

CHEMIE A ALCHYMIE V POČÍTAČOVÉ SIMULACI CHOVÁNÍ MOLEKUL

KAREL HUML^a a MARTIN KRATOCHVÍL^b

^aÚstav organické chemie a biochemie, Akademie věd České republiky, Flemingovo nám. 2, Praha 6, e-mail: huml@imc.cas.cz, ^bÚstav fyzikální chemie J. Heyrovského, Dolejškova 3, 182 23, Praha 8, e-mail: krata@indy.jh-inst.cas.cz

Došlo dne 23.VI.2000

Klíčová slova: počítačová simulace, molekulární dynamika, Monte Carlo, volná energie

Obsah

1. Úvod
2. Výpočet změny Helmholtzovy energie F a Gibbsovy energie G
3. Rozklad změny Gibbsovy funkce G
4. Termodynamický cyklus
5. Závěr

1. Úvod

Výpočty Gibbsovy, resp. Helmholtzovy energie (ve shodě se starší literaturou označované jako volná energie), se často přirovnávají ke hledání svatého grálu počítačové chemie¹. Důvodem je, že většina fyzikálních vlastností a dějů závisí na volné energii systému. Ze statistické mechaniky pak vyplývá, že rozdíl volné energie udává relativní pravděpodobnost nalezení systému v daném stavu oproti stavu druhému². Ze znalosti volné energie lze tak odvodit např. směr a rozsah chemické reakce, vliv prostředí na reakce v roztocích a v enzymech, atd.³ Proto metody stanovení volné energie jsou jednou z nejrychleji se rozvíjejících oblastí počítačové chemie.

Teoretickým základem všech těchto technik je statistická mechanika, tvořící most mezi mikroskopickým světem počítačové simulace a měřitelnými makroskopickými veličinami. K popisu systému mnoha částic lze použít různé statistické soubory. Nejpoužívanějším je kanonický soubor charakterizovaný počtem částic N , objemem V a teplotou T , označovaný symbolem NVT nebo izobaricko-izotermický soubor NPT , kde P je tlak. V dalším výkladu se nejprve omezíme na první z nich. Dá se však ukázat, že závěry, které odvodíme, jsou platné i pro soubor NPT (cit.⁴).

V systému o konstantním počtu částic, objemu a teplotě platí pro Helmholtzovu energii F

$$F = -k_B T \ln Q \quad (1)$$

kde Q je tzv. partiční funkce souboru NVT , definovaná vztahem

$$Q = (h^{3N} N!)^{-1} \iint \exp[-H(\mathbf{p}, \mathbf{r})/k_B T] d\mathbf{p} d\mathbf{r} \quad (2)$$

a kde \mathbf{p} je vektor impulsů a \mathbf{r} je vektor souřadnic všech N uvažovaných částic. Veličina $H(\mathbf{p}, \mathbf{r})$ je hamiltonián systému, k_B je Boltzmanova a h je Planckova konstanta¹.

Z termodynamiky víme, že ze znalosti Q můžeme určit všechny termodynamické veličiny. Výpočet partiční funkce je však obtížný, neboť i pro velmi malý systém integrál ve (2) konverguje pomalu a tak je třeba extrémně dlouhých časů k tomu, abychom Q získali s dostatečnou přesností. V praxi nás však většinou nezajímá hodnota F , ale rozdíl Helmholtzovy energie ΔF dvou stavů reakce, počátečního A a koncového B, tj.

$$\Delta F = F_B - F_A \quad (3)$$

kde výpočet konverguje k rozumnému odhadu veličiny ΔF v reálném čase. Podrobné zdůvodnění nalezneme například v práci¹.

V následujících odstavcích si krátce všimneme některých postupů vhodných pro výpočet změny Helmholtzovy energie ΔF s využitím metody molekulární dynamiky (MD), případně metody Monte Carlo (MC). Zmíníme se rovněž o problému rozkladu Gibbsovy energie G na jednotlivé složky a uvedeme si několik aplikací, včetně použití tzv. termodynamického cyklu, při srovnávání paralelních procesů.

2. Výpočet změny Helmholtzovy energie F a Gibbsovy energie G

Pro výpočet hodnoty ΔF existuje řada metod, vyznačujících se různou časovou náročností a přesností. Jmenujme na tomto místě především dvě nejrozšířenější: Kirkwoodovu^{5,6} metodu termodynamické integrace a Zwanzigovu⁷ metodu perturbační, kterou do počítačové simulace chování molekul zavedli Tembe a McCammon⁸. Obě metody se dají odvodit ze společné „mateřské“ rovnice a za použití jistých limitních předpokladů přechází jedna v druhou. Ukazuje se, že každá je vhodná pro jiný typ úloh. Z dalších postupů jsou často citovány metoda pomalého růstu⁹ a potenciál střední síly¹⁰.

Pro ilustraci si uvedeme hlavní myšlenku první z nich, tj. metody termodynamické integrace⁴. Předpokládejme, že známe dva různé stavy systému: A, B, charakterizované příslušným hamiltoniánem H . Předpokládejme dále, že hamiltonián H systému je funkcí jistého reálného parametru λ (angl. coupling parameter¹¹), a že H lze psát jako součet dvou členů: kinetické energie K a energie potenciální V

$$H(\mathbf{p}, \mathbf{r}, \lambda) = K(\mathbf{p}, \lambda) + V(\mathbf{r}, \lambda) \quad (4)$$

Parametr λ zvolme tak, aby pro $\lambda_A = 0$ bylo $H(\mathbf{p}, \mathbf{r}, \lambda_A) = H_A(\mathbf{p}, \mathbf{r})$ a pro $\lambda_B = 1$ platilo $H(\mathbf{p}, \mathbf{r}, \lambda_B) = H_B(\mathbf{p}, \mathbf{r})$.

Při spojitě změně λ , se pak spojitě mění Hamiltonián a tedy

i partiční funkce Q . Helmholtzova energie ve vztahu (1), je pak také funkcí λ . Jejím derivováním obdržíme^{1,4}

$$\frac{\partial F(\lambda)}{\partial \lambda} = \frac{-k_B T}{Q(\lambda)} \frac{\partial Q(\lambda)}{\partial \lambda} = \left\langle \frac{\partial H(\mathbf{p}, \mathbf{r}, \lambda)}{\partial \lambda} \right\rangle_{\lambda} \quad (5)$$

kde $\langle \rangle_{\lambda}$ znamená středování přes soubor konformací příslušný dané hodnotě λ , kde příslušný soubor je generován metodou molekulární dynamiky nebo Monte Carlo¹².

Integrací tohoto výrazu dostáváme pro rozdíl Helmholtzovy energie mezi stavy A a B

$$\Delta F_{BA} = \int_{\lambda_A=0}^{\lambda_B=1} \left\langle \frac{\partial H(\mathbf{p}, \mathbf{r}, \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (6)$$

což je tzv. termodynamická integrační formule metody termodynamické integrace³.

Volba λ závisí na typu úlohy. Přehled různých situací, kdy λ má charakter topografické tranzitní souřadnice, uvádí např. Beveridge¹⁰. Nejjednodušší situací je lineární závislost (např. spojitá změna torzního úhlu nebo meziatomové vzdálenosti) ve tvaru

$$V(\mathbf{r}, \lambda) = (1 - \lambda) V_A(\mathbf{r}) + \lambda V_B(\mathbf{r}) = V_A(\mathbf{r}) + \lambda \Delta V(\mathbf{r}) \quad (7)$$

Uvedený postup je analogií tzv. „morfinu“, známého z televize, kdy tvář jedné osoby se postupně mění ve tvář osoby druhé. Zavedením vztahu (7) do (6), a za předpokladu, že kinetická energie je konstantní¹³, následnou integrací získáme výsledný výraz

$$\Delta F = \int \langle \Delta V(\mathbf{r}) \rangle_1 d\lambda \quad (8)$$

kdy integrace probíhá od $\lambda = 0$ do $\lambda = 1$. Numerická integrace výrazu (7) pak vede k přibližnému vztahu

$$\Delta F \sim \sum \langle \Delta V(\mathbf{r}) \rangle_i \Delta \lambda \quad (9)$$

kde sumace probíhá přes všechna i od 1 do celkového počtu n kroků $\Delta \lambda$ (cit.¹⁴⁻¹⁶). Podrobnosti týkající se přesnosti odhadu ΔF a konvergence výpočtu nalezneme např. v publikacích^{4,17-19}. O použitelnosti výrazů (8) a (9), jestliže kinetická energie není konstantní, se zmíníme v následujícím.

3. Rozklad změny Gibbsovy energie G

Doposud jmenované metody výpočtu ΔF jsou v principu rigorózní, ale u složitějších struktur výpočetně velmi náročné. Z tohoto důvodu byla zavedena celá řada aproximativních postupů, které v mnoha případech dávají výsledky v dobré shodě s experimentem.

V metodě molekulární mechaniky (angl. forcefield method) se vychází ze vztahu, kdy potenciální energie V je dána

součtem dílčích příspěvků (vazebné, van der Waalovy, elektrostatické atd.) ve tvaru

$$V = \sum V_k \quad (10)$$

V analogii s tímto výrazem se předpokládá platnost výrazu¹⁹

$$\Delta F = \sum \Delta F_k \quad (11)$$

Kritice tohoto postupu byla věnována celá řada prací²⁰⁻²². Ukázalo se však, že rozklad je oprávněný, ale jednotlivé komponenty ΔF_k , na rozdíl od celkové změny Helmholtzovy energie ΔF , nejsou stavovými veličinami a proto závisí na cestě, po které reakce (či výpočet) probíhá^{3,19,23}. Vyplyvá to ze skutečnosti, že na rozdílných cestách se jednotlivé fyzikální (či chemické) vlivy mohou značně lišit, přestože jejich výslednice je stejná. Toho lze s výhodou využít nejen při diskusi výsledků, ale zejména pro experimentální nebo teoretický odhad jednotlivých členů.

Stejně úvahy platí i pro Gibbsovu energii a její rozdíly. Pro ilustraci si uvedeme rozklad změny Gibbsovy energie (tzv. master equation) v případě výpočtu vazební energie ΔG_{bind} ligandu s receptorem jak ji uvádí Ajay a Murcko²⁴

$$\Delta G_{\text{bind}} = \Delta G_{\text{solvent}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{motion}} \quad (12)$$

kde první příspěvek odpovídá vlivu rozpouštědla, druhý změně konformace, třetí specifickým interakcím protein-ligand a čtvrtý rotacím, translačně-rotacím pohybům a vibracím.

Různé způsoby rozkladu a jejich vyhodnocení nalezneme například v publikacích^{25,26}.

4. Termodynamický cyklus

V řadě případů nás zajímá rozdíl dvou (nebo více) podobných procesů charakterizovaných Gibbsovou energií. To znamená, že hledáme hodnotu funkce

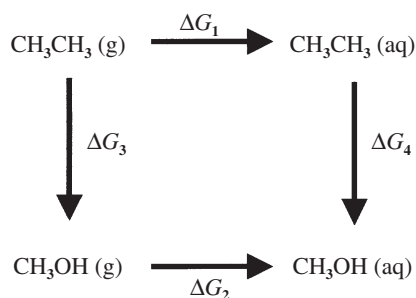
$$\Delta \Delta G = \Delta G_2 - \Delta G_1 \quad (13)$$

Jsou to úlohy typu stanovení rozdílů v solvataci, asociaci, stabilitě enzymů, katalýzy nebo aktivovaného komplexu²⁷. Jako příklad si uvedme výpočet Gibbsovy energie rozdílů v rozpouštění plynného (g) methanolu a plynného ethanu ve vodě (aq)^{13,28,29}. Postup výpočtu si můžeme znázornit následujícím termodynamickým cyklem, který zavedli do počítačové simulace Tembe a McCammon⁸.

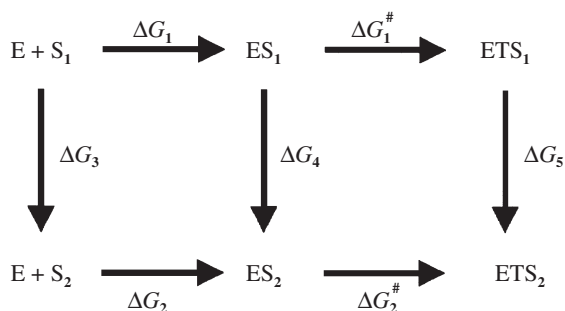
Vzhledem k tomu, že Gibbsova energie je stavová funkce, hodnota její změny při přechodu z počátečního rovnovážného stavu do konečného rovnovážného stavu je nezávislá na cestě mezi těmito dvěma stavy. To znamená, že bude rovněž platit

$$\Delta \Delta G = \Delta G_4 - \Delta G_3 \quad (14)$$

Jestliže (13) odpovídá reálné chemické (fyzikální) cestě,



Obr. 1. Termodynamický cyklus pro porovnání změny Gibbsovy energie rozpouštění plynného (g) methanolu a plynného ethanu ve vodě (aq). Chemické cesty: ΔG_1 , ΔG_2 ; alchymistické cesty: ΔG_3 , ΔG_4



Obr. 2. Složený termodynamický cyklus pro porovnání Gibbsovy energie vazby dvou různých substrátů: S_1 a S_2 ke stejnému enzymu E. Volný enzym a volný substrát: E a S; nekovalentní komplex: ES a tranzitní stav: ETS. Chemické cesty: ΔG_1 , $\Delta G_1^\#$, ΔG_2 , $\Delta G_2^\#$; alchymistické cesty: ΔG_3 , ΔG_4 , ΔG_5

pak (14) popisuje fiktivní (nechemickou, nefyzikální) cestu, při které obecně může docházet nejen ke změnám topologie, ale i transmucaci, vzniku či anihilaci prvků. Gao se svými spolupracovníky¹⁴ zavedli pro takovýto přechod označení alchymistická cesta. Dá se ukázat, že na počítači lze alchymistickou cestu obvykle simulovat mnohem snadněji, než cestu chemickou. V našem případě se tak např. dá vyloučit značně náročný výpočet příspěvku vody v kroku odpovídajícímu ΔG_1 a ΔG_2 .

Termodynamické cykly lze na sebe navazovat. Tak např. při interakci enzymu E s různými substráty S_1 a S_2 lze použít obr. 2.

Jestliže E + S značí volný enzym a volný substrát, pak ES je nekovalentní (van der Waalsův neboli Michaelisův) komplex a ETS je tranzitní stav. Potom pro vazebnou Gibbsovu energii platí

$$\Delta\Delta G_{\text{bind}} = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \quad (15)$$

a pro katalytickou Gibbsovu energii platí

$$\Delta\Delta G_{\text{cat}} = \Delta G_2^\# - \Delta G_1^\# = \Delta G_5 - \Delta G_4 \quad (16)$$

Reakci rozšířenou ještě o hydrataci diskutuje např. Cald-

well³⁰. Z rovnice (16) vyplývá možnost vyhnout se kvantově mechanickým výpočtům změny volné energie při přechodu přes tranzitní stav³. Rovněž tak metoda termodynamického cyklu dovoluje výpočet změny volné energie i v případě, kdy se mění hmoty atomů a tím i kinetické energie, takže lze použít postup uvedený v druhé kapitole i v tomto případě. Aplikace ve farmakochemii uvádí např.¹

5. Závěr

Gibbsova energie je významnou veličinou při studiu chování molekul a jejich interakcí. Teorie nabízí několik postupů jak tuto hodnotu vypočítat. Dosavadní praxe se však potýká s jistými potížemi. Relativně přesné výpočty jsou časově náročné a jsou proto vhodné pouze pro jednoduché systémy. Na druhé straně zrychlení výpočtů v případě složitých molekulárních systémů vyžaduje řadu aproximací, což vede pouze ke hrubým odhadům Gibbsovy energie. Počítačová chemie zde tak neřekla ještě poslední slovo. Rozvoj matematických programů a samotných počítačů dávají naději, že bude možno předpovídat průběh i složitých chemických procesů, které by „ve zkumavce“ byly technicky, časově i finančně velmi náročné.

Tato práce byla podpořena Grantovou agenturou České republiky: grant 307/96/K226 a Ministerstvem průmyslu a obchodu České republiky: grant #PP-Z1/24/1999–2000.

LITERATURA

1. Mark E. A., v knize: *Encyclopedia of Computational Chemistry*, str. 1070. Wiley, Chichester 1998.
2. Malijevský A.: *Chem. Listy* 91, 454 (1997).
3. Kollman P.: *Chem. Rev.* 93, 2395 (1993).
4. Straatsma T. P., Berendsen H. J. C., Postma J. P. M.: *J. Chem. Phys.* 85, 6720 (1986).
5. Kirkwood J. G.: *J. Chem. Phys.* 3, 300 (1935).
6. Kirkwood J. G.: *J. Chem. Phys.* 10, 394 (1942).
7. Zwanzig R. W.: *J. Chem. Phys.* 22, 1420 (1954).
8. Tembe B. L., McCammon J. A.: *Comput. Chem.* 8, 281 (1984).
9. Berenssen H. J. C., Postma J. P. M., van Gunsteren W. F., v knize: *Molecular Dynamics and Protein Structure* (Hermans J., ed.). Polycrystal Book Service, Western Springs 1985.
10. Beveridge D. L., DiCapua F. M., v knize: *Computer Simulation of Biomolecular Systems* (van Gunsteren W. F., Weiner P.K., ed.). ESCOM, Leiden 1989.
11. Kirkwood J. G., v knize: *Theory of Liquids* (Alder B. J., ed.). Gordon and Breach, New York 1968.
12. Huml K., Vondrášek J.: *Mater. Struct.*, v tisku.
13. DISCOVER: *Manual BIOSYM/Molecular Simulations 2.9.7/95.0/3.00*, sv. I. San Diego 1995.
14. Gao J., Kutzera K., Tidor B., Karplus M.: *Science* 244, 1069 (1989).
15. Wodak S. J., Prévost M., v knize: *Statistical Mechanics, Protein Structures and Protein Substrate Interactions* (Doniah S., ed.). Plenum Press, New York 1994.
16. Lau F. T. K., Karplus M.: *J. Mol. Biol.* 236, 1049 (1994).

17. Beveridge D. L., DiCapua F. M.: *Annu. Rev. Biophys. Chem.* **18**, 431 (1989).
18. Mezei M., Beveridge D. L.: *Ann. N. Y. Acad. Sci.* **482**, 1 (1986).
19. Boresch S., Archontis G., Karplus M.: *Proteins* **20**, 25 (1994).
20. Dill K. A.: *UCSF CAMD Course, San Francisco Jan. 16–18 1997*.
21. Mark A. E., van Gunsteren W. F.: *J. Mol. Biol.* **240**, 167 (1994).
22. Shi Y. Y., Mark A. E., Wang C.-X., Huang F., Berendsen H. J. C., van Gunsteren W. F.: *Protein Eng.* **6**, 289 (1993).
23. Boresch S., Karplus M.: *J. Mol. Biol.* **254**, 801 (1995).
24. Ajay, Murcko M. A.: *J. Med. Chem.* **38**, 4954 (1995).
25. Lee C. Y., Yang P. K., Tzou W. S., Hwanh M. J.: *Protein Eng.* **11**, 429 (1998).
26. Brooks III C. L., Karplus M., Pettitt B. M.: *Adv. Chem. Phys.* **71**, 59 (1988).
27. Madura J. D., Pettitt B. M., McCammon J. A.: *Chem. Phys. Lett.* **141**, 83 (1987).
28. Jorgensen W. L., Ravimohan C.: *J. Chem. Phys.* **83**, 3050 (1985).
29. Singh V. C., Brown F. K., Bash P. A., Kollman P. A.: *J. Am. Chem. Soc.* **109**, 1607 (1987).
30. Caldwell J. W., Agard D. A., Kollman P. A.: *Proteins* **10**, 140 (1991).

K. Huml^a and M. Kratochvíl^b (^a*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague*, ^b*J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Prague*): **Chemistry and Alchemy of Computer Simulations of Molecular Behaviour**

Methods of computer simulations in free energy calculations are briefly discussed. A method of thermodynamic integration is given in more detail to elucidate some features of the procedure. The so-called master equation is presented, including the criticism related to its application. Particularly, the thermodynamic cycle is described in connection with more complex tasks met in practice.