

LOSCHMIDT  
LABORATORIES



## Proteinové databáze



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



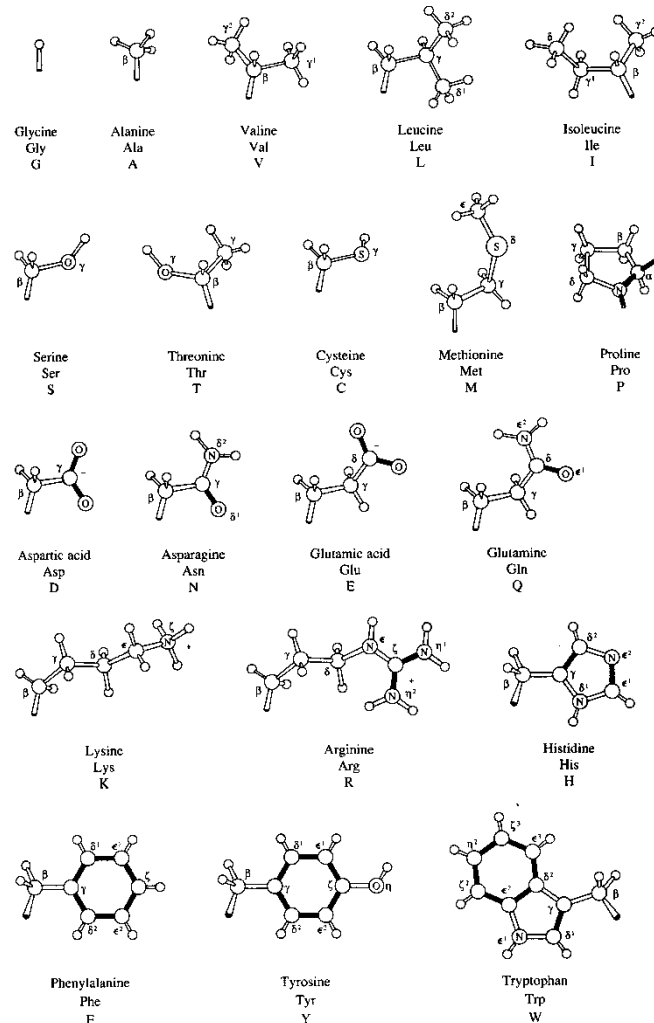
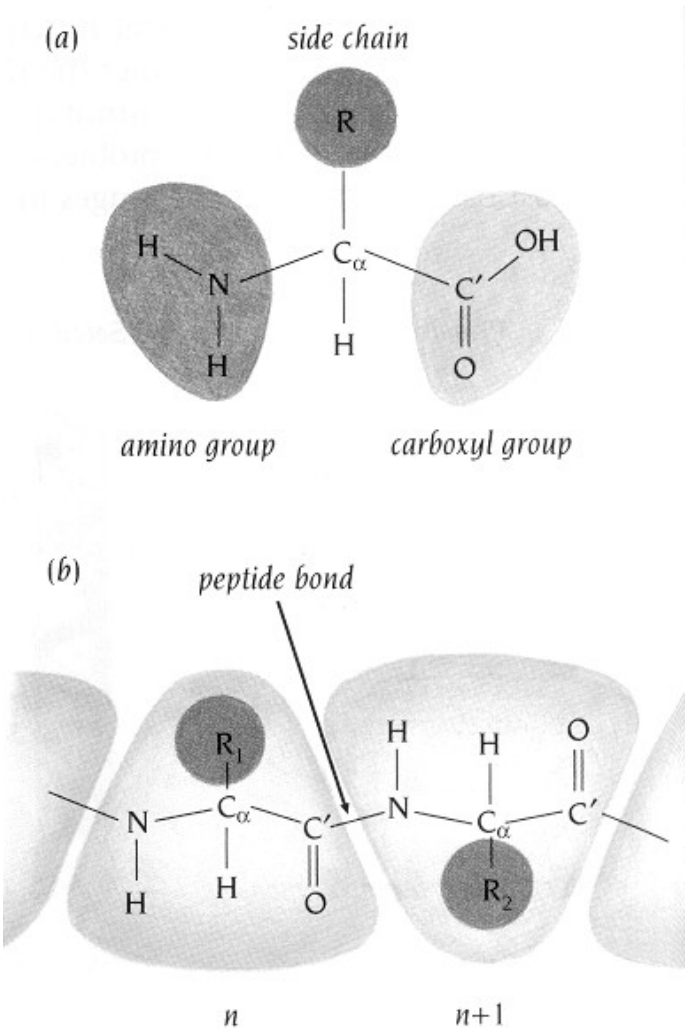
OP Vzdělávání  
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

- ❑ Základní stavební jednotky proteinů
- ❑ Hierarchie proteinové struktury
- ❑ Stanovení proteinové struktury
- ❑ Důležitost proteinové struktury
- ❑ Proteinové strukturní databáze
- ❑ Proteinové klasifikační databáze
- ❑ Strukturní soubory
- ❑ Vizualizace proteinových struktur

# Základní stavební jednotky



# Hierarchie proteinové struktury



- ❑ Primární struktura
- ❑ Sekundární struktura
- ❑ Terciární struktura
- ❑ Kvartérní struktura

# Hierarchie proteinové struktury



## □ Primární struktura

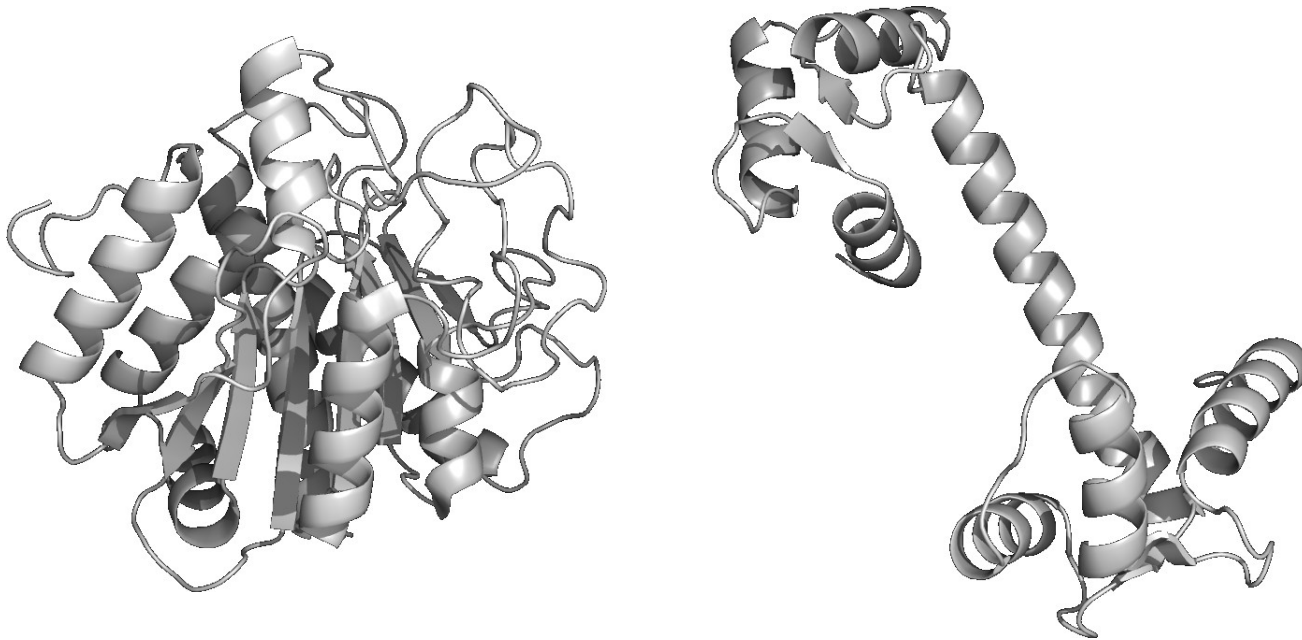
- Lineární sekvence aminokyselinových zbytků

MSLGAKPFGGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCAGLG  
RLIACDLIGMGDSKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVVDWGS  
ALGFDWARRHRERVQGIAYMEAIAMPIEWADFPEQDRDLFQAFRSQAGEELVLQD  
NVFVEQVLPGLILRPLSEAEMAAYREPFLAAGEARRPTLSWPRQIPIAGTPADVVAIA  
RDYAGWLSESPKLFINAEPGALTTGRMRDFCRTWPNQTEITVAGAHFIQEDSPD  
EIGAAIAAFVRRRLRPA

# Hierarchie proteinové struktury

## □ Sekundární struktura

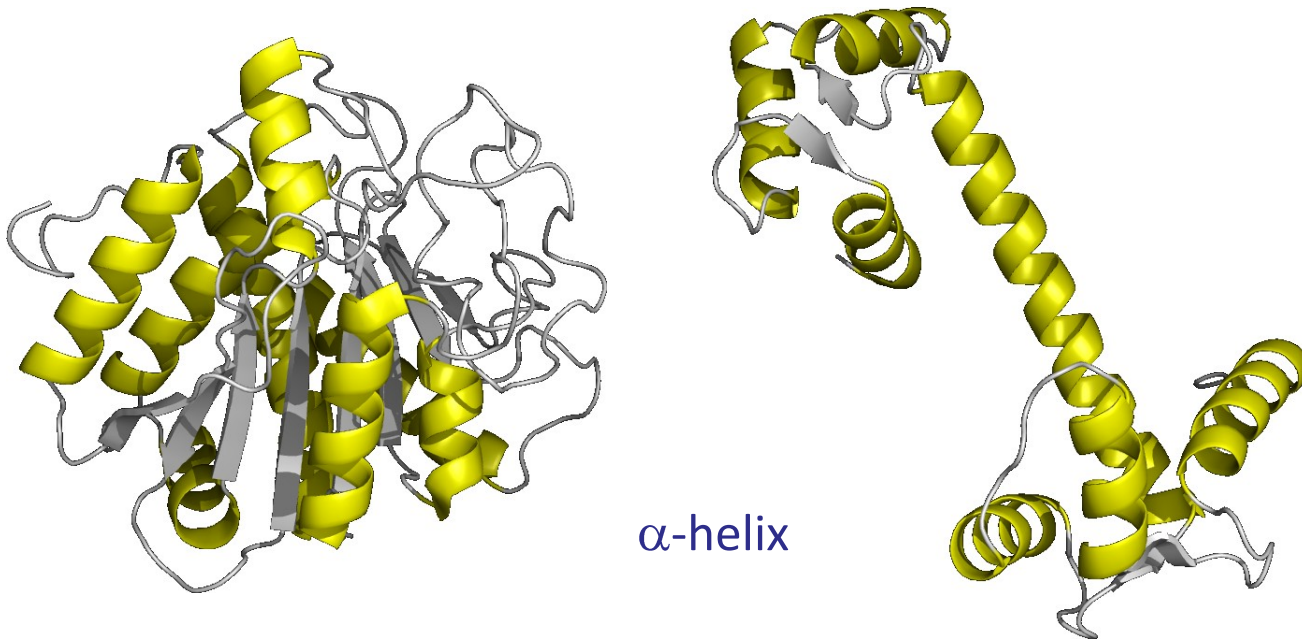
- Lokální prostorové uspořádání peptidového řetězce



# Hierarchie proteinové struktury

## □ Sekundární struktura

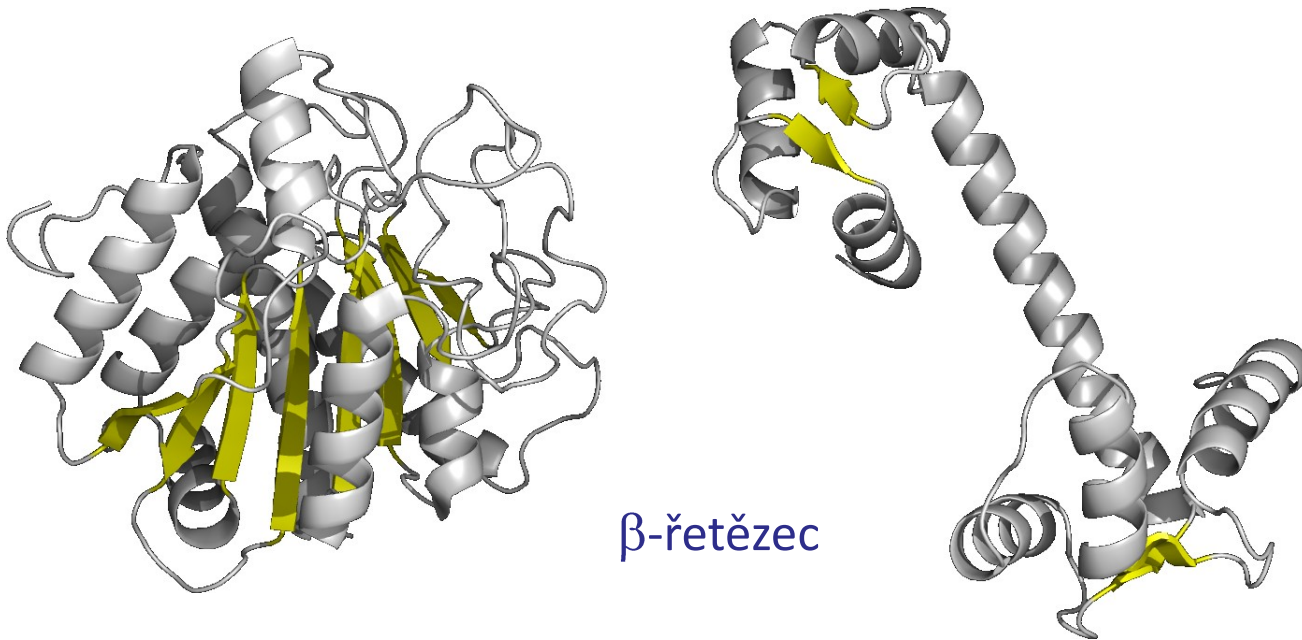
- Lokální prostorové uspořádání peptidového řetězce



# Hierarchie proteinové struktury

## □ Sekundární struktura

- Lokální prostorové uspořádání peptidového řetězce





# Hierarchie proteinové struktury

## □ Terciární struktura

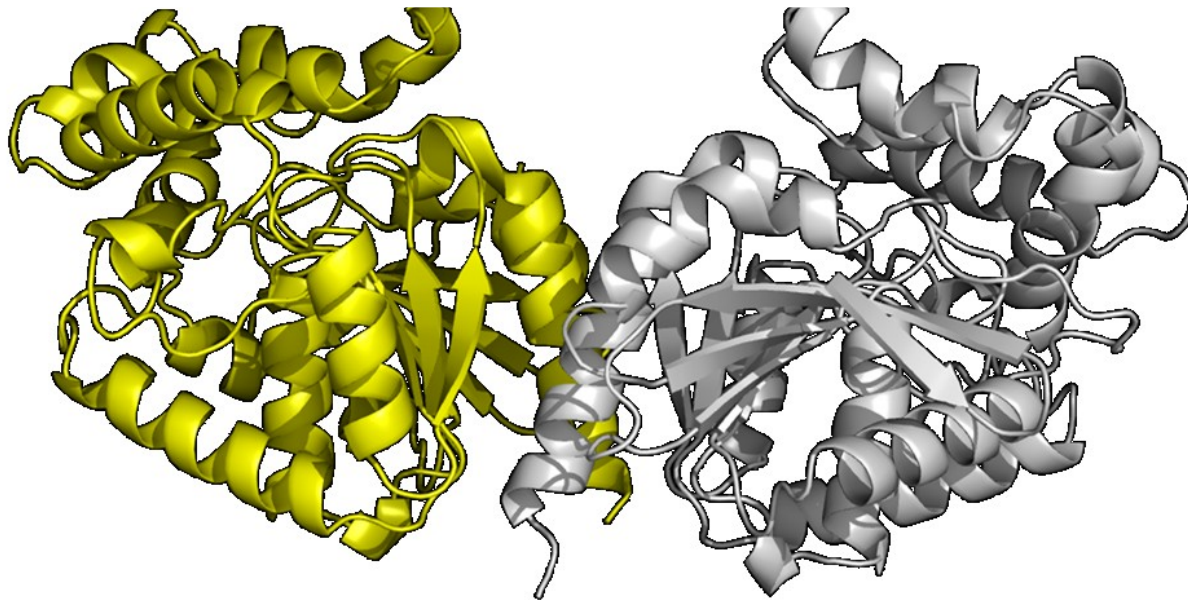
- Globální prostorové uspořádání polypeptidového řetězce



# Hierarchie proteinové struktury

## □ Kvartérní struktura

- Aglomerát několika polypeptidových řetězců



# Určení proteinové struktury



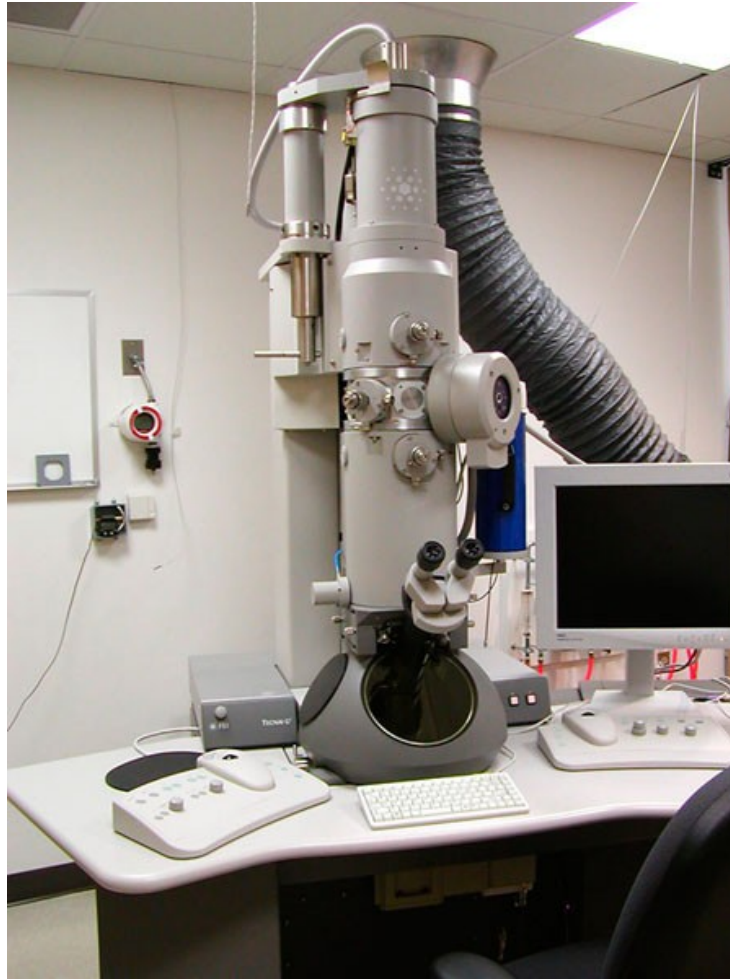
European Synchrotron Radiation Facility, Grenoble, France

# Určení proteinové struktury



600 MHz Ultrashield NMR Spetrometer

# Určení proteinové struktury



FEI Tecnai T12 Cryotransmission Electron Microscope

# Určení proteinové struktury

## □ Proteinová krystalografie

- Atomární rozlišení
- Bez omezení velikosti molekul (virových částic)
- Nutnost přípravy proteinových **krystalů**



# Určení proteinové struktury



## ❑ Proteinová krystalografie

- Atomární rozlišení
- Bez omezení velikosti molekul (virových částic)
- Nutnost přípravy proteinových krystalů

## ❑ Nukleární magnetická rezonance

- Atomární rozlišení
- Omezení velikostí molekul (40 kDa)
- Vzorek proteinu v roztoku = **dynamika**

# Určení proteinové struktury



## ❑ Proteinová krystalografie

- Atomární rozlišení
- Bez omezení velikosti molekul (virových částic)
- Nutnost přípravy proteinových krystalů

## ❑ Nukleární magnetická rezonance

- Atomární rozlišení
- Omezení velikostí molekul (40 kDa)
- Vzorek proteinu v roztoku = dynamika

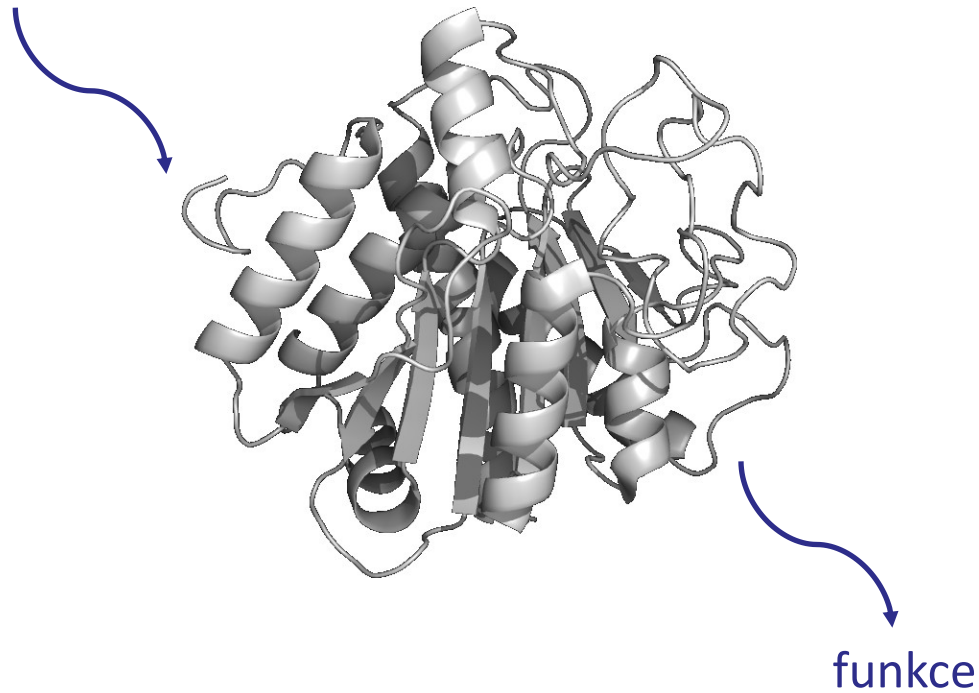
## ❑ Kryoelektronová mikroskopie

- Subatomární rozlišení

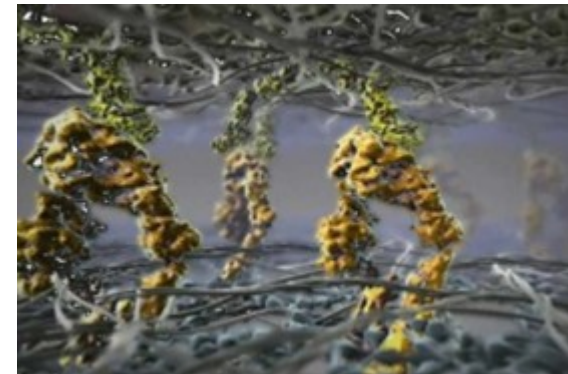
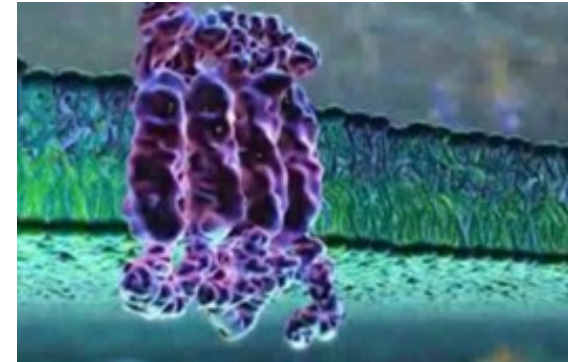
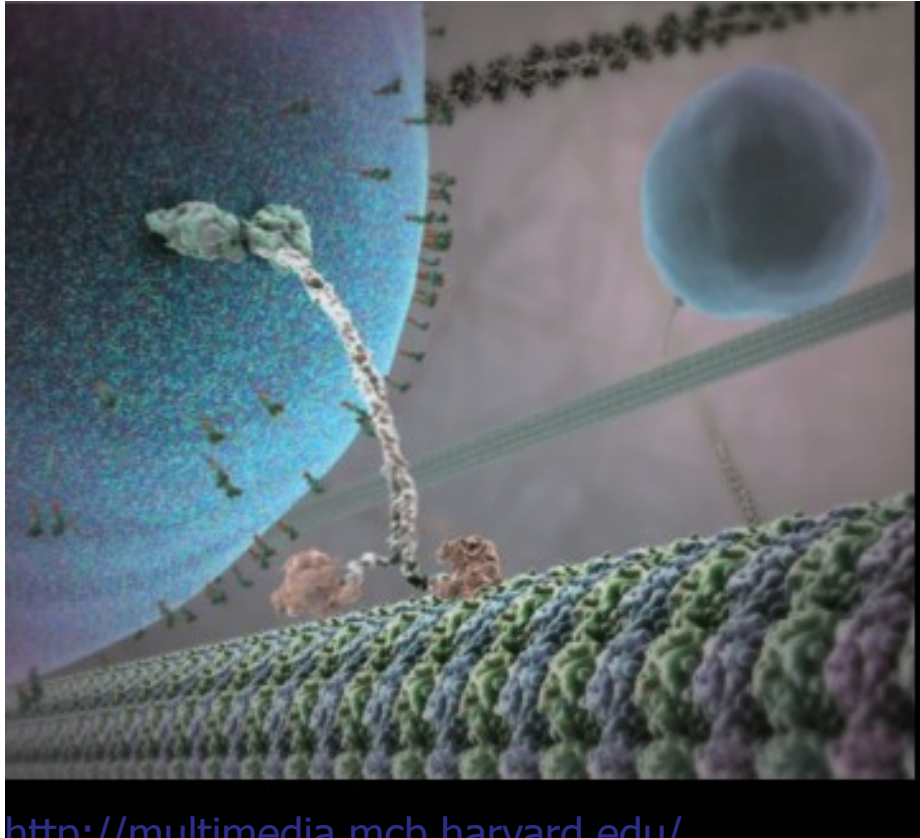


# Důležitost proteinové struktury

MSLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNPTSSYLWRNIMPHCA  
GLGRLIACDLIGMGDSKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVV  
HDWGSALGFDWARRHRERVQGIAYMEAIAMPIEWADFPEQDRDLFQAFRS  
QAGEELVLQD



# Důležitost proteinové struktury



<http://multimedia.mcb.harvard.edu/>

[http://www.wehi.edu.au/education/wehitv/molecular\\_visualisations\\_of\\_dna/](http://www.wehi.edu.au/education/wehitv/molecular_visualisations_of_dna/)

# Proteinové strukturní databáze

- RCSB PDB – Protein Data Bank
  - Světový **depozitář** struktur biomakromolekul



The screenshot shows the RCSB PDB website homepage. At the top left is the RCSB PDB logo. To the right, it says "An Information Portal to Biological Macromolecular Structures" and "As of Tuesday Nov 24, 2009 at 4 PM PST there are 61695 Structures". There are links for "PDB Statistics" and "MyPDB Login". Below the header is a navigation bar with "WHAT'S NEW", "HELP", and "PRINT". A search bar contains "PDB ID or keyword" and has "Search" and "Advanced Search" buttons. The main content area features a large heading "A Resource for Studying Biological Macromolecules" followed by a paragraph: "The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the **wwPDB**, the RCSB PDB curates and annotates PDB data according to agreed upon standards." Below this is another paragraph: "The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists." On the left side, there is a "Search" menu with options like "Advanced Search", "Latest Release", "Latest Publications", "Sequence Search", "Ligand Search", "Unreleased Entries", "Browse Database", and "Histograms". Below that is a "Home" menu with "News & Publications" and "Policies". On the right side, there is a "News" section with a list: "Complete News", "Newsletter", "Discussion Forum", and "Job Listings". Below the list is a date "24-November-2009" and a heading "New Website Features: Improved Tabular Reports" with a subtext "For any search result, users can examine individual PDB".



- RCSB PDB – Protein Data Bank
  - Research Collaboratory for Structural Bioinformatics (RCSB)
  - Služba pro odesílání struktur, prohlížení a vyhledávání dat
  - Sumární informace a struktura
  - Jedinečný identifikační kód **PDB-ID** (1EDE)
  - Interaktivní vizualizace
  - Hledání strukturních sousedů
  - Reference a **hyperlinky** do dalších databází

# Proteinové strukturní databáze

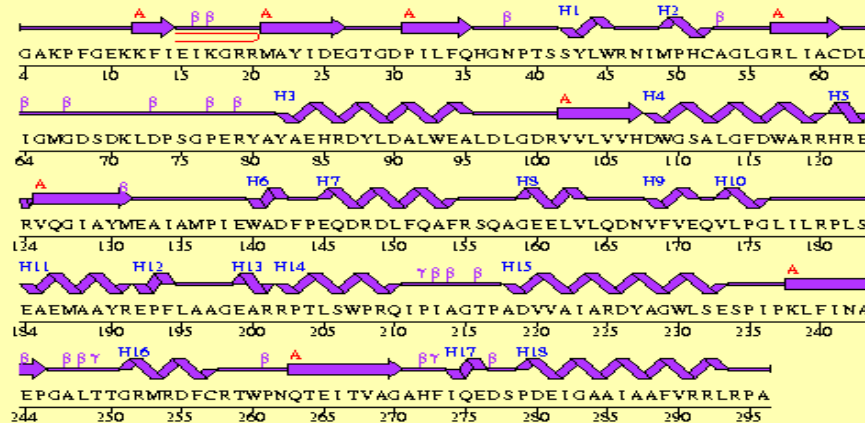
## □ PDBsum

- Přehled a **analýza** proteinových struktur

Chain A ( 293 residues )

- **CATH** structural classification ( 1 domain ) :

Links	CATH no.	Class	Architecture
<a href="#">CATH</a> <a href="#">DHS</a>	<b>3.40.50.950</b>	→ <i>AlphaBeta</i>	<i>3-Layer(aba) Sandwich</i>



Postscript version



View chain **A** alone.

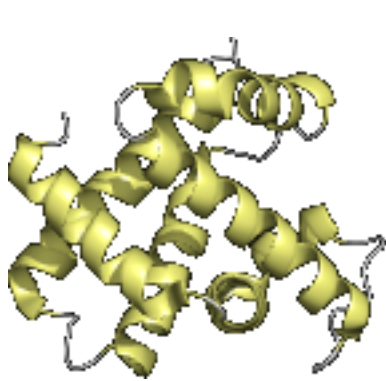
- **PROMOTIF** summary :  
1 [sheet](#), 8 [strands](#), 18 [helices](#), 19 [beta turns](#), 3 [gamma turns](#), 1 [beta bulge](#), 1 [beta hairpin](#), 4 [beta alpha beta units](#), 1 [psi-loop](#).

## □ PDBsum

- Souhrnné informace
- Sekundární struktury
- Katalytické a vazující aminokyselinové zbytky
- Ligandy
- Dutiny
- Protein-proteinové interakce
- Konzervovanost aminokyselinových zbytků
- Reference a hyperlinky do dalších databází

# Proteinové klasifikační databáze

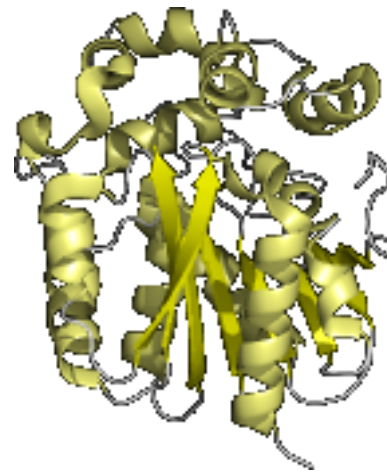
- Identifikace a kvantifikace strukturních podobností
  - Strukturní podobnost souvisí s **evolucí** a implikuje **funkci**



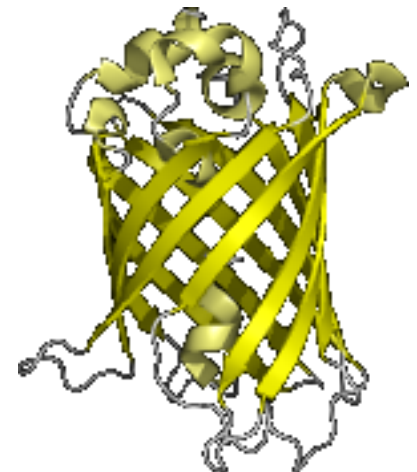
všechny alfa



všechny beta




alfa/beta



alfa+beta

- SCOP – Structural Classification of Proteins
  - Založena na **manuálním** srovnání struktur expertem (A. Murzin)

*Structural Classification of Proteins*



**Protein: Haloalkane dehalogenase from *Sphingomonas paucimobilis*, UT26, LinB [[TaxId: 13689](#)]**

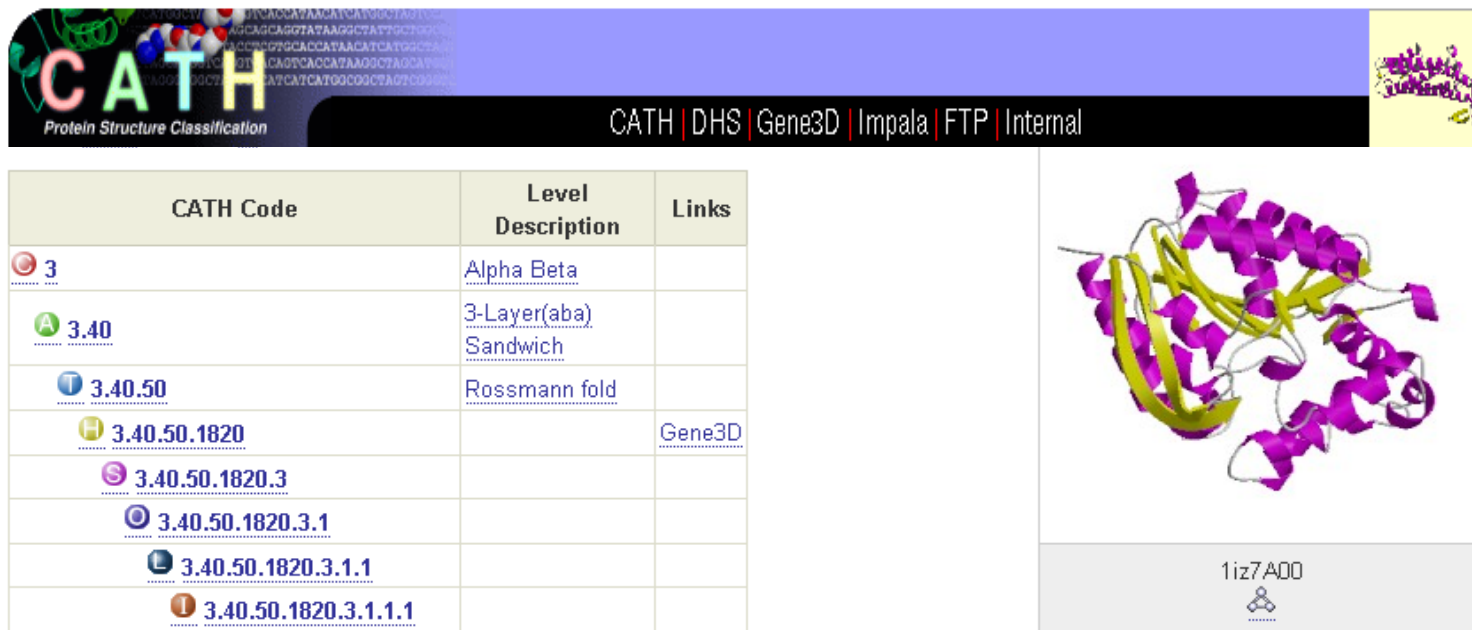
**Lineage:**

1. Root: [scop](#)
2. Class: [Alpha and beta proteins \(a/b\)](#) [51349]  
*Mainly parallel beta sheets (beta-alpha-beta units)*
3. Fold: [alpha/beta-Hydrolases](#) [53473]  
*core: 3 layers, a/b/a; mixed beta-sheet of 8 strands, order 12435678, strand 2 is antiparallel to the rest*
4. Superfamily: [alpha/beta-Hydrolases](#) [53474]  
*many members have left-handed crossover connection between strand 8 and additional strand 9*
5. Family: [Haloalkane dehalogenase](#) [53513]
6. Protein: Haloalkane dehalogenase [53514]
7. Species: [Sphingomonas paucimobilis, UT26, LinB \[\[TaxId: 13689\]\(#\)\]](#) [53517]



# Proteinové klasifikační databáze

- CATH – Class, Architecture, Topology, Homology
  - Založena na **kombinaci** automatického strukturního přiložení programem SSAP a manuálního srovnání, **E.C. kódování**



CATH Code	Level Description	Links
<a href="#">3</a>	<a href="#">Alpha Beta</a>	
<a href="#">A 3.40</a>	<a href="#">3-Layer(aba) Sandwich</a>	
<a href="#">T 3.40.50</a>	<a href="#">Rossmann fold</a>	
<a href="#">H 3.40.50.1820</a>		<a href="#">Gene3D</a>
<a href="#">S 3.40.50.1820.3</a>		
<a href="#">O 3.40.50.1820.3.1</a>		
<a href="#">L 3.40.50.1820.3.1.1</a>		
<a href="#">I 3.40.50.1820.3.1.1.1</a>		

1iz7A00

# Proteinové klasifikační databáze

PDB code: 1lys

<b>SCOP</b>		<b>CATH</b>	
<i>Class</i>	Alpha and Beta ( $\alpha+\beta$ )	<i>Class</i>	Mainly Alpha
		<i>Architecture</i>	Orthogonal Bundle
<i>Fold</i>	Lysozyme-like	<i>Topology</i>	Lysozyme
<i>Superfamily</i>	Lysozyme-like	<i>Homologous Superfamily</i>	Hydrolase (O-glycosyl)
<i>Family</i>	C-type lysozyme	<i>Homologous Family</i>	Hydrolase

## □ PDB formát

- Široce používaný
- Snadno čitelný
  
- Identifikátory struktury
- Literární reference
- Experimentální podmínky
- XYZ koordináty struktury

# Strukturní soubory

## □ PDB formát

	HEADER	LYASE (CARBON-CARBON)				03-JUL-95		1DNP				
structure annotation	TITLE	STRUCTURE OF DEOXYRIBODIPYRIMIDINE PHOTOLYASE										
	...	...										
	SOURCE	2 ORGANISM_SCIENTIFIC: ESCHERICHIA COLI										
	KEYWDS	DNA REPAIR, ELECTRON TRANSFER, EXCITATION ENERGY TRANSFER,										
	KEYWDS	2 LYASE, CARBON-CARBON										
	...	...										
	ATOM	21	ND1	HIS	A	3	55.365	27.866	62.971	1.00	11.07	N
	ATOM	22	CD2	HIS	A	3	57.200	28.354	61.894	1.00	13.12	C
	ATOM	23	CE1	HIS	A	3	56.124	26.783	62.981	1.00	13.03	C
	ATOM	24	NE2	HIS	A	3	57.243	27.052	62.334	1.00	8.19	N
	ATOM	25	N	LEU	A	4	55.580	32.694	59.656	1.00	12.61	N
	ATOM	26	CA	LEU	A	4	54.799	33.803	59.113	1.00	11.56	C
amino acid field	ATOM	27	C	LEU	A	4	53.552	33.269	58.374	1.00	7.76	C
	ATOM	28	O	LEU	A	4	53.650	32.363	57.532	1.00	6.99	O
	ATOM	29	CB	LEU	A	4	55.656	34.683	58.174	1.00	9.03	C
	ATOM	30	CG	LEU	A	4	54.946	35.887	57.518	1.00	2.00	C
	ATOM	31	CD1	LEU	A	4	54.623	36.920	58.550	1.00	6.21	C
	...	...										
cofactor filed	HETATM	7641	AN7	FAD	B	472	27.855	78.556	29.073	1.00	4.55	N
	HETATM	7642	AC5	FAD	B	472	28.524	78.026	27.955	1.00	2.00	C
	HETATM	7643	AC6	FAD	B	472	29.848	77.609	27.724	1.00	3.40	C
	HETATM	7644	AN6	FAD	B	472	30.787	77.757	28.664	1.00	6.22	N

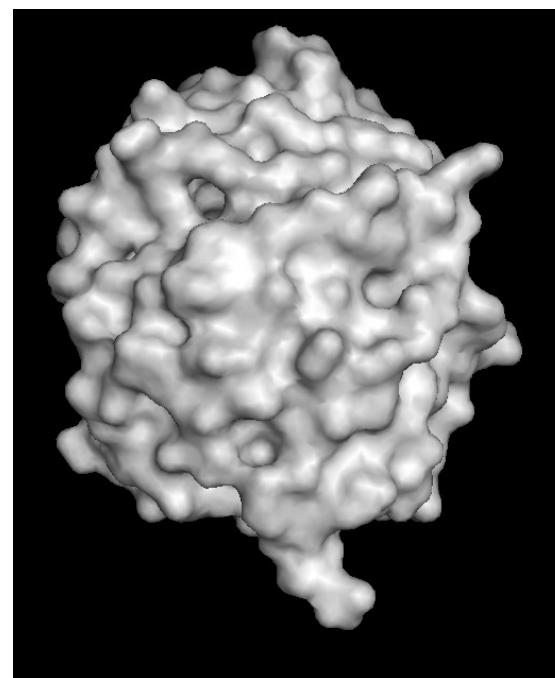
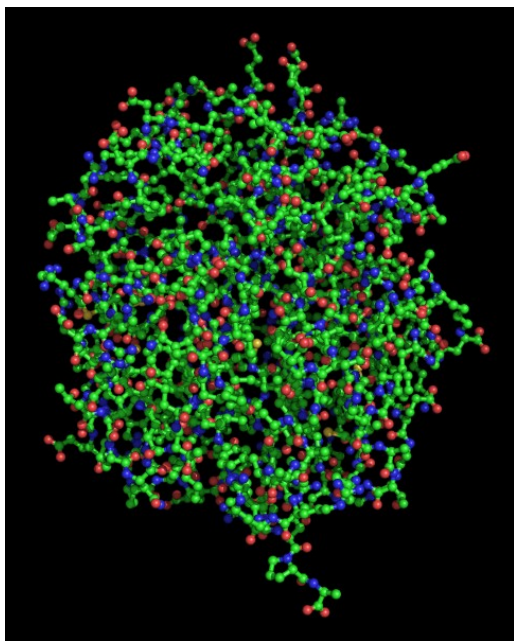
atom number / residue name | residue number | x, y, z coordinates | occupancy | temperature factor | atom type  
 atom name | polypeptide chain identifier

# Strukturní soubory

- mmCIF formát
  - Snadno analyzovatelný počítačovými programy

```
mmCIF  _struct.entry_id '1CBN'  
        _struct.title 'PLANT SEED PROTEIN'  
        _struct_keywords.entry_id '1CBN'  
        _struct_keywords.text 'plant seed protein'  
        _database_2.database_id 'PDB'  
        _database_2.database_code '1CBN'  
        _database_PDB_rev.rev_num 1  
        _database_PDB_rev.date_original '1991-10-11'
```

# Vizualizace proteinových struktur



# Vizualizace proteinových struktur

- ❑ PyMOL
- ❑ RasMol
- ❑ Swiss-PdbViewer
- ❑ Chimera
- ❑ Jmol
- ❑ Cn3D
- ❑ WebMol
- ❑ VMD
- ❑ Proteopedia

# Vizualizace proteinových struktur

Main Page - Proteopedia, life in 3D - Mozilla Firefox

Soubor Úpravy Zobrazení Historie Záložky Nástroje Nápověda

http://www.proteopedia.org/

Main Page - Proteopedia, life in ...

article discussion edit this page history

First time at Proteopedia? Click on the **green links**, they change the 3D image. Click and drag the molecules. Proteopedia is a 3D, interactive encyclopedia of proteins, RNA, DNA and other molecules. With a free user account, you can edit pages in Proteopedia. Visit the [Main Page](#) to learn more.

**Welcome to Proteopedia,**  
The free, collaborative 3D encyclopedia of proteins & other molecules

About • Editing • Help

Video Guide • Table of Contents • Content (Topic Pages) • What's New

**Exciting news!**

TheScientist "Labby Award" Winner

Publish your Proteopedia Page in *BAMBED*

*Journal of Biological Inorganic Chemistry*

Proteopedia is the winner of the 2010 TheScientist "Best Science Website" Labby Awards. Congratulations to all Proteopedia Users!

Proteopedia pages can now be submitted to the *Journal of Biochemistry and Molecular Biology Education (BAMBED)* for peer-review and publication. [Submit your page](#) and see the *Sept/Oct edition of BAMBED*.

The *Journal of Biological Inorganic Chemistry (JBIC)* now publishes Interactive 3D Complements in Proteopedia for each of their macromolecular structure papers. See the [first](#) and the [second](#) such Interactive 3D Complements.

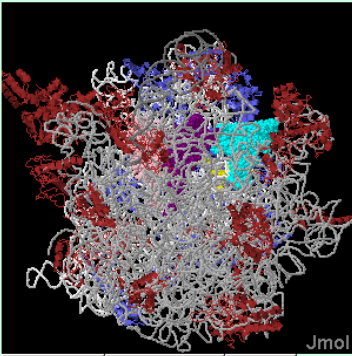
**Currently featured article**

Green links change the 3D image! Click and drag on the molecule!

**The Ribosome**

by Wayne Decatur

On October 7th, 2009 the Nobel Committee announced three structural biologists would share the [2009 Nobel Prize in Chemistry](#) for studies of the [The Ribosome](#). The ribosome is the machine in your cells that accurately and efficiently decodes the genetic information stored in your genome and synthesizes the corresponding polypeptide chain one amino acid at a time in the process of translation. Venkatraman Ramakrishnan of the M.R.C. Laboratory of Molecular Biology in Cambridge, England; Thomas A. Steitz of Yale University; and Ada E. Yonath of the Weizmann Institute of Science in Rehovot, Israel share the prize for the first atomic-resolution structures of the two subunits that come together to form an active ribosome. These structures are considered landmarks for the fact they showed clearly the major contributions to decoding and peptide bond synthesis come from RNA and not protein, as well as for the sheer size of the structures determined. These structures represent tour-de-force efforts in understanding fundamental processes in every organism on earth and will have direct impacts on how we fight pathogenic bacteria in the immediate future. Shown [here \(restore initial scene\)](#) are both subunits of the ribosome, as well as [mRNA](#) and [tRNA](#) that bind in the complex during the process of translation. [Read more...](#)



Jmol

toggle spin toggle quality popup

--large subunit on/off --small subunit on/off--

Jmol script terminated

navigation

- Main Page
- Table of Contents
- Structure Index
- Help

random

- Random article
- Random PDB entry

search

Go Search

Google Custom Search

Google Search

toolbox

- Export this page
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link



# Reference

- ❑ Claverie, J-M., & Notredame, C. (2006). **Bioinformatics For Dummies** (2nd ed.). Wiley Publishing, Hoboken, p. 436.
- ❑ Xiong, J. (2006). **Essential Bioinformatics**. Cambridge University Press, New York, p. 352.
- ❑ BioVisions na Harvardově univerzitě: <http://multimedia.mcb.harvard.edu/>
- ❑ **RCSB PDB**: <http://pdb.rcsb.org/pdb/home/home.do>
- ❑ **PDBsum**: <http://www.ebi.ac.uk/pdbsum/>
- ❑ **SCOP**: <http://scop.mrc-lmb.cam.ac.uk/scop/>
- ❑ **CATH**: <http://www.cathdb.info/>
- ❑ **PyMOL**: <http://pymol.sourceforge.net/>
- ❑ **ProteoPedia**: <http://www.proteopedia.org/>