Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins

Zhuqing Zhang and Hue Sun Chan[1]

Departments of Biochemistry, Molecular Genetics, and Physics, University of Toronto, Toronto, ON, Canada M5S 1A8

We compared folding properties of designed protein Top7 and natural protein S6 by using coarse-grained chain models with a mainly native-centric construct that accounted also for nonnative hydrophobic interactions and desolvation barriers. Top7 and S6 have similar secondary structure elements and are approximately equal in length and hydrophobic composition. Yet their experimental folding kinetics were drastically different. Consistent with experiment, our simulated folding chevron arm for Top7 exhibited a severe rollover, whereas that for S6 was essentially linear, and Top7 model kinetic relaxation was multiphasic under strongly folding conditions. The peculiar behavior of Top7 was associated with several classes of kinetic traps in our model. Significantly, the amino acid residues participating in nonnative interactions in trapped conformations in our Top7 model overlapped with those deduced experimentally. These affirmations suggest that the simple ingredients of native topology plus sequence-dependent nonnative interactions are sufficient to account for some key features of protein folding kinetics. Notably, when nonnative interactions were absent in the model, Top7 chevron rollover was not correctly predicted. In contrast, nonnative interactions had little effect on the quasi linearity of the model folding chevron arm for S6. This intriguing distinction indicates that folding cooperativity is governed by a subtle interplay between the sequence-dependent driving forces for native topology and the locations of favorable nonnative interactions entailed by the same sequence. Constructed with a capability to mimic this interplay, our simple modeling approach should be useful in general for assessing a designed sequence's potential to fold cooperatively.

chevron plot | desolvation | folding intermediate | S6 | Top7

The study of protein folding is important not only for deciphering the folding process and how misfolding can occur. The principles developed in theoretical investigations of folding (1–3) have provided insights into a broad range of molecular-recognition phenomena and dynamic behaviors in biology. Recent examples include protein–protein interactions (4), function of biomolecular machines (5), effects of desolvation in self-assembly (6, 7), and switch-like properties in binding (8). Among naturally evolved proteins, many fold cooperatively in a two-state-like manner (9), which is a remarkable feat from the vantage point of polymer physics (10). Although not all natural proteins share this property (11), its commonality argues that folding cooperativity may serve crucial biological functions such as guarding against harmful aggregation (12).

If cooperative folding can be a desirable trait under certain circumstances, a fundamental question arises: Can all folded globular structures attain a high degree of folding cooperativity? A revealing case is the designed protein Top7 (13), which folds to a de novo target structure that did not exist previously in the Protein Data Bank (PDB) but does so noncooperatively (14, 15). One possible reason for Top7's failure to fold cooperatively is that current sequence design techniques are inferior to natural selection in this regard (14, 15). A recent analysis suggested, however,

that a deeper cause might be that the Top7 topology itself is not conducive to cooperative folding (16). This view is consistent with several model studies indicating that native topology can constrain folding cooperativity (17–20). Thus, motivated by the same perspective that led to the considerations of a target structure's encodability (21) or designability (22), it is of interest to go a step further to assess a structure's designability for folding cooperativity.

The thermodynamic and kinetic manifestations of folding cooperativity are closely related because folding rates reflect the free energy barriers to folding. For instance, the negative correlation between native contact order and folding rate among small proteins (23) is consistent with predictions (18–20, 24) that proteins with higher native topological complexity tend to have higher folding barriers and thus fold more cooperatively. Here we focus mainly on folding kinetics, which affords more accurate characterizations of folding cooperativity than thermodynamics alone. A telling example is that even some Gō model proteins with an overall folding barrier are found to be short on kinetic folding cooperativity because the model chevron plots have significant rollovers (25, 26).

The experimental observation of a severe chevron rollover for Top7 indicates that nonnative interactions are at play in its complex, multiphasic folding kinetics (15). This finding prompted us to further pursue a recent native-centric model augmented by sequence-dependent nonnative hydrophobic interactions (27) and to use an improved version of this model to investigate the interplaying roles of native topology and placement of hydrophobic residues in Top7's noncooperative folding kinetics. As a control, we applied the same model to ribosomal protein S6 (28, 29), which folds much more cooperatively. Among several cooperatively folding proteins that have secondary structure elements similar to those in Top7 (e.g., acylphosphatase), we chose to study S6 because of the computational tractability engendered by its relatively fast folding rate.

Results

As detailed in *Methods* and *SI Text*, we used coarse-grained C^α chains to model the 92-residue Top7 (1qys) (13) and 97-residue S6 (1ris) (30) (Fig. 1). Two forms of native contact energy were considered: a 12-10 Lennard–Jones potential as in the common Gō-like model (31, 32) and a desolvation-barrier (db) potential that incorporates an energetic penalty against water expulsion (33–35). An improved excluded-volume term was implemented (Figs. S1 and S2). Favorable nonnative hydrophobic ($h\phi$) interactions specific to a protein's amino acid sequence were included
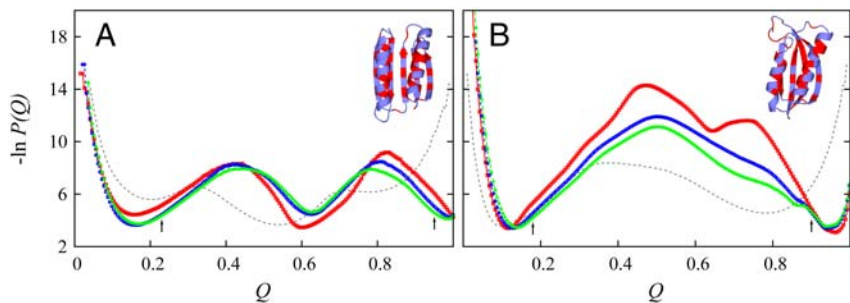
Fig. 1. Simulated midpoint free energy profiles and PDB structures of Top7 (*A*) and S6 (*B*). $P(Q)$ is normalized conformational population as a function of Q. In the ribbon diagrams, hydrophobic residues are in red; others are in blue. The profiles in red, blue, and green are for db + $h\phi$ models with, respectively, $\kappa^2/\varepsilon = 0$, 1, and 1.1. Profiles for the Gō models are in gray. The arrows indicate threshold Q_U and Q_F values used in our simulations of folding and unfolding kinetics. For Top7 and S6, respectively, $Q_U = 0.23$ and 0.18, and $Q_F = 0.95$ and 0.90. The profiles here are very similar to those obtained previously (16) with a slightly weaker excluded-volume repulsion (see *Methods*).

as well (16). The $h\phi$ compositions of Top7 (38/92 = 0.41) and S6 (43/97 = 0.44) are similar. Their different $h\phi$ packing patterns are highlighted in Fig. 1. Compared to S6, Top7 has more regularly alternating $h\phi$ and polar residues along its β-strands. However, Top7 has a 10-residue stretch (FAAILIKVFA from F63 to A72) within its C-terminal helix that contains 9 $h\phi$ residues. No such high local $h\phi$ density exists along the S6 sequence.

Native Topology Rationalizes Differences in Thermodynamic Folding Cooperativity of Top7 and S6. Fig. 1 provides the free energy profiles for several models of Top7 and S6 simulated near their respective transition midpoints. For Top7 (Fig. 1*A*), the Gō model free energy profile (no favorable nonnative interactions) shows a global minimum at $Q \approx 0.55$ with no barrier. With better accounting of desolvation effects, the Top7 profiles for the db + $h\phi$ models exhibit two barriers; i.e., folding is thermodynamically three-state. For the S6 models, every profile in Fig. 1*B* shows an overall barrier, indicating that folding is thermodynamically two-state. Because db enhances folding cooperativity (7, 33, 35), the overall barriers in the db + $h\phi$ S6 models are significantly higher than that in the Gō model. Consistent with experiment (14, 15, 28, 29), these model results demonstrate a lack of two-state folding cooperativity for Top7 and two-state-like folding for S6. Native topology is apparently a dominant factor that leads to the drastically different folding thermodynamics of Top7 and S6: Among the db + $h\phi$ models examined in Fig. 1, nonnative $h\phi$ interactions have only a minor impact on the free energy profiles of Top7. For S6, despite a rugged barrier (36) for the db ($\kappa^2 = 0$) model and a lowering of the folding barrier by nonnative $h\phi$ interactions, all free energy profiles are two-state-like.

We consider only the db + $h\phi$ models below because models that account for dbs are more realistic (6, 7, 33, 35, 37). In our analysis, the folded state and fully unfolded state were defined, respectively, by $Q \geq Q_F$ and $Q \leq Q_U$. These demarcations (arrows in Fig. 1) were chosen to be near either the unfolded (small Q) or folded ($Q \approx 1$) minimum but have a free energy (under midpoint conditions) $\approx 1.5 k_B T$ higher to allow for conformational fluctuations. (k_B is the Boltzmann constant, and T is absolute temperature.) To compare experimental data measured at variable denaturant concentration with results simulated at variable interaction strength ε/T, we fit theoretical equilibrium folding transition curves against their experimental counterparts to arrive at a linear relationship between [GuHCl] and ε/T (Fig. 2). This method is equivalent to matching theoretical and experimental free energies of unfolding ΔG_U (20) (see Fig. S3). The rationale is that, similar to ε/T, denaturants have a rather uniform effect on the entire protein (38). Discussion of related procedures is provided in *SI Text*. Fig. 2 shows the fits for $\kappa^2 = 1.1\varepsilon$ as well as the sigmoidal curves for fitting other κ^2 values. Fig. S3 shows that our fits entail extrapolating to zero denaturant from the transition region, because experimental data are not available for very low denaturant. In this regard, the premise of our linear fitting of ΔG_U against ε/T is no different from that on which the $\Delta G_U \sim 13$ kcal/mol estimate for Top7 (13) was based (Fig. S3*a*). It is instructive to note, however, that when partially folded con-

formations were included in the denatured population $1 - P_{folded}$ (including all $Q < Q_F$), both the simulated and experimental ΔG_Us in Fig. S3 exhibit nonlinear behaviors reminiscent of native-state hydrogen exchange isotherms (39).

Nonnative Hydrophobic Interactions Have Markedly Different Impacts on Top7 and S6 Folding Kinetics. Fig. 3 compares simulated and experimental folding/unfolding kinetics. Among the simulated folding chevron arms (filled symbols), those for the Top7 models display much more prominent rollovers (downward concavity) compared to those for the S6 models. (Simulation results for ε/Ts corresponding to [GuHCl] < 0 are considered to be experimentally inaccessible.) In the db models without nonnative $h\phi$ interactions (red squares), the folding arm of Top7 shows a rollover for [GuHCl] < 4 M, whereas that for S6 is essentially linear. When nonnative $h\phi$ interaction strength κ^2 is increased to 1.0ε (blue circles) and 1.1ε (green triangles), folding-arm rollover for Top7 becomes increasingly severe. Unlike the simple relationship between folding thermodynamics and kinetics for highly cooperative folders (35), the sensitivity of Top7 model kinetics to κ^2 was not apparent from the free energy profiles in Fig. 1*A* that show only minor variations with κ^2. For $\kappa^2 = 1.1\varepsilon$ in Fig. 3*A*, the Top7 model folding arm begins rolling downward at [GuHCl] ≈ 3 M. In
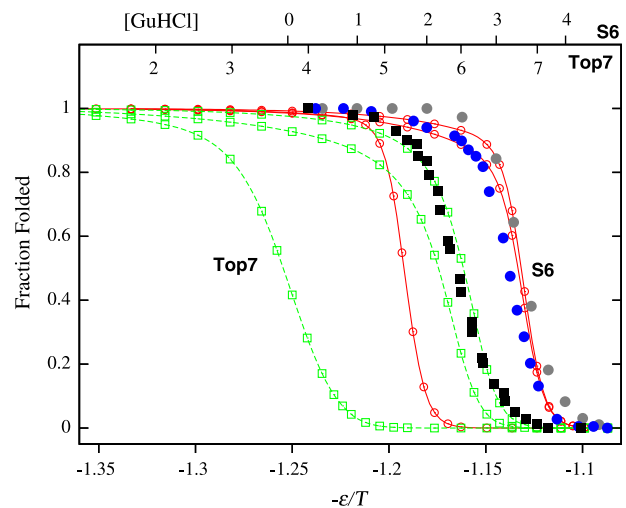

Fig. 2. Matching simulated and experimental folded fractions (P_{folded}). Filled symbols are experimental folded fractions of Top7 (black squares, from ref. 14) and S6 (blue and gray circles, from refs. 28 and 29, respectively) as functions of denaturant concentration [GuHCl] (top scales, in M). Curves through open symbols are simulated folded fractions of Top7 (green) and S6 (red) as functions of interaction strength $-\varepsilon/T$ (bottom scale) for db + $h\phi$ models with $\kappa^2/\varepsilon = 0$, 1, and 1.1 (from left to right for each set of curves). This figure shows the experimental [GuHCl] scales being fitted to the $\kappa^2 = 1.1\varepsilon$ models. Fits for other models were attained by moving the $-\varepsilon/T$ scale relative to the [GuHCl] scales. Midpoint ε/T value decreases with increasing κ^2 because nonnative $h\phi$ interactions destabilize the folded structure.
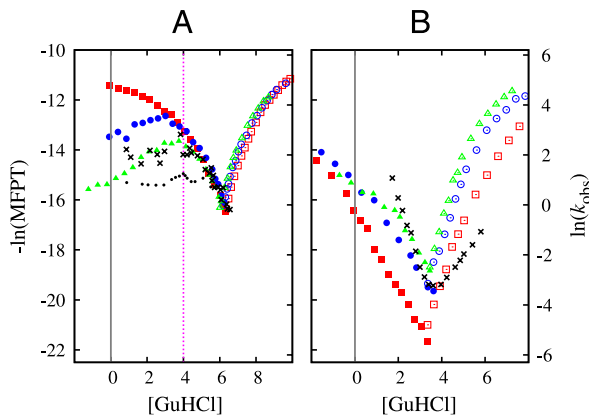
Fig. 3. Chevron plots for Top7 (*A*) and S6 (*B*). Data points in red, blue, and green provide negative logarithm of simulated mean first passage time (MFPT) of folding (filled symbols) and unfolding (open symbols), for db + hφ models with $\kappa^2/\varepsilon = 0$, 1, and 1.1, respectively. Dependence of model MFPT on ε/T is translated to that on [GuHCl] (in M; see Fig. 2). Black crosses are experimental data for S6 (28) and Top7 (single-exponential rate for [GuHCl] ≥ 4 M and fast-phase rate of the biexponential fit for [GuHCl] < 4 M in ref. 14; these data exhibit a trend similar to that of the middle phase in ref. 15). Vertical dashed lines mark [GuHCl] = 0 as well as the onset of severe experimental chevron rollover for Top7 at [GuHCl] = 4 M (14, 15). Black dots in (*A*) show $-\ln(A_{middle}/k_{middle} + A_{slow}/k_{slow})$ calculated from the experimental middle- and slow-phase data in ref. 15; this quantity corresponds to the negative logarithm of the MFPT contributed by these two phases. We did not include fast-phase data in this calculation because of uncertainties entailed by the negative A_{fast} values in ref. 15.

contrast, for the S6 models, although rollover increases somewhat with κ^2, the rollover is mild even for $\kappa^2 = 1.1\varepsilon$. Echoing effects of nonspecific nonnative interactions (40), the specific nonnative hφ interactions in our models speed up S6 folding, but they slow down Top7 folding under strongly folding conditions. In quantitative agreement with experiment (crosses in Fig. 3), transition-midpoint folding rates of Top7 models are ∼50 times faster than those of S6 models.

Fig. 3 argues strongly that nonnative hφ interactions are a critical part of real Top7 folding kinetics. In their absence, our simulation (red squares) failed utterly to mimic the experimental folding arm (black crosses). The onset of severe experimental rollover around [GuHCl] = 4 M (14) was reproduced only with a strong $\kappa^2 = 1.1\varepsilon$ (green triangles). This result suggests that the native-centric driving forces in Top7 are malleable; they can be overridden readily by nonnative forces. In contrast, the folding arms of the S6 models for different κ^2 values have only mildly different shapes (Fig. 3*B*). Thus, the native-centric driving forces in S6 are apparently robust against perturbations from the nonnative hφ interactions entailed by its own amino acid sequence. For Top7, as [GuHCl] decreases below 4 M, the chevron folding arm of the $\kappa^2 = 1.1\varepsilon$ model rolls downward, whereas the experimental middle and slow phases reported in ref. 15 show a nearly flat dependence. This apparent mismatch between theory and experiment might be partly caused by our model's uniform treatment of [GuHCl] dependence for all interaction types. Assessing the true extent of the mismatch is difficult, however, because of uncertainties regarding the fast phase in ref. 15. After all, the effective $-\ln(MFPT)$ calculated for the experimental middle and slow phases (black dots in Fig. 3*A*) does show a downward trend when [GuHCl] is decreased from 4 to ∼3 M because of increase in slow-phase amplitude. For S6, our models' chevron folding arms are similar to their experimental counterpart, but their slopes are not as steep. Thus, without many-body effects (10, 19, 20), the db + hφ models for S6 are still a bit short on reproducing the full folding cooperativity of real S6.

Complex Top7 Folding Arises from Transiently Trapped Misfolded Conformations. We now focus on the db + hφ, $\kappa^2 = 1.1\varepsilon$ Top7 model by which the overall experimental chevron trend was reasonably reproduced (Fig. 3*A*). Folding relaxation of this model is essentially single-exponential near the transition midpoint (1.16 ≤ ε/T ≤ 1.23, Fig. 4*A*) but clearly not single-exponential for stronger folding conditions ($\varepsilon/T > 1.25$, Fig. 4*B*). Following a common practice in analyzing experimental data (15), we fitted the unfolded population P(unfolded) to $\Sigma_i A_i \exp(-k_i t)$ with one, two, or three exponentials, where k_is are relaxation rates and t is time (Fig. 4*C*): Near the midpoint (1.16 ≤ ε/T < 1.25), the only reasonable fit is a single exponential. Then, if three exponentials were used to fit the rest of our data (ε/T ≥ 1.25), two of the fitted k_is turned out to be approximately equal for the ε/T values plotted in red in Fig. 4*C*. Thus, only two exponentials were used for those data. Fig. 4*C* summarizes our fits. As folding conditions become stronger (ε/T increases), k_1 for the fast phase (top curve) increases monotonically, but those of the middle and slow phases (k_2 and k_3, other curves) show a downward trend for ε/T ≥ 1.3 and ε/T ≥ 1.25, respectively. Consistent with experiment (15), the middle and slow phases in Fig. 4*C* display more prominent rollovers than the fast phase. Nonetheless, two issues that will require more effort should be noted. First, the middle- and slow-phase k_is in Fig. 4*C* decrease with increasing ε/T, but the corresponding experimental rates were essentially independent of [GuHCl] under strongly folding conditions (see above). Second, all fitted amplitudes in our model were positive (with middle-phase $A_2 < 0.3$, slow-phase $A_3 < 0.2$), whereas the experimental fast-phase amplitude was negative.
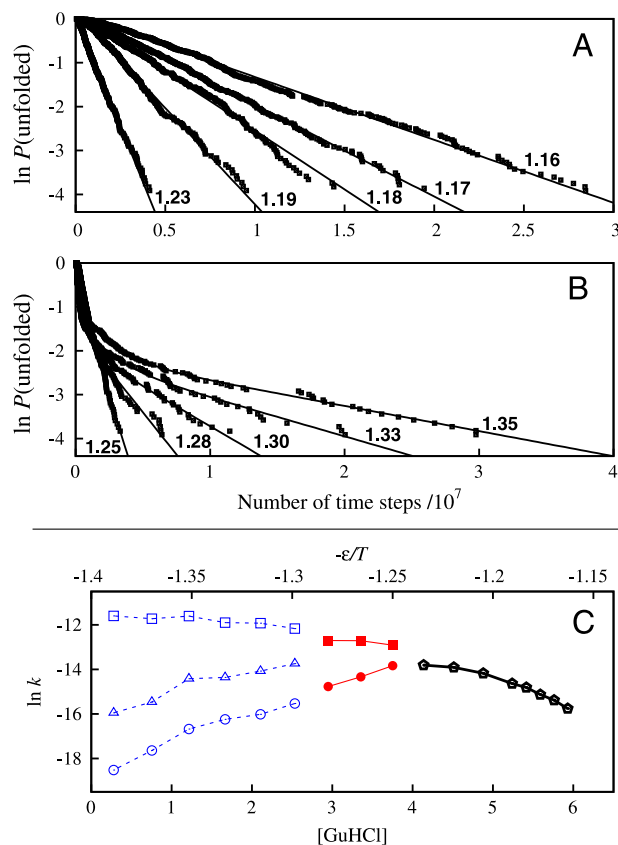

Fig. 4. Simulated folding relaxation for Top7 for the db + hφ model with $\kappa^2 = 1.1\varepsilon$. Data points in (*A*) and (*B*) show relaxation behaviors for selected ε/T values (as marked); curves are single or multiple exponential fits. Data points in (*C*) show the rates (k_is) from single- (black), two- (red), and three-exponential (blue) fits as functions of $-\varepsilon/T$ and [GuHCl] (using the match in Fig. 2).

Zhang and Chan

Although purely native-centric interactions can be consistent with mild chevron rollovers (25, 32) (see red squares for $\kappa^2 = 0$ in Fig. 3A), a severe rollover such as that exhibited by the green triangles ($\kappa^2 = 1.1\varepsilon$) in Fig. 3A are indicative of deep kinetic traps with nonnative contacts (3). We delineated the structural species underlying the complex Top7 model folding kinetics by monitoring both native and nonnative contacts along folding trajectories (Fig. 5A–D) and characterizing the transient intermediate structures (Fig. 5E–G). Near the midpoint (Fig. 5A, ε/T equivalent to [GuHCl] = 6 M), a partially folded ensemble with $Q \sim 0.6$ and $N_{\text{nonnative}} \sim 20$ nonnative contacts (green-shaded region) was populated occasionally. These conformations tend to have a folded C-terminal fragment and a disordered N-terminal fragment (Fig. 5E), similar to the equilibrium intermediate state under midpoint conditions (the "I state" in ref. 16). We now call this state I_0. When folding conditions were strengthened to an ε/T equivalent to [GuHCl] = 4 M (Fig. 5B), two intermediates with more nonnative contacts appeared more frequently (gray- and orange-shaded regions). Their prominence increases when folding conditions were further strengthened (Fig. 5C and D). Conformations in the gray-shaded regions have large $N_{\text{nonnative}} \sim 60$ but only $Q \sim 0.5$. Structurally, they are typified by Fig. 5F, which shows the C-terminal helix threading through an opening between the N-terminal helix and three β-strands. We call this ensemble I_1. These misfolded conformations clearly have to first unwrap before they can reach the Top7 native structure. In contrast, conformations in the orange-shaded regions have a high native content ($Q \sim 0.8$) and are less nonnative ($N_{\text{nonnative}} \sim 35$). Typified by Fig. 5G, these conformations, which we refer to collectively as I_2, may be viewed as mispacked, because both their N-terminal β-hairpin and C-terminal fragment are native-like but the N-terminal β-hairpin is far away from the C-terminal β-strands, and thus fail to form the native 5-strand β-sheet. More examples of folding trajectories are provided in Figs. S4–S6.

Nonnative Interactions Play Significant Roles in Top7 Folding. Fig. 6 gives the distribution of nonnative contacts in I_1 (upper left) and I_2 (lower right). Experiment has implicated L29, V48, F63, A64, A65, L67, and V81 in the formation and/or stabilization of nonnative states (15). Consistent with this finding, Fig. 6 shows that 6

of these 7 residues (except V81) participated in significant nonnative $h\phi$ interactions in the kinetic intermediates of our model: (*i*) L29 made contact with A65 (itself suggested by experiment to be involved in nonnative interactions), I66, and V68 in I_1; L29 took part in nonnative contact with I4, V6, and Y21 in I_2 with even higher probabilities. (*ii*) V48 had one strong nonnative contact with Y39 in I_2. (*iii*) F63, A64, A65, and L67 are part of the extended stretch of $h\phi$ residues in the C-terminal helix (see above). These four residues participated in numerous nonnative interactions in both intermediates, particularly in an extended region in I_1 between the C-terminal helix on one hand and the N-terminal helix and β-strands on the other (Fig. 5F). In Fig. 6, these contacts cluster along a horizontal band with a vertical span covering positions 63 to ~71. We have also characterized two transition states along Top7's free energy profile (Fig. 1A). Unlike small natural proteins that fold through transition states with native-like topologies (41), our model predicts that nonnative interactions are prevalent in Top7 folding transition states (see *SI Text* and Fig. S7).

Discussion

By modeling specific nonnative $h\phi$ interactions, we have added substantially to the advances made in previous atomic simulation of the C-terminal fragment of Top7 (42) and Gō-like modeling of S6 (43, 44). Our results suggest that the drastically different folding kinetics of Top7 and S6 is an outcome of an omnipresent interplay between native-centric driving forces and sequence-specific nonnative $h\phi$ interactions. This interplay is fundamentally a competition between native and alternative topologies, with many possible contributing factors. The scope of our study is limited. For example, the near-ideal geometry of the designed β-sheet in Top7 may make it easier to form nonnative strand arrangements, but tackling this question would require structural details beyond those considered in our model.

Despite our model's simplicity, it succeeded in rationalizing key differences in the folding kinetics of Top7 and S6. Many nonnative interactions in our Top7 model involve a long stretch of $h\phi$ residues, which is uncommon in natural globular proteins (45). We expect that much of Top7's experimental folding complexity is related to this peculiarity. By comparison, S6 does not have
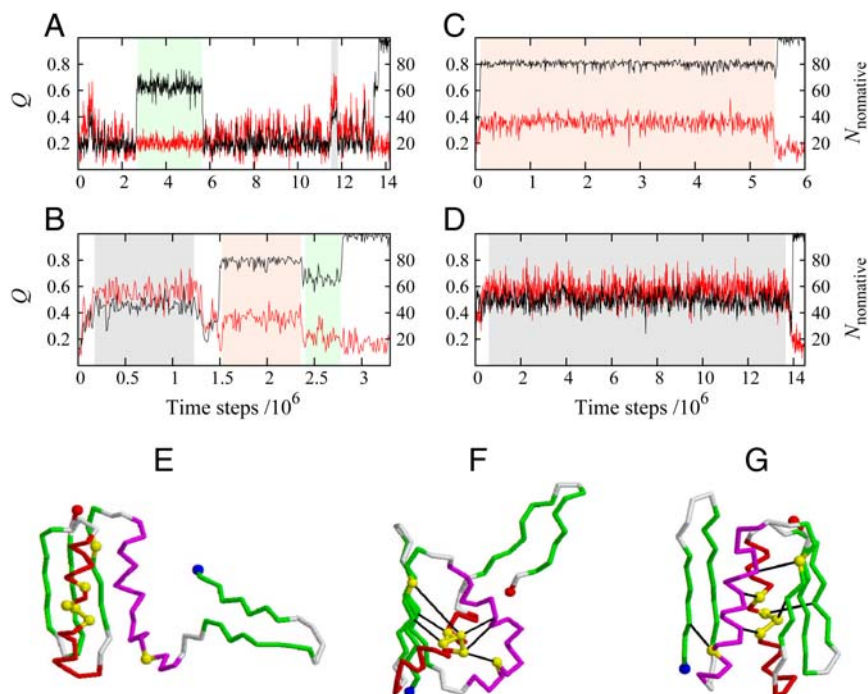
Fig. 5. Simulated trajectories, folding intermediates, and kinetic traps of Top7 for the db + $h\phi$ model with $\kappa^2 = 1.1\varepsilon$. (A–D) Fractional number of native contact Q (black traces, left scales) and number of nonnative contact $N_{\text{nonnative}}$ (red traces, right scales) as functions of time. Data points at the end of every 5,000 simulation time steps were tracked. Example trajectories were simulated near the transition midpoint at $\varepsilon/T = 1.16$ (A), around the top of the chevron rollover at $\varepsilon/T = 1.25$ (B), and under strongly folding conditions at $\varepsilon/T = 1.32$ (C and D). (E) A typical I_0 conformation sampled during time periods shaded in green in (A) and (B). (F) A trapped structure representative of the I_1 conformations sampled during time periods shaded in gray in (B) and (D). (G) Another trapped structure, representative of the I_2 conformations sampled during time periods shaded in orange in (B) and (C). The N and C termini of the structures are depicted as blue and red spheres, respectively. Six of the seven residues suggested by experiment to stabilize nonnative states (see text) are marked as yellow spheres. Black lines in (F) and (G) indicate examples of significant nonnative contacts involving these residues (chosen from the maps in Fig. 6).

such a long $h\phi$ stretch, and effects of nonnative $h\phi$ interactions (27, 39, 46) on experimental S6 folding are mild (47), possibly owing in part to the effects of "gatekeeper" residues (44). As a test, we have considered a db + $h\phi$ model for a hypothetical S6 mutant with four $h\phi$ substitutions to create a long $h\phi$ stretch similar to that in Top7. This model (Fig. S8) predicts a severe chevron rollover, suggesting in general that a long $h\phi$ stretch would likely lead to complex folding kinetics. However, it is not known experimentally whether such a sequence that increases the $h\phi$ composition by ∼10% yet still folds to the same S6 native structure actually exists. It remains to be elucidated as well whether the long $h\phi$ stretch or some similar $h\phi$ sequence patterns are necessary for encoding the Top7 native topology. Thus, it will be enlightening to explore the experimental viability of designing the Top7 fold with no long $h\phi$ stretch and designing the S6 fold with a long $h\phi$ stretch. Results from such experiments may go a long way toward deciphering the origin of folding cooperativity.

The present results are broadly in line with the perspective that intermediates in protein folding are common (46). The central challenge in the interpretation of multiphasic folding kinetics is to infer the intermediate state(s) along the folding pathway(s). For Top7, Watters et al. proposed a four-state model on the basis of a three-exponential fit of the experimental relaxation data (15). Because our model relaxation data can also be fitted by three exponentials (Fig. 4), we used our results to explore the relationship between multiphasic kinetics and folding intermediates by inspecting a total of ∼1,400 folding trajectories under three classes of conditions: near midpoint ($\varepsilon/T = 1.16, 1.19$), at the onset of severe chevron rollover ($\varepsilon/T = 1.25, 1.27$), and strongly favorable to folding ($\varepsilon/T = 1.32, 1.37$). Each trajectory was started from an open conformation in the unfolded state (U) and ended in the native state (N); examples are provided in Fig. 5A–D and Figs. S4–S6. Using Q and $N_{\text{nonnative}}$ of transient populations to identify them with either I_0, I_1, or I_2, we observed that a significant fraction (∼16%) of all trajectories may be described as direct U → N pathways (e.g., the middle panel in the third row of Fig. S4). These trajectories apparently passed through the I_0 minimum along the thermodynamic free energy profile (Fig. 1A) rather quickly. Among pathways that involved transiently populated intermediates, U → I_0 → N was most probable near the midpoint (>25%), but it became less likely under stronger folding conditions. In comparison, although U → I_1 → N and U → I_2 → N were rare near the midpoint, together they accounted for ∼12% of the folding trajectories under strongly folding conditions. Interestingly, in these pathways, the trapped I_1 or I_2 did not always have to go through an appreciable period of time in the U state before reaching N. Other less common pathways that we have observed include U → I_1 → I_2 → N, U → I_1 → I_0 → N, U → I_2 → I_0 → N, U → I_0 → I_2 → I_0 → N, U → I_2 → I_0 → I_2 → N, U → I_1 → I_2 → I_0 → N, and U → I_1 → I_0 → I_2 → N, but no pathway in which I_2 appeared before I_1 was observed. We also came across many trajectories that could not be identified with a pathway through a series of well-defined intermediate states. All in all, the diversity of these pathways indicates that although individual folding trajectories can be represented by various sequential pathways, there is no guarantee that complex folding kinetics such as that exhibited by Top7 is controlled by a simple network of transitions among a few discrete states.

In summary, we reemphasize that our model formulation should be useful as a tool in protein design to assess a target structure's potential for cooperative folding (16). Recent progress in NMR techniques has made it possible to obtain detailed structural information on "invisible," low-populated protein states (48, 49). In view of the agreement with experiment achieved here and our approach's previous success in predicting specific nonnative interactions in the folding transition state of Fyn SH3 mutants (27), the predicted misfolded and mispacked Top7 struc-

tures (Figs. 5 and 6) as well as the nonnative contacts predicted for its transition states (Fig. S7) are worthy candidates to be tested by structural experiments. The present study has opened a window into an intriguing interplay between native topology and nonnative hydrophobic interactions. Deeper insights await further efforts in both theory and experiment.

Methods

Our coarse-grained C^α models were modified from (i) the common Gō-like model (31, 32) and (ii) a native-centric model that accounted for dbs (33–35). The db model is now augmented by sequence-dependent nonnative $h\phi$ interactions (16, 27). Native contacts were determined from PDB structures by using a 4.5-Å separation cutoff between non-hydrogen atoms. Model potentials were parametrized by an energy $\varepsilon > 0$, a db height $\varepsilon_{\text{db}} = 0.1\varepsilon$, and a solvent-separated-minimum depth $\varepsilon_{\text{ssm}} = 0.2\varepsilon$ as before (16, 35). The only modification we have made to the formulation in ref. 16 is on the excluded-volume part of the nonnative interactions. In previous studies, our group followed refs. 31 and 33 in using a repulsive term $\varepsilon(r_{\text{rep}}/r_{ij})^{12}$ between residues not in contact in the native structure, where r_{ij} is the separation between C^α positions i and j, with $r_{\text{rep}} = 4.0$ Å (16, 32, 34, 35). This construct is adequate for many applications. But for predicting misfolded structures, $r_{\text{rep}} = 4.0$ Å is not always sufficient to avoid steric clashes in real proteins (Fig. S1). To alleviate this limitation, we adjusted r_{rep} on the basis of the following heuristic considerations.

(i) We retained 4.0 Å as the lower bound on r_{rep} because 4.0 Å is approximately the contact distance between two methanes (10), and thus it should correspond roughly to the closest possible approach between two C^α positions. (ii) The average volume of an amino acid residue ≈140 Å³ (50). If this volume were spherical, its radius would be ≈3.2 Å, and the separation between the centers of two such spheres would be ≈6.4 Å. Because real amino acid residues are not spheres, we stipulated a slightly smaller upper bound of 6 Å on r_{rep}. (iii) The maximum $r_{\text{rep}} = 6$ Å does not apply to all nonnative contact pairs because heterogeneity in residue sizes and protein core packing can result in a separation $r_0 < 6$ Å in the native structure between C^α positions that are not in native contact. For such cases, alternate packing between different rotamer pairs that allow for a closer approach between the C^α positions than that in the native structure might be possible; thus we chose $0.75r_0$ as the repulsion radius. Taking these considerations together, we arrived at a modified repulsive energy between nonnative pairs $E_{\text{rep}} = \varepsilon(\sigma_0/r_{ij})^{12}$, wherein
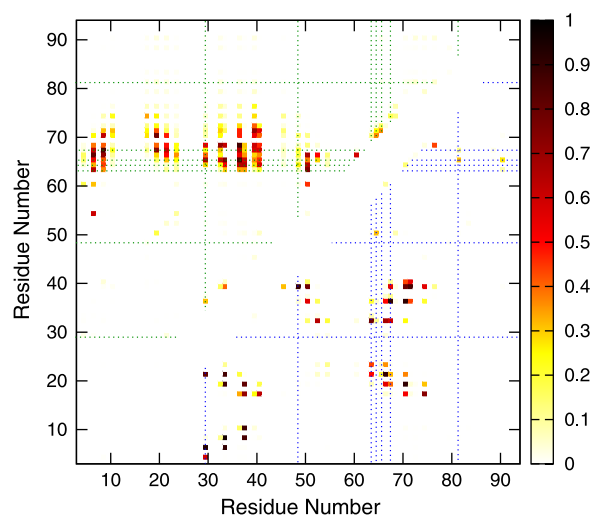

Fig. 6. Nonnative contact maps of simulated Top7 kinetic traps. Probabilities of nonnative contacts for I_1 and I_2: The upper left map is for I_1, determined from the gray-shaded regime in Fig. 5D; the sampled conformations are typified by Fig. 5F. The lower right map is for I_2, determined from the orange-shaded period in Fig. 5C; the sampled conformations are typified by Fig. 5G. Residues suggested by experiment to be involved in nonnative interactions (see text) are identified by dotted lines. Residue numbering in our contact maps is identical to that in the PDB.

$$\sigma_0 = \begin{cases} 4\ \text{Å} & \text{when } r_0\Delta^{(1/12)} < 4\ \text{Å}, \\ r_0\Delta^{(1/12)} & \text{when } 4\ \text{Å} < r_0\Delta^{(1/12)} < 6\ \text{Å}, \\ 6\ \text{Å} & \text{when } r_0\Delta^{(1/12)} > 6\ \text{Å}, \end{cases} \qquad [1]$$

and $\Delta^{(1/12)} = 0.75$ ($\Delta = 0.03$). E_{rep} is the only term in the $U_{\text{nonnative}}$ potential for nonnative interactions not involving hydrophobic residues (Fig. S2a).

As in our previous work, the attractive part of the nonnative $h\phi$ interactions is the sum $E_{\text{HP}} = -\Sigma_i\Sigma_j\kappa_i\kappa_j \exp[-(r_{ij} - \sigma_{h\phi})^2/2]$ over $h\phi$ residues, which include alanine, valine, leucine, isoleucine, methionine, tryptophan, phenylalanine, and tyrosine (27) in the Top7 or S6 sequence. We do not distinguish between hydrophobic residue types, so we set $\kappa_i = \kappa$. In view of the above modification on r_{rep}, we set $\sigma_{h\phi} = \sigma_0 + 1$ Å (instead of $\sigma_{h\phi} = 5$ Å in ref. 27). Further setting $K_{\text{HP}} = 1$ in the original formulation (27) to simplify notation, the total nonnative interaction between a pair of $h\phi$ residues is now given by

$$U_{\text{nonnative}} = \varepsilon(\sigma_0/r_{ij})^{12} - \kappa^2 \exp[-(r_{ij} - \sigma_0 - 1\ \text{Å})^2/2]. \qquad [2]$$

Examples of this $h\phi$ term are provided in Fig. S2b. Most of our results are for the db model augmented with nonnative $h\phi$ interactions, in which case the total potential is $E_0 - \kappa^2\Sigma_i\Sigma_j \exp[-(r_{ij} - \sigma_0 - 1\ \text{Å})^2/2]$, where E_0 is the db potential with $E_{\text{rep}} = \varepsilon(\sigma_0/r_{ij})^{12}$. We refer to this class of constructs as the db + $h\phi$ model. The purely native-centric db model is equivalent to a db + $h\phi$ model with $\kappa^2 = 0$. Further details of the simulation procedures are provided in SI Text.

1. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins*, 21:167–195.
2. Thirumalai D, Woodson SA (1996) Kinetics of folding of proteins and RNA. *Acc Chem Res*, 29:433–439.
3. Chan HS, Dill KA (1998) Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins*, 30:2–33.
4. Wang W, Xu W, Levy Y, Trizac E, Wolynes PG (2009) Confinement effects on the kinetics and thermodynamics of protein dimerization. *Proc Natl Acad Sci USA*, 106:5517–5522.
5. Hyeon C, Onuchic JN (2007) Mechanical control of the directional stepping dynamics of the kinesin motor. *Proc Natl Acad Sci USA*, 104:17382–17387.
6. Levy Y, Onuchic JN (2006) Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct*, 35:389–415.
7. Ferguson A, Liu Z, Chan HS (2009) Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J Mol Biol*, 389:619–636 and addendum (2009) 392:242.
8. Borg M, et al. (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci USA*, 104:9650–9655.
9. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor-2.1. Evidence for a 2-state transition. *Biochemistry*, 30:10428–10435.
10. Chan HS, Shimizu S, Kaya H (2004) Cooperativity principles in protein folding. *Methods Enzymol*, 380:350–379.
11. Tompa P, Fuxreiter M (2008) Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci*, 33:2–8.
12. Dobson CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci*, 24:329–332.
13. Kuhlman B, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302:1364–1368.
14. Scalley-Kim M, Baker D (2004) Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J Mol Biol*, 338:573–583.
15. Watters AL, et al. (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, 128:613–624.
16. Zhang Z, Chan HS (2009) Native topology of the designed protein Top7 is not conducive to cooperative folding. *Biophys J*, 96:L25–L27.
17. Abkevich VI, Gutin AM, Shakhnovich EI (1995) Impact of local and nonlocal interactions on thermodynamics and kinetics of protein folding. *J Mol Biol*, 252:460–471.
18. Zuo G, Wang J, Wang W (2006) Folding with downhill behavior and low cooperativity of proteins. *Proteins*, 63:165–173.
19. Cho SS, Weinkam P, Wolynes PG (2008) Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci USA*, 105:118–123.
20. Badasyan A, Liu Z, Chan HS (2008) Probing possible downhill folding: Native contact topology likely places a significant constraint on the folding cooperativity of proteins with ~40 residues. *J Mol Biol*, 384:512–530.
21. Chan HS, Dill KA (1996) Comparing folding codes for proteins and polymers. *Proteins*, 24:335–344.
22. Li H, Helling R, Tang C, Wingreen N (1996) Emergence of preferred structures in a simple model of protein folding. *Science*, 273:666–669.
23. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*, 277:985–994.
24. Knott M, Chan HS (2006) Criteria for downhill protein folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins*, 65:373–391.
25. Kaya H, Chan HS (2003) Origins of chevron rollovers in non-two-state protein folding kinetics. *Phys Rev Lett*, 90:258104.
26. Zhou Y, Zhang C, Stell G, Wang J (2003) Temperature dependence of the distribution of the first passage time: Results from discontinuous molecular dynamics simulations of an all-atom model of the second β-hairpin fragment of protein G. *J Am Chem Soc*, 125:6300–6305.
27. Zarrine-Afsar A, et al. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc Natl Acad Sci USA*, 105:9999–10004.
28. Miller EJ, Fischer KF, Marqusee S (2002) Experimental evaluation of topological parameters determining protein-folding rates. *Proc Natl Acad Sci USA*, 99:10359–10363.
29. Otzen DE, Oliveberg M (2004) Correspondence between anomalous m- and ΔC_p-values in protein folding. *Protein Sci*, 13:3253–3263.
30. Lindahl M, et al. (1994) Crystal structure of the ribosomal protein S6 from Thermus thermophilus. *EMBO J*, 13:1249–1254.
31. Clementi C, Nymeyer H, Onuchic JN (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol*, 298:937–953.
32. Kaya H, Chan HS (2003) Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: How adequate is native-centric topological modeling?. *J Mol Biol*, 326:911–931 and corrigendum: (2004) 337:1069–1070.
33. Cheung MS, García AE, Onuchic JN (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc Natl Acad Sci USA*, 99:685–690.
34. Liu Z, Chan HS (2005) Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J Mol Biol*, 349:872–889.
35. Liu Z, Chan HS (2005) Solvation and desolvation effects in protein folding: Native flexibility, kinetic cooperativity, and enthalpic barriers under isostability conditions. *Phys Biol*, 2:S75–S85.
36. Oliveberg M, Tan YJ, Silow M, Fersht AR (1998) The changing nature of the protein folding transition state: Implications for the shape of the free-energy profile for folding. *J Mol Biol*, 277:933–943.
37. MacCallum JL, Moghaddam MS, Chan HS, Tieleman DP (2007) Hydrophobic association of α-helices, steric dewetting, and enthalpic barriers to protein folding. *Proc Natl Acad Sci USA*, 104:6206–6210.
38. Myers JK, Pace CN, Scholtz JM (1995) Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci*, 4:2138–2148.
39. Feng H, Zhou Z, Bai Y (2005) A protein folding pathway with multiple folding intermediates at atomic resolution. *Proc Natl Acad Sci USA*, 102:5026–5031.
40. Plotkin SS (2001) Speeding protein folding beyond the Go model: How a little frustration sometimes helps. *Proteins*, 45:337–345.
41. Pandit AD, Jha A, Freed KF, Sosnick TR (2006) Small proteins fold through transition states with native-like topologies. *J Mol Biol*, 361:755–770.
42. Mohanty S, Meinke JH, Zimmermann O, Hansmann UHE (2008) Simulation of Top7 CFr: A transient helix extension guides folding. *Proc Natl Acad Sci USA*, 105:8004–8007.
43. Hubner IA, Oliveberg M, Shakhnovich EI (2004) Simulation, experiment, and evolution: Understanding nucleation in protein S6 folding. *Proc Natl Acad Sci USA*, 101:8354–8359.
44. Stoycheva AD, Brooks CL, Onuchic JN (2004) Gatekeepers in the ribosomal protein S6: Thermodynamics, kinetics, and folding pathways revealed by a minimalist protein model. *J Mol Biol*, 340:571–585.
45. Irbäck A, Sandelin E (2000) On hydrophobicity correlations in protein chains. *Biophys J*, 79:2252–2258.
46. Brockwell DJ, Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes?. *Curr Opin Struct Biol*, 17:30–37.
47. Otzen D (2005) Antagonism, non-native interactions and non-two-state folding in S6 revealed by double-mutant cycle analysis. *Protein Eng Des Sel*, 18:547–557.
48. Korzhnev DM, Kay LE (2008) Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding. *Acc Chem Res*, 41:442–451.
49. Ollerenshaw JE, Kaya H, Chan HS, Kay LE (2004) Sparsely populated folding intermediates of the Fyn SH3 domain: Matching native-centric essential dynamics and experiment. *Proc Natl Acad Sci USA*, 101:14748–14753.
50. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules*, 18:534–552.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY