Explicit-Chain Model of Native-State Hydrogen Exchange: Implications for Event Ordering and Cooperativity in Protein Folding

Hüseyin Kaya and Hue Sun Chan*

Department of Biochemistry, and Department of Medical Genetics & Microbiology, Faculty of Medicine, Protein Engineering Network of Centres of Excellence, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ABSTRACT **Native-state hydrogen exchange experiments on several proteins have revealed partially unfolded conformations with diverse stabilities. These equilibrium observations have been used to support kinetic arguments that folding proceeds via a sequential "pathway." This interpretative logic is evaluated here by analyzing the relationship between thermodynamic behavior and folding kinetics in a class of simplified lattice protein models. The chain models studied have varying degrees of cooperative interplay (coupling) between local helical conformational preference and favorable nonlocal interactions. When model cooperativity is high, as native conditions are weakened, "isotherms" of free energy of exchange for residues belonging to the same helix merge together before global unfolding. The point of merger depends on the model energetic favorability of the helix. This trend is similar to the corresponding experimental observations. Kinetically, we find that the ordering of helix formation in the very last stage of native core assembly tends to follow the stabilities of their converged isotherms. In a majority (but not all) of folding trajectories, the final assembly of helices that are thermodynamically more stable against exchange precedes that of helices that are less stable against exchange. These model features are in partial agreement with common experimental interpretations. However, the model results also underscore the ensemble nature of the folding process: the kinetics of helix formation is not a discrete, strictly "all-or-none" process as that envisioned by certain non-explicit-chain models. Helices generally undergo many cycles of partial formation and dissolution before their conformations are fixed in the final assembly stage of folding, a kinetic stage that takes up only ~2% of the average folding time in the present model; and the ordering of the helices' final assembly in some trajectories can be different from the dominant ordering stipulated by the exchange isotherms. Proteins 2005; 58:31–44.** © **2004 Wiley-Liss, Inc.**

Key words: sequential folding; partially unfolded form; two-process model; EX2 mechanism; global unfolding

INTRODUCTION

Since the classic experiment of Hvidt and Linderstrøm-Lang on pork insulin,[1] native-state hydrogen-deuterium exchange (often referred to simply as hydrogen exchange, HX in short) has long been recognized as a uniquely important tool in the investigation of dynamics and conformational diversity in the protein folded state.[2–4] HX rates are linked to protein conformational motility. In conjunction with NMR, HX data let us take a glimpse of the dynamic nature of protein native structures, offering a perspective complementary to the more static impression afforded by X-ray crystallography. In this context, a powerful technique has been to monitor the exchange rates $k_{HX}(i)$ of peptide amide protons (labeled by i) in different parts of the protein as a function of denaturant concentration or temperature.[5,6] By allowing HX to take place under conditions in which the native folded state is partially destabilized, these experiments have provided a wealth of information. Interpretations of such data often involve comparing the observed exchange rate $k_{HX}(i)$ with a reference rate $k_0(i)$ deduced from model compound measurements, stipulating that exchange is via the EX2 mechanism, and considering $k_0(i)$ to be the HX rate when the given proton is fully exposed. Under these assumptions, the degree to which a given proton is protected against HX because of its physico-chemical environment is often quantified by the factor $k_0(i)/k_{HX}(i)$. When denaturant concentration is varied under constant temperature, isotherms of the logarithm of protection factor $k_0(i)/k_{HX}(i)$ can be obtained. In these experiments, a salient observation is that for many—though not all—exchangeable protons, $\ln[k_0(i)/k_{HX}(i)]$ exhibits a significantly nonlinear dependence on denaturant concentration. A statistical

Grant sponsor: Canadian Institutes of Health Research; Grant number: MOP-15323; Grant sponsor: Canadian Protein Engineering Network of Centres of Excellence (PENCE); Grant sponsor: Canada Research Chair Programs.

H. Kaya's present address is Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138.

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

Received 14 April 2004; Accepted 8 July 2004

Published online 5 October 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20286

mechanical analysis[7] indicates that this feature is consistent with a two-process model[3] that envisions HX occurring as a result of localized conformational fluctuations as well as global unfolding of the entire protein. The smooth (differentiable) HX isotherms predicted by this analysis also rendered a previously postulated "global unlocking"[6] mechanism unnecessary.[7]

Protection factors in a protein can be quite diverse. The protection factors of many peptide amide protons are considerably lower than the few high protection factors that correspond to global unfolding. Native-state HX thermodynamic measurements have thus convincingly demonstrated that even for some apparent two-state proteins such as cytochrome *c,*[8] there is in fact a hierarchy of conformational states with free energies intermediate between the fully folded and fully unfolded states.[8–12]

Thermodynamic native-state HX data have important ramifications for folding kinetics as well. For bovine pancreatic trypsin inhibitor, Kim et al.[13] found that "the slow exchange core is the folding core," that is, "the most rigid segments of the native structure correspond to the most compact segments in early stages of folding" (pages 9600 and 9607 of ref. 13). For cytochrome *c,* Bai et al. noted that the partially unfolded forms detected by HX appear to represent the major unfolding intermediates and thus may define the major pathway of folding.[8] HX of ribonuclease H also identifies two partially folded species that resemble kinetic folding intermediates and the molten globule state of the protein.[14] However, such direct correspondences between HX thermodynamics and folding kinetics may not be universal. For instance, a subsequent experiment[15] shows that there is no correlation between slow exchange sites and residues involved in the folding nucleus of CI2. As well, in contrast to ribonuclease H, the partially unfolded states of T4 lysozyme detected by native-state HX do not mimic its kinetic folding intermediate.[11] For barnase, on the other hand, while there is no longer evidence (page 181 of ref. 17) for a well-structured burst-phase folding intermediate that is significantly more stable than the fully unfolded state,[17,18] a recent HX-detected partially unfolded form has been proposed to be a post-transition-state on-pathway folding intermediate.[19] In view of this apparent lack of generality, the relationship between native-state HX and folding kinetics remains controversial.[16,20–22] In this debate, it has been pointed out that kinetic mechanisms cannot in principle be derived from an equilibrium analysis because equilibrium properties are, by definition, pathway independent.[16] As well, partially unfolded conformations detected by native-state HX may actually be fluctuations of the native state, and therefore not necessarily on the dominant unfolding pathways.[15] (In other words, it may be possible to unfold via parallel pathways.)

Notwithstanding these general criticisms, the proposed cytochrome *c* folding pathway[8] based on initial native-state HX measurements has apparently been corroborated by further experiments. These include thermodynamic measurements comparing the stabilities against HX in the reduced and oxidized forms of the protein to narrow down the conformational possibilities of the partially unfolded states,[23] and kinetic measurements that exploit the EX1 mechanism at high pH.[24] An extensive compilation of HX data for 16 proteins also indicates a tendency for secondary structure elements that carry the amides most protected against native-state HX to be the same elements that are determined by pulse-labeling or competition experiments to be first protected during the kinetic folding process.[25] Taken together, although caution must be exercised in any kinetic interpretation of thermodynamic data,[11] existing data suggest an empirical,[25] statistical relation between native-state HX thermodynamics and the ordering of kinetic events during protein folding. The present work addresses the physical basis of this empirical connection.

THEORETICAL PERSPECTIVES

Previous theoretical works have addressed aspects of native-state HX: a statistical mechanical analysis that did not consider explicit chain representations provided a timely physical interpretation of the nonlinearity of HX isotherms.[7] Native-state HX thermodynamic patterns have been rationalized using a model with postulated strictly all-or-none subglobal cooperative folding units[26] as well as by elastic network models of proteins.[27,28] However, although these studies have provided valuable insights,[†] a broad range of questions concerning the implications of HX data on protein energetics and folding kinetics requires the consideration of self-contained polymer models. A hallmark of these models is that the conformational distribution is determined solely by the elementary intrachain interactions considered explicitly in the model. In this regard, an explicit-chain representation is crucial because it enables kinetic simulations of protein folding

[†]Important advances in protein folding have been made by approaches that do not consider chain conformations explicitly. These encompass traditional discourses using abstract "native," "denatured," and "intermediate" states,[29] models employing postulated free energy diagrams,[30] and folding rate prediction schemes based on applying Ising-type treatments to real protein structures.[31,32] Non-explicit-chain approaches also include native-structure-based methods for rationalizing protein stability and kinetics. Prime examples are an early thermodynamic model of Freire and Murphy[33] and a recent kinetic model of Weikl et al.[34] Both of these structure-based methods postulate strictly all-or-none folding and unfolding of preselected subglobal structural units; and chain conformational entropy is estimated by relatively simple formulas rather than computed directly using an explicit chain representation.

Non-explicit-chain calculation schemes can be extremely useful, but they are physically less informative in one important respect. This is because what form(s) of elementary intrachain interaction potential that may support the postulated calculation scheme (e.g., what form of inter-residue interactions would give rise to all-or-none subglobal folding units) is often not articulated. This uncertainty is confounded by these schemes' highly approximate account of conformational freedom. Thus, in most cases, non-explicit-chain approaches cannot provide a strong deductive logic from a set of clearly stated assumptions about fundamental physical interactions to the phenomena predicted by a model. For instance, whether a model protein would exhibit chevron rollover cannot be ascertained from simple free energy profiles alone.[35] And the above-described structure-based schemes[33,34] amount to very severe tacit restrictions on conformational freedom that would make overall thermodynamic cooperativity almost automatic. However, the question as to why so many a priori possible conformations are precluded remains to be more thoroughly addressed.[12]

and unfolding under the important physical constraints of chain connectivity and excluded volume, thus allowing the logical consequences of the (input) model interaction scheme to be better delineated.[12,36–38]

Multistate Protein Thermodynamics

From a polymer physics perspective, the multistate nature of protein thermodynamics[10] revealed by native-state HX is not surprising. In fact, the opposite—an absolute two-state situation—would have been hard to understand. Given the many conformations accessible to a chain molecule, it is conceptually difficult to envision how a set of physically plausible intrachain interactions could lead to only two conformational states (folded and unfolded) with very narrowly distributed free energies or enthalpies and *absolutely* no conformation with intermediate free energies (or enthalpies) in between.[12] Therefore, the existence of partially unfolded conformations is natural and expected, even if the magnitude of their population remains to be determined because of their limited detectability by existing experimental techniques. As a consequence of this elementary physical consideration that energetically intermediate conformations are a basic fact of polymers,[‡] the energetic distribution of denatured (non-native) conformations in self-contained polymer models of proteins can change significantly with native stability because the denatured conformations with lower free energies would be stabilized more, when overall native stability is increased, than the denatured conformations with higher free energies. This means that the denatured state can shift from a high-free-energy ensemble under weakly native conditions to a lower-free-energy ensemble under strongly native conditions (Fig. 3 of ref. 12). This trend readily leads to curved HX isotherms[9] that are *qualitatively* (though not necessarily quantitatively, see below) similar to those observed experimentally, and predicts significant variations in the structural stability of individual residues depending on their locations along the chain.[39]

Cooperativity Requirements

However, more detailed comparisons between polymer chain models and native-state HX experiments have been hampered by a fundamental statistical mechanical mismatch. Recent analyses indicate that many lattice models—these include a pioneering "hydrophobic-polar" (HP) two-letter lattice model for HX[9]—not only have conformations with intermediate enthalpies, but also have sizable populations of them. Consequently, these models' thermodynamic cooperativity is not sufficiently high[12,37,38,40,41] to mimic that of several proteins studied by native-state HX, although the models can produce curved HX isotherms. But the *apparent* two-state behavior of some of these real proteins is a major point of departure in the interpretation of the sparsely populated partially unfolded forms detected by native-state HX experiments.[42] In other words, although HX shows that conformations with intermediate enthalpies do exist in these proteins, their populations around the transition midpoint appear to be much lower than the high intermediate populations predicted by many chain models. This mismatch limits these models' ability to address experimental HX data. A case in point is the two-dimensional HP model: the prevalence of low-energy non-native conformations (termed "conformational distance relatives," CDRs[9]) is a main reason for this model's lack of calorimetric cooperativity.[12] In this light, although native-state HX can proceed via the CDRs in the model,[9] it seems unlikely that the CDRs in the model would share much similarity with the HX-detected partially unfolded forms of real, calorimetrically cooperative proteins.

The present study aims to overcome this mismatch by devising chain models with thermodynamic cooperativity comparable to real, apparent two-state proteins. We focus on one specific interpretative question, namely the proposal of Bai et al.,[8] that subglobal cooperative unfolding units can be identified by amino acid residues whose isotherms converge with increasing unfolding conditions (by increasing denaturant concentration), and that the relative stabilities of these cooperative units represent a pathway in a sequential model of folding.[8,10,23,24,42] To our knowledge, this central issue about the relationship between folding kinetics and the grouping of thermodynamic HX isotherms has not been addressed by explicit-chain modeling.

Meanings of "Intermediates"

Many partially unfolded forms (PUFs) detected by native-state HX are infinitesimally populated excited states that never acquired any sizable population during folding. Nevertheless, they have been referred to as folding or unfolding intermediates.[8] To avoid confusion, it is important to note that when PUFs are referred to as such, the meaning of "intermediates" becomes significantly different than that in the traditional protein folding literature. This is because intermediates have long been identified by their accumulated populations.[29,43] Indeed, as Sosnick et al. have stated, "it appears that intermediates characterized so far may be kinetically trapped by barriers that are optional rather than integral to the folding process."[44] In this respect, the physical picture offered by native-state HX data is quite akin to the energy landscape perspective that has emerged from polymer model simulations.[45–49] The landscape view is also consistent with the HX finding of a wide diversity of non-ground-state free energy levels under strongly native conditions.

The present investigation specifically addresses the role of equilibrium PUFs in the kinetics of folding. Although there is no general, necessary relationship between thermodynamics and kinetics of protein folding,[15,16] for many physical systems certain aspects of thermodynamic and kinetic behavior are correlated. If folding/unfolding thermodynamics and kinetics are determined by the same set of driving forces, it is not unreasonable to expect that "the first region to fold is the thermodynamically most stable

[‡]Intermediate-free-energy conformations cannot be eliminated entirely except by using highly contrived, physically implausible model potential functions.

portion of the protein."[14] On the other hand, it is important to realize that such correlations may not be universal, because there is evidence that for many proteins the driving forces for folding kinetics are partially divorced from those that dictate native stability.[40] In this context, the work reported here may be viewed as a first attempt to better delineate the conditions for validity of the thermodynamics-kinetics correlation identified by Bai et al.[8]

Simplified Models as Evaluation Tools

The logic of our critical evaluation is straightforward:

1. A simple lattice model is set up so that it exhibits thermodynamic properties—in particular calorimetric cooperativity and isotherms convergence—similar to that of several proteins studied by native-state HX. This model construction step will demonstrate, contrary to previous implications,[50–52] that simple lattice protein models can have thermodynamic properties similar to native HX observations.
2. By analyzing the time evolution of folding trajectories in the model, we assess the degree to which the thermodynamic ordering of the isotherms corresponds to the kinetic ordering of events during the folding process. Current protein kinetic experiments cannot yet provide such direct, detailed information about temporal order. Thus, it has been noted that "equilibrium data alone cannot define a kinetic sequence. In fact, neither can kinetic data."[24] In this regard, the ability to monitor the exact sequence of events during folding is an advantage that highlights the utility of an explicit-chain theoretical investigation.

The simple models we employ here are adequate for conceptual advances and testing of principles. This is in line with the fact that high-resolution structural information has not been invoked either in the formulation of the main interpretative paradigm[8] we seek to evaluate. Obviously, owing to our models' coarse-grained nature, they are agnostic about structurally detailed questions such as whether "penetration"[3,25] or "local unfolding"[4,42] is the mechanism of HX from the folded states. Nonetheless, our simple modeling approach allows several critical questions to be tackled. These include: (i) what is the role of the model interaction scheme and the resulting thermodynamic cooperativity in the validity of the proposed thermodynamics-kinetics correlation? (ii) If the thermodynamic HX data are correlated to a certain kinetic ordering of folding events, what is the nature of the kinetic ordering? Does it correspond to a sequential folding model? (iii) If the sequential picture is applicable to a protein model designed to exhibit a high degree of thermodynamic cooperativity and isotherms convergence, is such a sequential picture of folding absolute or probabilistic? In other words, are parallel pathways[53,54] or trajectories still possible?

METHODS

Models of Cooperative Protein Folding

The present investigation uses lattice models with a four-helix bundle native state (Fig. 1). This class of 55mer models was first constructed to provide a high degree of thermodynamic cooperativity similar to that observed experimentally for many real proteins.[55] When a set of many-body, local-nonlocal coupling interactions was introduced, thermodynamic trends exhibited by the original model (Fig. 1 of ref. 35) were reminiscent of experimental HX data on several cooperative proteins[5,8,42,56] (including, incidentally, a four-helix bundle[56]). In particular, as modeling conditions become more unfolding, many low-energy conformational states' stability curves merge before the global unfolding transition midpoint. In this respect, the thermodynamics of the original three-dimensional four-helix lattice model is akin to that of an earlier two-dimensional two-helix model with similar many-body interactions (Fig. 16 of ref. 12). However, HX of individual residues were not tackled in these earlier studies.

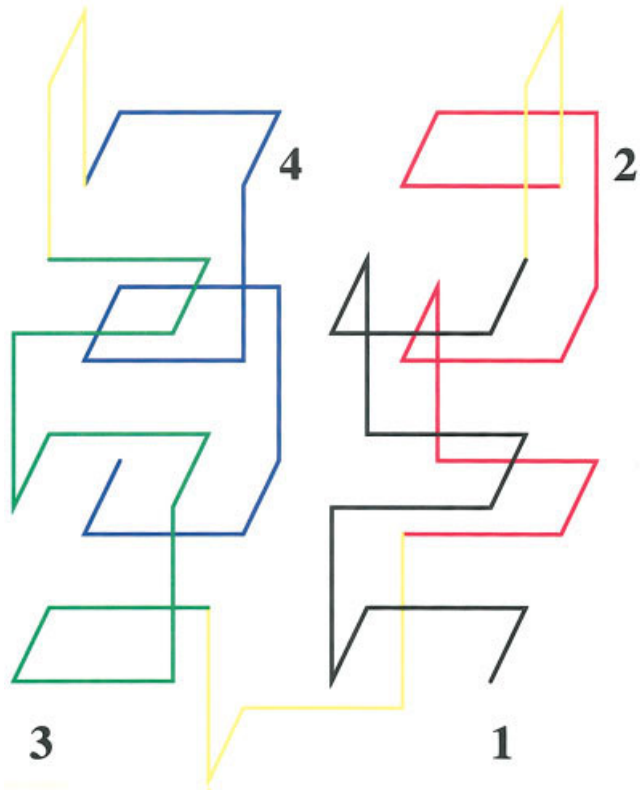

Fig. 1. A ground-state structure showing the native four-helix packing of the 55mer lattice protein model used in the present investigation. The numbers are labels for the four lattice helices. Though not depicted explicitly in the figure, residues are numbered starting from the beginning of helix 1 at the lower right. Bonds connecting residues in helices 1, 2, 3, and 4 (residues 1–12, 15–26, 30–41, and 44–55) are shown, respectively, in black, red, green, and blue. Bonds involving the connecting "turn" residues 13–14 (between helices 1 and 2), 27–29 (between helices 2 and 3), and 42–43 (between helices 3 and 4) are shown in yellow. The same color code is used in subsequent figures.

To explore effects of thermodynamic cooperativity on HX, we now consider three native-centric[57–59] variants[40] of the 55mer model. The basic properties of these models are derived from our original construct, which uses a five-letter alphabet for contact energies.[55] The three models in this work do not contain the E_{gs} term in ref. 40 (i.e.,

$E_{gs} = 0$). The first model studied here is identical to model (i) in ref. 40. Its energy function is given by

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

where the prime superscript on the $E'_{contact}$ term indicates that the sum of pairwise five-letter energies is restricted to native contacts in the ground-state conformation. All non-native contacts are assigned zero energy. The second term on the right disfavors left-handed helices. Aside from this repulsive term, the interaction scheme of this model is pairwise additive. Here we use the same contact energies and γ_{lh} parameter as in refs. 40 and 55. The ground-state energy of this model is equal to −36.1.

Our second model embodies the idea of a cooperative interplay between local conformational preference and the (nonlocal) packing of the protein core. It is identical to model (ii) in ref. 40. The construction of this model was motivated by the context-dependent formation of secondary structure in globular proteins. The corresponding energy function is given by

$$E = E'_{contact} + \gamma_{lh} N_{lh} + \varepsilon_{coop} \sum_{i=1}^{4} h_c^{(i)} \quad (2)$$

where $h_c^{(i)} = 1$ if native helix i (as labeled in Fig. 1 of this article) is fully formed, and, at the same time, has the local-nonlocal cooperative interactions defined by Figure 1 of ref. 40. Otherwise $h_c^{(i)} = 0$. Thus, the possible values for the summation in the above equation are 0, 1, 2, 3, or 4. When $\varepsilon_{coop} < 0$, this interaction scheme provides additional energetic favorability for the correct packing of two or more native helices. Here we set $\varepsilon_{coop} = -1$, the resulting model ground-state energy equals $-36.1 - 4 = -40.1$.

Our third model is a new variation introduced here to investigate the impact of stability and cooperativity of individual helices on the applicability of the proposed HX thermodynamics-kinetics correlation. To this end, instead of having a single cooperative term ε_{coop} that applies uniformly to all four helices [Eq. (2)], we now allow the local-nonlocal coupling to adopt different strengths for different helices. The resulting energy function becomes

$$E = E'_{contact} + \gamma_{lh} N_{lh} + \sum_{i=1}^{4} \varepsilon_{coop}^{(i)} h_c^{(i)} \quad (3)$$

where $\varepsilon_{coop}^{(i)}$ is the local-nonlocal cooperative parameter for helix i. For the present study, we set $\varepsilon_{coop}^{(1)} = \varepsilon_{coop}^{(4)} = -2$ and $\varepsilon_{coop}^{(2)} = \varepsilon_{coop}^{(3)} = -4$ because this choice of parameters apparently provides a straightforward separation of model HX isotherms into two main groups. This model is more stable (ground-state energy = −48.1) and more cooperative than the model defined by Equation (2).

To model the thermodynamics and kinetics of folding at different interaction strengths, we use an energetic scaling parameter ε. At a given ε, the effective energy of a conformation with energy function E [given by Eqs. (1), (2), or (3)] is equal to $-\varepsilon E$. Thus, as in ref. 40, interaction strength in the present work is parameterized by $\varepsilon/k_B T$, where $k_B T$ is Boltzmann constant times absolute temperature.

The three models considered here were chosen because they encompass significant differences in cooperativity properties. To underscore these differences, we refer to the models by their $\varepsilon_{coop}^{(i)}$ values, viz., as "0000", "1111", and "2442", respectively, for Equations (1), (2), and (3). The van't Hoff to calorimetric enthalpy ratio[37] $\Delta H_{vH}/\Delta H_{cal}$ of the 0000, 1111, and 2442 models are, respectively, κ_2 = 0.804, 0.888, and 0.925 without baseline subtractions and $\kappa_2^{(s)}$ = 1.004, 1.000, and 1.001 after empirical linear baseline subtractions.[37,55] Thus, the thermodynamic cooperativity of all three models may be viewed as comparable to that of real, calorimetrically cooperative proteins, but their degrees of cooperativity differ. Not surprisingly, models whose cooperative many-body interactions are stronger ($\varepsilon_{coop}^{(i)}$ more negative) have higher κ_2 values and are more stable. The transition midpoints are $\varepsilon/k_B T$ = −2.36, −2.10, and −1.71, respectively, for the 0000, 1111, and 2442 models. Folding and unfolding kinetics of these models are simulated using the same methodology as that in ref. 40.

Our interest in the 1111 and 2442 models was partly motivated by the intriguing idea of subglobal folding units.[8] The designs of these models were also prompted by our desire to capture, using physically plausible interactions, the extreme thermodynamic and kinetic cooperativity of real proteins that have linear folding and unfolding arms in their chevron plots. The 0000 model is not sufficient for this purpose. We have demonstrated that the 0000 model, like other common Gō models with only pairwise additive contact energies studied thus far,[35,38,60] exhibits a significant chevron rollover.[40] Therefore, it appears that many-body interactions beyond the pairwise contact energies envisioned by common Gō models are needed for kinetic cooperativity.[40] In this regard, it should be noted that we were unable to find evidence to support the suggestion (page 231 of ref. 51) that an earlier 48mer Gō model of Pande and Rokhsar[61] with only pairwise additive interactions contains "separable cooperative units."[51] Instead, we found that although this model is thermodynamically cooperative,[37] its folding/unfolding kinetics is not sufficiently cooperative to be two-state-like because it shows a severe chevron rollover.[38,60] In contrast, the chevron plot of the 1111 model has been shown to contain a substantial quasi-linear regime.[40] A generalization of the local-nonlocal coupling scheme in the 1111 model also leads to a more proteinlike correlation between folding rate and native topology.[38,62,63] A corresponding chevron plot for the present 2442 model has not been obtained. Owing to its high degree of cooperativity (which implies a high free energy barrier between the folded and unfolded states), the 2442 folding and unfolding rates near the transition midpoint are slow and thus the large number of trajectories required to determine a chevron plot is computationally costly to calculate. Nonetheless, because kinetic cooperativity is well correlated with thermodynamic cooperativity and the 2442 model (κ_2 = 0.925)

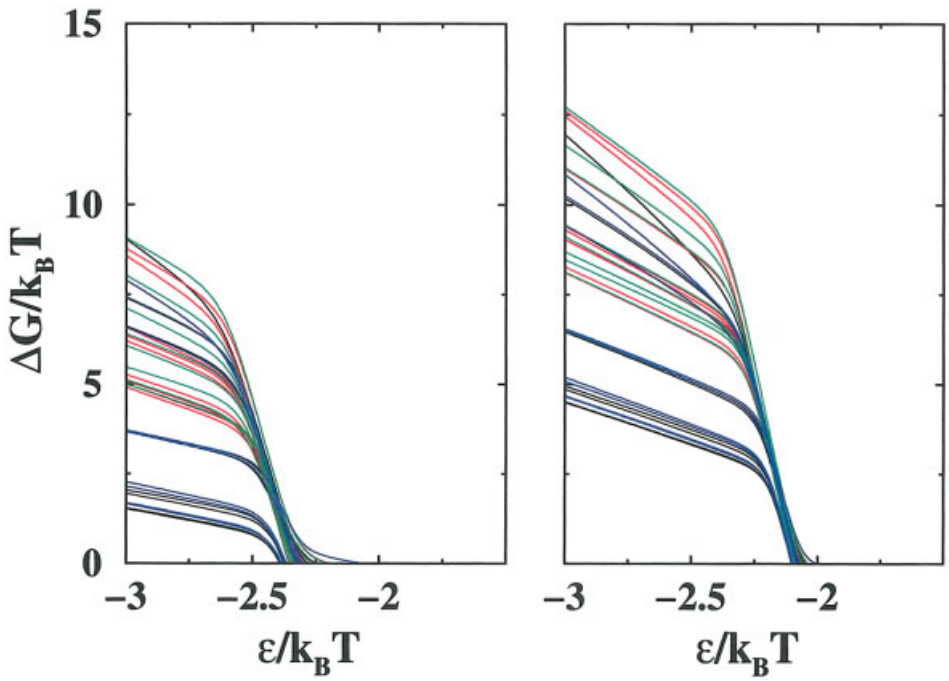

Fig. 2. Free energies of exchange ΔG in units of k_BT as functions of interaction strength ε/k_BT for the Gō-like five-letter 0000 model with pairwise additive interactions (left, $\varepsilon_{coop} = 0$) and the cooperative 1111 model (right, $\varepsilon_{coop} = -1$). The curves show the free energies of exchange for 36 residues (nine per helix). According to model definitions detailed in the text, for a given residue *i*, a necessary exchangeability requirement is that a specific intrahelical native contact (*i, j*) associated with the given residue has to be broken. These residues and their specific exchange-determining native contacts are as follows. Helix 1 (black curves): (4,1), (5,8), (6,1), (7,2), (8,3), (9,12), (10,5), (11,6), (12,7); helix 2 (red curves): (18,15), (19,22), (20,15), (21,16), (22,17), (23,26), (24,19), (25,20), (26,21); helix 3 (green curves): (33,30), (34,37), (35,30), (36,31), (37,32), (38,41), (39,34), (40,35), (41,36); helix 4 (blue curves): (47,44), (48,51), (49,44), (50,45), (51,46), (52,55), (53,48), (54,49), (55,50). The thermodynamics of exchange for the 0000 model (left) was computed using standard histogram techniques based on conformational sampling at its transition midpoint $\epsilon/k_BT = -2.35$, whereas that for the 1111 model (right) was computed using a combination of standard and parameter-space histogram techniques[55] based on sampling near this model's transition midpoint ($\epsilon/k_BT = -2.10$) at $\epsilon/k_BT = -2.0$ and -2.13.

is thermodynamically more cooperative than the 1111 model ($\kappa_2 = 0.888$), there are ample reasons to expect the 2442 model to be at least as kinetically cooperative as the 1111 model,[40] with no apparent accumulation of intermediates over an extended ε/k_BT regime.

The thermodynamics of the three models are investigated using extensive sampling and Monte Carlo histogram techniques.[37,40,55] To model hydrogen exchangeability, we monitor the conformational environment of 36 residues in the model (Fig. 2). A given residue is considered to be HX competent if both of the following conditions are satisfied: (i) a specific native contact involving the given residue is absent, and (ii) the given residue is accessible by solvent, that is, it has less than four (five for residues at the two chain ends) contacts with residues other than its immediate sequential neighbors along the chain. The specific contacts in condition (i) for the 36 residues are provided in the legend for Figure 2. They may be viewed as caricatures of native helical hydrogen bonds.

These model definitions are admittedly somewhat arbitrary. The models here are extremely coarse-grained and thus the correspondence between model and real protein properties is necessarily inexact. Nonetheless, as discussed above, the present setup should be adequate for exploring general physical principles. For simplicity, we sidestep an important experimental question as to whether the reference rates for folded-state HX have been accurately calibrated in view of possible complex effects afforded by the local native conformational environment.[25] In the present study, the model HX rate of a given residue under EX2 condition[64] is assumed to be directly determined[6,8] by the conformational population that allows the given residue to be HX competent as defined above. This means that the stability (ΔG in Figs. 2 and 3) of a given residue's HX competent conformational population relative to the ground state of the model is taken here to correspond to the HX free energy ΔG_{HX} reported in the literature.[11,23,51,56]

RESULTS AND DISCUSSION

Local-Nonlocal Coupling Leads to Convergence of HX Isotherms

Figures 2 and 3 show the HX free energies (isotherms) of individual residues as functions of the interaction parameter ε/k_BT. As we have discussed elsewhere,[35,38,40] the variation in ε/k_BT in these highly simplified models may be

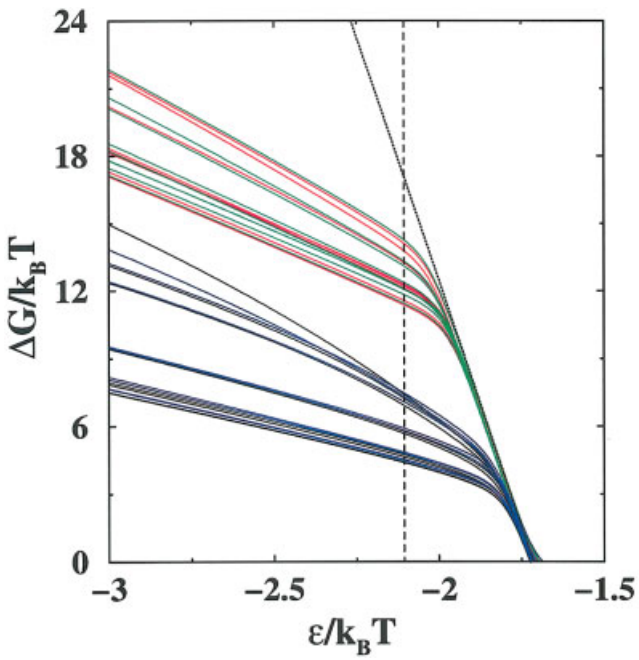

Fig. 3. Model free energies of exchange that are suggestive of "sequential folding." The color curves have the same meaning as that in Figure 2 but are now for the cooperative 2442 model. The inclined dotted line is the free energy of global unfolding (in units of k_BT computed as the difference between that of the $Q = 1$ ground-state conformations and the open $Q \leq 10/60$ conformations). The vertical dashed line marked the $\epsilon/k_BT = -2.105$ value at which folding kinetic simulations of the 2442 model are performed. Model HX thermodynamics was determined using parameter-space histogram techniques[55] based on conformational sampling of the 1111 model and an 1221 model at $\epsilon/k_BT = -2.0$ and -2.13, respectively.

viewed as roughly corresponding to the variation of denaturant under constant temperature. This is because the model global free energy of unfolding in these models is approximately linear in ε/k_BT (cf. inclined dotted line in Fig. 3), similar to the experimental dependence of many proteins' stability on denaturant concentration.

All three models we studied exhibit nonlinear HX isotherms with a wide range of HX free energies under strongly native conditions. HX ΔGs of residues belonging to the same helix can be vastly different, implying that individual helix-coil transition is not strictly all-or-none. The results also indicate notable differences in behavior between helices 1 and 4 on one hand, and helices 2 and 3 on the other. On average, the two helices at midchain (2 and 3, red and green) are more stable against HX than the two helices at the two chain ends (1 and 4, black and blue). Apparently, owing to the native topology of these models and conformational statistics, the two helices at the two ends of the chain are easier to unravel even when the interactions themselves in the 0000 and 1111 models are not biased against these two helices relative to those at midchain. This is because the coming together of two midchain helices entails formation of interhelix contacts with lower contact orders, which restricts the chain conformations less and therefore are entropically less costly.[65] For the 0000 model (left panel of Fig. 2), when conditions are strongly native at $\varepsilon/k_BT = -3.0$, the highest HX ΔG is approximately 8 k_BT for every helix. However, under the same condition, while all of the HX ΔGs for the residues in helices 2 and 3 (red and green) are approximately 5 k_BT or higher, helices 1 and 4 (black and blue) possess several easily exchangeable residues with substantially lower ΔG less than 4 k_BT. A similar trend is observed for the cooperative 1111 model (right panel of Fig. 2). Here the approximate range of HX ΔG at $\varepsilon/k_BT = -3.0$ for helices 2 and 3 is between 8 and 12.5 k_BT, whereas the corresponding approximate range for helices 1 and 4 is between 4 and 12 k_BT.

The weaker thermodynamic cooperativity of the 0000 model relative to the 1111 model is underscored by several low-ΔG isotherms of helices 1 and 4. Figure 2 (left panel) shows that as conditions become less native (ε/k_BT less negative), some isotherms of these two helices reach $\Delta G = 0$ before the overall transition midpoint. This observation is consistent with the 0000 model's less cooperative folding.[12,40] Similar features are absent in the more cooperative 1111 model (Fig. 2, right panel) and 2442 model (Fig. 3). For the 1111 model, all isotherms converge essentially to a global unfolding isotherm prior to the overall transition midpoint. This trend is qualitatively similar to experimental native-state HX data on several proteins,[5,8,11,56] indicating that the thermodynamic cooperativity of these real proteins are higher than that of the Gō-like 0000 model. However, the points of convergence for different helices in the 1111 model are not clearly distinguishable, although helices 1 and 4 (blue and black) appear to join the global unfolding isotherm very slightly later (i.e., at less negative ε/k_BT) than helices 2 and 3 (red and green). This is not surprising because this model affords equal cooperative energy (same ε_{coop}) to all four helices.

The 2442 model is an investigative tool designed to split this apparent degeneracy by allowing different $\varepsilon_{coop}^{(i)}$ values for different helices. Figure 3 shows the resulting isotherms. It would be useful to explore also other combinations of $\varepsilon_{coop}^{(i)}$'s, but that is beyond the scope of the present work. Unlike the 0000 and 1111 models in Figure 2, the 2442 isotherms in Figure 3 are clearly separable into two groups. The midchain helices 2 and 3 are now much more stable against HX and their isotherms also converge significantly earlier (at more negative ε/k_BT) than the chain-end helices 1 and 4. This comparison of the three models' thermodynamics suggests strongly that the stability ordering and convergence of HX isotherms depend critically on the heterogeneity of the interactions involving different sets of residues and how overall cooperativity is achieved. This means that when the underlying interactions are more homogeneous, as in the 0000 and 1111 models, a lack of clear ordering would ensue. Thus, it is possible that the HX isotherms of some proteins may not converge separately.

The rank orders of HX free energies of individual residues in the three models are well correlated. At $\varepsilon/k_BT = -3.0$, the HX ΔGs of helices 1 and 4 for the three models have essentially the same rank order. For helices 2 and 3, the three most stable residues against HX are

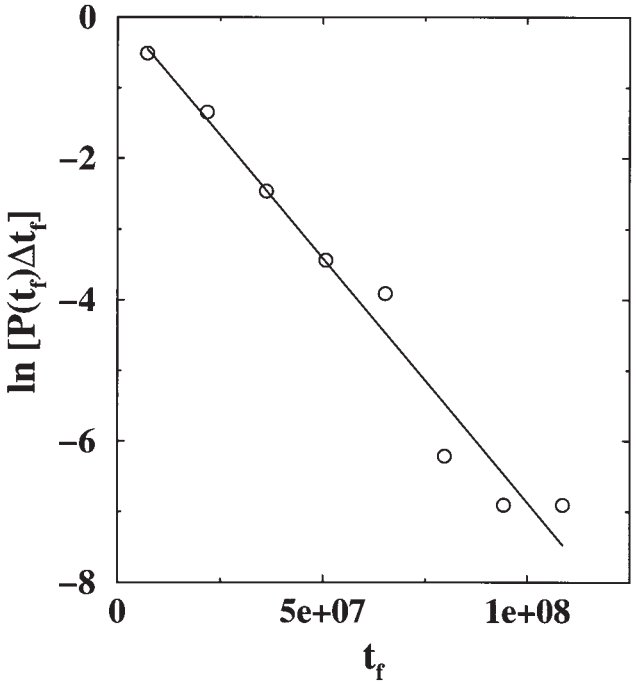

Fig. 4. Approximate single-exponential folding kinetics of the 2442 model at $\epsilon/k_BT = -2.105$. Open circles show the distribution of folding time t_f among 1000 independent folding trajectories. The continuous line is a least-square fit through the data points. Folding trajectories in this and other models in the present study are initiated from a random conformation, and t_f is defined as the first passage time (FPT) through the ground state. $P(t_f)\Delta t_f$ is the fraction of trajectories with $t_f - \Delta t_f/2 < \text{FPT} \leq t_f + \Delta t_f/2$. The distribution is shown here for bin size $\Delta t_f = 1.5 \times 10^7$, which is the standard deviation of the t_f distribution.

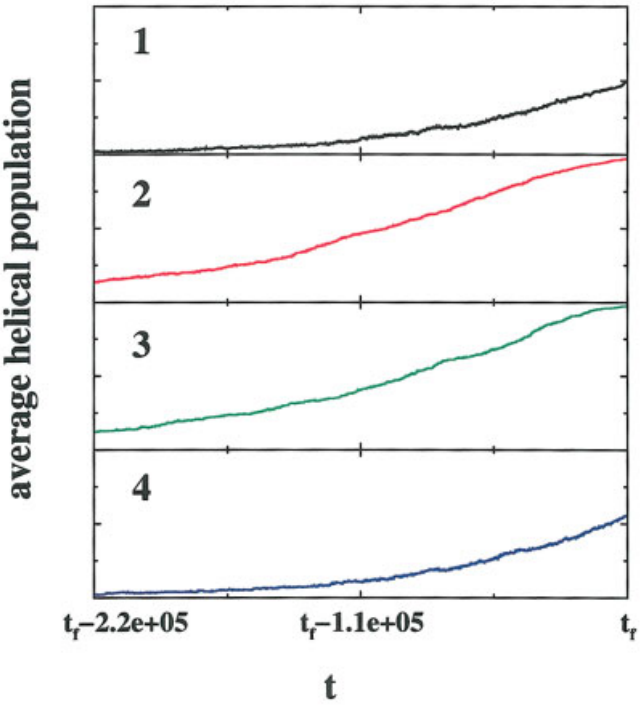

Fig. 5. Helix formation during the last stages of folding of the 2442 model is tracked by the fraction of time a fully formed helix 1, 2, 3, or 4 is present (vertical scale, averaged over 500 trajectories), as a function of the time interval before the ground state is reached at a given trajectory's folding time t_f. It should be noted that different folding trajectories have different t_f values (cf. Fig. 4).

identical for the three models. Among the 36 intrahelical native contacts we use to define exchangeability (Fig. 2), eight are completely buried in the native core. Consistent with expectation, six of these eight [(5,8), (8,3), (21,16), (40,35), (53,48), (48,51)] are among the most protected in that their ΔG values are high under strongly native conditions (detailed data not shown). The exceptions are (35,30) and (26,21), which have low HX ΔGs. The latter intrahelical native contacts are more readily broken partly because of their sequential proximity to the flexible three-residue (27–29) loop linking helices 2 and 3.

Thermodynamic Cooperativity Produces Approximate Single Exponential Relaxation

The folding kinetics of the three models are simulated by standard lattice Monte Carlo dynamics, using the method detailed in ref. 40. We focus here on the kinetics of folding under native conditions (Figs. 4–6). Folding trajectories for the three models were simulated at the ϵ/k_BT values provided in Figure 3 and in the legend for Figure 6. These interaction strengths are chosen because they lead to relatively high overall native stabilities as measured by the free energy differences between the ground state and an ensemble of highly open conformations[40] (≈10, 12.5, and 17 k_BT, respectively, for the 0000, 1111, and 2442 models). Figure 4 shows the logarithmic distribution of folding time for the 2442 model. The approximate linearity of this distribution—similar to that for the 1111 model[40]—implies that kinetic relaxation for folding is essentially single exponential[59] at the ϵ/k_BT value marked by the vertical dotted line in Figure 3. This behavior echoes that of a number of real proteins that have been investigated by native-state HX, including cytochrome *c* (refs. 8, 44) and Rd-apocyt b_{562} (ref. 56).

Thermodynamics-Kinetics Correlation Exists, but Need Not Be Universal

We monitor the formation of the four helices during the folding process of the 2442 model and find that on average, the formation of whole helices in this model occurs very late in folding. Figure 5 shows that when one or more completely formed native helices start to appear, the model protein is very close to reaching the native conformation. The mean (average) folding time estimated from Figure 4 is $\langle t_f \rangle = 1.51 \times 10^7$ (t_f ranging from 8.59×10^4 to 1.19×10^8 among the simulated trajectories). The time window in Figure 5, which records events immediately preceding the completion of folding, is less than 1.5% of the mean folding time. This implies that an overwhelming majority of the conformational population does not acquire one or more relatively stable and completely formed helices until, on average, ~98.5% of the folding time has lapsed, although this observation does not preclude the formation (and dissolution) of partial helices for more extended periods during the folding process (see below). A

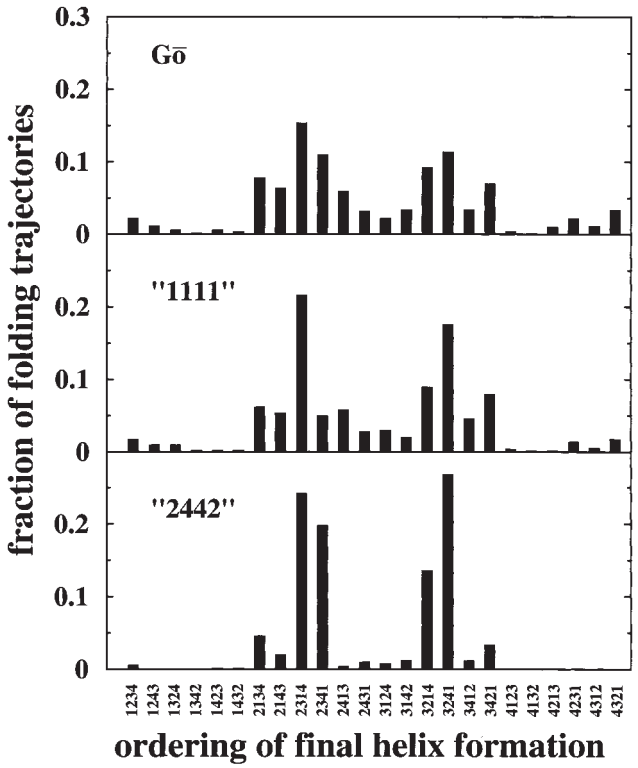

Fig. 6. Ordering of final helix formation. For each of the models, the histogram shows the distribution, among 500 independent folding trajectories, of the ordering (as defined in text) in which the four helices are fixed during the last stages of folding. Folding kinetics for the 0000 (Gō) model, the cooperative 1111 model, and the 2442 models were simulated at $\epsilon/k_B T = -2.703$, -2.50, and -2.105, respectively.

clear message from Figure 5 is that chains tend to acquire helices 2 and 3 before they complete the formation of helices 1 and 4 (though this does not apply to *every* individual trajectory; see below). At any given time preceding the final folding event at t_f, the *average* populations of helices 2 and 3 are always ahead of that of helices 1 and 4. Once helices 2 and 3 are formed, the formation of helices 1 and 4 tends to be relatively quick compared to the time scale in Figure 5. Thus, these events appear to be instantaneous in this plot (i.e., average populations of helices 1 and 4 jump from $\approx$0.5 to unity at t_f). Because the 2442 model HX isotherms for helices 2 and 3 show a clear convergence before that for helices 1 and 4 (Fig. 3), the kinetic simulation results in Figure 5 suggest that the behavior of this model is consistent with the HX thermodynamics-kinetics correlation hypothesis.[8] However, this does not necessarily imply that such properties are generally applicable to all proteins. A case in point is the 1111 model, which is calorimetrically and kinetically cooperative[40] but does not exhibit a clear stability ordering of HX isotherms (cf. Fig. 2).

Sequential Ordering of Late Folding Events Is Statistical, Not Absolute

Motivated by the average trend in Figure 5, we now examine individual folding trajectories to ascertain the order in which the last time a given helix is completely formed (fixed) before or at the very moment when the model protein is fully folded at t_f. Each trajectory can adopt 1 out of 24 possible ways to order the final formation ("fixation") of the four helices. The distributions of helix fixation order among the simulated folding trajectories of the three models are given in Figure 6, where the helix fixation order is represented by a string of four numbers denoting the helices. For example, the label "1234" (along the horizontal axis) means that helix 1 is fixed before helix 2, then helix 2 is fixed before helix 3, and finally helix 3 is fixed before helix 4. Figure 6 shows that the distribution of helix fixation order is highly nonuniform. Some orders are much more popular than others. For all three models studied in this work, it is clear that there is a preference for helices 2 and 3 to be fixed before helices 1 and 4. This effect is much more prominent for the 2442 model than for the 1111 model. Corresponding biases are not as strong for the 0000 (Gō) model, but helices 2 and 3 are still favored to be fixed before helices 1 and 4.

Two observations from these simulations are noteworthy: (i) under certain circumstances, as exemplified by the 2442 model here, a clear stability order of the convergence points of HX isotherms (Fig. 3) can be indicative of strong statistical preferences for a corresponding sequence of kinetic events during folding (Fig. 5 and the bottom panel of Fig. 6). (ii) On the other hand, the 0000 and 1111 model results also indicate that a lack of clear stability order of the convergence points of HX isotherms (Fig. 2) does not preclude statistical preferences for certain sequences of kinetic events (Fig. 6, top and middle panels).

Although Figure 6 shows that some kinetic orderings of helix fixation are much more dominant than others, the same set of data also indicates that other (nondominant) orderings of kinetic events are possible and folding can proceed via such alternate parallel pathways[53,54] (Fig. 7). For the 2442 model, approximately 84% of the folding trajectories follow one of the four most dominant temporal sequences of helix fixation: helix 2 $\rightarrow$ helix 3 $\rightarrow$ helix 1 $\rightarrow$ helix 4, helix 2 $\rightarrow$ helix 3 $\rightarrow$ helix 4 $\rightarrow$ helix 1, helix 3 $\rightarrow$ helix 2 $\rightarrow$ helix 1 $\rightarrow$ helix 4, or helix 3 $\rightarrow$ helix 2 $\rightarrow$ helix 4 $\rightarrow$ helix 1 (Fig. 6, bottom panel). Helices 2 and 3 are always fixed before helices 1 and 4 in these dominant pathways. In this sense, their orderings of late kinetic events do follow the thermodynamic ordering of HX isotherms convergence in Figure 3. However, this still leaves 16% of folding trajectories with kinetic orderings that differ from that suggested by the thermodynamic HX isotherms convergence pattern. Therefore, these results indicate that the validity of the thermodynamics-suggested "sequential" kinetic folding pathway (first helices 2, 3 then helices 1, 4) is statistical but not absolute. Moreover, Figure 6 shows that the same set of four temporal sequences of helix fixation also constitutes a major folding pathway of the 1111 model ($\approx$ 53% of the folding trajectories) and the 0000 (Gō) model ($\approx$47% of the folding trajectories). But this pathway is much less dominant in these models than in the 2442 model. Recalling that the HX isotherms of the 0000 and 1111 models do not have more than one clearly

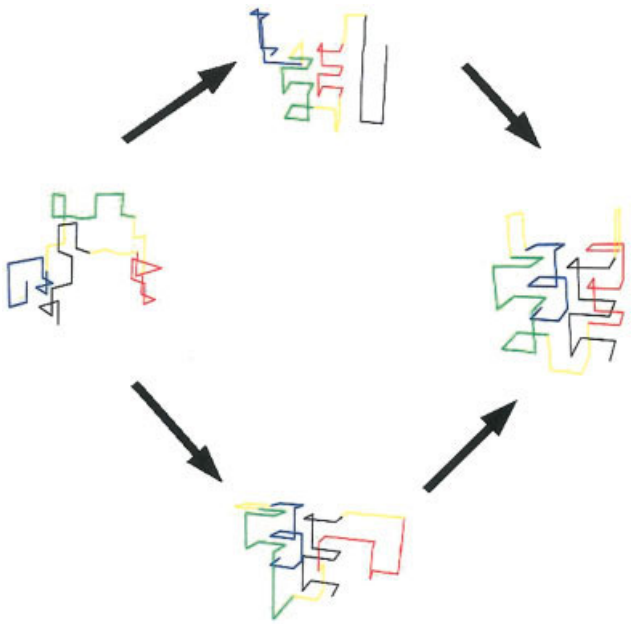

Fig. 7. Schematics showing possible parallel folding "pathways." To reach the native state (right), an unfolded conformation (left) can first form helices 2 and 3 (middle, upper pathway), or it can first form helices 1 and 4 (middle, lower pathway). For the 2442 model, the upper pathway is much more likely but the lower pathway is still possible (cf. Fig. 6).

distinguishable point of convergence (Fig. 2), the comparison in Figure 6 suggests that the dominance of a specific sequential folding kinetic pathway (as in the 2442 model) may in general be concomitant with a native-state HX pattern of well-separated points of isotherms convergence (Fig. 3). By the same token, for models with points of isotherms convergence even more separated than the present 2442 model, or for models with a higher overall thermodynamic cooperativity, it is conceivable that one particular sequence of late kinetic events can overwhelm alternate pathways even more than the dominant sequential order observed for the 2442 model here. It would be interesting to explore this prospect in future investigations.

A more detailed view of the ordering of kinetic events and alternate folding pathways is provided by Figures 8 and 9. These plots address late folding events; they cover only the last ~4% of their respective folding trajectories. The time evolution of the overall folding parameter Q and the fractional number of native helical contacts of individual helices are tracked. The information here is complementary to the average trend in Figure 5. In particular, these traces account not only for completely formed helices (as in Fig. 5) but also partially formed helices. It should be noted that the kinetics of folding transitions in these trajectories is two-state-like. The time evolution of the folding transition appears to be quite gradual on these plots (cf. top panels for Q) only because the time scales of the plots are short compared to the folding time. In this respect, Figures 8 and 9 highlight the fact that folding transitions that are considered to be "sudden" in terms of the folding time scale (e.g., perceived as a jump from low Q to high Q) do in fact consist of kinetic steps that traverse conformations with intermediate nativeness. This feature is natural and not surprisng, and is expected from an energy landscape view of protein folding.

The trajectory in Figure 8 belongs to the dominant pathway. Helices 2 and 3 are stabilized before helices 1 and 4. But the trajectory in Figure 9 does not follow this pattern. In this case, helices 1 and 4 are stabilized prior to helices 2 and 3, graphically illustrating the existence of alternate parallel folding pathways in this model. A salient feature in both sets of kinetic traces is their extensive fluctuations before the final folding event (the point at which all traces reach unity and stay flat). For each helix, another time point can be identified at which the helix is first completely formed and starts to maintain a high level of helicity (i.e., a high level of fractional helical contacts). Extensive fluctuations are observed both before and after this point. Before this point, the helix is partially formed and fluctuating between medium to low helicities, undergoing cycles of formation and dissolution of partial helices. Even after the point when a complete helix is achieved, for all helices except the one that forms last, fluctuations continue between a completely formed helix and partial helices with relatively high helicities. From a statistical mechanics perspective, we believe that this dynamic chain picture should reflect the physical situation of the late events in real protein folding.

It is clear from the above analysis of folding trajectories that although the "sequential" folding pathway suggested by native-state HX may be statistically valid in some situations, it applies only in a restricted sense. Temporally, HX-suggested kinetic ordering concerns only the very late stages of the folding, which in our model correspond to only a few percent of the total folding time. Structurally, Figures 8 and 9 show that the formation of individual helices—which are the constituents of PUFs[8] in the present model—does not conform to a sequential "all-or-none" paradigm. During folding, a native helix does not just form and stay completely formed until folding is completed. The present results suggest that in general it undergoes many cycles of partial dissolution and reformation prior to its fixation just before the eventual folding of the entire protein.

Denaturant-Independent HX

Finally, we reiterate that the models studied in this work are coarse-grained and highly simplified. In particular, the strengths of all interactions are taken to scale uniformly by a single parameter ϵ/k_BT that models denaturant concentration. As such, the present models may not capture some of the finer details of experimental native-state HX. For example, it is apparent that under strongly native conditions, HX can often proceed via denaturant-independent processes.[8,11] Possible mechanisms of such HX include "penetration"[3,25] and "local unfolding."[4,42] Our models are not equipped to address the validity or plausibility of these hypotheses because they involve proposed small conformational fluctuations that are probably below the structural and kinetic resolution of the present coarse-

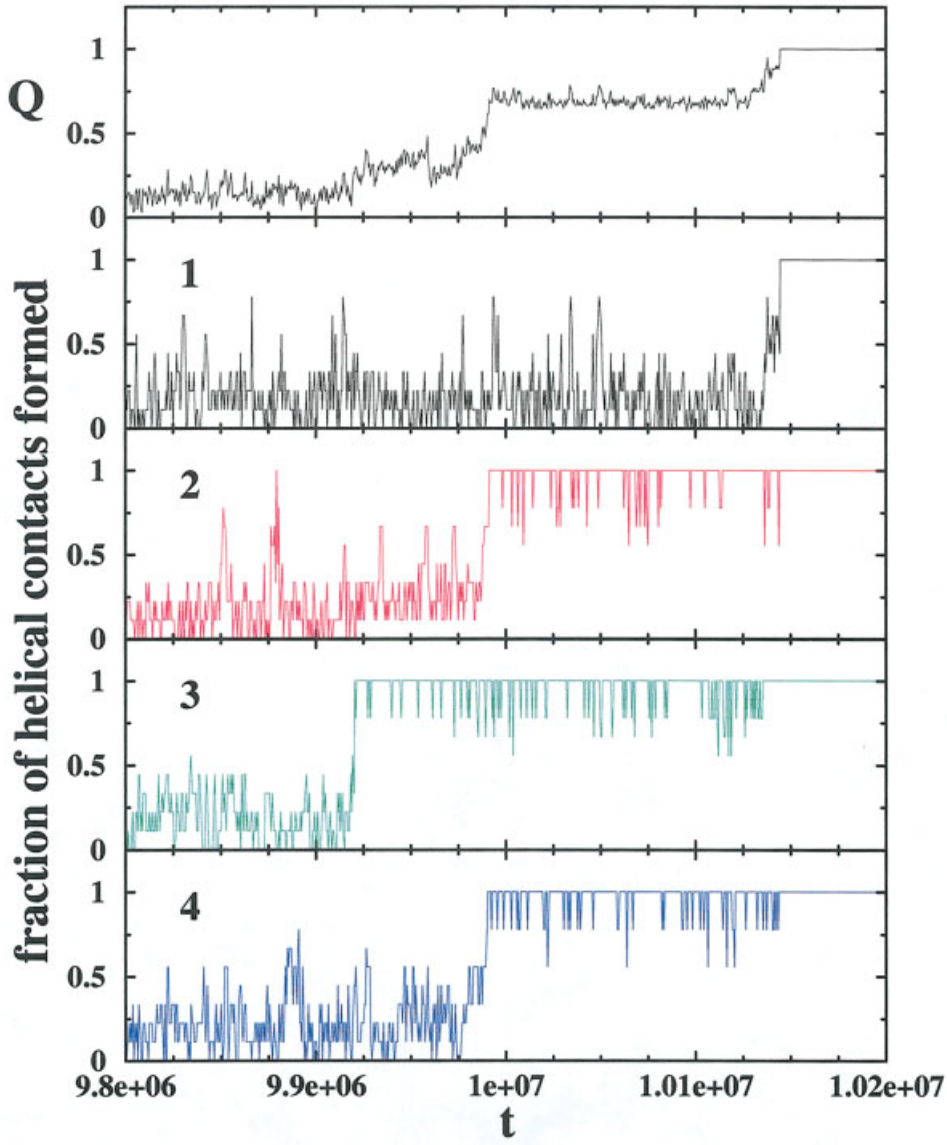

Fig. 8. A typical 2442 model folding trajectory at $\epsilon/k_BT = -2.105$ in which helices 2 and 3 are substantially populated prior to helices 1 and 4. Progress in overall folding is tracked by the fraction Q of total native contacts (top). The other four panels provide time evolution of the fraction of intrahelix native contacts for each of the four individual helices.

grained models. Nonetheless, Figure 10 shows that such effects may be accounted for qualitatively in our models using an approach similar to that of Qian et al.[7]—albeit in a somewhat ad hoc yet physical manner—by allowing for the existence of small ϵ-independent HX-competent conformational populations. Under such an assumption, the HX isotherms would naturally level off under strongly native conditions (dashed curve in Fig. 10), resembling what has been observed experimentally in some instances.[8,11] Similar effects may also be achieved by postulating a more complex potential function with terms having different dependencies on ϵ. We do not pursue these modeling possibilities further here because their impact on the issues at hand—that is, isotherms convergence and kinetic ordering—is minimal.

CONCLUSION

In this work, we have used explicit-chain models to probe the logic of native-state HX thermodynamics-kinetics correlation. Although the model interaction schemes are simplified, they are motivated by physical considerations. In our models, the nonlinearity of HX isotherms follows directly (and, as such, rather trivially) from the existence of conformations with energy levels intermediate between the ground state and the fully unfolded ensemble. This is the case in practically all self-contained, explicit-chain polymer models.[9,12,35] In this view, HX isotherms are expected to be curved even if the populations of all HX-competent conformations are dependent upon denaturant in a uniform manner, as in the

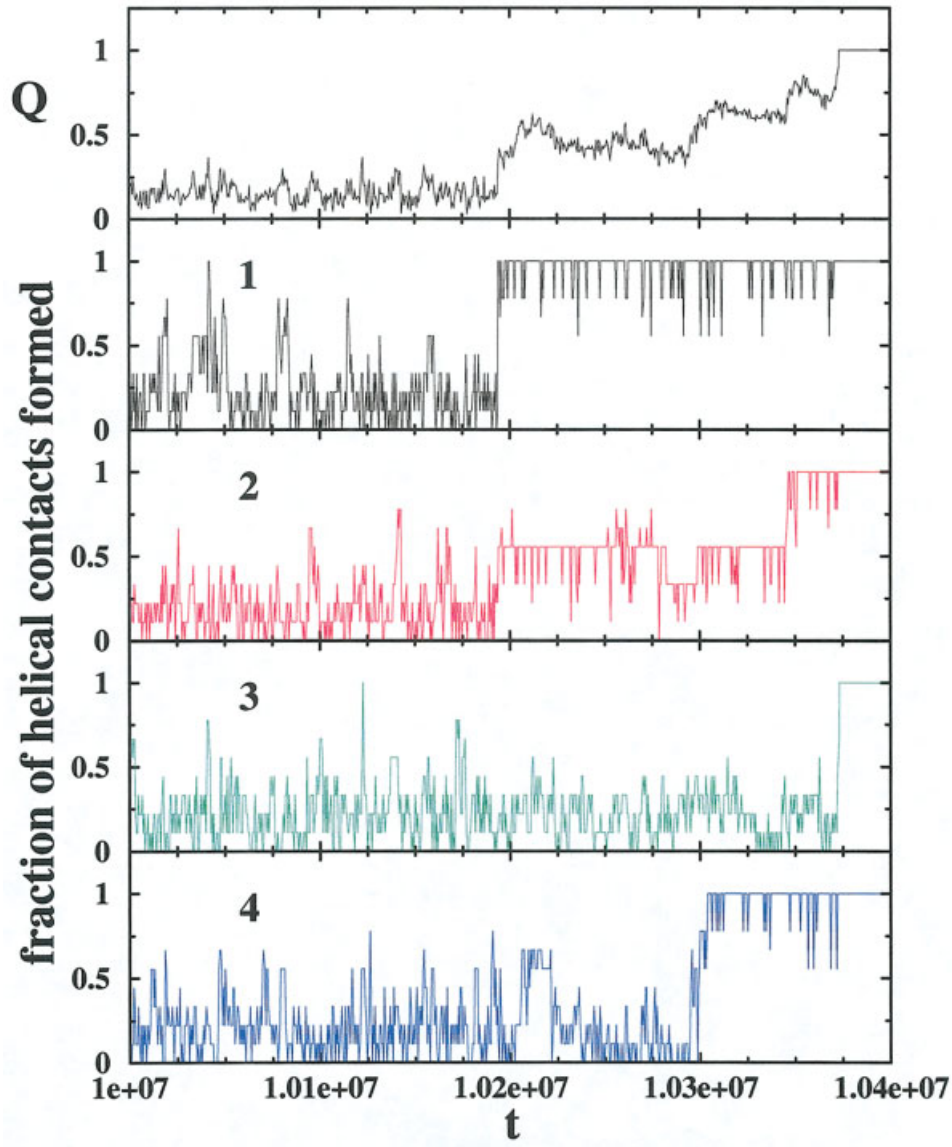

Fig. 9. Same as Figure 8, but for a folding trajectory in which helices 1 and 4 are substantially populated before helices 2 and 3.

present simplified modeling scenario (Figures 2 and 3). Hence, theoretically, HX isotherm nonlinearity per se does not require HX to proceed via multiple, physcially distinct processes. However, experimental observations that some HX isotherms are essentially flat at low denaturant do imply that, in addition to denaturant-induced shifts in conformational distribution, there are HX-enabling small denaturant-independent conformational fluctuations as well (Fig. 10).

For the most cooperative model studied in this work, we found a statistical connection between HX thermodynamics and folding kinetics. The theoretical trends discovered here share similarities with a previously proposed experimental interpretative framework linking the stabilities and convergences of HX isotherms to the kinetic ordering of late folding events that occur after the rate-limiting step of folding.[8] The present lattice model results demonstrate that experimental native-state HX observations are consistent with aspects of energy landscape perspectives of protein folding. The thermodynamics-kinetics correlation in our model is not absolute. While the kinetic "pathway" of late folding events suggested by the stability ordering of points of isotherm convergence is dominant, parallel folding pathways do exist. Hence, for our model, folding is not absolutely sequential. Moreover, the late folding events constitute only a very small percentage of the total folding time. Thus, knowledge about these events alone is insufficient to characterize the kinetics during a majority of the time a protein takes to fold, nor is it sufficient by itself for understanding the processes required to overcome the major rate-limiting barrier to folding.

A key factor that has emerged is that thermodynamic and kinetic cooperativity,[38] which proteins might have

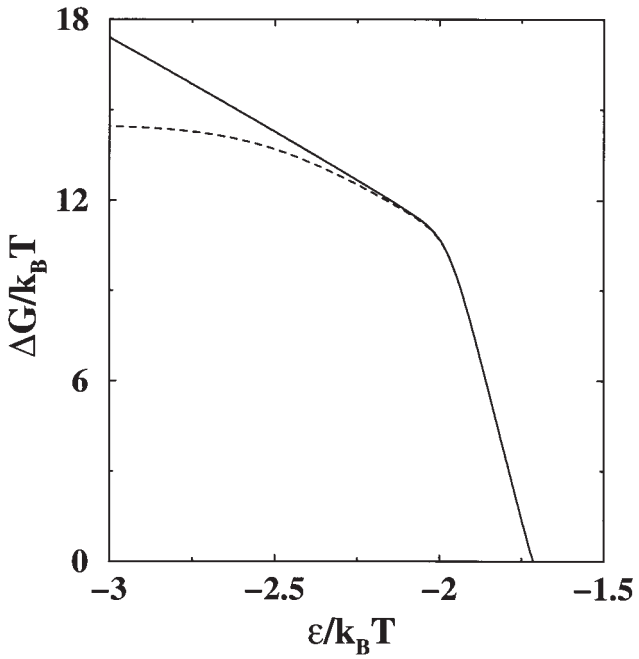

Fig. 10. A model HX isotherm that incorporates a hypothetical denaturant-independent HX mechanism. The solid curve is the (35,30) isotherm in Figure 3. The dashed curve is constructed by augmenting the HX thermodynamics of the 2442 model with a "background" population of HX-competent state whose free energy ΔG_0 relative to the ground state is independent of ϵ. The modified isotherm here is given by $-\ln\{\exp[-\Delta G(35,30)/k_BT] + \exp[-\Delta G_0/k_BT]\}$ versus ϵ/k_BT, where $\Delta G(35,30)$ is the ϵ-dependent HX free energy of the (35,30) isotherm; and we have used $\Delta G_0/k_BT = 14.5$ as an illustration.

evolved to guard against aggregation,[66] is key to understanding natural protein behavior.[67] Cooperativity criteria have not been sufficiently taken into account in earlier lattice protein models.[38] Here we found that they are critically important for establishing a quantitatively more adequate match between protein chain models and native-state HX experiments by appropriately reducing the conformational population with intermediate energies.

It should be noted that this work focuses only on an apparent two-state model helical protein. The basis of the present 2442 model's high degree of cooperativity is a particular form of local-nonlocal coupling mechanism.[40] The separation of points of HX isotherms convergence follows from the same cooperativity mechanism as well. However, there are other possible mechanisms that can induce high cooperativity,[40,68] including a more generalized local-nonlocal coupling.[38,63] Therefore, in lieu of a more exhaustive deciphering of the energetics of different proteins, at this moment there is no compelling reason to believe that the features observed here are universally applicable to all proteins. Physically, it would not be surprising that some other proteins, especially small, single-domain proteins, may not have chain segments that behave like separate cooperative folding units with clearly distinguishable stabilities.[15] Nevertheless, using a set of physically plausible model interaction schemes, here we have provided a physical scenario showing how, as an empirical correlation, native-state HX patterns can be predictive to a degree of the kinetic ordering of late folding events for some proteins. It would be extremely interesting to perform further evaluations using other modeling scenarios with different interaction schemes to better delineate the utility and limitations of such thermodynamics-kinetics correlations.

ACKNOWLEDGMENTS

We thank Yawen Bai, Walter Englander, Hong Qian, Tobin Sosnick, and Clare Woodward for very helpful discussions, Yawen Bai for a critical reading of an earlier version of this article, and Doug Barrick, Susan Marqusee, and Terry Oas for their comments on some of our preliminary results.

REFERENCES

1. Hvidt A, Linderstrøm-Lang K. Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions. Biochim Biophys Acta 1954;14:574–575.
2. Wüthrich K, Wagner G, Richarz R, Braun W. Correlations between internal mobility and stability of globular proteins. Biophys J 1980;32:549–560.
3. Woodward CK, Hilton BD. Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. Biophys J 1980;32:561–575.
4. Englander SW, Calhoun DB, Englander JJ, Kallenbach NR, Liem RKH, Malin EL, Mandal C, Rogero JR. Individual breathing reactions measured in hemoglobin by hydrogen-exchange methods. Biophys J 1980;32:577–589.
5. Kim K-S, Woodward C. Protein internal flexibility and global stability: Effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. Biochemistry 1993;32:9609–9613.
6. Mayo SL, Baldwin RL. Guanidinium chloride induction of partial unfolding in amide proton exchange in RNase A. Science 1993;262:873–876.
7. Qian H, Mayo SL, Morton A. Protein hydrogen exchange in denaturant: Quantitative analysis by a two-process model. Biochemistry 1984;33:8167–8171.
8. Bai Y, Sosnick T, Mayne L, Englander SW. Protein folding intermediates: Native-state hydrogen exchange. Science 1995;269:192–197.
9. Miller DW, Dill KA. A statistical mechanical model for hydrogen exchange in globular proteins. Protein Sci 1995;4:1860–1873.
10. Bai Y, Englander SW. Future directions in folding: The multi-state nature of protein structure. Proteins 1996;24:145–151.
11. Llinás M, Gillespie B, Dahlquist FW, Marqusee S. The energetics of T4 lysozyme reveal a hierarchy of conformations. Nature Struct Biol 1999;6:1072–1077.
12. Chan HS. Modeling protein density of states: Additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. Proteins 2000;40:543–571.
13. Kim K-S, Fuchs JA, Woodward CK. Hydrogen exchange identifies native-state motional domains important in protein folding. Biochemistry 1993;32:9600–9608.
14. Chamberlain AK, Handel TM, Marqusee S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. Nature Struct Biol 1996;3:782–787.
15. Clarke J, Itzhaki LS, Fersht AR. Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways? Trends Biochem Sci 1997;22:284–287.
16. Clarke J, Fersht AR. An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. Fold Des 1996;1:243–254.
17. Khan F, Chuang JI, Gianni S, Fersht AR. The kinetic pathway of folding of barnase. J Mol Biol 2003;333:169–186.
18. Chu RA, Bai YW. Lack of definable nucleation sites in the rate-limiting transition state of barnase under native conditions. J Mol Biol 2002;315:759–770.
19. Vu ND, Feng HQ, Bai YW. The folding pathway of barnase: The rate-limiting transition state and a hidden intermediate under native conditions. Biochemistry 2004;43:3346–3356.

20. Englander SW. Native-state HX. Trends Biochem Sci 1998;23: 378.
21. Woodward C, Li R. The slow-exchange core and protein folding. Trends Biochem Sci 1998;23:379.
22. Clarke J, Itzhaki LS, Fersht AR. A reply to Englander and Woodward. Trends Biochem Sci 1998;23:379–381.
23. Xu Y, Mayne L, Englander SW. Evidence for an unfolding and refolding pathway in cytochrome *c*. Nature Struct Biol 1998;5:774–778.
24. Hoang L, Bédard S, Krishna MMG, Lin Y, Englander SW. Cytochrome *c* folding pathway: Kinetic native-state hydrogen exchange. Proc Natl Acad Sci USA 2002;99:12173–12178.
25. Li R, Woodward C. The hydrogen exchange core and protein folding. Protein Sci 1999;8:1571–1591.
26. Hilser VJ, Freire E. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. J Mol Biol 1996;262:756–772.
27. Bahar I, Wallqvist A, Covell DG, Jernigan RL. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. Biochemistry 1998;37:1067–1075.
28. Rader AJ, Bahar I. Folding core predictions from network models of proteins. Polymer 2004;45:659–668.
29. Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 1982;51:459–489.
30. Zwanzig R. Simple model of protein folding kinetics. Proc Natl Acad Sci USA 1995;92:9801–9804.
31. Alm E, Baker D. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. Proc Natl Acad Sci USA 1999;96:11305–11310.
32. Muñoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. Proc Natl Acad Sci USA 1999;96:11311–11316.
33. Freire E, Murphy KP. Molecular basis of co-operativity in protein folding. J Mol Biol 1991;222:687–698.
34. Weikl TR, Palassini M, Dill KA. Cooperativity in two-state protein folding kinetics. Protein Sci 2004;13:822–829.
35. Kaya H, Chan HS. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: How applicable is the transition state picture to folding and unfolding? J Mol Biol 2002;315:899–909.
36. Karanicolas J, Brooks CL. The importance of explicit chain representation in protein folding models: An examination of Ising-like models. Proteins 2003;53:740–747.
37. Kaya H, Chan HS. Polymer principles of protein calorimetric two-state cooperativity. Proteins 2000;40:637–661 [Erratum: 2001; 43:523].
38. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. Methods Enzymol 2004;380:350–379.
39. Klimov DK, Thirumalai D. Is there a unique melting temperature for two-state proteins? J Comput Chem 2002;23:161–165.
40. Kaya H, Chan HS. Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity. Proteins 2003;52:510–523.
41. Knott M, Chan HS. Exploring the effects of hydrogen bonding and hydrophobic interactions on the foldability and cooperativity of helical proteins using a simplified atomic model. Chem Phys Forthcoming.
42. Englander SW, Mayne L, Bai Y, Sosnick TR. Hydrogen exchange: The modern legacy of Linderstrøm-Lang. Protein Sci 1997;6:1101–1109.
43. Kim PS, Baldwin RL. Intermediates in the folding reaction of small proteins. Annu Rev Biochem 1990;59:631–660.
44. Sosnick TR, Mayne L, Hiller R, Englander SW. The barriers in protein folding. Nature Struct Biol 1994;1:149–156.
45. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins 1995;21:167–195.
46. Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. Principles of protein folding—A perspective from simple exact models. Protein Sci 1995;4:561–602.
47. Thirumalai D, Woodson SA. Kinetics of folding of proteins and RNA. Acc Chem Res 1996;29:433–439.
48. Shakhnovich EI. Theoretical studies of protein folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
49. Chan HS, Dill KA. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. Proteins 1998;30:2–33.
50. Milne JS, Xu Y, Mayne LC, Englander SW. Experimental study of the protein folding landscape: Unfolding reactions in cytochrome *c*. J Mol Biol 1999;290:811–822.
51. Englander SW. Protein folding intermediates and pathways studied by hydrogen exchange. Annu Rev Biophys Biomol Struct 2000;29:213–238.
52. Rumbley J, Hoang L, Mayne L, Englander SW. An amino acid code for protein folding. Proc Natl Acad Sci USA 2001;98:105–112.
53. Wallace LA, Matthews CR. Highly divergent dihydrofolate reductases conserve complex folding mechanisms. J Mol Biol 2002;315: 193–211.
54. Ozkan SB, Dill KA, Bahar I. Fast-folding protein kinetics, hidden intermediates, and the sequential stabilization model. Protein Sci 2002;8:1958–1970.
55. Kaya H, Chan HS. Energetic components of cooperative protein folding. Phys Rev Lett 2000;85:4823–4826.
56. Chu R, Pei W, Takei J, Bai Y. Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. Biochemistry 2002;41:7998–8003.
57. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
58. Clementi C, Nymeyer H, Onuchic JN. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. J Mol Biol 2000;298: 937–953.
59. Kaya H, Chan HS. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: How adequate is native-centric topological modeling? J Mol Biol 2003;326:911–931 [Corrigendum: 2004;337:1069–1070].
60. Kaya H, Chan HS. Origins of chevron rollovers in non-two-state protein folding kinetics. Phys Rev Lett 2003;90:258104.
61. Pande VS, Rokhsar DS. Folding pathway of a lattice model of proteins. Proc Natl Acad Sci USA 1999;96:1273–1278.
62. Plaxco KW, Simons KT, Ruczinski I, Baker D. Topology, stability, sequence, and length: Defining the determinants of two-state protein folding kinetics. Biochemistry 2000;39:11177–11183.
63. Kaya H, Chan HS. Contact order dependent protein folding rates: Kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences. Proteins 2003;52:524–533.
64. Hvidt A, Nielsen SO. Hydrogen exchange in proteins. Adv Protein Chem 1966;21:287–386.
65. Chan HS, Dill KA. The effects of internal constraints on the configurations of chain molecules. J Chem Phys 1990;92:3118–3135.
66. Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci 1999;24:329–332.
67. Scalley-Kim M, Baker D. Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection. J Mol Biol 2004;338:573–583.
68. Jewett AI, Pande VS, Plaxco KW. Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. J Mol Biol 2003;326:247–253.