doi:10.1016/j.jmb.2005.03.084 J. Mol. Biol. (2005) 349, 872–889

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Desolvation is a Likely Origin of Robust Enthalpic Barriers to Protein Folding

Zhirong Liu and Hue Sun Chan*

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

Experimental data from global analyses of temperature (T) and denaturant dependence of the folding rates of small proteins led to an intrinsic enthalpic folding barrier hypothesis: to a good approximation, the T-dependence of folding rate under constant native stability conditions is Arrhenius. Furthermore, for a given protein, the slope of isostability folding rate *versus* $1/T$ is essentially independent of native stability. This hypothesis implies a simple relationship between chevron and Eyring plots of folding that is easily discernible when both sets of rates are expressed as functions of native stability. Using experimental data in the literature, we verify the predicted chevron–Eyring relationship for 14 proteins and determine their intrinsic enthalpic folding barriers, which vary approximately from 15 kcal/mol to 40 kcal/mol for different proteins. These enthalpic barriers do not appear to correlate with folding rates, but they exhibit correlation with equilibrium unfolding enthalpy at room temperature. Intrinsic enthalpic barriers with similarly high magnitudes apply as well to at least two cases of peptide–peptide and peptide–protein association, suggesting that these barriers are a hallmark of certain general and fundamental kinetic processes during folding and binding. Using a class of explicit-chain C_α protein models with constant elementary enthalpic desolvation barriers between C_α positions, we show that small microscopic pairwise desolvation barriers, which are a direct consequence of the particulate nature of water, can act cooperatively to give rise to a significant overall enthalpic barrier to folding. This theoretical finding provides a physical rationalization for the high intrinsic enthalpic barriers in protein folding energetics. Ramifications of entropy–enthalpy compensation in hydrophobic association for the height of enthalpic desolvation barrier are discussed.

Keywords: activation enthalpy; activation entropy; heat capacity; chevron plot; Arrhenius

*Corresponding author

Introduction

Deciphering the physico-chemical forces that drive proteins to fold is one of the central unresolved questions in molecular biology. As a challenge to the human intellect, this problem is difficult because many degrees of freedom, of the protein as well as the surrounding solvent molecules, need to be taken into account, but a precise description of the interactions involved is currently lacking. Mathematically, it is natural to formulate the folding process in terms of a high-dimensional energy landscape that provides the free energy for every protein conformation (with the energetics associated with the solvent degrees of freedom pre-averaged). It is clear from general polymer physics principles that the conformational search for the native structure, while stochastic to a certain extent, has to be partially directed by incremental energetic favorabilities[1–8] so as to allow the protein chain to circumvent the Levinthal paradox.[9–11] The energy landscape of protein folding is necessarily "funnel"-like in this particular respect.

In contrast to this funnel perspective, "barriers" or empirically defined transition states along free energy profiles (often functions of a single reaction

Abbreviations used: FPT, first passage time; PDB, Protein Data Bank.

E-mail address of the corresponding author: chan@arrhenius.med.toronto.edu

coordinate) have long been a very useful concept in the interpretation of experimental folding data.[12–15] This traditional "barrier-based" view and the funnel-like energy landscape perspective are not necessarily mutually exclusive. Indeed, logic dictates that they must be reconcilable in a valid theory of protein folding. To advance fundamental physical understanding, therefore, it is crucial that the microscopic nature of the empirical barriers be ascertained in models of protein folding that provide an explicit account of the chain connectivity and excluded volume constraints on the many degrees of freedom of a polypeptide molecule.

Obviously, a critical ingredient in explicit-chain protein folding models is the postulated potential function, or intrachain interaction scheme, that determines the shape of the multi-dimensional energy landscape. Different interaction schemes lead to different thermodynamics and kinetics. For instance, the incorporation of many-body intrachain interactions can increase the effective barrier to folding compared to that afforded by common models with only pairwise energy terms.[16–19] Some many-body interactions can lead to better agreements with the experimentally observed highly cooperative behavior[18] and trend of topology-dependent folding[20] for two-state proteins.[21,22] In this vein, one of our main goals is to explore what microscopic interaction schemes may provide a physical basis for the thermodynamics of the empirical barriers to protein folding.

Important energetic and structural information about the critical rate-limiting events in folding can be gleaned from the thermodynamic signatures, free energy, entropy, enthalpy, and heat capacity, of empirical folding barriers defined along a single or a small set of "reaction coordinates". Here, we focus primarily on folding barriers, but we note that thermodynamic data for empirical unfolding barriers are also available.[23] For folding, it has long been known that the rate is typically non-Arrhenius (logarithmic folding rate varies non-linearly with $1/T$, the reciprocal of absolute temperature).[24,25] This implies that the empirical free energy barrier to folding has non-zero enthalpic and entropic components as well as a non-vanishing heat capacity.

One major positive component of the empirical free energy barrier to folding originates from the reduction in conformational entropy (i.e. increase in entropic free energy) that accompanies the change from a large initial ensemble of open, unfolded conformations to a more restricted compact ensemble at the folding rate-limiting step. This entropic contribution has been underscored in the energy landscape picture[1–8] as well as other folding models that emphasize the role of conformational searches.[26,27] Indeed, even for a hypothetical smooth folding funnel without energetic ruggedness (no uphill folding path in free energy on the high-dimensional landscape surface as the native structure is approached from unfolded conformations), an effective free energy barrier can emerge as a result of non-perfect cancellation between energy and conformational entropy when the multi-dimensional coordinates of the landscape are projected onto a certain (collective) reaction coordinate. In the energy landscape picture itself, however, the entropic effects of conformational search is represented by the area of the high-dimensional funnel surface rather than barriers. Enthalpic barriers are different in this respect, because they may entail actual uphill folding paths on the high-dimensional energy landscape itself. Enthalpic folding barriers are consistent with a rugged energy landscape. But the heights of the enthalpic barriers considered in this work are much greater than that commonly posited for overall landscape ruggedness (see below).

The enthalpic component of the empirical folding barrier (i.e. "activation enthalpy") is expected to provide a more direct reflection of effective intrachain interactions at the rate-limiting step of folding, because by definition the enthalpic barrier is independent of the entropic effects of the conformational search. Its physical interpretation should be more definitive because the height of the enthalpic barrier can be directly determined by folding rate measurements. Unlike the entropic barrier, its value is not hampered by the uncertainties in appropriating front (pre-exponential) factors in common transition state analyses of protein folding.[5,28,29] Despite these advantages, the interpretation of folding activation enthalpy is often complicated by the substantial heat capacity associated with the folding barrier, meaning that the activation enthalpy is temperature dependent. Consequently, for some proteins at certain temperatures, the enthalpic "barrier" can vanish as the activation enthalpy can become negative.[30]

Against this backdrop, a more regular pattern of behavior was discovered in 1997 by Scalley & Baker for the apparent two-state protein L.[31] By adjusting both temperature T and denaturant concentration [d], they found that under isostability conditions (i.e. along a contour of T,[d] variables that maintains a constant $\Delta G_f/T$, where ΔG_f is the free energy of folding), the temperature dependence of folding rate is essentially Arrhenius. In other words, the dependence of logarithmic isostability folding rate on $1/T$ is essentially linear. By analyzing the experimental data available in the literature at the time,[32] these authors found that their observed pattern applied to cold shock protein B (CspB) as well. Subsequently, the same phenomenon was demonstrated for the apparent two-state N-terminal domain of the ribosomal protein L9 (NTL9) by Kuhlman *et al*.[33] using a similarly extensive study of T,[d] dependence of folding rates, a methodology they termed "global analysis". Realizing that this empirical pattern implies that the Eyring and chevron plots of a given protein are simply related by a high "intrinsic" enthalpic barrier of ≈20 kcal/mol or more, recently we found that this relationship may be generalizable to "non-two-state" proteins with chevron rollovers, at least for the case of barnase.[34] Taken together, these findings

mean that the complexity of the non-Arrhenius behavior of protein folding rate is, in large measure, reducible to that of the peculiar temperature dependence of native stability. This simplicity is of particular interest to our effort to understand the physics of the enthalpic folding barrier. Thus, the key questions now are: how general is the intrinsic enthalpic barrier phenomenon? If it is reasonably general, is there any significant correlation between the intrinsic enthalpic barrier height of a protein and its folding rate? We address both issues by performing new isostability analyses on published experimental data for a set of proteins spanning a wide range of folding rates.

The physical origin of the remarkable regularity embodied by the high intrinsic enthalpic folding barriers has barely been elucidated, although the possibility of a simple enthalpic relationship between Eyring and chevron plots for protein folding had been anticipated (in some cases prior to the experimental discovery) in lattice models of apparent two-state proteins[4,5,35,36] as well as proteins with folding-arm chevron rollovers.[35,36] In these lattice constructs, an intrinsic enthalpic barrier was introduced to control the rate of elementary chain moves, resulting in a temperature-dependent "Monte Carlo clock" for the model dynamics. The cause of such a temperature dependence was tentatively attributed to the barriers to dihedral rotations. However, the typical physical barrier height for bond rotations is of the order of k_BT, or its molar equivalence RT (Boltzmann constant or gas constant, respectively, times absolute temperature, $\sim$0.6 kcal/mol for $T \approx 300$ K), which is far from sufficient to account for experimental intrinsic enthalpic barrier heights of $\sim$20 kcal/mol. Indeed, as Scalley & Baker have commented in discussing their Model I,[31] it is implausible that the high intrinsic enthalpic folding barrier can be fully accounted for by a simple picture of conformational diffusion alone. Collective motions of the folding protein was invoked as a possible alternate rationalization, suggesting that the barriers might originate from "cooperative motions of the protein chain, perhaps involving many more-or-less simultaneous bond rotations".[35] But no quantitative physical model of such a mechanism has been proposed.

Considering that intrinsic enthalpic folding barriers are derived from underlying non-Arrhenius folding rates whose temperature dependence is clearly related[31,33] to the thermodynamic properties of hydrophobicity and other solvation effects,[37–40] we found it worthwhile to investigate the prospect that the intrinsic enthalpic barriers themselves are also mainly a consequence of solvation effects. To this end, this study employs continuum models with simplified chain representations[41–43] and native-centric (Gō-like) potentials, an approach that has proven useful for gaining insight into the thermodynamics and kinetics of protein folding[29,44–47] and native-state dynamics.[48,49] Motivated by atomistic considerations of potentials of mean force,[50] a desolvation barrier is incorporated into the native-centric pairwise contact-like potentials[51] to provide a rudimentary account of the particulate nature of the aqueous solvent. In this respect, the present model interaction scheme is similar to those in several recent studies of protein energetics[29,51–54] and protein structure prediction.[55]

Desolvation barriers in hydrophobic association have been recognized as a potential contributor to the rate-limiting step in folding.[29,51,56] In general, knowledge about the potential of mean force between a pair of small non-polar solutes can provide important insights into protein energetics.[57,58] However, the relationship between pairwise solvent-mediated interactions and folding energetics can often be complex. For instance, recent explicit-water simulations indicate that the positive heat capacity (ΔC_P) signature at the desolvation barrier between two small non-polar solutes is opposite to the negative ΔC_P for typical folding transition states,[39,40,59] and that deviations from pairwise additivity can be significant for many-body solvent-mediated interactions.[40,60] In this study, a pairwise model solvation potential is designed to capture the experimental trend. It is noteworthy in this regard that a recent more simplistic application of a similar model with different thermodynamic properties for the desolvation barriers did not produce any appreciable intrinsic enthalpic folding barrier.[34] In view of these subtleties, the particular pairwise native-centric potential postulated here should be regarded as a rudimentary "renormalized" construct, in that its ascribed thermodynamic properties should be viewed as originating from fundamentally many-body solvation effects.

Results and Discussion

Instrinsic enthalpic barriers for isostability folding

We begin by briefly summarizing a recent general formulation of enthalpic folding barriers. Motivated by experimental global-analysis findings[31,33] and earlier theoretical considerations based on lattice modeling,[5,35,36] our group has proposed that the combined temperature and denaturant dependence of folding rate k_f may be well approximated by:

$$\ln k_f(T,[d]) = F(\Delta G_f/T) - \frac{\Delta H_i^{\ddagger}}{RT} \qquad (1)$$

for a large class of apparent two-state proteins as well as certain "non-two-state" proteins.[34] Here, we aim to better elucidate the scope of applicability of this relation, which hypothesizes that folding rate under any combined temperature/denaturant condition is essentially a function only of $\Delta G_f/T$ (or, equivalently, the dimensionless $\Delta G_f/RT$) and T. The functional form of F is sensitive to the protein in question. However, whenever two or more

combined T,[d] conditions lead to the same $\Delta G_f/T$, their $\ln k_f$ contributions from the F term are identical. Consequently, the differences in their folding rates are determined entirely by their temperatures through a simple Arrhenius relation *via* a protein-dependent intrinsic enthalpic barrier $\Delta H_i^{\ddagger}$, which is defined as $-k_B\times$(slope) (or$=-R\times$(slope)), where slope is the rate of change of $\ln k_f$ with respect to $1/T$ along a constant $\Delta G_f/T$ contour. Sometimes the slope obtained experimentally in this manner may exhibit some slight variations for different values of $\Delta G_f/T$.[31,33] Nonetheless, to a good first approximation, and in view of possible experimental uncertainties, here we treat this slope and thus $\Delta H_i^{\ddagger}$ as a constant for a given protein. Owing to the scarcity of relevant experimental data, the sensitivity of $\Delta H_i^{\ddagger}$ to denaturant (urea or guanidine hydrochloride, GuHCl) has not been much explored, although there is an indication in one case that $\Delta H_i^{\ddagger}$ is relatively robust in this regard as well.[34]

It follows from equation (1) that the difference between temperature-dependent rates $\ln k_f(\Delta G_f/RT;T)$ at a constant denaturant concentration $[d]_0$ (Eyring plot, often $[d]_0=0$) and denaturant-dependent rates $\ln k_f(\Delta G_f/RT;T_0)$ at a fixed temperature T_0 (chevron plot) under isostability conditions (same $\Delta G_f/T$) is given by:[34]

$$\begin{aligned}&\Delta\ln k_f(\Delta G_f/RT;T)\\&\quad=\ln k_f(\Delta G_f/RT;T)-\ln k_f(\Delta G_f/RT;T_0)\\&\quad=\frac{\Delta H_i^{\ddagger}}{RT_0}-\frac{\Delta H_i^{\ddagger}}{RT}\end{aligned}\qquad(2)$$

In other words, $\Delta\ln k_f(\Delta G_f/RT;T)$ so defined has a simple linear dependence on $1/T$ with a slope equal to $-H_i^{\ddagger}/R$. It should be noted that in this formalism, the dependence of $\ln k_f(\Delta G_f/RT;T_0)$ on denaturant concentration [d] is *via* the [d]-dependence of ΔG_f. By the same token, the isostability difference in logarithmic folding rate, $\Delta\ln k_f$, between any two temperatures T_1 and T_2 is predicted by equation (1) to be:

$$\ln\left[\frac{k_f(T_1)}{k_f(T_2)}\right]_{\Delta G_f/T}=\frac{\Delta H_i^{\ddagger}}{R}\left(\frac{1}{T_2}-\frac{1}{T_1}\right)\qquad(3)$$

Experimental data support the proposed Eyring–chevron relationship

A thorough verification of the general equation (1) entails folding rate measurements for a large number of different T,[d] conditions. Thus far, such global analyses have been performed on only a handful of proteins.[31,33] Now equation (2) offers an avenue to corroborate the intrinsic enthalpic folding barrier idea[34] for a larger number of proteins for which both Eyring and chevron plots are available, pending more extensive verification of equation (1). Figure 1 provides such a corroboration for ten proteins. Experimental details for the folding rates analyzed here can be found in the original references in Table 1. In the $\ln k_f$ plots in Figure 1, the T-dependent $\ln k_f$ values are generally nonlinear in $\Delta G_f/RT$, but the [d]-dependent $\ln k_f$ values are all essentially linear for the range of $\Delta G_f/RT$ studied. Our isostability analysis produces an essentially linear dependence of $\Delta\ln k_f$ on $1/T$ for each protein in Figure 1, with the only exception of FKBP12 (see below). This result demonstrates that the Eyring–chevron equation (2) is a very good approximation for these proteins, lending support to the general intrinsic enthalpic folding barrier hypothesis.

For FKBP12, $\Delta\ln k_f$ *versus* $1/T$ shows a significant deviation from linearity (Figure 1, bottom right panel). As it stands, this example underscores that the intrinsic enthalpic folding barrier hypothesis is falsifiable, i.e. the corresponding linearity achieved by the other proteins in Figure 1 is not trivial. One possible reason for the non-conformity of the FKBP12 data is that its T-dependent $\Delta G_f/T$ might not be very reliable. Here, $\Delta G_f(T)$ is calculated from the T-dependence of the folding and unfolding rates. The folding rate was determined in zero denaturant; but the unfolding rate was actually measured in 7 M urea. The unfolding rate at 0 M urea was then assumed to be obtainable from the data at 7 M by a constant (T-independent) offset (Figure 6(B) of Main *et al.*[61]). However, such a procedure can introduce inaccuracies because although the logarithmic unfolding rate *versus* $1/T$ profiles for different denaturant concentrations are similar, they can also have appreciable differences (cf. results for the α subunit of tryptophan synthase in Figure 2 of Chen & Matthews[23] and the CspB example in Figure 3 of Schindler & Schmid[32]). Further experiments are necessary to ascertain whether the peculiar behavior of FKBP12 in Figure 1 is correctible by a more accurate $\Delta G_f(T)$, or is it a genuine property of this particular protein.

Figure 2 provides further isostability Eyring–chevron analyses for four other proteins whose intrinsic enthalpic folding barriers have been investigated before. Apparently, all four conform to the intrinsic enthalpic barrier hypothesis (essential linear behavior in the right panels). Within reasonable uncertainties, the $\Delta\ln k_f$ *versus* $1/T$ slopes obtained here by selected Eyring–chevron comparisons agree with previous values determined using somewhat different methods. Our estimated slope for protein L is essentially identical with that determined originally by Scalley & Baker, though our slope for CspB differs by ≈12% from theirs.[31] The slopes for CI2 and HPr (Figure 1) are within 14% and 4%, respectively, of their "calcu-calculated" slopes in Table 2 of Kuhlman *et al.*[33]

Interestingly, besides the folding of single-chain proteins, in at least two cases we found that the kinetics of bimolecular association/folding also appear to conform to an intrinsic enthalpic barrier paradigm (Figure 3). In these situations, the kinetic rate of interest is the association/folding rate k_{on} involving two molecules instead of the single-chain

Table 1. Sources of experimental protein folding data analyzed in Figures 1–5 of this study

Proteins	References	Thermodynamics[a]	Kinetics[a]	T_0 (K)[b]	Denaturant	$[d]_0$ (M)[c]	Slope[d] (-10^3)
Fyn SH3	83	Figure 1; legends to Figures 3 and 7	Figures 3 and 7	293.15	GuHCl	0.0	14.7
T4 lysozyme mutant	25, 84	Tables II and III[84]	Figures 3 and 4 and Table II[25]	285.15	GuHCl	3.0	11.2
CT AcP	85	Tables 1 and 2	Figures 6 and 7	301.15	Urea	1.1	15.7
M AcP	85	Tables 2 and 3	Figures 6 and 7	301.15	Urea	1.1	19.8
Apocyt b_5	86	Tables 1 and 2	Figures 4 and 5(B)	283.15	GuHCl	0.0	11.8
Cyt c[e]	87	Table 1	Table 2	328.7	GuHCl	2.94	16.3
Protein G (wild-type G B1)[f]	81, 88	Table I[88] and Table 1[81]	Figure 6(A) and Table II[88] and Figure 2 [81]	295	GuHCl	0.0	10.5
Protein A	77	Table 1	Table 1	310.15	GuHCl	0.0	10.9
Hen lysozyme[g]	89, 90	Legend to Figure 3[90] and Table 2[89]	Figure 3[90] and Figure 7[89]	293.15	GuHCl	0.0	14.3
FKBP12	61	Tables 1 and 5 and legend to Figure 6	Figure 6	298.15	Urea	0.0	7.4
Protein L	31	Table 1	Figure 1(B) and (C)	277	GuHCl	0.3	10.8
CspB	32	Table 1	Figure 2	299.15	Urea	0.0	12.3
CI2	30, 91	Figure 3[30] and Figure 1[91]	Figure 4[91] and Figure 1[30]	298.15	GuHCl	0.0	16.8
HPr[h]	92	Table 1	Figures 10(C) and 11(B)	293.15	GuHCl	0.0	13.7
RNaseS* (Pep-1F)[i]	93, 94	Table 2 and Figure 6[93], Table 2 and Figure 3[94]	Figure 5a[93] and Figure 2a[94]	283.15	Urea	0.0	11.8
GCN4-p1[i]	95	Table 3	Figures 5 and 6(A) and Table 3	283.15	Urea	0.0	10.7

[a] The Figure(s) and/or Table(s) of the reference(s) for the given protein that provide the dependence, displayed in Figures 1–3, of logarithmic kinetic folding rates $\ln k_f$ on native thermodynamic stabillity $\Delta G_f/RT$.
[b] The constant temperature, in kelvin, at which the denaturant dependence of k_f (chevron plot) is determined.
[c] The constant denaturant concentration at which the temperature dependence of k_f (Eyring plot) is measured.
[d] $(\partial[\Delta \ln k_f]/\partial[1/T])$ at constant $\Delta G_f/RT$ (with k_f in s^{-1}), in units of -1000 (cf. Figure 1).
[e] "Phase 4" of the folding kinetics of ferri-cytochrome c is analyzed.
[f] For the temperature-dependent folding rate, data for the pH 11.2 case[88] are used.
[g] Folding kinetics of the "fast phase"[89] (corresponding to the λ_2 rate[90]) is analyzed.
[h] Kinetics of the "faster" refolding phase of HPr is analyzed.
[i] For the two folding/bimolecular association reactions listed here, the folding/association rates k_{on} as functions of the thermodynamic stability of the folded complex, $\Delta G_b/RT = \ln$ [dissociation constant], are analyzed.

protein folding rate k_f considered so far. For the two bimolecular binding processes analyzed in Figure 3, the isostability Δk_{on} *versus* $1/T$ slopes are approximately 11,800 and 10,700 (Table 1), corresponding to $\Delta H_i^{\ddagger}$ values of 21–23 kcal/mol. Remarkably, these intrinsic enthalpic barrier heights are similar to that determined for the folding of quite a number of proteins (protein L, CspB, and protein G, for example). This finding suggests strongly that kinetic bottlenecks of a similar physical origin, such as the effects of desolvation to be explored below, may be at play for a large class of biomolecular folding and binding events.

In addition to the proteins plotted in Figures 1–3, isostability data for five other proteins are listed in Table 2. Details of the Eyring–chevron analyses of NTL9 and barnase have been provided elsewhere.[34] For the other proteins in Table 2, λ repressor and one of its mutants, villin 14T, and CspA, extensive temperature-dependent folding rates are not available. In each of these cases, an intrinsic enthalpic folding barrier is estimated by equation (3) using folding rates at two temperatures. Because the isostability analyses for this last set of proteins are so limited, it remains to be verified by more extensive rate measurements whether they conform to the intrinsic enthalpic folding barrier paradigm. Nonetheless, their isostability $\Delta \ln k_f$ slopes are included in Table 2 for completeness. However, because each of these slopes is estimated from only two datapoints (instead of multiple datapoints as in Figures 1 and 2), even if these proteins do conform to the paradigm, our current estimates of their $\Delta H_i^{\ddagger}$ values are not expected to be very accurate.

Altogether, excluding the two cases of bimolecular association in Figure 3, isostability analyses have been performed on 20 protein folding processes in this study (Figures 1 and 2, and Tables 1 and 2). The folding kinetics of most of these proteins is apparently two-state, but several proteins with folding-arm chevron rollovers are also included (Figure 4). The estimated intrinsic enthalpic folding barrier heights are quite diverse. Their isostability $\Delta \ln k_f$ slopes range approximately from $-8{,}000$ to $-20{,}000$, which translate into $\Delta H_i^{\ddagger}$ values of 16–40 kcal/mol. In view of this diversity, a natural question is whether $\Delta H_i^{\ddagger}$ is correlated with folding rate or certain simple topological properties of the native structure. Figure 4 offers an initial assessment: while the folding rates of the apparent two-state folders in the test set are, not surprisingly, well correlated with their relative contact orders,[20] the

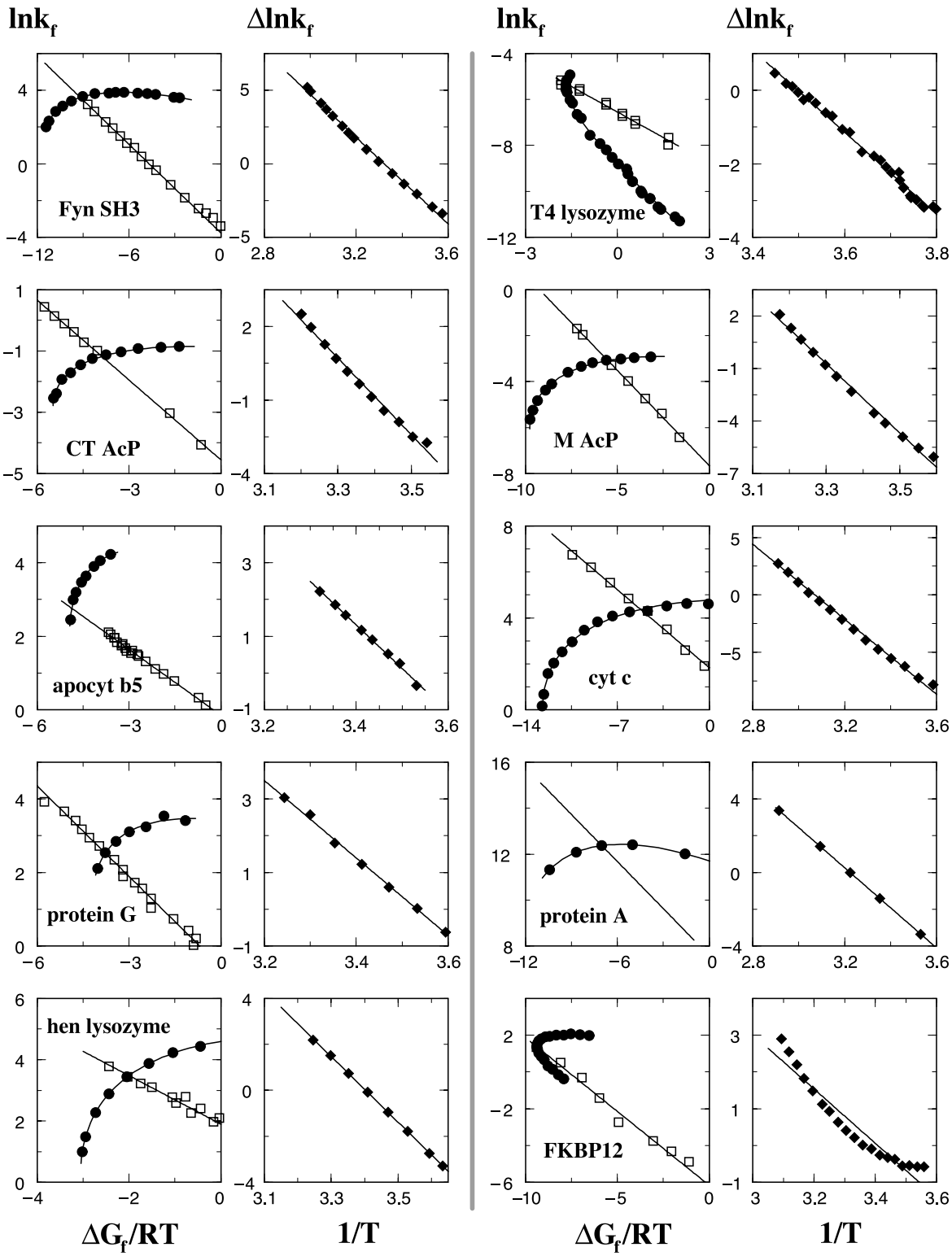

Figure 1. Isostability folding kinetic analysis of ten proteins: Fyn SH3 domain, the I3C-C97/C54T mutant of phage T4 lysozyme, common-type (CT) and muscle (M) acylphosphatase (AcP), apocytochrome b_5 (apocyt b_5), horse heart cytochrome *c* (cyt *c*), the IgG-binding domain of streptococcal protein G (protein G), the F13W G29A mutant of the B domain of staphylococcal protein A (protein A), hen lysozyme, and a 12 kDa human FK506-binding protein (FKBP12). References for the experimental data are given in Table 1. Two plots are shown for each protein. The left panel compares logarithmic folding rate ($\ln k_f$, where k_f is in units of s^{-1}, vertical scale) as a function of native stability ($\Delta G_f/RT$, horizontal scale) from (i) varying denaturant concentration $[d]$ at a constant temperature T_0 (open squares) and (ii) varying temperature T at a constant denaturant concentration $[d]_0$ (filled circles, where $[d]_0=0$ for a majority of the cases, see Table 1). Experimental data for the unfolding rate are used when appropriate to extract k_f values around transition midpoints from the total relaxation rates afforded by experimental chevron plots. In the case when the denaturant

Table 2. Additional sources of experimental protein folding data for Figures 4 and 5 in this study

Protein	References	Thermodynamics	Kinetics	T_0 (K) (or T_1, T_2)[a]	Denaturant	$[d]_0$ (M)	Slope (-10^3)
NTL9[b]	33	Table 1	Figures 2(A) and 5	292.15	GuHCl	0.0	9.3
Barnase (wild-type)[b]	96, 97	Table 1[96]	Figures 3 and 4[97]	298.15	Urea	0.0	14.9
λ_{6-85}	98, 99	Table 1[98] and Tables 2 and 3[99]	Table 1[98] and Table 2[99]	298.15, 310.15	Urea	Midpt[c]	18.5 (wild-type) 14.6 (WT*)[d]
Villin 14T	100	Table 1	Table 1	298.15, 310.15	Urea	Midpt[c]	10.6[e]
CspA	101	Table 1	Table 5	283.15, 298.15	Urea	Midpt[c]	8.1[f]

[a] The notation in this Table is essentially identical with that in Table 1 of this study. In the cases of the last three proteins for which only more limited experimental data are available, the two temperatures T_1, T_2 (in kelvin) used to estimate the intrinsic enthalpic folding barrier ($\Delta H_i^{\ddagger} = -R \times$ slope) in Figures 4 and 5 are listed instead of the single T_0 for $[d]$-dependent folding rates.
[b] Isostability folding kinetic analyses of these proteins as in the present Figures 1 and 2 are provided elsewhere.[34]
[c] Folding rates at two different denaturant concentrations $[d]_1$ (at T_1) and $[d]_2$ (at T_2) that satisfy the transition midpoint $\Delta G_f = 0$ condition are considered.
[d] The G46A/G48A mutant.
[e] The transition midpoints are determined by the folding and unfolding kinetic m-values provided in the reference.
[f] This slope is estimated for the two CspA variants H6-Xa-CspA and CspA-H6, which have essentially identical slopes. For wild-type CspA, however, the data provided in the reference lead to a rather anomalous estimated slope of ≈ -3000 with a much smaller absolute value.

$\Delta H_i^{\ddagger}$ values of the proteins exhibits no correlation with either experimental folding rate or relative contact order. Interestingly, Figure 5(a) shows a reasonable correlation between $\Delta H_i^{\ddagger}$ and the equilibrium enthalpy of unfolding ΔH_0 at room temperature for the apparent two-state proteins (Pearson coefficient $r = 0.85$). The physical origin of this somewhat surprising kinetics–thermodynamics correlation remains to be investigated. It should be noted, however, that while $\Delta H_i^{\ddagger}$ is T-independent, ΔH_0 depends on T (because of non-zero heat capacities of unfolding). The observed $\Delta H_i^{\ddagger} - \Delta H_0$ correlation is highest for $T \approx 25$–$40\,^{\circ}\mathrm{C}$, but the correlation deteriorates outside this temperature range, especially for lower T (e.g. $r < 0.5$ for $T = 10\,^{\circ}\mathrm{C}$, detailed data not shown). Figure 5(b) indicates that for the set of proteins studied, $\Delta H_i^{\ddagger}$ does not correlate with chain length.

Modeling the intrinsic enthalpic folding barriers using explicit-chain constructs with a constant desolvation barrier

To explore how high intrinsic enthalpic barriers to folding may arise from microscopic solvent-mediated intraprotein interactions, we now turn to an explicit-chain model based on a continuum C_α chain representation[29,45,46] with favorable pairwise native–centric interactions that incorporate desolvation barriers.[29,51] The potential function:

$$\begin{aligned} V_{\text{total}} &= V_{\text{stretching}} + V_{\text{bending}} + V_{\text{torsion}} \\ &\quad + V_{\text{non-bonded}} \\ &= \sum_{\text{bonds}}^{N-1} K_r (r - r_0)^2 + \sum_{\text{angles}}^{N-2} K_\theta (\theta - \theta_0)^2 \\ &\quad + \sum_{\text{dihedrals}}^{N-3} \{K_\phi^{(1)}[1 - \cos(\phi - \phi_0)] \\ &\quad + K_\phi^{(3)}[1 - \cos 3(\phi - \phi_0)]\} \\ &\quad + \sum_{i<j-3}^{\text{native}} U(r_{ij}; r_{ij}^{\text{n}}, \varepsilon, \varepsilon_{\text{db}}, \varepsilon_{\text{ssm}}) \\ &\quad + \sum_{i<j-3}^{\text{non-native}} \varepsilon \left(\frac{r_{\text{rep}}}{r_{ij}} \right)^{12} \end{aligned} \tag{4}$$

where N is the total number of C_α atoms in the given protein, r, θ, and ϕ are, respectively, the virtual bond length between successive C_α positions, virtual bond angle, and virtual torsion angle; whereas r_0, θ_0, and ϕ_0 are the corresponding native values obtained from the Protein Data Bank (PDB) structure. In the last two summations, r_{ij} is the spatial distance between two C_α atoms that are sequentially separated by at least three residues. For native

dependence was obtained from a formula with experimentally fitted parameters, only a continuous straight line is shown (without open squares). The sets of T and $[d]$-dependent experimental datapoints are then fitted, respectively, to (i) a continuous curve for $\ln k_f(\Delta G_f/RT; [d]_0, T)$ and (ii) a linear function for $\ln k_f(\Delta G_f/RT; [d], T_0)$. For each protein, the right panel shows the difference in constant-$[d]$ and constant-T logarithmic rate (vertical scale) $\Delta \ln k_f$ (equation (2)) as a function of inverse temperature ($1/T$ in units of $10^{-3}\mathrm{K}^{-1}$, horizontal scale). Here, each plotted datapoint (filled diamond) is the experimental T-dependent $\ln k_f$ value (filled circle in the left panel) minus the corresponding $[d]$-dependent $\ln k_f$ value at the same stability $\Delta G_f/RT$ (from the fitted straight line through the open squares in the left panel). Subsequently, the $\ln k_f$ *versus* $1/T$ datapoints are fitted to a linear function (as shown), the slope of which is reported in Table 1.

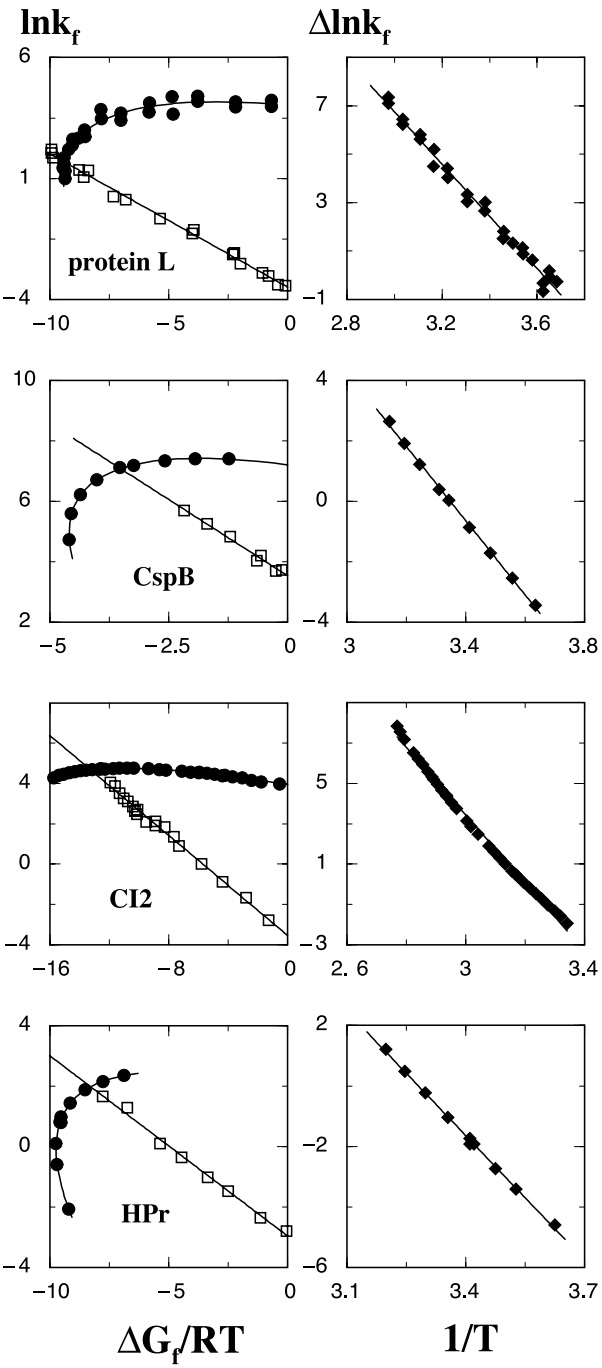

Figure 2. Same as Figure 1, now for the Y43W mutant of protein L (protein L), cold-shock protein B (CspB) from *Bacillus subtilis*, chymotrypsin inhibitor 2 (CI2), and the histidine-containing phosphocarrier protein (HPr). These proteins have been discussed in the context of global analysis,[31,33] but their Eyring and chevron plots have not been compared in the present manner. References for the experimental conditions and the $\Delta \ln k_f$ *versus* $1/T$ slopes determined here are listed in Table 1.

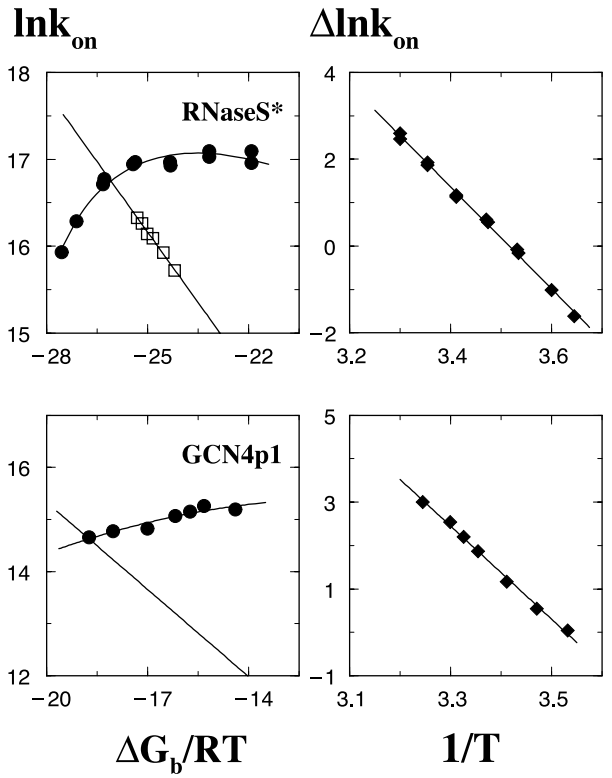

Figure 3. Essentially the same as Figure 1, but for the bimolecular association of the complex between fluorescein-labeled S-peptide and ribonuclease S (RNaseS*) (upper panels) and the homodimeric folding and association of coiled-coil peptide M2V GCN4-p1, the 33-residue leucine zipper domain of the yeast transcriptional activator GCN4 (lower panels). The folding/association rate k_{on} is in units of $s^{-1}\,M^{-1}$. The free energy ΔG_b is in units of kcal (mole of dimer)$^{-1}$, whereas the thermodynamic stability variable $\Delta G_b/RT$ of the folded complex is dimensionless. References for the experimental conditions and the $\Delta \ln k_{on}$ *versus* $1/T$ slopes determined here are listed in Table 1.

contacts, r_{ij}^n is the native C_α–C_α distance in the PDB structure between contacting residue i and j (r_{ij}^n here is equivalent to r'_{ij} in Kaya & Chan[29]). For non-native contacts, excluded volume repulsion is provided by an $(r_{ij})^{-12}$ term parameterized by r_{rep}.

The present pair potential U (Figure 6) is similar to but not identical with previous "with-solvation" potentials.[29,34] As in these earlier studies, overall thermodynamic stability is controlled by the variable $\varepsilon(>0)$, and the native-centric local bending and torsion parameters $K_\theta = 20\varepsilon$, $K_\phi^{(1)} = \varepsilon$, $K_\phi^{(3)} = 0.5\varepsilon$ scale linearly with ε. The bond stretching term is kept constant here at $K_r = 120$ (independent of ε), because it serves only to maintain an essentially constant C_α–C_α virtual bond length, which in principle should not vary with native stability,[34] although the differences in thermodynamic and kinetic results are negligible between using the present $K_r = 120$ *versus* $K_r = 100\varepsilon$ as used by Kaya & Chan[29] (comparison data not shown). In previous studies, the desolvation barriers and solvent-separated minima along the profiles of "with-solvation" pair potentials (corresponding to U) were controlled and scaled uniformly by a single parameter ε.[29,34,51] The key new feature here is that while the favorable interaction at the solvent-separated

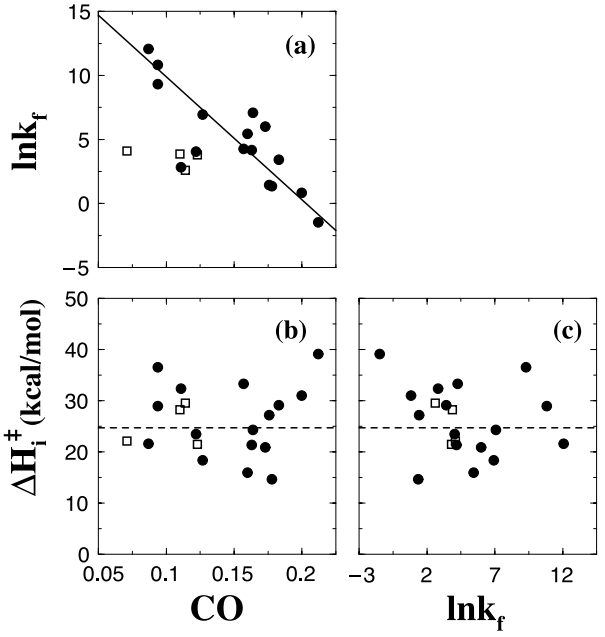

Figure 4. Lack of correlation between intrinsic enthalpic folding barrier $\Delta H_i^{\ddagger}$ and logarithmic folding rate $\ln k_f$. Folding rates for all proteins listed in Tables 1 and 2 except the two bimolecular reactions in Table 1 are considered. In addition to the proteins considered in Figures 1–3, this Figure includes data for NTL9, barnase, residues 6–85 of the N-terminal domain of λ repressor (λ_{6-85}), the N-terminal domain of chicken villin (villin 14T), and the major cold shock protein of *E. coli*. (CspA). Apparent two-state folders are represented by filled circles; apparent non-two-state folders (T4 lysozyme, hen lysozyme, barnase, and villin 14T) are denoted by open squares. Folding rates plotted were measured at 298.15 K and zero denaturant ($[d]=0$), except for CT and M AcP (301.15 K, $[d]=0$), CspB (299.15 K, $[d]=0$), cytochrome *c* (298.15 K $[d]=1.63$ M), and villin 14T (298.15 K, $[d]=1.3$ M). For the last two cases, the folding arms of the chevron plots are not linear, and directly measured $[d]=0$ rates are not available. For T4 lysozyme, we use the zero-denaturant folding rate (at 298.15 K in ^2H_2O) of the cysteine-free variant,[79] because extrapolation of the 3 M GuHCl rates[25] to $[d]=0$ is problematic.[80] For protein G, we use the pH 6.0 rate reported by McCallister *et al.*[81] For hen lysozyme, the rate for the fast phase is used. For CspA, the folding rates of the wild-type and two variants are very similar; thus only a single average of their folding rates is represented in this Figure. (a) The correlation between experimental $\ln k_f$ and relative contact order (CO) computed using the definition of Plaxco *et al.*[20] (i.e. equations (1) and (2) of Kamagata *et al.*)[82] The Pearson correlation coefficient for the 16 apparent two-state proteins is $r=-0.81$. The inclined line is the corresponding least-squares fit. (b) and (c) The intrinsic enthalpic barrier to folding $\Delta H_i^{\ddagger}$ does not correlate with either (b) CO or (c) $\ln k_f$. The horizontal broken lines in (b) and (c) mark the average $\Delta H_i^{\ddagger}$ value for all the protein folding and association processes in Tables 1 and 2.

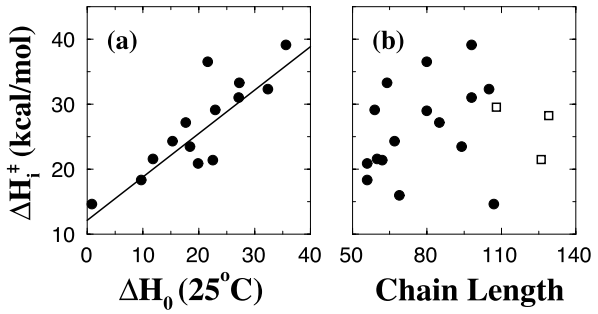

Figure 5. (a) Correlation between $\Delta H_i^{\ddagger}$ and the difference ΔH_0 in equilibrium enthalpy at 298.15 K (enthalpy in kcal/mol of the unfolded state minus that of the native state at 25 °C), calculated using the experimental references in Tables 1 and 2. Data for all the apparent two-state proteins in Figure 4 are shown, except λ^*_{6-85} and CspA for which equilibrium enthalpy data are not available. The inclined straight line is a least-squares fit. (b) $\Delta H_i^{\ddagger}$ *versus* chain length for all proteins in Figure 4 except T4 lysozyme (whose chain length = 164 is outside the plotted range).

minimum ($\varepsilon_{ssm}=0.2\varepsilon$, see Figure 6) remains native-centric and scales linearly with ε, the height of the desolvation barrier is made independent of the native stability parameter ε. Figure 7 illustrates the effects of this feature on the potential function U itself and the quantity U/T that governs conformational population distribution. Here, we explore the impact of such a constant microscopic desolvation barrier on the apparent rate-limiting step of overall folding by using mainly $\varepsilon_{db}=0.09$ and $\varepsilon_{db}=0.18$ as test cases.

Thermodynamic sampling and folding kinetic simulations are performed by Langevin dynamics[62] as specified in our earlier study, with the same integration time-step and identical parameterization for mass, viscosity, and the random force.[29] As a first test of our new modeling construct, here we apply it to chymotrypsin inhibitor 2 using the native contact set NCS2 as used by Kaya & Chan.[29] In other words, the model studied here is a modified version of the NCS2 "with-solvation" model in this reference, now with a constant K_r and the above-described ε-independent intrachain pairwise desolvation barriers.

The present model is thermodynamically cooperative, satisfying the calorimetric two-state criterion.[29] It is important, however, to note that this property does not imply a complete absence of conformations with intermediate enthalpies.[63–67] Kinetically, the present parameterizations of our model lead to folding-arm chevron rollovers,[29,34,68] implying that the models are insufficiently cooperative for apparent two-state protein folding kinetics.[36,69] Despite this limitation, the present construct should be adequate as a first continuum chain model to address possible microscopic origins of intrinsic enthalpic folding barriers, as the general pattern of behavior in question applies not only to two-state folders but to certain proteins with folding-arm chevron rollovers as well.

Inspired by explicit-water simulated potentials

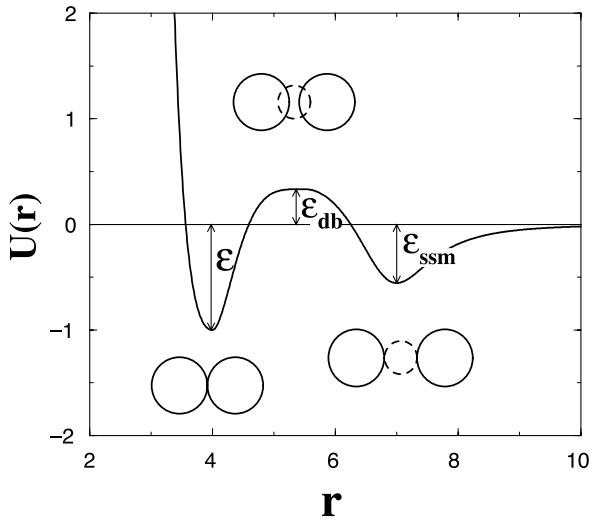

Figure 6. Potential energy function $U(r)$ between two amino acid residues that constitute a native contact pair in the present simulations, where r is the spatial separation between the two C_α positions. The function $U(r)$ corresponds to $U(r_{ij}; r_{ij}^{n}, \varepsilon, \varepsilon_{db}, \varepsilon_{ssm})$ in equation (4), with $r = r_{ij}$. The curves shown in this Figure and Figure 7 are examples of $U(r)$ functions with a particular choice of r_{ij}^{n} and the values of ε used in this study. The functional form of $U(r)$ is given by that of Cheung *et al.*[51] with the modification and specification for the "with-solvation" model (Figure 3 in Kaya & Chan[29]). The depth of the contact minimum (cm) ε has the same meaning as before, whereas the desolvation barrier (db) height ε_{db} and solvent-separated minimum (ssm) depth ε_{ssm} correspond respectively to the ε'' and ε' variables in previous studies,[29,51] i.e. $\varepsilon_{db} = \varepsilon''$, $\varepsilon_{ssm} = \varepsilon'$ (cf. Figure 1 in Cheung *et al.*[51]). The cartoons illustrate typical cm, db, and ssm configurations,[40] where water molecules are depicted by broken circles and amino acid residues are represented by continuous circles. The excluded-volume overlap of the two continuous circles with the broken circle for the db configuration conveys that in the desolvation barrier position a water molecule cannot be accomodated between the two amino acid residues.

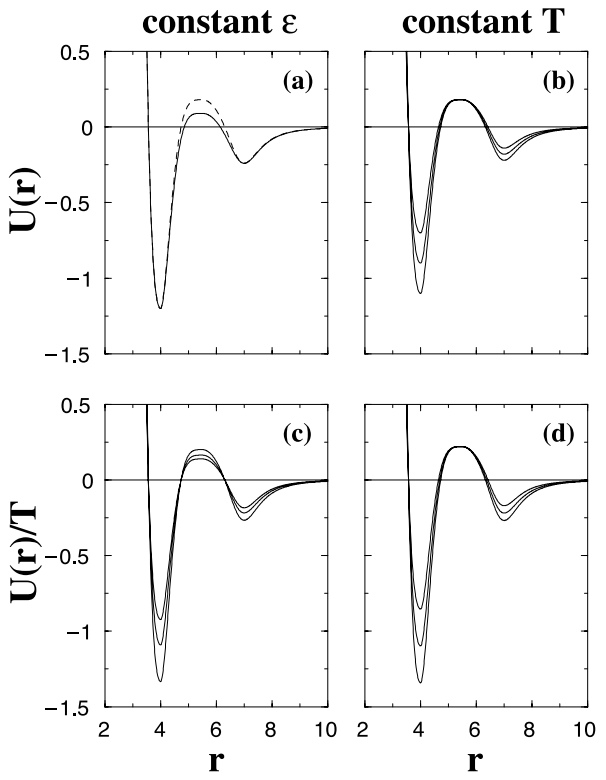

Figure 7. Model native-centric potentials with constant desolvation barrier heights. (a) Two $U(r)$ functions with ε-independent barrier heights $\varepsilon_{db} = 0.09$ (continuous curve) and $\varepsilon_{db} = 0.18$ (broken curve) are considered. In both cases, the solvent-separated minimum depth is ε-dependent, $\varepsilon_{ssm} = 0.2\varepsilon$. (b) Using $\varepsilon_{db} = 0.18$, this plot depicts examples for the variation of $U(r)$ when ε is changed while keeping temperature T constant. (c) and (d) For the same $\varepsilon_{db} = 0.18$ case, these plots show examples for the variation of $U(r)/T$ when (c) T is changed while keeping ε constant and (d) ε is changed while keeping T constant.

of mean force for small-solute hydrophobic interactions and clustering,[18,40,56,60,70–72] we focus on how the collective effects of pairwise desolvation barriers[73] between constituent groups of a protein may impact on its overall folding kinetics. This approach to solvation effects in protein energetics has proven useful.[29,51–54] The study of intrinsic enthalpic folding barriers entails comparisons of temperature-dependent *versus* denaturant-dependent rates. Here, these rates are simulated, respectively, by varying T while keeping the interaction strength ε constant, and varying ε while keeping T constant. In other words, variation in ε is used to model variation in denaturant concentration. This simple, coarse-grained approach to modeling denaturant concentration is motivated by the observation that constant-T native stability in the model is essentially linear in ε. This trend mimics the linear dependence of many real proteins' native stability on denaturant concentration,[29] even though the molecular basis of denaturant action can be rather complex.[58,74]

In essence, our model simulations compare the consequences of varying native stability in two different manners: one involves changing the average random kinetic energy of the system (varying T), whereas the other does not (varying ε). Specifically, we aim to ascertain the difference in folding rates when these two different means are used to achieve the same native stability. In matching our model results to experiments, it should be noted that solvent-mediated effective interactions are, in general, temperature-dependent. To simplify the analysis, however, ε is taken to be T-independent in the present treatment. This is only a minor limitation. If desired, it would be straightforward to incorporate temperature-dependent interactions (i.e. $\varepsilon \to \varepsilon(T)$) into our continuum model, as has been demonstrated in several lattice models.[5,35,36,60] It is clear from these earlier studies that augmenting the model with $\varepsilon \to \varepsilon(T)$ would only modify the temperature dependence of native

stability; it will not fundamentally affect the general conclusions regarding intrinsic enthalpic folding barriers computed under isostability conditions†.

A small enthalpic desolvation barrier between constituent groups of a protein can act cooperatively to produce a high overall enthalpic barrier to folding

Figure 8 shows the outcome of an isostability Eyring–chevron analysis (cf. Figures 1 and 2) on two versions of our model with different ε_{db} values for the pairwise desolvation barrier height. For the constant T, varying ε simulations (model chevron plots), $T=T_0=0.82$ and ε ranges from 0.93 to 1.25; for the constant ε, varying T simulations (model Eyring plots), $\varepsilon=1.2$ and T ranges from 0.75 to 1.06. The dependence of the logarithmic folding rate on ε or T is then translated into a dependence on native stability $\Delta G_f/T$ (Figure 8(a) and (b)), the latter is determined by conformational sampling using histogram techniques as described.[29,34] Folding relaxation is expected to be essentially single-exponential for all the datapoints in Figure 8 because their first passage time (FPT) distributions satisfy the "(median FPT)/ln 2=MFPT" test.[75]

Two different ε_{db} values are considered in Figure 8. The resulting $\Delta\ln k_f$ *versus* $1/T$ relationships are shown in Figure 8(c), where the error bars are deduced from the standard deviation of the MFPTs calculated using randomly chosen sets of half of the trajectories simulated for each datapoint. Figure 8(c) shows that for each ε_{db} value, there is an approximate linear relationship between $\Delta\ln k_f$ and $1/T$ for native stability $\Delta G_f/T > -25$, with slopes approximately equal to 4.9 ($\varepsilon_{db}=0.09$) and 8.9 ($\varepsilon_{db}=0.18$). Thus, the height of the apparent overall intrinsic enthalpic folding barrier $\Delta H_i^{\ddagger}$ in our model is positively correlated with the height of the intrachain pairwise desolvation barriers ε_{db}. If we identify k_BT_0 in the model with the room temperature value of ≈ 0.6 kcal/mol, this says that our models are producing apparent $\Delta H_i^{\ddagger}$ values when the free energy of unfolding <15 kcal/mol, a regime sufficient to cover the stability range of many small natural proteins.

It is instructive to compare the slopes in Figure 8(c) with the pairwise desolvation barrier heights in the model using the approximate scaling relation:

$$(\Delta\ln k_f)_{\Delta G_f/T} \sim -\alpha\frac{\varepsilon_{db}}{T} \tag{5}$$

where α is an enhancement factor and the $\Delta G_f/T$ subscript here indicates that each $\Delta\ln k_f$ is determined under isostability conditions at a particular $\Delta G_f/T$. The $1/T$ dependence in Figure 8(c) spans a wide range of different $\Delta G_f/T$ values (cf. Figure 8(a) and (b)), i.e. $(\Delta\ln k_f)_{\Delta G_f/T}$ at different $1/T$ values are determined at different $\Delta G_f/T$ values. The data plotted in Figure 8(c) indicate that $\alpha=-(\text{slope})/\varepsilon_{db}\approx 50$ (4.9/0.09=54.4 and 8.9/0.18=49.4 for $\varepsilon_{db}=$ 0.09 and 0.18, respectively). This result suggests that the apparent intrinsic enthalpic folding barrier determined from an Eyring–chevron analysis of the present model can be 50 times higher than the height of the individual intrachain pairwise enthalpic desolvation barrier assumed by the model. Therefore, a relatively small pairwise desolvation barrier between a protein's constituent groups can lead to a greatly enhanced effective overall enthalpic folding barrier.

By construction, equation (5) accounts only for the situation when the dependence on $1/T$ involves a range of different $\Delta G_f/T$ values and ε_{db} is kept constant. To better delineate the underlying physics, we consider also another approximate scaling relationship:

$$\ln k_f|_{\varepsilon/T=\text{constant}} \sim -\beta\frac{\varepsilon_{db}}{T} \tag{6}$$

for $\ln k_f$ itself (instead of $\Delta\ln k_f$), and in which ε_{db} is varied. Now $\varepsilon/T=$constant implies that $\Delta G_f/T$ is essentially constant; $\beta=-(\partial[T\ln k_f]/\partial\varepsilon_{db})_{\varepsilon/T}$ is the corresponding enthalpic barrier enhancement factor under this condition. To determine β, extensive simulations at $T=0.82$ have been performed for different ε and ε_{db} values. We found that $\ln k_f$ varies essentially linearly with ε_{db}, but the enhancement factor β changes considerably with ε/T and thus $\Delta G_f/k_BT$. Specifically, for $\Delta G_f/k_BT\approx -6, -9.5, -13, -20, -29$, and -45, $\beta\approx 10.3, 9.0, 7.7, 5.5, 3.8$, and 1.4, respectively (detailed data not shown).

These β factors suggest that at $\Delta G_f/k_BT\approx -6$, for example, the folding rate-limiting step involves surmounting approximately ten individual pairwise desolvation barriers at the same time ($\beta\approx 10.3$). Thus, even though the native–centric contact-like interactions (*U*s) in the model are pairwise additive, the energetic constraints of protein core formation and water expulsion nonetheless imparts a degree of cooperativity to the folding process (Figure 9). A possible physical reason is as follows. In our model, when contact energies are not highly favorable by themselves (small ε/T, less negative $\Delta G_f/k_BT$), formation of only one contact would not be particularly favorable overall, as it tends to constrain the sequential neighbors of the contacting residues into adopting unfavorable desolvation-barrier configurations. Conversely, the crossing of desolvation barriers tends to become less cooperative when individual contact energies

† $\Delta G_f/T$ is a function only of ε/T to a very good approximation. Hence, constant $\Delta G_f/T$ is essentially equivalent to constant ε/T conditions. It follows that if a single $\Delta H_i^{\ddagger}$ links every pair of conditions with the same $\Delta G_f/T$ (and therefore the same ε/T) when ε is only a function of denaturant concentration ($\varepsilon=\varepsilon([d])$) and T-independent (as in the present model), and for all $\varepsilon([d]=0)>0$, the same $\Delta H_i^{\ddagger}$ would apply to a generalized model with $\varepsilon([d])\to\varepsilon(T,[d])$. The main difference is that if a particular point defined by $\varepsilon([d])$ on a chevron plot at $T=T_0$ is isostable with $T=T_1$,$[d]=0$ in the present model, i.e. if $\varepsilon([d])/T_0=\varepsilon([d]=0)/T_1$, the same point on the chevron plot would be isostable instead with $T=T_2$,$[d]=0$ in the generalized model, where $\varepsilon(T_2,[d]=0)/T_2=\varepsilon([d]=0)/T_1$.

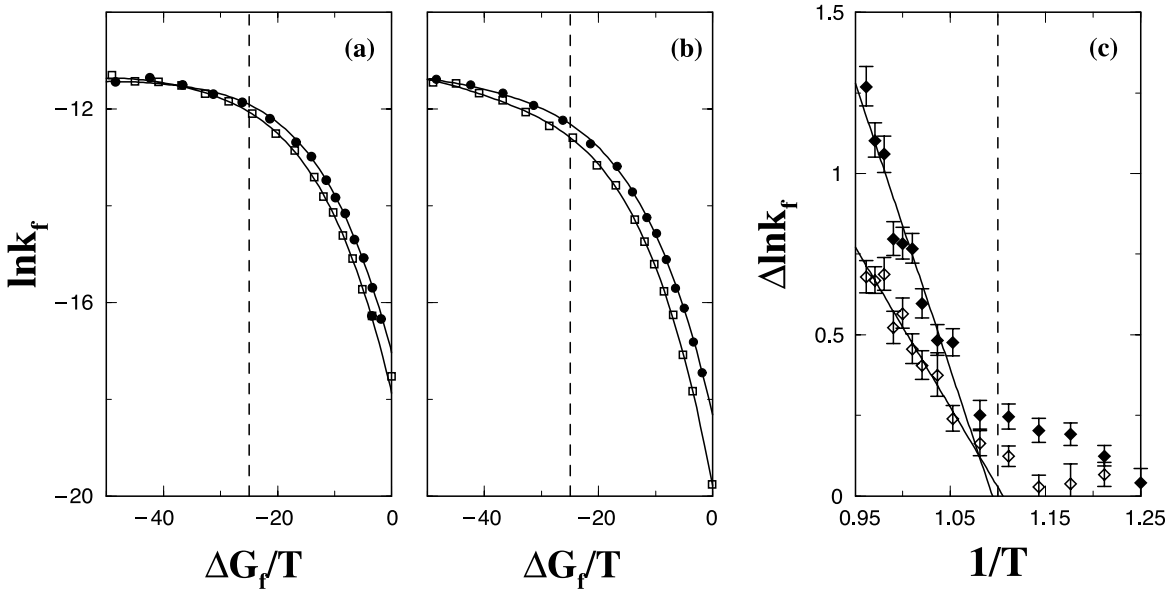

Figure 8. Intrinsic enthalpic folding barriers in the simulations of native-centric models of chymotrypsin inhibitor 2 using the present $U(r)$ function (cf. Figure 7) and the native contact set NCS2 by Kaya & Chan.[29] Each folding trajectory is initiated from a random conformation, and the first passage time for folding is the number of Langevin simulation time-steps needed to first achieve all native contacts (when fractional native contact $Q=1$). For most of the datapoints shown, approximately 400 trajectories are used to determine each mean first passage time (MFPT) of folding. Around the transition midpoint, ≈300 trajectories are used. The folding rate k_f is taken to be $(\text{MFPT})^{-1}$. (a) The constant ε, variable T logarithmic rates (filled circles) are compared with the constant T, variable ε logarithmic rates (open squares) for the model with $\varepsilon_{db}=0.09$. The continuous curves are fits to the simulated datapoints. The vertical broken line marks a stability level of $-25k_BT$ ($k_B=1$ in the model units), which translates into a free energy of unfolding ≈15 kcal/mol at room temperature. (b) Same as (a) but for $\varepsilon_{db}=0.18$. (c) Each datapoint is the logarithmic difference, at a given $\Delta G_f/T$, between the constant ε, variable T and the constant T, variable ε folding rates (the datapoints for the former minus the corresponding value on the fitted curve for the latter in (a) and (b)). Results for the $\varepsilon_{db}=0.09$ and 0.18 cases are shown respectively by open and filled diamonds. The error bars are estimated from standard deviations of the mean. A linear relation (continuous straight line) is fitted to each of the two sets of datapoints using only the $\Delta\ln k_f$ values for $\Delta G_f/T > -25$. The vertical broken line has the same meaning as that in (a) and (b).

become more favorable (larger ε/T), because the more favorable contact energies can overwhelm the unfavorable desolvation barriers, resulting in a reduced β for more negative values of $\Delta G_f/k_BT$ (see above).

This variation of β indicates that even though an Eyring-chevron analysis of our simulations leads to relatively high apparent $\Delta H_i^{\ddagger}$ values (Figure 8(c)), the present model may not strictly satisfy equation (1) because these barrier heights are probably $\Delta G_f/T$-dependent rather than essentially $\Delta G_f/T$-independent. It follows that the high slopes of the $\Delta\ln k_f$ *versus* $1/T$ plots in Figure 8(c) and the correspondingly high α values (equation (5)) achieved by our model may not reflect a universal cooperative desolvation barrier crossing process applicable to the entire observed range of native stabilities. Instead, these features are likely a combined effect of the tendency for moderately cooperative desolvation barrier crossing at relatively low native stability (e.g. $\beta\approx10$ at $\Delta G_f/k_BT\approx-6$, at small $1/T$) as well as the decrease of this cooperative tendency (decrease in β) as native stability is increased ($\Delta G_f/k_BT$ more negative at large $1/T$). By the same token, the very small β values at highly negative values of $\Delta G_f/k_BT$ account also for the significant decrease in the magnitude of the slope of $\Delta\ln k_f$ *versus* $1/T$ for $\Delta G_f/k_BT<-25$ ($1/T>1.1$ in Figure 8(c)).

As emphasized above, the present model mimics proteins that fold with chevron rollovers, but is insufficiently cooperative for apparent two-state proteins because chevron plots of similarly formulated models lack an extended quasi-linear regime.[29,34,36] Owing to this limitation, although our model offers the possibility that the enthalpic folding barrier determined by an Eyring–chevron analysis can be different from that determined using multiple datapoints at a particular $\Delta G_f/k_BT$, it remains to be determined whether this phenomenon applies to any real protein. Thus far, data from global analyses of several apparent two-state proteins[31,33] indicate that the enthalpic folding barrier height under isostability conditions is only minimally sensitive to native stability, if at all. Nonetheless, in view of the present model results, the possibility of a significant dependence of isostability enthalpic folding barrier height on $\Delta G_f/k_BT$ (i.e. $\Delta H_i^{\ddagger}\rightarrow\Delta H_i^{\ddagger}(\Delta G_f/T)$) should be explored experimentally, especially for less cooperative folders with chevron rollovers. As well, it would be interesting to investigate how our model can be brought more in line with real, apparent

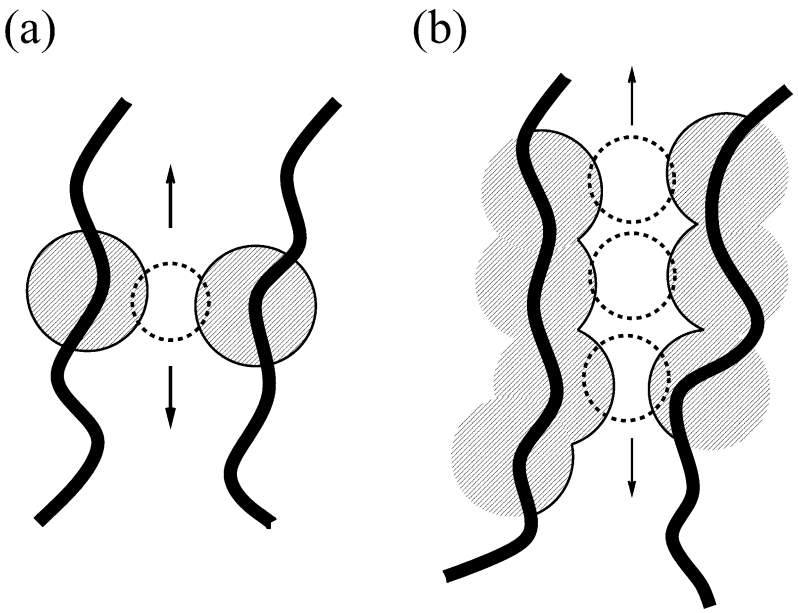

Figure 9. Schematics of desolvation scenarios that lead to enthalpic folding barriers. The dotted circles denote water molecules, the thick black curves represent parts of the protein backbone, and the shaded areas depict the van der Waals excluded volumes of the protein's constituent chemical groups. (a) A pairwise desolvation process (cf. Figure 6). The average number of water molecules that need to be expelled for contact formation is relatively small (arrows indicate water expulsion). (b) A cooperative desolvation process. In comparison with pairwise contact formation, in this case, more water molecules have to be expelled more or less simultaneously for multiple contact formation.

two-state proteins by more cooperative interaction schemes[17,36] including using higher pairwise desolvation barriers[34,51] (see below).

Using a mean field theory, Lum *et al.*[76] predicted that the association of hydrophobic units of sufficiently large sizes can proceed without encountering a desolvation free energy barrier. The experimental prevalence of high $\Delta H_i^{\ddagger}$ and the model results in this work, however, suggest that Lum *et al.*'s proposed kinetic process probably does not apply to the rate-limiting step in protein folding. Although a high enthalpic barrier can be consistent with a lack of free energy barrier (i.e. when there is sufficient entropic compensation; and in this regard it would be interesting to ascertain the enthalpic and entropic components of the free energy in the Lum *et al.* theory), it is unlikely that contact formation in the rate-limiting step in folding are devoid of desolvation free energy barriers, because such hypothetical processes would likely lead to folding rates much faster than that of real proteins (cf. without-solvation *versus* with-solvation models as in Kaya & Chan[29]).

Enthalpy–entropy compensation: a relatively low intrachain interaction free energy barrier can be consistent with a high enthalpic barrier at desolvation

Having demonstrated the physical principle that a significant overall enthalpic barrier to folding can emerge from relatively low intrachain pairwise desolvation barriers, we now turn to more quantitative issues regarding the magnitude of the overall enthalpic barrier. In Figure 8(c), the higher $\Delta \ln k_f$ *versus* $1/T$ slope for the $\varepsilon_{db} = 0.18$ case is approximately 8.9. Since this model quantity is in units of $k_B T$, it translates into an apparent intrinsic enthalpic folding barrier ≈ 5 kcal/mol. This means that although the present model does lead to a significant apparent $\Delta H_i^{\ddagger}$, the barrier height produced by our model is only about one-fifth of the average intrinsic enthalpic folding barrier height of ≈ 25 kcal/mol for the real proteins we have considered (Tables 1 and 2, and Figure 4). Can the physical perspective supplied by our desolvation model be augmented to afford higher overall enthalpic folding barriers that are in better quantitative agreement with experiments?

To answer this question, first, we note that in large measure the present modeling exercise serves only as a proof of principle. The relatively low ε_{db} values chosen here are for efficient simulation and conceptual advance, rather than for optimal quantitative agreement with experiments. It is clear from Figure 8 that if a larger ε_{db} was used, the height of the overall enthalpic folding barrier would increase. In that case, folding would be slower (and thus more difficult to simulate[34]). But such a slowdown in folding rate may actually make the model more realistic. In fact, the folding rates of "with-solvation" models with pairwise desolvation barrier heights similar to the present ε_{db} values are typically orders of magnitude faster than the folding rates of the corresponding real proteins.[29] It follows that one possible physical reason why the experimental $\Delta H_i^{\ddagger}$ are higher than those produced by the model here is that the effective intrachain pairwise enthalpic desolvation barriers for real proteins are higher than the ε_{db} values used in this work.

Second, the present native–centric contact-like interactions are pairwise additive (the $\sum_{i<j-3}^{\text{native}} U$ term in equation (4)). However, recent studies have shown that many-body effects beyond pairwise additivity are needed to better reproduce real protein properties in coarse-grained models.[16–19] Now, the above discussion of equation (6) suggests that for a given ε_{db}, an increase in interaction

cooperativity would likely lead to a higher $\Delta H_i^{\ddagger}$ (Figure 9). Also of interest are recent simulations of small non-polar solutes in explicit water indicating that under ambient conditions the unfavorable free energy at the desolvation barrier is probably anti-cooperative: the desolvation free energy barrier for three-methane association is apparently higher than the corresponding pairwise sum.[40,60] Although it remains to be ascertained whether the anti-cooperativity of the free energy at the desolvation barrier applies as well to the corresponding enthalpic component, taken together, these considerations suggest that $\Delta H_i^{\ddagger}$ is likely to be enhanced relative to the present model prediction when interaction non-additivity is taken into account.

Third, and perhaps the most important point is the possible interplay of solvent-related enthalpic and entropic components of the effective intrachain interactions. For simplicity, the present model considers only purely enthalpic intrachain pairwise desolvation barriers. In this setup, an increase in ε_{db} will invariably decrease the folding rate. In this case, then, if the pairwise enthalpic desolvation barriers were too high, folding would become exceedingly slow. However, the full picture can be more complex. Explicit-water simulations show that desolvation barriers have both enthalpic and entropic components. (The present model is limited in this respect, and should be augmented by incorporating such subtleties in future analyses.) Of direct relevance to the rates of contact formation is the unfavorable free energy at the desolvation barrier, not its enthalpic component. Whereas a very high desolvation free energy barrier would greatly impede folding, a very high enthalpic barrier to desolvation does not necessarily have the same effect. The reason is simple: if the high enthalpic barrier is largely compensated by a concomitant increase in entropy, the resulting free energy barrier can remain relatively low†. Indeed, this very behavior appears to apply in certain hydrophobic interactions: a recent simulation of three methane molecules in water shows that while the desolvation free energy barrier $\approx$0.6 kcal/mol at 25 °C, the nearby peak enthalpy $\approx$1.8 kcal/mol (cf. Figures 3 and 8 of Moghaddam *et al.*[40]). In other words, the enthalpic barrier to desolvation in this case is three times as high as the free energy barrier to desolvation, suggesting that in general the enthalpic signature of an elementary desolvation process can be much higher than the free energy barrier itself. This explicit-water simulation result thus offers a rationalization for high enthalpic folding barrier to originate from high enthalpic components of elementary desolvation barrier. At the same time, it allows the elementary desolvation free energy barriers themselves to remain sufficiently low so that folding would not be prohibitively slow.

† The favorable entropy here refers to solvent entropy, which is part of the effective intrachain interaction free energy represented by the vertical axis of an energy landscape. The entropy in this discussion does not involve conformational entropy, which is represented by the area of the high-dimensional surface, i.e. the "horizontal" coordinates, of the energy landscape.

Conclusions

Using experimental Eyring and chevron plots of folding kinetics data in the literature, we have shown for more than a dozen proteins that their temperature-dependent rate differences under isostability conditions exhibit a regularity consistent with the existence of high intrinsic enthalpic folding barriers. We emphasize, however, that although the proposed Eyring–chevron analysis is useful in lieu of global-analysis data, only the broad experimental coverage provided by global analyses[31–33,77] can verify or falsify conclusively the proposed intrinsic enthalpic folding barrier paradigm, and determine whether, and if so, the extent to which the height of the enthalpic folding barrier under isostability conditions depends on native stability. Here, in two cases, we find that intrinsic enthalpic barriers of similar heights are operative as well in kinetic processes of bimolecular association/folding, suggesting that a generic physical mechanism may be at play in both single-chain folding and binding involving two molecules.

To explore possible physical origins of high intrinsic enthalpic folding barriers, we devised a continuum explicit-chain construct that incorporates small enthalpic desolvation barriers.[29,51] Despite its limitations, this model has demonstrated that small intrachain pairwise desolvation barriers can act cooperatively to produce a significant effective enthalpic folding barrier for the overall folding process. It will be interesting to explore how this effect may also act in concert with other possible sources of protein folding cooperativity such as side-chain packing.[78] Our model results also suggest the possibility that if overall folding is not highly cooperative, enthalpic folding barriers under isostability conditions may depend on native stability (i.e. $\Delta H_i^{\ddagger} \rightarrow \Delta H_i^{\ddagger}(\Delta G_f/T)$ in equation (1)). It remains to be investigated whether this model feature captures similar properties in real proteins or it merely reflects the limited cooperativity of our model. This question deserves to be explored by further experimental global analyses, especially for natural and artificially designed[69] proteins with significant folding-arm chevron rollovers.

Explicit-water simulations conducted elsewhere showing enthalpy–entropy compensation around the desolvation free energy barrier in hydrophobic association[40] suggest that a high positive peak value of enthalpy can be consistent with a desolvation free energy barrier of moderate height. This is in line with the well-recognized understanding that hydrophobic association under ambient conditions is strongly favored by an increase in water entropy. Hence, it is physically possible for a protein folding energy landscape, which provides the free energy

for every conformation with solvent degrees of freedom pre-averaged, to exhibit both a high empirical enthalpic folding kinetic barrier and an essentially funnel-like topography such that Levinthal's paradox can be overcome. In such a scenario, barriers (uphill folding paths) do exist on an overall funnel-like energy landscape.[5,36] Some of such barriers originating from desolvation may take the form of a ridging rim around the native state. However, these landscape barriers in solvent-mediated intrachain interaction free energy can be relatively low yet still be consistent with a large positive enthalpic signature for the rate-limiting event in protein folding.

Acknowledgements

H.S.C. thanks David Baker and Dan Raleigh for enlightening dialogues at the time when part of the experimental data and ideas that led to this work first emerged in 1996–1998. The authors also thank Yawen Bai for recent insightful exchanges about enthalpic barriers, Hüseyin Kaya for useful discussions, two anonymous referees and the editor for helpful comments, and the Canadian Institutes of Health Research (CIHR grant no. MOP-15323) and the Canada Research Chairs Program for financial support.

References

1. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet.* **21**, 167–195.
2. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding–a perspective from simple exact models. *Protein Sci.* **4**, 561–602.
3. Thirumalai, D. & Woodson, S. A. (1996). Kinetics of folding of proteins and RNA. *Accts. Chem. Res.* **29**, 433–439.
4. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.
5. Chan, H. S. & Dill, K. A. (1998). Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics. *Proteins: Struct. Funct. Genet.* **30**, 2–33.
6. Dobson, C. M., Šali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angewandte Chem. Int. Ed.* **37**, 868–893.
7. Onuchic, J. N., Nymeyer, H., García, A. E., Chahine, J. & Socci, N. D. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advan. Protein Chem.* **53**, 87–152.
8. Mirny, L. & Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 361–396.
9. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45.
10. Levinthal, C. (1969). How to fold graciously. In *Mössbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois* (Debrunner, P., Tsibris, J. C. M. & Münck, E., eds), pp. 22–24, University of Illinois Press, Urbana, IL.
11. Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl Acad. Sci. USA*, **89**, 20–22.
12. Matthews, C. R. (1987). Effect of point mutations on the folding of globular proteins. *Methods Enzymol.* **154**, 498–511.
13. Matthews, C. R. (1993). Pathways of protein folding. *Annu. Rev. Biochem.* **62**, 653–683.
14. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **1**, 149–156.
15. Bilsel, O. & Matthews, C. R. (2000). Barriers in protein folding reactions. *Advan. Protein Chem.* **53**, 153–207.
16. Jewett, A. I., Pande, V. S. & Plaxco, K. W. (2003). Cooperativity, smooth energy landscapes and the origins of topology-dependent protein folding rates. *J. Mol. Biol.* **326**, 247–253.
17. Kaya, H. & Chan, H. S. (2003). Contact order dependent protein folding rates: kinetic consequences of a cooperative interplay between favorable nonlocal interactions and local conformational preferences. *Proteins: Struct. Funct. Genet.* **52**, 524–533.
18. Chan, H. S., Shimizu, S. & Kaya, H. (2004). Cooperativity principles in protein folding. *Methods Enzymol.* **380**, 350–379.
19. Ejtehadi, M. R., Avall, S. P. & Plotkin, S. S. (2004). Three-body interactions improve the prediction of rate and mechanism in protein folding models. *Proc. Natl Acad. Sci. USA*, **101**, 15088–15093.
20. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
21. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
22. Baker, D. (2000). A surprising simplicity to protein folding. *Nature*, **405**, 39–42.
23. Chen, X. & Matthews, C. R. (1994). Thermodynamic properties of the transition state for the rate-limiting step in the folding of the α subunit of tryptophan synthase. *Biochemistry*, **33**, 6356–6362.
24. Segawa, S.-I. & Sugihara, M. (1984). Characterization of the transition state of lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. *Biopolymers*, **23**, 2473–2488.
25. Chen, B.-L., Baase, W. A. & Schellman, J. A. (1989). Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry*, **28**, 691–699.
26. Makarov, D. E., Keller, C. A., Plaxco, K. W. & Metiu, H. (2002). How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl Acad. Sci. USA*, **99**, 3535–3539.
27. Wallin, S. & Chan, H. S. (2005). A critical assessment of the topomer search model of protein folding using a continuum explicit-chain model with extensive conformational sampling. *Protein Sci*. In the press.
28. Kaya, H. & Chan, H. S. (2002). Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding? *J. Mol. Biol.* **315**, 899–909.
29. Kaya, H. & Chan, H. S. (2003). Solvation effects and

driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **326**, 911–931 (Corrigendum: **337**, 1069–1070 (2004)).
30. Oliveberg, M., Tan, Y.-J. & Fersht, A. R. (1995). Negative activation enthalpies in the kinetics of protein folding. *Proc. Natl Acad. Sci. USA*, **92**, 8926–8929.
31. Scalley, M. L. & Baker, D. (1997). Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc. Natl Acad. Sci. USA*, **94**, 10636–10640.
32. Schindler, T. & Schmid, F. X. (1996). Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry*, **35**, 16833–16842.
33. Kuhlman, B., Luisi, D. L., Evans, P. A. & Raleigh, D. P. (1998). Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* **284**, 1661–1670.
34. Kaya, H., Liu, Z. & Chan, H. S. (2005). Chevron behavior and isostable enthalpic barriers in protein folding: successes and limitations of simple Gō-like modeling. *Biophys. J.* In the press.
35. Chan, H. S. (1998). Modelling protein folding by Monte Carlo dynamics: chevron plots, chevron rollover, and non-Arrhenius kinetics. In *Monte Carlo Approach to Biopolymers and Protein Folding* (Grassberger, P., Barkema, G. T. & Nadler, W., eds), pp. 29–44, World Scientific, Singapore.
36. Kaya, H. & Chan, H. S. (2003). Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity. *Proteins: Struct. Funct. Genet.* **52**, 510–523.
37. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Advan. Protein Chem.* **39**, 191–234.
38. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.
39. Shimizu, S. & Chan, H. S. (2000). Temperature dependence of hydrophobic interactions: a mean force perspective, effects of water density, and nonadditivity of thermodynamic signatures. *J. Chem. Phys.* **113**, 4683–4700 (Erratum: **116**, 8636 (2002)).
40. Moghaddam, M. S., Shimizu, S. & Chan, H. S. (2005). Temperature dependence of three-body hydrophobic interactions: potential of mean force, enthalpy, entropy, heat capacity, and nonadditivity. *J. Am. Chem. Soc.* **127**, 303–316 (Correction: **127**, 2363 (2005)).
41. Levitt, M. & Warshel, A. (1975). Computer simulation of protein folding. *Nature*, **253**, 694–698.
42. Thirumalai, D. & Klimov, D. K. (1999). Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **9**, 197–207.
43. Head-Gordon, T. & Brown, S. (2003). Minimalist models for protein folding and design. *Curr. Opin. Struct. Biol.* **13**, 160–167.
44. Micheletti, C., Banavar, J. R., Maritan, A. & Seno, F. (1999). Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Letters*, **82**, 3372–3375.
45. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). 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.* **298**, 937–953.
46. Koga, N. & Takada, S. (2001). Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model. *J. Mol. Biol.* **313**, 171–180.
47. Chavez, L. L., Onuchic, J. N. & Clementi, C. (2004). Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates. *J. Am. Chem. Soc.* **126**, 8426–8432.
48. Jacobs, D. J., Rader, A. J., Kuhn, L. A. & Thorpe, M. F. (2001). Protein flexibility prediction using graph theory. *Proteins: Struct. Funct. Genet.* **44**, 150–165.
49. Keskin, O., Bahar, I., Flatow, D., Covell, D. G. & Jernigan, R. L. (2002). Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry*, **41**, 491–501.
50. Pratt, L. R. & Chandler, D. (1977). Theory of the hydrophobic effect. *J. Chem. Phys.* **67**, 3683–3704.
51. Cheung, M. S., García, A. E. & Onuchic, J. N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA*, **99**, 685–690.
52. Karanicolas, J. & Brooks, C. L. (2002). The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **11**, 2351–2361.
53. Fernandez-Escamilla, A. M., Cheung, M. S., Vega, M. C., Wilmanns, M., Onuchic, J. N. & Serrano, L. (2004). Solvation in protein folding analysis: combination of theoretical and experimental approaches. *Proc. Natl Acad. Sci. USA*, **101**, 2834–2839.
54. Sessions, R. B., Thomas, G. L. & Parker, M. J. (2004). Water as a conformational editor in protein folding. *J. Mol. Biol.* **343**, 1125–1133.
55. Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z. & Wolynes, P. G. (2004). Water in protein structure prediction. *Proc. Natl Acad. Sci. USA*, **101**, 3352–3357.
56. Rank, J. A. & Baker, D. (1997). A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Sci.* **6**, 347–354.
57. Hummer, G., Garde, S., García, A. E., Paulaitis, M. E. & Pratt, L. R. (1998). The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl Acad. Sci. USA*, **95**, 1552–1555.
58. Shimizu, S. & Chan, H. S. (2002). Origins of protein denatured states compactness and hydrophobic clustering in aqueous urea: Inferences from nonpolar potentials of mean force. *Proteins: Struct. Funct. Genet.* **49**, 560–566.
59. Shimizu, S. & Chan, H. S. (2001). Configuration-dependent heat capacity of pairwise hydrophobic interactions. *J. Am. Chem. Soc.* **123**, 2083–2084.
60. Shimizu, S. & Chan, H. S. (2002). Anti-cooperativity and cooperativity in hydrophobic interactions: three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Proteins: Struct. Funct. Genet.* **48**, 15–30 (Erratum: **49**, 294 (2002)).
61. Main, E. R. G., Fulton, K. F. & Jackson, S. E. (1999). Folding pathway of FKBP12 and characterisation of the transition state. *J. Mol. Biol.* **291**, 429–444.
62. Veitshans, T., Klimov, D. & Thirumalai, D. (1997).

Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* **2**, 1–22.
63. Bai, Y., Sosnick, T., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
64. Wooll, J. O., Wrabl, J. O. & Hilser, V. J. (2000). Ensemble modulation as an origin of denaturant-independent hydrogen exchange in proteins. *J. Mol. Biol.* **301**, 247–256.
65. Feng, H., Vu, N.-D. & Bai, Y. (2004). Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b_{562}: native fold with non-native hydrophobic interactions. *J. Mol. Biol.* **343**, 1477–1485.
66. Ollerenshaw, J. E., Kaya, H., Chan, H. S. & Kay, L. E. (2004). Sparsely populated folding intermediates of the Fyn SH3 domain: matching native-centric essential dynamics and experiment. *Proc. Natl Acad. Sci. USA*, **101**, 14748–14753.
67. Kaya, H. & Chan, H. S. (2005). Explicit-chain model of native-state hydrogen exchange: implications for event ordering and cooperativity in protein folding. *Proteins: Struct. Funct. Bioinform.* **58**, 31–44.
68. Kaya, H. & Chan, H. S. (2003). Origins of chevron rollovers in non-two-state protein folding kinetics. *Phys. Rev. Letters*, **90**, 258104.
69. Scalley-Kim, M. & Baker, D. (2004). 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.* **338**, 573–583.
70. Raschke, T. M., Tsai, J. & Levitt, M. (2001). Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water. *Proc. Natl Acad. Sci. USA*, **98**, 5965–5969.
71. Yang, H. B. & Elcock, A. H. (2003). Association lifetimes of hydrophobic amino acid pairs measured directly from molecular dynamics simulations. *J. Am. Chem. Soc.* **125**, 13968–13969.
72. Ghosh, T., Kalra, A. & Garde, S. (2005). On the salt-induced stabilization of pair and many-body hydrophobic interactions. *J. Phys. Chem. B*, **109**, 642–651.
73. Hillson, N., Onuchic, J. N. & García, A. E. (1999). Pressure-induced protein-folding/unfolding kinetics. *Proc. Natl Acad. Sci. USA*, **96**, 14848–14853.
74. Wallqvist, A., Covell, D. G. & Thirumalai, D. (1998). Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *J. Am. Chem. Soc.* **120**, 427–428.
75. Gutin, A., Sali, A., Abkevich, V., Karplus, M. & Shakhnovich, E. I. (1998). Temperature dependence of the folding rate in a simple protein model: search for a "glass" transition. *J. Chem. Phys.* **108**, 6466–6483.
76. Lum, K., Chandler, D. & Weeks, J. D. (1999). Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, **103**, 4570–4577.
77. Dimitriadis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G. & Smith, D. A. (2004). Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. *Proc. Natl Acad. Sci. USA*, **101**, 3809–3814.
78. Clementi, C., García, A. E. & Onuchic, J. N. (2003). Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom study of protein L. *J. Mol. Biol.* **326**, 933–954.
79. Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **293**, 1195–1210.
80. Chen, B.-L., Baase, W. A., Nicholson, H. & Schellman, J. A. (1992). Folding kinetics of T4 lysozyme and nine mutants at 12 °C. Low-temperature unfolding of a mutant of phage T4 lysozyme. 2. Kinetic investigations. *Biochemistry*, **31**, 1464–1476.
81. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of β-hairpin formation in protein G folding. *Nature Struct. Biol.* **7**, 669–673.
82. Kamagata, K., Arai, M. & Kuwajima, K. (2004). Unification of the folding mechanisms of non-two-state and two-state proteins. *J. Mol. Biol.* **339**, 951–965.
83. Plaxco, K. W., Guijarro, J. I., Morton, C. J., Pitkeathly, M., Campbell, I. D. & Dobson, C. M. (1998). The folding kinetics and thermodynamics of the Fyn-SH3 domain. *Biochemistry*, **37**, 2529–2537.
84. Chen, B.-L. & Schellman, J. A. (1989). Low-temperature unfolding of a mutant of phage T4 lysozyme. 1. Equilibrium studies. *Biochemistry*, **28**, 685–691.
85. Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M. *et al.* (1999). Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme. *Biochemistry*, **38**, 2135–2142.
86. Manyusa, S. & Whitford, D. (1999). Defining folding and unfolding reactions of apocytochrome b_5 using equilibrium and kinetic fluorescence measurements. *Biochemistry*, **38**, 9533–9540.
87. Pascher, T. (2001). Temperature and driving force dependence of the folding rate of reduced horse heart cytochrome c. *Biochemistry*, **40**, 5812–5820.
88. Alexander, P., Orban, J. & Bryan, P. (1992). Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry*, **31**, 7243–7248.
89. Matagne, A., Jamin, M., Chung, E. W., Robinson, C. V., Radford, S. E. & Dobson, C. M. (2000). Thermal unfolding of an intermediate is associated with non-Arrhenius kinetics in the folding of hen lysozyme. *J. Mol. Biol.* **297**, 193–210.
90. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **270**, 294–304.
91. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
92. Van Nuland, N. A. J., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. T. & Dobson, C. M. (1998). Slow cooperative folding of a small globular protein HPr. *Biochemistry*, **37**, 622–637.
93. Goldberg, J. M. & Baldwin, R. L. (1998). Kinetic mechanism of a partial folding reaction. 1. Properties of the reaction and effects of denaturants. *Biochemistry*, **37**, 2546–2555.
94. Goldberg, J. M. & Baldwin, R. L. (1998). Kinetic mechanism of a partial folding reaction. 2. Nature of the transition state. *Biochemistry*, **37**, 2556–2563.
95. Ibarra-Molero, B., Makhatadze, G. I. & Matthews, C. R. (2001). Mapping the energy surface for the folding reaction of the coiled-coil peptide. *Biochemistry*, **40**, 719–731.
96. Oliveberg, M. & Fersht, A. R. (1996). Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry*, **35**, 2738–2749.

97. Dalby, P. A., Oliveberg, M. & Fersht, A. R. (1998). Folding intermediates of wild-type and mutants of barnase. I. Use of ϕ-value analysis and *m*-values to probe the cooperative nature of the folding pre-equilibrium. *J. Mol. Biol.* **276**, 625–646.
98. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W. & Oas, T. G. (1996). Microsecond protein folding through a compact transition state. *J. Mol. Biol.* **263**, 311–322.
99. Myers, J. K. & Oas, T. G. (1999). Contribution of a buried hydrogen bond to λ repressor folding kinetics. *Biochemistry*, **38**, 6761–6768.
100. Choe, S. E., Matsudaira, P. T., Osterhout, J., Wagner, G. & Shakhnovich, E. I. (1998). Folding kinetics of villin 14T, a protein domain with a central β-sheet and two hydrophobic cores. *Biochemistry*, **37**, 14508–14518.
101. Reid, K. L., Rodriguez, H. M., Hillier, B. J. & Gregoret, L. M. (1998). Stability and folding properties of a model β-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7**, 470–479.

Edited by C. R. Matthews

(Received 18 January 2005; received in revised form 30 March 2005; accepted 31 March 2005)