

LOSCHMIDT
LABORATORIES



Predikce proteinové struktury



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



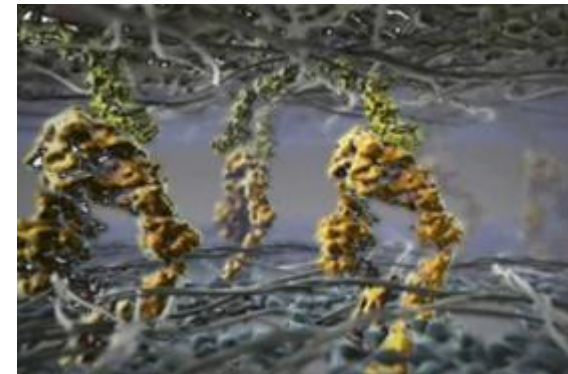
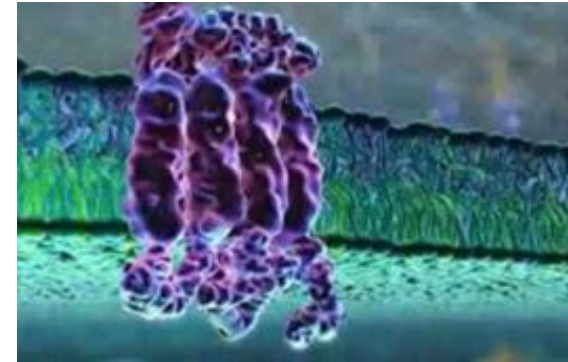
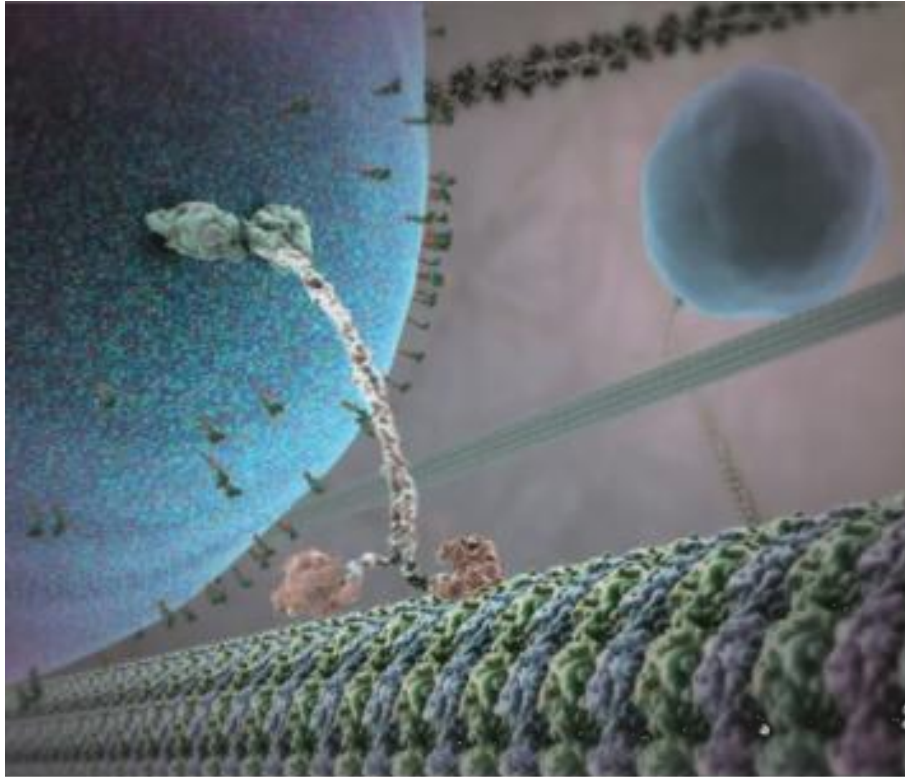
OP Vzdělávání
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

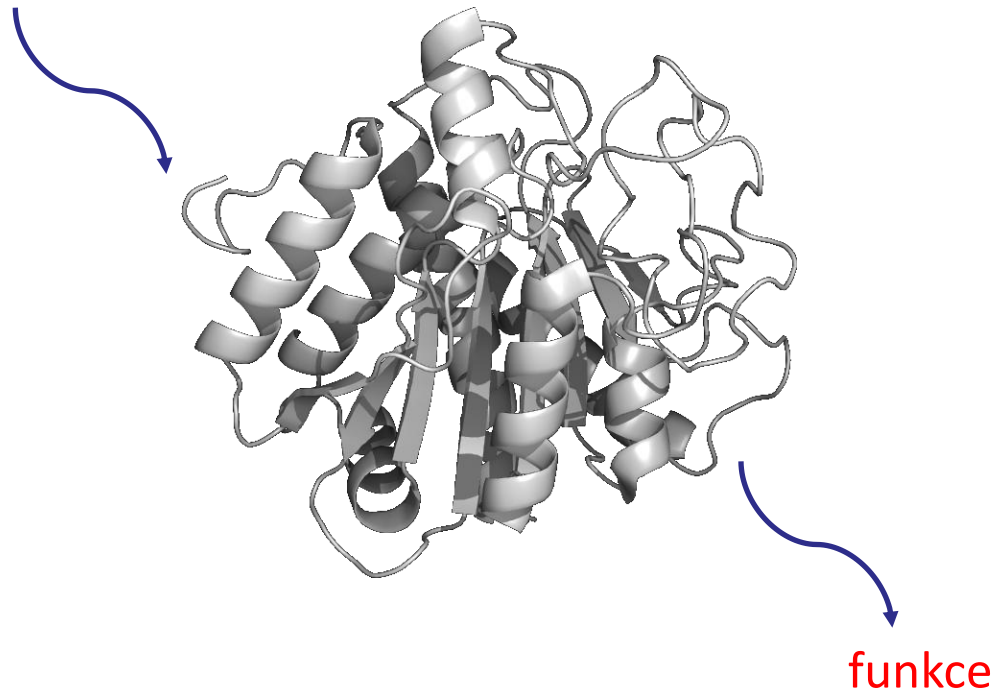
Důležitost proteinové struktury



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

Důležitost proteinové struktury

MSLGAKPFGEKKFIEIKGRRMAYIDEGTGDPILFQHGNTSSYLWRNIMPHCA
GLGRLIACDLIGMGDSKLDPSGPERYAYAEHRDYLDALWEALDLGDRVVLVV
HDWGSALGFDWARRHRERVQGIAYMEAIAMPIEWADFPEQDRDLFQAFRS
QAGEELVLQD



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**

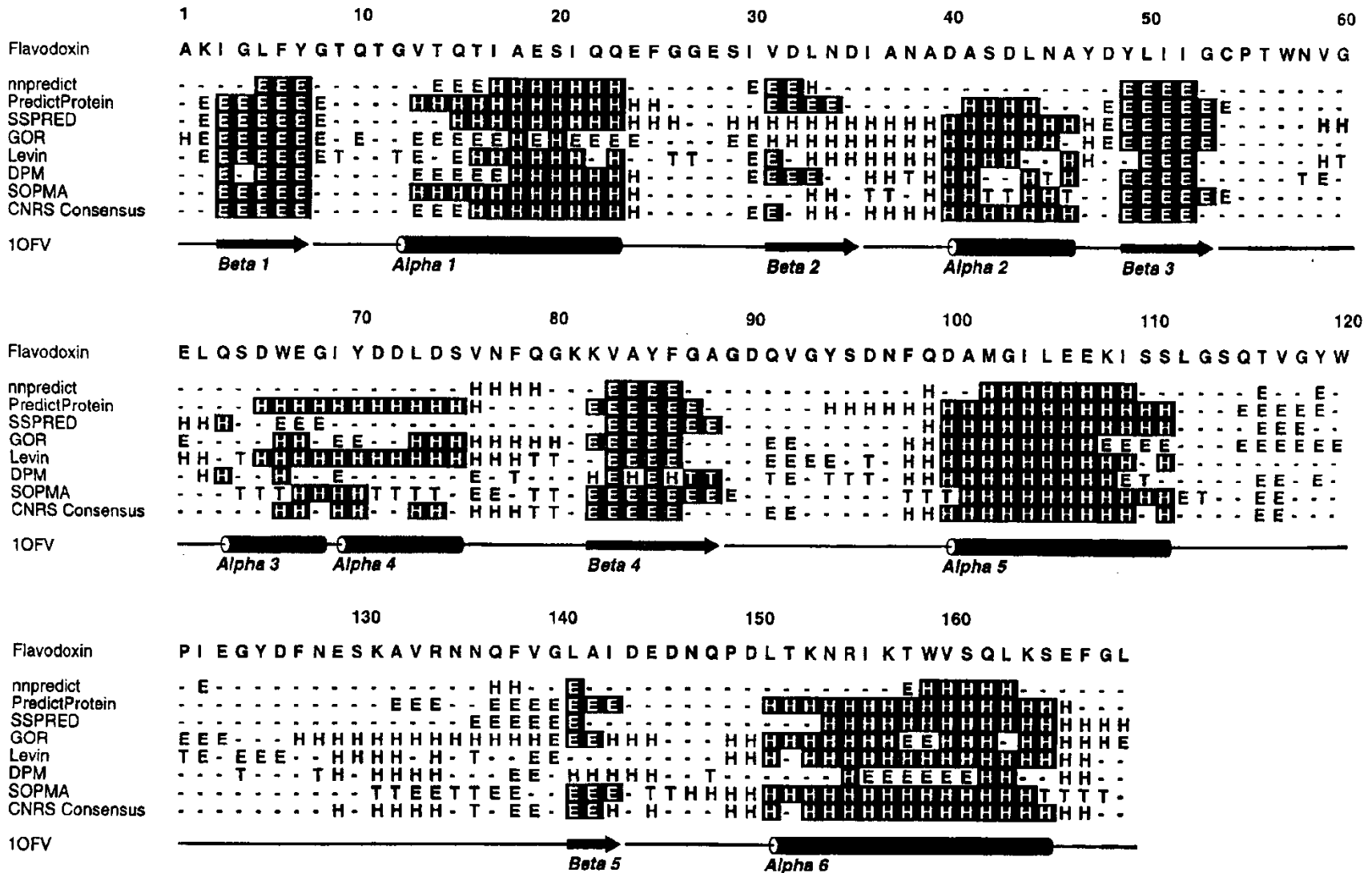


- Přiřazení jednoho **konformačního stavu** každému aminokyselinovému zbytku v proteinové sekvenci:
 - α -šroubovice (H, angl. helix)
 - β -řetězec (E, angl. strand)
 - otočka (C, angl. coil)



- 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

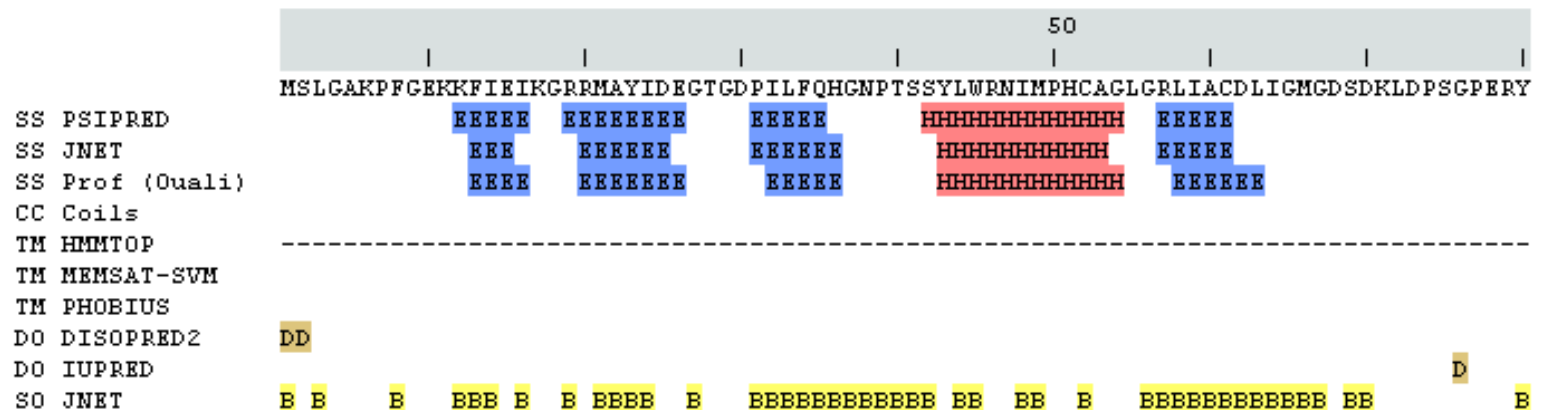
Predikce sekundární struktury



Predikce sekundární struktury

□ Quick2D

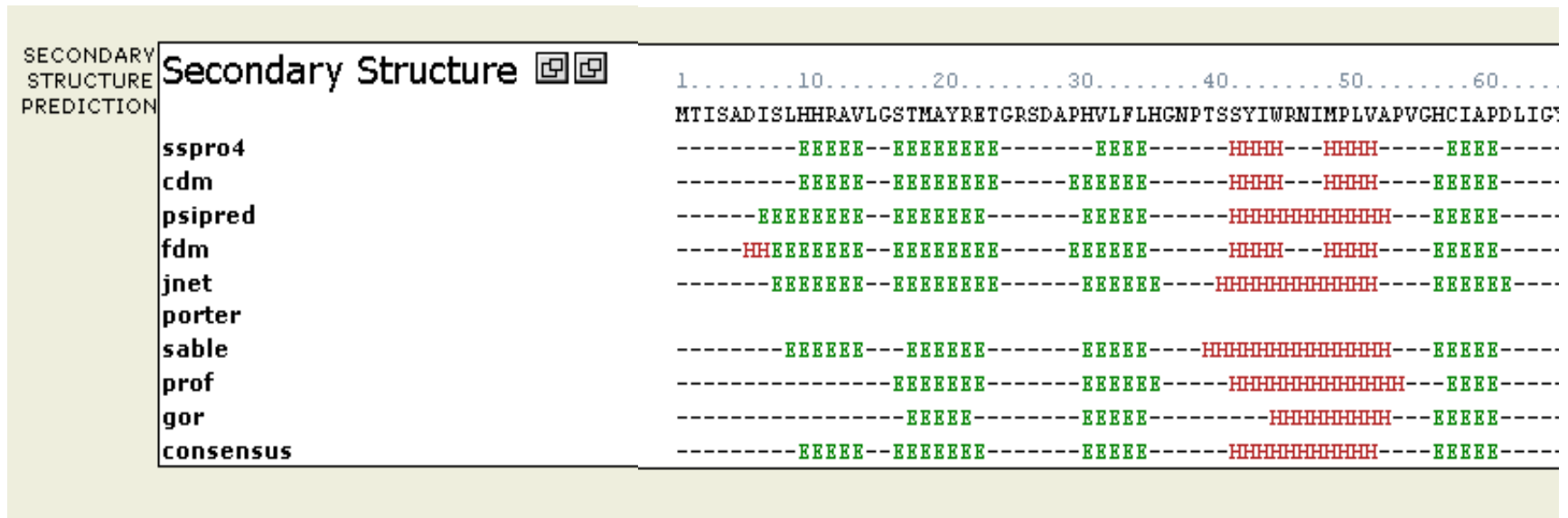
- Přiřazení sekundárních elementů: α -šroubovic, β -řetězců, otoček, transmembránových šroubovic a neuspořádaných regionů
- Metody PSI-PRED, JNET, Prof, Coils, MEMSAT2, HMMTOP, ...



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

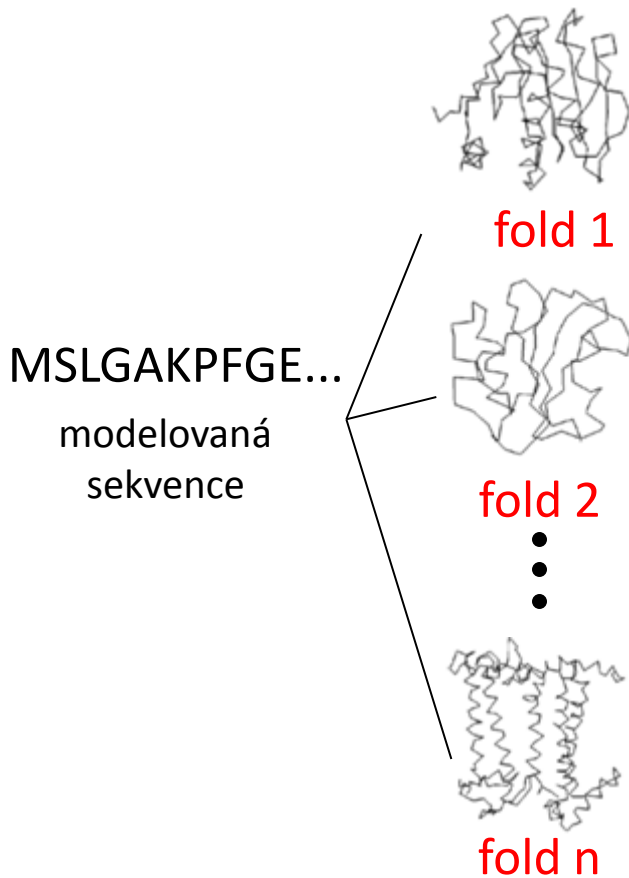
- Navlékání

MSLGAKPFGE...

modelovaná
sekvence

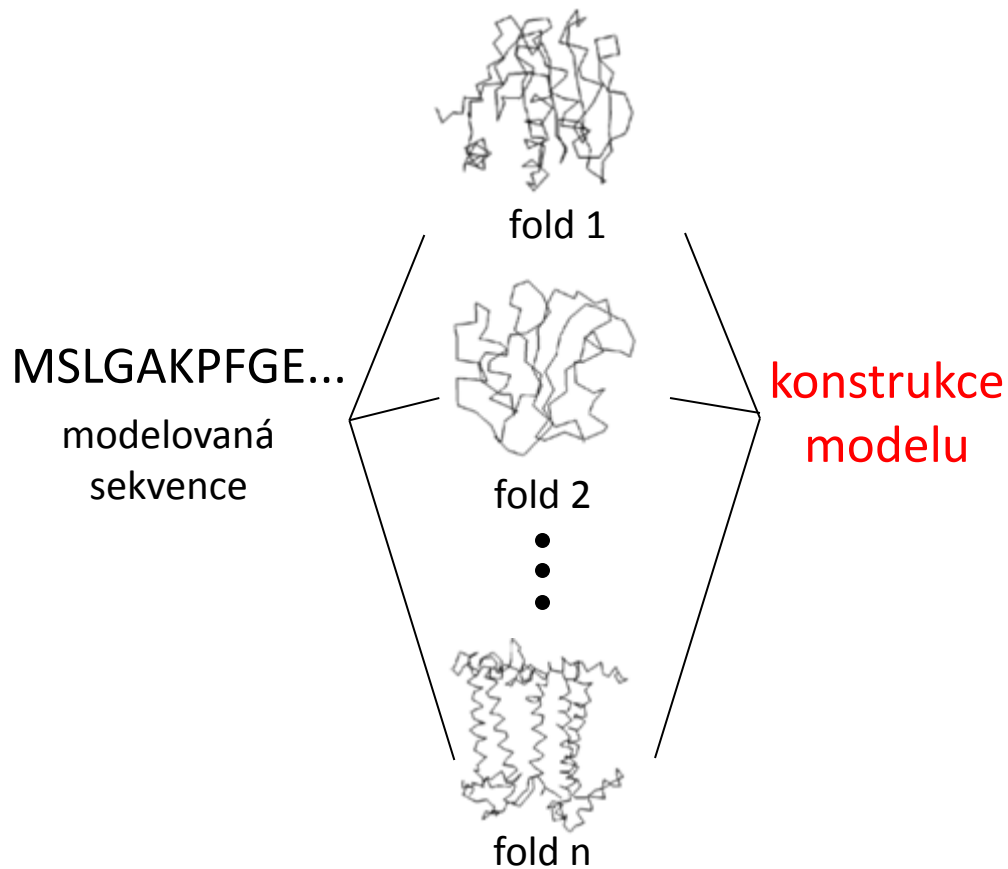
Predikce proteinového foldu

□ Navlékání



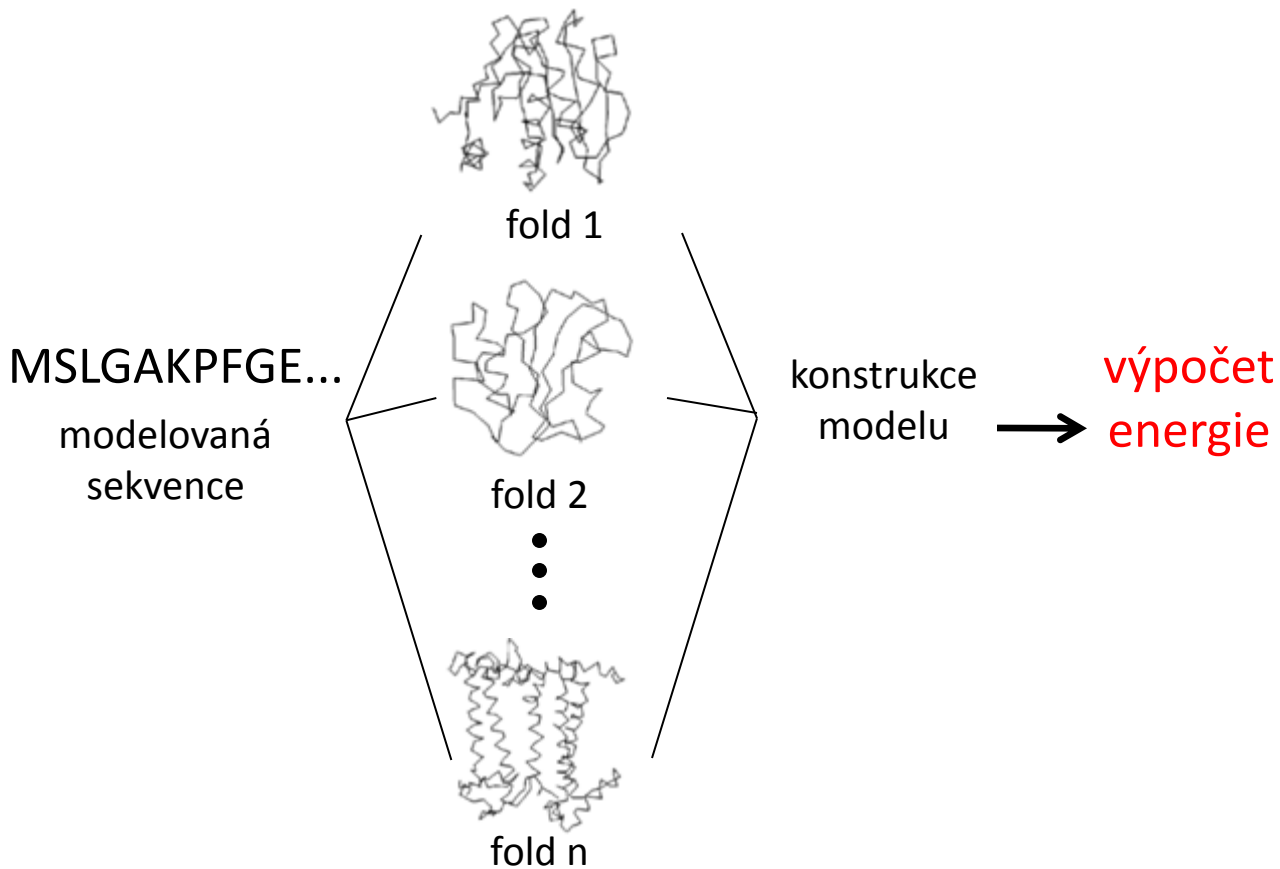
Predikce proteinového foldu

□ Navlékání



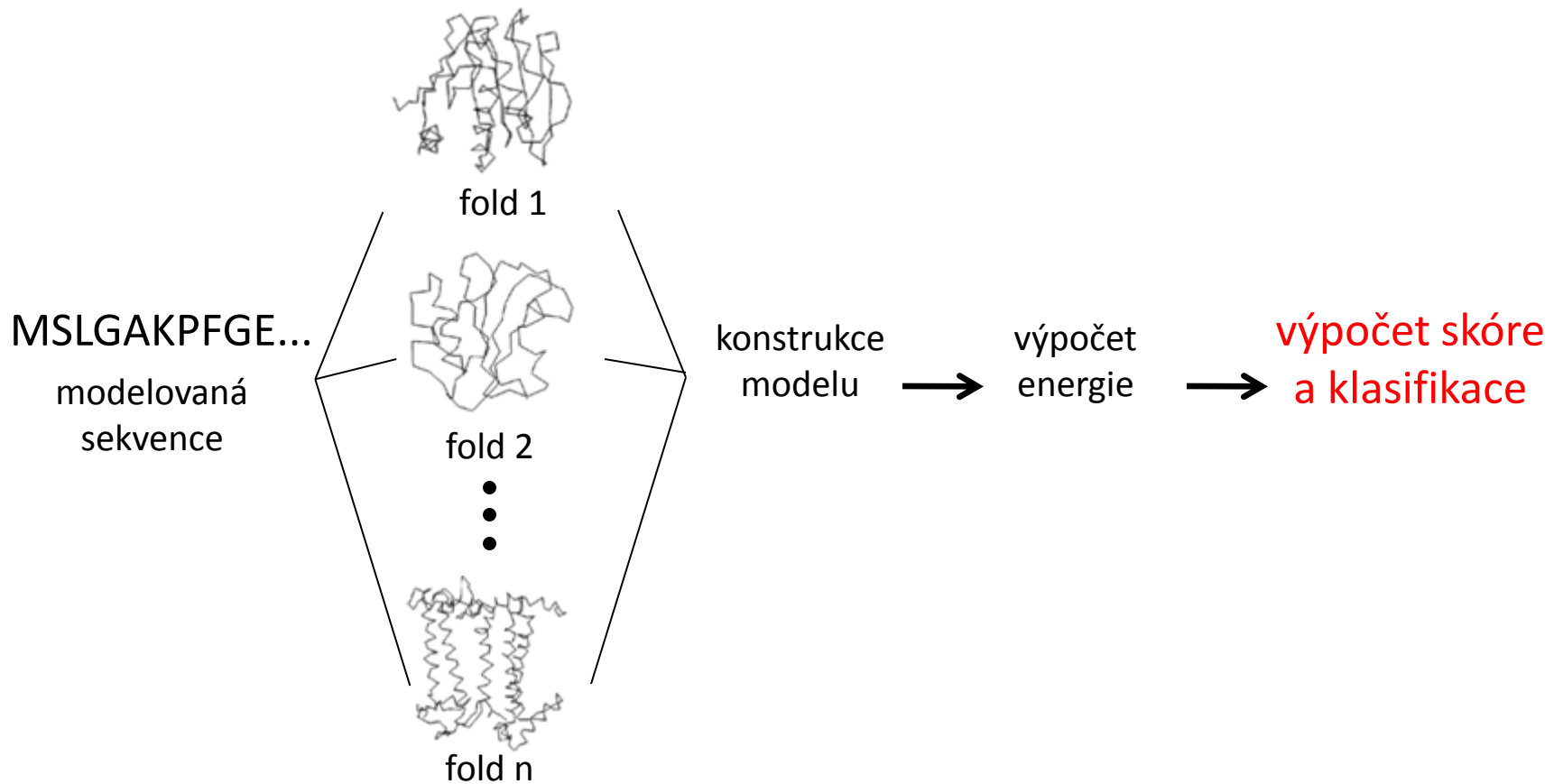
Predikce proteinového foldu

□ 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

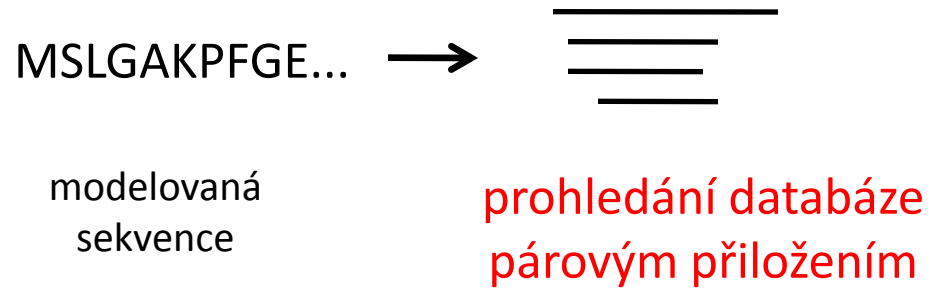
- Homologní modelování

MSLGAKPFGE...

modelovaná
sekvence

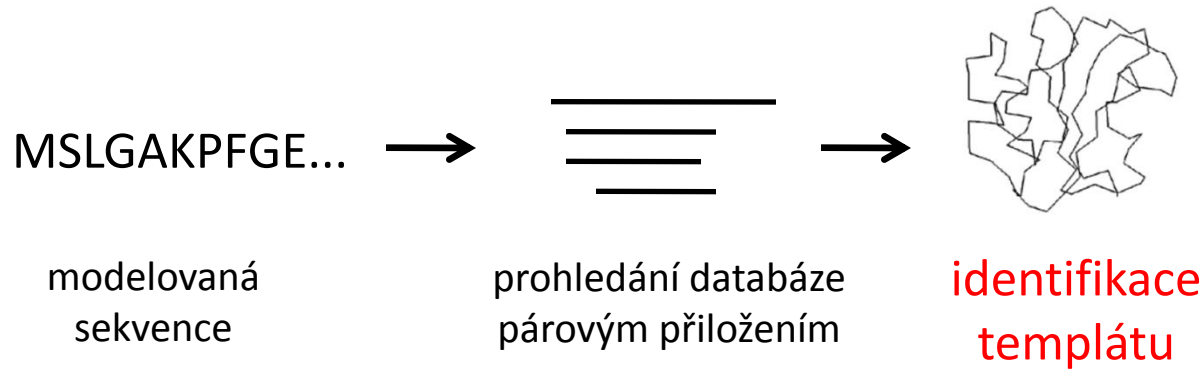
Predikce terciární struktury

□ Homologní modelování



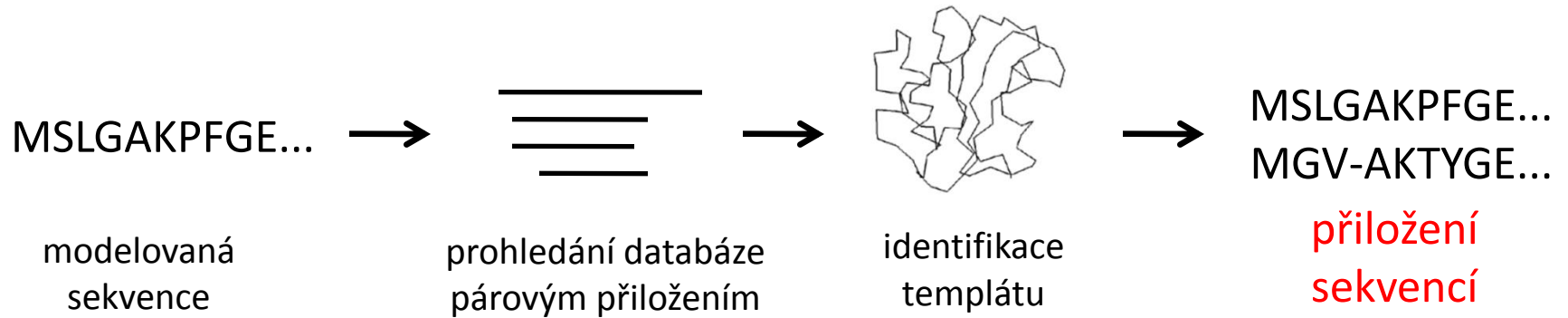
Predikce terciární struktury

□ Homologní modelování



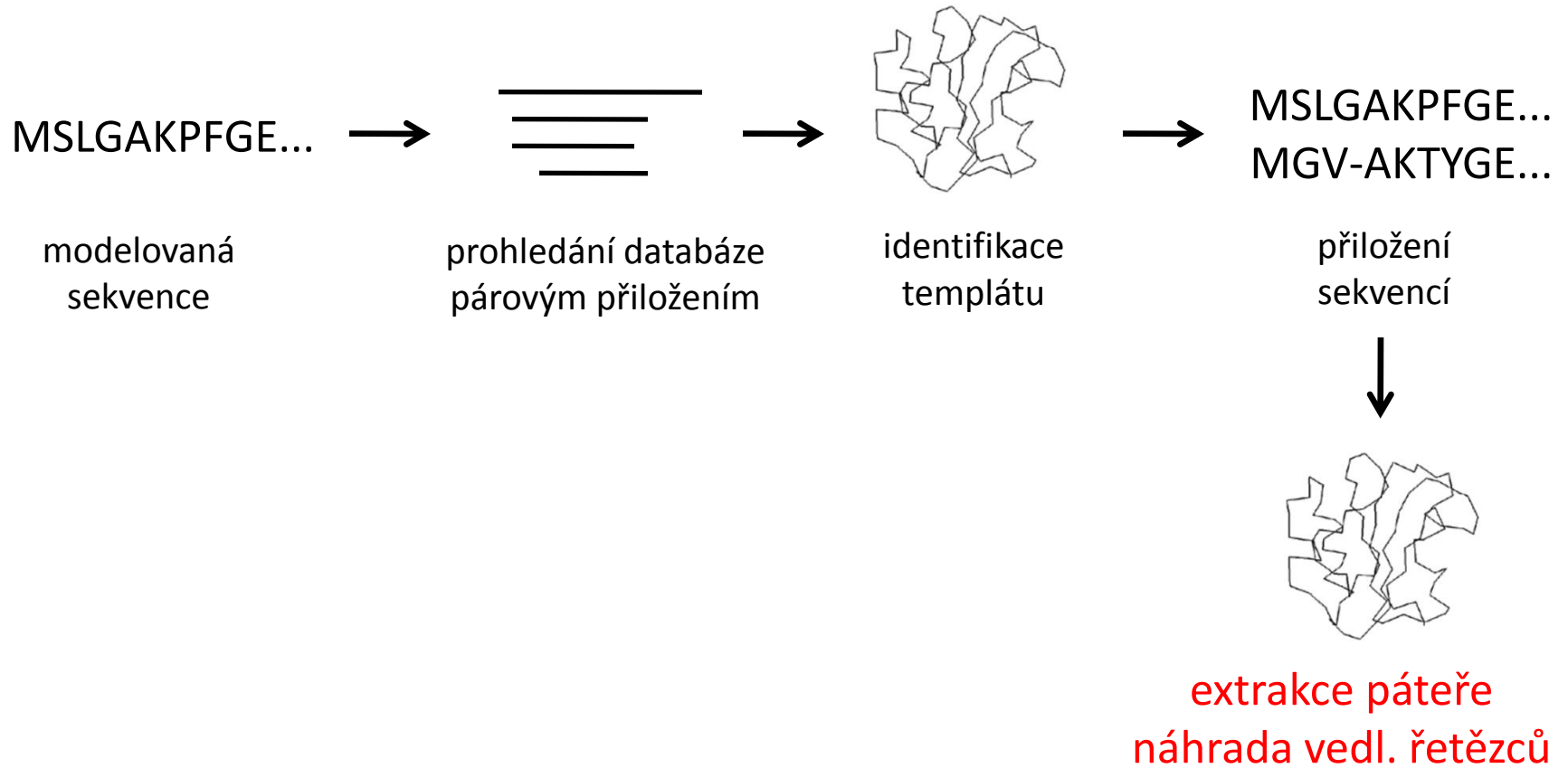
Predikce terciární struktury

□ Homologní modelování



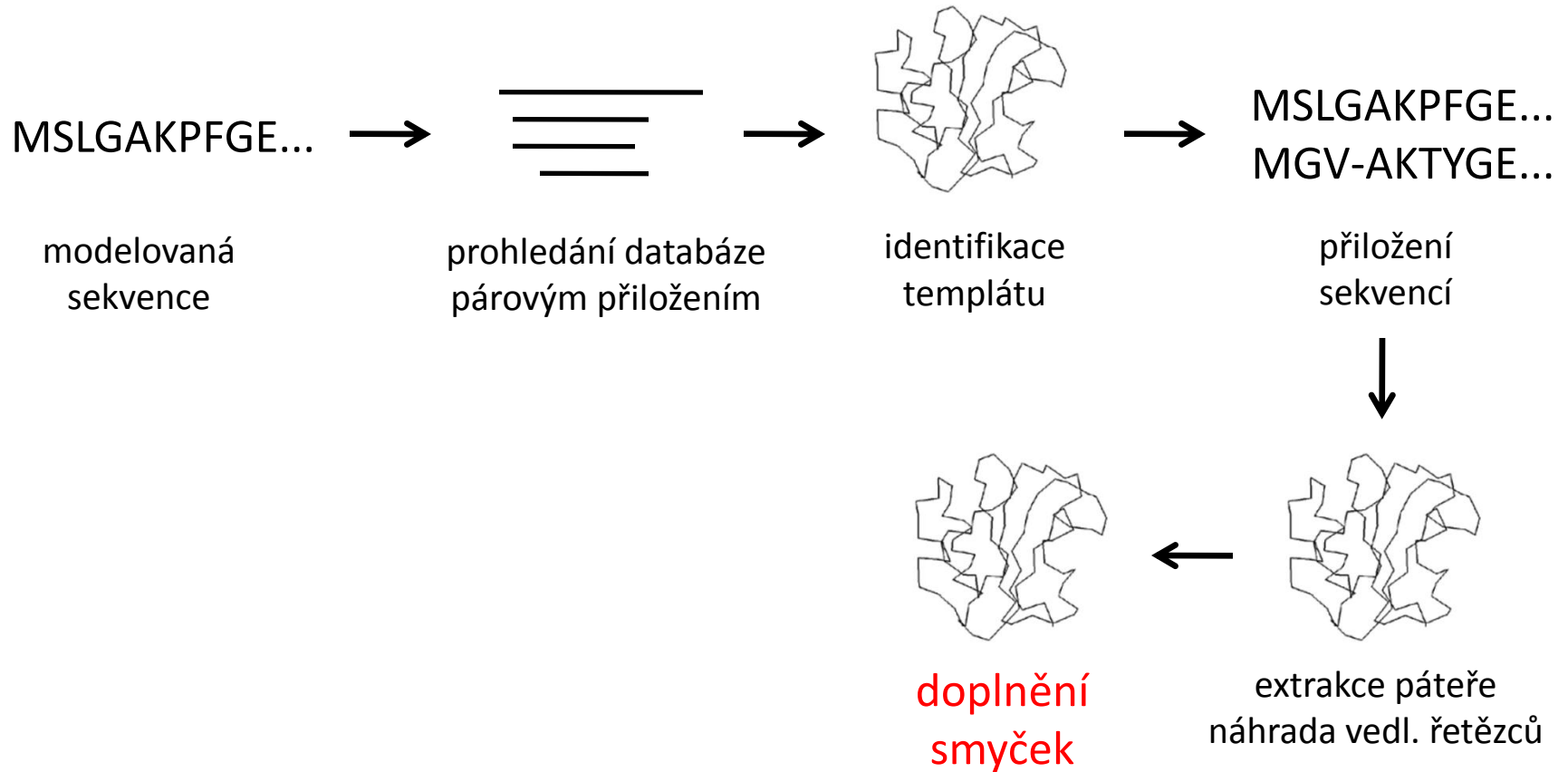
Predikce terciární struktury

□ Homologní modelování



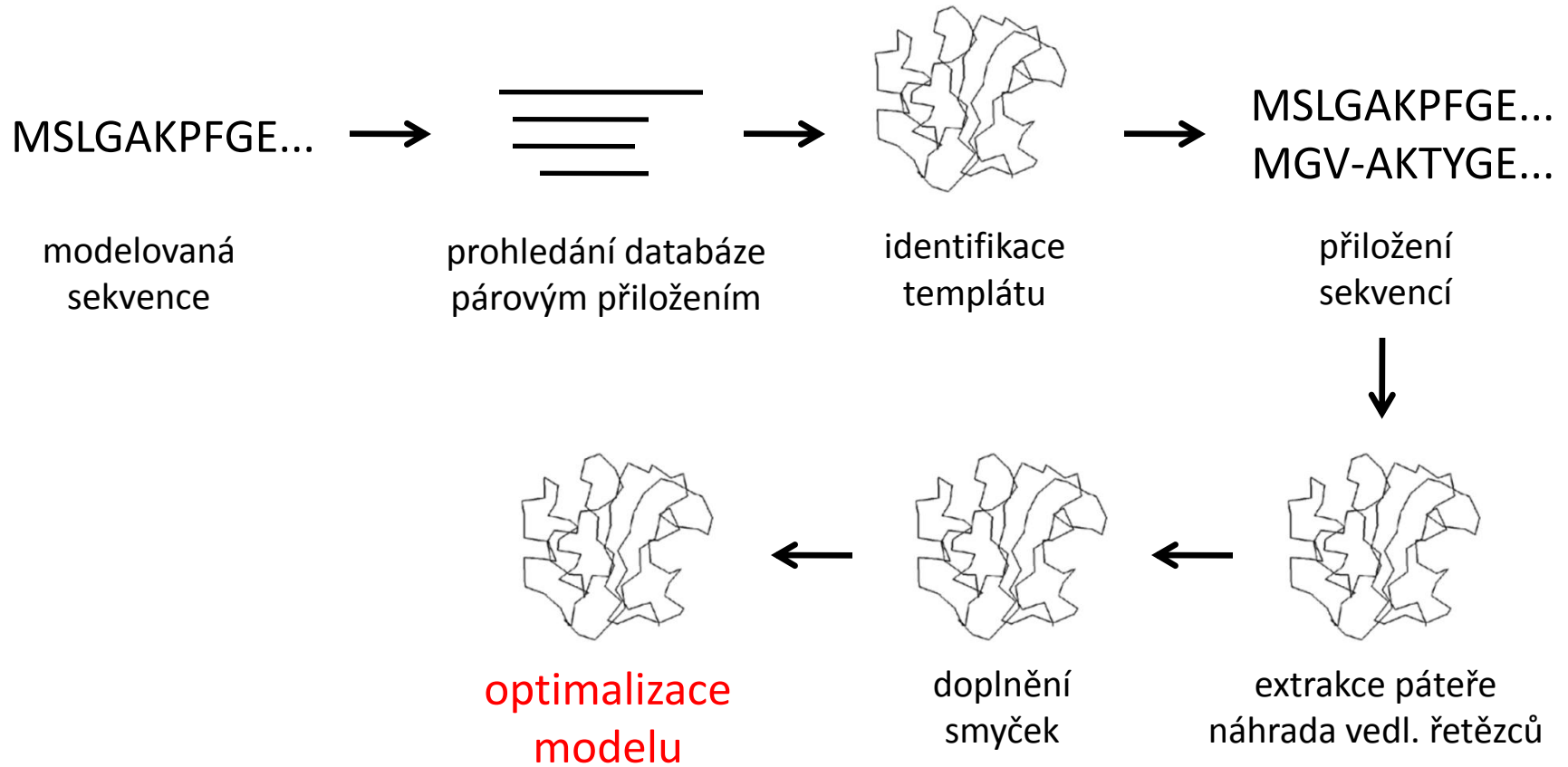
Predikce terciární struktury

□ Homologní modelování



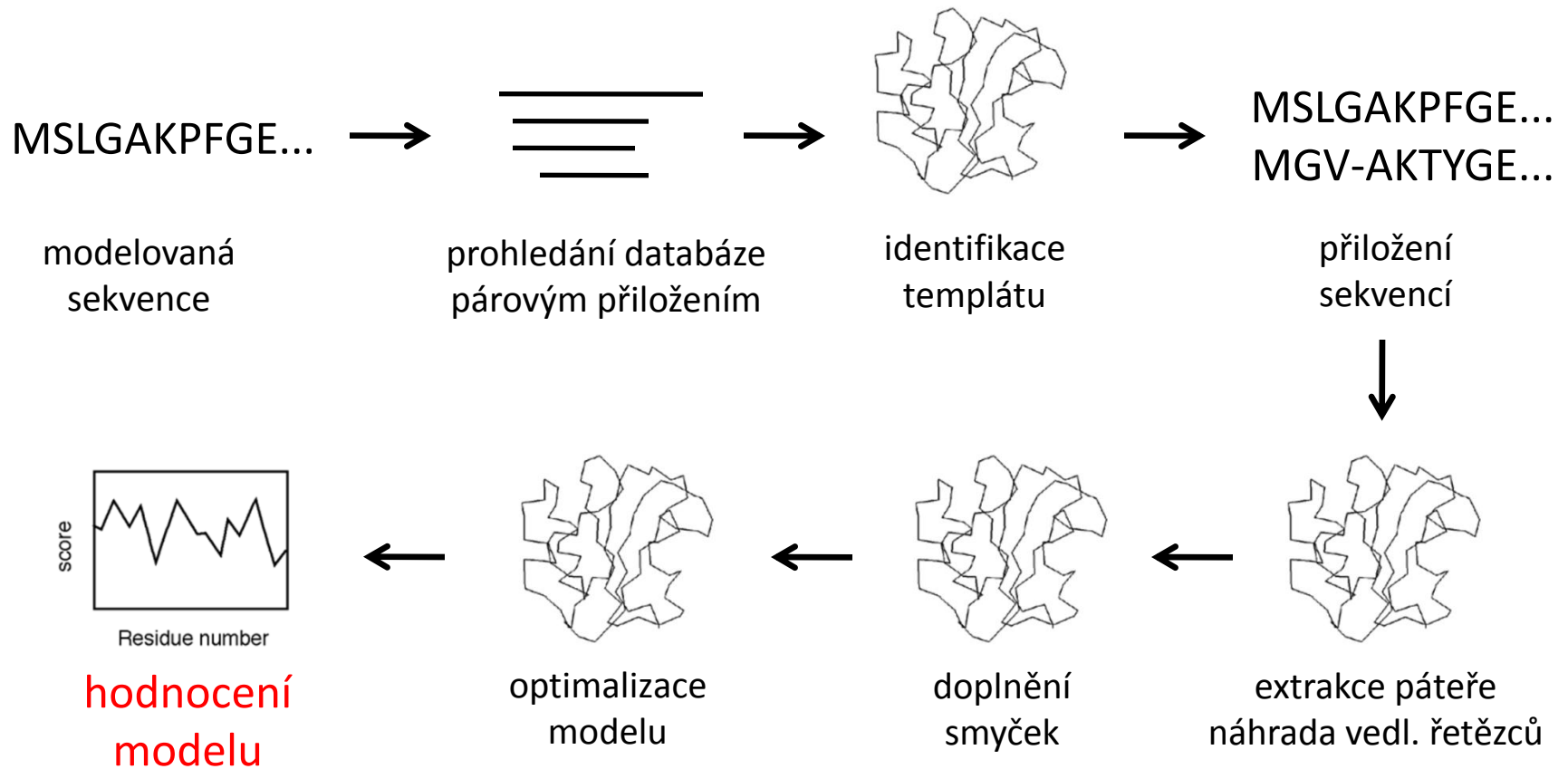
Predikce terciární struktury

□ Homologní modelování






Predikce terciární struktury

□ Homologní modelování



Predikce terciární struktury

- Homologní modelování
 - Swiss-Model
 - Modeller

   **SWISS-MODEL Workspace**
Modelling Tools Repository Document

[myWorkspace] [log]

SwissModel Automatic Modelling Mode ?

Email:
Project Title:

Provide a protein sequence or a UniProt AC Code: ?

Advanced options:

Use a specific template: ? PDB-ID: Chain:
or
Template file: ?

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



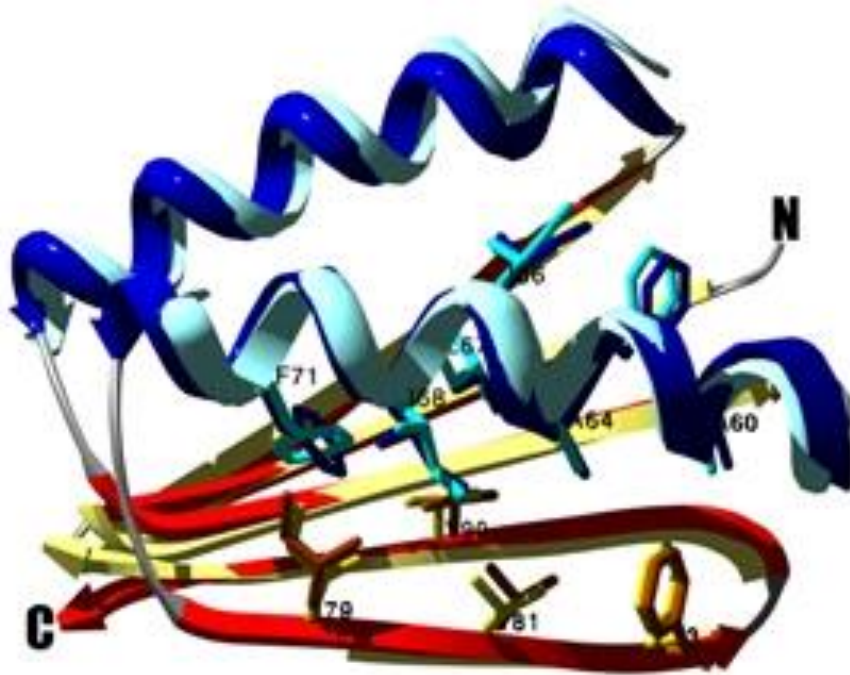
Predikce terciární struktury

- *Ab initio* predikce

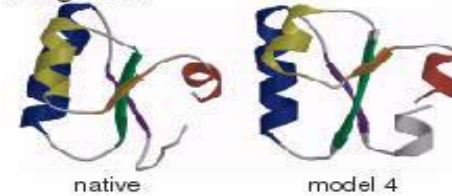


Predikce terciární struktury

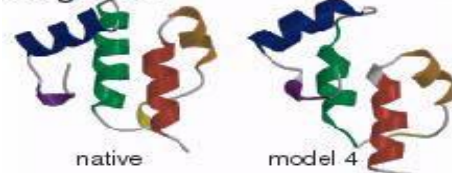
- *Ab initio* predikce
 - Rosetta, Robetta



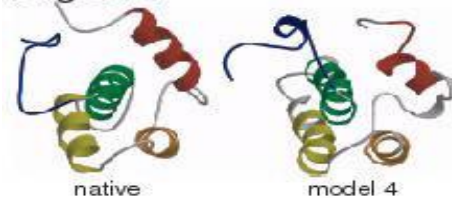
Target 77



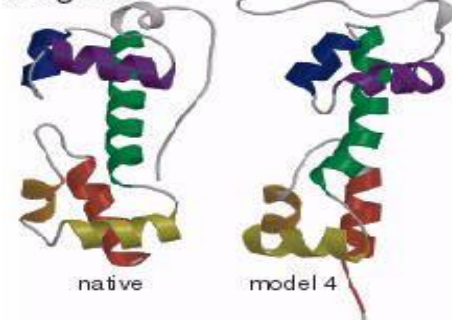
Target 56



Target 74



Target 79



Predikce terciární struktury

Meta-servery

- GeneSilico
- 3D-Jury

TERTIARY STRUCTURE PREDICTION		3dpssm				server not updated	1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....
	score	identity	scop	EC			
1d07 A	0.071	42%	c.69.1.8	3.8.1.5	MTISADISLHHRVLCSTMAYRETGRSDAPHVFLHGNPTSSYIWRNIMPLVAPVGHCIAPDLICYGQSGKPDISYRFFDQADYLDALIDELGAS		
1bn7 A	0.2	49%	c.69.1.8	3.8.1.5	--CARPFCEKRFIEIKGRMAYIDEG--TGDPIFLQHCNPTSSYLWRNIMPHCACLGRLLIACDLICMGDSKDPPEYAYARHHRDYLDALWEALDLDL		
1ek1 A2	0.2	21%	c.69.1.11	3.3.2.9	LPVPMDSVSHCYVTVKGRIRLHFVEMGS--GPAVCLCHCFPESWFSWRYPALAAACFRVLAIDMKCYGDSSSPPEEYAMELLCKEMVTFDRLKLGIPQ		
1s8o A	0.2	23%	c.69.1	3.3.2.9	MCNPSDMSHCYVTVKRVRLHFVVLGS--GPAVCLCHCFPESWFSWRYPALAAACFRVLAIDMKCYGESSAPPEEYCEVLECKEMVTFDRLKLGSLQ		
1cr6 A	0.42	21%	c.69.1	3.3.2.9	MPVPMDSVSHCYVTVKGRIRLHFVEMGS--GPAVCLCHCFPESWFSWRYPALAAACFRVLAIDMKCYGDSSSPPEEYAMELLCKEMVTFDRLKLGIPQ		
1ehy A	0.71	20%	c.69.1.11	3.3.2.9	--AIPEDFKHYEVQLPDVKIHYVREG--ACPTLLLHGWPCFVWWSKVICPLAEHYDVIVPDLRFGDSEKPPSKYSLDKAADDQAALLDALGIER		
1b6g	0.93	25%		3.8.1.5	MLDQYFSPNYLDYPLGRAHYLDEGNSDAEVFLCHGCEPTWSYLARKMIPVFAEGARVIAPDFFCFGKSDKPEEDYTFEFHRNFLALIERLDLRN		
1j1i A	2	20%	c.69.1.10	3.7.1.8	----AYVERRFVNAGCVETRYLEAG--RCQPVLLIHGCGASECNWRNVIPILARHYRVLIAMDMLCFGKTAKPDIEYTDQRRIRHLHDFIKAMNFDK		
1iuo A	2.2	20%	c.69.1.10	3.7.1.9	---NLEIG-KSILAACVLTNYHDVCG--RCQPVLLIHGSPCSAYAWRLTIPALSKFYRVLIAPDMVCGFGRDRPEYNYSKDSWVDHIIGIMDALIEIK		
1brt	2.7	19%		1.11.1.10	-----PFTIVQENSIDLYYEDHC--TGQPVLLIHGFPLSCHSWERQSAALDAGYRVITYDRRCFGQSSQPTGQYDYDFAADLMTVLETLDLQD		

TERTIARY STRUCTURE PREDICTION		phyre				1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....
	score	identity	scop	EC		
1bn7 A	1.8e-43	50%	c.69.1.8	3.8.1.5	MTISADISLHHRVLCSTMAYRETGRSDAPHVFLHGNPTSSYIWRNIMPLVAPVGHCIAPDLICYGQSGKPDISYRFFDQADYLDALIDELGAS	
1y37 B	5.8e-41	22%	c.69.1.11	3.8.1.3	----PFDPHYVEVLGERMHYVDVGPDRDGTVPVFLHGNPTSSYLWRNIMPHVAPSHRCIAPDLICMGKSDKPDLDYFDDHVRVYLDALFIEALGLEE	
1ehy A	2.7e-40	21%	c.69.1.11	3.3.2.9	IRRPEDFKHYEVQLPDVKIHYVREG--ACPTLLLHGWPCFVWWSKVICPLAEHYDVIVPDLRFGDSEKPPSKYSLDKAADDQAALLDALGIER	
2o2h A	4.2e-40	40%	c.69.1.8	3.8.1.5	MTAFPYGQPRYLEIACKRMAYIDEG--RGDAIVFQHCNPTSSYLWRNIMPHLEGLGRVACDLICMGASDKLSDRYSYGQRDFLFDALDLDLH	
1mj5 A	8.8e-40	43%	c.69.1.8	3.8.1.5	LCARPFCEKRFIEIKGRMAYIDEG--TGDPIFLQHCNPTSSYLWRNIMPHCACLGRLLIACDLICMGDSKDLERYAYAEHRDYLDALWEALDLDL	
1va4 A	2.6e-39	17%	c.69.1.12	3.1.1.2	-----TFVAKDQTQIYFKDWG--SGKPVVLSHGWLLDADMWYEQMEYLSGSCYRTIAFDRRFCGRSDQPWGNDYDFTADDIAQLIEHLDLKE	
1j1i A	4.1e-39	18%	c.69.1.10	3.7.1.8	----RAYVERRFVNAGCVETRYLEAG--RCQPVLLIHGCGASECNWRNVIPILARHYRVLIAMDMLCFGKTAKPDIEYTDQRRIRHLHDFIKAMNFDK	
2r11 A	4.7e-39	16%	c.69.1.12	3.1.1.1	--WPVRCSPYSTRFCQTHVIASCPEDAPPLVLLHGCALFSSMTWYPIADWSSSKYRTYAVDIIIGDKNKSIPENSCTRDTYANWLLDVPDNLGIER	
2vf2 A	5.4e-39	21%	c.69.1.10	3.7.1.8	-TFESTSRFAEVDVDCGLKLHYHEACVGMNDQTVVLLHCGCGPAAASWTNFIAVLARHFLVLAVDQPCYCHSDKRAEHQFNRYAAMALRGLDFQLGLGP	
1a8s A	5.6e-39	17%	c.69.1.12	1.11.1.10	-----TFTTRDCTQIYYKDWG--SQQPIVFSHCWPLNADSWESQMIFLAQCYRVLAHDRRCHGRSSQPWSGNDMDTYADDLAQLIEHLDLDR	

Predikce molekulárních komplexů

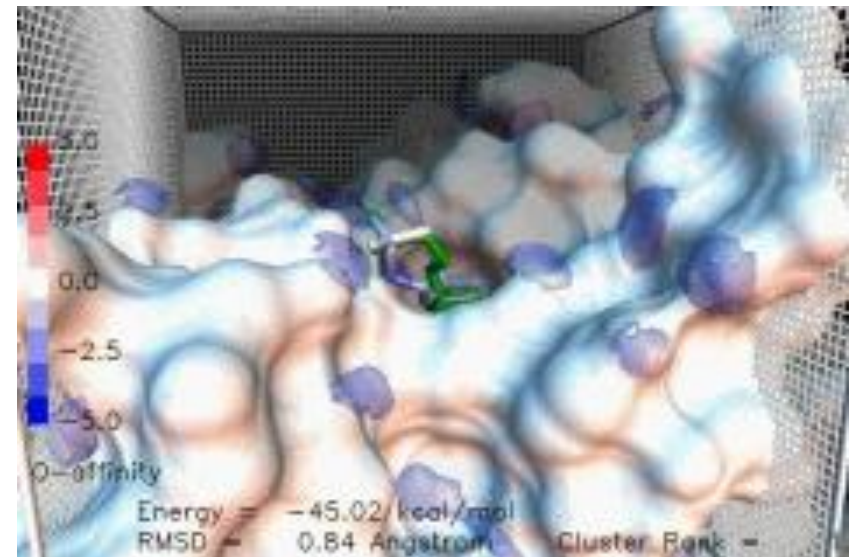
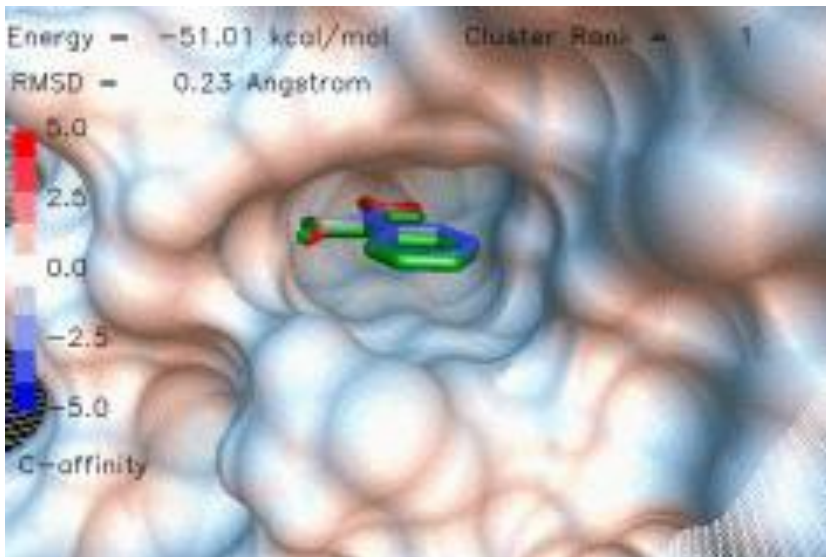
□ Molekulární dokování

- **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





□ 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





□ 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 S	Zhang-Server	164	129.070
2.	438 S	RAPTOR	164	103.274
3.	425 S	BAKER-ROBETTA	164	101.546
4.	409 S	pro-sp3-TASSER	164	99.062
5.	182 S	METATASSER	164	96.045
6.	322 S	Phyre_de_novo	164	91.897
7.	012 S	HHpred5	164	89.393
8.	020 S	MULTICOM-CLUSTER	164	87.470
9.	122 S	HHpred4	164	86.429
10.	256 S	SAM-T08-server	164	84.731
11.	013 S	MULTICOM-REFINE	164	83.832
12.	443 S	MUProt	164	81.140
13.	279 S	GS-KudlatyPred	161	78.994
14.	154 S	HHpred2	164	77.955

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
- ❑ **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
- ❑ **Rosetta@home**: <http://boinc.bakerlab.org/rosetta/>
- ❑ **CASP**: <http://predictioncenter.org/index.cgi>