JCTC Journal of Chemical Theory and Computation

Replica Exchange and Multicanonical Algorithms with the Coarse-Grained United-Residue (UNRES) Force Field

Marian Nanias,[†] Cezary Czaplewski,[†,‡] and Harold A. Scheraga*[,†]

*Baker Laboratory of Chemistry and Chemical Biology, Cornell University,
Ithaca, New York 14853-1301, and Faculty of Chemistry, University of Gdansk,
Sobieskiego 18, 80-952 Gdansk, Poland*

Received October 14, 2005

Abstract: Three algorithms, namely, a replica exchange method (REM), a replica exchange multicanonical method (REMUCA), and a replica exchange multicanonical method with replica exchange (REMUCAREM), were implemented with the coarse-grained united-residue force field (UNRES) in both Monte Carlo and molecular dynamics versions. The MD algorithms use the constant-temperature Berendsen thermostat, with the velocity Verlet algorithm and a variable time step. The algorithms were applied to one peptide (20 residues of alanine with free ends; ala_{20}) and two small proteins, namely, an α-helical protein of 46 residues (the B domain of the staphylococal protein A; 1BDD) and an $\alpha+\beta$ protein of 48 residues (the *Escherichia coli* Mltd Lysm Domain; 1E0G). Calculated thermodynamic averages, such as canonical average energy and heat capacity, are in good agreement among all simulations for poly-L-alanine, showing that the algorithms were implemented correctly and that all three algorithms are equally effective for small systems. For protein A, all algorithms performed reasonably well, although some variability in the calculated results was observed, whereas for a more complicated $\alpha+\beta$ protein (1E0G), only replica exchange was capable of producing reliable statistics for calculating thermodynamic quantities. Finally, from the replica exchange molecular dynamics results, we calculated free-energy maps as functions of the RMSD and radius of gyration for different temperatures. The free-energy calculations show correct folding behavior for poly-L-alanine and protein A, while for 1E0G, the native structure had the lowest free energy only at very low temperatures. Hence, the entropy contribution for 1E0G is larger than that for protein A at the same temperature. A larger contribution from entropy means that there are more accessible conformations at a given temperature, making it more difficult to obtain an efficient coverage of conformational space to obtain reliable thermodynamic properties. At the same temperature, ala_{20} has the smallest entropy contribution, followed by protein A, and then by 1E0G.

1. Introduction

Efficient sampling algorithms have been an essential component of methods for studying protein structure and dynamics in structural biology and theoretical chemistry. A variety of sampling algorithms have been used in our laboratory, and depending on whether the goal is global optimization or folding simulations, they can be be categorized in the following way.

For successful prediction of the three-dimensional structure of a protein (based solely on its amino acid sequence), several classes of algorithms have been used. The first class includes modifications of the Metropolis Monte Carlo procedure,[1,2] such as Monte Carlo with minimization,[3,4] electrostatically

* Corresponding author tel: (607) 255-4034; fax: (607) 254-4700; e-mail: has5@cornell.edu.

[†] Cornell University.

[‡] University of Gdańsk.

10.1021/ct050253o CCC: $33.50 © 2006 American Chemical Society
Published on Web 01/21/2006

driven Monte Carlo,[5,6] conformational family Monte Carlo,[7] and replica exchange Monte Carlo with minimization.[8] The second class includes deformation-based methods, such as the diffusion-equation method,[9] the distance-scaling method,[10] and the self-consistent basin-to-deformed-basin method.[11,12] The third class includes genetic algorithms such as the conformational space annealing (CSA) method.[13−15] For the study of protein-folding pathways, recently applied molecular dynamics (MD) with the united-residue (UNRES) force field[16−19] have been shown to be particularly effective. To evaluate thermodynamic properties, another class of sampling methods is necessary. This is because minimization-based methods violate the condition of microscopic reversibility required for producing Boltzmann statistics, and although methods such as molecular dynamics or Metropolis Monte Carlo can be used for estimating thermodynamic properties as well as for a global search, they easily become trapped for complex systems and, thus, are not the most effective methods for studying large systems.

The origins of one of the most popular advanced sampling methods, the replica exchange method (also known as exchange Monte Carlo[20] or parallel tempering[21]), can be traced back to the work carried out by Swendsen and Wang[22] for spin-glass systems, and the more familiar form of the algorithm was developed by Geyer[23] with his use of Metropolis-coupled Markov chain Monte Carlo. In the replica exchange method, several copies (replicas) of the system are simulated with standard Metropolis Monte Carlo[1,2] or molecular dynamics procedures (each replica differing from the others in a particular way, usually in temperature), while permitting an exchange among the replicas, and thus surmounting barriers in the rugged conformational energy landscapes. This method has been applied extensively in protein-folding simulations using both lattice[24−27] and off-lattice models.[28−32]

Recently, much attention has been paid to generalized ensemble algorithms whose advantage is efficient sampling of the conformational energy landscape. In this approach, efficient sampling does not mean locating the global minimum as quickly as possible but rather covering the landscape in such a way as to provide accurate statistics. Two well-known methods are the multicanonical algorithm[33,34] (also known as entropy sampling[35,36]) and simulated tempering[37] (also referred to as the method of expanded ensembles[38]). The multicanonical algorithm performs a one-dimensional random walk in energy space, while simulated tempering follows a random walk in temperature space, thereby inducing a random walk in the space of potential energy. Although these algorithms are generally too expensive for locating global minima,[39] they are useful for producing accurate statistics for thermodynamic averages of observed variables. However, the application of these algorithms is nontrivial and very tedious; in particular, the need to obtain the proper sampling weights often limits the use of generalized ensemble techniques.[40]

Due to the fact that the replica exchange method alleviates the problem of the tedious estimation of weight factors in the multicanonical algorithms, combinations of replica exchange with generalized ensemble methods have been developed, for example, REMUCAREM,[41] that is, replica exchange multicanonical algorithm with replica exchange; others include replica exchange simulated tempering or simulated tempering replica exchange.[42] Other modifications of replica exchange include replica exchange with solute tempering,[43] model hopping,[44] Hamiltonian replica exchange,[45] and the replica-exchange method using a generalized effective potential.[46]

Having demonstrated that the coarse-grained UNRES protein model is helpful in surmounting problems with all-atom models,[18,47] we apply the replica exchange method (REM), the replica exchange multicanonical method (REMUCA), and the REMUCAREM method, in both Monte Carlo (MC) and molecular dynamics versions, to the UNRES model in the present work. The advantage of replica exchange lies in its simplicity, and in contrast to other methods, it is not very sensitive to the few parameters involved therein (such as the cooling schedule in simulated tempering or the successful estimation of weight factors in multicanonical algorithms). The power of REMUCA lies in the effective estimate of the multicanonical weight factors from replica exchange simulations. REMUCAREM further exploits the idea of running several replicas of multicanonical simulations with different sets of multicanonical weights. The motivation behind the present work is to test the applicability of these algorithms to determine the thermodynamic properties of large systems. The ability to compute thermodynamic properties will thereby enable us to improve our UNRES model and, consequently, improve protein-folding simulations, that is, bring our simulated results closer to experimental ones.

2. Methods

2.1. UNRES Force Field. All the above-mentioned algorithms were implemented with the united-residue force field; hence, in this section, the UNRES model of polypeptide chains and the corresponding force field are described briefly. First, the UNRES model used with Monte Carlo procedures is described, followed by a description of the UNRES force field for molecular dynamics.

In the UNRES model,[48−58] a polypeptide chain is represented by a sequence of α-carbon (C^α) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α-carbons. Only these united peptide groups and the united side chains serve as interaction sites, the α-carbons serving only to define the chain geometry. All virtual bond lengths (i.e., $C^\alpha \cdots C^\alpha$ and $C^\alpha \cdots SC$) are fixed; the distance between neighboring C^α's is 3.8 Å, corresponding to trans peptide groups, while the side-chain angles (α_{SC} and β_{SC}) and virtual-bond (θ) and dihedral (γ) angles can vary. The UNRES force field has been derived as a restricted free-energy (RFE) function of an all-atom polypeptide chain *plus the surrounding solvent*, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system (i.e., the degrees of freedom of the solvent, the dihedral angles χ for rotation about the bonds in the side chains, and the torsional angles λ for rotation of the peptide groups about the $C^{\alpha}\cdots C^{\alpha}$ virtual bonds).[52,53,59] The RFE is further decomposed into factors arising from interactions within and between a given number of united interaction sites.[53] Expansion of the factors into generalized Kubo cumulants[60] facilitated the derivation of approximate analytical expressions for the respective terms,[52,53] including the *multibody* or *correlation* terms, which are derived in other force fields from structural databases or on a heuristic basis.[61] The theoretical basis of the force field is described in detail in ref 53. The energy of the virtual-bond chain for Monte Carlo simulations is expressed by eq 1.

$$U_{\mathrm{MC}} = \sum_{i<j} U_{\mathrm{SC}_i\mathrm{SC}_j} + w_{\mathrm{SCp}} \sum_{i\neq j} U_{\mathrm{SC}_i\mathrm{p}_j} + w_{el} \sum_{i<j-1} U_{\mathrm{p}_i\mathrm{p}_j} + w_{\mathrm{tor}} \sum_i U_{\mathrm{tor}}(\gamma_i) + w_{\mathrm{tord}} \sum_i U_{\mathrm{tord}}(\gamma_i,\gamma_{i+1}) + w_{\mathrm{b}} \sum_i U_{\mathrm{b}}(\theta_i) + w_{\mathrm{rot}} \sum_i U_{\mathrm{rot}}(\alpha_{\mathrm{SC}_i},\beta_{\mathrm{SC}_i}) + w^{(3)}_{\mathrm{corr}} U^{(3)}_{\mathrm{corr}} + w^{(4)}_{\mathrm{corr}} U^{(4)}_{\mathrm{corr}} + w^{(3)}_{\mathrm{turn}} U^{(3)}_{\mathrm{turn}} + w^{(4)}_{\mathrm{turn}} U^{(4)}_{\mathrm{turn}} \qquad (1)$$

The term $U_{\mathrm{SC}_i\mathrm{SC}_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{\mathrm{SC}_i\mathrm{p}_j}$ denotes the excluded-volume potential of the side-chain−peptide-group interactions. The peptide-group interaction potential ($U_{\mathrm{p}_i\mathrm{p}_j}$) accounts mainly for the electrostatic interactions (i.e., the tendency to form backbone hydrogen bonds) between peptide groups p_i and p_j. U_{tor}, U_{tord}, U_{b}, and U_{rot} are the virtual-bond dihedral angle torsional terms, virtual-bond dihedral angle double-torsional terms, virtual-bond angle bending terms, and side-chain rotamer terms, respectively; these terms account for the local propensities of the polypeptide chain. The terms $U^{(m)}_{\mathrm{corr}}$ represent *correlation* or *multibody* contributions from the coupling between backbone−local and backbone−electrostatic interactions, and the terms $U^{(m)}_{\mathrm{turn}}$ are correlation contributions involving m consecutive peptide groups; they are, therefore, termed turn contributions. The correlation contributions were derived[52,53] from a generalized-cumulant expansion[60] of the RFE of the system consisting of the polypeptide chain and the surrounding solvent. The multibody terms are indispensable for reproduction of regular α-helical and β-sheet structures.

The internal parameters of $U_{\mathrm{p}_i\mathrm{p}_j}$, U_{tor}, U_{tord}, $U^{(m)}_{\mathrm{corr}}$, and $U^{(m)}_{\mathrm{turn}}$ were derived by fitting the analytical expressions to the RFE surfaces of model systems computed by quantum mechanics at the MP2/6-31G** ab initio level,[57,58] while the parameters of $U_{\mathrm{SC}_i\mathrm{SC}_j}$, $U_{\mathrm{SC}_i\mathrm{p}_j}$, U_{b}, and U_{rot} were derived by fitting the calculated distribution functions to those determined from the PDB.[51] The w's are the weights of the energy terms, and they were determined (together with the parameters within each cumulant term and the well depths of the side-chain pairwise interaction potential $U_{\mathrm{SC}_i\mathrm{SC}_j}$) by hierarchical optimization[62] of the potential-energy function.

Molecular dynamics with UNRES requires an extra degree of freedom, namely, the vibrations of the virtual bonds, which are treated with an additional harmonic potential. The complete UNRES potential-energy function for molecular dynamics is then expressed by the following equation:[18]

$$U_{\mathrm{MD}} = U_{\mathrm{MC}} + w_{\mathrm{vib}} \sum_i U_{\mathrm{vib}}(d_i) \qquad (2)$$

where U_{MC} is the Monte Carlo UNRES potential energy described above (eq 1) and $U_{\mathrm{vib}}(d_i)$, d_i being the length of the ith virtual bond, are the simple harmonic potentials defined as $U_{\mathrm{vib}}(d_i) = (1/2)k_{d_i}(d_i - d^{\circ}_i)^2$, where k_{d_i} is the force constant of the ith virtual bond, currently set at 500 kcal/(mol Å^2), and d°_i is the average length (corresponding to that used in the fixed-bond UNRES potential) of the ith virtual bond; for example, $d^{\circ}_i = 3.8$ Å for a $C^{\alpha}\cdots C^{\alpha}$ virtual bond corresponding to a trans peptide group. As in previous work,[18] the weight w_{vib} was arbitrarily set at 1.

2.2. Replica Exchange Method (REM).

The replica exchange method is an extension of the Metropolis Monte Carlo, or molecular dynamics, methods. The underlying idea is to run different copies (replicas) of the system at different levels of a certain property (such as temperature). Although different properties have been considered in published work,[45,46] the property of change across different replicas (i.e., how replicas differ from one another) in the present context is temperature. To summarize the method, a MC or MD simulation is carried out on each selected conformation at its assigned temperature for a determined number of MC or MD steps, after which, the neighboring replicas undergo an exchange with the acceptance criterion described below (in eq 4). Let

$$\Delta \equiv [(\beta_m - \beta_n)\{E(\mathrm{Y}) - E(\mathrm{X})\}] \qquad (3)$$

where β_m is the inverse temperature defined as $1/(k_{\mathrm{B}}T_m)$ and $E(\mathrm{X})$ is the energy of conformation X. If one adopts the Metropolis method, the replica-exchange transition probability can be expressed as

$$W(\mathrm{X},\beta_m|\mathrm{Y},\beta_n) = \begin{cases} 1 & \text{for } \Delta \leq 0 \\ \exp(-\Delta) & \text{for } \Delta > 0 \end{cases} \qquad (4)$$

That is, if Δ is less than or equal to 0, the exchange is performed (since the probability is 1); otherwise, a random number between 0 and 1 is generated and compared to the factor $\exp(-\Delta)$. If the value of this factor is smaller, the exchange is performed; otherwise, the exchange is rejected.

To evaluate thermodynamic quantities at any temperature, it is essential to extract maximum information from all replicas. For this purpose, a multihistogram reweighting technique[63,64] can be used. For a replica exchange simulation with M replicas at M distinct temperatures, a set of M energy histograms $N_m(E)$ is obtained. The densities of states $[n(E)]$ are then obtained self-consistently from the following WHAM[63,64] equations:

$$n(E) = \frac{\sum_{m=1}^{M} g_m^{\ -1}\, N_m(E)}{\sum_{m=1}^{M} g_m^{\ -1} n_m \exp(f_m - \beta_m E)} \qquad (5)$$

and

$$\exp(-f_m) \equiv \sum_E n(E) \exp(-\beta_m E) \tag{6}$$

where $N_m(E)$ is the histogram at temperature T_m, $\beta_m = 1/(k_B T_m)$ is the inverse temperature, n_m is the total number of samples in the mth replica, $g_m = 1 + 2\tau_m$, and τ_m is the integrated autocorrelation time at temperature T_m. In biomolecular systems, g_m is approximately constant[64] and, therefore, can be canceled in eq 5. The WHAM eqs 5 and 6 are evaluated self-consistently, and the resulting densities of states are used to evaluate the expectation value of any observable A in eq 7:

$$\langle A\rangle_T = \frac{\sum_E A(E)\, n(E) \exp(-\beta E)}{\sum_E n(E) \exp(-\beta E)} \tag{7}$$

2.3. Multicanonical Algorithm (MUCA). A single canonical simulation (MC or MD) by definition samples a very restricted energy region. Furthermore, when sampling the conformations of the protein in low-energy regions, the multiple-minima problem is usually encountered and the simulation can be trapped in a particular local energy minimum, making it difficult to obtain a reliable estimate of the density of states of proteins. In determining the density of states of a large system by simulation procedures, a clear criterion is needed about the stage of simulations at which all of the conformational space of the protein has been sampled sufficiently. Traditional MC or MD procedures do not provide such a convergence criterion. For these reasons, a multicanonical algorithm[33,34] (also known as entropy sampling[35,36]) has been used for protein studies. In Section 2.4, we show why MUCA is combined with REM to produce REMUCA, whose efficiency is explored in the present work. For this purpose, we first outline MUCA. In the next paragraph, we present the background of entropy sampling and tie it together with the multicanonical algorithm notation.

In the present work, we use the term “conformation” to indicate a particular structure and the term “state” to denote all the conformations that either have a given energy or are within a small energy interval. The probability of occurrence of a conformation x with energy E, denoted as $P(\mathrm{x})$, and the probability of occurrence of a state with energy E, denoted as $P(E)$, are related to each other in a canonical ensemble by the following relations, with E being written for $E(\mathrm{x})$:

$$P(\mathrm{x}) \propto \exp(-\beta E) \tag{8}$$

$$P(E) \propto n(E) \exp(-\beta E) = \exp[S(E)/k_B - \beta E] \tag{9}$$

where k_B is the Boltzmann constant, $\beta = 1/k_B T$ with T being the temperature, $n(E)$ is the number of conformations with energy E (i.e., density of states), and $S(E) = k_B \ln[n(E)]$ is the entropy of the state with energy E.

The entropy sampling method is based on an artificial distribution of states, in which the probability of occurrence of a state with energy E is scaled by the exponential of the *negative* of the entropy of the state, $S(E)$. In entropy sampling, the probabilities of occurrence of a conformation x and a state with energy E, respectively, are defined as

$$P(\mathrm{x}) \propto \exp\{-S[E(\mathrm{x})]/k_B\} \tag{10}$$

$$P(E) \propto n(E) \exp[-S(E)/k_B] \tag{11}$$

where $n(E)$ and $S(E)$ have similar meanings as described above. Equations 10 and 11 can be related to eqs 8 and 9 by first setting $\beta = 0$ (i.e., temperature to infinity) in eqs 8 and 9 and then multiplying the resulting probabilities by the weight factor $\exp[-S(E)/k_B]$. The physical meaning of this modification is that the larger the conformational entropy of a state, the smaller is the weight given to the state. In this way, the probabilities of occurrence of all states with different energies are constant in the new distribution; that is, $P(E)$ of eq 11 is a constant, taken as 1.

To connect the entropy sampling formalism to the commonly used multicanonical algorithm, we can define a new variable, the multicanonical energy E_{mu}, in the following way

$$E_{mu}(E;T_0) = T_0 S(E) = k_B T_0 \ln[n(E)] \tag{12}$$

where T_0 is the reference temperature and $S(E)$ is the microcanonical entropy as above. The reference temperature is the temperature at which the MC or MD multicanonical simulation is carried out. It should be noted that the reference temperature theoretically plays no role in calculating thermodynamics, because the formula for obtaining thermodynamic quantities (eq 7) is independent of T_0; however, in practice, the value chosen for T_0 affects the sampling efficiency of numerical simulations. Equations 10 and 11 then become

$$P(\mathrm{x}) \propto \exp\{-E_{mu}[E(\mathrm{x});T_0]/T_0 k_B\} \tag{13}$$

and

$$P(E) \propto n(E) \exp[-E_{mu}(E;T_0)/T_0 k_B] \tag{14}$$

Consequently, the multicanonical Monte Carlo simulation is carried out with the following modified Metropolis acceptance criterion:

$$W(X|Y) = \begin{cases} 1 & \text{for } \Delta E_{mu} \le 0 \\ \exp(-\beta_0 \Delta E_{mu}) & \text{for } \Delta E_{mu} > 0 \end{cases} \tag{15}$$

where $\beta_0 = 1/k_B T_0$, T_0 being a reference temperature, and $\Delta E_{mu} \equiv E_{mu}[E(\mathrm{Y});T_0] - E_{mu}[E(\mathrm{X});T_0]$.

The multicanonical molecular dynamics simulation is carried out by integrating the following modified Newton equation;[65-67] see eq 21 of ref 65:

$$\dot{p}_k = -\frac{\partial E_{mu}(E;T_0)}{\partial q_k} = \frac{\partial E_{mu}(E;T_0)}{\partial E} f_k \tag{16}$$

where $\dot{p}_k$ is the momentum, q_k is the generalized coordinate of the kth atom, and f_k is the force on the kth atom. Specifically, the UNRES MD equation of motion (eq 32 of ref 16) is modified as

$$\ddot{q}(t) = -G^{-1} \frac{\partial E_{mu}(U;T_0)}{\partial U} \nabla_q U[q(t)] \tag{17}$$

where U [being $U(x)$] is the UNRES potential energy (U_{MD} of eq 2), $q(t)$ are the generalized coordinates at time t, and

G is the mass matrix (eq 26 of ref 16). In practice, one can use cubic splines to approximate $\partial E_{mu}(U;T_0)/\partial U$.

Because the density of states is usually not known a priori, the multicanonical weights are usually obtained by iterating short runs;[36,68−70] that is, E_{mu} is obtained such that eq 14 is constant for all energies E. For this purpose, one uses the single histogram reweighting technique to obtain a new estimate of the densities of states after each iteration:

$$n(E) = \frac{N_{mu}(E)}{\exp[-\beta_0 E_{mu}(E;T_0)]} \tag{18}$$

where N_{mu} is the histogram obtained from the multicanonical simulation (either MC or MD) and $\exp[-\beta_0 E_{mu}(E;T_0)] = 1/n(E)$ are the input multicanonical weights. The new estimates of the density of states are then used in eq 12 to obtain new values of E_{mu} and, hence, new input weights. This procedure is repeated until the histogram N_{mu} obtained from the multicanonical simulation is sufficiently flat (i.e., the probability of visiting any part of the energy space is constant). The resulting weights are then used for a long multicanonical simulation, from which thermodynamic quantities can be calculated.

To obtain expected averages from a multicanonical simulation, the single histogram reweighting technique (eq 18) is first used to obtain a new estimate of the densities of states. The new estimates of densities of states are then used in eq 7 to obtain the thermodynamic averages.

2.4. Replica Exchange Multicanonical Algorithm (REMUCA). MUCA without REM converges very slowly and consequently is inefficient.[71−73] Therefore, we have explored the use of REMUCA, which differs from MUCA in how the starting weights for the simulation are obtained. While MUCA requires short iterative multicanonical simulations, REMUCA obtains the starting weights from a short replica exchange simulation, by first obtaining the densities of states from REM, which are then used to estimate the multicanonical weights $\{\exp[-E_{mu}(E;T_0)/k_B T_0]\}$ with eq 12. In practice, the values for the multicanonical potential energy, $E_{mu}(E;T_0)$, obtained from replica exchange, are reliable only in the range of $\langle E\rangle_{T_{\min}} \le E \le \langle E\rangle_{T_{\max}}$, where $T_{\min}$ and $T_{\max}$ are the lowest and highest temperatures in REM, respectively, and $E_{\min} = \langle E\rangle_{T_{\min}}$ and $E_{\max} = \langle E\rangle_{T_{\max}}$ are the canonical expectation values at those temperatures; that is, we use multicanonical sampling only in the region between $E_{\min}$ and $E_{\max}$ and canonical sampling outside of this region. The reason the weights are reliable only between $E_{\min}$ and $E_{\max}$ is because $T_{\min}$ and $T_{\max}$ (which determine $E_{\min}$ and $E_{\max}$) are chosen arbitrarily for the REM simulation, such that the region sampled by overlapping replicas between $E_{\min}$ and $E_{\max}$ contains both the native structure and the most probable non-native structures. Therefore, the best region sampled by REM is the one between $E_{\min}$ and $E_{\max}$, which determines that the multicanonical input weights should be reliable only between $E_{\min}$ and $E_{\max}$. In principle, any sampling can be used below $E_{\min}$ and above $E_{\max}$ as long as the simulation returns back to the multicanonical region which should contain both the native structure and the most probable non-native structures; in practice, this calculation has been carried out with canonical sampling.

The only reason to explore the canonical region is to force a random walk from the multicanonical region, which may have wandered out of the multicanonical region, to return to the multicanonical region. In essence, by sampling for thermodynamic data only in the multicanonical region, it is being assumed that the multicanonical region is large enough to encompass both the native structure and the more probable (i.e., lower-energy) parts of the ensemble of non-native structures. In addition, at the upper ($E_{\max}$) and lower energy ($E_{\min}$) boundaries between the multicanonical and canonical regions, the constant probability in the multicanonical region decreases in the canonical region.

The canonical sampling is carried out by extrapolating the multicanonical energies $[E_{mu}(E;T_0)]$ linearly.[71] It should be noted that only data from the multicanonical region (between $E_{\min}$ and $E_{\max}$) are used for calculating thermodynamic properties. Hence, the energy space in REMUCA is divided into three regions as follows:

$$\epsilon^0_{mu}(E) \equiv \begin{cases} E_{mu}(E_{\min};T_0) + \left.\dfrac{\partial E_{mu}(E;T_0)}{\partial E}\right|_{E_{\min}} (E - E_{\min}) & \text{for } E \le E_{\min} \\ & \quad\text{(canonical)} \\ E_{mu}(E;T_0) & \text{for } E_{\min} \le E \le E_{\max} \\ & \quad\text{(multicanonical)} \\ E_{mu}(E_{\max};T_0) + \left.\dfrac{\partial E_{mu}(E;T_0)}{\partial E}\right|_{E_{\max}} (E - E_{\max}) & \text{for } E \ge E_{\max} \\ & \quad\text{(canonical)} \end{cases}$$

where $\epsilon^0_{mu}(E)$ is substituted for $E_{mu}(E;T_0)$ in eqs 15 (for MC) and 17 (for MD) and T_0 is the reference temperature for the Monte Carlo and molecular dynamics simulations (the temperature at which the MC or MD simulation is carried out). Again, the reference temperature bears no significance in the results of the thermodynamic quantities (because eq 7 is independent of T_0). The rest of the simulation for both MC and MD proceeds as in a traditional MUCA simulation (eq 15 for MC and eq 17 for MD) with ϵ^0_{mu} replacing E_{mu}.

2.5. Multicanonical Replica-Exchange Method (MUCAREM). We also explore the use of the REMUCAREM algorithm, whose core is the same as that of the MUCAREM algorithm. Therefore, we first present the theoretical background of MUCAREM and later extend the discussion to REMUCAREM. Just as REM consists of several replicas of canonical MC or MD simulations, MUCAREM consists of several replicas of multicanonical simulations. The difference between REM and MUCAREM is that the replicas in REM are associated with different temperatures whereas, in MUCAREM, the replicas are associated with different energy ranges over which multicanonical simulations are carried out. The advantage of the MUCAREM approach over the traditional REM is that the probability distributions of energies of different replicas are broader in MUCAREM than in REM; therefore, a smaller number of replicas is required to cover the entire energy range.

The starting weights are obtained by short iterations of MUCA simulations, as described earlier in Section 2.3. The following procedures are carried out in *each* cycle:

1. Select an energy range for each replica, for which the replica will carry out the MUCA simulation. This energy

range of a given replica should overlap the energy ranges of the neighboring replicas, and the combined energy range from all replicas should cover the whole energy space (i.e., the combined energy range should contain the native structure and the most probable non-native structures). Assign a different random protein conformation to each energy range.

2. A MUCA simulation with MC or MD is carried out on each selected conformation within its energy range for a determined number of MC or MD steps. The MC or MD simulations are carried out with eqs 15 or 17, respectively, where E_{mu} is replaced by ϵ_{mu}^{m} defined as follows:

$$\epsilon_{mu}{}^{m}(E) \equiv \begin{cases} E_{mu}(E^{m}_{\min};T_m) + \left.\dfrac{\partial E_{mu}(E;T_m)}{\partial E}\right|_{E^{m}_{\min}} (E - E^{m}_{\min}) & \text{for } E \le E^{m}_{\min} \\ & \quad\text{(canonical)} \\ E_{mu}(E;T_m) & \text{for } E^{m}_{\min} \le E \le E^{m}_{\max} \\ & \quad\text{(multicanonical)} \\ E_{mu}(E^{m}_{\max};T_m) + \left.\dfrac{\partial E_{mu}(E;T_m)}{\partial E}\right|_{E^{m}_{\max}} (E - E^{m}_{\max}) & \text{for } E \ge E^{m}_{\max} \\ & \quad\text{(canonical)} \end{cases}$$

where m is the replica index (m = min...max) and min and max are the lowest and highest temperature replicas. $E^{m}_{\min}$ is then the canonical expectation value of the energy of the mth replica at temperature $T^{m}_{\min}$ [$E^{m}_{\min} = \langle E\rangle_{T^{m}_{\min}}$], and similarly, $E^{m}_{\max}$ is the canonical expectation value of the energy of the mth replica at temperature $T^{m}_{\max}$ [$E^{m}_{\max} = \langle E\rangle_{T^{m}_{\max}}$] for the mth multicanonical replica. It should be noted that $T^{m}_{\min}$ and $T^{m}_{\max}$ are different for different replicas (for different m's) and, thus, determine a different multicanonical energy range $E^{m}_{\min}$ and $E^{m}_{\max}$ for different replicas. Therefore, the multicanonical simulation with each replica is carried out in a different energy range ($E^{m}_{\min}$ and $E^{m}_{\max}$).

3. After carrying out a selected number of MC or MD steps, stop the simulation of each replica and attempt an exchange of the whole conformations between neighboring replicas with the following transition probability:

$$W(\mathrm{Y}|\mathrm{X}) = \begin{cases} 1 & \text{for}\Delta\le 0 \\ \exp(-\Delta) & \text{for}\Delta > 0 \end{cases} \tag{19}$$

where $\Delta \equiv \beta_{m+1}\{\epsilon_{mu}^{m+1}[E(\mathrm{Y})] - \epsilon_{mu}^{m+1}[E(\mathrm{X})]\} - \beta_m\{\epsilon_{mu}^{m}[E(\mathrm{Y})] - \epsilon_{mu}^{m}[E(\mathrm{X})]\}$.

4. Continue the simulation with each newly formed conformation at each new energy range as in step 2.

5. Iterate points 3 and 4 until the system sufficiently covers the entire energy range.

As in REM, the densities of states are obtained from a self-consistent evaluation of the following modified WHAM equations:

$$n(E) = \frac{\sum_{m=1}^{M} g_m{}^{-1}\, N_m(E)}{\sum_{m=1}^{M} g_{\mathrm{m}}{}^{-1} n_m \exp[f_m - \beta_m \epsilon_{mu}^{m}(E)]} \tag{20}$$

and

$$\exp(-\, f_m) \equiv \sum_{E} n(E) \exp[-\beta_m \epsilon_{mu}^{m}(E)] \tag{21}$$

where $N_m(E)$ is the histogram at temperature T_m, $\beta_m = 1/(k_{\mathrm{B}}T_m)$ is the inverse temperature, n_m is the total number of samples in the mth replica, and g_m is defined as in Section 2.2. The resulting densities of states are then used to evaluate the expectation value of any observable in eq 7, with g_m canceling out, as in eq 5.

2.6. Replica Exchange Multicanonical with Replica-Exchange Method (REMUCAREM). MUCAREM without input weights from REM converges very slowly and, consequently, is inefficient.[71−73] Therefore, we have explored the use of REMUCAREM, which, as in REMUCA, obtains the starting weights from replica exchange simulations as opposed to iterative short MUCA simulations. Everything else proceeds in the same manner as in MUCAREM.

3. Implementation Details

All the simulations were carried out on one peptide (20 residues of alanine with free ends; ala$_{20}$) and two small proteins, namely, the B domain of staphylococal protein A (an α-protein; 46 residues; 1BDD)[74] and the *Escherichia coli* Mltd Lysm domain (an α+β protein; 48 residues; 1E0G).[75] The ala$_{20}$ peptide was used to check whether the algorithms perform correctly, and the proteins were chosen so that basic α and α+β topologies were tested, and their size was reasonable with respect to the computational time. As in our previous work,[76] the length of protein 1BDD was shortened from the original 60 residues in the PDB to 46 residues. The set of UNRES energy parameters, designated as the 4P force field[76] and used in the present work, was derived by optimizing the parameters for four proteins simultaneously: 1E0L[77] (a β protein; 37 residues), 1E0G[75] (an α+β protein; 48 residues), 1IGD[78] (an α+β protein; 61 residues), and 1GAB[79] (an α protein; 53 residues).

The MC simulations with REM, REMUCA, and REMUCAREM were carried out as follows. All four UNRES angles in every residue of the protein were subjected to a perturbation. One MC sweep consisted of updating all of these angles for each residue in the sequence, with a Metropolis evaluation after each perturbation. The MD simulations with these same algorithms were carried out with the Berendsen thermostat,[80] using the velocity Verlet algorithm[81] with a variable time step to integrate the equations of motion. The variable time step was accomplished by scaling the time step δt by powers of 2.[16] The cutoff change of acceleration δa_{cut} for the scaling procedure was increased to δa_{cut} = 4 Å/mtu,[16] to allow for the multiplication of the forces in the modified Newton equation (in eq 17, MUCA MD utilizes a factor that multiplies the forces, i.e., accelerations, which would cause the maximum change of acceleration $\delta a_{\max}$ to exceed the cutoff value δa_{cut}, and thus, the time step would be unnecessarily reduced). The time step was set at 4.89 fs to yield stable trajectories.[16] However, this is only a formal time step, and because of the reduction of the number of degrees of freedom in UNRES, the time step is several times larger compared with that of all-atom MD (see ref 16 for details). The coupling constant to the thermal bath was increased to

Table 1. Parameters Used in ala_{20} Simulations[a]

	MD			MC		
simulation	replicas	temp	steps	replicas	temp	sweeps
REM	16	400−2000	16,000,000	30	100−2000	2,000,000
REMUCA	1	100	10,000,000	1		1,000,000
REMUCAREM	2	100,101	20,000,000	2		1,000,000

[a] The replicas column shows the number of replicas used for each simulation. The temp shows the reference temperature (K) or range of temperatures for simulations (for REMUCA MC and REMUCAREM MC, the reference temperature cancels out in the equations; therefore, the corresponding fields are empty; this is because REMUCA and REMUCAREM depend only on the input weights which are independent of *T*, whereas in REM, the replicas differ from one another in temperature, and therefore, temperature does not cancel out). The step is the number of UNRES MD time steps, where the maximum time step was set to 4.9 fs in all MD simulations. A sweep is defined as perturbing all four angles at all the positions along the peptide sequence (for ala_{20}, one sweep is equal to 80 energy evaluations).

0.2445 ps to overcome the limitation of the Berendsen thermostat and produce a more Boltzmann-like distribution.[17] Replica exchange MD was carried out using multiplexing,[82] in which several replicas were simulated at each temperature. Since MC lacks the gradient and is consequently much less efficient at exploring the energy space than MD, the temperature range in the MC version of REM was lower than that of the REM MD simulations (so that the low-temperature replicas in REM MC would involve a sufficient number of moves to explore the low energy basins), and the number of replicas and the frequency of exchange in REM were much higher in MC. In all the simulations (both MC and MD), the system was equilibrated for 20% of the simulation length, and the last 80% of the simulation was used for the calculations. All Monte Carlo simulations were started from random conformations, and the starting point for all molecular dynamics simulations was an extended chain; because the system was equilibrated and because REM uses high-temperature replicas and both REMUCA and REMUCAREM perform a random walk in the energy space, the simulations were independent of the starting conditions.

4. Results and Discussion

4.1. Poly-L-alanine. First, to test the algorithms, a very simple poly-L-alanine system (20 residues) was chosen, and REM, REMUCA, and REMUCAREM simulations were carried out with both MC and MD. The parameters used in all simulations for ala_{20} are shown in Table 1. REM simulations were carried out first, from which the densities of states were obtained. It was found that the densities of states obtained from REM simulations were not precise enough for REMUCA, because REMUCA simulations did not perform a random walk (i.e., did not have flat energy histograms). Therefore, after the first iteration of REMUCA simulations, the densities of states were reweighted with eq 18, and with these weights, a second iteration of REMUCA simulations was carried out. The second set of weights used for REMUCA was also used for REMUCAREM simulations. The simulation weights for alanine are shown as a solid or dashed curve in Figure 1. The dashed line shows an example of the multicanonical energy function (eq 12), used in the modified Metropolis criterion in MC simulations (eq 15), while the solid line shows its derivative, a factor multiplying the force in the modified Newton equation (eq 17). The results are summarized in Figures 2 and 3.

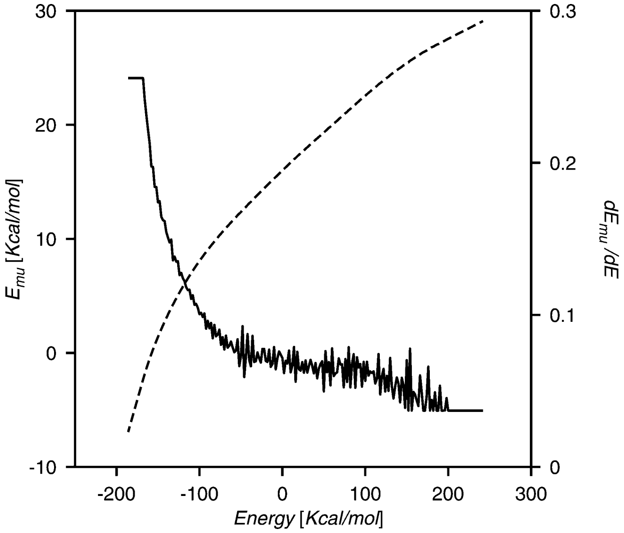

Figure 1. Parameters used for multicanonical simulations. The dashed line denotes the multicanonical energy function (eq 12), while the solid line denotes the derivative of this function fitted with cubic splines. The derivatives are used as a multiplicative factor [$\partial E_{mu}(U;T_0)/\partial U$] in the modified Newton equation (eq 17) in molecular dynamics. The flat regions of the derivative curve show where the multicanonical simulation changes to the canonical simulation.

Figure 2 consists of six plots. Three plots on the top correspond to MC simulations, whereas the three plots on the bottom correspond to MD simulations. The two plots in each column are for REM, REMUCA, and REMUCAREM simulations, respectively. Each plot depicts the logarithm of the probabilities ln[$P(E)$] as a function of energy (E) for the given simulation. By comparing the top row to the bottom row, it can be seen that MC simulations cover a smaller energy range than their MD counterparts. This is due to the fact that the MD energy function contains the extra vibration term (eq 2) adding to the energy range for MD simulations. It is evident from the plots that REMUCA MC and REMUCAREM MC are flatter {constant ln[$P(E)$]} than REMUCA MD and REMUCAREM MD. This discrepancy probably arises from the fact that the MD versions of multicanonical simulations utilize the derivative of the multicanonical energy function (eq 17), whereas the MC simulations use only the multicanonical energy function itself (eq 15; Figure 1). As mentioned in the Methods section, the derivatives are fitted using cubic splines, which can cause problems if the entropy function is not smooth (the derivative will be rough, which will cause numerical instabilities in the integration of eq 17).

By comparing the plots for REM MC and REMUCA MC,

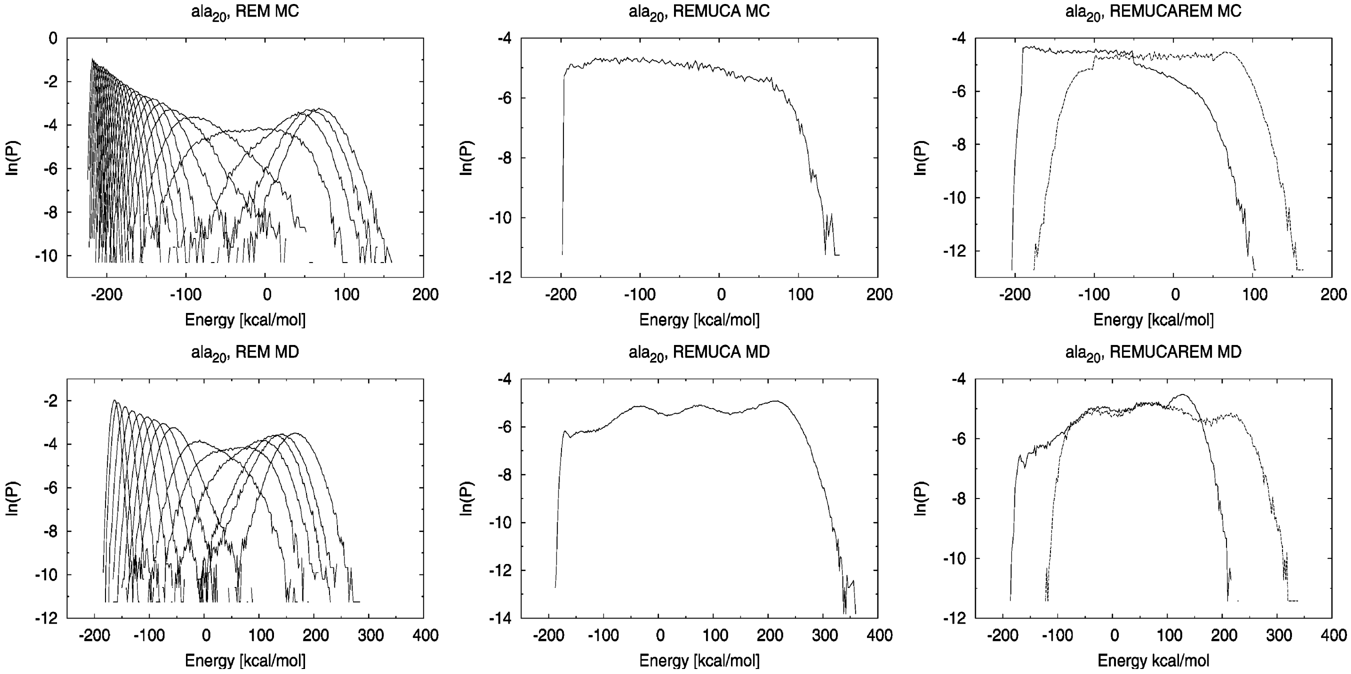

Figure 2. Histogram curves for simulations with alanine. The plots depict the logarithm of the probabilities ln[P(E)] as a function of energy (E). The top-row plots are from MC simulations (REM, REMUCA, and REMUCAREM, from left to right respectively). The bottom-row plots are from MD simulations. For REM and REMUCAREM (left and right columns, respectively), each curve corresponds to an individual replica at a different temperature (for REM) or different energy range (for REMUCAREM); see Table 1 for the number of such replicas.

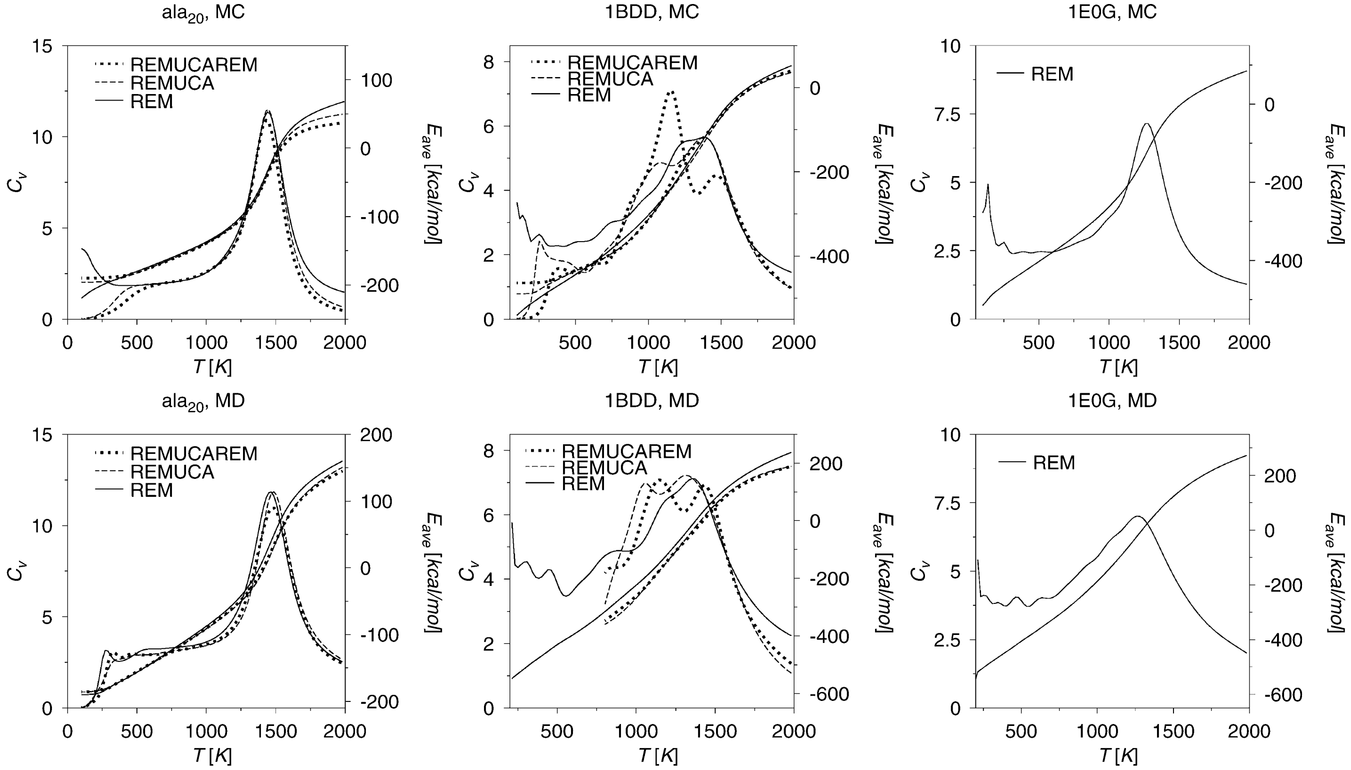

Figure 3. Heat capacity as well as average energy as a function of temperature for REM (solid line), REMUCA (dashed line), and REMUCAREM (dotted line) simulations with MC (top row) and MD (bottom row). The columns correspond to ala$_{20}$, 1BDD, and 1E0G, from left to right, respectively. The heat capacity curves are the ones with peaks at the folding temperatures. Good agreement for all three simulations for both MC and MD versions can be observed for ala$_{20}$; some overlap is observed for 1BDD, and only REM results (see text) are shown for 1E0G.

it can be seen that REMUCA MC does not cover the entire low-energy region but rather stops before −200 kcal/mol. This is because we shifted the low-energy boundary for multicanonical sampling up from the canonical average

Table 2. Parameters Used in 1BDD and 1E0G Simulations[a]

		MD			MC		
protein	simulation	replicas	temp	steps	replicas	temp	sweeps
1BDD	REM	30(×4)[b]	200−1800	240,000,000	50	50−1800	10,000,000
	REMUCA	1	50	30,000,000	1		1,000,000
	REMUCAREM	8	50−400	240,000,000	2		10,000,000
1E0G	REM	30(×4)[b]	200−1800	240,000,000	50	50−2000	10,000,000

[a] The replicas column shows the number of replicas used for each simulation. The temp shows the reference temperature (K) or range of temperatures for simulations (for REMUCA MC and REMUCAREM MC, the reference temperature cancels out in the equations; therefore, the corresponding fields are empty; this is because REMUCA and REMUCAREM depend only on the input weights which are independent of *T*, whereas in REM, the replicas differ from one another in temperature, and therefore, temperature does not cancel out). The step is the number of UNRES MD time steps, where the maximum time step was set to 4.9 fs in all MD simulations. A sweep is defined as 192 and 184 energy evaluations (four angles for each residue in the chain) for 1BDD and 1E0G, respectively. [b] Multiplexed replicas. 30(×4) means that four replicas for each temperature (with 30 temperatures) were simulated.

evaluated by the lowest temperature replica. The reason for doing this is that, when the boundary was lower in energy, the MC multicanonical simulations would walk in the entire energy range until they encountered the low-energy region, at which point the simulations would become trapped in deep local minima out of which they did not escape for the remainder of the simulation (data not shown). This issue was easily resolved for ala$_{20}$ MC simulations by simply raising the low-energy boundary, but the issue reappears during both MC and MD simulations with 1BDD and 1E0G and is discussed further when describing the results for 1BDD and 1E0G.

Figure 3 also shows two rows of plots, one for MC and one for MD simulations. The first column corresponds to simulations with poly-L-alanine. Each plot consists of two graphs; one is the heat capacity, and the other is the average energy as a function of temperature. Each graph contains three curves, individually corresponding to REM, REMUCA, and REMUCAREM simulations. The average energy was calculated with eq 7, and the heat capacity was evaluated according to the following formula:

$$C_V = \beta^2 \frac{\langle E^2 \rangle_T - \langle E \rangle_T{}^2}{N} \tag{22}$$

For both MC and MD simulations with ala$_{20}$, all the curves overlap, suggesting that the simulations converged to the same distribution. The main peak of the specific heat curve indicates the temperature of the peptide collapse. For a simple system such as ala$_{20}$, the collapse occurs simultaneously with folding to the native α-helical state. This temperature appears to be 1400 K for MC and 1500 K for MD. It is important to note that the UNRES temperature has no relevance to the experimental temperature because UNRES is a coarse-grained potential in which the nonessential degrees of freedom have been averaged out, and energy parameter optimization was carried out with a hierarchical procedure[56] to provide the steepest decrease of energy with increasing native likeness[62] while ignoring the correspondence between the simulated and experimental thermodynamic characteristics of folding. Moreover, the decoy sets were generated using the CSA method which walks only in the space of local minima, thus violating the detailed balance condition. As mentioned further in the Conclusions section, we are currently revising our hierarchical force field optimization procedure,[62] to introduce entropy using methods applied in the present work and, consequently, to capture as much physics as possible.

4.2. 1BDD. We repeated the same procedure for 1BDD as for ala$_{20}$. The parameters used for the simulations with 1BDD are described in Table 2. Similarly, as for ala$_{20}$, the results for 1BDD are shown in Figure 4. First, since 1BDD has more degrees of freedom than ala$_{20}$, we used a larger number of replicas in both REM MC and REM MD algorithms, and in REM MD, we additionally multiplexed each replica to have more trajectories from which to sample. Although it might appear that, by using more replicas, REM would perform much better than both REMUCA and REMUCAREM, the advantage of REMUCAREM (as mentioned in Section 2.5) is that a smaller number of replicas is required to cover the entire energy range. To provide a fair comparison, we used the same number of steps for both REM and REMUCAREM (see Table 2); although many more steps were used in REMUCAREM than in REMUCA, the results with REMUCAREM are not substantially improved over those with REMUCA, as discussed later in this section. As for poly-L-alanine, the density of states from the replica exchange simulations was insufficient to carry out a random walk with REMUCA and REMUCAREM; therefore, the densities of states were reweighted. The multicanonical histogram curves in Figure 4 correspond to one iteration of reweighting. Additionally, we encountered a trapping problem in the low-energy region for both MC and MD simulations. As for ala$_{20}$, we increased the low multicanonical energy boundary to escape the trapping regions (Figure 4 shows that REMUCA and REMUCAREM MC and MD do not sample all the way to the lowest energy, i.e., not beyond −500 kcal/mol). To verify whether moving the multicanonical energy boundary is acceptable, we show the RMSD results in Figure 5. The left column shows the energy versus RMSD profile for replica exchange simulations. As can be seen from this column, both REM MC and REM MD cover a wide conformational space, which includes the native structure (centered ~4.5 Å for REM MC and ~4.0 Å for REM MD). The middle and the right columns show an RMSD trajectory for REMUCA and REMUCAREM simulations, respectively. It can be seen that the system folds and unfolds several times over the course of the run, that is, attains the low-RMSD region. Even though the multicanonical simulation should perform a random walk in the energy space, it is more important that the simulation fully samples

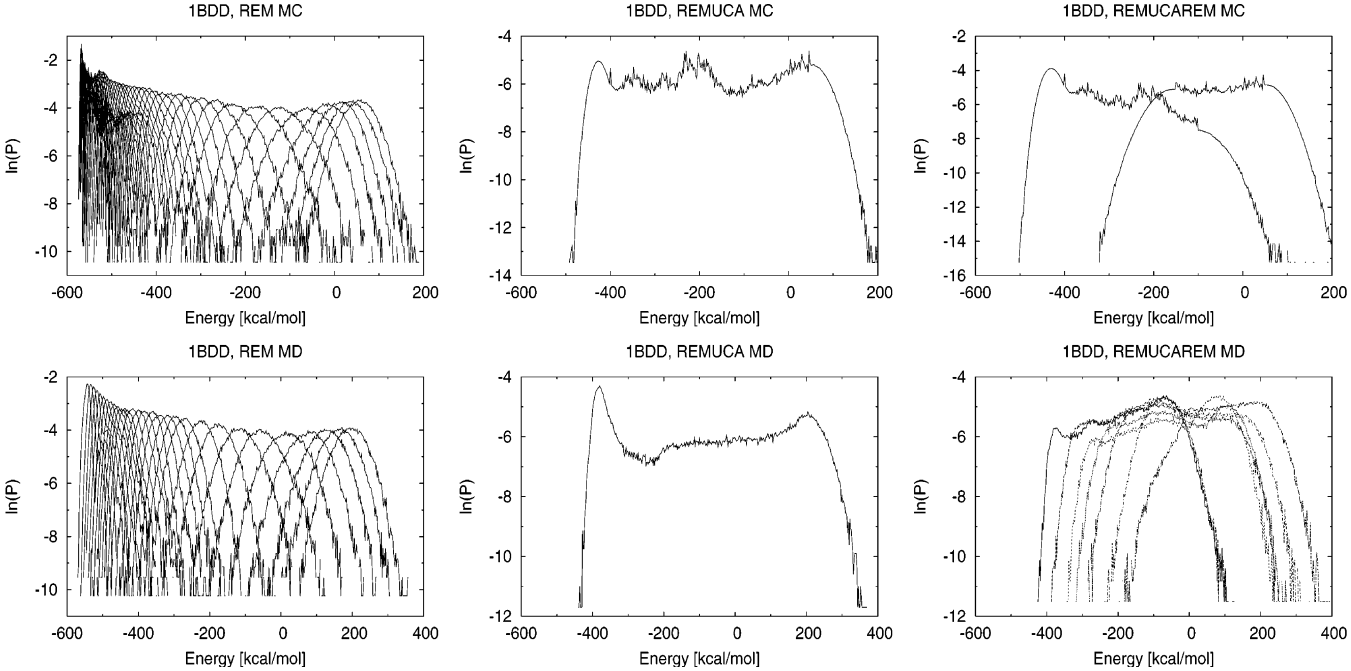

Figure 4. Histogram curves for simulations with 1BDD. The plots depict the logarithm of the probabilities as a function of energy. The top-row plots are from MC simulations (REM, REMUCA, REMUCAREM, from left to right respectively). The bottom-row plots are from MD simulations. For REM, and REMUCAREM (left, and right columns) each curve corresponds to an individual replica at a different temperature (for REM) or different energy range (for REMUCAREM); see Table 2 for the number of such replicas.

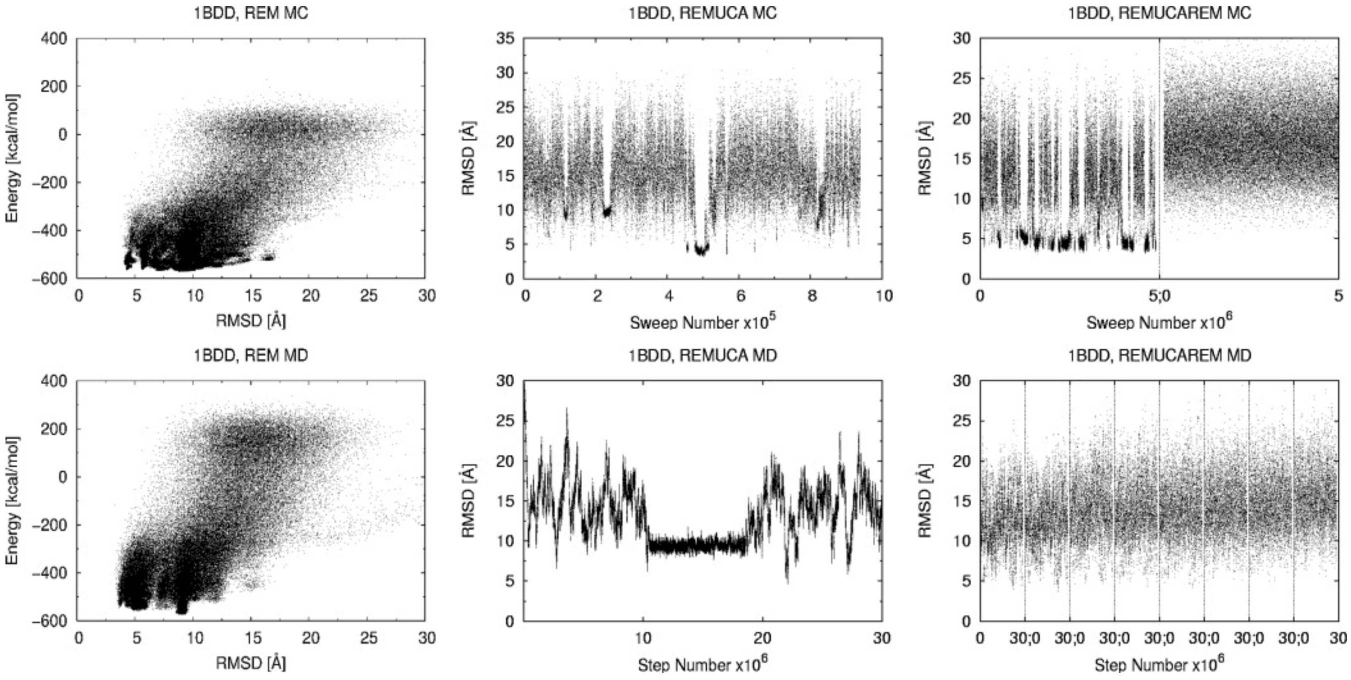

Figure 5. Simulation results for 1BDD. The top-row plots are from MC simulations (REM, REMUCA, and REMUCAREM, from left to right, respectively). The bottom-row plots are from MD simulations. The left column shows energy versus RMSD coverage of the energy space. The middle column shows the random walk of the REMUCA simulations, and the right column shows the random walk for all REMUCAREM replicas (one after another).

the conformational space, which can be observed in both the REMUCA and REMUCAREM RMSD trajectories.

The middle column of Figure 3 shows the calculated heat capacities and average energies for both MC and MD REM simulations with 1BDD. By contrast to the simulations with poly-L-alanine, 1BDD heat capacities have broad irregular peaks. The irregular peak is an overlap of two peaks, one corresponding to a collapse to a more compact state but without the final folding and one corresponding to a transition to the native state, as will be shown later. For 1BDD, REM, REMUCA, and REMUCAREM, peaks do not coincide as they do for poly-L-alanine. The fact that all simulations differ

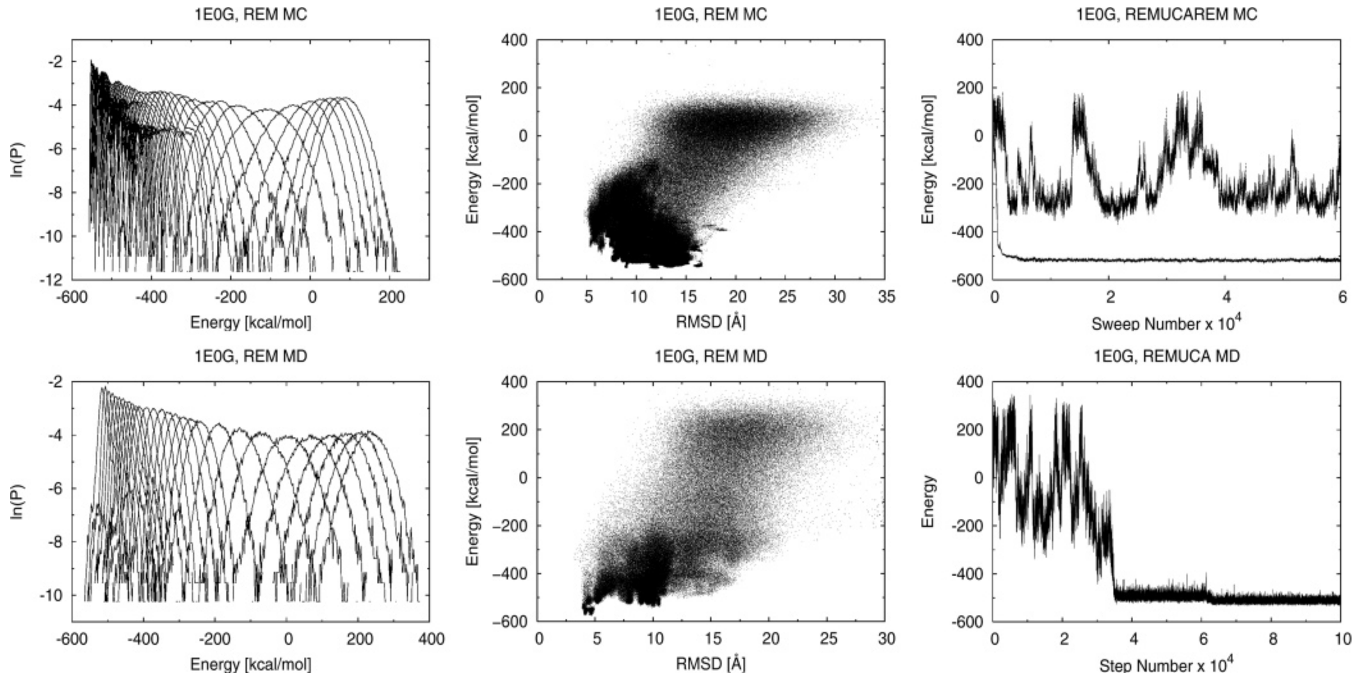

Figure 6. Simulation results for 1E0G. The top-row plots are from MC simulations, whereas the bottom-row plots are from MD simulations. The left-column shows the histogram curves for REM. Each curve corresponds to an individual replica at a different temperature. The middle column shows energy versus RMSD coverage of the energy space. The right-column shows energies at a series of steps of REMUCAREM for MC (top, with two replicas) and REMUCA for MD (bottom).

in the shape of their heat capacity curve suggests that all simulations have not converged to the same distribution. The reason the REMUCA and REMUCAREM curves do not cover the whole temperature range is that the multicanonical region was restricted to avoid trapping (i.e., the low multicanonical energy boundary was increased).

4.3. 1E0G. Finally, for 1E0G, replica exchange successfully sampled the energy space and produced reasonable statistics for thermodynamic quantities (Figure 6). The left column of Figure 6 shows the histograms for replica exchange simulations with both MC (top) and MD (bottom). The middle column depicts plots of energy as a function of the RMSD from the experimental structure, showing that the simulations cover an extended portion of the energy space. It can be seen that the REM MD simulation reaches the native state within an RMSD of around 4.5 Å and has low energy, whereas the REM MC simulation barely touches 5 Å RMSD, without reaching the low-energy region, which suggests incomplete N- and C-terminal β-strand contacts (correct β-strand packing provides a large contribution to decreasing the energy of the native structure and is necessary for the RMSD to be below 5 Å).

For multicanonical simulations (REMUCA and REMUCAREM), we were unable to obtain proper multicanonical weights, which would enable the system to carry out a random walk in the energy space. Even after several iterations of reweighting, the system would walk toward the low energy states, where it would stay for the remainder of the simulation. This behavior is shown in the right column of Figure 6, where a REMUCAREM simulation is shown for MC and a REMUCA simulation for MD. For REMUCAREM MC, it is evident that the lower-energy replica (replica 1) reaches low energies and remains trapped in a low-energy region, whereas the high-energy replica (replica 2) carries out a random walk. A similar behavior is observed for MD simulations (trapping of REMUCA MD is shown in Figure 6). This observation is similar to that from a study carried out by Bhattacharya and Sethna, who showed that, in the case of glassy systems, even multicanonical simulations have problems carrying out a random walk and instead become trapped in metastable states.[83] They implemented the entropy sampling version of the algorithm with Lennard-Jones glasses and observed that simulations that have dynamic updating of the microcanonical entropy function perform a random walk in the energy space, while the simulations with fixed weights (precomputed by iterative procedures) became trapped in metastable states. The dynamic updating of the weights (i.e., eq 5 of ref 36) is essentially a single histogram reweighting on the fly with the difference that not all regions might be visited, and typically the time between updates is much shorter. Dynamic updating ensures that the system does not remain in the same conformation for a long time. However, it also introduces discontinuities, and negative gradients into the E_{mu} function, which poses problems for the MD version of the REMUCA algorithm, with MD being more sensitive to the input weights because of its use of derivatives. The dynamic updating procedure pushes the system out of trapped states, but this violates the detailed balance condition and, thus, no longer guarantees convergence to the proper distribution or correct estimates of thermodynamic quantities. Because of the trapping problem, we did not calculate average energies and heat capacities from both REMUCA and REMUCAREM simulations for 1E0G (see Figure 3).

The third column of Figure 3 shows the calculated heat capacities and average energies for both MC and MD REM simulations with 1E0G. A sharp single peak for the heat capacity is observed for REM MC, whereas a broader peak is observed for REM MD simulations, and in both cases, it is centered at around 1270 K. As mentioned above (energy vs RMSD plot in Figure 6), the REM MC simulation does not quite sample the native region. This observation, and the fact that the heat capacity for REM MC has a sharper peak, suggests that REM MC predicts a collapse to a more compact state but without the final folding (i.e., there is no low-energy structure below a 5 Å RMSD as shown in the energy vs RMSD plot in Figure 6). On the other hand, the statistics from REM MD contain the native region (shown in the energy vs RMSD plot in Figure 6) and, thus, incorporates the contribution of the native region to the thermodynamic quantities. The collapse to a more compact structure and final folding do not seem to coincide (see the upcoming discussion about Figure 7), which broadens the heat capacity curve. For MC, the sharp peak is centered at 1270 K (Figure 3), which corresponds roughly to −130 kcal/mol of average energy. From the energy versus RMSD plot in Figure 6, it can be seen that the highest allowed energy for the collapsed structure (RMSD ∼ 5 Å) is also around −130 kcal/mol. Folding to the native state for MD occurs at lower energies, which broadens its heat capacity peak (see the discussion about Figure 7 in Section 4.4).

4.4. Free-Energy Diagrams. From our tests on ala$_{20}$, 1BDD, and 1E0G, we conclude that replica exchange molecular dynamics is the most efficient method for sampling and calculating thermodynamic quantities with a rugged energy landscape such as the 4P force field, applied to larger systems. Since the free energy is the most important quantity for the description of equilibrium properties of proteins, we used REM MD to calculate free-energy profiles for ala$_{20}$, 1BDD, and 1E0G. For this purpose, we used the densities of states obtained from the multihistogram analysis (eq 5). From the densities of states, we calculated the microcanonical entropy, $S(E_i) = k_B \ln[n(E_i)]$, for all conformations collected from the simulations and used it to compute the microcanonical free energies with the following expression: $F(E_i,T) = E_i - TS(E_i)$. To plot the restricted canonical free energy as a function of the RMSD (r) and radius of gyration (ρ), we calculated the restricted canonical free energy by evaluating the following expression for each grid point:

$$F(r,\rho,T) = -k_B T \ln \sum_{E_i \in N(r,\rho)} \exp\left(\frac{-F(E_i,T)}{k_B T}\right) \qquad (23)$$

where the index i enumerates conformations within the histogram bins, $N(r,\rho)$, for given ranges of the RMSD and radius of gyration.

Figure 7 shows the restricted canonical free-energy plots as a function of the RMSD and radius of gyration for various temperatures. Each column corresponds to simulations with ala$_{20}$, 1BDD, and 1E0G, from left to right, respectively. The temperatures are chosen so that the highest temperature is higher than that of the heat capacity peak (first row), within the peak (second row), below the peak (third row), and at zero K (fourth row) from top to bottom, respectively.

The highest-temperature free-energy plot for ala$_{20}$ shows that, at this temperature, the peptide is preferentially completely unfolded, as indicated by the high RMSD (greater than 5 Å) and the high radius of gyration (greater than 9 Å), whereas at the heat capacity peak temperature (1460 K), the lowest free-energy region connects both the native and the non-native basins (RMSD between 2 and 5 Å). For 1000 K, the free-energy surface already appears very similar to the free-energy surface at 0 K, which represents the potential energy surface. The native state (RMSD lower than 2 Å) is the lowest free energy at this temperature, confirming our observation from the heat capacity curve. It should be noted that the range of energies observed in the potential energy plot is much larger than the range observed with nonzero temperatures, showing that the search for the native state is very much facilitated in the restricted canonical free-energy surface. In other words, the restricted canonical free-energy differences do not need to be very large in order to pass from the unfolded to the folded state, whereas large potential energy barriers must be crossed to pass from the unfolded to the folded state in the potential energy surface. For ala$_{20}$, we conclude that, even though the force field was optimized without any thermodynamics, we still observe a correct folding behavior.

For protein A (1BDD), the restricted canonical free-energy plots look similar to the plots for ala$_{20}$. At high temperatures, unpacked, open structures with a high RMSD and radius of gyration are observed. At 1000 K, the low free-energy region connects unfolded non-native states with compact states (both native and non-native). At a much lower temperature (600 K), the lowest free-energy regions belong to the native basin (centered around 5 Å RMSD) and to the mirror image (centered around 9 Å RMSD). It should be noted that, for ala$_{20}$, the native region had the lowest free energy at 1000 K whereas, for 1BDD, the temperature had to be lowered to 600 K for this to occur. Finally, the potential energy plot is again similar to the low-temperature free-energy plot but has a much larger energy range. It should be noted that, at 600 K, the free energy has well-defined regions of low free energy whereas, for the potential energy, the native state is more evenly connected with compact but non-native states, which has been observed previously in MD studies with protein A in our laboratory (all 10 simulations successfully folded protein A with the 4P force field at 800 K).[18]

For 1E0G, the high-temperature plot again shows a preference for unfolded structures. For 1000 K, the compact structures are not quite preferential in free energy. From previous MD work with 1E0G in our laboratory,[18] it was found that the successful folding trajectory starts with the formation of noninteracting helical structures, which then collapse to a native HTH motif (15 Å RMSD) and finally to one with a 3.9 Å RMSD from that of the experimental structure. The HTH motif structures appear to be preferable in terms of free energy at 1000 K, which is still within the broad peak of the heat capacity for 1E0G. For low temperatures, such as 600 K, the low free-energy region connects the HTH motif to compact nativelike structures without

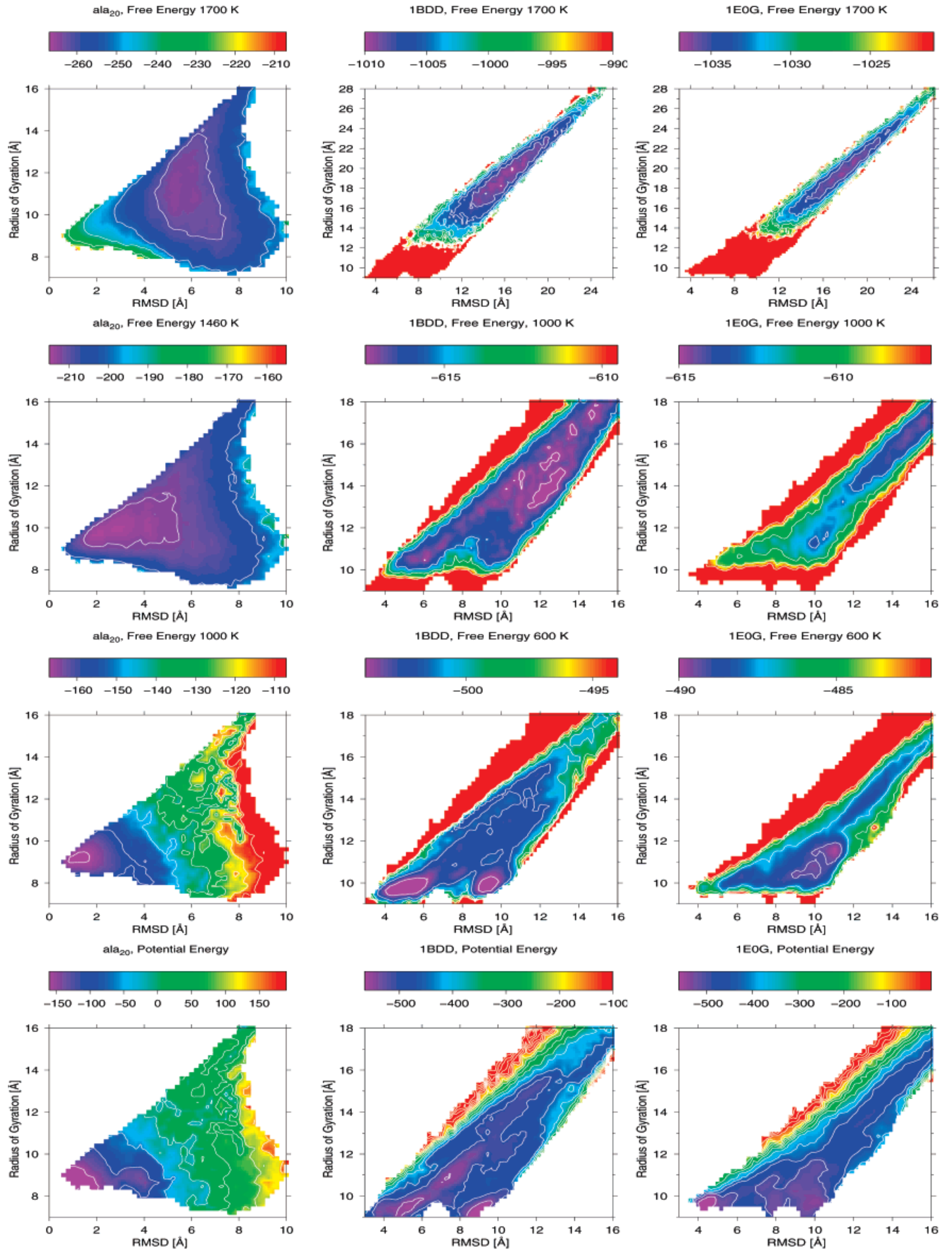

Figure 7. Restricted canonical free energy (in kcal/mol, indicated by the colored bars at the top of each graph) as a function of RMSD and radius of gyration for various temperatures. The free-energy surfaces were calculated from the REM MD simulations (see text). The columns correspond to simulations with ala$_{20}$, 1BDD, and 1E0G, from left to right, respectively. The temperatures are chosen so that the highest temperature is higher than that of the heat capacity peak (first row), within the peak (second row), below the peak (third row), and at 0 K (fourth row) for comparison.

β-strand contacts (around 6 Å RMSD). However, the fully formed native structure (centered at 4.5 Å RMSD) is at higher free energy, and it appears at the lowest free-energy region only at very low temperatures (where the free-energy plot is similar to the potential energy plot). Liwo et al. observed that only 6 out of 10 canonical MD simulations at 800 K yielded nativelike structures.[18] Our free-energy calculations show that the lowest free energy corresponds to non-native compact structures (i.e., with a low radius of gyration but a high RMSD); however, the native structures (with an RMSD less than 5 Å) have slightly higher free energy. Therefore, the non-native conformations are more probable, but the native structures still have a finite probability to occur. Thus, our free-energy calculations agree with the results obtained by Liwo et al.

Since the temperature must be extremely low in order for the native state to be the global minimum of the free energy, the entropy contribution is much larger than that for the same temperature in protein A and ala$_{20}$. A larger contribution from entropy means more accessible conformations for a given temperature. Therefore, the multicanonical simulations have to sample a larger number of accessible conformations, which becomes difficult for 1E0G.

From Figure 7, it can be seen that, for a simple system such as ala$_{20}$, the collapse occurs simultaneously (at 1460 K) with folding to the native α-helical state (RMSD values and radii of gyration for low free-energy regions decrease simultaneously with temperature from 1700 to 1460 to 1000 K). For protein A and 1E0G, the low free-energy region at 1000 K extends all the way to the low radius of gyration and high RMSD values. For protein A, two low free-energy regions remain as the temperature is decreased to 600 K, one being the native and one being the mirror image. For 1E0G, the low free-energy region at 600 K with a low radius of gyration but a high RMSD appears first, and as the temperature is lowered (not shown here), the native region becomes the lowest free-energy basin. However, this occurs at very low temperatures, as described above. This explains why the heat capacity peaks for both protein A and 1E0G are broad and irregular. The two main events, collapse and folding to the native state, occur at different temperatures.

5. Conclusions

In the present work, we implemented REM, REMUCA, and REMUCAREM algorithms with the UNRES force field, utilizing Monte Carlo and molecular dynamics techniques. First, we tested all the algorithms on a simple poly-L-alanine system. For both the MC and MD algorithms, we obtained good agreement for heat capacity and average energy curves, which shows that all the simulations converged to the same distribution and that our implementation works as expected.

Next, we applied the simulations to two proteins, namely, to 1BDD and 1E0G. First, the 1BDD simulations performed reasonably well. The best performance was observed for the replica exchange algorithm in both the MC and MD simulations, since REM appeared to be much less sensitive to the input parameters (the only important parameter is the distribution of temperatures). To carry out a random walk, REMUCA and REMUCAREM depend on a proper estimation of the input weights and, as for ala$_{20}$, both REMUCA and REMUCAREM simulations had to be reweighted in order to obtain reasonably flat histograms. A trapping problem occurred at low energies, which was alleviated by raising the lower energy boundary for multicanonical simulations. However, by excluding a certain energy region from being sampled, the agreement among the heat capacity curves for all simulations was not so good.

Since 1E0G has a more complicated fold than 1BDD, multicanonical simulations broke down, and only replica exchange simulations were capable of exploring the energy region and computing the thermodynamic averages. This observation agrees with that from the study by Aleksenko et al.,[84] who concluded that the generalized ensemble approach is a useful study tool for proteins up to 30−40 residues with simple topology such as the α-helix. Furthermore, since the MD version of REMUCA and REMUCAREM use the derivative of the entropy function, MD multicanonical simulations are even more sensitive than their MC counterparts; therefore, they are more difficult to implement. Conversely, MD is much more capable of exploring the energy landscape than MC; hence, MD simulations are much more useful for larger systems.

Finally, we analyzed data from our REM MD simulations for all three test systems and calculated free-energy maps as a function of the RMSD and radius of gyration. The free-energy calculations show the correct folding behavior for poly-L-alanine and protein A, while for 1E0G, the native structure had the lowest free energy only at very low temperatures; hence, the entropy contribution is much larger than that for the same temperature in protein A and ala$_{20}$. The larger contribution from entropy means more accessible conformations for a given temperature. For the same temperature, ala$_{20}$ has the smallest entropy contribution, followed by protein A, and then by 1E0G.

Although both REMUCA and REMUCAREM seem to have potential as sampling methods applied to smaller systems, replica exchange utilizing MD, coupled with multiplexing, appears to offer more insight into the behavior of protein folding for more complicated systems with a rough energy landscape. Moreover, since replica exchange is easy to implement and has few parameters to adjust, it is very suitable for implementation in the future revision of our hierarchical optimization procedure,[62] which is currently under development in our laboratory. The new optimization procedure is based on a hierarchical design of the potential-energy landscape such that the energy decrease follows the increase of native likeness[56] and utilizes MD as a sampling method to capture as much physics as possible. Preliminary tests (unpublished data) show that replica exchange together with umbrella sampling[85] (introduced when the native region is not sufficiently covered with the initial parameter set) covers a broader region of conformational space and, thus, produces better statistics for hierarchical optimization. Consequently, this will allow us to produce a coarse-grained force field suitable for molecular dynamics simulations, which will be capable of a more accurate evaluation of thermodynamic quantities.

Acknowledgment. We thank A. Liwo for valuable comments. This work was supported by grants from the National Institutes of Health (GM-14312), the National Science Foundation (MCB00-03722), the NIH Fogarty International Center (TW7913), and Grant 3 T09A 032 26 from the Polish Ministry of Scientific Research and Information Technology. This research was carried out by using the resources of (a) our 674-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University; (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center; and (c) the National Center for Supercomputing

Applications System at the University of Illinois at Urbana–Champaign.

References

(1) Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341.

(2) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(3) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.

(4) Li, Z.; Scheraga, H. A. *THEOCHEM* **1988**, *179*, 333–352.

(5) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283–1303.

(6) Ripoll, D. R.; Scheraga, H. A. *J. Protein Chem.* **1989**, *8*, 263–287.

(7) Pillardy, J.; Czaplewski, C.; Wedemeyer, W. J.; Scheraga, H. A. *Helv. Chim. Acta* **2000**, *83*, 2214–2230.

(8) Nanias, M.; Chinchio, M.; Oldziej, S.; Czaplewski, C.; Scheraga, H. *J. Comput. Chem.* **2005**, *26*, 1472–1486.

(9) Piela, L.; Kostrowicki, J.; Scheraga, H. A. *J. Phys. Chem.* **1989**, *93*, 3339–3346.

(10) Pillardy, J.; Olszewski, K. A.; Piela, L. *J. Phys. Chem.* **1992**, *96*, 4337–4341.

(11) Pillardy, J.; Liwo, A.; Groth, M.; Scheraga, H. A. *J. Phys. Chem. B* **1999**, *103*, 7353–7366.

(12) Pillardy, J.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **1999**, *103*, 9370–9377.

(13) Lee, J.; Scheraga, H. A.; Rackovsky, S. *J. Comput. Chem.* **1997**, *18*, 1222–1232.

(14) Lee, J.; Scheraga, H. A. *Int. J. Quantum Chem.* **1999**, *75*, 255–265.

(15) Czaplewski, C.; Liwo, A.; Pillardy, J.; Oldziej, S.; Scheraga, H. A. *Polymer* **2004**, *45*, 677–686.

(16) Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13785–13797.

(17) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13798–13810.

(18) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367.

(19) Khalili, M.; Liwo, A.; Scheraga, H. A. *J. Mol. Biol.* **2006**, *355*, 536–547.

(20) Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(21) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(22) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(23) Geyer, C. *Computing Science and Statistics Proceedings of the 23rd Symposium on the Interface*; American Statistical Association: New York, 1991.

(24) Gront, D.; Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **2001**, *115*, 1569–1574.

(25) Kolinski, A.; Gront, D.; Pokarowski, P.; Skolnick, J. *Biopolymers* **2003**, *69*, 399–405.

(26) Fenwick, M. K.; Escobedo, F. A. *J. Chem. Phys.* **2003**, *119*, 11998–12010.

(27) Romiszowski, P.; Sikorski, A. *Physica A* **2004**, *336*, 187–195.

(28) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(29) Zhou, R.; Berne, B.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931–14936.

(30) Sanbonmatsu, K.; Garcia, A. *Proteins* **2002**, *46*, 225–234.

(31) Garcia, A.; Onuchic, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.

(32) Lin, C.-Y.; Hu, C.-K.; Hansmann, U. H. E. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 436–445.

(33) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.

(34) Berg, B. A.; Neuhaus, T. *Phys. Rev. Lett.* **1992**, *68*, 9–12.

(35) Lee, J. *Phys. Rev. Lett.* **1993**, *71*, 211–214.

(36) Hao, M.; Scheraga, H. A. *J. Phys. Chem.* **1994**, *98*, 4940–4948.

(37) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.

(38) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.

(39) Gront, D.; Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **2000**, *113*, 5065–5071.

(40) Hansmann, U. *Phys. Rev. E* **1997**, *56*, 6200–6203.

(41) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664–6675.

(42) Mitsutake, A.; Okamoto, Y. *J. Chem. Phys.* **2004**, *121*, 2491–2504.

(43) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(44) Kwak, W.; Hansmann, U. *Phys. Rev. Lett.* **2005**, *95*, 138102.

(45) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

(46) Jang, S.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2002**, *91*, 058305.

(47) Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, J.; Khalili, M.; Arnautova, Y.; Jagielska, A.; Schafroth, H. D.; Kazmierkiewicz, R.; Ripoll, D.; Pillardy, J.; Saunders, J.; Kang, Y.; Gibson, K.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7547–7552.

(48) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1697–1714.

(49) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.

(50) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 849–873.

(51) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Ołdziej, S.; Scheraga, H. A. *J. Comput. Chem.* **1997**, *18*, 874–887.

(52) Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R. J.; Rackovsky, S.; Pincus, M. R.; Scheraga, H. A. *J. Comput. Chem.* **1998**, *19*, 259–276.

(53) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys.* **2001**, *115*, 2323−2347.

(54) Lee, J.; Ripoll, D. R.; Czaplewski, C.; Pillardy, J.; Wedemeyer, W. J.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7291−7298.

(55) Pillardy, J.; Czaplewski, C.; Liwo, A.; Wedemeyer, W. J.; Lee, J.; Ripoll, D. R.; Arłukowicz, P.; Ołdziej, S.; Arnautova, Y. A.; Scheraga, H. A. *J. Phys. Chem. B* **2001**, *105*, 7299−7311.

(56) Liwo, A.; Arłukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1937−1942.

(57) Ołdziej, S.; Kozłowska, U.; Liwo, A.; Scheraga, H. A. *J. Phys. Chem. A* **2003**, *107*, 8035−8046.

(58) Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kozłowska, U.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 9421−9438.

(59) Nishikawa, K.; Momany, F. A.; Scheraga, H. A. *Macromolecules* **1974**, *7*, 797−806.

(60) Kubo, R. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100−1120.

(61) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412−9426.

(62) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16934−16949.

(63) Ferrenberg, A.; Swendsen, R. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(64) Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(65) Hansmann, U.; Okamoto, Y.; Eisenmenger, F. *Chem. Phys. Lett.* **1996**, *259*, 321−330.

(66) Nakajima, N.; Nakamura, H.; Kidera, A. *J. Phys. Chem. B* **1997**, *101*, 817−824.

(67) Bartels, C.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 865−880.

(68) Berg, B. *Int. J. Mod. Phys. C* **1992**, *3*, 1083−1098.

(69) Hansmann, U.; Okamoto, Y. *J. Phys. Soc. Jpn.* **1994**, *63*, 3945−3949.

(70) Hansmann, U.; Okamoto, Y. *Physica A* **1994**, *212*, 415−437.

(71) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *329*, 261−270.

(72) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6664−6675.

(73) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *J. Chem. Phys.* **2003**, *118*, 6676−6688.

(74) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665−9672.

(75) Bateman, A.; Bycroft, M. *J. Mol. Biol.* **2000**, *299*, 1113−1119.

(76) Ołdziej, S.; Łagiewka, J.; Liwo, A.; Czaplewski, C.; Chinchio, M.; Nanias, M.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16950−16959.

(77) Macias, M. J.; Gervais, V.; Civera, C.; Oschkinat, H. *Nat. Struct. Biol.* **2000**, *7*, 375−379.

(78) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906−918.

(79) Johansson, M. U.; de Chateau, M.; Wikstrom, M.; Forsen, S.; Drankenberg, T.; Bjorck, L. *J. Mol. Biol.* **1997**, *266*, 859−865.

(80) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(81) Swope, W.; Andersen, H.; Berens, P.; Wilson, K. *J. Chem. Phys.* **1982**, *76*, 637−649.

(82) Rhee, Y.; Pande, V. *Biophys. J.* **2003**, *84*, 775−786.

(83) Bhattacharya, K.; Sethna, J. *Phys. Rev. E* **1998**, *57*, 2553−2562.

(84) Aleksenko, V.; Kwak, W.; Hansmann, U. *Physica A* **2005**, *350*, 28−37.

(85) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 2002.

CT050253O