[16] Cooperativity Principles in Protein Folding

By HUE SUN CHAN, SEISHI SHIMIZU, and HÜSEYIN KAYA

Introduction

Knowledge of the physical driving forces in proteins is essential for understanding their structures and functions. As polymers, proteins have remarkable thermodynamic and kinetic properties. A well-known observation is that the folding and unfolding of many small single-domain proteins, of which chymotrypsin inhibitor 2 is a prime example, appear to involve only two main states—N (native) and D (denatured).[1,2] These proteins' folding/unfolding transitions are often referred to as ''cooperative'' because of their phenomenological similarity to ''all-or-none'' processes. Traditionally, only N, D, and a small number of postulated intermediate states were invoked to account for experimental protein folding data. Under such an interpretative framework, two-state folding is described by the reaction $N \rightleftharpoons D$, and different properties are ascribed to N and D to account for different proteins.

Although useful, this approach does not address the microscopic origins of experimentally observed two-state–like behavior. Traditional analyses simply assume that there are a small number of conformational states. But proteins are chain molecules. Physically, it is obvious that a polymer chain can adopt many conformations, ranging from the most open to maximally compact, and all intermediate compactness in between. Thus, whether and how the multitude of conformations available to a protein may be grouped into two or more ''states''—as traditionally assumed—should be ascertained through a fundamental understanding of the effective intrachain interactions involved. In the protein literature, however, folding energetics are often discussed in terms of the sum of contactlike energies of a fully folded native structure versus that of a random-coil–like state or a certain other prespecified unfolded conformational ensemble.[3] Such analyses have yielded important insight. But they obscure the remarkable nature of protein cooperativities. This is because cooperativity has already been presumed in these discourses by their preclusion of many a priori possible conformations—notably compact nonnative conformations—from the energetic equation. To gain a consistent understanding

[1] S. E. Jackson and A. R. Fersht, *Biochemistry* **30,** 10428 (1991).
[2] D. Baker, *Nature* **405,** 39 (2000).
[3] E. Freire and K. P. Murphy, *Adv. Protein Chem.* **43,** 313 (1992).

0076-6879/04 $35.00

of the physical origins of proteinlike cooperativities, it is only logical that one should seek to reproduce them in *self-contained polymer models* in which the distribution of explicit-chain conformations is determined solely by the interactions considered explicity in the model.[4] This is the approach we take. Our rationale is based on the discovery that proteinlike cooperativities are nontrivial to achieve in chain models (see below). Consequently, we expect the stringent requirements of proteinlike cooperativities to provide us with important clues to protein energetics in general.

These considerations led us to ask: What effective intrachain interactions can give rise to the remarkable cooperativity of many small proteins?[5] Are traditional pictures based on additive interactions sufficient to rationalize experimental observations? At a more basic level, what insight can be gained from detailed molecular accounts of the elementary driving forces of protein folding? Using hydrophobic interactions as a test case, our investigation indicates that driving forces for protein folding are intrinsically nonadditive. Significantly, extensive evaluations of coarse-grained chain models suggest that the high degrees of thermodynamic and kinetic cooperativity of small single-domain proteins may likely originate from many-body interactions in the form of a coupling between local conformational preferences and favorable nonlocal interactions. This chapter summarizes our recent advances in these respects.

Multiple Meanings of Cooperativity in Biomolecular Processes

We begin with more precise definitions of cooperative behavior. Generally speaking, ''cooperativity'' refers to a particular type of deviation from a reference situation deemed not cooperative. Thus, whether a physical interaction or process is cooperative hinges on the reference interaction or process to which it is compared. Often the hypothetical reference noncooperative situation is one in which some form of additivity is satisfied (Fig. 1). For example, ligand binding to a protein is additive (not cooperative) if the free energy of association per ligand is independent of the number of ligands already bound to the same protein.[6] In this case, cooperative binding means that the favorability of binding one ligand increases with the number of ligands already bound until the binding sites of an allosteric proteins are saturated, as for binding of oxgyen to hemoglobin.[7]

[4] H. S. Chan, H. Kaya, and S. Shimizu, *in* ''Current Topics in Computational Molecular Biology'' (T. Jiang, Y. Xu, and M. Q. Zhang, eds.), p. 403. MIT Press, Cambridge, MA, 2002.
[5] H. S. Chan, *Nature* **392,** 761 (1998).
[6] C. R. Cantor and P. R. Schimmel, ''Biophysical Chemistry,'' Part III. Freeman, New York, 1980.
[7] G. K. Ackers, J. M. Holt, and A. L. Klinger, *Biochemistry* **39,** 117 (2000).

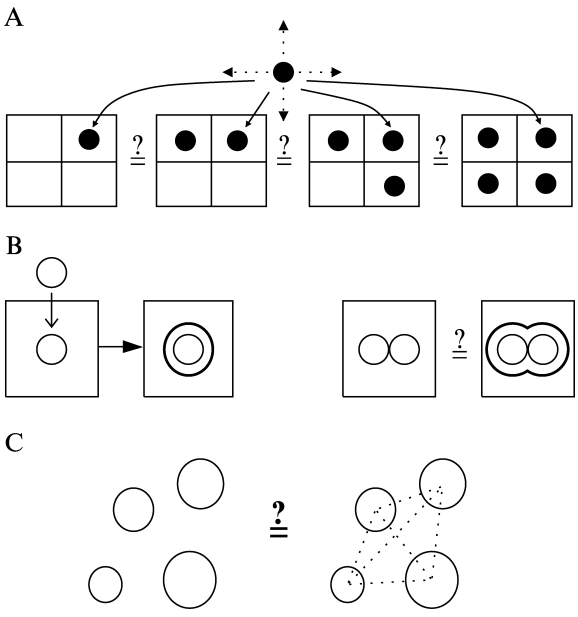

FIG. 1. "Cooperativity" and "anticooperativity" as deviations from additivity. Schematics of several questions of interest. (A) Multiple-site ligand binding: Is the intrinsic free energy of association per ligand dependent upon the number of bound sites? The dotted arrows here highlight the unbound ligand's translational freedom. (B) Implicit-solvent potentials and group additivity: Can solvent-accessible surface area (SASA) or other geometric parameterizations based on bulk-phase transfer data (left) accurately predict solvent-mediated interactions involving more than one solute (right)? (C) Many-body interactions and pairwise additivity: To what extent can interactions among three or more physical entities be accurately described as a pairwise sum of two-body interactions (dotted lines)?

On the other hand, binding is considered to be anticooperative if the favorability of binding one ligand decreases with the number of ligands already bound (Fig. 1A).

A different context of additivity is illustrated in Fig. 1B. Many empirical treatments of solvation employ experimental data from bulk-phase transfer, e.g., between aqueous and nonpolar phases. By itself, transfer data effectively characterize only the interactions between a single solute and its surrounding solvent molecules. Nonetheless, additivity assumptions are often invoked to apply transfer data to the energetics of more complex situations involving multiple-body interactions, including protein folding and protein–protein interactions. In these applications, free energy and other thermodynamic signatures are assumed to be proportional to geometric measures such as solvent-accessible surface area (SASA). Then,

many-body interactions are computed as the product of a given configuration's SASA with the energetic proportionality coefficient determined from (single-solute) transfer data. Using free energy of hydration ΔG and SASA as an example, this means that one first sets $\Delta G_1 = \gamma(\mathrm{SASA})_1$, where ΔG_1 is the hydration free energy of a single solute and $(\mathrm{SASA})_1$ is its SASA. Then the free energy of hydration ΔG_m of an interacting m-solute system is taken to be $\Delta G_m = \gamma(\mathrm{SASA})_m$, where $\gamma = \Delta G_1/(\mathrm{SASA})_1$ is determined from the single-solute experimental data and $(\mathrm{SASA})_m$ is the SASA of the m-solute system. This approach is often referred to as group additivity.[8] The idea is intuitive. But the question is, in view of the granularity or particulate nature of real solvents, how valid are such additivity assumptions (Fig. 1B)? In this context, multiple-solute interactions for a given solute configuration may be termed ''cooperative'' or ''anticooperative'' depending on whether the actual interactions are more or less favorable to the association of solutes than that predicted by SASA.

Figure 1C depicts yet another additivity condition. Now, instead of using the solvation interaction of a single solute (Fig. 1B, left) as a standard for additivity, the reference interaction is taken to be the solvent-mediated interactions between a pair of solutes (Fig. 1C). The question of cooperativity then is whether the solvent-mediated interaction free energy $W^{(m)}$ $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_m)$ among three or more solutes ($\mathbf{r}$s are the position vectors of the solutes) can be adequately described by the sum of independent two-solute interactions. In other words, whether $W^{(m)}$ $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_m) = \sum_{i<j}^{m} W^{(2)}(\mathbf{r}_i, \mathbf{r}_j)$ for $m \geq 3$? If so, the interaction is termed pairwise additive. If not, the interaction for a particular solute configuration is cooperative (or anticooperative) depending on whether the actual multiple-body interaction is more (or less) favorable to solute association than that predicted by the pairwise sum.[9] As an example, how hydrophobic interactions match up with the additivity conditions in Fig. 1B and C will be examined in the next section.

Besides binding and interaction nonaddivities (Fig. 1), the term ''cooperativity'' is also used extensively in biophysical chemistry to describe the sharpness of thermodynamic transitions (Fig. 2). One example is solute aggregation (Fig. 2A). Under appropriate conditions, the transformation of a solution with randomly dispersed solutes into a phase-separated mixture with large solute aggregates can occur over a very narrow range of solute mole fraction. A transition with such a sharp transformation from disorder to order caused only by a relatively small change in the physical parameters of the system is often characterized as ''cooperative.''[10] Another

[8] K. A. Dill, *J. Biol. Chem.* **272,** 701 (1997).
[9] S. Shimizu and H. S. Chan, *J. Chem. Phys.* **115,** 3424 (2001).

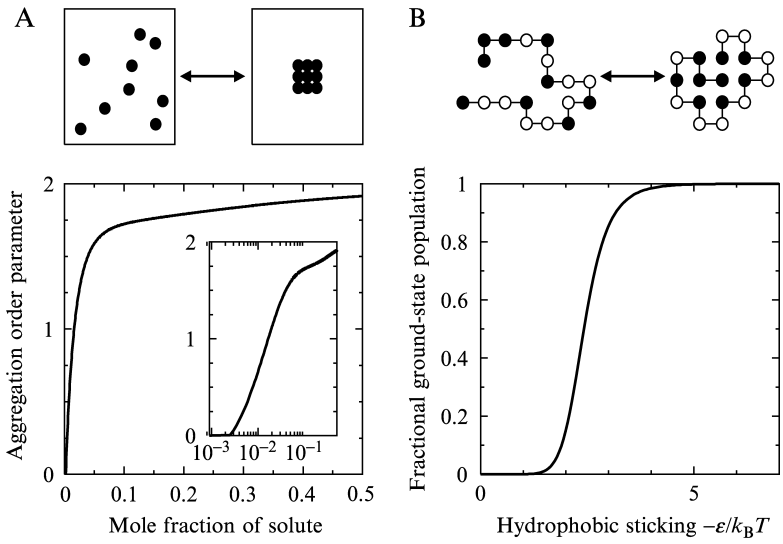

FIG. 2. "Cooperativity" as sharpness of configurational or conformational transitions. (A) Solvophobic aggregation: A model solute–solvent system undergoes a sharp transition from a dispersed solution state to an aggregated state when the solute mole fraction is increased beyond a certain threshold. Here the degree of aggregation is quantitated by an order parameter ranging from 0 to 2. The inset uses a logarithmic scale for solute mole fraction to underscore the sharpness of the transition. [Data from Fig. 1 of S. Shimizu and H. S. Chan, *J. Chem. Phys.* **115,** 3424 (2001), for the case with solute–solute contact interaction strength $\varepsilon/RT = -2.571$. See this reference for details.] (B) Folding transition: The conformational distribution of a two-dimensional HP model protein undergoes a sigmoidal transition as model hydrophobic strength increases, where k_BT is the Boltzmann constant times absolute temperature [cf. H. S. Chan and K. A. Dill, *J. Chem. Phys.* **100,** 9238 (1994)]. When hydrophobic sticking is weak, most of the chain population is distributed among the large number of non–ground-state conformations (one example of which is shown on the left). As hydrophobic sticking is strengthened, an increasing fraction of the chain population adopts the unique ground-state (native) conformation shown on the right.

example is chain collapse and folding of polymers (Fig. 2B), where a sigmoidal transition curve is often recognized as a hallmark of cooperativity.[11] For these thermodynamic transitions, "cooperativity" is associated with a certain degree of "all-or-none" character, a sudden nongradual change near the transition midpoint. Ultimately, thermodynamic transitions are dictated by the elementary interactions involved. Hence, interaction cooperativity or nonadditivity (Fig. 1) is expected to bear on transition cooperativity

[10] J. Tsai, M. Gerstein, and M. Levitt, *Protein Sci.* **6,** 2606 (1997).
[11] H. S. Chan, S. Bromberg, and K. A. Dill, *Phil. Trans. Roy. Soc. London B* **348,** 61 (1995).

(Fig. 2).[9,12] In this connection, it is noteworthy that interaction cooperativity is *not* necessary for sigmoidal (''cooperative'') transitions. This is immediately evident from Fig. 2, where the transitions are results of pairwise additive interactions. In these cases, even though the underlying contact energy is pairwise additive, the nongradual character of the transition originates geometrically from the fact that, on average, solutes or monomers well inside an aggregate or the core of a collapse conformation make more contacts per solute or per monomer than when the solutes or monomers are in more dispersed configurations or more open conformations. This naturally leads to a nonlinear increase in favorability of collapsed configurations once the favorability of the pairwise interactions reaches a threshold.[10] For protein folding, however, we found that a sigmoidal transition per se (such as that in Fig. 2B) is insufficient for the quantitatively more stringent experimental cooperativity criteria.[12–15] In this light, the relationship between interaction cooperativity and transition cooperativity needs to be better delineated. These issues will be analyzed below.

Nonadditivity Is Prevalent among Solvent-Mediated Interactions

Figure 3 highlights some of our recent atomic simulation efforts aimed at gaining a better grasp of interaction nonadditivity in protein folding. The results are obtained by extensive Monte Carlo sampling. Computational details are discussed elsewhere.[16–20] As a first step, interactions among methanes in water and other aqueous solvents are investigated as prototypical hydrophobic interactions. The interaction between a pair of methanes in water is distance dependent. Here salient features of the free energy of association (potential of mean force, PMF) include a favorable contact minimum at spatial separation $\xi \approx 3.8$ Å, an unfavorable desolvation barrier at $\xi \approx 5.7$ Å, and a slightly favorable solvent-separated minimum at $\xi \approx 7.0$ Å (Fig. 3A). In contrast, SASA predicts neither the desolvation barrier nor the solvent-separated minimum, and it greatly overestimates

[12] H. S. Chan, *Proteins Struct. Funct. Genet.* **40,** 543 (2000).
[13] H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet.* **40,** 637 (2000) [Erratum: **43,** 523 (2001)].
[14] H. Kaya and H. S. Chan, *Phys. Rev. Lett.* **85,** 4823 (2000).
[15] P. Pokarowski, A. Kolinski, and J. Skolnick, *Biophys. J.* **84,** 1518 (2003).
[16] S. Shimizu and H. S. Chan, *J. Chem. Phys.* **113,** 4683 (2000) [Erratum: **116,** 8638 (2002)].
[17] S. Shimizu and H. S. Chan, *J. Am. Chem. Soc.* **123,** 2083 (2001).
[18] S. Shimizu and H. S. Chan, *J. Chem. Phys.* **115,** 1414 (2001).
[19] S. Shimizu and H. S. Chan, *Proteins Struct. Funct. Genet.* **48,** 15 (2002) [Erratum: **49,** 294 (2002)].
[20] S. Shimizu and H. S. Chan, *Proteins Struct. Funct. Genet.* **49,** 560 (2002).

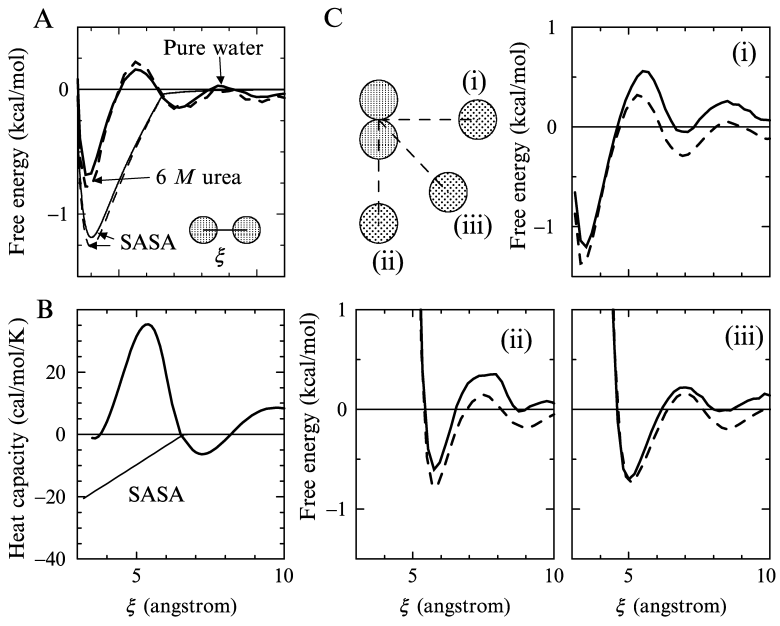

FIG. 3. Nonadditive effects of hydrophobic interactions are investigated by simulating solvent-mediated interactions among methanes in explicit water. Results shown are for 25° under atmospheric pressure using the TIP4P model of water. (A) The free energy of bringing a methane from infinity to a distance ξ from another methane in pure water (upper solid curve) and in 6 *M* aqueous urea (upper dashed curve) is compared with the SASA prediction based upon the corresponding solvation free energy of a single methane. (B) The heat capacity ΔC_P of two-methane association in pure water (upper curve) is compared with the SASA prediction based upon the hydration heat capacity of a single methane. [Data for (A) and (B) from S. Shimizu and H. S. Chan, *Proteins Struct. Funct. Genet.* **49,** 560 (2002).] (C) Free energies (solid curves) of bringing a methane from infinity along three directions (i, ii, and iii) to a distance ξ from the midpoint of a methane dimer are compared with the hypothetical free energies (dashed curves) that assume pairwise additivity [data from S. Shimizu and H. S. Chan, *Proteins Struct. Funct. Genet.* **48,** 15 (2002)].

the favorability of the contact minimum between two methanes in pure water and in 6 *M* urea. A desolvation barrier is predicted by some other geometric measures such as molecular surface area (not shown here). But their overall deviations from the actual two-body PMF remain significant.[19] Interestingly, Fig. 3A shows that urea does not destabilize the contact between two methanes.[20–22] This result suggests strongly that urea's effect

[21] A. Wallqvist, D. G. Covell, and D. Thirumalai, *J. Am. Chem. Soc.* **120,** 427 (1998).
[22] M. Ikeguchi, S. Nakamura, and K. Shimizu, *J. Am. Chem. Soc.* **123,** 677 (2001).

on hydrophobic interactions is sensitive to the sizes of hydrophobic solutes involved, because other simulations of ours indicate that urea destabilizes the contact between two neopentanes (not shown here). As we have pointed out,[20] this observation may provide a rationalization for why even with an increasing urea concentration, there is often little postdenaturational expansion of protein conformations. Such perspectives are not offered by traditional SASA approaches. Likewise, Fig. 3B shows that SASA does not account for the spatial dependence of two-methane heat capacity. SASA predicts that heat capacity of association should decrease monotonically (become more negative) as two methanes move closer, but the actual simulated heat capacity shows that heat capacity change is significantly positive in the desolvation barrier region and is approximately zero when the two methanes are in contact. The latter observation implies that the heat capacity signature of two methanes in contact is essentially the same as that of two methanes that are far apart. This may offer a reason for why the heat capacity signature of protein compact denatured states can be similar to that of more open unfolded states.[17,20] The results in Fig. 3A and B demonstrate clearly that for certain applications, the traditional group additivity assumption based on single-solute solvation (Fig. 1B) is not always adequate.

We have also studied how hydrophobic interactions might deviate from pairwise additivity (cf. Fig. 1C). Figure 3C compares the directly simulated three-methane PMFs in pure water (solid curves) with corresponding hypothetical PMFs (dashed curves) based upon the two-methane PMF in Fig. 3A and the assumption that many-methane interaction is pairwise additive. The results show that methanelike hydrophobic interactions are mostly anticooperative at 25° under atmospheric pressure. Association of three methanes is less favorable (has more positive or less negative free energy) than that predicted by assuming pairwise additivity.[19] The sign and extent of nonadditivity depend on the three-methane configurations. Our finding of anticooperativity at the contact minimum is consistent with a more recent study by Ghosh *et al.*[23] We have also deduced that the formation of an infinite methane cluster at 25° under atmospheric pressure is anticooperative, in that the cluster's stability is less than that predicted by assuming pairwise additivity.[18,19] The implications of these anticooperative aspects of hydrophobic interactions on the cooperativity of protein folding transition will be explored below.

[23] T. Ghosh, A. E. García, and S. Garde, *J. Phys. Chem. B* **107,** 612 (2003).

Thermodynamic Cooperativity of Protein Folding Transitions

The Calorimetric Criterion

We now turn to quantitative cooperativity criteria for the protein folding/unfolding transition (Fig. 4). A long-standing requirement[24,25] for protein thermodynamic cooperativity is that the van't Hoff to calorimetric enthalpy ratio $\Delta H_{vH}/\Delta H_{cal}$ for the transition is very close to unity. This enthalpy ratio is deduced from measurements of heat capacity as a function of temperature. An example of such experimental data is shown in Fig. 4A. In these considerations, the heat capacity of interest is the ''excess'' heat capacity, which is traditionally presumed to contain the thermodynamic information for the conformational transition alone. The excess heat capacity is obtained from raw calorimetric data by empirical baseline subtractions aimed at factoring out solvation effects related to the intrinsic difference in heat capacity of the native and denatured states. Recently, theoretical analyses[13,26] have raised questions as to whether all conformational contributions to protein heat capacity are properly accounted for by traditional methods of subtraction, or whether a small part of these contributions has been ''hidden'' below empirical baselines. Experimentally, this possibility has been suggested by nuclear magnetic resonance (NMR) measurements of heat capacity contributions from bond vector motions of staphylococcal nuclease and the N-terminal of the drk SH3 domain.[27] The scenario that traditional native baselines might have subtracted out conformational contributions to heat capacity is also consistent with native-state hydrogen exchange data[28] of cytochrome *c* and other proteins showing the presence of conformational transitions under strongly folding conditions. In our view, recent calorimetric data on a leucine zipper that fails to undergo a simple two-state transition[29] also lend support to this perspective.[13,26]

Thus, careful reevaluations of baseline subtractions are conceptually important, as their effects can be quantitatively significant. However, they are not critical for our present purposes. Hence they are not further pursued in this chapter. This is because for simple two-state proteins (which is the main focus here), our analysis indicates that baseline corrections of the type discussed above have to be relatively minor (see below).

[24] R. Lumry, R. Biltonen, and J. F. Brandts, *Biopolymers* **4,** 917 (1966).
[25] G. I. Makhatadze and P. L. Privalov, *Adv. Protein Chem.* **47,** 307 (1995).
[26] H. Kaya and H. S. Chan, *J. Mol. Biol.* **315,** 899 (2002).
[27] D. Yang, Y. K. Mok, J. D. Forman-Kay, N. A. Farrow, and L. E. Kay, *J. Mol. Biol.* **272,** 790 (1997).
[28] S. W. Englander, L. Mayne, Y. Bai, and T. R. Sosnick, *Protein Sci.* **6,** 1101 (1997).
[29] A. I. Dragan and P. L. Privalov, *J. Mol. Biol.* **321,** 891 (2002).

Therefore, for these proteins, we follow the traditional interpretation: calorimetric enthalpy ΔH_{cal} of the unfolding transition is taken to be the area under the excess C_P function obtained from empirical baseline substractions. The van't Hoff enthalpy ΔH_{vH} of the transition is deduced from the peak value $C_{P,\mathrm{max}}$ of the same function (Fig. 4A). In this regard, since virtually all of the heat capacities considered below correspond to excess quantities, references to ''excess'' will be dropped as the definition would be clear from the context.

The physical meaning of the calorimetric two-state criterion is elucidated in Fig. 4B. By definition, heat capacity $C_P = (\partial\langle H\rangle/\partial T)_P$ is the temperature derivative of the Boltzmann average of enthalpy H. It follows that ΔH_{cal}, which is defined to be the integral of C_P from a temperature T_0 well

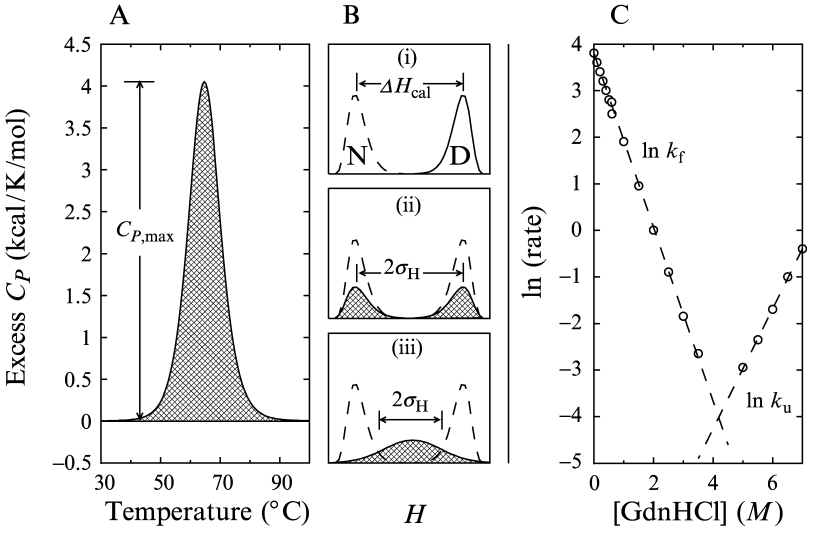

FIG. 4. Experimental criteria for protein thermodynamic and kinetic cooperativity. (A and C) Experimental measurements on the Ile → Val-76 mutant of chymotrypsin inhibitor 2 [data from S. E. Jackson, M. Moracci, N. elMasry, C. M. Johnson, and A. R. Fersht, *Biochemistry* **32,** 11259 (1993); S. E. Jackson, N. elMarsy, and A. R. Fersht, *Biochemistry* **32,** 11270 (1993)]. (A) Heat capacity scan after empirical baseline subtractions. (C) The chevron plot of logarithmic folding (k_{f}) and unfolding (k_{u}) rates versus denaturant concentration. The schematics in (B) compare two scenarios of enthalpy distribution. The vertical axes in i, ii, and iii represent chain population. In i the chain population is concentrated at low enthalpies under native (N) conditions (dashed curve) but shift to higher enthalpies under denaturing (D) conditions (solid curve). In ii and iii both of these reference distributions are sketched as dashed curves so as to compare them with the enthalpy distributions at the transition midpoint (shaded areas), and σ_{H} is the standard deviation of the midpoint distributions. The calorimetric two-state criterion is satisfied in ii but not in iii.

below the unfolding transition midpoint to a temperature T_1 well above, viz.,

$$\Delta H_{\mathrm{cal}} = \int_{T_0}^{T_1} dT C_P(T) = \int_{T_0}^{T_1} dT \frac{\partial \langle H(T) \rangle}{\partial T} = \langle H(T_1) \rangle - \langle H(T_0) \rangle = \langle H \rangle_{\mathrm{D}} - \langle H \rangle_{\mathrm{N}} \tag{1}$$

is equal to the difference between the average enthalpy of the denatured state $\langle H \rangle_{\mathrm{D}}$ and that of the native state $\langle H \rangle_{\mathrm{N}}$. This is schematically depicted in Fig. 4Bi, where the N and D peaks represent the enthalpy distributions at T_0 and T_1, respectively. For van't Hoff enthalpy, several different ΔH_{vH} expressions have been employed in the protein folding literature. Nonetheless, when the thermodynamic cooperativity of a protein is high, the commonly used definitions provide essentially the same restriction on enthalpy distribution.[13] Here we use

$$\Delta H_{\mathrm{vH}} = 2T_{\max} \sqrt{k_{\mathrm{B}} C_{P,\max}} \tag{2}$$

(in the κ_2 ratio[13]) as an example, where $T_{\max}$ is the temperature at which C_P attains its peak value $C_{P,\max}$. In general, heat capacity is proportional to the variance (square of standard deviation σ_{H}) of enthalpy:

$$C_P = \frac{\langle H^2(T) \rangle - \langle H(T) \rangle^2}{k_{\mathrm{B}} T^2} = \frac{\sigma_{\mathrm{H}}^2}{k_{\mathrm{B}} T^2} \tag{3}$$

Hence, Eqs. (2) and (3) imply that

$$\Delta H_{\mathrm{vH}} = 2(\sigma_{\mathrm{H}})_{\max} \tag{4}$$

i.e., the van't Hoff enthalpy is equal to two times the standard deviation of enthalpy at $T_{\max}$. Based on this consideration, Fig. 4Bii depicts a situation in which the calorimetric two-state criterion is satisfied. This is when $2(\sigma_{\mathrm{H}})_{\max}$ ($= \Delta H_{\mathrm{vH}}$) at $T_{\max}$ spans the entire shift in enthalpy ($= \Delta H_{\mathrm{cal}}$) that accompanies the unfolding transition. The only way for this to occur is to have the population distribution split into two halves at $T_{\max}$, one having essentially the same average enthalpy as N and the other as D. This does not mean that there is absolutely no conformation with enthalpy values in between, though the $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} \approx 1$ criterion does require their population to be very small. Conversely, if the protein has a significant conformational population with intermediate enthalpies at the transition midpoint, two times the standard deviation of enthalpy would fall short of spanning the entire range of enthalpy change caused by the transition (Fig. 4Biii). This would result in a violation of the calorimetric two-state criterion for thermodynamic cooperativity because $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} = 2(\sigma_{\mathrm{H}})_{\max}/\Delta H_{\mathrm{cal}}$ would then be considerably less than one.

Relationship with ''Minimal Frustration'' and ''Energy Gap'' Ideas

It follows from the above analysis that a protein's calorimetric two-state cooperativity implies that its conformational distribution is enthalpically well separated into two relatively sharply peaked populations—a low-enthalpy native population and a high-enthalpy denatured population, and very little conformational population with intermediate enthalpies in between.[12,13,24] We note that two of the determining factors of thermodynamic cooperativity, namely (1) the energetic separation between the native and denatured states and (2) the energetic variation among the denatured conformations, have also emerged as variables of fundamental importance in other theoretical considerations of protein folding. At this juncture, it is instructive to explore the relationship between various proposed theoretical criteria for proteinlike behavior with the experimental calorimetric two-state criterion.

One dimensionless parameter closely related to the above ingredients for thermodynamic cooperativity is the Z-score, which is widely used in protein structure prediction.[30] Z is defined as the width of energy variation among a given set of denatured structures (nonnative decoy conformations[31]) divided by the average energy separation between the native and the denatured structures. We found that if a simple random energy model of conformational distribution is assumed, and if the spread in denatured energies is relatively narrow, $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ is readily related to Z via the formula

$$\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} = \sqrt{1 - \frac{2 \ln g_{\mathrm{D}}}{Z^2}} \tag{5}$$

where g_{D} is the number of denatured conformations and $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ is the κ_0 ratio with the population-based van't Hoff enthalpy defined in Kaya and Chan.[13] This expression indicates clearly that $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ increases with increasing Z. It dictates that $Z \gg \sqrt{\ln g_{\mathrm{D}}}$ is needed for $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} \approx 1$ because in that case $1 - \Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} \approx \ln g_{\mathrm{D}}/Z^2$.

The $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} \approx 1$ condition is also intimately related to several foldability criteria for fast folding in recent energy landscape theories. One notable example is the ratio $T_{\mathrm{f}}/T_{\mathrm{g}}$ between the folding temperature T_{f} and the glass temperature T_{g}. The temperature T_{f} refers to the thermodynamic transition midpoint (the native state is more stable than the denatured state when $T < T_{\mathrm{f}}$), whereas T_{g} is the temperature under which folding kinetics is dominated by trapping mechanisms[32] (folding is ''glassy''

[30] J. U. Bowie, R. Luthy, and D. Eisenberg, *Science* **253,** 164 (1991).
[31] B. Park and M. Levitt, *J. Mol. Biol.* **258,** 367 (1996).

and sluggish for $T < T_g$). Glassy dynamics is related to the presence of ''frustrated'' intrachain interactions that are inconsistent with the native structure. Often these interactions are described pictorically in terms of a rugged energy landscape. Because natural small proteins can fold relatively fast to their native structures and in many cases folding kinetics is not glassy even under strongly native conditions, it has been stipulated that their intrachain interactions are evolutionarily designed to satisfy a principle of minimal frustration,[33] resulting in $T_f > T_g$. In this perspective, faster folding is associated with a higher T_f/T_g ratio, and the T_f/T_g criterion is taken as a quantification of the minimal frustration principle to achieve fast stable folding.[32]

In applying the T_f/T_g criterion, it is important to recognize that these temperatures are model parameters defined with temperature-independent interactions. In these models, native stability always increases with decreasing temperature. However, the effective intrachain interactions in real proteins are temperature dependent.[25] Thus, instead of identifying T_f and T_g with experimental temperatures,[34] physically it is more appropriate to interpret these temperatures as reciprocals of interaction strength (see below), and the $T_f > T_g$ condition as meaning that the onset of glassy dynamics occurs when native stability is higher than that at the transition midpoint.

A simple relation between T_f/T_g and $\Delta H_{vH}/\Delta H_{cal}$ can readily be established[12,13] in the random energy model (REM):

$$\Delta H_{vH}/\Delta H_{cal} = \sqrt{1 - 4\left(T_g/T_f\right)^2} \tag{6}$$

where we have used the REM definition of T_f/T_g of Onuchic *et al.*[32] [The quantity ln g_D in Eq. (5) is equivalent to the S_0/k_B expression in Eq. (12) of Onuchic *et al.*[32]] As in Eq. (5), the model interaction energies here are taken to be temperature independent, the population-based ΔH_{vH} is used, and the variation in energies among the denatured conformations is assumed to be relatively narrow. It is apparent from Eq. (6) that $\Delta H_{vH}/\Delta H_{cal}$ increases with increasing T_f/T_g.

Another parameter of interest is the energy gap. Fast folding of heteropolymer models has been linked to a large gap between the native energy and the energy of certain denatured conformations.[35] Despite problematic

[32] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, *Annu. Rev. Phys. Chem.* **48,** 545 (1997).
[33] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84,** 7524 (1987).
[34] B. Gillespie and K. W. Plaxco, *Proc. Natl. Acad. Sci. USA* **97,** 12014 (2000).
[35] A. Šali, E. Shakhnovich, and M. Karplus, *Nature* **369,** 248 (1994).

features of an early lattice formulation of this criterion,[32,36,37] the consideration in Fig. 4B indicates that a larger energy gap, or more appropriately a larger stability gap[38] between the native and denatured states, is generally conducive to a higher degree of thermodynamic cooperativity.

Thus, in summary, the experimental calorimetric requirement for two-state cooperativity is closely associated with theoretical criteria such as large Z-scores and large T_f/T_g ratios, and is seen to be qualitatively consistent with a larger energy gap. These observations imply that these theoretical criteria for a chain model's good folding behavior are fundamentally connected to the experimental observed thermodynamic cooperativity of real proteins. This is reassuring. Furthermore, the quantitative relationships above provide novel avenues to explore what values these theoretical parameters have to take to reproduce thermodynamic properties similar to that of real proteins. For instance, $T_f/T_g \approx 1.6$ has been proposed for a helical protein with approximately 60 residues.[39] But Eq. (6) suggests that T_f/T_g should be much higher for real proteins. Indeed, according to this formula,[13] $T_f/T_g = 6.4$ is required for $\Delta H_{vH}/\Delta H_{cal} = 0.95$. This issue will be further addressed below.

Kinetic Cooperativity of Protein Folding and Unfolding

Additional criteria for cooperativity are provided by the folding and unfolding kinetics of small single-domain proteins. A prominent characteristic of these proteins is that the data obtained using traditional optical probes on their reversible folding and unfolding are well described by the simple two-state reaction

$$\mathrm{N} \underset{k_f}{\overset{k_u}{\rightleftharpoons}} \mathrm{D}$$

where the logarithm of folding rate k_f and unfolding rate k_u are both essentially linear in denaturant concentration (Fig. 4C). In other words, their chevron plots[40] have linear folding and unfolding arms. Moreover, the two-state kinetics of these proteins are consistent with their thermodynamics. Their thermodynamically determined free energy of unfolding ΔG_u is an essentially linear function of denaturant concentration, and is well approximated by the kinetic quantity $k_B T \ln(k_f/k_u)$ obtained from comparing directly measured and extrapolated parts of the linear folding

[36] H. S. Chan, *Nature* **373,** 664 (1995).
[37] D. K. Klimov and D. Thirumalai, *Proteins Struct. Funct. Genet.* **26,** 411 (1996).
[38] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, *Proc. Natl. Acad. Sci. USA* **92,** 3626 (1995).
[39] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, *Science* **267,** 1619 (1995).
[40] C. R. Matthews, *Methods Enzymol.* **154,** 498 (1987).

and unfolding arms of the chevron plot (dashed lines in Fig. 4C). These properties constitute a more stringent requirement for cooperative behavior than thermodynamic cooperativity alone because they are not shared by all calorimetrically two-state proteins. For protein folding, it appears that thermodynamic cooperativity is necessary but not sufficient for kinetic cooperativity. Proteins that are moderately larger than the simple two-state variety often exhibit chevron rollovers, as for barnase[41] and ribonuclease A,[42] even though they are calorimetrically two-state.[25]

Linear Chevron Plots Require a High Degree of Thermodynamic Cooperativity

What types of intrachain interactions might be behind the remarkable cooperativities of the protein folding/unfolding transition? To explore this question, three representative heteropolymer models are compared in Figs. 5 and 6. In Fig. 5, model i is the 55-mer cooperative model of Kaya and Chan,[43] containing a physically motivated favorable coupling between helix formation and the packing of the native core as well as an extra stabilizing energy for the ground state. Model ii is the 48-mer Gō model of Pande and Rokhsar[44] with pairwise additive interaction.[44,45] Model iii is the 27-mer three-letter model of Socci *et al*.[32,46] Models i and ii here are ''native-centric'' in that their interactions are highly specific. Only interactions present in the native ground-state conformation are favored in these models. Model iii is based on general contact interactions of a three-letter alphabet such that nonnative contact interactions can be favored.

Figure 5A indicates that these models have very different thermodynamic cooperativities. The transition is sharp for models i and ii, but relatively broad for model iii. Quantitatively, the $\Delta H_{vH}/\Delta H_{cal}$ values (κ_2 ratio without empirical baseline subtractions[13]) for the three models are,[13,43] respectively, (i) 0.91, (ii) 0.87, and (iii) 0.46. We should point out here that the heat capacities in Fig. 5A were obtained using only temperature-independent interactions. But the general conclusion that additive nonspecific hydrophobic interactions (as modeled by chain sequences with small alphabets, cf. model iii) are insufficient for calorimetric cooperativity[12] is more broadly supported, including lattice model studies using

[41] A. Matouschek, J. T. Kellis, L. Serrano, M. Bycroft, and A. R. Fersht, *Nature* **346,** 440 (1990).
[42] W. A. Houry, D. M. Rothwarf, and H. A. Scheraga, *Nature Struct. Biol*. **2,** 495 (1995).
[43] H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet*. **52,** 510 (2003).
[44] V. S. Pande and D. S. Rokhsar, *Proc. Natl. Acad. Sci. USA* **96,** 1273 (1999).
[45] H. Kaya and H. S. Chan, *Phys. Rev. Lett*. **90,** 258104 (2003).
[46] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys*. **104,** 5860 (1996).

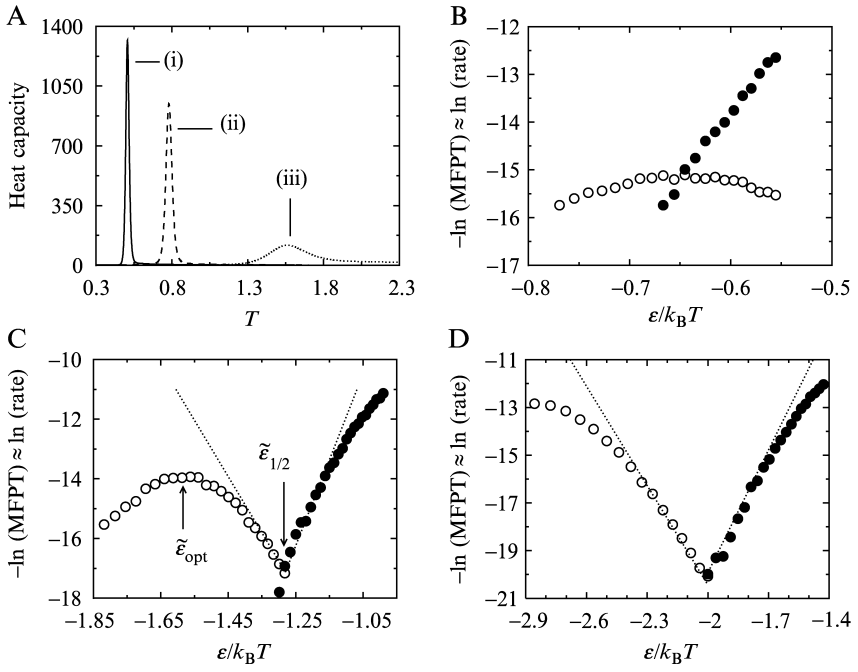

FIG. 5. Comparing the thermodynamic and kinetic cooperativity of three models. (A) Heat capacity scans for the (i) 55-mer cooperative, (ii) 48-mer Gō, and (iii) 27-mer three-letter models are obtained by standard Monte Carlo histogram techniques. The model interaction energies are taken to be temperature independent, with $\varepsilon = -1$. (B–D) Model chevron plots as negative logarithmic mean first passage time (MFPT) of folding (open circles) and unfolding (filled circles) versus intrachain interaction energy. (C) and (D) are for the 48-mer Gō (ii) and 55-mer cooperative (i) models. The dotted V shapes here depict hypothetical two-state chevron plots. (C) The $\tilde{\varepsilon}$ variables used in the text are indicated. (B) The 27-mer three-letter model (i). Each data point in (B) is averaged from 500 trajectories. For (B), folding starts from a randomly generated conformation and first passage is defined by achieving the ground-state conformation, whereas unfolding starts from the ground-state conformation and first passage is achieved when the chain has fewer than four native contacts. End flips are attempted for the monomers at the two chain ends of the 27-mer model, while corner flips (70%) and crankshafts (30%) are used to simulate the motion of other monomers. Model time in this figure and Fig. 7 is measured in units of attempted Monte Carlo moves. Part of the data shown in (A), (C), and (D) are from H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet*. **40,** 637 (2000); **52,** 510 (2003) as well as H. Kaya and H. S. Chan, *Phys. Rev. Lett.* **90,** 258104 (2003).

physically more realistic temperature-dependent effective interactions[12,19] (results not shown here). Nonetheless, because effective intrachain interactions in real proteins are temperature dependent, it is problematic to directly identify the *T*-dependent kinetics of a model having only temperature-independent interactions with the temperature effects on real protein

kinetics.[47,48] Instead, for kinetic considerations, it is often more appropriate to view the variation in interaction strength $\tilde{\varepsilon} \equiv \varepsilon/k_{\mathrm{B}}T$ in these models as corresponding to the variation of denaturant concentration at constant temperature.[26,47–49]

With this in mind, chevron plots of real proteins are modeled by the logarithmic folding and unfolding rates as functions of $\tilde{\varepsilon}$. Model chevron plots in Fig. 5 thus constructed show that thermodynamic cooperativity has a direct impact on kinetic cooperativity. Among the models shown, model iii with a three-letter alphabet is thermodynamically least cooperative. Concomitantly, its folding/unfolding kinetics (Fig. 5B) deviates most seriously from simple two-state behavior. As thermodynamic cooperativity increases for the native-centric models ii and i whose $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}}$ values are much higher, the corresponding chevron plots develop larger regions of quasilinear behavior resembling that of simple two-state proteins (Fig. 5C and D). The dotted V shapes in these plots depict hypothetical simple two-state linear chevron plots. In other words, the rates given by the dotted V shapes are consistent with a two-state account of the thermodynamic free energy of unfolding ΔG_{u}, as discussed above. The comparison in Fig. 5 indicates that for a chevron plot to possess a significant linear regime consistent with two-state thermodynamics, i.e., have a substantial $\tilde{\varepsilon}$ region of agreement between the simulated rates and the dotted V shape as in Fig. 5D, a high thermodynamic cooperativity with $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} > 0.9$ is most likely required.

Generic Statistical Mechanical Properties as Stringent Experimental Constraints on Possible Protein Energetics

It is clear from Fig. 5 that protein models with different interaction schemes can lead to very different predictions with respect to thermodynamic and kinetic cooperativities. This has an important ramification. It means that generic experimental statistical mechanical properties of simple two-state protein folding such as calorimetric cooperativity and linear chevron plots are useful for deciphering protein energetics. Because not all chain models can produce simple two-state behavior, important insights into how real proteins work may be gained by ascertaining what model intrachain interaction schemes can lead to experimental cooperativity properties, and what interaction schemes are deficient in those regards.

[47] H. S. Chan, *in* "Monte Carlo Approach to Biopolymers and Protein Folding" (P. Grassberger, G. T. Barkema, and W. Nadler, eds.), p. 29. World Scientific, Singapore, 1998.
[48] H. S. Chan and K. A. Dill, *Proteins Struct. Funct. Genet.* **30,** 2 (1998).
[49] H. Kaya and H. S. Chan, *J. Mol. Biol.* **326,** 911 (2003).

In evaluating a protein model's kinetic cooperativity against experiment and to characterize chevron rollovers or lack thereof, we found it useful to pay special attention to the behavior of the model at the transition midpoint (interaction strength $\tilde{\varepsilon}_{1/2}$) as well as at the interaction strength $\tilde{\varepsilon}_{opt}$ when the folding rate is fastest or optimal (cf. Fig. 5C). Simple two-state behavior requires $\tilde{\varepsilon}_{opt}$ to be significantly more negative (more favorable to the native state) than $\tilde{\varepsilon}_{1/2}$.

Interaction Specificity Enhances Cooperativity

Figure 5 shows that protein models with more specific interactions are more cooperative. The model with least specific interactions in Fig. 5 is the three-letter model. Figure 5B shows that the maximum folding rate of this model occurs at an interaction strength essentially identical to that of the transition midpoint ($\tilde{\varepsilon}_{opt} \approx \tilde{\varepsilon}_{1/2}$). This observation, which is consistent with earlier simulations of Onuchic *et al.*,[32] means that the folding rate to the left of the transition midpoint of this model decreases with increasing native stability. This trend is opposite to that of real small single-domain proteins. In Fig. 5C, intrachain interactions for the additive 48-mer Gō model[13,44,45] are more specific, as the energy function is constructed with explicit biases for the native structure. The resulting chevron plot is more proteinlike, in that the folding rate around the transition midpoint increases with increasing native stability ($\tilde{\varepsilon}_{opt} < \tilde{\varepsilon}_{1/2}$). But the severe chevron rollover exhibited by this model implies that its folding/unfolding kinetics still differ significantly from those of simple two-state proteins. In contrast, the 55-mer cooperative model[43] in Fig. 5D provides a better agreement with experimental simple two-state behavior. Intrachain interactions in this model are more specific than Gō models with only pairwise additive contact energies. The present 55-mer cooperative model energy function favors relatively large fragments of the native structure as a whole, resulting in a wider average energetic separation between the native and denatured conformations (cf. Fig. 6).

The 55-mer cooperative model incorporates many-body interactions embodying two main ideas. (1) The first is a cooperative interplay between local conformational preferences and favorable nonlocal interactions. This is motivated by experimental observations that secondary structure elements are not stable in isolation but are stable when packed against other parts of a protein in the folded core.[50] (2) The second is an extra favorable energy for the native conformation as a whole. This is motivated by experimental mutagenesis data[51] suggesting that driving forces for protein

[50] K. A. Dill, *Biochemistry* **29,** 7133 (1990).
[51] J. G. B. Northey, A. A. Di Nardo, and A. R. Davidson, *Nature Struct. Biol.* **9,** 126 (2002).

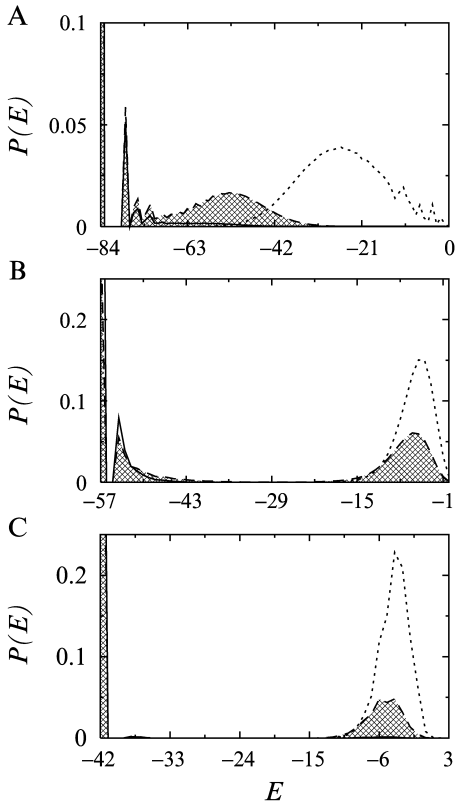

FIG. 6. Energy distributions and their implications for thermodynamic cooperativity (cf. Fig. 4B). $P(E)$ is the fraction of conformations with $E - 0.5 <$ energy $\leq E + 0.5$. The 27-mer three-letter (A), 48-mer Gō (B), and 55-mer cooperative (C) models in Fig. 5 are compared. Results are obtained by standard Monte Carlo histogram techniques. Distributions under strongly folding and strongly unfolding conditions are shown by solid and dotted curves, respectively. Distributions near the models' thermodynamic transition midpoints are shown by dashed curves with the areas underneath shaded. The $-\varepsilon/k_BT$ values used for strongly folding, transition, and strongly unfolding conditions for the three models are, respectively, (A) 0.77, 0.67, 0.33; (B) 1.47, 1.28, 1.14; and (C) 2.11, 2.0, 1.67.

folding kinetics are partially separated from the specific interactions that stabilize the native structure. Details of implementation of these ideas in a lattice model context are provided by Kaya and Chan.[43] These features of the 55-mer cooperative model lead to a chevron plot with an extensive linear regime ($\tilde{\varepsilon}_{opt}$ significantly more negative than $\tilde{\varepsilon}_{1/2}$). For this prototypical model with many-body interactions, the folding arm of the chevron plot is approximately linear for $\varepsilon/k_BT > -2.4$, corresponding to $\Delta G_u \leq$

$10k_BT$. Therefore, if $10k_BT$ is the maximum native stability that can be physically attained, the folding arm of this chevron plot would be linear for the entire physical regime. Because $\Delta G_u \approx 10k_BT$ is comparable to the native stabilities of many small single-domain proteins at zero denaturant at room temperature (e.g., $\Delta G_u \approx 9.0k_BT$ for protein L at 22°), Fig. 5D indicates how simple two-state folding/unfolding kinetics might arise from many-body effects similar to those postulated in this model and by virtue of the limited native stability of a small protein.[43]

Many-Body Interactions Needed for Linear Chevron Behavior

The existence of a maximum folding rate at a certain $\tilde{\varepsilon}_{opt}$ is a robust feature observed across many chain models. Physically, it has long been recognized that $\tilde{\varepsilon}_{opt}$ represents a balance between two opposing kinetic effects of making intrachain interactions thermodynamically more favorable to folding ($\tilde{\varepsilon}$ more negative).[52] On the one hand, more favorable intrachain interactions imply a stronger bias toward the native structure and thus faster folding. On the other hand, if intrachain interactions are too favorable, they would lead to deep kinetic traps, glassy dynamics, and slow folding.[32,53] Hence folding must be fastest or optimal at a certain intermediate interaction strength.[52] This perspective implies that chevron rollover is generally unavoidable. Nonetheless, as demonstrated by Fig. 5D, when thermodynamic cooperativity is enhanced by many-body interactions, kinetic trapping is reduced and $\tilde{\varepsilon}_{opt}$ can be pushed to more negative values relative to $\tilde{\varepsilon}_{1/2}$. In that event, folding-arm chevron rollover can be practically eliminated if the hypothetical native stability at the theoretical $\tilde{\varepsilon}_{opt}$ is much higher than the maximum native stability achievable in the given system.

Table I provides a broader comparison to further underscore the relationship between kinetic cooperativity and interaction specificity. In addition to the three chain models in Fig. 5, this table considers also a designed 20-letter 48-mer sequence[54] and a designed 2-letter 27-mer sequence (2LCa in ref. 32). Table I shows that for the 2-, 3-, and 20-letter models with finite alphabets, kinetic cooperativity increases (as characterized by larger $\tilde{\varepsilon}_{opt}/\tilde{\varepsilon}_{1/2}$ ratios) with increasing number of letters in the alphabet. This is consistent with the above observation that kinetic cooperativity increases with increasing interaction specificity. Sequences designed using

[52] R. Miller, C. A. Danko, M. J. Fasolka, A. C. Balazs, H. S. Chan, and K. A. Dill, *J. Chem. Phys.* **96,** 768 (1992).
[53] D. Thirumalai and S. A. Woodson, *Acc. Chem. Res.* **29,** 433 (1996).
[54] A. Gutin, A. Sali, V. Abkevich, M. Karplus, and E. I. Shakhnovich, *J. Chem. Phys.* **108,** 6466 (1998).

TABLE I
CHARACTERIZATION OF KINETIC COOPERATIVITY OF MODEL PROTEIN FOLDING AND UNFOLDING BY COMPARING PROPERTIES AT $\tilde{\varepsilon}_{opt}$ (WHEN FOLDING RATE IS MAXIMUM) TO THAT AT THE TRANSITION MIDPOINT $\tilde{\varepsilon}_{1/2}$

Protein chain model	$\tilde{\varepsilon}_{opt}/\tilde{\varepsilon}_{1/2}$	$\log_{10}[(k_f)_{opt}/(k_f)_{1/2}]^a$	$(\Delta G_u)_{opt}/k_BT^b$
55-mer cooperative[c]	1.48	3.2	30
48-mer Gō[d]	1.24	1.4	14
20-letter[e]	1.06	0.0	0.9
3-letter[f]	0.99	0.0	0.0
2-letter[f]	0.84	0.3	−3.0

[a] $(k_f)_{opt}$ and $(k_f)_{1/2}$ are the folding rates at interaction strengths $\tilde{\varepsilon}_{opt}$ and $\tilde{\varepsilon}_{1/2}$, respectively.
[b] $(\Delta G_u)_{opt}$ is the free energy of unfolding at interaction strength $\tilde{\varepsilon}_{opt}$.
[c] H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet.* **52,** 510 (2003).
[d] V. S. Pande and D. S. Rokhsar, *Proc. Natl. Acad. Sci. USA* **96,** 1273 (1999).
[e] A. Gutin, A. Sali, V. Abkevich, M. Karplus, and E. I. Shakhnovich, *J. Chem. Phys*. **108,** 6466 (1998).
[f] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, *Annu. Rev. Phys. Chem*. **48,** 545 (1997).

a larger alphabet tend to be more specific because they can exploit the higher degree of energetic heterogeneity afforded by the larger number of interaction types.[4,55–57]

Quantitative Characterizations of Chevron Plots

Two novel parameters are introduced in Table I to better characterize kinetic cooperativity: $\log_{10}[(k_f)_{opt}/(k_f)_{1/2}]$ compares the folding rates at $\tilde{\varepsilon}_{opt}$ with that at $\tilde{\varepsilon}_{1/2}$, and $(\Delta G_u)_{opt}$ is the free energy of unfolding at $\tilde{\varepsilon}_{opt}$. (By definition $\Delta G_u = 0$ at $\tilde{\varepsilon}_{1/2}$.) These parameters serve to facilitate more direct comparisons with experiments by eliminating references to the model interaction strength $\tilde{\varepsilon}$. The discussion above implies that kinetic cooperativity is associated with a large $(\Delta G_u)_{opt}$. The maximum folding rate generally occurs at a hypothetical native stability much higher than that covered by the linear regime of the model chevron plot. Therefore, for a model to behave like small single-domain proteins, $(\Delta G_u)_{opt}$ has to be significantly larger than typical zero-denaturant stabilities ($\approx 10k_BT$) of these proteins. Among the models listed in Table I, only the 55-mer cooperative model satisfies this requirement. For the same reason, $\log_{10}[(k_f)_{opt}/(k_f)_{1/2}]$ has to

[55] P. G. Wolynes, *Nature Struct. Biol*. **4,** 871 (1997).
[56] H. S. Chan, *Nature Struct. Biol*. **6,** 994 (1999).
[57] H. S. Chan and E. Bornberg-Bauer, *Appl. Bioinform*. **1,** 121 (2002).

be significantly larger than the common logarithm of the ratio between the zero-denaturant folding rate $(k_f)_0$ and transition midpoint folding rate $(k_f)_{1/2}$. The quantity $\log_{10}[(k_f)_0/(k_f)_{1/2}]$ is generally not small for real, small single-domain proteins, as their folding rates often span several orders of magnitude under different denaturant conditions. For example, $\log_{10}[(k_f)_0/(k_f)_{1/2}] \approx 3.1$ at 25° for wild-type chymotrypsin inhibitor 2.[1]

In view of these considerations, Table I shows that 2-, 3-, and 20-letter models are not kinetically cooperative in that their $\log_{10}[(k_f)_{opt}/(k_f)_{1/2}]$ and $(\Delta G_u)_{opt}$ are all very close to zero. This suggests strongly that model sequences designed using solely pairwise additive contact energies in a finite alphabet with ≤ 20 letters are not kinetically cooperative in general. Among them, the two-letter 27-mer sequence represents an extreme noncooperative case in which $\tilde{\varepsilon}_{opt}$ is less negative than $\tilde{\varepsilon}_{1/2}$ such that $\tilde{\varepsilon}_{opt}/\tilde{\varepsilon}_{1/2} < 1$ and fastest folding occurs under denaturing $[(\Delta G_u)_{opt} < 0]$ rather than native conditions. It should be emphasized here that the relatively high kinetic cooperativity of the 55-mer cooperative model in Table I is a direct consequence of its many-body interactions; it does not arise from its longer chain length per se. Much shorter chain models with many-body interactions can also achieve similar kinetic cooperativities. A case in point is a 27-mer (with relative contact order CO = 0.27) we have considered that has an extra favorable energy for the ground-state conformation as a whole.[43] We found that its $\tilde{\varepsilon}_{opt}/\tilde{\varepsilon}_{1/2} = 1.9$, $\log_{10}[(k_f)_{opt}/(k_f)_{1/2}] = 3.4$, and $(\Delta G_u)_{opt} = 35$, indicating that this particular 27-mer model has thermodynamic and kinetic cooperativities similar to that of the present 55-mer cooperative model.

Relationship between Thermodynamic and Kinetic Cooperativities

Figure 6 shows the distributions of conformational population that underlie the differences in thermodynamic cooperativity among the three models in Fig. 5. The three-letter model in Fig. 6A is calorimetrically noncooperative because the broad peak of its transition midpoint energy distribution ($E \approx -52$) lies approximately midway between that of the fully folded ($E = -84$) and fully unfolded ($E \approx -27$) states, corresponding to the noncooperative scenario in Fig. 4Biii. The significant conformational population with intermediate energy here is associated with substantial kinetic trapping, which is the root cause of the severe chevron rollover in Fig. 5B. As we pointed out, the continuous T-dependent shift of the energy distribution peak of this model contributes to a long high-temperature tail in the heat capacity function (Fig. 5Aiii) and implies a significant postdenaturational expansion of conformational dimension (as measured by root-mean-square radius of gyration, for example). But such an expansion is

not observed in small-angle X-ray scattering experiments on several small single-domain proteins.[13] This provides additional evidence that the thermodynamic behavior of this three-letter model is very different from these proteins. Therefore, in conjunction with the above consideration suggesting that $T_f/T_g > 6$ for simple two-state proteins—much higher than the $T_f/T_g = 1.6$ for the three-letter model, these observations cast doubt[13] on the proposal that ''real proteins resemble bead models in which only three kinds of residues are used to encode sequence.''[39] In this connection, our findings also argue against the hypothesis that the folding behavior of this particular three-letter sequence can be mapped onto that of a 60-residue helical protein through a ''law of corresponding states.''[39]

Figure 6B shows that the energy distribution of the 48-mer Gō model is more in line with the cooperative scenario in Fig. 4Bii. But even in this case, pairwise additive native-centric contact energies are insufficient for simple two-state kinetics (cf. Fig. 5C and Table I). This is because of the existence of a small yet not negligible non–ground-state population with near–ground-state energies[13] (small peak on the left of Fig. 6B). Some of these conformations would act as kinetic traps to slow folding and thus cause the folding chevron arm to roll over under conditions that are only mildly favorable to the native state.[45] In contrast, for the 55-mer cooperative model in Fig. 6C that has an extended regime of linear chevron behavior (Fig. 5D), the corresponding near–ground-state population is very much reduced. So, simple two-state kinetics appears to require a well-separated bimodal energy distribution quantitatively similar to that afforded by the cooperative model in Fig. 6C.

Additive Gō-Like Constructs Are Insufficient for Simple Two-State Kinetics: Implications for the Principle of Minimal Frustration

The 48-mer Gō model example in Figs. 5 and 6 and Table I indicates that additive interaction schemes envisioned by the common Gō potential cannot account for proteinlike simple two-state kinetics. This is a general deficiency, not an artifact of the lattice approach. Indeed, we have recently demonstrated that continuum (off-lattice) Gō-like models with essentially additive energies are also unable to produce simple two-state kinetics. As in Fig. 5C, the chevron plots of several continuum Gō models are seen to contain significant chevron rollovers, indicating that internal friction arising from kinetic trapping remains substantial in such constructs.[49] It is noteworthy that $T_f/T_g > 2$ has been reported for a 27-mer lattice Gō model.[58] Therefore, the failure of common Gō models to predict linear

[58] H. Nymeyer, N. D. Socci, and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **97,** 634 (2000).

chevron plots supports our contention above that the parameter T_f/T_g has to be significantly larger than 2 for a model protein to exhibit apparent simple two-state behavior.

This limitation of additive Gō models is basic. As such, it has far-reaching implications for the fundamental principles of protein folding energetics. Natural proteins are evolved molecules. What makes them different from other heteropolymers? A consistency principle was proposed by Gō 20 years ago.[59] It stipulates that different energetic components (e.g., local and nonlocal interactions) in naturally occurring proteins are evolutionarily designed to be consistent with one another when the protein adopts the native conformation. In other words, the native state is essentially free of energetic stress. This serves to ensure the stability of the native structure. A similar hypothesis was subsequentely offered by the principle of minimal frustration of Bryngelson and Wolynes[33] (see above) that recognizes energetic frustration is minimized but not nonexistent in real proteins,[60] i.e., ''consistency is not perfect.''[61]

As discussed above, we found that the minimal frustration principle is intimately related to the calorimetric criterion for thermodynamic cooperativity. The minimal frustration principle is extremely insightful in providing a first constraint on how real protein energetics should look. However, minimal frustration per se does not ensure proteinlike thermodynamic and kinetic cooperativities. In applications of the consistency principle and the principle of minimal frustration, the primary focus is often on the native state's energetic situation. But our analysis above shows clearly that cooperativity entails not only the stabilization of the native structure but also the destabilization of otherwise stable nonnative conformations (cf. Fig. 6). Therefore, inasmuch as the principle of minimal frustration is postulated to be well embodied by three-letter[32,39,58] or common Gō-like[62] models, results in Figs. 5 and 6 and Table I imply that the principle of minimal frustration is insufficient for the simple two-state thermodynamics and folding/unfolding kinetics of small single-domain proteins. In short, minimal frustration of the native structure appears to be *necessary but not sufficient* for proteinlike thermodynamic and kinetic cooperativities. It follows that for purposes of evaluating protein chain models and for delineating the remarkable differences between natural

[59] N. Gō, *Annu. Rev. Biophys. Bioeng*. **12,** 183 (1983).
[60] A. R. Panchenko, Z. Luthey-Schulten, R. Cole, and P. G. Wolynes, *J. Mol. Biol*. **272,** 95 (1997).
[61] N. Gō, *in* ''Old and New Views of Protein Folding'' (K. Kuwajima and M. Arai, eds.), p. 97. Elsevier, Amsterdam, 1999.
[62] C. Clementi, H. Nymeyer, and J. N. Onuchic, *J. Mol. Biol*. **298,** 937 (2000).

proteins and heteropolymers in general, the minimal frustration principle should be superseded by cooperativity principles based upon quantitative experimental criteria.

Chevron Rollover as a Consequence of Nonideal Thermodynamic Cooperativity

The present theoretical analysis of protein folding/unfolding kinetics offers a consistent perspective on both simple two-state kinetics as well as kinetics with chevron rollovers that are often operationally referred to as non–two-state. As discussed above, general considerations of polymer physics indicate that chevron rollover is probably unavoidable when interactions favoring intrachain sticking are sufficiently strong, i.e., when the stability of the native state is higher than a certain threshold. The phenomenon of chevron rollover is seen to arise from kinetic trapping that may be characterized as an internal friction effect related to the ''front factor'' in the transition state picture.[26,49] Microscopically, this means that there are increasing barrier recrossings and other impediments to conformational search with increasing native conditions, as is evident from the simulated folding trajectories under such conditions[45] (not shown here). From this vantage point, the experimentally pertinent question about chevron rollover becomes whether the omnipresent theoretical rollover occurs at a native stability accessible by experiments. Therefore, the simple two-state kinetics of small single-domain proteins imply that their theoretical rollovers occur at hypothetical native stabilities significantly higher than their real stabilities in zero denaturant. Our modeling effort above indicates that this already requires a high degree of thermodynamic cooperativity, and that many-body interactions beyond the additive contact interactions in common Gō models are necessary to achieve this feat. But obviously there are physicochemical limitations to the thermodynamic cooperativity achievable by a protein. As a result, chevron rollover would occur if the interactions in a particular protein are not sufficiently cooperative, or when native stability is relatively high at zero denaturant. Thus, chevron rollovers may be viewed as beginning signs of glassy dynamics that are expected by energy landscape theory to commence at higher native stabilities,[32,33] even though such hypothetical high-stability conditions are often not experimentally attainable for real proteins.[34]

Our simulations of the 48-mer Gō model and the 55-mer cooperative model show that when intrachain interactions are sufficiently specific, folding relaxation remains essentially single exponential in the rollover regime for native stabilities less than that at $\tilde{\varepsilon}_{\mathrm{opt}}$, i.e., for $\Delta G_{\mathrm{u}} < (\Delta G_{\mathrm{u}})_{\mathrm{opt}}$. (But folding relaxation often becomes non–single-exponential for $\Delta G_{\mathrm{u}} >$

$(\Delta G_u)_{opt}$.[26,45]) This feature echoes experimental rollover data on barnase[41] and ribonuclease A,[42] which exhibit essential single-exponential behavior after effects of proline isomerization are factored out.[41,42] Consistent with the present perspective, the zero-denaturant stability for barnase is $\approx$ 18 k_BT at 25°, which is considerably higher than that of many small single-domain proteins, but it is significantly lower than barnase's *hypothetical*[45] $(\Delta G_u)_{opt} \approx 40 k_BT$ at the same temperature. In situations in which intrachain interactions are less specific, however, chevron rollover can also be associated with non–single-exponential relaxation and kinetic partitioning.[48,53]

Coupling of Local and Nonlocal Interactions as a Possible Key to Contact-Order–Dependent Cooperative Folding

In addition to thermodynamic and kinetic cooperativities, we have also applied the remarkable empirical correlation between relative contact order (CO) and folding rate[63,64] to further narrow the types of cooperative energetics that are likely to be operating in real, small single-domain proteins. CO-dependent folding constitutes another nonredundant physical constraint on possible protein energetics because not all interaction schemes that provide for high thermodynamic and kinetic cooperativity can lead to significant correlation between CO and folding rate similar to that observed experimentally.[65] Figure 7A shows that the common Gō potential with pairwise contact energies, which is insufficient for simple two-state kinetics to begin with, is also not capable of producing CO-dependent folding. The folding rates span only approximately one order of magnitude, with much scatter and a very low correlation coefficient with CO. This and a similar finding by Jewett *et al.*[66] buttress our point that common Gō models are not adequate for certain basic generic properties of small single-domain proteins.

Recently, Jewett *et al.*[66] introduced a new interaction scheme aimed at enhancing thermodynamic cooperativity with a nonlinear relation between energy and the number of native contacts. Figure 7B shows that their cooperative interaction scheme leads to a better correlation between CO and folding rate than the common Gō potential, and the range spanned by the folding rates of the model proteins also increases to ~ 1.8 orders of magnitude. However, $r = 0.80$ for this interaction scheme is weak in comparison with the experimental correlation coefficient, and the divergence in folding

[63] K. W. Plaxco, K. T. Simons, and D. Baker, *J. Mol. Biol.* **227,** 985 (1998).
[64] D. Makarov and K. W. Plaxco, *Protein Sci.* **12,** 17 (2003).
[65] H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet.* **52,** 524 (2003).
[66] A. I. Jewett, V. S. Pande, and K. W. Plaxco, *J. Mol. Biol.* **326,** 247 (2003).

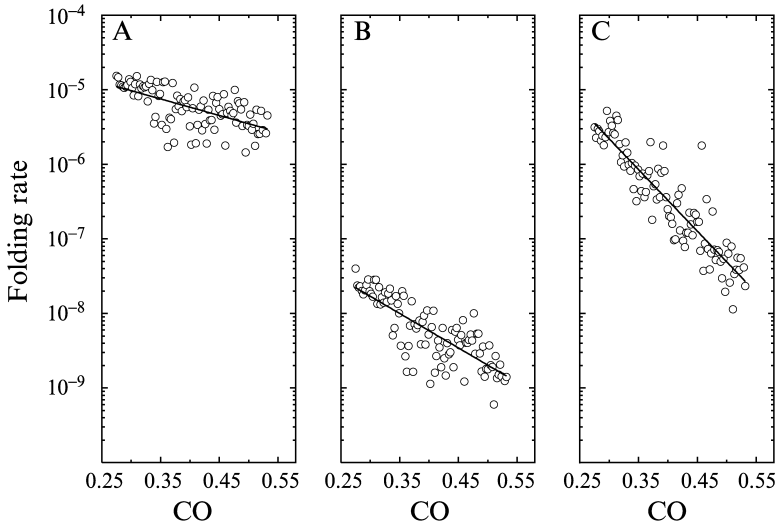

FIG. 7. Modeling CO-dependent folding. The folding rates (open circles) of a set of 97 three-dimensional 27-mer lattice model proteins with different CO values for their maximally compact native structures [H. Kaya and H. S. Chan, *Proteins Struct. Funct. Genet.* **52,** 524 (2003)] are determined under three different native-centric interaction schemes. Solid lines are least-square fits. The correlation between $\log_{10}$(folding rate) and CO is quantitated by the correlation coefficient square r^2. (A) Common Gō potential with pairwise additive contact energies; $r^2 = 0.39$. (B) The $s = 3$ cooperative scheme of A. I. Jewett, V. S. Pande, and K. W. Plaxco [*J. Mol. Biol.* **326,** 247 (2003)]; $r^2 = 0.65$. (C) The cooperative scheme of Kaya and Chan with $a = 0.1$ local–nonlocal coupling; $r^2 = 0.84$. All folding rates shown are simulated at $\varepsilon/k_BT = -1.47$, as in Kaya and Chan. Each data point in (A) and (C) is averaged from 500 trajectories, whereas each data point in (B) is averaged from 100 trajectories, using the same move set as that in Kaya and Chan.

rate between the low- and high-CO structures is still quite limited. It should be noted that the set of structures and chain moves used in the present study was chosen independently and is not exactly identical to that of Jewett *et al*. As a result, the $s = 3$ result of $r = 0.80$ ($r^2 = 0.65$) in Fig. 7B is different from the $r = 0.75$ ($r^2 = 0.57$) reported by Jewett *et al*.[66]

Our exploration thus far indicates that a cooperative interplay between local conformational preferences and the nonlocal favorable interactions responsible for protein core formation can lead to a significant CO/folding rate correlation.[65] This view, which is motivated by experimental observations (see above), differs from either the local-dominant or the nonlocal-dominant picture of protein folding.[50,67] Physically, it posits that nonlocal

[67] R. L. Baldwin and G. D. Rose, *Trends Biochem. Sci.* **24,** 26 (1999).

contact interactions cannot be strong unless local conformations sequentially near the contacting residues are essentially native so as to promote better packing between chain segments around the contacting residues. This is implemented in a model interaction scheme whereby a contact interaction is assigned a strongly favorable value when the pair of local chain segments centered around each of the contacting residues are in their native conformation, and assigned an attentuated value (by a factor a) otherwise. Figure 7C shows that this interaction scheme significantly expands the range of model folding rates to $\sim$2.6 orders of magnitude and leads to a high level of correlation with CO ($r = 0.91$) similar to that observed experimentally.[63,64] We note that the divergence between low- and high-CO folding rates in this model is still substantially lower than the six orders of magnitude among experimental folding rates of small single-domain proteins. This is probably because of the shortness of the model chains. Although this question remains to be investigated further, we have verified that this many-body interaction scheme with local–nonlocal coupling satisfies the thermodynamic and kinetic cooperativity criteria as well. Taken together, these observations lead us to hypothesize that similar local–nonlocal coupling mechanisms are at work in real proteins.

Concluding Remarks and Outlook

Results summarized in this chapter demonstrate that when generic, apparently mundane properties of simple two-state proteins are applied to evaluate self-contained polymer models, they can provide stringent constraints that lead to unexpected in-depth understanding of protein energetics. This is quite remarkable because so far only data acquired by traditional (structurally low-resolution) optical probes have been emphasized in our analysis. As discussed above, we emphasize here again that the present usage of the term ''two state'' should be appropriately construed in a structurally low-resolution context. Obviously, for a macromolecule such as a protein, it is physically inconceivable that there are only two discrete energy levels as envisioned in the ''Levinthal paradox.''[68] Conformations are expected to span a quasicontinuum of energies (or enthalpies), so it goes without saying that the ''two-state'' behavior of a polymer chain is not absolute like that of two quantum states. Nonetheless, from a polymer physics perspective, the low-resolution experimental hallmarks of small single-domain proteins, notably their calorimetric cooperativity and linear chevron plots, are remarkable feats that demand highly unusual energetics. Our model evaluation suggests strongly that in these proteins, while

[68] R. L. Baldwin, *BioEssays* **16,** 207 (1994).

conformations with intermediate energies or enthalpies do exist, their populations must be greatly reduced by specific many-body interactions. The thermodynamics of their energy landscapes may be characterized as ''near-Levinthal'' in this respect.[43]

Indeed, for our cooperative chain models, small populations of non–ground-state conformations always exist even under strongly native conditions[26] by virtue of the Boltzmann distribution. This feature is consistent with experimental data from native state hydrogen exchange.[28,69–71] As we have pointed out, some of the ''partially unfolded'' states revealed by native state hydrogen exchange may be viewed as part of a multiple-conformation native state implicitly defined by empirical baseline subtractions.[13,26] Here, as a first step, we have focused on small single-domain proteins because their near-ideal cooperative behavior is expected to provide more clear-cut information. Nevertheless, as shown above, our analysis is also pertinent to proteins that fold with chevron rollovers. Because many heteropolymer models are calorimetrically noncooperative, we expect self-contained polymer modeling to be useful in explaining downhill folding[72] as well.

Several foldability criteria pioneered by earlier researchers are extremely useful in distinguishing natural proteins from random heteropolymers. These include the consistency principle,[59] the principle of minimal frustration,[33] the energy gap[35] or stability gap ideas,[38] and the σ parameter that compares the conditions for folding versus that for chain collapse.[37] These criteria recognized key energetic ingredients that set natural proteins apart from random polypeptides. As such, these foldability criteria had to be critical in prebiotic evolution. However, these criteria alone do not address quantitatively the high degrees of thermodynamic and kinetic cooperativities of today's small single-domain proteins. The cooperativity principles discussed in this chapter build on these earlier foldability criteria. Apparently, among foldable proteins, there has been further evolutionary pressure to enhance folding/unfolding cooperativity. A probable biological impetus might be the ''avoidance of aggregation, particularly to highly insoluble amyloid fibrils'' in the crowded cellular environment, as has been pointed out by Dobson.[73] In this view, the lesser tendency for RNAs to aggregate may explain why RNAs have not evolved similar folding/unfolding cooperativities.

[69] S. Marqusee and D. Wildes, Chapter 15, this volume.
[70] H. Maity, W. K. Lim, J. N. Rumbley, and S. W. Englander, *Protein Sci.* **12,** 153 (2003).
[71] C. Woodward, N. Carulla, and G. Barany, Chapter 17, this volume.
[72] M. M. Garcia-Mira, M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Muñoz, *Science* **298,** 2191 (2002).
[73] C. M. Dobson, *Trends Biochem. Sci.* **24,** 329 (1999).

The work presented here shows that coarse-grained models are effective in delineating the general principles of protein folding/unfolding cooperativity. Ultimately, however, the physical bases (or lack thereof, for that matter) of the many-body interactions postulated in our coarse-grained models have to be ascertained. Some of these investigations are already under way. For example, we found that the temperature dependence of the hydrophobic effect (Fig. 3B) may account for the lack of postdenaturational chain expansion, and anticooperativity of certain hydrophobic interactions (Fig. 3C) may lessen the tendency of premature chain collapse and thus contribute to overall folding/unfolding cooperativity. In this pursuit, it would be extremely interesting to see how side chain packing, hydrogen bonding, and other atomic interactions may give rise to mechanisms of local–nonlocal coupling similar to those proposed above.

[17] Native State Hydrogen-Exchange Analysis of Protein Folding and Protein Motional Domains

By CLARE WOODWARD, NATÀLIA CARULLA, and GEORGE BARANY

Introduction

Slow hydrogen isotope exchange is a defining characteristic of the folded state of proteins. If a protein is water soluble, not self-aggregated, and has slow exchange of 10–25% of its backbone amide groups, it is reasonable to presume that folding to a biologically functional conformational ensemble has occurred. Slow exchange in this context refers to buried amide hydrogens that exchange with solvent hydrogens of different isotopic composition on the hour-to-day time scale at neutral pH and room temperature. Pioneering investigators of protein hydrogen exchange[1] recognized that slow exchange implies the existence of internal motions that expose buried amide NH groups to solvent and thereby permit isotope exchange. When viewing a diagram of protein hydrogen exchange as in Fig. 1A, the mind's eye should fill in not only a picture of an actual protein, but also the third and fourth dimensions of space and time, to envision an ensemble that fluctuates over diverse conformations and on many time scales, but for the most part populates native structure. Although hydrogen exchange is a result of internal motility, the hydrogen-exchange experiment does not usually yield the frequency or amplitude of a motion, but

[1] A. Hvidt and S. O. Neilsen, *Adv. Protein Chem.* **21,** 287 (1966).

0076-6879/04 $35.00