

A Simple Method To Estimate Free Energy from a Molecular Simulation: Renormalization on the Unit Interval

B. Jayaram[†] and D. L. Beveridge*

Department of Chemistry, Hall-Atwater Laboratories, Wesleyan University, Middletown, Connecticut 06457
(Received: January 3, 1990; In Final Form: April 4, 1990)

A theory to estimate the free energies of aqueous solutions expeditiously from a single conventional molecular simulation is proposed. Free energies are expressed as a function of the average internal energy and minimum energy by transforming the problem to the unit interval and evaluating much of the expression analytically. The theory is illustrated by applications to liquid water models and ionic solutes in water. The free energies of liquid water models (TIP4P, MCY-CI, and SPC/E) are in excellent accord with the results from other methods (average error is less than 3%). The free energies of ions (Li^+ , Na^+ , and Cl^-) in water are reasonably good (average error is less than 20%), considering the approximation involved. Applications to aqueous solutions of NaDNA in the presence of a simple salt indicate that the magnitude of the calculated free energies are consistent both qualitatively and quantitatively with polyelectrolyte theory for counterion condensation.

Introduction

There is considerable current research interest in the numerical calculation of the free energy of chemical and biomolecular systems via molecular simulation.¹ Various methods such as thermodynamic integration, perturbation method, and potential of mean force calculations have been applied to this problem, all of which to date involve a series of simulations, each of which are

individually costly and time consuming. There is a critical need for a theoretically robust and computationally efficient algorithm for free energies in which the result can be obtained much more rapidly. We explore herein the possibility of obtaining accurate estimates of free energy from a single conventional simulation.

The essence of the proposed procedure is a renormalization to the unit interval, whereby the configurational energy is trans-

[†] Present address: Department of Chemistry, Indian Institute of Technology, Hauz Kaus, New Delhi, 110016 India.

(1) Beveridge, D. L.; DiCapua, F. M. *Ann. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.

formed by using an auxiliary parameter ξ to the interval $0 \leq \xi \leq 1$. Here we discovered that a number of terms in the free-energy expression can be evaluated analytically, at what corresponds to be the limit of infinity in numerical sampling. What remains is a simple expression for free energy in terms of quantities routinely obtained from a single simulation.

Test cases for this procedure are the systems of liquid water, hydrated alkali metal cations, halide anions, and methane and aqueous solutions of a model of B-DNA in the presence of excess salt. Preliminary results demonstrate sufficient correlation with experiment and with other existing methods to generate confidence in the viability of the procedure.

Theory

The Helmholtz free energy A of a system is given¹⁻³ as

$$A = kT \ln \langle e^{E/kT} \rangle \quad (1)$$

where $E = E(X_i^N)$ is the total interaction energy of the system in configuration X_i^N , k is the Boltzmann constant, T is the temperature of the system, and angular brackets $\langle \rangle$ denote an ensemble average. The quantity A in the ensuing discussion refers to the excess Helmholtz free energy with ideal gas as the reference state.

Let $E_r = E_{\max} - E_{\min}$ represent the energy range spanned by the system with E_{\max} and E_{\min} standing for the maximum and minimum energies attained by the system. Note that E_{\max} is not a well-defined quantity in the context of a molecular simulation since it may depend on the run length. Thus, the true E_r per se cannot be evaluated reliably from a simulation. Nevertheless, let the energies E be expressed in dimensionless units through a variable ξ as

$$\xi = (E - E_{\min})/E_r \quad (2)$$

It follows from eq 2 that $0 \leq \xi \leq 1$. With these definitions, the total configurational energy of the system can be expressed as

$$E = E_{\min} + E_r \xi \quad (3)$$

Equation 1 may now be rewritten as

$$A = E_{\min} + kT \ln \langle e^{E_r \xi/kT} \rangle \quad (4)$$

We have thus far carried out only simple algebraic manipulations and eq 1 is not different from eq 4 except for certain definitions. Both eqs 1 and 4 give identical results if a mean energy simulation history of configurational energies is used for the direct evaluation of free energies, but these results are generally inaccurate due to an inadequate sampling of the configuration space, particularly the high-energy regions.

The problem as stated in eq 4, however, can be approached in the following manner: replace the ensemble (Boltzmann) averages in eq 4 of a uniformly distributed variable $E_r \xi$ with ensemble averages corresponding to a uniform distribution (symbolized below as $\langle \rangle_u$) of a Boltzmann distributed variable $E_r^* \xi^*$.

$$\langle e^{E_r \xi/kT} \rangle = \langle e^{E_r^* \xi^*/kT} \rangle_u$$

This implies that $e^{E_r^* \xi^*/kT}$ carries information on both the width and the shape of the energy distribution. Once this is done, the averages corresponding to a uniform distribution can be calculated as a simple integral over the whole unit interval; viz.,

$$\langle e^{E_r \xi/kT} \rangle = \langle e^{E_r^* \xi^*/kT} \rangle_u = \int_0^1 e^{E_r^* \xi^*/kT} d\xi^* \quad (5)$$

$$\int_0^1 e^{E_r^* \xi^*/kT} d\xi^* = \frac{e^{E_r^*/kT} - 1}{E_r^*/kT} \quad (6)$$

The free energy from eq 4 and 6 can then be estimated as

$$A \simeq E_{\min} + E_r^* - kT \ln (E_r^*/kT) \quad (7)$$

Additional clarification of eq 5 and 6 is given in Appendix A. The connection with Gibbs' formulation of canonical ensemble is outlined in Appendix B. The method proposed here to obtain eq 5 is analogous to the scheme of Metropolis et al.⁴ and is straightforward once the configurations are generated according to the Boltzmann distribution. Conceptually the steps taken so far have a simple interpretation: A transition from eq 1 to eq 7 is merely tantamount to a transformation of the problem of the determination of the excess free energy A to the problem of estimation of E_r^* , a function of the energy distribution.

We now focus on the evaluation of E_r^* . A multiplication of both sides of eq 3 by $e^{-\xi}$ results in

$$Ee^{-\xi} = E_{\min}e^{-\xi} + E_r \xi e^{-\xi} \quad (8)$$

Define now the following transform:

$$\langle X \rangle_t = \frac{\int_0^1 X e^{-\xi} d\xi}{\int_0^1 e^{-\xi} d\xi} \quad (9)$$

Application of this transform to both sides of eq 8 leads to

$$\begin{aligned} \langle Ee^{-\xi} \rangle_t &= \langle E_{\min}e^{-\xi} + E_r \xi e^{-\xi} \rangle_t \\ &= E_{\min} \langle e^{-\xi} \rangle_t + E_r \langle \xi e^{-\xi} \rangle_t \end{aligned} \quad (10)$$

Solving for E_r from eq 10 gives

$$E_r = \frac{\langle Ee^{-\xi} \rangle_t}{\langle \xi e^{-\xi} \rangle_t} - \frac{E_{\min} \langle e^{-\xi} \rangle_t}{\langle \xi e^{-\xi} \rangle_t} \quad (11)$$

We now wish to express $\langle Ee^{-\xi} \rangle_t$ in eq 11 in terms of the average internal energy of the system $\langle E \rangle$ as determined from the simulations. The strategy is to relate $\langle Ee^{-\xi} \rangle_t$ to Boltzmann averages of the energy and replace E_r by E_r^* . Let

$$\langle Ee^{-\xi} \rangle_t = \langle E \rangle f(E, \xi) \quad (12)$$

where $f(\equiv f(E, \xi))$ is as yet unknown. f is a distribution function. We hypothesize that f is a normal distribution function.⁵ Specifically f is assumed to be the probability of a normally distributed random variable, a function of energy, taking a value between $-\infty$ and $\langle \xi e^{-\xi} \rangle_t$ in the limit of infinite numerical realization. Then

$$f = \frac{\langle Ee^{-\xi} \rangle_t}{\langle E \rangle} = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\langle \xi e^{-\xi} \rangle_t} e^{-s^2/2} ds \quad (13)$$

or

$$\langle Ee^{-\xi} \rangle_t = \langle E \rangle \mathcal{N}(\langle \xi e^{-\xi} \rangle_t) \quad (14)$$

where $\mathcal{N}(x)$ denotes the normal distribution given in eq 13. Equation 11 may now be rewritten as

$$E_r^* = \frac{\langle E \rangle \mathcal{N}(\langle \xi e^{-\xi} \rangle_t)}{\langle \xi e^{-\xi} \rangle_t} - \frac{E_{\min} \langle e^{-\xi} \rangle_t}{\langle \xi e^{-\xi} \rangle_t} \quad (15)$$

The coefficients for $\langle E \rangle$ and E_{\min} in eq 15 are amenable to analytical treatment. The transforms $\langle e^{-\xi} \rangle_t$ and $\langle \xi e^{-\xi} \rangle_t$ can be evaluated as

$$\langle e^{-\xi} \rangle_t = (e + 1)/(2e) = 0.6839 \quad (16)$$

$$\langle \xi e^{-\xi} \rangle_t = (e^2 - 3)/(4e(e - 1)) = 0.2349 \quad (17)$$

and

$$\mathcal{N}(\langle \xi e^{-\xi} \rangle_t) = 0.5929 \quad (18)$$

From eqs 15-18, we have

$$E_r^* = 2.5241 \langle E \rangle - 2.9115 E_{\min} \quad (19)$$

This E_r^* is substituted in eq 6 to obtain free energies as

$$A = E_{\min} + kT \ln \{ [e^{E_r^*/kT} - 1]/(E_r^*/kT) \} \quad (20)$$

(2) Mezei, M.; Beveridge, D. L. Free Energy Simulations. In *Computer Simulations and Biomolecular Systems*; Beveridge, D. L., Jorgenson, W. L., Eds.; Annals of New York Academy of Sciences, 494; New York Academy of Sciences: New York, 1986.

(3) Valleau, J. P.; Torrie, G. M. *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 6, Chapter 5.

(4) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(5) Feller, W. *An Introduction to Probability Theory and its Applications*; John Wiley & Sons: New York, 1968; Vol. I.

TABLE I: Calculated Transfer Free Energies of Hydration (kcal/mol)

system	$\langle E \rangle$	E_{\min}	ΔA^h	$\Delta A(\text{expt})$	$\Delta A(\text{PM, TI})$	$\Delta A(\text{D})^a$
[H ₂ O] _{TIP4P}	-2180.39 ± 1.30	-2274.32	-5.37 ± 0.02	-5.74 ^b	-5.3, ^b -5.4 ^c	-9.3
[Li ⁺] _{aq}	-2329.76 ± 2.51	-2402.91	-158.0 ^f ± 7.6	-122.1 ^{e,g}		-289.9
[Na ⁺] _{aq}	-2293.86 ± 2.91	-2378.38	-114.3 ^f ± 8.5	-98.2 ^{e,g}	-135.0 ^{d,f}	-242.0
[Cl ⁻] _{aq}	-2257.77 ± 3.14	-2344.89	-87.2 ^f ± 9.0	-77.0 ^{e,g}	-79.3 ^{c,f}	-194.8
[CH ₄] _{aq}	-2190.96 ± 4.12	-2291.48	0.7 ± 11.3	2.0 ^{e,g}	2.27 ^{e,g}	-118.1

^a Free energies from direct evaluation using $kT \ln \langle e^{E_r^*/kT} \rangle$. ^b Reference 13. ^c Reference 14. ^d Reference 15. ^e Reference 16. ^f With Born correction. ^g Gibbs free energies. ^h ΔA for liquid water refers to free energy with ideal gas as the reference state. ΔA for solutes in water refers to free energy of transfer from vacuum to water.

This expression can be further simplified. From eq 20, if $e^{E_r^*/kT} \gg 1$ (and it generally is $> 10^{10}$ in a typical molecular simulation), then

$$e^{E_r^*/kT} - 1 \simeq e^{E_r^*/kT}$$

and

$$A = E_{\min} + E_r^* - kT \ln \left(\frac{E_r^*}{kT} \right) \quad (21)$$

Substituting for E_r^* from eq 19 in eq 21 leads to

$$A = 2.5241 \langle E \rangle - 1.9115 E_{\min} - kT \ln \{ (2.5241 \langle E \rangle - 2.9115 E_{\min}) / kT \} \quad (22)$$

or

$$A = \langle E \rangle - [C_1 E_{\min} - C_2 \langle E \rangle + kT \ln \{ (C_2 + 1) \langle E \rangle - (C_1 + 1) E_{\min} \} / kT] \quad (23)$$

with $C_1 = 1.9115$ and $C_2 = 1.5241$. The principal result of this work is contained in eq 23.

The new feature here is that closed-form solutions to the excess free energies for aqueous solutions are obtained as a function of the mechanical properties of the system via an expression obtained from transforms $\langle X \rangle_t$ to the unit interval. Both the average internal energy $\langle E \rangle$ and the minimum energy E_{\min} can be computed easily from any Metropolis Monte Carlo or molecular dynamics simulation and their statistical uncertainties specified.^{6,7} Given a knowledge of $\langle E \rangle$ and E_{\min} from the equilibrated phase of the system at a given temperature, the total free energy of the system can be estimated from eq 23 with little or no extra computational effort.

Numerical calculations are required to ascertain the accuracy of these estimates, as described in the following section.

Calculations

Illustrative free-energy calculations are performed on the TIP4P model of liquid water⁸ and on the theoretical models of a few monovalent ions⁹ and methane¹⁰ in water. For this study, new Metropolis Monte Carlo⁴ computer simulations are carried out on (1) liquid water, (2) [Na⁺]_{aq}, (3) [Li⁺]_{aq}, (4) [Cl⁻]_{aq}, and (5) [CH₄]_{aq} at a temperature of 298 K and a density of 1 g/mL. All of these systems have been the subject of previous Monte Carlo simulations from this group,^{7,11,12} and citations to the background literature for each case are contained therein. The system in each

case consisted of 216 particles. Periodic boundary conditions are imposed on the system. Solvent-solvent and solute-solvent interactions are truncated with a spherical cutoff of 7.75 Å. Spherical cutoff for ions allows an easy application of Born correction. In each case, 5 million configurations were generated starting with a preequilibrated system, with the acceptance ratio maintained around ~0.5. Ensemble averages of the internal energies ($\langle E \rangle = U$) were formed over the last 4 million configurations. The minimum energy E_{\min} attained by the system during the course of the run was recorded. In the case of liquid water, the runs were extended to 20 million configurations to provide a better reference point and also to check the efficacy of a direct evaluation of free energies from eq 1 with very long run lengths. These calculations were performed on CRAY Y-MP with a turnover rate of 1 million configurations per 25 min.

Results

The average internal energy of the whole system, the minimum energy, and the free energy (in kcal/mol) calculated by using eq 23 are collected for each system in Table I in columns 2–4, respectively. Also, in columns 5 and 6 the experimental free energies and those determined by other methods are listed where available.^{13–16}

The results for TIP4P water were found to be extremely satisfactory. In addition to the regular sources of error in the estimation of internal energy such as the quality of the potential functions, periodic boundary conditions, pairwise additivity in intermolecular interaction energies, and truncation of intermolecular interaction potentials, a likely source of error in free energies estimated via eq 23 is in the determination of minimum energies. Even if the error on E_{\min} were 50 kcal/mol, an unlikely proposition, the error in the free energies would be less than 10% for liquid water. This is highly encouraging.

It is interesting to estimate the free energy of the MCY-CI model¹⁷ of liquid water with this theory. Equation 23 in its simplest form reads as $A = 2.5 \langle E \rangle - 1.9 E_{\min}$. The average internal energy of MCY water is reported as -1868 kcal for a 216-particle system at a temperature of 298 K and a density of 1 g/mL.¹⁸ If we assume E_{\min} to be lower than the average internal energy by about 94 kcal (i.e., $E_{\min} \sim -1962$ kcal), based on the results of TIP4P water from Table I, the total free energy is estimated as $\{2.5(-1868) - 1.9(-1962)\} = -942.2$ kcal for a 216-particle system, which is $(-942.2/216) = -4.36$ kcal/mol. The reported value for the free energy of MCY water from thermodynamic integration is -4.31 kcal/mol.¹⁹ While such a coincidence is fortuitous, even if the assumed E_{\min} of -1962 above differs by 80 kcal from the actual E_{\min} in a MCY liquid water simulation, the estimated free energies would differ only by 20%. A similar treatment of SPC/E water²⁰ ($\langle E \rangle = -2137$ kcal²⁰ for 216 waters; $E_{\min} \sim -2231$ kcal

(6) Erpenbeck, J. E.; Wood, W. W. In *Modern Theoretical Chemistry*; Berne, B. J., Ed.; Plenum: New York, 1977; Vol. 6, Chapter 2.

(7) Beveridge, D. L.; Mezei, M.; Mehrotra, P. K.; Marchese, F. T.; Ravishanker, G.; Vasu, T. R.; Swaminathan, S. In *Molecular Based Study of Fluids*; Haile, J. M.; Mansoori, G. A., Eds.; Advances in Chemistry Series 204; American Chemical Society: Washington, DC, 1983; pp 297–351.

(8) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(9) Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903.

(10) Jorgensen, W. L.; Gao, J.; Ravimohan, C. *J. Phys. Chem.* **1985**, *89*, 3470.

(11) Mezei, M.; Beveridge, D. L. *J. Chem. Phys.* **1981**, *74*, 6902.

(12) Swaminathan, S.; Harrison, S. W.; Beveridge, D. L. *J. Am. Chem. Soc.* **1978**, *100*, 3705.

(13) Hermans, J.; Pathiaseril, A.; Anderson, A. *J. Am. Chem. Soc.* **1988**, *110*, 5982.

(14) Jorgensen, W. L.; Blake, J. F.; Buckner, J. K. *Chem. Phys.* **1989**, *129*, 193.

(15) Jayaram, B.; Fine, R.; Sharp, K.; Honig, B. *J. Phys. Chem.* **1989**, *93*, 4320.

(16) Rosseinsky, D. R. *Chem. Rev.* **1965**, *65*, 467.

(17) Matsuoka, O.; Clementi, E.; Yoshimine, M. *J. Chem. Phys.* **1976**, *64*, 1351.

(18) Mehrotra, P. K.; Mezei, M.; Beveridge, D. L. *J. Chem. Phys.* **1983**, *78*, 3156.

(19) Mezei, M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1978**, *100*, 3255.

TABLE II: Calculated Energetics of Hydration (kcal) of Methylamine, Methanol, and Ethane in TIP4P Water at 25 °C

Results of Perturbation Method Calculations ^a			
(1) $\Delta\Delta A$	$\text{CH}_3\text{NH}_2 \rightarrow \text{C}_2\text{H}_6$		8.47
(2) $\Delta\Delta A$	$\text{CH}_3\text{NH}_2 \rightarrow \text{CH}_3\text{OH}$		1.95
(3) $\Delta\Delta A$	$\text{CH}_3\text{OH} \rightarrow \text{C}_2\text{H}_6$		6.9
(4) stability trend	$\text{CH}_3\text{NH}_2 > \text{CH}_3\text{OH} > \text{C}_2\text{H}_6$		
	$\langle E \rangle^a$	E_{\min}^a	ΔA
(5) CH_3NH_2	-2175.00 ± 3.73	-2240.20	-57.37 ± 10.34
(6) CH_3OH	-2176.90 ± 3.63	-2251.63	-39.95 ± 10.16
(7) C_2H_6	-2147.06 ± 5.6	-2222.0	-21.26 ± 14.77
Results with the Present Method			
(8) $\Delta\Delta A$	$\text{CH}_3\text{NH}_2 \rightarrow \text{C}_2\text{H}_6$		36.11 ± 18.03
(9) $\Delta\Delta A$	$\text{CH}_3\text{NH}_2 \rightarrow \text{CH}_3\text{OH}$		17.42 ± 14.50
(10) $\Delta\Delta A$	$\text{CH}_3\text{OH} \rightarrow \text{C}_2\text{H}_6$		18.69 ± 17.93
stability trend	$\text{CH}_3\text{NH}_2 > \text{CH}_3\text{OH} > \text{C}_2\text{H}_6$		

^a Reference 22.

(-2137 - 94 as for TIP4P water); and $A \sim \{2.5(-2137) - 1.9(-2231)\}/216 = -1103.6/216 = -5.1$ kcal/mol gives -5.1 kcal/mol for the Helmholtz free energy compared to -5.5 kcal/mol¹³ reported earlier in the literature. These, in our opinion, demonstrate the utility of the present theory.

The direct evaluation of the excess free energy of liquid water using eq 1 was attempted previously.²¹ The free energies obtained from eq 1 on the diverse systems studied here are shown in the last column in Table I. Nonconvergence of the results obtained from eq 1³ is evident from the data. In this sense, the present methodology (eq 23) may be seen as an improvement over the direct evaluation (eq 1) of free energies.

The solute hydration free energies (given in Table I in kcal/mol) are calculated as follows:

$$\Delta A([X]_{\text{aq}}) = A([X]_{215\text{w}}) - A_{215\text{w}} = A([X]_{215\text{w}}) + 1155.2572 \quad (24)$$

The transfer free energy of a solute according to the above equation is the difference between the total free energy of the system with and without the solute. The free energy of a 215-particle TIP4P water system is -1155.2572 kcal (-1160.6305 \times 215/216) from row 1 and columns 2 and 3 of Table I. The Born correction for monovalent ions with a spherical cutoff of 7.75 Å equals ~ -21.42 kcal, and this has been added to the free energies of ions given in column 4. The transfer free energies of ions are in reasonable accord with estimates from other sources, and the average error incurred is less than 20% when compared with experiment. Thus, the method proposed here can be used for ionic systems with some measure of confidence. For methane, the error in absolute numbers is small although the percentage error is large. This is unavoidable if the absolute free energies are small, since the free energies then are estimated as small differences between large numbers as suggested by eqs 23 and 24. Other more elaborate methods such as the perturbation technique may have an edge over this method here in such an eventuality.

As a cautionary note, we would like to emphasize that for accurate estimates of $\Delta\Delta G^\circ$ or $\Delta\Delta A^\circ$ when the free-energy differences are small, the present method is not a substitute to the perturbation method or other existing free-energy methodologies. The hydration free-energy differences between methylamine, methanol, and ethane in TIP4P water were determined in our laboratory²² via the perturbation method (PM) by forming a three-legged thermodynamic cycle. The calculated free-energy differences are shown in rows 1–3 of Table II. The error on the thermodynamic cycle is ± 0.38 kcal, which is satisfactory. The stability trend obtained was $\text{CH}_3\text{NH}_2 > \text{CH}_3\text{OH} > \text{C}_2\text{H}_6$. The average internal energy ($\langle E \rangle$) and the minimum energy (E_{\min}) for the hydration of these three solutes in TIP4P water were also determined (with an 8.5-Å water–water cutoff and 1.5 million

TABLE III: Calculated Total Excess Free Energies (kcal) for Aqueous Solutions of NaDNA at Different Na⁺Cl⁻ Added Salt Concentrations

concn, mM	$\langle E \rangle$	E_{\min}	A
0	-191.52 ± 1.33	-199.74	-104.38 ± 3.35
25	-199.57 ± 1.51	-209.26	-106.55 ± 3.81
50	-206.97 ± 1.05	-218.73	-107.17 ± 2.65
100	-221.38 ± 1.07	-234.74	-112.99 ± 2.70
150	-235.84 ± 1.05	-245.32	-129.04 ± 2.65

MC steps as run length) during the course of this study.²² These are given in rows 5–7 of Table II. Using these two quantities, $\langle E \rangle$ and E_{\min} , and utilizing the present theory (eq 23 and 24 above), we have been able to estimate the absolute free energies of hydration (last column in rows 5–7) and hydration free-energy differences between these three solutes (recorded in rows 8–10 of Table II). The free-energy estimates in rows 5–7 deserve two comments. These are based on $\langle E \rangle$ and E_{\min} , which are obtained with different run characteristics, a larger water–water cutoff, and shorter run lengths than for TIP4P water and other systems reported in Table I. Thus, a strict determination of absolute free energies is not feasible using eq 24. In principle, the TIP4P number in eq 24 is to be redetermined with a larger cutoff. This, however, is not a cause for concern in estimating the free-energy differences. Also, shorter run lengths starting with a random configuration in general tend to give a less negative E_{\min} , which results (from eq 23) in a more negative free energy as seen in rows 5–7. Nonetheless, the stability trend for hydration in TIP4P water indicated by these numbers in rows 8–10 is $\text{CH}_3\text{NH}_2 > \text{CH}_3\text{OH} > \text{C}_2\text{H}_6$, in qualitative agreement with the PM results. Several points emerge from the results in Table II. Firstly, E_{\min} and hence the absolute free energies estimated via the present theory are sensitive to run characteristics. Secondly, the estimated $\Delta\Delta A^\circ$ values do not agree quantitatively in these three cases with the results of the perturbation method calculations. The free-energy differences in the perturbation method were evaluated via ensemble averages of the differences in solute binding energies, which were small and over several small perturbation steps. In the present method, the free-energy differences are obtained by first calculating the absolute free energies of hydration and then taking the differences. The error bounds thus as expected are large in the latter. This notwithstanding, there is a qualitative agreement between the results emerging from these two methods, which is interesting considering that no extra simulations are required in obtaining the stability trend with the present theory. Lastly, the free energies of hydration are not scaled either as average internal energies or minimum energies. For example, methanol hydration is seen (from rows 5–7, Table II) to be favored over methylamine with either $\langle E \rangle$ or E_{\min} alone, but free energies of hydration are seen to favor methylamine from PM calculations. This is captured by the present theory.

Application to DNA. We have recently reported $[T, V, N]$ ensemble Monte Carlo computer simulations on $[\text{NaDNA}]_{\text{aq}}$ at a temperature of 300 K in the presence of added simple salt.²³ The DNA concentration was fixed at 3 mM. The added salt (Na^+Cl^-) concentrations were at (a) 0, (b) 25, (c) 50, (d) 100, and (e) 150 mM. Different variations on the primitive model for the solvent were investigated in this work. In the following, an application of the free-energy methodology is outlined for the simple Coulombic model, in which electrostatic interactions were treated according to Coulomb's law with a dielectric constant of 80 for solvent water. Results on the energetics are summarized in Table III.

Excess free energies per ion of NaDNA in aqueous solutions are shown in Figure 1 as a function of the number of ions in the system. The results indicate that it is easier free energetically to remove/displace small ions from the NaDNA system at higher ionic strengths than at low salt concentrations, a qualitatively reasonable result.

(20) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(21) Sarsikov, G. N.; Dashevsky, V. G.; Molenkov, G. G. *Mol. Phys.* **1974**, *27*, 1249.

(22) Subramanian, P.; Beveridge, D. L. Unpublished results.

(23) Jayaram, B.; Swaminathan, S.; Beveridge, D. L.; Sharp, K.; Honig, B. *Macromolecules* **1990**, *23*, 3156.

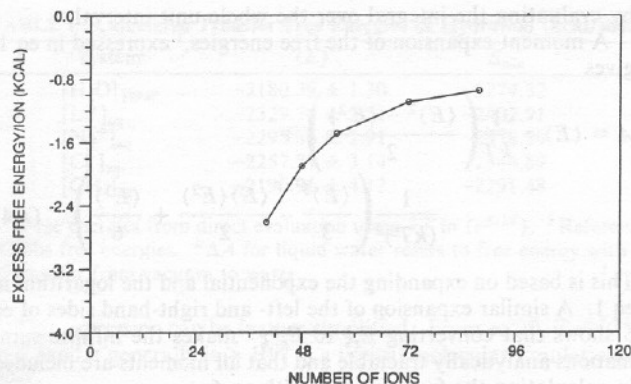


Figure 1. Excess total free energy/ion of NaDNA solutions as a function of added salt concentration shown as number of added salt ions.

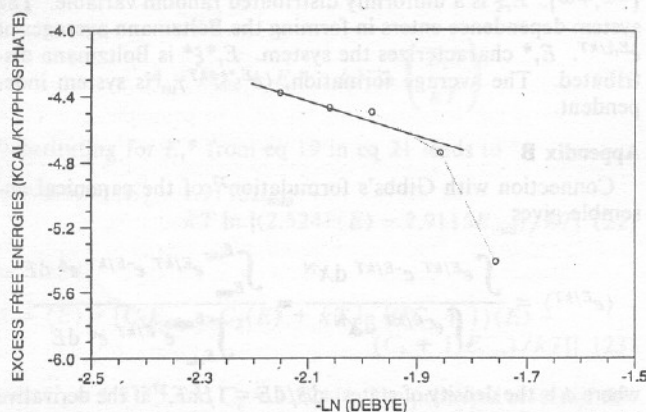


Figure 2. Excess total free energy/phosphate/ kT of aqueous solutions of NaDNA in the presence of added salt shown as a function of $\ln \kappa$, the logarithm of the inverse Debye length.

A quantitative test for the free energies of NaDNA in aqueous solutions containing added salt is necessary to justify the extension of the present free-energy methodology to DNA in aqueous solutions. Manning's limiting law (eq 17 of ref 24), for polyelectrolyte solutions ($A^{ex}/N_p kT = -\xi_M \ln \kappa$), is tested on colligative properties of polyelectrolyte solutions and melting temperatures of DNA,²⁵ and its validity is established at low salt concentrations. It follows from this limiting law that a plot of $A^{ex}/N_p kT$ versus $-\ln \kappa$ gives a straight line with a slope of ξ_M at low added salt concentrations. Note that the Debye length here includes all ionic species in the system. According to the counterion condensation theory,²⁴ ξ_M has a value of 4.2 for B-DNA under vacuum and counterions condense to reduce ξ_M to unity upon transfer to aqueous medium.^{24,26} Thus, the slope in the above plot, ξ_M , must equal 1. The calculated free energies per phosphate in units of kT are shown in Figure 2 as a function of the logarithm of the inverse Debye length. The slope of the straight line (best fit) passing through the points on the graph representing added salt concentrations from 0 to 100 mM is estimated to be 1.1 ± 0.1 . This suggests that the free-energy estimates are quantitatively of the correct magnitude.

Discussion

In this article, we describe a formalism to evaluate the free energies of aqueous solutions expeditiously. This approach is ideally suited for the estimation of free energies associated with a local minimum. In the absence of a rigorous justification for eqs 6 and 14, we make no pretense to having solved the problem of configurational partition function for aqueous solutions.

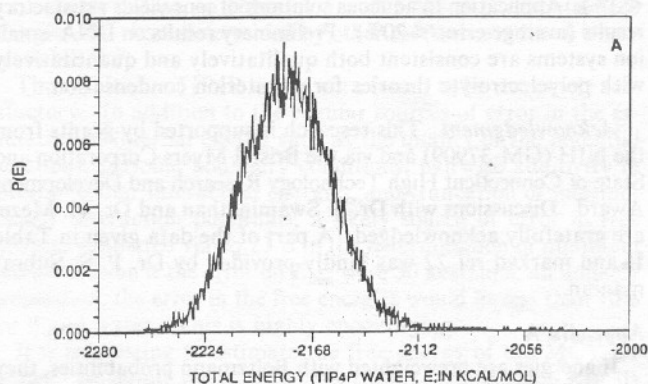
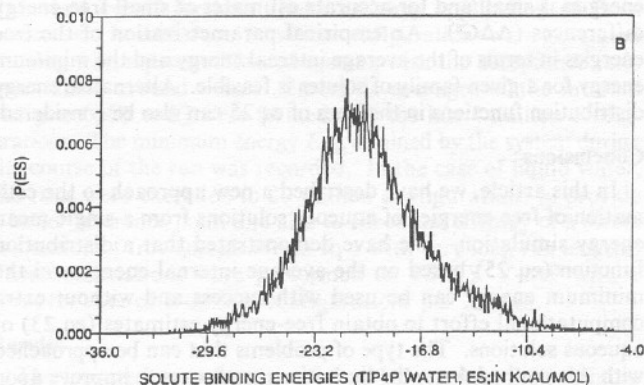
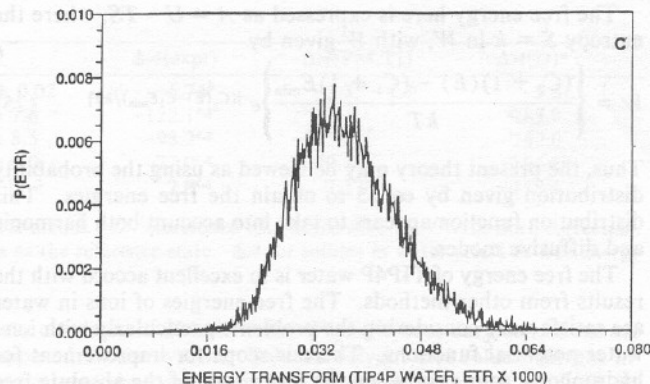


Figure 3. (A) Calculated total energy distribution for TIP4P water shown as a function of total energy. (B) Same as in A for solute binding energies. (C) Same as in A for the energy transform ($\langle Ee^{\xi} \rangle / \langle E \rangle$).

Equation 6 yields results that are derivable from Gibbs' formulation of the theory of canonical ensembles under the condition that the energies of the system are densely distributed closely around the most probable energy of the system (Appendix B). Equation 14 at this stage is a hypothesis motivated by the need to minimize the information to be used from finite length simulations. It is a closure relation cast in the language of the central limit theorem.⁵ Further analysis, as desired by referees, of the results of this theory vis-à-vis simulation is presented in Appendix C.

The minimum energy is a well-defined quantity theoretically, but a problem is anticipated in its realization in practice. It is identified here in our Monte Carlo simulations with the minimum most energy attained by the system (in the postequilibration phase) during the realization of a set of configurations consistent with the Boltzmann distribution at a specified temperature. It is sensitive to characteristics such as the run length and the acceptance ratio in the simulation. As a consequence, the errors are large, particularly when the absolute free energies are small as in the case of methane hydration. The problem is similar to that encountered in the determination of the average internal energies of transfer.⁷

(24) Manning, G. S. *J. Chem. Phys.* **1969**, *51*, 924.

(25) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry, Part III, The behavior of biological macromolecules*; W. H. Freeman and Company: San Francisco, 1980; Chapter 22.

(26) Manning, G. S. *Q. Rev. Biophys.* **1978**, *11*, 179.

The free energy here is expressed as $A = U - TS$, where the entropy $S = k \ln W$, with W given by

$$W = \left\{ \frac{(C_2 + 1)\langle E \rangle - (C_1 + 1)E_{\min}}{kT} \right\} e^{-(C_2\langle E \rangle - C_1 E_{\min})/kT} \quad (25)$$

Thus, the present theory may be viewed as using the probability distribution given by eq 25 to obtain the free energies. This distribution function appears to take into account both harmonic and diffusive modes.

The free energy of TIP4P water is in excellent accord with the results from other methods. The free energies of ions in water are satisfactory considering the problems particularly with ion-water potential functions. There is scope for improvement for hydrophobic solutes or when the magnitude of the absolute free energies is small and for accurate estimates of small free-energy differences ($\Delta\Delta G^\ddagger$). An empirical parametrization of the free energies in terms of the average internal energy and the minimum energy for a given family of solutes is feasible. Alternative energy distribution functions in the lines of eq 25 can also be considered.

Conclusions

In this article, we have described a new approach to the estimation of free energies of aqueous solutions from a single mean energy simulation. We have demonstrated that a distribution function (eq 25) based on the average internal energy and the minimum energy can be used with success and without extra computational effort to obtain free-energy estimates (eq 23) of aqueous solutions. The type of problems that can be approached with this methodology, the limitations, and ways to improve upon the free-energy estimates are discussed. The theory proposed performs extremely well on liquid water models (average error < 3%). Application to aqueous solutions of ions yields satisfactory results (average error < 20%). Preliminary results on DNA-small ion systems are consistent both qualitatively and quantitatively with polyelectrolyte theories for counterion condensation.

Acknowledgment. This research is supported by grants from the NIH (GM-37909) and via the Bristol Myers Corporation and State of Connecticut High Technology Research and Development Award. Discussions with Dr. S. Swaminathan and Dr. M. Mezei are gratefully acknowledged. A part of the data given in Table II and marked ref 22 was kindly provided by Dr. P. S. Subramanian.

Appendix A

If energies are preweighted with Boltzmann probabilities, they contribute to the ensemble averages with a probability of one

$$\langle E \rangle = \frac{1}{N} \sum_i^N E_i^* \quad (A1)$$

where N is the length of the Markov chain or N is the number of configurations. (The superscript * denotes the fact that the energies correspond to the Boltzmann distribution.) Similarly,

$$\langle e^{E_i/kT} \rangle = \frac{1}{N} \sum_i^N e^{E_i^*/kT} \quad (A2)$$

The same arguments apply to $E_r\xi$, the transformed variable of energy:

$$\langle e^{E_r\xi/kT} \rangle = \frac{1}{N} \sum_i^N e^{E_r^*\xi^*/kT} \quad (A3)$$

By replacing the summation in eq A3 with the integral in eq 5, we have accomplished two factors. Firstly, the integral above admits a closed-form solution to the desired averages. Secondly, we are supplementing the history of a finite length simulation with information corresponding to infinite length numerical realization

by evaluating the integral over the whole unit interval.

A moment expansion of the free energies,¹ expressed in eq 1, gives

$$A = \langle E \rangle - \frac{1}{kT} \left(\frac{\langle E \rangle^2 - \langle E^2 \rangle}{2} \right) + \frac{1}{(kT)^2} \left(\frac{\langle E \rangle^3}{3} - \frac{\langle E \rangle \langle E^2 \rangle}{2} + \frac{\langle E^3 \rangle}{6} \right) \dots \quad (A4)$$

This is based on expanding the exponential and the logarithm in eq 1. A similar expansion of the left- and right-hand sides of eq 6 shows that converting $E_r\xi$ to $E_r^*\xi^*$ makes the infinite summations analytically tractable and that all moments are included in calculating the free energies with eq 6.

Summarizing the discussion of eqs 5 and 6 briefly, E_r is not determined by the system. It may be anywhere in the interval $[-\infty, +\infty]$. $E_r\xi$ is a uniformly distributed random variable. The system dependence enters in forming the Boltzmann averages of $e^{E_r\xi/kT}$. E_r^* characterizes the system. $E_r^*\xi^*$ is Boltzmann distributed. The average formation, $\langle e^{E_r^*\xi^*/kT} \rangle_u$, is system independent.

Appendix B

Connection with Gibbs's formulation²⁷ of the canonical ensemble gives

$$\langle e^{E/kT} \rangle = \frac{\int e^{E/kT} e^{-E/kT} dX^N}{\int e^{-E/kT} dX^N} = \frac{\int_{E_{\min}}^{E_{\max}} e^{E/kT} e^{-E/kT} e^\phi dE}{\int_{E_{\min}}^{E_{\max}} e^{-E/kT} e^\phi dE}$$

where ϕ is the density of states. $d\phi/dE = 1/kT$,²⁷ if the derivative is evaluated at the most probable energy. On integration, $\phi = (E_{\max} - E_{\min})/kT = E_r/kT$. Substituting this in the above equation gives

$$\langle e^{E/kT} \rangle = \frac{\int_{E_{\min}}^{E_{\max}} e^{E_r/kT} dE}{\int_{E_{\min}}^{E_{\max}} e^{-E/kT} e^{E_r/kT} dE} = \frac{e^{E_{\min}/kT}}{\int_0^1 e^{-E_r\xi/kT} d\xi}$$

On carrying out the integration in the denominator, one obtains

$$\langle e^{E/kT} \rangle = e^{(E_{\min} + E_r)/kT} [(E_r/kT) / (e^{E_r/kT} - 1)]$$

The free energy then is given as

$$A = kT \ln \langle e^{E/kT} \rangle = E_{\min} + E_r - kT \ln [(e^{E_r/kT} - 1) / (E_r/kT)]$$

On expanding the exponential in the square brackets up to the first term in E_r/kT , $A \approx E_{\min} + E_r$. Note that this is very similar to eqs 7 and 21 above. Usage in eq 21 of E_r^* given by eq 15 presumably corrects for the approximations introduced in deriving the above equation.

Appendix C

In simulations on the systems examined here, all energy-related quantities are seen to tend to a normal distribution. The results on TIP4P water are shown in Figure 3 as an illustrative case. The distribution of the total configurational energy of the system is shown in Figure 3A. The distribution for the solute binding energies, defined as the interaction of a single water molecule (solute) with the rest of the system, is shown in Figure 3B. (Also see ref 7 for a detailed description of these quantities.) The distribution for the energy transform ($\langle e^{E_r\xi} \rangle / \langle E \rangle$) is shown in Figure 3C. On the basis of these results, a normal distribution approximation in eq 14 appears to be reasonable.

Improvements to this theory may be introduced via coefficients C_1 and C_2 in eq 23 or eq 25. Alternative functional forms for the probability distribution in eq 25 may also be explored.