

**Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence**

Wei Yang, Ryan Bitetti-Putzer, and Martin Karplus

Citation: *The Journal of Chemical Physics* **120**, 2618 (2004); doi: 10.1063/1.1638996

View online: <http://dx.doi.org/10.1063/1.1638996>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/120/6?ver=pdfcov>

Published by the [AIP Publishing](#)

---

**Articles you may be interested in**

[Conformational free energies of methyl-  \$\alpha\$  -L-iduronic and methyl-  \$\beta\$  -D-glucuronic acids in water](#)  
*J. Chem. Phys.* **132**, 104108 (2010); 10.1063/1.3355621

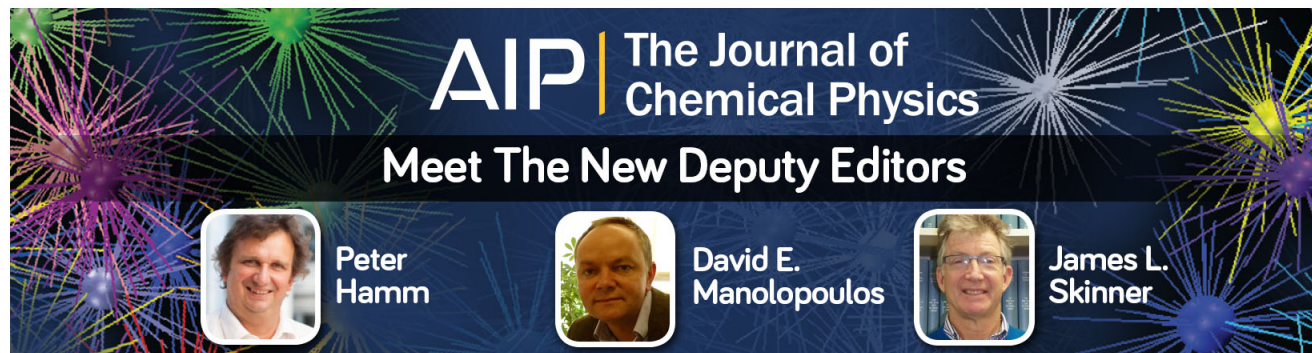
[Appropriate methods to combine forward and reverse free-energy perturbation averages](#)  
*J. Chem. Phys.* **118**, 2977 (2003); 10.1063/1.1537241

[Calculating free energies using average force](#)  
*J. Chem. Phys.* **115**, 9169 (2001); 10.1063/1.1410978

[Response to "Comment on 'New strategies to incorporate the solvent polarization in self-consistent reaction field and free-energy perturbation simulations'" \[\*J. Chem. Phys.\* 107, 1291 \(1997\)\]](#)  
*J. Chem. Phys.* **107**, 1293 (1997); 10.1063/1.474478


[Comment on "New strategies to incorporate the solvent polarization in self-consistent reaction field and free-energy perturbation simulations" \[\*J. Chem. Phys.\* 103, 10183 \(1995\)\]](#)  
*J. Chem. Phys.* **107**, 1291 (1997); 10.1063/1.474477

---




**AIP** | The Journal of  
Chemical Physics


**Meet The New Deputy Editors**



Peter Hamm



David E. Manolopoulos



James L. Skinner

# Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence

Wei Yang

*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138*

Ryan Bitetti-Putzer

*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138  
and Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, Massachusetts 02138*

Martin Karplus<sup>a)</sup>

*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138  
and Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, 67000 Strasbourg, France*

(Received 29 July 2003; accepted 12 November 2003)

A method is proposed for improving the accuracy and efficiency of free energy simulations. The essential idea is that the convergence of the relevant measure (e.g., the free energy derivative in thermodynamic integration) is monitored in the reverse direction starting from the last frame of the trajectory, instead of the usual approach, which begins with the first frame and goes in the forward direction. This simple change in the use of the simulation data makes it straightforward to eliminate the contamination of the averages by contributions from the equilibrating region. A statistical criterion is introduced for distinguishing the equilibrated (production) region from the equilibrating region. The proposed method, called reverse cumulative averaging, is illustrated by its application to the well-studied case of the alchemical free energy simulation of ethane to methanol. © 2004 American Institute of Physics. [DOI: 10.1063/1.1638996]

## INTRODUCTION

Free energy simulations are important tools for the study of many chemical and biological problems, including protein–ligand interactions, solvation effects and conformational changes.<sup>1–3</sup> Thermodynamic integration<sup>4</sup> and exponential averaging<sup>5</sup> are the most widely used techniques in free energy simulations. Both techniques use ensemble averages obtained by molecular-dynamics or Monte Carlo methods. Since finite time series consisting of correlated data points are involved in determining the averages, the question of convergence is an important one. Moreover, a series of simulations corresponding to different values of a parameter  $\lambda$  that connects the initial and final states are generally used, so that it is the convergence for each “window” that is important. There are two aspects to consider in deciding how long the simulations have to be, given that it is desirable to minimize the required computer time. For every window, it is necessary to determine first, how long a simulation is required to obtain an equilibrated system, and second, given an equilibrated system, how long a simulation is needed to obtain converged results. Accurate and precise ensemble averages for each window are essential for the free energy evaluation itself and for meaningful post processing<sup>6,7</sup> and component analysis<sup>8,9</sup> of the simulations. Estimating the time scale for equilibration and convergence can also help us to understand the dynamics (e.g., the relaxation behavior) of the systems under study.

In many free energy simulations, the equilibration time

appears to have been chosen somewhat arbitrarily; e.g., the same equilibration time is used for all windows, other than the first.<sup>10–13</sup> In other studies, the equilibration time has been determined by monitoring the energy and structure of the system according to certain qualitative criteria; e.g., the first plateau in the potential energy following thermalization, or when the structure begins to look “stable.” Usually, relatively short equilibration times are used; in Ref. 10, for example, 50 ps are used initially to prepare the system and then 1.5 ps equilibrating time followed by 1.5 ps production for each change in  $\lambda$  value; in Ref. 11, 10 ps equilibration time was followed by 20 ps data collection. This is in accord with the widespread opinion that an equilibrium state or, at least, that a state near-enough to equilibrium, is reached rapidly; in the latter case it commonly has been assumed that a longer production time will diminish, indeed eventually eliminate, any errors resulting from the inclusion of nonequilibrium data. Although in principle this is true, practically it is known that a deviation from the correct average takes a very long time to disappear into the noise (as in the famous example of the coin tossing problem<sup>14</sup>). Like the equilibration time, the convergence (production) time for the simulations usually is not monitored carefully and often is set to a constant value for all windows.<sup>10–13</sup> In a few applications, error estimate techniques, such as the block-averaging method,<sup>15</sup> have been employed to decide on the required production time, though without establishing rigorously when the production period should begin. The convergence has been examined by following the time development of the forward cumulative average of the free energy derivative starting from an arbitrarily initial chosen point.<sup>16</sup> The terms “forward” and

<sup>a)</sup>Author to whom correspondence should be addressed. Fax: 1-617-496-4793. Electronic mail: marci@tammy.harvard.edu

“reverse” cumulative average, which are essential for the present analysis, are defined as follows. Given a sequential set of data points  $\{X_j\}$ , where the index  $j$  runs from 1 (the starting point) to  $n$ , the sequence corresponding to the forward cumulative average (FCA) is the set  $\{\bar{X}_j^{\text{FCA}}\}$ , where  $\bar{X}_j^{\text{FCA}} \equiv (\sum_1^j X_j)/j$ , while that of the reverse cumulative average (RCA) is the set  $\{\bar{X}_i^{\text{RCA}}\}$ , where  $\bar{X}_i^{\text{RCA}} \equiv (\sum_i^{n-i+1} X_j)/i$ .

From this discussion, it is clear that an equilibration criterion to eliminate nonequilibrium contributions to the ensemble averages would be very useful. The FCA does not provide a well-defined procedure for distinguishing between deviations due to fluctuations at equilibrium and those due to contamination by nonequilibrated data. Thus, using the FCA to examine a window simulation, it is very difficult to determine rigorously when the system is equilibrated and the production can begin. Standard error estimates<sup>17</sup> are based on the assumption that the data are collected under stationary conditions, so that inaccurate results are obtained unless the equilibrated region has been determined. This explains how it is possible to have forward and reverse results that have the same calculated precision but differ in their values; i.e., it is the inclusion of nonequilibrium contributions that makes the free energy results different in the forward and the reverse directions, giving rise to the so-called “hysteresis” problem.<sup>18,19</sup> The issue is particularly important in slow-growth simulations,<sup>20</sup> in which the configurations always lag behind the Hamiltonian.<sup>21,22</sup> It should be mentioned, however, that nonequilibrium “fast-growth” simulations, if treated appropriately, can be used, in principle, to determine accurate free energies.<sup>23–25</sup> An essential difference between the two methods is that in slow-growth, the free energy calculation is formulated as if the system were at equilibrium.

Statistical methods have been proposed to locate the equilibration and production regions for molecular-dynamics simulations.<sup>26</sup> Schiferl and Wallace<sup>27</sup> describe several methods to examine the distribution of simulation data, all of which are applied in the forward direction in their study. Many of the methods proposed are expensive and somewhat impractical and thus have seldom been applied. The most useful method appears to be the  $W$  test,<sup>28,29</sup> which we employ in our development and review in the *Theory* section.

In this paper, we introduce a new approach, reverse cumulative averaging (RCA) combined with a statistical analysis. The method is computationally efficient and straightforward to apply and appears to be sufficiently robust for use in free energy analysis. In particular, by use of reverse cumulative averaging, the equilibrating and equilibrated regions can be distinguished in a rigorous manner, allowing the ensemble properties of the latter region (the production region) to be estimated with accurate and precise error bounds. We first describe the theoretical basis of reverse cumulative averaging, and then illustrate it by simulating the alchemical change of ethane to methanol in aqueous solution. We focus on thermodynamic integration, although the approach could be used with the exponential formula. We show how to determine the time required for equilibration and for converged averages in production. It is found that the slow step is the equilibration of the system, rather than the convergence of the free energy derivative in the equilibrated region.

## THEORY

For a sequential time series  $\{X\}$  of some property  $X$  that is an observable or a function of observables which can be evaluated from a molecular-dynamics or Monte Carlo simulation, the RCA function  $f(i, n)$  is defined as

$$f(i, n) = \bar{X}_i^{\text{RCA}} = \frac{1}{i} \sum_{j=n-i+1}^n X_j, \quad (1)$$

where  $X_j$  is the  $j$ th value in the complete data set  $\{X\}$  which contains  $n$  points ordered in the forward direction (i.e., that of the simulation). From Eq. (1), the RCA function at point  $i$  is the average of  $i$  data points starting with the *last* frame (i.e., the  $n$ th frame) of the trajectory. In free energy simulations,  $X$  is either the energy derivative (in thermodynamic integration) or the exponentially weighted energy difference (in exponential averaging). Using the RCA, we are able to find the minimum simulation time required to obtain approximately converged results for the ensemble average of  $X$ . The essential point is to determine the total data set length  $n$ , starting from an initial time, needed to obtain  $i$  stationary points which fulfill the chosen requirement for convergence.

### Determination of the segment length $\tau$

Since the calculated values of the variables obtained at each step of a molecular-dynamics simulation are highly correlated, a block-averaging method is used to coarse-grain the simulation results, such that sequential points become uncorrelated.<sup>26,27</sup> A set of values for  $\{X\}$  consisting of  $n$  steps is coarse-grained by dividing it into sequential, nonoverlapping blocks of the same size. If each block (segment) has length  $\tau$ , the set  $\{X\}$  is divided into  $l$  segments, where  $l = n/\tau$ ; for each segment, the average value is a member of a new, coarse-grained data set  $\{Y\}$  consisting of values  $Y_i$ . Starting in reverse (i.e., from the end of the simulation), we have

$$Y_i = \frac{1}{\tau} \sum_{j=n-i\tau+1}^{n-(i-1)\tau} X_j. \quad (2)$$

The value of  $\tau$  is determined empirically such that the members of set  $\{Y\}$  are statistically independent, and therefore, normally distributed; i.e., their deviations from the average  $\bar{Y}$  correspond to a random distribution. The method of “statistical inefficiency  $\Phi$ ”<sup>30</sup> is used to determine the minimum value (i.e., lower bound) for  $\tau$ . In this method, using the complete simulation data set  $\{X\}$ , one calculates the measure  $\Phi$

$$\Phi \equiv \frac{\tau \cdot \sigma^2(Y)_\tau}{\sigma^2(X)} \quad (3)$$

for a series of values of  $\tau$ . Here,  $\sigma^2(X)$  is the variance of the distribution  $\{X\}$ , and  $\sigma^2(Y)_\tau$  is the variance in the block average of  $X$  where the block length is  $\tau$  [Eq. (2)]. Once  $\tau$  is large enough that successive values of  $Y_i$  are statistically independent, the variance of the block-average distribution  $\{Y\}_\tau$  approximates the constant fold of the variance of the underlying distribution  $\{X\}$  divided by  $\tau$ , i.e.,  $\sigma^2(Y)_\tau \approx \Phi(\sigma^2(X)/\tau)$ . Thus, for values of  $\tau$  greater than this lower



bound, the  $\Phi$  in Eq. (3) is approximately constant, i.e., ( $\lim_{\tau \rightarrow \infty} \Phi = C$ ) and the members of the coarse grained series  $\{Y\}$  are statistically independent (i.e., uncorrelated).

Since the statistical inefficiency test assumes that the complete data set  $\{X\}$  of the simulation is a stationary series, we have to exclude the contribution from the nonequilibrium region. A possible rigorous procedure would start with the full simulation data, and would determine an initial minimum value for  $\tau$  as described above. Using this value of  $\tau$ , a coarse-grained data set  $\{Y\}$  would then be generated from the complete simulation according to Eq. (2). This coarse-grained data set would be subjected to normality checking (see below), which makes possible a provisional determination of the border between the equilibrating and equilibrated regions; i.e., the initial border is provisional because it is based on data that includes the nonequilibrium region. In the next step, the statistical inefficiency test would be applied *only* to the putative equilibrated region, and a new value of  $\tau$  would be determined, which in turn would be used to generate a new coarse-grained data set for the entire simulation. This process would be iterated to self-consistency; i.e., until the successive values of  $\tau$  are unchanged. Such a process would simultaneously determine the equilibrium value of  $\tau$  and the location of the border. However, we have found that the self-consistent procedure is not required to find the correct  $\tau$  value if initially  $\tau$  is chosen larger than the equilibrium value, which is true in all of the examples presented here. In this case, by regularly increasing the amount of the equilibrating region that is excluded (starting from the beginning of the simulation), the lower bound for  $\tau$  as determined by the statistical inefficiency test (i.e., the value of  $\tau$  at which  $\Phi$  converges to  $C$ ) will decrease monotonically until it converges. This way of determining  $\tau$  is faster than the self-consistent procedure described above. We illustrate the two procedures in the *Results* section.

### Determination of the border between the equilibrating and equilibrated regions

Following the generation of the reverse coarse-grained data set  $\{Y\}$  of size  $l$ , the sample mean  $\bar{Y}$  and variance  $s^2$  can be calculated for any subset of length  $k$  ( $k \leq l$ ),

$$\bar{Y} = \frac{1}{k} \sum_{j=1}^k Y_j, \quad (4)$$

$$s^2 = \frac{1}{k-1} \sum_{j=1}^k (Y_j - \bar{Y})^2.$$

Given the sample mean,  $\bar{Y}$ , calculated with Eq. (4) for a data set of length  $k$ , the true mean  $\langle Y \rangle$  falls within the range of values<sup>17</sup>

$$\langle Y \rangle = \bar{Y} \pm a_k \frac{s}{\sqrt{k}}, \quad (5)$$

with a confidence limit that depends on  $a_k$ , where  $a_k$  is a parameter of the Student's  $t$  distribution for  $k-1$  degrees of freedom. Typically, a confidence limit of at least 95% is

used. The desired precision of the free energy result is set in advance, i.e., the allowed error range is  $\pm \mu$ , such that

$$a_k \frac{s}{\sqrt{k}} \leq \mu. \quad (6)$$

For the calculation of a converged value, the efficiency is maximized when the condition of Eq. (6) is satisfied as an equality *and*, importantly, when the  $k$ th point, in the reverse coarse-grained data set, belongs to equilibrated region, while the  $(k+1)$ th point belongs to the equilibrating region; in this way, the minimum time required for equilibration is used. To determine the point  $k$ , a statistical criterion, which we refer to as “border checking,” is used.

### Border checking

#### Quantitative checking

The essential idea of the approach is based on the realization that an equilibrated data set, which is coarse-grained such that the members of the set are statistically independent, is normally distributed in most cases and that, once non-equilibrated (nonstationary) points are included, the normality of the equilibrium distribution is expected to be disrupted. This behavior is expected to be valid over the range of  $\lambda$  values, except at or near the end points when linear scaling of the potential is used. In this study, since a soft core van der Waals potential was used, a normal distribution should be valid for all the  $\lambda$  values; this is confirmed in the Results section. For a linearly scaled potential with positive or negative infinite free energy derivatives at the end points, extrapolation method can be used to avoid problems.<sup>6</sup> We remind the reader that points  $Y_i$  are included starting from the end of the trajectory, so that as the number of points included increases, one crosses the border between the equilibrated and nonequilibrated system. This is the essence of reverse cumulative averaging. The  $W$  test (also known as the Shapiro–Wilk test),<sup>28,29</sup> which is commonly used to check the normality of a distribution, is employed here to determine the first reverse coarse-grained data point where normality is no longer satisfied.

The  $W$  test is regarded as the best test for normality, particularly for relatively small samples (for which the method has been explicitly compared with other approaches,<sup>28</sup> although tables for samples up to 5000 are now available<sup>31</sup>). As we show below the size of a typical uncorrelated data set  $\{Y\}$  generated in free energy calculations is on the order of 100 points, of which only about 30 turn out to be in the equilibrated trajectory. In the  $W$  test, the null hypothesis (i.e., that the sample is taken from a normal population) is accepted with a certain confidence limit, which represents the lower-bound confidence on the normality, based on the value of  $W$ . To determine  $W$ , the  $k$  data points of the sample are ordered by magnitude, i.e.,  $Y_1 < Y_2 < \dots < Y_f < \dots < Y_k$ , and the slope  $b$  of the probability plot regression line is evaluated

$$b = \sum_{j=1}^m a_{k-j+1} (Y_{k-j+1} - Y_j), \quad (7)$$

where  $m = k/2$  if  $k$  is even, and  $(k-1)/2$  if  $k$  is odd. Using  $b$  and the sample variance  $s^2$ , the test statistic  $W$  is defined as

$$W = \frac{b^2}{s^2(k-1)}. \quad (8)$$

The coefficients  $a_{k-j+1}$ , which depend only on the size  $k$  of the sample, and the fractiles of  $W$  are listed in Ref. 28. The coefficients are chosen such that the deviation of  $W$  from unity is a measure of the non-normality of the distribution. To use the  $W$  test to determine the location of the border region,  $W$  is calculated as a function of  $k$  (i.e., the size of the reverse coarse-grained data set that is part of the equilibrated region), and for each value of  $k$ , the confidence with which the null hypothesis is accepted is determined by using a table that matches the calculated  $W$  value with its corresponding confidence.<sup>28</sup> As the RCA is extended to larger  $k$  so as to include earlier points in the trajectory, it eventually crosses the border to the equilibrating region. This causes the confidence limit to decrease because data in equilibrating region destroys the normality to some extent. As we discuss in the *Results* section, the “border” value of the confidence limit is set at 90%, i.e., the border is defined as the location where the confidence limit in the normality of  $\{Y\}$  falls below 90%; we note that 80% is usually regarded as a satisfactory confidence limit,<sup>28</sup> but we use 90% in the present work as a more conservative criterion.

### Qualitative checking

Border checking can also be done in a more simple and qualitative way by monitoring the plot of  $\langle Y \rangle$  versus  $k$ . Given that the same confidence holds for any two continuous subsets of data  $(Y_1, Y_2, Y_3 \dots Y_p)$  and  $(Y_1, Y_2, Y_3 \dots Y_q)$  where  $p < q \leq k$ , according to the Student's  $t$  distribution, the ranges most probably containing the expectation value (i.e., the confidence intervals) are  $[\bar{Y}_p - a_p(s_p/\sqrt{p}), \bar{Y}_p + a_p(s_p/\sqrt{p})]$  and  $[\bar{Y}_q - a_q(s_q/\sqrt{q}), \bar{Y}_q + a_q(s_q/\sqrt{q})]$ , respectively, where  $a_q, a_p$  correspond to a certain confidence given  $q-1$  and  $p-1$  degrees of freedom, respectively. Our general philosophy in this work is that the later points in the trajectory are more likely to be in the equilibrated region than earlier points. We also recognize that the magnitude of the confidence interval for the  $p$  subset constitutes an upper bound to that of the confidence interval for the converged data set, insofar as the amount of data contained in  $p$  may not be sufficient for convergence of  $\langle Y \rangle$ ; moreover, we recognize that, as long as the extra data contained in  $q$  relative to  $p$  remains in the equilibrated region, the size of the confidence interval for  $q$  will very likely be (1) smaller than that for  $p$  and (2) contained entirely within the interval of the latter. More precisely, given that the  $(Y_1, Y_2, Y_3 \dots Y_p)$  data set is thought to be located in the equilibrated region, we consider the larger  $(Y_1, Y_2, Y_3 \dots Y_q)$  data set; if the entirety of this data set, i.e., including the extra  $(Y_{p+1}, \dots, Y_q)$  data, still remains within the equilibrated region, then the likelihood that any part of the range  $[\bar{Y}_q - a_q(s_q/\sqrt{q}), \bar{Y}_q + a_q(s_q/\sqrt{q})]$  falls outside of the range  $[\bar{Y}_p - a_p(s_p/\sqrt{p}), \bar{Y}_p + a_p(s_p/\sqrt{p})]$  is very low, unless the some of the extra  $(Y_{p+1}, \dots, Y_q)$  data comes from equilibrating

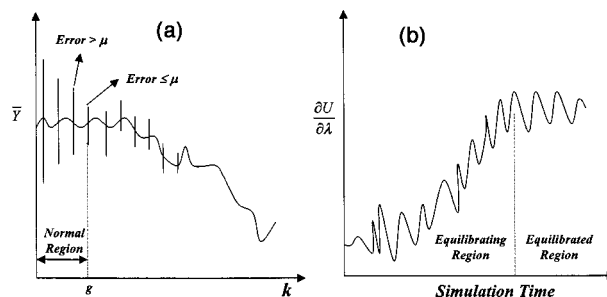


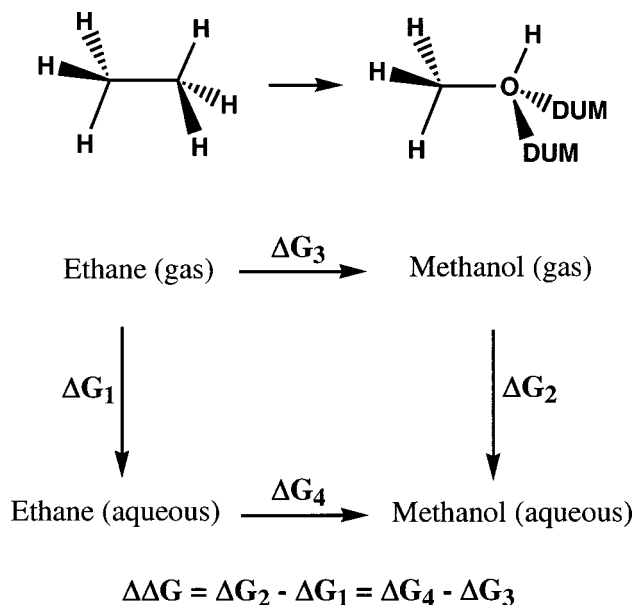
FIG. 1. Schematic diagrams illustrating the use of the RCA and FCA to determine the equilibration and production times. (a) shows the development in the RCA of a given measure (e.g., in this study,  $\partial U / \partial \lambda$ ) for a coarse-grained data set  $\{Y\}$ ; i.e., the value of  $\bar{Y}$  [see Eq. (3)] vs the index  $k$ , which indicates the length of the subset used to calculate  $\bar{Y}$ , is plotted. The point  $k'$  is determined to be the border between the equilibrating region (which in this diagram falls to the right of  $k'$ ) and the equilibrated region (which falls to the left of  $k'$ ). Since  $\{Y\}$  is an uncorrelated data set, it is normally distributed and its standard error  $\pm \mu$  can be determined using the Student's  $t$  distribution. (b) shows the forward development of the same measure  $\partial U / \partial \lambda$  as a function of simulation time. The border  $k'$  is shown, as in (a). We note that if the FCA is started from any point prior to the border, the resulting average is contaminated by inclusion of nonequilibrium data.

region. Thus a qualitative criterion for determining the border between the equilibrating and equilibrated regions is the following: as the RCA includes more points earlier in the trajectory (i.e.,  $p \rightarrow q$ ), the region at which any part of the confidence intervals for  $q, q+1$ , etc., begin to fall outside of the confidence intervals for  $q-1, q-2$ , etc., signals the crossing of the border, i.e., that data from the equilibrating region is now being included in the RCA. However, we note that the intervals associated with the Student's  $t$  test hold only with a certain confidence, i.e., they do not hold absolutely, but rather probabilistically, and thus exceptions are possible. As a result, this approach ultimately is qualitative, though as we show in the *Results*, it is useful in giving an initial picture of the convergence and equilibration periods.

The RCA approach is summarized in the schematic diagram in Fig. 1(a), which contrasts it with the FCA [Fig. 1(b)]. For the reverse coarse-grained data set  $\{Y\}$ , the simulation is approximately converged when the series from  $i = 1$  to  $k$  meets the following two criteria: (1) The data sets  $(Y_1), (Y_1, Y_2), \dots, (Y_1, Y_2, \dots, Y_k)$  all satisfy border checking done using the normality test; (2) the data set  $(Y_1, Y_2, \dots, Y_k)$  has an error range of  $\leq \mu$  [Eq. (6)], as shown in Fig. 1(a). The essential point is that reverse cumulative averaging (RCA) uncouples the convergence region from the equilibration region, while the forward cumulative average (FCA) starting from an arbitrary point can include uncertain information, as shown in Fig. 1(b).

### COMPUTATIONAL METHODS

To illustrate the theory, a free energy simulation of the alchemical change from ethane to methanol in aqueous solution was performed (see Scheme 1). The potential energy function for the simulation has the general form



Scheme 1. Thermodynamic cycle for determining the relative solvation free energies of ethane and methanol. The alchemical transformation  $\Delta G_3$  describes the transition from ethane in vacuum to methanol in vacuum. The alchemical transformation  $\Delta G_4$  describes the transition from ethane in aqueous solution to methanol in aqueous solution. The latter transformation is simulated in the present work. In the inset, the initial and final structures used in the free energy simulation are shown.

$$U(\mathbf{q}_s, \mathbf{q}_{\text{env}}, \lambda) = (1 - \lambda)U_A(\mathbf{q}_s, \mathbf{q}_{\text{env}}) + \lambda U_B(\mathbf{q}_s, \mathbf{q}_{\text{env}}) + U_{\text{env}}(\mathbf{q}_{\text{env}}), \quad (9)$$

where  $U_A$  and  $U_B$  are the potential functions of the individual solutes, which share the same set of coordinates  $\mathbf{q}_s$  (describing the solute structure) and  $\mathbf{q}_{\text{env}}$  (describing the solvent), and  $U_{\text{env}}$  is the potential functions of the solvent. The potential function in Eq. (9) describes an alchemical transformation in which one solute (unique part of A) is mutated to another solute (unique part of B) in the same solution, by varying  $\lambda$  from zero to one. By construction, we have  $U_A(\mathbf{q}_s, \mathbf{q}_{\text{env}}) = U_{AA}(\mathbf{q}_s) + U_{A,\text{env}}(\mathbf{q}_s, \mathbf{q}_{\text{env}})$  and  $U_B(\mathbf{q}_s, \mathbf{q}_{\text{env}}) = U_{BB}(\mathbf{q}_s) + U_{B,\text{env}}(\mathbf{q}_s, \mathbf{q}_{\text{env}})$ , where  $U_I$  includes the self-energy of the unique part of state I,  $U_{II}(\mathbf{q}_s)$ , plus its interaction with the environment  $U_{I,\text{env}}(\mathbf{q}_s, \mathbf{q}_{\text{env}})$ ; by contrast,  $U_{\text{env}}$  is solely the energy of the environment. Since the two solutes share the same coordinates  $\mathbf{q}_s$ , the implementation described by Eq. (9) is an example of a single-topology method.<sup>32</sup>

The parameters for ethane and methanol were taken from work of Boresch and Karplus,<sup>32</sup> and water was treated by the modified TIP3P model.<sup>33,34</sup> The van der Waals interactions were switched to zero between 8 and 12 Å, and the electrostatic interactions were treated with a constant dielectric of unity and with a shifted cutoff at 12 Å.<sup>35</sup> The system consisted of a cubic periodic box 31 Å on a side; it contained the solute and 997 water molecules. Molecular-dynamics simulations were done using the leapfrog integrator at a constant temperature of 300 K and a constant pressure of 10<sup>5</sup> Pa (1 atm) based on the Berendsen method.<sup>36</sup> A time step of 1 fs was used.

Free energy simulations were performed using thermodynamic integration<sup>4</sup> with the single-topology method in the

PERT module (B. Brooks, unpublished) of the CHARMM program.<sup>37</sup> The Helmholtz free energy difference between states A and B,  $\Delta A$ , is calculated using

$$\Delta A = \int_0^1 \frac{\partial A}{\partial \lambda} d\lambda = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda'} d\lambda' = \int_0^1 \langle U_B - U_A \rangle_{\lambda'} d\lambda' \quad (10)$$

where  $\langle \cdots \rangle_{\lambda'}$  is obtained as an ensemble average over the system corresponding to the potential  $U(\lambda)$  at the given value of  $\lambda'$  in Eq. (9); the rhs of the equality corresponds to linear scaling of the potential, as in Eq. (9). A series of “windows” are simulated at fixed  $\lambda$  values, and the resulting plot of the free energy derivative is numerically integrated. A total simulation time of 7.2 ns was used for the entire transformation; different simulation times were used for each window as determined by the reverse cumulative averaging algorithm. In changing the hydrogens in ethane to dummy atoms for methanol, the bond and angle terms applying to the dummy atoms were kept the same as in ethane.<sup>32,38</sup> A soft-core potential was used to deal with van der Waals end-point singularity problem;<sup>39</sup> this was introduced into the CHARMM program by S. Boresch (unpublished). The free simulation was initiated at  $\lambda=0.5$  and performed in two directions to  $\lambda=0$  and  $\lambda=1$ ; this starts with the stable system and permits two sets of calculations to be done in parallel. The gas-phase free energy difference between ethane and methanol ( $\Delta G_3$  in Scheme 1) was computed as in Ref. 32 by performing quasi-harmonic analyses on ethane and methanol individually; this gave a satisfactory result as shown in part A of Fig. 6 in Ref. 32.

## RESULTS AND DISCUSSION

We describe here the results obtained in the alchemical simulation of ethane to methanol in aqueous solution ( $\Delta G_4$  in Scheme 1).

### Generation of the reverse coarse-grained data set

The reverse coarse-grained data set was generated by following the self-consistent procedure described in *Theory* section. The aim is to determine the segment length  $\tau$  to be used in the coarse-graining of the equilibrated region to obtain uncorrelated data from which valid convergence statistics can be obtained. As expected from the *Theory* section, the segment length  $\tau$  and the inefficiency value  $\Phi$  are found to be sensitive to the amount of nonequilibrium data included in the statistical inefficiency test. The minimum length  $\tau$  for each equilibrated data set of  $\lambda$  to be uncorrelated, based on the “statistical inefficiency” test, is summarized in Table I; the value of  $\tau$  used for production was set equal to 8 ps, slightly larger than any of the values in Table I.

Figure 2 illustrates the use of the inefficiency test for the simulation with  $\lambda=1$ , which is run for 740 ps. Figure 2 plots the statistical inefficiency  $\Phi$  versus the segment length  $\tau$  for a series of calculations which differ only in the length of the initial equilibrating region that is excluded from the data set. The length of the latter region is incremented in units of 40 ps, i.e., 40 ps starting with the first calculated data point is excluded in the first calculation (not shown in Fig. 2), 80 ps



TABLE I. Summary of data in reverse cumulative averaging analysis of free energy simulation for ethane-methanol system.

$\lambda$	Length of equilibrating region <sup>a</sup> (ps)	Minimum $\tau$ (ps)	$\partial A/\partial \lambda$ with error (kcal/mol)
0.00	360	7.2	$13.03 \pm 0.28$
0.02	550	6.2	$12.47 \pm 0.34$
0.05	510	6.4	$11.35 \pm 0.27$
0.10	490	6.9	$9.36 \pm 0.28$
0.20	300	6.6	$5.39 \pm 0.31$
0.30	120	7.2	$1.28 \pm 0.34$
0.40	490	5.8	$-0.83 \pm 0.28$
0.50	270	5.8	$-1.02 \pm 0.21$
0.60	350	5.9	$0.28 \pm 0.25$
0.70	350	5.9	$-0.82 \pm 0.22$
0.80	250	6.4	$-5.39 \pm 0.34$
0.90	440	6.3	$-14.25 \pm 0.49$
0.95	430	6.9	$-19.65 \pm 0.44$
0.98	380	7.1	$-23.41 \pm 0.53$
1.00	540	7.3	$-26.19 \pm 0.52$

<sup>a</sup>As determined by the W test with a confidence limit of 90%.

in the second calculation (see Fig. 2), etc. In each calculation,  $\Phi(\tau)$  shows asymptotic convergence, as is typical for this measure.<sup>26</sup> We note that in Fig. 2, as more of the initial region is excluded, both the minimum segment length (i.e., the value of  $\tau$  at which  $\Phi$  converges) and the converged value of  $\Phi$  [i.e.,  $C$  in Eq. (3)] decrease. This is expected because the nonequilibrium process of relaxation following a change in  $\lambda$  is likely to involve lower-frequency motions than are found in the equilibrated region. Since such motions are correlated over longer periods, inclusion of the nonequilibrium data is expected to increase both the lower bound of  $\tau$  and the value of  $C$ , i.e., correlations over a longer time scale make the data set segments have a relatively larger variance. It is seen in Fig. 2 that when the initial 400 ps (or

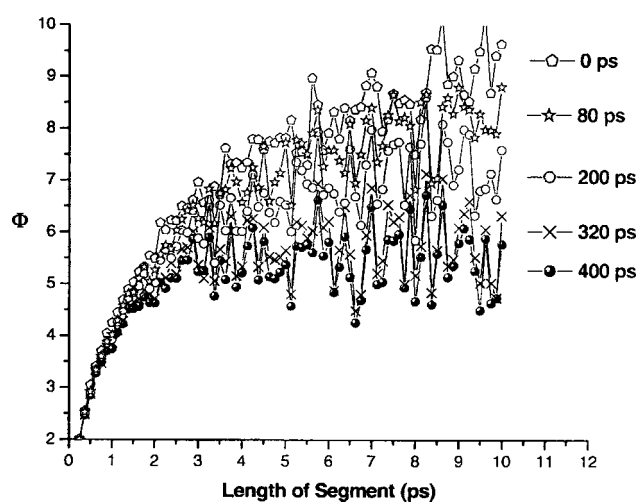


FIG. 2. Use of the statistical inefficiency test to determine the equilibrium segment length  $\tau$  over which the free energy derivative value decorrelates. As described in the text, the equilibrating region is progressively excluded using an iterative procedure. For the  $\lambda=1$  simulation, a series of calculations are shown, which differ based on the length of the initial region that is excluded from the total data in the calculation of  $\tau$ . The curve corresponding to each calculation is labeled using the length (in ps) of the excluded initial region. The minimum value of  $\tau$  converges when the excluded initial region is at least 400 ps.

more) of the simulation are excluded, the resulting  $\Phi(\tau)$  curves are similar and thus a converged minimum value of  $\tau$  has been reached. This indicates that the iterative procedure used to exclude the equilibrating region has converged. In our work, which includes more complex systems as well as the simulations described here, we have found this procedure to be adequate for determining the equilibrium segment length  $\tau$ . As described in *Methods*, a more precise (and more laborious) approach can be used, in which the increment by which the equilibrating region is extended following each step is determined by normality checking, using the value of  $\tau$  generated by that step, rather than by a regular incrementation, as used here. Given that the statistical inefficiency result using the regular (simple) incrementation (in which the lower bound for  $\tau$  converges with the initial 40 ps removed in the  $\lambda=1$  simulation) is more “tolerant” than the normality checking ( $W$  test), we conclude that the less expensive procedure is adequate. The border occurs after the initial 540 ps in the same simulation (see Table I); using that value, as in the self-consistent method,  $\tau$  is found to be the same.

According to the results using the statistical inefficiency test (Table I), about 8 ps is a suitable value for  $\tau$  to use for determining the statistics of the free energy derivatives at all  $\lambda$  values; i.e., this value is slightly larger than the set of minimum  $\tau$  values for all  $\lambda$  (the maximum calculated value of  $\tau$  is 7.3 ps for  $\lambda=1$ ). However,  $\tau=8$  ps is not so large a segment as to significantly reduce the efficiency of the calculation, relative to using the exact value for each  $\lambda$ . Having determined  $\tau$ , the reverse coarse-grained data set for use in border checking is generated according to Eq. (2).

### Determination of the border between the equilibrating and equilibrated regions, and convergence of the latter region

As described in the *Theory* section, there are two essential aspects to this phase of the analysis: (1) Determining the border between the equilibrating and equilibrated regions and (2) continuing the simulation until the data of the equilibrated region have converged to the desired precision. To fulfill condition (1), normality was examined, using the  $W$  test, as a function of the reverse process (i.e., starting with the final data point in the set  $\{Y\}$ ). As discussed in the *Theory* section, the confidence with which the null hypothesis of the  $W$  test is accepted is determined for each subset ( $Y_1 \dots Y_k$ ) of the reverse coarse-grained distribution, which is obtained by dividing the data into segments of length  $\tau$  in reverse direction and contains decorrelated data as determined from the “statistical inefficiency” analysis. The confidence limit calculated from the  $W$  test is a function of  $k$ , which we label  $\text{conf}_W(k)$ . Figure 3 shows the confidence limits calculated based on the  $W$  test for subsets ( $Y_1 \dots Y_k$ ) as a function of  $k$  for  $\lambda=1$ . After performing the  $W$  test and calculating the confidence limits for all of the subsets, a minimum confidence limit (MCL, called the “trust cutoff”) needs to be chosen to determine where normality is first broken. Here, it is defined so that, if  $\text{conf}_W(k-1) \leq \text{MCL}$  and  $\text{conf}_W(k) > \text{MCL}$ , then the point  $k$  is the place where normality is first broken. This point is taken to be the border between the equilibrating and equilibrated regions; i.e., it is assumed that

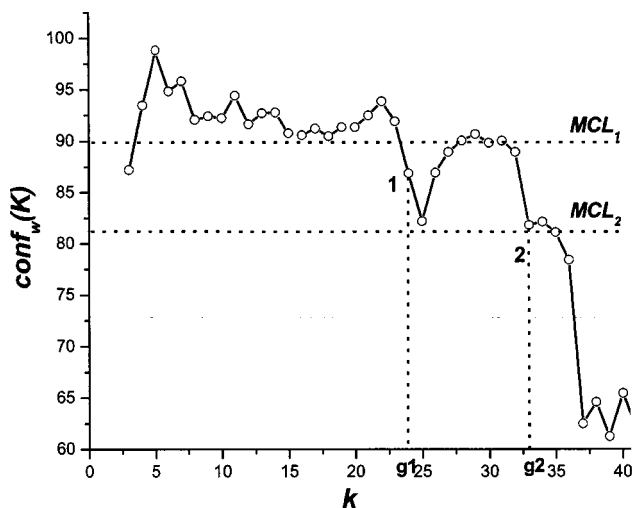


FIG. 3. Use of the  $W$  test to determine the border between the equilibrating and equilibrated (production) regions. For the  $\lambda=1$  simulation, using  $\tau=8$  ps, the confidence limit with which the null hypothesis of the  $W$  test is accepted [i.e.,  $\text{conf}_w(k)$ ] is plotted as a function of the subset size  $k$  of the coarse-grained data set  $\{Y\}$ . The minimum confidence limit (MCL) is set by the user; two MCLs are shown here; one at 90% ( $\text{MCL}_1$ ) and the other one at 82% ( $\text{MCL}_2$ ). The value of  $k$ , called  $g$ , at which the MCL is crossed is the border; thus the borders  $g_1$  and  $g_2$  are determined.

the contribution from nonequilibrium region is the reason for disrupting the normal distribution. An example of this procedure is given by Fig. 3. In this figure, for the  $\lambda=1$  simulation (in which the total simulation time is 740 ps), the subset ( $Y_1 \dots Y_k$ ) is started at  $t=740$  ps and is shown up to  $t=420$  ps (the region from  $t=420$  ps to  $t=0$  ps is not shown). The choice of trust cutoff influences the determination of the border; i.e., altering the trust cutoff (MCL) results in a different border. We have used a confidence limit of 90% as the trust cutoff for the  $W$  test. In the  $W$  test, for the initial subsets (e.g., the first points shown in Fig. 3), due to the small total number of points involved, the trust values obtained are unreliable<sup>28</sup> and we, therefore, ignore these points in determining the border. We note also that approximate normality may be re-established when some of the equilibrating data is included, as is seen in the region between borders 1 and 2 of Fig. 3. This region was judged to *not* be part of the real

equilibrated region because it is preceded by some points which are not in the equilibrated region based on the  $W$  test. The new normality is likely to result from several data points with small deviations that form a near-normal distribution. This shows that care has to be taken in choosing the border and that it is important to determine that the various criteria (inefficiency, normality) are consistent.

Having determined the border between the equilibrating and equilibrated regions, the last value of the RCA function counting from the end [Eq. (4)] in the equilibrium region is the optimal one to use because it is expected to have the smallest variance in equilibrium region. If the error of the result is not within the allowed range [ $\pm\mu$  in Eq. (6)], a decision to extend the simulation can be made. As examples of converged results, we show in Fig. 4(b) for several lambda values the actual equilibrium distribution of free energy derivative values and the expected normal distributions for the corresponding values of  $\mu$  listed in Table I. In Fig. 4(a), we show an example for the  $\lambda=1$  simulation where parts of the equilibrating region near the border are included. The more of this nonequilibrium region is included, the greater is the deviation of the distribution of values from a normal distribution (dash line bars) as shown in Fig. 4(a). The inclusion of more equilibrating data can generate transient pseudo normal distribution (filled bars), as shown in Fig. 4(a). The results do not form a real equilibrated data set because there are some subsets (black) that are not normally distributed.

Figure 5 shows the development of the reverse cumulative average, and its standard error [Eq. (6)] as a function of  $k$ , for the same simulation in the same region as is shown in Fig. 3. The borders determined in Fig. 3 are labeled “1” and “2,” and are correspondingly marked in Fig. 5. In Fig. 5, it can be seen that the standard error range of the RCA narrows continuously as  $k$  increases, and does so within both the upper and lower bounds of the ranges of the previous points, until the region of border “1” is reached, after which the error ranges begin to be outside the previous range on the high side of the average. As discussed in the *Theory* section, this effect is used as the criterion in the qualitative method of border checking to indicate the approximate location of the border. Given that border “1” was identified using the more

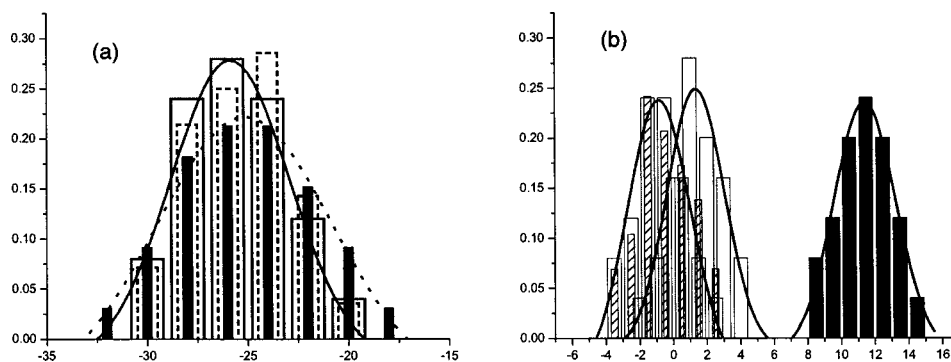


FIG. 4. (a) Distribution of the free energy derivative values over 200 ps in the converged region for the window simulations at  $\lambda=1$ , and other distributions that include differing amounts of data from the equilibrating region (in dash line bars and in filled bars). (b) Distribution of the free energy derivative values over 200 ps in the converged region for the window simulations at  $\lambda=0.05$ ,  $\lambda=0.3$ ,  $\lambda=0.7$ , and the expected (normal) distribution of values. For the simulations at  $\lambda=0.7$ , a second distribution is shown which includes some data from the equilibrating region; it can be seen that this distribution is no longer normal.



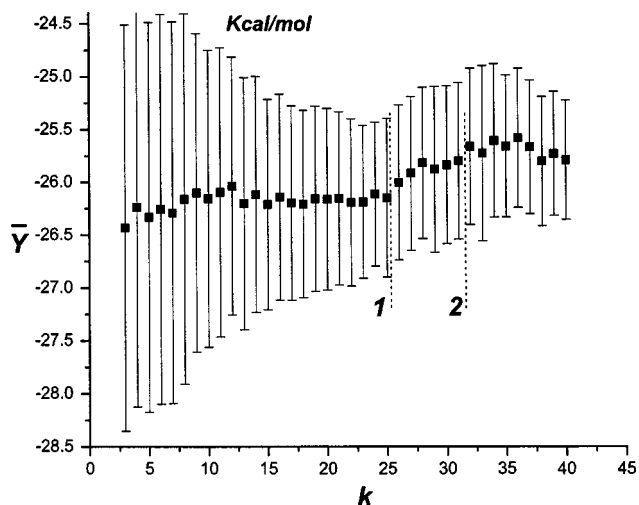


FIG. 5. Development of the RCA and its standard error range, the latter determined using the 99% confidence limit. The data shown is for the  $\lambda=1$  simulation using  $\tau=8$  ps, in Fig. 3. Also reproduced from Fig. 3 are the two borders “1” and “2.” As discussed in the text, near “1” the error bounds of the RCA begin to fall (at one end) outside of bounds of the previous values.

rigorous  $W$  test (Fig. 3), as described above, it is satisfying that it coincides with the more simple and qualitative method of border checking illustrated in Fig. 5. The result confirms that the qualitative method of border checking can be used

for an approximate estimate of the border. (The second border “2” was identified in Fig. 3 using a confidence limit for the  $W$  test of 82%.) In summary, the qualitative method of estimating the border agrees with the more rigorous method, as can be seen by comparing Figs. 3 and 4. In practice, it is not necessary to perform the  $W$  test at every step along the simulation; rather, qualitative criteria can be used to locate the likely border region to which the  $W$  test should be applied.

It is useful also to examine visually the trend of  $\partial A/\partial \lambda$  obtained by the RCA during a simulation, which essentially reproduces the results of the more quantitative statistical methods described above. In Fig. 6, we give several plots of the RCA starting from different points in a single simulation ( $\lambda=1$ ) that was run for 740 ps. Specifically, the RCA is calculated (1) starting from the end of the simulation (i.e., at  $t=740$  ps), and going in the reverse direction to  $t=244$  ps (a total length of 496 ps or 62 data points) (RCA points shown using circles); (2) starting from 56 ps before the end of the simulation, i.e., at  $t=684$  ps, and going to  $t=244$  ps (a total length of 440 ps or 55 data points) (RCA points shown using triangles); and (3) starting from 116 ps before the end (i.e., at  $t=628$  ps) and going to  $t=244$  ps (a total length of 384 ps or 48 data points) (RCA points shown using diamonds). In this figure, the progress of the RCA can be divided into three regions: The converging region (region A), the converged

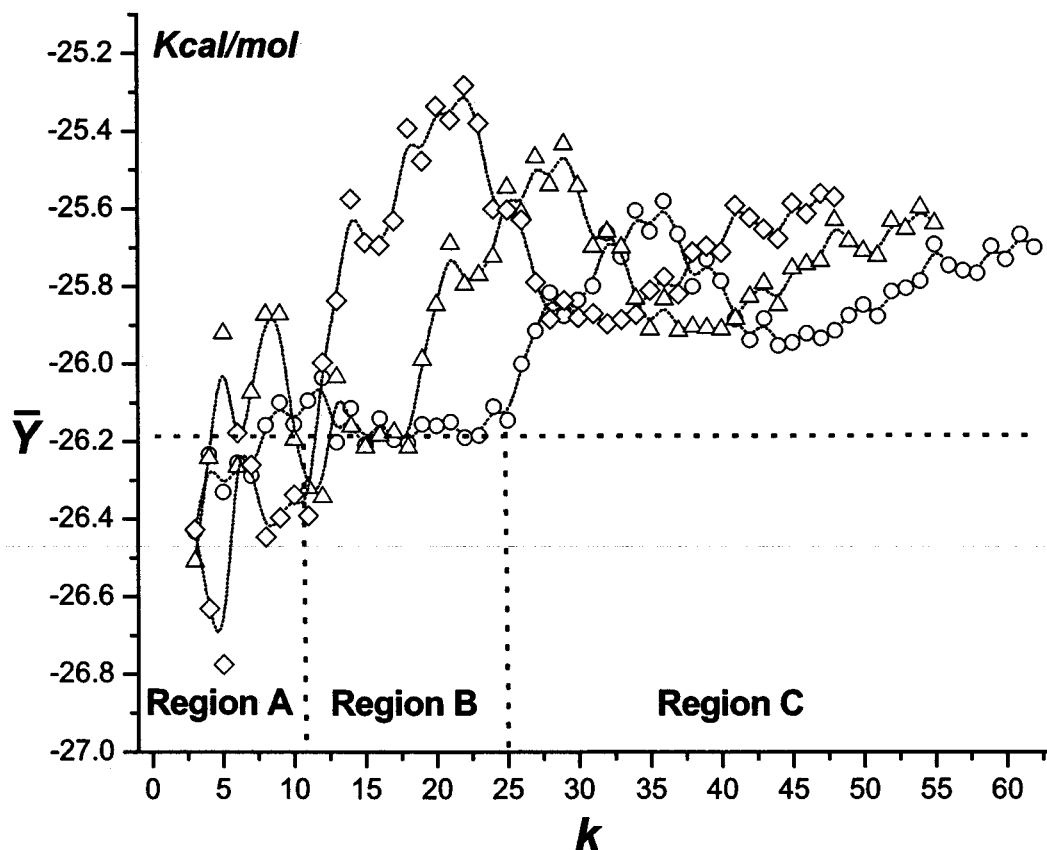


FIG. 6. Plots of the RCA for a single  $\lambda=1$  simulation of 740 ps, where the RCA is calculated (i) starting from the end of the simulation (i.e., at  $t=740$  ps) and extending over 440 ps in the reverse direction (RCA points shown using circles); (ii) starting from 56 ps before the end of the simulation (i.e., at  $t=684$  ps) and extending over 384 ps in the reverse direction (RCA points shown using triangles); (iii) starting from 112 ps before the end (i.e., at  $t=628$  ps) and extending over 328 ps (RCA points shown using diamonds). In this figure, the progress of RCA can be visually divided into three regions: The converging region (region A), the converged region (region B) and the equilibrating region (region C). The significance of each region is discussed in the text.

region (region B) and the equilibrating region (region C). In region A, RCA's from simulation subsets of different lengths fluctuate in the course of reaching a common border point where the RCA values from the three plots intersect; in region B, for the circle plot, the reverse cumulative average continues to fluctuate with small amplitudes around a relatively well-defined average values; in region C, the RCA's undergo dramatic transitions, which are caused by the lag of the configuration behind the Hamiltonian characteristic of the equilibrating region.

It can be seen in Fig. 6 that RCAs starting from different points meet for the first time at nearly the *same* position; i.e., where the averages of the free energy derivative converge. This is an expected result, since for normally distributed samples of different lengths, one expects that the average values will start to be the same in each case when a certain minimum number of points (namely, enough to meet the convergence requirement) is included in the average. As a result, the total time difference between two converged simulations of different lengths is roughly equal to the difference in the respective production times. We note that for each of the three different transitions (740, 684, and 628 ps), the equilibrium region begins approximately at the corresponding point ( $k \approx 25$  for the first,  $k \approx 18$  for the second, and  $k \approx 11$  for the third). The behavior shown here can be used as a simple qualitative criterion to monitor the RCAs; we have found that it is useful in providing an estimate of the likely border region. However, this needs to be checked by the quantitative methods. We note that the RCAs happen to meet subsequently in region C (the equilibrating region), as is also seen in Fig. 6; this is without physical significance, since this region is not at equilibrium.

### Free energy simulation of ethane-to-methanol in aqueous solution

The results of the free energy simulation of the alchemical transformation of ethane to methanol in solution obtained with RCA applied to thermodynamic integration are shown in Fig. 7; the error bars on the points represent the 99% confidence limit using Student's  $t$  distribution. We have collected data for 200 ps in the production region for all of the lambda values. The result is  $\Delta G_4 = -1.21 \pm 0.31$  kcal/mol; the overall error was estimated by calculating the difference between the integral of  $\partial A / \partial \lambda + \mu$  and the integral of  $\partial A / \partial \lambda - \mu$  [see Eq. (6)]. Combining the value of  $\Delta G_4$  with the gas phase free energy result  $\Delta G_3$ , which is 6.21 kcal/mol, slightly different from the one reported in Ref. 32 (5.70 kcal/mol), yields  $\Delta \Delta G = -7.42$  kcal/mol, which is close to the experimental result of  $-6.90$  kcal/mol.<sup>40</sup> The difference of the gas phase results is due to the use of different nonbond cutoffs and shifting of the electrostatic potential. We note that the present value is about 1.1 kcal/mol smaller in magnitude than a previous simulation done using the same parameters;<sup>32</sup> i.e. the solvation value obtained in Ref. 32 is  $-8.50$  kcal/mol, versus the value obtained here of  $-7.42$  kcal/mol. The difference is likely to arise from the fact that, although the parameters used were the same, the original study used a smaller water box (122 water molecules versus 997 water molecules used here) and a correspondingly

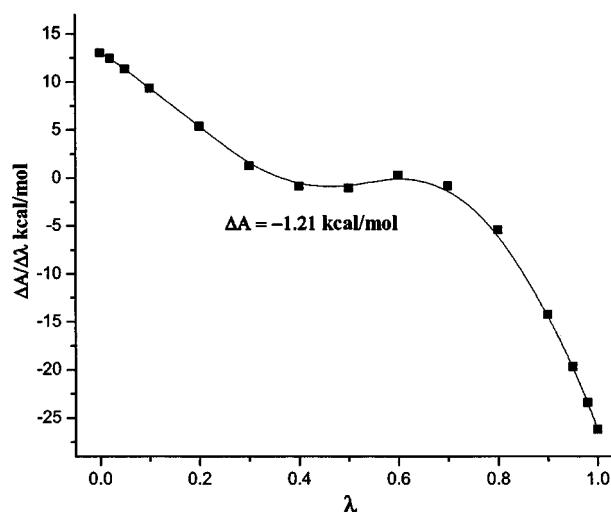


FIG. 7. Plot of the free energy derivative  $\partial A / \partial \lambda$  as a function of  $\lambda$  for the simulation of  $\Delta G_4$  (see Scheme 1). As discussed in the text, 200 ps of production were used for each  $\lambda$  value (shown using squares).

shorter cut-off radius (7.5 Å versus the value of 12.0 Å used here). As shown in Ref. 45, free energy results are sensitive to the cutoff for the electrostatic term, particularly when atom-based shifting is used. Also it is expected that significantly shorter equilibration and production periods are required under the original conditions. In the original study, the equilibration periods varied (depending on the  $\lambda$  value) from 5 to 10 ps, and the production periods ranged from 10 to 60 ps.

To compare the RCA method with the conventional approach of forward cumulative averaging (FCA) directly, we performed the following calculations. Using the simulation data set generated above, we assumed an equilibration period of 40 ps (which we note is quite long compared with those typically reported in the literature) and a production period of 200 ps for all lambda values. We then, as usual, integrated the free energy derivative to determine the free energy difference  $\Delta G_4$ , which was found to be 0.02 kcal/mol. This value is 1.23 kcal/mol smaller than the result obtained by RCA. We note that this difference in the results obtained by the two averaging methods, though substantial, underestimates the disparity between FCA and RCA, since the data set used for FCA is biased toward more equilibrated data because it was generated using the RCA procedure with the associated statistical criteria. As a result, the simulation at each  $\lambda$  value, is already more equilibrated than it would be if all the lambda simulations were each conducted serially for only 240 ps each.

### Analysis of equilibration

With 200 ps of production for each lambda value, the error is in the range of 0.2–0.6 kcal/mol, as is shown in Table I, and varies with the lambda value as shown in Fig. 7. The overall difference for each lambda value is small, in part because the single-topology method and a soft-core potential were used to enforce reasonable end-point conformations. It is expected that, with the dual-topology method or in the absence of a soft-core treatment, the end-point regions would

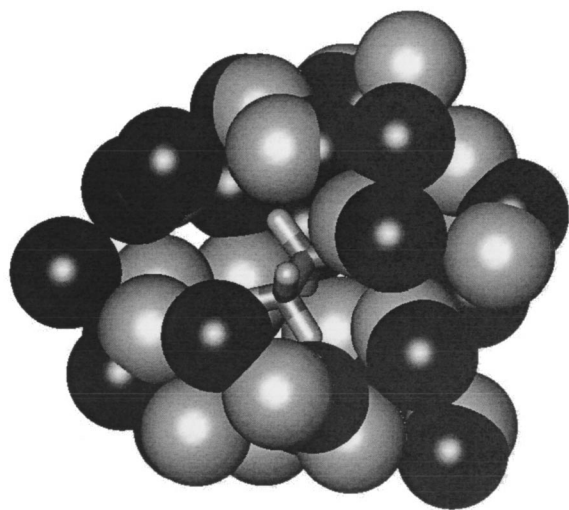


FIG. 8. Snapshots taken from the last recorded frame in the simulation at  $\lambda=0.2$  (white) and at  $\lambda=0.1$  (black). The region of the solute and the first solvation shell is shown. It can be seen that the first solvation shell is much tighter in  $\lambda=0.2$ , compared with  $\lambda=0.1$ , particularly around the region (stick model) of the solute that undergoes the alchemical change (i.e., from C–H to O–H).

require significantly longer production times to converge to the same level of precision as the central region.

The equilibrating regions for the different lambda values determined using the same  $W$  test trust cutoff value (90%) varied from 120 to 550 ps (see Table I); the distribution of times does not show a pattern based on the  $\lambda$  value. In the solution system, usually, there are three factors that govern the equilibration: Solute internal conformational change, solvent rearrangement caused by the solute conformation changes and solvent rearrangement caused by the change in the intrinsic property of the solute with  $\lambda$ . Because of the symmetry of ethane and methanol, solute internal conformational changes are expected to contribute little to the equilibration difference for different lambda values. However, solvent rearrangement coupled with the conformational change can make a larger difference when the system moves from one torsional angle to another. In addition to the solvent rearrangements caused by the conformational changes of the solute that may be generated when  $\lambda$  is changed, the fact that physical properties of the solute depend on  $\lambda$  also can play a role since the solute becomes less hydrophobic and more hydrophilic as  $\lambda$  increases (i.e., as ethane is transformed into methanol). As a result, the solvent reorientation will depend on  $\lambda$ . When  $\lambda$  is changed from 0.2 to 0.1, the large difference of the free energy derivative (i.e., from 5.39 kcal/mol versus 9.36 kcal/mol) indicates that solvent molecules have to undergo significant reorientations, which can cause the long equilibration time (490 ps). This can be seen in Fig. 8, which shows snapshots taken from the last frames of the simulation at  $\lambda=0.2$  and that at  $\lambda=0.1$ , respectively. From this figure, it is evident that the first solvation shell is much “tighter” in the simulation at  $\lambda=0.2$ . These results are in accord with the recent stress on the possible importance of a solvation coordinate for describing a reaction in solution.<sup>44</sup>

The above examples indicate that the equilibration process is rather complex and may require a long time, since it

may involve the equilibration of the solvent configuration and the coupling with the solute conformations. This is why the border checking method described here is so important for efficient and precise free energy simulations. We have found this to be true for treating very large complex systems, such as the calculation of absolute binding affinity of F<sub>1</sub>-ATPase for ATP, where the RCA approach has been found to be essential for obtaining meaningful results.

It is interesting to mention also free energy simulations based on “*ab initio*” molecular dynamics, in which very short simulations (2 to 3 ps of production without clear equilibration) have been used and excellent results obtained.<sup>41</sup> Given the present analysis, such an approach can work only if the alchemical transformation is made when the solvent configuration is such that the correct free energy is obtained. How such a solvent configuration can be found, in the absence of a knowledge of the desired answer, is not clear. This important question concerning “*ab initio*” molecular dynamics is not discussed in published work, as far as we have been able to find.

## CONCLUSIONS

A simple and practical method, called reverse cumulative averaging (RCA), is presented for locating the equilibrating and converging regions in free energy simulations. The essential point of the present method is the simple idea that cumulative averages are calculated and followed in reverse; i.e., the cumulation of data is done starting from the last point rather than the first point of the simulation, as in previous work. In addition, statistical criteria are used to determine the simulation time to obtain uncorrelated data segments and to monitor the border between the equilibrating and equilibrated region. Both of these are required to obtain converged free energy simulations with meaningful error bars. The RCA approach is illustrated here by applying it to the well-studied ethane-to-methanol alchemical transformation in solution. Both rigorous and more approximate statistical criteria are shown to be useful. Convergence to within  $\pm 0.2$  to 0.6 kcal/mol (depending on the  $\lambda$  value) requires from 210 to 790 ps equilibration and 200 ps of production. The large variation in the equilibration time is indicative of the importance of the present analysis.

The “coming of age” of free energy simulations<sup>1</sup> is leading to their utilization for a rapidly growing set of processes in complex systems. It is essential for such applications that the results of the free energy simulations be accurate and precise enough to answer the questions of interest. We believe that the RCA method with statistical criteria for determining the equilibration time and correlation time of the simulation data sets is an important step forward in free energy simulation methods. The accuracy of the results (systematic error) still depends on the available potential energy functions, whether of the quantum mechanics or molecular mechanics type. We have found reverse cumulative averaging to be effective in a free energy simulation of the differences in equilibrium of the hydrolysis reaction in the catalytic subunits of F<sub>1</sub>-ATPase,<sup>42</sup> as well as in a model study concerned with the utility of lambda dynamics.<sup>43</sup>



## ACKNOWLEDGMENTS

Discussions with Dr. Stefan Boresch, Dr. Ioan Andricioaei, Dr. Qiang Cui and Dr. Paul Maragakis, and computing support from Dr. Robert Yelle, are gratefully acknowledged. We thank Dr. Stefan Boresch, in particular, for a careful reading of the manuscript and for helpful comments on the ethane to methanol example. M.K. thanks G. Ciccotti for an illuminating discussion of free energy simulations with *ab initio* molecular dynamics. This work was supported in part by a grant from the National Institutes of Health and by a grant of computer time from National Energy Resource Supercomputing Center.

- <sup>1</sup>T. Simonson, G. Archontis, and M. Karplus, *Acc. Chem. Res.* **35**, 430 (2002).
- <sup>2</sup>C. L. III Brooks, M. Karplus, and M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics; Advance in Chemical Physics LXXI* (Wiley, New York, 1988).
- <sup>3</sup>P. A. Kollman, *Chem. Rev.* (Washington, D.C.) **93**, 2395 (1993).
- <sup>4</sup>J. G. Kirkwood, *J. Chem. Phys.* **3**, 300 (1935).
- <sup>5</sup>R. Zwanzig, *J. Chem. Phys.* **22**, 1420 (1956).
- <sup>6</sup>T. Simonson, G. Archontis, and M. Karplus, *J. Phys. Chem. B* **41**, 8347 (1997).
- <sup>7</sup>T. Simonson, G. Archontis, and M. Karplus, *J. Phys. Chem. B* **103**, 6142 (1999).
- <sup>8</sup>S. Boresch and M. Karplus, *J. Mol. Biol.* **254**, 801 (1995).
- <sup>9</sup>S. Boresch, G. Archontis, and M. Karplus, *Proteins: Struct., Funct., Genet.* **20**, 25 (1994).
- <sup>10</sup>J. Wang and P. A. Kollman, *J. Am. Chem. Soc.* **120**, 11106 (1998).
- <sup>11</sup>A. Pathiaseril and R. J. Woods, *J. Am. Chem. Soc.* **122**, 331 (2000).
- <sup>12</sup>O. Michlelin and M. Karplus, *J. Mol. Biol.* **324**, 547 (2002).
- <sup>13</sup>J. W. Essex, P. L. Severance, J. Tirado-Rives, and W. L. Jorgensen, *J. Phys. Chem.* **101**, 9663 (1997).
- <sup>14</sup>V. Uspensky, *Introduction to Mathematical Probability* (McGraw-Hill, New York, 1937).
- <sup>15</sup>H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- <sup>16</sup>G. Archontis, T. Simonson, D. Moras, and M. Karplus, *J. Mol. Biol.* **275**, 823 (1998).
- <sup>17</sup>M. Pagano and K. Gauvreau, *Principles of Biostatistics*, 2nd ed. (Duxbury, Pacific Grove, CA, 2000).
- <sup>18</sup>D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.* **91**, 7831 (1989).
- <sup>19</sup>D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.* **90**, 2460 (1989).
- <sup>20</sup>T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, *J. Chem. Phys.* **85**, 6720 (1986).
- <sup>21</sup>R. H. Wood, *J. Phys. Chem.* **95**, 4838 (1991).
- <sup>22</sup>J. Hermans, *J. Phys. Chem.* **95**, 9029 (1991).
- <sup>23</sup>C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- <sup>24</sup>D. A. Hendrix and C. Jarzynski, *J. Chem. Phys.* **114**, 5974 (2001).
- <sup>25</sup>G. Hummer, *Mol. Simul.* **28**, 81 (2002).
- <sup>26</sup>M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1989).
- <sup>27</sup>S. K. Schiferl and D. C. Wallace, *J. Chem. Phys.* **83**, 5203 (1985).
- <sup>28</sup>S. S. Shapiro and M. B. Wilk, *Biometrika* **52**, 591 (1965).
- <sup>29</sup>S. S. Shapiro, M. B. Wilk, and H. J. Chen, *J. Am. Stat. Assoc.* **63**, 1343 (1968).
- <sup>30</sup>R. Friedberg and J. E. Cameron, *J. Chem. Phys.* **52**, 6049 (1970).
- <sup>31</sup>P. Royston, *Stat. Comput.* **2**, 117 (1992).
- <sup>32</sup>S. Boresch and M. Karplus, *J. Phys. Chem. A* **103**, 119 (1999).
- <sup>33</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Phys. Chem.* **79**, 926 (1983).
- <sup>34</sup>E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
- <sup>35</sup>P. J. Steinbach and B. R. Brooks, *J. Comput. Chem.* **15**, 667 (1994).
- <sup>36</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- <sup>37</sup>B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- <sup>38</sup>S. Boresch and M. Karplus, *J. Phys. Chem. A* **103**, 103 (1999).
- <sup>39</sup>M. Zacharias, T. P. Straatsma, and J. A. McCammon, *J. Chem. Phys.* **100**, 9025 (1994).
- <sup>40</sup>A. Ben-Naim and Y. J. Marcus, *J. Chem. Phys.* **81**, 2016 (1984).
- <sup>41</sup>J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, and M. Sprik, *J. Am. Chem. Soc.* **124**, 6594 (2002).
- <sup>42</sup>W. Yang, Y. Q. Gao, Q. Cui, J. Ma, and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 874 (2003).
- <sup>43</sup>R. Bitetti-Putzer, W. Yang, and M. Karplus, *Chem. Phys. Lett.* **377**, 633 (2003).
- <sup>44</sup>P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- <sup>45</sup>M. Brunsteiner and S. Boresch, *J. Chem. Phys.* **112**, 6953 (2000).