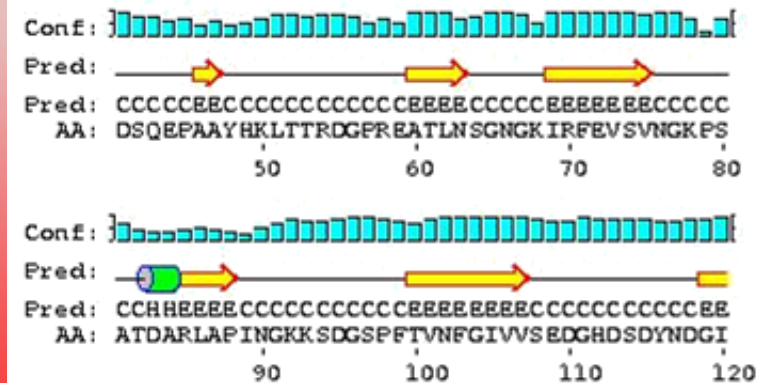




# Struktura proteinů

ADSQTSSNRAGEFSIPPNTDFRAIFFANAAE  
QQHIKLFIGDSQEPAAYHKLTTTRDGPREATL  
NSGNGKIRFEVSVNGKPSATDARLAPINGKK  
SDGSPFTVNFIVVSEdGHDSYNDGIVVL  
QWPIG

primární  
(sekvence)



sekundární



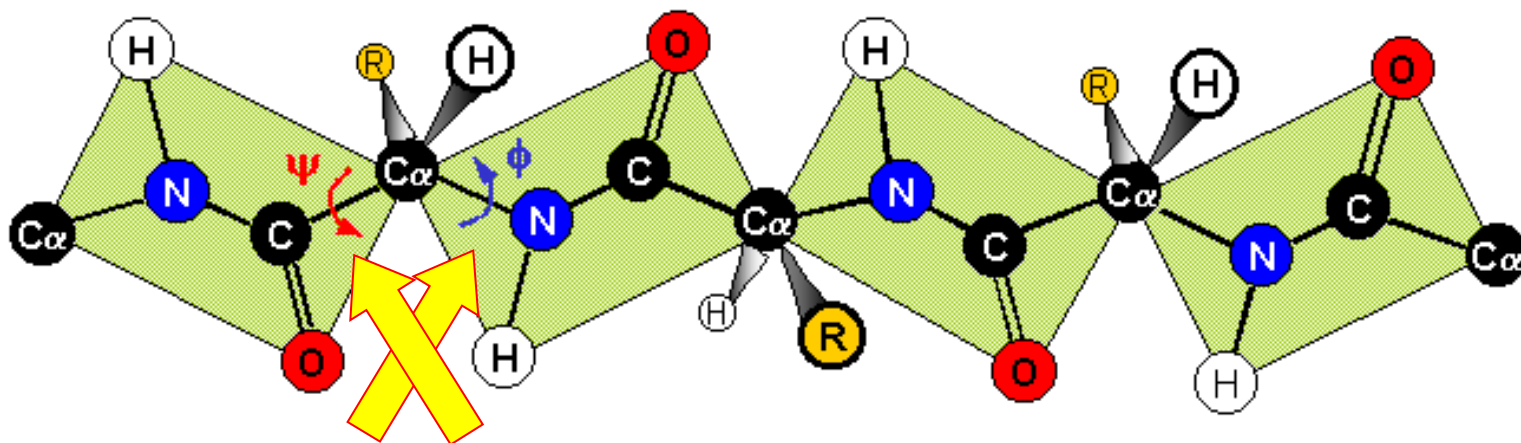
kvartérní



ter

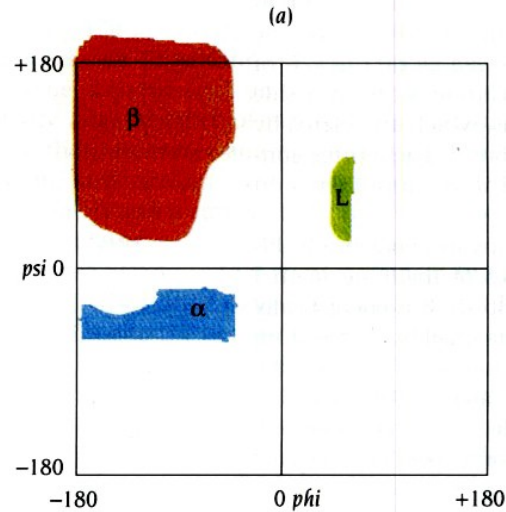
# Kostra polypeptidového řetězce

Peptidová vazba – planární



Konformaci kostry určují dva torzní úhly  $\phi$   
a  $\psi$  (úhel  $\omega$  je 180 )

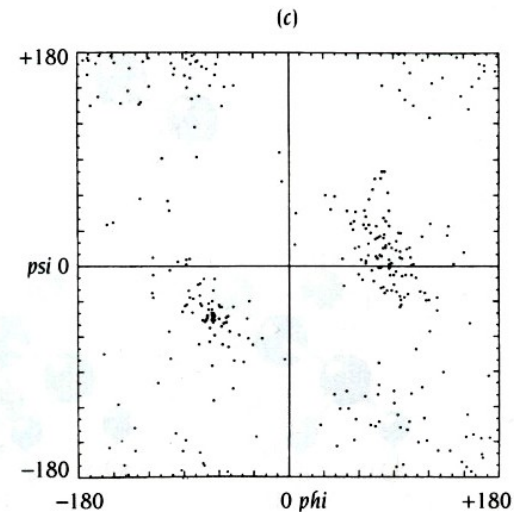
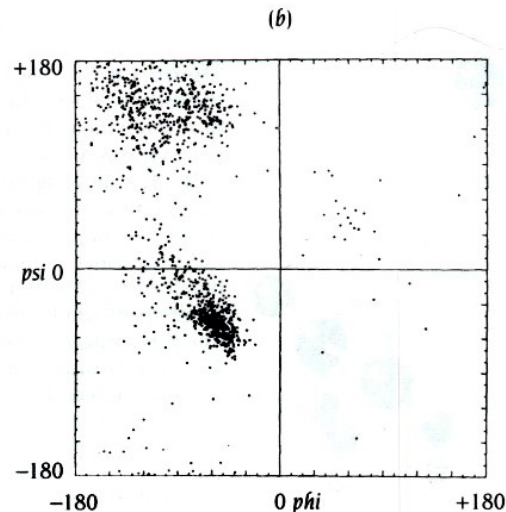
# Ramachandran plot diagram



a) Colored areas show sterically allowed combinations of the  $\phi$  and  $\psi$  angles

b) Observed values for all residue types except for glycine. Each point represents  $\phi$  and  $\psi$  values for an amino acid residue in a well-refined x-ray structure

c) Observed values for glycine

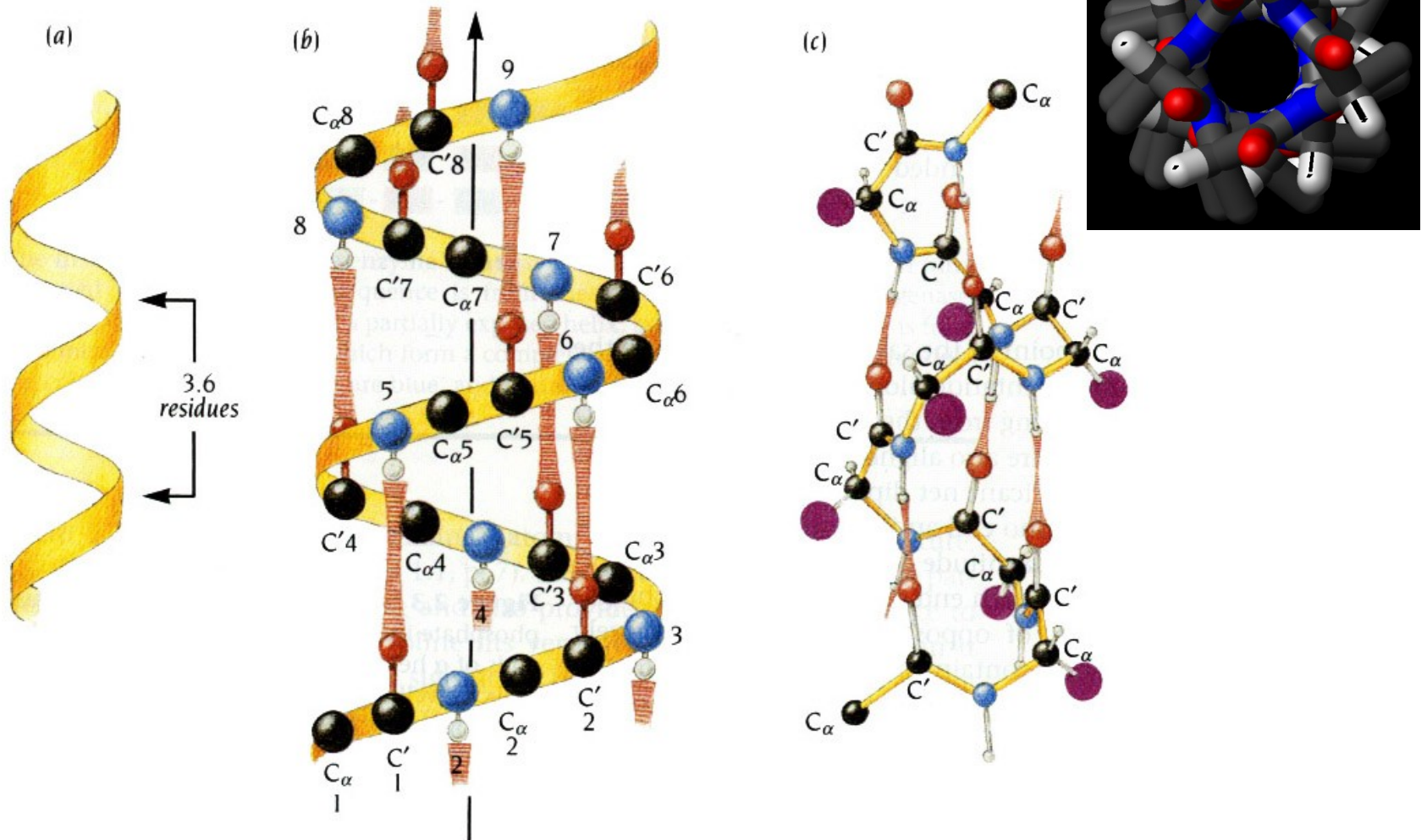


# Vazby zprostředkovávající vyšší struktury

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

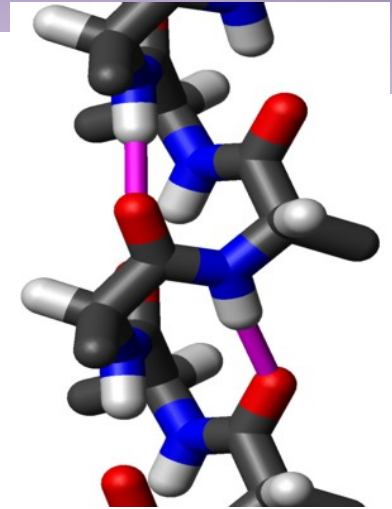
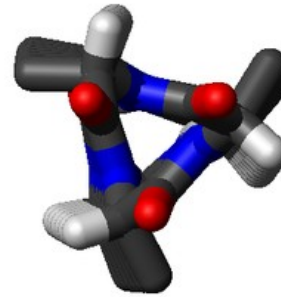


# $\alpha$ -helix

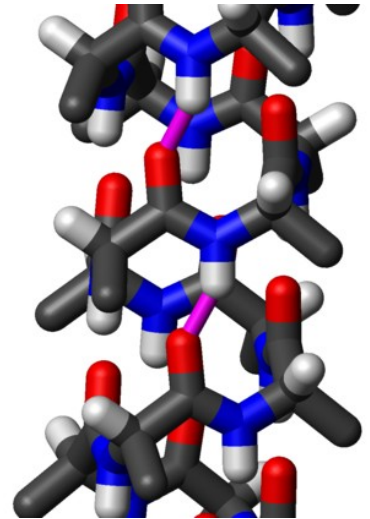
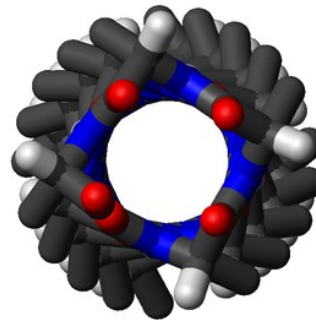


# Jiné helixové struktury

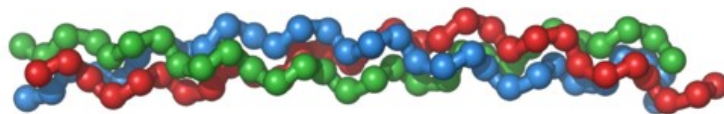
- $3_{10}$  helix



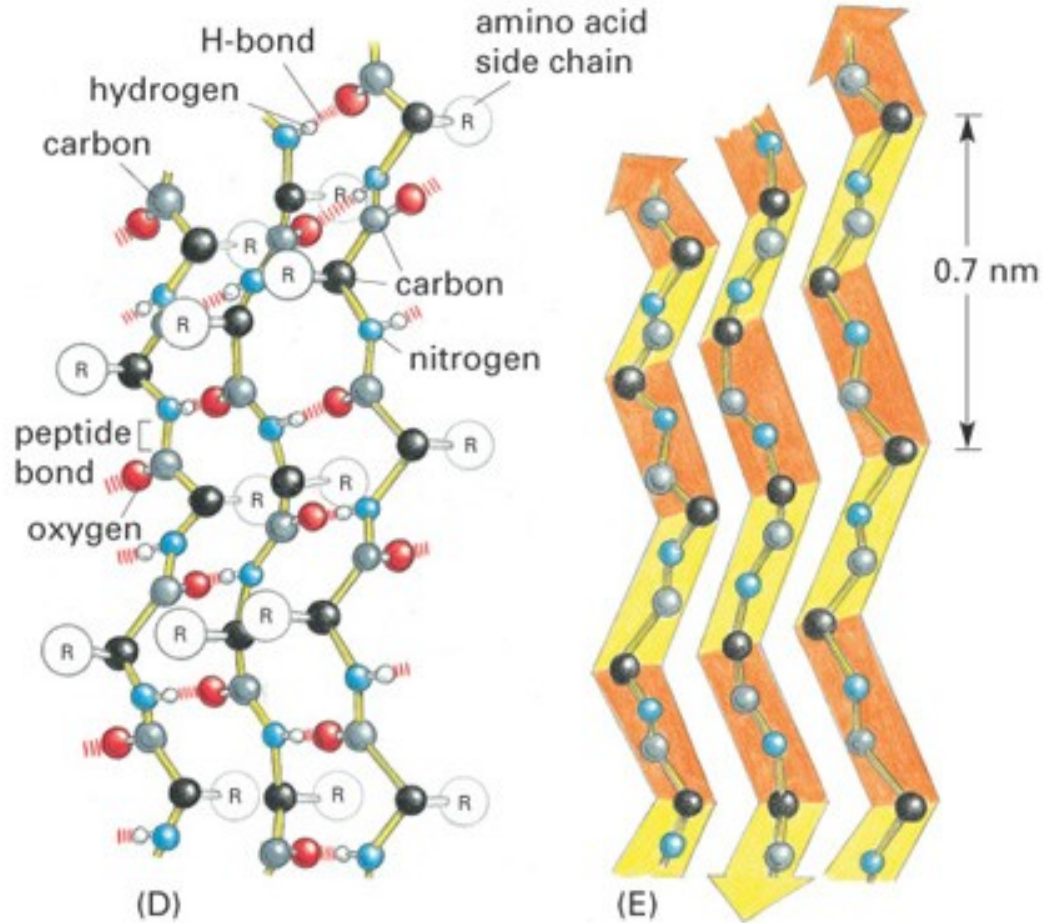
- $\pi$ -helix



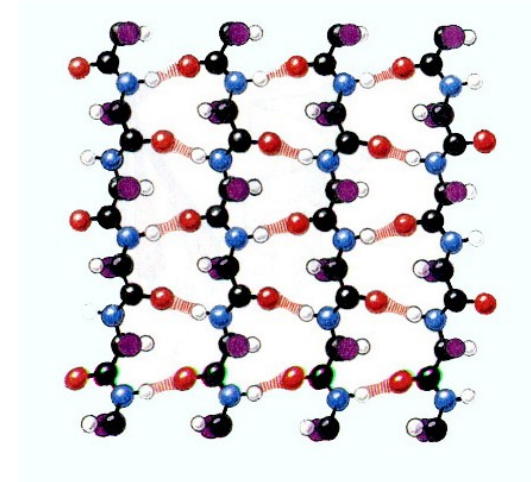
- kolagen – levotočivá šroubovice



# $\beta$ -list



Paralelní



Antiparalelní

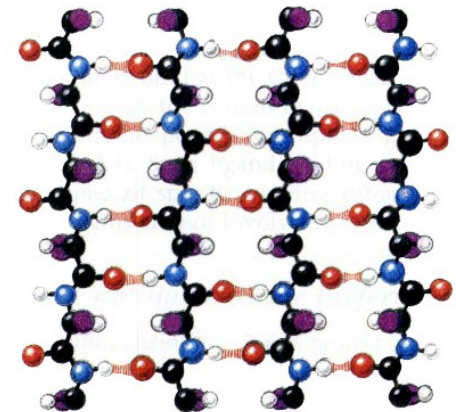
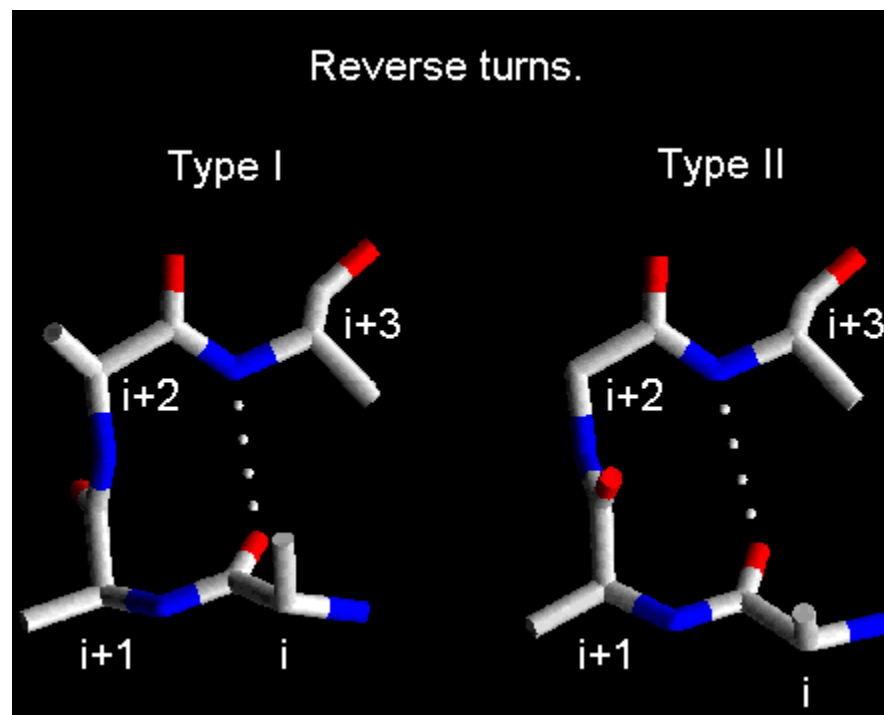


Figure 4-10 part 2 of 2 Essential Cell Biology, 2/e. (© 2004 Garland Science)



# Otáčky (turns)

- více typů otáček, různé počty AK
- příklad  $\beta$ -turn

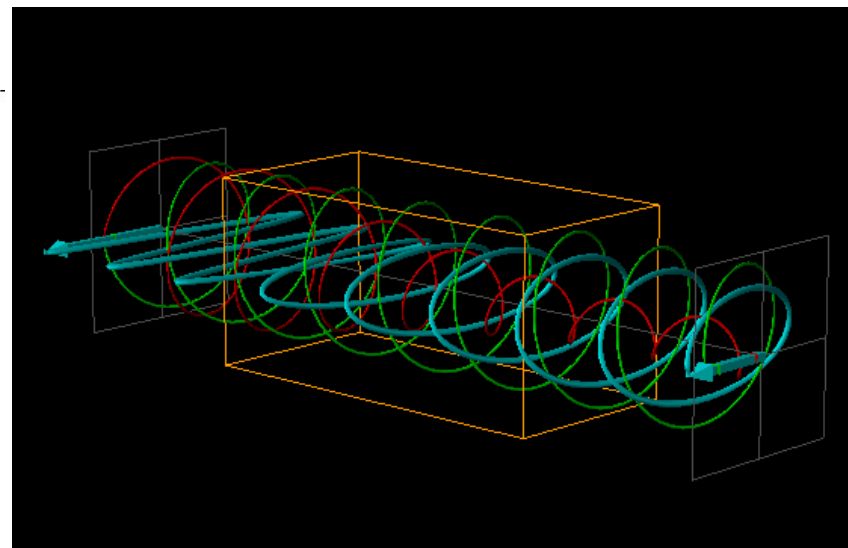


# Určení 2D struktury proteinů

- **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, ...)

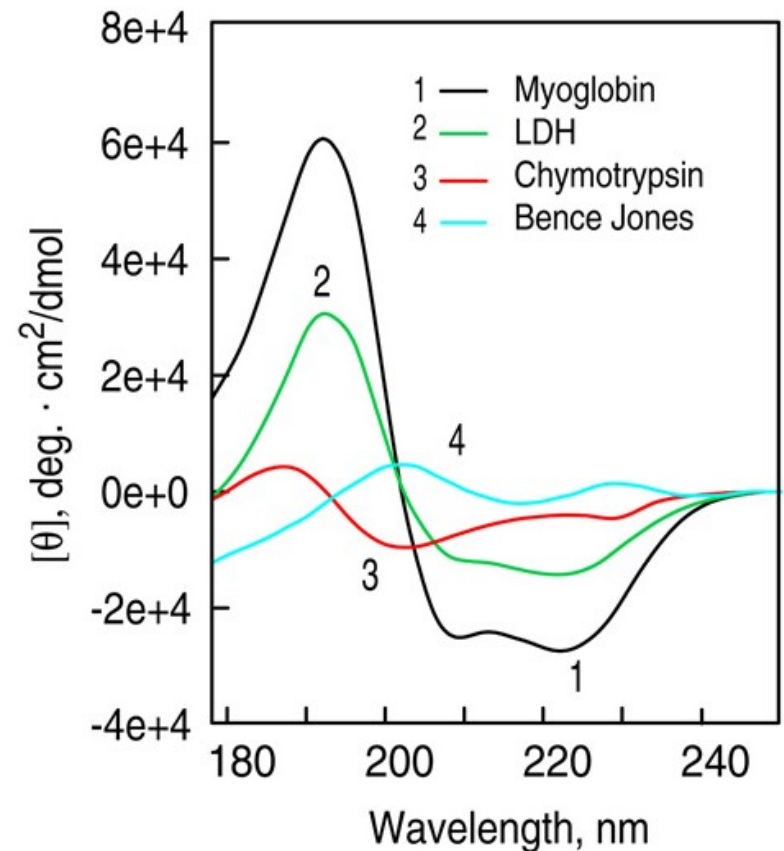
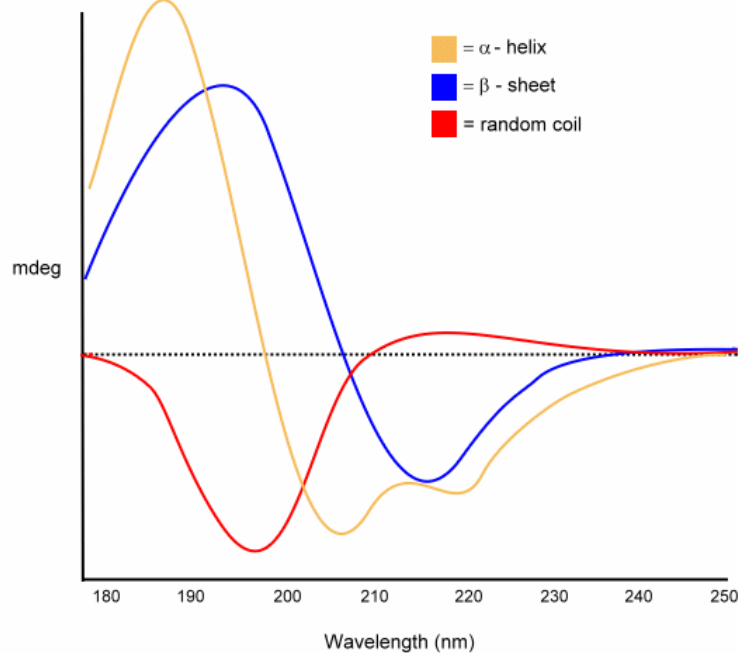


# Cirkulární dichorismus - veličiny

Veličina	Definice a vztahy	Jednotky
Absorbance	$A(\nu) = \log_{10} (I_0/I)$ <p><math>I_0</math> – intenzita světla vstupujícího do kyvety,  <math>I</math> – intenzita světla opouštějícího kyvetu</p>	bezrozměrná
Molární absorpční koeficient	$\varepsilon(\nu) = A(\nu)/cl$ <p><math>c</math> – molární koncentrace (<math>\text{mol L}^{-1}</math>), <math>l</math> – 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  $\alpha$ -helix,  $\beta$ -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



### Protein Circular Dichroism Data Bank

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

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

[Create account](#)

Login ID

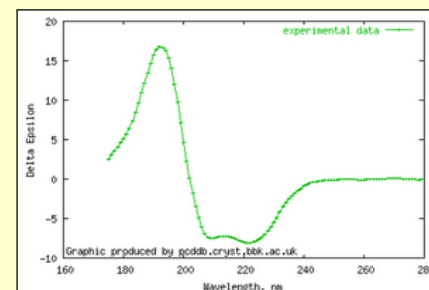
Password

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](mailto:pcddb@mail.cryst.bbk.ac.uk).

# Výpočet zastoupení sekundárních struktur

Různé SW dostupné  
jako freeware

k2d – aktuálně  
3. generace K2D3

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

## K2D3

Estimates protein secondary structure from CD spectra

*Proteins (2012) Vol 80-2*

about  
supplement  
use K2D2  
contact us

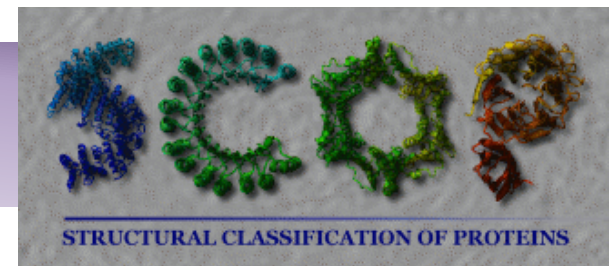
Input your CD data [help](#) [example](#)

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  $\alpha$ -helixu,  $\beta$ -skládaného listu nebo náhodného klubka (coil). K vyhodnocení využijte program **k2d3** (<http://k2d3.ogic.ca/>).

# 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 ( $\alpha/\beta$ )	147	244	803
Alpha and beta proteins ( $\alpha+\beta$ )	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
<b>Total</b>	<b>1195</b>	<b>1962</b>	<b>3902</b>

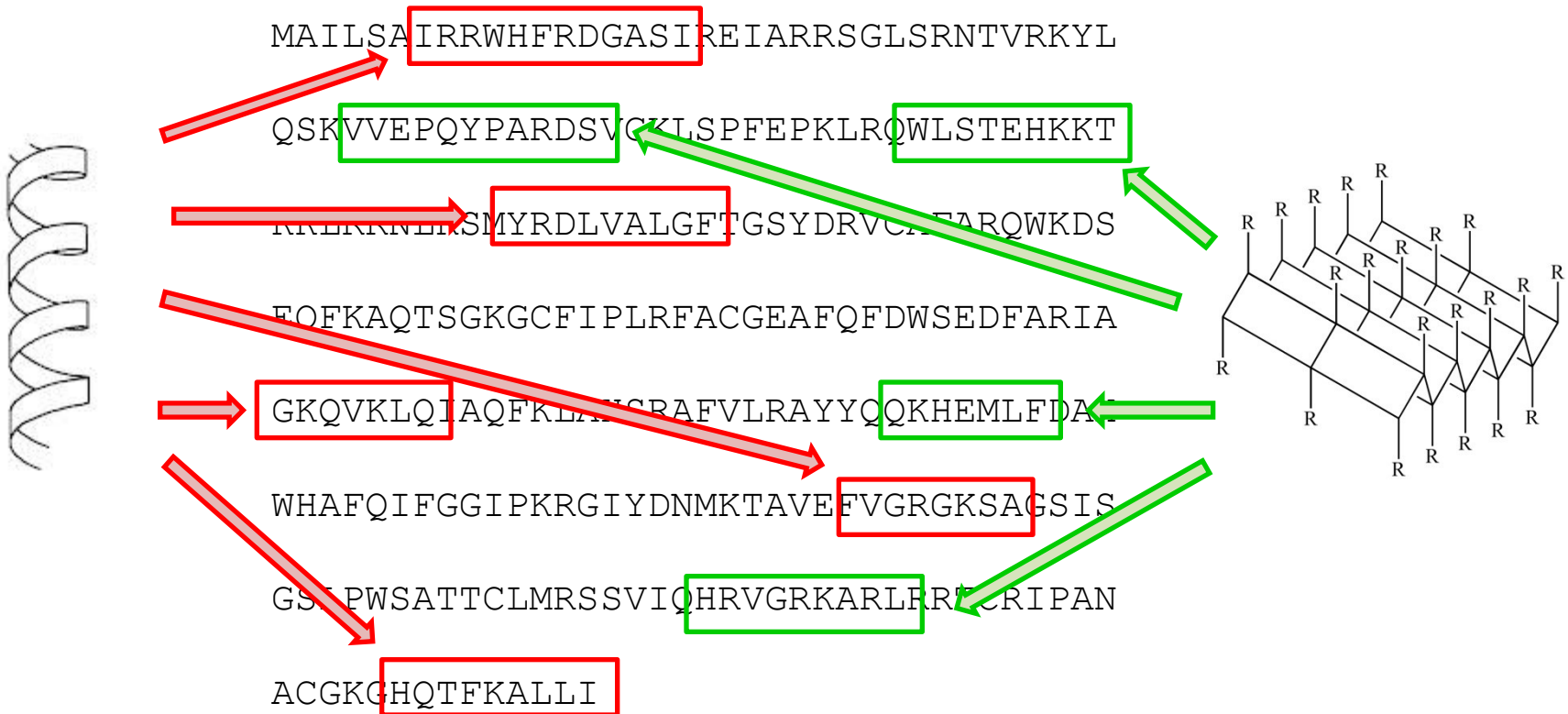


# Další čtení o cirkulárním dichorismu

- 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/pdf/cd/handout.pdf>

# Predikce 2D struktury

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



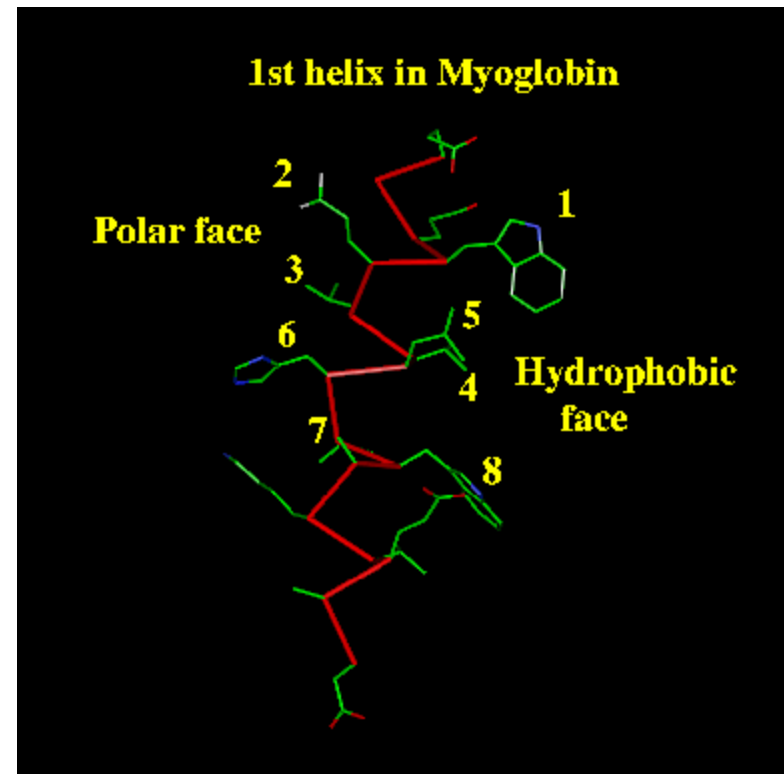
# Predikce 2D struktury

- Rozlišujeme tři základní typy 2D struktury
  - H – helix
  - E –  $\beta$ -list
  - C/(-) – smyčka/náhodné klubko (coil) – někdy jsou rozlišovány tyto dvě varianty
- S dobrou přesností lze určit helix, u  $\beta$ -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 $\alpha$ -helix

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

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



# Typické znaky $\beta$ -list

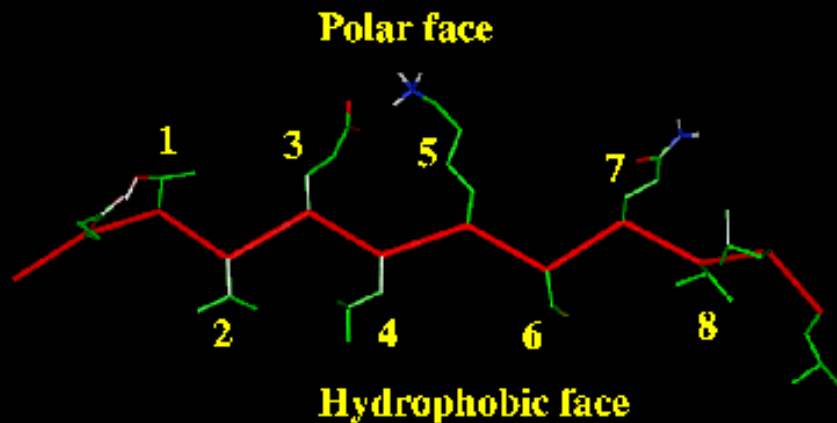
U  $\beta$  -listu se střídají rezidua po  $180^\circ$

Pro částečně zanořený  $\beta$  -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  $\beta$  -list (typicky u  $\alpha/\beta$  barelu) je tvořen nepolárními AK

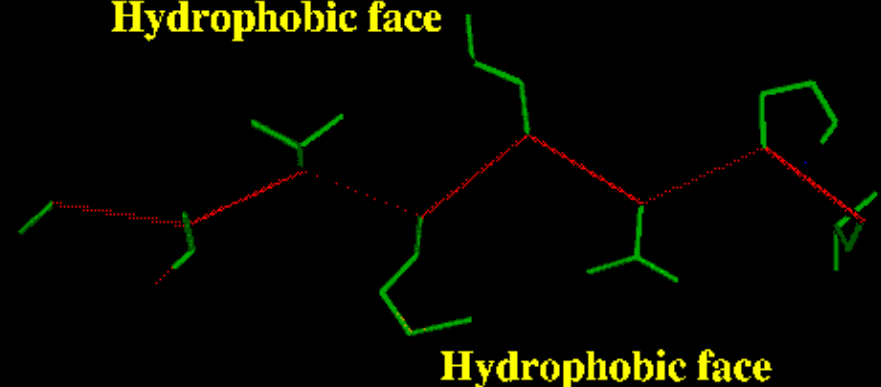
## Second strand in CD8



## 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  $\alpha$ -helixu,  $\beta$ -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	$\begin{array}{c} ^- \text{OOC} \\   \\ \text{CH} - \text{CH}_2 \\   \quad \backslash \\ \text{HN} \quad \text{CH}_2 \end{array}$		<b>POLAR BASIC</b> $\begin{array}{c} ^+ \text{NH}_3 - \text{CH}_2 - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\   \\ \text{N}^+ \text{H}_3 \end{array}$	Lysine Lys K MW = 146
Aspartic acid Asp D MW = 133	<b>POLAR ACIDIC</b> $\begin{array}{c} ^- \text{OOC} \\   \\ \text{H}_3\text{N}^+ - \text{CH} - \text{CH}_2 - \text{C}(=\text{O})\text{O}^- \end{array}$		$\begin{array}{c} \text{NH}_2 \\   \\ \text{N}^+ \text{H}_2 = \text{C} - \text{NH} - (\text{CH}_2)_3 - \text{CH} - \text{COO}^- \\   \\ \text{N}^+ \text{H}_3 \end{array}$	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  $\beta$ -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

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!

Enter the Protein Sequence (in fasta format)

```
>SEQ
MPRSVPFNISPLPQKDDSSLSASEYPNIASRTAPPSAVRRTTHSLLSEHTGYQSLESQMEAGETTSL
GKTRENHRGTPRRSYTISAIPTPDNYLRHSLTSGSLRRSRHHSRANSQSLRFRSSIDDEQDEDLPPS
AKDGMTASFLDERNWDYQFTSTDWVHDSIADGARLRELKRKDFRGRLLAAFDGAQGWILVALIGCITAA
IAYFVDVTEDFVFDLKEGFCTTRWFHNRESCCADTLDCSRWRWSQILSPSGSDNGWVDHSMFVLWVIL
SVISCYLTLFTKTVPVSSVSLTLDENLGAGTSRGTNHDASEDNSPASLINPKAHYPTISTRPAMTYISA
AGSGVAEVKVINSGFVLHGVLGFKTLVIKTIALVFSVSSGLSLGKEGYPVHIGACVGNIAACRLFSKYNDN
```

CLEAR

PREDICT

**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 MVCNISEFKWPYQPEDLNLTDALRLKYLGPQQMKQFVPICVTYLLIFVVGTLGNGLT

Peptide 2 PTNFYLFSLAVSDMLVLLVGLPLELYEMQQNYPFQLGASACYFRILLLETVCLASVLN

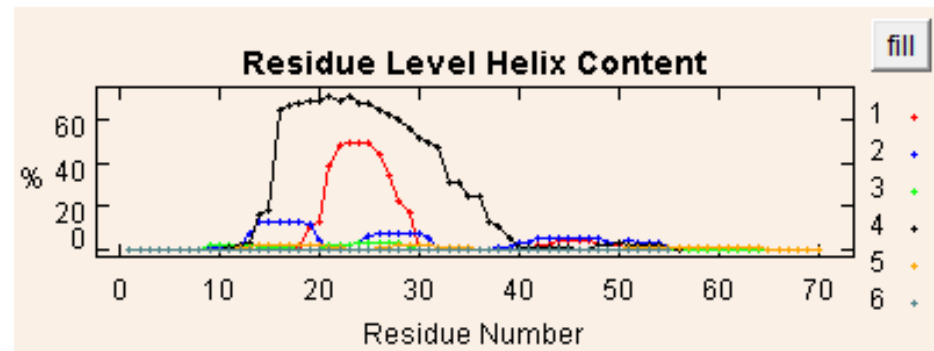
Peptide 3 VRPLQAKSVMTRAHVRRMVGAIWVLATLFLPNTSLHGLSQLTVPCRGPVPD SAIC

Peptide 4 TTALLFFCLPMVTISVLYLLIGLRLRRERMLLQVEVKGRKTAATQETSHRRIQLQDRGF

Peptide 5 VVFGICWAPFHADRIMWSLVYGHSTEGHLHAYQCVHIASGIFFYLGSAANPVLVLSLMS

Peptide 6 LGTQCCHRRQPYHGSHNHIRLTTGSTLCDVGHRSRDEPLAVNEDPGCQQQETDPS

[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](mailto: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 F 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 P 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



Sheet



Disordered



Dompred Boundary



DomSSEA Boundary





# Jpred

Predikce  
sekundární  
struktury  
proteinu

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

# Jpred 4

Incorporating Jnet

A Protein Secondary Structure Prediction Server

[Home](#) [REST API](#) [About](#) [News](#) [F.A.Q.](#) [Help & Tutorials](#) [Monitoring](#) [Contact](#) [Publications](#)

Input sequence<sup>(?)</sup>

```
MQVWPIEGIKKFETLSYLPPLTVEDLLKQIEYLLRSKWVPCLEFSKVGfVYRENHRSPGYDGRYWTMWKLPFMFGCTD  
ATQVLKELEEAKKAYPDAFVRIIGFDNVRQVQLISFIAYKPPGC
```

Advanced options (click to show/hide)

[Make Prediction](#)

[Reset Form](#)

**Primary citation:** Drozdetskiy A, Cole C, Procter J & Barton GJ. Nucl. Acids Res.  
(first published online April 16, 2015) doi: 10.1093/nar/gkv332 [link]  
More citations: link.



# Jpred

3D > 2D

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

## Jpred 4 Incorporating Jnet

A Protein Secondary Structure Prediction Server

[Home](#)[REST API](#)[About](#)[News](#)[F.A.Q.](#)[Help & Tutorials](#)[Monitoring](#)[Contact](#)[Publications](#)

### Match found in PDB

The sequence you submitted is similar to those with known structure. These may provide a more accurate secondary structure assignment than a JPred prediction.

If you still want to carry out a Jpred prediction click [continue](#)

### Hits found

Show  entries

PDB	Chain	Description	Blast E-value
3e22	C	Tubulin alpha-1C chain	0.0
3e22	A	Tubulin alpha-1C chain	0.0
3du7	C	Tubulin alpha-1C chain	0.0
3du7	A	Tubulin alpha-1C chain	0.0
1z2b	C	Tubulin alpha chain	0.0
1z2b	A	Tubulin alpha chain	0.0
5ca0	C	Tubulin alpha-1B chain	0.0
5ca0	A	Tubulin alpha-1B chain	0.0
4yj3	C	Tubulin alpha-1B chain	0.0
4yj3	A	Tubulin alpha-1B chain	0.0
4yj2	C	Tubulin alpha-1B chain	0.0
4yj2	A	Tubulin alpha-1B chain	0.0









```
UniRef90_B5RIA1 A L I G G F L I
UniRef90_Q6BT67 T L S R F M N F Q R Q E F F C V W T S M L Y R A V E T G Q F F D E Q I K E M R M
UniRef90_A7TLY2 A L T K F I D K Q R T E F F F V V W S S M R K R C V Q T A E H F N Y P I K Q M K M
UniRef90_UPI0000D56C93 S L A K H I Q A L N L P S L Q V W T S T L Q R T K A T S A H I N A P K Q H I P E
UniRef90_B0DAS3 R R A K S L E Q K G I D P L V I W T S T R R R A H H T A W P F F G P P V E K P Q
UniRef90_Q5KNM5 R L K A A V I A R R K N P L L I W T S A R R R A Y H T A W H S G Y K V V Q K P I
UniRef90_C1BMC9 E L G K Y I N G L K E N L S V W T S W M R R T L Q T A K H I S G I H E R W K T
UniRef90_B2B7C0 V M A E R L M R H R E Q P D V P I T S V S R F A T S T A D N S D K P S I H R S Q
UniRef90_Q0URM9 A M A R F V D R Q R K D W F C V W T S M L K R S I Q S A Q H F C Y E V K Q M R M
UniRef90_Q9UTE1 D L A N F M D R Q R V L W F N V W S S V R K R A I E T I E F F N Y N V K K I R L
UniRef90_Q4PFU8 E L A H F M M E Q D S K A F E V W T S Q R R R S S H T A A E A G Y K V I E R S Q
UniRef90_B6K0N8 S L T R F I N E Q R L L W E S V W S S M R R R G I E M T Q F F Y Y D V K A I R M
UniRef90_A6R990 K M T E R L Q H R L K P L M V W T S T R R R T V E T A Q F L K L G Y K M S R
UniRef90_A1D3Q5 A L T R F I D Y Q R P R N F C V W S S M M R R S V Q T V E G F N Y D V K Q M K M
UniRef90_Q6BVJ9 N L T E T L T E V D D N S F V V W T S V K T R T I E S T Q F F D I N I R H R I Q
UniRef90_Q9URZ7 L L P E Y V F S I G E K G L T V W T S S M A R T I I Q T A R H L N C O K L E W R A
UniRef90_C1HB45 A L T R F I N A E R Q R N F C V W S S M L I R S I Q T V Q Y F D Y D V K Q M R M
UniRef90_Q6C045 A L V R F M D Y Q R G E F D R
UniRef90_B8MKP4 A L A R F I D D Q R P R N F C V W S S M M K R A I Q T T S Y F N Y D I K Q M R M
UniRef90_Q1MS75 Q L A D H F M S I D I P Y I F T S T K I R S I Q T A E H N K T V C M E L T E
UniRef90_B9W719 N L T L T L F D H N S N S L E I W T S V K K R T L E T T Q F F N K N I H R V Q
UniRef90_B8J3M9 A L A R H M R H K D I N W V F T S T R L R S H E T A T H P G A H V M A F K E
UniRef90_Q24559 K L A N F I E K R C E K T A T I W T S T L Q R T I L T A S I V G F P K I Q W R A
UniRef90_A7EQN2 V L H N F I T D R D E K N F C V W T S M L K R S I Q T A E H F D Y D V K N W E M
UniRef90_B2B0M2 A Y P E L N Q E L E D K N F C V W T S M L Q R S V Q T A E Y F D Y D V K A W E M
UniRef90_Q9NKN9 A L K H F I Q D R Y V P H M V V L H A T N Y C V L S T L A P L M E G A T Q E E D
UniRef90_Q582Y3 A L K R F L L E R D A P N M V V L H G T N P N V M N T L R P L E Q A L E P D
UniRef90_B7FQZ7 R N K A N H E N E G P R P M K I L T S T M P R A A D T V N W T D Y E F Q Q M S N
UniRef90_B5YM34 A L F G F M K E E F T P P C H I M T S T M P R A R Q T W E D M P Y P I E V L S N
UniRef90_B8DLZ8 E L A A H F A D I E V P Y I F T S T R R R S A Q T A A P M R E A R P Q L P E
UniRef90_Q9UTK9 R I R K S F P D V S F K N L H V L S C M E D S V M S P F R E L G S V T S S M S S
UniRef90_C1STW1 R L A A T F R D T P L S M V F T S T R R R T R Y V L E A Q Q T A P P L Y S M E L
UniRef90_B2W304 A M T R F I H E L E A H N F C V W T S M L K R S I Q T G Q Y F C F E V K Q M R M
UniRef90_A4HEN8 V L F E Y F H K E I P T S L T V M T S C A K R C T Q T V H Y F A E E S I K H P T
UniRef90_C1FHQ5 I L A R F A A E H G A S P S R L W T S S M L R T I Q T A A L I P H P R V Y R N
UniRef90_Q4DA36 E I L G F L A S E K P V K V E V W T S Q L R R A I Q T V E N S E P K N V R W S S
UniRef90_A4HTA2 A L L E F L D S L K R H L E H V W T S Q L R A A R Q T T E L S N I R T L R W S S
UniRef90_A8JG80 K T R D M L S A V P F S A V F Q S P L A R A R Q T A D S A V F Q R V T L P C
UniRef90_B2VW18 E A G R Q L R S L L K P D D D Q F F T S P Y R R T R E T T E R N K I T V F E E P R
UniRef90_Q1E1L6 E A G R R L R A M L R P D D H F F T S P Y R R T R E T T E R N T I K V Y E E P R
UniRef90_Q5B397 D A G T R L R G L L R P D D H F F T S P Y R R T R E T T E R H T I K V Y E E P R
UniRef90_Q6C8W1 D C G K K I R E V N G P I S A V Y T S D L K R C V E T T D L L L A T A G I D P L
UniRef90_Q6FN95 N A G Q L I A Q H G I T L P Q I F T S R L L R T E Q T M D S N E I P V L R T W R
```

lupas\_21  
lupas\_14  
lupas\_28  
jnet  
conf  
sol25  
sol5  
sol0  
jhmm  
jpssm



# Úloha

- Seznamte se s programem **Jpred 4** (<http://www.compbio.dundee.ac.uk/jpred/>). Pro následující proteinové sekvence predikujte zastoupení sekundárních struktur. Patří tyto proteiny do skupiny tzv.  $\alpha/\beta$  proteinů, tj. proteinů, ve kterých jsou zastoupeny obě typické sekundární struktury ( $\alpha$ -helix,  $\beta$ -skládaný list)?

# Úloha

- Protein A:

GTYRQLFHPEQLISGKEDAANNYARGHYTIGKEIIDLVLDVRKLDQCTGLQGFLVFHSFGGGTGSFGFTSLLME  
RLSVDYGGKSKLEFSIYPAPQVSTAVVEPYNSILTTHTTLEHSDCAFMDNEAIYDICRRNLDIERPTYTNLRLMS  
QIVSSITASLRFDGALNVDLTFQTNLVPYPRIHFPLATYAPVISA EKAYHEQLSVAEITNACFEPANQMVKCDPRH  
GKYMACCLLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYQPPTVPPGGDLAKVQRAVCMLSNTT  
AVAEAWARLDHKFDLMYAKRAVHWYVGEEMEEGEFSEAREDMAALEKDYEEVGADSYEDEDEGE EY

- Protein B:

SIRLPAHLRLQPIYWSRDDVAQWLKWAENEFSLRPIDSNTFEMNGKALLLLTKEDFRYRSPHSGDVLYELLQHIL  
KQAGPNIFEMLRIDEGLRLKIYKDTEGYTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLCQDVDA  
AVRGILRNAKLKPVYDSLDCVRRRAALINMVFMGETGVAGFTNSLRMLQQKRWDEAAVNLA KSRWYNQTP  
NRAKRVITTFRTGTW DAYK

- Protein C:

ADVCMDPEPIVRIVGRNGLCVDVRDGRFHNGNAIQLWPCKSNTDANQLWTLKRDNTIRSNGKCLTTYGYSPG  
VYVMIYDCNTAATDATTRWQIWDNGTIINPRSSLVLAATSGNSGTTTLVQTNIYAVSQGWLPTNNTQPFVTTIVG  
LYGLCLQANSQVWIEDCSSEKAEQQWALYADGSIRPQQNRDNCLTSDSNIRETVVKILSCGPASSGQRWFMFK  
NDGTILNLYSGLVLDVRASDPSLQIILYPLHGDPNQIWLPLF



# 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](http://www.ch.embnet.org/software/TMPRED_form.html)

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

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](#) 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 17"/> <input type="text" value="maximum 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 formats)	<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 unknown

[EMBnet-Server] Date: Wed Mar 9 8:44:54 2016

tmmpred -par=matrix.tab -html -min=17 -max=33 -def -in=wwwtmp/TMPRED.26952.7025.seq -out=wwwtmp/TMPRED.26952.7025.out -out2=wwwtmp/TMPRED.26952.7025.out  
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
<b>S</b> 317 ( 317)	<b>S</b> 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
<b>S</b> 272 ( 272)	<b>S</b> 290 ( 290)	1692	282
<b>S</b> 310 ( 318)	<b>S</b> 337 ( 337)	1576	327

## 2.) Table of correspondences

Here is shown, which of the inside->outside helices correspond to which of the outside->inside helices.

Helices shown in brackets are considered insignificant.  
A "+"-symbol indicates a preference of this orientation.  
A "++"-symbol indicates a strong preference of this orientation.

inside->outside	outside->inside
40- 59 (20) 1635	37- 62 (26) 1983 ++
74- 96 (23) 1553 ++	74- 95 (22) 1269
112- 135 (24) 830	115- 134 (20) 1218 ++
158- 176 (19) 1652 ++	158- 179 (22) 1329
207- 227 (21) 2501	207- 229 (23) 2638 +
275- 293 (19) 2207 ++	272- 290 (19) 1692

# GOR4

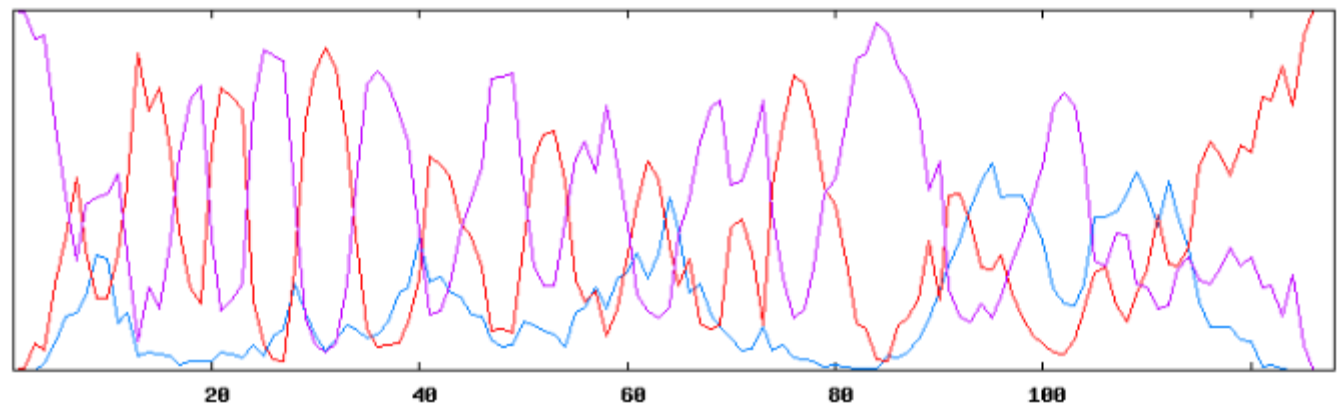
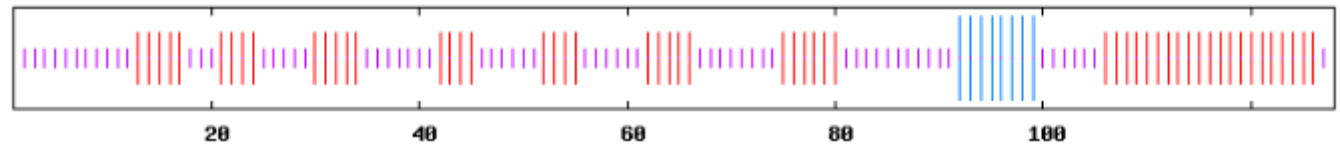
View GOR4 in: [\[AnTheProt \(PC\)\]](#), [\[Download...\]](#) [\[HELP\]](#)

```
          10          20          30          40          50          60          70
          |          |          |          |          |          |          |
MTPPTSSQLSKKTRWSYANGCTWSVKDDDHVLFIEGSGTSGMLRFKTFSGDFFTLVPGIHNLVSLVHLQG
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
DVYAPGVVSYGLNCSPLGSLHVVASIQQLSEKNSTIASSFYRLRHFFFTTKATSSGRS
ccccccccccccccccccccchhhhhhhhhccccccccccccccccccccccccccccccccccccccc
```

Sequence length : 128

GOR4 :

Alpha helix	(Hh)	:	8 is	6.25%
3 <sub>10</sub> 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):

CCCCCCHHHCHHCHEEEETTCEEEECTTEEEEEEECSCEEEEEEECSCEEEETTCCEEEEEE

H: alpha-helix

G:  $3_{10}$ -helix

I:  $\pi$ -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/tool/chou-fasman/>).  
Tutéž sekvenci analyzujte pomocí programu **GOR IV** ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_gor4.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html)) a výsledky porovnejte.

Sekvence:

```
MSTDKPLVIQSDKTLLLDVHSPFAQECRDSITAFSELVK SPEHVHTFLLTPLSLWNANAAG  
MTTEDIMGRLRTWSRYDIPEPVSYFITDISARFGSFVMTDIPDDADHYLLTVTIPRYAKEI  
SSHKT VSSLLFPRGNDTFLLNRYARGEVKLKLKLGFPVDDRIPLKKGFPVPMNLRQQTLS  
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