PROTEINS: Structure, Function, and Genetics 40:543–571 (2000)

Modeling Protein Density of States: Additive Hydrophobic Effects Are Insufficient for Calorimetric Two-State Cooperativity

Hue Sun Chan*

Department of Biochemistry and Department of Medical Genetics and Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

ABSTRACT **A well-established experimental criterion for two-state thermodynamic cooperativity in protein folding is that the van't Hoff enthalpy ΔH_{vH} around the transition midpoint is equal, or very nearly so, to the calorimetric enthalpy ΔH_{cal} of the entire transition. This condition is satisfied by many small proteins. We use simple lattice models to provide a statistical mechanical framework to elucidate how this calorimetric two-state picture may be reconciled with the hierarchical multistate scenario emerging from recent hydrogen exchange experiments. We investigate the feasibility of using inverse Laplace transforms to recover the underlying density of states (i.e., enthalpy distribution) from calorimetric data. We find that the constraint imposed by $\Delta H_{vH}/\Delta H_{cal} \approx 1$ on densities of states of proteins is often more stringent than other "two-state" criteria proposed in recent theoretical studies. In conjunction with reasonable assumptions, the calorimetric two-state condition implies a narrow distribution of denatured-state enthalpies relative to the overall enthalpy difference between the native and the denatured conformations. This requirement does not always correlate with simple definitions of "sharpness" of a transition and has important ramifications for theoretical modeling. We find that protein models that assume capillarity cooperativity can exhibit overall calorimetric two-state-like behaviors. However, common heteropolymer models based on additive hydrophobic-like interactions, including highly specific two-dimensional Gō models, fail to produce proteinlike $\Delta H_{vH}/\Delta H_{cal} \approx 1$. A simple model is constructed to illustrate a proposed scenario in which physically plausible local and nonlocal cooperative terms, which mimic helical cooperativity and environment-dependent hydrogen bonding strength, can lead to thermodynamic behaviors closer to experiment. Our results suggest that proteinlike thermodynamic cooperativity may require a cooperative interplay between local and nonlocal interactions. The prospect of using calorimetric data to constrain Z-scores of knowledge-based potentials is discussed. Proteins 2000;40:543–571.**

Key words: calorimetry; van't Hoff enthalpy; specific heat; hydrogen exchange; lattice model; Z-score

INTRODUCTION

What can we learn about the fundamental forces governing protein structure and stability from general, "generic" properties of protein folding[1]? Numerous statistical mechanical models have been proposed to rationalize experimentally observed thermodynamic behaviors of proteins (see, e.g., Refs. 2–5, and Refs. 6–14 for reviews). The typical starting point of these analyses is a model partition function describing the distribution of energies among various conformational states. The partition function depends on what protein interactions are assumed and how conformational freedom is treated by the given model. Theoretical predictions of thermodynamic quantities can then be deduced from the model partition function. The hope is that, through exploring and testing various model partition functions and comparing their predictions with experiments, useful links can be established between general thermodynamic properties and the underlying microscopic interactions in proteins. For this approach to be fruitful, therefore, it is crucial to determine to what degree model partition functions are consistent with experiments and, if current models are not satisfactory, how they may be improved by applying proper experimental constraints.

Calorimetric data are of immediate relevance to this endeavor, because it directly determines enthalpy changes as proteins undergo conformational transitions.[15,16] In principle, an effective partition function of the protein can be deduced by appropriate integration of the experimentally determined enthalpy as a function of temperature, offering the possibility that the density of states may then be obtained by deconvolution.[17–19]

One motivation for the development and refinement of calorimetric methods for protein folding was the notion that some proteins' folding-denaturation transitions were

Grant sponsor: Medical Research Council of Canada; Grant number: MT-15323.

*Correspondence to: Hue Sun Chan, Department of Biochemistry, Faculty of Medicine, University of Toronto Medical Sciences Building, 5th Floor, 1 King's College Circle, Toronto, Ontario, Canada M5S 1A8. E-mail: chan@arrhenius.med.toronto.edu

Received 7 January 2000; Accepted 11 April 2000

two-state[20,21] needed to be "firmly substantiated."[15] The quantitative criterion adopted by most calorimetrists for two-state transitions has been that the van't Hoff enthalpy, ΔH_{vH}, calculated at or around the peak of the specific heat, has to be equal or approximately equal to the calorimetric enthalpy ΔH_{cal} of the entire transition process. The observation of $\Delta H_{vH}/\Delta H_{cal} \approx 1$ for a protein provides direct evidence that the vast number of its accessible conformations can be grouped into essentially only two narrowly distributed enthalpic states at the transition midpoint. This condition is satisfied by many small single-domain proteins, whereas the transition of proteins with multiple domains can be more complex.[22] As a necessary first step, in this work we focus on modeling calorimetrically two-state proteins.

The $\Delta H_{vH}/\Delta H_{cal} \approx 1$ condition represents a stringent constraint on chain models of proteins, and, thus, it can serve as a powerful tool in designing proteinlike interaction potentials, yet its implications on theoretical modeling are often not fully appreciated. The two-state picture it dictates may appear straightforward in analyses that deal exclusively with abstract "states" and make no explicit reference to ensembles of chain conformations. But the two-state $\Delta H_{vH}/\Delta H_{cal} \approx 1$ cooperativity of real proteins is highly remarkable from a polymer perspective, because it implies that there are only very limited enthalpy variations among the denatured conformations. However, in heteropolymer models that attempt to capture the essential physics with flexible chains and pairwise contact interactions,[5,23–25] if the contact energies are enthalpic—as is often implicitly assumed (see discussions in Chan[26] and Chan and Dill[27]) it is not trivial to envision how wide variations in enthalpy among the large number of non-native conformations can be avoided.

In recent years, computational models have begun to provide useful insight into the physical origins of protein thermodynamic cooperativity. These investigations include introducing many-body contact energies to enhance thermodynamic cooperativity in high-coordination lattice models,[28,29] elucidating the differences between one-state and two-state transitions using simple exact models,[11,24] studying side-chain effects in two- and three-dimensional lattice models,[5,30] and simulating a simplified continuum chain model with hydrogen-bond-like orientation-dependent interactions.[31–33] However, these studies did not compare their predictions with the calorimetric requirement for two-state cooperativity.

More recently, Mohanty et al.[34] connected their interesting and elegant simulations of a two-chain system to calorimetric cooperativity and obtained a van't Hoff to calorimetric enthalpy ratio very close to unity. Because their main focus was the dimerization process, they did not address the question about whether a single (monomer) chain exhibits thermodynamic two-state cooperativity in their model. Their results do indicate, however, that thermal transitions within the dimeric state is not two-state. In a detailed and insightful study, Zhou et al.[35] compared the calorimetric criterion with other two-state criteria of folding and provided an in-depth discussion on the role of baseline subtractions (see Materials and Methods). They reached the important conclusion that, for an arbitrary system, the calorimetric criterion is necessary but not sufficient for two-state processes defined by well-separated distributions of appropriate order parameters.

The questions we address are different from that of Zhou et al.[35] We are primarily interested in using the calorimetric two-state criterion as an experimental constraint on protein models, regardless of whether these models are classified as "two-state" by other standards. We find that for proteins that can be completely denatured by heat, the experimental calorimetric two-state criterion is useful in evaluating whether model energetics are proteinlike. We show below that the quantitative $\Delta H_{vH}/\Delta H_{cal} \approx 1$ criterion is often more restrictive on protein models than the qualitative requirement for a relatively sharp transition[5] or the mere existence of a free energy barrier between the native and denatured states.[24,29,33,35] To be consistent with experimental observations on two-state cooperativity, we propose that chain models for single-domain proteins should be evaluated by the calorimetric standard. It is of interest that the constraint imposed by $\Delta H_{vH}/\Delta H_{cal} \approx 1$ on the density of states is formally similar to a spin-glass-inspired minimal frustration condition.[9,12,36] The original proposal for minimal frustration as an organizing principle of proteins has been motivated more by kinetic than by purely thermodynamic considerations.[3,37,38] This finding is noteworthy because it suggests a new connection between calorimetric two-state thermodynamic cooperativity and kinetic foldability.

As pointed out by Hilser and Freire,[39] it would be misleading to interpret calorimetric "two-state" in absolute terms and infer from $\Delta H_{vH}/\Delta H_{cal} \approx 1$ the total nonexistence of conformational states with enthalpies intermediate between the two dominant values. Clearly, such an hypothetical situation is physically untenable for polymers. Indeed, it is interesting to note, as Baldwin[40] has commented, that an early literal interpretation of the calorimetric two-state picture might have led to the belief or tacit assumption of a golf-course energy landscape that contributed to the formulation of the so-called Levinthal's paradox[41,42] (see also Privalov[43]). The main flaw of Levinthal's argument is its neglect of the funnel-like aspects of the folding landscape,[44–46] but the possibility of a funnel with conformations whose energies are intermediate between the fully unfolded state and the native state is precluded if one incorrectly interprets the two-state picture in absolute terms!

Calorimetric methods are not very sensitive to structural details. At best, the $\Delta H_{vH}/\Delta H_{cal} \approx 1$ condition implies only that the populations of conformations with intermediate enthalpies are negligible around the transition midpoint. It cannot rule out the existence of sparsely populated intermediate conformations, especially those that are populated under native conditions. Recent advances in hydrogen exchange and NMR techniques have revealed a multiple-state picture for a number of small proteins.[47–54] As well, denatured states of some proteins

under native conditions have been found to be relatively compact,[55,56] suggesting that their most stable conformations of the denatured ensemble under native conditions are not identical to that of the fully unfolded state. Although $\Delta H_{vH}/\Delta H_{cal} \approx 1$ remains a well-established experimental fact for many small single-domain proteins,[43] these new hydrogen-exchange results have shown the existence of intermediate states. Hence, both sets of observations need to be incorporated into a consistent statistical mechanical framework.[51,57] We propose such a perspective below.

Simple lattice modeling is a useful tool in our analysis. The present article focuses mainly on two-dimensional exact models. Three-dimensional models are considered in an upcoming article.[58] Theoretical insight has been gained by analyses based on hypothetical densities of states[20] and densities of states not derived from an explicit chain model, such as the Ising-type model considered by Ikegami.[59] Ultimately, however, it is the interplay between conformational entropy and effective intrachain interactions (i.e., with solvation effects included) that holds the key to understanding thermodynamic cooperativity. In this regard, heteropolymer models with explicit chain representations are indispensable for providing a physically plausible account of conformational entropy.

Proteinlike thermodynamic cooperativity has been predicted by a model in which a polypeptide chain is postulated to be made up of units of contiguous chain segments that fold and unfold in an all-or-none manner.[60,61] This is an elegant and successful approach. However, it is not a fully self-contained heteropolymer model because the conformational distribution is not dictated solely by the model potential, which is basically composed of additive contact energies. Large numbers of a priori viable conformations are precluded by the postulated all-or-none folding and unfolding of the units. The physical basis of this postulated behavior remains to be fully elucidated.

Therefore, notwithstanding the limitations of simple protein lattice models[54,62,63] (see also Shakhnovich[64]), we believe it is appropriate to take an incremental, constructive approach,[65] by first starting with minimalist heteropolymer models and then proceed progressively to add and test various plausible interactions for their abilities to produce proteinlike cooperativity. This bottom-up program allows for a step-by-step dissection of various energetic contributions, so that at each stage we can learn from both the models' proteinlike as well as non-protein[54] properties. This approach is complementary to methods that begin with full atomistic details. In any case, to our knowledge, no self-contained atomistic model of a single protein chain has yet been demonstrated to exhibit $\Delta H_{vH}/\Delta H_{cal} \approx 1$ cooperativity. We expect the present approach to supply rudimentary understanding of the essential ingredients for protein thermodynamic cooperativity, which should help guide searches for their fundamental physicochemical basis in more detailed models and by experiments.

RESULTS AND DISCUSSION

The Calorimetric Two-State Criterion Implies That the Enthalpy Distribution in the Denatured State Is Narrow

What constraint does the calorimetric condition $\Delta H_{vH}/\Delta H_{cal} \approx 1$ impose on protein models? Of most relevance is this criterion's implication on the predicted distribution of enthalpies (H's) among different chain conformations of a given protein. This distribution is called the density of states. Experimentally, thermodynamics that give rise to the observation of $\Delta H_{vH}/\Delta H_{cal} \approx 1$ for many small single-domain proteins are described by the "transition part" of the *excess* heat capacity function $C_{P,ex}(T)$ (i.e., the quantity $\langle \Delta C_{P,tr} \rangle$ defined by Freire;[19] see Materials and Methods). $C_{P,ex}$ is obtained by performing appropriate subtractions on raw calorimetric data. In Materials and Methods, we give the definition of the density of enthalpic states $g(H)$ [Eq. (26)] and discuss in detail the relation between the thermodynamics described by $C_{P,ex}$ and the fundamental physical partition function of the complete system of a protein plus its surrounding solvent and statistical mechanical issues pertinent to treating calorimetric data with baseline subtractions. Here we compare predicted enthalpy distributions among conformations in simple lattice chain models to evaluate their conformity to the calorimetric two-state requirement.

We consider the thermodynamics of a single chain. Unless otherwise stated, we use discretized enthalpy distributions in the form of the partition function in Eq. (26) in Materials and Methods. For our present purposes, it suffices (after proper calorimetric baseline subtractions) to assign only one single enthalpy level H_N to the native state. Except for some general constructs used in the Materials and Methods section, all thermodynamic quantities below refer to the system defined by $C_{P,ex}$. In nearly all applications, the partition function is normalized such that $g(H_N) = 1$.

We begin with the excess enthalpy function $\langle \Delta H(T) \rangle = \langle H(T) \rangle - H_N$ [Eqs. (28) and (30)]. This is the heat absorbed by a protein chain when its temperature is raised, leading to an increasing fraction of its population being transferred from the native state to conformations with higher enthalpies. The symbol $\langle \ldots \rangle$ denotes Boltzmann averaging over the *entire* ensemble, i.e., the single native (N) conformation and all other conformations that constitute the denatured (D) ensemble. To study heat denaturation, experimental data are needed to cover the entire transition region, between certain temperatures T_0 and T_1 ($> T_0$) that are, respectively, well below and well above the transition temperature. In other words, practically the entire chain population is in the native conformation at $T = T_0$, such that $\langle \Delta H(T_0) \rangle = 0$. In simple models that neglect freezing of the solvent, formally $\langle \Delta H(T) \rangle \to 0$ as $T \to 0$. In this study we do not consider cold denaturation.[26,27,66,67]

The calorimetric enthalpy ΔH_{cal} is the total heat absorbed by the entire transition process of heat denaturation, obtained by integrating the area under the $C_{P,ex}$

curve.[15–19,35] Therefore, in simple models, we may formally define

$$\Delta H_{\rm cal} = \lim_{T\to\infty}\langle \Delta H(T)\rangle. \tag{1}$$

In practice, T cannot be increased indefinitely, and the chain population is almost all in the denatured state at a certain finite temperature T_1 above the transition temperature, hence

$$\Delta H_{\rm cal} = \langle \Delta H(T_1)\rangle \tag{2}$$

experimentally. We now relate the calorimetric enthalpy to a van't Hoff enthalpy. The population-based van't Hoff enthalpy[16] is obtained by a van't Hoff analysis of the equilibrium constant

$$K_{\rm eq}(T) = \frac{[\rm N]}{[\rm D]} = \frac{e^{-H_N/(k_BT)}}{\sum_{\rm D} g(H_i)e^{-H_i/(k_BT)}}. \tag{3}$$

$K_{\rm eq}$ is the ratio of equilibrium population [N] in the native state versus [D] in the denatured state, and $\Sigma_{\rm D}$ only sums over denatured enthalpies. The van't Hoff enthalpy is then defined by a temperature derivative of the equilibrium constant,

$$\Delta H_{\rm vH}(T) = -k_BT^2\frac{\partial \ln K_{\rm eq}(T)}{\partial T} = \langle H(T)\rangle_{\rm D} - H_{\rm N} = \langle \Delta H(T)\rangle_{\rm D}, \tag{4}$$

resulting in a quantity that is temperature dependent in general. Note that Boltzmann averages denoted by $\langle \ldots\rangle_{\rm D}$ are only over the denatured ensemble, excluding the native state. For example, in the expression above, the average

$$\langle H(T)\rangle_{\rm D} = \frac{\sum_{\rm D} g(H_i)H_ie^{-H_i/(k_BT)}}{\sum_{\rm D} g(H_i)e^{-H_i/(k_BT)}}. \tag{5}$$

This is the average enthalpy of the denatured ensemble of all non-native conformations. In general, it is temperature dependent. When there are multiple enthalpy levels in the denatured ensemble, the population distribution of denatured conformations shifts as temperature is varied. In that case, the average enthalpy of the denatured ensemble would be lower under native conditions (at low temperatures) and increases monotonically with temperature. According to definition Eq. (4), the temperature-dependent population-based van't Hoff enthalpy is equal to the average enthalpy difference between the denatured state and the native state. It follows from Eqs. (3) and (5) that

$$\Delta H_{\rm vH}(T) = \langle \Delta H(T)\rangle_{\rm D} = [1 + K_{\rm eq}(T)]\langle \Delta H(T)\rangle. \tag{6}$$

This means that the van't Hoff enthalpy is always larger than the average enthalpy of the entire ensemble relative to that of the native state, because $K_{\rm eq} > 0$. However, for a protein whose native population is negligible at a sufficiently high temperature T_1 such that practically $K_{\rm eq}(T_1) = 0$, Eqs. (2), (4), and (6) imply that

$$\Delta H_{\rm vH}(T_1) = \Delta H_{\rm cal}. \tag{7}$$

Similarly, if $K_{\rm eq}$ of a model approches zero as $T \to \infty$, Eqs. (1), (4), and (6) imply that $\Delta H_{\rm cal} = \Delta H_{\rm vH}(\infty)$. Under these conditions, $\Delta H_{\rm vH}(T) < \Delta H_{\rm cal}$ for practically all temperatures. However, for systems whose $K_{\rm eq}$'s do not approach 0 as $T \to \infty$ (Zhou et al.[35]), Eqs. (1) and (6) imply that it is possible for $\Delta H_{\rm vH}(T)$ to be greater than $H_{\rm cal}$ at sufficiently high T. In this article, we are interested in protein models that satisfy $K_{\rm eq} \approx 0$ as $T \to \infty$.

The van't Hoff enthalpy is the general temperature-dependent $\Delta H_{\rm vH}(T)$ in Eq. (4) evaluated at the transition midpoint $T = T_{1/2}$, where $K_{\rm eq}(T_{1/2}) = 1$ (Privalov and Potekhin[16]),

$$\Delta H_{\rm vH} = \Delta H_{\rm vH}(T_{1/2}) = \langle \Delta H(T_{1/2})\rangle_{\rm D}. \tag{8}$$

To simplify notation, $\Delta H_{\rm vH}$ is understood to be evaluated at $T = T_{1/2}$ in this work when its T dependence is not indicated. Because it is difficult to determine the population midpoint, experimental $\Delta H_{\rm vH}$'s are often evaluated at the temperature when the $C_{P,\rm ex}$ function peaks[16–18] or alternatively at the enthalpy midpoint when $\langle \Delta H(T)\rangle = \Delta H_{\rm cal}/2$, i.e., when one half of the heat associated with the entire transition has been absorbed.[35]

It follows from Eqs. (7) and (8) and the above discussion on the temperature dependence of the van't Hoff enthalpy that the calorimetric two-state condition, $\Delta H_{\rm vH}/\Delta H_{\rm cal} \approx 1$, requires that: *the difference between the average enthalpy of the denatured ensemble (of all non-native conformations) and the native enthalpy at or around the heat denaturation transition midpoint has to be essentially the same as the corresponding difference when the transition is completed at a higher temperature.* This leaves open the possibility for this enthalpy difference to be smaller when the fractional denatured population is small at a temperature lower than the midpoint, especially when the protein is under native conditions. (Illustrations of this principle are provided below in the lower panels of Figs. 4 and 5.) Nonetheless, this criterion is a rather stringent constraint on protein densities of states, as will be shown in subsequent analyses. Other definitions of van't Hoff enthalpies have also been used for the calorimetric two-state criterion. Their relations with definition Eq. (8) will be discussed below in a separate section. Model results in this article indicate that the general implications of the calorimetric two-state criterion on protein densities of states are not sensitive to the choice among a number of commonly used formulas for van't Hoff enthapies.

Figure 1 shows a typical scan of a small single-domain protein that satisfies the calorimetric two-state criterion and a density of states (Fig. 1B) satisfying $\Delta H_{\rm vH}/\Delta H_{\rm cal} = 1.0$ with only two enthalpies, which fits the experimental data very well. However, the deduction of the underlying density of states from a $C_{P,\rm ex}$ function is not assumption-free, and the result is in general not unique (see Materials and Methods for details). Therefore, it is impossible to deduce with certainty protein densities of states from

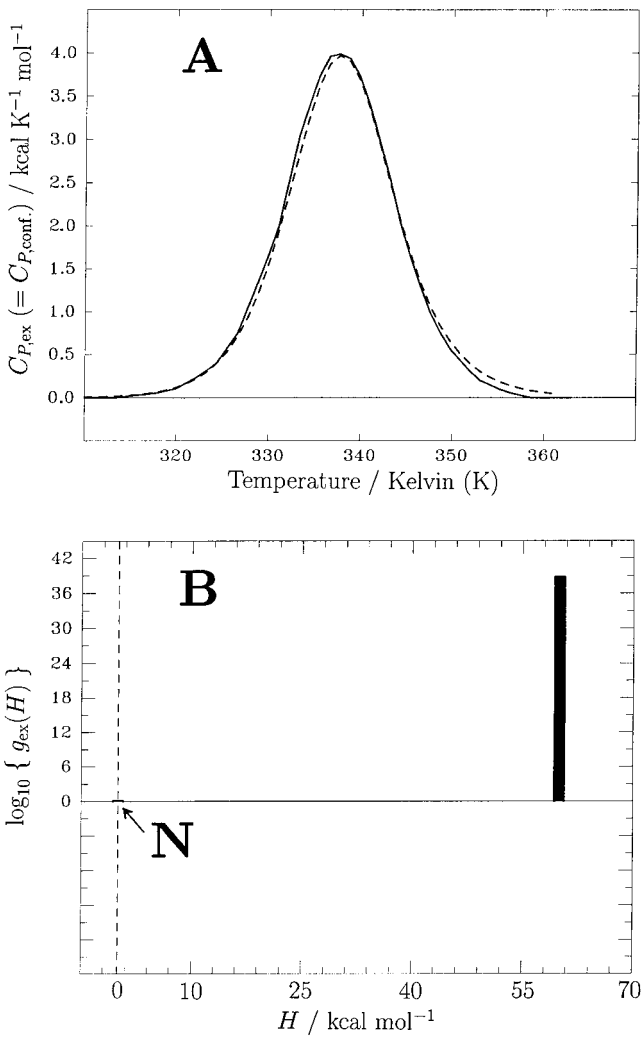

Fig. 1. A two-state fit of the transition excess heat capacity of the Ile → Val76 mutant of chympotrypsin inhibitor 2. **A:** $C_{P,ex}$ (solid curve) is the "excess C_P" of Jackson et al.,[144] i.e., normalized data after baseline subtractions in their Figure 3B, equivalent to the quantity $\langle \Delta C_{P,tr} \rangle$ in Freire.[19] The $C_{P,ex} = C_{P,conf.}$ label on the vertical axis emphasizes that the enthalpic variations described by $C_{P,ex}$ are exclusively those associated with conformational transitions (see Materials and Methods). The dashed curve shows a two-state fit by the density of states $g_{ex}(H)$ in (B). **B:** $g_{ex}(H) \neq 0$ only at two H values: $H = 0$ (native enthalpy, indicated by N and the vertical dashed line), and $H = 60.0$ kcal mol^{-1}. The latter was obtained by numerical integration of the area under the $C_{P,ex}$ curve in (A), which corresponds to the calorimetric enthalpy ΔH_{cal} of the entire transition. [Eq. (30) in the Materials and Methods section is used, with $T_0 = 310$K, and $\langle H(T_0) \rangle = 0$]. For the present analysis, the numerical difference between this value of H and the $\Delta H_{cal} = 61.9$ kcal mol^{-1} reported by Jackson et al.[144] is inconsequential. $g_{ex}(0) = 1$ and $g_{ex}(60.0) = 5.68 \times 10^{38}$ are used to compute the dashed curve in (A).

experiments and compare them with theoretical densities of states. Instead, we have to take the reverse approach. We first construct model densities of states, compute the resulting heat capacity and enthalpy functions, and then compare these predictions with the properties of experimental heat capacity and enthalpy functions to evaluate whether the models are sufficiently proteinlike. In a general sense, however, we do adopt essential elements of the basic picture in Figure 1 as the appropriate physical interpretation of calorimetric data on small single-domain proteins, in that we do not consider the possibilities of non-positive-definite effective densities of states discussed in the Materials and Methods section. Because the ratio $\Delta H_{vH}/\Delta H_{cal}$ is a key parameter governing the shapes of the specific heat and enthalpy functions, we now explore what deviations from the two-discrete-state picture in Figure 1 are allowed when $\Delta H_{vH}/\Delta H_{cal}$ is close to, but not exactly equal to unity. In the present work, we assume putatively that experimental observations of $\Delta H_{vH}/\Delta H_{cal} \approx 1$ (as for Fig. 1) are essentially accurate, while keeping in mind that various experimental uncertainties, including those in determining protein concentrations, may lead to errors of 5–10% in calorimetric data.[68]

Using the Calorimetric Two-State Criterion to Constrain a Random-Energy Model of Proteins

The two-state picture depicted by Figure 1B is absolute, with only two enthalpy levels. Although it fits the calorimetric data well, from the vantage point of polymer physics, it is inconceivable that no conformational state exists with enthalpies intermediate between the native and fully unfolded states. Here we show that in general they can exist under the $\Delta H_{vH}/\Delta H_{cal} \approx 1$ condition, as has been pointed out in similar contexts by Freire,[19] Zhou et al.,[35] and others. However, the calorimetric two-state condition places limits on their populations. We now apply a simple continuum model to illustrate the calorimetric constraint. Consider the density of states

$$g(H) = \delta(H) + \theta(H) \frac{g_D}{\sqrt{2\pi}\sigma_H} e^{-(H - H_D)^2/(2\sigma_H^2)}, \qquad (9)$$

where $\delta(H)$ is the Dirac delta function, the native enthalpy $H_N = 0$, the step function $\theta(H) = 1$ for $H \geq 0$, and $\theta(H) = 0$ for $H < 0$. g_D ($\gg 1$) and H_D are, respectively, the total number and average enthalpy of the denatured conformations, whereas the standard deviation σ_H specifies the width of the enthalpy distribution among them. This density of state is very similar to that of a Gaussian random-energy model used in early spin-glass-inspired analyses of protein folding.[3,38,69,70] The corresponding partition function is given by the continuum Eq. (25) in the Materials and Methods section. When σ_H is small compared to $H_D/\sqrt{2 \ln g_D}$, replacing $\theta(H)$ in Eq. (9) by unity and extending the lower limit of integration in Eq. (25) from zero to $-\infty$ incur only a negligible error, in which case the temperature-dependent van't Hoff enthalpy is given by

$$\Delta H_{vH}(T) = \langle \Delta H(T) \rangle_D = H_D - \frac{\sigma_H^2}{k_B T} \qquad (10)$$

for $T > \sigma_H^2/(k_B H_D)$, whereas the calorimetric enthalpy $\Delta H_{cal} = \lim_{T \to \infty} \langle \Delta H(T) \rangle_D = H_D$. Equation (10) indicates that if the distribution of denatured enthalpies is sharply peaked, the calorimetric two-state condition would be satisfied, notwithstanding the existence of a continuous nonvanishing distribution of enthalpic states from $H = H_N = 0$ to $H \to \infty$. More specifically, when $\sigma_H^2 \ll k_B T H_D$,

the temperature dependence of Eq. (10) is weak, and $\Delta H_{\rm vH}(T)/\Delta H_{\rm cal} \approx 1$ for practically all nonzero T.

The equilibrium constant of this model is

$$K_{\rm eq}(T) = \frac{[\rm N]}{[\rm D]} = \left\{ g_{\rm D} \exp\left[-\frac{H_{\rm D}}{k_B T} + \frac{\sigma_H^2}{2(k_B T)^2} \right] \right\}^{-1}. \quad (11)$$

The transition midpoint temperature $T_{1/2}$ is determined by equating this expression to unity. This results in a quadratic equation in T with two solutions. The physical solution is

$$T_{1/2} = \frac{\sigma_H^2}{k_B [H_{\rm D} - \sqrt{H_{\rm D}^2 - 2\sigma_H^2 \ln g_{\rm D}}]}, \quad (12)$$

whereas the other solution is discarded because it is an artifact of the approximation used to derive Eq. (11) by extending the enthalpy integration range to $-\infty$. Substituting Eq. (12) into Eq. (10),

$$\frac{\Delta H_{\rm vH}}{\Delta H_{\rm cal}} = \frac{\Delta H_{\rm vH}(T_{1/2})}{\Delta H_{\rm cal}} = \sqrt{1 - \frac{2\sigma_H^2 \ln g_{\rm D}}{H_{\rm D}^2}} \quad (13)$$

is obtained as a general relation between the calorimetric two-state cooperativity parameter and σ_H. It follows that $\Delta H_{\rm vH}/\Delta H_{\rm cal} \approx 1$ requires

$$\frac{H_{\rm D}}{\sigma_H \sqrt{2 \ln g_{\rm D}}} \gg 1. \quad (14)$$

Remarkably, the expression $H_{\rm D}/(\sigma_H \sqrt{2 \ln g_{\rm D}})$ is equal to one half of the ratio between the folding temperature $T_{\rm F}$ and the glass temperature $T_{\rm G}$ in the model of Onuchic et al.[12] (see their Eq. (12); and also Goldstein et al.[36]). If their energy distribution is taken as an enthalpy distribution, the calorimetric two-state condition Eq. (14) becomes $T_{\rm F}/T_{\rm G} \gg 2$ in their notation, which is consistent with their condition for fast folding kinetics. By the above analysis of the Gaussian model, $T_{\rm F}/T_{\rm G}$ = 3.3, 6.4, and 14.2 are required for $\Delta H_{\rm vH}/\Delta H_{\rm cal}$ = 0.8, 0.95, and 0.99, respectively. This finding shows that, in the context of this simple continuum model, the calorimetric two-state condition requires a narrow distribution of denatured-state enthalpies relative to the average difference in enthalpy between the native and the denatured states.

Evaluation of Simple Lattice Models by the Calorimetric Criterion

How do other commonly studied simple protein models measure up to the calorimetric standard? Figure 2 compares Gaussian-model densities of states [Eq. (9)] with the discrete density of states of an HP lattice model sequence whose folding transition has been classified previously as thermodynamically two-state by a more relaxed criterion that only requires a population minimum at an intermediate energy (enthalpy) separating the native and the denatured states.[11,24] However, by the more stringent calorimetric criterion, the transition of this HP sequence is not two-state, because its $\Delta H_{\rm vH}/\Delta H_{\rm cal}$ = 0.43 (see below). More proteinlike Gaussian-model densities of states in Figure 2 with $\Delta H_{\rm vH}/\Delta H_{\rm cal}$ = 0.8 and 0.95 have significantly narrower distributions of denatured (non-native) enthalpies.

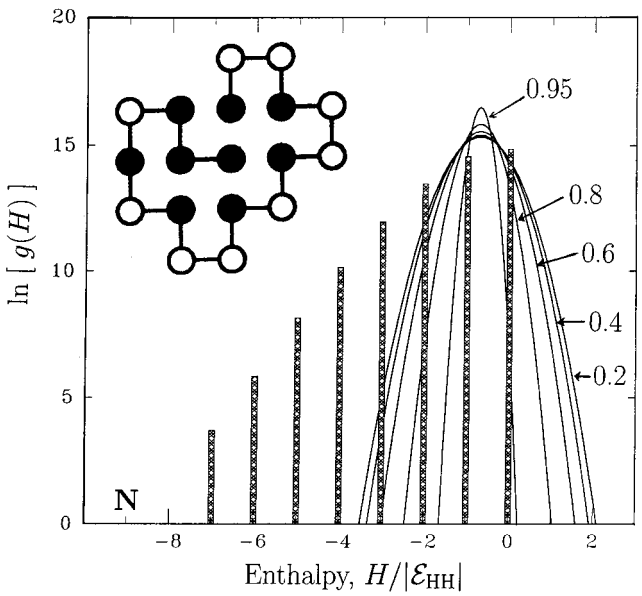

Fig. 2. Calorimetric two-state behavior requires a narrow distribution of denatured state enthalpies. A sequence in the two-dimensional HP model is shown in its native conformation. Filled and open circles are H and P monomers, respectively. Interaction energies are taken to be entirely enthalpic, with $\epsilon_{\rm HH} < 0$ for each HH contact, whereas HP and PP contacts are neutral (interaction enthalpy = 0). The histogram shows the natural logarithm of the number of conformations of this sequence at different enthalpy levels. The number of HH contacts is equal to $H/\epsilon_{\rm HH}$. The native conformation has 9 HH contacts (indicated by N).[11,24] The continuous curves show the natural logarithm of densities of states g(H)'s of five Gaussian models (see text), with σ_H = 0.505, 0.473, 0.413, 0.310, and 0.161. The curves are labeled by their $\Delta H_{\rm vH}/\Delta H_{\rm cal}$'s, which equal 0.2, 0.4, 0.6, 0.8, and 0.95, respectively. In these comparisons, the lattice model parameters $\Delta H_{\rm cal} = 8.29|\epsilon_{\rm HH}|$ and $g_{\rm D}$ = 5,808,334 are used for the Gaussian models.

As discussed above, the calorimetric criterion compares the enthalpy distribution near the transition midpoint with that when heat denaturation is completed. In this light, the dissection of the denatured ensemble in Figure 3 illustrates clearly why this HP sequence fails the calorimetric two-state test. Figure 3 shows how chains with different numbers of HH contacts contribute to the overall stability $(G_{\rm D} - G_{\rm N})/(k_B T) = \ln K_{\rm eq}$ [see Eq. (3)] of the entire denatured ensemble, as a function of the HH "sticking parameter" $-\epsilon_{\rm HH}/(k_B T)$. Under native conditions, i.e., when the sticking parameter $-\epsilon_{\rm HH}/(k_B T)$ is large, the most stable conformations in the denatured ensemble are those with the largest numbers of hydrophobic (HH) contacts. However, the relative stabilities of conformations shift as denaturation proceeds with decreasing sticking. This result is because as sticking becomes less important, conformational entropy plays an increasingly dominant role, and there are many more conformations in the high enthalpy levels. Hence, the rates at which the stabilities of different lines in Figure 3 shift depend on each line's balance between conformational entropy and favorable enthalpic interactions among the chains. In other words, they depend on the density of states of the

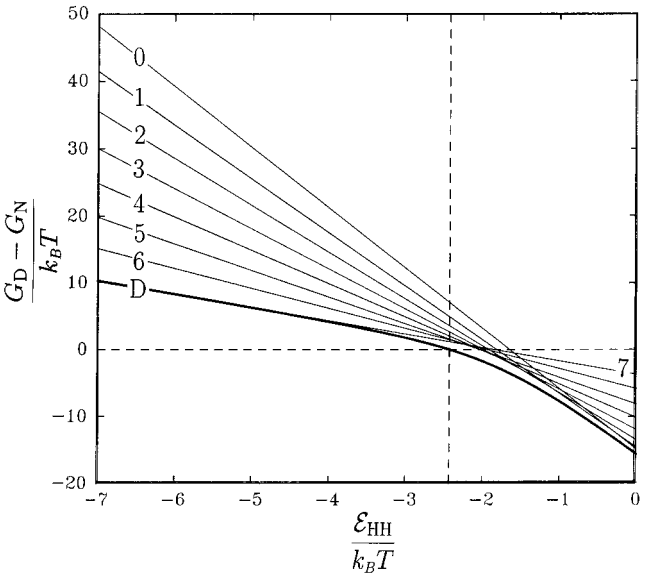

Fig. 3. The stability of denatured conformations for the HP sequence in Figure 2. More positive values of $(G_D - G_N)/(k_BT)$ means that a given state is more unstable relative to the native state. The stability of conformations with a given number h of HH contacts is described by the straight line $(G_D - G_N)/(k_BT) = -\ln g(h) - (h_N - h)\epsilon_{HH}/(k_BT)$, where $g(h)$ is the number of conformations with h HH contacts, ϵ_{HH} is the enthalpy of an HH contact, and h_N is the number of HH contacts in the native conformation[132]; $h_N = 9$ for this sequence. The stability lines are labeled by h. The thick curve with label "D" shows the stability of the entire denatured ensemble relative to the native conformation. The vertical dashed line indicates the transition midpoint.

protein chain. The crucial observation here is that, for this model sequence, at the transition midpoint, although the differences in stability among different lines are narrower than that under strongly native conditions, the denatured ensemble is still dominated by conformations with the largest number (in this case $h_N - 2$) of HH contacts: The $h = 7$ line intersects the vertical dashed line at a point lower than all other lines for smaller h's. This implies that the average enthalpy of the denatured state is low at the transition midpoint. However, when the transition is completed at weak HH sticking ($-\epsilon_{HH}/(k_BT) \to 0$), the vertical ordering of lines is mostly reversed and becomes dominated by chains with high enthalpies (small h's). Thus, the average enthalpy of the denatured ensemble is considerably lower at the transition midpoint than that when the denaturation is completed. By Eqs. (7) and (8), this means that the calorimetric two-state criterion is not satisfied.

This result is of experimental relevance. The layout of Figure 3 is motivated by similar plots of experimental hydrogen exchange data, because the HH sticking parameter $-\epsilon_{HH}/(k_BT)$ may be viewed as a crude model for denaturant concentration.[26,27] Analyses of this type will be used below to further elucidate the relation between the calorimetric two-state perspective and the multiple-state picture from recent native-state hydrogen exchange experiments.[51,53,54,57]

van't Hoff Enthalpies and Heat Capacity Functions: Gō and HP Models Are Not Calorimetrically Two-State

The basic calorimetric data are a heat capacity scan. To establish connections between theory and experiment, we now investigate how the functional forms of model densities of states affect the specific heat capacity and enthalpy as functions of temperature, especially at and around the transition midpoint. It is straightforward to show from Eqs. (5) and (29) that the transition excess heat capacity function $C_{P,ex}$ at the transition midpoint is given by

$$C_P(T_{1/2}) = \frac{\langle \Delta H(T_{1/2})\rangle_D^2}{4k_B T_{1/2}^2} + \frac{1}{2}(C_P)_D(T_{1/2}), \tag{15}$$

where the subscript "ex" has been dropped for notational simplicity, and

$$(C_P)_D = \frac{\partial \langle H(T)\rangle_D}{\partial T} \tag{16}$$

is the denatured heat capacity, i.e., the heat capacity that accounts for enthalpic transitions among the conformations in the denatured ensemble of all non-native conformations. Therefore, by definition Eq. (8), the population-based van't Hoff enthalpy

$$\Delta H_{vH} = 2T_{1/2}\sqrt{k_B[C_P(T_{1/2}) - (C_P)_D(T_{1/2})/2]}. \tag{17}$$

The upper panel of Figure 4 shows specific heat capacity functions for the HP sequence in Figure 2. The population midpoint temperature $T_{1/2}$ is very near, but not identical to the temperature T_{max} at the peak of the C_P function. Because the enthalpy distribution of the denatured ensemble of this sequence is wide, the denatured heat capacity $(C_P)_D$ is significant, contributing to more than a quarter of the peak of the overall heat capacity.*

Consistent with Figure 3, data in the lower panel of Figure 4 show that the average enthalpy of the denatured state at the transition midpoint is much lower than that when the ensemble is fully unfolded at high T. From this plot, $\Delta H_{vH}/\Delta H_{cal}$ is determined to be 0.43. Because $\Delta H_{vH}(T) = \langle \Delta H(T)\rangle_D$ is temperature dependent (dashed curve), if temperatures other than the population midpoint $T_{1/2}$ are used to mark the middle of the unfolding transition, different van't Hoff enthalpies would result. We consider two alternate transition midpoints, as noted above: (i) T_{max}, in which case $\Delta H_{vH}(T_{max})/\Delta H_{cal} = 0.46$ and (ii) the temperature at which $\langle \Delta H\rangle = \Delta H_{cal}/2$; in this case the van't Hoff to calorimetric enthalpy ratio becomes 0.58. Despite these variations, however, all three ratios are approximately 0.5 and significantly smaller than unity, indicating that this model HP sequence is quite far away from being calorimetrically two-state.

The HP model potential is nonspecific in that it contains only one type of favorable interaction. Can simple lattice

*It is useful for subsequent disucssion to note that the quantity in the square root in Eq. (17) is equal to the separation between the intersections of the vertical dashed line with C_P and $(C_P)_D$[D] in the upper panel of Figure 4.

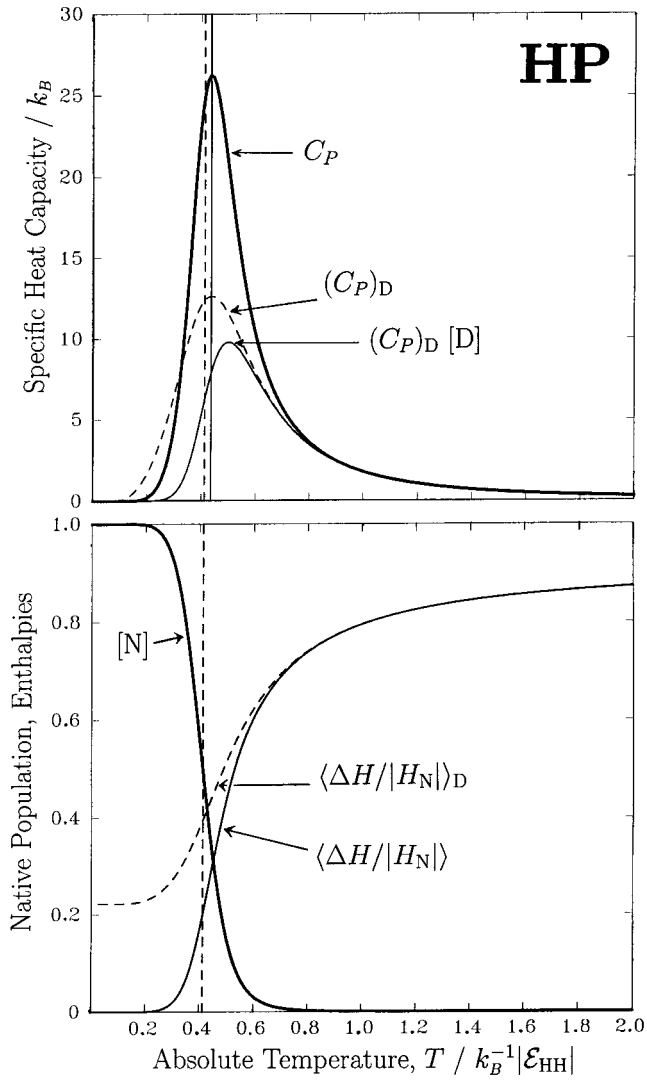

Fig. 4. Specific heat capacities and enthalpies of the HP sequence in Figure 2 versus temperature, $H_N = 9\epsilon_{HH}$. [N] and [D] are the fractional populations of the native and the denatured states, respectively ([N] + [D] = 1). The vertical solid line indicates the temperature T_{max} at which C_P is maximum, whereas the vertical dashed line indicates the temperature $T_{1/2}$ at which [N] = [D]. **Top:** C_P is for the entire conformational ensemble. $(C_P)_D$ is restricted to the denatured ensemble; whose contribution to the overall C_P is given by $(C_P)_D[D]$. **Bottom:** Both average enthalpies are relative to that of the native enthalpy, and normalized by $|H_N|$. $\langle \Delta H/|H_N| \rangle$ (solid curve) is for the entire ensemble, whereas $\langle \Delta H/|H_N| \rangle_D$ (dashed curve) is restricted to the denatured ensemble. $\Delta H_{vH}/\Delta H_{cal}$ is given by the ratio between the value of $\langle \Delta H/|H_N| \rangle_D$ at the transition midpoint (intersection of the vertical dashed line with the dashed curve) and the corresponding value at a sufficiently high temperature at which the transition is completed, i.e., when [N] $\approx$ 0; see Eq. (8).

models with more specific pairwise contact interactions produce much more cooperative folding by the calorimetric standard? We address this question by applying the Gō (Taketomi et al.[71]) and HP+ (Chan and Dill[27]) potentials to the same HP sequence. In these potentials, interaction specificity is enhanced by an ad hoc procedure. The favorable HH interaction and the neutral HP and PP interactions remain unchanged if a given contact is present

TABLE I. Exact Densities of States of the Sequence in Figure 2 Under Different Interaction Potentials are Obtained by Exhaustive Enumerations†

	$g(H)$		
H	HP	Gō	HP+
+10	0	0	154
+9	0	0	1,748
+8	0	0	7,193
+7	0	0	26,316
+6	0	0	72,662
+5	0	0	169,693
+4	0	0	366,828
+3	0	0	671,731
+2	0	0	1,095,375
+1	0	0	1,428,612
0	2,815,469	3,148,377	1,341,215
−1	2,100,897	2,052,228	518,168
−2	706,075	523,199	96,707
−3	156,218	75,242	10,607
−4	25,761	8,024	1,003
−5	3,530	1,110	266
−6	344	129	49
−7	40	25	7
−8	0	0	0
−9	1	1	1

†The $g(H)$ for the HP potential is from Dill et al.,[11] Enthalpy H is given in units of $|\epsilon_{HH}|$.

in the native structure, whereas interactions for all non-native contacts are assigned to be neutral (zero) for the Gō model and repulsive ($=|\epsilon_{HH}| > 0$) for the HP+ model. Table I gives their densities of states. Figure 5 shows that the introduction of these highly specific interaction schemes leads to higher native stabilities, the transition temperature $T_{1/2}$ (in units of $|\epsilon_{HH}|/k_B$) for the HP, Gō, and HP+ models are 0.41, 0.45, and 0.53, respectively. But the improvement in thermodynamic cooperativity is not significant. Whereas $\Delta H_{vH}/\Delta H_{cal}$ = 0.43 for the HP model, it equals 0.49 and 0.50 for the Gō and HP+ models, respectively. We also tested a number of popular three-dimensional lattice models of proteins. For every case tested, including a simple cubic lattice Gō model with 48 monomers, the model is found to fall short of the calorimetric two-state standard, although a three-dimensional Gō model appears to be more cooperative[58] than the two-dimensional one analyzed here. All these results suggest strongly that flexible chain models (with unique native conformations) that incorporate only pairwise additive contact energies are not capable of producing calorimetric two-state thermodynamics, for the simple reason that well-populated low-enthalpy non-native conformations cannot be avoided in these constructs.

Interpreting Experimental van't Hoff Enthalpies in Terms of Enthalpy Distributions Among Non-Native Conformations

In evaluating the thermodynamic cooperativity of protein models, heat capacity functions such as Figure 4 are direct links between theory and experiment, because van't

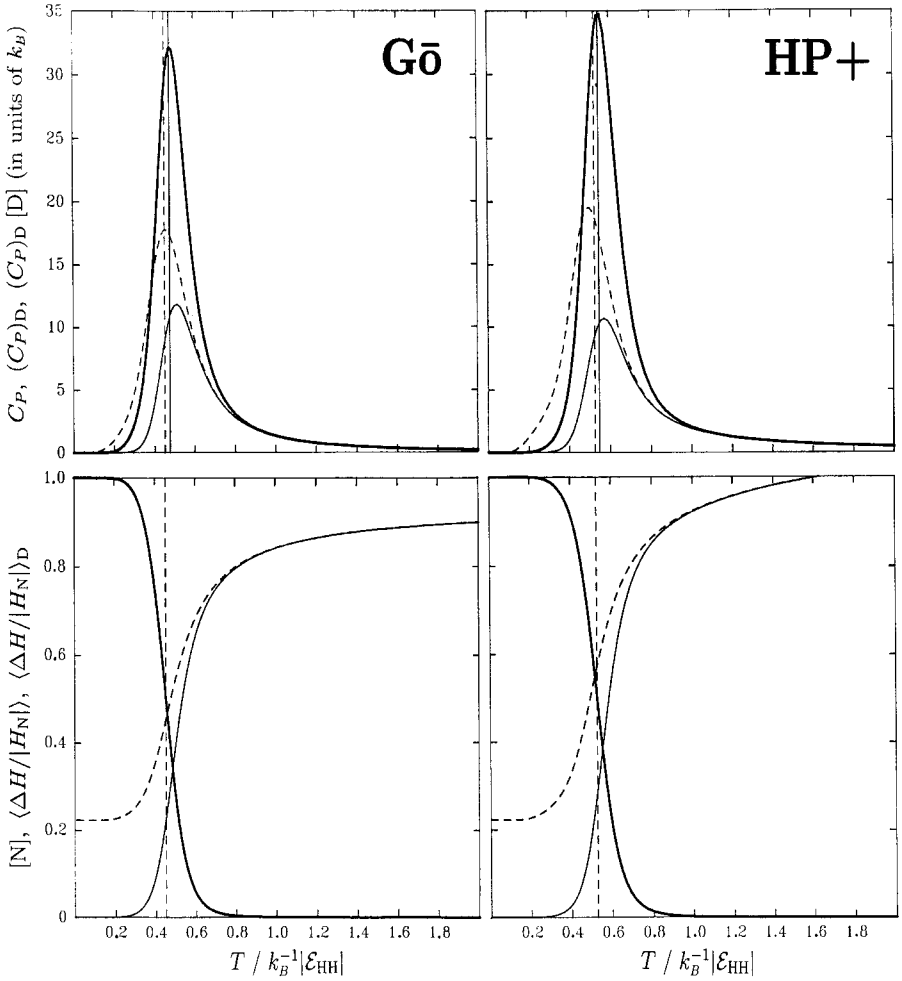

Fig. 5. Same as Figure 4, except the Gō (left column) and HP+ (right column) potentials are used instead of the HP potential for the sequence in Figure 2.

Hoff enthalpies of real proteins are extracted from heat capacity data. However, comparisons between theory and calorimetric experiment are sometimes not entirely straightforward, because a number of different formulas have been used to define experimental van't Hoff enthalpies. Here we discuss their physical meanings and their relations with ΔH_{vH}.

The van't Hoff enthalpy ΔH_{vH} used so far is based on the temperature dependence of the equilibrium constant K_{eq}, which is the population ratio between the native and denatured states.[16,35] Unlike other van't Hoff enthalpies (see below), the derivation of this population-based ΔH_{vH} and its physical interpretation in terms of Eq. (8) do not depend on whether the folding/unfolding transition is "two-state." A definition of van't Hoff enthalpy that does not depend on a "two-state" assumption offers conceptual clarity in theoretical analyses, because the calorimetric criterion for the term "two-state" is to be subsequently defined by the van't Hoff enthalpy itself. Moreover, by this definition, the relation $\Delta H_{vH} \leq \Delta H_{cal}$ can be shown in a totally straightforward manner for proteins that can be fully denatured by heat, involving only the simple condition that $K_{eq}(\infty)$ is practically zero [Eqs. (6) and (7)].

However, there is no experimental method to determine ΔH_{vH} directly from a heat capacity scan. Common formulas for calculating van't Hoff enthalpies from protein calorimetric data are derived by first assuming that the transition is "two-state," meaning that the enthalpy distribution is assumed to consist of only two sharp peaks. The calorimetric two-state criterion based on this type of van't Hoff enthalpies is essentially a self-consistency argument. The idea is that if such a "model-dependent" quantity (computed by assuming "two-state") is inconsistent with the actual, "model-independent" calorimetric enthalpy, it would imply that the transition is really not two-state. Conversely, if the two quantities are consistent, one may infer that the transition is indeed two-state. We now examine the relation between these experimentally derived van't Hoff enthaplies and the population-based van't Hoff enthalpy ΔH_{vH}.

One formula that has been used for calculating a van't Hoff enthalpy from the calorimetrically determined peak value $C_{P,max}$ of the transition excess specific heat function after baseline subtractions is the following[16]:

$$\Delta H_{vH}^{exp} = 2T_{max}\sqrt{k_B C_{P,max}}. \quad (18)$$

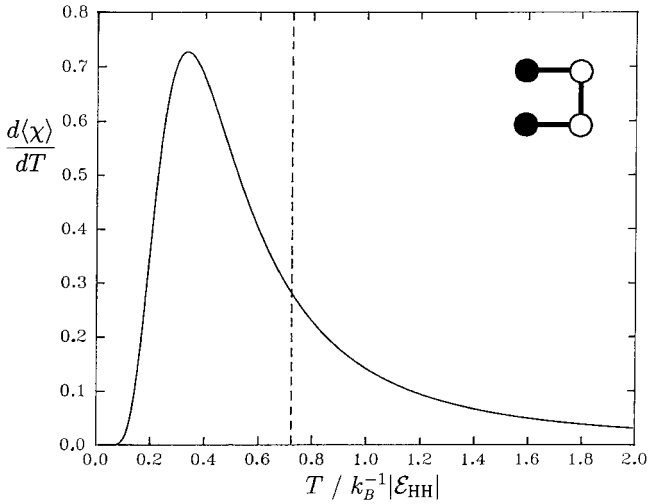

Fig. 6. Heat denaturation of an HP sequence with four monomers, shown in its native structure, and monitored by the temperature derivative of the average of an overlap function χ in Ref. 5. The transition midpoint $T_{1/2}$ is indicated by the vertical dashed line. For this model, $d\langle\chi\rangle/(dT) = \langle\chi\rangle_D C_V/|\epsilon_{HH}|$, where C_V is the heat capacity of the "tetramer toy model" in Ref. 11. $T_{1/2}$ is significantly different from T_{max} for this short chain.

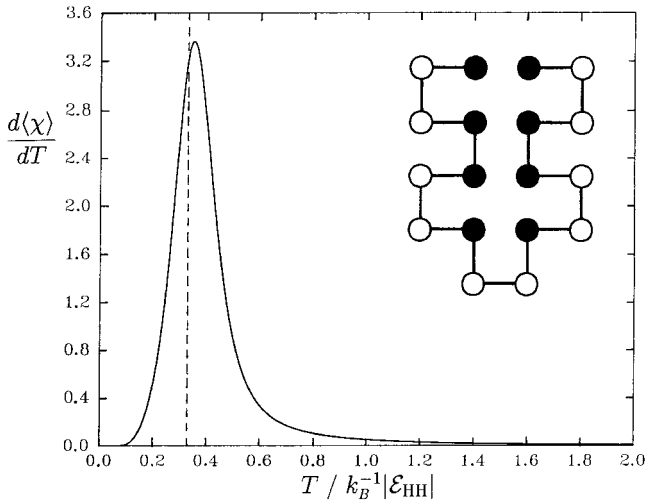

Fig. 7. Same as Figure 6, but for an HP sequence with 18 monomers. The sequence is shown in its native structure, its density of states is given in the HP column of Table II.

This expression differs from Eq. (17) in two respects. (i) T_{max} is used instead of $T_{1/2}$. In applications to real proteins, the offset between the two temperatures is sometimes taken to be not significant; alternatively, Eq. (18) may be evaluated at $T_{1/2}$ or some other midpoint temperature[15,16] instead of T_{max}. In all lattice models studied here, the offsets between the two temperatures are not big (Fig. 4), except for the very short chain in Figure 6. (ii) The term $(C_P)_D(T_{1/2})/2$ in Eq. (17) is absent, which implies that $\Delta H_{vH}^{exp} \neq \Delta H_{vH}$ in general. If $T_{max} \approx T_{1/2}$, the two enthalpies are approximately equal when $(C_P)_D(T_{1/2})/2 \approx 0$, i.e., when the transition is calorimetrically two-state, in which case $(C_P)_D[D] \approx 0$ for all temperatures.† For the HP, Gō, and HP+ models above, $\Delta H_{vH}^{exp}/\Delta H_{cal}$ equals 0.54, 0.64, and 0.63, respectively.

In general, if the protein density of states is assumed to have a form similar to that of the Gaussian models above, with $\sigma_H < H_D/\sqrt{2 \ln g_D}$ so that the native-state population is practically zero at high temperatures [$K_{eq}(\infty) = 0$], $\Delta H_{vH}^{exp} \leq \Delta H_{cal}$ is expected, and $\Delta H_{vH}^{exp}/\Delta H_{cal} \approx 1$ implies $\Delta H_{vH}/\Delta H_{cal} \approx 1$. This follows from the semiquantitative argument below, using the fact that $\sqrt{k_B T^2 C_P}$ is the standard deviation of the underlying enthalpy distribution [Eq. (29)], and that by Eq. (18) this standard deviation is capable of attaining a maximum $\approx \Delta H_{cal}/2$ if $\Delta H_{vH}^{exp}/\Delta H_{cal} \approx 1$.

For $K_{eq}(\infty) = 0$, we expect from the above analysis of the Gaussian model that the standard deviation of enthalpy among the denatured conformations to be smaller, or even considerably smaller than $\Delta H_{cal}/2$. Therefore, a standard deviation of enthalpy $\approx \Delta H_{cal}/2$ cannot be achieved without significant involvement of the native state. It can only be achieved when there are substantial populations in both the native and the denatured states, i.e., near the transition midpoint. At those temperatures, because the average enthalpy separation between native and denatured conformations is bounded from above by ΔH_{cal}, the maximum possible standard deviation is $\Delta H_{cal}/2$, which is possible only if the denatured enthalpy distribution is narrow; in other words, $\Delta H_{vH}/\Delta H_{cal} \approx 1$. Otherwise, enthalpies closer than $\Delta H_{cal}/2$ to the mean excess enthalpy $\langle\Delta H\rangle$ around the midpoint temperature would contribute substantially and reduce the standard deviation. Hence, $\sqrt{k_B T^2 C_P}$ can achieve a maximum $\approx \Delta H_{cal}/2$ only if $\Delta H_{vH}/\Delta H_{cal} \approx 1$. Moreover, when this condition is satisfied, the peak temperature of $\sqrt{k_B T^2 C_P}$ is expected to be very close to the peak temperature T_{max} of C_P. It follows that $\Delta H_{vH}^{exp}/\Delta H_{cal} \approx 1$ and $\Delta H_{vH}/\Delta H_{cal} \approx 1$ are essentially equivalent criteria for calorimetrically two-state transitions. On the other hand, when the system is far from calorimetrically two-state, $\Delta H_{vH}^{exp} > \Delta H_{vH}$ is expected from comparing Eqs. (17) and (18) and assuming $T_{1/2} \approx T_{max}$, although ΔH_{vH}^{exp} is likely to correlate with ΔH_{vH}. These inferences are valid for the models considered in this article. However, in the absence of any assumption on the form of the density of states, $\sqrt{k_B T^2 C_P}$ at some temperatures can be larger than $\Delta H_{cal}/2$ if $K_{eq}(\infty)$ is

†Physically, this means that the two-state derivation of ΔH_{vH}^{exp} in Eq. (18) neglects contributions to $C_{P,max}$ from the enthalpy distribution (after baseline subtractions) among different denatured conformations. The term $(C_P)_D(T_{1/2})/2$ in Eq. (17) is formally similar to the quantity $\Delta_0^1 C_P/2$ in the analysis of Privalov and Potekhin[16] [their Eq. (37)]. However, their $\Delta_0^1 C_P/2$ is assumed to be well described by a sigmoidal or similar baseline function interpolating between the heat capacities of the native and the denatured states and, therefore, subtractable by such a baseline function so that its contribution is absent in the final transition excess heat capacity[15] (see also the baseline function $\langle\Delta C_{P,bl}\rangle$ defined by Freire[19]). In other words, the $\Delta_0^1 C_P/2$ subtracted in conventional treatments is presumed to arise mainly from the difference in hydration effects of the two (native and denatured) states[72–74] and involve no substantial contribution from protein conformational transitions within the denatured ensemble. Therefore, it is expected that any significant deviation of ΔH_{vH}^{exp} from ΔH_{cal} would indicate that this assumption is wrong, that the denatured enthalpy distribution is relatively wide, and hence, the system is not really two-state.

significantly larger than zero, as was shown[35] recently.[‡] In subsequent discussions, we will continue to use $\Delta H_{\rm vH}$ for its conceptual clarity, but to make more direct connections with experimental data, $\Delta H_{\rm vH}^{\rm exp}$ values will also be provided. For all the protein models considered in this article, $\Delta H_{\rm vH}^{\rm exp}/\Delta H_{\rm cal} < 1$, and the general conclusions are not sensitive to whether $\Delta H_{\rm vH}$ or $\Delta H_{\rm vH}^{\rm exp}$ is used to define calorimetric two-state cooperativity. A more detailed comparison of these and other common definitions of van't Hoff enthalpy in protein calorimetry is provided elsewhere.[58]

The Calorimetric Two-State Criterion Is Different From the Free-Energy-Barrier and Sharpness Criteria for Cooperativity

Besides the calorimetric condition, other two-state criteria have been proposed. For instance, the existence of a less populated range of energies (taken here as enthalpies)[11,24,35] or equivalently a free energy barrier,[33] between the native and the bulk of the denatured ensemble have been used to define two-state transitions in proteins. This criterion, which may be casted as the mathematical condition[11] $d^2{\rm g}(H)/(dH^2)|_{H=H_{\rm N}} > 0$, is often less restrictive on the density of states than the $\Delta H_{\rm vH}/\Delta H_{\rm cal} \approx 1$ criterion, as discussed above (Fig. 2). Although a free energy barrier between the native state and denatured ensemble is necessary and essential, we conclude from the above discussion on the HP, Gō, and HP+ models that a free energy barrier per se is insufficient for calorimetric two-state cooperativity.

The term "cooperativity" is sometimes used to describe the sharpness of a transition. Is there a simple relationship between the sharpness of a transition and its calorimetric cooperativity? In a recent study on the role of sidechains in protein thermodynamic cooperativity, Klimov and Thirumalai[5] use an overlap function[75] χ to monitor the folding process. For any given conformation, χ is defined as

$$\chi = 1 - \frac{2}{(n-1)(n-2)} \sum_{i<j-1} \Delta(r_{ij} - r_{{\rm N},ij}), \qquad (19)$$

where n is the number of monomers along the chain, r_{ij} and $r_{{\rm N},ij}$ are the spatial separation between the ith and the jth monomers in the given and the native conformation, respectively; $\Delta(0) = 1$, and $\Delta(x) = 0$ for $x \neq 0$. Hence $\chi = 0$ for the native conformation. They then introduce a dimensionless cooperativity index Ω_c in terms of the Boltzmann average $\langle\chi\rangle$ of the overlap function over the entire conformational ensemble,

$$\Omega_c = \frac{T^2_{{\rm max},\chi}}{\Delta T} \max\left[\frac{d\langle\chi\rangle}{dT}\right], \qquad (20)$$

where ΔT is the full width at the half maximum of the peak of $d\langle\chi\rangle/(dT)$ and $T_{{\rm max},\chi}$ is the temperature at which $d\langle\chi\rangle/(dT)$ equals its maximum value $\max[d\langle\chi\rangle/(dT)]$ (Klimov and Thirumalai[5]).

We find that although Ω_c is very useful in characterizing the sharpness of a transition by how narrow the temperature range in which the bulk of the transition takes place, whether a transition is calorimetrically two-state does not necessarily correlate with the value of Ω_c. Figure 6 shows a toy model that is exactly two-state, yet its $\Omega_c = 0.183$ is small.[§] Because $K_{\rm eq}(\infty) = 1/4$ for this toy model, $\Delta H_{\rm vH}/\Delta H_{\rm cal} = 5/4$ [Eq. (6)]. If the experimental prescription Eq. (18) is used, $\Delta H_{\rm vH}^{\rm exp}/\Delta H_{\rm cal} = 0.92$. On the other hand, Figure 7 shows a longer HP model sequence that is quite far away from being calorimetrically two-state, because its $\Delta H_{\rm vH}/\Delta H_{\rm cal} = 0.31$, $\Delta H_{\rm vH}^{\rm exp}/\Delta H_{\rm cal} = 0.48$, (Fig. 8); yet its $\Omega_c = 2.27$ is more than 12 times larger than the strictly two-state sequence in Figure 6. In general, the sharpness of a transition as characterized by Ω_c increases with conformational entropy of the denatured ensemble. An increasing trend for Ω_c with increasing chain length has also been observed in Klimov and Thirumalai.[5]

This point is best illustrated by the experimental study of Tiktopulo et al.[76] on long homopolymers poly(*N*-isopropylacrylamide) with approximately 62,000 monomers per chain. The full width covering the entire C_P peak that signals a temperature-induced globule to coil transition is reasonably narrow, ~10°C, which is comparable with or even sharper than some small single-domain proteins[16] (see also Fig. 1). Yet this homopolymer transition is not calorimetrically two-state. Applying the formula $\Delta H_{\rm vH}^{\rm exp(a)} = 4k_BT^2_{\rm max}C_{P,{\rm max}}/\Delta H_{\rm cal}$, Tiktopulo et al. reported that $\Delta H_{\rm vH}^{\rm exp(a)}/\Delta H_{\rm cal} < 0.01$. Generally speaking, the sharpness of a transition tends to increase with the size of the system, whereas the calorimetric two-state condition concerns the form of the enthalpic distribution. Therefore, calorimetric cooperativity does not correlate with sharpness per se. Fundamentally, inasmuch as

[‡] A common alternate formula for an experimental van't Hoff enthalpy is $\Delta H_{\rm vH}^{\rm exp(a)} = 4k_BT^2_{\rm max}C_{P,{\rm max}}/\Delta H_{\rm cal}$ (Refs. 16 and 19), where the superscript "(a)" is used here for distinguishing it from Eq. (18). Hence, $\Delta H_{\rm vH}^{\rm exp(a)}/\Delta H_{\rm cal} = (\Delta H_{\rm vH}^{\rm exp}/\Delta H_{\rm cal})^2$. If $\Delta H_{\rm vH}^{\rm exp} H_{\rm cal} \leq 1$, $\Delta H_{\rm vH}^{\rm exp(a)} \leq \Delta H_{\rm vH}^{\rm exp} \leq \Delta H_{\rm cal}$. In that case $\Delta H_{\rm vH}^{\rm exp(a)}/\Delta H_{\rm cal} \approx 1$ is essentially equivalent to the criterion for calorimetrically two-state transitions using either $\Delta H_{\rm vH}$ or $\Delta H_{\rm vH}^{\rm exp}$. In some studies, another van't Hoff enthalpy, $\Delta H_{\rm vH}^{\rm exp(a)\prime}$, is defined as the quantity $4k_BT^2C_P/\Delta H_{\rm cal}$ evaluated at the temperature T when $\langle\Delta H(T)\rangle = \Delta H_{\rm cal}/2$ (e.g., Ref. 35). For chain models that are far from being calorimetrically two-state, this temperature can be significantly different from $T_{1/2}$ and $T_{\rm max}$. Zhou et al.[35] showed that it is possible for $\Delta H_{\rm vH}^{\rm exp(a)\prime}/\Delta H_{\rm cal}$ to be greater than one. For example, their $\lambda = 1.8$ square-well tetramer, with $K_{\rm eq}(\infty) = 0.305$, gives $\Delta H_{\rm vH}^{\rm exp(a)\prime}/\Delta H_{\rm cal} = 1.44$. (This is equal to the reciprocal of their γ.) Using the density of states they provided, we find that $\Delta H_{\rm vH}/\Delta H_{\rm cal} = 1.34$ for this system of theirs. It is of interest that the corresponding van't Hoff enthalpies calculated at the peak of the specific heat function are much smaller: $\Delta H_{\rm vH}^{\rm exp}/\Delta H_{\rm cal} = 0.73$ and $\Delta H_{\rm vH}^{\rm exp(a)}/\Delta H_{\rm cal} = 0.53$. For all lattice protein models considered in this article, $\Delta H_{\rm vH}^{\rm exp(a)\prime} < \Delta H_{\rm vH}$ except our tetramer toy model in Figure 6, for which $\Delta H_{\rm vH}^{\rm exp(a)\prime}/\Delta H_{\rm cal} = 3/2$. For the above HP, Gō, and HP+ models, $\Delta H_{\rm vH}^{\rm exp(a)\prime}/\Delta H_{\rm cal} = 0.29$, 0.43, and 0.37, respectively. Because the temperature at which $\langle\Delta H\rangle = \Delta H_{\rm cal}/2$ is expected to be very close to $T_{1/2} \approx T_{\rm max}$ for two-state ($\Delta H_{\rm vH}/\Delta H_{\rm cal} \approx 1$) systems with $K_{\rm eq}(\infty) = 0$, we infer that $\Delta H_{\rm vH}^{\rm exp(a)\prime}/\Delta H_{\rm cal} \approx 1$ is essentially equivalent to the other calorimetric two-state criteria considered in this section.

[§] In Ref. 5, $\Omega_c = (\Delta G_{\rm H_2O})^2\{4k_B^2T^2_{{\rm max},\chi}\ln[(3+2\sqrt{2})/(3-2\sqrt{2})]\}^{-1}$ is derived for a strictly two-state system to relate stability to Ω_c [Eq. (15) in Ref. 5]. Apparently, two assumptions were used to obtain this formula: $\langle\chi\rangle_{\rm D} = 1$, and $T_{{\rm max},\chi} \approx T_{1/2}$; the latter is expected to be valid for sufficiently long chains. However, because temperature derivatives affect only the enthalpic but not the entropic part of $\Delta G_{\rm H_2O}/(k_BT)$, the free energy of folding $\Delta G_{\rm H_2O}$ in this formula should be replaced by its enthalpic part $\Delta H_{\rm H_2O} = -T^2\partial(\Delta G_{\rm H_2O}/T)/(\partial T)$.

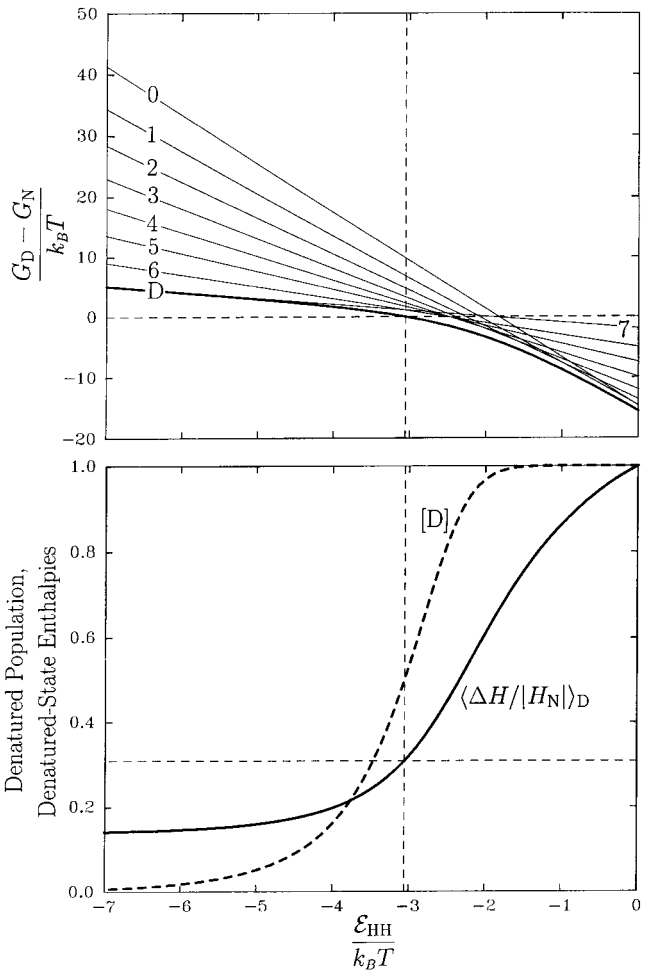

Fig. 8. **Top:** Same as Figure 3, but for the HP sequence in Figure 7. **Bottom:** [D] is the fractional denatured population (dashed curve). The average denatured enthalpy (solid curve) is low (horizontal dashed line) at the transition midpoint $T_{1/2}$ (vertical dashed line, when [D] = 1/2) compared to that when unfolding is completed at $\epsilon_{HH}/(k_B T) = 0$. $\Delta H_{vH}/\Delta H_{cal} = 0.31$ for this sequence.

$K_{eq}(\infty) = 0$, the calorimentric criterion $\Delta H_{vH}/\Delta H_{cal} \approx 1$ is very similar to the free-energy-barrier criterion. In fact, it may be viewed as a more restrictive form of the latter because it requires the width of the enthalpy distribution in the denatured ensemble to be narrow vis-à-vis the enthalpic width of the free energy barrier.

Restricting Model Denatured States Can Lead to Calorimetric Two-State Cooperativity: Effects of Assuming a Form of Capillarity Cooperativity

Our analysis so far has suggested strongly that additive pairwise contact-type interactions are in general insufficient for calorimetric two-state cooperativity when they are incorporated into models of flexible polymers. However, there is an apparent exception. Freire and coworkers applied a model of partially folded conformations to investigate the molecular basis of cooperativity in protein folding.[60,77] A large body of insightful studies were performed, including predictions on the properties of partially folded states and hydrogen exchange protection factors (see, e.g., Hilser and Freire,[39] Freire et al.,[61,78] and Xie and Freire[79]). It has also provided predictions that are in agreement with calorimetric data on protein heat denaturation.[61] The energetics of their model is based essentially on contact-type interactions, because the contribution from each interaction type varies linearly with a part of the model protein molecule's exposed solvent accessible surface area.[39] Therefore, to gain further insight into whether additive contact energies are sufficient for calorimetric two-state behaviors, it is crucial to ascertain the features in Freire and coworkers' model that underlie its success in reproducing proteinlike thermodynamic cooperativity.

The denatured ensemble in the model of Freire and coworkers includes both fully unfolded and "partially folded" conformations. However, not all possible conformations of a polypeptide chain were considered. Instead, their denatured conformations were generated by "selection rules"[39] that first divide the chain into a number of contiguous segments, and then *assume* that each segment behaves as a "cooperative folding unit" in that they can only be either fully folded (i.e., adopt the segment's conformation in the native state) or completely unfolded.[60] Although this construction is highly insightful, and may well capture an essential aspect of protein thermodynamic cooperativity, it is nonetheless an *additional assumption* on conformational distribution that does not follow from the contact-based interaction scheme. In this sense, this computational scheme is not a completely self-contained heteropolymer model. In this connection, it is interesting to note that the assumed existence of relatively sharp interfaces between folded and unfolded parts of a protein is conceptually very similar to recent capillarity theories of protein folding[80–82] and analytical models that use capillarity cooperativity.[83,84] Related ideas have also been applied to other models of protein folding kinetics.[85]

To delineate the ramification of the assumptions in the construction of denatured ensembles in the model of Freire and coworkers and to explore its effects when combined with a contact-based potential on the density of states, we now consider a highly simplifed version of their model. We do so by introducing rules similar to theirs for constructing denatured ensembles, applying them to the HP sequence in Figure 7 while retaining the original additive HP contact potential. Following the recent approach of Hilser and Freire,[39] who divide the chain into equal-length segments (except at the ends) independent of secondary structure definitions, we divide our $n = 18$ chains into m segments. We only consider cases in which n is divisible by m; so the length of each segment $l = n/m$. Each segment is either fully folded or fully unfolded. Monomers in these segments are referred to as folded or unfolded, respectively. We call these theoretical constructs "lattice capillarity models" (Fig. 9). When a segment is designated to be folded, all of its internal bond angles are set to their values in the native conformation. More specifically, if the monomers of a segment are numbered $i, i + 1, \ldots,$ and $i + l - 1$, the bond angles at monomers $i + 1, i + 2, \ldots,$ and

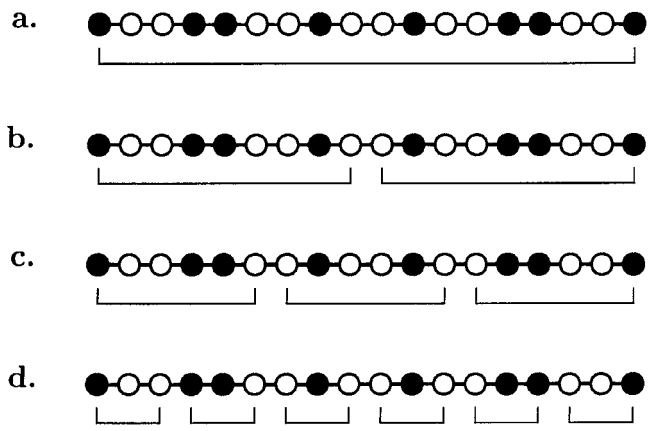

Fig. 9. Lattice capillarity models studied in this article. The braces indicate four different ways in which segments of the HP sequence in Figure 7 are divided into equal-size "cooperative folding units" that are assumed to be either fully folded or fully unfolded. a, b, c, and d correspond to $m = 1$, 2, 3, and 6, with $l = 18$, 9, 6, and 3, respectively.

$i + l - 2$ are fixed. It follows that there are a total of 2^m conformational "states." One is native (when all segments are folded), one is fully unfolded (when all segments are unfolded). In accordance with the terminology of Freire and Murphy,[60] we refer to the other $2^m - 2$ conformational states as "partially folded."

In the recent approach of Freire and coworkers, conformational entropies are determined by summing up contributions from different amino acid residues in unfolded segments, assuming that the segments are in random-coil-like conformations.[39,86] For the simple lattice capillarity models studied here, we introduce the following rules to achieve similar effects: (i) All contacts involving monomers in unfolded segments are forbidden. These include contacts between unfolded monomers and contacts between unfolded and folded monomers. (ii) If two folded segments are not contiguous (i.e., when they are separated by one or more unfolded segments along the chain), favorable interactions (HH contacts) between them are forbidden. This means that all conformations in a partially folded state have the same enthalpy. (iii) All contacts are forbidden in the ensemble of conformations with zero enthalpy. In other words, when a conformation contains no favorable contact, neutral contacts are also forbidden.

Densities of states for the lattice capillarity models in Figure 9 are determined by exhaustively enumerating all self-avoiding conformations that are consistent with these rules (Table II, columns a–d). To explore the robustness of these capillarity assumptions, we consider also a slightly different scheme for capillarity cooperativity that retains only rule (i) but discards rules (ii) and (iii). This causes changes in the densities of states of models (c) and (d) (see columns c′ and d′ in Table II) but does not affect models a and (b). Note that the $H = 0$ states for models (d) and (d′) have more conformations than the original HP model because they contain not only contributions from the fully unfolded state but also some of the partially folded states. This is possible because "folded" $l = 3$ segments have no internal contacts.

Figure 10 gives the stabilities of capillarity models with different segment lengths. Figure 11 shows their average denatured enthalpy functions and indicates their values at the transition midpoint where ΔH_{vH}'s are calculated. The corresponding $\Delta H_{vH}/\Delta H_{cal}$ ratios are given in row (i) of Table II. The capillarity model (a) is the trivial case that assumes that the entire chain can only be either fully folded or fully unfolded. Hence, it is calorimetrically two-state by default. The remarkable observation here, however, is that even when the segment length is systematically reduced to the shortest $l = 3$ (model d), the thermodynamic behavior is still very close to satisfying the calorimetric two-state criterion, with $\Delta H_{vH}/\Delta H_{cal} > 0.93$. This behavior contrasts sharply with that of the original HP model (Fig. 8).

Figure 10 illustrates the difference between the calorimetrically two-state-like capillarity models and the HP models (Figs. 3 and 8) for the stabilities of their denatured ensembles at different enthalpy levels. For the capillarity models (b–d) in Figure 10, the denatured ensembles are dominated by high-enthalpy conformations (with no or just one HH contact) at the transition midpoint rather than dominated by low-enthalpy conformations as in Figures 3 and 8. In other words, for the models in Figure 10, when hydrophobic sticking is gradually weakened $[-\epsilon_{HH}/(k_BT) \rightarrow 0]$, the ordering of the stabilities for different enthalpy levels are reversed before the transition midpoint. On the other hand, for the models in Figures 3 and 8 that are not calorimetrically two-state, this reversal does not occur until after the transition midpoint. As has been emphasized by Englander and coworkers, for calorimetrically two-state proteins, states that are more stable than the fully unfolded state under native conditions are overtaken ("crossed") by the fully unfolded state before the global denaturation transition.[51,57] To make further connections with experiments, Figure 12 applies a simple lattice modeling approach similar to that pioneered by Miller and Dill[87] to a lattice capillarity model and predicts general features that are quite similar to those observed in recent native-state hydrogen exchange experiments.[53,54]

These results lead us to conclude that the seemingly not drastic assumption of dividing a chain into fully folded and unfolded segments can produce a dramatic effect on thermodynamic cooperativity, even when the segments are relatively short. This is because this procedure places highly consequential restrictions on conformational freedom, by disallowing many low-enthalpy conformations (with more favorable contact enthalpies) that were hitherto allowed in a fully flexible polymer model with additive contact energies. This feature has been discussed by Freire and coworkers using a complementary line of reasoning. They emphasize that the energetic unfavorability of partially folded states in their model is due to the existence of "uncompensated exposure of apolar groups" belonging to the folded parts of the protein, and therefore "the unfolded state exposed to the solvent what is unfolded, whereas partially folded states necessarily expose more than what is unfolded."[77] It is interesting to note that capillarity models (c′) and (d′) with relaxed restrictions on interac-

TABLE II. Densities of States of HP and Lattice Capillarity Models†

	$g(H)$						
H	a	b	c	d	c′	d′	HP
0	646,909	646,909	646,909	3,095,714	646,909	3,742,623	2,479,880
−1	0	0	72,302	207,642	72,302	216,434	2,295,298
−2	0	4,272	438	11,278	438	12,067	840,490
−3	0	0	284	448	288	474	169,311
−4	0	0	0	0	0	15	21,427
−5	0	0	0	0	1	3	1,768
−6	0	0	0	0	0	2	153
−7	0	0	0	0	0	0	7
−8	1	1	1	1	1	1	1
(i) $\Delta H_{vH}/\Delta H_{cal}$	1.00	0.961	0.946	0.937	0.942	0.905	0.308
(ii) $\Delta H_{vH}/\langle\Delta H(T_{0.999})\rangle$	1.00	0.968	0.961	0.955	0.958	0.921	0.386
$\Delta H_{vH}/\langle\Delta H(T_{0.98})\rangle$	1.02	0.995	0.992	0.988	0.989	0.954	0.489
(iii) $\Delta H_{vH}/\Delta H_{vH}(T_{0.999})$	1	0.967	0.960	0.954	0.957	0.921	0.386
$\Delta H_{vH}/\Delta H_{vH}(T_{0.98})$	1	0.975	0.972	0.968	0.969	0.935	0.479

†Columns **a–d** are exact densities of states of the sequence in Figure 7 enumerated by using the four capillarity models **a–d** in Figure 9, in which favorable interactions between folded segments separated by one or more unfolded segments are forbidden, whereas columns **c′** and **d′** correspond, respectively, to the capillarity models **c** and **d** in which such interactions are allowed. The **HP** column is the exact density of states using the original HP potential. Enthalpy H is given in units of $|\epsilon_{HH}|$. ΔH_{cal} in row (i) is evaluated at $T \to \infty$ [Eq. (1)]. $T_{0.999}$ and $T_{0.98}$ in the last four rows are the temperatures at which the fractional denatured population equals 99.9% and 98%, respectively. The data in the two rows labeled by (ii) illustrate the effect on the van't Hoff to calorimetric enthalpy ratio of approximate calorimetric enthalpies obtained by integrating the specific heat function to a finite temperature is used instead of the ΔH_{cal} defined by Eq. (1) [i.e., substituting T_1 in Eq. (2) by $T_{0.999}$ and $T_{0.98}$]. The two rows labeled by (iii) underscore the temperature dependence of $\Delta H_{vH}(T)$ [see Eqs. (4) and (8)]; $\Delta H_{vH}(T_{[D]}) = \langle\Delta H(T_{[D]})\rangle/[D]$. Rows (ii) and (iii) highlight possible discrepancies between the theoretical $\Delta H_{vH}/\Delta H_{cal}$ that requires integrating the specific heat function to infinite temperature [Eqs. (1) and (30)] and $\Delta H_{vH}/\Delta H_{cal}$'s obtained by experimental or numerical techniques that integrate over finite temperature ranges.

tions between folded segments remain highly cooperative. Despite the decrease in their energy gaps (i.e., the difference in energy or enthalpy between the native and the lowest occupied state, see Šali et al.[88,89] and Chan[90]) relative to models (c) and (d) (from $5|\epsilon_{HH}|$ to $3|\epsilon_{HH}|$ and $2|\epsilon_{HH}|$, respectively), their $\Delta H_{vH}/\Delta H_{cal}$'s are reduced only

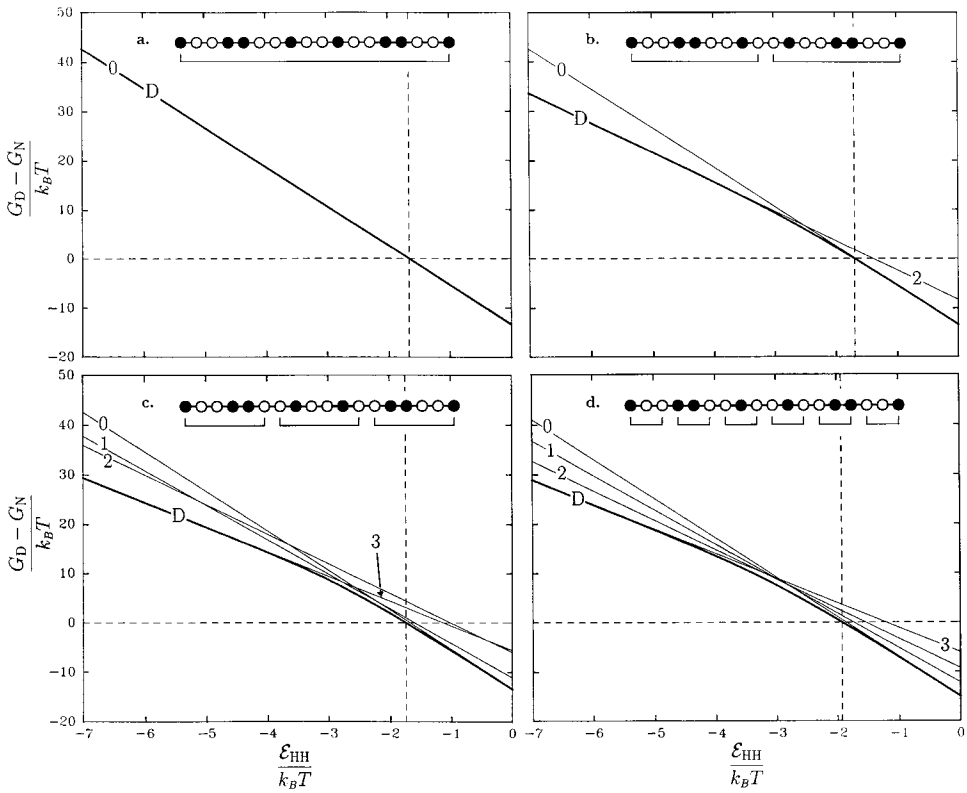

Fig. 10. Stabilities of denatured conformations in the four lattice capillarity models. Notations are the same as the upper panel of Figure 8.

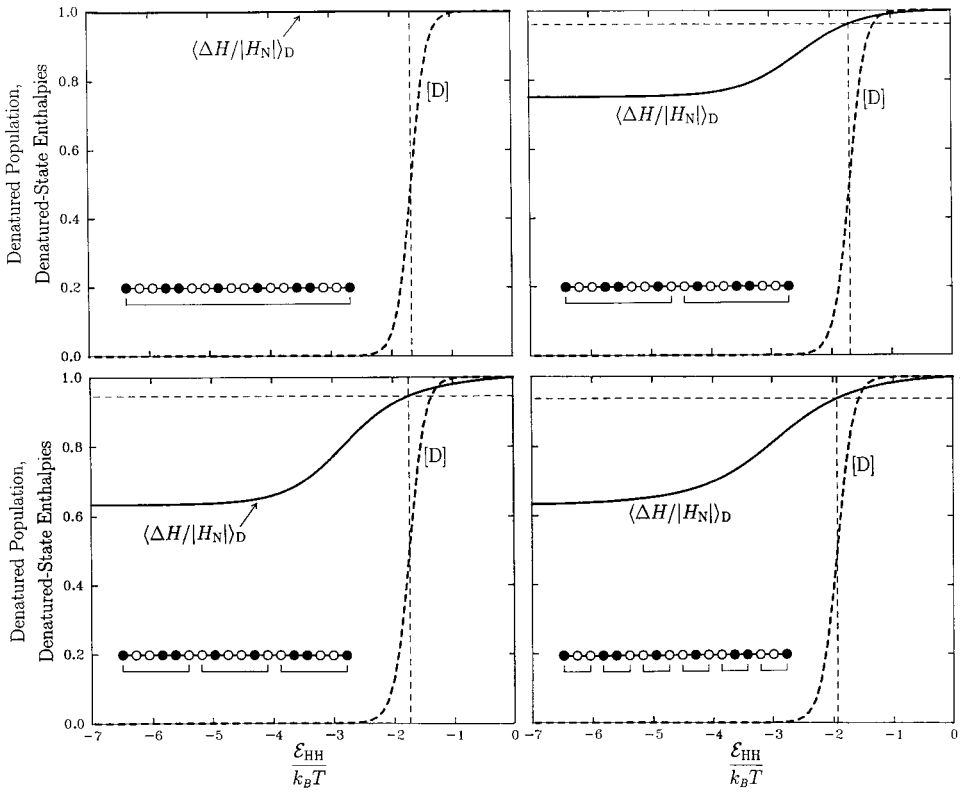

Fig. 11. Denatured populations and average denatured enthalpies of the four lattice capillarity models. Notations are the same as the lower panel of Fig. 8. $\Delta H_{vH}/\Delta H_{cal}$'s are given in Table II. $\Delta H_{vH}^{exp}/\Delta H_{cal}$ for a–d are 0.99, 0.94, 0.93, and 0.94, respectively.

slightly, from 0.946 and 0.937 to 0.942 and 0.905, respectively. (The corresponding $\Delta H_{vH}^{exp}/\Delta H_{cal}$'s change from 0.928 and 0.944 to 0.927 and 0.918.) It is not surprising that the shape of the entire energy/enthalpy distribution is much more important for understanding protein behaviors than the energy gap per se.[9,11]

Hence, the extensive investigation of Freire and coworkers and the present simple lattice capillarity models establish that in principle capillarity cooperativity can lead to global calorimetric two-state cooperativity and at the same time provide a physical microscopic multiple-state picture consistent with recent hydrogen exchange experiments.[39] Capillarity models[80,81] can provide deep physical insight. This perspective is also consistent with our finding that contact interactions alone are most likely insufficient to produce proteinlike cooperativity. The natural next step is to address the physical basis of various capillarity assumptions and to explore plausible physical reasons why proteins exhibit thermodynamic behaviors that are either consistent or similar to that predicted by capillarity cooperativity.

Plausible Energetic Ingredients for Calorimetric Two-State Cooperativity: A Scenario

Following the above line or reasoning, we now explore a class of models with plausible energetic ingredients that may enhance thermodynamic cooperativity. These models are constructed by adding new interactions to the original HP sequence in Figure 7, and the resulting densities of states are then determined by exact enumeration. The goal

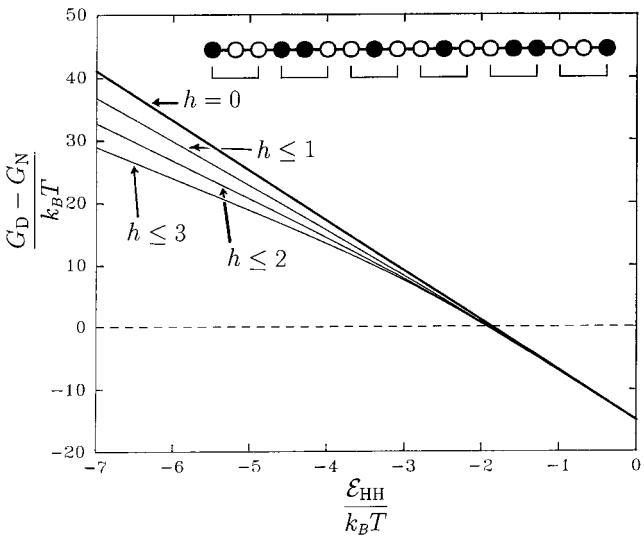

Fig. 12. Stabilities of denatured ensembles for the lattice capillarity model (d) in Figure 9. The line labeled by $h = 0$ is for all conformations with no HH contact. Curves represent collections of conformations with multiple numbers of HH contacts. For example, the curve labeled by $h \leq 3$ are for all conformations with 0, 1, 2, or 3 HH contacts. These curves are crude models for recent hydrogen exchange experiment results (see, e.g., Llinás et al.[53] and Milne et al.[54]). In this picture, if a certain monomer (residue) along the chain is exposed to the solvent after a certain threshold number $(h_N - h_i)$ of HH contacts are lost ($h_N = 8$ is the native number of HH contacts for this sequence), then the apparent stability of this monomer would be equal to the stability of the conformational ensemble with h_i or fewer HH contacts, as given by the $h \leq h_i$ curve.

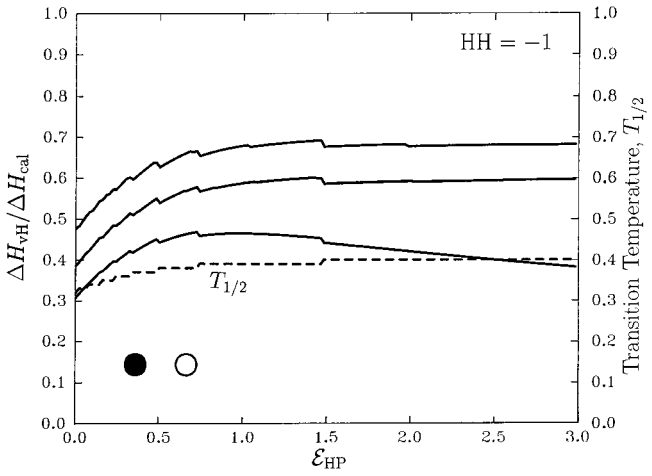

Fig. 13. Systematically modifying the HP sequence in Figure 7. The drawing at the lower left depicts the type of interaction being varied. In this plot, the HH contact enthalpy ϵ_{HH} is fixed at -1 (in units of $|\epsilon_{HH}|$, which is set to k_B here and in subsequent figures), whereas the HP enthalpy ϵ_{HP} (in the same unit) is modified to produce a range of different models. Within the range of ϵ_{HP} shown, thermodynamic cooperativity properties are determined for every model separated by a 0.02 interval in ϵ_{HP}. Solid curves show enthalpy data (left scale). The lowest solid curve is $\Delta H_{vH}/\Delta H_{cal}$ where ΔH_{cal} is calculated by Eq. (1) at $T = \infty$. The top and middle curves are respectively $\Delta H_{vH}/\Delta H_{vH}(T_{0.98})$ and $\Delta H_{vH}/\Delta H_{vH}(T_{0.999})$, as defined in the legend of Table II. The dashed curve shows the variation of transition midpoint temperature $T_{1/2}$ (right scale) with ϵ_{HP}.

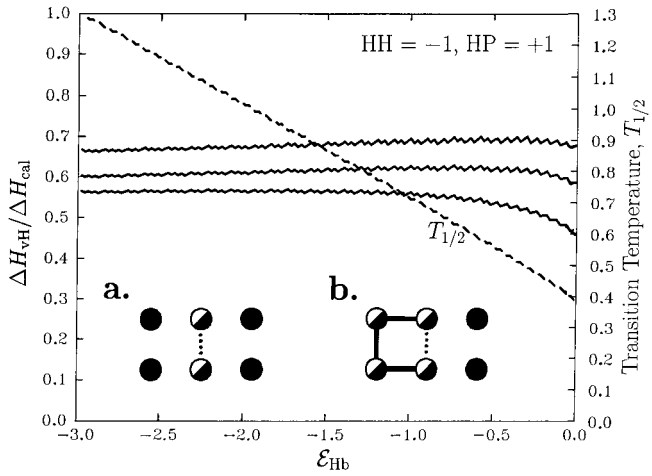

Fig. 14. Same as Figure 13, except the HP contact enthalpy ϵ_{HP} is also fixed, at $+1$, and a cooperative hydrogen-bond-like interaction term is varied. In the drawing at the lower left, a half-filled circle indicates that a given monomer can be either H or P. A lattice hydrogen-bond (labeled by dotted lines connecting a pair of monomers) is considered buried and assigned a favorable energy ϵ_{Hb} (in units of k_B) when one of the following two conditions is satisfied. **a:** When its environment is completely surrounded by H monomers; or **b:** when it is at a tight turn (as in a lattice helical conformation, see Fig. 16) and is shielded by two H monomers on the other side.

is to test the abilities of various physically plausible interactions in reducing the multiplicity of low-enthalpy non-native conformations. This exercise is for testing of principles. We do not intend it to be a detailed quantitative proposal for how real proteins interact, nor do we aim at achieving $\Delta H_{vH}/\Delta H_{cal} = 1$. As should be expected, many refined features of real proteins are not taken into account by simple models in two dimensions. The above analysis shows that calorimetric two-state cooperativity can be achieved trivially by assumption [model (a) in Fig. 9], or nontrivially by the invocation of a form of capillarity cooperativity [models (c–d) in Fig. 9]. This leads to the physical question: In the absence of these ad hoc assumptions, is it possible to improve a flexible chain model with pairwise contact interactions to achieve relatively high thermodynamic cooperativity by adding other energetic ingredients that are simple, general, and physically plausible?

We take a step-by-step approach (Figs. 13–15). We begin by assigning a repulsive interaction $\epsilon_{HP} > 0$ to every HP contact (between an H and a P monomer) to replace the zero (neutral) HP contact enthalpy in the original model. This assumes that hydrophobic residues dislike contacting polar residues more than they dislike exposing to water. In general, addition of repulsive interactions enhances stability.[27,91] Figure 13 tests a range of ϵ_{HP}'s. It shows that $\Delta H_{vH}/\Delta H_{cal}$ increases from 0.31 at $\epsilon_{HP} = 0$ (original HP model) to 0.4–0.5 for a modest value of ϵ_{HP} between 0.5 and 1.0. Further increase in ϵ_{HP} leads to a decrease of $\Delta H_{vH}/\Delta H_{cal}$ from its maximum value. This is because while the added interactions increase the average enthalpic separation between the native and denatured ensembles, they also increase the enthalpy distribution width of the denatured ensemble. The variation of $\Delta H_{vH}/\Delta H_{cal}$ with respect to ϵ_{HP} reflects this trade-off. This is a general feature that applies also to the other added interactions considered below.

This shows that addition of repulsive interactions enhances thermodynamic cooperativity. We now fix $\epsilon_{HP} = +1$ and consider a cooperative hydrogen-bonding-like term (Fig. 14). A number of mutagenesis and model compound experiments suggest that the strength of hydrogen bonds in real proteins is stronger when they are buried than when they are exposed to water.[92–95] Here this effect is tentatively modeled by the interaction in Figure 14, which assigns a favorable enthalpy $\epsilon_{Hb} < 0$ when a model hydrogen bond is buried, whereas the model hydrogen bond remains neutral if it is exposed (see also discussions in Refs. 96–101). The new lattice interaction in Figure 14 bears some similarity to an environment-dependent hydrogen bonding term introduced recently in a continuum protein folding model.[102] This is a nonadditive term. The importance of accounting for nonadditive effects in protein folding has been emphasized by Dill[103]; the effects of non-pairwise multiple-body interactions on the free energy barrier between the native and denatured states have been studied by Kolinski et al.[28,29] More recently, Betancourt and Thirumalai[104] pointed out that most probably purely pairwise contact potentials are insufficent for protein structure prediction.

We designate seven monomer pairs in our model sequence to be potentially hydrogen bonding (shown by dotted lines in the structure in Fig. 16). These are chosen to mimic hydrogen bonds that stabilize helices and turns

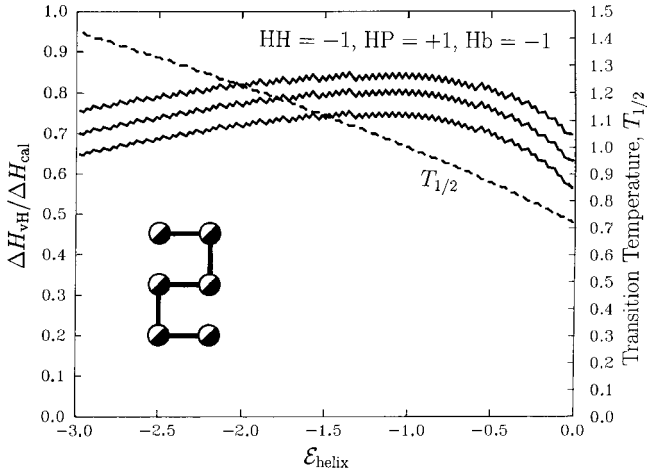

Fig. 15. Same as Figure 14, except ϵ_{Hb} is also fixed, at −1, while a helical cooperativity term is varied. Each occurrence of two consecutive lattice helical turn (as shown) that coincide with a helical segment in the native structure (Fig. 16) is assigned an extra favorable energy ϵ_{helix}, irrespective of the monomer types making up the lattice helix.[149–151]

in real proteins. Figure 14 shows that as ϵ_{Hb} is decreased (i.e., made more favorable) $\Delta H_{vH}/\Delta H_{cal}$ first increases then levels off, whereas the transition midpoint temperature increases linearly, demonstrating that calorimetric cooperativity does not necessarily correlate with native stability. Figure 14 shows that addition of the cooperative hydrogen bonding term leads to a modest increase in cooperativity, $\Delta H_{vH}/\Delta H_{cal}$ is increased from around 0.45 at $\epsilon_{Hb} = 0$ to larger than 0.55 for $\epsilon_{Hb} \leq -1$.

Next, we fix ϵ_{Hb} at −1 (with $\epsilon_{HH} = -1$ and $\epsilon_{HP} = +1$) and consider a helical cooperative parameter ϵ_{helix} (Fig. 15). This term is motivated by Zimm-Bragg type[105] helix-coil theory.[99,106] Figure 15 shows that with this additional contribution, $\Delta H_{vH}/\Delta H_{cal}$ can be enhanced to ≈ 0.75. This effect may be understood in light of the fact that for polymers in general, higher chain stiffness leads to more cooperative coil-globule transitions.[107–111] By encouraging formation of rodlike structure, $\epsilon_{helix} < 0$ may be viewed as contributing to some kind of effective stiffness of the chain.

We choose $\epsilon_{helix} = -1$ for further analysis. For this model, $\Delta H_{vH}/\Delta H_{cal} = 0.74$ (Fig. 17). This is a significant improvement from 0.31 of the original HP model. Moreover, we note that the ΔH_{cal} here is obtained at the theoretical athermal limit, $T = \infty$, at which point all conformations, including those that contain numerous unfavorable interactions, are equally probable. Realistically, however, although highly unfavorable structures with large positive enthalpies do exist as conformational possibilities for real proteins, their populations are expected to be negligible under experimental conditions, which by design and by necessity are always conducted in the presence of solvent, and restricted to a finite temperature range (see below). In view of this, if we take ΔH_{cal} to be $\langle \Delta H(T_{0.999}) \rangle$ (c.f. Table II) instead of $\langle \Delta H(\infty) \rangle$ [Eq. (1)], the van't Hoff to calorimetric enthalpy ratio would be further increased from 0.74 to 0.81. If we further substitute the population-based ΔH_{vH} by the experimental ΔH_{vH}^{exp} computed by using Eq. (18), the van't Hoff to calorimetric enthalpy ratio $\Delta H_{vH}^{exp}/\langle \Delta H(T_{0.999}) \rangle = 0.87$. This exercise shows that it is possible to construct simple models with thermodynamic cooperativity quite close to that of real proteins by including physically plausible cooperative elements in the interaction potential.

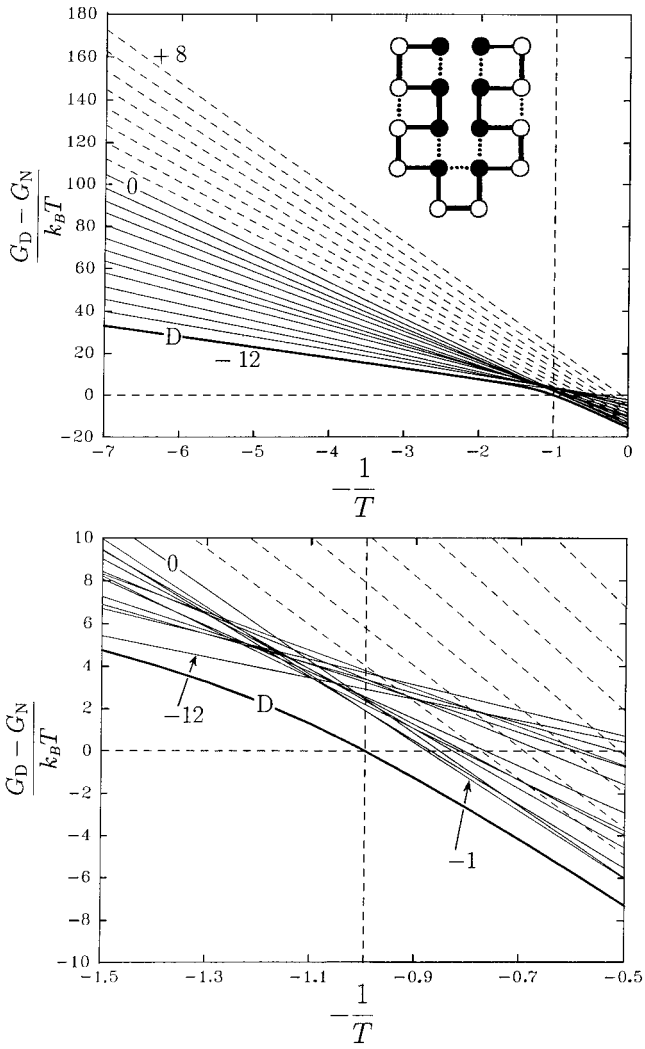

Fig. 16. The upper panel is the same as Figure 3, but for the cooperative model shown, with ϵ_{HH}, ϵ_{HP}, ϵ_{Hb}, and ϵ_{helix} equal −1, +1, −1, and −1, respectively (in units of k_B), and the lines are now labeled by enthalpy. The native (ground) state enthalpy is −17, the highest and lowest enthalpies of the denatured ensemble are, respectively, +8 and −12. Stabilities of conformations with positive and non-positive enthalpies are plotted by dashed and solid lines respectively. The lower panel is a enlarged version of the plot around the transition midpoint showing that this model is much closer to being calorimetrically two-state than the original HP sequence because its denatured population at the transition midpoint is dominated by conformations with relatively high enthalpies ≈ 0.

Figure 16 also shows that the positive enthalpy states (inclined dashed lines) are highly unstable except when $1/T \approx 0$, i.e., $T \to \infty$. This implies that although repulsive interactions (the HP contacts with $\epsilon_{HP} > 0$ in this example)

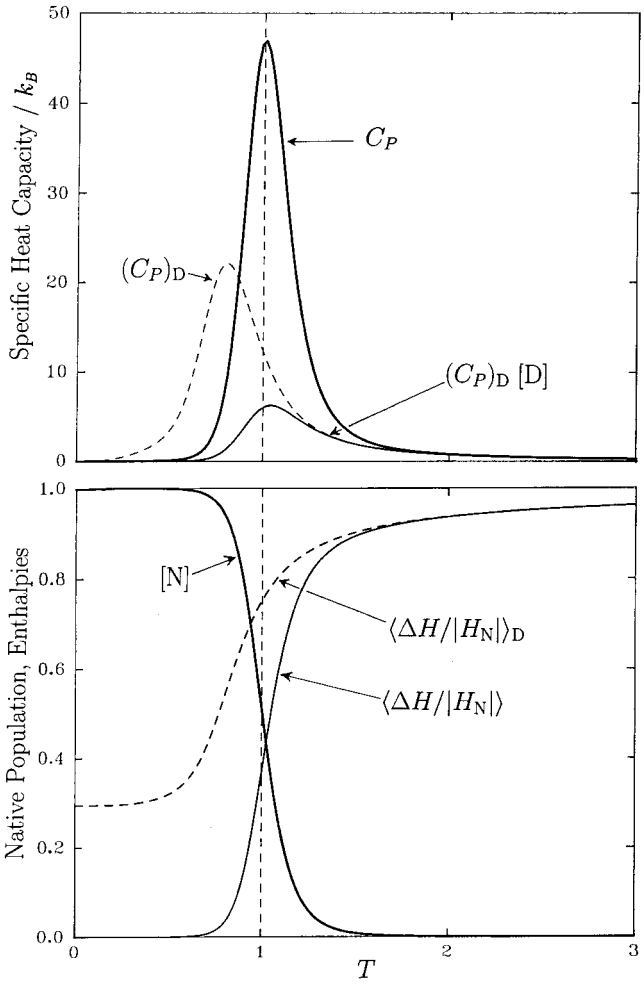

Fig. 17. Specific heat capacities and enthalpies. Same as Figure 4, but for the model described in Figure 16.

have important thermodynamic consequences because of their elimination of certain hitherto possible low-enthalpy conformations, the high-enthalpy conformations they create are quite irrelevant to the overall conformational distribution except when the system is near the theoretical athermal limit that most probably is experimentally unachievable. This is because these conformations have unfavorable enthalpies and their numbers are not sufficiently abundant.

For this model, as the denaturation process proceeds with increasing T, stabilities of low-enthalpy denatured conformations are overtaken by conformations with enthalpies ≈ 0 before the entire ensemble (D) reaches the transition midpoint (Fig. 16). This behavior is similar to that exhibited by capillarity models (b)–(d) in Figure 12. This accounts for this model's relatively high cooperativity. The corresponding heat capacity and enthalpy functions are given in Figure 17. It should be noted that the assignment of hydrogen bonding pairs and helical cooperativity contributions in this test model introduces local biases in favor of the native conformation. Therefore, like the Gō and HP+ models before it, this model potential is not fully self-sufficient in that it requires a prior knowledge of the native conformation. Nonetheless, the present results are significant because additive contact energies with strong native biases in Gō and HP+ models fail to produce a similarly high level of thermodynamic cooperativity. Taken together, this test study suggests that both local (hydrogen bonding, helical propensity) and nonlocal (attractive and repulsive contact-like) interactions and their cooperative interplay (environment-dependent hydrogen bonding) are likely to be needed for proteinlike thermodynamic cooperativity.

Using Entropic Interactions to Better Account for Heat Denaturation

In the course of developing the above scenario to illustrate how proteinlike thermodynamic cooperativity may be achieved, we have assumed so far for simplicity that all intrachain interactions are enthalpic. However, it is well established that protein-solvent interactions lead to effective or "implicit-solvent" protein intrachain interactions that contain entropic parts, especially for interactions that are classified as hydrophobic.[96,112–116] It has also been shown that incorporating entropic interactions can have significant consequences on the interpretation of simple model results.[26,27] Would the physical picture emerging from the above analysis be modified substantially when entropic elements of the intrachain interactions are incorporated?

To take a first step in more realistic modeling of temperature dependence, we briefly explore effects of entropic interactions. In the current cooperative model (Fig. 16), for a conformation with h HH contacts, the enthalpy is equal to $h\epsilon_{HH}$ plus contributions from other interactions (i.e., repulsive HP contacts, buried hydrogen bonds, and helical cooperativity). We now consider a modified model in which the hydrophobic HH contact interaction is designated to be entirely entropic, while all other interactions remain enthalpic. The interaction free energy of h HH contacts is now $Th\epsilon_{HH}$ (whereas it was $h\epsilon_{HH}$ in the unmodified original model), and numerically $\epsilon_{HH} = -k_B$ is unchanged. This is motivated by the observation that hydrophobic interactions are principally entropic around 25°C. Because this is intended only as a demonstration of principle,[¶] we neglect the fact that as temperature is increased, the hydrophobic interaction gradually crosses over from being principally entropic to increasingly enthalpic.[113]

If the original enthalpic density of states, now denoted as $g_o(H)$, is written as $g_o(H) = \Sigma_h g_{o,h}(H)$, where $g_{o,h}(H)$ is

[¶]Taking into account the entropic part of the hydrophobic interaction does not alter our conclusion that additive hydrophobic interactions alone are insufficient for calorimetric two-state cooperativity. Consider another model potential for the HP sequence in Figure 7, now with a mixed entropic-enthalpic HH contact free energy equals to $(1 + T)\epsilon_{HH}$ instead of the purely enthalpic ϵ_{HH} in the original HP model, but without the added repulsive HP interaction and the hydrogen-bonding-like and helical propensity terms. The resulting calorimetric two-state ratios remain low: $\Delta H_{vH}/\Delta H_{cal} = 0.36$ and $\Delta H_{vH}^{exp}/\Delta H_{cal} = 0.53$.

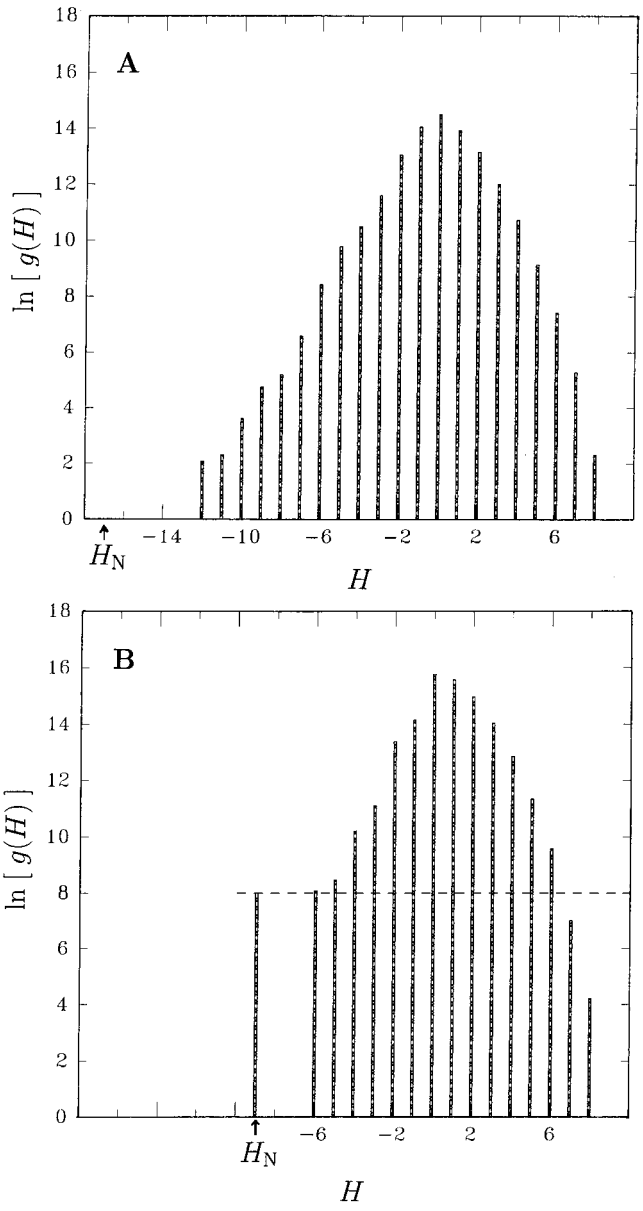

Fig. 18. Density of (enthalpic) states. Enthalpy H is in units of $|\epsilon_{HH}| = k_B$, H_N is the native enthalpy. **A:** $g_o(H)$ for the model in Figure 16 when all interactions are enthalpic. In this case $g(H_N) = 1$, hence ln $g(H_N) = 0$. **B:** $g_m(H)$ for the modified model in which the hydrophobic HH contact interaction is entropic, whereas all other interactions remain enthalpic. The horizontal dashed line indicates the native value of the density of states in this model; see text and Eq. (21) for details.

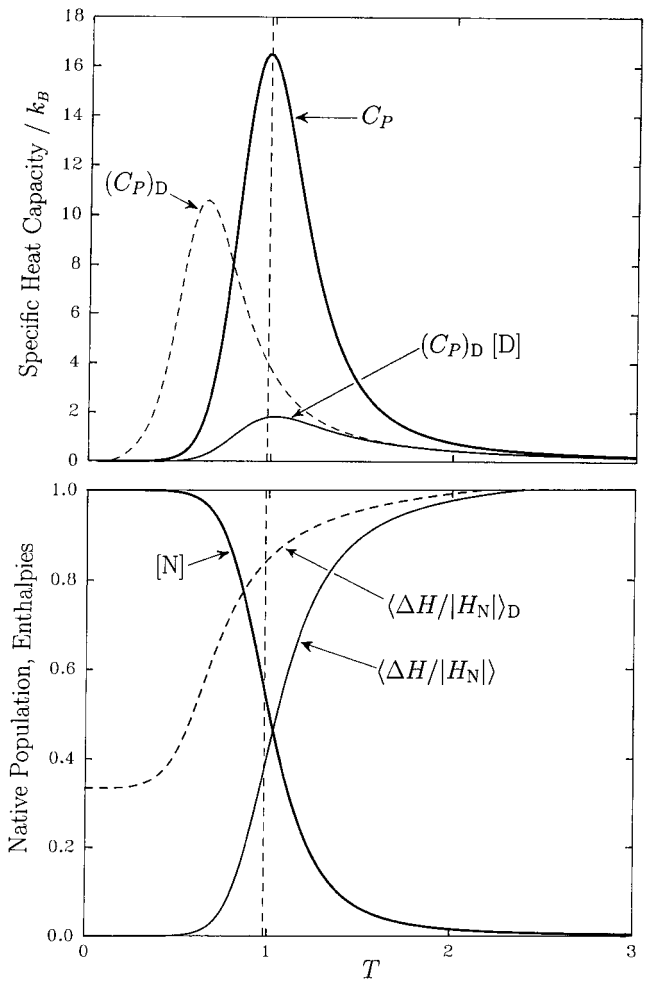

Fig. 19. Specific heat capacities and enthalpies. Same as Figure 17, but for the modified model in which the hydrophobic HH contact interaction is entropic.

the number of conformations with h HH contacts and a total enthalpy H in the original cooperative model, the enthalpic density of states $g_m(H)$ of the modified cooperative model (Fig. 18) is given by

$$g_m(H) = \sum_{H' = H + h\epsilon_{HH}} g_{o,h}(H') e^{-h\epsilon_{HH}/k_B} = \sum_{H' = H - h} g_{o,h}(H') e^{h}, \tag{21}$$

where H and H' are in units of $-\epsilon_{HH} = k_B$ in the last summation. One noteworthy feature of the density of states for the model with entropic interactions is that even when the native state is designated to be a single conformation, it is possible for its value of g to be larger than that of some other conformations ($H = 7, 8$ in Fig. 18B). This is not possible when intrachain interactions are entirely enthalpic. This is because entropic interactions take into account the entropy of the solvent. The combined entropy of the native state plus its interacting solvent can be larger than the combined entropy of some other conformation plus its interacting solvent, because protein-solvent interactions depend on what and how many chemical groups are exposed to the solvent by a given conformation.

Figure 19 shows that the modified and original cooperative models have similar behaviors (Fig. 17), although the modification causes a significant shift in the enthalpy distribution and a narrowing of the "energy gap" between the lowest denatured and native states from 5 to 3 enthalpy units (Fig. 18). The calorimetric cooperativity of the modified model, $\Delta H_{vH}/\Delta H_{cal} = 0.78$, is essentially the same as the original model. If we substitute ΔH_{cal} with $\langle \Delta H(T_{0.999}) \rangle$ (c.f. Table II) as for the original model, we

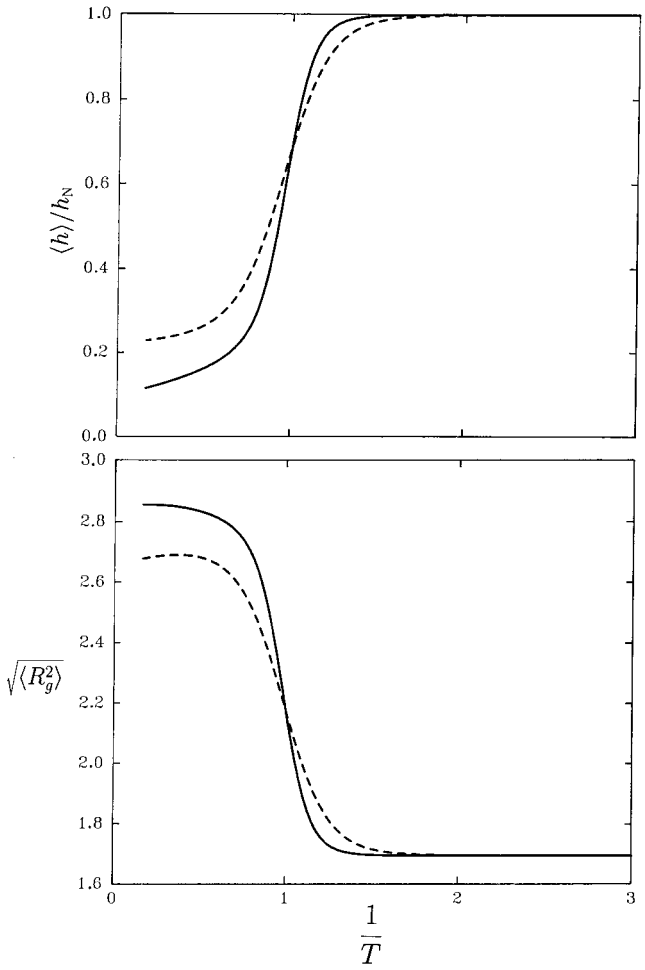

Fig. 20. Comparing the purely enthalpic model (solid curves) and the modified model with entropic hydrophobic (HH) interactions (dashed curves). As functions of inverse temperature, the upper panel shows the average number of HH contacts (normalized by the native number of HH contacts); the lower panel shows the root mean square radius of gyration of the model chains (in units of the lattice bond length).

obtain 0.80 as the van't Hoff to calorimetric enthalpy ratio for the modified model. If we further substitute the population-based $\Delta H_{\rm vH}$ by the $\Delta H_{\rm vH}^{\rm exp}$ computed by using Eq. (18), this ratio further increases, to $\Delta H_{\rm vH}^{\rm exp}/\langle \Delta H(T_{0.999}) \rangle = 0.85$.

It is of interest that this modification leads to significant changes in the conformational distribution in the denatured ensemble. Figure 20 shows that the denatured ensemble of the modified model is more compact and has more hydrophobic contacts than the original model. If we assume that the temperature dependence of the modified model is more realistic—because temperature increases do not significantly weaken a principally entropic hydrophobic interaction—then the denatured ensemble of this model would be a crude mimic of heat-denatured states of real proteins. On the other hand, the model temperature dependence of the original model may be viewed, in this perspective, as a crude model for denatured states of real proteins at high denaturant concentrations. This is because common denaturants do weaken the hydrophobic sticking factor, $-\epsilon_{\rm HH}/(k_B T)$, and cause it to decrease approximately linearly with increasing denaturant concentration.[26,27,117] If we neglect for simplicity the denaturant's different effects on other interaction types in the model, the $1/T \to 0$ results in Figure 20 may be interpreted as a comparison between unfolding by heat versus unfolding by denaturant. In this regard, the present prediction is consistent with the notion that denatured conformations unfolded by heat are in general more compact and have more residual hydrophobic contacts than those at high denaturant concentrations[118–120] (see also Lazaridis and Karplus[116] and Privalov[121]).

Calorimetric Contraints on Z-Scores of Knowledge-Based Potentials?

We have argued that calorimetric data should be used to place restrictions on enthalpy distributions in microscopic physical models. It follows naturally that calorimetric data can also be used to evaluate interaction energies derived statistically from native protein structure databases,[104,122–127] because these knowledge-based potentials are often believed to be reasonable approximations of the real physical interactions. Of particular relevance here is the Z-score commonly used in protein structure prediction.[128] Z-score quantitates the width of the interaction energy distribution of a collection of misfolded (non-native) structures vis-à-vis the energy separation between the native and non-native conformations. Therefore, in light of the above discussion, the Z-score is analogous to the variable $H_{\rm D}/\sigma_H$ that relates the enthalpic distribution of a model protein to its calorimetric cooperativity [Eq. (9)]. An interesting prospect of using calorimetric data to constrain Z-scores was recently raised by Zhang and Skolnick.[129] They concluded from their analysis that Z-scores from knowledge-based potentials might be too small. However, two issues have apparently been overlooked in their work.

First, Zhang and Skolnick used heat capacities of the unfolded and native states from Makhatadze and Privalov ($C_P^{\rm U}$ and $C_P^{\rm N}$ in Table VI of Makhatadze and Privalov[74]) to determine densities of states, which they then applied to determine energy distributions among different conformations. But according to Makhatadze and Privalov[72] themselves, the differences between native- and unfolded-state heat capacities reflect mainly hydration effects originating from solvent degrees of freedom, not enthalpic transitions among very different unfolded (non-native) conformations. They showed that their unfolded heat capacities could be determined by summing up contributions from "non-interacting" individual amino acids,[72,74] i.e., from "the amino acid composition" (page 326 of Makhatadze and Privalov[74]) while assuming that the unfolded chain is in a "random-coil" state.[73] Indeed, they stated that "the heat capacity effect due to the configurational freedom gain by the polypeptide chain was found to contribute only a small part of the overall heat capacity change on unfolding."[73] Therefore, these heat capacities do not originate from the

diverse conformational variation in the denatured states that causes variations in intrachain interaction energies which in turn give rise to the standard deviation denominator in the formula for the Z-score. In short, Makhatadze and Privalov's C_P^{U} is *not* physically equivalent to $(C_P)_{\mathrm{D}}$ defined above in Eq. (16). In fact, these heat capacities are precisely the parts of calorimetric data that are subtracted out to yield the transition excess heat capacity function.[19] As discussed in the Materials and Methods section below, it is the transition excess heat capacity—not the absolute heat capacities of the native and unfolded states—that is directly related to transitions among conformational states. Even in the alternate view proposed recently by Lazaridis and Karplus[116] that a part of C_P^{U} originates from variations in intrachain interactions, if contributions from vibrations of the model covalent bonds are excluded, this part constitutes less than one half of the heat capacity of the denatured state; and a significant part of C_P^{U} is still accounted for by solvation effects. Hence, Makhatadze and Privalov's C_P^{U} should not be used, at least not in its entirety, to describe energetic fluctuations originating from protein conformational diversity in the denatured state.

Second, Zhang and Skolnick used a random energy model similar to the above Eq. (9) in their analysis and assumed that the interaction free energies are purely enthalpic. But they did not use the calorimetric two-state criterion to constrain *Z*-scores. Here we investigate the consequence of applying this experimental criterion in conjunction with their assumption that interaction free energies are purely enthalpic. In the present notation, and using Eq. (13) above, their random-energy-model expression for the *Z*-score [their Eq. (12)] is given by

$$Z = \frac{H_{\mathrm{D}}}{\sigma_H} = \sqrt{\frac{2 \ln g_{\mathrm{D}}}{1 - (\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}})^2}}. \tag{22}$$

When applied to proteins that satisfy the calorimetric two-state criterion $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} \to 1$, Eq. (22) implies that $Z \to \infty$, in contrast to the maximum *Z*-score of $\approx$7.0 deduced by Zhang and Skolnick from analyzing C_P^{U} and C_P^{N}. For instance, if $\Delta H_{\mathrm{vH}}/\Delta H_{\mathrm{cal}} = 0.98$, $Z = 7.1\sqrt{\ln g_{\mathrm{D}}}$, where $g_{\mathrm{D}} \gg 1$. Is this requirement applicable?

The above argument would be valid if the effective intrachain interaction free energies in proteins are purely enthalpic, as was assumed by Zhang and Skolnick.[129] Because the heat capacity of a system is proportional to its enthalpy variance, the experimental calorimetric two-state condition does imply a narrow enthalpy distribution among the denatured conformations, as discussed above. However, the interaction free energies that contribute to *Z*-scores are not purely enthalpic, and a narrow distribution of enthalpies does not necessarily imply a narrow distribution of interaction free energies. The stability of a protein conformation is determined by a free energy that can have both enthalpic and entropic components.[26,27,46,115]

We may rewrite the partition function Eq. (26) as a sum over conformational states with individual Gibbs free energies $\mathcal{G}_i$ (Freire[19]).[‖] The Boltzmann-averaged interaction free energy is then

$$\langle \mathcal{G} \rangle = \frac{\sum_i \mathcal{G}_i e^{-\mathcal{G}_i/(k_B T)}}{\sum_i e^{-\mathcal{G}_i/(k_B T)}}. \tag{23}$$

To obtain an expression for the heat capacity in terms of distribution of interaction free energies, we substitute the standard relation $\langle H \rangle = \langle \mathcal{G} \rangle - T\langle \partial \mathcal{G}/(\partial T) \rangle$ between enthalpy and Gibbs free energy into the heat capacity expression Eq. (29), and arrive at

$$C_P = \frac{1}{k_B T^2}\left(\langle \mathcal{G}^2 \rangle - \langle \mathcal{G} \rangle^2\right) + \frac{2}{k_B T}\left(\langle \mathcal{G} \rangle \left\langle \frac{\partial \mathcal{G}}{\partial T} \right\rangle - \left\langle \mathcal{G} \frac{\partial \mathcal{G}}{\partial T} \right\rangle\right) - T\left\langle \frac{\partial^2 \mathcal{G}}{\partial T^2} \right\rangle + \frac{1}{k_B}\left[\left\langle \left(\frac{\partial \mathcal{G}}{\partial T}\right)^2 \right\rangle - \left\langle \frac{\partial \mathcal{G}}{\partial T} \right\rangle^2\right]. \tag{24}$$

This shows that the heat capacity is not proportional to the variance $\langle \mathcal{G}^2 \rangle - \langle \mathcal{G} \rangle^2$ of interaction free energies. Hence, the calorimetric two-state criterion does not necessarily imply a narrow distribution of interaction free energies, although it does imply that, even when the interaction free energies are diverse, their enthalpic parts have to be narrowly distributed. Because *Z*-scores describe interaction free energies that are not necessarily purely enthalpic, whether *Z*-scores from knowledge-based potentials are too small[129] remains to be further investigated. By the same token, to make better connections with experiments, it is important to study lattice protein models with temperature-dependent interactions.[26,27]

CONCLUSIONS

We have examined implications of the calorimetric two-state criterion on enthalpy distributions in denatured states of proteins. In all cases considered here, as long as a protein can be completely denatured (i.e., has practically zero native population) at high temperatures, the criterion that the van't Hoff enthalpy being approximately equal to the calorimetric enthalpy implies that the denatured enthalpy distribution is narrow relative to the average enthalpy difference between the native and denatured states. This is a general observation, despite certain nonuniformities arising from the fact that different van't Hoff enthalpies have been used in analyses of calorimetric data. Using simple exact chain models, we showed that calorimetric two-state cooperativity can be reconciled with the multiple-state nature of proteins as suggested by recent native-state hydrogen exchange experiments, an observation that is consistent with the polymeric nature of proteins. This is because conformational states with enthalpies intermediate between the native and the fully unfolded states can exist without violating the calorimetric two-state criterion as long as their populations are not

[‖]These $\mathcal{G}_i$'s should not be confused with the Gibbs free energy *G* in Eq. (27) for the entire system.

significant around the folding/unfolding transition midpoint. However, not all intrachain potential functions can produce these thermodynamic behaviors. Hence, this set of experimental observations should be used to evaluate potential functions in protein models. Using a primitive model that accounts for the entropic part of the hydrophobic interaction, we showed the importance of distinguishing between enthalpic and entropic components of effective intrachain interactions in modeling temperature-dependent properties of proteins.

We tested a number of exact two-dimensional lattice protein models and found that many of them, including highly specific Gō models, do not satisfy the calorimetric two-state criterion. These results strongly suggest that additive hydrophobic-like contact interactions alone are insufficient for proteinlike calorimetric cooperativity. However, incorporation of physically plausible energetic components, such as contributions from helical cooperativity and environment-dependent hydrogen bonding, can lead to thermodynamic properties much closer to that observed for calorimetrically two-state proteins. Effects of other interactions, such as sidechain packing,[5,30,130] may be tested in a similar manner. Thus, the macroscopic calorimetric two-state criterion can serve as a powerful constraint and a useful tool for constructing better microscopic protein chain models.

MATERIALS AND METHODS

We supply in this section the necessary background and technical details for the above treatment of van't Hoff and calorimetric enthalpies.

Partition Functions for Isothermal-Isobaric Ensembles

The specific heat capacities reported by protein calorimetry are measured at constant pressure P. The statistical mechanical analysis of such an isothermal-isobaric NPT ensemble begins with the partition function[131]

$$Q(T) = \int dH g(H) e^{-H/(k_B T)}, \tag{25}$$

where k_B is Boltzmann's constant, T is absolute temperature, $g(H)$ is the density of states, the integration is over all possible enthalpies H, $g(H)dH$ is the number of ways the system can be arranged to have its enthalpy between H and $H + dH$, and Q is equivalent to the quantity $\Delta(T, P, N)$ in Eq. (1.3.17) of Ben-Naim.[131] If the enthalpy distribution is treated as discrete, as in the seminal work of Freire and Biltonen[17] and in most recent lattice models, the partition function becomes

$$Q(T) = \sum_i g(H_i) e^{-H_i/(k_B T)}, \tag{26}$$

where H_i's are discrete enthalpies, and $g(H_i)$ is the discrete density of states, i.e., the number of ways the system takes on enthalpy H_i. The notation Q is the same as that in Freire and Biltonen's formulation. If we define ΔS_i to be the entropic part of their Gibbs free energy ΔG_i (c.f. the $\mathcal{G}_i$ used above) between the ith and the reference states, our $g_i(H)$ equals their $\omega_i \exp(\Delta S_i/k_B)$, where ω_i is the degeneracy of the ith state in their treatment.

Because $H = E + PV$, in addition to the usual summation over energies E in canonical constant-volume NVT formulations (see, e.g., Refs. 11 and 132), Eqs. (25) and (26) entail summation over possible volumes V. Hence, this formulation includes volume effects. Although PV contributions to liquid-state solvation free energies are often negligible under atmospheric pressure,[133,134] they are critical to protein denaturation at high pressure, at around 1,000 atm.[135–137] In the present application to protein folding, we make the physically reasonable assumption that Q is well approximated by summing only over contributions from different configurational states (as defined by the positions of all the particles in the system) at their respective average volumes, without additional summation over other volumes. This is based on the expectation that any appreciable volume variation from the average volume of each configurational state would lead to highly unfavorable arrangements with negligible contributions to Q. Thus, a given configuration of the system, in our case protein plus solvent, is expected to be associated with essentially only a single enthalpy. In other words, we proceed almost exactly as in the canonical NVT ensemble, except E is replaced by H, and the constant-volume specific heat C_V is now replaced by the constant-pressure specific heat C_P.

Basic Relations

We use mostly the discrete form Eq. (26) of the partition function in the analysis below. Equivalent results for the continuum Eq. (25) may be obtained by using integrations instead of summations. The Gibbs free energy of the entire system is given by

$$G(T) = -k_B T \ln Q(T). \tag{27}$$

The average enthalpy of the system

$$\langle H(T) \rangle = \frac{\sum_i g(H_i) H_i e^{-H_i/(k_B T)}}{\sum_i g(H_i) e^{-H_i/(k_B T)}} = k_B T^2 \frac{\partial \ln Q(T)}{\partial T}, \tag{28}$$

the temperature derivative of which is the constant-pressure specific heat

$$C_P(T) = \frac{\partial \langle H(T) \rangle}{\partial T} = \frac{\langle H^2(T) \rangle - \langle H(T) \rangle^2}{k_B T^2}, \tag{29}$$

where P is understood to be held constant in all differentiations, $\langle \ldots \rangle$ denotes Boltzmann averaging; $\langle H^2(T) \rangle = \Sigma_i g(H_i) H_i^2 \exp[-H_i/(k_B T)]/Q(T)$. Calorimetry measures $C_P(T)$, which yields $\langle H(T) \rangle$ upon integration,

$$\langle H(T) \rangle = \langle H(T_0) \rangle + \int_{T_0}^{T} dT' C_P(T'), \tag{30}$$

where T_0 is any reference temperature. By using Eq. (28), a further temperature integration gives the partition function,

$$\ln Q(T) = \ln Q(T_0) + \int_{T_0}^{T} dT' \frac{\langle H(T')\rangle}{k_B T'^2}. \qquad (31)$$

It follows that in principle the partition function as a function of T can be accessed numerically by calorimetry.[17,19]

The Statistical Mechanics of Baseline Subtractions—on the Positive Definiteness of Densities of States

The analysis of calorimetric data in terms of partition functions and densities of states usually begins with the excess enthalpy function, often denoted as $\langle \Delta H \rangle$, which is obtained by applying an empirical subtraction procedure on measured heat capacities.[17,19] Here we start at a more basic level, beginning conceptually with the fundamental physical partition function of the entire system of protein plus solvent, before any subtraction. We then examine how the changes incurred on the heat capacity function at different stages of the subtraction scheme may affect the mathematical properties of the various "partition functions" deduced by using Eqs. (30) and (31). Our purpose here is to clarify the relationship between experimental and simple model density of states, to come to grips with the complexities in a full account of calorimetric data in terms of physical densities of states, and to appreciate the statistical mechanical simplification provided by subtraction procedures.

Two essential subtractions are involved in the analysis of protein calorimetry data: (i) Subtraction of the buffer solution heat capacity to arrive at the "heat capacity of the protein" and (ii) subtraction of the extrapolated native and unfolded state heat capacities from the heat capacity of the protein to arrive at a "transition excess heat capacity."[19] Heat capacities computed from most protein lattice models—especially those with temperature-independent interactions—correspond directly to and should be compared with experimental transition excess heat capacities. We view this as a reasonable point of departure because both quantities are constructed to account for enthalpic variations associated with conformational transitions.

For brevity, we do not reiterate in detail the nature, accuracy, and experimental aspects of these procedures, because they can be found in excellent reviews.[16,19,68] It suffices to note that both subtractions entail studying *effective* thermodynamic systems defined by the *difference* between two C_P's, not the full thermodynamics of actual physical systems characterized by absolute C_P's. For a physical system describable by classical statistical mechanics, we expect its density of states to be positive definite, i.e., $g(H_i) \geq 0$ for all H_i's, because the multiplicity of states cannot be a negative number. However, for the *effective* density of states obtained by integrations of a difference between two physical heat capacities, there is no a priori general guarantee for positive definiteness. This can be seen most easily from the special case when $\langle H(T_0)\rangle = \ln Q(T_0) = 0$ in Eqs. (30) and (31). Let Q_{total} be the partition function corresponding to a total heat capacity $(C_P)_{\text{total}}$ that is subsequently decomposed into two parts,

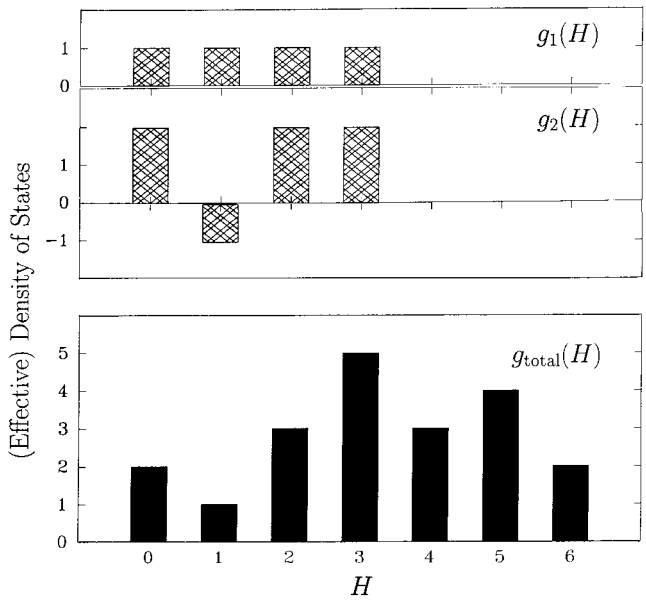

Fig. 21. An hypothetical case showing that calorimetric baseline subtractions can lead to negative terms in effective densities of states. $g_{\text{total}}(H)$ (bottom panel) is the convolution of $g_1(H)$ and $g_2(H)$ [top two panels, see Eq. (34)]. Although $g_1(H)$ and $g_{\text{total}}(H)$ are positive definite, $g_2(H)$ is not.

$$(C_P)_{\text{total}} = (C_P)_1 + (C_P)_2, \qquad (32)$$

and Q_1 and Q_2 are the effective partition functions for $(C_P)_1$ and $(C_P)_2$, respectively, then

$$Q_{\text{total}} = Q_1 Q_2, \qquad (33)$$

because $\ln Q_{\text{total}} = \ln Q_1 + \ln Q_2$. It follows that the total density of states g_{total} is a convolution of the two effective densities of states, g_1 and g_2:

$$g_{\text{total}}(H) = \sum_{H = H_1 + H_2} g_1(H_1) g_2(H_2). \qquad (34)$$

Figure 21 illustrates the possibility that the effective density of states deduced from the difference between two physical C_P's can take negative values. In Figure 21, both g_{total} and g_1 are positive definite. Hence, they can serve as densities of states of two hypothetical physical systems. The effective density of states g_2 calculated by Eq. (34) may be viewed as being obtained from a heat capacity differential ΔC_P; in this case $\Delta C_P = (C_P)_2 = (C_P)_{\text{total}} - (C_P)_1$. Figure 21 shows that g_2 is not positive definite.

The relevance of this to the analysis of calorimetric data is as follows. Consider first the subtraction of solvent (buffer solution) contribution. By standard procedures,[16,19] the "heat capacity of the protein" is a differential that equals to (i) the heat capacity of a protein plus its surrounding solvent with which it interacts *minus* (ii) the

heat capacity of a *bulk* solvent (not interacting with the protein) that have the same mass as the solvent interacting with the protein in (i). This differential heat capacity does not describe an isolated protein in the absence of solvent.[20] Because of the interaction between the protein and the solvent, integration of this differential heat capacity using Eqs. (30) and (31) gives an effective partition function, but not the partition function and density of states of a physical system that exists on its own. The protein-solvent interactions in (i) alter the enthalpy distribution in the solvent relative to the bulk solvent in (ii) (see, e.g., Scheraga[138]). Hence, the solvents in (i) and (ii) are different, and their effects do not cancel out by subtraction as they would if the protein were not interacting with the solvent in (i). Therefore, there is no simple guarantee that the effective density of states deduced from this differential heat capacity is positive definite.

Consider next the subtraction of the extrapolated heat capacities of the native and the denatured states.[16,19,68] The existence of these nonzero heat capacities implies that there are multiple enthalpy levels associated with any given protein conformation in the protein plus solvent system. The origins of these enthalpy levels include the enthalpy variation among the many possible configurations of the solvent molecules, collectively referred to as hydration effects,[72–74] and those sampled by vibrations, side-chain movements and other motions in the protein[139] whose amplitudes are sufficiently small to fit within the confine of the definition of a single "conformation." Some of these motions may be viewed as transitions between conformational substrates.[140] In this perspective, the goal of the subtraction of native and denatured state heat capacities is to factor out these multiple enthalpy levels associated with a conformation, so that essentially only one enthalpy level is left for each conformation in the resulting effective system. When combined with the buffer subtraction described above, a transition excess heat capacity, $C_{P,\mathrm{ex}}$, is obtained as the final product ($C_{P,\mathrm{ex}}$ is sometimes denoted as $\langle \Delta C_{P,\mathrm{tr}} \rangle$, see, e.g., Freire.[19]).

The assumption and expectation are that the effective partition function and the corresponding (excess) density of state g_{ex} derived from $C_{P,\mathrm{ex}}$ involves only enthalpy levels associated with conformational but not other transitions, and that $g_{\mathrm{ex}} \geq 0$. Because $C_{P,\mathrm{ex}}$ is obtained by two subtractions, there is no a priori guarantee for g_{ex}'s positive definiteness. Thus, the conventional assumption of $g_{\mathrm{ex}} \geq 0$ should be viewed as an invocation of Occam's razor. In this sense, what the subtraction scheme aims to achieve is an effective partition function much simpler, i.e., involving much fewer enthalpy levels than the original physical partition function of protein plus solvent. On the other hand, it is also assumed that no significant enthalpic fluctuation associated with conformational transitions is subtracted out by the conventional procedures. For experimental $C_{P,\mathrm{ex}}$ to be directly comparable with chain models with intrachain interactions, it should include enthalpic effects from all conformational transitions.

Therefore, in this article we do not consider nonconventional baseline construction schemes that subtract out not only enthalpic fluctuation in the solvent and those due to small-amplitude motions within a conformation but also contributions from larger scale conformational changes.[35] An example here would be to subtract $(C_P)_{\mathrm{D}}[\mathrm{D}]$ from the total C_P, or to subtract $(C_P)_{\mathrm{D}}$ from the total C_P for $T > T_{1/2}$ in the upper panel of Figure 4. Although these constructs are conceptually illuminating and may be useful practically in other contexts, such nonconventional baseline subtractions do not suit our purpose of seeking a fundamental understanding of densities of states, because the resulting effective partition function they lead to would not provide the full enthalpic density of states of the chain. In any case, the calorimetric two-state criterion satisfied experimentally by many small single-domain proteins applies to $C_{P,\mathrm{ex}}$'s obtained by conventional subtraction procedures. In other words, the experimental calorimetric two-state results were *not* obtained by nonconventional baselines that subtract significantly more from the area under the heat capacity function curve than conventional subtractions using either the linear extrapolation[121] or the sigmoidal-function weighted baseline[19] techniques. Hence, it is important to first conduct direct comparisons between theory and experiment without introducing nonconventional subtraction schemes in the analyses of model heat capacity functions. This is the approach we undertake in the present work. In an upcoming report,[58] we will analyze the physical implications of baseline subtractions in further detail.

From Partition Functions to Densities of States—on the Feasibility of "Assumption-Free" Inverse Laplace Transforms

We now turn to the possibility of deducing the underlying density of states g_{ex} from the transition excess heat capacity function $C_{P,\mathrm{ex}}$ and use an example to show that any such procedure involves assumptions about what functional form of the density of states is deemed acceptable. This is important for addressing the information content of calorimetric data. Numerical integrations of $C_{P,\mathrm{ex}}$ to yield an effective (excess) partition function is straightforward and well defined [Eqs. (30) and (31)]. We call this (effective) partition function Q, introduce the variable $\beta = 1/(k_B T)$, and, without loss of generality, shift the origin of enthalpy to the lowest enthalpy so that Eq. (25) becomes

$$Q(\beta) = \int_0^\infty dH g(H) e^{-\beta H}, \tag{35}$$

where $Q(\beta) = Q(T = \{k_B\beta\}^{-1})$. In this form, the partition function is easily recognized as the Laplace transform of the density of states,

$$Q(\beta) = \hat{\mathcal{L}}\{g(H)\}, \tag{36}$$

where H and β correspond respectively to the conventional variables t and s for Laplace transforms,[141]

$$\hat{\mathcal{L}}\{F(t)\} \equiv \int_0^\infty dt e^{-st} F(t). \tag{37}$$

This suggests that the density of conformational states may be obtained from the experimentally determined excess partition function by a general numerical inverse Laplace transform, viz.,

$$g(H) = \hat{\mathcal{L}}^{-1}\{Q(\beta)\}. \tag{38}$$

Attempts have been made at applying numerical inversion of Laplace transforms to a wide range of data analyses (see, e.g., Schnedermann[142]). Unfortunately, in the absence of additional constraints, general inverse Laplace transform is an ill conditioned problem, in that the solution is nonunique.[143] Because this issue is critical to the present analysis, we now illustrate this particular mathematical property of partition functions by attempting a general numerical inverse Laplace transform, using the calorimetric data of Jackson et al.[144] on chympotrypsin inhibitor 2 (CI2) for our test. These data have been discussed above in a two-state context (Fig. 1).

We take a matrix-inversion approach to Laplace-invert a discretized partition function. The discretization aspect of this method bears some similarities to the Legendre polynomial expansion technique discussed in Bellman et al.[145] In view of the viability of a two-state interpretation with a single non-native enthalpy level at $\Delta H_{\rm cal} \approx 60.0$ kcal mol^{-1} (Fig. 1), an upper bound $H_{\rm max}$ (>0) on the enthalpy of this system is assumed:

$$Q(\beta) = \int_0^{H_{\rm max}} dH g(H) e^{-\beta H}. \tag{39}$$

Next, we use a discretized version of Q that replaces the integration in Eq. (39) by a summation over N enthalpy values,

$$Q(\beta_j) = \sum_{i=1}^{N} \tilde{g}(H_i) e^{-\beta_j' H_i}, \tag{40}$$

to analyze available experimental data in a temperature range from T_0 to T_1. Let $\beta_0 = (k_B T_0)^{-1}$ and $\beta_1 = (k_B T_1)^{-1}$. In Eq. (40), $\beta_j' = \beta_j - \beta_1 = (j-1)(\beta_0 - \beta_1)/(N-1), H_i = (i-1)H_{\rm max}/(N-1)$, and

$$\tilde{g}(H_i) = g(H_i) e^{-\beta_1 H_i}. \tag{41}$$

Equation (40) may now be rewritten in matrix notations,

$$\mathbf{Q} = \hat{\mathbf{L}}\tilde{\mathbf{g}}, \tag{42}$$

in which $\mathbf{Q}$ and $\tilde{\mathbf{g}}$ are N-vectors, $\mathbf{Q}_j = Q(\beta_j)$, $\tilde{\mathbf{g}}_i = \tilde{g}(H_i)$, and $\hat{\mathbf{L}}$ is a symmetric $N \times N$ matrix, where $\hat{\mathbf{L}}_{ji} = \hat{\mathbf{L}}_{\mathbf{ij}} = \exp(-\beta_j' H_i)$. Then the inverse Laplace transform is provided by the relation

$$\tilde{\mathbf{g}} = \hat{\mathbf{L}}^{-1}\mathbf{Q}. \tag{43}$$

The symmetry of $\hat{\mathbf{L}}$ implies that it can always be diagonalized,

$$\hat{\mathbf{L}} = \hat{\mathbf{U}}^{\rm T}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}, \tag{44}$$

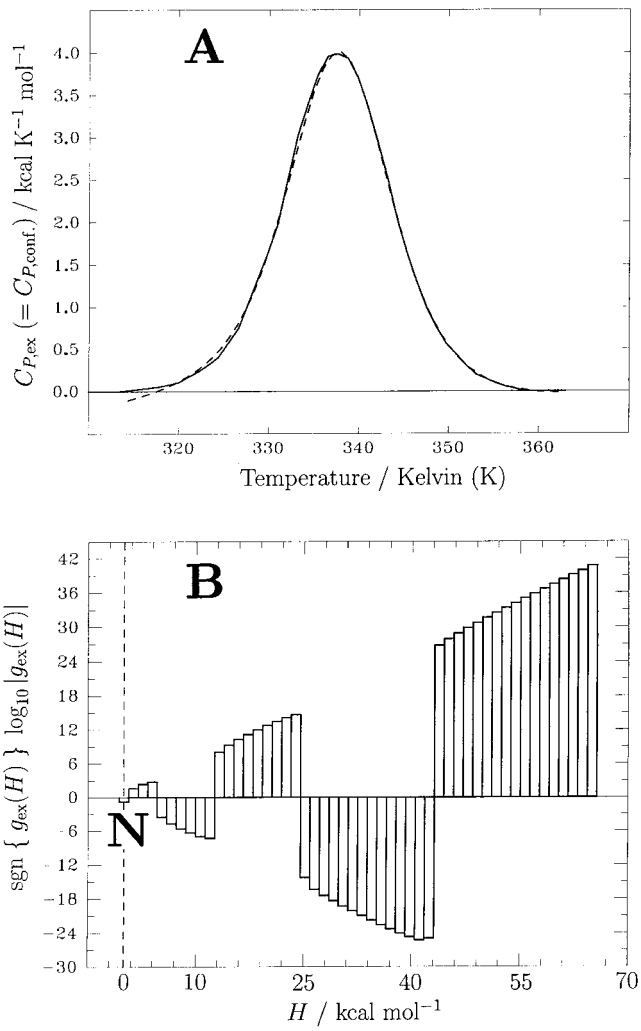

Fig. 22. Fitting the transition excess heat capacity of the Ile → Val76 mutant of chympotrypsin inhibitor 2 by a general matrix-based inverse Laplace transform technique. Notations are the same as that for Figure 1. Data from Jackson et al.[144] between T_0 = 41°C (314.15 K) and T_1 = 89°C (362.15 K) is analyzed. $H_{\rm max}$ = 65.0 kcal mol^{-1} is assumed, and N = 50 values of $g_{\rm ex}(H)$ are plotted. In the singular value decomposition, only reciprocals λ_i^{-1}'s of seven λ_i's with largest magnitudes are used. These λ_i's are 18.1, −7.71, 2.20, −0.500, 0.0942, −0.0150, 0.00206, for i = 1 − 7. λ_i^{-1}'s are set to zero for the other λ_i's with smaller magnitudes, ranging from $\lambda_8 = -2.45 \times 10^{-4}$ to $\lambda_{50} = 3.12 \times 10^{-18}$. In (**B**), sgn{$g_{\rm ex}(H)$} is the sign of $g_{\rm ex}(H)$, i.e., the logarithm of $g_{\rm ex}(H)$ is plotted if $g_{\rm ex} > 1$, but the negative logarithm of the absolute value of $g_{\rm ex}(H)$ is plotted if $g_{\rm ex} < -1$. ($|g_{\rm ex}| > 1$ for every $g_{\rm ex}$ obtained from this calculation.) Note that $g_{\rm ex}(H) < 0$ in (B) for H = 0, for the ranges H = 5.3 − 11.9 kcal mol^{-1} and H = 25.2 − 42.4 kcal mol^{-1}. Nevertheless, the quality of the fit in (**A**) (dashed curve) is comparable to the two-state fit in Figure 1.

where $\hat{\mathbf{U}}$ and $\hat{\mathbf{U}}^{\rm T}$ are a unitary matrix and its transpose. $\hat{\mathbf{\Lambda}}$ is a diagonal matrix with elements $\hat{\mathbf{\Lambda}}_{ij} = \delta_{ij}\lambda_i$, where $\delta_{ij} = 1$ for $i = j$, and zero otherwise. It follows that when the eigenvalues $\lambda_i \neq 0$ for all i, the inverse matrix of $\hat{\mathbf{L}}$ is determined by the reciprocals of λ_i's,

$$\hat{\mathbf{L}}^{-1} = \hat{\mathbf{U}}^{\rm T}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{U}}, \tag{45}$$

where $\hat{\mathbf{\Lambda}}_{ij}^{-1} = \delta_{ij}\lambda_i^{-1}$. Thus, we arrive at a solution to the inversion problem by first substituting $\hat{\mathbf{L}}^{-1}$ from Eq. (45)

into Eq. (43) to determine $\tilde{\mathbf{g}}$, then use Eq. (41) to obtain the density of states $g(H_i)$.

This method is applied to the CI2 data of Jackson et al.,[144] using $N = 50$. The fact that inverse Laplace transform is numerically ill conditioned is manifested by the very small magnitudes of an overwhelming majority of the eigenvalues λ_i's (see the legend to Fig. 22). This means that when a density of states is summed (or integrated) to yield a partition function, a lot of information is lost. In matrix/linear-transformation terms, a large subspace representing possible variations in the density-of-states function is effectively mapped onto the null space by $\hat{\mathbf{L}}$, i.e., giving practically zero contribution to the partition function.

This implies that any "assumption-free" inversion of $\hat{\mathbf{L}}$ is numerically unstable, because the λ_i's with exceedingly small magnitudes give rise to λ_i^{-1}'s with exceedingly large magnitudes in $\hat{\mathbf{\Lambda}}^{-1}$. To tackle this situation, we proceed to explore the standard singular value decomposition technique[146] as a tentative measure—by setting $\lambda_i^{-1} = 0$ for λ_i's with magnitudes smaller than a certain cutoff. Figure 22 shows one such solution. Although negative values for an effective g_{ex} should not be precluded according to the above discussion, the solution in Figure 22B is obviously artifactual and physically unreasonable. Nevertheless, it fits the original data very well (Fig. 22A) even in comparison with the two-state fit in Figure 1. The nonuniqueness of a general inverse Laplace transform is further highlighted by the fact that the g_{ex} function obtained by singular value decomposition is very sensitive to the numbers of nonzero λ_i^{-1}'s used. However, many of these very different g_{ex}'s provide equally good fits to the experimental $C_{P,\mathrm{ex}}$ (results not shown).

Therefore, there is no general "assumption-free" inversion technique to numerically recover with certainty the true underlying density of states from a given partition function. Additional assumptions, such as a maximum informational entropy principle,[147] are needed to constrain the solutions to some preferred forms. The robust and useful deconvolution technique that uses nonlinear least squares optimization[17,19,148] to obtain physically reasonable density of states is no exception, because this procedure is designed to "begin by assuming the simplest situation, i.e., the two-state case" and then "progressively increase the number of states"[19] (see also discussion in Zhou et al.[35]).

ACKNOWLEDGMENTS

I thank Sarina Bromberg and Ken Dill for introducing me to protein thermodynamic cooperativity. I thank them and Yawen Bai, Wayne Bolen, Ernesto Freire, Kent Kirshenbaum, Themis Lazaridis, Kip Murphy, Jose Sanchez-Ruiz, Tobin Sosnick, Maria-Luisa Tasayco, Dev Thirumalai, and Dong Xie for discussions and guidance in many occasions. I also thank Julie Forman-Kay, Lewis Kay and Régis Pomès for their critical reading of the manuscript and very helpful comments.

REFERENCES

1. Chan HS. Protein folding: matching speed and locality. Nature 1998;392:761–763.
2. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry 1985;24:1501–1509.
3. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci USA 1987;84:7524–7528.
4. Klimov DK, Thirumalai D. Factors governing the foldability of proteins. Proteins 1996;26:411–441.
5. Klimov DK, Thirumalai D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. Fold Design 1998;3:127–139.
6. Chan HS, Dill KA. Polymer principles in protein structure and stability. Annu Rev Biophys Biophys Chem 1991;20:447–490.
7. Dill KA, Shortle D. Denatured states of proteins. Annu Rev Biochem 1991;60:795–825.
8. Karplus M, Shakhnovich E. Protein folding: theoretical studies of thermodynamics and dynamics. In: Creighton TE, editor. Protein folding. New York: Freeman; 1992. p 127–195.
9. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–195.
10. Dill KA, Stigter S. Modeling protein stability as heteropolymer collapse. Adv Protein Chem 1995;46:59–104.
11. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–602.
12. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
13. Shakhnovich EI. Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
14. Thirumalai D, Klimov DK. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. Curr Opin Struct Biol 1999;9:197–207.
15. Privalov PL, Khechinashvili NN. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. J Mol Biol 1974;86:665–684.
16. Privalov PL, Potekhin SA. Scanning microcalorimetry in studying temperature-induced changes in proteins. Methods Enzymol 1986;131:4–51.
17. Freire E, Biltonen RL. Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and application to homogeneous systems. Biopolymers 1978;17:463–479.
18. Freire E. Statistical thermodynamic analysis of the heat capacity function associated with protein folding-unfolding transitions. Comments Mol Cell Biophys 1989;6:123–140.
19. Freire E. Differential scanning calorimetry. In: Shirley BA, editor. Methods in molecular biology. Vol 40. Protein stability and folding: theory and practice. Totowa, New Jersey: Humana Press Inc., 1995. p 191–218.
20. Lumry R, Biltonen R, Brandts JF. Validity of the "two-state" hypothesis for conformational transitions of proteins. Biopolymers 1966;4:917–944.
21. Brandts JF. Conformational transitions of proteins in water and in aqueous mixtures. In: Timasheff SN, Fasman GD, editors. Structure and stability of biological macromolecules. New York: Marcel Dekker, Inc., 1969. p 213–290.
22. Privalov PL. Stability of proteins: proteins which do not present a single cooperative system. Adv Protein Chem 1982;35:1–104.
23. Socci ND, Onuchic JN. Folding kinetics of protein-like heteropolymers. J Chem Phys 1994;101:1519–1528.
24. Chan HS, Bromberg S, Dill KA. Models of cooperativity in protein folding. Philos Trans R Soc London B 1995;348:61–70.
25. Gutin AM, Abkevich VI, Shakhnovich EI. A protein engineering analysis of the transition state for protein folding: Simulation in the lattice model. Fold Design 1998;3:183–194.
26. Chan HS. Modeling protein folding by Monte Carlo dynamics: Chevron plots, chevron rollover, and non-Arrhenius kinetics. In: Grassberger P, Barkema G, 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. Kolinski A, Godzik A, Skolnick J. A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: application to designed helical proteins. J Chem Phys 1993;98:7420–7433.
29. Kolinski A, Galazka W, Skolnick J. On the origin of the cooperativity of protein folding: implications from model simulations. Proteins 1996;26:271–287.
30. Bromberg S, Dill KA. Sidechain entropy and packing in proteins. Protein Sci 1994;3:997–1009.
31. Hao M-H, Scheraga HA. Monte Carlo simulation of a first-order transition for protein folding. J Phys Chem 1994;98:4940–4948.
32. Hao M-H, Scheraga HA. Statistical thermodynamics of protein folding: sequence dependence. J Phys Chem 1994;98:9882–9893.
33. Hao M-H, Scheraga HA. Molecular mechanisms for cooperative folding of proteins. J Mol Biol 1998;277:973–983.
34. Mohanty D, Kolinski A, Skolnick J. De novo simulations of the folding thermodynamics of the GCN4 leucine zipper. Biophys J 1999;77:54–69.
35. Zhou Y, Hall CK, Karplus M. The calorimetric criterion for a two-state process revisited. Protein Sci 1999;8:1064–1074.
36. Goldstein RA, Luthey-Schulten ZA, Wolynes PG. Optimal protein-folding codes from spin-glass theory. Proc Natl Acad Sci USA 1992;89:4918–4922.
37. Bryngelson JD, Wolynes PG. Intermediates and barrier crossing in a random energy model (with applications to protein folding). J Phys Chem 1989;93:6902–6915.
38. Bryngelson JD, Wolynes PG. A simple statistical field theory of heteropolymer collapse with application to protein folding. Biopolymers 1990;30:177–188.
39. Hilser VJ, Freire E. Structure-based calculation of the equilibrium folding pathway of proteins: correlation with hydrogen exchange protection factors. J Mol Biol 1996;262:756–772.
40. Baldwin RL. Folding intermediates in protein folding. BioEssays 1994;16:207–210.
41. Levinthal C. Are there pathways for protein folding? J Chim Phys 1968;65:44–45.
42. Levinthal C. How to fold graciously. In: Debrunner P, Tsibris JCM, Münck E, editors. Mössbauer spectroscopy in biological systems; Proceedings of a meeting held at Allerton House, Monticello, Illinois. Urbana, Illinois: University of Illinois Press, 1969. p 22–24.
43. Privalov PL. Intermediate states in protein folding. J Mol Biol 1996;258:707–725.
44. Zwanzig R, Szabo A, Bagchi B. Levinthal's paradox. Proc Natl Acad Sci USA 1992;89:20–22.
45. Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267:1619–1620.
46. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nature Struct Biol 1997;4:10–19.
47. Kim KS, Woodward C. Protein internal flexibility and global stability: effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. Biochemistry 1993;32:9609–9613.
48. Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. Proteins 1993;17:75–86.
49. Bai Y, Milne JS, Mayne L, Englander SW. Protein stability parameters measured by hydrogen exchange. Proteins 1994;20:4–11.
50. Bai Y, Sosnick T, Mayne L, Englander SW. Protein folding intermediates: native-state hydrogen exchange. Science 1995;269:192–197.
51. Bai Y, Englander SW. Future directions in folding: the multi-state nature of protein structure. Proteins 1996;24:145–151.
52. Chamberlain AK, Handel TM, Marqusee S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. Nat Struct Biol 1996;3:782–787.
53. Llinás M, Gillespie B, Dahlquist FW, Marqusee S. The energetics of T4 lysozyme reveal a hierarchy of conformations. Nat Struct Biol 1999;6:1072–1077.
54. Milne JS, Xu Y, Mayne LC, Englander SW. Experimental study of the protein folding landscape: unfolding reactions in cytochrome *c*. J Mol Biol 1999;290:811–822.
55. Gillespie JR, Shortle D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. J Mol Biol 1997;268:170–184.
56. Mok YK, Kay CM, Kay LE, Forman-Kay J. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. J Mol Biol 1999;289:619–638.
57. Englander SW, Mayne L, Bai Y, Sosnick TR. Hydrogen exchange: the modern legacy of Linderstrøm-Lang. Protein Sci 1997;6:1101–1109.
58. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000; in press.
59. Ikegami A. Structural changes and fluctuations of proteins. Biophys Chem 1977;6:117–130.
60. Freire E, Murphy KP. Molecular basis of co-operativity in protein folding. J Mol Biol 1991;222:687–698.
61. Freire E, Murphy KP, Sanchez-Ruiz JM, Galisteo ML, Privalov PL. The molecular basis of cooperativity in protein folding: thermodynamic dissection of interdomain interactions in phosphoglycerate kinase. Biochemistry 1992;31:250–256.
62. Honig B, Cohen FE. Adding backbone to protein folding: Why proteins are polypeptides. Fold Design 1996;1:R17–R20.
63. Crippen GM, Ohkubo YZ. Statistical mechanics of protein folding by exhaustive enumeration. Proteins 1998;32:425–437.
64. Shakhnovich EI. Modeling protein folding: the beauty and power of simplicity. Fold Design 1996;1:R50–R54.
65. Rapaport DC. The art of molecular dynamics simulation. New York: Cambridge University Press, 1995. p xi, 1–5.
66. Chen B-L, Schellman JA. Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. Biochemistry 1989;28:685–691.
67. Dill KA, Alonso DOV, Hutchinson K. Thermal stabilities of globular proteins. Biochemistry 1989;28:5439–5449.
68. Robertson AD, Murphy KP. Protein structure and the energetics of protein stability. Chem Rev 1997;97:1251–1267.
69. Garel T, Orland H. Mean-field model for protein folding. Europhys Lett 1988;6:307–310.
70. Shakhnovich EI, Gutin AM. Implications of thermodynamics of protein folding for evolution of primary sequences. Nature 1990;346:773–775.
71. 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 Peptide Protein Res 1975;7:445–459.
72. Makhatadze GI, Privalov PL. Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. J Mol Biol 1990;213:375–384.
73. Privalov PL, Makhatadze GI. Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects. J Mol Biol 1990;213:385–391.
74. Makhatadze GI, Privalov PL. Energetics of protein structure. Adv Protein Chem 1995;47:307–425.
75. Camacho CJ, Thirumalai D. Kinetics and thermodynamics of folding in model proteins. Proc Natl Acad Sci USA 1993;90:6369–6372.
76. Tiktopulo EI, Bychkova VE, Rička J, Ptitsyn OB. Cooperativity of the coil-globule transition in a homopolymer—microcalorimetric study of poly(*N*-isopropylacrylamide). Macromolecules 1994;27:2879–2882.
77. Murphy KP, Freire E. Thermodynamics of structural stability and cooperativity folding behavior in proteins. Adv Protein Chem 1992;43:313–361.
78. Freire E, Haynie DT, Xie D. Molecular basis of cooperativity in protein folding. IV. CORE: a general cooperative folding model. Proteins 1993;17:111–123.
79. Xie D, Freire E. Molecular basis of cooperativity in protein folding. V. Thermodynamic and structural conditions for the stabilization of compact denatured states. Proteins 1994;19:291–301.
80. Finkelstein AV, Badredtinov AY. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. Fold Design 1997;2:115–121.
81. Wolynes PG. Folding funnels and energy landscapes of larger proteins within the capillarity approximation. Proc Natl Acad Sci USA 1997;94:6170–6175.
82. Galzitskaya OV, Finkelstein AV. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. Proc Natl Acad Sci USA 1999;96:11299–11304.
83. Shoemaker BA, Wolynes PG. Exploring structures in protein

folding funnels with free energy functionals: the denatured ensemble. J Mol Biol 1999;287:657–674.
84. Shoemaker BA, Wang J, Wolynes PG. Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble. J Mol Biol 1999;287:675–694.
85. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
86. D'Aquino JA, Gómez J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. Proteins 1996;25:143–156.
87. Miller DW, Dill KA. A statistical mechanical model for hydrogen exchange in globular proteins. Protein Sci 1995;4:1860–1873.
88. Šali A, Shakhnovich E, Karplus M. Kinetics of protein folding: a lattice model study of the requirements for folding to the native state. J Mol Biol 1994;235:1614–1636.
89. Šali A, Shakhnovich E, Karplus M. How does a protein fold? Nature 1994;369:248–251.
90. Chan HS. Kinetics of protein folding. Nature 1995;373:664–665.
91. Chan HS, Dill KA. Comparing folding codes for proteins and polymers. Proteins 1996;24:335–344.
92. Shirley BA, Stanssens P, Hahn U, Pace CN. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. Biochemistry 1992;31:725–732.
93. Myers JK, Pace CN. Hydrogen bonding stabilizes globular proteins. Biophys J 1996;71:2033–2039.
94. Habermann SM, Murphy KP. Energetics of hydrogen bonding in proteins: a model compound study. Protein Sci 1996;5:1229–1239.
95. Zou Q, Habermann-Rottinghaus SM, Murphy KP. Urea effects on protein stability: Hydrogen bonding and the hydrophobic effect. Proteins 1998;31:107–115.
96. Dill KA. Dominant forces in protein folding. Biochemistry 1990; 29:7133–7155.
97. Honig B, Yang A-S. Free energy balance in protein folding. Adv Protein Chem 1995;46:27–58.
98. Lazaridis T, Archontis G, Karplus M. Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. Adv Protein Chem 1995;47:231–306.
99. Yang A-S, Honig B. Free energy determinants of secondary structure formation: I. α-helices. J Mol Biol 1995;252:351–365.
100. Yang A-S, Honig B. Free energy determinants of secondary structure formation. II. Antiparallel β-sheets. J Mol Biol 1995;252: 366–376.
101. Yang A-S, Hitz B, Honig B. Free energy determinants of secondary structure formation. III. β-turns and their role in protein folding. J Mol Biol 1996;259:873–882.
102. Takada S, Luthey-Schulten Z, Wolynes PG. Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and a random heteropolymer. J Chem Phys 1999;110: 11616–11629.
103. Dill KA. Additivity principles in biochemistry. J Biol Chem 1997;272:701–704.
104. Betancourt MR, Thirumalai D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. Protein Sci 1999;8:361–369.
105. Zimm BH, Bragg JK. Theory of phase transition between helix and random coil in polypeptide chains. J Chem Phys 1959;31:526–535.
106. Poland D, Scheraga HA. Theory of helix-coil transitions in biopolymers—statistical mechanical theory of order-disorder transitions in biological macromolecules. New York: Academic Press, 1970.
107. Ptitsyn OB, Kron AK, Eizner YuE. The models of the denaturation of globular proteins. I. Theory of globula-coil transitions in macromolecules. J Polymer Sci C 1968;16:3509–3517.
108. de Gennes P-G. Collapse of a polymer chain in poor solvents. Le Journal de Physique-Lettres (Paris) 1975;36:L55–L57.
109. Post CB, Zimm BH. Internal condensation of a single DNA molecule. Biopolymers 1979;18:1487–1501.
110. Grosberg AYu, Khokhlov AR. Physics of phase transitions in solutions of macromolecules. Physics Rev (Moscow) 1987;8:147–258.
111. Bastolla U, Grassberger P. Phase diagram of semi-stiff homopolymers. In: Grassberger P, Barkema G, Nadler W, editors. Monte Carlo approach to biopolymers and protein folding. Singapore: World Scientific, 1998. p 213–219.
112. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
113. Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. Adv Protein Chem 1988;39:191–234.
114. Roux B, Simonson T. Implicit solvent models. Biophys Chem 1999;78:1–20.
115. Lazaridis T, Karplus M. Effective energy function for protein in solution. Proteins 1999;35:133–152.
116. Lazaridis T, Karplus M. Heat capacity and compactness of denatured proteins. Biophys Chem 1999;78:207–217.
117. Alonso DOV, Dill KA. Solvent denaturation and stabilization of globular proteins. Biochemistry 1991;30:5974–5985.
118. Tanford C. Protein denaturation. Adv Protein Chem 1968;23:121–282.
119. Sosnick TR, Trewhella J. Denatured states of ribonuclease A have compact dimensions and residual secondary structure. Biochemistry 1992;31:8329–8335.
120. Hagihara Y, Hoshino M, Hamada D, Kataoka M, Goto Y. Chain-like conformation of heat-denatured ribonuclease A and cytochrome *c* as evidence by solution X-ray scattering. Fold Design 1998;3:195–201.
121. Privalov PL. Stability of proteins. Small globular proteins. Adv Protein Chem 1979;33:167–241.
122. Tanaka S, Scheraga HA. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. Macromolecules 1976;9:945–950.
123. Miyazawa S, Jernigan RL. Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
124. Sippl MJ. Knowledge-based potentials for proteins. Curr Opin Struct Biol 1995;5:229–235.
125. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: How accurate are they? J Mol Biol 1996;257:457–469.
126. Mirny LA, Shakhnovich EI. How to derive a protein folding potential? A new approach to an old problem. J Mol Biol 1996;264:1164–1179.
127. Mirny LA, Shakhnovich EI. Protein structure prediction by threading. Why it works and why it does not. J Mol Biol 1998;283:507–526.
128. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164–170.
129. Zhang L, Skolnick J. What should the *Z*-score of native protein structures be? Protein Sci 1998;7:1201–1207.
130. Shakhnovich EI, Finkelstein AV. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 1989;28: 1667–1680.
131. Ben-Naim A. Statistical thermodynamics for chemists and biochemists. New York: Plenum Publishing Corporation, 1992. p 10–12, 675–676.
132. Shortle D, Chan HS, Dill KA. Modeling the effects of mutations on the denatured states of proteins. Protein Sci 1992;1:201–215.
133. Ben-Naim A. Solvation thermodynamics. New York: Plenum Publishing Corporation, 1987. p 127–130, 201–203.
134. Chan HS, Dill KA. Solvation: Effects of molecular size and shape. J Chem Phys 1994;101:7007–7026.
135. Frye KJ, Perman CS, Royer CA. Testing the correlation between ΔA and ΔV of protein unfolding using m value mutants of staphylococcal nuclease. Biochemistry 1996;35:10234–10239.
136. Frye KJ, Royer CA. Probing the contribution of internal cavities to the volume change of protein unfolding under pressure. Protein Sci 1998;7:2217–2222.
137. Hummer G, Garde S, García AE, Paulaitis ME, Pratt LR. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. Proc Natl Acad Sci USA 1998;95:1552–1555.
138. Scheraga HA. Theory of hydrophobic interactions. J Biomol Struct Dyn 1998;16:447–460.
139. Yang D, Mok YK, Forman-Kay JD, Farrow NA, Kay LE. Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. J Mol Biol 1997;272: 790–804.

140. Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–1603.
141. Arfken G. Mathematical methods for physicists. New York: Academic Press, 1970. p 688–724.
142. Schnedermann E. The inverse Laplace transform as the ultimate tool for transverse mass spectra. Zeitschrift Für Physik C 1994;64:85–90.
143. Bertero M, Boccacci P, Pike ER. On the recovery and resolution of exponential relaxation rates from experimental data: a singular-value analysis of the Laplace transform inversion in the presence of noise. Proc R Soc Lond A 1982;383:15–29.
144. Jackson SE, Moracci M, elMasry N, Johnson CM, Fersht AR. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 1993;32:11259–11269.
145. Bellman R, Kalaba RE, Lockett JA. Numerical inversion of the Laplace transform: applications to biology, economics, engineering, and physics. New York: American Elsevier Publishing Company, Inc., 1966.
146. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in C. The art of scientific computing. Second Edition. New York: Cambridge University Press, 1997. p 59–65.
147. Plaza del Pino IM, Parody-Morreale A, Sanchez-Ruiz JM. Maximum entropy analysis of kinetic processes involving chemical and folding-unfolding changes in proteins. Anal Biochem 1997; 244:239–255.
148. Privalov G, Kavina V, Freire E, Privalov PL. Precise scanning calorimeter for studying thermal properties of biological macromolecules in dilute solutions. Anal Biochem 1995;232:79–85.
149. Chan HS, Dill KA. Intra-chain loops in polymers: effects of excluded volume. J Chem Phys 1989;90:492–509; Erratum: 1992;96:3361; 1997;107:10353.
150. Chan HS, Dill KA. Compact polymers. Macromolecules 1989;22: 4559–4573.
151. Thomas PD, Dill KA. Local and nonlocal interactions in globular proteins and mechanism of alcohol denaturation. Protein Sci 1993;2:2050–2065.