Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information

(ESI)

Conformational Sampling of Metastable States:

Tq-REM as a Novel Replica Exchange Method

MinJun Lee,[1] Jeseong Yoon,[1,*] Soonmin Jang,[2] Seokmin Shin[1,*]

* Corresponding author. E-mail: yjs0@snu.ac.kr; sshin@snu.ac.kr

[1] Department of Chemistry, Seoul National University, Seoul 151-747, Korea

[2] Department of Chemistry, Sejong University, Seoul 143-747, Korea

Quantitative Analysis of Sampling Efficiency

In Figure S1a, the evolution of free energy profile for different time windows during the simulations illustrates that *Tq*-REM more rapidly samples the basin topology of *T*-REM's free energy surface than *T*-REM. In order to quantitatively measure the convergence behavior of sampling for different simulations, we introduce a quantity $< \delta F(t) >$ defined as follows:

$$< \delta F(t) > \equiv \sqrt{\frac{1}{N_{threshold}} \sum_{\{i,j \mid FE_{threshold}\}}^{n,n} \left(FE_{i,j}^{obj}(t) - FE_{i,j}^{ref} \right)^2}$$

where PC space is discretized by dividing the PC1 and PC2 axes into $n \times n$ uniform cells, $FE_{i,j}^{obj}(t)$ is the ijth component of the object free energy profile of a particular simulation, calculated by the trajectory up to time t. Similarly, $FE_{i,j}^{ref}$ is the ijth component of the reference free energy profile of whole trajectory. We may consider convergence properties of free energy surfaces around local basin structures. $\{i,j \mid FE_{threshold}\}$ denotes the inclusion of cells which have their free energy values lower than $FE_{threshold}$. Similar 'ergodic measure' was used for measuring convergence efficiency of REST.[1] Ergodic measure measures how fast the free energy profiles of the two independent trajectories of the same system approach each other. In contrast, $< \delta F(t) >$ shows how rapidly the trajectory of the object system develops the basin topology of the free energy for the reference system. $< \delta F(t) >$ can be calculated for two independent trajectories of the same system or for two trajectories of different systems.

We calculated $< \delta F(t) >$ for the simulations of *T*-REM and *Tq*-REM at 300 K for Met-enkephalin. Here we set $FE_{threshold}$=-4 kcal/mol to consider only local minimum regions. For a benchmark system with small peptide, the intrinsic convergence properties of *T*-REM and *Tq*-REM are very similar as shown in Figure S1c where the reference for the calculation of $< \delta F(t) >$ is taken to be the free energy profile of whole trajectory for respective simulations. Figure S1b shows the behavior of $< \delta F(t) >$ for *T*-REM and *Tq*-REM when the reference is taken as the free energy profile of whole trajectory for *T*-REM. $< \delta F(t) >$ can be fit to a double exponential function: $y = a \times e^{-t/\tau_1} + (1-a) \times e^{-t/\tau_2}$. For *T*-REM, we obtained τ_1=3.39 and τ_2=11800, while τ_1=4.03 and τ_2=24742 for *Tq*-REM. It was observed that initial convergences are very fast up to t ~ 15ns. This early phase of sampling convergence can also be seen from Figure S1a. After t ~ 15ns, $< \delta F(t) >$ of *T*-REM continues to decrease while $< \delta F(t) >$ for *Tq*-REM seems to be almost flat, as shown more clearly in the logarithmic scale curve (inset of Figure S1b). This difference at t > 15ns suggests that *Tq*-REM trajectory completes sampling of the basic topology of the converged free energy profile for *T*-REM faster than *T*-REM trajectory itself. In other words, it can be argued that *Tq*-REM samples conformations in the region of free energy minimum more efficiently than conventional *T*-REM.

In order to examine the efficiency of sampling in conformational space, we calculated mean square displacement(MSD) in PC space, defined by MSD(τ) = $\langle (R(t+\tau) - R(t))^2 \rangle_t$ where $R(t)$ is coordinates of trajectory at time t in PC space. Figure S1d shows that the diffusion behavior in conformational space for *T*-REM and *Tq*-REM is similar, although *T*-REM seems to show faster diffusion in the intermediate time (region II). Increased MSD in region II can be attributed to faster diffusion between local basins, which may imply that *T*-REM's trajectory stays inside individual local basins for relatively shorter time. We calculated average residence times of the trajectories in local basin and non-basin regions of PC free energy surface. When we define the local basins as regions where the free energy is lower than -4 kcal/mol, residence times in local basin and non-basin regions are shorter (3.245 ps and 1.599 ps) for *T*-REM than *Tq*-REM (3.483 ps and 4.755 ps). When the energy criteria for local basin is changed to be -3 kcal/mol, basin residence time of *T*-REM is still shorter (33.21 ps) than *Tq*-REM (36.14 ps) while non-basin residence time of *T*-REM is comparable (0.9489 ps) to that of *Tq*-REM (0.9481 ps). One may conclude that *Tq*-REM samples local basin regions more intensively than *T*-REM, leading to improved basin sampling efficiency.

Sampling of dihedral angle space is examined by considering dihedral angle vector $\overrightarrow{d(t)}$ whose rth row components are $(\cos\varphi_r,\ \sin\varphi_r,\ \cos\psi_r,\ \sin\psi_r)$ with the is residue number r.[2] The sampling probability of local basin k can be calculated by $\mathrm{Probability}_k(t) = \langle b_k(t) \rangle$ where $b_k(t)$ is 1 if $\cos^{-1}\left(\overrightarrow{d(t)} \cdot \overrightarrow{d_k} / \left|\overrightarrow{d(t)}\right|\left|\overrightarrow{d_k}\right|\right) \leqq \delta\overrightarrow{d_k}$, or otherwise 0. $\delta\overrightarrow{d_k} = \langle \cos^{-1}\left(\overrightarrow{d_l} \cdot \overrightarrow{d_k} / \left|\overrightarrow{d_l}\right|\left|\overrightarrow{d_k}\right|\right) \rangle_l$ where $\overrightarrow{d_k}$ is dihedral angle vector with respect to the deepest free energy region for each basin and l denotes grids in PC space with values of free energy lower than -4 kcal/mol. This probability function shows the sampling probability around certain local basin in dihedral angle space. To examine the effects of different dihedral angle sampling pattern on overall conformational sampling, we calculated the time evolution of sampling probabilities of four basins in dihedral space[2] corresponding to basin 1 to 4 in Figure 2. Figure S2d shows that sampling probability of each basin and the overall basin sampling probability (inset of Figure S2d) converge faster in *Tq*-REM than *T*-REM. It was also shown that *Tq*-REM is capable of sampling metastable states (basin 2,3) in addition to the global minimum state (basin 1). In contrast, *T*-REM predominantly samples global minimum state much more than metastable states. It can be argued that *q*-replica with reduced potential barrier is more efficient than *T*-replica with higher temperature for balanced sampling of various local minimum dihedral angles, resulting in improved sampling of metastable states.

1. P. Liu, B. Kim, R. A. Friesner and B. J. Berne, *P Natl Acad Sci USA*, 2005, **102**, 13749-13754.
2. Y. G. Mu, P. H. Nguyen and G. Stock, *Proteins-Structure Function and Bioinformatics*, 2005, **58**, 45-52.

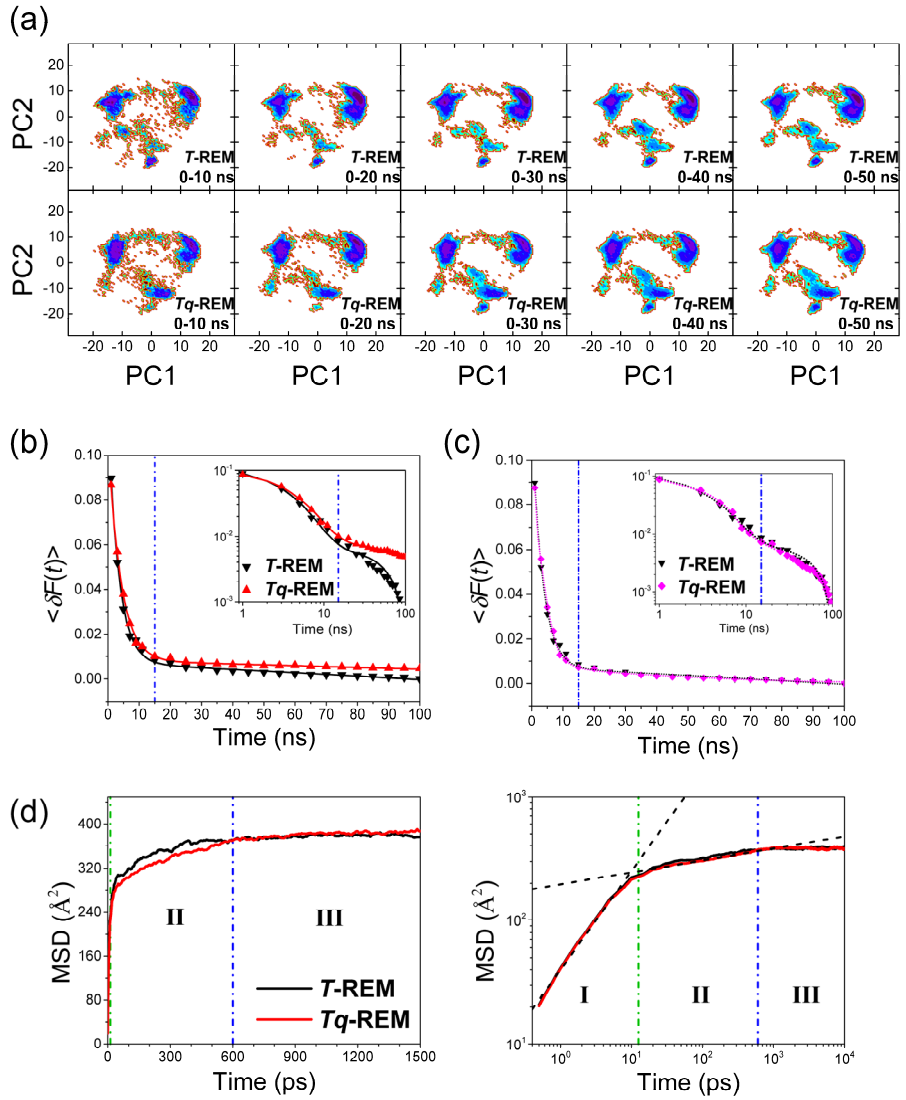

Figure S1. (a) Time evolutions of the free energy profiles of *T*-REM and *Tq*-REM in PC space at 300 K for Met-enkephalin. The free energy profile regions whose energies are lower than -4 kcal/mol are shown for clarity. (b) $< \delta F(t) >$ for *T*-REM and *Tq*-REM of which reference is free energy profile of whole trajectory for T-REM. The inset shows logarithm scale plots of the same graphs. (c) $< \delta F(t) >$ of *T*-REM and *Tq*-REM of which reference is free energy profile of whole trajectories for the respective simulations. (d) Time evolution of mean square displacement as calculated by MSD(τ) = $\langle (R(t+\tau) - R(t))^2 \rangle_t$ where $R(t)$ is coordinates of trajectory at time t in PC space. Linear scale (left) and logarithm scale (right) plots are shown. MSD can be fit to a power law function of time t^{α} in region I (0-12.5 ps) and II (12.5-600 ps), respectively, as denoted by the black dashed lines.

Sampling Behavior of Dihedral Angles

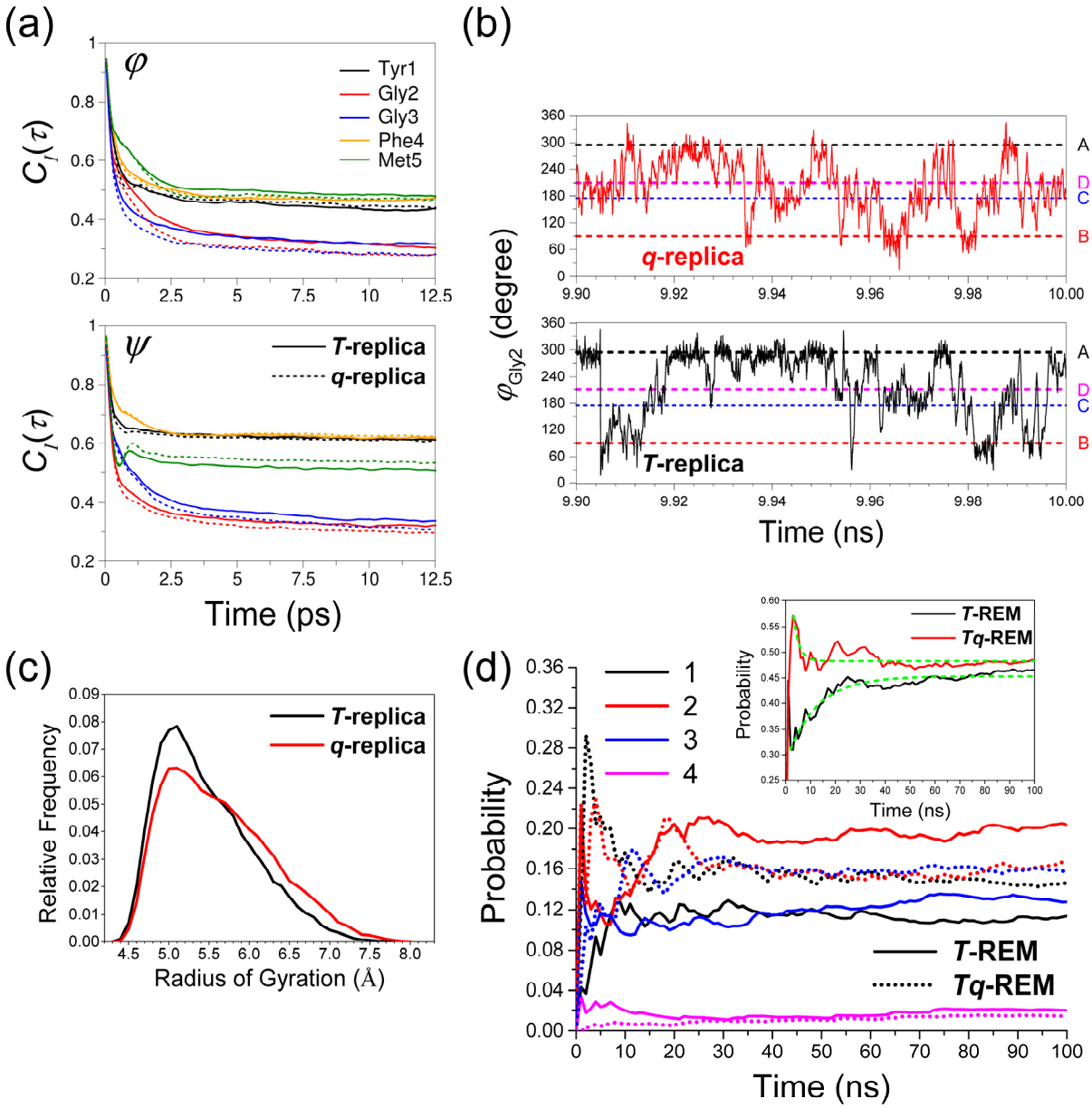

Figure S2. (a) Time auto-correlation functions of dihedral angles of Met-enkephalin, calculated for the trajectories corresponding to 5th replica of *T*-REM and *Tq*-REM. (b) Representative sampling behavior of φ_{Gly2} in *T*-replica and *q*-replica. Horizontal dash-dot lines indicate the φ values of local minima. (c) Distributions of radius of gyration of Met-enkephalin for *T*-replica and *q*-replica. (d) Time evolution of sampling probabilities of local basins.

Distributions of Structural Properties for Met-enkephalin

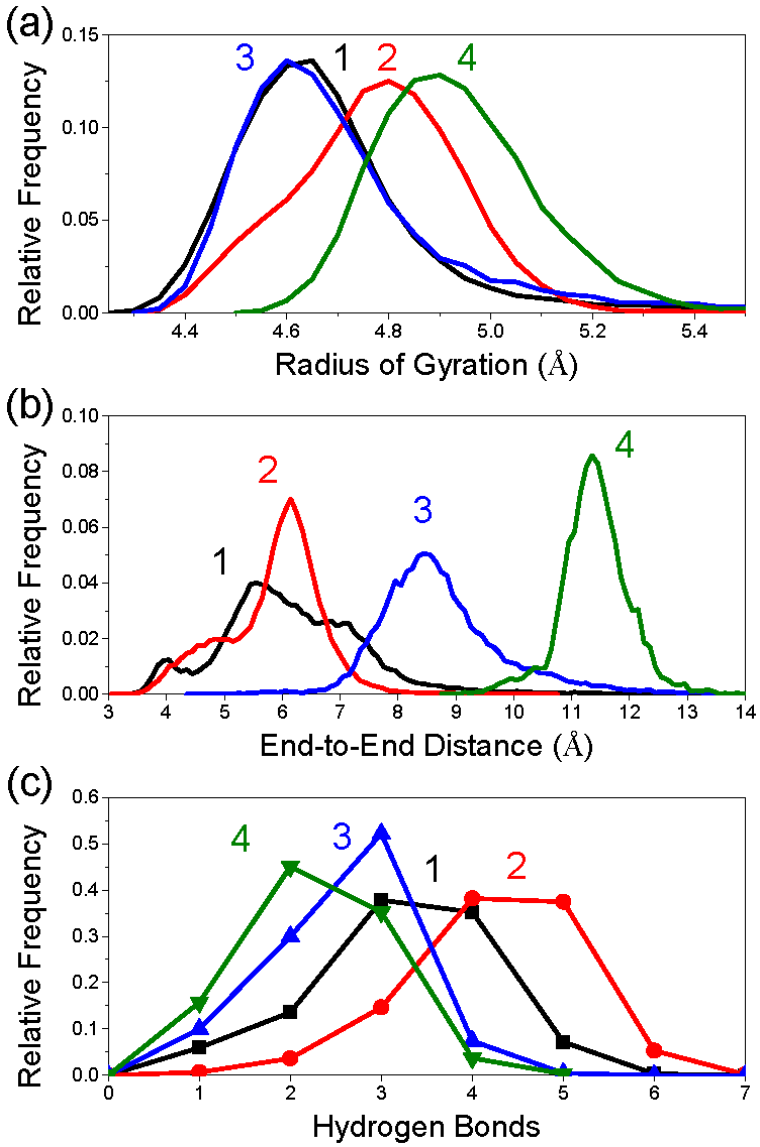

Figure S3. Distributions of (a) the radius-of-gyration, (b) the end-to-end distance, and (c) the number of hydrogen bonds for the four local basins in the free energy surfaces of *T*-REM and *Tq*-REM, as obtained from 100-ns trajectories at 300 K for Met-enkephalin.

Results for Multicanonical Simulations

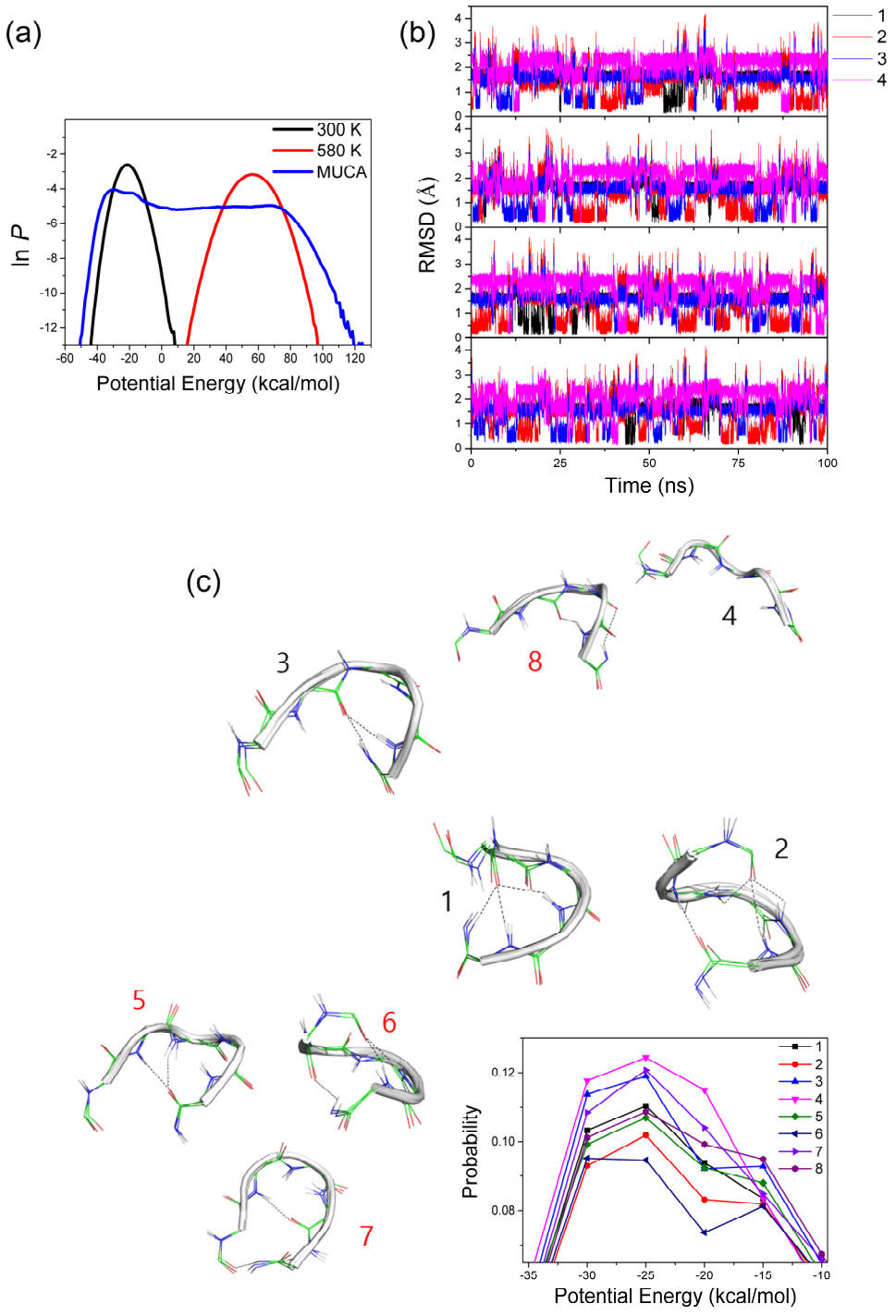

Figure S4. (a) Flat potential energy distribution we obtained for multicanonical simulation. (b) Backbone RMSD plots with respect to local basin structures from *T*-REM and *Tq*-REM. (c) Structures and potential energy probabilities of various basin structures. Hydrogen bonds are indicated by dashed lines.

Evolutions of Free Energy Surfaces for $(AAQAA)_3$

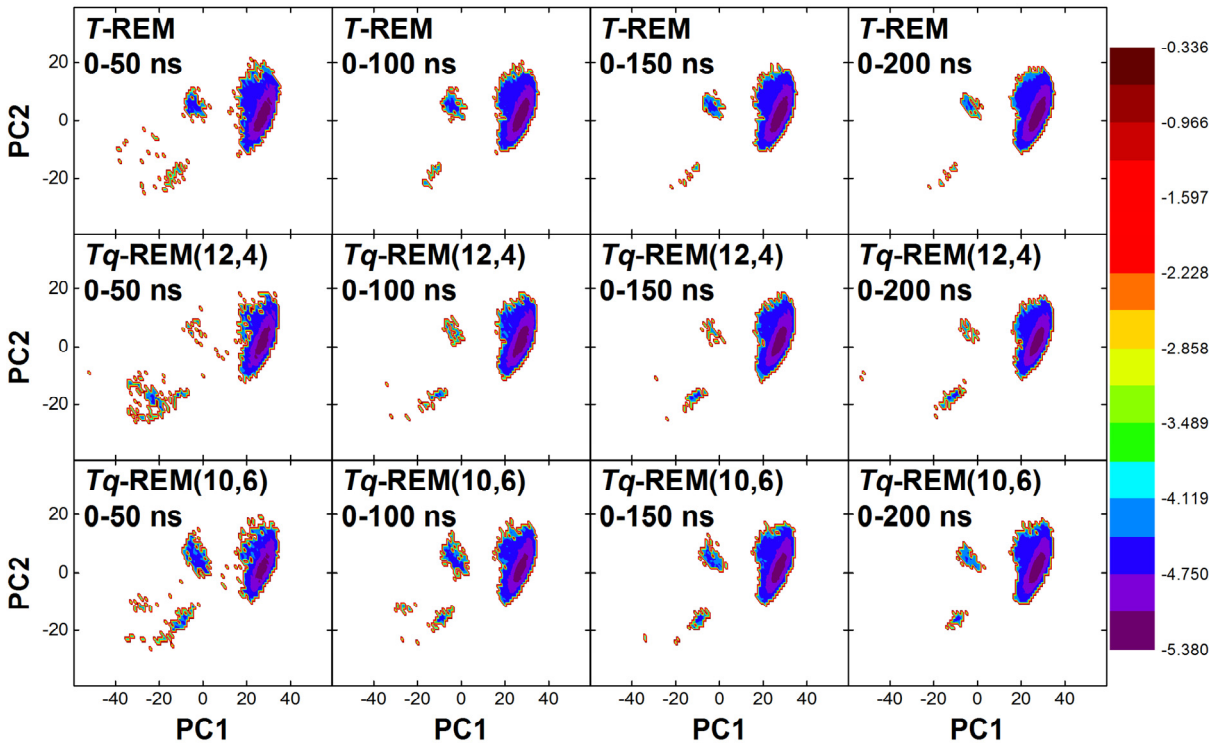

Figure S5. The evolutions of the free energy surfaces of *T*-REM and *Tq*-REM in PC space for different time windows during the 200-ns trajectories at 300 K for $(AAQAA)_3$. The free energy surface regions whose energies are lower than -4.35 kcal/mol are shown for clarity.

Distributions of Structural Properties for $(AAQAA)_3$

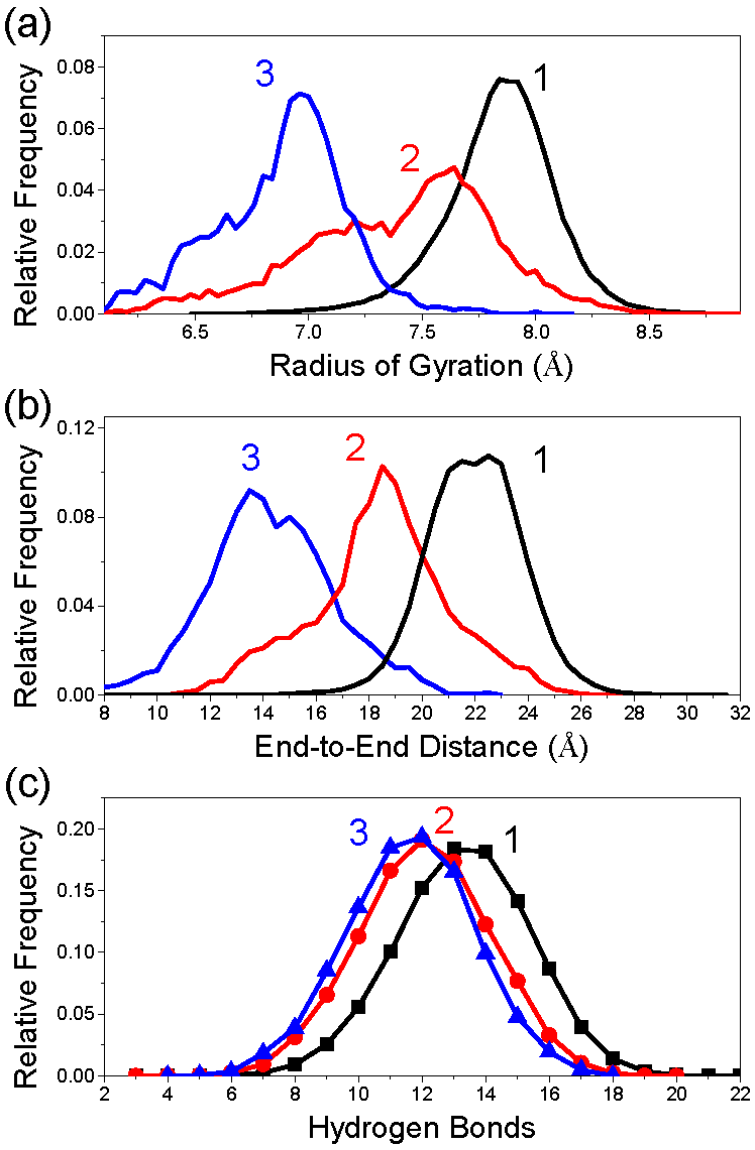

Figure S6. Distributions of (a) the radius-of-gyration, (b) the end-to-end distance, and (c) the number of hydrogen bonds for the three local basins in the free energy surfaces of *T*-REM and *Tq*-REM, as obtained from 200-ns trajectories at 300 K for $(AAQAA)_3$.

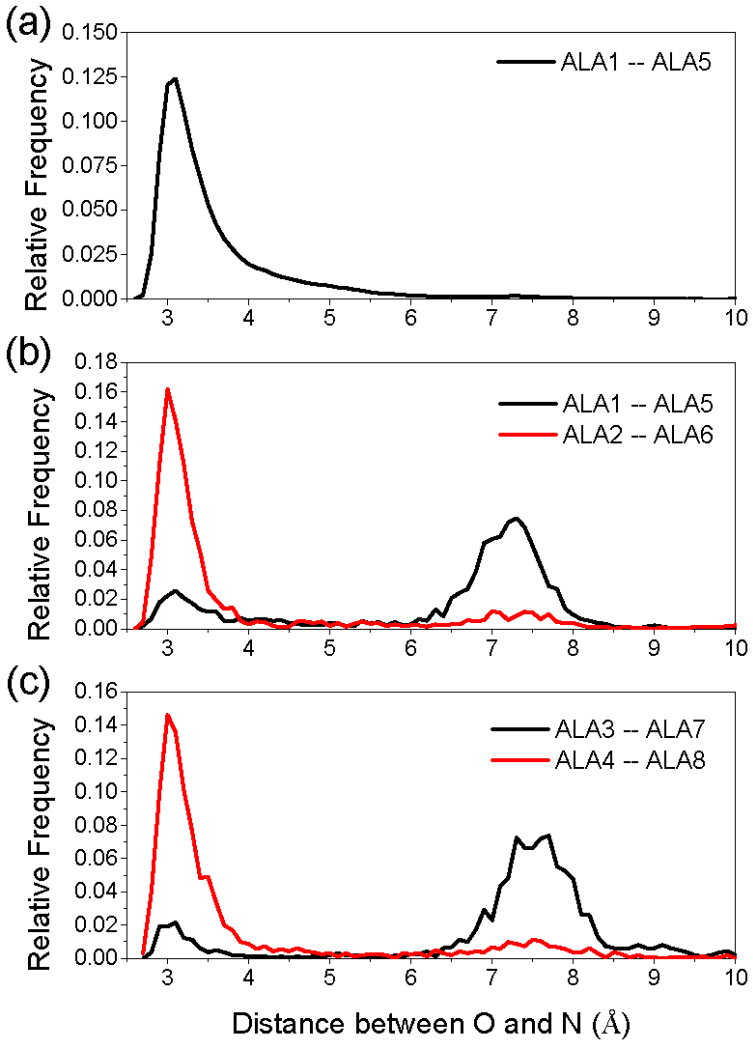

Figure S7. Distributions of distances between backbone oxygen and nitrogen of the residues participating in hydrogen bonds, for the local minimum structures corresponding to (a) basin 1, (b) basin 2, and (c) basin 3. Backbone hydrogen bonds are sequentially formed from middle to N-terminal direction along the pathway of 3→2→1 indicating the folding of $(AAQAA)_3$ helix.

Free Energy Surfaces in PC Space for $(AAQAA)_3$

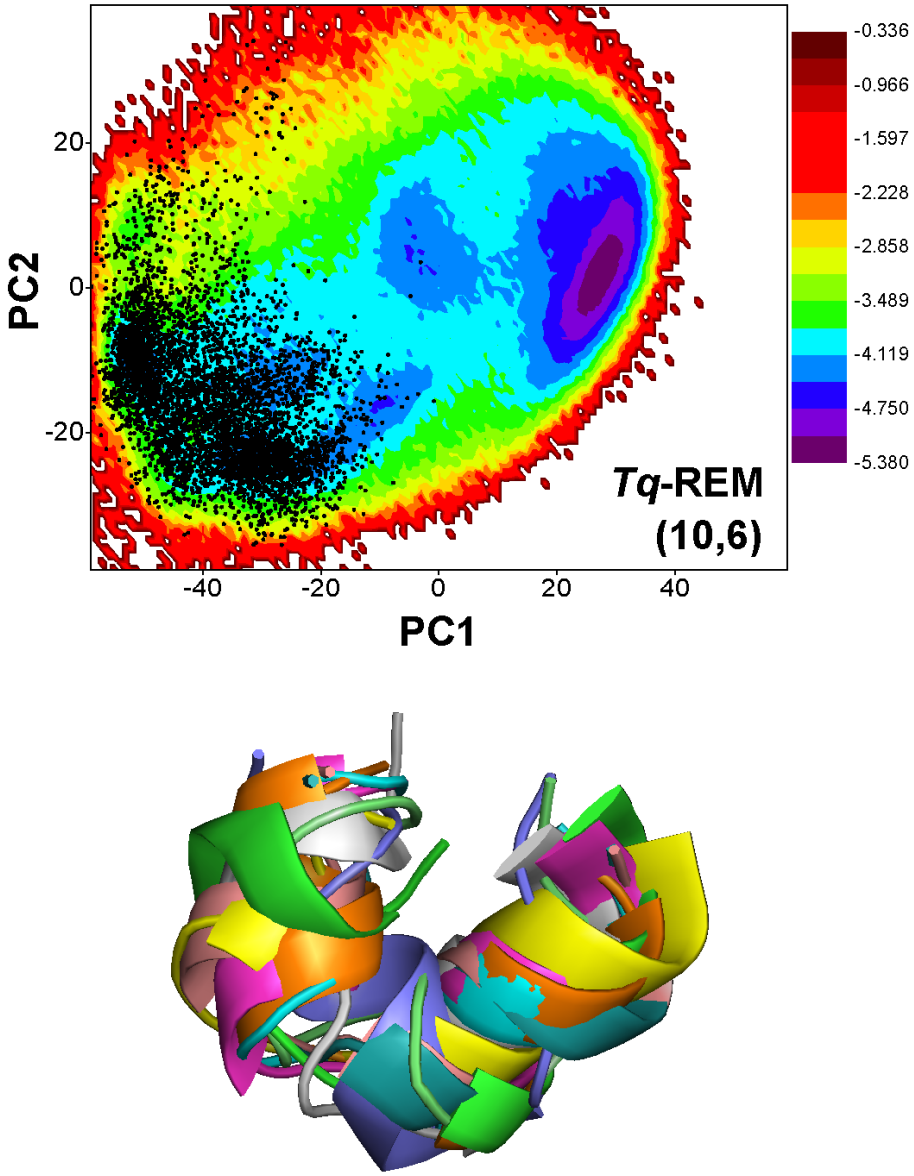

Figure S8. The free energy surfaces of *Tq*-REM(10,6) in PC space obtained from 200-ns trajectories at 300 K for $(AAQAA)_3$. Representative structures of $(AAQAA)_3$ for the local basin 4, as represented by black dots in the free energy surface.

Structural Properties for Trpzip2

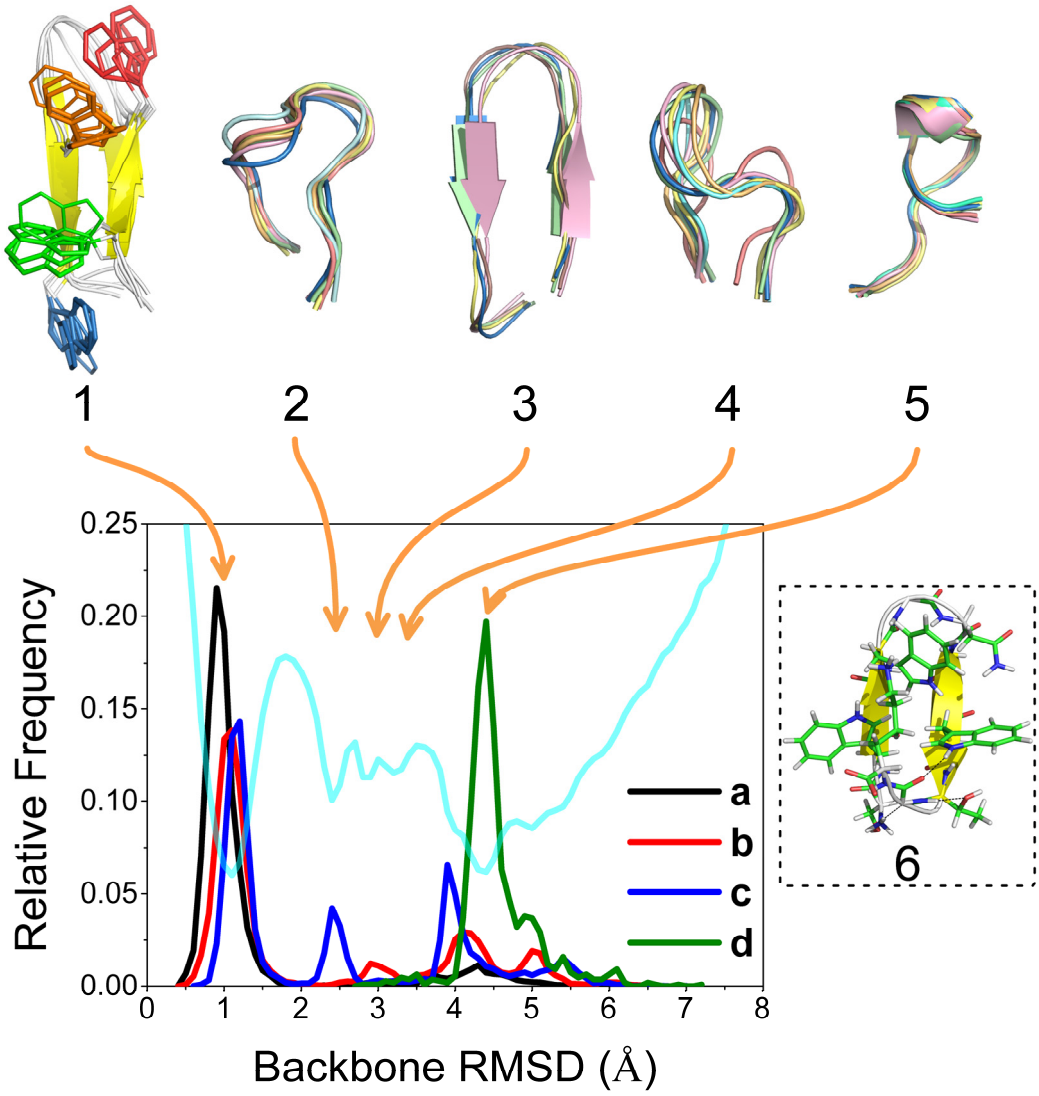

Figure S9. The distributions of the backbone RMSD of the structures corresponding to the four local basins (denoted **a** to **d**) in the free energy surface shown in Figure 10 of the article. The free energy as a function of the backbone RMSD, obtained from 200-ns trajectories at 300 K for Trpzip2, is also shown. The representative structures of Trpzip2, as shown in Figure 9 of the article, are related with the local minima in the free energy as a function of the backbone RMSD. The structure of off-folding pathway metastable state are denoted 6.

Results for Simulations of trialanine in Explicit Water

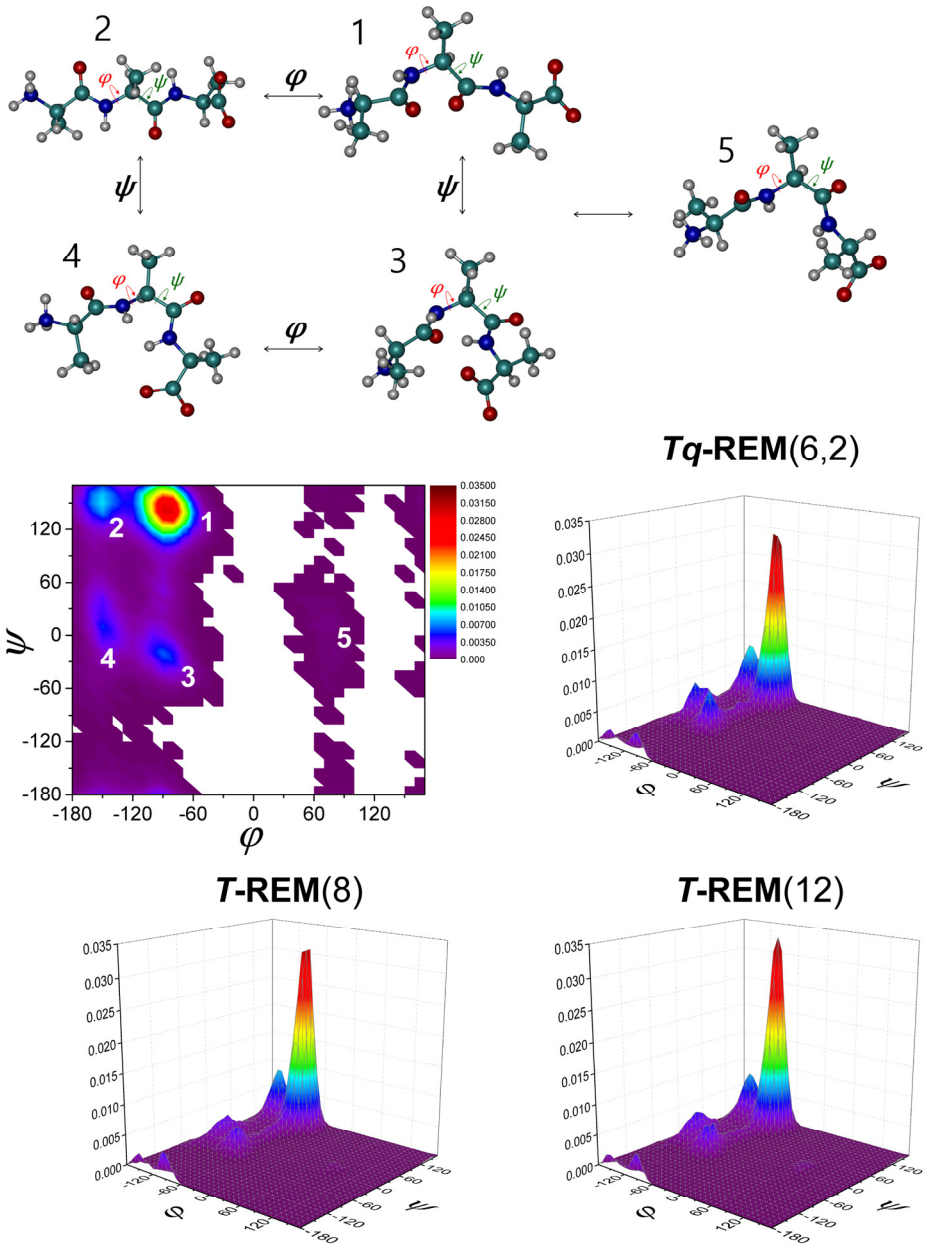

Figure S10. Representative structures of trialanine for the five local basins in the Ramachandran plots. Basins of five local minimum structures are denoted 1 to 5. Dihedral angles of central residue and transitions between basins are denoted. 3D graphs of Ramachandran distribution plots of each REMs show that *Tq*-REM(6,2) samples metastable α conformations more than two *T*-REMs.

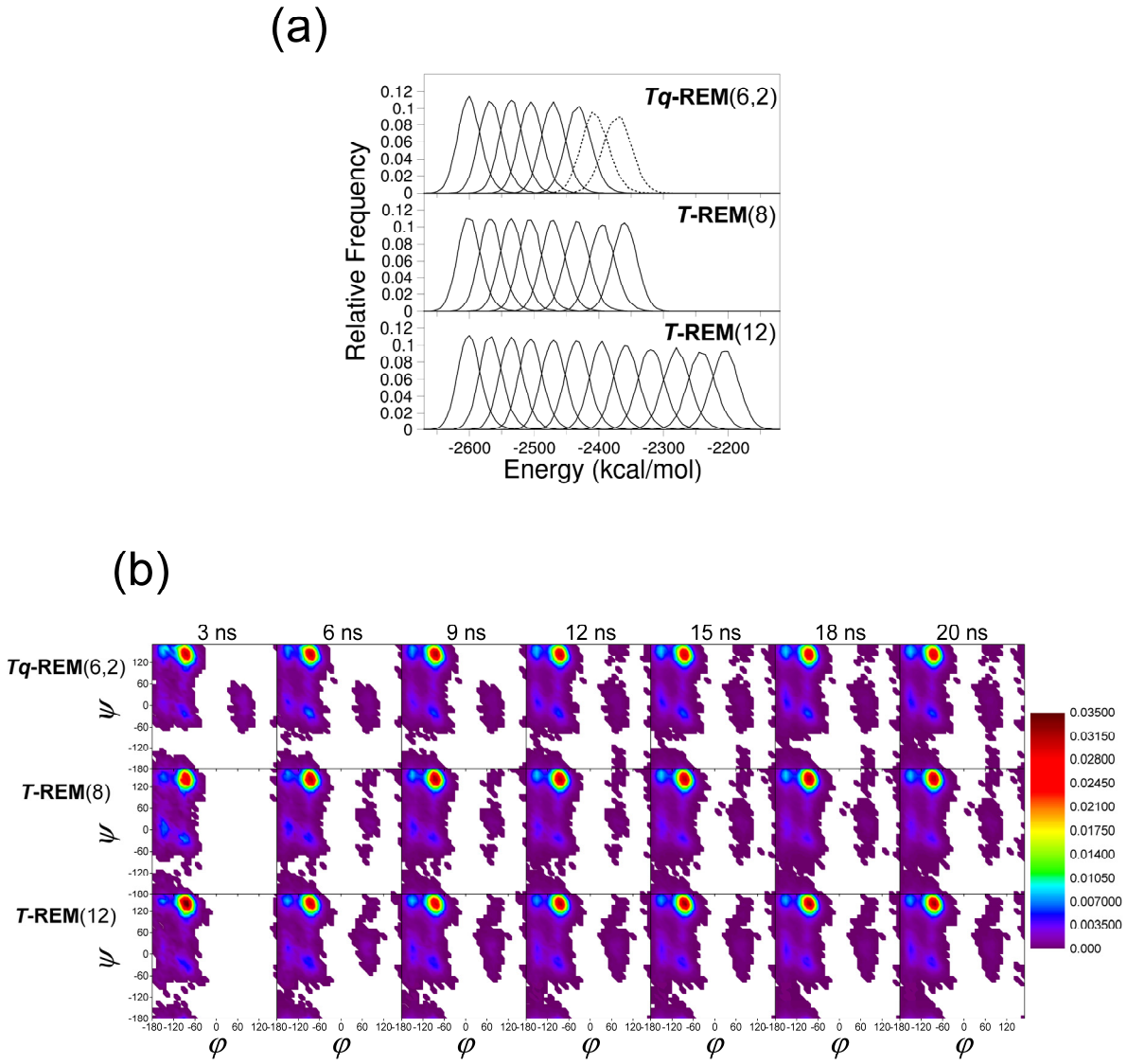

Figure S11. (a) The energy distributions of *Tq*-REM(6,2), *T*-REM(8), and *T*-REM(12). Solid lines are *T*-replicas and dot lines are *q*-replicas. (b) The time evolutions of the Ramachandran distributions of three REMs during 20-ns trajectories at 300 K for trialanine.