

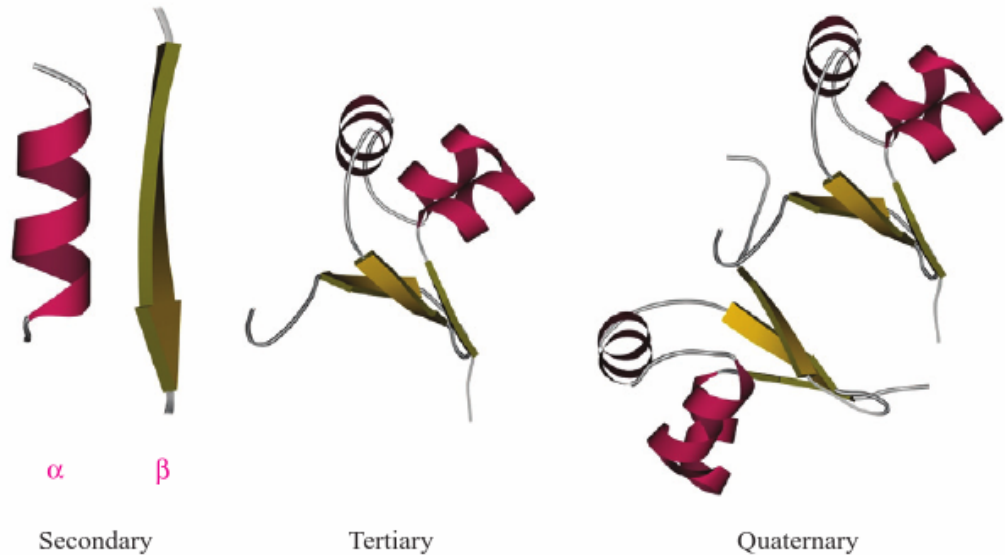
# **Proteiny – struktura a folding**

# Proteiny

- molekulární stroje, stavební bloky
- enzymatická katalýza
- regulační proteiny (řídí genovou expresi), receptory (akceptují intracelulární signály), strukturní proteiny (mikrofilamenta, mikrotubuly), membránové proteiny (transmembránový transport)
- enormní variabilita, vysoká specifita
- pro pochopení funkce nutná znalost 3D struktury
- složení buňky: **50% proteiny**, 15% sacharidy, 15% nukleové kyseliny, 10% lipidy, 10% ostatní

# Úrovně proteinové struktury

Primary ...-Gly-Val-Tyr-Gln-Ser-Ala-Ile-Asn-...



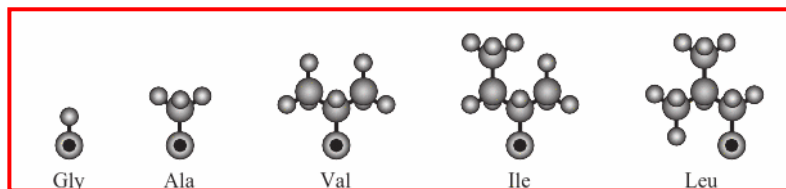
- primární
  - sekvence AK
- sekundární
  - pravidelná periodická konformace hlavního řetězce
- terciární
  - sbalení sekundárních struktur jednoho řetězce
- kvarterní
  - struktura tvořená více řetězci

# Aminokyseliny

rozdělení:

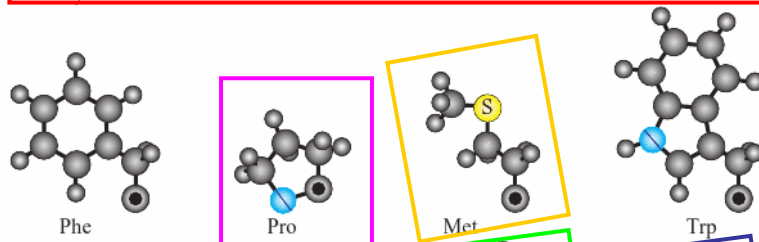
- **NPo: nepolární boční řetězec**
  - Ala, Val, Leu, Ile, Phe, Pro, Met, Gly, Trp, Tyr
- **CPo: nabitý polární boční řetězec**
  - Asp, Glu, His, Lys, Arg
- **UPo: nenabitý polární boční řetězec**
  - Ser, Thr, Cys, Asn, Gln

# Aminokyseliny

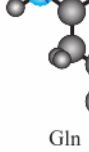
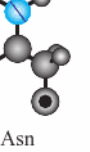
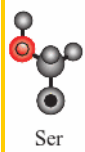
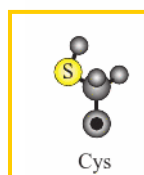


alifatický postranní řetězec

iminokyselina

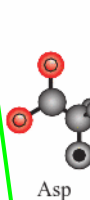
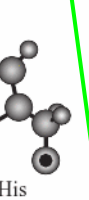
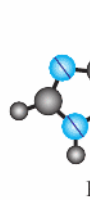
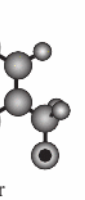
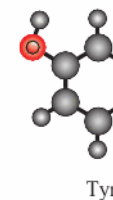


síra v postranním řetězci



karboxyl., amidová skupina (kyselé)

aromatické jádro, hydroxyl. skupina (ostatní)



amidová skupina (bazické)

**Figure 2.1.** The side chains of twenty standard amino acid residues (projecting from the main-chain C $\alpha$  atoms). Atoms forming the amino acids are shown on the right.

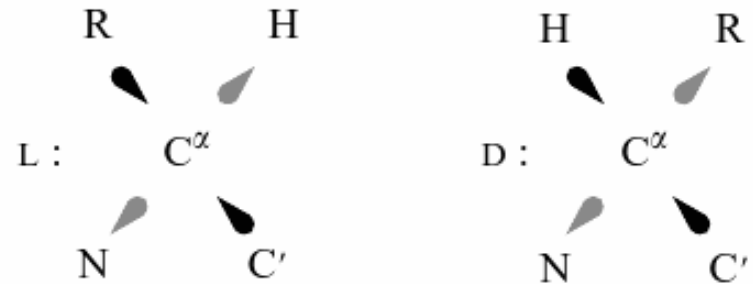
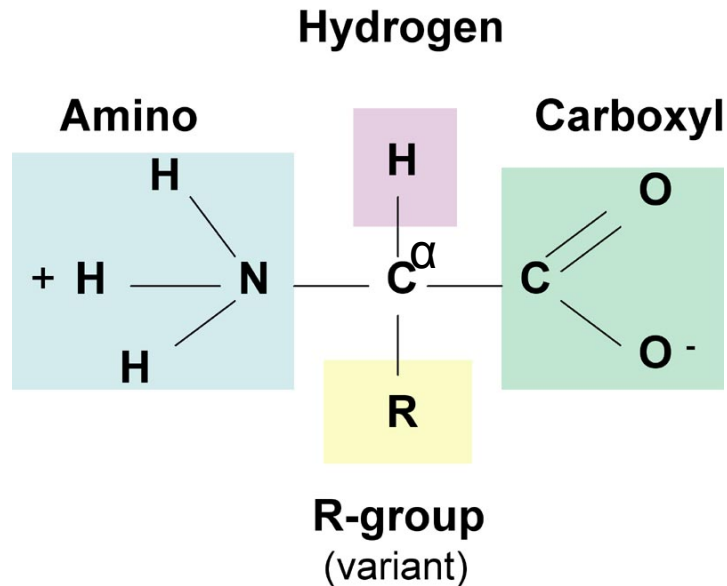
# Zastoupení AK v proteinech

Amino Acid	Freq. [%]
<b>Alanine</b> (Ala, A)	<b>8.1</b>
Arginine (Arg, R)	5.1
Asparagine (Asp, D)	5.2
Aspartic acid (Asn, N)	4.0
<i>Cysteine</i> (Cys, C)	1.2
Glutamine (Gln, Q)	3.8
Glutamic acid (Glu, E)	6.5
Glycine (Gly, G)	7.2
<i>Histidine</i> (His, H)	2.2
Isoleucine (Ile, I)	6.8
<b>Leucine</b> (Leu, L)	<b>10.3</b>
Lysine (Lys, K)	5.9
<i>Methionine</i> (Met, M)	2.5
Phenylalanine (Phe, F)	4.2
Proline (Pro, P)	4.3
Serine (Ser, S)	6.2
Threonine (Thr, T)	5.1
<i>Tryptophan</i> (Trp, W)	1.1
Tyrosine (Tyr, Y)	3.2
Valine (Val, V)	6.9

# Aminokyseliny

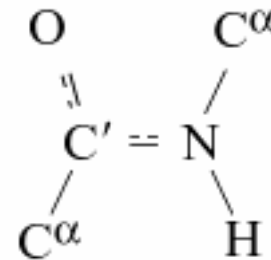
- jsou chirální (kromě Gly)
- v proteinech se vyskytuje pouze L-forma

## Amino Acid Structure



# Peptidová vazba

- je planární a má charakter částečné dvojné vazby
  - zajištěno  $sp^2$  hybridizací elektronů na atomech N a C'



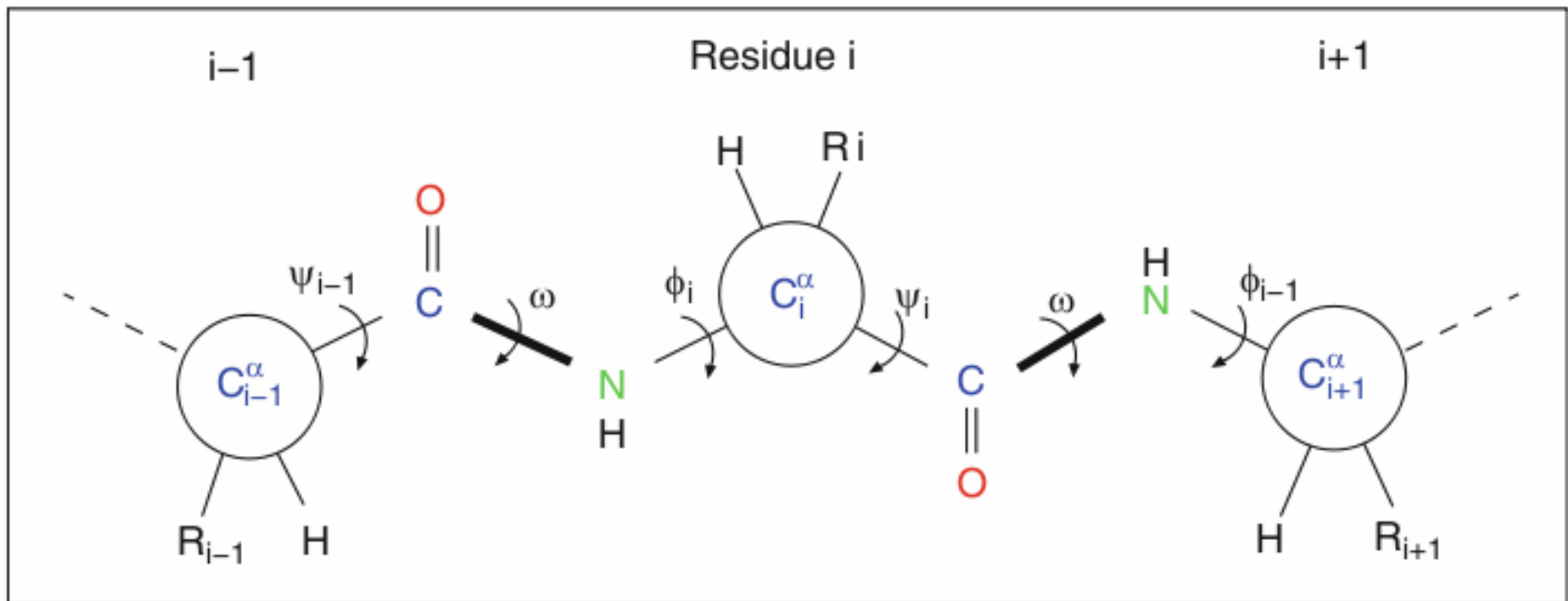
Kovalentní vazby – za normální teploty pevné, nevibrují

Kovalentní úhly – méně rigidní, vibrují za normální teploty, ale málo (cca 5 deg.)

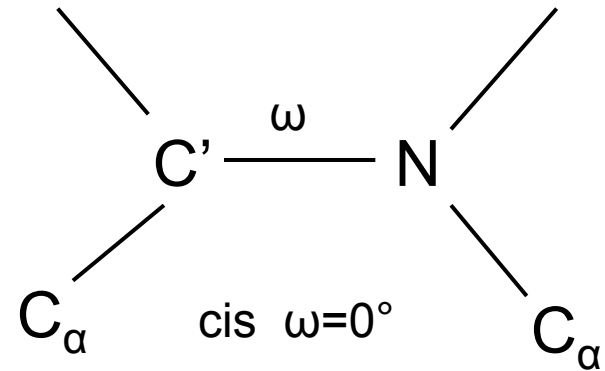
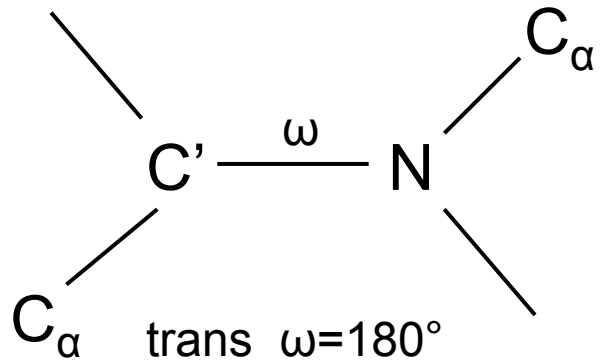
Dihedrální úhly – rotace kolem kovalentních vazeb, určují flexibilitu proteinů



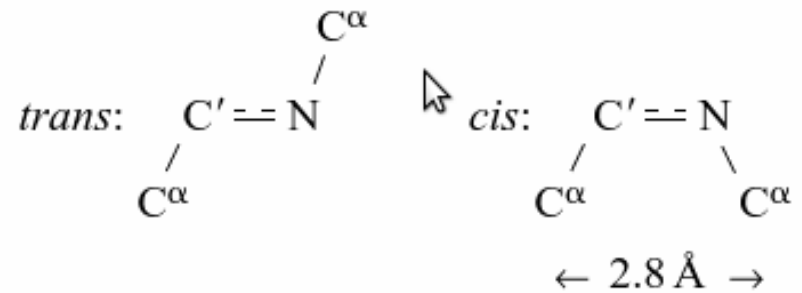
# Dihedrání úhly



# Dihedrální úhly



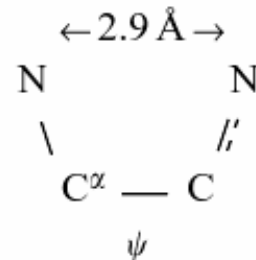
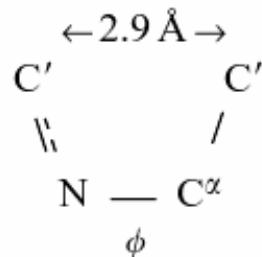
- energeticky je cis konformace nevýhodná ( $C_\alpha$  si překážejí)
- výjimka – prolin
  - trans zvýhodněna jen málo
  - globulární prot.: 90% trans, 10% cis



# Dihedrální úhly

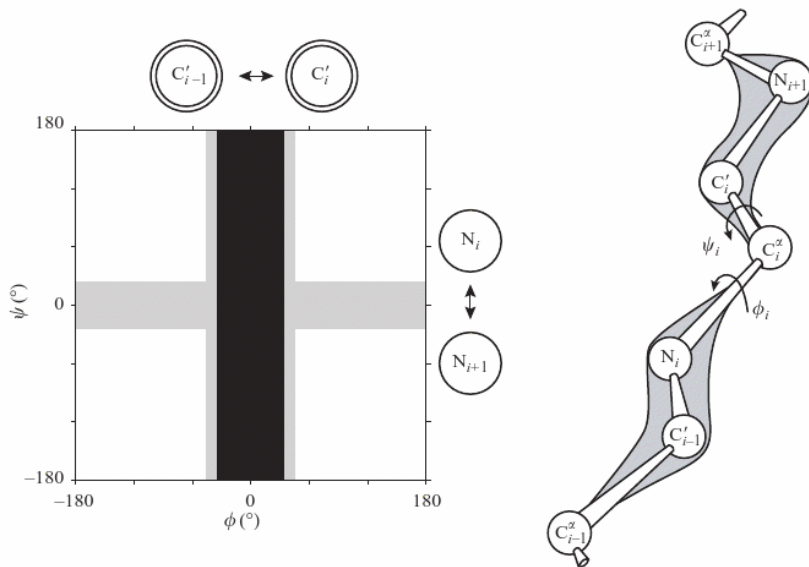
- na rozdíl od N-C', rotace kolem C<sub>α</sub>-C' a N- C<sub>α</sub> mají nízkou bariéru, a mohou tedy volně rotovat
- Ramachandranovy grafy
  - korelace úhlu φ (N- C<sub>α</sub>) a úhlu ψ (C<sub>α</sub>-C')
  - rotace kolem φ je obtížnější než rotace kolem ψ

$$2.9 \text{ \AA} < r_{\min}(\text{C}\cdots\text{C}) = 3.0 \text{ \AA} \quad 2.9 \text{ \AA} > r_{\min}(\text{N}\cdots\text{N}) = 2.7 \text{ \AA}$$



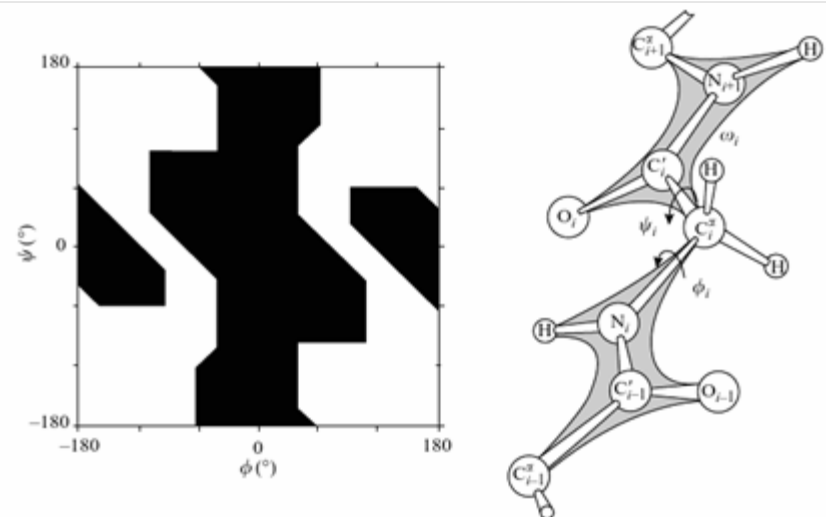
# Ramachandranovy grafy

obecný



**Figure 3.2.** This is how Ramachandran plots of the disallowed (■), strained (▨), and fully allowed (□) ( $\phi$ ,  $\psi$ ) conformations of the fragment  $C^{\alpha}C^{\prime}N-C^{\alpha}-C^{\prime}NC^{\alpha}$  would look, provided all these atoms had no other atoms attached (right) and atoms of residues  $i-1$  and  $i+1$  had no interactions.

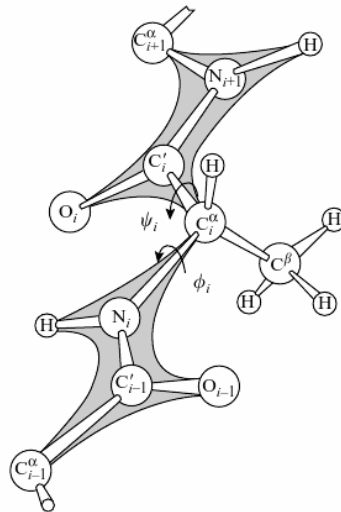
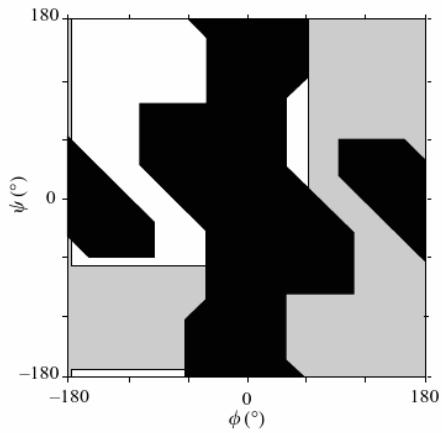
glycin



**Figure 3.3.** The map of disallowed (■) and allowed (□)  $\phi$ ,  $\psi$  conformations of glycine (Gly) in the protein chain.

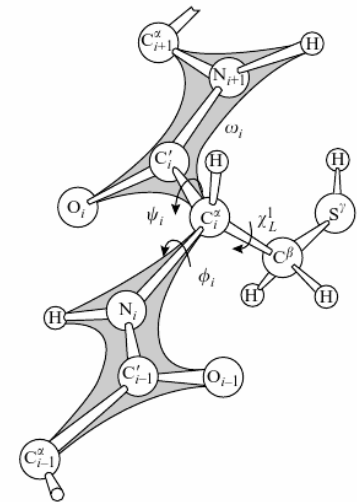
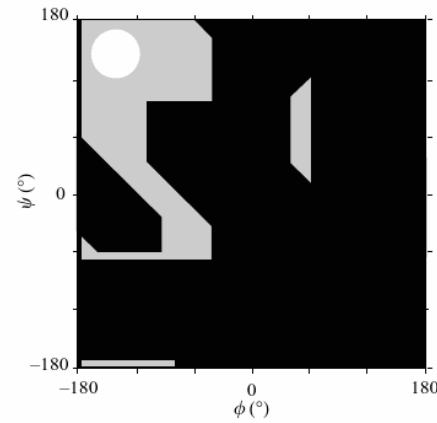
# Ramachandranovy grafy

alanin



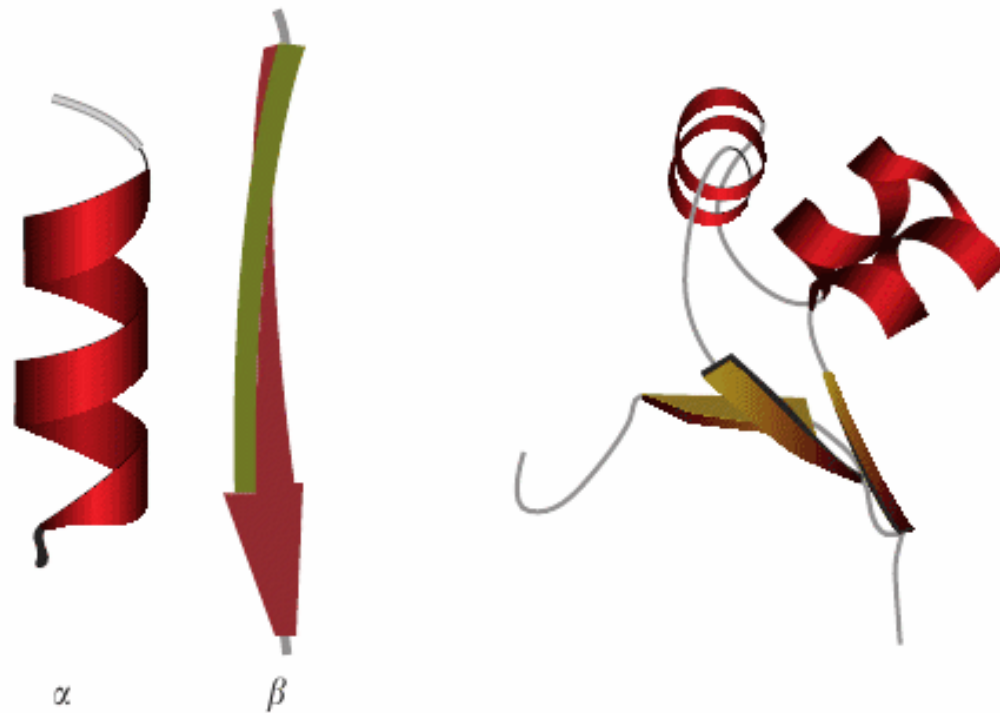
**Figure 3.4.** The map of allowed ( $\square$ )  $\phi$ ,  $\psi$  conformations of alanine (Ala) in the protein chain; ( $\square$ ), regions allowed for Gly only; ( $\blacksquare$ ), regions disallowed by main-chain interactions for all residues.

cystein



**Figure 3.5.** The map of disallowed ( $\blacksquare$ ) and allowed ( $\square$ ,  $\square$ )  $\phi$ ,  $\psi$  conformations of larger residues in the protein chain. ( $\square$ ), the region where all side-chain  $\chi^1$  conformations are allowed; ( $\square$ ), the region where some  $\chi^1$  conformations are disallowed.

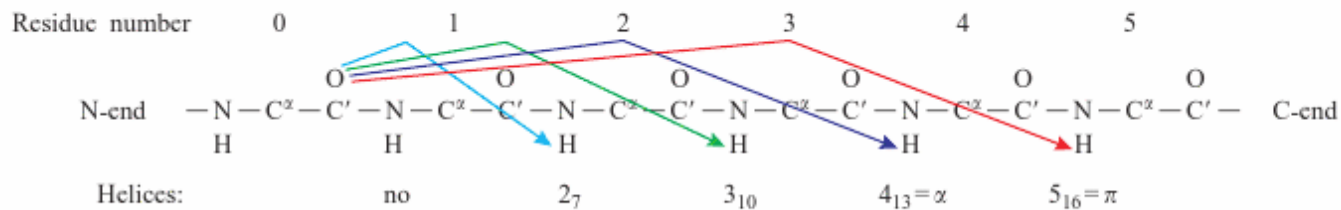
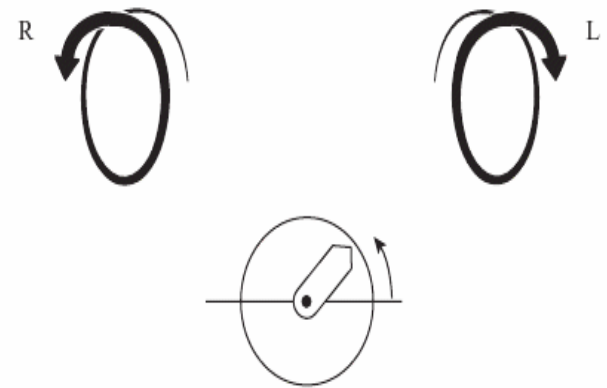
# Sekundární struktura



**Figure 7.1.** The secondary structures of a polypeptide chain ( $\alpha$ -helix and a strand of  $\beta$ -sheet) and the tertiary structure of a protein globule. Usually, taken together,  $\alpha$ - and  $\beta$ -structures make up about a half of the chain in a globular protein.

# $\alpha$ -helix

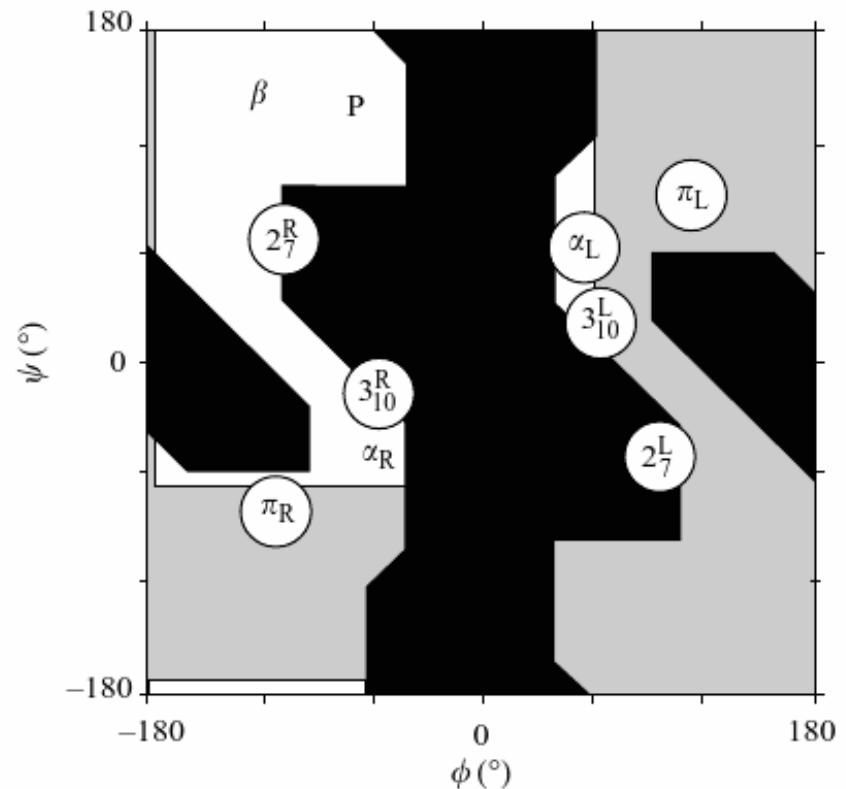
- obecně: pravotočivý, levotočivý
- stabilizace pomocí H-vazeb (mezi C=O a H-N jednoho řetězce)
  - existence řady helixů:  $2_7$ ,  $3_{10}$ ,  $4_{13}(=\alpha)$ ,  $5_{16}(\pi)$



**Figure 7.3.** Hydrogen bonds (shown with arrows) typical of different helices. The chain residues are numbered from the N- to the C-end of the chain.

# Typy helixů v proteinech

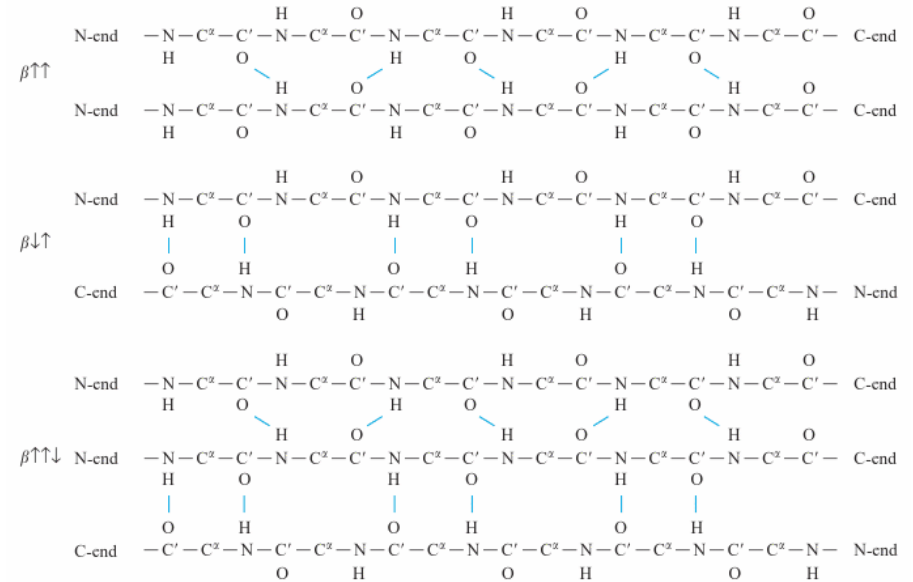
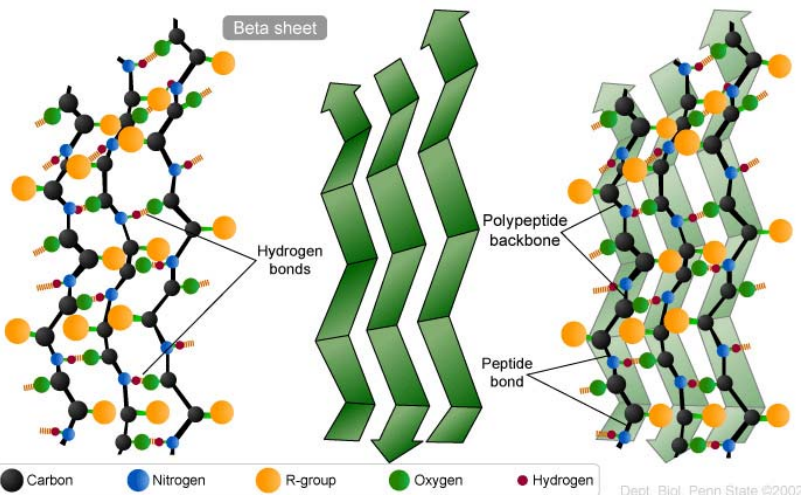
- pouze  $\alpha_R$ -helix je hluboko v oblasti povolených geometrií (ostatní buď na hraně, či povoleny jen pro Gly)
- $\alpha_L$ -helixy nejsou v proteinech pozorovány
- $2_7$  – velký úhel mezi N-H a O=C je nevýhodný pro H-vazby
- $\pi$ -helix – otáčky příliš široké, energeticky nevýhodné, vznik děr
- $3_{10R}$  ( $3_{10L}$  možný jen pro Gly) v proteinech přítomné, pouze ale na krátkých fragmentech (3-4 AK)



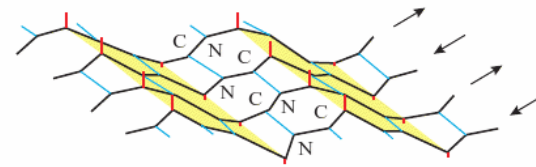


# $\beta$ -struktura

- paralelní nebo antiparalelní (nebo smíšená)
- stabilizace pomocí H-vazeb mezi jednotlivými řetězci  $\beta$ -struktury



**Figure 7.6.** Chain pathway and location of hydrogen bonds in the parallel ( $\beta\uparrow\uparrow$ ), antiparallel ( $\beta\downarrow\downarrow$ ) and mixed ( $\beta\uparrow\downarrow$ )  $\beta$ -structures. As shown, in each  $\beta$ -strand, the H-bonds (light blue) of one residue are directed oppositely to those of its neighbor in the chain.

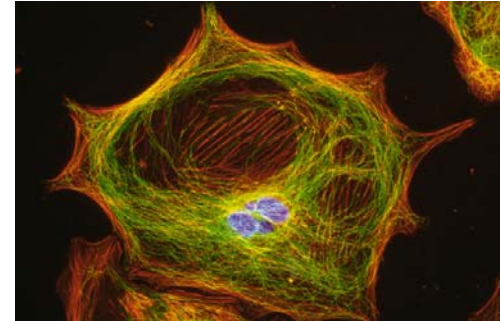


**Figure 7.7.** The  $\beta$ -sheet surface is pleated. The side-chains (shown as short red rods) are at the pleats and directed accordingly; i.e., the upward and downward side-chains alternate along the  $\beta$ -strand. The H-bonds are shown in light-blue. Adapted from [12].

# Rozdělení proteinů

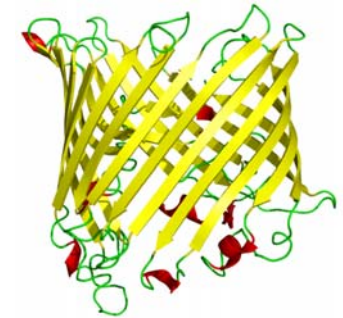
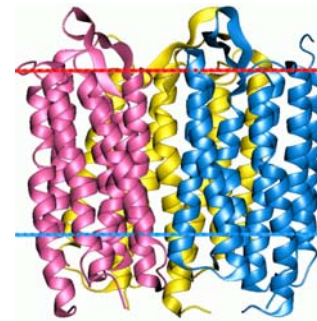
- **vláknité proteiny**

- tvoří veliké, obvykle dehydratované agregáty, struktura bohatá na H-vazby, pravidelná, drží díky interakcím mezi řetězci



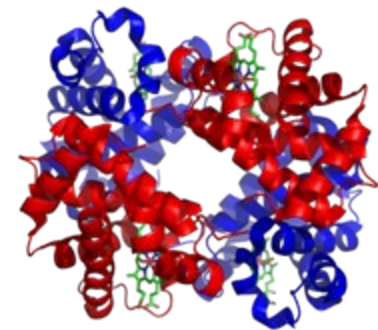
- **membránové proteiny**

- umístěny v membráně (dehydr.), pravidelná intramembránová část s vysokým podílem H-vazeb, omezení velikostí tloušťky membrány



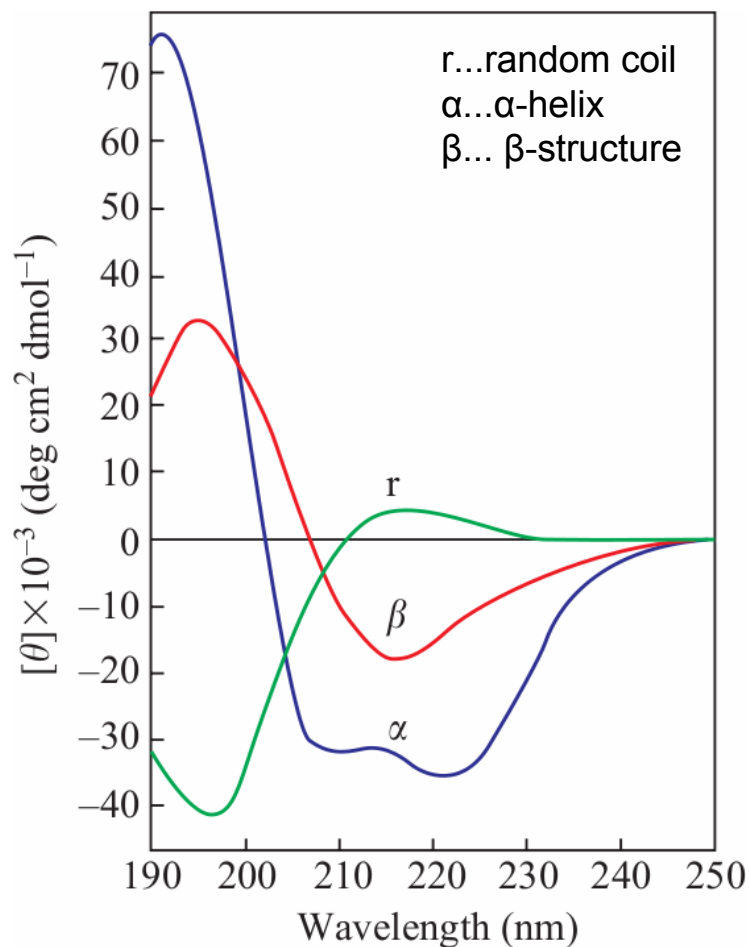
- **globulární proteiny**

- rozpustné ve vodě, méně pravidelné, strukturu stabilizují především meziřetězcové interakce

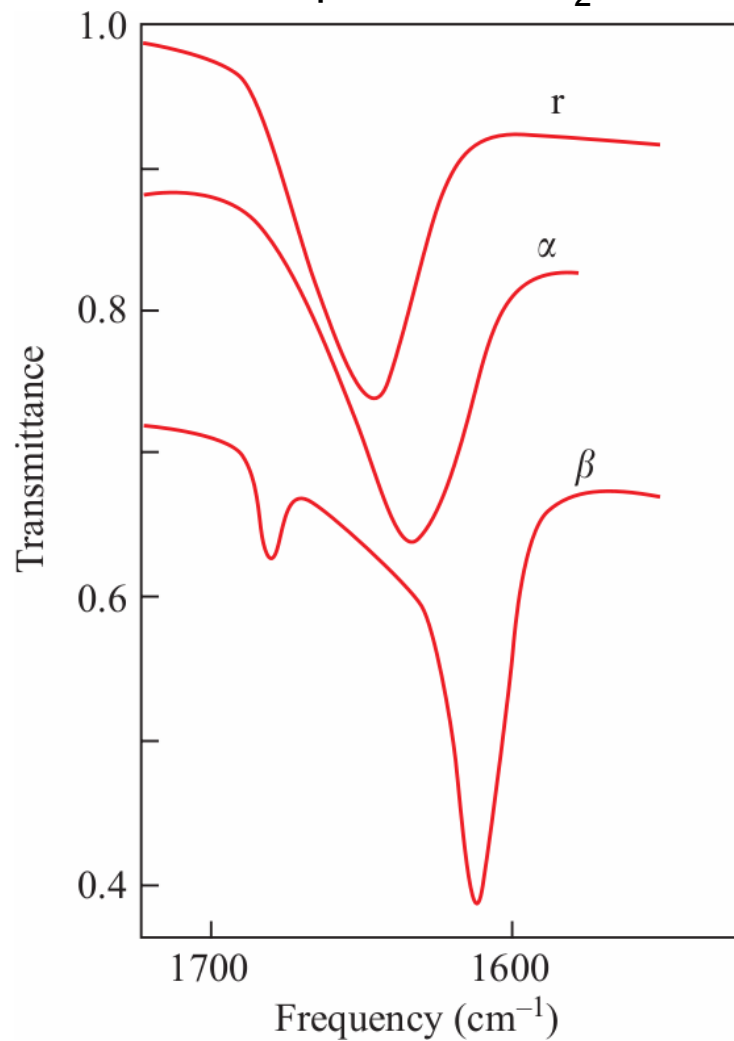


# Experimentální rozlišení sekundárních struktur

## UV CD spektrum



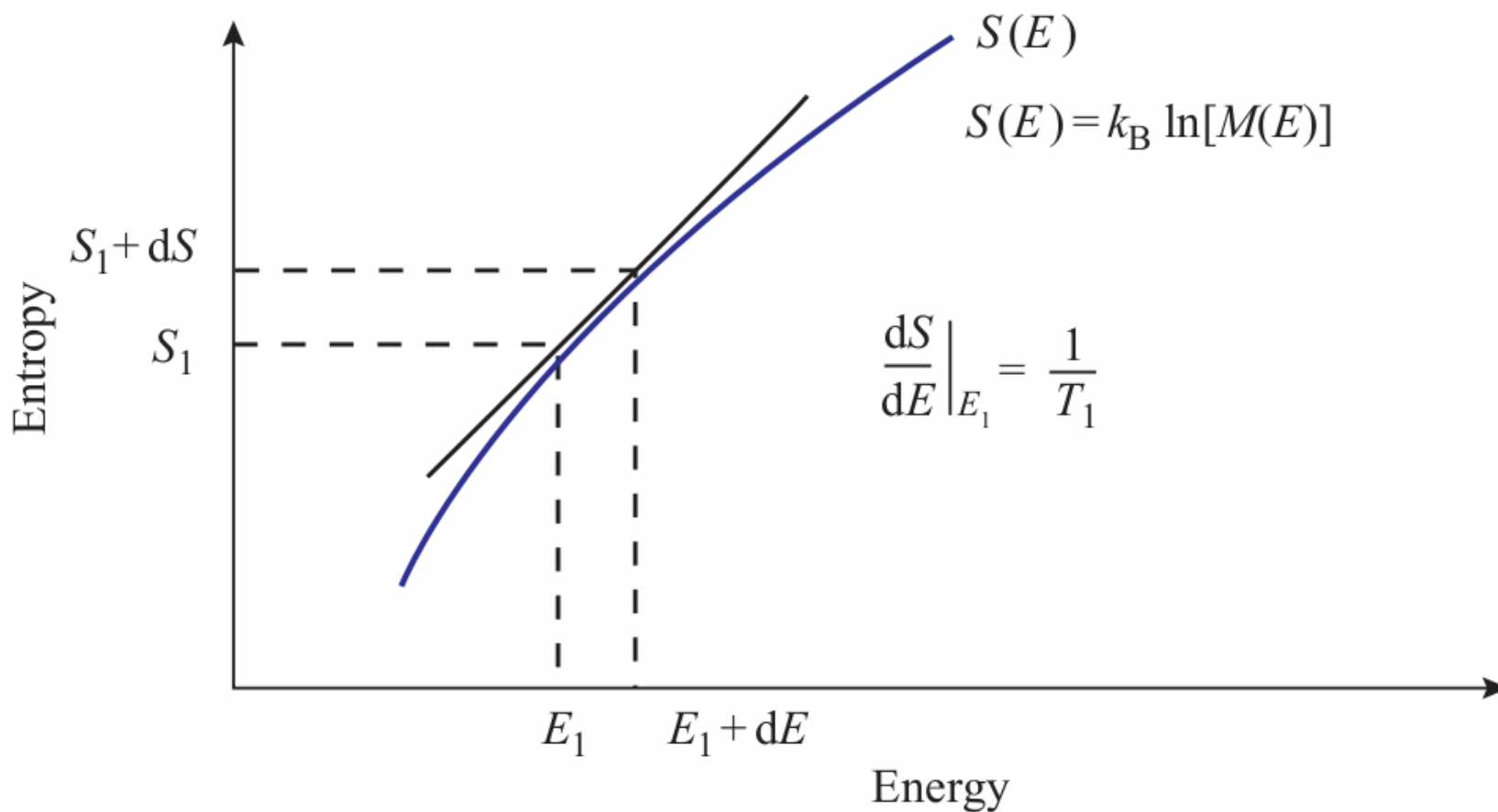
## IR spektrum v D<sub>2</sub>O



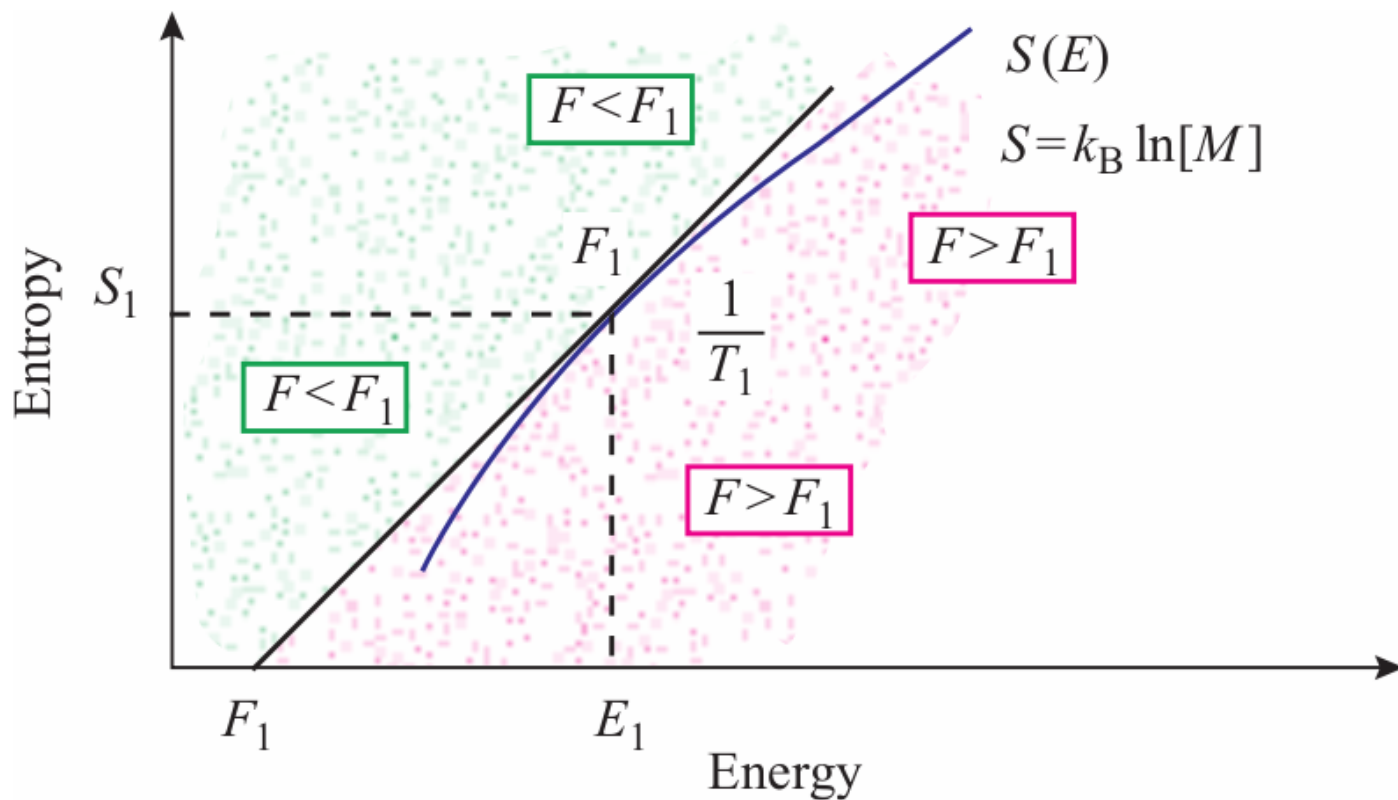
# System s mnoha stupni volnosti

- buď systém obsahující velké množství molekul, nebo systém tvořený jednou velkou a flexibilní molekulou
- nutný přístup pomocí statistické fyziky
- velké množství konfigurací
- entropie – kolik konfigurací (mikrostavů) odpovídá pozorovanému makrostavu
- teplota – úzce spojená s entropií

# Určení teploty velkého systému (termostatu)



Modrá křivka ukazuje závislost entropie  $S$  na energii systému  $E$ . Gradient  $dS/dE$  určuje teplotu  $T$  odpovídající energii  $E$ .  $M(E)$  je počet mikrostavů s energií  $E$ .

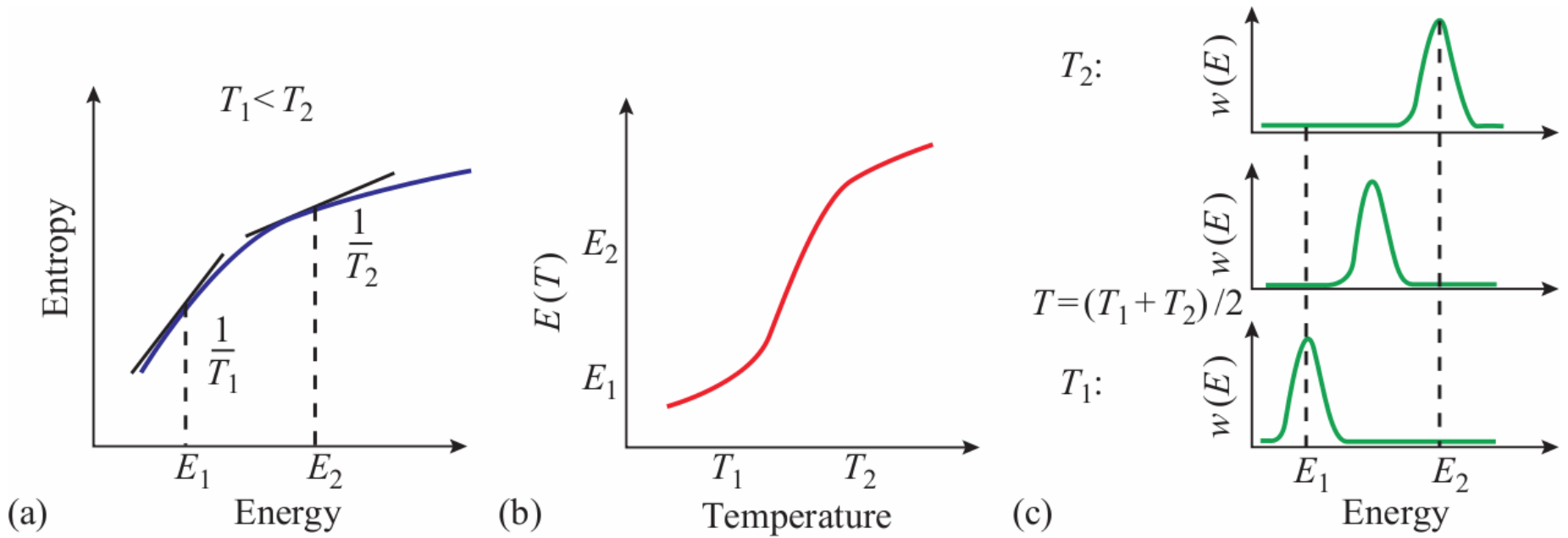


směrnice:  $S - S(E_1) = (E - E_1)/T_1$       $E - T_1 S = E_1 - T_1 S(E_1)$

$F_1 = E_1 - T_1 S(E_1)$  ... volná energie systému při teplotě  $T_1$  (konst. podél směrnice)

Protože je křivka  $S(E)$  konvexní (pod směrnici), odpovídá kontaktní bod minimu volné energie při teplotě  $T_1$

# Konformační změny: 1) pozvolná změna

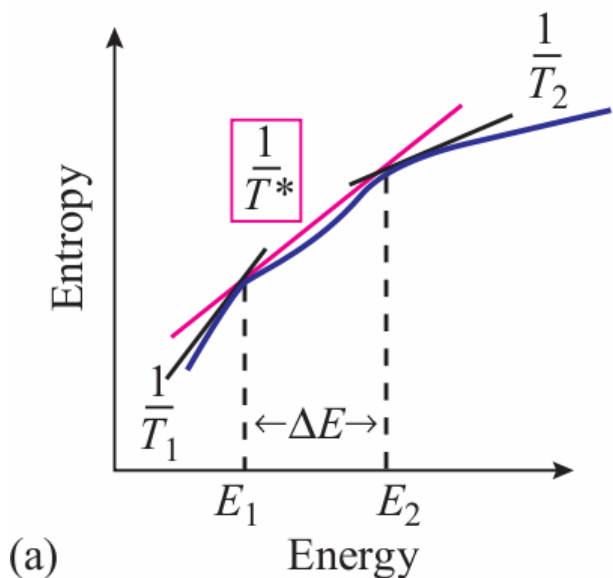


$w(E)$ ...pravděpodobnost energie  $E$  při dané teplotě  $T$

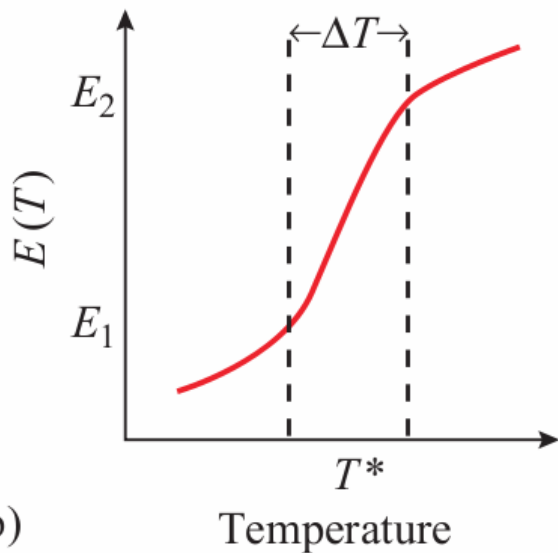
každá hodnota  $1/T$  odpovídá pouze jednomu bodu křivky  $S(E)$ , tedy stabilnímu stavu při dané teplotě  $T$ .

Při změně teploty systém postupně mění svůj termodynamický stav (entropii a energii)

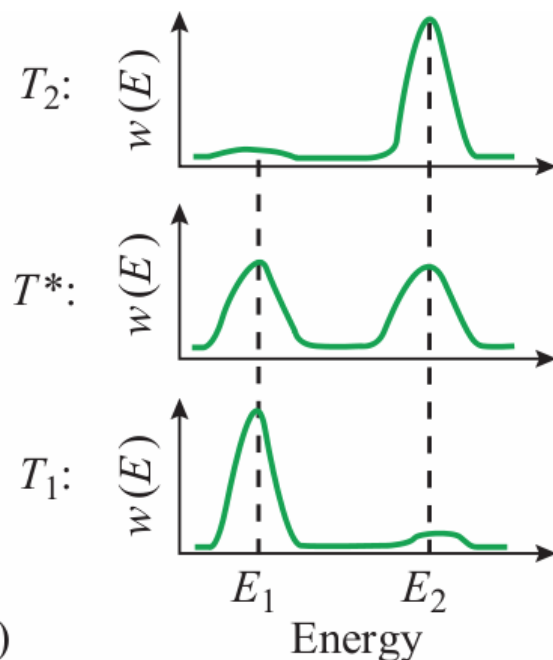
# Konformační změny: 2) prudká změna



(a)



(b)



(c)

Pokud gradient křivky  $S(E)$  s rostoucí  $E$  střídavě klesá a roste, pak existuje několik směrnic se stejným sklonem.

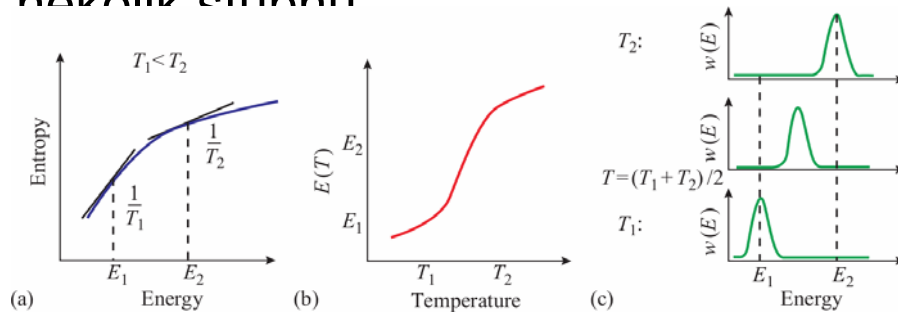
Při teplotě  $T^*$  budou mít struktury s nízkou a vysokou energií stejnou volnou energii a stejnou pravděpodobnost existence.

Každý systém se polovinu času nachází ve vysokoenergetickém a polovinu času v nízkoenergetickém stavu.

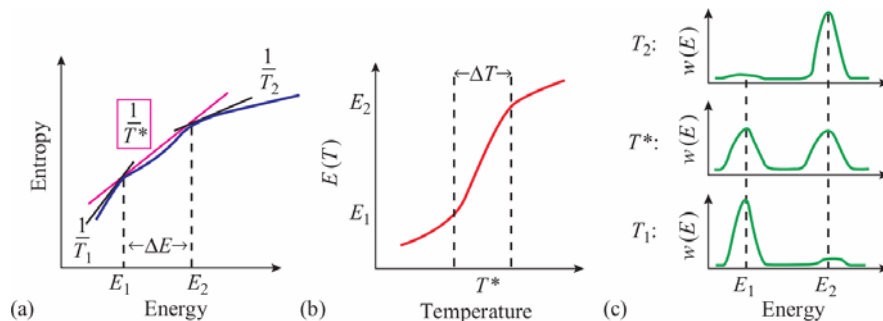


# Fázový přechod typu „všechno nebo nic“

- charakteristická je nepřítomnost intermediátů (oba stabilní stavy jsou odděleny energetickou bariérou)
- patří mezi fázové přechody prvního řádu (je to jeho mikroskop. analog)
- teplotní interval koexistence nízko- a vysokoenergetických stavů systému je pro malé systémy velmi malý, u makromolekul je to několik stupňů



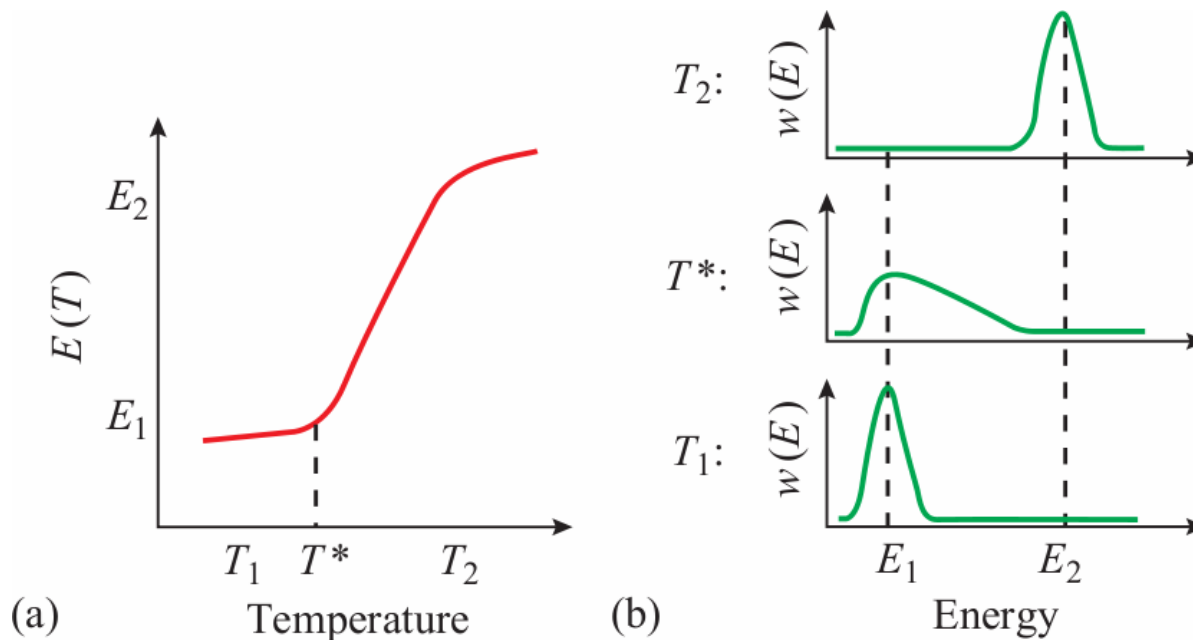
pozvolná změna konformace



fázový přechod 1. řádu

# Fázový přechod 2. řádu

- fázový přechod 1. řádu charakterizuje skoková změna energie (spolu s entropií, objemem a hustotou), typickou vlastností fázového přechodu 2. řádu je skoková změna tepelné kapacity



# Kinetika konformačních změn

## Co je pomalý a co rychlý proces?

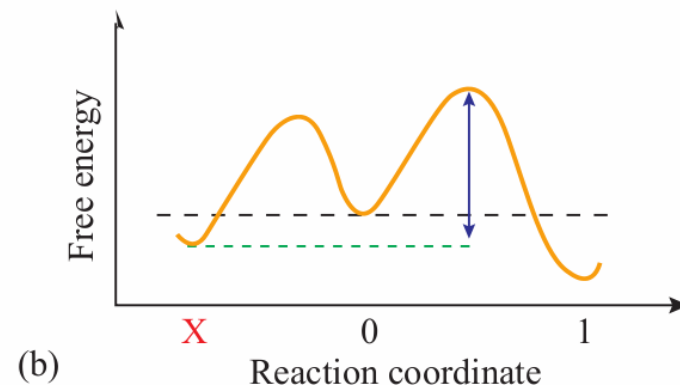
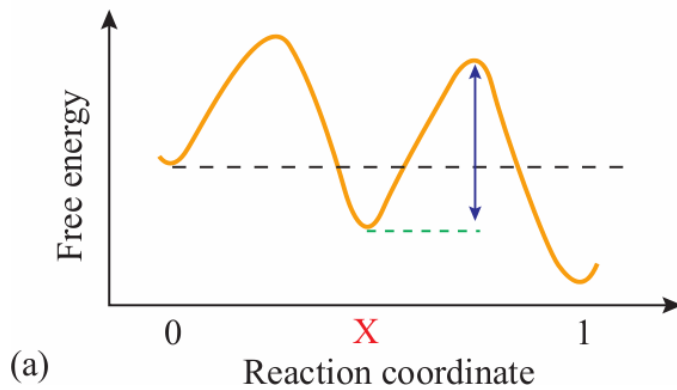
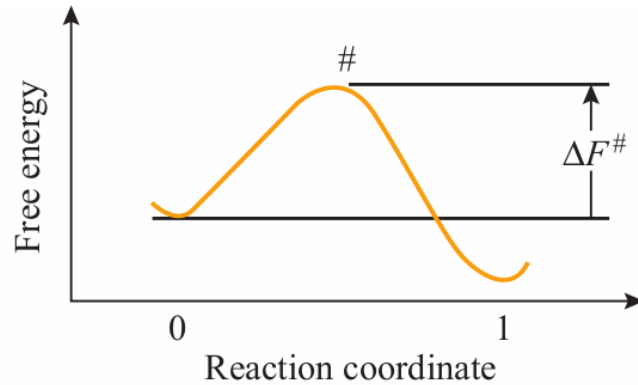
Proces je pomalý, pokud jeho rychlost je řádově pomalejší než výsledek získaný na základě rychlosti jednotlivých kroků a jejich počtu

Příklad: zapojení jednoho rezidua do sekundární struktury trvá 1 ns, řetězec obsahuje 1000 reziduí. Celý proces zapojení všech reziduí však netrvá očekávaných 1000 ns, ale 1000 s  $\Rightarrow$  proces je pomalý

Malou rychlost procesu mohou způsobovat:

- pomalá difúze při vysoké viskozitě
- nutnost překonání bariéry volné energie

# Bariéra volné energie

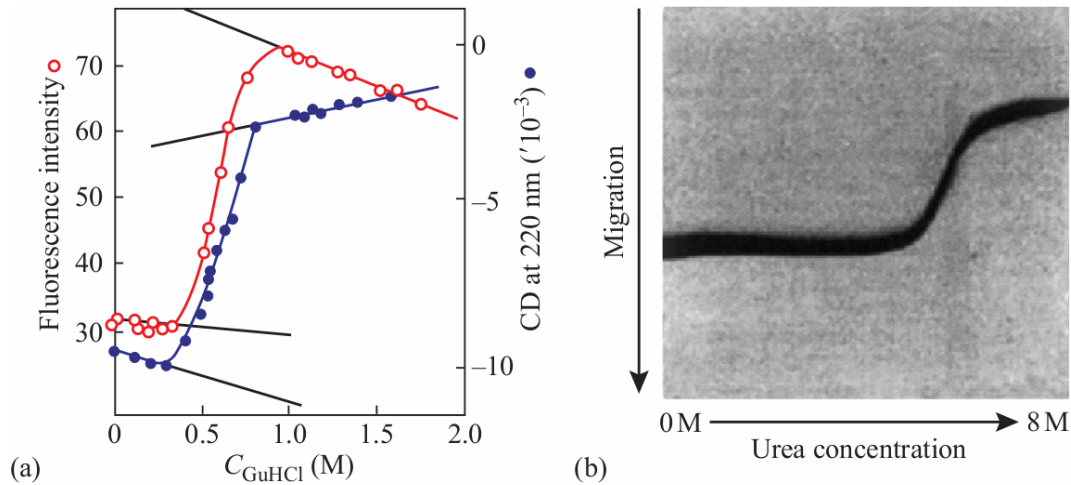


„pasti“ na křivce volné energie v podobě lokálních minim

# Denaturace proteinů

- in vitro: pomocí vysoké (nízké) teploty, denaturačních činidel apod.
- in vivo: umožňuje např. přechod přes membránu
- nejstabilnější formou proteinu nemusí být nativní forma
- protein nemusí mít fixovanou 3D strukturu (pouze při vazbách ligandů, DNA, RNA, apod.)
- denaturace: **fázový přechod „všechno nebo nic“** ⇒ pozorovatelnými kvantitami pouze nativní a denaturovaný stav
- renaturace je možná, pokud došlo pouze k mírným odchylkám od nativní struktury a protein není příliš velký
- reverzibilita procesu znamená, že informace o sekundární a terciární struktuře proteinu je obsažena již v sekvenci aminokyselin

# Denaturace proteinů



van't Hoffovo kritérium:  $\Delta E = \Delta H / N$

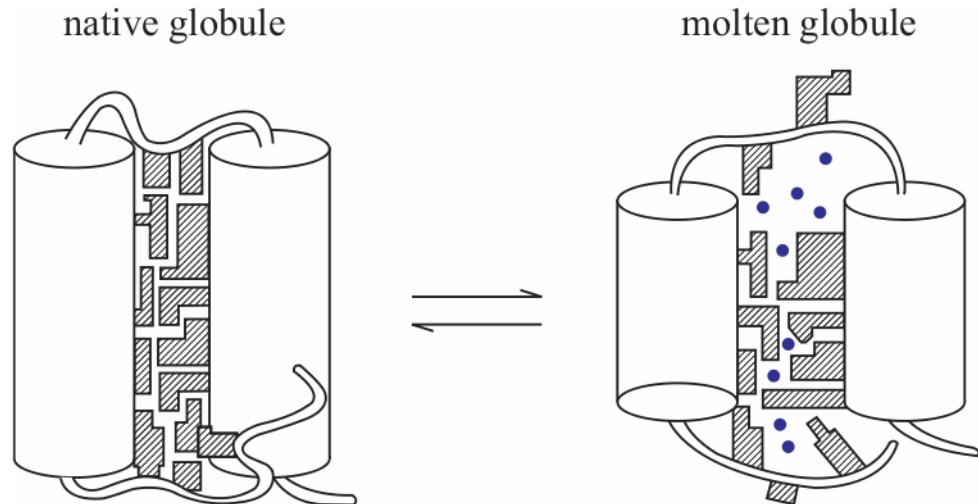
$\Delta E$ ...teplo spotřebované jednou „tající jednotkou“,  $\Delta H$ ...teplo spotřebované  $N$  molekulami

Tání celého proteinu probíhá jako fázový přechod „všechno nebo nic“

- $\Delta E < \Delta H / N$  ... tající jednotka je menší než celý protein, protein taje po částech
- $\Delta E > \Delta H / N$  ... tající jednotka je větší než celý protein, protein taje jako agregát molekul

# „Molten globule“

- představuje skoro nativní konformaci proteinu
- univerzální intermediát při sbalování a rozbalování proteinu (folding, unfolding)
- sekundární struktura je velmi blízká nativní formě proteinu
- malé uspořádání bočních řetězců
- méně kompaktní než nativní protein (jádro je však stejně kompaktní jak v nativním proteinu)
- výskyt solventu uvnitř globule

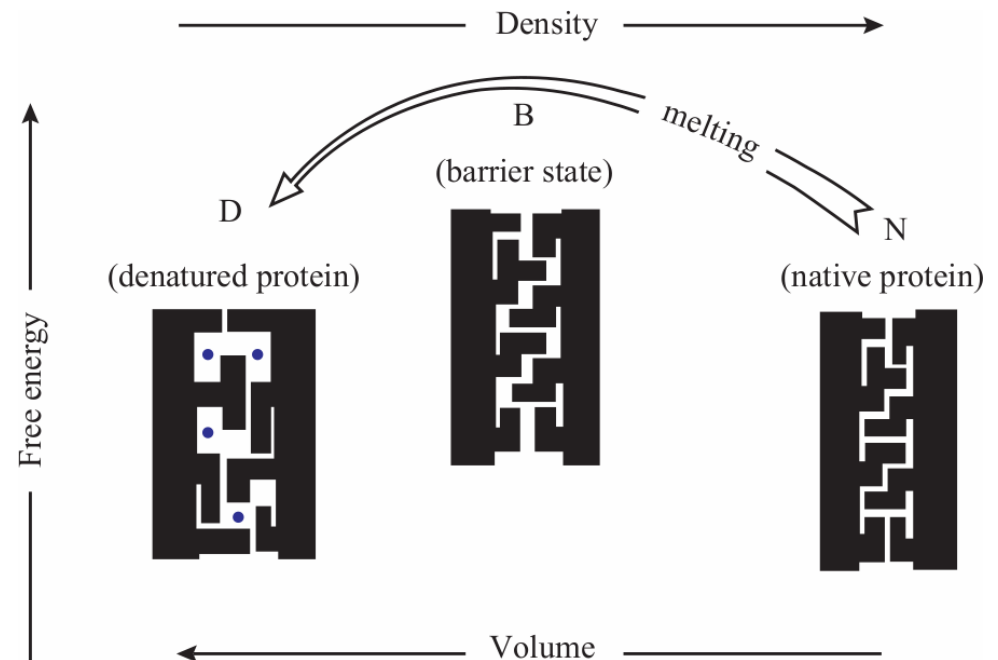


# „Molten globule“

Protože je tání proteinu fázový přechod 1. řádu, má mnohem vyšší energii a entropii než nativní stav (meziřetězcové interakce jsou mnohem slabší a mobilita řetězce mnohem vyšší)

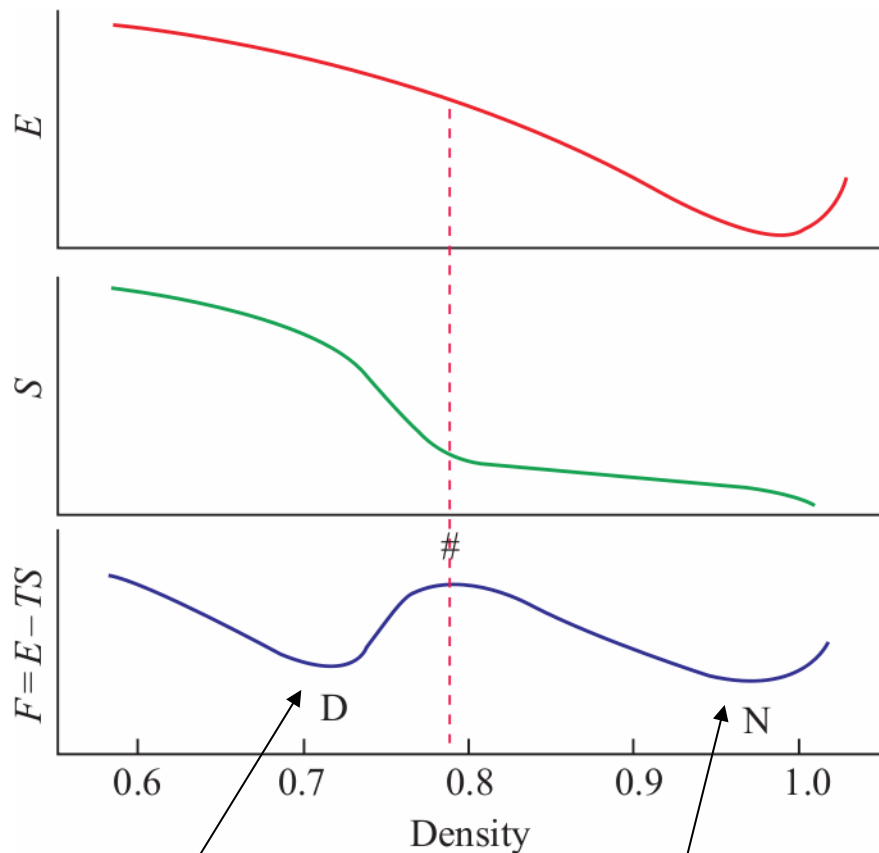
Většina stupňů volnosti proteinového řetězce spojena s malými fluktuacemi bočních řetězců, jejich osvobození je pro MG termodynamicky výhodné a vede k mírnému zvětšení MG (v porovnání s nativním proteinem)

Zvětšení objemu  $\Rightarrow$   
významný pokles VDW interakcí, nárůst volné energie



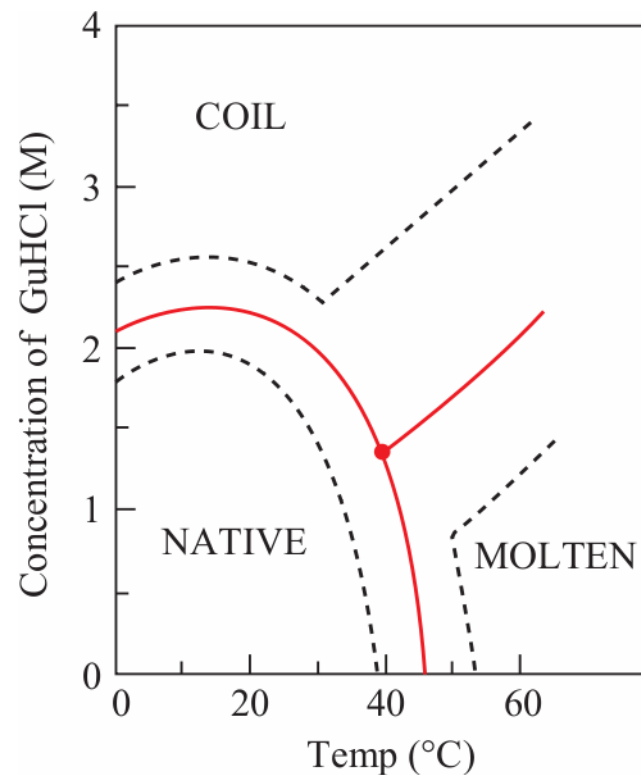


# Denaturace proteinů



denaturovaný stav

nativní stav



Fázový diagram konformačních stavů v lysozymu. Čárkovaně přechodové zóny.

# Protein folding

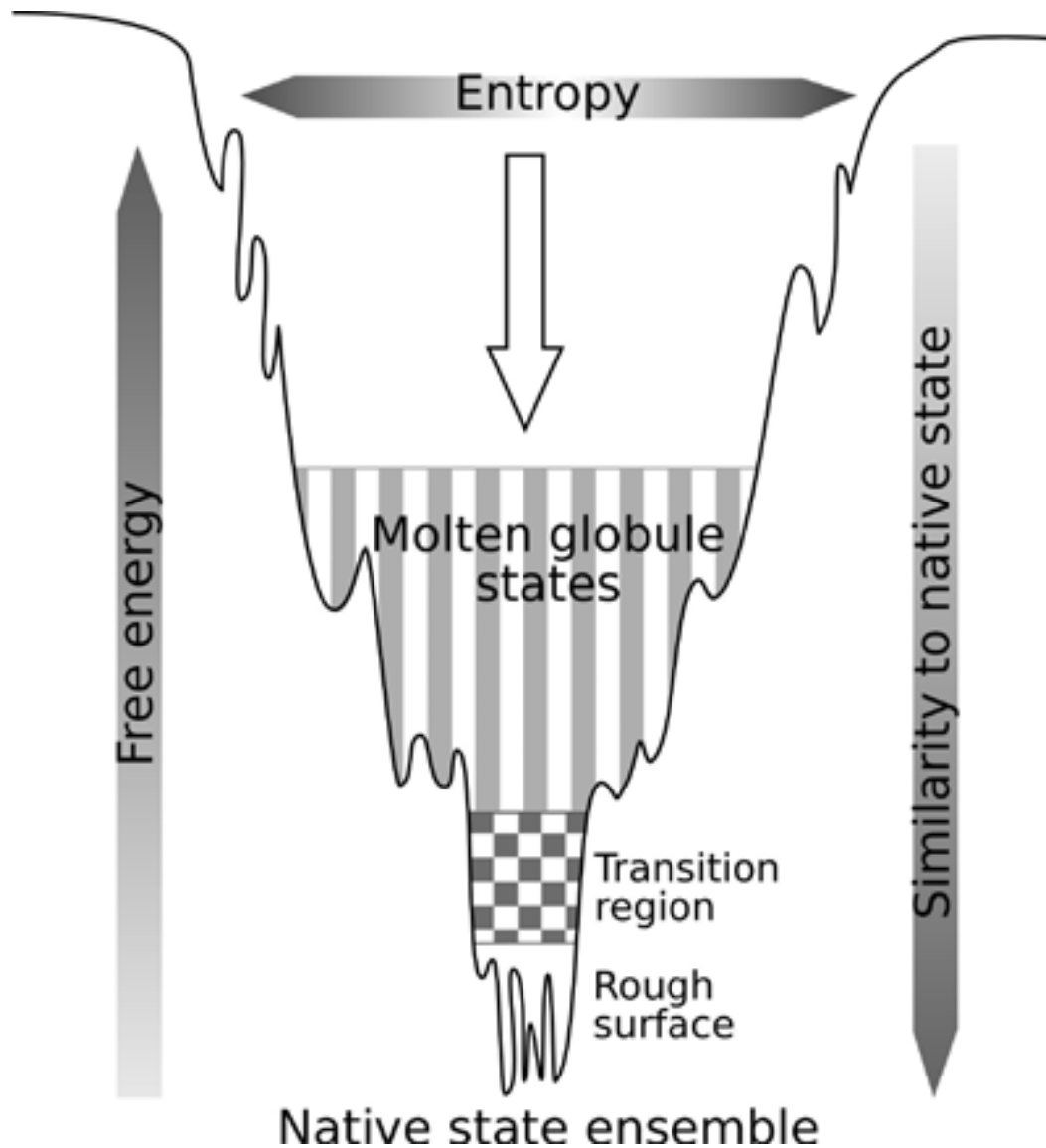
## Levinthalův paradox

- pokud by pro každé reziduum existovaly 2 možné konformace, pak pro řetězec se 100 rezidui existuje  $2^{100}$  alternativních struktur, a protože přechod z jedné konformace do druhé nemůže být rychlejší než 1 ps, prohledávání prostoru potenciální energie by trvalo nejméně  $\sim 2^{100}$  ps ( $\sim 10^{10}$  let)

**Otázka: Jak se dokáže protein sbalit do nativní formy během krátké doby (s-min)?**

Nativní forma proteinu je určena kineticky spíše než termodynamicky a jde cestou hledání snadno dosažitelného lokálního minima, než hledání globálního minima volné energie.

# Folding funnel



# Kinetika x termodynamika

- kinetika: velká rychlost foldingu, termodynamika: nejstabilnější struktura
- termodynamická kritéria:
  - folding je reverzibilní, probíhá jako fázový přechod 1. druhu
  - denaturovaný stav je nejčastěji náhodné klubko (random coil)
  - i za normálních podmínek je nativní stav proteinu jen o několik kcal/mol stabilnější než nesbalený

**Nativní stav: nízká energie**

**x**

**Náhodné klubko: velká entropie**

# Kinetika x termodynamika

Požadavky na rychlost foldingu:

- velká energetická mezera mezi nízkoenergetickými stavy a špatně uspořádanými (misfolded) stavy
- celý proces nesmí mít moc kroků
- nesmí obsahovat příliš vysoké energetické bariéry

Požadavky kinetiky i termodynamiky mohou být splněny současně: lze předpokládat, že v biologických procesech našly uplatnění právě ty proteiny, které se takto formovat dokáží.