Multiscale implementation of infinite-swap replica exchange molecular dynamics

Tang-Qing Yu, Jianfeng Lu, Cameron F. Abrams, and Eric Vanden-Eijnden

Abstract

Replica exchange molecular dynamics (REMD) is a popular method to accelerate conformational sampling of complex molecular systems. The idea is to run several replicas of the system in parallel at different temperatures that are swapped periodically. These swaps are typically attempted every few MD steps and accepted or rejected according to a Metropolis–Hastings criterion. This guarantees that the joint distribution of the composite system of replicas is the normalized sum of the symmetrized product of the canonical distributions of these replicas at the different temperatures. Here we propose a different implementation of REMD in which (i) the swaps obey a continuous-time Markov jump process implemented via Gillespie’s stochastic simulation algorithm (SSA), which also samples exactly the aforementioned joint distribution and has the advantage of being rejection free, and (ii) this REMD-SSA is combined with the heterogeneous multiscale method to accelerate the rate of the swaps and reach the so-called infinite-swap limit that is known to optimize sampling efficiency. The method is easy to implement and can be trivially parallelized. Here we illustrate its accuracy and efficiency on the examples of alanine dipeptide in vacuum and C-terminal β-hairpin of protein G in explicit solvent. In this latter example, our results indicate that the landscape of the protein is a triple funnel with two folded structures and one misfolded structure that are stabilized by H-bonds.

Significance

Efficiently sampling the conformational space of complex molecular systems is a difficult problem that frequently arises in the context of molecular dynamics (MD) simulations. Here we present a modification of replica exchange MD (REMD) that is as simple to implement and parallelize as standard REMD but is shown to perform better in terms of sampling efficiency. The method is used here to investigate the multifold structure of the C-terminal β-hairpin of protein G in explicit solvent, which to date has required much longer REMD simulations to produce converged predictions of conformational statistics. In particular, we identify a potentially important folded β structure previously undetected by other importance sampling methods.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: eve2@ims.nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1605089113/-/DCSupplemental.
be simulated via Gillespie’s stochastic simulation algorithm (SSA) (21), which is rejection-free: Given the current assignments of the temperatures across the replicas, the method computes directly the time at which the next swap occurs and proceeds with the MD until that time—this is in contrast with standard REMD, in which the swaps are proposed (and rejected or accepted) at fixed time lags. This REMD-SSA can then be combined with multiscale simulations schemes such as the heterogeneous multiscale method (HMM) (22–24) to effectively compute at the infinite-swap limit. Here, we discuss the theoretical and computational aspects of this reformulation of REMD involving a multiscale SSA, termed REMD-MSSA for short, and apply it to compute free-energy surfaces (FESs) of alanine dipeptide (AD) and a 16-residue β-hairpin from protein G. In both applications, we observe that REMD-MSSA is more efficient than standard REMD. In particular, the rapid convergence of the free energy calculation in the context of β-hairpin allows one to identify a β structure.

Methodology

REMD-SSA. Let us propose a reformulation of REMD such that (i) the temperature swaps are replaced by swaps in factors multiplying the forces acting on the replicas and (ii) these swaps occur via a continuous-time MJP—the generalization to other control parameters is straightforward and is considered briefly in Using Parameters Other Than Temperature. To this end, it is useful to take as a starting point the probability distribution that a REMD simulation is designed to sample. Suppose we use N replicas with positions denoted collectively as $X = \{x_1, \ldots, x_N\}$ and let $\beta_1 > \beta_2 > \cdots > \beta_N$ be the N inverse temperatures that are being swapped over these replicas. Denote also $\rho_\sigma(x) = Z_\sigma^{-1} e^{-\beta_\sigma V(x)}$ the distribution at inverse temperature $\beta_\sigma$ over the atomic potential $V(x)$ (with $Z_\sigma = \int e^{-\beta_\sigma V(x)} \, dx$). Then REMD samples the symmetrized equilibrium probability density (18, 19):

$$ q(X) = \frac{1}{N!} \sum_\sigma \rho_\sigma(x_1) \cdots \rho_\sigma(x_N), \quad [1] $$

where the sum is taken over all of the permutation $\sigma$ of the indices $\{1, \ldots, N\}$ (with $\sigma(i)$ denoting the index onto which $i$ is mapped by the permutation $\sigma$). The symmetrized density in Eq. 1 can be thought of as the marginal density on the positions $X$ alone of the following joint distribution for $X$ and the permutation $\sigma$:

$$ q(\sigma, X) = \frac{1}{N!} \rho_\sigma(x_1) \cdots \rho_\sigma(x_N). \quad [2] $$

Performing temperature swaps is equivalent to evolving the permutation $\sigma$ concurrently with the replica configurations $X$ in a way that is consistent with Eq. 2. In standard REMD this is done by proposing a new permutation $\sigma' \neq \sigma$ after a fixed time lag and accepting or rejecting it according to a Metropolis–Hastings criterion. However, it is easy to modify the method and make both $X$ and $\sigma$ continuous-time Markov processes in which the updates of $\sigma$ occur at random times. Introducing the symmetrized (mixture) potential

$$ V(\sigma, \sigma') = -\beta^{-1} \log q(\sigma, X) = \beta^{-1} \sum_{i=1}^{N} \beta_\sigma(x_i) \nabla V(x_i) + \text{cst}, \quad [3] $$

where we used $\beta \equiv \beta_1$ as reference temperature, and noting that $V_\sigma(\sigma, \sigma') = \beta^{-1} \sum_{i=1}^{N} \beta_\sigma(x_i) \nabla V(x_i)$, amounts to imposing that:

i) the replica positions evolve via standard MD (using, e.g., Langevin’s thermostat with friction coefficient $\gamma$) over the potential (Eq. 3),

$$ x_i = m^{-1}p_i; \quad p_i = -\beta^{-1} \beta_\sigma(x_i) \nabla V(x_i) - p_i + 2\gamma m^{-1} \eta; \quad [4] $$

where $\eta$ is white noise with mean zero and covariance $\langle \eta_\sigma(t) \eta_\sigma'(t') \rangle = 2\gamma \delta(t-t')\delta(\sigma, \sigma')$. And

ii) the permutation $\sigma$ is updated via the continuous-time MJP with rate

$$ r_{\sigma, \sigma'}(X) = a_{\sigma, \sigma'} e^{-\phi(\sigma, \sigma')(X)} \quad [5] $$

where $\gamma > 0$ is a constant that controls the swapping rate (the higher $\gamma$, the more frequently the jumps occur), and the symmetric matrix $a_{\sigma, \sigma'}$ indicates whether the permutation $\sigma'$ is accessible from $\sigma$, that is, $a_{\sigma, \sigma'} = 1$ if we allow the jump from $\sigma$ to $\sigma'$ and $a_{\sigma, \sigma'} = 0$ otherwise—below we will restrict ourselves to transposition moves in which two random indices are swapped and for which the total number of accessible states $\sigma'$ from $\sigma$ is $N(N-1)/2$.

These dynamics can also be reformulated as follows. Given that $\sigma(t) = \sigma$ at time $t$, the probability that it jumps out of $\sigma$ for the first time in the infinitesimal time interval $[t', t + dt]$ with $t' > t$ and reaches $\sigma' \neq \sigma$ is given by

$$ \exp \left( - \sum_{i} q_{\sigma, \sigma'}(X) |ds| \right) q_{\sigma, \sigma'}(X|\gamma) \, dt, \quad [6] $$

where $X(s)$ is the solution to Eq. 4 for $s \in [t, t']$ with $\sigma(t) = \sigma$ fixed. This reformulation of the process can be used to design straightforward variants of Gillespie’s SSA to simulate it—for this reason we will refer to it as REMD with SSA, or REMD-SSA for short. How to implement REMD-SSA with a finite $\gamma$ is explained in Supporting Information, and how to reach the limit $\gamma \to \infty$ is dealt with below. In both cases, these implementations are rejection-free and akin to kinetic Monte-Carlo schemes. For any swapping rate $\gamma$, the dynamics described in Eqs. 4 and 5 satisfies detailed balance with respect to $Q$ and therefore has Eq. 2 as its equilibrium distribution (see Supporting Information for details). As a result, the expectation of an observable $A$ at any temperature $1/\beta$ can be calculated as

$$ \langle A \rangle_\beta = \int A(x) \rho_\sigma(x) \, dx \quad = \int \sum_{\sigma} \left( \sum_{i=1}^{N} A(x_i) \rho_{1,\sigma}(x_i) \right) q(\sigma, X) \, dX \quad [7] $$

where $1_{\sigma \neq \sigma'} = 1$ if $\sigma \neq \sigma'$ and $1_{\sigma \neq \sigma'} = 0$ otherwise, and similarly for $1_{\sigma = \sigma'}$. Here $\sigma(t) \equiv \sigma$ denotes the index onto which $i$ is mapped at time $t$ by the time-dependent permutation $\sigma(t)$.

Infinite-Swap REMD. It can be shown that the sampling efficiency of REMD-SSA increases monotonically with $\gamma$ (Supporting Information). This suggests one should operate the scheme in the infinite-swap limit $\gamma \to \infty$, and so let us consider briefly this limit next—how to operate within it in practice will be described in the next section. As $\gamma \to \infty$, the permutation $\sigma$ evolves infinitely faster than $X$, meaning that $\sigma$ is always at equilibrium conditional on the current value of $X(t)$, and $X(t)$ only feels the average effect of $\sigma$. In other words, the dynamics of $X$ is captured by the limiting equation

$$ x_i = m^{-1}p_i; \quad p_i = -\beta_\sigma(x_i) \nabla V(x_i) - p_i + 2\gamma m^{-1} \eta. \quad [8] $$

Here

$$ R(\sigma) = \beta^{-1} \sum_{\sigma'} \beta_{\gamma}(\sigma) \rho_{\sigma'}(\sigma); \quad [9] $$

where $\rho_{\sigma'}(\sigma) = 1_{\sigma \neq \sigma'}$ denotes the index mapped onto $\sigma$ by the permutation $\sigma$, is the averaged rescaling parameter of the force, with the average taken with respect to the equilibrium distribution of $\sigma$ given $X$:

$$ \rho_{\sigma}(X) = \sum_{\sigma'} \rho_{\sigma'}(\sigma) = \frac{\rho(\sigma, X)}{\rho(\sigma')} \quad [10] $$

We note that Eq. 8 is exactly the infinite-swap REMD (ISREMD) formulated in ref. 20.

The equilibrium distribution sampled by the limiting equations (Eq. 8) is the mixed distribution $q(\sigma) \rho_\sigma(x)$ in Eq. 1. Therefore, the canonical average of $A$ at $\beta_\gamma$ can be estimated by

$$ \langle A \rangle_{\beta_\gamma} = \int A(x) \rho_{\gamma}(x) \, dx \quad = \int \sum_{\sigma} \left( \sum_{i=1}^{N} A(x_i) \rho_{1,\gamma}(x_i) \right) q(\sigma, X) \, dX \quad [11] $$

where

$$ \rho_{\gamma}(X) = \sum_{\sigma} \rho_{\gamma}(\sigma) \rho_{\sigma}(X) \quad [12] $$

is the probability that the $i$th replica is at the $\gamma$th temperature conditional on the replica positions being at $X$. 

Yu et al.
In ISREMD, the permutation σ is no longer a dynamical variable: We only need to evolve X(t) according to Eq. 8. Unfortunately, this involves calculating the factor R_κ(X) defined in Eq. 9, and the cost of these calculations increases very rapidly with the number of replicas N, because the number of permutations grows as a factorial of N. On top of this, to estimate averages we also need to compute r_κ(X), which is costly too. One way to address this problem is to introduce additional approximations, for example, partial swapping (18, 19). Next, we propose an alternative strategy that permits us to reach the infinite-swap limit while working directly with the original dynamics defined in Eqs. 4 and 5, without introducing additional approximations.

Implementation: REMD with Multiscale SSA. To simulate the process defined by Eqs. 4 and 5 at large ν, that is, in a regime where the evolution of slow variables X is effectively captured by the limiting equation (Eq. 8), we will use a variant of the HMM (22–24). HMM’s basic idea is to compute on-the-fly the coefficients in the limiting equation (Eq. 8) for the slow variables X, as they are needed to evolve these variables, via short bursts of simulation of the fast process — these simulations are performed in what is called the microsolver in HMM, whereas the routine used to evolve the slow variables is referred to as the macrosolver; the scheme is made complete by an estimator specifying how the data generated in the microsolver are used within the macrosolver (i.e., how the unknown coefficients in the limiting equation for X are being computed). In the present context, the HMM scheme will we use works as follows.

1. Microsolver: Evolve r_κ(X) via SSA from t_k to t_k+Δt using the rate in Eq. 5, that is, Set r_κ,0 = r_κ, t_κ = t_k, and for l ≥ 2, do:
   a. Compute the lag to the next reaction via
   \[ t_l = -\ln r / q_{r_\kappa l} \]  \[ \text{[13]} \]
   where r is a random number uniformly picked in the interval (0,1) and \[ q_{r_\kappa l} = \sum_v q_{r_\kappa v} (X_v) \]  \[ \text{[14]} \]
   b. Pick σ_l with probability
   \[ p_{r_\kappa} = q_{r_\kappa l} / q_{r_\kappa 1} \]  \[ \text{[15]} \]
   c. Set t_{k+1} = t_k + Δt and repeat till the first L such that t_{k+1} > t_k + Δt then set t_{k+1} = σ_l = t_k + Δt - t_{k+1}.

2. Estimator: Given the trajectory of σ, estimate \( \hat{r}_\kappa \) (X) and \( \hat{R}_\kappa \) (X) via
   \[ \hat{r}_\kappa (X_k) = 1 / \Delta t \int_{t_k}^{t_{k+1}} 1_{\nu r_\kappa > 0} ds \]  \[ \text{[16]} \]
   and
   \[ \hat{R}_\kappa (X_k) = \beta \sum_l \beta_l / \Delta t \int_{t_k}^{t_{k+1}} 1_{\nu r_\kappa > 0} ds \]  \[ \text{[17]} \]
   iii. Macrosolver: Evolve X_k to X_{k+1} using one time-step of size Δt in the MD integrator for Eq. 4 with \( \hat{R}_\kappa (X) \) replaced by the factor \( \hat{R}_\kappa (X_k) \) calculated in the estimator. Then repeat the three steps above for each time step k.

We will refer to this scheme as REMD with multiscale SSA, or REMD-MSSA for short. It gives an accurate approximation of the solution to the limiting equation (Eq. 8) if \( \nu \gg 1/\Delta t \). Indeed, in this regime the sampled \( \hat{R}_\kappa (X) \) and \( \hat{r}_\kappa (X) \) will be close to their averaged values in Eqs. 9 and 12. We also note that a different implementation of the microsolver, based on a Metropolis-Hastings algorithm can also be used; this implementation is discussed in Supporting Information. It is closer to what is done in standard REMD, except that many swaps are attempted at each MD step instead of one swap every \( \Delta t \) MD steps (i.e., this implementation still permits one to effectively compute in the infinite-swap limit).

Like standard REMD, REMD-MSSA is easily parallelizable. For example, the MD simulations of each replica can be performed on different nodes with an additional node used to evolve σ. At each MD step, the energies \( V(X) \) of the \( N \) replicas need to be communicated to the node evolving σ, and this node sends back to those evolving X, the \( N \) values of \( \hat{R}_\kappa (X) \). The calculation of averages can be handled separately and requires the \( N \) values of \( A(X) \) and the \( N(N-1) \) values of \( \hat{R}_\kappa (X) \).

A version REMD-MSSA has been implemented into Desmond (v3.4.0.2) (25) and used in the numerical examples presented next. To benchmark the method on a simple example, we also tested REMD-MSSA on a 1D asymmetric double-well potential. These results are reported in Figs. S1–S3 and show that REMD-MSSA combined with HMM permits one to reach the infinite-swap limit and is more efficient than standard REMD.

Numerical Results

**AD in Vacuum.** As a first test of REMD-MSSA, we studied block AD (ACE-ALA-CBX) in vacuum under the CHARMM force field with CMAP corrections (26, 27). The free energy profile along the two dihedral angles φ and ψ for this system has been mapped (e.g., in ref. 28). We performed tests of REMD-MSSA using either 4 or 12 replicas, using the values of ν specified in simulation details, and running the simulations for 50 ns. We compared these results with those from an ISREMD simulation of 50 ns with four replicas (in which there are only 24 permutations) using the analytic weights in ref. 10—such a calculation is ISREMD is not feasible with 12 replicas because the number of permutations (≈4.8 × 10^8) becomes too large to manage. Finally, we compared the results with those from a standard REMD with 12 replicas and swapping rate of 0.5 ps⁻¹.

The reconstructed free energy profiles at 300 K are shown in Fig. 4. The profiles calculated using REMD-MSSA with 4 and 12 replicas (Fig. 4A, Left, Upper and Lower) and ISREMD with four replicas (Fig. 4A, Upper Right) are in close agreement. These profiles also match those obtained with standard REMD (Fig. 4A, Lower Right) and they are in close agreement with those estimated by other enhanced sampling methods under the same force field, including those calculated with 100 ns of metadynamics and 250-ns accelerated MD (28). In particular, we correctly identified the four known local minima (using the same terminology as in Mackerell et al. (27)): C5 located at (−154,160), C7eq at (−81,66), C7α at (80,−63), and αR at (−99,12).

To assess the sampling efficiency of our scheme, we first looked at the times at which the temperatures of each replica take to make a round trip between their highest and lowest values; equivalently, this amounts to monitoring the values of σ(i,t) versus time for each replica index i. It has been argued (16, 29–31) that the shorter these round-trip times, the faster the scheme will converge. A similar conclusion was reached in ref. 17, using a quantity closely related to round-trip times: the lifetime, which gives the average time a temperature stays at a given value before being updated to another. In the context of ISREMD, the round-trip times (and the lifetime) should be zero because this method operates in the infinite-swap limit: This means that we can use this test to quantify how close REMD-MSSA is from this limit. In Fig. 1B we show a typical trajectory σ(i,t) from REMD-MSSA and one from standard REMD (with swap rate at 0.5 ps⁻¹) for the case of 12 replicas (the full set of trajectories can be found in Fig. S4). It can be seen that the round-trip times are indeed much shorter in REMD-MSSA than in standard REMD. Similarly, the lifetime in REMD-SSA is roughly 10 as compared with 105 ps for standard REMD. Another more stringent measure of efficiency is the flatness of the temperature distributions at each replica, or equivalently that of σ(i,t) for each replica index i. Indeed, it is known (30, 31) that these distributions should be uniform at equilibrium (i.e., for each i σ(i,t) should spend the same proportion of time at every index value j). Unlike the round-trip times, which can be short even if the replicas have not equilibrated yet, these distributions can only be flat if the replicas are themselves equilibrated.
simulation. These traces indicate that our REMD-MSSA simulation experienced several folding/unfolding events within 50 ns. This is a significant speed-up compared with a bare MD simulation, in which folding events are expected to take place every 500 ns to 1 ps (38). We used the REMD-MSSA simulation data to generate FES along the two order parameters, β-strand H-bonds N_H and backbone radius of gyration RG, from a 100-ns REMD-MSSA run (Fig. 24 and Fig. S7). This FES captures the main features present in FES obtained from previous longer REMD simulations (32, 34, 39, 40). Specifically, we observe basins corresponding to conformations that form no β-strand H-bonds, form one or two H-bonds as partially folded β-sheet, and fully β-sheet with more than three H-bonds. We also compare this FES with the one from a 120-ns standard REMD starting from the same initial structures (Fig. 2B). It can be seen that the folded basins are populated in REMD-MSSA whereas only a few samples are seen in standard REMD. The FES from REMD-MSSA is in better agreement with previous longer simulation studies, indicating that REMD-MSSA can give a more converged FES than standard REMD within a 100-ns run due to higher sampling efficiency.

To test convergence, in Fig. 2 C and D we show the trajectories and the distributions of σ(δi,t) for one representative replica, for both REMD-MSSA and standard REMD (more representative trajectories and distributions are shown in Figs. S8 and S9). The round-trip times and the temperature distribution from REMD-MSSA is almost uniform and significantly lower than those in REMD with swap rate 1 ps⁻¹, with values similar to those reported for AD, and the temperature distributions are also a significantly flatter in REMD-MSSA than in REMD. To measure how flat these distributions are across all of the replicas, we calculated the relative entropy of each, that is, \( \Delta S(p_i) = \sum p_i(j) \log p_i(j)/\tau_{ref}(j) \), where \( p_i(j) \) is the distribution \( \sigma(\delta_i,t) \) and \( \tau_{ref}(j) = 1/60 \) is the target uniform distribution: If \( p_i(j) = \tau_{ref}(j) \), \( S(p_i) = 0 \), otherwise it is positive, and the higher the value of \( S(p_i) = 0 \), the less flat \( p_i(j) \) is. These relative entropies are shown after ordering from smallest to largest in Fig. 2E: They are significantly smaller for REMD-MSSA than for standard REMD, indicating of faster mixing by the former than by the latter.

Remarkably, careful investigation of the FES in Fig. 2 leads us to find a folded state that is sampled in REMD-MSSA but not in standard REMD. For reference, the native state’s structure and H-bond registry are shown in Fig. 3A; the native state is characterized by the β-strand H-bond pairs (E14,T11), (T12,T3), and (D10,T5). We detect three metastable states whose representative structures are depicted in Fig. 3 as insets and labeled as \( \beta_N, \beta_D \), and \( \beta_M \). The representative backbone H-bond contacts of these three states are shown in Fig. 3B. From these H-bond registries, we recognize \( \beta_N \) as the native state, whereas \( \beta_D \) represents a distinct, nonnative β-hairpin state that is not sampled in our 100 ns of standard REMD, and \( \beta_M \) a misfolded state. Fig. 3C shows the FES at 270 K in the space of RG and \( N_H \) with states \( \beta_N, \beta_D, \) and \( \beta_M \) indicated. In this space, the three states are not well-distinguished, indicating that the (RG, \( N_H \)) space is likely not optimal for understanding the conformational distribution of the folded states. We show in Fig. 3D the FES at 270 K in the space of two new collective variables, \( d_{CMAP}^N \) and \( d_{CMAP}^D \), respectively, built from all samples in the REMD-MSSA at 270 K for which \( N_H > 2 \). Here, \( d_{CMAP}^N \) measures the \( L_1 \) distance from state \( \beta_N \) and \( d_{CMAP}^D \) from state \( \beta_D \) (see Supporting Information for their definition). In this space, the three states are easily distinguishable. To the best of our knowledge, the β-sheet state we observe here (\( \beta_D \)) has never been detected in any other simulations of this peptide. To verify that \( \beta_D \) is metastable, we launched a 220-ns standard MD simulation at 270 K from a \( \beta_D \) sample. The state remained stable for the duration of that simulation, as can be seen from the driftless traces of several collective variables that feature the conformational structure (Fig. 3E and Fig. S10). We also calculated the average potential energy difference between \( \beta_D \) and \( \beta_N \) to be 1.4 ± 0.7 kcal/mol, which indicates they are energetically indistinguishable.

Even though state \( \beta_D \) looks similar to \( \beta_N \), its H-bond registry is different, as evidenced by the O–H contact maps in Fig. 3B. Their distinction is revealed in the \( (d_{CMAP}^N, d_{CMAP}^D) \) space. To access \( \beta_D \) from

![Fig. 1](image-url). (A) Free energy surfaces along the two dihedral angles \( \phi \) and \( \psi \) for AD in vacuum at 300 K. (Upper Left) Fifty-nanosecond REMD-MSSA simulation with four replicas. (Upper Right) Fifty-nanosecond ISREMD simulation with four replicas using the analytical weights calculated from Eq. 10. (Lower Left) Fifty-nanosecond REMD-MSSA simulation with 12 replicas. (Lower Right) Fifty-nanosecond standard REMD simulation with 12 replicas and a swapping rate of 0.5 ps⁻¹. All plots have 60 × 60 bins and filtered by standard Gaussian kernel. Same level sets are used in the contour plots. (B) Trajectories \( \sigma(\delta_i,t) \) of one replica. (C) Distributions of \( \sigma(\delta_i,t) \) for the same replica after 0.8 ns of simulation. In B and C, the orange curve is from REMD-MSSA and the blue one is from standard REMD, both with 12 replicas.

at equilibrium. A representative distribution from REMD-MSSA and one from standard REMD, both calculated via time averaging of \( \sigma(\delta_i,t) \) over 0.8 ns of simulations, are shown in Fig. 1C (the full set of distributions can also be found in Fig. S5). As can be seen, the distribution from REMD-MSSA is almost uniform and significantly flatter than that from standard REMD, indicating that the former has converged after 0.8 ns, whereas the latter has not.

Folding of Protein G β-Hairpin in Explicit Solvent. As a second test we considered a β-hairpin peptide in explicit water, whose folding behavior resembles that of larger protein, and which has been investigated by many techniques, including standard REMD (32–37). Specifically, we studied the C-terminal fragment of the Ig binding domain B1 of protein G [Protein Data Bank (PDB) ID code 2gb1]. This capped peptide sequence contains 16 residues, 1-Asp-GEWYDDATKTFVTENM-NMe-16, with 256 atoms. This β-hairpin is known as a hard-to-fold protein. In our REMD-MSSA simulations, the initial structures for the replicas are chosen from an unfolded ensemble that is generated from a high-temperature MD run. We then run REMD-MSSA and check how fast folded state emerges and reasonable free energy profiles in various order parameters are obtained. In Fig. S6, we show traces of two order parameters, namely the number of β-strand H-bonds \( N_H \) (defined in Eq. S18) and the rmsd of the \( C_{\alpha} \) s from the native structure (PDB ID code 2gb1), obtained by monitoring the evolution for a few representative replicas in the REMD-MSSA
\[ \beta_N \text{, the amino-terminal strand must flip 180° so that its outside O} \]

and H can point inside to form a hydrogen bond with H and O on the carboxyl-terminal strand (C-strand). This implies that either state must unfold first to break all H-bonds to get a chance to fold into the other. We indeed observe that \( \beta_2 \) is populated predominantly by transitions from the unfolded pool rather than from the \( \beta_N \) pool. Another difference between the two folded states is the turn structure. \( \beta_N \) has a large \( \pi \)-turn (41) composed of D10, D9, A8, T7, and K6, whereas \( \beta_2 \) has a small \( \gamma \)-turn (42) only involving D9, A8, and T7. Previous studies proposed folding of this \( \beta \)-hairpin uses either a “zipper” mechanism (36, 43–46) where hydrogen bonds form sequentially from the turn, or a “hydrophobic collapse” in which the folded state arises from a collapsed globule (32, 33, 35, 39, 47, 48). It is not clear from our simulation results which of these is preferred, but the \( \gamma \)-turn of \( \beta_2 \) being so tight, may be unlikely to form before a few H-bonds first lock in a registry between the two strands. We therefore suggest a more detailed mechanistic study based on string method (49, 50) or transition path sampling (51) in the future. In addition to two folded \( \beta \)-sheet states, we observed one extremely stable misfolded state, labeled M. Its contact map and conformation are shown in Fig. 3. This misfolded state is stabilized by three \( \beta \)-strand hydrogen bonds and two salt bridges between the K6 ammonium group and the carboxylates on D10 and D9. Neither folded state displays any such side-chain salt bridging, and we therefore suggest that state M serves as a kinetic trap that slows the acquisition of authentic \( \beta \)-sheet structure, either in the native or secondary conformation. The picture that arises in light of these results is of a funnel-like landscape for the folding thermodynamics of this \( \beta \)-hairpin that has at least two distinct and deep minima and a misfolded kinetic trap.

**Concluding Remarks**

We have proposed a multiscale variant of REMD in which the swaps are performed via SSA, which effectively permits one to reach the infinite-swap limit known to optimize sampling efficiency. This REMD-MSSA is as simple to implement and parallelize as standard REMD and can be used with any set of temperatures, including sophisticated choices like those advocated, for example, in ref. 29 that would most likely improve the efficiency even more. It can also be used with parameters other than temperature (Figs. S11 and S12). Here its usefulness and efficiency were benchmarked on test cases of various complexity. In particular, in the context of folding the G \( \beta \)-hairpin, our REMD-MSSA method was shown to outperform standard REMD and other enhanced sampling techniques, as evidenced by the fact that it allowed us to identify a folded \( \beta \) structure previously undetected by these other methods. These results indicate that REMD-MSSA can be an efficient and accurate alternative to existing enhanced sampling methods and should be useful in a broad variety of MD applications.

**Simulation Details**

**AB.** We used REMD-MSSA runs with temperatures in geometric progression between 300 K and 1,200 K; other, perhaps more efficient, choices of temperatures could be used as well (29), but this one proved sufficient to our purpose. We used a time step of \( 2 \) fs for the Langevin dynamics and chose \( \gamma = 5 \) ps\(^{-1}\). The bond lengths between hydrogen and heavy atoms were kept constant via M-SHAKE (52). We choose \( \nu \) to obtain about \( 10^4 \) jumps of \( \sigma \) on average between two consecutive MD steps. This led to using \( \nu = 10000/\Delta \tau \) with 4 replicas and \( \nu = 500/\Delta \tau \) with 12 replicas\( \sigma \) must be larger with 4 rather than 12 replicas because the energy gaps between the replicas are larger in the former case, implying that the jump rate in Eq. 5 is smaller. In the
We solvated the protein with 1,549 water molecules, neutralized with three Na\(^+\) ions, resulting in a total system size of 4,906 atoms and box dimensions of 36.3 × 36.3 × 36.3 Å\(^3\). OPLS- AA is the force field for protein (53) and TIP3P for water (54). Long-range electrostatics are handled via particle-mesh Ewald summation with a 64 × 64 × 64 mesh. The cutoff for short-range nonbonded interactions was set to 12 Å. The integration time step is 2 fs with the bonds between hydrogen and heavy atoms constrained by the LINCS (52). The system is equilibrated at 270 K and 1 atm via 1-nS NVT Langevin dynamics then a 50-ns NVT simulation. REMD-Mssa simulations have 60 replicas with temperatures ranging from 270 K to 690 K in a geometric progression (55)—here, too, other choices of temperatures could be used (29). The initial structures for these replicas were generated from the 4-ns NVT simulation at 700 K. We used a friction coefficient of 5 ps\(^{-1}\) in the Langevin thermostat and choose \(\nu = 3/\Delta t\) for each replica under the same temperature setting we use random-neighbour swapping at the rate of 1 ps\(^{-1}\) and the same MD parameters as above. All MD simulations were carried out with Desmond (v3.4.0.2) (25).

\[\text{We solvated the protein with 1,549 water molecules, neutralized with three Na}\(^+\) ions, resulting in a total system size of 4,906 atoms and box dimensions of 36.3 } \times {36.3} } \times {36.3} } \text{Å}^3.\text{OPLS-AA is the force field for protein (53) and TIP3P for water (54). Long-range electrostatics are handled via particle-mesh Ewald summation with a 64 } \times {64} } \times {64} } \text{ mesh. The cutoff for short-range nonbonded interactions was set to 12 Å. The integration time step is 2 fs with the bonds between hydrogen and heavy atoms constrained by the LINCS (52). The system is equilibrated at 270 K and 1 atm via 1-nS NVT Langevin dynamics then a 50-ns NVT simulation. REMD-Mssa simulations have 60 replicas with temperatures ranging from 270 K to 690 K in a geometric progression (55)—here, too, other choices of temperatures could be used (29). The initial structures for these replicas were generated from the 4-ns NVT simulation at 700 K. We used a friction coefficient of 5 ps }^{-1}\text{ in the Langevin thermostat and choose } \nu = 3/\Delta t\text{ for each replica under the same temperature setting we use random-neighbour swapping at the rate of 1 ps }^{-1}\text{ and the same MD parameters as above. All MD simulations were carried out with Desmond (v3.4.0.2) (25).}\]