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        0.00000                  12CA  83
ORIGX2      0.000000  1.000000  0.000000        0.00000                  12CA  84
ORIGX3      0.000000  0.000000  1.000000        0.00000                  12CA  85
SCALE1      0.023419  0.000000  0.006100        0.00000                  12CA  86
SCALE2      0.000000  0.023981  0.000000        0.00000                  12CA  87
SCALE3      0.000000  0.000000  0.014156        0.00000                  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 propertiesand 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.
- diffrn 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: 10.1107/97809553602060000107

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
…

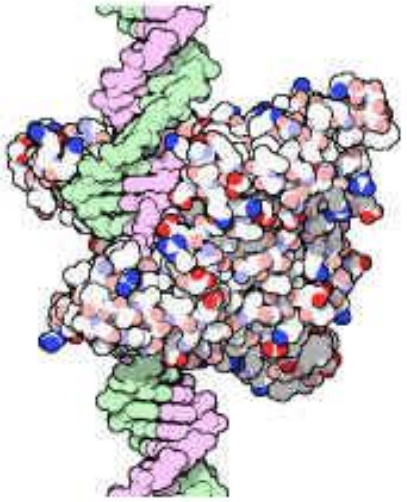
Databáze – není jich málo…

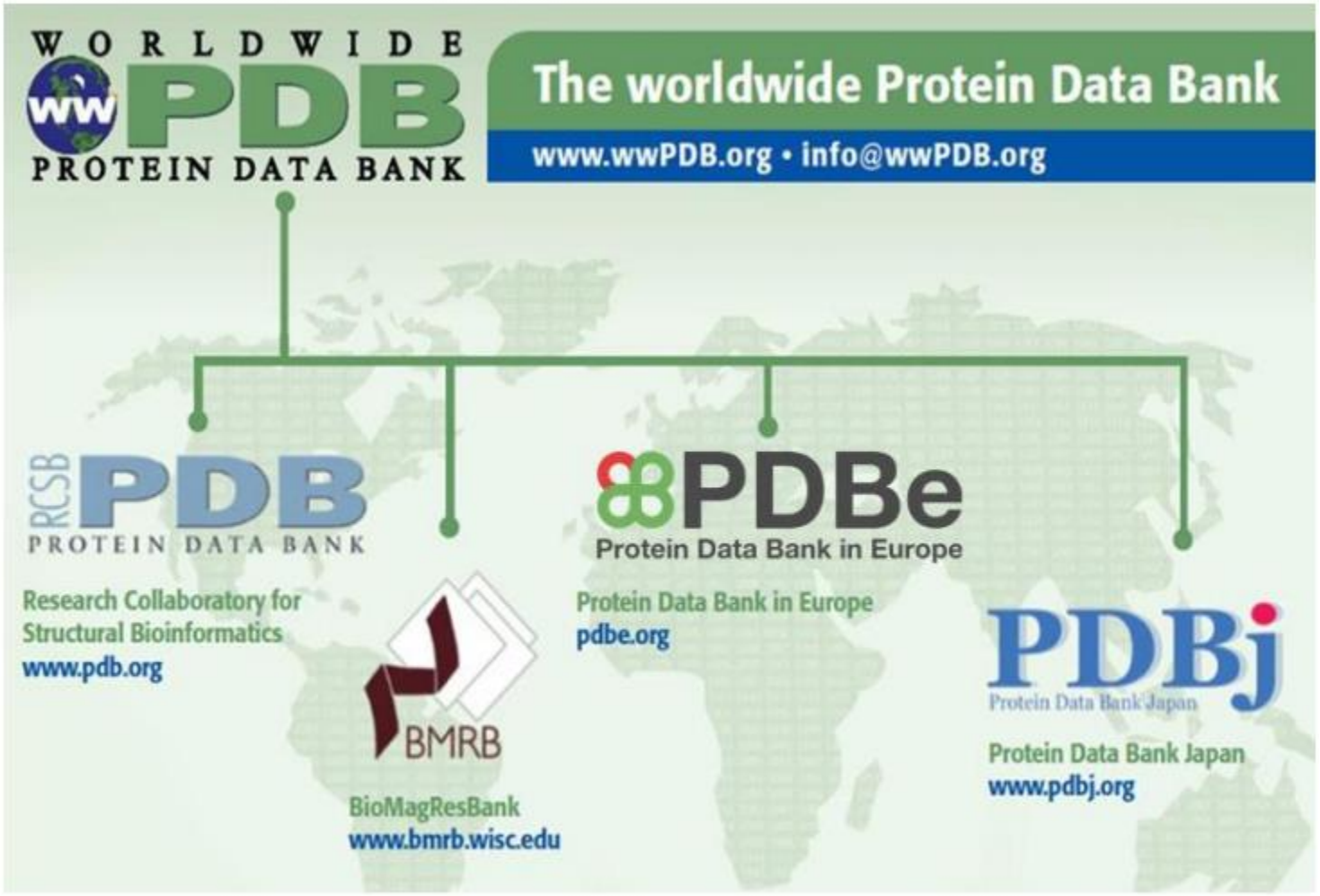
Structure and sequence/structure databases	
SCOP	http://scop.mrc-lmb.cam.ac.uk/scop/
CATH	http://www.biochem.ucl.ac.uk/bsm/cath/
FSSP	http://www2.ebi.ac.uk/dali/fssp/
Molecular Modeling Database	http://www.ncbi.nlm.nih.gov/Structure/
CAMPASS	http://www-cryst.bioc.cam.ac.uk/~campass/
ISSD	http://www.protein.bio.msu.su/issd/
Library of Protein Family Cores (LPFC)	http://WWW-SMI.Stanford.EDU/projects/helix/LPFC/
3D_ALI (a database of aligned protein structures and related sequences)	http://www.embl-heidelberg.de/argos/ali/ali_info.html
IDITIS (relational database and query tool for proteins)	http://www.oxmol.co.uk/prods/iditis/
HSSP	http://www.sander.embl-heidelberg.de/hssp/
Specialty databases	
HIV Protease Database	http://www-lbsc.ncifcrf.gov/HIVdb/
Nucleic Acid Database	http://ndbserver.rutgers.edu/
Prolysis (protease and protease inhibitor Web server)	http://delphi.phys.univ-tours.fr/Prolysis/
International Immunogenetics Database (IMGT)	http://imgt.cnusc.fr:8104/
Enzyme Structures Database	http://www.biochem.ucl.ac.uk/bsm/enzymes/
Features databases	
Molecular Movements Database	http://bioinfo.mbb.yale.edu/MolMovDB/
OLDERADO	http://neon.chem.le.ac.uk/olderado/
PROCAT	http://www.biochem.ucl.ac.uk/bsm/PROCAT/PROCAT.html
Protein Quaternary Structures (PQS)	http://pqs.ebi.ac.uk/
ReLIBase (receptor–ligand complexes database)	http://www2.ebi.ac.uk:8081/home.html
PROMISE	http://bioinf.leeds.ac.uk/promise/
PDBSum	http://www.biochem.ucl.ac.uk/bsm/pdbsum/
Biological Macromolecule Crystallization Database (BMCD)	http://h178133.nist.gov:4400/bmcd/bmcd.html
Resources	
Protein Data Bank	**htttp://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/





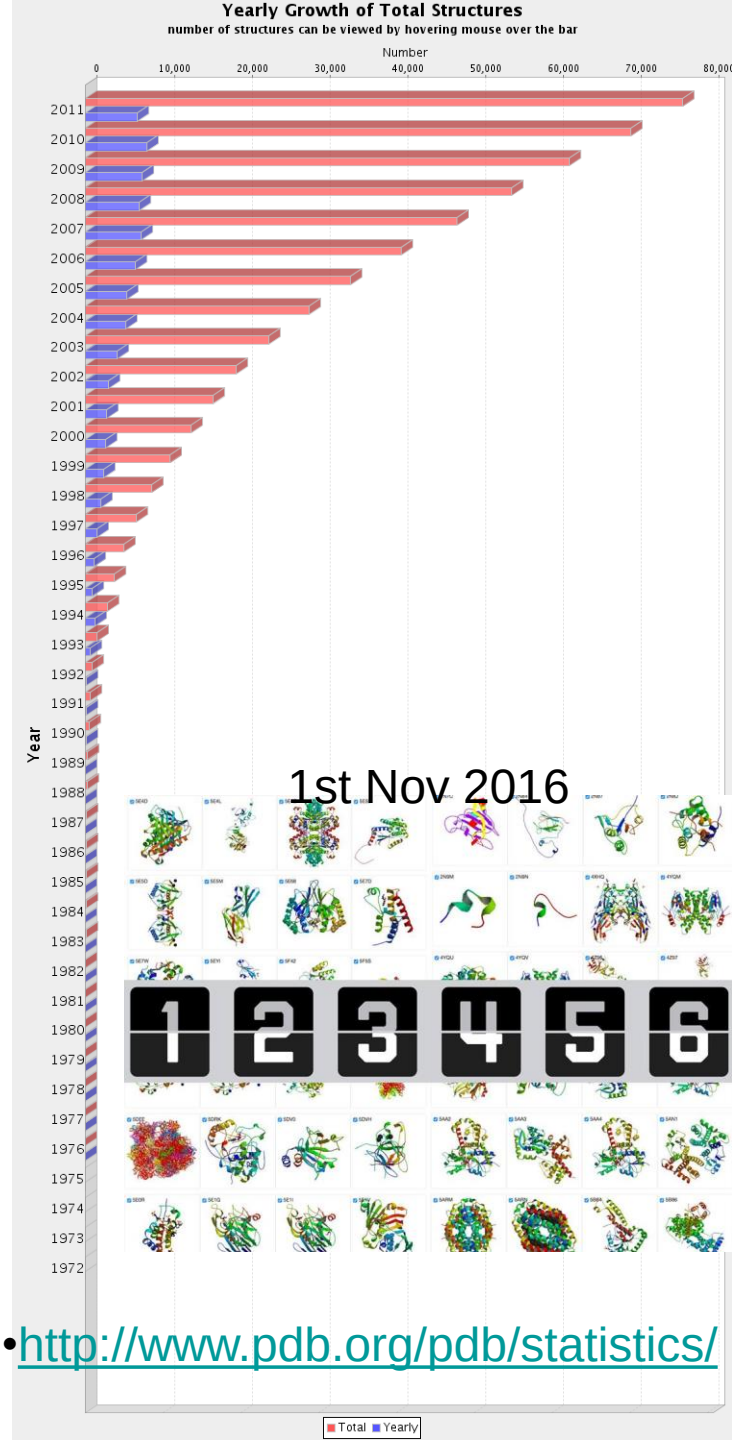
> 120k biomacromolecular structures

PDB statistics
říjen 2011

- Struktury 76814
  - X-ray 67125
  - NMR 9108
  - EM 380

  - Proteiny 71148
  - NA 2312
  - Protein/NA 3331

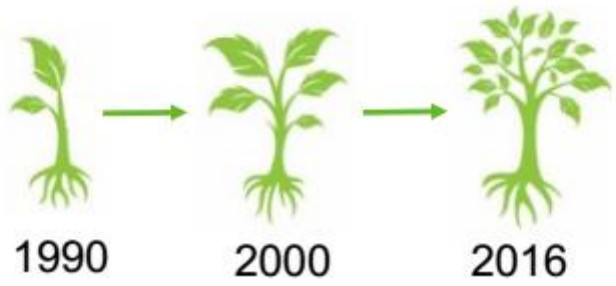
- zajímavosti, nejvíc struktur:
  - organism-Thermus thermophilus (4751)
  - gene source Homo Sapiens (20127)
  - R 1.5-2.0 A (26000)
  - Software – CNS (25276)
  - EC – 3.2.1.17 Lysozyme (1150)
  - Residues - 100-300 (30000)
  - Journal J.Mol.Biol. (9247)



- http://www.pdb.org/pdb/statistics/

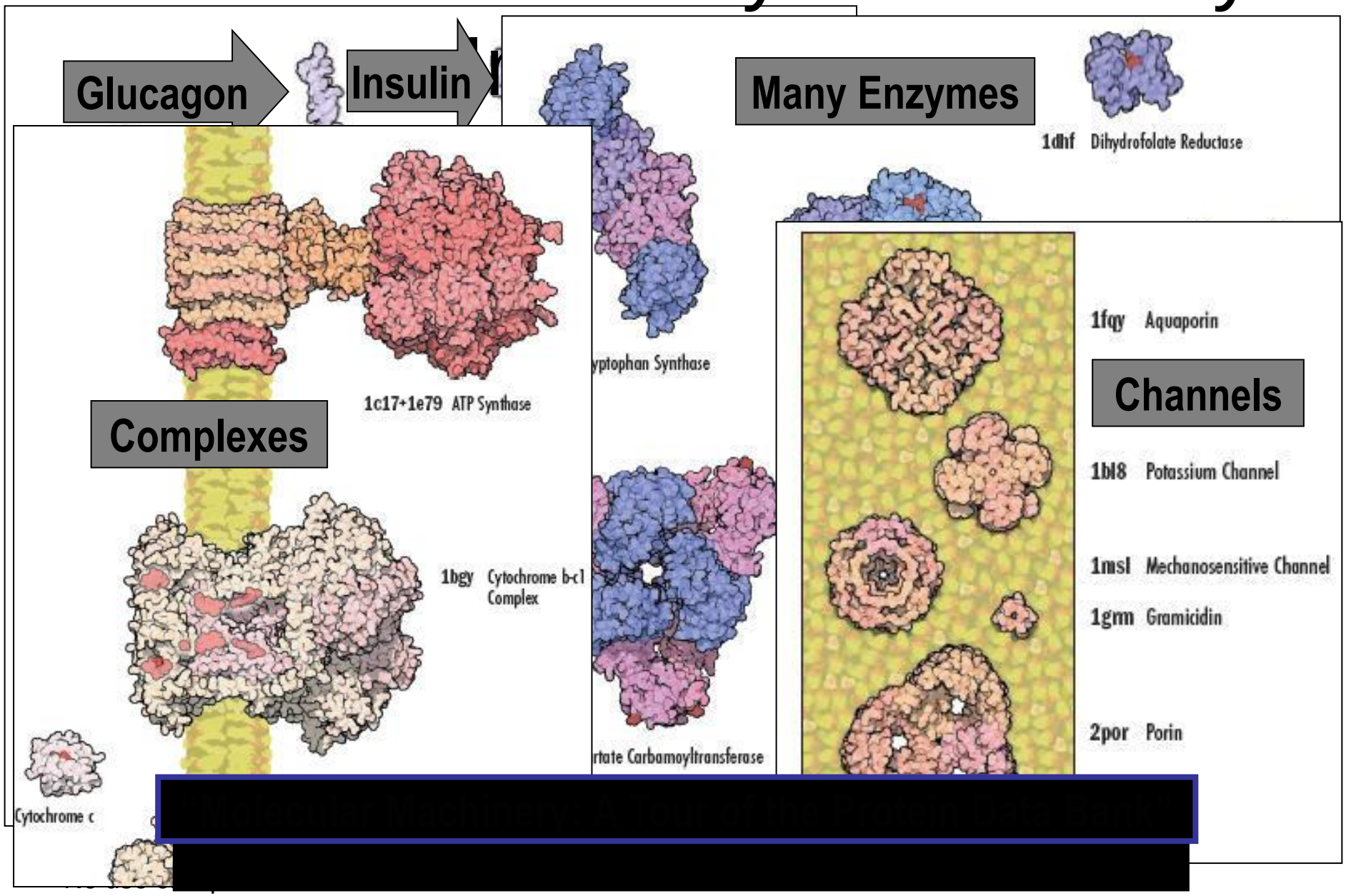
Growth of PDB

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

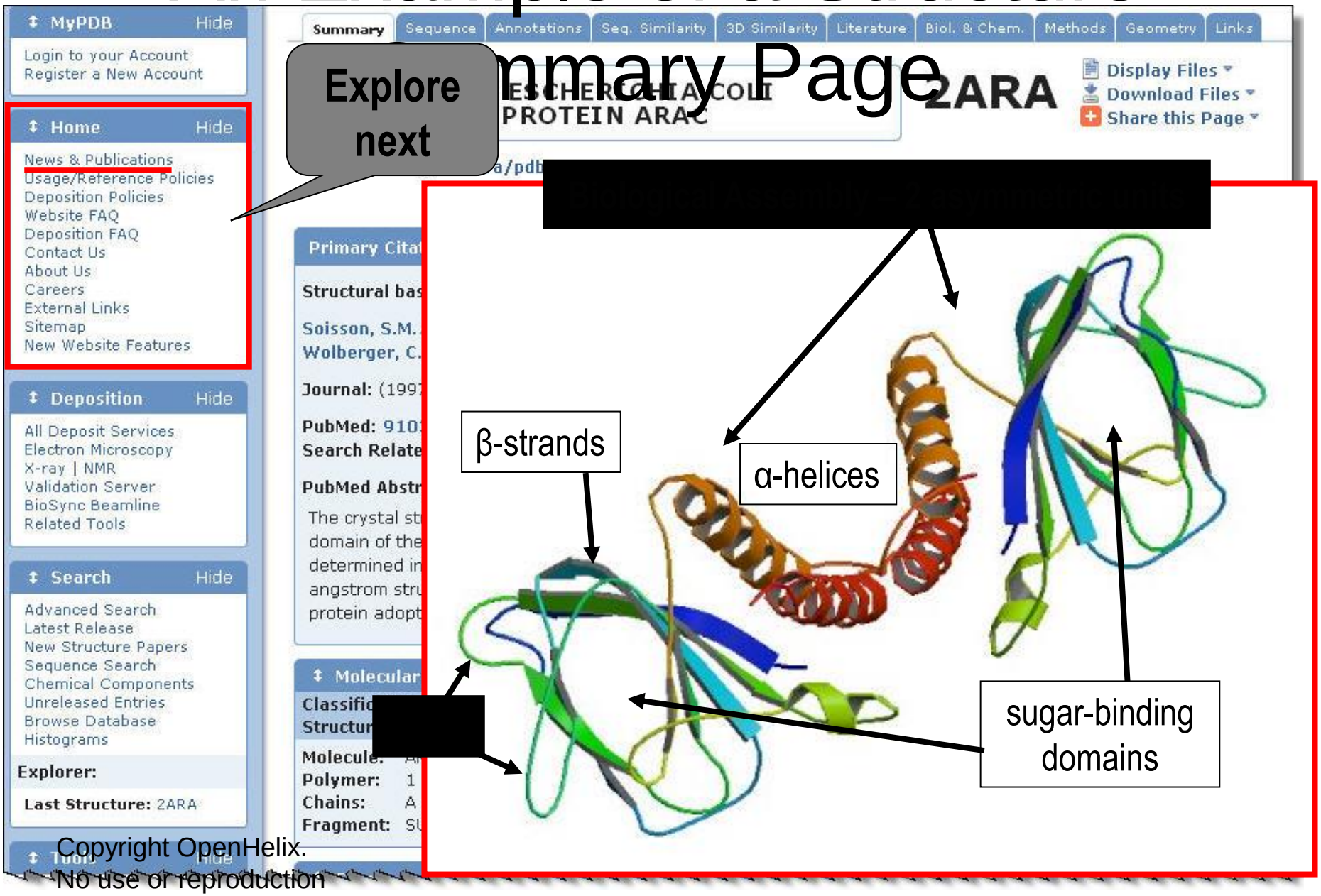


A richness of data about individual protein families / folds.

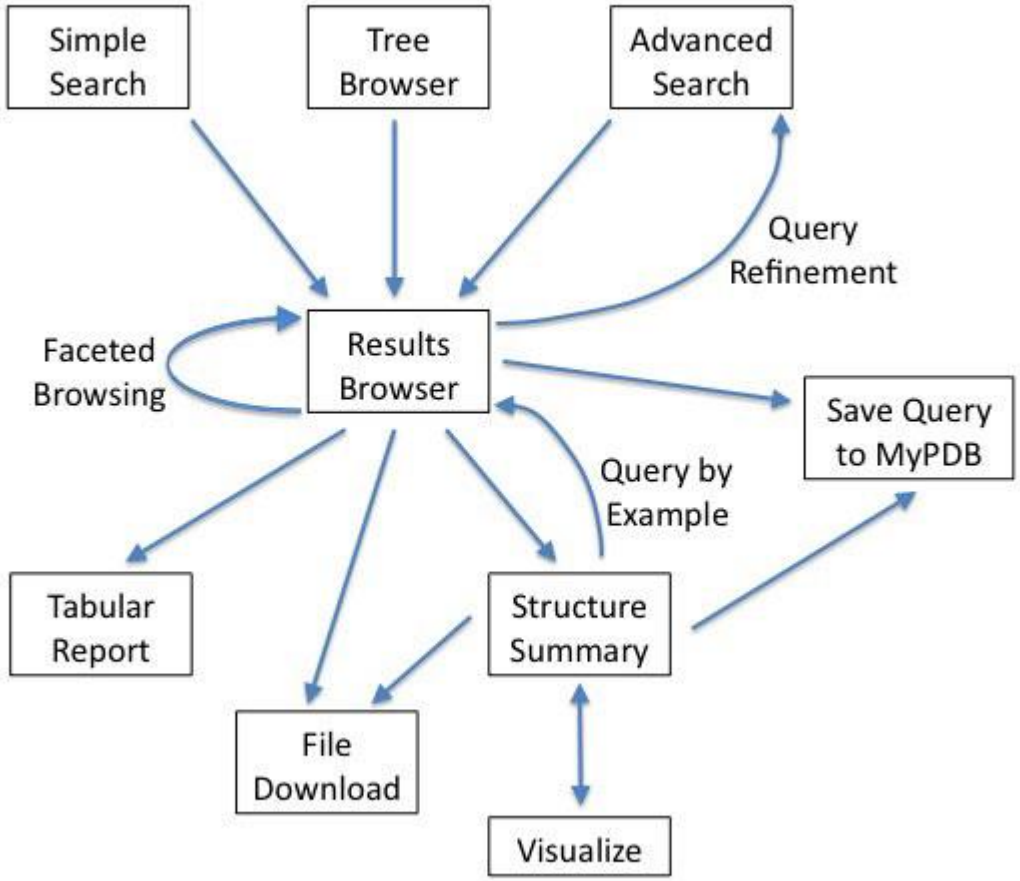
Structural Diversity - Constantly



An Example of a Structure Summary Page

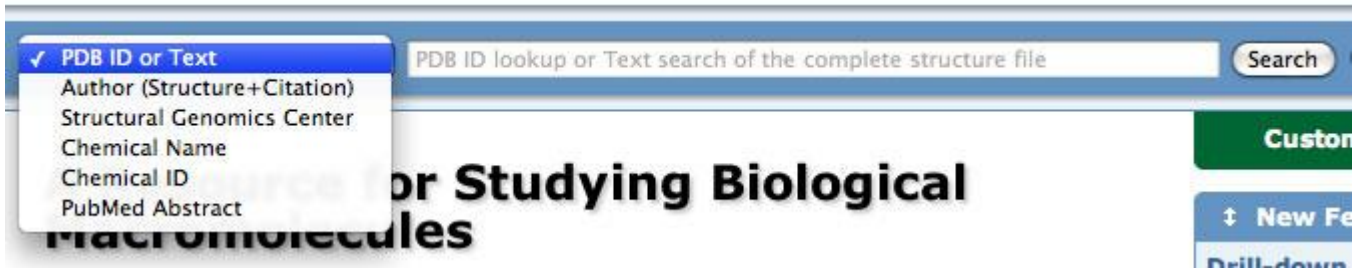


Hlavní složky RCSB PDB website



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

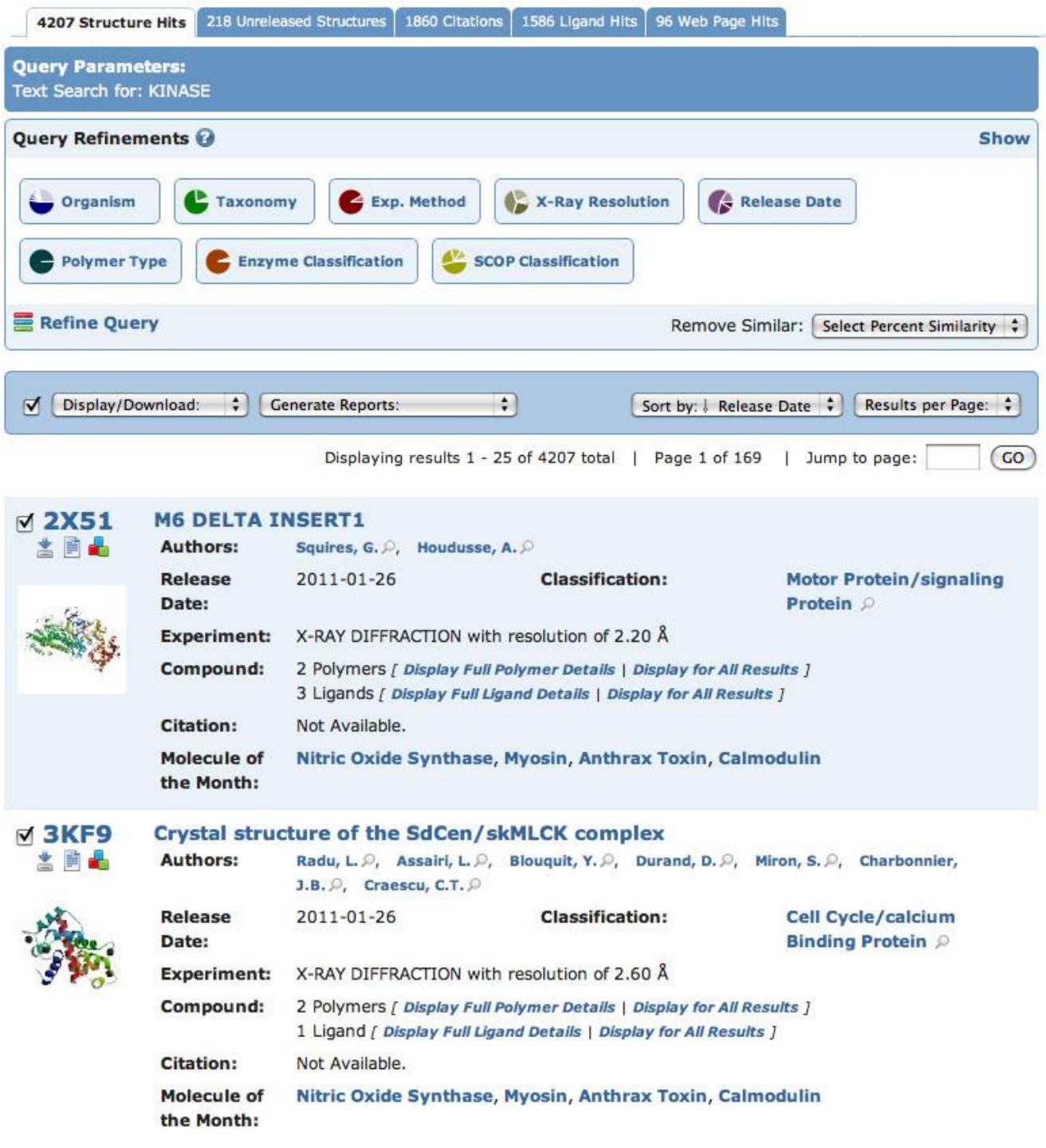


- top navigation bar
- PDB ID (e.g., 4HHB)
- full-text search
- Výběr z
  - “Author”
  - “Makromolecule”
  - “Sequence”
  - “Ligand”

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

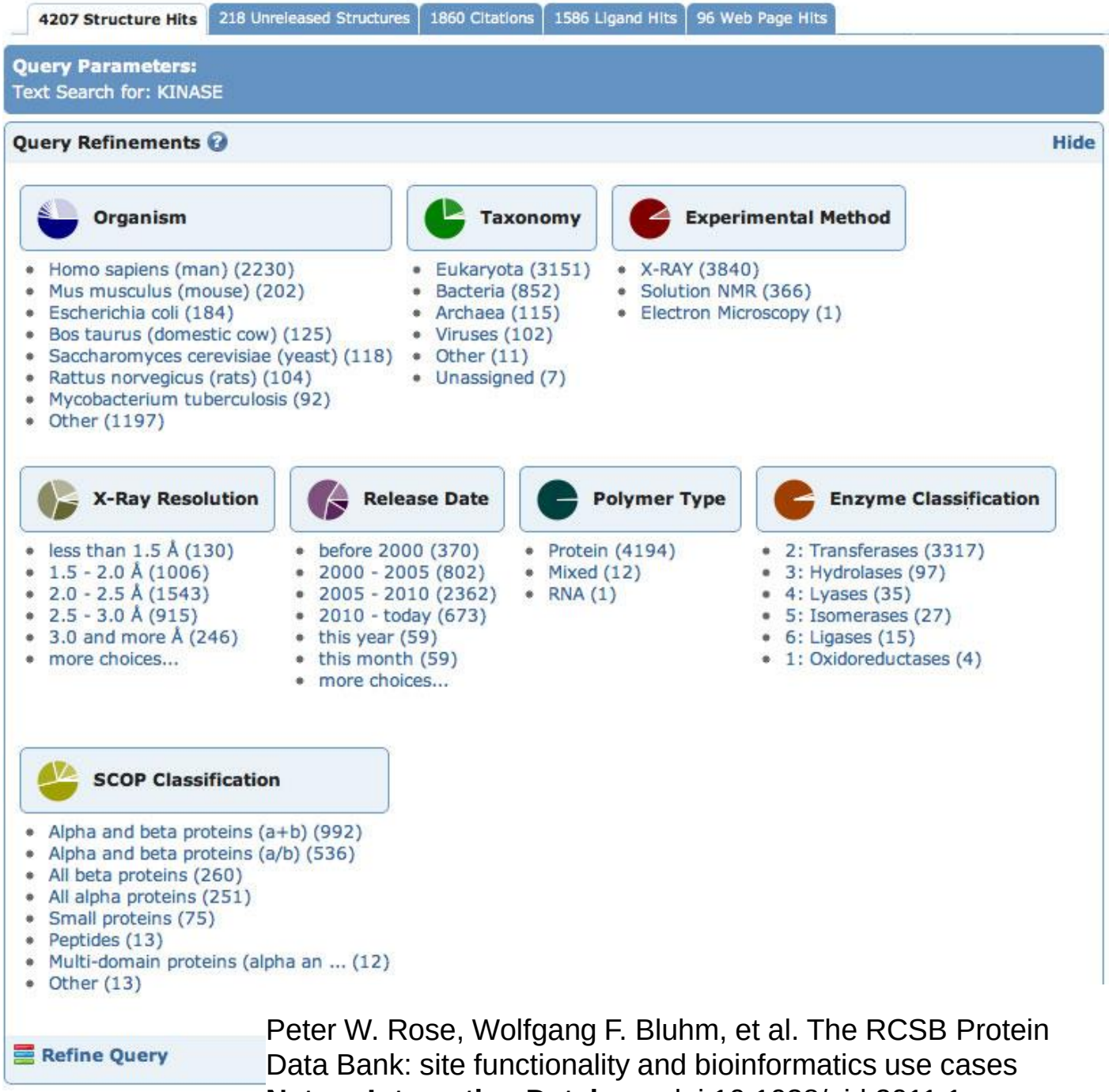
Results Browser

- 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



Refine search

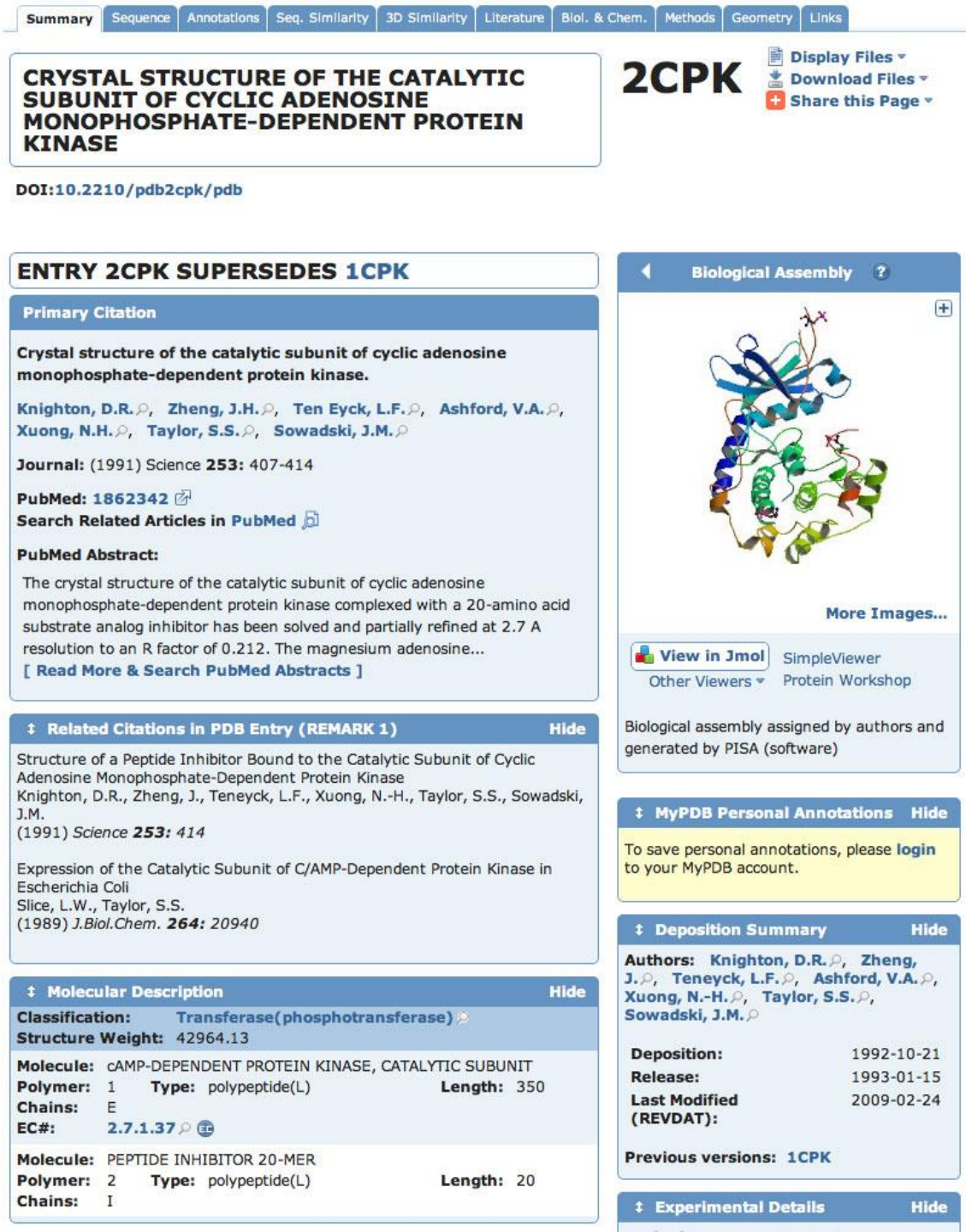
- zpřesnění výběrových kritérií



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

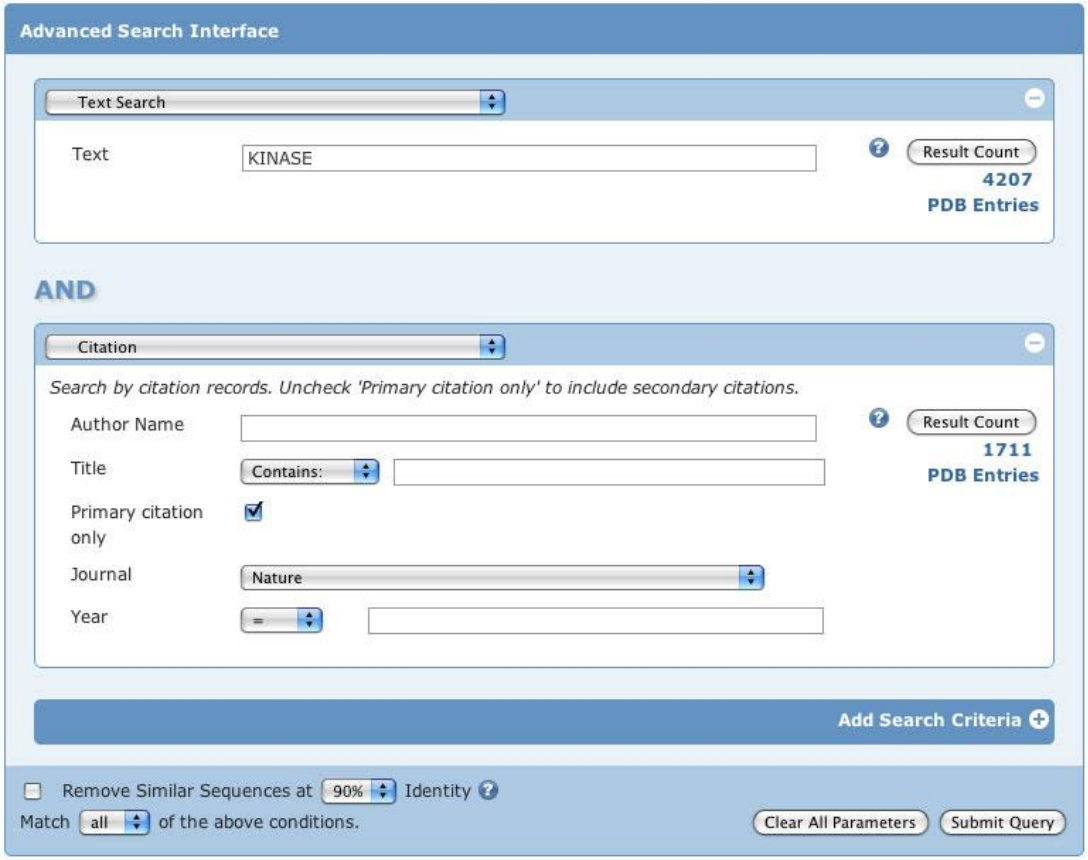
Structure Summary

- 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).



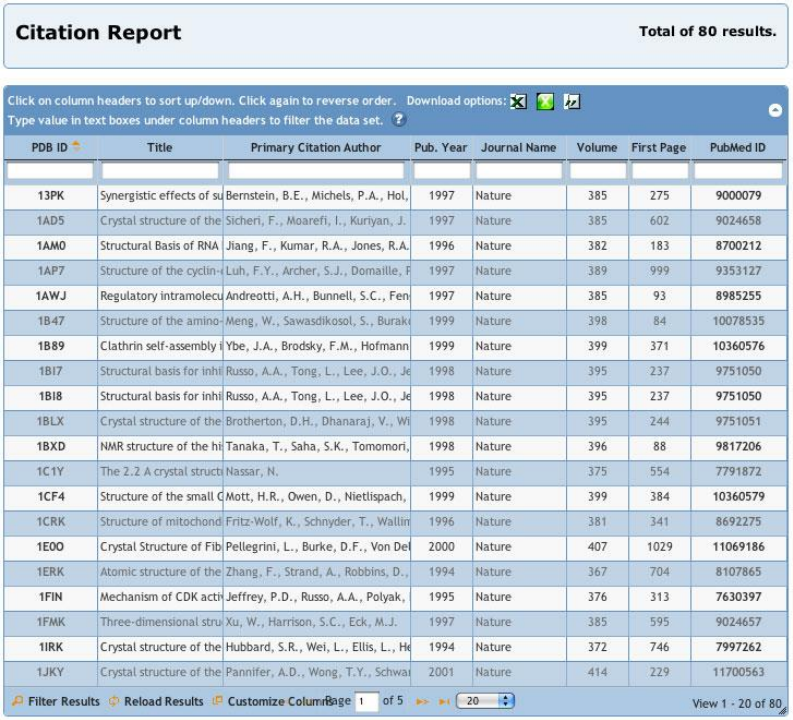
Advanced search

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



Report

- Tabulka s výsledky

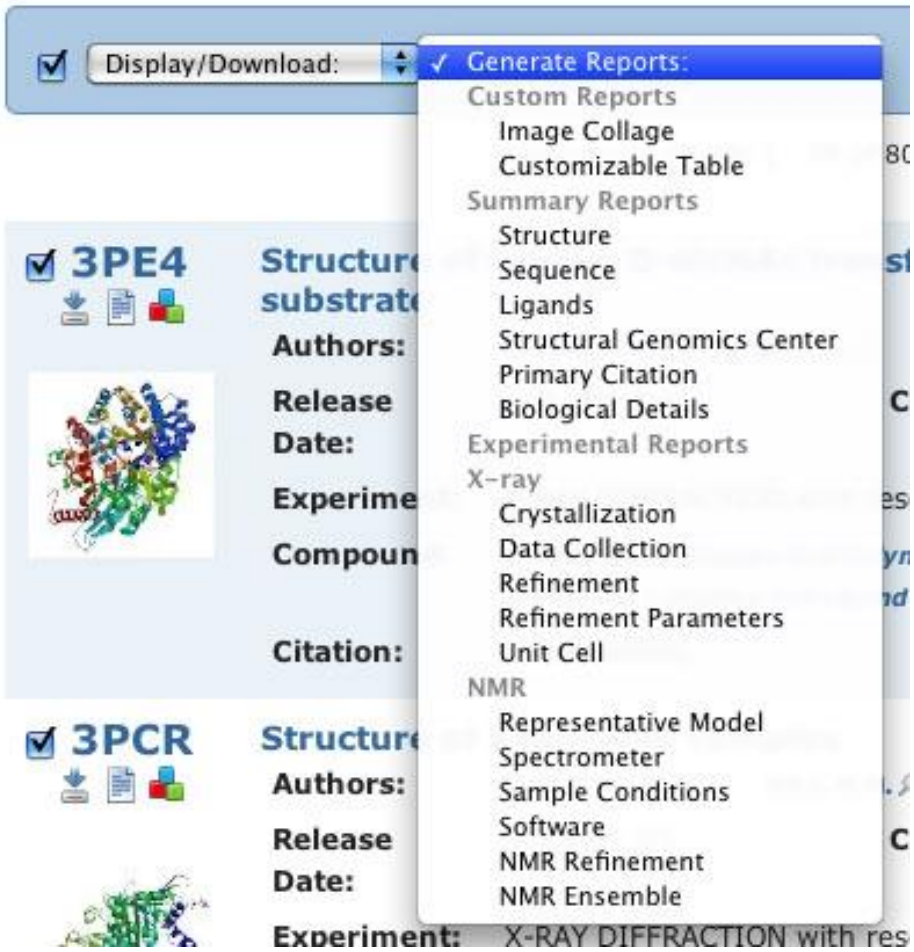


Citation Report — Total of 80 results.

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.

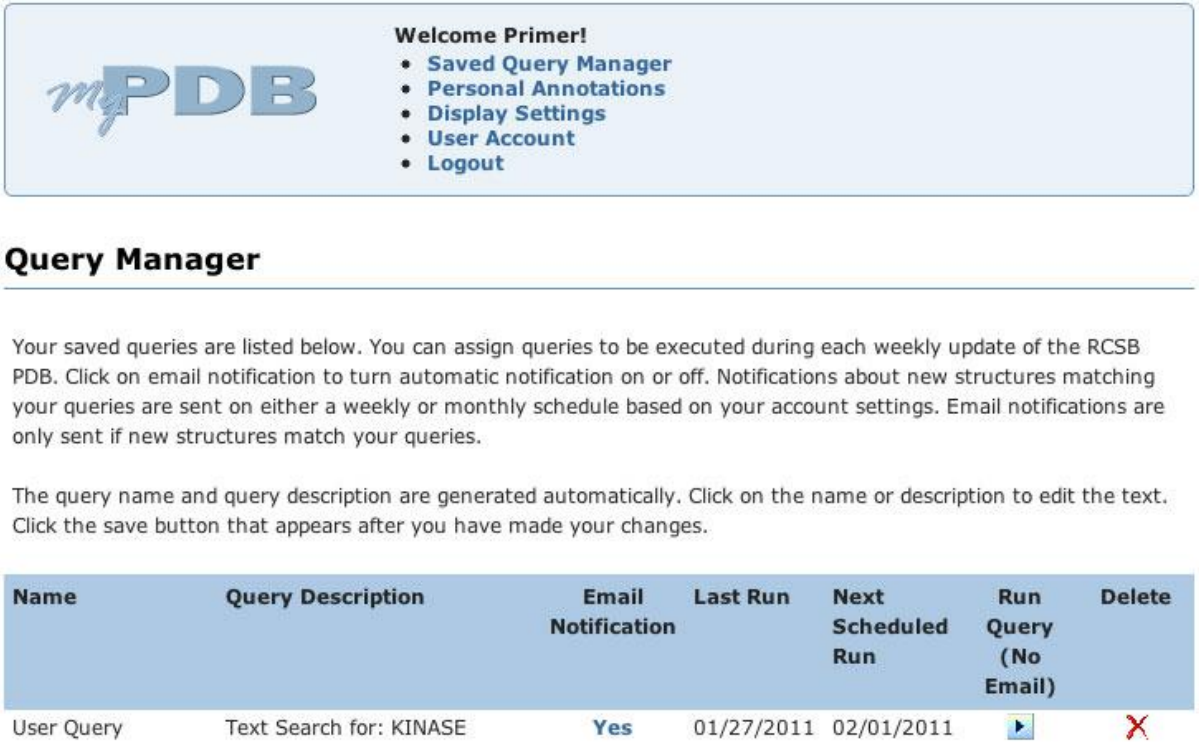
PDB ID	Title	Primary Citation Author	Pub. Year	Journal Name	Volume	First Page	PubMed ID
13PK	Synergistic effects of su	Bernstein, B.E., Michels, P.A., Hol,	1997	Nature	385	275	9000079
1AD5	Crystal structure of the	Sicheri, F., Moarefi, I., Kuriyan, J.	1997	Nature	385	602	9024658
1AM0	Structural Basis 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, F	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., Sawasdikosol, S., Burak	1999	Nature	398	84	10078535
1B89	Clathrin self-assembly i	Ybe, J.A., Brodsky, F.M., Hofmann	1999	Nature	399	371	10360576
1BI7	Structural basis for inhi	Russo, A.A., Tong, L., Lee, J.O., Je	1998	Nature	395	237	9751050
1BI8	Structural basis for inhi	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 hi	Tanaka, T., Saha, S.K., Tomomori,	1998	Nature	396	88	9817206
1C1Y	The 2.2 A crystal structi	Nassar, N.	1995	Nature	375	554	7791872
1CF4	Structure of the small C	Mott, H.R., Owen, D., Nietlispach,	1999	Nature	399	384	10360579
1CRK	Structure of mitochond	Fritz-Wolf, K., Schnyder, T., Wallin	1996	Nature	381	341	8692275
1E0O	Crystal Structure of Fib	Pellegrini, L., Burke, D.F., Von Del	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., He	1994	Nature	372	746	7997262
1JKY	Crystal structure of the	Pannifer, A.D., Wong, T.Y., Schwar	2001	Nature	414	229	11700563

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



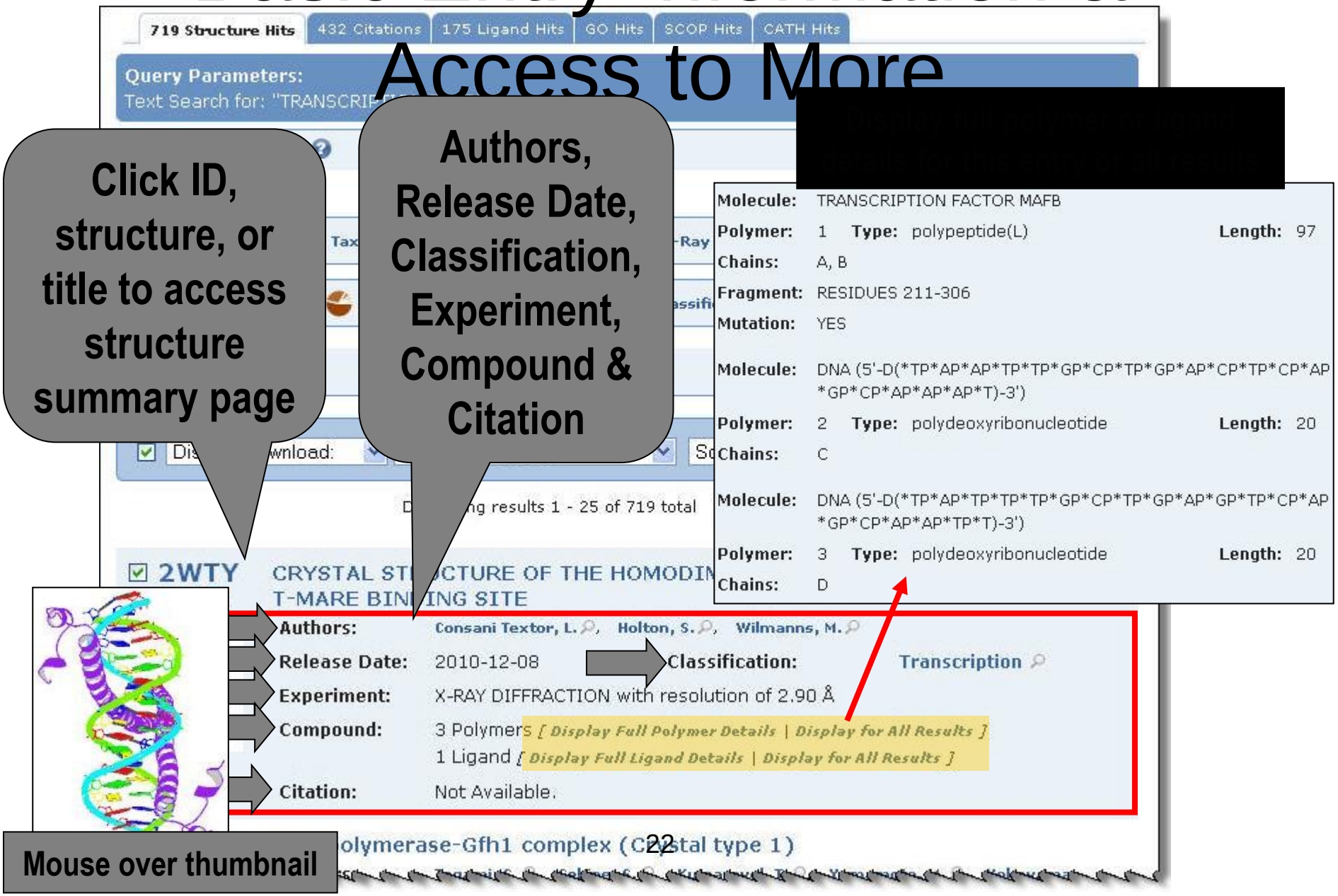
Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases **Nature Interaction Database** doi:10.1038/pid.2011.1

myPDB



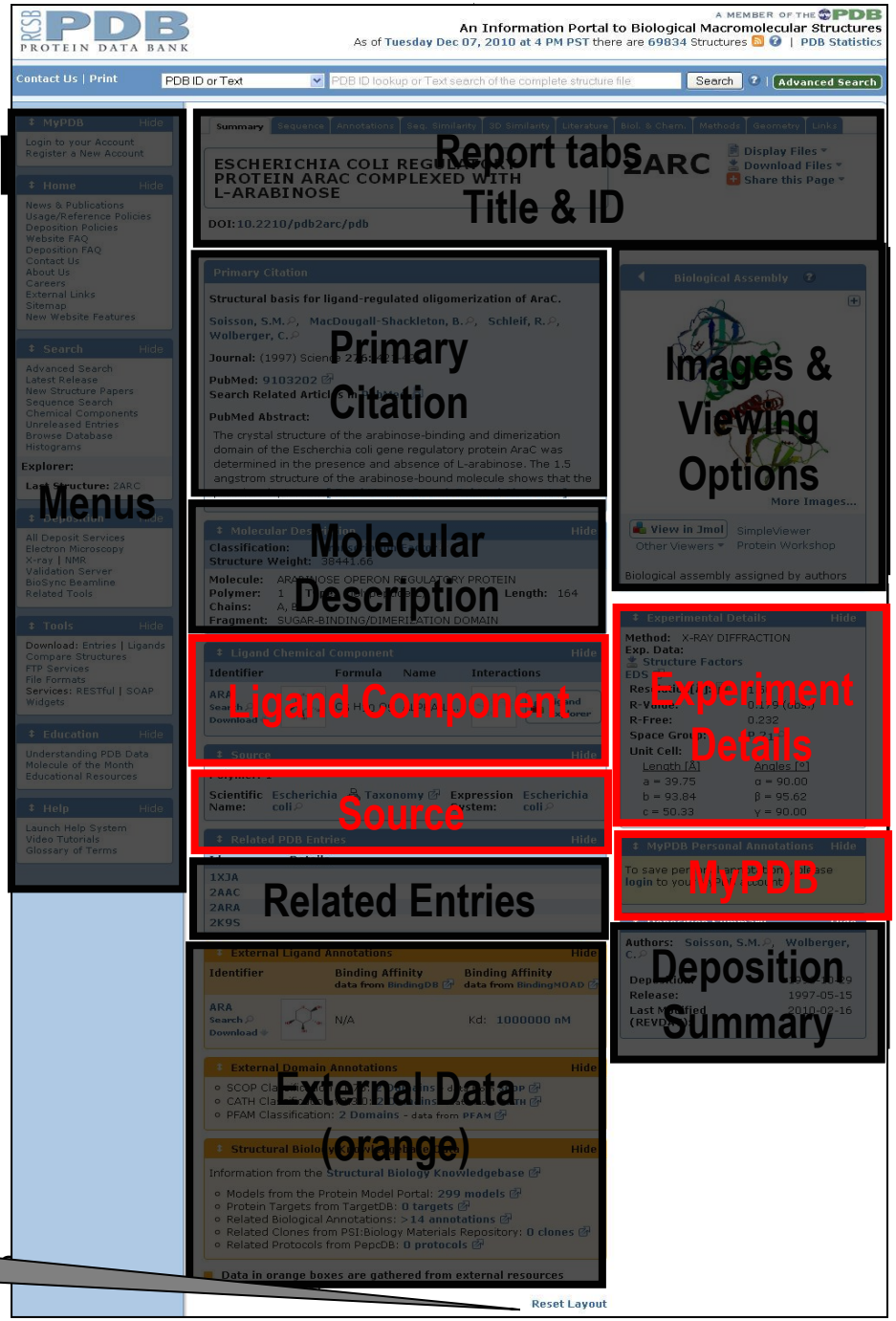
Peter W. Rose, Wolfgang F. Bluhm, et al. The RCSB Protein Data Bank: site functionality and bioinformatics use cases **Nature Interaction Database** doi:10.1038/pid.2011.1

Basic Entry Information & Access to More



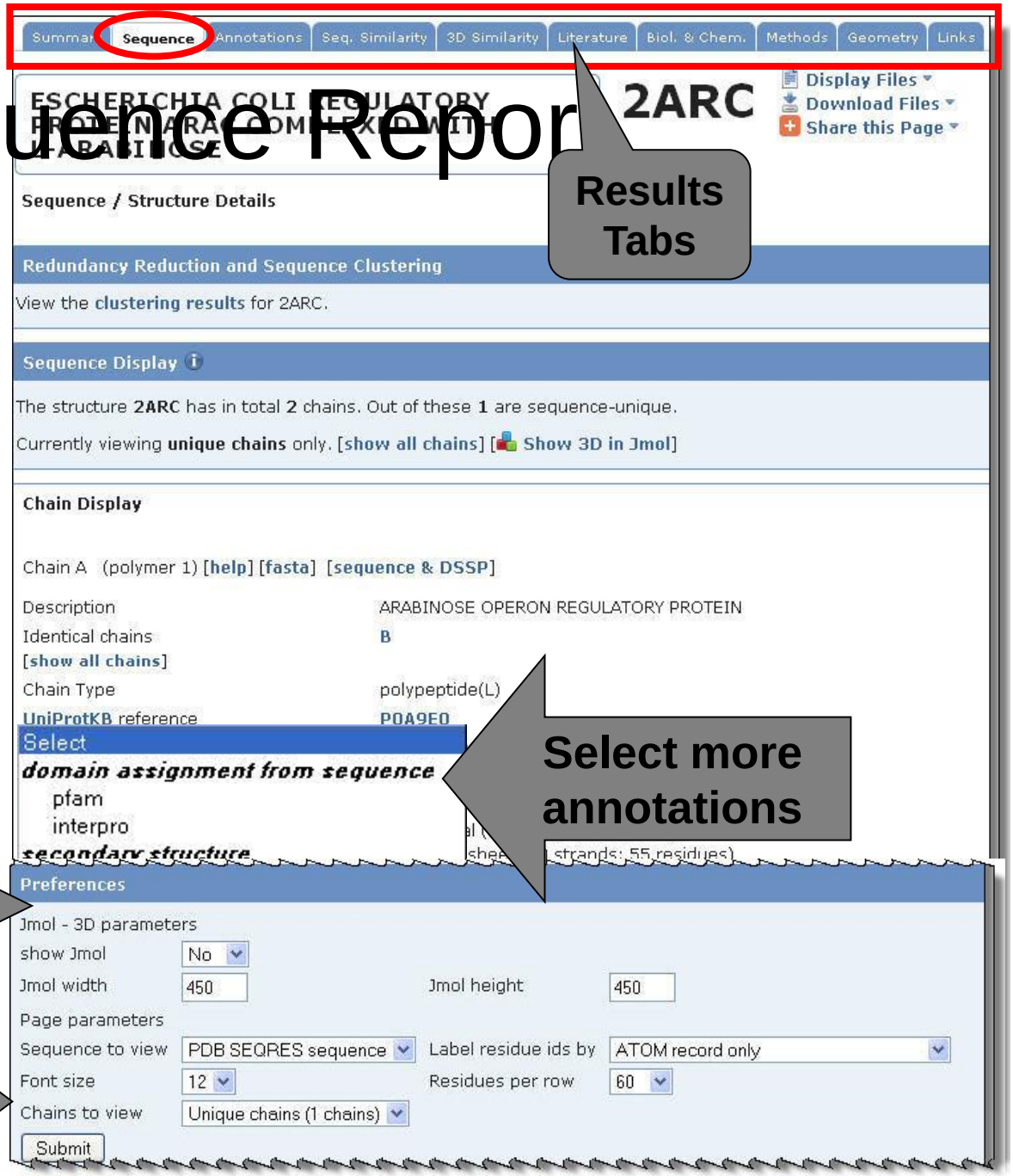
Structure Su

- 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



Sequence Repor

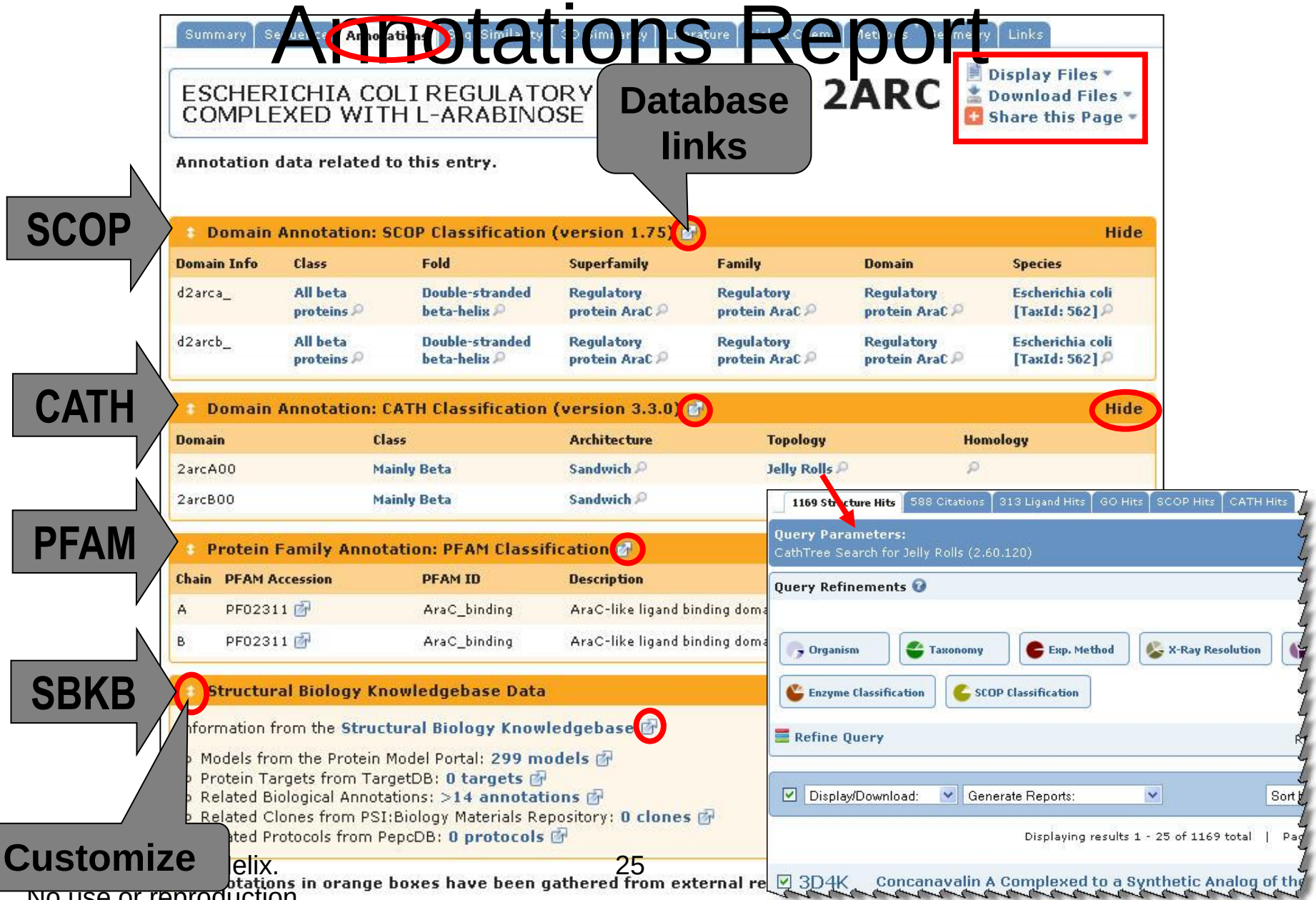
- **Graphic of secondary structures**
- **Choose domain assignments or annotations from dropdown menu**
- **Specify how you want to view graphic**



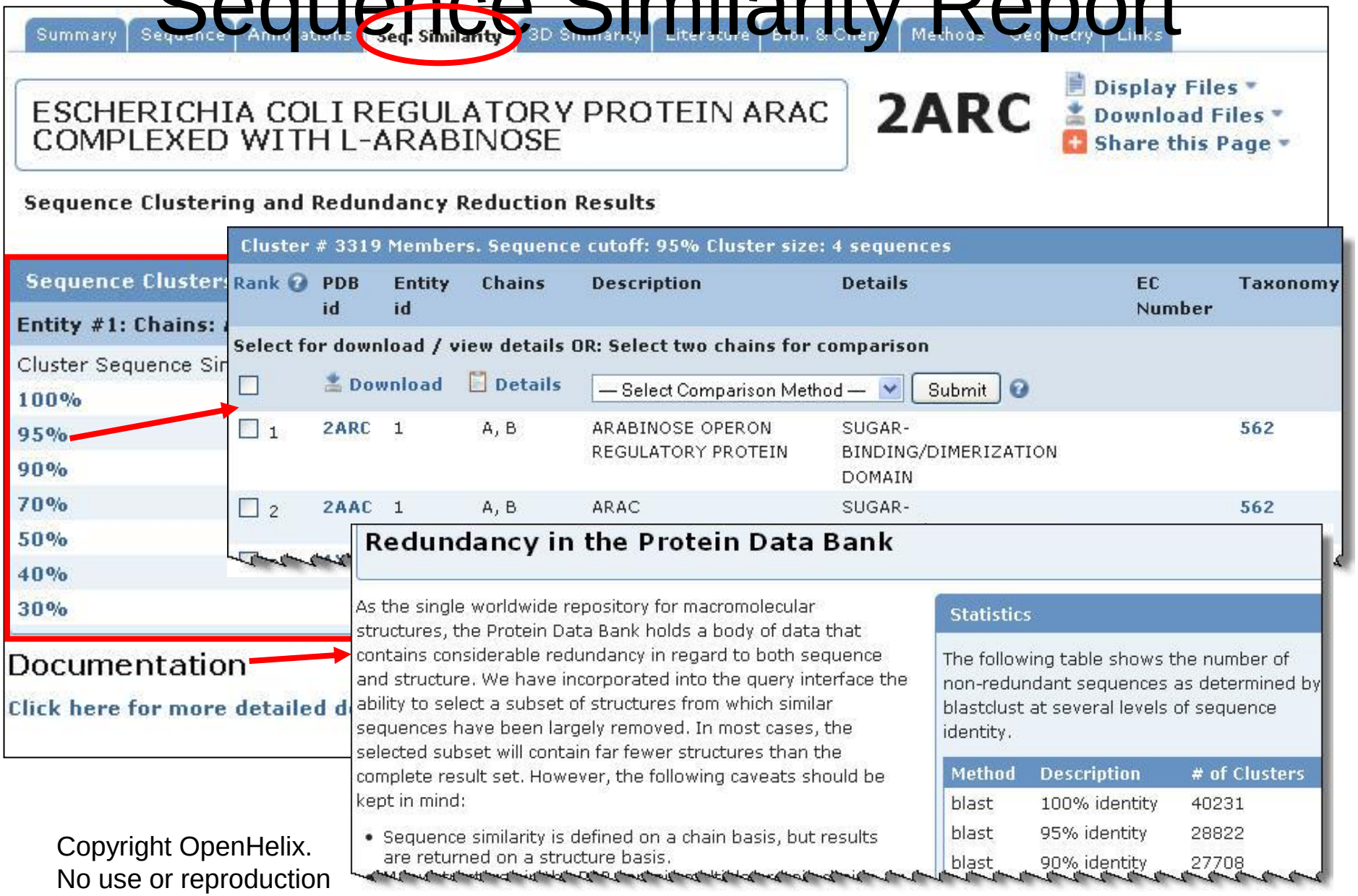
Set viewing preferences

References below

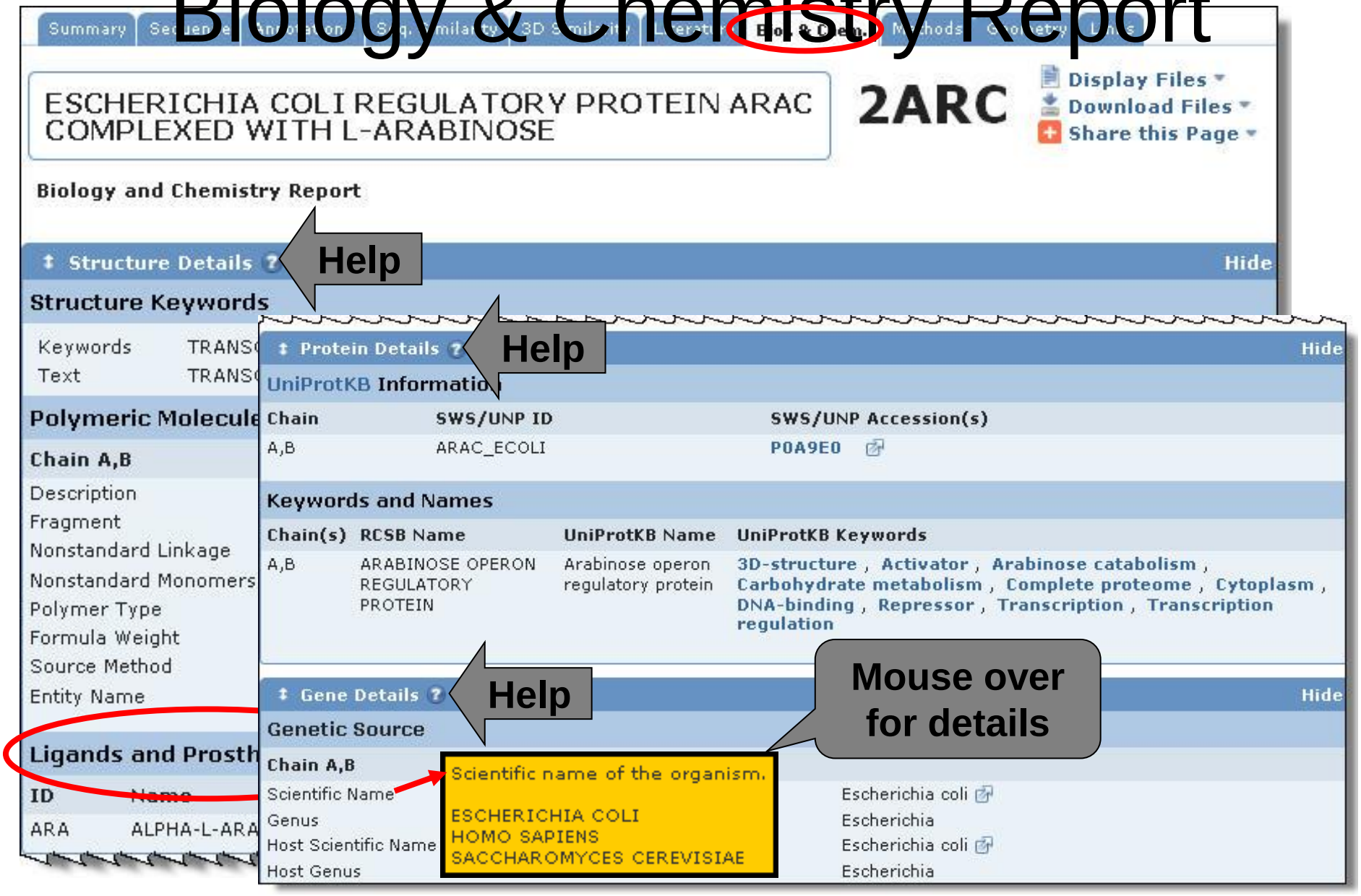
Annotations Report



Sequence Similarity Report

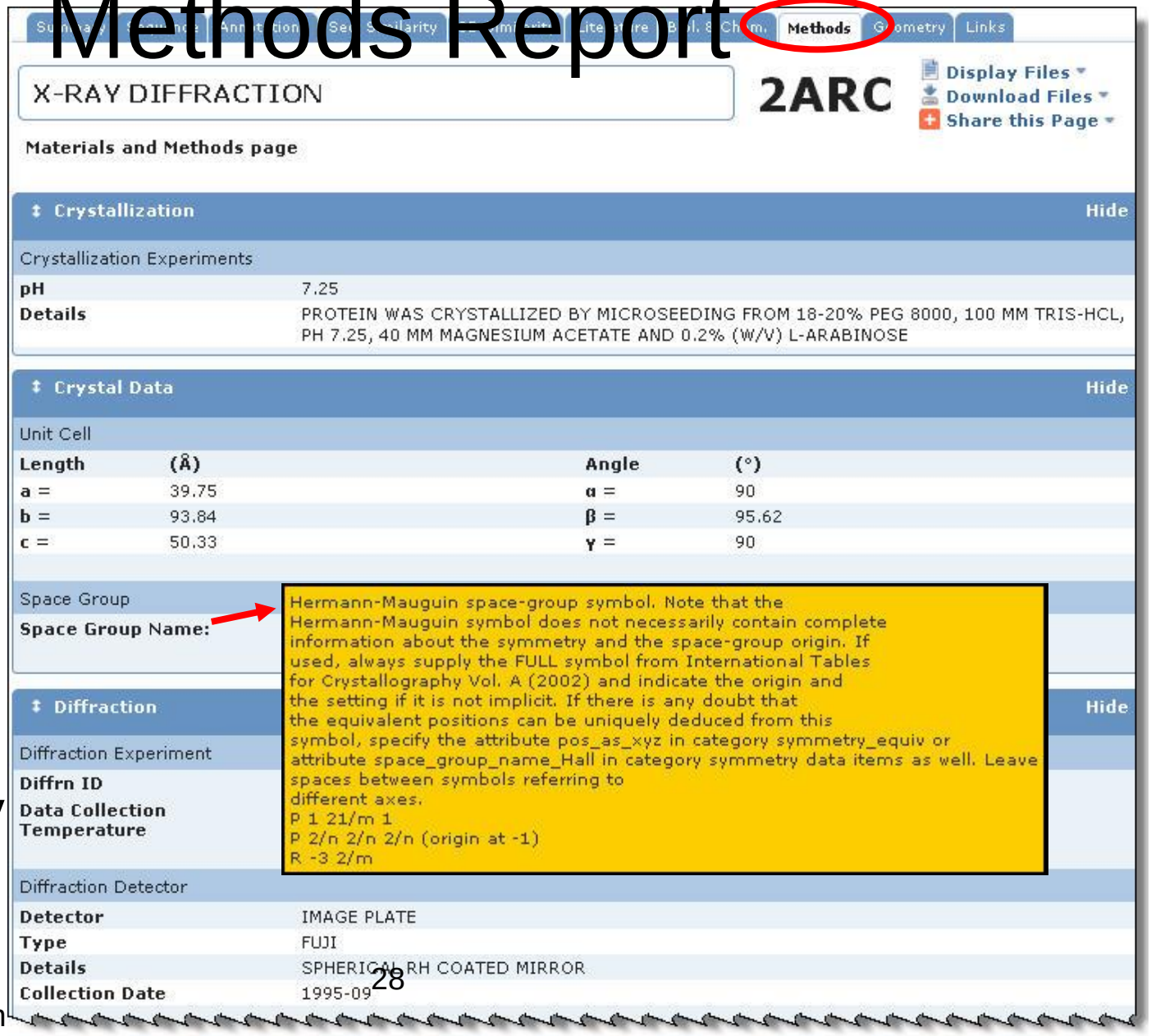


Biology & Chemistry Report

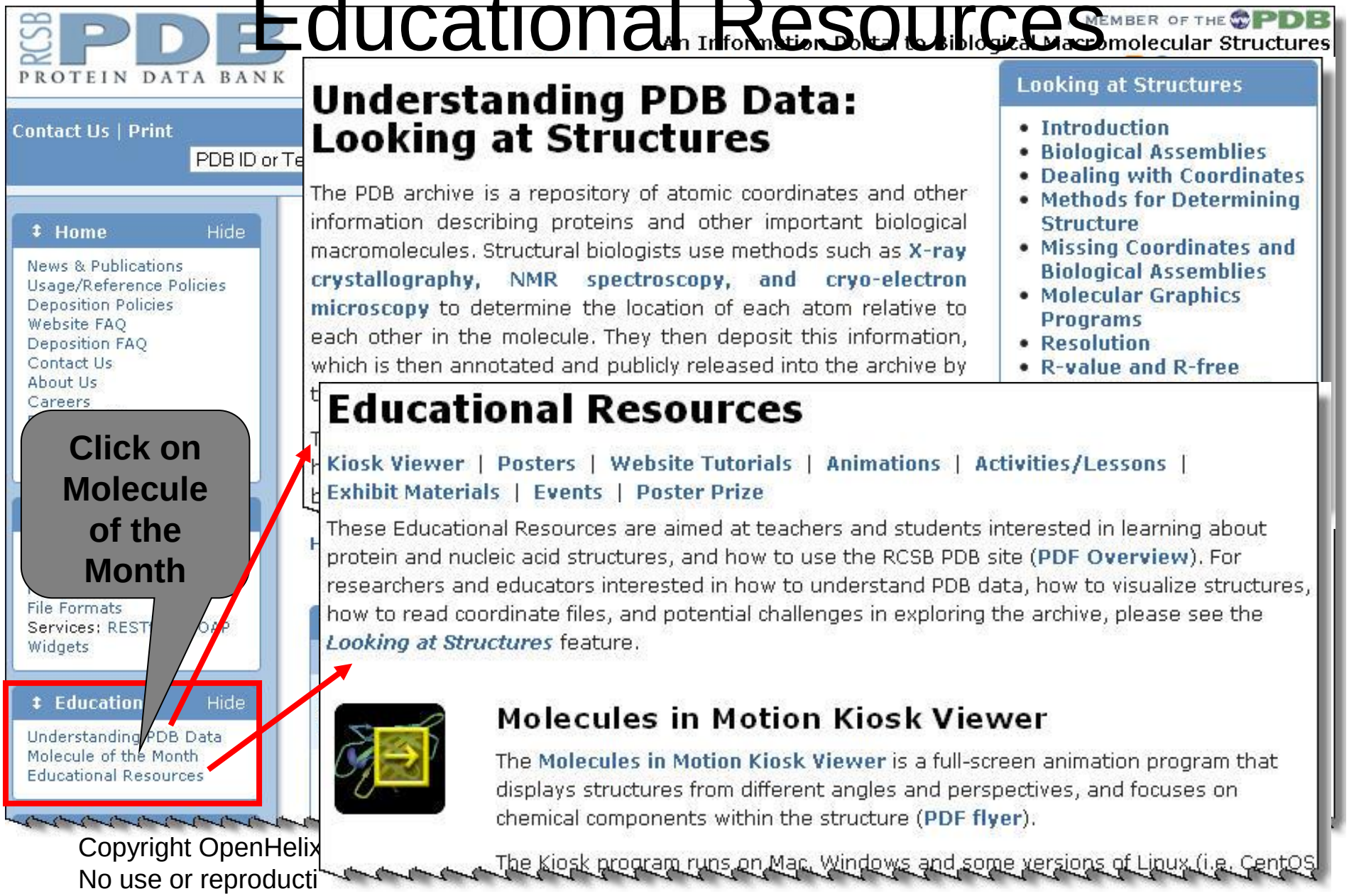


Methods Report

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



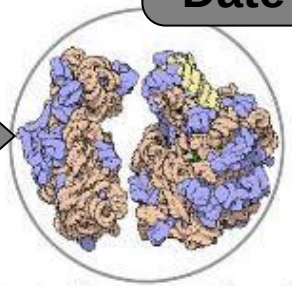
Educational Resources



Molecule of the Month

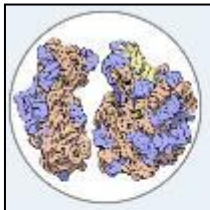
View by Title, Date or Category

Click on



Molecule of the Month: Structural View of Biology

List View of Archive By: **Title** | **Date** | **Category**



Protein Synthesis

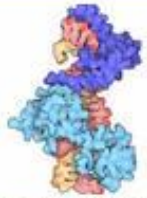
The major molecules of protein synthesis, ribosomes to folded proteins, are available

- Structure of Nucleic Acids
- DNA Maintenance
- Replication
- Transcription
- Translation
- Protein Folding - Modification - D

The *RCSB PDB's Molecule of the Month* series by ... intended for all types of users, but especially f... structure and function of the molecule, a discu...

Transcription

Enhanceosome

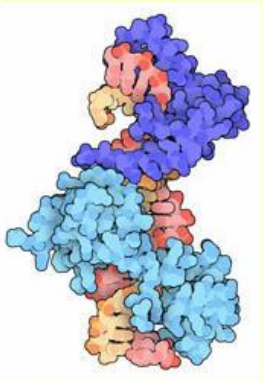


Oct and Sox Transcription

Poly(A) Polymerase

Aconitase and Iron Regulatory Protein 1

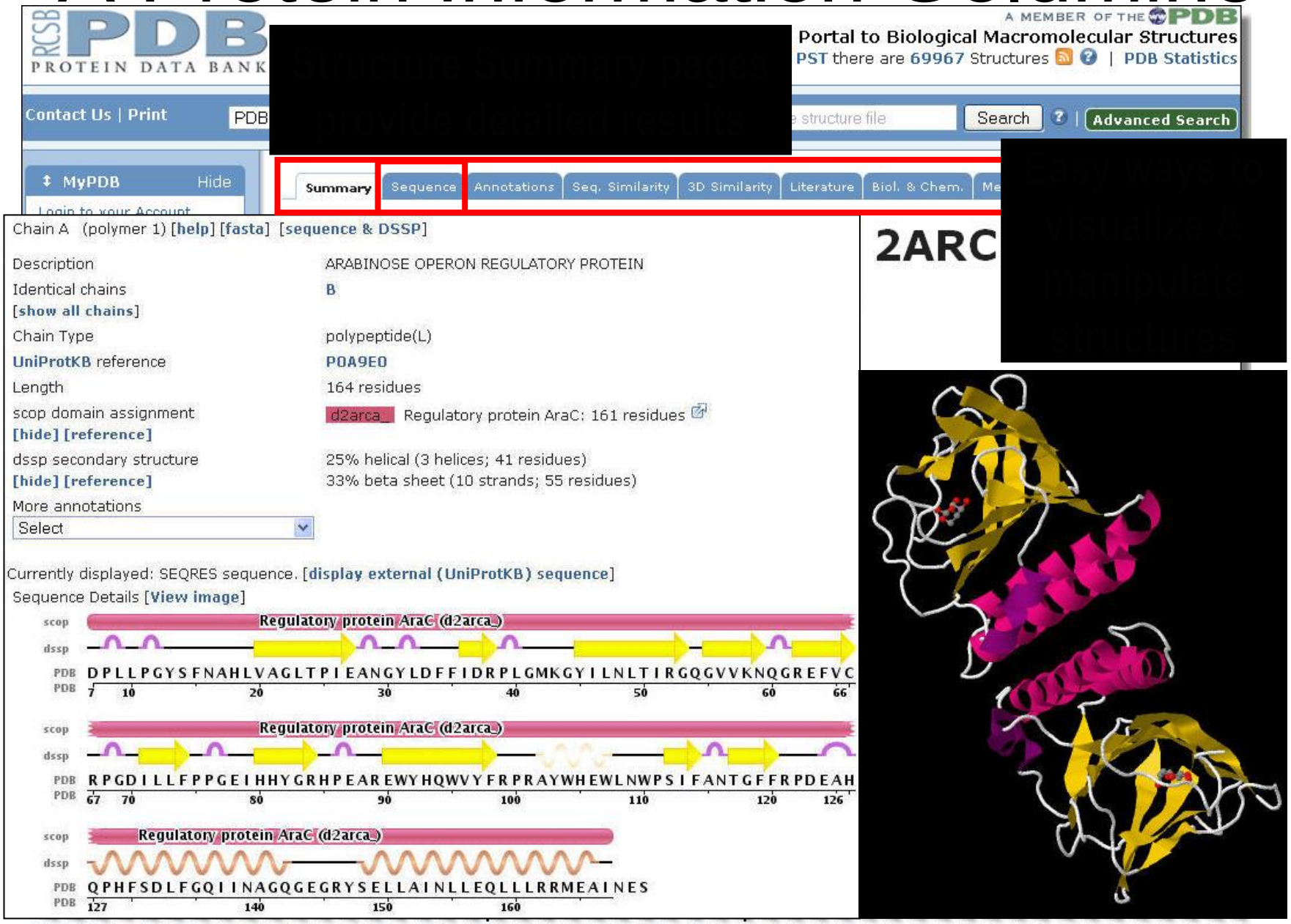
Oct and Sox Transcription Factors



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



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 MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters" and both of our Reports "Structure of the ABC transporter MsbA 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 MsbA structures (*1*, *2*, *5*) were incorrect in both the hand of the structure and the topology. Thus, our biological interpretations based on these inverted models for MsbA are invalid.

An in-house data reduction program introduced a change in sign for anomalous differences. This program, which was not part of a conventional data processing package, converted the anomalous pairs (I+ and I–) to (F– and F+), thereby introducing a sign change. As the diffraction data collected for each set of MsbA 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 MsbA structure was a consequence of the low resolution of the data as well as breaks in the electron density for the connecting loop regions. Unfortunately, the use of the multicopy refinement procedure still allowed us to obtain reasonable refinement values for the wrong structures.

The Protein Data Bank (PDB) files 1JSQ, 1PF4, and 1Z2R for MsbA and 1S7B and 2F2M for EmrE have been moved to the archive of obsolete PDB entries. The MsbA 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.

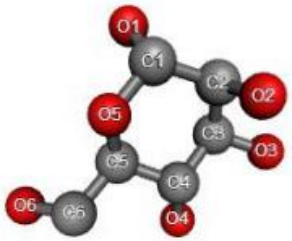
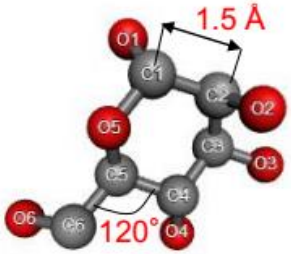
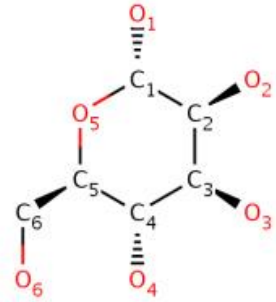
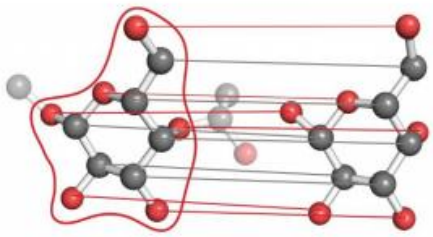
We very sincerely regret the confusion that these papers have caused and, in particular, subsequent research efforts that were unproductive as a result of our original findings.

GEOFFREY CHANG, CHRISTOPHER B. ROTH, CHRISTOPHER L. REYES, OWEN POURNELOS, YEN-JU CHEN, ANDY P. CHEN

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

References

1. G. Chang, C. B. Roth, *Science* **293**, 1793 (2001).
2. C. L. Reyes, G. Chang, *Science* **308**, 1028 (2005).
3. O. Pornillos, Y. J. Chen, A. P. Chen, G. Chang, *Science* **310**, 1950 (2005).
4. H. J. Dawson, K. P. Locher, *Nature* **443**, 180 (2006).
5. G. Chang, *J. Mol. Biol.* **330**, 419 (2003).
6. C. Ma, G. Chang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2852 (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**

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

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

Quality of data in PDB

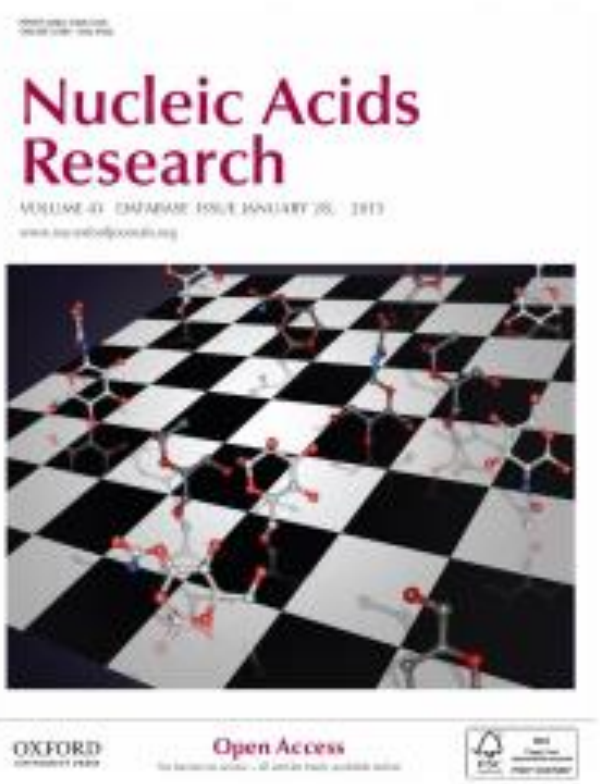
- Validation of ligands

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***

* Lütteke 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

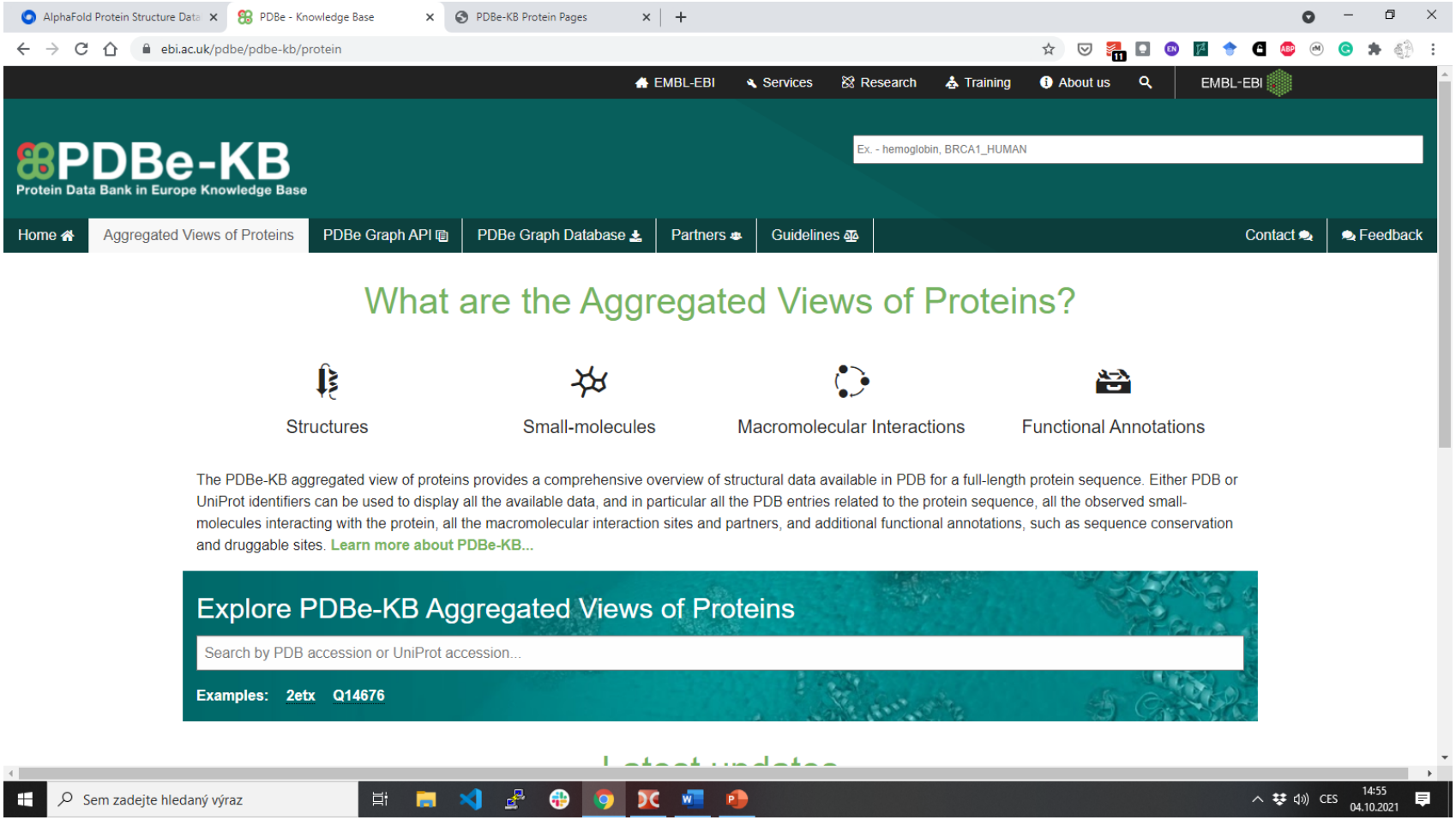
Structural bioinformatics tools for drug design

Extraction of biologically relevant information from structural databases

by

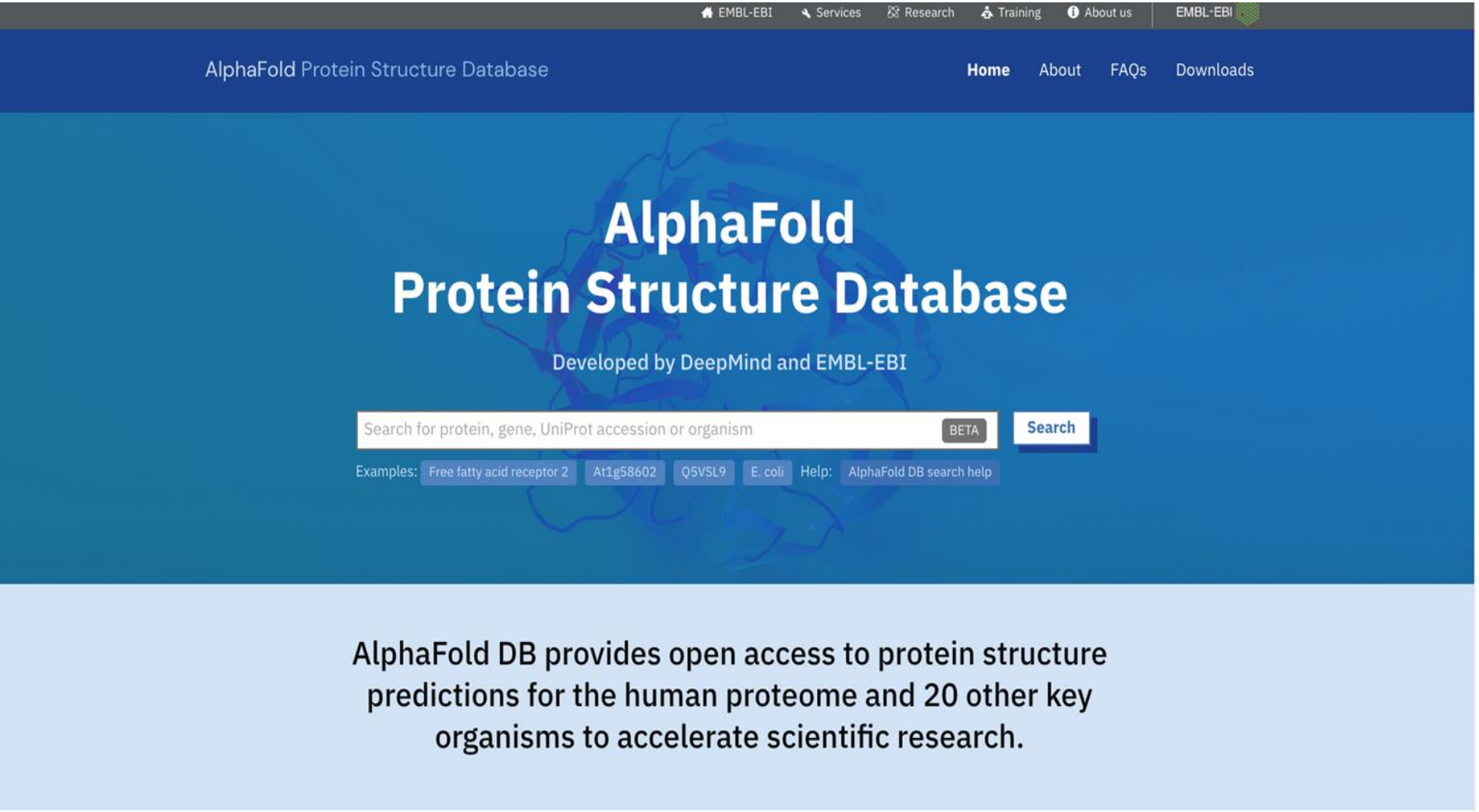
Jaroslav Koča, Radka Svobodová Vařeková, Lukáš Pravda, Karel Berka, David Sehnal, Stanislav Geidl and Michal Otyepka

PDBe-KB


AlphaFold Protein Structure Data
PDBe - Knowledge Base
PDBe-KB Protein Pages
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
CES
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)



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

Complete structures of 20 model organisms

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

SNW domain-containing protein 1

AlphaFold structure prediction

Download **PDB file** **mmCIF file** **Predicted aligned error**

Information

Protein	SNW domain-containing protein 1
Gene	SNW1
Source organism	Homo sapiens go to search
UniProt	Q13573 go to UniProt
Experimental structures	17 structures in PDB for Q13573 go to PDBe-KB
Biological function	(Microbial infection) Proposed to be involved in transcriptional activation by EBV EBNA2 of CBF-1/RBPJ-repressed promoters. go to UniProt

3D viewer

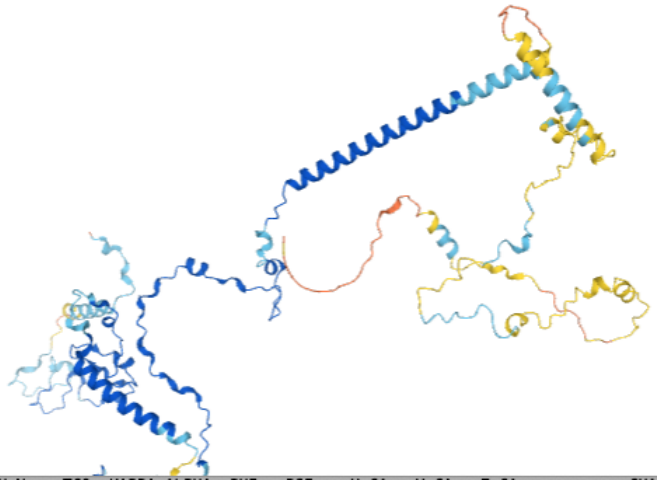
Model Confidence:

- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

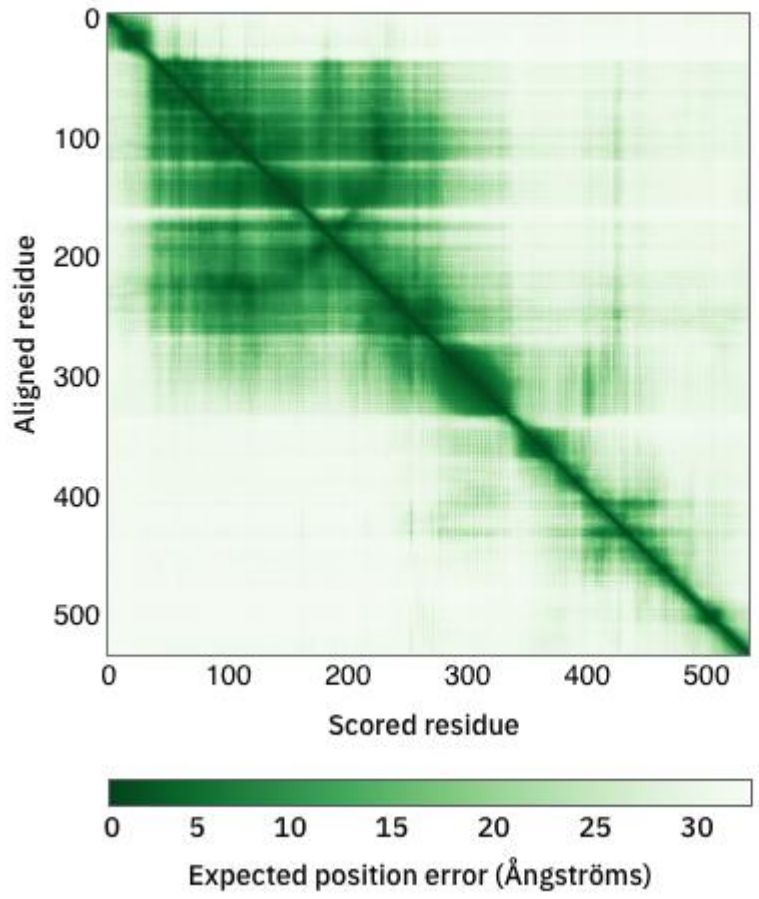
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

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

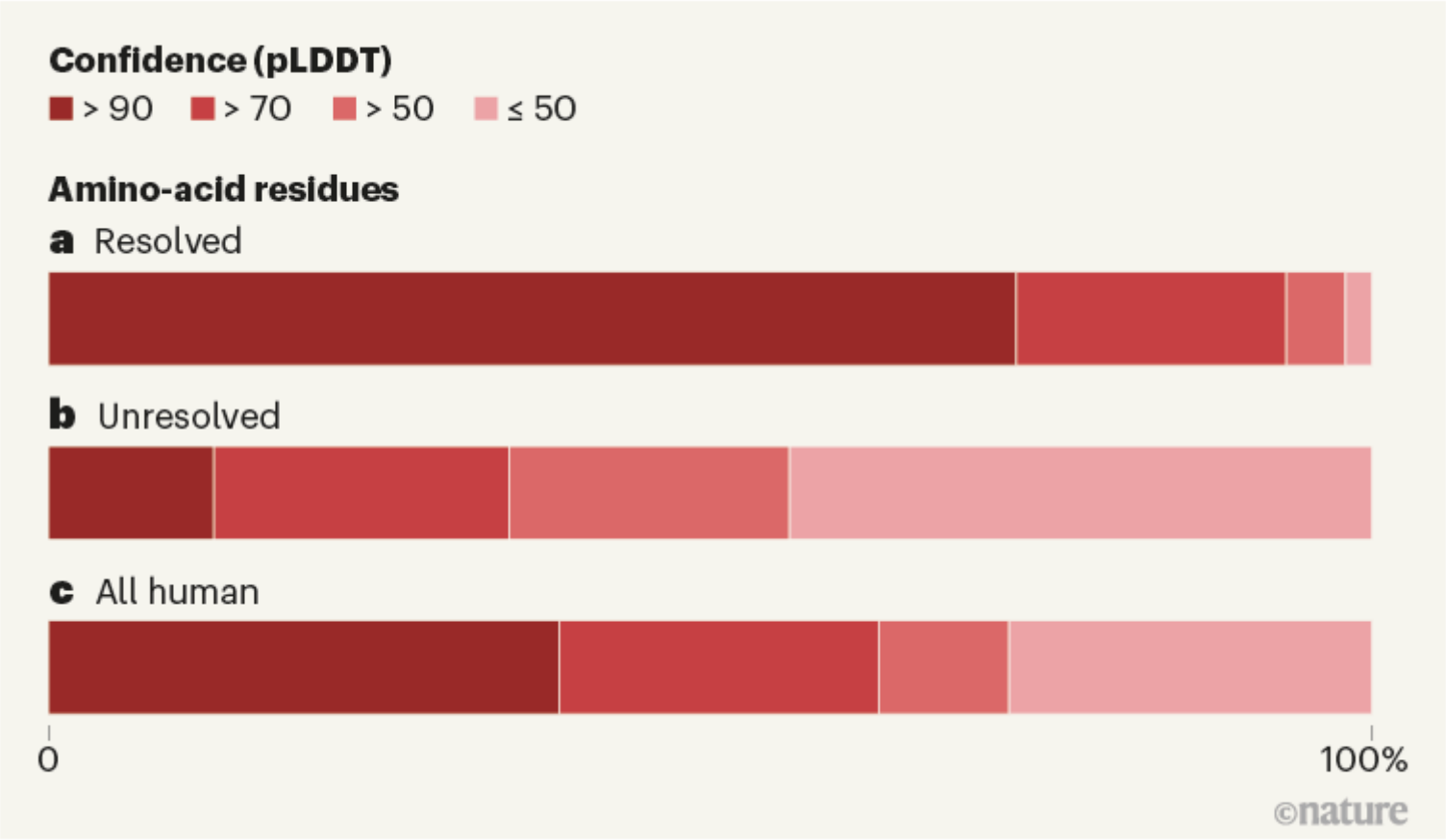
```
MALTSFLPAPTQLSQDQLEAEEKARSQRSRQTSLVSSRREPPPYGYRKGWIPRLLEDFGDGGAFPEIHVAQYPLDMGRKKKMSNALAIQVDSEGKIKYDAIARQGQSKDKVIYSKYTDLVPKEV
MNADDPDLQRPDEEAIKEITEKTRVALEKSVSQKVAAAMPVRAADKLAPAQYIRYTPSQQGVAFNSGAKQRVIRMVEMQKDPMEPPRFKINKKIPRGPPSPPAPVMHSPSRKMTVKEQQEWKIP
PCISNWKNAKGYTIPLDKRLAADGRGLQTVHINENFAKLAEALYIADRKAREAVEMRAQVERKMAQKEKEKHEEKLREMAQKARERRAGIKTHVEKEDGEARERDEIRHDRRKERQHDRNLSRA
```



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 lDDT-Cα metric (local superposition-free score for comparing protein structures and models using distance difference tests).

DĚKUJI ZA POZORNOST