Studies on free energy calculations. I. Thermodynamic integration using a polynomial path

Haluk Resat and Mihaly Mezei
Department of Physiology and Biophysics, Mount Sinai School of Medicine, CUNY, New York, New York 10029-6574

(Received 26 April 1993; accepted 2 July 1993)

The difference in the free energy of hydration between the C_7 and α_R conformations of alanine dipeptide at infinite dilution is computed using Monte Carlo method in a canonical ensemble with the united-atom AMBER force field. The recently introduced polynomial path with different exponents for the calculation of liquid state free energies using thermodynamic integration (TI) is tested. This is achieved by separating the interaction energy between the solute and the solvent molecules into the constituent r^{-12}, r^{-6} and Coulomb terms. The separate contribution of each term is calculated and the comparison shows that the shape of the TI integrand as a function of the coupling parameter is mainly determined by the Coulomb term. Analysis of the convergence characteristics shows that five-point Gaussian quadrature integration would be sufficient to obtain quantitatively reliable results when a polynomial path is employed.

I. INTRODUCTION

Since chemical equilibria is determined by free energy differences, the calculation of the free energies has long been of the utmost interest. This is particularly true for aqueous solutions. There already exists a vast amount of literature concerning free energy calculations, methodologies used and their implementation for certain systems.[1–6] The main idea behind all these calculations is that being an extensive property, the free energy difference between two points in phase space does not depend on the selection of the path which connects the two state points.

Although the idea and the theory behind the calculation of the free energies is rather simple, a reliable implementation of its numerical calculation proved to be nontrivial.[7–16] Throughout the last decade, several methods have been introduced to overcome, or at least to minimize, the difficulties encountered in free energy computer simulations. These methods have been well documented and most of them are extensively studied in Refs. 1–16. A promising but little tested method is a recent proposal made by Mezei[7] which can be applied to the thermodynamic integration (TI) approach.

In TI, one starts by choosing a *path* between two state points for which the free energy difference is to be calculated. The parameter which defines the path is called the coupling parameter. Although, in principle, the final results do not depend on the path employed, in reality (due to the inaccuracy of the numerical implementation) the reliability of the calculated results depend on the chosen path. Thus the appropriateness of the path chosen will be reflected in the integrand as a function of the coupling parameter, a good choice giving a smooth and almost linear curve. Since the integrand is calculated along the path at certain number of discrete points and then numerically integrated, the smoothness and the curvature-free nature of the integrand become important factors in obtaining reliable results. Since a smooth and almost linear curve can be sampled efficiently by using far fewer discrete points, these factors are also important in speeding up the numerical calculations.

In the early implementations of the TI approach, a linear coupling along the coupling parameter was employed. This implementation is also referred to as linear TI.[1] It was shown first numerically and then analytically[17] that when creation of particles is involved, linear TI approach gives rise to a divergent integrand. Fortunately an analytical formula can be derived for the observed singular behavior.[1,17] In the end, this known analytical form of the integrand can be used to "subtract out" the divergent part from the integrand, thus eliminating the problem of numerical integration of a singular function. Although this approach has been used often,[1–6] it is obvious that it is open to numerical instability problems.

To overcome this singular behavior, a nearly linear path TI (NL-TI) approach has been developed.[1] In the NL-TI approach, a path is chosen in a manner that incorporates the known analytical form of the integrand from the beginning. Therefore, the resulting TI integrand is a smooth, but, unfortunately, in most cases is still a curved function. Therefore, to obtain reliable results, or at least to numerically prove their reliability, quite a number of discrete integration points may be needed.

As we show later in this report, the curved behavior of the integrand within NL-TI arises due to the *overcorrection* imposed on the longer ranged interaction terms, especially the Coulomb term. To avoid this overcorrection a further generalization of the NL-TI approach has recently been introduced.[7] In this generalization, the interaction energy is separated into prospective terms, and different polynomial paths are used simultaneously for each energy term to calculate the integrand, and hence the free energy, of the thermodynamic integration approach. This generalized form will be referred to here as the polynomial path TI (PP-TI) approach.

In this report, we will investigate the advantages and the shortcomings of the above mentioned polynomial path TI approach. We will do so by studying the solvation free

 0021-9606/93/99(8)/6052/10/$6.00

energy difference between two conformations of a small biomolecule, alanine–dipeptide.

To briefly outline what follows, the next section presents a short summary of the mathematical details of the polynomial path approach. Section III deals with the numerical procedure, and details of the molecular models and employed simulations are discussed. The results of this study are reported and analyzed in Sec. IV. A summary of our findings and future work is given in Sec. V.

II. THEORY

In the coupling parameter approach to free energy calculations, the energy function at a point λ along the employed path can be expressed as[1,7]

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

Here λ is the coupling parameter, and as it varies from 0 to 1, the system is transformed from one end point to the other. Or in other words, the system $\lambda=0$ is gradually *turned off* while simultaneously the system $\lambda=1$ is gradually *turned on*. The functions $f_1(\lambda)$ and $f_0(\lambda)$ are continuous functions with limiting values $f_1(0)=f_0(1)=0$ and $f_1(1)=f_0(0)=1$; therefore $\lambda=0$ and 1 in Eq. (1) correspond to state points with energy functions $E_0(\mathbf{X}^N)$ and $E_1(\mathbf{X}^N)$, respectively.

Once the path is specified, the free energy difference between the two states can be obtained in several ways. In this paper, we use the thermodynamic integration (TI) approach, for which[18]

$$\Delta A=A_1-A_0=\int_0^1\left\langle\frac{\partial E(\lambda)}{\partial\lambda}\right\rangle_\lambda d\lambda, \tag{2}$$

where $\langle\cdots\rangle_\lambda$ corresponds to an ensemble average of the enclosed quantity with energy function $E(\lambda,\mathbf{X}^N)$ in the Boltzmann factor.

A convenient choice for the $f(\lambda)$ functions is

$$\begin{aligned} f_1(\lambda)&=\lambda^k,\\ f_0(\lambda)&=(1-\lambda)^k. \end{aligned} \tag{3}$$

When $k>1$, this choice is called nearly linear TI (NL-TI).[1] Substituting into Eq. (2) gives

$$\Delta A=\int_0^1 k\,[\lambda^{k-1}\langle E_1\rangle_\lambda-(1-\lambda)^{k-1}\langle E_0\rangle_\lambda]d\lambda. \tag{4}$$

This particular choice of the $f(\lambda)$ functions is shown[7–10,19] to give a smooth and monotonic integrand in the above equation. As discussed in the previous section, the smoothness of the integrand is especially convenient (and necessary) when the numerical integration is executed using a limited number of integrand points such as quadratures.

Another special case of TI, called linear TI, corresponds to the particular choice $k=1$ in Eq. (3). When creation or annihilation of atoms is involved, linear TI is shown[17,20] to lead to an improper integral, i.e., a definite integral where the integrand is singular for certain values of the integration variable, in this case at the end points. For a potential of the form $1/r^e$, the asymptotic behavior of the integrand is expected to be $\lambda^{(kd/e)-1}$, where d is the dimensionality of the space. Thus using a polynomial function of the type given in Eq. (3) and choosing a high enough k would eliminate the divergence of the integrand.[1] This observation forms the basis of the NL-TI approach. An interesting feature, not noted before, of the coupling given by Eq. (3) is that when k is assigned to have the lowest possible value that still gives a nondivergent TI integrand then the volume of the solute molecule would have a linear λ dependence.[21]

The generalized polynomial path approach to TI (PP-TI) goes as follows.[7] Rather than using a single k exponent, the individual terms in the interaction potential are separated and different values of k exponents are assigned to each interaction term. With the mixed-exponent proposal, the Eqs. (1) and (4), respectively, become[7]

$$E(\lambda,\mathbf{X}^N)=\sum_\nu \lambda^{k_\nu}E_{1,\nu}(\mathbf{X}^N)+(1-\lambda)^{k_\nu}E_{0,\nu}(\mathbf{X}^N) \tag{5}$$

and

$$\begin{aligned} \Delta A&=\int_0^1\sum_\nu k_\nu\,[\lambda^{k_\nu-1}\langle E_{1,\nu}\rangle_\lambda-(1-\lambda)^{k_\nu-1}\langle E_{0,\nu}\rangle_\lambda]d\lambda\\ &=\int_0^1\sum_\nu \mathscr{T}_{1,\nu}(\lambda)+\mathscr{T}_{0,\nu}(\lambda)d\lambda\\ &=\int_0^1 \mathscr{T}_1(\lambda)+\mathscr{T}_0(\lambda)d\lambda\\ &=\sum_{n=1}^{n_q}\mathscr{C}(\lambda_n)[\mathscr{T}_1(\lambda_n)+\mathscr{T}_0(\lambda_n)], \end{aligned} \tag{6a}$$

with

$$\begin{aligned} \mathscr{T}_{1,\nu}(\lambda)&\equiv k_\nu\lambda^{k_\nu-1}\langle E_{1,\nu}\rangle_\lambda,\\ \mathscr{T}_{0,\nu}(\lambda)&\equiv -k_\nu(1-\lambda)^{k_\nu-1}\langle E_{0,\nu}\rangle_\lambda,\\ \mathscr{T}_1(\lambda)&\equiv\sum_\nu\mathscr{T}_{1,\nu}(\lambda),\\ \mathscr{T}_0(\lambda)&\equiv\sum_\nu\mathscr{T}_{0,\nu}(\lambda). \end{aligned} \tag{6b}$$

In Eq. (6a), the last equality is valid when the integration is done using n_q point quadratures, and $\mathscr{C}(\lambda_n)$ is the weight coefficient for the nth-quadrature root point. Summation over ν goes over the respective interaction potential terms. In this report an interaction potential

$$U_{ij}(r)=\frac{A_{ij}}{r^{12}}-\frac{C_{ij}}{r^6}+\frac{q_iq_j}{r} \tag{7}$$

is utilized. Therefore, in the rest of this report $\nu=12,6$, and 1 will be the respective labels for the first, the second, and the last interaction terms of Eq. (7).

PP-TI has already been tested for bulk liquid water with very satisfactory results.[7,8] In this report, starting with this generalized form, we further test PP-TI by studying the hydration free energy of alanine–dipeptide in aqueous solution at infinite dilution. Our aim is also to find the

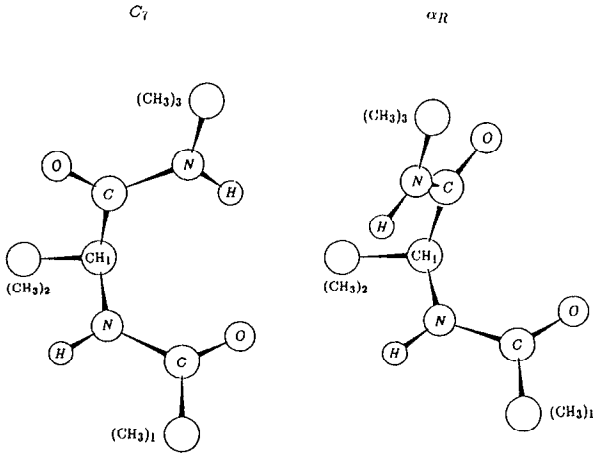

FIG. 1. C_7 and α_R conformations of alanine–dipeptide.

relative importance of the each interaction potential term, and the dependence of the smoothness of the TI integrand on the choice of the exponents set. By doing so, we will try to reflect on the selection of an optimal set of exponents $\{k_v\}$ so as to make the TI integrand a curve as close to a straight line as possible, thus reducing the magnitude of the error of quadrature integration and increasing the reliability.

III. COMPUTATIONAL DETAILS

Because of its small size and simplicity for a biomolecule, alanine–dipeptide was chosen to be our test system in this report. Another factor in this choice is the availability of earlier reports by several groups concerning its hydration properties.[22,23] Alanine–dipeptide is known to exist in several conformations with comparable probabilities. Using the standard representation (Ψ,Φ) in terms of the angles of torsion Ψ(N–C–C–N) and Φ(C–N–C–C), equatorial C_7 (90°,−90°) and α_R (−50°,−70°) conformations are believed to be the dominant structures both in gas phase and in solution. Because of their importance, we have chosen to investigate the hydration free energy difference between C_7 and α_R structures to test the mixed-exponent polynomial path choice in TI approach.

The molecular geometry of alanine–dipeptide at each conformation (see Fig. 1) was generated using the model builder of the INSIGHTII program (Biosym Technologies, Inc.) on a Silicon Graphics workstation. A rigid molecule was assumed, and bond length and angle values were taken from the system library. To keep the calculations at a simpler level, the united-atom representation was chosen. Since it is not important for this study, no *a priori* minimization with fixed torsion angles (adiabatic minimization) was applied. Therefore, the results of this study may include some enthalpic contribution. Site 6–12 potential parameters, A and C, as well as the site charges, q, in Eq. (7) were that of the AMBER force field.[24] To characterize the water, the TIP3P model[25] of Jorgensen was used. Water–solute interactions were calculated using the geometric mixing rules $A_{ij}=(A_iA_j)^{1/2}$ and $C_{ij}=(C_iC_j)^{1/2}$. A

TABLE I. Interaction potential and molecular geometry parameters.[a]

Site	A ($\times 10^{-3}$)	C	q
Solute:			
$(CH_3)_1$	2516.6	1228.8	−0.026
C	789.95	615.77	0.526
O	230.58	429.50	−0.500
N	540.68	588.25	−0.520
H	0.0819	0.0026	0.248
CH_1	592.47	461.83	0.215
$(CH_3)_2$	2516.6	1228.8	0.031
$(CH_3)_3$	2516.6	1228.8	0.272
Solvent:			
O	582.0	595.0	−0.834
H	0.0	0.0	0.417
Bond lengths			
Solute:	$l(CH_3^{(1)}\text{–}C)=l(CH_1\text{–}C)=1.51$		
	$l(CH_3^{(3)}\text{–}N)=l(CH_1\text{–}N)=1.46$		
	$l(CH_3^{(2)}\text{–}CH_1)=1.54$		
	l(C–N) = 1.48		
	l(C–O) = 1.24		
	l(N–H) = 1.08		
Solvent:	l(O–H) = 0.9572		
	l(H–H) = 1.5139		

[a]Site i of a molecule interacts with site j of a different molecule as $A/r^{12}-C/r^6+qq_j/r$ where the combination rules for A and C parameters are given in the text. A and C are in units of kcal/mol Å^{12} and kcal/mol Å^6, respectively. All distances, l, are in Angstroms.

systematic picture of the alanine–dipeptide may be seen in Fig. 1, and a list of the potential parameters used in this report are given in Table I.

TI integrand values were determined by (N,V,T) ensemble Metropolis Monte Carlo (MC) simulations. A preferential sampling of type $1/R_c$, where R_c is the distance between the center of masses of the solute and the solvent molecules, was used to accelerate the sampling. Face-centered cubic periodic boundary conditions were employed. Water–water short range interactions were truncated at a spherical cutoff of 7.75 Å, and the minimum image convention was used for the solute–water interactions. The system involved a single solute and 215 solvent molecules at $T=298$ K. The volume of the cubic unit cell, $V=6481.153$ Å^3, was determined using the density reported by Bose and Hudt.[26]

Simulations at each λ point involved an equilibration period of 2×10^6 (2 M) configurations, followed by a consequent run of 10 M configurations ($\approx 4.63\times10^4$ steps/molecules). We note that initial drifts in energies were considerable, therefore, long equilibration runs were necessary to obtain well equilibrated configurations. Due to this, use of shorter equilibration runs resulted in larger thermodynamic cycle errors. Simulations were performed on IBM RS6000/550 and Convex C220 machines. Further details of the simulations may be found in Table II.

IV. RESULTS

A. Thermodynamic integration integrand

Table II summarizes the results for the hydration free energy for the C_7 to α_R conformational change for alanine–

TABLE II. Simulation run information.[a]

λ	$\mathcal{T}$ $(C_7 \rightarrow \alpha_R)$	$\mathcal{T}_0$ $(C_7 \rightarrow \alpha_R)$	$\mathcal{T}_1$ $(C_7 \rightarrow \alpha_R)$
	Run #1. eight-point quadrature with $\{k_{12},k_6,k_1\}=\{4,4,4\}$.		
0.019 86	144.448±3.137	126.962	17.486
0.101 67	101.741±2.158	76.304	25.437
0.237 23	42.019±1.495	28.486	13.533
0.408 28	9.673±0.747	3.476	6.198
0.591 72	−10.270±0.889	−7.163	−3.107
0.762 77	−46.835±1.486	−13.003	−33.831
0.898 33	−114.087±2.687	−20.832	−93.255
0.980 14	−184.030±4.267	−19.942	−164.088
ΔA	−4.240±1.919	13.345	−17.584
	Run #2. eight-point quadrature with $\{k_{12},k_6,k_1\}=\{4,3,2\}$.		
0.019 86	81.727±3.291	60.126	21.601
0.101 67	69.167±2.781	49.449	19.718
0.237 23	40.727±1.608	31.119	9.608
0.408 28	11.749±1.712	13.675	−1.926
0.591 72	−16.275±1.372	1.427	−17.703
0.762 77	−50.999±1.663	−11.779	−39.219
0.898 33	−77.812±2.256	−18.601	−59.211
0.980 14	−92.621±3.044	−18.726	−73.894
ΔA	−3.944±1.968	11.298	−15.242
	Run #3. five-point quadrature with $\{k_{12},k_6,k_1\}=\{4,4,4\}$.		
0.046 91	122.858±2.742	104.855	18.002
0.230 76	40.733±1.285	28.611	12.122
0.500 00	0.658±0.839	−3.144	3.802
0.769 24	−51.257±1.711	−16.444	−34.812
0.953 09	−153.011±3.290	−22.955	−130.057
ΔA	−5.904±1.862	11.720	−17.623
	Run #4. five-point quadrature with $\{k_{12},k_6,k_1\}=\{4,3,2\}$.		
0.046 91	72.962±2.522	55.976	16.986
0.230 76	40.408±1.546	30.592	9.816
0.500 00	−0.200±1.234	7.203	−7.403
0.769 24	−48.639±1.946	−9.487	−39.152
0.953 09	−92.608±2.842	−19.881	−72.727
ΔA	−4.354±1.903	11.375	−15.729

[a]TI integrand, $\mathcal{T}$, and ΔA values are in kcal/mol. The estimated errors are calculated using method of batch means with batch sizes of 10^5 MC steps, and corresponds to two standard deviations. $\mathcal{T}_0$ and $\mathcal{T}_1$ are defined in Eq. (6).

dipeptide, and tabulates the integrand values at simulation λ points. Two sets of simulations for each k-exponent set were performed involving thermodynamic integration based on five-point and eight-point Gaussian quadratures. Convergence statistics of each simulation run are given in Fig. 2.

Even though three of the four simulation sets agree very well, the five-point Gaussian quadrature with $\{k\}=\{4,4,4\}$ run shows a systematic disagreement with the rest. While the disagreement between this run and the others are still within the estimated statistical error limit, extending the runlength for the five-point equal-exponent case with another 4 M steps (not shown) did not change the comparison. Therefore we partially attribute this observed disagreement to inadequacy of five-point Gaussian quadrature with equal exponents.

Figure 3 reports the respective TI integrands as a func-

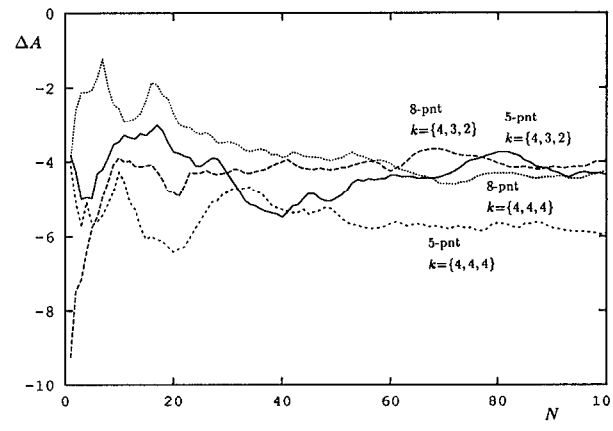

FIG. 2. Convergence characteristic of MC runs for transition $C_7 \rightarrow \alpha_R$. Additional details may be found in Table II. The ordinate scale is in kcal/mol and horizontal axis is in number of blocks with a block size of 10^5 MC steps. Run #1: dotted; run #2: long-dashed; run #3: short-dashed; and run #4: solid lines.

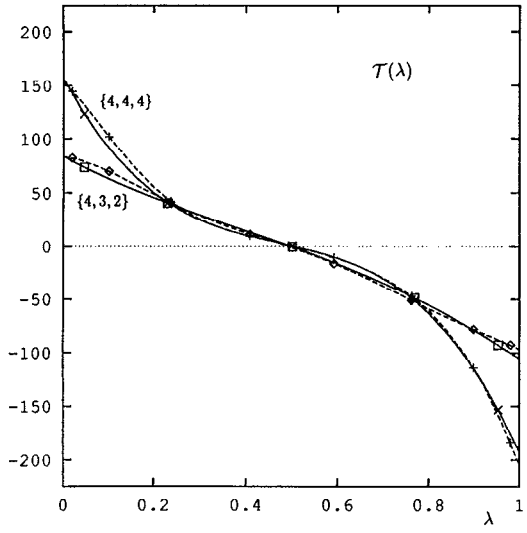

FIG. 3. Thermodynamic integration integrand for the transition $C_7 \rightarrow \alpha_R$. Horizontal scale is the coupling parameter and unitless. Ordinate scale is in kcal/mol. (+) and (×) are eight and five-point Gaussian quadratures MC results with $k=\{4,4,4\}$, respectively; and (◇) and (□), respectively, are eight- and five-point Gaussian quadratures MC results with $k=\{4,3,2\}$. Dotted and solid lines represents the eight- and five-point Gaussian quadrature fit polynomials to the simulation results.

tion of the coupling parameter. The PP-TI integrand is almost linear, and its y range is about half of that of the NL-TI. What is more important is that the slope at the end points is significantly reduced with PP-TI. This difference in slopes is important in the sense that a larger slope makes the extrapolation needed in taking the integral more vulnerable to errors due to possible statistical fluctuations. Hence, the difference in the slopes most likely would have been a factor in the agreement for the case of PP-TI and in the disagreement for the case of NL-TI between the five- and eight-point Gaussian quadrature based integrals. It is not so obvious from the figure but the results using eight-point Gaussian quadrature actually produce some weakly oscillatory integrands, an unreal feature. A similar experience with eight-point Gaussian quadrature has been encountered before (see Fig. 2 of Ref. 19). The oscillatory integrand most probably arises due to the inclusion of too many *fitting* parameters (n-parameters for n-point quadrature).

To further investigate the accuracy of numerical integration, we combine the results of the five- and eight-point Gaussian quadrature runs for each k-exponent group to obtain a set of 13-points along the thermodynamic integration path. Use of this combined set enables us to perform three additional tests.

The first test involves an *ordinary* 13-point quadrature fit to the combined 13-point set and then analytically integrating over λ. Here the word ordinary stands to point out that the λ points of the combined set were an *arbitrary* combination of quadrature points, rather than an optimal choice as in the case of *Gaussian* quadratures. For $k=\{4,3,2\}$ exponents, this test way of integrating gives -3.945 kcal/mol for the free energy difference between the

TABLE III. Different order polynomial fits to the free energy data.[a]

No. of fit parameters	$\Delta A(C_7 \rightarrow \alpha_R)$ $\{k\}=\{4,4,4\}$	$\Delta A(C_7 \rightarrow \alpha_R)$ $\{k\}=\{4,3,2\}$
3	−5.0018	−4.0737
4	−5.0019	−4.0737
5	−4.9156	−4.1001
6	−4.9156	−4.1001
7	−4.8926	−4.0995
8	−4.8927	−4.0995
9	−4.8925	−4.0994

[a]ΔA values are in kcal/mol.

C_7 and α_R conformations. This is to compare with -3.944 and -4.354 kcal/mol, respectively, for eight- and five-point Gaussian quadrature integrals (from Table II). Similarly for $k=\{4,4,4\}$, a 13-parameter fit gives -4.238 kcal/mol, compared to -4.240 and -5.904 kcal/mol for eight- and five-point Gaussian quadratures. Therefore the 13-parameter fit results are much closer to eight-point Gaussian quadrature results. This is to be expected on the basis that, since it involves more data points, eight-point Gaussian quadrature results will be favored in a functional fit. Note that the 13-parameter fit involves a 12th order polynomial and, based on experience with the eight-point Gaussian quadrature case, is expected to produce oscillations. Although not reported, the oscillations do appear implying the necessity of caution in using high-order quadratures.

In the second test, polynomials of varying order were least square fitted to 13 MC simulation points for each k-exponent set. This was followed by analytical integration with results tabulated in Table III. This approach would point to the lowest possible order polynomial which captures all the essentials of the TI integrand. Knowledge of this polynomial with the lowest possible order would be important in setting up future simulations for related or chemically similar systems. All these varying order fits are essentially in very good agreement with each other. Their differences are unnoticeable by eye on a single plot and therefore are not reported. This very good agreement is also reflected in Table III showing that a four-parameter (third order polynomial) fit is essentially sufficient to obtain quantitatively reliable results. As experienced before, using polynomials of order higher than six produce some weak oscillations in the integrand.[27]

We should point out that simulation run lengths of this report ($\approx 4.63\times10^4$ MC cycles$\approx$46 ps of molecular dynamics) are long compared with most of the other published simulation results for similar systems. This might have led to obtaining a smoother integrand curve than usual, and hence might be misleading. For this reason, a safer range of choosing the order of the fit polynomial would be 5–6, which avoids using too few simulation points and also avoids possible unphysical oscillatory integrands. But given the disagreement encountered in this study for the case of the five-point equal-exponent Gaussian quadrature approach, the above stated polynomial range can only be taken as a good starting guess for setting

up a simulation and needs to be supplemented with some additional tests. With the preceding caution in mind it still may be stated, based on the experience of this study, that five or six-point Gaussian quadratures would be the most reliable ones to use in TI approach with a fair degree of assurance.

Results of the second test lead us to our third test. In this test, for each k-exponent set, we proceed with the fourth order polynomial least squares fit of the above given second test. As mentioned before, this fit is in principle equivalent to a five-point Gaussian quadrature integration, but using a wider set of integrand points. Now using this fit function, we interpolate the TI integrand at the λ points which form the root points of the n-point Gaussian quadratures. We repeat this calculation for $n=3$ to 9. The aim of this test, as in the above case, is to find a convenient range for the number of quadrature points that one needs to employ to calculate the conformational free energy differences using the thermodynamical integration approach. The difference between ΔA values calculated using $n=3$ to 9 were less than 10^{-4} kcal/mol. Therefore, the results of the third test supports the conclusion reached in the second test.

Since they are based on a wider (13-point) set of simulation results, and because of the good consistency between them, we accept $n=5$ and 6 results of the second and the third tests to be the most reliable ones. Using these values, we see that the conformational free energy difference values for the separate runs with $\{k\}=\{4,4,4\}$ and $\{4,3,2\}$ exponent sets are in reasonably good agreement with each other. Averaging these two, we estimate the free energy of hydration difference between the C_7 and α_R conformations of alanine–dipeptide in liquid TIP3P model water to be $\simeq -4.5 \pm 1.9$ kcal/mol in favor of the α_R conformation. This calculated free energy difference between the C_7 and α_R conformations of alanine–dipeptide is quantitatively in the same range of earlier studies,[8,28] and it does not include the contributions from the internal degrees of freedom. Assuming that inter- and intramolecular degrees of freedom do not couple, and using the ΔA^{intra} value reported in Ref. 28 as an approximate value for our molecular model, we see that the intermolecular contribution, which is in favor of α_R conformation, to the conformational free energy difference is approximately nullified by the intramolecular contribution; thus giving comparable existence probabilities for C_7 and α_R conformations in aqueous solution.

What is also apparent in Table II is the fact that the quadrature points close to the end points carry relatively larger uncertainties compared with the ones near the center, $\lambda=0.5$. Thus, using the same total number of MC steps, i.e., same amount of numerical work, but distributing them among the quadrature points in an optimal way such that quadrature points close to end points would have longer run lengths than the ones in the vicinity of $\lambda=0.5$ would reduce the overall uncertainty. This finding supports argument given by Mezei,[8] and might be useful in obtaining numerical results with relatively less statistical error with the minimal amount of numerical work.

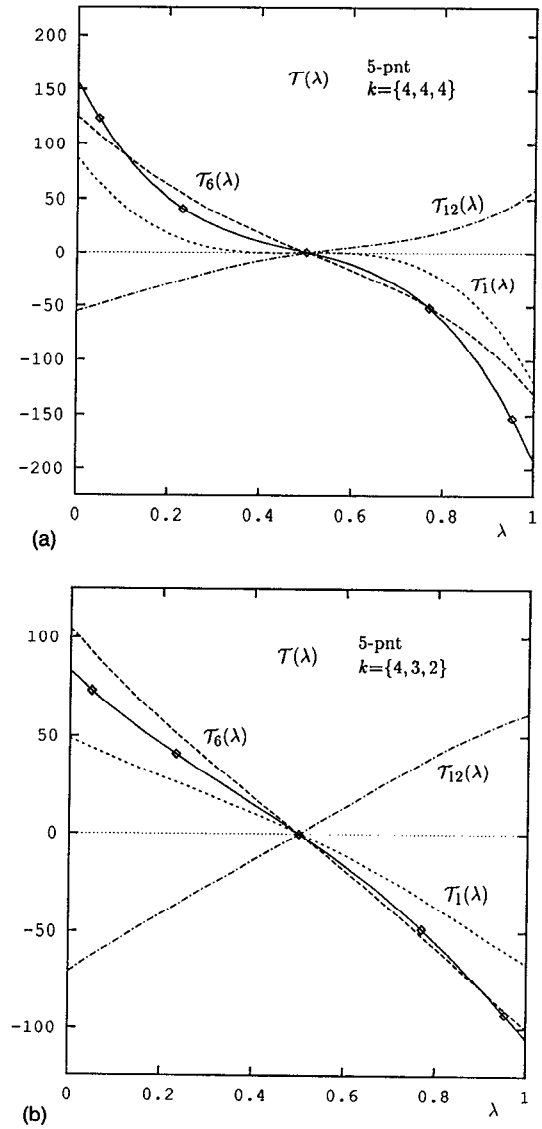

FIG. 4. (a) Contribution of each interaction term to the total TI integrand. Five-point Gaussian quadrature with $k=\{4,4,4\}$. Solid line: total TI integrand; dash–dot, dashed, and dotted lines, respectively, are the contributions of $1/r^{12}$, $1/r^6$ and Coulomb terms. (b) Same as in (a) but for $k=\{4,3,2\}$.

B. Separation into interaction energy terms

In this section, we investigate the contribution of individual interaction energy terms (see Sec. II) to pinpoint the term which gives rise to a curved TI integrand. To do this, in Table II, we first separate the TI integrand into contributions of $\lambda=0$ (C_7 configuration) and $\lambda=1$ (α_R configuration) solutes, $\mathcal{T}_0(\lambda)$ and $\mathcal{T}_1(\lambda)$, respectively. These two contributions are qualitatively similar (see Figs. 5, 6, and 8) and they closely resemble the TI integrand for going from an uninteracting point particle to a certain conformation of the solute particle.[19,20]

Figures 4–6 show the partition of the TI integrand into constituents according to the interaction potential terms. It can be seen in Fig. 4(a) that for the case of equal exponents $k_{12}=k_6=k_1=4$, the contribution of the r^{-12} and r^{-6} terms are close to straight lines; thus it is mainly the Cou-

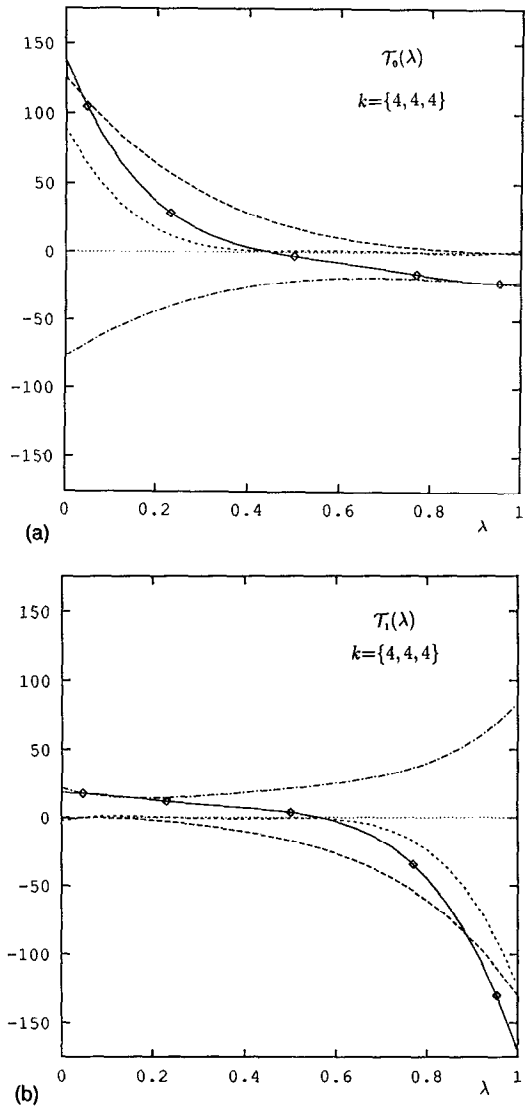

FIG. 5. (a) Contribution of C_7 conformation solute to the TI integrand of transition $C_7 \rightarrow \alpha_R$, and its separation into constituent interaction terms. Five-point Gaussian quadrature with $k=\{4,4,4\}$. Solid line: C_7 part of the TI integrand, $\mathcal{T}_0$; dash–dot, dashed, and dotted lines, respectively, are the contributions of $1/r^{12}$, $1/r^6$ and Coulomb terms. (b) Contribution of α_R conformation solute to the TI integrand of transition $C_7 \rightarrow \alpha_R$, and its separation into constituent interaction terms. Five-point Gaussian quadrature with $k=\{4,4,4\}$. Solid line: α_R part of the TI integrand, $\mathcal{T}_1$; dash–dot, dashed, and dotted lines, respectively, are the contributions of $1/r^{12}$, $1/r^6$ and Coulomb terms.

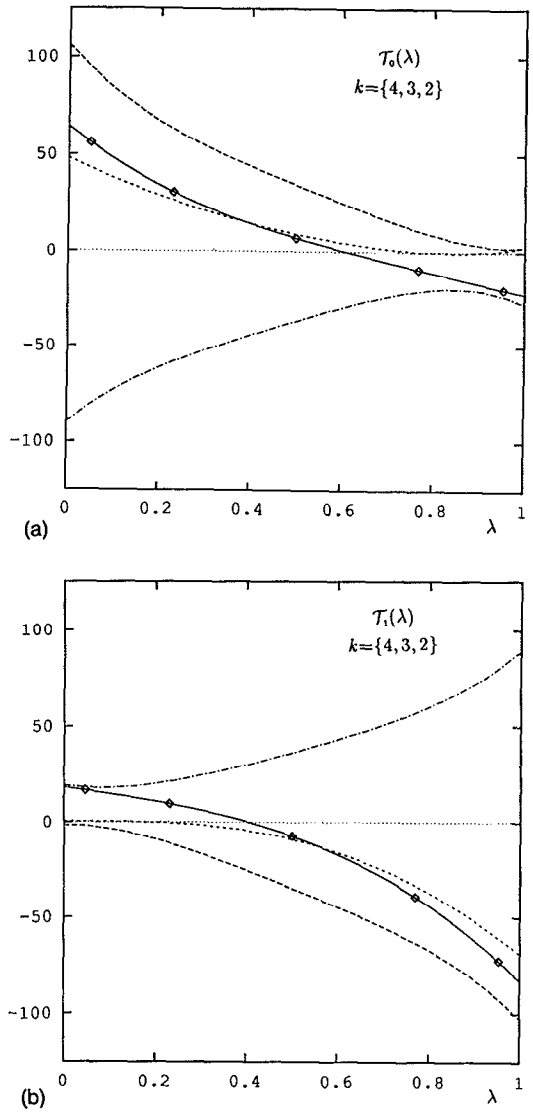

FIG. 6. (a) and (b) are same as in Figs. 5(a) and 5(b), but for the case $k=\{4,3,2\}$.

lomb term which gives rise to a curved TI integrand. In contrast, using a lower exponent for the $1/r$ term distributes the contribution of the Coulomb interaction more evenly through the coupling parameter range. One of the main objectives of this study was to show that by choosing a k-exponent set with different exponents for different interaction energy terms it would be possible to straighten and lower the slope of the TI integrand function. This would in turn be expected to decrease the magnitude of the error of integration, and increase the reliability of the results. As it is obvious from the comparison of Figs. 4(a) and 4(b) that the choice of $\{k_{12},k_6,k_1\}=\{4,3,2\}$ brings the TI integrand curve very close to a straight line, thus fulfilling the desired objective.

Success of the unequal exponent approach allows us to suggest the following recipe to increase the reliability and possibly lower the numerical error of the thermodynamical integration approach in the solvation free energy calculations: (a) First run a short simulation with a certain k'-exponent set (say $\{k'_{12},k'_6,k'_1\}$) to obtain the $\mathcal{T}_{j,\nu}(\lambda)$ values where $j=0$ or 1 and $\nu=12$, 6, or 1. (b) Then form the cumulants $\lambda^{m_\nu}\mathcal{T}_{1,\nu}(\lambda)$ and $(1-\lambda)^{m_\nu}\mathcal{T}_{0,\nu}(\lambda)$. (c) Choose the m_ν values which give a straight line for the cumulants. Using these m_ν obtain a new k-exponent set as $\{k\}=\{k'\}+\{m_\nu\}$. This new set would be a very good estimate for obtaining a straight TI integrand.[29] (d) After short test simulation runs, if necessary, make adjustments to the exponent set.

What is studied in this report is the free energy difference between two conformations of the same biomolecule. Therefore there is in a way a symmetry between the two λ

TABLE IV. Solvation free energies.

	five-point quadrature with $\{k_{12},k_6,k_1\}=\{4,3,2\}$	
λ	$\mathcal{T}$ ($C_7\rightarrow$IG)	$\mathcal{T}$ (IG$\rightarrow\alpha_R$)
0.046 91	54.367	46.146
0.230 76	33.189	17.985
0.500 00	6.487	−7.236
0.769 24	−19.282	−40.864
0.953 09	−39.645	−70.568
ΔA (kcal/mol)	6.917	−10.426

end points of the system. In this respect, this study complements a previous study.[7] In Ref. 7, the PP-TI approach was applied to study the solvation free energy of liquid water. For liquid water, the investigated quantity was the change in the free energy of a water molecule in going from an ideal gas state to a condensed phase. Therefore, there is an asymmetry between the two end points. Since the PP-TI approach seems to work very well for these two complementary test cases, it can be confidently stated that it is possible to obtain, for almost any system, an integrand which is very close to a straight line within the PP-TI approach. Since the error involved in integration (or extrapolation/interpolation) of a straight line is minimal, this achievement increases the accuracy of integration.

C. Thermodynamic cycle

To further study the convergence and the statistical error involved in the simulations, we make use of the thermodynamic circle idea.[30] For this, we defined a triangle. This requires two additional simulation runs: simulations of the ideal gas to C_7 and of the ideal gas to α_R conformation transitions. This forms a closed loop and the deviation of the free energy change of traversing one complete cycle from zero gives an idea about the statistical error of the simulations. The results of these runs are summarized in Table IV. Adding up the free energy values defines the

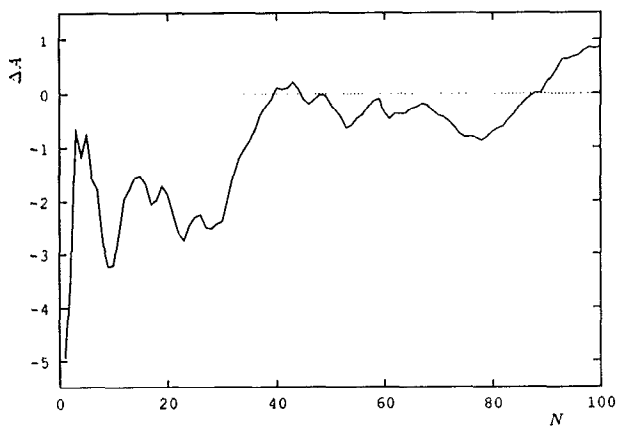

FIG. 7. Thermodynamic cycle error which is defined by Eq. (7) in the text. $k=\{4,3,2\}$ and five-point Gaussian quadratures is used. The ordinate scale is in kcal/mol and horizontal axis is in number of blocks with a block size of 10^5 MC steps.

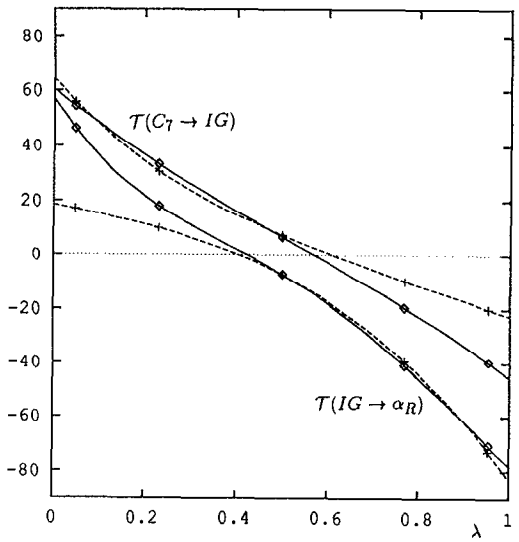

FIG. 8. Solid lines: TI integrand for transitions $C_7\rightarrow$IG and IG$\rightarrow\alpha_R$ (IG=ideal gas). Horizontal axis is the coupling parameter, and ordinate scale is in kcal/mol. Dotted lines are the $\mathcal{T}_0$ and $\mathcal{T}_1$ of the transition $C_7\rightarrow\alpha_R$ as tabulated in Table II, and, respectively, correspond to the contribution of C_7 and α_R conformation solute to the total TI integrand.

cycle error (IG=ideal gas)

$$\Delta A_{C_7\rightarrow\alpha_R}+\Delta A_{\alpha_R\rightarrow\mathrm{IG}}+\Delta A_{\mathrm{IG}\rightarrow C_7}=\text{cycle error.} \quad (8)$$

In Fig. 7, cycle error is reported as a function of run length. Although it oscillates with a period of approximately 2 million steps, the cycle error has a magnitude of less than 1 kcal/mol. This value is well within the statistical error estimates (approximately 2 kcal/mol) based on the fluctuations of the system. Another concern in Fig. 7 is what looks like a steady upward trend. Even though it is not reported, follow-up simulations showed that the reported cycle error takes a downturn and stays as less than 1 kcal/mol.

Thermodynamic cycle error and the convergence statistics reported in Figs. 7 and 2 show that the transitions studied here are fairly well converged around 6 M steps, but before that the fluctuations are considerable. This puts a lower limit to simulation run lengths. 6 M steps is rather long and emphasizes the well established[1-16] fact that simulations of biosystems (even the smallest ones) require quite lengthy computations.

D. Solvation free energies

Additional simulations that were performed to calculate the thermodynamic cycle error enable us to calculate the solvation free energies of each conformation. These simulations were done using PP-TI with an exponent set $k=\{4,3,2\}$, and involved a five-point Gaussian quadrature integration. The simulation results are tabulated in Table IV and are also reported in Fig. 8. As Fig. 8 reveals, the use of the PP-TI for the $C_7\rightarrow$IG transition results in a TI integrand which is again very close to a straight line. However, for the IG$\rightarrow\alpha_R$ transition, the deviation of the TI integrand from a straight line is still quite apparent. There-

fore, the $k=\{4,3,2\}$ exponent set which gives satisfactory results for the other cases studied in this report needs to be further modified for studying the IG$\rightarrow\alpha_R$ transition. This observation points to the fact that the optimal choice for the exponent set is not unique but is system dependent.

Results for the solvation free energies (Table IV) show that intermolecular interaction with the solvent molecules has a large stabilizing effect on both of the solute conformations, 6.92 and 10.43 kcal/mol for C_7 and α_R, respectively. To find out if the C_7 and the α_R conformations of alanine–dipeptide would favor either the gas phase or aqueous solution, contributions coming from the intramolecular degrees of freedom are also needed, and this is not done here.

V. SUMMARY AND DISCUSSION

Mixed exponents for the polynomial path have been employed for the calculation of the hydration free energy difference between the C_7 and α_R conformations of alanine–dipeptide with quite satisfactory and promising results. The role of different exponents has been investigated by separating the interaction between the solute and the solvent into constitutive terms. This enabled us to point out the importance of each particular term, especially the Coulomb term, and, by proper adjustment, helped us to reduce both the curvature of the integrand and its slope at the endpoints. Both of these effects are helpful in reducing the quadrature integration error.

It has also been shown that inclusion of more than the necessary fitting parameters sometimes produces some unexpected and possibly unphysical features in the TI integrand, signaling the need for extra caution when using fitting polynomials with large powers. The quantitative results with TI based on eight-point and five-point Gaussian quadratures are found to be within the statistical accuracy, but the example of this report showed that the agreement is improved when the TI integrand is linearized. Detailed study of our results showed that for the hydration of alanine–dipeptide case (which is simple but sophisticated enough to suggest that this finding is quite general) the use of five-point quadrature over the polynomial path is enough and possibly better than eight-point quadrature. Based on the findings of this report and an earlier study,[7] the choices $k_{12}=4$, $k_6=2$ or 3, and $k_1=2$ or 1 for the exponents would be a good starting point in the free energy of hydration calculations via thermodynamic integration based on polynomial path with mixed-exponents (PP-TI).

The calculations over the thermodynamic cycle IG$\rightarrow C_7\rightarrow\alpha_R\rightarrow$IG show that TI over the path of Eq. (4) (NL-TI) and especially over the path of Eq. (5) (PP-TI) are capable of handling the creation of larger solutes with the same ease as calculating solvation free energy differences, supporting the earlier suggestion[8] that it is the method of choice for calculating free energy differences between very different systems.

The aim of this study was to test study the ways to improve the quadrature integration. Of course, there are more fundamental sources of error, which we have not addressed here, in numerical calculations of statistical systems that also must be considered for an overall reduction of error in computations. But, our numerical results show that the error arising from improper quadrature integration might be on the same order as the statistical sampling error. To give some numbers, let us consider the $C_7\rightarrow\alpha_R$ transformation studied in this report. Reading from Table II, the sampling error for this transition is approximately 1.9 kcal/mol. This is to compare with the quadrature integration error. Taking the calculated solvation free energy using five- and eight-point quadratures, we see that the integration error might be of order 1.6 kcal/mol ($k=4,4,4$ case). Thus these two sources of error are of the same magnitude, and both need serious consideration.

In this study, we have not incorporated the protons (carbon hydrogens) into the model explicitly. This was done to reduce the numerical work in testing PP-TI. With the promising results of this report, it would be interesting to see if the explicit inclusion of hydrogens (all-atom model) would make any difference in the success of the unequal exponent approach. Another interesting point is of course to study the effect of the hydrogens on the hydration free energies. Work along these lines as well as the study of the dependence on site charges is currently in process in our laboratory and will be the subject of a future communication.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R55-GM43500. Computing resources were provided in part by the City University of New York, University Computing Center.

[1] M. Mezei and D. L. Beveridge, Ann. Acad. Sci. (N.Y.) **482**, 1 (1986).
[2] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Chem. **18**, 431 (1989).
[3] W. L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).
[4] M. Karplus and G. A. Petsko, Nature (London) **347**, 631 (1990).
[5] C. A. Reynolds, P. M. King, and W. G. Richards, Mol. Phys. **76**, 251 (1992); P. M. King, C. A. Reynolds, J. W. Essex, G. A. Worth, and W. G. Richards, Mol. Simul. **5**, 265 (1990).
[6] T. P. Straatsma and J. A. McCammon, Annu. Rev. Phys. Chem. **43**, 407 (1992).
[7] M. Mezei, J. Comp. Chem. **13**, 651 (1992).
[8] M. Mezei, Mol. Simul. **10**, 225 (1993).
[9] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[10] A. J. Cross, Chem. Phys. Lett. **128**, 198 (1986).
[11] J. Schlitter, Mol. Simul. **7**, 105 (1991); J. Schlitter and D. Husmeier, *ibid.* **8**, 285 (1992).
[12] S. H. Fleischman and D. A. Zichi, J. Chim. Phys. (Paris) **88**, 2617 (1991).
[13] A. Pohorille and L. R. Pratt, Methods Enzymol. **127**, 48 (1986).
[14] W. L. Jorgensen and C. Ravimohan, J. Am. Chem. Soc. **83**, 3050 (1985).
[15] M. J. Mitchell and J. A. McCammon, J. Comput. Chem. **12**, 271 (1991).
[16] M. H. Mazor and B. M. Pettitt, Mol. Simul. **6**, 1 (1991).
[17] M. R. Mruzik, F. F. Farid, F. Abraham, D. E. Screiber, and G. M. Pound, J. Chem. Phys. **64**, 481 (1976). For a similar earlier application see D. R. Squire and W. G. Hoover, *ibid.* **50**, 701 (1969).
[18] J. G. Kirkwood, *Theory of Liquids*, edited by B. J. Alder (Gordon and Breach, New York, 1968); J. Chem. Phys. **3**, 300 (1935).
[19] M. Mezei, Mol. Simul. **2**, 201 (1988); Mol. Phys. **47**, 1307 (1982); **67**, 1205(E) (1989).
[20] J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, Faraday Symp. Chem. Soc. **17**, 55 (1982).

[21] This can be seen from the following argument. Consider a spherical particle and define its radius r_λ as the value at which the shortest ranged repulsion term reaches a given threshold value B. For a A/r^e type repulsion this can be obtained as $\lambda^k A/r_\lambda^e = B$. Since the lowest possible value of k which gives a finite integrand is $k=e/d$, then it is straightforward to show that $r_\lambda^d \sim \lambda$.
[22] M. Mezei, P. K. Mehrotra, and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2239 (1985); M. Mezei, J. Comput. Phys. **68**, 237 (1987).
[23] (a) D. A. Brandt and P. J. Flory, J. Mol. Biol. **23**, 47 (1967); (b) P. J. Rossky, M. Karplus, and A. Rahman, Biopolymers **18**, 825 (1979); (c) Z. I. Hodes, G. Nemethy, and H. A. Scheraga, *ibid.* **18**, 1565 (1979).
[24] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profate, Jr., and P. Weiner, J. Am. Chem. Soc. **106**, 765 (1984).
[25] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[26] L. Bose and A. Hudt, J. Chem. Thermodyn. **3**, 663 (1971).
[27] For an argument in favor of possible *kinks* in the integrand, see Ref. 12. But, as Mezei discusses (Ref. 7), the integrand might be written in terms of fluctuations of energies. Therefore the integrand should be a smooth function without any kinks.
[28] (a) M. Mezei, P. K. Mehrotra, and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2239 (1985); (b) M. Mezei, J. Comput. Phys. **68**, 237 (1987).
[29] This estimate is based on the limiting behavior arguments given in Secs. II and IV and is straightforward to derive. Also note that even though, for simplicity, we have only used integer exponent values in this report, noninteger exponent values may also be utilized.
[30] B. L. Tembe and J. A. McCammon, Comput. Chem. **8**, 281 (1984).