

LOSCHMIDT  
LABORATORIES



## Predikce proteinové struktury



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

- ❑ Predikce sekundární struktury
- ❑ Predikce proteinového foldu
- ❑ Predikce terciární struktury
- ❑ Predikce molekulárních komplexů
- ❑ Hodnocení predikčních metod

# Predikce proteinové struktury



- ❑ Predikce sekundární struktury
- ❑ Predikce proteinového foldu
  - Navlékání - angl. **Threading**
- ❑ Predikce terciární struktury
  - Homologní modelování - angl. **Homology modelling**
  - *Ab initio* predikce - angl. **Ab initio prediction**
- ❑ Predikce molekulárních komplexů
  - Molekulární dokování - angl. **Molecular docking**

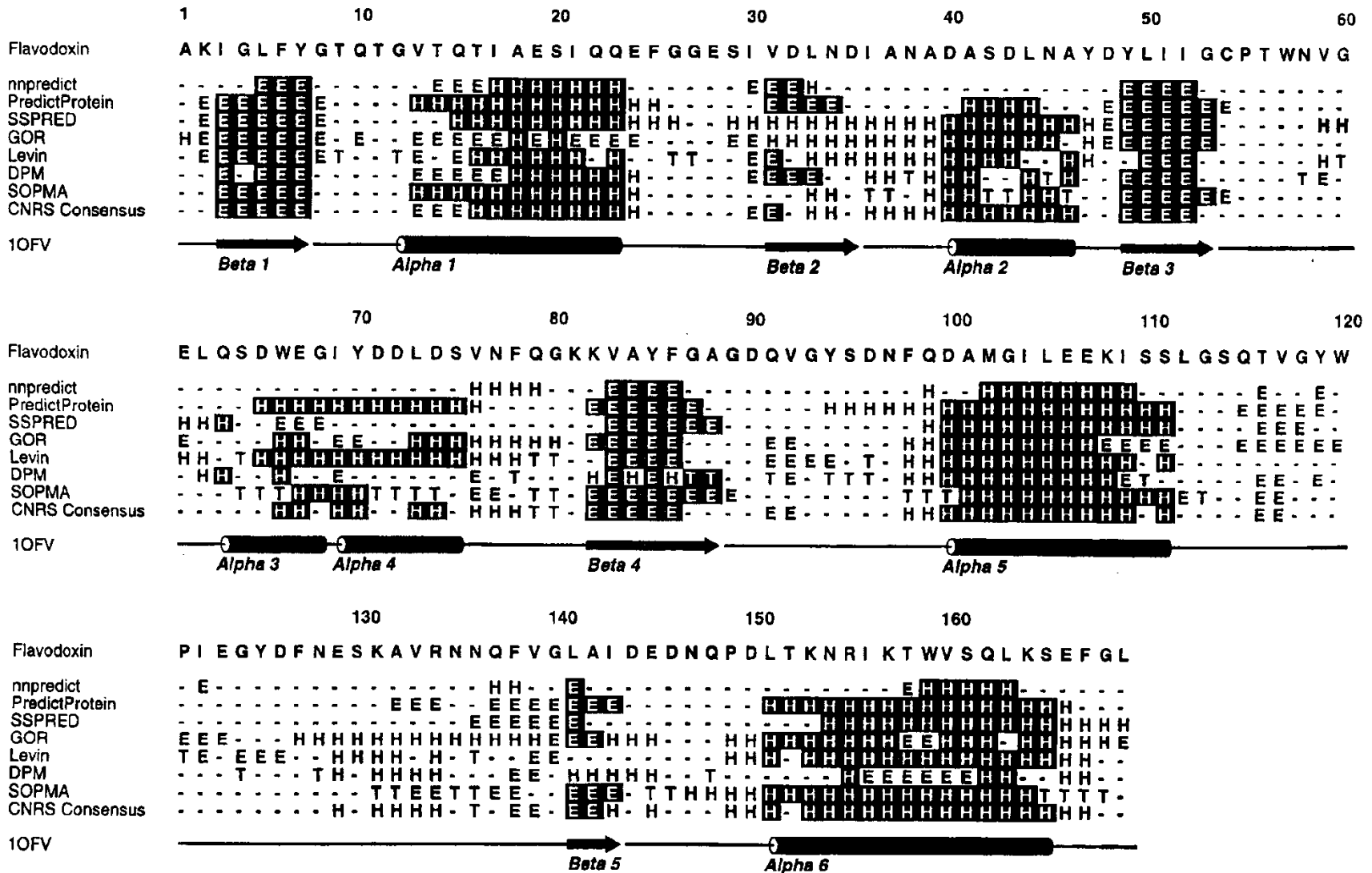


- Přiřazení jednoho **konformačního stavu** každému aminokyselinovému zbytku v proteinové sekvenci:
  - $\alpha$ -šroubovice (H, angl. helix)
  - $\beta$ -řetězec (E, angl. strand)
  - otočka (C, angl. coil)



- Přiřazení jednoho **konformačního stavu** každému aminokyselinovému zbytku v proteinové sekvenci:
  - **Přesnost >80%**
  - Klasifikace proteinů
  - Identifikace proteinových domén a funkčních motivů
  - Zlepšení spolehlivosti sekvenčních příložen
  - Příprava na predikci terciární struktury

# Predikce sekundární struktury

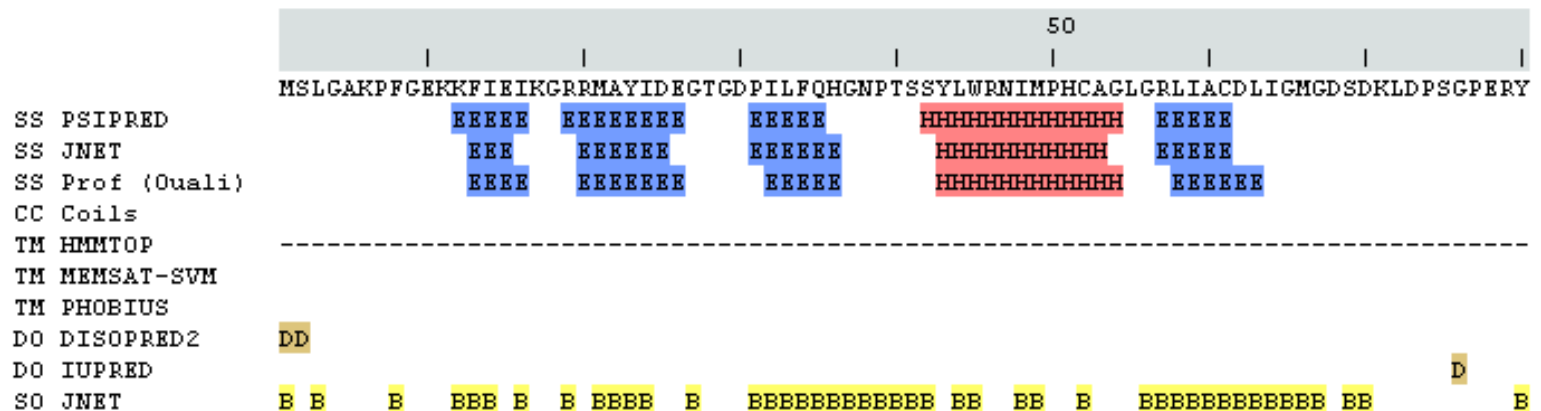




# Predikce sekundární struktury

## □ Quick2D

- Přiřazení sekundárních elementů:  $\alpha$ -šroubovic,  $\beta$ -řetězců, otcěk, transmembránových šroubovic a neuspořádaných regionů
- Metody PSI-PRED, JNET, Prof, Coils, MEMSAT2, HMMTOP, ...

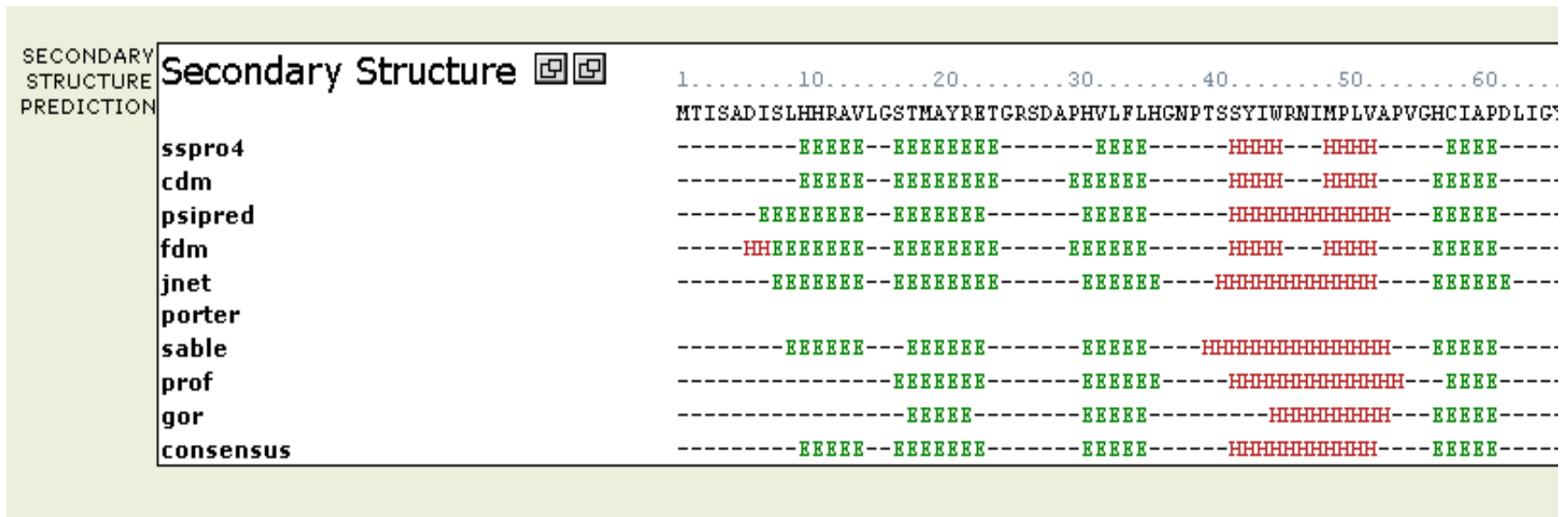




# Predikce sekundární struktury

## □ GeneSilico MetaServer

- **Meta-server** pro predikci struktury proteinů, včetně predikce sekundární elementů = **konsensus**



## □ Navlékání

- Rozpoznávání proteinového foldu
- Hledá strukturu, která nejlépe odpovídá proteinové sekvenci prohledáváním **knihovny známých foldů** a hodnocením **skóre**
- Používá se pro struktury, pro které **není** k dispozici vhodný templát pro homologní modelování
- Neposkytne výsledek, pokud správný fold není v knihovně





# Predikce proteinového foldu



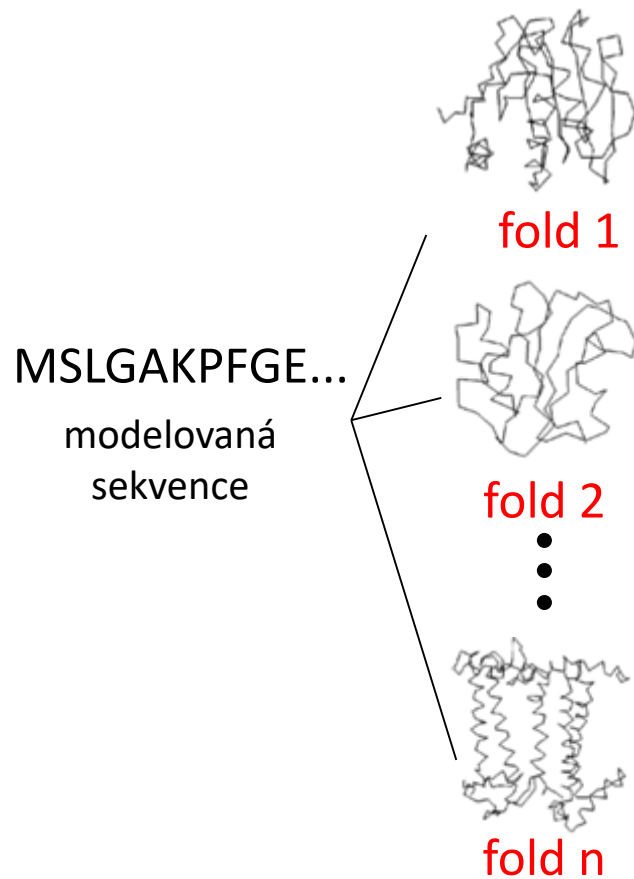
- Navlékání

MSLGAKPFGE...

modelovaná  
sekvence

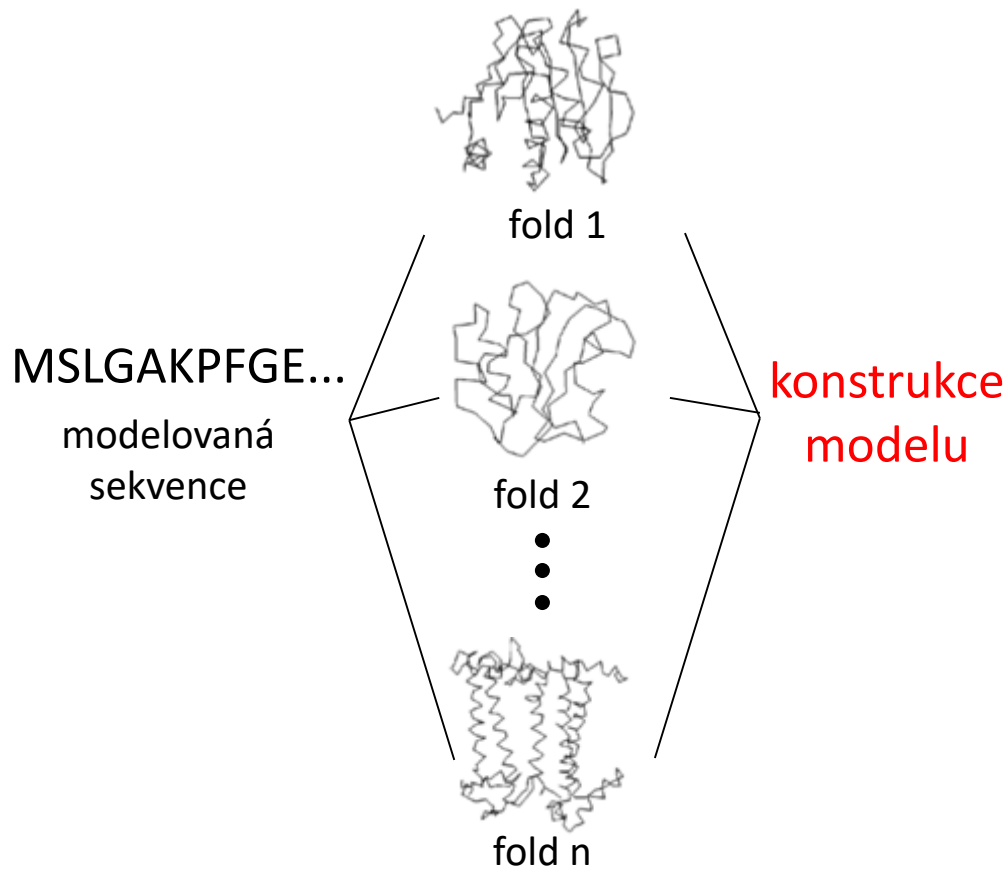
# Predikce proteinového foldu

## □ Navlékání



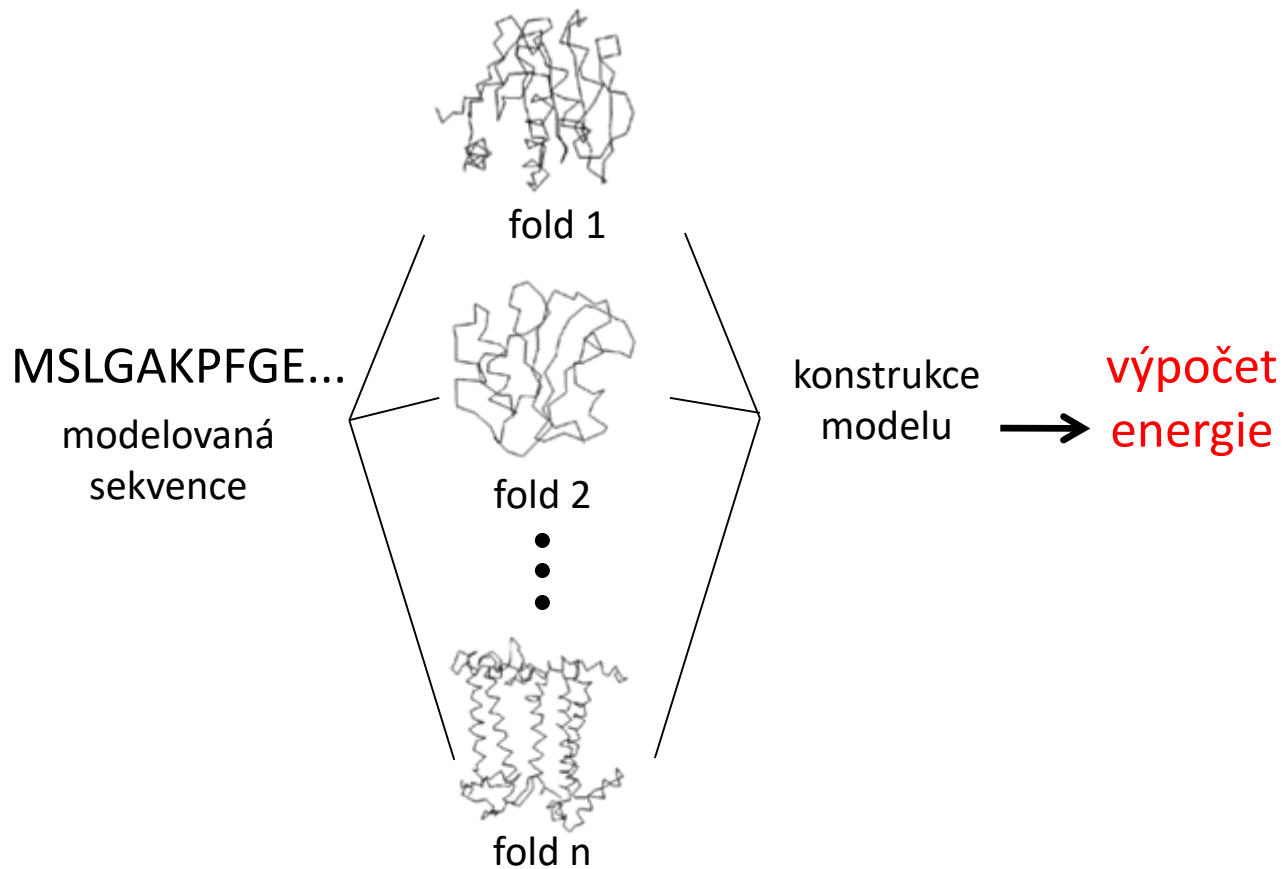
# Predikce proteinového foldu

## □ Navlékání



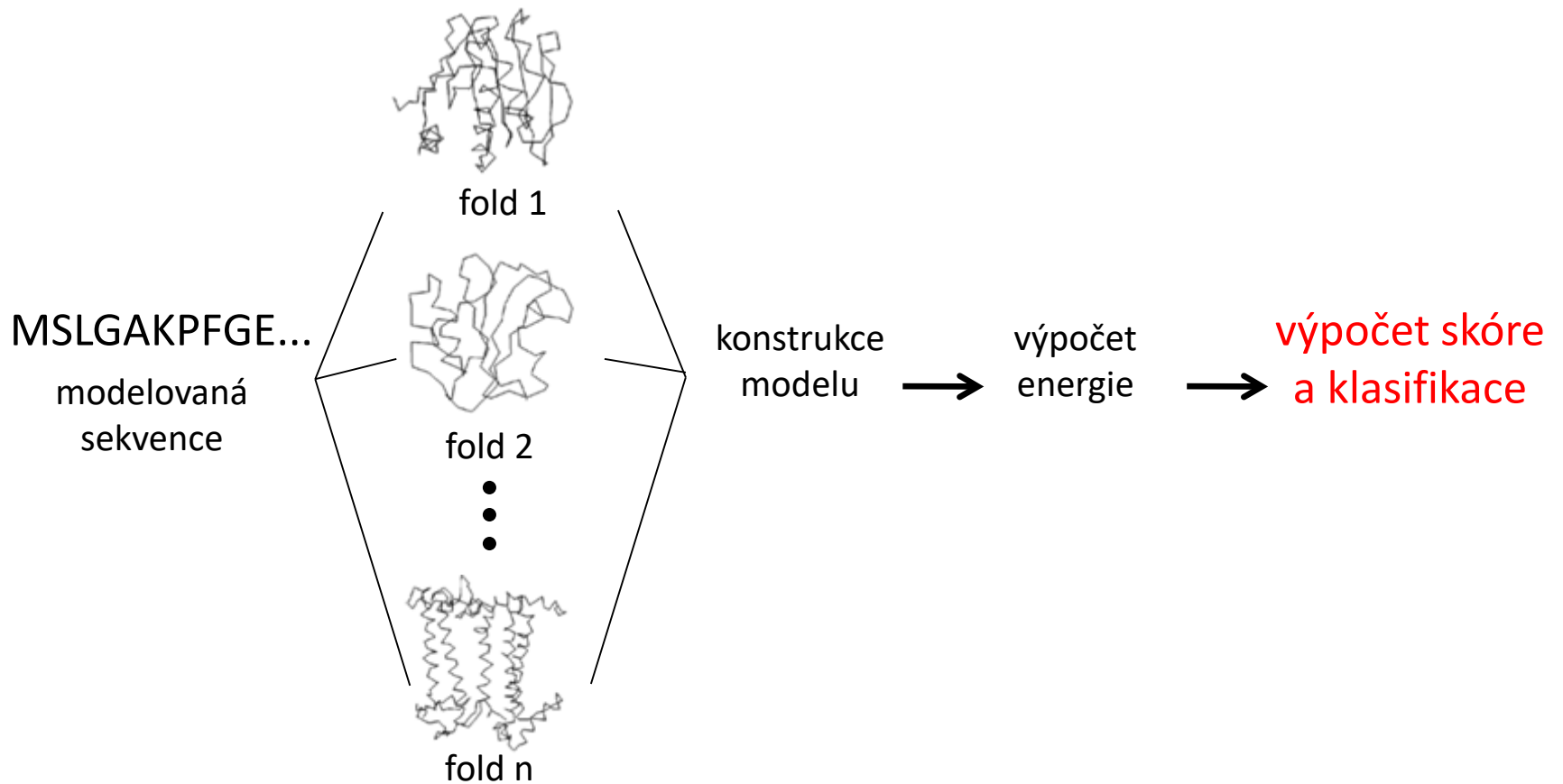
# Predikce proteinového foldu

## □ Navlékání



# Predikce proteinového foldu

## □ Navlékání



# Predikce proteinového foldu

## □ Navlékání

- PHYRE
- GenTHREADER

phyre

Protein Homology/analogY Recognition Engine

Version 0.2

Google groups

Subscribe to Phyre

Email:

Subscribe

[Visit this group](#)

The Phyre webserver is for **Academic use only**

E-mail Address

Optional Job description

Amino Acid Sequence

Quick Phyre Search





# Predikce terciární struktury



- Homologní modelování
- *Ab initio* predikce

## □ Homologní modelování

- Vytváří atomistický model založený na **experimentálně určené struktuře**, která je sekvenčně blízce příbuzná
- Vyžadovaná sekvenční identita **>25%**
- Základní princip = struktura je konzervována déle než sekvence





# Predikce terciární struktury



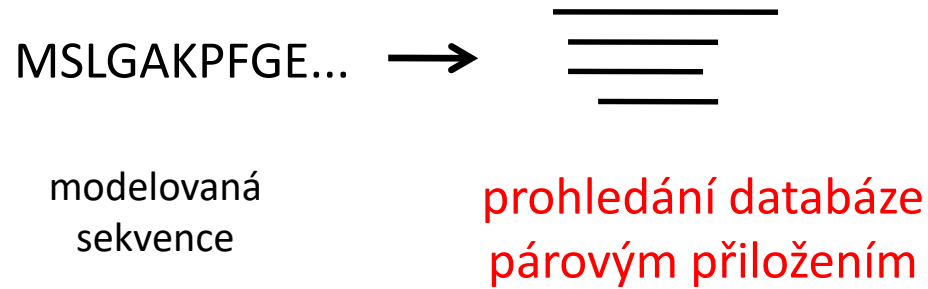
- Homologní modelování

MSLGAKPFGE...

modelovaná  
sekvence

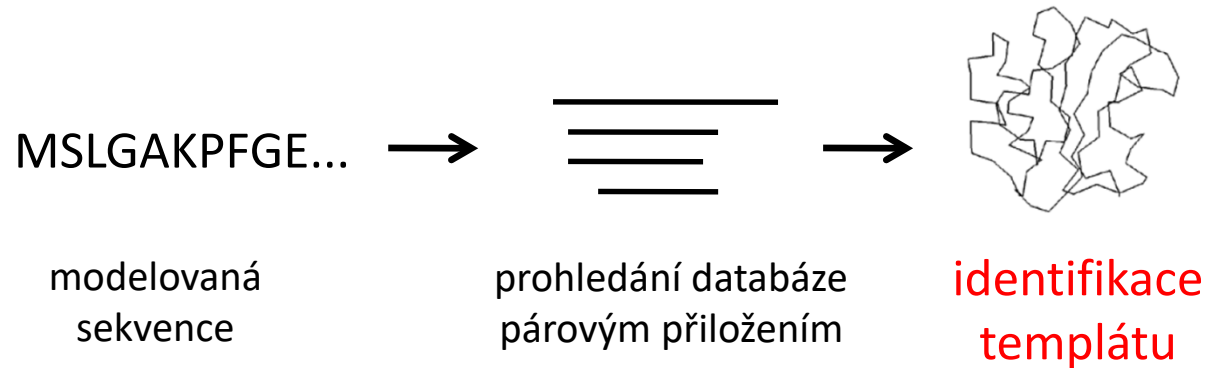
# Predikce terciární struktury

## □ Homologní modelování



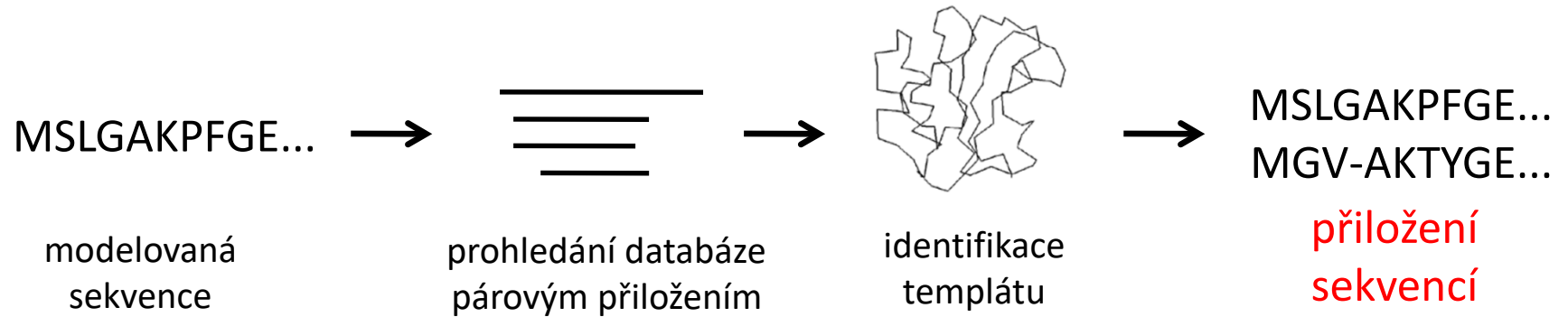
# Predikce terciární struktury

## □ Homologní modelování



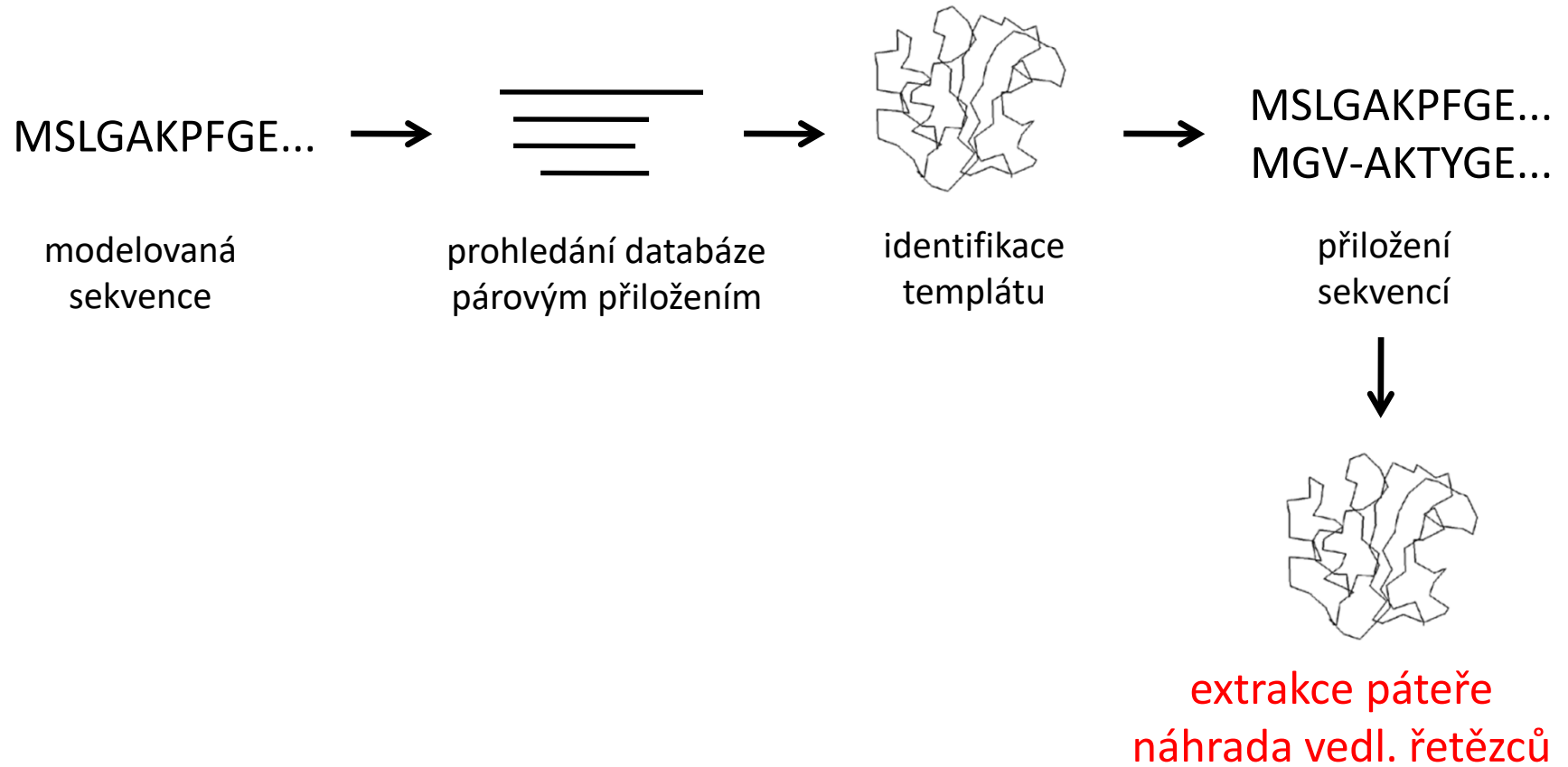
# Predikce terciární struktury

## □ Homologní modelování



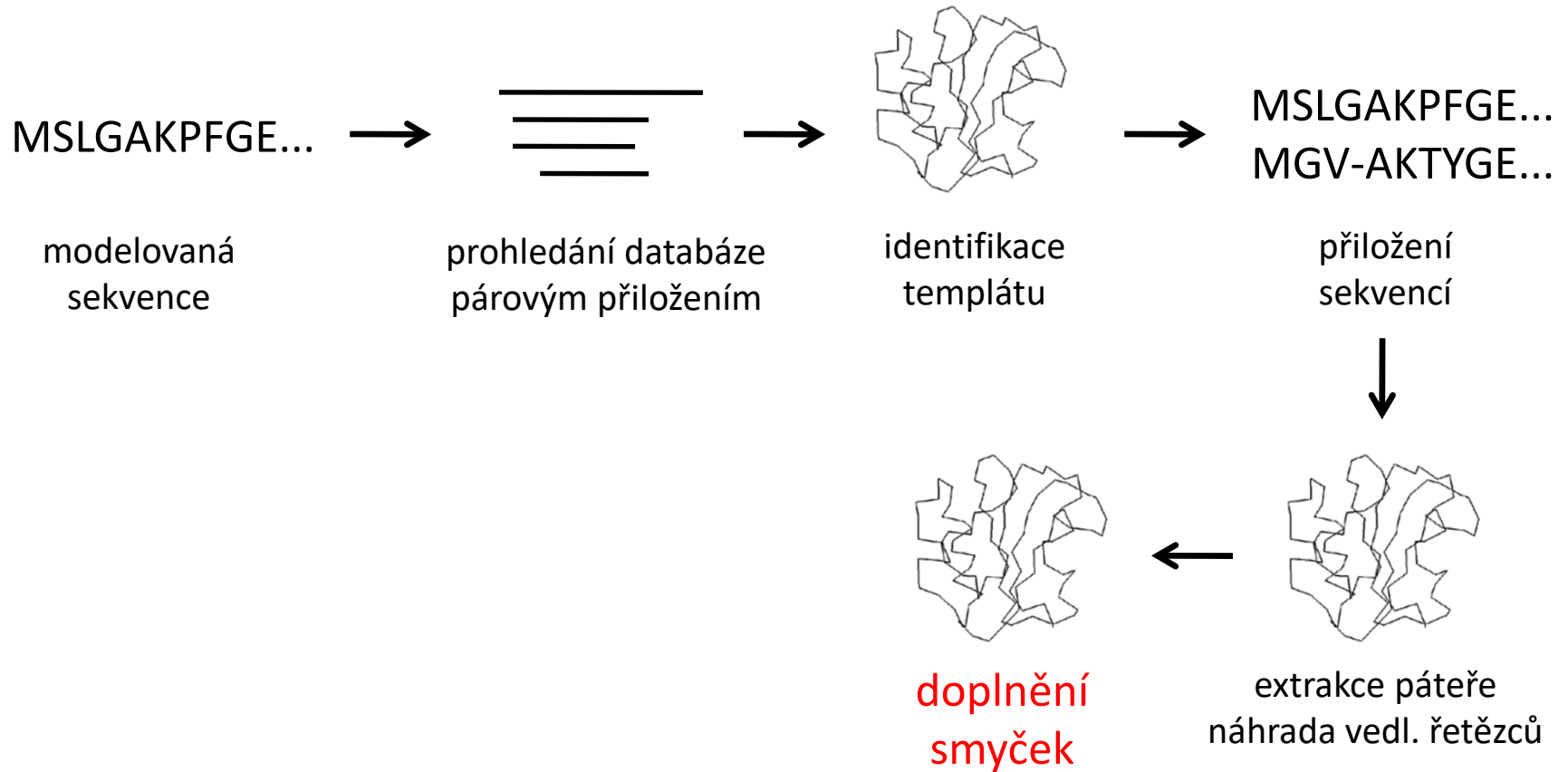
# Predikce terciární struktury

## □ Homologní modelování



# Predikce terciární struktury

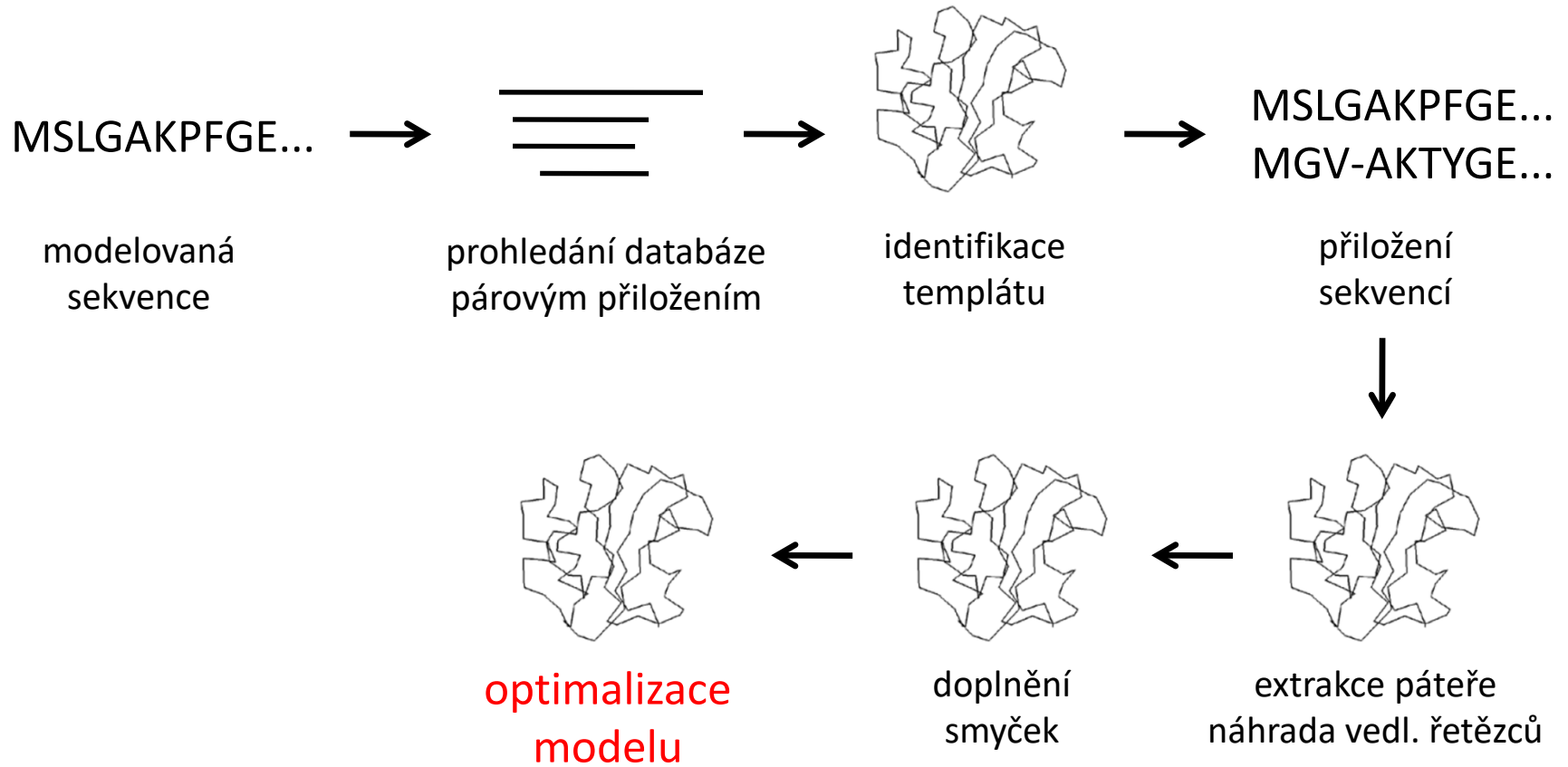
## □ Homologní modelování





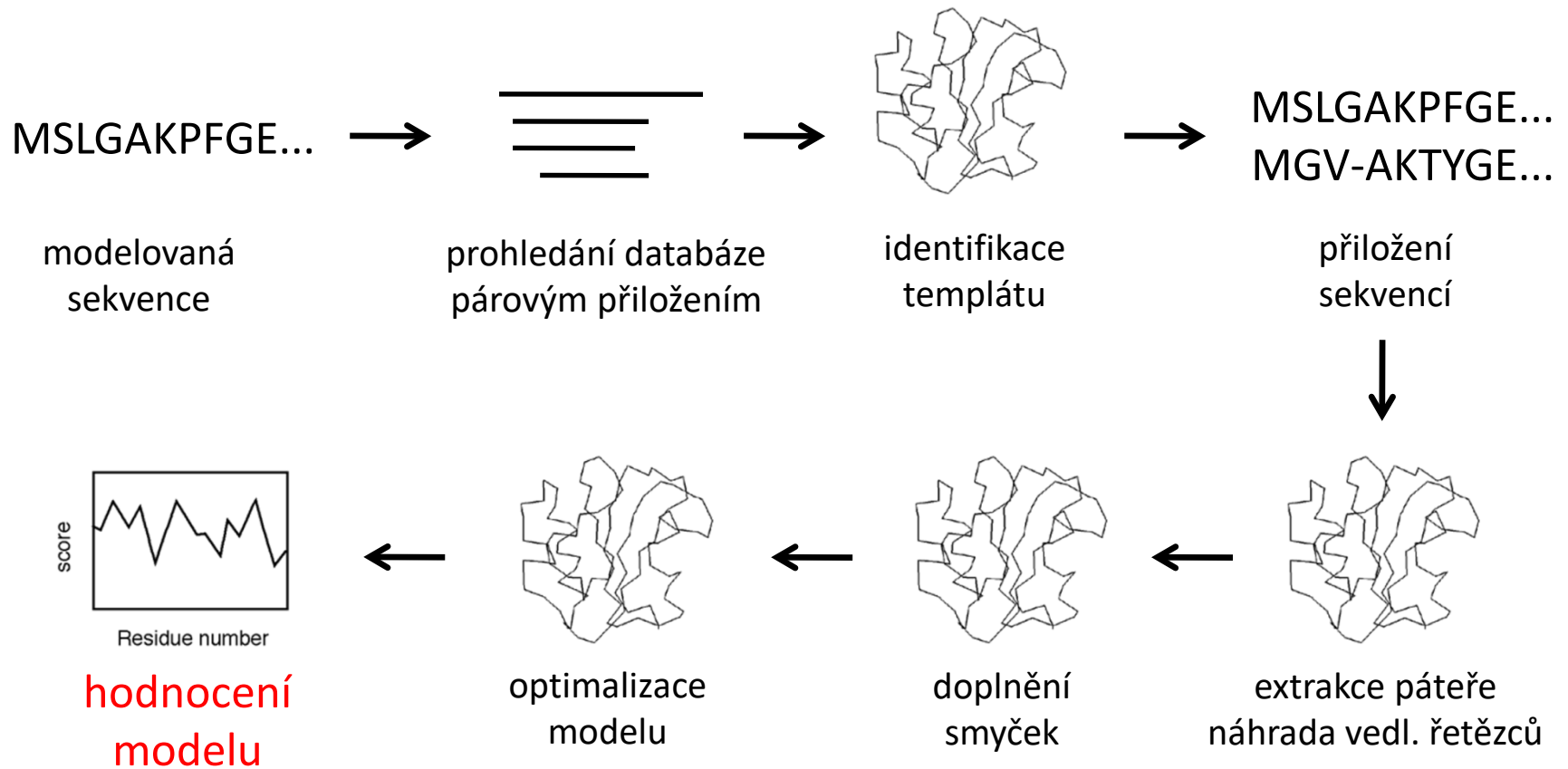
# Predikce terciární struktury

## □ Homologní modelování



# Predikce terciární struktury

## □ Homologní modelování



# Predikce terciární struktury

- Homologní modelování
  - Swiss-Model
  - Modeller

The screenshot shows the SWISS-MODEL Workspace interface. At the top, there are logos for SIB and BIOZENTRUM, and the text 'SWISS-MODEL Workspace' with navigation links for 'Modelling', 'Tools', 'Repository', and 'Document'. Below this is a user session indicator '[ myWorkspace ]' and a '[ log ]' link. The main section is titled 'SwissModel Automatic Modelling Mode' with a help icon. It contains input fields for 'Email:' and 'Project Title:'. Below these is a large text area for 'Provide a protein sequence or a UniProt AC Code:'. A 'Submit Modelling Request' button is located below the text area. At the bottom, there are 'Advanced options:' including 'Use a specific template:' with a PDB-ID and Chain field, and 'Template file:' with a file input field and a 'Procházet...' button.

# Predikce terciární struktury

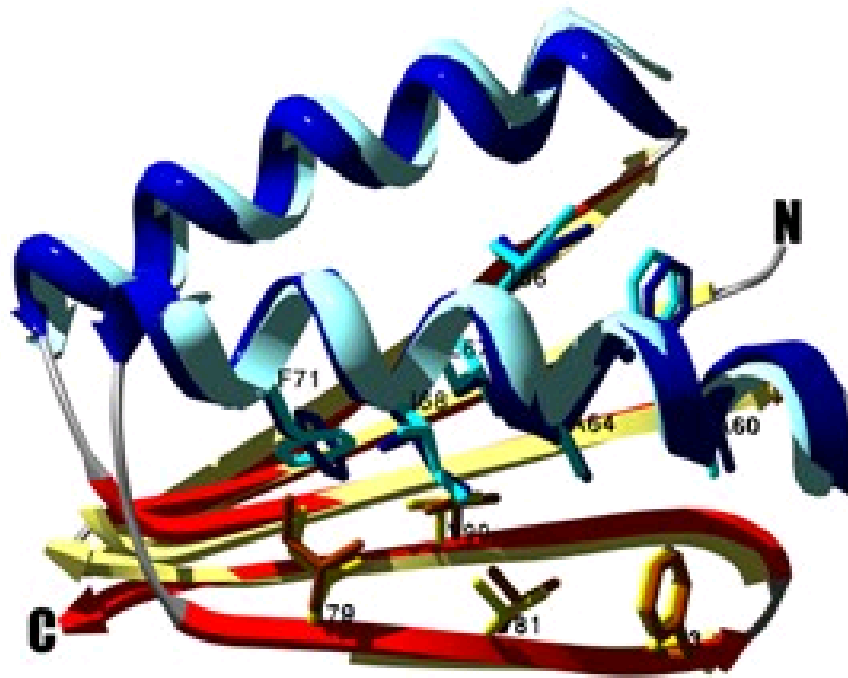
## □ *Ab initio* predikce

- Vytváří atomistický model založený na **základních fyzikálních principech**
- Hledá geometrii struktury v **globálním** energetickém minimu
- Umožňuje navrhovat struktury neexistující v přírodě
- “Svatý Grál” bioinformatiky



# Predikce terciární struktury

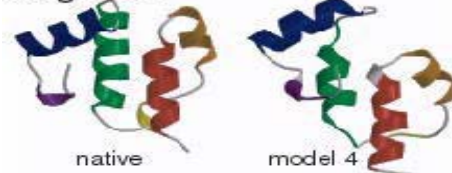
- *Ab initio* predikce
  - Rosetta, Robetta



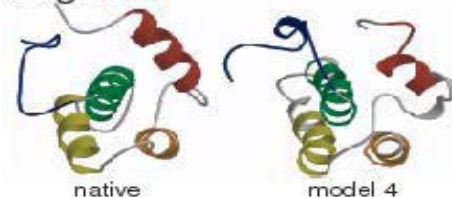
Target 77



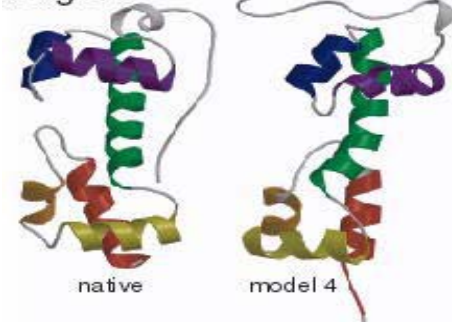
Target 56



Target 74



Target 79



# Predikce terciární struktury

## Meta-servery

- GeneSilico
- 3D-Jury

TERTIARY STRUCTURE PREDICTION		3dpsm server not updated				1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....
	score	identity	scop	EC	MTISADISLHHRVAVLGS	
1d07 A	0.071	42%	c.69.1.8	3.8.1.5	---CAKPFGEKKFIEIKGRMAYIDEC--TGDPIILFQHCNPTSSYLWRNIMPHCAGLGRLLIACDLIGMCDSDKPPERYAYAEHRDYLDALWEALDLDLP	
1bn7 A	0.2	49%	c.69.1.8	3.8.1.5	IGTCFPDFPHYVEVLGERMHYVDVGPDRDGTVPVFLHGNPTSSYLWRNIIPHVAPSHRCIAPDLIGMCKSDKPPDLDFDHDVRYLDFAFIRALGLEE	
1ek1 A2	0.2	21%	c.69.1.11	3.3.2.9	LPPVNDVSHGCVTVKCIIRLHFVEMCS--CPALCLCHCFPESWFSWRYQIPALAAQFRVLAIDMKCYGDSSSPPEYAMELLCKEMVTFIDKLGIPQ	
1s8o A	0.2	23%	c.69.1	3.3.2.9	MCNPSDMSHGCVTVKRVRLHFVLECS--CPAVCLCHCFPESWFSWRYQIPALAAQFRVLAIDMKCYGESSAPPEYCEVLECKEMVTFIDKLGCLSQ	
1cr6 A	0.42	21%	c.69.1	3.3.2.9	HPVNDVSHGCVTVKCIIRLHFVEMCS--CPALCLCHCFPESWFSWRYQIPALAAQFRVLAIDMKCYGDSSSPPEYAMELLCKEMVTFIDKLGIPQ	
1ehy A	0.71	20%	c.69.1.11	3.3.2.9	-AIPEDFKHYEVQLPDKIHYVREC--ACPTLLLLHGWPCFVWVWESKVICPLAEHYDVIIVPDLRGFGDSSEKPPSKYSLDKAADDQAALLDALGIER	
1b6g	0.93	25%		3.8.1.5	HLDQYFSPNYLDYPLCLRAHYLDEGNSDAEVFLCLHCEPTWVSYLRKMIIPVFAECARVIAADPFVFGKSDKPEEDYTFEPRNFFLALIERLDLRN	
1jli A	2	20%	c.69.1.10	3.7.1.8	----AYVERRFVNACGVETRYLEAC--RCQPVILIHCGAESSEGNWRNVIPILARHYRVIAMDLGFCCKTAKPDIETQDRRIRHLDHFIKAMNFDK	
1iuo A	2.2	20%	c.69.1.10	3.7.1.9	---NLEIG-KSILAACVLTNYHDVC--EGQPVILIHGSPCSAYAWRLTIPALSKFYRVIAADPMVFCFTDRPEYNYSKDSWVDHIIGIMDALEIER	
1brt	2.7	19%		1.11.1.10	-----PFTITVQENSIDLYYEDHC--TGQPVILIHGFPLSCHSWERQSAALDAGYRVITVDRRCFGQSSQPTTCGYDYDTFAADLNTVLETLDLQD	

TERTIARY STRUCTURE PREDICTION		phyre				1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....
	score	identity	scop	EC	MTISADISLHHRVAVLGS	
1bn7 A	1.8e-43	50%	c.69.1.8	3.8.1.5	----PDPHYVEVLGERMHYVDVGPDRDGTVPVFLHGNPTSSYLWRNIIPHVAPSHRCIAPDLIGMCKSDKPPDLDFDHDVRYLDFAFIRALGLEE	
1y37 B	5.8e-41	22%	c.69.1.11	3.8.1.3	--MFEGFERLLVDVCGDVTINCVVCG--SGPALLLLHGFQPNLHMWARVAPLLANETVWCADLRGCGSSKPHANYFRAMASDQRELMLTLCFEP	
1ehy A	2.7e-40	21%	c.69.1.11	3.3.2.9	IRRPEDFKHYEVQLPDKIHYVREC--ACPTLLLLHGWPCFVWVWESKVICPLAEHYDVIIVPDLRGFGDSSEKPPSKYSLDKAADDQAALLDALGIER	
2o2h A	4.2e-40	40%	c.69.1.8	3.8.1.5	MTAFPPYGPQPKYLEIACKRMAYIDEC--RGDAIVFQHCNPTSSYLWRNIMPHLEGLGRVACDLIGMCAASDKLSDRYSYGCEQRDFLFDALDLDLH	
1mj5 A	8.8e-40	43%	c.69.1.8	3.8.1.5	LGAKPFGEKKFIEIKGRMAYIDEC--TGDPIILFQHCNPTSSYLWRNIMPHCAGLGRLLIACDLIGMCDSDKLDERYAYAEHRDYLDALWEALDLDLP	
1va4 A	2.6e-39	17%	c.69.1.12	3.1.1.2	-----TFVAKDGTQIYFEDWC--SGKPVLFSHGWLLDADMWRYQMEYLSGCRYTIAFDRRGFRSDQPWTCNDYDTFADDLAQLIEHLDLKE	
1jli A	4.1e-39	18%	c.69.1.10	3.7.1.8	----RAYVERRFVNACGVETRYLEAC--RCQPVILIHCGAESSEGNWRNVIPILARHYRVIAMDLGFCCKTAKPDIETQDRRIRHLDHFIKAMNFDK	
2r11 A	4.7e-39	16%	c.69.1.12	3.1.1.1	--WPVRCKSFYITRFQTHVIASCPEDAPLVLLHGCALPSSMTWYPNIAWSSKRYRTYAVDIIIGKMKNSIPENSCTRDTYANWLLDVFVNDLIGIER	
2vf2 A	5.4e-39	21%	c.69.1.10	3.7.1.8	-TFESTSRFAEVDVDCIKLHYHEACVGNQDQTVVLLHCGGPGAAASWTNFIATVLAARHFVLAVDQPCYCHSDKRAEHQFNRYAAMALKGLFDQLGLCP	
1a8s A	5.6e-39	17%	c.69.1.12	1.11.1.10	-----TFTTRDGTQIYYEDWC--SGQPIVFSHCWPLNADSWESQMIFLAQCYRVIAHDRRGHGRSSQPWSCNDMDTYADDLAQLIEHLDLDD	

# Predikce molekulárních komplexů

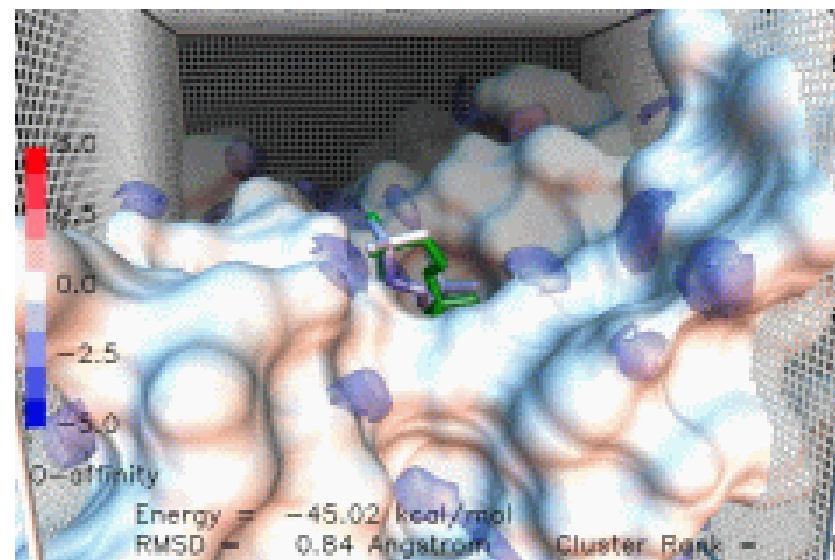
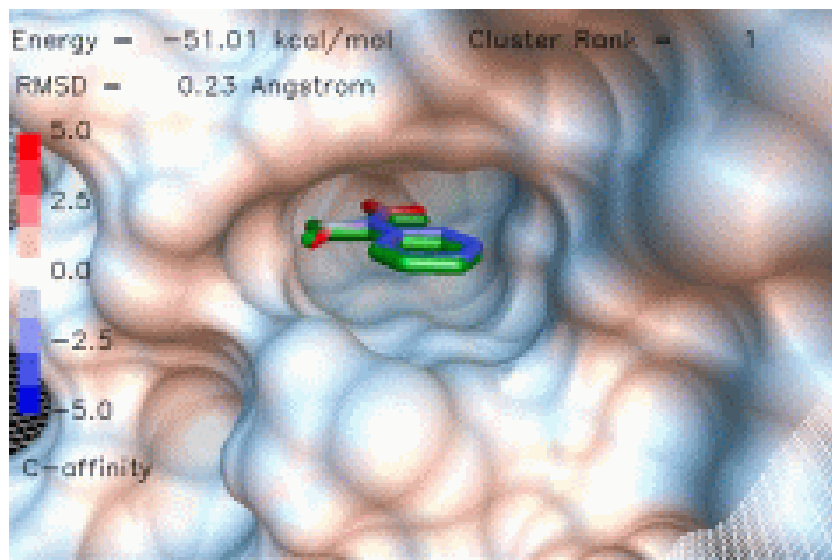
## □ Molekulární dokování

- **Umístování** malých organických molekul – **ligandů** – do vazebných domén receptorů, aktivních center enzymů nebo žlábků DNA
- Náhodně generované orientace a konformace ligandu v blízkosti biomolekuly jsou hodnoceny **energetickým skóre**
- Energetické skóre = **interakční energie** = van der Waalsova energie + elektrostatická energie + energie vodíkových vazeb + entropie



# Predikce molekulárních komplexů

- Molekulární dokování
  - DOCK
  - AUTODOCK





# Hodnocení predikčních metod

## □ CASP

- **C**ritical **A**ssessment of Techniques for Protein **S**tructure **P**rediction
- Mezinárodní soutěž **spolehlivosti** predikčních metod = umožňuje kritické a objektivní hodnocení
- K hodnocení jsou využívány **slepé predikce** = soutěžící obdrží proteinové sekvence se známou, avšak dosud nepublikovanou strukturou – organizátoři porovnají predikované a experimentální struktury





# Hodnocení predikčních metod



## □ CASP

- Predikce terciární struktury
- Predikce molekulárních komplexů
- Predikce kontaktů mezi zbytky
- Predikce neuspořádaných regionů
- Predikce domén
- Predikce funkce proteinů
- Hodnocení kvality modelů
- Upřesnění modelů

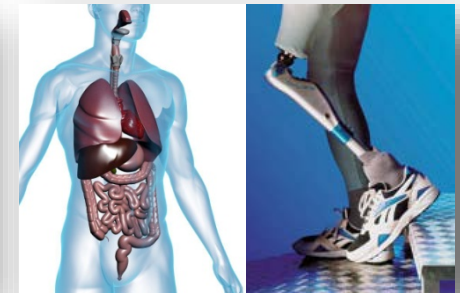
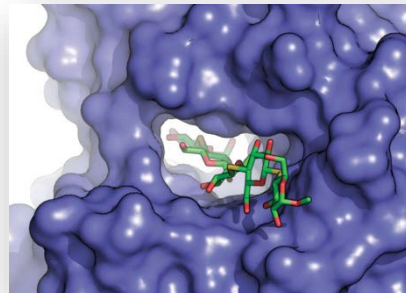
# Hodnocení predikčních metod

## □ CASP

#	▲ GR #	▲ GR Name	▼ Domains Count	▼ Cumulative Z-score (GDT_TS)
1.	426 <b>S</b>	Zhang-Server	164	129.070
2.	438 <b>S</b>	RAPTOR	164	103.274
3.	425 <b>S</b>	BAKER-ROBETTA	164	101.546
4.	409 <b>S</b>	pro-sp3-TASSER	164	99.062
5.	182 <b>S</b>	METATASSER	164	96.045
6.	322 <b>S</b>	Phyre_de_novo	164	91.897
7.	012 <b>S</b>	HHpred5	164	89.393
8.	020 <b>S</b>	MULTICOM-CLUSTER	164	87.470
9.	122 <b>S</b>	HHpred4	164	86.429
10.	256 <b>S</b>	SAM-T08-server	164	84.731
11.	013 <b>S</b>	MULTICOM-REFINE	164	83.832
12.	443 <b>S</b>	MUProt	164	81.140
13.	279 <b>S</b>	GS-KudlatyPred	161	78.994
14.	154 <b>S</b>	HHpred2	164	77.955

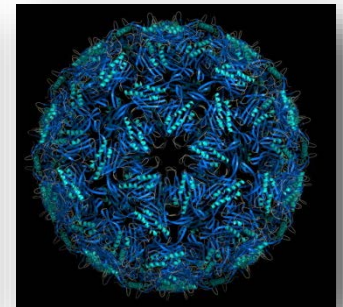
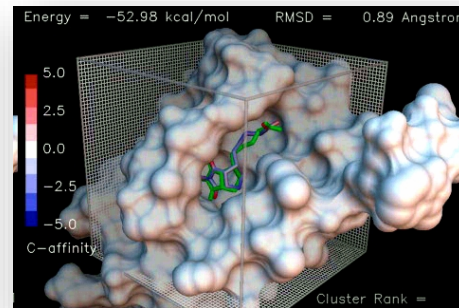
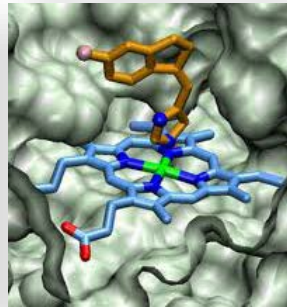
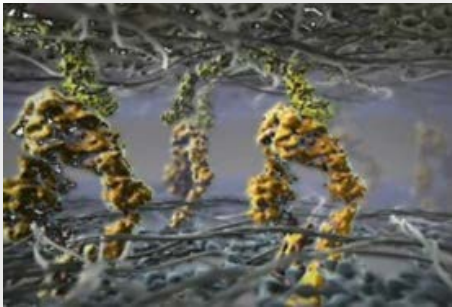
# Proteinové inženýrství Bi7410

- Období: jaro
- Rozsah: přednáška 1 hodina/týden
- Vyučující: Mgr. Radka Chaloupková, Ph.D.
- Osnova:
  - **strukturně-funkční vztahy proteinů**
  - **metody exprese a purifikace rekombinantních proteinů**
  - **metody strukturní a funkční analýzy proteinů**
  - **racionální design, semi-racionální design a řízená evoluce**
  - **příklady využití proteinového inženýrství**



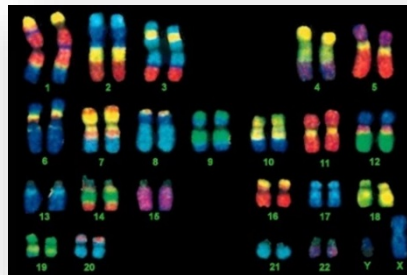
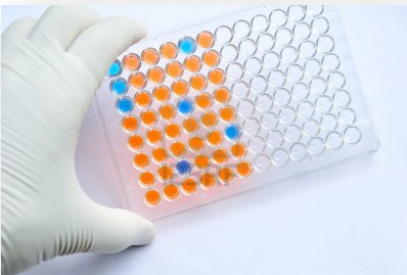
# Strukturní biologie Bi9410+9410c

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Vyučující: Mgr. David Bednář, Ph.D.
- Osnova:
  - struktura, stabilita a dynamika biologických makromolekul
  - makromolekulární interakce a komplexy
  - stanovení a předpověď struktury, identifikace důležitých oblastí
  - stanovení vlivu mutace na strukturu a funkci proteinu
  - aplikace v biologickém výzkumu, návrhu léčiv a biokatalyzátorů



# Molekulární biotechnologie Bi7430

- Období: podzim
- Rozsah: přednáška 2 hodiny/týden, cvičení 2 hodiny/týden
- Přednášky: Prof. Zbyněk Prokop, Ph.D.
- Cvičení: Dr. Šárka Bidmanová, Dr. Koen Beerens, Dr. Veronika Štěpánková, Mgr. Lukáš Chrást
- Osnova:
  - **proteinové, metabolické a tkáňové inženýrství**
  - **genetické inženýrství rostlin a živočichů**
  - **molekulární diagnostika, vakcíny, terapeutika**
  - **buněčná a genová terapie, regenerativní medicína**
  - **molekulární biotechnologie v průmyslu a zemědělství**



# 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.
- ❑ **PSI-PRED**: <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>
- ❑ **Quick2D (MPI Toolkit)**: [http://toolkit.tuebingen.mpg.de/quick2\\_d](http://toolkit.tuebingen.mpg.de/quick2_d)
- ❑ **Modeller**: <http://salilab.org/modeller/>
- ❑ **Modeller (GeneSilico)**: <https://genesilico.pl/toolkit/unimod?method=Modeller>
- ❑ **Swiss-Model**: <http://swissmodel.expasy.org/>
- ❑ **GenTHREADER**: <http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>
- ❑ **PHYRE**: <http://www.sbg.bio.ic.ac.uk/~phyre/index.cgi>
- ❑ **GeneSilico MetaServer**: <https://www.genesilico.pl/meta2/>
- ❑ **3D-Jury**: [http://meta.bioinfo.pl/submit\\_wizard.pl](http://meta.bioinfo.pl/submit_wizard.pl)
- ❑ **Rosetta@home**: <http://boinc.bakerlab.org/rosetta/>
- ❑ **CASP**: <http://predictioncenter.org/index.cgi>