University of Massachusetts Amherst

ScholarWorks@UMass Amherst

2018

Atomistic Peptide Folding Simulations Reveal Interplay of Entropy and Long-Range Interactions in Folding Cooperativity

Jianlin Chen
The Central Hospital of Taizhou

Xiaorong Liu
University of Massachusetts Amherst

Jianhan Chen
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/chem_faculty_pubs

Recommended Citation

Chen, Jianlin; Liu, Xiaorong; and Chen, Jianhan, "Atomistic Peptide Folding Simulations Reveal Interplay of Entropy and Long-Range Interactions in Folding Cooperativity" (2018). *Scientific Reports*. 1453.
https://doi.org/10.1038/s41598-018-32028-7

SCIENTIFIC REPORTS

OPEN

Atomistic Peptide Folding Simulations Reveal Interplay of Entropy and Long-Range Interactions in Folding Cooperativity

Jianlin Chen[1], Xiaorong Liu[2] & Jianhan Chen [2,3]

Understanding how proteins fold has remained a problem of great interest in biophysical research. Atomistic computer simulations using physics-based force fields can provide important insights on the interplay of different interactions and energetics and their roles in governing the folding thermodynamics and mechanism. In particular, generalized Born (GB)-based implicit solvent force fields can be optimized to provide an appropriate balance between solvation and intramolecular interactions and successfully recapitulate experimental conformational equilibria for a set of helical and β-hairpin peptides. Here, we further demonstrate that key thermodynamic properties and their temperature dependence obtained from replica exchange molecular dynamics simulations of these peptides are in quantitative agreement with experimental results. Useful lessons can be learned on how the interplay of entropy and sequentially long-range interactions governs the mechanism and cooperativity of folding. These results highlight the great potential of high-quality implicit solvent force fields for studying protein folding and large-scale conformational transitions.

The protein folding problem, i.e., understanding how natural proteins fold reliably into unique three-dimensional structures, remains one of the major tasks in molecular biology[1]. Significant advances in both theoretical and experimental aspects have been made in the last few decades and the general mechanism of protein folding is now considered to be known[2–5]. In particular, the statistical mechanics view, i.e., the energy landscape theory, has become a central framework for understanding protein folding, structure prediction and protein design[6–8]. The key ideas are that the global energy landscape of a naturally evolved protein resembles a partially rugged funnel and that the protein folds through multiple routes to the basin of global free energy minimum. The state-of-the-art explicit solvent protein force fields have also steadily improved over the years[9–13], and achieved remarkable successes in recent atomistic protein folding simulations[5,14]. Nevertheless, a complete and unambiguous picture of protein folding at the atomic level has yet to emerge. For example, there are still controversies even on how a 16-residue β-hairpin from the C-terminus of the protein G B1 domain (GB1p) folds in isolation[15–25]. This reflects some of the limitations in current computational and experimental methodologies. Direct all-atom molecular dynamics (MD) or Monte Carlo (MC) simulations can provide insight at levels of detail and fast timescales that cannot be currently reached by experiments, but remain limited by accessible simulation timescales and force field accuracy. The speed limit of the fastest folding proteins ranges from a few microsecond (μs) to tens of μs[26], mostly out of direct reach of statistically meaningful atomistic simulations when solvent is treated explicitly. As such, efficient implicit solvent models, where average properties of water are described macroscopically, can be particularly useful in reducing the timescale gap[27–32].

An important advance in the experimental frontier of protein folding studies is the discovery and design of ultrafast folding peptides and proteins[33]. These systems are more amenable to detailed simulations and provide a unique opportunity to bridge the gap between theoretical and experimental investigations. In particular, empirical protein force fields, originally calibrated to characterize conformations near the native basin, have important limitations for protein folding studies[34,35]; virtually all of them tend to generate overly compact ensembles for unfolded

[1]Department of Hematology, The Central Hospital of Taizhou, Taizhou, Zhejiang, 318000, P.R. China. [2]Department of Chemistry, University of Massachusetts Amherst, Amherst, MA, 01003, USA. [3]Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, MA, 01003, USA. Correspondence and requests for materials should be addressed to J.C. (email: jianhanc@umass.edu)
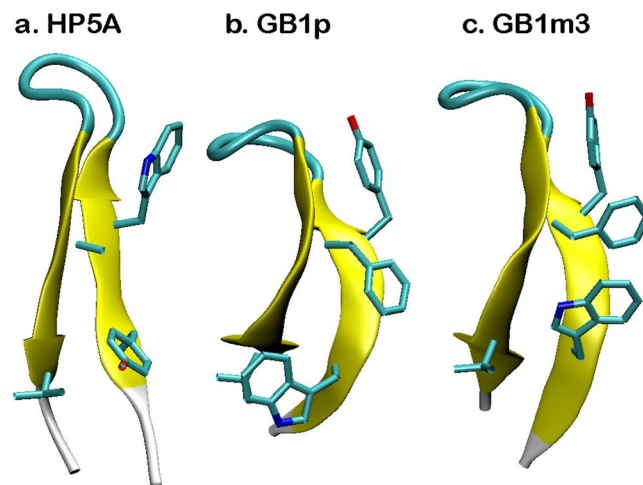
Figure 1. Representative folded structures of GB1p β-hairpins at 270 K. Key hydrophobic sidechains are shown in bonds. Peptide sequences and native contacts are given under Methods.

protein states[36–38]. Availability of fast folding peptides makes it more feasible to recalibrate these force fields to include the non-native/unfolded conformational space[39]. Such optimization requires extensive conformational sampling and can substantially benefit from enhanced sampling methods, particularly the replica exchange (REX) class of sampling techniques[40–42]. For example, we have previous parameterized a generalized Born with smooth switching (GBSW)[43] implicit solvent force field, by rebalancing solvation and intramolecular interactions to reproduce the experimental conformational equilibria of two helical peptides and a range of β-hairpins derived from GB1p[44,45]. These hairpins are sequentially homologous but display either reduced or enhanced stability compared to the native sequence[46], offering excellent controls for optimization of various implicit and explicit solvent force fields[9,13,47,48]. The optimized GBSW force field demonstrates a good level of transferability, successfully folding both trpzip2[49], a designed β-hairpin, and Trp-cage[50], a designed mini-protein, to ~1.0 Å accuracy. It has also been successfully applied to calculate the conformational equilibria of small proteins[51–53] and to describe non-trivial structural features of several IDPs[54–57].

In the present study, we apply the optimized GBSW protein force field to examine the detailed thermodynamics and mechanisms of β-hairpin formation using the set of GB1p-derived peptides. Despite their small size, these hairpins are good model systems for understanding the folding of larger proteins, with cooperative folding transitions at multi-μs timescales and involvement of sequentially long-range hydrophobic and backbone hydrogen-bonding (H-bond) interactions[15]. The findings here are also made more relevant with the use of a single physics-based force field tuned to preserve the delicate balance of competing interactions. Coupling implicit solvent and REX enhanced sampling allows derivation of key thermodynamic properties with good statistical convergence for direct comparison with available experimental data on these peptides. Together, the results reveal an interesting interplay between chain entropy and long-range interactions in determining the cooperativity and mechanism of peptide folding.
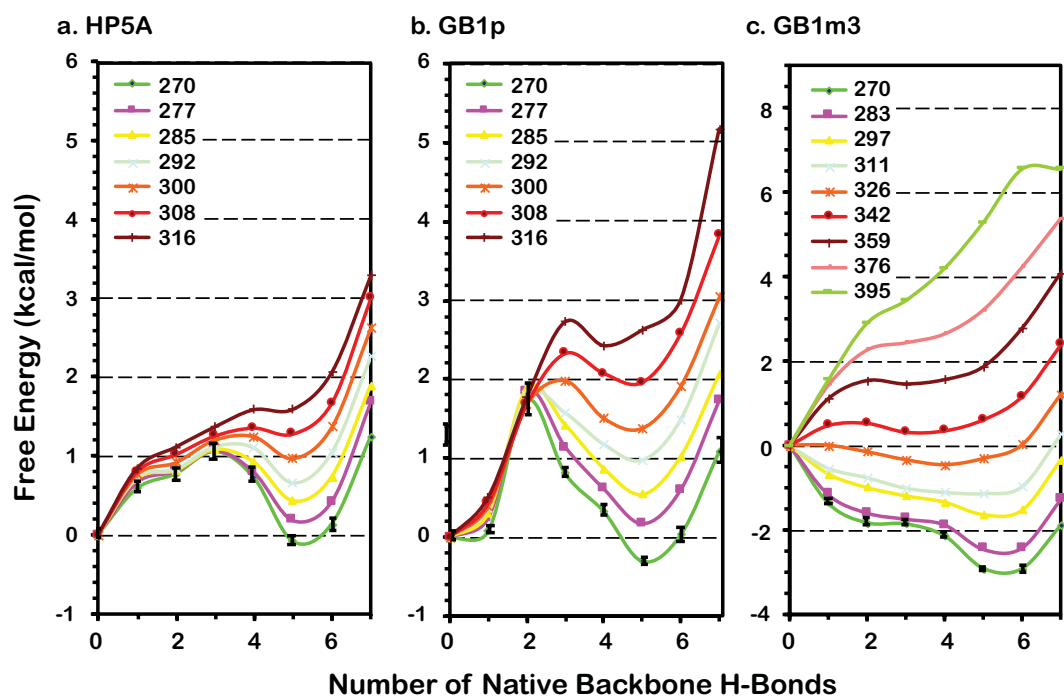
Results

Structure and energetics of GB1p series β-hairpin folding. All three GB1p-series peptides can fold into β-hairpin structures in the GBSW implicit solvent (see Fig. 1), with the same backbone H-bond registry as observed in the intact protein G B1 domain. Key hydrophobic sidechains also form cross-strand contacts as expected. Interestingly, the hairpin structures formed by these peptides display significant differences in the level of curving, with HP5A showing the least bending. As it will be discussed later, this is likely due to different relative importance of hydrophobic side chain interactions and backbone H-bonds. The force field also correctly recapitulates the experimental observation that the folded hairpin state is the most stable for GB1m3 and least stable for HP5A[49]. Examination of the free energy profiles as a function of the number of native backbone H-bonds formed (N_{HB}), shown in Fig. 2, suggests that all three hairpins fold cooperatively. It has been demonstrated that the number (or fraction) of native contacts formed provides an excellent reaction coordinate for describing protein folding using both coarse-grained[58] and atomistic[26] simulations. For both GB1p and HP5A, there are two well-defined folded and unfolded free energy minima, separated by a single major barrier. That is, both peptides display strong cooperative folding, with most native structural features formed together. This is highly consistent with previous experimental studies demonstrating that the isolated GB1p hairpin retains the folding cooperativity[15,18]. For GB1m3, there is a single free energy minimum that corresponds to the folded state at low temperatures, suggesting that it may follow 'downhill folding'[59]. However, a mild free energy barrier is present near the melting temperature (T_m), around 330 K (e.g., see the T = 342 K trace in Fig. 2c).

The cooperative folding behavior allows one to derive folding thermodynamics directly from the simulated ensembles sampled during REX-MD simulations (see Methods for details). The results for 270 K are summarized in Table 1 and compared with existing experimental data. Implicit solvent simulations were able to reproduce the order of stabilities of these hairpins, which is expected as they have been used in GBSW force field optimization[45]. Importantly, both folding enthalpy and folding entropy of GB1p derived from simulations are in

	ΔG (kcal/mol)		ΔU (kcal/mol)		ΔS (cal/mol/K)		T_m (K)	
	MD	Expr[a]	MD	Expr	MD	Expr	MD	Expr
HP5A	0.35 ± 0.1	0.82 ± 0.4	-2.4 ± 0.4		-7.6 ± 4	—	<270	<273[46]
GB1p	-0.44 ± 0.3	-0.8 ± 0.6	-9.9 ± 0.8	-11.6[15]	-38 ± 4	-39[15]	~273	<278[46]
				-12.6 ± 1.2[18]		-43 ± 4[18]		293 ± 4.3[18]
GB1m3	-1.3 ± 0.2	-1.1 ± 0.2	—	—	~-7.6	—	~330	333 ± 2[46]

Table 1. Key folding thermodynamic parameters of β-hairpins at 270 K. Standard errors were estimated based on the differences calculated using first and second half of the production ensembles (last 60 ns of REX-MD). [a]Estimated based on the folded population as $\Delta G = -RT \ln p_f/(1 - p_f)$[15,18,46]. Uncertainties are estimated using the range of reported p_f.


Figure 2. Free energy profiles of GB1p β-hairpins at different temperatures. All profiles were calculated directly from distributions sampled at various temperatures during the last 60 ns of REX-MD. Error bars are shown only for results at 270 K for clarity, which are standard errors estimated based on the differences calculated using first and second half of the production ensembles.

quantitative agreement with values derived from NMR and T-jump tryptophan fluorescence experiments[15,18,46]. We note that implicit-solvent derived folding entropies do not include contributions from solvent. Nonetheless, the agreement re-affirms that the optimized GBSW force field offers a realistic balance between solvent-mediated protein-protein interactions vs. conformational flexibility, which may be directly attributed to the optimization strategy that involves both calibration of pair-wise interactions of backbone and side chain moieties and calculation of conformational equilibria of carefully selected model peptides[45]. Consistent with realistic estimation of folding entropy, the melting temperatures derived from REX-MD are also highly consistent with experimental results for all three hairpins[46].

Even though both HP5A and GB1p hairpins are marginally stable (~35% vs ~65% folded), the underlying entropic and enthalpic contributions differ dramatically. The conformational entropic cost of folding for GB1p, ~38 cal/mol/K, is about 5-fold greater than for HP5A, which is only 7.6 cal/mol/K. This is mainly due to the proline-rigidified loop (-PATG-) in HP5A (and GB1m3), which was designed to replace the highly flexible -DATK- loop in the original GB1p sequence[46]. Analysis of the equilibrium conformational distributions, shown in Fig. 3, clearly shows that GB1p samples a wider conformational space compared to HP5A. In particular, HP5A samples two major nonnative basins beside the native state. Representative snapshots of these conformations are shown in Fig. S1, which suggests that the reduced conformational space sampled by HP5A can be primarily attributed to the loop. The loop forms a native-like turn in one of the nonnative states while adopting extended conformations in the other. This is also reflected in the loop distance distributions, where the -PATG- loop in HP5A occupied two major sub-states (dashed red line in Fig. 4) and has substantial probability of sampling folded

	vdW	ASP	Elec	GB Ener	Elec + GB	Total
HP5A	-3.5 ± 0.1	-0.10 ± 0.1	12.2 ± 1.6	-9.7 ± 2.1	2.5 ± 0.4	-2.4 ± 0.4
GB1p	-0.50 ± 0.06	-0.40 ± 0.1	-14.9 ± 1.7	7.9 ± 2.1	-7.0 ± 0.4	-9.9 ± 0.8

Table 2. Detailed folding energetics of HP5A and GB1p β-hairpins (all in kcal/mol). Standard errors were estimated based on the differences calculated using first and second half of the production ensembles (last 60 ns of REX-MD).
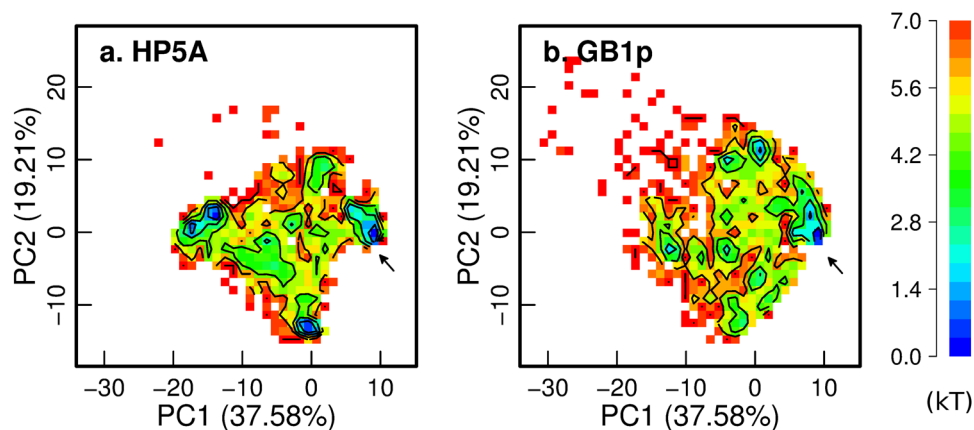

Figure 3. Free energy surfaces projected along the first two principal components. (**a**) HP5A and (**b**) GB1p. Only conformations sampled at 270 K during the last 60 ns of REX-MD were included and the principal components were derived by including backbone structures of both HP5A and GB1p. The arrows mark the location of folded basins. Contours are drawn at every kT.
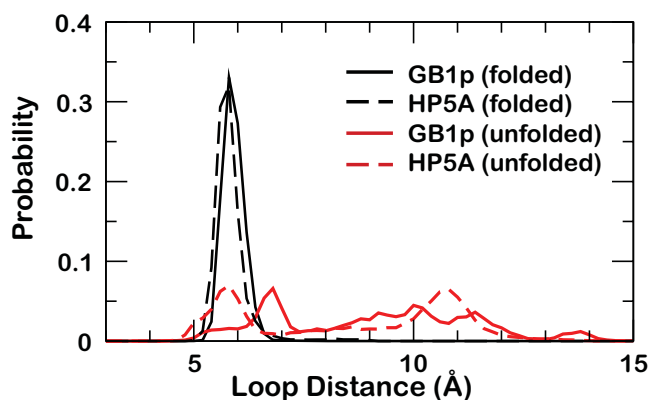
Figure 4. Distributions of loop distances in folded and unfolded GB1p and HP5A. The loop distance was calculated as the distance between Cα atoms of residues 6 and 11.

like conformations (black traces in Fig. 4). In comparison, the -DATK- loop in GB1p rarely adopts folded-like configurations and can sample a broad and largely continuous range of disordered states (solid red trace in Fig. 4).

Intriguingly, the enthalpic stabilization of folding, -2.4 ± 0.4 kcal/mol, is also much smaller for HP5A, which is only about one fourth of that of GB1p (-9.9 ± 0.8 kcal/mol). The design of HP5A sequence involves replacement of larger hydrophobic contacts (Trp-Val and Tyr-Phe in GB1p and GB1m3) with the smaller ones (Trp-Ala and Tyr-Val), and at the same time introduces two Lys residues in the N-terminus to potentially stabilize the folded state via salt-bridge interactions with negatively charged C-terminus[46]. Yet, analysis of the contributions of various interactions, summarized in Table 2, revealed that the largest contribution to larger folding enthalpy of GB1p came from solvent-screened electrostatic contributions, which is ~7.0 kcal/mol stabilizing vs ~2.5 kcal/mol de-stabilizing for HP5A. Contributions of salt-bridge interactions to protein stabilities are highly context dependent, and it has been shown that exposed ones (such as those could potentially form between N-terminal Lys residues and C-terminal Glu and the carboxyl group) are often destabilizing or contribute only slightly to protein stability[55,60,61]. This seems to be true for HP5A as well. Instead, presence of larger sidechains in GB1p (and GB1m3) reduces the solvent exposure of backbone H-bonds and enhances their stabilities, leading to a much larger total electrostatic contribution to the folding enthalpy of GB1p. These energetic effects also likely explain higher bending of the folded hairpin structures observed for GB1p and GB1m3 (Fig. 1), which apparently
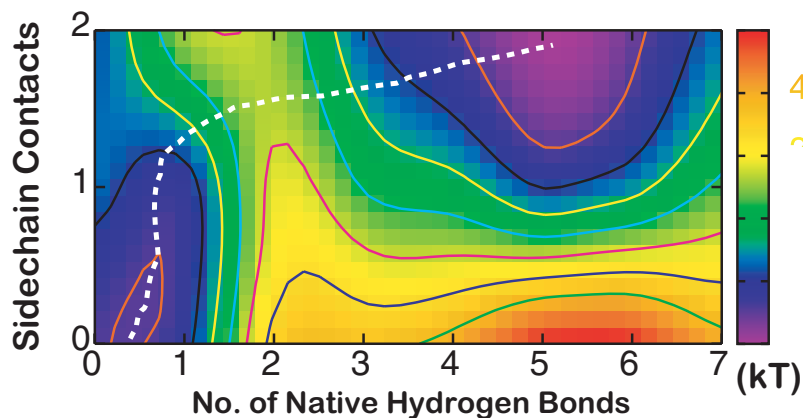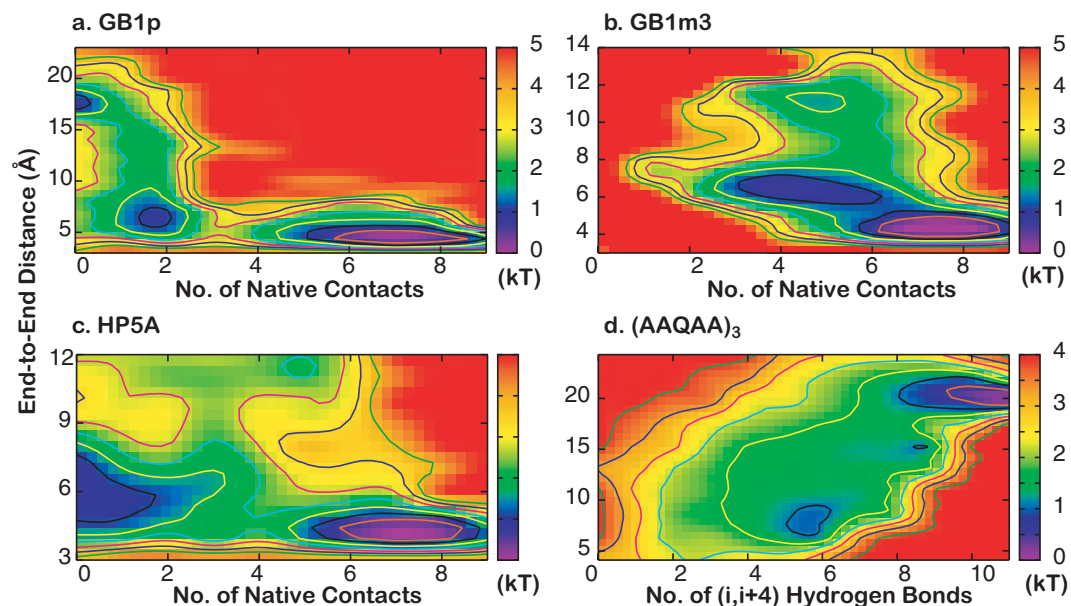
Figure 5. 2D Free energy profiles of GB1p as a function of the numbers of native side chain contacts formed and backbone H-bond formed. The dashed line illustrates the minimal energy path connecting unfolded and folded states. Contours are drawn at every kT.

increase the level of side chain-backbone contacts. We note that significant bending of GB1p hairpin structure was also observed in explicit solvent simulations[23].

Cooperative folding of GB1p: hydrophobic interactions vs backbone H-bonds. The ability of REX-MD simulations in GBSW implicit solvent to quantitatively recapitulate the stabilities and folding energetics of GB1p-series hairpins provides a unique opportunity to revisit several key questions regarding GB1p folding mechanism, particularly the roles of loop dynamics, hydrophobic side chain interactions and backbone H-bond formation. Folding of GB1p is clearly highly cooperative with a single major barrier separating the folded and unfolded states (Fig. 2), which is in agreement with existing experimental and computational studies[16,18,62,63]. Further examination of the 2D free energy surface demonstrates that the collapse of the peptide and formation of hydrophobic side chain contacts precede the backbone H-bond formation (e.g., see dashed line in Fig. 5). Such a sidechain driven cooperative folding mechanism is highly consistent previous NMR analyses[18,64]. A representative folding event sampled during REX-MD is shown in Fig. S2. Formation of native backbone H-bonds requires the loop making the correct turn and hydrophobic side chains making native-like contacts, which represent the rate limiting steps. Complete folding is almost always initiated by H-bonds near the turn, which leads to very fast zipping up of the rest of native backbone contacts (e.g., see Fig. S2). Consistent with simulations using latest optimized explicit solvent force fields[23], there are no partially helical intermediate state not well populated at equilibrium (Fig. S3). Previous observation of such states was likely a consequence of inherent biases in early force fields, which have now been shown to yield over-stabilized helical and nonspecific compact states[9–13]. We also note that, despite the appearance of a single dominant folding pathway in Fig. 5, specific native side chain and backbone contacts can form in different orders, giving rise to relatively diverse microscopic folding pathways even for these small hairpins.

Entropy, long-range interactions and folding cooperativity. The funneled energy landscape theory predicts that cooperativity of folding arises naturally as a consequence of imperfect cancellation of (conformational) entropy and enthalpy[6–8]. The three GB1b-series hairpins simulated here provide a nice case study that illustrates how the magnitude of conformational entropy and strengths and (topological) distribution of native (sequentially) long-range interactions together determines the level of folding cooperativity. Folding of GB1p is the most cooperative among the three hairpins. Formation of multiple energetically favorable contacts is required in order to overcome the large conformational entropy of folding due to the flexible loop, which enforces cooperative folding. As shown in Fig. 6a, the peptide can readily sample collapsed conformations in the unfolded state where both sidechain hydrophobic contacts are frequently formed. However, it requires formation of about two additional backbone H-bonds to fully overcome the entropic cost and precede to the fully folded state sharply downhill in free energy (Fig. 2b). For HP5A, the entropic cost of folding is much smaller (Table 1), but the net energetic stabilization effects of side chain and backbone native contacts are also much smaller (Table 2). This leads a more gradual compensation of the entropic cost through formation of native contacts and folding does not turn downhill until over three backbone H-bonds are formed together with both side chain contacts (Figs 2a and 6c). In contrast, GB1m3 contains the same rigidified loop of HP5A but retains large hydrophobic side chains of GB1p. As a result, the modest entropic cost of folding can be readily compensated by formation of side chain hydrophobic interactions alone and the folding is virtually downhill as a function of the number of backbone H-bonds (Fig. 2c). Examination of the 2D free energy surface as a function of the end-to-end distance and number of native contacts also reveal minimal barriers (Fig. 6b). It resembles the surface for folding of helical peptide (AAQAA)₃, where the entropic cost of helix-coil transitions is continuously compensated by formation of additional backbone H-bonds in a largely downhill fashion towards the folded state (Fig. 6d). However, there remains an important difference between folding of (AAQAA)₃ helix and GB1m3 β-hairpin. While helix formation may be initiated virtually anywhere along the sequence of (AAQAA)₃, the rigid loop of GB1m3 strongly restricts the accessible conformation space in the unfold state. Folding of the GB1m3 hairpin appears to largely follow similar

Figure 6. 2D Free energy profiles as a function of the end-to-end distance and number of native side chain and backbone H-bond contacts formed. (**a**) GB1p, (**b**) GB1m3, (**c**) HP5A and (**d**) (AAQAA)$_3$. Contours are drawn at every kT.

folding pathways as observed in GB1p and HP5A, which is initiated near the turn and involve formation of hydrophobic side chain contacts followed by rapid zipping up of backbone H-bonds (e.g., see Fig. S2). Note that the coil state of (AAQAA)$_3$ appears overly compact in GBSW compared to previous explicit solvent simulations[65], and the actual folding landscape of (AAQAA)$_3$ folding is thus likely more downhill than what is shown in Fig. 6d.

Discussion

While the energy landscape theory has provided a general statistical mechanics framework for understanding the thermodynamics and kinetics of protein folding, atomistic simulations using physics-based empirical energy functions are usually required to provide quantitative description of the detailed folding mechanism and energetics for a specific protein or peptide. This has increasingly become a reality with recent development of carefully optimized implicit and explicit solvent force fields[9,10,13,47,48] as well as development of advanced sampling methods and powerful hardware[4,66]. In this work, we combine an optimized implicit solvent force field and REX-MD enhanced sampling to analyze the folding of three related GB1p-series β-hairpins. The calculation was able to quantitatively recapitulate known mechanistic features as well as folding thermodynamics of these model systems. The results nicely illustrate how conformational entropy and long-range interactions together dictate various cooperative and downhill-like folding behaviors. This case study highlights the great potential of combining atomistic force fields and enhanced sampling for predictive studies of protein folding and large conformational transitions associated with various biological processes.

Methods

Model Peptides. Four peptides are used in this study, including helical (AAQAA)$_3$ and three GB1p-series β-hairpins. Consistent with the experimental conditions[18,46,67], termini of (AAQAA)$_3$ peptide were blocked with Ace and NH2 respectively, and all hairpins with unblocked termini. The β-hairpin sequences are: G$_{41}$EWTY DDATK TFTVT E$_{56}$ (GB1p), KKYTW NPATG KATVQ E (HP5A), and KKWTY NPATG KFTVQ E (GB1m3) (the loop segments are underlined). Note that HP5A and GB1m3 are derived from the native sequence of the C-terminal hairpin (residues 41–56) of the B1 domain of protein G (GB1p) with modified stabilities (folded populations at 298 K estimated from NMR chemical shirts[46] shown in parenthesis): HP5A (21 ± 10%) < GB1p (ca. 30%) < GB1m3 (86 ± 3%). Note that the folded population of GB1p has also been estimated to be ~42% at 278 K based on NMR[68] and ~80% at 273 K based on tryptophan fluorescence experiment[15]. Native side chain contacts (for Gb1p) include: Trp$_{43}$-Val$_{54}$ and Tyr$_{45}$-Phe$_{52}$; a total of 7 native backbone contacts for these hairpins include: Glu$_{42}$ NH-Thr$_{55}$ CO, Glu$_{42}$ CO-Thr$_{55}$ NH, Thr$_{44}$ NH-Thr$_{53}$ CO, Thr$_{44}$ CO-Thr$_{53}$ NH, Asp$_{46}$ NH-Thr$_{51}$ CO, Asp$_{46}$ CO-Thr$_{51}$ NH, and Asp$_{47}$ CO-Lys$_{50}$ NH. (AAQAA)$_3$ was estimated to be about 50% helical at 270 K based on NMR chemical shifts[67].

REX simulations in GBSW implicit solvent. As previously described[45], all peptides were simulated using the previously optimized GBSW protein implicit solvent force field[45], which was built on the CHARMM22/CMAP all-atom force field[69–71]. GBSW is generalized Born (GB)-class of implicit solvent that describes the solvent as a high dielectric continuum. GB offers a pair-wise, analytical approximation for calculating the electrostatic solvation free energy and is particularly suitable for molecular dynamics (MD) simulations[72]. GBSW in particular employs a van der Waals (vdW)-based surface with a smooth dielectric boundary, and the effective Born radii

(the key quantities in the GB approximation) are calculated by a rapid volume integration scheme that includes a higher-order correction term to the Coulomb field approximation[73]. Default GBSW parameters were used along with 50 Lebedev angular integration points and 24 radial integration points up to 20 Å for each atom[45]. The non-polar solvation energy was estimated from the solvent-exposed surface area with a phenomenological surface tension coefficient of 0.005 kcal/mol/Å^2.

REX-MD simulations in GBSW were performed using the MMTSB Tool Set[74] together with the CHARMM program[75,76]. REX involves multiple copies (replicas) of the peptide simulated at different temperatures independently and replicas attempt to exchange simulation temperatures periodically using Metropolis criteria that preserve the detailed balance. Replicas can travel up and down the temperature space during REX, which facilitates barrier crossing and reduces the probability of being trapped in states of local energy minima. 16 replicas were used in all simulations. The temperatures were distributed exponentially between 270 to 400 K for HP5A and GB1p and 270 to 550 K for (AAQAA)$_3$ and GB1m3[45]. SHAKE was applied to fix the lengths of all bonds with hydrogen atoms and a time-step of 2 femtoseconds (fs) was used. Temperature exchanges were attempted every 2.0 picoseconds (ps) of MD between neighboring replicas. The total simulation lengths were 20 nanoseconds (ns) for (AAQAA)$_3$ and 100 ns for the β-hairpins. In this work we only analyze ensembles derived from folding simulations that were initiated from fully extended structures. Previous comparison with results from control simulations initiated from folded structures suggested that the simulated ensembles were well converged within 100 ns[77].

Structural and energetic analysis. The post-analysis was done largely with CHARMM and the MMTSB Tool Set. The helicity was computed from the average 1–4 H-bond frequency, identified when the distance between the carbonyl oxygen of residue i, O_i, and the amide hydrogen of residue $i + 4$, HN_{i+4}, $d(O_i \cdots HN_{i+4}) \leq 2.6$ Å. Similar distance criteria were used to count the number of native backbone H-bonds in the β-hairpins. Sidechains are considered to be in contact if the shortest distance among heavy atoms is no greater than 4.2 Å. All conformations (or equivalently free energy profiles) were derived from conformations sampled during the last 60 ns of REX simulations. For principal component analysis (PCA), backbone structures of both GB1p and HP5A were first aligned using Cα atoms and then projected onto the first two principal components (PCs) to generate the 2D distributions/free energy profiles.

For folding energetic analysis, all conformations sampled during the last 60 ns of REX at 270 K were assigned to the folded state if $N_{HB} \geq 4$ and unfolded states if $N_{HB} \leq 1$. The free energy of folding was then calculated as $\Delta G = -RT \ln (P_f/P_u)$. The folding energy, ΔU, was calculated as the difference between average potential energies of the folded and unfolded sub-ensembles. Given ΔG and ΔU, the folding entropy was estimated as $\Delta S = (\Delta U - \Delta G)/T$. For GB1m3, the free energy profile (and equivalently the ensemble distribution) is dominated by the folded state (Fig. 2c) and the unfolded minimum is absent. ΔG was thus estimated by setting $P_u = 1 - P_f$. The low occupancy of the unfolded state at 270 K also prevents reliable estimation of ΔU directly from average potential energies. GB1m3 and HP5A contains the same rigidified loop and ΔS for GB1m3 was thus estimated to be similar to that of HP5A in Table 1. The folding transition temperature (T_m) was estimated as the temperature where $P_f = P_u$.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Anfinsen, C. B. Principles That Govern Folding of Protein Chains. *Science* **181**(4096), 223 (1973).
2. Onuchic, J. N., LutheySchulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
3. Shea, J. E. & Brooks, C. L. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52**, 499 (2001).
4. Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**(1), 52 (2012).
5. Lane, T. J., Shukla, D., Beauchamp, K. A. & Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **23**(1), 58 (2013).
6. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the Folding Routes. *Science* **267**(5204), 1619 (1995).
7. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**(1), 10 (1997).
8. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. Progress in modeling of protein structures and interactions. *Science* **310**(5748), 638 (2005).
9. Best, R. B. *et al.* Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **8**(9), 3257 (2012).
10. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**(8), 1950 (2010).
11. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**(9), L47 (2011).
12. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**(8), 3696 (2015).
13. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**(1), 71 (2017).
14. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**(6055), 517 (2011).
15. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**(6656), 196 (1997).
16. Dinner, A. R., Lazaridis, T. & Karplus, M. Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. USA* **96**(16), 9068 (1999).
17. Klimov, D. K. & Thirumalai, D. Mechanisms and kinetics of beta-hairpin formation. *Proc. Natl. Acad. Sci. USA* **97**(6), 2544 (2000).
18. Honda, S., Kobayashi, N. & Munekata, E. Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. *J. Mol. Biol.* **295**(2), 269 (2000).

19. Gnanakaran, S., Nymeyer, H., Portman, J., Sanbonmatsu, K. Y. & Garcia, A. E. Peptide folding simulations. *Curr. Opin. Struct. Biol.* **13**(2), 168 (2003).
20. Zhang, J., Qin, M. & Wang, W. Folding mechanism of beta-hairpins studied by replica exchange molecular simulations. *Proteins-Structure Function and Bioinformatics* **62**(3), 672 (2006).
21. Jas, G. S., Hegefeld, W. A., Middaugh, C. R., Johnson, C. K. & Kuczera, K. Detailed Microscopic Unfolding Pathways of an alpha-Helix and a beta-Hairpin: Direct Observation and Molecular Dynamics. *J. Phys. Chem. B* **118**(26), 7233 (2014).
22. Markiewicz, B. N., Yang, L. J., Culik, R. M., Gao, Y. Q. & Gai, F. How Quickly Can a beta-Hairpin Fold from Its Transition State? *J. Phys. Chem. B* **118**(12), 3317 (2014).
23. Zerze, G. H., Uz, B. & Mittal, J. Folding thermodynamics of beta-hairpins studied by replica-exchange molecular dynamics simulations. *Proteins-Structure Function and Bioinformatics* **83**(7), 1307 (2015).
24. Bille, A., Mohanty, S. & Irback, A. Peptide folding in the presence of interacting protein crowders. *J. Chem. Phys.* **144** (17) (2016).
25. Best, R. B. & Mittal, J. Microscopic events in beta-hairpin folding from alternative unfolded ensembles. *Proc. Natl. Acad. Sci. USA* **108**(27), 11087 (2011).
26. Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA* (2013).
27. Roux, B. & Simonson, T. Implicit solvent models. *Biophys. Chem.* **78**(1–2), 1 (1999).
28. Bashford, D. & Case, D. A. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.* **51**, 129 (2000).
29. Chen, J. & Brooks, C. L. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Phys. Chem. Chem. Phys.* **10**, 471 (2008).
30. Chen, J., Brooks, C. L. & Khandogin, J. Recent advances in implicit solvent based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.* **18**, 140 (2008).
31. Zheng, W., Andrec, M., Gallicchio, E. & Levy, R. M. Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc. Natl. Acad. Sci. USA* **104**(39), 15340 (2007).
32. Nguyen, H., Maier, J., Huang, H., Perrone, V. & Simmerling, C. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.* **136**(40), 13959 (2014).
33. Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14**(1), 76 (2004).
34. Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **25**(13), 1584 (2004).
35. Snow, C. D., Sorin, E. J., Rhee, Y. M. & Pande, V. S. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 43 (2005).
36. Best, R. B., Zheng, W. & Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **10**(11), 5113 (2014).
37. Nerenberg, P. S. & Head-Gordon, T. Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *Journal of Chemical Theory and Computation* **7**(4), 1220 (2011).
38. Palazzesi, F., Prakash, M. K., Bonomi, M. & Barducci, A. Accuracy of Current All-Atom Force-Fields in Modeling Protein Disordered States. *J. Chem. Theory Comput.* **11**(1), 2 (2015).
39. Huang, J. & MacKerell, A. D. Jr. Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **48**, 40 (2018).
40. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**(1–2), 141 (1999).
41. Hansmann, U. H. E. & Okamoto, Y. Numerical comparisons of three recently proposed algorithms in the protein folding problem. *J. Comput. Chem.* **18**(7), 920 (1997).
42. Liwo, A., Czaplewski, C., Oldziej, S. & Scheraga, H. A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **18**(2), 134 (2008).
43. Im, W. P., Lee, M. S. & Brooks, C. L. Generalized born model with a simple smoothing function. *J. Comput. Chem.* **24**(14), 1691 (2003).
44. Im, W., Chen, J. & Brooks, C. L. III. Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models. *Adv. Protein Chem.* **72**, 173 (2005).
45. Chen, J., Im, W. & Brooks, C. L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **128**(11), 3728 (2006).
46. Fesinmeyer, R. M., Hudson, F. M. & Andersen, N. H. Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* **126**(23), 7238 (2004).
47. Lee, K. H. & Chen, J. H. Optimization of the GBMV2 implicit solvent force field for accurate simulation of protein conformational equilibria. *J. Comput. Chem.* **38**(16), 1332 (2017).
48. Nguyen, H., Roe, D. R. & Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **9**(4), 2020 (2013).
49. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. Tryptophan zippers: Stable, monomeric beta-hairpins. *Proc. Natl. Acad. Sci. USA* **98**(10), 5578 (2001).
50. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9**(6), 425 (2002).
51. Khandogin, J., Chen, J. H. & Brooks, C. L. Exploring atomistic details of pH-dependent peptide folding. *Proc. Natl. Acad. Sci. USA* **103**(49), 18546 (2006).
52. Khandogin, J. & Brooks, C. L. Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proc. Natl. Acad. Sci. USA* **104**(43), 16880 (2007).
53. Khandogin, J., Raleigh, D. P. & Brooks, C. L. Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network. *J. Am. Chem. Soc.* **129**(11), 3056 (2007).
54. Chen, J. Intrinsically disordered p53 extreme C-terminus binds to S100B(betabeta) through "fly-casting". *J. Am. Chem. Soc.* **131**(6), 2088 (2009).
55. Ganguly, D. & Chen, J. Atomistic details of the disordered states of KID and pKID. implications in coupled binding and folding. *J. Am. Chem. Soc.* **131**(14), 5214 (2009).
56. Zhang, W., Ganguly, D. & Chen, J. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. *Plos Comput. Biol.* **8**(1), e1002353 (2012).
57. Ganguly, D. & Chen, J. Modulation of the disordered conformational ensembles of the p53 transactivation domain by cancer-associated mutations. *PLoS Comput. Biol.* **11**(4), e1004247 (2015).
58. Cho, S. S., Levy, Y. & Wolynes, P. G. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. USA* **103**(3), 586 (2006).
59. Dyer, R. B. Ultrafast and downhill protein folding. *Curr. Opin. Struct. Biol.* **17**(1), 38 (2007).
60. Luo, R., David, L., Hung, H., Devaney, J. & Gilson, M. K. Strength of solvent-exposed salt-bridges. *J. Phys. Chem. B* **103**(4), 727 (1999).
61. Yang, A.-S. & Honig, B. Electrostatic effects on protein stability. *Curr. Opin. Struct. Biol.* **2**(1), 40 (1992).
62. Zhou, R., Berne, B. J. & Germain, R. The free energy landscape for beta hairpin folding in explicit water. *Proc. Natl. Acad. Sci. USA* **98**(26), 14931 (2001).
63. Felts, A. K., Harano, Y., Gallicchio, E. & Levy, R. M. Free energy surfaces of beta-hairpin and alpha-helical peptides generated by replica exchange molecular dynamics with the AGBNP implicit solvent model. *Proteins* **56**(2), 310 (2004).

64. Kobayashi, N., Honda, S., Yoshii, H. & Munekata, E. Role of side-chains in the cooperative beta-hairpin folding of the short C-terminal fragment derived from streptococcal protein G. *Biochemistry (Mosc.)* **39**(21), 6564 (2000).
65. Huang, J. & MacKerell, A. D. Jr. Induction of peptide bond dipoles drives cooperative helix formation in the (AAQAA)$_3$ peptide. *Biophys. J.* **107**(4), 991 (2014).
66. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* **41**, 429 (2012).
67. Shalongo, W., Dugad, L. & Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl (Aaqaa) (3) Amide. *J. Am. Chem. Soc.* **116**(18), 8288 (1994).
68. Blanco, F. J., Rivas, G. & Serrano, L. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat. Struct. Biol.* **1**(9), 584 (1994).
69. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**(18), 3586 (1998).
70. Feig, M., MacKerell, A. D. & Brooks, C. L. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **107**(12), 2831 (2003).
71. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**(11), 1400 (2004).
72. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **112**(16), 6127 (1990).
73. Lee, M. S., Salsbury, F. R. & Brooks, C. L. Novel generalized Born methods. *J. Chem. Phys.* **116**(24), 10606 (2002).
74. Feig, M., Karanicolas, J. & Brooks, C. L. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graphics Modell.* **22**(5), 377 (2004).
75. Brooks, B. R. *et al.* Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**(2), 187 (1983).
76. Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **30**(10), 1545 (2009).
77. Zhang, W. & Chen, J. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. *J. Chem. Theory Comput.* **10**(3), 918 (2014).

Acknowledgements

Author Contributions

J.L. Chen and J.H. Chen, conception and design of the study; J.H. Chen, performing the simulation and analysis; J.L. Chen, X.L. and J.H. Chen, analysis and interpretation of data, drafting and revising the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-32028-7.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.