



KATEDRA FYZIKÁLNÍ CHEMIE
UNIVERSITY PALACKÉHO V OLOMOUCI



INSTITUTE OF MOLECULAR AND
TRANSLATIONAL MEDICINE



6th Advanced *in silico* Drug Design KFC/ADD

Structure Bioinformatics Tools for Drug Design

Karel Berka



EMBL-EBI



UP Olomouc, 30.1.-3.2. 2023



INSTITUTE OF PHYSICS
National academy of Sciences of Ukraine



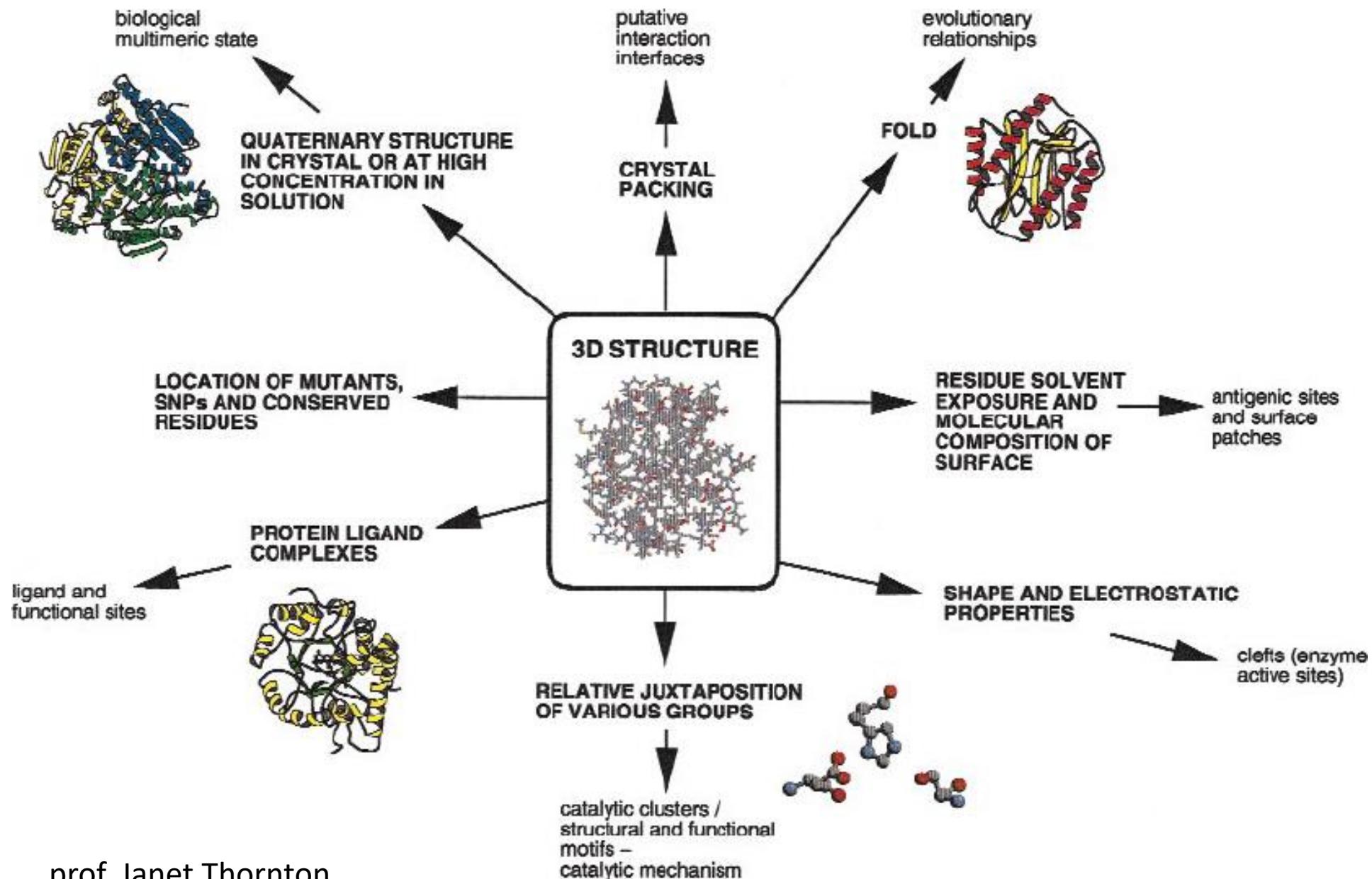
ÚOCHB AV
ČR
IOCB PRAGUE



Motto

Function of protein is given by its structure.

Use of Protein 3D Structure



STRUCTURAL DATABASES

“Century of Life Science Data”

Nucleic acid sequences



European Nucleotide Archive
ArrayExpress
Expression Atlas
...
...

Protein sequences



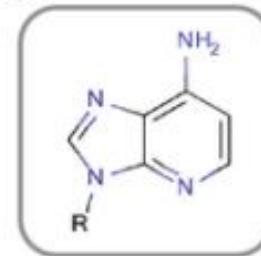
Uniprot

Protein (biomacromolecule) structures



Protein Data Bank

Drug-like molecules



PubChem
Zinc
DrugBank
PDB CCD
ChEBI
ChEMBL
...
...

<https://www.elixir-czech.cz/services>

R. Svobodová-Vařeková – habilitation lecture 2016

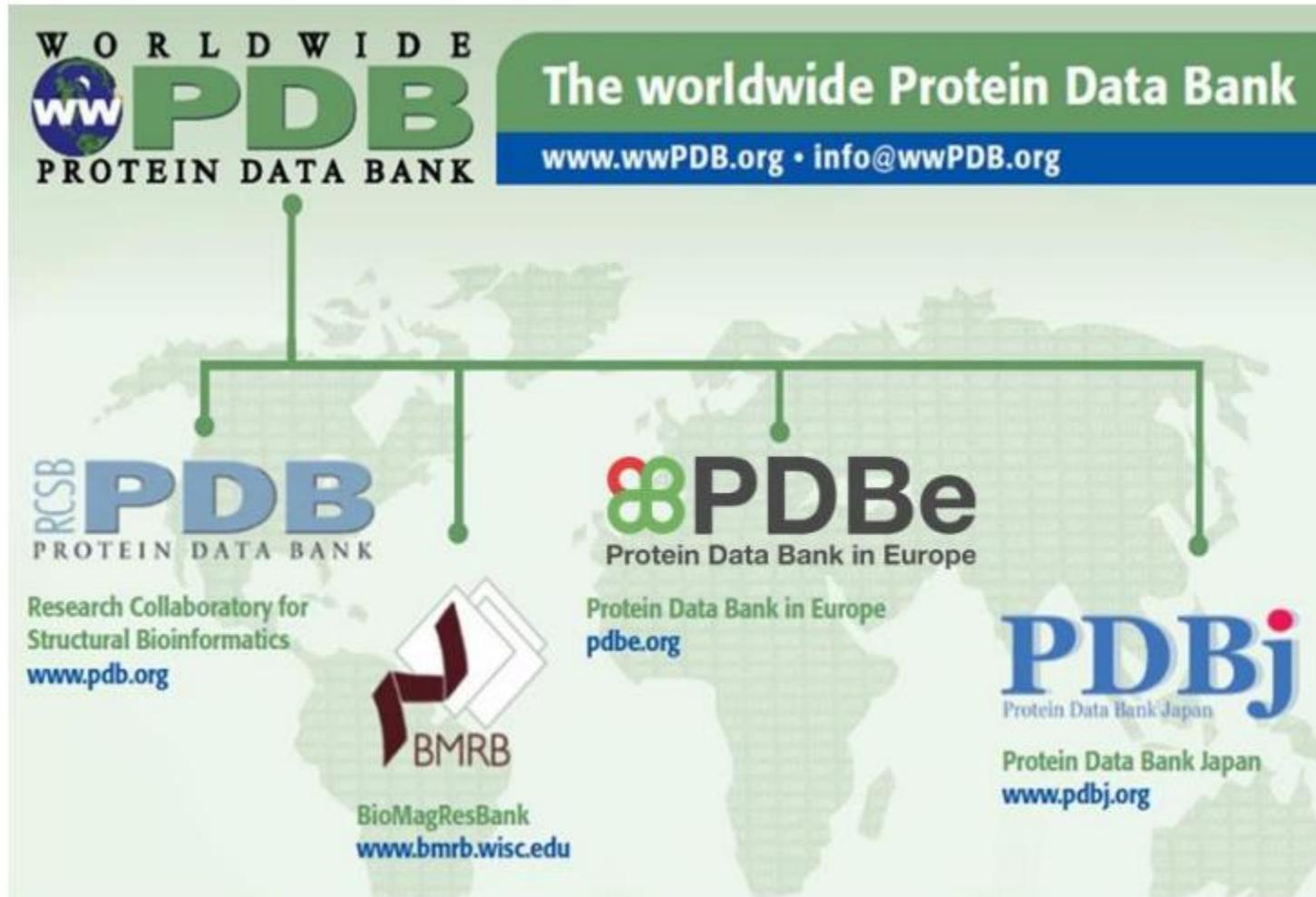


Databases, you ask?...

CATH	http://www.biochem.ucl.ac.uk/bsm/cath/
FSSP	http://www2.ebi.ac.uk/rdali/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/
Speciality databases	
HIV Protease Database	http://www-lbmc.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	http://www.rcsb.org/pdb/

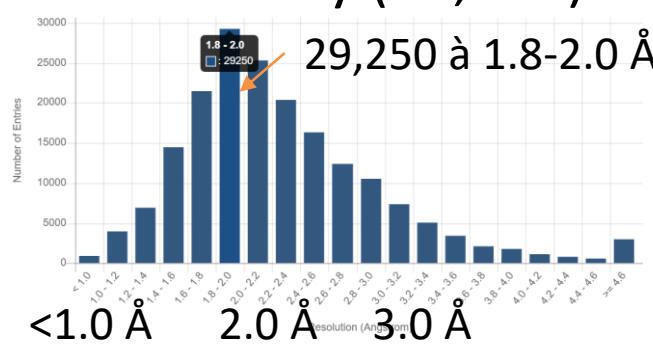
For more complete list of structural databases please refer to
[http://www.oxfordjournals.org/our_journals/nar/database/subcat/4/14.](http://www.oxfordjournals.org/our_journals/nar/database/subcat/4/14)

PDB – Protein Data Bank

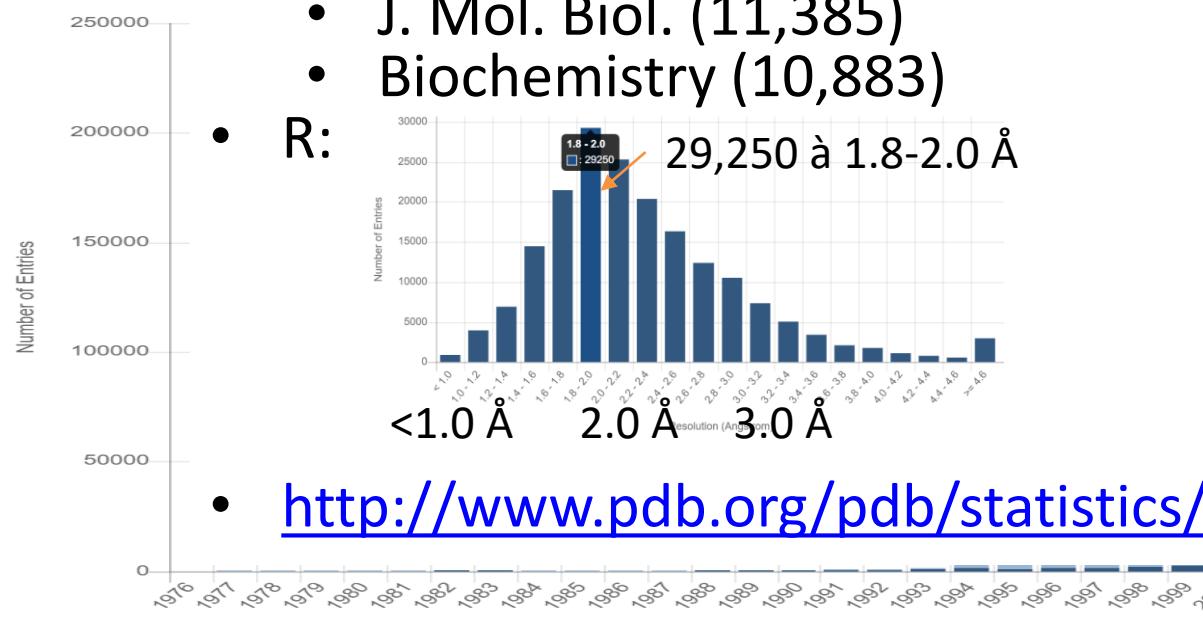


> 120k biomacromolecular structures

PDB Statistics (30.1.2023)

- Organism:
 - Homo Sapiens 60,931
 - Mus musculus 8,426
- Enzymes:
 - Hydrolases (44,563)
 - Transferases (37,626)
- Residues
 - 100-300 (60138)
- Journal
 - To be published (31,250)
 - J. Biol. Chem. (13,460)
 - J. Mol. Biol. (11,385)
 - Biochemistry (10,883)
- R:

- Structures **200,708**
 - X-ray 172,458 (+5.7%)
 - NMR 13,899 (+2.2%)
 - EM 14,022 (**+29.3%**)
 - Proteins 174,424
 - NA 2,284D + 1,691R
 - Protein/NA 11,253
 - +Ligands 39,079



PDBe

- PDB = Primary database
- Structures of proteins, nucleic acids, sugars, ligands,...
- Links to other databases (Uniprot)

The screenshot shows the PDBe website interface for PDBID 1r9o. The top navigation bar includes links for Services, Research, Training, and About us. A search bar is present, along with share and feedback buttons.

Search: X-ray diffraction, 2Å resolution, Quality

PDBe > 1r9o PDBID: Protein Data Bank in Europe - Bringing Structure to Biology

Primary Citation: Crystal Structure of P4502C9 with Flurbiprofen bound. Source organism: *Homo sapiens*. Primary publication: The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0 Å resolution. Wetter HB, Yano JK, Schoch GA, Yang C, Griffin KJ, Stout CD, Johnson EF. *J. Biol. Chem.* 279: 25630-7 (2004). PMID: 15182000.

Figures: X-ray diffraction, 2Å resolution, Quality

Ligands and Environments: 3 bound ligands: 3 x HBM, 3 x FLP, 1 x GOL.

Ligands: HBM, FLP, GOL.

Validation: No modified residues.

Experiments and Validation: Penetra Ratio, Stereoselectivity, Substrate selectivity, Enzyme kinetics, Enzyme mechanism.

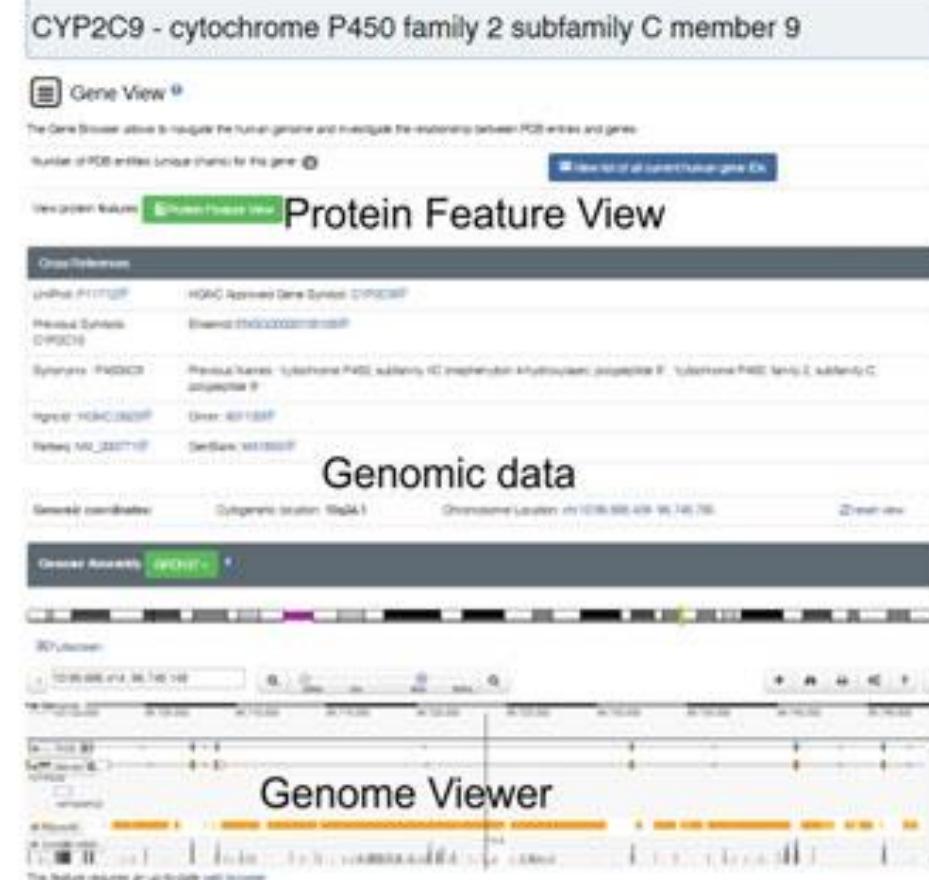
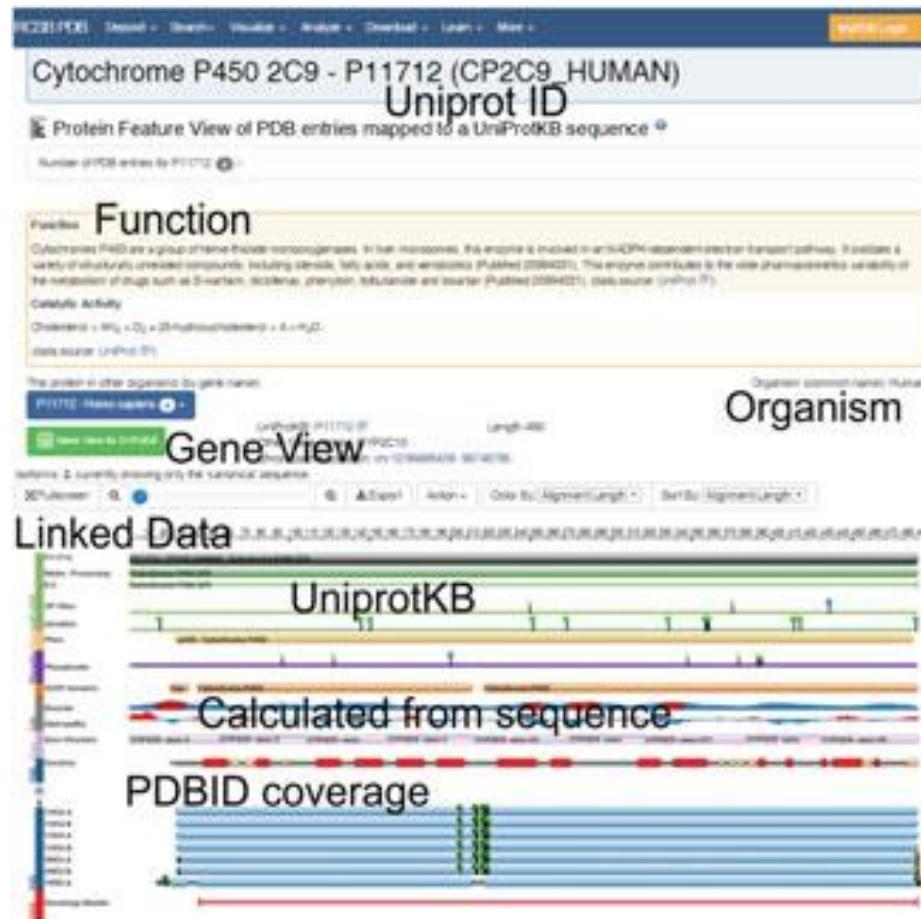
GO: heme binding, oxidation-reduction process, endoplasmic reticulum membrane.

Structure: Assembly composition: monomeric (preferred), Entry contents: 1 distinct polypeptide molecule, Macromolecule: Cytochrome P450 2C9.

Experimental Info: Spacegroups: R3, Unit cell: a: 91.047 Å, b: 91.047 Å, c: 169.481 Å, α: 90°, β: 90°, γ: 120°, R-values: R: 0.194, R_{work}: 0.191, R_{free}: 0.236, Expression system: *Escherichia coli*.

Annotation: Chain A, Length: 477 amino acids, Theoretical weight: 54.32 kDa, Source organism: *Homo sapiens*, Expression system: *Escherichia coli*, UniProt: P11712 (Residues: 18-490, Coverage: 97%), Gene names: CYPC10, CYPC9, Sequence domains: Cytochrome P450 [PF00067].

RCSB PDB



Other Structural DBs

	Classification	
CATH	Domain classification of structures	http://www.cathdb.info/
Pfam	Classification of sequence families	http://pfam.xfam.org/
	Flexibility and disorder	
PDB FLEX	Intrinsic flexibility in proteins	http://pdbflex.org/
DisProt	Database of Protein Disorder	http://www.disprot.org/
Pocketome	Encyclopedia of ensembles of druggable binding sites	http://www.pocketome.org/
	Detailed biomacromolecules	
OPM	Orientations of proteins in membranes	http://opm.phar.umich.edu/
NDB	Nucleic Acids Database	http://ndbserver.rutgers.edu
GFDB	Glycan Fragment Database	http://www.glycanstructure.org/

For more complete list of structural databases please refer to

http://www.oxfordjournals.org/our_journals/nar/database/subcat/4/14.

Summary Structures Ligands Interactions Annotations Similarity Publications Feedback ↗

What's new?

Cytochrome P450 3A4

Gene: CYP3A4 [Enzyme: EC 1.14.14.1](#) [Disease](#)

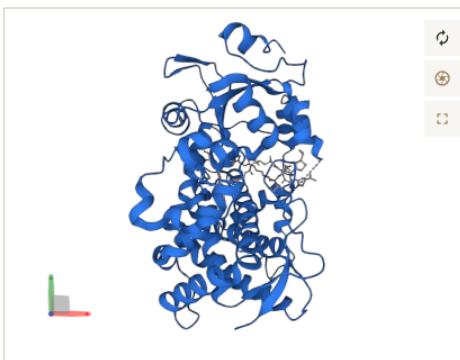
Organism: *Homo sapiens (Human)*

Synonyms: CYP3A3

Uniprot: P08684 [go to UniProt](#)

Biological function: A cytochrome P450 monooxygenase involved in the metabolism of sterols, steroid hormones, retinoids and fatty acids ([PubMed:10681376](#), [PubMed:11093772](#), [PubMed:11555828](#), [PubMed:14559847](#), [PubMed:12865317](#), [PubMed:15373842](#), [PubMed:15764715](#), [PubMed:20702771](#), [PubMed:19965576](#), [PubMed:21490593](#), [PubMed:21576599](#)). ... [+ \[show more\]](#) [go to UniProt](#)

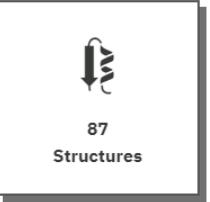
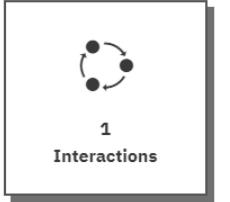
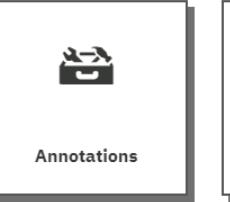
Representative structure for UniProt P08684
PDB chain with highest data quality, coverage and best resolution ⓘ



PDB chain shown: 4k9u A [go to PDBe](#)
UniProt residues 21 - 503
Coverage: 89%

3D view of superposed structures for region 1

Click on the icons below to view the relevant page:

-  87 Structures
-  83 Ligands
-  1 Interactions
-  Annotations
-  8 Similarity
-  281 Publications

[Download](#) [Download](#) [Download](#)

[3D view of superposed structures](#) [3D view of superposed ligands](#)

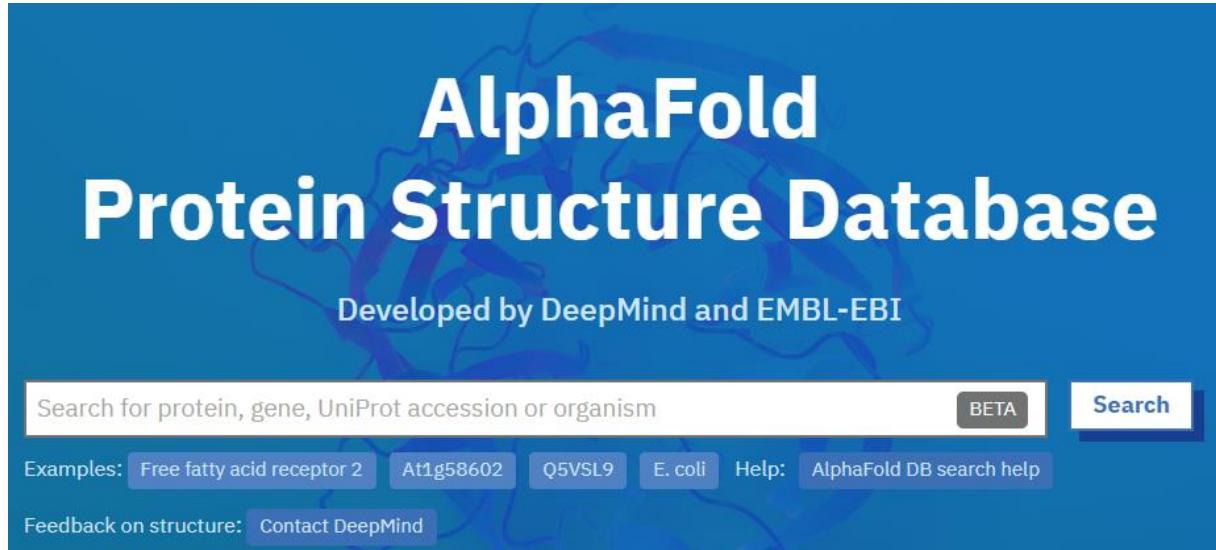
- Clusters by UNIPROT IDs
 - Structures
 - Ligands
 - Interactions
 - Annotations
 - Similarity
 - Publications

<https://www.ebi.ac.uk/pdbe/pdbe-kb>

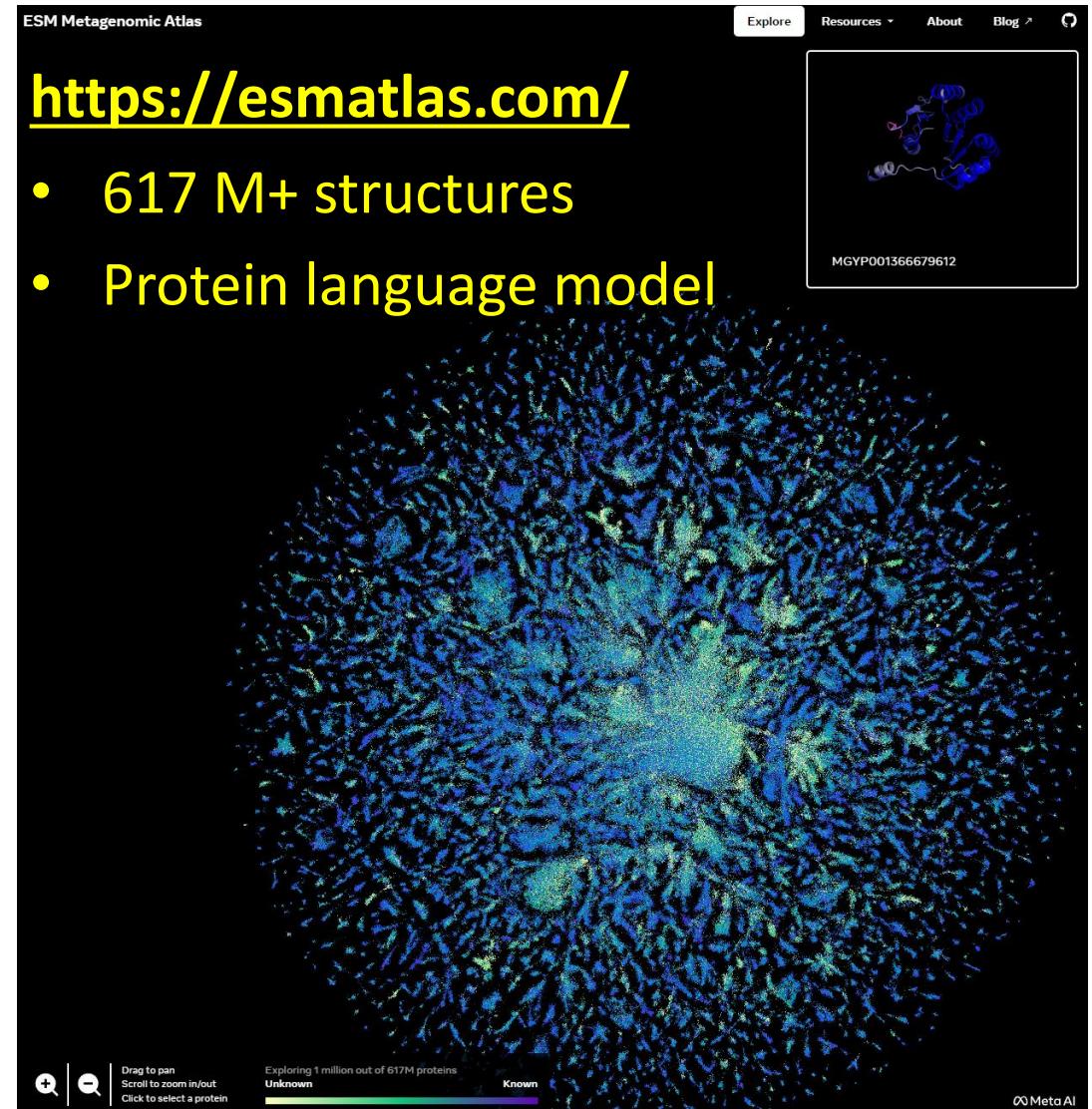
AlphaFoldDB + ESM Metagenomic Atlas

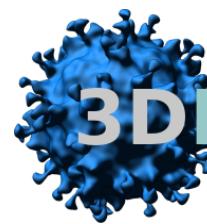
<https://wwwalphafold.ebi.ac.uk/>

- 200 M+ computed structures



- 16 model organisms (e.g. Homo Sapiens)
- 33 infective organisms (e.g. M. tuberculosis)
- 500k SwissProt, 17k MANE





3DBIONOTES-WS

West-Life
elixir INB
SPAIN

see annotations on structure

← → C ⓘ https://3dbionotes.cnb.csic.es/#query

3DBIONOTES-WS HOME - SUBMIT - NETWORK - QUERY - HELP elixir INB SPAIN

CRYSTAL STRUCTURE OF HUMAN MICROSOMAL P450 3A4

1TQN PROTEINS IN THIS MODEL: A - CYP3A4, Cytochrome P450 3A4

Domains & sites

Molecule processing

PTM

Sequence information

Structural features

Topology

Epitopes

Domain families

Disordered regions

Residue accessibility

Mol Probit

PDB-REDO

Structure coverage

Proteomics

Antigenic sequences

Variants

CYP3A4 - Cytochrome P450 3A4 - Homo sapiens - P08684

TOOLS VIEW

1 50 100 150 200 250 300 350 400 450 500 503

ProtVista i

The screenshot shows the 3DBIONOTES-WS web interface. On the left, there is a large ribbon diagram of the CYP3A4 protein structure. Below the ribbon diagram, there is a sequence alignment and a sequence logo. On the right, there is a detailed annotation track for residues 1 to 503. The annotation track includes various colored bars and symbols representing different types of annotations such as domains, PTM sites, and residue accessibility. The top right of the interface has a 'TOOLS' and 'VIEW' menu. The bottom of the interface shows a Windows taskbar with several open tabs and icons.

<https://3dbionotes.cnb.csic.es/ws/database>

PROTEIN FUNCTION DETECTION

Function Definition

EC – enzyme classification

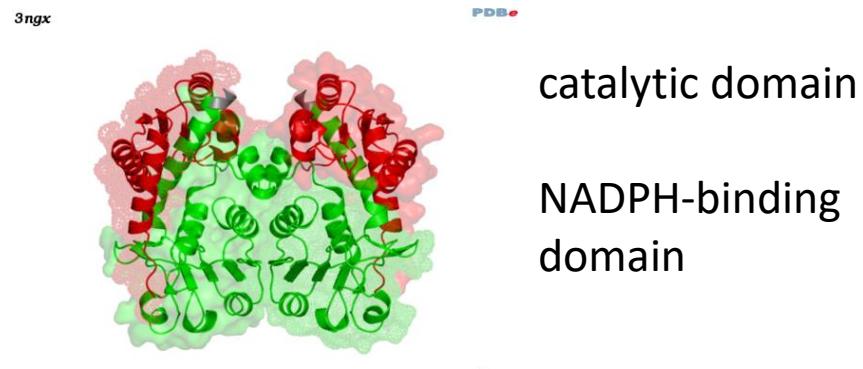
(4 levels of detail)

- Oxidoreductases (EC 1)
 - substrate is a donor of hydrogen or electrons
 - what substrate
 - acceptor-type
- only enzymes
- 1 enzyme – more functions
 - methylenetetrahydrofolate dehydrogenase/cyclohydrolase (EC 1.5.1.5 and EC 3.5.4.9)

GO - Gene Ontology

genome annotation – more generic

- **Biological process**
 - generic (cell growth)
 - specific (glycolysis)
- **Molecular function**
 - generic (enzyme)
 - specific (hexokinase)
- **Cell compartmentalisation**
 - where it works (ER, cytosol)



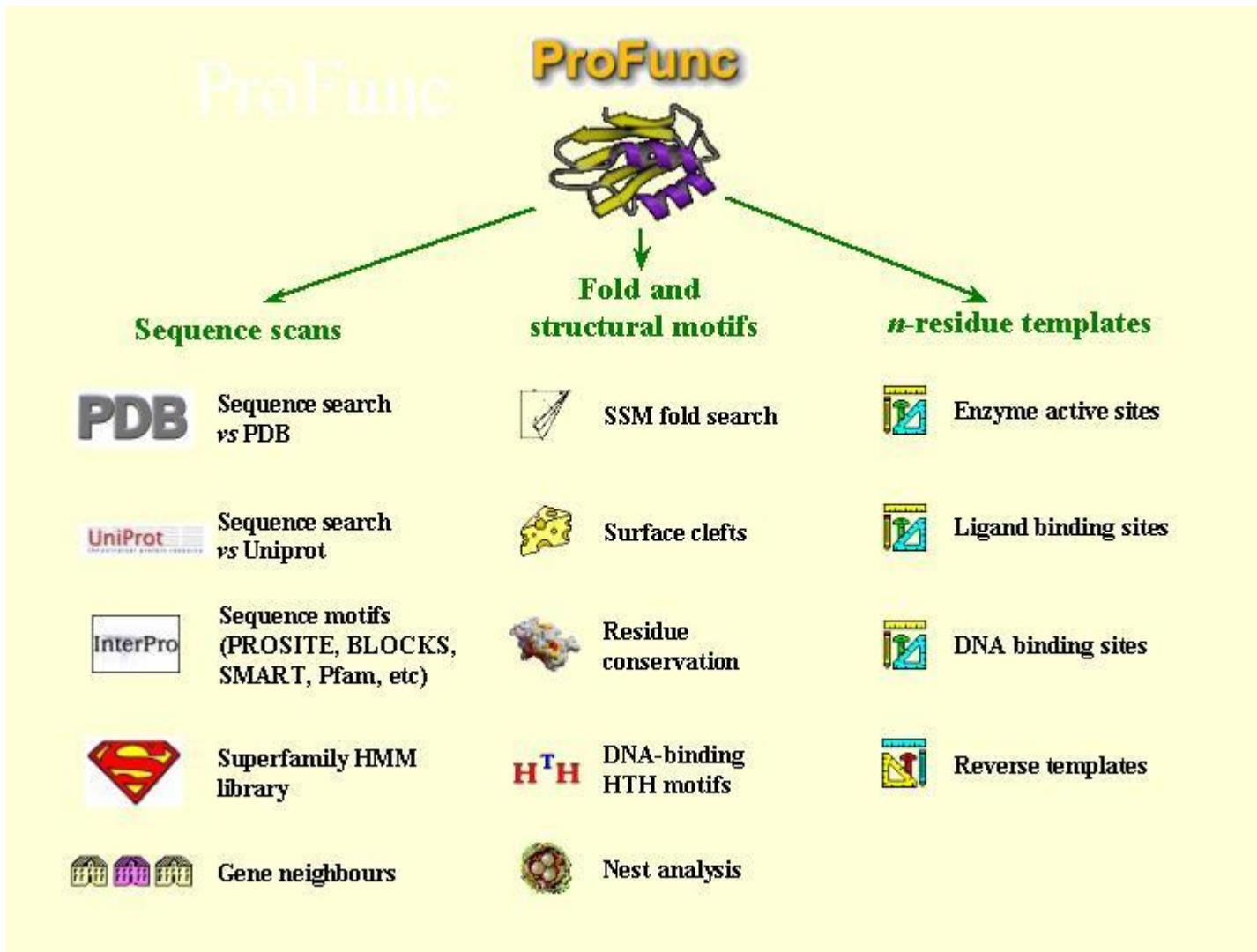
Function Search - Possibilities

- Sequence
 - below 40% identity is function less conserved
(exception CYP)
- Fold matching
 - structure alignment (DALI, FATCAT) against annotated structural DBs (CATH, SCOP)
- 3D motives
 - active site position (TESS, PDBSiteScan, PatternQuery)
- Active sites
 - geometric similarity (SURFNET)
 - physical properties (SURF'S UP!, SiteEngine)
 - ! same substrate can have multiple sites – ATP

Function Search – Possibilities II

- Phylogenetic analysis
 - search for conserved residues by multiple alignment
(ConSurf)
- DNA binding
 - Helix-turn-helix motiv (HTHquery)
- Ligands
 - similar sites as in other crystals in PDB (PDBsum, MSDsite)
- Channel annotation
 - Towards active site – substrate channelling (Caver, MOLE)

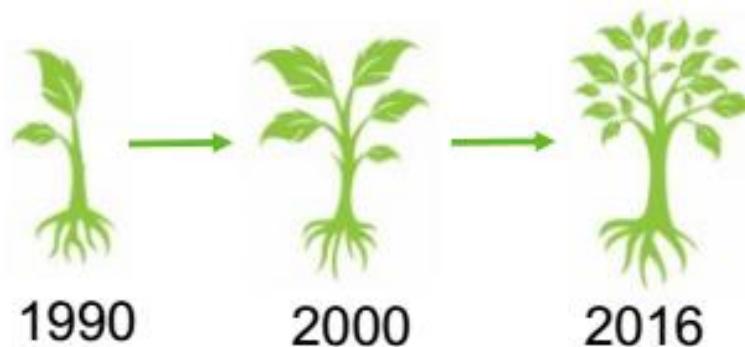
Function Search - ProFunc



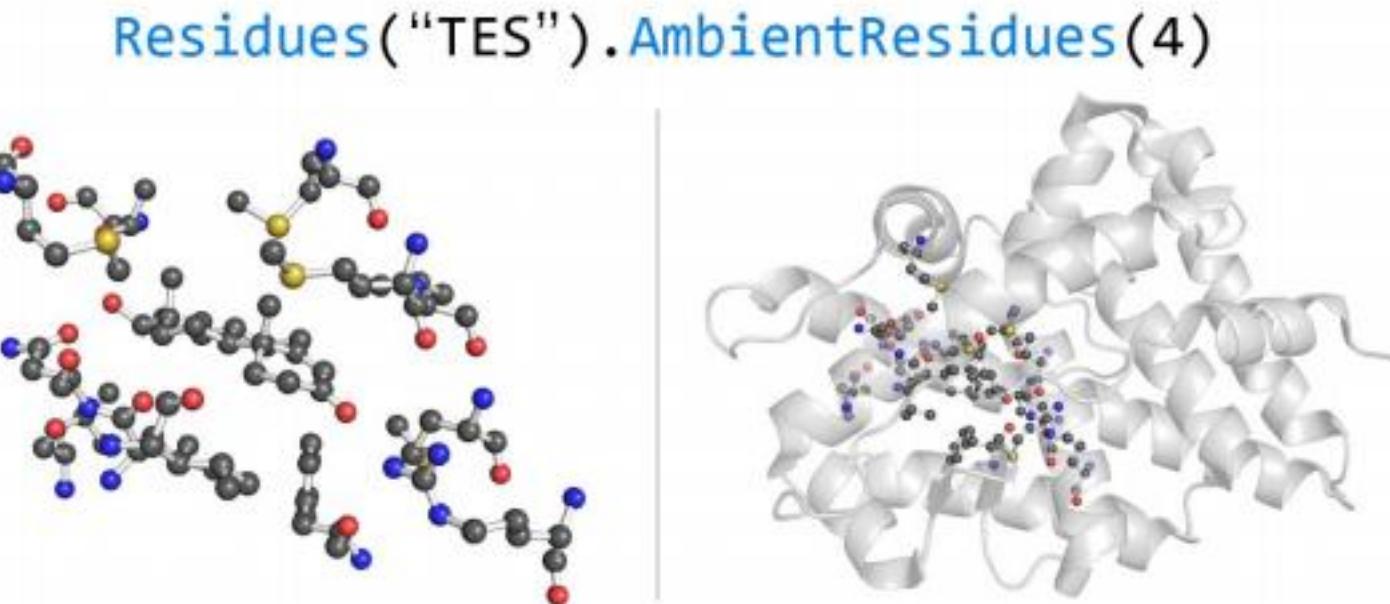
PATTERNS

Pattern Detection: Why?

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
Methemoglobin	4	11	37
Can we perform a comparison / analysis?	No	?	Yes



Pattern Detection: Approaches



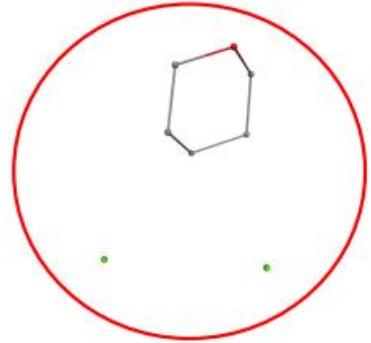
- Software: OpenStructure, Protein segment finder, PatternQuery*

<https://webchem.ncbr.muni.cz/Platform/PatternQuery>

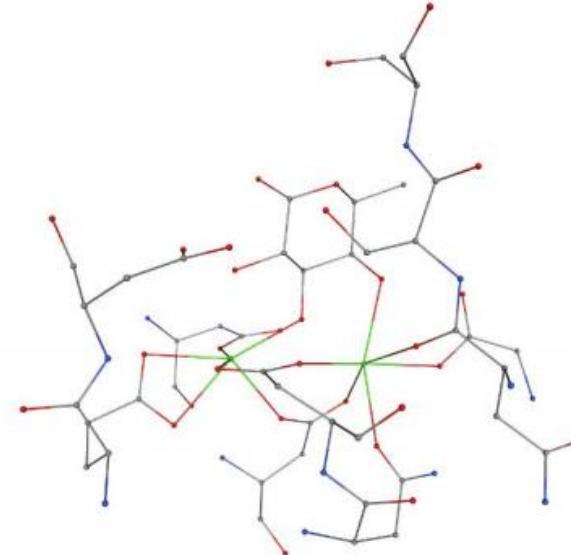
Sehnal et al., Nucl. Acids Res. (2015)

Pattern Detection Example

- sugar binding site from *Pseudomonas aeruginosa* lectin



```
Near(4,  
    Rings(5 * ["C"] + ["O"]),  
    Near(4, Atoms("Ca"), Atoms("Ca"))  
)  
.AmbientResidues(4)
```



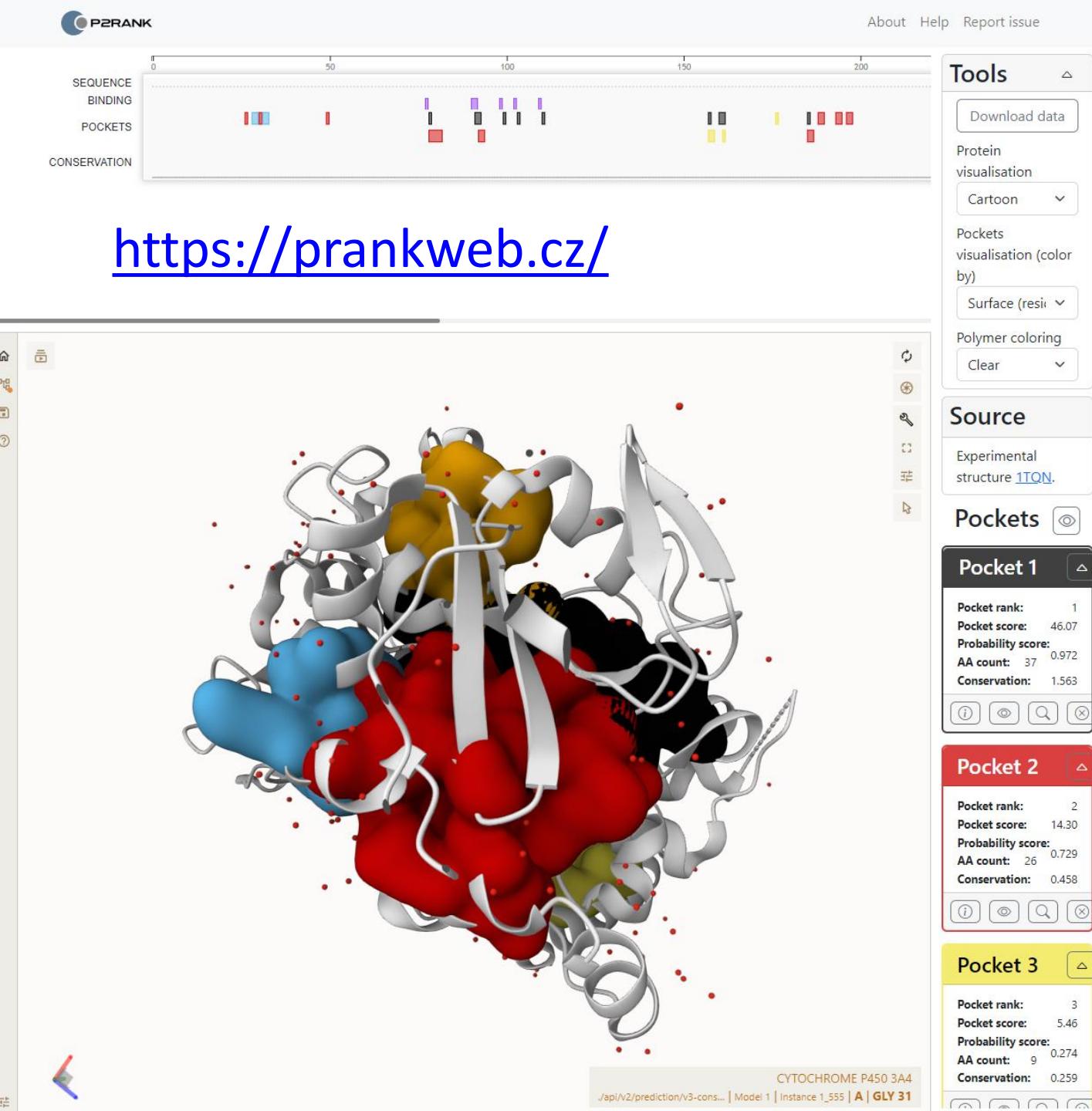
- Results (38 entries from PDB):
 - *Pseudomonas aeruginosa* (27 entries) or its synthetic (2 entries)
 - *Burkholderia cenocepacia* (4 entries)
 - *Ralstonia solanacearum* (2 entries)
 - *Chromobacterium violaceum* (2 entries)
 - *Bacillus subtilis* (1 entry)

ACTIVE SITE, CLEFTS, CHANNELS

Active Site

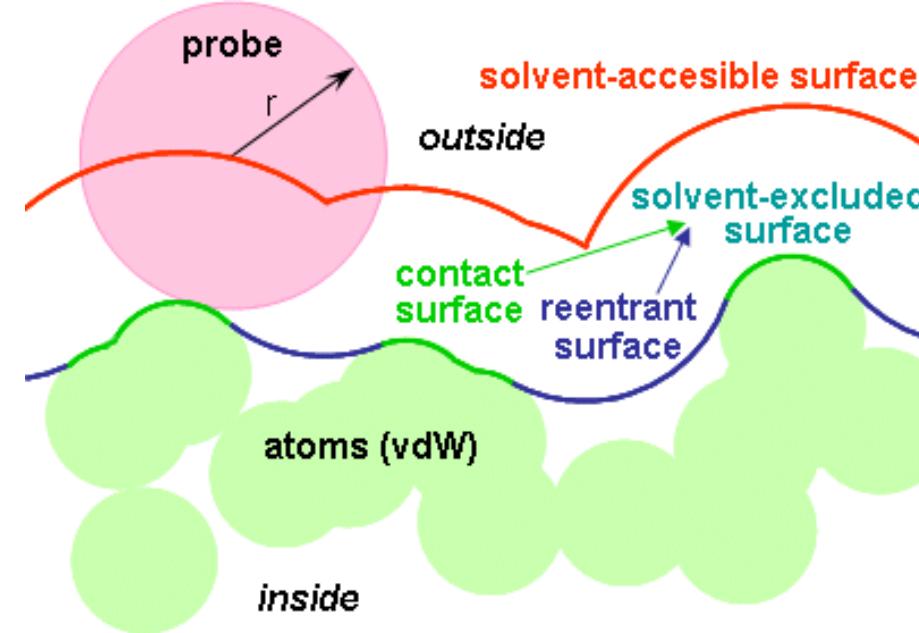
- Determines function

- Size and shape
 - Large (protein-protein, ev. protein-DNA)
 - PPI-PRED
 - small (clefts, ligands) - Surface, or Deep-inside (channels)
 - SURFNET, PocketFinder, Q-SiteFinder
- Phys chem properties
 - hydrophobic patches, charged surfaces
- Residue conservation – evolution trace method
 - ConSurf
- meta methods
 - aka ProFunc
- Machine learning
 - E.g. [Blast2GO](#), [P2Rank](#)



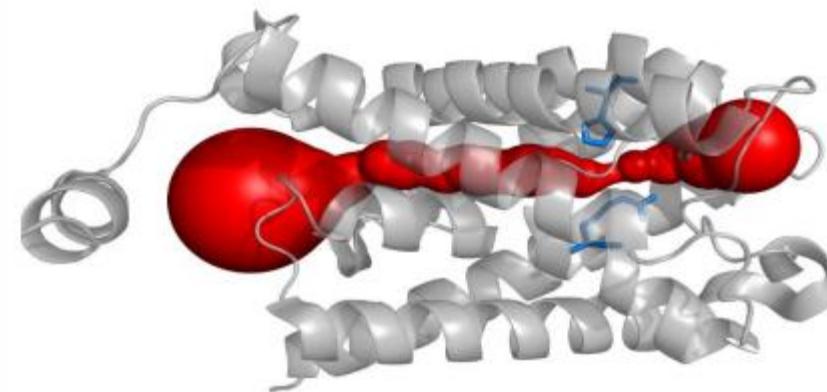
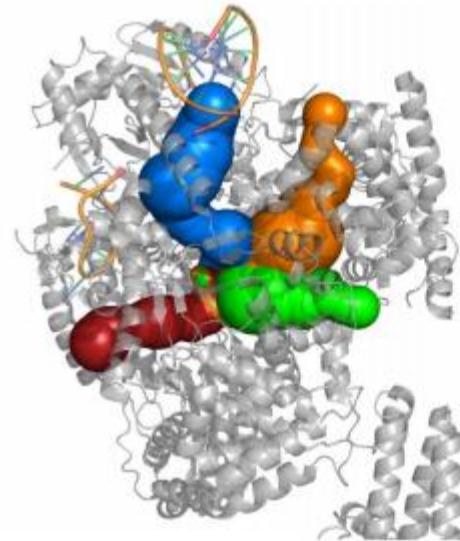
Cleft search

- SAS a SES
 - solvent accessible surface
 - solvent excluded surface
- Cleft search
 - Pocket-Finder
 - Slicing of receptor in several orientations
 - Closed spaces – cavities
 - Partially closed – pockets
 - QSiteFinder
 - grid search with generic probe (methan)



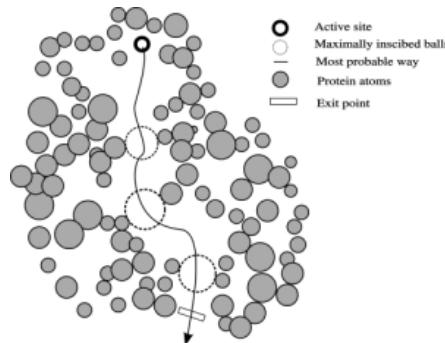
Channel Detection

- important role in biomolecular function – co-determine specificity



Channel Detection: Caver

- Caver



Dijkstra's algorithm
for optimal pathway search
on grid
or on Voronoi mesh

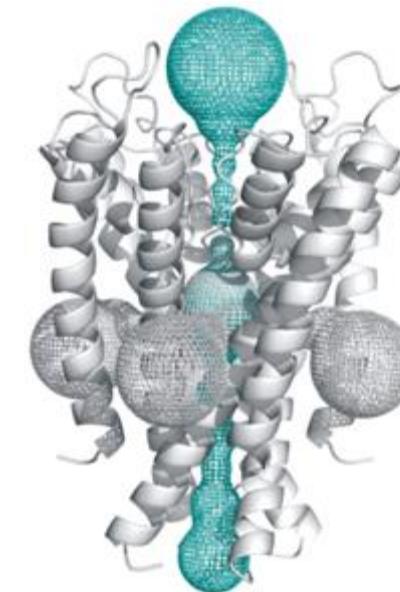
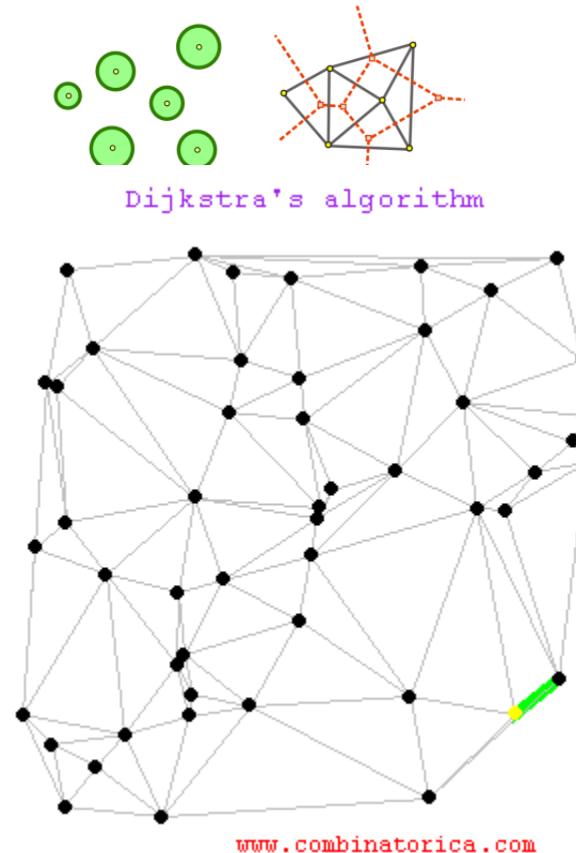
<http://loschmidt.sci.muni.cz/caver/index.php>

Petrek, M. et al - CAVER: a new tool to explore routes from protein clefts, pockets and cavities, *BMC Bioinformatics* 2006, 7:316
Beneš, P. et al. - CAVER 2.1 software, 2010.

Channel Detection: MOLE

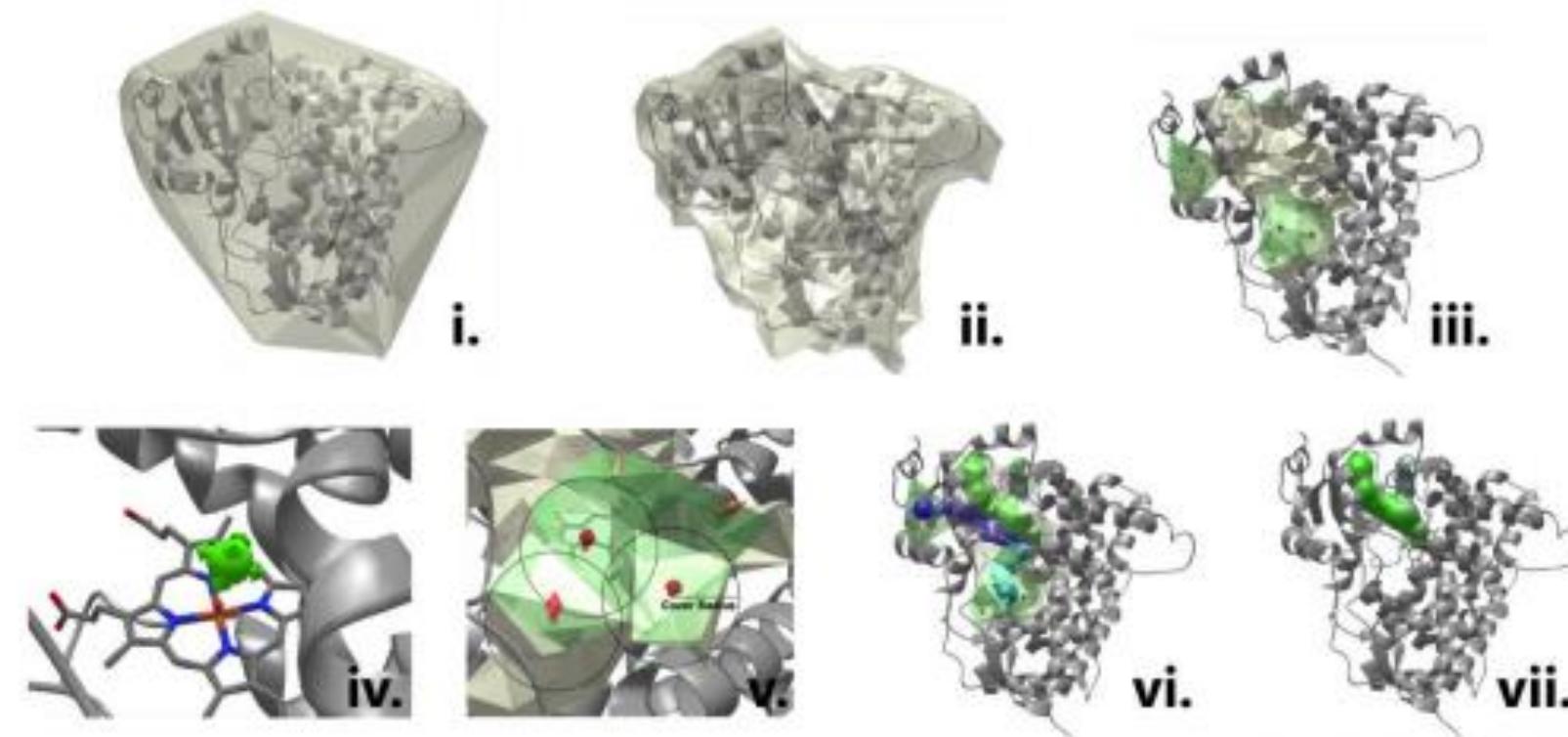
MOLE

Atoms maps => Voronoi diagram



<http://mole.chemi.muni.cz/online/current/>
<http://mole.upol.cz>

Channel Detection: MOLE 2.0



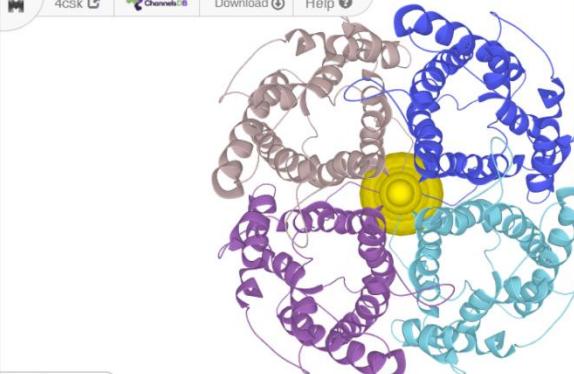
* Sehnal, Svobodová, et al., *J. Cheminf.* (2013)

** Berka, ..., Svobodová, et al., *Nucl. Acids Res.* (2012)

MOLEonline – mole.upol.cz

- Detection and analysis of channels, tunnels and pores

[Home](#) [ChannelsDB](#)
[PDBe](#) [Reports](#) [Help form](#)



Protein Sequence:

Chain A	3	28	43	68
SEFKKKLFWRRAVVAEFLATL	FV	ISIGSALGFKPVGN	NNTAVQD	VKVS
83	103	123	143	163
HLPNPAVTLLGQLLSCQISIF	FRALMYIIAQCVGAIVATAILS	GTTSLTONSLGRNDLADGVNS	SGQGLGIEII	183
183	203			

Channel profile: Tunnel radius: Radius Color by: Charge

Tunnel radius: Radius Color by: Hydropathy

LiteMol viewer

Selection Path (P1C0), Length: 50.9 Å

Channels Paths (1) All None Path (P1C0) Length: 50 Å

Origins Computed

Cavities Surface (1) Cavities (1)

Membrane Membrane

Current selection List of Channels

Interactive origins Surface, Cavities

Membrane position

Sequence

Submissions details

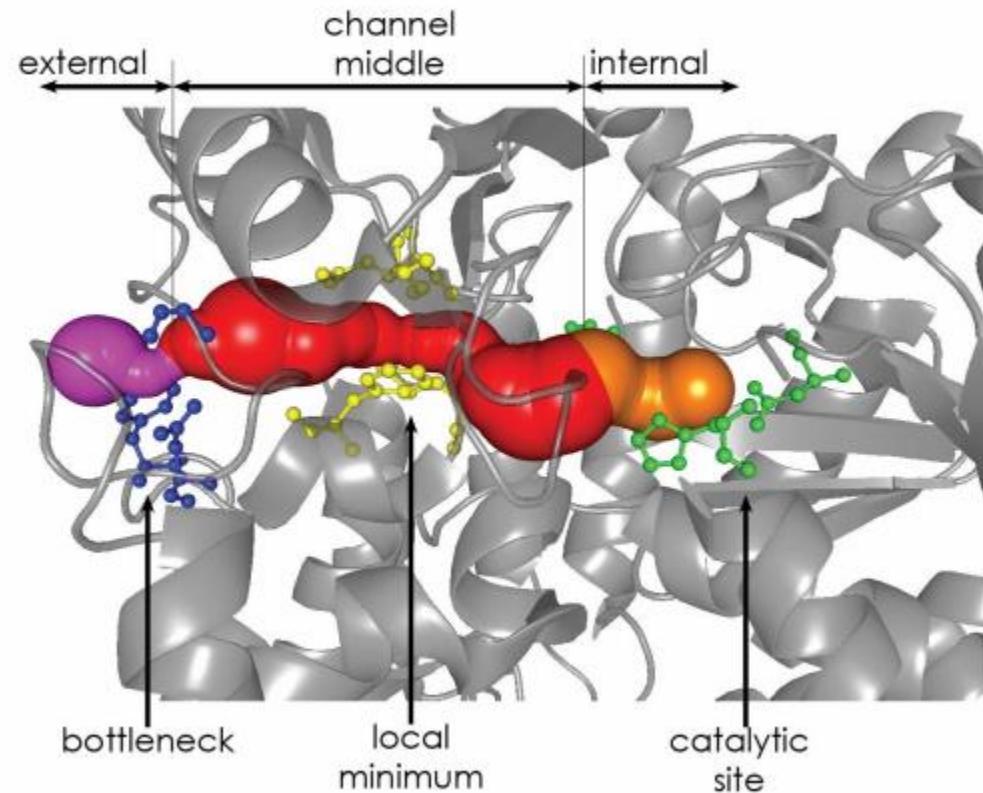
Mode switch

Parameters

Submissions and ChannelsDB data

Anatomy of Channels

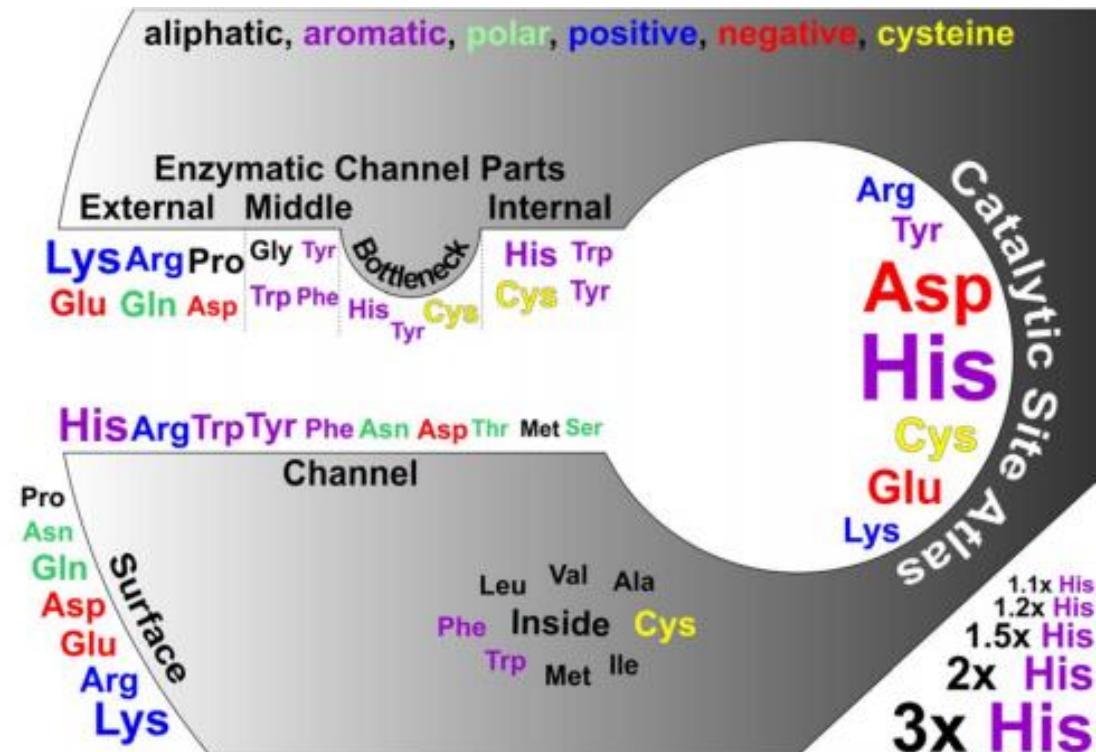
- **Methodology:**
 - Analysis of all the enzyme structures from PDB (4.306 structures)
 - Calculation of their properties via MOLE
 - Statistical analysis



Pravda L, Berka K, Svobodová Vařeková R,
Sehnal D, Banáš P, Laskowski RA, Koča J,
Otyepka M: Anatomy of enzyme
channels. *BMC Bioinf.*, 15(1), 379, 2014.

Anatomy of Channels

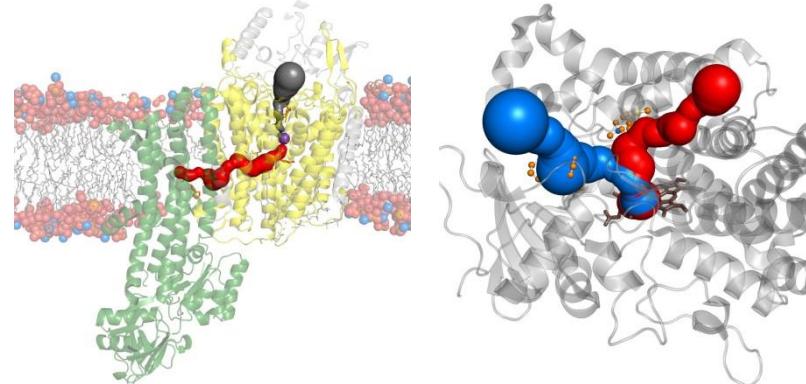
- **Results*:**
 - 64% of enzymes have at least one channel
 - Median channel length was 27.7 Å
 - Most of the channels are polar
 - An average channel is lined by 2 negative and 2 positive AAs
 - gating AAs enriched:



Pravda L, Berka K, Svobodová Vařeková R, Sehnal D, Banáš P, Laskowski RA, Koča J, Otyepka M: Anatomy of enzyme channels. *BMC Bioinf.*, 15(1), 379, 2014.

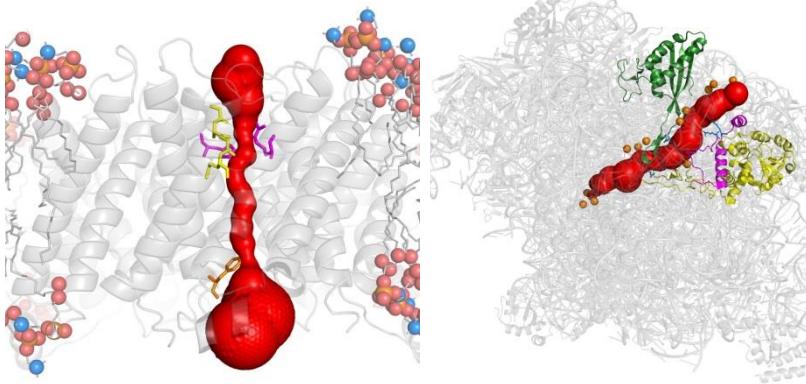
ChannelsDB

- Channel annotations for PDB
 - Reviewed from literature
 - Tunnels from [Catalytic Site Atlas](#)
 - Tunnels from enzyme cofactors (e.g. HEM or FAD)
 - Transmembrane pores
- Calculated with MOLE^{1,2}



KdpFABC complex
pore (5mrw)

Cytochrome P450
2D6 tunnel (3tbg)



Aquaporin (1ymg)

Ribosomal exit
tunnel (1jj2)

<http://ncbr.muni.cz/ChannelsDB>

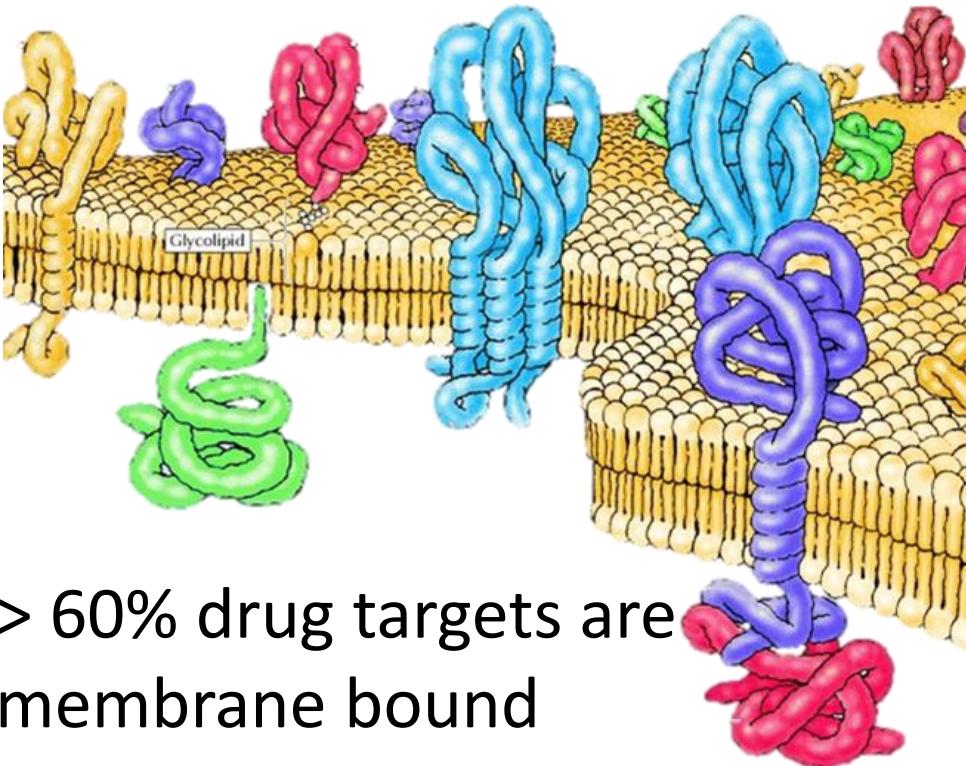
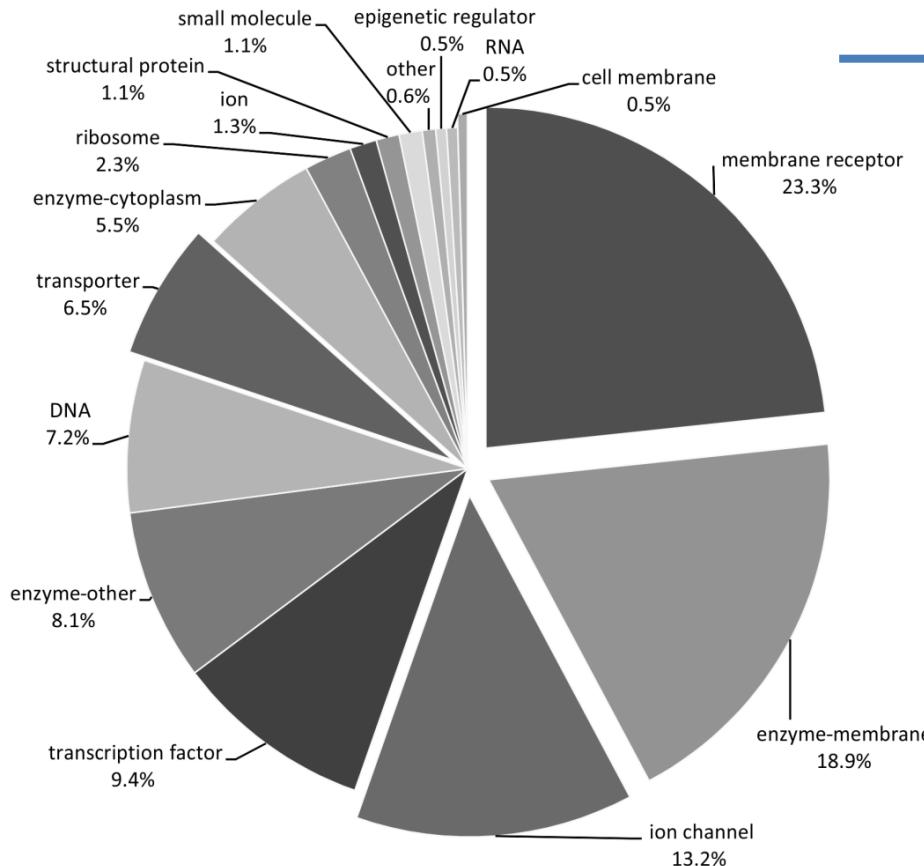
1. Berka K et al, *Nuc. Acids Res.*, 40(W1), W222-W227, 2012

2. Sehnal D et al., *J. Cheminform*, 5(1), 39, 2013

3. Pravda L et al. *Nuc. Acids Res.* 46(D1), D399–D405, 2018. doi: 10.1093/nar/gkx868

SPECIALTY SERVERS – MEMBRANE PROTEINS, DISORDER

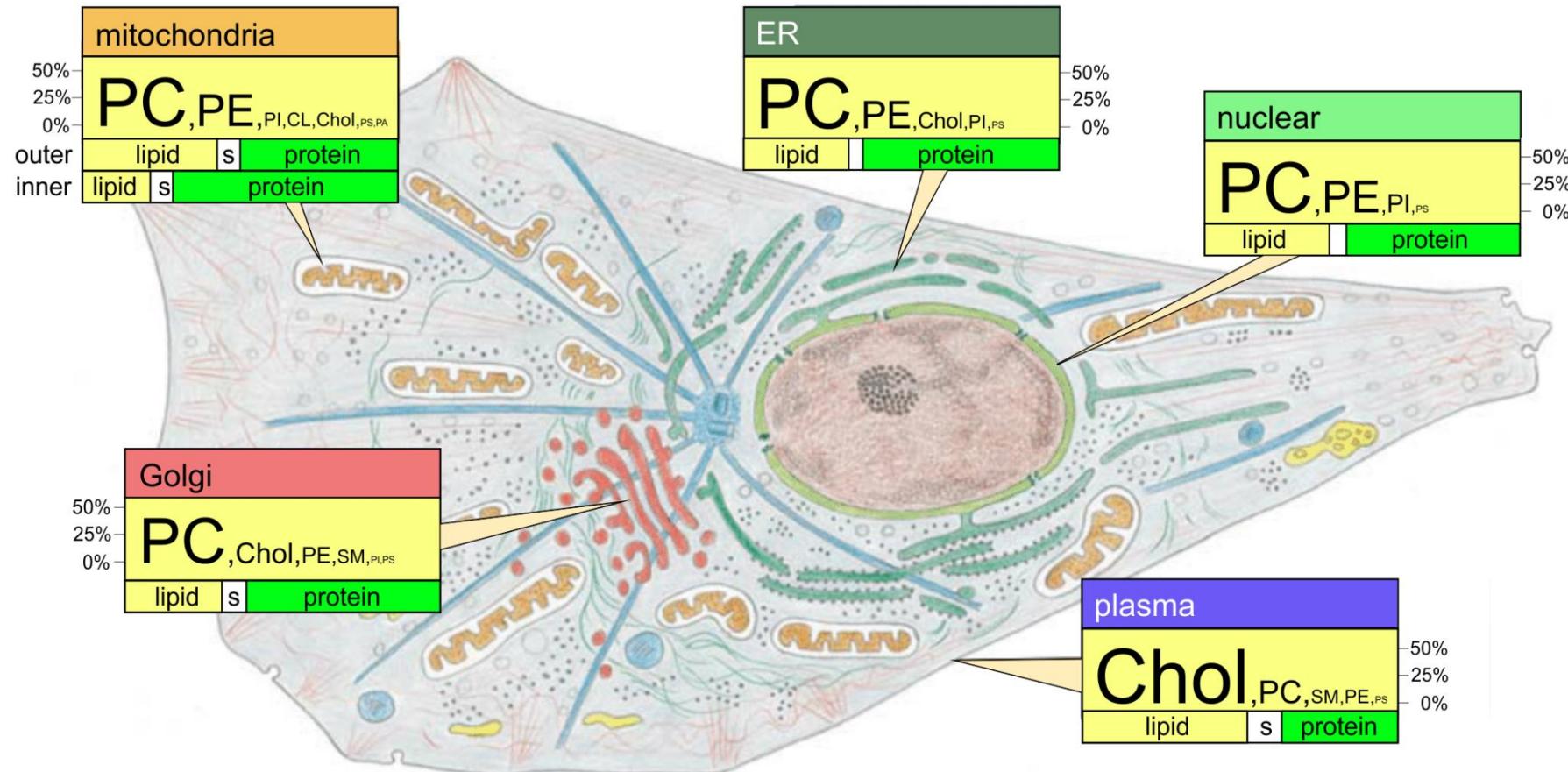
Membrane Proteins



> 60% drug targets are
membrane bound

Alas – not much membrane
protein structures are known

Cell Membranes



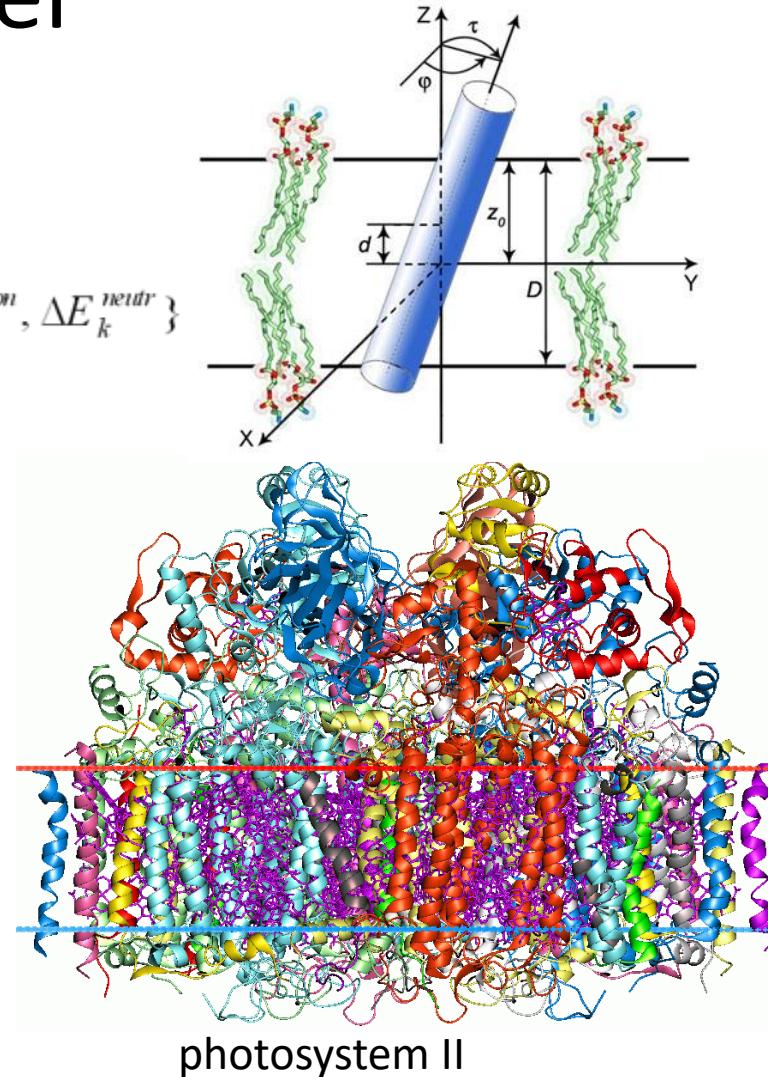
Orientations of Proteins in Membranes

OPM server

- protein = rigid body

$$\Delta G_{transf} = \sum_{i=1}^N \sigma_i^{wat \rightarrow bil}(z_i) ASA_i + \sum_{j=1}^M \eta_j^{wat \rightarrow bil}(z_j) \mu_j + \sum_{k=1}^L \min \{ \Delta E_k^{ion}, \Delta E_k^{neutr} \}$$

- Optimization of protein orientation in anisotropic implicit membrane model
+ membrane composition



Disorder – Intrinsically Disordered Proteins (IDP)

- **(Invisible)** – no structures
- estimates – B-faktor in Xray, NMR ensemble, DISOPRED, AlphaFold
- intristically disordered proteins – **no need of structure to function**
- Regulation functions:
 - molecular recognition (promiscuous)
 - molecular assembly (viral kapsids)
 - protein modification
 - entropic chain activities (entropic hourglasses)

 [Browse](#) [Ontology](#) [Release notes](#) [Download](#) [Help](#) [About](#) [Biocuration](#) 

Welcome to **DisProt**, the database of intrinsically disordered proteins

Disordered regions are manually curated from literature. DisProt annotations cover both structural and functional aspects of disorder detected by specific experimental methods. [Read more about DisProt](#)

Search in DisProt [Search](#)

Examples [p53](#) [CTNNB1](#)
[SARS-CoV-2](#) [Spike glycoprotein](#) [Nucleoprotein](#) [ORF3a protein](#) [ORF7a protein](#)
[Replicase polyprotein 1ab](#) [Envelope small membrane protein](#)

Organisms

 <i>H. sapiens</i> 969	 <i>M. musculus</i> 191	 <i>R. norvegicus</i> 74	 <i>S. cerevisiae</i> 171	 Neglected tropical diseases proteins 82	 Autophagy-related proteins 101	 Cancer-related proteins 145
 <i>E. coli</i> 122	 <i>A. thaliana</i> 77	 <i>D. melanogaster</i> 47	 <i>C. elegans</i> 45	 Viral proteins 191	 Extracellular matrix proteins 55	 Unicellular toxins and antitoxins 47

Datasets

How to cite

DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation Quaglia et al., (2022) *Nucleic Acids Research, Database Issue*.
DOI: [10.1093/nar/gkab1082](#)

 Version: 9.3
Release: 2022_12
Number of entries: 2470

 The DisProt database is part of the [ELIXIR](#) infrastructure and a service of the [IDP](#) Community.

TAKE HOME MESSAGE

Take Home Message

- There is an enormous amount of information available
- Primary databases - **PDB** and **Uniprot**
- Structure defines function
- Active site can have quite often channels
- Membrane proteins are one of the most important groups
- Disorder is also important especially in regulatory proteins (protein-protein interactions)