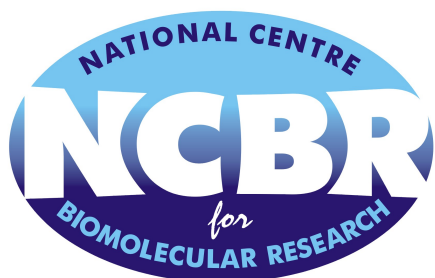


Výpočty volných energií (metody a aplikace)



Martin Petřek
(petrek@chemi.muni.cz)



National Centre for Biomolecular Research, Faculty of Science
Masaryk university in Brno, Kotlářská 2, 61137 Brno



Výpočty volných energií (metody a aplikace)

Obsah:

1. Termodynamika systému

(zákony, entropie, volná energie)

2. Volná energie v Biochemii

3. Metody výpočtů ze simulací

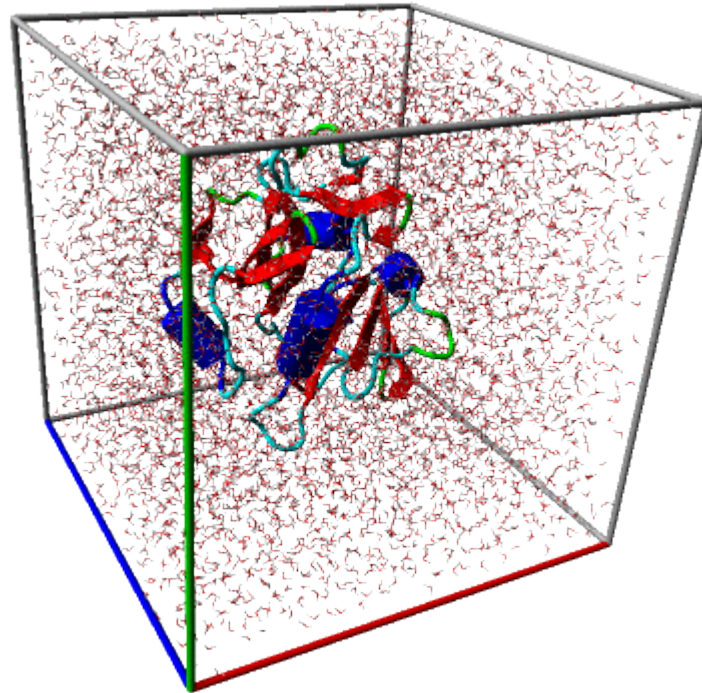
(Perturbační metoda, metoda Umbrella, metoda Bluemoon,
metoda ABF)

4. Ukázky a aplikace

5. Shrnutí

Termodynamika systému

Chemicko-biologický systém



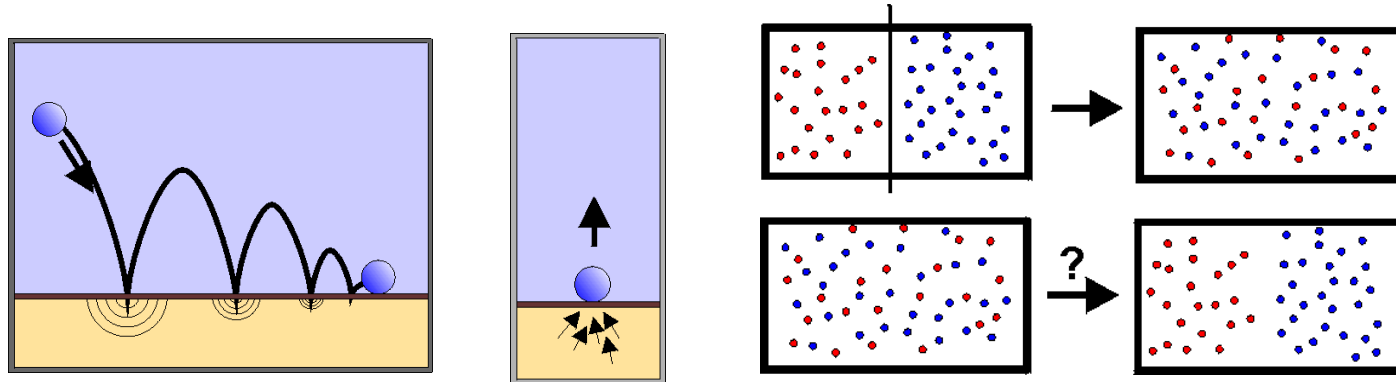
- vnitřní energie U
- teplota T , tlak p , objem V
- počet částic N
- interakce s okolím
 - přijímání/odevzdání tepla
 - konání/spotřeba práce

I. zákon termodynamiky $dU = dQ + dW$

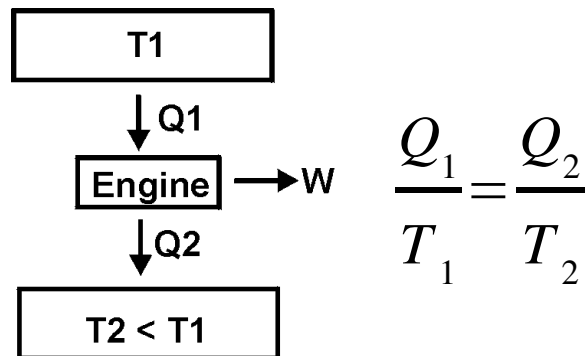
- v izolovaném systému je U konstantní
- U nic neříká o samovolných dějích (změnách) v systému

Entropie

- všechny jevy v systému splňující I. ZT jsou v principu možné
- samovolně nastávají ale jen určité jevy



- entropie zavedena při analýze vratných dějů s plynem (S. Carnot)



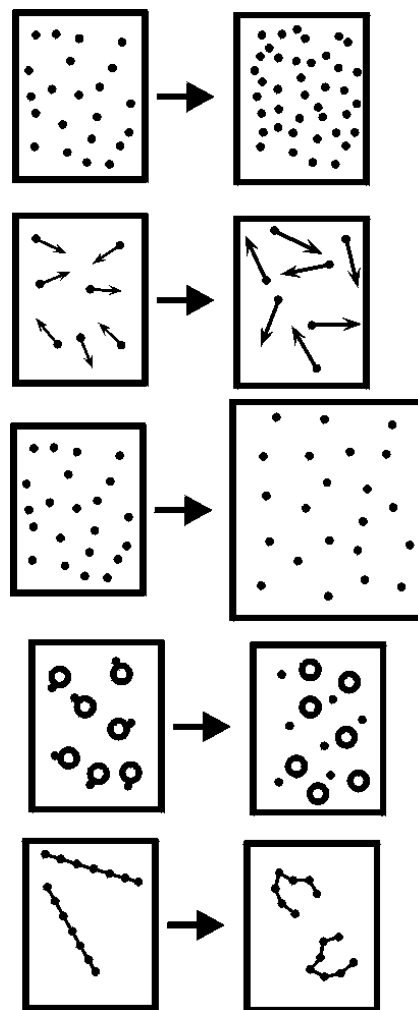
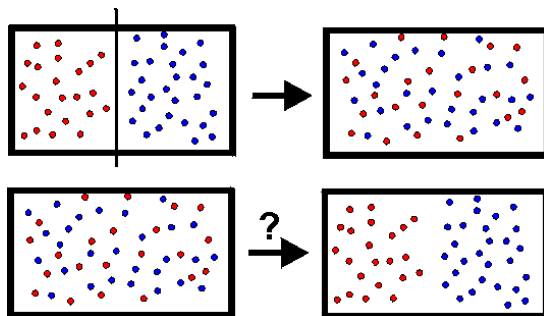
- Je-li teplo ΔQ vratně dodáno do systému při teplotě T , vzroste entropie systému o $\Delta S = \Delta Q/T$

- samovolné změny v izol. systému jsou spojené s růstem entropie S $dS \geq 0$ (II. zákon termodynamiky)

Růst entropie

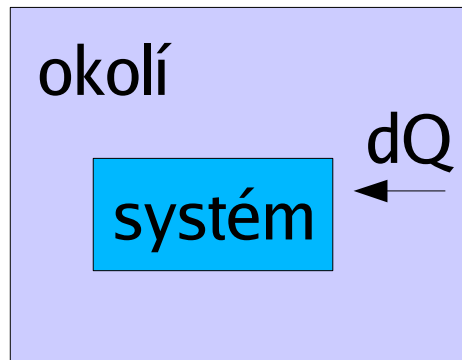
- *změny, kdy roste entropie:*

- *přidání molekul*
- *navýšení vnitřní energie*
- *zvětšení objemu*
- *mísení plynu*
- *...*



Entropie

- stavová funkce (změna S závisí jen na počátečním a koncovém stavu systému; nezávisí na tom „jak a kudy“ změna stavu probíhala)



Entropie nevratné změny

- systém a okolí (teplota T), ne v mech. rovnováze

$$dS + dS' \geq 0 \quad dS \geq -dS'$$

$$dS' = -dQ / T$$

$$dS \geq dQ / T \quad \text{Clausiova nerovnost}$$

- **Statistická fyzika: „entropie je míra neuspořádanosti systému“**
 - souvisí s počtem způsobů, jak realizovat vnější stav systému
 $S = k \log W$
 - v systému se realizují makrostavy, které jsou více pravděpodobné
 - uspořádaný systém – omezený počet způsobů
 - neuspořádaný systém – daleko více způsobů realizace
- => chování systému od uspořádaného k neuspořádanému

Volná energie

- pro popis dějů v systému je vhodnější použít stavové funkce A , G
Helmholtzova, Gibsova volná energie

- kombinace I. a II. zákona termodynamiky

$$dW=0, dU=dQ \rightarrow dS \geq dU/T \rightarrow 0 \geq dU - T dS$$

$$\text{Helmholtz: } A = U - TS \rightarrow 0 \geq dA \quad (V, T \text{ konstantní})$$

$$\text{Gibbs: } G = U + PV - TS \rightarrow 0 \geq dG \quad (p, T \text{ konstantní})$$

Význam volné energie:

- *kriterium spontánní změny v systému (samovolné děje)*
- *značí tu část vnitřní energie, kterou lze teoreticky přeměnit na užitečnou práci (maximální práce ve smyslu Carnot. vratného děje)*

Volná energie v biochemii

- souvislost s rovnovážnou konstantou reakce

$$\Delta G = -RT \ln K$$

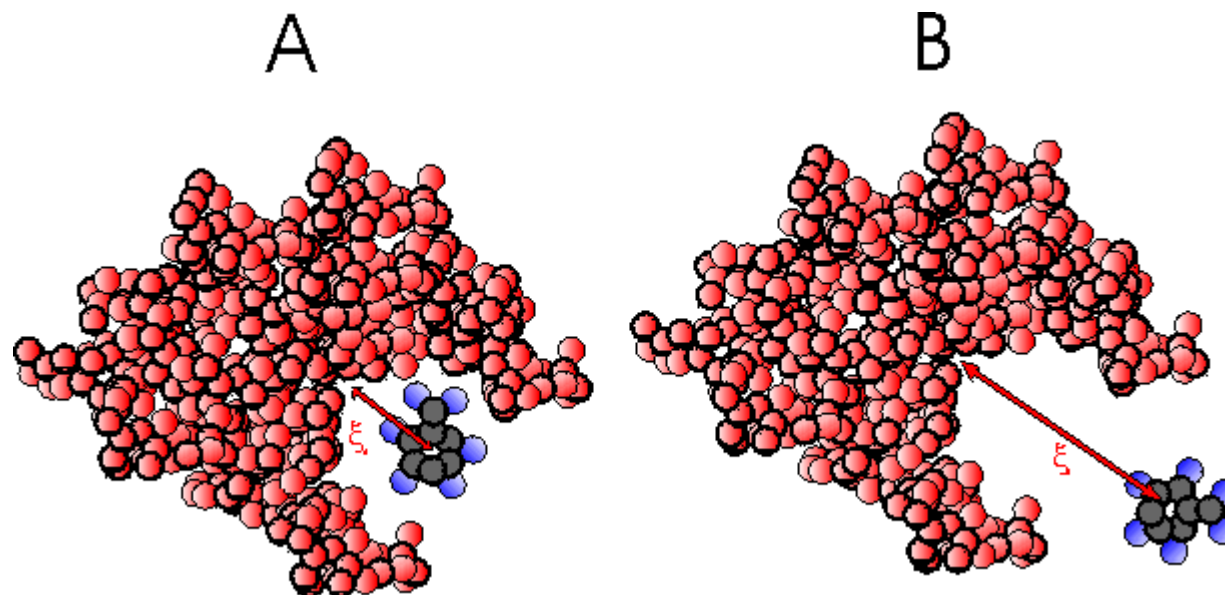
- souvislost s rychlostní konstantou a aktivační energií reakce

$$k = \frac{k_B T}{h} e^{-\frac{\Delta G_a}{RT}}$$

- znalost volné energie hraje roli
 - u návrhu nových reakčních mechanismů
 - při optimalizaci chemických procesů
 - návrhu nových katalyzátorů
 - ve snaze modifikovat funkce enzymů
- experimentální měření volné energie obtížné, někdy téměř nemožné (biologické systémy)

Volná energie při změně stavu

- dva stavy systému A, B, popíšeme pomocí proměnná ξ

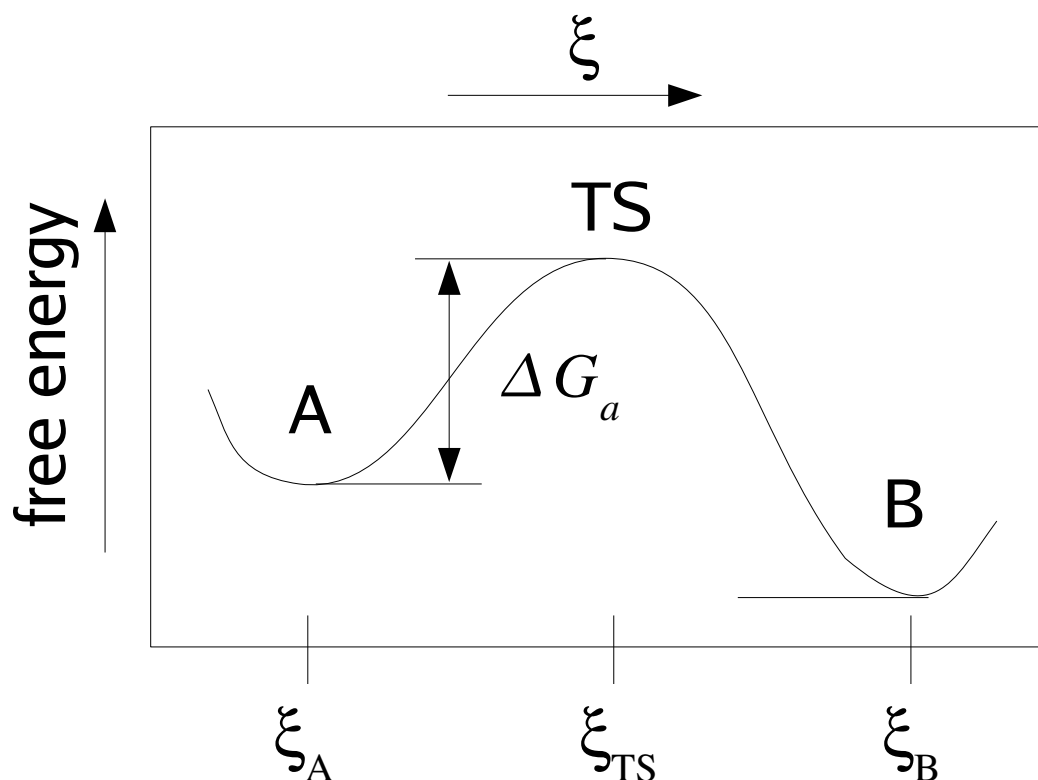


- např. vzdálenost těžišť 2 molekul, délka vazby, dihedralní úhel
- chceme odhadnout energetickou náročnost změny stavu z A do B (tu právě popisuje volná energie)
- pro konkrétní hodnotu ξ existuje spousta konfigurací systému

Volná energie při změně stavu

- *přechod systému ze stavu A do stavu B*
- *změna volné energie podél vybrané (obecné) souřadnice tzv. Reakční koordináta*

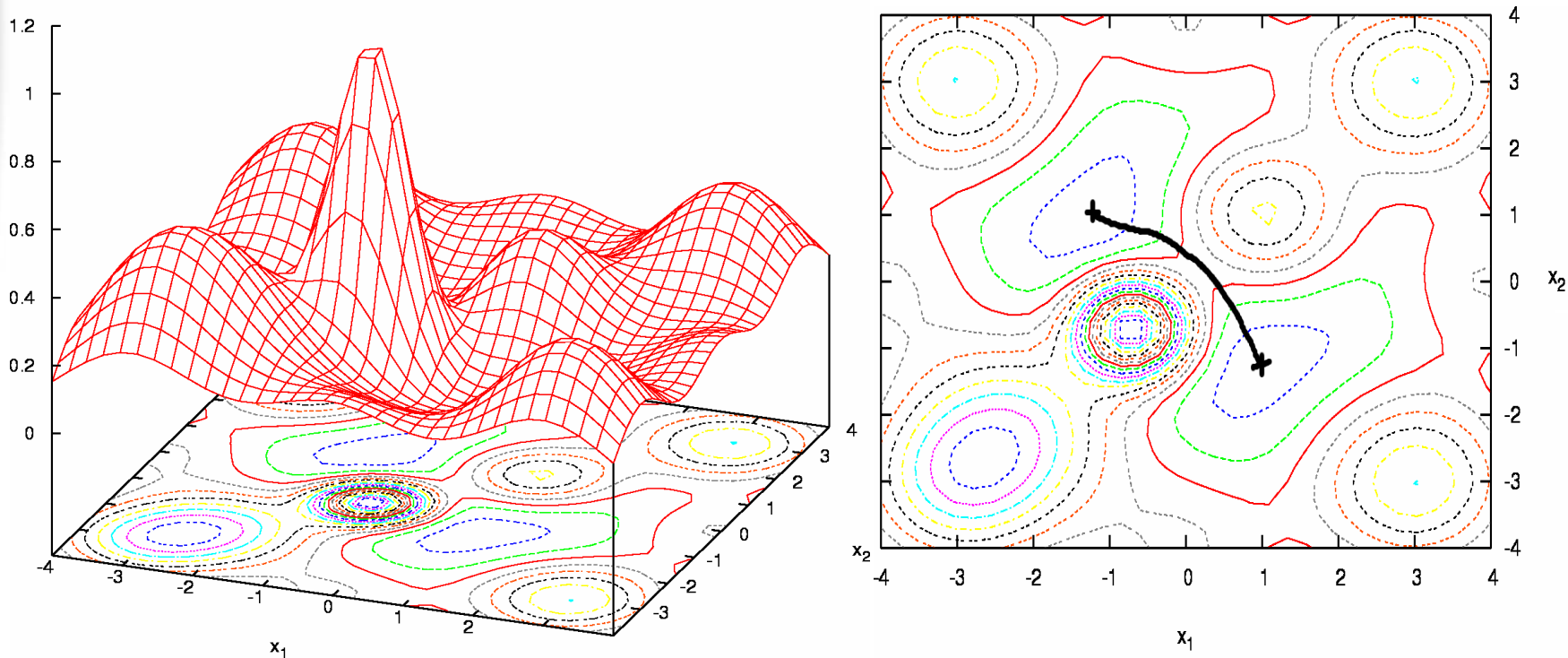
=> *graf profil volné energie v intervalu (ξ_A, ξ_B)*



- *odhady tranzitních stavů – odhad aktivační energie*

Volná energie při změně stavu

- *v praxi - reakční kordináta složitá (přiblížení, rotace), mnoha rozměrná reakční koordináta*
- *stavy je popsán např. (ξ, ϕ) => "hyperplocha" volné energie*
přechod z jednoho stavu do druhého \Leftrightarrow pohyb po této hyperploše z lokálního minima přes transitní stav (sedlový bod) do jiného lokálního minima



Volná energie v molekulové dynamice

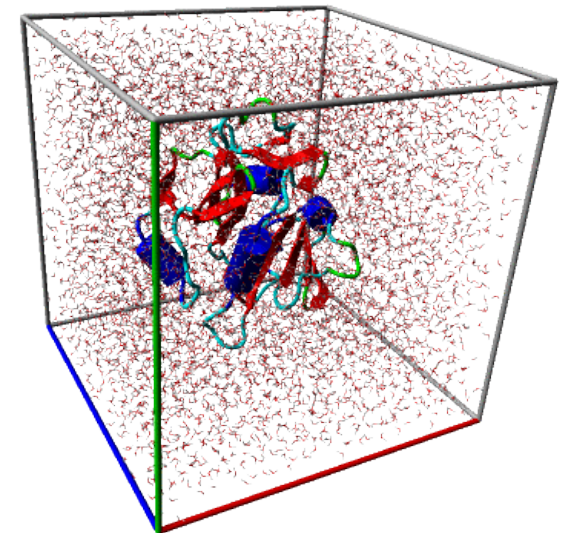
- systém obsahuje N částic (poloha, hybnost)

$$\mathbf{x} = (x_1, \dots, x_{3N}), \quad \mathbf{p} = (p_{x_1}, p_{x_2}, \dots, p_{x_{3N}})$$

- celková energie systému $H(\mathbf{x}, \mathbf{p}_x) = \frac{1}{2} \sum_i \frac{p_{x_i}^2}{m_i} + \Phi(\mathbf{x})$

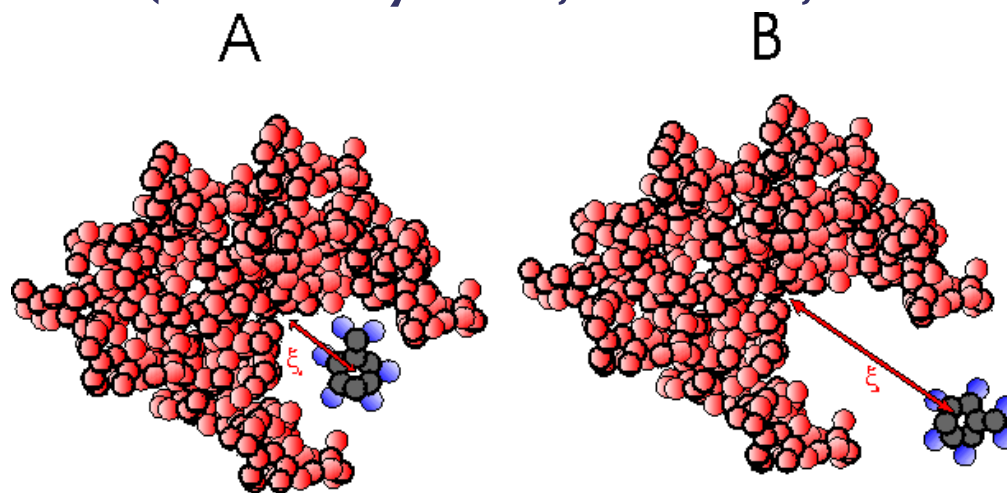
- Molekulová dynamika provádí numerické řešení pohybových rovnic v čase

► výpočet sil
úprava rychlostí
úprava souřadnic ■



Volná energie v molekulové dynamice

- reakční koordináta je definována některými souřadnicemi
- např. vzdálenost 1. a 10. částice $\xi = \sqrt{(x_1 - x_{30})^2 + (x_2 - x_{31})^2 + (x_3 - x_{32})^2}$
- hodnotu koordináty můžeme v průběhu simulace sledovat, resp. měnit (uměle zvyšovat, snižovat, udržovat, ...)



- *pro konkrétní hodnotu (např. $\xi = 1$) existuje spousta konfigurací systému*
=> volná energie pro konkrétní hodnotu ξ (např. $\xi = 1$) souvisí s pravděpodobností s jakou se stav ($\xi = 1$) může objevit

Volná energie v molekulové dynamice

$P(\xi = \xi^0)$... je pravděpodobnost toho, že systém je zrovna v konfiguraci, kdy hodnota koordináty je rovna ξ^0

$$G(\xi^0) = -k_B T \ln P(\xi = \xi^0) + G_0$$

(připomenutí: $G = H - T S$)

„Volná energie pro danou konkrétní hodnotu reakční koordináty ξ^0 je záporně vzatý logaritmus pravděpodobnosti P násobený teplotou a konstantou, kde P je pravděpodobnost toho, že se systém nachází v jedné z konfigurací, kdy koordináta je nastavena na ξ^0 ”

- některé metody pro výpočet G jsou založeny na odhadu této pravděpodobnosti
 - Monte-Carlo přístupy, Umbrella Sampling

Metody výpočtů volné energie

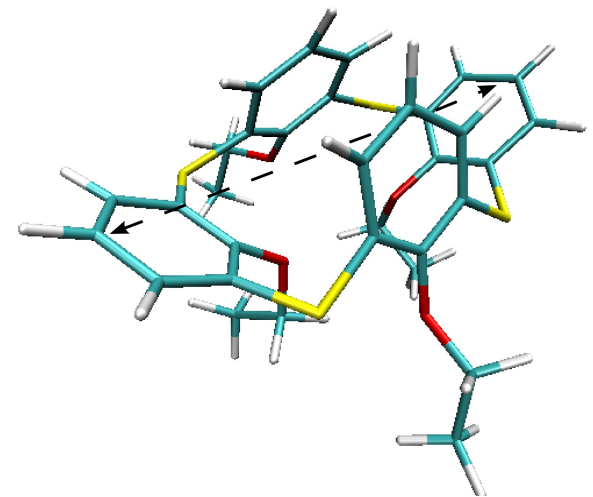
- „Jak odhadnout $P(\xi)$?”
- Klasický přístup – z dlouhého běhu MD simulace
 1. sledujeme ξ v průběhu simulace
 2. sestavíme histogram pro ξ v intervalu (ξ_1, ξ_2)
 3. odhadneme pravděpodobnost $P(\xi)$
 4. spočteme volnou energii $G(\xi)$

příklad:

pulsující molekula calixarenu

(výzkumná práce J. Matouška)

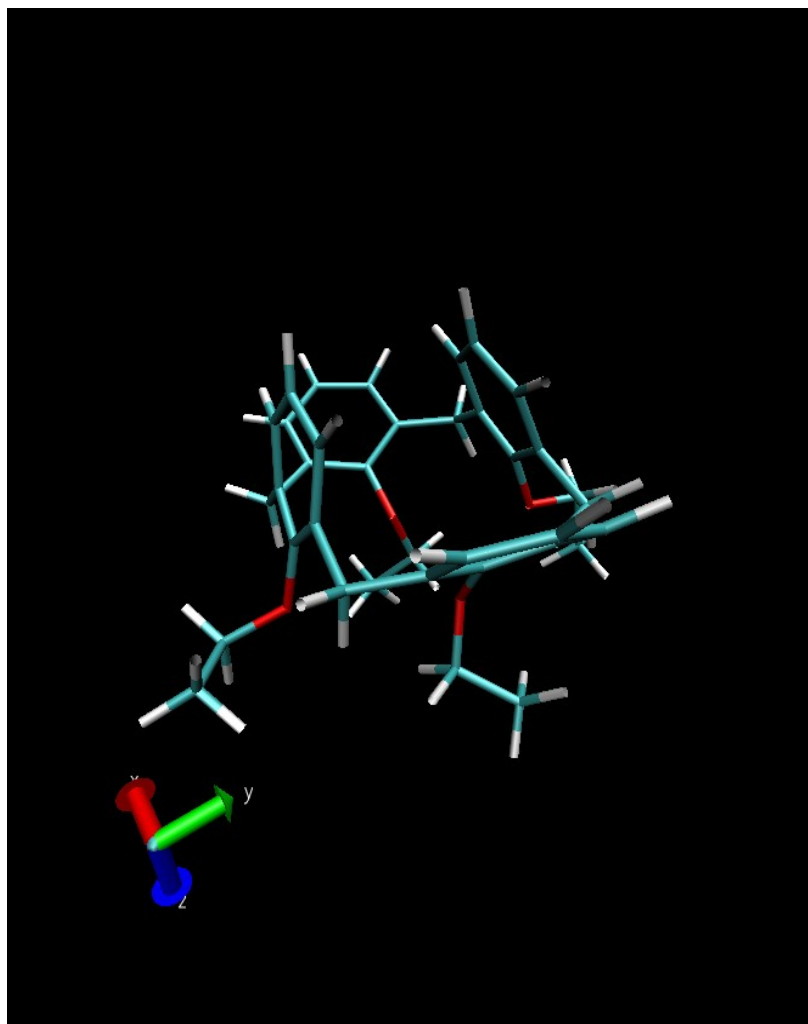
- 2 stavy, mezi kterými molekula volně přechází



Metody výpočtů volné energie

příklad: pulsující molekula tetraetoxycalix[4]arenu

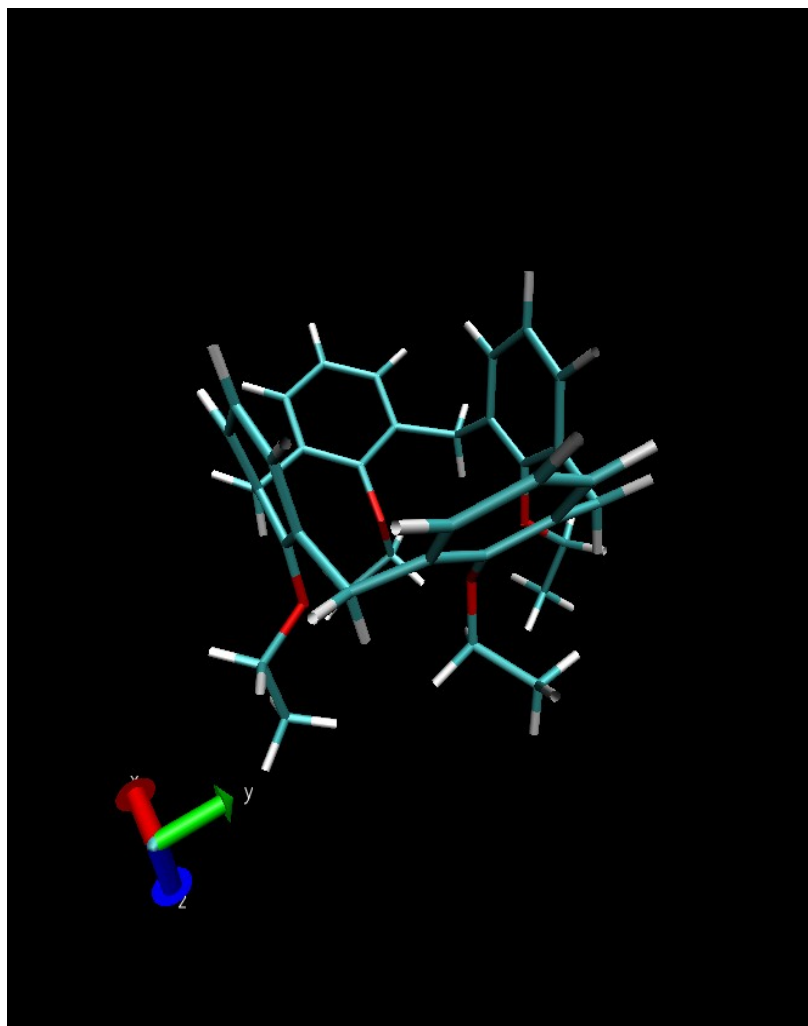
- 2 stavy, mezi kterými molekula volně přechází



Metody výpočtů volné energie

příklad: pulsující molekula tetraetoxycalix[4]arenu

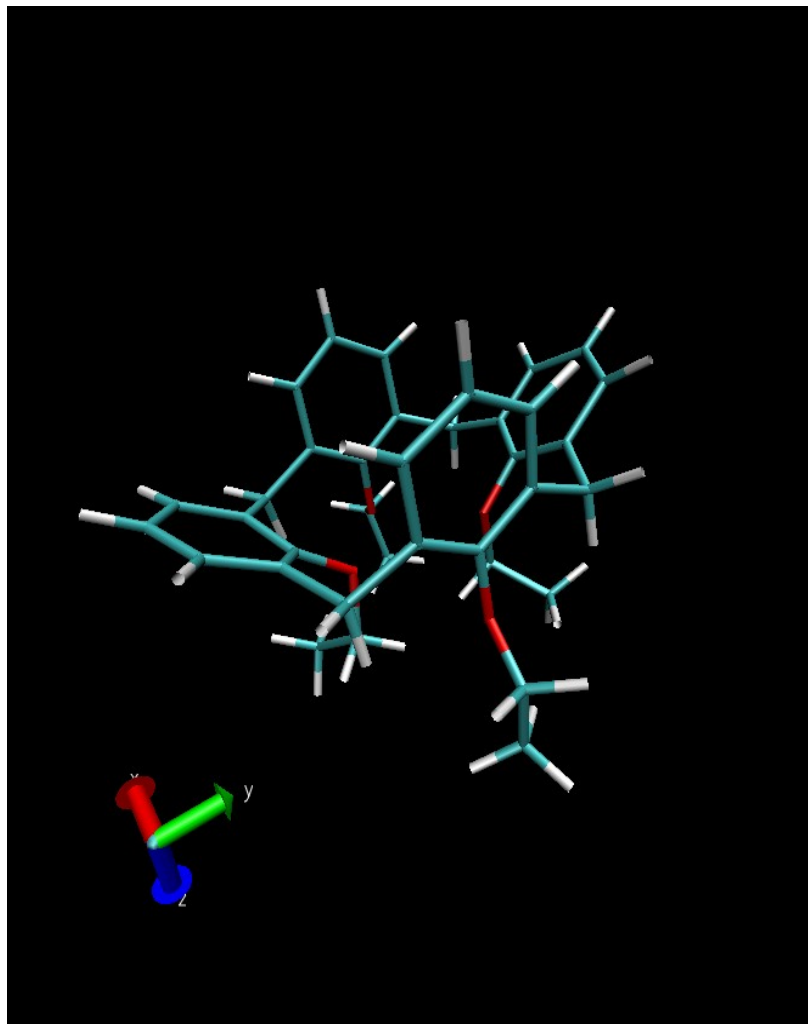
- 2 stavy, mezi kterými molekula volně přechází



Metody výpočtů volné energie

příklad: pulsující molekula tetraetoxycalix[4]arenu

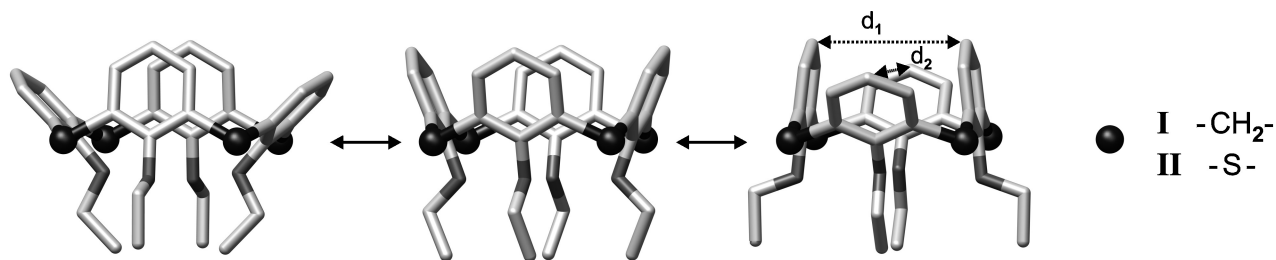
- 2 stavy, mezi kterými molekula volně přechází



Metody výpočtů volné energie

příklad: pulsující molekula calixarenu

- 2 stavy, mezi kterými molekula volně přechází
- jaká je energetická náročnost tohoto procesu, jaký je profil volné energie
- reakč. koordináta definovaná jako vzdálenost uhlíků C8 - C20 benzenových jader

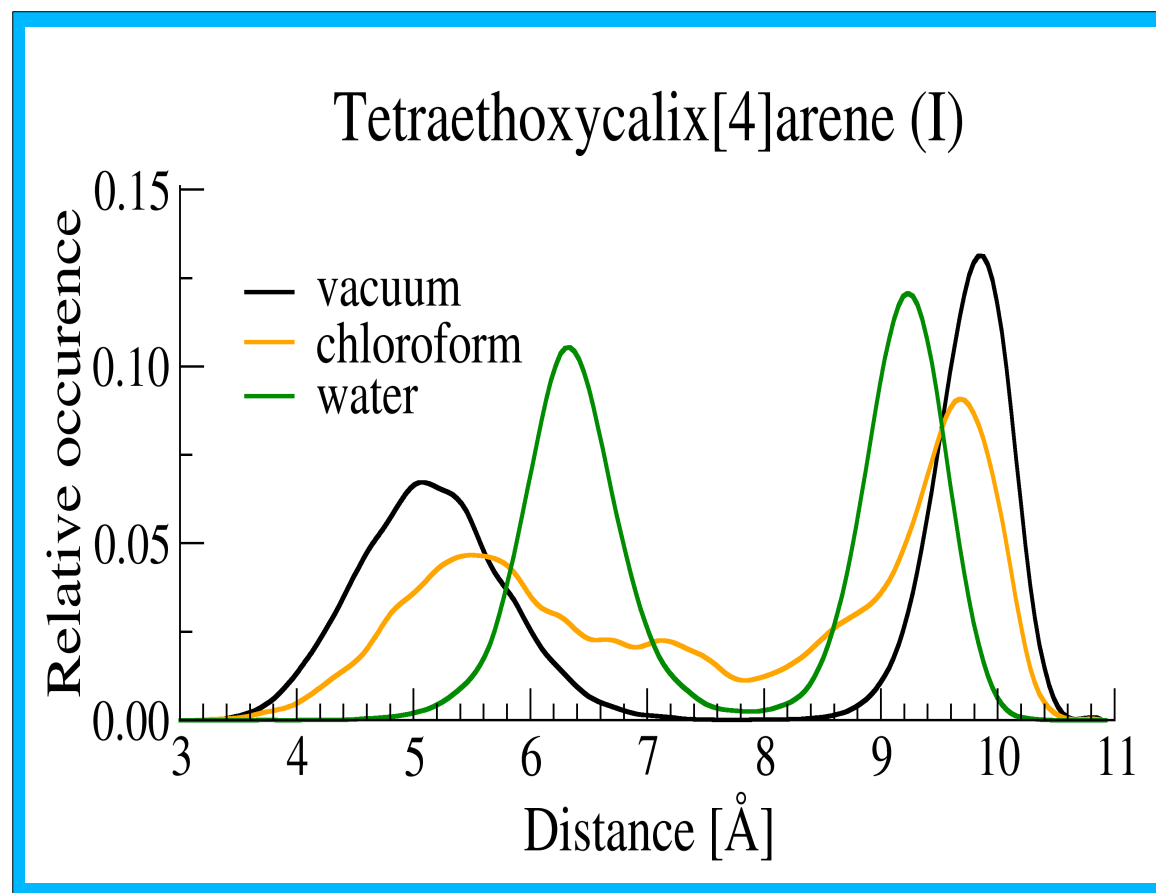


$$\xi = \sqrt{(x_1^{C8} - x_1^{C20})^2 + (x_2^{C8} - x_2^{C20})^2 + (x_3^{C8} - x_3^{C20})^2}$$

Metody výpočtů volné energie

příklad: pulsující molekula calixarenu

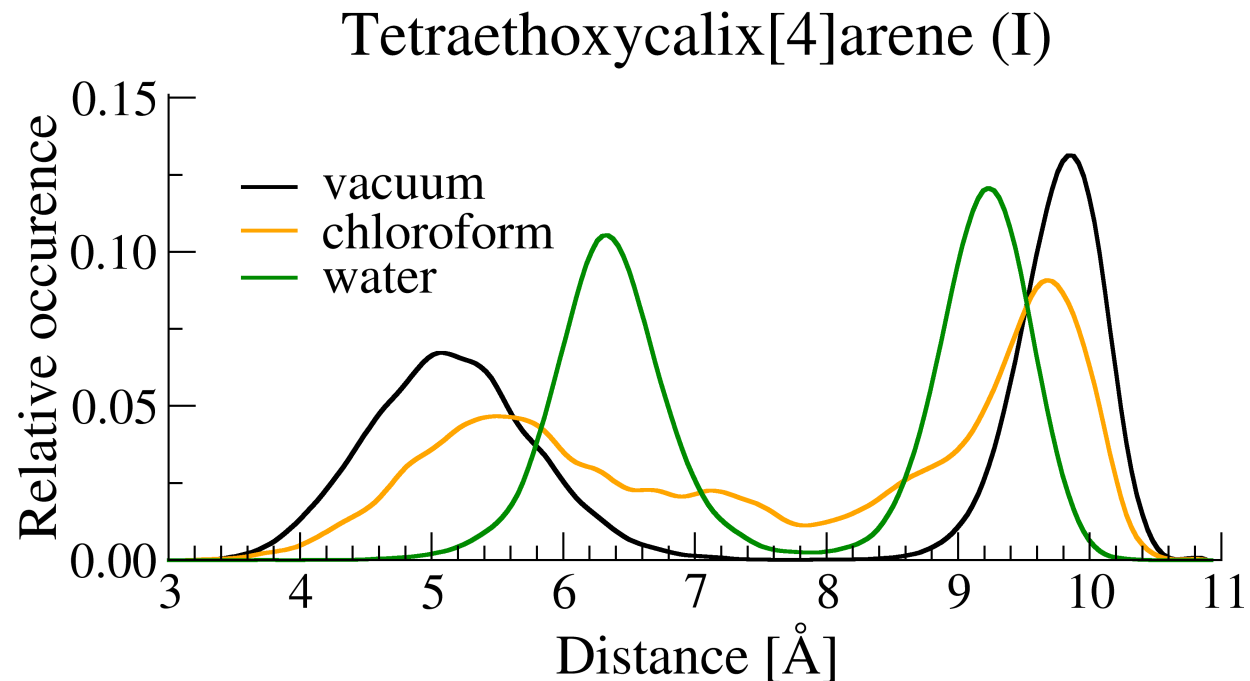
Krok.	Koordináta ξ [Å]
1.	4,87
10.	5,12
321.	5,42
1232.	4,77
4423.	4,73
5257.	5,46
5301.	5
5346.	4,59
5555.	4,96
5662.	5,11
5957.	4,85
6026.	5,06
6073.	7,5
6209.	5,11
6471.	8,65
6929.	9,49
7120.	9,24
7177.	8,68
7392.	9,24
7524.	8,98
7575.	8,59
7586.	9,13
7966.	8,68



Metody výpočtů volné energie – klasický přístup

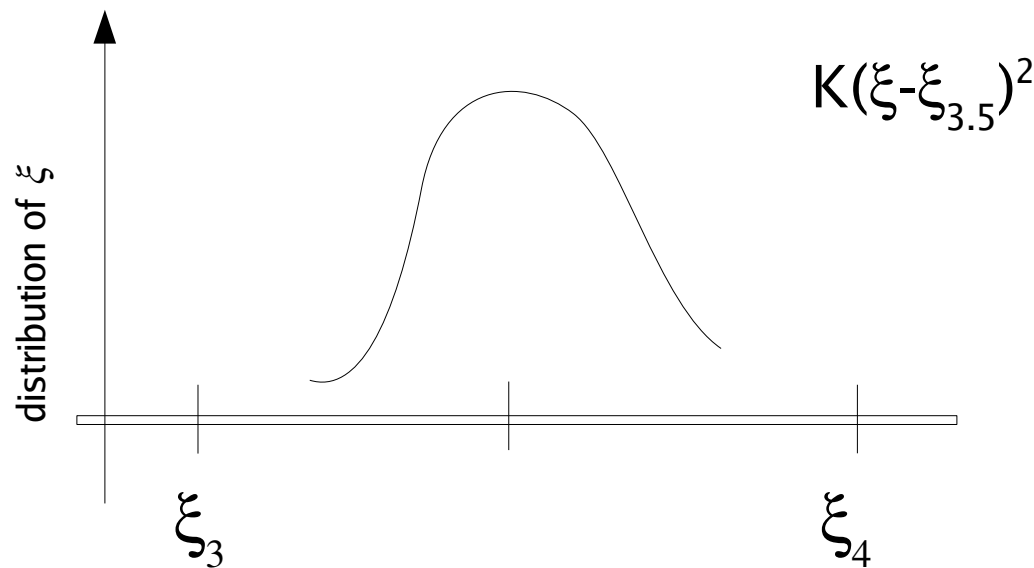
Problémy:

- špatné vzorkování ξ -prostoru
- většinu času je molekula poblíž minim ($\sim 5\text{Å}$., 9.5Å .)
- v ostatních bodech málo vzorků => nepřesný odhad P
- transitní stav ($\sim 7.5\text{Å}$.) navštíven jen zřídka



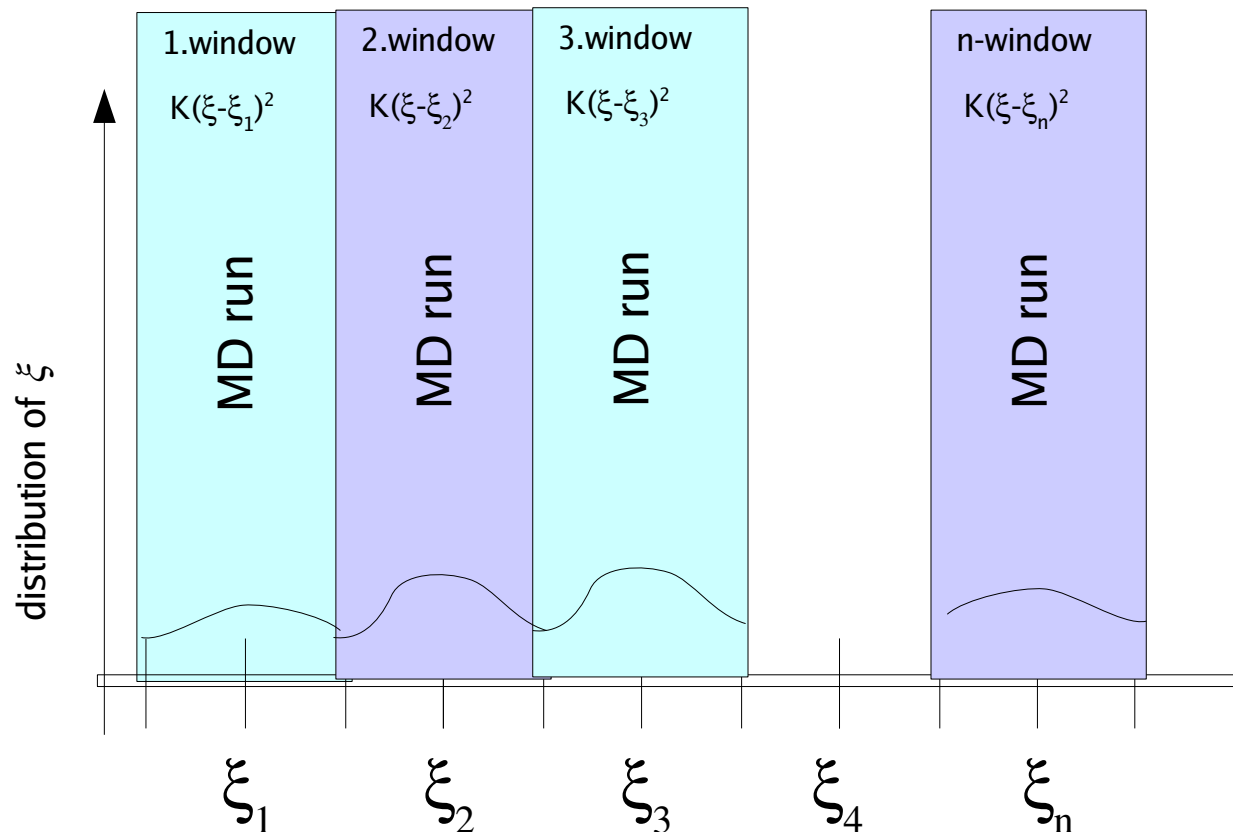
Metody výpočtů volné energie

- **Umbrella vzorkování [Torrie77]**
 - interval (ξ_A, ξ_B) se rozdělí na podintervaly (okna) a v každém se provede simulace a odhad pravděpodobnosti samostatně
 - je třeba zajistit, aby simulace běžela uvnitř okna
 - *do systému se dodá „umělá vazba“* $U_b(x, \xi_i) = K(\xi - \xi_i)^2$
 - *sledujeme distribuci ξ v každém okně*



Metody výpočtů volné energie

- Umbrella vzorkování [Torrie77]
 - „zvýšíme pravděpodobnost tam, kde to potřebujeme“
 - n-oddělených oken (pokrývají interval, překrývají se)



Metody výpočtů volné energie

Umbrella vzorkování [Torrie77]

- *dodaná „umělá vazba”* $U_b(x, \xi_i) = K(\xi - \xi_i)^2$
- *udržuje systém kolem koordináty ξ_i*
- *pravděpodobnost $P(\xi)$ je zvýšena touto dodatečnou vazbou*

☺ => *lepší provzorkování celého ξ -intervalu
(i tranzitních stavů)*



- *Přidaná vazba modifikuje volnou energii*
- *Odhady pravděpodobností musí být zkombinovány ze všech oken dohromady*
- *Nutné vhodně volit „tuhost vazby ... K”*

Umbrella vzorkování [Torrie77]

- kombinace histogramů pomocí metody

WHAM – weighted histogram analysis method

$$P_{\xi_i}^u(\xi) = P_{\xi_i}^b(\xi) e^{\frac{U_b(x, \xi_i) - g_i}{RT}}$$

$$P(\xi) = C \sum_i c_i(\xi) P_{\xi_i}^u(\xi)$$

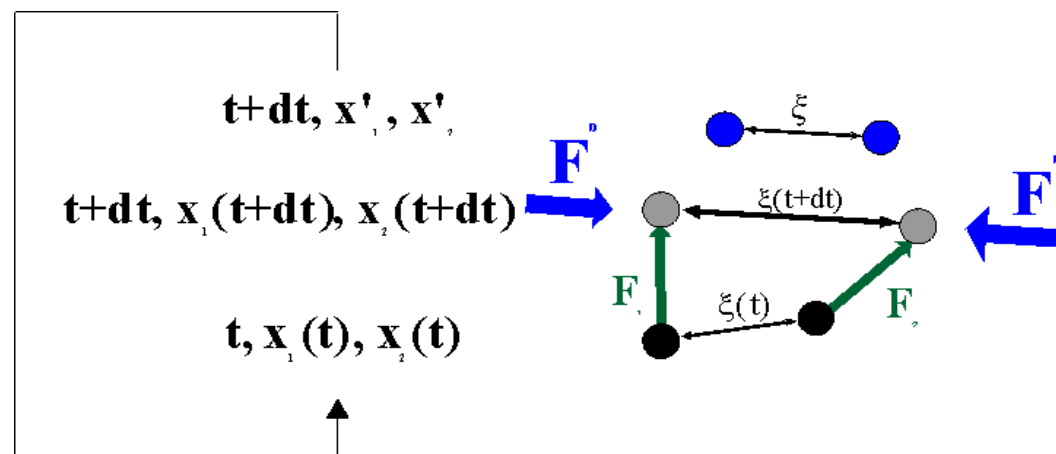
- získáváme odhad pravděpodobnosti
- $P(\xi) \Rightarrow$ přímý výpočet volné energie

$$G(\xi) = -k_B T \ln P(\xi) + G_0$$

Metody výpočtů volné energie

Bluemoon sampling [Otter98]

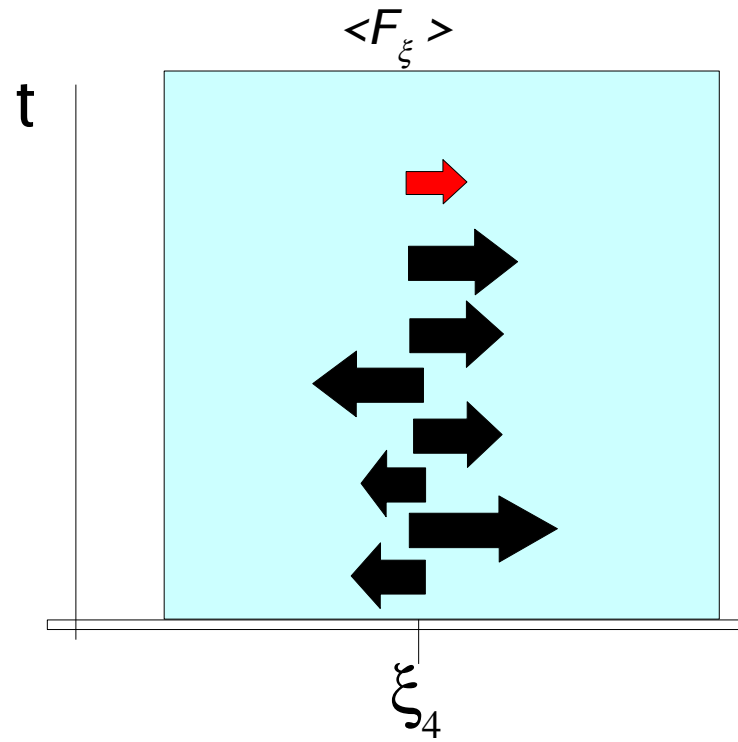
- odlišná metoda, neprovádí se odhad pravděpodobnosti
- zvolíme si body z intervalu (ξ_A, ξ_B)
- pro každý bod necháme běžet simulaci MD a zafixujeme reakční koordinátu
- => simulace „s omezením“ (constraint simulation)
- v průběhu výpočtu musíme uměle upravovat souřadnice, aby ξ zůstávalo konstantní („působíme silou“)



Metody výpočtů volné energie

Bluemoon sampling [Otter98]

- udržujeme hodnotu ξ konstantní
- v každém kroku simulace sledujeme sílu, kterou musíme působit, abychom předepsanou hodnotu udrželi



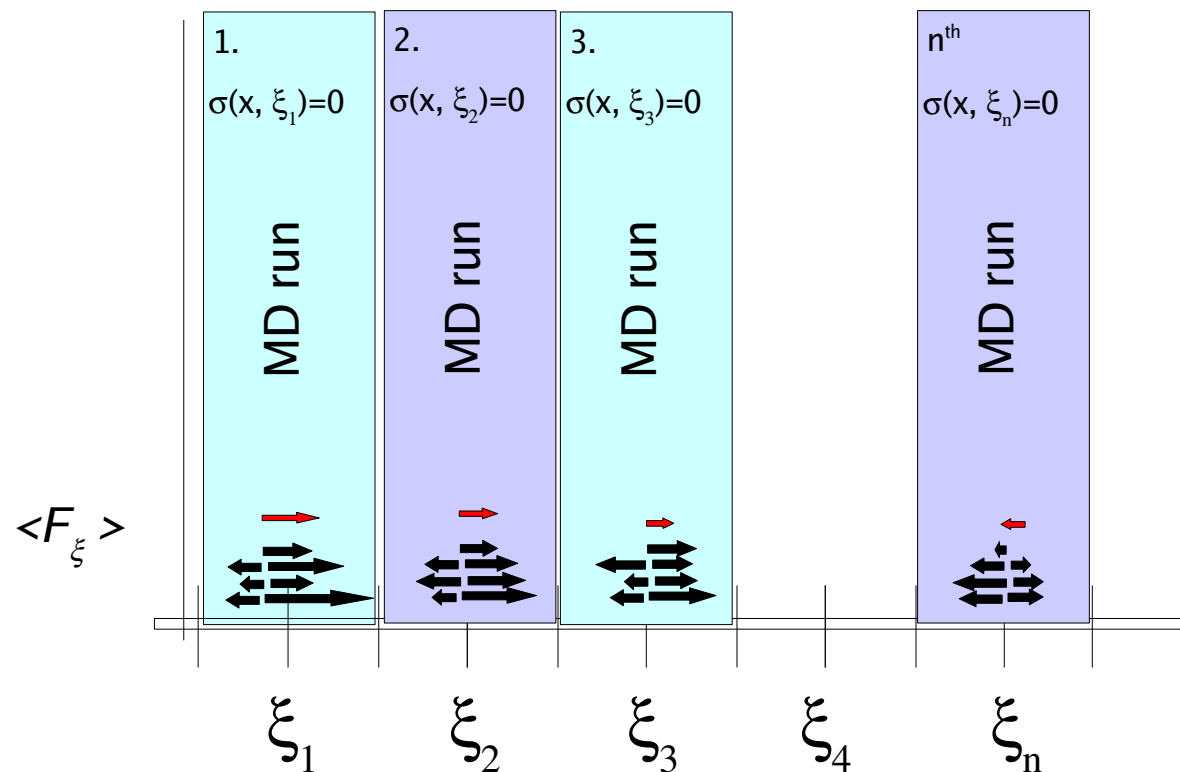
V každém kroku:
!! $\xi(x_1, x_2, \dots) = \xi_4$!!

- po dostatečně dlouhé době získáváme průměrnou sílu

Metody výpočtů volné energie

Bluemoon sampling [Otter98]

- n -oddělených MD běhů (v bodě je předepsán konstraint $\xi = \xi_i$)
- určí se *průměrné síly*, které držely koordinátu v jednotlivých bězích
- => na základě sil se spočte volná energie v celém intervalu

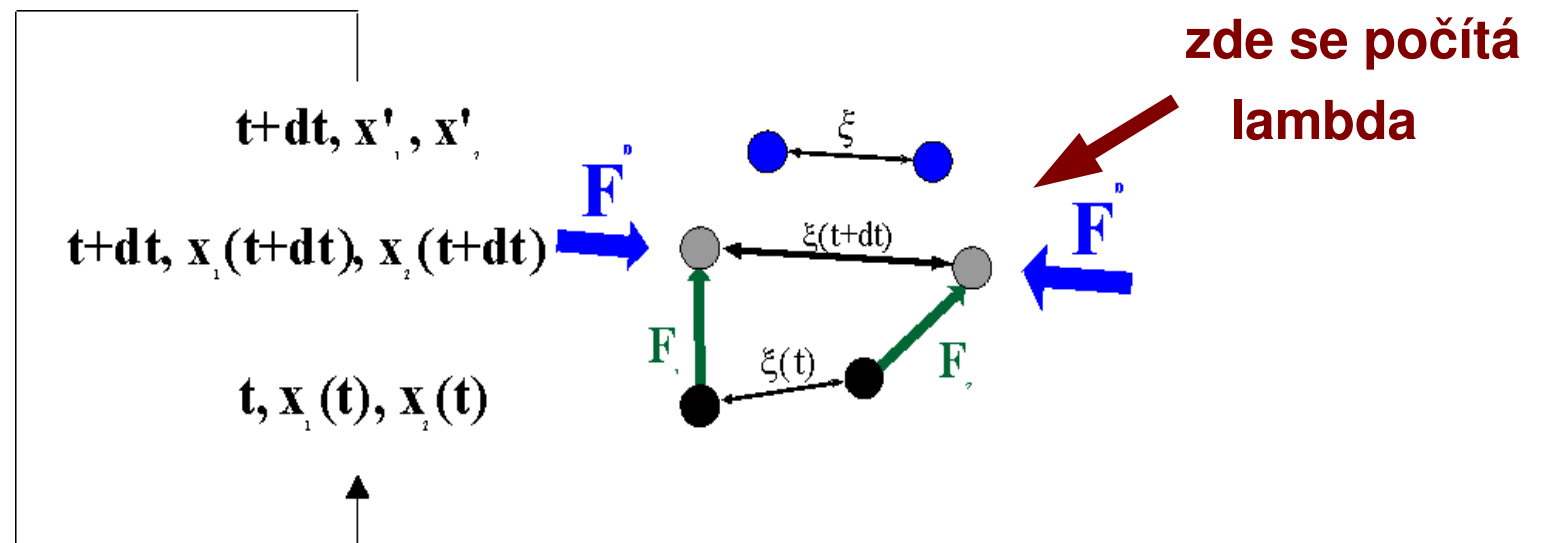


Bluemoon sampling [Otter98]

- Jak se určí potřebná síla k udržení konstantu?
- iteračně se určí tzv. *Lagrangeův koeficientem v pohybové rovnici*

$$m_i \ddot{x}_i(t) = -\frac{\partial}{\partial x_i} \left[\Phi(\mathbf{x}) - \sum_k \lambda_k(t) \sigma_k(\mathbf{x}) \right]$$

- *Lagr. koeficient* vyjadřuje, jak moc jsme museli opravit souřadnice, aby jsme splnili constraint



Bluemoon sampling [Otter98]

$$m_i \ddot{x}_i(t) = -\frac{\partial}{\partial x_i} [\Phi(\mathbf{x}) - \sum_k \lambda_k(t) \sigma_k(\mathbf{x})]$$

- **typy konstraintů: (vzdálenost, úhel, dihedrální úhel)**

Např. $\sigma_{dist}(\mathbf{x}, \xi) = r_{ij}^2 - \xi^2 = 0$

- **neznámé lambda se hledá iterativně v každém kroku MD,**

$$\lambda_n = \varphi(x, \lambda_{n-1}) \quad \text{aby platila podmínka konstraintu}$$

na základě lambda se spočte hledaná síla

(„síla k udržení konstraintu“)

technicky:
$$\frac{dG(\xi)}{d\xi} = \frac{\langle Z^{-1/2} [\lambda + RT Y] \rangle}{\langle Z^{-1/2} \rangle} = -\langle F_\xi \rangle$$

$$Z = \sum_i \frac{1}{m_i} \left(\frac{\partial \xi}{\partial \mathbf{x}_i} \right)^2 \quad Y = \frac{1}{Z^2} \sum_{i,j} \frac{1}{m_i m_j} \frac{\partial \xi}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{x}_i} \left(\frac{\partial \xi}{\partial \mathbf{x}_j} \cdot \frac{\partial \xi}{\partial \mathbf{x}_j} \right)$$

Bluemoon sampling [Otter98]

- v každém bodě tak máme „průměrnou sílu pro udržení ξ ”
(*potential mean force*)
- *souvislost síly s derivací volné energie podle ξ*

$$\frac{\partial G}{\partial \xi} = \left\langle \frac{\partial H}{\partial \xi} \right\rangle = \langle F_{\xi} \rangle$$

- => *numerickou integrací získáme profil*

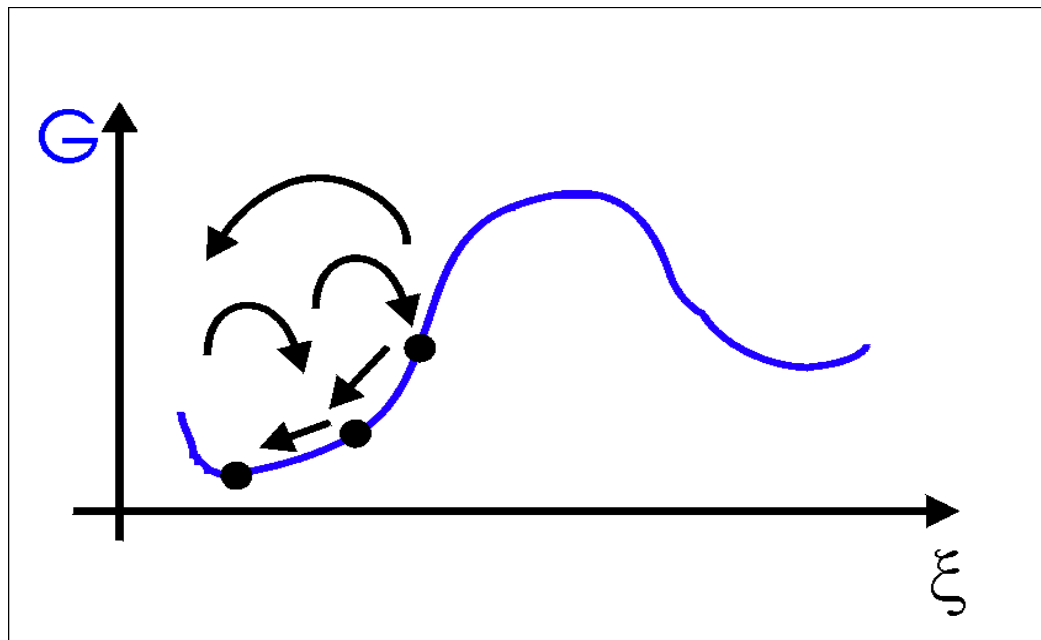
$$\Delta G = \int_{\xi_1}^{\xi_2} \langle F_{\xi} \rangle d\xi$$



- *potřeba mnoho bodů k pokrytí intervalu*
- *někdy pomalá konvergence při udržování konstraintu*
- *potřeba 2. derivací $\frac{\partial^2 \sigma}{\partial x_i \partial x_j}$ koordináty podle souřadnic*

Adaptive biasing force [Darve2001]

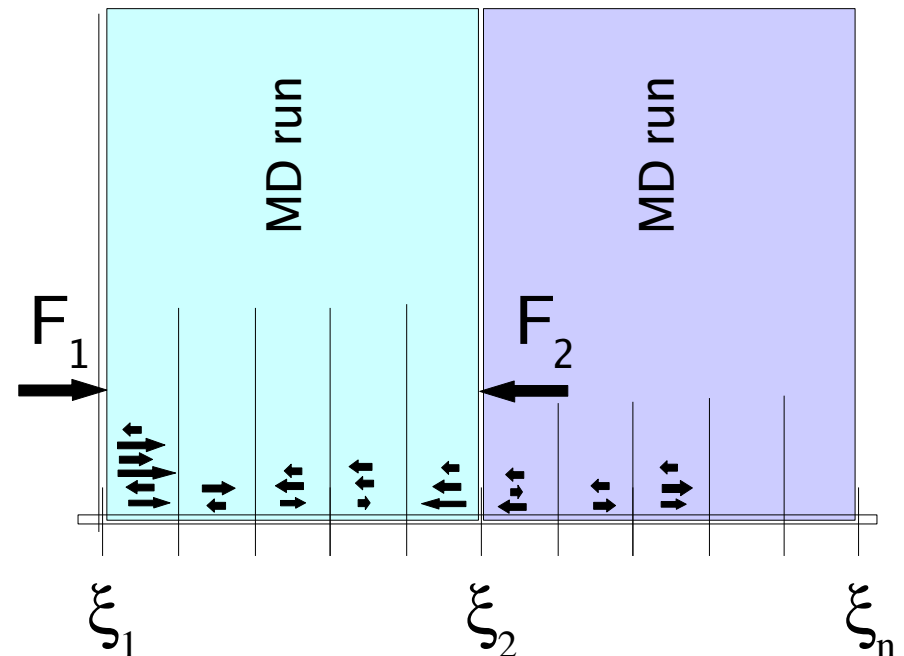
- „průměrná síla pro udržení ξ lze získat i z nekonstraintované simulace”
- několik málo oken pokrývajících interval
- v okně běží z počátku normální simulace, kde se sleduje síla, která by udržela koordinátu na hodnotě z předchozího časového kroku



Adaptive biasing force [Darve2001]

- (i) startuje se z minima => probíhá vzorkování v okolí minima (normální dynamika bez konstraintů)*
- (ii) průměrná síla se počítá v tzv. akumulátorech*
- (iii) po určité době (dostatečné vzorkování) se začne nastřádaná síla aplikovat na systém **v opačném směru***
=> systém začne vybíhat z minima do „neprobádaných oblastí“

- *tím se začnou lépe vzorkovat i další oblasti; síla se v nich střádá...*



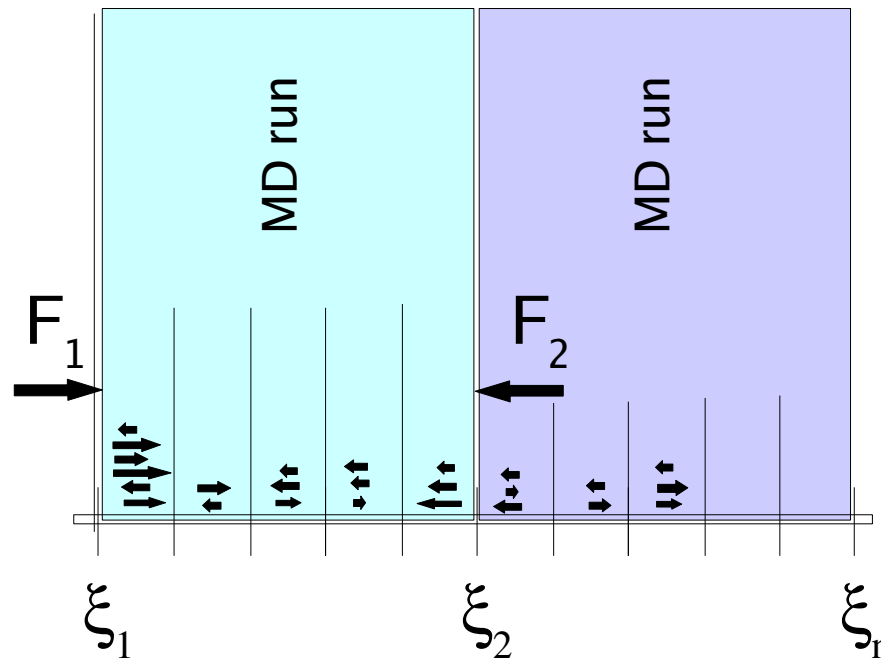
Adaptive biasing force [Darve2001]

Pohybová rovnice: $M \ddot{x}(t) = -\nabla \Phi(x) + F^{ABF}(\xi)$

- *hraniční „umělé síly“ (udržení simulace v okně)*
- *v průběhu simulace se u koordináty sleduje*

$$\xi(t), \dot{\xi}(t), \ddot{\xi}(t)$$

- *vlastní výpočet síly (F^{ABF}) je poněkud obtížnější (obecná reakční koordináta)*



Adaptive Biasing Force [Darve01]

$$\xi(t), \dot{\xi}(t), \ddot{\xi}(t) \quad Z\lambda := -\frac{d^2\xi}{dt^2}$$

- F^{ABF} výpočet

$$\frac{dG(\xi)}{d\xi} = \frac{\langle Z^{-1/2} [\lambda + X + RT Y] \rangle}{\langle Z^{-1/2} \rangle} = -\langle F_{\xi}^{ABF} \rangle \quad J_i = m_i^{-1/2} \frac{\partial \xi}{\partial x_i}$$

$$Z = \sum_i \frac{1}{m_i} \left(\frac{\partial \xi}{\partial \mathbf{x}_i} \right)^2 \quad Y = \frac{1}{Z^2} \sum_{i,j} \frac{1}{m_i m_j} \frac{\partial \xi}{\partial \mathbf{x}_i} \cdot \frac{\partial}{\partial \mathbf{x}_i} \left(\frac{\partial \xi}{\partial \mathbf{x}_j} \cdot \frac{\partial \xi}{\partial \mathbf{x}_j} \right)$$

$$[H]_{ij} = \frac{1}{m_i m_j} \frac{\partial^2 \xi}{\partial x_i \partial x_j} \quad X = Z^{-1} \dot{\xi} Z^{-1} J^T H J Z^{-1} \dot{\xi}$$

Adaptive Biasing Force [Darve01]

- na konci simulace je již celý prostor dobře provzorkován
- průměrné síly F^{ABF} jsou už dostatečně přesné

- => numerická integrace

$$\Delta G = \int_{\xi_1}^{\xi_2} \langle F_{\xi} \rangle d\xi$$



- *není příliš vhodné pro paralelizaci*
- *složitější implementace*

Metoda Termodynamické perturbace

- slouží k výpočtu rozdílu volné energie pro dva stavy
- volná energie ... stavová funkce
 - rozdíl ΔG mezi 2 stavy nezávisí na integrační cestě
=> lze volit i "nefyzikální cestu"

příklad: volná energie mezi molekulami

CH_3CH_2OH a CH_3CH_2SH ve vodním prostředí

- oba systémy jsou popsány svou energií (H_X , H_Y)
- platí vztah

$$\Delta A = k_B T \ln \langle \exp[-(H_X - H_Y)/k_B T] \rangle$$

- k výpočtu lze použít nefyzikální cestu záměny kyslíku a síry (resp. jejich Hamiltoniánu) => dá správný výsledek

Implementace

- **AMBER 8.0 – PMF knihovna (Kulhánek, Petřek)**
 - vícerozměrná reakční koordináta ξ
 - typy reakčních koordinát:
 - vzdálenost (2 atomů, 2 *těžišť*)
 - úhel (3 atomů, 3 *těžišť*)
 - diherální úhel (4 atomů, 4 *těžišť*)

Umbrella

- Tvorba histogramu v průběhu simulace
- WHAM program – načítání histogramů
- *Bluemoon*
 - automatická detekce kolidujících atomů v SHAKE algoritmu
- *ABF*

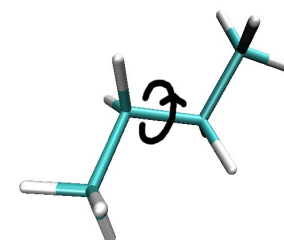
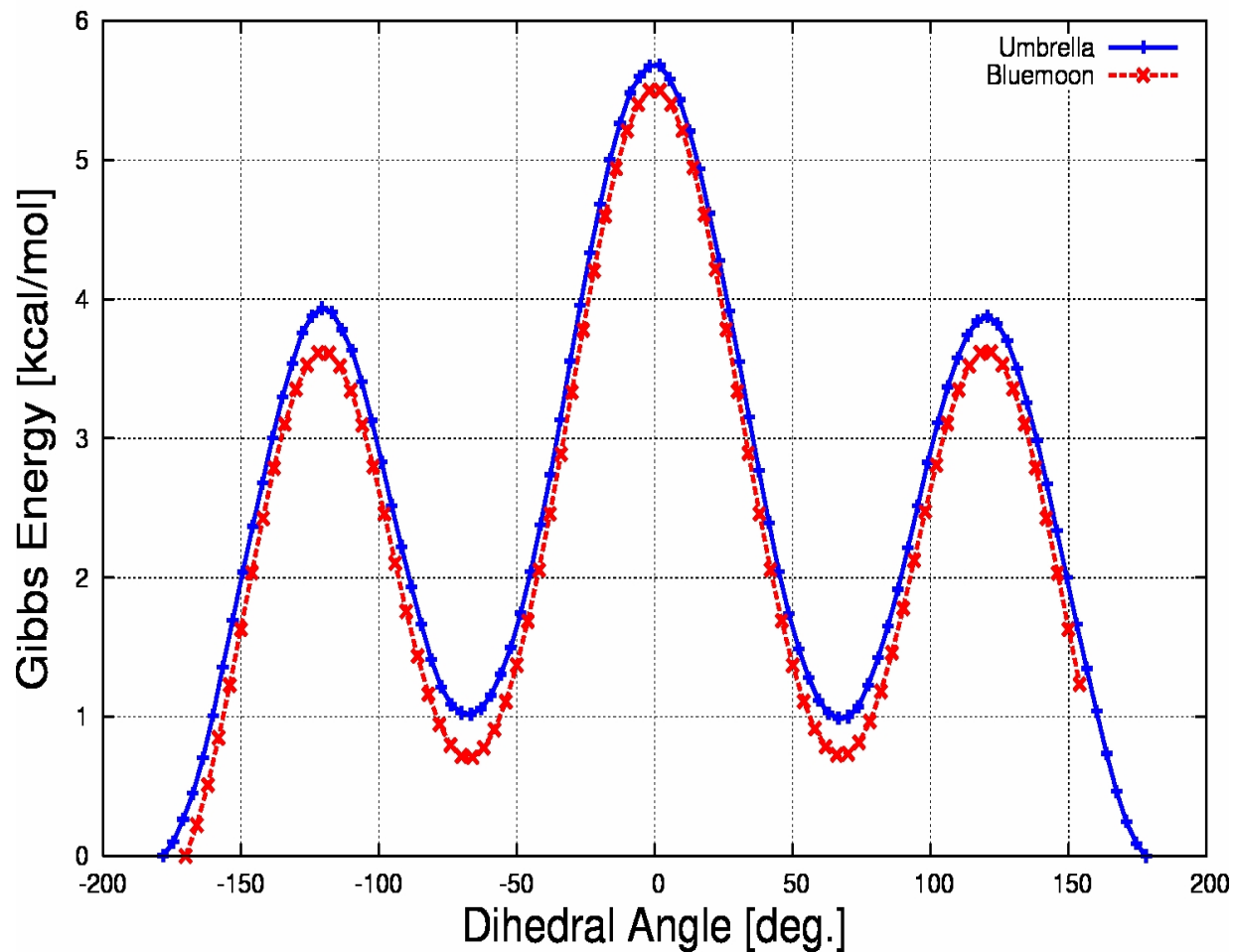
Implementace

- **AMBER 8.0 – PMF knihovna (Kulhánek, Petřek)**
 - Fortran90 (asi 25 000 řádek kódu)
 - sander, pmemd – háčky do zdrojových kódů
 - částečná paralelizace kódu v MPI
 - jednoduché použití v GRIDovém prostředí (METACentrum, VOCE)
- => module amber-pmf v systému CHARON**

Ukázka aplikací

- Test PMF knihovny
 - Butan (298K, 1Bar, dt = 1fs, 550ps, 82-86 oken)
 - ξ -koordináta – dihedrální úhel (C1-C2-C3-C4)

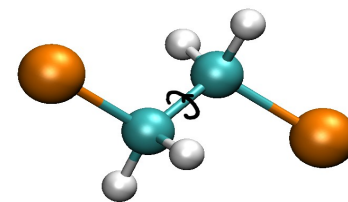
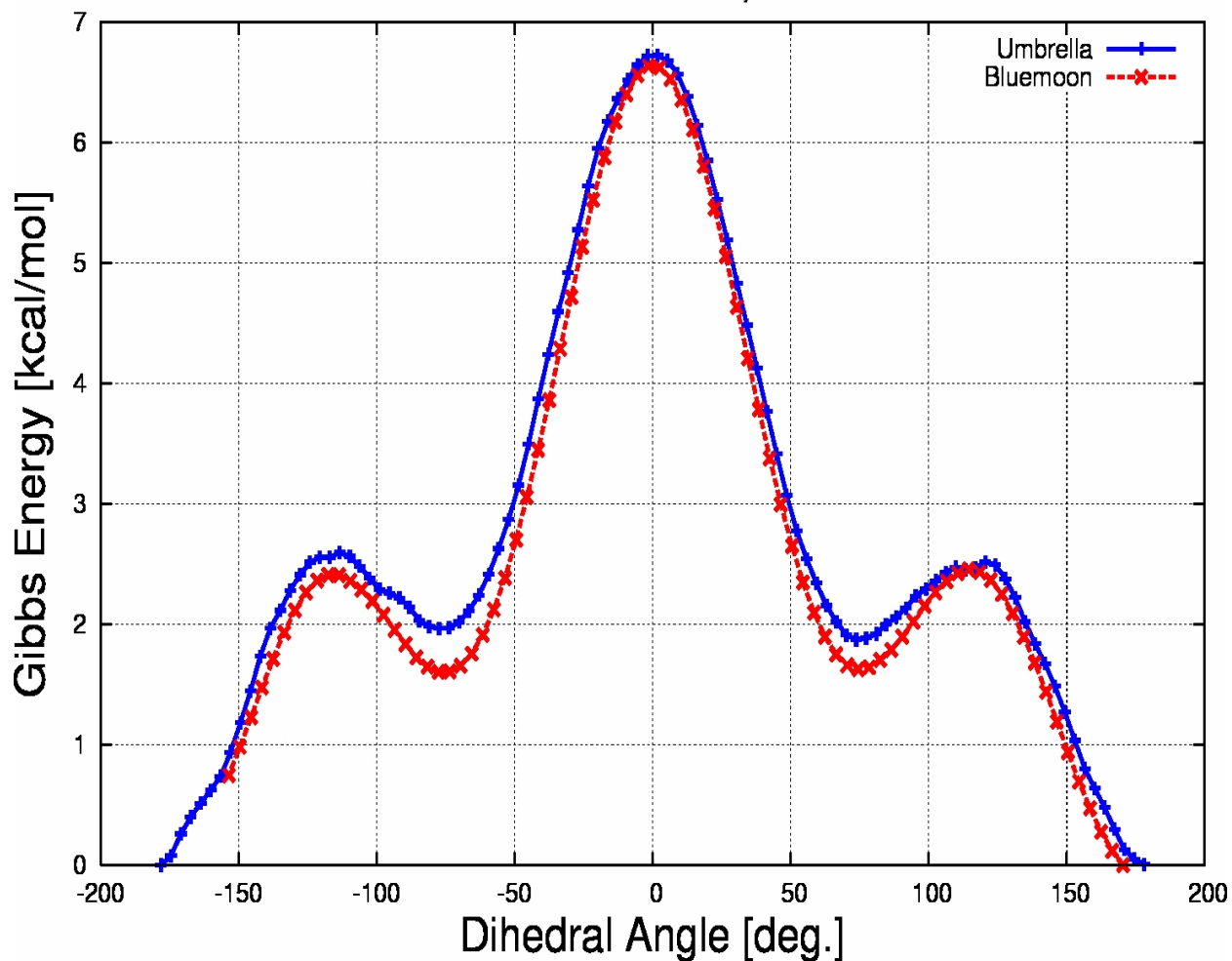
FEP - Dihedral C1-C2-C3-C4 Butan



Ukázka aplikací

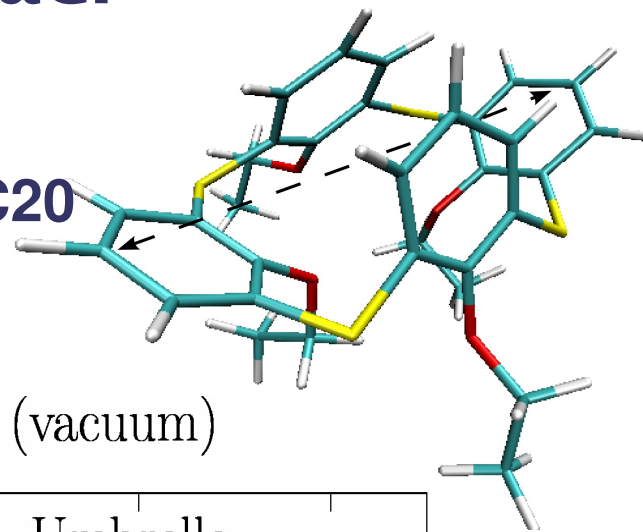
- 1,2-dichloroethane
- ξ -coordinate – dihedrální úhel (Cl-C-C-Cl)
(298K, 1Bar, dt = 1fs, 550ps, 82-85 windows)

FEP - Dihedral Cl-C-C-Cl 1,2-dichloroethane

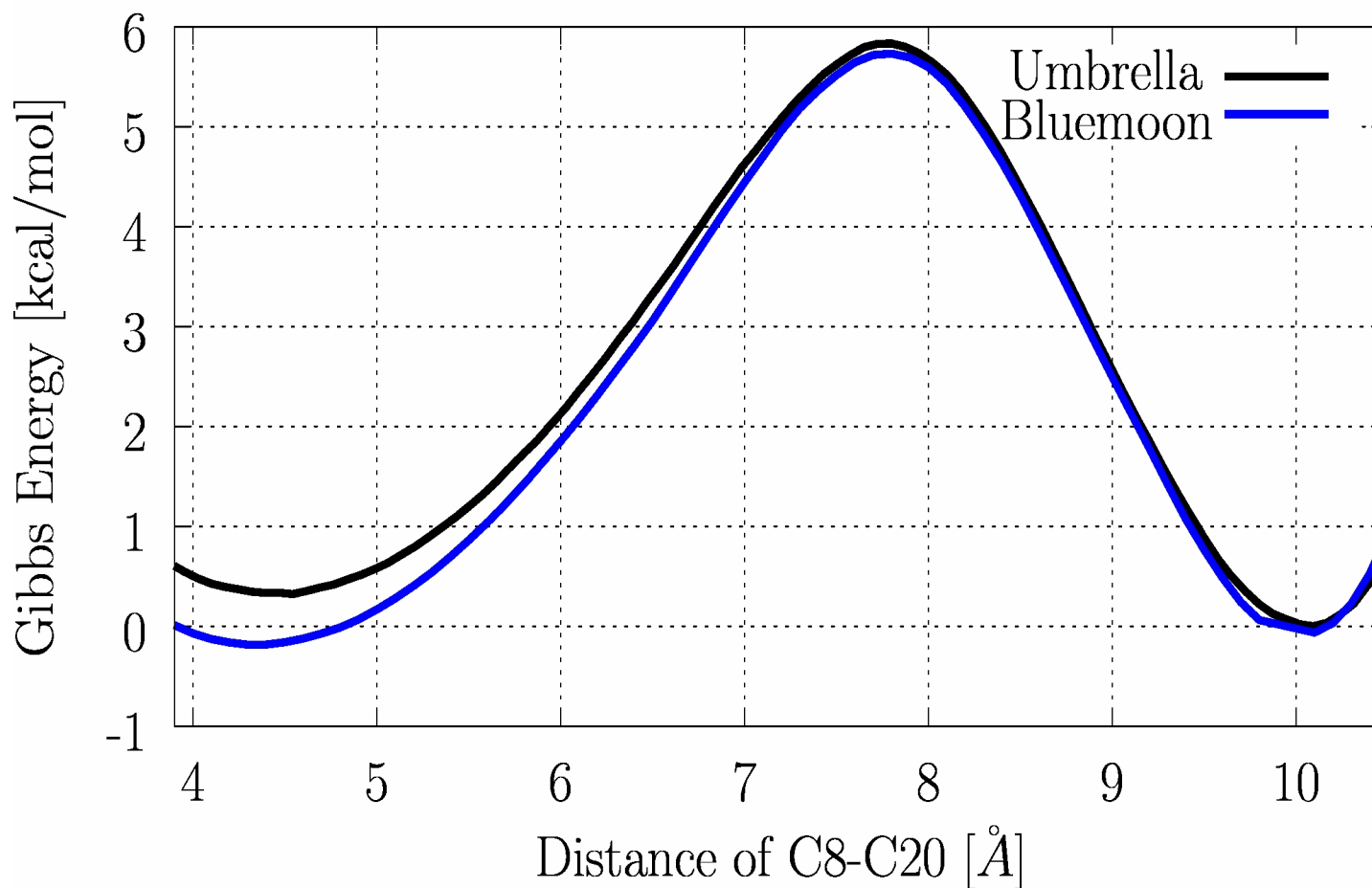


Ukázka aplikací

- tetraethoxythiacalix[4]arene
- ξ -koordináta - vzdálenost C8-C20
(298K, 1Bar, dt = 1fs, 200ps)

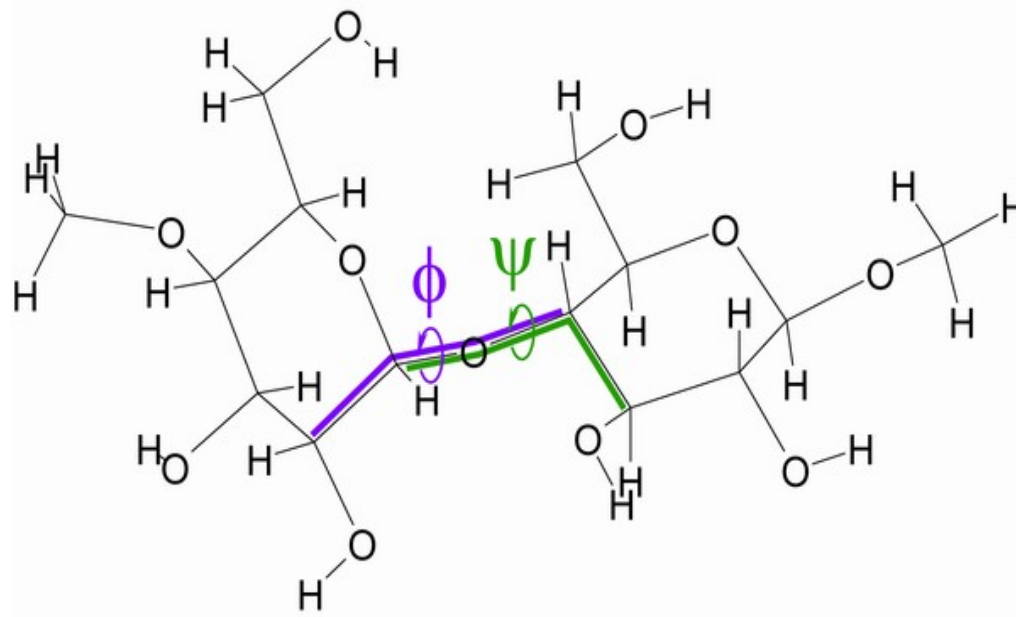


Tetraethoxythiacalix[4]arene (vacuum)



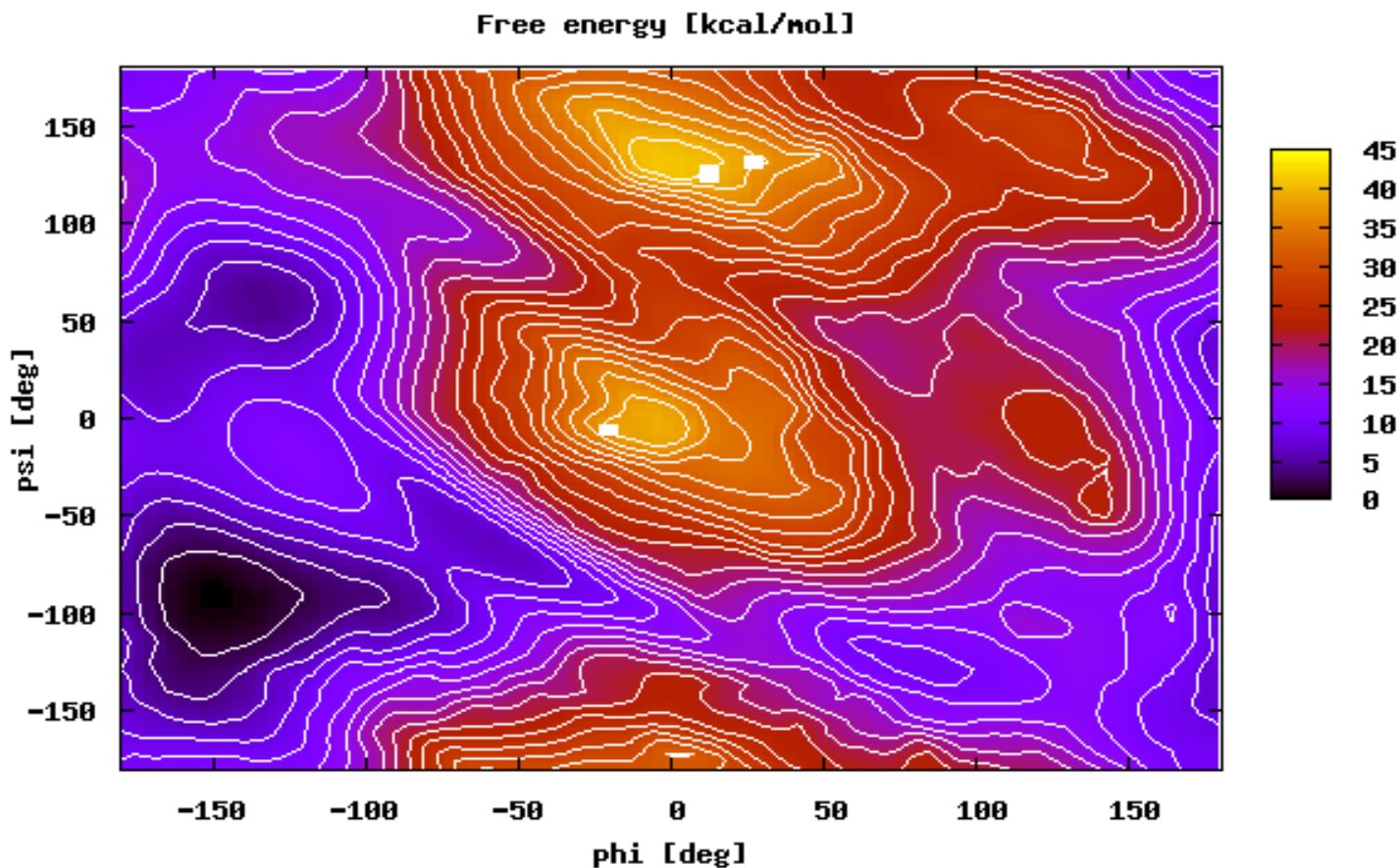
Ukázka aplikací

- cukry – hledání vhodných konformerů
- ξ -koordináta: 2 dihedrální úhly (ϕ, ψ)
- (298K, 1Bar, dt = 2fs, 200ps, SHAKE, 25 x 25 oken)
- umbrella sampling -> 2D-WHAM



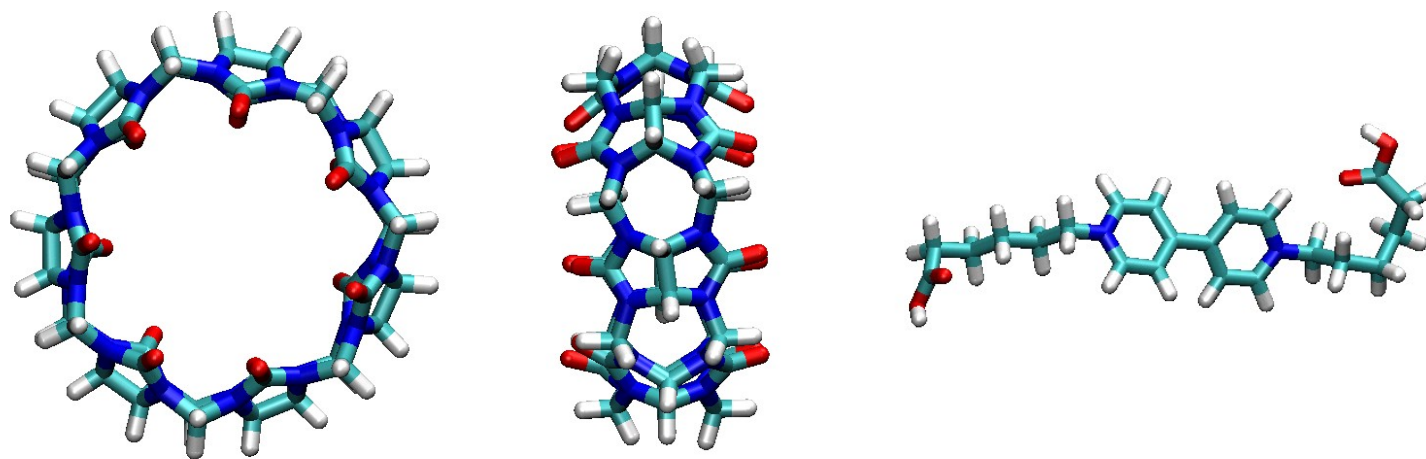
Ukázka aplikací

- cukry – hledání vhodných konformerů
- ξ -koordináta: 2 dihedrální úhly (ϕ, ψ)
- (298K, 1Bar, dt = 2fs, 200ps, SHAKE, 25 x 25 oken)
- umbrella sampling -> 2D-WHAM



Ukázka aplikací

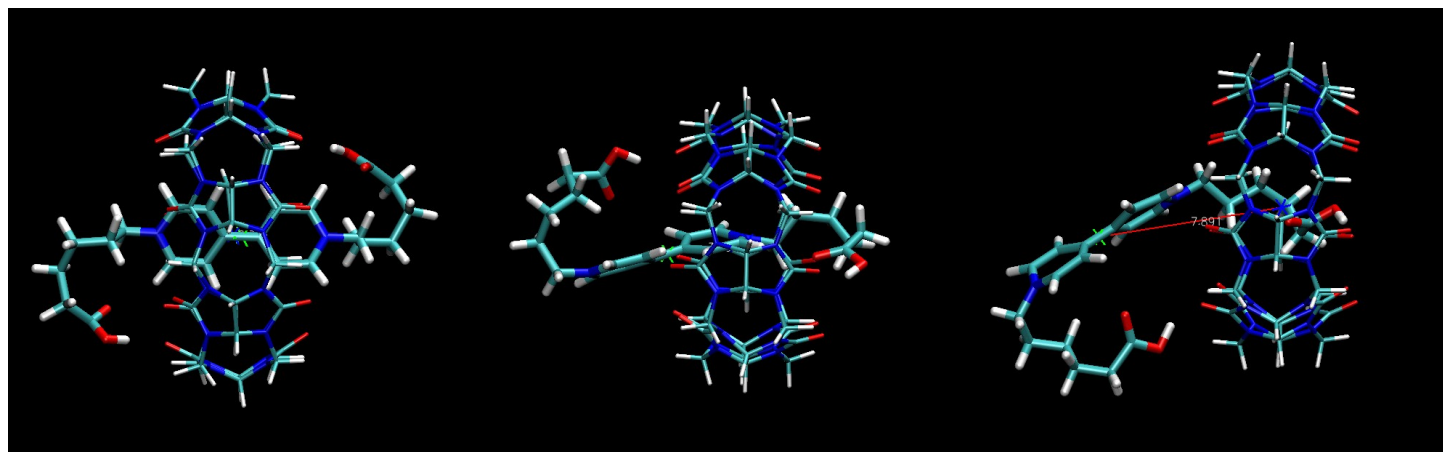
- rotaxany – energetický profil molekulárního switche
- ξ -koordináta: 2 vzdálenost těžišť
- (298K, 1Bar, dt = 2fs, 200ps, SHAKE)



- 2 stavy systému v závislosti na kyselosti prostředí
(tyčka navlečená na kolečku X tyčka vyvlečená ven)

Ukázka aplikací

- rotaxany – energetický profil molekulárního switche
- ξ -koordináta: 2 vzdálenost těžišť
- (298K, 1Bar, dt = 2fs, 200ps, SHAKE)



- 2 stavy systému v závislosti na kyselosti prostředí
(tyčka navlečená na kolečku X tyčka vyvlečená ven)



Literatura

- (1) den Otter, W. K.; Briels, W. J. J Chem Phys 1998, 109, 4139-4146.**
- (2) Torrie, G. M.; Valleau, J. P. J Comp Phys 1977, 23, 187.**
- (3) Darve, E.; Pohorille, A. J Chem Phys 2001, 115, 9169-9183.**
- (4) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Comput Phys Commun 1995, 91, 1-41.**
- (5) Matoušek, J; Kulhánek, P.; Čajan, M.; Koča, J. J Phys Chem A, 2006, 110, 861-867.**



Shrnutí

- **volná energie – zásadní role v chemických procesech**
- **v MD simulacích - 3 metody (umbrella, bluemoon, ABF)**
- **implementace pro MD balík AMBER 8.0**
- **typy koordinát: vzdálenost, úhel, dihedralní úhel a varianty s těžišti**
- **samostatná knihovna PMF, částečná podpora MPI**
- **důležité pro mapování hyperplochy potenciální energie, energetických bariér, vhodných konformerů ...**



Poděkování

prof. RNDr. Jaroslav Koča, DrSc.

RNDr. Petr Kulháněk

Mgr. Jiří Matoušek

Mgr. Lenka Šnajdrová

METACentrum, EGEE

GA Czech Republic (contract number: 204/03/H016)

MŠMT (contract number: MSM0021622413)