Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the Rop dimer-folding mystery

Yaakov Levy*†, Samuel S. Cho*‡, Tongye Shen*‡, José N. Onuchic*†, and Peter G. Wolynes*†‡§

*Center for Theoretical Biological Physics, and Departments of †Physics and ‡Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Contributed by Peter G. Wolynes, December 30, 2004

Protein folding has become one of the best understood biochemical reactions from a kinetic viewpoint. The funneled energy landscape, a consequence of the minimal frustration achieved by evolution in sequences, explains how most proteins fold efficiently and robustly to their functional structure and allows robust prediction of folding kinetics. The folding of Rop (repressor of primer) dimer is exceptional because some of its mutants with a redesigned hydrophobic core both fold and unfold much faster than the WT protein, which seems to conflict with a simple funneled energy landscape for which topology mainly determines the kinetics. We propose that the mystery of Rop folding can be unraveled by assuming a double-funneled energy landscape on which there are two basins that correspond to distinct but related topological structures. Because of the near symmetry of the molecule, mutations can cause a conformational switch to a nearly degenerate yet distinct topology or lead to a mixture of both topologies. The topology predicted to have the lower free-energy barrier height for folding was further found by all-atom modeling to give a better structural fit for those mutants with the extreme folding and unfolding rates. Thus, the non-Hammond effects can be understood within energy-landscape theory if there are in fact two different but nearly degenerate structures for Rop. Mutations in symmetric and regular structures may give rise to frustration and thus result in degeneracy.

Understanding the evolutionary capacity of proteins to overcome changes in sequence to retain their three-dimensional structures and function is an alternative way to formulate the protein-folding problem. The funneled energy landscape that arises to tolerate mutation(s) gives kinetically competent folders as long as stability is maintained (1, 2). The general experience from structural studies of mutant proteins with single substitutions is that the effect of mutations is usually rather minor and localized in the structure. Several proteins have even been shown to retain their global tertiary structure despite extensive redesign of their hydrophobic cores (3). The large sequence space giving a single protein topology is well illustrated by the huge families of structurally related proteins, sometimes found even with no discernible sequence homology (4–6), thus raising the question of possible convergent evolution. Structurally homologous proteins also usually show a conserved folding mechanism (3, 7), as would be expected for a simple funneled landscape. Proteins with similar native structures usually will have similar structures of the transition state. Thus, variations in their folding rates usually correlate with native-state stability (3, 8). On funneled landscapes, mechanistic intermediates and barriers arise from a tradeoff of chain entropy and stabilization energy that is reflected in native protein topology (1, 2, 9). As we shall see, the power of evolution to achieve a funneled landscape may be limited by the physics of symmetry (10).

This article focuses on the Rop protein, which, experiments show, displays anomalous mutational effects on its folding kinetics and on its structure. Prima facie, the anomalous folding behavior of Rop cannot be readily explained by the funneled energy-landscape concept. As we shall see, however, Rop may represent "the exception that proves the rule" for energy-landscape theory. Most critically, Rop is a homodimer. Its function is to bind two RNA hairpins in a key step regulating the replication of ColE1 plasmid in *Escherichia coli*. After association, both monomers, each of which is unstructured in the free form, adopt a helix–turn–helix structure. Together they define an antiparallel coiled-coil four-helix bundle (11, 12). Rop has been elegantly studied in the Regan laboratory (13–15) by systematically redesigning its hydrophobic core by mutating the residues in the "a" and "d" positions of the heptad repeats in its eight stacked layers. Several Rop variants were designed to differ in the number of mutated layers and their positions to investigate the effect of core packing perturbations on the stability of the protein and its folding kinetics. The 17 mutants of Rop show different core packing, ranging from what is termed an "underpacked" core (i.e., very low dimer stability) when all core residues are alanine to an "overpacked" core (i.e., tight interface) when all core residues were replaced by leucine (see Table 1 for a list of Rop variants and more details on the design scheme).

The extreme response of Rop to mutations is exemplified by there being two alternate crystal structures known for its mutants. In one case, Ala-31, which is located in the turn between the helices of each monomer, was replaced by a proline residue (A31P; mutant 13 in Table 1). The result is a dramatic conformational change that would not be expected from a single amino acid substitution. The structure of the dimer is changed from an *anti* topology into a *bisecting U* topology (Fig. 1) (16). This structure was then suggested to be actually a molten globule based on its thermodynamic properties (e.g., low stability and reduced ellipticity) as well as its fluctuations in molecular dynamic simulations (17). Unexpectedly, the A31P mutant displays some binding ability to the RNA target of Rop, presumably because helices 1 and 1′ remain adjacent as in the *anti* topology of the WT Rop. In the second case, redesigning the dimer to have two alanine and two isoleucine residues in each of the six central layers of the dimer interface (Ala_2Ile_2-6) results in a stable and highly α-helical dimer but with no ability to bind the RNA target of Rop. The loss of activity of Ala_2Ile_2-6 is surprising because Ala_2Leu_2-6, which differs from Ala_2Ile_2-6 only by having leucine in the "d" positions instead of isoleucine, does in fact show Rop activity. The crystal structure of Ala_2Ile_2-6 shows that this mutant adopts the *syn* topology (Fig. 1), obtained by a 180° flip of one monomer around an axis normal to the dimer interface. This reorientation of the two monomers splits the face formed by helices 1 and 1′, which is essential for RNA binding. The *anti* and *syn* forms of Rop dimer correspond to two different basins on the energy landscape (Fig. 1).

Beyond these dramatic structural changes, the folding kinetics of Rop mutants offers mysteries as well. The kinetic issues are observed mainly in the mutants comprising class I, which include two alanine and two leucine residues in a redesigned layer of the

Abbreviations: TSE, transition-state ensemble; rmsd, rms deviation.

§To whom correspondence should be addressed. E-mail: pwolynes@chem.ucsd.edu.

BIOPHYSICS

Table 1. The properties of the WT and mutant Rop dimer

Class	No.	Mutant	*In vitro* activity	*In vivo* activity	T_m, °C	ΔG^o, kcal/mol	Relative k_f	Relative k_u	Structure (method)
I	1	WT	Y	Y	64	−7.7	1	1	*anti* (X-ray, NMR)
	2	Ala_2Leu_2-4	Y	P	68	−5.8	1.5	28	*anti* (*in vitro* activity)
	3	Ala_2Leu_2-2	Y	Y	72	−7.7	3.2	18	*anti* (*in vitro* activity)
	4	Ala_2Leu_2-3+6	Y	—	72	−8.4	7.5	8.3	*anti* (*in vitro* activity)
	5	Leu_2Ala_2-2+7	Y	—	85	−12.8	10	18	*anti* (*in vitro* activity)
	6	Ala_2Leu_2-6-rev	Y	—	85	−10.3	85	6.7×10^2	*anti* (*in vitro* activity)
	7	Ala_2Leu_2-8-rev	Y	N	91	−9.9	92	2.7×10^3	*anti* (*in vitro* activity)
	8	Ala_2Leu_2-2+7	Y	—	85	−8.7	120	7.1×10^3	*anti* (*in vitro* activity)
	9	Ala_2Leu_2-1+8	Y	—	54	−6.3	160	1.1×10^2	*anti* (*in vitro* activity)
	10	Ala_2Leu_2-6	Y	N	82	−8.1	310	3.1×10^4	*anti* (*in vitro* activity)
	11	Ala_2Leu_2-8	Y	N	91	−7.5	610	5.0×10^4	*anti* (*in vitro* activity)
II	12	Ala31Pro	P	—	—	—	—	—	*bisecting U* (x-ray)
III	13	Ala_2Ile_2-6	N	N	83	−5.1	—	—	*syn* (x-ray)
	14	Leu_2Ala_2-8	N	—	—	−12.8	—	—	—
	15	Ala_2Met_2-8	N	N	48	−3.1	—	—	—
IV	16	Ala_2Val_2-8	—	—	—	—	—	—	—
	17	Ala_4-8	—	—	<2	—	—	—	—
V	18	Leu_4-8	N	N	—	—	—	—	—

The Rop variants differ in the number of the mutated layers and their positions. Most of the mutants are designed to have two residues with small side chains (as ''a'' residues) and two residues with larger side chains (as ''d'' residues). A set of mutants was designed in which two, four, six, or eight of the layers of the hydrophobic core were replaced by layers containing alanine at the ''a'' positions and leucine at the ''d'' positions. In other cases isoleucine, valine, and methionine were used to introduce a large side chain into the hydrophobic core instead of leucine. The antiparallel packing of the Rop monomers dictates symmetrical pattern of redesign of the core. Accordingly, redesigning layer 1 has to be accompanied with the redesign of layer 8, the same rule applies between layers 2 and 7 and layers 3 and 6. Each of the mutants is named according to the identity of the residues at the ''a'' and ''d'' positions of the repacked layers: for example, Ala_2Leu_2-6 has the six central layers repacked with alanine in the ''a'' positions and leucine in the ''d'' positions. The ''rev'' suffix refers to cases in which layers 2 and 7 have reversed pattern of packing (i.e., the small and large residues are at the ''d'' and ''a'' positions, respectively) to mimic their packing in the WT dimer. The Rop variants can be classified into five classes based on their folding thermodynamics, binding activity, the dimer topology, and their folding kinetics (these classes are similar to those defined in ref. 14). Based on the *in vitro* activity, it was concluded previously that all the mutants in class I have the *anti* topology. The folding rates of the Rop variants were measured at the same final fraction folded or unfolded. Class II contains the A31P mutant of Rop dimer that adopts the *bisecting U* topology. Class III is comprised of mutants that are highly α-helical; however, they completely lost their ability to bind RNA. The structure of Ala_2Ile_2-6 is the *syn* topology. The mutants in classes IV and V are less stable than the WT, and they do not bind RNA. Class IV is comprised of proteins which are underpacked (only Ala_2Met_2-8 forms dimer), and Leu_4-8 of class V is an overpacked protein that was suggested as forming a tetramer (14). The data on WT Rop as well as on the mutants of class I were taken from ref. 15, and the data on the mutant Ala_2Ile_2-6 were taken from ref. 34. For the other mutants, the data were taken from ref. 14. Y, Rop protein that binds RNA; P, partial active proteins; N, no activity; —, no experimental data are reported.

core and differ in the number and positions of the mutated layers (the folding kinetics of the mutants in classes II–V were not studied). These mutants have CD spectra similar to the WT Rop, and all show cooperative thermal denaturation. They also all show *in vitro* binding affinity comparable to WT Rop, an observation that supports the idea that all the mutants share a structure similar to the WT. However, it is problematic and most interesting that these mutants fold and unfold faster than the WT Rop. The increases in forward and backward rates depend on the number and position of repacked layers within the core, and for the mutant with all eight layers repacked (Ala_2Leu_2-8), the folding and unfolding is accelerated by more than two and four orders of magnitude, respectively (15). The thermal stability of these mutants is decoupled from this chemical stability (15). Adding to the puzzle, an *in vivo* assay for the Rop activity indicates that among the 10 mutants in class I, only the variant Ala_2Leu_2-2 is active *in vivo* and Ala_2Leu_2-4 is partially active (18). One may note that the folding and unfolding rates of these two mutants are similar to those for the WT Rop.

In this article we propose that the Rop mystery can be unraveled by assuming a double-funneled energy landscape for the Rop protein. The Rop mutants, which show different folding kinetics, adopt topologically different ground-state structures or at least have topologies that act as significant traps. This hypothesis is supported, of course, by the discrepancy between the *in vitro* and the *in vivo* binding activity of Rop mutants and the existence of variant crystal structures for other Rop-related sequences. Because of the symmetric sequences of these molecules, the *anti* and *syn* topologies have very similar core packing. They differ mainly in the relative position of the turns. The *bisecting U* topology is unlikely to be a strongly competing topology with the *anti* and *syn* topologies because of its poor packing and molten-globule properties, yet it can act as a high-energy intermediate in the folding of the other topologies. We thus suggest that for the redesigned sequences of Rop, mostly with repacking of the core edges, a switch of the Rop structure from *anti* into *syn* topology can and does occur. For other redesigned Rop proteins with moderate repacking of the core, these two topologies can be degenerate. As we shall see, energy-landscape theory suggests a lower free-energy barrier for adopting the *syn* topology in comparison to the *anti* topology. Thus, we argue that the mystery is resolved if the Rop mutants with the fast folding and unfolding adopt the *syn* topology and not the *anti* topology. This hypothesis would also explain the loss of *in vivo* activity for the fast-folding mutants.

Free-Energy Barrier for Rop Association: Folding to *anti* or *syn* Structural Topologies

Native Topology-Based Simulations. Theory, experiment, and simulation agree that protein topology is a key determinant of the speed of protein folding (19, 20) and a good predictor of the detailed protein-binding mechanism. The role of topology in binding was illustrated recently by native topology-based (Go) simulations accounting only for native contacts. Such simulations use a perfectly funneled energy landscape. These simulations reproduced the binding mechanism of many homooligomeric protein complexes by accurately predicting whether a folded monomer is prerequisite for their association (21–23). They also do a good job reproducing the

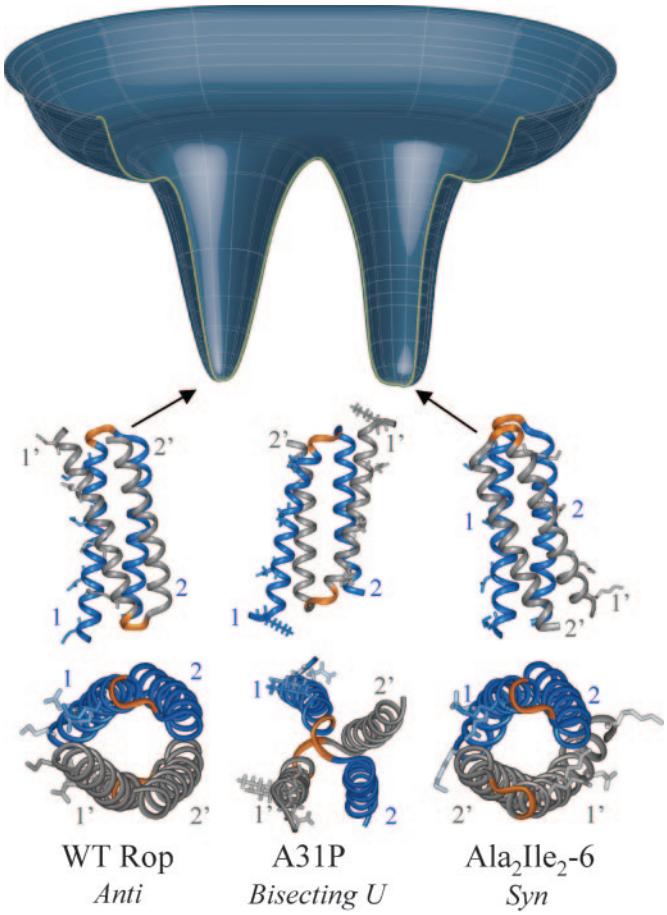

Fig. 1. A schematic representation of a double-funneled energy landscape of Rop dimer and its three structural topologies that correspond to the WT sequence and the mutants Ala_2Ile_2-6 and A31P. In these structures, one monomer is colored gray, and the other monomer is colored blue. The loop between the two helices in each monomer is colored orange. Residues Lys-3, Asn-10, Gln-18, and Lys-25 in helices 1 and 1′, which constitute the binding site to the RNA (39), are shown by stick representation. The structure of WT Rop dimer (11) and the Ala_2Ile_2-6 mutant (34) is a four-helix coiled coil and has been determined by x-ray crystallography [the structure of the WT Rop was also assigned by NMR (12)]. The WT Rop has an *anti* topology and the Ala_2Ile_2-6 has a *syn* topology, which is obtained by a 180° flip of one monomer around an axis normal to the dimer interface. The mutation Ala-31 into Pro is located in the loop and introduces a conformational change into a *bisecting U* topology (16). The *anti* and *syn* topologies are nearly degenerate and correspond to the two basins. The *bisecting U* is expected to be less stable than the *anti* and *syn* topologies and was not placed on the double-funneled energy landscape because its folding was not studied here.

Φ values characteristic of the binding transition state (24). One may ask whether, on the basis of such models, the *anti*, *syn*, and *bisecting U* topologies of Rop homodimer have different free-energy barrier heights and thus different kinetics of formation through folding/binding. The *anti* and *syn* topologies were simulated by using a perfectly funneled landscape model based on the crystal structures of the WT Rop and the Ala_2Ile_2-6 mutant, respectively. The structures of other Rop mutants have not yet been determined to our knowledge. The simulation protocol of the folding and association of Rop proteins is similar to the one used by us earlier in the studies of other protein complexes (21, 23, 24).

The free-energy surfaces for folding to *anti* or *syn* topologies of Rop dimer were projected onto three reaction coordinates: two corresponding to the folding of each monomer and the third corresponding to association (Fig. 2). The projected free-energy surfaces for both structures show coupling between monomer folding and association. Although these plots present similar mechanisms for forming either the *anti* or *syn* topologies, they give different binding transition-state ensembles (TSEs). The binding TSE for the *syn* topology has a lower free energy than does that for the *anti* topology. Plotting the free energy as a function of Q (the fraction of native contacts, which was shown previously to be successful in describing the structure of the TSE of both folding and binding reactions) in Fig. 2*C* shows that the *syn* barrier is lower by ≈1.6 kcal/mol (1 cal = 4.18 J) than the *anti* barrier at their respective folding temperatures in a model based on pairwise additive interactions. To quantify directly the effect of the barrier height on the folding kinetics, >1,000 folding and unfolding events were sampled for each structure. Fitting the kinetics of both folding and unfolding processes gave single exponential decays. The relative rates are shown in Fig. 2*C*, illustrating the faster folding and unfolding rates for the *syn* conformation.

The higher barrier for the *anti* topology and its slower kinetics at T_f indicate an intrinsic difference between the folding mechanisms of these two structures. A Go model, governed by native pairwise interactions, generally underestimates the barrier height because of their lack of nonadditive forces (25). Nonadditivity arises from solvent contribution or side-chain-reorientation effects. The nonadditivity corrections to the barrier heights for folding to the two structures were calculated by introducing such nonadditivity as a perturbation to the pairwise additive Go model. Plotkin and coworkers (26) showed that these perturbations yield higher barriers that more closely fit experiment. The Rop dimer corrections not only change the absolute barrier heights but also increase the relative rate differences. The energy of each sampled conformation is recalculated by using triplet interactions perturbatively in addition to the pairwise interactions. In the native state, the energy contributions of the two- and three-body interactions are $\varepsilon_2 N_2$ and $\varepsilon_3 N_3$, respectively, where N_2 and N_3 are the number of two- and three-body interactions in its native state. For a given conformation, the energies are as $E_2 = \varepsilon_2 Q_2 N_2$ and $E_3 = \varepsilon_3 Q_3 N_3$, where Q_2 and Q_3 are the fractions of native pairwise and triplet interactions in that conformation (ε_3 is assigned as $\varepsilon_2 N_2/N_3$ to preserve overall native stability). The total nonbonded energy of a conformation now is $E(\alpha) = (1 - \alpha)E_2 + \alpha E_3$; α controls the relative contribution of the two- and three-body interactions. The barrier for folding of the *anti* and *syn* topological structures as a function of α is shown in Fig. 2*D*. The barrier heights for both folding reactions increase with α; however, the increase for the *anti* topology is much sharper and is sufficient for $\alpha = 0.4$ to account for the difference between the fastest and slowest mutants.

Φ Value Analysis for the Folding TSE of *anti* and *syn* Topologies. The degree of structure formation in the folding TSE sheds light on the origin of the smaller *syn* barrier. The structure of the TSE is usually quantified by using Φ values (27, 28). Experimentally, a Φ value for a given residue is the ratio of the effect that a mutation at that position has on the stability of the TSE over its effect on the folded state stability, both relative to the denatured ensemble. A Φ value close to 1 means that the mutation similarly affects the thermodynamics of the TSE and the folded state, suggesting that the mutated residue is analogously structured in the TSE as it is in the folded state. Conversely, a Φ value close to 0 means that the mutation does not affect the stability of the TSE (relative to its unfolded state), indicating that the mutated residue is unstructured at the TSE. Recently the structures of the binding TSE for various association reactions have been evaluated by using native topology-based model simulations to calculate the contact and residue Φ values. A direct comparison between the simulated and experimental Φ values was possible for Arc-repressor and the tetramerization domain of p53 (24), giving a generally good correlation. This agreement supports the notion that the structure of the binding

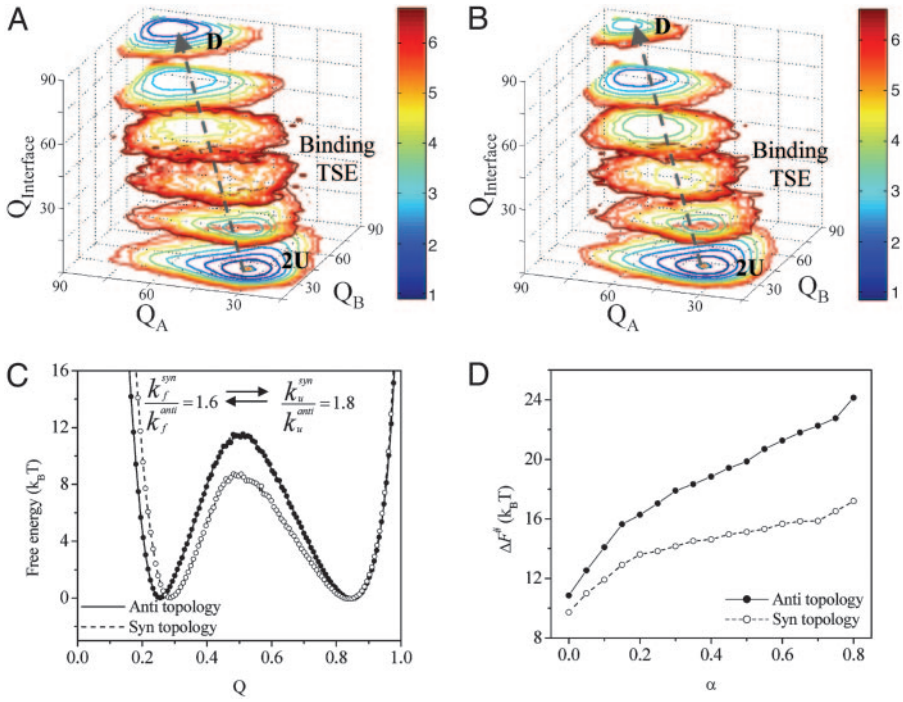

Fig. 2. The barrier for the folding of the *anti* and *syn* forms of Rop dimer. The folding free-energy landscapes for the *anti* (*A*) and *syn* (*B*) topologies of Rop dimer are shown. The reaction coordinates are the folding of the two monomers and the formation of the interface (i.e., association). U, an unfolded monomer; D, a folded dimer. The dashed arrow illustrates the coupling between folding and association. (*C*). Two-dimensional free-energy profiles for the folding and association of the two forms of the Rop dimer based on the additive native topology-based simulations. The rates for folding and unfolding for each topological structure were obtained from >1,000 events (using the additive model) that were fitted to a single exponential decay. (*D*) The folding barrier height, $\Delta F^{\#}$, as a function of α (the three-body contribution to the contact energy).

transition state indeed can, usually, be obtained by the knowledge of the final complex structure alone.

Supported by this success, we calculated the Φ values for the binding TSE for the *anti* and *syn* topologies. We must point out that for these two structures, Φ values have not yet been measured experimentally (indeed, this may be difficult because ideally measuring the kinetics of hetero association of these structurally labile species should be a problem!). The contact Φ values for the two topological structures are shown in Fig. 3*A*. The average contact Φ values of each residue are shown superimposed on the structure by using a color scale. The binding TSE for the *syn* topology is more structured than that of the *anti* topology. In both cases, the residues near the loop of each monomer (i.e., the C terminus of helix 1 and the N terminus of helix 2) are relatively structured and participate in the nucleus of the folding/association reaction. The Φ values show a symmetric pattern for the two monomers of each topology, which is anticipated for a homodimer but also can be the consequence of averaging over the sampled conformations of the TSE. It is possible that each conformation of the folding TSE of a homodimer will actually be asymmetric, with one chain being more structured than the other chain. This characteristic was observed recently in a simulation study for the TSE of several homooligomers (24). In fact, we see for both topologies some degree of decoupling between the folding of the two chains in the TSE. The *anti* topology is given a stronger asymmetric structure of the TSE (see Fig. 5, which is published as supporting information on the PNAS web site).

The Φ values for the *syn* topology and the existence of symmetry between its two chains suggest that the residues near the loops, which are all relatively structured, interact with each other and thus define its folding nucleus. In the *anti* topology, the residues near the loops are also part of the folding nucleus, but they face the helix termini of the other chain. This observation, in addition to the more asymmetric nature of the TSE conformations, suggests that the residues near the loop of one chain interact with a relatively flexible chain. The participation of the termini in the *anti* TSE introduces a sharper decrease of the entropy after folding. The larger entropy cost and the lower degree of structure in the folding TSE for the *anti* topology rationalize its higher free-energy barrier.

Free-Energy Functional. The free-energy profiles for folding Rop dimer into the *anti* and *syn* topologies were also calculated by using a detailed, residue-resolution free-energy functional method (29). This method, also based on a perfectly funneled landscape, is complementary to the simulations. The advantage of the free-energy functional-based variational numerical calculation is that it is free from sampling issues for high free-energy barrier regions and unfolded states. Thus, free-energy functional methods give more details about folding transition states such as rms deviation (rmsd) of residues without statistical uncertainty. Admittedly, such calculations must make approximations, but such calculations generally give reasonable barrier heights in comparison with experiment. The barriers are consistently higher than those from additive model simulations. The approximations made to describe a free-energy landscape by using site-specific variational localization parameters of residues around their native positions already have the effect of naturally accounting for the nonadditive interactions. Of course, we can explicitly add extra nonpairwise additive terms to the Hamiltonian to enhance the cooperativity, as was done by Eastwood and Wolynes (25).

Free-energy functional methods (29) were largely applied to folding of single domains. We extend these algorithms to the current multichain case by adding a virtual bond between monomers. Because we focus on the differences of folding pathways of different folded structures, we make the bond very weak. The relative folding profiles obtained are insensitive to the choice of intermolecular tethering interactions. As seen in Fig. 3*B*, the free-energy profiles of both *syn* and *anti* have quite high folding barriers. For comparison, we performed the present calculations also on U1A (an α/β globule), a typical single-domain protein of similar size. In comparison to that dimer, the Rop dimer has an extremely high and broad barrier (30), which can be explained by the elongated shape of the dimer and its nearly one-dimensional, dense contact interface. Comparing the profiles for *anti* and *syn* topologies, we see again that *syn* has a lower barrier than does the *anti* topology, consistent with the results from the topology-based simulation studies. The position of the barrier of the *syn* topology along the reaction coordinate is closer to the unfolded state than *anti* topology, again consistent with simulations. The free-energy profile

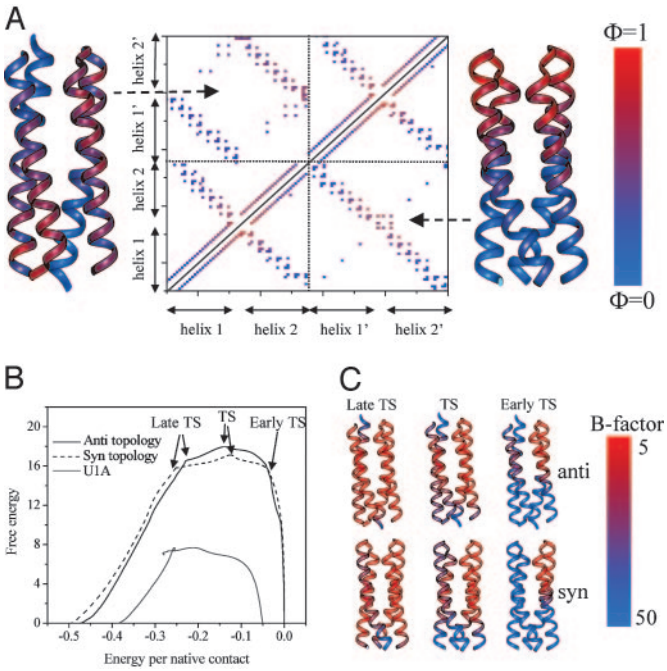

Fig. 3. The structure of the binding TSE of the Rop topologies. (*A*) Φ value analysis for the *anti* and *syn* topologies of Rop dimer. The contact and residue Φ values for the *anti* (upper triangle) and *syn* (lower triangle) forms of the Rop proteins. The contact Φ values, ϕ_{ij}, for each native contact pair between *i* and *j* is calculated from the probability of contact formation, P_{ij}, $\phi_{ij} = (P_{ij}^{TS} - P_{ij}^{U})/(P_{ij}^{F} - P_{ij}^{U})$. For Rop dimer, the superscripts F, U, and TS correspond to the folded dimer, unfolded monomers, and folding TSEs, respectively. The ϕ_i value of residue *i* is calculated from the contact values, ϕ_{ij}, by averaging all of the ϕ_{ij} values that are involved with residue *i*. The residue Φ values are presented by showing the corresponding residues in each topology in blue and red, corresponding to Φ equals to 0 and 1, respectively. (*B*) Free-energy profiles of Rop topologies. The profile of a typical single-domain protein U1A is also shown as a comparison. The *x* axis is normalized by the total number of contacts. The *y* axis is in the unit of monomer contact energy and is in approximate kilocalories/mole. (*C*) The *B* factor calculated for six important points along the two folding pathways (labeled by arrows in *B*).

plotted for either the *syn* or the *anti* topologies is only one of two parallel folding paths that the functional search uncovered. The two paths reflect the asymmetry of the transition states in that one monomer is more structured than the other.

Modeling the Rop Mutants as *anti*, *syn*, and *bisecting U* Topologies. Although 17 mutants of the Rop protein have been studied in the laboratory, the crystal structures for only two mutants (Ala_2Ile_2-6 and A31P) have been directly determined. The structures of the 10 mutants of class I (see Table 1) have been suggested to be very similar to the structure of the WT, because they all bind *in vitro* to the target RNA complex of Rop, with affinities comparable to that of WT Rop. The recent observation that only one of these mutants binds *in vivo* to the RNA complex raises some doubts about the validity of the comprehensive assigning of this structure to all the mutants in class I. As we have shown, structural heterogeneity among these mutants would provide a framework for explaining the pronounced speeding up observed for both folding and unfolding reactions of some of these mutants.

To check the assignment of a structure to each of the mutants, the sequence of each Rop mutant was threaded onto the three already observed Rop structures: the *anti*, *syn*, and *bisecting U* topologies. Each resulting designed structure was minimized and simulated by using molecular dynamics to investigate its stability and structural fluctuations. In these simulations, the proteins were

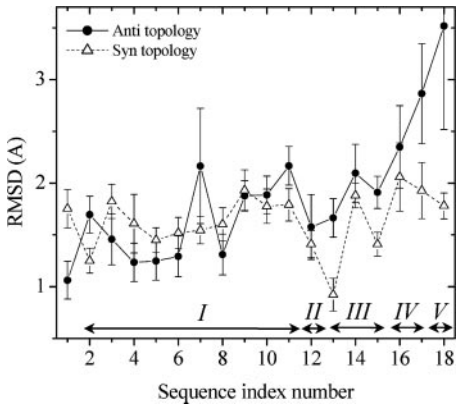

Fig. 4. The average rmsd of each designed Rop mutant as *anti* and *syn* topologies in respect to the x-ray structures of the WT and Ala_2Ile_2-6 mutants. Each designed structure was simulated with all-atom representation of the protein with explicit solvent model for 5 ns. To account for different packing of the two monomers, the rmsd was calculated after superimposing a single monomer. The arrows indicate the mutant classes as in Table 1.

simulated by using all-atom models, and the solvent was treated either with explicit solvent or by generalized Born (GB) models (31). The average rmsd values of the backbone-heavy atoms of each mutant from its redesigned structure are shown in Fig. 4 (see Fig. 6, which is published as supporting information on the PNAS web site for the GB-modeling results). The structural fluctuations of the WT Rop and the Ala_2Ile_2-6 from the *anti* and *syn* topologies act as control simulations, because their respective crystal structures have been resolved. As expected, the sequence of WT Rop displays a smaller rmsd for the *anti* topology than the *syn* topology when simulated with explicit solvent. When the solvent is represented with the GB model, the rmsd values from the *anti* and *syn* topologies are very similar. For the sequence of Ala_2Ile_2-6, a clear preference for the *syn* topology is found with both solvent models.

Although a weak test of integrity, the structural stabilities of all the Rop mutants were probed with respect to the *anti* and *syn* topologies by using explicit solvation dynamics by measuring rmsd values, which were found to be <1.7 Å. The rmsd values show lower values for some of the mutants of class I when modeled as the *anti* topology, but for the remaining mutants, the rmsd values are lower for the *syn* topology. The rmsd values of the designed mutants indicate the possibility of a conformational switching for the Rop sequences. The mutants in classes II–V consistently display lower rmsd values for the *syn* topology. A preference for the *syn* topology for these mutants can explain the lack of their binding activity.

More importantly, we find that the mutants Ala_2Leu_2-6 and Ala_2Leu_2-8, which belong to class I, also have lower rmsd values when simulated starting from the *syn* form than they do from the *anti* form (Fig. 4). These mutants show binding ability to RNA *in vitro* but not *in vivo*. They also have the highest folding and unfolding rates among the other mutants in class I. The preference for Ala_2Leu_2-6 and Ala_2Leu_2-8 to adopt the *syn* rather than the *anti* topology would explain their different thermodynamics and kinetics. The fast kinetics follows from the smaller barrier found in this study for the *syn* topology. We note that the rmsd values of all of the sequences from the *bisecting U*, when simulated starting from it, were always found to be much larger than those obtained for the *anti* and *syn* forms.

Conclusions

We propose that the non-Hammond kinetic effects found for mutants with similar stability as the WT Rop protein can be explained by a large conformational change caused by the mutation(s). A possible explanation is that the symmetry of the structure

allows for an element of frustration. Based on the *in vitro* activity of these mutants, it was assumed previously that they all adopt the same *anti* topology as the WT Rop, but their lack of *in vivo* binding ability leads us to question this hypothesis. We tested whether a few of these fast-folding/unfolding mutants might actually fold into the *syn* or *bisecting U* topologies that were found by x-ray to be the structure of other Rop mutants but have not yet been studied kinetically in the laboratory. We compared the predicted folding mechanism and rates of a protein folding into the *anti* topology with that for the *syn* topology by using a native topology-based model that was found recently to correctly capture the binding mechanism and the binding TSE of many other homooligomers. The *anti* topology has a higher barrier for folding than does the *syn* topology. The folding nucleus for both cases involves the residues around the loop (the C terminus of helix 1 and the N terminus of helix 2). In the *syn* topology, these residues from one chain interact with the corresponding residues of the other chain. In the *anti* topology, however, these residues interact with the termini of the other chain. The participation of the termini in the TSE accounts for its higher barrier for folding and unfolding. The all-atom modeling of the 17 Rop mutants as the *anti*, *syn*, and *bisecting U* topologies suggests that at least the two fastest folding mutants probably adopt the *syn* topology rather than the *anti* or the *bisecting U* topologies. This modeling would support our proposal that the origin for the folding speeding up is the presence of a different topological structure of nearly the same free energy but with a lower free-energy barrier for formation and breakup.

The Rop mutants that exhibit only a moderate speeding up of the folding and unfolding do not display in the all-atom modeling a strong preference toward the *syn* topology. However, the symmetry imposed in the sequences of the mutants when redesigning the hydrophobic core of the homodimers introduces energetic frustration that may lead to nearly degenerate structures. Accordingly, we expect that some of the mutants may adopt both the *anti* and *syn* topologies and may populate both structures in solution. Similar behavior of conformational nonspecificity due to symmetry in the sequence was observed previously in the design of the α_2 family of dimeric four-helix-bundle proteins (32). An early designed member of this family, α_2B, consisted of a pair of interconnected, identical helices, the sequences of which consisted of only Leu, Glu, and Lys. This sequence shows a molten globule-like conformation. When the sequence degeneracy of the helices was reduced (α_2C peptide), the protein showed an increased ability to adopt a unique conformation. However, additional changes in the sequence (α_2D peptide) were needed to induce a complete conformational specificity (33). In the present case, the redesigned hydrophobic cores introduce symmetry in the sequence that destroys the conformational specificity of the WT Rop, which causes it to adopt the *anti* topology. The existence of a mixture of at least two topological structures for the Rop protein variants can rationalize the *in vitro* activity and the lack of *in vivo* activity of some of the Rop mutants. A mixture of *anti* and *syn* topologies (even when the *syn* topology is much more populated than the *anti* topology) might show binding activity because the *anti* conformation, which is the active form of Rop, will become a dominant part of the population by virtue of the binding to RNA itself.

The existence of two competing topologies, each being preferred by different mutants or simultaneously populated in a mixture, resolves not only the mystery of Rop-folding kinetics but also can explain the biphasic kinetics found for the slow-folding mutants but not found for the fast mutants (15). Moreover, the decoupling between the thermal and chemical denaturation can be understood also by taking into consideration the possibilities of coexistence of two different structures (34). We see that recognizing the frustrated energy landscape of Rop caused by its symmetry provides a coherent framework to unravel the mystery of Rop folding. We feel this explanation is much more concrete and fits more comprehensively the observations than does the recent proposal that the key to understanding the Rop kinetics is a strong drying effect due to there being more hydrophobic residues in the core (35). The fact that hydrophobic dewetting and collapse have not been found to decouple for the binding of two rigid protein domains (36) makes it even less probable for such decoupling to occur in a folding/binding reaction involving flexible elements, as seems to be the case for Rop.

Variations in the folding and unfolding rates of homologous proteins are common and usually can be explained as Hammond effects, consistent with a simple funneled landscape. Non-Hammond effects of significant acceleration of both folding and unfolding while maintaining the stability of the WT protein are less common, but such effects have also been reported recently for repeat R15 of α-spectrin when compared with isolated repeats R16 and R17 (37). In addition, a recent *de novo* designed acylphosphatase protein exhibits a similar non-Hammond effect (38). In our view, it would not be unwise to consider the possibility of extreme conformational changes in these cases as well as in the present case of Rop dimer.

We thank Diego Fereirrio and Johan Ulander for helpful discussions and Lynne Regan for generously listening to, and sharing thoughts on, our resolution of the Rop-folding mystery. This work was funded by the National Science Foundation-sponsored Center for Theoretical Biological Physics (Grants PHY-0216576 and 0225630) with additional support from Grant MCB-0084797. Computations were carried out at the University of California at San Diego KeckII computing facility (partially supported by the National Science Foundation, Division of Molecular and Cellular Biosciences). S.S.C. is supported by a University of California at San Diego Molecular Biophysics training grant.

1. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997) *Annu. Rev. Phys. Chem.* **48,** 539–594.
2. Onuchic, J. N. & Wolynes, P. G. (2004) *Curr. Opin. Struct. Biol.* **14,** 70–75.
3. Gunasekaran, K., Eyles, S. J., Hagler, A. T. & Gierasch, L. M. (2001) *Curr. Opin. Struct. Biol.* **2001,** 83–93.
4. Murzin, A., Brenner, S., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540.
5. Orengo, C., Jones, D. & Thornton, J. (1994) *Nature* **372,** 631–634.
6. Koehl, P. & Levitt, M. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1280–1285.
7. Nishimura, C., Prytulla, S., Dyson, H. & Wright, P. (2000) *Nat. Struct. Biol.* **7,** 679–686.
8. Martinez, J. & Seraano, L. (1999) *Nat. Struct. Biol.* **6,** 1010–1016.
9. Onuchic, J., Nymeyer, H., Garcia, A., Chahine, J. & Socci, N. (2000) *Adv. Protein Chem.* **53,** 87–152.
10. Wolynes, P. G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14249–14255.
11. Benner, D., Kokkinidis, M. & Tsernoglou, D. (1987) *J. Mol. Biol.* **196,** 657–675.
12. Eberle, W., Pastore, A., Sander, C. & Rosch, P. (1991) *J. Biomol. NMR* **1,** 71–82.
13. Munson, M., O'Brien, R., Sturtevant, J. M. & Regan, L. (1994) *Protein Sci.* **3,** 2015–2022.
14. Munson, M., Balasubramniam, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M. & Regan, L. (1996) *Protein Sci.* **5,** 1584–1593.
15. Munson, M., Anderson, K. S. & Regan, L. (1997) *Folding Des.* **2,** 77–87.
16. Glykos, N. M., Cesareni, G. & Kokkinidis, M. (1999) *Structure (London)* **7,** 597–603.
17. Glykos, N. M. & Kokkinidis, M. (2004) *Proteins Struct. Funct. Genet.* **56,** 420–425.
18. Magliery, T. J. & Regan, L. (2004) *Protein Eng. Des. Sel.* **17,** 77–83.
19. Koga, N. & Takada, S. (2001) *J. Mol. Biol.* **313,** 171–180.
20. Chavez, L. L., Onuchic, J. N. & Clementi, C. (2004) *J. Am. Chem. Soc.* **126,** 8426–8432.
21. Levy, Y., Wolynes, P. G. & Onuchic, J. N. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 511–516.
22. Levy, Y., Caflisch, A., Onuchic, J. N. & Wolynes, P. G. (2004) *J. Mol. Biol.* **340,** 67–79.
23. Levy, Y., Papoian, G. A., Onuchic, J. & Wolynes, P. G. (2004) *Isr. J. Chem.* **44,** 281–297.
24. Levy, Y., Cho, S. S., Onuchic, J. N. & Wolynes, P. G. (2005) *J. Mol. Biol.*, in press.
25. Eastwood, M. P. & Wolynes, P. G. (2001) *J. Chem. Phys.* **114,** 4702–4716.
26. Ejtehadi, M. R., Avall, S. P. & Plotkin, S. S. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 15088–15093.
27. Fersht, A. R. (1994) *Curr. Opin. Struct. Biol.* **5,** 79–84.
28. Daggett, V. & Fersht, A. R. (2003) *Nat. Rev. Mol. Cell. Biol.* **4,** 497–502.
29. Portman, J. J., Takada, S. & Wolynes, P. G. (2001) *J. Chem. Phys.* **114,** 5069–5081.
30. Rosengarth, A., Rosgen, J. & Hinz, H.-J. (1999) *Eur. J. Biochem.* **264,** 989–995.
31. Feig, M., Onufriev, A., Lee, M. S., Im, W., Case, D. A. & Brooks, C. L, III (2004) *J. Comput. Chem.* **25,** 265–284.
32. Hill, R. B., Raleigh, D. P., Lombardi, A. & Degrado, W. F. (2000) *Acc. Chem. Res.* **33,** 745–754.
33. Hill, R. B. & Degrado, W. F. (1998) *J. Am. Chem. Soc.* **120,** 1138–1145.
34. Willis, M. A., Bishop, B., Regan, L. & Brunger, A. T. (2000) *Structure (London)* **8,** 1319–1328.
35. Lum, K., Chandler, D. & Weeks, J. D. (1999) *J. Phys. Chem. B* **103,** 4570–4577.
36. Zhou, R., Huang, X., Margulis, C. J. & Berne, B. J. (2004) *Science* **305,** 1605–1609.
37. Scott, K. A., Batey, S., Hooton, K. A. & Clarke, J. (2004) *J. Mol. Biol.* **344,** 195–205.
38. Scalley-Kim, M. & Baker, D. (2004) *J. Mol. Biol.* **338,** 573–583.
39. Predki, P. F., Nayak, L. M., Gottlieb, M. B. C. & Regan, L. (1995) *Cell* **80,** 41–50.