Solvation Effects and Driving Forces for Protein Thermodynamic and Kinetic Cooperativity: How Adequate is Native-centric Topological Modeling?

Hüseyin Kaya and Hue Sun Chan*

What energetic and solvation effects underlie the remarkable two-state thermodynamics and folding/unfolding kinetics of small single-domain proteins? To address this question, we investigate the folding and unfolding of a hierarchy of continuum Langevin dynamics models of chymotrypsin inhibitor 2. We find that residue-based additive Go-like contact energies, although native-centric, are by themselves insufficient for protein-like calorimetric two-state cooperativity. Further native biases by local conformational preferences are necessary for protein-like thermodynamics. Kinetically, however, even models with both contact and local native-centric energies do not produce simple two-state chevron plots. Thus a model protein’s thermodynamic cooperativity is not sufficient for simple two-state kinetics. The models tested appear to have increasing internal friction with increasing native stability, leading to chevron rollovers that typify kinetics that are commonly referred to as non-two-state. The free energy profiles of these models are found to be sensitive to the choice of native contacts and the presumed spatial ranges of the contact interactions. Motivated by explicit-water considerations, we explore recent treatments of solvent granularity that incorporate desolvation free energy barriers into effective implicit-solvent intraprotein interactions. This additional feature reduces both folding and unfolding rates vis-à-vis that of the corresponding models without desolvation barriers, but the kinetics remain non-two-state. Taken together, our observations suggest that interaction mechanisms more intricate than simple Go-like constructs and pairwise additive solvation-like contributions are needed to rationalize some of the most basic generic protein properties. Therefore, as experimental constraints on protein chain models, requiring a consistent account of protein-like thermodynamic and kinetic cooperativity can be more stringent and productive for some applications than simply requiring a model heteropolymer to fold to a target structure.

Introduction

A fundamental unresolved question in molecular biology is how solvent-mediated interactions conspire to produce the highly specific structures and dynamics of proteins. Recent experiments on highly cooperative “two-state” folding/unfolding kinetics of small single-domain proteins1–4 have, however, revealed an intriguing phenomenological simplicity. Most notably, the folding rates of these proteins are found to be well correlated with a simple contact order parameter deducible entirely from the native contact pattern, often referred to as a protein’s “topology”.1–7

From a reductionist viewpoint, protein behavior is ultimately determined by the large collection of atoms of the protein and those that constitute the

Keywords: calorimetric cooperativity; single-exponential kinetics; unfolding; chevron plot; desolvation barrier

*Corresponding author
simplified lattice 14 – 19 or continuum 19 – 21 representation 23 – 25 as well as explicit Go- like chain native-centric models without an explicit chain and folding rate have been provided by Ising-type of the observed relationship between contact order useful insights. 8 – 10 As well, recent developments solvent. Indeed, computational studies of protein folding/unfolding kinetics. But in a recent generalization of the Gaussian elastic network model that took into account chain connectivity, a significant correlation between experimental folding rates and the relaxation rates of the slowest vibrational modes was discovered, 44 suggesting an intimate connection between near-native vibrations and folding/unfolding kinetics.

Some of the successes of native-centric approaches have been attributed to the stipulation 28, 32 that Go-like potentials are protein-like to a certain degree because they serve to eliminate “to first order” 28 or to minimize 32 the “energetic frustration” that is presumed to be minimized in real proteins. 54 – 56 According to this view, native-centric models are thus left free to account for “topological frustration” 28, 32 alone, i.e. to capture the physics arising solely from chain connectivity, excluded volume, and the favorability of the native fold.

While we cherish the successes of the native-centric paradigm, it is also important to not lose sight of its limitations. In short, native-centric modeling entails: (i) admitting our ignorance of the basic physics of protein folding, at least for the time being; (ii) recognizing that a protein sequence’s known native structure may contain significant information about its actual energetics; (iii) assuming that a Go-like potential inferred from the known native structure is in some sense an approximate description of the underlying physical energetics; (iv) working out the logical consequences of these assumptions to gain insight into various aspects of the folding process. In this perspective, native-centric modeling should be taken as a tentative means to capture collective atomic/molecular effects that we do not yet understand. As such, its application may be physically more meaningful at a coarse-grained level (perhaps as a “renormalized” description). Although all-atom Go-like models 56, 44 are obviously superior in accounting for the important effects of side-chain packing (see, e.g. insightful discussion in Ref. 46), physically it is even harder to justify why the interaction between a pair of atoms in general would depend on whether they are in close spatial proximity in a particular protein’s native structure. Historically speaking, the renewed popularity of Go-like models in protein folding studies since the late 1990s may arguably represent a partial backtracking, albeit a very productive and well justified one, in modeling philosophy. This is so because the desire to supersede these earlier ad hoc interaction schemes appeared to be an impetus for the emergence since the late 1980s of simplified lattice protein folding models with general sequence-dependent potentials.

Here we endeavor to better delineate the utility and limitation of several common native-centric approaches to protein folding. In identifying their strengths as well as weaknesses, our goal is to pave the way for improving native-centric modeling and better reductionist approaches. It goes without saying that Go-like models are intrinsically incomplete, because (i) possible non-native interactions are in large measure neglected, 63, 64 and (ii) interaction (energetic) heterogeneities can be present in proteins with essentially identical topologies. 46 Practically speaking, even if a general usefulness of native-centric modeling is presumed, robustness of the predictions has to be ascertained.
Many native-centric interaction schemes can be postulated; not all of them have the same predictions. Some discrepancies are puzzling. For example, the combination of a Go-like potential with an explicit chain representation should, in principle, be a more adequate model of topological frustration than models without an explicit chain representation. Yet so far native-centric constructs with geometrically less realistic (non-explicit) chain representations appear to be more successful in reproducing experimental folding rates than direct folding kinetics simulations of Go-like models with explicit chain representations.

Specifically, the present work addresses two basic questions of robustness in native-centric modeling: (i) how much do the model predictions depend on the choice of native contacts for a given protein? (ii) to what extent would these predictions be modified when the effects of the protein’s aqueous solvent are taken into account with more sophistication? Pursuing a line of inquiry we have recently developed in the context of lattice models, we focus here on whether continuum coarse-grained Go-like energetics with an explicit chain representation can reproduce certain generic thermodynamic and kinetic cooperativities that have been experimentally observed across many real proteins. These statistical mechanics tests are stringent. For instance, the mere existence of a qualitatively sharp folding transition in a chain model does not necessarily imply that its underlying thermodynamics is protein-like. Homopolymers can have very sharp coil–globule transitions that are not calorimetrically two-state. Comparisons between simulated and experimental chevron plots show that even for chain models that satisfy the experimental thermodynamic two-state criteria, it is non-trivial to reproduce the highly cooperative non-glassy two-state folding kinetics of many small single-domain proteins. Therefore, applying these tests would, in due course, facilitate the improvement of existing models, suggest yet unexplored avenues of native-centric topological modeling, and ultimately help decipher the energetics of real proteins.

Comparing Different Native Contact Sets For CI2

We consider the 64-residue truncated form of chymotrypsin inhibitor 2 (CI2) using coarse-grained Cα representations with side-chain interactions accounted for by contacts between pairs of Cα positions separated by at least three Cαs along the chain sequence (contact order ≥ 4). CI2 is a widely studied small single-domain protein with no disulfide bond. It folds and unfolds as an apparent simple two-state system. CI2 is an ideal test case because a large body of experimental, all-atom molecular dynamics, and native-centric
modeling data is available). To investigate how coarse-grained native-centric model predictions may be sensitive to the definition of native contacts, here we examine two native contact sets, which we refer to as NCS1 and NCS2.

NCS1 is determined by the distance criterion of Shea et al., two amino acid residues i and j of a given protein are in contact if, in its native structure from the Protein Data Bank (PDB), either their Cα atoms are less than 8 Å apart, or any two heavy atoms one from each of their two side-chains are less than 4 Å apart, or both. Using this definition, there are 137 NCS1 contacts. NCS2 is borrowed from Clementi et al.’s native contact map for CI2 (Figure 2 of 32). NCS2 has 142 contacts. It was based upon the CSU software, which takes into account more detailed structural information such as contact surface area and solvent accessibility. There are considerable variations in native Cα–Cα distances among contacts in both NCS1 and NCS2. The minimum native contact distance is 4.325 Å for both sets, but the maximum are 12.255 Å and 15.558 Å for NCS1 and NCS2, respectively. The average native contact distance of NCS1 (6.528 Å) is smaller than that of NCS2 (7.288 Å).

However, the average sequence separations of NCS1 (23.1) and NCS2 (22.6) are almost identical.

Figure 1 compares the two native contact sets. They have 108 contacts in common (blue lines in Figure 1(b)). Among the native contacts that are not common to both sets, those belonging to NCS1 but not NCS2 (green lines in Figure 1(c)) tend to be between two ends of the chain or involve the β1 strand (residues 27–34). In contrast, contacts belonging to NCS2 but not NCS1 (red lines in Figure 1(d)) appear to be more uniformly distributed, involving more the α-helix (residues 13–23) and the region spanning residues 35–44. Specific examples of these differences are provided in Figure 2, showing that NCS1 identifies an hydrophobic–polar (alanine–arginine) contact but not an hydrophobic–hydrophobic (valine–phenylalanine) contact.

Models and Methods

Coarse-grained potentials without solvation/ desolvation barriers

The basic construct of our native-centric potentials follows that of Clementi et al. For a given model protein conformation specified by the positions of all its Cα atoms, the total Go-like potential energy:

\[
V_{\text{total}} = V_{\text{stretching}} + V_{\text{bending}} + V_{\text{torsion}} + V_{\text{non-bonded}}
\]

\[
= \sum_{\text{bonds}} K_r (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi (\phi - \phi_0)
\]

\[
\times \left\{ K_{\phi}^{(1)} (1 - \cos (\phi - \phi_0)) + K_{\phi}^{(3)} (1 - \cos (3\phi - 3\phi_0)) \right\}
\]

\[
+ \sum_{i<j<3} \sum_{\text{native}} \epsilon \left[ \frac{1}{(r_{ij}^\text{native})^{12}} - \frac{1}{(r_{ij}^\text{native})^{6}} \right] + \sum_{i<j<3} \sum_{\text{non-native}} \epsilon \left( \frac{r_{ij}^\text{rep}}{r_{ij}^\text{rep}} \right)^{12}
\]

(1)

where \( N \) is the total number of particles. This functional form has also been used by Koga and Takada. Here the first three summations are for local interactions, where \( r, \theta, \) and \( \phi \) are, respectively, the Cα–Cα virtual bond length between successive residues along the chain sequence, Cα–Cα virtual bond angles, and Cα–Cα virtual torsion angles; \( r_0, \theta_0, \) and \( \phi_0 \) are the corresponding native values in the PDB structure. These terms account for chain connectivity and presumed local conformational preferences for the native fold. The last two summations are for non-local interactions; \( r_0 \) is the spatial distance between two Cαs that have at least three residues between them along the chain sequence. In the summation over native contacts (as defined above for either NCS1 or NCS2), a 10–12 Lennard Jones (LJ) form is used, where \( r_{ij} \) is the Cα–Cα distance between the contacting residue \( i \) and residue \( j \) in the PDB structure.
the summation over non-native contacts, \( r_{op} \) parameterizes the excluded volume repulsion between residue pairs that do not belong to the given native contact set. As in Refs 32 and 35, we use \( r_{op} = 4 \) Å (whereas Ref. 28 uses \( r_{op} = 7.8 \) Å). The ratios between interaction parameters are \( K_5 = 100 \epsilon, K_6 = 20 \epsilon, K_8^{(1)} = \epsilon, \) and \( K_6^{(3)} = 0.5 \epsilon, \) as in Ref. 32. The interaction strength is thus controlled by a single parameter \( \epsilon \). We refer to the potential just described as the “without-solvation” model because it does not have a solvation/desolvation barrier (see below), although the terms in equation (1) may be interpreted as part of an implicit-solvent scheme that takes into account other aspects of solvent-mediated interactions.

Equation (1) assumes that native-centric favorable interactions have relatively long spatial ranges. In alternate square-well Gō models,\(^{40}\) however, favorable contact interactions have sharp cutoffs. Moreover, in many lattice models, contact interactions may be viewed as having infinitesimal spatial ranges. Thus, to investigate how the presumed spatial ranges of contact interactions may affect model predictions, we study a variation of the above model that restricts each of the pairwise 10–12 LJ native contact terms in equation (1) to \( r_{ij} \leq 1.2 r_{ij}^{\alpha} \) and sets the interaction to zero for \( r_{ij} > 1.2 r_{ij}^{\alpha} \), but all other aspects of the model stay the same. We call this the “without-solvation-SSR” (short spatial range) model.

An approximate account of solvation/desolvation barriers

We consider also coarse-grained “with-solvation” models designed to semi-quantitatively account for the solvation/desolvation free energy barriers encountered by a protein’s constituent groups as they cluster together in aqueous solvents (Figure 3). We refer to these barriers simply as “desolvation barriers” below. Desolvation barriers are a robust consequence of granularity or the particulate nature of the solvent.\(^{76}\) They have long been predicted by theory\(^{77}\) and atomic simulations.\(^{78,79}\) However, aside from an earlier study that used a square-well/square-shoulder form of desolvation barriers,\(^{80}\) until very recently,\(^{40,80,81}\) this salient physical feature was not taken into account in continuum coarse-grained protein models. While explicit-solvent molecular dynamics account for solvation effects directly, these simulations do not yet provide a definitive answer as to whether they can or cannot reproduce the experimentally observed thermodynamic and kinetic cooperativities in protein folding. Therefore, complementary “implicit-solvent” treatments\(^{40,84,80,81}\) like the present approach are needed. Indeed, the experimentally based cooperative tests conducted here should also be applied to all-atom models once their computational efficiency has improved to make it possible.

The scope of the present work is limited. In particular, the study of structural details, such as connections to the powerful experimental \( \Phi \)-value analysis of transition-state structures,\(^{47,72,74}\) is deferred to future applications of our investigative framework. We first tackle a little-explored but fundamental question: how deeply are protein folding thermodynamic and kinetic cooperativities affected by the introduction of generic desolvation

Figure 3. (a) Model with-solvation interactions between two amino acid residues belonging to a given native contact pair in the present study (defined by either NCSI or NCS2); \( r \) is their Cα–Cα separation. The potential energy \( U(r) \), shown here and in (b) in units of \( \epsilon \), depends on the native Cα–Cα distance \( r \) of a given contact in the PDB structure. The \( r \) values shown in this Figure are only for illustrative purposes. They do not correspond to actual contacts in the present NCSI or NCS2 models (see the text). \( U(r) \) here is defined by the functional form of Cheung et al.\(^{40}\) with \( k = 6, n = 2, m = 3, \epsilon' = 0.2 \epsilon, \) and \( \epsilon'' = 0.1 \epsilon \), where \( k, n, \) and \( m \) parameterize the functional form for \( r < r' \), \( r' \leq r < r'' \) and \( r'' \geq r' \), respectively (e.g., the excluded volume repulsion \( \sim r^{-2} \), see equation on page 689 of Ref. 40 for details), \( \epsilon' = \) the depth of water-separated minimum, and \( \epsilon'' = \) the height of the desolvation peak. The two potential functions shown in (a) are for \( r' = 4.0 \) Å (left) and \( r'' = 6.5 \) Å (right). The cartoons (for \( r'' = 6.5 \) Å) illustrate contact and water-separated minima configurations,\(^{86,87}\) where a water molecule is depicted as a solid circle of diameter \( = 3.4 \) Å. (b) With-solvation native potential for \( r' = 3.8 \) Å in the present study (as labeled) is compared with: (i) The \( r' = 3.8 \) Å functional form of Cheung et al. with the same values for \( k, n, m, \) and \( \epsilon \) as in (a), but with \( \epsilon' = \epsilon'/3 \) and \( \epsilon'' = 5 \epsilon'/9 \). (ii) The explicit-water simulated methanol-methane PMF at 25°C under atmospheric pressure obtained by Shimizu and Chan,\(^{60}\) in a unit such that the free energy at contact equals \( -\epsilon = -1 \). (LJ) The 10–12 Lennard-Jones without-solvation potential \( \epsilon(5(r'/r)^{12} - 6(r'/r)^{10}) \) (as in equation (1)) with \( r' = 3.8 \) Å. SSR. The corresponding LJ cutoff at 1.2\( r' \) in without-solvation-SSR models.
barriers? To this end, we employ the general implicit-solvent functional form introduced recently by Cheung et al.\(^{40}\) (Figure 3). The repulsive part of this potential (for \(r < r'\)) is similar, though not identical, to the repulsive part of the 10–12 LJ term in the without-solvation model above (equation (1)). The key difference is that now a free energy barrier is present at the midpoint \((r' + r)/2\) between the contact \((r')\) and water-separated \((r')\) free energy minima of a given pairwise interaction; \(r' = 3.0\) Å is the approximate diameter of a single water molecule. Shown in Figure 3(b)(i) is a potential with relative magnitudes of the barrier and minima similar to that in Ref. 40. This form has a relatively high desolvation barrier\(\dagger\). The \(U(r)\) function in the present study has a lower barrier (Figure 3(a)). As our goal is only to elucidate the generic implications of having a significant desolvation barrier, provided that the barrier is not negligible, a lower barrier is advantageous because it allows for faster kinetics and thus broader conformational sampling. Not the least, our choice is not inconsistent with recent explicit-water atomic simulations that predicted a lower pairwise desolvation barrier\(\ddagger\) (Figure 3(b)(iii)).

Now, for the with-solvation model, we simply replace the pairwise 10–12 LJ terms of the second summation over native contacts in the \(V_{\text{total}}\) equation (1) above with \(U(r)\)s (Figure 3(a)) for the corresponding native pairs. Other terms in equation (1) remain the same. We call the resulting potential function \(V_{\text{total}}^\beta\). Again, the interaction strength of a given model is controlled by one single parameter \(\epsilon\). In principle, terms in both the without-solvation and with-solvation potentials representing solvent-mediated interactions can depend on temperature.\(^{84–86}\) To simplify the formulation, however, and especially since most of the results in this report entail comparing kinetic trajectories under a constant given temperature, here \(V_{\text{total}}\) and \(U(r)\) are taken to be temperature independent, as in Refs. 32,40.

### Langevin dynamics

Folding and unfolding kinetics are simulated by Langevin dynamics,\(\ddagger\), using a formulation similar to Thirumalai and co-workers’.\(^{88,89}\) For each of the 3\(N\) degrees of freedom of the model protein \((x, y, z)\) coordinates of the C’s), the equation of motion is:

\[
\begin{align*}
 m \ddot{\mathbf{r}}(t) &= F_{\text{cont}}(t) - m \gamma \mathbf{v}(t) + \mathbf{\eta}(t) \\
 & \quad (2)
\end{align*}
\]

where \(m\), \(v\) and \(F_{\text{cont}}\) \(\gamma\) and \(\eta\) are respectively, mass, velocity, acceleration, conformational force, friction (viscosity) constant and random force. The conformational force is equal to the negative gradient of the total potential energy of the given model \((V_{\text{total}}\) or \(V_{\text{total}}^\beta\)). For the without-solvation-SSR models, conformational force from the pairwise 10–12 LJ native contact term in \(V_{\text{total}}\) between residues \(i\) and \(j\) is applied only if \(r_{ij} \leq 1.2r_i\). The random force has the autocorrelation function:

\[
\langle \eta(t)\eta(t') \rangle = 2m\gamma k_B T \delta(t - t')
\]

where \(k_B T\) is Boltzmann constant times absolute temperature. Every C’ is subject to a random force at each integration time step. The components of the random force are independently generated by setting \(\eta_i = (2m\gamma k_B T/\delta t)^{1/2} \xi_i\). Here \(\xi_i\) denotes the uncorrelated random force components in the \(x, y\) or \(z\) directions, \(\xi\) is a random value taken from a Gaussian distribution with zero mean and unit variance (obtained from a random number generator by standard techniques),\(^{90}\) and \(\delta t\) is the integration time step. At the commencement of a simulation at temperature \(T\), the initial velocities are assigned random values by setting \(v_i = (k_B T/m)^{1/2} \xi_i\).

We use the velocity-Verlet algorithm\(^{88–91}\) (equations (12) and (13) in Ref. 89) to integrate equation (2). Independent of simulation conditions such as variations in \(\epsilon\) and \(T\), the time scale of the model systems here is always controlled by the quantity \(\tau = \sqrt{m \gamma^2/\epsilon_0}\), with the length scale \(a = 4\) Å and a reference energy scale \(\epsilon_0 = 1\). We further set \(\gamma = 0.05 \text{ Å}^{-1}\) and use a molecular dynamics time step \(\delta t = 0.005 \text{ fs}\) in the numerical integration. Conformational sampling is performed by averaging over snapshots taken at every 400 time steps. Simulation times in this study are presented in units of \(\delta t\). The energy parameter \(\epsilon\) and temperature \(T\) are given, respectively in units of \(\epsilon_0\) and \(\epsilon_0/k_B\), and length is measured in units of Å.

To simplify notation, other units are chosen such that \(m = 1\) and \(k_B = 1\) in the present simulations, as in Veitshans et al.\(^{89}\) An approximate correspondence between model time and real protein kinetic time scales can be found in Ref. 89.

### Thermodynamic Cooperativity

#### Free energy profiles in different native-centric schemes

Using the progress variable \(Q\) (native contact fraction), Figure 4 shows that conformational distribution is significantly sensitive to the choice of native contact set and the presumed spatial ranges of native contact interactions. Consistent with the expectation for a two-state protein such as CI2 and a previous without-solvation study,\(^{32}\) the free energy profiles for NCS2 (continuous curves) exhibit a single peak at intermediate \(Q\) separating the native (high \(Q\)) and denatured (low \(Q\)) minima. In contrast, the NCS1 free energy profile has a plateau-like transition region in the without-solvation formulation (Figure 4(a), broken curve). More \(\dagger\) In order not to have a negative \(\epsilon' / \epsilon\) ratio, it appears that the relation \((\epsilon' - \epsilon')/(\epsilon' - \epsilon) = 1.33\) in the legend for Figure 1 in Cheung et al.,\(^{40}\) should read

\((\epsilon' + \epsilon')/(\epsilon' - \epsilon) = 1.33\).

\(\ddagger\) Alternatively, Newtonian dynamics in conjunction with the Berendsen et al. algorithm\(^{97}\) for coupling to a heat bath was used by several previous investigations\(^{32,53}\) of similar Go-like coarse-grained protein models.
at $\epsilon/k_BT = 1.165$ and 1.098 for the NCS1 and NCS2 with-solvation models. For without-solvation models in (a) and (b), the condition for contact is $r \leq r^* = (r_0 + r_1)/2$, i.e. when the C$^\alpha$–C$^\alpha$ distance $r$ is within the contact basin ($r$ not larger than that of the desolvation peak), as in Ref. 40.

remarkably, for the without-solvation-SSR and with-solvation models (Figure 4(b) and (c)), the NCS1 profiles develop a shallow minimum flanked by two peaks in the intermediate $Q$ region (broken curves), similar to certain postulated free energy profiles discussed previously, for example, by Fersht$^{32}$ and Chu and Bai$^{33}$ in the context of folding kinetics that apparently involves intermediates. Also notable is the progressive movement of the native minimum position from $Q \approx 0.9$ for the without-solvation models toward $Q = 1$ for the with-solvation models. The incorporation of desolvation barriers dramatically raises the overall folding/unfolding free energy barrier for NCS2, but only has a relatively subdued effect for NCS1 (cf. Figure 4(b) and (c)), suggesting that there is an intricate interplay between desolvation barrier effects and other aspects of solvent-mediated interactions in protein folding.

Calorimetric cooperativity: local conformational preferences are crucial

Figure 5 assesses the calorimetric cooperativity$^{47,48,68}$ of seven different native-centric models of CI2 by comparing their simulated van’t Hoff over calorimetric enthalpy ratios $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ to the experimental two-state requirement that $\Delta H_{\text{vH}}/\Delta H_{\text{cal}} \approx 1$. Model intraprotein interactions are taken to be temperature independent in this evaluation. Since vibrations along the bonds (equation (1)) contribute to heat capacity in these models outside the folding/unfolding transition region, and there is experimental evidence for heat capacity contributions from bond vector motion in real proteins,$^{34}$ the simulated $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ ratio without baseline subtractions does not correspond physically to the experimental $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ ratio obtained by empirical baseline subtractions.$^{47,48,68}$ Thus, only the baseline-subtracted $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ ratio $\kappa_0^{(r)}$ from the models are judged against the experimental calorimetric two-state criterion.$^{47,48,68}$ Figure 5 shows that $\Delta H_{\text{vH}}/\Delta H_{\text{cal}} = \kappa_0^{(r)} \approx 1$ is satisfied by all six models described in the last section. Apparently, similar Gō-like models in Refs 32 and 40 also exhibit calorimetric cooperativity. This is evident from their reported heat capacity scans although $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ ratios were not computed in these works.

The role of local interactions is addressed by a different coarse-grained model with Gō-like (through-space) contact interactions but very little local (through-bond) preference for the CI2 native structure. The setup of this “contact-dominant” model is similar to that of the NCS2 without-solvation-SSR model: It has the same virtual bond setup as well as in the (full LJ) without-solvation-SSR setup as well as in the (full LJ) without-solvation-SSR formulation. These models have even bigger difficulties reaching conformations with $Q \approx 1$ than the contact-dominant model.

---

$^\dagger$ We have also studied similar “contact-only” models with $K_1 = K_{1}^{(r)} = K_2^{(r)} = 0$ in the same without-solvation-SSR setup as well as in the (full LJ) without-solvation formulation. These models have even bigger difficulties reaching conformations with $Q \approx 1$ than the contact-dominant model.
was obtained from Langevin dynamics simulation near the transition midpoint. Figure 5 shows that it has a double hump, which is clearly dissimilar to the single-peak heat capacity scans of two-state proteins such as CI2\(^{72,95}\). Moreover, near this model’s temperature for the peak heat capacity value, the distribution of \(Q\) has only a single population maximum rather than being bimodal (data not shown). Indeed, a few highest and lowest \(Q\) values were so improbable that they were not sampled. These thermodynamically non-cooperative features are reflected by an exceedingly low van’t Hoff over calorimetric enthalpy ratio of \(k_2 = 0.33\).

One may conceivably argue from the “energetic versus topological frustration” perspective\(^{28,32}\) that energetic frustration has already been eliminated in the contact-dominant model because its potential favors native contacts, disfavors non-native contacts, and even slightly favors native bond angles and torsion angles. Yet the contact-dominant model’s thermodynamics is not protein-like. The non-cooperative behavior of this particular contact-dominant model might have been exaggerated by the exclusion of \((i,i+3)\) contacts in its formulation (see equation (1)). Nonetheless, the present result echoes several recent findings of less-than-protein-like thermodynamic cooperativity in continuum models with Go-like contact interactions but without local conformational preferences. These model studies include coarse-grained and all-atom discontinuous molecular dynamics models\(^{41,45}\) as well as a self-consistent field theory.\(^{37}\)

On the other hand, some three-dimensional lattice “contact-only” Go models are thermodynamically cooperative,\(^{67}\) probably because of default lattice restrictions on bond angles and torsion angles. However, in continuous space, the “negative design” afforded by Go-like contact interactions alone are apparently insufficient for protein-like thermodynamics. Indeed, a protein sequence’s ability to fold to a unique structure may be partially encoded in local signals.\(^{96,97}\) Protein-like behavior requires minimization of energetic frustration of the target native structure as well as enhanced frustration in the competing non-native conformations.\(^{98}\) A comparison between the contact-dominant model and the other models with local native propensities in Figure 5 suggests that an interplay between local conformational preference and non-local compactification forces\(^{47,48,67,68,99,100}\) are necessary for protein-like thermodynamic cooperativity. For this conclusion to be properly interpreted, we hasten to add that...
structural details of side-chain packing, hydrogen bonding, as well as general non-native-centric physical restrictions on bond angles and torsion angles (as in standard non-Gō-like force fields) have not been taken into account in the present coarse-grained (residue-based) contact-dominant model. But these effects are operative in real proteins. Clearly, these interactions must be part of the physical basis of any local propensity for the native fold in a more complete all-atom description.

**Kinetic Cooperativity**

**Sharp kinetic transitions between two thermodynamic states**

Folding kinetics in explicit-chain Gō-like models have been investigated using equilibrium sampling in conjunction with free energy profile analyses as well as direct dynamics simulations. Here, around their respective transition midpoints, all six native-centric C12 models (NCS1 or NCS2, with or without solvation) have kinetic characteristics consistent with their thermodynamic two-state cooperativity. Figure 6(a) and (b) show that the kinetic transitions between the native and denatured ensembles are sudden and sharp. Figure 6(c) and (d) show that the distributions of potential energy and Q are bimodally well separated into native and denatured regions, and the correlation between potential energy and Q is generally linear. A consistency check has also been made using Figure 6(c), which provided an average kinetic energy of 78.9. Equating this with $3NT/2$ for $N = 64$ (equipartition theorem) yields $T = 0.8219$, which is essentially identical to the input simulation temperature of $T = 0.82$, as it should. Figure 6(c) and (d) further indicate that after the initiation of folding around the transition midpoint, pre-equilibration of the denatured ensemble is rapid relative to the folding time scale.

**Chevron plots: matching kinetics with thermodynamics?**

Bearing in mind that protein thermodynamic cooperativity is necessary but not sufficient for simple two-state folding/unfolding kinetics, we proceed to evaluate model predictions against experimental stability curves and chevron plots. To do so, we determine model folding and unfolding rates using direct dynamics simulations over extensive ranges of native stability by varying the interaction parameter $\epsilon$ at constant temperature.
The natural logarithms of the rates are plotted as functions of \( \epsilon \) in Figures 7–9. Inasmuch as it was computationally feasible, first passage times (FPTs) of a large number of trajectories were used to provide reliable estimates of MFPTs (Tables 1–3). As one of us has argued, the variation of \( \epsilon \) may serve as a tentative model for varying denaturant concentration, though the detailed physics of how the effects of chemical denaturants should be incorporated into coarse-grained protein models is a subject of ongoing research. Here we view the upper panels of Figures 7–9 as model equivalences of chevron plots.

Native stability curves of the models as functions of \( \epsilon \) are plotted in the lower panels of Figures 7–9. They show that the free energy of unfolding between the native minimum and low-\( Q \) open conformations are approximately linear in \( \epsilon \) (upper continuous and broken curves). These quasi-linear stability curves estimated from simulation data around the transition midpoint correspond to those obtained experimentally by empirical linear extrapolation from directly measured data around the transition region. In contrast, the free energy difference between the native minimum and a denatured-state ensemble encompassing low-\( Q \) as well as intermediate-\( Q \) conformations (lower continuous and broken curves) is non-linear in \( \epsilon \), similar to that observed in previous lattice model studies. This is an expected feature intimately connected to the multiple-conformation nature of the native state and is consistent with recent native-state hydrogen exchange experiments. These characteristics of native stability underscore the fact that the operational definition of calorimetric two-state behavior (see above) does not necessarily imply that all denatured conformations have the same stability. Even for calorimetrically two-state proteins under native conditions well below the global folding/unfolding transition midpoint, the population of partially unfolded conformations can sometimes be non-negligible as long as it does not exceed a certain threshold.

Figures 7–9 show that the transition midpoints determined by thermodynamics and kinetics are quite close, with only minor discrepancies. The discrepancies for NCS1 models appear to be slightly larger in Figures 8 and 9. This is probably related to the high-free-energy minima in the transition regions of the corresponding NCS1 free energy profiles (Figure 4(b) and (c)). More surprisingly, however, is that even with their native-centric potentials, all six models fail to produce the type of simple two-state folding/unfolding kinetics observed experimentally for CI2 and many other small single-domain proteins. The operational definition for simple two-state folding/unfolding kinetics requires that the logarithmic folding and unfolding rates under constant temperature be approximately linear in native stability, and that the natural logarithm of the directly measured and linearly extrapolated (folding rate)/(unfolding rate) is a subject of ongoing research.
rate) ratio as a function of denaturant concentration matches the directly measured and linearly extrapolated equilibrium free energy of unfolding in units of $k_B T$. Here, the dashed-dotted V-shapes in the upper panels of Figures 7–9 show that as $-\epsilon/k_B T$ is changed at constant $T$ from the transition midpoint towards either more native or more denaturing conditions, the respective trends of increase in simulated folding or unfolding rate fall short of this requirement for the kinetics to be simple two-state. Instead, our models’ behavior is more akin to proteins that exhibit chevron rollovers, such as ribonuclease A and barnase, whose kinetics are operationally referred to as non-two-state. Comparisons with experimental chevron plots have not been made in other studies of continuum Go models, but the reported results indicate that they also do not predict simple two-state chevron behavior (see, e.g. Figure 2 of Ref. 34).

The four without-solvation and without-solvation-SSR models in Figures 7 and 8 show a clear rollover in both the folding and unfolding arms of their chevron plots. Reflecting the lower barriers along their free energy profiles (Figure 4), kinetics are generally faster for the without-solvation and without-solvation-SSR than the corresponding with-solvation models (Figure 9). For the with-solvation models, the rate at a given $-\epsilon/k_B T$ is substantially slower for NCS2 than that for NCS1. This trend is consistent with NCS2’s much higher free energy barrier in the transition region (Figure 4(c)). Most remarkably, comparing Figures 7 and 8 against Figure 9 demonstrates a dramatic impact of desolvation barriers on the folding/unfolding kinetics. In contrast to the chevron plots with significant curvatures for the
Table 1. Number of trajectories \( N \) used in the present study to determine the MFPT of folding and unfolding for the NCS1 and NCS2 without-solvation models (\( T = 0.82 \), Figure 7)

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>Unfolding NCS1</th>
<th>Unfolding NCS2</th>
<th>Folding NCS1</th>
<th>Folding NCS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.60</td>
<td>0.1734, 100</td>
<td>0.1763, 100</td>
<td>1.00, 0.6905</td>
<td>1.00, 0.9250</td>
</tr>
<tr>
<td>0.65</td>
<td>0.2472, 100</td>
<td>0.2547, 100</td>
<td>0.95, 0.9534</td>
<td>0.95, 1.1668</td>
</tr>
<tr>
<td>0.70</td>
<td>0.3825, 100</td>
<td>0.4212, 100</td>
<td>0.90, 1.4466</td>
<td>0.90, 1.3102</td>
</tr>
<tr>
<td>0.75</td>
<td>0.7807, 100</td>
<td>0.8950, 100</td>
<td>0.89, 1.5062</td>
<td>0.89, 1.3865</td>
</tr>
<tr>
<td>0.77</td>
<td>1.2540, 1100</td>
<td>1.9052, 1100</td>
<td>0.88, 1.8175</td>
<td>1.00, 1.5577</td>
</tr>
<tr>
<td>0.78</td>
<td>1.7684, 100</td>
<td>2.8292, 100</td>
<td>0.87, 2.1760</td>
<td>1.00, 2.0039</td>
</tr>
<tr>
<td>0.79</td>
<td>2.4983, 100</td>
<td>4.7546, 100</td>
<td>0.86, 2.2865</td>
<td>1.00, 2.1039</td>
</tr>
<tr>
<td>0.80</td>
<td>2.3120, 100</td>
<td>8.0901, 100</td>
<td>0.85, 3.0018</td>
<td>1.00, 2.7999</td>
</tr>
<tr>
<td>0.82</td>
<td>3.4755, 100</td>
<td>40.737, 100</td>
<td>0.81, 4.2971</td>
<td>1.00, 4.0108</td>
</tr>
</tbody>
</table>

Each MFPT listed is the average (arithmetic mean) over \( N \) first passage times for the given interaction strength \( \epsilon \). Time is measured from the start of a given simulation at \( t = 0 \) in units of \( dt \) (see the text).

Table 2. Same as Table 1, but for the without-solvation-SSR models (\( T = 0.64 \), Figure 8)

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>Unfolding NCS1</th>
<th>Unfolding NCS2</th>
<th>Folding NCS1</th>
<th>Folding NCS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.1804, 100</td>
<td>0.1856, 100</td>
<td>1.30, 1.0080</td>
<td>1.30, 0.9121</td>
</tr>
<tr>
<td>0.75</td>
<td>0.2450, 100</td>
<td>0.2730, 100</td>
<td>1.25, 1.1324</td>
<td>1.25, 1.0924</td>
</tr>
<tr>
<td>0.80</td>
<td>0.3362, 100</td>
<td>0.3759, 100</td>
<td>1.20, 1.4063</td>
<td>1.20, 1.0754</td>
</tr>
<tr>
<td>0.85</td>
<td>0.4489, 100</td>
<td>0.5304, 100</td>
<td>1.18, 1.4804</td>
<td>1.18, 1.2380</td>
</tr>
<tr>
<td>0.90</td>
<td>0.9161, 100</td>
<td>1.2724, 100</td>
<td>1.13, 2.3409</td>
<td>1.13, 1.6971</td>
</tr>
<tr>
<td>0.93</td>
<td>1.3364, 100</td>
<td>2.3473, 100</td>
<td>1.10, 2.8513</td>
<td>1.10, 2.2875</td>
</tr>
<tr>
<td>0.95</td>
<td>4.7098, 100</td>
<td>15.767, 100</td>
<td>1.08, 6.8365</td>
<td>1.08, 4.8426</td>
</tr>
<tr>
<td>0.97</td>
<td>9.4389, 100</td>
<td>40.261, 100</td>
<td>1.03, 10.385</td>
<td>1.03, 9.6890</td>
</tr>
<tr>
<td>0.99</td>
<td>16.268, 91</td>
<td>42.673, 91</td>
<td>1.00, 43.174</td>
<td>1.00, 25.440</td>
</tr>
<tr>
<td>1.00</td>
<td>51.148, 18</td>
<td>262.14, 18</td>
<td>0.97, --</td>
<td>0.97, --</td>
</tr>
</tbody>
</table>

Table 3. Same as Table 1, but for the with-solvation models (\( T = 0.82 \), Figure 9)

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>Unfolding NCS1</th>
<th>Unfolding NCS2</th>
<th>Folding NCS1</th>
<th>Folding NCS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>0.5559, 600</td>
<td>0.6549, 500</td>
<td>1.50, 1.2196</td>
<td>1.50, 2.1393</td>
</tr>
<tr>
<td>0.50</td>
<td>0.5419, 126</td>
<td>1.5792, 137</td>
<td>1.40, 2.4237</td>
<td>1.40, 3.5297</td>
</tr>
<tr>
<td>0.60</td>
<td>1.6697, 111</td>
<td>7.2719, 121</td>
<td>1.30, 3.9479</td>
<td>1.30, 5.5725</td>
</tr>
<tr>
<td>0.65</td>
<td>2.7900, 103</td>
<td>17.122, 50</td>
<td>1.25, 5.4452</td>
<td>1.25, 8.7000</td>
</tr>
<tr>
<td>0.70</td>
<td>5.4940, 1100</td>
<td>44.590, 205</td>
<td>1.20, 6.8649</td>
<td>1.20, 11.766</td>
</tr>
<tr>
<td>0.75</td>
<td>8.4002, 120</td>
<td>104.18, 34</td>
<td>1.18, 8.7936</td>
<td>1.18, 19.365</td>
</tr>
<tr>
<td>0.80</td>
<td>15.463, 91</td>
<td>293.56, 26</td>
<td>1.15, 10.823</td>
<td>1.15, 30.141</td>
</tr>
<tr>
<td>0.83</td>
<td>26.101, 51</td>
<td>--</td>
<td>1.10, 19.633</td>
<td>1.10, 48.000</td>
</tr>
<tr>
<td>0.85</td>
<td>42.512, 74</td>
<td>722.32, 15</td>
<td>1.08, 20.336</td>
<td>1.08, --</td>
</tr>
<tr>
<td>0.90</td>
<td>119.01, 32</td>
<td>1224.2, 6</td>
<td>1.05, 41.682</td>
<td>1.05, --</td>
</tr>
<tr>
<td>0.92</td>
<td>--</td>
<td>1922.0, 4</td>
<td>1.03, 43.365</td>
<td>1.03, --</td>
</tr>
<tr>
<td>0.93</td>
<td>168.05, 24</td>
<td>--</td>
<td>1.00, 75.048</td>
<td>1.00, --</td>
</tr>
<tr>
<td>0.95</td>
<td>216.32, 37</td>
<td>--</td>
<td>0.97, 93.742</td>
<td>0.97, --</td>
</tr>
</tbody>
</table>

Each MFPT listed is the average (arithmetic mean) over \( N \) first passage times for the given interaction strength \( \epsilon \). Time is measured from the start of a given simulation at \( t = 0 \) in units of \( dt \) (see the text).
Single-exponential relaxation

Experimentally, kinetic relaxation of many small single-domain proteins such as CI2 \(^{72,95}\) and some apparently non-two-state proteins with chevron rollovers \(^{108,114}\) are found to be essentially single-exponential. Therefore, it is of interest to ascertain whether the present models embody this hallmark, even though they are not kinetically simple two-state. For this purpose, we examine the distributions of first passage times (FPTs, as defined in Figures 7–9). Let \( P(t) dt \) be the probability for the FPT of a given kinetic process to lie within a range \( dt \) around time \( t \). If the relaxation is single-exponential:

\[
\int_{t_0}^{t} dt' P(t') = 1 - \exp(-k(t - t_0))
\]

where \( k \) is the kinetic rate, and \( t_0 \geq 0 \) is a minimum FPT to take into consideration the finite time needed for pre-equilibration after initiation of the kinetic process at \( t = 0 \). It follows that:

\[
\text{MFPT} = \int_{t_0}^{t} dt' P(t') = t_0 + 1/k
\]

To assess whether a given FPT distribution conforms to this description, a quantity \( P(t) \Delta t \) is computed by binning FPTs into time slots \(^{115}\) of size \( \Delta t \). If the kinetic process is single-exponential:

\[
\ln[P(t) \Delta t] = \left\{ \ln \left( \frac{\Delta t}{\text{MFPT} - t_0} \right) + \frac{t_0}{\text{MFPT} - t_0} \right\}
\]

\[
- \frac{t}{\text{MFPT} - t_0}
\]

i.e. \( \ln[P(t) \Delta t] \) versus \( t \) should be a straight line with slope \( = -(\text{MFPT} - t_0)^{-1} \).
Figure 10(a) shows that even under strongly native conditions concomitant with a significant chevron rollover, the NCS2 without-solvation-SSR model has approximately single-exponential relaxation. This behavior echoes that of a recent four-helix-bundle lattice model68 (Figure 10(b)). Consistent with equation (6), a comparison between the filled and open circles in Figure 10(a) indicates that while changing the bin size \( \Delta t \) naturally changes the \( \ln[P(t)|\Delta t] \) values, reasonable variations in \( \Delta t \) do not affect the slope of the \( \ln[P(t)|\Delta t] \) distribution. Figure 11 applies similar analyses to folding and unfolding in other models in the present study under representative native and denaturing conditions. Owing to computational limitations, the sample sizes for the FPT distributions are not very large, especially for the without-solvation models in Figure 11(c). Consequently, a certain level of statistical uncertainties ensued. Nonetheless, Figure 11 shows that for all cases tested, our data are consistent with single-exponential relaxation. As pointed out by Fersht,\(^\dagger\) the high-free-energy minima along the NCS1 free energy profiles (Figure 4(b) and (c)) do not preclude apparent single-exponential kinetics. The viability of equation (6) for our models is further buttressed by the relatively small differences between the slopes of the least-square-fitted lines in Figure 11 and the quantity \( -(\text{MFPT} - t_0)^{-1} \), where \( t_0 \) is taken to be the minimum FPT encountered in the simulated trajectories of a given model: For the models and their simulation conditions listed in the legend of Figure 11, and in the same order, \( \{10^6 \times (\text{MFPT} - t_0)^{-1}, [-10^6 \times \text{slope}]\} = \{10.0, 10.6\}, \{6.13, 6.71\}, \{6.11, 6.53\}, \{7.28, 7.56\}, \{28.6, 36.1\}, \{6.41, 7.02\}, \{19.5, 26.4\}, \{8.53, 9.36\}, \{2.05, 2.08\}, \{0.504, 0.431\}, \{0.242, 0.189\}, \{0.199, 0.161\}.\(^\ddagger\)

The native-centric formulations in the present Gō-like models lead to folding rates that are at least four orders of magnitude faster than the experimental CI2 folding rates. At 25°C and pH 6.3, the experimental CI2 folding rates at zero denaturant (native stability \( \Delta G = 12.0 k_B T \)) and the transition midpoint (in 3.92 M GdnHCl, native stability \( \Delta G = 0 \)) are, respectively, 47.8 s\(^{-1}\) and 0.035 s\(^{-1}\).\(^\ddagger\) If we use the physical argument of Veitshans et al.\(^\ddagger\) to identify the Langevin time unit \( \delta t \) with a real time scale of \( \sim 10^{-14} \) seconds, the folding rate of the NCS2 with-solvation model in Figure 9 is \( \sim 10^6 \) s\(^{-1}\) at \( \Delta G = 12.0 k_B T \) and \( 10^3 \) s\(^{-1}\) at \( \Delta G = 0 \). Corresponding folding rates of other models in Figures 7–9 are even faster by approximately two orders of magnitude. Despite these discrepancies, native-centric constructs do capture part of real protein energetics. This is evident from studies of extensive sets of real proteins using explicit-chain Gō models, wherein theoretically predicted folding and relaxation rates were found to correlate reasonably well with the experimental folding rates. However, it is noteworthy that the spread of these model-predicted rates among the set of proteins tested is apparently at least 1.5–2 orders of magnitude narrower than the diversity of experimental folding rates.\(^\ddagger\) (cf. Figure 5 of Ref. 44). This suggests that certain

\(\dagger\) Rates in the chevron plots in Figures 7–9 are computed by taking \( t_0 = 0 \). Our calculations indicate that using finite \( t_{\text{fs}} \) instead of \( t_0 = 0 \) to determine the rates \( k \) via equation (5) only leads to minimal modifications on the chevron plots (data not shown). The conclusions regarding rollovers and non-two-state kinetics remain unchanged.
basic aspects of protein energetics are yet to be
taken into account by common Gō-like models. In a similar vein, the chevron rollovers in Figures 7–9 represent a failure to account for the high degree of diversity in folding rates of a given protein under different native conditions. For real CI2, the folding rates at zero denaturant and at the transition midpoint differ by three orders of magnitude. But the Gō-like models in Figures 7–9 predict only one order of magnitude difference.

Chevron rollover: stability-dependent front factor?

To better understand the chevron rollovers, Figure 12 applies a protocol we recently developed\textsuperscript{68} to assess the models’ conformity to the commonly employed transition state picture in interpreting protein folding experiments. Model data are now fitted to the expression:

\[
\text{rate} = F(\epsilon, T)\exp\left[-\frac{\Delta G^\dagger(\epsilon, T)}{k_B T}\right]
\]

for folding or unfolding rate, taken as \((\text{MFPT})^{-1}\) from the direct dynamics simulations. On the other side of the above equation, \(\Delta G^\dagger\) is an activation free energy determined solely by thermodynamic Boltzmann weights\textsuperscript{68} using the method of Nymeyer et al.,\textsuperscript{32,116} \(F\) is the corresponding front factor.\textsuperscript{2,30,59,68,98,117} Figure 12 shows that, in contrast to the usual stipulation\textsuperscript{118} that the front factors of small single-domain proteins such as CI2 are essentially independent of intraprotein interaction strength and native stability, the \(F\) factors deduced from the present analysis are highly sensitive to \(\epsilon\). This implies that thermodynamic analyses of free energy profiles alone cannot predict the \(\epsilon\)-dependencies of the kinetic rates,\textsuperscript{38,39,68,117} and the chevron rollovers are underpinned by native-stability-dependent front factors in these models.\textsuperscript{68} This hypothesis regarding the physical origin of chevron rollover may soon be testable by single-molecule techniques.\textsuperscript{119} In addition to the definitions for unfolded, transition, and folded states (data not shown). Whereas the absolute value of \(F\) varies somewhat, the overall trend of dependence on \(\epsilon\) remains essentially unchanged. This resilience is similar to that observed in our previous analysis of the folding front factor of a 55mer lattice model (Figure 5 of Ref. 68).

Thus, for this key aspect of chevron behavior, the present native-centric models’ kinetics clearly do not resemble the simple two-state kinetics of CI2.\textsuperscript{72,95} The ramification of this finding is far-reaching, as it bears on the basic energetics of protein folding (see Discussion below). As they stand, the apparently non-two-state kinetics of these physical self-contained polymer models\textsuperscript{68} also shed light on the folding of other proteins that exhibit similar chevron rollovers as well.\textsuperscript{93,113,114} To date, rationalizations of chevron rollovers include deadtime intermediates,\textsuperscript{120} specific kinetic traps,\textsuperscript{98,102,121} peak-shifting on complex free energy profiles,\textsuperscript{93,122} burst
phase continuum,\textsuperscript{123} and internal friction as manifested by front factors that depend on native stability (Ref. \textsuperscript{68} and discussion therein). These perspectives are not necessarily mutually exclusive. For example, internal friction may arise from kinetic trapping mechanisms (H.K. & H.S.C., unpublished results). In any event, chevron rollover is an unequivocal prediction of the present models, irrespective of whether Q or other folding reaction coordinates are used for the transition state analysis.\textsuperscript{124} Figure 12(d) shows that the folding front factor decreases with more native conditions, and the unfolding front factor also decreases with more denaturing conditions. In short, there appears to be an aversion to speed in these models’ energetics. We tentatively attribute the slowing down in these models to a possible combination of effects of internal friction (conformational search) and external friction (implicit solvent viscosity). The origins of these effects remain to be better elucidated. For example, in some modeling situations,\textsuperscript{68} folding-arm rollovers are related to the onset of downhill folding.\textsuperscript{125,126} The chevron rollovers in the folding and unfolding arms of the NCS2 without-solvation model may be similarly related to downhill scenarios (see, e.g. the $\epsilon = 0.90$ and $\epsilon = 0.70$ profiles in Figure 12(a)). At least for the NCS2 with-solvation model in Figure 6, the fact that no deadtime intermediate was observed during our simulation suggests that such a mechanism is not necessary for chevron rollovers.\textsuperscript{68,93} In this example, chevron rollover emerges as a kinetic front-factor effect.

Discussion

We have compared two different native contact sets, and three different formulations of Gō-like interactions with and without desolvation barriers. The predictions of these native-centric models were evaluated against generic thermodynamic and kinetic properties of small single-domain proteins that these models were designed to mimic in the first place. We learnt several lessons. First, protein-like thermodynamic cooperativity requires non-local contact-like interactions acting in concert with local conformational probabilities for the native fold.\textsuperscript{47,48,67,68} (Figure 5). Second, some basic predictions of native-centric models, such as the free energy profiles in Figure 4, are significantly dependent on the native contact set and interaction scheme used, even if the choice is made among physically reasonable definitions. A recent study\textsuperscript{86,89} shows also that free energy profiles of native-centric models are sensitive to the chain’s presumed persistence length and energetic barriers to bond rotations.\textsuperscript{96,102} Third, we found that pairwise desolvation barriers in native-centric models could lead to some protein-like properties such as a higher free energy barrier separating the native and denatured states (cf. Figure 4(a), (b) and (c)) as well as more linear chevron plots (cf. Figures 7–9). These predictions are encouraging as they provide insight into corresponding features in real proteins. Fundamentally, however, the kinetics of all present native-centric models for CI2 do not resemble that of real CI2. The models with pairwise desolvation barriers, like those without, are kinetically non-two-state in the operational sense that they have large chevron rollovers.

Fourth, the significant differences between the predictions of with- and without-solvation models underscore the importance of proper accounting for the energetic cost of water expulsion in protein folding models, and that caution should be used when interpreting results obtained from effective potentials that do not have desolvation barriers.\textsuperscript{83,127} The barrier height in the present with-solvation models simulated at $T = 0.82$ is 0.24$k_BT$. For real proteins, the desolvation barrier heights encountered by the polypeptide chain as a part of the potential of mean force are expected to be sensitive to temperature. Thus, the present results should also bear on explicit-solvent unfolding simulations at high temperatures and the degree of dependencies of protein folding mechanisms on temperature.\textsuperscript{126} Of relevance here is the model system of a pair of methanes in water. Recent Monte Carlo simulations in the TIP4P water model indicate that their desolvation barrier is reduced from approximately 0.16 kcal/mol to 0.12 kcal/mol ($0.27k_BT$ to $0.16k_BT$) when temperature is increased from 298 K to 368 K under atmospheric pressure.\textsuperscript{84} Under typical high-temperature unfolding conditions of 498 K and a water density of 0.829 gm/ml,\textsuperscript{128} the desolvation barrier height is further reduced to $\approx 0.05$ kcal/mol or 0.05$k_BT$ (Figure 16.3 in Ref. 18; S. Shimizu and H.S.C., personal communication).

In a broader perspective, solvent-mediated interactions are known to be intrinsically pairwise non-additive,\textsuperscript{76,83} and the collapse of a hydrophobic chain may involve large length-scale dynamic effects.\textsuperscript{129} In this light, that the pairwise desolvation barriers here fail to produce simple two-state chevron plots is not too surprising. Indeed, recent explicit-water simulations show that the sign of heat capacity of the free energy barrier against folding is opposite to that against the association of a pair of methane molecules.\textsuperscript{84,85,106} Considerations of a three-methane model system further indicate that the height of desolvation barrier is clearly non-additive, and the sign and magnitude of this non-additivity is dependent upon the configuration of the non-polar solutes involved.\textsuperscript{83} Hence, solvation effects beyond the pairwise formulation considered here are likely needed to account for simple two-state protein folding/unfolding kinetics.

In summary, the present findings imply that the actual solvent-mediated interactions in real proteins are much more specific and well-designed than one would otherwise posit. Real small, single-domain proteins are more cooperative than common Gō-like models with pairwise additive
interactions. Nonetheless, recent innovations in
native-centric modeling have been immensely
valuable. As discussed above, they do capture
part of the essential physics. Many deep insights
would not have been gained without them. 1,29,32,33
But, at the same time, the limitations of common
Gō-like chain models 67,68 may be more basic than
previously appreciated. The present analysis implies
that more protein-like interaction schemes are yet
to be discovered. Every model considered here except
the contact-dominant variety can fold to the CI2
native structure. Qualitatively, the free energy pro-
files of the NCS2 models fit the expectation for that
of small single-domain proteins as well. Yet their kin-
etics are fundamentally different from that of CI2.
Thus, a protein model’s ability to fold to one single
target structure does not guarantee the adequacy of
its energetics; and the microscopic origin of simple
two-state folding/unfolding kinetics remains to be
elicitated. Our effort to address some of these ques-
tions is underway. Apparently, chevron rollovers can
be essentially eliminated in more cooperative chain
models with added energetic favorabilities for the
ground-state and near-ground-state structures
beyond that provided by the additive schemes in
common Gō models. These results will be presented
in a subsequent report (H.K. & H.S.C., unpublished
results). In the ongoing quest for a better under-
standing of protein energetics through the design
and interpretation of novel physical models, pro-
tein-like statistical mechanics properties such as
calorimetric two-state cooperativity 41,47,48,67,68,130,131
and simple two-state chevron behavior 68 should be
useful as stringent but necessary modeling

Acknowledgements
We thank Yawen Bai, Margaret Cheung, Cecilia
Clementi, Ken Dill, Angel García, Chinhui Guo,
Carol Hall, Anders Irbäck, Sophie Jackson, John
Karanicolas, Bob Matthews, Cristian Micheletti,
Hugh Nymeyer, Mikael Oliveberg, José Onuchic,
Kevin Plaxco, John Portman, Eugene Shakhnovich,
Joan-Emma Shea, Seishi Shimizu, Kim Sneppen,
Dev Thirumalai, Michele Vendruscolo, Peter
Wolynes, and Yaqqi Zhou for helpful discussions
during the period in which these ideas were
developed. This work was partially supported by the
Canadian Institutes of Health Research (CIHR
grant no. MOP-15323), a Premier’s Research
Excellence Award from the Province of Ontario,
and the Ontario Centre for Genomic Computing at
the Hospital for Sick Children in Toronto. H.S.C. is
a Canada Research Chair in Biochemistry.

References
folding reactions. Advan. Protein Chem. 53,
153–207.
3. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S.,
Fast kinetics and mechanisms in protein folding.
Annu. Rev. Biophys. Biomol. Struct. 29,
321–359.
unifying basis in protein-folding mechanisms:
contact order, chain topology, stability, and the
USA, 97, 1525–1529.
Contact order, transition state placement and the refolding
rates of single domain proteins. J. Mol. Biol. 277,
985–994.
392, 761–763.
7. Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker,
D. (2000). Topology, stability, sequence, and length:
defining the determinants of two-state protein-fold-
ing kinetics. Biochemistry, 39, 11177–11183.
protein folding intermediate observed in a 1-micro-
second simulation in aqueous solution. Science, 282,
740–744.
using replica exchange molecular dynamics. Pro-
protein folding, unfolding and fluctuations by com-
puter simulation. 1. The effect of specific amino acid
sequence represented by specific inter-unit inter-
Natl Acad. Sci. USA, 75, 554–558.
representations of globular proteins: how good are
15. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. &
energy landscape of protein folding: a synthesis.
16. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee,
of protein folding—a perspective from simple exact
433–439.
hierarchy of complexities. In Current Topics in
Computational Molecular Biology (Jiang, T., Xu, Y.
& Zhang, M. Q., eds), pp. 403–447, The MIT Press,
Cambridge, MA.
theory: from lattice to all-atom models. Annu. Rev.
the timescales and mechanisms of protein folding


Edited by C. R. Matthews

(Received 19 August 2002; received in revised form 2 December 2002; accepted 2 December 2002)