THE JOURNAL OF PHYSICAL CHEMISTRY A

 VOLUME 106, NUMBER 29, JULY 25, 2002

LETTERS

Rapid Calculation of the Structures of Solutions with ab Initio Interaction Potentials

Robert H. Wood,* Wenbin Liu, and Douglas J. Doren

Department of Chemistry and Biochemistry, and Center for Molecular and Engineering Thermodynamics, University of Delaware, Newark, Delaware 19716

Received: February 7, 2002; In Final Form: April 16, 2002

A non-Boltzmann sampling method is demonstrated for efficient calculation of structural properties of models having ab initio interaction potentials. A sample of independent configurations is generated with a molecular dynamics simulation using an approximate potential. Ab initio calculations, done only at the sampled configurations, are used to correct the thermal averages of any mechanical variable for the difference between the approximate and ab initio potentials. Tests of the method show that relatively few ab initio calculations are needed to obtain significant structural information and that bootstrap estimates of the uncertainties are reasonable. As a further test, we use results from a previous calculation of the hydration free energy of Na^+ at 973 K and 0.535 g/cm^3 to calculate the Na−O pair correlation function and coordination number for an ab initio model of Na^+-H_2O interactions.

1. Method

We have recently shown that free energies can be calculated with potential energies from first-principles electronic structure methods by a using the Free Energy Perturbation (FEP) approach.[1−3] Simulations with an approximate potential, V_0 (e.g., a Lennard-Jones plus charge model), are used to generate an ensemble of thermally accessible configurations. Calculations of the ab initio interaction potential, V_1, are needed only for a sample of statistically independent configurations taken from the simulation with V_0. In practice, this approach allows calculations of free energies that are several orders of magnitude faster than full ab initio simulations. Here we show that, in addition to calculating free energy by FEP, we can use non-Boltzmann weighting of the sample configurations to calculate the expectation value of any mechanical variable $M[p,q]$ (that is, a function of momenta, p, and coordinates, q) using the same first principles calculations needed for the free energy. As an example of the utility of the proposed method, we demonstrate that reasonably accurate coordination numbers and pair correlation functions of a solute can be obtained from accurate ab initio calculations of the energies of 100−200 configurations.

The requisite equations for calculating any mechanical variable are easily derived. Given any mechanical variable $M[p,q]$ (a function of coordinates and momenta only) and a Hamiltonian $H_1[p,q] = V_1[q] + T[p]$, the canonical ensemble average of M is

$$\langle M\rangle_1 = \frac{\int M e^{-\beta H_1}\, dp\, dq}{\int e^{-\beta H_1}\, dp\, dq} \tag{1}$$

where $\langle\ \rangle_1$ indicates an ensemble average from a simulation with Hamiltonian H_1. Consider an approximate Hamiltonian with the same masses as H_1, but a different potential $H_0[p,q] = V_0[q] + T[p]$. Using $\Delta V = H_1 - H_0$, eq 1 becomes

$$\langle M\rangle_1 = \frac{\int M e^{-\beta \Delta V} e^{-\beta H_0}\, dp\, dq}{\int e^{-\beta \Delta V} e^{-\beta H_0}\, dp\, dq} \tag{2}$$

10.1021/jp020361w CCC: $22.00
Published on Web 06/28/2002

Multiplying the numerator and denominator of eq 2 by $\int e^{-\beta H_0}$ dp dq and reducing to ensemble averages gives

$$\langle M \rangle_1 = \frac{\langle M e^{-\beta \Delta V} \rangle_0}{\langle e^{-\beta \Delta V} \rangle_0} \tag{3}$$

where $\langle\ \rangle_0$ indicates an ensemble average from a simulation with H_0 as the Hamiltonian. Similar equations hold for the *NPT* ensemble.

This derivation simply uses non-Boltzmann weighting[4,5] to calculate expectation values of a mechanical variable from simulations with an easily calculated model, V_0. The free energy perturbation method, used previously to calculate free energy, uses a similar approach to efficiently sample the phase space of two different Hamiltonians in order to calculate the corresponding difference in free energy, ΔA (or ΔG with the *NPT* ensemble).[4−6] The two methods are general and can be used to calculate any free energy by FEP and any M by non-Boltzmann weighting for any two potentials.

Here we use the technique to sample an ab initio potential V_1, using a potential, V_0, that is much more easily computed. This avoids ab initio simulations while still calculating the system properties with ab initio interactions. The saving in computer resources can be very large since the ab initio calculation is required only at a set of independent configurations from the simulation, not at every time step. The method is exact if all of configuration space is sampled. Previously, we have found that ab initio calculations at 50−100 configurations were sufficient to calculate the free energy of hydration of a solute X in solvent S if a reasonably accurate approximate Hamiltonian was used. This is possible because, for instance, if the approximate model has a free energy of hydration within 10% of the ab initio free energy and the FEP correction term is calculated with 10% uncertainty, then the resulting $\Delta_h G$ is accurate to 1%.

2. Tests of the Method

The derivation of this method shows that it is exact for an infinite number of independent configurations. The crucial questions are (1) how many configurations are needed to get significant structural information and (2) can we accurately estimate the uncertainty of our results? To investigate these questions, we have taken two different approximate models of Na^+-H_2O interactions and used the present method to calculate the pair correlation function (at 973 K and 0.535 g/cm^3) of the second model from a simulation of the first model. Since no ab initio calculations are involved, the calculations are relatively easy. The first approximate potential, V^{LJ}, is based on the Lennard-Jones plus charge model[7] for Na^+-water interactions [$\epsilon_{NaO} = 0.698\ 19$ kJ·mol^{-1}, $\sigma_{NaO} = 2.6865$ Å, and $q_{Na} = 1$], but with the charges on O and H determined by the polarizable TIP4P−FQ model for H_2O.[8] The second (more accurate) model was derived by Liu et al.[3] from a least-squares fit to the ab initio energies of 30 configurations of the V^{LJ} model. The potential for this "6−4" model is

$$V^{6-4} = \frac{A_{Na,j}}{R_{Na,j}^{\ 6}} - \frac{B_{Na,j}}{R_{Na,j}^{\ 4}} + \frac{q_{Na} q_j}{R_{Na,j}} \tag{4}$$

where $A_{NaO} = 7341.6$ kJ·mol^{-1}Å^6, $B_{NaO} = 720.0$ kJ·mol^{-1}Å^6, $A_{NaH} = 1627.2$ kJ·mol^{-1}Å^4, and $B_{NaH} = 258.8$ kJ·mol^{-1}Å^4. Liu et al.[3] found that the exponents 6 and 4 gave the best fit of any two-exponent model. These models have quite different short-range potentials and quite different pair correlation functions, so they provide a good test of the method.

To test the convergence of the method, we sampled independent configurations from a simulation of V^{LJ} and calculated the pair correlation function of the 6−4 model, g^{6-4}, for several choices of the number of configurations, N_{conf}. We first calculated $\Delta g^{LJ \to 6-4} = g^{6-4} - g^{LJ}$ by averaging over the radial distributions in the N_{conf} configurations sampled from the simulation with V^{LJ},

$$\Delta g_{NaO}^{LJ \to 6-4}(r;N_{conf}) = \frac{\langle g_{NaO} e^{-\beta \Delta V^{LJ \to 6-4}} \rangle_{LJ}}{\langle e^{-\beta \Delta V^{LJ \to 6-4}} \rangle_{LJ}} - \langle g_{NaO} \rangle_{LJ}$$

$$= \sum_{i=1}^{N_{conf}} g_{NaO,i}(w_i - 1) \tag{5}$$

Here, $g_{NaO,i}(r)$ is the discrete distribution of Na−O distances in configuration i, and $w_i = N_{conf}\ e^{-\beta \Delta V_i^{LJ \to 6-4}} / \sum_{j=1}^{N_{conf}} e^{-\beta \Delta V_j^{LJ \to 6-4}}$ is the configuration weight in the average over the V^{LJ} ensemble (eq 3). The radial distribution function was calculated using bins of width 0.074 Å; the final result was smoothed with a fourth-order, nine-point Savitzky−Golay filter.[9] Finally, g^{6-4} is calculated as

$$g^{6-4}(r;N_{conf}) = g^{LJ}(r) + \Delta g^{LJ \to 6-4}(r;N_{conf}) \tag{6}$$

with g^{LJ} taken from all the configurations of a long simulation of V^{LJ}, so that it has very little uncertainty. By using the small sample of N_{conf} configurations only to calculate the relatively small contribution of $\Delta g^{LJ \to 6-4}$, and a large sample to calculate the much larger contribution of g^{LJ}, the uncertainty in our final result is much smaller than if N_{conf} configurations had been used to calculate g^{6-4} directly from eq 3.

Figure 1 compares calculations of $g^{6-4}(r;N_{conf})$ with $N_{conf} =$ 50, 100, 200, and 500 (with and without smoothing) to a direct calculation of g^{6-4} from a long simulation with V^{6-4}. Figure 1 also shows the standard deviation at selected points, as calculated by bootstrap Monte Carlo resampling of the configurations.[10] Significant structural information is already obtained with only 50 configurations. When smoothed, the peak in $g^{6-4}(r;50)$ is 70 ± 10% as high as the peak in $g^{LJ}(r)$, whereas the actual peak in $g^{6-4}(r)$ is 62% as high. With 500 configurations, $g^{6-4}(r;500)$ has small error bars and is in good agreement with the accurate, direct calculation. These results also show that the Bootstrap method gives a reliable estimate of uncertainties, which will be essential when we do not have an accurate pair correlation function to compare with, i.e., when we use this method with ab initio energies.

It should be noted that for the first two nonzero bins at 1.961 and 2.035 Å, the bootstrap estimates of the uncertainties are not close to the actual error. This reflects the low frequency of finding molecules at these distances in the simulation of the approximate model (Figure 1). There are no waters in the 1.961 Å bin in the first 200 configurations and only 1 in the first 500. Similarly, there is only one water in the 2.035 Å bin in the first 200 configurations and only 7 in the first 500 configurations. The accuracy of these points could be improved by altering the potential of the approximate model to increase the pair correlation function in this region.

The smoothing reduces the noise in the curves, and is especially helpful with small samples. Figure 1b shows that smoothing apparently does not distort g^{6-4}. Bootstrap uncertainties are 20−30% lower for the smoothed g than for the

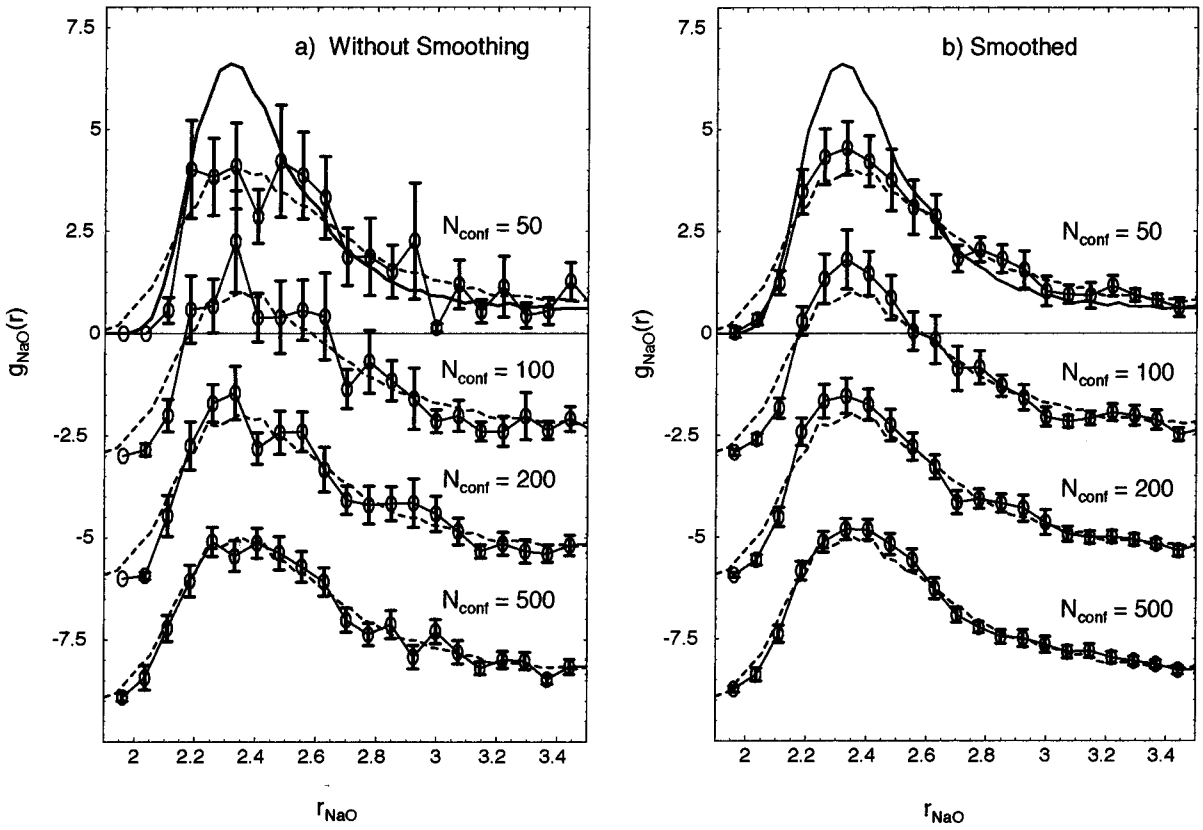

Figure 1. Calculations of $g_{NaO}(r)$ for various models of Na^+-H_2O at 973 K with 0.535 g/cm^3: (—) g^{LJ} from simulation of the LJ model; (- - -) g^{6-4} from simulation of the 6−4 model; (—Φ—) $g^{6-4}(r;N_{conf})$ calculated from eqs 5 and 6 using N_{conf} independent configurations of a simulation of the LJ model. The graphs are displaced by multiples of 3 for clarity. From top to bottom the curves correspond to N_{conf} = 50, 100, 200, and 500. Error bars are bootstrap estimates. (a) $g^{6-4}(r;N_{conf})$ shown without smoothing and (b) $g^{6-4}(r;N_{conf})$ smoothed with a Savitzky−Golay (4,9) filter.

unsmoothed results, giving a quantitative estimate of the reduction in error due to smoothing.

From the pair correlation function we can get the coordination number N_c of the Na^+ ion by integrating $g(r)$ out to the minimum in the pair correlation function (in this case we chose r_c = 3.4595 Å as the minimum),

$$N_c = \rho_B \int_0^{r_c} g(r) 4\pi r^2 \, dr \tag{7}$$

Here, ρ_B is the number density of water far from the solute. The resulting coordination numbers with bootstrap error estimates for N_{conf} = 50, 100, 200, and 500 are $N_c = 4.73 \pm 0.17$, 4.74 ± 0.11, 4.80 ± 0.09, and 4.79 ± 0.07, respectively. These values are significantly lower than the coordination number of V^{LJ} ($N_c^{LJ} = 5.0$), and they converge very nicely to the accurate value of $N_c^{6-4} = 4.77$ calculated directly from a long simulation of V^{6-4}.

We next demonstrate that significant structural information about an ab initio model can be obtained from the same calculations used to calculate free energies for that model. Specifically, for the case of a Na^+ ion in water at 973 K and 0.55 g/cm^3 density, we calculate the Na−O pair correlation function, g_{NaO}. We have previously calculated the change in free energy of the Na^+ ion in aqueous solution[3] when changing the solute−solvent interactions from the V^{6-4} model discussed above to an ab initio potential, V^Q. A simulation was done with this approximate potential, followed by ab initio calculations at 90 configurations, to calculate the free energy difference

$$\Delta G(V^{6-4} \rightarrow V^Q) = -k_B T \ln\langle e^{-\beta \Delta V^{6-4 \rightarrow Q}} \rangle \tag{8}$$

The details of the ab initio calculations have been reported previously and will not be repeated here.[1-3] Briefly, we have approximated the solute−solvent interaction energy as a sum of pairwise and multibody interaction energies:

$$V_{NaW} = V_{Na,W,n}[\text{pair}] + V_{Na,W,m}[\text{multi}] \tag{9}$$

Here $V_{Na,W,n}$[pair] is the sum of pair interaction energies between the solute (Na^+) and the nearest n solvent molecules (W_i), calculated at the MP2/6-311++G(3df,3pd) level of theory. The multibody interaction energy is the difference between the pairwise and total interaction energy of a small cluster with m solvent molecules, $V_{Na,W,m}$[multi] = $V_{Na,W,m}$ − $V_{Na,W,m}$[pair], calculated at the B3LYP/6-311++G (3df,3pd) level. Note that we are using the ab initio calculations only to perturb the ion−water interaction energy and rely on the TIP4P-FQ model to represent the water−water interactions. Thus we are calculating the structure of a QM/MM model with an ab initio model for the interaction of the Na^+ ion with the closest n water molecules. The result of this calculation is that $\Delta G(V^{6-4} \rightarrow V^Q) = -6.2 \pm 1.3$ kJ·mol^{-1}.[3]

Because our ab initio calculations perturb only the solute−solvent interactions, the resulting Hamiltonian has ab initio V_{NaO} and V_{NaH} for the closest waters, but more approximate water−water interactions. Thus, it is only appropriate to calculate solute−solvent structural properties, such as the pair correlation function g_{NaO} and coordination number N_c. We assume that the approximate water−water potential will have only second-order effects on the solute−solvent structure.

Figure 2 shows predictions of g_{NaO} based on simulations with the approximate V^{6-4} potential. The corrections to g_{NaO}^{6-4} from

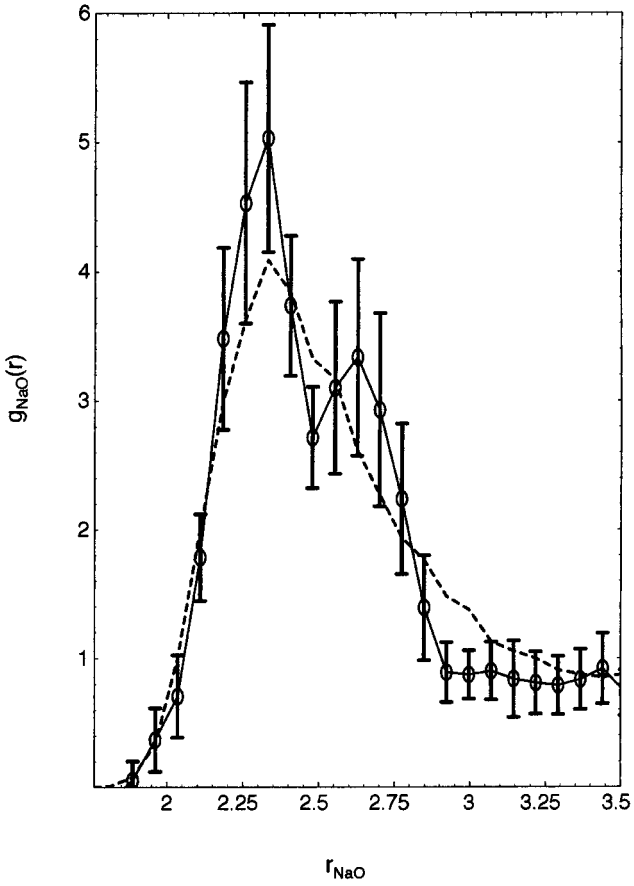

Figure 2. $g_{NaO}(r)$ for the ab initio model of Na^+-H_2O at 973 K with 0.535 g/cm^3: (- - -) $g^{6-4}(r)$ from simulation of the 6−4 model; (—Φ—) $g^Q(r;N_{conf})$ for $N_{conf} = 90$ independent configurations from a simulation of the 6−4 model. $g_{NaO}(r;N_{conf})$ has been smoothed with a Savitzky−Golay (4,9) filter. Error bars are bootstrap estimates.

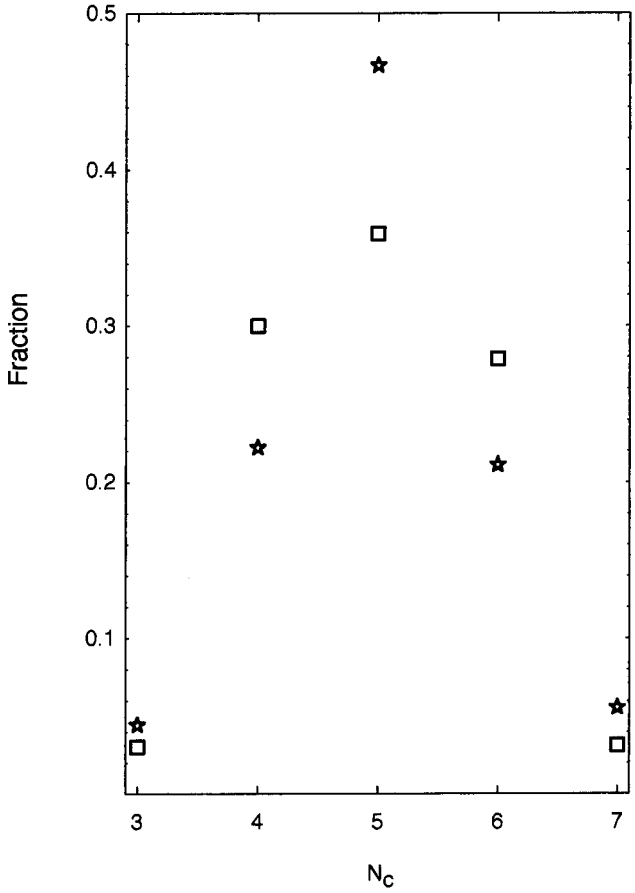

Figure 3. Fractions of configurations with various coordination numbers found in a sample of 90 independent configurations from a simulation of the 6−4 model. (☆) Fractions for 6−4 mode; (□) Fractions calculated for the ab initio model by weighting the 90 configurations from the 6−4 model.

ab initio calculations at 90 configurations are fairly small and comparable in magnitude at most points to the uncertainty. Thus, g_{NaO} for the ab initio model is not very different from $g_{NaO}{}^{6-4}$, although it probably has a slightly higher and broader first peak. The predicted N_c (with $r_c = 3.4595$ Å) is 4.98 ± 0.19, indicating that 90 configurations are adequate to get coordination numbers with reasonably small uncertainties. Since ab initio calculations were already done for these configurations in obtaining the free energy, the additional computational expense to obtain g_{NaO} is negligible.

If our simulation of the approximate model only included structures with a narrow range of coordination number, it is conceivable that structures with different coordination numbers could have much lower energies in the ab initio model, yet be missed by the present calculation. An examination shows that coordination numbers from 3 to 7 are sampled by our 90 configurations. Figure 3 shows the fraction of configurations with a given coordination number predicted for the V^{6-4} and V^Q models by these 90 configurations. These distributions are similar enough to indicate that the range of coordination numbers of V^Q is adequately sampled. The differences in the distributions are nevertheless physically significant, with the distribution for V^Q being less sharply peaked than that for V^{6-4}. This difference in the distributions has little effect on the average coordination number: $N_c{}^{6-4}$ is 4.77, while N_c^Q is 4.98 ± 0.19.

3. Discussion

This method is not limited to simple closed-form approximate potential models for the simulation. In principle, semiempirical quantum mechanical models or molecular mechanics/quantum mechanics hybrids could be used. We expect the present approach to be more accurate than a full ab initio simulation with the same computational resources because higher level ab initio calculations can be applied when a smaller number of configurations are calculated. For example, a recent density functional theory (DFT) simulation of 64 molecules of water for 10 ps required calculation of 69 000 electronic energies.[11] The structure predictions of this model are limited by the accuracy of DFT, which tends to underestimate dispersive forces. These structural predictions could easily be tested and improved by calculations on a few hundred independent configurations from this simulation using a more accurate electronic structure method. Alternatively, the same structural information obtained from the DFT simulation could be obtained at much lower cost from a simulation with an empirical potential, followed by application of our method with DFT energies at a few hundred configurations.

The technique presented here will allow accurate and computationally efficient answers to many questions. For instance, this method should be able to help resolve some of the continuing discrepancies between water structures obtained from neutron diffraction[12] and from simulations of potential models.[13,14] However, there are some significant approximations and limitations to the technique. We believe that the largest error in the present calculations is caused by the approximations in the quantum calculation. As implemented here, the solvent−solvent interactions are calculated with an approximate potential so that we cannot calculate accurate values of solvent−solvent

structure and the approximate solvent potential will cause indirect errors in the solute−solvent properties. We can, however, calculate the structures and free energy of pure water by considering one water as solute and perturbing its interactions with all other waters. Quantities such as velocity autocorrelation functions that involve p or q at different times cannot be calculated by this method.

The examples described here show that a reasonably accurate approximate potential and a modest amount of computation yields accurate values of N_c and $g(r)$ for a high level ab initio model. We have not yet investigated the error in $g(r)$ introduced by the approximations in our ab initio model. While there is an extensive literature on the accuracy of the energy calculated by various quantum methods, we have virtually no experience with the corresponding accuracy of $g(r)$. The present method will allow exploration of this question.

Acknowledgment. We are indebted to Eric Yezdimer for very helpful comments and for the suggestion of testing our method with analytical models. This research was supported by the National Science Foundation under Grant No. CHE9725163 and by the Department of Energy under Grant No. DEFG02-89ER-14080. We also acknowledge the computer facilities of the University of Delaware including facilities made available by the National Science Foundation Major Research Instumentation Program (Grant No. CTS-9724404) and the SGI Origin 2000 and Exemplar SPP2000 at the National Computational Science Alliance under Grant Nos. pzf and CHE-010029N.

References and Notes

(1) Wood, R. H.; Yezdimer, E. M.; Sakane, S.; Barriocanal, J. A.; Doren, D. J. *J. Chem. Phys.* **1999**, *110*, 1329.

(2) Sakane, S.; Yezdimer, E. M.; Liu, W.; Barriocanal, J. A.; Doren, D. J.; Wood, R. H. *J. Chem. Phys*. **2000**, *113*, 2583.

(3) Liu, W.; Sakane, S.; Wood, R. H.; Doren, D. J. *J. Phys. Chem. A* **2002**, *106*, 1409.

(4) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578.

(5) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys*. **1977**, *23*, 187.

(6) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K., 1987.

(7) Caldwell, J.; Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc*. **1990**, *112*, 9144.

(8) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys*. **1994**, *101*, 6141.

(9) Savitzky, A.; Golay, M. J. E. *Anal. Chem*. **1964**, *36*, 1627.

(10) Efron, B. *The jackknife, the Bootstrap and other resampling plans*; Society for Applied Mathematics: Philadephia PA, 1982.

(11) Silvestrelli, P. L.; Parrinello, M. *J. Chem. Phys*. **1999**, *111*, 3572.

(12) Sopor, A. K. *Chem. Phys*. **2000**, *258*, 121.

(13) Chialvo, A. A.; Yezdimer, E. M.; Driesner, T.; Cummings, P. T. *Chem. Phys*. **2000**, *258*, 109.

(14) Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T. *J. Chem. Phys*. **2000**, *113*, 9149.