January 13, 2014

Effects of Desolvation Barriers and Sidechains on Local-Nonlocal Coupling and Chevron Behaviors in Coarse-Grained Models of Protein Folding

Tao CHEN and **Hue Sun CHAN**

Departments of Biochemistry, of Molecular Genetics, and of Physics,
University of Toronto, Toronto, Ontario M5S 1A8, Canada

Running title: Local-Nonlocal Coupling in Protein Folding

Corresponding author information:
Hue Sun CHAN. E-mail: chan@arrhenius.med.toronto.edu
Tel: (416)978-2697; Fax: (416)978-8548
Mailing address: Department of Biochemistry, University of Toronto, Medical Sciences Building – 5th Fl., 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

Abstract

Local-nonlocal coupling is an organizational principle in protein folding. It envisions a cooperative energetic interplay between local conformational preferences and favorable nonlocal contacts. Previous theoretical studies by our group showed that two classes of native-centric coarse-grained models can capture the experimentally observed high degrees of protein folding cooperativity and diversity in folding rates. These models either embody an explicit local-nonlocal coupling mechanism or incorporate desolvation barriers in the models' pairwise interactions. Here a conceptual connection is made between these two paradigmatic coarse-grained interaction schemes by showing that desolvation barriers enhance local-nonlocal coupling. Furthermore, we find that a class of coarse-grained protein models with a single-site representation of sidechains also increases local-nonlocal coupling relative to mainchain models without sidechains. Enhanced local-nonlocal coupling generally leads to higher folding cooperativity and chevron plots with more linear folding arms. For the sidechain models studied, the chevron plot simulated with entirely native-centric intrachain interactions behaves very similarly to the corresponding chevron plots simulated with interactions that are partly modulated by sequence- and denaturant-dependent transfer free energies. In these essentially native-centric models, the mild chevron rollovers in the simulated folding arm are caused by occasionally populated intermediates as well as the movement of the unfolded and putative folding transition states. The strength and limitation of the models are analyzed by comparison with experiment. New formulations of sidechain models that may provide a physical account for nonnative interactions are also explored.

Introduction

A protein's behavior is encoded by its amino acid sequence and other aspects of its chemical composition, which collectively determine the physico-chemical interactions of the protein with itself and with its environment. Protein behaviors that are stipulated to be governed only by conservative forces may be described, as for any such physical system, by a potential energy that depends only on the positions of all constituent particles, i.e., those comprising the protein as well as the solvent. In biologically relevant aqueous settings, the dynamics of a protein is generally coupled to that of its surrounding solvent molecules. However, for dynamic processes such as protein folding that are significantly slower than the relaxation timescales of the solvent, the kinetic properties of a protein — along with its thermodynamic properties — may be characterized to a good approximation by an implicit-solvent energy landscape in which the solvent degrees of freedom are pre-averaged. In essence, such an energy landscape is a potential of mean force formulated as a function solely of the protein's conformational coordinates.

The shape of the energy landscape of a protein is thus dependent upon its amino acid sequence. For globular proteins, in accordance with the experimentally inspired consistency principle [1, 2] and principle of minimal frustration [3], the energy landscape is expected to be funnel-like [4, 5]. However, not all landscapes of natural proteins share this topography. Many proteins playing key roles in cellular signaling and regulation are intrinsically disordered, lacking the tendency to fold to a unique structure by themselves [6, 8–10]. Although some intrinsically disordered proteins undergo coupled folding and binding [12] and thus retain funnel-like features in the combined landscapes of these proteins and their binding targets [13–16], some intrinsically disordered proteins can remain highly dynamic with a "fuzzy" [17] conformational distribution even upon binding to their functional partners [18–20].

Folding Cooperativity. Here we restrict our attention to globular proteins with funnel-like landscapes. One property shared by many globular proteins is their high degrees of folding cooperativity, i.e., their folded and unfolded states are well separated in enthalpy [21] and related structural/energetic measures [22–25]. Folding cooperativity is not a corollary of a globular protein's ability to fold to a unique native structure [26]. Experiments showed that some "downhill folder" proteins can adopt an essentially unique native structure under strongly folding conditions but lack a free barrier separating the folded and unfolded states [27,28]. The degree of folding cooperativity achievable by a protein is likely constrained physically by its native topology [29–34]. Taken together, these observations suggest that folding cooperativity is probably an evolved property [35, 36].

Possible benefits of this trait include enhancing kinetic stability of folded proteins against their transformation into nonfunctional forms [37], minimizing risk of disease-causing amyloidogenesis [38, 39], and modulating or abrogating the potentially complicating effects of cotranslational folding [40–45]. In a broader perspective, folding cooperativity may be viewed as a manifestation, in one molecule, of the bistability and switch-like behaviors that are ubiquitous in biology [46–48].

Considerable effort has been taken in the past 15 years to elucidate the physical origin of folding cooperativity. Much advance has been made but our understanding is as yet incomplete. Early investigations uncovered that the issue is more complex than many protein scientists had surmised. For instance, the conformational densities of states [49] of common chain models that were set up to embody the presumably dominant role of hydrophobic effect in protein folding [50, 51] do not lend themselves to proteinlike folding cooperativity [52, 53]. This finding led to the view that many-body effects [54, 55] beyond pairwise additive interactions, such as a cooperative interplay between desolvation and hydrogen bonding strength [56,57], are required to account for folding cooperativity [23,52] as well as the tremendous diversity [58, 59] in experimentally observed folding rates of small, single-domain proteins [60, 61].

Local-Nonlocal Coupling. An inference from empirically successful theories of cooperative folding [59, 62–65] is that folding cooperativity is probably underpinned significantly by a local-nonlocal coupling mechanism [53, 59]. The local-nonlocal coupling hypothesis stipulates that energy landscapes of cooperatively folding proteins are organized in such a way that the tendency for a segment of a protein to adopt locally native structure is weak in isolation but is greatly enhanced by formation of proximate nonlocal native contacts. Likewise, it stipulates that nonlocal native contacts are less favorable unless the chain segments containing the contacting residues adopt locally native structure, which presumably would result in better packing in accordance with the consistency [1] and minimal frustration [3] principles. This hypothesized organizing principle is supported by experimental findings of subglobal cooperative units [66] as well as Ising-like behaviors in the folding of repeat proteins [67] (reviewed in Ref. [53]). The mechanism may be viewed as a sort of coupled folding and binding similar to that observed in the binding of some intrinsically disordered proteins [12] except the process is now between different parts of the same protein molecule.

Desolvation Barriers. Desolvation barriers have been identified as a likely physical origin of folding cooperativity [68–71]. Desolvation barriers represent the energetic penalty of water exclusion, as seen in the potentials of mean force between nonpolar solutes [72], when different parts of a protein molecule come together to form a well-packed core [73]. Desolvation barriers and the associated enthalpic barriers are a robust feature in intra- and intermolecular protein interactions [74] and should therefore be a staple in protein energy landscapes. Simulations of coarse-grained C_α chain models with desolvation barriers in their native-centric potentials showed that such models can rationalize to a large extent the experimentally observed folding cooperativity [69–71] and huge diversity in the folding rates of proteins with different native topologies [53, 75]. More recently, the cooperativity-enhancing effects of desolvation barriers are seen to be fundamentally related to their narrowing of the attractive range of intrachain interactions, which is a crucial determining factor of cooperativity in protein (heteropolymer) folding [76] as well as homopolymer coil-globule transitions [77]. Desolvation barriers are in essence a many-body effect resulting from averaging over solvent degrees of freedom. Nonetheless, from a formulational standpoint, it is noteworthy that pairwise additive native-centric potentials with desolvation barriers are quite adequate in accounting for proteinlike cooperativity without invoking explicit many-body terms.

Sidechain Effects. Another likely physical origin of folding cooperativity is sidechain packing [78, 79]. An early study of sidechain effects on folding cooperativity employed model chains configured on the simple cubic lattice. Each residue was modeled by two adjacent positions on the lattice, one for the mainchain and the other for the sidechain (i.e., a two-bead representation was used). Interactions between sidechain beads were governed by a transferrable potential [79]. Similar lattice models have also been used to investigate the role of nonnative interactions in protein folding [80]. The folding transitions of the lattice sidechain models were found to be significant sharper than the corresponding (mainchain) model without sidechains [79], confirming the expectation that sidechains tend to enhance folding cooperativity. However, the sidechain lattice models considered at the time do not satisfy the more rigorous van't Hoff/calorimetric enthalpy [23] or linear-chevron [81] criterion for cooperative folding, probably because those model chains with only 15 residues were too short to mimick more realistic protein behavior. In contrast, a class of recently developed continuum (off-lattice) two-bead sidechain models [82, 83] exhibits rather proteinlike cooperative behaviors [83,84] and has been applied successfully to rationalize many aspects of experimental folding behaviors [85–87].

Conceptual Questions and Coarse-Grained Modeling. In view of the above summary of what we have learned in recent years about folding cooperativity, it is natural to inquire about the relationship between the proposed organizing principle of local-nonlocal coupling on one hand, and the two likely physical origins of folding cooperativity — namely desolvation barriers and sidechain packing — on the other. We limit the scope of the present study to these two likely contributors, though undoubtedly there are other significant contributors such as hydrogen bonding to cooperative folding that we have to leave for future investigations. One of our main goals here is to ascertain whether and, if so, how desolvation barriers and sidechain effects enhance local-nonlocal coupling. As it stands, these conceptual questions are best addressed by coarse-grained protein chain models, such as those employed in the works discussed above, because the computational tractability of these models allows for efficient generating and testing of hypotheses. Accordingly, we utilize and compare results from existing coarse-grained models as well as develop new formulations for our present purposes. Coarse-grained protein chain models have provided much insights into protein folding [35, 53, 85, 87–98] and are complementary to models that contain higher structural and energetic details.

Recent advances in atomic simulations of protein folding by using distributed computing [99] or highly efficient special-purpose computers [100] are impressive. Notable success in ab initio folding has now been achieved for an increasing number [101–106], though not all [107], small proteins for which folding simulation was attempted. Although effort is still needed to further improve existing forcefields [108–110], including their ability to accurately describe not only the folded state but also unfolded conformational ensembles [111, 112], these recent developments hold great promises for modeling and understanding a wide range of biomolecular processes in atomistic detail [113, 114]. However, the computational capability of current all-atom simulations is not yet sufficient for our questions of interest because the issues we address require simulations of thermodynamics and kinetics of protein folding that are far more extensive than what is currently achievable in explicit-solvent atomic simulations. For instance, accurate determination of a chevron plot requires simulating thousands to tens of thousands of folding events under different folding conditions (see below), whereas successful all-atom, explicit-solvent simulations of folding reported to date involved only several to at most hundreds of folding events [100–106]. Nonetheless, with continuing advances in computational technology, it is quite possible that the questions addressed in the present work can become accessible to all-atom molecular dynamics simulations in the not-too-distance future.

Chevron Plots: Denaturant Dependence and Nonnative Interactions. As part of our study of folding/unfolding cooperativity, we have also examined the chevron properties of the desolvation-barrier and continuum sidechain models. Chevron plots provide logarithmic folding/unfolding or relaxation rates as functions of denaturant concentration [115, 116]. Through the efforts of many experimental groups, a large repertoire of chevron plots has been accumulated in past decades. It is therefore important to advance theoretical understanding of chevron plots because this form of data constitutes a major and rather comprehensive source of experimental information on protein folding/unfolding kinetics. Theoretical analyses indicate that cooperativity of a protein's folding/unfolding kinetics is generally linked to the protein's thermodynamic cooperativity [63, 70, 71]. Proteins that fold in a highly cooperative, simple two-state-like manner exhibit linear folding and unfolding arms in their chevron plots [117]. In contrast, nonlinear folding or unfolding chevron arms — showing either a downward curvature ("rollover") or a upward curvature — indicate more complex kinetics with transiently populated intermediates [116–120], sequential barriers [121], or parallel kinetic pathways with different dependence on denaturant [122, 123].

Previous studies have demonstrated that models with cooperativity-enhancing features of desolvation barriers [71, 124], local-nonlocal coupling [59] or other forms of many-body effects [30] can lead to model chevron plots with close-to-linear folding and unfolding arms that are qualitatively similar to those observed experimentally for small, single-domain proteins. Without an explicit account of denaturant effects, these model chevrons treated denaturant implicitly by using native stability as a proxy for denaturant concentration [30, 71, 124]. This limitation is now partially overcome by models that incorporate denaturant-dependent transfer free energies for the sidechains [83]. In light of this recent development, it is instructive to compare the model chevron plots obtained with an implicit vs. an explicit treatment of denaturant. We provide such a comparison below.

Another issue of interest is the effects of nonnative interactions on chevron behavior. Recent atomic simulations suggests that the role of nonnative interactions in the folding of small proteins is minimal [125], but nonnative interactions do affect the folding pathways of many proteins, especially larger ones, as is evident from experimentally observed transiently populated intermediates and chevron rollovers [126]. Recent studies showed that effects of specific nonnative interactions on protein folding can be modeled [124, 127–130] using a mixed native-centric and sequence-dependent formulation [131, 132]. In particular, this method was successful in accounting for the extreme chevron rollover [124] that has been observed experimentally for the de novo designed protein Top7 [36]. The tranfer free energies in the recently developed sidechain models of Thirumalai and coworkers [82, 83] offer a possibility for modeling nonnative interactions because the transfer free energies constitute a transferrable (not native-centric) potential. However, in the formulation to date [82, 83], the sidechain model for zero denaturant, i.e., under strongly folding conditions, was invariably native-centric. In those modeling constructs, the transfer free energies were used only to weaken intrachain interactions as denaturant concentration increases. Here we develop alternate formulations in which the sequence-dependent transfer free energies can strengthen – not only weaken – intrachain interactions to explore the possibility of using such models to capture nonnative interactions in protein folding.

Models and Method

We pursue the goals stated above by comparing select thermodynamic and kinetic properties predicted by several representative coarse-grained protein chain models. The models we consider are a C_α-Gō model, a desolvation-barrier (db) model, and three formulations of a class of sidechain (SC) models.

$\mathbf{C}_\alpha$-Gō model and db model. The potential energy function for the present C_α-Gō model is identical to that given in Eq. (1) of Ref. [70], wherein the contact energy E_c for a native pair i, j (a pair of residues belonging to the native contact set, see below) at a distance r_{ij} apart is given by a 12-10 potential (Fig. 1a). The potential energy function for the present db model is given by Eqs. (1)–(3) of Ref. [71]. For this model, the interaction energy E_c between a native pair is given by the potential energy $U(r_{ij}; r_{ij}^\mathrm{n}, \epsilon, \epsilon_\mathrm{db}, \epsilon_\mathrm{ssm})$ defined in this reference [71], with db height $\epsilon_\mathrm{db} = 0.1\epsilon$ and solvent-separated minimum well depth $\epsilon_\mathrm{ssm} = 0.2\epsilon$, where ϵ is the contact-minimum well depth and r_{ij}^n is the distance between residues i, j in the Protein Data Bank (PDB) structure (Fig. 1a). The present simulation parameters, including the force constants for the bond-length, bond-angle, and torsion-angle terms, are identical to those in Ref. [70]. As in Refs. [70, 71], a uniform length scale of $r_\mathrm{rep} = 4.0$ Å is used for the repulsive interactions between nonnative residues. Langevin dynamics is used for folding kinetics simulations and thermodynamic sampling in accordance with the formulation and parameters provided in Ref. [71]. As in our previous studies [70, 71], simulation time is reported in number of Langevin time steps. In the construction of our native-centric C_α-Gō and db models, any two residues i, j separated by at least three residues along the chain sequence ($|i - j| > 3$) and have at least one pair of nonhydrogen atoms, one from each residue, that are less than 4.5 Å apart in the PDB structure are defined to belong to the native contact set [29, 133]. This definition of native contact set is identical to that used in several recent studies from our group [29–31, 71, 75, 127] though it is slightly different from the definitions for native contact sets NCS1 and NCS2 in Ref. [70]. We use $\tilde{Q}_\mathrm{n}$ to denote the total number of contacts in the native contact set. During Langevin dynamics simulations, a pair of residues i, j belonging to the native contact set is considered to be forming a native contact in the C_α-Gō model if $r_{ij} \leq 1.2 r_{ij}^\mathrm{n}$. For the db model, the corresponding condition for native contact formation is $r_{ij} \leq r^\mathrm{db}$ where r^db is the position of the db peak in the pairwise native-centric potential [70, 71]. As before [29–31, 71, 75, 127], we use the fractional number of native contacts Q as a progress variable for folding [134, 135].

Sidechain (SC) models. Our approach to SC effects is largely adapted from the two-bead C_α-SCM formulations of Klimov and Thirumalai [136] and O'Brien et al. [82] in which amino acid residue i of a protein with n residues is represented by a C_α (backbone) atom with position vector $\mathbf{r}_{\alpha,i}$ and a pseudo-atom centered at the SC centroid with position vector $\mathbf{r}_{\mathrm{SC},i}$ (except when the amino acid residue is glycine, for which no $\mathbf{r}_{\mathrm{SC},i}$ is defined). Our SC models are not entirely identical to theirs, as will be described below. The minor variations between our and their SC models are technical in nature but these minor variations may nonetheless serve to assess the robustness of predictions from this class of coarse-grained SC models.

Following the modeling framework of O'Brien et al. [82], the total potential energy $E_{\mathrm{T}}(\{\mathbf{r}_\alpha\},\{\mathbf{r}_{\mathrm{SC}}\},C)$ of the model protein as a function of all its n backone positions $\{\mathbf{r}_\alpha\}$ and n SC positions $\{\mathbf{r}_{\mathrm{SC}}\}$ is given by two components, viz.,

$$E_{\mathrm{T}}(\{\mathbf{r}_\alpha\},\{\mathbf{r}_{\mathrm{SC}}\},C)=\sigma(E_{\mathrm{T}})_{\mathrm{SC-G\bar{o}}}(\{\mathbf{r}_\alpha\},\{\mathbf{r}_{\mathrm{SC}}\})+\Delta G_{\mathrm{trf}}(\{\mathbf{r}_{\mathrm{b}}\},\{\mathbf{r}_{\mathrm{SC}}\},C)\ , \tag{1}$$

where $(E_{\mathrm{T}})_{\mathrm{SC-G\bar{o}}}(\{\mathbf{r}_\alpha\},\{\mathbf{r}_{\mathrm{SC}}\})$ is the potential energy for the sidechain-Gō (SC-Gō) model and $\Delta G_{\mathrm{trf}}(\{\mathbf{r}_{\mathrm{b}}\},\{\mathbf{r}_{\mathrm{SC}}\},C)$ is a transfer free energy term that depends on the concentration C of cosolvent (denaturant in our case). The relative contributions of these two terms to (E_{T}) are controlled by the scaling factor σ. The SC-Gō potential is a sum of several contributions:

$$(E_{\mathrm{T}})_{\mathrm{SC-G\bar{o}}}=E_{\mathrm{bond}}+E_{\mathrm{angle}}+E_{HB}+E_{\mathrm{c}}+E_{NB}^{NN}\ , \tag{2}$$

where E_{bond}, E_{angle}, E_{HB}, E_{c}, and E_{NB}^{NN} are the bond-length, bond-angle, hydrogen-bonding, native-pair nonbonded (contact) potential, and nonnative repulsive terms, respectively. Here,

$$E_{\mathrm{bond}}=K_{\mathrm{b}}\left\{\sum_{i=1}^{n-1}\Big[|\mathbf{r}_{\alpha,i+1}-\mathbf{r}_{\alpha,i}|-|\mathbf{r}^{\mathrm{n}}_{\alpha,i+1}-\mathbf{r}^{\mathrm{n}}_{\alpha,i}|\Big]^2+\sum_{i=1}^{n}\Big[|\mathbf{r}_{\mathrm{SC},i}-\mathbf{r}_{\alpha,i}|-|\mathbf{r}^{\mathrm{n}}_{\mathrm{SC},i}-\mathbf{r}^{\mathrm{n}}_{\alpha,i}|\Big]^2\right\}\ , \tag{3}$$

where for computational efficiency we have used a stiff spring constant $K_{\mathrm{b}}=100.0$ kcal mol^{-1}Å^{-2} similar to that used in Ref. [136] to limit variations of virtual bond lengths instead of taking the approach in Ref. [82] that uses SHAKE [137] to fix them. In the above equation, the conformational coordinates $\mathbf{r}_\alpha$s and $\mathbf{r}_{\mathrm{SC}}$s are in units of Å. All summations over SC coordinates in this work, including that in Eq. (3), are restricted to non-glycine SCs. In Eq. (3), the difference $|\mathbf{r}^{\mathrm{n}}_{\alpha,i+1}-\mathbf{r}^{\mathrm{n}}_{\alpha,i}|$ is the reference ("equilibrium") distance between the C_α positions of residues $i+1$ and i in the PDB structure, and $|\mathbf{r}^{\mathrm{n}}_{\mathrm{SC},i}-\mathbf{r}^{\mathrm{n}}_{\alpha,i}|$ is the reference distance between the PDB position of the C_α atom of residue i and the centroid of its SC determined from the PDB coordinates of all nonhydrogen atoms belonging to the given SC. Thus, in contrast to the constant (= 3.8 Å) C_α-SC reference distance in Ref. [136], the present C$_\alpha$-SC reference distances are residue-type dependent and can vary among residues of the same type depending on the PDB structure. The bond-angle term is given by

$$\begin{aligned} E_{\rm angle} &= K_\theta \left\{ (\theta_{13} - \theta_{13}^{\rm n})^2 + (\theta_{n2} - \theta_{n2}^{\rm n})^2 + \sum_{i=2}^{n-1} \sum_{k=1}^{3} (\theta_{ik} - \theta_{ik}^{\rm n})^2 \right\} + \\ &\quad + \sum_{i=2}^{n-2} \left\{ K_\phi^{(1)} \left[1 - \cos(\phi_i - \phi_i^{\rm n}) \right] + K_\phi^{(3)} \left[1 - \cos 3(\phi_i - \phi_i^{\rm n}) \right] \right\} + \\ &\quad + K_{\rm ch} \sum_{i=2}^{n-1} (\psi_i - \psi_i^{\rm n})^2 \ . \end{aligned} \tag{4}$$

As before, symbols with the superscript “n” denote the PDB values of the corresponding variables. The bond angles θ_{i1}, θ_{i2}, and θ_{i3} at the C$_\alpha$ position of residue i are those defined by the positions $\{\mathbf{r}_{\alpha,i-1}, \mathbf{r}_{\alpha,i}\mathbf{r}_{\alpha,i+1}\}$, $\{\mathbf{r}_{\alpha,i-1}, \mathbf{r}_{\alpha,i}, \mathbf{r}_{{\rm SC},i}\}$, and $\{\mathbf{r}_{{\rm SC},i}, \mathbf{r}_{\alpha,i}, \mathbf{r}_{\alpha,i+1}\}$, respectively (Fig. 1b). All three angles are defined for residue 2 through residue $n-1$, whereas only one of the three angles is defined for each of the residues at the two chain ends ($i = 1$ or n). We use a spring constant $K_\theta = 30.0$ kcal mol^{-1}rad^{-2}, which is numerically equal to the spring constant K_A for an identical bond angle term in O’Brien et al. [82]. The mainchain torsion angle ϕ_i is that defined by $\{\mathbf{r}_{\alpha,i-1}, \mathbf{r}_{\alpha,i}, \mathbf{r}_{\alpha,i+1}, \mathbf{r}_{\alpha,i+2}\}$ (Fig. 1b). Here we use the same form of the potential energy term for ϕ_i as that in Refs. [70, 71] (with $K_\phi^{(1)} = 2K_\phi^{(3)}$), which is formally different from the dihedral potential in O’Brien et al. [82]. Nonetheless, we adopt force constants $K_\phi^{(1)} = 0.7$ kcal mol^{-1} and $K_\phi^{(3)} = 0.35$ kcal mol^{-1} that are numerically equal, respectively, to the dihedral force constants K_{D1} and K_{D3} in Ref. [82]. The last summation in the above equation, which takes the same form as that in Ref. [82], is for enforcing chirality by penalizing deviations of the improper dihedral angle ψ_i at residue i from its PDB value $\psi_i^{\rm n}$. Here ψ_i is the angle between the plane defined by $\{\mathbf{r}_{\alpha,i-1}, \mathbf{r}_{\alpha,i}, \mathbf{r}_{\alpha,i+1}\}$ and the plane defined by $\{\mathbf{r}_{\alpha,i-1}, \mathbf{r}_{{\rm SC},i}, \mathbf{r}_{\alpha,i+1}\}$ (Fig. 1b), i.e., $\psi_i = \cos^{-1}\{[(\mathbf{r}_{\alpha,i-1} - \mathbf{r}_{\alpha,i}) \times (\mathbf{r}_{\alpha,i+1} - \mathbf{r}_{\alpha,i-1})/|\mathbf{r}_{\alpha,i-1} - \mathbf{r}_{\alpha,i}||\mathbf{r}_{\alpha,i+1} - \mathbf{r}_{\alpha,i-1}|] \cdot [(\mathbf{r}_{\alpha,i+1} - \mathbf{r}_{\alpha,i-1}) \times (\mathbf{r}_{{\rm SC},i} - \mathbf{r}_{\alpha,i+1})/|\mathbf{r}_{\alpha,i+1} - \mathbf{r}_{\alpha,i-1}||\mathbf{r}_{{\rm SC},i} - \mathbf{r}_{\alpha,i+1}|]\}$. We use $K_{\rm ch} = 18.0$ kcal mol^{-1}rad^{-2} for this term. This value is comparable to that in the coarse-grained protein chain model of Takada et al. with explicit backbone atoms [56] and the lower values among the force constants for improper dihedral angles in the atomic model of Neria et al. [138]. We have also attempted to use the $K_{\rm ch}$ value of 18.013 kcal mol^{-1} degree^{-2} ($= 5.9 \times 10^4$ kcal mol^{-1}rad^{-2}) in O’Brien et al. [82]; but in that case the resulting energy associated with ψ was too large for our simulation to behave properly.

Following Ref. [82], we include an E_{HB} term [Eq. (2)], identical to theirs, to account approximately for the effects of native backbone hydrogen bonding:

$$E_{HB} = \epsilon_{HB} \sum_{(i,j)\in\{\mathrm{HB}\}} \left[\left(\frac{r^{\mathrm{n}}_{\alpha,ij}}{r_{\alpha,ij}} \right)^{12} - 2 \left(\frac{r^{\mathrm{n}}_{\alpha,ij}}{r_{\alpha,ij}} \right)^{6} \right] , \tag{5}$$

where the summation is only over the set {HB} of residue pairs i, j for which residue i and residue j are identified by the STRIDE algorithm [139] to be connected by backbone hydrogen bonding in the PDB structure of the given protein, $r_{\alpha,ij} \equiv |\mathbf{r}_{\alpha,j} - \mathbf{r}_{\alpha,i}|$ is the distance between the C_α positions of residues i and j, and we use the same energy scale $\epsilon_{HB} = 0.75$ kcal mol^{-1} for this term as in Ref. [82].

The E_{c} term in Eq. (2) accounts for native nonbonded interactions. It takes the form

$$\begin{aligned} E_{\mathrm{c}} \;\; = \;\; & \sum_{(i,j)\in\{\mathrm{NSC}\}} \Bigg\{ \epsilon^{\alpha-\alpha} \left[\left(\frac{r^{\mathrm{n}}_{\alpha,ij}}{r_{\alpha,ij}} \right)^{12} - 2 \left(\frac{r^{\mathrm{n}}_{\alpha,ij}}{r_{\alpha,ij}} \right)^{6} \right] + \epsilon^{\alpha-\mathrm{SC}} \left[\left(\frac{r^{\mathrm{n}}_{\alpha-\mathrm{SC},ij}}{r_{\alpha-\mathrm{SC},ij}} \right)^{12} - 2 \left(\frac{r^{\mathrm{n}}_{\alpha-\mathrm{SC},ij}}{r_{\alpha-\mathrm{SC},ij}} \right)^{6} \right] \\ & + \epsilon^{\mathrm{SC-SC}}_{ij} \left[\left(\frac{r^{\mathrm{n}}_{\mathrm{SC-SC},ij}}{r_{\mathrm{SC-SC},ij}} \right)^{12} - 2 \left(\frac{r^{\mathrm{n}}_{\mathrm{SC-SC},ij}}{r_{\mathrm{SC-SC},ij}} \right)^{6} \right] \Bigg\} , \end{aligned} \tag{6}$$

which is similar but not identical to the E^{N}_{NB} term in O'Brien et al. [82]. In this expression, the summation is only over the native contact set {NSC} containing residue pairs i, j for which at least two nonhydrogen atoms, one belonging to residue i and the other belonging to residue j, are less than 4.5 Å apart in the PDB structure. The summation is further restricted to interaction sites (C_α or SC beads) that are separated by four or more virtual bonds in the primary structure of the chain (thus SCs of sequentially adjacent residues do not interact via this term). Here $r_{\alpha,ij}$ is C_α-C_α distance as defined above; $r_{\alpha-\mathrm{SC},ij} \equiv |\mathbf{r}_{\mathrm{SC},i} - \mathbf{r}_{\alpha,j}|$ or $|\mathbf{r}_{\alpha,i} - \mathbf{r}_{\mathrm{SC},j}|$ denotes one of the two C_α-SC distances between residues i and j (terms for both instances are included in the summation); and $r_{\mathrm{SC-SC},ij} \equiv |\mathbf{r}_{\mathrm{SC},i} - \mathbf{r}_{\mathrm{SC},j}|$ is the SC-SC distance between residues i and j. The interaction parameters for C_α-C_α and C_α-SC are $\epsilon^{\alpha-\alpha} = 0.5$ kcal mol^{-1} and $\epsilon^{\alpha-\mathrm{SC}} = 0.37$ kcal mol^{-1}, respectively (both independent of i, j); whereas the SC-SC interaction parameters $\epsilon^{\mathrm{SC-SC}}_{ij}$ are dependent on the residue types of i and j. Following O'Brien et al., these native-centric SC-SC energies are taken to be proportional to a set of shifted energies based upon a statistical potential obtained by Miyazawa and Jernigan [140] (i.e., $\epsilon^{\mathrm{SC-SC}}_{ij} = 0.7(\Delta\epsilon_{ij} - 1.2)$ kcal mol^{-1} as for the C_α-SCM for Protein L in Ref. [82], where $\Delta\epsilon_{ij}$ is the statistical potential given in Table V of Ref. [140]). The formulation of our E_{c} term is essentially identical to that of the E^{N}_{NB} term in O'Brien et al. except that E_{c} contains favorable C_α-C_α interactions but such interactions are absent in their formulation.

Finally, the nonnative, nonbonded E^{NN}_{NB} term in Eq. (2) provides excluded-volume repulsion between interaction sites that do not belong to the native contact set:

$$E_{NB}^{NN} = \epsilon^{NN} \sum_{(i,j)\notin\{\mathrm{NSC}\}} \left[\left(\frac{2r_\alpha^{\mathrm{rep}}}{r_{\alpha,ij}} \right)^{12} + \left(\frac{r_\alpha^{\mathrm{rep}} + r_j^{\mathrm{vdW}}}{|\mathbf{r}_{\alpha,i} - \mathbf{r}_{\mathrm{SC},j}|} \right)^{12} + \left(\frac{r_\alpha^{\mathrm{rep}} + r_i^{\mathrm{vdW}}}{|\mathbf{r}_{\mathrm{SC},i} - \mathbf{r}_{\alpha,j}|} \right)^{12} + \left(\frac{r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}}}{r_{\mathrm{SC-SC},ij}} \right)^{12} \right], \quad (7)$$

where the summation is restricted to residue pairs i, j that do not belong to the native contact set and also pairs of interaction sites that are four or more bonds apart. The length scales we adopt for these repulsive interactions are equivalent to that in O'Brien et al. [82]: $r_\alpha^{\mathrm{rep}} = 1.37$ Å for C_α and the r_i^{vdW}s are SC van der Waals radii that depend on the amino acid type of residue i as given in Table S2 of Ref. [82]. Here, a relative small $\epsilon^{NN} = 0.01$ kcal mol^{-1} is used to minimize unphysically harsh steric clashes. The ratios between this ϵ^{NN} value we have chosen and the above energy scales $\epsilon^{\alpha-\alpha}$, $\epsilon^{\alpha-\mathrm{SC}}$, and $\epsilon_{ij}^{\mathrm{SC-SC}}$ for the native-centric interactions are comparable to the corresponding $(0.7)^{12} = 0.014$ ratio in previous coarse-grained models of Cheung et al. [141] and Azia and Levy [128]; but the present $\epsilon^{NN} = 0.01$ kcal mol^{-1} is ten orders of magnitude larger than the corresponding energy scale $\epsilon_i^{NN} = 10^{-12}$ kcal mol^{-1} for the nonnative repulsion term E_{NB}^{NN} in Ref. [82].

Following O'Brien et al. [82], the transfer free energy term $\Delta G_{\mathrm{trf}}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\}, C)$ in Eq. (1) takes the following form:

$$\Delta G_{\mathrm{trf}}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\}, C) = \sum_{i=1}^{n} \left[\delta g_{\mathrm{trf}}^{\alpha}(C) \left(\frac{\mathrm{SASA}^{\alpha}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\})}{\mathrm{SASA}^{\alpha}_{\mathrm{Gly}-\alpha-\mathrm{Gly}}} \right) + \delta g_{\mathrm{trf},t(i)}^{\mathrm{SC}}(C) \left(\frac{\mathrm{SASA}^{\mathrm{SC}}_{t(i)}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\})}{\mathrm{SASA}^{\mathrm{SC}}_{t(i),\mathrm{Gly}-t(i)-\mathrm{Gly}}} \right) \right], \quad (8)$$

where the summation is over all the residues (labeled by i) in the given protein, $t(i)$ is the amino acid type of residue i, $\delta g_{\mathrm{trf}}^{\alpha}(C)$ and $\delta g_{\mathrm{trf},t(i)}^{\mathrm{SC}}(C)$ are, respectively, the reference free energies of transfer of the polypeptide backbone and of the sidechain of amino acid type $t(i)$, embedded in the tripeptide Gly-$t(i)$-Gly, from an aqueous environment with zero denaturant to one with denaturant concentration C. $\mathrm{SASA}^{\alpha}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\})$ and $\mathrm{SASA}^{\mathrm{SC}}_{t(i)}(\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\})$ are the solvent accessible surface areas [142], respectively, of the backbone and the sidechain i [of type $t(i)$] in the conformation specified by the coordinates $\{\mathbf{r}_\mathrm{b}\}, \{\mathbf{r}_{\mathrm{SC}}\}$, whereas $\mathrm{SASA}^{\alpha}_{\mathrm{Gly}-\alpha-\mathrm{Gly}}$ and $\mathrm{SASA}^{\mathrm{SC}}_{t(i),\mathrm{Gly}-t(i)-\mathrm{Gly}}$ are the corresponding reference solvent accessible surface areas when an amino acid residue of type $t(i)$ is embedded in the tripeptide Gly-$t(i)$-Gly. In our simulations, solvent accessible surface areas and their derivatives with respect to Cartesian coordinates (the latter are needed for Langvein dynamics simulations, see below) are computed using the TINKER software [143], which is partly based upon the methodology developed in Refs. [144, 145]. We use the same expressions

$$\delta g^{\alpha}_{\rm trf}(C) = m^{\alpha}C + b^{\alpha} \tag{9}$$

$$\delta g^{\rm SC}_{{\rm trf},t(i)}(C) = m^{\rm SC}_{t(i)}C + b^{\rm SC}_{t(i)} \tag{10}$$

for the reference C-dependent transfer free energies and the same m and b parameters as those in O'Brien et al. Specifically, the present m^{α}, b^{α}, $m^{\rm SC}_{t(i)}$, and $b^{\rm SC}_{t(i)}$ are equivalent, respectively, to their m_{BB}, b_{BB}, m_k, and b_k parameters [where $k = t(i)$ labels amino acid type]; the numerical values of which are provided in Table S3 of Ref. [82]. The b^{α} and $b^{\rm SC}_{t(i)}$ parameters for urea are zero, which is expected if the $\delta g_{\rm trf}$s are indeed linear in denaturant concentration. However, the corresponding b parameters for guanidinium chloride (GdmCl) are nonzero [82]. This peculiar behavior is a reflection of the ionic nature of GdmHCl, which leads to the well-known property that [GdmHCl]-dependent transfer free energies do not extrapolate to the origin [146].

Three related models based upon the above SC formulation are considered in the present study. We refer to them as the SC-Gō, SC-urea and SC-GdmHCl models. The potential energy of the SC-Gō model is given entirely by the $(E_{\rm T})_{\rm SC-G\bar{o}}$ function in Eq. (2), whereas the potential energy functions of the SC-urea and SC-GdmHCl models are given by the expression $E_{\rm T} = \sigma(E_{\rm T})_{\rm SC-G\bar{o}} + \Delta G_{\rm trf}$ in Eq. (1) wherein the $\Delta G_{\rm trf}$ term is specified to account for urea and GdmHCl dependence respectively. In our SC-urea and SC-GdmHCl models, σ is adjusted so that the denaturant concentration at the folding-unfolding transition midpoint (at which point the folded and unfolded populations are equal) coincides with that determined from experiment. The tuning of σ is similar to the method of adjusting the simulation temperature in the recent works of Thirumalai and coworkers [82, 83]. The difference between the two approaches is minimal but they are not identical because Langevin dynamics (see below) is temperature dependent.

As for the Cα-Gō and db models described above, Langevin dynamics [70, 71, 147] is used for folding kinetics simulations and thermodynamic sampling in the SC models. The simulation is based on a set of equations of the form $m\dot{v}(t) = F_{\rm conf}(t) - m\gamma v(t) + \eta(t)$ (one equation for each Cartesian coordinate in $\{\mathbf{r}_{\rm b}\}, \{\mathbf{r}_{\rm SC}\}$), where m, v, $\dot{v}$, $F_{\rm conf}$, γ, η, and t are mass, velocity, acceleration, conformational force (including SASA-dependent forces in the SC-urea and SC-GdmHCl model), friction (viscosity) coefficient, random force, and time, respectively. The random force autocorrelation function $\langle\eta(t)\eta(t')\rangle = 2m\gamma k_{\rm B}T\delta(t-t')$, where $k_{\rm B}$ is Boltzmann constant, T is absolute temperature, and δ here denotes the Dirac delta function. Following the prescription of Veitshans et al. [147], we use an integration time step $\delta t = 15$ fs (1.5×10^{-14}s). In the results reported below, time is reported in units of integration time steps. The mass m for each bead is taken to be 2×10^{-22}g (120 g mol^{-1}) here, which is similar to the corresponding mass of 1.8×10^{-22}g per bead used in the two-bead SC model of Liu et al. [83] and comparable to the mass of 3×10^{-22}g per residue used in the C$_\alpha$ model of Veitshans et al. [147]. In view of an earlier test conducted by our group indicating that shapes of simulated chevron plots are insensitive to variation of the frictional coefficient over more than three orders of magnitude [30], here we have adopted a low frictional coefficient [147] $\gamma = 1.65 \times 10^{10}\text{s}^{-1}$ ($m\gamma = 3.3 \times 10^{-12}\text{gs}^{-1}$) to enhance computational efficiency as in our previous studies [70, 71].

A measure of local-nonlocal coupling. We define a local-nonlocal coupling [53,59] parameter $C_{\rm lnl}$ to quantify the degree to which two local chain segments centered respectively around each of two residues in a contact pair adopt local conformations consistent with those in the native structure. For a given native contact between residues i and j, we compute the root-mean-square deviation (rmsd) values [148] of the two 5-residue segments centered around each of the contacting residues (i.e., residues $i-2$, $i-1$, i, $i+1$, $i+2$ and residues $j-2$, $j-1$, j, $j+1$, $j+2$) from their respective local conformations in the native PDB structure. If both rmsd values are < 0.8 Å, we set $C_{\rm lnl} = 1$; otherwise we set $C_{\rm lnl} = 0$ (e.g., $C_{\rm lnl} = 1$ and 0, respectively, for the configurations in Fig. 1c and Fig. 1d). If one or both of the contacting residues is/are at or near the chain ends, the chain segment(s) considered is/are shortened by one or two residues accordingly (for i or $j = 1$, the chain segment considered consists of residues 1, 2 and 3; for i or $j = 2$, the chain segment considered consists of residues 1, 2, 3, and 4; similar rules apply to i or $j = n - 1$ or n).

Results and Discussion

To gain insight into the db and SC effects on local-nonlocal coupling and chevron behaviors, we apply — as a case study — the above model formulations to the IgG-binding B1 domain of Protein L from *Peptostreptococcus magnus* (PDB id: 1HZ6; residues 1–64) [149]. This domain (referred to as Protein L below) is a two-state folder [150]. One of our reasons for choosing Protein L is to compare our results with those of O'Brien et al. because it is one of the two proteins that have been investigated using their SC model [82]. To conduct a comparison of our SC Protein L models among themselves and with that of O'Brien et al. on as close to an equal footing as possible, we use a uniform $\sigma = 0.398$ [Eq. (1)] for all three of our SC models throughout this work with simulation temperatures $T_{\rm s} = 331$ K for the SC-urea model and $T_{\rm s} = 338$ K for the SC-GdmCl model.

Thermodynamic stability of the native state is modulated by concentration C of urea and of GdmCl, respectively, in these models; whereas native stability is modulated only by temperature in the SC-Gō model, for which the midpoint temperature is 356 K. The parameter choices for our SC models lead to midpoint urea and GdmCl concentrations that are essentially identical to the corresponding midpoint concentrations in the model of O'Brien et al., and simulation temperatures that are very similar to their $T_{\rm S} = 328$ K [82].

Energetics of local-nonlocal coupling: An example. We begin by addressing how local-nonlocal coupling arises in the models by comparing the energetic favorabilities of the example conformations in Fig. 2 for two pairs of 5-residue segments. Both pairs contain a native contact between the center residues of the two segments. However, one of the pairs satisfies local-nonlocal coupling ($C_{\rm lnl} = 1$, Fig. 2a and c) whereas the other does not ($C_{\rm lnl} = 0$, Fig. 2b and d). Using the potential energy function $(E_{\rm T})_{\rm SC-G\bar{o}}$ for the SC-Gō model in Eq. (2), we analyzed the total interaction energies between the segments and the contributions from individual terms belonging to $(E_{\rm T})_{\rm SC-G\bar{o}}$. We consider below three different methods of comparing the segment pairs' interaction energies. In all instances considered, the local conformations of the segment pair satisfying local-nonlocal coupling consistently lead to a lower total interaction energy within the pair than the local conformations of the segment pair not satisfying the aforementioned local-nonlocal coupling criterion.

We first consider the interaction energies between the two rigid segment pairs in isolation. In this comparison, only the 20 beads (backbone and SC positions) within each of the two segment pairs are compared, and all the C_α (backbone) and SC positions are fixed as shown in Fig. 2c and d. In this case, the total interaction energies within the segment pairs for the configurations in (c) and (d) are 9.54 and 12.26 kcal mol^{-1}, respectively. Thus, the energy difference between the segment-pair configuration in Fig. 2 that satisfies local-nonlocal coupling (c) and the one that does not (d) is $\Delta E = -2.72$ kcal mol^{-1}. Accordingly, the ratio of Boltzmann populations of the two configurations [population of (c) divided by that of (d)] at the simulation temperature $T = T_{\rm s} = 356$ K is $\exp(-\Delta E/k_{\rm B}T_{\rm s}) = 46.6$, indicating that configuration (c) is strongly favored over configuration (d). Part of ΔE is a consequence of the more favorable inter-segment nonbonded interactions in configuration (c) than those in configuration (d). Their contributions to $E_{\rm c}$, which depend on the distances between various interaction sites in the two configurations (Fig. 2e and f), are -0.81 and -0.19 kcal mol^{-1}, respectively, accounting for -0.62 kcal mol^{-1} in the overall energy difference ΔE. A major part of the large ΔE, however, is attributable to the terms in $E_{\rm bond}$ and $E_{\rm angle}$. For instance, the energies for the C_α-SC bond length, backbone bond angle, backbone dihedral angle, and improper dihedral angle terms in (c) are 3.57, 0.56, 0.048, and 0.84 kcal mol^{-1}, respectively, which are all significantly lower than the corresponding energies of 5.35, 1.48, 0.46, and 1.82 kcal mol^{-1} in (d). This comparison illustrates that more stable inter-segment nonbonded interactions as well as less strain in bond lengths, angles, and dihedral angles in configurations that satisfy local-nonlocal coupling relative to configurations that do not can all contribute to the tendency for the former to be energetically favored over the latter.

To assess the robustness of the above-observed favorability of local configurations consistent with local-nonlocal coupling, we conduct another comparison in which the SCs in configurations (c) and (d) in Fig. 2 are allowed to move while keeping only the backbone positions fixed. The resulting Boltzmann-averaged energy within the segment pairs for configurations (c) and (d) are 8.79 and 10.92 kcal mol^{-1}, respectively, leading to $\Delta E = -2.13$ kcal mol^{-1} and a [(c)/(d)] population ratio (defined above) that equals to $\exp(-\Delta E/k_{\rm B}T_{\rm s}) = 20.3$. Thus, on average, the mainchain configuration in (c) remains significantly favored over the mainchain configuration in (d). Because the SCs can now sample energetically more favorable positions, the differences in average bond-length, bond-angle and dihedral-angle energies here between the two configurations are reduced significantly from the corresponding energy differences in the preceding paragraph.

We also explore the interplay between local-nonlocal coupling of a local configuration and the rest of the protein conformation. To this end, we compare the average interaction energies within the two segments as in the preceding paragraph (SC positions allowed to vary); but instead of considering the two segments in isolation, the conformational sampling now also takes into account the interactions between the segment pairs and the rest of the conformations in which the two segment pairs are embedded (Fig. 2a and b). In other words, the Boltzmann averaging is now weighted also by the interactions between the segments and the rest of the protein chain. This computation results in averaged energies of 7.18 and 9.25 kcal mol^{-1}, respectively, for the mainchain configurations in Fig. 2c and d. It follows that $\Delta E = -2.07$ kcal mol^{-1} and the above-defined population fraction $\exp(-\Delta E/k_{\rm B}T_{\rm s}) = 18.6$, which is approximately equal to the corresponding population fraction in the preceding paragraph. Thus, in this example, the mainchain configuration in Fig. 2c that satisfies our local-nonlocal coupling criterion is consistently more favored over the mainchain configuration in Fig. 2d irrespective of whether their interactions with other parts of the chain conformations in Fig. 2a and b are taken into consideration.

Desolvation-barrier effects. The mainchain conformations of the segment pairs in Fig. 2c and d are also a good illustration of how db effects are conducive to local-nonlocal coupling (Fig. 3). In this example, the C$_\alpha$ positions of the center contacting residues in configuration (b) (Fig. 3b) that does not satisfy local-nonlocal coupling are a bit farther apart than the center C_α positions in configuration (a) (Fig. 3a) that satisfies local-nonlocal coupling. Because the bond-length and bond-angle terms of configurations (a) and (b) are essentially identical in this example, the energy difference ΔE between the two configurations is determined essentially by the difference in nonbonded interaction energy. For the Cα-Gō model in which db effects are absent, the nonbonded inter-segment interaction energies of the two configurations do not differ by too much (black arrow in Fig. 3c and blue arrow in Fig. 3d). However, in the db model in which the finite size of water molecules is taken into consideration, configuration (b) is strongly penalized because water is excluded in between the two contacting residues, leading to a much higher (unfavorable) energy (black arrow in Fig. 3d). Accordingly, the contribution to inter-segment nonbonded energy $E_{\rm c}$ from configuration (a) minus that from configuration (b) in Fig. 3 is $-0.98\epsilon - (-0.57\epsilon) = -0.41\epsilon$ for the C_α-Gō model but is $-0.98\epsilon - (+0.098\epsilon)$ $= -1.08\epsilon$ for the db model. In other words, by this inter-segment nonbonded term alone, the favorability of configuration (a) over configuration (b) is increased in the db model relative to the C_α-Gō model by $-1.08\epsilon - (-0.41\epsilon) = -0.67\epsilon$, where a more negative value in this difference means a stronger db-induced discrimination between configurations (a) and (b). When intra-segment $E_{\rm c}$ energies are also taken into account, the total $E_{\rm c}$ energy within the two segments in configuration (a) minus that in configuration (b) is $-1.87\epsilon - (-1.19\epsilon) = -0.68$ for the C_α-Gō model but is $-1.87\epsilon - 0.10\epsilon = -1.97\epsilon$ for the db model, leading to an even stronger db-induced discrimination between configurations (a) and (b) $[-1.97\epsilon - (-0.68\epsilon) = -1.29\epsilon]$. This comparison underscores a key consequence of db, which is that looser configurations of native nonlocal contacts that allow for more local conformational freedom — and hence weaker local-nonlocal coupling — are strongly penalized by a significantly shorter spatial range of attractive native interaction entailed by the db [76]. Deviations of sequentially local conformations are also strongly penalized by intra-segment db as well. All in all, the db-induced narrowing of the range of attractive interactions implies that even small deviations from native packing are strongly discouraged, thus leading to a more cooperative, or all-or-none-like transition, between the folded and unfolded states of a protein [76]. More evidence for this general trend is offered below.

Folding cooperativity and local-nonlocal coupling: Comparing models. The free energy profiles of our Protein L models simulated at their respective transition midpoint as functions of the progress variable Q are provided in Fig. 4a. Consistent with above considerations, SC and db effects enhance folding cooperativity. Among the models tested here, the db model appears to be thermodynamically most cooperative in that it has the highest free energy barrier ($\sim 9k_{\rm B}T$ at mid-Q) between the unfolded (low-Q) and the folded (high-Q) minima. The corresponding barriers in the SC models are not as high (at $\sim 4-6k_{\rm B}T$), but are nevertheless significantly higher than the free energy barrier for the C$_\alpha$-Gō model ($\sim 2k_{\rm B}T$), confirming the expected enhancement of folding cooperativity by SC packing. Notably, the SC-urea and SC-GdmCl models are slightly more cooperative (with slightly higher free energy barriers) than the SC-Gō model, suggesting that the SASA-dependent interaction potentials likely contain cooperative features that are absent in the simpler potential function for the SC-Gō model.

Figure 4b shows the degree of local-nonlocal coupling of the conformations sampled under the same modeling conditions as those in Fig. 4a. As indicated above, the local-nonlocal coupling parameter $C_{\rm lnl}$ measures the tendency of local conformations centered around a native contact to also adopt nativelike local conformations. It follows that, by definition, $C_{\rm lnl} = 1$ at $Q = 1$ for all models. However, the dependence of $C_{\rm lnl}$ on Q for $Q < 1$ can be different in different models. For the least cooperative C$_\alpha$-Gō model, the average $\langle C_{\rm lnl}\rangle$ varies in a roughly linear manner with the number of native contacts. In this case, the slope of $\langle C_{\rm lnl}\rangle$ vs. Q is approximately unity. In comparison, the db model exhibits higher local-nonlocal coupling over the entire range of $Q < 1$ (green curve). The highest degree of local-nonlocal coupling is seen for the SC models, with $\langle C_{\rm lnl}\rangle > 0.4$ for Q values as low as 0.1. In other words, close to one half of the native contacts that exist in the highly open conformations in the SC models are centered around two spatially adjacent segments of nativelike local conformations. Among the SC models, the SC-denaturant models (top two curves in Fig. 4b) exhibit a slightly higher local-nonlocal coupling than the SC-Gō model (blue curve). Taken together, the observations in Fig. 4b are consistent with our expectations that db and SC effects are conducive to local-nonlocal coupling. The quantitative trends obtained here serve to substantiate the above qualitative discussion of the two example segment-pair configurations analyzed in Figs. 2 and 3. Interestingly, the SC models entail higher local-nonlocal coupling than the db model (Fig. 4b) despite the lower cooperativity of the SC models relative to the db model (Fig. 4a). This result indicates that although local-nonlocal coupling is expected to be correlated with folding cooperativity, the physical origins of the two properties are not entirely identical. Thus, comparisons of $C_{\rm lnl}$ values between theory and experiment or between coarse-grained and atomic simulations should offer new physical understanding in addition to the insights gained by considerations of folding cooperativity [23, 35, 53].

Native-state fluctuations. A conspicuous difference between the db and other models in Fig. 4a is that the folded minimum for the db model is at $Q = 1$ whereas the folded minima for the other models are at a significantly lower $Q = 0.75 - 0.85$ with the folded minima of the SC models adopting a slightly lower Q values than that of the C_α-Gō model. This difference, which has been noted for the C_α and db models [31, 71], means that native-state packing in the db model is significantly tighter than that of the SC and C_α models. The situation is illustrated by the conformational drawings in Fig. 5. What degree of structural fluctuation in the native state is realistic? Ultimately this is a question that has to be settled by experiments; for example, by matching molecular dynamics simulation data with NMR measurements of bond vector dynamics [151]. As far as simulations are concerned, the mainchain rmsd at the native free energy minimum of the db model is ~ 1 Å (Fig. 5). Such level of structural fluctuation is in line with C_α rmsd values of ~ 1.3 Å observed in an explicit-solvent control simulation of ubiqutin in water by Alonso and Daggett [152]. This favorable comparison between the db model and atomic simulation suggests that the larger structural fluctuations characterized by the ~ 2 Å mainchain rmsd values in Fig. 5 for the SC and C_α-Gō models may be too large, in general, to mimick behaviors of cooperatively folding proteins in atomic simulations. In the recent atomic folding simulations of 12 proteins by Lindorff-Larsen et al., the trajectories of chignolin, villin, WW domain, NTL9, and protein G indicate that their native-state C_α rmsd is ~ 1 Å, which is again in line with the trend seen in our coarse-grained db model. However, their simulated trajectories for trp-cage, BBA, BBL, protein B, homeodomain, α3D, and λ-repressor show native states that are much more loosely packed with C_α rmsd value of ~ 4 Å or more [100].

Route measure. A simple yet informative characterization of the topography of energy landscapes is provided by the route measure [153]

$$R(Q) = \frac{1}{\tilde{Q}_{\rm n} Q(1-Q)} \sum_{k=1}^{\tilde{Q}_{\rm n}} [P(c_k|Q) - Q]^2 \; , \tag{11}$$

where $P(c_k|Q)$ is the probability of native contact c_k among conformations that have a given Q value. Here $\tilde{Q}_{\rm n} = 137$ is the total number of contacts in the native contact set of every one of our Protein L models. It follows from this definition that $0 \leq R(Q) \leq 1$. As discussed previously, larger $R(Q)$ means that the folding routes at a given Q value are more channeled [133, 153]. Figure 6 provides a comparison of the route measures of our Protein L models, which are all computed under the same midpoint conditions as those in Fig. 4. The main message from Fig. 6 is that the cooperativity-enhancing db and SC effects also lead to more route channeling. The $R(Q)$ function for the C_α-Gō model here is similar to the $R(Q)$ functions for other C_α-Gō models for small, single-domain proteins [75, 133]. In comparison, the $R(Q)$ function for the db model is generally higher, as has been seen in the study of db models of other proteins [75]. Interestingly, SC effects lead to even more channeled routes at low Q values; but $R(Q)$ for SC models falls below that of the db model at intermediate and high Q values. Echoing the trends in folding cooperativity (Fig. 4a) and local-nonlocal coupling (Fig. 4b), the folding routes of SC-urea and SC-GdmCl models are slightly more channeled than that of the SC-Gō model. The $R(Q)$ behavior of the db model is quite unique in that significant route channeling is observed at $Q \sim 0.9$ when the protein structure is very close to being fully native. A similar double-hump feature of the $R(Q)$ function has also been observed in other db models [75]. It will be instructive to elucidate how db effects can give rise to this peculiar property and to ascertain whether similar behaviors can be observed experimentally.

Native stability and chevron behaviors. We next turn to the relationship between thermodynamic and kinetic properties of these models. The thermodynamic stability of the folded state relative to the unfolded state as a function of denaturant concentration is shown in Fig. 7a for the SC-urea and SC-GdmHCl models. The free energy of folding in units of $k_{\rm B}T$ is given here by $\Delta G_{\rm f}/k_{\rm B}T = -\ln[P(Q > Q_{\rm F})/P(Q < Q_{\rm U})]$, where the threshold Q values $Q_{\rm F}$ and $Q_{\rm U}$ are chosen to provide physically reasonable demarcations for the folded and unfolded states (Fig. 7). As stated above in Models and Method, the σ parameters governing the relative weights of the SC-Gō and denaturant-dependent components in these models [Eq. (1)] are tuned for a given midpoint denaturant concentration at which $\Delta G_{\rm f} = 0$. For this purpose, we use the experimental transition midpoint for GdmHCl [150] to set the σ value for our SC-GdmHCl model. Corresponding experimental data for urea, however, are not available. In the absence of experimental data, we use the midpoint urea concentration in the Protein L model of O'Brien et al. [82] to set the σ parameter for the present SC-urea model to facilitate comparison of our results with those from their model.

Figure 7a shows that the $\Delta G_{\rm f}/k_{\rm B}T$ values for the two models are different at zero denaturant. This feature of the models appears counter-intuitive but it is a consequence of the nonzero b^{α} and $b^{\rm SC}_{t(i)}$ parameters in Eqs. (9) and (10) for GdmHCl [82], which in turn is a reflection of the peculiar effects of GdmHCl discussed above [146]. One may consider an alternative formulation in which the σ parameters for the models are tuned to achieve the same zero-denaturant $\Delta G_{\rm f}$, but in that case the midpoints would be changed. Taking the models' simulation temperatures ($T_{\rm s}$) into consideration, Fig. 7a shows zero-denaturant folding free energy $\Delta G_{\rm f} \approx -6.9$ and -3.8 kcal mol^{-1}, respectively, for our SC-urea and SC-GdmHCl models at $C = 0$. These values are different from the corresponding values of -5.7 and -6.0 kcal mol^{-1} reported by O'Brien et al. [82]. Part of the difference might be related to their definition of fractional native population, which is not identical to ours. It is not straightforward, however, to compare the C-dependence in the two studies because their folding/unfolding transition curves provide fractional native population rather than ΔG_{f} as a function of C.

The chevron plots for the two models (Fig. 7b) are simulated by Langevin dynamics as described above. Each mean first passage time (MFPT) data point in Fig. 7b is determined by $\sim 200-800$ trajectories. The chevron plots for both the SC-urea and SC-GdmHCl models exhibit a mild rollover in the folding and unfolding arms. The severity of rollover may be measured by the deviation of the actual simulated chevron (data points) from the ideal two-state chevron consistent with folding thermodynamics (dotted V-shape). By this criterion, the SC-GdmHCl model shows a lesser degree of folding-arm rollover than that of the SC-urea model. This trend is possibly related to the much smaller midpoint C for the SC-GdmHCl model, which means that the range of C that the chevron folding arm needs to cover is much narrower in the SC-GdmHCl model than in the SC-urea model.

We have also compared the chevron plots for all the models considered in this study on the same footing. In Fig. 7c, the logarithmic folding and unfolding rate ln(rate) $= -\ln(\mathrm{MFPT})$ for every model is plotted as a function of the native stability $\Delta G_{\mathrm{f}}/k_{\mathrm{B}}T$ of that model. As observed before in models for other proteins [71], the rollover exhibited by the chevron plot for the C_α-Gō model (red data points) is much more severe than the other models because the C_α-Gō model is the least thermodynamically cooperative among the models we consider (Fig. 4a). In the $\Delta G_{\mathrm{f}}/k_{\mathrm{B}}T$ representation in Fig. 7c, the degrees of rollover for the db and the three SC models are quite similar. The SC-urea, SC-GdmHCl and SC-Gō results afford a useful comparison between model chevron plots constructed in terms of ln(rate) as a function of native stability $\Delta G_{\mathrm{f}}/k_{\mathrm{B}}T$ and those constructed as a function of denaturant concentration C as in experiments. Because of the complexities and uncertainties in modeling denaturant concentration directly and the significantly higher computational cost it entails, following the approach in early simulations of chevron plots for two-dimensional lattice models [5, 154], analyses of chevron plots of lattice [59, 81, 155–157] and continuum [26, 30, 53, 70, 71, 74, 81, 124] three-dimensional protein chain models by our group since 2002 have until now relied solely on using either intrachain interaction strength (e.g., Refs. [70,155]) or native stability [30,53,59,71,74,81] as proxy for denaturant concentration [124]. To assess the experimental relevance of this body of work critically, it is instructive to observe that the chevron plot for the SC-Gō model has only slightly more folding-arm rollover than the chevron plots for the SC-urea and SC-GdmHCl models. This trend is consistent [71] with the slightly higher thermodynamic cooperativity of the SC-urea and SC-GdmHCl models than the SC-Gō model. Because the interactions in the SC-urea and SC-GdmHCl models are based upon SASA and transfer free energies

(Fig. 4a), their slightly higher degrees of thermodynamic cooperativity probably originates from the nonadditive aspects of SASA-type interactions [158]. It should also be recognized that the variation in $\Delta G_{\rm f}/k_{\rm B}T$ in the present SC-Gō model is brought about by varying T rather than by varying the intrachain interaction strength while keeping T constant. Nonetheless, as verified by a test simulation conducted by our group (Fig. 9 in Ref. [81]), within a range of $\Delta G_{\rm f}/k_{\rm B}T$ values typically spanned by real proteins in denaturant solutions of various concentrations, the model chevron plot obtained by varying T is expected to differ little from that obtained by varying interaction strength. In view of all of the above considerations, the similarity among the chevron plots of the SC-Gō, SC-urea and SC-GdmHCl models in Fig. 7c suggests quite convincingly that using $\Delta G_{\rm f}/k_{\rm B}T$ as proxy for denaturant concentration to model chevron plots of thermodynamically two-state folders is a reasonably accurate and computationally efficient approach.

Internal friction. In view of the prevalence of rollovers in our model chevron plots, we delve deeper into the origin(s) of this behavior. As outlined above in Introduction, a possible underlying cause for chevron rollover is transiently populated intermediates [116–120]. For the SC-urea and SC-Gō models, the example trajectories in Fig. 8a,b show that folding kinetics is two-state-like around the folding-unfolding transition midpoint in that the unfolded part of the kinetic trajectory under midpoint conditions is concentrated in a relatively narrow range of Q values. However, intermediate conformations are transiently and significantly populated under strongly folding conditions (Fig. 8c). Instead of a sharp kinetic separation between a folded state narrowly distributed around $Q \sim 0.8$ and an unfolded state narrowly distributed around $Q \sim 0.15$ as in Fig. 8a,b, we see in Fig. 8c repeated excursions from the unfolded state to intermediate $Q \sim 0.4$ conformations that fail to proceed directly to the folded state, in many cases the chain returns to the unfolded state. This deviation from two-state-like kinetic transitions under strongly folding conditions is concomitant with a decrease in folding rate relative to the ideal folding rate for a hypothetical two-state transition, i.e., a folding-arm rollover (Fig. 7c). A similar connection between appreciable transient population of intermediate conformations and folding-arm rollover was established in a lattice Gō model [156] (see Fig. 4 of this reference) and is observed in the present C_α-Gō model as well (Fig. 9). In the case of our C_α-Gō model, even the trajectory under essentially midpoint conditions (Fig. 9a) appears less two-state-like — i.e., it has more broadly distributed Q values for the unfolded part of the trajectory — than the midpoint trajectories of the SC models in Fig. 8a,b. This trend is consistent with our observation above that the C_α-Gō model is less cooperative than the SC models.

In general, folding-arm rollover entails a stability-dependent pre-exponential front factor in the transition state picture, with the front factor for folding decreasing with increasing native stability [70, 155]. The front factor corresponds to an effective conformational diffusion coefficient, and is an inverse measure of internal friction [70, 155]. It follows that folding-arm rollover may be interpreted [70, 155] as an increase in internal friction of the protein chain [159, 160] as native stability increases. Because the SC-Gō model does not admit favorable nonnative interactions, no deep kinetic traps involving attractive nonnative contacts are expected. Nonetheless, mild kinetic traps can arise in Gō-like models because of topological hindrance to directly reaching the native structures from certain partially folded conformations, as has been illustrated in a lattice context [156]. Such a mechanism may underlie some of the transiently populated intermediates in Fig. 8; but alternatively they may well correspond merely to partially folded conformations that are not sufficiently stable to incorporate the unfolded part of the chain into a globally folded structure before the partially folded part unravels itself.

Hammond behavior and ground state effect. The above focus on internal friction and front factor emphasizes kinetic effects that cannot be deduced soley from the shape of a free energy profile along a progress variable. In contrast, other suggested causes of chevron rollover are formulated in terms of proposed shapes for a protein's free energy profile. In the latter formulations, attention is largely given to the possible changing position of the transition-state peak along the progress variable while tacitly assuming constancy of the front factor(s). In such perspectives, folding-arm and unfolding-arm chevron rollovers are caused, respectively, by a native-stability-dependent change in the difference in SASA between the transition state and the unfolded state and between the transition state and the folded state. (In contrast, these SASA differences are taken to be constant for a protein that possesses a chevron plot with linear fold and unfolding arms.) Hammond behavior is a classic, intuitive paradigm for shifting transition state position in any reaction. Conventional Hammond behavior stipulates that the structural difference between the reactant and transition states decreases when the free energy difference between the two states decreases [161]. For example, Hammond behavior in protein unfolding means that "the transition state moves closer to the folded state along the reaction coordinate as a result of destabilization of that state" [162]. Inasmuch as Hammond behavior is a cause for chevron rollovers, it can emerge in the context of free energy profiles with a broad transition-state plateau [163] or sequential barriers [121, 164]. In this context, it has been pointed out that chevron rollovers can also arise from "ground state effect," i.e., a shift in position of the unfolded or folded states instead of a shift in position of the transition state [121], or, more generally, a combination of Hammond behavior and ground state effect [164].

As far as our Q-dependent free energy profiles in Fig. 4a are concerned, neither broad transition-state plateaus nor sequential barriers are observed for our models for Protein L, though sequential barriers have been featured in db models for other proteins [124, 165]. Nonetheless, Hammond behavior is apparent (Fig. 10). As highlighted in Fig. 10b, the position of the putative transition state — which corresponds to the peak position of the free energy barrier — is seen moving toward smaller Q values when the folding conditions become stronger, i.e., when the unfolded state becomes more destabilized (left arm of the light blue region). In the same vein, the free energy peak moves toward larger Q values when the folded state becomes more destabilized (right arm of the light blue region). A modest ground state effect is also observed in Fig. 10b: the unfolded minimum is shifted slightly to higher Q values as the unfolded state is destabilized, whereas the folded minimum is shifted slightly to lower Q values as the folded state is destabilized.

To what extent does this Hammond behavior account for the chevron rollovers in our SC-urea model? If one assumes a transition state theory (TST) formulation with an essentially constant front factor, folding and unfolding rates would be given by an equation in the form of (rate) $\propto \exp(-\Delta G^{\ddagger}/k_{\mathrm{B}}T)$ and thus $\ln(\text{rate}) = -\Delta G^{\ddagger}/k_{\mathrm{B}}T + \text{constant}$, where $\Delta G^{\ddagger}$ is the barrier height. Here we compute $\Delta G^{\ddagger}$ as a population ratio while taking into account the shifting Q-position of the folding/unfolding barrier as urea concentration C is varied (Fig. 10b). We obtain a TST-predicted chevron plot by choosing the constant in the above equation for ln(rate) such that the TST-predicted rate coincides with the actual simulated rate at the transition midpoint. Comparison between this TST-predicted chevron plot (blue curves in Fig. 10a) and the directly simulated kinetic data for the same model (filled and open squares in Fig. 10a) indicates that the TST-Hammond picture does provide a general rationalization for the curved chevron arms from simulation and may even be nearly adequate as an explanation for the curvature of the unfolding arm. While these trends are noteworthy, the TST-Hammond picture by itself is quite far from providing a full account of the observed folding-arm rollover. Specifically, Fig. 10a shows that as C decreases, the simulated folding rate becomes significantly slower than the TST-predicted folding rate, suggesting strongly that a diminishing effective front factor, i.e., increasing internal friction as discussed above, is at play. Indeed, the insufficiency of a TST-Hammond account of folding-arm chevron rollover in our model is not surprising in view of the general limitations of the common TST picture for protein folding [166].

Comparison with experiment: Rollovers and asymmetric chevron arms. To ascertain the strength and limitation of the denaturant-dependent SC models, we compare the simulated chevron plot of our SC-GdmHCl model with that determined by experiment [150] (Fig. 11). It is clear that although the simulated chevron plot (brown data points) exhibits a general trend similar to the experimental chevron plot (black data points), there are notable deviations in three respects. First, the overall thermodynamic stability $\Delta G_{\rm f}$ of our SC-GdmHCl model at zero denaturant is -3.8 kcal mol^{-1} as noted above, which is lower than the corresponding experimental $\Delta G_{\rm f} \approx -4.7$ kcal mol^{-1} reported in Ref. [150]. Second, whereas the simulated chevron arms exhibit small but appreciable rollovers, the experimental chevron arms are linear. Third, whereas the folding and unfolding arms of the simulated chevron plot appear symmetric (making a symmetric V-shape), they are asymmetric in the experimental chevron plot (making a skew V-shape). More specifically, whereas the slopes of the folding and unfolding arms ($-m_{\rm f}$ and $m_{\rm u}$ respectively) of the ideal chevron (dotted V-shape in Fig. 11) obtained from fitting simulated logarithmic rates have essentially the same absolute value, viz., $m_{\rm f} = 0.88$ kcal mol^{-1}M^{-1} and $m_{\rm u} = 0.83$ kcal mol^{-1}M^{-1} ($m_{\rm f}/m_{\rm u} \approx 1.06$), the corresponding kinetic m values for the experimental chevron plot are $m_{\rm f} = 1.5$ kcal mol^{-1}M^{-1} and $m_{\rm u} = 0.5$ kcal mol^{-1}M^{-1}; and thus a large ratio of $m_{\rm f}/m_{\rm u} \approx 3.0$ between the folding and unfolding kinetic m values.

An evaluation of the physical implications of the above-noted discrepancies between simulation and experiment is in order. The discrepancy in native stability per se may not represent a fundamental shortcoming of the model, because it likely arises from the difference between the present simulation temperature $T_{\rm s} = 331$ K for the SC-GdmHCl model and the experimental temperature $T_{\rm expt} = 295$ K. A higher simulation temperature weakens the denaturant sensitivity of the transfer free energies. If this effect is taken into account, our simulation result should predict a stability $\Delta G_{\rm f} \approx -3.8(T_{\rm s}/T_{\rm expt}) = -4.3$ kcal mol^{-1} at $T_{\rm expt}$. This predicted stability is not too far from the experimental value of -4.7 kcal mol^{-1}.

The presence of mild rollovers in the model chevron plot suggests that the SC-GdmHCl model is not as cooperative as real Protein L. In other words, SC effects and SASA-dependent interactions as formulated in this model are likely not quite sufficient to fully capture the cooperativity of real two-state proteins. In this connection, it is noteworthy that in the recent study of the src SH3 domain by Liu et al. that used a viscosity coefficient that corresponds to the viscosity of real water, the zero-denaturant folding rate simulated using their SC model is 16-fold faster than the experimental folding rate [83]. This result suggests that their SC model for src SH3 is also less cooperative than the real protein. Nonetheless, because the rollovers in the present model chevron plot are mild, it is likely that further addition of physical features such as db effects and direction-dependent hydrogen bonding would be able to bring the cooperativity of the model up to the level of real two-state folders. In this regard, it is valuable to contrast our GdmHCl-dependent SC model chevron plot for Protein L with the chevron plot for src SH3 simulated by

Liu et al. that appears linear [83]. Compared to the number of trajectories we used to determine folding and unfolding rates (Figs. 7b and 12), their rates were estimated using a smaller number of trajectories because they employed a higher viscosity coefficient to mimick the aqueous environment (50 trajectories for $C = 0$ and 60 trajectories for each of the data points for other C values). The smaller number of trajectories lead to more scatter in their data. For instance, their relaxation rate at 0.5 M is appreciably lower than the fitted chevron curve, suggesting a possible onset of rollover, if not for the higher relaxation rate at 0 M [83]. In contrast, to ensure better convergence of our simulated rates, we opted for a lower viscosity coefficient, $\gamma = 0.0125$ [29, 70], which corresponds to a low friction regime [147]. It is possible that the mild chevron rollover we observed might disappear if the Langevin dynamics simulations were conducted at higher viscosity. However, as mentioned above, a prior test conducted by our group showed for one example that the overall shape of model chevron plots remains essentially unchanged over a wide range of γ (1.25×10^{-4} to 22.5) [30] that encompasses the $\gamma = 12.5$ value suggested to be roughly equivalent to that for water [147]. Therefore, we expect the mild chevron rollover in the present SC-GdmHCl model to be a robust feature even at higher viscosity. As such, it will be extremely instructive to elucidate the seemingly different behaviors of our SC model and the Liu et al. model [83] to ascertain, for example, whether the difference in chevron behavior originates from the different native topologies of the proteins and/or the differences in the folding and unfolding criteria as well as in other aspects of the two modeling setups.

Possible difference in the rate-limiting steps of folding and unfolding. The failure of the SC-GdmHCl model to capture asymmetric chevron plot observed in experiment may represent a fundamental limitation of this class of models. Indeed, all chevron plots obtained thus far from simulation of native-centric models [30, 53, 70, 71, 81, 83] and native-centric models augmented with sequence-dependent interactions [53, 124] are invariably symmetric, i.e., with $m_{\rm f} \approx m_{\rm u}$. This phenomenon applies also to the above-mentioned recent model chevron plot for src SH3 simulated by Liu et al. [83]. However, in the case of src SH3, the experimental chevron plot is also quite symmetric (though it is slightly less symmetric than the simulated chevron plot) [167] and thus questions of discrepancy did not arise.

The inability of common coarse-grained protein chain models to produce asymmetric chevron plots is a basic deficiency that needs to be addressed because asymmetric chevron plots are commonplace in experiments. In the extreme case of CspB, the unfolding arm exhibits hardly any dependence on denaturant concentration ($m_{\rm u} \approx 0$) [168]. These observations imply that, for many two-state proteins, the activation events for folding and unfolding have very different denaturant dependence. At a deeper level, this asymmetry is likely related to the asymmetry observed experimentally for temperature dependent folding and unfolding kinetics: Whereas folding rates are significantly non-Arrhenius, unfolding rates are often Arrhenius or nearly so. To our knowledge, this phenomenon was first identified as early as in the 1960s for bovine chymotrypsinogen A [169, 170], as discussed by Lumry and Biltonen (Fig. 1 in Ref. [171]). A similar behavior was also observed in 1984 by Segawa and Sugihara for hen lysozyme [172] and for several other proteins since then (see discussion in Refs. [74,157]). Because non-Arrhenius folding rates are most probably attributable to solvent-mediated interactions and principally to hydrophobic effects, the Arrhenius or near-Arrhenius behaviors of the unfolding rates suggest that the rate-limiting step of unfolding tends to be less dependent on solvent and thus is less sensitive to variation of denaturant concentration. This interpretation is consistent with an early theory stipulating that the rate-limiting event in protein denaturation is the disruption of the tight SC packing before solvent enters the protein core [78], a physical picture that was discussed more recently in terms of a “dry molten globule” state [173]. However, the fact that our SC-GdmHCl model cannot reproduce the experimentally observed chevron-plot asymmetry indicates the the model cannot fully capture the SC effects proposed in Ref. [78]. Inspired by these observations, a successful physics-based approach to modeling chevron asymmetry would likely involve a partial separation between the interactions for thermodynamic stability and the driving forces for folding kinetics, in a formulation that follows the same vein as that implemented for a class of rudimentary lattice models to address folding-unfolding asymmetry [157].

Incorporating nonnative interactions in the SASA-based models. All the models considered so far in the present study are essentially native-centric, or Gō-like. In the SC-denaturant (SC-urea and SC-GdmHCl) models, the transfer free energies themselves are general physical quantities that do not bias any particular folded structure and therefore allow for nonnative interactions. However, in the above formulation and in the formulations of Thirumalai and coworkers [82, 83], the transfer free energies serve almost exclusively to weaken interactions rather than strengthen interactions — native or otherwise — because nearly all m^{α} and $m^{\rm SC}_{t(i)}$ values are negative [82] and therefore, by Eqs. (8)–(10), promote increase in SASA, and thus unfolding, of the model protein. In this respect, there is not much substantial difference between the SC-denaturant models and the SC-Gō model (Fig. 7c). To address potentially unphysical biases in native-centric formulations, here we explore alternative formulations of the SC-denaturant models that better utilize the physical possibility of nonnative interactions entailed by the transfer free energies. Such alternative formulations should be useful in modeling specific nonnative interactions in folding that are evident from experiments [126] and may provide a more physical account of such interactions in the same vein as models that combine Gō-like and sequence-dependent components [124, 127–132]. As a simple example, we consider a class of alternative formulations for the SC-urea model. In essence, the formulation of the SC-urea model in Eqs. (1)–(10) uses a reference denaturant concentration $C_0 = 0$. In other words, the SC-urea model is equivalent to the SC-Gō model at $C = C_0 = 0$. But there is no a priori reason, even within the limitations of these models' basic setup, why this is the best representation of physical reality. Alternatively, one may choose another reference denaturant concentration $C_0 > 0$ such that the SC-urea model becomes the SC-Gō model at this C_0. In that case, the transfer free energy contribution provided by Eqs.(9) and (10) is modified to

$$\delta g^{\alpha}_{\rm trf}(C) = m^{\alpha}(C - C_0) + b^{\alpha} \tag{12}$$

$$\delta g^{\rm SC}_{{\rm trf},t(i)}(C) = m^{\rm SC}_{t(i)}(C - C_0) + b^{\rm SC}_{t(i)} \; , \tag{13}$$

and the σ parameter in Eq. (1) is re-adjusted accordingly to provide for the same midpoint as before. In this alternative formulation, because $(C - C_0)$ can be negative, the denaturant dependence is able to promote reduction in SASA and contact formation. As a result, certain nonnative contacts can be favored by the physico-chemical properties embodied in the transfer free energies.

We study three such alternative formulations for the SC-urea model by choosing different C_0 values by monitoring the nonnative interactions in these models and their chevron behavior (Fig. 12). As expected from previous simulations of chevron plots for models that allow for favorable nonnative interactions [53, 124], $C_0 > 0$ leads to increased folding-arm chevron rollover (Fig. 12a). Not surprisingly, the number of nonnative contacts increases with increasing C_0, but the increases are modest and are largely confined to the unfolded state under transition midpoint conditions (Fig. 7c). As has been observed in similar models [124, 174], nonnative interactions in our modified SC-urea models for Protein L also tend to lower the folding/unfolding free energy barrier (Fig. 7b). These results suggest that nonnative interactions in real proteins may be investigated in modified SC-denaturant formulations in a physical yet straightforward manner.

Concluding Remarks

By conducting a detailed comparison of the thermodynamic and kinetic properties of a set of native-centric coarse-grained protein chain models that embody physical effects of desolvation barrier, sidechain packing, and sequence- and denaturant-dependent transfer free energies (Figs. 1–3), we have gained several fundamental insights into the general principles of protein folding, and delineated some of the strengths and limitations of the different modeling approaches.

Both desolvation barrier and sidechain packing enhance folding cooperativity and local-nonlocal coupling (Figs. 2–4). They also entail more channeled energy landscapes (Fig. 6). The impact of sidechain effects on promoting local-nonlocal coupling is particularly prominent (Fig. 4); but the sidechain models we considered lead to folded states that are much more flexible than models with desolvation barriers (Fig. 5). In future studies, these model features, especially the flexibility of native proteins, should be compared with pertinent experimental observations to provide a more rigorous evaluation of the models and to facilitate development of more physical coarse-grained protein chain models that may incorporate both desolvation barrier and sidechain effects as well as allowing for the physical possibility of favorable nonnative interactions (Fig. 12).

Sidechain models here and elsewhere [82,83] that incorporate sequence- and denaturant-dependent transfer free energies provide a reasonably good account of folding thermodynamics of real proteins (Fig. 7a). Conceptually, this success is not unexpected given that SASA and sequence-dependent transfer free energies have long been known to afford a reasonably accurate account of protein folding thermodynamics in non-explicit-chain analyses that assume a fully folded state and a relatively open unfolded state [175]. In this regard, the essential native-centric nature of the above-mentioned explicit-chain sidechain models serve well to provide folded and unfolded states that are clearly separated (with little intermediate conformations because of the Gō-like potential) as envisioned in — and needed for — the early non-explicit-chain calculations. Nonetheless, because the explicit-chain representation of the sidechain models provides a wealth of energetic and structural information, these recent models are very valuable tools for analyzing and interpretating data from protein folding experiments [87].

Notwithstanding the success of these sidechain models in thermodynamics, we observed several limitations in their kinetic properties. First, a mild rollover is seen in our model chevron plot for Protein L (Fig. 7b,c) even though the experimental chevron plot for this protein is linear. Our analysis indicates that the mild chevron rollover originates from a combination of factors that have been suggested in the literature, including transient intermediate and internal friction (Figs. 8 and 9), Hammond behavior, and ground state effect (Fig. 10). Second, a more basic limitation shared by the sidechain models and other models that have been used to simulate chevron behavior is their failure to reproduce asymmetric chevron plots that are often observed experimentally (Fig. 11). As discussed above, asymmetric chevron plots indicates that the rate-limiting steps of folding and unfolding are significantly different [74, 157]. Building on the advance in coarse-grained modeling of proteins made so far, to account for chevron asymmetry, future models will have to capture the physics of the energetic difference of the rate-limiting steps of folding and unfolding.

Acknowledgements

H.S.C. thanks Dev Thirumalai for a stimulating discussion during the "Dynamics of Biomolecular Machinery" workshop at the Kavli Institute for Theoretical Physics in China in 2009, and his suggestion to investigate the effects of sidechain interactions on chevron plots. This work was supported by the Canada Research Chair Program and research grant MOP-84281 from the Canadian Institutes of Health Research to H.S.C. The authors are also grateful to SHARCNET and SciNet of Compute Canada for their generous allotments of computational resources.

References

[1] N. Gō, *Annu. Rev. Biophys. Bioeng.*, 1983, 12, 183.

[2] N. Gō, in *Old and New Views of Protein Folding*, ed. K. Kuwajima and M. Arai, Elsevier Science BV, Amsterdam, The Netherlands, 1999, pp. 97–105.

[3] J.D. Bryngelson and P.G. Wolynes, *Proc. Natl. Acad. Sci. USA*, 1987, 84, 7524.

[4] P.E. Leopold, M. Montal and J.N. Onuchic, *Proc. Natl. Acad. Sci. USA*, 1992, 89, 8721.

[5] K.A. Dill and H.S. Chan, *Nature Struct. Biol.*, 1997, 4, 10.

[6] B.A. Shoemaker, J.J. Portman and P.G. Wolynes, *Proc. Natl Acad. Sci. USA*, 2000, 97, 8868.

[7] K. Gunasekaran, C.-J. Tsai, S. Kumar, D. Zanuy and R. Nussinov, *Trends Biochem. Sci.*, 2003, 28, 81.

[8] V.N. Uversky and A.K. Dunker, *Biochim. Biophys. Acta*, 2010, 1804, 1231.

[9] P. Tompa, *Structure and Function of Intrinsically Disordered Proteins*. Chapmen & Hall/CRC Press, Boca Raton, Florida, 2010.

[10] T. Mittag and J.D. Forman-Kay, *Curr. Opin. Struct. Biol.*, 2007, 17, 3.

[11] P. Csermely, K.S. Sandhu, E. Hazai, Z. Hoksza, H.J.M. Kiss, F. Miozzo, D.V. Veres, F. Piazza and R. Nussinov, *Curr. Protein Peptide Sci.*, 2012, 13, 19.

[12] K. Sugase, H.J. Dyson and P.E. Wright, *Nature*, 2007, 447, 1021.

[13] Y. Huang and Z. Liu, *J. Mol. Biol.*, 2009, 393, 1143.

[14] D. Ganguly and J. Chen, *Proteins*, 2011, 79, 1251.

[15] J. Wang, X Chu, Y. Wang, S. Hagen, W. Han and E. Wang, *PLoS Comput. Biol.*, 2011, 7, e1001118.

[16] A. Bhattacherjee and S. Wallin, *Biophys. J.*, 2012, 102, 569.

[17] P. Tompa and M. Fuxreiter, *Trends Biochem. Sci.*, 2008, 33, 2.

[18] M. Borg, T. Mittag, T. Pawson, M. Tyers, J.D. Forman-Kay and H.S. Chan, *Proc. Natl Acad. Sci. USA*, 2007, 104, 9650.

[19] A.B. Sigalov, *Prog. Biophys. Mol. Biol.*, 2011 106, 525.

[20] J. Song, S.C. Ng, P. Tompa, K.A.W. Lee and H.S. Chan, *PLoS Comput. Biol.*, 2013, 9, e1003239.

[21] R. Lumry, R. Biltonen and J.F. Brandts, *Biopolymers*, 1966, 4, 917.

[22] D. Baker, *Nature*, 2000, 405, 39.

[23] H. Kaya and H.S. Chan, *Proteins*, 2000, 40, 637; Erratum: 2001, 43, 523.

[24] D. Barrick, *Phys. Biol.*, 2009, 6, 015001.

[25] J.J. Portman, *Curr. Opin. Struct. Biol.*, 2010, 20, 11.

[26] M. Knott and H.S. Chan, Proteins, 2006, 65, 373.

[27] M.M. Garcia-Mira, M. Sadqi, N. Fischer, J.M. Sanchez-Ruiz, and V. Muñoz, *Science*, 2002, 298, 2191.

[28] M. Sadqi, D. Fushman and V. Muñoz, *Nature*, 2006, 442, 317.

[29] S. Wallin and H.S. Chan, *J. Phys. Condens. Matt.*, 2006, 18, S307. Corrigendum: *J. Phys. Condens. Matt.*, 2009, 21, 329801.

[30] A. Badasyan, Z. Liu and H.S. Chan, *J. Mol. Biol.*, 2008, 384, 512.

[31] A. Badasyan, Z. Liu and H.S. Chan, *Int. J. Quantum Chem.*, 2009, 109, 3482.

[32] Z. Zhang and H.S. Chan, *Biophys. J.*, 2009, 96, L25.

[33] E.A. Shank, C. Cecconi, J.W. Dill, S. Marqusee and C. Bustamante, *Nature*, 2010, 465, 637.

[34] M. Rustad and K. Ghosh, *J. Chem. Phys.*, 2012, 137, 205104.

[35] H.S. Chan, S. Shimizu and H. Kaya, *Methods Enzymol.*, 2004, 380, 350.

[36] A.L. Watters, P. Deka, C. Corrent, D. Callender, G. Varani, T. Sosnick and D. Baker, *Cell*, 2007, 128, 613.

[37] J.M. Sanchez-Ruiz, *Biophys. Chem.*, 2010, 148, 1.

[38] C.M. Dobson, *Trends Biochem. Sci.*, 1999, 24, 329.

[39] A.K. Buell, A. Dhulesia, M.F. Mossuto, N. Cremades, J.R. Kumita, M. Dumoulin, M.E. Welland, T.P.J. Knowles, X. Salvatella and C.M. Dobson, *J. Am. Chem. Soc.*, 2011, 133, 7737.

[40] G. de Prat Gay, J. Ruiz-Sanz, J.L. Neira, L.S. Itzhaki and A.R. Fersht, *Proc. Natl. Acad. Sci. USA*, 1995, 92, 3683.

[41] J.L. Neira and A.R. Fersht, *J. Mol. Biol.*, 1999, 287, 421.

[42] M.P. Morrissey, Z. Ahmed and E.I. Shakhnovich, Polymer, 2004, 45, 557.

[43] A.H. Elcock, *PLoS Comput. Biol.*, 2006, 2, e98.

[44] E.P. O'Brien, J. Christodoulou, M. Vendruscolo and C.M. Dobson, *J. Am. Chem. Soc.*, 2011, 133, 513.

[45] H. Krobath, E.I. Shakhnovich and P.F.N. Faísca, *J. Chem. Phys.*, 2013, 138, 215101.

[46] J.J. Tyson, R. Albert, A. Goldbeter, P. Ruoff and J. Sible, *J. Royal Soc. Interface*, 2008, 5, S1.

[47] H. Qian, *Annu. Rev. Biophys.*, 2012, 41, 179.

[48] T. Sikosek, E. Bornberg-Bauer and H.S. Chan, *PLoS Comput. Biol.*, 2012, 8, e1002659.

[49] H.S. Chan, S. Bromnberg and K.A. Dill, *Phil. Trans. Royal Soc. London Ser. B*, 1995, 348, 61.

[50] K.A. Dill, *Biochemistry*, 1985, 24, 1501.

[51] K.A. Dill, *Biochemistry*, 1990, 29, 7133.

[52] H.S. Chan, *Proteins*, 2000, 40, 543.

[53] H.S. Chan, Z. Zhang, S. Wallin and Z. Liu, *Annu. Rev. Phys. Chem.*, 2011, 62, 301.

[54] S.S. Plotkin, J. Wang and P.G. Wolynes, *J. Chem. Phys.*, 1997, 106, 2932.

[55] M.P. Eastwood and P.G. Wolynes, *J. Chem. Phys.*, 2001, 114, 4702.

[56] S. Takada, Z. Luthey-Schulten and P.G. Wolynes, *J. Chem. Phys.*, 1999, 110, 11616.

[57] H. Kaya and H.S. Chan, *Phys. Rev. Lett.*, 2000, 85, 2823.

[58] A.I. Jewett, V.S. Pande and K.W. Plaxco, *J. Mol. Biol.*, 2003, 326, 247.

[59] H. Kaya and H.S. Chan, *Proteins*, 2003, 52, 524.

[60] K.W. Plaxco, K.T. Simons and D. Baker, *J. Mol. Biol.*, 1998, 277, 985.

[61] H.S. Chan, *Nature*, 1998, 392, 761.

[62] E. Freire and K.P. Murphy, *J. Mol. Biol.*, 1991, 222, 687.

[63] H. Kaya and H.S. Chan, *Proteins*, 2005, 58, 31.

[64] V.J. Hilser, B. Garcia-Moreno, T.G. Oas, G. Kapp and S.T. Whitten, *Chem. Rev.*, 2006, 106, 1545.

[65] K. Ghosh and K.A. Dill, *J. Am. Chem. Soc.*, 2009, 131, 2306.

[66] Y. Bai, T.R. Sosnick, L. Mayne and S.W. Englander, *Science*, 1995, 269, 192.

[67] C.C. Mello and D. Barrick, *Proc. Natl. Acad. Sci. USA*, 2004, 101, 14102.

[68] J.A. Rank and D. Baker, *Protein Sci.*, 1997, 6, 347.

[69] M.S. Cheung MS, A.E. García and J.N. Onuchic, *Proc. Natl. Acad. Sci. USA*, 2002, 99, 685.

[70] H. Kaya and H.S. Chan, *J. Mol. Biol.*, 2003, 326, 911. Corrigendum: *J. Mol. Biol.*, 2004, 337, 1069.

[71] Z. Liu and H.S. Chan, *Phys. Biol.*, 2005, 2, S75.

[72] L.R. Pratt and D. Chandler, *J. Chem. Phys.*, 1977, 67, 3683.

[73] Y. Levy and J.N. Onuchic, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, 35, 389.

[74] Z. Liu and H.S. Chan, *J. Mol. Biol.*, 2005, 349, 872.

[75] A. Ferguson, Z. Liu and H.S. Chan, *J. Mol. Biol.*, 2009, 389, 619. Corrigendum: *J. Mol. Biol.*, 2010, 401, 153.

[76] H. Kaya, Z. Uzunoğlu and H.S. Chan, *Phys. Rev. E* 2013, 88, 044701.

[77] M.P. Taylor, W. Paul and K. Binder, *Phys. Rev. E*, 2009, 79, 050801(R).

[78] E.I. Shakhnovich and A.V. Finkelstein, *Biopolymers*, 1989, 28, 1667.

[79] D.K. Klimov and D. Thirumalai, *Fold. Des.*, 1998, 3, 127.

[80] L. Li, L.A. Mirny and E.I. Shakhnovich, *Nature Struct. Biol.*, 2000, 7, 336.

[81] H. Kaya, Z. Liu and H.S. Chan, *Biophys. J.*, 2005, 89, 520.

[82] E.P. O'Brien, G. Ziv, G. Haran, B.R. Brooks amd D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2008, 105, 13403.

[83] Z. Liu, G. Reddy, E.P. O'Brien and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2011, 108, 7787.

[84] Z. Liu, G. Reddy and D. Thirumalai, *J. Phys. Chem. B*, 2012, 116, 6707.

[85] D. Thirumalai, E.P. O'Brien, G. Morrison and C. Hyeon, *Annu. Rev. Biophys.*, 2010, 39, 159.

[86] G. Reddy, Z. Liu and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2012, 109, 17832.

[87] D. Thirumalai, Z. Liu, E.P. O'Brien and G. Reddy, *Curr. Opin. Struct. Biol.*, 2013, 23, 22.

[88] J.N. Onuchic and P.G. Wolynes, *Curr. Opin. Struct. Biol.*, 2004, 14, 70.

[89] E. Shakhnovich, *Chem. Rev.*, 2006, 106, 1559.

[90] C. Clementi, *Curr. Opin. Struct. Biol.*, 2008, 18, 10.

[91] K.A. Dill, S.B. Ozkan, M.S. Shell and T.R. Weikl, *Annu. Rev. Biophys.*, 2008, 37, 289.

[92] D.L. Pincus, S.S. Cho, C. Hyeon and D. Thirumalai, *Prog. Mol. Biol. Trans. Sci.*, 2008, 84, 203.

[93] R.D. Hills and C.L. Brooks, *Int. J. Mol. Sci.*, 2009, 10, 889.

[94] J. Zhang, W. Li, J. Wang, M. Qin, L. Wu, Z. Yan, W. Xu, G. Zuo and W. Wang, *IUBMB Life*, 2009, 61, 627.

[95] V. Tozzini, *Q. Rev. Biophys.*, 2010, 43, 333.

[96] S. Takada, *Curr. Opin. Struct. Biol.*, 2012, 22, 130.

[97] A.N. Naganathan, *Wiley Interdisciplinary Reviews: Comput. Mol. Sci.*, 2013, 3, 504.

[98] W.G. Noid, *J. Chem. Phys.*, 2013, 139, 090901.

[99] C.D. Snow, H. Nguyen, V.S. Pande and M. Gruebele, *Nature*, 2002, 420, 102.

[100] K. Lindorff-Larsen, S. Piana, R.O. Dror and D.E. Shaw, *Science*, 2011, 334, 517.

[101] H. Lei, X. Deng, Z. Wang and Y. Duan, *J. Chem. Phys.*, 2008, 129, 155104.

[102] P.L. Freddolino and K. Schulten, *Biophys. J.*, 2009, 97, 2338.

[103] V.A. Voelz, G.R. Bowman, K. Beauchamp and V.S. Pande, *J. Am. Chem. Soc*, 2010, 132, 1526.

[104] C. Zhang and J. Ma, *J. Chem. Phys.*, 2010, 132, 244101.

[105] S. Piana, K. Lindorff-Larsen and D.E. Shaw, *Proc. Natl. Acad. Sci. USA*, 2012, 109, 17845.

[106] S. Piana, K. Lindorff-Larsen and D.E. Shaw, *Proc. Natl. Acad. Sci. USA*, 2013, 110, 5915.

[107] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan and W. Wriggers, *Science*, 2010, 330, 341.

[108] J.R. Allison, M. Bergeler, N. Hansen and W.F. van Gunsteren, *Biochemistry*, 2011, 50, 10965.

[109] S. Piana, K. Lindorff-Larsen and D.E. Shaw, *Biophys. J.*, 2011, 100, L47.

[110] R.O. Dror, H.F. Green, C. Valand, D.W. Borhani, J.R. Valcourt, A.C. Pan, D.H. Arlow, M. Canals, J.R. Lane, R. Rahmani, J.B. Baell, P.M. Sexton, A. Christopoulos and D.E. Shaw, *Nature*, 2013, doi:10.1038/nature12595.

[111] W.F. van Gunsteren, R. Burgi, C. Peter and X. Daura, *Angew. Chem.*, 2001, 113, 363.

[112] W.F. van Gunsteren, R. Burgi, C. Peter and X. Daura, *Angew. Chem. Int. Ed.*, 2001, 40, 351.

[113] R.B. Best, *Curr. Opin. Struct. Biol.*, 2012, 22, 52.

[114] D.W. Borhani and D.E. Shaw, *J. Comput. Aided Mol. Des.*, 2012, 26, 15.

[115] C.R. Matthews and M.R. Hurle, *BioEssays*, 1987, 6, 254.

[116] M.R. Hurle, G.A. Michelotti, M.M. Crisanti and C.R. Matthews, *Proteins*, 1987, 2, 54.

[117] S.E. Jackson and A.R. Fersht, *Biochemistry*, 1991, 30, 10428.

[118] A. Matouschek, J.T. Kellis, L. Serrano, M. Bycroft and A.R. Fersht, *Nature*, 1990, 346, 440.

[119] K.A. Scott and J. Clarke, *Protein Sci.*, 2005, 14, 1617.

[120] M. Tsytlonok and L.S. Itzhaki, *Archive Biochem. Biophys.*, 2013, 531, 14.

[121] I.E. Sanchez and T. Kiefhaber, *J. Mol. Biol.*, 2003, 325, 367.

[122] C.F. Wright, K. Lindorff-Larsen, L.G. Randles and J. Clarke, *Nature Struct. Biol.*, 2003, 10, 658.

[123] N. Aghera and J.B. Udgaonkar, *Biochemistry*, 2013, 52, 5770.

[124] Z. Zhang and H.S. Chan, *Proc. Natl. Acad. Sci. USA*, 2010, 107, 2920.

[125] R.B. Best, G. Hummer and W.A. Eaton, *Proc. Natl. Acad. Sci. USA*, 2013, 110, 17874.

[126] D.J. Brockwell and S.E. Radford, *Curr. Opin. Struct. Biol.*, 2007, 17, 30.

[127] A. Zarrine-Afsar, S. Wallin, A.M. Neculai, P. Neudecker, P.L. Howell, A.R. Davidson and H.S. Chan, *Proc. Natl. Acad. Sci. USA*, 2008, 105, 9999.

[128] A. Azia and Y. Levy, *J. Mol. Biol.*, 2009, 393, 527.

[129] A. Zarrine-Afsar, Z. Zhang, K.L. Schweiker, G.I. Makhatadze, A.R. Davidson and H.S. Chan, Proteins, 2012, 80, 858.

[130] R.B. Best, *J. Phys. Chem. B*, 2013, 117, 13235.

[131] J.-E. Shea, Y.D. Nochomovitz, Z. Guo and C.L. Brooks, *J. Chem. Phys.*, 1998, 109, 2895.

[132] T.V. Pogorelov and Z. Luthey-Schulten, *Biophys. J.*, 2004, 87, 207.

[133] L.L. Chavez, J.N. Onuchic and C. Clementi, *J. Am. Chem. Soc.*, 2004, 126, 8426.

[134] A. Šali, E.I. Shakhnovich and M. Karplus, *Nature*, 1994, 369, 248.

[135] S.S. Cho, Y. Levy and P.G. Wolynes, *Proc. Natl. Acad. Sci. USA*, 2006, 103, 586.

[136] D.K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, 2000, 97, 2544.

[137] W.F. van Gunsteren and H.J.C. Berendsen, *Mol. Phys.*, 1977, 34, 1311.

[138] E. Neria, S. Fischer and M. Karplus, *J. Chem. Phys.*, 1996, 105, 1902.

[139] D. Frishman and P. Argos, *Proteins*, 1995, 23, 566.

[140] S. Miyazawa and R.L. Jernigan, *Proteins*, 1999, 34, 49.

[141] M.S. Cheung, J.M. Finke, B. Callahan and J.N. Onuchic, *J. Phys. Chem. B*, 2003, 107, 11193.

[142] B. Lee and F.M. Richards, *J. Mol. Biol.*, 1971, 55, 379.

[143] J.W. Ponder, TINKER4.2, June 2004, Department of Chemistry, Washington University, St. Louis, Missouri, U.S.A.

[144] T.J. Richmond, *J. Mol. Biol.*, 1984, 178, 63.

[145] L. Wesson and D. Eisenberg, *Protein Sci.*, 1992, 1, 227.

[146] G.I. Makhatadze, *J. Phys. Chem. B*, 1999, 103, 4781.

[147] T. Veitshans, D. Klimov and D. Thirumalai, *Fold. Des.*, 1997, 2, 1.

[148] E.A. Coutsias, C. Seok and K.A. Dill, *J. Comput. Chem.*, 2004, 25, 1849.

[149] J.W. O'Neill, D.E. Kim, D. Baker and K.Y. Zhang, *Acta Crystallogr., Sect. D*, 2001, 57, 480.

[150] M.L. Scalley, Y. Qian, H. Gu, A. McCormack, J.R. Yates III and D. Baker, *Biochemistry*, 1997, 36, 3373.

[151] D.W. Yang and L.E. Kay, *J. Mol. Biol.*, 1996, 263, 369.

[152] D.O.V. Alonso and V. Daggett, *J. Mol. Biol.*, 1995, 247, 501.

[153] S.S. Plotkin and J.N. Onuchic, *J. Chem. Phys.*, 2002, 116, 5263

[154] H.S. Chan and K.A. Dill, *Proteins*, 1998, 30, 2.

[155] H. Kaya and H.S. Chan, *J. Mol. Biol.*, 2002, 315, 899.

[156] H. Kaya and H.S. Chan, *Phys. Rev. Lett.*, 2002, 90, 258104.

[157] H. Kaya and H.S. Chan, *Proteins*, 2003, 52, 510.

[158] S. Shimizu and H.S. Chan, Proteins, 2002, 48, 15. Erratum: Proteins, 2002, 49, 294.

[159] K.W. Plaxco and W.A. Eaton, *Proc. Natl. Acad. Sci. USA*, 1998, 95, 13591.

[160] H.S. Chung and W.A. Eaton, *Nature*, 2013, doi:10.1038/nature12649.

[161] G.S. Hammond, *J. Am. Chem. Soc.*, 1955, 77, 334.

[162] J.M. Matthews and A.R. Fersht, *Biochemistry*, 1995, 34, 6805.

[163] M. Oliveberg, Y.-J. Tan, M. Silow and A.R. Fersht, *J. Mol. Biol.*, 1998, 277, 933.

[164] M. Schätzle and T. Kiefhaber, *J. Mol. Biol.* 2006, 357, 655.

[165] Z. Zhang and H.S. Chan, *Biophys. J.*, 2009, 96, L25.

[166] Z. Zhang and H.S. Chan, *Proc. Natl. Acad. Sci. USA*, 2012, 109, 20919.

[167] V.P. Grantcharova and D. Baker, *Biochemistry*, 1997, 36, 15685.

[168] T. Schindler and F.X. Schmid, *Biochemistry*, 1996, 35, 16833.

[169] M. Eisenberg and G. Schwert, *J. Gen. Physiol.*, 1951, 34, 583/

[170] J.F. Brandts, *J. Am. Chem. Soc.*, 1964, 86, 4302.

[171] R. Lumry and R. Biltonen, in *Structure and Stability of Biological Macromolecules*, ed. S.N. Timasheff and G.D. Fasman, Marcel Dekker, Inc., New York, 1969, ch. 2, pp. 65–212.

[172] S. Segawa and M. Sugihara, *Biopolymers*, 1984, 23, 2473.

[173] R.L. Baldwin, C. Frieden and G.D. Rose, *Proteins*, 2010, 78, 2725.

[174] C. Clementi and S.S. Plotkin, *Protein Sci.*, 2004, 13, 1750.

[175] J.K. Myers, C.N. Pace and J.M. Scholtz, *Protein Sci.*, 1995, 4, 2138.

Figure Captions

Figure 1. Energetic and geometric features of the models used in this study. (a) $E_{\rm c}$ is the potential energy between a pair of C_α atoms i, j belonging to the native contact set and at distance r_{ij} in the C_α-Gō (blue), desolvation-barrier (db) (black), and sidechain (SC) (red) models. The model interaction energies are given in units of ϵ, ϵ, and kcal mol^{-1}, respectively. (b) Basic geometry of the sidechain model. Positions of C_α atoms and sidechain pseudo-atoms (centered at the centroids of the sidechains) are indicated by gray and bronze spheres, respectively. Note, however, that the relative sizes of the C_α atoms and sidechain pseudo-atoms are not necessarily drawn to scale. Angles labeled by θ_1, θ_2, θ_3, ϕ_1, and ψ_1 are examples of the bond angles θ_{ik}s, torsional angles ϕ_is, and improper dihedral angles ψ_is in the model. (c) Schematic of an example native contact (indicated by dotted line) favored by local-nonlocal coupling. The C_α rmsd values (in Å) indicate small root-mean-square deviations of each of the two 5-residue protein segments from their corresponding local conformations in the native structure. (d) An example of native contacts between the same two residues in (c) that are disfavored by local-nonlocal coupling because in this case the local structures around the two residues have large deviations from their corresponding native conformations.

Figure 2. Energetic analysis of local-nonlocal coupling in the SC model. (a, b) Two different conformations of the Protein L sequence from snapshots of the SC-Gō model sampled under midpoint conditions at simulation temperature 356 K. According to the distance criterion in our SC model, the residue pair 30-58 (Thr-Leu) is in contact in both snapshots. The mainchain rmsd from native for the two 5-residue segments (enclosed in the dotted boxes) centered at Thr30 and at Leu58 is small for snapshot (a), viz., 0.207Å and 0.775Å respectively, whereas the corresponding values are larger at 0.336Å and 1.782Å for snapshot (b). In these drawings, only the C_α atoms and sidechain pseudo-atoms of the two 5-residue segments (positions 28–32 and 56–60) are shown as spheres. Other parts of the protein sequence are shown merely as mainchain traces. The pairs of contacting C_α atoms and sidechain pseudo-atoms are further highlighted by their depiction as red and green spheres respectively. The size of these spheres are larger and more reflective of their excluded volumes in the model than the size of spheres used for the rest of the two segments (which is drawn in the same style as that in Fig. 1b). (c, d) Magnified representation of the two 5-residue segments in (a) and (b) respectively. (e, f) Contributions from various native-centric terms to the interaction energy in the SC model between the pair of contacting residues in (c) and (d) respectively. The underlying potential energy functions for C_α-C_α, sidechain-sidechain, C_α-sidechain, and sidechain-C_α of the interaction between residues 30 and 58, respectively, are shown in the same order by the red, green, light blue, and magenta curves. For each energy term, the corresponding distance between the C_α atom(s) and/or sidechain pseudo-atom(s) is indicated by an arrow of the same color. The corresponding differences in interaction in the db model between this pair of residues in the two conformations are also included as black curves with the corresponding arrows to facilitate comparison with Fig. 3.

Figure 3. Energetic analysis of local-nonlocal coupling in the db model. (a) and (b) are the mainchain conformations of the two 5-residue segments in Fig. 2c and d. Here the contacting residues 30 and 58 are shown as large red spheres whereas C_α positions of the rest of the two 5-residue segments are shown as smaller gray spheres. The size of the large red spheres highlights the optimal (native) C_α-C_α distance between the two residues in the model. A water molecule is represented in (b) by a blue sphere of radius 1.5Å to underscore a hypothetical excluded-volume clash that underlies the unfavored interaction free energy between the two residues at the db distance. (c, d) The native-centric potential energy functions for the given residue pair in the C_α-Gō model and the db model are plotted as blue and black curves respectively. The arrows in (c) and (d) indicate, respectively, the C_α-C_α distances between the two residues in (a) and (b).

Figure 4. Folding cooperativity and local-nonlocal coupling. (a) Free energy profiles of the models (as marked) for Protein L in this study, where $P(Q)$ is normalized conformational population as a function of Q and hence $-\ln P(Q)$ is free energy in units of $k_{\rm B}T$. Here, Q is the fraction of the 137 C_α–C_α native contacts in the PDB structure of Protein L that are present in a given conformation. The $-\ln P(Q)$ profiles shown are computed under each model's transition-midpoint condition, which can be different for different models. In units of $k_{\rm B}$ with $\epsilon = 1$, the midpoint temperature for the C_α-Gō model is 1.01 and that for the db model (where $\epsilon_{\rm db} = 0.1\epsilon$ and $\epsilon_{\rm ssm} = 0.2\epsilon$) is 0.849. The midpoint temperature for the sidechain-Gō (SC-Gō) model is 356 K ($k_{\rm B}T = 0.707$ kcal/mol). The simulation temperatures and midpoint denaturant concentrations for the SC-urea and SC-GdmCl models are 331 K ($k_{\rm B}T = 0.658$ kcal/mol), [C] = 6.6 M, and 338 K ($k_{\rm B}T = 0.672$ kcal/mol), [C] = 2.5 M, respectively. (b) The local-nonlocal coupling parameter averaged over the native contacts formed, $\langle C_{\rm lnl}\rangle$, is simulated under the models' respective midpoint conditions and shown as functions of Q using the same color code for the models as that in (a).

Figure 5. Comparing model native ensembles. For each of the sidechain-Gō, C_α-Gō, and db (Cα-db) models, ten representative conformations chosen randomly around the native minimum are depicted in blue, red, or green, viz., the Q values of the conformations shown for a given model are at or nearly at the high-Q minimum of $-\ln P(Q)$ in Fig. 4. The average C_α rmsd values (in Å) of each of these model native ensembles from the PDB structure 1HZ6 for Protein L (depicted in black) are indicated below the conformational drawings.

Figure 6. Route measure. The route measure $R(Q)$ as a function of fractional native contact Q for each of the models in this study is plotted using the same color code as that in Fig. 4. A higher route measure at a given Q implies that the conformational distribution is more restricted at that Q value.

Figure 7. Thermodynamic stability and model chevron plots. (a) Native-state stability (free energy of folding $\Delta G_{\rm f}$ in units of $k_{\rm B}T$) of SC-GdmCl and SC-urea models of Protein L as a function of GdmCl (brown data points) or urea (magenta data points) concentration (C in units of M) at temperatures $T = 338$ K and 331 K respectively. In our thermodynamic analysis, the folded state and unfolded state for these models are defined, respectively, by $Q > Q_{\rm F}$ and $Q < Q_{\rm U}$, where $Q_{\rm F} = 101/137$ and $Q_{\rm U} = 21/137$. (b) The corresponding chevron plots. Data points for negative logarithm of the simulated mean first passage time (MFPT) for folding and unfolding are shown by filled and open symbols, respectively, using the same colors as in (a) to denote GdmCl and urea dependence. Folding simulations start with randomly generated conformations with the Q value at the unfolded minimum, and the model protein is considered to be folded when $Q > Q_{\rm F}$. Unfolding simulations are initiated from a native $Q = 1$ conformation, and the model protein is considered to be unfolded when $Q < Q_{\rm U}$. The dotted V-shapes are hypothetical two-state chevron plots that are consistent with the denaturant dependence of thermodynamic stability in (a). Transition midpoints of these two SC models are marked by the vertical and horizontal dashed lines in (a) and (b). (c) Unified comparison of model chevron plots, with $-\ln$(MFPT) plotted as a function of native stability $\Delta G_{\rm f}/k_{\rm B}T$. Here variation of $\Delta G_{\rm f}$ is a result of varying denaturant concentration C for the SC-GdmCl (brown circles) and SC-urea (magenta squares) models [as in (a)], but is a result of varying T for the C_α-Gō (red inverted triangles), db (green diamonds), and SC-Gō (blue triangles) models. As in (b), data points for folding and unfolding are represented, respectively, by filled and open symbols; and each of the dotted V-shapes is a hypothetical two-state chevron plot for the model denoted by the same color. $Q_{\rm F} = 101/137$ for all three SC models; $Q_{\rm U} = 21/137$ for the SC-urea and SC-GdmCl models [as in (a) and (b)], and $Q_{\rm U} = 17/137$ for the SC-Gō model. For models without an explicit sidechain representation, $Q_{\rm F} = 37/137$, $Q_{\rm U} = 112/137$ for the C$_\alpha$-Gō model, and $Q_{\rm F} = 23/137$, $Q_{\rm U} = 135/137$ for the db model.

Figure 8. Transient intermediates can be an origin of mild chevron rollovers. (a) An example trajectory of the SC-urea model simulated at the folding-unfolding transition midpoint of the model ($C = 6.6$ M, $T = 331$ K). (b) An example trajectory of the SC-Gō model simulated under midpoint conditions ($T = 356$ K). (c) An example trajectory of the SC-urea model simulated for zero denaturant ($C = 0$, $T = 331$ K). This model is equivalent to the SC-Gō model simulated under the same temperature. The dependence of Q on simulation time (number of Langevin time steps) of this trajectory illustrates the physical connection between transient populations with $Q \gtrsim 0.4$ and the mild chevron rollover in the folding arms of these models observed in Fig. 7b,c. (d) Snapshots of transient structures (green conformational traces) sampled at three time points [(i), (ii), and (iii)] along the trajectory shown in (c). To facilitate structural comparison with the native PDB structure, the PDB structure is depicted by black traces positioned with minimum rmsd from each of the green transient conformations.

Figure 9. Transient intermediates in the C$_\alpha$-Gō model. (a) An example trajectory of the C$_\alpha$-Gō model simulated near the folding-unfolding transition midpoint ($\Delta G_{\rm f}/k_{\rm B}T = 0.198$). (b) An example trajectory in the same model simulated under strongly folding conditions ($\Delta G_{\rm f}/k_{\rm B}T = -10.41$). (c) Snapshots of transient structures (green conformational traces) sampled at three time points [(i), (ii), and (iii)] along the trajectory shown in (b). As in Fig. 8d, the superposing black traces are the reference native PDB structure with minimum rmsd from each of the green transient conformations.

Figure 10. Movement of putative transition state along progress variable Q provides a partial but not a full account of model chevron rollover. (a) Plotted in black [$-\ln$(MFPT), left vertical scale] is the SC-urea chevron plot (filled and open squares) and the corresponding hypothetical two-state chevron behavior (dotted lines) in Fig. 7b. Plotted as blue solid curves ($-\Delta G^{\ddagger}/k_{\rm B}T$, right vertical scale) are minus ($-1\times$) folding and unfolding free energy barrier in units of $k_{\rm B}T$ (blue curves spanning the folding and unfolding arms, respectively) deduced from the populations of the shifting peak of the free energy profiles. (b) Free energy profiles of the SC-urea model at different urea concentrations ($C =$ [urea]) are shown in black. For the profiles plotted as solid curves, $C = 0$, 1.5, 3.0, 4.5 and 6.0 M from top to bottom. For the profiles plotted as dashed curves, $C = 6.6$, 7.4, 8.2, 9.0, and 9.8 M from bottom to top. The two vertical grey strips are the low-Q unfolded (left) and the high-Q folded (right) regions, whereas the curved area shaded in light blue around $Q \approx 0.3 - 0.5$ is the putative denaturant-dependent transition-state region used in the present analysis. The plotted $-\Delta G^{\ddagger}/k_{\rm B}T$ values [blue curves in (a)] are determined by the unfolded (U), folded (F), and putative transition-state (TS) populations in these conformational regions ($P_{\rm U}$, $P_{\rm F}$, and $P_{\rm TS}$ respectively), whereby $-\Delta G^{\ddagger}/k_{\rm B}T = \ln(P_{\rm TS}/P_{\rm U})$ for folding and $-\Delta G^{\ddagger}/k_{\rm B}T = \ln(P_{\rm TS}/P_{\rm F})$ for unfolding. If $\Delta G^{\ddagger}$ for folding is alternately defined as the free energy difference between the top of the barrier and the bottom of the unfolded-state minimum, the deviation between the TST-predicted and simulated folding rate would be even larger because there is a gradual shift of the unfolded-state minimum toward higher Q values as C decreases.

Figure 11. The chevron plot for Protein L predicted by the SC-GdmCl model differs significantly from that obtained from experiment. Plotted in brown is the chevron plot predicted by our SC-GdmCl model. As in Fig. 7b, the natural logarithm of folding and unfolding MFPT (right vertical scale) are represented by filled and open squares, respectively; whereas the dotted brown V-shape is a hypothetical two-state chevron plot constructed to be consistent with the denaturant dependence of native stability in the same SC-GdmCl model. Plotted in black is the experimental chevron plot obtained by Scalley et al. Data points shown as filled and open squares are natural logarithm of relaxation rates ($k_{\rm obs}$, left vertical scale) determined using folding and unfolding experiments, respectively, by Scalley et al. in the absence of 0.4 M Na_2SO_4 as reported in Fig. 6 of Ref. [150]. The black dotted V-shape here is constructed in accordance with the experimentally determined $m_{\rm f} = 1.5$ kcal mol^{-1} M^{-1} and $m_{\rm u} = 0.5$ kcal mol^{-1} M^{-1} provided in Table 1 of the same reference. It is clear from this comparison that the experimental folding rate is significantly more sensitive to GdmCl concentration than that predicted by the present SC-GdmCl model. The inset provides further statistical analysis of simulated folded times using the method in Fig. 4 of Ref. [124]. It shows that the dependence of logarithmic fractional unfolded population $\ln P$(unfolded) on simulation time t is approximately linear. The plotted data points are for $C = 0$ M (red), 1.6 M (green), and 2.4 M (blue). These results indicate that kinetic relaxation of folding simulated using the SC-GdmCl model is essentially single-exponential, thus folding MFPT = (folding rate)$^{-1}$.

Figure 12. Native-centric SC models augmented by solvent-dependent free energies of transfer of amino acids may be used to study sequence-specific nonnative interactions.

The reference denaturant concentration ([urea]) in the (default) SC-urea model is $C = 0$, i.e., the model is a pure SC-Gō model at $C = C_0 = 0$. This model is hereby denoted as SC-[urea]$^{C_0=0\mathrm{M}}$. For the rationale provided in the text, we generalize the SC-urea formulation to consider SC models with any given C_0. (a) Simulated chevron plots [$-\ln(\mathrm{MFPT})$ data points plotted as squares, triangles, circles or diamonds] and their corresponding hypothetical two-state folding and unfolding chevron arms (dotted V-shapes, same style as in the figures presented above) for the original SC-[urea]$^{C_0=0\mathrm{M}}$ model (black) and an SC-[urea]$^{C_0=6.2\mathrm{M}}$ model with $C_0 = 6.2$ M (red). To ensure numerical accuracy of our model predictions, results from three sets of independent simulations are shown for the SC-[urea]$^{C_0=0\mathrm{M}}$ model (black symbols). For the folding arm (filled symbols), the minimal numbers of trajectories we have used to determine any folding MFPT are 2497, 159, and 410, respectively, for the data points plotted as circles, squares, and triangles. For the unfolding arm (open symbols), the minimal numbers of trajectories we have used to determine any unfolding MFPT are 2443, 172, and 197, respectively, for the corresponding data sets. For the SC-[urea]$^{C_0=6.2\mathrm{M}}$ model, one set of $-\ln(\mathrm{MFPT})$ values simulated using ≥ 2500 trajectories for each data point is shown (red diamonds). (b) Free energy profiles for the SC-[urea]$^{C_0=0\mathrm{M}}$ (black) SC-[urea]$^{C_0=6.2\mathrm{M}}$ (red), and SC-[urea]$^{C_0=8.9\mathrm{M}}$ (blue) models simulated at or near the models' respective transition midpoints. When C_0 is increased from 0 M to 8.9 M, there is an appreciable lowering of the free energy barrier at the putative transition state. (c) Nonnative contacts increase with increasing C_0. Here a pair of sidechains i, j separated by ≥ 4 virtual bonds in a given conformation is considered to be in a nonnative contact if and only if (i) the distance r_{ij} between their centers satisfies $r_{ij} < r_i^{\mathrm{vdW}} + r_j^{\mathrm{vdW}} + 1\text{Å}$, (ii) this condition is not satisfied in the native PDB structure, and (iii) residue pair i, j is not in the native contact set defined above. We use $N_{\mathrm{nn,SC}}$ to denote the total number of such nonnative sidechain-sidechain contacts in a given conformation. Using the same color code as that in (b), $N_{\mathrm{nn,SC}}$ is shown as a function of Q for the three SC models we have considered.

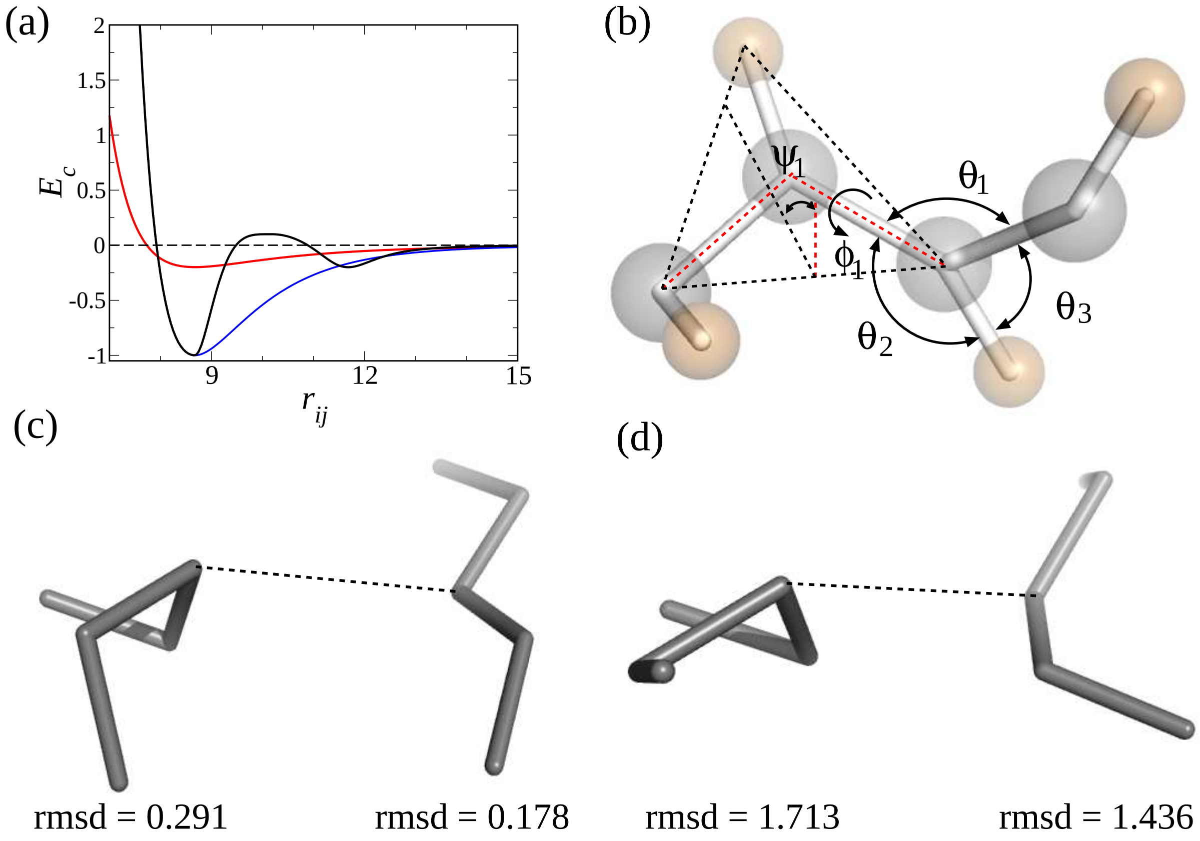
(a)
2
1.5
1
0.5
0
-0.5
-1
E_c
9
12
15
r_{ij}
(b)
ψ_1
θ_1
ϕ_1
θ_3
θ_2
(c)
(d)
rmsd = 0.291
rmsd = 0.178
rmsd = 1.713
rmsd = 1.436

Figure 1

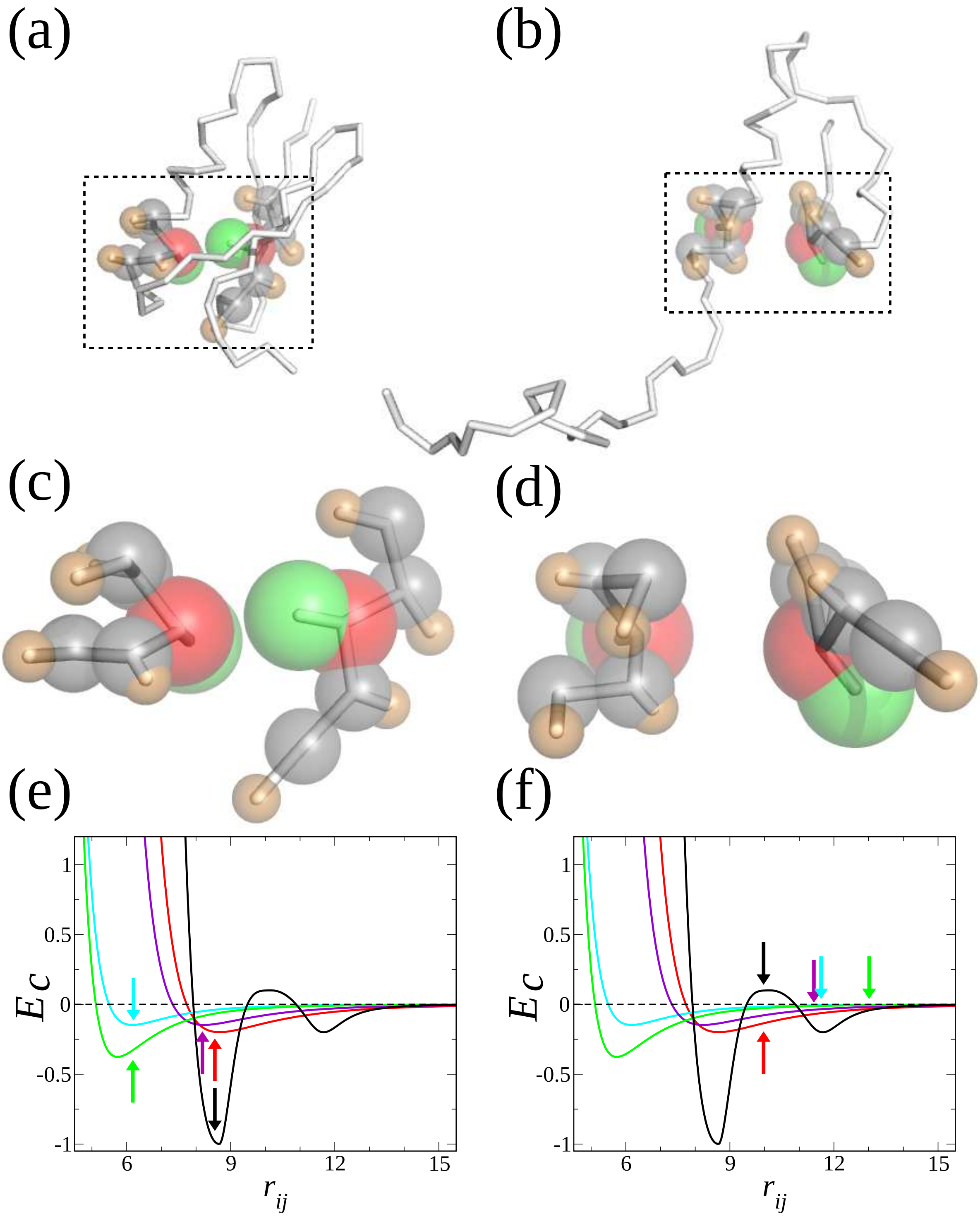

Figure 2

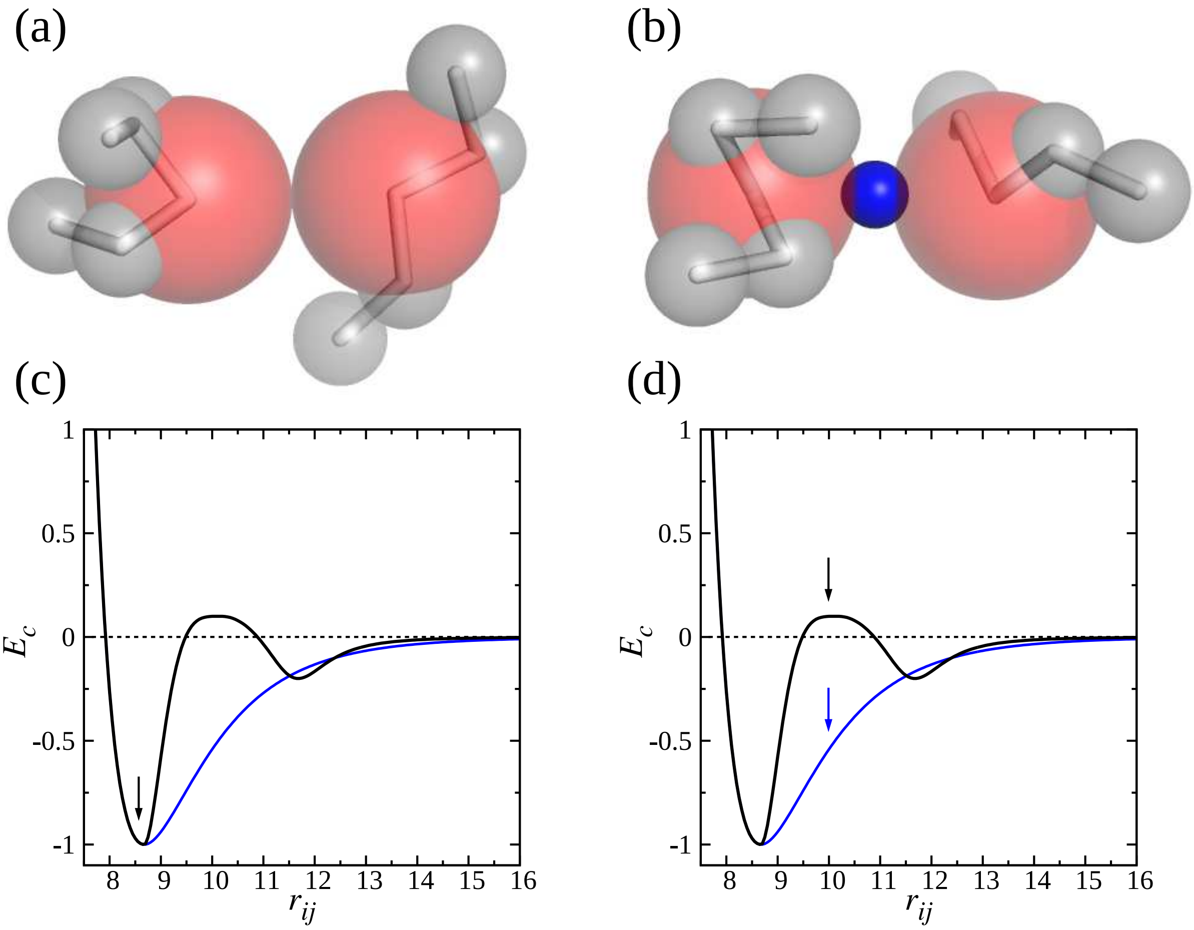

Figure 3

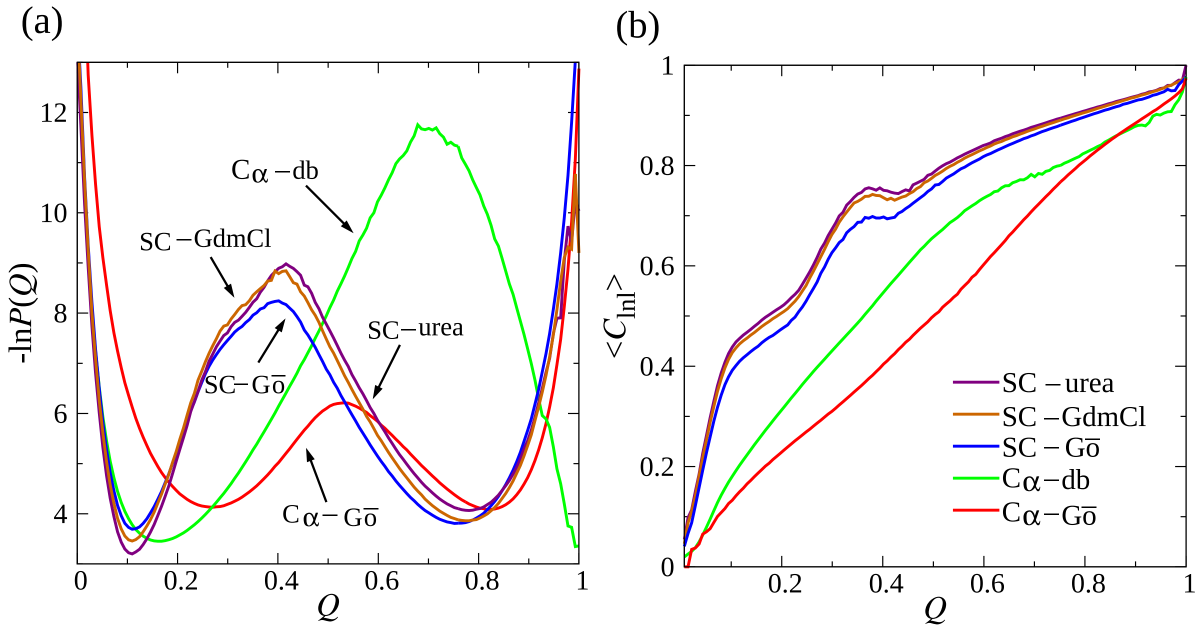
(a)
12
10
8
6
4
0
0.2
0.4
0.6
0.8
1
-ln$P(Q)$
Q
Cα −db
SC −GdmCl
SC−Gō
SC−urea
Cα− Gō
(b)
1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1
$<C_{\text{lnl}}>$
Q
SC −urea
SC −GdmCl
SC − Gō
Cα−db
Cα−Gō

Figure 4

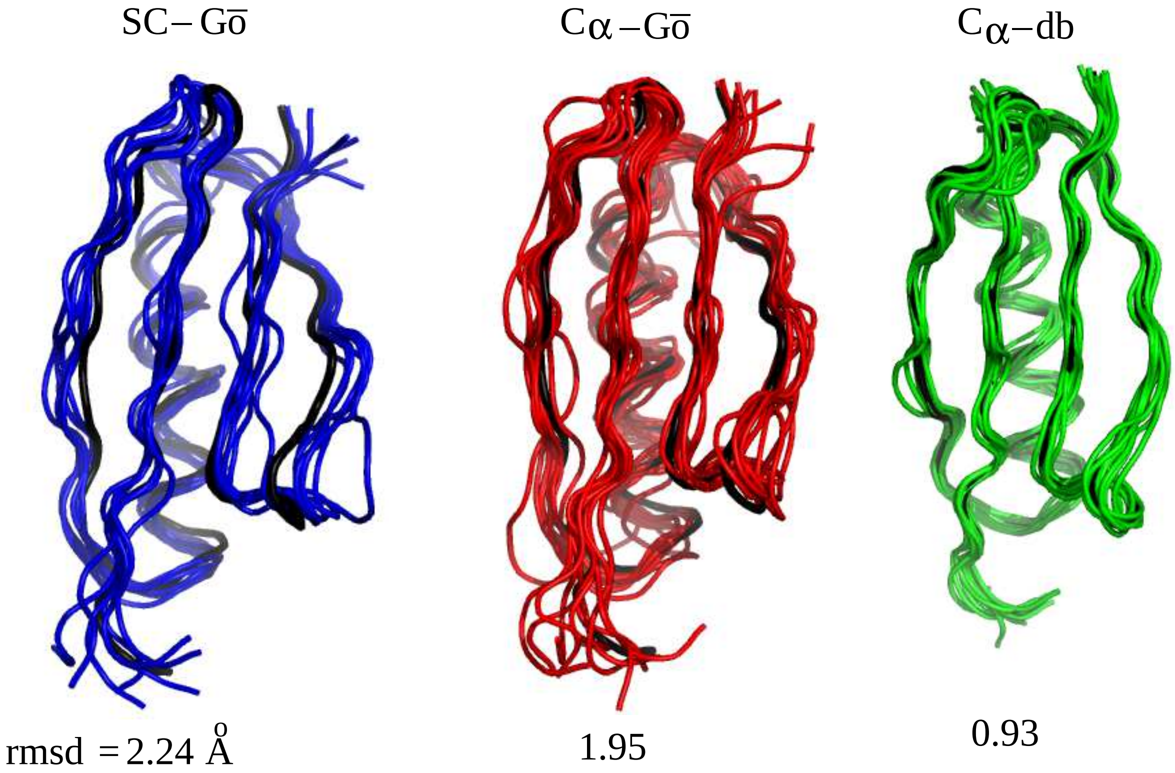

Figure 5

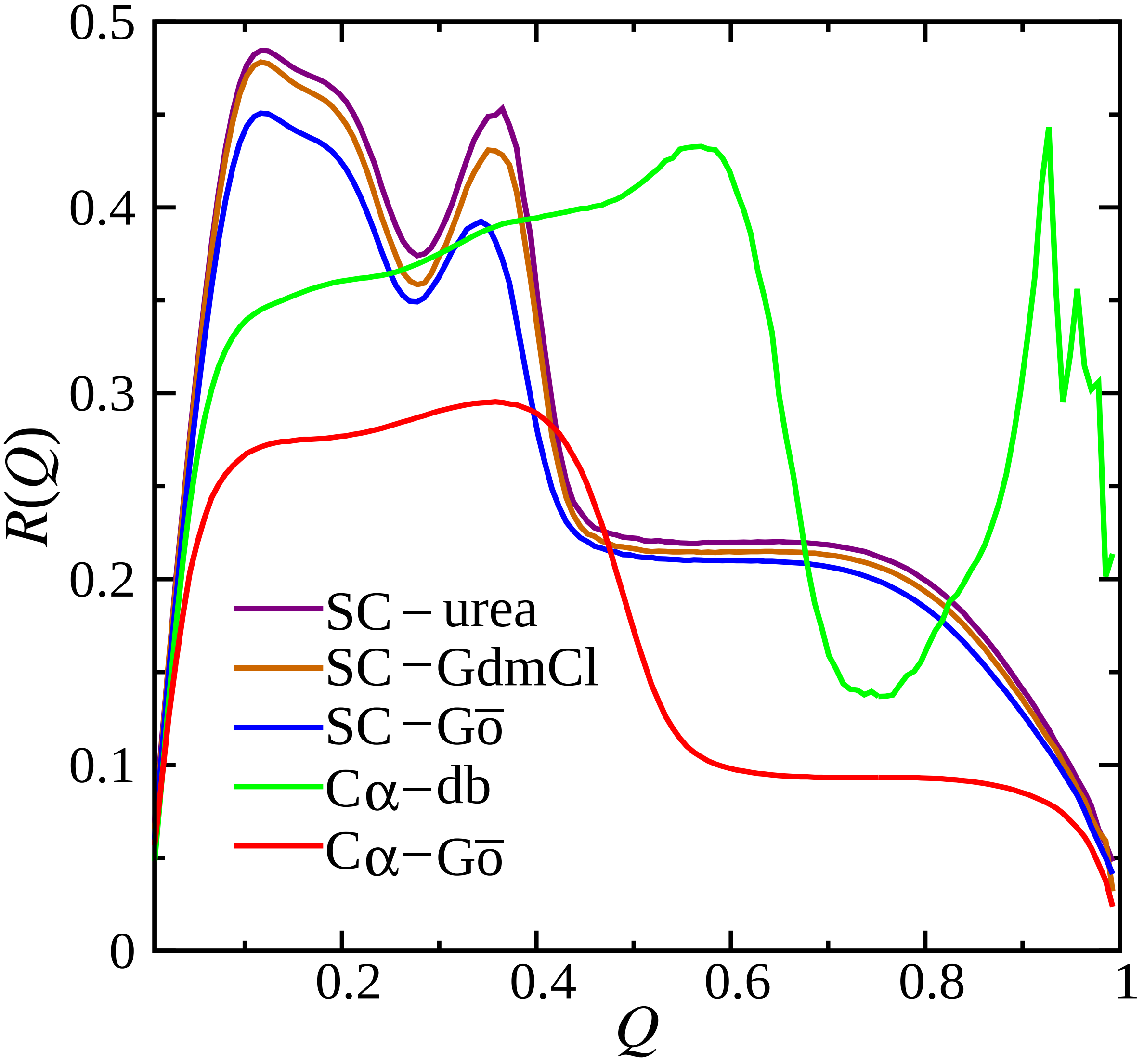

Figure 6

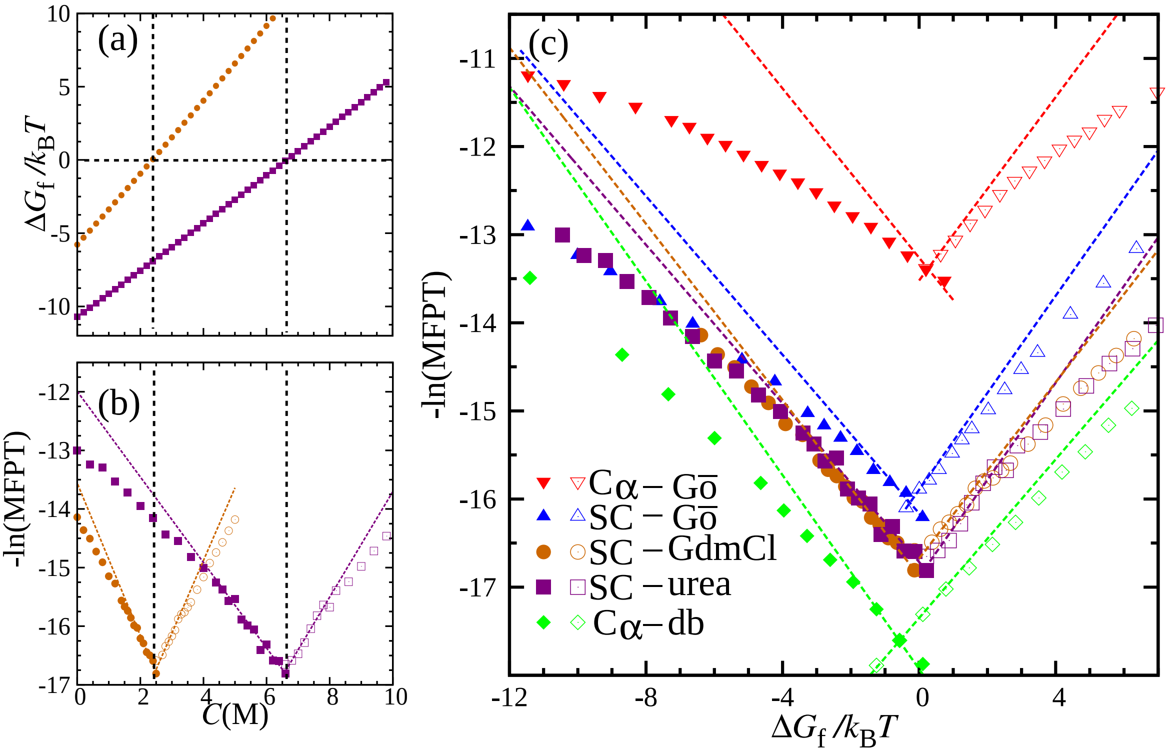
(a)
10
5
0
-5
-10
$\Delta G_f / k_B T$
(b)
-12
-13
-14
-15
-16
-17
-ln(MFPT)
0
2
4
6
8
10
C(M)
(c)
-11
-12
-13
-14
-15
-16
-17
-ln(MFPT)
-12
-8
-4
0
4
$\Delta G_f / k_B T$
Cα− Gō
SC − Gō
SC −GdmCl
SC −urea
Cα−db

Figure 7

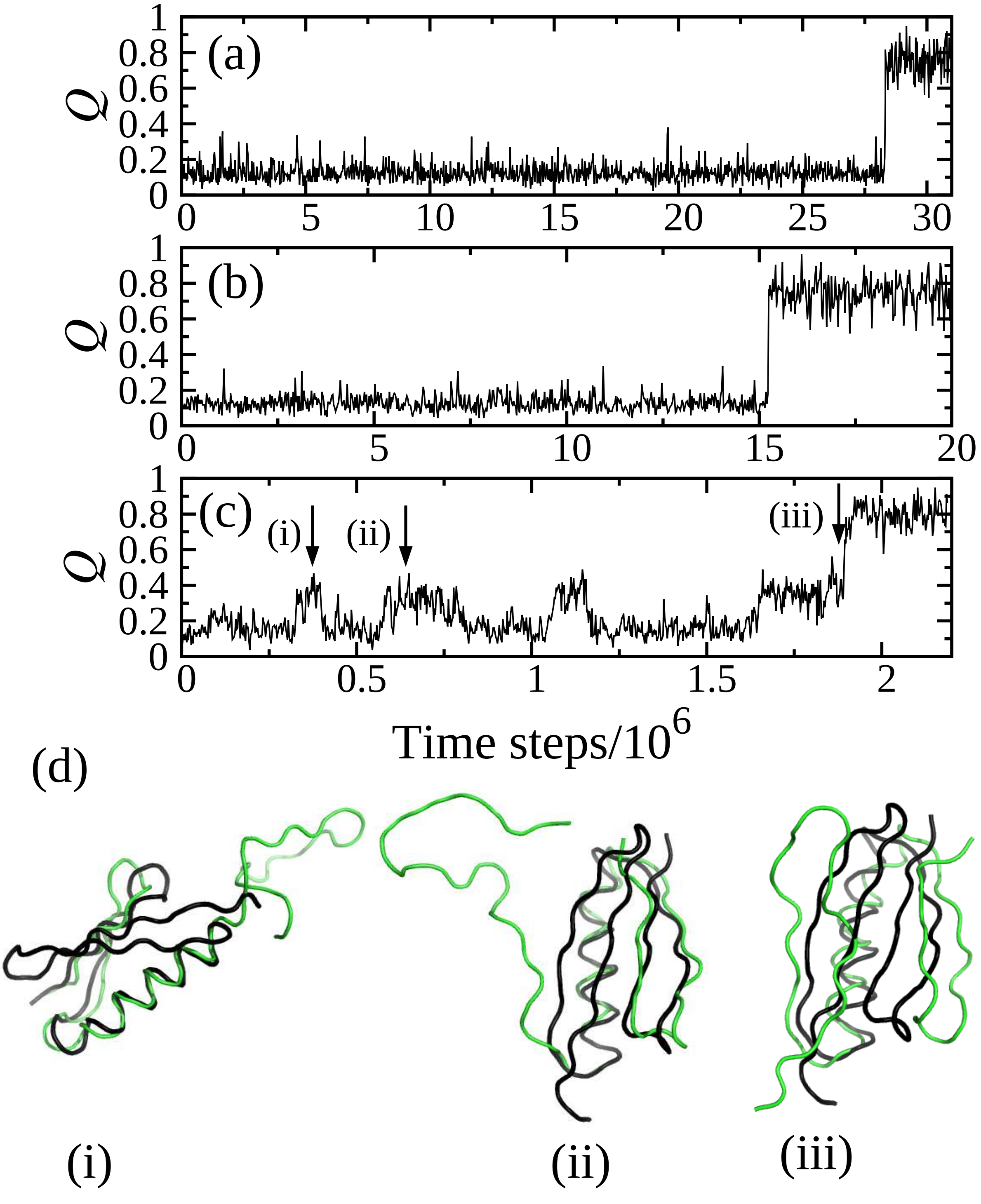
Q
1
0.8
0.6
0.4
0.2
0
(a)
0
5
10
15
20
25
30
(b)
0
5
10
15
20
(c)
(i)
(ii)
(iii)
0
0.5
1
1.5
2
Time steps/10^6
(d)
(i)
(ii)
(iii)

Figure 8

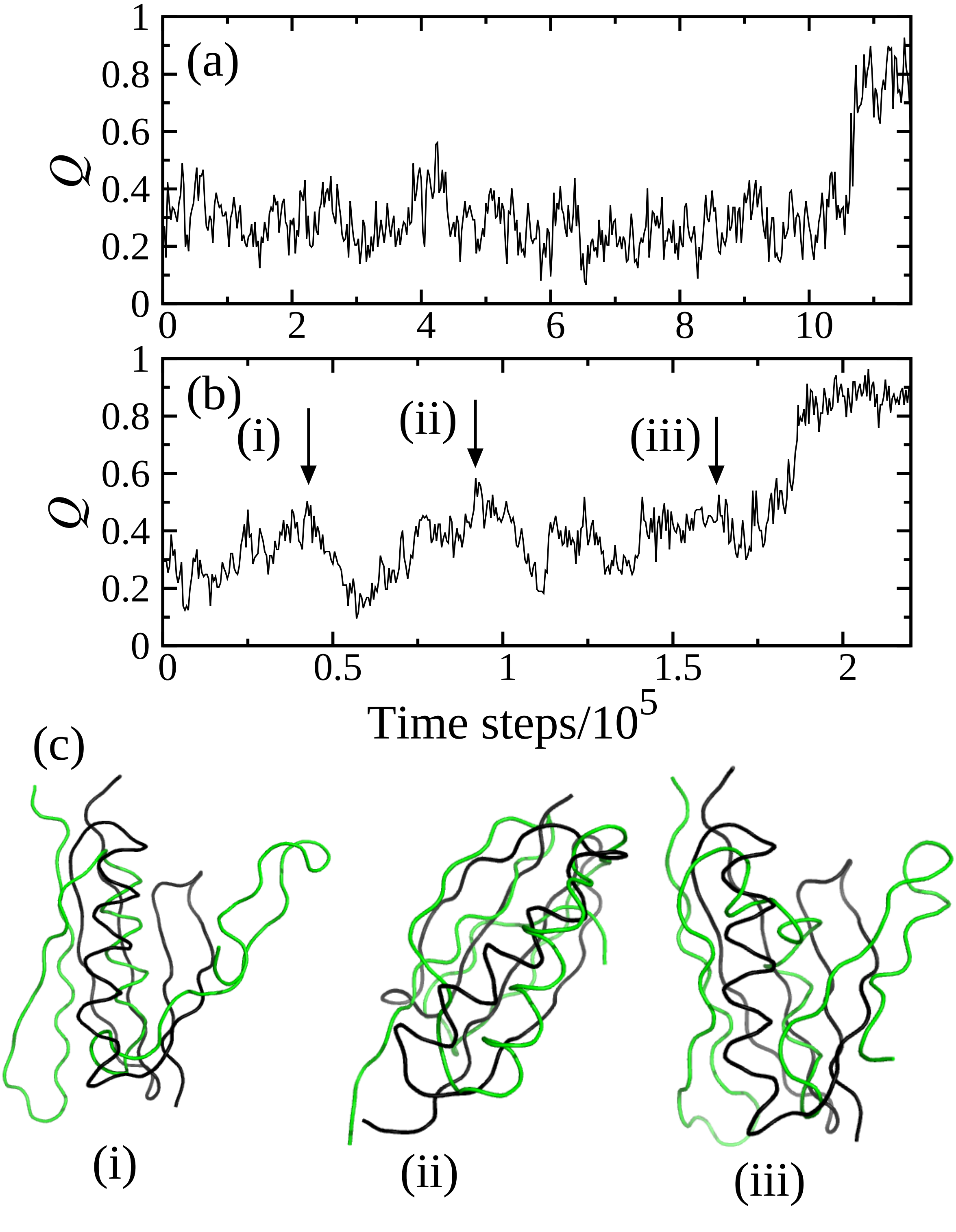
(a)
Q
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
(b)
(i)
(ii)
(iii)
Q
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
Time steps/10^5
(c)
(i)
(ii)
(iii)

Figure 9

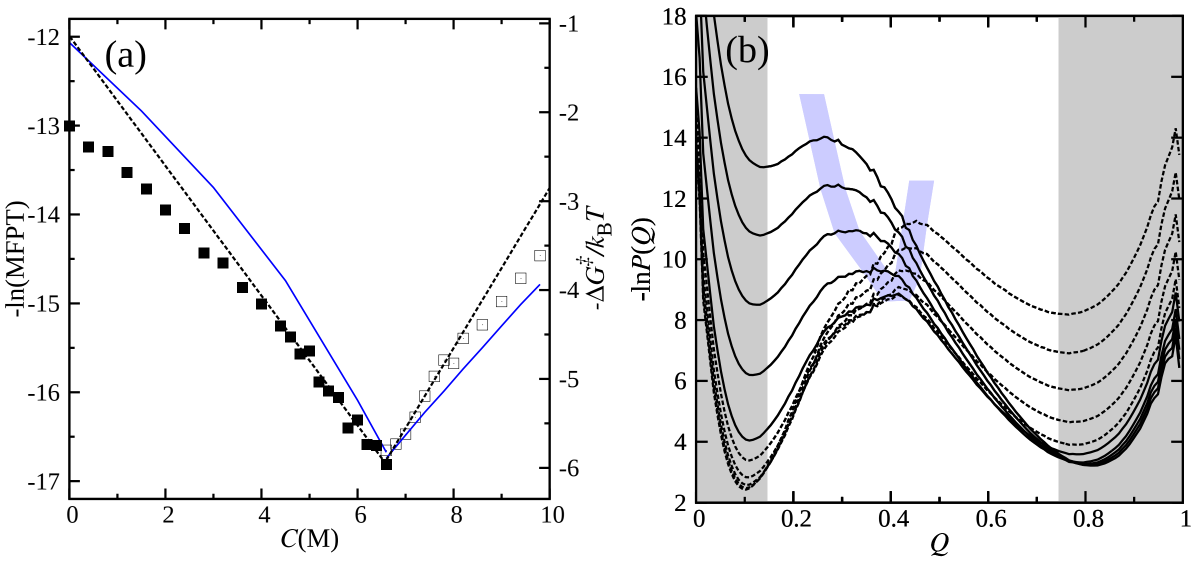
(a)
-ln(MFPT)
-12
-13
-14
-15
-16
-17
0
2
4
6
8
10
C(M)
$-\Delta G^{\ddagger}/k_BT$
-1
-2
-3
-4
-5
-6
(b)
$-\ln P(Q)$
18
16
14
12
10
8
6
4
2
0
0.2
0.4
0.6
0.8
1
Q

Figure 10

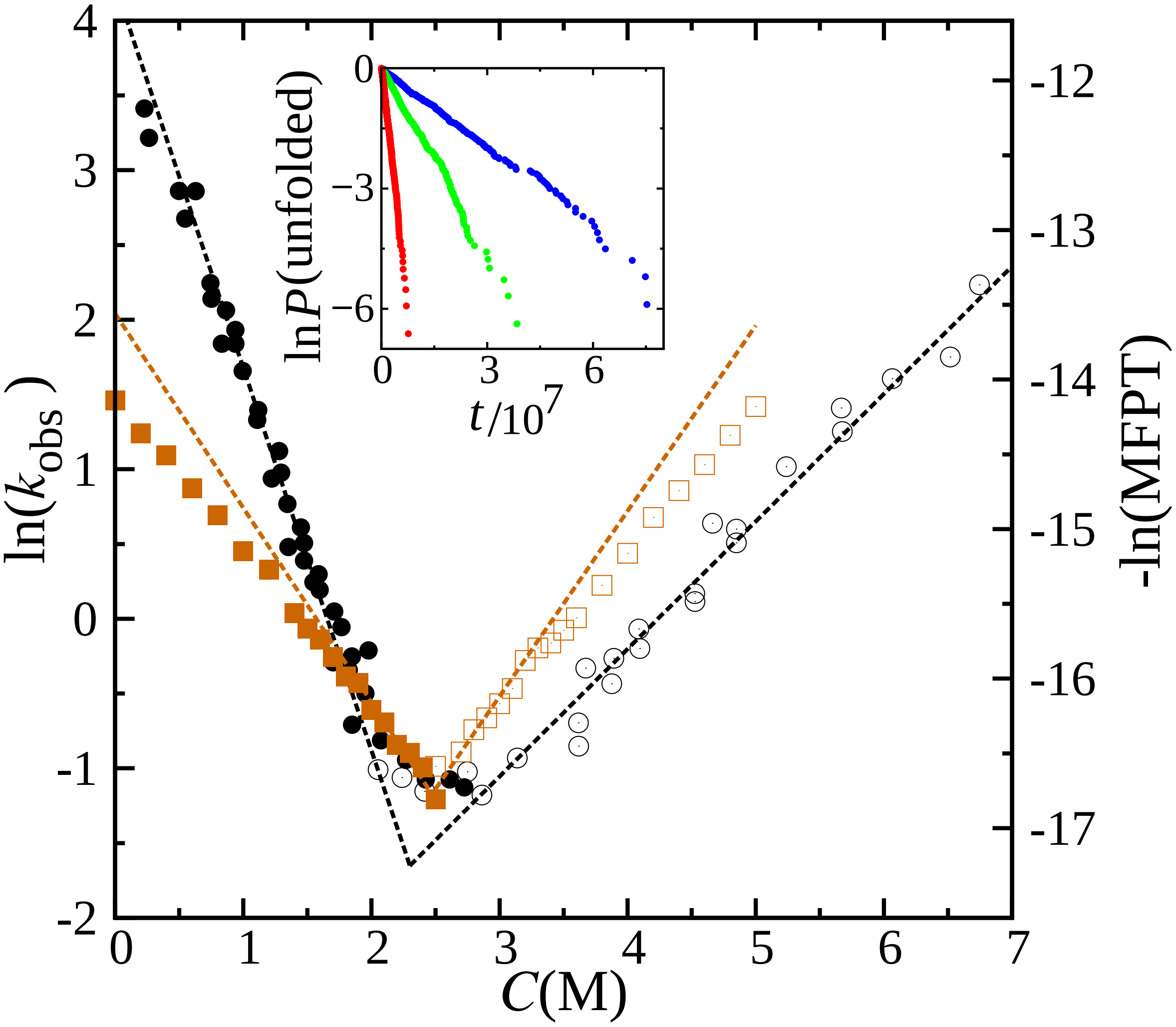
ln(k_{obs})
-ln(MFPT)
C(M)
4
3
2
1
0
-1
-2
0
1
2
3
4
5
6
7
-12
-13
-14
-15
-16
-17
ln$P$(unfolded)
t/10^7
0
−3
−6
0
3
6

Figure 11

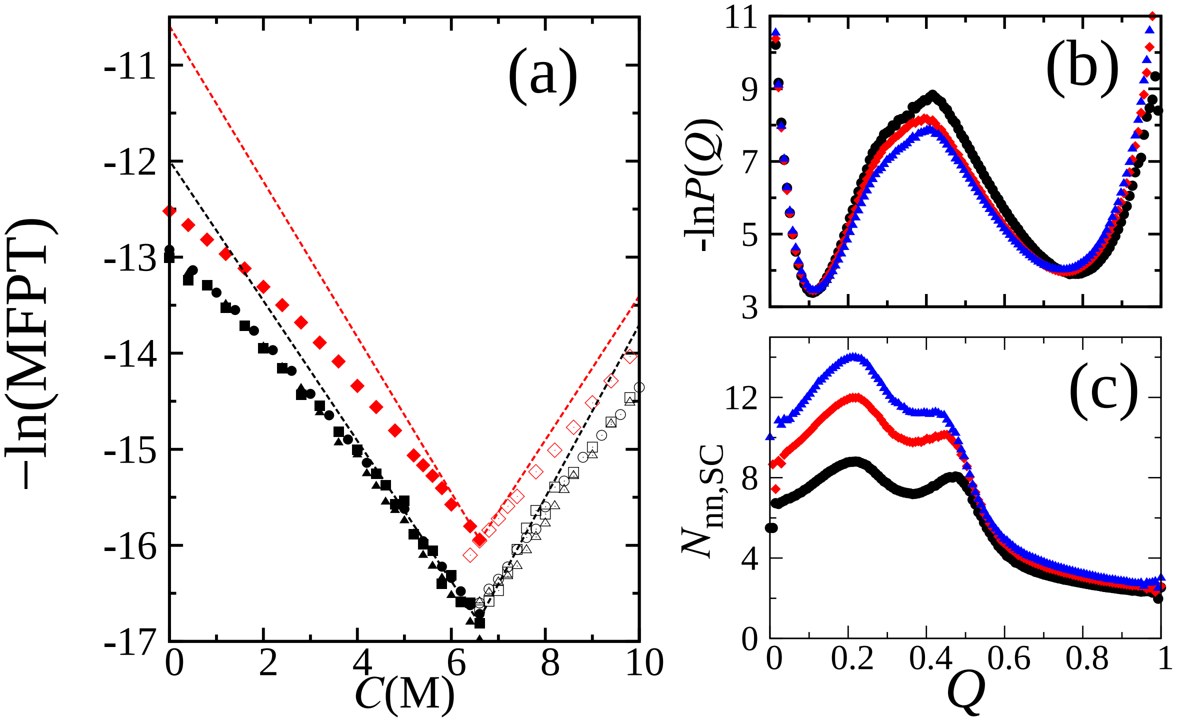
(a)
−ln(MFPT)
-11
-12
-13
-14
-15
-16
-17
0
2
4
6
8
10
C(M)
(b)
-ln$P(Q)$
11
9
7
5
3
(c)
$N_{\mathrm{nn,SC}}$
12
8
4
0
0
0.2
0.4
0.6
0.8
1
Q

Figure 12