Corrections

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

The authors note that two protein names appeared incorrectly. All instances of “2out-knot” should instead appear as “2ouf-knot,” and all instances of “2out-ds” should instead appear as “2ouf-ds.” Both the online article and the print article have been corrected.

ENVIRONMENTAL SCIENCES

The authors note that additional GenBank accession numbers for all the mitochondrial sequences reported in their paper are JX475454–JX475467.

IMMUNOLOGY

The authors note that, due to a printer’s error, the statement in the Acknowledgments, “we thank the Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Development, for continued support and funding” should instead appear as “we thank the Division of Intramural Research, National Institute of Dental and Craniofacial Research, for continued support and funding.”

MEDICAL SCIENCES

The authors note that the Acknowledgments appeared incorrectly. The corrected Acknowledgments appear below.

ACKNOWLEDGMENTS. This work was funded by National Institutes of Health Grants R01 HL093234 (to S.M.P.), R01 AI058107 (to S.K.), U19 AI062629 (to S.K.), and U01 AI075386 (to S.K.), and the German Research Foundation (S.D.).

PHYsiOLOGY

The authors note that the author name Faisal Cheema should instead appear as Faisal H. Cheema. The corrected author line appears below. The online version has been corrected.

XiaoBo Gao, Dominique Eladari, Francoise Leivel, Ben Yi Tew, Cristina Miró-Julia, Faisal H. Cheema, Lance Miller, Raoul Nelson, Teodor G. Păunescu, Mary McKee, Dennis Brown, and Qais Al-Awqati
Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot

Wenfei Li1,2, Tsuyoshi Terakawa3, Wei Wang1, and Shoji Takada1,4

1National Laboratory of Solid State Microstructure and Department of Physics, Nanjing University, Nanjing 210093, China; 2Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan; and 3Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan

Edited by William A. Eaton, National Institutes of Health-NIDDK, Bethesda, MD, and approved June 4, 2012 (received for review March 14, 2012)

While fast folding of small proteins has been relatively well characterized by experiments and theories, much less is known for slow folding of larger proteins, for which recent experiments suggested quite complex and rich folding behaviors. Here, we address how the energy landscape theory can be applied to these slow folding reactions. Combining the perfect-funnel approximation with a multiscale method, we first extended our previous atomic-interaction based coarse grained (AICG) model to take into account local flexibility of protein molecules. Using this model, we then investigated the energy landscapes and folding routes of two proteins with complex topologies: a multidomain protein adenylate kinase (AKE) and a knotted protein 2ouf-knot. In the AKE folding, consistent with experimental results, the kinetic free energy surface showed several substates between the fully unfolded and native states. We characterized the structural features of these substates and transitions among them, finding temperature-dependent multiroute folding. For protein 2ouf-knot, we found that the improved atomic-interaction based coarse grained model can spontaneously tie a knot and fold the protein with a probability up to 96%. The computed folding rate of the knotted protein was much slower than that of its unknotted counterpart, in agreement with experimental findings. Similar to the AKE case, the 2ouf-knot folding exhibited several substates and transitions among them. Interestingly, we found a dead-end substate that lacks the knot, thus suggesting backtracking mechanisms.

In the last two decades, chemico-physical knowledge of protein folding has been greatly deepened primarily focusing on small and fast-folding proteins (1, 2). Many such proteins were identified as two-state folders and their transition state ensembles were experimentally characterized by site-directed mutagenesis via the ΔΦ-value analysis (3, 4). Ultrafast experimental technology enabled direct observation of fast-folding kinetics unambiguously (5–7). Single molecule experiments directly visualized cooperative two-state folding transitions (8–10). In parallel, statistical physics theory clarified that proteins have evolved their sequences to have globally funnel-like energy landscapes, providing a general framework of understanding (11, 12). Many simulations inspired by the theory were compared with experimental observations and gave feedback to the theory (13–17). In particular, structure-based models that can concisely realize perfect funnel landscapes of given native topologies showed great promise to capture many essential features for small and fast-folding proteins (13, 14, 18–22). Recently, a specialized computer enabled folding simulations of small proteins with an atomistic model, which were consistent with the two-state folding on funnel-like landscapes (23).

In contrast, much less is understood for slow folding of proteins with medium or large size (e.g., proteins larger than 150 residues.). Generally, the folding of such proteins is much more complex with more than two states. Do such proteins still have globally funnel-like energy landscapes? More specifically, can we use similar perfect funnel models to characterize such complex folding mechanisms of individual proteins? Here, we address these issues focusing on two specific cases: folding of a multidomain protein, adenylate kinase, and that of a designed knotted protein, 2ouf-knot.

Adenylate kinase (AKE) is a medium-sized and slow-folding protein with three domains. Very recently, its folding has been characterized by the high-throughput single molecule fluorescence resonance energy transfer (smFRET) (24). Long time smFRET trajectories were fed into the hidden Markov method to identify approximately six metastable states and state-to-state transition rates. Their analysis revealed that multiple intersecting folding routes coexist in the AKE folding, and denaturant concentration can modulate the probabilities of these folding routes. A somewhat similar combination of single molecule experiments and the learning algorithm were used to obtain a complex multiroute folding networks of calmodulin, a two-domain protein (25). Such complex folding schemes provide us an ideal opportunity to examine the view of funnel-like landscape for multidomain proteins.

The knotted proteins are another class of proteins in which the structures show complexity by making knots. About 10 knotted protein folds have been identified up to now (26, 27). Their folding involves the precise threading of a chain segment through a knotting loop and therefore poses new challenge to the current protein folding theories. Shakhnovich et al. simulated the folding of a pair of related knotted proteins, Yibk and YbeA, using a structure-based Go model (28) and suggested that addition of specific nonnative interactions can greatly facilitate knot formation and thus successful folding, which calls for some alternation in the funnel landscape. Later, Onuchic and coworkers found that the similar Go model, without nonnative interactions, is able to guide the folding of the knotted protein to the native state via slipknot intermediate (29), although the probability of successful folding was very low (≈1–2%) even after selectively strengthening/weakening the native interactions between certain residue pairs. Apparently, further studies are necessary. Recently, by connecting two monomers with a linker or by a disulfide bond, a knotted protein 2ouf-knot and its unknotted counterpart 2ouf-ds were designed (26). The 2ouf-knot and 2ouf-ds have almost identical sequence and tertiary structures and therefore provide ideal

Author contributions: W.L. and S.T. designed research; W.L. and S.T. performed research; W.L., T.T., and S.T. contributed new reagents/analytic tools; W.L., T.T., W.W., and S.T. analyzed data; and W.L., T.T., W.W., and S.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence may be addressed. E-mail: takada@biophys.kyoto-u.ac.jp or wangwei@nju.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201807109/-/DCSupplemental.
model systems to study the role of a knot in the protein folding and to identify the critical energetic factors contributing to the successful folding of the knotted proteins. Here, folding of this pair of proteins, 2ouf-knot and 2ouf-ds, were investigated by computer simulations.

To compare simulations vis-a-vis experiments for the complex folding of larger proteins, we need accurate yet concise simulation methodology. Here, while sticking on the globally funnel-like landscape view, we include sequence-specific and atomic interactions as much as possible. For the purpose, we previously developed an atomic-interaction-based coarse-grained (AICG) model for proteins (30), and here, we extend it to improve description of local flexibility (the extended model designed as AICG2).

In the present work, we first briefly describe the AICG2 model that takes into account sequence-specific local and nonlocal interactions and further optimizes the relative weights of them via a multiscale protocol. Using the AICG2 model, we studied the folding of AKE and a knotted protein 2ouf-knot. We were able to capture all the experimentally found major features in the folding of AKE and further provide structural characterization of and transitions among substates. For the knotted protein, AICG2 realized efficient folding of 2ouf-knot with a probability up to 96%.

Thus, the present study suggests that the perfect funnel model can capture essential features of the folding of these proteins with complex topologies once native interactions were carefully tuned and local flexibility was included.

Results and Discussions

Accurate yet Concise Modelling of Folding. Previously, we developed an AICG using a multiscale protocol (30). AICG uses one-bead-per-residue representation, but its interresidue nonlocal contact energies were derived from atomic interactions. Here, we extended AICG by introducing a flexible local potential (AICG2).

The local interactions are described by combination of generic statistical potentials of the residue types and local contact potentials that are based on the native structure. The relative weight and average strength of the local contact potentials were derived by a multiscale protocol, the same as for the nonlocal contact interactions. Incorporating sequence-specificity and the locally flexible nature of chains, AICG2 can reproduce folding energy landscape better than the simple Go model. As an illustration, the asymmetric folding pathways of the topologically symmetric proteins, protein G and protein L, were well reproduced (SI Appendix, Figs. S1 and S2).

High Free Energy Barrier and Substates in the AKE Folding. AKE is a 214-residue kinase that plays important roles in maintaining the balance of ATP and ADP in the cell and, more importantly here, it has been used as a model multidomain protein (31). AKE consists of three domains, i.e., LID domain, NMP domain, and CORE domain (31) (Fig. 1A). AKE folds relatively slowly and a recent single molecule experiment revealed complex multiroute folding (24).

We first tried to realize reversible folding and unfolding transitions using AICG2 model and unbiased molecular dynamics (MD) simulations, which turned out to be difficult due to its intrinsically slow folding. Then, we used the modified multicanonical method (32) for better conformational sampling. The converged one-dimensional free energy profiles near the folding temperature $T_F$ (Fig. 1B) show relatively high and broad free energy barrier, consistent with slow folding in vitro (33) and in silico. Albeit no apparent intermediate states in Fig. 1B, the two shoulders labelled by the arrows suggest that the folding/unfolding is not a simple two-state transition (see SI Appendix, Fig. S3). Some substates, virtually hidden in Fig. 1B, may exist between denatured and native states.

Because the high temperature simulations tend to hide the possible substates, we next conducted folding simulations starting from fully unfolded structures at the three temperatures, 0.90$T_F$, 0.92$T_F$, and 0.94$T_F$. A representative time course of formation of native contacts, i.e., so-called Q-score, at 0.90$T_F$ (Fig. 1C) shows successive transitions among different substates. Particularly, the CORE domain encounters at least five substates beside the fully unfolded and native states. Such transitions among multiple substates closely resemble smFRET trajectories observed (24), although directly mapping the substates by the experiment with those by the simulation is difficult due to the lack of structural information in the experiment.

To identify these substates robustly and statistically, we plotted the kinetic free energy landscape on the reaction coordinates $Q_{tot}$ and $Q_{core}$ at 0.90$T_F$ and its one-dimensional projections. The seven substates and completely unfolded (U) and native states (N) are labeled. The substate 1 may correspond to the denatured state (D).

![Fig. 1. Free energy surfaces and a representative trajectory of the folding of AKE.](image)

(A) Crystal structure of the AKE [Protein Data Bank (PDB): 4ake]. (B) Free energy profiles along $Q_{tot}$ at temperatures $T = 0.98T_F$, 0.99$T_F$, 1.00$T_F$, 1.01$T_F$, and 1.02$T_F$. (C) A representative folding trajectory monitored by $Q_{tot}$ (black), $Q_{LID}$ (gray), $Q_{NMP}$ (red), and $Q_{CORE}$ (green). (D) Two-dimensional kinetic free energy landscape on the reaction coordinates $Q_{tot}$ and $Q_{core}$ at 0.90$T_F$, and its one-dimensional projections. The seven substates and completely unfolded (U) and native states (N) are labeled. The substate 1 may correspond to the denatured state (D).

Structural Features of the Substates in the AKE Folding. An advantage of the MD simulation is the ability to provide detailed structural
Multiroute Folding of AKE and Its Modulation by Temperature. To characterize folding routes, we analyzed the transitions among the above defined substates, identifying five most probable folding routes (arrows of each color in Fig. 2 represent one route). In the figure, the breadth of the arrows represents the probabilities of the routes. (see also SI Appendix, Table S2 and Fig. S8 for more details). At 0.907$T_F$ (Fig. 2A), the dominant folding route is $U \rightarrow 1(D) \rightarrow 3 \rightarrow 4 \rightarrow N$ (green arrows). In this route, the LID domain folds first (or it is folded in the denatured state). Then the central $\beta$ sheet of the CORE domain folds (substate 3), which is followed by folding of the NMP domain (substate 4). At last, N- and C-terminal $\alpha$ helices of the CORE domain dock to make the native state. There was a slightly less probable alternative route, $U \rightarrow 1(D) \rightarrow 6 \rightarrow 7 \rightarrow N$ (black arrows), where the LID domain, which was folded in 1(D) and substate 3, transiently unfolded in substate 6, and 7.

This second route became dominant at higher temperatures, 0.92$T_F$ and 0.944$T_F$ (Fig. 2B and C). Thus, the AKE folding took heterogeneous routes, and the populations of the folding routes are sensitive to temperatures. These results are in agreement with the single molecule observations in which a few folding routes were identified and were found sensitive to the denaturant concentration. We note that although we can approximately model the effect of denaturant by high temperatures, the actual temperature dependence of folding can be very complicated as shown in a theoretical work (37). Both the equilibrium simulations around the folding temperature and the folding simulations at low temperatures showed that the LID domain frequently transits between the folded and unfolded states during the whole folding trajectory (Fig. 1C, SI Appendix, Fig. S9). As higher temperature tends to destabilize the LID domain (SI Appendix, Fig. S10), increasing the temperature promotes the folding route $U \rightarrow 1(D) \rightarrow 3 \rightarrow 6 \rightarrow 7 \rightarrow N$ as observed in Fig. 2. Interestingly, such multiroute folding was also observed in some simple proteins (38), and has been generally linked to topological complexity of the folded state and the structure symmetry (39).

Slow Folding of Knotted Protein 2ouf-knot. We move to a protein with another type of topological complexity, namely, knot formation. Here we focus on a designed knotted protein, 2ouf-knot, and its unknotted counterpart, 2ouf-ds, which have almost identical sequences and tertiary structures, thus being ideal model systems to study the role of a knot in folding (Fig. 3A and B). These proteins were designed based on a natural dimeric protein, in which each subunit has four helices. Connecting two subunits with a nine-residue loop (L1 in Fig. 3A) led to 2ouf-knot, with left-handed trefoil knot of 28-residue depth in N terminus and 57-residue depth in C terminus. In contrast, the two subunits were connected with a disulfide bond to make 2ouf-ds that lacks a knot. Experimentally, both proteins reversibly folded, but 2ouf-knot folded an order of magnitude slower than 2ouf-ds (26).

Here, using AICG2, we conducted extensive unbiased MD simulations at a wide temperature range for both 2ouf-knot and 2ouf-ds. In simulations, both proteins exhibited reversible folding and unfolding transitions (Fig. 3C and D) and we confirmed that the knot is indeed formed for 2ouf-knot. Consistent with the experiment, the folding of the knotted one was much slower than the unknotted counterpart (Its folding rate is nearly seven times slower than that of the 2ouf-ds). The higher free energy barrier of 2ouf-knot in the one-dimensional free energy profile (Fig. 3E) also indicates slower folding of the knotted protein than that of 2ouf-ds. Apparently, such slowing of 2ouf-knot folding purely results from the knot because it is the only major difference between 2ouf-knot and 2ouf-ds.

Interaction Heterogeneity and Chain Flexibility Promote the Knot Formation. The reversible folding/unfolding transitions in Fig. 3C surprised us because previous folding simulations of another knotted protein with a Go model showed quite a low ratio of successful knot formation (29). To quantify the ratio of successful folding, we conducted extensive independent folding simulations in a range of temperatures below $T_F$. The successful folding ratio

Fig. 2. Folding routes and structural features of the substates of AKE. (A) Folding routes and residue-resolved contact scores of every substate at $T = 0.907T_F$. Color code: red, unstructured; blue: fully structured. The five most probable folding routes that reach the native state are represented by colored arrows with the line widths representing the abundance of each folding route. Different colors represent different folding routes. (B, C) Same as A but at $T = 0.92T_F$ (B) and 0.947$T_F$ (C) with the substates represented by circled numbers.
can be improved to a large extent (29), indicating the importance of the sequence specific interactions. Yet, it is of some surprise because 2ouf-knot is a designed protein and has not evolved to achieve folding ability.

**Folding of the 2ouf-knot Is More Topologically Frustrated Than Its Unknotted Counterpart.** To compare folding mechanisms of knotted and unknotted proteins, we next drew a number of two-dimensional free energy surfaces (Fig. 4, SI Appendix, Figs. S13 and S14). In Fig. 4, although the overall shapes of the free energy surfaces of two proteins show some similarity, the knotted one possesses some unique features. In early stages of folding ($Q_{\text{tot}} \sim 0.2$ and $Q_{\text{tot}} \sim 0.5$), the H1-H4 contacts can be formed partly (Fig. 4A). However, because the formation of such H1-H4 contacts hinders the knot formation, they need to be transiently broken up ($Q_{\text{tot}} \sim 0.35$ and $Q_{\text{tot}} \sim 0.6$). Thus, the knot formation introduces topological frustrations, which was also observed in the folding of other proteins with complex topology (32). In contrast, the free energy landscape of 2ouf-ds is much smoother (Fig. 4B). The folding order of the H4-H6 contacts also has large differences. For the 2ouf-knot, the H4-H6 contacts are predominantly formed early on (Fig. 4C), whereas for the 2ouf-ds, the H4-H6 contacts can be formed in a later stage with certain probability (Fig. 4D). In addition, for the 2ouf-ds, the folding of the two halves is symmetric due to the symmetry in sequences and the tertiary structures (SI Appendix, Fig. S14). In comparison, for the 2ouf-knot, the pathway with the H1-H4 contacts form before the H5-H8 contact formation is much more probable. Such asymmetry in folding routes is crucial for the efficient folding of the knot, which will be discussed in more detail below.

**Structural Features of the Substates in 2ouf-knot Folding.** Similar to the case of AKE, we conducted extensive folding simulations of 2ouf-knot starting from fully unfolded structures at the three temperatures, $0.90T_F$, $0.925T_F$, and $0.95T_F$. Fig. 5 A and B shows two representative trajectories. The stepwise folding of different parts of the 2ouf-knot implies the existence of several substates. We plotted the kinetic free energy landscape projected onto the reaction coordinate $Q_{\text{tot}}$ and rmsd in Fig. 5C. Six substates besides the unfolded state and the native state were

---

**Fig. 3.** Folding of a designed knotted protein 2ouf-knot and its unknotted counterpart 2ouf-ds (A, B). Crystal structures of 2ouf-knot (PDB: 3mlg) (A) and 2ouf-ds (PDB: 3mli) (B). The eight helices are labeled by H1, H2, ... H8, and the loop is labeled by L1. Two cysteines forming the disulfide bond in 2ouf-ds are explicitly shown. (C, D) Time courses of the Q score of the entire protein for the knotted protein (C) and the unknotted protein (D) at the folding temperature $T = T_F$. (E) One-dimensional free energy profiles for the knotted (solid line) and unknotted (dashed line) proteins at $T = 0.98T_F$, 1.00$T_F$, and 1.02$T_F$. (F) The ratios of the successful folding with four different models when started from fully unfolded structures.

---

**Fig. 4.** Two-dimensional free energy surfaces of knotted (A, C) and unknotted (B, D) proteins on the reaction coordinates $Q_{\text{tot}}$ and $Q_{\text{H1-H4}}$ (A, B) and on coordinates $Q_{\text{tot}}$ and $Q_{\text{H4-H6}}$ (C, D) at $T_F$. The unit of the free energy is $k_BT$. 
Folding of Two Other Knotted Proteins. Because folding mechanisms of natural proteins and designed proteins could be very different as clarified in ref. 40, we also conducted the folding simulations of zouf-knot. In this case, the folding routes of zouf-knot at 0.90T_F, 0.925T_F (Fig. 6B), and 0.95T_F (Fig. 6C) are analyzed. More detailed analysis is given in SI Appendix, Table S3 and Fig. S16. Not surprisingly, multiple folding routes and structures with high probabilities, suggesting the heterogeneity of the zouf-knot folding routes. At 0.90T_F, the dominant folding route U → 1 → 2 → 3 → 4 → N is the most sequential one. The major feature of this folding route is that the knot formation is not just the threading of H1 itself, but it is accompanied by high flexibility of L1 and H5. Thus, we call it “concerted threading.” Detailed analysis shows that in all the trajectories with this folding route at the temperature of 0.90T_F, the knot forms by the concerted threading mechanism. In comparison, if the pathway involves the substate 6, the knot mostly forms via a slipknot intermediate (Fig. 5B, Movie S2). For example, among the trajectories with the folding route U → 1 → 2 → 6 → N, which is the most probable route via substate 6, approximately 80% of the trajectories follow the “slipknot intermediate” knotting mechanism. This case, the H5 is well packed to H5 before the knot formation (substate 6), which makes the L1 and H5 well fixed. The rigid knotting loop makes the direct threading of H1 difficult. Consequently, the formation of the knot by the slipknot intermediate is preferred (Fig. 5B, Movie S2). Such results suggest that the knotting mechanism depends on the folding order of other parts of the protein. Notably, no route leading to the folded state via the substate 5, which requires the threading of the much longer C-terminal segments (H8 and H9) through a different knotting mechanism (H2-H3-H4-L1), is observed. We found that most routes involving substate 5 terminate at the substate 5, which is a dead-end trap.

In Fig. 6, we also observe some shortcuts of the folding routes, e.g., U → 1 → 2 → N in which the placements of the N-terminal segment (H1) and the C-terminal segments (H7 and H8) are cooperative. These shortcut routes become more probable at higher temperatures, perhaps because higher temperature usually tends to make the folding transition more cooperative.
for two natural proteins with trefoil knot, i.e., MJ0366 and YibK (28, 29, 41), which we briefly describe here. The trefoil knot of YibK is much deeper than those of the 2ouf-knot and MJ0366 (SI Appendix, Figs. S17–S19). For the MJ0366, we found that the maximal ratio of successful folding with AICG2 was 99%, a much higher ratio than that with the simple Go model (50%). However, the maximal ratios of successful folding for YibK by both simple Go model and AICG2 were very low (0.1–0.2%), which may suggest that in addition to the interaction ingredients included in the current AICG2, more realistic interaction features, e.g., non-native interactions (28, 40), need to be considered in order to achieve efficient folding of YibK.

**Material and Methods**

**Reaction Coordinates and Data Analysis.** In describing the folding of AKE, the $Q_{\text{core}}$, $Q_{\text{LID}}$, and $Q_{\text{NMB}}$, defined as the fraction of native contacts for the full-length protein, the CORE domain, the LID domain, and the NMP domain, respectively, were used as reaction coordinates. Similarly, in describing the folding of 2ouf-knot and its unknotted counterpart, besides $Q_{\text{core}}$, we used $Q_{\text{LID1}}$, $Q_{\text{LID2}}$, $Q_{\text{HIN}}$, $Q_{\text{HIV}}$, and $Q_{\text{HIV1}}$, defined as the fraction of native contacts between the corresponding segments. We also used the rmsd of $C_{\alpha}$ atoms. The Koniaris–Muthukumar–Taylor algorithm was used in detecting the knot (42). In constructing the folding routes, the snapshots of the folding trajectories were assigned to the substrates according to the values of the reaction coordinates. See SI Appendix for details.

**Simulation Details.** We used CafeMol (43) developed in our laboratory for general CG simulations (44). The all-atom MD simulations were conducted by AMBER 11 (45). The protein structures were visualized by VMD software (46). See SI Appendix for details.

**ACKNOWLEDGMENTS.** This work was based on a joint project between the Japan Society for the Promotion of Science and the National Natural Science Foundation of China, partly supported by Grant-in-Aid for Scientific Research on Innovative Areas “Molecular Science of Fluctuations Toward Biological Functions,” by Research and Development of the Next-Generation Integrated Simulation of Living Matter of the Ministry of Education, Culture, Sports, Science, and Technology, by the National Natural Science Foundation of China (Grants Nos. 11174134, 91127026, and 10834002), and by a project funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.