Contact Order Dependent Protein Folding Rates: Kinetic Consequences of a Cooperative Interplay between Favorable Nonlocal Interactions and Local Conformational Preferences

Hüseyin Kaya and Hue Sun Chan[*]
Protein Engineering Network of Centres of Excellence (PENCE), Department of Biochemistry, and Department of Medical Genetics and Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ABSTRACT Physical mechanisms underlying the empirical correlation between relative contact order (CO) and folding rate among naturally occurring small single-domain proteins are investigated by evaluating postulated interaction schemes for a set of three-dimensional 27mer lattice protein models with 97 different CO values. Many-body interactions are constructed such that contact energies become more favorable when short chain segments sequentially adjacent to the contacting residues adopt native-like conformations. At a given interaction strength, this scheme leads to folding rates that are logarithmically well correlated with CO (correlation coefficient r = 0.914) and span more than 2.5 orders of magnitude, whereas folding rates of the corresponding Gō models with additive contact energies have much less logarithmic correlation with CO and span only approximately one order of magnitude. The present protein chain models also exhibit calorimetric cooperativity and linear chevron plots similar to that observed experimentally for proteins with apparent simple two-state folding/unfolding kinetics. Thus, our findings suggest that CO-dependent folding rates of real proteins may arise partly from a significant positive coupling between nonlocal contact favorabilities and local conformational preferences. Proteins 2003;52:524–533.
© 2003 Wiley-Liss, Inc.

Key words: calorimetry; chevron plot; Gō models; simple two-state kinetics; single-domain proteins; nonadditivity

INTRODUCTION

Generic Protein Properties as Energetic Constraints

The folding of many small single-domain proteins is well approximated by simple two-state thermodynamics and kinetics.[1,2] In the past several years, we have shown that fundamental insights into protein energetics can be gained by using these general, apparently mundane properties as experimental constraints on protein chain models.[3–10] This approach is based on the recognition that model interaction schemes capable of producing these commonly observed experimental properties are, somewhat surpris-

ingly, not entirely straightforward to come up with. To date, much advance has been made by coarse-grained modeling of protein folding.[7,11–15] Nonetheless, the interactions postulated by many existing models are insufficient for calorimetric two-state cooperativity.[3,4] Furthermore, even common Gō models are not cooperative enough for simple two-state kinetics, their explicit native biases notwithstanding. Specifically, we recently found that several lattice[6,9,10] and continuum (off-lattice)[8] Gō-like formulations with essentially additive interaction schemes all led to chevron rollovers—a hallmark of folding kinetics that are often operationally referred to as non-two-state.[9] Apparently, many-body interactions are needed to produce chevron plots with linear folding and unfolding arms consistent with a two-state description of equilibrium thermodynamics.[10]

Small single-domain proteins are characterized as well by a significant correlation between relative contact order (CO) and folding rate.[16] Therefore, it is only logical to require a model protein interaction scheme to produce a similar correlation.[17,18] Ising-like[19,20] and other[21,22] constructs without explicit chain representations have had successes in this regard. However, as for thermodynamic and kinetic cooperativities, achieving the CO dependence requirement in models with explicit chain representations appears to be a nontrivial task. Notably, an early lattice model study using a 20-letter alphabet suggested that proteins with higher CO should fold faster,[23] thus predicting a trend opposite[17] to that for real single-domain proteins.[16,18] A more recent 20-letter lattice model investigation, on the other hand, found modest correlations between increasing CO and longer logarithmic folding

time (correlation coefficient $r \approx 0.70-0.79$ for chain lengths ≥ 54).[24] An earlier continuum Gō model studies of 18 proteins also found a modest correlation between increasing CO and slower logarithmic folding rates ($r = 0.69$).[25] But the corresponding dispersion in simulated folding rates covers only ≈ 1.5 orders of magnitude, which is much narrower than the ≈ 5 orders of magnitude covered by the real folding rates of the proteins in the given dataset. When a different potential function was used in a more recent continuum Gō model analysis, however, no correlation between CO and simulated folding rates was discerned.[26]

Recently, based on lattice 27mer simulations, Jewett et al.[27] have proposed that enhanced thermodynamic cooperativity and many-body interactions—which are basic properties of individual two-state proteins to begin with[1–10]—may also be a key to understand the correlation between CO and folding rate across different proteins. This is an attractive and insightful idea. However, the particular way in which thermodynamic cooperativity was enhanced by these authors led only to modest increases in folding rate dispersion relative to that for the corresponding lattice Gō models with pairwise additive contact energies. Both the dispersion in folding rates and the correlation of logarithmic folding rate with CO ($r = 0.75$) for the most cooperative interaction scheme they reported were similar to that obtained from an earlier continuum Gō model study[25] as well as that from a recent simulation of 20-letter lattice models[24] with only pairwise additive contact energies (see above). In our view, these results suggest that while CO-dependent folding may well derive from certain intraprotein interactions that are also responsible for high thermodynamic cooperativity, CO-dependent folding does not arise from thermodynamic cooperativity *per se*. In other words, how cooperativity is achieved can be critically important. Many *a priori* many-body mechanisms are consistent with high thermodynamic cooperativity. An example is the two rather different interaction schemes we considered in Ref. 10—one involves local-nonlocal coupling, whereas the other assigns an extra favorable energy to the ground-state structure as a whole. But perhaps not all such mechanisms can mimic experimentally observed CO dependencies to the same degree. Therefore, to shed light on the physical mechanisms of CO-dependent folding, we endeavor to construct an interaction scheme that would provide larger dispersions in folding rates and better correlations with CO.

MODELS AND METHODS

The present study focuses on the idea of a cooperative interplay between local conformational preferences and the contact-like interactions that drive the packing of the protein core.[3,5,6,10] We have shown that chain models embodying this idea can lead to calorimetric cooperativity and simple two-state kinetics,[10] although our exploration thus far has been limited to model proteins that are mostly helical.[3,5,6,10] Here we consider a general formulation of this idea, the basic ingredients of which are described by Figure 1(A). This hypothesis may be viewed as a synthesis of the local-dominant and the nonlocal-dominant perspectives.[28] We were motivated by the recognition that both local[29,30] and nonlocal[31,32] intraprotein interactions are important determinants of protein structure and stability. Yet local conformational preferences alone are often insufficient for stable secondary structures under physiological conditions. Secondary structure formation is known to be context dependent[33]; they are stable when packed in the core of a protein but are usually not stable in isolation (Ref. 32 and references therein). Furthermore, conformational space grows exponentially with chain length, even when preferences arising from local excluded volume effects are taken into account.[34] It follows that a large part of the stability and uniqueness of protein native structures cannot be explained by local interactions alone.[35] On the other hand, our recent Gō-model studies have shown that nonlocal contact-like interactions by themselves are not cooperative enough for simple two-state kinetics[6,8–10] if they are not coupled to local conformational propensities.

A Simple Model of Local-Nonlocal Coupling

Here we explore the hypothesis in Figure 1(A) by incorporating its form of local-nonlocal coupling into a new interaction scheme in Figure 1(B) for explicit-chain models configured on three-dimensional simple cubic lattices. This allows the idea to be tested quantitatively. Figure 1(B) may be viewed as a generalization of similar constructs we have used previously in the context of helical proteins.[3,5,6,10] As a first step in our inquiry, we make the simplifying assumption that the interactions are native-centric,[25–27,31,36–38] in that only native interactions are favored, whereas nonnative interactions are neutral (have zero energy). The local-nonlocal coupling in Figure 1(B) involves nonadditive many-body interactions. A chain segment that is locally nativelike (with native bond and torsion angles) but makes no native contact is not stabilized (contributing zero energy). On the other hand, nonlocal contact interactions between monomers far apart along the chain sequence are more favorable when the chain segments around the contacting residues are in their native conformations than when they are not. As such, the present model differs from models that additively combine contact energies and local favorabilities.[39] The importance of nonadditive many-body effects in protein folding has been recognized,[3,5,6,10,40–44] but they have not been used extensively to model calorimetric two-state cooperativity and linear chevron plots.[3–10] Our aim here is to utilize extremely coarse-grained representations as a computationally efficient means to explore the general principles linking CO-dependent folding and proteinlike cooperativities. Many structural and energetic details of real proteins are beyond the scope of this work. In particular, the present work does not deal with the microscopic physical origins of local-nonlocal coupling. Instead we just presume that its presence in naturally occurring proteins could arise from evolutionary design. Because of these, the simple interaction scheme in
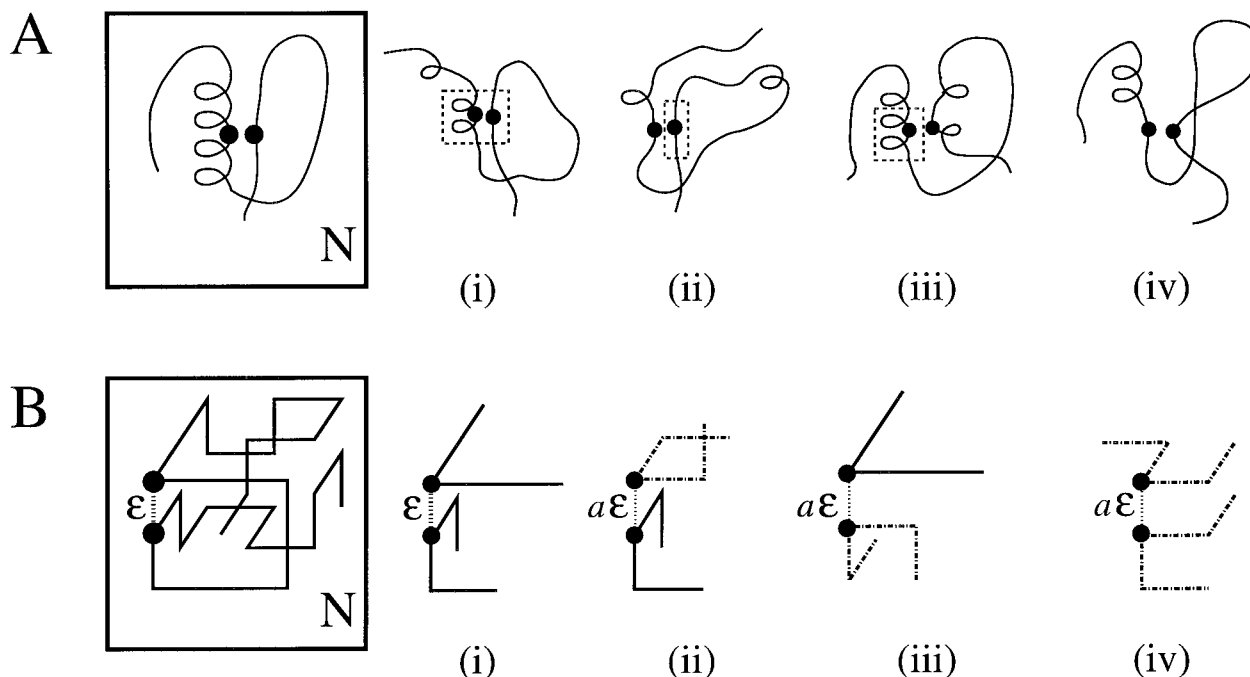
Fig. 1. (A) Schematics of local-nonlocal cooperative energetics in protein folding. The conformation in the solid box represents the native (N) structure; the two filled circles depict a pair of nonlocal residues interacting favorably in the native state. The interaction strength between a residue pair is strong and essentially the same as that in the native structure if the chain segments sequentially local to both residues are nativelike, as in (i). [Dotted boxes in (A) are used to mark nativelike chain segments.] However, the interaction strength is weakened if one or two chain segments sequentially local to the interacting residues are not nativelike, as in examples (ii)–(iv). (B) A lattice implementation of this protein folding scenario. Here the favorable energy for every contact (between residues i and j, $|j - i| \geq 3$) in the ground-state native (N) structure is ε (< 0) when the relative positions of i and the four residues sequentially closest to it (if i is not at a chain end or a sequential neighbor of a chain end, this refers to the five residues centered at i, i.e., residues $i - 2$, $i - 1$, i, $i + 1$, and $i + 2$) as well as the relative positions of j and the four residues sequentially closest to it (if j is not at a chain end or a sequential neighbor of a chian end, this refers to the five residues centered at j, i.e., $j - 2$, $j - 1$, j, $j + 1$, and $j + 2$) are the same as that in N (solid lines in (i)], irrespective of the relative orientations of the two five-residue chain segments. However, if the local conformation of one or both sets of five contiguous residues is nonnative, the contact energy is weakened by an attenuation factor a ($0 \leq a < 1$). Examples of the latter situation is given by (ii)–(iv), where nonnative local chain segments are drawn as broken lines.

Figure 1(B) should be viewed only as a tentative model in this regard.

In order to examine the folding rates of a set of model proteins whose native structures cover a diverse range of CO values, we now consider chains of length $n = 27$ configured on simple cubic lattices. For these 27mers, there are 103,346 distinct maximally compact conformations (not related by rotations or inversions)[45,46] confined to a $3 \times 3 \times 3$ cube. The distribution of CO among these maximally compact conformations covers 97 different values[27] from CO = 208/756 = 0.275 to 402/756 = 0.532 [inset of Fig. 2(A), where CO is computed using Equation 1 of Ref. 16]. For each CO value, we randomly choose a maximally compact 27mer conformation as the native structure of a model protein (Table 1).*

Folding and unfolding kinetics are modeled by standard Monte Carlo simulations using the Metropolis criterion and the elementary chain moves of end flips, corner flips, crankshafts, and rigid rotations. The relative frequencies of attempting these moves are 4.7%, 58.3%, 27%, and 10%

respectively (cf. Ref. 6).† Time is measured by the number of attempted Monte Carlo moves for a given process. The set of elementary chain moves is chosen to mimic physically plausible processes. Lattice model kinetics are dependent on the choice of move set.[12] Nonetheless, we expect the general trend predicted by the model is less sensitive to move set when kinetics are not dominated by trapping events,[12] as is the case here and has been verified by Jewett et al.[27] Progress toward the native state is tracked by the fractional number of native contacts Q (Refs. 3–6). To ascertain the implications of the local-nonlocal coupling we proposed, results from a highly cooperative interaction scheme with $a = 0.1$ are compared with that from the additive scheme ($a = 1$) of common Gō models [cf. Fig. 1(B)]. Folding trajectories are initiated at a randomly generated conformation; folding first passage time is defined by the formation of the $Q = 1$ ground-state conformation. Unfolding trajectories are initiated at the ground-state

---

*Since the present choices of structures are independent of that by Jewett et al.,[27] the structures listed in Table I do not necessarily coincide with those used in their study.

†The following typographical error in Ref. 6 should be corrected. The relative attempt frequencies of corner flips and crankshafts used in this prior study of ours were, respectively, 60.6% and 27%, not the 27% and 60.6% stated on p. 901 of Ref. 6.
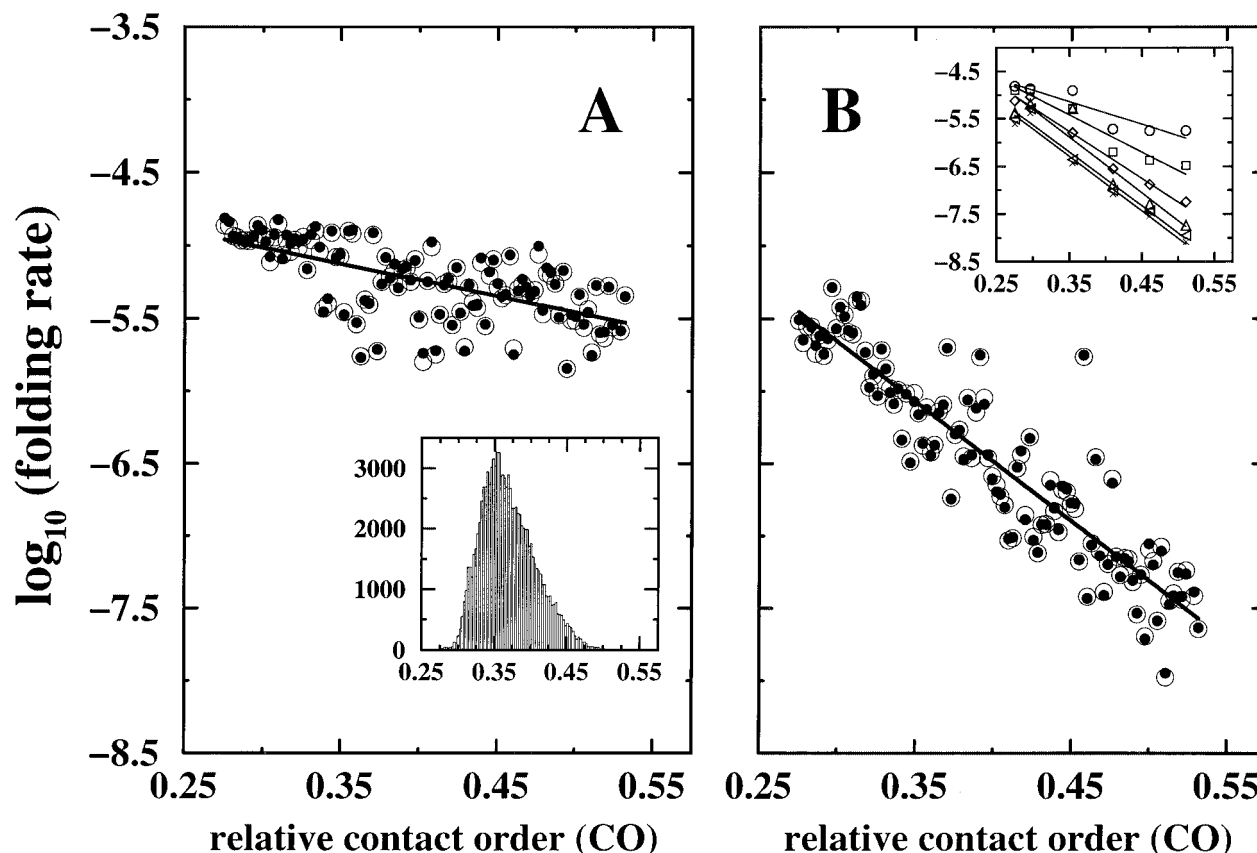
Fig. 2.   Correlation between the common (base 10) logarithm of folding rate and CO for the 97 structures in Table I under moderately folding conditions at $\varepsilon/k_BT = -1.47$, using (A) the common additive Gō potential and (B) the local-nonlocal cooperative interaction scheme with $a = 0.1$. Solid lines are least-square fits. Here folding rate is the reciprocal of mean folding first passage time (folding rate = 1/MFPT). Each MFPT is averaged from 500 trajectories. Associated with each value of $\log_{10}(1/\text{MFPT})$ (filled circle) is an open circle marking the common logarithm of the median folding first passage time (FPT) divided by ln 2. If the kinetics is single-exponential, MFPT = (median FPT)/ln 2. The inset in (A) is the distribution of CO among the 103,346 maximally compact 27mer conformations, wherein the number of conformations (vertical scale) is shown as a function of CO (horizontal scale). The inset in (B) uses six representative structures with different CO values ($\Sigma \Delta S_{ij}$ = 208, 224, 268, 310, 348, and 386 entries in Table I) to illustrate that $\log_{10}$(folding rate) (vertical scale) is more sensitive to CO (horizontal scale) when the local-nonlocal coupling is stronger. In this inset, different symbols denote different a values; the lines fitted through the symbol are, from top to bottom, for a = 1, 0.75, 0.5, 0.25, 0.1, and 0.0.

conformation; unfolding first passage time is the time it takes for the chain to be left with three or fewer native contacts ($Q \leq 3/28$); $Q = 3/28$ is chosen to define unfolding because it corresponds approximately to the free energy minimum for the denatured state.

RESULTS
Sensitivity of Folding Rate on CO Enhanced by Local-Nonlocal Coupling

Figure 2 provides the correlation between CO and folding rate among our 27mer models. It shows clearly that the local-nonlocal coupling mechanism postulated in Figure 1 can lead to a significant enhancement of correlation as well as much increased sensitivity of folding rate to CO. Whereas the dispersion in folding rates among the common additive Gō models in Figure 2(A) covers only approximately one order of magnitude (a factor of 10) and the logarithmic folding rates exhibit only a relatively weak correlation with CO (correlation coefficient $r = 0.63$), the corresponding dispersion among

the $a = 0.1$ cooperative models in Figure 2(B) covers approximately 2.5 to 3 orders of magnitude, with a strong correlation between CO and logarithmic folding rate ($r = 0.914$) comparable to that observed among a selection of real, small single-domain proteins.[18] Similar to the corresponding experimental situations,[16,18] the comparisons in Figure 2 were performed under conditions for which folding relaxation is essentially single-exponential, as is evident from the good agreements in Figure 2 between median first passage time divided by ln 2 and the corresponding mean first passage time.[6,47] To better delineate the effects of having weakened contact interactions when the chain segments locally adjacent to the contacting residues are nonnative, several a values other than the $a = 0.1$ used for the main plot are compared in the inset of Figure 2(B). It shows CO-dependent folding at different levels of local-nonlocal coupling (different a values) for several 27mers with representative COs. The $a = 0$ case here corresponds to complete interdependence between nonlocal

contact and local structure. This inset indicates that sensitivity of folding rate to CO increases (the fitted line has a more negative slope) with decreasing a, and that the behavior of the $a = 0.1$ models is very similar to that of the $a = 0$ models. These results further affirm that local-nonlocal coupling is a key ingredient for the good correlation between CO and fold rate in these models. Nevertheless, as for real proteins,[16,18] despite the good correlation, CO by itself cannot predict folding rates of the present models with high accuracy. Folding rates here can vary significantly for different structures with the same CO as well. For example, for the particular 27mer with CO = 346/756 = 0.458 in Figure 2(B), the datapoint $\log_{10}$(folding rate) = -5.75 may be viewed as an "outlier" vis-à-vis the fitted line. However, for two other 27mers with the same CO but do not belong to the randomly chosen set in Table I [and therefore not plotted and not used in the correlation analysis of Fig. 2(B)], we found $\log_{10}$(folding rate) = -7.26 and -7.60, which happen to be much closer to the fitted line in Figure 2(B). The reasons behind variations in folding rates among structures with same CO remain to be elucidated.

A Consistent Model of Thermodynamic and Kinetic Cooperativity

Figure 3 provides further analyses of the folding/unfolding kinetics of one example 27mer structure we choose to study in more detail. Consistent with our previous results,[6,8–10] it shows that the model chevron plot[48] predicted by the common additive Gō potential (upper plot) deviates significantly from simple two-state kinetics in that it exhibits a severe rollover under only moderately native conditions. More specifically, for this case rollover becomes significant at $\varepsilon/k_B T$ values that are only slightly more negative (more favorable to folding) than that of the transition midpoint ($\varepsilon/k_B T \approx -1.43$). In contrast, the chevron plot predicted by the model with a substantial local-nonlocal coupling (lower plot) is qualitatively similar to that of real, small single-domain proteins that fold and unfold with simple two-state kinetics.[10] In particular, it has essentially linear folding and unfolding arms over an extended range of $\varepsilon/k_B T$ values. We have also obtained for this model the equilibrium free energy of unfolding ΔG_u as a function of $\varepsilon/k_B T$, where ΔG_u here is taken to be that between the unique $Q = 1$ conformation and those with $Q \leq 3/28$. (The same definition is used for unfolding kinetics as stated above.) Because ΔG_u is essentially linear in $\varepsilon/k_B T$, the linearity of the chevron arms over an extended $\varepsilon/k_B T$ range implies an essentially linear relationship between folding/unfolding rates and ΔG_u within the corresponding regime (i.e., the model parameter ε may be eliminated in favor of the lower horizontal scale in Fig. 3). Furthermore, comparing the mean first passage times in Figure 3 versus the corresponding median first passage times divided by ln 2 shows that folding or unfolding relaxation for this model is essentially single exponential[6,47] for $\Delta G_u < 10k_B T$. Essentially single-exponential folding under moderately folding conditions is further

demonstrated by an approximately linear logarithmic distribution of first passage time[8,9,49] shown in the inset. Similar to the cooperative models we recently investigated,[10] for the model with local-nonlocal coupling in Figure 3, the thermodynamic ΔG_u values match well with the kinetically obtained quantity $k_B T$ ln[(folding rate)/(unfolding rate)] for ΔG_u ranging from $10k_B T$ to $-6k_B T$ (lower V-shape). In other words, the folding/unfolding kinetics of this model is apparently simple two-state[6,8–10] within a ΔG_u range quite similar to that experimentally accessible to small single-domain proteins.[10] Finally, the cooperative model in Figure 3 is also calorimetrically two-state. Assuming that the interactions are temperature independent, the model's van't Hoff to calorimetric enthalpy ratio $\Delta H_{vH}/\Delta H_{cal}$ (κ_2 without baseline subtraction[4]) is determined to be 0.992 (detailed calculation not shown), satisfying the requirement of $\Delta H_{vH}/\Delta H_{cal} \approx 1$ for apparent two-state thermodynamics.[3–5] Taken together, the above considerations imply that the local-nonlocal coupling mechanism for enhanced CO-dependent folding in Figure 2(B) also provides—as it should—a consistent account of thermodynamic and kinetic cooperativities[6,8–10] in simple two-state proteins (Fig. 3).

As it stands, the transition midpoints of all 27mers considered here with the local-nonlocal coupling parametrized by $a = 0.1$ are very close to one another. This is because the interaction scheme in Figure 1(B) assigns the same energy ($=28\varepsilon$) to every ground-state conformation. This is a simplifying assumption in the present modeling setup. Because the thermodynamic stabilities of real, small single-domain proteins are quite diverse,[16,18] it is important to note that, in a broader perspective, our hypothesis that significant CO-dependent folding can emerge from local-nonlocal coupling is not contingent upon the different proteins in question having very similar thermodynamic stabilities. In more sophisticated models, for example, an extra favorable energy that differs from one 27mer to another may be assigned to the ground-state conformation (i.e., a different E_{gs} term as defined in Ref. 10 for each 27mer). In that case, the thermodynamic stabilities of different 27mers can be very different, but their folding rates would not be affected by this extra feature of the model. In other words, the correlation between CO and folding rate in Figure 2(B) would remain unchanged. As we have recently argued,[10] such extra stabilizing energies for the ground state as a whole are physical plausible because experimental evidence[50] indicates that in real proteins there is a partial separation between the driving forces for folding kinetics and the interactions responsible for thermodynamic stability.

DISCUSSION

Energy landscapes of the present models are further characterized in Figure 4 for three representative structures with low, intermediate, and high CO values. In this figure, the low- and high-CO structures are, respectively, the fastest and slowest folding among the 97 structures in Table I, whereas the intermediate-CO structure is the one analyzed in Figure 3. For the common additive Gō poten-

TABLE I. The Ground-State 27mer Conformations ($n = 27$) Used in This Investigation[†]

$\Sigma \Delta S_{ij}$	Conformation	$\Sigma \Delta S_{ij}$	Conformation
208	uufddfuurddbuubddruufddfuu		
210	uufddfuurbbdffdbbrffuubdbu	306	uufrrbldrfdflurullddburdbr
212	ufdfuubbrddffubufrddbuubdd	308	uufdfrbrbulddrfllfrruublfl
214	uuffdbdfrbufubbddrffuubdbu	310	ufrulblfddrrbbllfuburdrfub
216	ufdfuubbrddfuufddruubbdddu	312	uufrbbllffdrrdllbubdrurfdb
218	ufdfuubbrdfufddbbruuffddbu	314	uufdrubbdfdfllbbuffubbrddr
220	uuffddburfdbbuuffrddbuubdd	316	uffrddblbruufdllbdffrulubb
222	uuffddburdfuubbddruuffdbdf	318	uufrbddbuullffdrrdllbubdru
224	uufddrbufubrfdbdfflurulldd	320	uufddfrruubbdfdbluuffldrbd
226	uufddfuurddbubdrffubbulfrf	322	uffdbrbrfufullbbrdrufldfdr
228	uffdbrbuffdrbbuffubbllfrfl	324	ufrbddlfrflurullbbddffubrr
230	ufdrbufublffddruurddbuubdd	326	ufrfddluulddbbuufdrdbrfubu
232	uufrrblddrufdluldfurdruull	328	ufrubdbulddrrfllbufurblb
234	uufddfuurddrbluurfdbbulddr	330	uufdrubblddllfubuffddrrbbuf
236	ufddbbuurrflfrdlbdfrbubldr	332	uuffrddruubbdfllfdbrrbluuf
238	uuffdbdfrrblbrulffrulbbrfd	334	ufrrdbdfluldbbrruuflbldrfd
240	uffdbrfurdbblufrbufflbbrf	336	uffrbrbuflblffrrddllbrrblu
242	uufrbddfflburflurrbbdffdbb	338	uffurrdldrbblurullfrrdldlf
244	ufdrrbluulfrdrbuflfddrrul	340	ufrrbbdffdlbrbllffurbubldr
246	ufdfurbdfruullbrrddblurull	342	ufrullbrrblldrrdllfufdrrbu
248	ufdfrbuflurblbrrdldrffuubd	344	uffdrdllbrbluuffdbrrdbuuff
250	ufddbrfruublfdbrdblluurdru	346	uuffddrrbbuufdfuldbubddflu
252	uffdbrrflurbbdlufufrbbllff	348	uufrbbdluldrrfllbuufdrrb
254	ufddbrblurrdfflubrfulbrbll	350	ufubrrdfdfuldblfuurrbldbdr
256	ufdfurdruullbbrddrfluurdbu	352	uffddrbllurrfubbddlluuffdd
258	ufrbdflfrrbbuullfrrdfulldr	354	ufrfdrbufubbllfrflddbrburd
260	uufddfurbbrdlffrbufubblffl	356	uffrrbdbuullffrrbldbdflfrr
262	uuffdbrbufrfldrdllbrbrfubu	358	uufdrrubddffuulldrdlbrbuuf
264	ufdrurddbuuldblurrddllffrb	360	ufrddllfrruulldrblubddrruu
266	ufdrurddllbrbluuurrfldbrdfu	362	ufubrrdffldrbblflfuurrbldb
268	uuffdbrubrfddbuldflfrrulur	364	ufrfddlbbllffubbuffrdbrdbuu
270	uuffrrdllbdrbufrulbrddffll	366	ufrfdbdfllbbuufdfurdbdbruu
272	uufrdfuldbdfrruubblddfrubd	368	uffurrbbddfflllbrbuulfrdrfl
274	ufdrubrfddllbbuurrdldrfuld	370	uffurrddbbuufllbrddflfrubr
276	ufdrbdlfrrubdbllurffrbbdl	372	uufdrfdruubbddlluufflddbrru
278	ufddrrbllbrrullurrfflbdfrb	374	uffrdrbbuullffrrdbuldbdflf
280	uffrddlubdruubddllfubuffdd	376	uffrddlllbuubddrfebuufdlflu
282	ufrbdffuldlubbddrrfflbuldf	378	ufdrfdlluubbdfdbrfrbuuffld
284	uufrfldrrubbldrfdblfuldfrr	380	ufrfddlbrbllffubbuffrdbrbu
286	uffubbrddrffuldlbrurbufflb	382	uufdrfdrbbuufdfullddbrbuuf
288	ufdfrrubufldlubbrfdbdfrbuu	384	ufrbbullddfuurrfflddrrbblu
290	uufrbbldrfdbllfubuffddrurd	386	uffddrbllfuubbddrruuffdbll
292	uffrddbbuufdldblffrulubbdf	388	uffrburbddfflllbrbuulffrrdb
294	ufdrdfulurbbddlluufddrfluu	390	ufrufddrbbuffubbllffddbrbu
296	ufdfrbdflbbruuffllddbbuufd	392	ufrubbddllfuubdruffllddrrbu
298	ufrrdblblurfrbddffluldbrbl	394	ufrrddlbburuflblddffurbrdb
300	ufdfrulllddbuubddrffrbuubdd	396	uffrddblffbufubbddrruufdllf
302	ufdfurddllluubbdfdbrfrbuufd	398	ufrfddlllubdrrblluuffrdbrbu
304	ufrdlluurrbbdfdbllfuubdruf	400	ufrufrbbddfflllbrbuulffdrrb
		402	

[†]The conformations are given by sequences of 26 bond directions, where r = right ($+x$), l = left ($-x$), f = forward ($+y$), b = backward ($-y$), u = up ($+z$), d = down ($-z$). A structure is randomly selected for each of the 97 possible CO values amongst the compact 27mer structures with $t_{max} = 28$ contacts. Each integer $\Sigma \Delta S_{ij}$ is the sum of $|j - i|$ over the (i, j) nearest-neighbor contacts in the given conformation ($j - i \geq 3$). Here CO = $\Sigma \Delta S_{ij}/(nt_{max}) = \Sigma \Delta S_{ij}/756$.

tial, energy E is directly proportional to Q ($E = \varepsilon Q$). However, for the cooperative models with local-nonlocal coupling, there are multiple energy levels for each Q, with $E = \varepsilon Q$ as the lower bound (left panels of Fig. 4). This

means that, on average, the energetic separations between non-ground-state and ground-state conformations in the cooperative models with local-nonlocal coupling are larger than that in the additive Gō models. This feature is
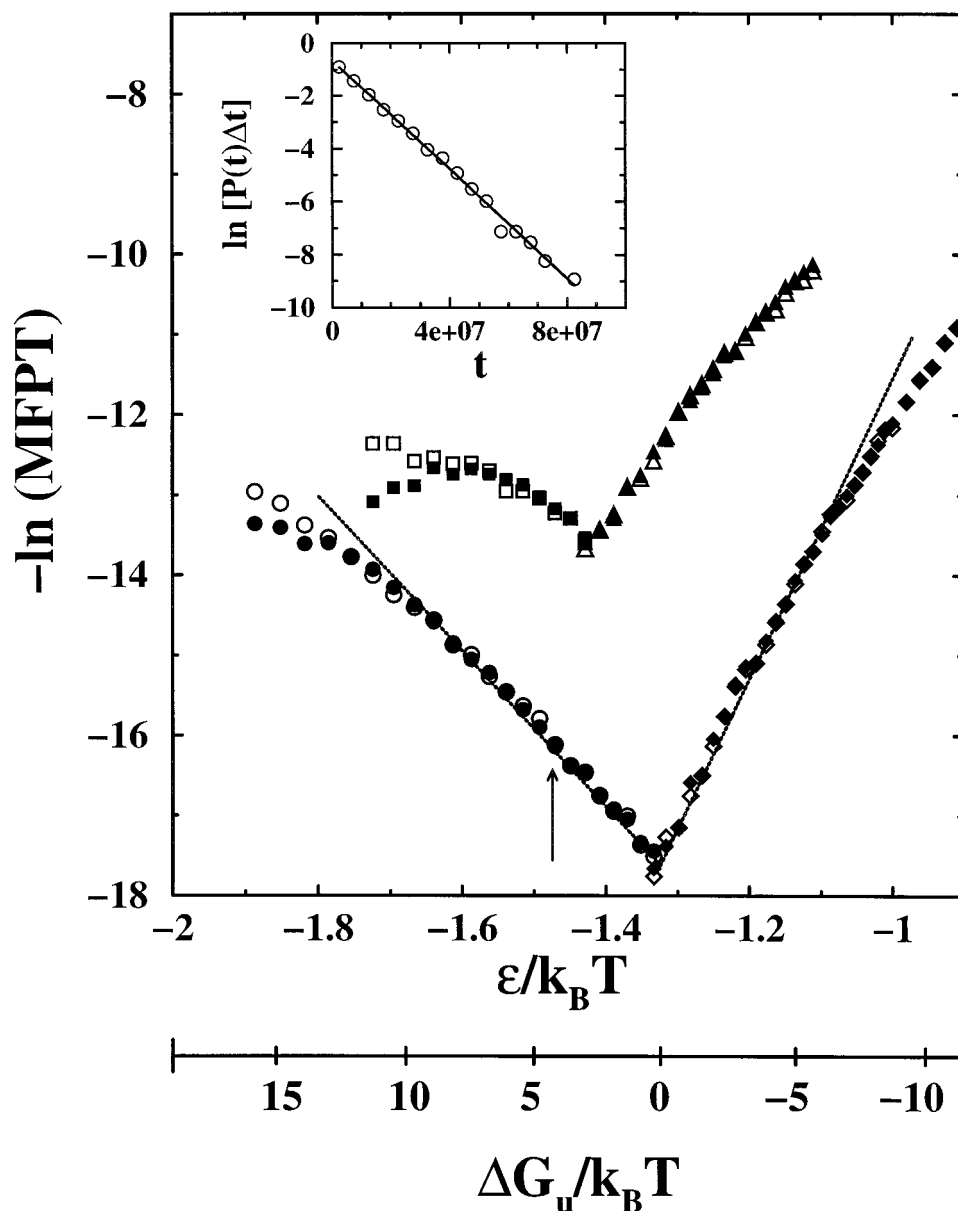
Fig. 3. Model chevron plots for a CO = 0.410 structure ($\Sigma \Delta S_{ij}$ = 310 entry in Table I) are given by negative natural logarithm of MFPT as a function of $\varepsilon/k_B T$ (filled symbols). Values of (median FPT)/ln 2 are shown by the open symbols. Squares (folding) and triangles (unfolding) are for the additive Gō potential ($a = 1$, upper plot), whereas circles (folding) and diamonds (unfolding) are for the $a = 0.1$ local-nonlocal cooperative interaction scheme (lower plot). Each MFPT is averaged from 500 trajectories, except for the model with local-nonlocal coupling at $\varepsilon/k_B T = -1.47$ (arrow). For this particular case, 7500 folding trajectories were simulated to provide enriched statistics for the FPT distribution in the inset, wherein $P(t)\Delta t$ is the fraction of trajectories with $t - \Delta t/2 <$ FPT $\leq t + \Delta t/2$, and the bin size Δt for FPT is equal to 5×10^6. The free energy of unfolding ΔG_u for the $a = 0.1$ cooperative model is computed using Monte Carlo histogram techniques based on sampling at the transition midpoint $\varepsilon/k_B T = -1.33$. ΔG_u is essentially linear in ε (lower horizontal scale). The dotted V-shape, which fits well to the kinetic datapoints of the $a = 0.1$ cooperative model over an extended regime, is an hypothetical simple two-state chevron plot consistent with the dependence of ΔG_u on ε.

demonstrated directly in the right panels of Figure 4, which show that the number of non-ground-state conformations within a given energy range is smaller for the cooperative models than for the additive Gō models except for the highest energies ($E \approx 0$). It follows that the overall thermodynamic cooperativities of the models with local-nonlocal coupling are substantially higher than that of the corresponding additive Gō models. This behavior is expected as well from our recent finding that simple two-state folding/unfolding kinetics (Fig. 3 above) requires "near-Levinthal" thermodynamic cooperativity.[10] Indeed, for the three models in Figure 4 with local-nonlocal coupling, the van't Hoff to calorimetric enthalpy ratios $\Delta H_{vH}/\Delta H_{cal}$ are, from top to bottom, $\kappa_2 = 0.972, 0.992$, and
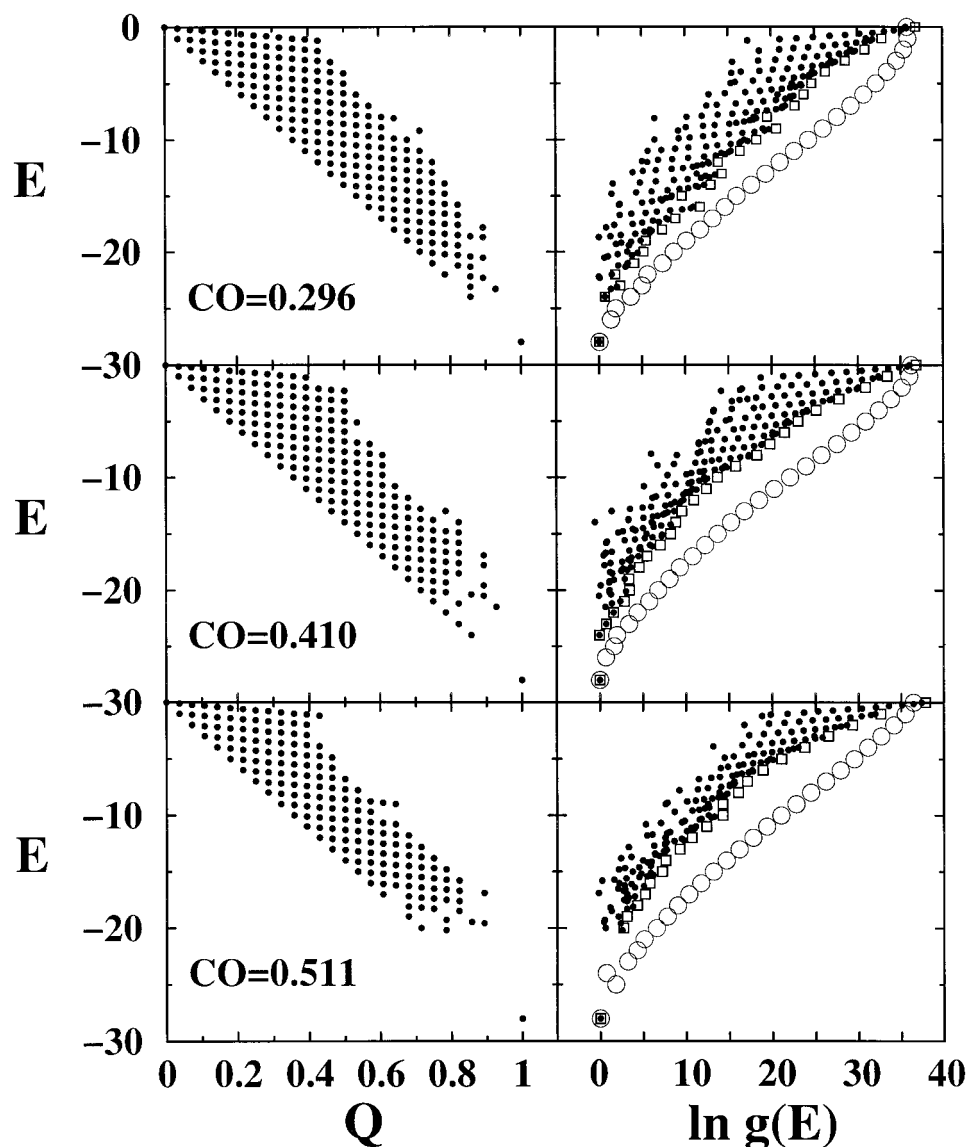
Fig. 4. Energy landscapes of three representative models with local-nonlocal coupling ($a = 0.1$, $\Sigma \, \Delta S_{ij} =$ 224, 310, and 386 entries in Table I; $\varepsilon = -1$). Left panels: the correlation between E and Q; each dot indicates that at least one conformation with the given (E,Q) was encountered in our sampling. Right panels: the structures' logarithmic densities of states, where $g(E)$ is the number of conformations with energy E for the cooperative models ($a = 0.1$, dots). Included for comparison are the ln $g(E)$ values of the corresponding additive Gō models ($a = 1$, open circles; $\varepsilon = -1$). The densities of states here are estimated by Monte Carlo sampling at the models' transition midpoints $\varepsilon/k_B T = -1.33$ ($a = 0.1$) and $\varepsilon/k_B T = -1.43$ ($a = 1$). Note that the cooperative models have more energy levels than the additive models. Therefore, to compare their densities of states on an equal footing, the open squares provide the natural logarithm of the number of conformations in the $a = 0.1$ cooperative models with energies in the range $m - 0.5 \le E < m + 0.5$, where $m = 1,0, -1, -2, \ldots$ is an integer. Now the densities of states represented by the open squares ($a = 0.1$) are directly comparable to that represented by the open circles ($a = 1$) because their values are based upon the same unity bin size for E.

0.998. These values are extremely high for model enthalpy ratios without baseline subtractions.[4] In contrast, the corresponding additive Gō models are less cooperative, with $\kappa_2 = 0.751$, 0.861, and 0.878. Here it is noteworthy that the additive Gō models' $\Delta H_{vH}/\Delta H_{cal}$ ratios even after empirical baseline subtractions,[4] $\kappa_2^{(s)} = 0.885$, 0.961, and 0.962, are lower than the $\Delta H_{vH}/\Delta H_{cal}$ ratios of the cooperative models in the absence of baseline subtractions.

Contact-order Dependence Indicative of Special Mechanisms of Cooperativity

Obviously, thermodynamic cooperativity is a necessary ingredient for any protein chain model that purports to rationalize the generic properties of small single-domain proteins.[3–10] For the particular interaction scheme we consider, the above analysis shows that features that give rise to significant CO-dependent folding also lead to high

thermodynamic cooperativity. However, the converse is not necessarily true. More in-depth considerations and a comparison of the present results with that of Jewett et al.[27] indicate that higher thermodynamic cooperativity *per se* does not necessarily give rise to more enhanced dependence of folding rate on CO. Our reasoning is as follows. First, for the present set of 27mer structures we have chosen randomly, the correlation between logarithmic folding rate and CO is quantified by $r = 0.63$ ($r^2 = 0.39$) for the additive Gō interaction scheme. Despite that this correlation happens to be weaker than that of Jewett et al.'s collection of additive Gō models (their $r^2 = 0.51$), after cooperativity is enhanced by local-nonlocal coupling, the correlation between logarithmic folding rate and CO for our $a = 0.1$ models is much higher ($r^2 = 0.84$, see Fig. 2 above, an improvement in r^2 value of 0.33)* than the best case reported by Jewett et al.[27] ($r^2 = 0.57$ for their $s = 3$, an improvement in r^2 value of 0.06 over that for their additive Gō models).†

Second, the folding rates of our cooperative models are much more sensitive to CO, covering 2.5 to 3 orders of magnitude, whereas those of Jewett et al. cover only approximately 1.3 orders of magnitude. This means that the present local-nonlocal coupling mechanism is significantly more effective in enhancing CO dependence than the nonlinear E-Q relationship postulated by Jewett et al. (Equation 1 of Ref. 27). Their interaction scheme does not make direct reference to chain conformations as such. Thermodynamic cooperativity is enhanced in their models by stipulating that the total contact energy E (for a given conformation as a whole) does not decrease (does not become more favorable) linearly with increasing Q as in common Gō models; but rather decreases at progressively faster and faster rates when Q is closer to unity.¶

Third, in fact, if thermodynamic cooperativity is further increased in the interaction scheme of Jewett et al. by increasing their s parameter, the energy landscape will eventually become a Levinthal golf course in the $s \to \infty$ limit. In that case, folding would be rate-limited by random conformational search and CO-dependence would be all but eliminated. Fourth, in this connection, we have recently considered three 27mer models with CO = 0.28, 0.40, and 0.51 in a separate study. The thermodynamic cooperativity of these models are enhanced by assigning an extra stabilizing energy to the ground state but without local-nonlocal coupling.[10] For the energetic parameters we considered, the folding rates of these models cover less than an order of magnitude.[10] The same set of results also indicated that dispersion in folding rates under moderately

folding conditions would decrease if thermodynamic cooperativity is increased by assigning an even stronger stabilizing energy to the ground state, in a manner similar to greatly increasing s in Jewett et al.'s formulation. Taken together, these observations lead us to the conclusion that while thermodynamic cooperativity is certainly necessary, by itself it is not sufficient to guarantee CO-dependent folding rates similar to that observed experimentally[16,18] if the underlying mechanism for thermodynamic cooperativity is not specified.

CO-dependent folding highlights the important role of local interactions in determining folding rates.[16–18] It suggests that the mechanism of folding may involve relatively fast formation of local structure. We have also established a connection between our proposed mechanism and the transition-state picture of protein folding (see cover image), although details of the relationship remain to be further explored. In this regard, we note that under the general lattice scheme in Figure 1(B), formation of strong (unattenuated) native contacts with contact order $|j - i| = 3$ is relatively easier than formation of strong native contacts with higher contact orders. This is because in the $|j - i| = 3$ case there is an overlap between parts of the two local segments that have to be nativelike in order for the contact to be strong. Physically, how a general mechanism similar to that in Figure 1 may arise in real proteins from solvent-mediated atomic interactions such as sidechain packing and hydrogen bonding remains to be elucidated. Many basic issues will have to be tackled to address this question. For example, correlations between backbone and sidechain rotamer conformations[53] may contribute to such a mechanism. Another possibility is that *anti-cooperativity* of certain hydrophobic interactions[54] may disfavor premature nonspecific hydrophobic collapse (which would lead to kinetic trapping[14]) when the sidechains are locally less well packed than that in the native state. If this is the case, it could give rise to local-nonlocal coupling mechanisms similar to that postulated in Figure 1.

In summary, although the models used in the present study are rudimentary, they provide strong evidence that a cooperative interplay between local conformational preferences and nonlocal favorable contact-like interactions is an important mechanism in accounting for experimentally observed CO-dependent folding of small single-domain proteins. We are optimistic that more rigorous applications of the CO-dependence constraint as well as the thermodynamic and kinetic cooperativity requirements would help further narrow down theoretical possibilities and thus contribute to a more realistic understanding of protein energetics.

NOTE

Because of their small size, uniform density and identical overall shape, the present set of lattice 27mer ground-state conformations by themselves do not address the recently proposed dependencies of folding rate on the topological descriptor "cliquishness" which measures the clustering of inter-residue contacts (Ref. 55).

ACKNOWLEDGMENTS

---

*Because all the model chains in the present study have the same length and the same number of native contacts, their correlation coefficient between folding rate and CO is the same as that between folding rate and the total contact distance (TCD) defined in Ref. 51.

†If the $s = 3$ interaction scheme of Jewett et al. is applied to the present set of structures and kinetic models, we found $r^2 = 0.65$ for the correlation between CO and folding rate. In this case, the folding rates span ≈ 1.8 orders of magnitude; see Ref. 52 for details.

¶Jewett et al. suggested that the "extraordinary cooperativity in protein folding" may originate from "three-body interactions." But how three-body interactions might lead to their E-Q relationship remains to be elucidated.

helpful discussions, and Vijay Pande and Kevin Plaxco for kindly sharing their work (Ref. 27) before publication.

REFERENCES

1. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 1991;30:10428–10435.
2. Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
3. Chan HS. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 2000;40:543–571.
4. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661 [Erratum: Proteins 2001;43:523].
5. Kaya H, Chan HS. Energetic components of cooperative protein folding. Phys Rev Lett 2000;85:4823–4826.
6. Kaya H, Chan HS. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding? J Mol Biol 2002;315:899–909.
7. Chan HS, Kaya H, Shimizu S. Computational methods for protein folding: scaling a hierarchy of complexities. In: Jiang T, Xu Y, Zhang MQ, editors. Current Topics in Computational Molecular Biology. Cambridge, MA: The MIT Press; 2002. p 403–447.
8. Kaya H, Chan HS. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? J Mol Biol 2003;326:911–931.
9. Kaya H, Chan HS. Origins of chevron rollovers in non-two-state protein folding kinetics. Phys Rev Lett 2003;90:258104.
10. Kaya H, Chan HS. Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity. Proteins 2003;52:510–523.
11. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–195.
12. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–602.
13. Thirumalai D, Woodson SA. Kinetics of folding of proteins and RNA. Acc Chem Res 1996;29:433–439.
14. Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
15. Mirny L, Shakhnovich E. Protein folding theory: from lattice to all-atom models. Annu Rev Biophys Biomol Struct 2001;30:361–396.
16. Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;227:985–994.
17. Chan HS. Matching speed and locality. Nature 1998;392:761–763.
18. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177–11183.
19. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
20. Muñoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
21. Debe DA, Goddard WA. First principles prediction of protein folding rates. J Mol Biol 1999;294:619–625.
22. Makarov DE, Keller CA, Plaxco KW, Metiu H. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. Proc Natl Acad Sci USA 2002;99:3535–3539.
23. Abkevich VI, Gutin AM, Shakhnovich EI. Impact of local and nonlocal interactions on thermodynamics and kinetics of protein folding. J Mol Biol 1995;252:460–471.
24. Faisca PFN, Ball RC. Topological complexity, contact order, and protein folding rates. J Chem Phys 2002;117:8587–8591.
25. Koga N, Takada S. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. J Mol Biol 2001;313:171–180.
26. Cieplak M, Hoang TX. Universality classes in folding times of proteins. Biophys J 2003;84:475–488.
27. Jewett AI, Pande VS, Plaxco KW. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. J Mol Biol 2003;326:247–253.
28. Uversky VN, Fink AL. The chicken-egg scenario of protein folding revisited. FEBS Lett 2002;515:79–83.
29. Baldwin RL, Rose GD. Is protein folding hierarchic? I. Local structure and peptide folding. Trends Biochem Sci 1999;24:26–33.
30. Shortle D. Composites of local structure propensities: evidence for local encoding of long-range structure. Protein Sci 2002;11:18–26.
31. Gō N, Taketomi H. Respective roles of short- and long-range interactions in protein folding. Proc Natl Acad Sci USA 1978;75:559–563.
32. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29: 7133–7155.
33. Minor DL, Kim PS. Context-dependent secondary structure formation of a designed protein sequence. Nature 1996;380:730–734.
34. Feldman HJ, Hogue CWV. Probabilistic sampling of protein conformations: new hope for brute force? Proteins 2002;46:8–23.
35. Shimizu S, Chan HS. Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. Proteins 2002;49:560–566.
36. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
37. Clementi C, Nymeyer H, Onuchic JN. 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 2000;298:937–953.
38. Linhananta A, Zhou Y. The role of sidechain packing and native contact interactions in folding : discontinuous molecular dynamics folding simulations of an all-atom Gō model of fragment B of *Staphylococcal* protein A. J Chem Phys 2002;117:8983–8995.
39. Thomas PD, Dill KA. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. Protein Sci 1993;2:2050–2065.
40. Kolinski A, Galazka W, Skolnick J. On the origin of the cooperativity of protein folding: implications from model simulations. Proteins 1996;26:271–287.
41. Plotkin SS, Wang J, Wolynes PG. Statistical mechanics of a correlated energy landscape model for protein folding funnels. J Chem Phys 1997;106:2932–2948.
42. Liwo A, Kazmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. United-residue force field for off-lattice protein structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. J Comput Chem 1998;19:259–276.
43. Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and a random heteropolymer. J Chem Phys 1999;110:11616–11629.
44. Eastwood MP, Wolynes PG. Role of explicitly cooperative interactions in protein folding funnels: a simulation study. J Chem Phys 2001;114:4702–4716.
45. Chan HS, Dill KA. The effect of internal constraints on the configurations of chain molecules. J Chem Phys 1990;92:3118–3135 [Erratum: J Chem Phys 1997;107:10353].
46. Chan HS, Bornberg-Bauer E. Perspectives on protein evolution from simple exact models. Appl Bioinformatics 2002;1:121–144.
47. Gutin A, Sali A, Abkevich V, Karplus M, Shakhnovich EI. Temperature dependence of the folding rate in a simple protein model: search for a "glass" transition. J Chem Phys 1998;108:6466–6483.
48. Matthews CR. Effect of point mutations on the folding of globular proteins. Methods Enzymol 1987;154:498–511.
49. Abkevich VI, Gutin AM, Shakhnovich EI. Free energy landscape for protein folding kinetics: intermediates, traps, and multiple pathways in theory and lattice model simulations. J Chem Phys 1994;101:6052–6062.
50. Northey JGB, Di Nardo AA, Davidson AR. Hydrophobic core packing in the SH3 domain folding transition state. Nature Struct Biol 2002;9:126–130.
51. Zhou HY, Zhou YQ. Folding rate prediction using total contact distance. Biophys J 2002;82:458–463.
52. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2003; in press.
53. Dunbrack RL. Rotamer libraries in the 21st century. Curr Opin Struct Biol 2002;12:431–440.
54. Shimizu S, Chan HS. Anti-cooperativity and cooperativity in hydrophobic interactions: three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. Proteins 2002;48:15–30. [Erratum: Proteins 2002;49:294].
55. Micheletti C. Prediction of folding rates and transition-state placement from native-state geometry. Proteins 2003;51:74–84.