Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method

SHANKAR KUMAR and JOHN M. ROSENBERG*

Department of Biology and the W. M. Keck Center for Advanced Training in Computational Biology at the University of Pittsburgh, Carnegie Mellon University, and the Pittsburgh Supercomputing Center, Pittsburgh, Pennsylvania 15260

DJAMAL BOUZIDA and ROBERT H. SWENDSEN

Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

PETER A. KOLLMAN

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received 23 February 1994; accepted 23 August 1994

ABSTRACT

The recently formulated weighted histogram analysis method (WHAM)[1] is an extension of Ferrenberg and Swendsen's multiple histogram technique for free-energy and potential of mean force calculations. As an illustration of the method, we have calculated the two-dimensional potential of mean force surface of the dihedrals gamma and chi in deoxyadenosine with Monte Carlo simulations using the all-atom and united-atom representation of the AMBER force fields. This also demonstrates one of the major advantages of WHAM over umbrella sampling techniques. The method also provides an analysis of the statistical accuracy of the potential of mean force as well as a guide to the most efficient use of additional simulations to minimize errors.

*Author to whom all correspondence should be addressed at the Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260.

Journal of Computational Chemistry, Vol. 16, No. 11, 1339–1350 (1995)

CCC 0192-8651 / 95 / 111339-12

Introduction

The generation of multidimensional free-energy and potential of mean force (PMF) surfaces from computer simulations is one of the most challenging problems in computational chemistry and biology. Because of the difficult nature of the problem, multidimensional PMF/free-energy surfaces have rarely been computed. This article deals with the application of the recently formulated weighted histogram analysis method (or WHAM) for the generation of multidimensional PMF profiles.[1] The problem of calculating free-energy differences in ligand–protein or ligand–DNA interactions with WHAM will be dealt with in a future communication.

The method generally used so far for PMF generation is umbrella sampling.[2-6] In the umbrella sampling method, the Hamiltonian $\hat{H}(x)$ is replaced by a modified potential, $\hat{H}_{(\lambda)}(x)$, of the form

$$\hat{H}_{(\lambda)}(x) = \hat{H}_0(x) + \sum_{i=1}^{L} \lambda_i \hat{V}_i(x) \qquad (1)$$

The functions $\hat{V}_i(x)$ $(i = 1, 2, \ldots, L)$ are restraining (or biasing) potentials that are chosen to shift the sampling distribution along a (multidimensional) generalized coordinate of interest, $\tilde{\xi}$. The tilde over the reaction coordinate ξ denotes the (possible) multidimensionality of the reaction coordinate.

The unperturbed Hamiltonian $\hat{H}_0(x)$ can be included in this set of functions if it is required to vary the temperature (see the following section). When ambiguities might arise, circumflexes are used over the symbols to denote functions; thus, for example, $\hat{V}_i(x)$ denotes the function and V_i a particular value the function takes. The $\lambda_i (i = 1, 2, \ldots, L)$ are the coupling parameters. The symbol in braces, $\{\lambda\}$, denotes the set of values $\lambda_1, \lambda_2, \ldots, \lambda_L$.

The difficulties inherent in the umbrella sampling method for free-energy or PMF calculations are outlined in Ref. 1, and the difficulties become more pronounced if multidimensional free-energy or PMF maps are to be calculated. This is why multidimensional free energies have rarely been calculated. Two-dimensional free energies have been calculated by connected regions that are low in energy by a one-dimensional path[7,8]; however, such methods are not really multidimensional and the technique of using one-dimensional connecting paths could lead to erroneous results. The estimation of multidimensional free-energy values is one of the critical areas in which WHAM outperforms methods previously used.

As would be expected from counting statistics, the statistical errors are guaranteed to go down with WHAM with an increased number of simulations,[1] although this is not necessarily the case with umbrella sampling or free-energy perturbation methods. The adequacy of statistical sampling will be discussed briefly in the "Results and Discussion" section and will be detailed in a subsequent article.

The problem is twofold: (1) Simulations will have to be carried out for long periods of time to obtain a good sampling of $\tilde{\xi}$; and (2) once the simulations of the required length(s) have been carried out, proper merging of the data from the various simulations to obtain unique values of free energy has been difficult. In this article we use the WHAM[1] equations to deal with the latter problem; and by using the dynamically optimized Monte Carlo method,[9,10] we also deal with the former to some extent to generate two-dimensional PMF profiles of the γ–χ dihedral angles in deoxyadenosine.

The γ dihedral is that formed by the quartet of atoms $O_{5'}—C_{5'}—C_{4'}—C_{3'}$ and χ by $O_{4'}—C_{1'}—N_9—C_4$ (Fig. 1). Pearlman and Kollman[11] have calculated the γ–χ PMF profiles for deoxyadenosine using the all atom (AA) and united atom-hydroxyl (united atom OH) representations using the method of Tobias and Brooks.[12] They have also calculated the PMF profiles for adenosine using these two models and for the united

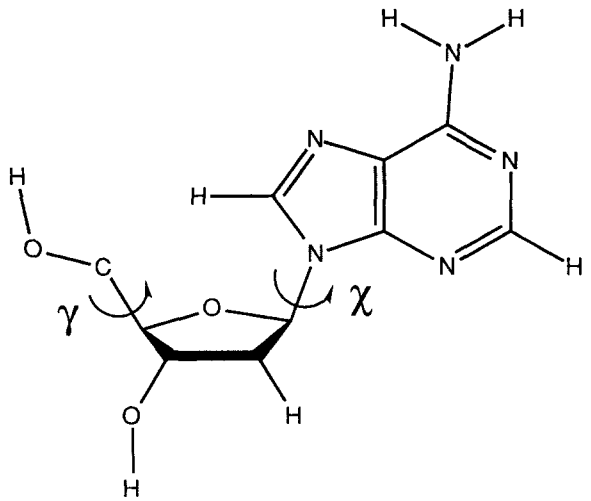

FIGURE 1. The γ and χ dihedrals of adenosine.

atom (UA) model. In this study we have used the all atom and united atom models of deoxyadenosine so that qualitative and, when possible, quantitative comparisons can be made.

The Weighted Histogram Analysis Method

The derivation of the WHAM equations is detailed in ref. 1, and the results are given here. Consider R simulations, with the ith simulation being carried out at temperature $T_i = 1/k_B\beta_i$ and with the coupling parameters in eq. (1) set to $\{\lambda\}_i$.† Let n_i be the number of configurations archived from the ith simulation. Then the (unnormalized) probability histogram $P_{\{\lambda\}',\beta'}(\{V\}, \tilde{\xi})$ is given by[1]

$$P_{\{\lambda\}',\beta'}(\{V\}, \tilde{\xi}) = \frac{\sum_{k=1}^{R} N_k(\{V\}, \tilde{\xi})\exp\left[-\beta'\sum_{j=1}^{L}\lambda'_j V_j\right]}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta_m\sum_{j=1}^{L}\lambda_{jm}V_j\right]} \quad (2)$$

and

$$\exp(-f_j) = \sum_{\{V\},\tilde{\xi}} P_{\{\lambda\}_j,\beta_j}(\{V\}, \tilde{\xi}) \quad (3)$$

where $N_k(\{V\}, \tilde{\xi})$ is the value taken by the histogram (frequency of observation) at $(\{V\}, \tilde{\xi})$ during the kth simulation and f_m is the (dimensionless) free energy of the system described by the Hamiltonian of eq. (1) with coupling parameters $\{\lambda\}_m$; $f_m = \beta_m A_m$, where A_m is the Helmholtz free energy of the system during the mth simulation. $P_{\{\lambda\}',\beta'}(\{V\}, \tilde{\xi})$ is the corresponding probability histogram that would result if a simulation were to be carried out at temperature $T' = 1/k_B\beta'$ and with coupling constants $\lambda'_1, \lambda'_2, \ldots, \lambda'_L$ (denoted by $\{\lambda\}'$). The initial step in WHAM is the determination of the free energies f_m. This is done by setting $\{\lambda\}' = \{\lambda\}_m$, $\beta' = \beta_m$ and assigning arbitrary values (usually zeroes) to the f_m in eq. (2). Equations (2) and (3) are then iterated until convergence in the free energies, f_q, is obtained.[1] Once the free energies are determined, the probability distribution $P_{\{\lambda\}',\beta'}(\{V\}, \tilde{\xi})$ is calculated using eq. (2) by setting the coupling parameters to $\{\lambda\}'$ and the temperature to $T' = 1/k_B\beta'$.

Equation (2) represents a general form of the WHAM equation that can be used to extract data from simulations carried out at various temperatures without any restriction on the form of the restraining potentials. Using eq. (2) in its most general form entails the generation of $(L + D_\xi)$ dimensional histograms, where D_ξ is the dimensionality of $\tilde{\xi}$. Note that in the most general form of the WHAM equations, we have implicitly included the unperturbed Hamiltonian among the restraining potentials. Various strategies can be used to reduce computational cost. If PMFs of $\tilde{\xi}$ are desired at a single temperature T, then carrying out all the simulations at the same temperature T obviates the necessity of histogramming along the unperturbed energy $\hat{H}_0(x)$, thus reducing the dimensionality of the histograms by unity.[13] This result was rediscovered recently by integrating the unperturbed Hamiltonian out of the WHAM equations and was the strategy employed in the generation of PMF profiles of the O_1—H_5 hydrogen bonding distance in the blocked tripeptide $(ALA)_3$—CBX.[14]

The dimension of the histograms can be reduced further by a judicious selection of the biasing potentials. If all the biasing potentials are explicit functions of the reaction coordinate $\tilde{\xi}$ only, then the dimensionality of the histograms can be reduced by $L - 1$.[1] Thus, by carrying out simulations at the temperature $T = T'$ and by using biasing potentials that are explicit functions of $\tilde{\xi}$, the dimensionality of the histograms can be reduced to D_ξ.[1] Not all problems lend themselves easily to such dimension reduction strategies. For example, it was easier to incorporate a potential that was not an explicit function of the reaction coordinate in the generation of the PMF profile of the pseudorotation phase angle[1] of the sugar ring in deoxyadenosine than to include an explicit function because this would have slowed the simulation itself.

If the biasing potentials are chosen to be explicit functions of $\tilde{\xi}$ only and if $T_i = T'$ (for $i = 1, 2, \ldots, R$), then eqs. (2) and (3) can be simplified further to the following:

$$P_{\{\lambda\}',\beta'}(\tilde{\xi}) = \frac{\sum_{k=1}^{R} N_k(\tilde{\xi})\exp\left[-\beta'\sum_{j=1}^{L}\lambda'_j\hat{V}_j(\tilde{\xi})\right]}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta'\sum_{j=1}^{L}\lambda_{jm}\hat{V}_j(\tilde{\xi})\right]} \quad (4)$$

and

$$\exp(-f_j) = \sum_{\tilde{\xi}} P_{\{\lambda\}_j,\beta'}(\tilde{\xi}) \quad (5)$$

†$\{\lambda\}_k$ refers to the value of the coupling parameters during the kth simulation; $\{\lambda\}_k$ denotes the set $\{\lambda_1, \lambda_2, \ldots, \lambda_L\}_k$, which is identical to $\{\lambda_{1,k}, \lambda_{2,k}, \ldots, \lambda_{L,k}\}$.

Note that the unperturbed Hamiltonian, $\hat{H}_0(x)$, is now *not* included among the restraining functions.

The Generation of the 2D γ–χ PMF Surfaces

We will now demonstrate the use of the WHAM equations to generate the two-dimensional γ–χ PMF profiles using the all atom model of Kollman and co-workers[15] for deoxyadenosine. By carrying out all the simulations at the temperature of 300 K —the temperature at which PMF profiles are desired—the necessity of histogramming the data along the unperturbed energy $\hat{H}_0(x)$ was eliminated. The restraining potentials were chosen to be explicit functions of the dihedral angles γ and χ, thus reducing the dimensionality of the histograms in eq. (3) to two. Furthermore, PMF profiles of γ–χ were desired for deoxyadenosine without any biasing potentials added to $\hat{H}_0(x)$; thus, $\lambda_i = 0$ $(i = 1, 2, \ldots, L)$.

For the determination of the free energies f_m, the WHAM equations become [see eqs. (4) and (5)]

$$P_{\{\lambda\}_q,\beta'}(\gamma,\chi) = \frac{\sum_{k=1}^{R} N_k(\gamma,\chi)\exp\left[-\beta'\sum_{j=1}^{L}\lambda_{jq}\hat{V}_j(\gamma,\chi)\right]}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta'\sum_{j=1}^{L}\lambda_{jm}\hat{V}_j(\gamma,\chi)\right]} \tag{6}$$

and

$$\exp(-f_j) = \sum_{\gamma,\chi} P_{\{\lambda\}_j,\beta'}(\gamma,\chi) \tag{7}$$

Once the free energies, f_m, were determined by iterating eqs. (6) and (7), the probability distribution $P_{\{\lambda=0\},\beta}(\gamma, \chi)$ was calculated using eq. (6) by setting all the coupling constants to zero. Thus,

$$P_{\{0\},\beta'}(\gamma,\chi) = \frac{\sum_{k=1}^{R} N_k(\gamma,\chi)}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta'\sum_{j=1}^{L}\lambda_{jm}\hat{V}_j(\gamma,\chi)\right]} \tag{8}$$

Monte Carlo (MC) simulations were performed on both the united atom and all atom models of deoxyadenosine. Pearlman and Kollman[11] noted that during their molecular dynamics (MD) simulations of the all atom model, the hydroxyl hydrogen was attracted to a nearby negative charge and got trapped in this conformation for a long period of time, thus resulting in inefficient sampling of the rotamers of the hydroxyl bonds. To overcome this problem, the acceptance ratio method (ARM) and the dynamically optimized Monte Carlo (DOMC)[9,10] were used in this study. These techniques are simple variations of the Metropolis algorithm,[16] which for certain problems can be profitably used to increase sampling efficiency in MC simulations; this is brought about by optimizing the parameters governing the random MC jumps, as described next.

The acceptance ratio method is used to carry out simulations with different dynamically determined step sizes for each particle. The approximately exponential behavior of the acceptance ratio, $\langle P \rangle$, is used to calculate the ideal maximum step size δ_{id}. The simulation is set up as a sequence of cycles; for any given particle, if the maximum jump is δ_{old} and the corresponding acceptance ratio is $\langle P_{old} \rangle$, then the new jump size δ_{new} for the next cycle is calculated by[9,10]

$$\delta_{new} = \delta_{old}\frac{\ln(a\langle P_{id}\rangle + b)}{\ln(a\langle P_{old}\rangle + b)} \tag{9}$$

where $\langle P_{id} \rangle$ is the ideal desired acceptance ratio that yields the highest root mean square (rms) deviations and is obtained by carrying out a series of preliminary simulations. For biomolecular systems, $\langle P_{id} \rangle$ is approximately 0.4 when the MC moves are generated from a sphere. Moreover, a and b are real parameters such that δ_{old} is either multiplied or divided by a convenient scale factor whenever $\langle P \rangle$ is 0 or 1 (to prevent any overflow/underflow errors). The factors a and b were chosen to be 0.37 and 0.296. In the simulations reported here, each cycle comprised 200 sweeps (one sweep being one MC move for every atom in the system).

In the three-dimensional dynamically optimized Monte Carlo method, simulations are done in cycles, as in the acceptance ratio method, but the MC moves are sampled from ellipsoids whose axes are chosen to be short along the "hard" directions and long along the "soft" directions.[9,10] In this study the one-dimensional dynamically optimized Monte Carlo method is used and the estimate of the maximum jump size δ is given by

$$\delta = F\left\{\frac{[(\Delta x)^2]}{\beta[\Delta E]}\right\}^{1/2}$$

Here square brackets indicate averages over all the attempted MC moves (whether or not they were

accepted) in a cycle. $[(\Delta x)^2]$ is the average of the square of the actual MC moves, and $[\Delta E]$ is the average of the energy changes resulting from the MC moves along the direction x. F is a scale factor that is adjusted to yield the minimum energy correlation according to Bouzida et al.[9]; this automatically yields a simulation that is optimized to maximize conformational sampling. A value of F around 3.0 was used for the simulations carried out here.

Bond rotations were introduced around the γ and the χ dihedrals in the all atom model to increase the efficiency of conformational sampling. The single-particle Cartesian space moves were carried out using acceptance ratio method, and the bond rotations were done with the one-dimensional dynamically optimized Monte Carlo method. Thus both the acceptance ratio method and the dynamically optimized Monte Carlo methods were used simultaneously in this study. The initial conformations of the all atom and united atom models of deoxyadenosine were taken from the AMBER[15,17] database. Because simulations with just the AMBER force field did not yield sufficient sampling of the higher energy regions, restraining potentials had to be used. Restraints of the form

$$V = 1.0 + \cos(\theta - \phi)$$

were added to the AMBER force field. Here θ is the dihedral angle (either γ or χ) and ϕ is a phase difference that controls where the minimum of V occurs.

One-dimensional histograms of the counting statistics of both γ or χ were calculated as the simulations developed. If the minimum count in any bin was not at least of the order of 1000 (approximately equivalent to a 3% error), then more simulations were carried out with restraints added to improve sampling in statistically poor regions. The restraining potentials along with the coupling parameters λ that were used in the simulations are detailed in Table I. Eleven simulations were done with the all atom model and six with the united atom model.

Results and Discussion

Two-dimensional PMF contours of the γ–χ surface with both the united atom and all atom models for deoxyadenosine were easily obtained from the simulational data using WHAM. The WHAM calculations were carried out on an IBM RISC/6000 model 560 workstation, as were the MC simulations. The CPU requirements for WHAM were modest and needed much less than 1% of the total time taken by the MC simulation.[‡] This shows that WHAM is not a computational bottleneck.

The PMF of γ at 300 K (obtained by integrating the γ–χ PMF along χ) exhibits the threefold staggered pattern of ethane (Fig. 2). The minima for γ are seen to occur at around 66° (g^+ or gauche plus), 180° (anti), and 300° (g^- or gauche minus).

[‡]A FORTRAN program of WHAM is available upon request.

TABLE I.
Restraint Potentials Used in the Simulations.

Simulation No.	Restraining Potentials for the All Atom Model (kcal / mol)	Restraining Potentials for the United Atom Model (kcal / mol)
1	0.0	0.0
2	5.0[1.0 + cos(γ − 153°)]	5.0[1.0 + cos(γ − 153°)]
3	2.5[1.0 + cos(γ − 170°)]	5.0[1.0 + cos(γ − 63°)]
4	2.5[1.0 + cos(γ − 170°)]	5.0[1.0 + cos(γ − 297°)]
5	3.5[1.0 + cos(γ − 63°)]	5.0[1.0 + cos(γ − 170°)]
6	3.5[1.0 + cos(γ − 170°)]	7.5[1.0 + cos(γ − 153°)]
7	0.0	
8	0.0	
9	5.0[1.0 + cos(γ − 153°)]	
10	5.0[1.0 + cos(γ − 297°)]	
11	5.0[1.0 + cos(γ − 63°)]	

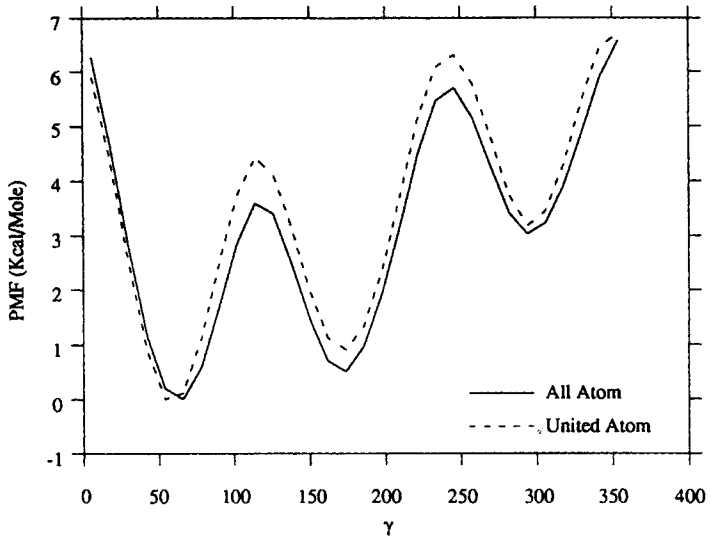

FIGURE 2. PMF profiles of gamma at (γ) with the all atom and united atom models.

The global minimum for γ occurs at g^+, a manifestation of the "gauche effect."[18] In the all atom model, the PMF at the anti minimum (E_a) is 0.5 kcal/mol higher than that of the g^+ minimum (E_{g^+}), and the g^- minimum (E_{g^-}) is higher than E_{g^+} by 3.0 kcal/mol. The results for the united atom model (Fig. 2) are similar except that the anti minimum is higher by about 0.9 kcal/mol than the global g^+ minimum.

The united atom and all atom PMF profiles of χ at 300 K, obtained by integrating the γ–χ PMF profiles along γ, are shown in Figure 3. Both the united atom and all atom models exhibit the familiar bimodal pattern for χ. The minima are seen to occur at the syn (around 40°) and anti (around 200°) conformations, with the global minima at the syn conformation for both the united atom and all atom models. For the all atom model, the difference between the two is around 0.8 kcal/mol, whereas for the united atom model this difference is around 1.2 kcal/mol. The "double peak" centered around 320° that is exhibited by the united

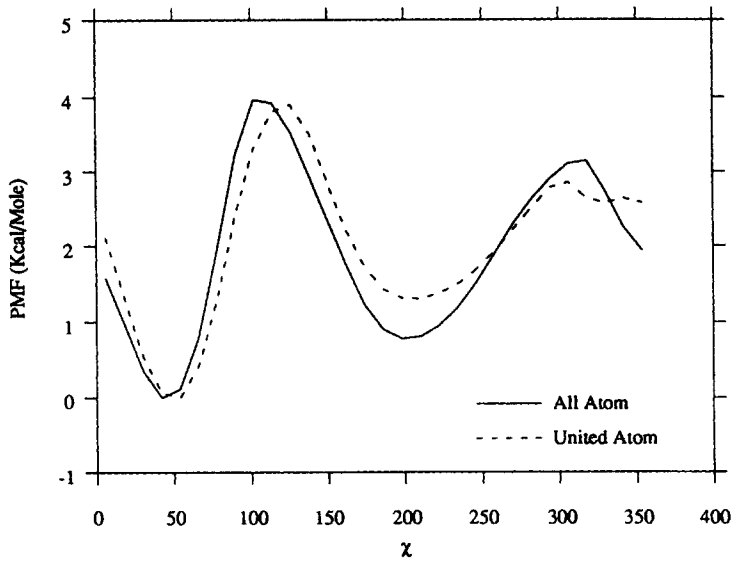

FIGURE 3. PMF profiles of chi (χ) with the all atom and united atom models.

atom map is probably an artifact caused by the united atom model. Nuclear magnetic resonance (NMR) data suggest that both syn and anti conformations of χ are equally probable in purine nucleosides.[18]

The all atom and the united atom γ–χ PMF contours at 300 K are shown in Figures 4 and 5. At the qualitative level, the two contour maps look very similar. The minimum of the PMF on the γ–χ surface is seen to occur around (66°, 42°) or (g^+, syn) for both the models; the probability surfaces (Figs. 6 and 7) exhibit an overall maximum around this point. The γ–χ probability contours are shown in Figures 8 and 9. The other significant minimum of the PMF on the γ–χ plane occurs in the (anti, anti) region around (180°, 200°). It is the base—$O_{5'}$ interactions, specifically the hydrogen bond between the N_3 and the $O_{5'}$ atoms, that stabilize (g^+, syn) global minimum. The g^- and anti minima of γ are stabilized by the favorable entropic conditions that result from the $O_{5'}$ atom swinging away from the base and sugar moieties. In longer strands of DNA, the anti conformation of χ is energetically more favorable than the syn conformation due to steric considerations.

The difference of 3.0 kcal/mol between E_{g^+} and E_{g^-} causes the g^- region to be sparsely populated by γ in simulations of DNA at 300 K. In molecular dynamics simulations of the dodecamer CGCGAATTCGCG at 300 K, the g^- region was not sampled at all, indicating the E_{g^-} is higher by more than 3 kcal/mol than the global g^+ minimum in longer pieces of DNA.[19] These results are consistent with experimental observations, in which the g^- conformation is only rarely seen.[18] In MD simulations of the dodecamer CGCGAATTCGCG, the glycosidic angles χ were all anti, consistent with experimental observations.[18]

The qualitative features of the results compare favorably with those calculated by Pearlman and Kollman[11] using the method of Tobias and Brooks.[12] In addition to the united atom and all atom models, Pearlman and Kollman also used the united atom-hydroxyl model, in which only the hydroxyl groups were represented by united atom oxygens; they carried out simulations with the all

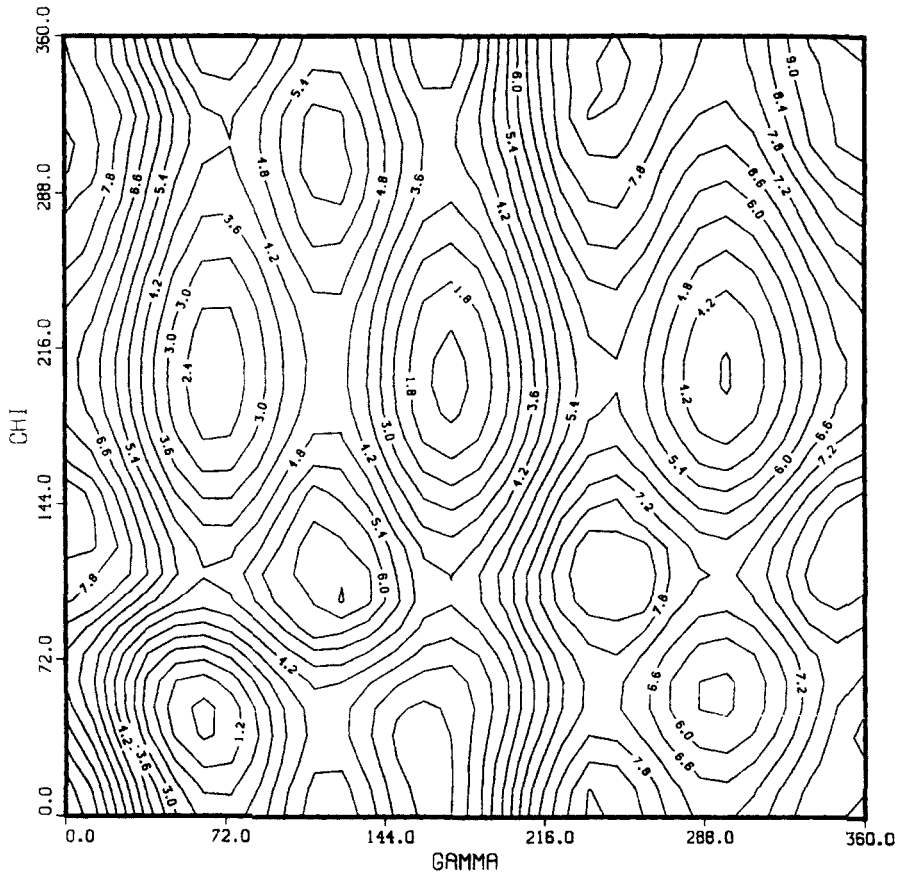

FIGURE 4. Two-dimensional γ–χ PMF contours for the all atom model.

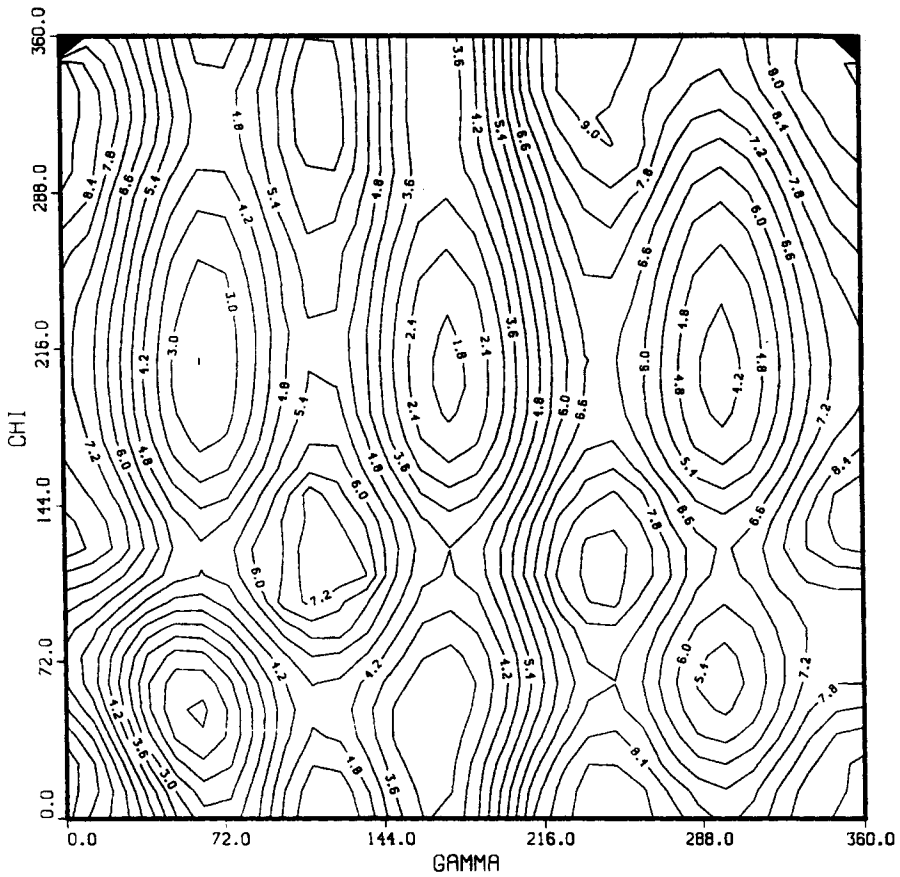

FIGURE 5. Two-dimensional γ–χ PMF contours for the united atom model.

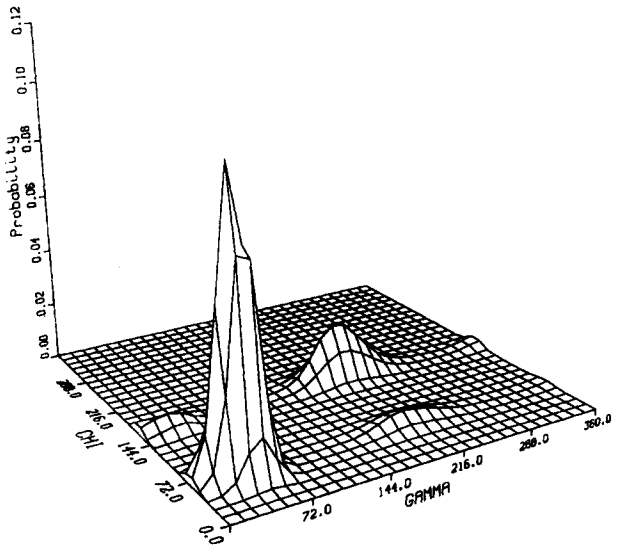

FIGURE 6. γ–χ probability surface for the all atom model.

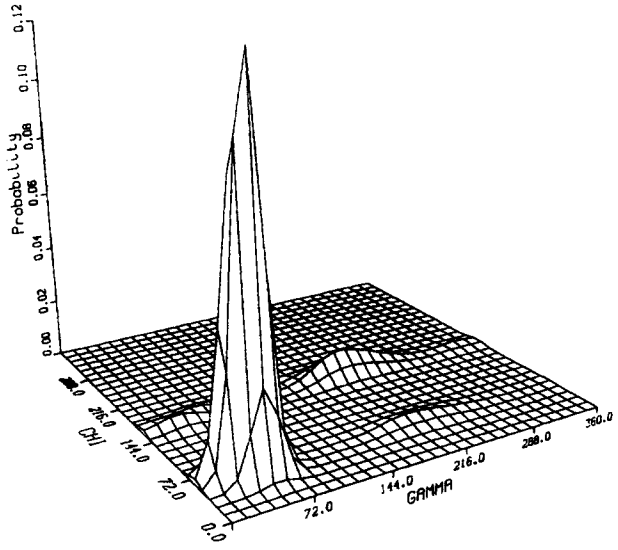

FIGURE 7. γ–χ probability surface for the united atom model.

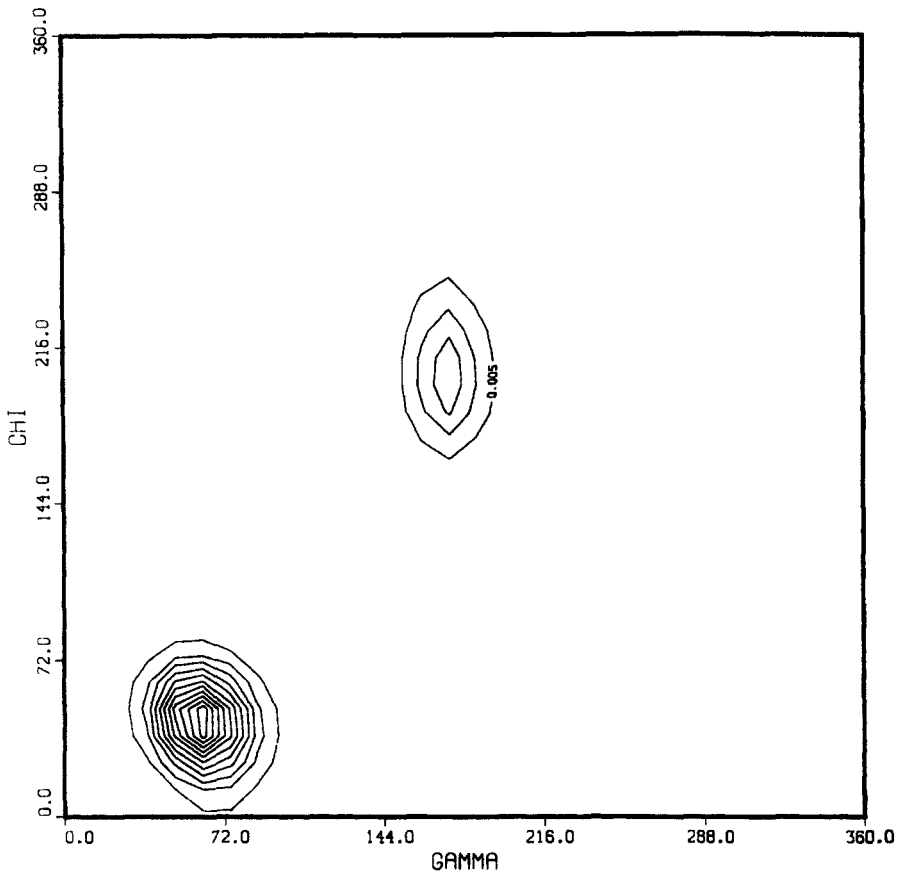

FIGURE 8. γ–χ probability contours for the all atom model.

atom, united atom, and united atom-hydroxyl models on adenosine and with the all atom and united atom-hydroxyl models on deoxyadenosine. Note that Pearlman and Kollman use the quartet of atoms, $O_{4'}—C_{1'}—N_9—C_8$ to define χ, which differs from the definition given here by 180°.

The same set of free-energy minima for the dihedral γ as calculated here was also observed by Pearlman and Kollman with both adenosine and deoxyadenosine. However, with deoxyadenosine in the united atom-hydroxyl representation, they found that E_{g^-} was lower than E_{g^+} and E_a, in direct contradiction with experimental observations. This is surprising because the presumably less accurate united atom model produces a more reasonable PMF map for γ. Pearlman and Kollman also note that their minimized potential energy results for γ in adenosine are in better agreement with experiment than their free-energy results. The PMF profiles of γ generated by WHAM agree well on a qualitative level with those calculated by Pearlman and Kollman for adenosine; this is not surprising because the nature of the sugar is known not to cause any qualitative changes in conformational behavior.

For the dihedral χ, Pearlman and Kollman found that the united atom model for adenosine did not exhibit the bimodal pattern, and the PMF had only one significant syn minimum. This lack of bimodal behavior may be due to improper conformational space sampling caused by constraints (which are necessary in PMF calculations by the method of Tobias and Brooks) on both the γ and χ dihedrals. Here, WHAM is seen to have produced a result in better agreement with experiment. The all atom χ PMF profiles for adenosine calculated by Pearlman and Kollman also exhibit bimodal behavior, but with the anti conformer as the global minimum. Given the good stereochemistry of the $O_{5'}—H\cdots N_3$ hydrogen bond, (g^+, syn) probably is the correct global minimum for adenosine. Hence WHAM is giving a better answer here.

The features of the two-dimensional PMF contour maps found by WHAM are similar to the united atom-hydroxyl γ–χ plot for deoxyadenosine calculated by Pearlman and Kollman, al-

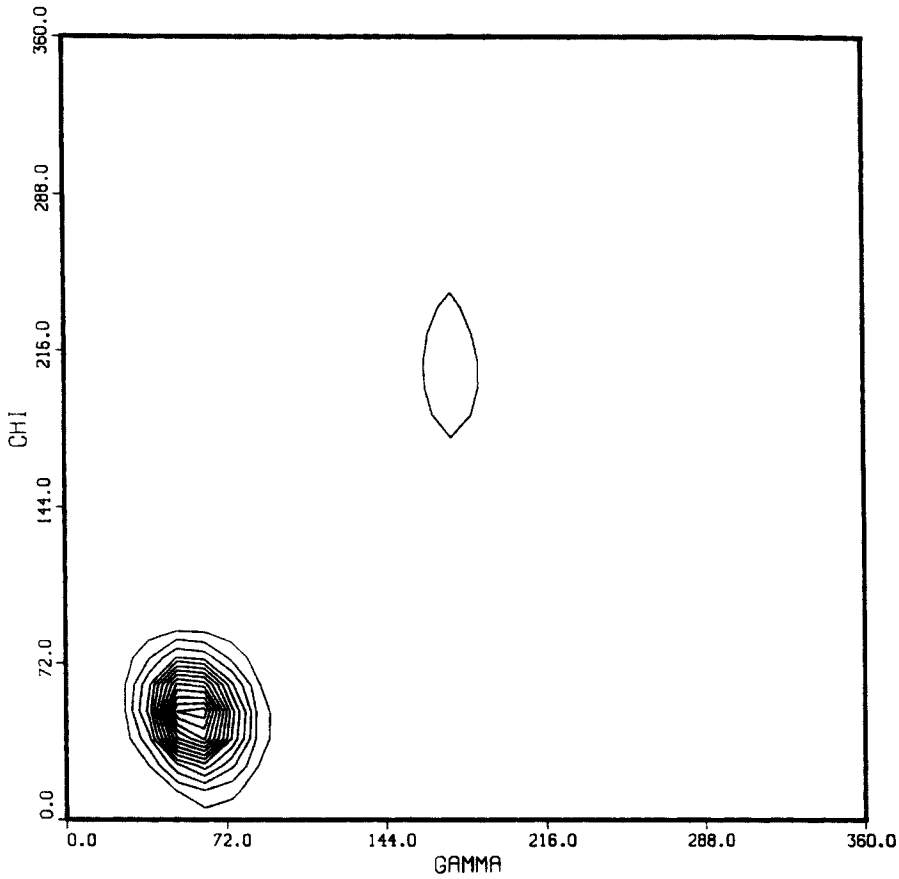

FIGURE 9. γ–χ probability contours for the united atom model.

though the bimodality of χ is not obvious from their contours. When the γ–χ PMF contours for the all atom model (Fig. 4) are compared with Pearlman and Kollman's minimized potential energy γ–χ maps, it can be seen that qualitatively the maps are largely similar. This is consistent with the broad similarity Pearlman and Kollman found between the γ–χ PMF contours and minimized energy contours in the united atom-hydroxyl representation of deoxyadenosine.

The inverse of the relative errors,[1] which in the present case is simply the square root of the total number of counts in a particular bin, is shown in Figure 10. The logarithm of the total number of snapshots in each of the bins is shown as a contour plot in Figure 11. With WHAM, it should be noted that the contributions from all bins are taken into account to estimate probabilities (unlike umbrella sampling) so that the errors due to the low countings in the high-energy regions [e.g., the (72°, 108°) in Figs. 10 and 11] are minimized; in fact, it is by minimizing such errors that the WHAM equations are derived.[1] Naturally, when doing extrapolations

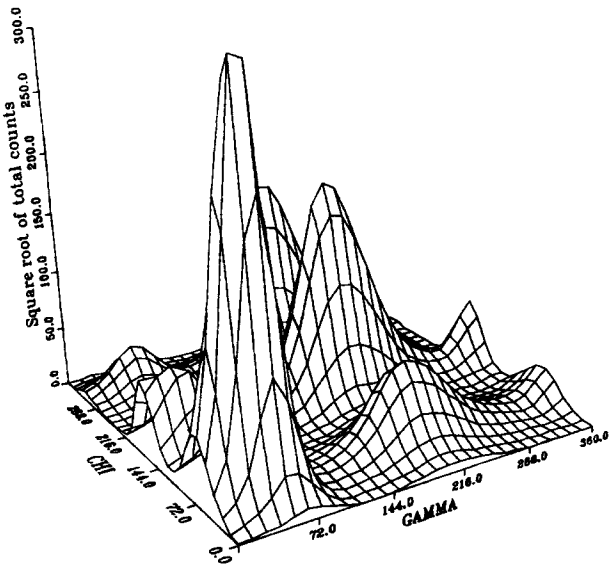

FIGURE 10. Square root of the total counts in the individual bins on the γ–χ plane (all atom model).

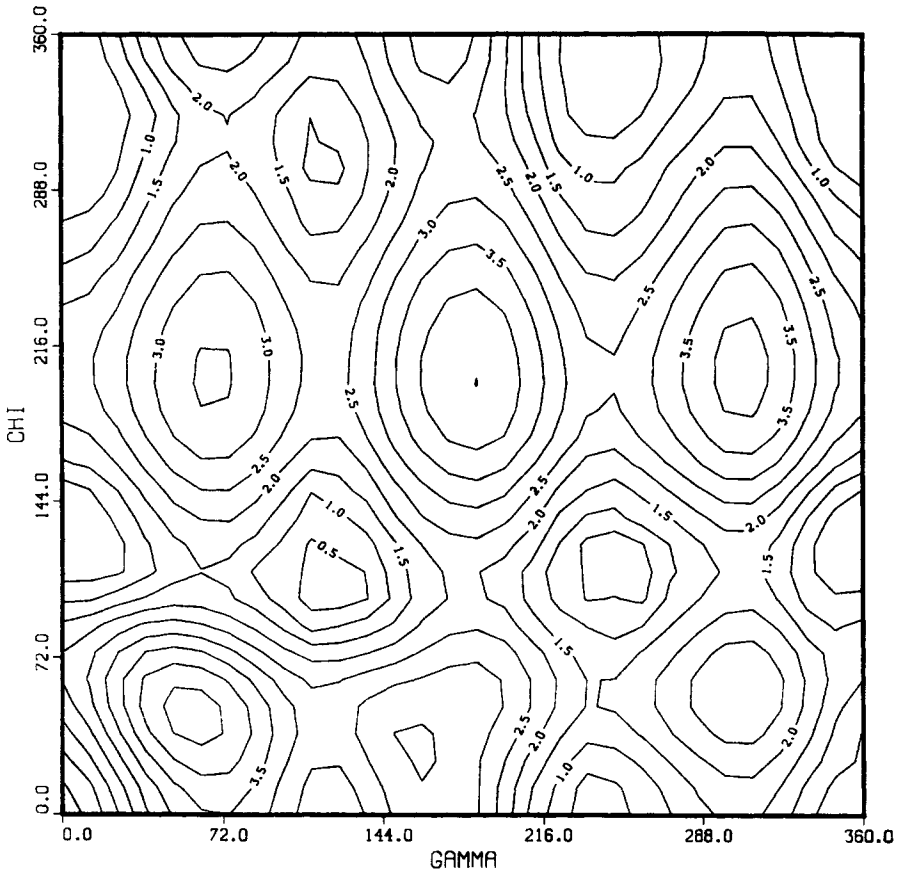

FIGURE 11. Contour plot of the logarithm of the total number of counts (all atom model).

or varying the temperature, it is particularly important to analyze the errors and to determine the range of validity of the extrapolations.[21] This matter is discussed in greater detail in the new preprint by Ferrenberg, Landau, and Swendsen.[22]

The simple form of the error equation makes WHAM very useful in determining with what degree of confidence one can make estimates of probabilities and free energies. Also, the simulations can be tailored to address the specific problem that is to be solved. For instance, if one is interested in near-exact description of high-energy transition regions [e.g., (72°, 108°)], then more sampling must be achieved in these regions; on the other hand, if an overall picture of the free energy surface is desired, the PMF plots shown in Figures 2–5 are adequate. Further analyses of errors and sampling statistics will be explained in a subsequent article.

If one were interested in correlations between the behavior of the γ–χ surface and any other parameter (e.g., the pseudorotation phase angle of the furanose ring), the dimensionality of the PMF surfaces can be easily increased to include the pseudorotation phase angle. The use of such correlations can be useful in the evaluation and reparameterization of force fields and in the investigation of structure/function relationships in biomolecules.

WHAM can be effectively used to do a free-energy component analysis in molecular interactions. For instance, if one were interested in finding out the free-energy contribution of one (or many) particular hydrogen bond(s), then partitioning the Hamiltonian of the system under investigation (into self-explanatory components) as

$$H = H_{\substack{\text{rest of} \\ \text{the system}}} + \lambda H_{\substack{\text{hydrogen} \\ \text{bond(s)}}}$$

and varying λ from zero to unity the hydrogen bond(s) contribution to the total free energy can be determined.

To summarize, the WHAM technique is a straightforward and powerful method to compute multidimensional PMF/free-energy maps. It can

be used for generating multidimensional PMFs of complex reaction coordinates (like pseudorotation phase angles[1]) without having to constrain any physical parameter (like dihedral angles). This is a distinct advantage over the methods that are currently used, especially when the effects of constraining conformational parameters on sampling cannot be known *a priori*. The WHAM technique incorporates multiple overlaps of probability distributions to exploit all statistical data and enhance overall accuracy.

Acknowledgments

This work was supported by grants from NIH (GM25671) and the W. M. Keck Center for Advanced Training in Computational Biology at the University of Pittsburgh, Carnegie Mellon University, and the Pittsburgh Supercomputing Center. Supercomputer time was provided by the Pittsburgh Supercomputing Center through grant 90026P. P. A. K. acknowledges support received from NIH grant CA-25644 and R. H. S. from NSF grant DMR-9221333.

References

1. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, *J. Comp. Chem.*, **13,** 1011 (1992).
2. J. P. Valleau and D. N. Card, *J. Chem. Phys.*, **57,** 5457 (1972).
3. G. N. Patey and J. P. Valleau, *J. Chem. Phys.*, **63,** 2334 (1975).
4. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1987.
5. C. L. Brooks, M. Karplus, and M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Advances in Chemical Physics*, vol. LXXI, John Wiley & Sons, New York, 1988.
6. M. Mezei, P. K. Mehrotra, and D. L. Beveridge, *J. Am. Chem. Soc.*, **107,** 2239 (1985).
7. A. Anderson, M. Carson, and J. Hermans, *Ann. NY Acad. Sci.*, **482,** 51 (1986).
8. A. G. Anderson and J. Hermans, *Proteins*, **3,** 262 (1988).
9. D. Bouzida, S. Kumar, and R. H. Swendsen, *Phys. Rev. A*, **45,** 8894 (1992).
10. D. Bouzida, S. Kumar, and R. H. Swendsen, In *Computer Simulation Studies in Condensed Matter Physics III, Springer Proceedings in Physics 53*, D. P. Landau, K. K. Mon, and H.-B. Schuttler, Eds., Springer-Verlag, Berlin, 1991, p. 193.
11. D. A. Pearlman and P. A. Kollman, *J. Am. Chem. Soc.*, **113,** 7167 (1991).
12. D. J. Tobias and C. L. Brooks III, *J. Chem. Phys.*, **89,** 5115 (1988).
13. A. M. Ferrenberg, Ph.D. thesis, Efficient Use of Monte Carlo Simulation Data, Carnegie Mellon University, Pittsburgh, PA, 1989.
14. E. M. Boczko and C. L. Brooks III, *J. Phys. Chem.*, **97,** 4509 (1993).
15. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, **7,** 230 (1986).
16. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21,** 1087 (1953).
17. P. K. Weiner and P. A. Kollman, *J. Comp. Chem.*, **2,** 287 (1980).
18. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
19. S. Kumar, Ph.D. thesis, Dynamical Behavior of DNA, University of Pittsburgh, Pittsburgh, PA, 1990.
20. AMBER, Version 4.0, user's manual, University of California, 1991.
21. A. M. Ferrenberg and R H. Swendsen, *Phys. Rev. Lett.*, **61,** 2635 (1988).
22. A. M. Ferrenberg, D. P. Landau, and R. H. Swendsen, manuscript in preparation.