JOURNAL OF COMPUTATIONAL PHYSICS 68, 237–248 (1987)

Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias

MIHALY MEZEI

Department of Chemistry, Hunter College, C.U.N.Y., New York, New York 10021

Received May 7, 1985; revised December 30, 1985

DEDICATED TO THE MEMORY OF ÉVA MEZEI

A self-consistent procedure is described for the determination of the non-Boltzmann bias for the umbrella sampling technique of Valleau, Patey, and Torrie. The new procedure offers more reliable results with less human interference. The problem of matching several differently normalized probability distributions on overlapping domains has been treated in detail. The algorithm has been tested on the calculation of the solvent contribution to the free energy difference between the C_7 and α_R conformation of the alanine dipeptide, treated earlier with the conventional umbrella sampling technique.

1. INTRODUCTION

Non-Boltzmann or umbrella sampling [1, 2] is frequently used in computer simulation when the sampling of a specific region of the configuration space is required that may not be sampled adequately in a direct Boltzmann sampling. The technique calls for a sampling with a modified potential. The major difficulty in the successful application of umbrella sampling is the search for the appropriate modification, usually determined by trial and error. The purpose of this paper is to demonstrate the feasibility of generating a potential modification from the simulation itself in a self-consistent manner that performs the required sampling more efficiently than the ones obtained by trial and error.

2. BACKGROUND

The successes of both the Monte Carlo and molecular dynamics methods in describing the liquid state of matter is based on the ability of these methods to restrict the sampling of the configuration space to the extremely small fraction that contributes significantly to most properties of interest. However, for calculations of free energy differences, adequate sampling of a much larger subspace is required. Also, for the related problem of potential of mean force calculations the structural parameter along which the potential of mean force is calculated has to be sampled

0021-9991/87 $3.00

uniformly. These goals can be achieved by performing the calculation using an appropriately modified potential energy function V

$$V = E + E_W \tag{1}$$

where E is the original potential function and E_W is the modification [1, 2, 3]. The subscript W indicates that the Boltzmann factor is modified by a "weighting function" $\exp(-E_W/kT)$ when V is used. From a calculation performed with V for any quantity Q the Boltzmann average $\langle Q \rangle_E$ can be recovered as

$$\langle Q \rangle_E = \langle Q \exp(E_W/kT) \rangle_V / \langle \exp(E_W/kT) \rangle_V \tag{2}$$

where the subscript V indicates that the average is obtained with the modified potential of Eq. (1). While there is no serious theoretical limitation on the choice of E_W, practical considerations limit it to functions that do not show too large variation over their domain since their exponential is involved in Eq. (2) [4].

Generally, E_W is a function of a few parameters λ only, either energetic (when the free energy difference between two types of particles is computed [2, 5, 6]) or structural (when the potential of mean force is calculated [1, 3, 7–10]). $E_W(\lambda)$ is to be chosen in such a way that the domain of λ is adequately sampled.

The usual way of determining the non-Boltzmann bias $E_W(\lambda)$ proceeds by trial and error. At first a trial $E_W(\lambda)$ is assumed, a short simulation is run and the region sampled is examined. Next, $E_W(\lambda)$ is modified to give larger weight to undersampled or unsampled regions of the λ-space and a new run is performed. This process is repeated until an $E_W(\lambda)$ is found the samples the desired region of the λ-space. In most of the cases it was found practical to target subspaces of the λ-space for a calculation, necessitating the "matching" of the obtained probability distributions, since results from the calculations in different λ-subspaces are undetermined up to a normalization factor. This matching is usually done by examining the calculated probability distributions $P(\lambda)$ for the different regions and determining the normalization factor in such a way to obtain the best "match" in the overlapping regions ($P(\lambda)\, d\lambda$ is the Boltzmann probability of occurrence of configurations with parameter $\lambda \in [\lambda, \lambda + d\lambda]$).

The optimal choice of $E_W(\lambda)$ is clearly

$$E_W(\lambda) = kT \ln P(\lambda) \qquad (= -W(\lambda)). \tag{3}$$

Equation (3) also defines the potential of mean force $W(\lambda)$ along the coordinate λ. Unfortunately, $P(\lambda)$ is obtained only as a result of the calculation with the proper $E_W(\lambda)$, thereby creating a vicious cycle.

We propose to break this cycle by an iterative approach where $P(\lambda)$ is first estimated on the small set of λ that would be sampled using E and this estimate is used in the next step to enlarge the set of λ sampled. This refinement is performed until the adequate $E_W(\lambda)$ is found. There are several problems to be solved before

such an algorithm would work successfully on complex systems such as molecular liquids:

a. The algorithm has to provide an automatic matching procedure that works even when the individual estimates of $P(\lambda)$ are not too precise (otherwise the iterations would have to be too long for the method to be practical).

b. The algorithm has to able to recognize iterations that should be dropped as equilibration (otherwise too many iterations would be "wasted" in correcting the error).

c. The algorithm should monitor the degree of the coverage of the λ-space required and be able to "steer" the sampling towards the undersampled or unsampled region. This is particularly important since the estimates of $P(\lambda)$ are likely to be the most imprecise near the boundary of the λ-region sampled and easily can have the effect of preventing the extension of the region sampled.

d. Provisions should be made to recognize likely errors in the calculated $P(\lambda)$ and the temporary replacement of these parts.

While the present work was in progress, Paine and Scheraga published a calculation determining the probability distribution of the dihedral angles in a free dipeptide molecule [11] that was based on the iterative use of Eq. (3), as proposed here. Due to the small dimensionality of the problem, none of the difficulties described above have been encountered and the straight application of the iterative scheme worked successfully.

3. Theory

We describe here an algorithm for the self-consistent determination of $E_W(\lambda)$ over a domain D. The description will be in terms of a multidimensional λ. However, the numerical example given is for the case of a 1-dimensional parameter λ and some steps in the algorithm will contain specific references to the 1-dimensional case.

3.1. *Definitions*

Assume that $P(\lambda)$ is computed on a finite grid $\{\lambda_k\}$. Let S_n denote the set of λ_k that was already sampled after iteration n and P_n^k the estimate of $P(\lambda_k)$, based on the first n iterations, undetermined up to a normalization factor. For iteration n, let s_n be the set of λ_k sampled during the iteration, f_n^k be the number of configurations on which the probability estimate p_n^k, obtained from this iteration only (again, undetermined up to a normalization factor), is based. Furthermore, for any iteration n, let

$$F_n^k = \sum_{j=1}^{n} f_j^k, \tag{4}$$

$$r_n^k = f_n^k / F_n^k \tag{5}$$

and

$$w_n^k = f_n^k \Big/ \sum_{\{h \mid \lambda_h \in s_n\}} f_n^h. \tag{6}$$

r_n^k gives the relative contribution of iteration n to the estimate P_n^k and w_n^k gives the relative contribution of the gridpoint λ_k to p_n^k in iteration n. An indicator to the degree of sampling of the kth gridpoint in n iteration SI_n^k can be defined as

$$SI_n^k = F_n^k / [\max_k F_n^k]. \tag{7}$$

Finally, the function $\lambda^c(\lambda, S)$ is defined as

$$\lambda^c(\lambda, S) \in S, \qquad |\lambda - \lambda^c(\lambda, S)| = \min_{\lambda' \in S} |\lambda - \lambda'| \tag{8}$$

that is, $\lambda^c(\lambda, S)$ is the point in the set S closest to λ.

3.2. *The Outline of the Algorithm*

The algorithm consists of the following steps:

1. Set the iteration counter n to 0, set $S_0 = \{0\}$, $P_0(\lambda) = 1$, and $E_W(\lambda) = 0.0$.
2. Increment the iteration counter n.
3. Run the simulation with the latest E_W for a "reasonable" length.
4. Compute p_n^k for each grid (using Eq. (2)).
5. Decide, if the iteration is to be considered an equilibration step. If so, repeat from step 3.
6. For $n > 1$, match $\mathbf{p}_n$ and $\{\mathbf{p}_i \mid i = 1, \ldots, n-1\}$ to obtain the best estimate $\mathbf{P}_n$. The problem is to find the appropriate normalization factors N_i for each $\mathbf{p}_i$ so that they can be correctly combined to form $\mathbf{P}_n$. The matched probability distribution P_n^k is obtained as

$$P_n^k = \sum_{i=1}^{n} r_i^k N_i p_i^k. \tag{9}$$

The r_i^k factor in Eq. (9) gives higher weight to better sampled gridpoints.

7. If the sampling of the parameter set D appears to be adequate, stop.
8. For each grid $\lambda_k \in S_n$, set

$$E_W(\lambda_k) = kT \ln P_n^k. \tag{10}$$

9. Assign values for $\lambda \notin S_n$. This step involves: (a) extrapolation of $E_W(\lambda_k)$; (b) possible temporary modification of $E_W(\lambda_k)$ to promote the sampling of under-sampled regions; and (c) modification of $E_W(\lambda_k)$ to contain the sampling within the set D.

10. Modify temporarily $E_W(\lambda_k)$ to rectify suspected errors in P_n^k.
11. Repeat from step 2.

3.3. *Detailed Description of the Individual Steps*

Step 5. *Equilibration check.* In order to recognize iterations that are to be considered equilibration and discarded as such, one has to compare the set s_n with the previously sampled regions. In the present version we discarded an iteration n if the set s_n contained grids that were "significantly outside" the overall region sampled before, (including the discarded iterations) and continue from step 3. The reason for this is that if the simulation reached a new region then the iteration should be definitely considered an equilibration period. By "significantly outside" we meant that the set s_i contained points that differ from any previously sampled points by more than 0.06.

As an enhancement, depending on the length of an iteration, or on the distance of the newly sampled ponts from S_{i-1}, it might be useful to eliminate the subsequent N_D iterations. It is also likely, that if the set s_n differs too much from s_{n-1} (even though points in s_i were already sampled at an earlier phase of the calculation) then the calculation was in an unequilibrated phase and should be similarly discarded.

Step 6. *Matching.* The first question here is the establishment of a matching criterion that can be applied without the need of human intervention. There are two obvious choices:

(1) minimize the appropriately weighted sum of deviation squares

$$\sum_{i=1}^{n} \sum_{\{k \mid \lambda_k \in s_n\}} w_i^k (N_i p_i^k - P_n^k)^2 \tag{11}$$

or

(2) minimize the analogous relative deviation square sum

$$DS = \sum_{j=1}^{n} \sum_{\{k \mid \lambda_k \in s_n\}} w_j^k [(N_j p_j^k - P_n^k)/P_n^k]^2. \tag{12}$$

The w_i^k weighting in Eqs. (11), (12) is introduced to reflect the accuracy of the estimate p_i^k. This choice gives equal weight to each iteration and is thus meaningful only if the iteration length is kept constant. A more general choice, allowing for variable iteration length is $w_i^k(L_i/L)$, where L_i and L are the number of configurations in iteration i and the whole run, respectively.

The minimization of the deviation square leads to a system of linear equations (given in the Appendix), whose solution is rather straightforward while the minimization of the relative deviation squares results in a system of nonlinear equations. The solution of the latter can be done by a numerical minimizer and requires also a reasonable initial estimate as described below. Unfortunately, the

probabilities can vary over several orders of magnitude, thus it is necessary to use the relative deviation square to ensure that the quality of the fit will not depend on magnitude of the fitted function.

A common feature of both expressions that $N_i = 0$ is a solution to it, although clearly not the one we are looking for. This reflects the fact that the whole system still will contain an undetermined overall normalization factor. To avoid obtaining this trivial solution, we always fix one of the N_i-s (the first one) to one. Also, if the s_i's form pairwise disjoint classes the minimization is only necessary for the class which contains iteration n.

A tempting proposal here is to retain N_i once it is computed and for each iteration n determine only N_n. Such procedure could be called 1-step optimized matching as opposed to the n-step optimized matching where all N_i's are (re)determined at each iteration.

The procedure used in the present work used the nonlinear n-step optimizing matching, that is, all N_i's were determined in each step by minimizing the DS of Eq. (12). The initial estimate of the N_i's were obtained by using the N_i's of the previous iteration and determining the initial estimate of N_n from a linear one step-optimized matching by minimizing

$$\sum_{\{k \mid \lambda_k \in S_n\}} \left\{ \left[F^k_{n-1} \Big/ \left(\sum_{\{h \mid \lambda_h \in S_n\}} F^h_{n-1} \right) \right] (P^k_{n-1} - P^k_n)^2 + w^k_n (N_n p^k_n - P^k_n)^2 \right\} \tag{13}$$

where P^k_n is given as

$$P^k_n = (1 - r^k_n)\, P^k_{n-1} + r^k_n N_n p^k_n, \tag{14}$$

a special case of Eq. (9). Differentiation with respect to N_n yields the equation

$$N_n = \left[\sum_{\{k \mid \lambda_k \in S_n\}} P^k_{n-1} p^k_n c_k \right] \Big/ \left[\sum_{\{k \mid \lambda_k \in S_n\}} (p^k_n)^2 c_k \right] \tag{15}$$

where

$$c_k = \left[F^k_{n-1} \Big/ \left(\sum_{\{h \mid \lambda_h \in S_i\}} F^h_{n-1} \right) \right] (r^h_n)^2 + w^h_n (1 - r^h_n)^2. \tag{16}$$

Use of the non-linear 1-step optimized matching to minimize the corresponding relative deviation square sum should be even better but the simplicity of Eqs. (15), (16) would be lost.

The derivatives of DS, required for the numerical minimizers, are given as

$$\begin{aligned} \frac{\partial}{\partial N_i} DS = 2 \sum_{j=1}^{n} \sum_{\{k \mid \lambda_k \in s_n\}} & w^k_j \{ \delta^i_j p^k_j (N_j p^k_j - P^k_n)/(P^k_n)^2 \\ & - r^k_i p^k_i [(N_j p^k_j - P^k_n)^2/(P^k_n)^3 + (N_j p^k_j - P^k_n)/(P^k_n)^2] \} \end{aligned} \tag{17}$$

where δ^i_j is the Kroenecker delta.

For calculations requiring a large number of iterations, it might be necessary to group the iterations into smaller sets and use fully optimized matching first within the sets and then optimize the matching of the grouped iteration estimates.

Step 9. *Assignment of* $E_W(\lambda_k)$ *outside* S_n. The simplest assignment for $\lambda_k \notin S_n$ is

$$E_W(\lambda_k) \Rightarrow E_W(\lambda^c(\lambda_k, S_n)) \tag{18}$$

that is outside S_n set $E_W(\lambda_k)$ to the value that it has at the point in S_n nearest to λ_k.

Use of Eq. (18) will lead to possibly large discontinuities in $E_W(\lambda_k)$ whenever the s_i's can be partitioned into pairwise disjoint classes. This can be avoided if one applies a "group normalization factor" to each disjoint class that brings P_n to the (nearly) same scale for all classes. For the 1-dimensional λ, used in the computational example given in this paper, this can be simply achieved. Assume that the parameter λ has been sampled in the intervals $[\lambda_1^0, \lambda_1^1]$, $[\lambda_2^0, \lambda_2^1]$,..., such that $\lambda_i^1 < \lambda_{i+1}^0$. Starting with $i=1$, for each "gap" i, do the following: (a) set $E_W(\lambda_k) = E_W(\lambda_i^1)$ for $\lambda_i^1 < \lambda_k < \lambda_{i+1}^0$; (b) set $E_W(\lambda_k) = E_W(\lambda_k) + E_W(\lambda_i^1) - E_W(\lambda_{i+1}^0)$ for $\lambda_k \geqslant \lambda_{i+1}^0$. For multidimensional λ, however, this task leads to another minimization problem since, in general, there is no unequivocal way of defining the analog of the endpoints of the classes.

To encourage the extension of the sampling one should again compare the region sampled in iteration n with the previous history of the calculation. If the set s_n shrunk from s_{n-1}, that is, $s_{n-1} \notin s_n$ and the abandoned region was "undersampled," then set

$$E_W(\lambda_k) \Rightarrow E_W(\lambda^c(\lambda_k, s_{n-1})) - C|\lambda_k - \lambda^c(\lambda_k, s_{n-1})|. \tag{19}$$

for $\{\lambda_k \mid \lambda_k \in s_{n-1} \wedge \lambda_k \notin s_n\}$.

By an "undersampled" gridpoint we mean that the sampling indicator SI_n^k is less than 0.5. Use of Eq. (19) will encourage sampling of the region suddenly abandoned in iteration n. It is also possible to use a set $s'_{n-1} > s_{n-1}$ in the test to decide if Eq. (19) is needed.

Finally, for $\lambda_k \notin D$, set

$$E_W(\lambda_k) \Rightarrow E_W(\lambda_k, D) + K|\lambda_k - \lambda^c(\lambda_k, D)| \tag{20}$$

to discourage sampling outside the targeted domain D.

Step 10. *Temporary modification of* $E_W(\lambda_k)$. If there is some information about the range of $W(\lambda)$ then one can set a limit $DW_{\max}$ to the difference between the values of the weighting function at neighbouring gridpoints. For the 1-dimensional case this translates into replacing $E_W(\lambda_k)$ with $E_W(\lambda_k) + d$ for $k > j$ whenever

$$DW = |E_W(\lambda_j) - E_W(\lambda_{j-1})| > DW_{\max} \tag{21}$$

where d is given as

$$d = \text{sign}(DW) * (DW_{\max} - DW). \tag{22}$$

It is reasonable to restrict this replacement to gridpoints that were not sampled too well, that is, where the sampling indicator SI_n^k is smaller than a threshold value, chosen to be 0.2 in the present work.

4. CALCULATIONS

The algorithms described in Section 3 was tested on the calculation of the solvent contribution to the free energy difference between two conformations of the Alanine dipeptide in aqueous solution at 25°C. This system has been studied recently by the original umbrella sampling method and the free energy difference was determined by computing the potential of mean force along a straight line connecting the two conformations in the torsion angle space [8]. The solution was modeled by one dipeptide molecule with flexible torsion angles and 201 rigid water molecules under face-centered cubic periodic boundary conditions. The water–water interactions were modeled by the MCY-CI potential [12] and the dipeptide–water interactions were described by the potential library of Clementi and coworkers [13]. The combination of force bias [14] and preferential sampling [15] techniques has been employed on the molecular displacement whose joint application was shown to enhance significantly the convergence of solute–solvent properties [16].

The conformation of the dipeptide is traditionally described by the torsion angles ψ and ϕ, defined by the backbone atoms NCCN and CNCC, respectively. An unsuccessful attempt has been made earlier to generate an $E(\lambda) = E(\psi, \phi)$ that samples the entire torsion angle space using $E(\lambda)$ in different functional forms [17]. However, since several conformations have been identified as possible minima of the potential of mean force surface, calculations described in [8] sampled conformations connecting two pairs of these possible minima. One calculation sampled the line connecting $C_7(90°, -90°)$ and $\alpha_R(-50°, -70°)$ and another sampled the line connecting C_7 and $P_{II}(150°, -80°)$. These calculations gave the solvent contribution to the free energy difference between the two conformations, $\Delta A_{ij}^{\text{hyd}}$, as

$$\Delta A_{ij}^{\text{hyd}} = kT \ln(P_i/P_j) \tag{23}$$

where P_i and P_j are the Boltzmann probabilities of occurrence of conformations i and j, respectively.

In these calculations E_W was the function of a single parameter λ that defines the conformation along the line connecting i and j,

$$(\psi, \phi) = (1 - \lambda)(\psi_i, \phi_i) + \lambda(\psi_j, \phi_j) \tag{24}$$

and E_W was used in an analytical form

$$E_W(\lambda) = c(\lambda - \lambda_0)^2. \tag{25}$$

The calculation of $\Delta A_{ij}^{\text{hyd}}$ required the sampling of a total of ~ 3000 K and

$\sim$2000 K configurations for the two lines studied, respectively. This includes 5 and 3 separate runs of 500–700 K length in addition to several shorter ($\sim$50 K) trials to determine the appropriate c and λ_0 values for each run.

The algorithm was tested on the sampling of the parameter λ of Eq. (24) in the domain $D = [0, 1]$ for the $C_7 \rightarrow \alpha_R$ transition. This appears to be a good test case since $P(\lambda)$ has a minimum at $\lambda \cong 0.25$. The calculation was started from a configuration with $\lambda = 0.18$. One iteration consisted of 20K configurations. The C and K parameters in Eqs. (19) and (20) were both chosen to be 0.5 kT. The [0, 1] interval has been divided into a uniform grid of 50 intervals and E_W has been computed by linear interpolation between the grid centers.

5. Results

At the outset of the study, both the linear 1-step optimized matching and the linear n-step optimized matchings were tried. The 1-step optimized matching, after having crossed the barrier remained near the $\lambda = 1.0$ range, which corresponds to the minimum of $W(\lambda)$. The linear fully optimized matching provided reasonable coverage for the [0, 0.8] range but was unable to reach $\lambda = 1.0$. Furthermore, the matched curves varied qualitatively even after 30 iterations (of 10 K length). This demonstrates the effect of the large variations in the order of magnitude of $P(\lambda)$ over the [0, 1] interval.

The calculation using the non-linear n-step optimized matching crossed the barrier in 6 iterations and reached $\lambda = 1.0$ in 12 iterations. However, the estimate near $\lambda = 0.45$ and $\lambda = 1.0$ was very bad at first: The $W(\lambda)$ values showed a sudden increase as λ was approaching 1.0 and a sudden drop as λ was approaching 0.45 from larger λ values. It took 25 iterations to correct the first error. Since the error was in the direction of overestimating $W(\lambda)$, it had the effect of "trapping" the system until the $W(\lambda)$ estimate was corrected. The subsequent 25 iterations were spent in the range [0.5, 1.0]. The correction of the second error turned out to be more difficult since it had the effect of "repelling" the system from the problematic region. However, use of Eq. (19) in step 9 finally forced the system to sample smaller λ values and the second error was also corrected. This allowed the system to recross the barrier and sample again the region around $\lambda = 0.0$. The calculation was stopped after 118 iterations, requiring 2480 K Monte Carlo steps.

The resulting $W(\lambda)$ is shown in Fig. 1 together with the $W(\lambda)$ estimates from the previous work using the harmonic umbrella sampling. The original work performed the matching by using points that lie in an interval where the two curves are the "most parallel" and no quantitative account of the precision of $P(\lambda)$ at the matched point was used. The $W(\lambda)$ curve obtained in the present work parallels the matched harmonic umbrella sampling result rather well, with the exception of run 2. This run was around the maximum of $W(\lambda)$, and was experiencing some ergodic difficulties, as remarked in [8]: The first part of the calculation was spent mainly on the C_7 side of the $W(\lambda)$ curve and the second part around the α_R side. In view of

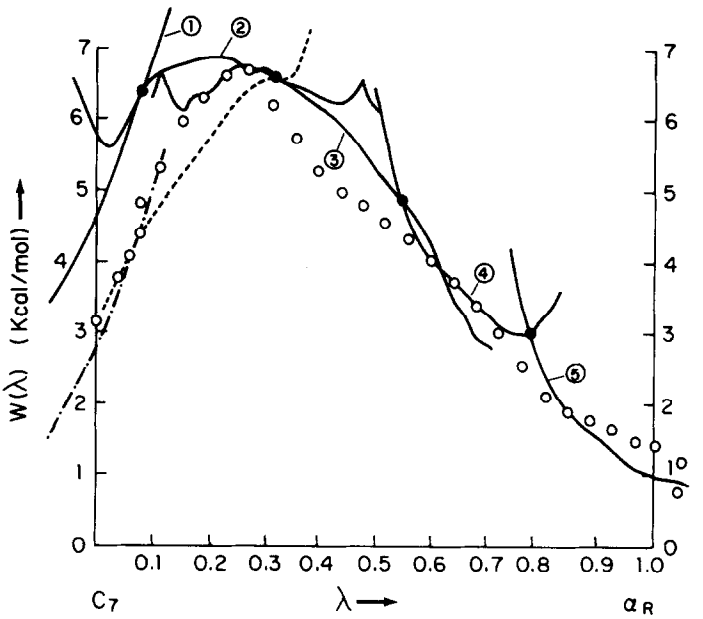

FIG. 1. Potential of mean force along the correlated conformational coordinate λ. Full lines: results of the harmonic umbrella sampling of [8]. Filled circles show point used for matching; open circles, the results of the adaptive procedure using the non-linear n-step optimized matching; – – –, the result of the new harmonic umbrella sampling calculation replacing run 2; – · – · –, the result of run 1 shifted to match the new run 2.

the present, rather large discrepancy, we performed a new harmonic umbrella sampling calculation. After several unsuccessful tries we found that $E_W(\lambda) = 150^*(\lambda - 0.25)^2$ sampled consistently both sides of the peak of $W(\lambda)$. We performed a 700 K long run. The result of this run is also shown on Fig. 1.

The free-energy difference was obtained in the present work using the adaptive algorithm as 1.4 Kcal/mole. The previous work gave 3.6 Kcal/mole. After replacing the second run with the one calculated here, we obtain 1.8 Kcal/mole. The error of $\Delta A_{ij}^{\text{hyd}}$ was estimated previously to be 0.3–0.6 Kcal/mole, based on the stability of the $P(\lambda)/P(\lambda')$ ratios examined in the longest individual run.

The fully optimized matching procedures require computer time that is roughly proportional to the third power of the number of iterations. Optimizing 50 iterations added about 15% to the Monte Carlo computation effort.

6. DISCUSSION

In comparison to the standard umbrella sampling procedure the algorithm presented here has the following advantages:

1. The sampling region can be set a priori and the possibility of having to do "one more run" to fill a gap is absent. The need for human intervention is reduced.

2. There is a built-in self check on the matched probability distribution: If the sampling stays too close to a region or avoids sampling a region, then the $P(\lambda)$ is incorrect there (and vice versa). This is a particularly important point as it was demonstrated that the harmonic umbrella sampling result may be seriously in error on runs of medium length without giving any obvious sign of lack of convergence.

3. The matching of the individual runs are mostly done based on the middle of the sampled interval. The standard procedure has to use the most unreliable regions for matching, thereby reducing its precision.

4. The current work required about the same computational effort as the conventional procedure. However, it appears that with relatively simple modification significant computational gains can be obtained as discussed below.

In summary, the present work demonstrated the feasibility and the improved reliability of the adaptive procedure and pinpointed the issues critical to its success: (1) Use of a matching criterion that gives uniform a priori weight to all λ values (that is, use of the non-linear matching instead of the linear); (2) flexibility in assigning the normalization constants N_i (that is, use n-step optimized matching instead of the computationally more appealing 1-step optimized matching); and (3) provide filters to eliminate iterations that constitute an equilibration phase. Future work, however, should refine and improve it on several points: (1) Step 5 should be enhanced to allow for better recognition of iterations that are to discarded as equilibration. In particular, additional filters in step 10 could recognize some of the obviously wrong estimates in a later stage of the calculation and remove them from the averaging. (2) The algorithm is likely to be sensitive to the choice of C value in Eq. (19). Too small a value will not have the desired effect and too large a value will result in incorrect estimates since the weighting function was changed so drastically that an equilibration step would be needed first. (It appears that our choice was a little on the high side for this system.) (3) More work is needed to determine the optimal choice of iteration length, the best way to group iterations into "super iterations" (as suggested at the end of step 6), and the optimal size of the interval D to be sampled in a single calculation.

APPENDIX: DETERMINATION OF THE NORMALIZATION CONSTANTS WITH THE LINEAR n-STEP OPTIMIZED MATCHING

Differentiation with respect to each N_i yields a system of linear equations with coefficients

$$A_{ij} = \sum_{\{k|\lambda_k \in s_i \wedge s_j\}} (G^k r_j^k r_i^k - w_j^k r_i^k - w_i^k r_j^k)\, p_i^k p_j^k + \delta_j^i w_j^k (P_j^k)^2 \qquad \text{(A1)}$$

where

$$G^k = \sum_{i=1}^{n} w_i^k. \tag{A2}$$

The system is homogeneous, reflecting the fact that there is an arbitrary normalization factor for the composite probability distribution. Fixing one of the N_i's to 1.0 yields an inhomogeneous system that will provide definitive values for the N_i's. Notice, however, that the solution will depend on the i chosen since fixing one of the N_i's slightly alters the minimization problem. (It would be independent of the choice of i only if the determinant of **A** were zero.)

ACKNOWLEDGMENTS

This research was supported by NIH Grant GM 24914 and by NSF Grant CHE-8203501. Professor D. L. Beveridge is thanked for many helpful discussions.

REFERENCES

1. G. N. PATEY AND J. P. VALLEAU, *J. Chem. Phys.* **63**, 2334 (1975).
2. G. TORRIE AND J. P. VALLEAU, *J. Comp. Phys.* **23**, 187 (1977).
3. C. S. PANGALI, M. RAO, AND B. J. BERNE, *J. Chem. Phys.*, **71**, 2975 (1979).
4. H. L. SCOTT AND C. Y. LEE, *J. Chem. Phys.* **73**, 4591 (1980).
5. S. OKAZAKI, K. NAKANISHI, H. TOUHARA, AND Y. ADACHI, *J. Chem. Phys.* **71**, 2421 (1979).
6. M. MEZEI, *Mol. Phys.* **47**, 1307 (1982).
7. G. RAVISHANKAR, M. MEZEI, AND D. L. BEVERIDGE, *Faraday Symp. Chem. Soc.* **17**, 79 (1982).
8. M. MEZEI, P. K. MEHROTRA, AND D. L. BEVERIDGE, *J. Amer. Chem. Soc.* **107** (1985).
9. M. BERKOWITZ, J. A. KARIN, A. MCCAMMON, AND P. J. ROSSKY, *Chem. Phys. Lett.* **105**, 577 (1984).
10. J. CHANDRASEKHAR, S. F. SMITH, AND W. JORGENSEN, *J. Amer. Chem. Soc.* **106**, 3049 (1984); J. CHANDRASEKHAR AND W. L. JORGENSEN, *J. Amer. Chem. Soc.*, in press.
11. G. H. PAINE AND H. A. SCHERAGA, *Biopolymers* **24**, 1391 (1985).
12. O. MATSUOKA, E. CLEMENTI, AND M. YOSHEMINE, *J. Chem. Phys.* **64**, 1351 (1976).
13. E. CLEMENTI, F. CAVALLONE, AND R. SCORDAMAGLIA, *J. Amer. Chem. Soc.* **99**, 5531 (1977).
14. M. RAO, C. S. PANGALI, AND B. BERNE, *Mol. Phys.* **37**, 1773 (1979).
15. J. C. OWICKI AND H. A. SCHERAGA, *Chem. Phys. Latt.* **47**, 600 (1979); J. C. OWICKI, in *Computer Modeling of Matter*, P. G. Lykos, Ed. (Amer. Chem. Soc., Washington, D.C., 1978).
16. P. K. MEHROTRA, M. MEZEI, AND D. L. BEVERIDGE, *J. Chem. Phys.* **78**, 3156 (1983).
17. P. K. MEHROTRA, unpublished results.