Probing the folded state and mechanical unfolding pathways of T4 lysozyme using all-atom and coarse-grained molecular simulation

Wenjun Zheng and Paul Glenn

Citation: The Journal of Chemical Physics **142**, 035101 (2015); doi: 10.1063/1.4905606
View online: http://dx.doi.org/10.1063/1.4905606
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/3?ver=pdfcov
Published by the AIP Publishing

Probing the folded state and mechanical unfolding pathways of T4 lysozyme using all-atom and coarse-grained molecular simulation

Wenjun Zheng[a] and Paul Glenn
Department of Physics, University at Buffalo, Buffalo, New York 14260, USA

(Received 15 September 2014; accepted 26 December 2014; published online 16 January 2015)

The Bacteriophage T4 Lysozyme (T4L) is a prototype modular protein comprised of an N-terminal and a C-domain domain, which was extensively studied to understand the folding/unfolding mechanism of modular proteins. To offer detailed structural and dynamic insights to the folded-state stability and the mechanical unfolding behaviors of T4L, we have performed extensive equilibrium and steered molecular dynamics simulations of both the wild-type (WT) and a circular permutation (CP) variant of T4L using all-atom and coarse-grained force fields. Our all-atom and coarse-grained simulations of the folded state have consistently found greater stability of the C-domain than the N-domain in isolation, which is in agreement with past thermostatic studies of T4L. While the all-atom simulation cannot fully explain the mechanical unfolding behaviors of the WT and the CP variant observed in an optical tweezers study, the coarse-grained simulations based on the Go model or a modified elastic network model (mENM) are in qualitative agreement with the experimental finding of greater unfolding cooperativity in the WT than the CP variant. Interestingly, the two coarse-grained models predict different structural mechanisms for the observed change in cooperativity between the WT and the CP variant—while the Go model predicts minor modification of the unfolding pathways by circular permutation (i.e., preserving the general order that the N-domain unfolds before the C-domain), the mENM predicts a dramatic change in unfolding pathways (e.g., different order of N/C-domain unfolding in the WT and the CP variant). Based on our simulations, we have analyzed the limitations of and the key differences between these models and offered testable predictions for future experiments to resolve the structural mechanism for cooperative folding/unfolding of T4L. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4905606]

INTRODUCTION

It is well known that many single-domain globular proteins fold in a cooperative all-or-none fashion.[1] However, the folding behavior of multi-domain modular proteins[2] is more complex (for example, see Ref. 3)—some may involve relatively independent folding of individual domains followed by their assembly, while others may require cooperative communication between domains during folding to ensure efficient folding without kinetic trapping or pathological misfolding. The Bacteriophage T4 lysozyme (T4L) has been extensively investigated as a prototype system for understanding the folding/unfolding mechanism of modular proteins. T4L consists of two globular domains (see Fig. 1)—an α/β N-terminal domain and an all-α C-terminal domain. Most ensemble studies found T4L undergoes an all-or-none two-state unfolding transition[4] (although one study found T4L undergoing apparent three-state denaturation[5]). Such folding cooperativity can be attributed to the inter-domain coupling mediated by two key structural elements—an N-terminal helix A forming part of the C-domain, and a long helix C bridging between the N- and C-domain (see Fig. 1). There is also evidence for some degree of independence between the folding of N- and C-domain—a partially unfolded/folded intermediate of T4L exists with the C-domain folded and the N-domain unfolded during chemical and thermal unfolding.[6–11] The presence of such an intermediate was supported by a native state hydrogen exchange study,[12] which found T4L can be divided into an N-terminal subdomain with low stability and a C-terminal subdomain with high stability. This is also consistent with a thermodynamic

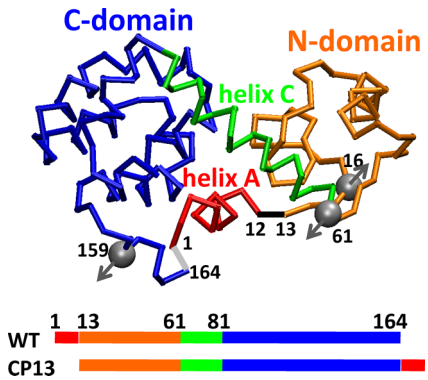

FIG. 1. Structural architecture of T4L comprised of the N-domain (orange) and the C-domain (blue) linked by helix A (red) and helix C (green). In the WT, helix A is connected to the N-domain (via a bond between residues 12 and 13). In the CP13 variant, helix A is disconnected from the N-domain and connected to the C-domain (via a bond between residues 1 and 164). Force-pulled residues 16, 61, and 159 are shown as gray spheres with arrows representing pulling forces. The sequence architecture of the WT and the CP13 variant is also shown (using same color schemes as above).

[a]E-mail: wjzheng@buffalo.edu. Telephone: (716) 645-2947. Fax: (716) 645-2507.

analysis of T4L fragments that found the C-domain (but not the N-domain) capable of folding in isolation.[13] Additionally, a pulse-labeling hydrogen exchange study found an early folding intermediate with residues from both domains contributing to the structure of this intermediate,[14] which was questioned by a later study.[9] Taken together, T4L seems to follow a complex folding/unfolding mechanism, featuring some degree of coupling between the N- and C-domain (leading to cooperative folding) coexisting with some degree of independence (resulting in an intermediate with the two domains folded to different extent).

In complement with ensemble measurements of thermal and chemical unfolding, single-molecule mechanical unfolding studies of T4L have been performed using atomic force microscope (AFM)[15,16] and optical tweezers.[17] By selectively pulling and unfolding different parts of a single protein molecule, one can access diverse unfolding pathways through different regions of the energy landscape, and quantitatively probe unfolding cooperativity involving multiple protein domains. For example, the unfolding cooperativity between the N- and C-domain of T4L can be assessed by addressing the following two questions:[17] To what extent do the two domains unfold simultaneously (rather than sequentially) when both are being pulled? How much does the C-domain unfold when only the N-domain is being pulled? To elucidate the role of chain connectivity in dictating the folding cooperativity of T4L, a circular permutation (CP) variant of T4L (named CP13 where the N-terminal 12 residues are truncated and re-linked to the C terminus, see Fig. 1) was constructed and analyzed by using both optical tweezers[17] and AFM.[15,16] The CP13 variant exhibited less cooperative and more diverse unfolding behaviors than the wild type (WT) T4L[15–17] (e.g., mixture of two-state and three-state transition events). This finding suggests weaker coupling between the N- and C-domain in the CP13 variant than the WT. To ultimately decipher the structural basis of the mechanical unfolding transition in T4L, it is important to analyze single-molecule pulling experiments in structural terms with high spatial and temporal resolution.

Structure-based molecular simulation is increasingly employed to complement biophysical and biochemical studies of protein transitions. The molecular dynamics (MD) simulation is capable of probing the dynamics of proteins under physiological conditions (i.e., in the presence of solvent and ions) with atomistic details.[18] MD is in principle well suited for simulating the folding/unfolding transition of T4L with high spatial and temporal resolution. However, the all-atom MD simulation with explicit solvent is computationally expensive and often limited to tens of nanoseconds (ns) in simulation time which is much shorter than the experimental time scale of T4L folding/unfolding in single-molecule measurements.[15–17] Although it is becoming feasible to perform long MD simulation (from microseconds to milliseconds) using newly developed specialized hardware or massively parallelized supercomputer (see Ref. 19), such facility has not yet been readily available to the community. Because of the time-scale limitation, it is difficult for MD simulation to adequately sample the conformational space and access transient intermediates of T4L during folding/unfolding. Additionally, the accuracy of MD force field remains uncertain for long-time simulation of protein dynamics.[20]

To accelerate MD simulation while retaining atomistic details, many venues have been attempted such as using implicit solvent model[21] instead of explicit solvent or applying a driving force in steered MD[22] and targeted MD.[23] Such mechanical manipulations nicely mimic the puling force exerted on single protein molecules in AFM and optical tweezers experiments, but the simulated pulling is often too fast (with simulation time limited to nanoseconds-microseconds) to accurately capture the experimentally observed dynamics of protein mechanical unfolding (with time scale ~ seconds). An alternative strategy is to use a coarse-grained (CG) model, which is based on a reduced protein representation (e.g., treating an amino-acid residue as a single bead) and/or a simplified energy function. By greatly reducing computing cost, a CG model allows much longer simulation time needed for simulating slow pulling as in single-molecule experiments. A variety of CG models[24] have been developed to simulate various aspects of protein conformational dynamics. For example, the elastic network model (ENM)[25–27] represents a protein structure as a network of C_α atoms with neighboring ones connected by springs with a uniform force constant.[28] The Go model uses a structure-based potential function, which is based on the native contacts of a folded protein structure, to simulate protein folding/unfolding.[29] The use of ENM in conjunction with normal mode analysis has led to many applications, such as simulation of single-molecule protein stretching,[30,31] analysis of protein conformational transitions,[32–34] and thermal fluctuations in protein crystals.[35–41] However, it remains uncertain if the CG models are sufficiently accurate to explore intermediate conformations during protein folding/unfolding and conformational transitions. In recent studies,[42,43] an all-atom Go model was used to investigate the folding transition and intermediate states on the folding pathways of single-domain proteins, which offered detailed insights to protein-folding intermediates captured by experimental probes like NMR.[44] In another recent study,[45] Krobath *et al.* used CG lattice and off-lattice models to probe the role of N-terminal to C-terminal coupling in the folding cooperativity of small single-domain proteins.

In this study, we have performed extensive equilibrium and steered MD simulations of both the WT and the CP13 variant of T4L using all-atom and two CG force fields (Go model and modified elastic network model (mENM), see Methods). Although equilibrium MD simulation of T4L was performed in the past,[46–51] to our knowledge no sMD simulation has been reported for the mechanical unfolding of T4L. Our objective is two-fold: first, we will use simulations (based on three different models) to understand the experimentally observed native-state stability and cooperative unfolding behaviors of the WT and the CP13 variant,[17] which will improve our mechanistic understanding of multi-domain protein folding/unfolding; second, we will critically evaluate the quality of these various models in reproducing the experimentally observed results, which will help us to identify the limitations of all-atom and CG models and assess the feasibility of capturing complex folding/unfolding behaviors of multi-domain proteins using simplified CG models. The

unfolding pathways predicted by our simulations will motivate future experiments to resolve the structural mechanism of T4L mechanical unfolding and its cooperativity.

METHODS

All-atom MD and sMD simulation of T4L

The all-atom MD and sMD simulation was carried out using the CHARMM22[52] force field and an implicit solvent model—Fast Analytical Continuum Treatment of Solvation (FACTS),[53] which is computationally less expensive than explicit solvent models. An X-ray structure of T4L (PDB id: 3LZM) was chosen as the native conformation for WT T4L. This T4L structure was initially minimized in energy for 5000 steps using the adopted basis Newton-Raphson algorithm, and then subject to an MD or sMD run. During a sMD run, a pair of pulling forces were applied (following Ref. 17), via a spring of force constant 100 pN/Å, to the C_α atoms of residues 16 and 159 (or 61) to gradually increase their distance to 430 Å (or 135 Å) at a speed of 21.5 Å/ns using the AFM command[54] of CHARMM program.[55] The simulation time of sMD was 20 ns for the 16-159 pulling and 6.3 ns for the 16-61 pulling. The simulation time of MD was 20 ns. For both MD and sMD run, a Langevin dynamics simulation was run using the CHARMM program with a time step of 1 fs, a heat bath at a temperature of 300 K and a friction coefficient of 0.1 ps^{-1}. The atomic coordinates of T4L and pulling force were saved every 10 ps. Ten independent MD (sMD) trajectories were generated and then combined to form an ensemble of folded-state (partially unfolded) conformations of T4L for further analysis (see below).

The MD and sMD simulations were also conducted for the CP13 variant of T4L. To model the change of chain connectivity from WT to CP13 variant (see Fig. 1), we added a harmonic distance restraint (with force constant 1 kcal/mol $Å^{-2}$) between the C_α atoms of residues 1 and 164 and deleted the chemical bond, bond angles, and dihedral angles between residues 12 and 13.

Go-model-based CG MD and sMD simulation of T4L

The Go Model Server[56] (http://mmtsb.org/webservices/gomodel.html) was used to generate the topology, parameter, and sequence files in CHARMM format for the Go model based on the T4L structure 3LZM. We used the same sMD pulling setup as in the all-atom sMD (see above), except at a lower pulling speed of 2.2 Å/ns. The simulation time of sMD was 200 ns for the 16-159 pulling and 63 ns for the 16-61 pulling. The simulation time of MD was 100 ns. For both MD and sMD run, we performed the same Langevin dynamics simulation as in the all-atom simulation (see above), except using a larger time step of 0.01 ps. The C_α coordinates of T4L and pulling force were saved every 100 ps. Ten MD (sMD) trajectories were generated to form an ensemble of folded-state (partially unfolded) conformations of T4L.

The simulations were also conducted for the CP13 variant of T4L. The change of chain connectivity from WT to CP13 variant was modeled in the same way as in the all-atom simulation (see above).

mENM

An ENM can be constructed from the C_α coordinates of a protein native structure, where each residue is represented by its C_α atom. The original form of the ENM potential energy[28] is

$$E_{ENM} = \frac{1}{2}\sum_{i<j} C_{ij}\theta(R_c - d_{ij,0})(d_{ij} - d_{ij,0})^2, \quad (1)$$

where d_{ij} is the distance between residues i and j, and $d_{ij,0}$ is the value of d_{ij} given by the native structure, $\theta(x)$ is the Heaviside function, R_c is the cutoff distance (chosen to be 15 Å here), and C_{ij} is the force constant of the spring connecting residues i and j. C_{ij} can be set to a uniform constant,[26] or two different values for bonded and non-bonded residue pairs,[57] or varying as a function of distance $d_{ij,0}$.[58]

To allow non-bonded residues to move apart readily while maintaining the pseudo-bonds between bonded residues, we modified the ENM energy in Eq. (1) to the following form (named mENM energy):[59,60]

$$\begin{aligned} E_{mENM} &= E_b + E_{nb}, \\ E_b &= \frac{1}{2}\sum_{\langle ij\rangle \in P_b} C_b(d_{ij} - d_{ij,0})^2, \\ E_{nb} &= \frac{1}{2}\sum_{\langle ij\rangle \notin P_b} C_{nb}\theta(R_c - d_{ij,0}) \times \frac{d_{ij,0}^2}{36}\left(1 - \frac{d_{ij,0}^6}{d_{ij}^6}\right)^2, \end{aligned} \quad (2)$$

where E_b is the pseudo-bonded energy (P_b is the set of pseudo-bonded residue pairs, the bonded force constant $C_b = 10$, in arbitrary unit); E_{nb} is the non-bonded energy described by the Lennard-Jones 6-12 potential—it has a minimum at $d_{ij,0}$, saturates as d_{ij} goes to infinity, and diverges as d_{ij} approaches zero (here, the non-bonded force constant $C_{nb} = (4/d_{ij,0})^2$, following Ref. 58). Therefore, unlike the harmonic potential in Eq. (1), the mENM energy allows two non-bonded residues to move apart at a finite energy cost. To set the absolute energy scale, both C_b and C_{nb} are rescaled so that the total non-bonded energy matches the value given by the Go model of T4L.[56]

mENM-based CG MD and sMD simulation of T4L

To explore the mechanical unfolding pathways of T4L upon constant-speed pulling using the mENM force field (see Eq. (2)), we performed sMD simulation following the Go-model-based sMD protocol described above (except for a change in the parameter file to redefine all native contacts and the non-bonded parameters). We also conducted equilibrium MD simulation using the mENM force field.

Ensemble-based analysis of folded-state or partially unfolded conformations of T4L

To analyze the mechanical unfolding pathways of T4L, it is useful to calculate a reaction coordinate (RC) as the pulling distance between the C_α atoms of the two residues being pulled apart (i.e., residues 16 and 61 or 159). All partially unfolded conformations of T4L were grouped by their RC

values into a discrete set of RC bins (with bin width of 10 Å). For the conformations in each RC bin, we calculated an average conformation by separately superimposing the residues of N-domain and C-domain onto the native structure and then averaging over the superimposed coordinates of each domain. An average unfolding pathway was constructed using a series of these average conformations in the order of ascending RC. An average pulling force was calculated for each RC bin and plotted as a function of RC to obtain the force-distance relation akin to the force-extension curve measured in single-molecule optical tweezers experiments.[17] To assess the extent of local folding at individual residue positions, we calculated the following fraction of native contacts at residue i for a given conformation of T4L

$$f_{nc}(i) = \frac{\sum_j \delta_{ij}\theta(1.1d_{ij,0} - d_{ij})}{\sum_j \delta_{ij}}, \quad (3)$$

where d_{ij} is the distance between residues i and j, $d_{ij,0}$ is the value of d_{ij} given in the native structure, $\theta(x)$ is the Heaviside function, δ_{ij} is 1 if residues i and j form native contact (as defined in the Go model[56]) and 0 otherwise. $f_{nc}(i)$ was averaged over the N-domain residues (denoted $f_{nc,N}$) to assess how well the N-domain is folded. The same calculation was done for the C-domain to obtain the fraction of native contacts in the C-domain (denoted $f_{nc,C}$) and for all residues to obtain the fraction of native contacts in T4L (denoted $f_{nc,all}$). The mean and standard deviation of these fractions were then calculated over a set of T4L conformations (i.e., a RC bin from sMD simulation or a folded-state ensemble generated by equilibrium MD simulation).

RESULTS AND DISCUSSION

All-atom MD simulation of T4L in folded state

To study the structural stability of folded state in T4L, we conducted all-atom MD simulation of T4L at room temperature (300 K) using an implicit solvent model (see Methods). We simulated both the WT and the CP13 variant to evaluate the effect of changing chain connectivity on the folded-state stability of T4L. For each system, ten 20-ns-long MD trajectories were generated and combined to construct a structural ensemble of T4L in the folded state (for equilibration assessment of the trajectories, see Fig. S4[64]). Due to limited simulation time, this ensemble only represents a subset of conformations accessible to T4L in the folded state. Despite such caveat, it should allow us to compare the relative stability between the WT and the CP13 variant and between the N- and C-domain (see below). We note that "stability" refers to the structural stability of the folded-state ensemble as assessed by the distribution of fraction of native contacts, which is different from but related to the thermodynamic stability of a folded state.

We first calculated the Root Mean Squared Deviation (RMSD) relative to the native structure of T4L and the fraction of native contacts in T4L (denoted $f_{nc,all}$, see Methods) for the structural ensembles. As indicated by broader distribution of RMSD and $f_{nc,all}$ (see Fig. 2(a)), the CP13 variant is structurally more variable and less stable than the WT in the folded state. This is consistent with previous thermodynamic studies of T4L that found the CP13 variant to be less stable than the WT.[13,61]

To assess the stability of the N- and C-domain in T4L, we calculated the fraction of native contacts in the N- and C-domain (denoted $f_{nc,N}$ and $f_{nc,C}$, see Methods) for the structural ensembles of the WT and the CP13 variant (see Fig. 2(b)). For the WT, both domains are equally well folded on average ($f_{nc,N} = 0.861 \pm 0.044$ and $f_{nc,C} = 0.853 \pm 0.037$) although the N-domain seems more variable (with higher standard deviation in $f_{nc,N}$). Inspection of the distribution of $f_{nc,N}$ and $f_{nc,C}$ revealed a minor cluster of partially unfolded conformations in the WT ensemble with $f_{nc,N} < f_{nc,C}$ (circled in Fig. 2(b)). Indeed, such partially unfolded conformations were accessible to native state hydrogen exchange measurement, which found the N-domain was less stable and more prone to local unfolding than the C-domain in the folded state of T4L.[12] For the CP13 variant, the N-domain is less well folded and more unstable than the C-domain ($f_{nc,N} = 0.763 \pm 0.099$ and $f_{nc,C} = 0.840 \pm 0.045$). Indeed, while the C-domain is well folded in all conformations, the N-domain is unfolded in ~0.5% of conformations (with $f_{nc,N} < 0.5$, see Fig. 2(b)). Therefore, the circular permutation (see Fig. 1) destabilizes the N-domain more than the C-domain in the CP13 variant. We further analyzed the fraction of native contacts at individual residues of the N- and C-domain (see Fig. 2(c)). Thirty seven residues are less stable in the CP13 variant, which are distributed in the N-domain, helix A, helix C, and the C-terminal region of C-domain (see Fig. 2(c)).

To further assess the stability of the N- and C-domain in isolation, we ran ten 20-ns-long MD trajectories for each domain alone (with the other deleted) and combined them to form a structural ensemble of the N- or C-domain fragment. The C-domain remains well-folded in the absence of N-domain ($f_{nc,C} = 0.842 \pm 0.055$, with $f_{nc,C} > 0.5$ in all conformations, see Fig. 2(d)), whereas the N-domain is partially unfolded in isolation ($f_{nc,N} = 0.680 \pm 0.125$, with $f_{nc,N} < 0.5$ in ~10% of conformations, see Fig. 2(d)). This finding agrees with the experimental observation that the C-domain can fold in isolation while the N-domain fragment is predominantly unfolded.[13] Our finding supports the importance of C-domain in conferring stability to the N-domain and T4L as a whole.

In sum, our all-atom MD simulation of T4L in the folded state and as fragments has revealed higher stability of the WT than the CP13 variant, and higher stability of the C-domain than the N-domain, which is in qualitative agreement with past thermodynamic studies of T4L. This gives us confidence in the accuracy of all-atom force field in simulating the equilibrium dynamics of T4L in the folded state, while it remains uncertain if it is accurate enough to simulate the non-equilibrium dynamics of T4L during mechanical unfolding.

All-atom sMD simulation of T4L during mechanical unfolding

To explore the mechanical unfolding pathways of T4L, we conducted all-atom sMD simulation of T4L being pulled apart by a pair of forces acting at residue pair (16, 61) or (16, 159) at room temperature (see Methods). These pulling setups (see Fig. 1) follow a recent optical tweezers study of

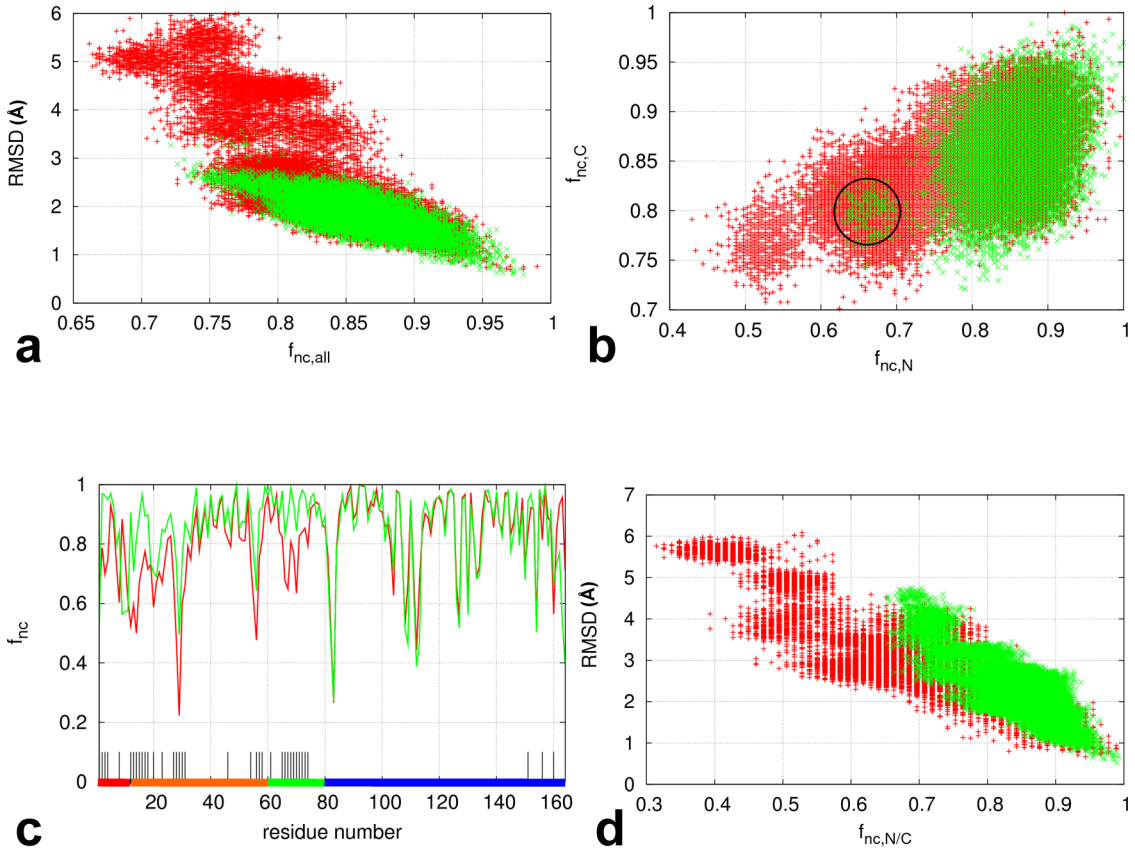

FIG. 2. Stability analysis of folded-state structural ensembles of the WT (green) and the CP13 variant (red) generated by all-atom MD simulation: (a) scattered plot of RMSD vs. $f_{nc,all}$; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$; (c) f_{nc} at individual residue positions; (d) scattered plot of RMSD vs. $f_{nc,N}$ (for the N-domain fragment of T4L, red) and $f_{nc,C}$ (for the C-domain fragment of T4L, green). In panel (c), ranges of residue numbers for the N/C-domain, helix A, and helix C are colored (using same color schemes as Fig. 1), residue positions where f_{nc} is reduced by >0.1 from the WT to the CP13 variant are shown by vertical dashes.

T4L[17]—the (16, 61) pulling specifically unfolds the N-domain and explores its effect on the stability of the C-domain, while the (16, 159) pulling unfolds the entire T4L and explores the unfolding sequence of the N- and C-domain. We simulated both the WT and the CP13 variant to evaluate the effect of circular permutation on the unfolding pathways of T4L. For each system, ten sMD trajectories (6 or 20 ns per trajectory, see Methods) were generated and combined to construct a structural ensemble of partially unfolded conformations of T4L. Based on this ensemble, we calculated an average unfolding pathway and analyzed the evolution of fraction of native contacts and pulling force as a function of the pulling distance (i.e., the distance between residues 16 and 159 or 61) which is used as the RC for the unfolding transition (see Methods).

We first analyzed the (16, 159) pulling. In the WT, the average unfolding pathway indicates cooperative unfolding of both the N- and C-domain with $f_{nc,N}$ and $f_{nc,C}$ decreasing simultaneously (see Fig. 3(a)) and the C-domain leading the N-domain (i.e., $f_{nc,N} > f_{nc,C}$). In contrast, the average unfolding pathway of the CP13 variant shows less cooperativity between the two domains with $f_{nc,N}$ decreasing faster than $f_{nc,C}$ (see Fig. 3(a)) and the N-domain leading the C-domain (i.e., $f_{nc,N} < f_{nc,C}$). The greater cooperativity in the WT than the CP13 variant was also evident from the finding of smaller difference between $f_{nc,N}$ and $f_{nc,C}$ in the WT ($f_{nc,N} - f_{nc,C} = 0.15 \pm 0.20$ for the WT and -0.22 ± 0.16 for the CP13 variant). We ran two additional sMD trajectories of the WT and the CP13 variant at five-fold slower pulling speed, and observed similar unfolding pathways (see Fig. S1(a)[64]), which indicates the above finding is not an artifact of fast pulling simulation. Our finding is in qualitative agreement with the observation that the WT unfolds more cooperatively than the CP13 variant by an optical tweezers study[17] and an AFM study.[16] In particular, Ref. 17 found that most unfolding events in the CP13 variant were three-state transitions with the N-domain unfolding before the C-domain, while only two-state transitions were observed in the WT.

To structurally visualize distinct unfolding pathways in the WT and the CP13 variant, we inspected the average unfolding pathways (see Figs. 4(a) and 4(b)) and sMD trajectories (see supplementary material for coordinate files[64]).

In the WT, the pulling of residues 16 and 159 triggers early unfolding of nearby regions (including helix A, helix C, and the C-terminal region of C-domain) (see snapshot 1 in Fig. 4(a)), followed by detachment of helix A from the C-domain and separation of the N- and C-domain (see snapshot 2 in Fig. 4(a)), and then unfolding of the two domains (see snapshot 3 in Fig. 4(a)). Despite sharing the above sequence of unfolding events, the individual sMD trajectories exhibited large variations (e.g., in the detailed order of N- and C-domain unfolding), which is consistent with the finding of broad distribution in $f_{nc,N}$ and $f_{nc,C}$ (see Fig. 3(a)).

In the CP13 variant, the pulling of residues 16 and 159 causes early unfolding of the N-domain (see snapshot 1 in

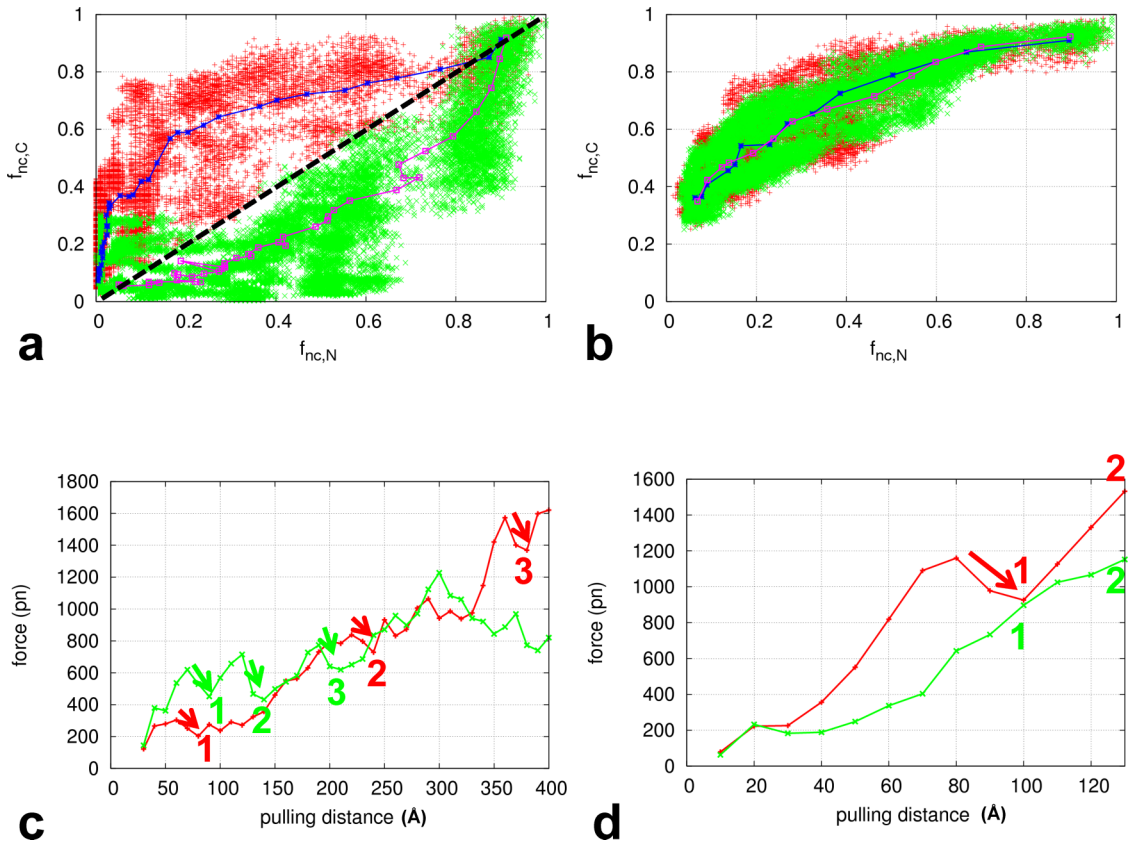

FIG. 3. Analysis of mechanical unfolding of the WT (green) and the CP13 variant (red) of T4L based on all-atom sMD simulation: (a) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-159 pulling; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-61 pulling; (c) force-distance curves for the 16-159 pulling; (d) force-distance curves for the 16-61 pulling. In panels (a) and (b), the average unfolding pathways are depicted by averages of $f_{nc,C}$ and $f_{nc,N}$ over bins of pulling distances (colored purple for the WT and blue for the CP13 variant). In panels (c) and (d), key rip events are highlighted by arrows, and the intermediates following those rip events (or near the end of unfolding) are numerically labeled (the arrows and labels are colored green for the WT and red for the CP13 variant).

Fig. 4(b)), followed by partial unfolding of the C-domain in the C-terminal region, helix A, and helix C (see snapshot 2 in Fig. 4(b)), and then unfolding of the rest of C-domain (see snapshot 3 in Fig. 4(b)).

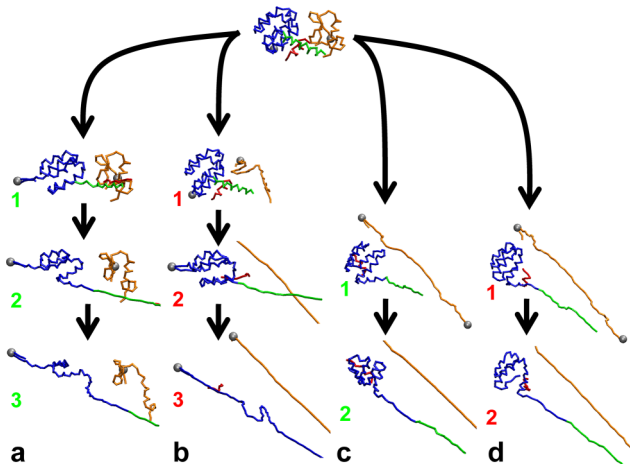

FIG. 4. Snapshots of intermediate conformations along the average unfolding pathways based on all-atom sMD simulation: (a) the 16-159 pulling of the WT; (b) the 16-159 pulling of the CP13 variant; (c) the 16-61 pulling of the WT; (d) the 16-61 puling of the CP13 variant. The T4L structure is colored the same way as in Fig. 1. The snapshots are numerically labeled in the same way as in Figs. 3(c) and 3(d). The unfolding pathways are directed by downward-pointing arrows (starting from the native structure of T4L at the top).

Therefore, as revealed by the structural visualization of unfolding pathways, the higher cooperativity of the WT was conferred by the attachment of helix A to the C-domain which protects the N-domain from early unfolding and favors early partial unfolding of the C-domain. The circular permutation in the CP13 variant removes such protection and allows the less stable N-domain to unfold before the more stable C-domain (see Fig. 2(d)). This explains the above finding of dramatic change in unfolding pathways between the WT and the CP13 variant (see Fig. 3(a)).

We then calculated the pulling force as a function of pulling distance (see Methods) for the (16, 159) pulling of the WT and the CP13 variant. At pulling distance up to 150 Å, we found larger force in the WT than the CP13 variant, featuring two force peaks (each followed by a sudden decrease of force corresponding to a rip event, see Fig. 3(c)). The two rips correspond to early partial unfolding of the C-domain (including the C-terminal region, helix A, and helix C, see snapshot 1 in Fig. 4(a)), followed by detachment of helix A from the C-domain (see snapshot 2 in Fig. 4(a)). At pulling distance beyond 150 Å, there are more rips corresponding to further unfolding of the C-domain and the N-domain (for example, see snapshot 3 in Fig. 4(a)). In sum, our force-distance calculation is consistent with the optical-tweezers finding of larger unfolding force in the WT than the CP13 variant,[17] although our calculation predicts multiple rips and unfolding intermediates which were not resolved experimentally.

Next, we analyzed the (16, 61) pulling. In both the WT and the CP13 variant, the average unfolding pathways indicate that the C-domain unfolds to similar extent ($f_{nc,C} \sim 0.35$) following the N-domain unfolding (see Fig. 3(b)). Indeed, as revealed by structural visualization of the average unfolding pathways in the WT and the CP13 variant, unfolding of the N-domain is followed by partial unfolding of the C-domain in helix C, while helix A remains in proximity to the C-domain (see Figs. 4(c) and 4(d)). Therefore, the (16, 61) pulling primarily causes unfolding of helix C rather than the detachment of helix A, resulting in similar unfolding pathways between the WT and the CP13 variant. The above finding is inconsistent with the observation that the WT and the CP13 variant unfold to different extent with different cooperativity during the (16, 61) pulling by an optical tweezers study.[17] To understand the reason for such discrepancy between simulation and experiment, we ran two additional sMD trajectories of the WT and the CP13 variant at five-fold slower pulling speed. Encouragingly, we found the WT was further unfolded in the C-domain compared with the CP13 variant (with $f_{nc,C}$ lower by ~0.1, see Fig. S1(b)[64]). This preliminary result indicates that longer sMD simulation may be needed to resolve the different unfolding pathways of the WT and the CP13 variant in the (16, 61) pulling. Due to limited computing resource, we will seek for alternative CG sMD simulation of the 16-61 pulling (see below).

We then compared the force-distance curves for the (16, 61) pulling between the WT and the CP13 variant. At pulling distance up to 100 Å, we found larger force in the CP13 variant than the WT, featuring a broad force peak at 80 Å (see Fig. 3(d)). This is inconsistent with the experimental finding of smaller unfolding force in the CP13 variant,[17] which confirms the inadequacy of our all-atom sMD simulation to account for the 16-61 pulling of T4L.

In summary, our all-atom sMD simulation of T4L supports the importance of circular permutation in changing the cooperativity of mechanical unfolding of T4L during the 16-159 pulling. However, our simulation does not agree with the experimental finding of a more cooperative WT than the CP13 variant during the 16-61 pulling.[17]

CG MD simulation of T4L in folded state based on the Go model

To overcome the time-scale limit of all-atom MD simulation, it is desirable to perform CG simulation which is computationally cheap and can access longer time scales (e.g., up to μs). The C_α-based Go model is a popular CG model for simulating protein folding/unfolding.[62] Based on the Go-model parameters obtained from a Go model server,[56] we first conducted ten 100-ns-long MD runs of T4L (both the WT and the CP13 variant) in the folded state at room temperature (300 K). However, the initial MD simulation did not yield a stable folded state—the N-domain is largely unfolded ($f_{nc,N} = 0.226 \pm 0.087$ for the WT and 0.208 ± 0.062 for the CP13 variant) while the C-domain is well folded ($f_{nc,C} = 0.788 \pm 0.068$ for the WT and 0.819 ± 0.040 for the CP13 variant). To ensure the stability of T4L in the folded state, we repeated MD simulations after gradually decreasing temperature in steps of 10 K starting from 300 K until both domains are folded (i.e., $f_{nc,N} > 0.5$ and $f_{nc,C} > 0.5$) in the WT and the CP13 variant. This goal was achieved at 230 K. So, we decided to perform all Go-model-based simulations of T4L at 230 K.

To probe the effect of circular permutation on the folded-state stability of T4L, we used the Go model to conduct MD simulation for the WT and the CP13 variant (see Methods). For each system, ten 100-ns-long MD trajectories were generated and combined to construct a folded-state ensemble of T4L (for equilibration assessment of the trajectories, see Fig. S5[64]). Then, we compared the relative stability between the WT and the CP13 variant and between the N- and C-domain (see below).

We found the distribution of RMSD and $f_{nc,all}$ largely overlaps between the WT and the CP13 variant, although a small fraction of CP13 conformations show very large RMSD (see Fig. 5(a)). This is in contrast to the finding of much larger difference between the WT and the CP13 variant by all-atom MD simulation (see Fig. 2(a)).

By analyzing the distributions of $f_{nc,N}$ and $f_{nc,C}$, (see Fig. 5(b)), we found that the N-domain is less well-folded and more variable than the C-domain in both the WT and the CP13 variant ($f_{nc,N} = 0.757 \pm 0.056$ and $f_{nc,C} = 0.921 \pm 0.021$ for the WT, $f_{nc,N} = 0.734 \pm 0.063$ and $f_{nc,C} = 0.918 \pm 0.022$ for the CP13 variant). Therefore, the circular permutation in the CP13 variant has little effect on the stability of the N- and C-domain (with slightly reduced stability in the N-domain). This is contrary to the finding of much larger change in the stability of N-domain between the WT and the CP13 variant by all-atom MD simulation (see Fig. 2(b)). The distribution of f_{nc} at individual residues is highly similar between the WT and the CP13 variant (see Fig. 5(c)), with only 5 residues in helix A and the N-domain having reduced stability in the CP13 variant. This is in contrast to the finding of more extensive differences in f_{nc} between the WT and the CP13 variant by all-atom MD simulation (see Fig. 2(c)).

Next, we generated ten 100-ns-long MD trajectories for the N- or C-domain fragment to form a structural ensemble for each fragment. The C-domain remains well-folded in the absence of N-domain ($f_{nc,C} = 0.919 \pm 0.023$, see Fig. 5(d)), whereas the N-domain is largely unfolded in isolation ($f_{nc,N} = 0.354 \pm 0.099$, see Fig. 5(d)). This is in contrast to the finding of smaller difference in stability between the N- and C-domain by all-atom MD simulation (see Fig. 2(d)).

In sum, our Go-model-based MD simulation of T4L in the folded state and as fragments has found slightly higher stability of the WT than the CP13 variant, and much higher stability of the C-domain than the N-domain, which is in qualitative agreement with past thermodynamic studies of T4L.

CG sMD simulation of T4L during mechanical unfolding based on the Go model

To explore the mechanical unfolding pathways of T4L, we used the Go model to conduct CG sMD simulation of T4L undergoing (16, 61) or (16, 159) pulling at 230 K (see Methods). We also performed the sMD simulation at 10-fold slower pulling speed and repeated the analysis to make sure the results are not sensitive to the pulling speed (see Fig. S2[64]).

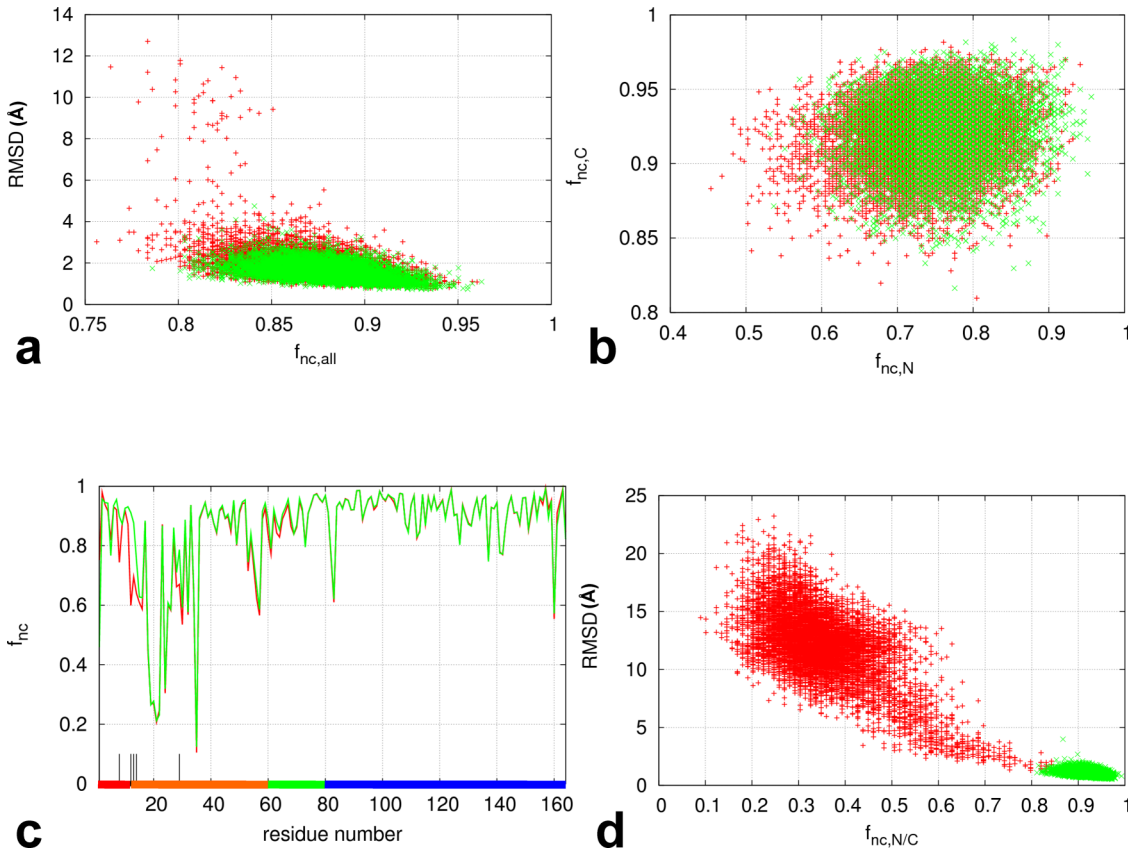

FIG. 5. Stability analysis of folded-state structural ensembles of the WT (green) and the CP13 variant (red) generated by Go-model-based MD simulation: (a) scattered plot of RMSD vs. $f_{nc,all}$; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$; (c) f_{nc} at individual residue positions; (d) scattered plot of RMSD vs. $f_{nc,N}$ (for the N-domain fragment of T4L, red) and $f_{nc,C}$ (for the C-domain fragment of T4L, green). In panel (c), ranges of residue numbers for the N/C-domain, helix A, and helix C are colored (using same color schemes as Fig. 1), residue positions where f_{nc} is reduced by >0.1 from the WT to the CP13 variant are shown by vertical dashes.

We first analyzed the (16, 159) pulling. In the WT, the average unfolding pathway exhibits three distinct stages (see Fig. 6(a)): (1) Early partial unfolding of the C-domain with $f_{nc,C}$ decreasing from ~0.9 to ~0.7. (2) Subsequent unfolding of the N-domain while the C-domain remains folded. (3) Late unfolding of the rest of C-domain while the N-domain partially refolds and then unfolds. In the CP13 variant, the average unfolding pathway is similar to the WT except that the unfolding of the N-domain starts earlier than the partial unfolding of the C-domain (see Fig. 6(a)). For both the WT and the CP13 variant, the N-domain leads the C-domain during unfolding (with $f_{nc,N} < f_{nc,C}$), although the unfolding of the N- and C-domain is more intertwined in the WT than the CP13 variant, hinting for higher cooperativity in the WT. The above finding is in qualitative agreement with the observation that the WT unfolds more cooperatively than the CP13 variant by an optical tweezers study[17] and an AFM study.[16]

Then, we visualized the average unfolding pathways (see Figs. 7(a) and 7(b)) and sMD trajectories (see coordinate files in supplementary material[64]).

In the WT, the pulling of residues 16 and 159 triggers early unfolding of nearby regions (including helix A, helix C, and the C-terminal region of C-domain) and detachment of helix A (see snapshot 1 in Fig. 7(a)), followed by unfolding of the N-domain (see snapshot 2 in Fig. 7(a)), and then unfolding of the rest of C-domain (see snapshot 3 in Fig. 7(a)).

In the CP13 variant, the pulling of residues 16 causes early unfolding of the N-domain (see snapshot 1 in Fig. 7(b)), followed by partial unfolding of the C-domain in helix C (see snapshot 2 in Fig. 7(b)), and then unfolding of the rest of C-domain (see snapshot 3 in Fig. 7(b)).

Therefore, while sharing overall similar unfolding pathways (i.e., the N-domain unfolds before the C-domain unfolds), the unfolding of N-domain is slightly delayed in the WT until the detachment of helix A from the C-domain. The circular permutation in the CP13 variant removes such delay and allows the less stable N-domain to unfold even earlier before the more stable C-domain (see Fig. 5(d)). This explains the above finding of minor change in unfolding pathways between the WT and the CP13 variant (see Fig. 6(a)).

We then compared the force-distance curves for the (16, 159) pulling between the WT and the CP13 variant. At pulling distance up to 100 Å, we found larger force in the WT than the CP13 variant, featuring two force peaks and associated rip events (see Fig. 6(c)). The two rips correspond to early partial unfolding of the C-domain in the C-terminal region, helix A, and helix C (see snapshot 1 in Fig. 7(a)), followed by unfolding of the N-domain (see snapshot 2 in Fig. 7(a)). This is consistent with the experimental finding of greater cooperativity and higher unfolding force in the WT.[17] Both the WT and the CP13 variant have a maximum force peak at 220 Å (see Fig. 6(c)) corresponding to the unfolding of C-domain (see

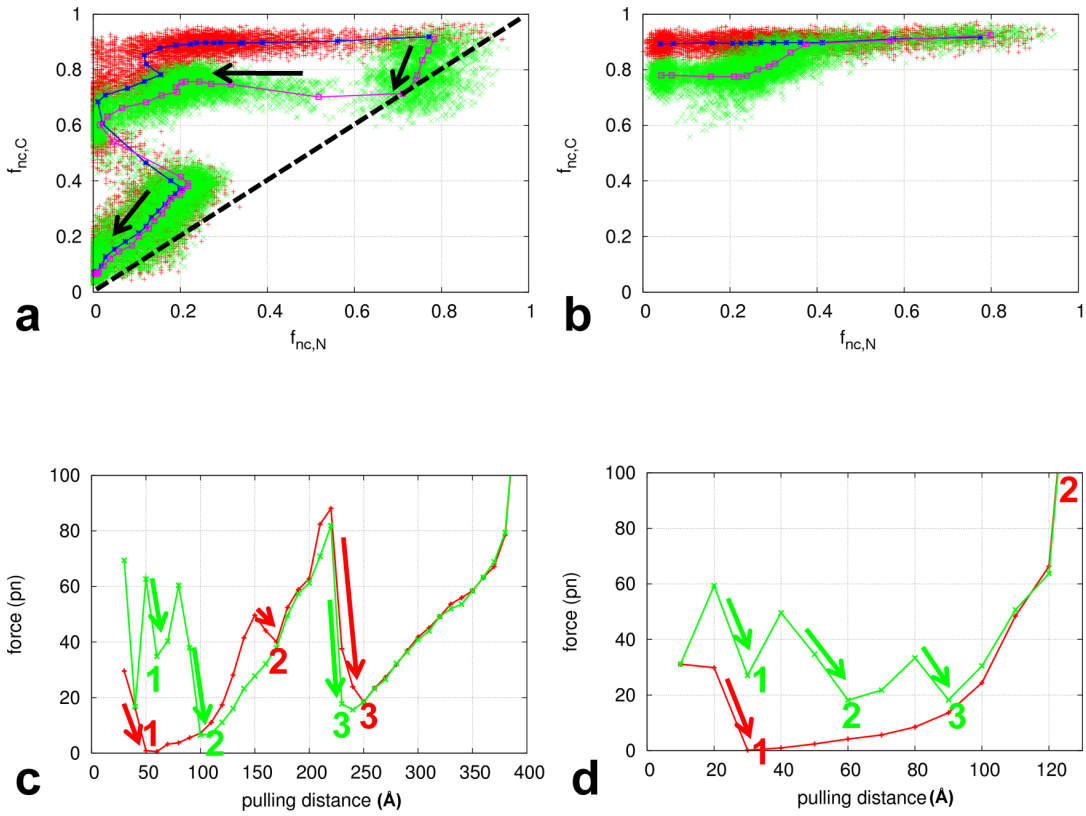

FIG. 6. Analysis of mechanical unfolding of the WT (green) and the CP13 variant (red) of T4L based on sMD simulation of the Go model: (a) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-159 pulling; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-61 pulling; (c) force-distance curves for the 16-159 pulling; (d) force-distance curves for the 16-61 pulling. In panels (a) and (b), the average unfolding pathways are depicted by averages of $f_{nc,C}$ and $f_{nc,N}$ over bins of pulling distances (colored purple for the WT and blue for the CP13 variant). In panel (a), three stages of the WT unfolding pathway are marked by three black arrows. In panels (c) and (d), key rip events are highlighted by arrows, and the intermediates following those rip events (or near the end of unfolding) are numerically labeled (the arrows and labels are colored green for the WT and red for the CP13 variant).

snapshot 3 in Figs. 7(a) and 7(b)), supporting the existence of a stable intermediate with the N-domain unfolded and the C-domain partially folded in both systems (see snapshot 2 in Figs. 7(a) and 7(b)). However, this is inconsistent with the finding of major energy barrier for mechanical unfolding of the WT close to the native state.[16,17] The finding of an intermediate with a more stable C-domain in both the WT and the CP13 variant is consistent with the native state hydrogen exchange data,[7] but it is at odds with the experimental observation of a single-rip transition (i.e., without intermediate) in the WT.[17]

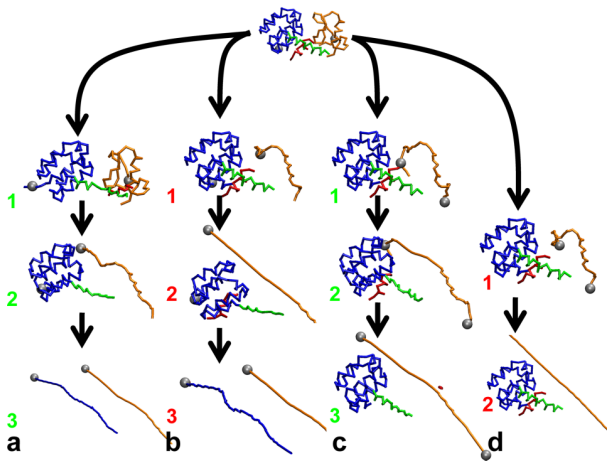

FIG. 7. Snapshots of intermediate conformations along the average unfolding pathways based on sMD simulation of the Go model: (a) the 16-159 pulling of the WT; (b) the 16-159 pulling of the CP13 variant; (c) the 16-61 pulling of the WT; (d) the 16-61 puling of the CP13 variant. The T4L structure is colored the same way as in Fig. 1. The snapshots are numerically labeled in the same way as in Figs. 6(c) and 6(d). The unfolding pathways are directed by downward-pointing arrows (starting from the native structure of T4L at the top).

Next, we analyzed the (16, 61) pulling. Similar to the (16, 159) pulling, we found slightly higher cooperativity in the WT than the CP13 variant: in the WT, the C-domain is slightly unfolded after the unfolding of N-domain (with $f_{nc,C}$ reduced by ~0.1, see Fig. 6(b)); in the CP13 variant, the C-domain remains folded after the N-domain is unfolded. Indeed, as shown by visualization of the average unfolding pathway in the WT, we found that the N-domain unfolds first (see snapshot 1 in Fig. 7(c)), followed by partial unfolding of the C-domain in helix C (see snapshot 2 in Fig. 7(c)), and then detachment of helix A from the C-domain (see snapshot 3 in Fig. 7(c)). In contrast, neither helix A nor helix C unfolds during the unfolding of N-domain in the CP13 variant (see Fig. 7(d)). The above finding is consistent with the experimental observation that the WT unfolds further with greater cooperativity than the CP13 variant during the (16, 61) pulling.[17]

We then compared the force-distance curves for the (16, 61) pulling between the WT and the CP13 variant. At pulling distance up to 100 Å, we found larger force in the WT than the CP13 variant, featuring three force peaks and associated rip events (see Fig. 6(d)) corresponding to the N-domain un-

folding, helix C unfolding, and helix A detachment, respectively (see snapshot 1-3 in Fig. 7(c)). This is consistent with the experimental finding of greater cooperativity and higher unfolding force in the WT than the CP13 variant.[17]

In summary, our Go-model-based sMD simulation supports the importance of circular permutation in changing the cooperativity of mechanical unfolding in T4L, which qualitatively agrees with the optical tweezers finding that the WT is more cooperative than the CP13 variant in both 16-61 and 16-159 pulling.[17] The simulation indicates some degree of coupling between the N- and C-domain so that the unfolding of N-domain partially destabilizes the C-domain. In contrast, the optical tweezers study found stronger coupling between the two domains so that the C-domain completely unfolds after the unfolding of N-domain.[17]

CG MD simulation of T4L in folded state based on the mENM

Given the diversity of CG models in the literature (e.g., different forms of "Go-like" models), it is interesting to compare their predictions in light of experimental data.[31] To this end, we used an alternative CG model to simulate mechanical unfolding of T4L, which is based on a modified form of ENM (mENM, see Methods) featuring a Lennard-Jones potential for non-bonded residue-residue interactions. Compared with the Go model from Ref. 56, the mENM is essentially a different "Go-like" model with a simpler form of non-bonded potential. We previously used mENM to simulate the mechanical unfolding of the SNARE complex (a four-helix bundle, see Ref. 31) which revealed new unfolding pathways distinct from that predicted by the Go model. To allow direct comparison with the simulation of Go model (see above), we calibrated mENM by ensuring that the total non-bonded energy is equal to that of Go model and distributed uniformly among residue-residue contacts within a C_α-C_α distance cutoff of 15 Å.

We used the mENM to conduct MD simulation that probes the folded state of the WT and the CP13 variant (see Methods). For equilibration assessment of the trajectories, see Fig. S6.[64] Unlike the Go-model-based MD simulation (see above), the mENM-based MD simulation at 300 K yielded a stable folded state for both the WT and the CP13 variant (see below).

We found the distribution of RMSD and $f_{nc,all}$ overlaps very well between the WT and the CP13 variant (see Fig. 8(a)), which is similar to the Go-model-based simulation, but in contrast to the all-atom simulation.

By comparing the distributions of $f_{nc,N}$ and $f_{nc,C}$, (see Fig. 8(b)), we found that for both the WT and the CP13 variant, the N-domain is less well-folded and more variable than the C-domain ($f_{nc,N} = 0.563 \pm 0.079$ and $f_{nc,C} = 0.668 \pm 0.050$ for the WT, $f_{nc,N} = 0.535 \pm 0.078$ and $f_{nc,C} = 0.670 \pm 0.050$ for the CP13 variant). Therefore, similar to the Go-model-based simulation, circular permutation has little effect on the stability of the N- and C-domain (with slightly reduced stability in the N-domain). Similar to the Go-model-based simulation, the distribution of f_{nc} at individual residues overlaps well between

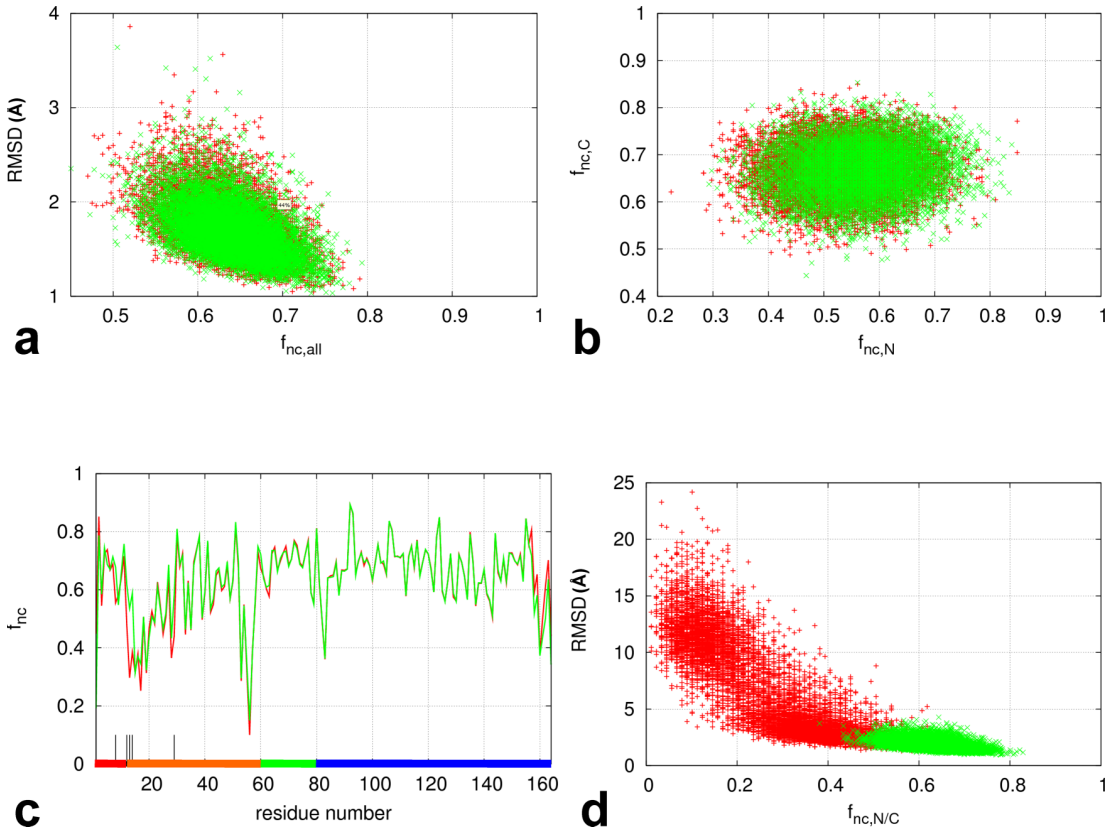

FIG. 8. Stability analysis of folded-state structural ensembles of the WT (green) and the CP13 variant (red) generated by mENM-based MD simulation: (a) scattered plot of RMSD vs. $f_{nc,all}$; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$; (c) f_{nc} at individual residue positions; (d) scattered plot of RMSD vs. $f_{nc,N}$ (for the N-domain fragment of T4L, red) and $f_{nc,C}$ (for the C-domain fragment of T4L, green). In panel (c), ranges of residue numbers for the N/C-domain, helix A, and helix C are colored (using the same color schemes as Fig. 1), residue positions where f_{nc} is reduced by >0.1 from the WT to the CP13 variant are shown by vertical dashes.

the WT and the CP13 variant, with only 5 residues in helix A and the N-domain having reduced stability in the CP13 variant (see Fig. 8(c)).

Based on MD simulations of the N- and C-domain in isolation, we found the C-domain remains folded in isolation ($f_{nc,C} = 0.631 \pm 0.054$, see Fig. 8(d)), whereas the N-domain is largely unfolded in the absence of C-domain ($f_{nc,N} = 0.286 \pm 0.14$, see Fig. 8(d)). This is similar to the finding of large difference in stability between the N- and C-domain by the Go-model-based simulation (see Fig. 5(d)).

In sum, our mENM-based MD simulation of T4L in the folded state and as fragments has revealed similar stability between the WT and the CP13 variant and highly different stability between the N- and C-domain of T4L.

CG sMD simulation of T4L during mechanical unfolding based on the mENM

To explore the mechanical unfolding pathways of T4L, we used the mENM to conduct CG sMD simulation of T4L undergoing (16, 61) or (16, 159) pulling at 300 K (see Methods). We also performed the sMD simulation at 10-fold slower pulling speed (see Fig. S3[64]) and at 230 K (see Fig. S9[64]) to make sure the results are not sensitive to the pulling speed and temperature.

We first analyzed the (16, 159) pulling. In the WT, we found two distinct classes of unfolding pathways (see Fig. 9(a)): (1) Early partial unfolding of the C-domain, followed by unfolding of the N-domain, and then unfolding of the rest of C-domain. (2) Early unfolding of the C-domain, followed by unfolding of the N-domain. The 2nd class is more dominant, resulting in an average unfolding pathway with the C-domain leading the N-domain in unfolding (i.e., $f_{nc,N} > f_{nc,C}$, see Fig. 9(a)). In contrast to the WT, the CP13 variant follows a single class of unfolding pathways with the N-domain unfolding followed by the C-domain unfolding (see Fig. 9(a)), which agrees well with the finding of dominant two-rip unfolding transitions in Ref. 17. Therefore, the circular permutation causes dramatic change in unfolding pathways between the WT and the CP13 variant, which is in contrast to the finding of similar unfolding pathways in the WT and the CP13 variant by the Go-model-based sMD simulation (see Fig. 6(a)). Similar to the Go-model-based simulation, the unfolding of the N- and C-domain is more intertwined in the WT than the CP13 variant, hinting for higher cooperativity in the WT than the CP13 variant as observed in an optical tweezers study[17] and an AFM study.[16] The existence of multiple unfolding pathways was also observed in the AFM study.[16]

We then inspected the average unfolding pathways (see Figs. 10(a) and 10(b)) and sMD trajectories (see coordinate files in supplementary material[64]).

In the WT, the pulling of residue 159 causes early unfolding of the C-terminal region of C-domain (see snapshot

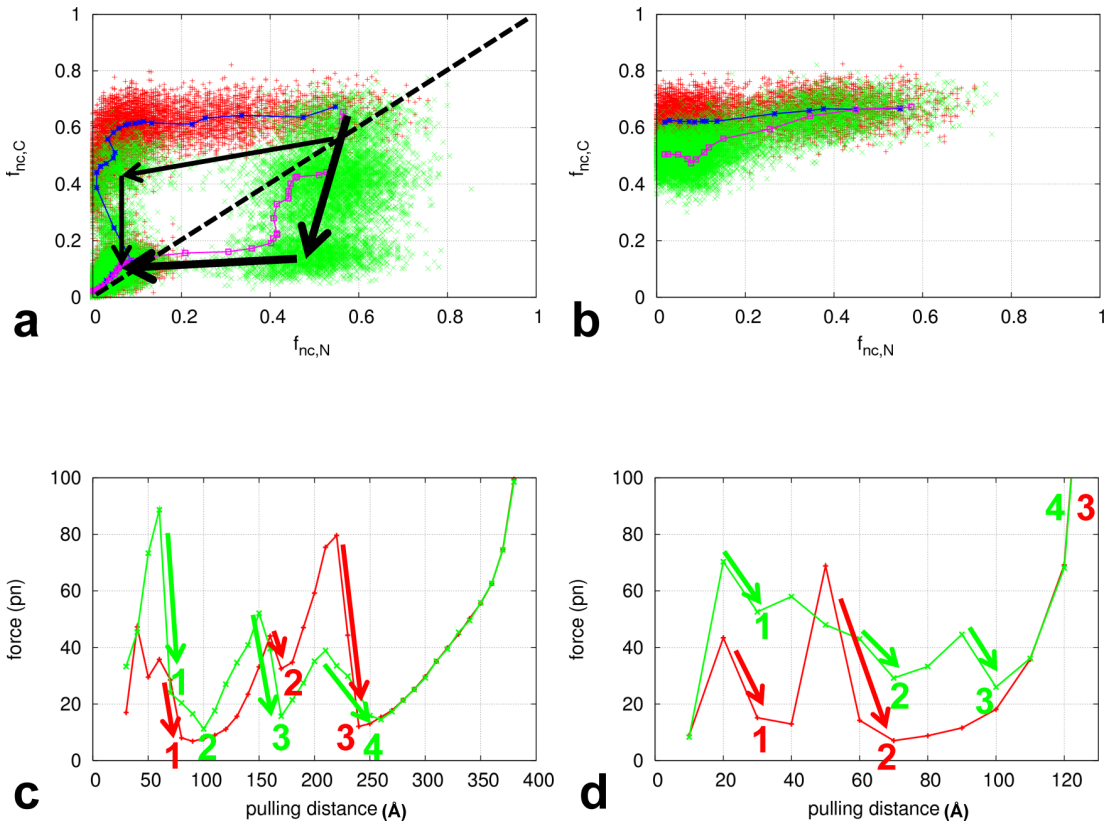

FIG. 9. Analysis of mechanical unfolding of the WT (green) and the CP13 variant (red) of T4L based on sMD simulation of the mENM: (a) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-159 pulling; (b) scattered plot of $f_{nc,C}$ vs. $f_{nc,N}$ for the 16-61 pulling; (c) force-distance curves for the 16-159 pulling; (d) force-distance curves for the 16-61 pulling. In panels (a) and (b), the average unfolding pathways are depicted by averages of $f_{nc,C}$ and $f_{nc,N}$ over bins of pulling distances (colored purple for the WT and blue for the CP13 variant). In panel (a), two classes of WT unfolding pathways are shown by thick and thin arrows. In panels (c) and (d), key rip events are highlighted by arrows, and the intermediates following those rip events (or near the end of unfolding) are numerically labeled (the arrows and labels are colored green for the WT and red for the CP13 variant).

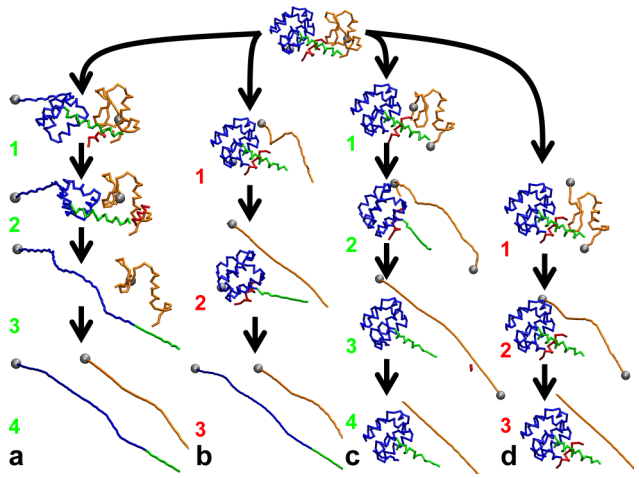

FIG. 10. Snapshots of intermediate conformations along the average unfolding pathways based on sMD simulation of the mENM: (a) the 16-159 pulling of the WT; (b) the 16-159 pulling of the CP13 variant; (c) the 16-61 pulling of the WT; (d) the 16-61 puling of the CP13 variant. The T4L structure is colored the same way as in Fig. 1. The snapshots are numerically labeled in the same way as in Figs. 9(c) and 9(d). The unfolding pathways are directed by downward-pointing arrows (starting from the native structure of T4L at the top).

1 in Fig. 10(a)), followed by separation of the two domains (after detachment of helix A from the C-domain, see snapshot 2 in Fig. 10(a)) and partial unfolding of the N-domain (due to contribution from the class-1 pathways, see snapshot 2 in Fig. 10(a)), and then unfolding of the rest of C-domain (see snapshot 3 in Fig. 10(a)), and finally unfolding of the N-domain (see snapshot 4 in Fig. 10(a)).

In the CP13 variant, the pulling of residue 16 causes early unfolding of the N-domain while helix A remains attached to the C-domain (see snapshot 1 in Fig. 10(b)), followed by unfolding of helix C (see snapshot 2 in Fig. 10(b)), and finally unfolding of the rest of C domain (see snapshot 3 in Fig. 10(b)).

In sum, the WT and the CP13 variant exhibit highly different unfolding pathways—the unfolding of N-domain is significantly delayed in the WT until the detachment of helix A from the C-domain (and after the unfolding of C-domain in most unfolding transitions). The circular permutation in the CP13 variant removes such delay and allows the N-domain to unfold before the C-domain (see Fig. 8(d)). Compared with the Go model, the mENM predicts a more dramatic effect of circular permutation on restructuring the unfolding pathways which is akin to the all-atom sMD simulation (see Fig. 3(a)).

We then compared the force-distance curves for the (16, 159) pulling between the WT and the CP13 variant. In the WT, the maximum force peak at 60 Å and associated rip event (see Fig. 9(c)) correspond to the early partial unfolding of the C-domain and the N-domain (see snapshots 1 and 2 in Fig. 10(a)), followed by two minor force peaks and rips at 150 Å and 210 Å (see Fig. 9(c)) corresponding to further unfolding of the C- and N-domain, respectively (see snapshots 3 and 4 in Fig. 10(a)). This is consistent with the finding of a major energy barrier for mechanical unfolding of the WT close to the native state,[16,17] which is associated to the breaking of interactions between helix A and the C-domain.[16] In the CP13 variant, the maximum force peak shifts to 220 Å (see Fig. 9(c)) corresponding to the late unfolding of C-domain (see snapshot 3 in Fig. 10(b)), which is preceded by two minor peaks at 40 Å and 160 Å (see Fig. 9(c)) corresponding to the early unfolding of N-domain and helix C (see snapshots 1 and 2 in Fig. 10(b)). This is consistent with the experimental finding of greater cooperativity and higher unfolding force in the WT.[17] The above finding implies that a stable intermediate with the N-domain unfolded and the C-domain folded is only populated during mechanical unfolding of the CP13 variant. This is in good agreement with the observation of single-rip transitions (without intermediate) in the WT and two-rip transitions in the CP13 variant in Ref. 17.

Next, we analyzed the (16, 61) pulling. Similar to the (16, 61) pulling of Go model, we found slightly higher cooperativity in the WT than the CP13 variant—the WT is slightly unfolded in the C-domain after the unfolding of N-domain (with $f_{nc,C}$ lower by ~0.1, see Fig. 9(b)), while the CP13 variant remains folded in the C-domain. As revealed by visualization of the average unfolding pathways, in the WT, the N-domain unfolds early at N-terminus (see snapshot 1 in Fig. 10(c)), followed by further unfolding of the N-domain and partial unfolding of the C-domain in helix C (see snapshot 2 in Fig. 10(c)), and then detachment of helix A from the C-domain (see snapshot 3 in Fig. 10(c)). In contrast, in the CP13 variant, neither helix A nor helix C unfolds as the N-domain unfolds (see Fig. 10(d)). This observation highlights the key role of helix A and C in conferring greater cooperativity to the WT than the CP13 variant in the (16, 61) pulling. The above finding is qualitatively consistent with the observation that the WT unfolds while the CP13 variant remains folded in the C-domain upon the (16, 61) pulling.[17]

We also compared the force-distance curves for the (16, 61) pulling between the WT and the CP13 variant. We found larger force in the WT than the CP13 at pulling distance up to 100 Å (except at a maximum force peak at 50 Å in the CP13 variant, see Fig. 9(d)). The above finding is consistent with the greater cooperativity and higher unfolding force in the WT than the CP13 variant as observed in Ref. 17. In the WT, we found three rips (see Fig. 9(d)) corresponding to progressive unfolding of the N-domain, helix C unfolding, and helix A detachment (see snapshots 1-3 in Fig. 10(c)). In the CP13 variant, we observed two rips (see Fig. 9(d)) corresponding to progressive unfolding of the N-domain starting at N terminus (see snapshots 1 and 2 in Fig. 10(d)).

In summary, our mENM-based simulation supports the importance of circular permutation in changing the cooperativity of mechanical unfolding during both 16-159 pulling and 16-61 pulling in T4L as observed experimentally.[17] The simulation indicates some degree of coupling between the N- and C-domain so that the unfolding of N-domain partially destabilizes the C-domain.

CONCLUSION

Before ending, we will further compare and analyze the different dynamic behaviors revealed by the all-atom, Go model, and mENM based simulations of mechanical unfolding of T4L.

All-atom simulation

The sMD simulation of the 16-159 pulling predicted distinct unfolding pathways between the WT and the CP13 variant (see Fig. 3(a)), in agreement with the experimental finding of a more cooperative WT than the CP13 variant during the 16-159 pulling.[17] However, the sMD simulation of the 16-61 pulling, while revealing cooperative coupling between the N- and C-domain (i.e., large reduction in $f_{nc,C}$ as the N-domain is unfolded, see Fig. 3(b)), failed to explain the experimental finding of a more cooperative WT than the CP13 variant during the 16-61 pulling.[17] The above discrepancy can be traced to the early unfolding of helix C while helix A remains attached to the C-domain (see Figs. 4(c) and 4(d)), which may be due to limited simulation time (i.e., pulling too fast), or inaccurate force field that underestimates helix C's stability or overestimates the packing interaction between helix A and the C-domain. This is in line with our recent finding that all-atom sMD simulation did not correctly capture the mechanical unfolding pathway of a four-helix-bundle protein (possibly due to overestimation of the packing interactions between helices).[31] Therefore, the all-atom force field may need to be further improved and validated before one can use it to run sMD simulation and accurately interpret mechanical unfolding experiments.

Go-model simulation

The sMD simulation of the 16-159 and 16-61 pulling predicted slightly different unfolding pathways between the WT and the CP13 variant (see Figs. 6(a) and 6(b)), which qualitatively agrees with the experimental finding of a more cooperative WT than the CP13 variant.[17] Notably, the unfolding pathways during the 16-159 pulling are partially similar between the WT and the CP13 variant (see Fig. 6(a)), both involving a stable intermediate with the C-domain folded and the N-domain unfolded. Therefore, rather than drastically restructuring the energy landscape and unfolding pathways as observed in the all-atom sMD simulation (see Fig. 3(a)), the Go model predicted a more subtle modification of the unfolding pathways to confer slightly different cooperativity between the WT and the CP13 variant. Similar finding was made in a recent simulation study of the free-energy landscape of T4L based on the Go model.[63] The similar unfolding behaviors of the WT and the CP13 variant hint for a more dominant role of the non-bonded native contacts (common to the WT and the CP13 variant) than chain connectivity in dictating the unfolding dynamics of Go model.

mENM simulation

Unlike the Go model, the mENM-based sMD simulation predicted highly distinct unfolding pathways for the WT and the CP13 variant during the 16-159 pulling (see Fig. 9(a)). To further understand the different unfolding behaviors between the Go model and mENM, we have compared the contact maps of the two models (see Fig. S7[64]): the Go model features fewer contacts at shorter distances (up to ~12 Å) while the mENM features more contacts at longer distances (up to the 15 Å cutoff). We reason that the addition of more long-distance contacts in mENM may alter the coupling between the N- and C-domain, and therefore change the order by which these two domains unfold during the 16-159 pulling (see Figs. 6(a) and 9(a)). To verify the above reasoning, we conducted new sMD simulation of mENM with a lower cutoff distance (R_c = 13 Å). Indeed, we found the new unfolding pathways to be more similar to the Go model than the old mENM with R_c = 15 Å (see Fig. S8[64]). Therefore, the long-range interactions play a key role in dictating the unfolding pathways and cooperativity in T4L.

In sum, while neither model gives a fully satisfactory explanation of all findings of the pulling experiment,[17] the two CG models gave reasonable results in qualitative agreement with the finding of distinct cooperativity between the WT and the CP13 variant. While the all-atom simulation is limited by short simulation time and force-field inaccuracy, the CG models also have their limitations (e.g., the native-centric character of Go model and mENM limits their accuracy in sampling partially folded/unfolded conformations). Future development of the CG models (e.g., with proper treatment of long-range interactions) are needed to improve their accuracy in simulating the mechanical unfolding of multi-domain proteins.

In conclusion, we have performed extensive equilibrium and steered MD simulations of both the WT and the CP13 variant of T4L using all-atom and CG force fields. Our all-atom and CG simulations of the folded state have consistently found greater stability of the C-domain than the N-domain in isolation. While the all-atom pulling simulation cannot fully explain the different unfolding behaviors of the WT and the CP13 variant as observed experimentally,[17] the CG pulling simulations based on the Go model or mENM are in qualitative agreement with the experimental finding of greater unfolding cooperativity in the WT than the CP13 variant.[17] The two CG models predict different unfolding pathways for the WT T4L. Future experiments will be needed to resolve the specific structural mechanism (e.g., unfolding pathways) underlying the folding/unfolding cooperativity of T4L.

ACKNOWLEDGMENTS

We thank funding support from the National Science Foundation (Grant No. 0952736) and American Heart Association (Grant No. 14GRNT18980033). The computer simulations were conducted using the supercomputing facility at the Center for Computational Research at University at Buffalo and the LOBOS cluster at NIH.

[1]J. J. Portman, Curr. Opin. Struct. Biol. **20**, 11 (2010).
[2]J. Fitter, Cell. Mol. Life Sci. **66**, 1672 (2009).
[3]E. Freire, K. P. Murphy, J. M. Sanchez-Ruiz, M. L. Galisteo, and P. L. Privalov, Biochemistry **31**, 250 (1992).
[4]M. Elwell and J. Schellman, Biochim. Biophys. Acta **386**, 309 (1975).
[5]F. Meersman and K. Heremans, Biochemistry **42**, 14234 (2003).
[6]J. Cellitti, R. Bernstein, and S. Marqusee, Protein Sci. **16**, 852 (2007).
[7]J. Cellitti, M. Llinas, N. Echols, E. A. Shank, B. Gillespie, E. Kwon, S. M. Crowder, F. W. Dahlquist, T. Alber, and S. Marqusee, Protein Sci. **16**, 842 (2007).
[8]H. Kato, H. Feng, and Y. Bai, J. Mol. Biol. **365**, 870 (2007).

[9]H. Kato, N. D. Vu, H. Feng, Z. Zhou, and Y. Bai, J. Mol. Biol. **365**, 881 (2007).
[10]M. Desmadril and J. M. Yon, Biochemistry **23**, 11 (1984).
[11]N. C. Gassner, W. A. Baase, J. D. Lindstrom, J. Lu, F. W. Dahlquist, and B. W. Matthews, Biochemistry **38**, 14451 (1999).
[12]M. Llinas, B. Gillespie, F. W. Dahlquist, and S. Marqusee, Nat. Struct. Biol. **6**, 1072 (1999).
[13]M. Llinas and S. Marqusee, Protein Sci. **7**, 96 (1998).
[14]J. Lu and F. W. Dahlquist, Biochemistry **31**, 4749 (1992).
[15]G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist, and C. Bustamante, Proc. Natl. Acad. Sci. U. S. A. **97**, 139 (2000).
[16]Q. Peng and H. Li, Proc. Natl. Acad. Sci. U. S. A. **105**, 1885 (2008).
[17]E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee, and C. Bustamante, Nature **465**, 637 (2010).
[18]M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[19]J. L. Klepeis, K. Lindorff-Larsen, R. O. Dror, and D. E. Shaw, Curr. Opin. Struct. Biol. **19**, 120 (2009).
[20]S. Piana, J. L. Klepeis, and D. E. Shaw, Curr. Opin. Struct. Biol. **24**, 98 (2014).
[21]M. Feig and C. L. Brooks, 3rd, Curr. Opin. Struct. Biol. **14**, 217 (2004).
[22]H. Lu, B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten, Biophys. J. **75**, 662 (1998).
[23]J. Schlitter, M. Engels, and P. Kruger, J. Mol. Graphics **12**, 84 (1994).
[24]V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005).
[25]K. Hinsen, Proteins **33**, 417 (1998).
[26]A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[27]F. Tama and Y. H. Sanejouand, Protein Eng. **14**, 1 (2001).
[28]M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[29]N. Go, Annu. Rev. Biophys. Bioeng. **12**, 183 (1983).
[30]J. I. Sulkowska, A. Kloczkowski, T. Z. Sen, M. Cieplak, and R. L. Jernigan, Proteins **71**, 45 (2008).
[31]W. Zheng, Proteins **82**, 1376 (2014).
[32]P. Maragakis and M. Karplus, J. Mol. Biol. **352**, 807 (2005).
[33]W. Zheng, B. R. Brooks, and G. Hummer, Proteins **69**, 43 (2007).
[34]M. Tekpinar and W. Zheng, Proteins **78**, 2469 (2010).
[35]J. Hafner and W. Zheng, J. Chem. Phys. **132**, 014111 (2010).
[36]J. Hafner and W. Zheng, J. Chem. Phys. **135**, 144114 (2011).
[37]W. Zheng, Biophys. J. **94**, 3853 (2008).
[38]W. Zheng, Biophys. J. **98**, 3025 (2010).
[39]D. A. Kondrashov, Q. Cui, and G. N. Phillips, Jr., Biophys. J. **91**, 2760 (2006).
[40]D. A. Kondrashov, A. W. Van Wynsberghe, R. M. Bannen, Q. Cui, and G. N. Phillips, Jr., Structure **15**, 169 (2007).
[41]D. Riccardi, Q. Cui, and G. N. Phillips, Jr., Biophys. J. **96**, 464 (2009).
[42]H. Krobath, S. G. Estacio, P. F. Faisca, and E. I. Shakhnovich, J. Mol. Biol. **422**, 705 (2012).
[43]S. G. Estacio, H. Krobath, D. Vila-Vicosa, M. Machuqueiro, E. I. Shakhnovich, and P. F. Faisca, PLoS Comput. Biol. **10**, e1003606 (2014).
[44]P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundstrom, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay, Science **336**, 362 (2012).
[45]H. Krobath, A. Rey, and P. F. Faisca, "How determinant is N-terminal to C-terminal coupling for protein folding?," Phys. Chem. Chem. Phys. (published online).
[46]G. Mann and J. Hermans, J. Mol. Biol. **302**, 979 (2000).
[47]B. L. de Groot, S. Hayward, D. M. van Aalten, A. Amadei, and H. J. Berendsen, Proteins **31**, 116 (1998).
[48]T. M. Gray, E. J. Arnoys, S. Blankespoor, T. Born, R. Jagar, R. Everman, D. Plowman, A. Stair, and D. Zhang, Protein Sci. **5**, 742 (1996).
[49]G. E. Arnold, J. I. Manchester, B. D. Townsend, and R. L. Ornstein, J. Biomol. Struct. Dyn. **12**, 457 (1994).
[50]G. E. Arnold and R. L. Ornstein, Proteins **18**, 19 (1994).
[51]G. E. Arnold and R. L. Ornstein, Protein Eng. **5**, 703 (1992).
[52]A. D. Mackerell, Jr., M. Feig, and C. L. Brooks, 3rd, J. Comput. Chem. **25**, 1400 (2004).
[53]U. Haberthur and A. Caflisch, J. Comput. Chem. **29**, 701 (2008).
[54]E. Paci and M. Karplus, Proc. Natl. Acad. Sci. U. S. A. **97**, 6521 (2000).
[55]B. R. Brooks, C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, J. Comput. Chem. **30**, 1545 (2009).
[56]J. Karanicolas and C. L. Brooks, 3rd, J. Mol. Biol. **334**, 309 (2003).
[57]L. W. Yang and C. P. Chng, Bioinform Biol Insights **2**, 25 (2008).
[58]L. Yang, G. Song, and R. L. Jernigan, Proc. Natl. Acad. Sci. U. S. A. **106**, 12347 (2009).
[59]W. Zheng, Biophys J **100**, 478 (2011).
[60]W. Zheng and M. Tekpinar, Biophys. J. **101**, 2981 (2011).
[61]M. Sagermann, W. A. Baase, B. H. Mooers, L. Gay, and B. W. Matthews, Biochemistry **43**, 1296 (2004).
[62]C. Clementi, A. E. Garcia, and J. N. Onuchic, J. Mol. Biol. **326**, 933 (2003).
[63]G. Radou, M. Enciso, S. Krivov, and E. Paci, J. Phys. Chem. B **117**, 13743 (2013).
[64]See supplementary material at http://dx.doi.org/10.1063/1.4905606 for 9 supplementary figures and 12 C_α coordinate files for selected snapshots of sMD trajectories .