FREE-ENERGY COMPUTER SIMULATIONS FOR THE STUDY OF PROTON TRANSFER IN SOLUTIONS

Mihaly Mezei*
Department of Chemistry and Center for Study
in Gene Structure and Function
Hunter College and the Graduate Center, CUNY
New York, NY 10021, USA

*Present address: Department of Physiology and Biophysics, Mount Sinai School of Medicine, CUNY, New York, NY 10029.

INTRODUCTION

The understanding of the effect of a solvent on proton transfer requires the elucidation of the solvent contribution to the free energy of solvation. Due to the high dimensionality of the configuration space involved in the problem, the solvent effect can only be modeled via computer simulation for a complex polyatomic solute. While simulation of rather large systems have recently become relatively routine, the methods for the calculation of free energy from computer simulation has been still recognized as a computationally exacting task[1-4]. The purpose of this paper is to give a brief overview of the free energy simulation methodology and demonstrate its capabilities on the study of the thermodynamics of the formation of the glycine zwitterion in water.

FREE-ENERGY SIMULATION METHODOLOGY

Unless the calculations are performed in the grand-canonical ensemble, free energy computer simulation techniques are characterized by the path used to connect the two systems between which the free-energy difference is to be calculated in the configuration space and by the quantity chosen whose Boltzmann average is related to the free energy. The path is described by the introduction of a coupling parameter λ into the energy function. The various choices of the coupling parameter generally fall into the following two categories[3]:

$$E(\lambda, \underline{X}^{\mathrm{N}}) = \lambda^{\mathrm{k}} * E_1(\underline{X}^{\mathrm{N}}) + (1 - \lambda)^{\mathrm{k}} * E_{\mathrm{o}}(\underline{X}^{\mathrm{N}}) \quad \text{or} \tag{1}$$

$$E(\lambda, \underline{C}, \underline{X}^{\mathrm{N}}) = E(\lambda * \underline{C}_1 + (1 - \lambda) * \underline{C}_{\mathrm{o}}, \underline{X}^{\mathrm{N}}). \tag{2}$$

Here E_{o} and E_1 are the energy functions for the two systems and λ is chosen in such a way that $\lambda = 0$ and $\lambda = 1$ in equations (1) and (2) describe systems with energy function $E_{\mathrm{o}} and E_1$, respectively. Mathematically speaking, equations (1-2) describe a homeomorphism between E_{o} and E_1. The path described by equation (1) is linear for k=1 and 'nearly

linear' for $k > 1$. The introduction[3] of k> 1 is a generalization of the integral transform introduced by Abrams et al.[5] The symbol $\underline{C}$ in equation(2) stands for the collection of potential parameters (including molecular geometries) whose values differentiate systems 0 and 1. Moving along the path described by equation (1) the system 0 is gradually "fading away" while system 1 is simultaneously being "turned on" — a kind of Chesire cat approach. Equation(2), on the other hand, describes a path where system 0 is continuously deformed into system 1.

As the free energy is a state function, several different paths can be used to calculate a given free-energy difference. A creative example for this is the use of a thermocycle by McCammon and coworkers to replace the "obvious" but computationally unfavorable path with a seemingly more complex but computationally tractable one[6]. The nonlinear path defined by Equation (2) was introduced in conjunction with the perturbation method by Jorgensen[7]. The possibility that a nonlinear path can be singularity-free for thermodynamic integration was noted by Mezei and Beveridge[3] and by Cross[8].

Once the path is specified, the free energy difference between the two states can be obtained in various ways. Thermodynamic integration uses the expression of Kirkwood[9]:

$$\Delta A = A_1 - A_o = \int_0^1 \partial A(\lambda)/\partial \lambda d\lambda \tag{3}$$

$$= \int_0^1 \langle \partial E(\lambda)/\partial \lambda \rangle_\lambda \tag{4}$$

where k is the Boltzmann constant, T is the absolute temperature and the symbol $\langle \ \rangle_\lambda$ stands for the Boltzmann average of the quantity enclosed using $E(\lambda)$ in the Boltzmann factor. The integration can be carried out with a quadrature[5,10,11], (preferably Gaussian) or using the slow-growth method[12] where λ is continuously varied during the simulation. The quadrature method with the exponent k in Equation (1) set to 4 was found to perform reliably and efficiently for the calculation of the free energy of liquid water[13] and for the solvation free energy of lithium and sodium ions[14]. The larger than 1 value for k is required to eliminate singularities in the integrals of equations (3,4) that would otherwise arise from the repulsive core of the potential[3]. The slow growth method was used recently to obtain the free-energy profile (i.e. to calculate the potential of mean force) of the proton transfer in aqueous solution between formic acid and methylamine by Mavri and Hadic[15]. While this latter technique ensures that all λ values will be sampled it was found that for this system there are serious ergodic difficulties ('large hysteresis').

The perturbation method[6,16,17] is based on the expression

$$\Delta A = -kT \ln \langle \exp[-(E_1 - E_o)/kT] \rangle_o \tag{5}$$

where the symbol $\langle \ \rangle_\lambda$ stands for the Boltzmann average of the quantity enclosed using E_o as the energy in the Boltzmann factor. While the perturbation method is widely used, the presence of the exponential in Equation(5) warrants caution for larger changes in λ since exponentiation drastically enlarges statistical fluctuations and thereby can easily introduce large numerical errors[3,18]. Finite difference thermodynamic integration[19] combines these

two methods: the integral of equation(3) is evaluated by approximating the integrand with a finite difference ratio over a small λ interval. The small change in the free energy, needed in the finite difference ratio is calculated with the perturbation method. As only small changes are required in λ, the perturbation method results will be reliable.

The solvation free energy can also be related to various probabilities. These methods include the acceptance ratio method of Bennett[16], the overlap ratio method developed by Jacucci and Quirke[20] (based on the ideas put forward by Bennet[16] and shown to perform well in aqueous systems[21]) and the probability ratio method.

The probability ratio method was originally developed for the determination of the potential of mean force[21], and first applied to the determination of free energy differences by Mezei, Mehrotra and Beveridge[23]:

$$\Delta A = -kT \ln[(P(\lambda)_{\lambda=1}/V_1)/(P(\lambda)_{\lambda=0}/V_o)], \tag{6}$$

where $P(\lambda)$ is the Boltzmann probability of the system to be at the intermediate stage λ when λ is also a variable during the simulation and V_o, V_1 represent the configuration space volume corresponding to the $\lambda = 0$ and 1 state, respectively. A familiar example for $P(\lambda)/V_\lambda$ is the radial distribution function $g(r)$ where V is the well known $4\pi r^2$ factor. Valleau, Patey and Torrie have recognized that equation(6) translates small free energy differences into large ratios in the probability of sampling and thus this method requires non-Boltzmann sampling with a modified Hamiltonian, $E'(\underline{X}^{\mathrm{N}}, \underline{R}(\lambda))$ (usually referred to as "umbrella sampling")[17,24] to sample λ values whose probability is small:

$$E'(\underline{X}^{\mathrm{N}}, \underline{R}(\lambda)) = E(\underline{X}^{\mathrm{N}}, \underline{R}(\lambda)) + E_{\mathrm{W}}(\lambda). \tag{7}$$

The Boltzmann average $\langle Q \rangle_{\mathrm{B}}$ of any quantity Q can be recovered as

$$\langle Q \rangle_{\mathrm{B}} = \langle Qw(\lambda) \rangle_{\mathrm{W}}/\langle w(\lambda) \rangle_{\mathrm{W}} \quad \text{where} \tag{8}$$

$$w(\lambda) = \exp[E_{\mathrm{W}}(\lambda)/kT] \tag{9}$$

and $\langle \rangle_{\mathrm{W}}$ implies configurational average using the modified Hamiltonian given by equation(7). Most previous calculations determined $E_{\mathrm{W}}(\lambda)$ empirically, either in tabular or in analytical form. The fact that the best choice for $E_{\mathrm{W}}(\lambda)$ is $W(\lambda)$ suggested iterative approaches that not only provided computational efficiency but resulted in a method that is inherently self- checking. Paine and Scheraga[25] obtained the gas-phase conformational free energy map of the alanine dipeptide and Mezei recalculated the free energy difference between the C_7 and α_{R} conformations of the alanine dipetide in aqueous solution[26,18]. For the aqueous system several technical difficulties had to be overcome: matching of iterations with large statistical noise, recognition of equilibration phase, guiding the simulation to undersampled regions and others. It turned out to be important that the normalization factors of the estimated probability distributions be continually redetermined through the solution of a nonlinear minimization problem as the calculation proceeds (the "matching" problem of the previous studies). This iterative scheme is called adaptive umbrella sampling.

In recent work on the dimethyl phosphate anion the adaptive umbrella sampling method proved to be significantly more reliable than the use of the harmonic weighting function[27]: on a thermodynamic cycle consisting of three distinct solute conformations the closure error was 8 kcal/mol for the harmonic method and 0.6 kcal/mol with the adaptive method. The potential of mean force between sodium and chloride ions in water has also been calculated with the adaptive method[28]. It is thus expected to perform well also for the study of a proton transfer free-energy profile.

GLYCINE ZWITTERION FORMATION

Glycine forms zwitterion in water at neutral pH. As described in the Appendix the free energy of the zwitterion formation can be estimated as -11 kcal/mol, favoring the zwitterionic form. The gas-phase energy difference has been calculated by ab initio methods most recently by Langlet, Caillet, Evleth and Kassab[29] at the 6-31G* level with geometry optimization (see references therein for earlier ab initio work). As they estimated that at the best level the neutral form is more stable by about 20 kcal/mol, the solvation free energy must favor the zwitterionic form by about 30 kcal/mol. In the study reported here ab initio calculations were performed for the estimation of the gas phase contribution up to 2nd order Moller-Plesset level and the difference between the solvation free energies was calculated by thermodynamic integration. A major limitation of the calculations reported here is the use of a prefixed geometry for both the neutral and the zwitterionic form. Thus the results of the present calculation could be refined by exploring the effect of intramolecular conformations, requiring both additional ab-initio studies and potential of mean force calculation along the various torsion angles describing the two molecules.

Table 1
Calculated ab-initio energies

Basis set	Glycine	Zwitterion	Difference	
			a.u.	kcal/mol
STO-3G	-279.1029	-278.9666	0.1363	85.5
6-31G	-282.6697	-282.6424	0.0273	17.1
6-311G**	-282.8862	-282.8515	0.0347	21.8
6-311G^{++}	-282.7537	282.7288	0.0249	15.6
6-311G**/MP2	-283.7546	-283.7182	0.0364	22.8

Ab Inito Calculations

The molecular geometries were taken form the work of Clementi and coworkers[30,31]. Calculations were done at the STO-3G, 6-31G, 6-311G++ and 6-311G** levels[32]. The 6-311G** calculation also included correlation contributions at the MP2 level. The calculated

energies are given in Table 1. The highest level calculations confirm the estimate of Langlet et al. The fractional charges on the atoms obtained by Mulliken population analysis of the ab-initio calculations are given in Table 2.

Table 2
Fractional charges obtained by Mulliken population analysis

	Glycine				Glycine zwitterion		
	STO-3G	6-31G	6-311**		STO-3G	6-31G	6-311**
O(H)	-0.2786	-.6935	-0.4249	O	-.4850	-.7233	-0.6246
O(C)	-0.2607	-.5718	-0.4741	O	-.4294	-.6757	-0.5804
C(O)	0.2951	.7663	0.5230	C(O)	.2406	.7269	0.4882
C(A)	-0.0407	-.2001	-0.0361	C(A)	-.0483	-.2775	-0.1428
N	-0.3855	-.8107	-0.5242	N	-.3583	-.8788	-0.3194
H(N)	0.1476	.3182	0.1940	H(N)	.3326	.4789	0.3160
H(N)	0.1584	.3326	0.2038	H(N)	.2901	.4325	0.2736
H(O)	0.2161	.4368	0.2761	H(N)	.2901	.4325	0.2736
H(C)	0.0617	.1916	0.1188	H(C)	.0837	.2422	0.1578
H(C)	0.0865	.2307	0.1434	H(C)	.0837	.2422	0.1578

Free Energy Simulations

The water-water interactions were described by the TIP4P potential[33] and the solute-water interactions used the AMBER parameter set[34]. The atomic charges were taken from the Mulliken population analysis described above. As the fractional charges are strongly basis set dependent, calulations were performed with both the charges from the STO-3G and the 6-311G** results. The glycine was surrounded by 215 waters in a cell corresponding to the FCC close packing (truncated octahedron).

The simulations used the path described by equation(1) and evaluated the free energy differences by thermodynamic integration, i.e. using equation (4). The integrals were approximated by Gaussian quadratures. At first, the free-energy difference between the two glycine forms was calculated using the charges derived from the STO-3G calculations with a 5-point quadrature and λ exponent $k = 4$. Next for both glycine forms the free energy differences between the models using the STO-3G and 6-311G** charges were calculated with 3-point quadratures and λ exponent $k = 1$ (as the change from one model to an other did not involve the introduction of new repulsion centers). The diagram below gives the calculated free energy differences calculated from the simulations as well as the free energy differences derived form them and the ab initio calculations, in kcal/mol. The error estimates were obtained by the method of batch means[35,36] and represent 95% confidence intervals (2 S.D.).

Glycine	-32.0±2.3	Glycine zwitterion
(STO-3G charges)	$\longrightarrow$	(STO-3G charges)

$$
\begin{array}{ccc}
\text{Glycine} & -32.0\pm2.3 & \text{Glycine zwitterion} \\
\text{(STO-3G charges)} & \longrightarrow & \text{(STO-3G charges)} \\
\big| & & \big| \\
\big|\ -3.8\pm0.2 & & \big|\ -9.2\pm0.2 \\
\text{V} & & \text{V} \\
\text{Glycine} & -37.4\pm2.3 & \text{Glycine zwitterion} \\
\text{(6-311G** charges)} & \longrightarrow & \text{(6-311G** charges)} \\
& \text{(indirect)} &
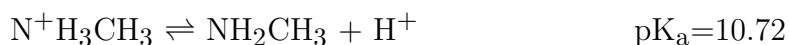\end{array}
$$

Total free energy difference: -9.2±2.3   (STO-3G charges)
                                  -14.6±2.3 (6-311G** charges)

Considering the approximations made in describing the intermolecular interactions as well as the lack of complete geometry optimization, the final result compares rather well with the experimental estimate of -11 kcal/mol, showing that the methods desribed here are capable of a reasonable treatment of the thermodynamics of proton transfer.

Appendix

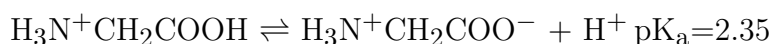Estimate of the experimental free energy of the zwitterion formation

The comparison of the pK_a of several amines show surprisingly little variation[37]:

$NH_4^+ \rightleftharpoons NH_3 + H^+$ pK_a=9.26

$N^+H_3(CH_2)_2OH \rightleftharpoons NH_2(CH_2)_2OH + H^+$ pK_a=9.50

$N^+H_3CH_3 \rightleftharpoons NH_2CH_3 + H^+$ pK_a=10.72

$N^+H_3CH_2CH_3 \rightleftharpoons NH_2CH_2CH_3 + H^+$ pK_a=10.67

$H_3N^+CH_2COO^- \rightleftharpoons H_2NCH_2COO^- + H^+$ pK_a=9.6

Thus, it is reasonable to assume that

$H_3N^+CH_2COOH \rightleftharpoons H_2NCH_2COOH + H^+$ pK_a=9.6 .

Combining this estimate with

$H_3N^+CH_2COOH \rightleftharpoons H_3N^+CH_2COO^- + H^+$ pK_a=2.35

gives (by dividing the two equilibrium constant expressions)

$[H_3N^+CH_2COO^-]/[H_2NCH_2COOH] = 10^{-8.2}$.

Using $\Delta A = -kT \ln K$ results in $\underline{\Delta A = \text{-11 kcal/mol}}$.

References

1. A. Pohorille, and L.R. Pratt, Methods in Enzymology, Biomembranes. 127:64 (1986).

2. D. Frenkel, In Molecular-Dynamics Simulation on Statistical Mechanical Systems, Soc. Italiana di Fisica, Bologna, (1986).

3. M. Mezei, and D.L. Beveridge, Ann. Acad. Sci. N.Y. 482:1 (1986).

4. D.L. Beveridge and F.M. DiCapua, Annu. Rev. Biophys. Chem. 18: 431 (1989).

5. M.R. Mruzik, F.F. Abraham, D.E. Schreiber, and G.M. Pound, J. Chem. Phys. 64:481 (1976).

6. J.A. McCammon, Science 238:486 (1987).

7. W.L. Jorgensen, and C. Ravimohan, J. Chem. Phys. 83:3050 (1985).

8. A. Cross, Chem. Phys. Letters 128:98 (1986).

9. J.G. Kirkwood, In Theory of Liquids. B.J. Alder, Ed. Gordon and Breach. New York, NY, (1968).

10. M. Mezei, Molec. Phys. 47:1307 (1982); Erratum: Ref. 13.

11. M. Mezei, S. Swaminathan, and D.L. Beveridge. J. Am. Chem. Soc. 100:3255 (1978).

12. H.J.C. Berendsen, J.P.M. Postma, and W.F. van Gunsteren, in Molecular Dynamics and Protein Structure, edited by J. Hermans, (Polycrystal Book Service, Western Springs, Illinois, 1985) p43. Chem. Soc. 17:55 (1982).

13. M.Mezei, Molec. Phys. 67:1205 (1989) (erratum to Ref. 10).

14. P.V. Maye and M. Mezei, submitted.

15. J. Mavri and D. Hadic, J. Comp. Chem. in print; this volume.

16. C.H. Bennet, J. Comp. Phys., 22:245 (1976).

17. G.M. Torrie, and J.P. Valleau, J. Comp. Phys. 23:187 (1977).

18. M. Mezei, <u>Molecular Simulation</u> 3301 (1989).

19. M. Mezei, <u>J. Chem. Phys</u>. 86:7084 (1987).

20. G. Jacucci, and N. Quirke, <u>Molec. Phys</u>. 40:1005 (1980).

21. M.Mezei, <u>Molec. Phys</u>. 65:219 (1988).

22. A. Ben-Naim, Water and Aqueous Solutions, Plenum Press, New York (1974).

23. M. Mezei, P.K. Mehrotra, and D.L. Beveridge, <u>J. Am. Chem. Soc</u>. 107:2239 (1985).

24. G. Patey and J.P. Valleau, <u>J. Chem. Phys</u>. 63:2334 (1977).

25. G.M. Paine, and H.A. Scheraga, <u>Biopolymers</u> 24:1391 (1985).

26. M. Mezei, J. Comp. Phys., 68:237 (1987).

27. B. Jayaram, M. Mezei, and D.L. Beveridge. 1988. <u>J. Am</u>. <u>Chem. Soc</u>. 110:1691 (1988).

28. R. Friedman and M. Mezei, to be published.

29. J. Langlet, J. Caillet, E. Evleth, and E. Kassab, in "Modeling of Molecular Structures and Properties", J.L. Rivail, Ed., <u>Studies in Physical and Theoretical</u> <u>Chemistry</u> 71:345 (1990), Elsevier, Amsterdam.

30. E. Clementi, F. Cavallone, and R. Scordamaglia, <u>J. Am.</u> <u>Chem. Soc</u>. 99:5531 (1977).

31. L. Carozzo, G. Corongiu, C. Petrongolo, and E. Clementi, <u>J. Chem. Phys</u>. 68:787 (1978).

32. GAUSSIAN-82 from the laboratory of J.A. Pople, adapted to IBM by S. Topiol and R. Osmond.

33.W.L. Jorgensen, J. Chandrashekar, J.D. Madura, R. Impey, and M. Klein, <u>J. Chem. Phys</u>. 79:926 (1983).

34.S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P.K. Weiner, <u>J Am. Chem. Soc</u>. 106:765 (1984).

35.R.B. Blackman, and J.W. Tuckey, The Measurement of Power Spectra. Dover (1958).

36.W.W. Wood, Physics of Simple Liquids. H.N.V Temperly, F.S. Rowlinson, and G.S. Rushbrooke, Eds. North-Holland (1968).

37. D.G. Peters, J.M. Hayes and G.M. Hieftje, Chemical Separations and Measurements, Saunders, Philadelphia (1979).