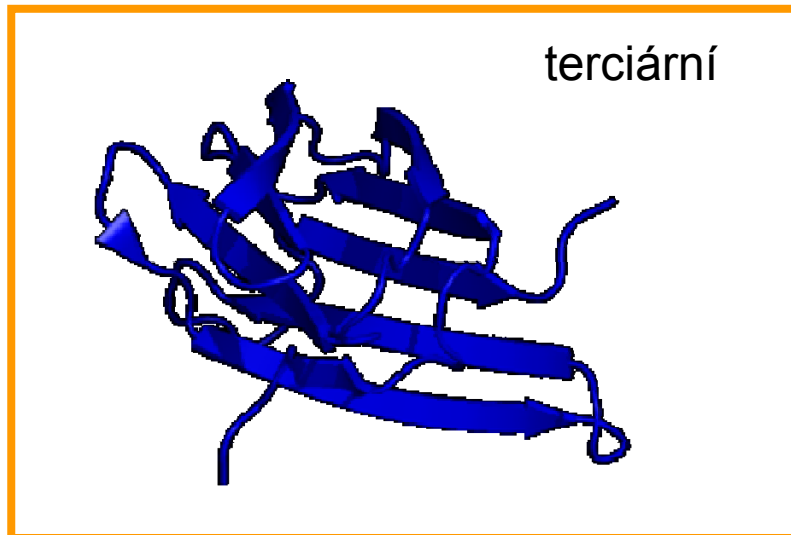
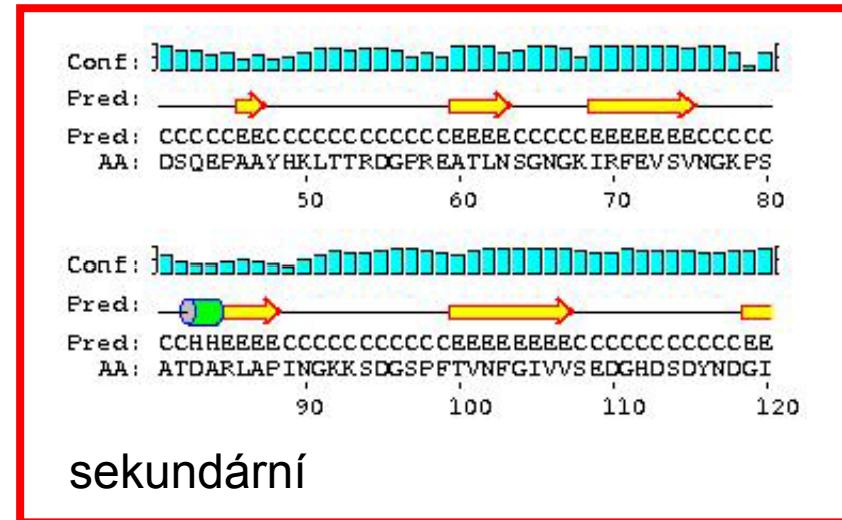


# Predikce sekundárních struktur proteinů

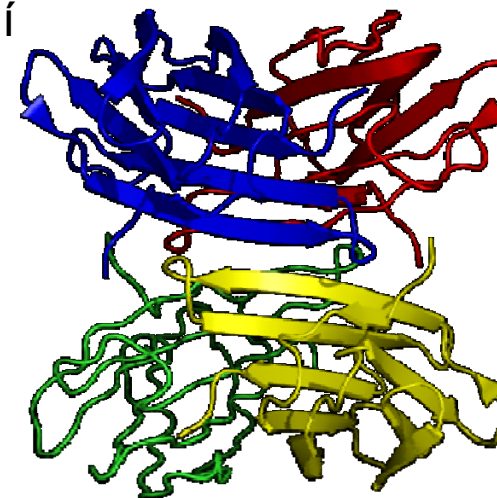
# Struktura proteinů

ADSQTSSNRAGEFSIPPNTDFRAIF  
FANAAEQQHILKFIGDSQEPAAYHK  
LTTRDGPREATLNSGNGKIRFEVSV  
NGKPSATDARLAPINGKKSDGSPF  
TVNFGIVVSEDGHDSYNDGIVVL  
QWPIG

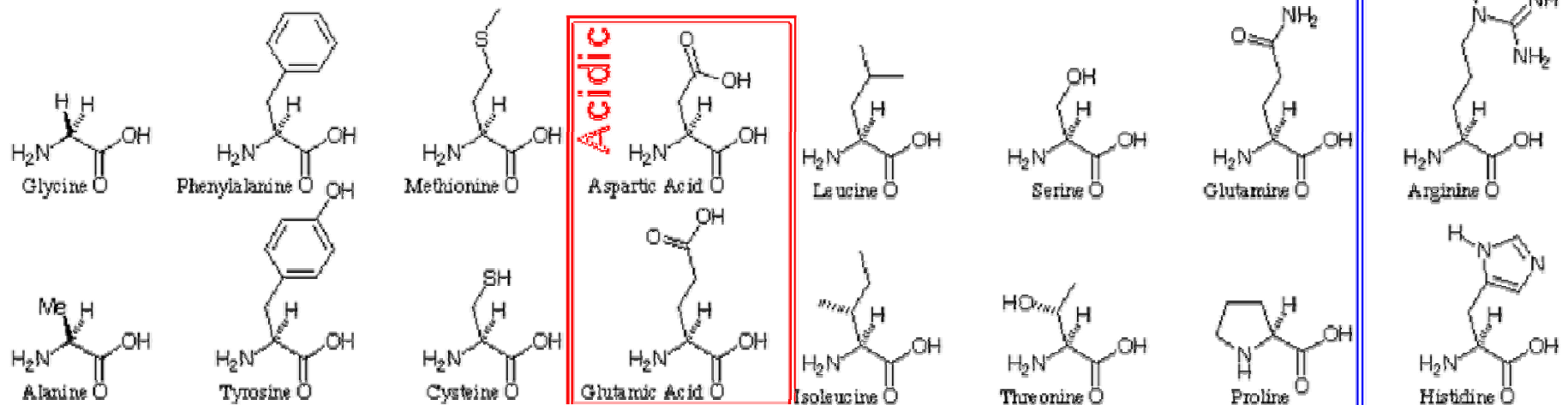
primární  
(sekvence)



kvartérní



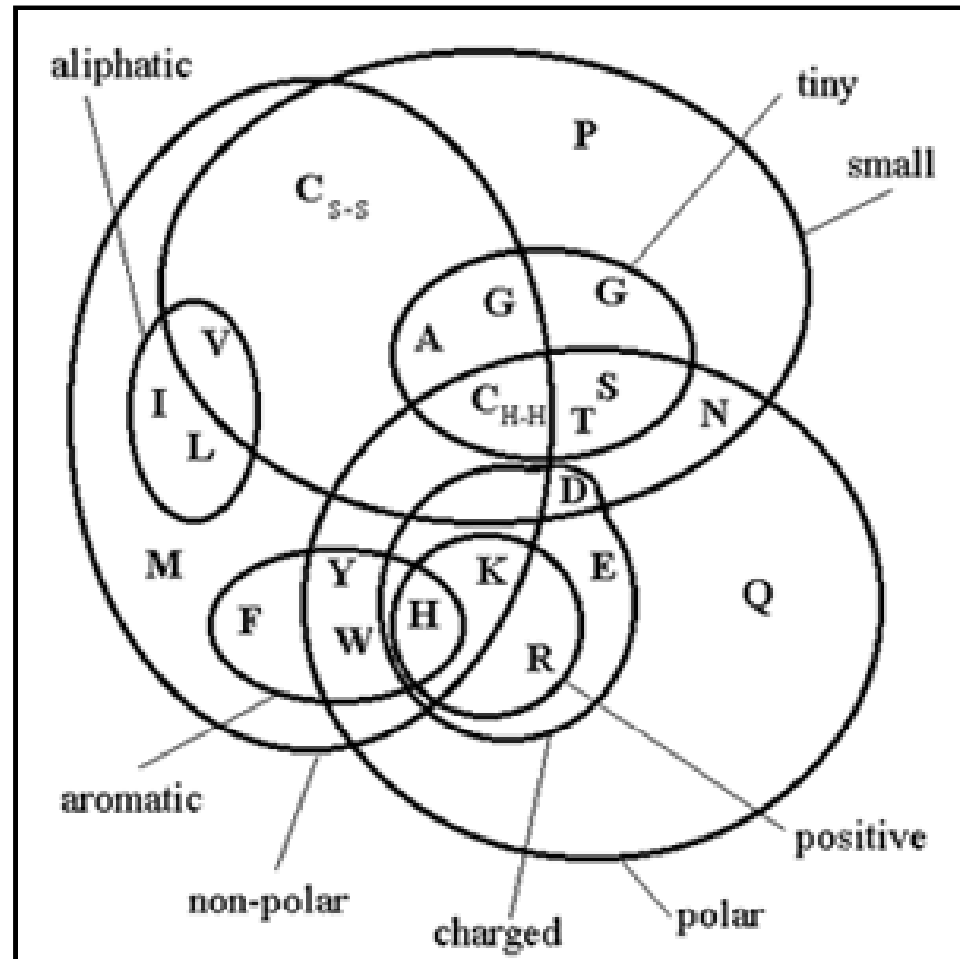
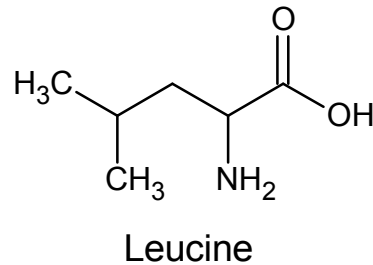
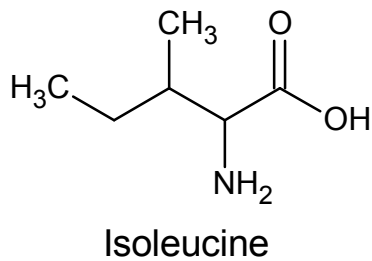
# Aminokyseliny



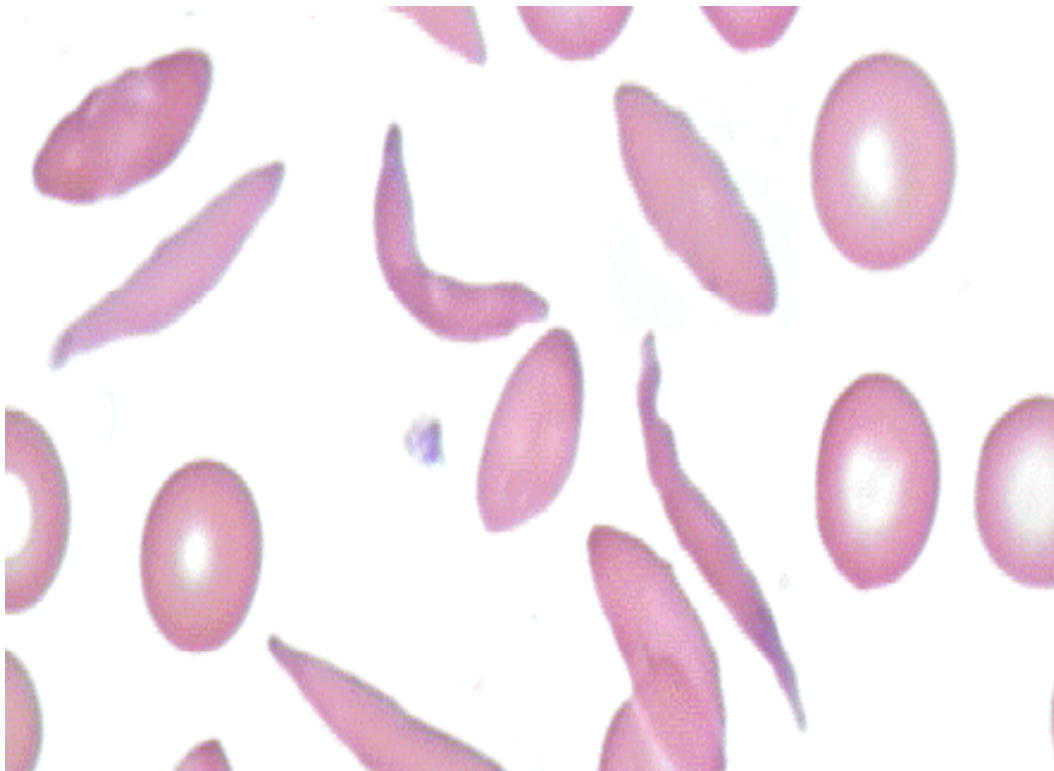
glycin	alanin	valin	leucin	izoleucin	asparagová kys.	asparagin	glutamová kys.	glutamin	arginin	lysin	histidin	fenylalanin	serin	threonin	tyrozin	tryptofan	methionin	cystein	prolin	selenocystein
Gly	Ala	Val	Leu	Ile	Asp	Asn	Glu	Gln	Arg	Lys	His	Phe	Ser	Thr	Tyr	Trp	Met	Cys	Pro	Sec
G	A	V	L	I	D	N	E	Q	R	K	H	F	S	T	Y	W	M	C	P	U

# Třídění aminokyselin

Aminokyseliny s podobnými vlastnostmi mohou plnit v proteinu stejné funkce – bývají vzájemně zastupitelné



# Hemoglobin S - vliv jediné substituce

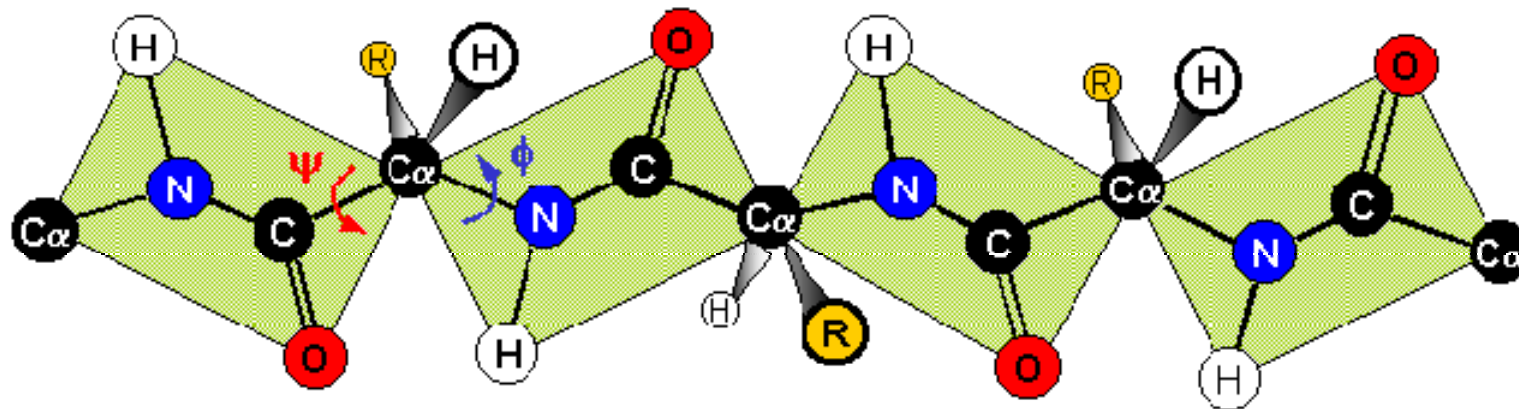


## HbS (Glu6Val)

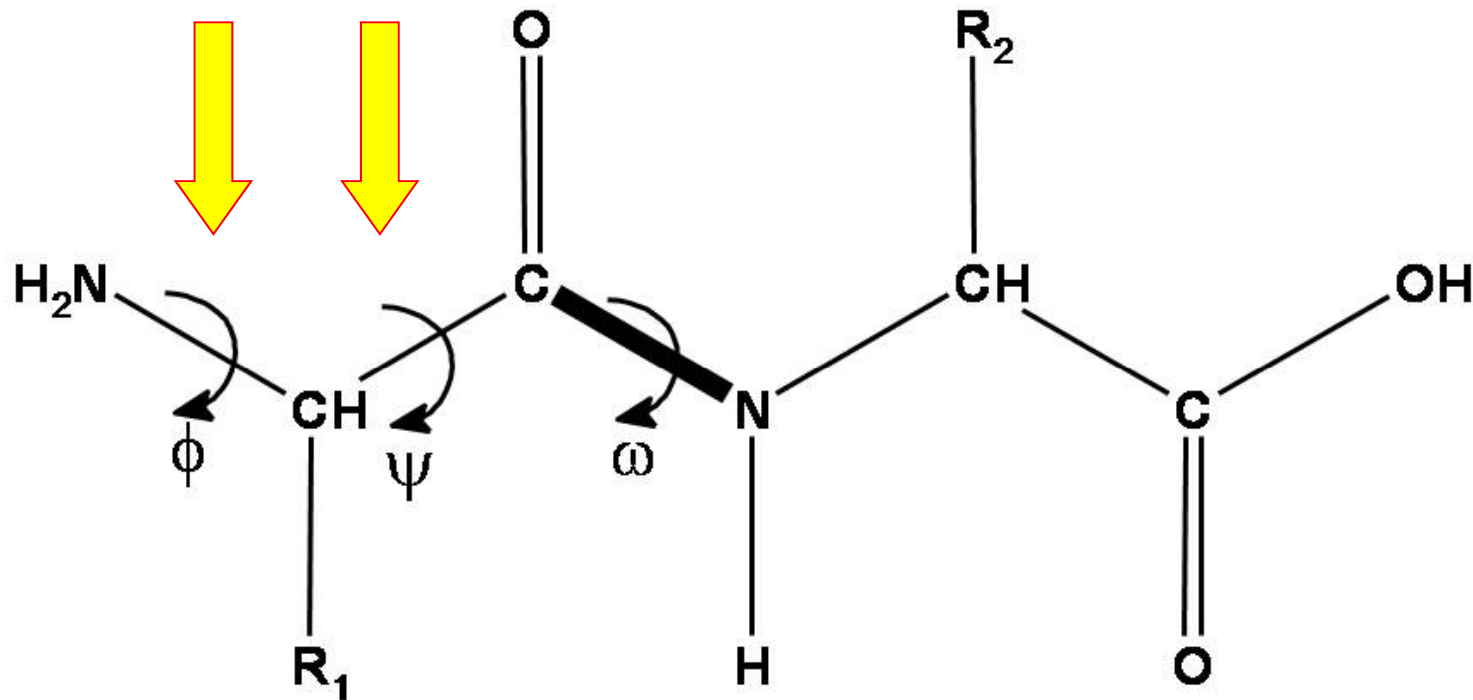
- deoxyHb je 50x méně rozpustný než oxyHb
- dlouhé asociované řetězce

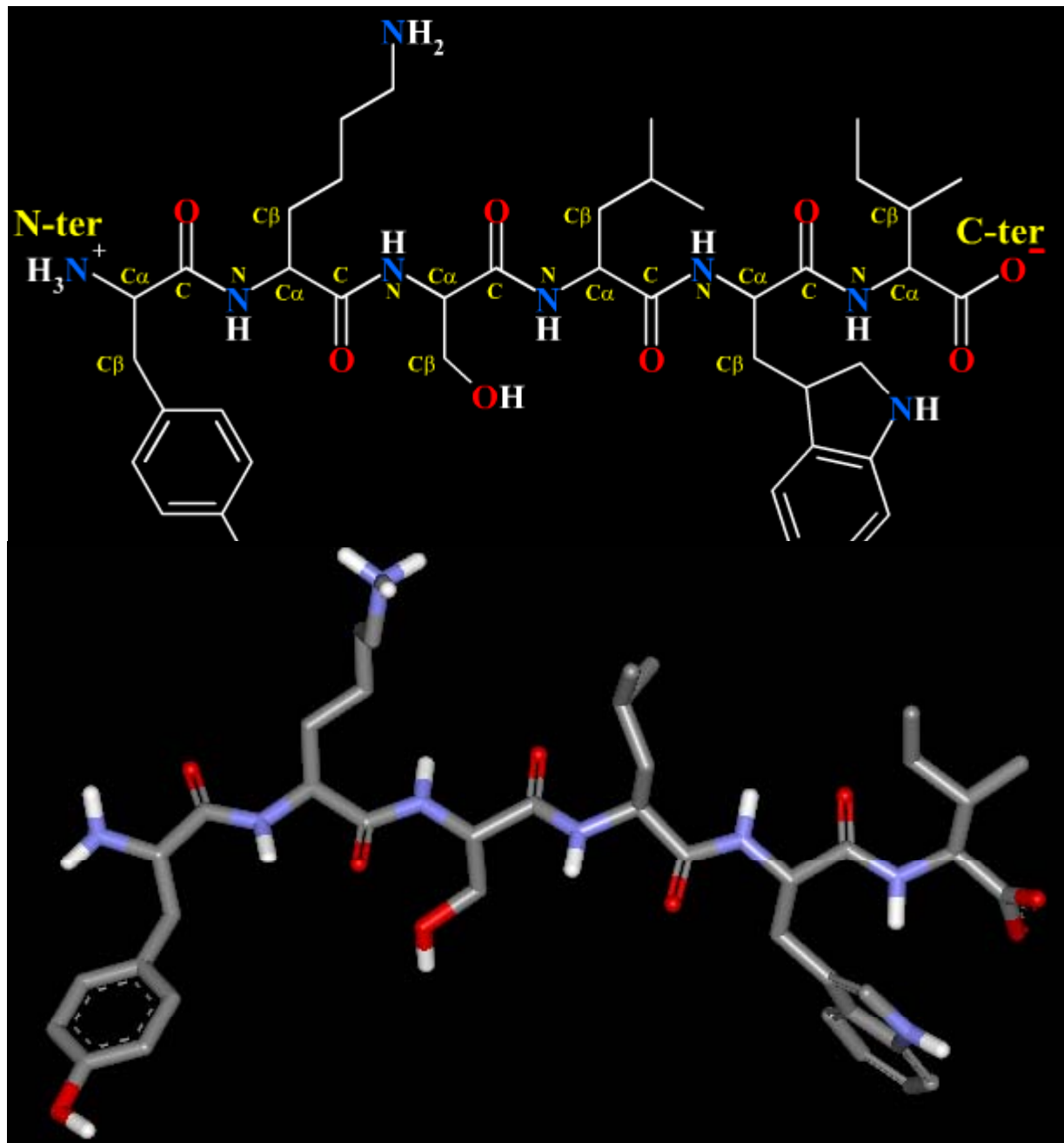
# Kostra polypeptidového řetězce

Peptidová vazba – planární

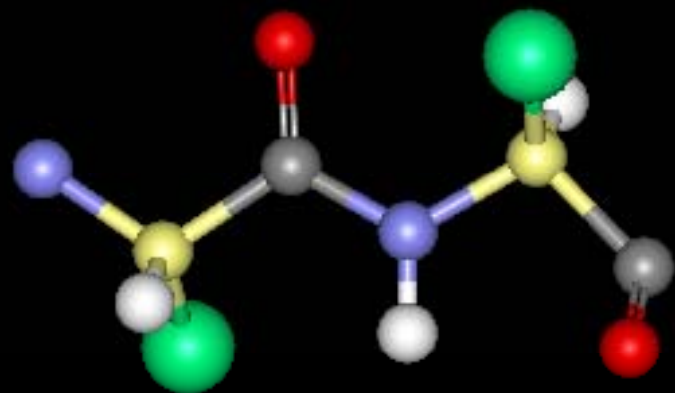


- Konformaci kostry určují dva torzní úhly  $\phi$  a  $\psi$  (úhel  $\omega$  je  $180^\circ$ )

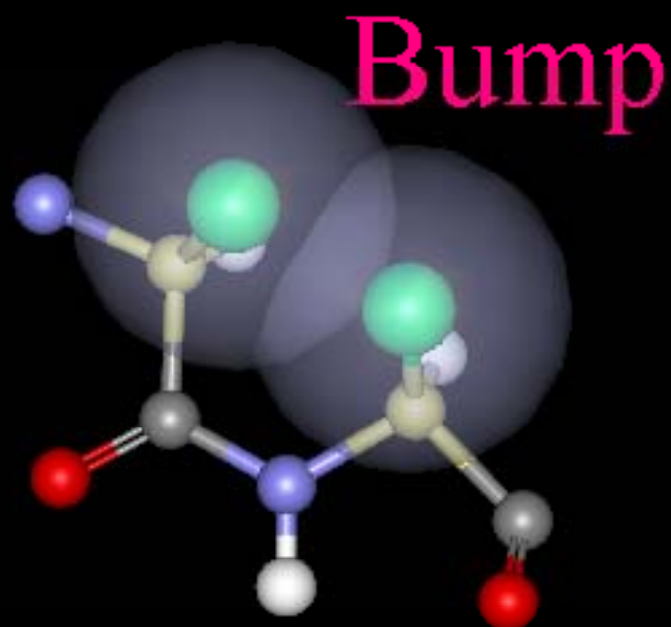






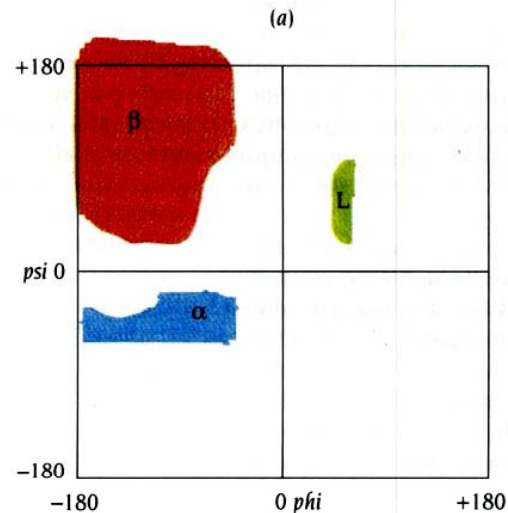


s-trans

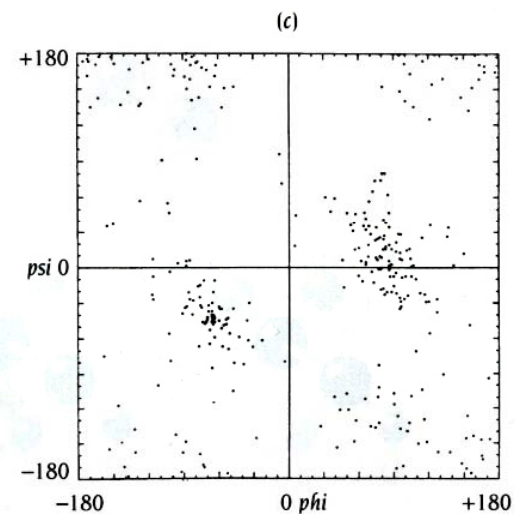
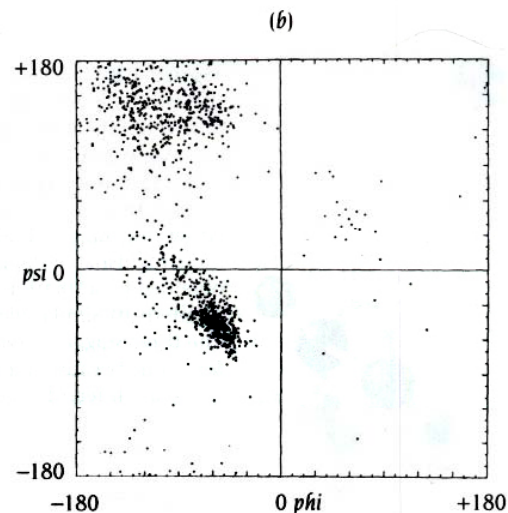


s-cis

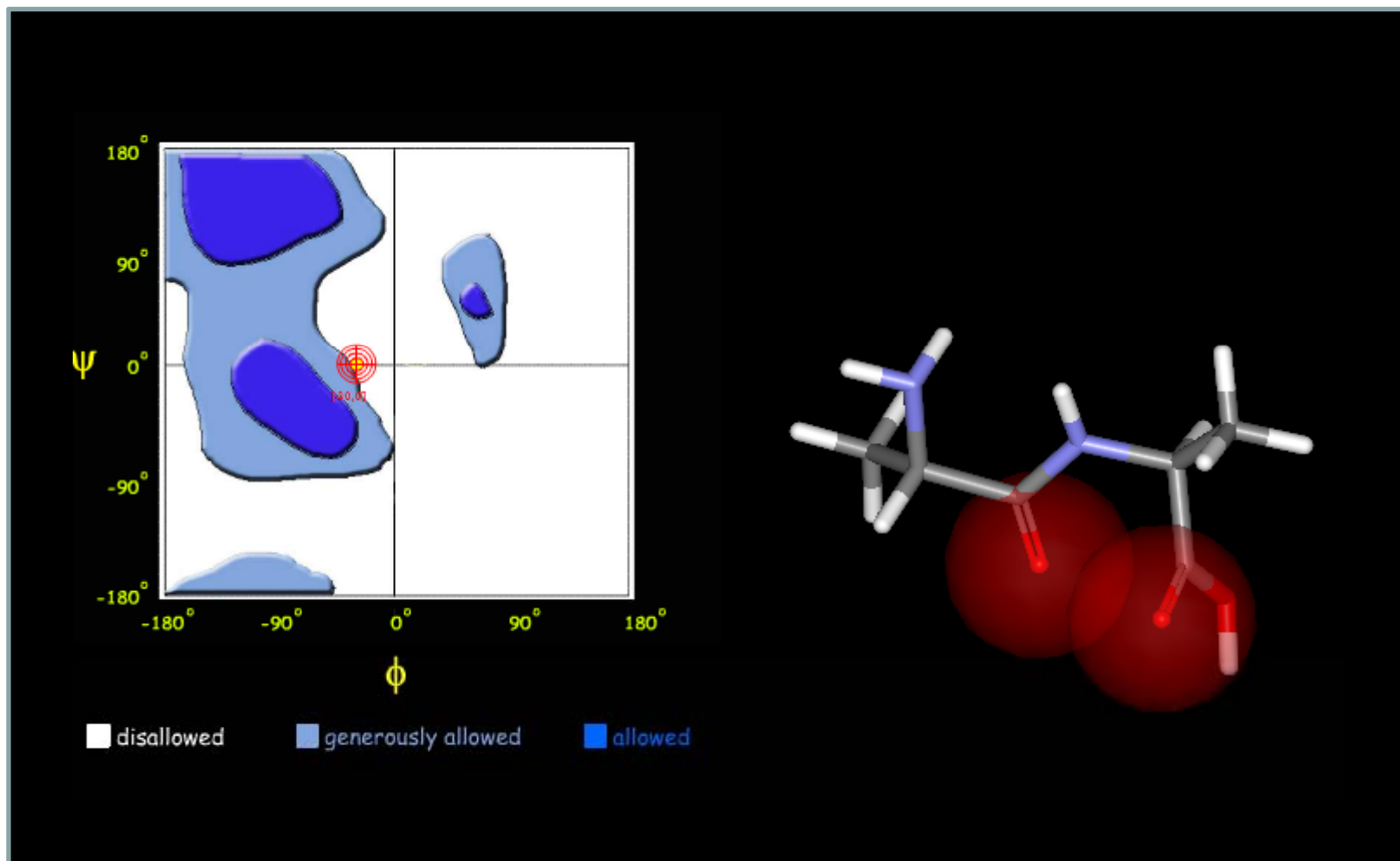
# Ramachandran diagram



- 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

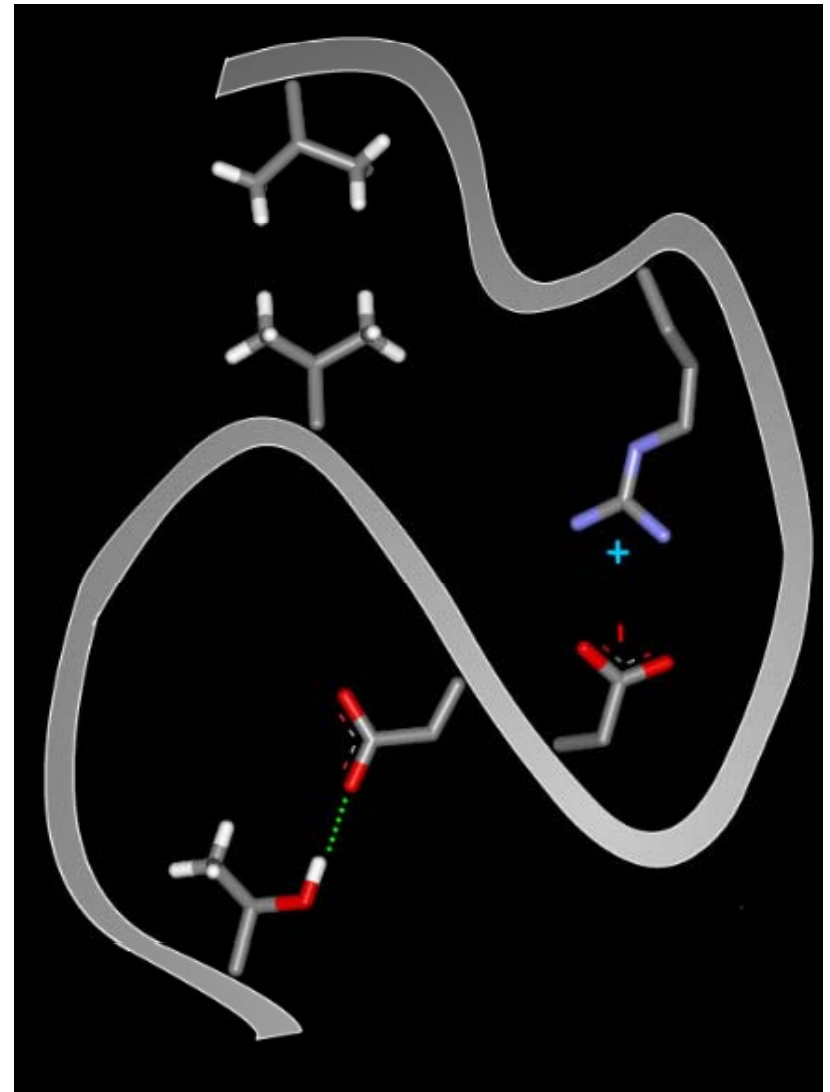


# Ramachandranův diagram



# Vazby zprostředkovávající vyšší struktury

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



# 2-D struktury

- Stabilní konformace polypeptidového řetězce
- Důležité pro udržení proteinové 3-D struktury
- Cca 50 % aa residuí je součástí  $\alpha$ -helixů nebo  $\beta$ -skládaných listů
- Predikce sekundárních struktur znamená předpověď zda residuum spadá mezi H (helix), E (list) nebo C (smyčka)

# Predikce 2D struktury

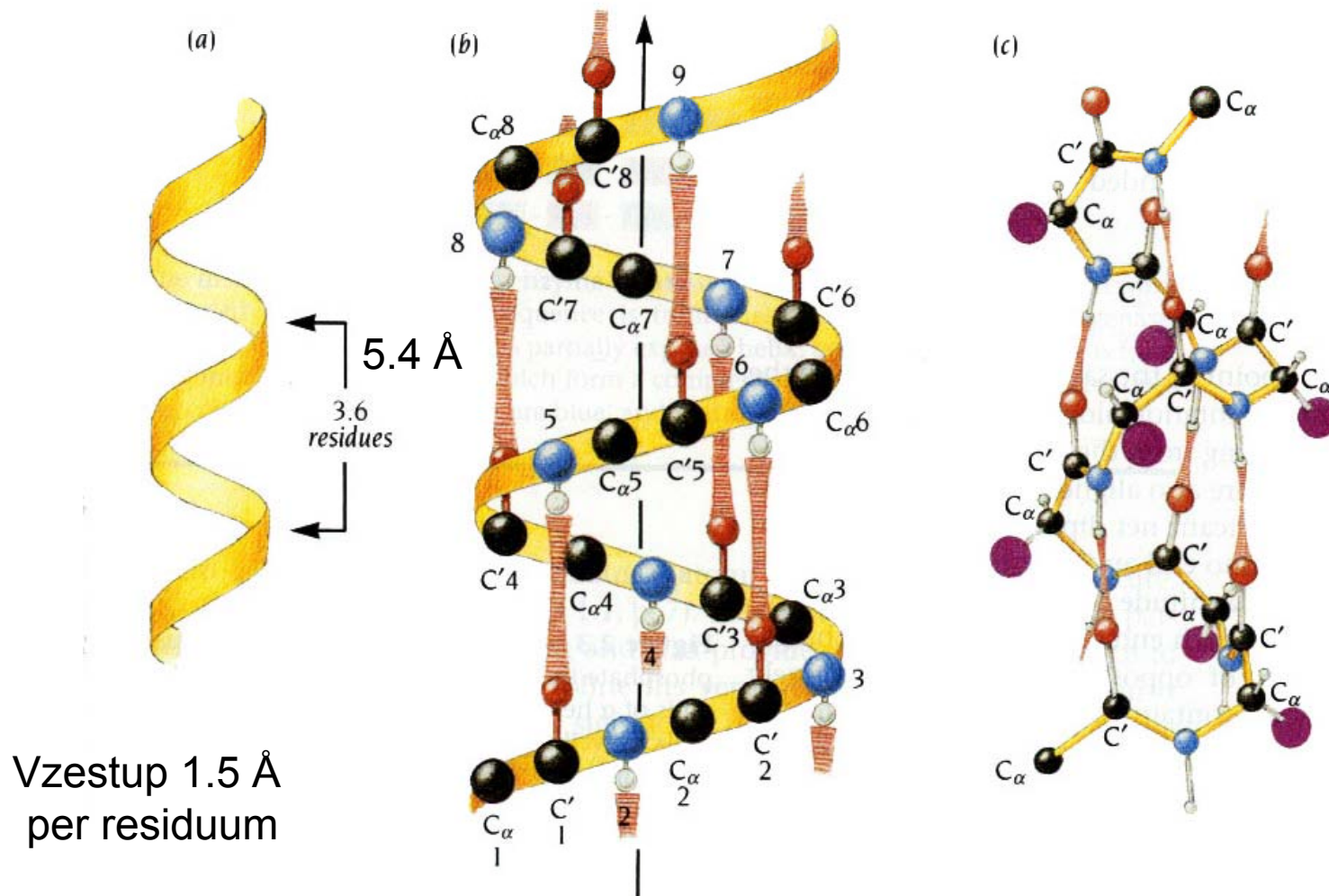
- Důležité pro klasifikaci proteinů
- Separace domén a funkčních motivů
- SS jsou mnohem konzervovanější než aminokyselinová sekvence
- Předpověď SS předchází obvykle jako mezikrok při předpovědi terciární struktury při threadingových metodách.

# Predikce 2D struktury

- Rozlišujeme tři základní typy
  - H – helix
  - E –  $\beta$ -list
  - C/(-) – smyčka/náhodné klubko (coil) – někdy jsou rozlišovány tyto dvě varianty
- S dobrou přesností lze určit helix (jejich tvorba je určena interakcemi „krátkého“ dosahu), u  $\beta$ -listu (interakce „dlouhého“ dosahu) 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)

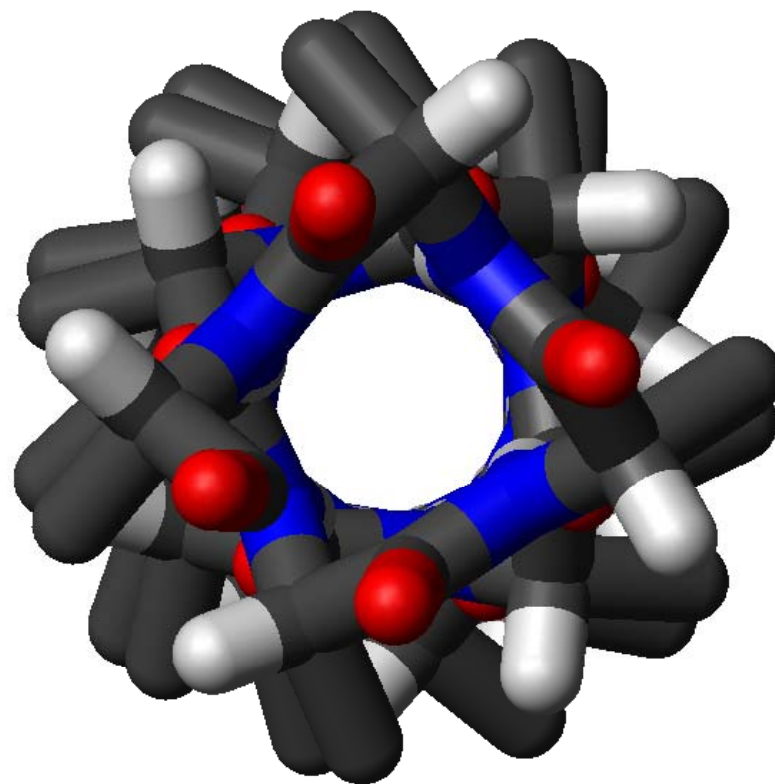
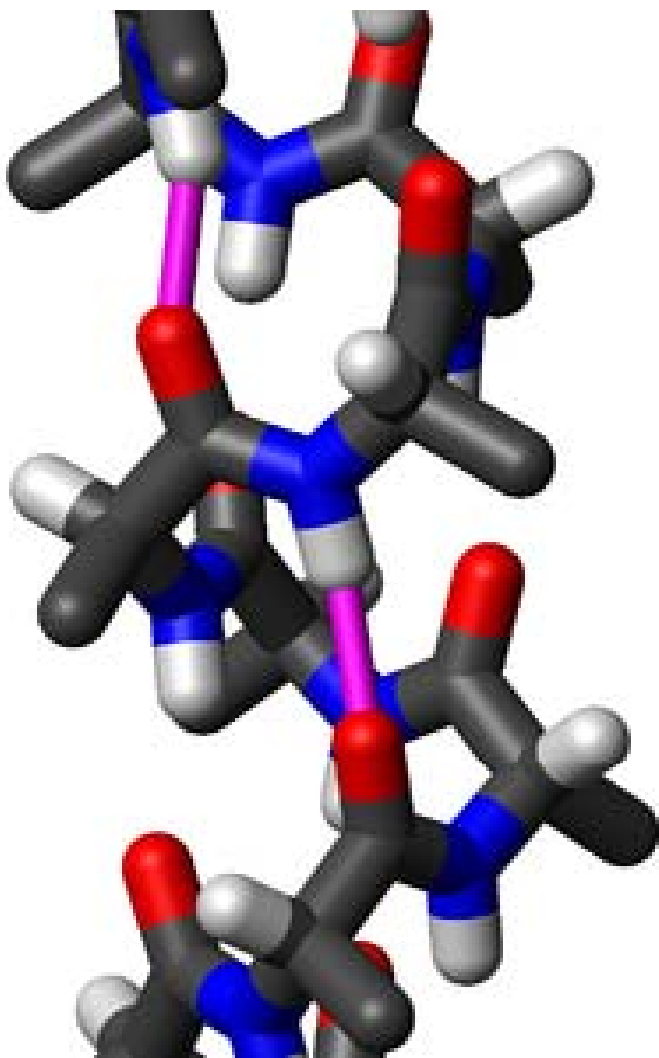
# $\alpha$ -helix

stabilizace mezi rezidui  $i+4$  a  $i-4$  H-vazbami



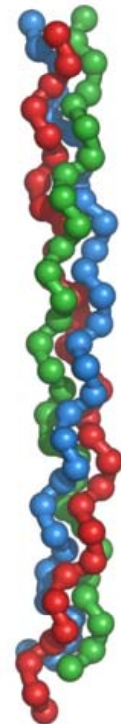


# Boční a horní pohled

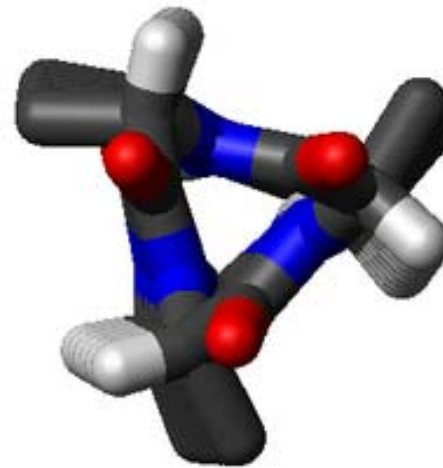
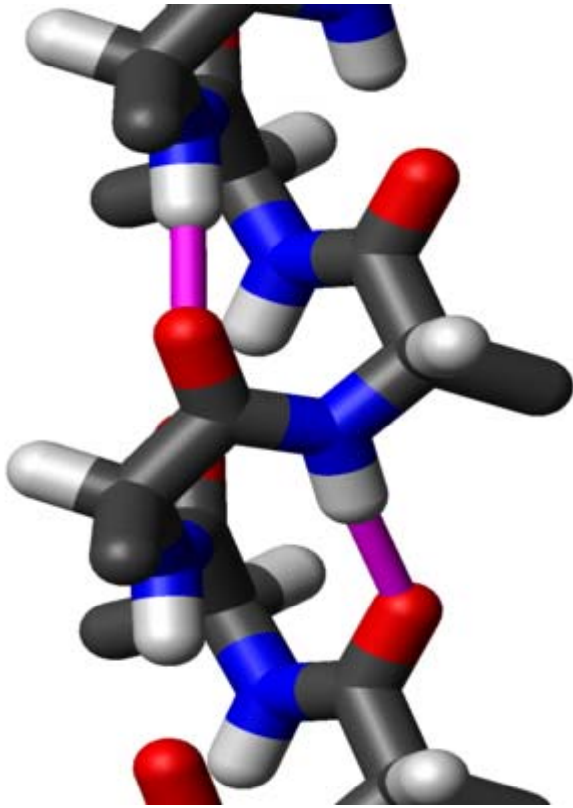


# Jiné helixové struktury

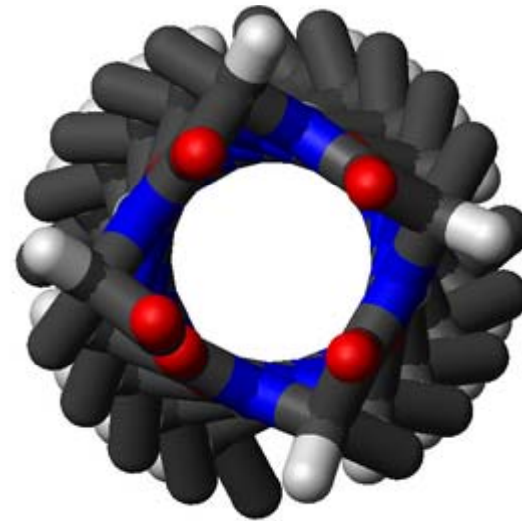
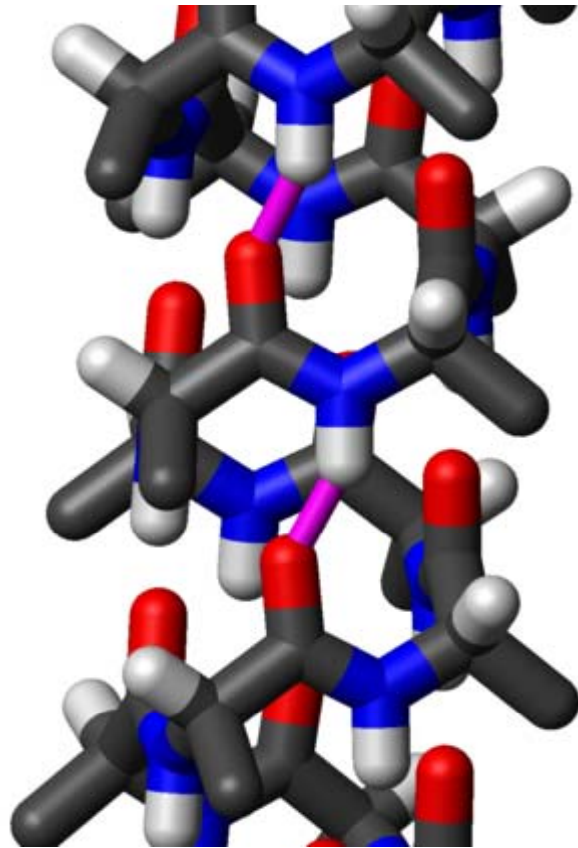
- $3_{10}$  helix – obvykle na začátku nebo konci  $\alpha$ -helixu
- $\pi$ -helix – zřídka, považován za málo stabilní
- kolagen – levotočivá šroubovice



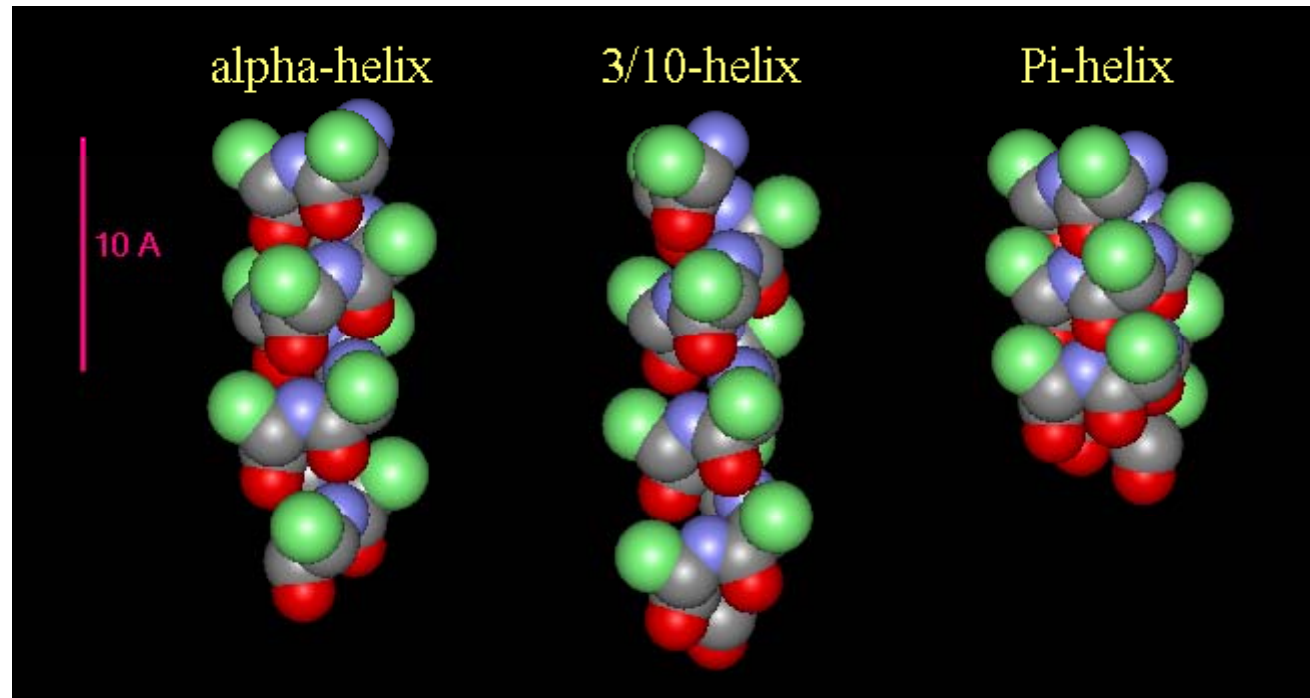
# $3_{10}$ helix



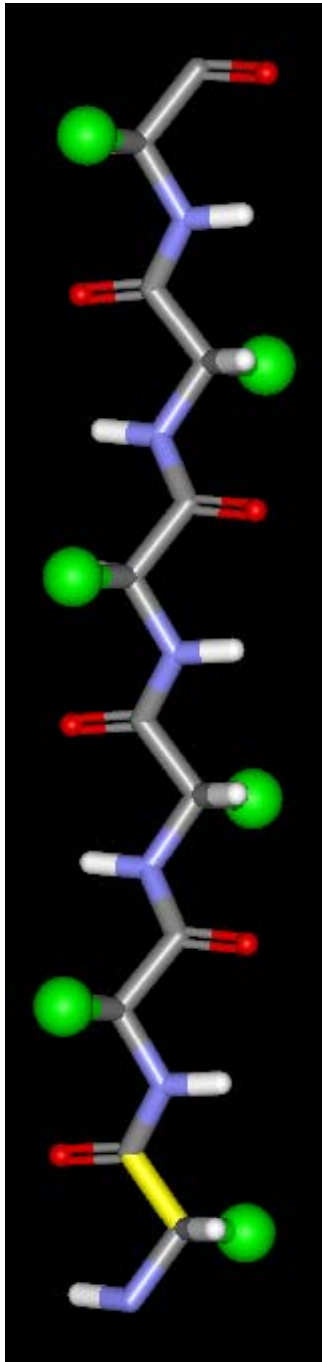
# $\pi$ -helix



# Porovnání 16 aa v typických helixech



	$\alpha$ -helix	$3_{10}$ -helix	$\pi$ -helix
Vodíkové můstky	$0_i$ k $N_{i+4}$	$0_i$ k $N_{i+3}$	$0_i$ k $N_{i+5}$
Počet residuí na otáčku	3.6	3	4.4
Počet atomů na otočku	13	10	16
Vinutí (na 1 aa)	1.5	2.0	1.15



# $\beta$ -list

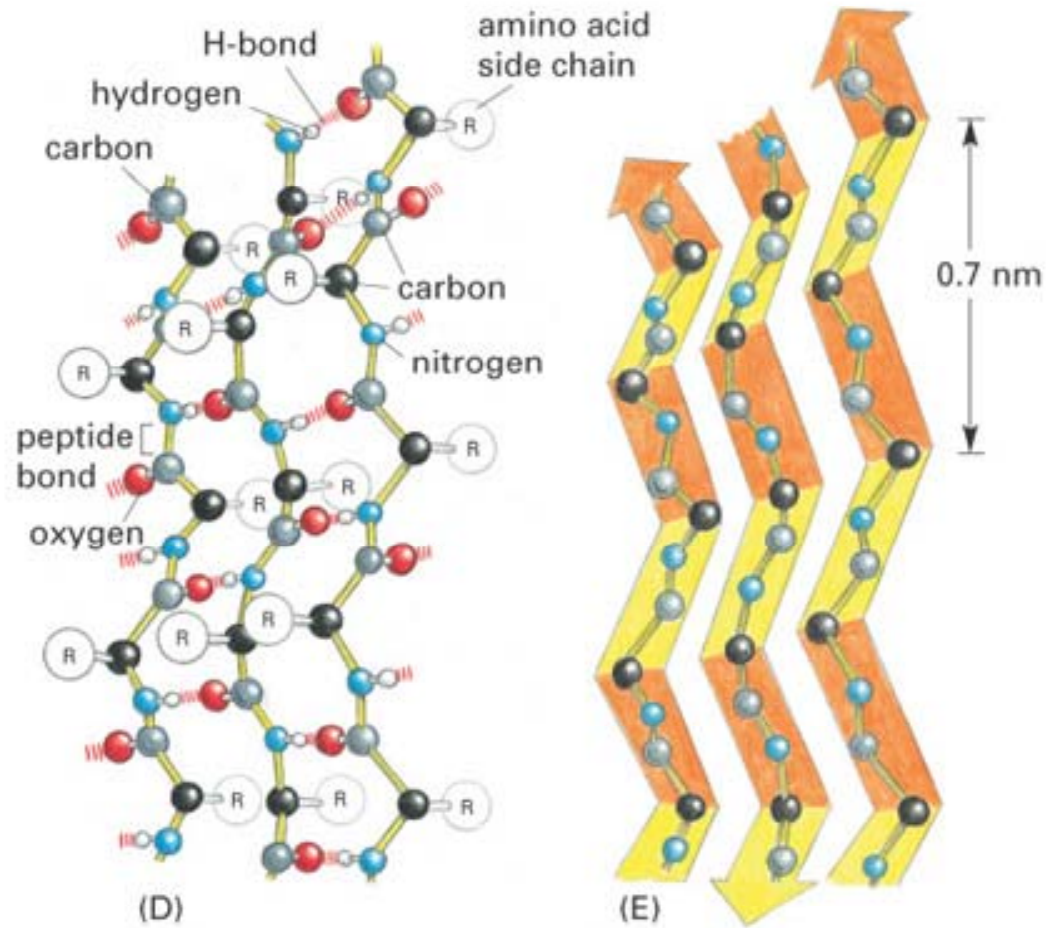
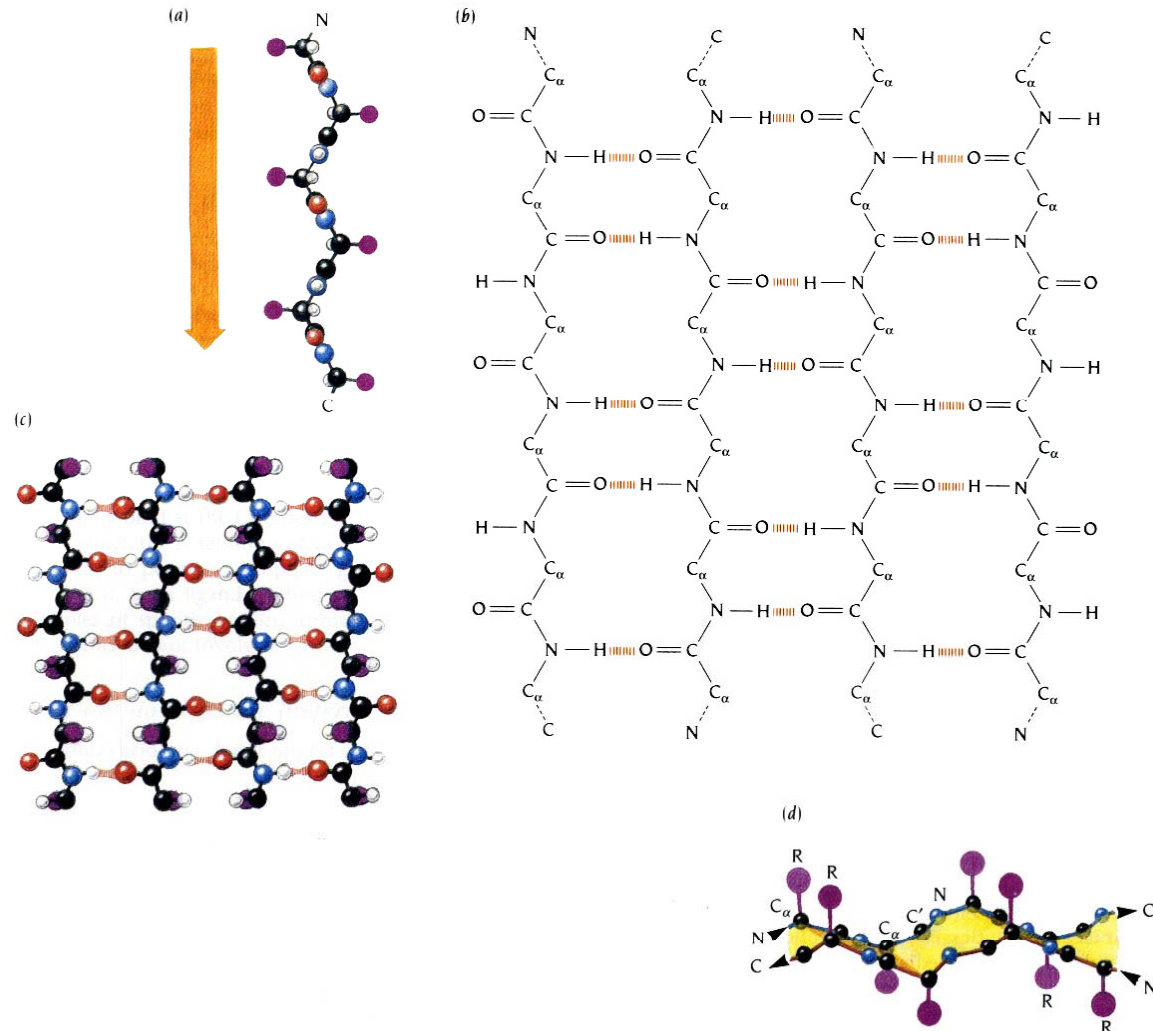


Figure 4-10 part 2 of 2 Essential Cell Biology, 2/e. (© 2004 Garland Science)

# Antiparalelní $\beta$ -list

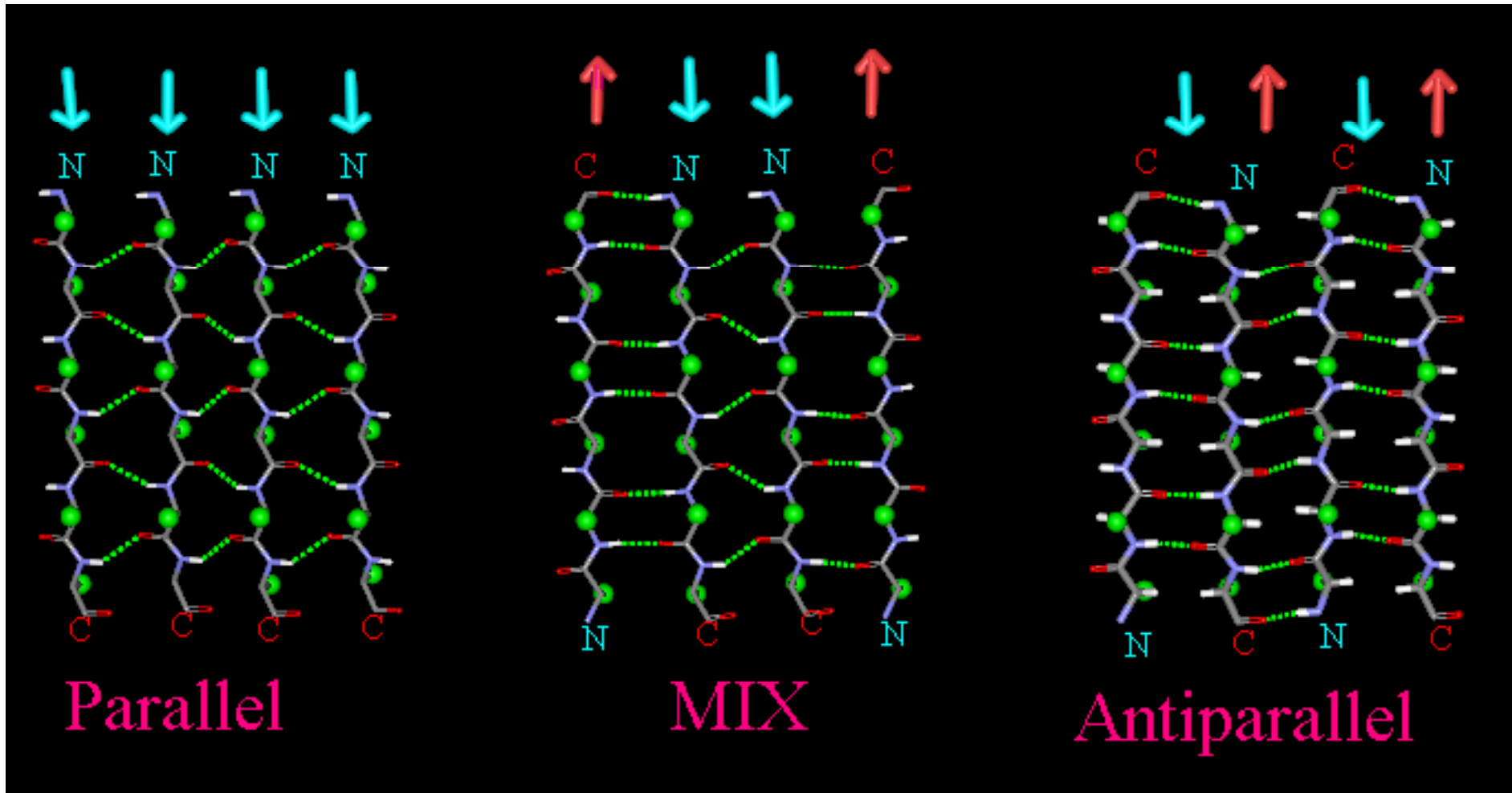








# Výskyt $\beta$ -listů v proteinech



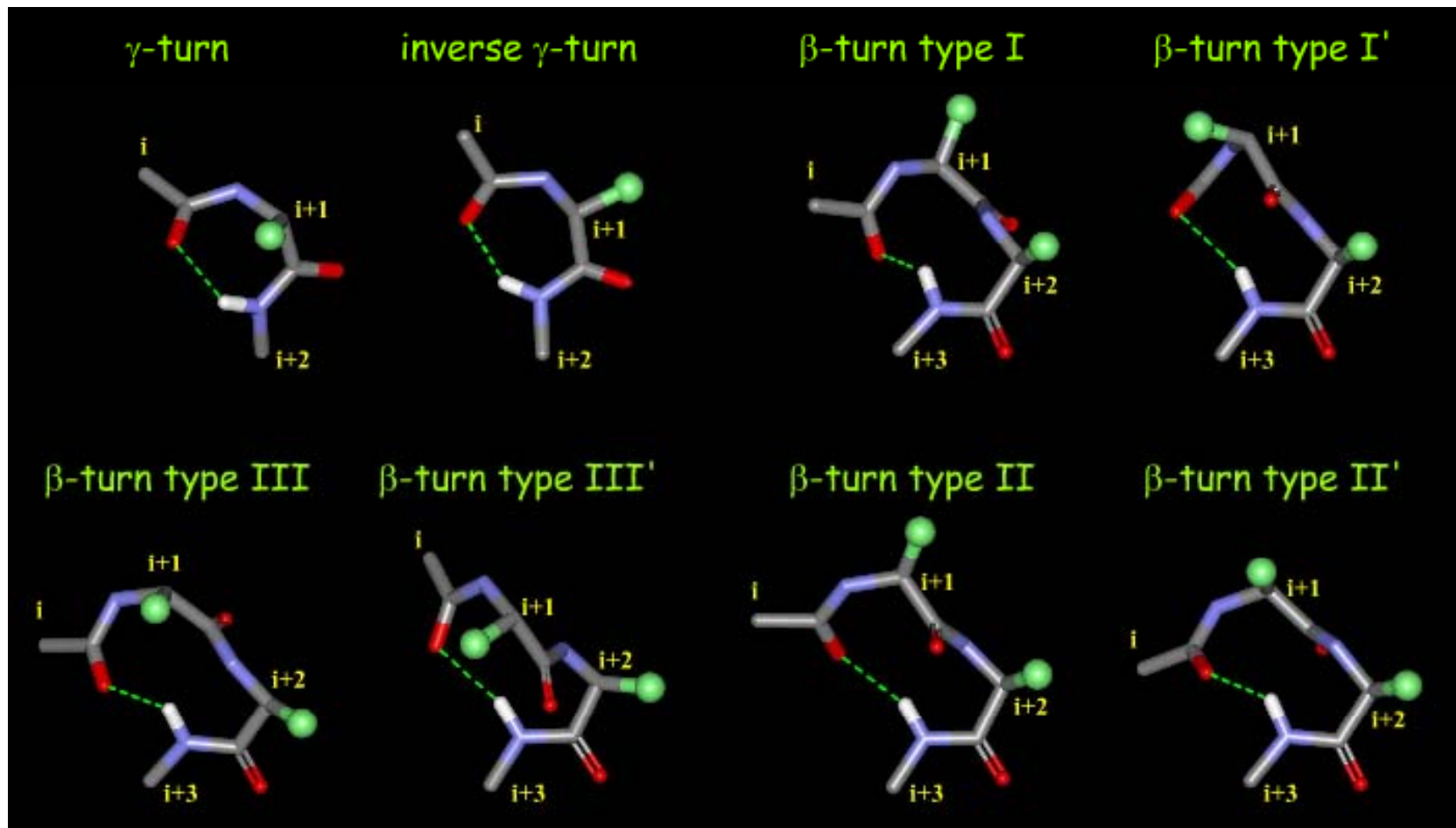
# Pootočení $\beta$ -listů v realitě

Torze řetězce  $0^\circ - 30^\circ$  na reziduum



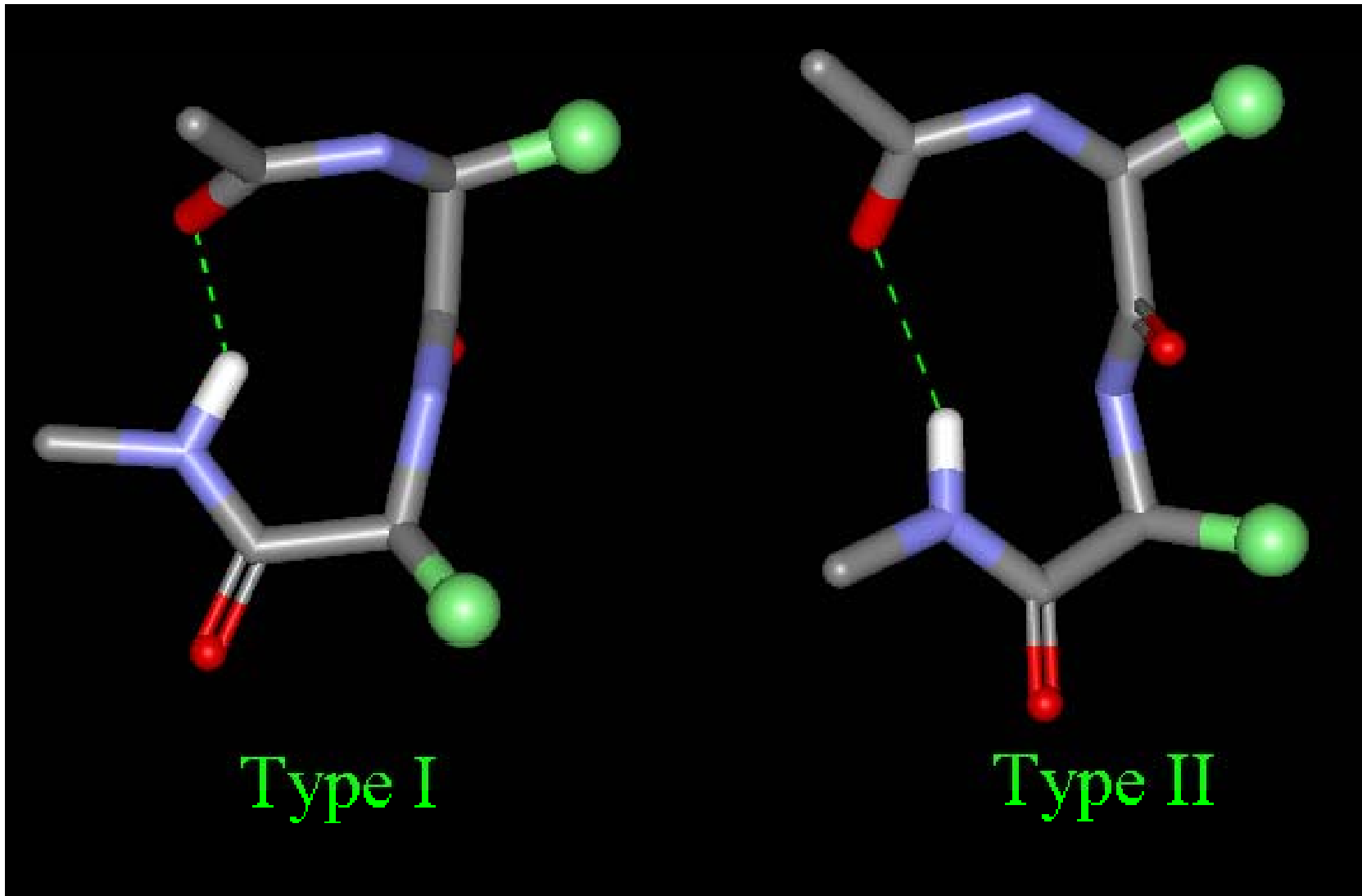
# Otáčky (turns)

- větší množství otáček (obrácení směru polypeptidového řetězce), různé počty AK



# $\beta$ -otočky

nejčastější H-vazby ob reziduum



# Coils – smyčky, neuspořádané klubko

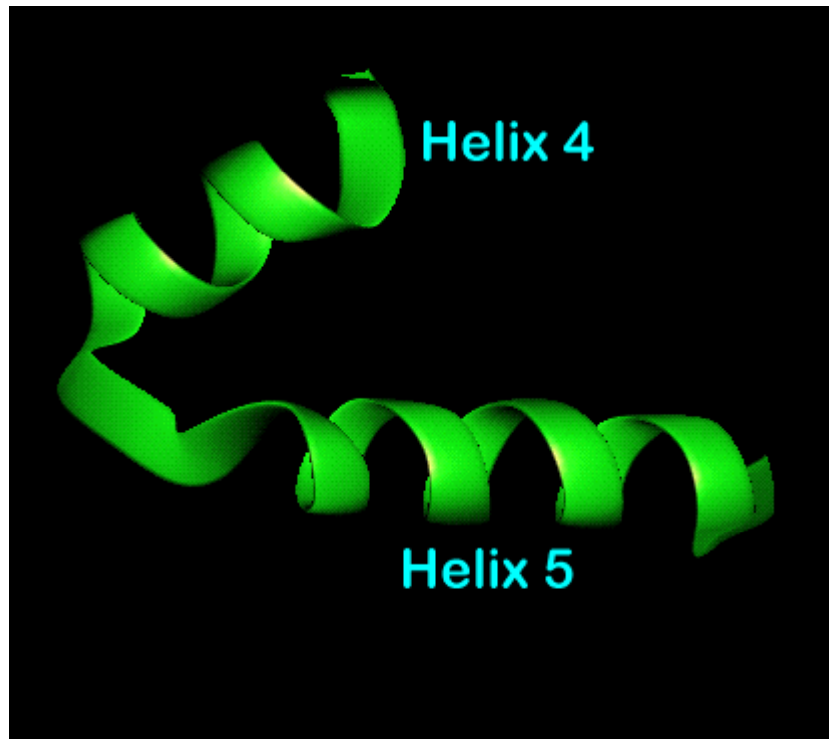
- Vše ostatní, co nespadá mezi  $\alpha$ -helix,  $\beta$ -list či otočky

# 3D vizualizace 2D struktur

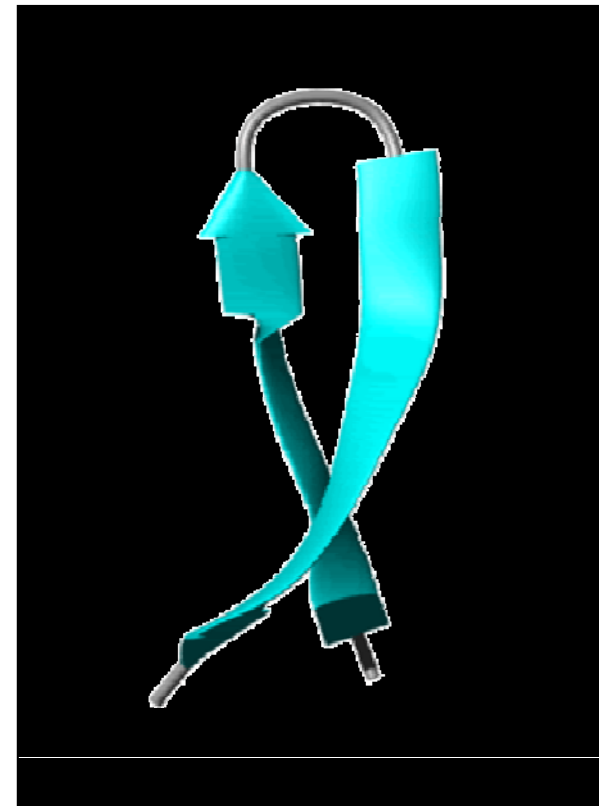


# Jednoduché motivy

helix-otáčka-helix

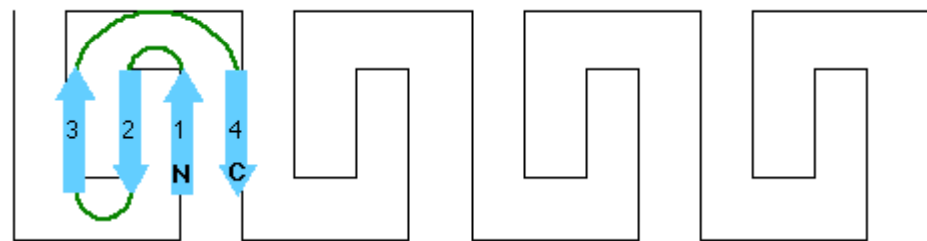


$\beta$  -vlásenka



# Jednoduché motivy

Řecký klíč (greek key)



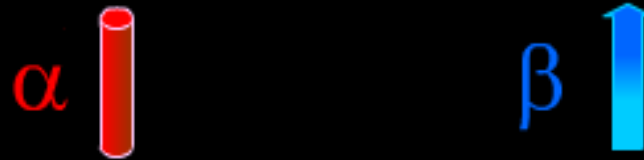


# Motivy - Domény

- Kombinací jednoduchých motivů jsou tvořeny **motivy**
- Motivy vytvářejí proteinové **domény**
- Dle zastoupení 2D struktur dělíme proteiny na:
  - $\alpha$  -proteiny
  - $\beta$ -proteiny
  - $\alpha / \beta$  proteiny – kombinace  $\beta$ - $\alpha$ - $\beta$  motivů
  - minoritní skupiny – oddělené domény tvořené jen  $\alpha$  či jen  $\beta$  strukturami, domény bohaté na kovy,...

There are three main classes of **super-secondary structures** all- $\alpha$ , all- $\beta$  and mixed  $\alpha/\beta$  structures. Simple super-secondary structures consist of 2 to 3 secondary elements. These can be assembled into larger super-secondary structures that are sometimes referred as "folds".

**Secondary Structure Elements**



**Super-Secondary Structure**

$\alpha\alpha$



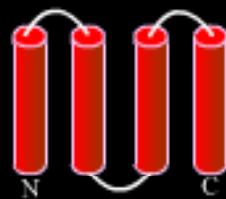
$\beta\alpha\beta$



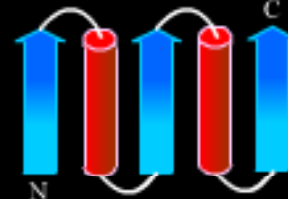
$\beta\beta$



**Examples of larger associations (Folds)**



*4-helix bundle*



*Rossmann fold*



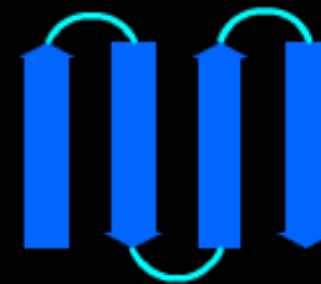
*greek key*

Representative  $\beta$ -super-secondary structures are illustrated in the following pages. The super-secondary structures are represented by both their 3D architecture and by their 2D schematic topological arrangement.

- $\beta$ -Hairpin



- $\beta$ -Meander

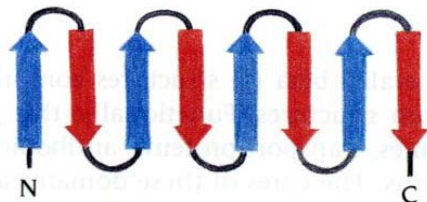
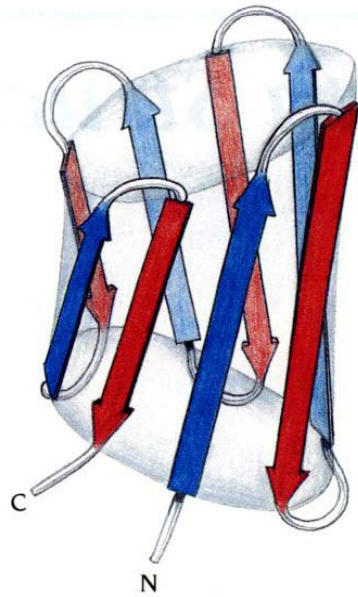


- Greek Key

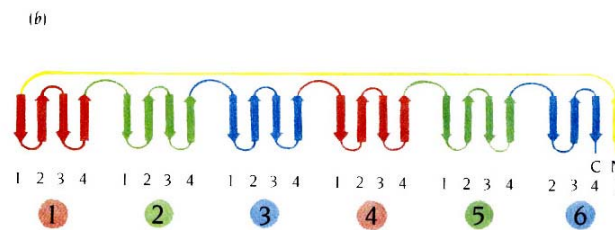
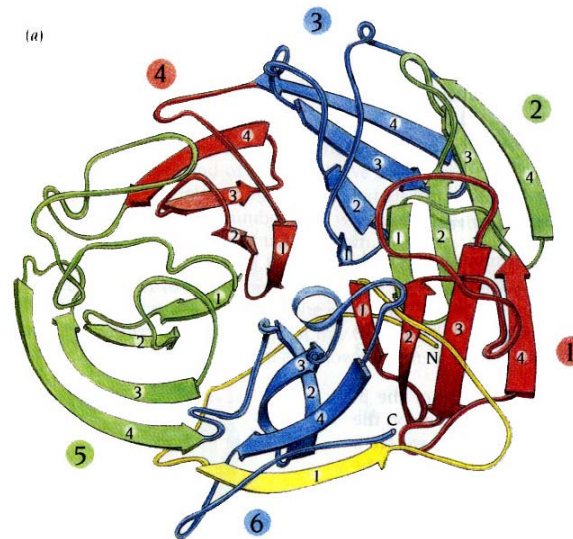


# $\beta$ -struktury

$\beta$  -barel

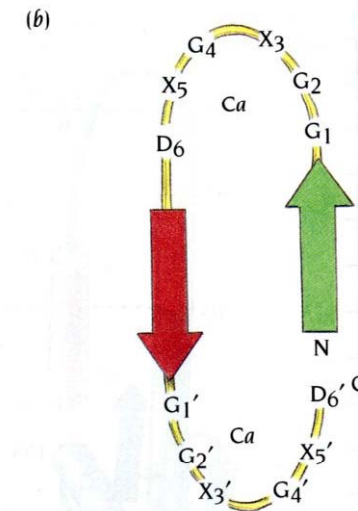
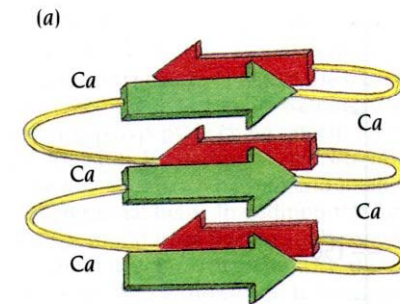
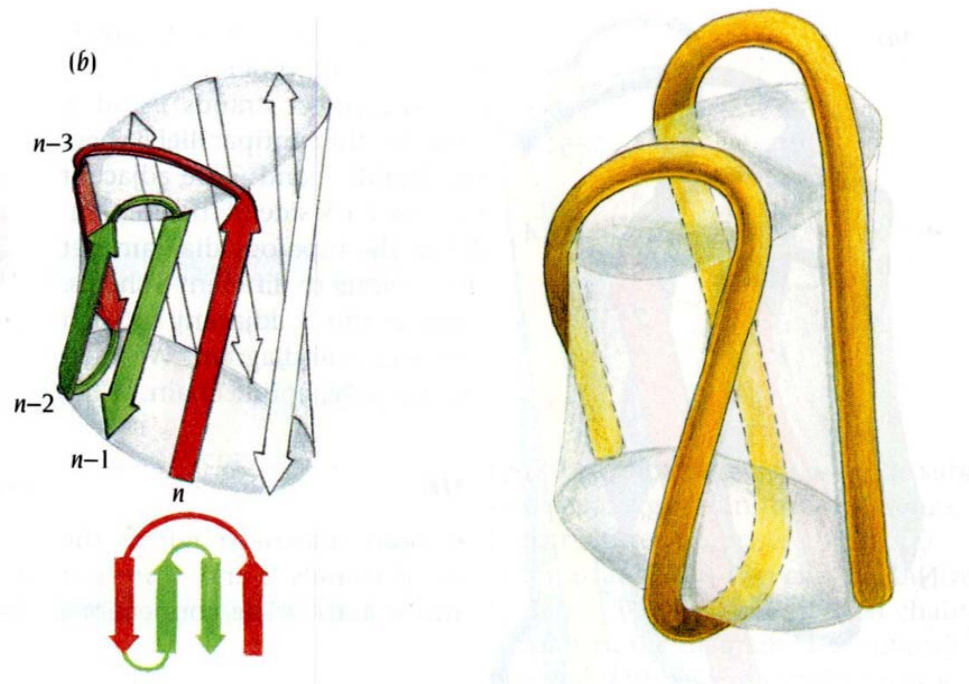


propeller blade



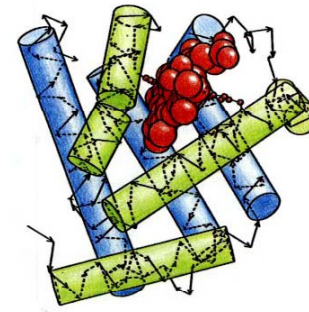
# $\beta$ -struktury

- Greek key barrel
- Jelly roll barrel
- $\beta$  -helix



# $\alpha$ -proteiny/motivy

- Globinový fold

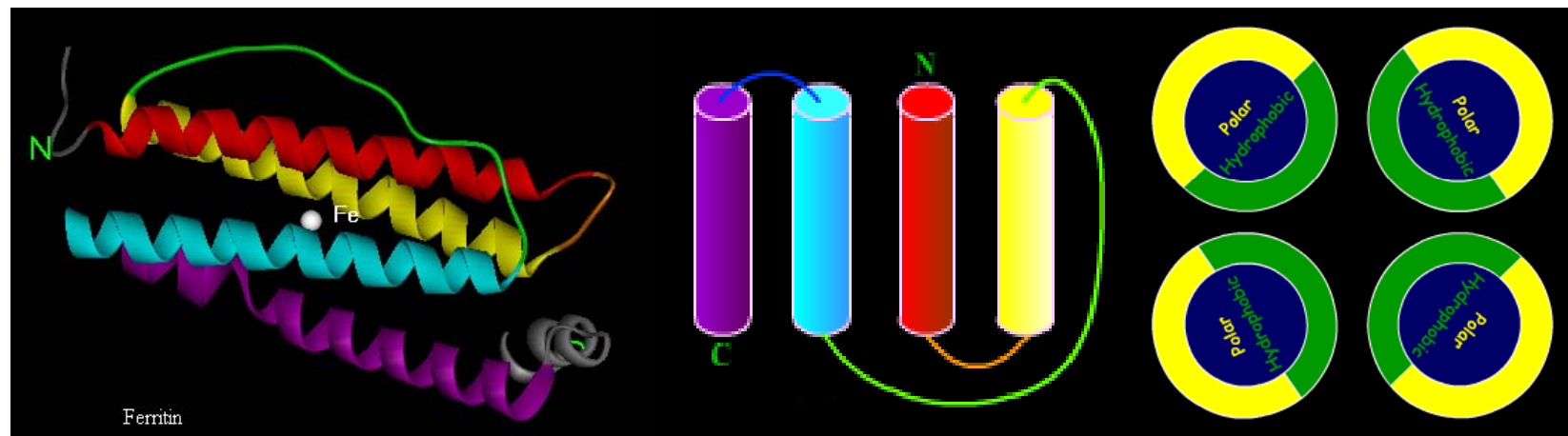
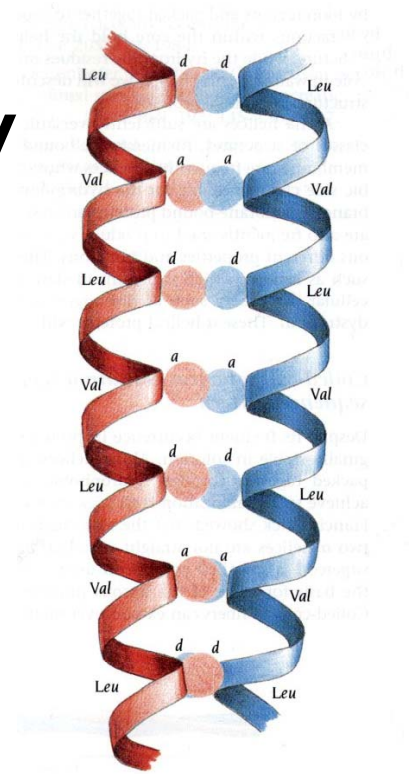


- Membránové proteiny



# $\alpha$ -proteiny/motivy

- Strukturní proteiny – keratiny, cytoskelet, ...
  - coiled-coil
  - čtyřhelixový svazek (Four-Helix Bundle)

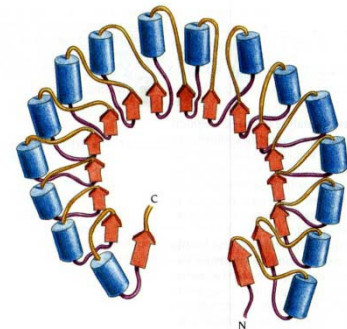
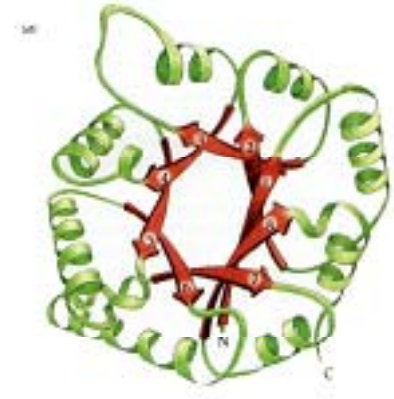




# $\alpha$ / $\beta$ motivy

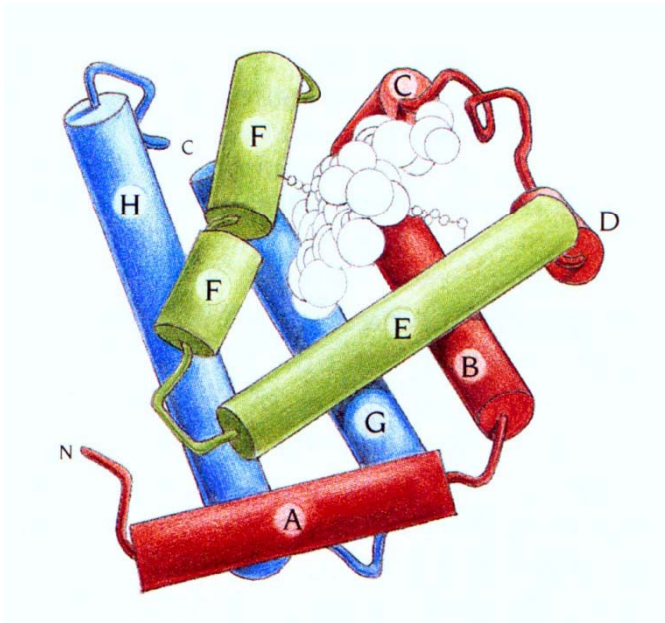
Centrální jádro tvořené  $\beta$  -listy  
obklopené  $\alpha$  -helixy

- TIM barrel
- Rossmannův fold
- Horseshoe fold – leucin-rich motivy

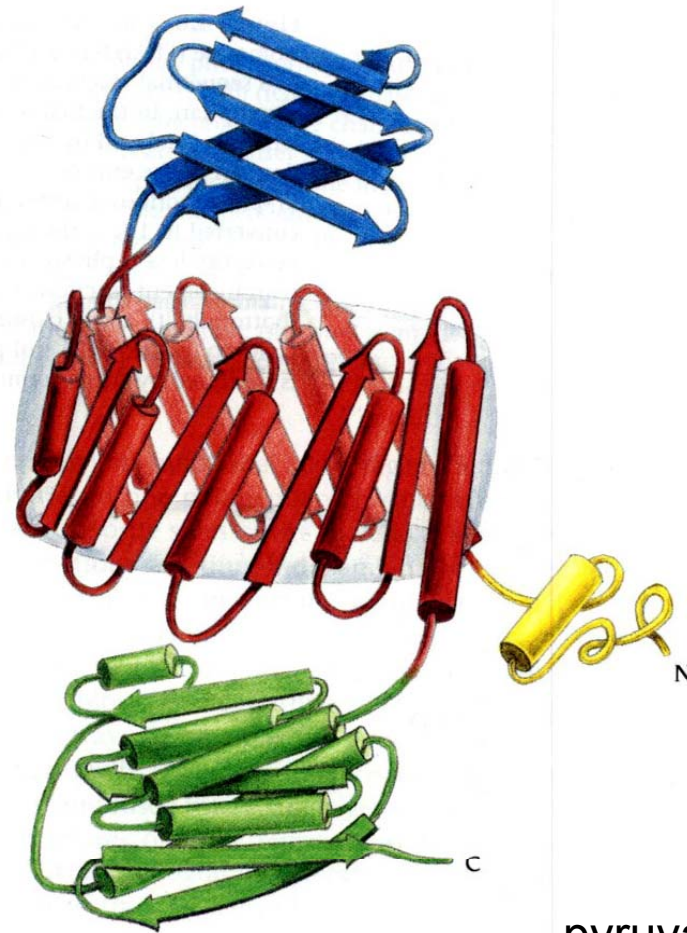




# Organizace proteinů – jedno- i vícedoménové proteiny



myoglobin

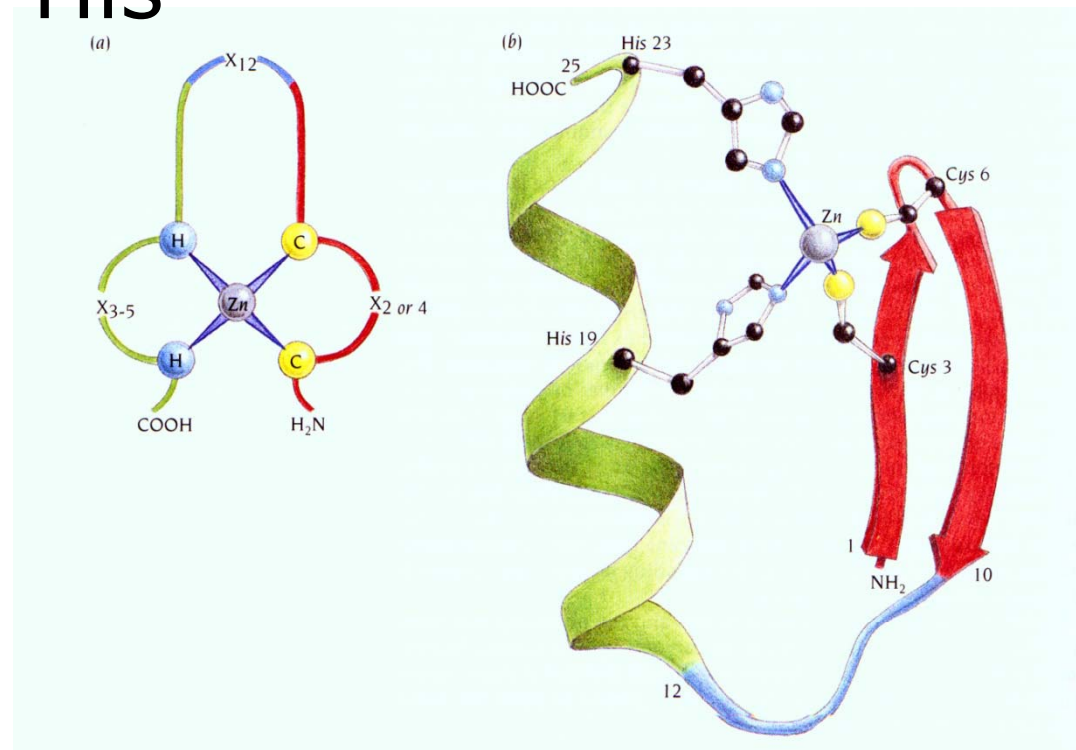


pyruvátkináza

# Zinkový prst

Výskyt: transkripční faktory

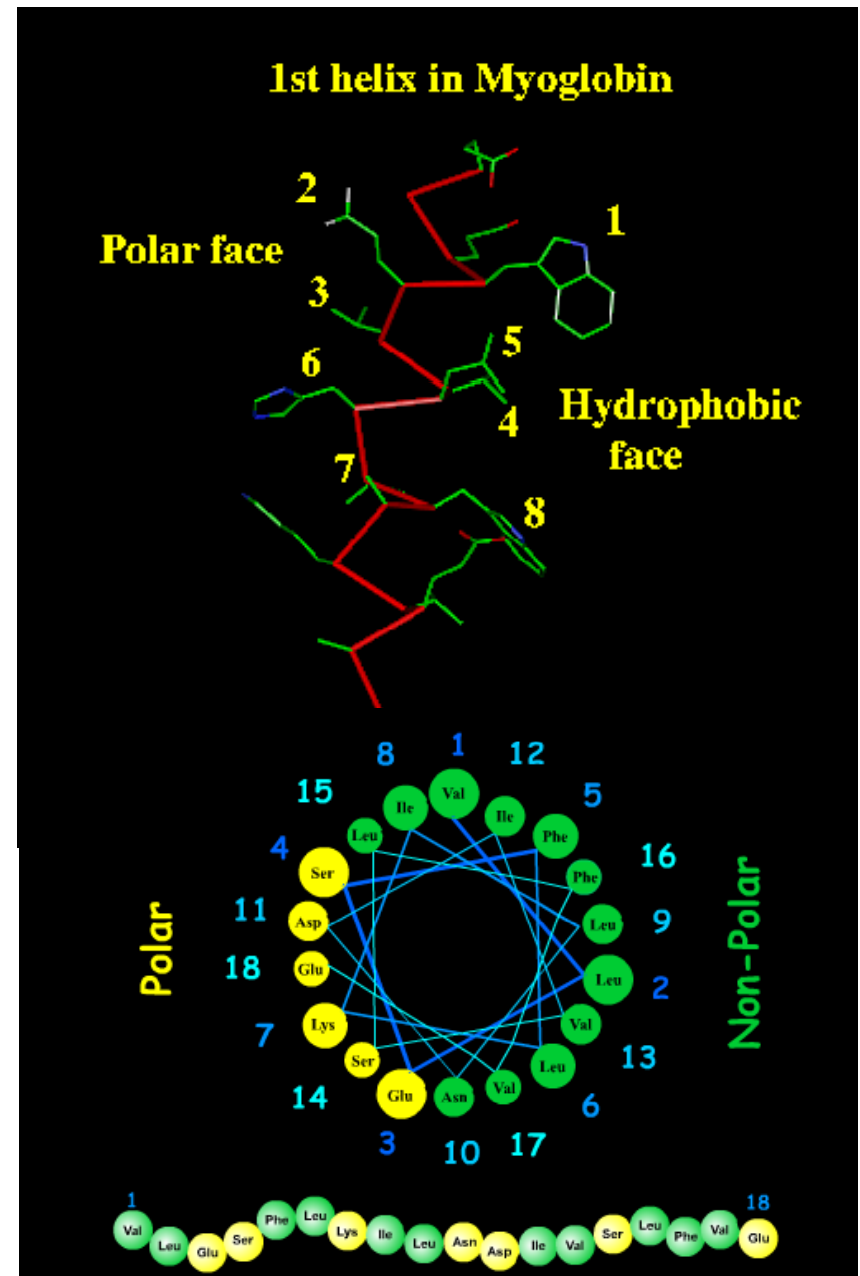
Několik typů – Zn atom je vázán AK  
zbytky Cys a His



# Typické znaky $\alpha$ -helix

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

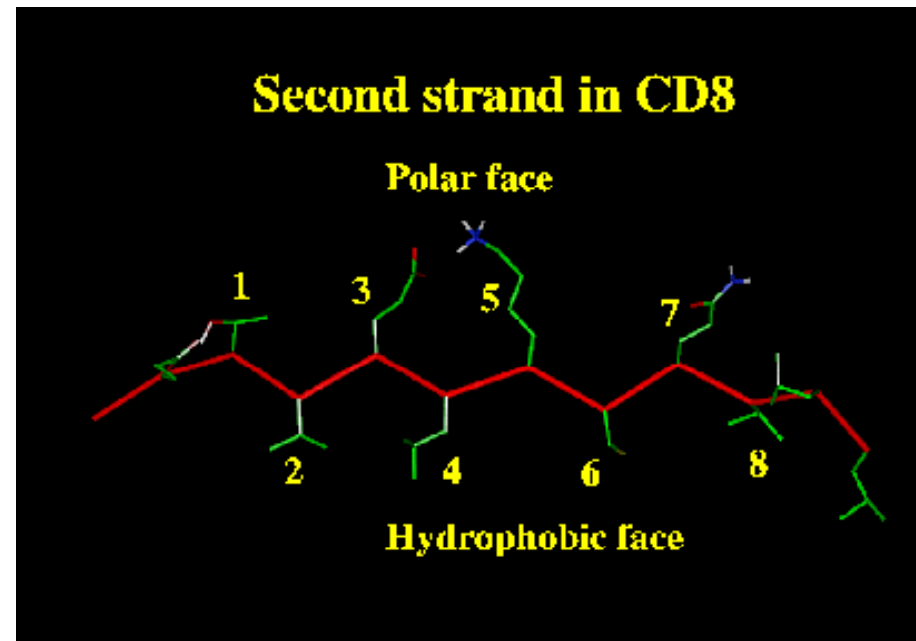
Potom pro 3.6 helix ( $\alpha$ -helix) platí, že  $i$ ,  $i+3$ ,  $i+4$  &  $i+7$  -té reziduum míří na tutéž stranu. Jsou-li všechna hydrofobní či naopak hydrofilní = zřejmě  $\alpha$  -helix



# Typické znaky $\beta$ -list

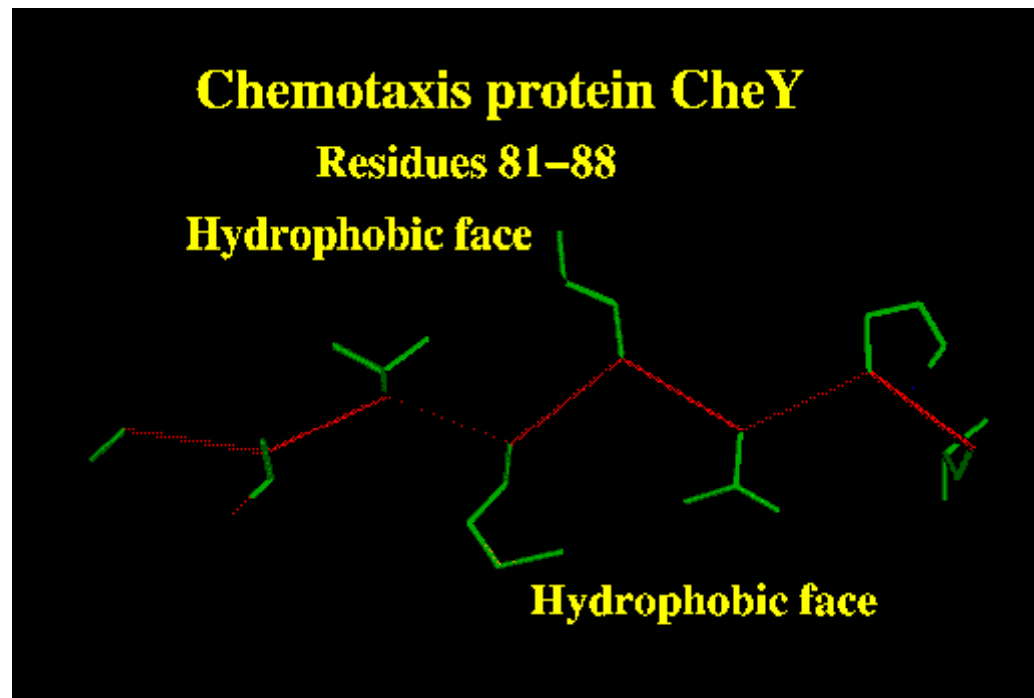
U  $\beta$  -listu se střídají rezidua po  $180^\circ$  a pro částečně zanořený  $\beta$  -list platí analogicky:

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



# Typické znaky $\beta$ -list

- Zcela zanořený  $\beta$  -list (typicky u  $\alpha / \beta$  barelu) je tvořen řadou nepolárních AK



# Predikční algoritmy

1. generace: *ab-initio*, vycházela z fyzikálně-chemických vlastností a ze statistiky pro jednotlivá rezidua (Chou-Fasman, GOR (Garnier, Osguthorpe, Robson))
2. generace: *plus incorporation of more local residue interactions*, zahrnovala i vliv nejbližších AK na zkoumané reziduum – předpověď max. 60% správnost, u  $\beta$ -listu do 40%
3. generace: *homology-base models*, zahrnuje navíc multiple sequence alignment a využívá skutečnosti, že 2D struktura se zachovává déle než sekvenční podobnost – až 80% spolehlivost (závisí na metodě)

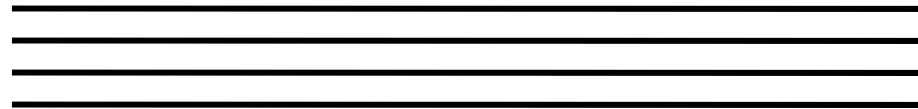
# Relative Amino acid Propensity Values for Secondary Structure Elements Used in the Chou-Fasman Methods

Amino Acid	( $\alpha$ -Helix)	$P$ ( $\beta$ -Strand)	$P$ (Turn)
Alanine	1.42	0.83	0.66
Arginine	0.98	0.93	0.95
Asparagine	0.67	0.89	1.56
Aspartic acid	1.01	0.54	1.46
Cysteine	0.70	1.19	1.19
Glutamic acid	1.51	0.37	0.74
Glutamine	1.11	1.11	0.98
Glycine	0.57	0.75	1.56
Histidine	1.00	0.87	0.95
Isoleucine	1.08	1.60	0.47
Leucine	1.21	1.30	0.59
Lysine	1.14	0.74	1.01
Methionine	1.45	1.05	0.60
Phenylalanine	1.13	1.38	0.60
Proline	0.57	0.55	1.52
Serine	0.77	0.75	1.43
Threonine	0.83	1.19	0.96
Tryptophan	0.83	1.19	0.96
Tyrosine	0.69	1.47	1.14
Valine	1.06	1.70	0.50

$$\frac{R_i(\text{SS})}{R_t(\text{SS})} = \frac{\sum R_i}{\sum R_t}$$

# 3. Generace - Homology-based methods

MSA



Predikce sekundárních struktur pro každou sekvenci



fitování předpovězené sekundární struktury do AA příložen

HHHCHCCEEEECCHH  
HHHHHCCEEEECCHH  
ECCHHCCEEEECCEE  
HHHHHCCCCEEEECCH  
HHHHCCCEEEECHHC

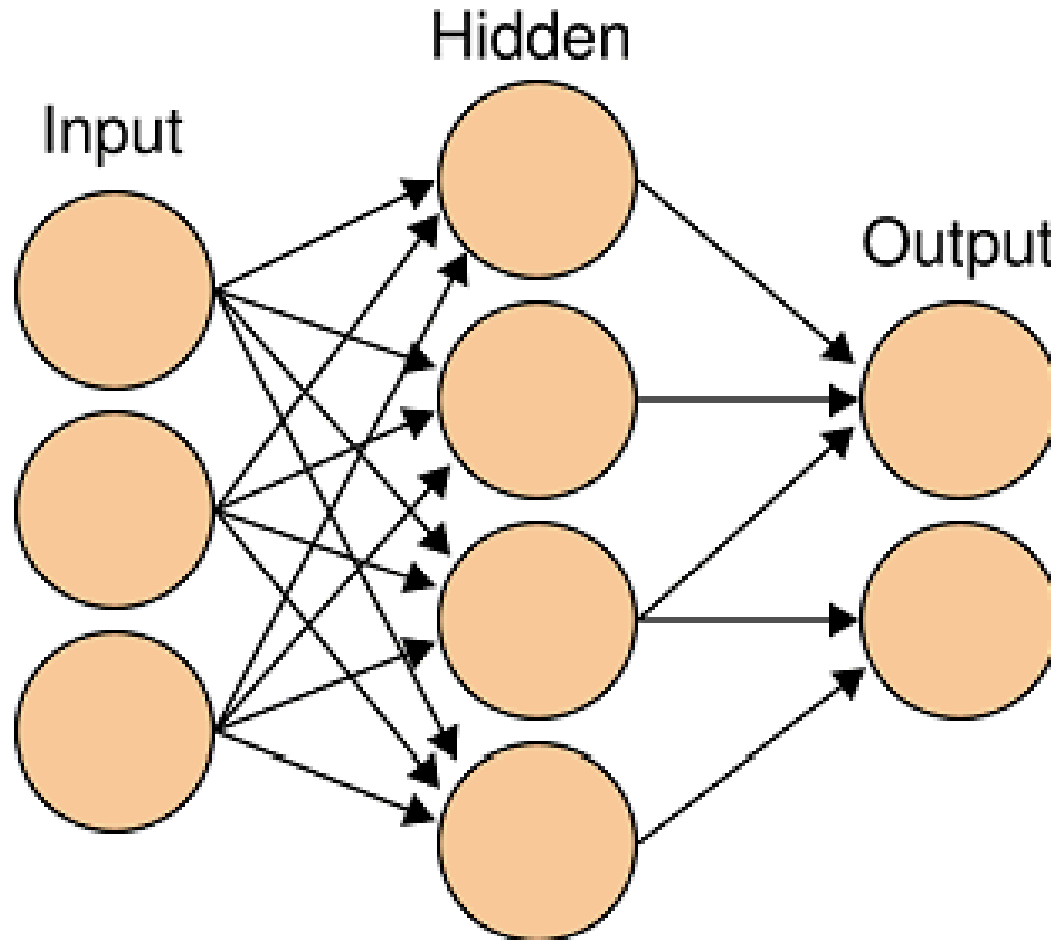


Konečná předpověď  
Založená na konsenzuální sekvenci

HHHHHCCEEEECCHH



# 3. Generace – neuronové sítě



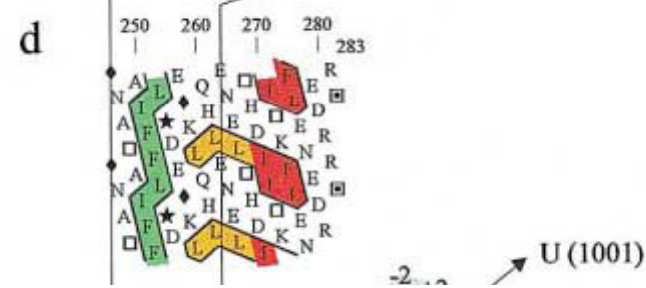
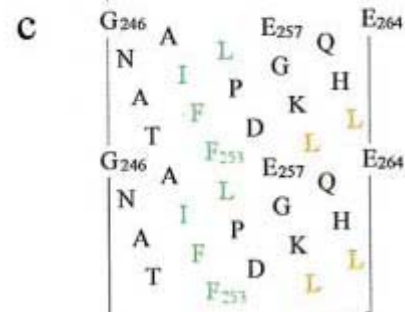
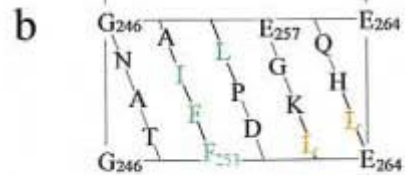
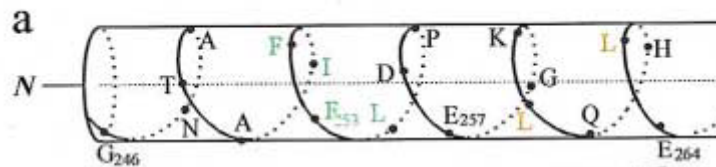
Sekvence se známou  
sek. strukturou

Trénink, přiřazování  
Váh jednotlivým funkcím

Aplikace nalezených  
algoritmů na neznámou sekvenci

# Hydrophobic cluster analysis

2D



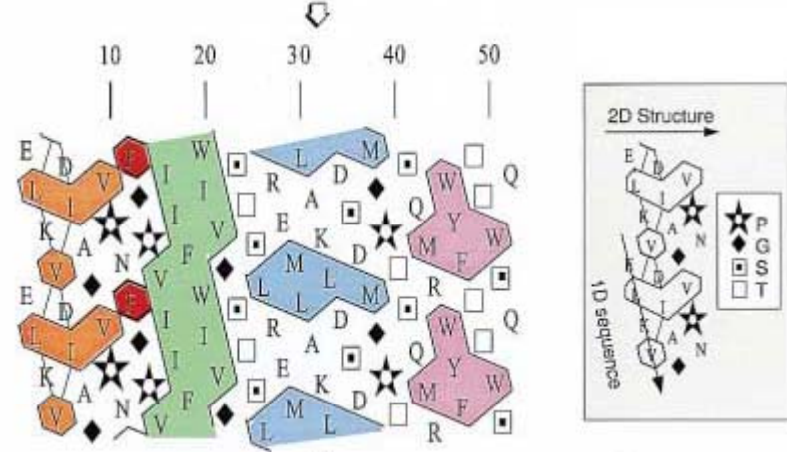
1D

ELKVDIAGV<sup>1</sup>PNF<sup>2</sup>GPV<sup>3</sup>IIFWIVG<sup>4</sup>STSLREML<sup>5</sup>  
 AKLDSDMGPTSQMRWYF<sup>6</sup>TTWSQ

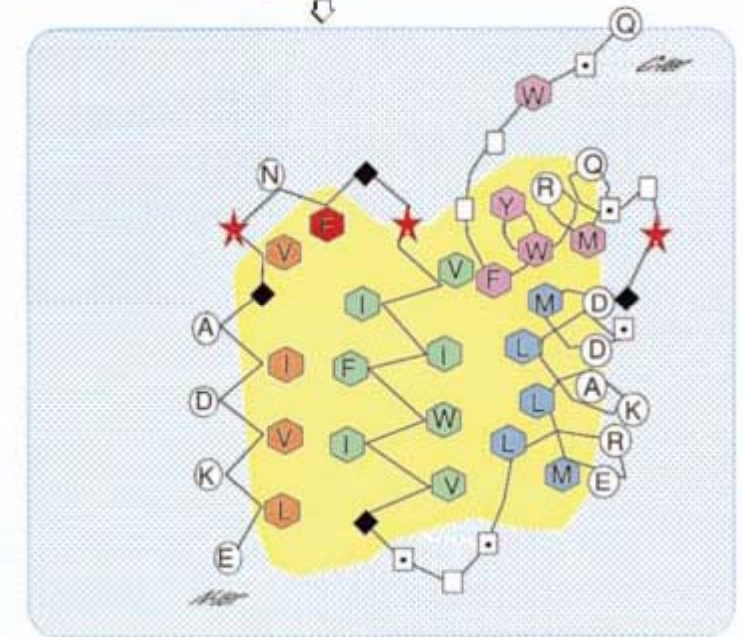
Coded  
1D

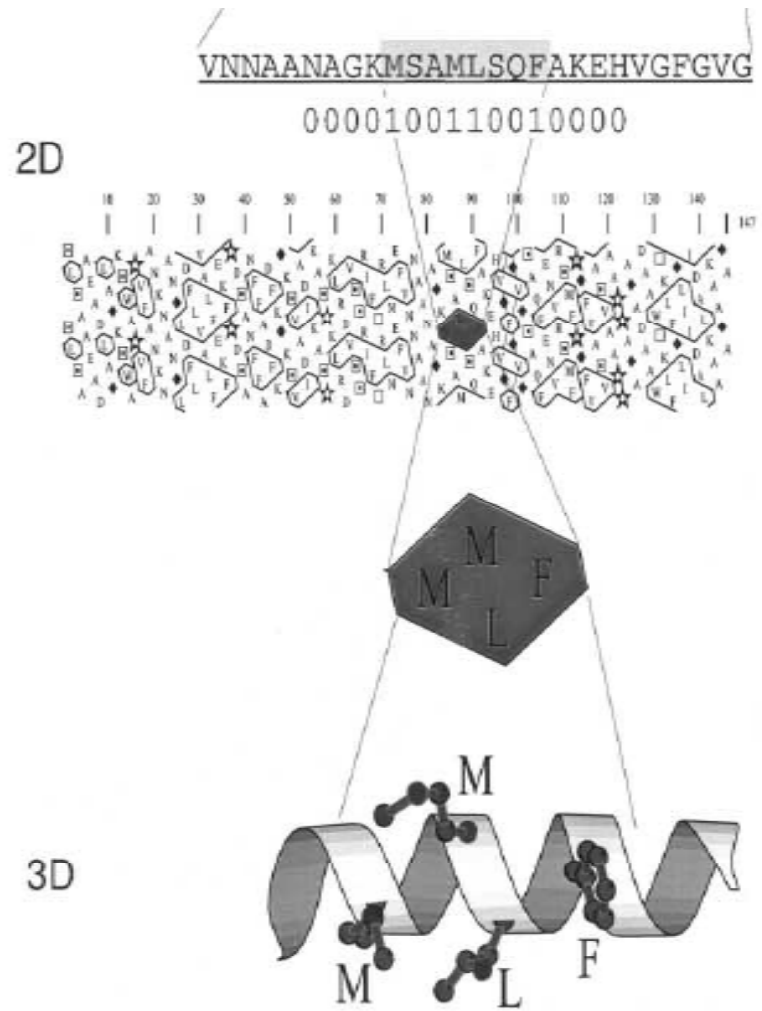
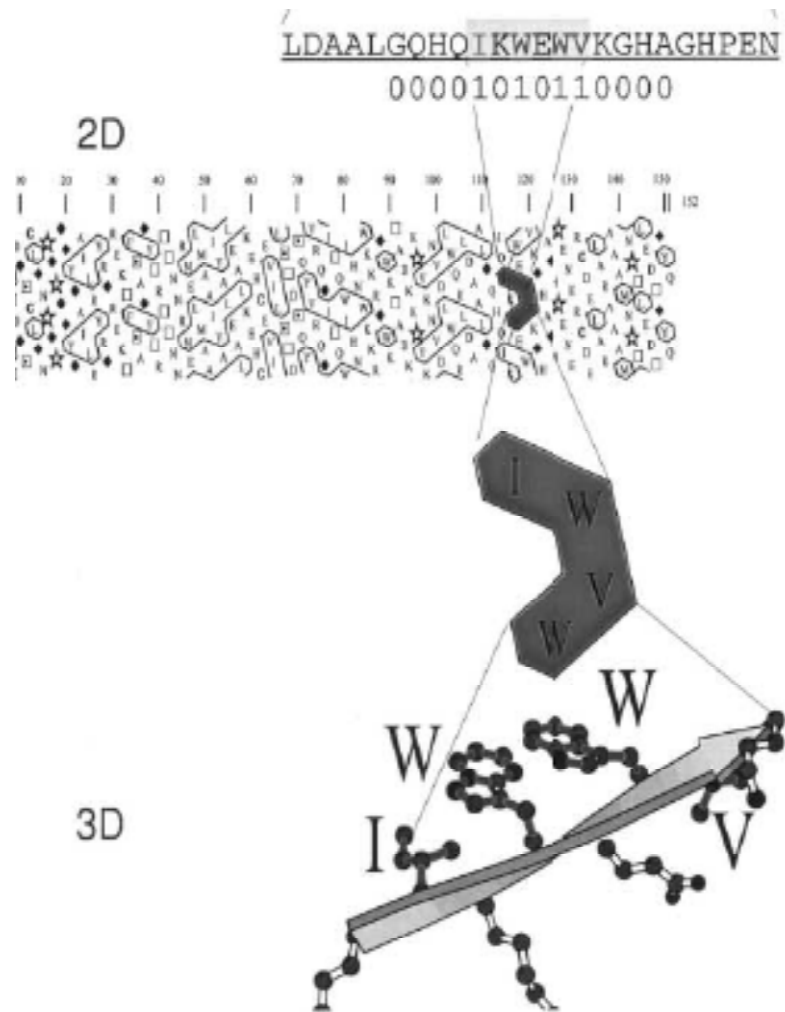
ELKVDIA<sup>1</sup>★V★NF<sup>2</sup>★V<sup>3</sup>IIFWIV<sup>4</sup>□□□LREML<sup>5</sup>  
 AKLD□DM<sup>6</sup>★□□QMRWYF□□W□Q

2D  
HCA



3D



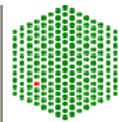


# Programové balíky

- [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, PHDthreader, 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

# Predikce obsahu helixu ve struktuře – Agadir

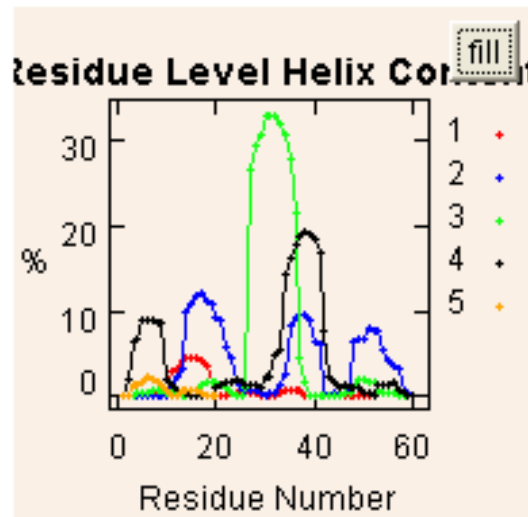
(<http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html>)



EMBL WWW Gateway to AGADIR Service

A service from the [Serrano's group](#)

August 6th 2003: A new version of AGADIR has been implemented.

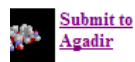


## AGADIR

*algorithm to predict the helical content of peptides*

[description](#)  
[software upgrades](#)  
[application](#)  
[instructions](#)

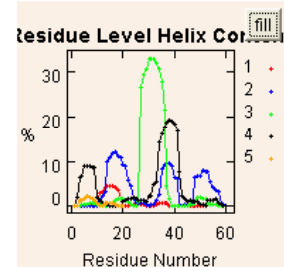
7 version  
 1st 6th, 2003



Thanks for visiting the Agadir service.

Peptide 1	MNTRSFHRIDVHKARELLQRPDTVLLDCRHPSPDFRAGHIAGASPLGDYNADDHVLNIAKH
Peptide 2	RPVLIYCYHGNASQMRAQLFADFGFAEVYSLDGGYEAWRKVHTPANSQLTEALQCWLMAQ
Peptide 3	EFPAADIHARTRDGVTPLMRAAGEGDFARVAELLAAGADPHQRNNDGNQALWFACVSENL
Peptide 4	DTLDLLVAVGAHLNHQNDNGATCLMYAASAGKTAVVERLLAFGADRSLLSLDDFTALDMA
Peptide 5	ANLECLNLLRETPRIKAVT

Check the [output](#)



# PSIpred

[PSIPRED](#)  
[home>](#)

## The PSIPRED Protein Structure Prediction Server

### Info

We suggest that you do not bookmark this page as it is liable to move. It is best to access the server via the [PSIPRED home page](#), which has more information about the methods and a full reference list.

### Input Sequence

[Help](#)

Input sequence (single letter code)

### Choose Prediction Method

[Help](#)

- Predict Secondary Structure (PSIPRED v2.6)
- Predict Transmembrane Topology (MEMSAT3)
- Fold Recognition(GenTHREADER - quick)
- Fold Recognition (mGenTHREADER - with profiles and predicted secondary structure)
- Fold Domain Recognition (domGenTHREADER - with profiles and predicted secondary structure for domains within proteins)

### Filtering Options

[Help](#)

- Mask low complexity regions
- Mask transmembrane helices
- Mask coiled-coil regions

Warning: Turn off all filtering if you are running MEMSAT3

### Submit Sequence

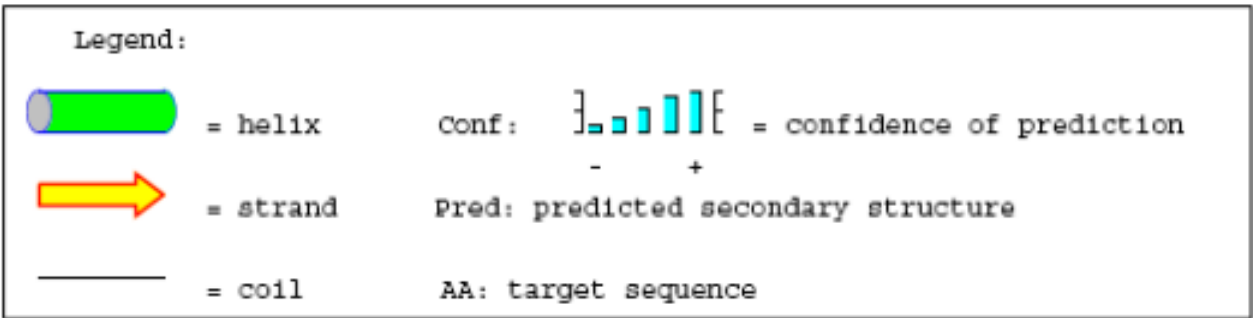
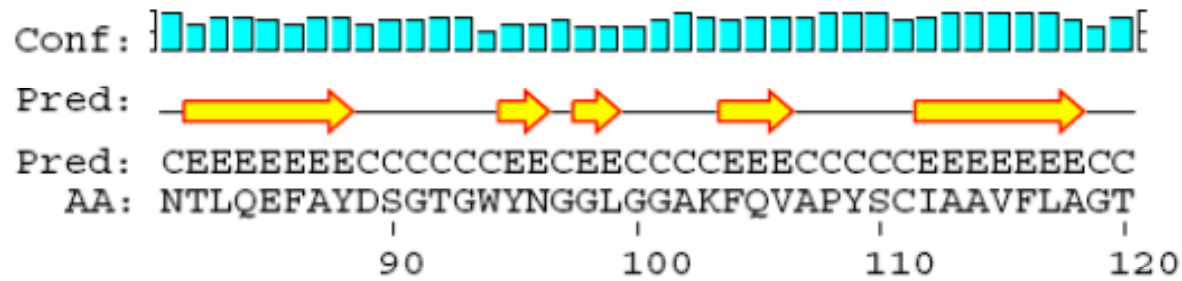
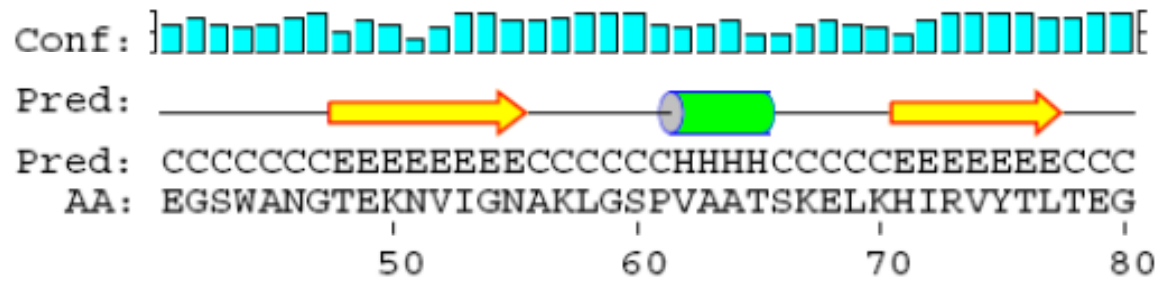
E-mail address [Help](#)

Password (only required for commercial e-mail addresses) [Help](#)

Short name for sequence [Help](#)

Predict

Clear form





# APSSP

## APSSP: Advanced Protein Secondary Structure Prediction Server

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

This server allow to predict the secondary structure of protein's from their amino acid sequence. This is an advanced version of our PSSP server, which participate in [CASP3](#) and in [CASP4](#). PSSP is also part of [CAFASP2](#). Raghava, G. P. S. (2000) Protein secondary structure prediction using nearest neighbor and neural network approach. CASP4: 75-76. This server is also [participating in world-wide Live-Bench competition EVA](#), so you can get the performance of methods including APSSP from EVA Server. This server is also part of [Meta II Prediction](#) server. Please visit, [ExpASY Tools](#) for more protein structure prediction tools.

## Request Form

Target/name of protein (optional):

Paste your sequence data here:

```
ATQQGVFTLPANTRFGVIAFANSSGTQIVNVLVNNETAATFSGQSTNNAVIGTQVLNSGS  
SGKVQVQVSVNGRPSDLVSAQVILTNELNFALVGSSEDTDNDYNDVAVVWINWPLG
```

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!

Contact Email Address: [raghava@imtech.ernet.in](mailto:raghava@imtech.ernet.in)

Většina programů má  
uživatelsky jednoduché  
rozhraní

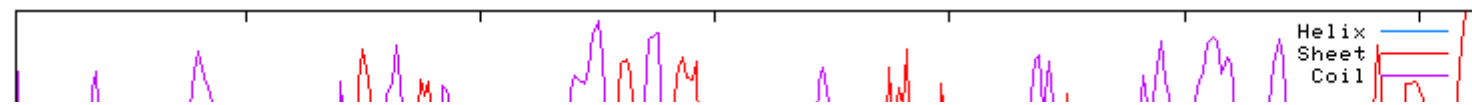
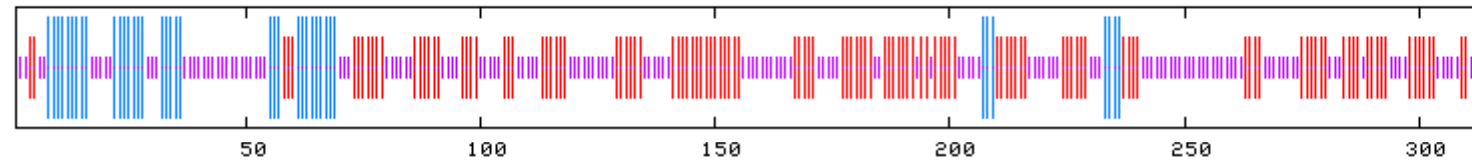


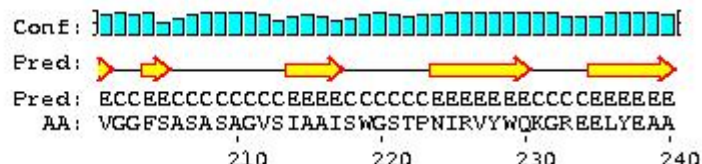
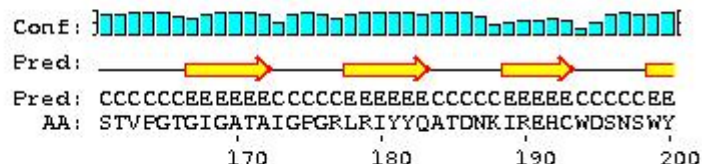
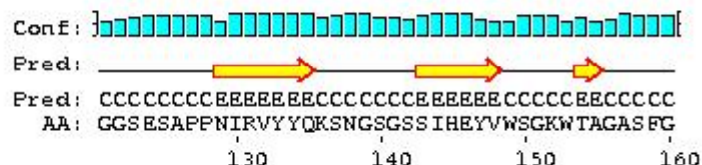
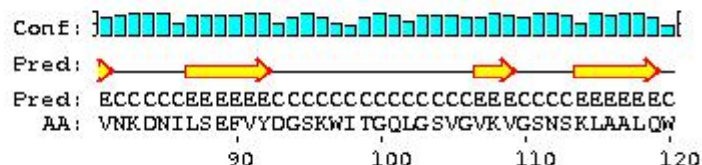
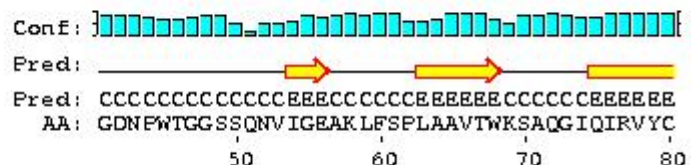
```
      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
PTEFLYTSKIAAISWAATGGRQQRVYFQDLNGKIREAQRRGGDNPWTGGSSQNVIGEAKLFSPLAAVTWKS
ccccceecchhhhhhhhhccccchhhhhhhccchhhhhccccccccccccccccccccchhheeehhhhhhhhh
AQGIQIRVYCVNKDNLSEFVYDGSKWITGQLGSGVGVKVGSNKLAALQWGGSESAPPNIRVYYQKSNGS
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
GSSIHEYVWVGKWTAGASFGSTVPGTGIGATAIGPGLRIYYQATDNKIREHCWDSNSWYVGGFSASASA
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
GVSIAAISWGSTPNIRVYWQKGREELYEAAYGGSWNTPGQIKDASRPTPSLPDTFIAANSSGNIDISVFF
eeeeeeccccccccccccccccchhhheeecccccccccccccccccccccccccccccccccccccccc
QASGVSLQQWQWISGKGWSIGAVVPTGTGAPAGW
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
```

Sequence length : 312

GOR4 :

Alpha helix	(Hh)	:	40	is	12.82%
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)	:	122	is	39.10%
Beta turn	(Tt)	:	0	is	0.00%
Bend region	(Ss)	:	0	is	0.00%
Random coil	(Cc)	:	150	is	48.08%
Ambiguous states (?)		:	0	is	0.00%
Other states		:	0	is	0.00%







## GOR4 result for : UNK\_78160

[Abstract](#) GOR secondary structure prediction method version IV, J. Garnier, J.-F. Gibrat, B. Robson, Methods in Enzymology, R.F. Doolittle Ed., vol 266, 540-553, (1996)

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

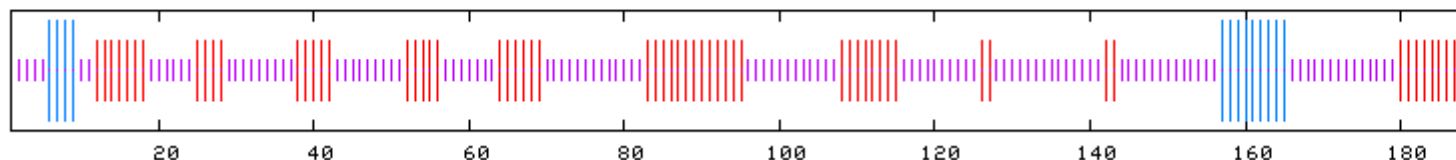
```

      10      20      30      40      50      60      70
      |      |      |      |      |      |      |
SDVDIEAQDAGQTLVQVISIPSETWVAIQLP SQRYRFDVVFENVSPTSSG SVLVAQMAPQSGGVYGSNY
ccccccchhhcccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
SGSGWGN DLGGGFYGYSEAKWMCLWPANRSGPSSKTGLYGTCKLMNLNQSSAVPSVTSNLFAPTAYKNE
cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
PGYANVGGCCQKIRGLASSIQFAFALAGGNVPQNTDTFNGGTIKVYGWN
ccccccccccccccccccccchhhhhhhhhcccccccccccccccccccccccccccccccccccccccc
    
```

Sequence length : 189

GOR4 :

Alpha helix	(Hh) :	13 is	6.88%
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) :	60 is	31.75%
Beta turn	(Tt) :	0 is	0.00%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	116 is	61.38%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%







# Rozšíření možností 2D predikce

Vedle předpovědi 2D struktury je rovněž analyzována

- Přístupnost pro **solvent**
- Předpověď **transmembránového** helixu
  - Určení zda je/není transmembránový
  - Podíl hydrofobních řetězců (AK zbytků) na povrchu – umožňuje postihnout i částečně zanořené membrány

# Rozšíření možností 2D predikce

**SSpro8** (<http://www.igb.uci.edu/tools/scratch/> )

Nepoužívá pouze základní tři typy 2D struktury (helix,  $\beta$ -list a ostatní), ale využívá kompletní klasifikaci dle DSSP (Database of Secondary Structure Assignments) – tj. 8 typů struktury:

- H: alpha-helix
- G:  $3_{10}$ -helix
- I:  $\pi$ -helix (extremely rare)
- E: extended strand
- B: beta-bridge
- T: turn
- S: bend
- C: the rest

Výsledná přesnost cca 63%



# SSpro8 – ukázka výstupu

```
SSpro and SSpro8 predictions (http://www.igb.uci.edu/tools/scratch/)  
* See below for notes!
```

```
Query_name: srp21_cer
```

```
Query_length: 166
```

```
Prediction:
```

```
SVKPIDNYIITNSVRLFEVNPSTLFSISYKPTQKIDTKVSRFRHNSHLSLNYKFTTNKS  
CCCCCCCCCCCCCEEEEECECCCCCEEEEECCCCCCCCCEEEEECCCCCEEEEECCCC  
.....TEEESEEEEE.TT.EEEEE.....EEEEEE...EEEEEE...S.
```

```
-----+++++-----+-----+-----+-----+-----+-----  
-----++-+++-----+-----+-----+-----+-----  
-----+-----  
-----+-----
```

```
eeeeeeee-eee-e--eeeeeeee-eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee  
eeee-eee-ee--e--e-eeee--e-eeeeeeeeeeeeeeee-eeeeee-eee-e-eeee  
eeee-eee-ee--e--e-eeee--e-eeeeeeeeeeeeee-e-eeeeee-e-e-e-eeee  
eeee-ee--e--e--e-eeee--e-eeeeeeeeeeee-e-eeeeee-e-e-e-eeee  
eeee-ee-----e--e-----e-eeeeeeee--ee-e-----e-e-eee-
```

```
KDVSRLLSALGPRGVSITPGKIEKIAQSKKKNNKIKESSKKIKGKSIQDIVGLATLIVNT  
HHHHHHHHHCCCCCECCCCCHHHHHHHHHHHHCCCCCECCCCCCCCCHHHHHHHHHHHHCC  
HHHHHHHHHH..T..EE..TTHHHHHHHHHHTTEEEE..HH..TT..HHHHHHHHHEEEE.
```

```
---+++++-----+-----+-----+-----+-----+-----  
-----+-----+-----+-----+-----+-----+-----  
-----+-----+-----+-----+-----+-----+-----  
-----+-----+-----+-----+-----+-----+-----
```

```
ee-ee--eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee-eee-e---ee  
ee-ee--eeeeeee-e-eeee-ee-eeeeeeeeeeeeeeeeeeee-ee-----ee  
ee-ee--eeeeeee-e-eeee-ee--eeeeeeee-eeeeee-eeee-ee-----e  
ee-ee--ee-eeee-e-eeee-ee--eeeeeeee-eeeeee-eeee-ee-----  
ee--e--ee--eee-e--eee-ee--eeeeeeee-eeeeee-eee--e-----
```

# Rozšíření možností 2D predikce

Predikce **3D struktury**?

Viz. Další přednáška:

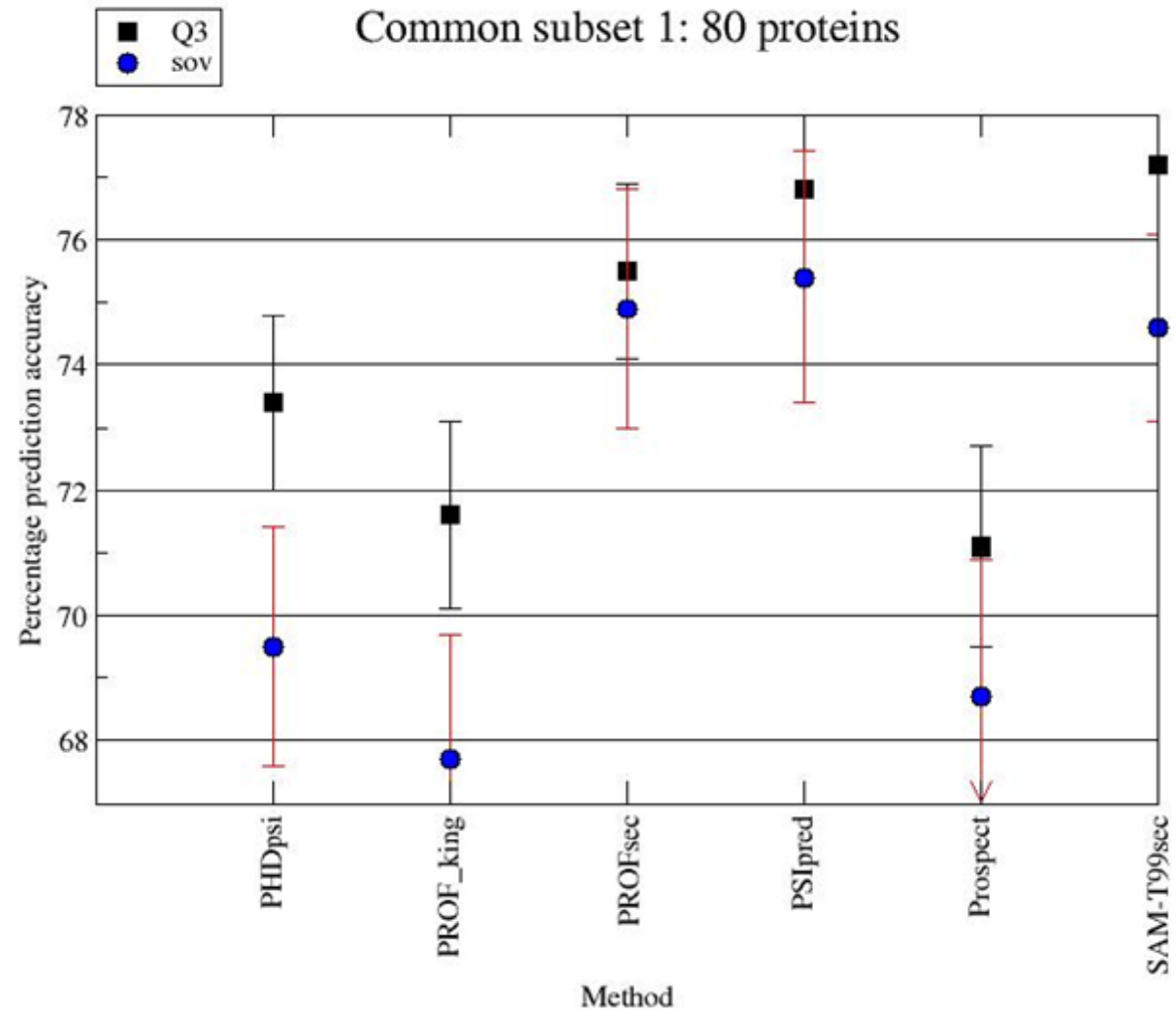
3-D topologie proteinů,  
předpověď 3-D struktury,  
threading

# Benchmark – porovnání algoritmů

Porovnání výsledků predikce se skutečně zjištěnou strukturou.

Benchmark **EVA** (<http://cubic.bioc.columbia.edu/eva/>)

- průběžné testování existujících serverů
- v současnosti v provozu přes 300 týdnů...
- testování 2D i 3D predikce



# Benchmark

Dle benchmarku EVA v současnosti na předních místech:

- PROFsec  
(<http://cubic.bioc.columbia.edu/predictprotein> )
- PSIPred  
(<http://insulin.brunel.ac.uk/psiform.html> )
- SABLE  
(<http://sable.cchmc.org/> )

# Zopakování / shrnutí

- ▼ **2D struktura** – vzájemné uspořádání AK jednotek proteinu ( $\alpha$ -helix,  $\beta$ -list)
- ▼ **3D struktura** – celkové rozmístění atomů proteinu v prostoru
- ▼ **Strukturní motivy** – na rozhraní mezi 2D a 3D, vznikají kombinací 2D struktur

# Zopakování / shrnutí

- ▼ Řada **programů** využívajících rozdílné přístupy – použití závisí na vstupních datech a účelu
- ▼ Neexistuje univerzální program – nejlepší je **kombinace výsledků** z více programů
- ▼ **Predikce 2D struktury je do značné míry spolehlivá** – zejména pro proteiny s  $\alpha$ -helix; vzrůstá při podobnosti s proteinem se známou strukturou

Most natural polypeptide chains contain between 50 and 2000 amino acid residues. Shorter chains are called oligopeptides or simply peptides. The following histogram shows the length of the proteins in the E-coli proteome.

