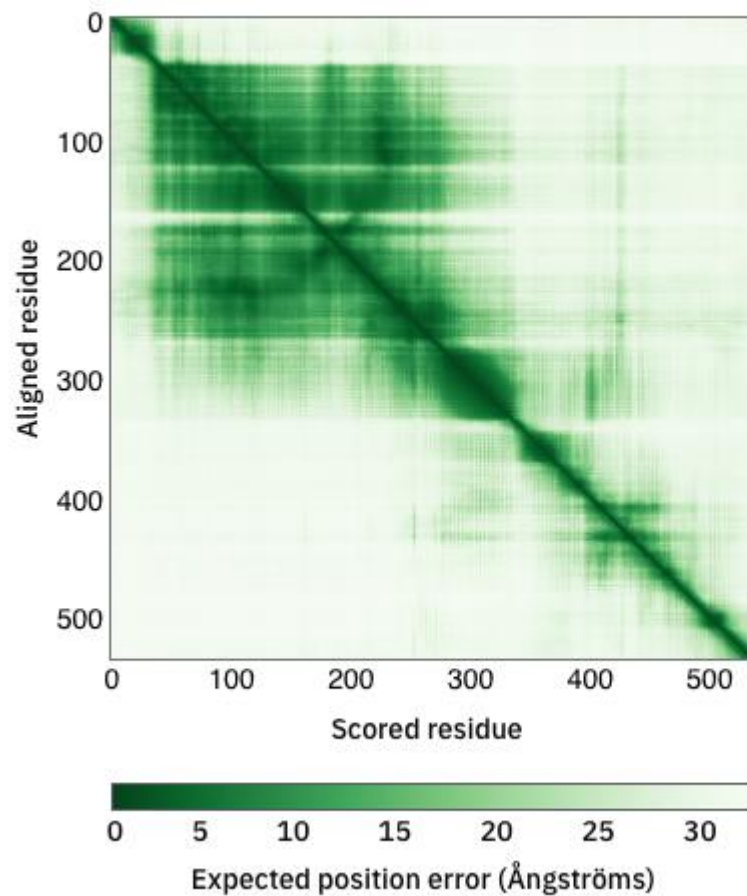
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q13573-... 1: SNW do... A

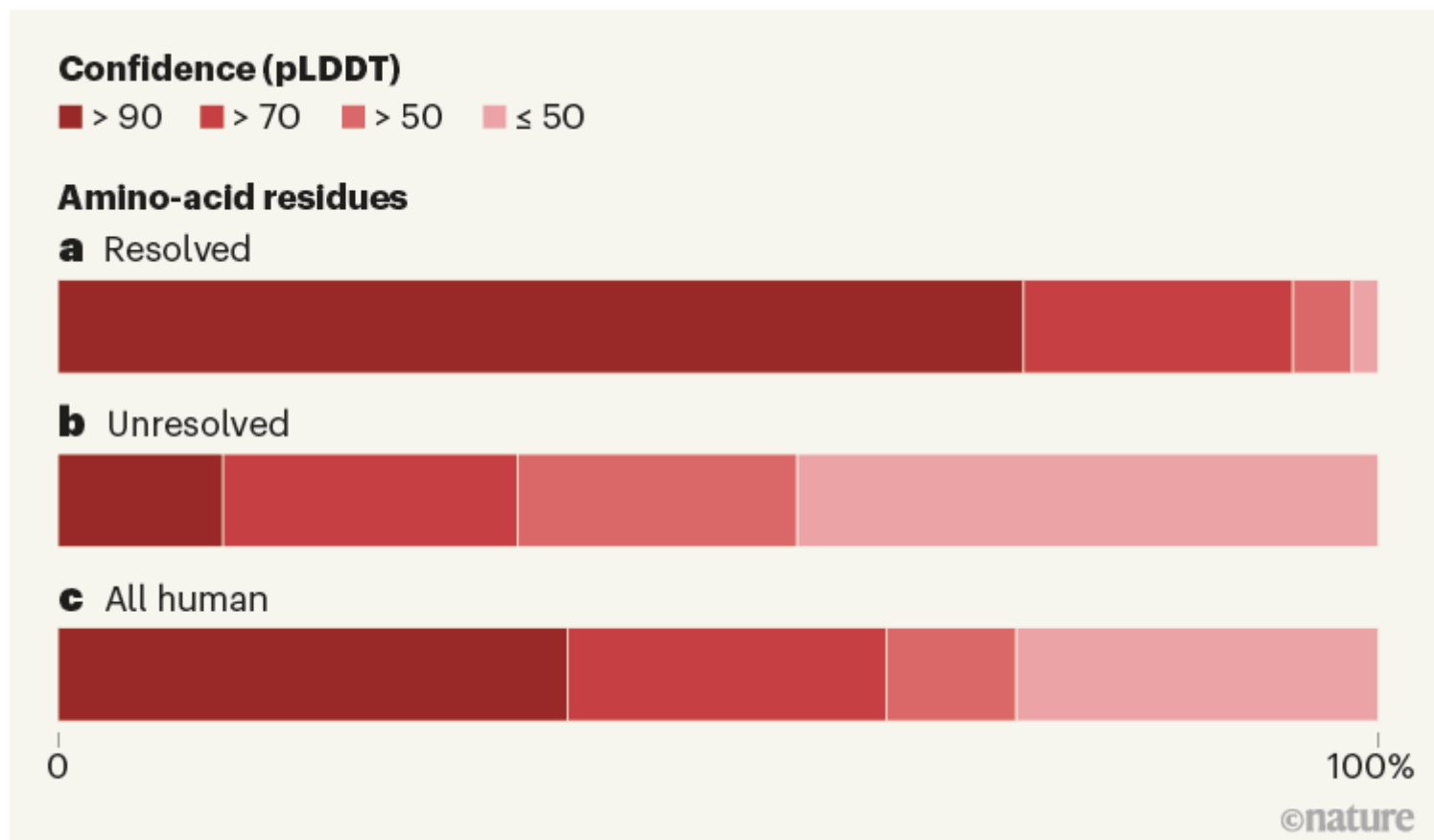
1 MALTSFLPAPTQLSQDQLEAAEKARSQSRQTSLVSSRRPEPYGRKNGIPLLEDGFGDGAFFIEHVAQYPLDMGRKKKMSNALAIQVDSEGKIKYDAIARQGSQKDKYISKYITDLVPKEV
131 21 31 41 51 61 71 81 91 101 111 121
NMADDPLQRPDEEAIEKTEIKTRVLEKSSQSVKVAAMPRAADKLAPQAIYIRYTPSQQGVAFNSGAKQVIRVMVEQKDPMEPPRFKINKIIPRGPPSPAPVMHSPSRKMTKVEQKEQWKIP
251 261 271 281 291 301 311 321 331 341 351 361 371
PCISNWNKNAKTYIPLDKRLAADRGLOTVINHENFKALEALYIADRKAREAVEMRAOVERKKMAOEKEKEHEEKLREMAOKARERRAGIKTHVEKEDGEAREDEIRHDDRKEROHDRNLSRA



AlphaFold tells you where is it right!



How good are the predictions of human proteins?



pLDDT - per-residue estimate of its confidence on a scale from 0 - 100 model's predicted score on the [IDDT-C \$\alpha\$ metric](#) (local superposition-free score for comparing protein structures and models using distance difference tests).

DĚKUJI ZA POZORNOST