J Biol Phys (2007) 33:255–270
DOI 10.1007/s10867-008-9062-7

ORIGINAL PAPER

Nonadditive Interactions in Protein Folding: The Zipper Model of Cytochrome *c*

A. N. Morozov · Y. J. Shiu · C. T. Liang · M. Y. Tsai · S. H. Lin

Received: 19 December 2007 / Accepted: 25 February 2008 /
Published online: 12 April 2008

Abstract Hydrogen exchange experiments (Krishna et al. in *J. Mol. Biol.* 359:1410, 2006) reveal that folding–unfolding of cytochrome *c* occurs along a defined pathway in a sequential, stepwise manner. The simplified zipper-like model involving nonadditive coupling is proposed to describe the classical "on pathway" folding–unfolding behavior of cytochrome *c*. Using free energy factors extracted from HX experiments, the model can predict and explain cytochrome *c* behavior in spectroscopy studies looking at folding equilibria and kinetics. The implications of the proposed model are discussed for such problems as classical pathway vs. energy landscape conceptions, structure and function of a native fold, and interplay of secondary and tertiary interactions.

Keywords Sequential folding · Pathway · Nonadditivity · The Zipper model · Cytochrome *c* · Foldon · Circular dichroism

1 Introduction

Proteins are synthesized as linear chains of amino acid residues with specific primary sequences encoded at the genetic level. In a living environment, the linear polypeptide chain undergoes a rapid transition into a three-dimensional structure that is stable and capable of carrying out a biological function [1]. The search for the native fold in hyperdimensional conformational space is governed by a vast number of heterogeneous

A. N. Morozov (✉) · Y. J. Shiu · C. T. Liang · M. Y. Tsai
Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei, Taiwan, Republic of China
e-mail: morozov@gate.sinica.edu.tw

S. H. Lin
National Chiao Tung University, 1001 Ta Hsuen Road, Hsinchu, Taiwan, Republic of China

A. N. Morozov
Institute of Surface Chemistry, National Academy of Sciences of Ukraine, Generala Naumova str. 17, Kyiv 03164, Ukraine

interactions involving the surrounding solvent and amino acid residues. The folding potential is rough due to frustration among the numerous possible interactions [2]. However, proteins are able to find their native folds on a very fast time scale compared with that of a random search in conformational hyperspace [3]. Simplified quantitative models [4, 5] have shown that the folding potential has to be sloped enough toward the native state to overcome the roughness of the energy function and to compensate for the gradual loss of conformational entropy as the polypeptide chain adopts increasingly ordered structures. In the classical Levinthal formulation, the slope toward the native fold exists on a unique "pathway" [6], which is followed by all protein molecules as they undergo essentially the same sequence of events to reach the native state quickly through discrete intermediates [7–9]. The problem with the classical conception is the puzzle of how the unique pathway can be found fast enough among all possible pathways. The classical pathway concept can be viewed as a limiting case of the newer "energy landscape" concept. The latter one eliminates the aforementioned puzzle by assuming a funnel shape of the folding potential with multiple routes toward the native structure [2, 10–12]. Although the energy landscape approach currently dominates as the theoretical principle of protein folding, studies of several proteins are more consistent with the classical scenario. Such proteins fold through distinct intermediates along defined pathways [13–17]. In this class of proteins, cytochrome *c* (cyt *c*) has been the most extensively investigated experimentally. Early calorimetric [18] and spectroscopy studies [19] suggested that cyt *c* exhibits a two-state behavior. The existence of folding intermediates was revealed by the multiphase folding kinetics [20–22] and by different folding–unfolding behaviors of equilibrium spectroscopy measurements on different wavelengths [23–25]. It was mainly the work of Englander's group using hydrogen exchange (HX) experiments that provided detailed structural information about the intermediate states of cyt *c* [13, 14, 26–31]. HX experiments consistently show that cyt *c* folds by sequentially putting into place five groups of amino acid residues, called foldons. The residues of the *i*th foldon have approximately the same thermodynamic [13, 14, 28] and kinetic characteristics [26, 29]. The foldons appear in a definite order on the pathway from the native to the unfolded state and vice versa. The sequential character of cyt *c* folding was reaffirmed by stability-labeling experiments in which the stability of a given foldon is altered and the related effect on other foldons is measured [27, 30, 31].

Elucidating the mechanisms of classical "on pathway" folding is an open and appealing problem in the theory of protein folding. Because of the size and complex behaviors of protein–solvent systems, it is not yet feasible to investigate this problem using a strictly microscopic approach. Simplified statistical models coupled with coarse-grained descriptions of the conformational space are widely used for investigating folding thermodynamics and kinetics of polypeptides [32–38]. Indeed, the energy landscape concept stems from theoretical and computational studies utilizing coarse-grained models with pairwise additive interactions [10–12, 33]. At the same time, biologically important hydrophobic or biochemical effects can give rise to nonadditive coupling if a coarse-grained approach is used. In this paper, we propose a phenomenological zipper-like model of cyt *c* folding to show that nonadditive coupling is a plausible cause of the classical scenario in protein folding.

The model is closely related to a structural viewpoint gained in HX studies [13, 14] and draws its free energy factors from these experiments. To validate the model, we have checked its ability to explain and predict the results of spectroscopy measurements. It successfully explains a difference between the thermodynamic denaturant effect on the circular dichroism behavior at 222 (*cd222*) and 695 nm (*cd695*) [24]. In addition, the proposed model successfully predicts the thermodynamic temperature effect on *cd222* (see Section 6). The kinetic solution of the Zipper model can be obtained using either the mean-

field approximation (MFA) of the generalized kinetic Ising model (GKI) [35] or Monte Carlo (MC) kinetic simulations [39]. We have used MFA to explain qualitatively how the foldon structure of cyt *c* manifests itself in two-exponential *cd222* and three-exponential fluorescence (fl) kinetics, while MC kinetic simulations were used to obtain the related numerical solutions.

The Zipper Hamiltonian with pairwise additive interactions of the peptide bonds is used in the Muñoz and Eaton model (ME) to describe folding of secondary structures [34]. In the ME model, the zipper term serves as a topological constraint. In the case of alpha helix folding, this topological constraint is equivalent to the short-range approximation [36]. The idea to use zipper terms as topological constraints in a hierarchical scheme of protein folding was introduced by Hansen et al. [38], but no reasons were proposed for the existence of such constraints. In effect, their model is limited to the special case when only the coupling interactions between the structural units are important. We use a similar but more general formalism to account for solvent interactions, and we assume that on a coarse-grained level of description, zipper terms describe the effects of nonadditivity. In the case of cyt *c*, the heme group is the most plausible source of such interactions [40].

The paper is organized as follows. In Section 2, we present the Zipper model to treat the thermodynamics and kinetics of cyt *c* folding–unfolding. Model calculations are used in Section 3 to analyze and fit the experimental results. The last section discusses the implications of the proposed model for understanding the classical folding scenario, the separation of the secondary and tertiary folding time scales, and the specific roles of the different foldons for the structural and functional properties of cyt *c*. It should be noted that because of the simplicity of the proposed model, the calculations presented serve mainly to illustrate the ideas in the paper and cannot be considered a detailed numerical exposition.

2 The Zipper Hamiltonian

In the proposed model, one can view the cyt *c* protein as a collection of interacting foldon units. These units, in turn, are collections of amino acid residues in the secondary structures [30]. The fact that the free energy of the folding–unfolding transition for the different residues of a given foldon unit is the same [13, 14] indicates a high level of cooperativity at the level of the secondary structure. Thus, cyt *c* folding can be approximated as an "assembling" of the tertiary structure from two-state structural units (foldons). The conformation of the cyt *c* molecule is represented, then, by a set of binary variables:

$$\mathbf{x} = \{x_i\},\ i = 1, .., N\ x_i \in (0, 1) \tag{1}$$

where N=5 is the number of foldon units, x_i=1 or x_i=0 characterizing, respectively, the folded or unfolded state of the ith unit. The experimentally observed pathway of cyt *c* folding–unfolding via a series of partially unfolded intermediate conformations is presented schematically in Fig. 1. The most recent HX experiments [14] showed that the first three foldons fold by a stepwise sequential process, whereas the fourth and fifth do not follow a predetermined sequence. The assumed Hamiltonian, which is actually a free energy potential, reads:

$$H = \sum_{i=1}^{N} \varepsilon_i x_i + \sum_{i=1}^{N-1} j_i \prod_{k=1}^{i} x_k + j_N \prod_{k=1, k\neq N-1}^{N} x_k, \tag{2}$$

where ε_i and j_i denote the free energy factors of the ith unit due to the solvent and coupling interactions, respectively. In other words, j_i describes all cooperative effects, while ε_i is the sum of the residue–solvent effects over all residues of ith foldon. One can see that addition of the ith unit to the native structure is favored by coupling only if the zipper product $\prod_{k=1}^{i} x_k$ is equal to unity. The latter condition is fulfilled if all x^k units are in the native states, i.e., x_k=1, k=1, ..., i. In this way, the sequence of cyt c folding is assumed to be governed by the nonadditive many-body interactions of foldons. According to the minimal frustration principle [33], only native coupling is taken into account by the Hamiltonian in (2). The j^N zipper product has been modified to account for the independent behavior of the fourth and fifth units.

For N=5, the summation over 32 possible states to find the partition function or the equilibrium averages of physical values can be done directly; for the general case, the analytical solution of the Zipper model [41] can be used. By employing a substitution $x_i = (\sigma_i + 1)/2,\ \sigma_i \in (-1, 1)$, the Zipper model is transformed into the GKI model:

$$H_{\mathrm{GKI}} = -\sum_{i=1}^{N} \sigma_i E_i\left(\sigma_1, .., \sigma_{i-1}, \sigma_{i+1}, ..., \sigma_N\right). \tag{3}$$

For a system interacting with a heat bath and described by H^{GKI}, the master equation in the single-flip approximation reads [35, 42]:

$$\begin{aligned} \frac{\mathrm{d}P(x_1, ., x_i, ., x_N, t)}{\mathrm{d}t} = &-\sum_i w(x_i \to 1 - x_i) P(x_1, ., x_i, ., x_N, t) \\ &+ \sum_i w(1 - x_i \to x_i) P(x_1, ., 1 - x_i, ., x_N, t), \end{aligned} \tag{4}$$

where $P(x_1,., x_i,., x_N, t)$ is the probability of occupying state $\boldsymbol{x}=\{x_1,., x_i,., x_N\}$ at time t, and the local flipping rate w is determined by the Fermi golden rule [35]. In kinetic experimental measurements, one observes the time evolution of an average of a physical value over all possible kinetic trajectories $\langle f(t)\rangle = \sum_{\mathbf{x}} f(\mathbf{x}) P(\mathbf{x}, t)$. To analyze experimental results, it is especially important to know the time evolution of the folded fraction of the ith unit $\langle x_i(t)\rangle$, i=1,.., N. This is given by the solution of the set of differential equations [42]:

$$\tau_i \frac{\mathrm{d}}{\mathrm{d}t}\langle x_i\rangle = \frac{1}{2} - \langle x_i\rangle + \frac{1}{2}\tanh\left(\frac{\langle E_i\rangle}{k_{\mathrm{B}}T}\right), \tag{5}$$

where $\langle E_i\rangle$

$$\begin{aligned} \langle E_i\rangle &= -\left(\varepsilon_i + \sum_{k\geq i}^{N-1} j_k \prod_{\substack{m=1 \\ m\neq i}}^{k} \langle x_m(t)\rangle + j_N \prod_{\substack{m=1 \\ m\neq i, N-1}}^{N} \langle x_m(t)\rangle\right) \Big/ 2,\ i = 1, 2, 3 \\ \langle E_i\rangle &= -\left(\varepsilon_i + \sum_{m=1}^{N-2} \langle x_m(t)\rangle\right) \Big/ 2, \quad i = 4, 5, \end{aligned} \tag{6}$$

and τ_i depends on the heat bath interactions and $\langle E_i\rangle$ [35, 42]. It should be noted that, for heterogeneous structural units, (5) generally predicts multiphase nonexponential kinetics.

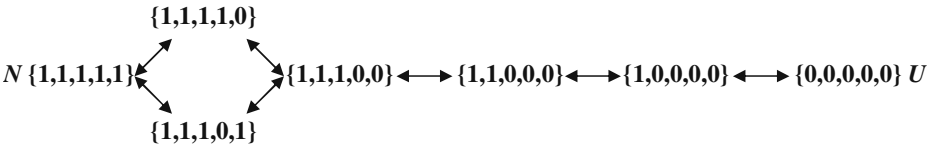

Fig. 1 The pathways of cyt *c* folding–unfolding via partially unfolded forms, as detected in HX experiments: *N*, The native fold; $\boldsymbol{x}=\{x_i\}$, the conformation of the partially unfolded state; *U*, the fully unfolded state

3 Results

3.1 The Denaturant Effect at Equilibrium

First, we shall show that folding–unfolding along the pathway observed in the equilibrium HX experiments at low denaturant concentrations (Fig. 1) [13, 14] can be duplicated by the numerical calculations using the Hamiltonian in (2). The effect of a denaturant on the coupling interactions is usually neglected. In accordance with the experimental results [13, 14, 43], the linear dependence of the foldon–solvent effect on denaturant concentration, described by m_i, is assumed to be

$$\varepsilon_i = \varepsilon_i^{(0)} + m_i[\mathrm{den}], \tag{7}$$

where [den] denotes the denaturant concentration and $\varepsilon_i^{(0)}$ is $\boldsymbol{\varepsilon}_i$ at [den]=0 and room temperature. Massive numerical calculations at the atomistic level show that at room temperature, $\boldsymbol{\varepsilon}$ for an alanine residue in water is close to zero [44]. Alanine side chain interactions are usually neglected [45]. The addition of a specific side chain of a given residue usually does not favor native conformations either. This is because of additional repulsive interactions and additional losses of conformational entropy related to the side chain ordering. These considerations, along with the well-known marginal stability of secondary structures, enable us to take $\varepsilon_i^{(0)} = 0$, i=1,.., N as a good approximation for the aggregate foldon–solvent effect. Using this assumption, the free energy factors $\boldsymbol{\varepsilon}_i$ and j_i in the Hamiltonian of (2) were calculated for an oxidized form of horse cyt *c* using the experimental dependence of G_i on denaturant concentration at 300 K [13] (see Appendix A for details). The values of m_i and j_i factors are summarized in Table 1.

In our simplified description, the "on pathway" conformations, i.e., $\boldsymbol{x}=\{1,1,0,0,0\}$, are presented by seven states (see Fig. 1), while the "off pathway" conformations, i.e., $\boldsymbol{x}=\{0,1,0,1,0\}$, correspond to another 25 states. Using (2) and (7), one can calculate the denaturant-dependent probabilities of occupying the "on pathway" $P(\boldsymbol{x}_{\text{on pathway}})$ and "off

Table 1 Parameters of the foldon units

i of the foldon unit	1	2	3	4	5
j_i kJ/ mol	−11.7	−10.886	−5.86	−17.57	−7.58
m_i kJ/(mol M)	6.28	4.18	0.65	2.72	2.09
$\Delta S_i^{(0)}$ J/(mol K)	171	−342	−71	−77.4	−93.7
$\Delta C_{p,i}$ kJ/(mol K)	5.44	0	0	0	0
α_i^{cd222}	0.82	0.18	0	0	0
α_i^{fl}	0.7	0.24	0.06	0	0
α_i^{cd695}	0	0	0	1	0
τ_i	38 ms	0.65 s	7 s	10 s	7 s

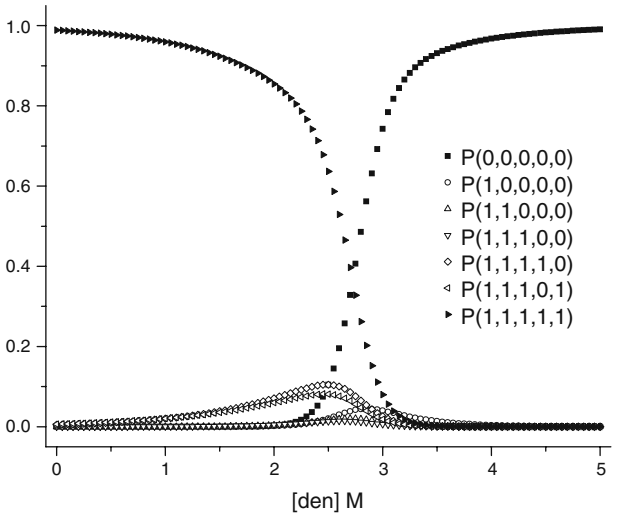

Fig. 2 Denaturant-dependent (guanidinium chloride, GdmCl) probabilities of occupying the "on pathway" states (T=300 K)

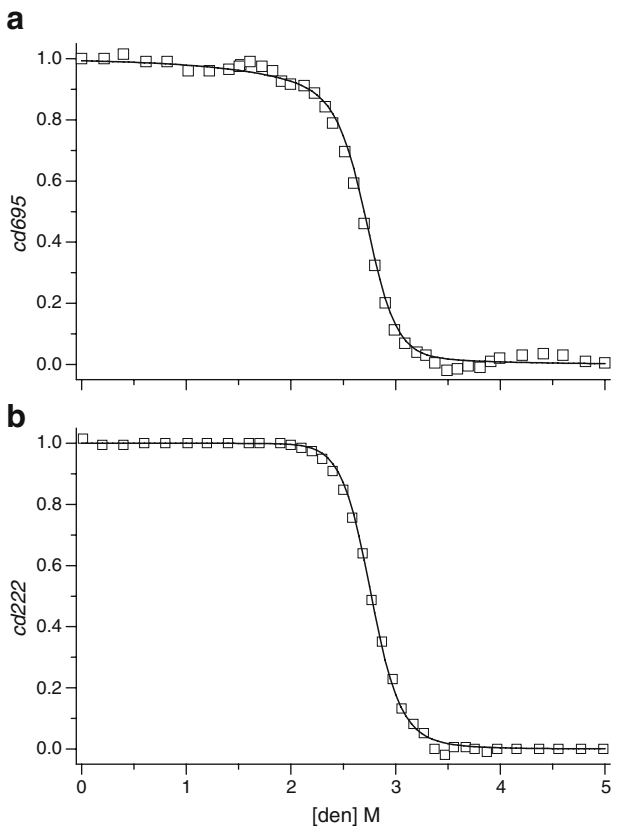

Fig. 3 The denaturant effect: comparison of the normalized optical signals due to the experimental measurements (*open squares*) and model calculations (*solid lines*) of cyt *c* equilibrium folding–unfolding in GdmCl: **a** *cd695*; **b** *cd222*

pathway" $P(\boldsymbol{x}_{\text{off pathway}})$ states. With the parameters listed in Table 1, we have found that the inequality

$$\sum_{x_{\text{on pathway}}} P(x) \gg \sum_{x_{\text{off pathway}}} P(x), \tag{8}$$

is valid at all denaturant concentrations typically used in folding–unfolding experiments (0< [den]<5 M guanidinium chloride (GdmCl)). Inequality (8) implies that the system described by the Zipper Hamiltonian in (2) folds and unfolds along the pathway depicted in Fig. 1.

We show now that the model can explain the observed difference between the *cd222* and *cd695* signals in equilibrium studies of denaturant-induced folding–unfolding of cyt *c* [24]. Figure 2 presents the calculated populations for the "on pathway" states. The model calculations show that a deviation from two-state behavior is due mainly to the lower stability of the fourth and fifth foldon units (red and brown symbols in Fig. 2). The structural changes of the fourth foldon unit, due to a methionine (Met80)–heme absorbance band [21, 46], would be reflected by the *cd695* signal. The structural changes of the more stable first and second units can be monitored by *cd222* because of the alpha helices in these units. Earlier unfolding of the fourth unit therefore explains the experimentally observed difference between *cd222* and *cd695* signals. To reproduce these behaviors numerically, the linear expansion of the normalized optical signal $\widetilde{f}_{obs}$ over equilibrium averages $\langle x_i \rangle$, i= 1,.., N was used for the *cd222* and *cd695* signals (see Appendix B):

$$\widetilde{f}_{obs} = \sum \alpha_i \langle x_i \rangle. \tag{9}$$

It should be noted that, in the general case, (9) is valid provided that inequality (8) is true. The theoretical estimates of the α_i coefficients for the *cd222* and *cd695* signals (see Appendix B) are presented in Table 1. The denaturant-dependent equilibrium behavior of *cd222* and *cd695* (Fig. 3a, b) from these calculations is in good agreement with the experimental data [24]. The midpoint (f_{obs}=1/2) of *cd695* occurs at [GdmCl]≅2.5 M, compared with [GdmCl]≅2.75 M for the midpoint of *cd222*.

3.2 The Temperature Effect at Equilibrium

We report the equilibrium measurements of the *cd222* signal as a function of temperature (Fig. 4). We show that, using the free energy factors from HX experiments [13, 47], the model successfully predicts the spectroscopic measurements. We assume that the effect of temperature on coupling interactions can be neglected. The temperature effect on ε_i can be described as follows:

$$\begin{aligned} \varepsilon_i = \varepsilon_i^{(0)} - \left(T - T^{(0)}\right) \Delta S_i^{(0)} \\ + \Delta C_{p,i} \left(\left(T - T^{(0)}\right) - T \ln \left(T \Big/ T^{(0)} \right) \right), \end{aligned} \tag{10}$$

where $\Delta S_i^{(0)}$ and $\Delta C_{p,i}$ are, respectively, the entropy and specific heat change in a folding transition of the ith unit. The values of $\Delta S_i^{(0)}$ and $\Delta C_{p,i}$ were extracted using the HX experimental results [47] (see Appendix A) and are presented in Table 1. From the positive value of the entropy change $\Delta S_1^{(0)}$ and nonzero value of $\Delta C_{p,1}$, it can be assumed that the hydrophobic effect plays an important role in the folding of the first unit. The temperature

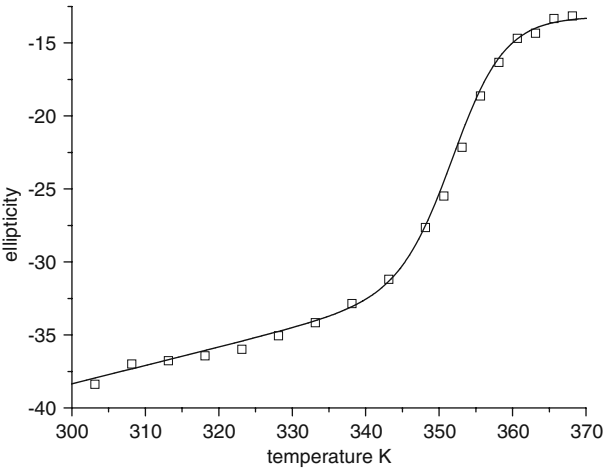

Fig. 4 The temperature effect: equilibrium unfolding of cyt *c* measured by *cd222* (*open squares*) and model calculations (*solid line*). The linear approximation for the native and unfolded signals was used in the calculations $f_{cd222}^{(N,U)}(T) = A^{(N,U)} + B^{(N,U)}T$ ($A^{(N)} = -76.0254$, $B^{(N)} = 0.12559$; $A^{(U)} = -9.8$, $B^{(U)} = -0.009$)

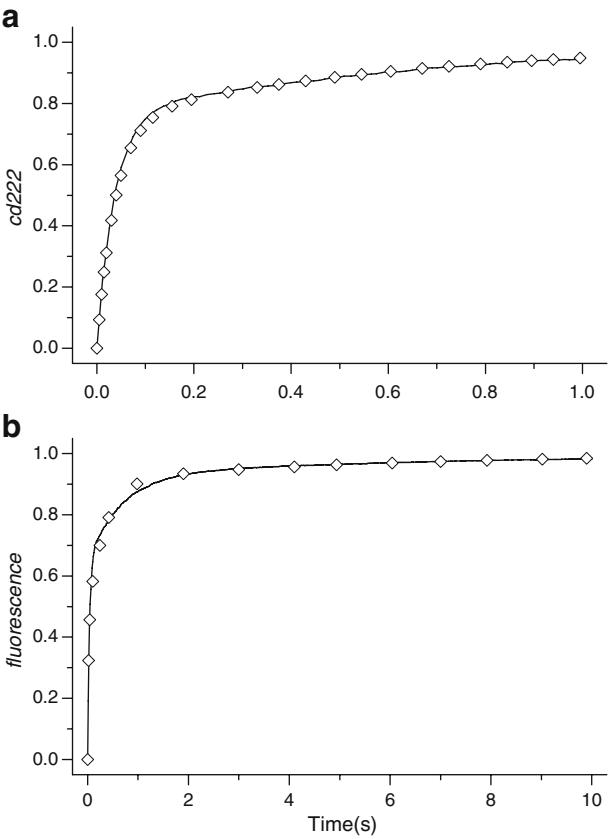

Fig. 5 Experimental (*scatter*) and model (*solid lines*) refolding kinetics of fully unfolded cyt *c* at 4.2 M [GdmCl] and by sudden dilution to 0.7 M [GdmCl]: **a** normalized *cd222*; **b** normalized fluorescence

dependence of *cd222* in the native $f^{(N)}_{cd222}(T)$ and unfolded $f^{(U)}_{cd222}(T)$ states is usually assumed to be linear [50], which is supported as well by our experimental results (see Fig. 4). Without normalization, (9) for the observable *cd222* signal reads:

$$\langle f_{cd222}\rangle = f^{(N)}_{cd222}(T)\left(\alpha_1^{cd222}\langle x_1\rangle + \alpha_2^{cd222}\langle x_2\rangle\right) + f^{(U)}_{cd222}(T)\left(1 - \left(\alpha_1^{cd222}\langle x_1\rangle + \alpha_2^{cd222}\langle x_2\rangle\right)\right). \tag{11}$$

Using the parameters from HX experiments (see Table 1), calculations from the model showed good agreement with the *cd222* experimental behavior (Fig. 4).

3.3 Refolding Kinetics

Just as for the equilibrium case, the observable optical signal $\widetilde{f}_{obs}(t)$ can be expanded into linear combinations over $\langle x_i(t)\rangle$:

$$\widetilde{f}_{obs}(t) = \sum \alpha_i \langle x_i(t)\rangle. \tag{12}$$

Because $\alpha_i^{cd222} = 0$, i=3,4,5, it follows from (5) and (12) that the *cd222* signal is bound to reveal two kinetic phases characterized by the time scales τ_1 and τ_2 of the first and second foldon units, respectively. The same reasoning predicts the existence of three kinetic phases for fluorescence ($\alpha_i^{fl} = 0$, i=4,5). Moreover, in the case $\tau_1 \ll \tau_2 \ll \tau_3$, the solution of (5) is close to a simple exponential behavior. For this case, (12) predicts two- and three- exponent kinetics, respectively, for *cd222* and fluorescence, in line with experimental results reported in [22, 29]. To reproduce the experimental observations of refolding kinetics by numerical calculations, the magnitudes of the time scales τ_1, τ_2, and τ_3 were adopted from [22]. The master equation (4) was solved using MC dynamic calculations [39]. The probabilities of the "on pathway" transitions were calculated using the relationships:

$$w(x_i \to 1 - x_i) = w(1 - x_i \to x_i)\exp\left(-\frac{H(x_1,..x_i,..x_N) - H(x_1,..,1-x_i,..x_N)}{k_B T}\right), \qquad w(x_i = 0 \to x_i = 1) = \tau_i \tag{13}$$

while the probabilities of "off pathway" transitions were set to zero. The model calculations (see Fig. 5) successfully duplicate the refolding experiments in [22]. The values of τ_i used in calculations (see Table 1) are due to fitting the experimental dependences of *cd222*. The coefficients $\alpha_i^{fl} = 0$, i=1,2,3 were found by fitting the experimental results.

It should be noted that the experimental results reported in [22, 29], in addition to the aforementioned kinetic phases of the *cd222* and fluorescence signals, reveal a very fast ($\tau <$ 4 ms) burst phase. Since this phase occurs without actual folding–unfolding of the foldon units and represents the system's response to a change of denaturant concentration [29], we did not consider the burst phase in the proposed model. There is a different viewpoint on the burst phase [51].

4 Discussion

Despite its simplicity, the present model describes many of the essential features of cyt *c* folding–unfolding from a standpoint that cyt *c* consists of five nonadditively interacting foldon units. The heterogeneous thermodynamic and kinetic properties of the foldon units were found to be important for describing the folding–unfolding behavior. Under mildly denaturing conditions, the first three foldon units are stable, whereas the last two are prone

to unfolding (Fig. 2). Such a deviation from the two-state behavior is an important feature of cyt *c* because it reflects the different structural and functional roles of foldons. The fifth foldon unit is the Ω loop, which plays important roles in cyt *c* function [52, 53]. One can conclude that in a mildly changing living environment, the tertiary structure of the cyt *c* protein is stable because the first three units form a nucleus, but is also highly reactive because the fourth and fifth units are unstable. This instability becomes especially evident if pH changes are used as a denaturant (see Fig. 5c in [54]).

The energy landscape concept generally predicts protein folding to show complex dynamics with a broad distribution of time scales [55, 56]. However, the kinetics of cyt *c* folding appears to be a simple two- or three-exponent process, depending on the type of spectroscopic measurement. Many models have been proposed to explain this phenomenon (see for example [20]). In the present work, we show that two- and three-exponent refolding kinetics is a manifestation of sequential folding of the foldon units, which occurs over different time scales. The characteristic time $\tau_1 \sim 50$ ms for the refolding of denatured cyt *c* (see Table 1) is determined by side chain packing of the amino acid residues of the first foldon unit [57]. The time scale $\tau_2 \sim 1$ s is determined by the predominant, at higher denaturant concentrations, heme–histidine misligation [58]. The slowest time, $\tau_1 \sim 10$ s (i= 3, 4, 5) is thought to be due to misisomerization of proline residues [59] namely, Pro 30 (third foldon), Pro 44 and 71 (fourth), and Pro76 (fifth).

The protein-folding theory seeks to understand the dynamic interplay of tertiary and secondary structures [45, 60, 61]. The folding time scales of cyt *c* foldon units (50 ms$\leq\tau\leq$ 10 s) are much slower than the related time scales of secondary structural motifs, i.e., $\tau_\alpha \sim 100$ ns for a 21-residue alpha helix [62] and $\tau_\beta \sim 10\ \mu$s for a 17-residue beta hairpin [63]. The difference in the magnitude of the secondary and tertiary time scales implies that establishing native tertiary interactions are rate-limiting events in the folding of cyt *c*. Makarov and Plaxco came to the same conclusion when investigating two-state folding of single-domain proteins using the topomer search model [64]. An exploration of the relatively small parts of a protein's conformational space to establish the proper tertiary contacts is a plausible answer to the question, "How can the cyt *c* molecule quickly find the unique folding pathway among all possible random trajectories?" To illustrate this hypothesis, proposed for the first time in [13], we recall the famous Levinthal estimation for the time scale of protein folding τ_{fold} due to a random walk in conformational space:

$$\tau_{\text{fold}} \sim 3^M \tau \tag{14}$$

where M is the number of amino acid residues of a protein and τ is the characteristic time of the transition between conformational macrostates of a residue. With M=104 for cyt *c* and $\tau \approx 50$ ps [36, 65] for helix–coil transitions, the estimation gives $\tau_{\text{fold}} \sim 10^{39}$ s and is known as Levinthal's paradox. Dividing the cyt *c* phase space into N foldons can be taken into account in (14) as follows:

$$\tau_{\text{fold}} \sim N 3^{M/N} \tau. \tag{15}$$

Substituting N=5 into (15) gives $\tau_{\text{fold}} \sim 1$ s, which is in strikingly good agreement with experimental observations.

5 Conclusion

Using the Zipper model, the nonadditive many-body interactions of experimentally observed foldon units were introduced in a phenomenological way to describe the "on

pathway" classical folding of cyt *c*. Such nonadditive interactions can play an important role in describing protein folding on a coarse-grained level.

6 Material and Methods

Horse heart cyt *c* in an oxidized form was purchased from Sigma (C-7752, M_W=12,384 g/mole) and used without further purification. Aqueous solutions of 40 μM cyt *c* were prepared in 7 mM potassium phosphate buffer at pH 7, where the ionic strength was adjusted by adding 7–21 mM KCl. The pH value of the protein was performed by using a pH meter (Beckman φ390).

For the temperature dependence of the circular dichroism (CD) spectra, the spectrometer (Jasco J715) was operated at a bandwidth of 2 nm, response time 0.5 s, scanning speed 100 nm/min, and ten spectra average. During the measurements, cyt *c* was prepared in a 1-mm light path quartz cell (Hellma, 165-QS), whose temperature was adjusted by a water cooler system (Neslab RTE-111). The temperature signal was monitored and fed back to the controller via two thermal couple sensors. In the studies described above, the temperature dependence of alpha helix folding–unfolding of cyt *c* was obtained from the various intensities of the 222 nm peak in the CD spectra.

Acknowledgments The authors are grateful to Prof. V. M. Rozenbaum for useful discussions related to the subject of this paper. We acknowledge Prof. T. C. Chang for kindly providing us the CD instrument in this work. This work was supported by the National Science Council (grant no. NSC 95-2113-M-001-044).

Appendix A

By subtracting the Hamiltonian function in (2) from the sequential conformations as depicted in Fig. 1, one finds that

$$\varepsilon_i + j_i = G_i - G_{i-1} \quad i = 1, .., N, \tag{16}$$

where $G_i = H(x_1 = 1, .., x_i = 1, x_{i+1} = 0, .., x_N = 0)$. At low concentrations of GdmCl and a temperature of 30 C, the experimental measurements of G_i [13] show a linear dependence on the denaturant concentration

$$G_i = G_i^{(0)} + m_{G_i}[\text{den}] \tag{17}$$

Taking into account (7) and approximating $\varepsilon_i^{(0)} = 0$ (Section 3), one can rewrite (16) as:

$$j_i = G_i^{(0)} - G_{i-1}^{(0)},\ m_i = m_{G_i} - m_{G_{i-1}} \quad i = 1, .., N. \tag{18}$$

Using (18) and the experimental values G_i, i=1,..,3 from [13], one can find m_i and j_i values unequivocally for the first three units. For the fourth and fifth foldon units, only the sum $G_{4+5} = G_4 + G_5$ was found in the experiments. Summation of (18) for the fourth and fifth foldon units gives:

$$j_5 + j_4 = G_{4+5}^{(0)} - G_3^{(0)},\ m_5 + m_4 = m_{G_{4+5}} - m_{G_3} \tag{19}$$

Based on the results in [43], we have assumed that values of m_4 and m_5 are proportional to the numbers of residues in the related foldon units. This assumption is roughly confirmed

by the values of m_i i=1,2,3 and allows us to find m_4, m_5 using (19) and the experimental value of $m_{G_{4+5}}$. The free energy factors j_4 and j_5 were found by numerical fitting of the equilibrium *cd222* signal [24] using (9) and (2). Using the HX experimental results [47], the values of $\Delta S_i^{(0)}$ in (10) were found in a similar way as we have done for m_i. Values of $\Delta C_{p,i}$ are taken from [46]; the values are zero except when i=1.

Appendix B

The average of a physical value f is given by:

$$\langle f \rangle = \sum_{\mathbf{x}} f(\mathbf{x}) P(\mathbf{x}), \tag{20}$$

where $P(\boldsymbol{x})$ is the probability to occupy the state $\boldsymbol{x}$. Spectroscopic measurements like absorption spectra, fluorescence spectra, and CD measure the sum of the local contributions to the signal:

$$f(x) = \sum_{i=1}^{N} f_i(x), \tag{21}$$

where $f_i(\boldsymbol{x})$ denotes the specific contribution of the ith unit.

Taking into account the "on pathway" character of cyt c folding–unfolding, one can approximate (20) by the sum over "on pathway" states:

$$\langle f \rangle \cong \sum_{x_{\text{on pathway}}} f(x) P(x). \tag{22}$$

First, let us consider the *cd222* signal. This signal is due to the contributions of the alpha helices of the first and the second foldon units, i.e., f_{cd222}=f_{cd222} (x_1, x_2). For "on pathway" states, one can decompose f_{cd222} into a linear combination over x_1, x_2:

$$\begin{aligned} f_{cd222}(x_1, x_2) = {} & f_{cd222}(0,0) + [f_{cd222}(1,0) - f_{cd222}(0,0)]x_1 \\ & + [f_{cd222}(1,1) - f_{cd222}(1,0)]x_2. \end{aligned} \tag{23}$$

The specific contribution of the ith unit to the overall *cd222* signal of cyt c depends only on the local conformation of this unit. Accordingly, for *cd222*, (21) reads:

$$f_{cd222}(x_1, x_2) = f_{cd222}(x_1) + f_{cd222}(x_2). \tag{24}$$

Substituting (23) in (22) gives:

$$\langle f_{cd222} \rangle = f_{cd222}(0,0) + [f_{cd222}(1,0) - f_{cd222}(0,0)]\langle x_1 \rangle + [f_{cd222}(1,1) - f_{cd222}(1,0)]\langle x_2 \rangle. \tag{25}$$

From this, the relation for the normalized signal can be written as:

$$\begin{aligned} \langle \widetilde{f}_{cd222} \rangle &\equiv \frac{\langle f_{cd222} \rangle - f_{cd222}^{(U)}}{f_{cd222}^{(N)} - f_{cd222}^{(U)}} = \alpha_1^{cd222} \langle x_1 \rangle + \alpha_2^{cd222} \langle x_2 \rangle \\ \alpha_1^{cd222} &= \frac{f_{cd222}(1,0) - f_{cd222}(0,0)}{f_{cd222}^{(N)} - f_{cd222}^{(U)}} \quad \alpha_2^{cd222} = \frac{f_{cd222}(1,1) - f_{cd222}(1,0)}{f_{cd222}^{(N)} - f_{cd222}^{(U)}}, \end{aligned} \tag{26}$$

where $f_{cd222}^{(N)}, f_{cd222}^{(U)}$ designate, respectively, the *cd222* signal of the folded and unfolded cyt *c* molecule. The number of helical residues contributing to the *cd222* signal should approximately equal the number of peptide carbonyls buried in the alpha helix [48]. In the single helix, this number equals the number of the helix backbone hydrogen bonds. Under the two-state description of the foldon units, and taking into account (24), the α_i^{cd222} coefficients can be calculated as the ratio of the number of alpha helix backbone hydrogen bonds in the *i*th unit to the total number of the alpha helix backbone hydrogen bonds:

$$\alpha_i^{cd222} = n_i^{(h)} \Big/ \Big(\sum n_i^{(h)}\Big), \quad n_i^{(h)} = n_i^{(\alpha)} - 2,\ i = 1, 2, \tag{27}$$

where $n_i^{(\alpha)}$ is the number of the helical residues in the *i*th foldon unit namely, $n_1^{(\alpha)} = 25$ and $n_2^{(\alpha)} = 7$ [49]. The calculated α_i coefficients of cd_{222} signal are presented in Table 1. Another useful representation of (24) is

$$\langle f_{cd222}\rangle = f_{cd222}^{(N)}\big(\alpha_1^{cd222}\langle x_1\rangle + \alpha_2^{cd222}\langle x_2\rangle\big) + f_{cd222}^{(U)}\big(1 - \big(\alpha_1^{cd222}\langle x_1\rangle + \alpha_2^{cd222}\langle x_2\rangle\big)\big). \tag{28}$$

The *cd695* signal depends on the local conformation of the fourth foldon unit and is contributed solely by this unit:

$$f_{cd695}(2x) = f_{cd695}(x_4). \tag{29}$$

Similar to *cd222* and using (29), one can obtain the expression for the normalized signal:

$$\big\langle \widetilde{f}_{cd695}\big\rangle \equiv \frac{\langle f_{cd695}\rangle - f_{cd695}^{(U)}}{f_{cd695}^{(N)} - f_{cd695}^{(U)}} = \langle x_4\rangle. \tag{30}$$

The fluorescence signal of cyt *c* is due to Förster energy transfer between the tryptophan residue of the third unit (Trp 59) and the heme group of the first unit [20]. The distance between tryptophan and the heme depends on the conformations of all the residues in between. In the frame of the proposed model, it can be expressed as

$$f_{\mathrm{fl}}(x) = f_{\mathrm{fl}}(x_1, x_2, x_3). \tag{31}$$

Similar to *cd222*, one can obtain from (31) the relation for the normalized fluorescence signal

$$\big\langle \widetilde{f}_{\mathrm{fl}}\big\rangle = \alpha_1^{\mathrm{fl}}\langle x_1\rangle + \alpha_2^{\mathrm{fl}}\langle x_2\rangle + \alpha_3^{\mathrm{fl}}\langle x_3\rangle \tag{32}$$

where α_i^{fl} has a structure similar to α_i^{cd222} in (26). The theoretical calculation of the α_i^{fl} coefficients is beyond the limits of the proposed model.

References

1. Frauenfelder, H., Wolynes, P.G., Austin, R.H.: Biological physics. Rev. Mod. Phys. **71**, 419–430 (1999)
2. Onuchic, J.N., Luthey-Schulten, Z., Wolynes, P.G.: Theory of protein folding: the energy landscape perspective. Annu. Rev. Phys. Chem. **48**, 545–600 (1997)
3. Levinthal, C.: How to fold graciously. In: Mossbauer Spectroscopy in Biological Systems: Proceedings of the University of Illinois Bulletin, vol. 67, pp. 22–24 (1969)
4. Zwanzig, R., Szabo, A., Bagghi, B.: Levinthal's paradox. Proc. Natl. Acad. Sci. USA **89**, 20–22 (1992)
5. Konermann, L.: Exploring the relationship between funneled energy landscapes and two-state protein folding. Proteins **65**, 153–163 (2006)

6. Levinthal, C.: Are there pathways for protein folding? J. Chim. Phys. **65**, 44–45 (1968)
7. Kim, P.S., Baldwin, R.L.: Intermediates in folding reactions of small proteins. Annu. Rev. Biochem. **59**, 631–660 (1990)
8. Englander, S.W.: Protein folding intermediates and pathways studied by hydrogen exchange. Annu. Rev. Biophys. Biomol. Struct. **29**, 213–238 (2000)
9. Bai, Y.: Protein folding pathways studied by pulsed- and native-state hydrogen exchange. Chem. Rev. **106**, 1757–1768 (2006)
10. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G.: Funnels, pathways and the energy landscape protein folding: a synthesis. Proteins **21**, 167–195 (1995)
11. Chan, H.S., Dill, K.A.: Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. Proteins **30**, 2–33 (1998)
12. Onuchic, J.N., Wolynes, P.G.: Theory of protein folding. Curr. Opin. Struct. Biol. **14**, 70–75 (2004)
13. Bai, Y., Sosnick, T.R., Mayne, L., Englander, S.W.: Protein folding intermediates: native-state hydrogen exchange. Science **269**, 192–197 (1995)
14. Krishna, M.M.G., Maity, H., Rumbley, J.N., Lin, Y., Englander, S.W.: Order steps in the cytochrome *c* folding pathway: evidence for a sequential stabilization mechanism. J. Mol. Biol. **359**, 1410–1419 (2006)
15. Cecconi, C., Shank, E.A., Bustamante, C., Marqusee, S.: Direct observation of the three state folding of a single protein molecule. Science **309**, 2057–2060 (2000)
16. Feng, H., Zhou, Z., Bai, Y.: A protein folding pathway with multiple folding intermediates at atomistic resolution. Proc. Natl. Acad. Sci. USA **102**, 5026–5031 (2005)
17. Baldini, G., Cannone, F., Chirico, G., Collini, M., Campanini, B., Bettati, S., Mozzarelli, A.: Evidence of discrete substates and unfolding pathways in green fluorescent protein. Biophys. J. **92**, 1724–1731 (2007)
18. Privalov, P.L., Khechinashvilli, N.N.: A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. J. Mol. Biol. **86**, 665–684 (1974)
19. McLendon, G., Smith, M.: Equilibrium and kinetic studies of unfolding of homologous cytochromes c. J. Biol. Chem. **253**, 4004–4008 (1978)
20. Tsong, T.Y.: Detection of three kinetic phases in the thermal unfolding of ferricytochrome *c*. Biochemistry **12**, 2209–2214 (1973)
21. Myer, Y.P.: Ferricytochrome *c* refolding and the methionine 80–sulfur–iron linkage. J. Biol. Chem. **259**, 6127–6133 (1984)
22. Elöve, G.A., Chaffotte, A.F., Roder, H., Goldberg, M.E.: Early steps in cytochrome *c* folding probed by time-resolved circular dichroism and fluorescence spectroscopy. Biochemistry **31**, 6876–6883 (1992)
23. Knapp, J., Pace, C.N.: Guanidine hydrochloride and acid denaturation of horse, cow and *Candida krusei* cytochrome c. Biochemistry **13**, 1289–1294 (1974)
24. Maine, L., Englander, S.W.: Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. Protein Sci. **9**, 1873–1877 (2000)
25. Latypov, R.F., Cheng, H., Roder, N.A., Zhang, J., Roder, H.: Structural characterization of an equilibrium unfolding intermediate in cytochrome *c*. J. Mol. Biol. **357**, 1009–1025 (2006)
26. Roder, H., Elöve, G.A., Englander, S.W.: Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. Nature **335**, 700–704 (1988)
27. Xu, Y., Mayne, L., Englander, S.W.: Evidence for an unfolding and refolding pathway in cytochrome *c*. Nat. Struct. Biol. **5**, 774–778 (1998)
28. Hoang, L., Bedard, S., Krishna, M.M.G., Lin, Y., Englander, S.W.: Cytochrome *c* folding pathway: kinetic native-state hydrogen exchange. Proc. Natl. Acad. Sci. USA **99**, 12173–12178 (2002)
29. Krishna, M.M.G., Lin, Y., Mayne, L., Englander, S.W.: Intimate view of a kinetic protein folding intermediate: residue-resolved structure, interactions, stability and unfolding rates, homogenity. J. Mol. Biol. **334**, 501–513 (2003)
30. 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)
31. Maity, H., Maity, M., Krishna, M.M.G., Mayne, L., Englander, S.W.: Protein folding: the stepwise assembly of foldon units. Proc. Natl. Acad. Sci. USA **102**, 4741–4746 (2005)
32. Lifson, S., Roig, A.: On the theory of helix–coil transitions in polypeptides. J. Chem. Phys. **34**, 1963–1974 (1961)
33. Bryngelson, J.D., Wolynes, P.G.: Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. USA **84**, 7524–7528 (1987)
34. 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)

35. Liang, K.K., Hayashi, M., Shiu, Y.J., Mo, Y., Shao, J., Yan, Y., Lin, S.H.: Thermodynamics and kinetics of protein folding: a mean field theory. Phys. Chem. Chem. Phys. **5**, 5300–5308 (2003)
36. Morozov, A.N., Lin, S.H.: Modeling of folding unfolding mechanism in alanine based alpha helical polypeptides. J. Phys .Chem. B **110**, 20555–20561 (2006)
37. Hairyan, S.A., Mamasakhlisov, E.S., Morozov, V.F.: The helix–coil transition in polypeptides: a microscopic approach. II. Biopolymers **35**, 75–84 (1995)
38. Hansen, A., Jensen, M.H., Sneppen, K., Zocchi, G.: A hierarchical scheme for cooperativity and folding in proteins. Physica A **250**, 355–361 (1998)
39. Fichthorn, K., Weinberg, W.H.: Theoretical foundations of dynamical Monte Carlo simulations. J. Chem. Phys. **95**, 1090–1096 (1991)
40. Weinkam, P., Zong, C., Wolynes, P.G.: A funneled energy landscape for cytochrome *c* directly predicts the sequential folding route inferred from hydrogen exchange experiments. Proc. Natl. Acad. Sci. USA **102**, 12401–12406 (2005)
41. Bruscolini, P., Pelizzola, A.: Exact solution of the Muñoz–Eaton model for protein folding. Phys. Rev. Lett. **88**, 258101 (2002)
42. 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)
43. Makhatadze, G.I., Privalov, P.L.: Protein interactions with urea and guanidinium chloride: a calorimetric study. J. Mol. Biol. **226**, 491–505 (1992)
44. Smith, P.E.: The alanine dipeptide free energy surface in solution. J. Chem. Phys. **111**, 5568–5579 (1999)
45. Baldwin, R.L., Rose, G.D.: Is protein folding hierarchic? I. Local structure and peptide folding. Trends. Biochem. **24**, 26–33 (1999)
46. Dragomir, I., Hagarman, A., Wallace, C., Schweitzer-Stenner, R.: Optical band splitting and electronic perturbations of the heme chromophore in cytochrome *c* at room temperature probed by visible electronic circular dichroism spectroscopy. Biophys. J. **92**, 989–998 (2007)
47. Milne, J.S., Xu, Y., 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)
48. Madison, V., Schelmann, J.: Optical activity of polypeptides and proteins. Biopolymers **11**, 1041–1076 (1972)
49. Bushnell, G.W., Louie, G.V., Brayer, G.D.: High-resolution three-dimensional structure of horse heart cytochrome *c*. J. Mol. Biol. **214**, 585–595 (1990)
50. Scholtz, J., Marqusee, S., Baldwin, R.L., York, E.J., Stewart, J.M., Santoro, M., Bolen, D.W.: Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. Proc. Natl. Acad. Sci. USA **88**, 2854–2858 (1991)
51. Shastry, M.C., Roder, H.: Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat. Struct. Biol. **5**, 385–392 (1998)
52. Krishna, M.M.G., Lin, Y., Rumbley, J.N., Englander, S.W.: Cooperative omega loops in cytochrome *c*: role in folding and function. J. Mol. Biol. **331**, 29–36 (2003)
53. Fetrow, J.S.: Protein motifs 6. Omega-loops—nonregular secondary structures significant in protein function and stability. FASEB J. **9**, 708–717 (1995)
54. Shiu, Y.J., Jeng, U.S., Su, C., Huang, Y.S., Hayashi, M., Liang, K.K., Yeh, Y.L., Lin, S.H.: A modified Ising model for the thermodynamic properties of local and global protein folding–unfolding observed by circular dichroism and small angle X-ray scattering. J. Appl. Crystallogr. (Suppl) **40**, 195–199 (2007)
55. Bryngelson, J.D., Wolynes, P.G.: Intermediates and barrier crossing in a random energy model (with applications to protein folding). J. Phys. Chem. **93**, 6902–6915 (1989)
56. Metzler, R., Klafter, J., Jortner, J.: Hierarchies and logarithmic oscillations in the temporal relaxation patterns of proteins and other complex systems. Proc. Natl. Acad. Sci. USA **96**, 11085–11089 (1999)
57. Colon, W., Elöve, G.A., Wakem, L.P., Shermann, F., Roder, H.: Side chain packing plays a critical role in the kinetics of cytochrome *c* folding. Biochemistry **35**, 5538–5549 (1996)
58. Sosnick, T.R., Mayne, L., Hiller, R., Englander, S.W.: The barriers in protein folding. Nat. Struct. Biol. **1**, 149–156 (1994)
59. Wood, L.C., White, T.B., Ramdas, L., Nall, B.T.: Replacement of a conserved proline eliminates the absorbance-detected slow folding phase of iso-2-cytochrome *c*. Biochemistry **27**, 8562–8568 (1988)
60. Dill, K.A., Fiebig, K.M., Chan, H.S.: Cooperativity in protein-folding kinetics. Proc. Natl. Acad. Sci. USA **90**, 1942–1946 (1993)
61. Voelz, V.A., Dill, K.A.: Exploring zipping and assembly as a protein folding principle. Proteins **66**, 877–888 (2007)

62. Lednev, I.K., Karnoup, A.S., Sparrow, M.C., Asher, S.A.: Transient UV Raman spectroscopy finds no crossing barrier between the peptide R-helix and fully random coil conformation. J. Am. Chem. Soc. **123**, 2388–2392 (2001)
63. Muñoz, V., Ghirlando, R., Blanco, F.J., Jas, G.S., Hofritcher, J., Eaton, W.A.: Folding and aggregation kinetics of a β-hairpin. Biochemistry **45**, 7023–7035 (2006)
64. Makarov, D.E., Plaxco, K.: The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Protein Sci. **12**, 17–26 (2003)
65. Young, W.S., Brooks III, C.L.: A Microscopic view of helix propagation; N and C terminal growth in alanine helices. J. Mol. Biol. **259**, 560–572 (1996)