

Lehký úvod do modelování proteinů

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky.



- struktura proteinu
- proteinové inženýrství
- predikce struktury
- potenciální energie
- knihovny rotamerů
- dead end elimination
- k čemu je mi to dobré

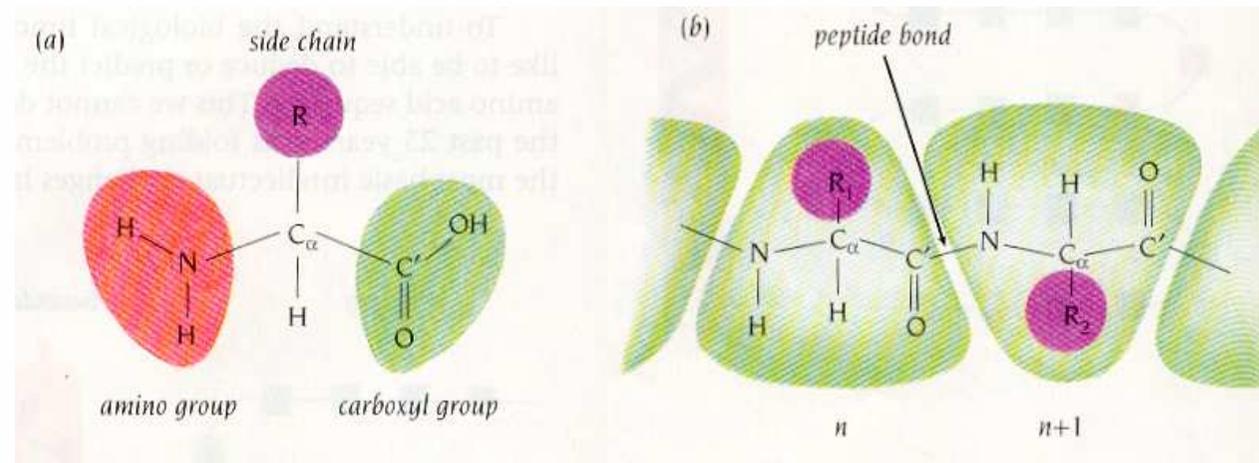
Proteiny



- složeny ze 20 základních aminokyselin
- protein - lineární řetězec aminokyselin
- aminokyseliny spojeny peptidickou vazbou
- pouze atomy H, O, N, C, S

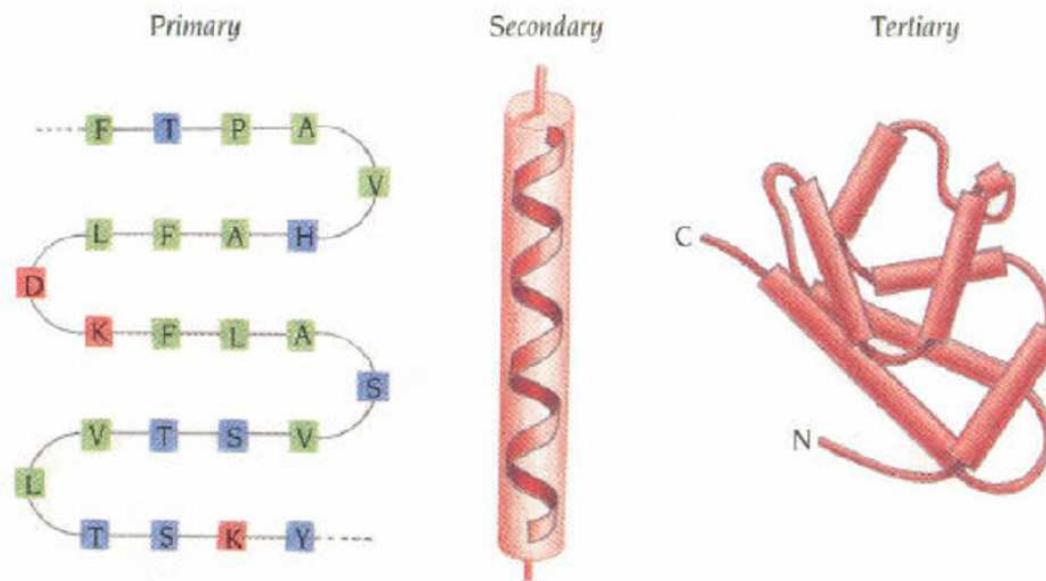
Pojmy

- páteř
- boční řetězce
- reziduum



Struktura

- primární – pořadí aminokyselin v řetězci
- sekundární – uspořádání řetězce
- terciární – výsledné prostorové uspořádání, poloha jednotlivých atomů





Predikce struktury

- dána sekvence aminokyselin (primární struktura)
- hledáme výsledné prostorové uspořádání (terciální strukturu)

Návrh proteinů (protein design)

- inverzní problém
- dána terciální struktura
- hledáme sekvenci aminokyselin



Přístupy

- simulace „sbalení proteinu“
 - velmi náročné (Folding@home)
- porovnávací modelování
- zjednodušení - často rozděleno do dvou fází:
 1. predikce páteře
 2. predikce bočních řetězců
- většina přístupů založena na minimalizaci nějaké energetické funkce



Kvantová mechanika

- + přesné
- použitelné jen pro malé molekuly

Klasická mechanika

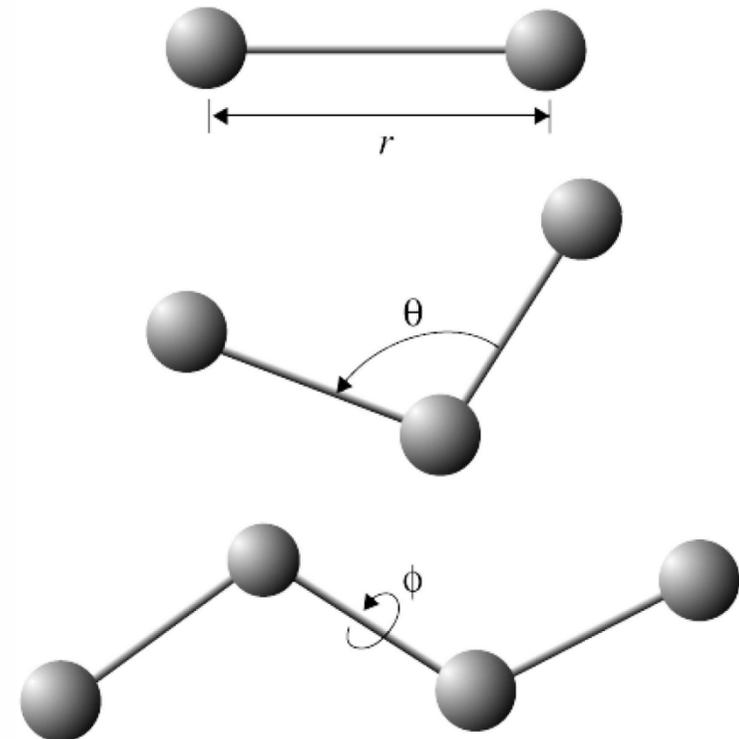
- molekula modelována jako soustava koulí spojených pružinami
- + „rychlé“ (použitelné i pro molekuly s desítkami tisíc atomů)
- nepřesné



Co zahrnuje výpočet potenciální energie U

- délka vazeb
- úhel mezi vazbami
- dihedral úhly

- nevazebné interakce
- Coulombovy interakce
- vodíkové můstky
- polarizace
- ...a co si kdo vymyslí





$$E_{\text{total}} = \sum_{\text{bonds}} K_b (b - b_{eq})^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_{eq})^2$$
$$+ \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Kde se berou parametry

- měření na reálných molekulách
- simulace malých molekul pomocí kvantové mechaniky



Vazby

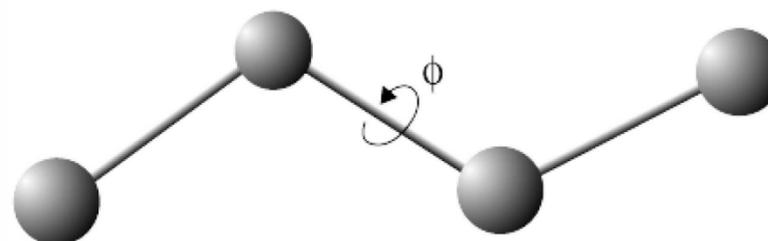
- počet vazeb v molekule s n atomy $\approx n$
- tvar $K_b(b - b_{eq})^2$ odvozen z Hookova zákona
 - $F = -K_b(b - b_{eq})$

Úhly

- obdobně jako vazby
- pro vazby i úhly existují další varianty potenciálů s více parametry

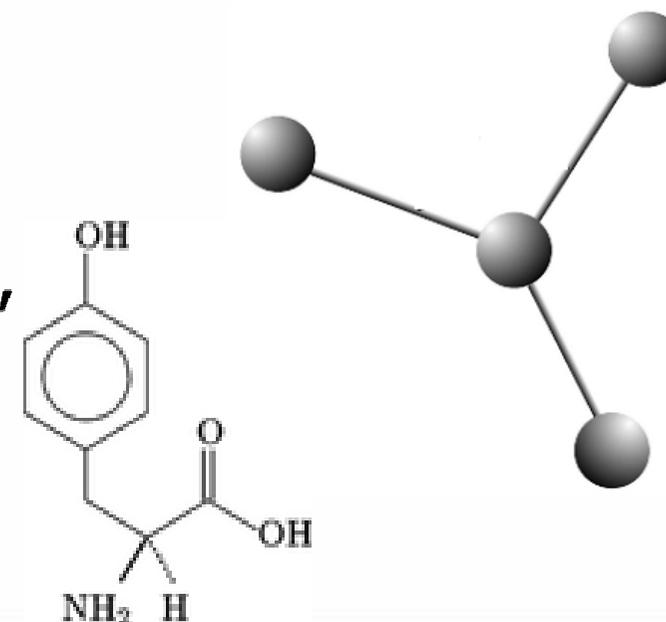
$$U_{dihedral} = \frac{V_n}{2} (1 - \cos(n\Phi - \gamma))$$

- V_n – silová konstanta
- n – periodičita (1 až 4)
- γ – fázový posun (0° nebo 180°)
- Φ – skutečný úhel



Nepravé dihedrály

- používají se pro vnucení tvaru, typicky pro vnucení planarity planárnímu jádru



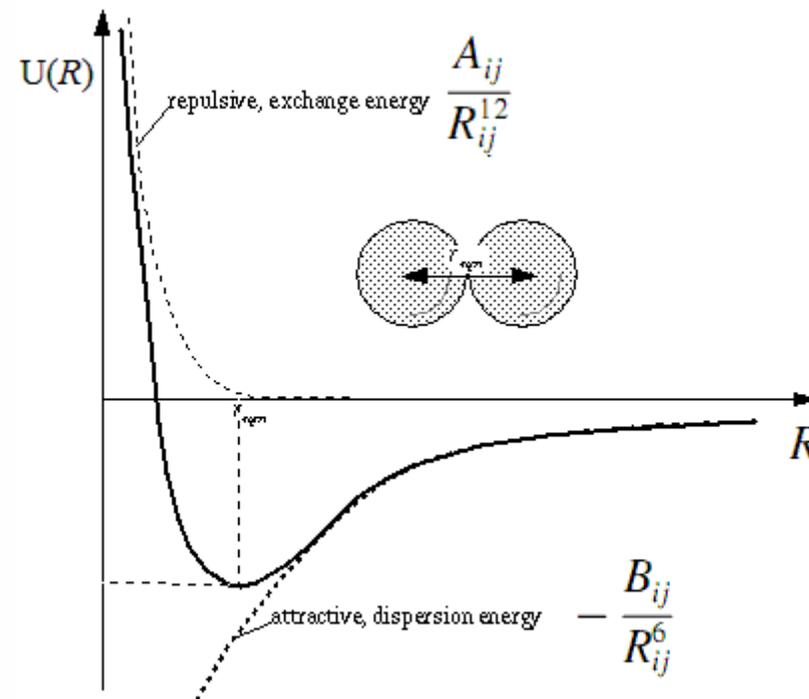
Nevazebné interakce



- atomy (molekuly) se:
 - na velkou vzdálenost přitahují, $U(R) \approx 1/R^6$
 - na malou vzdálenost odpuzují, $U(R) \approx -1/R^{12}$
- Lennard-Jonesovi 12-6 potenciály

$$U_{vdw} = \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6}$$

- výpočet koeficientů



Coulombovy interakce



- zjednodušení – atom aproximován bodovým nábojem

$$U_{coulomb} = \frac{q_i q_j}{\epsilon R_{ij}}$$

- q_i, q_j – částečné atomové náboje
- R_{ij} – vzdálenost atomů
- ϵ – permitivita prostředí (aproximace chybějících molekul okolí)



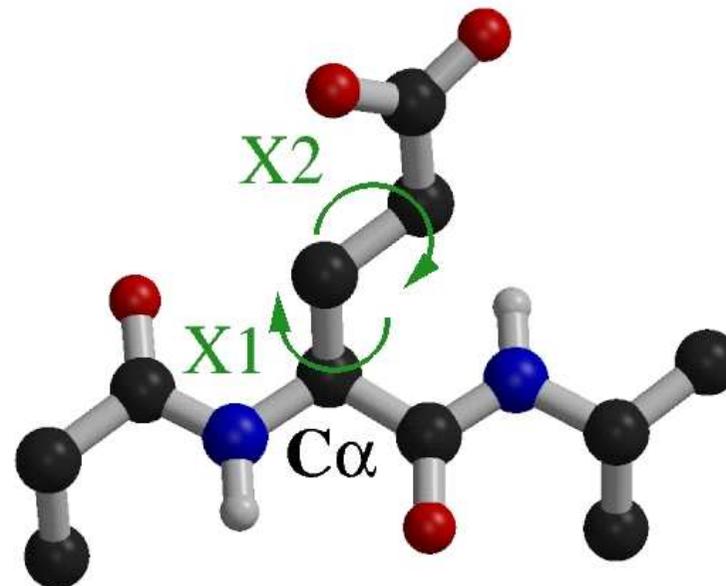
Poznámky

- zavedení typů atomů (např. různé typy uhlíku)
- $\vec{F} = -grad U, U = \int \vec{F} d\vec{s}$
- zjednodušující předpoklad – pair wise additivity
- spojitá změna energie
- minimální potenciální energie vs. stabilní stav proteinu
- výpočet stále složitý – minimalizace funkce s $3n$ proměnnými (vazby, úhly, dihedrály)

Další zjednodušení



- délky vazeb a velikosti úhlů se příliš nemění
→ použijeme střední hodnoty a neoptimalizujeme
- dihedrály – několik typických hodnot
→ vznik knihoven rotamerů
 - rotamer – reziduum (zbytek aminokyseliny)
s konkrétními hodnotami dihedral úhlů



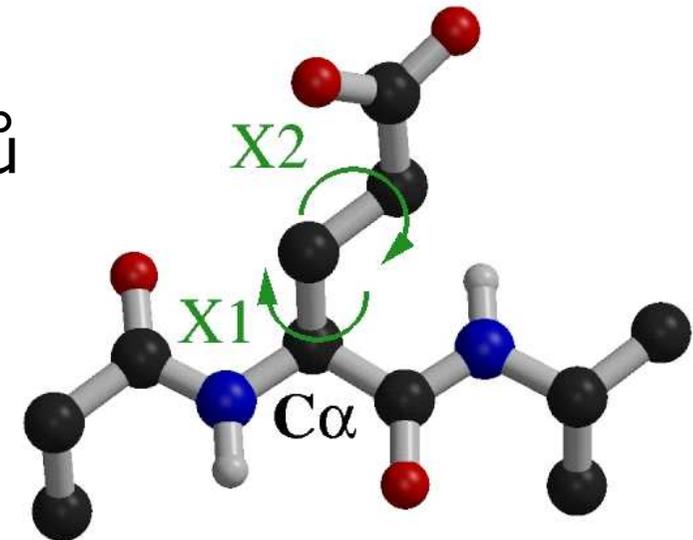
Knihovny rotamerů



- sada n-tic typických dihedral úhlů pro každé reziduum
- četnost výskytu každé n-tice
- odvozeny ze známých proteinů

Typy

- backbone-independent
- backbone-dependent
- secondary-structure-dependent





Vstup

- páteř proteinu (považována za pevnou)
- pro každou pozici (v primární struktuře) dána množina přípustných rotamerů (jednoho nebo více reziduí)
- potenciální funkce E

Úkol

- pro každou pozici i najít rotamer i_r minimalizující celkovou potenciální funkci

$$E_{\text{total}} = E_{\text{template}} + \sum_i E(i_r) + \sum_i \sum_{j, j < i} E(i_r, j_u)$$

- NP-těžké

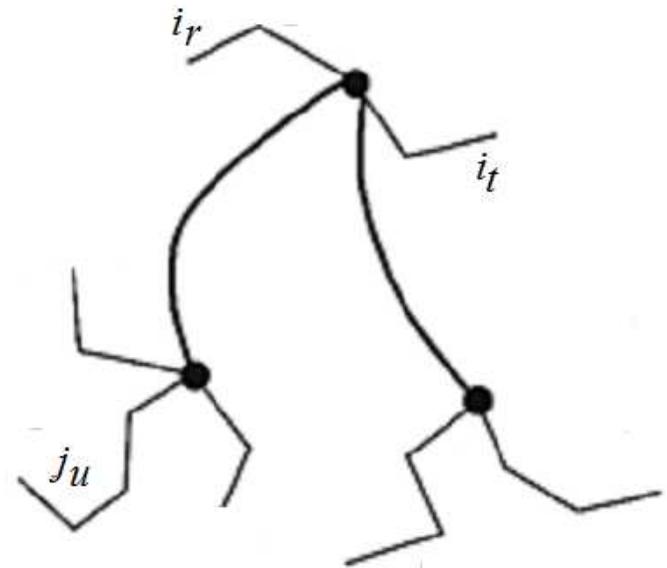
Dead end elimination (DEE)



- NP-těžké => nemůžeme procházet všechny možnosti
- snaha eliminovat „drahé“ rotamery

Princip DEE

- rotamer i_r může být eliminován rotamerem i_t , pokud platí



$$E(i_r) - E(i_t) + \sum_{j, j \neq i} \min_u [E(i_r, j_u) - E(i_t, j_u)] > 0$$

- $O(n^3 p^2)$, p – počet pozic, n – průměrný počet rotamerů na pozici

Dead end elimination



- existují další, silnější (a pomalejší) varianty DEE

Použití

```
while (se daří něco eliminovat){  
    while (simple DEE se daří něco eliminovat)  
        simple DEE;  
    while (DEE vyššího řádu se daří něco eliminovat)  
        DEE vyššího řádu;  
}  
Monte Carlo, nebo exhaustive search;
```



Optimalizace tunelů

- nalezen tunel v proteinu
- pro některá která rezidua obklopující tunel jsou vybráni možní náhradníci (jiné typy reziduí)
- hledáme věrohodnou mutaci proteinu, která zlepšuje tunel

+ nový nástroj, který by se prý chemikům hodil
- spousta práce, vzdálený obor



CENTRE
OF COMPUTER GRAPHICS
AND VISUALIZATION

UNIVERSITY
OF WEST BOHEMIA

PLZEŇ
CZECH REPUBLIC

<http://herakles.zcu.cz>

Díky za pozornost

Otázky?