J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

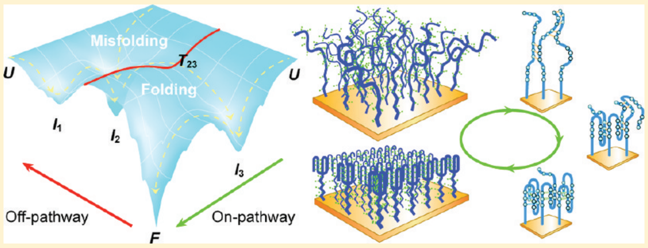
Understanding Thermodynamic Competitivity between Biopolymer Folding and Misfolding under Large-Scale Intermolecular Interactions

Youdong Mao*,† and Jinzhong Zhang‡

†Dana-Farber Cancer Institute, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02215, United States

‡Department of Physics, Northeastern University, Boston, Massachusetts 02115, United States

Ⓢ *Supporting Information*



ABSTRACT: Cooperativity is a hallmark of spontaneous biopolymer folding. The presence of intermolecular interactions could create off-pathway misfolding structures and suppress folding cooperativity. This raises the hypothesis that thermodynamic competitivity between off-pathway misfolding and on-pathway folding may intervene with cooperativity and govern biopolymer folding dynamics under conditions permitting large-scale intermolecular interactions. Here we report direct imaging and theoretical modeling of thermodynamic competitivity between biopolymer folding and misfolding under such conditions, using a two-dimensional array of proton-fueled DNA molecular motors packed at the maximal density as a model system. Time-resolved liquid-phase atomic force microscopy with enhanced phase contrast revealed that the misfolding and folding intermediates transiently self-organize into spatiotemporal patterns on the nanoscale in thermodynamic states far away from equilibrium as a result of thermodynamic competitivity. Computer simulations using a novel cellular-automaton network model provide quantitative insights into how large-scale intermolecular interactions correlate the structural dynamics of individual biomolecules together at the systems level.

INTRODUCTION

Since Anfinsen's pioneering work in 1961, biopolymer folding has elicited tremendous interest and extensive studies in many scientific communities.[1] It was eventually recognized as one of the most significant problems in physical chemistry and molecular biophysics.[2−5] Almost all kinds of biopolymer-based molecular machines, produced by either natural evolution or artificial engineering, are associated with or driven by certain degrees of folding behaviors.[6−22] The past decade has seen rapidly growing, widespread interest in constructing nanostructures,[23] nanomachines,[24] and stimuli-responsive nanomaterials[25−27] from biopolymers. In many of these systems, more than millions of biopolymer molecules are often arranged in a particularly crowded environment on the nanoscale that entails large-scale, complex, strong intermolecular interactions. Herein the term "large-scale" is defined as the property of very high copy numbers (e.g., 10^6) with respect to a biopolymer system, in which complex intermolecular interactions connect all of the molecules in the system together into a single inseparable "interaction network". In living cells, newly synthesized biopolymers of high copy numbers also have to fold in an extremely crowded environment. At present, little is known about how biopolymers fold under these complex conditions. Such a quest for understanding the folding mechanism of millions of strongly interacting biopolymers represents the highest degree of folding complexity studied to date, therefore presenting a significant challenge.

Biopolymer folding in the bulk of a dilute solution or under single-molecule conditions is largely governed by cooperativity in conformational changes.[6,7] The presence of large-scale intermolecular interactions could create multiple misfolding structures or make the folding energy landscape more frustrated.[8−11] This raises the hypothesis that thermodynamic competition between off-pathway misfolding and on-pathway folding may strongly suppress the folding cooperativity and govern the folding dynamics under conditions permitting large-scale intermolecular interactions.[8] To what extent and in what way could the folding cooperativity be suppressed by thermodynamic competitivity arising from off-pathway misfolding? Addressing this problem in a quantitative manner is essential for comprehending the relationship between off-pathway misfolding and on-pathway folding and, more importantly, the mechanistic underpinning. In this work, we attempted to address this issue quantitatively by direct imaging and theoretical modeling of thermodynamic competitivity between biopolymer folding and misfolding in the presence of large-scale intermolecular interactions, using a model system consisting of a two-dimensional array of a typical proton-fueled DNA molecular motor[19−21] packed at the maximal density with respect to the folded state of the molecule. We achieved a quantitative agreement between experiments and theoretical modeling in characterizing the misfolding competitivity. Our study suggests a novel competition−cooperation−condensation

Received: October 11, 2011
Published: November 29, 2011

dx.doi.org/10.1021/ja209534c | *J. Am. Chem.Soc.* 2012, 134, 631−639

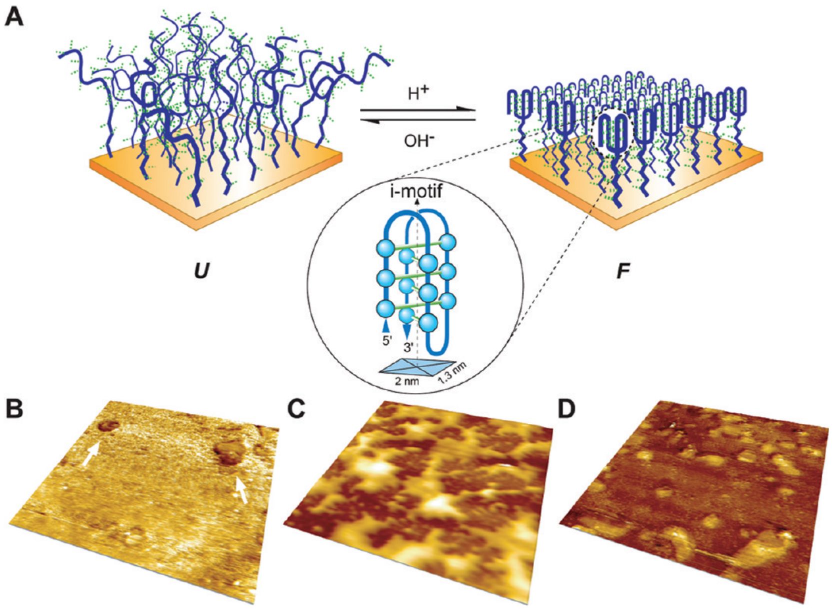


Figure 1. Direct AFM observations of the folding process of the DNA molecular motor array. (A) Schematic of the model system constructed from a DNA molecular motor array. The molecules fold into the *i*-motif structure (*F*) at pH 4.5 but unfold into single strands (*U*) at pH 8.5 at room temperature. The inset shows the native *i*-motif structure as a simplified stick-and-bead model; only cytosine residues are shown as blue beads. (B) At the beginning of folding (pH 7), some folding nuclei with lower heights and sizes of ~0.5 μm (highlighted by the arrows) show up. The nucleation behavior is a signature of the first-order phase transition. (C) At the middle stage (pH 6), the folded areas begin to coalesce and get connected. (D) At the late stage (pH 5), the folded areas coalesce into a connected region, setting off the unfolded areas to appear as extruding islands. The AFM topographic images shown in (B−D) are 5 μm × 5 μm and shown as three-dimensional perspective views.

mechanism for biopolymer folding under large-scale intermolecular interactions.

■ RESULTS AND DISCUSSION

Experimental Observations. The model system was prepared by self-assembly of synthetic thiol-terminated DNA molecular motors[19−21] onto a Au(111) surface via covalent gold−sulfur bonds at the maximal area density of the folded structure[21] (see Methods). Each DNA chain can fold from a single-stranded coil (denoted as *U*) into a four-stranded structure called an *i*-motif (denoted as *F*) upon triggering by protonation (Figure 1A).[28] In the structure of the *i*-motif, four tracts of cytidine form two base-paired parallel-stranded duplexes having their hemiprotonated cytosine−cytosine base pairs ($C:C^+$) fully intercalated at acid pH.[29,30] Because of the three hairpin loops in the folded structure (Figure 1A inset), its folding behavior in the absence of intermolecular interactions already bears higher topological complexity and more resemblance to protein folding than does that of hairpin DNA [Figure S1 in the Supporting Information (SI)].[12−14] Several variants of the model system have been widely used to fabricate nanomechanical devices,[20,21] switchable nanochannels,[22] and stimuli-responsive materials.[27]

The direct, real-time, in situ observation of biopolymer folding in the model system was achieved by tapping-mode atomic force microscopy (AFM) in the liquid phase.[31] At the beginning of folding, a few folding nuclei[15] with lower topographic heights were clearly observed (Figure 1B). The nucleation was followed by the domain growth of folded areas (Figures 1−3). In the *F* domain, where the DNA molecules were well-folded into native structures, a rhombic lattice of molecular footprints was identified by high-resolution AFM imaging (Figure 2F). The observed lattice constant of 26 ± 3 Å corresponds to the upper limit of the *i*-motif's area density. The measured molecular footprint size of 2.0 nm × 1.3 nm is highly consistent with the crystal structure of *i*-motif.[30] These results were further confirmed by a 45° rotation of the AFM scanning direction (Figure S2 in the SI). The AFM observations provide direct evidence that the DNA molecules can fold into the native *i*-motif structure at the molecular crowding limit. The results suggest that the overall folding transition resembles the behavior of crystallization.

Because AFM at high spatial resolution suffers from low time resolution, we slowed the folding process to match its time scale with the time resolution of AFM by providing insufficient protons, allowing the spatiotemporal behavior of folding intermediate states to be captured (see Methods). A typical topographic image obtained by this approach (Figure 2A) exhibits a self-organized pattern with ample nanoscale topographic features, such as flowerlike nanoclusters, wirelike nanowalls, and canal-like nanotrenches (Figure 2C−E and Figure S3 in the SI). The simultaneously obtained phase image[32] (Figure 2B) shows an enhanced image contrast between the intermediate-folding and well-folded domains; the intermediate-folding molecules in the brighter areas are softer in stiffness and higher in topography than the well-folded molecules in the darker areas. This result reflects the fact that the overall stiffness of a biopolymer increases dramatically as it folds into a compact structure. Taking advantage of the enhanced phase contrast, we were able to track unambiguously in real time the mesoscopic coarsening of the *F* domains, which shows a strong anisotropy in growth rate (Figure 3A and Figures S4 and S5 in the SI).

Misfolding and Folding Intermediates. To facilitate the analysis of folding pathways, we define an on-pathway folding intermediate as a metastable state having no intermolecular hydrogen bonds and an off-pathway misfolding intermediate as

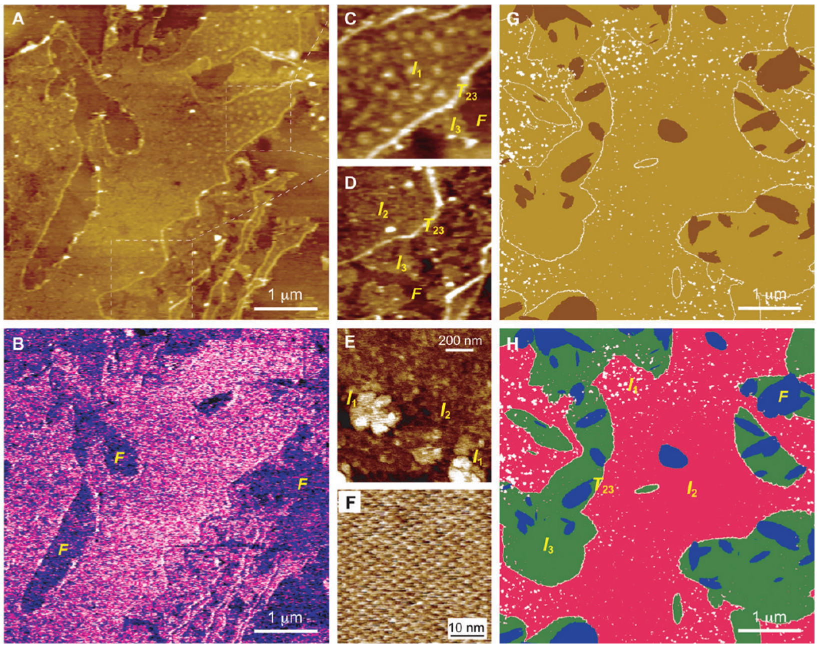


Figure 2. Nanoscale transient self-organization of the misfolding and folding intermediates arising from thermodynamic competitivity. (A) High-resolution AFM topographic image showing the coexistence of I_1, I_2, T_{23}, I_3, and F and their self-organized pattern with several distinctive topological features. (B) Simultaneously obtained phase image of the same area as shown in (A). The T_{23} interface between the I_2 and I_3 domains appears softer than the three intermediates (I_1, I_2, I_3) and F, while the three intermediates are all softer than F. It should be noted that AFM phase imaging is sensitive to stiffness changes but not topographic changes. (C, D) Enlargements of the indicated areas in (A) that highlight the coexistence of folding intermediates. (E) High-resolution AFM topographic image showing in detail the flowerlike I_1 nanoclusters surrounded by the I_2 domains. (F) High-resolution AFM topographic image observed in an F domain, showing the rhombic lattice of the *i*-motif's footprint. (G) Self-organized pattern of the misfolding and folding intermediates computationally simulated by using the cellular-automaton network model, with coloration resembling that in the AFM topographic image. The simulation reproduces the topological features observed in (A). (H) The same simulated pattern as (G) with a different coloration highlighting the areas occupied by different folding states. The qualitative agreement in topological features between (A) and (G) indicates that the flowerlike nanoclusters, wirelike nanowalls, and canal-like nanotrenches in (A–E) correspond to I_1, T_{23}, and F, respectively; the regions surrounding the I_1 nanoclusters but surrounded by the T_{23} nanowalls correspond to I_2; the regions outside of the T_{23} nanowalls and directly interfacing with the F domains correspond to I_3. T_{23} emerges at the interface between the I_2 and I_3 domains and extends to the interface between the I_2 and F domains.

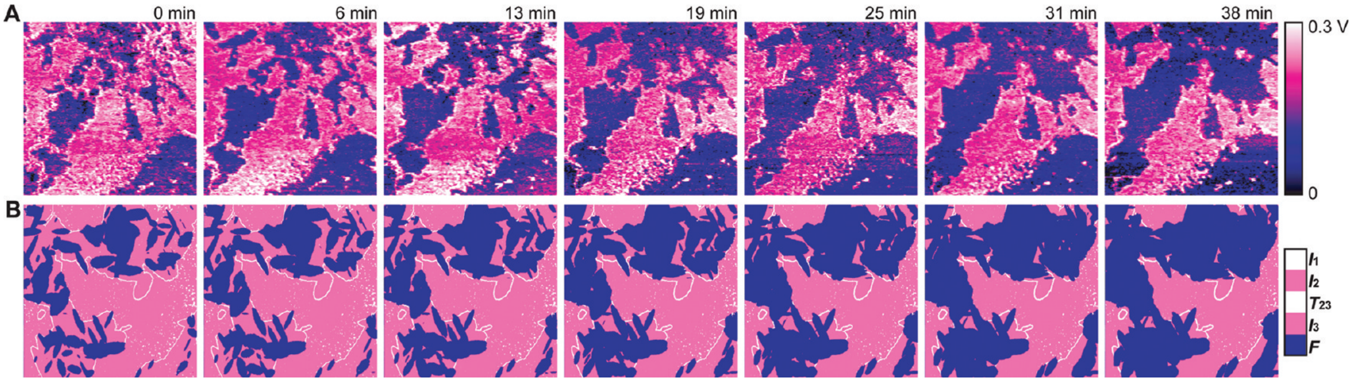


Figure 3. Real-time phase-contrast AFM imaging of the folding process. (A) Coarsening of F domains (blue areas) observed in real time by AFM phase imaging with enhanced contrast. Image size: 8 μm × 8 μm. (B) Computer simulation of the mesoscopic coarsening of F domains, which qualitatively reproduces the strongly anisotropic growth of the F domains observed in (A). In both (A) and (B), the F domains show a palmlike topology and several straight edges arising from the growth anisotropy. Image size: 5 μm × 5 μm. The geometric discrepancy between (A) and (B) is attributed to the stochastic nature of folding dynamics.

a metastable state having nearly no intramolecular hydrogen bonds. On the basis of conformational analysis of the possible interactions between neighboring molecules, we hypothesize two off-pathway intermediates and one on-pathway intermediate (Figure 4A). The first off-pathway intermediate, I_1, represents the unfolded conformations of DNA chains that form metastable aggregations by intermolecular hydrogen bonds between neighbors. The second off-pathway intermediate, I_2, represents the semifolded conformations of DNA chains looping one or two times and forming aggregated molecular networks by intermolecular hydrogen bonds. The on-pathway intermediate, I_3, represents the semifolded conformations of DNA chains looping one or two times by non-native intramolecular hydrogen bonds and without intermolecular

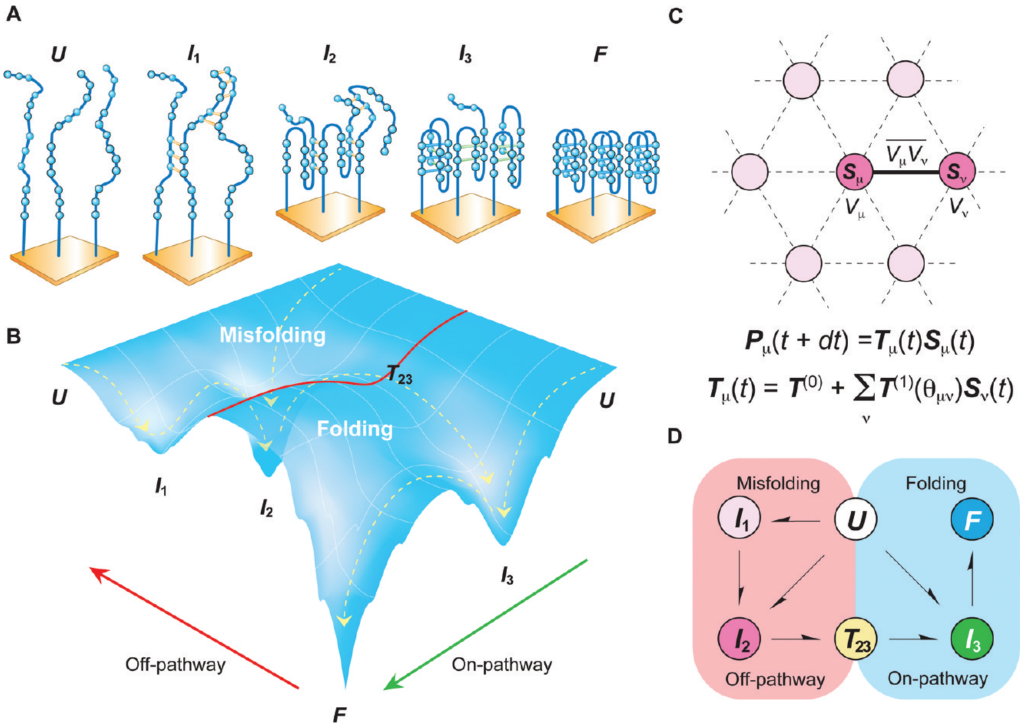


Figure 4. Misfolding and folding pathways and energy landscape. (A) Schematics of the typical conformations of U, I_1, I_2, I_3, and F, displayed as stick-and-bead models. For clarity, only three molecules are shown for each case, and only cytosine residues are shown as blue beads. Intermolecular, intramolecular non-native, and native C:C$^+$ base pairs are shown as orange, green, and blue sticks, respectively. (B) Simplified three-dimensional perspective view of the misfolding and folding energy landscape of individual molecules in the model system. The red line separates the misfolding regime governed by intermolecular interactions and the folding regime governed by intramolecular interactions. The yellow dashed lines highlight the significant folding pathways. (C) Geometry of the cellular-automaton network (CAN) model and the core mathematical formulas. (D) Multiplex misfolding (left frame) and folding (right frame) pathways resolved from experiments and simulations.

hydrogen bonds. The thermodynamic stability of I_1 is expected to be lower than those of the other two intermediates because of the relatively higher conformational entropy. I_2 and I_3 are both stabilized by an increased degree of molecular crowding relative to I_1. When a molecular conformation changes from I_2 to I_3, it has to break most of its intermolecular hydrogen bonds. We hypothesize that this should lead to an unstable transition state, T_{23}, that approximates an unfolded coil in conformation and emerges at the interface between the domains of I_2 and I_3. Figure 4B schematically depicts the energetic relationships among U, I_1, I_2, T_{23}, I_3, and F.

Cellular-Automaton Network Model. To gain insight into the experimental observations, we performed a computational modeling study. Because the conformational space of the model system comprising more than millions of biomolecules is at least 10^6 times larger in dimensions than that of a single biomolecule, any computer simulation based on the existing Gō-like[2,3,8,9] or all-atom models[17] of folding dynamics would largely exceed the capability of modern supercomputers. Instead, we established a novel two-dimensional cellular-automaton[33] network (CAN) model in which each biomolecule is a vertex, V_m, and each intermolecular interaction between two neighbors V_μ and V_ν is an undirected edge, $\overline{V_\mu V_\nu}$ (Figure 4C). The state of each vertex, $\mathbf{S}_m(t)$, evolves among six basis vectors $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4, \mathbf{s}_5, \mathbf{s}_6\}$ in a six-dimensional state space, which represent the folding states $\{U, I_1, I_2, T_{23}, I_3, F\}$, respectively:

$$\mathbf{s}_1 = \begin{bmatrix} 1\\0\\0\\0\\0\\0 \end{bmatrix} \rightarrow U,\quad \mathbf{s}_2 = \begin{bmatrix} 0\\1\\0\\0\\0\\0 \end{bmatrix} \rightarrow I_1,\quad \mathbf{s}_3 = \begin{bmatrix} 0\\0\\1\\0\\0\\0 \end{bmatrix} \rightarrow I_2,\quad \mathbf{s}_4 = \begin{bmatrix} 0\\0\\0\\1\\0\\0 \end{bmatrix} \rightarrow T_{23},\quad \mathbf{s}_5 = \begin{bmatrix} 0\\0\\0\\0\\1\\0 \end{bmatrix} \rightarrow I_3,\quad \mathbf{s}_6 = \begin{bmatrix} 0\\0\\0\\0\\0\\1 \end{bmatrix} \rightarrow F$$

The stochastic dynamics of the state transitions in V_m from $\mathbf{S}_m(t)$ to $\mathbf{S}_m(t + \mathrm{d}t)$ are defined by a six-dimensional probability vector $\mathbf{P}_m(t + \mathrm{d}t)$, whose components $P_{m,i}(t + \mathrm{d}t)$ represent the probabilities that $\mathbf{S}_m(t)$ changes to $\mathbf{S}_m(t + \mathrm{d}t) = \mathbf{s}_i$, where $i \in \{1, 2, ..., 6\}$. The dependence of $\mathbf{P}_m(t + \mathrm{d}t)$ on $\mathbf{S}_m(t)$ is given by

$$\mathbf{P}_m(t + \mathrm{d}t) = \mathbf{T}_m(t)\mathbf{S}_m(t) \tag{1}$$

Here $\mathbf{T}_m(t)$ is a 6×6 transition matrix whose components can be expanded in the first-order approximation (see the next section for the derivation) as

$$T_{\mu,mn}(t) = T^{(0)}_{mn} + \sum_{k} \sum_{\nu} T^{(1)}_{mnk}(\theta_{\mu\nu}) S_{\nu,k}(t) \tag{2}$$

where $T^{(0)}_{mn}$ is the probability that $\mathbf{S}_\mu(t)$ transitions from $\mathbf{s}_n$ to $\mathbf{s}_m$ as a result of the intramolecular interactions; $T^{(1)}_{mnk}(\theta_{\mu\nu})$ is the $\mathbf{s}_n \rightarrow \mathbf{s}_m$ transition probability contributed by the intermolecular interactions between V_μ in state $\mathbf{s}_n$ and its neighbor V_ν in state $\mathbf{s}_k$; $\theta_{\mu\nu}$ is the azimuth of edge $\overline{V_\mu V_\nu}$; and m, n, and $k \in \{1, 2, ..., 6\}$. Under a similar treatment in the first-order approximation, the total free energy barrier for the $\mathbf{s}_n \rightarrow \mathbf{s}_m$ transition of $\mathbf{S}_\mu(t)$ can be expressed as

$$\Delta G_{\mu,mn}(t) = \Delta G^{(0)}_{mn} + \sum_{k} \sum_{\nu} \Delta G^{(1)}_{mnk}(\theta_{\mu\nu}) S_{\nu,k}(t) \tag{3}$$

where $\Delta G^{(0)}_{mn}$ is the portion of the free energy barrier contributed by the intramolecular interactions; $\Delta G^{(1)}_{mnk}(\theta_{\mu\nu})$ is the portion of the free energy barrier contributed by the intermolecular interactions between V_μ in state $\mathbf{s}_n$ and its neighbor V_ν in state $\mathbf{s}_k$; $\theta_{\mu\nu}$ is the azimuth of edge $\overline{V_\mu V_\nu}$; and m, n, and $k \in \{1, 2, ..., 6\}$. Here the summation over ν is taken through all the vertices that are interacting neighbors of V_μ. This expression reflects the fact that the folding energy landscape of individual molecules strongly varies with time and is highly dependent on changes in the intermolecular interactions and state changes of their neighbors. For instance, when all of the interacting neighbors of V_μ fold into F, the folding energy landscape of V_μ would degenerate to the three states for the case of single-molecule conditions (Figure S1 in the SI).

Previous studies involving energy landscape theory have shown that biopolymer folding is best described as a Kramers-like diffusive process[34] across the energy landscape and that the folding rate constant is dependent on several aspects of the folding landscape. For a one-dimensional free energy surface with harmonic wells, the folding rate constant is related to the shape of the energy landscape, $\omega_{\mathrm{D}}\omega^*/2\pi$, the height of the energy barrier, ΔG^*, and the diffusion coefficient, D_0, by the expression $k_{\mathrm{f}} = [(\omega_{\mathrm{D}}\omega^* D_0)/(2\pi RT)]\ \exp(-\Delta G^*/RT)$, where ω_{D} and ω^* describe the curvatures of the energy landscape in the denatured state and the transition state, respectively, R is the molar gas constant, and T is the absolute temperature. As k_{f} dt represents the folding probability across the energy barrier ΔG^* in the time interval dt, from eqs 2 and 3 one may obtain

$$T^{(0)}_{mn} + \sum_{k} \sum_{\nu} T^{(1)}_{mnk}(\theta_{\mu\nu}) S_{\nu,k}(t) \propto \frac{\omega_{\mathrm{D}}\omega^*_{mn} D_0}{2\pi RT} \exp\left(\frac{-\Delta G^{(0)}_{mn} - \sum_k \sum_\nu \Delta G^{(1)}_{mnk}(\theta_{\mu\nu}) S_{\nu,k}(t)}{RT}\right) \tag{4}$$

where ω^*_{mn} ($m \neq n$) is the curvature in the transition state between $\mathbf{s}_n$ and $\mathbf{s}_m$ in the energy landscape. Equations 2−4 enable computer simulations of the entire process of the folding transition in the model system that may be directly used to interpret the real-time AFM observations (see Methods).

The CAN model extends energy landscape theory to the regime of strong intermolecular interactions in complex molecular folding systems. It provides an analytic framework for describing quantitatively how the complex intermolecular interactions modulate the energy landscapes of individual molecules and therefore strongly couple their folding dynamics together at the systems level. It should be noted that this model can readily be generalized to N-state folding without mathematical modifications of eqs 2−4 simply by expanding the basis vectors to $\mathbf{R}^N$ space, where $N > 2$ is an integer representing the total number of stable and metastable folding states in a folding energy landscape. In the computer simulations of the full folding process, three types of intermolecular interactions, namely, hydrogen-bonding, excluded-volume,[35−38] and electrostatic interactions,[39−45] were taken into account in a coarse-grained fashion (for details, see Methods, Table 1, and note S1 in the SI).

First-Order Approximation in the CAN Model. Using the Taylor series expansion of $\mathbf{T}_m(t)$ around $\mathbf{S}_m(t)$, the analysis of the biomolecular interactions involving the folding dynamics yields

$$\mathbf{T}_\mu(t) = \mathbf{T}^{(0)}_\mu + \sum_{\nu} \mathbf{T}^{(1)}_{\mu\nu} \mathbf{S}_\nu(t) + \sum_{\nu} \sum_{\gamma} \mathbf{T}^{(2)}_{\mu\nu\gamma} \mathbf{S}_\nu(t) \mathbf{S}_\gamma(t) + ... \tag{5}$$

where $\mathbf{T}^{(0)}_\mu$ is a 6 × 6 probability matrix that describes the effect of intramolecular interactions on the transition of $\mathbf{S}_\mu(t)$, $\mathbf{T}^{(1)}_{\mu\nu}$ is a 6 × 6 × 6 third-order probability tensor that describes the effect of intermolecular interactions between two nearest neighbors V_μ and V_ν on the transition of $\mathbf{S}_\mu(t)$, $\mathbf{T}^{(2)}_{\mu\nu\gamma}$ is a 6 × 6 × 6 × 6 fourth-order probability tensor that describes the effect of intermolecular interactions among three neighbors V_μ, V_ν, and V_γ on the transition of $\mathbf{S}_\mu(t)$, and so on; here, $\mu \neq \nu$, $\mu \neq \gamma$, and $\nu \neq \gamma$. Only terms for the first three orders (zeroth, first, and second) are shown in eq 3, and other higher-order items are not displayed. Equation 5 describes how the intramolecular and intermolecular interactions determine the cellular-automata rules of our CAN model in a general mathematical form. Although including higher-order items would increase precision in computer simulations using this model, it would substantially increase the computational complexity. We found that to capture the major physical characteristics of the folding transition, the first-order approximation, which truncates eq 5 after the second term, is sufficient to produce simulation results that are well-comparable to experiment. The component $T^{(0)}_{\mu,mn}$ of $\mathbf{T}^{(0)}_\mu$ represents the probability that $\mathbf{S}_\mu(t)$ transitions from $\mathbf{s}_n$ to $\mathbf{s}_m$ ($m, n \in \{1, 2, ..., 6\}$) as a result of the intramolecular interactions. This means that all $\mathbf{T}^{(0)}_\mu$ are the same (i.e., $\mathbf{T}^{(0)}_\mu = \mathbf{T}^{(0)}$ for $\mu \in \{1, 2, ..., M\}$, where M is the total number of biomolecules). Because the intermolecular interactions might not be isotropic, the third-order tensors $\mathbf{T}^{(1)}_{\mu\nu}$ may be functions of the azimuth $\theta_{\mu\nu}$ of edge $\overline{V_\mu V_\nu}$;[40−44] that is, $\mathbf{T}^{(1)}_{\mu\nu} = \mathbf{T}^{(1)}(\theta_{\mu\nu})$ for any V_μ and its interacting neighbor V_ν. The component $T^{(1)}_{mnk}(\theta_{\mu\nu})$ of $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ represents the probability that $\mathbf{S}_\mu(t)$ transitions from $\mathbf{s}_n$ to $\mathbf{s}_m$ contributed by the intermolecular interactions between V_μ in state $\mathbf{s}_n$ and its neighbor V_ν in state $\mathbf{s}_k$ (m, n, $k \in \{1, 2, ..., 6\}$). Thus, eq 5 may be simplified in the first-order approximation as

$$\mathbf{T}_\mu(t) \approx \mathbf{T}^{(0)} + \sum_{\nu} \mathbf{T}^{(1)}(\theta_{\mu\nu}) \mathbf{S}_\nu(t) \tag{6}$$

Here the summation over ν includes all of the vertices V_ν that are the interacting neighbors of V_μ. This is the tensor representation of eq 2.

Comparison between Experiments and Simulations. We simulated the entire folding process using the CAN model to track in real time the folding-state changes of 16 million

DNA molecules in an area of 8 μm × 8 μm at single-molecule resolution (see Methods). Comprehensive comparisons between experiments and simulations (Figures 2, 3, and 5; also see note S2 and Figure S6 in the SI) revealed that the transiently self-organized pattern is a manifestation of the coexistence of I_1, I_2, T_{23}, I_3, and F emerging in nonequilibrium thermodynamic states. The topological features of the self-organized pattern were well-captured in the computer simulations (Figure 2G,H). Each folding state shows its own distinctive topological features (see the Figure 2 legend and Table S1 in the SI). The qualitative features of time-dependent evolutions of the folding states observed in AFM imaging were also reproduced in the simulations (Figure 3A,B and Figure S6; see note S2 in the SI for a detailed description). Taken together, these allowed us to measure experimentally their structural and dynamic properties. For example, the topographic heights of I_1, I_2, T_{23}, I_3, and F were measured to be 9.5 ± 0.5, 6.5 ± 0.5, 10.0 ± 0.5, 6.5 ± 0.5, and 4.5 ± 0.5 nm, respectively (Figure S3 in the SI). These are well-consistent with the theoretical expectations (10.2, 6.8, 10.2, 6.8, and 4.9 nm, respectively). The dynamic evolution of the percent population of each state remarkably achieved quantitative agreement between the experiments and computer simulations (Figure 5A), confirming the applicability of the CAN model.

Misfolding and Folding Pathways. The experimental identification of the misfolding and folding pathways (Figure 4D), which were verified in the computer simulations (see the movie in the SI), suggests that the nanoscale self-organization behavior is a direct consequence of thermodynamic competitivity between the off-pathway and on-pathway intermediates. Our results revealed that those individuals in the molecular population fold into off-pathway and on-pathway intermediates simultaneously (Figures 4B,D and 5A). Remarkably, the three intermediates and F show simultaneous nucleation and growth (see Figures S4−S6 and the movie in the SI), but the pathways of neighboring molecules are strongly coupled. As a result, the F nuclei prefer to emerge from the I_3 domains rather than the I_2 domains (Figures 2 and 3 and Figure S4 in the SI), confirming that only I_3 is on-pathway. Because I_1 is less stable than I_2 and I_3, their thermodynamic competition strongly suppresses the growth of I_1 nuclei, limiting the sizes of I_1 nanoclusters to ∼150 nm. Interestingly, the competition between I_2 and I_3 gives rise to the quasi-static T_{23} interfaces between their domains as a result of their comparable thermodynamic stabilities.

Misfolding Competitivity. In the absence of intermolecular interactions, folding cooperativity merely makes sense of the intramolecular conformational dynamics. In the presence of large-scale intermolecular interactions, folding cooperativity may be understood on two levels. One is the intramolecular cooperation between the foldings of distinct intramolecular domains. The other is the intermolecular cooperation between the foldings of neighboring biomolecules, which reflects the scenario that the on-pathway folding of a biomolecule facilitates and promotes the on-pathway folding of its neighbors. In this study, we mainly focused on the latter, in that the former would best be addressed under conditions of no strong intermolecular interactions. Figure 5B depicts the dynamic quantification of off-pathway misfolding behavior, the degree of which is indicated by the percent population trapped in the off-pathway intermediates relative to that for on-pathway folding (see the Figure 5 legend for details). The first-order derivatives of the dynamics approximately indicate the relative degree of misfolding competitivity versus folding cooperativity (Figure 5C). The second-order discontinuity in

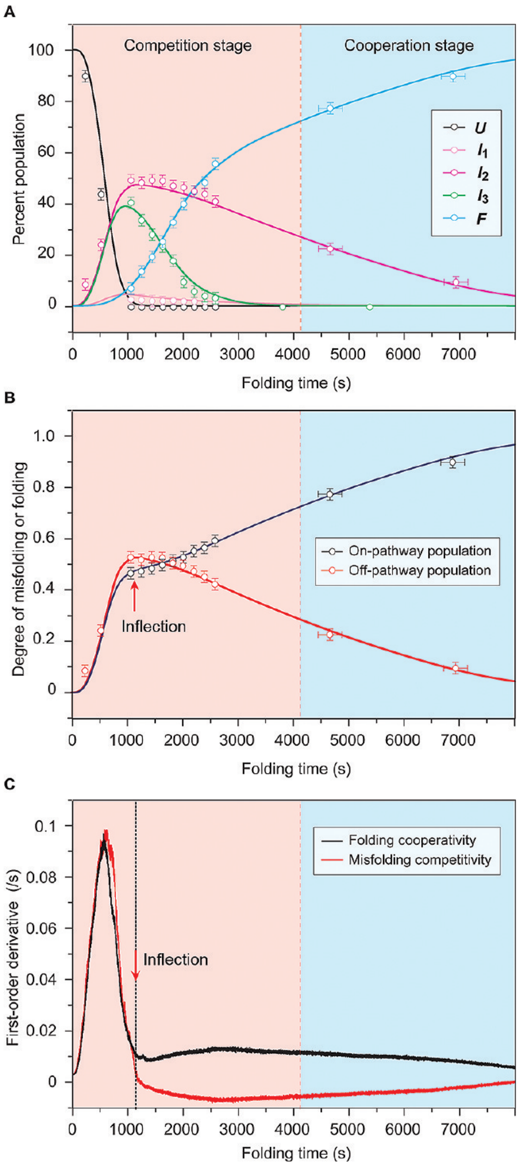


Figure 5. Dynamics of misfolding competitivity vs folding cooperativity. (A) Percent populations of the misfolding and folding intermediate states as functions of folding time obtained from experimental measurements (open circles with error bars) and computer simulations using the CAN model (solid lines), showing good agreement between experiment and theory. (B) Quantification of misfolding vs folding. The degree of misfolding (red curve) was quantified by summing the percent populations of the off-pathway intermediates I_1, I_2, and T_{23}. The degree of folding (black curve) was similarly quantified by summing the percent populations of the on-pathway intermediate I_3 and F. Solid lines and open circles represent results from computer simulations and experimental measurements, respectively. (C) First-order derivatives of the two curves shown in (B) obtained by computer simulations, demonstrating the second-order discontinuity with respect to the time-dependent population dynamics of misfolding and folding shown in (B). The first-order derivatives approximately indicate the relative degree of misfolding competitivity (red curve) vs folding competitivity (black curve). The corresponding inflection, where the second-order discontinuity occurs, is marked by the red arrow. This effect is attributed to the strong intervention of misfolding competitivity against folding cooperativity. The overall folding process may be understood as a competition stage followed by a cooperation stage. The end of the competition stage is defined by the percent population of misfolding intermediates that is half of its maximum.

the dynamics is a hallmark of the strong intervention of misfolding competitivity against intermolecular folding cooperativity (Figure 5B,C). It has been a puzzle whether the off-pathway misfolding would be irreversible and lead to dead ends.[9] Our results clarify that this may not necessarily be the case: to our surprise, the folding cooperativity may still dominate the molecular population at the late stage of the folding transition (Figure 5B,C). In this case, the free energy of the native structure is sufficiently lower than that of the off-pathway misfolding structures (Figure 4B) so that the intermolecular interactions do not fully compensate for the loss in intramolecular energy during the enthalpically driven aggregation of neighboring chains. Notably, within the same off-pathway misfolding intermediates, the formation of intermolecular interaction networks shows enthalpy-driven cooperation, which nevertheless has an adverse effect that promotes the thermodynamic competition between the off-pathway and on-pathway intermediates. Taken together, our findings suggest that the entire folding transition obeys a paradigm in which thermodynamic competition dominates the early stage of folding and is then followed by an intermolecular cooperation stage. This may be succinctly called a competition—cooperation—condensation mechanism, in accordance with the nucleation—condensation mechanism previously proposed for biopolymer folding under conditions with no intermolecular interactions.[2−5] However, it remains unclear how the suppression of intermolecular folding cooperation affects the intramolecular folding cooperation; this question is expected to be addressed in the future.

This study has unraveled several novel aspects of biopolymer folding properties that were missing in the absence of large-scale intermolecular interactions. Previous studies underscored the crucial role of folding cooperativity.[6−17] This study has revealed how complex intermolecular interactions induce strong thermodynamic competitivity. This highlights the misfolding competitivity as a factor that is equally important as the folding cooperativity in determining the folding dynamics when strong interpathway coupling happens. Furthermore, the phase transition behaviors in those well-studied many-body systems[39−45] often involve simple, elemental interactions between particles or molecules. In this system, the perplexing variation of intermolecular interactions with conformational changes in misfolding and folding confer multiplex phase transitions in which phase separations occur along multiple pathways simultaneously and are transiently self-organized into spatiotemporal patterns. These behaviors are expected to reflect the general properties of frustrated biomolecular systems in the regime of strong intermolecular interactions, although the complex folding system in this study is different from in vivo biopolymer folding systems in cells. The folding of proteins and nucleic acids in cells proceeds in a highly heterogeneous and complex composition and is in many cases biologically governed by chaperones that individually facilitate the correct folding process of the biopolymer molecules.[46] Hence, even at high copy numbers, the nearest neighbors during folding in cells are most likely to be chaperones or heterogeneous molecules. In contrast, the system in this study is highly homogeneous in composition, and the large numbers of identical molecules make it closely relevant for biopolymer nanomaterials or crystallization of highly pure and concentrated biopolymer systems.

SUMMARY

In this work, we have directly visualized the nanoscale thermodynamic competitivity of biopolymer folding arising from large-scale intermolecular interactions. A simple statistical-mechanical model was established to explain diverse experimental observations with a single set of parameters (see Table 1). Our results sculpture a novel concept of misfolding competitivity, in contrast to the established concept of folding cooperativity. We show that in the presence of large-scale intermolecular interactions, the intervention of misfolding competitivity against folding cooperativity makes the folding behavior strongly deviate away from purely cooperative dynamics. Looking forward, the CAN model formulation provides a mathematical framework for investigating other complex biomolecular systems. This may be of general utility in understanding how large-scale complex intermolecular interactions correlate with structural dynamics of individual biomolecules together at the systems level. The quantitative insight into the folding and misfolding mechanisms under such complex conditions should assist biologically inspired engineering of high-order nanostructures and nanomachines.

METHODS

Sample Preparation. Au(111) surfaces were obtained by evaporating gold on newly cleaved mica surfaces, annealed at 350 °C for 60 min. The mica was cleaved before use. The DNA sequence used was 5′-AAAAAAAAAACCCTAACCCTAACCCTAACCC-3′ with the 5′ end modified by a thiol group (−SH). The DNA molecules were synthesized by the standard procedure and purified by polyacrylamide gel electrophoresis. 5′-Thiolated DNA was dissolved in 100 mM phosphate-buffered saline (pH 4.5) with 100 mM NaCl to give a final concentration of 1 μM. The DNA solution was heated to 80 °C for 5 min, allowed to anneal to 20 °C, and then applied to the newly prepared Au(111) surface over an area of ~1 cm^2 for 24 h at 4 °C to achieve a compactly self-assembled monolayer of *i*-motif structures.

Atomic Force Microscopy. AFM imaging was performed in PBS using a Molecular Imaging PicoPlus II system (Agilent) with MACLever cantilevers fixed on a small-range scanner controlled by a Picoscan 2500 controller. The cantilevers used (force constant, 0.1 N/m; length, 140 μm; thickness, 0.6 μm; mean width, 18 μm; resonant frequency in buffer solution, 28−30 kHz) were type-IV MACLevers (Molecular Imaging, Agilent). All images were obtained in tapping mode driven by magnetic fields [magnetic AC (MAC) mode]. The use of MAC mode has been shown to result in a decreased extent of disturbance of the underlying biomolecules and the solution around them by the cantilever in comparison with tapping mode driven by acoustic mechanical oscillation.[31]

Refolding Experiments and Real-Time Observations. After the model system was self-assembled at pH 4.5, it was first scrutinized by liquid-phase AFM at both pH 4.5 and pH 8.5 to ensure quality of the molecular self-assembly. Samples with insufficient self-assembly density or surface coverage were abandoned. The DNA molecules were denatured at pH 8.5 before the beginning of the refolding experiments. High-resolution liquid-phase AFM required 5−7 min to complete the scanning of each image. However, the single molecules fold on a time scale of seconds,[19] and the entire DNA array may complete folding in several minutes when supplied with sufficient protons.[21] This means that the folding rate had to be slowed by 1−2 orders of magnitude to allow AFM to capture the folding intermediates. By repeating the experiments about a hundred times, we found three methods to perform this task, each resolving one aspect of the experimental issues and providing information complementary to the others. In the first method, we repeated AFM scanning of only two images each cycle immediately after the pH value of the folding buffer was decreased by 0.5 over the range from pH 8.5 to 4.5. The pH titration was done by fully exchanging the folding buffer in situ at the AFM imaging site with 500 μL of folding buffer with a decreased pH

Table 1. Components of $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ Used in the Computer Simulations[a]

		mn								
	k	12	13	15	16	23	34	43	45	56
$\mathbf{T}^{(0)}$		0	0	6×10^{-10}	10^{-9}	0	0	0	0	8×10^{-9}
$\mathbf{T}^{(1)}$	1	8×10^{-8}	9×10^{-11}	0	0	0	0	0	0	0
	2	9.9×10^{-4}	7×10^{-9}	0	0	0	0	0	0	0
	3	0	7×10^{-2}	0	0	1×10^{-4}	0	10^{-6}	0	0
	4	0	0	0	0	0	6×10^{-4}	0	0	0
	5	0	0	$1.5 \times 10^{-1} f(\theta_{\mu\nu})$	0	2×10^{-3}	9×10^{-3}	0	5×10^{-6}	0
	6	0	0	4×10^{-2}	10^{-6}	4×10^{-3}	5×10^{-2}	0	10^{-1}	$4 \times 10^{-2} g(\theta_{\mu\nu})$

[a] The labels m, n, and k refer to the component subscripts of $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ according to the representations: $T^{(0)}_{mn}$ and $T^{(1)}_{mnk}(\theta_{\mu\nu})$. The diagonal components (not shown) were calculated according to the normalization constraint. Each nonzero component of $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ represents an accessible state transition on the folding pathways. Components corresponding to thermodynamically insignificant or impossible state transitions were set to zero. Although there are 63 off-diagonal components in total for $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$, only 21 are nonzero. One of the advantages of our CAN model is that it allows a possible folding pathway to be tested simply by setting its corresponding $\mathbf{T}^{(0)}$ or $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ component to either zero (negative test) or a significant probability (positive test). One may readily find out that the significant probabilities shown in this table strictly correspond to the folding pathways shown in Figure 4D. These significant probabilities were estimated from $\Delta\mathbf{G}^{(0)}$ and $\Delta\mathbf{G}^{(1)}(\theta_{\mu\nu})$ according to eq 4. There are rich existing publications providing basic (though partial) thermodynamic data for the estimation of $\Delta\mathbf{G}^{(0)}$ and $\Delta\mathbf{G}^{(1)}(\theta_{\mu\nu})$ for the model system.[47−52] The anisotropic growths of the I_3 and F domains are reflected in the dependence of $T^{(1)}_{155}(\theta_{\mu\nu})$ and $T^{(1)}_{566}(\theta_{\mu\nu})$ on $\theta_{\mu\nu}$. $T^{(1)}_{155}(\theta_{\mu\nu})$ and $T^{(1)}_{566}(\theta_{\mu\nu})$ were empirically fit by the expressions $f(\theta_{\mu\nu}) = \exp[3.6 - 3.6/\cos(\theta_{\mu\nu})]$ and $g(\theta_{\mu\nu}) = \exp[4.2 - 4.2/\cos(\theta_{\mu\nu})]$, respectively.

value. This method allowed the overall topographic changes of the folding transition to be captured (Figure 1). However, it was not possible to capture the detailed microscopic coarsening of the folding intermediates by this method because of the dramatic topographic change upon the decrease in pH by 0.5. In the second method, we solved this problem by directly changing the pH value from 8.5 to 5.5 and simultaneously decreasing the volume of the folding buffer to 100 μL in a single cycle of buffer exchange. We then rapidly searched across the areas around the original imaging site in a range of 50 μm × 50 μm. The folding rates across different local areas were not homogeneous because the in situ exchange of the folding buffer could only be realized by injecting new buffers from one side of the buffer reservoir, which gave rise to inhomogeneous pH changes across the imaging areas. For this reason, we speculated that there were certain areas where the folding transition was slowed by 2 orders of magnitude at the middle stage of folding but still kept proceeding at a rate matching the time resolution of AFM. Following this idea, we did find such areas and kept repeating AFM imaging at the same areas, though most other areas were observed to complete folding on the time scale of the first frame of AFM imaging after the pH adjustment. This enabled the real-time observation of the mesoscopic coarsening of the F domains by AFM phase imaging (Figure 3A). However, the topographic images were found to be unclear because of dramatic motions of the molecules, which gave rise to false z-feedback of AFM control system (Figure S5 in the SI). In the third method, we solved this problem by fully stopping the folding at its middle stage. After finding out the areas that showed the mesoscopic coarsening of the F domains, we decreased the volume of pH 5.5 folding buffer to 30 μL. The insufficient number protons in the smaller volume of folding buffer caused certain areas to stop folding at the middle stage. This allowed clear high-resolution topographic imaging of the folding intermediates (Figure 2).

Computer Simulations. The implementation algorithm for the computer simulations was strictly based on the CAN model as described in the Results and Discussion. In brief, the algorithm first calculates $\boldsymbol{P}_m(t)$ for each vertex using the current state $\boldsymbol{S}_m(t)$ and then updates the state to give $\boldsymbol{S}_m(t + \mathrm{d}t)$ at the next time step using the calculated $\boldsymbol{P}_m(t)$. This process starts with all vertex states being set as $\mathbf{s}_1$ and is iterated until all vertex states are changed to $\mathbf{s}_6$. The off-diagonal components of $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ used in the simulations were estimated from eq 4 and are provided in Table 1. To visualize the simulations, the CAN was mapped to a bitmap image in which a unit array of 2 × 2 pixels represents a molecule. The unit pixel arrays were packed into a rhombic lattice in a square of the bitmap. Each folding state was represented by a different color, and the colors were chosen for the convenience of comparison with experiments. The simulation program was written in the standard C language. To simulate the folding transition in an 8 μm × 8 μm area, a bitmap size of 8000 × 8000 pixels was used, corresponding to 4000 × 4000 DNA molecules.

All of the simulation results presented in this paper were based on a single set of parameters (Table 1). The simulation program is available upon request. The parameters used in the CAN simulations fall into two categories: $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}$. $\mathbf{T}^{(0)}$ represents the portion of the state-transition probability contributed solely by the intramolecular interactions, whereas $\mathbf{T}^{(1)}$ represents the part contributed by the intermolecular interactions. Each nonzero component of $\mathbf{T}^{(0)}$ and $\mathbf{T}^{(1)}(\theta_{\mu\nu})$ represents an accessible state transition on the folding pathways. Since there are only three on-pathway states ($U \rightarrow I_3 \rightarrow F$), only $T^{(0)}_{15}$, $T^{(0)}_{16}$, and $T^{(0)}_{56}$ should be nonzero, because the intramolecular interactions alone account for the state-transition probability of on-pathway folding. Specifically, $T^{(0)}_{15}$ and $T^{(0)}_{16}$ correspond to the nucleation rates of I_3 and F from U, respectively, and $T^{(0)}_{56}$ corresponds to the folding rate from I_3 to F. The parameters of $\mathbf{T}^{(1)}$ can also be further classified into two major categories: (a) nucleation rates of folding/misfolding intermediates and (b) the growth rates of intermediates and F. The parameters of the form $T^{(1)}_{1nk}$ are in category (a). If a molecule folds into an on-pathway state (I_3 or F), it can be considered less probable that this molecule will form intermolecular hydrogen bonds with its neighbors, and thus, it should not affect the nucleation rate of the off-pathway intermediates among its neighbors because of the tight structure of the i-motif and short quadruplex. This means that when k = 4, 5, or 6, $T^{(1)}_{1nk}$ should be zero for n = 2 and 3. On the other hand, for a molecule that has folded into an on-pathway state (I_3 or F), its neighbors that are in off-pathway intermediate states should have no effect on the nucleation rate of I_3 or F in this molecule. This means that when k = 1, 2, 3, or 4, $T^{(1)}_{1nk}$ should be zero for n = 5 and 6. These rationales can be generalized to a basic rule: the components corresponding to thermodynamically insignificant or impossible state transitions are set to zero. It is not difficult to apply this rule to the other $T^{(1)}_{mnk}$ parameters in category (b). Specific attention was paid to the anisotropic intermolecular interactions of the molecules in the on-pathway states (I_3 and F). The anisotropic growth of the I_3 and F domains is reflected in the dependence of $T^{(1)}_{155}(\theta_{\mu\nu})$ and $T^{(1)}_{566}(\theta_{\mu\nu})$ on $\theta_{\mu\nu}$. $T^{(1)}_{155}(\theta_{\mu\nu})$ and $T^{(1)}_{566}(\theta_{\mu\nu})$ were empirically fit to the expressions $f(\theta_{\mu\nu}) = \exp[3.6 - 3.6/\cos(\theta_{\mu\nu})]$ and $g(\theta_{\mu\nu}) = \exp[4.2 - 4.2/\cos(\theta_{\mu\nu})]$, respectively. We also tested the sensitivity of the simulation outcomes against each parameter. We found that the results are highly sensitive to $\mathbf{T}^{(1)}$: whenever a $\mathbf{T}^{(1)}$ parameter was changed by more than 1 order of magnitude, there were always essential changes in features that were remarkably inconsistent with the experimental observations. Only the single set of parameters shown in Table 1, which strictly correspond to the folding/misfolding pathways summarized in Figure 4D, was used in all of the simulations presented in this paper that achieved agreement in all measurable aspects. After more than 100 runs of the simulation program with changed nonzero parameters, we did not find a nonzero parameter that could be changed freely (or by more than 1 order of magnitude) without destroying particular features corresponding to our experimental observations.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Notes on intermolecular interactions and the energy landscape and on the comparison between experimental observations and computer simulations; seven supporting figures; a table of the physical features of the misfolding and folding states; and a movie showing the complete folding transition as obtained from a computer simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

youdong_mao@dfci.harvard.edu

■ ACKNOWLEDGMENTS

The authors are grateful to D. Han, Q. Ouyang, L. Jiang, D. Liu, and J. G. Sodroski for helpful discussions and support in experiments. We thank S. Wang and K. Wu for assistance in sample preparation and Q. Ouyang and J. G. Sodroski for feedback and comments on the manuscript.

■ REFERENCES

(1) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309−1314.

(2) Pande, V. S.; Grosberg, A. Y.; Tanaka, T. *Rev. Mod. Phys.* **2000**, *72*, 259−314.

(3) Onuchic, J. N.; Schulten, Z. L.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(4) Daggett, V.; Fersht, A. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497−502.

(5) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.

(6) Shank, E. A.; Cecconi, C.; Dill, J. W.; Marqusee, S.; Bustamante, C. *Nature* **2010**, *465*, 637−640.

(7) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123−127.

(8) Hamada, D.; Tanaka, T.; Tartaglia, G. G.; Pawar, A.; Vendruscolo, M.; Kawamura, M.; Tamura, A.; Tanaka, N.; Dobson, C. M. *J. Mol. Biol.* **2009**, *386*, 878−890.

(9) Cellmer, T.; Bratko, D.; Prausnitz, J. M.; Blanch, H. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11692−11697.

(10) Wensley, B. G.; Batey, S.; Fleur, A. C. B.; Chan, Z. M.; Tumelty, N. R.; Steward, A.; Kwa, L. G.; Borgia, A.; Clarke, J. *Nature* **2010**, *463*, 685−688.

(11) Solomatin, S. V.; Greenfeld, M.; Chu, S.; Herschlag, D. *Nature* **2010**, *463*, 681−684.

(12) Wallace, M. I.; Ying, L.; Balasubramanian, S.; Klenerman, D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5584−5589.

(13) Ansari, A.; Kuznetsov, S. V.; Shen, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771−7776.

(14) Ma, H.; Wan, C.; Wu, A.; Zewail, A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 712−716.

(15) Galzitskaya, O. V.; Finkelstein, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11299−11304.

(16) Chen, S. J.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 646−651.

(17) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.

(18) Caliskan, G.; Hyeon, C.; Perez-Salas, U.; Briber, R. M.; Woodson, S. A.; Thirumalai, D. *Phys. Rev. Lett.* **2005**, *95*, No. 268303.

(19) Liu, D.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 5734−5736.

(20) Shu, W.; Liu, D.; Watari, M.; Riener, C. K.; Strunz, T.; Welland, M. E.; Balasubramanian, S.; McKendry, R. A. *J. Am. Chem. Soc.* **2005**, *127*, 17054−17060.

(21) Mao, Y.; Liu, D.; Wang, S.; Luo, S.; Wang, W.; Yang, Y.; Ouyang, Q.; Jiang, L. *Nucleic Acids Res.* **2007**, *35*, e33.

(22) Xia, F.; Guo, W.; Mao, Y.; Hou, X.; Xue, J.; Xia, H.; Wang, L.; Song, Y.; Ji, H.; Ouyang, Q.; Wang, Y.; Jiang, L. *J. Am. Chem. Soc.* **2008**, *130*, 8345−8350.

(23) Seeman, N. C. *Nature* **2003**, *421*, 427−431.

(24) Bath, J.; Turberfield, A. J. *Nat. Nanotechnol.* **2007**, *2*, 275−284.

(25) Stuart, M. A. C.; Huck, W. T. S.; Genzer, J.; Muller, M.; Ober, C.; Stamm, M.; Sukhorukov, G. B.; Szleifer, I.; Tsukruk, V. V.; Urban, M.; Winnik, F.; Zauscher, S.; Luzinov, I.; Minko, S. *Nat. Mater.* **2010**, *9*, 101−113.

(26) Mao, Y.; Chang, S.; Yang, S.; Ouyang, Q.; Jiang, L. *Nat. Nanotechnol.* **2007**, *2*, 366−371.

(27) Wang, S.; Liu, H.; Liu, D.; Ma, X.; Fang, X.; Jiang, L. *Angew. Chem., Int. Ed.* **2007**, *46*, 3915−3917.

(28) Guéron, M.; Leroy, J. L. *Curr. Opin. Struct. Biol.* **2000**, *10*, 326−331.

(29) Gehring, K.; Leroy, J. L.; Guéron, M. *Nature* **1993**, *363*, 561−565.

(30) Kang, C.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11636−11640.

(31) Han, W.; Lindsay, S. M.; Jing, T. *Appl. Phys. Lett.* **1996**, *69*, 4111−4113.

(32) Stark, M.; Moller, C.; Muller, D. J.; Guckenberger, R. *Biophys. J.* **2001**, *80*, 3009−3018.

(33) Wolfram, S. *Rev. Mod. Phys.* **1983**, *55*, 601−644.

(34) Kramers, H. A. *Physica* **1940**, *7*, 284−304.

(35) Carlon, E.; Orlandini, E.; Stella, A. L. *Phys. Rev. Lett.* **2002**, *88*, No. 198101.

(36) Chang, S.; Mao, Y.; Luo, C.; Ouyang, Q. *J. Phys.: Conf. Ser.* **2006**, *29*, 18−26.

(37) Kafri, Y.; Mukamel, D.; Peliti, L. *Phys. Rev. Lett.* **2000**, *85*, 4988−4991.

(38) Baiesi, M.; Orlandini, E.; Stella, A. L. *Phys. Rev. Lett.* **2003**, *91*, No. 198102.

(39) Ha, B. Y.; Liu, A. J. *Phys. Rev. Lett.* **1997**, *79*, 1289−1292.

(40) Lorman, V.; Podgornik, R.; Zeks, B. *Phys. Rev. Lett.* **2001**, *87*, No. 218101.

(41) Harreis, H. M.; Kornyshev, A. A.; Likos, C. N.; Löwen, H.; Sutmann, G. *Phys. Rev. Lett.* **2002**, *89*, No. 018303.

(42) Kornyshev, A. A.; Leikin, S. *Phys. Rev. Lett.* **2001**, *86*, 3666−3669.

(43) Kornyshev, A. A.; Leikin, S. *Phys. Rev. Lett.* **1999**, *82*, 4138−4141.

(44) Kornyshev, A. A. *Phys. Rev. Lett.* **2000**, *84*, 2537−2540.

(45) Qiu, X.; Kwok, L. W.; Park, H. Y.; Lamb, J. S.; Andresen, K.; Pollack, L. *Phys. Rev. Lett.* **2006**, *96*, No. 138101.

(46) Fenton, W. A.; Horwich, A. L. *Q. Rev. Biophys.* **2003**, *36*, 229−256.

(47) Leroy, J. L.; Gueron, M.; Mergny, J. L.; Helene, C. *Nucleic Acids Res.* **1994**, *22*, 1600−1606.

(48) Mergny, J. L.; Lacroix, L.; Han, X.; Leroy, J. L.; Helene, C. *J. Am. Chem. Soc.* **1995**, *117*, 8887−8898.

(49) Mergny, J. L.; Lacroix, L. *Nucleic Acids Res.* **1998**, *26*, 4797−4803.

(50) Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J. L. *Biophys. J.* **2003**, *84*, 3838−3847.

(51) Mathur, V.; Verma, A.; Maiti, S.; Chowdhury, S. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1220−1227.

(52) Zhao, Y.; Zeng, Z. X.; Kan, Z. Y.; Hao, Y. H.; Tan, Z. *ChemBioChem* **2005**, *6*, 1957−1960.

Supporting Information for

Understanding Thermodynamic Competitivity between Biopolymer Folding and Misfolding under Large-Scale Intermolecular Interactions

Youdong Mao†* and Jinzhong Zhang§

†Dana-Farber Cancer Institute, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02215, USA. §Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA.

*To whom correspondence should be addressed. E-mail: youdong_mao@dfci.harvard.edu.

Table of Contents

Notes

S1. Intermolecular interactions and the energy landscape

There are three major types of intermolecular interactions in the model system: hydrogen bonding, excluded-volume, and electrostatic interactions. The hydrogen bonds between neighbours is the major determinant of the off-pathway misfolding intermediates and of their stability. The excluded-volume interaction has been shown to play an important role in the phase transitions of DNA hybridization or denaturation[35-37], as well as structures of RNA molecules[38]. This interaction has an impact on both the off-pathway and on-pathway intermediates. The electrostatic interaction, especially, the counterion-induced attraction potential between the well-folded *i*-motif structures[40-45], mainly influences the stability of *F*.

The excluded-volume interaction describes the spatial exclusion of the volumes of neighbouring chains. This type of interaction exists for the folding of single biomolecules, but is expected to have more impact when millions of biomolecules fold at the crowding limit. The typical phenomenological outcome, which we attribute to this interaction, is the anisotropic growth of I_3 domains. As I_3 is a semi-folded on-pathway intermediate, incomplete folding does not permit the molecules to fill the in-plane space very tightly, which gives them a chance to stack against their neighbours. The excluded-volume interaction results in that the space available for the neighbours to fold (or chain diffusion) is wide in certain direction (x) but is narrow in the orthogonal direction (y). Thus, the chains can have a faster folding rate in the x-direction than in the y-direction.

A recent simulation study on protein folding in few-chain system suggested that the intermolecular interactions increase the chance of protein aggregation[9]. A similar situation was directly observed in our model system. The intermediates I_1 and I_2 are the typical biopolymer aggregation. This raises the question of why millions of biomolecules in the model system are still able to fold into the native structure. A reasonable answer revealed in this study lies in the third interaction. The electrostatic interaction between neighbouring folded structures deepens the global energy minimum in the energy landscape relative to that for a single molecule in the absence of the intermolecular interactions. The observation presented in Fig. S2 shows that the azimuths of the *i*-motif structures (i.e., the orientation of the long axis of its footprint) in the model system are well aligned in the *F* domains. The *i*-motif structure with its cross-section of 1.3×2.0 nm^2 only has $P2$ symmetry (Fig. 1A of the paper), which may account for the azimuth-dependent electrostatic interactions between neighbouring quadruplex. It has major grooves much wider than its minor grooves, which results in that the counterions accumulated in the major grooves largely exceed those in the minor grooves. Another evidence for the anisotropy of the electrostatic interactions is that the growth rate of *F* domains is highly anisotropic, as directly observed in the real-time coarsening of *F* domains (Fig. 3 of the paper and Fig. S4). This was well verified in the computer simulations (Fig. S7 and Movie S1). Similar electrostatic interactions were extensively studied and proposed for double helix DNA structures[39-45].

S2. Comparison between experimental observations and computer simulations

When comparing computer simulations with AFM experiments, the image convolution effect by the AFM cantilever tips should be taken into account. The AFM cantilever tips usually has a radius of 10̃50 nm, which results in the x-y widening over the real x-y dimension of the imaging object. The degree of the widening effect is also dependent of the z height of the object, due to the pyramidal or conical shapes of AFM cantilever tips; the higher the object, the wider its size in AFM image. Thus, the measured width of ~50 nm of the T_{23} interfaces in Fig. S3 should only reflect the width of few DNA molecules. The sizes of the I_1 clusters in Fig. 2 of the paper should reflect the size of about no more than tens of molecules. With consideration of the image convolution effect, we find that all experimental observations about the topological features were qualitatively reproduced in the computer simulations; this includes comparisons shown in Figs 2 and 3 of the paper, Figs S6 and S7. The simulations remarkably reach a quantitative agreement with experiments with respect to the time-dependent evolution of the percent population of each folding state, as shown in Fig. 5 of the paper. We summarize in the following several major aspects of the phenomenological agreement between experiments and simulations with respect to the detailed topological and dynamic features of the folding pathways and the phase separations. See Movie S1 for a dynamic visualization of the features.

1) I_1 can only nucleate into small molecular clusters. All three intermediates nucleate simultaneously, leaving them in a strong competition against others. Because I_1 is less stable than both I_2 and I_3, the growth of I_2 and I_3 domains suppresses the growth of I_1 nuclei. In both simulations and experiments, they appear as small flower-like nano-clusters with a size around 10̃100 nm.

2) The localization of I_1 clusters results from the spatiotemporal suppression and thermodynamic competition from I_2 and I_3. One may notice that in both experiments and simulations, the I_1 nano-clusters mainly appear inside the I_2 domains, and rarely show in the I_3 domains. When the I_1 nuclei are wrapped inside the I_3 domains, they incur stronger thermodynamic competition from I_3 compared to the situation that they are wrapped inside the I_2 domains, which results in a faster reaction rate to break their intermolecular hydrogen bonds. Movie S1 shows that the I_1 nuclei rapidly shrink and disappear when surrounded by I_3 domains. When the I_1 nuclei are wrapped inside the I_2 domains, they can still form intermolecular hydrogen bonds with their I_2 neighbours. This may markedly slow down their transition to lower energy intermediates. Inside the I_2 domains, there is also a notable localization of the I_1 nuclei; the I_1 nuclei barely show up in the areas around the emerging origins of I_2 nuclei (nucleation origins), because of the competition and suppression from I_2. This behaviour has two prominent features shown in both experiments and simulations. First, there is a gradient in the size of the I_2 clusters along the growth direction of I_2 domains. The size of I_1 clusters decrease as their positions approach to the I_2 nucleation origins. Second, those large I_1 clusters are localized either in the areas around T_{23} interfaces or in the areas where two I_2 domains meet each other upon

domain growth, because these areas are the farthest away from the I_2 nucleation origins.

3) The off-pathway intermediates I_1 and I_2 show isotropy in their growth. This strongly contrasts with the strong anisotropy in the growth of the on-pathway intermediate I_3 and F domains. This phenomenon may be attributed to that a molecule has the equal probability to form intermolecular hydrogen bonds with its neighbours in any direction.

4) The I_2 and I_3 domains both emerge simultaneously before the F nuclei begin to grow up. Because the I_1 nuclei are not stable enough to compete with the other two intermediates, most of the molecular populations are either in the domain of I_2 or I_3. We observed that those F nuclei are not able to grow to micrometer sizes before the I_2 or I_3 domains invade most of the areas and form the quasi-static interfaces. The strong intermolecular interactions not only prevent the molecules from directly folding into F, but also stabilize the on-pathway intermediates by the excluded-volume interactions that actually deepen the on-pathway energy wells in the folding landscape.

5) The T_{23} interfaces are prone to be quasi-static. The nuclei of I_2 and I_3 emerge simultaneously, and grow up simultaneously. When the two domains encounter each other, their interfaces form the T_{23} state. Because of the relative high energy barrier from I_2 to $I_{3,}$ and the similar free-energy levels between I_2 and I_3, the transition from I_2 to I_3 is much slower than the transition from I_3 to F, making the T_{23} interfaces have nearly no movement before F invades the I_2 domains. This may be regarded as a consequence of strong competition between the off-pathway and on-pathway intermediates.

6) The I_3 domains have anisotropic growth rates. One may notice from the self-organized patterns of the misfolding and folding intermediates that the geometry of I_3 domain is highly asymmetric and shows stripe-like shape. This indicates the anisotropic growth of the I_3 domain. We attribute this effect to the excluded-volume interactions, as discussed in the previous section.

7) F nucleates inside the I_3 domains rather than the I_2 regions. This is not only shown in the Fig. 2 of the paper, where many small F regions are found in the I_3 domains, but also more clearly shown in Fig. S4, where the F nuclei were coarsening inside the I_3 regions. This indicates that only I_3 is the on-pathway intermediate and I_1 and I_2 are off-pathway intermediates.

8) The F domains have anisotropic growth rates. The F domain grows faster in certain direction than in others. This is attributed to the counterion-induced intermolecular attraction potential between neighbouring *i*-motif structures that is highly dependent on the azimuths of the molecules, as discussed in the previous section.

9) The F domain grows faster against the I_3 domains than against the I_1 and I_2 domains. This provides another aspect of evidence that I_3 is the on-pathway intermediate and I_1 and I_2 are off-pathway intermediates. The conformations of I_1 and I_2 must first break all of their intermolecular hydrogen bonds before they can fold into F, which are energetically less favourable and need more time to approach the correct folding structure. The slower rate of F domain growth against I_1 and I_2 is a consequence of multiple state-transitions along the folding pathway ($I_2 \rightarrow T_{23} \rightarrow I_3 \rightarrow F$).

Supporting Figures

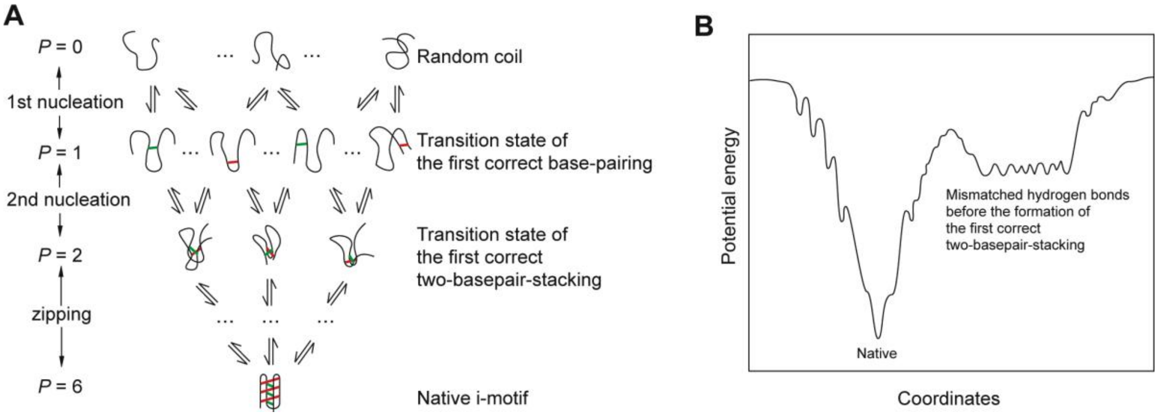


Figure S1. Schematics showing three-state approximation of single i-motif DNA folding in the absence of intermolecular interactions. (**A**) The folding pathways of single *i*-motif DNA molecules. The number of native C:C$^+$ base pairs is denoted as *P*. The folding is speculated to be consistent with the nucleation-condensation mechanism[4]. To fold into the native structure, the DNA chain must form the first correct C:C$^+$ base pair (the first nucleation with *P* =1), which then nucleates to the first correct two-basepair-stacking (two nearest correct C:C$^+$ base pairs stacking together; the second nucleation with *P* = 2). Once the first correct two-basepair-stacking is formed, where the entropy is largely consumed, the molecular chain will undergo a fast zipping-like base-pairing that leads to the final native *i*-motif structure. (**B**) The schematic energy landscape of the DNA folding in the absence of intermolecular interactions. A recent study on the hairpin-DNA folding using ultrafast spectroscopy[14] showed that, before the first folding nucleation, the molecular conformations have to change to a globular, compact structure that favours the formation of the folding nucleus (i.e., the first correct base pair). But the transition of this globular state to the conformation of the first folding nucleus suffers a small energy barrier owing to the existence of mismatched base-pair or local energy minima. The native *i*-motif DNA structure used in this study has three hairpin loops. Thus, it is expected that the same situation happens to the folding of this molecule, and its folding in the absence of intermolecular interactions is also described by a three-state folding model. Namely, the molecular conformation must first change to a globular, compact structure to facilitate the folding nucleation, the formations of the first correct base-paring and the first correct two-basepair-stacking. This argument is well consistent with the study presented in this paper, where only one on-pathway intermediate (I_3) was well observed. In other words, the energy landscape of individual molecule in our folding system degenerates into three states, when all of its neighbours fold into the *i*-motif structures. This analysis also suggests that the large-scale intermolecular interactions transform a three-state folding under single-molecule conditions into a five-state folding. This is a hallmark arising from the large-scale, complex intermolecular interactions.

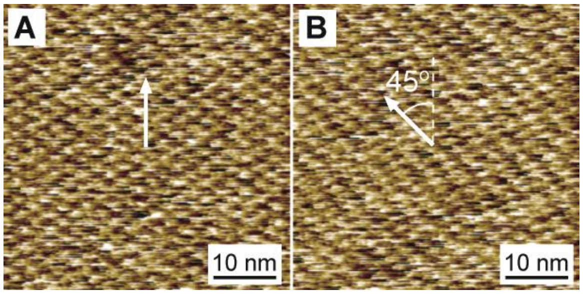


Figure S2. High-resolution liquid-phase AFM images in the *F* domain. (**A**) The rhombic lattice of the DNA molecules observed after the molecules were well folded into the *i*-motif structures at pH 4.5. The footprint size of each molecule in this AFM image can be directly measured to be 2.0 × 1.3 nm^2, which is well consistent with the crystal structure of *i*-motif[30]. (**B**) The same area as **A** obtained by rotating the AFM scanning direction by 45°. The slight twist of molecular lattice direction may result from the perturbation of AFM scanning on the azimuths of the *i*-motif structures, because of the mismatch between AFM scanning direction and the molecular packing direction. The agreement between **A** and **B** confirms that the observed rhombic lattice results from the well-packed DNA molecules rather than AFM artifacts. Any AFM artifacts are supposed to be independent of the scanning direction. The arrows in **A** and **B** indicate the same direction with respect to the DNA lattice.

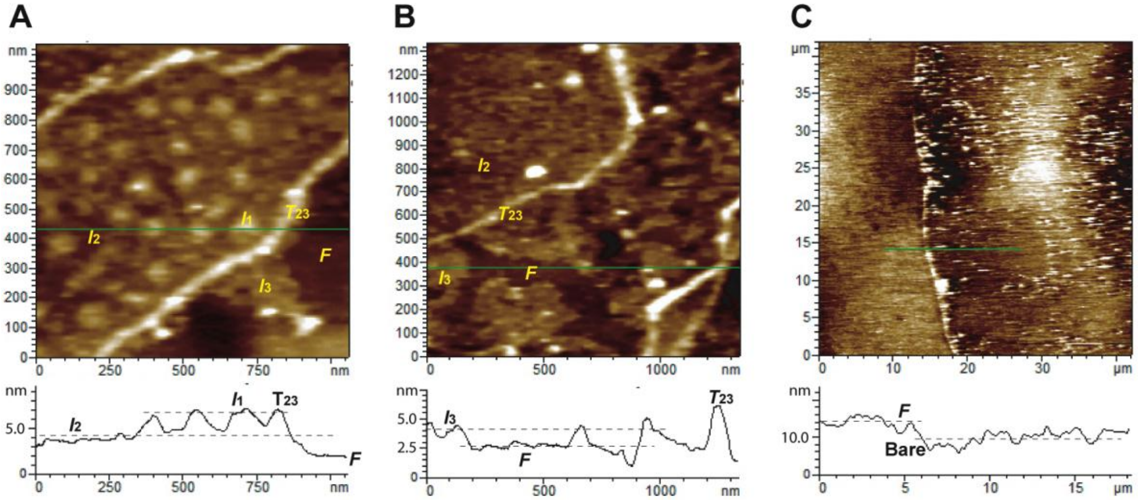


Figure S3. The topographic height measurement of the misfolding and folding intermediate and *F* states in AFM images. (**A**) The measurement of the topographic heights of I_1, I_2, and T_{23}. (**B**) The measurement of the topographic height of I_3. In **A** and **B**, the folding states in the local areas are marked according to the theoretical explanations. (**C**) The measurement of the topographic height of *F*. The DNA molecules in the right areas were disassembled and removed by directly scratching the AFM tip against the bare gold surface. Then the height of *F* was measured at the interface between the intact and scratched areas. In all images, the green lines in the upper panels mark the position and width of the measured height profiles. The lower panels show the corresponding height profiles. The topographic heights of T_{23}, I_1, I_2, I_3, and F were measured to be 10.0 ± 0.5, 9.5 ± 0.5, 6.5 ± 0.5, 6.5 ± 0.5, and 4.5 ± 0.5 nm, respectively. These results also indicate that the DNA molecules are standing up on the Au(111) surface, which is well consistent with the previous studies[21].

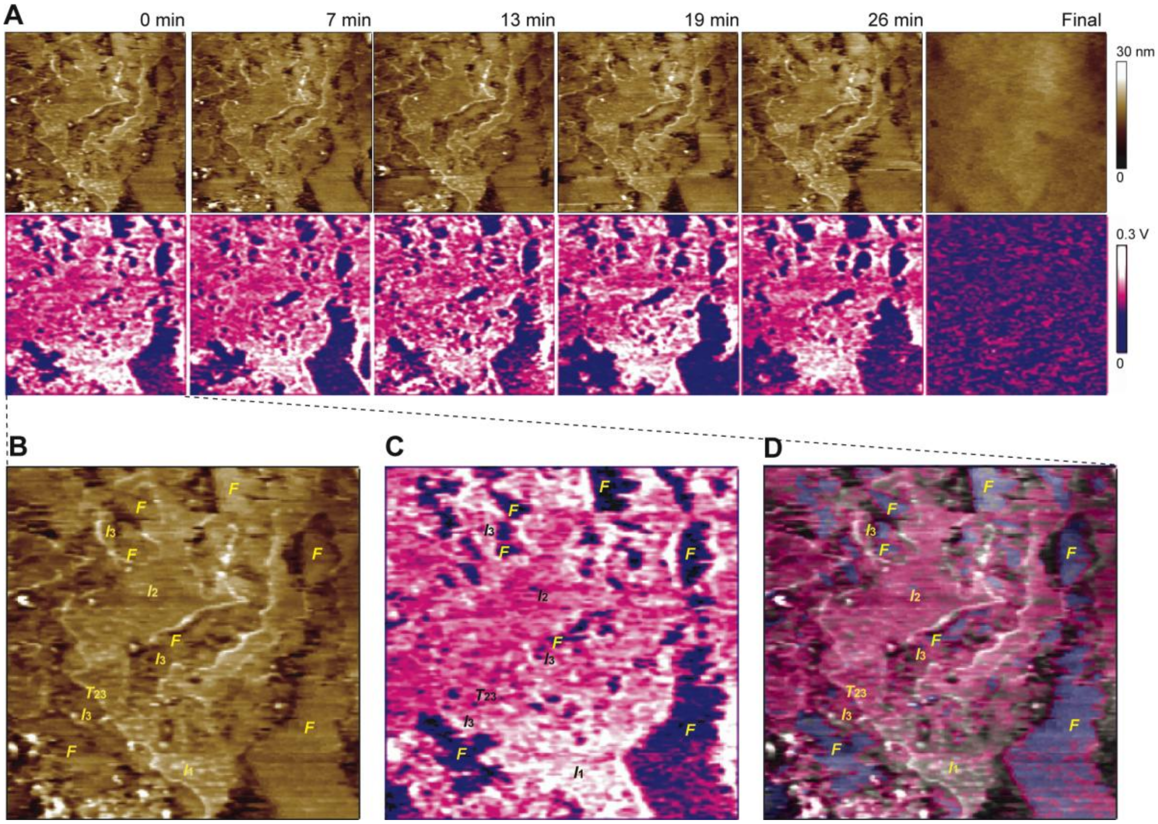


Figure S4. Analysis of the folding pathways between the intermediates and *F* state. (**A**) The consecutive AFM observation on the intermediate folding transition. The upper panels show the topographic images. The lower panels show the simultaneously obtained phase images. The last image in each row show the image of the same area after the folding transition was fully completed. Each image has a size of 5 × 5 μm^2. (**B**) and (**C**) respectively, show the topographic and phase images of the first column of **A** in a larger size and with each intermediate marked by the corresponding symbols. (**D**) The superposition of images **B** and **C**, for a better comparison between the topographic and phase images. The comparison shows that the *F* nuclei mostly emerge from the I_3 domains, whereas it is hard to definitely distinguish any *F* nucleus in the I_2 domains. The real-time observation shown in **A** further confirms that the *F* nuclei encapsulated inside I_3 domains grow up continuously. Because there is not much difference in stiffness between the intermediates I_2 and I_3, the phase signals show no difference between I_2 and I_3. However, the phase signals of the I_1 state are marginally distinguishable and are higher in voltage than I_2 and I_3.

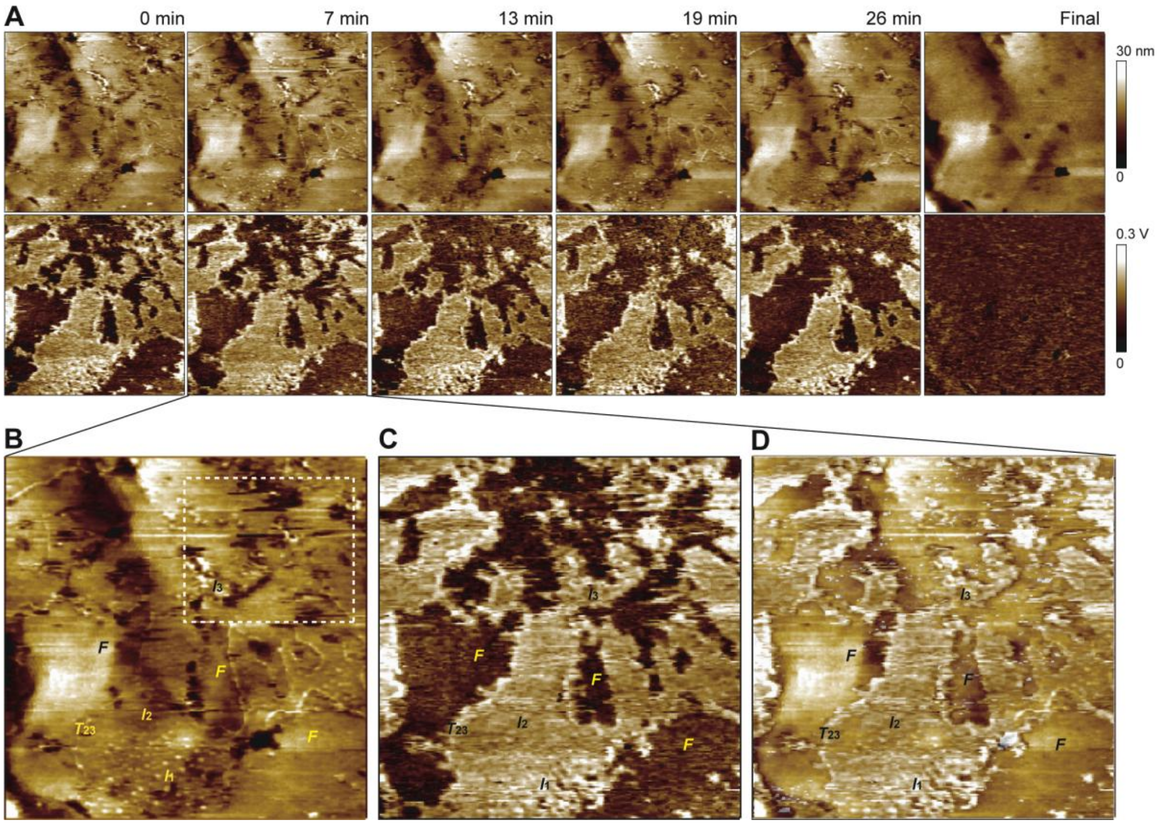


Figure S5. Evidence that the real-time topographic AFM imaging fails in imaging the mesoscopic coarsening of the intermediates and *F* domains. (**A**) The consecutive AFM observation on the intermediate folding transition. The upper panels show the topographic images. The lower panels show the simultaneously obtained phase images. The last image in each row show the image of the same area after the folding transition was fully completed. Each image has a size of 6 × 6 μm^2. The equilateral triangle or 60° angled orientations of the topographic dislocations is a signature of the Au(111) surface. The Au(111) dislocations are more clearly shown in the finally folded topography. Because the AFM phase imaging is only sensitive to stiffness changes, but not sensitive to topographic changes, the dislocations of Au(111) surface shown in the topographic images are not observed in the phase image. The phase images use the same colour bar as the topographic images, in order to demonstrate how much the imaging contract is enhanced. (**B**) and (**C**) respectively, show the topographic and phase images of the first column of **A** in a larger size and with each folding intermediate marked by the corresponding letters. (**D**) The superposition of images **B** and **C**, for a better comparison between the topographic and phase images. Although the phase imaging provides a clear-cut contrast between the folding and folded areas, we were unable to clearly track the microscopic coarsening behaviours of *F* domains from their corresponding topographic images. This results from that the dramatic change of the topographic height and the rapid, stochastic molecular motion during the folding transition make the *z*-axis feedback of the AFM cantilever fail in correctly predicting the topographic changes in local areas. This is clearly indicated in **B**, in which the areas under dramatic folding motion (marked by dashed box) show false information of the topographic height. However, because the phase imaging of the tapping-mode AFM is not sensitive to the topographic change, it can still capture the stiffness contrast between the *F* domains and those folding intermediates, thus being able to capture the detailed microscopic coarsening of *F* domains.

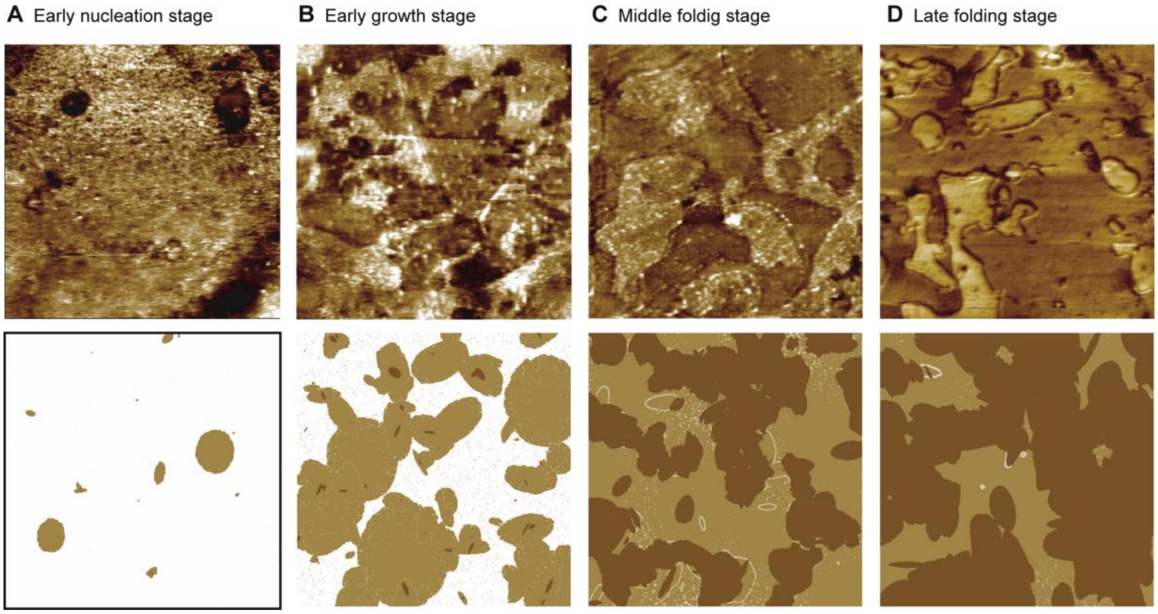


Figure S6. **Comparison between experiments and simulations on the topological features at the different stages of the folding transition**. The upper panels show the AFM topographic images; the lower panels show the computer-simulated topographic images. The size of each image is 5 × 5 μm^2. The image pairs of **A** – **D**, respectively, show four types of different topological features that represent the nucleation of semi-folded intermediates (**A**), the phase competition between off-pathway and on-pathway intermediates (**B**), the invading of *F* into the off-pathway intermediates (**C**), the final stage of transition to *F* (**D**). Note that the topology of *F* domains in the on-pathway intermediate is different from that in the off-pathway intermediate. The former shows much stronger anisotropy than the latter. This is well reproduced in the computer simulations. The simulations suggest that the reduced anisotropy of *F* domain growth against I_2 domains is a consequence of multiple state-transitions along the folding pathway ($I_2 \rightarrow T_{23} \rightarrow I_3 \rightarrow F$). The multistep changes of folding states slow down the folding transition in individuals and thus decrease the degree of the anisotropy. See Supplementary Movie S1 for more details.

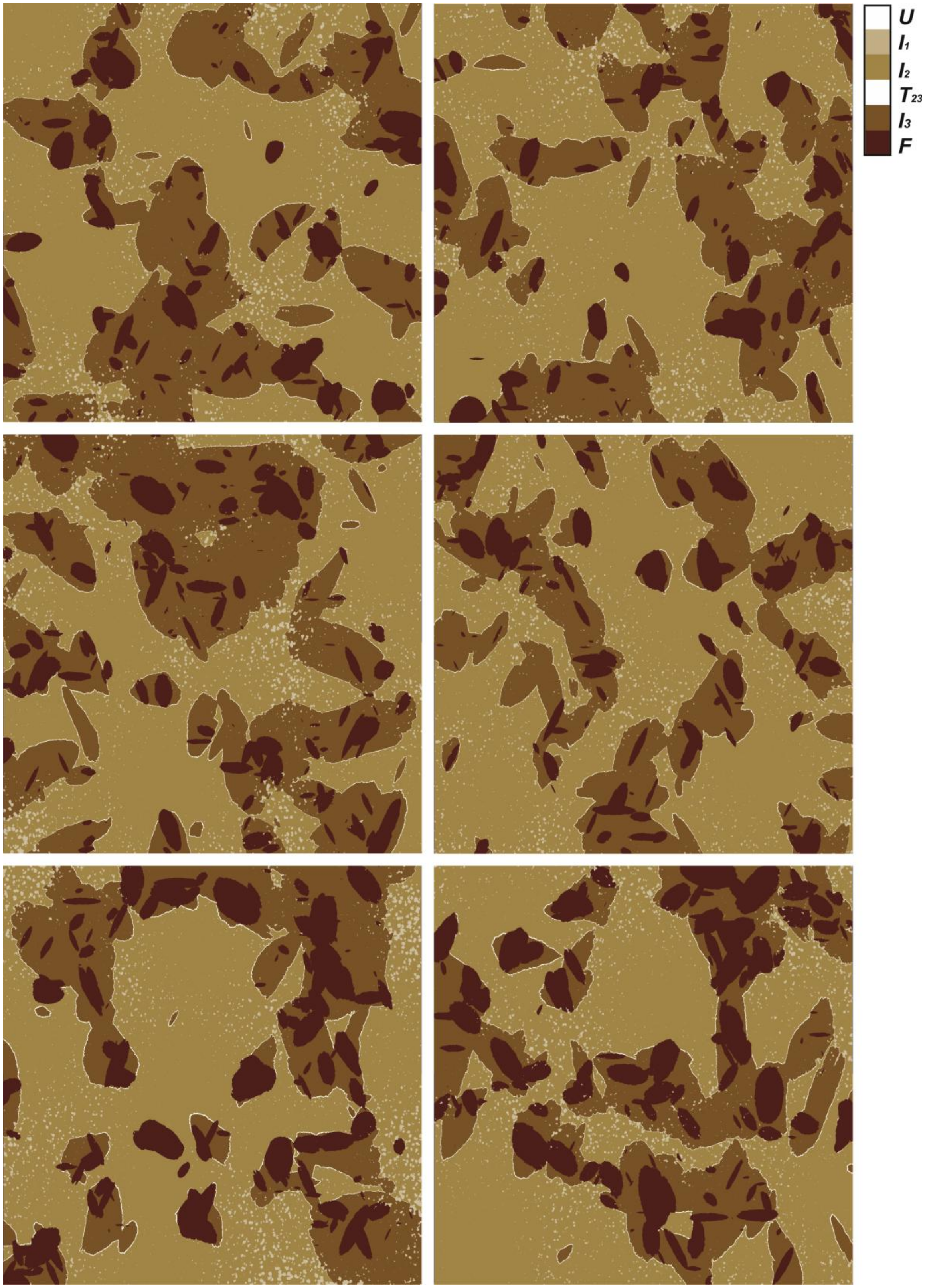


Figure S7. **A gallery of computer-simulated self-organization patterns arising from the misfolding and folding intermediates**. Each panel is a snapshot from an independent simulation. The diversity of the geometry results from the stochastic nature of the folding dynamics. The topological features observed in Fig. 2A of the paper and Figs S4 and S5 are well reproduced in these images. The size of each image is 8 × 8 μm^2.

Table S1. The physical features of the misfolding and folding states

State	(*P*, *W*, *Q*, *L*)	Theoretical height /Observed height (nm)	Topological features in AFM and simulated images	ΔG (kcal mol^{-1})
U	(0, 0, 0, 0)	10.2 /10.0 ± 1.0	Higher topographic fluctuation with surface roughness of 3~4 nm	0
I_1	(0, 0, 1–6, 0)	10.2 /10.0 ± 1.0	Flower-like extruding clusters with lateral sizes of 10-200 nm; mostly localized in I_2 domains	–4.0
I_2	(0, 0, 1–6, 1)	6.8 /6.0 ± 1.0	Lower topographic fluctuation with surface roughness of 1~2 nm; negative curvature interfacing I_3 and *F* domains	–9.0
I_3	(0, 1–6, 0, 1–2)	6.8 /6.0 ± 1.0	Lower topographic fluctuation with surface roughness of 1~2 nm; positive curvature interfacing I_2 domains and negative curvature interfacing *F* domains; anisotropic growth	–9.6
T_{23}	(0, 0, 1–3, 0–1)	10.2 /9.0 ± 1.0	Wire-like nano-walls at the interface between I_2 and I_3 regions, with widths of few molecules; quasi-static	–2.0
F	(6, 0, 0, 3)	4.9 /4.5 ± 1.0	Very small topographic fluctuation, showing rhombic lattice; positive curvature interfacing I_2 and I_3 domains; anisotropic growth	–20.8

Note: Each misfolding or folding intermediate represents an ensemble of conformations sharing the similar topological and thermodynamic properties listed herein. The coordinate (*P*, *W*, *Q*, *L*) represents the order parameters of a folding molecule, where *P* is the number of native intramolecular hydrogen bonds, *W* the number of non-native intramolecular hydrogen bonds, *Q* the number of intermolecular hydrogen bonds, and *L* the number of the hairpin loops. The values of Δ*G* are estimated from the strength of hydrogen bonds, conformational entropy, and electrostatic energy. The Δ*G* of the single *i*-motif structure is –19.15 kcal mol^{-1} (ref. 46–51). The Δ*G* of duplex formation of the same sequence is –8.03 kcal mol^{-1} (ref. 19). These provide the basis for the estimation of the Δ*G* of these above states. The Δ*G* of intermolecular electrostatic potential between neighbouring *i*-motif structures is estimated to be 0.09 kcal mol^{-1} per base pair .[42-45] This gives the intermolecular Δ*G* of –1.62 kcal mol^{-1} for each molecule at the *F* state, when considering six neighbours.

Movie S1 Legend

The complete folding transition from a computer simulation. The movie shows that the molecular population takes several folding pathways in individuals simultaneously. The major folding pathways shown in this movie substantiate the experimentally identified folding pathways summarized in Fig. 4D of the paper. The colours for the folding states are using the same colour bar shown in Fig. S7. The size of the folding area is 8 × 8 μm^2.