Structure optimization and folding mechanisms of off-lattice protein models using statistical temperature molecular dynamics simulation: Statistical temperature annealing

Jaegil Kim,* John E. Straub, and Thomas Keyes

Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA

(Received 11 January 2007; revised manuscript received 28 March 2007; published 18 July 2007)

The recently proposed statistical temperature molecular dynamics (STMD) algorithm [Kim *et al.*, Phys. Rev. Lett. **97**, 050601 (2006)] is used as the core of an optimization algorithm, statistical temperature annealing (STA), for finding low-lying energy minima of complex potential energy landscapes. Since STMD realizes a random walk in energy, the idea is simply to initiate repeated minimizations from configurations in the low-energy segments of STMD trajectories. STA is tested in structural optimization of various off-lattice *AB* and extended *AB* protein models in two and three dimensions with different chain lengths. New putative ground states were found for the two- and three-dimensional *AB* 55-mer, and for the three-dimensional extended *AB* 21-mer and 55-mer. The distinct folding features of the models are analyzed in terms of the statistical temperature and other representations of the structure of the potential energy landscape. It is shown that the characteristic behavior of the statistical temperature undergoes a qualitative change with the inclusion of a torsional potential in the extended *AB* model, as the more rigid backbone makes the potential energy landscape more funnel-like.

DOI: 10.1103/PhysRevE.76.011913 PACS number(s): 87.15.−v, 87.10.+e, 05.70.−a, 36.20.−r

I. INTRODUCTION

The prediction of a protein structure from its primary amino acid sequence is one of the most challenging problems in biological and computational science. Based on Anfinsen's thermodynamic hypothesis [1], which states that the native structure of a protein corresponds to the global minimum of the free energy surface of protein plus solvent, the structure prediction problem is translated into a global optimization problem. For the case of an implicit solvent model, one may use the potential energy as the target energy function, recognizing that the native state must now be identified as a group of configurations sharing the structural motif of the global minimum. Thus folding and optimizing are not identical, but remain closely related.

However, the intrinsic complexity of the potential energy landscape of a protein, with a multitude of local minima separated by high-energy barriers, makes it difficult to find the global minimum, or ground state, within a reasonable computational time. Even in a highly simplified *HP* lattice model [2] consisting of hydrophobic (*H*) and polar (*P*) amino acids, the number of conformations increases enormously as the chain length grows, and finding the global minimum is nontrivial [3]

To overcome the multiple minimum problem, and to enhance computational efficiency in the search for the global minimum, several sophisticated algorithms have been developed such as simulated annealing (SA) [4], basin hopping [5], energy landscape paving (ELP) [6], conformational space annealing (CSA) [7], metadynamics [8], conformational flooding [9], and various generalized ensemble techniques [10–13]. Despite substantial differences in detail, these methods are basically Monte Carlo (MC) algorithms, except for metadynamics and conformational flooding. Combined with clever, nonphysical trial moves, they have been successful in finding global minima in both lattice and off-lattice protein systems. However, as the system size increases, the design of effective MC trial moves becomes harder, due to steric hinderance in compact conformations. This difficulty is more serious in an explicit solvent, due to the protein-water interactions. Therefore, it is a significant challenge to develop an efficient optimization algorithm that retains the merit of advanced sampling techniques and does not suffer from the difficulty of designing trial moves in condensed phases.

Recently, we proposed the "statistical temperature molecular dynamics" (STMD) algorithm, which combines ingredients of multicanonical molecular dynamics [14] and Wang-Landau (WL) sampling [15] through the concept of the statistical temperature. The basic strategy, employing a sampling weight inversely proportional to the density of states, is similar to multicanonical sampling [16] or the annealing contour MC [17] (ACMC) method. However, STMD is distinguished from optimization algorithms adapting generalized ensemble techniques or WL sampling in that it uses a dynamic modification scheme for the statistical temperature and does not require a histogram accumulation, which greatly accelerates a conformational search through the self-adjusting sampling weight.

Since STMD has been designed to perform a random walk in energy without trapping in local minima, and, being a molecular dynamics algorithm, generates natural collective motions of the particles, it can be a powerful optimization tool for biomolecules. Here we present a STMD-based optimization algorithm, "statistical temperature annealing" (STA). The performance of STA is tested on implicit solvent model proteins, the off-lattice *AB* protein model [18] in two and three dimensions, and the extended *AB* model, which includes a torsional potential [19] and is inherently three dimensional. STA reproduces previously known global minima and finds further putative ground states in several cases.

*jaegil@bu.edu

The folding processes of *AB* and extended *AB* model proteins are investigated in terms of their statistical temperatures. It is also shown that inclusion of a torsional potential makes the energy landscape more funnel-like in the extended *AB* model. Our study reveals that the folding of *AB* proteins proceeds by a sharp nonspecific collapse, followed by slow relaxation into the global minimum in glassy energy landscapes, arising from the energetic dominance of nonbonded interactions over local interactions in collapsed states. On the other hand, the torsional potential in the extended *AB* model prevents an initial rapid collapse and allows a progressive folding into the ground state, by increasing chain stiffness and reducing the accessible conformational space.

II. STATISTICAL TEMPERATURE MOLECULAR DYNAMICS ALGORITHM

The formulation of STMD relies [21] on the well-known thermodynamic relationship between the microcanonical entropy $S(E)$ and the statistical temperature $T(E)$ [29],

$$T(E) = [\partial S(E)/\partial E]^{-1}, \tag{1}$$

where $S(E)=\ln\Omega(E)$, in units such that $k_B=1$, and $\Omega(E)$ is the density of states. From the one-to-one correspondence in Eq. (1), STMD achieves a flat energy distribution via the systematic refinement of the statistical temperature estimate $\widetilde{T}(E)$,

$$\widetilde{T}'_{j\pm1} = \frac{\widetilde{T}_{j\pm1}}{1 \mp \delta f \widetilde{T}_{j\pm1}}, \tag{2}$$

where j represents the index for the energy grid defined as $E_j=G(E/\Delta)\Delta$, with bin size Δ and $G(x)$ returning the nearest integer to x, $\delta f=\ln f/(2\Delta)\ll 1$, and $f\geq 1$ is the temperature modification factor. It should be emphasized that the statistical temperature $\widetilde{T}(E)$ in Eq. (2) is dynamically modified every time the system visits the energy state E_j, and this process accelerates the determination of the entropy estimate $\widetilde{S}(E)=\int^E 1/\widetilde{T}(E')dE'$ compared to multicanonical sampling requiring histogram accumulation.

With the dynamic update scheme of Eq. (2) STMD achieves a flat energy distribution by transforming an initially constant $\widetilde{T}(E)$ to the true statistical temperature $T(E)$ through iterative refinements. The modification factor is reduced, $f\rightarrow\sqrt{f}$, when the energy histogram reaches a specified flatness; see Ref. [21] for details.

The primary advantage of STMD is its implementation as a molecular dynamics simulation, which integrates equations of motion of the generalized ensemble sampling [14,23] coupled to a Nosé-Hoover thermostat [22]. By considering the generalized ensemble sampling characterized by the sampling weight $w(E)=\exp[-\widetilde{S}(E)]=\exp[-\beta_0 v_{eff}(E)]$, $\beta_0=1/k_BT_0$, as the canonical ensemble sampling associated with the effective potential $v_{eff}=T_0\widetilde{S}(E)$ at the fixed kinetic temperature T_0, the equations of motion are obtained as

$$\dot{\mathbf{q}}_i = \mathbf{p}_i,$$

$$\dot{\mathbf{p}}_i = -\nabla_{\mathbf{q}_i} v_{eff}(E) - \xi\mathbf{p}_i = \gamma(E)\mathbf{f}_i - \xi\mathbf{p}_i,$$

$$\dot{\xi} = [K(\mathbf{p}_i) - N_f T_0]/Q, \tag{3}$$

where $K(\mathbf{p}_i)=\Sigma_i \mathbf{p}_i^2/2$, $\gamma(E)=T_0/\widetilde{T}(E)$, and $\mathbf{q}_i$, $\mathbf{p}_i$, and $\mathbf{f}_i$ correspond to the coordinate, the momentum, and the force of ith particle, respectively. Here, ξ and Q represent the conjugate momentum and fictional mass of the thermostat, determining the strength of thermal coupling to a system having N_f degrees of freedom.

Equation (3) corresponds to an ordinary molecular dynamics simulation combined with the energy-dependent force scaling, and the average kinetic energy is maintained at the fixed temperature T_0. The velocity-Verlet integration algorithm with the fixed force scaling factor $\gamma(E)=T_0/\widetilde{T}(E)$ produces a configurational sampling with the weight $w(E)=e^{-\widetilde{S}(E)}$, and determines the probability density function

$$P(E) \sim e^{S(E)-\widetilde{S}(E)} = \exp\left(\int^E \delta\widetilde{\beta}(E')dE'\right), \tag{4}$$

where $\delta\widetilde{\beta}(E)=[\widetilde{T}(E)-T(E)]/(\widetilde{T}(E)T(E))$.

The dynamical driving force of STMD, pushing the system to escape from a confined energy region and continue a random walk in energy, is the systematic bias in $\delta\widetilde{\beta}(E)$. When the system gets trapped in some energy region E_j, the accumulated operations of Eq. (2) upon the repeated visits to E_j produce statistical temperature gradients of $\delta\widetilde{\beta}_{j-1}<0$ and $\delta\widetilde{\beta}_{j+1}>0$, and create a concave local curvature in $P'(E)=dP(E)/dE=P(E)\delta\widetilde{\beta}(E)$ i.e., $P'(E_{j-1})<0$ and $P'(E_{j+1})>0$, which generates an outgoing probability flux from E_j and assists the system to escape. We emphasize that the fluctuations in $\delta\widetilde{\beta}(E)$ are self-adjusting, allowing the system to constantly wander through the whole energy space with a modification factor $f>1$. As the modification factor f approaches unity and $\widetilde{T}(E)$ is refined to $T(E)$ with the vanishing δf, the system reaches true thermodynamic equilibrium and attains a flat energy distribution in energy without any further update of $\widetilde{T}(E)$.

STA optimization protocols are outlined as follows. (i) Determine a sampling region by selecting low- and high-temperature limits T_l and T_h, respectively, and choose Δ, T_0 (Usually T_h), and the initial modification factor δf. (ii) Run STMD, applying the dynamic operation of Eq. (2) every time step, starting from $\widetilde{T}(E)=T_h$, with the low-energy flattening $\widetilde{T}(E)=T_{\min}$ for $E<E_{\min}$, $T_{\min}=\widetilde{T}(E_{\min})=\min\{\widetilde{T}(E)\}$. Low-energy flattening assists the system to access unexplored lower-energy regions more quickly through canonical sampling at $T_{\min}$. (iii) Repeat step (ii) with a reduced $f\rightarrow\sqrt{f}$ once the histogram fluctuations are less than 20% of the mean. (iv) Calculate low-lying local minima, also called inherent structures (IS) [24], and possibly the global minimum, by systematically quenching low-energy configurations.

Finally, note the difference in the application of STMD to optimization vs equilibrium sampling. Originally, STMD was designed to facilitate equilibrium sampling by generat-

TABLE I. Lowest-lying minima of the 2D *AB* model determined by high-temperature Monte Carlo (HTMC) simulation [18], improved pruned-enriched-Rosenbluth method with importance sampling (nPERMis) [20], annealing contour Monte Carlo (ACMC) simulation [17], conformational space annealing (CSA) [7], and statistical temperature annealing (STA).

Sequence	HTMC	nPERMis	ACMC	CSA	STA
$S_{13,2D}$	−3.2235	−3.2939	−3.2941	−3.2941	−8.2941
$S_{21,2D}$	−5.2881	−6.1976	−6.1979	−6.1980	−6.1980
$S_{34,2D}$	−8.9749	−8.9749	−10.7001	−10.8060	−10.8060
$S_{55,2D}$	−14.4089	−18.5154	−18.7407	−18.9110	−18.9202

ing a flat energy distribution, overcoming energy barriers, which is achieved in the asymptotic limit of $\delta f \to 0$. On the other hand, optimization is based on the comprehensive sampling of low-energy configurations, which does not require vanishing δf. In some cases, we continued STMD simulations with a fixed δf to accelerate the search of the whole configuration space.

III. STATISTICAL TEMPERATURE ANNEALING

A. *AB* model

We first consider STA optimization in the *AB* protein model, an off-lattice version of the *HP* model consisting of hydrophobic (A) and hydrophilic (B) monomers, which has been studied by various existing algorithms. The potential energy of an N-monomer chain is a sum of bending and nonbonded energies,

$$E_I = E_{bend} + E_{nonb} = \sum_{i=1}^{N-2} \frac{1}{4}(1 - \cos\,\theta_i) + \sum_{j>i+1}^{N} 4(r_{ij}^{-12} - C_{\sigma_i\sigma_j} r_{ij}^{-6}), \qquad (5)$$

where θ_i is the angle between the ith and $(i+1)$th bonds with length unity, and r_{ij} is the distance between monomers i and j; $C_{\sigma_i\sigma_j}$ is $+1$, $+1/2$, and $-1/2$, respectively, for AA, and BB, and AB pairs. We studied four Fibonacci sequences, S_{13} $=ABBABBABABBAB$, $S_{21}=BABABBAB^{*}S_{13}$, $S_{34}=S_{21}*S_{13}$, and $S_{55}=S_{34}*S_{21}$, where the asterisk indicates concatenation. To handle a fixed bond length the SHAKE/RATTLE algorithm has been applied with the velocity-Verlet integrator.

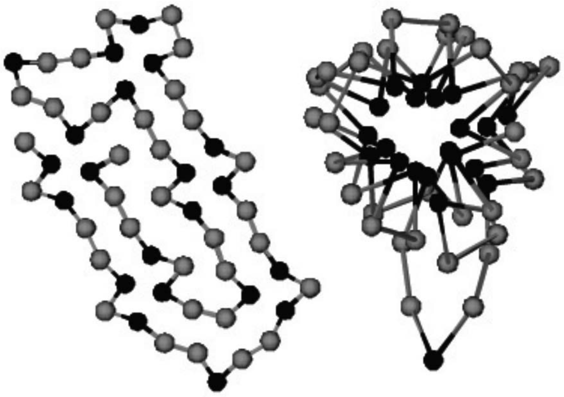

FIG. 1. Lowest-energy structures of $S_{55,2D}$ (left) and $S_{55,3D}$ (right) for *AB* model. Gray and black balls correspond to hydrophilic and hydrophobic monomers, respectively.

Natural dimensionless units are used throughout the paper. Simulation parameters are a time step of 0.0025, $\Delta=1$, $T_l=0.05$, $T_h=T_0=1.0$, and the initial $f=1.0005$. Inherent structures [24] are determined by quenching low-lying STMD configurations, using the conjugate-gradient algorithm with bond constraints [25].

Tables I and II compare STA results for the lowest-energy states of the *AB* model with those of other optimization algorithms in two and three dimensions. STA reproduces the minima found by CSA in both dimensions for the shorter sequences $N=11$, 21, and 34, which correspond to the lowest-energy values in the literature. On the other hand, for the longer chain $S_{55,2D}$ and $S_{55,3D}$, STA finds a lower minimum than the putative ground state values of CSA in two dimensions and ELP in three dimensions, and is thus more likely to find the global minimum in a more complex potential energy landscape. Here, the subscripts n and m in $S_{n,m}$ indicate the sequence and the dimension, respectively. The lowest-energy structure of $S_{55,2D}$ in Fig. 1 shows a similar geometrical pattern, but a different clustering of hydrophobic monomers, compared to that of CSA. The structural difference of the STA lowest minimum from that of ELP in $S_{55,3D}$ is remarkable, with one hydrophobic monomer separated from a tubelike hydrophobic core at the center. We certainly do not claim that this is the global minimum.

As confirmed in Tables I and II, the comprehensive sampling of low-lying configurations with STMD and the subse-

TABLE II. Lowest-lying minima of the 3D *AB* model determined by nPERMis [20], multicanonical Monte Carlo sampling (MUCA) [16], energy landscape paving (ELP) [16], CSA [7], and STA. The energy values in the last column (STMD) correspond to the lowest-energy values sampled by STMD simulations restricted to low-temperature regions with $T_l=0.01$ and $T_h=T_0=0.2$.

Sequence	nPERMis	MUCA	ELP	CSA	STA	STMD
$S_{13,3D}$	−4.9616	−4.967	−4.967	−4.9746	−4.9746	−4.9667
$S_{21,3D}$	−11.5238	−12.296	−12.316	−12.3266	−12.3266	−12.3176
$S_{34,3D}$	−21.5678	−25.321	−25.476	−25.5113	−25.5113	−25.4932
$S_{55,3D}$	−32.8843	−41.502	−42.428	−42.3418	−42.5781	−42.4503

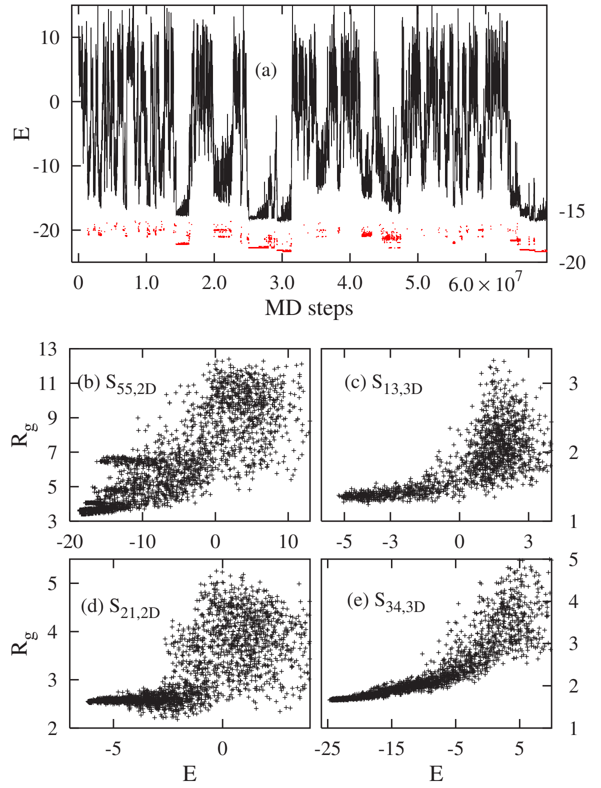

FIG. 2. (Color online) (a) Time series of STMD energies, and inherent structure energies E_{is} [gray (red) dots, shifted by -3 for visualization], determined by quenching low-lying configurations of $S_{55,2D}$ with $E<-16$, and scatter plots of simulated configurations in (E,R_g) for (b) $S_{55,2D}$, (c) $S_{13,3D}$, (d) $S_{21,2D}$, and (e) $S_{34,3D}$ in the *AB* model.

quent local minimization is very promising to locate a global minimum in complex energy landscapes. This strategy of combining enhanced sampling with a local minimization is particularly effective when the basin attraction of the global minimum is very narrow, and is extensively used in other optimization algorithms, such as basin hopping [5], conformational space annealing [7], and the improved pruned-enriched-Rosenbluth method with importance sampling (nPERMis) [20]. Furthermore, it should be stressed that the extra aiding step of a local minimization is not computationally demanding at all in STA, since only low-lying STMD configurations, which are very close to each local minimum, are minimized.

We performed additional STMD simulations for the three-dimensional (3D) *AB* model in which the system is strongly restricted to low temperature, T_l=0.01 and T_h=T_0=0.2, to demonstrate the ability of STMD in sampling low energies, with no minimization. The lowest energies found (see the column of STMD in Table II) are almost the same for short chains of $S_{13,3D}$ and $S_{21,3D}$, and even smaller for longer chains of $S_{34,3D}$ and $S_{55,3D}$, compared to the previous best values of MUCA and ELP in Table II.

The characteristic behavior of STMD in the search for the global minimum is demonstrated in Fig. 2(a), the time series of energies for $S_{55,2D}$. From a randomly chosen high-energy configuration, the simulation begins to sample an unexplored low-energy region with a typical random walk in energy and suddenly gets trapped in one of the low-lying basins of attraction at 1.5×10^7 MD steps. However, with the accumulated operations of Eq. (2) creating a bias in $\tilde{T}(E)$, the simulation soon escapes and continues a random walk. This systematic process is repeated until the system reaches the global minimum for the first time after 3×10^7 MD steps, with a second visit at 7×10^7 MD steps.

The time series of inherent structure energies for $S_{55,2D}$, represented by dots in Fig. 2(a), shows glassy dynamics with several intermittent trappings in different low-lying IS. The corresponding rough structure of the energy landscape is illustrated by the scatter plot of the sampled configurations in (E,R_g) in Fig. 2(b), R_g being the radius of gyration representing the compactness of the chain. The scatter plot contains purely static information, but provides a strong intuitive connection to dynamics. Clearly, the folding of $S_{55,2D}$ proceeds by a collapse and a subsequent rearrangement to several structurally dissimilar compact states. The densely populated branches in the low-energy region with $E<-5$ in Fig. 2(b) represent kinetic traps, which hamper the search for the ground state and cause a slowing down of folding.

All the scatter plots in Fig. 2 are in accord with the widely accepted two-step folding process, in which a fast nonspecific collapse [26] is followed by a slow relaxation to the native state. The R_g values are highly dispersed in the unfolded high-energy regions due to a large flexibility of the chain, but are almost frozen after collapse. This folding behavior of the *AB* model is directly reflected in the profile of the temperature estimate. For $S_{55,2D}$ in Fig. 3(a), the temperature estimate $\tilde{T}_i(E)$ at the *i*th iteration step shows an initial sharp drop and a very long tail for a broad low-energy region.

The sharp drop of $\tilde{T}(E)>0.2$ is associated with the sampling of unfolded, high-energy states with a large value of R_g, as shown in Fig. 3(b), in which the scatter plot in (E,R_g) has been divided into three separate domains depending on the values of $\tilde{T}(E)$. The long tail region with $0.07<\tilde{T}(E)<0.2$ corresponds to a nonspecific collapse of the chain to several compact states, which is followed by a freezing into the glassy regime for $\tilde{T}(E)<0.07$, characterized by structurally dissimilar IS in Fig. 2(a). Furthermore, as demonstrated in the inset of Fig. 3(a), long-lived, intermittent trappings in different low-lying IS for $\tilde{T}(E)<0.07$ cause a considerable fluctuation in the statistical temperature estimate near the global minimum as a function of the iteration. On the other hand, $\tilde{T}(E)$ shows a smooth variation with a good convergence at both intermediate- and high-energy regions.

The initial sharp drop and subsequent slow decay of $\tilde{T}(E)$ appear to be generic features of Fibonacci sequence *AB* proteins in both dimensions, as seen in Fig. 4(a), in which the temperature estimates for $S_{13,3D}$, $S_{21,2D}$, and $S_{34,2D}$ are fully convergent up to $f_d=f-1=10^{-8}$ ($f_d \approx \delta f \Delta$), and the temperature estimates of $S_{34,3D}$ and $S_{55,3D}$ are optimized with f_d =0.000 016. The characteristic broadening of $\tilde{T}(E)$ in the collapse region is due to the dominance of E_{nonb} over E_{bend} in Eq. (5) for compact states. As seen in the decomposed energy

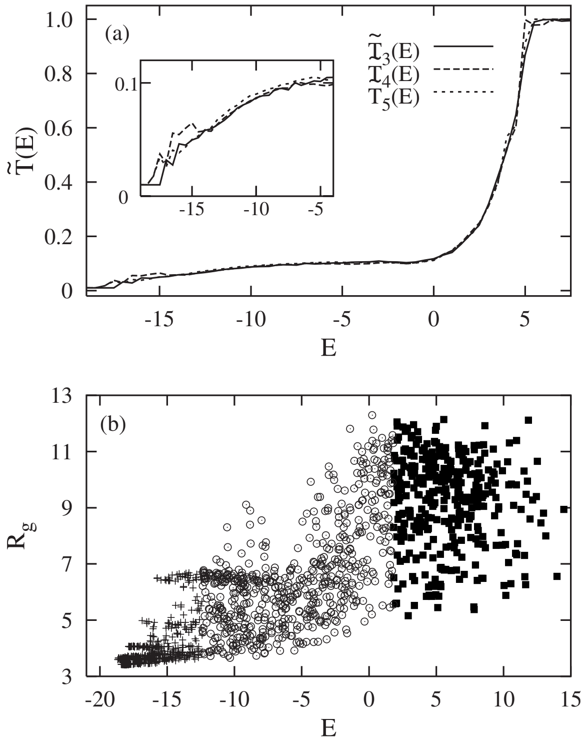

FIG. 3. (a) Statistical temperature estimate $\tilde{T}_i(E)$ at ith iteration step, and (b) scatter plots in (E,R_g) with points labeled by the values of $\tilde{T}(E)$ in $S_{55,2D}$ of AB model; filled squares, open circles, and crosses correspond to configurations with $\tilde{T}(E)>0.2$, $0.07<\tilde{T}(E)<0.2$, and $\tilde{T}(E)<0.07$, respectively.

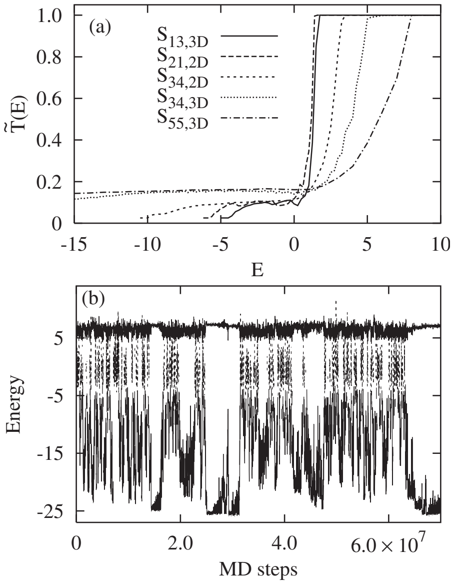

FIG. 4. (a) Statistical temperature estimates from STMD simulations of various chains $S_{n,m}$ in AB model, and (b) time series of E_{bend} (upper solid lines), E_{nonb} with $\tilde{T}(E)>0.2$ (dotted lines), and E_{nonb} with $\tilde{T}(E)<0.2$ (lower solid lines) in $S_{55,2D}$ of the AB model.

profiles in Fig. 4(b), the bend energy is centered around +6 with variation $\Delta E_{bend}\approx 3$ for a broad range of conformations, while the variation of the nonbonded interaction shows a dramatic change from $\Delta E_{nonb}\approx 9$ for unfolded, extended states with $\tilde{T}(E)>0.2$ to ≈ 20 for collapsed states with $\tilde{T}(E)<0.2$. This energetic dominance of the nonlocal interactions over the local interactions strengthens the nonspecific collapse tendency of AB proteins and creates a highly rugged energy landscape. The flat variation of $\tilde{T}(E)$ also explains why the conventional simulated-annealing-type optimization often fails in these models. In the temperature region where $\tilde{T}(E)$ is almost flat, the time required to reach local equilibrium in each annealing step increases steeply, due to a delocalized canonical energy distribution [27], which becomes very broad with a vanishing $\delta\tilde{\beta}(E)$ in Eq. (4), and the fast cooling of the system necessarily leads to trapping in one of the low-lying compact states.

B. Extended *AB* model

The other test model is the extended AB model [19], adding a torsional potential to the bend and nonbonded interactions,

$$E_{II}=-\kappa_1\sum_{i=1}^{N-2}\mathbf{u}_k\cdot\mathbf{u}_{k+1}-\kappa_2\sum_{k=1}^{N-3}\mathbf{u}_k\cdot\mathbf{u}_{k+2}+4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}C_{\sigma_i\sigma_j}(r_{ij}^{-12}-r_{ij}^{-6}),\qquad(6)$$

where $\mathbf{u}_k$ is the bond vector between monomers k and $k+1$, and $C_{\sigma_i\sigma_j}$ is +1 for AA pairs, and +1/2 for BB and AB pairs. Here, local interaction parameters $(\kappa_1=-1,\kappa_2=0.5)$ have been chosen to capture essential local properties of functional proteins. In the simulations we used $T_l=0.03$, but different T_h and bin size Δ depending on the chain length. To perform an intensive search of low-energy conformations, the temperature sampling has been restricted to $T_h=T_0=0.4$ for longer chains S_{34} and S_{55}; otherwise, $T_h=T_0=0.8$. We used the same initial modification factor $f=1.0005$ for all simulations. With torsions, the model is inherently three dimensional.

With respect to the original AB model, the inclusion of a torsional potential increases chain stiffness, diminishes the accessible conformational space, and opposes an initial rapid collapse to compact states. The resulting potential energy landscape is more funnel-like, as in functional proteins, and directly affects the folding thermodynamics and the search process for a global minimum. The comparison of the original and extended AB models may illustrate the importance [28] of the backbone in protein folding.

The energy time series for $S_{21,\mathrm{II}}$ (II denotes the extended model) in Fig. 5(a) clearly shows two separate sampling regions, corresponding to high-energy, extended states and distinct folded states. The inherent structure energies [gray (red) dots] demonstrate that the simulation can easily find a putative global minimum at 8×10^6 MD steps with no particular

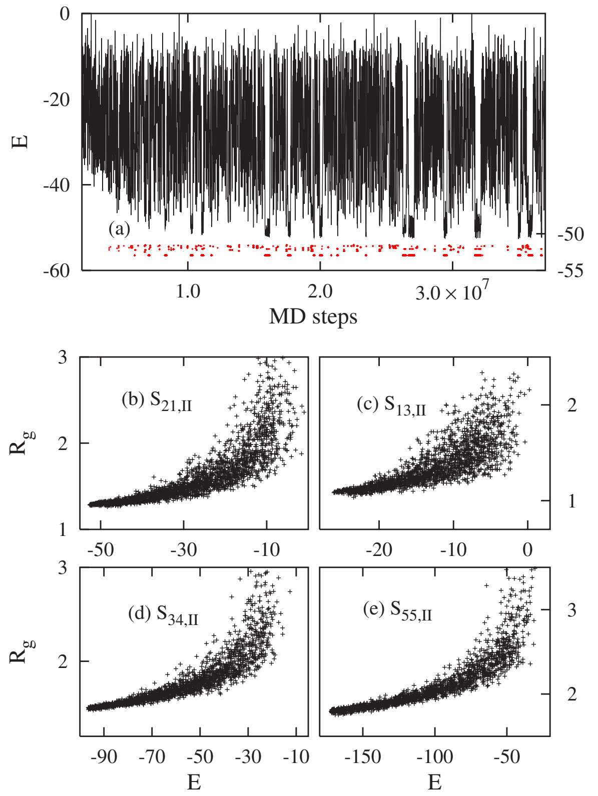

FIG. 5. (Color online) (a) Time series of STMD simulation and inherent structure energies E_{is} [gray (red) dots, shifted by −4 for visualization] determined by quenching low-lying STMD configurations of $S_{21,\mathrm{II}}$ with $E<-45$, and (b) scatter plots of simulated configurations in (E,R_g) for (b) $S_{21,\mathrm{II}}$, (c) $S_{13,\mathrm{II}}$, (d) $S_{34,\mathrm{II}}$, and (e) $S_{55,\mathrm{II}}$ in the extended *AB* model.

computational effort. The systematic escapes from local minima combined with a random walk in energy enable repeated visits to the global minimum through the self-adjusting statistical temperature. This is in sharp contrast to the case of $S_{55,2D}$ of the original *AB* model, in which the search for the global minimum is frustrated by kinetic trapping in low-lying IS, and the global minimum is accessible only through extensive sampling of low-energy states (see Fig. 2).

The scatter plots in (E,R_g) in Figs. 5(b) ($S_{21,\mathrm{II}}$), 5(c) ($S_{13,\mathrm{II}}$), 5(d) ($S_{34,\mathrm{II}}$), and 5(e) ($S_{55,\mathrm{II}}$) also illustrate the funneled shape of energy landscapes of the extended *AB* model, in which the collapse of the proteins proceeds gradually with decrease of energy, leading to the global minimum without a significant interruption by trapping.

Torsions also change the behavior of $\tilde{T}(E)$. Instead of the sharp drop and long tail of $\tilde{T}(E)$ in the original *AB* model, the statistical temperature of $S_{21,\mathrm{II}}$ in Fig. 6(a) shows a smooth variation for the whole energy region and a characteristic rounding at low energy, $E\approx-48$, corresponding to the folding transition and consistent with the folding energy region in Fig. 5(a). Since $C_v=[\partial T(E)/\partial E]_U^{-1}$, from the equivalence of microcanonical and canonical ensembles [29], $U(T)$ being the canonical average energy at T, it can be easily seen that the specific heat will give a peak at the temperature corresponding to the folding energy region. The smooth variation of $\tilde{T}(E)$ crossing the collapse and the rounding near the folding region are found in all studied sequences of the extended *AB* model, as seen in Fig. 6(b), in which the statistical temperatures of $S_{13,\mathrm{II}}$ and $S_{21,\mathrm{II}}$ are fully convergent up to $f_d=10^{-8}$, and those of $S_{34,\mathrm{II}}$ and $S_{55,\mathrm{II}}$ have been optimized with f_d=0.000 12.

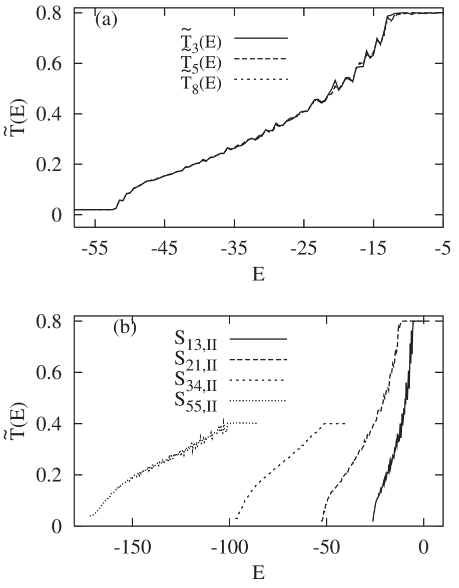

FIG. 6. (a) Statistical temperature estimate $\tilde{T}_i(E)$ at *i*th iteration step in $S_{21,\mathrm{II}}$ and (b) temperature estimates for $S_{13,\mathrm{II}}$, $S_{21,\mathrm{II}}$, $S_{34,\mathrm{II}}$, and $S_{55,\mathrm{II}}$ in the extended *AB* model.

The advantage of STA in finding low-lying, and hence global, minima is more prominent in the extended *AB* model, as shown in Table III. STA finds new lowest-energy values for $S_{21,\mathrm{II}}$ and $S_{55,\mathrm{II}}$, and confirms the putative ground state for $S_{13,\mathrm{II}}$ determined by the ACMC method. Our result for $S_{34,\mathrm{II}}$ is well below those from ACMC and ELP simulations, and slightly higher than that of CSA by about 0.11%. STA lowest-energy structures in Fig. 7 show a single hydrophobic core, with more spherical compact structures than with the original *AB* model in 3D. The structure of $S_{13,\mathrm{II}}$ is almost identical to that determined by ACMC calculations, corresponding to a small energy difference, but other confomers show a considerable structural difference from known putative ground states. Note that the lowest-energy value sampled by STMD restricted to low-temperature regions (T_l=0.01 and $T_h=T_0$=0.2) is even lower for $S_{13,\mathrm{II}}$, $S_{34,\mathrm{II}}$, and $S_{55,\mathrm{II}}$, and almost the same for $S_{21,\mathrm{II}}$, compared to the best values from ELP not employing a local minimization (see the values for STMD in Table III).

Except for the difference in the dynamic update scheme for the sampling weight, the basic strategy of STA optimization is similar to that of other generalized ensemble variants, e.g., the ACMC, MUCA, and ELP methods. All these methods are designed to steer the search away from previously visited energy regions by imposing a self-adjusting penalty (or modification) in the sampling weight (or the statistical temperature). The minor differences of the energy values of the STA lowest minima and those of ACMC and ELP in

TABLE III. Lowest-lying energy minima of the extended *AB* model determined by ACMC [17], ELP [16], CSA [7], and STA. The energy values in the STMD and FBMCMD [30] columns correspond to the lowest-energy values sampled by STMD and FBMCMD simulations, respectively, restricted for low-temperature regions with $T_l=0.01$ and $T_h=T_0=0.2$.

Sequence	ACMC	ELP	CSA	STA	STMD	FBMCMD
$S_{13,\mathrm{II}}$	−26.506	−26.498	−26.4714	−26.5066	−26.5052	−26.4354
$S_{21,\mathrm{II}}$	−51.7575	−52.917	−52.7865	−52.9339	−52.9100	−52.7040
$S_{34,\mathrm{II}}$	−94.0431	−97.261	−97.7321	−97.6171	−97.5570	−97.3281
$S_{55,\mathrm{II}}$	−154.505	−172.696	−173.9803	−174.5681	−174.4890	−172.8869

Tables I–III indicate comparable performance for the short 12-mer and 21-mer, in which Monte Carlo trial moves are still effective for sampling compact, collapsed conformations. However, STA outperforms ACMC and ELP consistently as the system size or the dimension increases, and the potential becomes more complicated. This implies that the success of STA optimization should be attributed to the combination of the enhanced sampling with the collective movements of particles generated by MD simulation, which allows more effective conformational changes even in a highly condensed phase.

Finally, the optimization performance of STMD has been compared with that of the well-established multicanonical MD algorithm in Fig. 8 for both long chains of $S_{34,\mathrm{II}}$ and $S_{55,\mathrm{II}}$. Instead of conventional multicanonical MD [14], here we used the force-biased multicanonical MD (FBMCMD) [30], which accelerates the convergence of the simulation by determining the sampling weight automatically. The simulation protocol of FBMCMD is almost the same as that of STMD except for the update scheme for the statistical temperature. The relative performance has been examined by plotting the sampled lowest-energy value as a function of MD steps, with an update every time the system finds a lower energy. We used the same simulation parameters in both STA and FBMCMD, $\Delta=1.0$, $T_l=0.01$, and $T_h=T_0=0.2$.

As seen in Fig. 8, STMD finds lower-energy states for both chains, and, for low energies found with both methods, finds them sooner. The reason is that the dynamic modification of the statistical temperature in STMD allows an instant self-adjustment of the sampling weight to escape a trapping, while FBMCMD requires time for a histogram accumulation. Note that we have performed a test for finding minima, the topic of this paper, not a comprehensive comparison of STMD and FBMCMD. However, based on Fig. 8 and the arguments given about the sampling weight, we expect that STMD is also superior for other applications in rough energy landscapes.

IV. SUMMARY AND CONCLUSIONS

In summary, the optimization performance of STA has been examined in various confomers of the off-lattice *AB* and extended *AB* protein models. The quenching of low-lying STMD configurations, generated by the self-adjusting sampling weight combined with the dynamic update scheme for the statistical temperature, shows a superior ability to find low-lying minima in rough energy landscapes compared to other optimization algorithms; thus there is a greater possibility of finding the global minimum. Since STA employs a collective movement of the beads through molecular dynamics simulation, low-energy, compact states are more effectively sampled in a manner that grows in importance with

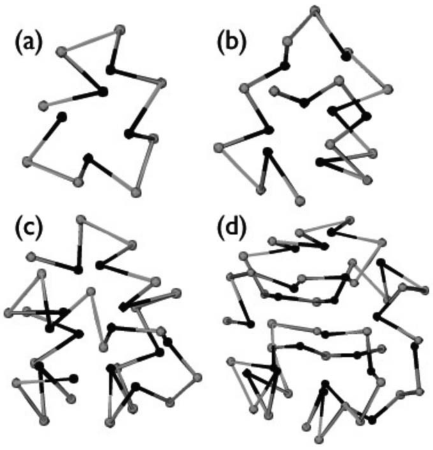

FIG. 7. Lowest-energy structures for (a) $S_{11,\mathrm{II}}$, (b) $S_{21,\mathrm{II}}$, (c) $S_{34,\mathrm{II}}$, and (d) $S_{55,\mathrm{II}}$ in the extended *AB* model determined by STA. Gray and black balls correspond to hydrophilic and hydrophobic monomers, respectively.

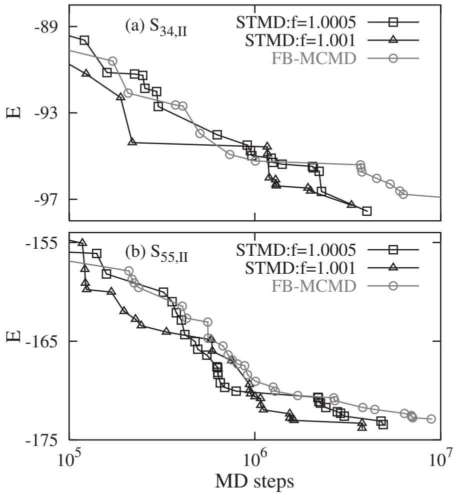

FIG. 8. Time series of lowest energy values sampled by STMD with different initial modification factor *f* and FBMCMD [30] simulations for (a) $S_{34,\mathrm{II}}$ and (b) $S_{55,\mathrm{II}}$ in the extended *AB* model.

increased chain length and dimension, and the inclusion of torsions.

We also found that the folding of Fibonacci sequences of the original *AB* model proceeds by a two-step process of a nonspecific collapse, followed by slow relaxation to the ground state in a glassy regime. This characteristic folding behavior is strongly correlated with the initial sharp drop and the low-energy tail of the statistical temperature crossing the collapse region, which is due to the energetic dominance of nonbonded interactions over the local interactions. On the other hand, scatter plots of sampled configurations in (E,R_g) show that the inclusion of a torsional potential in the extended *AB* model makes the global shape of the potential energy landscape more funnel-like by restricting the accessible conformational space with an increased chain stiffness, attenuating the nonspecific collapse, and producing a smoothly varying statistical temperature.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. CHE-0312959 (ITR) and CHE-0352026.

[1] C. B. Anfinsen, Science **181**, 223 (1973).

[2] K. F. Lau and K. A. Dill, Macromolecules **22**, 3986 (1989); D. Shortle, H. S. Chen, and K. A. Dill, Protein Sci. **1**, 201 (1992).

[3] C. I. Chou, R. S. Han, S. P. Li, and T. K. Lee, Phys. Rev. E **67**, 066704 (2003).

[4] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[5] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **84**, 6611 (1987).

[6] U. H. E. Hansmann and L. T. Wille, Phys. Rev. Lett. **88**, 068105 (2002).

[7] J. Lee, H. A. Scheraga, and S. Rackovsky, J. Comput. Phys. **18**, 1222 (1997).

[8] C. Micheletti, A. Laio, and M. Parrinello, Phys. Rev. Lett. **92**, 170601 (2004).

[9] O. F. Lange, L. V. Schafer, and H. Grubmuller, J. Comput. Chem. **27**, 1693 (2006).

[10] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992).

[11] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).

[12] B. J. Bern and J. E. Straub, Curr. Opin. Struct. Biol. **7**, 181 (1997).

[13] A. Mitsutake, Y. Sugita, and Y. Okamoto, Biopolymers **60**, 96 (2001).

[14] N. Nakajima, H. Nakamura, and A. Kidera, J. Phys. Chem. B **101**, 817 (1997); U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, Chem. Phys. Lett. **259**, 321 (1996).

[15] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001).

[16] M. Bachmann, H. Arkin, and W. Janke, Phys. Rev. E **71**, 031906 (2005).

[17] F. Liang, J. Chem. Phys. **120**, 6756 (2004).

[18] F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, Phys. Rev. E **48**, 1469 (1993); F. H. Stillinger and T. Head-Gordon, *ibid.* **52**, 2872 (1995).

[19] A. Irback, C. Peterson, F. Potthast, and O. Sommelius, J. Chem. Phys. **107**, 273 (1997).

[20] H.-P. Hsu, V. Mehra, and P. Grassberger, Phys. Rev. E **68**, 037703 (2003).

[21] J. G. Kim, J. E. Straub, and T. Keyes, Phys. Rev. Lett. **97**, 050601 (2006).

[22] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[23] E. J. Barth, B. B. Laird, and B. J. Leimkuhler, J. Chem. Phys. **118**, 5759 (2003).

[24] F. H. Stillinger and T. A. Weber, Phys. Rev. A **28**, 2408 (1983); Science **225**, 983 (1984).

[25] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN* (Cambridge University Press, Cambridge, U.K., 1992).

[26] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **32**, 136 (1998).

[27] J. G. Kim, Y. Fukunishi, and H. Nakamura, Phys. Rev. E **67**, 011105 (2003).

[28] G. Rose, P. Fleming, J. Banavar and A. Maritan, Proc. Natl. Acad. Sci. U.S.A. **103**, 16623 (2006).

[29] K. Huang, *Statistical Mechanics* (Wiley, New York, 1972).

[30] J. G. Kim, Y. Fukunishi, A. Kidera, and H. Nakamura, Phys. Rev. E **68**, 021110 (2003); J. G. Kim, Y. Fukunishi, and H. Nakamura, *ibid.* **70**, 057103 (2004).