Theor Chem Acc (1997) 98:171–181

Theoretical
Chemistry Accounts

Regular Article

Overcoming barriers in macromolecular simulations: non-Boltzmann thermodynamic integration

Nobuyuki Ota[1], **Axel T. Brünger**[1,2]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, 266, Whitney Avenue, New Haven, CT 06520, USA
[2] Howard Hughes Medical Institute, Yale University, New Haven, CT 06520, USA

Received: 16 June 1997 / Accepted: 29 August 1997

Abstract. A new combination of non-Boltzmann sampling (umbrella sampling) and thermodynamic integration for the computation of free energies of flexible systems is described and compared with other methods in terms of its accuracy, efficiency, and viability for simulations of macromolecules. The non-Boltzmann sampling is achieved through the definition of a surrogate potential function that uniformly reduces barriers between energy minima of a given molecule while retaining the overall features of its torsional profile. The influence of the surrogate potential on the free energy calculation is taken into account by a correction similar to umbrella sampling. This work is tested by computation of the free energy of hydration difference between butane and propanol. The free energy difference converges to the experimentally measured value in one tenth of the computer time required for a simulation employing conventional thermodynamic integration without the umbrella potential. Moreover, the deviations of the computed free energies for similar sampling periods are greatly reduced with the new method. Finally, an accurate free energy difference is obtained in one third of the number of simulations that are necessary for the multi-state isomeric method.

Key words: Free energy calculation – Umbrella sampling – Thermodynamic integration – Non-Boltzmann sampling

1 Introduction

The probability that a molecular system of interest will be found in one state or another is given by the free energy difference between those states, such that biochemical events like binding and folding can be expressed in terms of free energies. Among the methods to compute free energy differences of molecules of known structure are free energy perturbation (FEP) [1] and thermodynamic integration (TI) [2]. Both methods have been used to compute accurately free energy differences of small model compounds and fairly rigid systems [3–6]. However, for flexible macromolecules, this accuracy is compromised because of insufficient sampling of the many minima on the potential energy surface [7, 8]. To remedy this situation, umbrella (or non-Boltzmann) sampling techniques [9] were introduced to augment FEP or TI. Here, the potential energy function is replaced with a surrogate potential that shifts the distribution of molecular states such that sampling a wider range of the potential energy surface is possible. The free energy averages computed from this surrogate potential are then corrected to obtain their true values. The biasing potential must be defined such that it is different enough to enhance sampling of relevant conformational space but similar enough to avoid convergence problems. Two early studies employed this methodology to study solvent effects on the rotational barrier in butane: one set the torsional potential about the central C—C bond to zero [10–12], while the other chopped the barrier heights but maintained the troughs through the introduction of a penalty function [13]. While these efforts were successful in terms of both accuracy and convergence, this approach has not been used generally for systems with multiple torsions, presumably owing to the difficulties associated with defining a surrogate potential general enough to cover all the rotamers effectively.

A derivative of this approach, importance sampling [14–19], has enjoyed some attention as well. In this case, several simulations, where a restraining potential is introduced to explore more of the conformational space in a specific region of the potential energy function, are connected to cover the entire profile. This method has been applied successfully for cases where relatively few intramolecular degrees of freedom are considered, e.g. alaninedipeptide [14], 1,2-dimethoxyethane [15], and 18-crown-6 [16]. Nonetheless, application of this method to larger and more flexible systems becomes difficult because of the explosion of substates of the system [8].

Correspondence to: A.T. Brünger

Mark et al. described a combination of non-Boltzmann sampling and TI which consists of three stages [20]. First, the surrogate potential is introduced; then the corresponding free energy difference is evaluated. Second, the actual conformational change or mutation is performed. Third, the surrogate potential is slowly eliminated.

Here, a new combination of non-Boltzmann sampling and TI is described, referred to as NBTI. In contrast to the method of Mark et al., it consists of a single stage. In order to reduce barriers across the board, the biasing potential is generated by direct modification of the force field prior to the simulation. In particular, the dominant term in the Fourier series describing each torsion angle is scaled down by a factor of four, and the non-bonded 1,4 interactions are scaled down by five in order to encourage barrier crossings. The advantages of this choice of surrogate potential over setting it to zero are that the barriers are reduced while the overall features of the torsional profile are retained. Moreover, in contrast to using a ceiling of, for instance, 1–3 kcal/mol for all torsions, this uniformly lowers barrier heights but also preserves the character of the torsion angle profile. Molecular dynamics calculations with such a modified energy function should, then, allow better sampling of conformational space while preserving the integrity of the system. The influence of this surrogate potential on the free energy is corrected by a procedure commonly used in umbrella sampling calculations.

We have tested NBTI by calculating the relative free energy difference in hydration when mutating butane to propanol. To estimate the free energy difference, it is required to properly sample the rotamer distribution of both molecules. This is not a simple task because the distribution of rotamers of butane is different from that of propanol. With the appropriate surrogate potentials, NBTI shows faster convergence and better sampling than conventional TI.

2 Method

2.1 Theory of NBTI

The computation of free energy differences between two states using TI involves mapping the Hamiltonian of the system from an initial state, A, to a final state, B. This is conveniently expressed through the definition of a coupling parameter, λ, where $\lambda = 0$ corresponds to state A and $\lambda = 1$ describes state B. The Helmholtz free energy is then given by

$$\Delta A = \int_0^1 \left\langle \frac{\partial E(X^N, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{1}$$

where $\langle\ \rangle_\lambda$ indicates an NVT ensemble average governed by $E(\lambda)$.

Torrie and Valleau have devised a procedure for extracting a Boltzmann-weighted ensemble average from non-Boltzmann sampling, also referred to as an umbrella sampling [9]. The improved sampling of conformational space can be achieved by a simulation with the surrogate potential, E_s, whose barrier potentials are reduced by ΔE in order to facilitate the sampling of all the possible configurations of the molecular system of interest.

$$E_s(X^N, \lambda) = E(X^N, \lambda) - \Delta E(X^N, \lambda) \tag{2}$$

The ensemble average $\langle Q \rangle$ of any property Q of the original system is then given as

$$\langle Q \rangle = \frac{\langle Q(X^N, \lambda) \exp[-\beta \Delta E(X^N, \lambda)] \rangle_S}{\langle \exp[-\beta \Delta E(X^N, \lambda)] \rangle_S} \tag{3}$$

By the direct application of Eq. (3) into the equation of thermodynamic integration, Eq. (1), one can obtain a new exact equation, non-Boltzmann thermodynamic integration (NBTI).

$$\Delta A = \int_0^1 \frac{\left\langle \frac{\partial E(X^N, \lambda)}{\partial \lambda} \exp[-\beta \Delta E(X^N, \lambda)] \right\rangle_S}{\langle \exp[-\beta \Delta E(X^N, \lambda)] \rangle_S} d\lambda \tag{4}$$

The idea of NBTI is that one can estimate the free energy of the original system directly from the simulation using the surrogate potential energy which generates a non-Boltzmann distribution. In principle, the surrogate potential could be any arbitrary function. However, in practice, there could be numerical problems. First, the surrogate potential should preserve the location of minima and lower the barriers between them. Otherwise, the simulation would spend more time in sampling unimportant configurations rather than important configurations.

The barriers are lowered by preparing two different parameter sets. One of the sets corresponds to the original potential energy function of the system, while the other corresponds to the surrogate potential energy. In the case of the mutation from butane to propanol, the surrogate potential is one in which the parameters of the third coefficients of the Fourier series expressing the dihedral angle terms are reduced to one fourth of the original parameters, and the intramolecular van der Waals interactions are reduced to one fifth of the original interactions. As shown in Fig. 1, the barriers around the dihedral angle of butane are reduced with little change to the positions and energy levels of the three minima.

The generation of surrogate potentials is relatively simple for most organic compounds which make use of a Fourier series expansion. In the case of 3-fold dihedral angle potentials, the third Fourier coefficient, V_3, can be reduced as shown below in the mutation from butane to propanol. In the same fashion, the second coefficient, V_2, can be reduced for 2-fold potentials. In addition, van der Waals and/or electrostatic interactions and the first coefficient, V_1, might be reduced for enhancement of *gauche-gauch*e transition. By this procedure, one can easily build up surrogate potentials for each dihedral of larger molecules.

By modification of potential energy parameters, one can generate a gradually changing surrogate potential which is differentiable everywhere as λ goes from 0 to 1. Thus, for configurations generated by the non-Boltzm-

ann simulation, one can calculate the energy difference between the original potential energy and the surrogate potential energy, and the derivative of the original energy with respect to λ in order to obtain the free energy difference, Eq. (4).

Equation (4) can be applied to any segment of a thermodynamic cycle. The relevant thermodynamic cycle [21] for the solvation free energy difference between butane and propanol is given in Fig. 2. The difference in the experimentally measured free energies of transfer of the solutes into water ($\Delta A2-\Delta A1$) is equal to the computationally tractable difference in free energies afforded by mutation of butane to propanol both in gas phase and in water ($\Delta A4-\Delta A3$).

The surrogate potential is used for the computation of both $\Delta A3$ and $\Delta A4$. A coupling parameter, λ, is used

$$E_S(\lambda) = (1-\lambda)E_S(\text{butane}) + \lambda E_S(\text{propanol}), \tag{5}$$

where λ varies from 0 ($E_s = E_s$(butane)) to 1 ($E_s = E_s$(propanol)). In this way, a series of windows connecting the λ values is carried out, and the incremental free energies are accumulated to yield the total free energy difference for the process in the presence of the surrogate potential. Equation (4) is applied to correct for the surrogate potential. The derivatives, $\partial E/\partial\lambda$, the energy differences between the original potential energy and the surrogate potential energy, ΔE, and therefore, $\langle(\partial E/\partial\lambda)\exp[-\beta\Delta E]\rangle/\langle\exp[-\beta\Delta E]\rangle$, are computed, from which the free energy is evaluated as shown below.

$$\begin{aligned}\Delta A &= \sum_i \Delta A_i \\ &= \sum_{i=1}^{n} \frac{\left\langle \frac{\partial E(X^N,\lambda)}{\partial\lambda}\exp[-\beta\Delta E(X^N,\lambda)]\right\rangle_{s,i}}{\langle\exp[-\beta\Delta E(X^N,\lambda)]\rangle_{s,i}}(\Delta\lambda_i)\end{aligned} \tag{6}$$

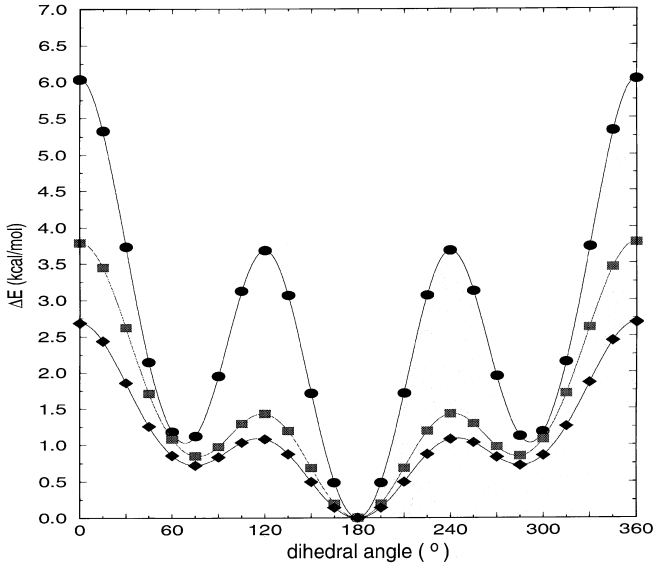

Fig. 1. Energy profiles around the central dihedral angle of butane in gas phase, as obtained from the force field, Eq. (7), with the standard OPLS parameter set (*circles*), with the modified parameters in which the third coefficients of the Fourier series are modified (*squares*), and with the surrogate parameter set in which the van der Waals interaction is reduced in addition to the modification of the Fourier series (*diamonds*)

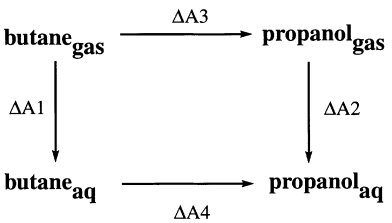

Fig. 2. Thermodynamic cycle for solvation free energy difference between butane and propanol

2.2 Computational procedure

2.2.1 Force field

Intra- and intermolecular interactions were evaluated using classical potential functions [22, 23] featuring harmonic terms for the bonded and valence interactions, a Fourier series to describe torsional motion, and a common Coulomb plus 6–12 Lennard-Jones functional format for the non-bonded terms,

$$\begin{aligned}E_{\text{potential}} = &\sum_{\text{pairs}(i,j)}\left(\frac{q_iq_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right) \\ &+ \sum_{\text{bonds}} K_b(r-r_0)^2 \\ &+ \sum_{\text{angles}} K_\theta(\Theta-\Theta_0)^2 \\ &+ \sum_{\text{torsions}}\sum_{n=0}^{3}\frac{V_n}{2}[1-(-1)^n\cos(n\phi-\gamma)]\end{aligned} \tag{7}$$

where $A_{ii} = 4\varepsilon_i\sigma_i^{12}$ and $C_{ii} = 4\varepsilon_i\sigma_i^{6}$. The van der Waals combining rules are expressed as $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ and $e_{ij} = (e_i + e_j)^{1/2}$ [24, 25]. Furthermore, all intramolecular 1,4 non-bonded interactions were scaled down by a factor of 2. The associated OPLS all-atom parameters [26–28] were used for the solutes, and the TIP3P model [29] for water was employed. All parameters are given in Tables 1 and 2.

The surrogate potential for the non-Boltzmann sampling was generated by dividing by 4 the threefold term of the X—C—C—X torsion,

$$\begin{aligned}E^{\text{dihedral}}_{\text{surrogate}}(\phi) = &\frac{V_1}{2}(1+\cos\phi) + \frac{V_2}{2}(1-\cos 2\phi) \\ &+ \frac{(V_3/4)}{2}(1+\cos 3\phi)\end{aligned} \tag{8}$$

and by scaling down by 5 the intramolecular 1,4, Lennard-Jones interactions shown in Eq. (9),

$$E^{vdw}_{\text{surrogate}} = \sum_{\text{pair}(i,j)}\left(\frac{q_iq_j}{r_{ij}} + \frac{A^{\text{sur}}_{ij}}{r_{ij}^{12}} - \frac{C^{\text{sur}}_{ij}}{r_{ij}^{6}}\right) \tag{9}$$

Table 1. OPLS parameters for butane and propanol

(a) Non-bonded parameters (Eq. 7)

Atom[a]	$q(e)$	σ(Å)	ε (kcal/mol)
H(C)	+0.06	2.50	0.030
$C(H_3)$	−0.18	3.50	0.066
$C(H_2)$	−0.12	3.50	0.066
$C(H_2OH)$	+0.145	3.50	0.066
O(H)	−0.683	3.12	0.170
H(O)	+0.418	0.00	0.000

[a] Atoms are listed with their covalently bonded neighbors in parentheses

(b) Torsion parameters (kcal/mol) (Eq. 7)

Torsion	V_1	V_2	V_3
H—C—C—C	0.00	0.00	0.366
H—C—C—H	0.00	0.00	0.318
C—C—C—C	1.740	−0.157	0.279
C—C—C—O	1.711	−0.500	0.663
C—C—O—H	−0.356	−0.174	0.492
H—C—C—O	0.00	0.00	0.468
H—C—O—H	0.00	0.00	0.450

(c) Bond parameters (Eq. 7)

Bond	K_b (kcal/mol Å^2)	r_0 (Å)
H—C	340	1.09
C—C	268	1.529
C—O	320	1.41
O—H	553	0.945

(d) Angle parameters (Eq. 7)

Angle	$K\Theta$ (kcal/mol degree^2)	Θ_0 (degree)
H—C—C	37.5	110.7
H—C—H	33.0	107.8
C—C—C	58.35	112.7
C—C—O	50.0	109.5
H—C—O	35.0	109.5
C—O—H	55.0	108.5

Table 2. TIP3P parameters for water (rigid geometry)

(a) Non-bonded parameters (Eq. 7)

Atom	$q(e)$	σ(Å)	ε (kcal/mol)
H	+0.417	0.00	0.00
O	−0.834	3.15061	0.1521

(b) Bond parameters (Eq. 7)

Bond	r_0 (Å)
H—H	1.5174
O—H	0.9572

(c) Angle parameters (Eq. 7)

Angle	Θ_0 (degree)
H—O—H	104.52
O—H—H	37.74

where $A_{ij}^{sur} = 4(\varepsilon_i/5)\sigma_i^{12}$ and $C_{ij}^{sur} = 4(\varepsilon_i/5)\sigma_i^6$. The resultant gas-phase torsional profiles from the original and surrogate potentials are shown in Fig. 1.

2.2.2 Molecular dynamics

All molecular dynamics simulations employed periodic boundary conditions in an 18.8563 Å^3 box that contained the solute plus 214 water molecules. Group-based cutoffs, where the solutes were divided into methyl, methylene, and hydroxymethylene groups and each water molecule was treated as a group, were employed; electrostatic interactions were truncated at 8.5 Å, and van der Waals interactions were truncated at 9.0 Å. A dielectric constant of 1 was used throughout.

The Verlet algorithm was used to integrate the equations of motion [30]. In the aqueous phase, temperature coupling [31] was used to keep the temperature at 298 K with a coupling constant of 10 ps^{-1}. However, in the gas phase this temperature coupling algorithm produced an undesirable effect [32]. The motion of each atom of the solute became gradually synchronized and the solute started rotating without any conformational transitions after 50 ps equilibration [32]. To alleviate this effect, Langevin dynamics with a friction coefficient of 10 ps^{-1} was employed instead of temperature coupling.

In order to facilitate the use of a 2 fs timestep, the water molecule and all bonds involving hydrogens were held rigid with the SHAKE [33] algorithm, and the solute hydrogen masses were increased to 10 amu [34].

2.2.3 NBTI

Two series of molecular dynamics simulations were performed in order to simulate the mutation of butane to propanol in the gas phase and in water. In one, the free energies were computed using NBTI; in the other, conventional TI was used in order to assess their relative performances. The initial configuration of butane in water was thermalized gradually from 10 K to 298 K and equilibrated for 100 ps according to the schedule given in Table 3. The mutation then proceeded in 15 unevenly spaced windows, where λ_i = 0.0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.975, 0.99, and 1.0. To avoid convergence problems often associated with the solute-solvent energetics, more closely spaced windows were chosen for the endpoints [35]. This was particularly important as the alcohol group was "created" for the path from butane to

Table 3. Thermalization and equilibration protocol of the initial configuration of butane in water

Step	Simulation time (ps)	Δt (fs)	Temperature (K)[a]	τ (ps)
1	0.5	0.5	10	0.001
2	2.0	2.0	10	0.01
3	2.0	2.0	10 → 200	1.0
4	10	2.0	298	1.0
5	100	2.0	298	0.1

[a] The temperature coupling method [31] was employed

propanol. In each window, geometric and parametric features of the system were gradually modified as the simulation progressed from butane to propanol, and the "disappearing" hydrogens were converted to dummy atoms ($q = \varepsilon = \sigma = 0$) and retracted to final bond lengths of 0.3 Å. No special treatment was made to correct for the small free energy differences associated with changing the bond lengths [36–38].

The simulations were begun by assigning velocities to all atoms via a Maxwellian distribution at 298 K. With the TI and NBTI approaches, two distinct simulation experiments were carried out. First, simulations with equal length equilibration and sampling stages were performed. Here, the issue of longer timescales was the focus and involved lengthening these periods from 10 ps/10 ps (equilibration/sampling) to 50 ps/50 ps in five even increments and then performing one simulation for 100 ps/100 ps. Second, the convergence of the simulations was assessed by keeping the equilibration time constant at 50 ps and increasing the sampling stage from 10 ps to 200 ps.

2.2.4 Multi-isometric state free energy calculations

The NBTI method was validated through comparisons with multi-isomeric state free energy calculations [15, 16], which have been shown to yield accurate free energies of hydration for small, flexible organic molecules while sampling adequately rotational probabilities. Briefly, to obtain the free energy of hydration difference between butane and propanol from this method, six simulations – three in gas phase and an equivalent three in aqueous phase – were performed. As indicated in Eq. (10), two potentials of mean force (PMFs) were computed to explore the conformational space defined by the central torsion of each endpoint (butane and propanol), and a mutation of the conformationally restricted *trans*-butane to *trans*-propanol was computed. The relative free energy of mutation from butane to propanol, then, was computed by summing these free energy changes [15].

$$\begin{aligned}\Delta A &= A_B - A_P \\ &= \left(A_B - A_B^t\right) + \left(A_B^t - A_P^t\right) + \left(A_P^t - A_P\right)\end{aligned} \tag{10}$$

where $A_B - A_B^t$ is the free energy difference between the *trans* form of butane and the full system with all of the possible isomers, $A_B^t - A_P^t$ is the free energy difference between the *trans* form of butane and the *trans* form of propanol, and $A_B^t - A_P^t$ is the free energy difference between the *trans* form of propanol and the full system with all of the possible isomers.

In both the gas phase and water, 3 ns MD simulations (following a thermalization phase as described in Table 3) in which the central X—C—C—X torsion of the molecule (butane and propanol) was turned off were performed to obtain the PMFs, W(ϕ) [10],

$$W(\phi) = -RT \ln \tilde{\rho}(\phi) - U(\phi) + \text{constant}, \tag{11}$$

where $\tilde{\rho}(\phi)$ is the normalized probability density of the dihedral angle obtained from removal of the potential energy associated with motion about this bond. In practice, this is equivalent to employing an umbrella potential, $U(\phi)$, in order to generate the biased ensemble of configurations. The free energy difference between the *trans* form of the solute and its conformationally unrestricted counterpart was computed as

$$\Delta A = -k_B T \ln \left[\left[\int \exp(-\Delta A_{\text{rot}}(\phi)/k_B T) d\phi \right] \Big/ 2\pi \right], \tag{12}$$

where k_B is the Boltzmann constant and T is absolute temperature of the simulation. ΔA_{rot} gives the potential of mean force around the central dihedral angle, ϕ, with respect to the free energy of the *trans* state.

Consistent with defined protocols for the multi-isomeric state approach [15–16], molecular dynamics simulations using conventional TI with a "flat-bottom" restraining potential, U_{res} were then performed for the *trans*-butane to *trans*-propanol mutation. The functional form of U_{res} is

$$U_{\text{res}}(\phi) = \begin{cases} k(\phi - \phi_0 - \Delta\phi)^2 & \text{if } \phi > \phi_0 + \Delta\phi \\ k(\phi - \phi_0 + \Delta\phi)^2 & \text{if } \phi > \phi_0 - \Delta\phi \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where ϕ_0 is the trans angle 180°, k = 50 kcal/mol degree^2 and $\Delta\phi$ determines the allowed range of dihedral angle excursions, 30°. As above, no special treatment was made for eliminating the effects of the restraining potential on the free energies. Otherwise, the MD conditions and procedures outlined above for NBTI and conventional TI were employed here as well. The system was subjected to equilibration periods of 50 ps followed by sampling over 50 ps. In all cases, coordinates of the central cell of the periodic lattice were stored every 0.5 ps of the sampling stage, and the free energies were computed as ensemble averages over these coordinates.

Finally, although statistical errors derived from ensemble averages have conventionally been used for error estimation in the free energy calculations [3–6], it is found that the statistical errors are sometimes fortuitously small and can be attributed to the molecule being trapped in a local minimum [39]. Accordingly, all simulations were performed three times and were begun from different randomly drawn velocities. The reported free energies and errors are the averages of these trials and standard deviations [39], respectively.

All calculations were carried out with the program X-PLOR [24] running on a Hewlett-Packard APOLLO 735.

3 Results

The free energy results for the multi-isomeric state benchmark are given in Table 4 and Fig. 3. The effects of hydration on the butane and propanol torsional profiles, which are given by the PMFs, follow expected trends. Specifically, with butane, the *gauche* conformer is better hydrated by about 0.25 kcal/mol, and the *cis* rotamer is lower than the *gauche* by an equivalent amount. These results are consistent with a hydrophobic stabilization of the more compact rotamers and are in agreement with

Table 4. Free energies of solvation difference between butane and propanol (in kcal/mol) at 298 K obtained by multi-isomeric state free energy calculation. All free energies are reported as averages and standard deviations obtained by the three separate simulations with different initial velocities

Path	ΔA_{gas}	ΔA_{aq}	ΔA_{sol}
Butane → t-butane	−1.27 ± 0.04	−1.21 ± 0.05	
t-Butane → t-propanol	−2.00 ± 0.2	−8.99 ± 0.2	
t-Propanol → propanol	0.82 ± 0.04	0.96 ± 0.06	
Butane → propanol	−2.45 ± 0.2	−9.16 ± 0.2	−6.7 ± 0.3

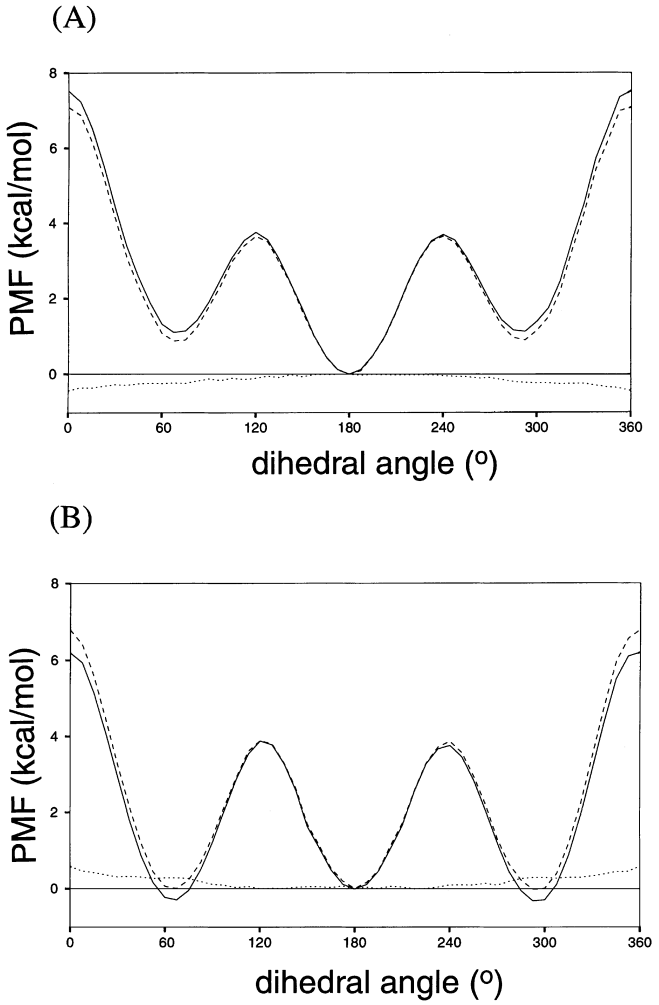

Fig. 3A,B. Hydration effect around the central dihedral angle of the solute, **A** butane and **B** propanol. The PMFs are plotted with respect to the free energy level of the *trans*-state. The *solid line* and the *dashed line* indicate the PMFs in gas phase and in water respectively. The hydration effects (*dotted line*) are calculated as the difference of PMFs between gas phase and aqueous phase

earlier studies [40–44]. With propanol, the situation is reversed, with the *trans* conformer being the most well-hydrated species. This can be explained in terms of a loss of full hydrogen-bonding potential as the molecule moves from *trans* to *cis* [40, 43]. The free energy difference between butane and propanol in the gas phase and in water was −2.45 ± 0.2 kcal/mol and −9.16 ± 0.2 kcal/mol, respectively. Therefore, the relative free energy of hydration was estimated as −6.7 ± 0.3 kcal/mol, which is in good agreement with the experimental result of −6.9 kcal/mol [44].

The data for the first experiment with TI and NBTI are shown in Tables 5A and B and Fig. 4. In the final analysis, both methods produce good results; mutation of butane to propanol yields free energy differences of hydration of −6.8 ± 0.2 kcal/mol for conventional TI and −6.8 ± 0.1 kcal/mol for NBTI. Consideration of the progression in the free energy differences and associated errors from short to long simulation times reveals interesting results. While in the end simulations using TI converge to the target free-energy difference, the shorter runs get trapped in local minima of the energy surface where the relative free energy of hydration is 0.5 kcal/mol less favorable for propanol than expected. As a consequence, the *gauche* conformers are improperly sampled, which would increase the preference for propanol in the gas phase and lessen its preference in water due to restricted hydrogen bonding. Inspection of the errors in the TI simulations lends further support to this conclusion. Specifically, in any one run the statistical error is small (0.1 kcal/mol for 20 ps/20 ps), whereas for the average of three runs the standard deviation is much larger (0.4 kcal/mol for 20 ps/20 ps). This is not the case for NBTI, where the relative free energy of hydration taken at any of the time intervals is in excellent agreement with the target value. Moreover, both the statistical errors and the standard deviations from the three simulations are equivalent and small (in the 0.1–0.2 kcal/mol range for all). For short simulations (Tables 5 and 6), the statistical errors of NBTI are larger than those of TI.

The convergence studies shown in Tables 6A and B and Fig. 5 emphasize the differences between NBTI and TI. Following an equilibration time of 50 ps, NBTI moves quickly towards the target free-energy difference; after 10 ps, the relative free energy of hydration is −6.8 ± 0.2 kcal/mol and by the end of the 200 ps sampling stage it remains at −7.0 ± 0.1 kcal/mol. Conventional TI, on the other hand, stays near −6.3 kcal/mol for short sampling times. Furthermore, it is clear that TI requires at least 100 ps to converge to the expected free-energy difference; at the end of the 100 ps sampling stage TI gives a relative free energy of hydration of −6.8 ± 0.3 kcal/mol.

The individual convergence of ΔA_{gas} and ΔA_{aq} by both methods is also compared in Tables 6A and B and Figs. 6 and 7. In theory, the free energies by TI and NBTI must converge to the same value regardless of the difference of the internal potential energies. In the gas phase, more than 100 ps sampling is necessary for both methods to converge to the same value. In the aqueous phase, even after 200 ps, there is still a significant difference in the values between TI and NBTI, −9.38 ± 0.1 kcal/mol and −9.52 ± 0.1 kcal/ mol, respectively. The free energy obtained by TI is still decreasing even after 100 ps and it fluctuates, which indicates the slow convergence by TI. In contrast, the free energy by NBTI stays around −9.55 kcal/mol after 40 ps sampling. The reason for the superior convergence of ΔA_{sol} compared to ΔA_{gas} or ΔA_{aq} is unclear.

Table 5. Free energy of solvation differences between butane and propanol (in kcal/mol) at 298 K

(A) Free energy differences obtained by TI

Simulation time[a] (equilibration, sampling)	ΔA_{gas}[b]	ΔA_{aq}[b]	ΔA_{sol}[c]
(5 ps, 5 ps)	−2.78 ± 0.5 (0.08)	−9.19 ± 0.3 (0.5)	−6.4 ± 0.6 (0.5)
(10 ps, 10 ps)	−2.55 ± 0.4 (0.06)	−8.83 ± 0.3 (0.2)	−6.3 ± 0.5 (0.2)
(20 ps, 20 ps)	−2.67 ± 0.3 (0.03)	−8.89 ± 0.3 (0.1)	−6.2 ± 0.4 (0.1)
(30 ps, 30 ps)	−2.62 ± 0.2 (0.03)	−8.99 ± 0.2 (0.1)	−6.4 ± 0.3 (0.1)
(40 ps, 40 ps)	−2.68 ± 0.2 (0.03)	−9.01 ± 0.2 (0.09)	−6.3 ± 0.3 (0.09)
(50 ps, 50 ps)	−2.38 ± 0.2 (0.02)	−9.06 ± 0.2 (0.08)	−6.7 ± 0.3 (0.08)
(100 ps, 100 ps)	−2.35 ± 0.2 (0.02)	−9.13 ± 0.1 (0.04)	−6.8 ± 0.2 (0.04)

(B) Free energy differences obtained by NBTI

Simulation time[a] (equilibration, sampling)	ΔA_{gas}[b]	ΔA_{aq}[b]	ΔA_{sol}[c]
(5 ps, 5 ps)	−2.83 ± 0.2 (0.2)	−9.47 ± 0.3 (0.6)	−6.6 ± 0.4 (0.6)
(10 ps, 10 ps)	−2.80 ± 0.1 (0.1)	−9.61 ± 0.2 (0.4)	−6.8 ± 0.2 (0.4)
(20 ps, 20 ps)	−2.64 ± 0.1 (0.07)	−9.52 ± 0.2 (0.2)	−6.9 ± 0.2 (0.2)
(30 ps, 30 ps)	−2.63 ± 0.1 (0.06)	−9.41 ± 0.2 (0.1)	−6.8 ± 0.2 (0.1)
(40 ps, 40 ps)	−2.83 ± 0.1 (0.04)	−9.58 ± 0.1 (0.1)	−6.8 ± 0.1 (0.1)
(50 ps, 50 ps)	−2.81 ± 0.1 (0.04)	−9.62 ± 0.1 (0 1)	−6.8 ± 0.1 (0.1)
(100 ps, 100 ps)	−2.73 ± 0.1 (0.03)	−9.55 ± 0.07 (0.06)	−6.8 ± 0.1 (0.07)

[a] Simulation times are composed of two parts: equilibration and sampling (see method)
[b] The free energy values are reported as the average values and standard deviations of three separate simulations with different initial velocities. The average statistical errors derived from the ensembles are given in parentheses
[c] ΔA_{sol} is calculated as $\Delta A_{aq} - \Delta A_{gas}$

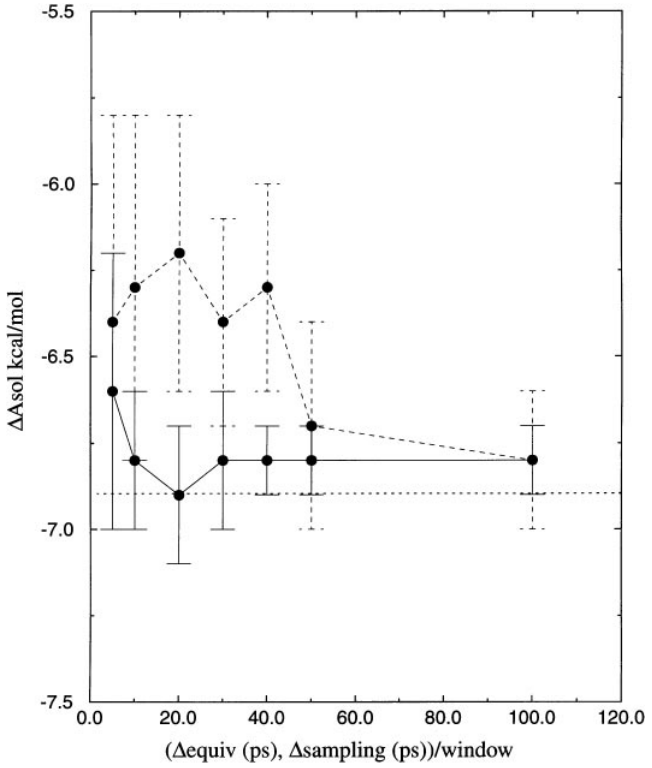

Fig. 4. Convergence of free energy difference in solvation between butane and propanol by conventional TI (*dashed line*) and NBTI (*solid line*) with various lengths of equilibration and sampling time. The calculated free energy values are plotted as a function of simulation time of the equilibration and sampling stages; for instance, 30 ps on the *x* axis indicates the simulation with 30 ps equilibration and 30 ps sampling per window. The *horizontal dotted line* indicates the experimentally measured value (−6.9 kcal/mol) [44]. The free energies and the error bars are calculated as averages and standard deviations of three distinct simulations with different initial velocities

As expected, the accuracy and efficiency of NBTI can be explained in terms of the enhanced sampling and is demonstrated in Figs. 8–10. From transition state theory [45], the predicted transition rate for *trans*/*gauche* butane, given a 3.75 kcal/mol barrier at 298 K and a transmission coefficient of unity, is 0.011 ps^{-1}. The average transition rate from a gas phase simulation of butane at 298 K using the unmodified potential is in close agreement at 0.012 ps^{-1}, or once every 83 ps. With NBTI, the transition rate is increased by an order of magnitude of 0.149 ps^{-1}, or once every 6.7 ps. TI would fail to sample adequately in a 50 ps sampling stage; this is clearly not the case with NBTI. The same behavior is observed in water (data not shown).

The effects of sampling periods on the gas-phase dihedral angle distributions of the central torsion angle in butane are shown in Figs. 9 and 10. The most striking feature is that even at a sampling period of 200 ps, a simulation with the standard OPLS force field completely fails to sample the *gauche*+ conformation, which should be found with the same probability as the *gauche*– rotamer. This is not the case with the modified force field, which samples all conformers with reasonable frequency. Thus it appears that prohibitively long sampling periods – in the order of a nanosecond – are required for TI using the standard force field to sample with the same efficiency as NBTI with the modified force field. Clearly, the ability of these methods to compute quickly and accurately the relative free energy of hydration of butane and propanol is a direct consequence of their ability to visit conformational probabilities on the energy surface.

4 Conclusion

The accuracy of free energy perturbation methods depends on the validity of empirical energy functions and on sampling of conformational space. NBTI

Table 6. Free energy of solvation differences between butane and propanol (in kcal/mol) at 298 K

(A) Free energy differences obtained by TI with a constant equilibration interval (50 ps)

Simulation time[a] [equiv (50 ps), sampling]	ΔA_{gas}[b]	ΔA_{aq}[b]	ΔA_{sol}[c]
(50 ps, 10 ps)	−2.36 ± 0.2 (0.05)	−8.70 ± 0.3 (0.3)	−6.3 ± 0.4 (0.3)
(50 ps, 20 ps)	−2.44 ± 0.2 (0.04)	−8.71 ± 0.3 (0.2)	−6.3 ± 0.4 (0.2)
(50 ps, 30 ps)	−2.56 ± 0.2 (0.04)	−8.82 ± 0.2 (0.1)	−6.3 ± 0.3 (0.1)
(50 ps, 40 ps)	−2.51 ± 0.1 (0.03)	−9.04 ± 0.2 (0.1)	−6.5 ± 0.2 (0.1)
(50 ps, 50 ps)	−2.41 ± 0.1 (0.02)	−9.02 ± 0.2 (0.1)	−6.6 ± 0.2 (0.1)
(50 ps, 100 ps)	−2.52 ± 0.2 (0.03)	−9.31 ± 0.2 (0.1)	−6.8 ± 0.3 (0.1)
(50 ps, 200 ps)	−2.57 ± 0.1 (0.02)	−9.38 ± 0.1 (0.1)	−6.8 ± 0.2 (0.1)

(B) Free energy differences obtained by NBTI with a constant equilibration interval (50 ps)

Simulation time[a] [equiv (50 ps), sampling]	ΔA_{gas}[b]	ΔA_{aq}[b]	ΔA_{sol}[c]
(50 ps, 10 ps)	−2.98 ± 0.1 (0.1)	−9.73 ± 0.2 (0.4)	−6.8 ± 0.2 (0.4)
(50 ps, 20 ps)	−2.80 ± 0.1 (0.09)	−9.82 ± 0.2 (0.2)	−7.0 ± 0.2 (0.2)
(50 ps, 30 ps)	−2.78 ± 0.1 (0.07)	−9.70 ± 0.2 (0.2)	−6.9 ± 0.2 (0.2)
(50 ps, 40 ps)	−2.73 ± 0.1 (0.06)	−9.55 ± 0.1 (0.1)	−6.8 ± 0.1 (0.1)
(50 ps, 50 ps)	−2.69 ± 0.1 (0.05)	−9.53 ± 0.1 (0.1)	−6.8 ± 0.1 (0.1)
(50 ps, 100 ps)	−2.67 ± 0.1 (0.05)	−9.58 ± 0.1 (0.1)	−6.9 ± 0.1 (0.1)
(50 ps, 200 ps)	−2.65 ± 0.1 (0.04)	−9.52 ± 0.1 (0.1)	−7.0 ± 0.1 (0.1)

[a] Simulation times are composed of two parts: equilibration and sampling (see method)
[b] The free energy values are reported as the average values and standard deviations of three separate simulations with different initial velocities. The statistical errors are given in parentheses
[c] ΔA_{sol} is calculated as $\Delta A_{aq} - \Delta A_{gas}$

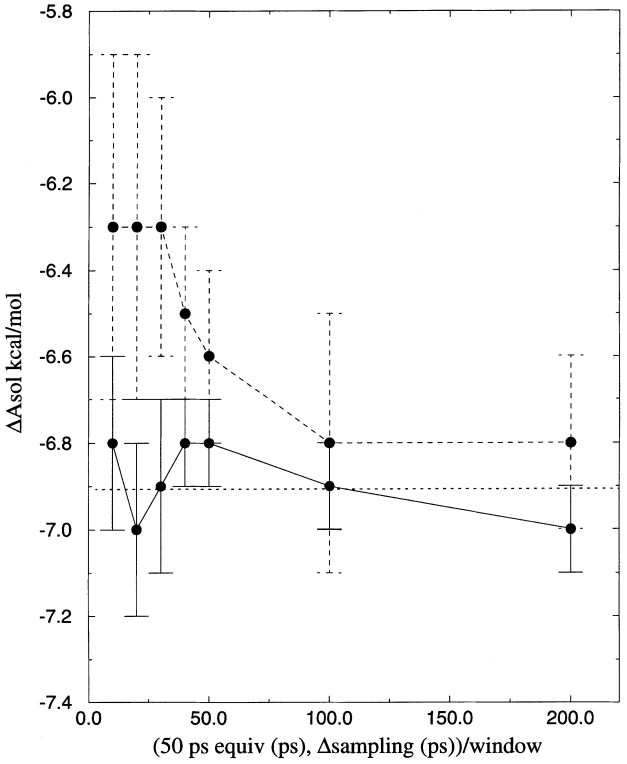

Fig. 5. Convergence of the free energy difference of solvation between butane and propanol by conventional TI (*dashed line*) and NBTI (*solid line*) with a constant length of equilibration (50 ps) and several lengths of sampling time per window. The calculated free energy values are plotted as a function of the sampling time; for instance, 30 ps on the *x* axis indicates the simulation with 50 ps equilibration and 30 ps sampling per window. The *horizontal dotted line* indicates the experimentally measured value (−6.9 kcal/mol) [44]. The free energies are calculated using a single simulation; therefore the error bars are estimated by the statistical error of each ensemble

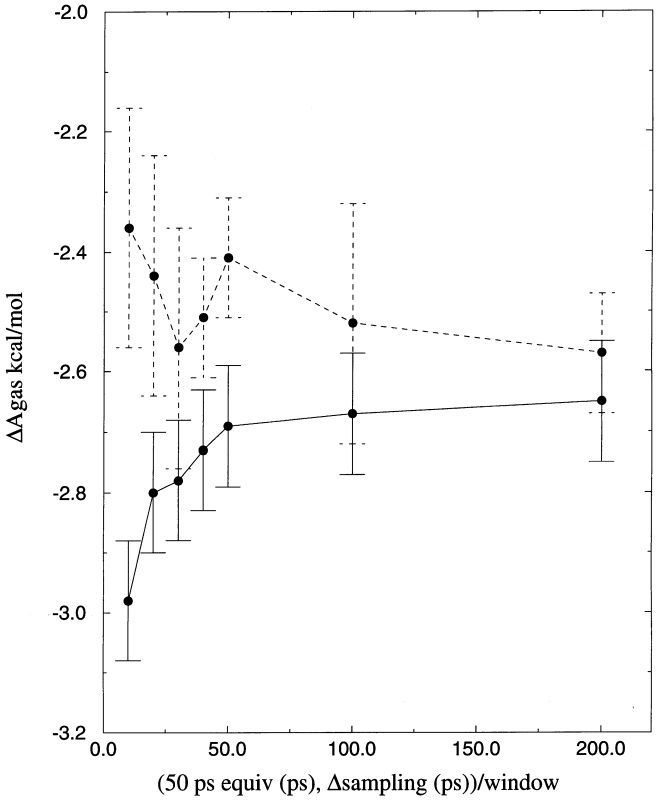

Fig. 6. Convergence of the free energy difference in gas phase between butane and propanol obtained by conventional TI (*dashed line*) and NBTI (*solid line*) with a constant length of equilibration (50 ps) and several lengths of sampling time per window. The calculated free energy values are plotted as a function of the sampling time; for instance, 30 ps on the *x* axis indicates the simulation with 50 ps equilibration and 30 ps sampling per window. The free energies are calculated using a single simulation; therefore the error bars are estimated by the statistical error of each ensemble

addresses the latter problem. The modified force field used with NBTI can improve sampling of conformational space by reducing energy barriers without changing the location of the energy minima. In principle, the energy function could be "flattened," but a great penalty would be paid for frequently sampling rotamers that otherwise would be considered high in energy and unimportant. The net effect would be a gain in error in the computed free energies. The influence of the modification of the force field is corrected by the non-Boltzmann sampling (umbrella sampling) procedure without the need for additional free energy perturbation stages.

The barrier reduction made for the butane/propanol test case significantly improves sampling of conforma-

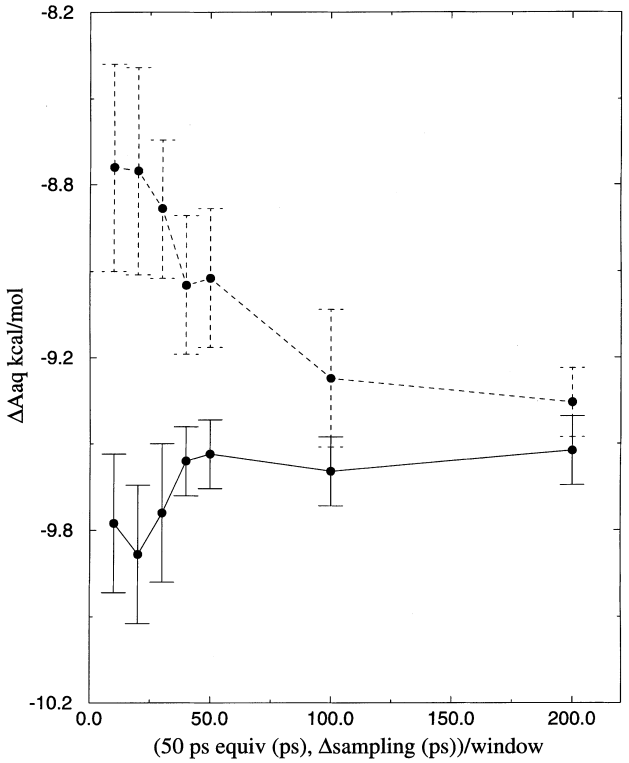

Fig. 7. Convergence of the free energy difference in aqueous phase between butane and propanol obtained by conventional TI (*dashed line*) and NBTI (*solid line*) with a constant length of equilibration (50 ps) and several lengths of sampling time per window. The calculated free energy values are plotted as a function of the sampling time; for instance, 30 ps on the *x* axis indicates the simulation with 50 ps equilibration and 30 ps sampling per window. The free energies are calculated using a single simulation; therefore the error bars are estimated by the statistical error of each ensemble

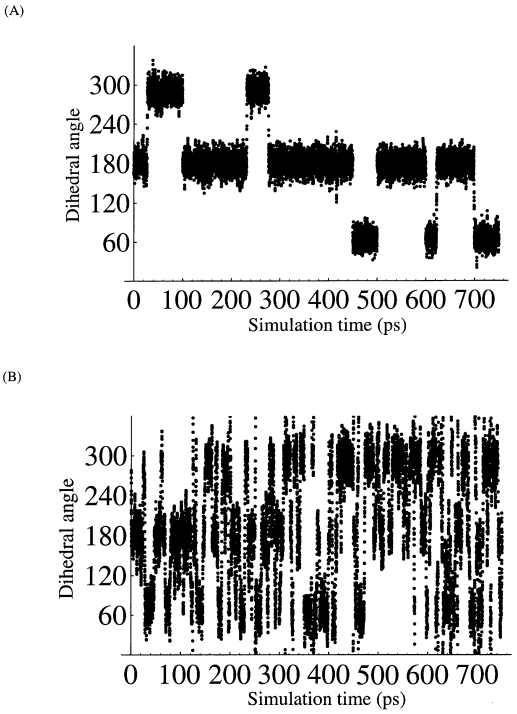

Fig. 8A,B. Time course of the central dihedral angle of the solute during mutation from butane to propanol in gas phase obtained by TI (**A**) and NBTI (**B**). The distributions are plotted every 0.05 ps only for the sampling stages (750 ps)

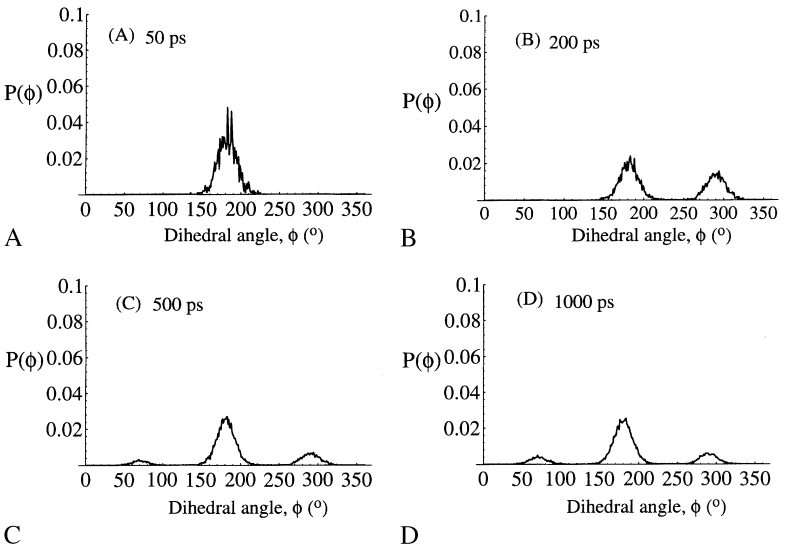

Fig. 9A-D. Normalized dihedral angle probability distribution of the central dihedral angle of butane in gas phase obtained by the standard OPLS force field for various simulation times: **A** 50 ps, **B** 200 ps, **C** 500 ps, **D** 1000 ps

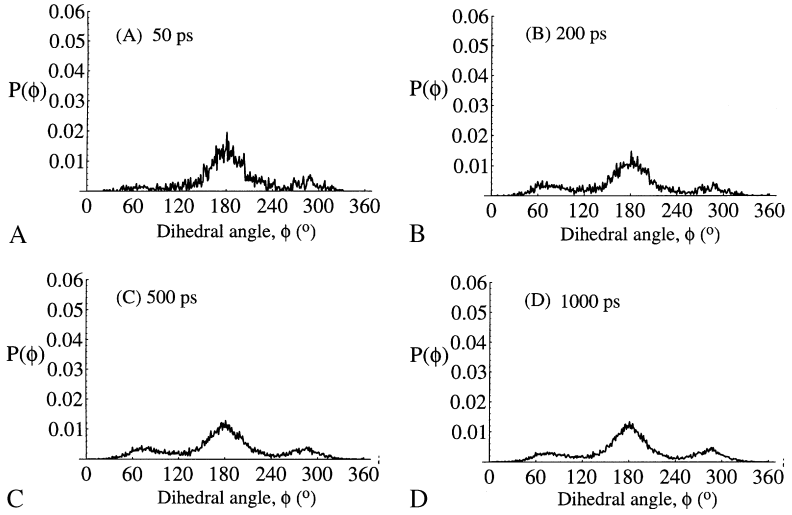

Fig. 10A-D. Normalized dihedral angle probability distribution of the central dihedral angle of butane in gas phase obtained by the modified OPLS force field with reduced dihedral angle barriers, Eq. (7), for various simulation times: **A** 50 ps, **B** 200 ps, **C** 500 ps, **D** 1000 ps

tional space. Convergence to accurate free energy values is achieved with one tenth of the simulation time for TI. Additionally, NBTI is competitive with the multi-isomeric state approach when one considers the numbers of simulations involved in computing the free energy difference; that is, NBTI involves only two simulations for the difference of the free energy of hydration between butane and propanol; the multi-isomeric state and the methods of Mark et al. [20] require six.

An important feature of NBTI is the ability to perform free energy calculations of a system with multiple isomeric states without having to calculate PMFs. In the case of a larger molecule with many degrees of freedom, the calculation of PMFs is computationally very expensive [8]. Therefore, it is difficult to use the multi-isomeric approach for proteins [8]. NBTI could be useful for this case.

While the focus here is on enhanced sampling of intramolecular degrees of freedom, it is also possible to envisage its application to ligand-receptor binding computations with the goal of reducing the intermolecular barriers between various binding modes. Work in this area is ongoing.

Acknowledgements. We thank Professor William L. Jorgensen and Dr. Erin M. Duffy for stimulating discussions and critical reading of the manuscript. This work was supported by a grant from the National Institute of Health to A.T.B. (NIH POL GM39546-06).

References

1. Zwanzig RW (1954) J Chem Phys 22:1420
2. Kirkwood JG (1935) J Chem Phys 3:300
3. Beveridge DL, DiCapua FM (1989) Annu Rev Biophys Chem 18:431
4. Jorgensen WL (1989) Acc Chem Res 22:184
5. Straatsma TP, McCammon JA (1992) Annu Rev Phys Chem 43:407
6. Kollman P (1993) Chem Rev 93:2395
7. Simomson T, Brünger AT (1992) Biochemistry 31:8661
8. Hodel A, Simonson T, Fox RO, Brünger AT (1993) J Phys Chem 97:3409
9. Torrie GM, Valleau JP (1977) J Comput Phys 23:187
10. McCammon JA, Harvey SC (1987) Dynamics of proteins and nucleic acids. Cambridge University Press, Cambridge
11. Tobias DJ, Brooks CL III, Fleischman SH (1989) Chem Phys Lett 156:256
12. Rebertus DW, Berne BJ, Chandler D (1979) J Chem Phys 70:3395
13. Bigot B, Jorgensen WJ (1981) J Chem Phys 75:1944
14. Anderson AG, Hermans J (1988) Proteins 3:262
15. Straatsma TP, McCammon JA (1989) J Chem Phys 90:3300
16. Straatsma TP, McCammon JA (1989) J Chem Phys 91:3631
17. Dang LX, Pettitt BM (1987) J Am Chem Soc 109:5531
18. Northrup SH, Pear MR, Lee C-Y, McCammon JA, Karplus M (1982) Proc Natl Acad Sci USA 79:4035
19. Haydock C, Sharp JC, Prendergast FG (1990) Biophys J 57:1269
20. Mark AE, van Gunsteren WF, Berendsen HJC (1991) J Chem Phys 94:3808
21. Tembe B, McCammon JA (1984) Comput Chem 8:281
22. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657
23 Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P (1984) J Am Chem Soc 106:765
24. Brünger AT X-PLOR version 3.1. Yale University, New Haven, Conn (1992)
25. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187
26. Kaminski G, Duffy EM, Matsui T, Jorgensen WL (1994) J Chem Phys 98:13077
27. Maxwell DS, Tirado-Rives J, Jorgensen WL (1995) J Comput Chem 16:984
28. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225
29. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
30. Verlet L (1967) Phys Rev 159:98
31. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684
32. Ota N, Unpublished results
33. Ryckaert J, Ciccotti G, Berendsen H (1977) J Comput Phys 23:327
34 Pomès R, McCammon JA (1990) Chem Phys Lett 166:425
35. Jorgensen WL, Ravimohan C (1985) J Chem Phys 83:3050
36. Pearlman DA, Kollman PA (1991) J Chem Phys 94:4532
37. Sun Y, Spellmeyer D, Pearlman DA, Kollman P (1992) J Am Chem Soc 114:6798
38. Severance DL, Essex JW, Jorgensen WL (1995) J Comput Chem 16:311
39. Hermans J, Yun RH, Anderson AG (1992) J Comput Chem 13:429

40. Jorgensen WL (1986) J Phys Chem 90:1276
41. Jorgensen WL, Buckner JK (1987) J Phys Chem 91:6083
42. Pratt LR, Chandler D (1977) J Chem Phys 66:147
43. Fleischman SH, Brooks CL III (1987) J Chem Phys 87:3029
44. Ooi T, Oobatake M, Némethy G, Scheraga HA (1987) Proc Natl Acad Sci USA 84:3086
45. Glasstone S, Laidler K, Eyring H (1941) The theory of rate processes. McGraw-Hill, New York, p 153