Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding

Tao Chen[1,2,3], Jianhui Song[1,2,3] and Hue Sun Chan[1,2,3]

The diverse biological functions of intrinsically disordered proteins (IDPs) have markedly raised our appreciation of protein conformational versatility, whereas the existence of energetically favorable yet functional detrimental nonnative interactions underscores the physical limitations of evolutionary optimization. Here we survey recent advances in using biophysical modeling to gain insight into experimentally observed nonnative behaviors and IDP properties. Simulations of IDP interactions to date focus mostly on coupled folding-binding, which follows essentially the same organizing principle as the local–nonlocal coupling mechanism in cooperative folding of monomeric globular proteins. By contrast, more innovative theories of electrostatic and aromatic interactions are needed for the conceptually novel but less-explored 'fuzzy' complexes in which the functionally bound IDPs remain largely disordered.

Addresses
[1] Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[2] Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[3] Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada

Corresponding author: Chan, Hue Sun (chan@arrhenius.med.toronto.edu)

Current Opinion in Structural Biology 2015, **30**:32–42

This review comes from a themed issue on **Folding and binding**

Edited by **Annalisa Pastore** and **Guanghong Wei**

For a complete overview see the Issue and the Editorial

Available online 24th December 2014

http://dx.doi.org/10.1016/j.sbi.2014.12.002

Introduction

Our picture of how proteins function was historically dominated by the properties of folded globular proteins. Although a rigid 'lock-and-key' mechanism has long been found to be inadequate, the conformational diversity envisioned by early proponents of more dynamic modes of interactions was still rather limited. Only recent advances in experimental methods have led to a much deeper appreciation of functional protein dynamics [1]. A notable discovery in this regard is that intrinsically disordered proteins (IDPs) are involved in a wide range of biological functions [2–4].

In essence, folding and binding entail a discrimination in favor of native against nonnative interactions. Therefore, deciphering the energetics that disfavors nonnative interactions — those that are absent in folded proteins or populated negligibly in IDPs — is a key to the physical basis of molecular recognition in biology. It is intuitive, as has been evident since the earliest simulation of folding [5], that some nonnative interactions can be favorable. Although nonnative effects are sometimes not manifest experimentally, one must not lose sight of the necessity for an explanation for their non-prevalence. Indeed, as summarized below, it is probably easier to rationalize nonnative interactions that have been detected experimentally than to understand their general rarity in natural proteins.

Nonnative interactions in protein folding and misfolding

The possibility of nonnative effects was suggested as early as 1971 by Ikai and Tanford, whose analysis of cytochrome c pointed to an alternate dead-end folding pathway [6]. But such effects were not much pursued for many years. Instead, an alternate interpretation was offered for their experimental data [7]. By 1990, Kim and Baldwin concluded that 'the initial results indicate that structures in folding intermediates are nativelike' [8]. This outlook was largely affirmed by the consistency [9] and minimal frustration [10] principles in the 1980s, though both of these theoretical frameworks allow for nonnative effects: consistency between local and nonlocal interactions is imperfect [9] and energetic frustration exists albeit with minimal effects [10]. In the 1990s, the native-centric view was reinforced by the discovery of two-state proteins that fold without populated intermediates and the native-interaction-based Φ-value analysis developed to study these proteins [11].

The native-centric perspective led Gō and coworkers to develop models that have no favorable nonnative interactions [12]. These 'Gō models' are useful for exploring general folding principles even though they do not address the physico-chemical properties of a protein's amino acid sequence. When more attention was turned to the latter question in the early 1990s, non-Gō chain models embodying sequence-dependent physical effects such as hydrophobicity invariably produced significant nonnative interactions, suggesting that folding intermediates can in

fact be kinetic traps. However, quite surprisingly, the nonnative interactions predicted by these presumably more physical non-Gō models (reviewed in [13]) are much more prevalent than that observed experimentally in small, single-domain proteins. This mismatch between experiment and the early non-Gō models prompted a revival of Gō-like modeling, which has since witnessed many successful applications but has also left the cause for the failure of intuitively physical potentials to produce proteinlike cooperative behaviors largely unresolved (reviewed in [14•]).

The physical significance of potential nonnative effects is often masked by apparent success of native-centric methods

Besides explicit-chain Gō-like models, some structure-based nonexplicit-chain approaches are also native-centric in essence. One example is thermodynamic modeling based upon solvent accessible surface area (SASA) that accounts only for native but not nonnative burials [15,16], in which the change in free energy or other properties upon folding is given by the difference between their values in the Protein Data Bank (PDB) structure and in an unfolded state assumed to be either uniformly open [15] or consisting of subglobal folding units that undergo all-or-none folding–unfolding transitions [16]. These models are often successful [15,16] despite their neglect of nonnative effects and preclusion of many *a priori* possible conformations. Logically, this success hinges on, and is a clear reflection of many natural proteins' highly cooperative folding because the models cannot be successful if folding is less cooperative and nonnative effects are prevalent. Unfortunately, empirical success can also lead to complacency in that the need to address the physical origin of this remarkable folding cooperativity was often overlooked [14•].

Although prediction by models with and without favorable nonnative interactions can be very different [13], the two classes of models may produce very similar predictions for certain aspects of folding. A recent comparison of lattice models suggests that 'the order in which native contacts accumulate' during folding 'is profoundly insensitive to omission of nonnative interactions' [17]. An independent analysis of atomic simulation trajectories of several natural proteins arrives at a similar conclusion that nonnative interactions are 'irrelevant to the mechanism of folding in most cases' [18•]. What the authors [18•] mean here by 'mechanism' is quite restrictive, however. It refers primarily to the kinetic events along transition paths that go directly from the unfolded to the folded state. Because the conformational distribution along transition paths is different from that of the larger unfolded ensemble [18•,19], the authors' findings do not negate, nor do those of the lattice study [17], the significant impact nonnative interactions can have on other aspects of folding such as its overall rate [18•,20]. As such, the main message from Refs [17,18•] is that there are probably few obligatory nonnative interactions in folding. In this respect, these new findings are consistent with earlier observations that 'throughway' folding paths devoid of nonnative contacts are almost always possible, even when there are ample nonnative interactions in the unfolded ensemble such that a significant fraction of the folding trajectories is partitioned into paths with transient kinetic traps [13,21].

Nonnative interactions in the unfolded, intermediate and transition states

With improved experimental techniques (reviewed in [22]), nonnative effects have become more apparent. For instance, nonnative interactions were detected in the unfolded states of the N-terminal domain of ribosomal protein L9 (NTL9) [23] and bacterial immunity protein Im7 [24]. Nonnative interactions were also found recently in the folding intermediates of an FF domain mutant [25], ubiquitin [26], an unfolding intermediate of the SH3 domain of PI3 kinase [27], and the folding transition state of protein L [28].

Nonnative interactions are not necessarily functionally detrimental. For Im7, they are likely a consequence of functional constraints [29,30]. Nonnative interactions often slow but some can speed up folding [31]. One example is that nonnative interactions can result in more efficient folding of knotted proteins [32]. More broadly, molecular evolution can be facilitated by nonnative conformations if they perform selectable promiscuous functions that differ from a protein's dominant function(s) (reviewed in [33]).

Nonnative interactions in atomic and coarse-grained simulations

Recent success in *ab initio* folding simulation of small proteins [34] indicates that current molecular dynamics forcefields can capture reasonably well the pertinent physics. Nonetheless, nonnative effects are still significantly overestimated by current forcefields because the unfolded states they predicted are far more collapsed than those observed experimentally [35••]. Indeed, a recent comparison between experimental and simulation data on the NuG2b variant of protein G finds excess nonnative hydrogen bonding in the simulated conformational ensemble [36••], underscoring that improvements of forcefields are needed, especially if they are to capture the high folding cooperativity of natural proteins.

Much recent progress in understanding nonnative effects was achieved by using complementary coarse-grained modeling. These efforts include accounting for the nonnative effects in Im7 [24] in terms of 'localized frustration' [29] and a sequential stabilization algorithm for predicting folding pathways [37]. Consistent with experiments showing significant nonnative effects in Im7 but not its

Im9 homolog [24], the *TerItFix* algorithm fails to reach the native structure of Im7 but is successful in folding Im9 [37].

Nonnative interactions as perturbations in a native-centric framework

An intuitive approach to model nonnative effects is treating them as perturbations in native-centric models [31,38–40,41•,42•,43] by using a 'hybrid' potential (reviewed in [14•,44•])

$$E_T = E_N + E_{SD}, \quad (1)$$

where the total energy E_T is the sum of a Gō-like protein-specific E_N and a sequence-dependent E_{SD} that is transferrable across different proteins. By augmenting E_N by E_{SD}, one posits that nonnative physics cannot be eliminated entirely by evolutionary optimization [31]. Examples of proteins that have been successfully modeled by this method are provided in Figure 1, wherein the original predictions are illustrated by new simulation data from models with more energetic heterogeneity [45] and cooperativity-enhancing features [46,47•] such as desolvation barriers [14•,48–50] in the interaction potential.

Hybrid modeling has accounted for the mutational effects on Fyn SH3 folding [31,42•], offering a testable prediction that folding can be speeded up by a nonnative hydrophobic contact in the transition state between residue positions 3 and 53 [31] (Figure 1a). Nonnative interactions are seen to lower folding barriers in myristoylated hisactophilin as well [41•] (Figure 1b). The method was also applied to study nonnative electrostatics in NTL9 (Figure 1c), ribonuclease SA, β-lactalbumin, egg-white lysozyme [38], and Fyn SH3 [42•]. For the designed protein Top7, which is of considerable recent theoretical interest [44•,51•], the method rationalizes its complex folding kinetics [52,53] in terms of intermediates transiently trapped by nonnative hydrophobic interactions [39] (Figure 1d).

Nonnative effects and folding intermediates are often detected experimentally by rollovers in chevron plots of logarithmic folding/unfolding rates as functions of denaturant concentration [22]. Native stability may serve as a proxy for denaturant effects in simulated chevron plots [14•,39,47•]. A more direct treatment of denaturant dependence was enabled by a class of coarse-grained models that employ sidechain and backbone transfer free energies [54•]. By extending the largely native-centric original model [54•], a recent alternate formulation allows nonnative contacts to be promoted by decreasing denaturant [47•]. Applied to the Top7 example, the extended model produces a GdmCl-dependent chevron that exhibits a severe folding-arm rollover (Figure 2a) similar to experiment [52], a feature that cannot be captured by native-centric interactions alone (Figure 2b). Although the GdmCl dependence of native stability in the alternate molecular transfer free energy model (Figure 2a) is somewhat weaker (has a smaller m_{eq}) than that observed experimentally [52], the similarity between simulated and experimental chevrons in Figure 2a is a good indication that the physical information embodied in sidechain transfer free energies can be harnessed to rationalize experimentally detected nonnative effects.

Theory and simulation of IDPs

Interest in nonnative effects is part of a recent trend toward a better appreciation of protein conformational diversity. This development opens the conceptual possibility that Nature can exploit any protein conformation for biological function. Perhaps the best illustration of this notion is functional IDPs and intrinsically disordered regions (abbreviated collectively here as IDPs). IDPs perform many critical functions especially in signaling and regulation [3,4] by making good use of conformational disorder [55–57,58•]. Many IDPs undergo large-scale ordering or folding upon binding their partners [59], but some IDPs function in fuzzy complexes wherein the bound IDP is only locally ordered around the binding site(s) while the rest of the IDP chain remains flexible [60–63].

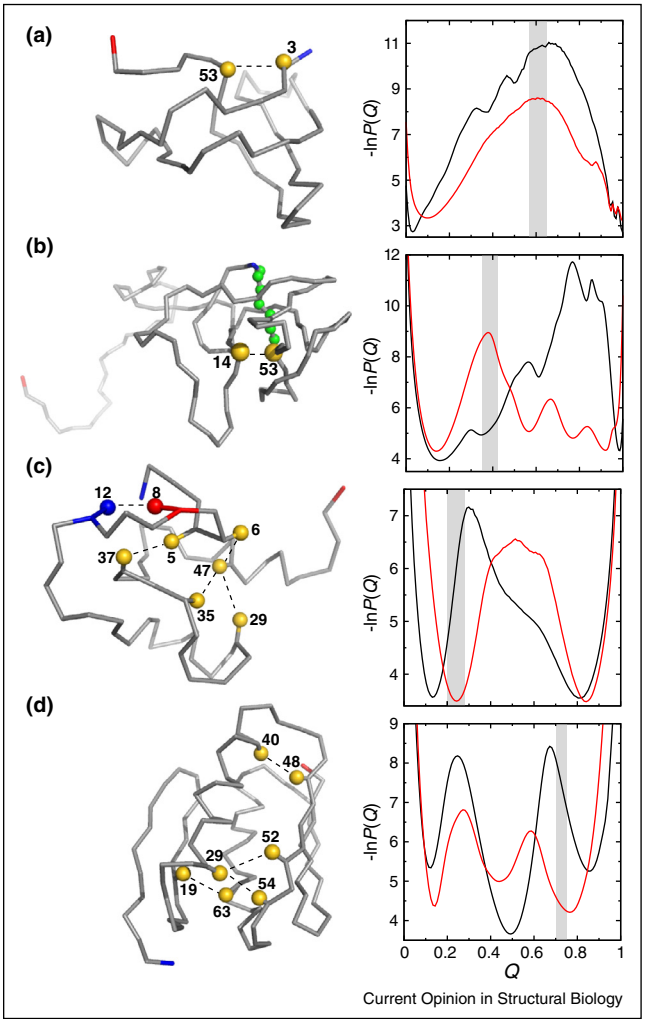
IDP studies have benefited from coordinated experimental and computational efforts [64•]. Because of the high cost of explicit-solvent molecular dynamics [65–68], many IDP simulations employ coarse-grained and/or implicit-solvent models [69,70•]. Multi-scaled techniques combining all-atom and coarse-grained simulations have also been developed [71,72•,73]. Some of the methods are adapted from folding studies: Gō-like and hybrid models (Eqn 1) are used for coupled folding-binding [74,75] and nonnative IDP interactions [76]. A coarse-grained transferrable potential for folding was successfully applied to simulate coupled folding-binding [77]. Techniques for simulating unfolding at super-high temperatures are adopted for the study of IDP unbinding as well [67].

An extensive summary of IDP simulation studies is now available in an excellent recent review [70•]. Our emphasis here is complementary. Through a brief (non-exhaustive) sketch of recent progress that includes but also goes beyond the much-studied coupled folding-binding phenomenon, we will highlight the theoretical challenges posed by IDP fuzzy complexes and the pertinent new concepts that should be further explored.

Ampholytic IDP conformational ensembles and their characterization

IDPs are described by conformational ensembles derived from experimental data and simulations (Figure 3a). A public database of IDP ensembles is now in place

Figure 1



Hybrid modeling of nonnative interactions. Each hybrid-model snapshot (*left*) contains one or more favorable nonnative contacts between hydrophobic (golden) or positively (blue) and negatively (red) charged residues. A selection of nonnative contacts are identified by dashed lines. The free energy profiles (*right*) in red (hybrid model) and black (a corresponding Gō-like model) are negative logarithm of normalized population $P(Q)$ as a function of fractional number of native contact Q computed near each model's transition midpoint. The vertical shades indicate the approximate Q values of the snapshots. **(a)** The N53I-L3I double mutant of the Fyn SH3 domain (PDB ID: 1SHF). **(b)** Myristoylated hisactophilin (PDB ID: 1HCD) with a C14 acyl chain (green beads) covalently linked to its N-terminus. The C_α Gō-like formulations in (a) and (b) follow that in Ref [39], whereas the nonnative energies are the Miyazawa–Jernigan (MJ) energies in Table 3 of Ref [45]. The snapshots in both (a) and (b) belong to the transition state. The behaviors seen here are in line with the results of Zarrine-Afsar *et al.* [31] and Shental-Bechora *et al.* [41•]. **(c)** NTL9 (PDB ID: 1CQU). The snapshot is an unfolded conformation with a nonnative D8-K12 contact that is indirectly stabilized by a cluster of nonnative hydrophobic contacts. This picture is consistent with experiment [23] and simulation by Azia and Levy [38]. **(d)** The designed protein Top7 (PDB ID: 1QYS). The snapshot shows a kinetically trapped conformation similar to I_2 in Figure 5G of Ref [39]. The sidechain (SC) models used in (c) and (d) are based on Ref [47•] but with weakened through-bond interactions ($K_\theta = K_{ch} = 5.0$ kcal mol^{-1} rad^{-2}; $K_\phi^{(1)} = 2K_\phi^{(3)} = 0.5$ kcal mol^{-1}) and purely repulsive C_α–SC terms to reduce native biases. The spatial ranges of favorable native-centric C_α–C_α and SC–SC interactions are now truncated at distance $r_{ij} = 1.2r_{ij}^0$ and $1.6r_{ij}^0$ for (c) and (d), respectively, to enhance folding cooperativity [46]. In (c), MJ energies and the form in Refs [31,39] are used for native and nonnative SC–SC interactions, respectively, whereas the additional interactions between charges at the tips of SCs are those in Ref [38]. In (d), the Gō-like black profile is from the SC-GdmCl model [47•]. The hybrid model (red profile) is a SC-$[GdmCl]^{C_0=15M}$ formulation [47•] that reduces to a SC-Gō model at a hypothetical $C_0 = 15$ M (see Figure 2).

([78] and references therein). For fuzzy complexes, an ensemble is needed not only for unbound but also for bound IDPs [62,79]. Compared to globular proteins, IDPs have more proline, glycine, polar and charged residues and fewer hydrophobic residues. Many IDPs are polyampholytes, that is, polymer chains carrying charges of both signs [80]. Recent experiments [81,82] and simulations [81,83•] on IDP polyampholytes show that their conformational dimension depends not only on their net charge [81] but also the charge pattern along the protein sequence [82,83•], whereas heat-induced collapses of IDP conformations can be rationalized by the solvation properties of their more hydrophilic residues [84•].

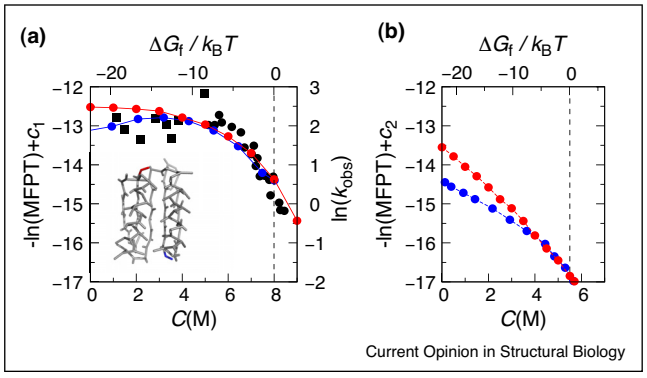
IDPs have preferred conformations and transient intrachain interactions even in their unbound states. IDPs with more hydrophobic residues tend to be relatively more ordered [85]. For IDPs in environments with a high concentration of other macromolecules, molecular crowding is expected to reduce IDP conformational dimension [86]. However, the modulation of conformational properties by crowding is not uniform along the IDP chain, thus allowing certain target-binding motifs to be selectively stabilized [87].

Local–nonlocal coupling in coupled folding-binding

There is now a rather extensive literature on simulation studies of coupled folding-binding of IDPs [69,70•] (Figure 3a–d). Several simulations of unbound IDPs show enhanced preformed structures consistent with their bound conformations [66,68,88] (Figure 3a). In line with this trend, experiments show that an increase in the helical propensity of ACTR accelerates its binding to NCBD [89], and that the binding affinities of the hub protein Keap1 to its multiple IDP partners are determined mainly by the population of bound-state-like structures in the unbound IDP ensembles [88]. However, bound-state-like structures in the unbound IDP PUMA are not necessary for its binding to MCL-1 [90].

Recent hybrid models similar to those used for folding predict that coupled folding-binding of several IDPs is significantly facilitated by electrostatics [91,92] (Figure 3c). Nonnative effects are common in all-atom IDP simulations [65,71]. Across several coarse-grained [76], implicit-solvent [93], and all-atom [71] simulations,

Figure 2



Nonnative interactions and chevron rollover. Folding arms of chevron plots for Top7 (PDB ID: 1QYS, inset) are computed in SC models with **(a)** and without **(b)** strongly favorable nonnative interactions using Langevin dynamics, where MFPT is mean first passage time of folding in units of simulation time steps [47•]. The folded and unfolded states are defined, respectively, by $Q > 0.85$ and $Q < 0.18$. Here $-\ln(\text{MFPT})$ is determined in the transfer-free-energy SC models in Figure 1d [red data points, left vertical scale with $c_1 = 0$ in (a) and $c_2 = 0$ in (b) for the SC-[GdmCl]$^{C_0=15M}$ and SC-GdmCl models, respectively] as a function of denaturant (GdmCl) concentration C (bottom horizontal scales) as well as in contact-energy-based SC models [blue data points, left vertical scale with $c_1 = 3.6$ in (a) and $c_2 = 1.8$ in (b)] as a function of native stability $\Delta G_f/k_BT$ (free energy of folding in units of Boltzmann constant k_B times absolute temperature $T = 298$ K; top horizontal scales). The C-dependent native stability of the SC models is also provided approximately by the top $\Delta G_f/k_BT$ scales. Each simulated $-\ln(\text{MFPT})$ is an average over at least 269 trajectories near the transition midpoint and up to ≈9000 trajectories for strongly folding conditions. The sampling uncertainties are at most ±0.05 and as small as ±0.0075 for large $-\ln(\text{MFPT})$s. Lines joining simulated data points are guides for the eye. The contact-energy-based SC models are constructed by the formulation in Figure 1c except explicit electrostatic interactions are not considered here. In the SC-[GdmCl]$^{C_0=15M}$ model in (a), the model [GdmCl] that we identify with experimental zero denaturant is that at which the model $\Delta G_f/k_BT$ matches the experimental Top7 value (−22.3) at [GdmCl] = 0 [53]. Accordingly, we re-define this model concentration as $C = 0$. The model transition midpoint on this scale is at $C \approx 8.0$ M [vertical dashed line in (a)], which is higher than the experimental 6.2 M [52,53]. Nonetheless, the chevron rollover of this model compares favorably with the logarithmic folding rates of Scalley-Kim and Baker [black data points from Figure 4 of [52], rate in s^{-1}, right vertical scale in (a)], which we re-plot here as a function of $\Delta G_f/k_BT$ (top scale). By contrast, for the SC-GdmCl model in (b), which is similar to the recent SC models of Thirumalai and coworkers [54•], the midpoint C is more accurately predicted at ≈5.5 M [vertical dashed line in (b)]; but this Gō-like construct fails to capture the experimental chevron rollover.

nonnative interactions are seen to lower the free energy barrier to coupled folding-binding, reminiscent of similar effects in folding [31,41•]. A recent mutation experiment on ACTR binding to NBCD suggests that the transition state contains weak native contacts but is otherwise disordered [94]. A multi-scaled simulation study of N_{TAIL} in MeV (Figure 3d) leads also to a picture that the majority of the folding is induced after the initial binding recognition [72•].

These observations suggest that coupled folding-binding is mostly folding after binding instead of selection of sparsely populated conformations [59]. This behavior is almost identical to the local–nonlocal coupling in cooperative protein folding [14•,47•]. On the basis of experimental evidence, the local–nonlocal coupling perspective posits a cooperative interplay between local structural propensity and favorable nonlocal interactions such that unfolded segments are largely disordered until native-like nonlocal contacts are essentially in place (reviewed in [14•]). In this respect, folding may be viewed as a single-chain version of 'mutually induced folding-binding' of separate peptides [77]. Because of their mechanistic similarity, insights from the study of coupled folding-binding should advance understanding of cooperative folding and vice versa.

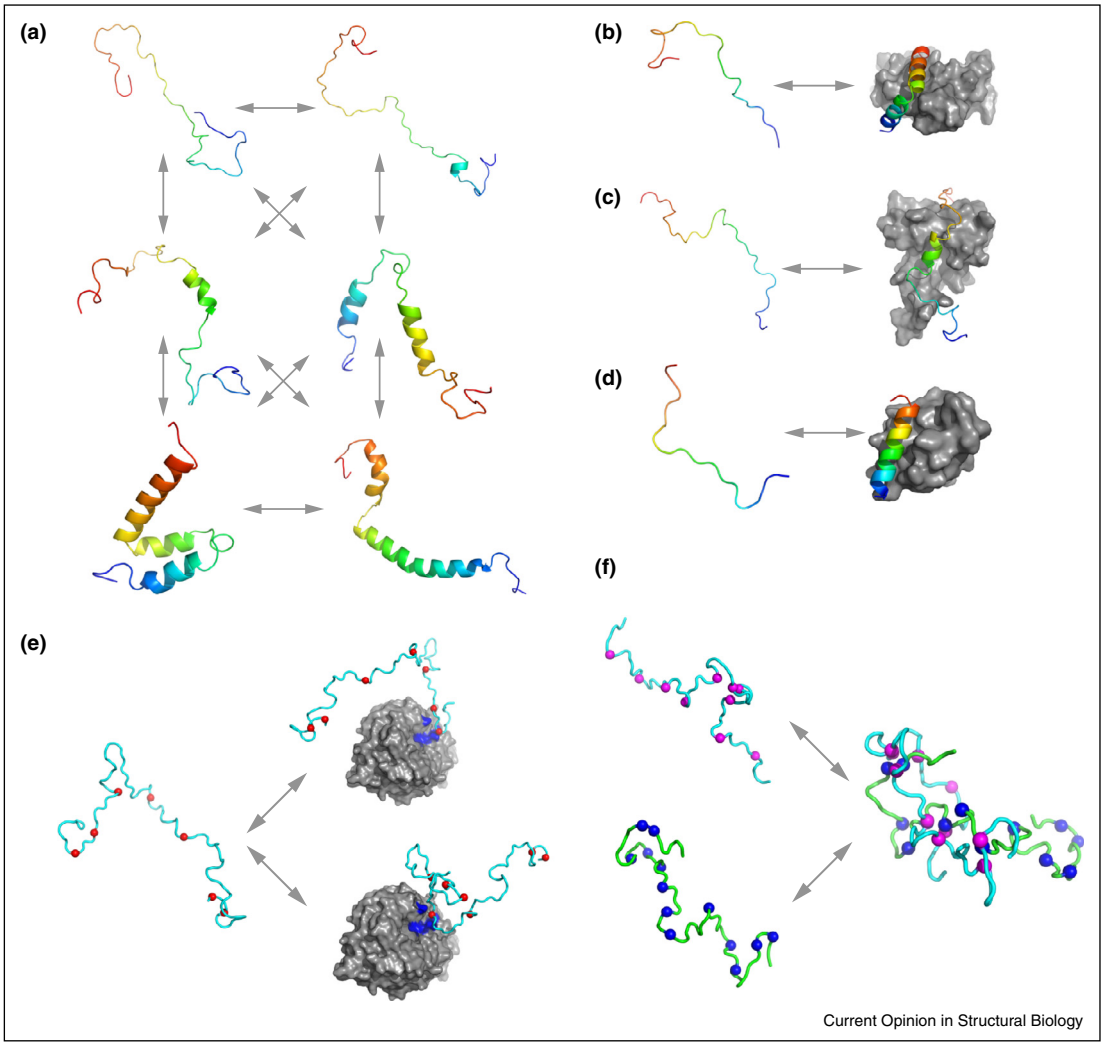
Polyelectrostatic and polyaromatic interactions in fuzzy complexes

Coupled folding-binding is not the only mode of IDP interactions. There is wide variability in IDP bound-state flexibility [4]. Fuzzy or dynamic complexes are functional IDP-folded-protein complexes in which the bound IDP remains largely disordered [60–63]. It is an open question whether fuzzy complexes of multiple disordered IDPs exist. Two fuzzy homodimers have been proposed [95,96], but the case for the cytoplasmic domain of the T-cell receptor ζ subunit [95] has recently been called into question [97].

A well-known fuzzy complex is that between the cyclin-dependent kinase inhibitor Sic1 (an IDP) with the SCF ubiquitin ligase (folded) subunit Cdc4 in yeast [62,98]. A single pocket on Cdc4 can bind to any one of the nine Cdc4-binding sites on Sic1 if the site is phosphorylated. The Sic1–Cdc4 complex is a dynamic ensemble of interconverting sub-ensembles of disordered Sic1 bound to Cdc4 via different phosphorylated sites [62]. Although only one Sic1 site can engage directly in binding at any given time, a threshold number of six phosphorylations on Sic1 is required for functional interaction with Cdc4. This phosphorylation dependence, which provides ultrasensitive cell-cycle regulation [98], has been rationalized by a mean-field model of 'polyelectrostatic' effect. In this scenario, binding is favored by multisite electrostatic attractions between the positively charged Cdc4 binding pocket and the multiple, negatively charged phosphorylated Sic1 sites even when the sites are not in direct contact with Cdc4 [60] (Figure 3e).

Multisite interactions are also implicated in oncogenic transactivation by the Ewing Sarcoma (EWS) activation domain (EAD). Mutational and *in vivo* functional experiments indicate that oncogenic activity of EAD relies on the aromatic but not the hydrophobic nature of its multiple tyrosine residues [99,100•]. The precise mechanism of EAD activity is unknown. Notwithstanding, a wide range

Figure 3



Examples of recent computational modeling of IDPs. **(a)** An ensemble of unbound NCBD, the nuclear coactivator binding domain of the transcriptional coactivator CBP (CREB-binding protein). The bottom-left conformation is similar to that in complex with the activation domain from the p160 transcriptional coactivator (ACTR; PDB ID: 1KBH). The simulation by Knott and Best [66] suggests that the two NCBD helices that provide most stabilization for the NCBD/ACTR complex are found to form readily, though transiently, in the unbound state. **(b)–(d)**: Coupled folding-binding. (b) Binding of a phosphorylated kinase-inducible domain (pKID), an IDP, with the kinase-inducible domain interacting domain (KIX; PDB ID: 1KDX). Huang and Liu [74] used explicit-chain simulation of this system to assess the fly-casting mechanism [55]. (c) Binding of IDP p53-TAD1 with TAZ2 (PDB ID: 2K8F). The simulation by Ganguly *et al.* [92] shows that electrostatic interactions probably play a significant role in accelerating the formation of this complex. (d) Binding of the IDP N_{TAIL} within the C-terminal domain of the measles virus (MeV) nucleoprotein to the X domain of the MeV phosphoprotein (PDB ID: 1T6O). The simulation by Wang *et al.* [72•] suggests that after the initial recognition events, the induced folding process of the IDP is much less cooperative than the all-or-none docking of two interacting globular proteins. **(e)** Polyelectrostatic interactions between residues 1 through 90 of the IDP cyclin-dependent kinase inhibitor Sic1 [62] and the SCF ubiquitin ligase subunit Cdc4 as envisioned by Borg *et al.* [60] (the molecular surfaces show the WD40 domain of Cdc4, PDB ID: 1NEX). The red spheres mark the negative charges at the seven phosphorylated sites along the Sic1 IDP chain, whereas the positively charged Sic1-binding pocket of WD40 is indicated by the blue patch. **(f)** Hypothetical polycation-π interactions between the aromatic (tyrosine) residues (magenta spheres) in a transactivation domain of the Ewing Sarcoma oncoprotein family (EAD) and the positively charged (arginine) residues (blue spheres) in the RGG3 sequence in the Ewing Sarcoma RNA-binding domain. The possibility that these interactions may lead to a fuzzy complex of two IDP chains (right) that remain disordered was discussed by Song *et al.* [100•] as a mechanism for out-competing disease-causing EAD interactions [100•,101].

of experimental EAD data can be rationalized, through coarse-grained explicit-chain modeling, by assuming a hypothetical binding partner with positive charges that can form favorable cation-π contacts with the EAD tyrosines [100•]. Such a 'polycation-π' perspective is consistent as well with the auto-suppression of EAD activity by the EWS RNA-binding domain [101] (Figure 3f).

Whether polycation-π effects play a role in normal, non-oncogenic EAD function is unknown. Nonetheless, in view of the biomolecular significance of cation-π and π–π interactions [102], particularly in IDP coupled folding-binding [103•], it is likely that many cases of functional multisite, polyaromatic and/or polyelectrostatic IDP interactions will be discovered. A recently noted example is that in the auto-inhibition of Ets-1, a fuzzy complex between the DNA-binding ETS domain and the disordered serine-rich region is facilitated by the aromatic residues and phosphoserines in the latter [104].

Primacy of explicit-chain simulations in theoretical protein biophysics

Several functional advantages of IDPs have been proposed through analytical modeling in the absence of an explicit representation of the protein chain. In those cases, explicit-chain models should be and have been used to delineate the physical assumptions involved.

For instance, it has been suggested via an analytical model that intrinsic disorder enhances the allosteric coupling between two ligand-binding domains of a protein provided each domain can bind only when it is folded [56]. Accordingly, the free energy differences between the binding-competent and incompetent forms are identified as folding stabilities in the model; but, as noted by the authors, other interpretations of the free energy differences are not precluded [56]. Therefore, to address the degree to which intrinsic disorder is necessary — not merely sufficient—for strong allosteric coupling, one has to explore the possibility, using explicit-chain models, of achieving the same free energy differences with binding-competent and incompetent conformations that are both folded.

Another example is the 'fly-casting' mechanism, which posits that because an unfolded protein has a larger capture radius, it binds faster than its folded counterpart. The mechanism was first proposed via an analytical model that assumes that the translational diffusion processes of folded and unfolded proteins proceed at the same rate [55]. A subsequent explicit-chain evaluation confirms that IDPs bind faster than folded proteins, but not because IDPs have larger capture radii *per se* since a larger capture radius implies slower diffusion [74]. Instead, the faster binding of IDPs is due largely to their conformational flexibility, resulting in less constraint on the binding kinetics [74,105].

Yet another often-quoted conjecture is that intrinsic disorder enhances binding specificity, even though the notion is in apparent contradiction with the perceived advantage for some IDPs to bind multiple partners (reviewed in [57,58•]). A recent comparison between explicit-chain simulation and experimental enthalpy–entropy compensation data suggests that binding interactions of IDP are more malleable than those of folded proteins and hence IDP binding in fact has lower specificity [106]. An issue that deserves further attention, however, is that the experimental folding entropy contains solvent entropy [49,50] and therefore may not correspond entirely to the simulated conformational entropy [106].

Conclusions

Nonnative effects, their selective avoidance, and IDP interactions are central to biomolecular recognition. By adding some physics back to structure-based formulations, hybrid models can assess what plausible interactions can reproduce experimental behaviors and therefore might exist in real proteins. However, while useful, the approach relies on Gō-like modeling to achieve folding and therefore should be viewed as an interim measure. Ultimately, an atomistic account of nonnative interactions — and lack thereof — is needed. In this regard, critical comparisons between nonnative interactions in simulations and experiments will be extremely useful in assessing and improving current atomic forcefields.

We have also highlighted concepts and simulation methods that apply to both protein folding and IDP interactions. IDP coupled folding-binding is seen as a case of local–nonlocal coupling, which is also a likely general mechanism of cooperative folding. Besides behaviors kindred to folding, some IDPs can form fuzzy, dynamic yet functional complexes that are probably stabilized partly by polyelectrostatic and polyaromatic interactions. Theoretical efforts to understand these new modes of interactions have barely begun but the potential impact of these efforts is tremendous. Intriguingly, multisite electrostatic and aromatic interactions of low-complexity IDP sequences are also a significant driving force for liquid–liquid phase separation [107] in the formation of hydrogels [108] or protein-dense droplets [109] that lead to functional membraneless organelles in the cell [108,107,109]. Insights gained from study of fuzzy complexes are thus likely applicable to this newly recognized relationship between cellular function and structural organization as well.

Conflict of interest

None declared.

Acknowledgements

We thank Julie Forman-Kay and Tobin Sosnick for helpful discussions and sharing their work before publication. We also benefited greatly from interactions with colleagues in the Protein Folding Consortium funded by the US National Science Foundation. This work was supported by grants to HSC from the Canadian Institutes of Health Research and the Natural Science and Engineering Research Council of Canada, which we grateful acknowledge. We also thank SHARCNET and SciNet of Compute Canada for their generous allotments of computational resources.

References and recommended reading

Papers of particular interest, published within the period of review,

have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Csermely P, Palotai R, Nussinov R: **Induced fit, conformational selection and independent dynamic segments: an extended view of binding events**. *Trends Biochem Sci* 2010, **35**:539-546.
2. Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D^2 concept**. *Annu Rev Biophys* 2008, **37**:215-246.
3. Tompa P: **Intrinsically disordered proteins: a 10-year recap**. *Trends Biochem Sci* 2012, **37**:509-516.
4. Marsh JA, Teichmann SA, Forman-Kay JD: **Probing the diverse landscape of protein flexibility and binding**. *Curr Opin Struct Biol* 2012, **22**:643-650.
5. Levitt M, Warshel A: **Computer simulation of protein folding**. *Nature* 1975, **253**:694-698.
6. Ikai A, Tanford C: **Kinetic evidence for incorrectly folded intermediate states in refolding of denatured proteins**. *Nature* 1971, **230**:100-102.
7. Hagerman PJ: **Kinetic analysis of reversible folding reactions of small proteins: application to folding of lysozyme and cytochrome c**. *Biopolymers* 1977, **16**:731-747.
8. Kim PS, Baldwin RL: **Intermediates in the folding reactions of small proteins**. *Annu Rev Biochem* 1990, **59**:631-660.
9. Gō N: **Theoretical studies of protein folding**. *Annu Rev Biophys Bioeng* 1983, **12**:183-210.
10. Bryngelson JD, Wolynes PG: **Spin glasses and the statistical mechanics of protein folding**. *Proc Natl Acad Sci U S A* 1987, **84**:7524-7528.
11. Fersht AR, Matouschek A, Serrano L: **The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding**. *J Mol Biol* 1992, **224**:771-782.
12. Taketomi H, Ueda Y, Gō N: **Studies on protein folding, unfolding and fluctuations by computer simulation. 1. The effect of specific amino acid sequence represented by specific inter-unit interactions**. *Int J Pept Protein Res* 1975, **7**:445-459.
13. Chan HS, Dill KA: **Protein folding in the landscape perspective: chevron plots and non-Arrhenius kinetics**. *Proteins* 1998, **30**:2-33.
14. • Chan HS, Zhang Z, Wallin S, Liu Z: **Cooperativity, local–nonlocal coupling, and nonnative interactions: principles of protein folding from coarse-grained models**. *Annu Rev Phys Chem* 2011, **62**:301-326.

A review that addresses (i) the importance of folding cooperativity and its conceptual link to the 'Levinthal paradox', (ii) the experimental evidence for local–nonlocal coupling, (iii) desolvation and steric dewetting as a likely physical origin of the enthalpic barrier to protein folding, and (iv) hybrid models for studying nonnative effects. This article also touches briefly upon folding of knotted proteins and IDP interactions.

15. Myers JK, Pace CN, Scholtz JM: **Denaturant *m*-values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding**. *Protein Sci* 1995, **4**:2138-2148.
16. Hilser VJ, Garcia-Moreno B, Oas TG, Kapp G, Whitten ST: **A statistical thermodynamic model of the protein ensemble**. *Chem Rev* 2006, **106**:1545-1558.
17. Gin BC, Garrahan JP, Geissler PL: **The limited role of nonnative contacts in the folding pathways of a lattice protein**. *J Mol Biol* 2009, **392**:1303-1314.
18. • Best RB, Hummer G, Eaton WA: **Native contacts determine protein folding mechanisms in atomistic simulations**. *Proc Natl Acad Sci U S A* 2013, **110**:17874-17879.

The authors analyze long equilibrium molecular dynamics trajectories of nine proteins. Nonnative contacts are present along transition paths but these nonnative contacts tend to have shorter lifetimes than those in the unfolded ensemble.

19. Zhang Z, Chan HS: **Transition paths, diffusive processes, and preequilibria of protein folding**. *Proc Natl Acad Sci U S A* 2012, **109**:20919-20924.
20. Faisca PFN, Nunes A, Travasso RD, Shakhnovich EI: **Non-native interactions play an effective role in protein folding dynamics**. *Protein Sci* 2010, **19**:2196-2209.
21. Thirumalai D, Klimov DK, Woodson SA: **Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules**. *Theor Chem Acc* 1997, **96**:14-22.
22. Brockwell DJ, Radford SE: **Intermediates: ubiquitous species on folding energy landscapes?** *Curr Opin Struct Biol* 2007, **17**:30-37.
23. Cho J-H, Meng W, Sato S, Kim EY, Schindelin H, Raleigh DP: **Energetically significant networks of coupled interactions within an unfolded protein**. *Proc Natl Acad Sci U S A* 2014, **111**:12079-12084.
24. Friel CT, Beddard GS, Radford SE: **Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design**. *J Mol Biol* 2004, **342**:261-273.
25. Korzhnev DM, Vernon RM, Religa TL, Hansen AL, Baker D, Fersht AR, Kay LE: **Non native interactions in the FF domain folding pathway from an atomic resolution structure of a sparsely populated intermediate: an NMR relaxation dispersion study**. *J Am Chem Soc* 2011, **133**:10974-10982.
26. Vallée-Bélisle A, Michnick SW: **Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis**. *Nat Struct Mol Biol* 2012, **19**:731-736.
27. Dasgupta A, Udgaonkar JB: **Transient non-native burial of a Trp residue occurs initially during the unfolding of a SH3 domain**. *Biochemistry* 2012, **51**:8226-8234.
28. Yoo TY, Adhikari A, Xia Z, Huynh T, Freed KF, Zhou R, Sosnick TR: **The folding transition state of protein L is extensive with nonnative interactions (and not small and polarized)**. *J Mol Biol* 2012, **420**:220-234.
29. Sutto L, Latzer J, Hegler JA, Ferreiro DU, Wolynes PG: **Consequences of localized frustration for the folding mechanism of the Im7 protein**. *Proc Natl Acad Sci U S A* 2007, **104**:19825-19830.
30. Friel CT, Smith DA, Vendruscolo M, Gsponer J, Radford SE: **The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints**. *Nat Struct Mol Biol* 2009, **16**:318-324.
31. Zarrine-Afsar A, Wallin S, Neculai AM, Neudecker P, Howell PL, Davidson AR, Chan HS: **Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding**. *Proc Natl Acad Sci U S A* 2008, **105**:9999-10004.
32. Škrbić T, Micheletti C, Faccioli P: **The role of non-native interactions in the folding of knotted proteins**. *PLoS Comput Biol* 2012, **8**:e1002504.
33. Sikosek T, Chan HS: **Biophysics of protein evolution and evolutionary protein biophysics**. *J R Soc Interface* 2014, **11**:20140419.
34. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE: **How fast-folding proteins fold**. *Science* 2011, **334**:517-520.
35. •• Piana S, Klepeis JL, Shaw DE: **Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations**. *Curr Opin Struct Biol* 2014, **24**:98-105.

This review summarizes recent success in simulating folding rates and known native structures. It also highlights insufficiency of current force-fields in rationalizing the large enthalpies of unfolding and high degrees of disorder of unfolded conformations for some proteins. Consequently, the authors emphasize using experimental 'structural properties of disordered states and the temperature dependence of folding stability' as one of the criteria to assess future forcefield modifications. From our vantage point, the observations made in this review indicate clearly that current potential functions need to be improved before they can fully capture the high degrees of folding cooperativity of many natural proteins.

36. •• Skinner JJ, Yu W, Gichana EK, Baxa MC, Hinshaw J, Freed KF, Sosnick TR: **Benchmarking all-atom simulations using hydrogen exchange**. *Proc Natl Acad Sci U S A* 2014, **111**:15975-15980.
This study compares the extent of intra-molecular hydrogen bonding predicted by long-time molecular dynamics against experimental data from hydrogen exchange and other biophysical measurements for a protein G variant. The simulated unfolded state is found to be more collapsed and contains excess intra-molecular hydrogen bonding when benchmarked against experiment. Thus, simulated folding is apparently less cooperative than the real process. On the basis of these observations, the authors suggest that the forcefields require improvement and that comparisons such the ones they conducted should provide a general protocol for evaluating simulation results.

37. Adhikari AN, Freed KF, Sosnick TR: **De novo prediction of protein folding pathways and structure using the principle of sequential stabilization**. *Proc Natl Acad Sci U S A* 2012, **109**:17442-17447.

38. Azia A, Levy Y: **Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt**. *J Mol Biol* 2009, **393**:527-542.

39. Zhang Z, Chan HS: **Competition between native topology and nonnative interactions in simple and complex folding kinetics of natural and designed proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:2920-2925.

40. Enciso M, Rey A: **Improvement of structure-based potentials for protein folding by native and nonnative hydrogen bonds**. *Biophys J* 2011, **101**:1474-1482.

41. • Shental-Bechor D, Smith MTJ, MacKenzie D, Broom A, Marcovitz A, Ghashut F, Go C, Bralha F, Meiering EM, Levy Y: **Nonnative interactions regulate folding and switching of myristoylated protein**. *Proc Natl Acad Sci U S A* 2012, **109**:17839-17844.
A combined experimental–computational study of the effects of myristoyl-related hydrophobic interactions and nonnative electrostatic interactions on the behaviors of hisactophilin. Stabilization of the folded protein by myristoylation is demonstrated consistently by hybrid model simulation and experiment.

42. • Zarrine-Afsar A, Zhang Z, Schweiker KL, Makhatadze GI, Davidson AR, Chan HS: **Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain**. *Proteins* 2012, **80**:858-870.
A combined experimental–computational study of nonnative electrostatic interactions that rationalizes the increased folding rates of six Fyn SH3 domain variants.

43. Contessoto VG, Lima DT, Oliveira RJ, Bruni AT, Chahine J, Leite VBP: **Analyzing the effect of homogeneous frustration in protein folding**. *Proteins* 2013, **81**:1727-1737.

44. • Yadahalli S, Rao VVHG, Gosavi S: **Modeling non-native interactions in designed proteins**. *Israel J Chem* 2014, **54**:1230-1240.
A review of three hybrid modeling approaches to study the nonnative effects in the folding of designed protein Top7. Despite variations across models, the consensus structural properties of the folding intermediates predicted by the three models are found to be consistent with experiment.

45. Miyazawa S, Jernigan RL: **Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading**. *J Mol Biol* 1996, **256**:623-644.

46. Kaya H, Uzunoğu Z, Chan HS: **Spatial ranges of driving forces are a key determinant of protein folding cooperativity and rate diversity**. *Phys Rev E* 2013, **22**:044701.

47. • Chen T, Chan HS: **Effects of desolvation barriers and sidechains on local–nonlocal coupling and chevron behaviors in coarse-grained models of protein folding**. *Phys Chem Chem Phys* 2014, **16**:6460-6479.
This study explores several possible physical origins of local–nonlocal coupling in cooperative folding, and provides new alternate formulations of the molecular transfer model that allow for significantly favorable, physics-based nonnative interactions.

48. Cheung MS, García AE, Onuchic JN: **Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse**. *Proc Natl Acad Sci U S A* 2002, **99**:685-690.

49. Moghaddam MS, Shimizu S, Chan HS: **Temperature dependence of three-body hydrophobic interactions: potential of mean force, enthalpy, entropy, heat capacity, and nonadditivity**. *J Am Chem Soc* 2005, **127**:303-316.

50. Dias CL, Chan HS: **Pressure-dependent properties of elementary hydrophobic interactions: ramifications for activation properties of protein folding**. *J Phys Chem B* 2014, **118**:7488-7509.

51. • Truong HH, Kim BL, Schafer NP, Wolynes PG: **Funneling and frustration in the energy landscapes of some designed and simplified proteins**. *J Chem Phys* 2013, **139**:121908.
Energy landscapes of designed sequences are found to be more complex than that of a natural sequence. Compared to the designed sequences considered, the natural sequence appears to be more apt to exploit subtle physical differences among different types of amino acids to achieve a funnel-like landscape. Consequently, structure prediction using reduced amino-acid alphabets is less accurate for the natural sequence than for the designed sequences.

52. Scalley-Kim M, Baker D: **Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection**. *J Mol Biol* 2004, **338**:573-583.

53. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, Baker D: **The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection**. *Cell* 2007, **128**:613-624.

54. • Thirumalai D, Liu Z, O'Brien EP, Reddy G: **Protein folding: from theory to practice**. *Curr Opin Struct Biol* 2013, **23**:22-29.
An insightful review of recent advances in matching coarse-grained protein folding simulations with experiments. It summarizes several applications of the original molecular transfer model (MTM) that provides a direct modeling of denaturant dependence, including an MTM-simulated chevron plot.

55. Shoemaker BA, Portman JJ, Wolynes PG: **Speeding molecular recognition by using the folding funnel: the fly-casting mechanism**. *Proc Natl Acad Sci U S A* 2000, **97**:8868-8873.

56. Hilser VJ, Thompson EB: **Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins**. *Proc Natl Acad Sci U S A* 2007, **104**:8311-8315.

57. Zhou H-X: **Intrinsic disorder: signaling via highly specific but short-lived association**. *Trends Biochem Sci* 2012, **37**:43-48.

58. • Liu Z, Huang Y: **Advantages of proteins being disordered**. *Protein Sci* 2014, **23**:539-550.
An extensive summary of nine often-invoked proposed functional advantages of IDPs with updated pertinent experimental and simulation data.

59. Kiefhaber T, Bachmann A, Jensen KS: **Dynamics and mechanisms of coupled protein folding and binding reactions**. *Curr Opin Struct Biol* 2012, **22**:21-29.

60. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS: **Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity**. *Proc Natl Acad Sci U S A* 2007, **104**:9650-9655.

61. Tompa P, Fuxreiter M: **Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions**. *Trends Biochem Sci* 2008, **33**:2-8.

62. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD: **Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase**. *Structure* 2010, **18**:494-506.

63. Fuxreiter M, Simon I, Bondos S: **Dynamic protein–DNA recognition: beyond what can be seen**. *Trends Biochem Sci* 2011, **36**:415-423.

64. • Chen J: **Towards the physical basis of how intrinsical disorder mediates protein function**. *Arch Biochem Biophys* 2012, **524**:123-131.
An informative review of both experimental and computational approaches for studying IDPs.

65. Higo J, Nishimura Y, Nakamura H: **A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations**. *J Am Chem Soc* 2011, **133**:10448-10458.

66. Knott M, Best RB: **A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations**. *PLoS Comput Biol* 2012, **8**:e1002605.

67. Qin F, Ye W, Chen Y, Chen X, Li Y, Zhang J, Chen H-F: **Specific recognition between intrinsically disordered LEF and DNA**. *Phys Chem Chem Phys* 2012, **14**:538-545.

68. Mittal J, Yoo TH, Georgiou G, Truskett TM: **Structural ensemble of an intrinsically disordered polypeptide**. *J Phys Chem B* 2013, **117**:118-124.

69. Zhang W, Ganguly D, Chen J: **Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins**. *PLoS Comput Biol* 2012, **8**:e1002353.

70. • Baker CM, Best RB: **Insights into the binding of intrinsically disordered proteins from molecular dynamics simulation**. *WIREs Comput Mol Sci* 2014, **4**:182-198.
An extensive and timely review of the expanding field of computational studies of IDP coupled folding–binding.

71. Wang J, Wang Y, Chu X, Hagen S, Han W, Wang E: **Multi-scaled explorations of binding-induced folding of intrinsically disordered protein inhibitor IA3 to its target enzyme**. *PLoS Comput Biol* 2011, **7**:e1001118.

72. • Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E, Wang J: **Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein**. *Proc Natl Acad Sci U S A* 2013, **110**:E3743-E3752.
Simulation results from this study suggest that the evolution from the IDP encounter complex to the final bound complex can resemble an essentially 'downhill' process with low free energy barriers.

73. Terakawa T, Higo J, Takada S: **Multi-scale ensemble modeling of modular proteins with intrinsically disordered linker regions: application to p53**. *Biophys J* 2014, **107**:721-729.

74. Huang Y, Liu Z: **Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism**. *J Mol Biol* 2009, **393**:1143-1159.

75. Knott M, Best RB: **Discriminating binding mechanisms of an intrinsically disordered protein via a multi-state coarse-grained model**. *J Chem Phys* 2014, **140**:175102.

76. Huang Y, Liu Z: **Nonnative interactions in coupled folding and binding processes of intrinsically disordered proteins**. *PLoS ONE* 2010, **5**:e15375.

77. Bhattacherjee A, Wallin S: **Coupled folding-binding in a hydrophobic/polar protein model: impact of synergistic folding and disordered flanks**. *Biophys J* 2012, **102**:569-578.

78. Varadi M *et al.*: **pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins**. *Nucleic Acids Res* 2014, **42**:D326-335.

79. Jin F, Yu C, Lai L, Liu Z: **Ligand clouds around protein clouds: a scenario of ligand binding with intrinsically disordered proteins**. *PLoS Comput Biol* 2013, **9**:e1003249.

80. Higgs PG, Joanny JF: **Theory of polyampholyte solutions**. *J Chem Phys* 1991, **94**:1543-1554.

81. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV: **Net charge per residue modulates conformational ensembles of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:8183-8188.

82. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D, Schuler B: **Charge interactions can dominate the dimensions of intrinsically disordered proteins**. *Proc Natl Acad Sci U S A* 2010, **107**:14609-14614.

83. • Das RK, Pappu RV: **Conformations of intrinsically disordered proteins are influenced by linear sequence distribution of oppositely charged residues**. *Proc Natl Acad Sci U S A* 2013, **110**:13392-13397.
A simulation study of thirty 50mer sequences of equal numbers of lysines and glutamic acids. Different arrangements of the two oppositely charged residues along the chain sequence lead to significantly different patterns of dependence of inter-residue spatial distance on sequence separation.

84. • Wuttke R, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, Schuler B: **Temperature-dependent solvation modulates the dimensions of disordered proteins**. *Proc Natl Acad Sci U S A* 2014, **111**:5213-5218.
A combined experimental–computational study of five unfolded proteins or IDPs showing that the temperature dependence of their radii of gyration can be modeled approximately using experimentally determined temperature-dependent solvation free energies of their constituent amino acid residues.

85. Espinoza-Fonseca LM: **Leucine-rich hydrophobic clusters promote folding of the N-terminus of the intrinsically disordered transactivation domain of p53**. *FEBS Lett* 2009, **583**:556-560.

86. Qin S, Zhou H-X: **Effects of macromolecular crowding on the conformational ensembles of disordered proteins**. *J Phys Chem Lett* 2013, **4**:3429-3434.

87. Cino EA, Karttunen M, Choy W-Y: **Effects of molecular crowding on the dynamics of intrinsically disordered proteins**. *PLoS ONE* 2012, **7**:e49876.

88. Cino EA, Killoran RC, Karttunen M, Choy W-Y: **Binding of disordered proteins to a protein hub**. *Sci Rep* 2013, **3**:2305.

89. Iešmantavičius V, Dogan J, Jemth P, Teilum K, Kjaergaard M: **Helical propensity in an intrinsically disordered protein accelerates ligand binding**. *Angew Chem Int Ed* 2014, **53**:1548-1551.

90. Rogers JM, Wong CT, Clarke J: **Couple folding and binding of the disordered protein PUMA does not require particular residual structure**. *J Am Chem Soc* 2014, **136**:5197-5200.

91. Levy Y, Onuchic JN, Wolynes PG: **Fly-casting in protein–DNA binding: frustration between protein folding and electrostatics facilitates target recognition**. *J Am Chem Soc* 2007, **129**:738-739.

92. Ganguly D, Zhang W, Chen J: **Electrostatically accelerated encounter and folding for facile recognition of intrinsically disordered proteins**. *PLoS Comput Biol* 2013, **9**:e1003363.

93. Naganathan AN, Orozco M: **The conformational landscape of an intrinsically disordered DNA-binding domain of a transcription regulator**. *J Phys Chem B* 2013, **117**:13842-13850.

94. Dogan J, Mu X, Engström A, Jemth P: **The transition state structure for coupled binding and folding of disordered protein domains**. *Sci Rep* 2013, **3**:2076.

95. Sigalov AB, Zhuravleva AV, Orekhov VYu: **Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form**. *Biochimie* 2007, **89**:419-421.

96. Danielsson J, Liljedahl L, Barany-Wallje E, Sonderby P, Kristensen LH, Martinez-Yamout MA, Dyson HJ, Wright PE, Poulsen FM, Maler L, Graslund A, Kragelund BB: **The intrinsically disordered RNR inhibitor Sml1 is a dynamic dimer**. *Biochemistry* 2008, **47**:13428-13437.

97. Nourse A, Mittag T: **The cytoplasmic domain of the T-cell receptor zeta subunit does not form disordered dimers**. *J Mol Biol* 2014, **426**:62-70.

98. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, Sicheri F, Pawson T, Tyers M: **Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication**. *Nature* 2001, **414**:514-521.

99. Ng KP, Potikyan G, Savene ROV, Denny CT, Uversky VN, Lee KAW: **Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins**. *Proc Natl Acad Sci U S A* 2007, **104**:479-484.

100 • Song J, Ng SC, Tompa P, Lee KAW, Chan HS: **Polycation-π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family**. *PLoS Comput Biol* 2013, **9**:e1003239.
Functional experiments and model simulation data on an extensive set of EAD sequence variants provide evidence that the oncogenic transactivation activity of EAD is largely underlied by favorable multisite cation-π interactions.

101. Alex D, Lee KAW: **RGG-boxes of the EWS oncoprotein repress a range of transcriptional activation domains**. *Nucleic Acids Res* 2005, **33**:1323-1331.

102. Meyer EA, Castellano RK, Diederich F: **Interactions with aromatic rings in chemical and biological recognition**. *Angew Chem Int Ed* 2003, **42**:1210-1250.

103 • Espinoza-Fonseca LM: **Aromatic residues link binding and function of intrinsically disordered proteins**. *Mol Biosyst* 2012, **8**:237-246.
An extensive survey of 77 PDB structures of protein–protein complexes involving IDPs. Approximately 40% of the complexes are found to have at least one intermolecular π–π interaction.

104. Desjardins G, Meeker CA, Bhachech N, Currie SL, Okon M, Graves BJ, McIntosh LP: **Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor**. *Proc Natl Acad Sci U S A* 2014, **111**:11019-11024.

105. Trizac E, Levy Y, Wolynes PG: **Capillarity theory for the fly-casting mechanism**. *Proc Natl Acad Sci U S A* 2010, **107**:2746-2750.

106. Huang Y, Liu Z: **Do intrinsically disordered proteins possess high specificity in protein–protein interactions?** *Chem Eur J* 2013, **19**:4462-4467.

107. Li P, Banjade S, Cheng H-C, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, Russo PS, Jiang Q, Nixon BT, Rosen MK: **Phase transitions in the assembly of multivalent signalling proteins**. *Nature* 2012, **483**:336-340.

108. Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL: **Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels**. *Cell* 2012, **149**:753-767.

109. Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs T, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ: **Phase transition of a disordered Nuage protein generates environmentally responsive membraneless organelles**. *Mol Cell* 2014. (in press).