JOURNAL

OF THE AMERICAN CHEMICAL SOCIETY

VOLUME 109, NUMBER 6 MARCH 18, 1987

An Approach to the Application of Free Energy Perturbation Methods Using Molecular Dynamics: Applications to the Transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, Glycine → Alanine, and Alanine → Phenylalanine in Aqueous Solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the Gas Phase

U. C. Singh,[†] Frank K. Brown,[‡] Paul A. Bash,[⊥] and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. *Received May 12, 1986*

Abstract: A general method for the application of free energy perturbation methods using molecular dynamics has been developed which implements both a time dependent and a time independent procedure. The free energy difference between two states is determined by slowly perturbing state A into state B via a coupling factor λ. This method has been applied to the transformation of methanol into ethane, ammonium into oxonium, glycine into alanine, and alanine into phenylalanine. The calculated changes in free energy for all of these transformations are in reasonable agreement with the related experimental data, which suggests the method can be usefully applied to a wide range of chemical and biochemical problems.

One of the fundamental difficulties in the extraction of useful properties from the simulation of complex molecules has been the enormous statistical fluctuations found when calculating relative internal energies and enthalpies. These uncertainties are an inherent property of the simulation, given that one must evaluate small differences between large numbers, each of which has large statistical fluctuations. It was thus very exciting to us to realize that such immense statistical uncertainty is *not* an intrinsic property of simulations of free energy differences, particularly when one is dealing with closely related systems. Thus, when the question can be formulated in terms of solvation free energy differences, the evaluation turns out to involve evaluating ensemble averages of differences in solute–solvent energies, which converge much faster than solvent–solvent energies and thus can be evaluated much more precisely. In addition, free energy is generally a quantity of greater utility than internal energy/enthalpy.

The free energy perturbation approach applied to cavity formation in water and solubility of noble gases in water by Postma et al.,[1] to the solvation of CH_3OH and CH_3CH_3 by Jorgensen and Ravimohan,[2] and to differences of Cl^- and Br^- solvation and SC24/H^+ ion association by Lybrand et al.[3,4] leads to results in encouraging agreement with relevant experiments. These results further suggest that the approach is general in the sense that it can be implemented either through Monte Carlo or molecular dynamics methods. Only Jorgensen's $CH_3OH \rightarrow CH_3CH_3$ conversion involves a *large* change in the system, and this used the Monte Carlo method. Molecular dynamics has thus far been applied to systems with relatively small perturbations. For example, the perturbation of Cl^- to Br^- involved only a change in the van der Waals radius of 0.15 Å. This simulation required only two steps, a calculation with Cl^- radius and another with Br^- radius. No intermediate values between the two radii were necessary to adequately characterize the free energy change. In the formation of cavities in water which are filled by noble gas atoms, molecular dynamics approaches were used, but again the molecular nature of the system studied was rather limited. More recently Wong and McCammon[5] studied the change of benzamidine to *p*-fluorobenzamidine and glycine to alanine in trypsin. These calculations showed how one could get reasonable free energy differences in large systems, but the difference in the perturbed and the unperturbed systems was small enough that separate simulations of the end points of the perturbations were adequate to calculate the free energy differences. This will not be true for general and large perturbations.

It is important to place the study presented here in perspective with previous work in free energy simulations. Mezei et al.[6] were

[†] Current address: Laboratory of Molecular Biology, Scripps Clinic, La Jolla, CA 92037.

[‡] Current address: Smith, Kline and French Labs, Philadelphia, PA 19101.

[⊥] Current address: Department of Chemistry, Harvard University, Cambridge, MA 02138.

(1) Postma, J. P. M.; Berendsen, H. J.; Haak, J. R. *Faraday Symp.* **1981**, *17*, 55–67.

(2) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(3) Lybrand, T.; Indira, G.; McCammon, A. *J. Am. Chem. Soc.* **1985**, *107*, 7793.

(4) Lybrand, T.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833.

(5) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* **1986**, *108*, 3830.

0002-7863/87/1509-1607$01.50/0

the first to calculate the free energy of liquid water using "the coupling" approach. Mezei and Beveridge[7] give a current review of the various approaches to study free energy changes in arbitrary systems, and a nice pedagogical summary of many of these approaches is given by Valleau and Torrie.[8]

As noted above, however, Postma et al.[1] first demonstrated the usefulness of the "perturbation method" approach used here in studying cavity formation in water using modern computer simulation methodology. Tembe and McCammon showed the advantage of the perturbation method (PM) approach over umbrella sampling in a simple Lennard–Jones model system.[9] Jorgensen and Ravimohan[2] then demonstrated the usefulness of the PM approach on real molecular solutes. It was the receipt of the preprint of ref 2 in April 1985 that convinced us of the power of the perturbation method, and at that point, one of us (U.C.S.) began to implement this approach with molecular dynamics. Here we show the implementation of the PM approach in a general way in a molecular simulation package and demonstrate its usefulness in a variety of chemical systems. In particular, we apply the approach to a number of systems of moderate chemical complexity, including aqueous solutions of CH_3OH, CH_3CH_3, OH_3^+, NH_4^+, glycine, alanine, and phenylalanine within a molecular dynamics framework. Elsewhere, we show the application of the method to protein–inhibitor interactions[10] and to the determination of the differences in solvation free energy for many amino acids.[11]

Methods

In this section we will outline the statistical perturbation theory and its implementation into the molecular dynamics program to evaluate free energy differences. Statistical mechanics perturbation theory was first developed by Zwanzig[12a] and has been applied by several workers to dense fluids.[12a-d] The Hamiltonian of a system is separated into two parts

$$H = H_0 + H_1 \tag{1}$$

where H_0 is the reference state and H_1 is a small perturbation from the reference system. When the configurational partition functions of the reference and the total states are used, the perturbation free energy is

$$G - G_0 = G_1 = -RT \ln \langle \exp(-H_1/RT) \rangle_0 \tag{2}$$

where $\langle\rangle_0$ denotes the ensemble or time average over the reference system.

We will describe a general method to calculate free energy differences, which can be used to determine free energy differences of solvation as well as relative changes in binding free energy of protein–ligand complexes. The essential part of the calculation is to define a Hamiltonian for solute states A and B, which are linked by the coupling parameter λ, as has been done by Tembe and McCammon.[9]

$$H_\lambda = \lambda H_A + (1 - \lambda) H_B \qquad 0 \le \lambda \le 1 \tag{3}$$

H_A is the Hamiltonian for A, and H_B is that for B; when $\lambda = 1$, $H_\lambda = H_A$, and when $\lambda = 0$, $H = H_B$. At intermediate values of λ, the solute is a hypothetical mixture of A and B. This type of coupling ensures the smooth conversion of the solute A into B through several stages which allows the surrounding system to readjust its configuration accordingly.

The total Hamiltonian at a given λ is

$$H_0(\lambda) = H_{sol} + H_\lambda \tag{4}$$

where H_{sol} is the Hamiltonian for the solvent or, more generally, the part of the system where the properties are not changing. We assume that $H_0(\lambda)$ is the reference state. The perturbation is introduced into the system as follows. If we divide the range of λ into N windows, $\{\lambda_i, i = 1, N\}$, at each window λ_i, the solute state is perturbed to λ_{i+1} and λ_{i-1} states by taking the reference state as $H_0(\lambda_i)$. Thus, the perturbation Hamiltonian in eq 1 becomes

$$H_1(\lambda_i) = H_{\lambda_i \pm_1} - H_{\lambda_i} \tag{5}$$

and which yields the corresponding free energy from eq 2 as

$$\begin{aligned} G_{\lambda_i \pm_1} - G_{\lambda_i} &= G_1(\lambda_i) \\ &= -RT \ln \langle \exp(-H_1(\lambda_i)/RT) \rangle_{\lambda_i} \end{aligned} \tag{6}$$

and eq 6 leads immediately to the difference in the free energy of the two solute states A and B as the summation over all the windows.

$$\Delta G = \sum_{i=1}^{N} G_1(\lambda_i) \tag{7}$$

The backward and forward evaluation of G_1 is a check for possible hysteresis in the calculation.

An alternate approach, which is referred to as time dependent perturbation or slow growth, is to define the coupling parameter λ as a function of time which is varied continuously as the system evolves with time. When $H_1(\lambda_i) << kT$, eq 6 can be reduced to

$$G_1(\lambda) = H_1(\lambda) \tag{8}$$

where the ensemble average has been replaced by a single value, and eq 7 becomes

$$\Delta G = \sum_{\lambda} H_1(\lambda) \tag{9}$$

According to eq 9, the difference in free energy is simply the sum of the perturbation energy for each time step.

The above two approaches have been implemented into the molecular simulation program AMBER-UCSF (Version 3.0)[13] in a general way, because our primary interest is to use this technique for the calculation of free energy changes for mutations in proteins and protein–substrate interactions.

The AMBER potential function has the form

$$\begin{aligned} E_{\text{Total}} = &\sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \\ &\sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \\ &\sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right] \end{aligned} \tag{10}$$

consistent with the notation below.

The group of atoms, denoted by P, to be converted from state A to state B is defined in the PARAM module of the program. Then the potential function parameters for the A and B states are generated and passed to the dynamics module. The various terms in the potential functions for P are calculated in the following way:

(1) For any bonded pair ij of atoms $i \in$ P or $j \in$ P, the bond parameters are defined by

$$\begin{aligned} K_r &= \lambda k_r^A + (1 - \lambda) K_r^B \\ r_0 &= \lambda r_0^A + (1 - \lambda) r_0^B \end{aligned} \tag{11}$$

(6) Mezei, M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1978**, *100*, 3255.
(7) Mezei, M.; Beveridge, D. L. *Proc. N.Y. Acad. Sci.*, in press.
(8) Valleau, J. P.; Torrie, G. M. *Modern Theoretical Chemistry*; Berne, B. J., Ed; Plenum: New York, 1977; Vol. 5, Statistical Mechanics, Part A, Equilibrium Techniques.
(9) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281.
(10) Bash, P. A.; Singh, U. C.; Brown F. K.; Langridge, R.; Kollman, P. A. *Science*, in press.
(11) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science*, submitted for publication.
(12) (a) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420. (b) Barker, J. A.; Henderson, D. *J. Chem. Phys.* **1967**, *47*, 2856. (c) Barker, J. A.; Henderson, D. *Annu. Rev. Phys. Chem.* **1972**, *23*, 439. (d) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(13) Amber-UCSF, version 2.0 and 3.0 have been largely developed by U. C. Singh, starting from version 1.0 described by P. Weiner and P. Kollman: *J. Comp. Chem.* **1981**, *2*, 287.

Table I. The Charges (q) and R^* and ϵ Parameters for the Perturbation of Methanol to Ethane[a] and Ammonium to Oxonium[b]

molecule	group	q	R^*, Å	ϵ, kcal/mol
CH_3OH	CH_3	0.265	2.119	0.207
CH_3OH	O	−0.700	1.723	0.170
CH_3OH	H	0.435	0.0	0.0
CH_3	CH_3	0.00	2.119	0.207
H_2O	O	−0.834	1.768	0.152
H_2O	H	0.417	0.0	0.0
NH_4^+[a]	N	−0.896	1.90	0.20
NH_4^+	H	0.474	0.0	0.0
OH_3^+	O	−0.571	1.70	0.15
OH_3^+	H	0.524	0.0	0.0

[a] See ref 2 and 16. [b] See ref 24 and 25.

(2) For any angle whose atoms $i \in$ P or $j \in$ P or $k \in$ P, the parameters are defined by

$$K_\theta = \lambda K_\theta^A + (1 - \lambda)K_\theta^B \qquad (12)$$

$$\theta_0 = \lambda\theta_0^A + (1 - \lambda)\theta_0^B$$

(3) For any dihedrals whose atoms $i \in$ P or $j \in$ P or $k \in$ P or $l \in$ P, the dihedral energy is defined as

$$E_\phi = \lambda E_\phi^A + (1 - \lambda)E_\phi^B \qquad (13)$$

(4) For any atom $i \in$ P, the partial charge is

$$q_i = \lambda q_i^A + (1 - \lambda)q_i^B \qquad (14)$$

(5) For the nonbonded energy calculation two methods have been used. In the first case, the coefficients of the potential function have been coupled and in the other the VDW radii R_i and the well depth ϵ_i of the atoms have been coupled. For any nonbonded pair ij where $i \in$ P or $j \in$ P, the coefficients are defined by

$$A_{ij} = \lambda A_{ij}^A + (1 - \lambda)A_{ij}^B \qquad (15)$$

$$B_{ij} = \lambda B_{ij}^A + (1 - \lambda)B_{ij}^B$$

The same is true for the 10–12 potential. In the other case, for any atom $i \in$ P, the VDW radius, R_i^*, and well depth $\in_i$ are defined by

$$R_i^* = \lambda R_i^{*A} + (1 - \lambda)R_i^{*B} \qquad (16)$$

$$\epsilon_i = \lambda\epsilon_i^A + (1 - \lambda)\epsilon_i^B$$

from which the coefficients have been recalculated for a given value of λ. These values are related to the A and B parameters by

$$A_{ij} = \epsilon_{ij} R_{ij}^{*12} \qquad (17)$$

$$B_{ij} = 2\epsilon_{ij} R_{ij}^{*6}$$

where $\epsilon_{ij} = (\epsilon_i^*\epsilon_j)^{1/2}$ and $R^*_{ij} = (R^*_i + R^*_j)$.

Equations 12–17 define completely the coupling of the potential functions involving A and B and have been implemented into the energy and force calculation of the program. Either method may be used for collecting the data. In the windowing technique, one has to define the number of windows, and for each window the number of equilibration steps and the number of data collection steps. The program tabulates the free energy change for each window and also the total accumulated free energies. In the slow growth technique one has to define the conversion time during which the solute is interconverted.

Perturbation Models

Both the slow growth and windowing procedures were used to determine aqueous solvation free energy differences between methanol and ethane, ammonium and oxonium, glycine and alanine, and alanine and phenylalanine. For all the simulations TIP3P[14] water was used.

For the methanol to ethane perturbation the OPLS parameters[15] were adopted and adapted to AMBER.[16] These parameters are given in Table I. The methyl groups were treated as united atoms; only the hydrogens bonded to oxygen were included explicitly. Standard geometries were employed for the monomers: for water, r(OH) = 0.9572 Å, ∠HOH = 104.52°; for ethane, r(C–C) = 1.53 Å; for methanol, r(C–O) = 1.43 Å, r(O–H) = 0.945 Å, and ∠COH = 108.5°. A bond was defined between the two hydrogens in water, r(H–H) = 1.514 Å, the force constant for the bond angle ∠HOH was set equal to zero, and the force constants for the bond lengths of water were set equal to 400 kcal/Å^2. This ensured a rigid but geometrically correct water model for SHAKE.[17] (SHAKE is an algorithem which allows fixed internal geometries during molecular dynamics simulations.) For the continuous growth procedure, an artificial bond was placed between the hydroxy hydrogen and the carbon of methanol to minimize the deviation in the perturbed energy due to angle fluctuations. For the simulation, the oxygen VDW, oxygen charge, and hydrogen charge are slowly perturbed, as λ goes from 1 to 0, into their respective parts in ethane. The O–H bond is not scaled since the hydrogen effectively disappears as the charge goes to zero. What remains at λ = 0 is two united atom CH_3 groups and a "dummy" atom in the place of the hydroxy hydrogen. The force constants associated with "dummy" atoms are set equal to those of the unperturbed state.

For the perturbation of ammonium into oxonium, the parameters adopted are given in Table I.[18] The geometry of the monomers is as follows: ammonium, r(N–H) = 1.022 Å, ∠HNH = 109.45°,[19] oxonium, r(H–O) = 0.976 Å, ∠(HOH) = 110.7°.[20] The molecular charges were determined using the above geometries by the UCSF G80 electrostatic potential routine[21] at the 6-31G* basis set level.[22] The R^* and ϵ values for the nitrogen atom of NH_4^+, (N), and the oxygen atom of H_3O^+, (O), were empirically modified to fit the experimental difference in the heats of hydration and to reproduce the equilibrium distance between the heavy atoms and the water oxygen, (OW). The experimental $\Delta\Delta H$ between $NH_4^+(H_2O)_3$ and $OH_3^+(H_2O)_3$ is −23.7 kcal/mol,[23] while our calculated value is −24.7 kcal/mol. The average ab initio distance for N–OW is 2.72 Å and that for O–OW is 2.50 Å.[24] The calculated values with AMBER for the N–OW and O–OW distances are 2.74 and 2.47 Å, respectively. The calculated ΔE for the interaction of oxonium with 3 waters is −84.4 kcal/mol, while the experimental value is −69.0, but the calculated value is very similar to what Kochanski found in a Monte-Carlo simulation.[25] The error in the calculated ΔE for the interaction of three waters with oxonium is likely due to the neglect of three and higher body interactions in the calculations, as suggested by Kochanski. In this simulation Lennard–Jones potentials are only given to the heavy atoms of each ion (N and O). The hydrogens are included as locations for the partial charges. As ammonium is perturbed into oxonium, the charge on one of the hydrogens is slowly lowered to zero, which becomes a "dummy" atom, (DU).

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

(15) (a) For ethane: Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489. (b) For methanol: Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 1276.

(16) For AMBER the OPLS parameter σ must be converted to R^*. σ is the Lennard–Jones diameter and R^* the van der Waals radius of a given atom (i.e., half of the position of the minimum in the van der Waals curve for a given atom interacting with a similar atom type). They are related by $R^* = 2^{1/6}(\sigma/2)$.

(17) van Gunsteren, W. F.; Berendsen, H J. C. *Mol. Phys.* **1977**, *34*, 1311–1327.

(18) After the ammonium and oxonium parameters were determined and used we received OPLS parameters for NH_4^+ from Professor Jorgensen. The values for N correspond to q = −0.40, R^* = 1.95, and ϵ = 0.170. Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174.

(19) Yamaguchi, Y.; Schaefer, H. F., III *J. Chem. Phys.* **1980**, *73*, 2310.

(20) Bunker, P. R.; Amano, T.; Spirko, V. *J. Mol. Spectrosc.* **1984**, *107*, 208.

(21) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129.

(22) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *20*, 213.

(23) (a) Kebarle, P. *Annu. Rev. Phys. Chem.* **1972**, *28*, 445. (b) Lau, Y. K.; Ikuta, S.; Kebarle, P. *J. Am. Chem. Soc.* **1982**, *104*, 1462. (c) Payzant, J. D.; Cunningham, A. J.; Kebarle, P. *Can. J. Chem.* **1973**, *51*, 3242.

(24) Ikuta, S. *Mass Spectrosc.* **1982**, *30*, 297.

(25) Kochanski, E. *J. Am. Chem. Soc.* **1985**, *107*, 7869.

Table II. The STO-3G Charges for the Perturbed Amino Acids[a]

group	atom	charge	group	atom	charge	group	atom	charge
glycine	O1	−0.624	alanine	O1	−0.618	phenylalanine	O1	−0.604
	C1	0.573		C1	0.535		C1	0.569
	O2	−0.650		O2	−0.656		O2	−0.665
	CA	0.044		CA	0.229		CA	0.054
	HA1	0.052		HA1	−0.014		HA1	0.015
	HA2	0.056		CB	0.008		CB	0.223
	N	−0.264		N	−0.591		CG	−0.483
	HN1	0.289		HN1	0.352		CD1	0.233
	HN2	0.306		HN2	0.379		CE1	−0.025
	HN3	0.306		HN3	0.376		CZ	0.042
							CEZ	−0.030
							CD2	0.093
							N	−0.490
							HN1	0.352
							HN2	0.346
							HN3	0.370

[a] The geometry of the atoms is given in ref 27; in each case O1, C1, O2 is the carboxyl group and N, HN1, HN2, and HN3 the ammonium group, and the other atoms are named according to standard amino acid nomenclature given in ref 27.

Table III. The Procedure Used for the Simulations of the Ammonium–Oxonium and Methanol–Ethane Perturbations

			windowing procedure[a]					slow growth[b]		
simulation	starting/ ending λ	Δλ[c]	no. of steps[d] for equil	step size[d], ps	no. of steps[e] for data	step size,[e] ps	ΔG, kcal/mol forward/backward[f]	no. of steps	step size, ps	accumulated ΔG
1	1.0/0.5	0.05	500	0.002	2000	0.0005	7.08/−6.86			
	0.5/0.0	0.10	500	0.002	2000	0.0005				
2	1.0/0.0	0.005	300	0.002	100	0.002	6.94/−6.40			
3	1.0/0.98							1000	0.002	0.71
	0.98/0.94							1000	0.002	1.82
	0.94/0.875							2000	0.002	3.21
	0.875/0.75							2000	0.002	6.12
	0.50/0.18							1000	0.002	6.93
	0.18/0.0									
4	1.0/0.0	0.05	500	0.002	2000	0.002	−19.8/20.2			
5	1.0/0.0	0.05	250	0.002	1000	0.002	−20.6/20.7			
6	1.0/0.0	0.05	200	0.002	200	0.002	−21.6/21.9			
7	1.0/0.0							10,000	0.002	−21.1

[a] Using eq 7 to calculate the total ΔG. [b] Using eq 9 to calculate the total ΔG. [c] Incremental change in λ. [d] Number of steps of equilibration at a given λ and size of the molecular dynamics time step for these. [e] After equilibration, the number of data collection points and the molecular dynamics time step for these. [f] Calculated free energy forward corresponds to going from starting → ending λ; backward is the reverse.

The other three hydrogens assume the charge and geometry of the oxonium hydrogens. The "dummy" atom does not have its H–N bond scaled, but the Θ_0 for the DU–O–H angle is 108.5°.

For the amino acid simulations standard AMBER[26] parameters are adopted for the solute, and the solvent is TIP3P water. As noted elsewhere,[26] the 10–12 parameters are used for solute–solute interactions, but not for water–water or water–solute interactions. The amino acid monomers were represented as zwitterions. The charges were determined[21] with the STO-3G basis set with use of crystallographic geometries,[27] and are given in Table II. For the glycine to alanine perturbation a hydrogen is changed into a united atom CH_3 group. The perturbation of alanine to phenylalanine is more demanding. This is accomplished by starting with a greatly reduced phenyl ring, one with C–C bond lengths of 0.2 Å. The C–C bond between C_β and the phenyl group is also reduced to 0.5 Å. At λ = 1 the carbons in the phenyl group have Lennard–Jones parameters set equal to zero; effectively they are "dummy" atoms. As the simulation runs from λ = 1 to 0, the Lennard–Jones parameters and bond lengths slowly increase, allowing for a smooth introduction of the phenyl group.

For the amino acid simulations, 8–14, we consider the entire amino acid as the "perturbed group". Thus, when we change, e.g., alanine to phenylalanine, any changes in the internal energy of the solutes were not included in the free energy change. This was done to avoid any artifacts from interactions between the terminal groups of the amino acids and the side chains. In all the simulations in aqueous solution, an 8 Å molecular based cutoff was used.

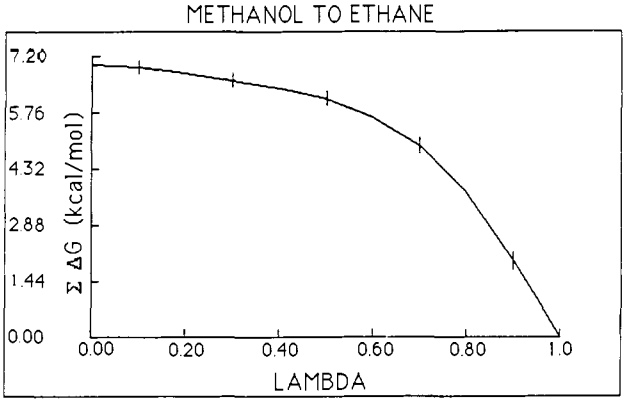

Figure 1. A plot of the average accumulated free energy change vs. λ for the perturbation of methanol (λ = 1) to ethane (λ = 0) for simulation 1. The bars represent the fluctuations at that step.

Computational Procedure and Results

The perturbation of methanol into ethane (λ = 1 → 0) was studied with both the slow growth and windowing procedures, and these perturbations are given as simulations 1–3 in Table III. Simulation 1 was run with 42 500 steps for a total simulation time of 34 ps. The calculation was conducted by employing 214 water molecules in a cubic box at constant pressure (1 atm) and temperature (300 K) and scaling R^* and ε. It was clear from Jorgensen's results[2] that the greatest change in free energy was in

(26) Weiner, S. J.; Kollman, P.; Case, D.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(27) (a) Glycine: Jonsson, P.-G.; Kuick, A. *Acta Crystallogr.* **1972**, *B28*, 1827. (b) L-Alanine: Simpson, H. J., Jr.; Marsh, K. E. *Acta Crystallogr.* **1966**, *20*, 550. (c) L-Phenylalanine: Al-Karaghouli, A. R.; Koetzle, T. F. *Acta Crystallogr.* **1975**, *B31*, 2461.

Table IV. The Procedure Used for the Simulations of the Amino Acid Perturbations

			windowing procedure[a]					slow growth[b]		
simulation	starting/ ending λ	Δλ[c]	no. of steps[d] for equil	step size,[d] ps	no. of steps[e] for data	step size,[e] ps	ΔG, kcal/mol[f] forward/backward	no. of steps	step size, ps	accumulated ΔG
8	1.0/0.0	−0.05	1000	0.001	1000	0.001	1.18–1.14			
9	1.0/0.7	0.10	1000	0.002	4000	0.0005	0.39/−0.29			
	0.7/0.0	0.20	2000	0.002	8000	0.0005	0.70/−0.65			
10								60000	0.001	0.76
11	1.0/0.75	0.01	100	0.001	100	0.001	−0.66/0.62			
	0.75/0.0	0.01	100	0.002	100	0.001	−1.93/2.23			
12	0.0/0.25	0.025	250	0.002	250	0.001	1.89/−1.36			
	0.0/0.25	0.01	100	0.002	100	0.001	1.53/−1.22			
	0.25/1.0	0.025	250	0.002	250	0.001	1.76/−1.61			
13	1.0/0.0							63000	0.001	−2.29
14	0.0/0.25							15750	0.001	1.61
	0.0/0.25							31500	0.001	1.40
	0.25/1.0							47250	0.001	2.44

[a] Using eq 7 to calculate the total ΔG. [b] Using eq 9 to calculate the total ΔG. [c] Incremental change in λ. [d] Number of steps of equilibration at a given λ and size of the molecular dynamics time step for these. [e] After equilibration, the number of data collection points and the molecular dynamics time step for these. [f] Calculated free energy forward corresponds to going from starting → ending λ; backward is the reverse.

the initial stage of the perturbation (λ near 1). To ensure adequate sampling at the beginning of the calculation a large number of steps were used, in fact, 66% of the simulation time was conducted over the first Δλ = 0.5. The calculated free energy changes for the forward and backward directions were 7.08 and −6.86 kcal mol, respectively, for an average absolute free energy change of 6.97 kcal/mol. The experimental free energy change is 6.93 kcal/mol.[28] A plot of the accumulated free energy vs. λ (λ = 1 → 0) is given in Figure 1. Simulation 2 was conducted under the same conditions given above, but no weighting was introduced and fewer steps (8400) were used. The outcome was a change in free energy in the forward and backward direction of 6.94 and −6.40 kcal/mol, respectively, resulting in an average absolute ΔG of 6.67 kcal/mol. Considering the shorter computational time needed for this calculation and the fact that equal time was spent at each interval of λ, the amount of hysteresis in the forward and backward energies is still rather small. The slow growth procedure was conducted by employing 215 waters and a rectangular box at constant pressure (1 atm) and temperature (300 K) and scaling the A and B parameters. In a third simulation, more steps were performed in the initial stages of the simulation than in the latter part, with 50% of the simulation time consumed over the first Δλ = 0.25. The result was an accumulated free energy change of 6.93 kcal/mol. Plots of the accumulated free energy change vs. λ for simulations 1–3 are given in Figure 2.

The perturbation of ammonium into oxonium was performed in the gas phase and in solution with use of both the slow growth and windowing techniques. Initially, the gas-phase study of $NH_4(H_2O)_3^+ \rightarrow OH_3(H_2O)_3^+$ was conducted in order to check the parameters. Simulation 4 used a total of 52 500 steps for a total simulation time of 42 ps and was performed in less than 1 min on a Cray-XMP. The outcome was an average free energy change of −20.0 kcal/mol, and the experimental vallue is −21.3 kcal/mol.[23] Since the calculated change in free energy was in reasonable agreement with experiment, we carried out a simulation of $NH_4^+(aq) \rightarrow H_3O^+(aq)$.

The solution studies, simulations 5–7, used two windowing procedures and one slow growth procedure. For these simulations and all others which are conducted in solution, the simulation was run at constant pressure (1 atm) and temperature (300 K) and scaling R^* and ϵ. These simulations were run with 215 water molecules starting with an 18.77 × 18.77 × 18.77 Å box. A plot of the change in free energy vs. λ for all three of these simulations is given in Figure 3. The accumulated change in free energy for all of these simulations lies between −20.6 and −21.8 kcal/mol. Experimentally the enthalpy of solvation has been estimated to be between 75 and 81 kcal/mol for NH_4^+ and approximately 100 kcal/mol for OH_3^+.[29] Thus, the $\Delta\Delta H_s$ is roughly 19–25 kcal/mol,

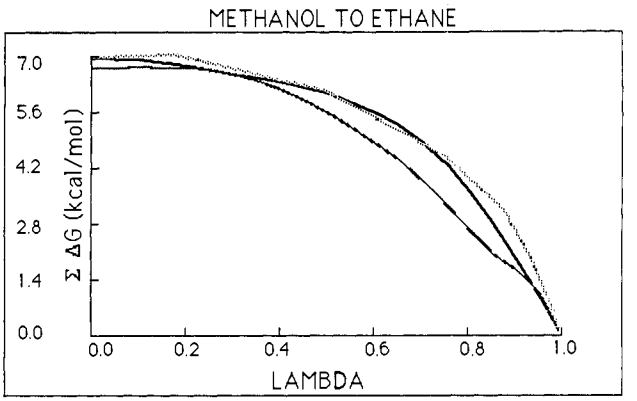

Figure 2. A plot of the average accumulated free energy change vs. λ for the perturbation of methanol (λ = 1) to ethane (λ = 0) for simulation 1 (bold), simulation 2 (dotted), and simulation 3 (dashed).

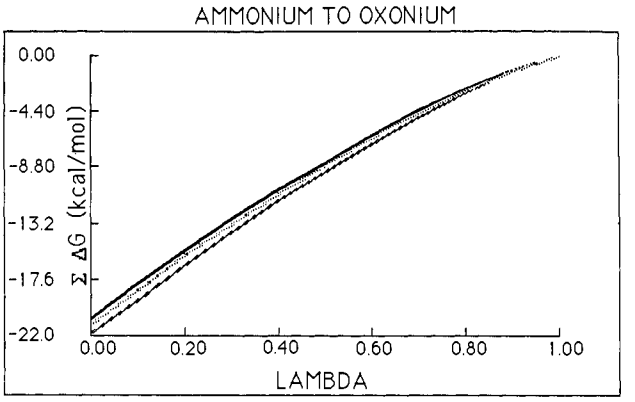

Figure 3. A plot of the average accumulated free energy change vs. λ for the perturbation of ammonium (λ = 1) to oxonium (λ = 0) for simulation 5 (bold), simulation 6 (dotted), and simulation 7 (dashed).

and as in the gas-phase studies the $\Delta\Delta G_s$ would be expected[23] to be about 2–3 kcal/mol less than the $\Delta\Delta H_s$. Our calculated values are within the estimated experimental value. These simulations again show that both the slow growth and windowing procedures do equally well.

The success of these perturbations was very exciting; however, it is the goal of the research to be able to apply this technique to macromolecular systems, especially proteins and nucleic acids. Thus, the perturbation of single amino acids into others was undertaken to determine if sensible differences in the free energy of solvation could be obtained for large changes in molecular structure. We chose to perturb alanine into glycine and phenylalanine because these represent both a small and a much larger perturbation and we wished to determine whether the method was

(28) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2106.
(29) Aue, D. H.; Webb, H. M.; Bowers, M. T. *J. Am. Chem. Soc.* **1976**, *98*, 318.

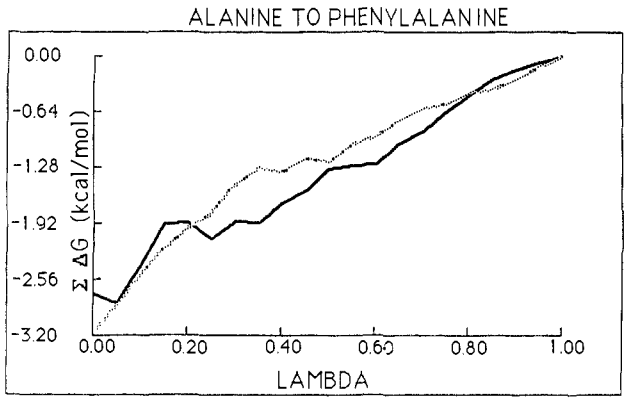

Figure 4. A plot of the average accumulated free energy change vs. λ for the perturbation of alanine (λ = 1) to phenylalanine (λ = 0) for simulation 11 (bold) and simulation 12 (dotted).

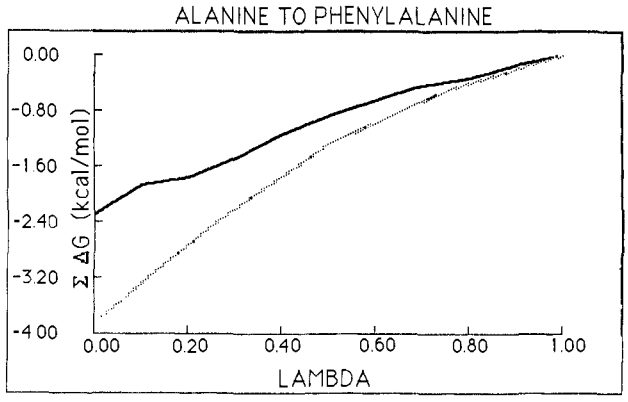

Figure 5. A plot of the average accumulated free energy change vs. λ for the perturbation of alanine (λ = 1) to phenylalanine (λ = 0) for simulation 13 (bold) and simulation 14 (dotted).

capable of working on such changes. The results for the amino acid simulations are given in Table IV.

Simulations 8–10 represent the perturbation of alanine into glycine, with simulations 8 and 9 using the windowing procedure and simulation 10 using the slow growth procedure. These simulations were run with 367 water molecules starting with a 22.8 × 22.1 × 21.0 Å box. The outcome of simulations 8 and 9 was an average absolute free energy change of 1.16 and 1.01 kcal/mol, respectively. Simulation 10 empoyed the slow growth procedure and used 60 000 steps for a total simulation time of 60 ps. This is more than 15 ps longer than either windowing procedure and was the most extensive calculation performed. The result of this calculation was an accumulated free energy difference of 0.76 kcal/mol. As noted by Jorgensen and Ravimokan,[2] incomplete sampling would tend to lead to a free energy change too positive; thus, it is not surprising that the slow growth value of ~0.8 kcal/mol is smaller than the value of ~1.1 kcal/mol in the two windowing calculations. Still, the variation in calculated free energies of ~0.3 kcal/mol is comparable to that found above for the methanol-to-ethane perturbation.

Simulations 11–14 represent the perturbation of alanine into phenylalanine. These calculations employed 545 water molecules starting with a 23.6 × 26.9 × 25.6 Å box. For simulations 11 and 12 the windowing procedure was used, and for simulations 13 and 14 the slow growth procedure was used. To check for path dependence simulations 11 and 13 were run from λ = 1 to 0, and simulations 12 and 14 were run in the reverse direction. The average absolute free energy change for simulations 11–14 is −2.72, 3.18, −2.29, and 3.84 kcal/mol, respectively. The average of the window and slow growth procedures is 2.95 and 3.07 kcal/mol, respectively, which gives an overall average of 3.01 kcal/mol. During simulations 12 and 14 it was noticed that the free energy changed rapidly from λ = 0.0 to 0.25. To test step size dependence the calculation was restarted at λ = 0.0 and run to λ = 0.25. The step size was reduced but the total number of steps was maintained for simulation 12, while the number of steps were doubled for simulation 14 (see Table IV). The restarts had only a slight affect on the change in free energy, but both show a reduction in the change in free energy of ~0.2 kcal/mol. Plots of the average absolute accumulated change in free energy vs. λ for simulations 11–12 are given in Figure 4 and simulations 13 and 14 are given in Figure 5.

Although there are no experimental data for the free energies of solvation of glycine, alanine, and phenylalanine in water, Wolfenden[30] has suggested that "model" systems for the side chains are reasonable. Although he suggests that the relative solvation free energy of H_2 and CH_4 is a good model for the relative free energy of glycine → alanine, it may be that the model of methane → ethane is closer to reality. For the relative solvation free energy of alanine and phenylalanine the model of methane to toluene seems appropriate.[30] It should be noted that our calculated free energies for the relative solvation of glycine vs. alanine are −0.76 to −1.16 kcal/mol, compared to −0.2 kcal/mol for experiments on the $CH_4 \rightarrow CH_3CH_3$ "model"[28] and −0.3 kcal/mol for experiments on relative solvation of propane and isobutane.[28] Our calculated free energies of alanine → phenylalanine are −2.29 to −3.84 kcal/mol compared to the experimental value of −2.7 kcal/mol for the methane to toluene "model".[28] The calculated numbers are certainly in qualitative agreement with experiments on the model systems, but we stress that our focus here has not been to try to reproduce those but to show that the free energy perturbation approach can be reasonably applied to "large" perturbations.

Nonetheless, it is interesting that the calculated free energy for glycine to alanine is noticeably more negative than the experimental free energy of the methane-to-ethane model. This may be a real difference, because there are two components that dominate the free energy of hydration of nonpolar groups:[1,10,31] The first is the cavity term, which is unfavorable when one increases the size of the group, e.g., in going from methane to ethane, and the second is the change in dispersion attraction,[31] which is favorable for the change of methane to ethane. It is likely that for the change of glycine to alanine where the CH_2 group may be partially shielded from solvent, the cavity term is less than that for methane to ethane, where the group is fully exposed. It is also encouraging that the free energy change for alanine to phenylalanine is calculated to be closer to the value for the model methane to toluene. This is reasonable because the phenyl side chain in the amino acid is nearly completely exposed to solvent, as in the model system, toluene.

Discussion

The free energies calculated here are generally in good agreement with available experimental data. The success of these calculations and others[1-5] implies that this procedure will be applicable to a variety of studies in many areas of chemistry. Elsewhere, we present a calculation on a protein–inhibitor interaction, which confirms the power of this approach for studying free energy changes in large macromolecular systems.[10] The variety of systems perturbed and the number of simulations conducted for each system has taught us that the approach to any simulation is governed by the size of the structural and electrostatic change in the system. For the perturbation of methanol into ethane the change in the free energy of solvation is almost completely due to electrostatic effects. In this perturbation the tightly coordinated waters associated with the hydroxy group must become reorganized when the hydroxy group is perturbed into the methyl group. Thus, the water structure must reorganize, which requires an adequate simulation time to ensure sufficient sampling of conformational space. It is important to equilibrate for sufficiently long periods of time when the electrostatics are changing the most.

(30) Wolfenden, R.; Anderson, L.; Cullis, P. M.; Southgate, C. C. B. *Biochemistry* **1981**, *20*, 849.

(31) Straatsma, T.; Berendsen, H.; Postma, J. *J. Chem. Phys.*, in press.

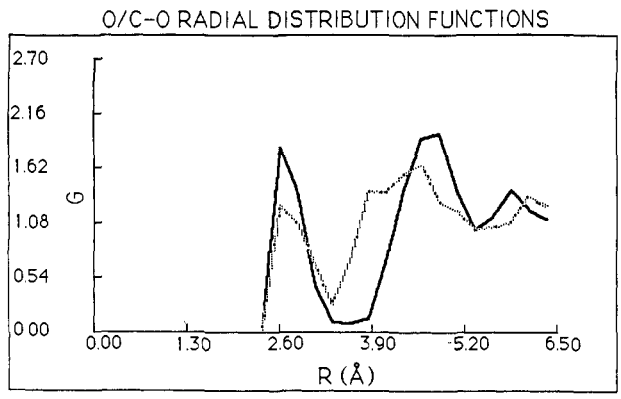

Figure 6. The radial distribution of the hydroxy oxygen–water oxygen distance at $\lambda = 0.75$ for simulation 1 (dotted) and simulation 2 (bold).

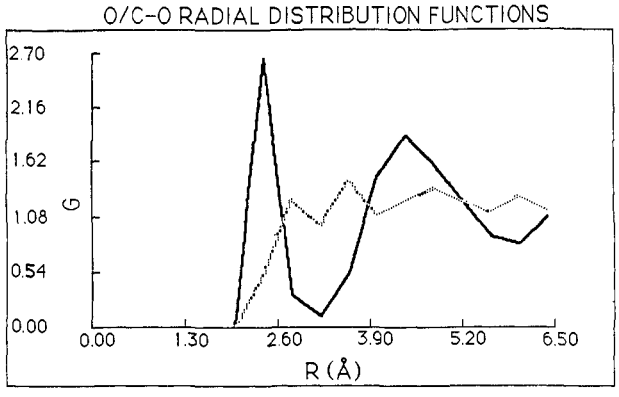

Figure 7. The radial distribution of the hydroxy oxygen–water oxygen distance for the equilibrations at $\lambda = 1.0$ (bold) and at $\lambda = 0.75$ (dotted).

This is evident from the methanol-to-ethane calculations in which the majority of the calculation is performed in the first $\Delta\lambda = 0.5$. This weighting ensures that the solvent reorganization will be more complete and that the predicted change in free energy for this value of λ will better represent the "true" change. For the nonweighted simulation, 2, the solvent is not reorganized adequately in the initial stages of the calculation; therefore, the predicted change in free energy is too small. The radial distribution functions (rdfs) between the oxygen of water and the atom that is transformed from oxygen in methanol to carbon in ethane at $\lambda = 0.75$ for simulations 1 and 2 are given in Figure 6, and the rdfs for a 10 ps equilibration at $\lambda = 0.75$ and a 4 ps equilibration at $\lambda = 1.0$ are given in Figure 7. The rdfs for the equilibrated runs, those given Figure 7, are similar to those reported by Jorgensen[2] for $\lambda = 0$ and 0.25 (Jorgensen set methanol as $\lambda = 0$ and ethane as $\lambda = 1$, which is the opposite of our convention) despite the fact that our rdfs have a high noise level because of the small number of samples available. The 10-ps equilibration at $\lambda = 0.75$ represents the "fully" equilibrated structure at $\lambda = 0.75$. The radial distribution for the 10-ps equilibration at $\lambda = 0.75$ indicates that the tightly coordinated first solvation shell is lost and the second shell is not as distinct as it is for the radial distribution at $\lambda = 1.0$. The radial distribution for simulation 1 resembles the 10-ps equilibration at $\lambda = 0.75$ more than does simulation 2. The radial distribution for simulation 2 shows that the first solvation shell is $\sim 1/2$ what it was at $\lambda = 1$, but the second solvation shell is intact. Thus, the water has not completely reorganized in the second simulation, which is reflected by the fact that the free energy change at this point is lower than in simulation 2. However, since the change in free energy decreases dramatically for the second half of the calculation, simulation 2 has a chance to obtain the proper solvent–solute configuration by $\lambda = 0.30$, and the final change free energy only differs by approximately 0.3 kcal/mol.

For simulations in which there is a large structural change, especially the alanine to phenylalanine system, the most important factor is the size of $\Delta\lambda$. This is clear from the results of the windowing procedure in which the smaller $\Delta\lambda$ values gave a better result for $\lambda = 0.25 \rightarrow 0.0$ in simulation 11 and from the slow growth calculations which consistently gave good results. This is due to the fact that statistics are taken for $\lambda \pm \Delta\lambda$. At λ and $\Delta\lambda$, where $\Delta\lambda$ is large, there will be a large VDW interaction caused by the increase in size of the perturbed group interacting with the solvent and at $\lambda - \Delta\lambda$ there will be a gap between the solvent and the perturbed group with both effects destabilizing.

The VDW interaction scaling procedure used is also of importance when conducting a large structural change. Two scaling procedures were implemented, A,B and R^*,ϵ. The R^*,ϵ method scales the radii according to eq 16, whereas the A,B method scales the VDW interaction energy according to eq 15. A simple example to demonstrate the difference in using these two procedures is to place a hydrogen and carbon atom at different distances from each other and then change the hydrogen into a carbon atom. The VDW interaction energy is then calculated on the basis of either A,B or R^*,ϵ scaling as a function of λ ($\lambda = 0$ corresponds to a hydrogen–carbon VDW interaction and $\lambda = 1$ a carbon–carbon interaction) and is given in Figure 8. It is clear that A,B scaling has a larger slope than the R^*,ϵ scaling. Thus, as the nonbond energies are determined for $\lambda \pm \Delta\lambda$ the energies will be much larger for the A,B scaling which can lead to large errors in the calculation. A,B scaling will work when $\Delta\lambda$ is extremely small and the structural change is small, as in the initial stages of the methanol–ethane slow growth perturbation where $\Delta\lambda = 0.00002$. The R^*,ϵ scaling allows for a smooth introduction of the perturbing group regardless of whether the windowing or slow growth procedure is used; thus, R^*,ϵ scaling and the use of a linear dependence of λ appear to be computationally more convenient.

The results for both the "slow growth" and "windowing procedures" to calculate the free energy differences are similar. However, the windowing procedure allows the enthalpy differences to be obtained at each interval, which may facilitate the understanding of the origin of the energy differences between states A and B. The windowing procedure also offers a greater assurance that the statistics taken are for the lowest energy configuration due to the initial equilibration at each window, and statistical errors in the forward and backward directions are easily noticed, whereas, slow growth must be checked for hysteresis by running the calculation in the opposite direction. However, the slow growth technique gives a better understanding of systematic error. From the alanine ($\lambda = 1$) to phenylalanine ($\lambda = 0$) studies it is clear that the Hamiltonian lags behind the perturbation and in each simulation the change in the free energy is too positive as pointed out by Jorgensen and Ravimohan.[2] For the $\lambda = 1 \rightarrow 0$ simulation the energy associated with formation of the cavity needed for the growing phenyl group is too high and for the reverse run the loss of dispersion attraction is overestimated. However, when the two simulations are averaged these effects should cancel out. The analysis of systematic error is more difficult for the windowing procedure because there would be a systematic error at each window that would not necessarily be related to the previous window.

The estimate of the overall statistical error for the simulations is another concern. When one uses many windows in the windowing procedure, as has been done here, a simple summation of the fluctuations found at each window for the windowing procedure gives erroneously large statistical errors because the fluctuation at each windows is often of the same magnitude as the change in free energy. Furthermore, Straatsma et al.[32] have recently suggested that the estimate of the statistical error should be obtained from the deviation in the average over subseries that are considered uncorrelated. They state that the estimation of the statistical error is too low as usually determined by Monte Carlo and molecular dynamic simulations because the statistical error is being calculated from a highly correlated series. To implement such an approach in an analysis of statistical errors, many more steps per window would be necessary and such an approach cannot be used at all with the slow growth method.

(32) Straatsma, T. P.; Berendsen, H. J. C.; Stam, A. J. *Mol. Phys.* **1986**, *57*, 89.

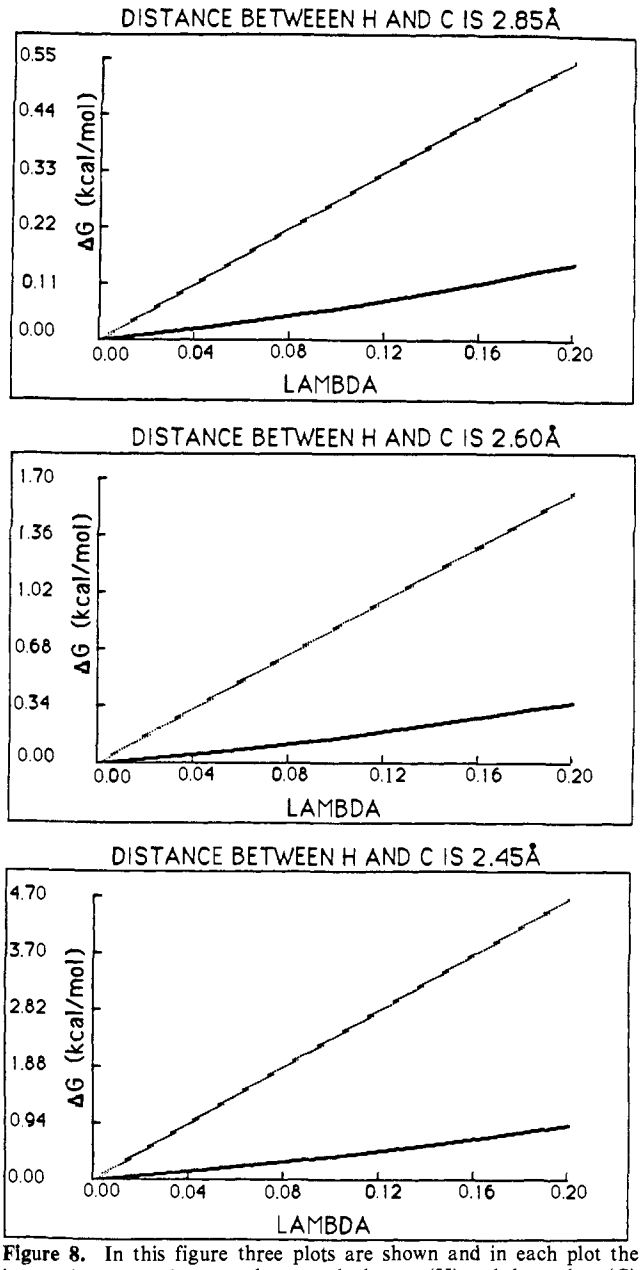

Figure 8. In this figure three plots are shown and in each plot the interaction energy between the perturbed atom (H) and the carbon (C) atom is given as a function of $\Delta\lambda$. Within each plot the interaction energy for R^* and ϵ scaling is given by a bold rule and for A,B growth a dotted rule. The different distances between H and C for the three plots represent three possible distances that could be found during the dynamics, and it is quite clear that as the distance decreases the difference between the two methods increases.

Thus, this type of error estimation does not seem suitable for these simulations. In our opinion the best true estimate of the statistical errors in the calculations is given by the standard deviation found in independent simulations. These are presented in Table V for the systems studied here. One could easily get a lower standard deviation for the alanine to phenylalanine perturbation if the slow growth runs (numbers 13 and 14) were extended to a time of 100 ps or longer. Such runs are computationally feasible (10 ps takes ~1.5 h on the Cray XMP48) but in our opinion are not worth doing unless the scientific question addressed requires the lower statistical uncertainty.

It is probably worth summarizing why the method works so well and what are its limitations. The fact that eq 6 involves only differences in interaction involving the groups that change going from system A to system B (eq 3) is the reason for the relatively low statistical error inherent in the method. Thus, for relative solvation free energies, one can focus on the change in solute–solvent interactions, which converge much faster[33] than the changes in solvent–solvent energies required to evaluate changes in enthalpy of solvation. The fundamental considerations in the application of the method are first ensuring that the Hamiltonians used are correct and second obtaining statistics and samplings which are adequate to evaluate the free energy term in brackets in eq 6. The first consideration often requires that one find appropriate parameters on representative model systems similar to the system of interest. This allows one to use the accuracy of the perturbation method to improve and refine intermolecular potential functions, as we[11] and others[34] are currently doing. The second consideration requires careful analysis of each system. Our criteria for adequacy in sampling is the amount of statistical error found in independent simulations for the same systems. Each type of problem will have its own sampling limitations, but as the power of computers grows, more and more of these problems can be overcome.

Table V. Summary of Free Energy Calculations

system	simulations[a]	ΔG values[b]	av ΔG[c]
MeOH(aq) → ethane(aq)	1	6.93	6.86 ± 0.12
	2	6.67	
	3	6.97	
$NH_4^+(H_2O)_3$ → $H_3O^+(H_2O)_3$	4	−20.0	−20.0
NH_4^+(aq) → H_3O^+(aq)	5	−20.6	−21.2 ± 0.4
	6	−21.8	
	7	−21.1	
Ala(aq) → Gly(aq)	8	1.16	0.98 ± 0.18
	9	1.01	
	10	0.76	
Ala(aq) → Phe(aq)	11	−2.72	−3.01 ± 0.50
	12	−3.18	
	13	−2.29	
	14	−3.84	

[a] See Tables III and IV for details on the simulations. [b] Calculated free energies in kcal/mol (1 kcal = 4.184 kJ). [c] Average and standard deviation of free energies in kcal/mol.

Conclusion

We have presented a general methodology for carrying out calculations to determine free energy changes in complex molecules using molecular dynamics and have applied it to a number of molecules. The results for both gas phase and solution are in encouraging agreement with available experiments. Thus, it is likely that the approach used here will be a versatile and useful method for studying a wide variety of chemical and biochemical systems.

Acknowledgment. We are pleased to acknowledge support from the NIH (GM-29072) and computer time through the NSF (DMB) to P.A.K. and grants awarded to P.A.B. and R. Langridge from the San Diego Supercomputer Center (SDSC). We thank Doug Hutt for Boeing Computer Services and R. Hilderbrant of SDSC for their assistance in using the respective facilities, as well as the UCSF Computer Graphics Lab, supported by NIH (RR01081), for the use of its facilities. We are grateful to W. van Gunsteren for the chance to run a comparative simulation of the methanol–ethane perturbation with GROMOS, and P.A.K. is grateful to van Gunsteren and H. Berendsen for their hospitality during his sabbatical year, 1985–1986, where he learned much about the perturbation approach. P.A.K. is also grateful to CECAM for sponsoring a Free Energy Workshop, organized by van Gunsteren and A. McCammon in September 1985. At that meeting, A. Cross noted the advantages of not using A and B scaling in the nonbonded interactions.

Registry No. Methanol, 67-56-1; ethane, 74-84-0; ammonium, 14798-03-9; oxonium, 13968-08-6; glycine, 56-40-6; alanine, 56-41-7; phenylalanine, 63-91-2.

(33) Chandresekhar, J.; Spellmeyer, D.; Jorgensen, W. J. Am. *Chem. Soc.* **1984**, *106*, 903.
(34) van Gunsteren, W. F.; Kollman, P. A.; Berendsen, H., unpublished results.