Replica exchange molecular dynamics simulations of reversible folding

Francesco Rao and Amedeo Caflisch[a)]
Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

(Received 28 April 2003; accepted 21 May 2003)

The replica exchange molecular dynamics (REMD) approach is applied to a 20-residue three-stranded antiparallel β-sheet peptide. At physiologically relevant temperature REMD samples conformational space much more efficiently than constant temperature molecular dynamics (MD) and allows reversible folding (312 folding events during a total simulation time of 32 μs). The energetic and structural properties during the folding process are similar in REMD and conventional MD at the temperature values where there is enough statistics for the latter. The simulation results indicate that the unfolded state contains a significant amount of non-native interactions especially at low temperature. The folding events consist of a gradual replacement of non-native contacts with native ones which is coupled with an almost monotonic decrease of the REMD temperature. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1591721]

I. INTRODUCTION

To accurately describe the thermodynamics and kinetics of complex systems, such as biological macromolecules, a thorough sampling of the relevant conformations is required. Since such systems have energetic and entropic barriers that are higher than the thermal energy at physiological temperature standard molecular dynamics (MD) techniques often fail to adequately sample the conformational space. A number of approaches to enhance sampling of phase space have been introduced.[1,2] They are based on multiple time steps,[3] modified Hamiltonians,[4–6] or generalized ensembles e.g., entropic sampling, multicanonical methods, replica exchange methods (REM).[7] REM is an efficient way to simulate complex systems at low temperature and is the simplest and most general form of simulated tempering.[8] Sugita and Okamoto have extended the original formulation into an MD based version (REMD) and tested it on the pentapeptide Met-enkephalin *in vacuo*.[9] Sanbonmatsu and Garcia have applied REMD to investigate the structure of Met-enkephalin in explicit water[10] and the α-helical stabilization by the arginine side chain which was found to originate from the shielding of main chain hydrogen bonds.[11] Furthermore, the energy landscape of the C-terminal β-hairpin of protein G in explicit water has been investigated by REMD.[12,13] Recently, a multiplexed approach with multiple replicas for each temperature level has been applied to large-scale distributed computing for the folding of a 23-residue miniprotein.[14] Starting from a completely extended chain, conformations close to the NMR structures were reached in about 100 trajectories (out of a total of 4000) but no evidence of reversible folding (i.e., several folding and unfolding events in the same trajectory) was presented.[14]

Even for a small protein it is currently not yet feasible to simulate reversible folding with a high-resolution approach, e.g., MD simulations with an all-atom model. The practical difficulties in performing such brute force simulations have led to several types of computational approaches and/or approximative models to study protein folding. An interesting approach is to unfold starting from the native structure[15,16] but a detailed comparison with experiments[17] is mandatory to make sure that the high temperature sampling does not introduce artefacts. Another possibility is offered by very small protein fragments for which the conformational space is sufficiently small so that full searches can be accomplished and/or transitions of interest occur on a manageable time scale.[18] The thermodynamic properties of two peptides (an α-helix and a β-hairpin of 13 and 12 residues, respectively) have been determined using an implicit solvation model and adaptive umbrella sampling.[19] Furthermore, the free energy surface of Betanova, an antiparallel three-stranded β-sheet peptide, has been constructed starting from conformations obtained during unfolding simulations in explicit water at elevated temperatures (between 350 and 400 K).[20]

In previous studies we have shown that it is possible to simulate the reversible folding of structured peptides at relatively high temperature values (330–360 K)[21–25] using an implicit model of the solvent based on the accessible surface area.[26] In this work we use REMD to explore, at temperature values of 275–465 K, the conformational space of Beta3s a designed peptide (Thr_1-Trp_2-Ile_3-Gln_4-Asn_5-Gly_6-Ser_7-Thr_8-Lys_9-Trp_{10}-Tyr_{11}-Gln_{12}-Asn_{13}-Gly_{14}-Ser_{15}-Thr_{16}-Lys_{17}-Ile_{18}-Tyr_{19}-Thr_{20}) whose solution conformation has been studied by NMR.[27] Nuclear Overhauser enhancement (NOE) and chemical shift data indicate that at 10 °C Beta3s populates a single structured form, the expected three-stranded antiparallel β-sheet conformation with turns at Gly6-Ser7 and Gly14-Ser15, in equilibrium with the random coil. Furthermore, Beta3s was shown to be monomeric in aqueous solution by equilibrium sedimentation and NMR dilution experiments.[27] The folding behavior and energy surface of Beta3s are investigated here using the same implicit model of the solvent[26] and a comparison is made with previous constant temperature MD simulations.[25] This approximation is justified by explicit water MD studies which have shown

[a)]Author to whom correspondence should be addressed. Phone: (41 1) 635 55 21; fax: (41 1) 635 68 62; electronic mail: caflisch@bioc.unizh.ch

that the solvent does not play a detailed role in the folding of Betanova.[20]

The present study was motivated by two main questions: Does REMD allow a thorough sampling of the relevant conformations of a three-stranded antiparallel β-sheet peptide at physiologically relevant temperatures? Do the energetic and structural properties during the folding events sampled by REMD correspond to those observed in constant temperature MD simulations? The simulation results indicate that both questions can be answered affirmatively.

II. METHODS

A. Model

The MD simulations and part of the analysis of the trajectories were performed with the CHARMM program.[28] The peptide was modeled by explicitly considering all heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 potential function[28,29]). The remaining hydrogen atoms are considered as part of the carbon atoms to which they are covalently bound (an extended atom approximation). The effective energy, whose negative gradient corresponds to the force used in the dynamics (see also below), is of the form

$$E(\mathbf{r}) = E_{vacuo}(\mathbf{r}) + G_{\mathrm{solv}}(\mathbf{r}); \tag{1}$$

for a molecule with N atoms at Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$. The PARAM19 *vacuo* energy function is

$$E_{vacuo}(\mathbf{r}) = \frac{1}{2}\sum_{\text{bonds}} k_b(b-b_0)^2 + \frac{1}{2}\sum_{\substack{\text{bond}\\ \text{angles}}} k_\theta(\theta-\theta_0)^2 + \frac{1}{2}\sum_{\substack{\text{dihedral}\\ \text{angles}}} k_\phi[1+\cos(n\phi-\delta)]$$
$$+ \frac{1}{2}\sum_{\substack{\text{improper}\\ \text{dihedrals}}} k_\omega(\omega-\omega_0)^2 + \sum_{i>j} \varepsilon_{ij}^{\min}\left[\left(\frac{d_{ij}^{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{d_{ij}^{\min}}{r_{ij}}\right)^{6}\right] + \sum_{i>j}\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}},$$

where b is a bond length, θ a bond angle, ϕ a dihedral angle, ω an improper dihedral, r_{ij} is the distance between atoms i and j, q_i and q_j are partial charges, $d_{ij}^{\min}$ and $\varepsilon_{ij}^{\min}$ are the optimal van der Waals distance and energy, respectively, and $\epsilon(r_{ij})$ is a screening function. Parameters are given in Ref. 29.

An implicit model based on the solvent accessible surface was used to describe the main effects of the aqueous solvent on the solute.[26] In this approximation, the solvation free energy is given by

$$G_{\mathrm{solv}}(\mathbf{r}) = \sum_{i=1}^{M} \sigma_i A_i(\mathbf{r}), \tag{2}$$

for a molecule having M heavy atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_M)$. $A_i(\mathbf{r})$ is the solvent-accessible surface computed by an approximate analytical expression[30] and using a 1.4 Å probe radius. The model contains only two surface-tension-like parameters: one for carbon and sulfur atoms ($\sigma_{\mathrm{C,S}} = 0.012$ kcal/mol Å^2), and one for nitrogen and oxygen atoms ($\sigma_{\mathrm{N,O}} = -0.060$ kcal/mol Å^2).[26] Furthermore, ionic side chains were neutralized[31] and a linear distance-dependent screening function [$\epsilon(r_{ij}) = 2r_{ij}$] was used for the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long range interactions were used, i.e., a shift function[28] was employed with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. This cutoff length was chosen to be consistent with the parametrization of the force-field and implicit solvation model. The model is not biased toward any particular secondary structure type. In fact, exactly the same force field and implicit solvent model have been used recently in MD simulations of folding of structured peptides (α-helices and β-sheets) ranging in size from 15 to 31 residues,[22–24] and small proteins of about 60 residues.[32,33] Despite the lack of friction due to the absence of explicit water molecules, the implicit solvent model yields a separation of time scales consistent with experimental data: helices fold in about 1 ns[21] (≈ 0.1 μs, experimentally[34]), β-hairpins in about 10 ns[21] (≈ 1 μs),[34] and triple-stranded β-sheets in about 100 ns[25] (≈ 10 μs).[27]

B. REMD simulations

The basic idea of REMD is to simulate different copies (*replicas*) of the system at the same time but at different temperature values. Each replica evolves independently by MD and every 1000 MD steps (2 ps), states i,j with neighbor temperatures are swapped (by velocity rescaling) with a probability $w_{ij} = \exp(-\Delta)$,[9] where $\Delta \equiv (\beta_i - \beta_j)(E_j - E_i)$, $\beta = 1/kT$ and E is the effective energy [Eq. (1)]. During the 1000 MD steps the Berendsen thermostat is used to keep the temperature close to a given value with a coupling of 0.1 ps. This rather tight coupling and the length of each MD segment (2 ps) allows the kinetic and potential energy of the system to relax. High temperature simulation segments facilitate the crossing of the energy barriers while the low temperature ones explore in detail the conformations present in the minimum energy basins. The result of this swapping between different temperatures is that high temperature replicas help the low temperature ones to jump across the energy barriers of the system.

In this study 8 replicas were used with temperatures (in K): 275, 296, 319, 344, 371, 400, 431, 465. The acceptance ratio of exchange between neighbor temperatures was around

TABLE I. Simulations performed.

#	Length (μs)	T (K)	Method	Starting conformation	Folding events	Folding time (μs)
2	8×1	275–465	REMD	native	152	0.064
2	8×1	275–465	REMD	unfolded	160	0.067
8	1	275	MD	unfolded	0	⋯
4	1	296	MD	unfolded	2	>0.44
4	2.7,2.7,2.8,4.4	330	MD	native	72	0.085
1	1	465	MD	unfolded	0	⋯

20% to 30%. Four REMD runs each with 8 replicas were performed: two started from the native structure and two from an unfolded structure obtained in a run at 330 K (see below). Each trajectory has a length of 1 μs for a total of 32 μs of simulation time (see Table I).

C. Constant temperature MD simulations

A series of control runs of 1 μs each were performed at the two lowest (275 K, 8 runs and 296 K, 4 runs) and the highest (465 K, 1 run) temperature value used in REMD. The Berendsen thermostat was used and the starting structure was the same unfolded conformation as the one employed in two of the four REMD runs. In addition, four simulations at the melting temperature of 330 K[25] started from the folded state (a total of 12.6 μs which were available from another study[35]) were used as a comparison for the folding process between conventional MD and REMD (see Table I).

For both REMD and constant temperature MD, the SHAKE algorithm[36] was used to fix the length of the covalent bonds involving hydrogen atoms, which allows an integration time step of 2 fs. Furthermore, the nonbonded interactions were updated every 10 dynamics steps and coordinate frames were saved every 20 ps for a total of 5×10^4 conformations/μs. A 1 μs run requires approximately 2 weeks on a 1.4 GHz Athlon processor and the REMD simulations were run in parallel on a Linux Beowulf cluster.

D. Native contacts and progress variables

The conformations sampled previously at 300 K were used to define a list of 26 native contacts, of which 11, 11, 2 and 2 involve residues in strands 1–2, 2–3, 2–2 and 3–3, respectively (see Table 2 of Ref. 23). These include 10 backbone hydrogen bonds (five on each β-hairpin) and 16 contacts between side chains. The folding progress variable Q_{nat} is defined as the fraction of contacts common to both the current conformation and the native structure. It can be plotted as a function of simulation time to monitor the amount of folded structure. For a given snapshot along a trajectory, a native hydrogen bond is considered formed if the O$\cdots$H distance is smaller than 2.6 Å. A native side chain contact is considered formed if the distance between geometrical centers is smaller than 6.7 Å. A conformation is considered folded when the fraction of native contacts Q_{nat} is larger than 0.85 ($Q_{\mathrm{nat}}>22/26$) and unfolded when it is smaller than 0.15 ($Q_{\mathrm{nat}}<4/26$).[23,25] The folding time is defined as the temporal interval between the first time point with $Q_{\mathrm{nat}}>22/26$ and the first time point with $Q_{\mathrm{nat}}<4/26$ just after the preceding folding event. The following subsets of native contacts were used for monitoring the folding pathways: Q_{1-2} is defined as the fraction of the 11 native contacts (5 hydrogen bonds and 6 side chain interactions) formed between strands 1 and 2, while Q_{2-3} as the fraction of the 11 native contacts between strands 2 and 3.[23]

The total number of contacts (N_{tot}) includes native and non-native ones, and is computed counting all hydrogen bonds and side-chain contacts between all pairs of residues at least three positions apart in the sequence. In addition, the contacts between side chains of the pairs of residues 8–10, 16–18 and 18–20 are included because they are considered native contacts. The fraction of total contacts Q_{tot} is $N_{\mathrm{tot}}/26$ where the denominator was chosen to facilitate the comparison with Q_{nat} (note that Q_{tot} can be larger than 1 but this happens sporadically).

E. Effective energy and free energy

The effective energy and free energy surfaces, determined by simulations and experiments, play an important role for the understanding of the protein folding reaction.[37] The effective energy is the sum of the intramolecular energy (CHARMM PARAM19 force field energy) and the solvation free energy [Eq. (1)]. The latter is approximated by the solvent accessible surface term[26] [Eq. (2)] and contains the free energy contribution of the solvent within the approximations of an implicit model of the water molecules. The effective energy does not include the configurational entropy of the peptide which consists of conformational and vibrational entropy contributions.[31] For a system in thermodynamic equilibrium, the difference in free energy in going from state A to state B is proportional to the natural logarithm of the quotient of the probability of finding the system in state A divided by the probability of state B.

Due to the complexity of the protein folding process, it is necessary to group states and project the energy onto one or two order parameters that characterize the system. The value of the effective energy is averaged within a bin defined by discretizing the reduced space. For the free energy profiles the probability of finding the system in a given bin is assumed to be proportional to the number of MD snapshots belonging to that bin. An arbitrarily chosen reference point (e.g., the fully unfolded state) is used as the denominator of the probability quotient.

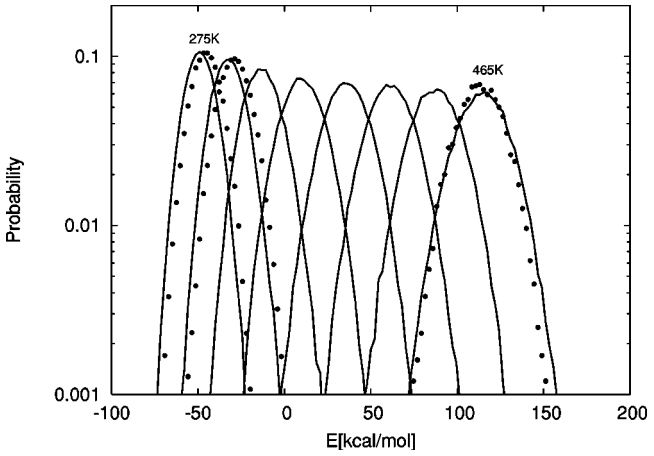

FIG. 1. Probability distribution of the effective energy in one of the two REMD runs starting from unfolded (solid lines) and constant temperature MD runs (filled circles). The REMD distributions correspond to the following temperatures (from left to right): 275, 296, 319, 344, 371, 400, 431, and 465 K. The constant temperature MD distributions (from left to right) correspond to 8 μs at 275 K, 4 μs at 296 K and 1 μs at 465 K.

III. RESULTS AND DISCUSSION

A. REMD diagnostics

The temperature values used in a REMD simulation have to be chosen carefully for an efficient sampling of the properties of interest. The highest value has to be high enough to jump over the energy barriers of the system, while the lowest value has to allow the exploration of the details of the energy minima. On the other hand, given a fixed number of replicas the range of temperature values cannot be too large because the acceptance probability for a temperature exchange has to allow a reasonable number of swaps during a simulation run. This implies that the temperature values need to be close enough to each other to guarantee a good overlap of the energy histograms (Fig. 1). An optimal distribution implies that the acceptance ratio for a swap between neighbor temperatures is nearly constant, resulting in a free random walk in temperature space. The filled circles in Fig. 1 show the results from the constant temperature MD simulations (at 275, 296, and 465 K) which were started from the same unfolded conformation used as a starting structure in two of the four REMD runs. The distributions agree at high temperature but tend to shift towards less favorable energies at low temperature values. This shows that conventional MD at low temperature can get trapped in local energy minima while REMD is superior in sampling conformational space. The time series of the exchanges of temperature values for one replica is shown in Fig. 2. It is clear that the trajectory visits all temperature levels several times within the 1 μs of simulation time. The other replicas show similar exchange patterns in all of the four REMD runs.

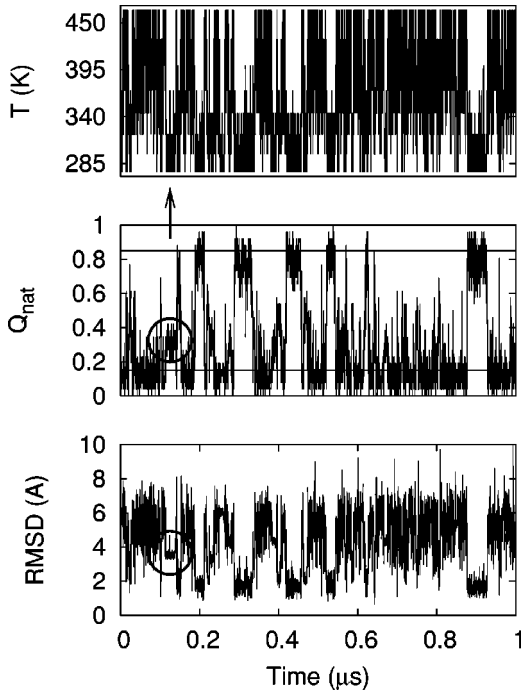

FIG. 2. Time series of (from top to bottom) the temperature T, fraction of native contacts Q_{nat}, and the C_α RMSD from the folded structure for a replica of a REMD run started from unfolded. The circles and arrow denote a simulation interval where the temperature is low but the peptide is not folded (see the text).

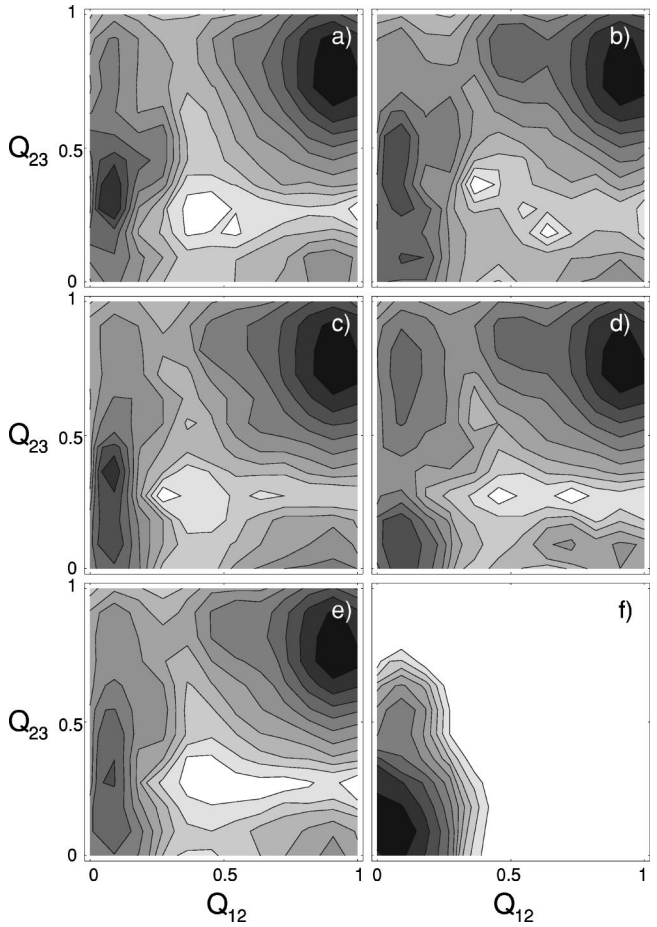

FIG. 3. Projection of the free energy surface onto the progress variables Q_{1-2} and Q_{2-3} at 275 K. (a)–(b) Two REMD runs started from unfolded. (c)–(d) Two REMD runs started from folded. In (a)–(d) each surface is plotted using 1 μs of data (5×10^4 conformations) sampled at 275 K. (e) Average over the four REMD runs. (f) Average over the eight 1-μs constant temperature MD runs at 275 K.

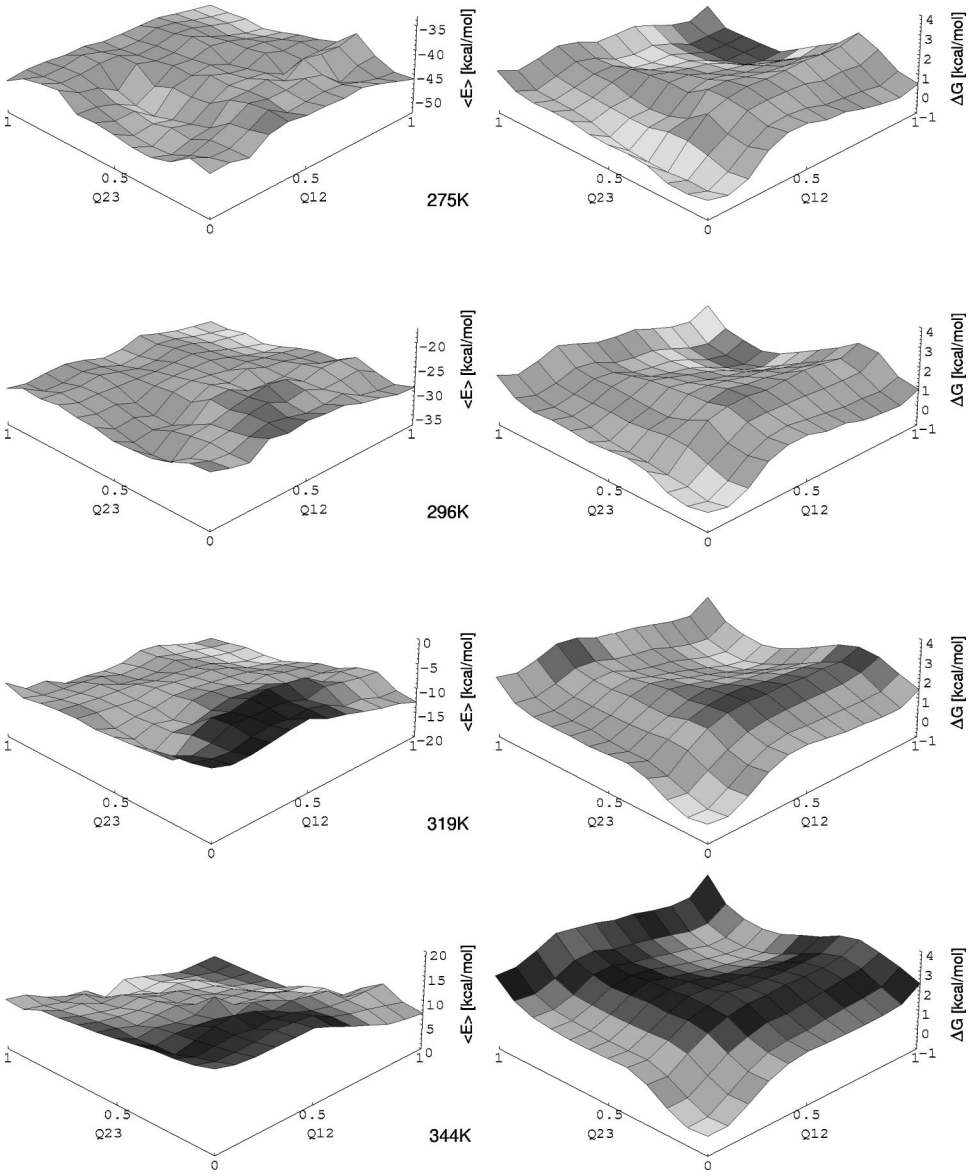

FIG. 4. Average effective energy (left) and free energy surface (right) as a function of the progress variables Q_{1-2} and Q_{2-3} in the REMD runs. The plots correspond to the following temperatures (from top to bottom) 275, 296, 319, and 344 K. A total of 2×10^5 conformations sampled during the 4 REMD runs were used for each value of the temperature.

B. Reversible folding

The time series of the fraction of native contacts (Q_{nat}) and C_α RMSD from the average NMR conformation indicate that several folding events are sampled along the REMD trajectories (Fig. 2). The C_α RMSD from native averaged over the time intervals where the structure is folded (e.g., 0.88–0.92 μs for the replica shown in Fig. 2) is 1.7 Å. A total of 312 folding events are sampled along the total simulation time of 32 μs. This corresponds to an average folding time of 0.065 ± 0.006 μs which is about 20% shorter than the value obtained by averaging over the 72 folding events sampled at 330 K constant temperature MD. It is important to note that there were only two folding events in the four 1 μs constant temperature runs at 296 K from unfolded (Table I). Moreover, in the eight 1 μs constant temperature runs at 275 K the value of Q_{nat} never exceeded 0.4 and the C_α RMSD from native was always above 3.5 Å.

Interestingly, in the REMD simulation intervals where the temperature is moderate (below 320 K) the peptide is folded most of the time but can also assume non-native conformations with values of $Q_{\text{nat}}<0.4$ and RMSD>3.5 Å. This indicates that the unfolded state is explored at all temperature values (see the circles and arrow in Fig. 2).

C. Energy surfaces

Figure 3 shows a projection of the free energy surface for the REMD conformations sampled at 275 K and the constant temperature MD runs. It is clear that while the latter explores only a small portion of the unfolded state, the former samples both the folded and unfolded states. Moreover, if one neglects the noise due to frustration in the unfolded state the REMD surfaces at 275 K look similar among each other although two different initial conformations were used in the four REMD runs [Figs. 3(a)–3(d)]. This indicates that the choice of the initial conformation does not affect the REMD sampling.

Figures 4 and 5 show the average effective energy and free energy surfaces at the four low temperature values used in REMD. Raising the temperature results in a less rugged

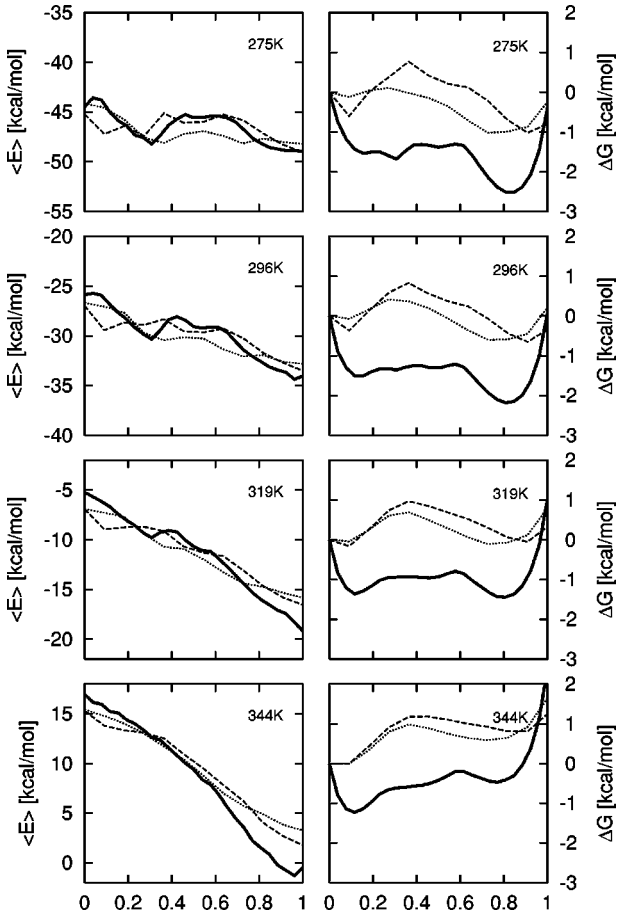

FIG. 5. Average effective energy (left) and free energy surface (right) as a function of the fraction of native contacts Q_{nat} (thick lines), Q_{1-2} (dashed lines), and Q_{2-3} (dotted lines). The same temperature values as in Fig. 4.

and more downhill profile of the effective energy. At 275 K and 296 K there is a pronounced minimum of the effective energy with values of Q_{2-3} in the range 0.3–0.7 and Q_{1-2} <0.3. This minimum was not observed in our previous constant temperature simulations and indicates that the β-hairpin 2–3 has an intrinsic enthalpic stabilization due to the interactions between the side chains of Trp10 and Tyr19 and Trp10 with the nonpolar part of the Lys17 side chain. This is consistent with an NMR analysis of two shorter peptides encompassing the single β-hairpins, i.e., peptides TWIQNGSTKWYQ and KWYQNGSTKIYT corresponding to Beta3s residues 1–12 and 9–20, respectively.[27] The NMR data at 10 °C indicate that the latter is as stable as Beta3s while the former is less stable.

The projection of the effective energy and free energy into Q_{1-2} and Q_{2-3} are consistent with our previous simulation results at constant temperature (330 K[25] and 360 K[23]). The free energy surfaces are more symmetric at higher temperature values (Fig. 4, right) because the entropic contribution starts to dominate and the conformational entropy penalty during folding is similar for both hairpins. The projections of the free energy along Q_{1-2} and Q_{2-3} (Fig. 5, right, dashed and dotted lines, respectively) indicate that the barrier is higher for the former especially at 275 and 296 K where the enthalpic contribution is more favorable for the formation of the second hairpin (Fig. 5, left). This effect is still present though less pronounced at higher temperatures and is consistent with previous constant temperature MD simulations at 330 K[25] and 360 K.[23]

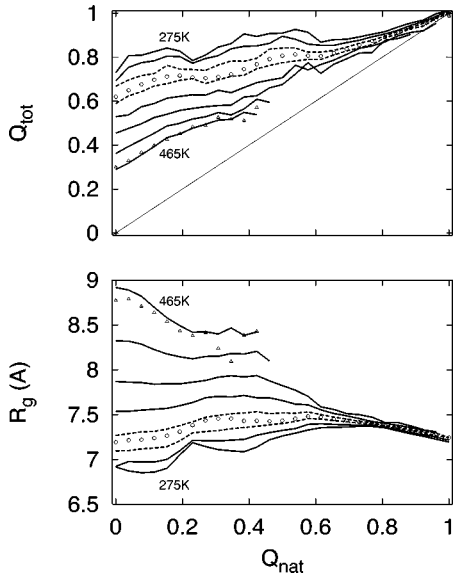

FIG. 6. Average fraction of total contacts (top) and average value of the radius of gyration (bottom) as a function of the fraction of native contacts Q_{nat}. REMD data (solid lines for all temperature values except for 319 K and 344 K in dashed lines) were obtained from the four runs, i.e., 4 μs for each temperature value. Constant temperature MD data are shown with symbols (12.6 μs at 330 K, circles; 1 μs at 465 K, triangles).

D. Non-native contacts and radius of gyration

At low temperature, non-native contacts in the unfolded state can act as traps. During folding, the total number of contacts grows moderately and non-native contacts are replaced by native ones [Fig. 6(a)]. At high temperature, the unfolded state is less compact [Fig. 6(b)] and has few non-native interactions so that the number of contacts during folding shows a more pronounced increase than at low temperature. From Fig. 6 it is evident that the behavior of total number of contacts and radius of gyration during folding is essentially the same in REMD and constant temperature MD at both medium (i.e., close to the melting temperature) and high temperature values. This provides additional evidence that the sampling of conformational space in REMD is correct.

E. Folding events

The previous analysis was based on the projections of energetic and structural properties along progress variables defined by the number of contacts present in the folded state or subsets thereof. Since folding is a complicated process with several degrees of freedom involved the choice of an adequate progress variable is not straightforward.[38] For this reason, it is useful to supplement the previous projections with the analysis of the folding mechanism during the time interval before reaching the folded structure. Figure 7 shows the behavior of temperature, number of native contacts and total number of contacts averaged over the 75 folding events that took longer than 5 ns along one of the two REMD runs

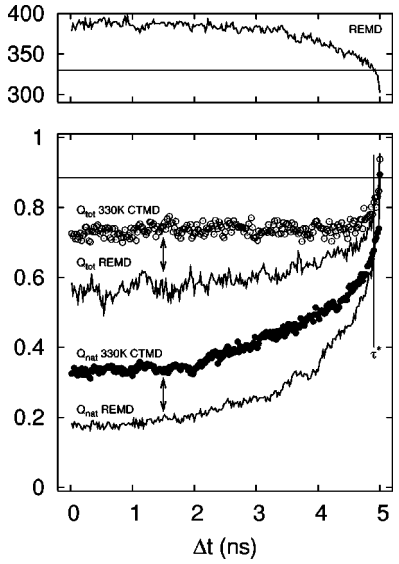

FIG. 7. The temperature (top) and average fraction of contacts (bottom) plotted during the 5 ns before reaching the folded structure for the REMD runs (solid lines) and the 330 K constant temperature MD simulations (circles). The REMD data are averages over 75 folding events along one of the two REMD runs started from unfolded. The constant line in the top marks the melting temperature of 330 K while the one in the bottom the definition of a folded structure, i.e., $Q_{\text{nat}} > 22/26$. The plot shows that the discrepancy in the total number of contacts between REMD and constant temperature MD simulations originates from the difference in native contacts (arrows) which implies that the number of non-native contacts is approximately the same. The time point τ^* at which the REMD temperature approaches 330 K is marked by a vertical line (bottom).

started from unfolded. In the last 5 ns before reaching the folded structure (defined by $Q_{\text{nat}} > 22/26$), the temperature on average decreases almost monotonically and the decrease is more pronounced the closer the system approaches the folded state. The conformations sampled during the folding process are characterized by minor variations in the number of contacts, i.e., an almost monotonic replacement of non-native interactions with native ones. At the beginning of the 5 ns time interval before reaching the folded structure the system is on average at elevated temperature in REMD with a value of $Q_{\text{nat}} \sim 0.18$ which is smaller than in the 330 K constant temperature MD ($Q_{\text{nat}} \sim 0.34$) because less contacts are formed at high temperature in REMD between strands 2 and 3 (not shown). Interestingly, the difference in the number of native contacts is the cause of the difference in the Q_{tot} curves (the arrows in Fig. 7) which shows that the two simulation types yield the same amount of non-native contacts during the folding process. This indicates that the REMD approach does not produce an artificial Go-like dynamics where non-native contacts are explicitly penalized. The REMD curves approach the same value of the constant temperature ones at the time point where the temperature is about 330 K (τ^* in Fig. 7).

IV. CONCLUSIONS

Four main results emerge from the REMD simulations of Beta3s with an implicit solvent. First, it is possible to sample the reversible folding of a structured peptide of 20 residues at physiologically relevant temperatures. This allows us to extract equilibrium properties even at low temperature and yields an atomic level description of the most populated conformations which is not (yet) feasible with conventional MD. The REMD approach is useful for an efficient sampling of the phase space and unlike other methods (like entropic sampling or the multicanonical method) REMD does not require the evaluation of the density of the states in a not obvious and tedious iterative procedure. Moreover, it can be implemented in a straightforward way on a parallel computer giving a scalability almost linear in the number of replicas used. Second, the important energetic and structural properties (e.g., average effective energy, number of non-native contacts, radius of gyration) monitored along the folding process are the same in REMD and constant temperature MD. The discrepancies at low temperature values are due to the limitations in sampling by constant temperature MD. Third, the effective energy surface at low temperature is more rugged and less symmetric with respect to the formation of the two hairpins than at high temperature. A similar but less pronounced temperature dependence is observed for the free energy surface. Fourth, the unfolded state can be investigated under folding conditions, namely physiologically temperatures, and contains a significant portion of non-native structures whose amount is inversely related to the temperature. The high amount of non-native interactions in the unfolded state at low temperature might be valid, in general, for structured peptides and will be analyzed in more detail by further simulation studies. In conclusion, REMD seems particularly useful to study the reversible folding of structured peptides (and probably small proteins in the near future) at the atomic level of detail.

ACKNOWLEDGMENTS

We are grateful to Dr. A. Cavalli, M. Cecchini, E. Paci, and G. Settanni for helpful discussions. We thank Dr. M. Seeber for developing software tools used to analyze the trajectories. We thank A. Widmer (Novartis Pharma, Basel) for providing the molecular modeling program Wit!P which was used for a visual analysis of the trajectories. The simulations were performed on a Beowulf cluster running Linux and we thank Urs Haberthür for his help in setting up the cluster. This work was supported by the Swiss National Competence Center in Structural Biology (NCCR) and the Swiss National Science Foundation (Grant No. 31-64968.01 to A. C.).

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulations* (Academic, San Diego, 2002).
[2] B. J. Berne and J. E. Straub, Curr. Opin. Struct. Biol. **7**, 181 (1997).
[3] T. Schlick, E. Barth, and M. Mandziuk, Annu. Rev. Biophys. Biomol. Struct. **26**, 181 (1997).
[4] X. Wu and S. Wang, J. Phys. Chem. B **102**, 7238 (1998).
[5] J. Apostolakis, P. Ferrara, and A. Caflisch, J. Chem. Phys. **110**, 2099 (1999).
[6] I. Andricioaei, A. R. Dinner, and M. Karplus, J. Chem. Phys. **118**, 1074 (2003).
[7] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).
[8] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[9] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[10] K. Sanbonmatsu and A. Garcia, Proteins: Struct., Funct., Genet. **46**, 225 (2002).
[11] A. E. Garcia and K. Sanbonmatsu, Proc. Natl. Acad. Sci. U.S.A. **99**, 2782 (2002).

[12] A. E. Garcia and K. Sanbonmatsu, Proteins: Struct., Funct., Genet. **42**, 345 (2001).
[13] R. Zhou, B. Berne, and R. Germain, Proc. Natl. Acad. Sci. U.S.A. **98**, 14931 (2001).
[14] Y. M. Rhee and V. S. Pande, Biophys. J. **84**, 775 (2003).
[15] A. Caflisch and M. Karplus, J. Mol. Biol. **252**, 672 (1995).
[16] T. Lazaridis and M. Karplus, Science **278**, 1928 (1997).
[17] U. Mayor, N. R. Guydosh, C. M. Johnson *et al.*, Nature (London) **421**, 863 (2003).
[18] D. Mohanty, R. Elber, D. Thirumalai, D. Beglov, and B. Roux, J. Mol. Biol. **272**, 423 (1997).
[19] M. Schaefer, C. Bartels, and M. Karplus, J. Mol. Biol. **284**, 835 (1998).
[20] B. D. Bursulaya and C. L. Brooks III, J. Am. Chem. Soc. **121**, 9947 (1999).
[21] P. Ferrara, J. Apostolakis, and A. Caflisch, J. Phys. Chem. B **104**, 5000 (2000).
[22] A. Hiltpold, P. Ferrara, J. Gsponer, and A. Caflisch, J. Phys. Chem. B **104**, 10080 (2000).
[23] P. Ferrara and A. Caflisch, Proc. Natl. Acad. Sci. U.S.A. **97**, 10780 (2000).
[24] P. Ferrara and A. Caflisch, J. Mol. Biol. **306**, 837 (2001).
[25] A. Cavalli, P. Ferrara, and A. Caflisch, Proteins: Struct., Funct., Genet. **47**, 305 (2002).
[26] P. Ferrara, J. Apostolakis, and A. Caflisch, Proteins: Struct., Funct., Genet. **46**, 24 (2002).
[27] E. De Alba, J. Santoro, M. Rico, and M. A. Jiménez, Protein Sci. **8**, 854 (1999).
[28] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[29] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[30] W. Hasel, T. F. Hendrickson, and W. C. Still, Tetrahedron Comput. Methodol. **1**, 103 (1988).
[31] T. Lazaridis and M. Karplus, Proteins: Struct., Funct., Genet. **35**, 133 (1999).
[32] J. Gsponer and A. Caflisch, J. Mol. Biol. **309**, 285 (2001).
[33] J. Gsponer and A. Caflisch, Proc. Natl. Acad. Sci. U.S.A. **99**, 6719 (2002).
[34] W. A. Eaton, V. Munoz, J. Hagen, S. G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter, Am. Lab. (Boston) **29**, 327 (2000).
[35] A. Cavalli, U. Haberthür, E. Paci, and A. Caflisch, Protein Sci. (to be published).
[36] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[37] A. R. Dinner, A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus, TIBS **25**, 331 (2000).
[38] A. Dinner and M. Karplus, J. Phys. Chem. B **103**, 7976 (1999).