

Blok 2

Sekundární struktura proteinů

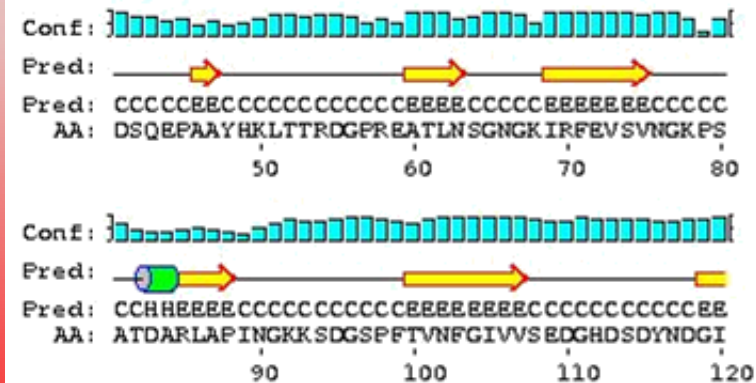
C3211 Aplikovaná bioinformatika
Přednášející: Josef Houser



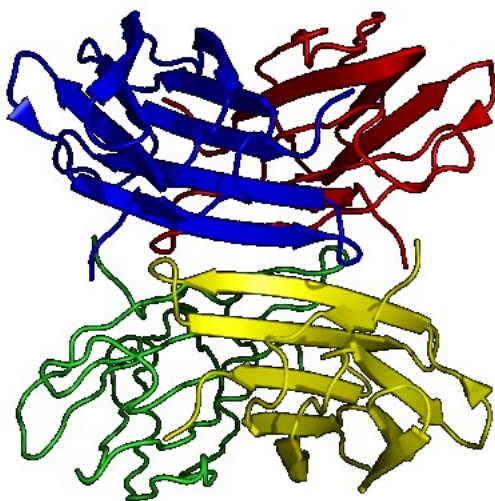
Struktura proteinů

ADSQTSSNRAGEFSIPPNTDFRAIFFANAAE
QQHIKLFIGDSQEPAAYHKLTTTRDGPREATL
NSGNGKIRFEVSVNGKPSATDARLAPINGKK
SDGSPFTVNFIVVSEDGHDSYNDGIVVL
QWPIG

primární
(sekvence)

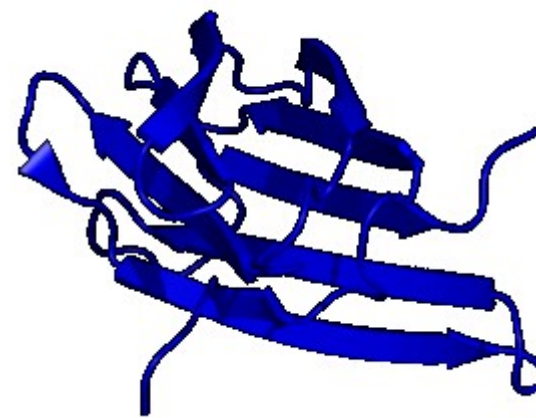


sekundární



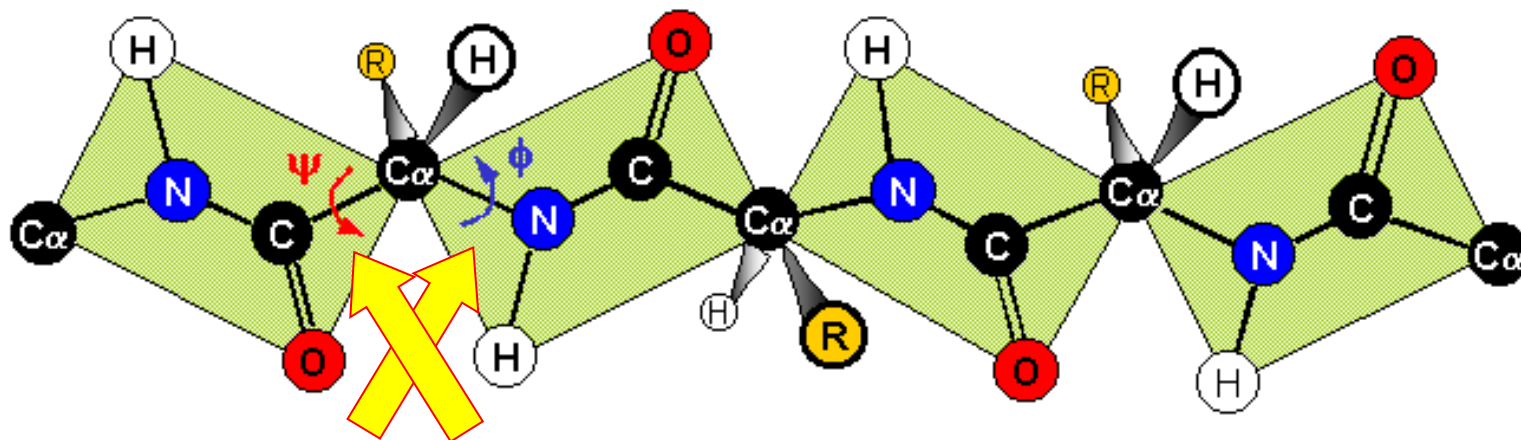
kvartérní

ter



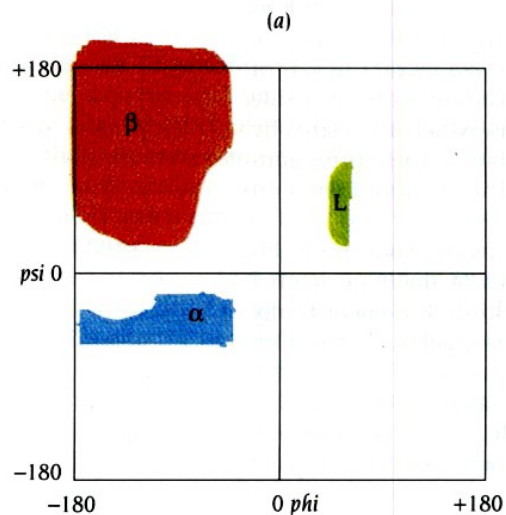
Kostra polypeptidového řetězce

Peptidová vazba – planární

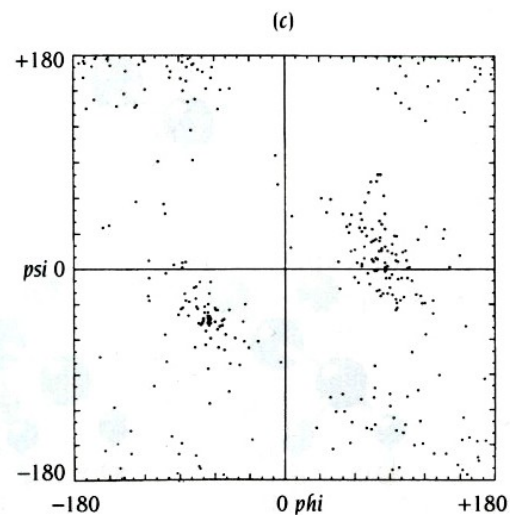
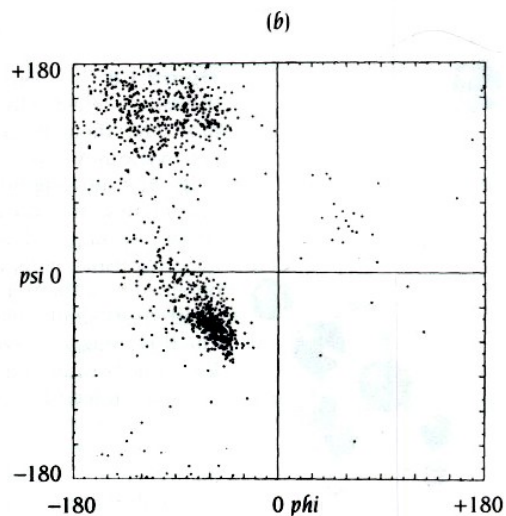


Konformaci kostry určují dva torzní úhly ϕ a ψ (úhel ω je 180)

Ramachandran plot diagram



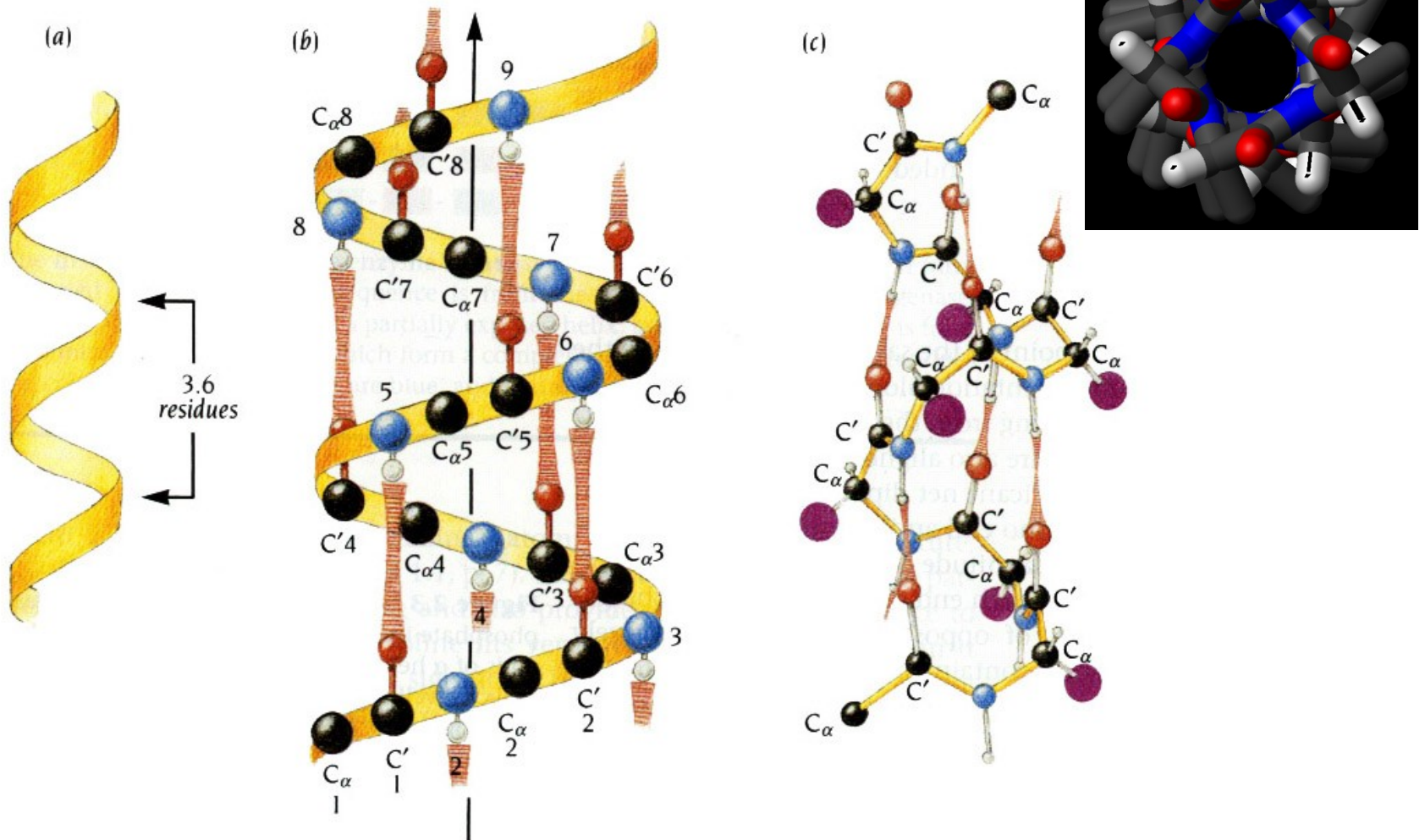
- a) Colored areas show sterically allowed combinations of the ϕ and ψ angles.
- b) Observed values for all residue types except for glycine. Each point represents ϕ and ψ values for an amino acid residue in a well-refined x-ray structure.
- c) Observed values for glycine



Vazby zprostředkovávající vyšší struktury

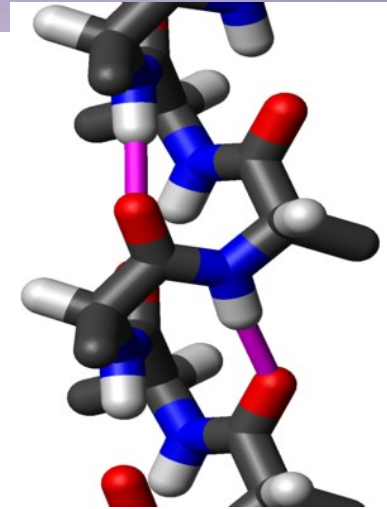
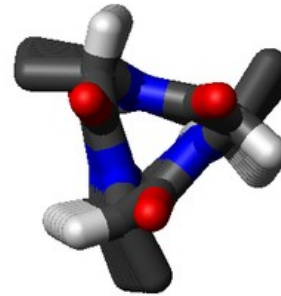
- **Vodíková vazba (H-můstek)**
- Nabité AK
- Kontakty polárních AK
- Nepochární / hydrofobní AK
- Stacking – aromatické AK
- Cystein / cystin – vazba síra-síra
- Vazba iontů kovů

α -helix

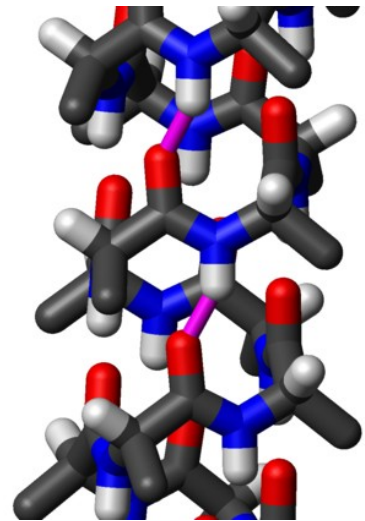
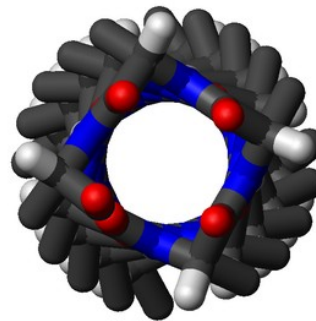


Jiné helixové struktury

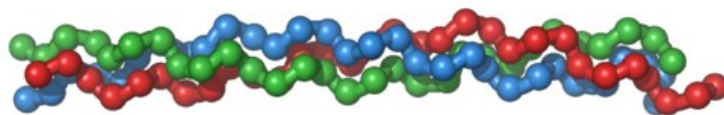
- 3_{10} helix



- π -helix



- kolagen – levotočivá šroubovice



β -list

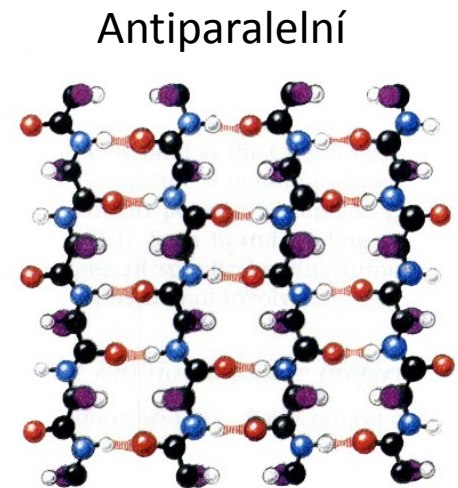
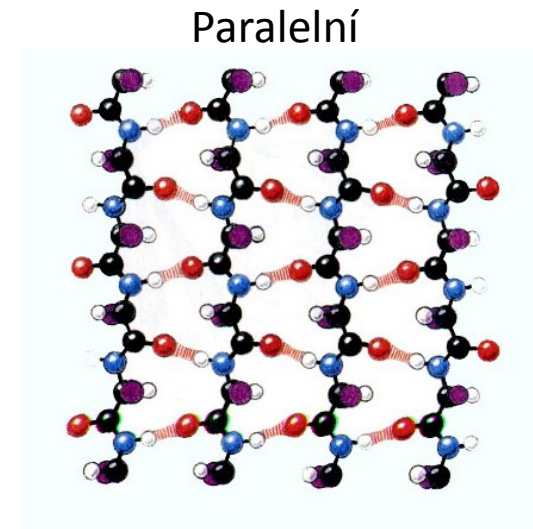
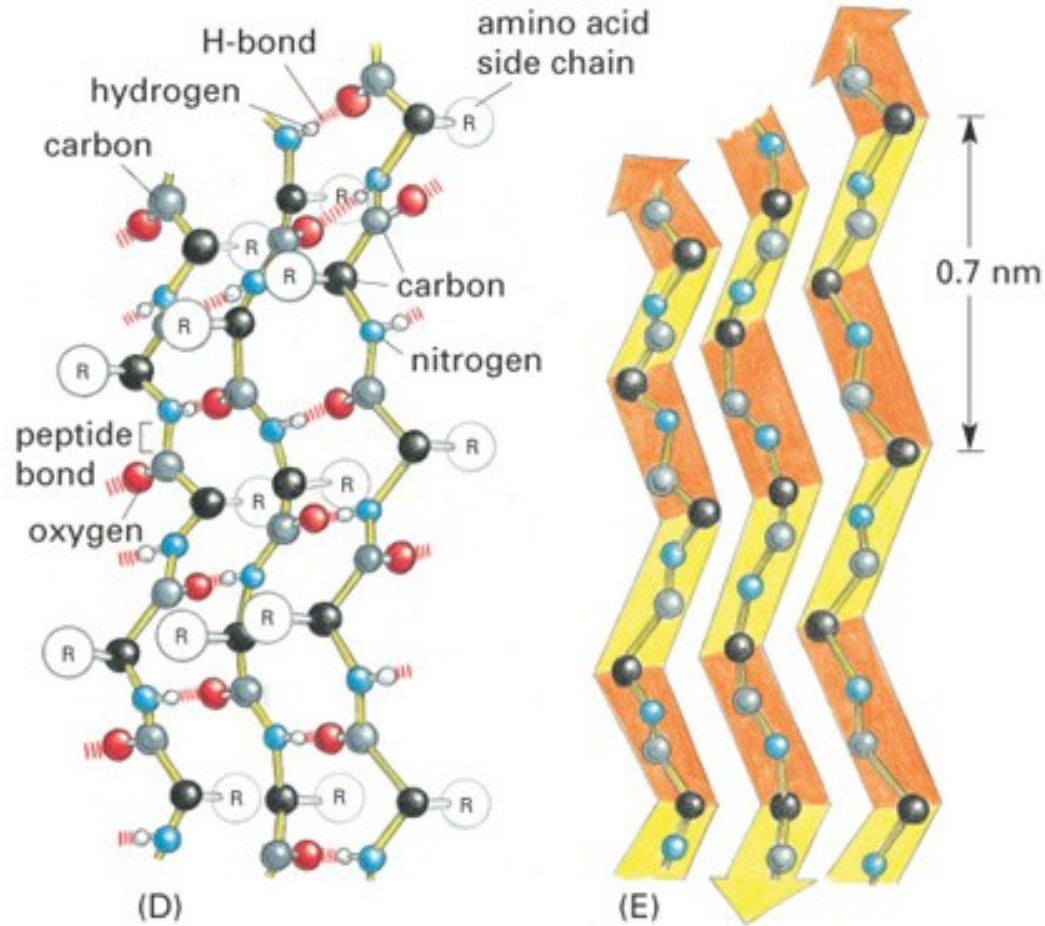
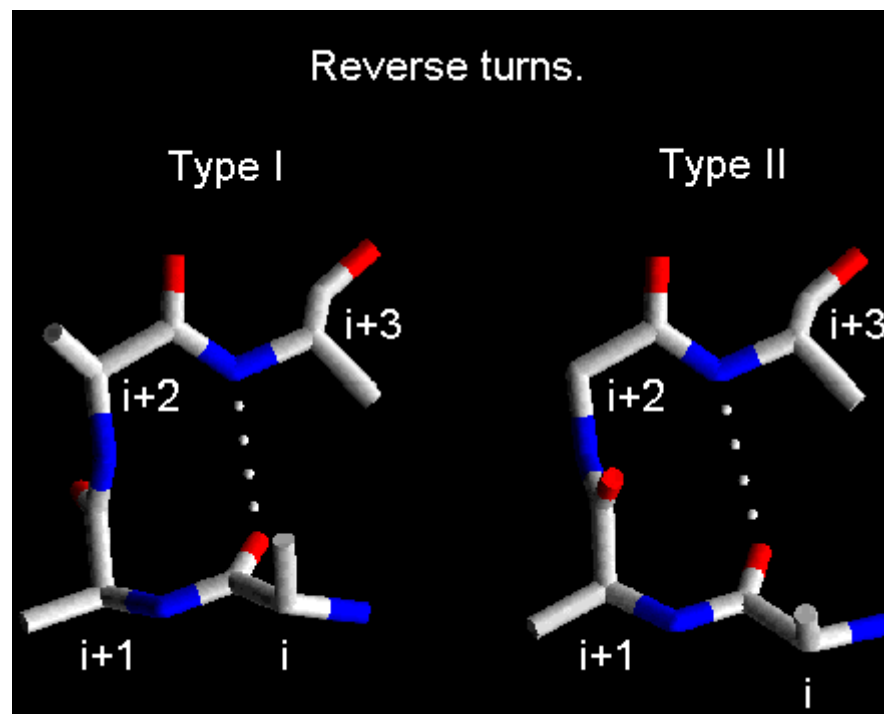


Figure 4-10 part 2 of 2 Essential Cell Biology, 2/e. (© 2004 Garland Science)

Otáčky (turns)

- více typů otáček, různé počty AK
- příklad β -turn

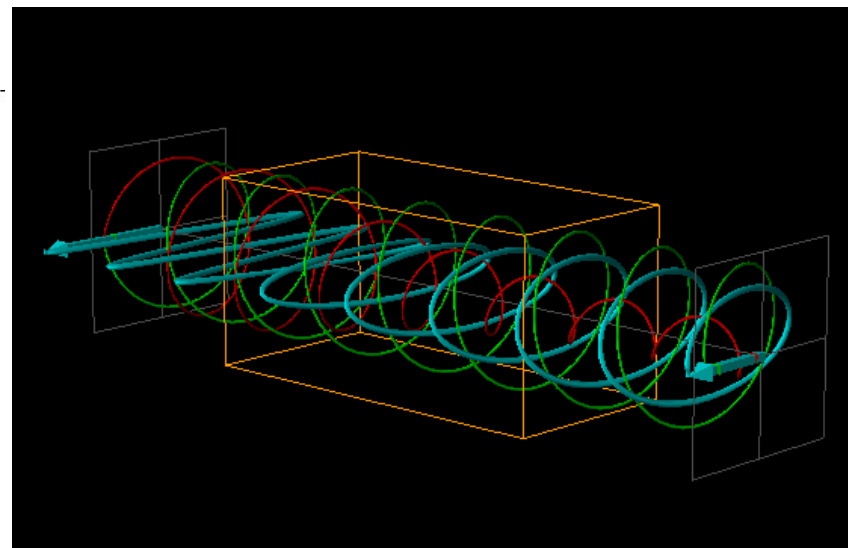


Určení 2D struktury proteinů

- **Cirkulární dichroismus (CD)**
- **Infračervená spektroskopie (IR)**
- **Nukleární magnetická resonance (NMR)**
- **Predikce *in silico***

Cirkulární dichorismus

- **Cirkulární Dichroismus (CD)** - různá absorpce levé a pravé složky cirkulárně polarizovaného světla; $\Delta\varepsilon \equiv \varepsilon_L - \varepsilon_R$
- Opticky aktivní jsou chirální molekuly
- Určení sekundární a terciární struktury biomakromolekul
- Určení stability a konformačních změn proteinů v různých podmínkách (teplota, iontová síla, pH, polarita rozpouštědla, ...)

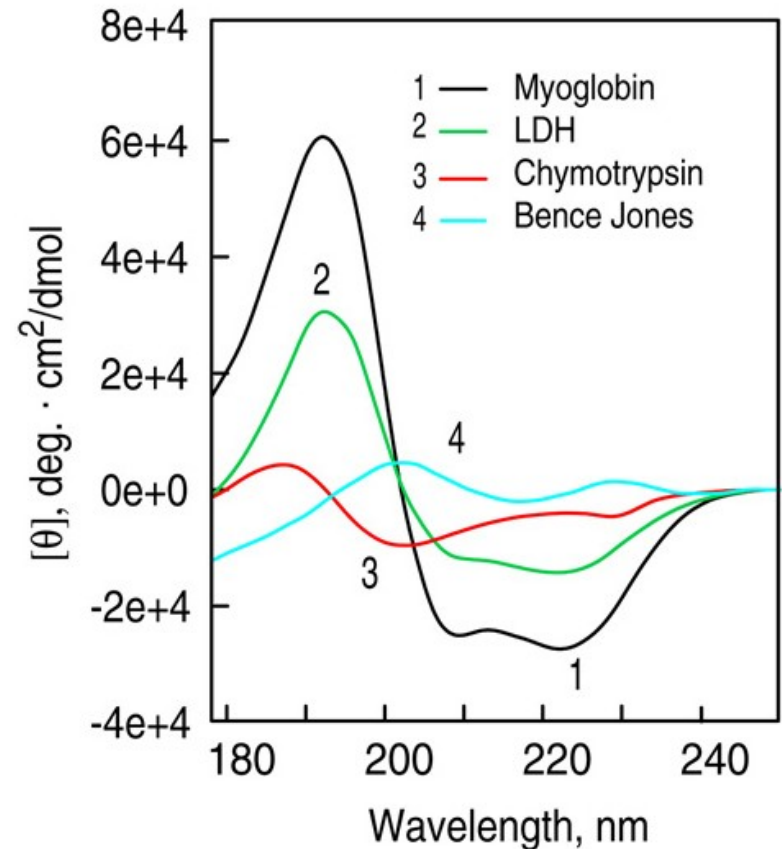
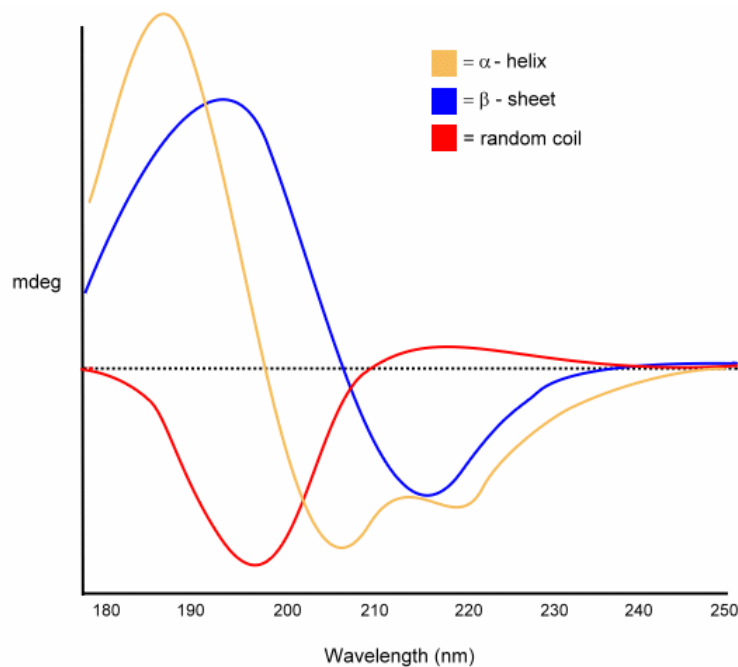


Cirkulární dichorismus - veličiny

Veličina	Definice a vztahy	Jednotky
Absorbance	$A(\nu) = \log_{10} (I_0/I)$ <p>I_0 – intenzita světla vstupujícího do kyvety, I – intenzita světla opouštějícího kyvetu</p>	bezrozměrná
Molární absorpční koeficient	$\varepsilon(\nu) = A(\nu)/cl$ <p>c – molární koncentrace (mol L^{-1}), l – délka kyvety (cm)</p>	$\text{L mol}^{-1} \text{ cm}^{-1}$
Cirkulární dichroismus	$\Delta A(\nu) = A_L(\nu) - A_R(\nu), A_L(\nu), A_R(\nu)$ <p>absorbance pro levotočivé a pravotočivé kruhově polarizované světlo</p>	bezrozměrné
Molární CD	$\Delta\varepsilon(\nu) = \Delta A(\nu)/cl$	$\text{L mol}^{-1} \text{ cm}^{-1}$
Elipticita	$\theta(\nu) = 32\,980 \Delta A(\nu)$	mdeg
Molární elipticita	$[\theta](\nu) = \theta(\nu)/(10\, c\, l)$	$\text{deg cm}^2 \text{ dmol}^{-1}$

Určení sekundární struktury pomocí CD

- Charakteristická CD spektra ve vzdálené UV oblasti pro α -helix, β -sheet a random coil



Význam určení 2D struktury

- Folding (skládání) proteinu
- Interakce proteinu s vazebným partnerem
- Sledování stability proteinu – teplotní, pH, polarita prostředí,...
- Identifikace proteinu (PCDDDB)
- Zařazení proteinu v rámci databáze SCOP
(Structural Classification of Proteins)

PCDDDB

<http://pcddb.cryst.bbk.ac.uk/home.php>

Protein Circular Dichroism Data Bank

Možnost porovnání CD spekter se záznamy v databázi



DichroMatch



Protein Circular Dichroism Data Bank

[Home](#) | [Search](#) | [Download](#) | [Deposit](#)

[Lost Password](#) | [Help](#)
[Create account](#)

Login ID

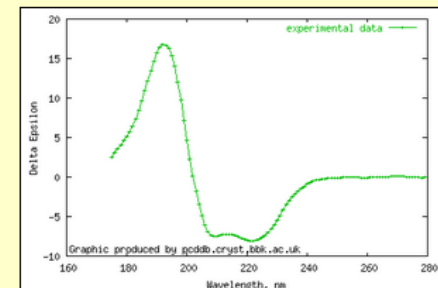
Password

Information
About
Contact
News
Related Videos
Site Map
Glossary
Terms and
Conditions
Cookies
DichroWeb
2Struc
DichroMatch
CreateRDS

Current holdings (live data):
Released entries: 233
Entries in pre-release: 330

Citing the PCDDDB :
Whitmore, L., Woollett, B., Miles, A.J., Klose, D.P., Janes, R.W. and Wallace, B.A., PCDDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata.
[Nucleic Acids Research \(2011\) 39 D480-D486](#)

CD0000047000 - Myoglobin (horse)



For feedback, please email : pcddb@mail.cryst.bbk.ac.uk.

Výpočet zastoupení sekundárních struktur

Různé SW dostupné
jako freeware

k2d – aktuálně
3. generace K2D3

<http://www.ogic.ca/projects/k2d3/>


K2D3

Estimates protein secondary structure from CD spectra

Proteins (2012) Vol 80-2

about
supplement
use K2D2
contact us

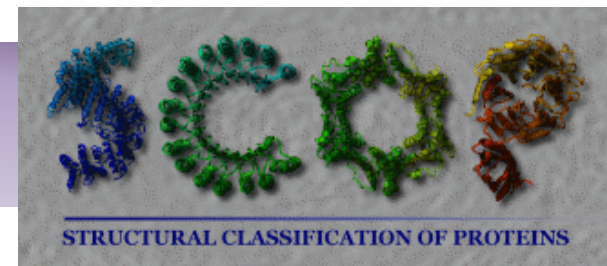
Input your CD data [help](#) [example](#)

1. **Wavelength range in nm**: from to
2. **CD units**:
3. **Protein size**: length in aa OR weight in kDa
4. **CD spectrum**:


Úloha

- Vyhodnoťte data získaná měřením cirkulárního dichroismu roztoku neznámého proteinu (*Input_CD_data.txt*). Určete, zda se ve vzorku nachází protein převážně ve stavu α -helixu, β -skládaného listu nebo náhodného klubka (coil). K vyhodnocení využijte program **k2d3** (<http://www.ogic.ca/projects/k2d3/>).

SCOP databáze



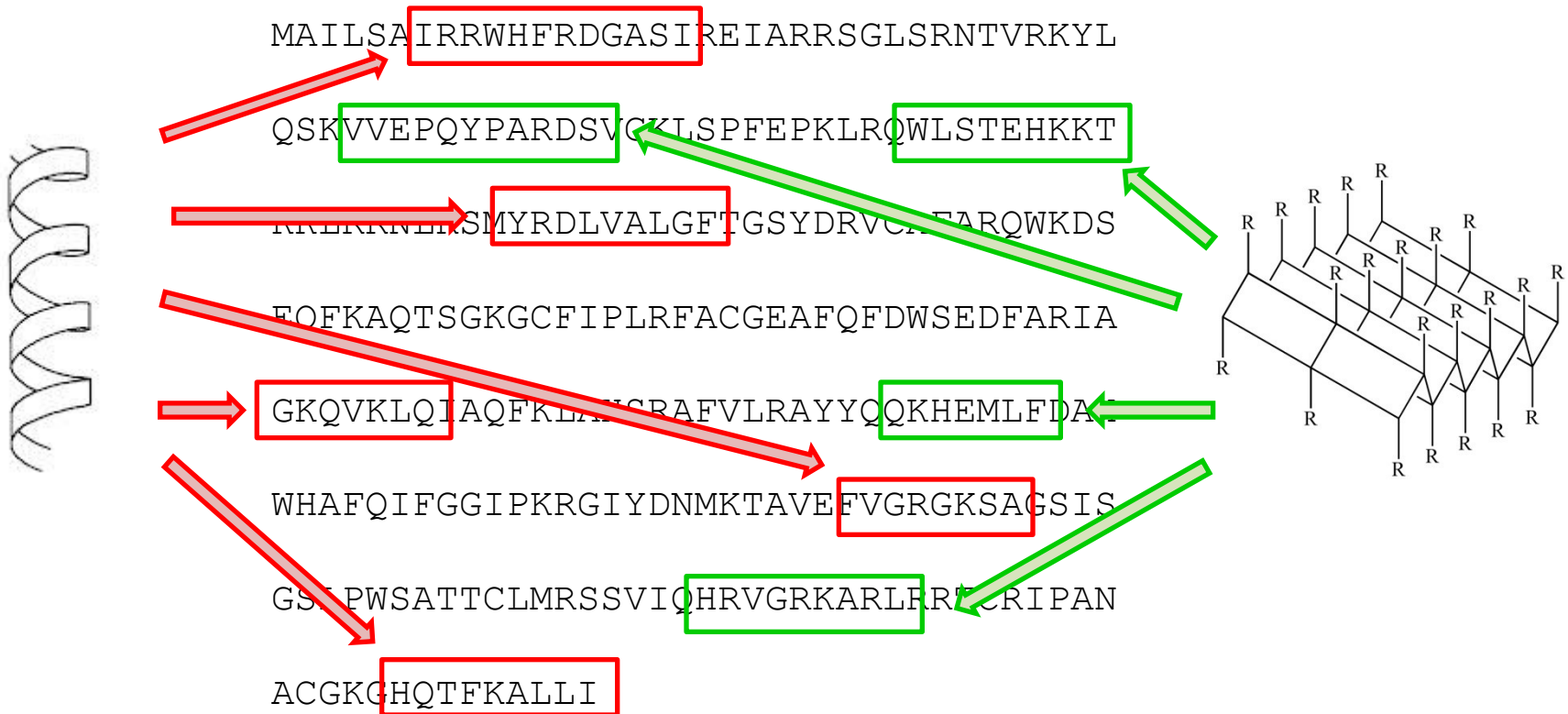
Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902

Další čtení o cirkulárním dichorismu

- Norma J. Greenfield **Using circular dichroism spectra to estimate protein secondary structure** Nat Protoc. 2006; 1(6): 2876–2890. doi: [10.1038/nprot.2006.202](https://doi.org/10.1038/nprot.2006.202)
- <http://www.niu.edu/analyticallab/cd/handout.pdf>

Predikce 2D struktury

Cíl: Určit umístění 2D elementů v rámci konkrétní sekvence *in silico*



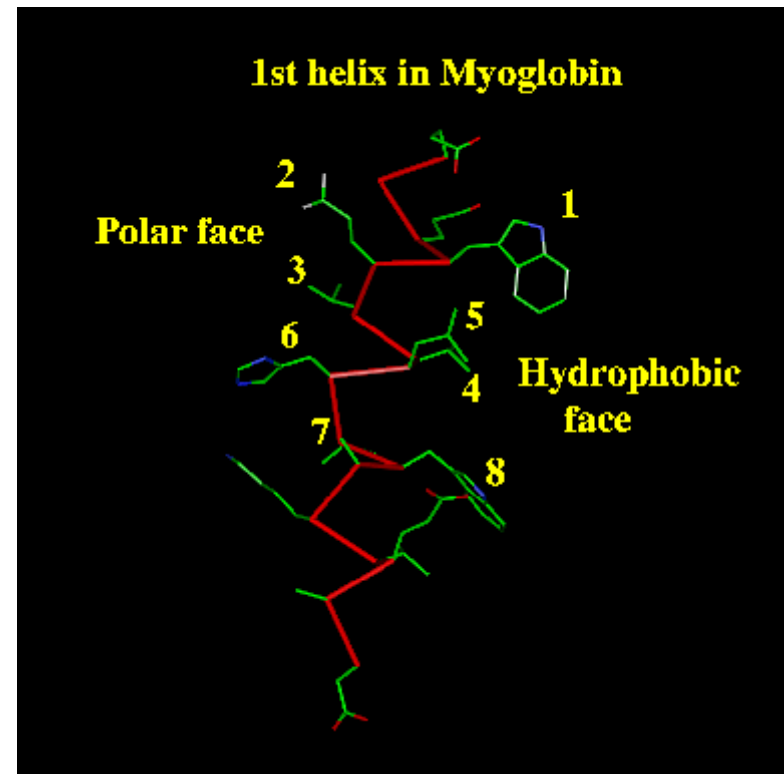
Predikce 2D struktury

- Rozlišujeme tři základní typy 2D struktury
 - H – helix
 - E – β -list
 - C/(-) – smyčka/náhodné klubko (coil) – někdy jsou rozlišovány tyto dvě varianty
- S dobrou přesností lze určit helix, u β -listu úspěšnost určení 2D struktury klesá
- Některé programy přidávají i číslo vyjadřující pravděpodobnost pro daný AK zbytek (např. H 60% - znamená, že s 60% pravděpodobností se jedná o helix)

Typické znaky α -helix

Často je helix částečně exponovaný – tj. jedna strana je otočena dovnitř proteinu (hydrofobní), druhá ven (hydrofilní)

Potom pro 3.6 helix (α -helix) platí, že i , $i+3$, $i+4$ & $i+7$ -té reziduum míří na tutéž stranu. Jsou-li všechna hydrofobní či naopak hydrofilní = zřejmě α -helix



Typické znaky β -list

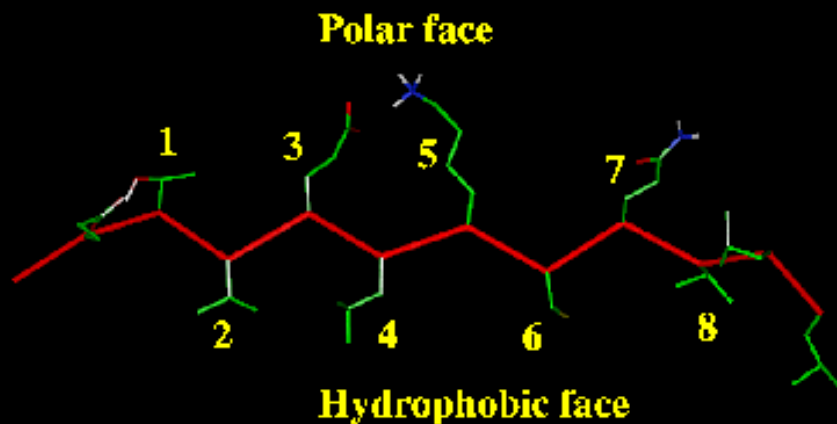
U β -listu se střídají rezidua po 180°

Pro částečně zanořený β -list platí:

- $i, i+2, i+4, i+8$ –té reziduum je polární a zároveň
- $i+1, i+3, i+5$ –té je nepolární

Zcela zanořený β -list (typicky u α/β barelu) je tvořen nepolárními AK

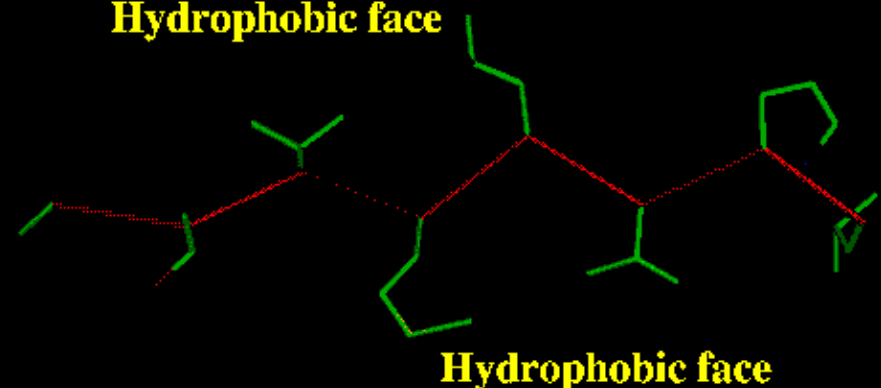
Second strand in CD8



Chemotaxis protein CheY

Residues 81–88

Hydrophobic face



Úloha

- Na základě základních znalostí o uspořádání sekundárních struktur odhadněte **bez použití predikčních programů**, zda následující krátké peptidy jsou součástí α -helixu, β -skládaného listu nebo náhodného klubka. Můžete využít přehlednou tabulku aminokyselin.

Úloha

Peptid 1

KALSQVIRL

Peptid 2

GGASPAPPK

Peptid 3

LELQIRVYA

	NONPOLAR, HYDROPHOBIC	R GROUPS	POLAR, UNCHARGED	
Alanine Ala A MW = 89	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_3 \end{array}$		$\begin{array}{c} \text{H} - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Glycine Gly G MW = 75
Valine Val V MW = 117	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH} \begin{array}{l} / \text{CH}_3 \\ \backslash \text{CH}_3 \end{array} \end{array}$		$\begin{array}{c} \text{HO} - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Serine Ser S MW = 105
Leucine Leu L MW = 131	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH} \begin{array}{l} / \text{CH}_3 \\ \backslash \text{CH}_3 \end{array} \end{array}$		$\begin{array}{c} \text{OH} \\ \\ \text{CH}_3 - \text{CH} - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Threonine Thr T MW = 119
Isoleucine Ile I MW = 131	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH} \begin{array}{l} / \text{CH}_3 \\ \backslash \text{CH}_2 - \text{CH}_3 \end{array} \end{array}$		$\begin{array}{c} \text{HS} - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Cysteine Cys C MW = 121
Phenylalanine Phe F MW = 131	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}_6\text{H}_5 \end{array}$		$\begin{array}{c} \text{HO} - \text{C}_6\text{H}_4 - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Tyrosine Tyr Y MW = 181
Tryptophan Trp W MW = 204	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}_8\text{H}_6\text{N}_2 \end{array}$		$\begin{array}{c} \text{NH}_2 \\ \\ \text{C} = \text{O} - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Asparagine Asn N MW = 132
Methionine Met M MW = 149	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{S} - \text{CH}_3 \end{array}$		$\begin{array}{c} \text{NH}_2 \\ \\ \text{C} = \text{O} - \text{CH}_2 - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Glutamine Gln Q MW = 146
Proline Pro P MW = 115	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{CH} - \text{CH}_2 \\ \quad \backslash \\ \text{HN} \quad \text{CH}_2 \end{array}$		POLAR BASIC $\begin{array}{c} ^+ \text{NH}_3 - \text{CH}_2 - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Lysine Lys K MW = 146
Aspartic acid Asp D MW = 133	POLAR ACIDIC $\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}(=\text{O})\text{O}^- \end{array}$		$\begin{array}{c} \text{NH}_2 \\ \\ \text{N}^+ \text{H}_2 = \text{C} - \text{NH} - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Arginine Arg R MW = 174
Glutamine acid Glu E MW = 147	$\begin{array}{c} ^- \text{OOC} \\ \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{CH}_2 - \text{C}(=\text{O})\text{O}^- \end{array}$		$\begin{array}{c} \text{HN}^+ \text{NH} \\ \\ \text{C} - \text{CH}_2 - \text{CH} - \text{COO}^- \\ \\ \text{N}^+ \text{H}_3 \end{array}$	Histidine His H MW = 155

Predikční algoritmy

1. generace: vycházela z fyzikálně-chemických vlastností a ze statistiky pro jednotlivá rezidua
2. generace: zahrnovala i vliv nejbližších AK na zkoumané reziduum – předpověď max. 60% správnost, u β -listu do 40%
3. generace: zahrnuje navíc multiple sequence alignment a využívá skutečnosti, že 2D struktura se zachovává déle než sekvenční podobnost – až 80% spolehlivost (závisí na metodě)

Programové balíky

- [AGADIR](#) - An algorithm to predict the helical content of peptides
- [APSSP](#) - Advanced Protein Secondary Structure Prediction Server
- [GOR](#) - Garnier et al, 1996
- [HNN](#) - Hierarchical Neural Network method (Guermeur, 1997)
- [HTMSRAP](#) - Helical TransMembrane Segment Rotational Angle Prediction
- [Jpred](#) - A consensus method for protein secondary structure prediction at University of Dundee
- [JUFO](#) - Protein secondary structure prediction from sequence (neural network)
- [nnPredict](#) - University of California at San Francisco (UCSF)
- [Porter](#) - University College Dublin
- [PredictProtein](#) - PHDsec, PHDacc, PHDhtm, PHDtopology, PHDthreeder, MaxHom, EvalSec from Columbia University
- [Prof](#) - Cascaded Multiple Classifiers for Secondary Structure Prediction
- [PSA](#) - BioMolecular Engineering Research Center (BMERC) / Boston
- [PSIpred](#) - Various protein structure prediction methods at Brunel University
- [SOPMA](#) - Geourjon and Deléage, 1995
- [SSpro](#) - Secondary structure prediction using bidirectional recurrent neural networks at University of California
- [DLP-SVM](#) - Domain linker prediction using SVM at Tokyo University of Agriculture and Technology
- ...

User friendly

This server predicts secondary structure of protein from the amino acid sequence. In this server, Chou & Fasman algorithm has been implemented.

APSSP: Advanced Protein Secondary Structure Prediction Server

Dr. G P S Raghava, Scientist & Head, [Bioinformatics Centre](#)
[Institute of Microbial Technology](#), Sector 39A, Chandigarh, INDIA

Request Form

Enter the Protein Sequence (in fasta format)

```
>SEQ
MFRSVFPNISPLPQKDDSSLSASEYPNIASRTAPPSPSAVRRTHSLLSETHGTGYSQMEAGETTSL
GKTRFNHRGTPRRSYTSISAIPTFDNYLRHSLTSGSLRRSRHRSRANSQSLRFSRRSSIDDEQLP
AKDGMTASFLDERNWDYQFTSTDWVHDSIADGARLRELKRKDFRGLLAADFQAQGWILVALIGCIT
IAYFVDVTEDFVFDLKEGFCTTRWFHNRESCCADTLDCSRWRWSQILSPGSDNGWVDHSMFVLW
SVISCYLTLFTKTVPVSSVSLTLDENLGAGTSRGTNHDASEDNPSASLINPKAHYPTISTRPAMY
AGSGVAEVKVINSGFVLHGFLGFKTLVIKTIALVFSVSSGLSLGKEGPYVHIGACVGNIAICRLF
```

CLEAR PREDICT

Target/name of protein (optional):

Paste your sequence data here:

Please Select the Format of your Sequence (File or Paste)

Single sequence - amino acids only

Please Enter your Email:

(Please enter your email address if you want to receive your result via email)

Clear All

Run predictions!

Většina programů
má uživatelsky
jednoduché rozhraní



Agadir

(<http://agadir.crg.es/>)

Predikce obsahu helixu ve struktuře peptidů

Nevhodné pro delší proteiny (max 200-300 AK)

Nutná registrace

Agadir Results:

Peptide 1 MVCNISEFKWPYQPEDLNLTDALRLKYLGPQQMKQFVPICVTYLLIFVVGTGNGLT

Peptide 2 PTNFYLFSLAVSDMLVLLVGLPLELYEMQQNYPFQLGASACYFRILLLETVCLASVLM

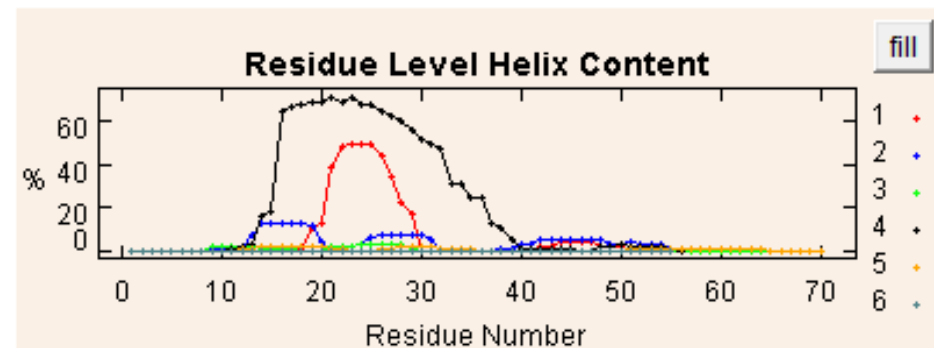
Peptide 3 VRPLQAKSVMTRAHVRRMVGAIWVLATLFSLPNTSLHGLSQLTVPCRGPVPSAIC

Peptide 4 TTALLFFCLPMVTISVLYLLIGLRLRRERMLLQVEVKGRKTAATQETSHRRIQLQDRG

Peptide 5 VVFGICWAPFHADRIMWSLVYGHSTEGHLAYQCVHIASGIFFYLGSAANPVLVSLMS

Peptide 6 LGTQCCHRRQPYHGSHNHIRLTTGSTLCDVGHNRNDRDEPLAVNEDPGCQQETDPS

[Check the output](#)



PSIPRED

bioinf.cs.ucl.ac.uk/psipred/

Chcete-li mít své záložky vždy po ruce, umístěte je sem na lištu záložek. [Importovat záložky...](#)

Site Navigation

- Introduction
- People
- Projects
- Publications
- Web Servers
- Downloads
- Vacancies
- Contact
- Group Intranet

Server Navigation

- PSIPRED Server
- PSIPRED help
- Server Overview
- Server Citation
- News
- History
- Software Download
- Login

The PSIPRED Protein Sequence Analysis Workbench

The PSIPRED Protein Sequence Analysis Workbench aggregates several UCL structure prediction methods into one location. Users can submit a protein sequence, perform the predictions of their choice and receive the results of the prediction via e-mail or the web. For a summary of the available methods you can read [More...](#)

The PSIPRED Team

Current Contributors David T. Jones, Daniel Buchan, Tim Nugent, Federico Minneci & Kevin Bryson

Previous Contributors Anna Lobley, Sean Ward, Liam J. McGuffin

For queries regarding PSIPRED: psipred@cs.ucl.ac.uk

Input Sequence Filter

Choose Prediction Methods

<input checked="" type="checkbox"/> PSIPRED v3.3 (Predict Secondary Structure)	<input type="checkbox"/> DISOPRED2 (Disorder Prediction)
<input type="checkbox"/> pGenTHREADER (Profile Based Fold Recognition)	<input type="checkbox"/> MEMSAT3 & MEMSAT-SVM (Membrane Helix Prediction) <i>Updated!</i>
<input type="checkbox"/> BioSerf v2.0 (Automated Homology Modelling) <i>Updated!</i>	<input type="checkbox"/> DomPred (Protein Domain Prediction)
<input type="checkbox"/> FFPred v2.0 (Eukaryotic Function Prediction) <i>Updated!</i>	<input type="checkbox"/> GenTHREADER (Rapid Fold Recognition)
<input type="checkbox"/> MEMPACK (SVM Prediction of TM Topology and Helix Packing)	<input type="checkbox"/> pDomTHREADER (Fold Domain Recognition)

[Help...](#)

Input Sequence (Single sequence or Multiple Sequence alignments; as raw sequence or fasta format)

[Help...](#)

If you wish to test these services follow this link to retrieve a [test fasta sequence](#).

Submission Details

Email Address for job completion alert (optional)

PSIPRED

Sequence analysis results for job: AB

ID: a2588d36-654d-11e2-adf3-00163e110593

Summary

PSIPRED

Downloads

Secondary Structure Map

Feature predictions are colour coded onto the sequence according to the sequence feature key shown below.

```
1 M P R S V P P N I S P L P Q K D D S S L S A S E Y P N I A S R T A P P S P S A V R R T H S L L S E T 50
51 H T G Y Q S L E S Q M E A G E T T S L L G K T R E N H R G T P R R S Y T S I S A I P T P D N Y L R H 100
101 S L T S G S L R R S R H H S R A N S Q S L R F S R R S S I D D E Q D E D L P P S A K D G M T A S F L 150
151 D E R N W Y D Q F T S T D W V H D S I A D G A R L R E L R K R K D F R G R L L A A F D G A Q G W I L 200
201 V A L I G C I T A A I A Y F V D V T E D F V F D L K E G F C T T R W F H N R E S C C A D T L D C S R 250
251 W R S W S Q I L S P S G S D N G W V D H S M F V L W V V I L S V I S C Y L T L F T K T V V P S S V S 300
301 L T T L D E N L G A G T S R G T N H D A S E D N S P A S L I N P K A H Y P T I S T R P A M T Y Y S A 350
351 A G S G V A E V K V I N S G F V L H G Y L G F K T L V I K T I A L V F S V S S G L S L G K E G P Y V 400
401 H I G A C V G N I A C R L F S K Y N D N D G K R R E V L S A S A A S G V A V A F G A P I G G V L F S 450
451 L E E V S Y Y F P P K T L F R T F F C C I A A A L S L K F L N P Y G T G K I V L F Q V R Y V T D W E 500
501 I F E I V V F A L L G V L G G A A G A L F I K A S S L W A K S F R K L S I I K R W P M L E V I L V A 550
551 L V T G V V S F W N R Y A K L P V S E L L F E L A S P C D P E S V T S T G L C P T E D G I G E I I S 600
601 D L L V A F V I K S L L T V V T F G I K V P A G I Y V P S M V V G G L M G R I V G H V V Q F L V V K 650
651 F P N F F L F S T C P V Y S G M E S C V V P G V Y A M V A A G A T M C G V T R L S V T L A V I L F E 700
701 L T G S L D H V L R F E S L A V L C A K W T A D A I E P R S I Y D L L T D M N S Y P F L D N K I Q L L 750
751 S D D E L G D I V R P V R K S R V I D I S E S P F V P A T E L R S K L Q H L L M A G E L D S G L P I 800
801 L R N D I L C G I I P A P D L E Y A L D T I E D E E H T M C L M S M D T A S A V V D S E D S N G N S 850
851 W V D F R R Y I D P A P I S L D I H S P I D L V Y Q C F A K L G L R Y L C V L R D G Q Y A G L V H K 900
901 K A F V K Y V K E N E
```

Key

Annotations

Helix

H

Sheet

S

Disordered

D

Dompruned Boundary

B

DomSSEA Boundary

B

Jpred

Predikce
sekundární
struktury
proteinu

<http://www.compbio.dundee.ac.uk/www-jpred/>

Jpred 3
Incorporating Jnet

A Secondary Structure Prediction Server

Sequence:

[Help](#)
[Advanced](#)

Make Prediction

Clear

[The Barton Group - The University of Dundee](#)

Citation: Cole C, Barber JD & Barton GJ. *Nucleic Acids Res.* 2008. 35 (suppl. 2) W197-W201 [\[link\]](#)

[More citations](#)



Jpred

3D > 2D

*PDB (Protein
Data Bank)
– databáze
3D struktur*



Jpred 3
Incorporating Jnet

The
Barton
Group

[Home](#)

[About](#)

[News](#)

[FAQ](#)

[Help](#)

[Contact](#)

Match found in PDB

You might want to reconsider the accuracy and what you might gain from secondary structure prediction, if close sequence homologues exist in the structural database.

If you still want to carry out a Jpred prediction click

Hits found

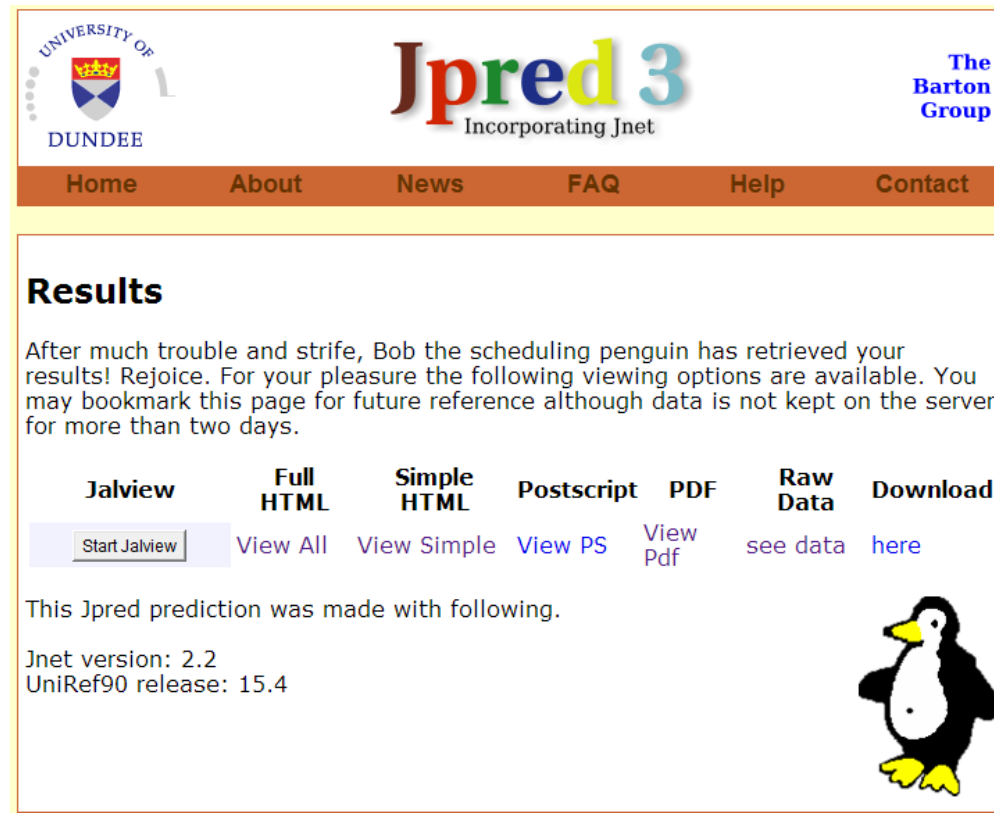
PDB	Chain	Description	Blast E-value
3qpw	A	6-phosphofructo-2-kinase/fructose...	0.0
3qpv	A	6-phosphofructo-2-kinase/fructose...	0.0
3qpu	A	6-phosphofructo-2-kinase/fructose...	0.0
2i1v	B	6-phosphofructo-2-kinase/fructose...	0.0
2dwp	A	6-phosphofructo-2-kinase/fructose...	0.0
2dwo	A	6-phosphofructo-2-kinase/fructose...	0.0

Jpred

PSI-Blast > Predikce 2D struktur > Konsensus

Výstupní formáty:

- Jednoduchý
- HTML
- PostScript
- PDF



The screenshot shows the Jpred 3 web interface. At the top left is the University of Dundee logo. The main title is "Jpred 3 Incorporating Jnet". On the top right is "The Barton Group". A navigation bar contains links for Home, About, News, FAQ, Help, and Contact. The "Results" section contains a message: "After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days." Below this message are several viewing options: "Jalview" (with a "Start Jalview" button), "Full HTML" (with a "View All" link), "Simple HTML" (with a "View Simple" link), "Postscript" (with a "View PS" link), "PDF" (with a "View Pdf" link), "Raw Data" (with a "see data" link), and "Download" (with a "here" link). At the bottom, it states "This Jpred prediction was made with following." followed by "Jnet version: 2.2" and "UniRef90 release: 15.4". A cartoon penguin is in the bottom right corner.


```

QUERY : MPLELTQSRVQKIWVVDHRPSLPRSCGPKLTNSPTVIVMVGLPARGKTYISKKLTRYLNWIGVPTKVFNVGVEYRREAVKQYSSYNFFFRPDNEEAMKVRKQCALAALRDVKSYLA
UniRef90_UPI00016E380D : MPRELTQNRIQKIWIWPNKDKPAPRRAGAPHFANPPTVIVMVGLPARGKTYISKKLTRYLNWIGMPTKVFNVGVEYRREAVKNYSSYDFFRPDNECAVKIRQQCALAALRDVKSYLK
UniRef90_UPI0000F2BD4C : -----LPTASDPSLLAAWASYMTNSPTLIVMIGLPARGKTYVSKKLTRYLNWIGVPTKVFNLGVYRREAVKSYKSYDFFRHDNEEAMKIRKQCALVALEDVKAYLT
UniRef90_COHAN1 : -----SWASYMTNSPTVIVMIGLPARGKTYMSKKLTRYLNWIGVPTKVFNLGVYRREAVKAYKSYDFFRHDNKEAMEIRKQCALVALEDVKGYLT
UniRef90_Q6NRYO : -PLELTQNPLKKIWVFPNNGHPVQHSARQKMTNCPTLIVMVGLPARGKTYISKKLTRYLNWIGVPTKEFNVGQYRRDLVQTFKSFEFFLPDNEEGQKIRKQCALLALNDVRRYLG
UniRef90_P16118 : ---ELTQTRLQKIWIWPHSSGSSRLQRRIPQFTNSPTMIVMVGLPARGKTYISTKLTRYLNWIGTPTKVFNLGQYRREAV-SYKNYEFFLPDNEALQIRKQCALAALKDVHNYLS
UniRef90_A2AR70 : --KELTQTPLLKIWVWPWGSGLGQRGSSRPQFTNCPTMIVMVGLPARGKTYISKKLTRYLNWIGVPTQVFNLGQYRREAVVQTYKNYEFFLPDNEEAMTIRRSACAALDKIGNYFA
UniRef90_UPI0000ECA0DB : -----SWASYMTNSPTLIVMIGLPARGKTYVSKKLTRYLNWIGVPTRVFNLGVYRREAVKSYKSYDFFRHDNKEAMEIRKRCALVALQDVKAYLL
UniRef90_UPI00016E31A6 : -PRELTQNPLKKIWMPFKNGHISQRRV--CMTNCPTLIVTVGLPARGKTYISKKLTRYLNWIGVPTKEFNVGVEYRRECLKIYKSEFFFRPDNEEGLKIRRQCAMAALNDVRQYLS
UniRef90_Q7PZN8 : -----YVNSPHVIAMVGLPARGKTYISKKLSRYLNWIGINTKVFNLGEYRRHATDAYRSHEFFFRPDNEEAMAIRQHCAELALELDCVNWL-
UniRef90_B7PT92 : -----YVNVPHI IAMVGLPARGKTYIAKKLTRYLNWIGIQTRVFNVGEYRRQATEAYRSHDFFRADNAEAMAIRSKCALDALEDMCKWLE
UniRef90_B8MIQ5 : -----MSSGESGRTESASKSDDTKICVVMVGLPARGKSLIAGKALRYLAWVGISARIFNVGTYRRLKSP-HPLANFFDPHNPEGEKMRWAAAEAAVSDMLKWFK
UniRef90_B3RUK6 : -----PKLCNVPTVIVMVGLPARGKTFIARKLTHYLNWIGYKTKVFNLGEYRRKNFGATKPPSFYHPDNKEAVASRLKCCALALEDMTEFLK
UniRef90_A1CDR3 : -----SAQRGSGQKSFASDDSKI CVVMVGLPARGKSLIAGKAMRYLGVWGI PARVFNVTGTYRRQDTPQ-PTATFFDPHNPEGERMRAAAEAALSMDLNNWFK
UniRef90_B2Z894 : -----SWASYMTNSPTLIVMIGLPARGKTYVSKKLTRYLNWIGVPTKVFNLGVYRREAVKSYQSYDFFRHDNEEAMKIRKQCALVALEDVKAYFT
UniRef90_COHA25 : -PRELTQNPLKKIWMPWKNGLFVHILQRRCTNCPTLIVTVGLPARGKTYISKKLTRYLNWIGVPTKEFNVGQYRRECVKIYKSEFFFRPDNEEGLKIRRQCALTALKDVHQYLT
UniRef90_A4R8B3 : -----VGVRAEDTRLCVVMVGLPARGKSYIAQKAQRYLRWLSIEAKTFNVGNRYRRKDA-AHPPADFFDTQNEDEGKRRRAAAEAALNDMLSWFR
UniRef90_B4PZH7 : -----YRRQATSAYKSHEFFRADNEEAMAIRSRCANQALHDS CDWLL
UniRef90_Q6FOH9 : -----MPVTTISNMNDSRICVVMVGLPARGKSFISQKILRYLSWLSINAKCFNVGNRYRKYAGVIKPDAKFFDFNNPEGYRVREAAQQAVSDMLKWYE
UniRef90_Q6DTY7-3 : -PRELTQNPLKKIWMPYSNRPALHASQGVCMTCNPTLIVMVGLPARGKTYISKKLTRYLNWIGVPTREFNVGQYRRDIVKTYKSEFFLPDNEEGLKIRKQCALAALSDVRKFLS
UniRef90_A7TGQ5 : -----ENLKV CVVMVGLPARGKSFISQKIVRFLSWISINARSFNVGQYRRDAGTLEPSSDFDFANEENYKLRKQAESFAVEDILKWFNI
UniRef90_A9GKW0 : -----RDSSPIVLAMVGLPARGKTSIARKIARYLSWLGHPTRLFNVGNYRRAHLGTAQPHQFFDPNNTTEGRQALLGMAMTALDDLIGWLS
UniRef90_Q6FWD2 : -----KRRPTTIDVPTKSKTSPDGLISKEDPGSKLIVMVGLPATGKSFITNKLSRYLNYSMYCQVFNVGNTRRKYAKQEQDSFFDPTNTEYALRDKWALDITLEELMDYLL
UniRef90_Q0DKG9 : -----EDRKLAIVLVGLPARGKTFIAAKLTRYLRWLGHETRHFNVGKYRRLKHGANQSADFFRDDNPEGIEARNEVAALAMEDMDI DWM-
UniRef90_P32604 : -----STISNDNDTKVCVIMVGLPARGKSFISOKTIRYLSWLSIKAKCFNVGNRYRRDVSQNVDPAEFFNFENTDNFKTRELAAONATKDTVNFFT
UniRef90_A4HEN8 : -----PDFSAVQPMLGSLIVIMVGLPGRGKTYVAEQIRRYFQWNGLACRIFSHQACRRRVEGVDRDASVLSPTGVAEVEF--RIAKELAHDLITYIC
UniRef90_C1FHQ5 : -----
UniRef90_Q4DA36 : -----DAKSFYA
UniRef90_A4HTA2 : -----
UniRef90_A8JG80 : -----
UniRef90_B2VW18 : -----
UniRef90_Q1E1L6 : -----
UniRef90_Q5B397 : -----
UniRef90_Q6C8W1 : -----
UniRef90_Q6FN95 : -----

OrigSeq : 1-----11-----21-----31-----41-----51-----61-----71-----81-----91-----101-----111-----
          : MPLELTQSRVQKIWVVDHRPSLPRSCGPKLTNSPTVIVMVGLPARGKTYISKKLTRYLNWIGVPTKVFNVGVEYRREAVKQYSSYNFFFRPDNEEAMKVRKQCALAALRDVKSYLA

Jnet      : -----EEEE-----EEEEEE-----HHHHHHHHHH-----EEEE--HHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
jhmm      : -----EEEE-----EEEEEE-----HHHHHHHHHHHH-----EEEE--HHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
jpssm     : -----EE-----EEEEEE-----HHHHHHHHHH-----EEEE-----EEEE-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

Lupas 14  : -----
Lupas 21  : -----
Lupas 28  : -----

Jnet_25   : -----B-BBB-----BB-----BBBBBBBBBB-B-BBBB-BBB-BB-B--B-BBBB-B--BB--B-----B-BB--B--BB-BB--BB--BB--BB-BB--
Jnet_5    : -----BBBBBB-----B-B--BB--B--B--B-----B-----B--BB--BB--B--B--
Jnet_0    : -----BBB-----B-----B-----B-----B-----B-----B-----B-----B-----B-----B-----B-----B-----
Jnet Rel  : 998888877700202577777777522477787458888437887770002677764200886127761570000006777777766776468999999999999999987

```


Úloha

- Seznamte se s programem **Jpred 3** (<http://www.compbio.dundee.ac.uk/www-jpred/>). Pro následující proteinové sekvence predikujte zastoupení sekundárních struktur. Patří tyto proteiny do skupiny tzv. α/β proteinů, tj. proteinů, ve kterých jsou zastoupeny obě typické sekundární struktury (α -helix, β -skládaný list)?

Úloha

- Protein A:

GTYRQLFHPEQLISGKEDAANNYARGHYTIGKEIIDLVLDVRKRLADQCTGLQGFLVFHSFGGGTGSFGFTSLLME
RLSVDYGGKSKLEFSIYPAPQVSTAVVEPYNSILTTHTTLEHSDCAFMDNEAIYDICRRNLDIERPTYTNLNLRLMS
QIVSSITASLRFDGALNVDLTEFQTNLVPYPRIHFPLATYAPVISA EKAYHEQLSVAEITNACFEPANQMVKCDPRH
GKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYQPPTVVPGGDLAKVQRAVCMLSNTT
AVAEAWARLDHKFDL MYAKRA FVHWYVGEGMEEGEFSEAREDMAALEKDYE EVGADSYEDEDEGE EY

- Protein B:

SIRLPAHLRLQPIYWSRDDVAQWLKWAENEFSLRPIDSNTFEMNGKALLLLTKEDFRYRSPHSGDVL YELLQHIL
KQAGPNIFEMLRIDEGLRLKIYKDTEGYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLCQDVDA
AVRGILRNAKLKPVYDSLDCVRRRAALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLA KSRWYNQTP
NRAKRVITTFRTGTW DAYK

- Protein C:

ADVCM DPEPIVRIVGRNGLCVDVRDGRFHNGNAIQLWPCKSNTDANQLWTLKRDNTIRSNGKCLTTYGYSPG
VYVMIYDCNTAATDATRWQIWDNGTIINPRSSLVLAATSGNSGTTLVQTN IYAVSQGWLPTNNTQPFVTTIVG
LYGLCLQANSQVWIEDCSSEKAEQQWALYADGSIRPQQNRDNCLTSDSNIRETVVKILSCGPASSGQRWMFK
NDGTILNLYSGLVLDVRASDPSLKQIILYPLHGDPNQIWLPLF

Rozšíření možností 2D predikce

- Přístupnost pro **solvent**
Jpred, NetSurfP
- Předpověď **transmembránového** helixu
TMpred, HTMSRAP
- Predikce **dalších typů 2D struktur**
GOR4, SSpro8

TMpred

http://www.ch.embnet.org/software/TMPRED_form.html

Predikce přítomnosti
transmembránových
helixů, včetně
orientace (zevnitř ven
nebo zvenku dovnitř)

[ch.EMBnet.org](#)

[Home](#) [Services](#) [Courses](#) [Links](#) [Contacts](#)

TMpred - Prediction of Transmembrane Regions and Orientation

The TMpred program makes a prediction of membrane-spanning regions and their orientation. The algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring.

[K. Hofmann & W. Stoffel](#) (1993)

TMbase - A database of membrane spanning proteins segments
Biol. Chem. Hoppe-Seyler **374**,166

For further information see the [TMbase](#) and [TMpredict](#) documentation.

Usage: Paste your sequence in one of the supported [formats](#) into the sequence field below

and press the "Run TMpred" button.

Make sure that the format button (next to the sequence field) shows the correct format

Choose the minimal and maximal length of the hydrophobic part of the transmembrane helix

Output format	<input type="text" value="html"/> <input type="text" value="minimum"/> <input type="text" value="17"/> <input type="text" value="maximum"/> <input type="text" value="33"/>
Query title (optional)	<input type="text"/>
Input sequence format	<input type="text" value="Plain Text"/>
Query sequence: or ID or AC or GI (see above for valid	<input type="text"/>

Úloha

- Pokuste se pomocí programu **TMpred** předpovědět, kolik transmembránových helixů obsahuje protein NMU-R1 z myši (UniProtKB/Swiss-Prot: O55040.1). Bude podle predikce jeho C–terminální doména orientována spíše dovnitř buňky nebo ven?

TMpred output for gi_81908388

[ISREC-Server] Date: Mon Feb 11 14:52:26 2013

tmpred -par=matrix.tab -html -min=17 -max=33 -def -in=wwwtmp/.TMPRED.26869.2906.seq -out=wwwtmp/.TMPRED.26869.2906.out -out3=wwwtmp/.TMPRED.26869.2906.txt >wwwtmp/.TMPRED.26869.2906.err

Sequence: MVC...DPS, length: 405

Prediction parameters: TM-helix length between 17 and 33

1.) Possible transmembrane helices

The sequence positions in brackets denominate the core region.

Only scores above 500 are considered significant.

Inside to outside helices : 7 found

from	to	score	center
40 (40)	59 (59)	1635	48
74 (78)	96 (96)	1553	86
112 (114)	135 (132)	830	122
158 (158)	176 (176)	1652	168
207 (210)	227 (227)	2501	218
275 (275)	293 (293)	2207	285
317 (317)	334 (334)	1372	325

Outside to inside helices : 7 found

from	to	score	center
37 (37)	62 (55)	1983	47
74 (77)	95 (95)	1269	86
115 (115)	134 (134)	1218	125
158 (161)	179 (177)	1329	169
207 (209)	229 (227)	2638	218
275 (275)	293 (293)	2207	285
317 (317)	334 (334)	1372	325

GOR4

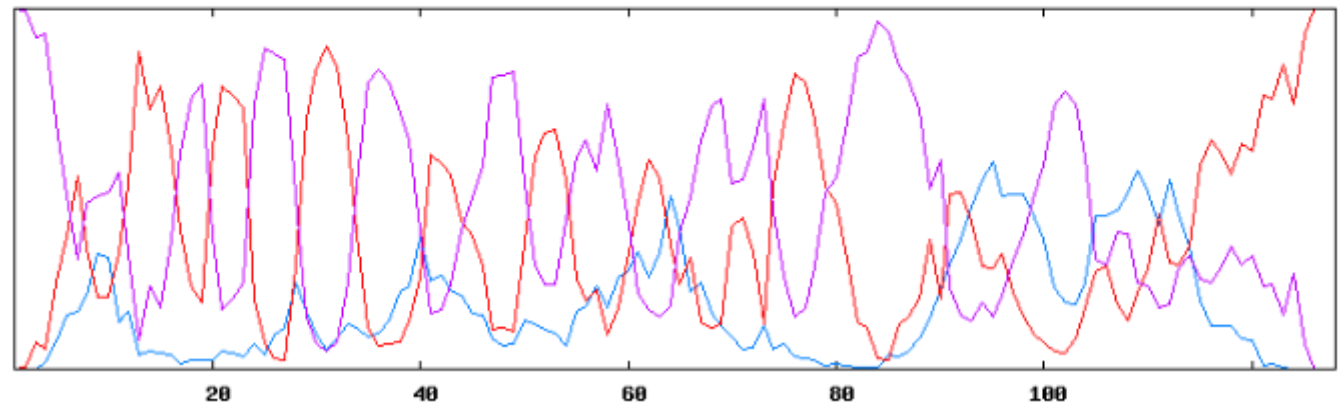
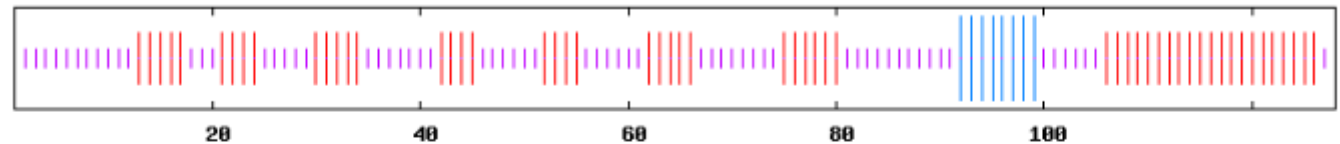
View GOR4 in: [\[AnTheProt \(PC\)\]](#), [\[Download...\]](#) [\[HELP\]](#)

```
          10          20          30          40          50          60          70
          |           |           |           |           |           |           |
MTPPTSSQLSKKTRWSYANGCTWSVKDDDHVLFIEGSGTSGMLRFKTFSGDFFTLVPGIHNLVSLVHLQG
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
DVYAPGVVSYGLNCSPPLGSLHVVASIYQLSEKNSTIASSFYRLRHFFTTKATSSGRS
ccccccccccccccccccccchhhhhhhcccccccccccccccccccccccccccccccccccccccc
```

Sequence length : 128

GOR4 :

Alpha helix	(Hh) :	8 is	6.25%
3 ₁₀ helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	54 is	42.19%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	66 is	51.56%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%



SSpro8 – výstup

E-mailem

Name: AB2

Amino Acids:

MTPPTSSQLSKKTRWSYANGCTWSVKDDDHVLFIEGSGTSGMLRFKTFSGDFFTLVPGIHNLVSLV

Predicted Secondary Structure (8 Class):

CCCCCCHHHCHHCEEEETTCEEEECTTEEEEEEECSCEEEEEEECSCEEEETTTCCEEEEEEE

H: alpha-helix

G: 3_{10} -helix

I: π -helix

E: extended strand

B: beta-bridge

T: turn

S: bend

C: the rest

Úloha

- Následující sekvenci analyzujte pomocí programu **CFSSP** (<http://www.biogem.org/cgi-bin/cho-fas.pl>).
Tutéž sekvenci analyzujte pomocí programu **GOR4** (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) a výsledky porovnejte.

Sekvence:

MSTDKPLVIQSDKTLLLDVHSPFAQECRDSITAFSELVKSPHEVHTFLLTPLSLWNANAAG
MTTEDIMGRLRTWSRYDIPEPVSYFITDISARFGSFVMTDIPDDADHYLLTVTIPRYAKEI
SSHKTVSSLLFPRGNDTFLLNRYARGEVKKLKLKLGFPVDDRIPLKKGFPVPMNLRQQTLS
GKDFSIRDYQEAAARSLLGDRGPGTGY

2D je málo pro 3D svět...



Predikce **3D struktury**
je komplikovanější – viz
samostatný blok

Úloha

- Nejspolehlivější informace o sekundární struktuře můžeme získat analýzou 3D struktury, je-li tato známa. 3D struktury jsou ukládány v tzv. **PDB databázi**. (např.). Vyhledejte na webových stránkách <http://pdb.rcsb.org> libovolný protein a podívejte se na informace o jeho sekundární struktuře.

<http://2struc.cryst.bbk.ac.uk/twostruc>

- Přiřazení sekundární struktury na základě známé 3D struktury
- Možnost vybrat z několika programů, včetně tvorby tzv. „konsenzu“

Úloha

- Stejný protein, který jste již analyzovali v PDB databázi analyzujte rovněž pomocí nástrojů na serveru **2struct** (<http://2struc.cryst.bbk.ac.uk/twostruc>). Porovnejte přiřazení sekundárních struktur v jednotlivých programech.



The Secondary Structure Server

[Front Page](#)[2Struc](#)[Compare The Protein](#)[Compare The Protein Help](#)[About The Methods](#)[COOKIES](#)[Make Contact](#)

Welcome to the Control Panel for **4agi**,

From here you can select the method you want to run, pick the models and chains you want to analyse as well as select which analysis to run.

Each analysis will load in a new window leaving this one untouched so you can easily jump from one analysis to another.

Available Methods

- | | |
|---|-------------------------------------|
| <input checked="" type="checkbox"/> DSSP* | <input type="checkbox"/> DSSP Cont* |
| <input type="checkbox"/> Stride* | <input type="checkbox"/> PSEA |
| <input type="checkbox"/> PALSSE | <input type="checkbox"/> Sticks |
| <input type="checkbox"/> XTLSSTR | <input type="checkbox"/> Kaksi |

* can be used with Swiss Model files

[Select All](#)[Deselect All](#)

Available Structures

Model	Chains	Selected
1	ABCD	<input checked="" type="checkbox"/>

[Protein Structure Summary](#)[Structure Summary By Chain](#)[Multiple Structure Alignments](#)[Original Structure Alignments](#)[Sequence Structure Alignments](#)

Porovnání predikce a experimentu

- **Predikce:**

- Rychlá (sekundy), levná/dostupná (Freeware)
- Spolehlivá pro známé (!) proteiny a pro proteiny s vysokou homologií
- Málo spolehlivá pro neznámé proteiny
- Nepoužitelná pro unikátní případy

- **Experiment:**

- Teoreticky použitelný pro libovolný protein
- Finančně (i miliony Kč) a časově (týdny až roky) náročná příprava proteinu a optimalizace měření