J Biol Phys (2012) 38:543–571
DOI 10.1007/s10867-012-9271-y

ORIGINAL PAPER

Thermodynamics of protein folding using a modified Wako-Saitô-Muñoz-Eaton model

Min-Yeh Tsai · Jian-Min Yuan · Yoshiaki Teranishi · Sheng Hsien Lin

Received: 26 July 2011 / Accepted: 7 May 2012 / Published online: 21 June 2012

Abstract Herein, we propose a modified version of the Wako-Saitô-Muñoz-Eaton (WSME) model. The proposed model introduces an empirical temperature parameter for the hypothetical structural units (i.e., foldons) in proteins to include site-dependent thermodynamic behavior. The thermodynamics for both our proposed model and the original WSME model were investigated. For a system with beta-hairpin topology, a mathematical treatment (contact-pair treatment) to facilitate the calculation of its partition function was developed. The results show that the proposed model provides better insight into the site-dependent thermodynamic behavior of the system, compared with the original WSME model. From this site-dependent point of view, the relationship between probe-dependent experimental results and model's thermodynamic predictions can be explained. The model allows for suggesting a general principle to identify foldon behavior. We also find that the backbone hydrogen bonds may play a role of structural constraints in modulating the cooperative system. Thus, our study may contribute to the understanding of the fundamental principles for the thermodynamics of protein folding.

Keywords WSME model · Protein · Beta-hairpin · Backbone hydrogen bond · Thermodynamics · Probe-dependent thermodynamic behavior · Site-dependent behavior · Foldon

M.-Y. Tsai (✉) · Y. Teranishi · S. H. Lin
National Chiao Tung University, 1001 Ta Hsuen Road, Hsinchu, Taiwan, Republic of China
e-mail: mytsai886@gmail.com

M.-Y. Tsai
Department of Chemistry, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, Republic of China

J.-M. Yuan
Department of Physics, Drexel University, Philadelphia, PA 19104, USA

1 Introduction

Proteins are complex macromolecules that are synthesized as linear chains of amino acid residues in cells. Specific primary sequences are encoded at the genetic level, where DNA is transcribed into RNA; the RNA is then translated into amino acid sequences at the ribosome, a process that generates polypeptide chains. This biological process is known as "the central dogma of molecular biology". Within the physiological environment, a linear polypeptide chain, made of amino acid residues, can spontaneously fold into self-organized, usually compact, globular and stable three-dimensional structures, referred to as the native protein. The protein native state allows these polypeptide chains to perform intricate biological functions. Anfinsen's thermodynamic hypothesis states that the native state is the most stable conformation among a large number of possible conformations available for a given polypeptide chain [1]. Beginning with Anfinsen's insights, many experimental and theoretical researches have been devoted to protein folding and significant advances have been made over the last few decades [2–8]. However, because of the size and extreme complexity of the heterogeneous interactions between the surrounding solvent and amino acid residues, the protein folding mechanism is still not fully understood.

The microscopic approaches to the protein folding problem on a quantum or molecular mechanics level were especially fruitful [9]. However, microscopic simulations of protein mechanics are still limited to timescales that are orders of magnitude shorter than biologically relevant timescales. Therefore, simplified statistical mechanical models have been used to investigate the thermodynamics and kinetics of protein folding [10–18]. These models can be viewed as generalizations of the classical Ising model [19], proved to be useful in thermodynamic and kinetic studies of complex systems [15, 20–23]. The tertiary structures of a protein's native folds have repeating sequences of structural motifs, α-helices, β-sheets, or hierarchically more complex foldons [5, 24]. At the same time, experimental studies have revealed that proteins often show two-state behavior [25]. Hence, Ising-like statistical models can form a theoretical framework to study protein folding. It is worth noting that more complex multistate models, based on the generalization of the classical Potts model [19, 26], can also be employed, though rare so far, to study protein folding [27, 28]. Among the above-mentioned models, the Wako-Saitô-Muñoz-Eaton (WSME) model has been applied to study the folding of many proteins [6, 7, 18, 29, 30, 33] and RNA molecules [31, 32]. The WSME model is a topology-based model in which a protein state is represented topologically by $\{x_1, x_2, \cdots, x_i, \cdots x_N\}$; x_i is a binary variable and $x_i = 1$ and $x_i = 0$ indicate, respectively, the folded (native) or unfolded state at the i^{th} local peptide bond. A two-state description for a single structural unit (e.g., a peptide bond or a residue side-chain) in a protein is used to study the folding and unfolding of a whole protein from a coarse-grained point of view. "Coarse-grained" herein refers to the fact that the molecular degrees of freedom for each residue in the protein such as vibrational, rotational and dihedral motions, among others, have been reduced such that only two states need to be specified physically (e.g., defined by residue's ϕ, ψ dihedral angles [6, 34]). Note that the folded and unfolded states are used to describe the local physical status for a given structural unit. This two-state description has been used and interpreted physically for approximately a decade [6, 35]. The WSME model can be briefly described as follows:

$$H(\{x\}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \epsilon_{ij} \Delta_{ij} \prod_{k=1}^{j} x_k - T \sum_{i=1}^{N} \Delta s_i x_i. \tag{1}$$

The first term denotes the contact energy between peptide bonds i and j; this is primarily attributed to the energetic (or enthalpic; PV work is ignored here) contribution ($\epsilon_{ij} < 0$) in the Hamiltonian, an effective free energy function. This stabilization energy is gained provided that a contact forms between peptide bonds i and j ($\Delta_{ij} = 1$ in this case; $\Delta_{ij} = 0$ otherwise) and all intervening peptide bonds, from bond i to j, are native. Note that Δ_{ij} denotes the element (i, j) for a contact matrix Δ, which embodies the geometric properties for a protein [36] (e.g., α-helices, β-hairpins). The second term represents the conformational entropy; this is primarily due to the entropic cost ($\Delta_{si} < 0$) by ordering peptide bond i in the native state.

The essence of the WSME model resides in the zipper-like energy component, which was first proposed by Wako and Saitô as an intra-island-interaction approximation [37, 38]. This model was further considered by Muñoz and coworkers, who introduced single, double, and triple sequence approximations [6, 7, 18, 39]. Since the work of Muñoz et al., many papers on the WSME model have appeared in the literature, including those seeking an exact solution for the equilibrium thermodynamics [14] and kinetics [40] of this model. While some of the literature on the WSME model studies the exact solution and its applications, we re-examine this model using experimental data. As a result, heuristic insights were obtained that result in the modification of the WSME model. The presented modified WSME model may be useful in providing experimental insights into the thermodynamics of protein folding. The concepts underlying this modification are detailed in the next paragraph.

In protein folding, researchers often analyze the thermodynamics of proteins using two-state models [41–43]. This approach assumes that, from a global structural point of view, a protein can occupy only two states: folded and unfolded. If the experimental fit for a result via the simple, two-state model is unsatisfactory, one may attribute the deviation to an intermediate state. However, when protein folding-unfolding is examined using various experimental techniques, such as absorption (Abs), circular dichroism (CD), fluorescence (Flu) and differential scanning calorimetry (DSC), to extract thermodynamic properties of a given protein (e.g., free energy difference and folding-transition temperature), these thermodynamic analyses may yield different results from experiment to experiment, depending on the protein studied. Some proteins, which have been examined as multi-state protein folders, e.g., Cytochrome c, show probe-dependent /site-specific properties and multi-exponential kinetics due to the existence of folding intermediates, while others are distinguished as two-state folders, e.g., GB1, SH3, which are characterized by their probe-independent thermodynamics as well as single-exponential kinetics. However, there has been a debate on the classification for two-state and downhill folders, which concerns primarily the existence of a folding barrier. Previously, the probe dependence issue has been analyzed and discussed, primarily for detecting downhill protein folding [39, 44–47]. Although the community used multi-probe spectroscopy and other techniques all together to elucidate local and global structures of proteins, the relationship between experimentally measured local structures and their corresponding thermodynamic properties was rarely discussed. Herein, a brief discussion on this issue follows. Cytochrome c (Cyt c) is a small, single domain, soluble protein, which has been widely studied using various experimental techniques and its thermodynamic properties are highly dependent on the particular technique used [48–55]. The free energy difference ($\Delta G = G_u - G_n$) obtained from the literature [48–55] is typically from 4 to 13 (kcal/mol) under physiological conditions, depending on the experimental techniques. The enthalpy difference ($\Delta H = H_u - H_n$), which can be directly measured from DSC, is approximately 100 (kcal/mol). One may conclude, according to the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$), that the entropy difference

($\Delta S = S_u - S_n$) is large enough to compensate for the gain in ΔH. However, upon careful examination of the ΔS calculated via spectroscopic measurements (e.g., Abs, CD, Flu), the value of ΔS is less than inferred from the DSC measurement (e.g., ~0.3 kcal/mol K from DSC, ~0.2 kcal/mol K from Flu and CD, ~0.15 kcal/mol K from Heme Abs). It is this contradiction between the DSC-inferred and spectroscopy-inferred ΔS that drew our attention to experimental probe dependence. Thus, we hypothesize that, for protein folding, the derived ΔG and ΔH measured using spectrometry actually describe local properties because a spectroscopic probe investigates only local properties. Thus, it is likely that different spectroscopic methods probe different positions in a protein and thus may generate different thermodynamic results. The ΔH measured by DSC is, however, a global property and its derived ΔG and ΔS should be discussed and interpreted on a global scale. Another example includes a β-hairpin. In their recent review, Scheraga and coworkers discussed a disagreement on the determination of the folding-transition temperature for β-hairpin-forming peptides [8]. They concluded that the disagreement arose because different structure-related features were monitored using different methods. All the above examples are heuristic observations that motivated us to include probe-dependent properties into a revised WSME model. This may serve as a model framework to clarify the relationship between multi-state, two-state, and downhill protein folders.

Five foldons in Cyt *c* were identified by Englander's group using hydrogen exchange (HX) experiments with site-specific monitoring of the exchange rates for the backbone amide hydrogen [5, 56–58]. A foldon originally referred to a discrete, contiguous section of a polypeptide chain consistent with the principle of minimum frustration [59]. This term was later extended to include any nucleation-competent sub-motif in a protein [60]. Herein, a foldon refers to a highly cooperative group that shows independent thermodynamic and kinetic behaviors; thus, one may classify, according to its thermodynamic behavior, the residues belonging to the same cooperative group as a macro-unit, referred to as a foldon by Englander [5, 56–58]. The foldon behavior for Cyt *c* was also confirmed by Shiu et al. using spectroscopy, such as UV-vis Abs, Flu, CD and small angle X-ray scattering (SAXS) [48]. Since the discovery of the foldon structure, there have been some related studies in the literature, including a zipper-like model involving non-additive coupling for Cyt *c* [16], a molecular dynamics (MD) simulation for Cyt *c* foldon behavior [61] as well as a hierarchical thermodynamic analysis of the global folding-unfolding transition and local cooperative fluctuations [62]. All of these results suggested that protein folding may be characterized as a hierarchy, that is, as a global and local folding scheme [48, 57]. To examine this hierarchical point of view, both the probe-dependent studies and the corresponding site-specific thermodynamic properties should be considered. The "protein foldon theory", though a decade old, is likely to close the gap between the probe-dependent and site-specific thermodynamic disparities. In other words, the site-specific thermodynamic behavior shown in the HX experiments should be related to the probe-dependent results [48, 63]. It should be noted, however, that so far the statistical mechanical foundation of the foldon theory has rarely been investigated. For this purpose, we propose to modify the original WSME model so that it can provide more insights into the probe-dependent and site-specific thermodynamic behaviors. This is equivalent to providing a statistical mechanical theory for the foldon picture.

This article primarily concerns the difference in the site-specific thermodynamic behavior between the two models: the WSME model and the model proposed herein; the latter accounts for site-dependent properties in terms of foldon units. The global folding scheme refers to the free energy balance throughout the entire protein ($\Delta G = \Delta H - T\Delta S$)

[6, 7, 18, 39, 44–47]. Within the local folding scheme, this balance is described for effective units (e.g., foldon units in the protein). The proposed model introduces an empirical temperature parameter $T_{1/2}$ for foldon units in the protein [see Appendix A for details], and the resulting enthalpic factor is used to balance the entropic cost from ordering each foldon. The balance for foldons gives rise to the local folding phenomenon. Thus, in principle, the temperature parameter can be measured directly from experiments. This type of information is missing from the original WSME model, which may render it difficult to properly associate local energetics with probe-dependent experiments. One may argue that a local foldon unit cannot fold or unfold without interacting with other units. Indeed, if one has to map a foldon physically to a residue unit such as a peptide bond or a side chain, its local folding seems to be unphysical. However, this physical-insight driven concept may not be easily linked to the interpretation of these probe-dependent experimental results since the relationship between the correlation (coupling) effect among residue units and experimental signals is unclear. On top of this, the original WSME model does not take into account the solvation effects, which so far have never been explicitly treated in the model. To empirically include these complicated solvation effects, the novel coarse-grained concept, foldon, is used to represent an effective unit, i.e., a structural unit, a motif, or even a whole protein, which implicitly accounts for all solvation-related and other fundamental forces that result in a folding-unfolding phenomenon. Our approach makes use of phenomenological parameters; thus, protein folding can be discussed practically and with greater experimental insights. Furthermore, the folding curve [see Fig. 1] for local folding shows nothing but the relative folded fraction of a foldon, not a 'real' folded state of the whole protein. In other words, all the coupling interactions such as electrostatic interactions, van der Waals force and H-bonds are implicitly included in the thermodynamic parameters. Therefore, these two approaches (the WSME model and our proposed model) should not contradict each other on this point—whether a single unit can fold without interactions. Note that the source of the entropic cost balancing discussed above differs from the WSME model [see Section 2], which globally attributes the entropic cost balancing to the interaction contact energy. Moreover, our model also retains the zipper-like contact energy from the original WSME model. Before examining more complicated protein systems, we first investigate a relatively simple system: β-hairpins, which have been extensively studied for their simple structural topology [64].

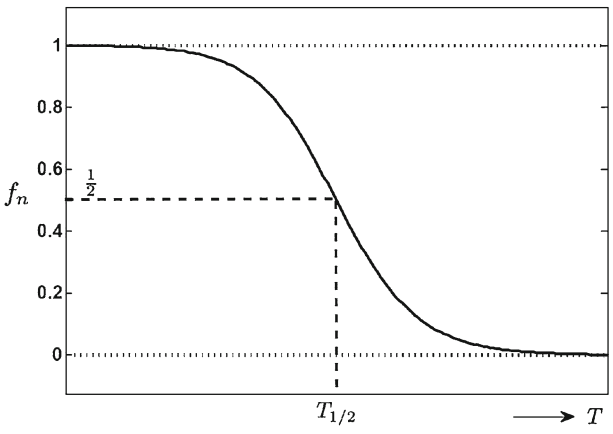

Fig. 1 The schematic diagram for the temperature-dependent two-state folding-unfolding process. The protein folding-unfolding behavior is associated with a sigmoidal transition, in which the mid-point temperature is characterized by $T_{1/2}$ (as commonly observed from experiments). At $T_{1/2}$, the fractions of both the folded and unfolded states are half of unity (i.e., 1/2). Note that the condition $\Delta S^0\left(T_{1/2}\right) > 0$ is used for this plot

The primary topics to address are: 1. We hope to gain a better understanding of protein local folding and the site-dependent thermodynamic behavior and, in particular, to compare the predictions for the thermodynamic behaviors of the two models mentioned above. The thermodynamic investigation is conducted by examining the thermodynamic folding fraction generated from both models using a different number of effective foldon units [see Section 2]. 2. For the comparison with experiments, the fluorescence and calorimetric data for the GB1 C-terminal β-hairpin (41–56) and the FRET data for a GB1 variant, GB1-m3p β-hairpin peptide, were numerically fit to both the WSME model and the model proposed herein. The goal is to show that the modified model provides a better connection to probe-dependent thermodynamic behaviors.

The paper is organized as follows. Section 2 introduces the basic concepts in protein folding thermodynamics and their relation to our proposed model; in Section 3, we report our numerical results; the implications of this study are discussed in Section 4; and finally, the conclusion is given in Section 5.

2 The model

Conventional protein folding thermodynamics Before we discuss the details of the proposed model, we briefly review the conventional treatment of the thermodynamics of protein folding. In the study of protein folding thermodynamics, the two-state model commonly used is characterized by a folded and unfolded state and can be written as a simple two-state reaction:

$$N \rightleftharpoons U \tag{2}$$

where N denotes the folded (or native) state and U, the unfolded state. It follows that

$$\frac{f_u}{f_n} = \frac{1 - f_n}{f_n} = K, \tag{3}$$

and that

$$f_n = \frac{1}{1 + K} = \frac{1}{1 + e^{-\Delta G^0/RT}}, \tag{4}$$

where $K = \exp\left(-\Delta G^0/RT\right)$; f_n is the fraction of the folded state and K is the equilibrium constant of the reaction. Note that the temperature dependence of ΔG^0 can be expanded as follows:

$$\Delta G^0\left(T\right) = \Delta G^0\left(T_{1/2}\right) + \left(\frac{\partial \Delta G^0}{\partial T}\right)_{T_{1/2}} \left(T - T_{1/2}\right) \tag{5}$$

or

$$\Delta G^0\left(T\right) = -\Delta S^0\left(T_{1/2}\right)\left(T - T_{1/2}\right) \tag{6}$$

where $\Delta G^0\left(T_{1/2}\right)$ is defined as the free energy of transition. From (6) and (4), it follows that

$$f_n = \frac{1}{1 + e^{\Delta S^0\left(T_{1/2}\right)\left(T - T_{1/2}\right)/RT}}. \tag{7}$$

A schematic diagram for temperature-dependent protein folding and unfolding is shown in Fig. 1.

The conventional treatment discussed above provides a crucial implication in formulating our proposed model, which includes a consideration of the local folding-unfolding behavior for each foldon unit in a protein [see Appendix A for details]. Our model is detailed in the following paragraph.

The modified WSME (M-WSME) model and contact-pair treatment of the β-hairpin The modified WSME (M-WSME) model is formulated as follows:

$$H(\{x\}) = \sum_{m=1}^{N-1} \sum_{n=m+1}^{N} \epsilon_{mn} \Delta_{mn} \prod_{k=m}^{n} x_k + \sum_{i=1}^{N} \Delta s_i \left(T - T_{1/2}\right) x_i, \tag{8}$$

where N (distinct from the native state) is the total number of foldon units (e.g., number of effective units), not the number of residues (or peptide bonds) and is expected to be a small number herein; ϵ_{mn} denotes the interaction between the m^{th} and the n^{th} foldon units (it is assumed that this energy is enthalpic, just like an enthalpic constraint, e.g., backbone H-bond formation/disruption and its accompanying effect such as burial/exposure of -NH and -CO to the interior/solvent); Δ_{mn} is the contact matrix that defines the protein's geometrical properties. For the second term, a characteristic temperature $T_{1/2}$ is included for each foldon unit; there is no physical equivalent of $T_{1/2}$ in the WSME model.[1] Herein, $T_{1/2}$ is physically related to the mid-point temperature observed in two-state protein folding thermodynamics and is considered the same for all foldon units. It is possible that $T_{1/2}$ can be generalized as T_i so that T_i can be different for different foldons to allow flexibility. In other words, we generalize the two-state thermodynamic behavior for each foldon unit in a protein to provide a link to probe-dependent experimental results. Furthermore, Δs_i is defined with a sign opposite ($\Delta s_i > 0$) to that in the WSME model, indicating the entropic cost for maintaining two-state folding-unfolding behavior for the foldon unit within a relevant temperature range (not a conformational entropy as appeared in the WSME model). Note that the second term in (8) was formulated based on (6) and that all foldon units involved should be characterized with an individual local folding-unfolding property. It should be noted that, under a thermal equilibrium condition for a given protein, neither of the two states (the folded and unfolded states) exist exclusively; instead, a thermodynamic fraction of the folded (or unfolded) state can be specified for *foldon units* in the protein. The fraction is then adopted to examine the relative stability of the states on the free energy level from the local folding scheme. The local folding-unfolding behavior of a foldon unit can be analogous to applying an external field to an effective spin unit, generating a degeneracy factor (more microstates are available) for the spin unit [65] [see (6) in the given reference]. Moreover, it should be noted that the contact energy terms in the WSME model shown in (1) were primarily used to account for hydrophobic interactions [6, 7, 18]. For the M-WSME model, the contact energy terms account for the enthalpic constraints, primarily backbone H-bonds; thus, only where the contact forms does a contribution from the binding energy appear. It is well-known that β-hairpins form a considerable number of backbone H-bonds. Additional interactions, such as hydrophobic and electrostatic salt-bridge interactions, among others, are implicitly attributed to the second term in (8). In other words, Δs_i in the M-WSME model is not only a conformational entropy, as in the WSME model, but also a phenomenological parameter

[1] Garcia-Mira et al. [39] adopted a concept of temperature of convergence in the WSME model, similar to the free energy expansion used herein. However, our approach emphasizes the local folding behavior for foldon units rather than native stretches.

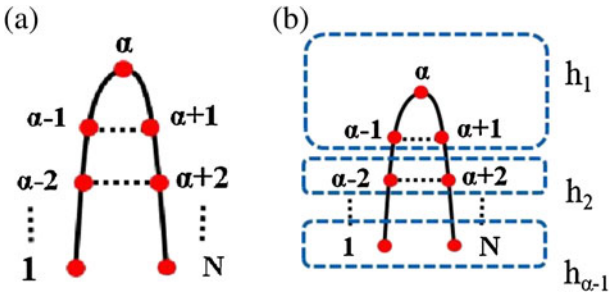

Fig. 2 A schematic representation of the geometrical shape for a β-hairpin with an odd N. The β-hairpin is described via (**a**) contact pair notation (see text) and (**b**) the classified regions

that implicitly accounts for all other interactions beyond H-bonding during folding. In this simplified model, one can assess the effect of the enthalpic constraint (most likely backbone H-bonds) on the system. This is the first step toward understanding the effect from one of the fundamental interactions in protein folding using statistical mechanics. Note that when $T_{1/2}$ approaches zero, we obtain the original WSME model as a limiting case (with its sign of Δs_i opposite to that of the M-WSME model).

For β-hairpins, taking *foldons* as effective units, the following criteria describe the contact matrix [36]:

$$\Delta_{mn} \begin{cases} = 1, & \text{if } m+n = N+1 \\ = 0, & \text{otherwise} \end{cases} . \tag{9}$$

Given that N is odd,[2] a schematic representation of the system is illustrated in Fig. 2(a). In Fig. 2(a), α is a new index used to symbolize the position where the foldon, located in the middle of the turn, mediates the two parallel β-strands on the sides. Thus, the first contact forms between the foldon pair $\alpha-1$ and $\alpha+1$ and the second contact forms between the pair $\alpha-2$ and $\alpha+2$ and so forth. It is known that the condition $N = 2\alpha-1$ is satisfied. The M-WSME model for the β-hairpin (with an odd N) can be recast as follows:

$$H = \sum_{j=1}^{\alpha-1} \left[\epsilon_{\alpha-j,\alpha+j} x_\alpha \prod_{k=1}^{j} x_{\alpha-k} \cdot x_{\alpha+k} + (T - T_{\alpha-j}) \Delta s_{\alpha-j} x_{\alpha-j} + (T - T_{\alpha+j}) \Delta s_{\alpha+j} x_{\alpha+j} \right] + (T - T_\alpha) \Delta s_\alpha x_\alpha \tag{10}$$

where subscript j denotes the index in which the j^{th} foldon contact pair is formed and the summation accounts for corresponding configurations of the system with $1, 2, \ldots, \alpha-1$ sequential contact pairs formed, respectively. Note that in (10), $\Delta_{\alpha-j,\alpha+j} = 1$ is used according to (9) and the subscript of $T_{\alpha-j}$, $T_{\alpha+j}$ and T_α emphasizes a possible generalization for a heterogeneous case; we assume that $T_{\alpha-j} = T_{\alpha+j} = T_\alpha = T_{1/2}$ herein. For the M-WSME model, (10) will be useful in calculating the partition function for the β-hairpin. The treatment for (10) is referred to as the contact-pair treatment, which is detailed in the following paragraph.

[2] Where N is even, and the foldons in the turn region of the β-hairpin may form an energetically unfavorable steric hindrance. Therefore, this case is not of interest (at least for the study herein).

The partition function with the Hamiltonian from (10) is expressed as

$$Z = \left(\sum_{x_\alpha} B_\alpha^{x_a}\right)\left(\sum_{x_{\alpha-1}}\sum_{x_{\alpha+1}} C_1^{x_\alpha x_{\alpha-1} x_{\alpha+1}} B_{\alpha-1}^{x_{\alpha-1}} B_{\alpha+1}^{x_{\alpha+1}}\right)\left(\sum_{x_{\alpha-2}}\sum_{x_{\alpha+2}} C_2^{x_\alpha x_{\alpha-1} x_{\alpha+1} x_{\alpha-2} x_{\alpha+2}} B_{\alpha-2}^{x_{\alpha-2}} B_{\alpha+2}^{x_{\alpha+2}}\right)\cdots$$
$$\cdots\left(\sum_{x_1}\sum_{x_N} C_{\alpha-1}^{x_\alpha x_{\alpha-1} x_{\alpha+1} x_{\alpha-2} x_{\alpha+2} \ldots x_1 x_N} B_1^{x_1} B_N^{x_N}\right) \tag{11}$$

where

$$B_i = \left(e^{\Delta s_i/k}\right)^{T_{1/2}/T-1} \tag{12}$$

and

$$C_j = e^{-\epsilon_{\alpha-j,\alpha+j}/kT}. \tag{13}$$

Note that in (11), the partition function, Z, includes many parentheses and each parenthesis includes factors that represent weights for the configurations for either a foldon or a foldon pair. For instance, in the first parenthesis, the summation $(1 + B_\alpha)$ indicates the sum of weights over the configurations 0 and 1, respectively, for the α^{th} foldon. Next, the second parenthesis accounts for the configurations of the foldon pair $(\alpha - 1, \alpha + 1)$ and its summation reads $\left(1 + B_{\alpha-1} + B_{\alpha+1} + C_1^{x_\alpha} B_{\alpha-1} B_{\alpha+1}\right)$; the third parenthesis describes the configurations for the foldon pair $(\alpha - 2,\ \alpha + 2)$, which reads $\left(1 + B_{\alpha-2} + B_{\alpha+2} + C_2^{x_\alpha x_{\alpha-1} x_{\alpha+1}} B_{\alpha-2} B_{\alpha+2}\right)$ and so forth. Thus, clearly the value from the latter parenthesis depends on the configuration in the former parenthesis. For example, the weights for the pair $(\alpha - 1,\ \alpha + 1)$ depend on the configuration of the α^{th} foldon and the weights for the pair $(\alpha - 2,\ \alpha + 2)$ depend on the configuration of the α^{th}, $(\alpha - 1)^{th}$ and $(\alpha + 1)^{th}$ foldons and so forth. This dependency is a key observation, which facilitates an analytical formulation and calculation of the partition function [see Appendix B for details]. Therefore, the partition function can be expressed in terms of individual components, which unambiguously describe their physical values. All of these components are summarized as follows:

$$Z = Z_0 + Z_1 + Z_2 + \cdots + Z_j + \cdots + Z_{\alpha-2} + Z_{\alpha-1} \tag{14}$$

$$Z_0 = \prod_{k=1}^{\alpha-1}(1 + B_{\alpha-k})(1 + B_{\alpha+k}) + B_\alpha (1 + B_{\alpha-1} + B_{\alpha+1}) \cdot \prod_{k=2}^{\alpha-1}(1 + B_{\alpha-k})(1 + B_{\alpha+k}) \tag{15}$$

$$Z_1 = C_1 B_\alpha B_{\alpha-1} B_{\alpha+1}(1 + B_{\alpha-2} + B_{\alpha+2}) \cdot \prod_{k=3}^{\alpha-1}(1 + B_{\alpha-k})(1 + B_{\alpha+k}) \tag{16}$$

$$Z_2 = C_1 C_2 B_\alpha (B_{\alpha-1} B_{\alpha+1})(B_{\alpha-2} B_{\alpha+2})(1 + B_{\alpha-3} + B_{\alpha+3}) \cdot \prod_{k=4}^{\alpha-1}(1 + B_{\alpha-k})(1 + B_{\alpha+k}) \tag{17}$$

$$\vdots$$

$$Z_j = \left(\prod_{i=1}^{j} C_i B_{\alpha-i} B_{\alpha+i}\right) B_\alpha \left(1 + B_{\alpha-j-1} + B_{\alpha+j+1}\right) \cdot \prod_{k=j+2}^{\alpha-1}(1 + B_{\alpha-k})(1 + B_{\alpha+k}) \tag{18}$$

[for $j = 1$ to $(\alpha - 3)$]

$$\vdots$$

$$Z_{\alpha-2} = \left(\prod_{i=1}^{\alpha-2} C_i B_{\alpha-i} B_{\alpha+i}\right) B_\alpha \left(1 + B_1 + B_N\right) \tag{19}$$

$$Z_{\alpha-1} = \left(\prod_{i=1}^{\alpha-1} C_i B_{\alpha-i} B_{\alpha+i}\right) B_\alpha \tag{20}$$

where $Z_0, Z_1, Z_2, \ldots, Z_j, \ldots, Z_{\alpha-2}$ and $Z_{\alpha-1}$ denote the reduced partition functions that have, respectively, $0, 1, 2, \ldots, , \ldots j$, $(\alpha - 2)$ and $(\alpha - 1)$ sequential contact pairs.

Finally, we derive an averaged quantity $\langle x_i \rangle$, which is a very important physical quantity that describes the site-dependent properties of the folding-unfolding process for each foldon unit. This quantity is expressed as

$$\langle x_i \rangle = \frac{z_i\,(1)}{Z} = \frac{z_i\,(1)}{z\,(0) + z_i\,(1)} = \frac{1}{1 + \frac{z_i(0)}{z_i(1)}}, \tag{21}$$

where $z_i(0)$ and $z_i(1)$ denote the reduced partition function where the configuration for the i^{th} foldon unit is $x_i = 0$ and $x_i = 1$, respectively. Note that

$$Z = z_i\,(0) + z_i\,(1)\,. \tag{22}$$

It follows from (21) that $N = 1$

$$\langle x \rangle = \langle x_1 \rangle = \frac{(1)}{1 + B_1^{-1}}. \tag{23}$$

For $N = 3$, given that $B_1 = B_2 = B_3 = B$ (Δs_i is the same for all the three units),

$$\begin{aligned} \langle x \rangle &= \langle x_1 \rangle = \langle x_2 \rangle = \langle x_3 \rangle \\ &= \frac{1}{1 + \frac{(1+B)^2}{B(1+B) + B^2(1+C_1 B)}}. \end{aligned} \tag{24}$$

Equations (21) to (24) relate the thermal quantities for a system in relation to its partition function. The expression for the case with higher N is somewhat complicated and is not detailed herein.

Although the description of the system with a different number of foldon units seems to be arbitrary, the number is important to present some special thermodynamic behaviors pertaining to the β-hairpin topology of the system [see Fig. 5 for an example when $N = 13$]. In fact, the case $N = 1$ is equivalent to conventional two-state folding thermodynamics [see Appendix C], from which more detailed thermodynamics for the β-hairpin can be further investigated using a different N (i.e., $N = 3, 5, \ldots$). Note that the maximum value for N should not exceed the number of peptide bond units (or residues) in the system; otherwise, the excess of foldon number may become meaningless. In addition, a more complicated quantity, such as the correlation between units (e.g., $\langle x_1 x_2 \ldots \rangle$) can, accordingly, be treated numerically.[3]

[3] The numerical calculations for this study were performed using MATLAB version 7.6.0 (The MathWorks, Inc.).

Note that in the formulation and calculation of the partition function as well as its related thermodynamic quantities, the WSME model is equivalent to the M-WSME model where A_i is substituted for B_i, that is,

$$A_i = e^{\Delta s_i/k}, \quad \text{for the WSME model,} \tag{25}$$

and

$$B_i = A_i^{T_{1/2}/T-1} \quad \text{for the M-WSME model.} \tag{26}$$

3 Results

Three-foldon description ($N = 3$) for the folding of the β-hairpin Since a standard protocol (or experimental support) to identify the optimal number of effective foldon units for the β-hairpin is still not available, it is of interest to investigate thermodynamics using a different N. It turns out that the case $N = 3$ meets the minimal requirement for the foldon description with a coupling constant (between unit 1 and 3). In addition, the minimal case allows us to provide an analytical expression for the heat capacity equation [see Appendix D]. Thus, it is of interest to begin with this particular case. To understand the difference in thermodynamic behavior between the WSME and M-WSME models for the β-hairpin, we investigate the thermodynamic behavior for both models. First, a homogeneous version for both models was investigated; we assumed that $\Delta s_1 = \Delta s_2 = \Delta s_3 = \Delta s$. This resulted in an identical thermodynamic quantity $\langle x_i \rangle$; that is, $\langle x_1 \rangle = \langle x_2 \rangle = \langle x_3 \rangle = \langle x \rangle$ for both the WSME and the M-WSME model. This average quantity $\langle x \rangle$ as a function of temperature for the β-hairpin ($N = 3$) is presented in Fig. 3. As shown in Fig. 3, each panel, given a specific value for the interaction[4] $\epsilon_{13} = 0$, -1.1 or -3.0 (kcal/mol), contains a plot of the curves with $|\Delta s| = 0.0032$, 0.01 and 0.04 (kcal/mol K) [see Appendix C], indicating that, respectively, small, middle, and large values are assumed in the system. Note that ϵ_{13} accounts for the coupling interaction (enthalpic constraint) between unit 1 and 3 in the study. The results indicated that where $\epsilon_{13} = 0$, there is no sigmoidal curve for the WSME model and no folding-unfolding transition [see Fig. 3(a)]; however, for the M-WSME model, we obtained a sigmoidal curve and the smoothness (or sharpness) of the transition was characterized by the selected Δs [see Fig. 3(d)]. Note that no coupling interaction between units means the absence of the enthalpic constraint, while other interactions such as electrostatic, van der Waals and H-bonding (both backbone-water and backbone-backbone) are implicitly included in the thermodynamic parameters, as has been explained in Sec. II. In this case, the H-bonding can be accounted for through the entropic effect. For example, it was found that the net effect of switching from backbone-water H-bond to backbone-backbone H-bond is mainly entropic for the α-helix [17].

Given that $\epsilon_{13} \neq 0$, the impact of the enthalpic constraint (i.e., the backbone H-bond) on both the WSME and M-WSME models was also examined. The results showed that the mid-point of the sigmoidal curve (the temperature at which the fraction of the native state is 1/2) for the WSME model was shifted significantly toward high temperatures,

[4]The value $\epsilon_{13} = -1.1$ (kcal/mol) was referenced to the paper published by Muñoz et al. [7].

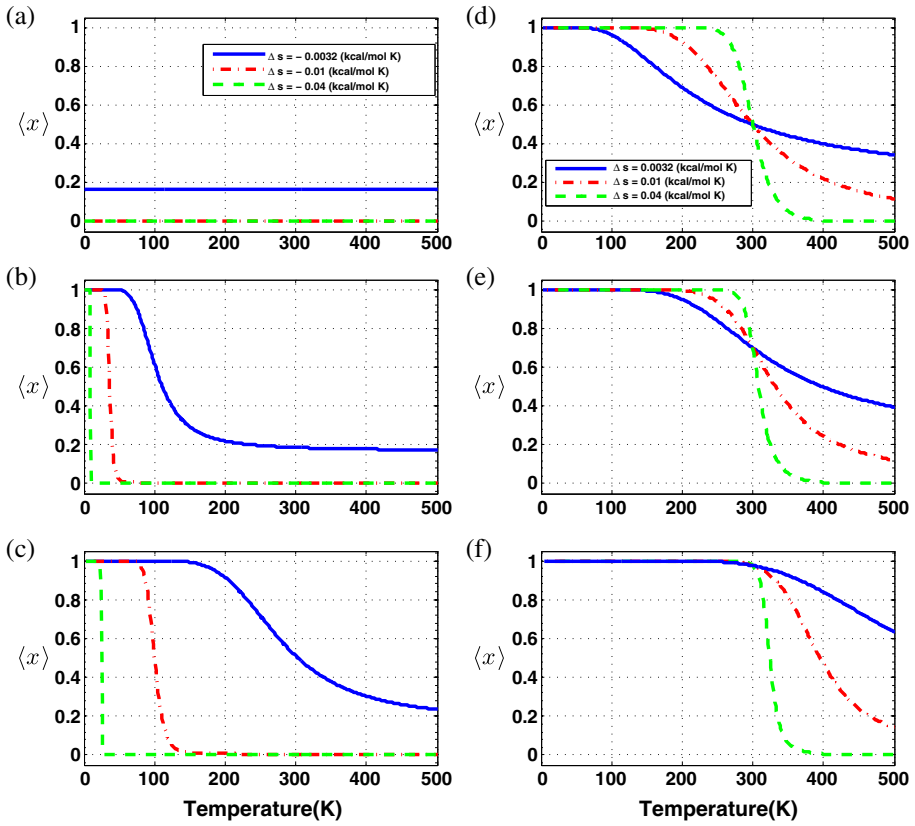

Fig. 3 Comparison of the thermodynamic native state fraction of each foldon unit for the WSME and M-WSME models with the minimal requirement ($N = 3$) for the β-hairpin. In each panel, the *blue solid*, *red dash-dotted* and *green dashed lines* denote, respectively, $|\Delta s| = 0.0032$, 0.01, and 0.04 (kcal/mol K). The WSME model: (**a**), (**b**) and (**c**) with $\epsilon_{13} = 0$, -1.1 and -3.0 (kcal/mol), respectively. The M-WSME model: (**d**), (**e**) and (**f**) with $\epsilon_{13} = 0$, -1.1 and -3.0 (kcal/mol), respectively; $T_{1/2} = 300$ K was used. Note that the homogeneous condition (see text) is applied to (**a**)–(**f**)

with decreasing $|\Delta s|$ and characterized by a distinct change in the transition sharpness. However, for the M-WSME model with the same $|\Delta s|$ and ϵ_{13}, we observed a sigmoidal curve that behaved smoother than that from the WSME model. In addition, we observed a temperature-crossing point, characterized by $T_{1/2}$ in the M-WSME model. These results demonstrated the significant differences resulting from our modifications for the M-WSME model. Our analysis suggests that the folding behavior from the WSME model relies on the interactions between units; therefore, the folding phenomenon requires a 'trigger' from these interactions. However, for the M-WSME model, the folding behavior is attributed to the free energy balance of each foldon unit; that is, the energetic and entropic factors simultaneously govern the local folding behavior of foldons in the system.

Presumably, the coupling interactions between units in the M-WSME model play an auxiliary role in forming the global native state and these interactions do not change the sigmoidal behavior of the curve significantly. This statement is consistent with the result

[see Appendix C], as the enthalpic constraints are not necessary for folding in the M-WSME model. Interestingly, it can be argued that the existence of backbone H-bonds may not play a deterministic role in the formation of native proteins (if backbone H-bonds contribute to most of the enthalpic constraints). From our investigation, it may only have a minor effect on the structural stabilization for β-hairpins.

Size effect As the total number of foldon units, N, increases (suggesting a longer β-hairpin), it can be expected that the thermodynamic behavior (the fraction of the native state) for each foldon varies with it. This dependency also reflects the special topology for the β-hairpin. In a homogeneous condition, $\Delta s_1 = \Delta s_2 = \cdots \Delta s$ and $\epsilon_{\alpha-j,\alpha+j} = \epsilon$, all of the foldons in the β-hairpin can be classified into different regions according to their thermodynamic behavior. For example, the thermodynamic behavior at the $(\alpha-1)^{th}$, α^{th} and $(\alpha+1)^{th}$ foldons share the same folding-unfolding pattern, as does the pair $(\alpha-2)^{th}$ and $(\alpha+2)^{th}$ and so forth. This consistent thermodynamic behavior stems from the symmetric shape and topology of the β-hairpin itself.

Thus, the β-hairpin can be divided into the following different regions: the h1 region (the turn region, where the first enthalpic constraint forms), the h2 region (where the second enthalpic constraint forms) and the $\ldots, h_\alpha - 1$ region (where the $(\alpha-1)^{th}$ enthalpic constraint forms) [see Fig. 2(b)]. The fraction of the native state of each foldon with a different N for both the WSME and M-WSME models are shown in Fig. 4 (only the behavior at h1 and h2 is shown). Figure 4, in principle, shows that, given the following parameters: $|\Delta s| = 0.04$ (kcal/mol K), $\epsilon = -1.1$ (kcal/mol) and $T_{1/2} = 300$ K, as suggested in Appendix C, the temperature range for the sigmoidal transition in the WSME model [Fig. 4(a) and (b), 5–20 K] is significantly different compared with the M-WSME model [Fig. 4(c) and (d), 200–400 K], regardless of the h1 and h2 regions. Moreover, our results also show that the curve shifts in the case of the WSME model [Fig. 4(a) and (b)] as the number of units increases, whereas in the M-WSME model [Fig. 4(c) and (d)], the slope of the curve becomes larger (near the lower temperature area, the higher temperature area remains nearly intact) when the number of units increases. The convergent thermodynamic behavior observed for the M-WSME model may provide insights into the special topology of the β-hairpin, which deserves further study. Note that in Fig. 4(b) and (d), the case $N = 3$ is not specified because the h2 region does not exist in this case.

In addition to investigating the effect on increasing the total number of units, we also examined thermodynamic behavior at different positions for a system with a fixed number of foldon units ($N = 13$). The idea is to investigate a system with its number of foldons equivalent to the number of peptide bonds existing in a short β-hairpin peptide as a limiting case; a foldon in the system can be seen as a peptide bond. The result is shown in Fig. 5, which shows that a convergent thermodynamic behavior, similar to that shown in Fig. 4, was also observed. This result further illustrates the influence of the special topology of the β-hairpin on its thermodynamic behavior and that experimental results with single residue resolution, such as NMR experiments, can be explained using our model. Under this assumption, we also suggest that the different thermodynamic behavior at different positions in the low-temperature range is associated with the foldon behavior discovered by Englander et al. [5, 56–58] using HX experiments. In other words, the free energy diversity observed for Cyt *c*'s foldon behavior in the low-temperature range [58] may be associated with the special topology of local folding in the protein. Herein, our result shows that the M-WSME model can provide a statistical mechanics interpretation to a foldon's divergent behavior as a function of temperature.

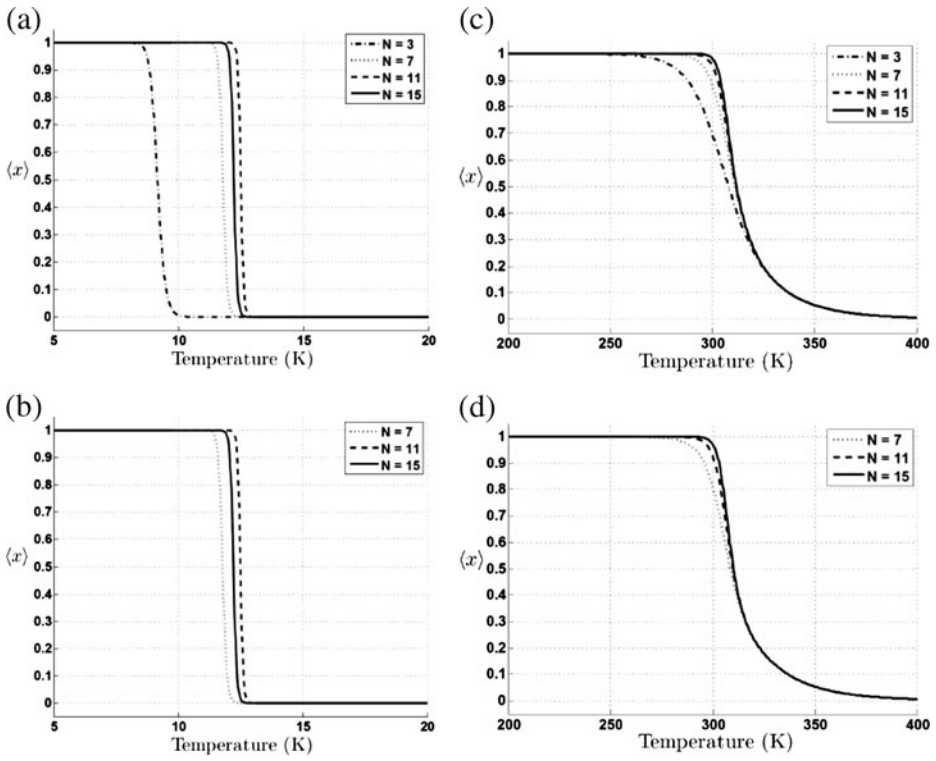

Fig. 4 The native state fraction of foldon unit in the β-hairpin at the h1 and h2 regions (see text). In each plot, the cases with a different N are compared; the *solid*, *dashed*, *dotted* and *dash-dotted* lines denote $N =$ 15, 11, 7, and 3, respectively. The WSME model: (**a**) the h1 region and (**b**) the h2 region. The M-WSME model: (**c**) the h1 region and (**d**) the h2 region. The parameters used in the calculations are as follows: $|\Delta s| = 0.04$ (kcal/mol K), $\epsilon = -1.1$ (kcal/mol) and $T_{1/2} = 300$ (K)

Comparison between the WSME and M-WSME models in the folding of a real β-hairpin
In the experimental studies of β-hairpin folding, the GB1 C-terminal β-hairpin (41–56) and its derivatives have been widely investigated [6, 7, 66–73]. The formation of the β-hairpin was thought to be associated with the formation of a hydrophobic cluster, which consists in the following four residues: Trp43, Tyr45, Phe52, and Val54. Thus, the population of the cluster (i.e., the correlation between these four residues) was used to approximate the population of the β-hairpin at an equilibrium condition [6, 7, 74]. Note that whether the experimental signals can reflect faithfully the correlation between the four residues in the cluster is beyond our discussion.

Herein, we compared the WSME and M-WSME models in the folding of the hairpin peptide using the approximation mentioned above. We compared the site-dependent thermodynamic properties from both models. The experimental data were from Muñoz et al. [7] and were digitized for numerical fitting. In our treatment, the cluster population was represented by $\langle x_3 x_5 x_{11} x_{13} \rangle$ because peptide bonds were used as our foldon units.[5]

[5]The method is somewhat different from that of Bruscolini et al. [74], who used $\langle x_3 x_5 x_{12} x_{14} \rangle$ as the population of the hydrophobic cluster formation, as they adopted the residue as the index.

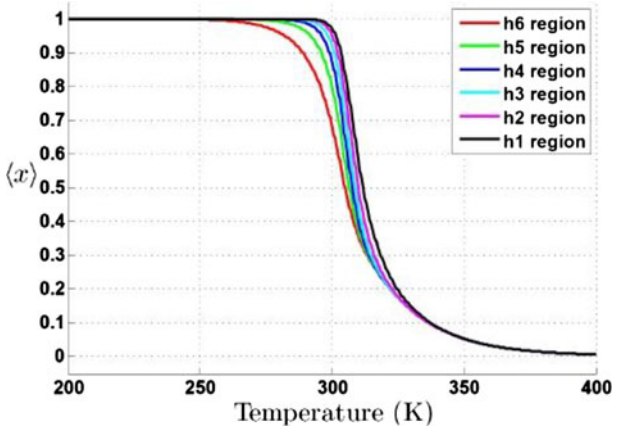

Fig. 5 The native fraction of the β-hairpin ($N = 13$) at different positions for the M-WSME model. The fractions from the different positions are color-coded sequentially; the *red line* (leftmost) denotes the h6 region, which includes the 1st and 13th foldon units; the *green line* denotes the h5 region, which includes the 2nd and 12th units; the *blue line* denotes the h4 region, which includes the 3rd and 11th units; the *cyan line* denotes the h3 region, which includes the 4th and 10th units; the *magenta line* denotes the h2 region, which includes the 5th and 9th units; and the *black line* (rightmost) includes the h1 region, which includes the 6th and 8th units. Note that $\Delta s = 0.04$ (kcal/mol K), $\epsilon = -1.1$ (kcal/mol) and $T_{1/2} = 300$ K for all units (the homogeneous condition)

A schematic diagram of the β-hairpin with the peptide bond as the index is shown in Fig. 6(a) and the fitting results[6] are shown in Fig. 6(b) and (c). For the WSME model, $\Delta s = -0.0036$ kcal/mol K and $\epsilon = -2.47$ kcal/mol were calculated and for the M-WSME model, $\Delta s = 0.0139$ kcal/mol K, $T_{1/2} = 342.4$ K and $\epsilon = -0.30$ kcal/mol were obtained [see Table 1]. Note that the value of $T_{1/2}$ predicted for the M-WSME model is somewhat unreasonable. This may be due to the use of the cluster assumption that we discussed above; thus, the prediction for $T_{1/2}$ is not of interest. To physically compare between both models, the number shown in the parentheses in Table 1 denotes the absolute value of the conformational entropy, which is originally defined in the WSME model while it is not in the M-WSME model. However, an inferred value can be obtained according to the two-state nature of the entire system, which assumes homogeneous Δs for each unit. Thus, the conformational entropy for the M-WSME model can be approximately obtained from the ratio Δs/N. The result shows that the conformational entropy obtained from the M-WSME model is almost four times smaller than that from the WSME model, suggesting a significant difference in the interpretation of the folding-unfolding sigmoidal transition. The WSME model emphasizes the coupling effect between units (-2.47 for GB1), which serves as an essential driving force in cooperative folding-unfolding; thus, its free energy balance requires more entropic cost (0.0036 for GB1). However, the M-WSME model attributes the

[6]The least-square fitting procedure was conducted using Matlab Curve Fitting Toolbox 1.2.1 (The MathWorks, Inc.).

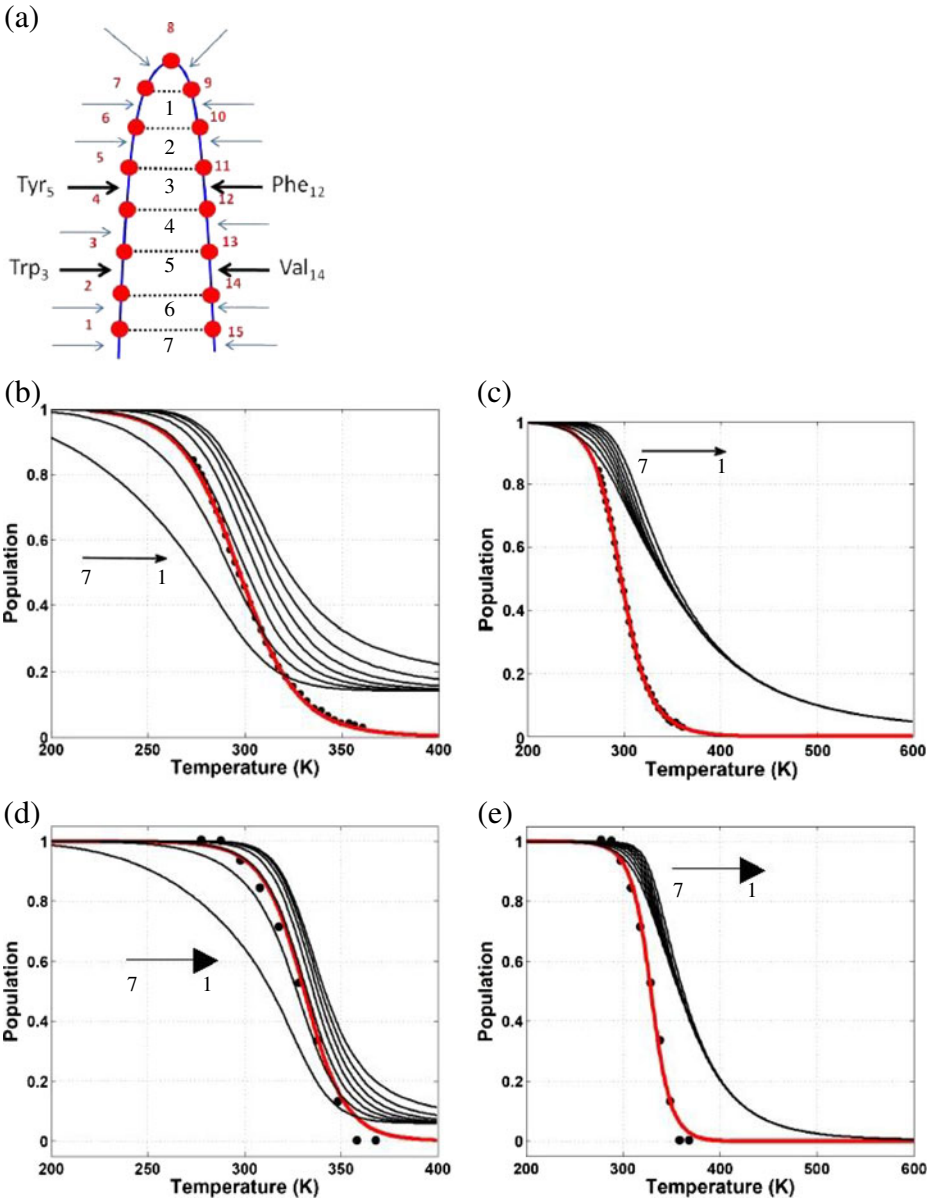

Fig. 6 Comparison between the WSME and M-WSME models in the unfolding of the β-hairpins selected (see below). (**a**) The schematic diagram of the β-hairpin with the peptide bond index. Note that the *arrows* highlight the positions of side-chain residues. (**b**) The WSME model. (**c**) The M-WSME model. The *dots* are digitized directly from Muñoz's paper (GB1 peptide) [7], which were derived from a two-state analysis of their fluorescence experiments and associated with the population of the hydrophobic cluster: Trp43, Tyr45, Phe52 and Val54. The *red line* denotes the numerical fit results from the thermal quantity, which represents the correlation among these corresponding peptide bond units. The *black lines* represent the site-dependent native fractions for all sites, which are aligned in order, according to the direction of the *arrow*: region h_7, ..., h_2, h_1 [see Fig. 2(b)]. Similarly, the fit results for the other peptide, the GB1-m3p, are given in (**d**) the WSME model and (**e**) the M-WSME model. Note that the *dots* are normalized from FRET efficiency from Du's paper (GB1-m3p) [73] (see Table 1 for the parameters fit by both models)

Table 1 The thermodynamic parameters fit for β-hairpin peptides using the WSME and M-WSME models

Model/Peptide	$T_{1/2}$ (K)	Δs (kcal/mol K)	ϵ (kcal/mol)
WSME/GB1	–	−0.0036 (0.0036)	−2.47
M-WSME/GB1	342.4	0.0139 (0.0009)	−0.30
WSME/GB1-m3p	–	−0.0055 (0.0055)	−3.96
M-WSME/GB1-m3p	357.8	0.0256 (0.0017)	−0.30

*Goodness of fit:
WSME/GB1: R-square = 0.9990
M-WSME/GB1: R-square = 0.9995
WSME/GB1-m3p: R-square = 0.9942
M-WSME/GB1-m3p: R-square = 0.9936
*Note that ϵ was fixed at bound −0.3 in the M-WSME/GB1-m3p fitting

cooperative transition to the two-state nature of the protein system on the foldon level, that is, the coupling effect is minor (−0.3 for GB1) and so is the entropic cost (0.0009 for GB1). Which of these two versions is more reasonable? Herein, we offer an alternative way to examine this question. Figure 6(b) and (c) reproduce the site-dependent thermodynamic properties of the WSME and M-WSME models, respectively, fit from the hydrophobic cluster population. The WSME model shows a pattern of divergent behavior, suggesting an entirely different behavior among all the sites; however, the M-WSME model shows a more convergent behavior. Although both models demonstrate site dependence, the WSME model does not show a complete unfolding curve at local sites (i.e., the curve levels off to a non-zero constant). The lack of complete unfolding may be due to high dependence on the coupling effect among all units, as can be understood by comparing the black lines [Fig. 6(b)] with the complete unfolding curve (red line) [Fig. 6(b)] generated from the population of the hydrophobic cluster. It can be seen that the complete unfolding behavior, to some extent, requires exact correlation (a product, $\langle x_3 x_5 x_{11} x_{13} \rangle$) among the four hydrophobic residues (units); this can explain the feature of the high coupling effect of the WSME model. We think these differences are interesting, as the formation of the β-hairpin is thought to be cooperative and certain corresponding site-dependent folding-unfolding completion is an expected observation. However, the results from the WSME model suggest that this model may not have considered the site-dependent behavior that we examined. Therefore, we believe that our proposed M-WSME model may provide an alternative way to study protein folding, as this modified model can capture the two-state nature along with protein's site-dependent features. As to the coupling effect in the M-WSME model, it may play a role of non-ideal behavior in thermodynamics [15] (with respect to the no-coupling, ideal case). This statement is based on our effective treatment on the foldon level. Note that we did not include parameters for the hydrophobic interactions in the WSME model fitting, as was originally included in Muñoz's approach [7]. These parameters were omitted in an attempt to directly relate the WSME model to the M-WSME model, where no correspondence to the hydrophobic interactions is specified [see Section 2]. However, by examining the Δs value (−0.0036) from our fitting results and comparing it with the value (−0.0032) fit by Muñoz et al. [7], it seems that our fitting is reasonable to a certain extent, despite the omission of a hydrophobic interaction description.

In addition to the GB1 C-terminal β-hairpin, a similar comparison was made for its variant, the GB1-m3p β-hairpin. The GB1-m3 peptide was designed with a much better thermal stability than the parent GB1 peptide [70]. Tucker et al. designed a new FRET pair on the GB1-m3 peptide by replacing the single Phe residue with a non-natural residue,

Phe_{CN}, where a cyano group is added at the para position of the Phe side chain [75]. The FRET pair is therefore formed between the Phe_{CN} (donor) and the Trp (acceptor). We compared the site-dependent thermodynamic properties of GB1-m3p from both models via the same protocol as the fit for the GB1 peptide. The experimental data (FRET efficiency) were from Du et al. [73] and the data were further normalized to the folding population for numerical fitting.[7] The fitting results are shown in Fig. 6(d), (e) and the parameters fit by both models are also summarized in Table 1.

The goodness of all the fits is given below in Table 1. The high R-square value ($\sim$0.99) calculated from the fitting of both the WSME and M-WSME models suggest that both models can successfully capture the feature of the folding-unfolding transition. Thus, one may not be able to judge which of the two versions is more reliable merely from the measures of fit quality.

Fitting a β-hairpin peptide to calorimetric data To provide a direct connection between the experiments and models discussed, calorimetric data from the GB1 C-terminal hairpin (41–56) were used to perform the least-square fit[8] to the models. Calorimetric data are from a differential scanning calorimetry (DSC) [55] experiment, which provides an important physical quantity: heat capacity as a function of temperature. The DSC data for the β-hairpin peptide were digitized[9] from the paper published by Honda et al. [68]. Results from the numerical fit to the heat capacity formulas, derived from the models [see Appendix D], are shown in Fig. 7. A comparison of the parameters fit by both the WSME and M-WSME models are listed in Table 2. Note that herein we assumed that the entire peptide behaves effectively as the number of foldons that we used for the system (i.e., $N = 1, 3, \ldots$). Although this is unrealistic compared with a real β-hairpin peptide, it serves as a preliminary test to examine the difference in the thermodynamic interpretation between the M-WSME and WSME models. Our results showed that, when $N = 3$, the values fit for ϵ_{13} from WSME (-11.27) and M-WSME (-0.74) are different, suggesting an entirely distinct thermodynamic interpretation for the peak transition in the DSC diagram. For the WSME model, the heat of transition (total enthalpy absorbed during transition) is obviously attributed to the interaction ϵ_{13}, while ϵ_{13} is small for the M-WSME model and the heat of transition is primarily due to the intrinsic enthalpic contribution from each foldon unit [see Appendix D]. This result is consistent with the discussion in the previous paragraph, where the enthalpic binding constraint may play only an auxiliary role in peptide folding. By examining results fit using a different N ($N = 1$ and 3) in the M-WSME model, we found that Δs and Δh decrease as N changes from 1 to 3. A change in these values is intuitive given that the heat of transition is equally distributed among all of the foldon units available in the system. Thus, Δh for each unit should decrease upon increasing the number of units.[10]

[7] The normalization procedure was performed according to the equation, $(Y_u - Y_{obs})/(Y_u - Y_n)$, where Y_{obs} denotes the observed FRET efficiency; Y_u (unfolded baseline) and Y_n (folded baseline) are assumed to be linearly dependent on temperature.

[8] The least-square fitting procedure was conducted using MATLAB Curve Fitting Toolbox 1.2.1 (The MathWorks, Inc.).

[9] The software Engauge Digitizer 4.1 was used to extract data points from the published papers.

[10] Whether the equivalence of heat and van't Hoff enthalpy is valid is not discussed herein.

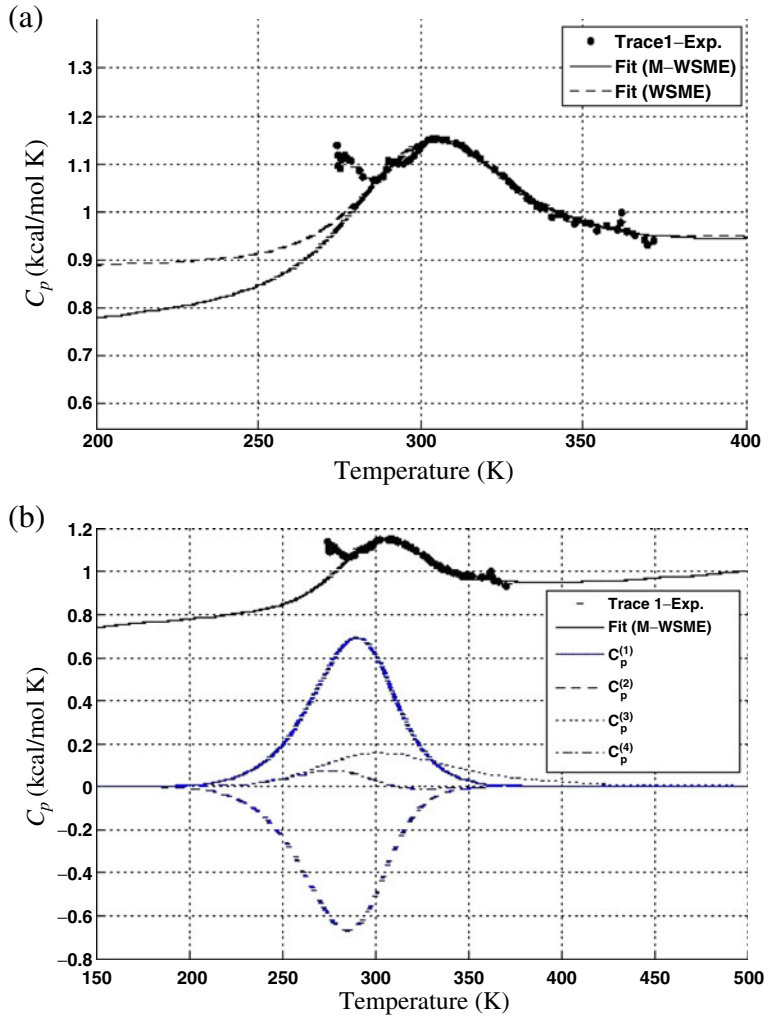

Fig. 7 The numerical fit results from both the WSME and M-WSME models ($N = 3$) for the GB1 β-hairpin peptide. (**a**) The *dots* denote the experimental data adapted from the paper published by Honda et al. [68]. Only the first trace (heating) is shown in the figure (there were six DSC traces in total, including three heating and three cooling experiments). The *solid line* denotes the numerical curve fit for the M-WSME model; the *dashed line* denotes the fit for the WSME model. (**b**) A demonstration of all the heat capacity components in the M-WSME model ($N = 3$) [see Appendix D]. The *black dots* and *solid line* are, respectively, the experimental data and fit results, which are also shown in (**a**). The *solid*, *dashed*, *dotted* and *dashed-dotted blue lines* represent, respectively, the first, second, third, and fourth components in the heat capacity equation (see Table 2 for the parameters fit by both models)

We also found that the ϵ_{13} value had only a minor effect (-0.74) in magnitude compared with Δh (6.2). This result supports the proposition that the interaction between the foldon units may not be a determining factor in forming global native topology (discussed in the beginning of this section for the thermodynamic investigation of the $N = 3$ case).

Table 2 The thermodynamic parameters fit for the calorimetric data from the GB1 β- hairpin peptide (41–56) using the WSME and M-WSME models. Note that only the cases for $N = 1$ and $N = 3$ are considered in this study

Model	WSME ($N = 1$)	WSME ($N = 3$)	M-WSME ($N = 1$)	M-WSME ($N = 3$)
Δs (kcal/mol K)	–	−0.012	0.038	0.021
$T_{1/2}$ (K)	–	–	304	300
Δh (kcal/mol)	–	–	11.5	6.2
ϵ_{13} (kcal/mol)	–	−11.27	–	−0.74

Note that $\Delta h = T_{1/2} \times \Delta s$

Note that the first column is absent (marked by "–"), as there is no fit equivalent for the WSME model when $N = 1$; this is also true for the other entries in the second and third columns

The fit parameters are the averaged values from the fitting results, in which experimental data were from six DSC traces [68]

Each fit was performed by adding a linear temperature-dependent function $a + b \cdot T$ to the heat capacity equations [see Appendix D] to adjust the baseline. a and b are constants (fit results not shown)

4 Discussion

Although the findings from this study are preliminary for the proposed M-WSME model, with the β-hairpin topology, several thermodynamic implications can be drawn. From the investigation of the β-hairpin, different regions can be identified according to their thermodynamic behavior. This behavior is closely related to the symmetric property of the β-hairpin. These findings suggest that the mean-field approach could be used to further simplify the system and that the possibility of its application should be examined. The mean-field theory has been widely used to study protein systems, including the relationship between probe-dependent experiments and models [15, 16, 48]; the free energy landscape and the dynamic behavior of the WSME model [36]; and the thermodynamic properties of the Galzitskaya–Finkelstein (GF) model and its applications [74, 76]. It is worth noting that, in some cases, the mean-field approach can relate thermodynamics and statistical mechanical models, thereby adding insight into the relationship between experimentally determined parameters and theoretical models [15]. It will be interesting to study the mean-field version of the present study.

Moreover, the findings on the diversity of the thermodynamic behavior in different regions of the β-hairpin concur with previous temperature-dependent HX studies [58] in explaining the thermodynamic behavior of each residue within a foldon, although quantitative comparison is not yet available. In addition, a MD simulation study on a β-hairpin further demonstrated the diversity of the thermodynamic behavior among all the backbone H-bond pairs, based on the calculation of their end-to-end distances (Tsai, M.Y., Yuan, J.M., Yamaki, M., Lin, S.H.: Molecular dynamics insight into thermodynamics of a beta-hairpin peptide (2012, unpublished)). All these results suggest that β-hairpins may be described thermodynamically at the foldon level. Further investigation is required to examine a quantitative description of the local thermodynamic behavior and its related energetics. Although the system (β-hairpin) in this study is small, the following recommendations could serve as general principles to identify foldon behavior using Ising-like statistical mechanical models. 1. Perform an investigation on the thermodynamic behavior of the local units in the system. 2. Group the units that show similar thermodynamic behavior. However, the model in its current form cannot predict which stretches belong to different foldons. In this study, we observed that similar thermodynamic behavior could be identified within the regions that maintain a specific geometrical symmetry (β-hairpin). The result

could be used to build a thermodynamic pattern for the site-dependent behavior of a β-hairpin, which we believe may provide an insight into the identification of a β-hairpin in proteins. It is thought that the relatively simple model system can be further improved and generalized to investigate more complex systems, such as a system with α/β motifs, thereby providing greater insight into the thermodynamic interpretation of detailed foldon behavior. The findings herein are a step in the direction of establishing a theoretical approach to interpret the experimental results for probe-dependent protein folders. As potential multiple pathways of protein folding can be understood by the energy landscape theory [77], it is worth noting that some proteins, such as the lambda repressor and WW domain, can be tuned by mutations to become different types of folders [78, 79] e.g., two-state, three-state and downhill etc. Because some of the mutants have been experimentally confirmed to demonstrate probe-dependent thermodynamics, it would be interesting to apply the M-WSME model to study the thermodynamic properties of these mutants.

Finally, a cooperative effect from the formation of the β-hairpin should be discussed. It was assumed that the source of the cooperativity of a β-hairpin primarily results from the contact interaction between the peptide bond units (e.g., backbone H-bond interactions and hydrophobic clustering) in the WSME model. However, this explicit treatment for the hydrophobic effect may yield an imperfection in the description of the site-dependent local properties. As shown in Fig. 6(b) and (d), the unfolding curves from the WSME model at the local sites generate rather divergent and incomplete folding-unfolding behaviors, which may not explain the cooperative behavior at the thermodynamic level. Our results from the M-WSME model introduce an assumption that the cooperative effect may be attributed to the two-state thermodynamic properties for each foldon unit (an enthalpic factor as the compensation for the entropic cost at the foldon level), which demonstrates the validity of the local folding behavior, verified from the probe-dependent thermodynamic behavior measured from experiments. In other words, in the M-WSME model, the cooperative phenomenon for the β-hairpin may be explained by the two-state nature of protein folding on the thermodynamic level, a combination of complex interactions (e.g., hydrophobic and salt-bridge interactions) implicitly considered in Δs_i for each foldon as a result of solvation effects. Using this approach, as the backbone H-bond interactions were defined solely as enthalpic constraints, our results show that backbone H-bonds only have a minor effect on the sigmoidal behavior (cooperative behavior) because of their small free energy contribution. This result is consistent with Scheraga and his coworkers, who pointed out that no clear evidence supports the presence of backbone H-bonds in non-turn regions [8, 66, 67, 80–84]. In other words, the backbone H-bond interaction may not contribute as much as expected in the cooperative formation of β-hairpins but may play only a structural as well as energetic constraint in modulation of the cooperative system [see Fig. 6(c) and (e)]. Recently, Bruscolini et al. [30] extended the original model to the WSME-S model, which includes solvation effects explicitly. The authors addressed the solvation effects using an expression of the temperature-dependent parameters in the original WSME model, which can be adjusted to the phenomenological behavior of the heat capacity of any protein conformation. Interestingly, the model includes structural quantities (e.g., solvent accessible surface areas) in the novel expression for the heat capacity of any model configuration. However, the M-WSME model does not address the solvation effects in terms of structural information, which may play intricate roles during protein folding. Instead, the solvation effects during folding-unfolding are implicitly treated in terms of the free energy change up to the first order of temperature at local foldons (Taylor expansion with respect to the transition temperature). The advantage is that this approach makes use of probe-

dependent properties measured directly from various spectrometers. Thus, a general feature about the protein's local folding behavior can be easily obtained. Although the WSME-S model uses detailed structural quantities and the novel heat capacity expression to account for solvation effects, it may not be sufficient in providing quantitative folding-unfolding features (e.g., shape of curves), from various kinds of probe-dependent spectroscopic techniques, with sufficient accuracy. This is due to the fact that different residue-dependent (or site-dependent) thermodynamic behaviors are not easily reproduced from a single fit to the heat capacity, unless the model itself ensures the inclusion of sufficient structural-related features for any given experimental probes. Another important point to address is hydrophobic clustering. Many studies have examined the hydrophobic effect in the stability of β-hairpins: experimental studies [66, 73, 85], MD simulation [86] and MC simulation [87]. A proper explicit description of the hydrophobic clustering effect will help to clarify its role in cooperativity and the M-WSME model may also benefit from such a description to provide a local view of the hydrophobic effects in the β-hairpin.

5 Conclusions

In this study, we proposed a modified version of the Wako-Saitô-Muñoz-Eaton (WSME) model, which includes a phenomenological parameter $T_{1/2}$ in the description of protein folding. The new parameters should provide a connection between the model and various probe-dependent thermodynamic behaviors, which can elucidate site-specific thermodynamic behavior of proteins. It is worth mentioning that the heterogeneous case of $T_{1/2}$ for different foldons may be a possible generalization used to treat more complicated systems. In a system with β-hairpin topology, the thermodynamics of both the proposed model (the M-WSME model) and the original WSME model were investigated and compared. Our results showed that, without the coupling interactions (enthalpic constraints) between foldon units, no complete folding-unfolding transition (1-to-0 sigmoidal curve) is observed for the WSME model, while the M-WSME model showed complete folding-unfolding behavior. This result suggests that the folding behavior from the WSME model strongly relies on the interactions between units, such as backbone H-bonds or hydrophobic clustering, which are explicitly treated as contact energy. However, the M-WSME model primarily attributes the folding transition behavior to the intrinsic properties of the foldon units (i.e., effective units), in which all interactions under solvated conditions are implicitly included in Δs_i and the coupling interactions between units are used to account for some enthalpic constraints such as backbone H-bonds. Furthermore, our comparison suggests that the backbone H-bond interactions may not play a determinate role in the stabilization of β-hairpins; instead, it may be involved in modulating the cooperative formation of β-hairpins. Furthermore, our results showed that a small homogeneous β-hairpin demonstrates diversity in thermodynamic behavior at different regions; in addition, convergent thermodynamic behavior was observed at high temperature. These findings imply the possibility that mean-field approaches can be applied to this system and that β-hairpins may be thermodynamically described at the foldon level. Based on the results studied herein, we recommend an approach that defines a foldon by examining the thermodynamic behavior of the local effective units in the system, when they function as a unified group. To conclude, the thermodynamics of the M-WSME model reported in this paper have demonstrated that this new model may be practically implemented and has provided adequate results for the small β-hairpin examined herein. These findings may highlight the need for research to investigate many of the

issues raised above and, in particular, the physical interpretations of the proposed M-WSME model.

Acknowledgements We wish to thank the National Science Council (Taiwan) for financial support. We also appreciate the help from Dr. Oleksandr Morozov from Florida International University for providing useful suggestions about this paper.

Appendix A: The phenomenological parameters

From (6), the free energy equation can be recast as

$$\Delta G^0(T) = \Delta H^0\left(T_{1/2}\right) - T\Delta S^0\left(T_{1/2}\right), \tag{27}$$

where $\Delta H^0\left(T_{1/2}\right) = T_{1/2} \cdot \Delta S^0\left(T_{1/2}\right)$. Both the enthalpy and entropy differences are constant and are characterized by the transition mid-point $T_{1/2}$. These two factors regulate the linear temperature dependence of free energy, thereby facilitating the folding-unfolding behavior. In the statistical mechanical models discussion, we replaced $\Delta S^0\left(T_{1/2}\right)$ and $\Delta H^0\left(T_{1/2}\right)$ with Δs and Δh, respectively, for convenience; thus, Δh can also be represented as $T_{1/2} \cdot \Delta s$. It should be noted that in the WSME model, entropy is the only concern for individual peptide bond units, as illustrated in (1). However, the M-WSME model uses the free energy equation from (6) to account for the folding of foldons; therefore, for local folding, the entropic cost is balanced by the newly added enthalpic factor for each foldon unit in the system. This is referred to as the "local-folding scheme" in terms of foldon units. In short, the enthalpy part and entropy part both have entropic interpretation and the resulting free energy change is generalized for each effective foldon unit in the system.

Appendix B: The contact-pair treatment—supplement

From (11), we briefly enumerate the configurations for the system as follows.

$$\begin{aligned} Z = (1) \cdot & \left\{ \left(\sum_{x_{\alpha-1}} \sum_{x_{\alpha+1}} B_{\alpha-1}^{x_{\alpha-1}} B_{\alpha+1}^{x_{\alpha+1}} \right) \cdots\cdots \left(\sum_{x_1} \sum_{x_N} B_1^{x_1} B_N^{x_N} \right) \right\} \\ & + (B_\alpha) \cdot \left\{ \left(\sum_{x_{\alpha-1}} \sum_{x_{\alpha+1}} C_1^{x_{\alpha-1} x_{\alpha+1}} B_{\alpha-1}^{x_{\alpha-1}} B_{\alpha+1}^{x_{\alpha+1}} \right) \cdots \right. \\ & \left. \cdots \left(\sum_{x_1} \sum_{x_N} C_{\alpha-1}^{x_{\alpha-1} x_{\alpha+1} x_{\alpha-2} x_{\alpha+2} \cdots x_1 x_N} B_1^{x_1} B_N^{x_N} \right) \right\}. \end{aligned} \tag{28}$$

The above equation accounts for the α^{th} peptide bond configurations; that is, the first term (see the first curly braces) denotes that $x_\alpha = 0$ and the remaining configurations $(x_{\alpha-1}, x_{\alpha+1}, \cdots, x_1, x_N)$ are to be determined, while the second term (see the second curly braces) denotes that $x_\alpha = 1$ and the remaining configurations are to be determined. Note that the first bracket in the second curly braces includes the configuration weights for the first contact pair $(1 + B_{\alpha-1} + B_{\alpha+1} + C_1 B_{\alpha-1} B_{\alpha+1})$, which denote, respectively, the weights for the configurations (0, 0), (1, 0), (0, 1) and (1, 1). These weights are then divided into two

groups: the group with the configuration (1, 1) and the other group with the remaining configurations: (0, 0), (1, 0) and (0, 1). Multiplying them out, it follows that

$$
\begin{aligned}
Z = (1) \cdot \left\{ \left(\sum_{x_{\alpha-1}} \sum_{x_{\alpha+1}} B_{\alpha-1}^{x_{\alpha-1}} B_{\alpha+1}^{x_{\alpha+1}} \right) \cdots\cdots \left(\sum_{x_1} \sum_{x_N} B_1^{x_1} B_N^{x_N} \right) \right\} \\
+ (B_\alpha)\,(1 + B_{\alpha-1} + B_{\alpha+1}) \cdot \left\{ \left(\sum_{x_{\alpha-2}} \sum_{x_{\alpha+2}} B_{\alpha-2}^{x_{\alpha-2}} B_{\alpha+2}^{x_{\alpha+2}} \right) \cdots\cdots \left(\sum_{x_1} \sum_{x_N} B_1^{x_1} B_N^{x_N} \right) \right\} \\
+ B_\alpha C_1 B_{\alpha-1} B_{\alpha+1} \left\{ \left(\sum_{x_{\alpha-2}} \sum_{x_{\alpha+2}} C_2^{x_{\alpha-2} x_{\alpha+2}} B_{\alpha-2}^{x_{\alpha-2}} B_{\alpha+2}^{x_{\alpha+2}} \right) \cdots \right. \\
\left. \cdots \left(\sum_{x_1} \sum_{x_N} C_{\alpha-1}^{x_{\alpha-2} x_{\alpha+2} \cdots x_1 x_N} B_1^{x_1} B_N^{x_N} \right) \right\}. \qquad (29)
\end{aligned}
$$

where Z_0 denotes the summation of the first and second terms, as shown in Section 2. Similarly, the same expansion performed on the curly braces of the third term and accounting for the configurations of the second contact pair $(\alpha - 2, \alpha + 2)$, Z_1 is identified and so forth.

Appendix C: The single foldon unit case

The thermodynamic two-state model is one of the simplest thermodynamic models for protein folding thermodynamics [see Fig. 1]. It correlates with statistical thermodynamics via (7) and (23). In other words, the two-state model in protein thermodynamics corresponds effectively to the single foldon unit case ($N = 1$) in the M-WSME model.

In the case where $N = 1$, the interaction terms are excluded and only Δs is specified. Three different values[11] ($|\Delta s| = 0.0032$, 0.01 and 0.04 kcal/mol K) were used in our investigations, which indicate, respectively, small, middle, and large values assumed for the protein system. Note that $\Delta s < 0$ is defined for the WSME model and $\Delta s > 0$ for the M-WSME model and the definition does not change the thermodynamic interpretation of $\langle x_i \rangle$, given that their absolute values are consistent. More specifically, $\Delta s = -0.0032$, -0.01 and -0.04 (kcal/mol K), when the WSME model was investigated; while $\Delta s = 0.0032$, 0.01 and 0.04 (kcal/mol K), when the M-WSME model was studied. The thermodynamic fraction of the native state compared between the WSME and M-WSME models is given in Fig. 8. Our results show that there is no sigmoidal curve for the WSME model in any case that we studied; instead, a constant value for the fraction was observed and determined using Δs. However, the M-WSME model showed a sigmoidal curve and the Δs can be used to manipulate the sharpness (or smoothness) of the transition; the larger the $|\Delta s|$ we use, the steeper the transition curve we obtain. The above results suggest that the WSME model cannot be used to describe protein folding in the single unit case due to the imbalance in the entropic cost for a single foldon unit. Therefore, the folding behavior is not shown. However, if the free energy balance is considered for each single foldon unit in the system (the local folding scheme), the folding behavior can be observed for a single

[11] The value $|\Delta s| = 0.0032$ (kcal/mol K) was referenced to the paper published by Muñoz et al. [7].

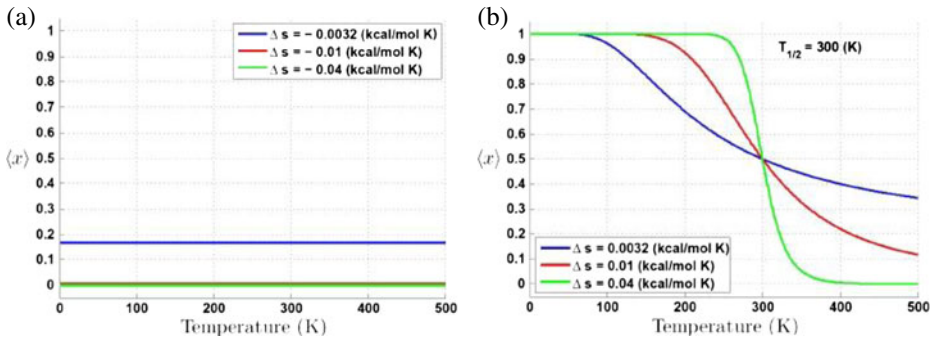

Fig. 8 Comparison of the thermodynamic native state fraction for the WSME (*left*) and M-WSME (*right*) models ($N = 1$). In each panel, the *blue*, *red*, and *green solid lines* denote, respectively, $|\Delta s| = 0.0032$, 0.01 and 0.04 (kcal/mol K). (**a**) The WSME model. (**b**) The M-WSME model with $T_{1/2} = 300$ K

foldon. Herein, we emphasize that the sigmoidal behavior from our treatment of a single foldon does not describe the exact form for the folded state in either a local or global sense.

In other words, the sigmoid is a thermodynamic behavior, and it describes the folded fraction, which fits the behavior observed from experiments, thus it can be used to trace the equilibrium folding-unfolding process. Thus, this description does not contradict the argument that a single foldon never folds without the aid of interactions with other foldon units. It should be noted that, given $|\Delta s| = 0.04$ (kcal/mol K), the M-WSME model showed a complete sigmoidal curve within the temperature range for biological relevance. This value could be used as a reference for other investigations in this paper.

Appendix D: The heat capacity equations

The heat capacity is a thermal-related physical quantity that can be obtained directly from the statistical mechanical relation

$$C = \left(\frac{dE}{dT}\right) = k_B \left[\frac{1}{Z} \cdot \frac{d}{dT}\left(T^2 \frac{dZ}{dT}\right) - \left(\frac{T \cdot dZ/dT}{Z}\right)^2\right] \tag{30}$$

where E is the internal energy of the system,[12] Z denotes the partition function of the system and k_B is the Boltzmann constant. Note that $E = k_B T^2 \left(d \ln Z/dT\right)$ is used in (30). According to the set of reduced partition functions derived and shown from (14) through (20), the heat capacity equations for systems with a different N can, in principle, be derived.

[12] We assume that the effect of volume change on the energy ($H = E + PV$), during protein folding-unfolding is negligible; therefore, H $\approx E$. Similarly, there is no need to distinguish $C_p = \left(\frac{dH}{dT}\right)_p$ from $C_v = \left(\frac{dE}{dT}\right)_v$; therefore, $C = C_p = C_v$.

For $N = 1$ (the M-WSME model),

$$\overline{C} = \frac{(\Delta h)^2}{RT^2} P_1 \left(1 - P_1\right), \tag{31}$$

where R is the gas constant (in kcal/mol K), $\Delta h = T_{1/2} \cdot \Delta s$ and $\overline{C}$ denotes the heat capacity per mole. P_1 denotes the probability of the folded state ($x_1 = 1$); thus, the product $P_1(1 - P_1)$ indicates the transition probability from the folded ($x_1 = 1$) to the unfolded state ($x_1 = 0$).

For $N = 3$ (the M-WSME model),

$$\overline{C} = \overline{C}_1 + \overline{C}_2 + \overline{C}_3 + \overline{C}_4, \tag{32}$$

where

$$\overline{C}_1 = \frac{(\epsilon_{13} - \Delta H)^2}{RT^2} \left[P_{111} \times (1 - P_{111})\right], \tag{33}$$

and $\Delta H = 3\Delta h$. The product $P_{111} \times (1 - P_{111})$ denotes the transition probability from the state $(x_1, x_2, x_3) = (1, 1, 1)$ to $(x_1, x_2, x_3) = (0, 0, 0)$. Note that the other components, $\overline{C}_2$, $\overline{C}_3$ and $\overline{C}_4$, which denote certain complicated transitions among all the configurations, are not detailed here.

For $N = 3$ (the WSME model),

$$\overline{C} = \frac{(\epsilon_{13})^2}{RT^2} P_{111}(1 - P_{111}). \tag{34}$$

Herein, we do not intend to show the detailed derivation of $\overline{C}_2$, $\overline{C}_3$ and $\overline{C}_4$ shown in (32). Instead, we have focused on our preliminary derivation, which may be used to distinguish the thermodynamic interpretations of the M-WSME model from those of the WSME model. The detailed derivations are now in progress and will be published elsewhere.

References

1. Anfinsen, C.B.: Principles that govern folding of protein chains. Science **181**, 223–230 (1973)
2. Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R.: The protein folding problem. Ann. Rev. Biophys. **37**, 289–316 (2008)
3. Barrick, D.: What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding? Phys. Biol. **6**, 015001 (2009)
4. Thirumalai, D., O'Brien, E.P., Morrison, G., Hyeon, C.: Theoretical perspectives on protein folding. Annu. Rev. Biophys. **39**, 159–183 (2010)
5. Bai, Y.W., Sosnick, T.R., Mayne, L., Englander, S.W.: Protein-folding intermediates—native-state hydrogen-exchange. Science **269**, 192–197 (1995)
6. Muñoz, V., Henry, E.R., Hofrichter, J., Eaton, W.A.: A statistical mechanical model for beta-hairpin kinetics. Proc. Natl. Acad. Sci. USA **95**, 5872–5879 (1998)
7. Muñoz, V., Thompson, P.A., Hofrichter, J., Eaton, W.A.: Folding dynamics and mechanism of beta-hairpin formation. Nature **390**, 196–199 (1997)
8. Lewandowska, A., Oldziej, S., Liwo, A., Scheraga, H.A.: Beta-hairpin-forming peptides; models of early stages of protein folding. Biophys. Chemist. **151**, 1–9 (2010)
9. Senn, H.M., Thiel, W.: QM/MM methods for biological systems. In: Reiher, M. (ed.) Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations. Topics in Current Chemistry, vol. 268, pp. 173–290 (2007)
10. Zimm, B.H., Bragg, J.K.: Theory of the phase transition between helix and random coil in polypeptide chains. J. Chem. Phys. **31**, 526–535 (1959)
11. Lifson, S.: Theory of helix-coil transition in polypeptides. J. Chem. Phys. **34**, 1963–1974 (1961)

12. Doig, A.J., Chakrabartty, A., Klingler, T.M., Baldwin, R.L.: Determination of free energies of N-capping in alpha-helices by modification of the Lifson–Roig helix-coil theory to include N-capping and C-capping. Biochemistry **33**, 3396–3403 (1994)
13. Thompson, P.A., Muñoz, V., Jas, G.S., Henry, E.R., Eaton, W.A., Hofrichter, J.: The helix-coil kinetics of a heteropeptide. J. Phys. Chem. B **104**, 378–389 (2000)
14. Bruscolini, P., Pelizzola, A.: Exact solution of the Muñoz–Eaton model for protein folding. Phys. Rev. Lett. **88**, 258101 (2002)
15. Liang, K.K., Hayashi, M., Shiu, Y.J., Mo, Y., Shao, J.S., Yan, Y.J., Lin, S.H.: Thermodynamics and kinetics of protein folding: a mean field theory. Phys. Chem. Chem. Phys. **5**, 5300–5308 (2003)
16. Morozov, A.N., Shiu, Y.J., Liang, C.T., Tsai, M.Y., Lin, S.H.: Nonadditive interactions in protein folding: the Zipper model of cytochrome c. J. Biol. Phys. **33**, 255–270 (2007)
17. Morozov, A.N., Lin, S.H.: Modeling of folding and unfolding mechanisms in alanine-based alpha-helical polypeptides. J. Phys. Chem. B **110**, 20555–20561 (2006)
18. Muñoz, V., Eaton, W.A.: A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc. Natl. Acad. Sci. USA **96**, 11311–11316 (1999)
19. Baxter, R.J.: Exactly Solved Models in Statistical Mechanics. Academic Press, New York (1982)
20. Boeglin, A., Zhang, X.G., Lin, S.H.: On the microscopic approach of the mean field kinetic Ising model. Physica A **137**, 439–453 (1986)
21. Rozenbaum, V.M., Morozov, A.N.: Fluctuation interaction of Ising subsystems. JETP Lett. **75**, 631–634 (2002)
22. Rozenbaum, V.M., Morozov, A.N., Lin, S.H.: Orientational regularities in two-dimensional quasidipole system with degenerate ground states. Phys. Rev. B **68**, 155405 (2003)
23. Rozenbaum, V.M., Morozov, A.N., Lin, S.H.: Generalized Ashkin-Teller model on the Bethe lattice. Phys. Rev. B **71**, 195411 (2005)
24. Fersht, A.: Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W.H. Freeman, New York (1999)
25. Kloss, E., Courtemanche, N., Barrick, D.: Repeat-protein folding: new insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys. **469**, 83–99 (2008)
26. Wu, F.Y.: The Potts model. Rev. Mod. Phys. **54**, 235–268 (1982)
27. Ananikyan, N.S., Hajryan, S.A., Mamasakhlisov, E.S., Morozov, V.F.: Helix-coil transition in polypeptides: a microscopical approach. Biopolymers **30**, 357–367 (1990)
28. Schreck, J.S., Yuan, J.M.: Exactly solvable model for helix-coil-sheet transitions in protein systems. Phys. Rev. E **81**, 061919 (2010)
29. Faccin, M., Bruscolini, P., Pelizzola, A.: Analysis of the equilibrium and kinetics of the ankyrin repeat protein myotrophin. J. Chem. Phys. **134**, 075102 (2011)
30. Bruscolini, P., Naganathan, A.N.: Quantitative prediction of protein folding behaviors from a simple statistical model. J. Am. Chem. Soc. **133**, 5372–5379 (2011)
31. Imparato, A., Pelizzola, A., Zamparo, M.: Equilibrium properties and force-driven unfolding pathways of RNA molecules. Phys. Rev. Lett. **103**, 188102 (2009)
32. Tokar, V.I., Dreysse, H.: Transfer matrix solution of the Wako-Saitô-Muñoz-Eaton model augmented by arbitrary short-range interactions. J. Stat. Mech.: Theory and Experiment, 08028 (2010)
33. Chung, H.S., Tokmakoff, A.: Temperature-dependent downhill unfolding of ubiquitin. II. Modeling the free energy surface. Proteins Struct. Funct. Bioinform. **72**, 488–497 (2008)
34. Chodera, J.D., Swope, W.C., Pitera, J.W., Dill, K.A.: Long-time protein folding dynamics from short-time molecular dynamics simulations. Multiscale Model. Simul. **5**, 1214–1226 (2006)
35. Zwanzig, R.: Simple model of protein folding kinetics. Proc. Natl. Acad. Sci. USA **92**, 9801–9804 (1995)
36. Flammini, A., Banavar, J.R., Maritan, A.: Energy landscape and native-state structure of proteins—a simplified model. Europhys. Lett. **58**, 623–629 (2002)
37. Wako, H., Saitô, N.: Statistical mechanical theory of protein conformation. 1. General considerations and application to homopolymers. J. Phys. Soc. Jpn. **44**, 1931–1938 (1978)
38. Wako, H., Saitô, N.: Statistical mechanical theory of protein conformation. 2. Folding pathway for protein. J. Phys. Soc. Jpn. **44**, 1939–1945 (1978)
39. Garcia-Mira, M.M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J.M., Muñoz, V.: Experimental identification of downhill protein folding. Science **298**, 2191–2195 (2002)
40. Zamparo, M., Pelizzola, A.: Kinetics of the Wako-Saitô-Muñoz-Eaton model of protein folding. Phys. Rev. Lett. **97**, 068106 (2006)
41. Jackson, S.E.: How do small single-domain proteins fold? Fold. Des. **3**, R81–R91 (1998)
42. Makhatadze, G.I., Privalov, P.L.: Energetics of protein structure. Adv. Protein Chem. **47**, 307–425 (1995)

43. Santoro, M.M., Bolen, D.W.: Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry **27**, 8063–8068 (1988)
44. Fung, A., Li, P., Godoy-Ruiz, R., Sanchez-Ruiz, J.M., Muñoz, V.: Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with alpha+beta topology that folds downhill. J. Am. Chem. Soc. **130**, 7489–7495 (2008)
45. Naganathan, A.N., Doshi, U., Muñoz, V.: Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. J. Am. Chem. Soc. **129**, 5673–5682 (2007)
46. Sadqi, M., Fushman, D., Muñoz, V.: Atom-by-atom analysis of global downhill protein folding. Nature **442**, 317–321 (2006)
47. Muñoz, V.: Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. Int. J. Quantum Chem. **90**, 1522–1528 (2002)
48. Shiu, Y.J., Jeng, U.S., Huang, Y.S., Lai, Y.H., Lu, H.F., Liang, C.T., Hso, I.J., Su, C.H., Su, C., Chao, I., Su, A.C., Lin, S.H.: Global and local structural changes of cytochrome c and lysozyme characterized by a multigroup unfolding process. Biophys. J. **94**, 4828–4836 (2008)
49. Moza, B., Qureshi, S.H., Islam, A., Singh, R., Anjum, F., Moosavi-Movahedi, A.A., Ahmad, F.: A unique molten globule state occurs during unfolding of cytochrome c by $LiClO_4$ near physiological pH and temperature: structural and thermodynamic characterization. Biochemistry **45**, 4695–4702 (2006)
50. Pletneva, E.V., Gray, H.B., Winkler, J.R.: Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome c. J. Mol. Biol. **345**, 855–867 (2005)
51. Marmorino, J.L., Lehti, M., Pielak, G.J.: Native tertiary structure in an A-state. J. Mol. Biol. **275**, 379–388 (1998)
52. Segel, D.J., Fink, A.L., Hodgson, K.O., Doniach, S.: Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c. Biochemistry **37**, 12443–12451 (1998)
53. Hamada, D., Kuroda, Y., Kataoka, M., Aimoto, S., Yoshimura, T., Goto, Y.: Role of heme axial ligands in the conformational stability of the native and molten globule states of horse cytochrome c. J. Mol. Biol. **256**, 172–186 (1996)
54. Hagihara, Y., Tan, Y., Goto, Y.: Comparison of the conformational stability of the molten globule and native states of horse cytochrome c: effects of acetylation, heat, urea and guanidine-hydrochloride. J. Mol. Biol. **237**, 336–348 (1994)
55. Privalov, P.L., Khechinashvili, N.N.: A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. J. Mol. Biol. **86**, 665–684 (1974)
56. Krishna, M.M.G., Maity, H., Rumbley, J.N., Lin, Y., Englander, S.W.: Order of steps in the cytochrome c folding pathway: evidence for a sequential stabilization mechanism. J. Mol. Biol. **359**, 1410–1419 (2006)
57. Maity, H., Maity, M., Englander, S.W.: How cytochrome c folds, and why: Submolecular foldon units and their stepwise sequential stabilization. J. Mol. Biol. **343**, 223–233 (2004)
58. Milne, J.S., Xu, Y.J., Mayne, L.C., Englander, S.W.: Experimental study of the protein folding landscape: unfolding reactions in cytochrome c. J. Mol. Biol. **290**, 811–822 (1999)
59. Panchenko, A.R., Luthey-Schulten, Z., Wolynes, P.G.: Foldons, protein structural modules, and exons. Proc. Natl. Acad. Sci. USA **93**, 2008–2013 (1996)
60. Nickson, A.A., Stoll, K.E., Clarke, J.: Folding of a LysM domain: entropy-enthalpy compensation in the transition state of an ideal two-state folder. J. Mol. Biol. **380**, 557–569 (2008)
61. Tsai, M.Y., Morozov, A.N., Chu, K.Y., Lin, S.H.: Molecular dynamics insight into the role of tertiary (foldon) interactions on unfolding in cytochrome c. Chem. Phys. Lett. **475**, 111–115 (2009)
62. Qian, H.: Thermodynamic hierarchy and local energetics of folded proteins. J. Mol. Biol. **267**, 198–206 (1997)
63. Mayne, L., Englander, S.W.: Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. Protein Sci. **9**, 1873–1877 (2000)
64. Hughes, R.M., Waters, M.L.: Model systems for beta-hairpins and beta-sheets. Curr. Opin. Struct. Biol. **16**, 514–524 (2006)
65. Bakk, A., Hoye, J.S.: One-dimensional Ising model applied to protein folding. Physica A **323**, 504–518 (2003)
66. Cochran, A.G., Skelton, N.J., Starovasnik, M.A.: Tryptophan zippers: stable, monomeric beta-hairpins. Proc. Natl. Acad. Sci. USA **98**, 5578–5583 (2001)
67. Kobayashi, N., Honda, S., Yoshii, H., Munekata, E.: Role of side-chains in the cooperative beta-hairpin folding of the short C-terminal fragment derived from streptococcal protein G. Biochemistry **39**, 6564–6571 (2000)
68. Honda, S., Kobayashi, N., Munekata, E.: Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. J. Mol. Biol. **295**, 269–278 (2000)

69. Blanco, F.J., Rivas, G., Serrano, L.: A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. Nat. Struct. Biol. **1**, 584–590 (1994)
70. Fesinmeyer, R.M., Hudson, F.M., Andersen, N.H.: Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. J. Am. Chem. Soc. **126**, 7238–7243 (2004)
71. Streicher, W.W., Makhatadze, G.I.: Calorimetric evidence for a two-state unfolding of the β-hairpin peptide Trpzip4. J. Am. Chem. Soc. **128**, 30–31 (2006)
72. Olsen, K.A., Fesinmeyer, R.M., Stewart, J.M., Andersen, N.H.: Hairpin folding rates reflect mutations within and remote from the turn region. Proc. Natl. Acad. Sci. USA **102**, 15483–15487 (2005)
73. Du, D.G., Tucker, M.J., Gai, F.: Understanding the mechanism of beta-hairpin folding via phi-value analysis. Biochemistry **45**, 2668–2678 (2006)
74. Bruscolini, P., Cecconi, F.: Mean-field approach for a statistical mechanical model of proteins. J. Chem. Phys. **119**, 1248–1256 (2003)
75. Tucker, M.J., Oyola, R., Gai, F.: Conformational distribution of a 14-residue peptide in solution: A fluorescence resonance energy transfer study. J. Phys. Chem. B **109**, 4788–4795 (2005)
76. Bruscolini, P., Cecconi, F.: Analysis of PIN1WW domain through a simple statistical mechanics model. Biophys. Chem. **115**, 153–158 (2005)
77. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G.: Funnels, pathways, and the energy landscape of protein-folding—a synthesis. Proteins Struct. Funct. Gene. **21**, 167–195 (1995)
78. Liu, F., Du, D., Fuller, A.A., Davoren, J.E., Wipf, P., Kelly, J.W., Gruebele, M.: An experimental survey of the transition between two-state and downhill protein folding scenarios. Proc. Natl. Acad. Sci. USA **105**, 2369–2374 (2008)
79. Liu, F., Gruebele, M.: Tuning lambda(6–85) towards downhill folding at its melting temperature. J. Mol. Biol. **370**, 574–584 (2007)
80. Lewandowska, A., Oldziej, S., Liwo, A., Scheraga, H.A.: Mechanism of formation of the C-terminal beta-hairpin of the B3 domain of the immunoglobulin-binding protein G from streptococcus. IV. Implication for the mechanism of folding of the parent protein. Biopolymers **93**, 469–480 (2010)
81. Lewandowska, A., Oldziej, S., Liwo, A., Scheraga, H.A.: Mechanism of formation of the C-terminal beta-hairpin of the B3 domain of the immunoglobulin binding protein G from Streptococcus. III. Dynamics of long-range hydrophobic interactions. Proteins Struct. Funct. Bioinform. **78**, 723–737 (2010)
82. Skwierawska, A., Makowska, J., Oldziej, S., Liwo, A., Scheraga, H.A.: Mechanism of formation of the C-terminal beta-hairpin of the B3 domain of the immunoglobulin binding protein G from Streptococcus. I. Importance of hydrophobic interactions in stabilization of beta-hairpin structure. Proteins Struct. Funct. Bioinform. **75**, 931–953 (2009)
83. Skwierawska, A., Imudzinska, W., Oldziej, S., Liwo, A., Scheraga, H.A.: Mechanism of formation of the C-terminal beta-hairpin of the B3 domain of the immunoglobulin binding protein G from Streptococcus. II. Interplay of local backbone conformational dynamics and long-range hydrophobic interactions in hairpin formation. Proteins Struct. Funct. Bioinform. **76**, 637–654 (2009)
84. Espinosa, J.F., Syud, F.A., Gellman, S.H.: Analysis of the factors that stabilize a designed two-stranded antiparallel beta-sheet. Protein Sci. **11**, 1492–1505 (2002)
85. Wu, L., McElheny, D., Huang, R., Keiderling, T.A.: Role of Tryptophan-Tryptophan interactions in Trpzip beta-hairpin formation, structure, and stability. Biochemistry **48**, 10362–10371 (2009)
86. Juraszek, J., Bolhuis, P.G.: Effects of a mutation on the folding mechanism of a beta-hairpin. J. Phys. Chem. B **113**, 16184–16196 (2009)
87. Imamura, H., Chen, J.Z.Y.: Dependence of folding dynamics and structural stability on the location of a hydrophobic pair in beta-hairpins. Proteins Struct. Funct. Bioinform. **63**, 555–570 (2006)