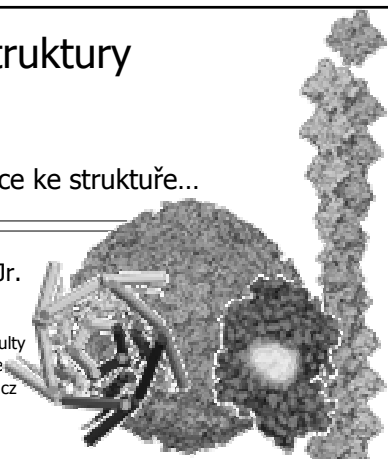


Úvod do struktury proteinů I

aneb od sekvence ke struktuře...

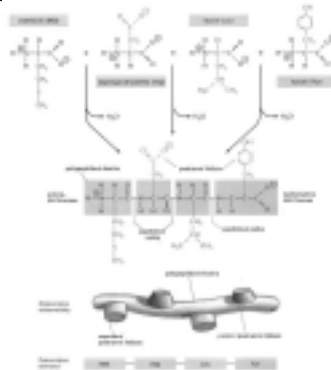
Vladimír Kopecký Jr.

Fyzikální ústav
Matematicko-fyzikální fakulty
Univerzity Karlovy v Praze
kopecky@karlov.mff.cuni.cz



Co je a jak vzniká protein?

- Vzniká v ribosomech pomocí tRNA v procesu zvaném translace – překladem tripletního kódu
- Peptidový řetězec je **orientovaný od N-konce k C-konci**
- Existují peptidy od 40 do 10 000 aminokyselin
- Průměrná hmotnost aminokyselinového zbytku je 110 Da

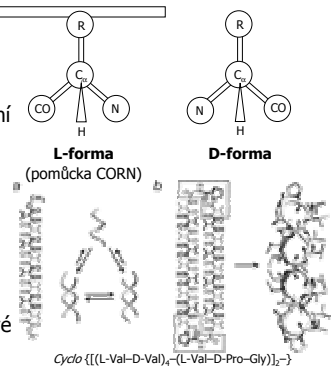


Genetický kód je když...

| První pozice | Druhá pozice | | | | Třetí pozice |
|--------------|------------------|---------|---------------|-------------------|--------------|
| | U | C | A | G | |
| | U Fenylalanin | Serín | Tyrosin | Cystein | U |
| | C Leucin | Prolin | STOP | STOP Tryptofan | C |
| A | Izoleucin | Treonin | Histidin | Arginin | A |
| | Metionin | | Glutamin | | G |
| G | Valin | Alanin | Asparagin | Serín | G |
| | | | Lysin | Arginin | C |
| | | | Kys. asparag. | Glycin | Třetí pozice |
| | | | Kys. glutam. | | |

L vs. D enantiomery aminokyselin

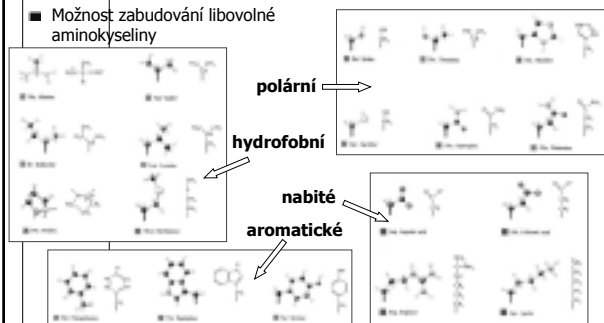
- Všechny aminokyseliny mimo Gly jsou chirální
- Život na zemi užívá **L-enantiomery** (chirální formu)
- Neexistuje fyzikálně-chemický důvod pro preferenci L před D-formou
- V umělých peptidech lze kombinovat L i D formu
- D forma vytváří zrcadlové obrazy sek. struktur



M. Sastry et al., J. Am. Chem. Soc. 128 (2006) 10650–10651.

20 standardních aminokyselin

- 21 aminokyselinou je selenocystein (přítomna od Archae po savce)
- 22 aminokyselinou je pyrrolysine (výskyt v Archae a Eubacteria)
- Možnost zabudování libovolné aminokyseliny



Kódování aminokyselin

- 1984 – Mezinárodní unie pro biochemii (IUB) stanovuje doporučení pro užívání jednopísmenných a třípísmenných zkratk standardních aminokyselin
- Třípísmenný kód** je snadno srozumitelný v textu článků
- Jednopísmenný kód** usnadňuje zápis a analýzu primární struktury proteinu do genomových a strukturních databází

>2BP4:A|PDBID|CHAIN|SEQUENCE

DAEFREDSGYEVHHQK

Asp-Ala-Glu-Phe-Arg-His-Asp-Ser-Gly-Tyr-Glu-Val-His-His-Gln-Lys

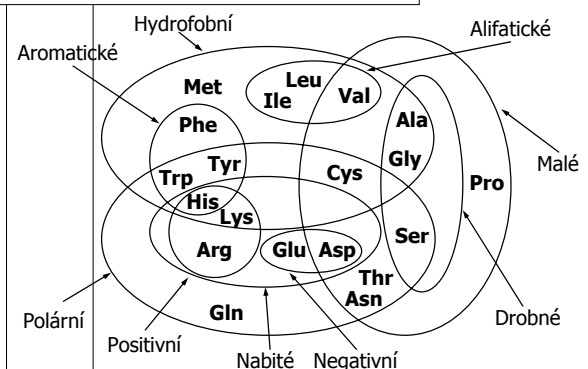
Jednopísmenný (fasta formát) a třípísmenný kód pro Zn-vazebnou doménu amyloidu β (16-mer)

Kódy aminokyselin – 3 vs. 1

| | |
|--|------------------------------------|
| ■ A = Ala = alanin (9,0) | ■ P = Pro = prolin (4,6) |
| ■ C = Cys = cystein (2,8) | ■ Q = Gln = glutamin (3,9) |
| ■ D = Asp = k. asparagová (5,5) | ■ R = Arg = arginin (4,7) |
| ■ E = Glu = k. glutamová (6,2) | ■ S = Ser = serin (7,1) |
| ■ F = Phe = fenyalanin (3,5) | ■ T = Thr = threonin (6,0) |
| ■ G = Gly = glycin (7,5) | ■ V = Val = valin (6,9) |
| ■ H = His = histidin (2,1) | ■ W = Trp = tryptofan (1,1) |
| ■ I = Ile = isoleucin (4,6) | ■ Y = Tyr = tyrosin (3,5) |
| ■ K = Lys = lysin (7,0) | ■ B = Asx = Asp nebo Asn |
| ■ L = Leu = leucin (7,5) | ■ Z = Glx = Glu nebo Gln |
| ■ M = Met = methionin (1,7) | ■ X = Xxx = jakákoli |
| ■ N = Asn = asparagin (4,4) | ■ * = --- = stop |

■ Čísla v závorkách udávají průměrný procentuální výskyt dané aminokyseliny v proteinech

Typy třízení aminokyselin



Šok ze struktury proteinů

- V roce 1958 rozřešili **J. Kendrew** a **M. Perutz** (Medical Research Council Laboratory of Molecular Biology, Cambridge) strukturu myoglobinu v nízkém rozlišení
- V roce 1962 oba dostali Nobelovu cenu
- Struktura se však jevila záhadnou bez špetky logiky...



John Cowdery Kendrew (1917–1997)

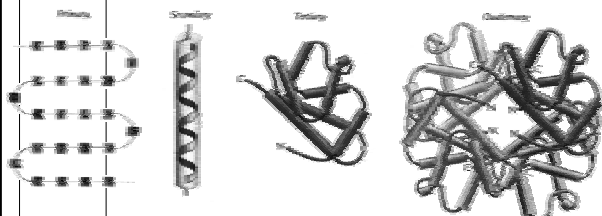


Max Ferdinand Perutz (1914–2002)



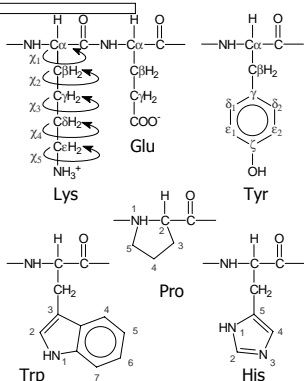
Od primární ke kvartérní struktuře

- **Primární** – pořadí aminokyselin v řetězci proteinu
- **Sekundární** – lokální 3D konformace hlavního řetězce
- **Terciální** – globální 3D struktura hlavního řetězce
- **Kvartérní** – globální 3D uspořádání všech podjednotek dané bílkoviny, tj. všech hlavních řetězců



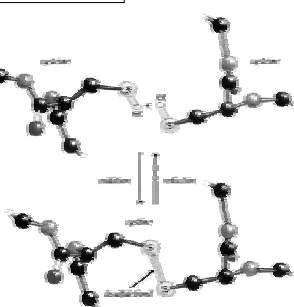
Konformace postranních řetězců

- V názvu aminokyselinového zbytku je přípona **-in** nahrazena **-yl**
- **Atomy** v postranním řetězci jsou často značeny písmeny **řecké abecedy** počínaje C nejbližším k CO
- Toto značení může být problematické a tak se užívá též **běžného chemického číslování**
- **Dihedrální úhly** se v postranním řetězci **značí χ** a **číslují počínaje úhlem mezi C_α - C_β**



Disulfidické můstky cysteinu

- Cysteiny mohou v oxidačním prostředí tvořit **sírné můstky**
- Většinou se **nenachází u intracelulárních** proteinů
- Jsou **časté u extracelulárních** proteinů
- Úlohou **S-S** můstků je:
 - **stabilizace struktury** (např. před degradací u inhibitorů proteáz)
 - **spojování řetězců** bílkovin (např. A a B řetězce insulinu)



Můstky – trans, gauche⁺, gauche⁻

■ Dihedrál ní úhel S-S vazby

je vždy ca. $\chi_1 = -60^\circ$
($+60^\circ$ působí kolize S, s C_α)

■ Levotočivý disulfidický

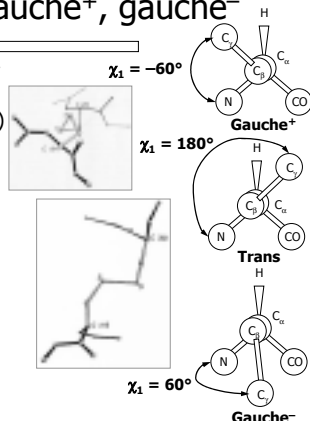
můstek $\chi_1 = -60^\circ$,
 $\chi_2 = -90^\circ$, $\chi_3 = -90^\circ$,
 $\chi_2' = -90^\circ$, $\chi_1' = -60^\circ$

■ Pravotočivý disulfidický

můstek $\chi_1 = -60^\circ$,
 $\chi_2 = +120^\circ$, $\chi_3 = +90^\circ$,
 $\chi_2' = -50^\circ$, $\chi_1' = -60^\circ$

■ Pravděpodobnost výskytu konformací S-S můstku:

GGG > GGT > TGT



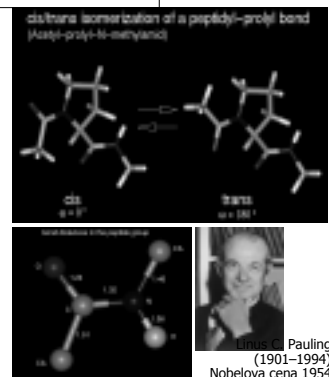
Cis/Trans forma peptidické vazby

■ Peptidická vazba má 40% **charakter dvojné vazby** a je víceméně **planární** (rezonanční efekt π -vazby s max. 85 kJ/mol vede ke zkrácení C-N o 14 pm a prodloužení o 2 pm u C=O)

■ Existuje **cis** ($\omega = 0^\circ$) a **trans** ($\omega = 180^\circ$) forma peptidické vazby

■ Pozice **trans** je díky **sterickým zábránám** o 8 kJ/mol **stabilnější** než cis

■ **Stabilní isomerisace** vazby se projevuje **pouze u prolinu!**



Geometrické parametry

■ Úhel omega ω – úhel peptidické vazby C'-N

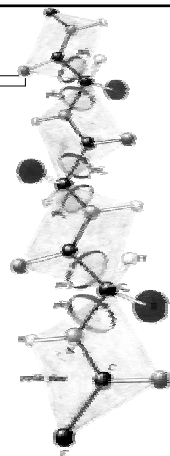
■ Úhel fi ϕ – úhel vazby N – C_α

■ Úhel psi ψ – úhel vazby C_α – C'

■ Konformace ψ vs. ϕ určuje typ sekundární struktury

■ Na ψ vs. ϕ je založen Ramachandranův diagram

■ **Glycin** může zaujmout mnoho rozdílných konformací (R = H), nepodléhá sterickému bránění, je silně konzervován v homologních proteinech



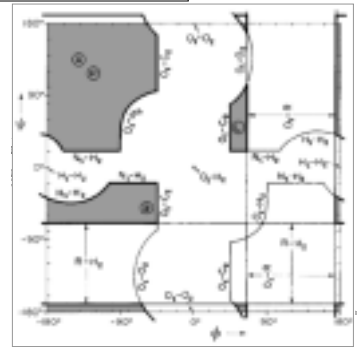
Ramachandranův diagram

■ Závislost torzního úhlu ϕ na ψ definuje **stericky dovolené oblasti v konformačním prostoru proteinu**

■ Pojmenován po indickém biofyzikovi z University Madras, G. N. Ramachandranovi (1922–2001) objeviteli struktury kolagenu (1954)



Gopalasamundran Narayana Ramachandran (1922–2001)



■ G. N. Ramachandran et al.: Stereochemistry of polypeptide chain configur. J. Mol. Biol. 7 (1962) 95–99.

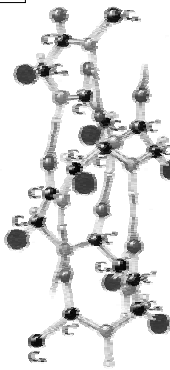
Proč existuje sekundární struktura?

■ Amino kyseliny uvnitř globulárních proteinů mají **hydrofobní postranní řetězce**

■ Vzniká **hydrofobní jádro** a **hydrofilní povrch** bílkoviny – to je ovšem problém...

■ **Hlavní řetězec je silně hydrofilní**, je **donorem vodíkové vazby NH**, a jejím **akceptorem C=O**, na každé peptidové jednotce

■ Hlavní řetězec je **neutralizován vazbou sama na sebe**, což **vede ke vzniku sekundární struktury** ve které participují NH a C=O skupiny



α -helix – základní informace

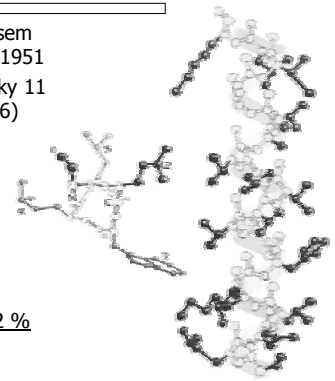
■ Teoreticky navržen Linusem Paulingem (MIT) v roce 1951

■ V průměru dosahuje délky 11 reziduí (od ca. 4 po až 56)

■ Přibližně 25 % struktury globulárních proteinů je tvořeno α -helixem

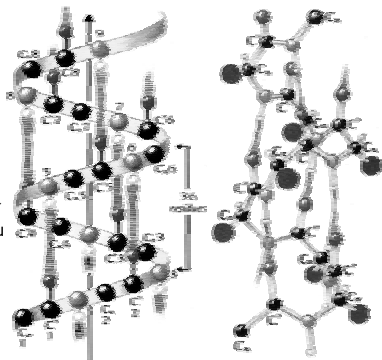
■ Helix je **pravotočivý** (levotočivost je pro L-aminokyseliny stericky bráněná)

■ V proteinech **existují 1–2 % krátkých** (3–5 reziduí) **levotočivých** helixů



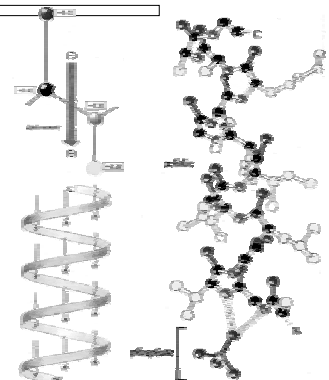
α -helix – technické parametry

- Úhel $\phi = -57^\circ$ (ca. -50° až -60°)
- Úhel $\psi = -47^\circ$ (ca. -40° až -60°)
- Počet reziduí na otáčku $n = 3,6$
- Výška závitů $p = 0,54$ nm
- n-tý C'=O se váže s n+4 NH zbytkem helixu, tj. výjma první a poslední skupiny jsou všechny C=O a NH skupiny vyvázané – konce helixů jsou polární a většinou umístěné na povrchu proteinu



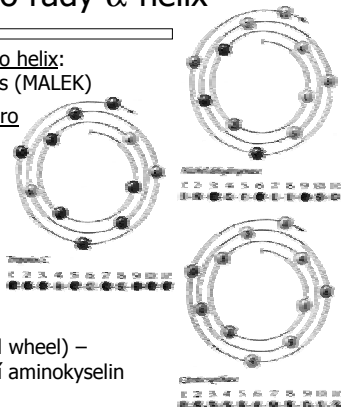
α -helix – fyz.-chemické vlastnosti

- Helix má s ohledem na nábojové parametry peptidické vazby **dipólový moment rovnoběžný s osou**
- C-konec je tak částečně **záporně** a N-konec částečně **kladně** nabitý
- Negativně nabitě ligandy se často **váží na N-konec** (díky geometrii a volným NH skupinám ideální pro fosfátové skupiny)
- Kladně nabitě ligandy se však na C-konec váží jen zřídka!



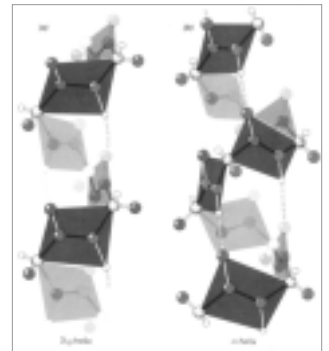
Aminokyseliny to rády α -helix

- **Aminokyseliny dobré pro helix:** Met, Ala, Leu, Glu, a Lys (MALEK)
- **Aminokyseliny špatné pro helix:** Pro (způsobuje ohyb v helixu), Gly, Tyr a Ser
- Nejčastější je umístění **helixu na povrchu proteinu** s jednou stranou obrácenou do roztoku
- **Helikální kruh** (helical wheel) – schématické znázornění aminokyselin v helixu (tj. po 100°)



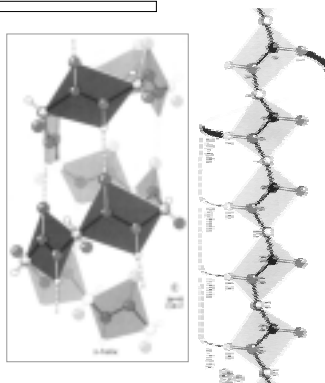
3_{10} -helix – technické parametry

- Úhel $\phi = -49^\circ$
- Úhel $\psi = -26^\circ$
- Počet reziduí na otáčku **3,0 a 10 atomů** je vázáno v kruhu (včetně H)
- Výška stoupání 0,6 nm, tj. je užší a protáhlejší než α -helix
- V otcích **končí α -helix 3_{10} -helixem**, či tento předchází β -struktúře



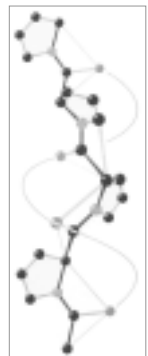
Exotické helixy

- **π -helix** – též $4,4_{16}$ -helix
 - úhly $\phi = -57^\circ$ a $\psi = -70^\circ$
 - výška stoupání 0,52 nm
- **2,2'-helix** – nikdy nebyl v proteinech pozorován
 - úhly $\phi = -78^\circ$ a $\psi = +59^\circ$
 - 2,2 rezidua na obrátku a 7 atomů v kruhu
- **3_{10} -helix a π -helix se vyskytují v globulárních proteinech vzácně**



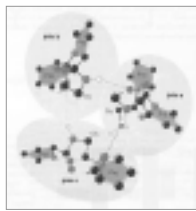
Poly-L-prolinový helix

- Většina nenativních poly-homopeptidů zaujímá **helikální strukturu**
- Pro má silné sterické bránění, nemá NH v kruhu – netvoří H-můstky
- **Helix polyprolin II** – za jistých podmínek precipituje jako levotočivý helix all-trans
 - 3 zbytků na otáčku a výška závitů 0,94 nm
 - Úhly $\phi = -75^\circ$ a $\psi = +150^\circ$
- **Helix polyprolin I** – konformace all-cis vede na pravotočivý helix
 - 3,3 zbytků na otáčku a výška závitů 0,63 nm
 - Úhly $\phi = -75^\circ$ a $\psi = +160^\circ$
- **Poly-L-glycin** překvapivě precipituje jako helix shodných parametrů s polyprolin II
 - pravotočivý i levotočivý (Gly není chirální)



Kolagenový helix

- **Trojitá šroubovice levotočivých helixů, svinutých do pravotočivé struktury superhelixu**
- Jde vlastně o kvartérní strukturu
- Konzervativní sekvence tripletů **–Gly–(X = Pro)–(Y = 4-hydroxyPro)–**
- Gly tvoří 30 %, Pro a Pro tvoří 15–30 %
- Úhly $\phi = -51^\circ$ a $\psi = -153^\circ$

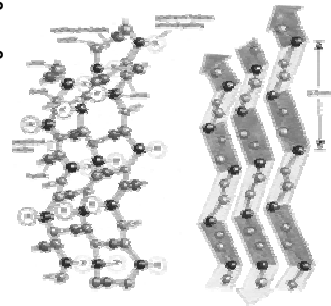


Každý 3. zbytek prochází osou superhelixu a musí být Gly. Vodíková vazba vzniká mezi NH glycinu a kyslíkem X. Pro uděluje rigiditu a trojšroubovice pevnost v tahu.

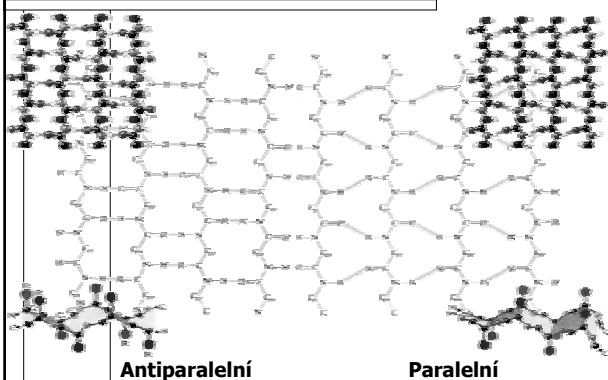


β-skládaný list – technické parametry

- **Antiparalelní**
 $\phi = -139^\circ, \psi = +135^\circ$
- **Paralelní**
 $\phi = -119^\circ, \psi = +113^\circ$
- **β-sheet** je de facto prvek terciální struktury, sekundární je **β-strand**
- Průměrná délka listu je 6 reziduí – tj. 2,1 nm (max. 15), v průměru je složen ze 6 vláken, tj. 2,5 nm (max. 15)
- Průměrný obsah v globulárním proteinu dosahuje 15 %

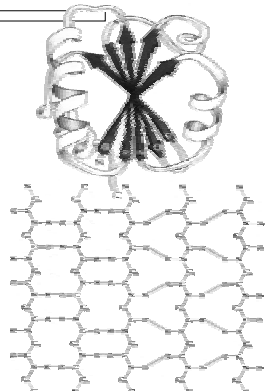


β-skládaný list – paralelní vs. anti



β-skládaný list – obecné vlastnosti

- Většina β-sheetů má **levotočivou torzi (twist) 10–25° a výsledný útvar je pravotočivý**
- **Směšené β-sheets** (na jedné straně β-strandu paralelní a na druhé antiparalelní) se vyskytují **pouze v 20 %**
- Průměrný obsah β-sheetu v proteinu je 15 %
- **Aminokyseliny dobré pro β-sheet uprostřed:** Tyr, Phe, Trp, Thr, Val, Ile (okraje jsou různorodé z důvodu možné agregace)



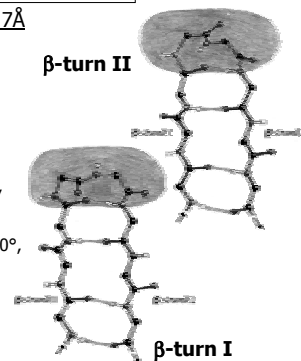
Nerepetitivní struktury

- Existují **otočky (turns)** a **smyčky (loops)** a **β-výduě (bulge)**
- Teoreticky předpovězeny Venkatachalamem (1968)
- Neexistuje zcela jednotné názvosloví (tight turns, **reverse turns**, β-turns, β-bends, hairpin bends, 3_{10} bends, kinks, widgets, etc.), ale ani klasifikace
- Je známo ca. 9 druhů otoček a smyček (často sporných) + jejich **zrcadlově symetrické formy**
- Charakterizace pomocí Ramachandranova diagramu je problematická...
- **Vlásačka (hairpin)** – otočka či smyčka **spojující spolu interagující protiběžné segmenty** sekundární struktury (např. loop mezi antiparalelními β-sheets se nazývá **β-vlásačka**)

■ C.M.Venkatachalam: Conformation of system of 3 linked peptide units. Biopolymers 6 (1968) 1425–1436.

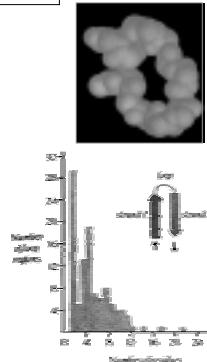
Nerepetitivní struktury – otočky

- **Obecně vzdálenost konců $< 7\text{\AA}$**
- **γ-turn** – vazba mezi CO(i) a NH(i+2), **velmi vzácné** v proteinech (častější v peptidech)
- **β-turn** – vazba mezi CO(i) a NH(i+3), **nejčastější**
 - Typ I – $\phi_2 = -60^\circ, \psi_2 = -30^\circ, \phi_3 = -90^\circ, \psi_3 = 0^\circ$, blízký deformovanému 3_{10} -helixu
 - Typ II – $\phi_2 = -60^\circ, \psi_2 = +120^\circ, \phi_3 = +90^\circ, \psi_3 = 0^\circ$
- **α-turn** – vazba mezi CO(i) a NH(i+4)
- **π-turn** – vazba mezi CO(i) a NH(i+5)



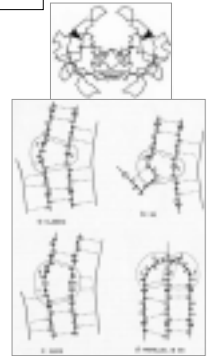
Nerepetitivní struktury – smyčky

- **Ohyb (bend)** – segment o délce 5 zbytků s otočením řetězce $> 70^\circ$
- **Ω -smyčka** – připomíná tvarem písmeno Ω , obsahuje 6–16 zbytků, vzdálenost konců je < 1 nm, téměř všechny proteiny s více než 60 aminokyselinami obsahují alespoň jednu smyčku
- **Smyčky (loops)** – vyskytují se vždy na povrchu, mají rozpoznávací úlohu – jsou často neuspořádané a volné a po vazbě ligandu se uspořádají



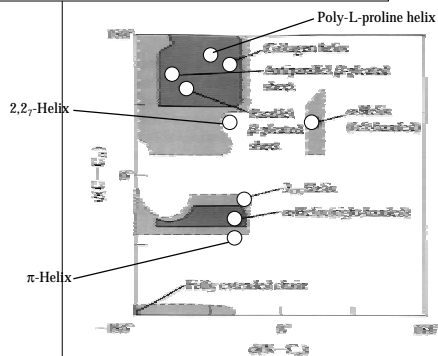
Nerepetitivní struktury – β -výduť

- **Zlom (výduť – bulge) v β -sheetu** způsobený rezidui s **torzí blízkou α -helixu**
- Oblast v β -sheetu se dvěma rezidui na jednom vlákně (1 a 2) a jedním na protilehlém (X)
 - Častá vidlicová vazba X_{CO} na 1_{NH} a 2_{NH}
- Pouze 5 % β -výdutí je mezi paralelními β -sheety (typ GX)
- Výduť zvyšují pravotočivou torzi struktury, často se vyskytují v aktivním místě, mohou být důsledkem mutací



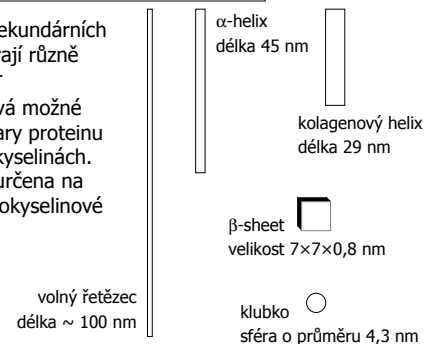
■ J. S. Richardson, E. D. Getzoff, D. C. Richardson, Proc. Natl. Acad. Sci. USA, 75 (1978) 2574–2578.

Ramachandran... – vyplněno!



„Velikost“ jednotlivých struktur

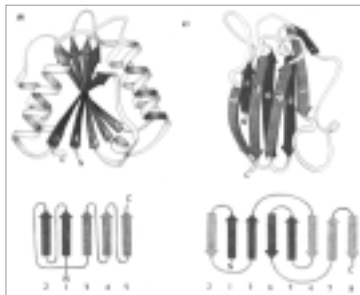
- Různé typy sekundárních struktur zabírají různé veliký prostor
- Obrázek udává možné velikosti a tvary proteinu o 300 aminokyselinách. Struktura je určena na základě aminokyselinové sekvence.



■ D. E. Metzler: Biochemistry, Academic Press, New York, 1977.

Jak se vyznat ve změní atomů?

- **Schematické diagramy** pochází od Arthura Laska (Cambridge, UK) a Jane Richardsové z roku 1969 – slouží k 3D schematické reprezentaci struktury
- **Topologické diagramy** publikovány v Nature 1976 Michalem Levitem a Cyrusem Chothiaou (Cambridge, UK) – určeny k 2D reprezentaci 3D uspořádání proteinu



Doporučená literatura

- C. Branden, J. Toonze: Introduction to protein structure. Garland Publishing, New York – London, 1991.
- D. Eisenberg: The discovery of the α -helix and β -sheet, the principal structural features of proteins. Proc. Natl. Acad. Sci. USA 100 (2003) 11207–11210.
- W. Kabsch, C. Sander: Dictionary of protein secondary structure. Biopolymers 22 (1983) 2577–2637.
- J. S. Richardson: The Anatomy and Taxonomy of Protein Structure. Advances in Protein Chemistry, vol. 34, Academic Press. <http://kinemage.biochem.duke.edu/~jsr/index.html>
- D. Voet, J. G. Voetová: Biochemie. Victoria Publishing, Praha, 1995.