Simple Two-State Protein Folding Kinetics Requires Near-Levinthal Thermodynamic Cooperativity

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, Canada

ABSTRACT **Simple two-state folding kinetics of many small single-domain proteins are characterized by chevron plots with linear folding and unfolding arms consistent with an apparent two-state description of equilibrium thermodynamics. This phenomenon is hereby recognized as a nontrivial heteropolymer property capable of providing fundamental insight into protein energetics. Many current protein chain models, including common lattice and continuum Gō models with explicit native biases, fail to reproduce this generic protein property. Here we show that simple two-state kinetics is obtainable from models with a cooperative interplay between core burial and local conformational propensities or an extra strongly favorable energy for the native structure. These predictions suggest that intramolecular recognition in real two-state proteins is more specific than that envisioned by common Gō-like constructs with pairwise additive energies. The many-body interactions in the present kinetically two-state models lead to high thermodynamic cooperativity as measured by their van't Hoff to calorimetric enthalpy ratios, implying that the native and denatured conformational populations are well separated in enthalpy by a high free-energy barrier. It has been observed experimentally that deviations from Arrhenius behavior are often more severe for folding than for unfolding. This asymmetry may be rationalized by one of the present modeling scenarios if the effective many-body cooperative interactions stabilizing the native structure against unfolding is less dependent on temperature than the interactions that drive the folding kinetics. Proteins 2003;52:510–523.   © 2003 Wiley-Liss, Inc.**

Key words: **calorimetric cooperativity; single-domain proteins; chevron plot; chevron rollover, non-Arrhenius kinetics, Gō models; contact order**

INTRODUCTION

A logical test of any general conception about the driving forces in protein folding is to ascertain whether polymer models incorporating the given idea can predict generic behavior of real proteins.[1,2] In these considerations, self-contained polymer models with explicit chain representations[3] are of particular importance. Quite obviously, the relationship between model energetics and conformational distribution can only be addressed in a physically plausible manner when chain connectivity and excluded volume are adequately taken into account. Using this analytical framework, we found that even mundane protein properties, such as calorimetric two-state cooperativity[4–6] and simple two-state folding/unfolding kinetics,[7,8] are remarkable feats from a polymer standpoint. Simply put, it is nontrivial to construct heteropolymer models with commonly used model interaction schemes to reproduce such behavior. These include popular 2-, 3-, 20-letter models and traditional Gō models (see below).[4–8] Thus, generic protein properties present severe constraints on modeling. Hence, insight into real protein energetics can be gained by requiring self-contained polymer models to satisfy such constraints.

Motivated by the proposed consistency principle[9] or principle of minimal frustration[10] for protein energetics, Gō[11] and Gō-like models (e.g., see Refs. 8, 12–16 and references therein) have long been used in protein folding investigations. These models postulate that only intrachain interactions found in the native (ground-state) conformation are favorable; all other possible intrachain interactions are assumed to be either neutral or unfavorable. Recently, this native-centric approach to modeling has often been justified as well by the discovery that folding rates of natural small single-domain proteins are well correlated with the contact order[17] of their native structures. How well do common Gō models mimic the generic properties of small single-domain proteins? For chain models configured on two-dimensional square lattices, we found that even with their explicit native biases, the common Gō interaction scheme falls far short of producing the type of calorimetric two-state cooperativity observed for many small proteins.[4] Three-dimensional Gō-like lattice[5–7] and continuum[8] models are more protein-like in this regard, because many of them may be consid-
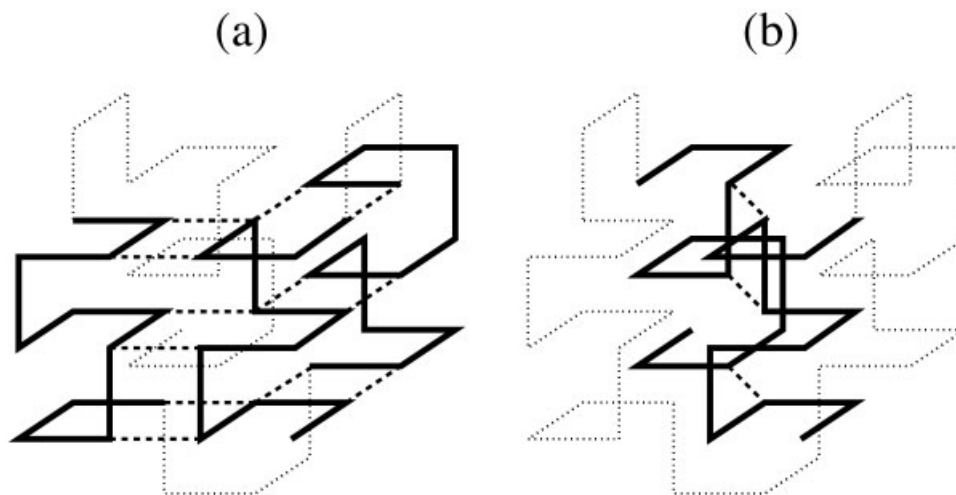
Fig. 1.   Modeling a cooperative interplay between local conformational preference and protein core packing in a 55mer native-centric four-helix-bundle model. Certain native helices are shown as thick solid lines for illustrative purposes. Their sequence positions are indicated by the thin dotted lines depicting the rest of the full native structure. Given a native helix is completely formed [the front right helix appearing in both (**a**) and (**b**) in this example], a favorable cooperative energy $\varepsilon_{\mathrm{coop}}$ (<0) is assigned if either (**a**) four or more of the six native contacts (thick dotted lines) between the given fully formed native helix and each of the two chain segments for the two flanking native helices are present or (**b**) at least two of the three "diagonal" native contacts are present between the given fully formed native helix and the chain segment for the diagonally neighboring native helix, or both (**a**) and (**b**). Thus, condition (**a**) requires at least eight inter-helix nearest-neighbor native contacts, whereas condition (**b**) requires at least two inter-helix next-nearest-neighbor native contacts on the simple cubic lattice. It follows that the maximum total contribution from these cooperative energies is $4\varepsilon_{\mathrm{coop}}$ when all four native helices are completely formed and correctly packed against one another.

ered calorimetrically cooperative if certain lattitude is allowed for empirical baseline subtractions.[5]

However, common Gō-like schemes are apparently not capable of producing simple two-state folding/unfolding kinetics with linear chevron plots.[18] We recently demonstrated this finding in several examples,[7,8] including lattice and continuum (off-lattice) models as well as models with a rudimentary implicit-solvent treatment of desolvation barriers.[8,19] Thus, the inability of common Gō-like constructs to predict simple two-state folding/unfolding kinetics is not an artifact restricted only to lattice Gō models. Most likely, it is a fundamental problem arising from the additive nature of common Gō-like interaction schemes. Our results indicate strongly that such model interaction schemes—on-lattice and otherwise—afford insufficient cooperativity to capture real two-state protein energetics. Here we address this basic question by constructing and testing novel native-centric lattice models that go beyond common additive Gō-like schemes, with intrachain potentials that can lead to simple two-state folding/unfolding kinetics. The ultimate goal of this line of inquiry, of which the present lattice exercise is only a first step, is to decipher the many-body cooperative interactions underlying the behavior of small single-domain proteins.

MODELING A COOPERATIVE INTERPLAY BETWEEN LOCAL CONFORMATIONAL PREFERENCE AND PROTEIN CORE FORMATION

We first explore several native-centric variants of a 55mer model (Figs. 1–4). Their basic features are derived from the original model we put forth recently.[6] For the Gō-like constructs studied here, we retain contributions from the term disfavoring the initiation of left-handed helices (Eq. 1 of Ref. 6). Contributions from the five-letter contact energies are retained for the native contacts in the ground-state conformation, whereas all non-native contacts are assigned zero energy as in common Gō models. The resulting native-centric model has the energy function

$$E = E'_{\mathrm{contact}} + \gamma_{\mathrm{lh}}N_{\mathrm{lh}} \qquad (1)$$

where the prime superscript on the E'_{contact} term indicates that the sum of pairwise five-letter energies is restricted to native contacts, and the second term on the right disfavors left-handed helices. Here we use the same contact energies and γ_{lh} parameter as in Ref. 6. The ground-state energy of the present model equals -36.1. We refer to this as model (i).

Next we consider a model that embodies the idea of a cooperative interplay between local conformational preference and the (nonlocal) packing of the protein core.[4,6,7] This is motivated by the observation that secondary structure formation in globular proteins is often context dependent and that short helices are often not stable in isolation but are stable when packed in the core of a protein (e.g., see Ref. 20 and references therein). Here we mimic this effect by assigning a favorable energy $\varepsilon_{\mathrm{coop}}$ to each incidence of the conformational situation described in Figure 1, leading to the energy function

$$E = E'_{\mathrm{contact}} + \gamma_{\mathrm{lh}}N_{\mathrm{lh}} + \varepsilon_{\mathrm{coop}}h_{\mathrm{c}} \qquad (2)$$
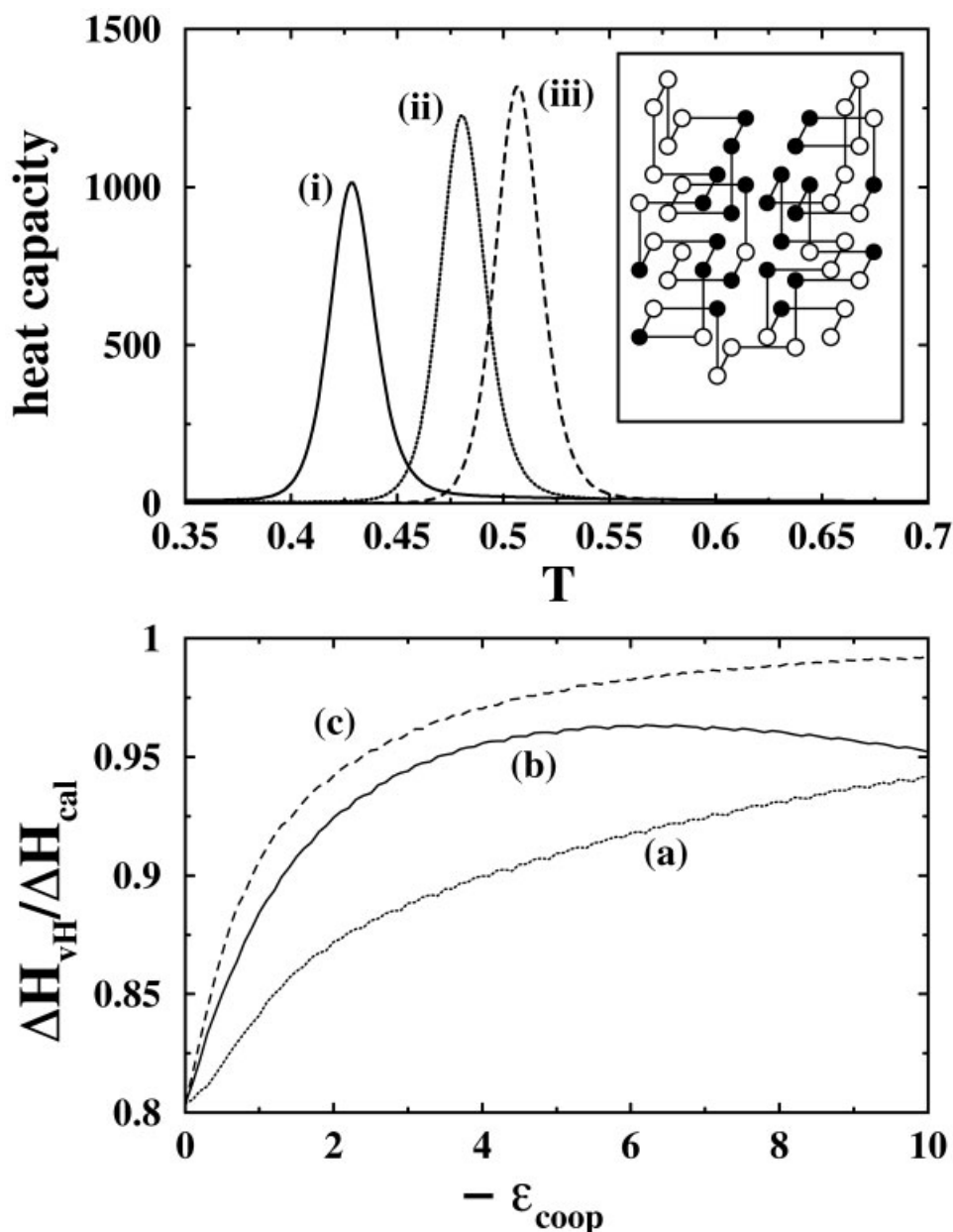
Fig. 2. The overall thermodynamic cooperativity of a model protein is boosted by the many-body interactions described in Figure 1 and in the text. **Top**: Heat capacity as a function of temperature for (i) the pairwise additive native-centric 55mer model (Eq. 1), (ii) a model having the interactions in (i) plus the cooperative interaction scheme in Figure 1 with $\varepsilon_{coop} = -1.0$ (Eq. 2), and (iii) a model with the interactions in (ii) plus an extra favorable energy of $E_{gs} = -2.0$ for the ground-state conformation (Eq. 3). The heat capacity scans were obtained by using Monte Carlo histogram techniques based on conformational sampling around each model's transition midpoint.[5–7] The inset (from Ref. 7) shows the 55mer model ground-state structure with (nominally) hydrophobic and polar residues depicted, respectively, as filled and open circles. **Bottom**: Van't Hoff to calorimetric enthalpy $\Delta H_{vH}/\Delta H_{cal}$ ratios are given by κ_2 defined in Ref. 5 (without empirical baseline subtractions) for three classes of 55mer models whose interaction schemes are parametrized by an ε_{coop} variable. (a) As in (i) above plus an extra favorable energy of ε_{coop} for the ground-state conformation (Eq. 4). (b) As in (ii) above but, instead of fixing $\varepsilon_{coop} = -1.0$, a variable ε_{coop} is used for the helix packing contribution defined by Figure 1 and Eq. 2. (c) As in (b) plus an extra favorable energy of $E_{gs} = \varepsilon_{coop}$ for the ground-state conformation (Eq. 3).

where h_c counts the incidences of a fully formed native helix having the cooperative interactions defined by Figure 1, and $0 \leq h_c \leq 4$ for the present 55mer model. This we refer to as model (ii).

To account for the possibility that optimal packing of the protein core as a whole can impart significant thermodynamic stability to the native structure, we consider yet another model with an extra favorable energy assigned
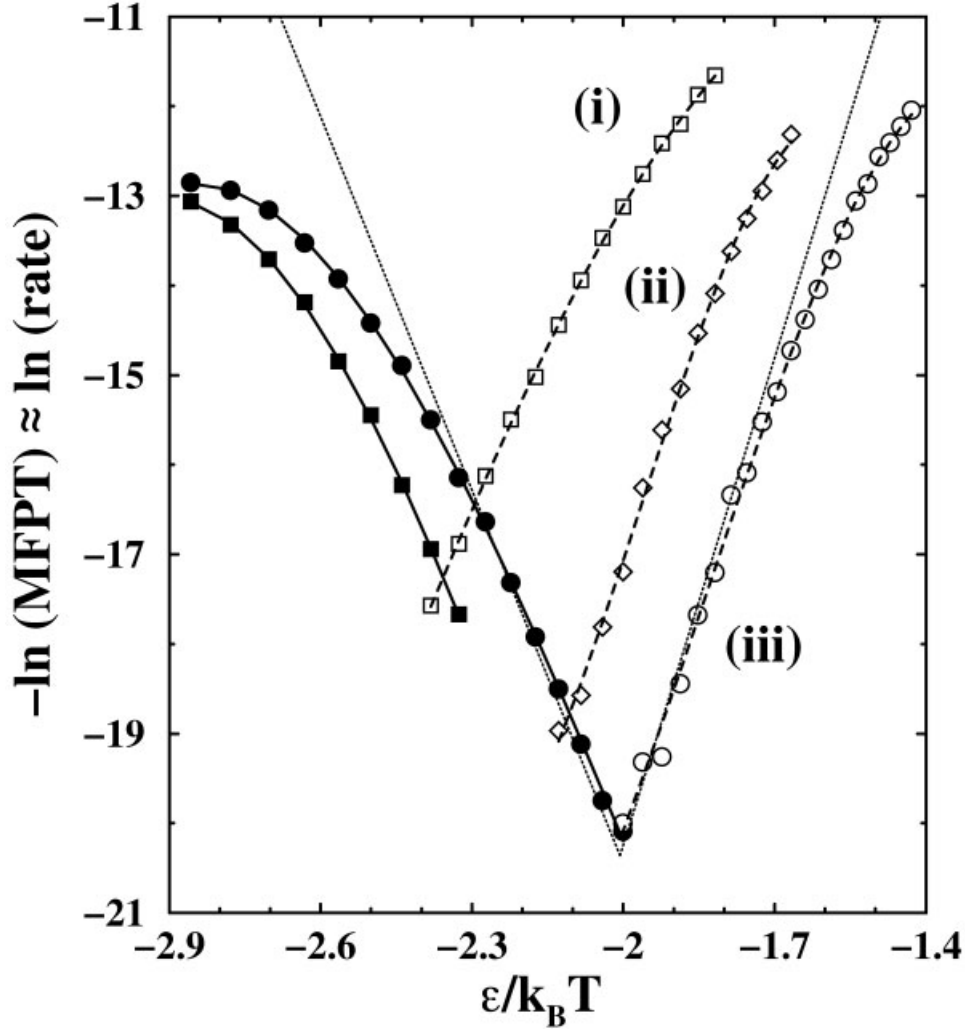
Fig. 3. Chevron plots for the 55mer models (i), (ii), and (iii) in Figure 2 are provided by the negative natural logarithm of mean first passage time (MFPT) as functions of interaction strength ϵ/k_BT. The present Monte Carlo (MC) dynamics simulations use the same general procedure as that in Ref. 7, now with a move set consisting of end flips (4.7%), corner flips (58.3%), crankshafts (27%), and rigid rotations (10%). Folding (filled symbols) starts from a randomly generated conformation. First passage time (FPT) for folding is the number of attempted MC moves needed to reach the ground-state conformation. Unfolding (open symbols) starts from the ground-state conformation. Here unfolding FPT is the number of attempted MC moves needed to reach a conformation with fewer than seven native contacts. Each plotted MFPT is averaged from 1000 trajectories. Solid and dashed curves through the data points are mere guides for the eye. The dotted V-shape for model (iii) is a hypothetical simple two-state chevron plot consistent with the ϵ/k_BT dependence of thermodynamic stability as given by the free energy difference ΔG_u between the ground state and the unfolded conformational ensemble with less than seven native contacts.

only to the native conformation. The energy function now becomes

$$E = E'_{\text{contact}} + \gamma_{\text{lh}}N_{\text{lh}} + \varepsilon_{\text{coop}}h_c + E_{\text{gs}} \qquad (3)$$

where the augmented E_{gs} term takes a nonzero favorable value only when the chain is in its unique ground-state conformation. We refer to this as model (iii). We note that the $\varepsilon_{\text{coop}}$ and E_{gs} terms introduced in Eqs. 2 and 3 are many-body in nature. Many-body interactions have been investigated in the context of protein folding (e.g., see Refs. 6, 21, and 22). However, their relationship with linear chevron plots and simple two-state kinetics has not been much explored.

Figure 2 shows that the many-body cooperative interactions introduced above enhance thermodynamic cooperativity. In calculating the heat capacities of these models, we made the simplifying assumption that the interactions are temperature independent, and set enthalpy equal to the model energy, as in our previous investigations.[5-8] For calorimetric two-state behavior, the van't Hoff to calorimetric enthalpy ratio $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ has to be close to 1.[4-6] Now the $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ ratio (κ_2 in Ref. 5 without empirical baseline subtraction) equals 0.804 for model (i), which does
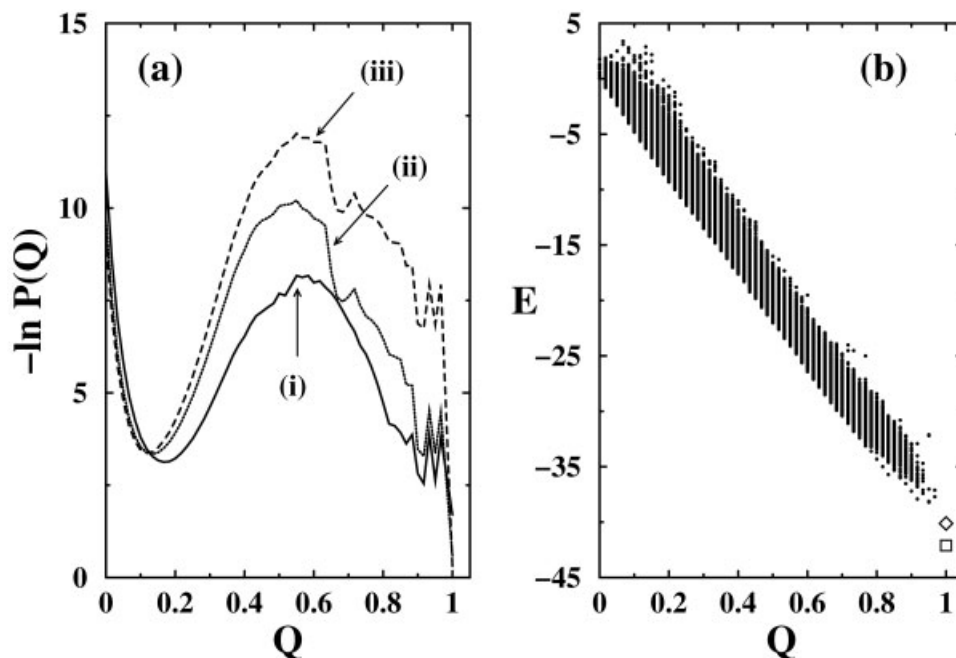
Fig. 4. Distribution of native contacts in the 55mer models. The total (maximum) number of common spatial nearest-neighbor native contacts in the ground-state conformations is 60 [diagonal contacts in Fig. 1(b) are not included in this accounting]. Q is the fractional number of native contacts for a conformation, defined as the number of native contacts it contains divided by the maximum. **a**: The free energy profiles of the models (i), (ii) and (iii) defined in Figure 2 are given by the negative logarithmic distributions of Q. The profiles shown are for $\epsilon/k_B T = -2.34$, -2.11, and -2.0, respectively, for (i), (ii), and (iii), near each model's transition midpoint and were obtained by standard MC histogram techniques.[5–7] **b**: Correlation between energy E and Q for models (ii) and (iii). Dots indicate the existence of conformations with the given E and Q values. The open diamond and square mark the ground-state energies of -40.1 and -42.1 for models (ii) and (iii), respectively.

not contain many-body cooperative interactions. But it is considerably higher at 0.88 and 0.91, respectively, for models (ii) and (iii) with favorable values of ϵ_{coop} and E_{gs} (Fig. 2, top).* This is not too surprising because the many-body cooperative interactions defined above tend to increase the energetic (enthalpic) separation between the ground-state and near ground-state conformations on one hand and the open unfolded conformations on the other, or disfavor conformations with intermediate energy (enthalpy), or both. Both of these effects would lead to higher calorimetric cooperativity.[4,5] The thermodynamic ramifications of these interactions are further explored in the lower panel of Figure 2, which covers a broad range of values for ϵ_{coop} and E_{gs}. An additional scenario in which an extra favorable energy for the ground-state is augmented to model (i), that is,

$$E = E'_{contact} + \gamma_{lh} N_{lh} + E_{gs} \qquad (4)$$

is also studied [curve (a)]. The trend observed in the lower panel of Figure 2 is that stronger many-body cooperative interactions of the type defined above generally lead to higher calorimetric cooperativity. However, there appears to be an upper limit on κ_2 ($\approx$0.96) achievable by the

helix-packing term alone [curve (b)], because at very high $-\epsilon_{coop}$ values, it is possible that some intermediate non-ground-state conformations can become relatively stable (c.f. Fig. 1).

Figure 3 presents the chevron plots for the three models considered in the upper panel of Figure 2. To model folding and unfolding kinetics at different interaction strengths, an energetic scaling parameter ϵ is introduced. At a given ϵ, the effective energy of a conformation with energy E (given by Eqs. 1, 2, or 3) is equal to $-\epsilon E$; and variation in $\epsilon/k_B T$ (at constant T) serves as a model denaturant concentration variation, as in Ref. 7. Figure 3 shows that at sufficiently strong intrachain interaction (more negative $\epsilon/k_B T$), every folding arm of the three chevron plots exhibits a rollover. This finding suggests that chevron rollover is practically unavoidable in polymer models with physically plausible interactions, because when intrachain interactions become generally very favorable, kinetic trapping is bound to increase in importance.[7,8] However, native thermodynamic stability would be extremely high when the model parameter $\epsilon/k_B T$ becomes extremely negative. Many such situations are not physically realizable in real proteins,[7] whose native stabilities even in zero denaturant are often marginal. In this light, in comparing Figure 3 with experiments, the relevant question is whether there is a quasi-linear regime of the model chevron plots that is consistent with the two-state thermodynamics of

---

*For every model considered in Figure 2, the $\kappa_2^{(s)}$ (Ref. 5) value for the $\Delta H_{vH}/\Delta H_{cal}$ ratio after empricial baseline subtraction equals 1.0.

the given model and covers a range of thermodynamic stability similar to that of real, simple two-state proteins.

Pursuing this logic, we note that folding rollover occurs quite near to the transition midpoint for model (i), but the ϵ/k_BT range of a quasi-linear regime is more extended for the two more cooperative models (ii) and (iii). The folding arms of models (ii) and (iii) are identical because, by construction, although the E_{gs} term in Eq. 3 slows unfolding, it does not affect the kinetics of folding. For model (iii), we have used standard histogram techniques and extensive conformational sampling[7] at $\epsilon/k_BT = -2.105$ to determine the dependence of the free energy of unfolding ΔG_u on ϵ/k_BT (detailed data not shown). The resulting thermodynamic relation, which is essentially linear and has approximately the same transition midpoint as determined from the kinetic chevron plot, was applied to construct the dotted V-shape in Figure 3. The close agreement between the dotted V-shape and the simulated chevron plot for model (iii) from $\epsilon/k_BT \approx -2.3$ to -1.6 implies that the folding/unfolding kinetics of model (iii) is consistent with a simple two-state description within this regime. The strongest intrachain interaction in the two-state regime is at $\epsilon/k_BT \approx -2.31$, which corresponds to a native stability $\Delta G_u \approx 10 k_BT$ for this particular model. It is clear from comparing the chevron plots of models (ii) and (iii) that the linear regime can be readily extended by increasing the magnitude of E_{gs} beyond that in model (iii). But even as it stands, model (iii)'s behavior in Figure 3 should provide a semi-quantitative rationalization for the simple two-state kinetics of many small single-domain proteins. Indeed, more than half of the 24 two-state proteins listed by Plaxco et al.[2] have native stabilities around 25°C comparable to or lower than 10 k_BT. For example, $\Delta G_u = 3.6\ k_BT$ (2.1 kcal/mol) for CspB at 25°C and pH 7.0,[23] and $\Delta G_u = 9.0\ k_BT$ (5.3 kcal/mol) for protein L at 22°C and pH 7.0.[24] Therefore, in our view, simple two-state folding/unfolding kinetics emerges as a limiting case phenomenon when the hypothetically high native stability at which chevron rollover would occur is not attainable by a small single-domain protein. Conversely, rollover becomes observable when a protein fails to achieve a sufficiently high thermodynamic cooperativity commensurate with its native stability.

A comparison between the kinetic properties in Figure 3 and the thermodynamic properties in Figure 2 indicates that the thermodynamic requirement of simple two-state behavior is stringent, allowing only for a small adjustment from empirical baseline subtraction.[5] Apparently, a model has to be nearly or as cooperative as model (iii) or more to achieve a reasonable reproduction of simple two-state protein folding/unfolding kinetics. This finding suggests strongly that, in modeling situations when the heat capacity contributions from bond vibrations are not considered (as in the present cases) and intrachain interaction energies are taken to be temperature-independent, a without-baseline-subtraction $\Delta H_{vH}/\Delta H_{cal}$ ratio of $\kappa_2 > 0.9$ would likely be required for simple two-state kinetics [Fig. 2, upper panel, model (iii)].

We have also checked that folding kinetics is essentially single exponential within the quasi-linear regime by confirming that the logarithmic distribution of folding first passage times is approximately linear.[7,8,25] In fact, extensive testing for 10 values of ϵ/k_BT between -2.22 and -2.78 covering the quasi-linear regime and beyond indicates that they are consistent with single exponential relaxation. Unfolding kinetics is essentially single exponential as well (detailed data not shown). As in our previous study,[7] the onset of nonexponential folding relaxation at interaction strength $\epsilon/k_BT \approx -2.9$ is concomitant to that of a drastic chevron rollover.

Not surprisingly, Figure 4(a) shows that the free energy barrier separating the native and denatured states is higher for a more cooperative model, consistent with its slower folding and unfolding rates at the transition midpoint (c.f. Fig. 3). Figure 4(b) shows that the relation between energy and the number of native contacts is approximately linear for the two cooperative models. In this regard, the present exercise suggests that certain many-body interactions embodying a local-nonlocal cooperative interplay ($\epsilon_{coop} = -1.0$), and an added ground-state stability ($E_{gs} = -2.0$) in proteins can lead to remarkable improvements in kinetic cooperativity even when the magnitudes of these terms are relatively small.

MODELING A PARTIAL SEPARATION BETWEEN THE INTERACTIONS FOR THERMODYNAMIC STABILITY AND THE DRIVING FORCES FOR FOLDING KINETICS

Having established a plausible scenario for simple two-state protein folding/unfolding kinetics, we proceed to broaden our exploration and to better delineate how various energetic components might contribute to this remarkable behavior. As a first step, we consider in this section a somewhat different class of models in which a local-nonlocal cooperative interplay is absent but the unique ground-state conformation is favored by an extra strong energy. The interaction scheme is a simplified version of Eq. 4 above, with the energy function

$$E = E_{Gō} + E_{gs} \qquad (5)$$

where $E_{Gō}$ is the usual lattice Gō potential that assigns a favorable energy ϵ_0 (<0) to every contact in the native structure of the model and assigns zero energy to all other (nonnative) contacts, and E_{gs} applies only to the ground-state conformation, as in Eqs. 3 and 4. To explore the effect of native topology, we study three cubic-lattice 27mer models with E given by Eq. 5 for three different ground-state structures (Figs. 5 and 6).

As discussed above, the E_{gs} term serves to increase native stability and enhance thermodynamic stability, leading to a reduced unfolding rate. But it has no effect on the folding kinetics modeled by Monte Carlo dynamics with the Metropolis acceptance criterion.[26,27] Thus, the energetics described by Eq. 5 entails a partial separation between the interactions that drive the protein to fold kinetically (the pairwise contact energies $E_{Gō}$) and the interactions that stabilize the ground-state structure ($E_{Gō}$
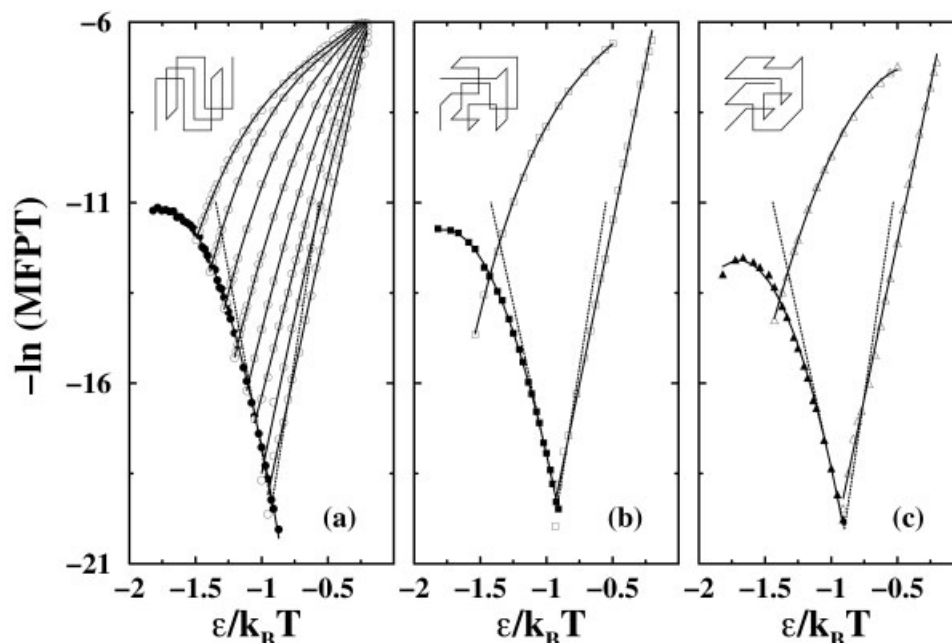
Fig. 5.   Chevron plots for three different 27mer Gō models and for their corresponding models with an extra favorable energy E_{gs} assigned to the native (ground-state) structure (as shown), with $\epsilon_0 = -1$ (Eq. 5). Folding MFPT is independent of E_{gs}, and ϵ has the same meaning as in Figure 3. Here each MFPT is averaged from 500 trajectories. Folding (filled symbols) and unfolding (open symbols) simulations were performed as for Figure 3 except only local chain moves were used for the present 27mer MC dynamics simulation (no rigid rotations), and unfolding FPT is now defined by the time needed to reach a conformation with fewer than four native contacts from a given starting ground-state conformation with 28 native contacts. Solid curves are mere guides for the eye. **a**: Unfolding chevron arms for eight models with different degrees of cooperativity are shown; from top to bottom, $E_{gs} = 0$ (common additive Gō model), $E_{gs} = -2, -4, -6, -8, -10, -12$, and -14. **b, c**: Unfolding chevron arms for the common additive Gō model ($E_{gs} = 0$, upper curves) are compared with the unfolding arms for $E_{gs} = -14$ (lower quasi-linear curves). The dotted V-shapes are hypothetical simple two-state chevron plots consistent with the $\epsilon/k_B T$ dependence of ΔG_u between the ground state and the unfolded conformational ensemble with less than four native contacts of a given model; ΔG_u values are determined by standard histogram techniques[5–7] based on conformational sampling at $\epsilon/k_B T = -0.935$, -0.917 and -0.917 for (a), (b), and (c), respectively (detailed data not shown).

and the many-body E_{gs}). Because $E_{Gō}$ contributes partially to native stability, the role separation just described becomes a more predominant feature of the model when E_{gs} is large compare to ϵ_0. A similar mechanism of partial separation between folding-kinetics and native-stabilizing interactions is also presumed by Eq. 3 [model (iii)] and Eq. 4 above. Our interest in this scenario was partly motivated by experimental studies showing that mutants of a wild-type protein are much less likely to have a slower unfolding rate than to have a faster folding rate. For example, among the 41 mutants of Fyn SH3 domain studied by Northey et al.,[28] only 3 have slightly reduced unfolding rates relative to that of the wildtype, whereas 5 times as many (15) mutants have folding rates faster than that of the wildtype. This means that interactions that accelerate folding do not necessarily lead to higher native stability (only four mutants are more stable than the wildtype), presumably because some mutants that fold fast do not pack well when folded[28] (c.f. discussion of conformational strain by Ventura et al.[29]). These observations suggest that a partial separation of folding-kinetics and native-stabilizing intraprotein interactions, as envisioned by the E_{gs} term, is physically plausible.

Figure 5 shows that combining a pairwise Gō potential

with an E_{gs} term can also lead to simple two-state protein folding/unfolding kinetics, although for these relatively short chains the E_{gs}/ϵ_0 ratio needed to achieve simple two-state behavior is large. Figure 5(a) provides a series of unfolding chevron arms for different E_{gs} values, showing clearly that the quasi-linear regime of the chevron plot can be extended by a more negative E_{gs}. For all three models considered in Figure 5 with $E_{gs}/\epsilon_0 = 14$, approximately simple two-state behavior persists to $\epsilon/k_B T \approx -1.15$ (c.f. simulated chevron plots and dotted V-shapes), corresponding to native stabilities $\Delta G_u \approx 10 \ k_B T$. As for model (iii) above (Fig. 3), for each model in Figure 5, we have verified that folding relaxation is essentially single exponential for $\epsilon/k_B T > -1.6$ by obtaining linear logarithmic first passage time distributions for six $\epsilon/k_B T$ values from -1.67 to -0.91. Unfolding relaxation is also essentially single exponential (detailed data not shown). Folding relaxation becomes nonexponential (for $\epsilon/k_B T < -1.8$) only when native stability is much higher than that spanned by the simple two-state regime between $\epsilon/k_B T \approx -1.15$ and $\epsilon/k_B T \approx -0.723$.

Figure 6 compares chevron plots of the three cooperative 27mer models. The rank ordering of their folding rates is consistent with a correlation between slower folding rate
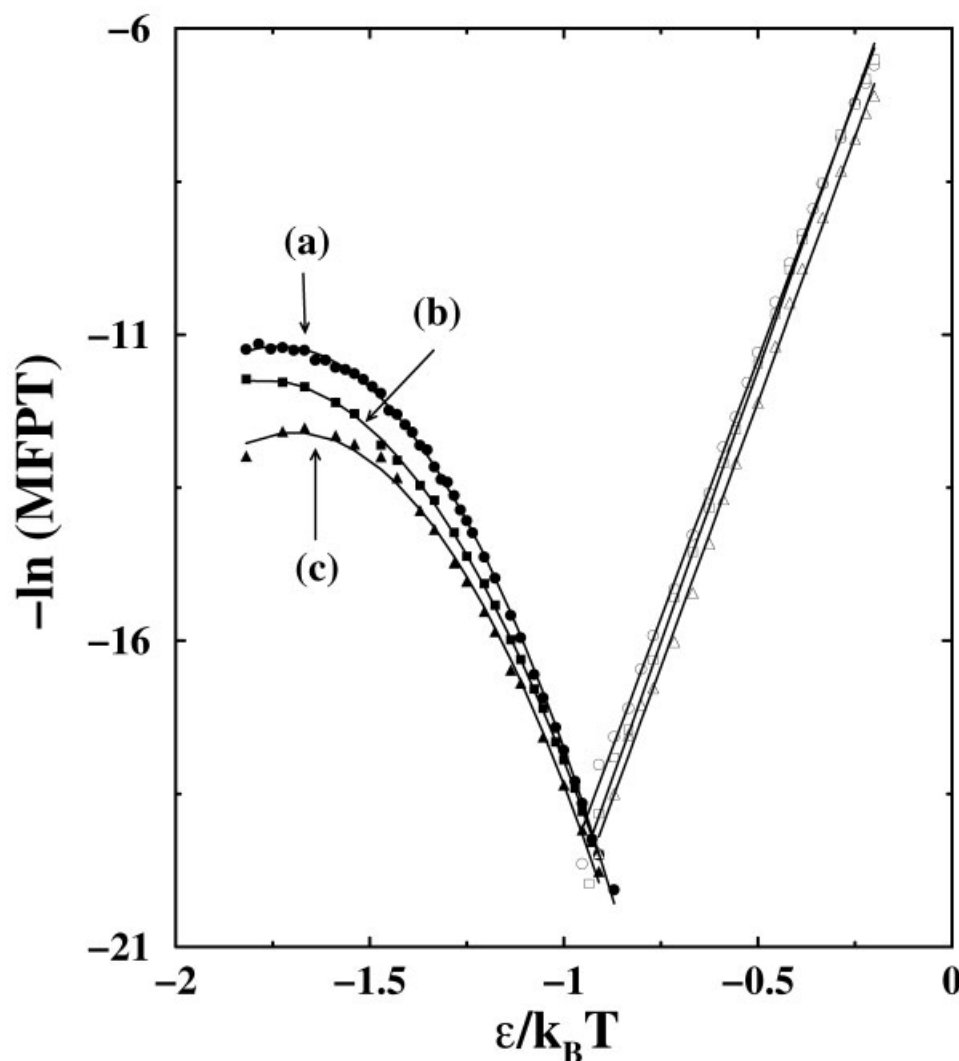
Fig. 6.   Comparing chevron plots for the three different models in Figure 5 with $E_{gs} = -14$ shows that a model with lower native contact order (CO, as defined in Ref. 17) tends to fold slightly faster. Here CO = 0.28, 0.40, and 0.51 for models (a), (b), and (c), respectively.

and higher relative contact order (CO).[17] However, for these models, the dependence of folding rate on CO is weak. Near the onset of drastic chevron rollover and nonexponential folding relaxation ($\epsilon/k_BT \approx -1.75$), the 27mer model with CO = 0.51 folds only approximately 4 times slower than the 27mer model with CO = 0.28. The dispersion in folding rate is even smaller within the simple two-state regime. This is a far cry from the six orders of magnitude of variation in folding rates observed among real, small, single-domain proteins.[2] Recently, CO-dependent folding rates have been addressed by using explicit-chain models with limited yet encouraging successes. Using a Gō-like potential for 18 small single-domain proteins, Koga and Takada[30] obtained a correlation between CO and folding rate, but the variation in rates covered only $\approx$ 1.5 orders of magnitude. More recently, Jewett et al[31] conducted an extensive lattice 27mer simulation study using cooperative Gō-like models with a nonlinear E–Q relation. A correlation between CO and folding

rate was found, but again the dispersion in folding rates spanned only 1 to 1.5 order of magnitude. Although the mechanisms and energetics of CO-dependent folding remain to be better elucidated,[32] our more recent investigation shows that models with a local-nonlocal cooperative interplay similar to that in models (i) and (ii) (Eqs. 2 and 3) above can lead to a relatively large dispersion in folding rates and a better correlation between CO and folding rate.[33,34]

A RATIONALIZATION OF NON-ARRHENIUS PROTEIN FOLDING/UNFOLDING KINETICS

The physics embodied by the extra favorable ground-state energy E_{gs} in the models described by Eqs. 3–5 above implies that there is a fundamental asymmetry between folding and unfolding kinetics.[28] This led us to ask whether the same physical picture may shed light on the significant difference in the degree of deviation from Arrhenius kinetics for folding versus unfolding that are often ob-

TABLE I. Deviations From Arrhenius Behavior in Protein Folding and Unfolding

Protein name	"$(\Delta C_P^{\ddagger})_{\mathrm{f}}/(\Delta C_P^{\ddagger})_{\mathrm{u}}$"
T4 lysozyme mutant[a]	−2.99
CI2[b]	−2.95
CspB[c]	−9.0
Protein L[d]	−1.68
NTL9[e]	−1.33

[a]Chen et al.[36]
[b]Jackson and Fersht[37]
[c]Schindler and Schmid[23]
[d]Scalley and Baker[24]
[e]Kuhlman et al.[38]

served in experiments. Early measurements by Segawa and Sugihara[35] showed that the folding kinetics of hen eggwhite lysozyme was significantly non-Arrhenius (logarithmic folding rate $\ln k_{\mathrm{f}}$ nonlinear in $1/T$), whereas the unfolding kinetics was essentially Arrhenius (logarithmic unfolding rate $\ln k_{\mathrm{u}}$ linear in $1/T$). Table I summarizes more recent experimental data from the literature for several proteins with simple two-state folding/unfolding kinetics and whose temperature-dependent rates of both folding and unfolding have been measured directly. For the proteins listed, the trend that folding is more non-Arrhenius than unfolding is quantified by reported activation heat capacities for folding $(\Delta C_P^{\ddagger})_{\mathrm{f}}$ that are significantly larger in magnitude than the corresponding activation heat capacity for unfolding $(\Delta C_P^{\ddagger})_{\mathrm{u}}$. Table I puts the "$(\Delta C_P^{\ddagger})_{\mathrm{f}}/(\Delta C_P^{\ddagger})_{\mathrm{u}}$" ratio in quotation marks because the common approach of using temperature-independent activation heat capacities to analyze folding/unfolding kinetics data may be problematic.[17] We note that another potential source of the difficulty is that possible temperature dependencies of the heat capacities[39] associated with protein folding/unfolding transitions were not considered in such analyses. Nonetheless, as an empirical parameter, "$(\Delta C_P^{\ddagger})_{\mathrm{f}}/(\Delta C_P^{\ddagger})_{\mathrm{u}}$" serves well to show that $\ln k_{\mathrm{f}}$ is often significantly more curvilinear in $1/T$ than $\ln k_{\mathrm{u}}$.

This trend can be captured qualitatively in the present modeling context if the solvent-mediated driving forces for folding kinetics and the many-body native-stabilizing interactions are taken to have different temperature dependencies. This is a physically plausible assumption because some intraprotein solvent-mediated forces such as the hydrophobic effect are known to be sensitive to the sizes and shapes of the interacting groups.[40–42] Here we use a collection of cooperative 27mer models in Figure 5(a) with different values of E_{gs} to expound the principles involved. Temperature-dependent interactions are now introduced by letting the pairwise contact energy ϵ_0 to vary with temperature in a hydrophobic-like manner while leaving E_{gs} temperature-independent.

The following schematic analysis of T-dependent folding and unfolding rates (Fig. 7) is similar to that introduced by Chan[26] and Chan and Dill.[27] However, the present focus on folding/unfolding asymmetry in the context of three-dimensional protein chain models was not addressed in these earlier studies of short-chain two-dimensional mod-

els.[26,27] The first step in the present analysis is to obtain from Figure 5(a) the logarithmic folding and unfolding rates $\ln k_{\mathrm{f}}$ and $\ln k_{\mathrm{u}}$ [which are taken to be their respective $-\ln(\mathrm{MFPT})$] as functions of ϵ_0 and E_{gs}. Because the effective energy is given by $-\epsilon E$ in Figure 5 (see above) where E is given by Eq. 5 with ϵ_0 set to -1, each (ϵ, E_{gs})-dependent data point in Figure 5(a) may be regarded as the folding or unfolding rate for the energy function E itself with an ϵ_0 value equals that of ϵ and an E_{gs} value equals ϵ times the E_{gs} for the given unfolding chevron arm. We note that within the quasi-linear single-exponential regime,

$$\ln k_{\mathrm{f}} = \alpha_{\mathrm{f}} + \beta_{\mathrm{f}} \frac{\epsilon_0}{k_{\mathrm{B}}T} \tag{6}$$

holds approximately for constant α_{f} and β_{f} because folding kinetics is independent of E_{gs}. A least-square fit yields $\beta_{\mathrm{f}} = -15.4$. For unfolding within the quasi-linear single-exponential regime, the approximate linear relation

$$\ln k_{\mathrm{u}} = \alpha_{\mathrm{u}} + \beta_{\mathrm{u}} \frac{\epsilon_0}{k_{\mathrm{B}}T} + \beta_{\mathrm{u}}' \frac{E_{\mathrm{gs}}}{k_{\mathrm{B}}T} \tag{7}$$

is expected, where α_{u}, β_{u}, and β_{u}' are constants. Extensive analyses indicate that $\beta_{\mathrm{u}} \approx 4.0$ and $\beta_{\mathrm{u}}' \approx 1.0$. We use $\beta_{\mathrm{u}} = 3.9$, $\beta_{\mathrm{u}}' = 1$ below. Figure 7(a) shows that these values fit the simulated unfolding rates of the more cooperative models extremely well.

Next, a hypothetical temperature-dependent $\epsilon_0 = \epsilon_0(T)$ is introduced in Figure 7(b). This functional form for $\epsilon_0/k_{\mathrm{B}}T$ (solid curve, left scale) was motivated by the temperature dependence of hydrophobic effects[39] and is similar to that explored in Refs. 26 and 27. It follows that the temperature-dependent folding rate in the quasi-linear single-exponential regime is now given by Eq. 6 above with $\epsilon_0 \to \epsilon_0(T)$ provided by Figure 7(b), that is

$$\ln k_{\mathrm{f}}(T) = \alpha_{\mathrm{f}} + \beta_{\mathrm{f}} \frac{\epsilon_0(T)}{k_{\mathrm{B}}T} \tag{8}$$

Similarly, the temperature-dependent unfolding rate in the quasi-linear single-exponential regime of an effective $E_{\mathrm{gs}} = -14$ cooperative model is given by Eq. 7 above with $\epsilon_0 \to \epsilon_0(T)$ provided by Figure 7(b), whereas E_{gs} remains temperature independent:

$$\ln k_{\mathrm{u}}(T) = \alpha_{\mathrm{u}} + \beta_{\mathrm{u}} \frac{\epsilon_0(T)}{k_{\mathrm{B}}T} + \frac{E_{\mathrm{gs}}}{k_{\mathrm{B}}T} \tag{9}$$

where the $E_{\mathrm{gs}}/k_{\mathrm{B}}T$ term is set equal to -14 for a reference temperature (T^*) in Figure 7(c) at which $\epsilon_0/k_{\mathrm{B}}T = -1$. Hence, $E_{\mathrm{gs}}/k_{\mathrm{B}}T = -14(T^*/T)$ is linear in $1/T$. These temperature-dependent folding and unfolding rates are plotted in the upper part of Figure 7(c). It is clear that folding is significantly more non-Arrhenius than unfolding because the only source of non-Arrhenius behavior in the present formulation of the model is $\epsilon_0(T)$, and $\ln k_{\mathrm{f}}$ depends more strongly on $\epsilon_0(T)$ ($\beta_{\mathrm{f}} = -15.4$) than $\ln k_{\mathrm{u}}$ ($\beta_{\mathrm{u}} = 3.9$).

One missing physical ingredient in the consideration thus far is that intrinsic conformational transition rates should accelerate at higher temperature. This is not taken
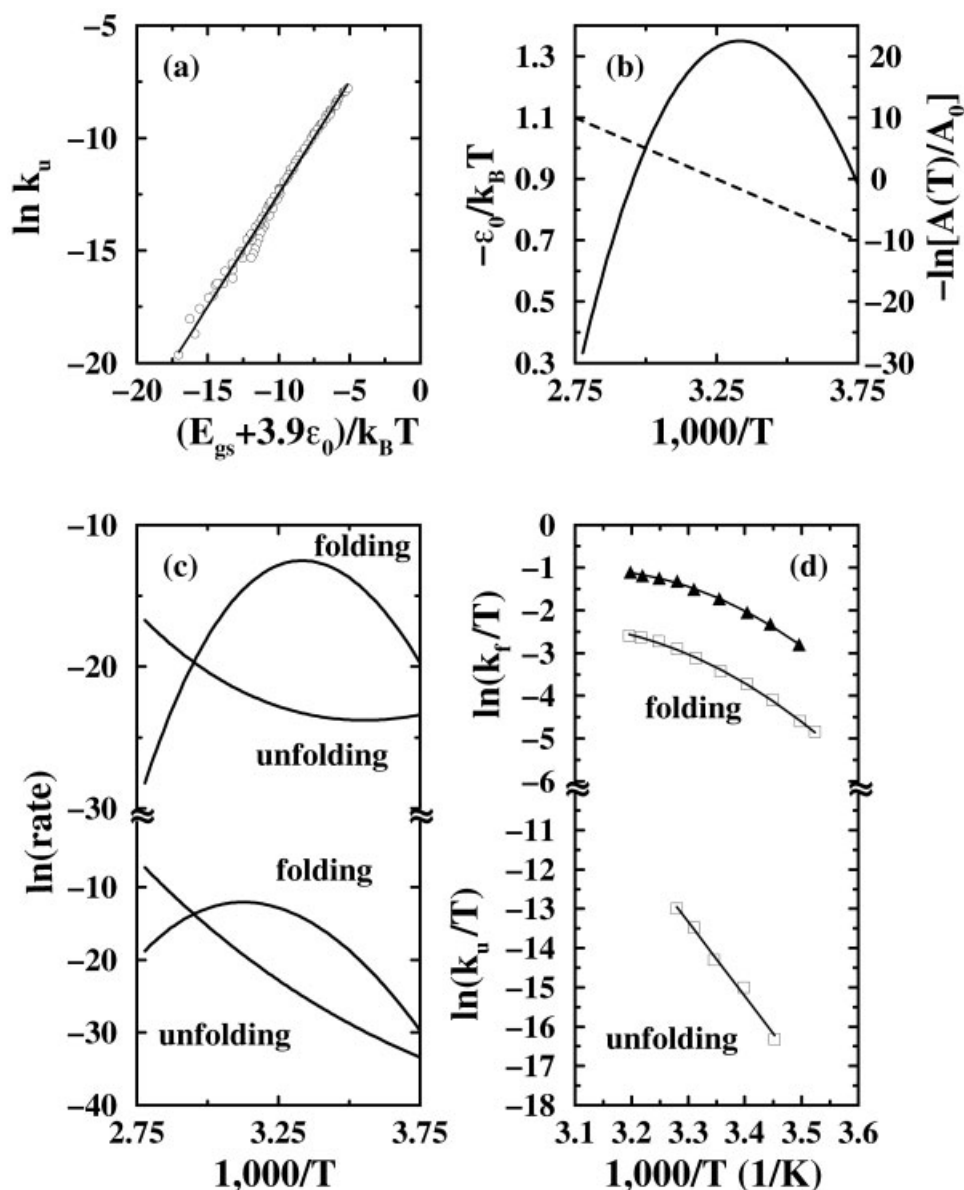
Fig. 7. Rationalizing non-Arrhenius protein folding and unfolding kinetics. **a**: A linear fit of the logarithmic unfolding rates (vertical axis) of the 27mer models in Figure 5(a) with $-E_{gs} \geq 6$ to the expression shown (horizontal axis). For the data points plotted, the correlation coefficient $r = 0.998$. **b**: Solid curve: a hypothetical hydrophobic-like temperature dependence of the model interaction strength $-\epsilon_0/k_B T$ that drives folding kinetics (left vertical scale). Dashed line: a hypothetical temperature dependence of the intrinsic conformational transition rate $1/A(T)$ relative to a reference rate $1/A_0$ (right vertical scale). More analytical details of this physical picture are provided in Refs. 26 and 27. (Note that a typographical error should be corrected in the caption for Figure 4 of Ref. 26: "$\Delta S_0 = 5.2$" should read "$\Delta S_0 = -5.2$.") **c**: Upper curves: Temperature-dependent folding and unfolding rates obtained by combining the solid curve in (b) and data from Figure 5(a) (Eqs. 8 and 9, temperature-dependent intrinsic conformational transition rate not taken into account). Lower curves: Temperature-dependent folding and unfolding rates obtained by combining the solid curve and dashed line in (b) and data from Figure 5(a) (Eq. 10, temperature-dependent intrinsic conformational transition rate taken into account). **d**: Included for comparison are the temperature-dependent CI2 folding and unfolding rates at 25°C and pH 6.3 from the experiments of Jackson and Fersht (adapted from Fig. 4 of Ref. 37). The upper and lower folding curves were for 0 M and 1.5 M GdnHCl, respectively, the unfolding curve was measured at 0 M GdnHCl.

into account if physical time is simply identified with number of attempted Monte Carlo moves, as in the analysis above. This issue has been identified and addressed in some detail in Refs. 26 and 27. As in these references,

Figure 7(b) introduces an adjustment factor (dashed line, right scale) to better mimic physical time. Here $A(T)$ is the temperature-dependent time needed for a given kinetic process and A_0 is a reference time. Thus, the hypothetical

$-\ln[A(T)/A_0]$ function in Figure 7(b) stipulates that the intrinsic logarithmic rate $[-\ln A(T)]$ is higher at higher temperatures (i.e., the Monte Carlo clock should run faster at higher T). This adjustment factor is readily incorporated[26,27] by setting

$$\ln(\text{folding rate}) = \ln k_f(T) - \ln[A(T)/A_0]$$

$$\ln(\text{unfolding rate}) = \ln k_u(T) - \ln[A(T)/A_0]$$

$$(10)$$

where $\ln k_f$ and $\ln k_u$ on the right-hand side are the expressions given, respectively, by Eqs. 8 and 9. Macroscopically, this amounts to introducing an additional enthalpic contribution[24,26] to the free energy barrier of protein folding. The lower part of Figure 7(c) shows that incorporating a $-\ln[A(T)/A_0]$ term can lead to a more realistic description of temperature-dependent protein folding and unfolding rates [c.f. experimental data in Figure 7(d); see also Fig. 3 of Ref. 23, Fig. 1(C) of Ref. 24, Fig. 3 of Ref. 36, and Fig. 5 of Ref. 38].

DISCUSSION: A NEAR-LEVINTHAL SCENARIO FOR SIMPLE TWO-STATE PROTEIN FOLDING/ UNFOLDING KINETICS

The results of the present investigation suggest strongly that the physical interactions underlying the simple two-state folding/unfolding kinetics of small single-domain proteins should involve many-body effects beyond that stipulated by common additive Gō models, even though the physico-chemical origins of these effects remain to be elucidated. Therefore, with regard to protein thermodynamic and kinetic cooperativity, common Gō models with pairwise additive contact energies are not ideal. Apparently, the type of many-body interactions that are conducive to simple two-state kinetics also lead to higher thermodynamic cooperativity and entail a partial separation between folding-kinetics and native-stabilizing interactions. Historically, an impetus to formulate the Levinthal paradox might have been the discovery in the late 1960s that some proteins were calorimetrically two-state[43] (see Ref. 4 and references therein). Naturally, an extreme interpretation of the $\Delta H_{vH}/\Delta H_{cal} = 1$ property would imply that only two enthalpy levels exist (native and denatured), and thus the landscape should resemble a golf course [Fig. 8(a)]. However, a golf course landscape dictates that folding would be exceedingly slow, but the folding of real proteins is relatively fast. To address the "why is folding fast" question, recent theoretical discussions emphasize the funnel-like nature of the protein-folding energy landscape as a solution to the Levinthal paradox[44,45] [Fig. 8(b)], and common Gō potentials are often used to model a relatively smooth funnel-like energy landscape.[12] We found that calorimetric two-state cooperativity can be consistent with the funnel-like landscapes of three-dimensional Gō models, provided some lattitude is allowed for empirical baseline subtractions. This is because in these native-centric models, the conformational populations with intermediate energies (enthalpies)—though not zero—are relatively low.[5–8]
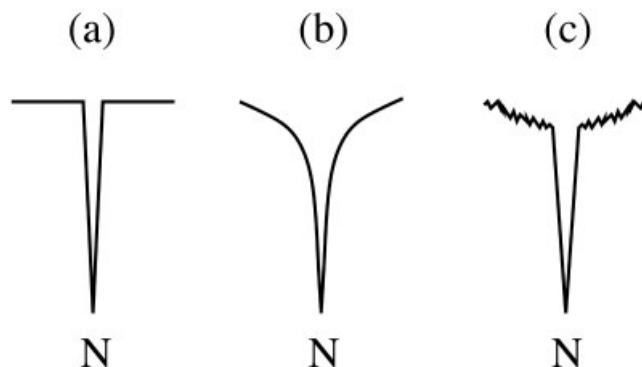

Fig. 8. Schematics of hypothetical and proposed energy landscapes for protein folding. **a**: A golf-course or "Levinthal" landscape. **b**: A funnel landscape. **c**: A "near-Levinthal" scenario.

However, as discussed above, common Gō models are insufficient for simple two-state protein-folding kinetics.[7,8] Because kinetic traps are still significant in these constructs under native conditions, the chevron plots they predicted have severe rollovers.[46] Another shortcoming of common Gō models is that their predicted folding rates are often too fast compared to that of real proteins.[8,14] In one example, it is at least four orders of magnitude faster.[8] So, in a sense, in the context of recent Gō-modeling efforts, the critical question has shifted from "why is folding fast" to "why is folding slow." The present study concludes that a thermodynamic cooperativity higher than that afforded by common additive Gō models is necessary for simple two-state kinetics (Figs. 2 and 3). This scenario also offers a clue to the "why is folding slow" question. For real, small, single-domain proteins, it appears that the key to avoiding kinetic traps and chevron rollover is to have only weakly favorable intrachain interactions during the folding process [gentle upper slopes of the funnel in Fig. 8(c)] until a significant fraction of the chain is native-like and ready to come together to form a large number of native contacts at once, at which point strong cooperative many-body effects kick in to stabilize the structure [steep lower slopes of the funnel in Fig. 8(c)]. This idea is implemented in the cooperative models studied here. Indeed, in Figures 3 and 5, the quasi-linear folding regimes of the cooperative models are in the weakly interacting unfolding regimes (small $-\epsilon/k_B T$) of the corresponding additive Gō models. Thus, for folding of the cooperative models, the energetic bias is not very strong during most of the conformational search. This feature serves to diminish the effects of kinetic traps [shallow minima on the gentle upper slopes of Figure 8(c)] because the depths of kinetic traps in heteropolymers are often correlated with the overall energetic bias toward the native structure.[27] As a result, folding is faster in the cooperative models relative to other heteropolymer models with deep kinetic traps. But at the same time, the very feature of a weakened energetic bias toward the native structure during most of the conformational search also leads to slower folding in comparison with common Gō models, because the latter have stronger native biases during the corresponding kinetic process.

Nonetheless, the reduction in folding rate relative to common Gō models does not make the cooperative models less proteinlike, because even a small bias is sufficient to circumvent the Levinthal paradox, as all of our models fold. In this way, the cooperative model scenario provides a physical plausible answer to the "why is folding slow" question posed above. In fact, for real proteins, folding rates may be further reduced by the inevitable presence of favorable nonnative interactions, which are not taken into account by the native-centric cooperative models here. Anti-cooperativity of certain hydrophobic interactions[47] may also play a role in discouraging premature chain collapse. In this scenario, the energy landscape of a simple two-state protein is still funnel-like but with a narrow bottleneck [Fig. 8(c)]. Thus, the resulting highly bimodal distribution of energy and high thermodynamic cooperativity approach (although never equal) that of a hypothetical Levinthal golf course.[†]

Part of the present view is similar to that of Jewett et al.,[31] who recently introduced a native-centric model in which the energy E of a conformation—unlike that in common Gō models—does not decrease linearly with the number of native contacts. (Conformations with more negative E values are more favorable.) In their model, the rate of decrease in E with increasing fractional number of native contact Q becomes progressively higher as the native structure is approached (when $Q \to 1$). This means that the energetic bias toward the native structure is not strong during the initial stages of folding when there are few native contacts (small Q) but becomes stronger when the native structure is approached during the final stages of folding ($Q \to 1$). Insofar as this general trend is concerned, the physical picture of cooperative folding discussed above (especially the model defined by Eq. 5) is very much similar to that of Jewett et al. Nevertheless, although both the model of Jewett et al. and the present cooperative models have high degrees of thermodynamic cooperativity, their underlying mechanisms are not identical. More recent investigations indicate that detailed kinetic features, such as the correlation between CO and folding rate, do depend significantly on how thermodynamic cooperativity is achieved microscopically.[33,34] Because the lattice model of Jewett et al. was inspired by the more general topomer search model of folding,[48] it would be extremely interesting to compare in future investigations the relationships between the topomer search model and the several different scenarios of thermodynamic cooperativity explored here.

As we have emphasized,[4,5] the experimental calorimetric two-state criterion, which has proven useful for evaluating protein chain models,[4–8,33,34,49–52] does not imply that

there are only two infinitely sharp energy (or enthalpy) levels. In other words, the calorimetric two-state criterion does not preclude the existence of "partially unfolded" conformations with energies intermediate between the energy distribution peaks under strongly folding and strongly denaturing conditions (c.f. Fig. 16 of Ref. 4, Fig. 10 of Ref. 5, Fig. 1 of Ref. 7, and Figs. 7–9 of Ref. 8). These theoretical findings are consistent with native-state hydrogen exchange experiments.[53] Conformational populations with intermediate energies in several calorimetrically cooperative models tested thus far (see figure references above) share some similarities with that in calorimetrically non-cooperative constructs such as certain HP models[4] and a 15mer 20-letter side-chain model.[5,54,55] However, the critical difference between calorimetric cooperative and non-cooperative models is that, at the transition midpoint, conformational populations with intermediate energies are not significant for calorimetrically cooperative models but are significant for calorimetrically non-cooperative models. This difference is well characterized by the $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ ratio.[4,5,49–52] Our approach of evaluating chain models by experimental cooperativity criteria is designed to address what kind of elementary intrachain interactions may be needed to produce the generic cooperative features of proteins, while taking into account as much as possible that proteins are polymers and therefore chain connectivity and excluded volume are severe constraints. In this respect, investigations using self-contained polymer models such as that conducted here are fundamentally more informative than those that use postulated free energy profiles in the absence of explicit chain representations (e.g., see Ref. 56).

In summary, our results suggest that intramolecular recognition in real two-state proteins is highly specific. As well, the role of many-body interactions in providing a larger average energetic difference between "native" and "denatured" conformations than that afforded by common pairwise additive interaction schemes have potentially important implications for the discrimination of decoys in protein structure prediction.[57] In principle, the many-body interactions proposed in the present work can be characterized quantitatively through careful experiments and extensive atomic simulations. How sidechain packing, sidechain/mainchain correlation,[58] and interactions such as hydrogen bonding contribute to this mechanism remains to be investigated. To help address these questions, the ramifications of the different scenarios explored in this work need to be first delineated in greater detail.[33,34] Lattice model studies of Klimov and Thirumalai[54] showed that sidechain degrees of freedom increase the sharpness of the thermodynamic folding/unfolding transition relative to that of the corresponding (mainchain) model with no sidechains. However, their short-chain 20-letter sidechain models configured on three-dimensional simple cubic lattices do not appear to be calorimetrically cooperative. The $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ ratio of one of the sidechain sequences studied in Refs. 54 and 55 is equal to $\kappa_2 = 0.38$ without baseline subtraction and increases only to $\kappa_2^{(\mathrm{s})} = 0.54$ after reasonable baseline subtractions.[5] These values are far from the

---

[†]We emphasize that the relatively smooth funnel drawings in Figure 8 should be viewed only as pictorial devices for underscoring the smoothness of the energy landscapes of native-centric models relative to that of models with deeper kinetic traps. Even for Gō and Gō-like models, energy landscapes cannot be completely smooth because of repulsive interactions (including excluded volume effects) and other microscopic energy barriers due to bond rotations and solvation, for example (Refs. 8, 26, 27; c.f. Eq. 10).

$\Delta H_{vH}/\Delta H_{cal} \approx 1$ required for calorimetrically cooperative behavior. Kinetically, even a Gō-like version of their sidechain model exhibits a severe chevron rollover (Fig. 3 of Ref. 59). These results imply that although sidechain contributions are expected to be important for protein cooperativities,[52] their role has yet to be better elucidated.

Finally, although the present study focuses on the behavior of small single-domain proteins, we hasten to emphasize that not all proteins have simple two-state folding/unfolding kinetics. Hence, the high-cooperativity requirement deduced in the above analysis may not apply to other proteins. In fact, one distinct advantage of the energy landscape perspective and self-contained polymer modeling is their ability to cover a wide spectrum of possible protein behavior under a unified physical framework. For instance, although common additive Gō models are insufficient for simple two-state kinetics, they are useful for understanding real proteins with chevron rollovers.[7,8,46,60] And even calorimetrically non-cooperative models (see discussion in Ref. 5) may prove to be helpful in rationalizing downhill protein folding[61] as well.

ACKNOWLEDGMENTS

REFERENCES

1. Chan HS. Matching speed and locality. Nature 1998;392:761–763.
2. 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.
3. 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.
4. Chan HS. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 2000;40:543–571.
5. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661 [Erratum: Proteins 2001;43:523].
6. Kaya H, Chan HS. Energetic components of cooperative protein folding. Phys Rev Lett 2000;85:4823–4826.
7. 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.
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. Gō N. Theoretical studies of protein folding. Annu Rev Biophys Bioeng 1983;12:183–210.
10. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84:7524–7528.
11. Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions. Int J Pept Protein Res 1975;7:445–459.
12. 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.
13. Li L, Shakhnovich EI. Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. Proc Natl Acad Sci USA 2001;98:13014–13018.
14. Portman JJ, Takada S, Wolynes PG. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. J Chem Phys 2001;114:5082–5096.
15. Micheletti C, Lattanzi G, Maritan A. Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures. J Mol Biol 2002;321:909–921.
16. 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.
17. 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.
18. Matthews CR. Effect of point mutations on the folding of globular proteins. Methods Enzymol 1987;154:498–511.
19. Cheung MS, García AE, Onuchic JN. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc Natl Acad Sci USA 2002;99:685–690.
20. Dill KA. Dominant forces in protein folding. Biochemistry 1990;29:7133–7155.
21. Kolinski A, Galazka W, Skolnick J. On the origin of the cooperativity of protein folding: implications from model simulations. Proteins 1996;26:271–287.
22. 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.
23. Schindler T, Schmid FX. Thermodynamic properties of an extremely rapid protein folding reaction. Biochemistry 1996;35:16833–16842.
24. Scalley ML, Baker D. Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. Proc Natl Acad Sci USA 1997;94:10636–10640.
25. 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.
26. Chan HS. Modelling protein folding by Monte Carlo dynamics: chevron plots, chevron rollover, and non-Arrhenius kinetics. In: Grassberger P, Barkema GT, Nadler W, editors. Monte Carlo Approach to Biopolymers and Protein Folding. Singapore: World Scientific; 1998. p 29–44.
27. Chan HS, Dill KA. Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
28. Northey JGB, Di Nardo AA, Davidson AR. Hydrophobic core packing in the SH3 domain folding transition state. Nat Struct Biol 2002;9:126–130.
29. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L. Conformational strain in the hydrophobic core and its implications for protein folding and design. Nat Struct Biol 2002;9:485–493.
30. 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.
31. 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.
32. Miller EJ, Fischer KF, Marqusee S. Experimental evaluation of topological parameters determining protein-folding rates. Proc Natl Acad Sci USA 2002;99:10359–10363.
33. Kaya H, Chan HS. Contact order dependent protein folding rates: kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences. Proteins 2003;52:524–533.
34. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2003. Forthcoming.
35. Segawa S, Sugihara M. Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. Biopolymers 1984;23:2473–2488.
36. Chen B-L, Baase WA, Schellman JA. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. Biochemistry 1989;28:691–699.
37. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1.

Evidence for a two-state transition. Biochemistry 1991;30:10428–10435.

38. Kuhlman B, Luisi DL, Evans PA, Raleigh DP. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. J Mol Biol 1998;284:1661–1670.

39. Makhatadze GI, Privalov PL. Energetics of protein structure. Adv Protein Chem 1995;47:307–425.

40. Lee CY, McCammon JA, Rossky PJ. The structure of liquid water at an extended hydrophobic surface. J Chem Phys 1984;80:4448–4455.

41. Lum K, Chandler D, Weeks JD. Hydrophobicity at small and large length scales. J Phys Chem B 1999;103:4570–4577.

42. Shimizu S, Chan HS. Origins of protein denatured states compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. Proteins 2002;49:560–566.

43. Baldwin RL. Folding intermediates in protein folding. BioEssays 1994;16:207–210.

44. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–195.

45. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.

46. Kaya H, Chan HS. Origins of chevron rollovers in non-two-state protein folding kinetics. Phys Rev Lett 2003;90:258104.

47. 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].

48. Makarov DE, Plaxco KW. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Protein Sci 2003;12:17–26.

49. Crippen GM, Chhajer M. Lattice models of protein folding permitting disordered native states. J Chem Phys 2002;116:2261–2268.

50. Fan K, Wang J, Wang W. Folding of lattice protein chains with modified Gō potential. Eur Phys J B 2002;30:381–391.

51. Clementi C, Garcia AE, Onuchic JN. Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L. J Mol Biol 2003;326:933–954.

52. Pokarowski P, Kolinski A, Skolnick J. A minimal physically realistic protein-like lattice model: designing an energy landscape that ensures all-or-none folding to a unique native state. Biophys J 2003;84:1518–1526.

53. Englander SW, Mayne L, Bai Y, Sosnick TR. Hydrogen exchange: the modern legacy of Linderstrϕm-Lang. Protein Sci 1997;6:1101–1109.

54. Klimov DK, Thirumalai D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. Fold Des 1998;3:127–139.

55. Klimov DK, Thirumalai D. Is there a unique melting temperature for two-state proteins? J Comput Chem 2002;23:161–165.

56. Zwanzig R. Simple model of protein folding kinetics. Proc Natl Acad Sci USA 1995;92:9801–9804.

57. Park BH, Huang ES, Levitt M. Factors affecting the ability of energy functions to discriminate correct from incorrect folds. J Mol Biol 1997;266:831–846.

58. Dunbrack RL. Rotamer libraries in the 21st century. Curr Opin Struct Biol 2002;12:431–440.

59. Li MS, Klimov DK, Thirumalai D. Folding in lattice models with side chains. Comput Phys Comm 2002;147:625–628.

60. Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV. Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. Proteins 2003;51:162–166.

61. Garcia-Mira MM, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V. Experimental identification of downhill protein folding. Science 2002;298:2191–2195.