INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY

Phys. Biol. **2** (2005) S75–S85 doi:10.1088/1478-3975/2/4/S01

Solvation and desolvation effects in protein folding: native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions

Zhirong Liu and Hue Sun Chan[1]

Department of Biochemistry, and Department of Medical Genetics & Microbiology, Faculty of Medicine, University of Toronto, Toronto, Ontario M5S 1A8, Canada

E-mail: chan@arrhenius.med.toronto.edu

Received 28 June 2005
Accepted for publication 27 July 2005
Published 9 November 2005
Online at stacks.iop.org/PhysBio/2/S75

Abstract

As different parts of a protein chain approach one another during folding, they are expected to encounter desolvation barriers before optimal packing is achieved. This impediment originates from the water molecule's finite size, which entails a net energetic cost for water exclusion when the formation of compensating close intraprotein contacts is not yet complete. Based on recent advances, we extend our exploration of these microscopic elementary desolvation barriers' roles in the emergence of generic properties of protein folding. Using continuum Gō-like C_α chain models of chymotrypsin inhibitor 2 (CI2) and barnase as examples, we underscore that elementary desolvation barriers between a protein's constituent groups can significantly reduce native conformational fluctuations relative to model predictions that neglected these barriers. An increasing height of elementary desolvation barriers leads to thermodynamically more cooperative folding/unfolding transitions (i.e., higher overall empirical folding barriers) and higher degrees of kinetic cooperativity as manifested by more linear rate-stability relationships under constant temperature. Applying a spatially non-uniform thermodynamic parametrization we recently introduced for the pairwise C_α potentials of mean force, the present barnase model further illustrates that desolvation is a probable physical underpinning for the experimentally observed high intrinsic enthalpic folding barrier under isostability conditions.

1. Introduction

Many proteins of biological interest fold in aqueous environments. Major driving forces for protein folding—hydrophobicity, electrostatic interactions and hydrogen bonding—are solvent-mediated. An elucidation of the role of water is key to any fundamental understanding of the folding process. As far as theory and computer simulation are concerned, it has long been recognized that a more realistic account of the finite size and hydrogen-bonding geometry of water molecules is of critical importance. When 'explicit-water' simulations of a globular protein immersed in a large number of atomistic model water molecules became computationally viable, they were found to produce more realistic vibrational amplitudes and hydrogen-bonding patterns [1] than that predicted previously by simulations *in vacuo* [2] (see [3] for a review). Nonetheless, since explicit-water simulations of protein dynamics are computationally intensive, implicit-water or 'implicit-solvent' models that are computationally less costly have also been developed to account for solvation effects. These models do not consider individual water molecules. Instead, various approximate

[1] Address for correspondence: Department of Biochemistry, University of Toronto, Medical Sciences Building, 5th Floor, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

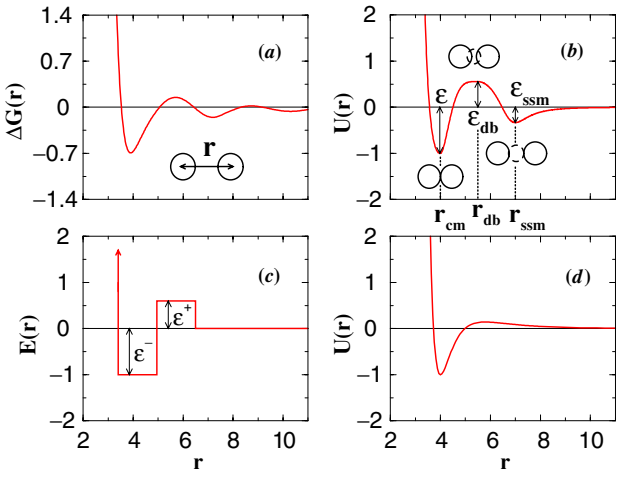

Figure 1. Explicit-water simulations and pairwise implicit-solvent model interactions. (*a*) The potential of mean force $\Delta G(r)$ (in units of kcal mol^{-1}) between a pair of methane-like Lennard–Jones spheres separated by distance r (in Å), simulated by Shimizu and Chan [18] using the TIP4P model of water at temperature 298.15 K under 1 atm. The desolvation barrier for this case is at $r \approx 5.7$ Å. (*b*) A pairwise implicit-solvent model interaction potential with desolvation barrier, using the original prescription of Cheung *et al* [21] and the subsequent 'with-solvation' model formulation of Kaya and Chan [24]. Here, ϵ is the depth of the contact minimum (cm), whereas $\epsilon_{\rm db}$ and $\epsilon_{\rm ssm}$ are, respectively, the desolvation barrier (db) height and solvent-separated minimum (ssm) depth. We have used $\epsilon_{\rm db} = 5\epsilon/9$ and $\epsilon_{\rm ssm} = \epsilon/3$ for the example shown. The diagrams illustrate typical cm, db and ssm configurations [19], where water molecules are depicted by dashed circles and amino acid residues are represented by solid circles (drawings adapted from Liu and Chan [31]). (*c*) An earlier square-well model pairwise interaction potential with a square-shoulder desolvation barrier, adapted from Hillson *et al* [20]. (*d*) A model pairwise interaction potential introduced by Karanicolas and Brooks [23] with desolvation penalty (their equation (1) is applied using $\epsilon_{ij} = 1$ and $\sigma = 4$ for the present depiction).

continuum treatments for the aqueous solvent are used [4–7]. Results obtained from implicit-solvent models and from explicit-solvent simulations do not always agree, however. There are explicit-water-induced effects that are notably not captured by certain implicit-solvent approaches, such as a possible mechanism for concurrent core collapse and desolvation during protein folding [8, 9] as highlighted by a recent comprehensive simulation study [10]. In other words, how solvation is modeled has an significant impact on the topography of the protein folding energy landscape [11–15].

A salient effect of desolvation may be illustrated by considering a pair of small nonpolar solutes in water (figure 1(*a*)). Before a close contact between the solutes can be formed, finite-sized water molecules have to be first excluded from the space between the two solutes. This necessitates a decrease in favorable water-solute interactions as the solutes approach each other, a loss that is to be compensated by solute–solute interaction only when a closer association between the solutes is achieved. Hence, there is an energetic barrier to contact formation (i.e., a barrier to desolvation), as reflected in the solute pair's potential of mean force [16–19]. Physically, it is intuitive to expect that a similar transient energetic cost to apply as well, if not more so, to the process of squeezing water out from the protein interior as the core of a globular protein is formed [8, 9, 17]. Ramifications of this basic energetic consideration for protein folding have been explored computationally in the context of implicit-solvent approaches by incorporating elementary desolvation barriers in the pairwise effective interaction potentials between different parts of a protein [20–26]. An advantage of these models is their computational tractability. They are valuable for conceptual development because many critical questions in protein thermodynamics and kinetics [27] cannot yet be addressed by models with full atomic details. It is particularly instructive to compare these models and other implicit-solvent models that do not embody elementary desolvation barriers, such as the volume-exclusion formulation for solvation in the original version of the effective force field EEF1 [5] and common approaches based on solvent accessible surface area (SASA) [18, 28]. For instance, desolvation barriers tend to endow protein models with a higher degree of thermodynamic cooperativity (as manifested by a sharper heat capacity peak at the folding/unfolding transition midpoint) than their no-desolvation-barrier counterparts [21, 24, 29], and thus provide a better mimicry for the behavior of many small, single-domain proteins [27].

More recently, the height of the pairwise desolvation barriers in these models have also been suggested to correlate significantly with kinetic cooperativity, as characterized by an extended quasi-linear regime in the chevron plot of constant-temperature folding and unfolding rates versus denaturant concentration [29]. This possibility underscores the importance of desolvation in the physics of folding because kinetic cooperativity was not an achievable feat to date for lattice [30] or coarse-grained C_α Gō models with essentially pairwise additive interactions devoid of desolvation barriers [24]. Furthermore, desolvation barriers provide a promising physical framework to rationalize the high intrinsic enthalpic signatures of 15–40 kcal mol^{-1} [31] observed experimentally for the empirical barriers to folding under isostability conditions [32, 33]. Remarkably, these high intrinsic enthalpic folding barriers exist not only for single-domain 'two-state' proteins [34, 35], but apply to certain 'non-two-state' proteins with chevron rollovers as well [29, 31, 36]. The apparent generality of this phenomenon is not inconsistent with our perspective that high intrinsic enthalpic folding barriers are likely a consequence of certain generic desolvation mechanisms. Building on these advances, this paper explores further the impact of solvation/desolvation effects on protein conformational distribution and fluctuation, particularly their role in folding cooperativity. We assess the robustness of some of our findings by comparing new simulation results from a coarse-grained model of barnase with our previous results for CI2. Certain useful technical details not covered in prior publications will also be provided.

2. Methods

2.1. Native-centric Gō-like constructs

The present study employs a coarse-grained simplified chain representation [37–39] and a native-centric approach. Despite their basic limitations [24, 40], native-centric Gō-like models have provided valuable insights—currently not accessible by other approaches—into the thermodynamics and kinetics of protein folding [41–51] as well as native-state dynamics [52, 53]. As in our recent works [24, 29, 31], we consider the following potential [44, 45] with a C_α chain representation and favorable pairwise native-centric interactions:

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

where N is the total number of amino acid residues in the given protein, r, θ and ϕ are, respectively, the virtual bond length, bond angle and torsion angles defined by C_α positions; r_0, θ_0 and ϕ_0 are the corresponding native values in the Protein Data Bank (PDB) structure. The spatial distance r_{ij} is that between two C_α positions i and j separated sequentially by at least three residues. If residues i and j are in contact in the native structure, r_{ij}^{n} is the corresponding native C_α–C_α distance in the PDB structure (second last summation). Otherwise, excluded-volume repulsion for non-native contacts is parametrized by r_{rep} (last summation). Here, the pairwise native-centric interaction $U\left(r_{ij}; r_{ij}^{\text{n}}, \epsilon, \epsilon_{\text{db}}, \epsilon_{\text{ssm}}\right)$ is a function of the strength ϵ of the most favorable interaction at contact (depth of the contact minimum), the desolvation barrier height ϵ_{db} and the depth ϵ_{ssm} of the solvent-separated minimum (see figure 1(*b*) and detailed description below). To illustrate the diversity of modeling possibilities in this general approach, two other examples of implicit-solvent potentials with desolvation barriers are depicted in figures 1(*c*) and (*d*).

2.2. Desolvation potentials

In this paper, we use only the potential function in figure 1(*b*). This functional form is essentially identical to that introduced by Cheung *et al* [21]. Setting the native contact distance r_{ij}^{n} equal to the contact-minimum (cm) separation of this potential and dropping the ij subscript in r_{ij} for notational simplicity (i.e., $r_{ij} \to r, r_{ij}^{\text{n}} \to r^{\text{n}} = r_{\text{cm}}$), the native-centric pairwise interaction U in the above equation for V_{total} is given by

$$\begin{aligned} &U(r; r_{\text{cm}}, \epsilon, \epsilon_{\text{db}}, \epsilon_{\text{ssm}}) \\ &= \begin{cases} \epsilon Z(r)[Z(r)-2] & \text{for} \quad r < r_{\text{cm}} \\ CY(r)^n[Y(r)^n/2 - (r_{\text{db}}-r_{\text{cm}})^{2n}]/2n + \epsilon_{\text{db}} & \\ \quad \text{for} \quad r_{\text{cm}} \leqslant r < r_{\text{db}} & \\ -B[Y(r)-h_1]/[Y(r)^m + h_2] & \text{for} \quad r \geqslant r_{\text{db}} \end{cases} \end{aligned} \tag{2}$$

where

$$\begin{aligned} Z(r) &= (r_{\text{cm}}/r)^k \\ Y(r) &= (r-r_{\text{db}})^2 \\ C &= 4n(\epsilon+\epsilon_{\text{db}})/(r_{\text{db}}-r_{\text{cm}})^{4n} \\ B &= m\epsilon_{\text{ssm}}(r_{\text{ssm}}-r_{\text{db}})^{2(m-1)} \\ h_1 &= (1-1/m)(r_{\text{ssm}}-r_{\text{db}})^2/(\epsilon_{\text{ssm}}/\epsilon_{\text{db}}+1) \\ h_2 &= (m-1)(r_{\text{ssm}}-r_{\text{db}})^{2m}/(1+\epsilon_{\text{db}}/\epsilon_{\text{ssm}}). \end{aligned} \tag{3}$$

Here, $r_{\text{ssm}} = r_{\text{cm}} + 3$ Å, following from the consideration that $r_{\text{ssm}} - r_{\text{cm}} = 3$ Å which corresponds approximately to the diameter of a water molecule, and $r_{\text{db}} = (r_{\text{ssm}} + r_{\text{cm}})/2$, as in the original [21]. We use $k = 6, m = 3$ and $n = 2$ as in our previous studies [24, 29, 31]. We note that the repulsive interaction $Z(r)[Z(r)-2]$ above for $r < r_{\text{cm}}$ coincides with the repulsive part of a 6-12 Lennard–Jones potential. If a (10-12)-type repulsive potential is preferred, the expression $Z(r)[Z(r)-2]$ may be replaced by $5(r_{\text{cm}}/r)^{12} - 6(r_{\text{cm}}/r)^{10}$. However, the effect is negligible for such a minor variation in the strongly repulsive regime of the potential.

The interaction strength ϵ here has the same meaning as that in [21], whereas the present variables $\epsilon_{\text{db}}, \epsilon_{\text{ssm}}, r_{\text{cm}}, r_{\text{db}}$ and r_{ssm} correspond, respectively, to the variables $\epsilon'', \epsilon', r', r^\dagger$ and r'' in Cheung *et al*. Two apparent typographical errors in this original [21] should be corrected. First, in the expression for C, the exponent in the denominator should be $4n$ (as in the above equation), not $2n$ as given by Cheung *et al* [21]. Second, the expression $(\epsilon''+\epsilon')/(\epsilon'-\epsilon)$ in the caption for figure 1 in [21] should read $(\epsilon''+\epsilon')/(\epsilon-\epsilon')$ [29].

2.3. Langevin dynamics

The dynamics of our model system is simulated using a set of Langevin equations [54],

$$m\dot{v}(t) = F_{\text{conf}}(t) - m\gamma v(t) + \eta(t), \tag{4}$$

where $m, v, \dot{v}, F_{\text{conf}}, \gamma, \eta$ and t are, respectively, mass, velocity, acceleration, conformational force, friction (viscosity) constant, random force and time; F_{conf} is the negative gradient of V_{total} with respect to the coordinate for the given spatial direction. The random force satisfies the autocorrelation function

$$\langle\eta(t)\eta(t')\rangle = 2m\gamma k_{\text{B}}T\delta(t-t'), \tag{5}$$

where $k_{\text{B}}T$ is Boltzmann constant times absolute temperature. As described in detail elsewhere [24], for simplicity, units are chosen such that $m = 1$ and $k_{\text{B}} = 1$, as in [54]. Simulation times are reported in units of Langevin time steps. We also use the same Langevin dynamics for thermodynamic conformational sampling, in conjunction with standard histogram techniques and bias potentials as appropriate.

3. Results and discussion

3.1. Solvation/desolvation barriers reduce native flexibility

We apply the above model formulation to the 64-residue truncated form of chymotrypsin inhibitor 2 (CI2; residues

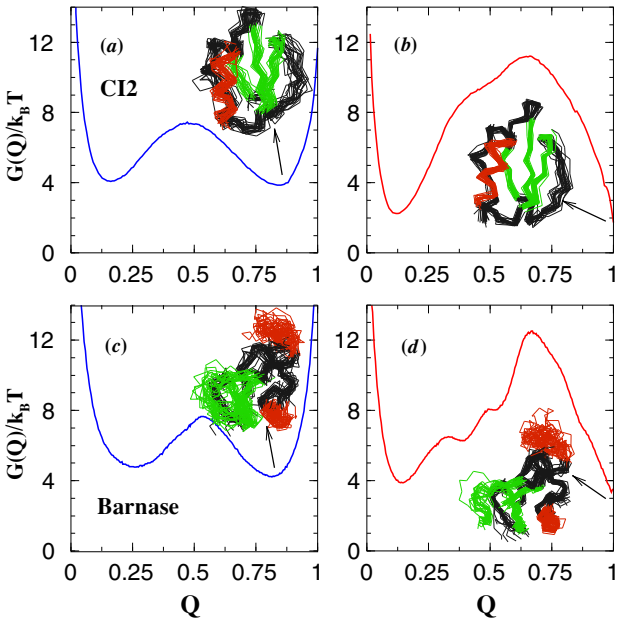

Figure 2. Free-energy profiles for CI2 (upper panels) and barnase (lower panels) in the without-solvation (left panels) and with-solvation (right panels) models. The progress variable Q here is the fractional number of native contacts, and $G(Q)/k_BT = -\ln P(Q)$, where $P(Q)$ is the normalized population distribution as a function of Q, and the definition of Q in the simulation is identical to that in our group's previous work [24]. For CI2, we use the native contact set NCS2 [24]. For barnase, the native contact set is determined by the criterion of Chavez *et al* [50]: two residues are considered to be in contact if at least a pair of their heavy atoms are less than 4.5 Å apart in the Protein Data Bank (PDB) structure of the protein. Bias potentials and histogram techniques are used to obtain the free-energy profiles from conformational sampling at $T = 0.82$ at or near each system's transition midpoint ($\epsilon = 0.806$, 0.916, 0.787 and 0.893 for (*a*), (*b*), (*c*) and (*d*), respectively). The with-solvation model results shown here are computed using $\epsilon_{db} = 0.1\epsilon$. In this regard, (*b*) is a corrected [29] and more well-sampled version of figure 4(*c*) in Kaya and Chan [24]. For each of the four models, a superposition of 20 randomly chosen conformations from the native free-energy minimum is shown (helical, β-sheet and other regions are in red, green and black, respectively). The root-mean-square deviation (rmsd) of these conformations from the corresponding PDB native structures are, respectively, (*a*) 2.02 Å, (*b*) 0.85 Å, (*c*) 2.84 Å and (*d*) 2.22 Å.

20–83 of 2CI2) and a slightly truncated 108-residue form of barnase (residues 3–110 of 1BNI:A). To better delineate the role of the pairwise desolvation barriers, we compare the simulated free-energy profiles from these 'with-solvation' models with those generated from the corresponding 'without-solvation' models (figure 2). The 'without-solvation' formulation here is based on a native-centric 12-10 interaction that does not include a desolvation barrier, as described before [24]. Figure 2 shows that, even by incorporating a relative low pairwise desolvation barrier ($\epsilon_{db} = 0.1\epsilon$), the with-solvation models achieve considerably higher overall free-energy barriers (figures 2(*b*) and (*d*)) than their without-solvation counterparts (figures 2(*a*) and (*c*)). The increase in Q-based folding barrier height is $\approx 6k_BT$ for both the CI2 and barnase models. Interestingly, the desolvation barriers also shift the overall free-energy barrier peaks toward the native state, from $Q \approx 0.5$ to $Q \approx 0.7$ for these cases. In this connection, it is noteworthy that in contrast to the earlier simulation of Clementi *et al* using a different native contact set (figure 5 of [44]), the present without-solvation free-energy profile for barnase (figure 2(*c*)) exhibits only a single peak and no folding intermediate on the Q-based profile[2]. This suggests that caution should be used when interpreting detailed predictions from Gō-like models, as they can be sensitive to subtleties in model construction [24].

A conspicuous feature exhibited by the comparison in figure 2 is that the native free-energy minima are at $Q \approx 1$ for the with-solvation models (figures 2(*b*) and (*d*)), but are less similar to the PDB structures, at $Q \approx 0.85$, for the corresponding without-solvation models (figures 2(*a*) and (*c*)). The conformational drawings in figure 2 and the root-mean-square deviations from the PDB structures described in the caption of this figure further underscore this fundamental difference. Clearly, the presence of a desolvation barrier significantly shortens the spatial range of the attractive basin associated with the contact minimum, rendering the native conformations less free to fluctuate, and therefore less flexible, than that for the hypothetical situation in which the desolvation barrier is non-existent. Inasmuch as modeling is concerned, merely reducing the spatial range of the native-centric attractive interactions without introducing a desolvation barrier can achieve a similar effect, though to a lesser degree (e.g., in the short-spatial-range 'SSR' models [24]). Physically, however, the existence of solvation/desolvation barriers appear to be the most natural candidate for the origin of restricted spatial ranges for certain favorable intraprotein interactions.

3.2. Desolvation barriers enhance thermodynamic cooperativity

To further explore the effect of elementary desolvation barriers on the overall thermodynamic cooperativity, figure 3 compares the temperature dependence of heat capacity of three models with different pairwise desolvation barrier heights. We first evaluate these models using the calorimetric criterion for two-state cooperativity, that the van't Hoff to calorimetric enthalpy ratio $\Delta H_{vH}/\Delta H_{cal}$ should be approximately unity [55]. As has been recently elucidated [27], with a proper choice of calorimetric baselines, this criterion implies a bimodal distribution in enthalpy with well-separated double peaks at the folding/unfolding transition midpoint (see [56, 57] and references therein), although this feature does not imply a complete absence of conformations with intermediate enthalpies [58–62]. The calorimetric two-state criterion has been demonstrated to be a rather stringent modeling constraint, because not all commonly studied protein chain models satisfy this seemingly mundane requirement [27, 63, 64]. Here, we apply the enthalpy ratio definition

$$\Delta H_{vH}/\Delta H_{cal} = \kappa_2^{(s)} \equiv \frac{2T_{max}\sqrt{k_B C_{P,max}^{(s)}}}{\Delta H_{cal}^{(s)}}, \qquad (6)$$

[2] Significant chevron rollovers are possible in the absence of Q-based intermediates [24].

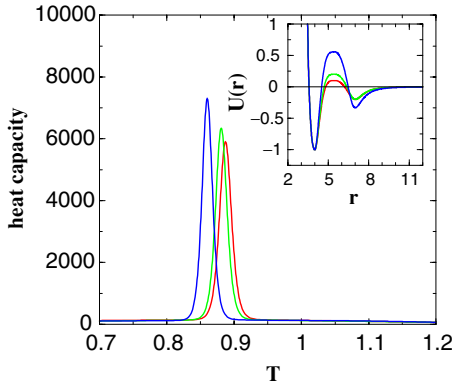

Figure 3. Heat capacity as a function of temperature for with-solvation CI2 models with (from right to left) $\epsilon_{\rm db} = 0.1\epsilon$, $\epsilon_{\rm ssm} = 0.2\epsilon$ (red); $\epsilon_{\rm db} = 0.2\epsilon$, $\epsilon_{\rm ssm} = 0.2\epsilon$ (green) and $\epsilon_{\rm db} = 5\epsilon/9$, $\epsilon_{\rm ssm} = \epsilon/3$ (blue). The corresponding pairwise interaction potentials $U(r)$ of the native-centric C_α models are shown in the inset. Note that the last prescription (blue $U(r)$ curve, inset) for the pairwise interaction is identical to that of Cheung *et al* [21], and the present red heat capacity curve is an enhanced and corrected [29] version of curve (v) in figure 5 of Kaya and Chan [24]. The heat capacity functions here are computed from densities of states for $V_{\rm total}$ using bias potentials and standard histogram techniques.

where $C^{(\rm s)}_{P,\rm max}$ and $\Delta H^{(\rm s)}_{\rm cal}$ are the peak heat capacity value and calorimetric enthalpy, respectively, after empirical baseline subtractions and $T_{\rm max}$ is the temperature at the heat capacity peak [57]. This analysis yielded $\kappa^{(\rm s)}_s \approx 1.00$ for all three models, indicating that the models are calorimetrically cooperative. This result is not surprising, as even the corresponding Gō-like models without desolvation barriers are in compliance with this calorimetric criterion [24].

Although all three models in figure 3 satisfy the calorimetric two-state criterion, the sharpness of their transition is obviously correlated with the height of the pairwise desolvation barrier. As the desolvation barrier height increases from $\epsilon/10$ to $\epsilon/5$ then $5\epsilon/9$, the peak heat capacity value increases, and the model becomes slightly less stable (transition midpoint shifts slightly to lower temperatures). Concomitantly, the temperature width of the transition region narrows, with the full width at half maximum decreasing from $\Delta T \approx 0.021$ to 0.020 then 0.018. Although a sharp transition does not necessarily imply that the transition is calorimetrically cooperative [57, 65], given that the underlying transitions are calorimetrically cooperative as for the models in figure 3, sharpness may be used to further quantitate their different degrees of cooperativity. The cooperativity parameter Ω_c of Klimov and Thirumalai [66–68], which is essentially a sharpness measure, is useful in this regard. In the notation of [67],

$$\Omega_c = \frac{T_F^2}{\Delta T}\left|\frac{{\rm d}f_N}{{\rm d}T}\right|_{T=T_F} \tag{7}$$

where f_N is a measure of fractional native population, ΔT is the full width at half maximum of $|{\rm d}f_N/{\rm d}T|$ and T_F is the folding transition temperature identified with the maximum in ${\rm d}f_N/{\rm d}T$. If f_N is now taken to be the fractional difference in enthalpy between the fully denatured and the native state, i.e., $f_N = \Delta H/\Delta H_{\rm cal}$, it follows that ${\rm d}f_N/{\rm d}T = \Delta C_P/\Delta H_{\rm cal}$, and a Klimov–Thirumalai cooperativity parameter

$$\Omega_c = \frac{(T_{\rm max})^2}{\Delta T}\frac{C^{(\rm s)}_{P,\rm max}}{\Delta H^{(\rm s)}_{\rm cal}} \tag{8}$$

may then be defined by the heat capacity related quantities in equation (6). Using this expression, Ω_c is computed to be ≈1620, 1750 and 2030, respectively, for $\epsilon_{\rm db} = \epsilon/10$, $\epsilon/5$ and $5\epsilon/9$. Thus, when measured in terms of transition sharpness, the thermodynamic cooperativity of these calorimetrically two-state models do increase with elementary pairwise desolvation barrier height.

3.3. Kinetic cooperativity correlates with desolvation barrier height

Next we use the same three models in figure 3 to study the relationship between desolvation barrier height and kinetic cooperativity. The upper panels of figure 4 show the models' Q-based free-energy profiles. Expectedly, and consistent with figure 3, the overall free-energy barrier at $Q \approx 0.65$ for these CI2 models becomes higher with increasing height of the underlying elementary pairwise desolvation barrier. The lower panels of figure 4 show the corresponding model chevron plots. As we have recently suggested [29], the more extensive kinetic simulation data displayed here demonstrate clearly that the quasi-linear regime of the folding chevron arm covers an increasing range of native stability as the pairwise desolvation barrier height $\epsilon_{\rm db}$ of the model is increased. Since the models' thermodynamic cooperativity by a sharpness measure also increases with $\epsilon_{\rm db}$ (figure 3), this result from figure 4 adds further substance to the proposal that folding/unfolding kinetic cooperativity is correlated with thermodynamic cooperativity [69].

Kinetic folding/unfolding cooperativity is a non-trivial modeling constraint, as many common protein chain models fall short of predicting this phenomenon [27]. Kinetic cooperativity appears also to have been an evolved property for certain proteins, as it is not straightforward to design polypeptides with the high degrees of kinetic cooperativity that are routinely observed among many real proteins [70].

Figure 4 shows that for $\epsilon_{\rm db} = \epsilon/10$, $\epsilon/5$ and $5\epsilon/9$, respectively, the quasi-linear chevron regime extends to native stabilities of $\Delta G_{\rm f} \approx -6$, -8 and $-11 k_{\rm B}T$. It is noteworthy here that for the $\epsilon_{\rm db} = 5\epsilon/9$ case, the native stability range of the quasi-linear chevron regime is comparable to the zero-denaturant native stability of $\approx 10 k_{\rm B}T$ for many small, single-domain proteins at room temperature. As well, the good match between the logarithmic mean first passage times (mean FPTs, or MFPTs) and the corresponding values of logarithmic (median FPT)/ln 2 in the lower panels of figure 4 is consistent with an essentially single-exponential kinetic relaxation [71], as is often determined experimentally for linear chevrons. Taken together, these model results offer a physical perspective to rationalize the experimentally observed linear

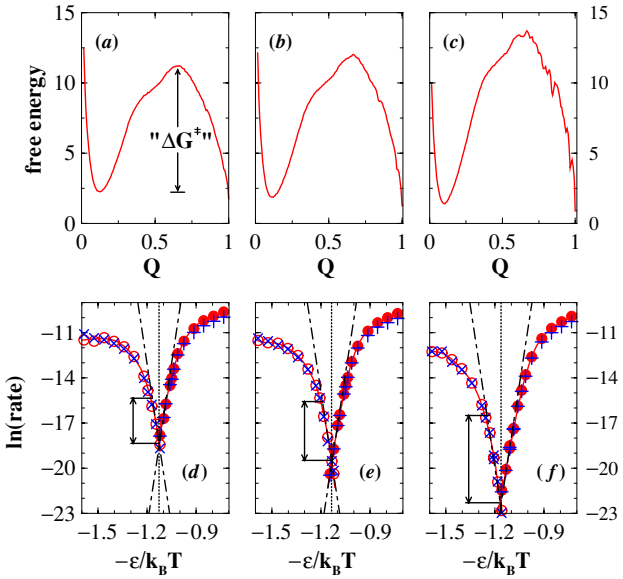

Figure 4. Folding/unfolding kinetics of with-solvation CI2 models computed using different pairwise elementary desolvation barrier heights. Upper panels: free-energy profiles ($-\ln P(Q)$ as in figure 2) simulated at $T = 0.82$ for (*a*) $\epsilon_{\rm db} = 0.1\epsilon$, $\epsilon_{\rm ssm} = 0.2\epsilon$, (*b*) $\epsilon_{\rm db} = 0.2\epsilon$, $\epsilon_{\rm ssm} = 0.2\epsilon$ and (*c*) $\epsilon_{\rm db} = 5\epsilon/9$, $\epsilon_{\rm ssm} = \epsilon/3$ at their approximate transition midpoints: $\epsilon = 0.916$, 0.920 and 0.942 for (*a*), (*b*) and (*c*), respectively. The quantity '$\Delta G^{\ddagger}$' defines the height of the overall free-energy barrier (in units of $k_{\rm B}T$) along these Q-based profiles between the denatured (low Q) and native ($Q \approx 1$) minima. Lower panels: model chevron plots simulated at $T = 0.82$ for multiple values of ϵ. The with-solvation CI2 models simulated in (*d*), (*e*) and (*f*) correspond, respectively, to that in (*a*), (*b*) and (*c*). Each folding (open circles) or unfolding (filled circles) datapoint is the negative logarithm of mean first passage time (MFPT), defined as in our previous study of these models (figure 12 of Kaya *et al* [29]) and averaged from about 400 trajectories, except 20–100 trajectories per datapoint are used for MFPTs around the transition midpoints. The dashed V-shapes are hypothetical chevron plots constructed to match the equilibrium free-energy difference between the folded and unfolded states as determined by standard histogram techniques. The vertical dotted lines mark the midpoint $-\epsilon/k_{\rm B}T$ values defined by the heat capacity peaks in figure 3. The cross (×) and plus (+) signs denote the negative logarithm of the median folding/unfolding first passage time (FPT) divided by ln 2. The double arrows indicate the approximate quasi-linear regimes of the folding chevron arms. Part of the results in (*d*) and (*e*) are adapted from figure 12 of Kaya *et al* [29].

chevron behavior of many real, small, single-domain proteins [69].

Figure 5 shows that the folding rates at the transition midpoints of this model correlate very well with an overall free-energy barrier '$\Delta G^{\ddagger}$', a quantity we have empirically and rather simplistically defined (hence the quotation marks) as the difference between the $-\ln P(Q)$ peak at intermediate Q and the $-\ln P(Q)$ value of the denatured minimum at low Q in the upper panels of figure 4 (see illustration in figure 4(*a*)). The fitted line in figure 5 has a slope of -1.16, which is very close to the ideal value of -1.0 when the folding rates of these models are described by a transition state theory expression

$$k_{\rm f} = \mathcal{F} \exp\left[-\frac{\Delta G^{\ddagger}}{k_{\rm B}T}\right] \qquad (9)$$

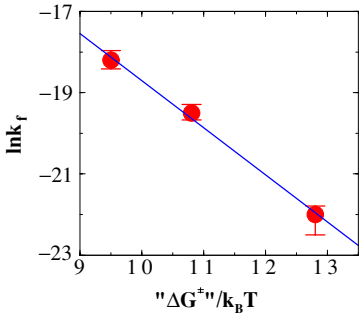

Figure 5. Correlation between the logarithmic folding rate ($\ln k_{\rm f}$) at transition midpoint and the Q-based folding free-energy barrier of the with-solvation CI2 models in figure 4. The solid line is a least-square fit to the three datapoints. The error bars are estimated from the spread of MFPTs computed from ∼1000 randomly chosen sets of half of the trajectories that contribute to each of the datapoints.

with an identical pre-exponential 'front' factor $\mathcal{F}$ for all three models and '$\Delta G^{\ddagger}$' $= \Delta G^{\ddagger}$. Our previous investigations have indicated that front factors defined empirically in a similar manner are in general not constants, but rather sensitive to native stability [24, 29, 72]. In this context, it is interesting that when native stability is now held fixed at zero (under transition midpoint conditions), $\mathcal{F}$ exhibits only a minimal decrease (fitted line with a slope of -1.16 instead of -1.0), if any, with increasing elementary pairwise desolvation barrier height. In other words, within the range of $\epsilon_{\rm db}$ considered, there is little dependence of the effective front factor on $\epsilon_{\rm db}$. Moreover, the upper panels of figure 4 indicate that the Q-position of the transition-state-like peak along the Q-based free-energy profile is essentially constant over the range of $\epsilon_{\rm db}$ investigated. It is important to realize, however, that this set of simple regular behaviors is restricted only to a stability regime at or very close to the transition midpoint; thus figure 5 does not imply that equation (9) with a constant $\mathcal{F}$ is generally applicable to a broad range of stabilities in the present models.

3.4. Desolvation barriers rationalize intrinsic enthalpic folding barriers

Owing to the intuitive insights they provide, empirical folding barriers and transition states—akin to those in the theoretical constructs of figures 2 and 4—have long been a staple in the protein folding literature [73–75]. In these interpretative frameworks, barrier properties are of interest because they offer clues for deciphering folding energetics. Barrier properties are determined by rate measurements. In particular, the thermodynamic signatures—free energy, enthalpy, entropy and heat capacity—of folding barriers are deduced from the temperature dependence of folding rates. These rates are generally non-Arrhenius, i.e., the logarithmic folding rate varies nonlinearly with $1/T$. This observation is often cast in terms of heat capacities of the folding transition states [76–78]. However, when both temperature T and denaturant concentration $[d]$ are adjusted to maintain constant native stability (isostability condition, $\Delta G_{\rm f}/T =$ constant), the logarithmic folding rate was found to depend

linearly and rather strongly on $1/T$ along constant $\Delta G_f/T$ contours for several apparent two-state proteins [32, 33]. As we have pointed out, this pattern of experimental behavior is remarkable and deserves theoretical attention, especially the question as to why the intrinsic enthalpic barriers for isostability folding are so high [29, 31].

Prompted by these considerations, we have recently proposed the general relation

$$\ln k_f(T,[d]) = F(\Delta G_f/T) - \frac{\Delta H_i^{\ddagger}}{T} \tag{10}$$

as an approximate expression for the combined temperature and denaturant dependence of folding rate, where the function F is protein dependent. Based on available direct experimental measurements on several apparent two-state proteins [32, 33], here the intrinsic enthalpic folding barrier $\Delta H_i^{\ddagger}$ is taken, as a first approximation, to be a constant (independent of $\Delta G_f/T$) for a given protein, but may differ for different proteins [29, 31]. (Note that units are chosen in the above equation such that $k_B = 1$, as for the rest of this paper.) In general, however, the possibility that $\Delta H_i^{\ddagger}$ may depend on $\Delta G_f/T$ cannot be precluded (see below). Now, assuming $\Delta H_i^{\ddagger}$ is constant, it follows from equation (10) that, for a given protein, the difference under isostability conditions between logarithmic temperature-dependent rate $\ln k_f(\Delta G_f/T;T)$ at a constant denaturant concentration $[d]_0$ (Eyring plot) and logarithmic denaturant-dependent rate $\ln k_f(\Delta G_f/T;T_0)$ at a fixed temperature T_0 (chevron plot) is given by

$$\Delta \ln k_f(\Delta G_f/T;T) = \frac{\Delta H_i^{\ddagger}}{T_0} - \frac{\Delta H_i^{\ddagger}}{T}, \tag{11}$$

which is linear in $1/T$ with slope $= -\Delta H_i^{\ddagger}$. Equation (11) forms the basis of our proposed Eyring-chevron analysis, a method designed to assess the validity of equation (10) when extensive global-analysis data [32, 33, 79–82] on T and $[d]$ dependence are not available. We have verified equation (11) for about a dozen mostly two-state proteins, with fitted $\Delta H_i^{\ddagger}$ values among them varying quite widely between 15 and 40 kcal mol^{-1} [31]. A noteworthy outcome of this exercise is that the intrinsic enthalpic folding barrier paradigm appears to apply as well to several 'non-two-state' proteins (with chevron rollovers) including barnase [29].

To provide a plausible physical account of this behavior and to overcome unsatisfactory aspects in earlier lattice-model rationalizations [13, 36, 69], we have recently introduced a particular functional form for the native-centric pairwise potential $U\left(r_{ij};r_{ij}^{n},\epsilon,\epsilon_{db},\epsilon_{ssm}\right)$ (cf equation (1)) in which the desolvation barrier height ϵ_{db} is kept constant and does not scale with ϵ (see [31] for details). In contrast to the $\epsilon_{db} \propto \epsilon$ models that do not produce a nonzero $\Delta H_i^{\ddagger}$ [29], this formulation led to appreciable intrinsic enthalpic folding barriers when applied to with-solvation models of CI2 [31]. The physical basis of the particular postulated form of U in [31] remains to be elucidated. In any event, U should be considered as a 'renormalized' pairwise potential that implicitly involves many-body effects [83]. Nonetheless, as indicated by explicit-water simulations, it is not at all surprising that the functional form of a potential of mean force (free energy of association)

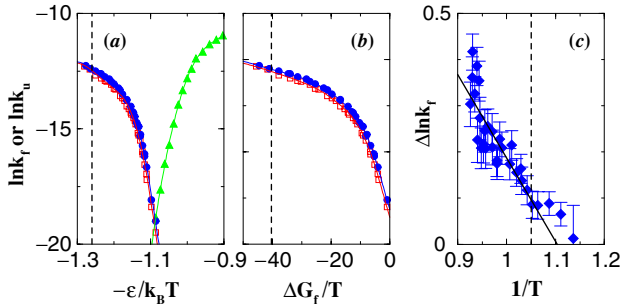

Figure 6. Intrinsic enthalpic folding barriers. The Eyring-chevron procedure of Liu and Chan for analyzing a with-solvation model for CI2 [31] is now applied to a similar model for barnase, with $\epsilon_{db} = 0.09$ being independent of both ϵ and T. (*a*) Logarithm of folding/unfolding rates at constant T and variable ϵ (as a model for denaturant dependence; red squares and green triangles) are compared with folding rates at constant ϵ and variable T (as a model for constant denaturant temperature dependence; blue circles). Each datapoint is averaged from at least 400 trajectories. (*b*) The T- and ϵ-dependent folding rates in (*a*) as functions of folding stability $\Delta G_f/T$. (*c*) Logarithmic difference $\Delta \ln k_f$ between the constant-ϵ, variable-T and the constant-T, variable-ϵ folding rates (former minus latter) that have the same $\Delta G_f/T$ as provided in (*b*). The inclined solid line in (*c*) is a least-squares fit to the datapoints left of the vertical dashed line, whereas the vertical dashed lines in (*a*) and (*c*) mark the interaction strength and temperature that correspond to a stability level of $\Delta G_f = -40k_BT$ as shown in (*b*). The error bars in (*c*) are estimated from statistical uncertainties of the folding rates as in figure 5.

can differ significantly from the spatial dependence of other thermodynamic signatures [19].

Here, we apply the constant-ϵ_{db} formulation in [31] to barnase. Similar to the previous CI2 case, the formulation produces an effective $\Delta H_i^{\ddagger} > 0$ for the Eyring-chevron comparison equation (11). Specifically, figure 6(*c*) shows a slope ≈ -1.8 for the $\Delta \ln k_f$ versus $1/T$ plot. This slope can be somewhat more negative if only low-stability datapoints (with smaller $1/T$) are considered. But in any case, the slope for barnase is considerably lower in magnitude than the ≈ -4.9 slope for the CI2 model that was based on the same U with an identical ϵ_{db}. The underlying reason for this substantial difference remains to be explored.

The slope estimated from figure 6(*c*) suggests $\Delta H_i^{\ddagger} \approx 1.8k_BT$ for the present barnase model. This is much lower than the experimental value of $\approx$30 kcal mol$^{-1} \approx 50k_BT$ (at $T = 298$ K) for this protein [29]. Nonetheless, relative to the individual desolvation barrier height $\epsilon_{db} = 0.09$ in this model, the slope ≈ -1.8 in figure 6(*c*) already represents a 20-fold magnification in impact on the overall Eyring-chevron behavior. In this sense, the present barnase model is successful in further demonstrating the principle that low elementary pairwise enthalpic desolvation barriers can act cooperatively to achieve a higher overall empirical enthalpic barrier to folding. As for a possible resolution of the significant numerical discrepancy between the experimentally deduced high $\Delta H_i^{\ddagger}$ and the low effective $\Delta H_i^{\ddagger}$ obtained from the present model, several factors should be taken into account as we have discussed before, including considerations of entropy–enthalpy compensation during the desolvation process and the

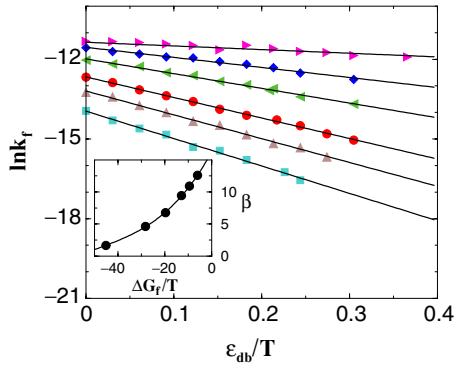

Figure 7. Interplay of the effects of desolvation barrier height (horizontal axis) and native stability on the logarithmic folding rate $\ln k_f$ of with-solvation CI2 models. All folding simulations here are conducted at $T = 0.82$ and $\epsilon_{ssm} = 0.2\epsilon$. Simulations at different ϵ values are represented by different symbols (joined by fitted lines). From top to bottom: $\epsilon = 1.2$, 1.1, 1.05, 1.01, 0.99 and 0.97. Each MFPT datapoint in this figure is averaged from at least 400 trajectories. The datapoints are well described by a linear relationship between $\ln k_f$ and ϵ_{db}/T (fitted lines), with a slope β that reflects the enhancement effect caused by cooperative desolvation processes. The inset shows the variation of β as a function of native stability. An abridged version of this analysis was discussed on p 882 of [31].

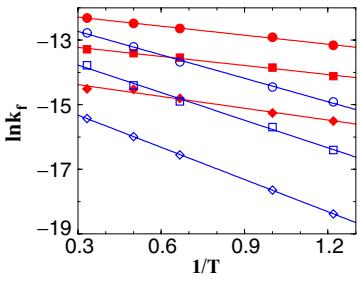

Figure 8. Effect of desolvation barrier height on the temperature dependence of folding rate under various isostability conditions. With-solvation CI2 models are considered. In each set of five datapoints (each set denoted by one symbol) in these folding simulations, ϵ/T is kept fixed while ϵ is varied with T to maintain a constant equilibrium native stability: $\Delta G_f/T = -15.0$ (circle), -10.0 (squares) or -5.0 (diamonds). The elementary pairwise desolvation barrier height $\epsilon_{db} = 0.09$ and 0.3, respectively, for the solid red and open blue symbols. Each datapoint here is averaged from about 400 trajectories except ∼100 trajectories are used for the slowest folding events plotted near the right-bottom corner of this figure.

adoption of a higher value of ϵ_{db} for the pairwise desolvation barrier in the model [31]. These issues require further analyses. It should also be noted that the Eyring-chevron analysis is formulated to compare only two folding rates for each ΔG_f value. As a result, this analysis alone cannot ascertain whether $\Delta H_i^{\ddagger}$ is essentially a constant for the entire native stability regime of interest, as assumed, or $\Delta H_i^{\ddagger}$ is sensitive to native stability, i.e., $\Delta H_i^{\ddagger} \to \Delta H_i^{\ddagger}(\Delta G_f/T)$ [31]. Aspects of this question are briefly addressed below.

3.5. Interplay of desolvation barrier height and native stability

Figure 7 considers the effects of pairwise desolvation barrier height on folding rate when native stability is held approximately constant at various levels. Folding rates on each fitted straight line are simulated with different ϵ_{db} but with the same ϵ/T. This means that $\Delta G_f/T$ is approximately constant along each fitted line because native stability depends only weakly on ϵ_{db} (cf figure 3). The linear fits in figure 7 indicates that the scaling relation

$$\ln k_f|_{\epsilon/T=\text{constant}} \sim -\beta\frac{\epsilon_{db}}{T} \tag{12}$$

holds to a very good approximation, with β corresponding to the magnitude of the negative slopes of the fitted lines. As we have discussed [31], β may be roughly interpreted as an enhancement factor accounting for the number of pairwise desolvation barriers that have to be simultaneously (cooperatively) overcome at the rate-limiting step of folding. This enhancement factor decreases with increasing level of native stability (figure 7, inset). Physically, this result is not surprising. In the presence of elementary desolvation barriers, a single isolated native contact tends not to be particularly favorable because such a configuration may most likely be 'frustrated' in that residues neighboring the contact pair would then be constrained into unfavorable relative positions coinciding with their desolvation barriers [31]. This is the basic reason for the enhanced degrees of cooperativity of the with-solvation models relative to the without-solvation models, because the above energetic consideration implies that neighboring native contacts would tend to form at the same time or not at all in the with-solvation formulations. However, this consideration clearly hinges on the strength of the favorable interaction at contact minimum *vis-à-vis* that of the unfavorable interaction at the desolvation barrier. Therefore, when native stability becomes higher (more negative $\Delta G_f/T$) as a result of an increasing magnitude of ϵ relative to that of ϵ_{db}, the cooperativity-enhancing effects of the desolvation barrier is expected to be mitigated. The results in figure 7 are consistent with this analysis.

Figure 8 addresses the degree to which the present versions of our model formulation are sufficient for fully mimicking the intrinsic enthalpic folding barrier phenomenon by producing an essentially constant $\Delta H_i^{\ddagger}$ over a native stability regime comparable to that of real proteins. Three ΔG_f values covering the typical native stability range of small proteins are considered, and two different values of ϵ_{db} are used to probe the effects of pairwise desolvation barrier height. Figure 8 shows that $\Delta H_i^{\ddagger}$ is not a constant for these models, as the slope of the logarithmic folding rate $\ln k_f$ under isostability condition versus $1/T$ depends on native stability, i.e., $\Delta H_i^{\ddagger} \to \Delta H_i^{\ddagger}(\Delta G_f/T)$. For $\epsilon_{db} = 0.09$, the slope for $\Delta G_f/T = -5.0$, -10.0 and -15.0 are, respectively, -1.22, -0.94 and -0.93. For $\epsilon_{db} = 0.3$, the corresponding slopes are -3.32, -2.85 and -2.42. So $\Delta H_i^{\ddagger}$ is sensitive to $\Delta G_f/T$ for both values of ϵ_{db}. One could argue that a higher pairwise desolvation barrier may lead to a weaker dependence of $\Delta H_i^{\ddagger}$ on $\Delta G_f/T$, at least for intermediate native stability, as the drop in slope from $\Delta G_f/T = -5.0$ to -10.0 is 14% for $\epsilon_{db} = 0.3$ whereas the corresponding percentage drop is

larger at 23% for $\epsilon_{db} = 0.09$. But more extensive data would be needed to establish a definitive trend.

Thus, in conjunction with the results in figure 7, the results in figure 8 indicate that although the present formulation of our model is sufficient to produce intrinsic enthalpic folding barriers (figure 6(*c*) above and [31]), it is not sufficiently cooperative to give rise to an essentially constant $\Delta H_i^{\ddagger}$ over an extended range of ΔG_f. Further investigations will be necessary to resolve this mismatch. Nonetheless, at the same time, it would be interesting to investigate whether the $\Delta H_i^{\ddagger} \rightarrow \Delta H_i^{\ddagger}(\Delta G_f/T)$ phenomenon has its experimental counterparts, especially for proteins with less cooperative folding/unfolding transitions.

4. Conclusion and outlook

Through model construction and analysis, we have demonstrated the critical roles of solvation/desolvation barriers in the high degrees of thermodynamic and kinetic cooperativity as well as the high intrinsic enthalpic folding barriers that have been experimentally observed for many small proteins. Thus, our investigation suggests that, for certain applications of implicit-solvent models to protein folding and dynamics, the effects of desolvation barriers are essential to incorporate. In addition to documenting the successes of our approach in capturing several generic protein folding properties, deficiencies in our models are also identified. In particular, further work would be required to quantitatively capture the remarkably high $\Delta H_i^{\ddagger}$ values of real proteins. Ultimately, it goes without saying that the crucial physical question needed to be addressed for any simplified chain model that appears to reproduce generic protein properties is how the postulated interactions in the model might have arisen atomistically.

Acknowledgments

We thank Doug Barrick (Johns Hopkins), Charlie Brooks (Scripps), Thomas Kiefhaber (Basel) and Sheena Radford (Leeds) for very helpful recent discussions, and the Canadian Institutes of Health Research (CIHR grant no MOP-15323) for financial support. HSC holds a Canada Research Chair in Proteomics, Bioinformatics and Functional Genomics.

References

[1] Levitt M and Sharon R 1988 Accurate simulation of protein dynamics in solution *Proc. Natl Acad. Sci. USA* **87** 7557–61

[2] McCammon J A, Gelin B R and Karplus M 1977 Dynamics of folded proteins *Nature* **267** 585–90

[3] Duan Y and Kollman K A 2001 Computational protein folding: from lattice to all-atom *IBM Syst. J.* **40** 297–309

[4] Roux B and Simonson T 1999 Implicit solvent models *Biophys. Chem.* **78** 1–20

[5] Lazaridis T and Karplus M 1999 Effective energy function for proteins in solution *Proteins: Struct. Funct. Genet.* **35** 133–52

[6] Feig M and Brooks C L 2004 Recent advances in the development and application of implicit solvent models in biomolecule simulations *Curr. Opin. Struct. Biol.* **42** 217–24

[7] Jaramillo A and Wodak S J 2005 Computational protein design is a challenge for implicit solvation models *Biophys. J.* **88** 156–71

[8] Sheinerman F B and Brooks C L 1998 Molecular picture of folding of a small α/β protein *Proc. Natl Acad. Sci. USA* **95** 1562–7

[9] Sheinerman F B and Brooks C L 1998 Calculations on folding of segment B1 of streptococcal protein G *J. Mol. Biol.* **278** 439–56

[10] Rhee Y M, Sorin E J, Jayachandran G, Lindahl E and Pande V S 2004 Simulations of the role of water in the protein-folding mechanism *Proc. Natl Acad. Sci. USA* **101** 6456–61

[11] Bryngelson J D, Onuchic J N, Socci N D and Wolynes P G 1995 Funnels, pathways, and the energy landscape of protein folding: a synthesis *Proteins: Struct. Funct. Genet.* **21** 167–95

[12] Dill K A and Chan H S 1997 From Levinthal to pathways to funnels *Nat. Struct. Biol.* **4** 10–19

[13] Chan H S and Dill K A 1998 Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics *Proteins: Struct. Funct. Genet.* **30** 2–33

[14] Wales D J 2003 *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge: Cambridge University Press)

[15] Komatsuzaki T, Hoshino K, Matsunaga Y, Rylance G J, Johnston R L and Wales D J 2005 How many dimensions are required to approximate the potential energy landscape of a model protein? *J. Chem. Phys.* **122** 084714

[16] Pratt L R and Chandler D 1977 Theory of the hydrophobic effect *J. Chem. Phys.* **67** 3683–704

[17] Rank J A and Baker D 1997 A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding *Protein Sci.* **6** 347–54

[18] Shimizu S and 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

Shimizu S and 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.* **49** 294 (erratum)

[19] Moghaddam M S, Shimizu S and 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** 303–16

Moghaddam M S, Shimizu S and 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** 2363 (erratum)

[20] Hillson N, Onuchic J N and García A E 1999 Pressure-induced protein-folding/unfolding kinetics *Proc. Natl Acad. Sci. USA* **96** 14848–53

[21] Cheung M S, García A E and 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** 685–90

[22] Guo C L, Cheung M S, Levine H and Kessler D A 2002 Mechanisms of cooperativity underlying sequence-independent β-sheet formation *J. Chem. Phys.* **116** 4353–65

[23] Karanicolas J and Brooks C L 2002 The origins of asymmetry in the folding transition states of protein L and protein G *Protein Sci.* **11** 2351–61

[24] Kaya H and 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** 911–31

Kaya H and Chan H S 2004 Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *J. Mol. Biol.* **337** 1069–1070 (Corrigendum)

[25] Fernandez-Escamilla A M, Cheung M S, Vega M C, Wilmanns M, Onuchic J N and Serrano L 2004 Solvation in protein folding analysis: combination of theoretical and experimental approaches *Proc. Natl Acad. Sci. USA* **101** 2834–9

[26] Sessions R B, Thomas G L and Parker M J 2004 Water as a conformational editor in protein folding *J. Mol. Biol.* **343** 1125–33

[27] Chan H S, Shimizu S and Kaya H 2004 Cooperativity principles in protein folding *Methods Enzymol.* **380** 350–79

[28] Lee B and Richard F M 1971 The interpretation of protein structures: estimation of static accessibility *J. Mol. Biol.* **55** 379–400

[29] Kaya H, Liu Z and Chan H S 2005 Chevron behavior and isostable enthalpic barriers in protein folding: successes and limitations of simple Gō-like modeling *Biophys. J.* **89** 520–35

[30] Kaya H and Chan H S 2003 Origins of chevron rollovers in non-two-state protein folding kinetics *Phys. Rev. Lett.* **90** 258104

[31] Liu Z and Chan H S 2005 Desolvation is a likely origin of robust enthalpic barriers to protein folding *J. Mol. Biol.* **349** 872–89

[32] Scalley M L and 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** 10636–40

[33] Kuhlman B, Luisi D L, Evans P A and 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** 1661–70

[34] Jackson S E 1998 How do small single-domain proteins fold? *Fold. Des.* **3** R81–91

[35] Baker D 2000 A surprising simplicity to protein folding *Nature* **405** 39–42

[36] Chan H S 1998 Modelling protein folding by Monte Carlo dynamics: chevron plots, chevron rollover, and non-Arrhenius kinetics *Monte Carlo Approach to Biopolymers and Protein Folding* ed P Grassberger, G T Barkema and W Nadler (Singapore: World Scientific) pp 29–44

[37] Levitt M and Warshel A 1975 Computer simulation of protein folding *Nature* **253** 694–8

[38] Thirumalai D and Klimov D K 1999 Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models *Curr. Opin. Struct. Biol.* **9** 197–207

[39] Head-Gordon T and Brown S 2003 Minimalist models for protein folding and design *Curr. Opin. Struct. Biol.* **13** 160–7

[40] Clementi C and Plotkin S S 2004 The effects of nonnative interactions on protein folding rates: theory and simulation *Protein Sci.* **13** 1750–66

[41] Micheletti C, Banavar J R, Maritan A and Seno F 1999 Protein structures and optimal folding from a geometrical variational principle *Phys. Rev. Lett.* **82** 3372–5

[42] Shea J-E, Onuchic J N and Brooks C L 1999 Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A *Proc. Natl Acad. Sci. USA* **96** 12512–7

[43] Zhou Y and Karplus M 1999 Interpreting the folding kinetics of helical proteins *Nature* **401** 400–3

[44] Clementi C, Nymeyer H and 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** 937–53

[45] Koga N and 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** 171–80

[46] Li L and Shakhnovich E I 2001 Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations *Proc. Natl Acad. Sci. USA* **98** 13014–8

[47] Cieplak M and Hoang T X 2003 Universality classes in folding times of proteins *Biophys. J.* **84** 475–88

[48] Clementi C, García A E and 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** 933–54

[49] Hu L H, Wang J and Wang W 2003 Folding simulations for a three-helix bundle protein *Int. J. Mod. Phys.* B **17** 49–54

[50] Chavez L L, Onuchic J N and Clementi C 2004 Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates *J. Am. Chem. Soc.* **126** 8426–32

[51] Zhang J, Meng Q and Wang W 2005 Multiple folding mechanisms of protein ubiquitin *Proteins: Struct. Funct. Bioinform.* **59** 565–79

[52] Jacobs D J, Rader A J, Kuhn L A and Thorpe M F 2001 Protein flexibility prediction using graph theory *Proteins: Struct. Funct. Genet.* **44** 150–65

[53] Keskin O, Bahar I, Flatow D, Covell D G and Jernigan R L 2002 Molecular mechanisms of chaperonin GroEL–GroES function *Biochemistry* **41** 491–501

[54] Veitshans T, Klimov D and Thirumalai D 1997 Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties *Fold. Des.* **2** 1–22

[55] Privalov P L and Khechinashvili N N 1974 A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study *J. Mol. Biol.* **86** 665–84

[56] Chan H S 2000 Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity *Proteins: Struct. Funct. Genet.* **40** 543–71

[57] Kaya H and Chan H S 2000 Polymer principles of protein calorimetric two-state cooperativity *Proteins: Struct. Funct. Genet.* **40** 637–61

Kaya H and Chan H S 2001 Polymer principles of protein calorimetric two-state cooperativity *Proteins: Struct. Funct. Genet.* **43** 523 (erratum)

[58] Bai Y, Sosnick T, Mayne L and Englander S W 1995 Protein folding intermediates: native-state hydrogen exchange *Science* **269** 192–7

[59] Qian H 1997 Thermodynamic hierarchy and local energetics of folded proteins *J. Mol. Biol.* **267** 198–206

[60] Ollerenshaw J E, Kaya H, Chan H S and 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** 14748–53

[61] Kaya H and 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

[62] Feng H, Zhou Z and Bai Y 2005 A protein folding pathway with multiple folding intermediates at atomic resolution *Proc. Natl Acad. Sci. USA* **102** 5026–31

[63] Irbäck A, Sjunnesson F and Wallin S 2000 Three-helix-bundle protein in a Ramachandran model *Proc. Natl Acad. Sci. USA* **97** 13614–8

[64] Knott M and Chan H S 2004 Exploring the effects of hydrogen bonding and hydrophobic interactions on the foldability and cooperativity of helical proteins using a simplified atomic model *Chem. Phys.* **307** 187–99

[65] Tiktopulo E I, Bychkova V E, Rička J and Ptitsyn O B 1994 Cooperativity of the coil-globule transition in a homopolymer—microcalorimetric study of poly(*N*-isopropylacrylamide) *Macromolecules* **27** 2879–82
[66] Klimov D K and Thirumalai D 1998 Cooperativity in protein folding: from lattice models with sidechains to real proteins *Fold. Des.* **3** 127–39
[67] Li M S, Klimov D K and Thirumalai D 2004 Finite size effects on thermal denaturation of globular proteins *Phys. Rev. Lett.* **93** 268107
[68] Li M S, Klimov D K and Thirumalai D 2005 Finite size effects on calorimetric cooperativity of two-state proteins *Physica* A **350** 38–44
[69] Kaya H and Chan H S 2003 Simple two-state protein folding kinetics requires near-Levinthal thermodynamic cooperativity *Proteins: Struct. Funct. Genet.* **52** 510–23
[70] Scalley-Kim M and 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** 573–83
[71] Gutin A, Sali A, Abkevich V, Karplus M and Shakhnovich E I 1998 Temperature dependence of the folding rate in a simple protein model: search for a 'glass' transition *J. Chem. Phys.* **108** 6466–83
[72] Kaya H and 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
[73] Matthews C R 1993 Pathways of protein folding *Annu. Rev. Biochem.* **62** 653–83
[74] Bilsel O and Matthews C R 2000 Barriers in protein folding reactions *Adv. Protein Chem.* **53** 153–207
[75] Akmal A and Muñoz V 2004 The nature of the free energy barriers to two-state folding *Proteins: Struct. Funct. Bioinform.* **57** 142–52
[76] Segawa S-I and Sugihara M 1984 Characterization of the transition state of lysozyme unfolding: I. Effect of protein-solvent interactions on the transition state *Biopolymers* **23** 2473–88
[77] Chen B-L, Baase W A and Schellman J A 1989 Low-temperature unfolding of a mutant of phage T4 lysozyme: 2. Kinetic investigations *Biochemistry* **28** 691–9
[78] Jaswal S S, Truhlar S M E, Dill K A and Agard D A 2005 Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases *J. Mol. Biol.* **347** 355–66
[79] Schindler T and Schmid F X 1996 Thermodynamic properties of an extremely rapid protein folding reaction *Biochemistry* **35** 16833–42
[80] Bachmann A and Kiefhaber T 2001 Apparent two-state tendamistat folding is a sequential process along a defined route *J. Mol. Biol.* **306** 375–86
[81] Dimitriadis G, Drysdale A, Myers J K, Arora P, Radford S E, Oas T G and 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** 3809–14
[82] Otzen D E and Oliveberg M 2004 Correspondence between anomalous m- and ΔC_p-values in protein folding *Protein Sci.* **13** 3253–63
[83] Lum K, Chandler D and Weeks J D 1999 Hydrophobicity at small and large length scales *J. Phys. Chem.* B **103** 4570–7