Monte Carlo simulation of differences in free energies of hydration

William L. Jorgensen and C. Ravimohan

Citation: The Journal of Chemical Physics **83**, 3050 (1985); doi: 10.1063/1.449208
View online: http://dx.doi.org/10.1063/1.449208
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/83/6?ver=pdfcov
Published by the AIP Publishing

Monte Carlo simulation of differences in free energies of hydration

William L. Jorgensen and C. Ravimohan
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

(Received 21 May 1985; accepted 12 June 1985)

Perturbation theory has been applied to calculate the relative free energies of hydration of methanol and ethane in dilute soluton. It is demonstrated that only two or three Monte Carlo simulations using double-wide sampling are necessary to obtain results with high precision. The small statistical uncertainty in the computed change in free energy of hydration and the good accord with experimental thermodynamic data are most encouraging for application of the procedure to a wide range of problems. Structural effects accompanying the mutation of methanol to ethane in water are also discussed; hydrogen bonding to the solute is essentialy eliminated by only a 25% reduction in the atomic charges of methanol.

I. INTRODUCTION AND BACKGROUND

The ability to calculate relative free energies for different solutions has important applications in chemistry and biochemistry. Relative solubilities could be derived, and comparisons could also be made for series of reactants, transition states, and products. This can yield conformer populations, relative free energies of activation, and relative binding affinities for donor–acceptor complexes. Several statistical mechanical procedures have evolved for computing free energy differences, as reviewed previously.[1–6] Two particularly promising approaches feature umbrella sampling and a perturbation procedure in which one solute is mutated into another.

In umbrella sampling, one is often attempting to calculate a distribution function $g(r)$ that is related to the free energy change or potential of mean force by $w(r) = -kT \ln g(r)$. Use of a biasing function allows sampling a wider range of r than would otherwise be possible. Subsequently, correction is made for the influence of the biasing function and proper averages are obtained. The procedure was developed by Valleau and co-workers[1] and has been applied to many problems including the calculation of conformer populations,[5,7] cavity distributions in a Lennard-Jones liquid,[8] and the temperature dependence of the free energy of water.[9] The method has also been extended through importance sampling to cover even greater ranges of r; multiple simulations are performed for overlapping regions of r and the results are spliced together to obtain the overall $g(r)$. This procedure has been used to obtain potentials of mean force for important systems including two ions in polar solvents,[10,11] two hydrophobic solutes in water,[12–14] internal rotation in biomolecules,[15,16] and S_N2 reactions in water and DMF.[17,18]

The perturbation approach centers on the easily derived relationship

$$A_1 - A_0 = -kT \ln\langle \exp[-\beta(E_1 - E_0)]\rangle_0 . \qquad (1)$$

The equation expresses the free energy difference between systems 0 and 1 by an average of a function of their energy difference that is evaluated by sampling based on E_0. Naturally, the two systems cannot be radically different or the average will be too slowly convergent. This problem can be overcome by performing multiple simulations over intermediate systems between 0 and 1. In this context, it is convenient to define a coupling parameter λ that allows the smooth conversion of system 0 to 1. Then for many possible features ζ of the systems including geometrical and potential function parameters, Eq. (2) can be used to represent the mutation of system 0 to 1 as λ goes from 0 to 1:

$$\zeta(\lambda) = \zeta_0 + \lambda(\zeta_1 - \zeta_0) . \qquad (2)$$

Under these circumstances, the energy E is not typically a simple function of λ.

The perturbation procedure has been applied to several problems including calculation of the free energy of hydration of methane,[19] and the free energy of cavity formation in water.[20] Also, recently Tembe and McCammon have pointed out the potential utility of the method for computing relative binding constants for enzyme–substrate complexes.[21] In model calculations, they obtained good accord between results from the perturbation and umbrella sampling approaches. They also noted some general advantages of the perturbation method for studying donor–acceptor complexes.

A subset of the uses for the perturbation procedure is to compute the free energy change for converting one solute into another in solution. This has numerous important applications as noted above and an intriguing parallelism to mutagenic experiments in biochemistry. In the case of rigid solutes, i.e., having no intramolecular degrees of freedom, relative free energies of solvation are directly obtained. Using the standard pairwise additive intermolecular potential functions, the energy change between two rigid systems, $E_1 - E_0$, is then just the difference between the total solute–solvent interaction energies, $E_1^{sx} - E_0^{sx}$. In order to test the perturbation method for the interconversion of two solutes in a nonidealized system, results are reported here for the mutation of a methanol solute to ethane in dilute aqueous solution. This could also be considered as a model for conversion of a threonine residue to valine in water. The utility of the method is critically dependent on statistical issues such as the computed uncertainty in the free energy change and the number of simulations required to span between the two solutes. Encouraging findings on these points are presented below along with data on the thermodynamic and structural changes that accompany the mutation.

 0021-9606/85/183050-05$02.10

II. COMPUTATIONAL PROCEDURE

Series of Monte Carlo simulations were carried out for the interconversion of methanol and ethane in dilute aqueous solution. A coupling parameter λ was used to smoothly transform methanol ($\lambda = 0$) to ethane ($\lambda = 1$). Simulations were run for $\lambda_i = 0.0$, 0.125, 0.25, 0.50, 0.75, and 1.0. In order to check the self-consistency, the simulations were run in both directions, i.e., $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_{i+1} \rightarrow \lambda_i$ except at the two end points. This is known as "double-ended" sampling.[3] It is facilitated by using what may be termed "double-wide" sampling. Specifically, the free energy differences for $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_i \rightarrow \lambda_{i-1}$ can be obtained simultaneously since both require sampling based on the λ_i system. Operationally, the double-wide sampling was implemented by monitoring three solutes in the computer program corresponding to λ_{i-1}, λ_i, and λ_{i+1}.

The simulations were run in the NPT ensemble at 25 °C and 1 atm using standard procedures including periodic boundary conditions.[22] The systems consisted of the solute plus 125 water molecules in a cube. Metropolis sampling was augmented by preferential sampling in which the probability of attempting to move a solvent molecule was made proportional to $1/(r^2 + c)$ where r is a solute–water oxygen distance. The constant c was fixed at 120 Å^2 which caused the solvent molecules nearest the solute to be moved about twice as often as the most distant water molecules. An attempt to move the solute was made on every 50th configuration and a change in volume was tried on every 600th configuration. The ranges for the attempted moves were the same in each simulation and provided a ~40% acceptance probability for new configurations. Each simulation consisted of an equilibration phase for 0.5×10^6 configurations, followed by averaging for properties over an additional 1.5×10^6 configurations. There was no statistically significant change in the computed free energy differences after averaging for $\sim 0.8 \times 10^6$ configurations. Thus, convergence of this quantity is relatively rapid.

The intermolecular potential functions are described in detail elsewhere.[22–25] The TIP4P model was used for water[23] and the OPLS parameters were adopted for ethane[24] and methanol.[25] The validity of these potential functions has been demonstrated in prior Monte Carlo simulations of pure liquid water, alkanes, and alcohols, and in simulations of dilute aqueous solutions.[22–27] In fact, results have been reported previously from Monte Carlo calculations for both methanol and ethane in TIP4P water.[26,27] The potential functions and conditions for the earlier simulation of aqueous ethane are identical to the present ones; however, the previous model for methanol has the hydrogens on carbon represented explicitly in order to study the internal rotation,[27] while a united atom model is used here in the OPLS description for methyl groups.

The monomers are represented by interaction sites usually located on nuclei. The interaction energy between monomers a and b is then determined by Coulomb and Lennard-Jones interactions between all intermolecular pairs of sites [Eq. (3)]. Standard combining rules are

$$\Delta\epsilon_{ab} = \sum_{i}^{\text{on } a} \sum_{j}^{\text{on } b} (q_i q_j e^2/r_{ij} + A_{ij}/r_{ij}^{12} - C_{ij}/r_{ij}^{6}) \qquad (3)$$

TABLE I. Parameters for the intermolecular potential functions.

Molecule	Group	q	σ (Å)	ϵ (kcal/mol)
CH_3OH	CH_3	0.265	3.775	0.207
CH_3OH	O	−0.700	3.070	0.170
CH_3OH	H	0.435	0.0	0.0
CH_3CH_3	CH_3	0.0	3.775	0.207
H_2O	O	0.0	3.1536	0.155 04
H_2O	H	0.52	0.0	0.0
H_2O	M[a]	−1.04	0.0	0.0

[a] M is a point on the HOH bisector, 0.15 Å from oxygen toward the hydrogens.

used such that $A_{ij} = (A_{ii}A_{jj})^{1/2}$ and $C_{ij} = (C_{ii}C_{jj})^{1/2}$. Furthermore, the A and C parameters may be expressed as $A_{ii} = 4\epsilon_i \sigma_i^{12}$ and $C_{ii} = 4\epsilon_i \sigma_i^6$ where σ and ϵ are the Lennard-Jones radius and energy terms. The parameters for the present study are summarized in Table I. Spherical cutoffs were used to truncate the intermolecular interactions at separations greater than 7.5 Å.

Standard geometries were adopted for the monomers: for ethane, r(CC) = 1.53 Å; for methanol, r(CO) = 1.43 Å, r(OH) = 0.945 Å, ∠ COH = 108.5°; and, for water, r(OH) = 0.9572 Å, ∠ HOH = 104.52°.[23–25] The monomers were kept fixed in these geometries during the simulations so intramolecular vibrational effects have not been considered. The geometries and potential function parameters $(q_i, \sigma_i, \epsilon_i)$ for the intermediate systems were determined from Eq. (2). The atom that is the oxygen of methanol was kept coincident with one of the atoms that becomes a carbon in ethane. The carbon in methanol is then grown into the other carbon of ethane as λ goes from 0 to 1. The bond length gradually increases from 1.43 to 1.53 Å. The bond length to the atom that is the hydroxyl hydrogen in methanol was not scaled. This atom has no Lennard-Jones parameters since it is effectively covered by the adjacent atom and its charge disappears as λ goes to 1. In general, smooth interconversion of the geometries of the two systems, like the opening and closing of a flower, is anticipated to be advantageous for convergence of such calculations.

III. RESULTS AND DISCUSSION

A. Thermodynamics

The computed free energy changes are presented in Table II and Fig. 1. They are Gibbs free energies since the iso-

TABLE II. Free energy change for the interconversion of methanol and ethane in water.

			ΔG (kcal/mol)	
λ_i		λ_j	$i \rightarrow j$	$j \rightarrow i$
0.000		0.125	2.77 ± 0.10	−2.45 ± 0.06
0.125		0.250	1.58 ± 0.07	−1.70 ± 0.08
0.250		0.500	1.44 ± 0.14	−1.69 ± 0.10
0.500		0.750	0.67 ± 0.08	−0.70 ± 0.08
0.750		1.000	0.24 ± 0.05	−0.27 ± 0.05
	Total		6.69 ± 0.21	−6.81 ± 0.17
	Exptl.[a]		6.93	−6.93

[a] Reference 28.

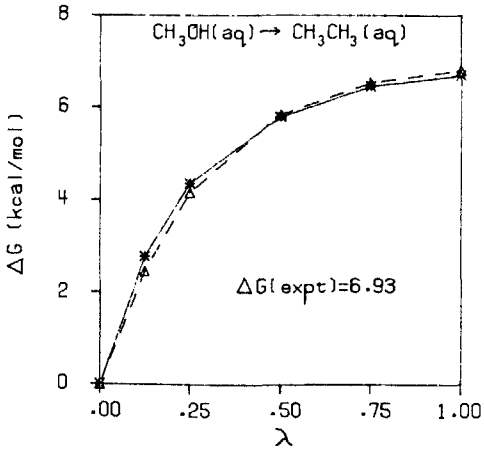

FIG. 1. Computed free energy changes for the interconversion of methanol ($\lambda = 0$) and ethane ($\lambda = 1$) in water. The solid and dashed lines correspond to the forward ($\lambda = 0 \rightarrow 1$) and reverse ($\lambda = 1 \rightarrow 0$) processes, respectively.

thermal–isobaric ensemble has been used. The statistical uncertainties in the tables are $\pm 1\sigma$ and were obtained from separate averages over blocks of 5×10^4 configurations.

Several key observations may be made. (1) The overall ΔG's obtained by proceeding from methanol to ethane (6.69 kcal/mol) or vice versa (6.81 kcal/mol) are in good accord and within the estimated uncertainties of ± 0.2 kcal/mol. (2) These values are also in close agreement with the experimental difference in the free energies of hydration of methanol and ethane from the gas phase (6.93 kcal/mol).[28] This provides further verification of the utility of the OPLS potential functions. (3) The incremental free energy changes compare well and are predominantly within the error bars except for the $0.0 \leftrightarrow 0.125$ increment. This step shows by far the largest change in free energy, roughly 40% of the total. The results suggest that the perturbation becomes too large with a concomitant reduction in the precision of the free energy calculation for differences greater than roughly 3–4 kT. However, with double-wide sampling, a range of as much as ~ 6–$8\ kT$ could be covered in one simulation. (4) There is no constant pattern in the incremental free energies for $i \rightarrow j$ or $j \rightarrow i$. That is, the magnitudes of the computed differences are larger for $j \rightarrow i$ for the last four steps, but not for the first. This issue is related to the Gibbs–Bogoliubov inequality which actually suggests that, if there is any bias, it should be for the magnitude of the $i \rightarrow j$ differences to be larger than for $j \rightarrow i$ in Table II.[3] Basically, for a system 1 at higher free energy than system O, $\langle E_1 - E_0 \rangle_0$ will be increasingly overestimated as the difference between the systems increases. Along these lines, a simulation was also run for $\lambda = 0.0 \rightarrow 0.25$. The computed ΔG is 4.71 ± 0.15 kcal/mol which is substantially larger than the sum for the first two steps in Table II. The trend is consistent with the Gibbs–Bogoliubov pattern which becomes a greater effect as the perturbation increases.

Some additional energetic results are given in Table III. The average, total solute–solvent interaction energies are recorded. The sampling was based on the solute corresponding to λ_i and directly yielded E_i^{sx}. The solute–solvent energies for the perturbed solutes λ_{i-1} and λ_{i+1} were also averaged and provided E_{i-1}^{sx} and E_{i+1}^{sx}. Some points to note from these data are as follows. First, the E^{sx} values become progressively less exothermic as methanol ($\lambda = 0$) goes to ethane ($\lambda = 1$) due to the loss of solute–solvent hydrogen bonding. The statistical uncertainties in the solute–solvent energies are modest; the fluctuations in $\ln\langle \exp[-\beta(E_{i+1}^{sx} - E_i^{sx})] \rangle_i$ are apparently even smaller and account for the small uncertainties for the computed ΔG's. Of course, the total energies are much greater in magnitude (~ -1260 kcal/mol) due to the solvent–solvent contributions and have much larger uncertainties ($\sim \pm 3$ kcal/mol).

TABLE III. Computed solute–solvent energies (kcal/mol) for the interconversion of methanol and ethane in water.[a]

λ_{i-1}	λ_i	λ_{i+1}	E_{i-1}^{sx}	E_i^{sx}	E_{i+1}^{sx}
	0.000	0.125		-19.8 ± 0.4	-16.4 ± 0.3
0.000	0.125	0.250	-14.9 ± 0.3	-12.7 ± 0.3	-10.5 ± 0.2
0.125	0.250	0.500	-11.6 ± 0.3	-9.8 ± 0.3	-6.8 ± 0.1
0.250	0.500	0.750	-7.4 ± 0.2	-6.0 ± 0.1	-4.6 ± 0.1
0.500	0.750	1.000	-5.3 ± 0.1	-5.0 ± 0.1	-4.4 ± 0.1
0.750	1.000		-4.6 ± 0.1	-4.6 ± 0.1	

[a] The sampling is based on the λ_i system.

It should also be noted that there are seeming inconsistencies in Table III. For example, from the first row, $E_{0.125}^{sx}$ is -16.4, but from the second row it is -12.7. This occurs because the first value is obtained from configurations based on sampling with $\lambda = 0.0$, while the second is from sampling with $\lambda = 0.125$. With $\lambda = 0.0$ there is greater solute–solvent structure and order which yields a lower $E_{0.125}^{sx}$, an effect similar to what would be obtained by running the $\lambda = 0.125$ simulation at a lower temperature.

A final piece of thermodynamic data that can be obtained from the results is the heat of hydration of the solutes from the ideal gas.[22,26] The total energies from the simulations of aqueous methanol and ethane are -1267 ± 3 and -1250 ± 2 kcal/mol, while the energy for pure TIP4P water at 25 °C and 1 atm with $N = 125$ is -1256 ± 2. This yields predicted heats of hydration of -12 ± 4 and 5 ± 3 kcal/mol for methanol and ethane which may be compared with experimental values of -10.8 and -4.7.[29,30] The greater precision and accuracy for the free energy differences is encouraging and follows from the avoidance of taking differences in the total energies.

B. Energy distributions

Additional details on the solute–solvent energetics are revealed in distributions for the total solute–solvent interaction energies (Fig. 2) and individual solute–solvent interaction energies (Fig. 3). The averages of the distributions in Fig. 2 are the E^{sx} values for λ_i in Table III. The solutes experience a continuum of energetic environments which becomes narrower as the hydrogen bonding is removed. There is substantial overlap of the adjacent distributions. This seems desirable for applying the perturbation relation in Eq. (1), though the issue is actually more complicated. What is critical is that the configurations that are sampled for λ_i not be uniformly unreasonable for λ_{i-1} and λ_{i+1}.

The change in hydrogen bonding is well revealed in Fig. 3. For $\lambda = 0.0$, the classic bimodal shape is obtained with the

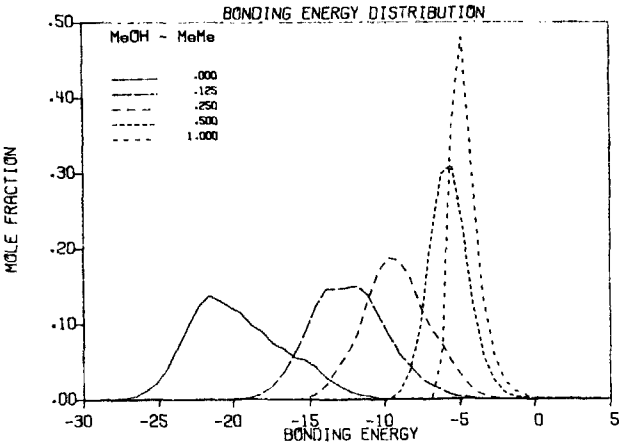

FIG. 2. Computed distributions for the total solute–water interaction energies for different values of the coupling parameter λ. The units for the ordinate are mole fraction of solute per kcal/mol.

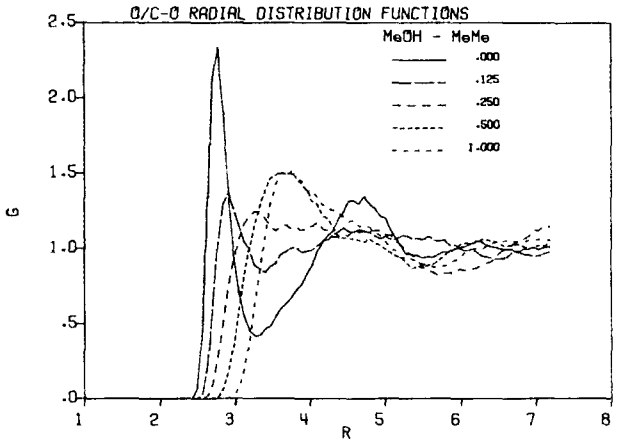

FIG. 4. Computed radial distribution functions between oxygens of water and the atom that is converted from oxygen in methanol to carbon in ethane. Distances are in Å throughout.

methanol–water hydrogen bonds represented by the low energy band up to about -2.7 kcal/mol. Integration of the band yields an average of 2.6 methanol–water hydrogen bonds. The large spike centered at 0.0 kcal/mol is due to the many relatively weak interactions with more distant water molecules. Figure 3 shows that the solute–solvent hydrogen bonding is no longer a distinct entity when $\lambda = 0.25$ is reached. This corresponds to a reduction of the atomic charges for methanol by only 25%.

C. Radial distribution functions

Structural results for aqueous methanol and ethane have been described previously,[26,27] so trends as a function of λ will be the focus here.

Figure 4 contains the radial distribution functions (rdfs) between the oxygen of water and the atom that is transformed from oxygen in methanol to carbon in ethane. The methanol–water hydrogen bonding is reflected in the high, narrow first peak for $\lambda = 0$. Integration reveals roughly three water molecules within the range of the peak to 3.3 Å. The second band centered near 4.5 Å can be assigned to water molecules in the second shell around the hydroxyl group and to those in the first shell around the methyl group.

As λ increases, the hydrogen bonding features are rapidly replaced by the broad peaks at greater separation characteristic of hydrophobic solutes. The distinct structure is diminished and the number of water molecules in the first solvent shell increases to about 23 for ethane.[26] The same patterns are found in the other rdfs associated with the hydroxyl end of methanol. For example, the hydroxyl hydrogen–water oxygen rdfs are shown in Fig. 5. Again, the hydrogen bonding is essentially eliminated when λ reaches 0.25.

The final rdfs shown in Fig. 6 are between the oxygen of water and the carbon of methanol that becomes a carbon in ethane. The changes are much less, but the qualitative trend is the same. The first peak is flattened and shifted to larger separation. As the solute becomes more hydrophobic, the water structure near the solute becomes more amorphous. It should be realized that the changes in the first peak for the

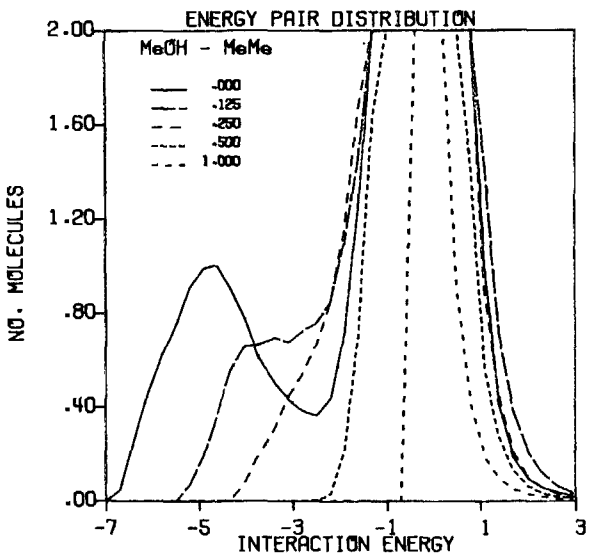

FIG. 3. Computed distributions for the individual solute–water interaction energies (kcal/mol). The units for the ordinate are the number of water molecules per kcal/mol. Results are shown for five values of the coupling parameter λ.

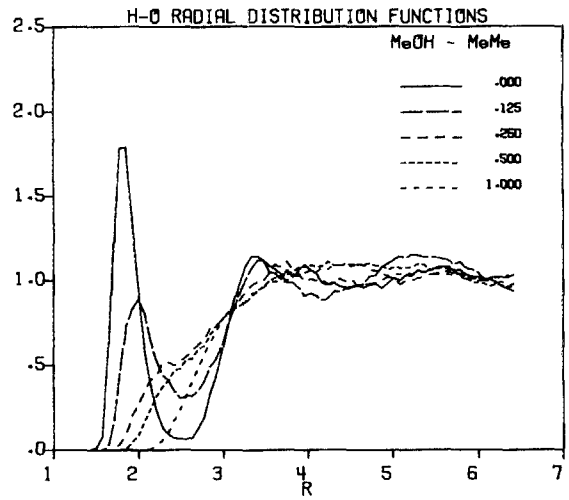

FIG. 5. Computed radial distribution functions between oxygens of water and the atom that is the hydroxyl hydrogen in methanol.

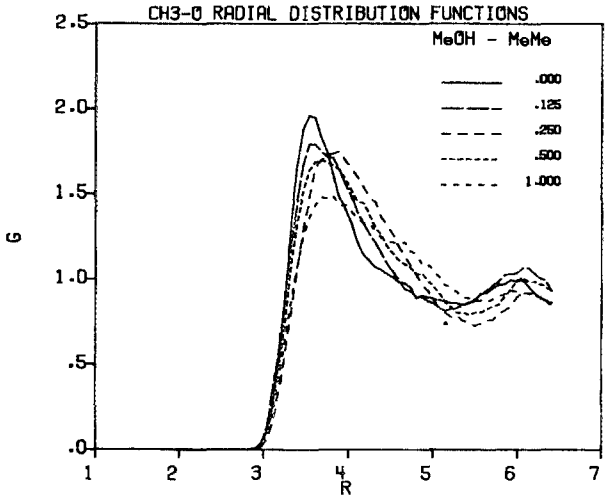

FIG. 6. Computed radial distribution functions between oxygens of water and the atom that is transformed from the carbon of methanol to a carbon in ethane.

C–O rdfs are not just due to water molecules nearest the methyl groups. The water molecules hydrogen bonding to the hydroxyl group for $\lambda = 0$ also contribute to the sharpness of the first peak in Fig. 6.

IV. CONCLUSION

Methodology for performing perturbation theory calculations to compute relative free energies for solutions has been investigated for the interconversion of methanol and ethane in water. The results are most encouraging for further chemical and biochemical applications. In particular, the computed uncertainty for this significant mutation is only 0.2 kcal/mol or 3% of the total free energy change. The results were also shown to be independent of the direction of the mutation and in good accord with the experimental difference in free energies of hydration for methanol and ethane. Furthermore, with double-wide sampling only two or three Monte Carlo simulations would be necessary to obtain an estimate of the overall free energy change with good precision. However, additional simulations at the two end points may generally be desirable to obtain structural and thermodynamic data on these key systems. In the present case, it was also found that the dominant changes occur for small perturbations near the solute with largest dipole moment. This is probably a general trend and should be a useful observation for selecting values of the coupling parameter for future work.

ACKNOWLEDGMENTS

Gratitude is expressed to the National Science Foundation and National Institutes of Health for support of this research.

[1]J. P. Valleau and G. M. Torrie, in *Statistical Mechanics, Part A*, edited by B. J. Berne (Plenum, New York, 1977), p. 169.
[2]K. S. Shing and K. E. Gubbins, Adv. Chem. **204**, 73 (1983).
[3]C. H. Bennett, J. Comp. Phys. **22**, 245 (1976).
[4]N. Quirke and G. Jacucci, Mol. Phys. **45**, 823 (1982).
[5]For a review, see: W. L. Jorgensen, J. Phys. Chem. **87**, 5304 (1983).
[6]M. Mezei and D. L. Beveridge (to be published).
[7]D. W. Rebertus, B. J. Berne, and D. Chandler, J. Chem. Phys. **70**, 3395 (1979).
[8]J. C. Owicki and H. A. Scheraga, J. Phys. Chem. **82**, 1257 (1978).
[9]F. Sussman, J. M. Goodfellow, P. Barnes, and J. L. Finney, Chem. Phys. Lett. **113**, 372 (1985).
[10]G. N. Patey and J. P. Valleau, J. Chem. Phys. **63**, 2334 (1975).
[11]M. Berkowitz, O. A. Karim, J. A. McCammon, and P. J. Rossky, Chem. Phys. Lett. **105**, 577 (1984).
[12]C. Pangali, M. Rao, and B. J. Berne, J. Chem. Phys. **71**, 2975 (1979).
[13]G. Ravishanker, M. Mezei, and D. L. Beveridge, Faraday Symp. Chem. Soc. **17**, 79 (1982).
[14]G. Ravishanker and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2565 (1985).
[15]S. H. Northrup, M. R. Pear, C.-Y. Lee, J. A. McCammon, and M. Karplus, Proc. Natl. Acad. Sci. U. S. A. **79**, 4035 (1982).
[16]M. Mezei, P. K. Mehrotra, and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2239 (1985).
[17]J. Chandrasekhar, S. F. Smith, and W. L. Jogensen, J. Am. Chem. Soc. **107**, 154 (1985).
[18]J. Chandrasekhar and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 2974 (1985).
[19]S. Okazaki, K. Nakanishi, H. Touhara, and Y. Adachi, J. Chem. Phys. **71**, 2421 (1979).
[20]J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, Faraday Symp. Chem. Soc. **17**, 55 (1982).
[21]B. L. Tembe and J. A. McCammon, Comput. Chem. **8**, 281 (1984).
[22]W. L. Jorgensen and C. J. Swenson, J. Am. Chem. Soc. **107**, 1489 (1985).
[23]W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[24]W. L. Jorgensen, J. D. Madura and C. J. Swenson, J. Am. Chem. Soc. **106**, 6638 (1984).
[25]W. L. Jorgensen (in preparation).
[26]W. L. Jorgensen, J. Gao, and C. Ravimohan, J. Phys. Chem. (in press).
[27]W L. Jorgensen and J. D. Madura, J. Am. Chem. Soc. **105**, 1407 (1983).
[28]A. Ben-Naim and Y. Marcus, J. Chem. Phys. **81**, 2016 (1984).
[29]D. M. Alexander and D. J. T. Hill, Aust. J. Chem. **22**, 347 (1969); C. Jolicoeur and G. Lacroix, Can. J. Chem. **54**, 624 (1976).
[30]S. F. Dec and S. J. Gill, J. Solution Chem. **13**, 27 (1984).