

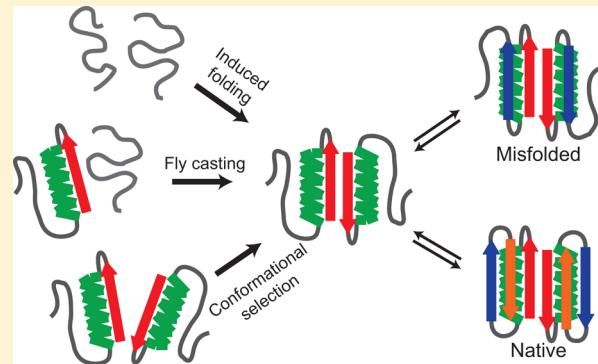
Atomistic Description of the Folding of a Dimeric Protein

Stefano Piana,^{*,†} Kresten Lindorff-Larsen,^{†,§} and David E. Shaw^{*,†,‡}

[†]D. E. Shaw Research, New York, New York 10036, United States

[‡]Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, United States

ABSTRACT: Equilibrium molecular dynamics simulations are increasingly being used to describe the folding of individual proteins and protein domains at an atomic level of detail. Isolated protein domains, however, are rarely observed *in vivo*, where multidomain proteins and multimeric assemblies are far more common. It is clear that the folding of such proteins is often inextricably coupled with the process of dimerization; indeed, many protein monomers and protein domains are not stable in isolation, and fold to their native structures only when stabilized by interactions with other members of a protein complex. Here, we use equilibrium molecular dynamics simulations with an aggregate simulation length of 4 ms to elucidate key aspects of the folding mechanism, and of the associated free-energy surface, of the Top7-CFr dimer, a 114-amino-acid protein homodimer with a mixed α/β structure. In these simulations, we observed a number of folding and unfolding events. Each folding event was characterized by the assembly of two unfolded Top7-CFr monomers to form a stable, folded dimer. We found that the isolated monomer is unstable but that, early in the folding pathway, nascent native structure is stabilized by contacts between the two monomer subunits. These contacts are in some cases native, as in an induced-folding model, and in other cases non-native, as in a fly-casting mechanism. Occasionally, folding by conformational selection, in which both subunits form independently before dimerization, was also observed. Folding then proceeds through the sequential addition of strands to the protein β sheet. Although the long-time-scale relaxation of the folding process can be well described by a single exponential, these simulations reveal the presence of a number of kinetic traps, characterized by structures in which individual strands are added in an incorrect order.



INTRODUCTION

Proteins are synthesized in the cell or *in vitro* as unstructured polypeptide chains that self-assemble into their functionally active three-dimensional shapes. This self-assembly process, known as protein folding, is one of the fundamental processes of life, and has been the subject of many experimental studies. Advances in computer hardware and algorithms, together with improvements in the accuracy of the physical models used in molecular dynamics simulations, have made it feasible to perform equilibrium, explicit-solvent simulations of the folding of small protein domains at an atomic level of detail.^{1,2}

To date, such atomic-level simulations have generally been limited to the folding of isolated, monomeric protein domains.^{3–9} Most of the proteome, however, consists of multidomain proteins, often in multimeric assemblies. With few exceptions (e.g., refs 10–13), such assemblies have been studied using coarse-grained, native-topology-based models.^{14–23} The results of these studies have suggested alternative mechanisms that may be operative in the folding of various protein dimers:^{24,25} In the simplest scenario, which is often referred to as “conformational selection”, dimerization takes place only after the interfacial regions of the individual monomers have independently folded into their (dimeric) native states. In more complex cases, sometimes described as “induced folding”, the folding and dimerization processes are

temporally and mechanistically intertwined; in this scenario, which is typical of assemblies with large dimeric interfaces, the monomers are unstable in isolation, and begin to dimerize early in the folding process.

Here we report the results of physics-based, fully atomistic equilibrium molecular dynamics (MD) simulations of the folding of a protein homodimer, Top7-CFr. The Top7-CFr monomer is the 57-amino-acid C-terminal fragment of Top7, a protein designed in the Baker lab.²⁶ During expression of Top7 in *E. coli*, the Top7-CFr monomer results from a mistranslation.²⁷ Analysis of Top7-CFr shows that it folds into a very stable symmetric dimer composed of two identical fragments (in contrast, Top7 is a monomeric protein under a very wide range of concentrations²⁷). Kinetic studies of the folding of Top7-CFr in denaturant indicate that this protein folds on the millisecond time scale at 10 μM concentration.²⁸ The folding rate is likely to be even faster at the protein concentrations characteristic of MD simulations, which are typically much higher (in the millimolar range). Under these

Special Issue: Peter G. Wolynes Festschrift

Received: February 28, 2013

Revised: July 19, 2013

Published: July 24, 2013

conditions, sub-millisecond folding is expected, allowing the observation of multiple folding/unfolding events in millisecond atomistic MD simulations. We performed four equilibrium MD simulations of the Top7-CFr monomer and four of the Top7-CFr dimer close to the dimer melting temperature using a state-of-the-art physics-based force field and an explicit representation of the solvent molecules. In the 4 ms of aggregate simulation time, we observed several folding and unfolding events. We used these simulations to reconstruct the folding free-energy surface of this protein and to characterize its folding mechanism.

We find that, under the conditions used in our simulations, Top7-CFr behaves as an obligatory dimer.²³ The isolated monomer is unstable, and the dimer can only fold when, in the early steps of folding, the dimer interface is stabilized by native and/or non-native intermonomer interactions. The mechanisms observed in these initial folding steps varied from one case to another, and ranged from induced folding to conformational selection. Folding then proceeds from the interface toward the remainder of the dimer. In the process, however, folding errors can occur,²⁹ consisting of incorrect additions of β -strand structure. These β -strand insertions produce persistent misfolded structures, which act as kinetic traps.

METHODS

Simulations of the Top7-CFr monomers were started from an extended conformation. Simulations of the Top7-CFr dimer were started from either an extended conformation or from PDB entry 2GJH.²⁷ Initial structures of the monomer were solvated in cubic $61 \times 61 \times 61 \text{ \AA}^3$ boxes, containing 6574 water molecules and 0.1 M NaCl. The dimers were solvated in cubic $75 \times 75 \times 75 \text{ \AA}^3$ boxes, containing 12 497 water molecules and 0.1 M NaCl. The dimer concentration in simulation was ~ 8 mM. The CHARMM22* force field³⁰ was used to represent the system. Initial test runs of the Top7-CFr dimer, starting from the experimental structure, were performed to estimate the protein's melting temperature. The resulting value of 400 K is consistent with the experimental observation that the melting temperature for the Top7-CFr dimer is >373 K.²⁷ The systems were equilibrated at 400 K and 1 bar in the NPT ensemble; production runs at 400 K were performed in the NVT ensemble^{31,32} with the Anton specialized hardware³³ using a 2.5 fs time step. The choice of the NVT ensemble for the production run was made for performance reasons. A previous investigation indicated that this has a negligible effect on the results.³⁴ Bonds involving hydrogen atoms were restrained to their equilibrium lengths using the M-SHAKE algorithm.³⁵ Nonbonded interactions were truncated at 9.5 \AA with a shifted-force approach.^{34,36}

Four simulations, of 850, 880, 780, and 760 μs , were performed for the Top7-CFr dimer (two starting from an extended conformation and two starting from the native conformation), for a total of 3.2 ms of simulation time. Four simulations, of 210, 208, 198, and 172 μs , were performed for the Top7-CFr monomer, for a total of 0.8 ms of simulation time. Folded states, unfolded states, transition pathways for Top7-CFr monomer formation, and transition pathways for Top7-CFr dimer formation were defined using a transition-based assignment algorithm.^{30,37–39} This analysis requires the identification of one or more descriptors (e.g., RMSD from native or fraction of native contacts) that can distinguish the folded from the unfolded state, along with the definition of two

cutoffs: one for the folded state and one for the unfolded state. The results are usually robust with respect to the exact choice of the two cutoffs, in particular when smoothing of the time series is performed.³⁰ In the case of the Top7-CFr monomer simulations, this analysis was performed on the RMSD time series of the $C\alpha$ atoms of residues 2–50 (these residues are numbered as they are in PDB entry 2GJH). The $C\alpha$ RMSD time series were smoothed by taking a running average over a 5 ns window. A cutoff of 3.0 \AA was used to define the folded state, and 6.0 \AA was used for the unfolded state. To identify the misfolded state, cutoffs of 2.5 and 6.0 \AA were used on the $C\alpha$ RMSD time series of residues 2–40 only, as residues 41–57 are disordered in this state. The reference structure was the average monomer structure of state I₁ in the dimer simulation. In the case of the Top7-CFr dimer simulations, this analysis was performed simultaneously on the RMSD time series of the $C\alpha$ atoms of the interface β sheet (residues 3–9; cutoffs of 1.0 and 4.0 \AA), the $C\alpha$ atoms of the full interface (residues 3–29; cutoffs of 2.5 and 5.0 \AA), and the $C\alpha$ atoms of residues 3–29 of each monomer (cutoffs of 2.0 and 3.0 \AA). The $C\alpha$ RMSD time series were smoothed by taking running averages over a 25 ns window. Errors of the calculated rates of folding/unfolding were estimated with a Bayesian analysis.⁴⁰ To calculate the fraction of native contacts formed in simulation, we considered a contact between any two heavy atoms formed when the distance between the atoms was <5 \AA and the atoms were in different residues. A Markov state model of the metastable states on the free-energy surface was generated from a fit of the autocorrelation functions of 700 $C\alpha$ – $C\alpha$ contacts on time scales ranging from 0.25 to 100 μs .^{39,41,42} In this case, we considered a contact between $C\alpha$ atoms formed when the two atoms were within 8 \AA of each other. Detailed balance was enforced on the resulting rate matrix,⁴³ but the intrinsic 2-fold symmetry of the system was not exploited in the analysis.

RESULTS AND DISCUSSION

Simulations of the Top7-CFr Monomer. As a baseline for understanding the folding of the dimeric protein, we first examined the folding and stability of the isolated monomer. We performed four simulations of the Top7-CFr monomer at 400 K, for a total of ~ 0.8 ms of simulation. This temperature was chosen because test simulations indicated that 400 K is close to the Top7-CFr dimer's melting temperature. In the four equilibrium simulations, the Top7-CFr monomer folded eight times to conformations <3 \AA root-mean-square deviation (RMSD) from the experimentally determined structure (folding time: $76 \pm 27 \mu\text{s}$) but unfolded after a very short time (unfolding time: $0.09 \pm 0.03 \mu\text{s}$), indicating that the isolated monomer is not stable under these conditions (folding free energy of the monomer: $5.4 \pm 0.4 \text{ kcal mol}^{-1}$). About 40 transitions to a near-native, misfolded conformation were also observed (folding time: $15 \pm 2 \mu\text{s}$). This conformation features a non-native β sheet in which the first β strand (residues 2–9) and the second β strand (residues 34–40) directly interact to form a non-native β sheet. This state is characterized by a lower contact order and thus a lower chain entropy; it is thus unsurprising that it could form faster than the native state.⁴⁴ In the monomer simulations, this state was also very unstable and generally converted back into the unfolded state (unfolding time: $0.06 \pm 0.01 \mu\text{s}$), although in one instance we observed a transition from this misfolded state to the folded state. Previous simulations of the Top7-CFr monomer using a simplified, though transferable, potential⁴⁵ indicated that, at room

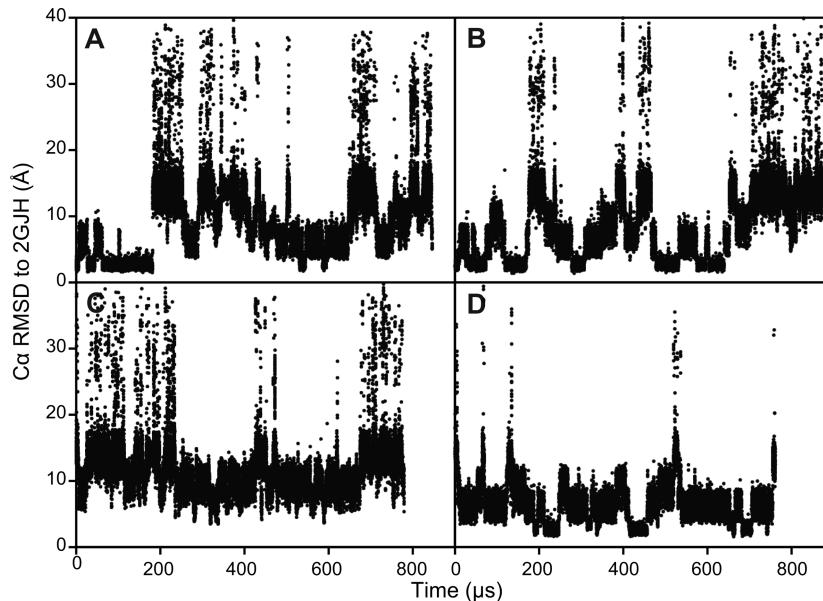


Figure 1. RMSD time series in simulations of the Top7-CFr dimer. RMSD of the Ca atoms (residues 2–50) with respect to the experimentally determined dimeric structure (PDB entry 2GJH). Simulations A and B were started from the experimentally determined structure, while simulations C and D were started from an extended conformation.

temperature, its N-terminal residues (residues 2–13) have a very strong tendency to form helical structures in the unfolded state, and that this region is helical in the early steps of folding, transforming into the native β strand only later in the folding pathway.⁴⁶ In our simulation conditions, in which the Top7-CFr monomer is unfolded most of the time, the N-terminal residues are slightly more likely to form sheets rather than helices (they have a 15% probability of forming a sheet versus a 10% probability of forming a helix); their probability of forming helices is much smaller than the 40% observed for residues 20–30, which form a helix in the native structure. The early steps of the folding pathway involve the formation of the β strand, including the contacts between the C- and N-terminal strands. While some helical structure forms early on, the full helix typically forms last. This is the reverse of the mechanism observed in Mohanty et al.,⁴⁶ in which the full helix formed early in the folding pathway, with the C- and N-terminal contacts forming in the last steps. The differences between these two sets of simulations are likely due to differences in the temperatures and force fields used.

Simulations of the Top7-CFr Dimer. To study the formation of the dimer, we performed four equilibrium simulations of systems containing two copies of the Top7-CFr monomer, corresponding to a monomer concentration of ~8 mM. In two of the simulations, the monomers started in extended conformations, and in the other two, they started in the native-state dimer structure determined experimentally by NMR spectroscopy.²⁷ In the four simulations, we observe a total of 12 transitions between unfolded structures (with Ca RMSDs >20 Å from the native dimeric state) and folded structures [with Ca RMSDs <2 Å from the native state (Figure 1)]. The lowest Ca RMSD observed in a simulation that started from an unfolded conformation was 1.2 Å. Note that the trajectory in Figure 1C should be considered nonfolding, as a Ca RMSD of <2 Å from the native state was not achieved in this simulation. We find approximately equal populations of folded and extended structures in simulation. Indeed, the relative stability of the native dimeric species with respect to the

unfolded state is $0.1 \pm 0.3 \text{ kcal mol}^{-1}$, suggesting that the simulation temperature is close to Top7-CFr's melting temperature. This observation is consistent with the protein's experimentally measured melting temperature of >373 K.²⁷ A direct two-state analysis to estimate the folded-to-unfolded ratio would be inappropriate for this system: The Ca RMSD traces strongly suggest that the folding free-energy surface is characterized by a number of metastable states with significant populations. Our analysis of a two-dimensional projection of the folding free-energy surface along the $Q_{\text{A}+\text{B}}$ (i.e., the fraction of intrasubunit native contacts for the two monomers, A and B) and $Q_{\text{interface}}$ (i.e., the fraction of intersubunit native contacts) reaction coordinates (Figure 2A) corroborates the presence of multiple states. Indeed, even in this simplified projection, multiple minima can be distinguished, separated by barriers of a few kcal mol^{-1} . In agreement with the results of the Top7-CFr isolated-monomer simulations, the folded monomer is unstable in simulation in the presence of an unfolded binding partner, with a free energy ~2.5 kcal mol^{-1} higher than that of the unfolded monomer. This value is ~3 kcal mol^{-1} lower than the folding free energy of 5.4 kcal mol^{-1} calculated from the isolated-monomer simulations, indicating that non-native intermonomer interactions provide a significant stabilizing contribution. These results suggest that Top7-CFr is an obligatory dimer,⁴⁷ at least under these simulation conditions. This finding is somewhat unexpected, as the dimer has a relatively small interface: The number of interface native contacts in the dimer is ~25% that of the contacts formed within each individual subunit.⁴⁷ Nevertheless, our finding is in agreement with experiments, which also suggested that the monomer is very unstable and that the folding transition and dimer formation are strongly coupled.²⁷

To provide a high-level characterization of the metastable states of the system and their rates of interconversion, we performed a dynamic-clustering analysis^{9,39} of the four dimer simulations (Figure 2B). Using this analysis, we attempted to find a set of macrostates that optimally describes the dynamics of the system on the microsecond time scale. We find that six

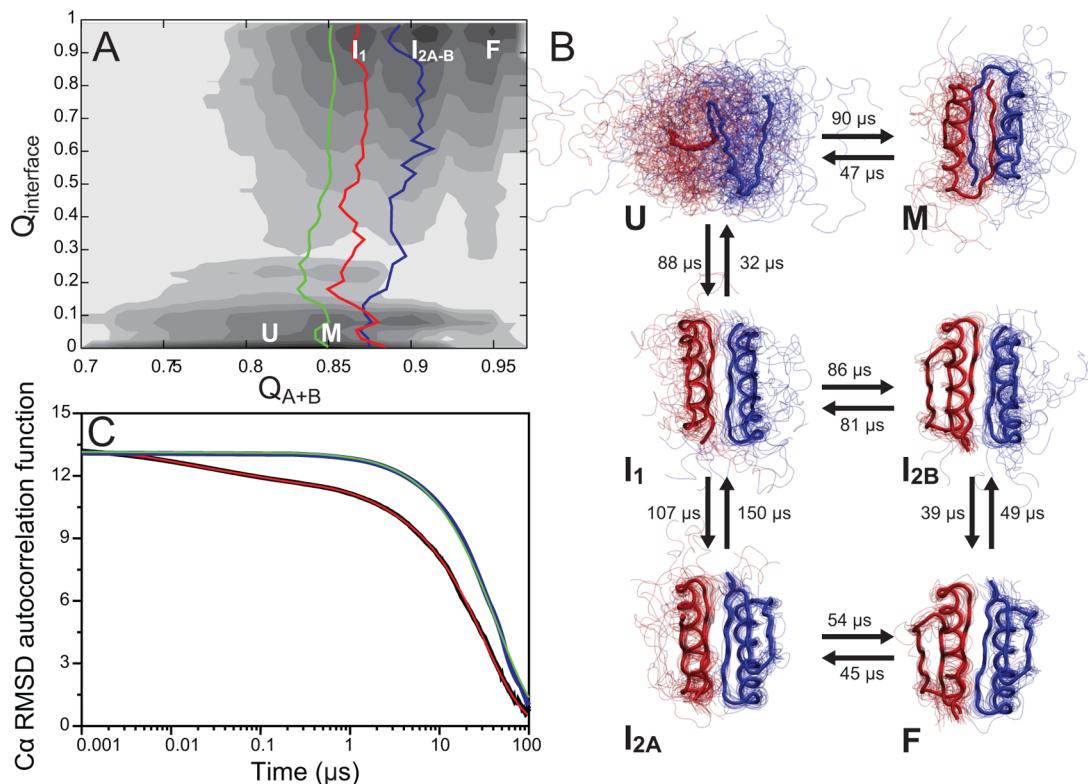


Figure 2. Folding free-energy surface and kinetics. (A) Two-dimensional projection of the folding free-energy surface as a function of the fraction of intrasubunit native contacts of monomers A and B (Q_{A+B}) and the fraction of intersubunit native contacts ($Q_{\text{interface}}$). Each shade of gray represents an energy increment of 1 kcal mol⁻¹. The position of the free-energy minima corresponding to the metastable states identified by a dynamic-clustering analysis are indicated with white letters. The average transition pathway from U to I is reported for each of the three dimerization mechanisms observed in simulation as blue (conformational selection), red (fly casting), and green (induced folding) lines. (B) Markov state model of the folding free-energy surface. The average structures of the metastable states of the system that are responsible for the multimicrosecond dynamics, as identified by a dynamic-clustering analysis, are reported (bold lines) together with 30 structures selected at random from the structural ensemble of each state (thin lines). In the interest of clarity, residues 51–57, which are disordered in the native state, are not displayed. The time scales for interconversion between the different states are also reported. A few transitions were also observed between the unfolded state (U) and the intermediates I_{2B} and I_{2A}. These transitions have a typical time scale of several hundred microseconds and are not reported for the sake of clarity. (C) (black) Autocorrelation function of the Ca RMSD from the experimental NMR structure 2GJH; (red) four-exponential fit with time scales of 10 ns, 78 ns, 1.9 μ s, and 32 μ s; (blue) autocorrelation function calculated from the Markov state model with a lag time of 250 ns; (green) single-exponential fit with a time scale of 45 μ s.

clusters is the smallest number that provides a good description of the dimer's long-time scale dynamics. Indeed, there is good agreement between the time scale of the decay of the Ca RMSD autocorrelation function obtained from the raw simulation data (32 μ s) and the longest relaxation time scale coming from the Markov state model based on the cluster analysis (45 μ s; Figure 2C). The agreement is particularly striking because RMSD information was not used to generate the Markov state model. As in a previous study,⁹ most of the states identified in the cluster analysis correspond to distinct minima on a low-dimensional free-energy surface (Figure 2A). This analysis reveals that the first metastable state encountered in the folding process is a symmetric dimer in which the native interface, encompassing the first β strand (residues 2–9) and the helix (residues 13–30), is fully formed (state I₁ in Figures 2B and 3). In some cases, the two strands that are in contact at the interface (residues 2–9) may form in reverse order, producing an off-pathway, misfolded dimeric state (state M in Figures 2B and 3). The smooth progress toward the native state is obstructed by residues 34–40, which in I₁, I_{2A}, and I_{2B} interact with N-terminal residues 2–9 to form the non-native β sheet already observed in the monomer simulations (Figures 2B and 3). In the dimer simulations, this non-native β sheet is

stable on a time scale of tens of microseconds and acts as a kinetic trap, thus resulting in a highly frustrated folding free-energy surface. This kinetic trap appears to be different from the kinetic traps observed in simulation studies of full Top7, which were generated by interactions with a patch of hydrophobic residues on the C-terminal helix.⁴⁸ While a frustrated energy landscape is rarely observed in naturally occurring proteins, it has been proposed that it may be more common in designed proteins, for which achieving a minimally frustrated folding free-energy landscape was not part of the fold optimization.⁴⁹ Despite the presence of a large number of slowly interconverting states, the long-time scale relaxation of the system, beyond the microsecond time scale, can still be adequately described by a single exponential with a relaxation time of \sim 30 μ s (Figure 2C). This may be ascribed partly to the similar relaxation time scales of 20–40 μ s observed for all macrostates and partly to the fact that the signal is largely dominated by the initial dimerization step. This observation is also in agreement with the single-exponential kinetics observed experimentally for Top7-CFr folding.²⁸

From a methodological perspective, the 2-fold symmetry of this enzyme allows us to perform a robust internal control of the precision of the simulation and analysis results. The

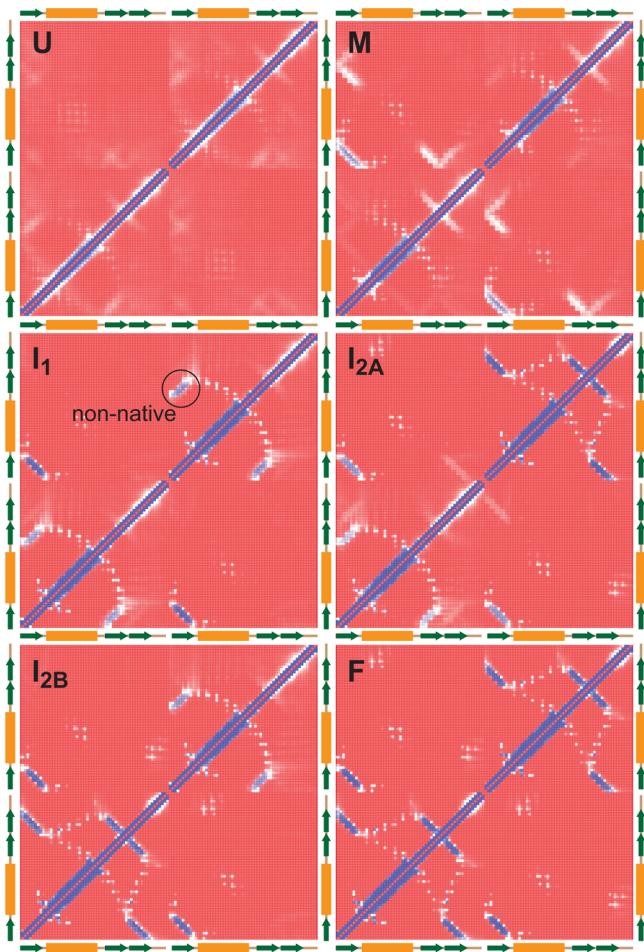


Figure 3. Time-averaged contact map of the metastable states. The time-averaged contact map reported for each of the metastable states identified with a kinetic-clustering analysis as described in Figure 2. Residues are defined to be in contact when the distance between the C α atoms is $<8\text{ \AA}$. Coloring goes from red (indicating that a contact never formed) to blue (a contact formed 100% of the time). The positions of the native-state secondary-structure elements are indicated by green arrows (sheets) and orange boxes (helices). The position of the non-native β sheet that is observed in states I₁, I_{2A}, and I_{2B} is highlighted on state I₁ with a circle.

dynamic-clustering method used in this work identifies two structurally identical macrostates (states I_{2A} and I_{2B} in Figure 2B; note that no symmetry information was provided in the analysis). In principle, the rates of interconversion of states I_{2A} and I_{2B} with the other substates of the system (I₁ or F) should be identical. The differences in the rates reported in Figure 2B can thus be taken as an estimate of the statistical accuracy of the simulations and the robustness of the analysis. In this respect, it is reassuring that most of the differences are within 10–30% of the rates, with the only exception being the I_{2A}-to-I₁ and I_{2B}-to-I₁ transitions, in which the rates differ by a little less than a factor of 2.

Characterization of the Initial Folding Steps. Two limiting mechanisms can be envisioned for the folding of dimeric species: a conformational-selection mechanism, in which two prefolded monomers encounter and bind, and a pure induced-folding mechanism, in which formation of the dimer interface initiates the folding process of the two subunits. A third mechanism has more recently been proposed, in which a fully or partially structured monomer acts as a template to

initiate the folding of the other subunit: the so-called “fly-casting” mechanism.⁵⁰ As a first step in characterizing the dimer-formation mechanism, we used a dual-cutoff method to separate the native-like states of the dimer from the unfolded states.^{30,37–39} This analysis results in the identification of 52 transition paths leading from the unfolded state to a stable, native dimer interface, generally corresponding to the transition between states U and I₁ and, in a few cases, transitions between U and I_{2A} or I_{2B}. A visual inspection of the transition paths reveals that, in the very early stages of dimerization, the two monomers interact nonspecifically through hydrophobic patches,⁵¹ usually exposed by partial or complete formation of the native helices. The formation of native contacts between the two N-terminal β strands at the dimer interface triggers folding. At this point, most of the interface’s native contacts are rapidly established, and folding proceeds until the formation of a non-native β strand locks the system in one of the intermediate dimeric states (I₁, I_{2A}, and I_{2B}).

Previous analysis of the folding mechanism of a number of dimeric proteins indicated that proteins with large hydrophobic interfaces are expected to be two-state folders, while proteins with small hydrophilic surfaces are expected to be three-state folders.⁴⁷ The dimer interface of Top7-CFr is relatively small (about 25% of the total native contacts) but rather hydrophobic (average hydrophobicity of 0.42).^{52,53} Top7-CFr is thus predicted not to have a strong preference toward either a two-step or a three-step folding mechanism.⁴⁷ To establish the type of folding mechanism followed by Top7-CFr, we quantified the amount of native structure present in each monomer at the moment of dimerization (here defined as the formation of the interface β strand composed by residues 3–9). To do this, we compared normalized integrals over the 52 transition paths⁶ of the RMSD to native of different parts of the protein. In particular, we compared the RMSD of interface residues 3–9 to that of the full set of the interfacial residues (residues 3–29) of each individual subunit (Figure 4). We find that the amount of native structure already present in each monomer at the moment of formation of the first native contacts at the interface varies. In 40% of the transitions, native-structure formation in both monomers follows the formation of the native interface, as in a pure induced-folding mechanism, while in over 50% of cases we observe a substantial amount of native structure in one of the two subunits, with the other being partially or completely unstructured. The latter scenario resembles a fly-casting folding mechanism,⁵⁰ in which the native interface presented by one subunit assists the folding of the other. In only 6% of the cases are both monomers substantially structured at the moment of the encounter, as in a conformational-selection mechanism. The average transition-path times (τ_{tp}) for the three dimerization mechanisms are similar, with the fly-casting mechanism ($\tau_{tp} = 0.7 \pm 0.3\text{ }\mu\text{s}$) possibly being slightly faster than either conformational selection ($1.3 \pm 1\text{ }\mu\text{s}$) or induced folding ($1.1 \pm 0.6\text{ }\mu\text{s}$). A larger statistical sample would be required, however, to firmly establish how different the transition-path times of the three mechanisms are. To further characterize the different folding mechanisms, we report the average transition path on the $Q_{\text{interface}}/Q_{\text{A+B}}$ free-energy surface (Figure 2A). This projection reinforces the notion that, in the induced-folding events, a smaller amount of structure is present in the monomers at the dimerization step, while progressively more preexisting structure is required for initiating dimerization in fly-casting and conformational-selection events. Conformational-selection

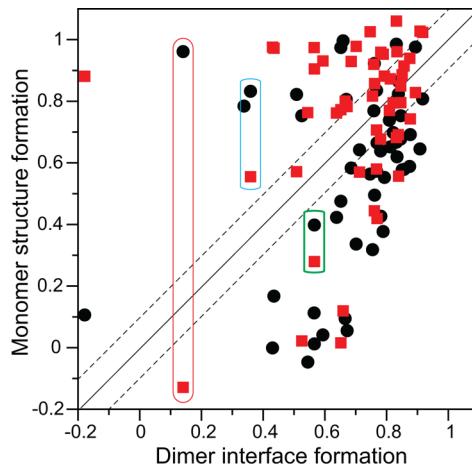


Figure 4. Order of structure formation during folding. Normalized integrals of the $\text{C}\alpha$ RMSD traces of the interface β strand (residues 3–9) plotted versus the normalized integrals of the $\text{C}\alpha$ RMSD traces of the interface residues (residues 3–29) of monomer A (black circles) or monomer B (red squares). Larger values of the integrals indicate earlier structure formation on the transition path. The points under the diagonal line correspond to transitions in which native structure at the interface formed before native structure in the monomer. Points near the diagonal (roughly defined as the region between the dashed lines) indicate transitions in which structure in the monomer and at the interface formed at the same time. Points above the diagonal represent transitions in which native structure in a monomer was present before the native interface formed. In a pure induced-folding transition, both the red and black points should be found on or below the diagonal (green-box example), while, in a pure conformational-selection transition, all points should be above the diagonal (blue-box example). In a fly-casting transition, only one of the two point colors should be above the diagonal (red-box example).

events are more likely to directly connect I_{2A} or I_{2B} with the unfolded state, but these events require a substantial preorganization and are thus characterized by a larger free-energy barrier, making them less frequent than induced-folding or fly-casting events. In summary, Top7-CFr appears to exhibit the full spectrum of possible folding mechanisms, with a preference for an intermediate mechanism, by which some native structure is formed in one of the two monomers in the early steps of dimerization.

CONCLUSIONS

We have reported the results of physics-based, fully atomistic equilibrium simulations of the folding of a dimeric protein, Top7-CFr. We find that Top7-CFr is characterized by a complex free-energy landscape, with a number of slowly interconverting kinetic traps generated by the formation of non-native β sheets. The simulations reveal a marked preference for the induced-folding and fly-casting mechanisms, although a few conformational-selection folding events are also observed. Top7-CFr is a very stable dimeric protein, and the simulations reported here were performed near the melting temperature and at relatively high monomer concentration, far from physiological conditions. It is possible that, at a lower temperature and lower concentration, one of the three mechanisms may predominate. More generally, the observation of alternative dimerization mechanisms in a single protein dimer suggests that the contribution of these mechanisms to overall folding may depend on external conditions, thus providing a means to control the assembly of protein

complexes in living organisms. A common feature of the early stages of folding is the interaction of the two subunits through non-native patches of hydrophobic residues. The simulation results indicate that these interactions help stabilize nascent native structure, and that the monomers' proximity facilitates the formation of intersubunit native contacts at the dimer interface. While this effect may have been somewhat amplified by the simulation conditions (i.e., a high monomer concentration and the use of a shifted cutoff³⁴), we speculate that it may be a more general feature of the folding of obligatory dimers with hydrophobic dimer interfaces.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Stefano.Piana-Agostinetti@DEShawResearch.com (S.P.); David.Shaw@DEShawResearch.com (D.E.S.). Phone: (212) 403-8165 (S.P.); (212) 478-0260 (D.E.S.). Fax: (646) 873-2165 (S.P.); (212) 845-1286 (D.E.S.).

Present Address

[§]Structural Biology and NMR Laboratory (SBiNLab), Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark.

Notes

The authors declare the following competing financial interest(s): This study was conducted and funded internally by D. E. Shaw Research, of which D.E.S. is the sole beneficial owner and Chief Scientist, and with which all authors are affiliated.

ACKNOWLEDGMENTS

We thank John Klepeis for a critical reading of the manuscript and Berkman Frank for editorial assistance.

REFERENCES

- (1) Best, R. B. Atomistic Molecular Simulations of Protein Folding. *Curr. Opin. Struct. Biol.* **2012**, 22 (1), 52–61.
- (2) Prigozhin, M. B.; Gruebele, M. Microsecond Folding Experiments and Simulations: A Match is Made. *Phys. Chem. Chem. Phys.* **2013**, 15 (10), 3372–3388.
- (3) Freddolino, P. L.; Schulten, K. Common Structural Transitions in Explicit-Solvent Simulations of Villin Headpiece Folding. *Biophys. J.* **2009**, 97 (8), 2338–2347.
- (4) Beauchamp, K. A.; Ensign, D. L.; Das, R.; Pande, V. S. Quantitative Comparison of Villin Headpiece Subdomain Simulations and Triplet-Triplet Energy Transfer Experiments. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, 108 (31), 12734–12739.
- (5) Bowman, G. R.; Voelz, V. A.; Pande, V. S. Atomistic Folding Simulations of the Five-Helix Bundle Protein $\lambda(6-85)$. *J. Am. Chem. Soc.* **2011**, 133 (4), 664–667.
- (6) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, 334 (6055), 517–520.
- (7) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein Folding Kinetics and Thermodynamics from Atomistic Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109 (44), 17845–17850.
- (8) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; et al. Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc.* **2012**, 134 (30), 12565–12577.
- (9) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-Level Description of Ubiquitin Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, 110 (15), 5915–5920.
- (10) Chong, L. T.; Snow, C. D.; Rhee, Y. M.; Pande, V. S. Dimerization of the p53 Oligomerization Domain: Identification of a

- Folding Nucleus by Molecular Dynamics Simulations. *J. Mol. Biol.* **2005**, *345* (4), 869–878.
- (11) Bonomi, M.; Barducci, A.; Gervasio, F. L.; Parrinello, M. Multiple Routes and Milestones in the Folding of HIV-1 Protease Monomer. *PLoS One* **2010**, *5* (10), e13208.
- (12) Higo, J.; Nishimura, Y.; Nakamura, H. A Free-Energy Landscape for Coupled Folding and Binding of an Intrinsically Disordered Protein in Explicit Solvent from Detailed All-Atom Computations. *J. Am. Chem. Soc.* **2011**, *133* (27), 10448–10458.
- (13) Bhattacharjee, A.; Wallin, S. Coupled Folding-Binding in a Hydrophobic/Polar Protein Model: Impact of Synergistic Folding and Disordered Flanks. *Biophys. J.* **2012**, *102* (3), 569–578.
- (14) Ding, F.; Dokholyan, N. V.; Buldyrev, S. V.; Stanley, H. E.; Shakhnovich, E. I. Molecular Dynamics Simulation of the SH3 Domain Aggregation Suggests a Generic Amyloidogenesis Mechanism. *J. Mol. Biol.* **2002**, *324* (4), 851–857.
- (15) Levy, Y.; Caflisch, A.; Onuchic, J. N.; Wolynes, P. G. The Folding and Dimerization of HIV-1 Protease: Evidence for a Stable Monomer from Simulations. *J. Mol. Biol.* **2004**, *340* (1), 67–79.
- (16) Piana, S.; Taylor, Z.; Rothlisberger, U. Folding Pathways for Initiator and Effector Procapases from Computer Simulations. *Proteins* **2005**, *59* (4), 765–772.
- (17) Levy, Y.; Onuchic, J. N. Mechanisms of Protein Assembly: Lessons from Minimalist Models. *Acc. Chem. Res.* **2006**, *39* (2), 135–142.
- (18) Simler, B. R.; Levy, Y.; Onuchic, J. N.; Matthews, C. R. The Folding Energy Landscape of the Dimerization Domain of Escherichia coli Trp Repressor: A Joint Experimental and Theoretical Investigation. *J. Mol. Biol.* **2006**, *363* (1), 262–278.
- (19) Patel, B.; Finke, J. M. Folding and Unfolding of Gamma TIM Monomers and Dimers. *Biophys. J.* **2007**, *93* (7), 2457–2471.
- (20) Turjanski, A. G.; Gutkind, J. S.; Best, R. B.; Hummer, G. Binding-Induced Folding of a Natively Unstructured Transcription Factor. *PLoS Comput. Biol.* **2008**, *4* (4), e1000060.
- (21) Wang, W.; Xu, W. X.; Levy, Y.; Trizac, E.; Wolynes, P. G. Confinement Effects on the Kinetics and Thermodynamics of Protein Dimerization. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (14), 5517–5522.
- (22) Xia, F.; Thirumalai, D.; Gräter, F. Minimum Energy Compact Structures in Force-Quench Polyubiquitin Folding are Domain Swapped. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (17), 6963–6968.
- (23) Zheng, W.; Schafer, N. P.; Davtyan, A.; Papoian, G. A.; Wolynes, P. G. Predictive Energy Landscapes for Protein-Protein Association. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (47), 19244–19249.
- (24) Wright, P. E.; Dyson, H. J. Linking Folding and Binding. *Curr. Opin. Struct. Biol.* **2009**, *19* (1), 31–38.
- (25) Kieffaber, T.; Bachmann, A.; Jensen, K. S. Dynamics and Mechanisms of Coupled Protein Folding and Binding Reactions. *Curr. Opin. Struct. Biol.* **2011**, *22*, 1–9.
- (26) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302* (5649), 1364–1368.
- (27) Dantas, G.; Watters, A. L.; Lunde, B. M.; Eletr, Z. M.; Isern, N. G.; Roseman, T.; Lipfert, J.; Doniach, S.; Tompa, M.; Kuhlman, B.; et al. Mis-Translation of a Computationally Designed Protein Yields an Exceptionally Stable Homodimer: Implications for Protein Engineering and Evolution. *J. Mol. Biol.* **2006**, *362* (5), 1004–1024.
- (28) Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. The Highly Cooperative Folding of Small Naturally Occurring Proteins Is Likely the Result of Natural Selection. *Cell* **2007**, *128* (3), 613–624.
- (29) Krishna, N. M.; Englander, S. W. A Unified Mechanism for Protein Folding: Predetermined Pathways with Optional Errors. *Protein Sci.* **2007**, *16* (3), 449–464.
- (30) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47–L49.
- (31) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (32) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.
- (33) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J., et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proceedings of the Conference on High Performance Computing, Networking, Storage, and Analysis (SC09)*; ACM: New York, 2009.
- (34) Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Dror, R. O.; Shaw, D. E. Evaluating the Effects of Cutoffs and Treatment of Long-Range Electrostatics in Protein Folding Simulations. *PLoS One* **2012**, *7* (6), e39918.
- (35) Lippert, R. A.; Bowers, K. J.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Shaw, D. E. A Common, Avoidable Source of Error in Molecular Dynamics Integrators. *J. Chem. Phys.* **2007**, *120* (4), 046101.
- (36) Steinbach, P. J.; Brooks, B. R. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comput. Chem.* **1994**, *15*, 667–683.
- (37) Northrup, S. H.; Hynes, J. T. The Stable States Picture of Chemical Reactions. I. Formulation for Rate Constants and Initial Condition Effects. *J. Chem. Phys.* **1980**, *73* (6), 2700–2714.
- (38) Buchete, N. V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112* (19), 6057–6069.
- (39) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330* (6002), 341–346.
- (40) Ensign, D. L.; Pande, V. S. Bayesian Single-Exponential Kinetics in Single-Molecule Experiments and Simulations. *J. Phys. Chem. B* **2009**, *113* (36), 12410–12423.
- (41) Hummer, G. Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations. *New J. Phys.* **2005**, *7*(34).
- (42) Best, R. B.; Hummer, G. Reaction Coordinates and Rates from Transition Paths. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6732–6737.
- (43) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105.
- (44) Plaxco, K. W.; Simons, K. T.; Baker, D. Contact Order, Transition State Placement and the Refolding Rates of Single Domain Proteins. *J. Mol. Biol.* **1998**, *277* (4), 985–994.
- (45) Irbäck, A.; Mohanty, S. Folding Thermodynamics of Peptides. *Biophys. J.* **2005**, *88* (3), 1560–1569.
- (46) Mohanty, S.; Meinke, J. H.; Zimmermann, O.; Hansmann, U. H. Simulation of Top7-CFR: A Transient Helix Extension Guides Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (23), 8004–8007.
- (47) Levy, Y.; Wolynes, P. G.; Onuchic, J. N. Protein Topology Determines Binding Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (2), 511–516.
- (48) Zhang, Z.; Chan, H. S. Competition Between Native Topology and Nonnative Interactions in Simple and Complex Folding Kinetics of Natural and Designated Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (7), 2920–2925.
- (49) Scalley-Kim, M.; Baker, D. 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.* **2004**, *338* (3), 573–583.
- (50) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. Speeding Molecular Recognition by Using the Folding Tunnel: The Fly-Casting Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (16), 8868–8873.
- (51) Bachmann, A.; Wildemann, D.; Praetorius, F.; Fischer, G.; Kieffaber, T. Mapping Backbone and Side-Chain Interactions in the Transition State of a Coupled Protein Folding and Binding Reaction. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (10), 3952–3957.

(52) Pacios, L. F. Distinct Molecular Surfaces and Hydrophobicity of Amino Acid Residues in Proteins. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (5), 1427–1435.

(53) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. A Survey of Flexible Protein Binding Mechanisms and Their Transition States Using Native Topology Based Energy Landscapes. *J. Mol. Biol.* **2005**, *346* (4), 1121–1145.