

# Microcanonical replica exchange molecular dynamics simulation of proteins

Parimal Kar,<sup>1,\*</sup> Walter Nadler,<sup>2,†</sup> and Ulrich H. E. Hansmann<sup>1,‡</sup>

<sup>1</sup>*Department of Physics, Michigan Technological University, Houghton, Michigan 49931, USA*

<sup>2</sup>*Institute for Advanced Simulation (IAS), Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich, D-52425 Jülich, Germany*

(Received 17 September 2009; published 18 November 2009)

We present microcanonical replica exchange molecular dynamics simulations as an alternative to canonical ones. Its advantage is the easily tunable high acceptance rate for replica exchange. We present the theory, comment on its actual implementation, and demonstrate its application for a common test case, the trp-cage protein.

DOI: [10.1103/PhysRevE.80.056703](https://doi.org/10.1103/PhysRevE.80.056703)

PACS number(s): 02.70.Ns, 87.10.Tf, 87.15.Cc

## I. INTRODUCTION

In the last years, we have seen remarkable progress in modeling the folding, aggregation, and interaction of proteins. For instance, a recent investigation of a 49-residue C-terminal fragment of the artificial protein TOP7, relying on an all-atom force field and an implicit solvent, found not only a lowest-energy configuration within 2 Å to the experimentally determined structure but also a novel folding mechanism that relies on “caching” of a N-terminal “chameleon” segment [1]. These successes are mainly due to the advances in sampling techniques. Generalized-ensemble and replica exchange techniques [2] are now routinely used to enhance the sampling of low-energy configurations and—especially in their optimized forms [3–7]—have led to much faster convergence at physiological temperatures than achieved in regular Monte Carlo or molecular dynamics simulations.

While these techniques have alleviated the sampling problem, a number of difficulties remain. Most prominent here are simulations of proteins with explicit water. This is because in replica exchange, the probability for an exchange between two temperatures decreases not only with the temperature difference  $\Delta T$  between two replicas but also with the number of degrees of freedom  $N$ . Hence, because of the large number of water molecules needed in protein simulations, the temperature intervals  $\Delta T$  have to be chosen small and, therefore, a large number  $M$  of replicas is needed to cover the range between the temperature of interest (the lowest one) and the highest temperature which should correspond to the largest relevant barrier in the system. On the other hand, the number of round trips between lowest and highest temperatures and back defines a lower bound on the number of independent configurations sampled at the lowest temperature (i.e., the one of interest). However, the number of round trips decreases as  $\sqrt{M}$  with the number of replicas  $M$ . As a consequence, protein simulations with explicit water do not only require a large number of replicas but also long simulation times for each replica in order to reach equilibrium and obtain sufficient statistics.

In a recent brief communication [5], we suggested to circumvent this problem of low acceptance rate and resulting

large number of replicas through the use of a microcanonical replica exchange method that is rejection free and therefore optimizes the flow along the temperature ladder. Molecular dynamics simulations are usually done in the canonical ensemble ( $T=\text{const}$ ) instead of a microcanonical ensemble ( $E=\text{const}$ ). One reason is that the canonical ensemble is often more closer to the experimental settings (albeit not always, constant energy surface simulations are of interest in their own right [8,9], e.g., for comparison with recent molecular-beam experiments [10]). The other reason is that integration errors can accumulate in microcanonical molecular dynamics and easily lead to numerical instabilities and uncontrolled behavior; the use of a thermostat usually washes out the effect of these errors. Our assumption is that these integration errors are also averaged out in microcanonical replica exchange molecular dynamics (MREMD) through the exchange moves and velocity reweighting. As it is possible, in principle, to connect back from a microcanonical ensemble to the canonical ensemble, the rejection-free microcanonical replica exchange molecular dynamics becomes a promising alternative in cases, such as simulations, in explicit solvent that otherwise suffer from low acceptance rates.

The purpose of the present paper is to test the suitability of this idea in a practical application. We have chosen as our system the trp-cage protein [11,12] as it has become a common model to test numerical methods [13,14]. The protein has been also previously investigated [15,16] with replica exchange methods [17–19], allowing a comparison with our results. As the present paper describes a proof-of-concept study, we simulate the molecule with an implicit solvent allowing for a faster evaluation of our approach. In the following section, we first describe our method in detail before presenting our results. We finally discuss possible applications and modifications of our approach.

## II. METHODS

### A. Statistical physics of microcanonical molecular dynamics

In microcanonical molecular dynamics, the equations of motion are solved numerically for a particular system, generating states of constant energy  $E$  for that system. Assuming ergodicity, the hypersurface of states with constant energy  $E$  is connected and all states on the constant energy hypersur-

\*pkar@mtu.edu

†wnadler@fz-juelich.de

‡hansmann@mtu.edu

face are sampled uniformly. For observables that depend only on kinetic and potential energies  $M(E_{pot}, E_{kin})$ , the microcanonical averages are given by

$$\langle M \rangle_E = \frac{1}{|\Omega_{tot}(E)|} \int dE_1 \int dE_2 \delta(E - E_1 - E_2) \times \Omega_{pot}(E_1) \Omega_{kin}(E_2) M(E_1, E_2), \quad (1)$$

with  $\Omega_{pot}(E)$  and  $\Omega_{kin}(E)$  being the respective densities of states for the potential energy and for the kinetic energy; the total state space volume of the energy shell at  $E$  is used as normalization

$$|\Omega_{tot}(E)| = \int dE_1 \int dE_2 \delta(E - E_1 - E_2) \Omega_{pot}(E_1) \Omega_{kin}(E_2). \quad (2)$$

Usually, we are interested in canonical averages, i.e.,

$$\langle M \rangle_\beta = Z^{-1} \int dE M(E) \Omega_{pot}(E) e^{-\beta E}, \quad (3)$$

with  $\beta$  as the inverse canonical temperature, and the partition function  $Z$  is used as normalization,

$$Z = \int dE \Omega_{pot}(E) e^{-\beta E}. \quad (4)$$

In order to evaluate such properties from microcanonical simulations, we need to estimate the density of states for the potential energy  $\Omega_{pot}(E)$  from them. Since the distribution of potential energies observed in a microcanonical simulation is given by

$$P(E_{pot}; E) \propto \int dE_1 \int dE_2 \delta(E - E_1 - E_2) \Omega_{pot}(E_1) \times \Omega_{kin}(E_2) \delta(E_{pot} - E_1) = \Omega_{pot}(E_{pot}) \Omega_{kin}(E - E_{pot}), \quad (5)$$

the density of potential energies has to be separated from the kinetic-energy part. This is straightforward as the kinetic energy is given by

$$E_{kin} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i}, \quad (6)$$

with  $\mathbf{p}_i$  as the momentum vector and  $m_i$  as the mass of atom or group  $i$ ; the density of states for the kinetic energy therefore can be determined analytically,

$$\Omega_{kin}(E) \propto E^{(3N-f-2)/2}, \quad (7)$$

where  $f$  counts the constraints on the system (i.e., the true number of degrees of freedom is not  $3N-2$  but reduced by  $f$ ) Hence, up to the normalization constant, the distribution of potential energies observed in a microcanonical simulation is given by

$$P(E_{pot}; E) \propto \Omega_{pot}(E_{pot}) (E - E_{pot})^{(3N-f-2)/2}. \quad (8) \quad \text{and}$$

Since both functions on the right-hand side grow strongly with their arguments,  $P(E_{pot}; E)$  is a sharply peaked function. Consequently, microcanonical averages of the energies are given by the most probable value, e.g.,

$$\langle E_{pot} \rangle_E \approx \hat{E}_{pot}, \quad (9)$$

note that  $\hat{E}_{pot} + \hat{E}_{kin} = E$  holds. A saddle-point approximation of Eq. (8) leads to the well-known relation between kinetic energy and microcanonical temperature  $[\beta_E = 1/k_B T_E \equiv d \ln \Omega_{pot}(E) / dE]$

$$\hat{E}_{kin} = \frac{M}{2} T_E, \quad (10)$$

where  $M = 3N - f - 2$  is the number of degrees of freedom in the system, and one obtains

$$P(E_{pot}; E) \propto \Omega_{pot}(E_{pot}) \exp \left\{ -\beta_E E_{pot} - \frac{M}{2} \left( \frac{E_{pot} - \hat{E}_{pot}}{\hat{E}_{kin}} \right)^2 + O \left[ \left( \frac{E_{pot} - \hat{E}_{pot}}{\hat{E}_{kin}} \right)^3 \right] \right\}. \quad (11)$$

Therefore, to leading order, the microcanonical energy distribution is given by the Boltzmann distribution, with the canonical temperature equal to the microcanonical temperature.

## B. Microcanonical replica exchange

In canonical replica exchange [17–19], two configurations with energies  $E_1$  and  $E_2$ , sitting at temperatures  $T_1$  and  $T_2$ , are exchanged with probability  $\exp(\Delta\beta\Delta E)$ , with the inverse temperature  $\beta = 1/k_B T$ . In a microcanonical replica exchange, one uses that

$$E(x, v) = E_{pot}(x) + E_{kin}(v) \quad \text{with} \quad E_{kin}(v) = \frac{1}{2} \sum_i m_i v_i^2, \quad (12)$$

where the potential energy  $E_{pot}$  depends only on the coordinates  $x$ , and the kinetic energy  $E_{kin}$  solely on the velocities  $v$ . Scaling all velocities by a factor  $r$  therefore changes the kinetic energy by

$$E_{kin}(rv) = r^2 E_{kin}(v). \quad (13)$$

Hence, assuming  $E^{(1)} < E^{(2)}$  and choosing suitable scaling parameters  $r_1$  and  $r_2$ , one can exchange the two configurations with probability one

$$\begin{aligned} E^{(1)}(x^{(1)}, v^{(1)}) &= E_{pot}(x^{(1)}) + E_{kin}(v^{(1)}) \\ \rightarrow E^{(2)}(x^{(1)}, r_1 v^{(1)}) &= E_{pot}(x^{(1)}) + E_{kin}(r_1 v^{(1)}) \\ &= E_{pot}(x^{(1)}) + r_1^2 E_{kin}(v^{(1)}), \end{aligned} \quad (14)$$

TABLE I. Replicas and their corresponding energies and temperatures.

Energy shell	Total energy (kcal/mol)	$T$ (K)
E1	-368.5	250
E2	-360.9	260
E3	-340.0	273
E4	-311.2	290
E5	-271.7	315
E6	-252.3	325
E7	-223.2	350
E8	-192.2	373
E9	-151.2	393
E10	-119.8	413
E11	-81.9	433
E12	-46.9	450
E13	-15.2	473
E14	19.2	493
E15	51.6	513
E16	91.3	533
E17	130.3	555
E18	184.3	580

$$\begin{aligned}
E^{(2)}(x^{(2)}, v^{(2)}) &= E_{pot}(x^{(2)}) + E_{kin}(v^{(2)}) \\
\rightarrow E^{(1)}(x^{(2)}, r_2 v^{(2)}) &= E_{pot}(x^{(2)}) + E_{kin}(r_2 v^{(2)}) \\
&= E_{pot}(x^{(2)}) + r_2^2 E_{kin}(v^{(2)}), \quad (15)
\end{aligned}$$

where the two rescaling factors  $r_1$  and  $r_2$  are given by

$$r_{1,2} = \sqrt{\frac{E^{(2),(1)} - E_{pot}(\mathbf{x}_{1,2})}{E^{(1),(2)} - E_{pot}(\mathbf{x}_{1,2})}}. \quad (16)$$

Such moves are possible for  $E_{pot}(\mathbf{x}_2) < E^{(1)}$ , a restriction that does not violate detailed balance. On the other hand, ergodicity is ensured because of the regular microcanonical molecular dynamics between exchange moves. The acceptance probability for an allowed move is always one since both weight functions are constant.

### C. Technical details and setting

We test the efficiency of this microcanonical replica exchange molecular dynamics in all-atom simulations of the 20-residue trp-cage miniprotein, which has become a com-

monly used test system for the evaluation of the sampling schemes. The AMBER9 package is used with the FF99SB force field, approximating the interaction between protein and surrounding solvent by the generalized Born implicit solvent. 18 replicas are used with the total energies—and corresponding temperatures—given in Table I. These values were chosen to enable us a comparison with Ref. [23]. After generating linear configurations with the module xLEAP and minimizing these with 500 steps of steepest descent followed by another 500 steps of conjugate gradient, we heat the molecule to the respective target temperatures of Table I. Here, and in the canonical replica exchange simulations with that we compare our results, we use the AMBER subroutine SHAKE and a Berendsen thermostat for temperature control (coupling constant 1.0 ps). The resulting 18 structures serve as our initial starting configurations for both microcanonical and canonical replica exchange molecular dynamics (CREMD) simulations. Each structure consists of 304 atoms; however, the number of degrees of freedom is not  $3N-2=910$  but 757 as SHAKE constraints the length of certain bonds. For each algorithm, we perform runs of 15 ns, with an exchange move attempted every 5 ps. We had written an external driver script for the replica exchange scheme. Only the last 10 ns are used for analysis.

## III. RESULTS

The inherent roughness of protein free-energy landscapes leads to slow sampling at low temperatures (or in microcanonical simulations at low energies). In order to demonstrate this sampling problem, we have performed for our test system canonical molecular dynamics runs at  $T=250$  K and microcanonical molecular dynamics at the corresponding energy ( $E_{tot}=-368.5$  kcal/mol). The two runs are over a time of 270 ns, which corresponds to the total effort in the replica exchange simulations ( $18 \times 15$  ns). In Fig. 1, we show as function of time the RMSD of the actual configuration to the experimentally determined one (Protein Data Bank Id: 1L2Y). Over the whole length of the simulation, the RMSD is around or larger than 6 Å, indicating that the simulations never thermalized and got stuck in local minima structurally very different from the native configuration.

A common approach to overcome this sampling problem is parallel tempering, also known as replica exchange sampling [17–19]. In Fig. 2, we display the resulting time series of RMSD at  $T=250$  K from a replica exchange simulation of the trp-cage protein with the temperature distribution given by Table I. As in the canonical and microcanonical

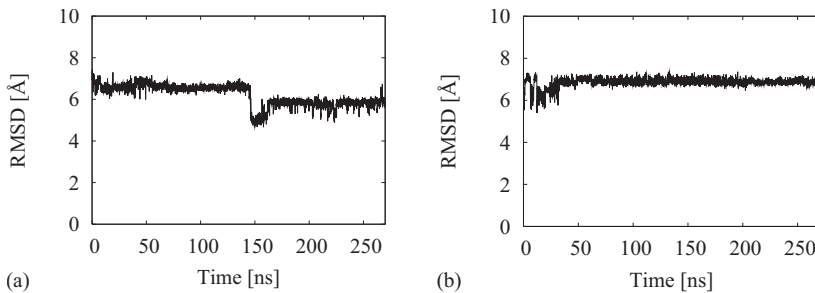


FIG. 1. Root-mean-square deviation (RMSD) to the experimentally determined structure as a function of time for (a) a canonical molecular dynamics simulation at  $T=250$  K and (b) a microcanonical molecular dynamic simulation at the corresponding energy  $E_{tot}=-368.5$  kcal/mol.

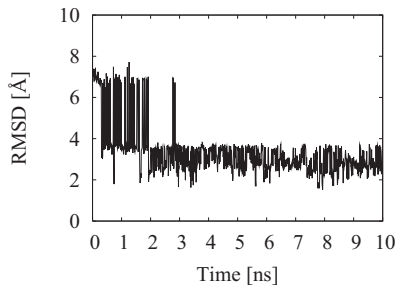


FIG. 2. RMSD to the experimentally determined structure as a function of time. The data are from a canonical replica exchange simulation with a temperature distribution given in Table I and measured at  $T=250$  K.

runs of Fig. 1, the RMSD starts at around 7 Å, indicating a starting configuration very different from the native one. However, the replica exchange sampling process leads soon to configurations that are within 3 Å RMSD and therefore similar to the experimentally determined structure. Our results are comparable to the ones obtained by Simmerling *et al.* [11], who have performed 50 ns long all-atom fully unrestrained folding simulation of this protein at 325 K in implicit GB solvent [20,21] using the AMBER FF99SB force field [22]. Without showing data, we also remark that the transition temperature of  $\approx 413$  K (see also Fig. 9) is comparable to the melting temperatures of  $\approx 400$  K found by Pitera and Swope [23]. Albeit diverging from the experimentally determined transition temperature of 315 K [24], both results show that our data are comparable with previous simulations relying on the AMBER force field and an implicit solvent.

The reason for the enhanced sampling of low-energy configurations is the excursions to high temperatures that allow a replica to escape from local minima. As an example, in Fig. 3(a), we show this walk through temperature space for one of the 18 replicas. A lower limit for the number of independent structures observed at lowest temperature  $T=250$  K is the number of round trips between this temperature and the highest temperature (in our case,  $T=580$  K) and back. In our example, only one such round trip is observed and only a total of three round trips for all replicas together. The difficulty in ensuring a sufficient number of round trips (and therefore sufficient statistics), especially for the case of protein simulations in explicit solvent, has been described in the introduction and is the starting point for our investigation. Our proposed algorithm replaces a replica exchange in temperature by an exchange of replicas between different energy levels in microcanonical molecular dynamics. As the exchange move is rejection free, it leads to much faster round

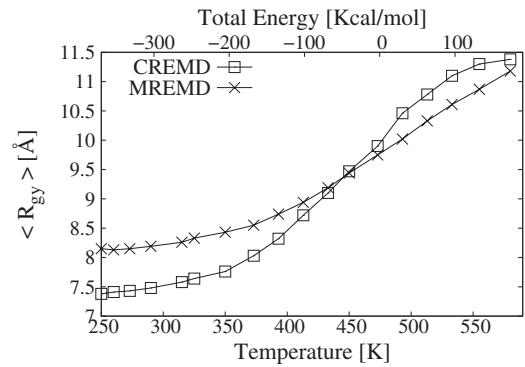


FIG. 4. Radius of gyration  $\langle r_{gy} \rangle$  as function of temperature (in a CREMD simulation) and as a function of the corresponding energy levels in a MREMD simulation with the updates proposed in Ref. [5].

trip times. This can be seen also in Fig. 3(b), where we show the walk of one replica through energy space. Note that the various energy levels correspond to the temperatures of the canonical replica exchange run and are also listed in Table I. Because of the large number of round trips, we could show here only 2 ns of the 10 ns long run, for otherwise the figure would no longer be readable.

However, while the MREMD of Ref. [5] leads to a 50-fold decrease in round trip times when compared to the CREMD, this gain in efficiency does not translate into improved sampling. This is obvious from Fig. 4, where we plot the average radius of gyration  $\langle r_{gy} \rangle$ , as a function of temperature  $T$ , when calculated from the canonical replica exchange molecular dynamics, and as a function of the corresponding total energies when calculated from the microcanonical replica exchange molecular dynamics. As a measure for the compactness of protein structures and its change, this quantity indicates structural transitions. Clearly, the two curves differ considerably. Together with similar behavior for other physical quantities (data not shown), the difference between the two curves indicates sampling problems in the approach.

The difference between the two simulations is puzzling as the microcanonical replica exchange method is formally correct and, therefore, should yield the same results as the canonical replica exchange. Hence, this difference indicates that despite the increased flow through temperature space, the sampling is still slower than in the canonical case, not faster as was expected.

A fundamental assumption behind the idea of optimizing a replica exchange simulation through maximizing the flow through temperature space is that relaxation at a given temperature is fast compared with the time scale of flow through

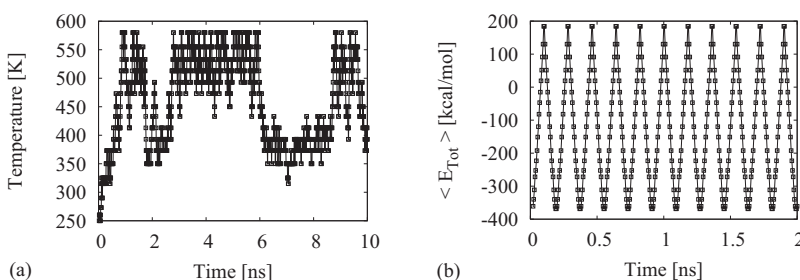


FIG. 3. Walk of a specific replica (a) through temperature in a CREMD simulation and (b) through energy in a MREMD simulation with the updates proposed in Ref. [5]. Note that the large number of roundtrips observed for the later case allowed us only to show a short segment of the 10 ns run.



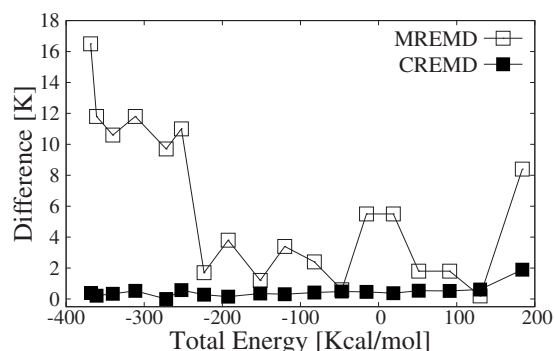


FIG. 5. Difference between canonical temperature (see Table I) and microcanonical temperature calculated from the kinetic energy via Eq. (10), as function of total energy. The figure shows this quantity as measured in CREMD simulations as well as in the MREMD simulation with updates proposed in Ref. [5].

temperatures. In the present case, this seems not to be the case. An indicator for this lack of kinetic-energy equilibration is the difference between the microcanonical temperature and the canonical temperature. We have plotted this quantity in Fig. 5, as a function of total energy, comparing data from the canonical replica exchange molecular dynamics simulation with those from our microcanonical replica exchange approach. While the temperature difference fluctuates around zero for the canonical run, it differs strongly in the case of microcanonical replica exchange.

Hence, the assumptions behind Eqs. (8) and (11) do not hold on the time scales of our simulations. The equivalence can be expected to be restored for very long simulation times, see, for instance, in Fig. 6, the time evolution of the frequency of nativelylike configurations with simulation time; however, the required long simulation times would defy the purpose of our investigation.

In order to overcome this bottleneck, one can think of two approaches. The microcanonical replica exchange molecular dynamic leads for finite times to quasicyclic motions in phase space. Introducing randomness in the system will destroy these deterministic motions and allow for sampling of a wider area in phase space. One possibility to introduce this

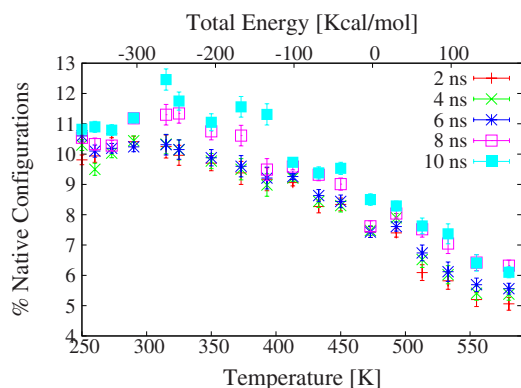


FIG. 6. (Color online) Frequency of nativelylike configurations (RMSD < 3.7 Å) as function of simulation time as measured in a microcanonical replica molecular dynamics simulation with the updates proposed in Ref. [5].

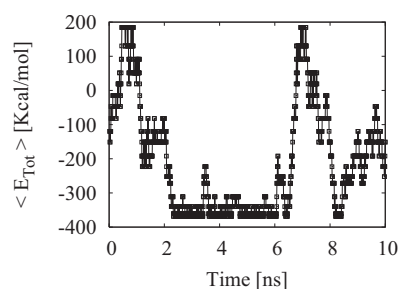


FIG. 7. Walk of a specific replica through energy in a microcanonical replica molecular dynamics with trial of exchange moves given by Eq. (8)

randomness is by periodic refreshing of the velocities at the highest-energy shell. This is justified as the underlying assumption of replica exchange methods is that a given replica can cross any relevant barrier, and therefore loses history, once it reaches the highest temperature/energy. As our data show, this is not the case in the MREMD but can be enforced by such randomization of velocities at this energy shell. We call this version randomized microcanonical replica exchange molecular dynamics (RMREMD). By the definition of the method, the walk of replicas through the various energy shells for RMREMD is still deterministic and does not differ from that of the original method (MREMD) displayed in Fig. 3(b).

A second possibility to introduce randomness in the motion is by way of the replica exchange move, i.e., giving up the rejection-free exchange moves in microcanonical replica exchange molecular dynamics. A possible approach is to enforce validity of Eq. (8) by exchanging replicas between energy shells according to this distribution. We name this version of our approach weighted microcanonical replica exchange molecular dynamics (WMREMD). The resulting random walk through the energy shells is displayed in Fig. 7.

We have performed simulations of both variants with same statistics as in the case of canonical replica exchange molecular dynamics and the original version of microcanonical replica exchange molecular dynamics. For a comparison of the various methods, we show in Fig. 8 the percentage of nativelylike configurations for all four methods. Note the difference between the original MREMD and RMREMD on one side and canonical replica exchange molecular dynamics

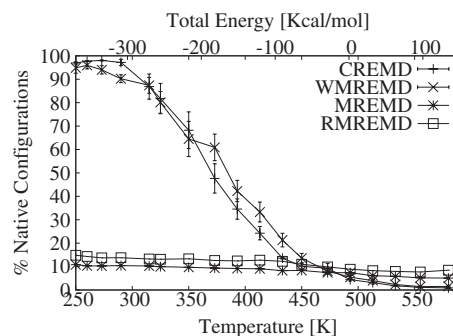


FIG. 8. Frequency of configurations with a RMSD smaller than 3.7 Å as measured in canonical (CREMD) and various versions of microcanonical replica exchange molecular dynamics.

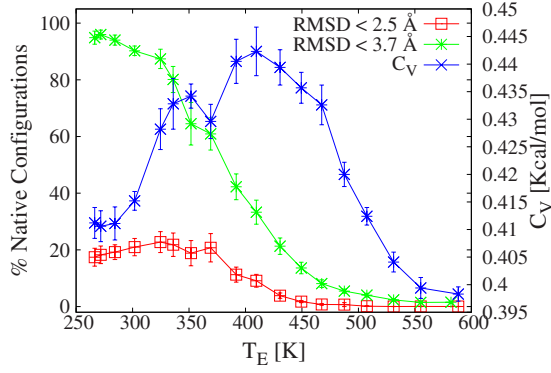


FIG. 9. (Color online) Frequency of nativelike configurations as measured according to two criteria (see text) and specific heat, as measured in simulations with our weighted microcanonical replica exchange molecular dynamics.

and WMREMD on the other side. While the first two lead at lowest energy (and corresponding temperature) to less than 20% of nativelike structures, the weighted microcanonical replica exchange molecular dynamics leads essentially to the same frequency as the canonical replica exchange molecular dynamics, i.e., about 90% of nativelike configurations. However, while the data in canonical replica exchange molecular dynamics rely on solely three round trips, WMREMD let to nine round trips, i.e., three times higher statistics.

So far, our investigation has shown that the weighted, i.e., modified, microcanonical replica exchange molecular dynamics (WMREMD) leads to correct averages and exhibits at least three times faster sampling than canonical replica exchange molecular dynamics. Having demonstrated the improved sampling, we want to show now how this allows us to study in detail the thermodynamics of the trp-cage protein. In Fig. 9, we show the frequency of configurations with a RMSD smaller than 3.7 Å and those with RMSD smaller than 2.5 Å. Approaching from high energies (temperatures) a critical energy (temperature) of  $\approx -120$  kcal/mol (corre-

sponding to  $T \approx 413$  K), the frequency of configurations with RMSD smaller than 3.7 Å increases dramatically and stays constant after approaching its maximum. On the other hand, configurations with RMSD smaller than 2.5 Å, i.e., those very close to the experimentally determined one, also first increase rapidly but decrease again after reaching its maximum value at  $\approx -272$  kcal/mol ( $T \approx 315$  K). Note that the increase in both curves is correlated with the position of the peak in specific heat

$$C = \frac{\langle E_{pot}^2 \rangle - \langle E_{pot} \rangle^2}{k_B T^2} \quad (17)$$

shown also in Fig. 9. On the other hand, the decrease observed for the frequency of configurations with RMSD smaller than 2.5 Å seems to be correlated with a shoulder in the specific heat curve.

A natural interpretation for the steep increase in the nativelike configuration (according to both definitions) and the peak in specific heat is that of a folding transition. The decrease in frequency of configurations, whose similarity to the native structure is measured according to the more stringent criteria of an RMSD smaller than 2.5 Å, requires a more detailed analysis. For this purpose, we show in Fig. 10 histograms of configurations as function of RMSD for four values of energy. At the highest-energy shell [184.3 kcal/mol, Fig. 10(a)], we observe a broad single-peaked distribution centered around a RMSD of  $\approx 6-7$  Å, indicating that at this energy (and corresponding temperature) configurations have little resemblance with the native structure. The distribution shown in (b) is drawn for  $E_{tot} = -119.8$  kcal/mol, the energy level corresponding to the peak in specific heat. Here, we find a distribution centered around a RMSD of  $\approx 4$  Å that covers both structures with large RMSD and such that resemble the native one (small RMSD). Hence at this energy level, which corresponds to a microcanonical folding temperature of  $\approx 413$  K, we have an equilibrium of unfolded and folded configurations (these with RMSD smaller than

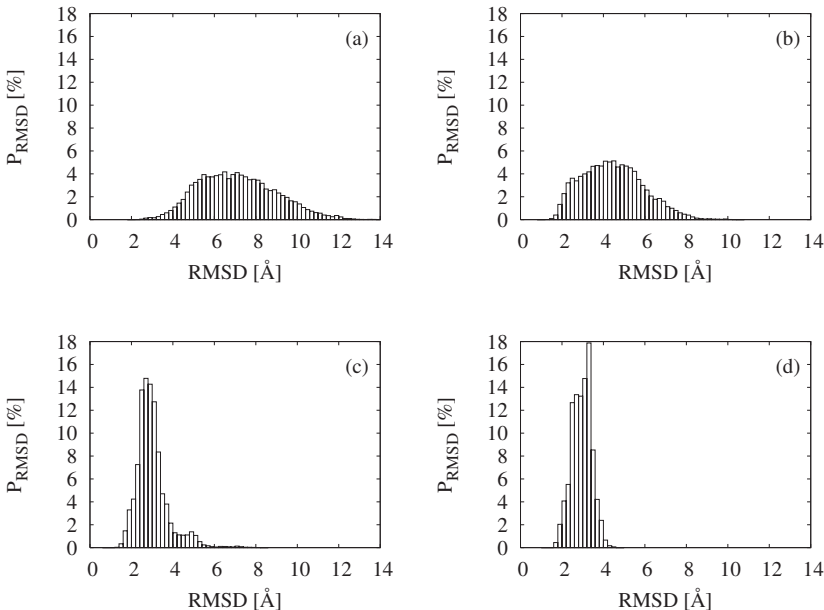


FIG. 10. Histograms of configurations as function of RMSD calculated for four different energy levels.

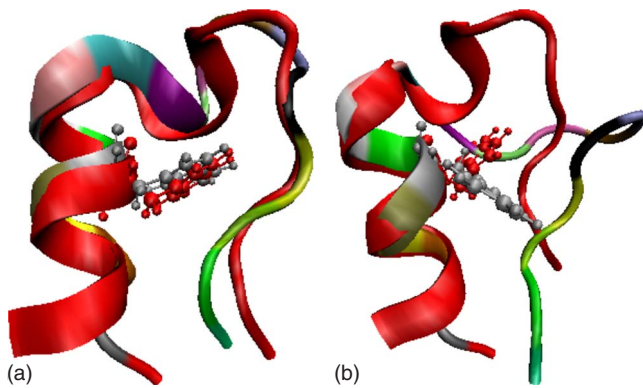


FIG. 11. (Color online) The two dominant low-energy structures (light color) shown in overlay with the native structure (red/dark).

3.7 Å). The third distribution [Fig. 10(c)] is calculated for  $E_{tot} = -271.7$  kcal/mol ( $T = 315$  K), i.e., the position of the shoulder in specific heat and maximum of the curve in Fig. 7 that displays the frequency of configurations with RMSD smaller than 2.5 Å. Again, we observe a single-peaked distribution centered around  $\approx 3$  Å that is almost exclusively made up of nativelike structures (such with a RMSD smaller than 3.7 Å). Surprisingly, this distribution does not become narrower when going to the lowest-energy level  $E_{tot} = -368.5$  kcal/mol nor does its center moves to smaller values of RMSD. Instead, the distribution has now a broad shoulder around a RMSD of  $\approx 2.5$  Å, extending into a large peak centered around a RMSD of 3.3 Å. We remark that with a finer bin size, the distribution becomes bimodal. Hence, this distribution indicates an equilibrium between configurations with RMSD around and smaller than 2.5 Å and such with RMSD between 3 and 4 Å. An example for both types of configurations is shown in Fig. 11.

In connection with Fig. 9, we interpret the series of histogram as follows. At a temperature of  $\approx 413$  K, we have a folding transition that separates unfolded configurations from an ensemble of configurations that are similar to the native structure. This ensemble is made up of two clusters of structures shown in Fig. 10. Both configurations are stabilized by a salt bridge between ASP9 and ARG16 that is responsible for the fast folding kinetics of this protein. Decreasing the temperature further increases the frequency of the configurations of Fig. 11(a). The overlay with experimentally deter-

mined structure emphasizes how closely the configuration resembles the native structure ( $\approx 2$  Å) not only in the backbone but also in the orientation of the tryptophan side chain. However, below a certain temperature, the frequency of configurations of this type decreases again and dominant now are the slightly different configurations of Fig. 11(b). These configurations differ from the native structure by RMSD about 3–4 Å and are characterized by a wrongly positioned tryptophan side chain and divergent backbone orientation at residue 9 that leads to this structure. Unlike in the native structure, the chain terminals are connected by hydrogen bonds that energetically favor this structure over the native form. The increase in frequency of these structure in lieu of the native one with decreasing temperature may indicate limitations in the accuracy of our energy function (see also Ref. [25]) but could also indicate a partial “cold unfolding.” In the later case, this would demonstrate again the well-known fact that the native state of a protein is the global minimum in free energy at physiological temperatures but not necessarily the global minimum in potential energy.

#### IV. CONCLUSIONS

We have tested a recently proposed microcanonical replica exchange molecular dynamics approach in simulations of the trp-cage protein in implicit solvent. We evaluated the performance of this method and introduced a variant that lead to improved sampling for this protein. Using this sampling technique, we could not only find the native structure of this protein within 2 Å RMSD but also show that the folding thermodynamics of this protein is surprisingly rich, with not only a folding transition but also indications for a partial cold unfolding.

#### ACKNOWLEDGMENTS

This work has been supported, in part, by the National Science Foundation under research Grant No. CHE-0809002. This work was started while P.K. was visiting the “Computational Biology and Biophysics” group (headed at that time by U.H.) at the John-von Neumann Institute for Computing (NIC), Jülich, Germany. He would like to thank NIC for its hospitality and support, especially the use of computers at NIC, where part of the simulations were done.

- [1] S. Mohanty, J. H. Meinke, O. Zimmermann, and U. H. E. Hansmann, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8004 (2008).
- [2] O. Zimmermann and U. H. E. Hansmann, *Biochim. Biophys. Acta* **1784**, 252 (2008).
- [3] S. Trebst, M. Troyer, and U. H. E. Hansmann, *J. Chem. Phys.* **124**, 174903 (2006).
- [4] W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **75**, 026109 (2007).
- [5] W. Nadler and U. H. E. Hansmann, *Phys. Rev. E* **76**, 057102 (2007).
- [6] W. Nadler, J. H. Meinke, and U. H. E. Hansmann, *Phys. Rev. E* **78**, 061905 (2008).
- [7] D. Gront and A. Kolinski, *J. Phys.: Condens. Matter* **19**, 036225 (2007).
- [8] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1989).
- [9] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, 2002).
- [10] M. Kohtani, T. C. Jones, J. E. Schneider, and M. F. Jarrold, *J. Am. Chem. Soc.* **126**, 7420 (2004).

- [11] C. Simmerling, B. Strockbine, and A. Roitberg, *J. Am. Chem. Soc.* **124**, 11258 (2002).
- [12] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Anderson, *Nat. Struct. Biol.* **9**, 425 (2002).
- [13] A. Schug, T. Herges, and W. Wenzel, *Phys. Rev. Lett.* **91**, 158102 (2003).
- [14] Y. He, Y. Xiao, A. Liwo, and H. A. Scheraga, *J. Comput. Chem.* **30**, 2127 (2009).
- [15] A. Schug, T. Herges, A. Verma, K. H. Lee, and W. Wenzel, *ChemPhysChem* **6**, 2640 (2005).
- [16] D. Paschek, H. Nymeyer, and A. E. Garcia, *J. Struct. Biol.* **157**, 524 (2007).
- [17] C. J. Geyer and A. Thompson, *J. Am. Stat. Assoc.* **90**, 909 (1995).
- [18] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [19] U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- [20] W. C. Still, A. Tempczyk, R. C. Hawley, and T. J. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1990).
- [21] V. Tsui and D. A. J. Case, *J. Am. Chem. Soc.* **122**, 2489 (2000).
- [22] J. M. Wang, P. Cieplak, and P. A. J. Kollman, *J. Comput. Chem.* **21**, 1049 (2000).
- [23] J. W. Pitera and W. Swope, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7587 (2003).
- [24] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. J. Hagen, *J. Am. Chem. Soc.* **124**, 12952 (2002).
- [25] F. Eisenmenger and U. H. E. Hansmann, *J. Phys. Chem. B* **101**, 3304 (1997).