Native Topology of the Designed Protein Top7 is Not Conducive to Cooperative Folding

Zhuqing Zhang and Hue Sun Chan*
Departments of Biochemistry and of Molecular Genetics, University of Toronto, Faculty of Medicine, Toronto, Ontario, Canada

ABSTRACT Many single-domain proteins with <100 residues fold cooperatively; but the recently designed 92-residue Top7 protein exhibits clearly non-two-state behaviors. In apparent agreement with experiment, we found that coarse-grained, native-centric chain models, including potentials with and without elementary desolvation barriers, predicted that Top7 has a stable intermediate state in which the C-terminal fragment is folded while the rest of the chain remains disordered. We observed noncooperative folding in Top7 models that incorporated nonnative hydrophobic interactions as well. In contrast, free energy profiles deduced from models with desolvation barriers for a set of thirteen natural proteins with similar chain lengths and secondary structure elements suggested that they fold much more cooperatively than Top7. Buttressed by related studies on smaller natural proteins with chain lengths of ~40 residues, our findings argue that the de novo native topology of Top7 likely imposed a significant restriction on the cooperativity achievable by any design for this target structure.

Received for publication 2 October 2008 and in final form 19 November 2008.

*Correspondence: chan@arrhenius.med.toronto.edu

Experiments showed that many small, single-domain natural proteins fold cooperatively in a two-state-like manner (1). Theoretical analyses, however, indicated that cooperativity is a remarkable biophysical property, not readily achievable in chain models with only pairwise-additive effective interactions, and that its mimicry often requires many-body terms in the model potential (2). In conjunction with experimental observations, e.g., that the de novo designed α/β protein Top7 (3) failed to fold cooperatively (4,5), these considerations have led to the view that folding cooperativity is likely a result of natural selection for biological functions, such as resistance against harmful protein aggregation (2,4,5).

Native-centric, coarse-grained, explicit-chain modeling is useful for understanding protein folding energetics (6–11). Notably, such models have recently been used to rationalize the differences in folding cooperativity among several proteins with ~40 residues (12). Here, we apply a similar approach to gain insight into the folding behavior of the larger, 92-residue Top7. This designed protein has a very stable novel fold that has not been observed among natural proteins. At the same time, the folding kinetics of Top7 is far more complex than many natural proteins of comparable size. Top7 exhibited severe chevron rollovers and other hallmarks of noncooperative, non-two-state folding (4,5), although their physical origin is not well understood.

This study models Top7 by using C^{α} chains with four different interaction schemes: the common Gō-like potential (6,8), a native-centric potential with desolvation barrier (db) terms (7,8,11), by themselves as well as augmented by sequence-dependent nonnative hydrophobic interactions as introduced by Zarrine-Afsar and colleagues (13). Langevin dynamics is used for conformational sampling (12). Model behaviors of Top7 are contrasted against those of several natural proteins that are of comparable size and have similar secondary structure elements in their folded structures.

Fig. 1 assesses folding cooperativity by determining free energy profiles (8) of Top7 near each model's transition midpoint, where $P(Q)$ is the probability of a conformation having fractional number of native contacts Q (defined by a threshold separation of 4.5 Å, as in ref. 12). Remarkably, all four models stipulate consistently that Top7 is not a two-state protein. In the common Gō-like model, an intermediate (I) state more stable than both the folded and unfolded states appears around $Q \approx 0.56$. The behavior in the db model is clearly three-state, with an I-state around $Q \approx 0.6$. These predicted features are qualitatively robust even if the native-contact threshold is changed from 4.5 Å to 5.5 Å or 6.5 Å. When nonnative hydrophobic interactions are allowed, the peculiar free energy profile for the common Gō-like model becomes essentially barrierless in Q, whereas their effect on the free energy profile for the db model is rather minimal and amounts to only small decreases in the heights of the two overall barriers.

Fig. 2 analyzes the contact patterns in the model native folded structure and various I-states of Top7. The native fold of Top7 contains two α-helices and a five-strand antiparallel β-sheet (Fig. 2 *d*). It is quite clear from its native contact map (Fig. 2 *a*, upper-left map) that the C-terminal fragment (CFr) of Top7 is significantly more compact than the N-terminal part because there are more contacts between the second α-helix and the last three β-strands.

To gain structural information about the I-states manifested in Fig. 1, we determined contact probabilities in three

Editor: Gerhard Hummer.

doi: 10.1016/j.bpj.2008.11.004

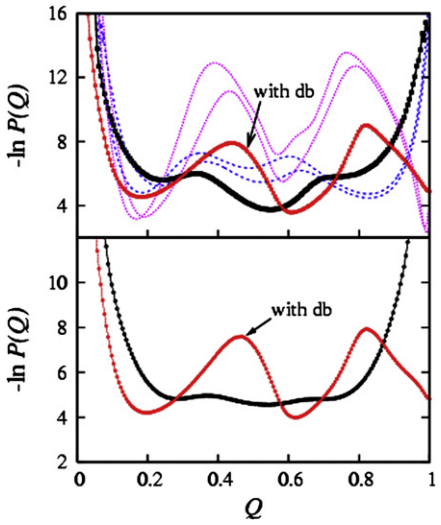

FIGURE 1 Model free energy profiles for Top7. Top: Common Gō-like (*black*) and db (*red*) models. The blue and magenta curves are profiles, respectively, for the Gō and db models, based on different distance thresholds of 5.5Å (*lower*) and 6.5Å (*upper*) for native contact definition. Bottom: Free energy profiles when nonnative hydrophobic interactions are incorporated into the Gō-like (*black*) and db (*red*) models.

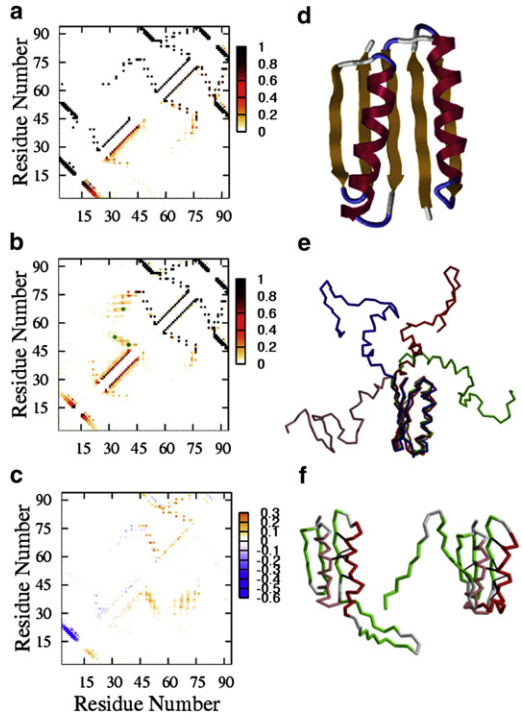

FIGURE 2 Contact patterns in Top7 models. (*a*) (*upper left* (*ul*)) Contact map of native structure in (*d*) (PDB code 1QYS); (*lower right* (*lr*)) contact probabilities of the I-state in the Gō-like model ($0.54 < Q < 0.58$). (*b*) (*ul*) Contact probabilities of the I-state in the db model with nonnative hydrophobic interactions ($0.605 < Q < 0.645$); (*lr*) contact probability of the I-state in the db model ($0.59 < Q < 0.63$). (*c*, *ul*) *lr* of *b* minus *lr* of *a*. (*c*, *lr*) *ul* of *b* minus *lr* of *b*. (*e*) Backbone representation of two I-state conformations each for the Gō-like (*red* and *pink*) and db (*blue* and *green*) models. (*f*) Two representative I-state conformations in the db model with nonnative hydrophobic interactions (α-helices in *red* or *pink*, β-strands in *green*). Selected low-energy nonnative contacts are indicated by black lines; the thicker black lines among them mark the nonnative contacts common to both conformations (residue pairs (31, 50), (35, 65) and (38, 46) marked by *green solid circles* in *b*).

sets of 500 randomly selected conformations with Q values characteristic of each of the I-states. The resulting contact probability maps in Fig. 2 show that CFr is well-formed in all three I-states predicted by our models. (The common Gō model with nonnative hydrophobic interactions does not predict a thermodynamic I-state.) In contrast, although the first α-helix and the β-strands in the N-terminal are sometimes partially formed in the I-states, they tend not to attain a stable tertiary structure. As illustrated by Fig. 2 *e*, typical I-state conformations often consist of a more disordered N-terminal tail flying around a compact CFr core.

Fig. 2 *c* compares contact probabilities in the I-state ensembles of the common Gō and db models (upper-left map) as well as the influence of nonnative hydrophobic interactions in the db model (lower-right map). Relative to the without-db Gō model, the db model significantly destabilizes the N-terminal β-hairpin while imparting some additional stability to the already stable CFr in the Top7 I-state. Nonnative hydrophobic interactions appear in the I-state when they are allowed in the db model, as shown by the orange-color points in the upper-left map in Fig. 2 *b* that are not found in the lower-right maps of Fig. 2, *a* and *b*.

Our results further indicate that nonnative contacts in the Top7 I-state tend to occur between the two helices and also between the first helix and the middle β-strand (Fig. 2 *b*, upper left and Fig. 2 *c*, lower right). Fig. 2 *f* shows two typical conformations with favorable (low energy, <0.8) nonnative contacts indicated by black lines. In view of recent success in predicting nonnative contacts in folding (13), it would be instructive to study experimentally the roles of these contacts in the folding kinetics of Top7.

To explore whether the predicted thermodynamics of Top7 is typical or atypical in Nature, we selected 13 natural proteins of similar size (85-100 residues) with similar secondary elements and compared their db model-predicted midpoint free energy profiles with that of Top7 (Fig. 3 and Table 1). A shorter protein, gpW (PDB code 1HYW), recently identified experimentally as a possible downhill folder (14), is also considered (see also Liu and Gruebele (15)). Fig. 3 shows that although the natural proteins' free energy barrier heights vary, they all have an appreciable barrier separating the folded and unfolded states, and none of them has a stable I-state as predicted for Top7 by the same db model. 1HYW exhibits a low free energy barrier, consistent with a lower folding cooperativity than other natural proteins studied here. Table 1 compares the relative contact order, absolute contact order (16), and the number of nonlocal contacts per residue N_N^c (9) of the proteins. The values of these topological parameters for Top7 are

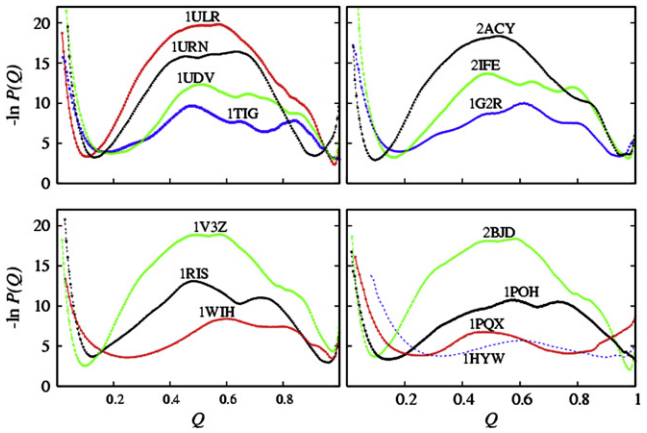

FIGURE 3 db-model free energy profiles for natural proteins with PDB codes as marked.

low, but they are not the lowest in Table 1. These parameters are thus insufficient to predict the severe noncooperative behavior of Top7 captured by our explicit-chain modeling.

Taken together, our results suggest strongly that the target structure of Top7 plays a dominant role in the designed protein's noncooperative behavior. The native topology of Top7 may be a more intrinsic impediment to cooperative folding (12) than the presumably limited competence of artificial design in accomplishing such a feat (4,5). In fact, one recent view holds that the Top7 sequence was already quite well designed when evaluated by a ''local frustration'' (17) criterion (P.G. Wolynes, University of California at San Diego, personal communication, 2008), even though many fundamental questions regarding artificial versus evolutionary design remain open. A lesson from our results and earlier related findings (9,10,12,18) is that native-centric, coarse-grained, explicit-chain modeling can be a potentially powerful tool in experimental protein design to screen target structures for their likelihood of achieving folding cooperativity.

TABLE 1 Comparing the native topology parameters for Top7 and fourteen natural proteins

PDB code	N	n_α/n_β	RCO	ACO	N_N^c
1QYS	92	2/5	10.77	9.91	1.86
1ULR	87	2/5	21.63	18.82	2.29
1V3Z	90	2/5	21.68	19.51	2.27
2BJD	90	2/5	21.73	19.56	2.31
2ACY	98	2/4	20.02	19.62	2.31
1RIS	97	2/4	18.95	18.38	2.10
1UDV	88	2/4	11.94	10.50	1.84
1PQX	91	2/4	10.62	9.66	1.45
1TIG	88	2/4	11.99	10.55	1.84
1POH	85	3/4	17.64	14.99	2.05
1URN	96	3/4	16.90	16.23	2.00
2IFE	91	3/3	12.66	11.52	1.99
1G2R	94	4/3	9.89	9.30	1.63
1WIH	84	2/6	8.93	7.50	1.52
1HYW	58	2/2	13.35	7.74	1.29

ACO, absolute contact order; N, number of amino acid residues; n_α, n_β, number of α-helices and β-strands; N_N^c, number of nonlocal contacts per residue; and RCO, relative contact order.

ACKNOWLEDGMENTS

We thank Artem Badasyan, Allison Ferguson and Zhirong Liu for technical assistance. We thank Peter Wolynes for sharing his unpublished work on Top7 during a recent ACS meeting in Philadelphia.

Financial support of this work was provided by the Canadian Institutes of Health Research (grant MOP-84281) to H.S.C., who is a Canada Research Chair holder.

REFERENCES and FOOTNOTES

1. Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*. 30:10428–10435.
2. Chan, H. S., S. Shimizu, and H. Kaya. 2004. Cooperativity principles in protein folding. *Methods Enzymol*. 380:350–379.
3. Kuhlman, B., G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, et al. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 302:1364–1368.
4. Scalley-Kim, M., and D. Baker. 2004. Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. *J. Mol. Biol*. 338:573–583.
5. Watters, A. L., P. Deka, C. Corrent, D. Callender, G. Varani, et al. 2007. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*. 128:613–624.
6. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and ''en-route'' intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol*. 298:937–953.
7. Cheung, M. S., A. E. García, and J. N. Onuchic. 2002. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA*. 99:685–690.
8. Kaya, H., and H. S. Chan. 2003. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric modeling? *J. Mol. Biol*. 326:911–931.
9. Zuo, G., J. Wang, and W. Wang. 2006. Folding with downhill behavior and low cooperativity of proteins. *Proteins*. 63:165–173.
10. Knott, M., and H. S. Chan. 2006. Criteria for downhill folding: calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins*. 65:373–391.
11. Liu, Z., and H. S. Chan. 2005. Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol*. 349:872–889.
12. Badasyan, A., Z. Liu, and H. S. Chan. 2008. Probing possible downhill folding: native contact topology likely places a significant constraint on the folding cooperativity of proteins with ~40 residues. *J. Mol. Biol*. 384:512–530.
13. Zarrine-Afsar, A., S. Wallin, A. M. Neculai, P. Neudecker, P. L. Howell, et al. 2008. Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. USA*. 105:9999–10004.
14. Fung, A., P. Li, R. Godoy-Ruiz, J. M. Sanchez-Ruiz, and V. Muñoz. 2008. Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with $\alpha+\beta$ topology that folds downhill. *J. Am. Chem. Soc*. 130:7489–7495.
15. Liu, F., and M. Gruebele. 2008. Downhill dynamics and the molecular rate of protein folding. *Chem. Phys. Lett*. 461:1–8.
16. Ivankov, D. N., S. O. Garbuzynskiy, E. Alm, K. W. Plaxco, D. Baker, et al. 2003. Contact order revisited: influence of protein size on the folding rate. *Protein Sci*. 12:2057–2062.
17. Ferreiro, D. U., J. A. Hegler, E. A. Komives, and P. G. Wolynes. 2007. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. USA*. 104:19819–19824.
18. Cho, S. S., P. Weinkam, and P. G. Wolynes. 2008. Origins of barriers and barrierless folding in BBL. *Proc. Natl. Acad. Sci. USA*. 105:118–123.