JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Folding Thermodynamics and Mechanism of Five Trp-Cage Variants from Replica-Exchange MD Simulations with RSFF2 Force Field

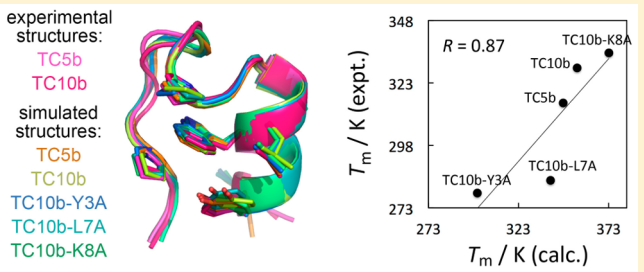
Chen-Yang Zhou,[†,‡] Fan Jiang,[*,†] and Yun-Dong Wu[*,†,‡]

[†]Laboratory of Computational Chemistry and Drug Design, Laboratory of Chemical Genomics, Peking University Shenzhen Graduate School, Shenzhen 518055, China

[‡]College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Supporting Information

ABSTRACT: To test whether our recently developed residue-specific force field RSFF2 can reproduce the mutational effect on the thermal stability of Trp-cage mini-protein and decipher its detailed folding mechanism, we carried out long-time replica-exchange molecular dynamics (REMD) simulations on five Trp-cage variants, including TC5b and TC10b. Initiated from their unfolded structures, the simulations not only well-reproduce their experimental structures but also their melting temperatures and folding enthalpies reasonably well. For each Trp-cage variant, the overall folding free energy landscape is apparently two-state, but some intermediate states can be observed when projected on more detailed coordinates. We also found different variants have the same major folding pathway, including the well formed P_{II}-helix in the unfolded state, the formation of W6-P12/P18/P19 contacts and the α-helix before the transition state, the following formation of most native contacts, and the final native loop formation. The folding mechanism derived here is consistent with many previous simulations and experiments.



INTRODUCTION

The functional structure of a natively folded protein is encoded in its amino acid sequence and achieved during the folding process. The protein folding problem includes three main questions:[1] (1) What physical interactions drive a protein to its native structure (the force field)? (2) How does a protein fold so rapidly (the folding mechanism)? (3) How can we predict protein structures from their sequences (the structure prediction)? Although recently there have been remarkable achievements toward resolving these challenges by using molecular dynamics (MD) simulations,[2] protein folding remains an active research area. Studies on relatively small model systems play a key role, especially in testing protein force fields[3,4] and revealing the folding free-energy landscapes and mechanisms.[2,5,6]

Trp-cage, a designed 20-residue mini-protein reported by Andersen and co-workers in 2002,[7] has become a popular model system for folding studies. In the native structure of Trp-cage (Figure 1), residues 3−9 form an α-helix and residues 16−19 form an extended polyproline-II (P_{II}) helix. The loop region of residues 10−15 connecting the α-helix and the P_{II}-helix contains successive β-turns (3_{10}-helix). The three segments together form a well-packed hydrophobic core surrounding the indole ring of the tryptophan W6 with several key nonlocal contacts. It also has a stabilizing salt-bridge between D9 and R16[8] and a side-chain-to-backbone hydrogen bond (H-bond) between W6 and R16. Its small size, very fast folding,[9] remarkable stability (Table 1), and protein-like structural features make it the ideal model system to study protein folding, especially via computer simulations.

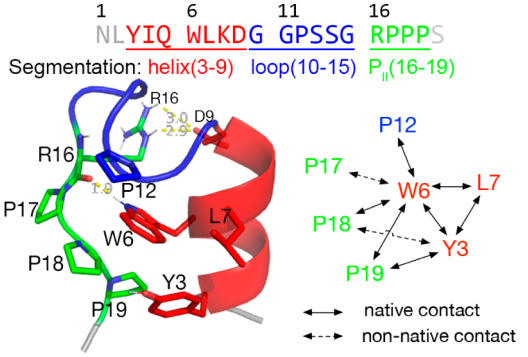


Figure 1. Sequence and experimental structure of the original Trp-cage (TC5b). The three secondary structure segments are in different colors: red for α-helix, blue for loop, and green for P_{II}-helix. The cartoon structure is from the first model in the NMR ensemble (1L2Y) in which the key side-chains forming the hydrophobic core are shown in sticks.

Despite numerous experimental[7,9−26] and computational[20,25,27−44] studies, the folding mechanism of the Trp-cage has not been fully understood with seemingly conflicting pictures. One important issue is whether there is a significantly populated folding intermediate state. Thermal-induced[7,14,17] and denaturant-induced[20] unfolding experiments using various characterization techniques (such as CD, DSC, NMR, and Trp fluorescence) support a two-state behavior of the Trp-cage. For

Received: June 20, 2015

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Folding Kinetics and Thermodynamic Parameters of Some Trp-Cage Variants from Experiments

variant	year	ref	τ_f (μs)	T_m (cage)	T_m (helix)	ΔH_f (kJ/mol)
TC5b	2002	7		315		
TC5b	2002	9	4.1	317		−49
TC5b	2011	18	3.7	315		−56
TC5b	2008	17		315	319	
TC10b	2011	18	1.6	328		−58
TC10b	2008	17		329	334	−65 ± 2
TC10b	2014	26	1.4			
TC10b-Y3A	2008	17		279	282	−32
TC10b-L7A	2008	17		284	296	−42
TC10b-K8A	2008	17		335	339	

most variants in Table 1, in particular, the melting temperatures (T_m) of the secondary structure (α-helix) and the tertiary structure (cage) are quite close. Single-exponential relaxation kinetics after a temperature jump also suggest two-state folding.[9] Conversely, UV-Raman spectroscopy suggests the existence of a compact folding intermediate with partially folded α-helix, where G11, Pro12, and Trp6 are closer than in the folded structure.[10] Early formation of a compact globule-like intermediate state was revealed by fluorescence correlation spectroscopy.[11] A more recent T-jump IR spectroscopic study indicated different relaxation behavior of different secondary structure segements.[18] Some experiments suggest that the intermediate is highly similar to the native structure with completely formed α-helix.[19] Some previous simulations also indicated the existence of an intermediate state before[40] or after the transition state.[33] Noticeably, a Markov state model with multiple metastable states and complex kinetic network can give out apparent two-state folding behavior.[45]

Another important issue is whether the formation of the α-helix occurs early during folding. NMR experiments have indicated the existence of significant nonlocal hydrophobic contacts even in the urea-denatured state.[15,22] An experimental study also suggested a folding intermediate of a collapsed structure without a well-formed α-helix.[46] However, very early hydrophobic collapse may not contradict the early formation of an α-helix, as indicated in a pioneering folding simulation by Duan and co-workers.[28] A combined study of T-jump experiment and MD simulations indicates that the formation of α-helix precedes the formation of the hydrophobic cage structure.[25] In a previous simulation using transition path sampling, parallel folding pathways were observed with the dominant one forming the tertiary contacts before helix formation.[31]

For a reliable folding mechanism from MD simulations, an accurate protein force field is a crucial requirement. For example, although the native structure of TC5b has been successfully predicted using an ab initio folding simulation before the release of its NMR structure,[47] some previous REMD simulations gave its T_m as much higher than the experimental value.[29,32,40,48] Also, some previous simulations gave considerable α-helix population in the unfolded state,[33,36] contradicting the experimental observations.[7,13]

Recently, we modified the OPLS-AA/L and the AMBER 99SB force fields with new backbone and side-chain dihedral angle (ϕ, ψ, χ_1, χ_2, etc.) potentials.[49,50] Different parameters were used for the 20 amino acid residues, and they were obtained by fitting the free energy surfaces of the dipeptide models from molecular dynamics simulations to those obtained from the statistical analysis of residues outside regular secondary structures (coil library) of high resolution protein structures.[51] Our residue-specific force field 1 (RSFF1) from OPLS-AA/L has successfully folded a set of 14 small proteins with systematic overstabilization of native structures.[52] The RSFF2 force field improved from AMBER 99SB also give balanced secondary structure preferences and better reproduce the experimental melting curves of some model systems.[50] In a very recent benchmark study, RSFF2 showed improvements in the modeling of conformational behavior of peptides and proteins.[53]

In addition, to predict the effect of a given mutation on the structure and stability of a protein is also important, such as in understanding protein evolution[54] and disease-related variations.[55] For Trp-cage, a series of mutations in the original TC5b were reported in 2008.[17] Among them, TC10b with stabilizing N1D/L2A/I4A triple mutations has attracted special attention. Very recently, English and Garcia used the Amber ff99SB force field to correctly reproduce the melting temperature (T_m) of TC10b to be higher than that of TC5b.[56] In their study, a replica-exchange molecular dynamics (REMD)[57] simulation was used to enhance the conformational sampling and to achieve folding/unfolding equilibrium. However, simulating the effect of mutation on folding thermodynamics is still quite limited.[58]

In this work, by long-time REMD simulations on five Trp-cage variants, we show that our RSFF2 force field can not only correctly reproduce the thermodynamic effects of mutations but also gives a general model of its folding mechanism that is consistent with previous studies. The differences among the five variants in their folding pathways are also studied.

MATERIALS AND METHODS

Simulation Details. As listed in Table 2, five Trp-cage variants were selected for our simulations, including three single-site mutants of TC10b.[17] All three mutants have one Ala substitution on one residue in the α-helix segment of TC10b. The TC10b-K8A has the highest melting temperature (T_m) of 335 K, whereas the TC10b-Y3A has the lowest T_m of 284 K. Only the structures of TC5b and TC10b have been resolved in NMR experiments.

Table 2. Trp-Cage Variants Simulated in This Work

name	sequence (1−9)	t_{trj} (μs)[a]	N_F[b]	N_U[b]
TC5b	NLYIQWLKD	1.27	71	56
TC10b	DAYAQWLKD	1.52	77	55
TC10b-Y3A	DAAAQWLKD	1.22	46	40
TC10b-L7A	DAYAQWAKD	1.42	69	54
TC10b-K8A	DAYAQWLAD	1.55	64	44

[a]Simulation time of each of the 36 replicas. [b]Total number of folding (F) and unfolding (U) events.

The initial NMR structures were solvated in truncated octahedron boxes with ∼1530 TIP3P[45] water molecules. All side-chains of Arg, Lys, and Asp were ionic at pH 7. Cl− and Na+ ions were added to neutralize the system and achieve a salt concentration of 50 mM. For each system, the initial periodic box was equilibrated using a 3 ns NPT MD simulation at 300 K and 1 atm. Then, a 5−10 ns NVT MD simulation at 600 K was carried out to generate fully unfolded structures for the subsequent REMD simulation. For each REMD simulation, 36 replicas ranging from 273 to 460 K were used, and the intermediate temperatures were chosen to give uniform exchange rates.[59] Each replica was simulated for >1.2 μs such that folding-unfolding

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

equilibrium can be reached.[39] All simulations were performed using *Gromacs* 4.5.4.[60]

Consistent with the settings in our previous works,[49,50] electrostatics were treated using the particle-mesh Ewald (PME) method with a real-space cutoff of 9 Å, and van der Waals interactions were truncated at 9 Å with the long-range dispersion correction for energy and pressure. A velocity-rescaling thermostat[61] with τ_T = 0.2 ps and a Berendsen barostat[62] with τ_P = 0.5 ps were used to maintain constant temperature and constant pressure (for NPT simulations), respectively. All bonds involving hydrogen were constrained using LINCS,[63] and a time step of 3 fs was used. The mass of the water oxygen atom was reduced from 16 to 2 amu to increase the sampling efficiency[64] without altering the thermodynamic equilibrium properties.

Analysis. We used the "gromos" clustering algorithm of Daura et al.[65] to obtain representative structures with a cutoff of 1.5 Å. We found that the obtained structures are not sensitive to the cutoff values. The cluster centers of the largest cluster under 300 K were used as the reference structure for the native state in the following analysis because there are no PDB structures for K8A, Y3A, or L7A mutants of TC10b. All of the RMSD values were calculated on the backbone atoms of various segments (residues 3−19, 3−9, 10−15, and 16−19). The fraction of all-atom native contact (Q_{aa}) was calculated using the following equation, as in the previous works of Best et al.[66] and our own[52]

$$Q_{aa} = \frac{1}{N}\sum_{(i,j)} \frac{1}{1 + \exp[\beta(r_{ij} - \lambda r_{ij}^0)]} \quad (1)$$

The summation includes all pairs of non-hydrogen atoms i and j separated by at least three residues with the distance between atoms i and j in the native state (r_{ij}^0) < 4.5 Å. The β is set to 5.0 Å^{-1}, and we use a slightly smaller λ value of 1.4, leading to a more strict criteria for native contact.

For each system, the melting curve $f_F(T)$, which is the fraction of folded as the function of temperature, was calculated using trajectories after 0.9 μs to ensure folding/unfolding equilibrium. The obtain $f_F(T)$ was then converted to the folding free energies

$$\Delta G_F(T) = -RT\ln\left(\frac{f_F(T)}{1 - f_F(T)}\right) \quad (2)$$

The obtained $\Delta G_F(T)$ around the T_m (folding midpoint) were fitted to $\Delta H_F - T\Delta S_F$ to obtain the folding entropy and enthalpy.

RESULTS AND DISCUSSION

Prediction of Experimental Structures and Thermodynamics. Starting from unfolded structures, REMD simulations using the RSFF2 force field can fold all five Trp-cage variants very well. As shown in Figure 2, the representative structures from the simulations of TC5b and TC10b are quite similar to their NMR structures (1L2Y and 2JOF) with RMSD values of 1.0 and 1.5 Å, respectively. In our simulations, the two N-terminal residues and one C-terminal residue are relatively flexible, possibly because the Trp-cage folding motif is defined by the hydrophobic staple interaction between Y3 and P19.[17] When excluding these three terminal residues, the simulated structures have RMSD < 1.0 Å from NMR structures. Interestingly, the representative structures of the three TC10b mutants are very similar to that of TC10b with RMSDs < 0.5 Å even for the Y3A mutant with an incomplete hydrophobic core. Indeed, the rigidity of the Trp-cage structure has been indicated in a recent experimental study of its cyclized

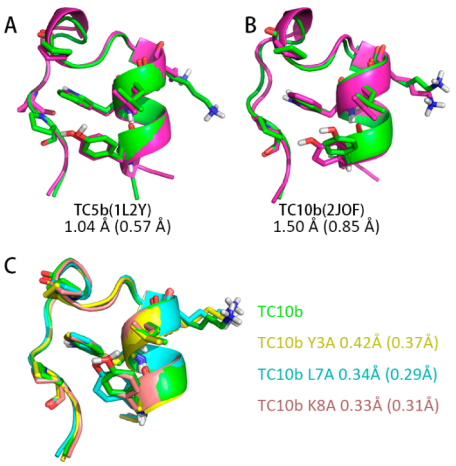


Figure 2. Representative structures of the largest clusters from the REMD simulations (300 K replica) of (A) TC5b, (B) TC10b, and (C) TC10b mutants. The predicted structures of TC5b and TC10b (green) are superimposed on corresponding PDB structures (magenta, PDB IDs in parentheses). The predicted structures of the three TC10b mutants are superimposed on the predicted TC10b structure. The RMSD values in parentheses are without the flexible two N-terminal and one C-terminal residues. The Y3, W6, L7, K8, P12, and P18 residues are shown in sticks.

variant.[67] This may also imply that the protein structure can be quite conserved upon point mutations.

Figure 3 shows the folding free energy landscapes (FELs) projected on two folding reaction coordinates: the $RMSD_{3-19}$

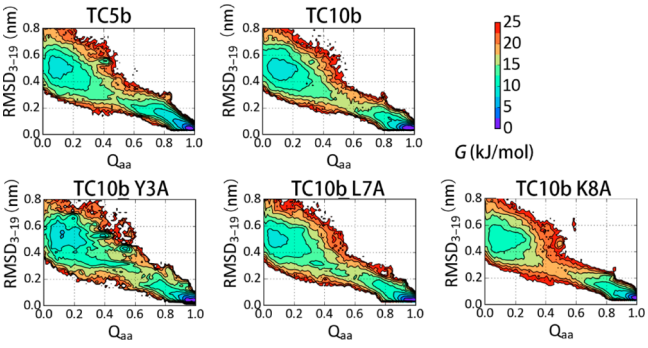


Figure 3. Folding free energy landscapes of the Trp-cage variants projected on the fraction of native contact (Q_{aa}) and the backbone RMSD of residues 3−19 ($RMSD_{3-19}$) calculated from structures sampled in replicas near their simulated T_m. Neighboring contours are separated by 2.5 kJ/mol, and the same scale is used throughout this work.

(the RMSD of backbone atoms within residues 3−19) and the fraction of native contacts (Q_{aa}). The five Trp-cage variants have similar FELs, each with two major basins separated by a barrier of 15−20 kJ/mol around Q_{aa} = 0.5. Compared with the other four variants, the most stable TC10b-K8A variant has a higher free energy barrier between the two basins. A deep and narrow basin around Q_{aa} > 0.9 and $RMSD_{3-19}$ < 1 Å indicates the low conformational flexibility of the Trp-cage fold. The basin of unfolded structures locates around Q_{aa} = 0.1 and $RMSD_{3-19}$ = 5 Å. Indeed, previous simulations found that extended starting structures can rapidly collapse into structures with RMSD values around 5 Å.[27,28,30] In some previous studies,[11] this equilibrated unfolded state was regarded as an "intermediate state" before the folding transition state (TS). However, a recent study indicates

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

that the Trp-cage unfolded state ensemble does not contain long-lived metastable states.[42] This agrees with the apparent two-state folding FELs from our simulations.

On the basis of the apparent two-state behavior of Trp-cage folding, we calculated the fractions folded ($Q_{aa} > 0.5$) at different temperatures, as shown in Figure 4B. The obtained melting

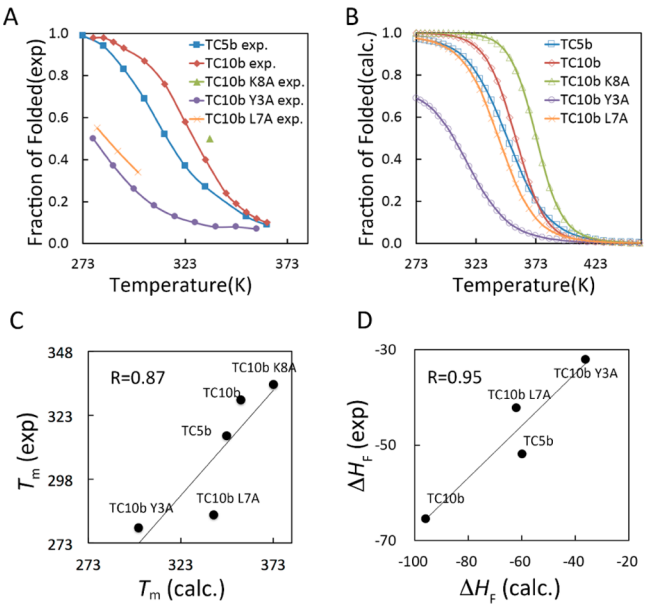


Figure 4. Folding thermodynamics from REMD simulations compared with experimental data: (A) experimental melting curves (only T_m was reported for TC10b-K8A), (B) calculated melting curves, (C) calculated folding midpoint temperatures T_m, and (D) folding enthalpies ΔH_F for the five variants plotted against the corresponding experimental data.

curves are reliable because there are many folding/unfolding events for each system (Table 2). For comparison, Figure 4A shows the melting curves characterized in experiments.[7,17] Our simulations give narrow temperature ranges of transition, indicating the cooperativity of Trp-cage folding/unfolding is well-described with the RSFF2 force field. As shown in Figure 4C and D, we compared the T_m and folding enthalpies ΔH_F fitted from simulated melting curves with the corresponding experimental data. A good linear correlation between calculated T_m and corresponding experimental T_m is observed with correlation coefficient $R = 0.87$. The systematic overestimation of T_m is also observed in our previous simulations using the RSFF1 force field.[49,52] This may be related to some imbalance between protein−protein and protein−water interactions resulting from current nonbonded parameters,[68,69] which are not improved in our RSFF1 and RSFF2 parametrization. The calculated ΔH_F also correlate very well with experimental values ($R = 0.95$), although the calculated absolute values of ΔH_F are systematically higher. This is different from previous REMD simulations using AMBER ff99SB and ff99SB*-ILDN force fields, which underestimate the $|\Delta H_F|$ of TC5b and TC10b.[38,56,58] The residue-specific conformational parameters used in our force field may give better compatibility between sequence and structure, resulting in more favorable energetics for folding.

The sequences of all five variants differ in the N-terminal (α-helix) part. To study the effects of these mutations on the intrinsic stability of the helix segment, we carried out additional REMD simulation of the fragment of residues 1−10 for each variant. In each fragment, the N-terminal was left uncapped, and the C-terminal was amidated. The overall helicities of the five fragments are in the order: TC10b-K8A ≈ TC10b > TC10b-Y3A ≈ TC5b > TC10b-L7A (Figure S1). In the hydrophobic core of the Trp-cage, Y3, W6, and L7 are all in contact with each other (Figure 1). These interactions will be preserved in the isolated helix formation. W6 ↔ L7 interaction may not stabilize the helix structure, but Y3 ↔ W6 and Y3 ↔ L7 are helix-stabilizing $i \leftrightarrow i+3$ and $i \leftrightarrow i+4$ interactions.[70] Although the Ala residue has a significantly higher helix propensity than the Tyr residue, the Y3A mutation will abolish the two helix-stabilizing side-chain packings. For the L7A mutation, the helix propensity of Ala is only slightly higher than that of Leu, but this mutation will abolish the Y3 ↔ L7 packing. This order of intrinsic stability of the helix segment is in rough agreement with the experimentally observed T_m values, except for TC10b-Y3A. Both experimental and calculated T_m of TC10b-Y3A are much lower than those of TC5b, although they have a similar intrinsic helix stability. This indicates the loss of nonlocal side-chain packings of Y3 with the poly-Pro segment also contributes to the low stability of the Y3A mutant.

Order of Local Structure Formation during Folding. The results described above indicate that our force field is reasonably accurate, allowing us to further study the folding mechanism of the Trp-cage. To quantify the order of the secondary structure formation, inspired by the previous work of Lindorff-Larsen et al.,[2] we calculated the backbone RMSD of every consecutive five-residue segment ($RMSD_5$) for structures in five different Q_{aa} ranges. It has been shown that Q_{aa} can be used as a good folding reaction coordinate for mechanistic studies.[66] As shown in Figure 5, for all five variants, the $RMSD_5$ of the C-terminal P_{II}-helix region is already quite low for $Q_{aa} < 0.2$ (fully unfolded) and does not change much compared with $Q_{aa} > 0.8$ (fully folded). As shown in Figure 6C, the backbone RMSD of residues 16−19 ($RMSD_{16-19}$) in the unfolded state of TC5b has a distribution quite similar to that in the folded state, except for a much less populated region of $RMSD_{16-19} > 1$ Å. The quite small difference between unfolded and folded structures in the poly-Pro region was also observed in a previous simulation.[2] Indeed, the decrease of conformational entropy of the unfolded state due to the relatively rigid poly-Pro segment has been regarded as one key contributing factor to the unusual stability of the Trp-cage fold.[7] This also explains the experimental finding that the ϕ-value

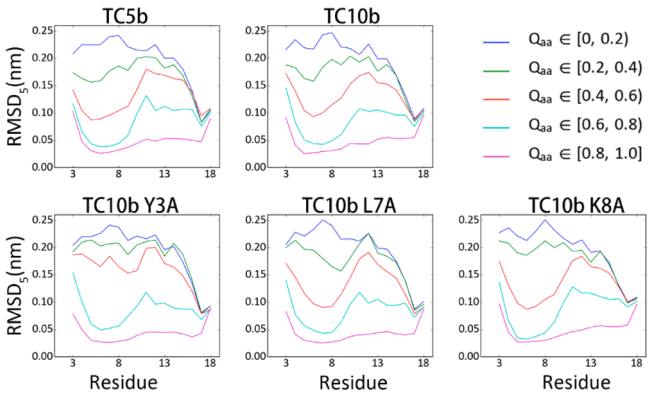


Figure 5. Average backbone RMSD of all consecutive 5-residue segments ($RMSD_5$) from structures within different Q_{aa} ranges sampled from near-T_m replicas. The $Q_{aa} < 0.2$ and $Q_{aa} > 0.8$ (top and bottom lines in each plot) correspond to well-unfolded and -folded structures, respectively, and the red lines correspond to the transition state. The x axis represents the index of the middle residue of each segment.

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

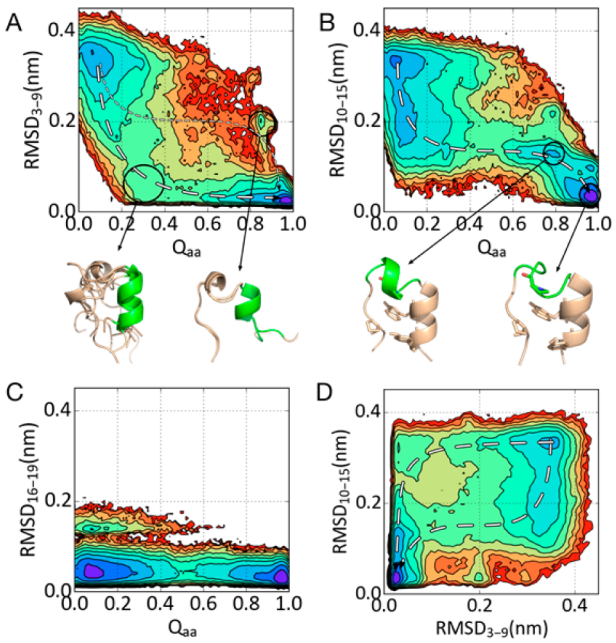


Figure 6. Free energy landscape of TC5b projected on the Q_{aa} and (A) $RMSD_{3-9}$, (B) $RMSD_{10-15}$, or (C) $RMSD_{16-19}$. The $RMSD_{3-9}$, $RMSD_{10-15}$, and $RMSD_{16-19}$ are the backbone RMSD values of residues 3−9, residues 10−15, and residues 16−19, respectively. Representative structures of some selected regions are also given. For each cartoon structure, the segment used for the RMSD calculation is shown in green. The arrows point from the unfolded state to the folded state.

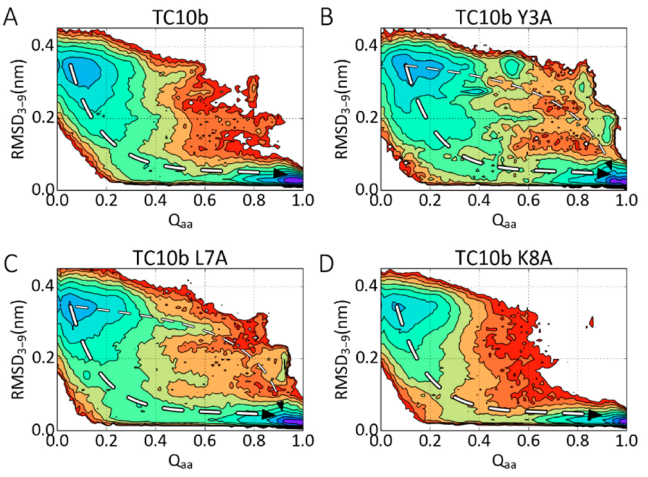


Figure 7. Free energy landscape of TC10b and its three mutants projected on the Q_{aa} and the $RMSD_{3-9}$. The arrows point from the unfolded to the folded state.

of the P19A mutant of TC10b is near zero[18] because such perturbation can have similar effects on the unfolded and transition states. Therefore, a small ϕ-value may not always indicate that the structure is formed after the folding TS.

From Figure 5, both the α-helix and the loop regions have average $RMSD_5$ > 2 Å for the well-unfolded structures and have $RMSD_5$ < 0.5 Å for the well-folded structures, indicating significant ordering during folding. Noticeably, the RMSD of the α-helix part has been found to be a key reaction coordinate for Trp-cage folding.[71] Here, in Figure 6A, we also give the free energy landscape projected on $RMSD_{3-9}$ and Q_{aa}. In agreement with Figure 5, $RMSD_{3-9}$ is usually quite large (>3 Å) for the unfolded basin. Also, the formation of α-helix occurs with a free energy cost of 8−10 kJ/mol from the most stable unfolded structures. The simulated low α-helicity in the unfolded state of the Trp-cage is consistent with the small difference between the experimental T_m of helix unfolding and that of tertiary structure (cage) unfolding (Table 1). Indeed, experimental studies of the N-terminal fragment of the Trp-cage did not show significant helicity.[72] However, except for TC10b-Y3A, the α-helix on average forms earlier than the loop, as indicated by a significant reduction of $RMSD_5$ of the α-helix segment when $0.4 < Q_{aa} < 0.6$ (Figure 5). In the dominant folding pathway of TC5b shown in Figure 6A, $RMSD_{3-9}$ decreases to <1 Å before Q_{aa} reaches 0.4 with the α-helix being well-formed in the representative structures. Besides the dominant pathway, there is still a possibility for the full α-helix to form quite late during folding. From Figure 6A, there is a very small basin with Q_{aa} > 0.8 and $RMSD_{3-9}$ around 2 Å within this minor folding pathway. The representative structures of this basin are quite similar to the native fold but with the N-terminal α-helix unraveled.

Figure 7 gives the $RMSD_{3-9}/Q_{aa}$ plots for TC10b and its mutants, and all are similar to TC5b with α-helix forming earlier as the major folding pathway. Like TC5b, there is also a possibility of folding through a minor pathway for TC10b-Y3A and TC10b-L7A. However, this minor pathway is not observed for TC10b and TC10b-K8A folding. As we found, the Y3A and L7A mutations make the α-helix part intrinsically less stable. Nevertheless, the barriers for this minor pathway seem to be much higher, indicating a quite low possibility of going through this pathway.

We also calculated the $RMSD_{10-15}$ to measure the formation of the native loop structure (Figure 6B). Interestingly, the loop region adopts the native structure quite late after the TS, only when Q_{aa} > 0.8. This explains the experimental finding that the stabilizing effect of the G10 to D-amino acid mutations mainly originates from a decrease in the unfolding rate, instead of accelerating the folding.[24] It is possible that the native loop structure is intrinsically unstable, which can only be stabilized by well-formed native contacts. An intermediate state very similar to the native state except for the non-native loop can be observed around Q_{aa} = 0.8 and $RMSD_{10-15}$ = 1.4 Å. This state has considerable population with free energy only ~3 kJ/mol higher than the folded state. This agrees with a recent experiment implying a near-native intermediate state of TC5b[19] and the unfolding of 3_{10}-helix at a temperature lower than its overall T_m.[18] Some previous simulations already indicated that the destabilization of 3_{10}-helix occurs very early during overall unfolding.[38] Interestingly, near-native metastable states have also been observed in the simulations of other small proteins.[73] It has been proposed that the existence of a post-TS intermediate state does not contradict experimentally observed two-state folding kinetics.[74] However, for the Trp-cage, there seems to be no significant barrier between this state and the native state. In addition, the loop structure in this state is also different from the most favored structures adopted in the unfolded state, which are more extended with $RMSD_{10-15}$ > 3 Å. Compared with TC5b, this near-native intermediate state is less populated for TC10b and its mutants (Figures S2 and S3). This agrees with previous equilibrium REMD simulations by Garcia and co-workers[38,56] in which a minor substate with RMSD of ~1.7 Å is observed for TC5b but not for TC10b.

Formation of Nonlocal Contacts during Folding. To provide further understanding of Trp-cage folding, we analyzed the distance between the side-chain center of Trp-6 and that of

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

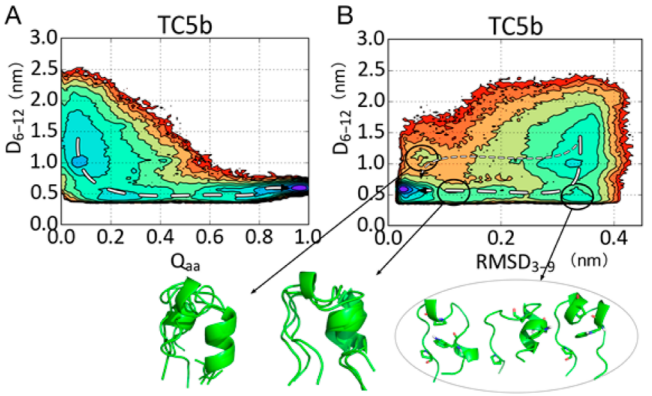


Figure 8. Free energy landscape of TC5b projected on the distance between side-chain centers of W6 and P12 (D_{6-12}) and (A) Q_{aa} or (B) $RMSD_{3-9}$.

Pro-12 (D_{6-12}). As shown in Figure 8A, we can find two free energy basins in the unfolded state. One with D_{6-12} > 7 Å, indicating no W6−P12 contact. The other one has D_{6-12} around 5 Å, which is even slightly smaller than that in the native state. Also, this W6−P12 contact occurs even earlier than the α-helix formation (Figure 8B). This agrees with a previous NMR experiment of unfolded TC5b by Mok et al., where a smaller proton−proton distance between W6 and P12 is observed compared with the native structure.[15] However, unlike the folding TS, most unfolded structures still seem to lack the W6−P12 contact in our simulations. Clustering analysis of the region with D_{6-12} < 6 Å and $RMSD_{3-9}$ > 3 Å did not give a single dominant structure, indicating diverse conformations. However, from the representative structures, the native topology seems to be formed before the α-helix formation. Indeed, in the unfolded state, the short α-helix structure is quite unstable, whereas the formation of hydrophobic contacts can be relatively easy. Also from Figure 8B, a minor folding path with α-helix formation before the W6−P12 contact is still possible, but the population of the key intermediate is rather low. For TC10b and its three mutants, the W6−P12 contact is also well formed before the α-helix formation in the dominant folding pathway (Figure S4). Interestingly, as Q_{aa} increases from 0.7 to 1.0 after the folding TS, D_{6-12} increases slightly to ~6 Å (the native basin). By inspecting corresponding structures, we found that the formation of native β-turns in the loop pushed the P12 slightly away from W6. This is in agreement with the previous experimental finding by Ahmed et al.[10] that the W6 becomes more exposed in the folded state than the intermediate state.

From the FELs projected on the Q_{aa} and the distances between W6 and P17/P18/P19 (Figure 9), both W6−P18 and W6−P19 distances decrease to the distances in the native state before reaching the TS (Q_{aa} = 0.5). Similarly for TC10b and its mutants, the D_{6-19} decreases to the native-like values at Q_{aa} = 0.5 in the major folding pathway (Figure S5). Interestingly, in Figure 9, we can see that the P17 side-chain can have close contact with W6 (state A) in the unfolded state and up to Q_{aa} = 0.6. This contact is non-native because in the native structure P17 is away from W6. It seems that the non-native P17−W6 contact should break to go through the TS on the folding path (state B). When W6 forms face-to-face contact with P17 in state A, it cannot be in contact with P19. In the representative structure of state A, the polar hydrogen in the W6 side-chain forms a H-bond with the G15 backbone O atom instead of forming a native H-bond with R16. Still, in both state A and state B, W6 is in contact with P18. The

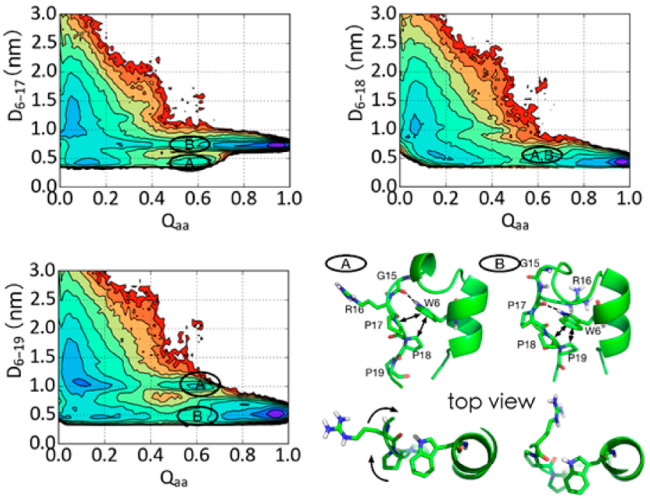


Figure 9. Free energy landscape of TC5b projected on the Q_{aa} and the distances between side-chain centers of W6 and P17/P18/P19. The representative structures with hydrophobic contact between W6−P17 (state A) and W6−P19 (state B) are shown below. The arrows point from the unfolded to the folded state.

off-pathway intermediate state also exists for TC10b and its three single mutants but with less populations (Figure S6).

From our simulations, both the α-helix and the W6−P19 contact usually form well at the folding TS. Which one forms earlier? In the FELs projected on D_{6-19} and $RMSD_{3-9}$ (Figure S6) of all five variants, two parallel pathways are clearly observed. This is different from the situation for W6−P12 contact, which usually forms earlier than α-helix. Understandably, P19 is more distant from W6 than P12 in sequence, so the W6−P19 contact is entropically less favored than the W6−12 contact. Nevertheless, these key hydrophobic contacts can form quite well at Q_{aa} = 0.5 in the major folding pathway. Conversely, the formation of some key polar contacts is relatively late. As shown in Figure 10A, the probability of forming the native side-chain-to-backbone H-bond between W6 and R16 at Q_{aa} = 0.5 is less than 40%. Also, the probability of forming a salt-bridge between D9 and R16 is relatively low until Q_{aa} > 0.7 (Figure 10B). This agrees well with

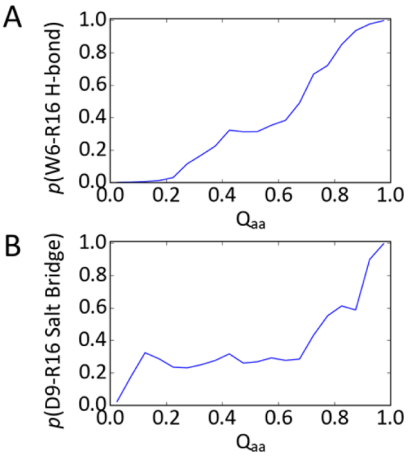


Figure 10. Fractions of structures with (A) hydrogen bonding between W6 side-chain H atom and R16 backbone O atom and (B) salt bridge between D9 and R16 side-chains at different Q_{aa} values obtained from the REMD simulation of TC5b.

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

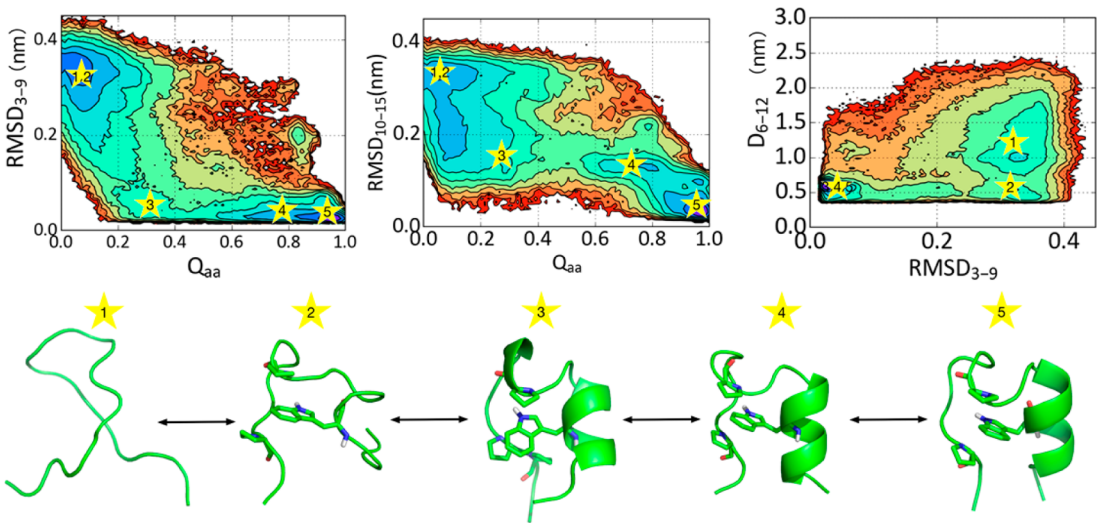


Figure 11. Five stages of Trp-cage folding on the free energy landscapes (FELs) projected on three different sets of order parameters with the five corresponding representative structures shown below. There are roughly four steps on the major pathway: (1) the formation of key hydrophobic contacts, (2) the formation of α-helix, (3) crossing the TS with the formation of most native contacts, and (4) the formation of a native loop structure. The FELs of TC5b are used as an example.

the previous finding that there is no significant change in folding time upon the protonation of D9.[26]

SUMMARY

We have carried out microsecond REMD simulations of five Trp-cage variants using our recently developed RSFF2 force field. Starting from unfolded structures, the experimental structures of TC5b and TC10b can be excellently reproduced, and the native structures of the three TC10b mutants (Y3A, L7A, K8A) are nearly identical to that of TC10b. All five Trp-cage variants have overall two-state folding behavior. Nevertheless, this does not preclude the existence of metastable intermediates along the folding pathway.[74] From the equilibrated melting curves, our simulations slightly overestimate the melting temperatures T_m and absolute folding energies $-\Delta H_F$ but give good linear correlations with the corresponding experimental data of the five variants. Our study indicates that it is now possible to predict the effect of mutations on the stability of a protein through physics-based folding simulations using the enhanced sampling method and improved force field.

We further explored the folding free energy landscapes, and a detailed folding mechanism can be proposed. As a simplified model, the major folding pathway can be divided into five stages (Figure 11). Folding starts with unfolded structures (stage 1) without the α-helix formation and the native contacts but with readily formed C-terminal P_{II}-helix. Then, within the unfolded state, the central Trp residue (W6) makes contacts with some Pro residues (stage 2), driving the formation of the α-helix. The α-helix can be well-formed just before the TS is reached (stage 3). In this stage, although the W6 is in contact with some surrounding hydrophobic residues, the side-chains are not well packed ($Q_{aa} < 0.5$) with low probability of forming the native salt-bridge and the native H-bond from the W6 side-chain. Still, most important structure-ordering events occur before the TS. After the TS, the structure becomes very close to the native one except for the loop region (stage 4). The final step is the formation of the native loop structure and the D9−R16 salt-bridge from this near-native intermediate state (stage 5).

Comparing between different mutants, the less stable mutants (TC5b, TC10b-Y3A, TC10b-L7A) possibly have minor folding pathways in which the formation of the native α-helix occurs rather late ($Q_{aa} > 0.8$). It is possible that the mutation destabilizing a secondary structure element may also make it form relatively late in the process of folding. In addition, we found that TC5b has higher populations of both the near-native intermediate state and the W6−P17 mispacked (off-pathway) intermediate state compared with TC10b and its single mutants.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00581.

The simulated helicities of fragment 1−10 for each variant and more free energy landscapes for TC10b and its three mutants (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jiangfan@pku.edu.cn.
*E-mail: wuyd@pkusz.edu.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Y.-D.W. and F.J. are supported by the National Natural Science Foundation of China (Grant No. 21133002 and 21203004). Y.-D.W. is also supported by the Shenzhen Peacock Program (KQTD201103) and Peking University Shenzhen Graduate School. We are also thankful for financial support from the Nanshan District in Shenzhen (KC2014ZDZJ0026A).

REFERENCES

(1) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042−1046.
(2) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517−520.
(3) Das, R. *PLoS One* **2011**, *6*, e20044.
(4) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *PLoS One* **2012**, *7*, e32131.

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(5) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4925−4930.
(6) Yang, L.; Shao, Q.; Gao, Y. Q. *J. Phys. Chem. B* **2009**, *113*, 803−808.
(7) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(8) Williams, D. V.; Byrne, A.; Stewart, J.; Andersen, N. H. *Biochemistry* **2011**, *50*, 1143−1152.
(9) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(10) Ahmed, Z.; Beta, I. A.; Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943−10950.
(11) Neuweiler, H.; Doose, S.; Sauer, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16650−16655.
(12) Iavarone, A. T.; Parks, J. H. *J. Am. Chem. Soc.* **2005**, *127*, 8606−8607.
(13) Bunagan, M. R.; Yang, X.; Saven, J. G.; Gai, F. *J. Phys. Chem. B* **2006**, *110*, 3759−3763.
(14) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876−2880.
(15) Mok, K. H.; Kuhn, L. T.; Goez, M.; Day, I. J.; Lin, J. C.; Andersen, N. H.; Hore, P. J. *Nature* **2007**, *447*, 106−109.
(16) Hudáky, P.; Stráner, P.; Farkas, V.; Váradi, G.; Tóth, G.; Perczel, A. *Biochemistry* **2008**, *47*, 1007−1016.
(17) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W. *Protein Eng., Des. Sel.* **2008**, *21*, 171−185.
(18) Culik, R. M.; Serrano, A. L.; Bunagan, M. R.; Gai, F. *Angew. Chem., Int. Ed.* **2011**, *50*, 10884−10887.
(19) Rodriguez-Granillo, A.; Annavarapu, S.; Zhang, L.; Koder, R. L.; Nanda, V. *J. Am. Chem. Soc.* **2011**, *133*, 18750−18759.
(20) Heyda, J.; Kožíšek, M.; Bednárova, L.; Thompson, G.; Konvalinka, J.; Vondrášek, J.; Jungwirth, P. *J. Phys. Chem. B* **2011**, *115*, 8910−8924.
(21) Halabis, A.; Zmudzinska, W.; Liwo, A.; Oldziej, S. *J. Phys. Chem. B* **2012**, *116*, 6898−6907.
(22) Rogne, P.; Ozdowy, P.; Richter, C.; Saxena, K.; Schwalbe, H.; Kuhn, L. T. *PLoS One* **2012**, *7*, e41301.
(23) Rovó, P.; Stráner, P.; Láng, A.; Bartha, I.; Huszár, K.; Nyitray, L.; Perczel, A. *Chem. - Eur. J.* **2013**, *19*, 2628−2640.
(24) Culik, R. M.; Annavarapu, S.; Nanda, V.; Gai, F. *Chem. Phys.* **2013**, *422*, 131−134.
(25) Meuzelaar, H.; Marino, K. A.; Huerta-Viga, A.; Panman, M. R.; Smeenk, L. E.; Kettelarij, A. J.; van Maarseveen, J. H.; Timmerman, P.; Bolhuis, P. G.; Woutersen, S. *J. Phys. Chem. B* **2013**, *117*, 11490−11501.
(26) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. *Biochemistry* **2014**, *53*, 6011−6021.
(27) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.
(28) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711−717.
(29) Zhou, R. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13280−13285.
(30) Chowdhury, S.; Lee, M. C.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 13855−13865.
(31) Juraszek, J.; Bolhuis, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15859−15864.
(32) Paschek, D.; Nymeyer, H.; Garcia, A. E. *J. Struct. Biol.* **2007**, *157*, 524−533.
(33) Paschek, D.; Hempel, S.; Garcia, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17754−17759.
(34) Xu, W.; Mu, Y. *Biophys. Chem.* **2008**, *137*, 116−125.
(35) Hu, Z.; Tang, Y.; Wang, H.; Zhang, X.; Lei, M. *Arch. Biochem. Biophys.* **2008**, *475*, 140−147.
(36) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Genet.* **2009**, *76*, 448−460.
(37) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A. *PLoS Comput. Biol.* **2009**, *5*, e1000452.
(38) Velez-Vega, C.; Borrero, E. E.; Escobedo, F. A. *J. Chem. Phys.* **2010**, *133*, 105103.
(39) Day, R.; Paschek, D.; Garcia, A. E. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1889−1899.
(40) Zheng, W.; Gallicchio, E.; Deng, N.; Andrec, M.; Levy, R. M. *J. Phys. Chem. B* **2011**, *115*, 1512−1523.
(41) Shao, Q.; Shi, J.; Zhu, W. *J. Chem. Phys.* **2012**, *137*, 125103.
(42) Deng, N. J.; Dai, W.; Levy, R. M. *J. Phys. Chem. B* **2013**, *117*, 12787−12799.
(43) Han, W.; Schulten, K. *J. Phys. Chem. B* **2013**, *117*, 13367−13377.
(44) Kannan, S.; Zacharias, M. *PLoS One* **2014**, *9*, e88383.
(45) Bowman, G. R.; Pande, V. S. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10890−10895.
(46) Rovó, P.; Farkas, V.; Hegyi, O.; Szolomájer-Csikós, O.; Tóth, G. K.; Perczel, A. *J. Pept. Sci.* **2011**, *17*, 610−619.
(47) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.
(48) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7587−7592.
(49) Jiang, F.; Zhou, C.; Wu, Y. *J. Phys. Chem. B* **2014**, *118*, 6983−6998.
(50) Zhou, C. Y.; Jiang, F.; Wu, Y. D. *J. Phys. Chem. B* **2015**, *119*, 1035−1047.
(51) Jiang, F.; Han, W.; Wu, Y. D. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3413−3428.
(52) Jiang, F.; Wu, Y. D. *J. Am. Chem. Soc.* **2014**, *136*, 9536−9539.
(53) Li, S.; Elcock, A. H. *J. Phys. Chem. Lett.* **2015**, *6*, 2127−2133.
(54) Sikosek, T.; Chan, H. S. *J. R. Soc., Interface* **2014**, *11*, 20140419.
(55) Kroncke, B. M.; Vanoye, C. G.; Meiler, J.; George, A. J.; Sanders, C. R. *Biochemistry* **2015**, *54*, 2551−2559.
(56) English, C. A.; Garcia, A. E. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2748−2757.
(57) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(58) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17845−17850.
(59) Prakash, M. K.; Barducci, A.; Parrinello, M. *J. Chem. Theory Comput.* **2011**, *7*, 2025−2027.
(60) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(61) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(63) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(64) Lin, I.-C.; Tuckerman, M. E. *J. Phys. Chem. B* **2010**, *114*, 15935−15940.
(65) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(66) Best, R. B.; Hummer, G.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.
(67) Scian, M.; Lin, J. C.; Le Trong, I.; Makhatadze, G. I.; Stenkamp, R. E.; Andersen, N. H. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 12521−12525.
(68) Best, R. B.; Zheng, W.; Mittal, J. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(69) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(70) Padmanabhan, S.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 1992−1997.
(71) Juraszek, J.; Bolhuis, P. G. *Biophys. J.* **2008**, *95*, 4246−4257.
(72) Stewart, J. M.; Lin, J. C.; Andersen, N. H. *Chem. Commun.* **2008**, 4765−4767.
(73) Beauchamp, K. A.; McGibbon, R.; Lin, Y. S.; Pande, V. S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17807−17813.
(74) Brockwell, D. J.; Radford, S. E. *Curr. Opin. Struct. Biol.* **2007**, *17*, 30−37.

DOI: 10.1021/acs.jctc.5b00581
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX