Many cellular proteins are intrinsically disordered and undergo folding, in whole or in part, upon binding to their physiological targets. The past few years have seen an exponential increase in papers describing characterization of intrinsically disordered proteins, both free and bound to targets. Although NMR spectroscopy remains the favored tool, a number of new biophysical techniques are proving exceptionally useful in defining the limits of the conformational ensembles. Advances have been made in prediction of the recognition elements in disordered proteins, in elucidating the kinetics and mechanism of the coupled folding and binding process, and in understanding the role of post-translational modifications in tuning the biological response. Here we review these and other recent advances that are providing new insights into the conformational propensities and interactions of intrinsically disordered proteins and are beginning to reveal general principles underlying their biological functions.

Introduction

Since the previous review on this subject in this journal [1], significant progress has been made in the characterization of processes that involve the coupling of folding of intrinsically disordered (sometimes termed unstructured) domains with binding to a target molecule, which may itself be more or less disordered. Intrinsically disordered proteins (IDPs) are frequently associated with cellular control mechanisms and signaling, and have been identified at the center of ‘hubs’ in protein interaction networks [2]. Interaction with a wide variety of targets therefore appears to be a hallmark of functional disordered proteins. Analyses of correlations between predicted protein disorder and protein function, as assessed by Gene Ontology or Swiss-Prot functional keywords, reveals that disorder is associated with a broad repertoire of biological functions and processes [3–6]. In the present review we survey the most exciting advances that have occurred over the past two years.

Technical advances

Intrinsically disordered proteins were initially identified experimentally through spectroscopic means [7], as well as theoretically by sequence analysis [8]. Prediction of disordered regions through bioinformatic searches has reached a high degree of sophistication and numerous web servers are available for prediction of disordered regions of protein sequences (reviewed in [9]).

Although information on the conformational propensities of intrinsically disordered domains both free in solution and in complex with targets has been mostly obtained by CD, fluorescence and NMR spectroscopy [10], other experimental methods have recently been employed to provide important insights into the conformational ensemble in the disordered state and into the coupled folding and binding process. These include small angle X-ray scattering (SAXS) [11,12*], paramagnetic spin labeling in combination with ensemble molecular dynamics [13] and single molecule fluorescence [14*]. High-speed atomic force microscopy [15] and Raman optical activity [16] have been used to characterize and classify intrinsically disordered proteins and provide insight into the nature of their conformational ensembles and dynamic properties. The combination of NMR spectroscopy and SAXS provides an especially powerful approach to characterize the structural ensemble of intrinsically disordered proteins, as illustrated by the recent determination of the structure of the full-length p53 tumor suppressor and its intrinsically disordered transactivation domain [17**]. While X-ray crystallography and NMR remain the primary techniques used to characterize the structures formed by intrinsically disordered proteins when bound to their targets, spectroscopic techniques such as EPR can be applied to study folding of a disordered protein labeled at specific sites with a paramagnetic nitroxide spin probe [18,19].

Coupled folding and binding of IDPs

Although it has long been recognized that many intrinsically disordered protein domains fold upon binding to their targets [20], the molecular principles and broad repertoire of interactions are only now becoming understood. While some proteins are completely disordered in the absence of their physiological partner and fold into globular structures only upon binding, most coupled
folding and binding events involve relatively short amphipathic motifs contained within longer disordered sequences [1,21]. Indeed, it is frequently possible to identify these folding motifs, which have been termed molecular recognition elements or MoRFs, by bioinformatic analysis of the protein sequence [22]. These recognition motifs can fold into helix, β-strand, or form irregular structure on binding to a target protein. Binding of disordered linear motifs by β-strand addition has been reviewed recently [23]. A survey of the intermolecular interfaces in IDP complexes indicated that IDPs rely more on hydrophobic interactions than their globular counterparts and form a more intimate fit with their binding partners [24]. An excellent example is the competitive binding of the bacterial toxin colicin to the peptidoglycan-associated lipoprotein (Pal) site of TolB; the intrinsically unfolded colicin can be molded to fit intimately the TolB surface, giving a greater degree of shape complementarity and comparable binding affinity to the much larger and folded Pal [25]. Colicin mimics key interactions within the TolB–Pal interface, while preventing the conformational change in TolB that is induced by binding of its cognate partner Pal.

A feature of many disordered recognition elements is that they can fold into different structures on binding to different target proteins. Examples include the intrinsically disordered regulatory region near the C-terminus of p53, which folds into helical, β-strand, and extended structures on binding to different protein partners [26], and the C-terminal activation domain of the hypoxia inducible factor HIF-1α which adopts helical and extended structures, respectively, in its complexes with CBP/p300 and the asparagine hydroxylase FIH [21].

Perhaps an even more extreme example is the nuclear coactivator binding domain (or IBiD domain) of CBP/p300, which forms a disordered molten globule state in the absence of a binding partner [27] and folds into helical structures with different topologies upon binding to a p160 nuclear receptor coactivator [28] or the interferon regulatory factor IRF-3 [29] (Figure 1). This structural plasticity, which is displayed by many intrinsically disordered proteins, raises questions about the role of ‘preformed structural elements’ [30] in the recognition process. Even when a disordered protein displays a measurable propensity to populate helical or other secondary structures in the conformational ensemble formed in the unbound state, it would seem unwise to assume that this conformation will necessarily be favored upon binding to a partner protein.

Mechanism of coupled folding and binding

One of the most intriguing questions related to the coupling of folding and binding of disordered domains is the mechanism of the process, that is does folding occur before binding or does binding occur before folding? Two extreme mechanistic possibilities can be envisaged, induced folding or conformational selection (Figure 2). In the first mechanism, the protein associates with its binding partner in a fully disordered state and subsequently folds in association with the target protein; folding is induced by association with the target. In the conformational selection mechanism, the target protein ‘selects’ a conformation closely approximating that of the bound form from the ensemble of conformations populated by the intrinsically disordered protein when free in solution. Clearly, any real system may utilize either one of these extreme mechanisms or some combination of the two; for example, it is to be expected that some degree of conformational adaptation will be required even when a preformed structural element in an IDP binds to its target.

NMR spectroscopy is especially well suited to the elucidation of the mechanisms of coupled folding and binding. Not only do the chemical shifts, residual dipolar couplings, and other NMR parameters give important

---

Figure 1

Structural comparison of the NCBD (IBiD) domain of CBP in complex with different partners. (A) NCBD in complex with the ACTR domain of p160 [28]. (B) NCBD in complex with IRF3 [29].

insights into the ensemble of structures that are sampled by the disordered protein in the absence of its binding partner, but the structure and dynamics of the complex with the target protein can also be determined. Relaxation dispersion NMR spectroscopy provides a particularly powerful approach for investigating millisecond time scale dynamics and protein folding mechanisms [31]. Relaxation dispersion methods have been applied to investigate the mechanism by which the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB folds into a helical structure on binding to its target, the KIX domain of the transcriptional coactivators CBP/p300 [32]. pKID is largely unstructured in solution [33] and folds upon binding into a pair of helices (\(\alpha_A\) and \(\alpha_B\)) that contact a particular face of the folded KIX domain [34]. Upon binding to KIX, pKID forms a weak encounter complex, stabilized by hydrophobic contacts, which exchanges rapidly with the free pKID. The encounter complex is largely unstructured and evolves by way of a partly folded intermediate to the fully bound state, without dissociating from the KIX surface [32]. This mechanism is summarized in Figure 3. Thus, binding of pKID involves an ‘induced’ folding mechanism.

The mechanism by which pKID folds upon binding to KIX has recently been investigated by coarse-grained simulations using a Go-type model [35]. The simulations are remarkably consistent with the experimental results [32], indicating that folding is induced after binding of the largely unstructured pKID to the surface of KIX and proceeds by way of a partly folded intermediate. The simulations also provide additional insights into the coupled folding and binding process that are not available from the present experiments. The pKID is largely unfolded in the transition state, and hydrophobic residues in the \(\alpha_B\) helix of pKID appear to play a crucial role in formation of the transition state and in the association process. Increasing the content of helical structure in the uncomplexed state of pKID leads to a decrease in the rate of association consistent with the ‘fly-casting’ mechanism [36], while formation of transient encounter complexes enhances the association rate without changing the binding mechanism [35]. Taken together, the relaxation dispersion experiments and the simulations provide a remarkably consistent and detailed view of the coupled folding and binding process for binding of pKID to KIX. A similar induced folding mechanism has been revealed by coarse-grained molecular dynamics and single molecule spectroscopic studies of binding of the disordered CBD domain of the Wiskott–Aldrich syndrome protein (WASP) to Cdc42 [37]. The unstructured CBD domain first binds to its target and then folds cooperatively into the native structure.

The disordered acidic transactivation domains of c-Myc, Gal4, and VP16 also appear to bind their targets by an induced folding mechanism [38]. However, given the diverse nature and wide ranging conformational propensities of intrinsically disordered proteins, from almost complete lack of structure, through pronounced propensity for formation of local elements of secondary structure, to formation of compact molten globules [21], it is to be expected that such a mechanism will not be universal and that many intrinsically disordered proteins may utilize different mechanisms. Indeed, kinetic studies of ligand-induced folding of a series of mutants of staphylococcal nuclease, considered as a model of an intrinsically disordered protein, demonstrate both folding before binding and binding before folding [39]. Further evidence that conformational selection may play a role in binding of an IDP comes from studies of the retinal phosphodiesterase inhibitory \(\gamma\)-subunit (PDE\(\gamma\)) [40]. The uncomplexed PDE\(\gamma\) is highly disordered but the conformational ensemble includes a loosely folded state that resembles the structure formed by PDE\(\gamma\) when bound to a GTPase-activating protein complex.

In apparent contradiction to the conformational selection mechanism, examples have been reported where increasing the stability of pre-organized secondary structure fails to enhance the binding affinity or even slows the kinetics
of complex formation [41,42], suggesting that disorder in the free state may actually be advantageous for the binding process. It seems likely that even in extreme cases where there is a strong propensity for transient formation of the folded conformer in the uncomplexed IDP, a certain amount of structural adaptation will be required when a preformed structural element binds, for which intrinsic flexibility is advantageous. This conclusion is supported by the simulations of the interaction of the p21 kinase inhibitor with its target; excessive stabilization of local secondary structural elements in the disordered p21 is detrimental to binding and folding to the native structure [43,44].

Some IDPs remain disordered even after binding to their targets. An example is the interaction between the Smad2 Mad homology domains and the Smad binding domain (SBD) of Smad anchor for receptor activation (SARA). The complex formed between the SBD and Mad homology domain appears to be heterogeneous, with no one region of the SBD making a dominant contribution to the hydrophobic interactions between the proteins [45]. A dynamic complex is also formed between the regulatory (R) region of the cystic fibrosis transmembrane conductance regulator and a nucleotide-binding domain (NBD) of the same protein [46]. Interactions between the nucleotide-binding domain and the intrinsically disordered R-region are mediated by multiple sites in R, each of which has helical propensity. The complex is dynamic and all of the helical elements in R interact transiently with the NBD and are stabilized, with no one element being favored over the others. Phosphorylation reduces the helicity of the R-region elements and weakens binding to the NBD. This system represents a variation on the coupled folding and binding scheme, in that the complex formed is highly dynamic and the recognition element in the IDP does not become fully ordered upon binding to the target.

A novel manifestation of coupled folding and binding has been observed in interactions between the Arf tumor suppressor and the ubiquitin E3 ligase Hdm2 [47]. Arf inhibits the Hdm2-dependent degradation of p53 by binding to Hdm2. The interaction is mediated by short intrinsically disordered domains of Arf and Hdm2, which fold synergistically to bimolecular oligomeric structures that resemble amyloid. This represents an unprecedented example of a molecular complex assembled by folding of small domains from two proteins to form a β-strand-rich oligomer.

Folding and binding kinetics

It has been suggested that the rate of macromolecular association is enhanced by the presence of disorder [36,48]. An unstructured protein has a large capture radius that facilitates the diffusive search for a binding target through the ‘fly-casting’ mechanism [36]. While several theoretical simulations suggest that fly-casting plays an important role in binding and folding of disordered proteins, for example [49], experimental support for this model has been lacking. A recent laser temperature jump analysis of the kinetics of binding of the intrinsically disordered inhibitor IA3 to the yeast protease YPrA [50] indicated a mechanism in which a transient complex is rapidly formed between IA3 and its target, followed by folding into a helical structure within the...
active site of the enzyme. This result suggested a fly-casting mechanism, in which the initial interaction between the disordered protein and its target facilitates folding and binding without affecting the equilibrium constant.

To date, comparatively few studies of the kinetics of coupled folding and binding processes have been reported. The kinetics of binding of the intrinsically disordered E-cadherin cytoplasmic domain to β-catenin have been investigated using surface plasmon resonance [51]. The association rate is fast (ca. $4 \times 10^5 \text{M}^{-1} \text{s}^{-1}$) and is enhanced more than 10-fold by phosphorylation. Formation of the complex, which involves structuring of nearly 100 residues, is highly disfavored entropically and is driven by a large decrease in enthalpy. The NMR relaxation dispersion analysis of binding of pKID to KIX also revealed a fast association (ca. $6 \times 10^6 \text{M}^{-1} \text{s}^{-1}$) followed by a slower structural rearrangement of the partly folded intermediate [32]. Relaxation dispersion studies have been reported for binding of the intrinsically disordered C-terminal activation domain of hypoxia inducible factor, hydroxylated at Asn803 (HIF-OH), to the TAZ1 domain of the transcriptional coactivator CBP [52]. In this case, the rate of association is diffusion limited ($\sim 10^7 \text{M}^{-1} \text{s}^{-1}$). Thus, disordered proteins appear to display a wide variety of binding rates.

**Cooperative binding**

It has recently been proposed on theoretical grounds that, in cases where folding accompanies binding, intrinsic disorder can maximize allosteric coupling between binding sites [53]. This suggests a mechanism whereby proteins with intrinsically disordered binding motifs can transmit signals that are triggered by interaction with different types of ligands. The fact that disordered proteins frequently bind to their targets through relatively short amphipathic sequences means that even quite small target proteins may bind more than one IDP simultaneously to form a ternary complex. This characteristic provides the potential for synergy and cross talk between different signaling pathways, mediated by interactions of disordered domains with a common binding partner. For example, the activation domains from the c-Myb and mixed lineage leukemia (MLL) transcription factors bind cooperatively to the KIX domain of CBP to form a ternary complex [54]. Bivalent interactions of the transcription factor CSL with intrinsically disordered (RAM) and partly folded (ankyrin) domains in the Notch intracellular domain have been described recently [55]. The RAM domain binds tightly to CSL and enhances ankyrin binding by increasing the effective concentration. A polymer physics model was used to model the binding and suggests that the disordered linker between the RAM and ankyrin domains is of optimal length to enhance the bivalent interaction.

**Role of phosphorylation and flexibility in signal transduction**

One of the most common means of signal transduction is through phosphorylation at defined sites within signaling proteins. Analysis of the sequences surrounding known phosphorylation sites reveals a strong propensity towards intrinsic disorder in the neighboring regions of the protein [56]. Phosphorylation can act at different levels, affecting either the conformational propensities in the uncomplexed state or the interactions between the IDP and its partner proteins. Further, phosphorylation can function as a switch or as a ‘rheostat’ to fine-tune the biological response [57].

Perhaps it might be expected that phosphorylation should cause a change in the intrinsic propensity of the protein for structure formation in the free form. Such a change in conformational preference was not observed experimentally for the pKID domain [33], although recent computer simulations appear to suggest that the turn region incorporating the phosphorylated serine residue of pKID may be favored in the uncomplexed state when the phosphoryl group is present [58]. A comprehensive thermodynamic analysis of the pKID/KIX system [42] showed that the effect of the phosphoryl group on the binding of pKID to KIX was primarily enthalpic, an indication that prior formation of secondary structure in the free protein has a negligible effect on the binding interaction.

A variation on the theme of phosphorylation and conformational switching is shown by the Ets-1 transcriptional activator and its interaction with DNA [57]. In the active, unphosphorylated state, the Ets domain binds DNA, an interaction that is allowed by the unfolded and conformationally labile H1 helix. Upon phosphorylation at multiple sites, the H1 helix becomes progressively more highly structured, disfavoring the conformational change that permits the binding of Ets-1 to DNA. The disordered region thus functions as an allosteric effector, essentially a ‘rheostat’ that fine-tunes transcription in response to variations in the level of phosphorylation. The cyclin-dependent kinase inhibitor Sic1 is a further example of an intrinsically disordered protein that is regulated by multiple site phosphorylation. In this case, phosphorylation of a threshold number of sites is required for receptor binding, which is driven by cumulative electrostatic interactions with the phosphorylated Sic1 [59].

A general entropic model has been presented to account for highly cooperative binding driven by multiple site phosphorylation in intrinsically disordered regions [60].

**Are these domains disordered *in vivo*?**

For many, the evidence that important transcriptional and signaling domains could be as disordered *in vivo* as they appear to be in the *in vitro* experiments described above is hard to countenance. Aside from the issue of stability within the cell, it is has been postulated that the crowded
environment within the cell might induce folding of IDPs. This issue has been recently addressed through NMR experiments, with mixed results: while the eukaryotic α-synuclein remains disordered within the cytoplasm in Escherichia coli cells [61], the bacterial protein FliG appears to gain some structure in the cellular milieu [62].

One of the signature characteristics of a disordered protein is that it is highly susceptible to proteolytic degradation in vitro, and it has been suggested that sensitivity to proteolysis may be important to the signaling function of many IDPs [21]. A recent survey shows that while there is a correlation between protein disorder and short half-life, the correlation is weak suggesting that while there is a correlation between protein disorder and sensitivity to proteolysis may be important to the signaling function of IDPs, other factors may also play a role.

Conclusions
The occurrence of disordered regions in important cellular signaling and regulatory proteins continues to be a common theme as the recently published genome sequences of many organisms are mined for information. Folding upon binding of these regions to their partners has been a fertile field for the elucidation of the structural bases of a number of physiologically relevant protein–protein interactions, and will doubtless remain relatively accessible experimentally. Database and computer simulation, together with increasingly sophisticated algorithms for the identification of sequence motifs characteristic of certain binding interactions, should also provide information on systems that may be too large or heterogeneous for initial experimental characterization, but which can be approached experimentally after they have been parsed in silico.

Acknowledgments
This work was supported by grants CA96865, DK34909 and GM57374 from the National Institutes of Health, and by the Skaggs Institute for Chemical Biology.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


SAXS is used as the basis for analysis of the conformational ensemble of flexible proteins using an ensemble optimization method.


Investigation of the structural ensemble and dynamics of the prion-determining domain of the yeasts prion Sup35 using single-molecule fluorescence resonance energy transfer and fluorescence correlation spectroscopy.
Linking folding and binding

Wright and Dyson 37


32. Sugase K, Dyson HJ, Wright PE: Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 2007, 447:1021-1025. NMR methods, including chemical shift titration and relaxation dispersion, are used to determine the kinetics and mechanism by which the intrinsically disordered pKID activation domain of CREB folds on binding to the KIX domain of CBP.


Simulations using G0-type models to characterize the mechanism by which pKID folds on binding to KIX (see [32] above). An induced-folding mechanism incorporating non-native initial contacts and a partially folded intermediate is consistent with the experimental results.


42. Zor T, Mayr BM, Dyson HJ, Montminy MR, Wright PE: Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myc) and inducible (CREB) activators. J Biol Chem 2002, 277:42241-42248.


47. Sivakolundu SG, Nourse A, Moshiach S, Bothner B, Ashley C, Satumba J, Lahti J, Kriwacki RW: Intrinsically unstructured domains of Arf and Hm2 form bimolecular oligomeric structures in vitro and in vivo. J Mol Biol 2008, 384:240-254. This paper reports the novel finding that intrinsically disordered regions in Arf and Hm2 fold together to form amyloid-like structures that function in the regulation of Hm2 and p53.


