Polynomial Path for the Calculation of Liquid State Free Energies from Computer Simulations Tested on Liquid Water

---

Mihaly Mezei

Department of Physiology and Biophysics, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029

Received 13 September 1991; accepted 9 January 1992

The polynomial path is introduced for the calculation of liquid state free energies. The well-characterized SPC, TIP4P, and MCY water models were used to demonstrate its efficiency, as well as its range of applicability in conjunction with Monte Carlo computer simulations using thermodynamic integration based on Gaussian quadratures. The technique employed is compared with the slow-growth method (another variant of thermodynamic integration), the perturbation method, and the use of the grand-canonical ensemble.

INTRODUCTION AND BACKGROUND

The calculation of the free energy, long recognized as a computationally exacting task, is of special importance due to the role of the free energy in determining chemical equilibria. As reviewed recently,[1-4] free energy simulation techniques are characterized by the path selected that connects the two systems in the configuration space (e.g., the choice of the coupling parameter, *vide infra*) and by the quantity chosen whose Boltzmann average is related to the free energy—with the exception of calculations in the grand-canonical ensemble, where the chemical potential is fixed at the outset. The various choices of the coupling parameter, described in detail in ref. 2, fall into two distinct classes:

$$E(\lambda, \mathbf{X}^N) = f_1(\lambda) * E_1(\mathbf{X}^N) + f_0(\lambda) * E_0(\mathbf{X}^N) \quad (1)$$

or

$$E(\lambda, \mathbf{C}, \mathbf{X}^N) = E[\mathbf{C}(\lambda), \mathbf{X}^N] \quad (2)$$

Here E_0 and E_1 are the energy functions for the two systems between which the free energy difference is to be computed and λ is the so-called coupling parameter that varies from 0 to 1. The functions $f_1(\lambda)$ and $f_0(\lambda)$ are continuous functions with

$$f_1(0) = f_0(1) = 0 \quad (3)$$

and

$$f_1(1) = f_0(0) = 1 \quad (4)$$

so that $\lambda = 0$ or 1 in eq. (1) describes systems with energy function E_0 and E_1, respectively. Moving along the path described by eq. (1), the system 0 is gradually "fading away" while system 1 is simultaneously being "turned on"—a kind of Chesire cat approach. The symbol **C** in eq. (2) stands for the collection of potential parameters (including molecular geometries) and their dependence on λ is such that **C**(0) describes system 0 and **C**(1) describes system 1. Thus, eq. (2) describes a path in the configuration space where system 0 is continuously deformed into system 1. A convenient choice for f_1 and f_0 is

$$f_1(\lambda) = \lambda^k \quad \text{and} \quad f_0(\lambda) = (1 - \lambda)^k \quad (5)$$

Once the path is specified, the free energy difference between the two states can be obtained in various ways. The calculations described in this article used thermodynamic integration (based on ideas of Kirkwood[5]):

$$\Delta A = A_1 - A_0 = \int_0^1 \langle \partial E(\lambda)/\partial\lambda \rangle_\lambda \, d\lambda \quad (6)$$

and reference will be made to the so-called perturbation method:

$$\Delta A = -kT \ln\langle \exp[-(E_1 - E_0)/kT] \rangle_0 \quad (7)$$

where k is the Boltzmann constant, T is the absolute temperature, and the symbols $\langle\,\rangle_\lambda$ $\langle\,\rangle_0$ stand for the Boltzmann average of the quantity enclosed using $E(\lambda)$ and E_0, respectively, as the energy in the Boltzmann factor. It is to be stressed that eq. (7) is exact and thus its use is different from the use of thermodynamic perturbation theories that are based on a truncated expansion.

Using eqs. (1) and (5) for the coupling parameter simplifies eq. (6) into

$$\Delta A = \int_0^1 k[(1 - \lambda)^{k-1}\langle E_0\rangle_\lambda + \lambda^{k-1}\langle E_1\rangle_\lambda] \, d\lambda \quad (8)$$

It has been shown[6] that for $k = 1$ the integrand

CCC 0192-8651/92/050651-06$04.00

in eq. (8) is monotonous, a desirable quality when the use of a quadrature is anticipated. Similar arguments can show that both $\langle E_0 \rangle_\lambda$ and $\langle E_1 \rangle_\lambda$ are monotonous functions of λ; thus, it is reasonable to expect that the integrand of eq. (8) is still going to be "well behaving" (i.e., not have many unexpected extrema). Paths described by eq. (2), however, may contain unspecified number of oscillations, although for the liquid water problem this did not appear to be the case (the integrand was found to have only one maximum in refs. 7 and 8). Furthermore, when the reference system is the ideal gas eq. (1) describes a transcritical path and thus avoids any possible numerical instability that may arise when a phase transition is encountered along the path—a distinct possibility with the use of eq. (2).

The need for $k > 1$ arises for transitions involving the creation and/or annihilation of atoms. Such a situation occurs when the number of atoms in the two systems differ or the conformation of the two systems are different enough that atoms are moved to previously unoccupied positions. In such cases, thermodynamic integration with $k = 1$ leads to a so-called improper integral [i.e., a definite integral where the integrand is singular at the endpoint(s)] and the other methods become numerically unstable. The source of the divergence in the integrand or the instability is the divergence of the interaction potential at $r = 0$. However, for a potential of the form $1/r^e$ the asymptotic behavior of the integrand is known to be $\lambda^{(kd/e)-1}$, where d is the dimensionality of the space[3]; thus, k can be selected high enough that the integrand remains finite everywhere. For the three-dimensional case, the assymptotic behavior was given earlier by Mruzik et al.[9] and Swope et al.[10] The overall behavior of the integrand is dominated by the terms with the highest e. In particular, a $1/r^{12}$ repulsion in three dimensions requires $k \geq 4$. Calculations on the dense Lennard-Jones fluid[11] and on liquid water[12] have demonstrated the efficiency of this approach. Couplings using eq. (2) in general do not encounter this singularity problem.[2,13] Notice also that the path described by eq. (1) results in the same calculation as the integral transform suggested by Mruzik et al.[9] when one of the systems is the ideal gas.

The purpose of this article is to present a generalization of eq. (1), the polynomial path, and demonstrate its efficiency and range of applicability on liquid water models. Liquid water was chosen for the test since it has been the subject of several studies and the results are well established. Once the efficiency of the proposed technique is demonstrated on liquid water, it can then be used for the calculation of the free energies of other liquids, as well as for the calculation of solvation free energies. The article also presents comparisons of the efficiency of thermodynamic integration using Gaussian quadrature, thermodynamic integration with the slow-growth method, the perturbation method, and the use of grand-canonical ensemble for the calculation of liquid free energies.

THEORY

The proposed polynomial path generalizes eq. (1) in such a way that different exponents will be used for different parts of the potential. For a potential energy involving terms with n different powers of interatomic distances $e(i)$, $i = 1, n$:

$$E(\mathbf{X}^N) = \sum_{i=1}^{n} E_{e(i)}(\mathbf{X}^N) \tag{9}$$

the polynomial coupling replaces eq. (1) by

$$E(\lambda, \mathbf{X}^N) = \sum_{i=1}^{n} \lambda^{k[e(i)]} * E_{1,e(i)} + (1-\lambda)^{k[e(i)]} * E_{0,e(i)} \tag{10}$$

where $k[e(i)]$ is the λ exponent to be used with the terms involving the $e(i)$th inverse distance powers. The idea behind this proposal is that the lower the exponent the more evenly will the integration sample the [0, 1] λ interval; thus, the high λ exponent should only be used with the repulsion term. The use of different λ exponents incurs negligible additional computational expense. Calculations described in this article used integer exponents only, but eq. (10) can be used with fractional exponents as well. Hermans et al.[7] used the same formalism to break up the calculation into individual segments where only one of the terms was "grown" to avoid collapsation of the system.

CALCULATIONS

The calculations reported here were run at 298 K temperature and at experimental density (0.997 g/mL). Sixty-four waters were used and face-centered cubic periodic boundary conditions were imposed, allowing a 7.0-Å cutoff. All simulations used the Metropolis algorithm[14] with the force-bias sampling technique of Pangali et al.[15] The integration over the coupling parameter λ was performed using Gaussian quadratures.

The error estimates for the calculated free energies were derived from the error estimates on the integrands calculated at each quadrature point using the method of batch means[16,17] based on 10^5 Monte Carlo step blocks. The estimates given represent 2 standard deviations.

Previous calculations using the exact same setup

Table I. Free energy differences calculated with thermodynamic integration.

					Run length (K)			
					100	300	600	1000
Sys	n_q	$k(1)$	$k(6)$	$k(12)$	A'	A'	A'	A'
SPC	5	4	4	4	−5.66	−5.67 ± 0.06	−5.66 ± 0.04	−5.65 ± 0.05
SPC	8	4	4	4	−5.63	−5.68 ± 0.06	−5.68 ± 0.05	−5.70 ± 0.03
MCY	5	4	4	4	−4.00	−4.03 ± 0.05	−4.03 ± 0.04	−4.04 ± 0.03
SPC	3	4	4	4	−5.65	−5.61 ± 0.12	−5.61 ± 0.06	−5.61 ± 0.04
SPC	3	2	2	4	−5.71	−5.73 ± 0.06	−5.71 ± 0.06	−5.69 ± 0.05
MCY	3	4	4	4	−3.97	−3.97 ± 0.03	−4.02 ± 0.24	−4.04 ± 0.16
TP4	3	4	4	4	−5.48	−5.43 ± 0.08	−5.44 ± 0.07	−5.42 ± 0.06
TP4	3	3	2	4	−5.46	−5.46 ± 0.07	−5.47 ± 0.05	−5.45 ± 0.03

A' is the excess free energy of system, labeled Sys, in kcal/mol; n_q is the number of quadrature points; K refers to 1000 Monte Carlo steps; SPC, MCY, and TP4 stand for SPC,[18] MCY,[19] and TIP4P[20,21] waters, respectively; results in the first three rows are from ref. 12.

as described above obtained the excess free energy of the SPC[18] and MCY[19] waters as −5.65 ± 0.05 kcal/mol and as −4.04 ± 0.03 kcal/mol, respectively,[12] using five-point Gaussian quadratures. An eight-point quadrature gave −5.70 ± 0.03 kcal/mol for the SPC free energy, essentially confirming the five-point quadrature result. All these calculations used the path of eq. (1) with $k = 4$. Their remarkable precision indicated by the small error estimates was confirmed by the direct calculation (three-point quadrature, $k = 1$) of the difference of the free energies of the SPC and MCY waters as −1.63 ± 0.04 kcal/mol: The sum of the free energy differences over the thermodynamic cycle IG-MCY-SPC-IG gave 0.02 kcal/mol (instead of zero). In the present article, calculations are reported using three-point quadratures, requiring simulations at $\lambda = 0.1125$, 0.5, and 0.8875, both using the path described above and the polynomial path of eq. (10).

RESULTS AND DISCUSSION

Using $k = 4$ for all terms, the excess free energies of the SPC and MCY waters were obtained as −5.61 ± 0.04 and 4.01 ± 0.11 kcal/mol, respectively. Table I gives details of the convergence. Only for the SPC water is the difference between the three- and five-point quadrature results statistically significant.

When trying the polynomial path, several interesting results were obtained. Using $k(1) = 1$, $k(6) = 2$, and $k(12) = 4$ for the SPC water, the simulation at $\lambda = 0.1125$ "got stuck," i.e., the acceptance rate of trial moves progressively decreased to near zero. This was the result of oxygens and hydrogens being allowed to get too close as their contribution to the acceptance decision was not scaled down as much as the repulsions. Using $k(1) = 2$, $k(6) = 2$, and $k(12) = 4$ gave −5.69 ± 0.05 kcal/mol, in excellent agreement with the five- and eight-point quadrature results—actually closer to the eight-point quadrature result. The confidence in the three-point quadrature on the polynomial path is also justified by Figure 1, showing the integrands over both paths: The polynomial path integrand is almost linear, while the integrand of the calculation using the exponent 4 overall has a maximum and its curvature undergoes significant variations. Clearly, a reduced curvature means that the creation of the particle was more evenly spread over the [0, 1] coupling parameter interval.

The calculations on the MCY water proved to be a case where the polynomial path offered no improvement. For the MCY water, the repulsion term is represented by a sum of exponentials and also the electrostatic term is larger than for the SPC water. Calculations with $k = 4$ uniformly on the SPC water showed that the ratio of the repulsive and electrostatic contribution at $\lambda = 0.1125$ is 400 and for the polynomial path it was 90. However, for the $k(1) = k(\exp) = 4$ MCY water calculation this ratio was 1/2. Varying k_1 and $k_{\exp}$ resulted in similarly small ratios. Furthermore, when $k(1) = 3$ and $k(\exp) = 4$ was used (i.e., near the uniform $k = 4$ case), the integrands were obtained as −13, −45, and −15 kcal/mol, yielding a free energy way too low. Other combinations gave even lower values for the integrand at the first quadrature point, indicating an even steeper integrand. However, with an integrand that has such steep variation the extrapolation to the endpoint regions are clearly ill defined. For comparison, the uniform $k = 4$ case gave 1.6, 2.3, and −19.6 kcal/mol, similar to the corresponding SPC integrand shown in Figure 1. Other exponent combinations fared even worse.

Finally, calculations were also performed on the TIP4P[20,21] water that has a 6-12 interaction on the oxygen but the negative charge center is offset by

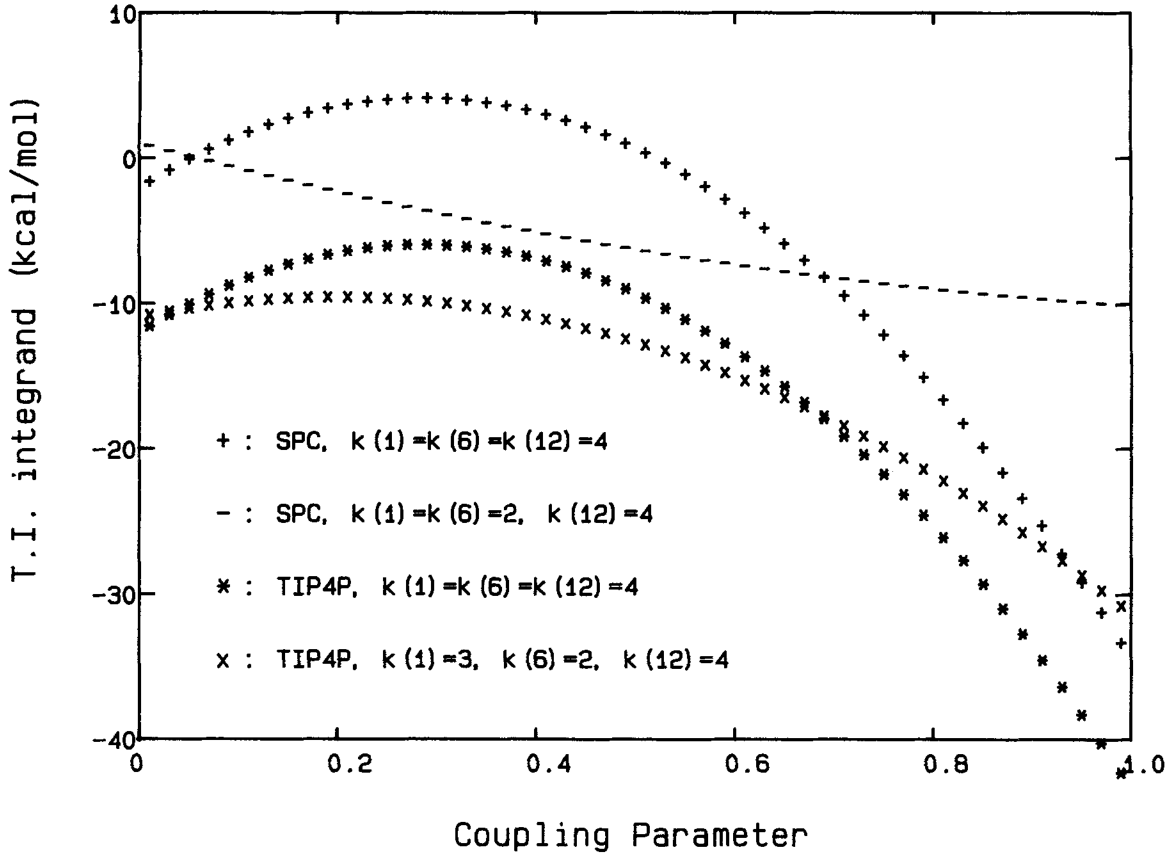

Figure 1. Polynomials fitted to the integrand as generated by the Gaussian quadrature for the calculation of the SPC and TIP4P waters' free energy. TIP4P results are shifted down by 10 kcal/mol.

0.15 Å from the oxygen, and its charge is cca 25% larger than the SPC oxygen. Using $k(1) = k(6) = 2$ and $k(12) = 4$ gave a rather large negative integrand at λ_1 and the acceptance rate was steadily decreasing, indicating an inadequate calculation. Using $k(1) = 3$, $k(6) = 2$, and $k(12) = 4$ gave the excess free energy of the TIP4P water as -5.45 ± 0.05 kcal/mol. A subsequent calculation with $k = 4$ overall gave -5.42 ± 0.06 kcal/mol. Previously, Jorgensen et al.[8] obtained -5.9 ± 0.4 kcal/mol for a system of 216 waters using an 8-Å cutoff and Hermans et al.[7] obtained -5.3 kcal/mol with 80 molecules and a 6-Å cutoff, both under simple cubic periodic boundary conditions. Allowing for the differences in the system size, cutoffs, and boundary conditions, our results are thus supported by the previous work. The ratio of the repulsive and the electrostatic contributions at the smallest λ run was 220 and 400 for the two calculations, respectively. Comparison of the integrands of these two calculations, also displayed in Figure 1, shows that the polynomial path was again able to reduce significantly the curvature of the integrand, making the numerical quadrature more reliable.

Thus, it is concluded that for the polynomial path to be of advantage (in fact, of use) the contribution of the $1/r$ terms at the small λ end must be significantly smaller than the repulsive contribution. A good guide is the minimum ratio of $k(1)/k(\mathrm{rep})/\lambda^{k(\mathrm{rep})-k(1)}$ [see eq. (8)] since this assures that the repulsive term dominates the calculated integrand.

COMPARISON OF LIQUID FREE ENERGY SIMULATION TECHNIQUES

Hermans et al. calculated the excess free energy of several water models, including the SPC and TIP4P waters, using a nonlinear path and a variant of the thermodynamic integration called slow growth.[7] Jorgensen et al. showed that the excess free energy of the liquid can be approached as a problem of solvating a water molecule in liquid water and used the perturbation method to calculate this solvation free energy[8] for the TIP4P water. Also, simulation of water in the grand-canonical ensemble was shown to be feasible at liquid densities using the cavity-biased technique[22,23] with the SPC and ST2[24] models and by the Grand Molecular Dynamics method of Pettit.[25]

Comparison of the thermodynamic integration (even without the polynomial path) and perturbation method results conforms to earlier statements about the relative merits of the thermodynamic integration and the perturbation method[4,13] and shows unequivocally the superiority of the thermodynamic integration for calculating liquid free energies, both in terms of cost and in terms of accuracy: In the calculations performed by Jorgensen et al.[8] for the TIP4P water, even when 10 separate runs (in each directions) were used the statistical uncertainty of the result was estimated to be 0.4 kcal/mol. The SPC and MCY results quoted above were obtained in only five separate calculations (each) using thermodynamic integration, with an uncertainty of about 0.05 kcal/mol. The thermodynamic integration calculations of Hermans were similarly very economical, *vide infra*. The difference between the efficiencies of the two techniques is particularly large for the calculation of pure liquid free energies since the use of the perturbation method is only feasible when a single molecule is mutated into water (as opposed to the whole system when using thermodynamic integration) with a concomitant increase in the statistical error by an approximate factor of $N^{-1/2}$. For solvation free energy calculations, the comparison depends on the extent of the mutation involved.

The comparison between the slow-growth technique and the Gaussian quadrature in general is less clear cut. The calculations of Hermans et al. required a total of 100,000 molecular dynamics time steps, which is about the same computational effort as the separate equilibrations and the 300K long runs using five-point quadratures or 500K long runs and three-point quadratures (adequate when the polynomial path is applicable). The data in Table I shows that for each system studied the free energy difference was well converged after 300K, even though the final results reported are based on 1000K long runs. However, the slow-growth technique, despite its impressive performance on the liquid water problem, is inherently ''dangerous'' since it is relying on ''almost converged'' data. As long-range correlations have been shown to exist in liquid water[15,26] they may manifest themselves in unexpected ways. This reservation is also supported by the recent results of Mitchell and McCammon,[27] who found that to obtain precise results for a dipeptide–tripeptide mutation several hundred ps-long slow-growth runs are required, and by the calculations of Mavri and Hadic, where large hysteresis error was found for the free energy profile of the proton transfer between carboxylic acid and methylamine in water.[28] Furthermore, the customary estimate of the error by calculating the free energy difference along the same path in two directions, used both in refs. 7 and 8, is not a strong enough test: There is an example where it underestimated the error by several kcal/mol,[29] although better ways of estimating the error of the slow-growth result are also appearing.[30,31] This factor, combined with the consistently high precision of our calculations, led us to prefer the Gaussian quadrature technique.

The calculation of the excess free energy of the SPC and ST2 water in the grand-canonical ensemble using the cavity-biased method were based on two 6000K long calculations for both models.[23] As the insertion and deletion attempts increased the computation time by a factor of 2.5 for each iteration, as far as the calculation of the free energy is concerned these calculations are significantly less economical than thermodynamic integration. Calculations with the Grand Molecular Dynamics method proved to be similarly very computing intensive.

The good agreement between the SPC water free energy calculated directly and from the MCY water suggests that the calculation of the free energy of a new model could very economically be obtained, even when the polynomial path offers no improvement, by calculating the free energy difference between the new model and one of the older ones for which the free energy is already known since this path does not involve the creation or annihilation of atoms and a three-point Gaussian quadrature with $k = 1$ overall is adequate.

SUMMARY

The results presented show that the introduction of the polynomial path was able to improve the precision of the calculated liquid free energy for the SPC and TIP4P waters, both a 1-6-12 type of potential, by significantly reducing the curvature of the integrand, especially for the SPC water, where the integrand became nearly linear. However, it was of no use for the MCY water, which has an exponential repulsion that is nonsingular. Thus, the method should only be applied to the truly singular potentials and even there care must be taken in the choice of the exponents: For the run with the smallest coupling parameter value, the acceptance rate is to be monitored carefully and the dominance of the repulsion contribution must be ensured. It is not unreasonable to expect that similar improvements can be obtained for solvation free energy calculation when the solute is described with a singular repulsion term. Also, the results confirm the robustness of the thermodynamic integration based on Gaussian quadratures: For the two water models studied, even without the polynomial path the error of a three-point quadrature was found to be of the order of a tenth of a kcal/mol.

This work was supported by RCMI Grant SRC5G12RR0307 from NIH to Hunter College and a CUNY/PSC grant while the author worked at Hunter College, CUNY. Computing resources were provided by the City University of New York, University Computing Center. Several useful discussions with Prof. K. Bencsath are gratefully acknowledged.

References

1. A. Pohorille and L.R. Pratt. *Meth. Enzymol., Biomembr.*, **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. NY Acad. Sci.*, **482**, 1 (1986).
4. D.L. Beveridge and F.M. DiCapua, *Annu. Rev. Biophys. Chem.*, **18**, 431 (1989).
5. J.G. Kirkwood, in *Theory of Liquids*, B.J. Alder, Ed., Gordon and Breach, New York, 1968.
6. M. Mezei, *Mol. Phys.*, **7**, 307 (1982).
7. J. Hermans, A. Pathiaseril, and A. Anderson, *J. Am. Chem. Soc.*, **110**, 5982 (1988).
8. W.L. Jorgensen, J.F. Blake, and J.K. Buckner, *Chem. Phys.*, **129**, 193 (1989).
9. M.R. Mruzik, F.F. Abraham, D.E. Schreiber, and G.M. Pound, *J. Chem. Phys.*, **64**, 481 (1971).
10. W.C. Swope, H.C. Andersen, P.H. Berens, and K.R. Wilson, *J. Chem. Phys.*, **76**, 637 (1982).
11. M. Mezei, *Mol. Simul.*, **2**, 201 (1989).
12. M. Mezei, *Mol. Phys.*, **67**, 1205 (1989), erratum to ref. 6.
13. A. Cross, *Chem. Phys. Lett.*, **128**, 198 (1986).
14. N.A. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
15. M. Rao, C.S. Pangali, and B.J. Berne, *Mol. Phys.*, **37**, 1779 (1979).
16. R.B. Blackman and J.W. Tuckey, *The Measurement of Power Spectra*, Dover, New York, 1958.
17. W.W. Wood, in *Physics of Simple Liquids*, H.N.V. Temperly, F.S. Rowlinson, and G.S. Rushbrooke, Eds., North-Holland, Amsterdam, 1968.
18. H.J.C. Berendsen, J.M.P. Postma, W.F. van Gunsteren, and J. Hermans, in *Jerusalem Symposia on Quantum Chemistry and Biochemistry*, B. Pullman, Ed., Reidel, Dordrecht, Holland, 1981.
19. O. Matsuoka, E. Clementi, and M. Yoshimine, *J. Chem. Phys.*, **64**, 1351 (1976).
20. W.L. Jorgensen, J. Chandrashekar, J.D. Madura, R. Impey, and M.L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
21. W.L. Jorgensen and J.D. Madura, *Mol. Phys.*, **56**, 1381 (1985).
22. M. Mezei, *Mol. Phys.*, **40**, 901 (1980).
23. M. Mezei, *Mol. Phys.*, **61**, 565 (1987); erratum: **67**, 1207 (1989).
24. F.H. Stillinger and A. Rahman, *J. Chem. Phys.*, **60**, 1545 (1974).
25. J. Ji, T. Cagin, and B.M. Pettitt, *J. Chem. Phys.*, **96**, 1333 (1992).
26. M. Mezei, S. Swaminathan, and D.L. Beveridge, *J. Chem. Phys.*, **71**, 3366 (1979).
27. M.J. Mitchell and J.A. McCammon, *J. Comp. Chem.*, **12**, 271 (1991).
28. J. Mavri and D. Hadic, in *Proton Transfer in Hydrogen-Bonded Systems*, T. Bountis, Ed., Plenum, New York, 1992.
29. B. Jayaram, M. Mezei, and D.L. Beveridge, *J. Am. Chem. Soc.*, **110**, 1691 (1988).
30. R.H. Wood, *J. Phys. Chem.*, **95**, 4838 (1991).
31. J. Hermans, *J. Phys. Chem.*, **95**, 9029 (1991).