

Služby pro predikci struktury proteinů

Josef Pihera

Struktura proteinů

- ▶ Primární – sekvence aminokyselin
- ▶ Sekundární – stáčení a spojování vodíkovými vazbami
- ▶ Supersekundární struktura – přechod, opakovaná geometrická uspořádání
- ▶ Terciární – Skládání proteinů
- ▶ Kvaternární – kompozice více proteinů

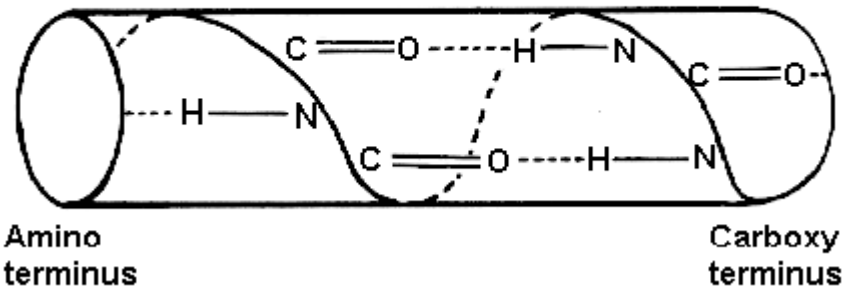
- ▶ Formování struktury záleží na rotaci molekul kolem uhlíků C-alpha a C-beta v peptidech
- ▶ Rozdělení na Alfa a Beta formace (kladné/záporné úhly)



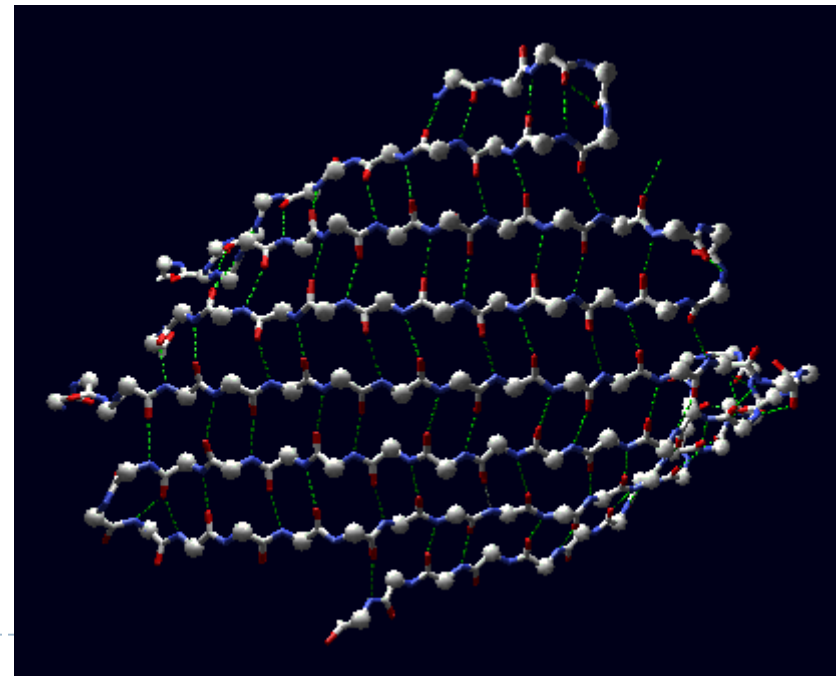
Sekundární struktura

- ▶ Alpha-helix
- ▶ Beta-sheet (Beta-skládaný list)

Toilet roll representation of the main chain hydrogen bonding in an alpha-helix.

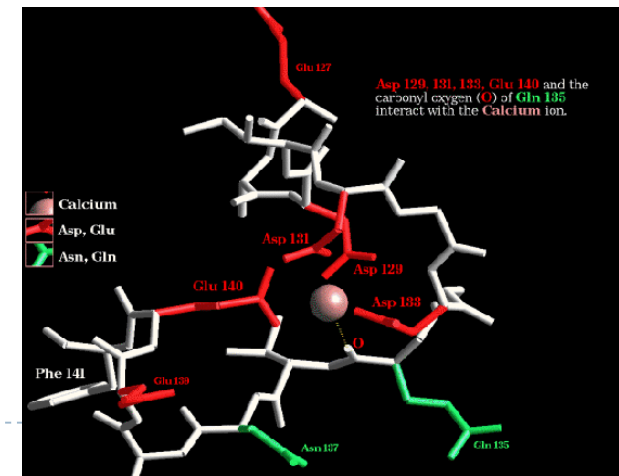
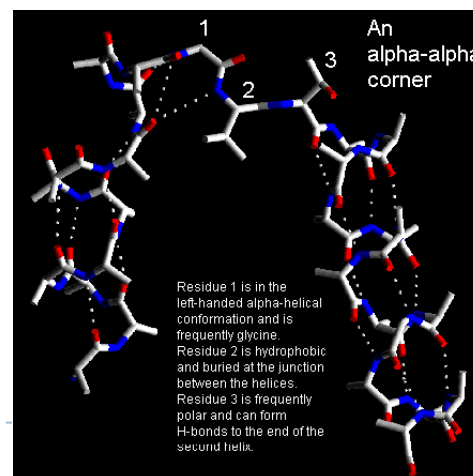
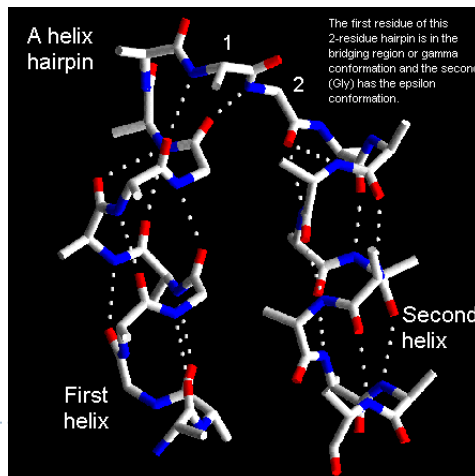
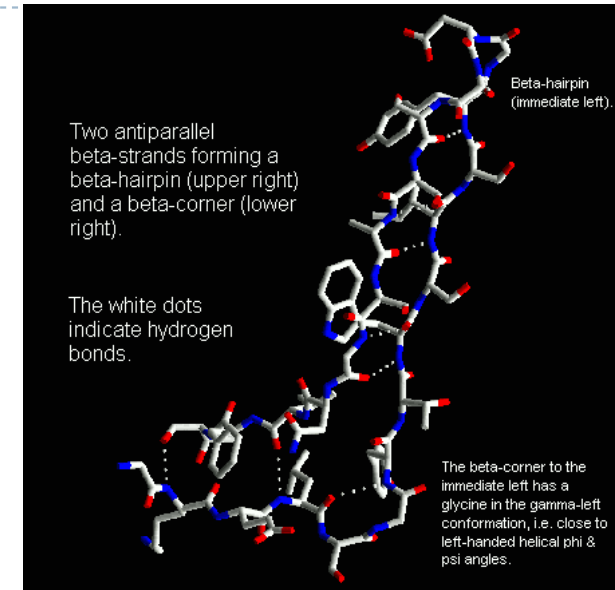
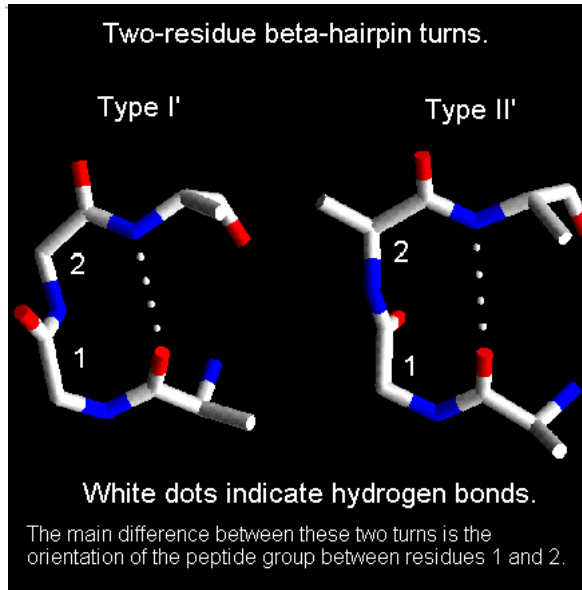


3-vláknový antiparalelní b-sheet



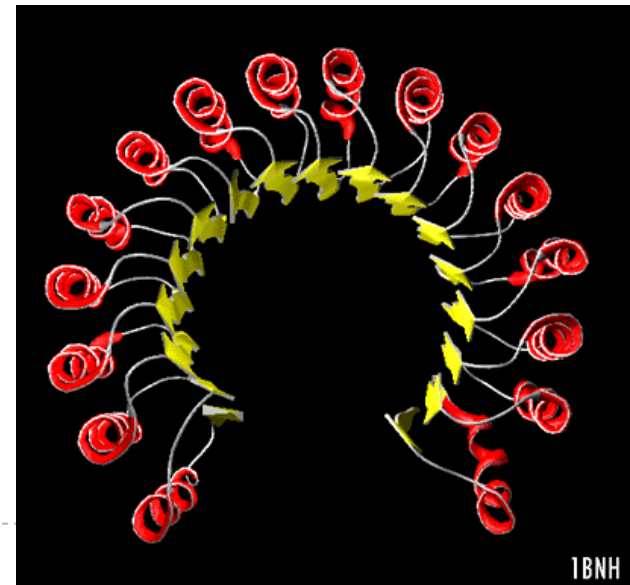
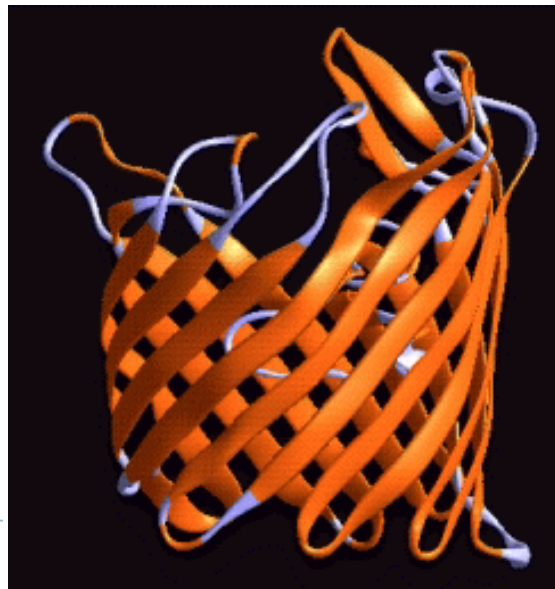
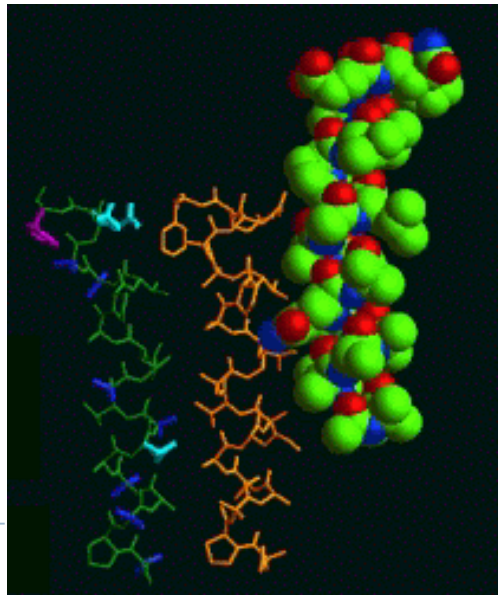
Supersekundární struktura

- ▶ Beta-hairpin
- ▶ Beta-corners
- ▶ Helix-hairpin
- ▶ A-a corner
- ▶ Helix-turn-helix



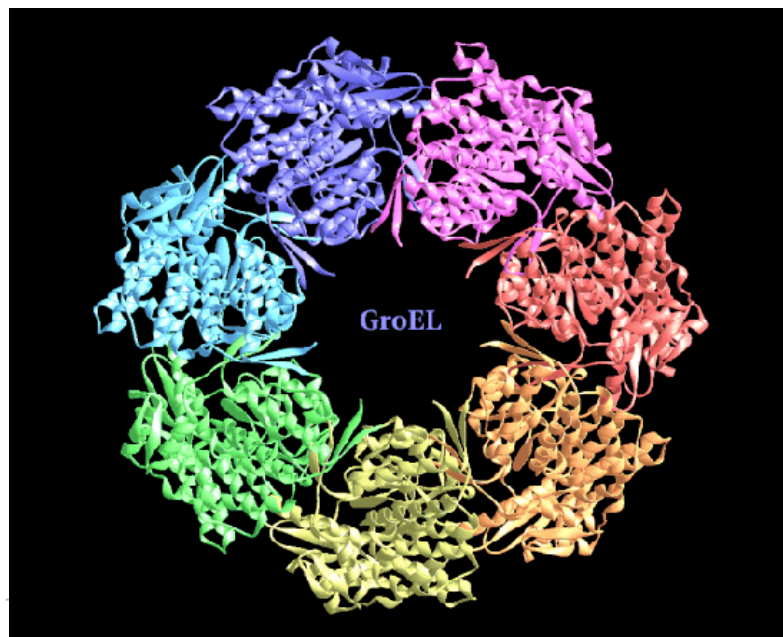
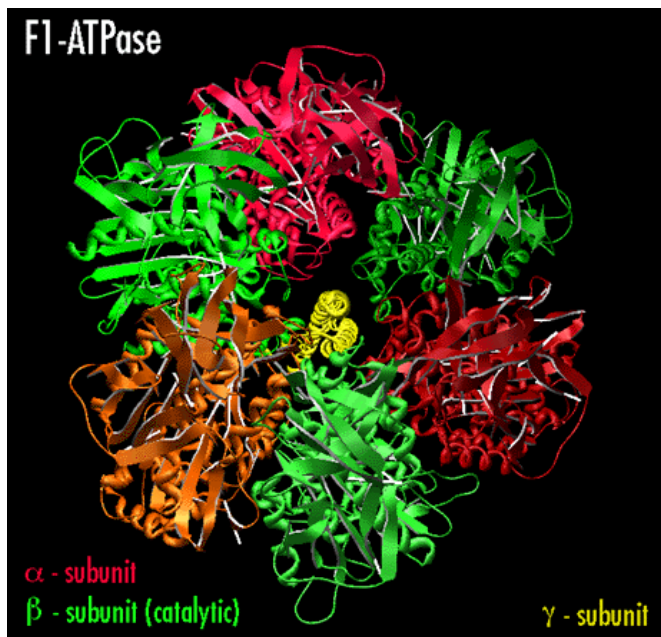
Terciární struktura

- ▶ Alfa topologie (cytokin, transkripční faktory, globiny)
- ▶ Beta topologie – sendviče a sudy
- ▶ Alfa-Beta topologie

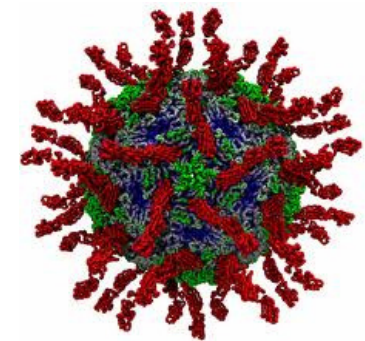


Kvarternární struktury

- ▶ Hetero-multimery
- ▶ Homo-multimery
- ▶ Větší struktury (viry, mikrotubuly, histony...)



imaging.beckman.illinois.edu
Virus obrny



Význam znalosti struktury proteinů

- ▶ Objevování a výzkum léčiv
- ▶ Experimentální toxikologie
- ▶ Predikce funkce proteinů
- ▶ Zkoumání genetického driftu
- ▶ Exprese genu



Metody určení struktury

▶ Sekundární struktura

- ▶ Bayesovské modely
- ▶ Strojové učení – SVM, neuronové sítě

▶ Terciární struktura

- ▶ Ab Initio – výpočetně náročné, používání tzv. kontaktních map
- ▶ Komparativní metody
 - ▶ Homologie – „podobné geny mají podobnou strukturu“
 - ▶ Folding recognition – v přírodě se vyskytuje cca 1300 základních šablon, zbytek jsou modifikace
 - minimalizace skórovací funkce, hledání nejvhodnější šablony
 - ▶ Využívání existujících databází (např. PFAM, SMART, CDD, COG, KOG, nebo PDB, SCOP)



Dostupné služby

- ▶ Placené / **Zadarmo** – těmi se budeme zabývat
- ▶ Stažitelný software
 - ▶ Např:
 - ▶ MODELLER
- ▶ Webové servery
 - ▶ Různé metody, často 1 server agreguje více funkcí
 - ▶ Např:
 - ▶ SWISS-MODEL
 - ▶ PSI-PRED
 - ▶ I-TASSER



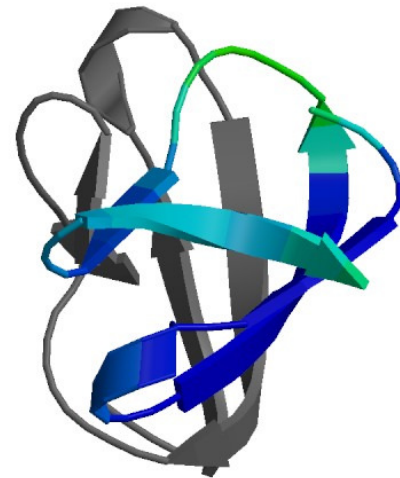
Srovnání služeb

- ▶ **CASP - Critical Assessment of Techniques for Protein Structure Prediction**
 - ▶ <http://predictioncenter.org/>
 - ▶ Populární automatizovaný benchmark
 - ▶ Vývojáři se sami přihlašují do benchmarku
 - ▶ Použití GDT_TS a GDT_HA metriky
- ▶ **LiveBench**
 - ▶ <http://bioinfo.pl>
 - ▶ vyhodnocení na nejnovějších vzorcích z PDB
 - ▶ hlavně srovnání predikce sekundární struktury
- ▶ **Ne vše se na webu tváří aktivně**
 - ▶ CAFASP (x CAFASP3), EVA



SWISS-MODEL

- ▶ <http://swissmodel.expasy.org>
- ▶ Podporuje plně automatický mód
- ▶ Homology modelling
- ▶ QMEAN4 – lin. Kombinace 4 ukazatelů (0-1)
- ▶ Benkert P, Biasini M, Schwede T. (2011).
"Toward the estimation of the absolute quality of individual protein structure models."
Bioinformatics, 27(3):343-50.



Solution structure of the Zn(II) form of Desulfiredoxin
QMEAN4 score: 0.857



SWISS-MODEL

[SIB](#) | The Center for Molecular Life Sciences | [Modelling](#) | [Tools](#) | [Repository](#) | [Documentation](#)

[myWorkspace] [login]

Workunit: P000001 P00273 - Overview

1 37

Print/Save this page as

Model Summary ?

<p>Image</p> <p>Model information: Modelled residue range: 2 to 37 Based on template: [1dxgA]* (1.80 Å) Sequence Identity [%]: 100 Evaluate: 1.23e-7</p> <p>Quality information: [details]* QMEAN Z-Score: -0.01 </p>	<p>Quaternary structure information: [details]* Template (1dxg): DIMER Model: DIMER</p> <p>Ligand information: [details]* Ligands in the template: FE: 2. Ligands in the model: FE: 2</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

[logs: \[Templates\]*](#) [\[Alignment\]*](#) [\[Modelling\]*](#)
[display model: as \[pdb\]*](#) - as [\[DeepView project\]*](#) - in [\[AstexViewer\]*](#)
[download model: as \[pdb\]*](#) - as [\[Deepview project\]*](#) - as [\[text\]*](#)

Global Model Quality Estimation ? [\[+/-\]](#)

QMEAN4 global scores: ?			Local scores	
QMEANscore4 ?	Estimated absolute model quality ?	Score components ?	Coloring by residue error ?	Residue error plot ?
0.86	 Z-Score: -0.01 Plot 1: [save png]* Plot 2: [save png]*	 [save png]*	 [save jpg]* Chain: ← A → Z-score = 1.10	 [save png]*
			Coloring (all chains): [save jpg]* [save pdb]*	Energy profile: ? [save raw scores]*

QMEAN4 global scores: ?

PSIPRED

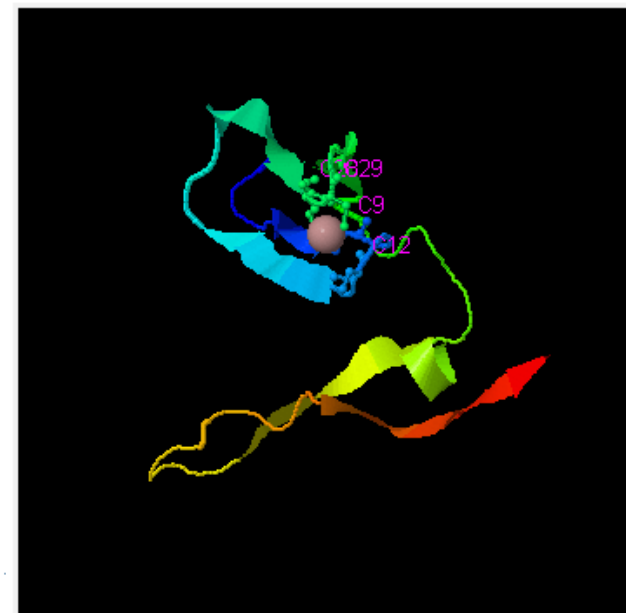
- ▶ <http://bioinf.cs.ucl.ac.uk/psipred/>
- ▶ Různé metody
 - ▶ Pro terciární struktury
GenTHREADER
 - ▶ Fold Recognition
- ▶ Javovský applet,
který vždy zasekl prohlížeč...



I-TASSER

- ▶ <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>
- ▶ Vysoce hodnocený v CASP pro terciární struktury
- ▶ Založen na Fold Recognition a Ab Initio
- ▶ Asi nejucelenější a nejinformativnější pro méně znalého uživatele
- ▶ Zpracování trvá několik hodin

Predicted Binding Site



I-TASSER

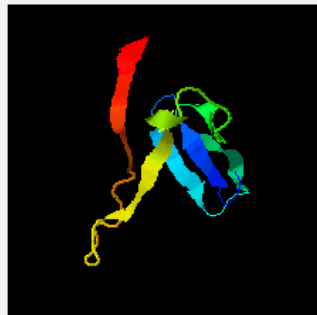
Predicted Secondary Structure

	20	40	60
Sequence	ANEGDVYKCELCGQVVKVLEEGGGTLVCCGEDMVKQANEGDVYKCELCGQVVKVLEEGGGTLVCCGEDMVKQ		
Prediction	CCCCSSSSCHCCSSSSSCCCCCSSCCCCSSCCCCCHHCCCCSSSSSSCCSSSSSCCCCCC		
Conf.Score	975408983016829999988974788879966950157015422578738998867998999959831749		

Predicted Solvent Accessibility

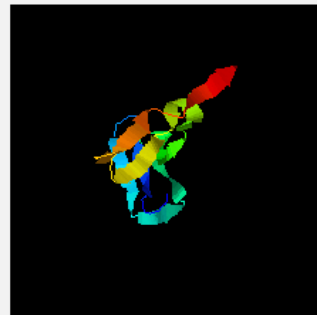
	20	40	60
Sequence	ANEGDVYKCELCGQVVKVLEEGGGTLVCCGEDMVKQANEGDVYKCELCGQVVKVLEEGGGTLVCCGEDMVKQ		
Prediction	876442230531220120146472421014461524564444244641431241147762220102351668		
	Values range from 0 (buried residue) to 9 (highly exposed residue)		

Top 5 Models predicted by I-TASSER



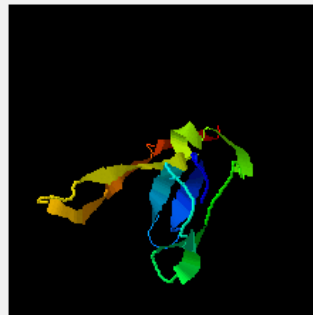
[Download Model 1](#)

C-score=0.14



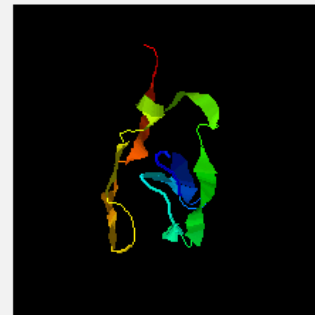
[Download Model 2](#)

C-score=-1.68



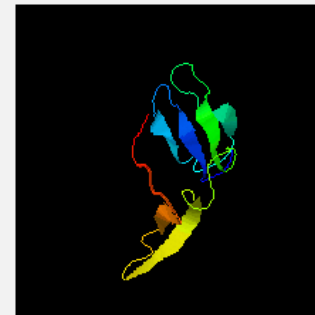
[Download Model 3](#)

C-score=-2.68



[Download Model 4](#)

C-score=-2.78



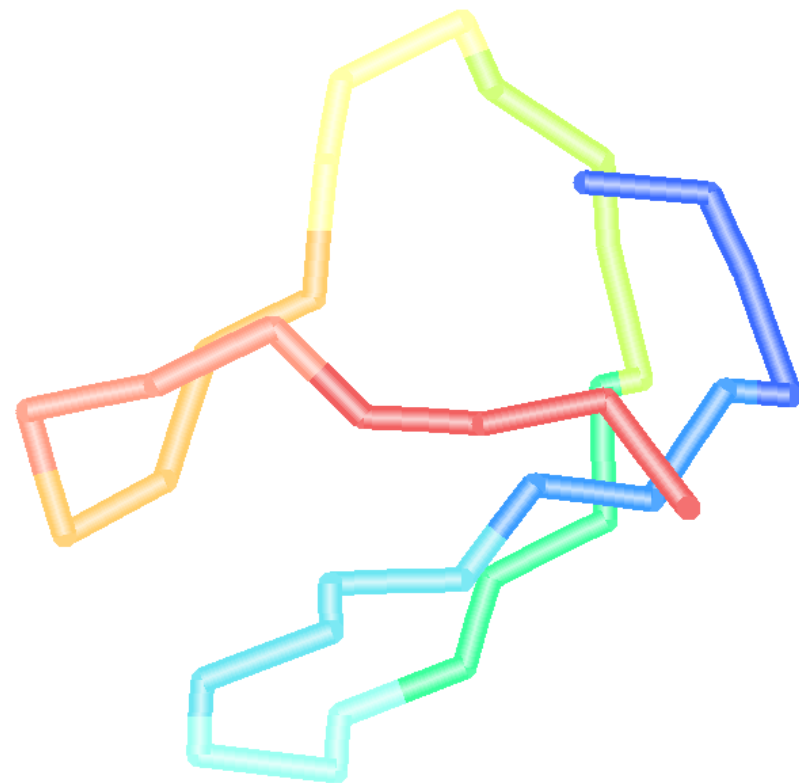
[Download Model 5](#)

C-score=-1.34

HHPRED

- ▶ <http://toolkit.tuebingen.mpg.de/hhpred>
- ▶ Od Max-Planck Institute for Developmental Biology
- ▶ Homology
- ▶ Založeno na HMM

Šablona 3D modelu



ROBETTA

- ▶ <http://robetta.bakerlab.org/>
- ▶ Vysoce hodnocený v CASP
- ▶ Meta-server – využívá výsledků různých ostatních serverů
- ▶ Momentální beta-verze bohužel při ani jednom z pokusů nevydala výsledek
- ▶ Odmítání více proteinového vstupu
- ▶ Homology a Ab Initio



ROBETTA

- ▶ ...ale protože výsledky jiných uživatelů lze zobrazit...
- ▶ Ukázka výstupu pro lidský Herpes virus 5

Features and Secondary Structure

1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160

GNNRFVHSGPERLLLSQLPVERQALTELEYQAMGAVVRAAFLANSTGRAMRKYSGRDAGTLLPLGSRPYGYARVTPRSQMNGVATDLRQLSPRDANIIVLVATVWHEVDPAADPTVGRKAGHPELCAQDGLYLALGAGFRVYDLANNTLLILAAADADEN

tmhmm (0)

low complexity (0%)

coiled-coils (0%)

disordered (2%) xxx

psipred

-----HH-----EE-----HHHHHHHHHHHHHH-----HHHHHHHHHH-----EE-----EEEEE-----HH-----EEEEEE-----EEEEE-----EEEEE-----HH-----

Domain Repeats Prediction ▲

Boundary	STD (+/-)	Consensus
--	--	--

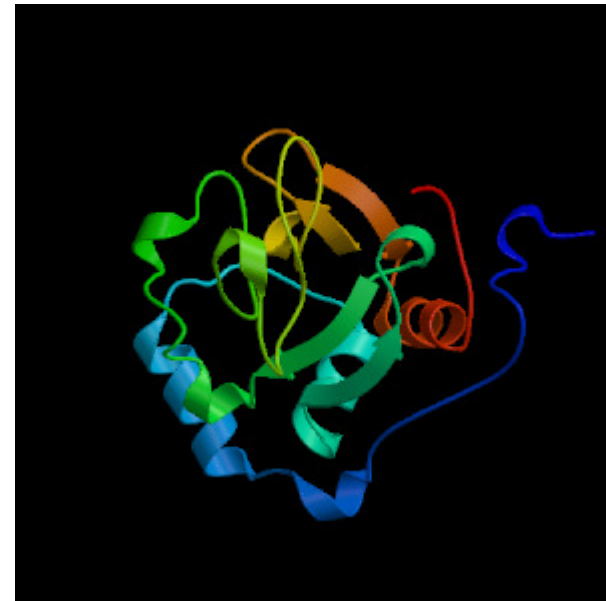
Ginzu Domain Prediction 1 (RECOMMENDED: click on domain number for details: alignments, model energies, and MAMMOTH detections) ▲

Domain	Span	Source	Reference Parent	Parent Span	Confidence	Annotations
domain 1	1-175	hhsearch	2papA	10-135	1.720471	STRUCTURAL GENOMICS, UNKNOWN FUNCTION

0 50 100 150

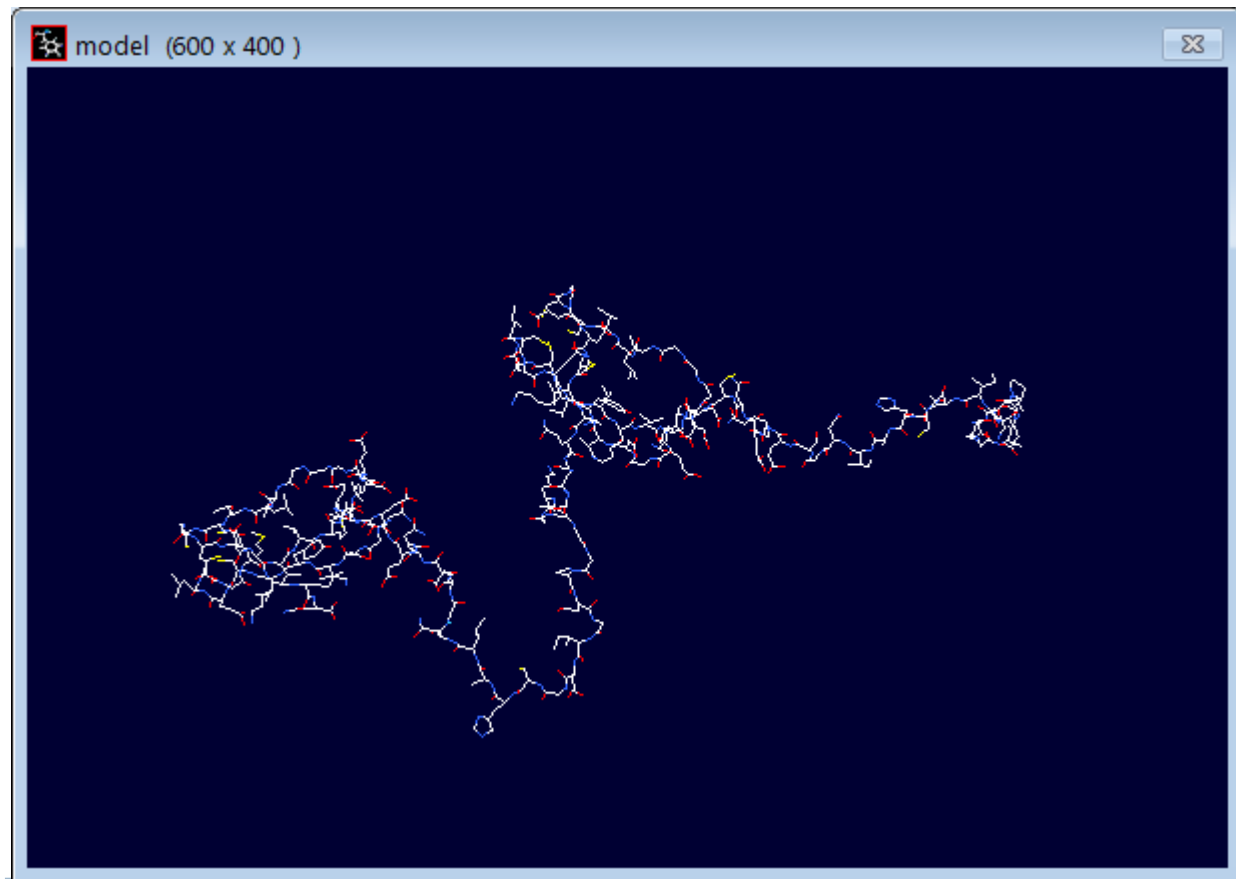
hhsearch

Top 250 of 250 HR PSI-Blast alignments.



MULTICOM

- ▶ http://cas.p.rnet.missouri.edu/multicom_3d.html
- ▶ Výstup v řádu minut
- ▶ Pouze na mail
- ▶ Formát PDB, lze zobrazit dostupnými nástroji (Swiss-PDB Viewer)
- ▶ Fold-Recognition



Hodnocení vybraných služeb

- ▶ Všechny služby byly zdarma (pro akademické použití), často ale vyžadovaly bezplatnou registraci
- ▶ Doba zpracování od pár sekund po několik hodin
- ▶ Množství rozličných metod, CASP pak dobře pomáhá orientaci
- ▶ Ne všechny výstupy přehledné
- ▶ Asi nejpoužitelnější je I-TASSER, který je i doporučován v CASP



Číselné srovnání některých služeb

► Převzato z

http://predictioncenter.org/casp9/groups_analysis.cgi?type=server&tbm=on&submit=Filter

Služba	GDT_TS
MULTICOMM	68.99
I-TASSER	86.737
HHPredA	87.57



Zdroje

- www.wikipedia.org
 - <http://swissmodel.expasy.org/>
 - <http://toolkit.tuebingen.mpg.de/>
 - <http://robeta.bakerlab.org/>
 - <http://predictioncenter.org/>
 - <http://bioinfo.pl>
 - <http://robeta.bakerlab.org/>
 - <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>
 - <http://bioinf.cs.ucl.ac.uk/psipred/>
 - http://casp.rnet.missouri.edu/multicom_3d.html
 - <http://www.rcsb.org/pdb/explore/explore.do?structureId=2LK5>
-



Konec a čas na dotazy

