Interplaying Roles of Native Topology and Chain Length in Marginally Cooperative and Noncooperative Folding of Small Protein Fragments

ARTEM BADASYAN,[1,2] ZHIRONG LIU,[1,2] HUE SUN CHAN[1,2,3]

[1]*Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A8, Canada*
[2]*Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada*
[3]*Department of Physics, University of Toronto, Toronto, ON M5S 1A7, Canada*

Received 18 December 2008; accepted 23 March 2009
Published online 19 May 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/qua.22272

ABSTRACT: Coarse-grained chain simulations were used to study fragments of two homologous proteins of the peripheral subunit-binding domain (PSBD) family, *Bacillus stearothermophilus* PSBD (E3BD) and *Escherichia coli* 2-oxo-glutarate dehydrogenase PSBD (BBL). To ascertain a robust rank order of folding cooperativity, native-centric intraprotein interactions were modeled by (i) a common Gō-like potential, and (ii) native-centric potentials with desolvation barriers or (iii) many-body terms. Homologous proteins can possess substantially different folding cooperativity. Consistent with experiment, our calculations indicated that E3BD fragments fold more cooperatively than BBL fragments of approximately the same chain length. For a given fragment, native contacts deduced from Protein Data Bank structures can vary significantly depending on the number of residues that the structure encompasses in addition to those of the fragment itself, resulting in variation in model folding cooperativity predicted using different native contact sets for the

Correspondence to: H. S. Chan; e-mail: chan@arrhenius.med.toronto.edu

Artem Badasyan is currently at Dipartimento di Chimica Fisica, Facoltà di Scienze, Università Ca' Foscari di Venezia, Calle Larga S. Marta DD2137, I-30123 Venezia, Italy.

Zhirong Liu is currently at College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

Contract grant sponsor: The Canadian Institutes of Health Research (CIHR).

Contract grant number: MOP-84281.

same fragment. This observation underscores that folding cooperativity of these fragments can be extremely sensitive to change in chain length. Thus, a ∼40-residue protein fragment's folding cooperativity, or lack thereof, does not necessarily imply essentially identical behaviors for super- or sub-fragments with only several residues more, or less, than the given fragment. Ramifications for experimental investigations of downhill folding are discussed. Int J Quantum Chem 109: 3482–3499, 2009

Key words: Gō model; desolvation barriers; downhill folding; many-body interactions; nonadditivity

1. Introduction

Protein folding is a focal point of diverse areas of investigation in biophysical chemistry and biomedical sciences. As in any scientific endeavor, to elucidate the physical basis of the folding process, theory and computational modeling are indispensable. Indeed, significant theoretical advances have been made during the past 20 years (see, e.g., Ref. [1, 2] for historical perspectives). Nonetheless, our understanding of certain fundamental questions in protein folding is still far from adequate. One intriguing aspect of protein behavior is the high degrees of cooperativity exhibited in the folding of some proteins. Since the 1960s [3], it has been known experimentally that many proteins with $\lesssim$180 amino acid residues populate essentially only two thermodynamic states — folded and unfolded. These proteins possess little, though not nonexistent, intermediate conformational population at equilibrium [4, 5], even though kinetic intermediates could transiently accumulate during folding [6]. Subsequently, in the early 1990s, the possibility of even more switch-like folding came to light [7]. From that time onward, an increasing number of single-domain proteins with $<$100 residues have been found to undergo essentially two-state folding/unfolding kinetics with no appreciable, or extremely low, population of transient intermediates [8, 9].

Inspired by these experimental discoveries, it has become progressively clear from recent theoretical analyses that protein folding cooperativity is a physically remarkable phenomenon. As a polymer property, experimental two-state–like folding cooperativity does not readily follow from intuitive notions of protein energetics based upon pairwise additive terms [10]. Rather, cooperativity constitutes a strong indication that substantial many-body effects are in play in the driving forces for folding [11, 12]. Thus, a polypeptide's ability to fold to a stable, essentially unique structure does not necessarily mean that it has to do so cooperatively [13]. This theoretical insight was confirmed by the noncooperative folding behavior of the de novo designed protein Top7 [14], raising the possibility that evolutionary selection for biological functions might have led to folding cooperativity [15, 16].

In light of a better conceptual grasp of protein folding cooperativity, especially the recognition that cooperativity is not automatic, recent experimental findings of possible noncooperative, nontwo-state, barrierless or "downhill" folding [17, 18] are exciting (see Refs. [19, 20] for recent reviews). Downhill folding was first stated explicitly in an insightful exposition as a new theoretical possibility [1], even though — unbeknownst at the time — most of the coarse-grained protein chain models studied in the 1990s actually had low degrees of folding cooperativity and thus they either folded "downhill" or had tendencies to do so (see model evaluation in Refs. [11, 12]). Theoretical and experimental studies identified two types of downhill folding: First, some proteins such as certain mutants of λ-repressor behave in a two-state manner near their transition midpoint and under mildly folding conditions, but they can fold downhill under strongly folding conditions [17, 21]. Second, some proteins can undergo global downhill folding, which means that the protein always behaves in a one-state, unimodal manner for the entire range of experimental conditions [22, 23].

Motivated by the proposal that the *Escherichia coli* protein BBL is a global downhill folder [18], much of the experimental investigation of global downhill folding has focused on ∼40-residue fragments of the peripheral subunit-binding domain (PSBD) family [24–28]. During the past several years, significant information has been gained about the folded structures, thermodynamics, and folding kinetics of various BBL constructs as well as BBL homologues from other organisms [24–28], even though whether BBL is indeed a global downhill folder remains controversial [29–31]. More recent progress following these lines of investigation includes an extensive folding study of the BBL homologue *Pyrobaculum aerophilum* PSBD (POB) [32], a study involving high-resolution

temporal and structural monitoring of BBL folding dynamics [33], an analysis of pH effects on BBL folding [34], as well as an extensive study of folding kinetics of several BBL mutants each with a single tryptophan substitution [35]. Beyond the PSBD family, the 62-residue gpW protein was recently identified as a likely candidate to fold downhill [36]. Other studies delineated the conditions governing the crossover from two-state-like to downhill folding in λ_{6-85} mutants [37] and in WW domain sequences [38].

In our view, data from these experiments and the theoretical considerations they inspired (see below) indicated quite convincingly that global downhill folding is probable for at least certain mutated forms of natural proteins such as BBL. As we have noted [22, 39], the possibility of global downhill folding is conceptually in line with the recently broadened view of the structure-function relationship of proteins. Many proteins participating in cellular regulation in higher organisms are now known to function in intrinsically disordered states, involving broad conformational ensembles rather than an essentially unique structure [40, 41]. The behaviors of intrinsically disordered proteins demonstrate that biology can exploit physical interactions to serve a wider variety of purposes than those in globular, cooperatively folding proteins (see, e.g., Refs. [42, 43]). Conformational diversity could also allow a protein to tackle multiple tasks and facilitate evolution of new biological functions [44, 45].

Recent coarse-grained chain modeling of PSBD and other proteins have demonstrated that native topology (i.e., native contact pattern) plays a significant role in folding cooperativity [22, 39, 46–48]. (For a historical perspective of the usage of the term "topology" in protein folding, see pp. S308–309 of Ref. [49]). By comparing the folding cooperativity of models for different proteins, it is quite clear that chain models with more nonlocal contacts tend to fold more cooperatively [22, 46]. This finding is in line with earlier theoretical predictions [50, 51] and the model results provided plausible rationalizations for the apparent difference in folding cooperativity (or lack thereof) between different BBL constructs [48] as well as between PSBD proteins from different organisms [39]. In fact, the trend is consistent with the seminal discovery that among two-state proteins, those with more nonlocal native contacts tend to fold slower [8, 52] (see Ref. [53] for theoretical ramifications of this discovery). The consistency follows from the consideration that a higher folding cooperativity means a higher overall free energy barrier between the unfolded and folded states, and thus a higher barrier to folding kinetics as well. Therefore, proteins with higher degrees of folding cooperativity should fold slower [39, 49]. The recent studies discussed above showed that this correlation between the predominance of nonlocal contacts and folding cooperativity applies not only to two-state folders but also to marginal cooperatively or noncooperatively (non-two-state) folding short protein fragments. In fact, the constraining effect of native topology on folding cooperativity is not restricted only to proteins with ~40 residues. A notable example is the designed protein Top7 with >90 residues. The noncooperative folding of this protein was thought to be caused primarily by limitations of current artificial design algorithms [14, 16]. In view of the field's rudimentary knowledge of protein energetics, it is certainly reasonable to assume that current artificial design would be less competent in achieving folding cooperativity than natural evolutionary design. However, recent model simulations showed that a more basic physical origin of the non-two-state folding of Top7 may be the constraining effects of its very de novo target native topology [54].

Beside explicit-chain modeling, nonexplicit-chain Ising-like treatments [55] have also been applied to study putative downhill folding [56, 57]. These approaches have proved useful for comparing folding cooperativity of different proteins. However, because of their intrinsic limitations, these abstract constructs tend to overestimate folding cooperativity (see below). This is one of the reasons for our unyielding emphasis on the primacy of explicit-chain methodologies in folding studies [10]. Among explicit-chain approaches, all-atom simulations would appear to be most realistic. Downhill folding was addressed recently by three independent atomic simulation studies [58–60]. Simulation data obtained by Zhang et al. indicated that the free energy profile of BBL as a function of fractional number of native contacts, Q, was one-state [58], similar to those observed in coarse-grained modeling [22, 46] and in agreement with the experimental interpretation that this protein is a global downhill folder [18, 27]. Simulation data obtained by Pitera et al. for several different BBL sequence constructs and the homologue *Bacillus stearothermophilus* E3BD showed no sigmodial folding transition as a function of temperature [59]. Although the authors noted that the simulated variation of E3BD properties as a function of temperature was slightly sharper than those for the BBL constructs

[59], the difference is much less discernible than the clear distinction in folding cooperativity between BBL and E3BD in experiment [25], a distinction that native-centric coarse-grained modeling was able to capture [39]. The more recent simulation by Settanni and Fersht produced the intriguing result that the BBL native state was heterogeneous and that the population ratio of the two predominant conformations changed with folding conditions, a feature the authors proposed as a possible underpinning of the apparently asynchronous unfolding behavior of BBL [60].

How should these all-atom simulation efforts be evaluated? For the one-state free energy profiles computed from all-atom simulations of putative downhill folders to carry weight, control simulations need to be performed. The reason is basic: The finding of a one-state profile for a putative downhill folder would be informative only if, using the same algorithm, a two-state profile can be found for a cooperatively folding protein. Otherwise, the meaning of a predicted one-state profile would be ambiguous if equivalent simulations for proteins that are experimentally known to be two-state were also predicted to have one-state transitions. The logic is analogous to that for control experiments. Unfortunately, as it stands, it is not clear whether current atomic simulations are capable of predicting an appreciable free energy barrier for proteins that are known to be two-state folders. For instance, in an extensive all-atom simulation study of a 46-residue fragment [61] of the 58-residue B domain of Protein A, which likely folds in a two-state–like manner experimentally [62], no clear free energy barrier was observed along Q at the model chain's transition midpoint [Fig. 2(C) of Ref. [61]]. Thus, the question of control atomic simulation remains to be elucidated, and atomic simulation results on putative downhill folding should be interpreted with caution in the meantime.

Now, to gain further insight into global downhill folding, we use coarse-grained, explicit-chain modeling with multiple native-centric interaction schemes [39]. In evaluating whether a protein may fold downhill, coarse-grained modeling has the obvious advantage, owing to its computational tractability, that the protein's predicted behaviors can be explored using physically motivated variations in model construction [13, 63] and compared in detail with control cases. For example, our previous work has shown clearly that several ~40-residue fragments of members of the PSBD family fold significantly less cooperatively than a 39-residue fragment of the N-terminal domain of ribosomal protein L9 (NTL9), thus providing a plausible rationalization for the corresponding experimental observations [22, 39].

We focus this study on BBL and E3BD. Our previous multiple-interaction-scheme study modeled one 45-residue fragment for each of these proteins. Those models were based on structures listed under Protein Data Bank (PDB) codes 2BTH and 1W4E for single-point mutants with a tryptophan substitution at position 166. Our previous effort predicted that the 45-residue E3BD fragment should fold more cooperatively than the 45-residue BBL fragment [39]. This result is in accord with the trend exhibited by experimental chevron plots of these fragments [25] and consistent with the calorimetric observation of a significantly sharper heat capacity peak for E3BD than for BBL [Fig. 8(a) of Ref. [25]]. Building on this theoretical advance, the present study takes a broader view. We note that folding cooperativity can be significantly different for native-centric models of BBL fragments of different chain lengths and/or based upon different PDB structures [48]. Experimental data on BBL and E3BD fragments with different chain lengths are available [18, 64] and should be taken into account. Therefore, we extend our approach [39] to consider three PDB structures for BBL sequences with 40, 47, and 51 residues (same structures studied in Ref. [48]); and two PDB structures for E3BD sequences with 44 and 55 residues. Additional results on NTL9 are included also for comparison.

As in Ref. [22], we examine below the free energy profiles, sigmoidal transitions, and distributions of radius of gyration of our models to evaluate and compare their folding cooperativity. We also apply the theoretical equivalent [10–13] of the experimental calorimetric criterion [3–5], which has seen useful applications in analyses of barrier heights and other aspects of putative global downhill folding [22, 65, 66]. Interestingly, similar to a prior observation for two different BBL structures in the PDB [48], we found that folding cooperativity of native-centric models for the same BBL sequence or the same E3BD sequence truncated from different PDB structures can be substantially different. Because the different PDB structures encompass different total number of residues (see above), this apparent uncertainty in native-centric model prediction underscores the fact that core packing and folding cooperativity for these small protein fragments are highly sensitive to even small variation in total chain length [64]. Taking this into consideration, a more coherent physical picture

emerges: Our calculations suggest that when homologous BBL and E3BD sequences of equal length (in the range ≈40–45 residues) are compared, the E3BD fragment generally folds more cooperatively than the BBL fragment. This trend agrees with experiment and is apparently driven by sequence-specific interactions that underpin a higher native contact density in E3BD than that in BBL.

2. Models and Method

Recognizing the predominant role of native interactions in natural protein folding, the present methodology is limited to that of native centric modeling. Favorable nonnative interactions are not considered although experiments showed that nonnative interactions exist in protein folding [67–69]. Because some of the effects of nonnative interactions may be treated as a perturbation on coarse-grained native-centric models [70, 71], the present approach may be viewed as a zeroth-order approximation in that regard.

We use the multiple native-centric interaction scheme method developed recently [39]. For each PDB structure investigated, continuum C^{α} chain models were constructed using three different interaction schemes: a common Gō-like [72] model [73, 74], a physically motivated potential with desolvation barriers (db model), and a potential with a formal many-body energy term (mb model) [39]. Thermodynamic sampling were conducted by Langevin dynamics [75] using the same parameters as those in Refs. [13, 49]. Common Gō models tend to underestimate folding cooperativity of natural proteins [10, 11, 13], whereas the db and mb models entail higher degrees of folding cooperativity. Using two well-studied examples, Figure 1 provides an overview of the general trend of behaviors afforded by our three native-centric interaction schemes, in which

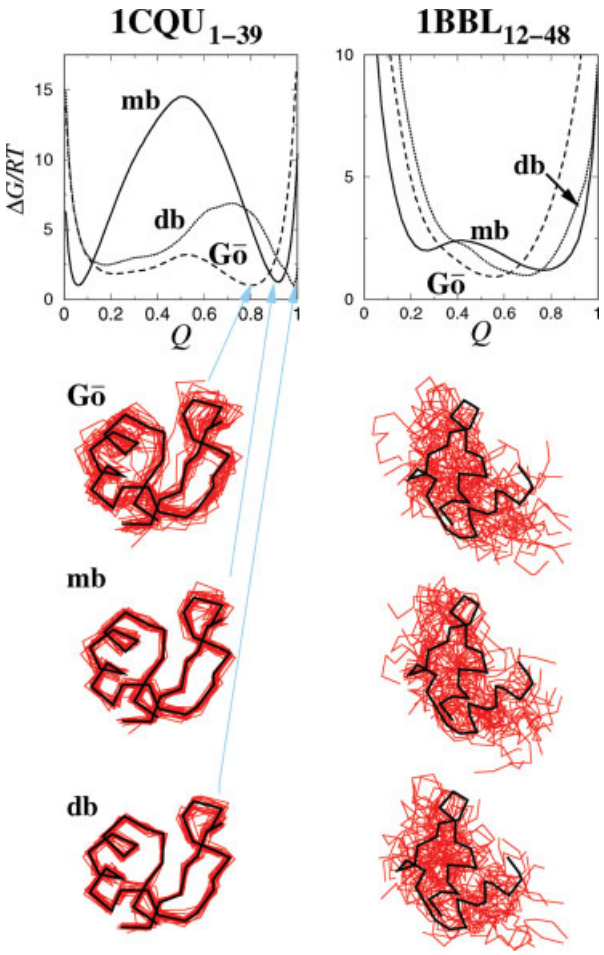

FIGURE 1. Free energy profiles and conformational fluctuations. Profiles as functions of fractional number of native contact Q (top plots) are shown near each model's folding/unfolding transition temperature for the 39-residue NTL9 fragment ($1CQU_{1-39}$) and 37-residue BBL fragment ($1BBL_{12-48}$) modeled native-centrically using the common Gō (Gō, dashed curves), desolvation-barrier (db, dotted curves) and many-body (mb, solid curves) interaction schemes. $\Delta G/RT = -\ln P(Q) + \text{constant}$, where $P(Q)$ is the probability of a conformation having Q. Note that the $\Delta G/RT$ scales for 1CQU and 1BBL are different. The $1CQU_{1-39}$ profiles are similar to that in Ref. [39]. The chain drawings depict and compare conformational fluctuations around the native (high-Q) minima for all three models of 1CQU (blue arrows) and for the mb model of 1BBL, as well as around the global minima for the Gō and db models of 1BBL. To characterize thermal fluctuation around each of these free energy minima, we denote the Q value at the minimum as Q_m and consider the ensemble of conformations around Q_m spanning the range $Q > Q_m^{+RT}$ for each model, where Q_m^{+RT} is defined as the Q value at which ΔG is RT higher, for the first time, than that at Q_m as Q decreases from Q_m in the given model ($Q_m^{+RT} < Q_m$). The root-mean-square deviation (rmsd) of the conformations [77] in these ensembles from their respective PDB structures are, for 1CQU: 1.94 Å (Gō), 0.87 Å (mb), 0.64 Å (db); and for 1BBL: 8.06 Å (Gō), 6.32 Å (mb), 7.63 Å (db). For each model, 1,000 randomly selected conformations were used to determine rmsd, and 20 randomly selected conformations are shown (red C^{α} traces) superposed upon the PDB structure (black C^{α} trace).

free energy as a function of Q is plotted in units of RT, where R is the gas constant and T is absolute temperature.

2.1. PHYSICALLY-MOTIVATED DESOLVATION/SOLVATION BARRIERS ATTENUATE CONFORMATIONAL FLUCTUATIONS

We consider the db model because desolvation and solvation effects are essential in folding and unfolding [76] (see introductory discussion in Ref. [77] and references therein). Desolvation underlies many experimentally observed biophysical properties of folding such as pressure effects, activation volumes [78], and enthalpic barriers [79, 80]. At the same time, a high solvation barrier to unfolding is seen as key to kinetic stability of proteins [81]. Here, we use an implicit-solvent approach to incorporate desolvation barriers (db's) [13, 63, 79, 82]. A detailed description of our formulation is provided in Ref. [77]. As discussed in Ref. [82], we consider temperature-independent terms with a functional form motivated by atomistically simulated (pairwise) two-methane potentials of mean force (PMF's). It should be noted that this simplified approach accounts for neither the temperature dependence of PMF's [77, 79] nor their nonadditivity [83]. In spite of these limitations, the resulting models with db's can still capture useful, essential physics of solvation and desolvation at a coarse-grained level (reviewed in Ref. [84]). For instance, these models are effective in enhancing folding cooperativity [63, 82] and in accounting for the remarkable diversity of folding rates [85] observed in experiments [52].

For the 39-residue 1CQU fragment considered in Figure 1 (left), we see that db's significantly reduce conformational fluctuation in the native state (native free energy minimum at $Q \approx 1$) vis-à-vis that allowed by the common Gō model (native free energy minimum at $Q \approx 0.8$). This salient feature of the db model is important and physically realistic; and the trend here is in line with that observed previously for db models of larger proteins including chymotrypsin inhibitor 2 (CI2) and barnase [77]. For the same protein, the db model is more constraining on native-state conformational fluctuation than the mb model (native free energy minimum at $Q \approx 0.93$), although the mb model has a higher free energy barrier and thus folds more cooperatively [39].

On the other hand, results for the 37-residue 1BBL fragment in Figure 1 (right) indicate that db's cannot enhance folding cooperativity of a native-centric model that folds globally downhill under the common Gō potential. For this protein, there is no overall free energy barrier in both the common Gō and db models. This result suggests that the ability of db's to enhance folding cooperativity is conditional and likely involves a subtle interplay with native topology. Db's can enhance folding cooperativity if a given native topology already affords a tendency to fold cooperatively (as for 1CQU). However, if the native topology is so inconducive to cooperative folding that there is no free energy barrier at the common Gō model level (as for 1BBL), adding db's would still not create an overall folding free energy barrier.

2.2. FORMAL MANY-BODY TERMS AS A CONCEPTUAL TOOL

To explore more broadly how model folding cooperativity of a given protein varies with native-centric interaction scheme, we have also considered a many-body (mb) term. Whereas potential energy E is approximately proportional to $-Q$ in the common Gō potential and the db model ($E \sim -Q$), our mb potential energy $E \sim -Q^2$ [39]. Here we use the same mb formulation that was described recently in Ref. [39]. Because Q counts pairwise contacts, the quadratic form $-Q^2$ constitutes a nonadditive mb effect by which contacts reinforce one another. The consideration of mb effect has been motivated by experimental data indicating that while the energy landscapes of many proteins are funnel-like, their high degrees of folding cooperativity imply a "near-Levinthal" scenario [86]. Because chain models with the common Gō potential are insufficient to mimick highly cooperative folding [11, 13], an emerging consensus is that mb effects beyond pairwise interactions are needed to account for the physical driving forces in small, single-domain proteins [87–90].

Physical considerations suggest that mb effects in cooperatively folding protein likely take the general form of a coupling between local and nonlocal interactions [88, 91]. As has been argued in Ref. [88], such a mechanism might arise from well-designed sidechain packing. Another possible physical origin for local-nonlocal coupling could be an enhancement of hydrogen bonding strength by hydrophobic burial [92, 93]. Local-nonlocal coupling can lead to cooperative folding behaviors reminiscent of that exhibited by small, single-domain proteins, as has

been demonstrated in extensive lattice modeling [88, 91, 93]. More recently, proteinlike cooperative folding was also achieved in an analytic theory for helix bundle folding that coupled local Zimm-Bragg zipping of helical segments with nonlocal native contact formation between two or more helices [94]. This theory is native-centric (no favorable nonnative interactions) and is an elegant formulation of a local–nonlocal coupling mechanism similar to that proposed for the lattice four-helix bundle models in Refs. [91, 93].

In seeking a deeper physical understanding, it is important to recognize that the trend of protein folding properties predicted by different forms of mb effects varies. Not all experimental trends of cooperative folding follow from mb effects per se. For instance, the correlation between native topology and folding rate [52] and the convergence of isotherms in native-state hydrogen exchange data [15, 95] were rationalized by local–nonlocal coupling but not necessarily by other forms of mb effects [12, 88, 91]. Results from recent experiments on the cooperative folding of ankyrin [96] and other repeats (reviewed in Ref. [97]) suggest strongly that local–nonlocal coupling may be a rather general mechanism operative not only in small, single-domain globular proteins but in larger repeat proteins as well.

In contrast to local–nonlocal coupling, the $E \sim -Q^2$ mb term used in this study is formal. We designate it as such beause $E \sim -Q^2$ is a mathematical construct invoked merely to increase the overall folding cooperativity of a protein chain model without regard to how cooperativity might arise from plausible physical interactions between different parts of the chain. Even so — and similar to other formal mb terms [98], our mb term is useful for conceptual exploration. However, it should always be remembered that formal mb terms such as ours (unlike the db term) was not derived from a microscopic physical perspective on how constituent parts of a protein interact. For instance, our mb model does not consider db's, which are physically important. Therefore, results from our mb model [39] and other similar constructs [98] should be interpreted with extra caution.

Figure 1 shows that our mb model is significantly more cooperative than the db model for 1CQU (left part of the figure). For this protein, the mb model produced an overall free energy barrier significantly higher than that of the db model. On the other hand, for 1BBL (Fig. 1, right), the mb interaction scheme produced only a barely existent overall free energy barrier, rendering the model marginally cooperative. Similar to the conditional ability of db's to enhance folding cooperativity, as noted above, even the mb scheme apparently cannot beget highly cooperative folding if a given protein is a global downhill folder when modeled using the common Gō potential. All in all, these results are consistent with a certain degree of robustness in the rank order of folding cooperativity across different native-centric interaction schemes [39].

TABLE I

Native contacts in the protein fragments studied in the present work.

Protein	Fragment	N	$\tilde{Q}_n$	$2\tilde{Q}_n/N$	N_N
BBL	$1BBL_{12-48}$	37	39	2.11	0.73
	$1W4H_{131-167}$	37	57	3.08	0.97
	$2CYU_{2-38}$	37	30	1.62	0.59
E3BD	$2PDD_{3-43}$	41	56	2.73	0.85
	$1W3D_{131-170}$	40	63	3.15	0.90
NTL9	$1CQU_{1-39}$	39	93	4.77	2.08
BBL	$2BTH_{126-170}$	45	66	2.93	1.09
E3BD	$1W4E_{126-170}$	45	75	3.33	1.07

N is the length of the protein fragment (number of amino acid residues); $\tilde{Q}_n$ is the number of native contacts determined according to a 4.5 Å separation cutoff (see text) between residues in the protein fragment; $2\tilde{Q}_n/N$ is the corresponding number of native contacts per amino acid residue. N_N is the number of nonlocal native contacts ($|i-j| > 6$) per amino acid residue computed using the definition of Zuo et al. with a 5.0 Å cutoff for native contacts [46]. (The N_N value for $1BBL_{12-48}$ is identical to that in Ref. [46].) The bottom two rows of data are included here for reference only. We have studied the 2BTH and 1W4E models previously [39] and they are not being further investigated in the present work.

2.3. PDB STRUCTURES USED IN NATIVE-CENTRIC MODELING

Figure 2 shows the PDB structures (thick C^α traces) of four PSBD protein fragments (1BBL [99], 1W4H [25], 2PDD [100], 1W3D [101]) and their native contacts (thin traces) used for our native-centric modeling. The numbers of native contacts in these structures are given in Table I with those in two other structures in this study (the cooperatively-folding [102] 1CQU [103] and a more recent BBL structure 2CYU [27]; drawings shown in subsequent figures) as well as two additional structures we studied previously. As in Ref. [39], only native contacts between amino acid residues i, j separated by three or more residues are considered ($|i-j| > 3$). For a given PDB

structure, a pair of residues constitutes a native contact if a pair of nonhydrogen atom, one from each residue, are less than 4.5 Å apart.

From a native-centric modeling perspective, the chain length of a PDB structure can affect model folding cooperativity via at least two effects: First, all else being equal, models with longer chain lengths tend to fold more cooperatively [11]. Second, even if one focuses only on a given fragment of a protein (i.e., the fragment length is fixed), the PDB contact pattern of the fragment, and thus the native-centric model interactions, may depend on the total chain length of the PDB structure. For instance, a longer total chain length may increase local packing and result in more native contacts. We are interested in the second effect, which is less understood. For this reason, we consider model BBL fragments of equal length (37 residues) extracted from three different PDB structures, and model E3BD fragments of essentially equal length (41 and 40 residues) extracted from two different PDB structures (see Table I). The 41-residue 2PDD fragment is identical to the one studied previously using only the common Gō potential [22]. A 40-residue 1W3D fragment is chosen here because the C^{α} position of the last residue in the 41-residue 2PDD fragment (alanine, no. 43 in the 2PDD file) was not available for 1W3D (the same alanine is labeled no. 171 in the 1W3D file). For notational simplicity, when the meaning is clear from the context of the discussion, we will omit below the subscripts specifying the ranges of residues for the fragments in Table I.

All six PDB structures used in this study were determined by NMR. For 1BBL, we used the only structure provided in its PDB file. For PDB files with multiple NMR structures, we used the "best representative conformer" (model no. 1) for all cases in which such a conformer was identified (1W4H, 1W3D, and 2CYU). For the remaining two PDB files, we used model no. 1 for 2PDD and model no. 5 for 1CQU as in our previous investigations [22, 39].

3. Results

3.1. RANK ORDERING FOLDING COOPERATIVITY BY FREE ENERGY PROFILES

Figure 3 compares the model free energy profiles of the BBL and E3BD fragments modeled using five different PDB structures. Following the approach we developed [22, 39], it is clear from the heights of the overall free energy barriers, or lack thereof, that the following rank order of folding cooperativity, 1W3D > 1W4H > 2PDD > 1BBL > 2CYU, holds consistently across the three different native-centric interaction schemes. From a modeling standpoint, this rank order of folding cooperativity is readily rationalized by an identical rank order of the corresponding number of native contacts per residue in these structures, viz., 3.15 > 3.08 > 2.73 > 2.11 > 1.62, as can be seen from the column for $2\tilde{Q}_n/N$ in Table I. A positive correlation between folding cooperativity with native contact density $2\tilde{Q}_n/N$ is consistent with our previous findings [22, 39]. This trend is essentially, though not completely, in line with a correlation between folding cooperativity and "number of nonlocal contacts per residue (N_N)" noted by Zuo et al. [46] (the ranking of 1W3D and 1W4H by $2\tilde{Q}_n/N$ was switched by N_N). It should be noted that N_N includes only $|i-j| > 6$ contacts whereas $2\tilde{Q}_n/N$ counts all contacts with $|i-j| > 3$, and that N_N was defined using a slightly larger cutoff for native contacts (see Table I). As exemplified by a recent study on Top7, the effectiveness of N_N in predicting folding cooperativity needs to be further evaluated [54].

Figure 3 shows that the 2PDD models (blue) of E3BD fold more cooperatively than the 1BBL models (red) for BBL. This result confirms and generalizes our previous finding based upon only the common Gō potential [22]. However, on the face of it, Figure 3 fails to provide an unambiguous prediction that the 40- or 41-residue fragment of E3BD folds more cooperatively than the 37-residue fragment of BBL. The apparent ambiguity is exhibited by the results that the 2PDD models (blue) for E3BD fold less cooperatively than the 1W4H models (green) for BBL, even though the E3BD models based upon 1W3D (black) fold more cooperatively than the BBL models based upon 1W4H (green).

This observation raises a basic question in native-centric modeling of protein fragments. To resolve the paradox, it is crucial to realize that the subset of native contacts in a 45-residue 1W4H structure for our 37-residue BBL fragment can differ significantly from the native contacts actually exist in an independently folded 37-residue BBL fragment. Accordingly, we posit that the higher native contact density deduced from the 1W4H structure for the 37-residue BBL fragment (see Table I) is not representative of that of the fragment as an independent entity — that is, when it is not part of a larger structure. The eight additional residues in 1W4H likely led to a folded structure with a higher average compactness than that prevails in the folded structure of an independent 37-residue BBL fragment. In

contrast, we deem the 1BBL and 2CYU structures more representative of the actual folded structure of the 37-residue BBL fragment. This is because the 37-residue fragment encompasses the entire ordered structure in 1BBL (unlike for 1W4H, no truncation of ordered structure was needed), and it is only three residues shorter than the 40-residue BBL fragment used to determined 2CYU (which provides coordinates for 39 residues). These physical considerations argue against using the 1W4H native contacts to model our 37-residue BBL fragment. Taken together, we view the available computational evidence as supportive of our expectation that the 40- or 41-residue fragment of E3BD should fold more cooperatively than the 37-residue fragment of BBL.

Beside effects of total chain length, another possible source for the appreciable differences in the native contacts of the 37-residue BBL fragments deduced from 1BBL, 2CYU, and 1W4H is the different pH conditions under which these structures were determined (pH 5.3 for 1BBL and 2CYU [27, 99] and pH 7.0 for 1W4H [25]), as has been pointed out [34, 48]. This question is worthy of further investigation but is beyond the scope of our present work.

3.2. RANK ORDERING FOLDING COOPERATIVITY BY SHARPNESS OF SIGMOIDAL TRANSITIONS

Figure 4 compares the temperature variation of Boltzmann-averaged potential energy $\langle E(T)\rangle$ in the common Gō model for 1BBL against that for 1CQU. The results show that the sigmoidal transition of $\langle E(T)\rangle$ as a function of temperature T for 1CQU is significantly sharper than that for 1BBL. As for the trend exhibited in Figure 3, this difference is readily rationalizable by native contact density. Data for $2\tilde{Q}_n/N$ in Table I show that the more cooperatively folding 1CQU has a native contact density of 4.77 (see contact pattern on the right of Fig. 4), which is more than double that of 2.11 for 1BBL [22]. As part of the 1CQU model's more sigmoidal behavior, its slope of $\langle E(T)\rangle$ outside the transition region is not as steep as that for 1BBL. This means that the construction of folded- and unfolded-state baselines (by extrapolating, respectively, from low- and high-T behaviors of $\langle E(T)\rangle$) is relatively unproblematic and less ambiguous for 1CQU than for 1BBL.

With regard to the question about truncation discussed above, the present model 39-residue fragment of 1CQU was truncated from a 56-residue NMR structure [103], and thus our model might overestimate the fragment's folding cooperativity. Despite this caveat, because the 1CQU native topology is so significantly distinct from that of the PSBD fragments, we have confidence in the validity of our prediction (which agrees with experiment) that $1CQU_{1-39}$ is a more cooperative folder than the PSBD fragments we studied [22, 39].

Figure 5 tracks the temperature-dependent folding/unfolding transition of the model proteins in Figure 2 by their normalized potential energy

$$\langle \tilde{E}(T)\rangle \equiv \frac{\langle E(T)\rangle - \langle E(T)\rangle_N}{\langle E(T)\rangle_D - \langle E(T)\rangle_N} \tag{1}$$

and average fractional number of native contacts, $\langle Q(T)\rangle$. In the above definition of $\langle \tilde{E}(T)\rangle$, $\langle E\rangle_N$ and $\langle E\rangle_D$ are both linear functions of T corresponding, respectively, to the baseline average potential energy of the native (folded) and denatured (unfolded) states estimated from the low- and high-T dependence of the $\langle E(T)\rangle$ function outside the transition region (see Fig. 4). Figure 5 shows that, irrespective of interaction scheme, the sharpness of the sigmoidal transitions of the model proteins follow the rank order 1W3D > 1W4H > 2PDD > 1BBL for both $\langle \tilde{E}(T)\rangle$ and $\langle Q(T)\rangle$.

This rank order of sharpness of sigmoidal transition in Figure 5 is identical with the rank order of folding cooperativity ascertained from the Q-based free energy profiles in Figure 3. For protein models of approximately equal chain length, as is the case here, a correlation is expected between transition sharpness and the height of overall free energy barrier — or equivalently the degree to which the transition is two-state–like [77]. However, it is important to note that sharpness per se [104] of a transition does not necessarily imply that it is two-state–like [10]. In general, sharpness tends to increase with chain length even if the transition is not two-state–like. A case in point is the sharp yet non-two-state coil-globule transition of the 62,000-unit homopolymer poly(*N*-isopropylacrylamide) [105].

3.3. RANK ORDERING FOLDING COOPERATIVITY BY THE CALORIMETRIC CRITERION

The folding cooperativities of our protein fragment models were further compared using the calorimetric criterion [3–5, 10–12]. To this end, the heat capacity function

$$C_p(T) = \frac{1}{RT^2}[\langle E^2(T)\rangle - \langle E(T)\rangle^2] \tag{2}$$

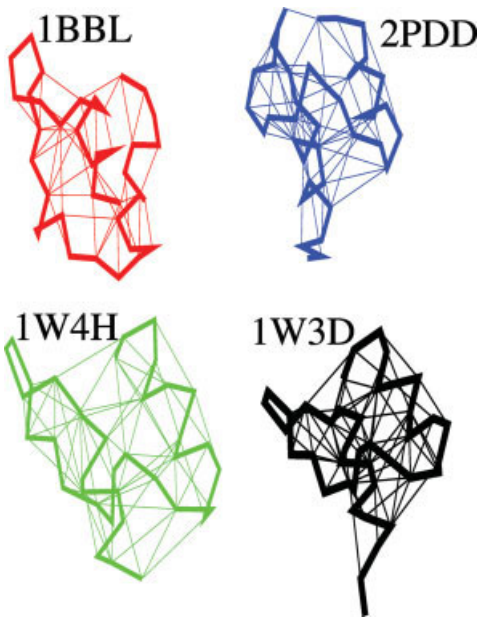

FIGURE 2. Native contacts used for modeling folding cooperativity of protein fragments. Shown here are four different PDB structures: 1BBL and 1W4H for the protein BBL, 2PDD and 1W3D for the protein E3BD. Thick lines are the native C^{α} traces of the PDB conformations chosen for modeling; thin lines joining a pair of C^{α} positions indicate native contacts defined by the 4.5Å criterion (see text). Additional information about these structures and their contacts can be found in Table I. The structures are color coded for clarity in subsequent discussion.

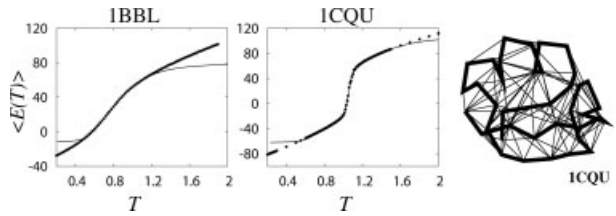

FIGURE 4. Cooperative folding entails a sharper variation of potential energy with temperature near the folding/unfolding transition temperature. Average potential energy $\langle E(T)\rangle$ as a function of temperature T was computed using the common Gō potential for the 1BBL and 1CQU fragments. Datapoints plotted were determined by direct simulations at each given temperature, whereas continuous curves were estimates by using histogram techniques based upon simulations near the transition temperature. Shown on the right is the native structure and native contacts (same style as in Fig. 2) used for $1CQU_{1-39}$ modeling in our present and previous efforts [22, 39].

was computed for every model. Examples are shown in Figure 6. In the present formulation, only the potential energy enters into the heat capacity. For our purpose, the difference between this formulation and that of Langevin dynamics studies that also included kinetic energy in the calculation of model heat capacities [22] is inconsequential. This is because it can be shown in general that insofar as the potential energy is not a function of momenta (which is the case here), at any T, the $C_p(T)$ defined in Eq. (2) using only potential energy is smaller by

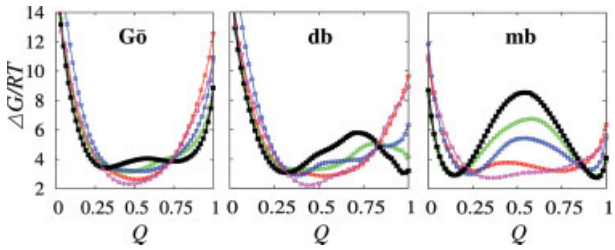

FIGURE 3. Comparing folding cooperativity of the protein fragments. Free energy profiles computed using the three native-centric interaction schemes (Gō, db, mb) are shown for 1BBL (red), 2PDD (blue), 1W4H (green), and 1W3D (black) using the color code in Figure 2. Here, $\Delta G/RT = -\ln P(Q)$. Included also for comparison are the magenta curves (with lowest minima) providing the corresponding free energy profiles for a 37-residue fragment of BBL modeled using the PDB structure 2CYU (see Table I).

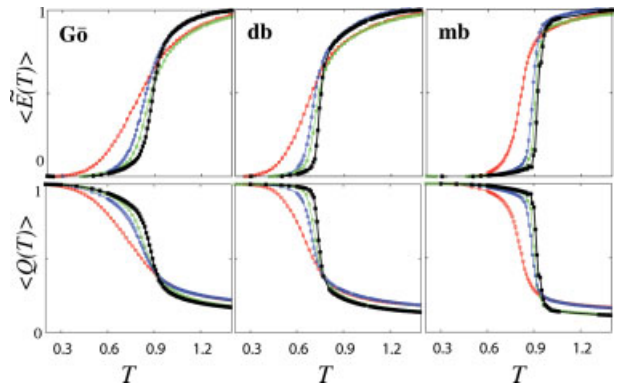

FIGURE 5. Average normalized potential energy $\langle\tilde{E}(T)\rangle$ and average fractional number of native contacts $\langle Q(T)\rangle$ as functions of temperature. Results for the three model interaction schemes (Gō, db, mb) are shown for 1BBL (red), 2PDD (blue), 1W4H (green), and 1W3D (black). Datapoints plotted for each T were obtained from direct simulations at the given T. Curves joining the datapoints are merely a guide for the eye. The normalized $\tilde{E}(T)$ was obtained from $E(T)$ using fitted baselines for the folded and unfolded states.

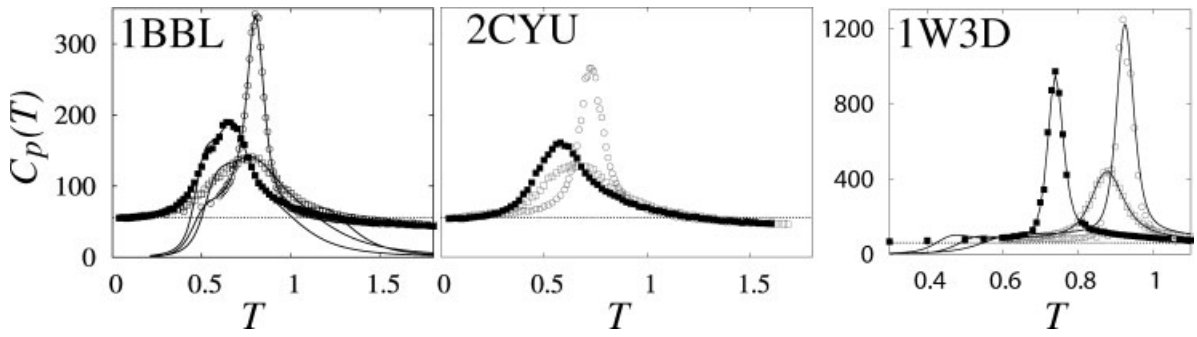

FIGURE 6. Characterizing folding cooperativity by the model heat capacity functions $C_p(T)$. Example results are shown for 1BBL, 2CYU, and 1W3D. C_p is in units of R. The C_p scale for 2CYU is identical to that for 1BBL. In all three panels, the heat capacity functions for the Gō, db, and mb models are invariably those with, respectively, the lowest, middle, and highest C_p peak; and the horizontal dotted lines marked the low-T baseline heat capacity of $3N/2$ (see text). Every plotted datapoint was obtained by direct simulation at the given temperature. Included for comparison in the 1BBL and 1W3D panels are continuous curves representing heat capacity functions estimated using histogram techniques based only upon simulations near each model's transition temperature.

a constant value of $3NR/2$ than the full heat capacity that includes kinetic energy. (N is the number of particles in the system; N is equal to the number of C^{α}'s in our models.) It should also be noted that the difference between constant-pressure and constant-volume heat capacities (C_p and C_v, respectively) is negligible for proteins under ambient conditions. Thus, even though our model heat capacity is effectively computed under constant volume, we denote our heat capacity as C_p to underscore its role as the model mimick of the constant-pressure heat capacity meausred experimentally by differential scanning calorimetry.

Figure 6 also compares directly simulated C_p values with histogram-technique estimates based on simulations near each model's transition temperature T_m at which C_p is maximum. The results indicate clearly that such histogram techniques are somewhat reliable outside the transition regime only when the model is relatively cooperative (1W3D) but are much less reliable when the model is non-cooperative (1BBL). In view of these test results, only directly simulated C_p's are used in the present analysis.

The calorimetric criterion for two-state–like protein folding requires $\Delta H_{vH}/\Delta H_{cal} \approx 1$ [3–5, 10]. In this expression, $\Delta H_{cal} \equiv \int dT C_p(T)$ is the calorimetric enthalpy and

$$\Delta H_{vH} = 2\sqrt{RT_m^2 C_p(T_m)} \qquad (3)$$

is the van't Hoff enthalpy determined for the C_p maximum in accordance with the κ_2 and $\kappa_2^{(s)}$ definitions for $\Delta H_{vH}/\Delta H_{cal}$ in Ref. [11]. As in experimental calorimetry, the determination of a model $\Delta H_{vH}/\Delta H_{cal}$ ratio suitable for quantifying folding cooperativity often necessitates baseline subtractions [11]. In this regard, the simulation results (see e.g., Fig. 6) suggested that a folded-state baseline of $C_p = 3NR/2$ may be appropriate for our models. Fundamentally, this is because the behavior of every one of our models at low temperatures ($T \to 0$) resembles that of $3N$ one-dimensional harmonic oscillators, which entails a full heat capacity of $3NR$ [106, 107]. Thus, after subtracting the $3NR/2$ contribution from kinetic energy precluded in Eq. (2), a low-T baseline of $3NR/2$ ensued. This is the folded-state C_p baseline we adopted for all of our models. On the other hand, the unfolded-state C_p baseline was not the same for different models. For unfolded states, we considered C_p values above T_m separately for each model and used the directly-simulated C_p values at the 20 highest T's investigated for the given model to construct the unfolded-state C_p baseline for the model.

Table II shows that 1CQU is most cooperative among the protein fragments studied here, irrespective of the native-centric interaction scheme used. Even so, because of its small size, models of $1CQU_{1-39}$ appear less cooperative than longer protein chains such as CI2. Whereas db models of CI2 exhibit $\Delta H_{vH}/\Delta H_{cal} = \kappa_2^{(s)} \approx 1$ [77], the corresponding $\kappa_2^{(s)}$ value for the present db $1CQU_{1-39}$ model is only 0.78. As reported in Table II, for $1CQU_{1-39}$, a more cooperative mb interaction scheme is needed for $\kappa_2^{(s)} \approx 1$. Models for the PSBD fragments

TABLE II

Folding/unfolding cooperativity of the model protein fragments.

Protein fragment	$\Delta H_{vH}/\Delta H_{cal}$		
	Gō	db	mb
$1BBL_{12-48}$	0.36	0.31	0.55
$1W4H_{131-167}$	0.44	0.47	0.85
$2CYU_{2-38}$	0.38	0.37	0.56
$2PDD_{3-43}$	0.46	0.48	0.76
$1W3D_{131-170}$	0.60	0.65	0.84
$1CQU_{1-39}$	0.73	0.78	0.97

Cooperativity is quantified by the van't Hoff versus calorimetric enthalpy ratio $\Delta H_{vH}/\Delta H_{cal}$, computed here as the $\kappa_2^{(s)}$ parameter in Refs. [11, 22, 93] with baseline subtractions (see text). Results shown are for native-centric C^{α} models of the protein fragments using, respectively, the common Gō (Gō), desolvation-barrier (db), and many-body (mb) interaction schemes.

are even less cooperative. None of their computed $\Delta H_{vH}/\Delta H_{cal}$ values is higher than 0.85 even after reasonable baseline subtractions as described earlier.

Consider the expression for ΔH_{vH} in Eq. (3) before baseline subtractions. Clearly, ΔH_{vH} increases with increasing T_m and with increasing peak heat capacity value $C_p(T_m)$. For the five PSBD models, we found that both T_m and $C_p(T_m)$ follow the rank order 1W3D > 1W4H > 2PDD > 1BBL > 2CYU (detailed data not shown). This rank order is identical to the rank order of folding cooperativity deduced from the free energy profiles in Figure 3 and rationalized above by the native contact densities in Table I. This coincidence of rank order of E-based and Q-based cooperativity parameters is not too surprising given the native-centric nature of the model interactions. Nonetheless, it underscores the robustness of these constructs as predictive models for folding cooperativity.

After baseline subtractions, Table II shows that the resulting $\kappa_2^{(s)}$ values of the PSBD models follow the rank order 1W3D > 2PDD > 1W4H > 2CYU > 1BBL for the common Gō and db models, and 1W4H > 1W3D > 2PDD > 2CYU > 1BBL for the mb models. For the common Gō and db models, the $\kappa_2^{(s)}$ values of 1W4H and 2PDD are essentially the same, and the $\kappa_2^{(s)}$ for 1BBL and 2CYU also take quite similar values. The $\kappa_2^{(s)}$ values for 1BBL and 2CYU are essentially identical for their mb models as well. However, for the mb interaction scheme, the $\kappa_2^{(s)}$ values for 1W4H and 1W3D are approximately equal, which is not the case for the common Gō and db interaction schemes.

Two clear trends emerged despite these subtle effects of baseline subtractions, a procedure that is inevitably somewhat subjective [11] both experimentally and theoretically (note that different sets of baselines were used in Ref. [22]). First, recall the results in Figures 3 and 5 indicating that model of a fragment based upon truncation of a larger PDB structure tends to fold more cooperativity, a trend we rationalized by chain-length effects on packing (see above). Consistent with this trend and irrespective of native interaction scheme, Table II shows that for E3BD, the 1W3D model is calorimetrically more cooperative than the 2PDD model. Similarly, for BBL, the 1W4H model is calorimetrically more cooperative than the 1BBL and 2CYU models. Second, Table II demonstrates that, by the calorimetric criterion, the 41-residue 2PDD model is consistently more cooperative than the 37-residue 1BBL or 2CYU model. Thus, the heat capacity consideration here is seen to provide additional theoretical support to the contention that the 41-residue E3BD fragment should fold more cooperatively than the 37-residue BBL fragment.

3.4. DISTRIBUTION OF RADIUS OF GYRATION AND FOLDING COOPERATIVITY

Figure 7 provides the distribution of radius of gyration for all the protein fragment models studied in this work, simulated near their respective transition temperature T_m. Raidus of gyration $R_g = \sqrt{\sum_{j=1}^{N} |\mathbf{r}_j - \langle \mathbf{r} \rangle|^2 / N}$ where $\mathbf{r}_j$ is the position vector of the jth C^{α} and $\langle \mathbf{r} \rangle \equiv \sum_{j=1}^{N} \mathbf{r}_j / N$. Results in Figure 7 show that more cooperatively folding models tend to have bimodal R_g distributions [$P(R_g)$'s] whereas less cooperatively or noncooperatively folding models have unimodal $P(R_g)$'s. For the least cooperative models 2CYU and 1BBL, the mb interaction scheme is needed for a bimodal $P(R_g)$. For the other four models whose native topologies are more conducive to cooperative folding, $P(R_g)$ begins to become bimodal at the level of the db interaction scheme. Expectedly, the more cooperative a model – which means that the model's folded and unfolded conformations are more clearly separated in E or Q, the more bimodal is its $P(R_g)$. Thus, as has been proposed [22], if single-molecule measurement of R_g is possible, $P(R_g)$ may provide a definitive means to distinguish between

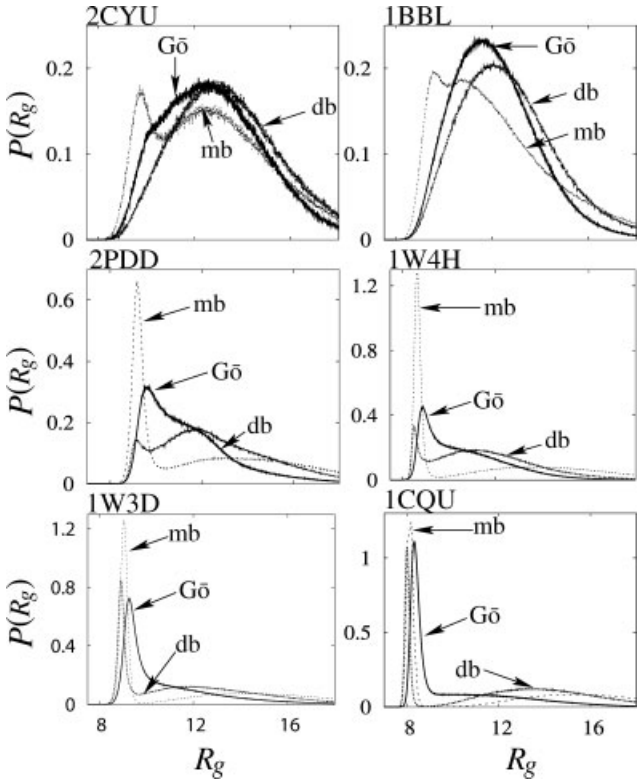

FIGURE 7. Distribution $P(R_g)$ of radius of gyration R_g for all of the model protein fragments studied in this work. Each distribution was computed at the transition temperature of the given model.

two-state–like, bimodal cooperative folding and one-state, unimodal, noncooperative folding.

4. Discussion

4.1. EFFECT OF CHAIN LENGTH ON PACKING DENSITY

The above-noted impact of overall chain length of a protein on the native contact (packing) density of a truncated fragment is an effect that deserves further examination. To address this issue, Figure 8 compares the native contact maps of the same 37-residue sequence truncated from 1BBL and 2CYU (wherein contacts between residues i,j in their respective PDB structures are marked). Even though the 2CYU and 1BBL structures were determined by NMR under similar solvent conditions (both at pH 5.3, 2CYU was determined at 278 K, whereas 1BBL was determined at 288 K and 298 K) [27, 99], there are significantly fewer contacts in the 37-residue 2CYU fragment than in the corresponding 1BBL fragment, resulting in a native contact density for $2CYU_{2-38}$ ($2\tilde{Q}_n/N = 1.62$), that is, 23% lower than that for $1BBL_{12-48}$ ($2\tilde{Q}_n/N = 2.11$; see Table I). Because the NMR structural determinations for 2CYU and 1BBL were performed on peptides with total lengths of 40 and 51 residues, respectively, the present observation suggests strongly that a longer overall chain length tends to increase native contact density. This hypothesis is consistent with a recent result [48] showing that native-centric models with different levels of mb interactions based on a 39-residue truncation of 1W4H (termed "truncated-Fersht") folded more cooperatively than the corresponding models based upon the 39-residue 2CYU (termed "Muñoz").[1] In this regard, the comparison in the present Figure 8 may be even more revealing and provide stronger support for the chain length-packing hypothesis because 1BBL and 2CYU were determined under more similar conditions, whereas 1W4H was determined under a different pH of 7.0 [25, 48].

4.2. SENSITIVITY OF NATIVE-CENTRIC MODELING TO CHOICE OF CONTACT SET: QUESTIONS OF ROBUSTNESS

Next we turn to a perennial question in native-centric modeling: To what extent are our conclusions robust with respect to how we define native contacts from a given PDB structure? As a matter of principle, our group has long advocated the need to explore physically plausible variations in native-centric formulations as a means to avoid over-interpretation and to ascertain robust conclusions [13]. We have also paid attention to the ramification of multiple NMR-determined PDB structures on native-centric modeling of putative downhill folding [39]. Modeling issues related to multiple PDB structures have recently been pursued in other contexts as well [108, 109]. For instance, a comparison was made between the Gō model behaviors (no consideration of db and mb effects) of immunoglobulin-binding domain of streptococcal protein G based upon an average structure from NMR (PDB ID: 2GB1) and those based

[1]The criterion for native contact was more permissive in Cho et al. [48], with "a distance cutoff of 9.0 Å." Accordingly, the numbers of native contacts, $\tilde{Q}_n$, in these authors' "Muñoz," "Fersht," and "truncated-Fersht" models are, respectively, 61, 117, and 101 [Fig. 2(b) in Ref. [48]]. Because $N = 39$, 45, and 39, their $2\tilde{Q}_n/N =$ 3.13, 5.2, and 5.18, respectively. The corresponding N_N computed using the definition in Ref. [46] (with a different criterion for native contacts) are 0.56, 0.98, and 0.95, respectively.

on a structure from X-ray crystallography (PDB ID: 1PGB) [108].

As stated, results presented so far in this work were obtained using a 4.5 Å cutoff between nonhydrogen atoms to extract native contacts from PDB structures.[2] To address the robustness question, we have also studied alternate models for the four PDB structures in Figure 2, now constructed with native contacts extracted using a 6.0 Å cutoff. With this more permissive criterion, there are more native contacts in the alternate models than that depicted by the thin lines in Figure 2. Specifically, the numbers of native contacts for 1BBL, 1W4H, 2PDD, and 1W3D increase from those in Table I to 74, 100, 99, and 106, respectively.

We compared the free energy profiles for these model fragments in Figure 3 (native contacts defined by a 4.5 Å cutoff) with those constructed using native contacts defined by a 6.0 Å cutoff (examples shown in Fig. 9). By inspecting the free energy profiles, we found that the above rank order of folding cooperativity 1W3D > 1W4H > 2PDD > 1BBL remains invariant under the more permissive criterion for native contacts. With the 6.0 Å native-contact cutoff, the overall free energy barrier height for every model increases a little relative to that for the corresponding model with the 4.5 Å native-contact cutoff. But the differences in behavior between the BBL models (1BBL and 1W4H) and that between the E3BD models (2PDD and 1W3D) remain largely unchanged by the shift in native-contact cutoff. The only notable exceptions are those in Figure 9, which shows the creation of an overall free energy barrier for 2PDD comparable in height to that for 1W3D in the db interaction scheme, and the narrowing of the difference in overall free energy barrier height between 2PDD and 1W3D in the mb interaction scheme (cf. corresponding profiles in Fig. 3). Nonetheless, albeit the overall free energy barrier heights for 2PDD and 1W4H become almost identical in the db interaction scheme with the 6.0 Å native-contact cutoff, the overall free energy barrier heights in both the db and mb interaction schemes still follow the rank order 1W3D > 1W4H > 2PDD > 1BBL, lending credence to the robustness of this predicted trend of behavior.

[2]We note that folding rates of the four BBL mutants each with a single tryptophan substitution in Ref. [35] (PDB structures for these new mutants are not yet available) are similar to that of the BBL H166W mutant 2BTH in Ref. [25], suggesting that the overall folding free energy barrier heights are similar for the new mutants and for 2BTH.

4.3. IMPORTANCE OF EXPLICIT-CHAIN MODELING IN ADDRESSING PHYSICS OF FOLDING COOPERATIVITY

As has been touched upon in the Introduction, besides explicit-chain simulations [22, 39, 46–48, 58, 59], nonexplicit-chain Ising-like constructs [55] have recently been applied also to study putative downhill folding [56, 57]. However, key conclusions reached in some of these nonexplicit-chain studies are inconsistent with those obtained from explicit-chain modeling. For instance, using an extended version of the Ising-like formulation of Muñoz and Eaton [55], Yu et al. asserted that BBL folds cooperatively. In particular, their computed chevron plots for models based upon the 1BBL and 1W4E structures had linear arms that showed no rollover [57], directly contradicting results from our explicit-chain native-centric models based upon the same PDB structures [22, 39].

We are of the opinion that some of the main results of Ising-like modeling of protein folding are artifactual. Although Ising-like models are useful for rank ordering folding rates and thus folding cooperativities [55], these nonexplicit-chain constructs often preclude existent physical interactions that are detrimental to two-state–like folding. Consequently, these constructs tend to overestimate folding cooperativity, for the following specific reasons. First, in these constructs, native contacts are deemed favorable only when all the residues (peptide bonds) between them are in the native state. As pointed out by Bruscolini et al. [56], this is a drastic assumption. It precludes many conformations of intermediate compactness, thus artifically enhancing folding cooperativity. In this regard, this assumption is very similar to that of the COREX algorithm [110, 111], which boosts global folding cooperativity by presuming all-or-none order-disorder conformational transitions of local chain segments (see pages 554–557 of Ref. [10]). Second, applications of master-equation formulation to a small number of conformational states constructed by grouping conformations with similar properties tend to minimize trapping and landscape ruggedness effects, and thus lead to an artificial increase in kinetic cooperativity. Consider the two-dimensional lattice HP+ model introduced in Ref. [112]. This model does not satisfy the calorimetric two-state criterion, as is evident from its exact heat capacity function in Figure 5 of Ref. [10] and the calculation of the resulting $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ in the same reference. However, a master-equation approximation of the folding kinetics based on grouping

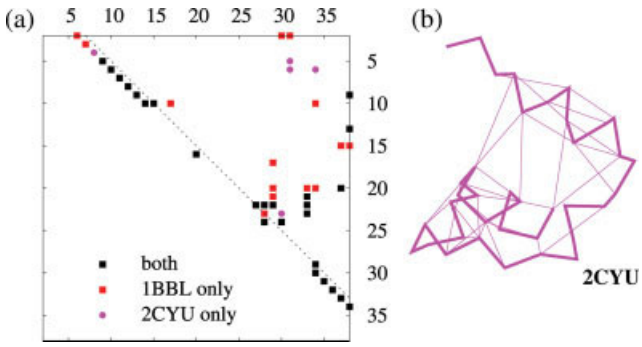

FIGURE 8. Comparing native contacts from two different PDB structures for BBL. (a) Contact map showing 25 contacts common to both 1BBL and 2CYU (black squares), 14 contacts in 1BBL but not in 2CYU (red squares), and 5 contacts in 2CYU but not in 1BBL (magenta circles). As in Ref. [39], the $|i - j| = 5$ dashed diagonal may be used to demarcate between local and nonlocal contacts. (b) Native structure and native contacts used here to model $2CYU_{2-38}$ (same style as in Fig. 2).

of conformations [113] yielded a chevron plot with no rollovers for this model (Fig. 7, upper left, of Ref. [2]). This shows that master-equation approximations can produce linear chevron plots even for cases in which the underlying models are thermodynamically non-two-state. Furthermore, for the explicit-chain HP+ model itself, the linear chevron plot obtained by master-equation approximation was also likely inaccurate because direct simulations have demonstrated that even thermodynamically much more two-state–like three-dimensional Gō models have chevron rollovers [13, 114, 115]. In view of these basic mismatches between physical expectations on one hand, and Ising-like model assumptions as well as master-equation kinetics on the other, predictions of folding cooperativity of individual proteins based solely on nonexplicit-chain modeling cannot be considered conclusive in the absence of explicit-chain evaluation.

5. Concluding Remarks

Using multipe native-centric interaction schemes, we have documented a diverse range of cooperative, marginally cooperative and noncooperative folding behaviors among model protein fragments with ∼40 residues. Our results once again underscore that significant information regarding how cooperatively a protein may fold can be gleaned from its native contact pattern. As manifested in their sometimes subtly different native topologies, the folding cooperativity or noncooperativity of homologous proteins such as E3BD and BBL can be quite different. More generally, members of a protein family sharing the same fold can have significantly different packing densities in order to, perhaps, serve different biological functions [116]. Therefore, the folding cooperativity or lack of folding cooperativity of a member of a protein family may not provide accurate inference for the cooperative or noncooperative behavior of another member of the same family. Not surprisingly, folding cooperativity of small protein fragments are sensitive to chain length. For instance, experiments have shown clearly that a 33-residue E3BD fragment folds less cooperatively than either a 41- or

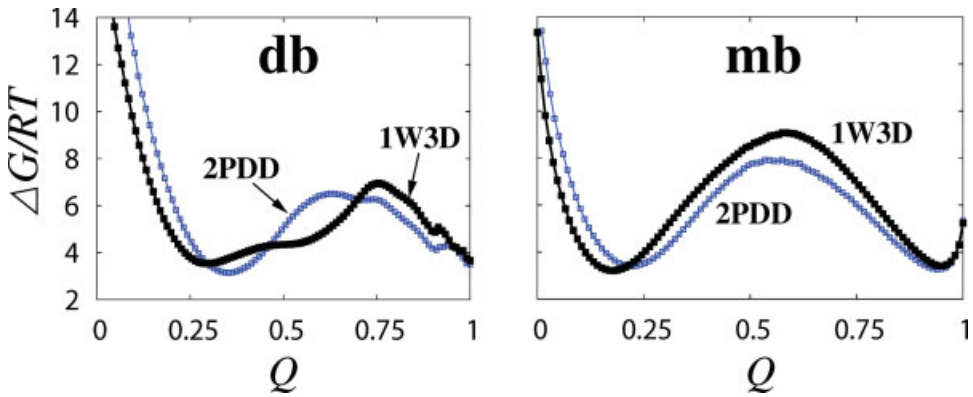

FIGURE 9. Free energy profiles under a varied definition of native contacts. Shown are $\Delta G/RT$ versus Q plots for 2PDD (blue) and 1W3D (black) computed using a 6.0 Å contact criterion (see text) instead of the 4.5 Å criterion used for results presented in all other figures in this article. The profiles are for the db (left) and mb (right) models at each model's transition temperature.

a 36-residue E3BD fragment [64]. We found that total chain length can affect native contact density as well. For the cases investigated, longer total chain length increases native contact density. This observation implies that in the construction of native-centric models, it should be more appropriate to extract native contacts from PDB structures that require least truncation. Taking the present results for 40- or 41-residue E3BD fragments and 37-residue BBL fragments together with our previous results for 45-residue E3BD and BBL fragments [39] — results that agree with experiments [25], it suggests quite strongly that an E3BD fragment always folds more cooperatively than a homologous BBL fragment of approximately equal chain length.

ACKNOWLEDGMENTS

We are grateful to the organizers of the 6th Congress of the International Society for Theoretical Chemical Physics for their invitation to present part of the work reported in this article in Vancouver, B.C. in July 2008. Some of our results were also presented in preliminary form at the Conference on Structure and Dynamics in Soft Matter and Biomolecules held at the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy in June 2007. We thank Yawen Bai, Allison Ferguson, Neil Ferguson, Alan Fersht, Martin Gruebele, Michael Knott, Victor Muñoz, Jose Sanchez-Ruiz, Eugene Shakhnovich, Stefan Wallin, and Zhuqing Zhang for helpful discussions, and Zhuqing Zhang for double checking some of the contact counts. H.S.C. holds a Canada Research Chair in Proteomics, Bioinformatics, and Functional Genomics.

References

1. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Proteins Struct Funct Genet 1995, 21, 167.
2. Dill, K. A.; Chan, H. S. Nat Struct Biol 1997, 4, 10.
3. Lumry, R.; Biltonen, R.; Brandts, J. F. Biopolymers 1966, 4, 917.
4. Privalov, P. L.; Khechinashvili, N. N. J Mol Biol 1974, 86, 665.
5. Privalov, P. L. Adv Protein Chem 1979, 33, 167.
6. Kim, P. S.; Baldwin, R. L. Annu Rev Biochem 1990, 59, 631.
7. Jackson, S. E.; Fersht, A. R. Biochemistry 1991, 30, 10428.
8. Baker, D. Nature 2000, 405, 39.
9. Barrick, D. Phys Biol 2009, 6, 015001.
10. Chan, H. S. Proteins Struct Funct Genet 2000, 40, 543.
11. Kaya, H.; Chan, H. S. Proteins Struct Funct Genet 2000, 40, 637 (Erratum: 2001, 43, 523).
12. Chan, H. S.; Shimizu, S.; Kaya, H. Methods Enzymol 2004, 380, 350.
13. Kaya, H.; Chan, H. S. J Mol Biol 2003, 326, 911 (Corrigendum: 2004, 337, 1069).
14. Scalley-Kim, M.; Baker, D. J Mol Biol 2004, 338, 573.
15. Bai, Y. Biochem Biophys Res Comm 2006, 340, 976.
16. Watters, A. L.; Deka, P.; Corrent, C.; Callender, D.; Varani, G.; Sosnick, T.; Baker, D. Cell 2007, 128, 613.
17. Sabelko, J.; Ervin, J.; Gruebele, M. Proc Natl Acad Sci USA 1999, 96, 6031.
18. Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V. Science 2002, 298, 2191.
19. Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi, M.; Muñoz, V. Biochemistry 2006, 45, 8466.
20. Liu, F.; Gruebele, M. Chem Phys Lett 2008, 461, 1.
21. Yang, W. Y.; Grueble, M. Nature 2003, 423, 193.
22. Knott, M.; Chan, H. S. Proteins Struct Funct Bioinform 2006, 65, 373.
23. Muñoz, V. Annu Rev Biophys Biomol Struct 2007, 36, 395.
24. Ferguson, N.; Schartau, P. J.; Sharpe, T. D.; Sato, S.; Fersht, A. R. J Mol Biol 2004, 344, 295.
25. Ferguson, N.; Sharpe, T. D.; Schartau, P. J.; Sato, S.; Allen, M. D.; Johnson, C. M.; Rutherford, T. J.; Fersht, A. R. J Mol Biol 2005, 353, 427.
26. Naganathan, A. N.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. Biochemistry 2005, 44, 7435.
27. Sadqi, M.; Fushman, D.; Muñoz, V. Nature 2006, 442, 317.
28. Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L.; Fersht, A. R. Proc Natl Acad Sci USA 2007, 104, 123.
29. Ferguson, N.; Sharpe, T. D.; Johnson, C. M.; Schartau, P. J.; Fersht, A. R. Nature 2007, 445, E14.
30. Zhou, Z.; Bai, Y. Nature 2007, 445, E16.
31. Sadqi, M.; Fushman, D.; Muñoz, V. Nature 2007, 445, E17.
32. Sharpe, T. D.; Ferguson, N.; Johnson, C. M.; Fersht, A. R. J Mol Biol 2008, 383, 224 (Corrigendum: 2009, 387, 519).
33. Li, P.; Oliva, F. Y.; Naganathan, A. N.; Muñoz, V. Proc Natl Acad Sci 2009, 106, 103.
34. Arbelya, E.; Rutherford, T. J.; Sharpe, T. D.; Ferguson, N.; Fersht, A. R. J Mol Biol 2009, 387, 986.
35. Neuweiler, H.; Sharpe, T. D.; Johnson, C. M.; Teufel, D. P.; Ferguson, N.; Fersht, A. R. J Mol Biol 2009, 387, 975.
36. Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanchez-Ruiz, J. M.; Muñoz, V. J Am Chem Soc 2008, 130, 7489.
37. Liu, F.; Gruebele, M. J Mol Biol 2007, 370, 574.
38. Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. Proc Natl Acad Sci USA 2008, 105, 2369.
39. Badasyan, A.; Liu, Z.; Chan, H. S. J Mol Biol 2008, 384, 512.
40. Tompa, P.; Fuxreiter, M. Trends Biochem Sci 2007, 33, 2.
41. Mittag, T.; Forman-Kay, J. D. Curr Opin Struct Biol 2007, 17, 3.

42. Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S. Proc Natl Acad Sci USA 2007, 104, 9650.
43. Lu, Q.; Lu, H. P.; Wang, J. Phys Rev Lett 2007, 98, 128105.
44. Amitai, G.; Gupta, R. D.; Tawfik, D. S. HFSP J 2007, 1, 67.
45. Wroe, R.; Chan, H. S.; Bornberg-Bauer, E. HFSP J, 2007, 1, 79.
46. Zuo, G.; Wang, J.; Wang, W. Proteins Struct Funct Bioinform 2006, 63, 165.
47. Prieto, L.; Rey, A. J Chem Phys 2007, 127, 175101.
48. Cho, S. S.; Weinkam, P.; Wolynes, P. G. Proc Natl Acad Sci USA 2008, 105, 118.
49. Wallin, S.; Chan, H. S. J Phys Condens Matt 2006, 18, S307.
50. Gō, N.; Taketomi, H. Proc Natl Acad Sci USA 1978, 75, 559.
51. Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. J Mol Biol 1998, 252, 460.
52. Plaxco, K. W.; Simons, K. T.; Baker, D. J Mol Biol 1998, 227, 985.
53. Chan, H. S. Nature 1998, 392, 761.
54. Zhang, Z.; Chan, H. S. Biophys J 2009, 96, L25.
55. Muñoz, V.; Eaton, W. A. Proc Natl Acad Sci 1999, 96, 11311.
56. Bruscolini, P.; Pelizzola, A.; Zamparo, M. J Chem Phys 2007, 126, 215103.
57. Yu, W.; Chung, K.; Cheon, M.; Heo, M.; Han, K.-H.; Ham, S.;, Chang I. Proc Natl Acad Sci USA 2008, 105, 2397.
58. Zhang, J.; Li, W.; Wang, J.; Qin, M.; Wang, W. Proteins Struct Funct Bioinform 2008, 72, 1038.
59. Pitera, J. W.; Swope, W. C.; Abraham, F. F. Biophys J 2008, 94, 4837.
60. Settanni, G.; Fersht, A. R. J Mol Biol 2009, 387, 993.
61. García, A. E.; Onuchic, J. N. Proc Natl Acad Sci USA 2003, 100, 13898.
62. Bai, Y.; Karimi, A.; Dyson, H. J.; Wright, P. E. Protein Sci 1997, 6, 1449.
63. Kaya, H.; Liu, Z.; Chan, H. S. Biophys J 2005, 89, 520.
64. Spector, S.; Young, P.; Raleigh, D. P. Biochemistry 1999, 38, 4128.
65. Muñov, V.; Sanchez-Ruiz, J. M. Proc Natl Acad Sci USA 2004, 101, 17646.
66. Godoy-Ruiz, R.; Henry, E. R.; Kubelka, J.; Hofrichter, J.; Muñoz, V.; Sanchez-Ruiz, J. M.; Eaton, W. A. J Phys Chem B 2008, 112, 5938.
67. Capaldi, A. P.; Kleanthous, C.; Radford, S. E. Nat Struct Biol 2002, 9, 209.
68. Feng, H.; Takei, J.; Lipsitz, R.; Tjandra, N.; Bai, Y. Biochemistry 2003, 42, 12461.
69. Cho, J. H.; Sato, S.; Raleigh, D. P. J Mol Biol 2004, 338, 827.
70. Zarrine-Afsar, A.; Wallin, S.; Neculai, A. M.; Neudecker, P.; Howell, P. L.; Davidson, A. R.; Chan, H. S. Proc Natl Acad Sci USA 2008, 105, 9999.
71. Chan, H. S.; Zhang, Z. J Biol 2009, 8, 27.
72. Taketomi, H.; Ueda, Y.; Gō, N. Int J Pept Protein Res 1975, 7, 445.
73. Micheletti, C.; Banavar, J. R.; Maritan, A.; Seno, F. Phys Rev Lett 1999, 82, 3372.
74. Clementi, C.; Nymeyer, H.; Onuchic, J. N. J Mol Biol 2000, 298, 937.
75. Veitshans, T.; Klimov, D.; Thirumalai, D. Fold Des 1997, 2, 1.
76. Sheinerman, F. B.; Brooks, C. L. Proc Natl Acad Sci USA 1998, 95, 1562.
77. Liu, Z.; Chan, H. S. Phys Biol 2005, 2, S75.
78. Hummer, G.; Garde, S.; García, A. E.; Paulaitis, M. E.; Pratt, L. R. Proc Natl Acad Sci USA 1998, 95, 1552.
79. Liu, Z.; Chan, H. S. J Mol Biol 2005, 349, 872.
80. MacCallum, J. L.; Sabaye Moghaddam, M.; Chan, H. S.; Tieleman, D. P. Proc Natl Acad Sci USA 2007, 104, 6206 (Correction: 2008, 105, 19561).
81. Rodriguez-Larrea, D.; Minning, S.; Borchert, T. V.; Sanchez-Ruiz, J. M. J Mol Biol 2006, 360, 715.
82. Cheung, M. S.; García, A. E.; Onuchic, J. N. Proc Natl Acad Sci USA 2002, 99, 685.
83. Moghaddam, M. S.; Shimizu, S.; Chan. H. S. J Am Chem Soc 2005, 127, 303 (Correction: 2005, 127, 2363).
84. Levy, Y.; Onuchic, J. N. Annu Rev Biophys Biomol Struct 2006, 35, 389.
85. Ferguson, A.; Liu, Z.; Chan, H. S. J Mol Biol 2009, DOI: 10.1016/j.jmb.2009.04.011.
86. Kaya, H.; Chan, H. S. Proteins Struct Funct Genet 2003, 52, 510.
87. Jewett, A. I.; Pande, V. S.; Plaxco, K. W. J Mol Biol 2003, 326, 247.
88. Kaya, H.; Chan, H. S. Proteins Struct Funct Genet 2003, 52, 524.
89. Ejtehadi, M. R.; Avall, S. P.; Plotkin, S. S. Proc Natl Acad Sci USA 2004, 101, 15088.
90. Qi, X.; Portman, J. J. Proc Natl Acad Sci USA 2007, 104, 10841.
91. Kaya, H.; Chan, H. S. Proteins Struct Funct Bioinform 2005, 58, 31.
92. Takada, S.; Luthey-Schulten, Z.; Wolynes, P. G. J Chem Phys 1999, 110, 11616.
93. Kaya, H.; Chan, H. S. Phys Rev Lett 2000, 85, 4823.
94. Ghosh, K.; Dill, K. A. J Am Chem Soc 2009, 131, 2306.
95. Bai, Y.; Sosnick, T.; Mayne, L.; Englander, S. W. Science 1995, 269, 192.
96. Mello, C. C.; Barrick, D. Proc Natl Acad Sci USA 2004, 101, 14102.
97. Kloss, E.; Courtemanche, N.; Barrick, D. Arch Biochem Biophys 2008, 469, 83.
98. Eastwood, M. P.; Wolynes, P. G. J Chem Phys 2001, 114, 4702.
99. Robien, M. A.; Clore, G. M.; Omichinski, J. G.; Perham, R. N.; Appella, E.; Sakaguchi, K.; Gronenborn, A. M. Biochemistry 1992, 31, 3463.
100. Kalia, Y. N.; Brocklehurst, S. M.; Hipps, D. S.; Appella, E.; Sakaguchi, K.; Perham, R. N. J Mol Biol 1993, 230, 323.
101. Allen, M. D.; Broadhurst, R. W.; Solomon, R. G.; Perham, R. N. FEBS J 2005, 272, 259.
102. Horng, J.-C.; Moroz, V.; Raleigh, D. P. J Mol Biol 2003, 326, 1261.
103. Luisi, D. L.; Kuhlman, B.; Sideras, K.; Evans, P. A.; Raleigh, D. P. J Mol Biol 1999, 289, 167.
104. Klimov, D. K.; Thirumalai, D. Fold Des 1998, 3, 127.
105. Tiktopulo, E. I.; Bychkova, V. E.; Rička, J.; Ptitsyn, O. B. Macromolecules 1994, 27, 2879.

106. Landau, L. D.; Lifshitz, E. M. Course of Theoretical Physics; Pergamon: London, 1976; Vol. 1 (Mechanics), Chapter 2.
107. Landau, L. D.; Lifshitz, E. M. Course of Theoretical Physics; Pergamon: Oxford and New York, 1980; Vol. 5 (Statistical Physics), Chapter 6.
108. Prieto, L.; Rey, A. J Chem Phys 2008, 129, 115101.
109. Rey-Stolle, M. F.; Enciso, M.; Rey, A. J Comput Chem 2009, 30, 1212.
110. Freire, E.; Murphy, K. P. J Mol Biol 1991, 222, 687.
111. Hilser, V. J.; Freire, E. J Mol Biol 1996, 262, 756.
112. Chan, H. S.; Dill, K. A. Proteins Struct Funct Genet 1998, 30, 2.
113. Chan, H. S.; Dill, K. A. J Chem Phys 1994, 100, 9238.
114. Kaya, H.; Chan, H. S. Phys Rev Lett 2003, 90, 258104.
115. Zhou, Y.; Zhang, C.; Stell, G.; Wang, J. J Am Chem Soc 2003, 125, 6300.
116. Hills, R. D.; Brooks, C. L. Biophys J 2008, 95, L57.