Quarterly Reviews of Biophysics **44**, 4 (2011), pp. 467–518. © Cambridge University Press 2011
doi:10.1017/S0033583511000060 Printed in the United States of America

Expanding the proteome: disordered and alternatively folded proteins

H. Jane Dyson

Department of Molecular Biology MB2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Abstract. Proteins provide much of the scaffolding for life, as well as undertaking a variety of essential catalytic reactions. These characteristic functions have led us to presuppose that proteins are in general functional only when well structured and correctly folded. As we begin to explore the repertoire of possible protein sequences inherent in the human and other genomes, two stark facts that belie this supposition become clear: firstly, the number of apparent open reading frames in the human genome is significantly smaller than appears to be necessary to code for all of the diverse proteins in higher organisms, and secondly that a significant proportion of the protein sequences that would be coded by the genome would not be expected to form stable three-dimensional (3D) structures. Clearly the genome must include coding for a multitude of alternative forms of proteins, some of which may be partly or fully disordered or incompletely structured in their functional states. At the same time as this likelihood was recognized, experimental studies also began to uncover examples of important protein molecules and domains that were incompletely structured or completely disordered in solution, yet remained perfectly functional. In the ensuing years, we have seen an explosion of experimental and genome-annotation studies that have mapped the extent of the intrinsic disorder phenomenon and explored the possible biological rationales for its widespread occurrence. Answers to the question 'why would a particular domain need to be unstructured?' are as varied as the systems where such domains are found. This review provides a survey of recent new directions in this field, and includes an evaluation of the role not only of intrinsically disordered proteins but also of partially structured and highly dynamic members of the disorder–order continuum.

1. Introduction 469

2. Physiological roles for IDPs 471
2.1. Association of IDPs with disease 472

3. Why would proteins be disordered? 473
3.1. Promiscuous basal activity 473
3.2. Enhanced specificity 474
3.3. Surface area burial and affinity 474
3.4. Complex connections 475
3.5. Facilitate regulation by post-translational modification 475
3.6. Regulation by proteolysis 475
3.7. Regulation by competition between IDP ligands 476

Author for correspondence: Dr H. J. Dyson, Department of Molecular Biology MB2, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
Tel.: 1-858-784-2223; Fax: 1-858-784-9822; Email: dyson@scripps.edu

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

3.8. Higher capture radius for formation of complexes 476
3.9. Bulk physicochemical effects 476
3.10. Increased disorder required for function 476

4. Special new concepts – Molecular Recognition Features (MoRFs), fuzzy complexes, hub proteins 477
4.1. MoRFs, MoREs and more 477
4.2. Fuzzy complexes 477
4.3. Malleable machines 478
4.4. Hub proteins 478

5. Intrinsic disorder as function: linkers and repeats 478
5.1. Linkers 478
5.2. Scaffolding proteins 480
5.3. Repeat sequences 480
5.3.1. Examples 481
5.4. Alternative splicing, evolution and evolvability 481

6. Intrinsic disorder as part of function: coupled folding and binding 482
6.1. Transcriptional activation: partners of the KIX domain of CBP 482
6.2. Disordered regions of the tumor suppressor p53 483
6.3. Mutual synergistic folding 483
6.4. Calmodulin signaling 484
6.5. Proline-rich sequences and their partners 485
6.6. Binding without folding 485

7. Thermodynamics and kinetics 485
7.1. Binding-induced folding versus conformational selection 486
7.2. Kinetic discrimination in antibody recognition 487
7.3. Thermostability and intrinsic disorder 487
7.4. Allostery and intrinsic disorder 489
7.5. Compensation for entropy changes in binding 489

8. Kinetic control of signaling by competition for binding sites 489
8.1. Binding of ligands to the KIX domain of CBP 489
8.2. Binding of ligands to the TAZ domains of CBP 490
8.3. Action of disordered domains of viral proteins 491
8.4. Kinetic control in signal transduction 491
8.5. Mutual synergistic folding and kinetic control 492
8.6. Control of the cell cycle by IDPs 493

9. Role of post-translational modification 494
9.1. Direct relationship between chemistry and function: Hif-1α and TAZ1 494
9.2. Phosphorylation of IDPs and IDRs 496
9.3. Ubiquitination and neddylation 497
9.4. Ubiquitin-independent protein degradation 498
9.5. IDPs and metal ion binding 498

10. Partly folded and molten globule-like states 498
10.1. AR proteins 499
10.2. Characterization of binding-induced molten globule-like state 499
10.3. Coiled coil and leucine zipper proteins 500

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

11. Strategies for study of disordered proteins 500
11.1. Bacterial expression of IDPs 501
11.2. Biochemistry of IDPs 501
11.3. NMR techniques used for IDPs 501
11.4. Fluorescence methods 503
11.5. Other experimental techniques 503
11.6. Computational studies of IDPs 503

12. Conclusions 503

13. Acknowledgments 504

14. References 504

1. Introduction

Proteins that are folded into defined three-dimensional (3D) structures have provided the basis for our present understanding of the means whereby cellular metabolism is conducted in all cells, from bacteria to complex eukaryotes. The paradigm of the transcription of the genetic information encoded in deoxyribonucleotide sequences in DNA to the ribonucleotide sequence of messenger RNA, whence it is translated by ribosomes to a sequence of amino acids in a protein, has been axiomatic in the last 50 years at least. Yet the last step in the process, the folding of the linear amino acid sequence to form a well-structured protein is not completely understood yet, and it appears from recent work in a number of fields that the formation of a stable 3D structure may not be absolutely required in all cases in order for a protein to be functional. Indeed, the absence of stable 3D structure appears in some cases to be required for the correct function of the protein.

Some 15 years ago, two parallel sets of observations made it clear that 3D structure was likely not a prerequisite for the function of some protein systems. On the one hand, the availability of large numbers of gene sequences, culminating in the publication of the complete human genome in 1999, allowed a computational and bioinformatic approach to the analysis of sequence propensities. Sequence analysis is used, for example, as one of the methods to approach the prediction of unknown structure in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) exercises. Following such an analysis, several groups, particularly that of Dunker (then at Washington State University and now at Indiana University) noted that the published genomes contained sequences that would code (if they coded for anything at all) for unstructured proteins (Romero *et al.* 1998). The astonishing statistics were that 6–33% of bacterial proteins, 9–37% of archeal proteins and 35–50% of eukaryotic proteins should contain stretches of 40 or consecutive disordered amino acids (Dunker *et al.* 2000). At the same time, and quite independently, several experimental labs had noted that some proteins, particularly in the areas of transcriptional and translational control, cell cycle control and signaling (Daughdrill *et al.* 1998; Kriwacki *et al.* 1996) remained without a stable 3D structure even when subjected to all the ingenuity of the experimentalist's repertoire, including optimization of pH, temperature, buffer conditions, salt concentrations and the presence of additives such as chaotropic agents and osmolytes. A pattern began to emerge. These proteins remained unstructured in solution in the absence of their physiological partners, but would fold in the presence of the partner, forming stable complexes that could be purified and characterized (Daughdrill *et al.* 1997; Kriwacki *et al.*

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

1996; Radhakrishnan *et al.* 1997). The first compendium of these early experimental results was published in 1999 (Wright & Dyson, 1999).

Once it had been accepted that proteins could possibly be unfolded yet still functional, an analysis of the sequences of a number of pivotal proteins in the cell revealed that large stretches contained amino acid compositions that were clearly not conducive to the formation of a normal globular protein fold (Garner *et al.* 1998; Romero *et al.* 1999, 2001; Williams *et al.* 2001). The sequences frequently contained a very high proportion of Ser, Gly, Pro, Asn and Gln, with very low frequency of hydrophobic amino acids that would normally form part of the hydrophobic core of a folded globular protein. Some of the sequences contained multiple tandem repeats of these amino acids, while others were enriched in charged amino acids, Lys, Arg, Glu and Asp. These apparently anomalous sequence compositions prompted the development of several algorithms for prediction of disordered sequences in proteins (Linding *et al.* 2003a, b; Romero *et al.* 1997; Uversky *et al.* 2000; Ward *et al.* 2004; Weathers *et al.* 2004). More recently, a number of refinements and additions to these methods have been developed, including Disprot (Sickmeier *et al.* 2007), CDF (Xue *et al.* 2009b), FoldIndex (Prilusky *et al.* 2005), TopIDP (Campen *et al.* 2008) and contact prediction (Schlessinger *et al.* 2007). It is not the purpose of this review to explain in detail these computational and annotational approaches. Each likely has strengths, and the common practice in the field is to apply several of these predictors to a sequence of interest (Lieutaud *et al.* 2008).

More focused prediction methods have been applied specifically to certain systems, such as the prediction of loop structures (Deryusheva *et al.* 2008), amyloid formation (Galzitskaya *et al.* 2006), transcription factors (Fukuchi *et al.* 2009) and transcriptional activation networks (Singh & Dash, 2007). Differences in compositional biases and hence the predicted proportions of disordered regions were found between water-soluble proteins and the membrane-integral proteins of the helical and β-barrel classes (Xue *et al.* 2009a), hinting that different prediction methods may be needed for membrane proteins; the authors (Xue *et al.* 2009a) note that the paucity of structural information for membrane proteins renders these conclusions preliminary. Annotation of published non-human genomes has begun (Forbes *et al.* 2010; Galea *et al.* 2009; Mohan *et al.* 2008; Tompa *et al.* 2006; Xue *et al.* 2010a) and libraries (Lobanov *et al.* 2010a) and databases (Chen *et al.* 2006a, b; Goh *et al.* 2008; Lobanov *et al.* 2010b; Peng *et al.* 2006) of disordered proteins have been compiled. Methods have been developed for calculation of likely binding sites and solvent accessibilities (Bernado *et al.* 2006; Estrada *et al.* 2009) in unfolded ensembles. Proteomic approaches have been applied not only to disordered proteins themselves (Csizmok *et al.* 2007) but also to their interaction partners (the so-called interactome) (Fong *et al.* 2009; Haynes *et al.* 2006; Lobanov *et al.* 2010b). Interestingly, it appears that the abundance of expressed proteins is directly correlated with their predicted level of disorder (Paliy *et al.* 2008), and the lifetimes of disordered proteins and their mRNAs are tightly regulated (Gsponer *et al.* 2008; Ma & Nussinov, 2009) and variable (Edwards *et al.* 2009). A power law distribution of disordered sequences within the human proteome has been used to infer the functional importance of disordered sequences (Tompa & Kalmar, 2010).

As the number of examples of intrinsically disordered proteins (IDPs) began to grow, those interested in this newly discovered quirk of protein chemistry began to form theories as to what advantages intrinsic disorder might confer upon a protein. Particularly if the protein formed part of an interaction network, it might, for example, be advantageous for it to interact with a number of partners, and, being disordered, the protein would be capable of using different structures to achieve these interactions. Multiple binding sites or motifs on a single disordered domain might

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

be conducive to the formation of higher-order complexes, where disparate partners were brought into close proximity by independent binding to the same domain. In hindsight, it is easy to see why such mechanisms would be operative in systems such as transcriptional activation or cell cycle control, where the assembly of large multi-protein complexes is necessary for the process to occur.

Although the bioinformatic and *in vitro* experimental evidence for disorder in functional protein domains were unequivocal, there were doubts expressed that disordered proteins could function *in vivo*. The well-known protease sensitivity of unfolded proteins was an apparent problem, as well as the crowded nature of the cellular environment, which ought to induce folded states. One of the best explanations of the disorder phenomenon in these terms is given in a 2002 review by Dunker *et al.* (2002). Firstly, proteases are compartmentalized and sequestered in the cell (Frankel & Kim, 1991), and their reactions are highly regulated (Dunker *et al.* 2001; Wright & Dyson, 1999). Secondly, the disordered regions themselves may be inaccessible or lack protease sensitive sites, and some, particularly those associated with transcriptional regulation, for example, may be present in the free, disordered state only transiently as they pass from one partner to another or are associated with chaperones (Dunker *et al.* 2002). Candidate 'nanny' chaperones have recently been identified (Tsvetkov *et al.* 2009b), although disordered proteins as a class display no particular preference for chaperone binding in general (Hegyi & Tompa, 2008). Chaperones themselves frequently contain a high proportion of disordered regions: 54% of residues in RNA chaperones were shown to fall in such regions (Tompa & Csermely, 2004), which were suggested to function as recognition elements and loosening agents for misfolded regions via an 'entropy transfer' mechanism that has also been invoked in other protein–protein interactions (Sue *et al.* 2008). Specific disordered regions of proteins may have essential roles in folding and assembly of molecular complexes. An example of this is the assembly of IgG antibodies, which requires the presence of an unfolded heavy-chain constant domain (CH1 domain) of the heavy chain to interact with the light chain light-chain constant domain (CL domain) to form the intact IgG (Feige *et al.* 2009).

Molecular crowding has been found to favor folded states (Iakoucheva & Dunker, 2003), but this necessarily applies only to proteins that are capable of folding independently. If the energy landscape of the polypeptide is not compatible with an independently folded state, even a crowded environment cannot induce the formation of such a state (Dunker *et al.* 2002). From a topological point of view, many complexes of disordered domains have structures that can be formed only if one or both the partners are disordered before it is formed, for example, the complexes of the transcriptional adapter zinc-binding (TAZ) domains of CREB binding protein (CBP; see later section).

In this review, I will provide examples for both the types of IDPs and their myriad functional attributes. Clearly this is a field that is far from mature, and numerous examples are even now being studied and will doubtless be published before this article goes to press. I have endeavored to be as comprehensive as possible in at least mentioning seminal studies, but the literature increases every day, and I apologize for any omissions.

2. Physiological roles for IDPs

One of the earliest observations about the occurrence of IDPs was that they were present with highest frequency in certain physiological niches. Thus, IDPs were found to be enriched among cancer related and signaling proteins (Iakoucheva *et al.* 2002). Disordered domains seem to be

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

present in high frequency in regulatory proteins associated with transcription and translation, the cell cycle, signal transduction and protein phosphorylation, all processes that are in some way involved in physiological control. The central role of proteins that contain ID domains (IDDs) in crucial cellular processes argues both for the importance of disordered domains as a general structural type and for the fundamental importance of intrinsic disorder in the metabolism of living cells.

2.1 Association of IDPs with disease

One of the intriguing connections that have been made by a number of groups is that IDPs and IDDs are frequently associated with diseases. From the standpoint of genetic diseases, this is relatively easy to understand. Increases in the numbers of repeat sequences frequently presages susceptibility to hereditary neurodegenerative diseases such as Huntington's and Parkinson's diseases (Midic *et al.* 2009; recent reviews include Uversky & Eliezer, 2009; Uversky, 2009a). At first sight it may be puzzling that the lesions that precipitate disease in these cases occur in disordered regions with generally low requirements for exact amino acid sequences. Why would the disease-causing mutations not been seen preferentially in the structured regions, where they would presumably have more effect? The answer to this question lies in the likelihood that the effects of deleterious mutations in structured domains of vital proteins (for example, leading to lowered stability or impaired function) would likely be so great as to prove lethal, usually in the embryonic stage. Genetic diseases caused by mutations in disordered regions can thus be regarded as a lesser evil. Mutations in disordered regions can also result in the loss of important post-transcriptional modification sites, leading to disease (Li *et al.* 2010).

IDPs are also prevalent in diseases that are not apparently primarily inherited. The connection between disorder and diseases such as cancer, cardiovascular disease, amyloidoses, neurodegenerative diseases and diabetes has been extensively explored in recent reviews (Midic *et al.* 2009; Uversky *et al.* 2008). The authors conclude that these diseases, which have in common failures of protein signaling and structure and have been loosely grouped as 'conformational diseases', may be characterized not only by protein misfolding but also by failures of post-translational modification and inability to interact correctly with physiological partners (Uversky *et al.* 2008). A possible rationale for the toxicity of amyloid-like aggregates was recently suggested to be the resulting sequestration of important cellular control factors, which frequently include intrinsically disordered regions (Olzscha *et al.* 2011). Cancer in particular has come under scrutiny: cancers frequently arise due to chromosomal translocations that result in the formation of fusion proteins, which might be expected to have a high degree of intrinsic disorder. An intriguing hypothesis to explain the uncontrolled growth of cancer cells in these cases is that the disordered segments are qualitatively different from the unfolded forms of globular proteins, thus enabling the fusion proteins to escape the normal surveillance and housekeeping functions of the cell (Hegyi *et al.* 2009). Another connection of IDPs to cancers occurs in the human papillomaviruses, where variants with high risk for oncogenic transformation appear to contain a higher proportion of intrinsic disorder (Uversky *et al.* 2006). Indeed, it appears that more virulent strains of disease-causing viruses may be characterized by their increased levels of intrinsic disorder (Goh *et al.* 2009), and viruses in general appear to have a high proportion of genes potentially coding for disordered proteins (Xue *et al.* 2010b). Recognition of the crucial role of IDPs in a number of diseases prompts new ways of thinking in the design of drugs (Cheng *et al.* 2006).

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

While the majority of the explicit connections of IDPs with disease states remain a matter of bioinformatic correlations of the occurrence of disorder in disease-related proteins, a considerable amount of basic biophysical and biochemical work has also been done to elucidate the possible molecular basis of IDP interactions. A number of possible answers to the question 'why would proteins be disordered in order to function?' have occurred to the community as these studies have proceeded, and these thoughts are summarized in Section 3 of this review. In later sections, I will summarize some of the experimental work that has been done in the last 10 years to elucidate the biochemical and biophysical behavior of IDPs that are known to be important in physiological processes.

3. Why would proteins be disordered?

Both bioinformatic and experimental approaches have identified intrinsic disorder in an unexpectedly large number of proteins. Given the prevalence of this observation, it behooves us to suggest possible reasons why such a repertoire should have evolved, particularly among higher organisms. The present state of the field (reviewed in Dyson & Wright, 2005; Gsponer & Babu, 2009, among many other examples) gives the following possible reasons why a protein or part of a protein might be disordered in order to function. The list is not exhaustive, and will no doubt be expanded as more examples are discovered.

3.1 Promiscuous basal activity

One feature that instantly comes to mind is the possibility that a disordered domain could bind in different conformations to different partners, as must occur, for example, when the cyclin-dependent kinase inhibitor p21 binds to different cyclin–CdK complexes during the cell cycle (Kriwacki *et al.* 1996). This concept was explored many years ago in the context of antibody affinities, when it was noticed that the combining sites of antibodies were frequently highly flexible (Tainer *et al.* 1984). Experimental tests aimed at observing different structures of antibody combining sites in complex with different antigens generally proved inconclusive (Wilson *et al.* 1985), as the designs were based on crystallographic structures that showed different (static) structures in the same amino acid sequences when they were present in different proteins. With IDPs, however, the constraints imposed by the presence of ordered structure are no longer operative, and a true conformational ensemble can be explored by the sequence. The result is that a number of examples have now been found of disordered sequences that take up completely different conformations when complexed to different partners. One of the best examples is the carboxyl-terminal activation domain of the hypoxia-inducible factor α domain, Hif-1α. In complex with the TAZ1 domain of the CBP, the Hif-1α C-terminal activation domain (CAD) takes up a largely helical configuration (Dames *et al.* 2002; Freedman *et al.* 2002), but when bound to the enzyme that catalyzes the oxygen-dependent hydroxylation of Hif-1α, this same sequence is present as an extended structure (Elkins *et al.* 2003; Lee *et al.* 2003). Other examples include chromatin remodeling proteins, which interact with a wide variety of proteins and nucleic acids (Sandhu, 2009), and different isoforms of nuclear hormone receptors (Kumar & Thompson, 2010; Nocula-Lugowska *et al.* 2009). An interesting side issue is the appearance of disorder as a result of splice variation. Splice variants that have been structurally characterized generally differ by not more than six amino acids, although variations of much greater lengths are known. It was

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

recently shown that the probability that one or other of the variants would be disordered increased with the size of the amino acid sequence difference between the variants, likely accounting for their absence from structure databases (Hegyi *et al.* 2011). Alternative splicing provides a facile route to the evolution of functional diversity (Romero *et al.* 2006; Tokuriki & Tawfik, 2009), pointing toward another, more subtle role for disorder in proteins.

3.2 Enhanced specificity

One of the intriguingly common features of published structures of complexes of proteins or protein domains that are disordered in the free state is the highly complementary nature of the binding to the target. The interactions between the partners are highly complicated, and involve many different surfaces. In many cases, one partner may appear to be 'wound around' the other partner, an interaction that simply could not occur unless the amino acid sequence in question was present before complex formation in a state that lacked its own stable 3D folded structure. An early example of such a structure includes the lymphoid enhancer-binding factor 1 high mobility group (LEF1 HMG) domain complex with DNA, where an $\sim 100^{\circ}$ bend is induced in the DNA structure by binding of the protein (Love *et al.* 1995), which is largely disordered in solution (Love *et al.* 2004). Further examples of such complexes between protein partners are described in detail in Sections 6 and 8 of this review. Another class of interactions where unstructured regions are instrumental in raising the sequence-specific affinity between partners is in the interactions between a number of transcription factors and DNA. Besides the 'snap-lock' specificity switch that operates with structured zinc finger domains (Laity *et al.* 2000; Wuttke *et al.* 1997), a number of transcription factors contain unstructured sequences at N- or C-termini that make all-important contributions to increasing the affinity of the interaction (Gearhart *et al.* 2003; Hill *et al.* 2009; Holmbeck *et al.* 1998). This phenomenon has recently been observed in RNA–protein interactions (Kucera *et al.* 2011). The end result of the use of disordered segments in such interactions is that highly specific binding can be accomplished at modest affinity. Perhaps even more importantly for complexes that are formed as a result of the reception of discrete signals, the complexes must be capable of dissociation once the signal has been turned off. Ubiquitin-independent degradation of disordered segments (see below, this section) and competition for binding sites (see Section 8) form important components of this 'off-switch' mechanism.

3.3 Surface area burial and affinity

One of the most intriguing rationales for the employment of intrinsically disordered domains in interactions involving signal transduction and cellular control is the observation that the complexes formed manifest the burial of anomalously large surface area for the number of amino acids involved (Gunasekaran *et al.* 2003). The consequence of this for the cell is that a high degree of specificity (and affinity) can be achieved for relatively little metabolic outlay: to achieve the same surface area burial employing only folded proteins would require the proteins to be much larger, imposing an unacceptable metabolic burden on the cell (Gunasekaran *et al.* 2003). A general model has been suggested where catalytic and low-affinity binding proteins might preferentially take up ordered structures, while high-affinity binding proteins could 'tolerate' disorder (Liu *et al.* 2009). These authors suggest that the model explains the role of disorder in tuning binding affinity to maximize specificity, providing a means whereby protein function can be optimized through natural selection mechanisms.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

3.4 Complex connections

Many of the central functions of the cell, such as transcription of genes, require the assembly of multi-protein complexes. One of the major roles of a transcriptional coactivator such as CBP is to connect the various members of the multi-protein complex, by interacting with many different proteins at once, perhaps at many different sites on the DNA. Given this role, it seems logical that CBP contains a number of long linker sequences to allow for differences in architecture at these sites. This aspect is described in more detail in Section 5. Another intriguing aspect is the role played by disordered segments in the assembly of multi-subunit signaling complexes (Sigalov, 2010).

3.5 Facilitate regulation by post-translational modification

So much of metabolism appears to involve the switching on and off of complex pathways that it is no wonder that a majority of these pathways involve chemical switches, some of which are highly ingenious (the direct influence of oxygen chemistry on the regulation of the hypoxia response by the Hif factors is one example). IDPs provide an excellent vehicle for the control of biological switches by post-translational modification. Sites of phosphorylation, methylation, hydroxylation, etc., frequently occur in disordered regions of otherwise highly structured proteins – the Hif-1α system provides one rationale for such an arrangement, where the structural requirements for enzymatic modification and target binding are different, and can be accommodated because the interaction domain containing the hydroxylation site is disordered and thus capable of taking up either conformation. Post-translational modification and its relation to disordered proteins are explored in detail in Section 9.

3.6 Regulation by proteolysis

As mentioned previously, all IDPs are not necessarily particularly susceptible to proteolysis, and, since proteolysis is generally highly regulated in viable cells, IDPs probably have the potential for cellular lifetimes appropriate to their function. Nevertheless, the probability of rapid degradation of a disordered region in the free state has been carefully explored by several groups. The abundance of IDPs in the cell, which may precipitate disease states if perturbed, has been shown to be tightly regulated by several cellular processes, including mRNA transcript clearance, translational rate and proteolytic clearance (Gsponer *et al.* 2008). Degradation of several cancer-associated proteins, which are frequently disordered, was shown to be proteasome-mediated, but occur independent of ubiquitination (Jariel-Encontre *et al.* 2008). Ubiquitin-independent degradation would allow for rapid clearance of potentially harmful proteins, without the necessity for specific intact ubiquitination sites, which could potentially be inactivated by mutation. The involvement of the molten globule-like C-terminal region of free IκBα in its rapid ubiquitin-independent degradation by the 20S proteasome has been recently documented (Mathes *et al.* 2010; O'Dea *et al.* 2007). The disordered N-terminal transactivation domain (N-TAD) of the tumor suppressor p53 was recently shown to be degraded by a rapid ubiquitin-independent pathway involving the 20S proteasome, as well as a slower ubiquitin-dependent process mediated by the 26S proteasome (Tsvetkov *et al.* 2009a). Interestingly, the rapid process is blocked by proteins such as Hdmx that bind to the p53 N-terminus (Tsvetkov *et al.* 2009a), suggesting that complex formation that sequesters the disordered domain in a stable structure abrogates the ubiquitin-independent degradation process. A more detailed exploration of the p53-Hdm interaction network is presented in Section 8.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

3.7 Regulation by competition between IDP ligands

Much thought has been given by a number of groups to the thermodynamic and kinetic consequences of the lack of structure in IDPs, in particular for the energetics of complex formation. These studies are described in Section 8. Why does not the process of coupled folding and binding involve an unacceptable entropy penalty from the folding of the IDP? It has now been demonstrated in a number of experimental systems that the entropy penalty for folding of an IDP is not unacceptably high when there is sufficient enthalpy driving the reaction. Indeed, the enthalpy compensation of, for example phosphorylation, can provide a precise conformational switch, such as occurs in the phosphorylated kinase-interaction domain (pKID)/KIX complex (Radhakrishnan *et al.* 1997). These and other examples are discussed at greater length in Sections 8 and 9. One of the most intellectually satisfying justifications for disorder in domains that compete for the same binding sites or partner molecules is the concept of competition for scarce regulatory molecules by various disordered domains (De Guzman *et al.* 2004b). In some systems, the binding sites for different domains on the same partner differ slightly, so that one could envisage the competition for the binding site at the molecular level as a 'stripping' or 'peeling' reaction. This concept, and the experimental data that support it, are described in Section 8.

3.8 Higher capture radius for formation of complexes

Originally proposed by the Wolynes group (Bryngelson *et al.* 1995; Shoemaker *et al.* 2000), the concept of 'fly-casting' suggests that disorder in a polypeptide chain enhances the capture radius of the chain, with consequences for the kinetics of complex formation. This attractive concept has not received unequivocal support from later experiments or simulations, at least in part following the realization that the conformational ensemble of an intrinsically disordered domain would on average have quite a small radius, due to the presence of significant numbers of relatively compact conformers, that is, the probability that a given conformer would be fully extended and therefore have a capture radius greater than average, would be quite low. Nevertheless, one can imagine that at least the possibility of such a conformer exists for the disordered protein, whereas if the polypeptide were present as a globular folded domain, the probability of such conformer would be extremely small.

3.9 Bulk physicochemical effects

Since the solvent-accessible surface of a disordered ensemble is greater than that of a globular folded protein of the same size, it might be expected to have a proportionally greater effect on the bulk physical chemistry of the solution. Such an effect has been demonstrated, for example in the cryoprotective functions of dehydrins (Hughes & Graether, 2010) and in proteins associated with desiccation tolerance (Chakrabortee *et al.* 2010).

3.10 Increased disorder required for function

Although this behavior may ultimately prove to be widespread, there are at present only a few systems known where the unfolded form is the functional one and the folded form is non-functional. Examples include the redox-regulated chaperone Hsp33, which contains a C-terminal zinc-binding site (Graf *et al.* 2004). In the absence of oxidative stress, the reduced form of Hsp33

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

binds zinc to form a well-structured domain (Won *et al.* 2004). Oxidation of the zinc-binding cysteines causes release of the zinc and loss of structure in the C-terminal domain, leading to dimerization and activation of the chaperone. Such a requirement for conformational freedom is also seen in the nematode anticoagulant NAPc2 (Duggan *et al.* 1999) and in the IκBα inhibitor of nuclear factor κB (NF-κB), where stabilizing mutations in the poorly ordered parts of the ankyrin repeat (AR) domain result in loss of the stripping function of IκBα (Bergqvist *et al.* 2009; Ferreiro *et al.* 2007).

4. Special new concepts – Molecular Recognition Features (MoRFs), fuzzy complexes, hub proteins

Because the whole idea of the functional disordered protein is new and difficult to describe, this infant field has given rise to a number of new technical terms that should perhaps be introduced and defined early in this review. The first term is one which has a number of synonyms – 'intrinsically disordered protein' itself. Because the concept arose in several fields independently at about the same time, the terminology in each field began slightly differently, with 'intrinsically unstructured' (Wright & Dyson, 1999), 'natively unfolded' (Weinreb *et al.* 1996), 'natively or intrinsically disordered' (Dunker *et al.* 2002) and even 'rheomorphic' (Holt & Sawyer, 1993). At a recent conference (Barcelona BioMed Conference 'Intrinsically Disordered Proteins in Biomedicine', October 2010), a consensus was reached among the participants that the terminology 'intrinsically disordered' was the most descriptive and the least open to misinterpretation, and this term is accordingly used throughout the text of this review.

4.1 MoRFs, MoREs and more

The term Molecular Recognition Feature (MoRF) or Molecular Recognition Element (MoRE) was introduced in 2005 to describe a particular type of relatively short, perhaps marginally structured protein region within a longer, largely disordered sequence (Mohan *et al.* 2006; Oldfield *et al.* 2005; Vacic *et al.* 2007). The key characteristics of a MoRF are that it is the site of binding of the disordered protein to a partner, and that it undergoes some form of disorder-to-order transition upon binding. MoRFs were classified as α-MoRFs, β-MoRFs and ι(iota)-MoRFs, according to whether α-, β-, or irregular secondary structure type was formed upon binding (it is not clear how to classify a sequence such as Hif-1α, which forms α-structure in one complex (Dames *et al.* 2002; Freedman *et al.* 2002) and β-structure in another (Elkins *et al.* 2003; Lee *et al.* 2003)). MoRFs may have a propensity for formation of residual structure in the free state (Kim *et al.* 2009; Mohan *et al.* 2006), although this may be hard to detect experimentally (Zor *et al.* 2002). At least in the case of the short linear motifs recognized by SH2, SH3 and Ser/Thr kinase domains, these sequences may be conserved between species (Ren *et al.* 2008).

4.2 Fuzzy complexes

The term 'fuzzy complexes' was coined to emphasize the fluid nature of protein–protein interactions (Tompa & Fuxreiter, 2008), and to point out that this attribute of many, if not all, important protein complexes provides a reminder that high-throughput methods to define the structures of all complexes (the 'interactome') may be doomed to frustration, if not to failure. Four modes of structural disorder or fuzziness in complexes were defined, covering the major

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

forms of static and dynamic disorder that might be expected in complexes. A noteworthy dialogue (Chatr-Aryamontri *et al.* 2008; Fuxreiter & Tompa, 2009; Welch, 2009; Wilkins & Kummerfeld, 2008), published over several months has refined and expanded the 'fuzzy' concept, with references both to early papers (Alberts, 1998; Srere, 2000) and to more recent reviews pointing out the pitfalls of too great a reliance on the results of high-throughput experiments, particularly in the realm of protein–protein interactions (Mackay *et al.* 2007).

4.3 Malleable machines

The important processes in the cell frequently require the assembly and operation of molecular machines. Those machines whose operation is not constant, but which must be turned on in response to signals or other metabolic requirements of a given cell, frequently contain components that are less than fully structured, and sometimes completely disordered, within a matrix of other, more or less rigid components. Such a 'malleable machine' (Fuxreiter *et al.* 2008) would presumably be better capable than an entirely rigid entity of responding to different conditions, recognizing multiple targets that may have different structures, and allowing or facilitating conformational rearrangements.

4.4 Hub proteins

The full inventory of all the interactions in a given cell is termed the interactome (Sanchez *et al.* 1999). This information can be systematically classified in network diagrams; within such a diagram, the proteins comprise the network nodes, most of which are connected to only a few other proteins. However, some proteins, termed 'hub proteins' (Albert *et al.* 2000; Han *et al.* 2005), are connected to many more partners than the average. Hub proteins determine the organization of the network, since their removal or inactivation would cause severe disruption. Bioinformatic studies have consistently identified a higher-than-average proportion of intrinsic disorder in hub proteins (Dosztanyi *et al.* 2006; Dunker *et al.* 2005; Haynes *et al.* 2006; Kim *et al.* 2008), particularly when the proteins are not correlated in their mRNA expression (Singh *et al.* 2007). Several recent review articles provide an interesting analysis of the actual nature of hub proteins, distinct from their positions in interactome maps (Stein *et al.* 2009; Tsai *et al.* 2009b; Tyagi *et al.* 2009).

5. Intrinsic disorder as function: linkers and repeats

As mentioned in Section 3, one of the major functional roles of intrinsically disordered segments appears to be in the provision of linker sequences between interaction domains that may or may not be well-structured. For example, the CBP protein (Fig. 1) contains well-structured domains such as KIX and the TAZ domains, and unstructured interaction domains such as the nuclear coactivator binding domain (NCBD), and in between these sequences there are frequently lengthy segments that are not conserved in amino acid sequence or even necessarily in length, but are conserved in their amino acid composition (Dyson & Wright, 2005).

5.1 Linkers

Although the role of linker sequences is likely to be primarily topological, allowing distant parts of the polypeptide chain to interact with diverse partner sequences that might be far apart or

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

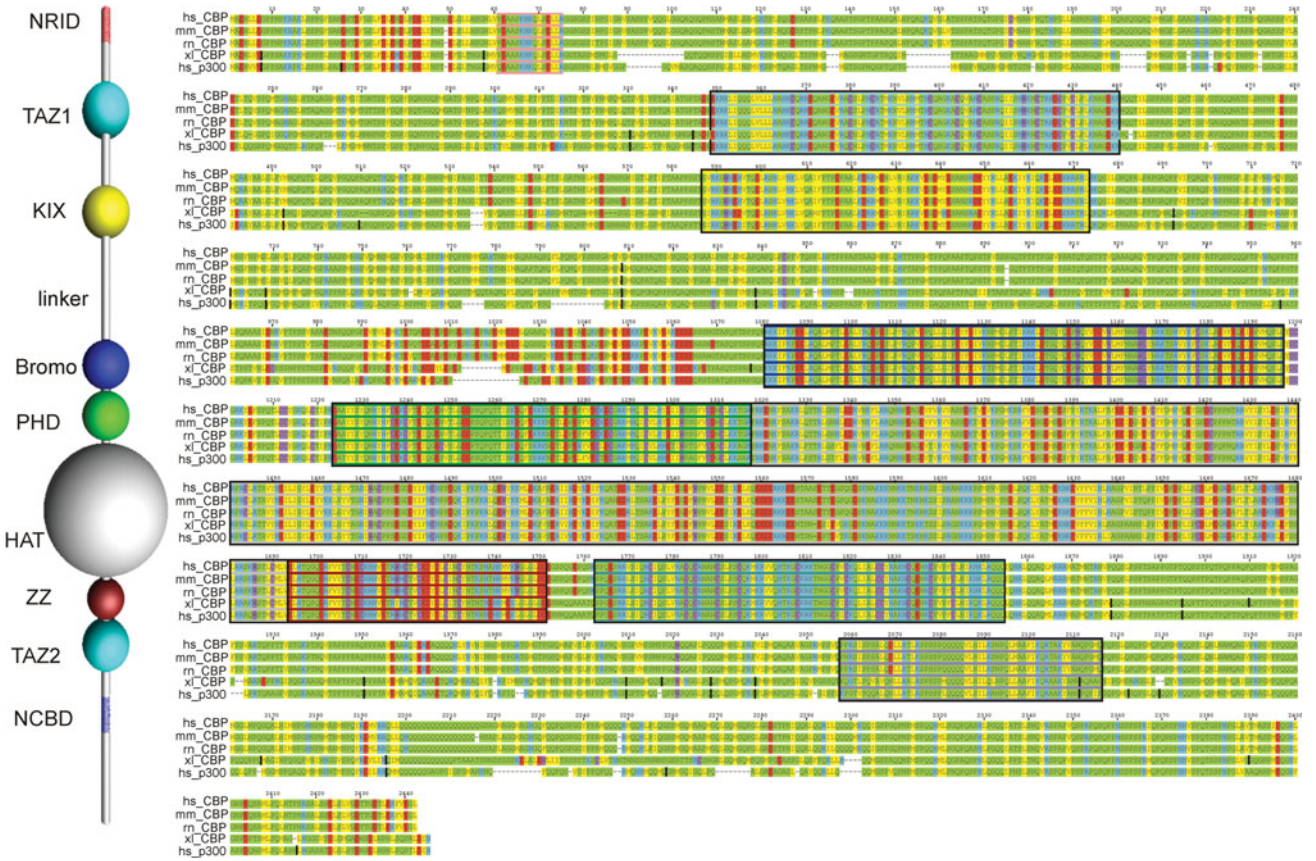


Fig. 1. Schematic figure showing (left) the domain structure of CBP and (right) alignment of amino acid sequences of CBP from human (hs), mouse (mm), rat (rn), frog (xl) and its paralog p300 from human (hs). Amino acids are colored according to the classification acidic Glu, Asp (red); basic Lys, Arg, His (blue), hydrophobic Val, Leu, Ile, Phe, Tyr, Met (yellow), rare Cys, Trp (purple) and disorder-promoting Gly, Ala, Ser, Thr, Pro, Asn, Gln (green). Vertical black lines indicate the presence of short insertions relative to the sequence of human CBP. Aligned sequences corresponding to the structured domains denoted by spheres on the left are boxed with the corresponding color.

close together, linkers and unstructured tail sequences play quite specific roles in a number of systems. The critical role of a linker sequence in the operation of the E2-RING finger switch in the ubiquitylation complex was recently demonstrated by a combination of biophysical measurements (Das, 2009), and the further role of linker motions in the Cullin-RING E3 ligases was suggested by molecular dynamics calculations (Liu & Nussinov, 2010a). Further details on the role of intrinsic disorder in the ubiquitin system are presented in Section 9. Computational examination of the influence of the presence of a flexible linker on the affinity of multi-domain transcription factors for DNA (Vuzman *et al.* 2010a, b) showed a significant increase in expected affinity, which was ascribed to an enhanced ability to slide along the DNA sequence to find specific binding sites. This sliding from non-specific to specific binding was mediated by NA-binding domains of different affinities and by the presence of unstructured tails on the proteins. The detection of secondary structure preferences in a disordered peptide derived from aggrecan (Jowitt *et al.* 2010) and the apparent presence of a semi-ordered structure in this sequence provides another variation on the linker function of disordered domains. In this case, there appears to be a requirement for more rigidity than would be available for a fully disordered chain.

5.2 Scaffolding proteins

The detection of intrinsic disorder in scaffolding proteins using bioinformatic methods has been extensively reviewed (Cortese *et al.* 2008). The role of disorder in scaffold proteins was hypothesized to involve many of the classic rationales, including fly-casting and masking of intramolecular interactions. Experimental verification of the presence of extensive disorder in an important scaffolding protein, Axin, was recently published (Noutsou *et al.* 2011).

5.3 Repeat sequences

Among the most common fully disordered regions in the proteome are those that consist of repeats of short sequence motifs. In addition, however, repeat sequences also comprise a number of structured domains, such as zinc finger repeats, ARs and armadillo repeats. Structured repeat sequences will not be considered further in this review, but are briefly discussed, with extensive references, by Matsushima *et al.* (2008). This excellent review focuses in detail on the various, largely unstructured, tandem repeat sequences presently available in sequence databases, as well as cataloguing the various conformational preferences that tandem repeat sequences may take up. The reader is referred to the 'outstanding questions' and 'concluding remarks' sections at the end of this review for a cogent summary of the main issues surrounding disordered repeat sequences.

The connection of repeat sequences, especially when they are expanded beyond normal numbers, to diseases such as Huntington's disease highlights an important characteristic of IDPs – their tendency to form aggregates. Detection of fibrils and plaques are used to diagnose degenerative diseases, although the actual toxic species may rather be a smaller oligomer than a large, presumably inert aggregate (Merlini & Bellotti, 2003). The aggregation mechanism of the expanded polyglutamine repeat sequences that occur in Huntington's disease has been shown to be complex and strongly length dependent (Thakur *et al.* 2009). Repeat sequences are associated with a disease of a different nature: both the merozoite and sporozoite forms of the malaria parasite *Plasmodium falciparum* contain highly hydrophilic repeat sequences, which can form both

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

local secondary structures similar to reverse turns (Dyson *et al.* 1990) and amyloid-like fibrils (Adda *et al.* 2009).

5.3.1 Examples

Long intrinsically disordered regions that most likely function as linkers are no longer being ignored as a nuisance in structure determination of folded domains though they remain an important bar to the crystallization of proteins. Structural characterization of proteins containing disordered regions is more facile using solution methods such as nuclear magnetic resonance (NMR) or computational methods such as molecular dynamics, and a number of systems have recently been reported. These include G protein-coupled receptors, where the disordered regions appear to contain an unusual preponderance of positively charged residues (Lys, Arg and His), thought to be involved either in positioning of the transmembrane helices or in interacting with specific signaling partners (Jaakola *et al.* 2005).

Human replication protein A has been characterized both structurally and dynamically by NMR (Olson *et al.* 2005). NMR relaxation and relaxation dispersion measurements provide biophysical information on dynamics and segmental motion that is unparalleled (Palmer, 2004). However, the analysis of the data for linkers and other disordered regions attached to more ordered or structured domains has been somewhat problematic: the presence of the disordered chain precludes the employment of a number of simplifying assumptions that are routinely used for analysis of relaxation data for folded proteins. Further details and examples of NMR studies of IDP dynamics are presented in Section 11.

Another example that has received both experimental and computational attention is the interaction between two domains, the PDZ and BAR domains, of PICK1, a conserved membrane protein (He *et al.* 2011; Xu & Xia, 2006). PICK1 is one of the few proteins that contain both a PDZ (known for interacting with membrane proteins) and a BAR (which binds to lipids). The two domains are separated by a linker sequence of some 40 residues. Molecular dynamics simulations suggest that the role of the linker in PICK1 is to facilitate initial weak hydrophobic interactions between the two domains, subsequently allowing sliding and repositioning of the domains for the most favorable interaction (He *et al.* 2011).

5.4 Alternative splicing, evolution and evolvability

An important source of intrinsic disorder in protein variants is the existence of alternative splicing of mRNA, which might be expected to increase the likelihood of disorder in one or more splice variants by the addition or excision of a peptide segment that may disrupt the 3D structure present in the alternative variant (Pentony & Jones, 2010). Alternative splice sites frequently occur within intrinsically disordered segments, which allows for the requisite diversity of function and interaction without causing major structural disruption (Romero *et al.* 2006). Many of the factors that mediate alternative splicing of mRNA are themselves intrinsically unstructured, which appears to be necessary for spliceosome assembly and for interaction with many partners (Haynes & Iakoucheva, 2006; Wang *et al.* 2010). Alternative splicing confers variability, with possible advantages for evolutionary fitness, but may also be a prelude to disease, as in the case of oncogenic fusion proteins (Hegyi *et al.* 2009, 2011).

Dual coding in alternate reading frames of the same gene frequently correlates with intrinsic disorder (Kovacs *et al.* 2010), which may provide a mechanism for the generation

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

of novel IDPs. New proteins are also thought to arise through overlapping genes (Rancurel *et al.* 2009), and the rate of evolution appears to be greater for proteins containing intrinsically disordered or repeat sequences (Alba *et al.* 2007; Brown *et al.* 2002). Interestingly, the frequency of repeated and intrinsically disordered sequences is much greater in higher organisms than in prokaryotes (Alba *et al.* 2007); the highest degree of disorder overall appears to reside in the lower eukaryotes (Mohan *et al.* 2008) (see, for example the surface proteins of the malaria parasite mentioned above), perhaps as an aid to protein–protein interactions (Mohan *et al.* 2008).

6. Intrinsic disorder as part of function: coupled folding and binding

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 (otherwise known as MoRFs or MoREs, see Section 4). The past few years have seen a meteoric increase in the number and richness of the systems where coupled folding and binding are seen. This field has been reviewed several times recently (Dyson and Wright, 2002, 2005; Uversky, 2010; Wright & Dyson, 2009), but the concept is so central to the entire rationale for the existence of IDPs that work on several systems mentioned in these previous reviews will be expanded and updated in the following paragraphs. A recent rather comprehensive compendium of disorder-related complexes can be found in Uversky (2010).

6.1 Transcriptional activation: partners of the KIX domain of CBP

Among the earliest domains that were identified as disordered in their functional state were many of the interacting partners of the transcriptional coactivator CBP and its paralog p300 (Fig. 1). This molecule, 2441 amino acids in length, serves as a central enabler in the transcription of activated genes, providing a bridge between the upstream activation site and the transcriptional initiation complex. CBP/p300 contains a number of structured domains, linked by disordered regions of varying length. Some of the interacting domains of CBP are themselves intrinsically disordered, and fold upon interaction with their partners in a process that we have termed 'mutual synergistic folding' (Demarest *et al.* 2002), which is described more fully in a later paragraph of this section. One of the best-characterized interaction domains of CBP is the KIX domain, spanning residues 587–673. This domain is independently folded in the absence of binding partners, but most of its partners are disordered, and fold upon binding. This was first demonstrated for the pKID of CREB (Radhakrishnan *et al.* 1997), which is disordered in the absence of KIX (Radhakrishnan *et al.* 1998). The unphosphorylated domain has a much lower affinity for KIX; an interaction between unphosphorylated KID and KIX can be detected by NMR (Zor *et al.* 2002) though previously unseen by less sensitive methods. Importantly, phosphorylation does not affect the conformational equilibrium in free KID (Radhakrishnan *et al.* 1998). The affinity of pKID for KIX is dominated by enthalpy terms related to the charge of the phosphorylated Ser133 of CREB (Zor *et al.* 2002), and is not significantly affected by the presence or absence of residual secondary structure in the free pKID or KID or an alternative partner c-Myb (Zor *et al.* 2002). The status of KIX as a multiple-site binding module was established by a comparison of the binding of a third partner, the mixed lineage leukemia (MLL) interaction domain, at a completely independent site on KIX (Goto *et al.* 2002). Interactions of

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

two partners with KIX to form ternary complexes (KIX/MLL/c-Myb or KIX/MLL/pKID) showed a 2-fold increase in affinity for pKID or c-Myb in the presence of KIX and MLL, compared with KIX alone (Goto *et al.* 2002). The structural basis for this cooperative behavior appears to be the presence of small structural changes and additional ordering of the KIX domain in the ternary complex (De Guzman *et al.* 2006).

An extensive NMR study, employing relaxation dispersion and chemical shift titrations, showed that the mechanism of coupled folding and binding of pKID to KIX employed 'folding upon binding' under NMR conditions in preference to the 'conformational selection' model (Sugase *et al.* 2007a). This observation has profound implications for the thermodynamic and kinetic basis for coupled folding and binding, and is discussed, with several other examples showing different behavior, in Section 7.

6.2 Disordered regions of the tumor suppressor p53

As a central player in the regulation of cell growth and division, as well as the site of many known mutations in transformed cells leading to cancers, the tumor suppressor p53 has received an enormous amount of attention. This molecule contains both ordered and disordered regions (Fig. 2). The ordered DBD binds DNA, while the TD mediates tetramerization of the protein. p53 is comprehensively regulated by post-translational modification, principally phosphorylation and acetylation, which modulate the function and the lifetime of the protein in the cell, in response to external signals. The pre-eminent region that undergoes phosphorylation is the N-terminal TAD, which contains seven major phosphorylation sites. The TAD is disordered in solution, but becomes ordered to various degrees and in various sites when bound to partner molecules such as MDM2 (Kussie *et al.* 1996) and the KIX, TAZ1, TAZ2 and NCBD domains of CBP/p300 (Feng *et al.* 2009; Ferreon *et al.* 2009b; Lee *et al.* 2009, 2010b). Phosphorylation changes the affinity of p53 for partners such as MDM2 and the CBP domains (Ferreon *et al.* 2009b), and multiple phosphorylations enhance and modify the interactions (Lee *et al.* 2010a). Residual structure in the AD1 and AD2 regions of the p53 TAD has been detected by NMR (Ferreon *et al.* 2009b; Lee *et al.* 2000) and inferred from principal component analysis of the conformational ensemble of the free protein (Lowry *et al.* 2008a). The C-terminus of p53, another intrinsically disordered region, adopts multiple structures upon binding different partners (Oldfield *et al.* 2008), apparently mediated both by charge and hydrophobic interactions (Chen, 2009).

6.3 Mutual synergistic folding

An extreme form of coupled folding and binding occurs when both partners are more or less disordered in their free state, but become well-ordered upon interaction. Such an interaction occurs between the p160 nuclear hormone receptor coactivator and the NCBD of CBP/p300 (Demarest *et al.* 2002). The free form of the p160 interaction domain (termed ACTR) is almost completely disordered in solution, whereas the NCBD of CBP retains some helical structure according to the CD spectrum, but is not apparently cooperatively folded (Demarest *et al.* 2002) and thus may be termed a molten globule-like state (Demarest *et al.* 2002; Kjaergaard *et al.* 2010). The complex is cooperatively folded and highly helical. Other complexes of the NCBD with different partner proteins show differences in the NCBD structure, which is perhaps not

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

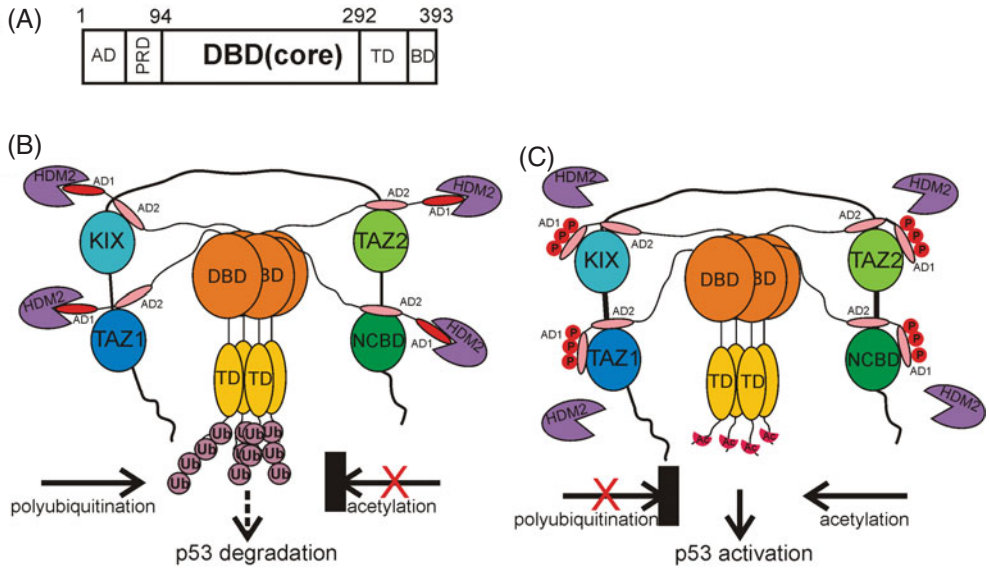


Fig. 2. A. Schematic diagram showing domains of the tumor suppressor p53. AD: activation domain; PRD: proline-rich domain; DBD: DNA binding domain; TD: tetramerization domain; BD: C-terminal regulatory domain. B. Schematic figure illustrating the model for the interactions of p53 with CBP and HDM2 (adapted from Ferreon *et al.* 2009b).

surprising, given the malleability of the free state molten globule. The partner proteins take up widely differing positions on the surface of the NCBD. This subject is described more fully in Section 8.

Another variation on the theme of synergistic folding is provided by the NF-κB–IκBα system. In the absence of extracellular signals, NF-κB is present in the cytoplasm in complex with IκBα. The premier sources of binding energy for this complex are interactions that occur at either end of the IκBα AR domain (Bergqvist *et al.* 2008), both involving coupled folding and binding. The C-terminal region of IκBα, consisting of ARs 5 and 6 and the proline-glutamic acid-serine-threonine (PEST)-containing sequence, folds upon binding to NF-κB (Sue *et al.* 2008; Truhlar *et al.* 2006). Conversely, it is the intrinsically disordered nuclear localization sequence of the p65 subunit of NF-κB that undergoes a binding-associated folding upon interacting with IκBα (Bergqvist *et al.* 2006; Cervantes *et al.* 2010). Clearly in this case the coupled folding and binding of the two interacting molecules represent a sensitive adjustment of the formation of the complex to both thermodynamic and kinetic demands in a system that must respond rapidly and correctly to signals as they arrive.

6.4 Calmodulin signaling

Calcium signals are transduced by calmodulin through its interactions with a number of partner proteins. Bioinformatic studies revealed that these protein–protein interactions likely involved the coupled binding and folding of disordered regions in both calmodulin and its partner proteins (Dunker *et al.* 1998; Radivojac *et al.* 2006). The flexible linker between the two calcium-binding domains mediates the interactions with partner proteins, allowing the calcium-binding domains to wrap around the partner.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

6.5 Proline-rich sequences and their partners

Eukaryotic signal transduction frequently involves the interaction of proline-rich sequences with small folded domains such as the SH3 (Mayer *et al.* 1988; Stahl *et al.* 1988) and WW (Bork & Sudol, 1994) domains. The interactions are dynamic and weak, as befits their role in signaling (Gu & Helms, 2005).

6.6 Binding without folding

Although the classic idea of 'binding' implies an increase in order, a number of interactions between disordered proteins appear to retain significant disorder in one or both binding partners. Such interactions, which have been termed 'fuzzy' (Tompa & Fuxreiter, 2008) or 'cloud contacts' (Uversky, 2010) may consist of multiple weak interactions: the IDP uses multiple binding sites, while contacting multiple sites of the partner. Some examples include the interactions of the cystic fibrosis transmembrane conductance regulator (CFTR) regulatory region with the NBD1 (Baker *et al.* 2007), the Sic1 protein with an ubiquitin ligase subunit (Borg *et al.* 2007; Mittag *et al.* 2010), and the cytoplasmic domain of the T-cell receptor ζ chain with the SIV nef protein (Sigalov *et al.* 2008).

7. Thermodynamics and kinetics

The thermodynamic underpinning of the IDP phenomenon retains a fascination for protein chemists of all persuasions. If the protein is disordered in its free state, then there are many low-energy states that are significantly populated, in contrast to the single (or small number of closely-related) states populated by folded proteins. Thus, the primary operative feature of an IDP would appear to be elevated entropy, which must be counteracted (at some energy cost) when the IDP folds upon binding to its partner. It follows that the enthalpy change associated with the folding and binding process must be large enough to compensate, and this appears to be the case, for example, in the binding of the phosphorylated form of the kinase-inducible domain of CREB (pKID) to the KIX domain of CBP (Zor *et al.* 2002). Nevertheless, there are other factors that are needed to fill out this simplistic picture, and a number of groups have addressed the problem. A quantitative thermodynamic theory for IDPs (Liu *et al.* 2009) was used to predict the types of functional proteins that might have a greater likelihood of employing intrinsically disordered regions. While catalytic proteins and low-affinity binding proteins appear to have a preference for ordered structures, disorder was predicted to be significantly more prevalent for binding proteins with high affinities (Liu *et al.* 2009). The thermodynamic basis for the variation in binding-induced biomolecular switches was examined (Vallee-Belisle *et al.* 2009), and was formulated in terms of a trade-off between the potential amplitude of the signal change and the affinity of the interaction. This trade-off was also invoked to explain the results of a computational study on a structural protein–protein interaction network dataset (Carbonell *et al.* 2009). The question 'How can peptides overcome the entropic cost involved in switching from an unstructured, flexible peptide to a rigid well-defined bound structure?' (London *et al.* 2010) was addressed by analyzing a database of high-resolution structures of peptide–protein complexes. The entropy cost appeared to be minimized when there was little conformational change of the partner protein upon binding. Other contributory factors included the optimization of packing at the interface, the increased number of hydrogen bonds formed at the interface and the

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

disproportionately large contribution to the binding energy of particular residues termed 'hot spots' (London *et al.* 2010). The interaction of proteins with DNA provides another example where enthalpy and entropy appear to have different roles in the interactions: binding of proteins in the DNA major groove is mostly driven by enthalpy, while minor groove binding is entropically driven (Privalov *et al.* 2007). This distinction appears to be related to the difference in hydration properties between the major and minor grooves of DNA, which gives a clue as to why the entropic penalty for coupled folding and binding is not prohibitively large. While an IDP does not favor a single-folded conformation in isolation, it may nevertheless contain hydrophobic residues (possibly specifically related to the 'hot spots' of binding) which in general have unfavorable hydration properties. When these hydrophobic side chains are folded into a core at the interface of a complex, their hydration water is released, with consequent increase in the entropy of the system. This influence of water and the hydration of IDPs can be seen in the behavior of some (but not all) IDPs in the presence of osmolytes. For example, addition of trimethylamine-N-oxide (TMAO) to a solution of the protein component of bacterial ribonuclease P, which folds upon binding to the RNA portion, resulted in folding of the protein even in the absence of its partner (Chang & Oas, 2010).

7.1 Binding-induced folding versus conformational selection

How do the two partners of a coupled-folding-and-binding interaction make contact and form the complex? This problem has been formulated in terms of two limiting models, binding-induced folding, where interaction between the partners occurs as the first step, followed by folding on the surface of the partner, and conformational selection, where appropriately structured members of the ensemble of the IDP are selected to form the complex. In reality, most coupled folding and binding reactions will take place using elements of each of these mechanisms, and it has been suggested that the models have so much in common that they should be merged into a 'synergistic model' (Espinoza-Fonseca, 2009a). Another treatment argued that an extended conformational selection model could be used to encompass both 'conformational selection' and any elements of 'induced fit' that might be necessary for the complex to form (Csermely *et al.* 2010). The two limiting cases were distinguished by calculating the flux through each pathway (Hammes *et al.* 2009). This analysis showed that conformational change of flavodoxin upon binding of its cofactor flavin mononucleotide (FMN) was dominated by conformational selection at low ligand concentration and by binding-induced folding at high concentrations, with a significant range of conditions where both pathways were operative. Both models have also been invoked to explain the folding of various mutants of staphylococcal nuclease upon binding of substrate (Onitsuka *et al.* 2008). A system that shows more concrete evidence for conformational selection is the discrimination between various disordered inhibitors of protein phosphatase 1, where a comparison of ensemble models generated for three disordered peptide ligands with structures of the complexes implicated transient, preformed structure in the interaction (Marsh *et al.* 2010).

The most comprehensive study of the mechanism of a particular coupled folding and binding interaction was the analysis by NMR chemical shift perturbation and relaxation dispersion of the binding of pKID to KIX (Sugase *et al.* 2007a). The binding of pKID to KIX is quite tight, with a K_d of $\sim$700 nM (Zor *et al.* 2002), and the exchange of pKID between free and bound states is slow on the NMR chemical shift timescale. Nevertheless, at molar ratios close to 1:1, additional effects are seen in the NMR spectrum, consistent with an additional

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

process occurring on a fast-exchange timescale (Fig. 3) (Sugase *et al.* 2007a). This was interpreted as evidence of an intermediate state that was populated during the binding process. Relaxation dispersion measurements were used to delineate the structure of this state; the results indicated that pKID binds initially to KIX in a series of encounter complexes that resolve initially into a partially folded intermediate state and then into the fully-folded complex, in a process that is largely identifiable as 'binding followed by folding' rather than the selection of a pre-formed conformer in the 'conformational selection' mechanism. Molecular dynamics simulations of this system (Huang and Liu, 2010a, b; Turjanski *et al.* 2008) largely confirm the general outline of this mechanism. The role of non-native interactions during binding and folding was noted as a potentially important operative factor in the kinetics of the process (Huang & Liu, 2010a).

7.2 Kinetic discrimination in antibody recognition

It has long been recognized that the antigen-binding sites of antibodies are relatively mobile (Tainer *et al.* 1984). A more recent body of work seeks to address the question 'How can antibodies be both promiscuous and specific?' (James *et al.* 2003). One example is a monoclonal IgE antibody raised to the hapten 2,4-dinitrophenol (DNP), which is sufficiently specific to discriminate between nitrophenol and DNP, but which can also bind protein ligands (James & Tawfik, 2003). It appears that the antigen-binding site of this antibody is in a conformational equilibrium between structures that can bind DNP and undergo an induced-fit conformational change to form a long, narrow-binding pocket specific for the hapten, and wider, shallower pocket structures that can accommodate protein antigens (James & Tawfik, 2005; James *et al.* 2003). Such multi-specificity may well be the explanation for the phenomena of auto-immunity and allergy, where antibodies raised against pathogenic antigens can also bind components of the organism itself or innocuous environmental molecules.

7.3 Thermostability and intrinsic disorder

Thermophilic organisms, those that prefer to live at temperatures above the normal physiological temperatures preferred by mesophilic organisms, contain proteins that are adapted in their structure and, particularly, in their dynamic properties, to higher temperatures (Petsko, 2001). Even in a mesophilic cell, that is, one that normally lives and grows at around room temperature, there nevertheless exist a repertoire of proteins that are resistant to higher temperatures, and this repertoire is enriched in disordered proteins (Galea *et al.* 2006). A systematic study of proteins that are enriched after heat treatment of mouse fibroblast cells (Galea *et al.* 2009) showed preponderance of soluble disordered proteins and of proteins with large segments that are predicted to be disordered. Even some sizable, fully-folded proteins were shown to be resistant to heat treatment (Galea *et al.* 2009).

A contrasting result was shown in a survey of proteins from prokaryotic thermophilic organisms, which were found to employ fewer disordered regions in several protein types, compared to their analogues in mesophilic organisms (Burra *et al.* 2010). Transcription factors, for example, were shown to be less disordered in thermophiles, but proteins required for translation and ribosome biogenesis retained their disordered characteristics. The authors concluded that the lower incidence of disorder in thermophilic organisms was likely the result of functional simplification as a result of adaptation to extreme conditions.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

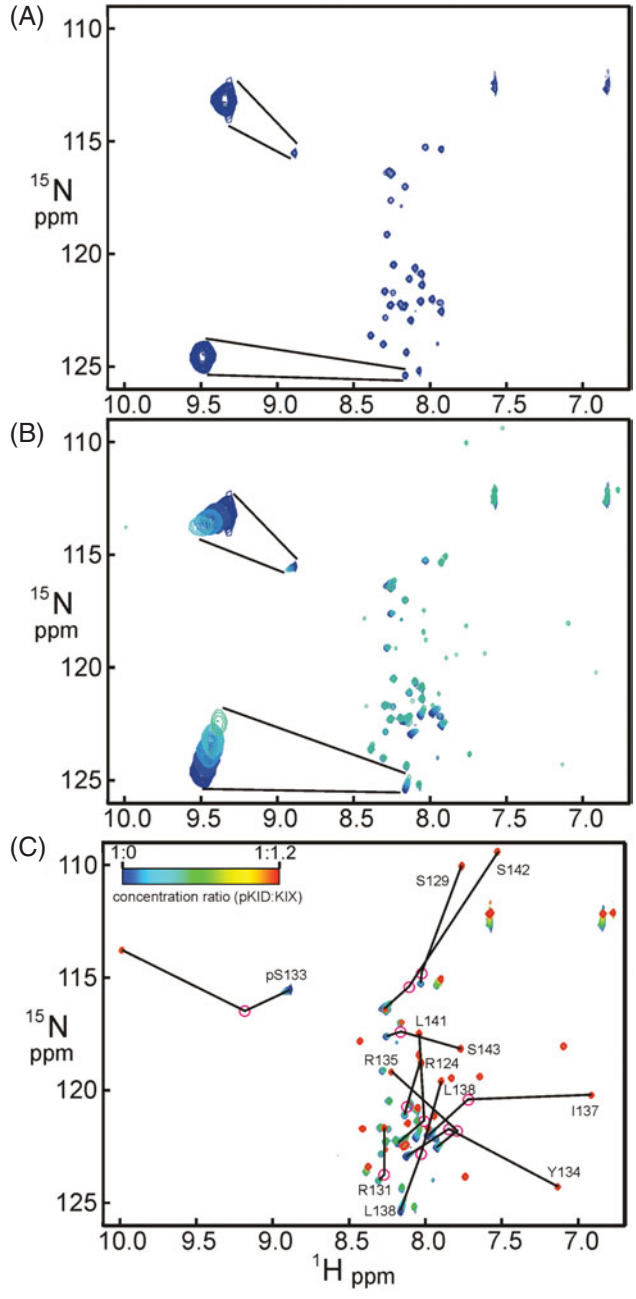


Fig. 3. NMR ^{1}H-^{15}N HSQC spectra illustrating the fast- and slow-exchange processes occurring as the KIX domain of CBP is added to the intrinsically disordered pKID domain of CREB. (A) HSQC spectrum of pKID in the absence of KIX. Insets show enlarged versions of (top) the cross peak belonging to the phosphorylated Ser133 and (bottom) the cross peak of Leu138. (B) Superimposed HSQC spectra of pKID in the absence of KIX (blue) and in the presence of 1:0.1, 1:0.2, 1:0.3 and 1:0.4 mole ratios of KIX (colors progressing toward green). Insets show enlarged versions of the pSer133 and Leu138 cross peak sets. (C) Superimposed spectra for the complete titration. Corresponding assignments in the spectrum of the free and complexed pKID are linked by arrows. The pink circles represent cross peak positions calculated on the basis of the crosspeak movements illustrated in part B for the partly folded intermediate state (Sugase *et al.* 2007a).

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

7.4 Allostery and intrinsic disorder

Allostery is the term used to denote the transmission of regulatory information through conformational changes in proteins. It follows that this phenomenon is intimately connected with dynamics in proteins and with the observation of disordered segments in many important regulatory proteins; these ideas have recently been reviewed (Gibson, 2009; Hilser, 2010; Smock & Gierasch, 2009; Tsai *et al.* 2009a). Experimental results are now available for several systems. The cooperative response of the bacterial flagellar switch appears to be related to its stochastic multistate nature (Bai *et al.* 2010). Another system where cooperativity arises in the presence of intrinsic disorder is the toxin–antitoxin repression operon in bacteriophage P1, where interactions between the disordered C-terminus of the antitoxin Phd and the toxin Doc mediate repression of transcription of the on (Garcia-Pino *et al.* 2010). Other toxin–antitoxin modules show variable levels of structural change upon binding (Drobnak *et al.* 2009; Lah *et al.* 2005).

7.5 Compensation for entropy changes in binding

One possible way that a disordered protein may compensate for the loss of entropy as it folds upon binding is to transfer the flexibility to another region of the complex. This mechanism was invoked to interpret NMR data observed for the AR domain of IκBα in the free state and in complex with its partner NF-κB (Sue *et al.* 2008), where a region of high mobility, resembling a molten globule state, was observed in the two C-terminal ARs in the free protein, but in the complex a different part of the molecule, the third AR, showed increased mobility (Sue *et al.* 2008). A similar redistribution of more dynamic regions of the protein upon binding was seen for the complex of the VTS1p-SAM domain with an RNA hairpin (Ravindranathan *et al.* 2010) and for the binding of stromelysin-1 to tissue inhibitor of metalloproteinases 1 (TIMP-1) (Arumugam *et al.* 2003).

8. Kinetic control of signaling by competition for binding sites

Although the coupled folding and binding of IDPs can (and must) be justified in thermodynamic terms, it is likely that much of the impetus for the occurrence of IDPs in nature is kinetic. Metabolic control involves the ability of a living system to change in response to stimuli. The fundamental chemical mechanism of response to environmental change appears to be competition of signaling factors for receptors, a kinetic phenomenon.

8.1 Binding of ligands to the KIX domain of CBP

The potential for competition between partners for binding sites on scarce transcriptional activator molecules is illustrated by the CBP system (shown in Fig. 1). CBP consists of a long single polypeptide chain that contains a number of variously-sized folding domains, including a histone acetyl transferase domain and a series of smaller folded interaction domains, including KIX (mentioned above in Section 7), and the TAZ domains. The KIX domain has two major binding sites for partners, one in the center of the molecule that accommodates ligands such as the pKID of CREB and the interaction domain of c-Myb (Zor *et al.* 2002), and the other at one end of the molecule that binds the activation domain of the mixed-lineage leukemia (MLL) protein (Goto *et al.* 2002). Interestingly, the two sites are allosterically coupled, resulting in tighter binding for ligands in ternary complexes than in the corresponding binary complexes (De Guzman *et al.*

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

2006; Goto *et al.* 2002). The KIX domain also interacts with the tumor suppressor p53 (Lee *et al.* 2009). The FOXO3a transcription factor provides a variation on the competition for binding to the KIX domain (Wang *et al.* 2009). This protein consists of a well-folded forkhead (FH) domain that interacts with DNA, and a disordered domain containing three conserved regions (CR1, CR2 and CR3). The CR3 sequence interacts both with the KIX domain of CBP and intramolecularly with the FH domain. In the absence of DNA, the interaction of CR3 with FH prevents its interaction with KIX, but when FH binds DNA, the CR3–KIX interaction can occur, thus recruiting the CBP and its histone acetyltransferase activity to the transcriptional machinery at the transcription site.

8.2 Binding of ligands to the TAZ domains of CBP

The TAZ domains exhibit a different form of multiple-ligand binding. There are two TAZ domains in CBP, termed TAZ1 and TAZ2 (De Guzman *et al.* 2004b). Both domains are predominantly helical, with three structural zinc binding sites. The 3D structures of these domains show significant differences, and there is a corresponding difference in their ligand affinity. For example, both TAZ1 and TAZ2 bind the N-TAD of p53. Binding occurs in two separate sites on each of the TAZ domains, but with significantly different affinities (Ferreon *et al.* 2009b), and the affinities for the two domains are changed in different ways when p53 is phosphorylated (Ferreon *et al.* 2009b; Lee *et al.* 2010a). The interaction of CBP with p53 provides an excellent example of the synergy of thermodynamic and kinetic factors in determining the response of transcriptional systems to signals. In unstressed cells, the N-TAD is unphosphorylated, and binds to the ubiquitin ligase human double minute protein 2 (HDM2) as well as to the CBP domains TAZ1, TAZ2, KIX and NCBD (Ferreon *et al.* 2009b). As a consequence, p53 is polyubiquitinated and degraded. Upon phosphorylation of p53 as a result of genotoxic stress, the affinity of HDM2 is significantly reduced, and the HDM2-binding segment of p53 N-TAD becomes bound exclusively to the CBP domains. p53 is no longer ubiquitinated, but instead is acetylated and activated for the transcription of stress-related genes (Ferreon *et al.* 2009b) (Fig. 2b).

The p53 system provides an example of the competition of folded domains (HDM2, the TAZ and KIX domains) for an IDP ligand. The TAZ domains also illustrate the competition between numerous IDP ligands for a folded domain. Complexes between TAZ1 and the activation domains of a number of ligands have been solved, mainly in solution by NMR methods. The TAZ1 structures include the free domain (De Guzman *et al.* 2005), complexes with the activation domains of the transcription factor Hif-1α (Dames *et al.* 2002; Freedman *et al.* 2002), the transcriptional repressor CITED2 (De Guzman *et al.* 2004; Freedman *et al.* 2003) and the signal transducer and activator protein STAT2 (Wojciak *et al.* 2009). These activation domains all bind in overlapping regions of the central part of the TAZ1 domain. However, the binding sites are by no means identical, and the sense of the binding (N to C-terminal) of the ligands are in some cases completely different. This point is illustrated in Fig. 4. The interaction between the N-TAD of Hif-1α and TAZ1 (Fig. 4, red ribbon) bears some resemblance to the binding of the N-terminus of CITED2 (Fig. 4, blue ribbon). However, their C-termini are bound quite differently, with Hif-1α wound around in a short helical structure onto the opposite side of the TAZ1 domain, while the CITED2 C-terminus remains unstructured on the same face. The structure of the TAZ1–STAT2 complex (Fig. 4, green ribbon) provides a further significant structural variation: the N- and C-termini of STAT2 are swapped with respect to the Hif-1α and CITED2 complexes. The differences between the interactions of these competing domains suggests

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

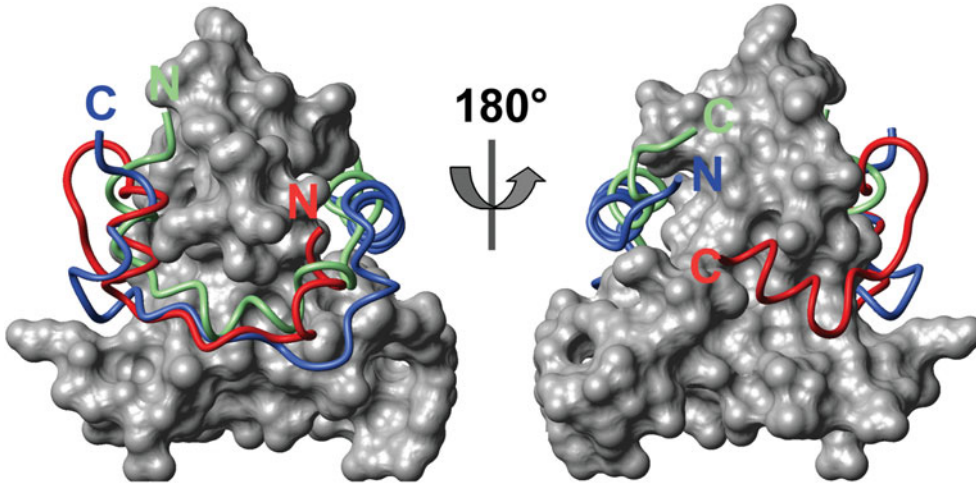


Fig. 4. Comparison of the structures of ligands bound to the CBP TAZ1 domain. The surface of TAZ1 (almost identical in all three complexes) is shown in gray, with the backbone of STAT2-TAD (Wojciak *et al.* 2009) in green, the HIF-1α-CTAD (Dames *et al.* 2002) in red and the CITED2-TAD (De Guzman *et al.* 2004a) in blue. The left and right images represent a 180° rotation around the vertical axis in the plane of the page. The N- and C-termini of each ligand are labeled: note that STAT2 binds in the opposite sense to the other two ligands (adapted from Wojciak *et al.* 2009).

a possible mechanism for the replacement of one domain by another: a competing domain could conceivably bind even in the presence of an already-bound domain, in one of the sites that the bound domain does not occupy, and then proceed to peel off and replace the bound domain by a process of local mass action.

8.3 Action of disordered domains of viral proteins

Viruses frequently utilize IDPs to subvert the control of infected cells (Davey *et al.* 2011). Competition for scarce resources such as CBP/p300 in the cell appears to be a major strategy used by viruses. The adenovirus early region 1A (E1A) protein utilizes its disordered domains to hijack the transcriptional machinery of the cell. The E1A protein interacts with the TAZ2 domain of CBP, forming a complex similar to those described in the previous paragraph for TAZ1 (Ferreon *et al.* 2009c). An N-terminal CR of E1A mediates the interaction, forming short helical segments in complex with TAZ2 (Ferreon *et al.* 2009c). E1A competes successfully with p53 for binding to TAZ2, disrupting transcriptional control by p53. Interestingly, E1A also mediates formation of a ternary complex with the retinoblastoma protein pRb and TAZ2, suggesting a mechanism for one of the physiological consequences of adenovirus infection, inhibition of phosphorylation of pRb and promotion of MDM2 binding. These events affect cell differentiation and promote permanent exit from the cell cycle (Nguyen *et al.* 2004). One of the absolute requirements for ternary complex formation is the presence of a long disordered sequence between the TAZ2 binding site at CR1 and the LXCXE sequence that binds to the B domain of pRb (Fig. 5).

8.4 Kinetic control in signal transduction

Kinetic control is also key to the understanding of other signal transduction pathways. Inhibition of NF-κB signaling by the inhibitor IκBα involves signaling-related phosphorylation and

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

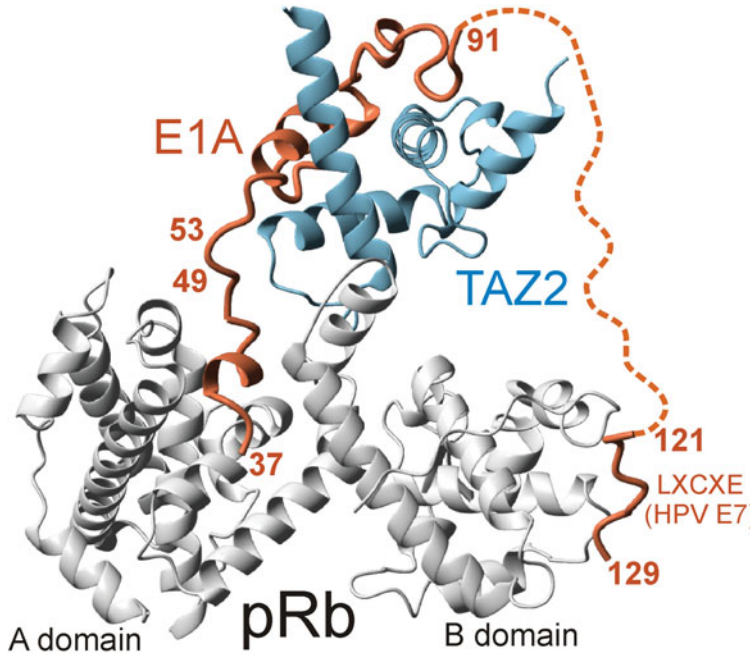


Fig. 5. Structural model of the ternary complex between the retinoblastoma protein pRb, the TAZ2 domain of CBP and the adenoviral E1A protein. The model was generated using the crystal structure of the complex of pRb with E1A (CR1, residues 37–49) (PDB entry 2R7G) (Liu & Marmorstein, 2007), the NMR structure of the complex between the TAZ2 domain of CBP and residues 53–91 of E1A (PDB entry 2KJE) (Ferreon *et al.* 2009c) and the crystal structure of the HPV E7 peptide (DLYCYEQLN, homologous to CR2 residues 121–129 of E1A) containing the LXCXE motif that interacts with pRb (PDB entry 1GUX) (Lee *et al.* 1998). The flexible linker between residues 83 and 120 of E1A is indicated schematically as a dotted line. The backbone structures of pRb, E1A and TAZ2 are represented as ribbons colored gray, coral and blue, respectively (adapted from Ferreon *et al.* 2009c).

degradation of IκBα in a process that incorporates interactions between folded, unfolded and partly folded domains of both proteins (described in more detail in Section 10). Another aspect of this system that has hitherto received less attention is the IκBα-mediated stripping of NF-κB from DNA following the transcriptional event. It was known that IκBα genes were frequently transcribed downstream of NF-κB-regulated genes. It has now been shown that the newly synthesized IκBα participates actively in the acceleration of NF-κB removal from the DNA (Bergqvist *et al.* 2009). The observation that a ternary complex was formed between NF-κB, IκBα and DNA in solution by NMR and stopped flow fluorescence (Sue *et al.* 2011) gives important insights into the possible molecular mechanism of the removal of NF-κB from DNA.

8.5 Mutual synergistic folding and kinetic control

As mentioned in Section 6, the term mutual synergistic folding was coined to describe the process whereby a folded complex could be formed when both components were disordered. The complex between the NCBD domain of CBP and the interaction domain of the nuclear receptor coactivator p160 (Demarest *et al.* 2002) represents one example of this class of complex. Free NCBD domain contains some helical structure, although it is not cooperatively folded (Demarest *et al.* 2002). Consistent with its identity as an incompletely folded domain, complexes

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

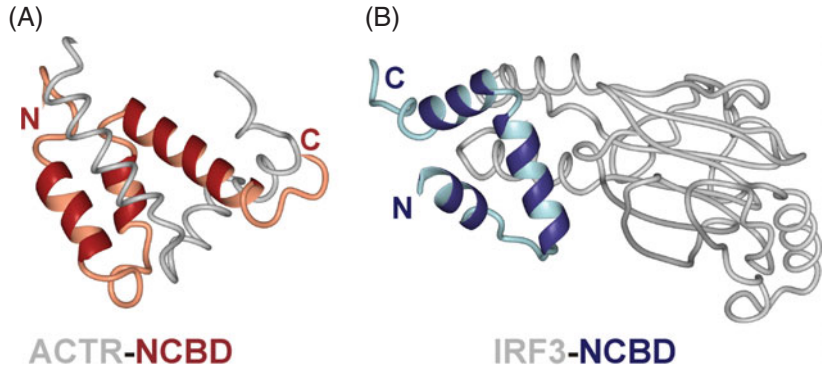


Fig. 6. Structural comparison of the NCBD domain of CBP in complex with (left) the ACTR domain of p160 (PDB entry 1KBH) (Demarest *et al.* 2002) and (right) IRF3 (PDB entry 1Z0Q) (Qin *et al.* 2005). Adapted from Wright & Dyson (2009). Note that free ACTR is intrinsically disordered, whereas IRF3 is a globular protein in the free state.

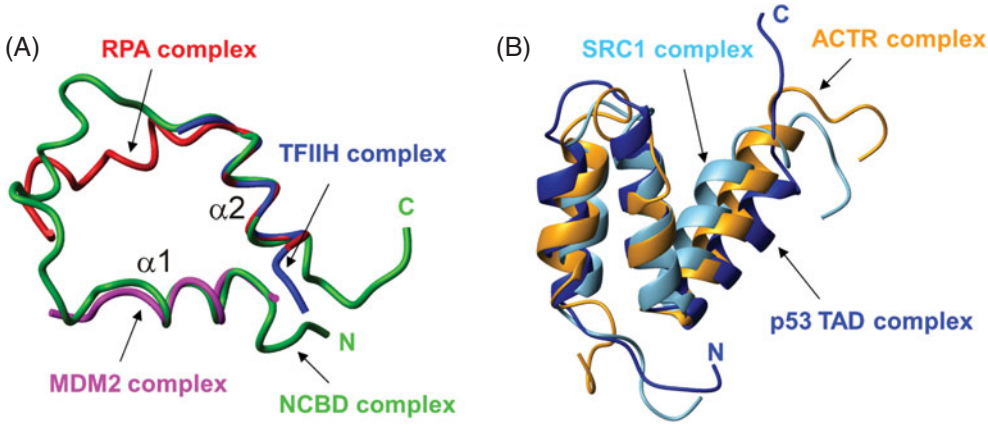


Fig. 7. (A) Superposition of p53 TAD structures in various complexes, with the NCBD domain of CBP (green) (PDB entry 2L14) (Lee *et al.* 2010b), with MDM2 (magenta) (PDB entry 1YCQ) (Kussie *et al.* 1996), with RPA (red) (PDB entry 2B29) (Bochkareva *et al.* 2005) and with TFIIH (blue) (PDB entry 2GS0) (Di Lello *et al.* 2006). The corresponding residues for each complex are aligned on the backbone of heavy atoms of the two well-defined helices, $\alpha 1$ and $\alpha 2$ of the NCBD complex. (B) Superposition of NCBD structures in various complexes, with the p53 TAD (blue) (PDB entry 2L14) (Lee *et al.* 2010b), with ACTR (PDB entry 1KBH) (Demarest *et al.* 2002) and with SRC1 (PDB entry 2C52) (Waters *et al.* 2006).

formed between NCBD and other interaction partners show considerable structural diversity, as illustrated in Fig. 6. The partners of NCBD in several complexes provide another illustration of the structural diversity accessible as a result of disorder (Fig. 7).

8.6 Control of the cell cycle by IDPs

The cell cycle consists of a series of folded proteins, the cyclins and cyclin-dependent kinases that are regulated in almost all respects by small domains or proteins that are largely if not completely disordered. One of the most important of these regulatory proteins is p53, which is implicated in a very high proportion of mutations leading to cancer. Other examples include the

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

cyclin-dependent kinase inhibitors p21, which was one of the first proteins to be designated 'intrinsically disordered' (Kriwacki *et al.* 1996), $p27^{Kip1}$ and $p14^{Arf}$.

The inhibitor $p27^{Kip1}$ mediates the formation of ternary complexes, with Cdk2 and cyclin A, which are significantly more thermally stable than any of the corresponding binary complexes, thus appearing to act as a 'thermodynamic tether' (Bowman *et al.* 2006). Phosphorylation of a tyrosine residue in the Cdk-binding region of p27 has been directly implicated in oncogenic transformation, as it changes the structure of the ternary complex, allowing partial activity of the Cdk and resulting progression of the cell cycle (Grimmler *et al.* 2007).

The inhibitor $p14^{Arf}$ interferes with p53 control by binding to HDM2 and preventing the normal degradation of p53 (Bothner *et al.* 2001). The interaction between Arf and HDM2 appears to be heterogeneous, with the formation of oligomers and amyloid-like fibrils (Bothner *et al.* 2003; Sivakolundu *et al.* 2008). Structural characterization of this system was achieved through studying the interaction of a 9-residue peptide to represent the interaction domain of Arf and the Arf-interacting domain of HDM2, which form bimolecular oligomers that were characterized by CD spectroscopy, sedimentation velocity measurements and NMR (Sivakolundu *et al.* 2008). These complexes distinctively showed the signature of β-strand formation (Bothner *et al.* 2001, 2003; Sivakolundu *et al.* 2008), unlike many of the other complexes so far studied, where the IDP components tend to form helices or loops, or may have heterogeneous 'fuzzy' characteristics.

9. Role of post-translational modification

As mentioned in Sections 3 and 4, one of the more obvious attributes of IDPs, particularly in cellular signaling systems, is that they are frequently the site of post-translational modifications. This makes sense if the primary sites of binding to partners are the small sequence motifs termed MoRFs: modification of a critical interacting residue might well serve to abrogate an interaction, for example. The occurrence of many different types of post-translational modification, sometimes at the same sites or in competing sites, argues that the picture is more complex than this simple analysis. This section will introduce a few examples of post-translational modifications on IDPs that are known to introduce specific physiological effects.

9.1 Direct relationship between chemistry and function: Hif-1α and TAZ1

TAZ1 is known to bind to a large number of proteins, through interaction domains that are frequently disordered in the free state. One of the most interesting interactions from the point of view of the direct relationship between chemistry and function is that between the TAZ1 domain and the hypoxia-inducible factor Hif-1α (Lando *et al.* 2002). Under normal conditions of oxygen concentration, the Hif-1α protein is hydroxylated at two positions, Pro564 and Asn803 (Fig. 8). Hydroxylation at Pro564 promotes binding of the Von Hippel–Lindau (VHL) ubiquitin ligase that results in the degradation of the protein. Hydroxylation at Asn803 provides a second tier of control by inhibiting the recruitment of CBP/p300 and hence halting potential transcriptional activation. Under conditions of low oxygen concentration (hypoxia), these two sites are no longer hydroxylated. Ubiquitin-dependent degradation of Hif-1α no longer occurs, and the protein dimerizes with the constitutive hypoxia factor ARNT and is translocated to the nucleus. There the heterodimer binds to CBP, with an interaction between the C-terminal disordered region of Hif-1α and the TAZ1 domain of CBP, to promote transcription of oxygen stress

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

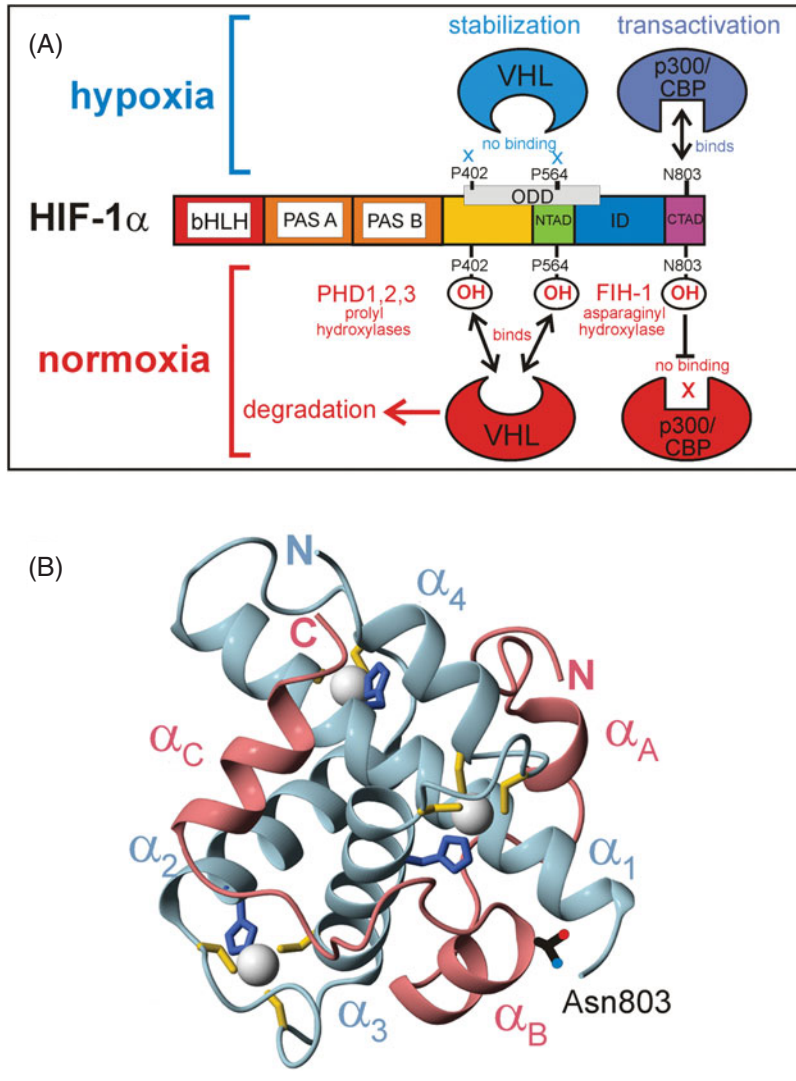


Fig. 8. (A) Schematic diagram showing the regulation of the HIF-1α transcription factor under normal oxygenation conditions (bottom), where proline hydroxylation in the central ODD domain recruits the Von Hippel–Lindau factor, leading to degradation, and asparagine hydroxylation in the C-terminal activation domain lowers the affinity for transcriptional activators. In hypoxic conditions (top), neither the prolines nor the asparagine are hydroxylated, with the result that HIF-1α is stabilized and binds to CBP/p300 to promote transcription of hypoxia-response genes. Adapted from Hirota & Semenza (2005). (B) Backbone structure of one member of the family of NMR structures of the complex of HIF-1α CTAD with the TAZ1 domain of CBP (PDB entry 1L8C) (Dames *et al.* 2002). The HIF-1α hydroxylation site on at Asn803 on α_B is indicated.

genes. This process is an important target for anti-tumor drugs, since the inhibition of hypoxia-inducible genes should prevent the formation of new vasculature around tumors, disallowing uncontrolled growth (Takenaga, 2011). The structural basis for the interaction between TAZ1 and Hif-1α was reported by two groups (Dames *et al.* 2002; Freedman *et al.* 2002). The intrinsically disordered C-terminal sequence of Hif-1α binds to the folded TAZ1 domain through the formation of three short helical segments that are located in grooves on opposing faces of the

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

TAZ1 surface and connected with loops. The regulatory Asn803 is located on the C-terminus of the second helical segment (Fig. 8), where it makes hydrogen-bonding contacts with the Hif-1α Asp799 carbonyl and the TAZ1 Asp346 backbone NH. The entire side chain is deeply buried in the interface between the two proteins, and is packed against hydrophobic side chains of TAZ1. The addition of a hydroxyl group to the Asn side chain would presumably cause disruption of this hydrophobic interaction, resulting in a lowered affinity of (hydroxylated) Hif-1α for TAZ1. (Interestingly, the affinity of hydroxylated Hif-1α for TAZ1, though greatly reduced from that of the unhydroxylated form, is still quite substantial. Attempts to replicate the relaxation dispersion experiments of Sugase *et al.* (2007a) with TAZ1 and Hif-1a were unsuccessful, as the affinity is simply too great for the fast-exchange interaction to be observable. However, the hydroxylated form of Hif-1α was amenable to this technique (Sugase *et al.* 2007b), with the proviso that much lower concentrations had to be used, since the affinity of TAZ1 even for hydroxylated Hif was substantially greater than that of pKID for KIX).

9.2 Phosphorylation of IDPs and IDRs

Covalent addition of a phosphate group to the side chain hydroxyl groups of serine, threonine or tyrosine is one of the most important regulatory mechanisms in eukaryotic cells. Phosphorylation adds a relatively large negatively charged group, which may contribute to the enthalpy of binding of a ligand and hence enhance binding, or alternatively to the disruption of an otherwise hydrophobic binding pocket and the consequent dissociation of the complex. Numerous experimental studies of IDPs include phosphorylated forms, and examples are given below. The effects of phosphorylation are amenable to computational study (Narayanan & Jacobson, 2009), and bioinformatic studies have also pinpointed a connection between the location of phosphorylation sites and the probability that the surrounding sequences would be disordered, utilizing this observation as a basis for the prediction of phosphorylation sites (Iakoucheva *et al.* 2004).

The level of phosphorylation of p53 is a critical determinant of the gatekeeper function of p53 in the cell. Under normal conditions, p53 is unphosphorylated and degraded through its interaction with the E3 ubiquitin ligase HDM2. Genotoxic stress causes phosphorylation of p53, and a greater level of stress results in a greater level of multisite phosphorylation, primarily in the intrinsically disordered N-TAD (Lee *et al.* 2010a). This behavior was interpreted as a graded response: since p53 must compete with other ligands for binding to the transcriptional activator CBP, the addition of multiple phosphorylations raises the affinity of the p53 for its binding sites on CBP (the TAZ1, KIX, TAZ2 and NCBD domains). A greater efficiency of binding to CBP with increased numbers of phosphorylated sites provides a molecular rationale for the increase in the p53 response following prolonged or severe genotoxic stress (Lee *et al.* 2010a). A graded response was also invoked in the analysis of the 'ultrasensitive' response of the cyclin-dependent kinase inhibitor Sic1 of yeast (Borg *et al.* 2007). Above a threshold level of phosphorylation, the dissociation and subsequent elimination of the inhibitor may simply be driven by cumulative electrostatic interactions (Borg *et al.* 2007). Another example of control by multiple phosphorylation is in the remodeling of chromatin by FACT, where it is thought that phosphorylation of an acidic disordered region on one of the domains provides sufficient electrostatic repulsion to enable dissociation of the protein from DNA (Tsunaka *et al.* 2009).

Although it has been suggested on the basis of molecular dynamics calculations that phosphorylation may cause an increase in the propensity for helix formation in free IDPs (Solt *et al.* 2006), there has been little support for this notion from most experimental studies. No difference

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

in the conformational preference for helix was observed upon phosphorylation of a peptide representing the kinase-inducible domain (KID) of CREB (Radhakrishnan *et al.* 1998), and a comprehensive thermodynamic analysis of the binding of the KIX domain of CBP to pKID on the one hand, and unphosphorylated KID and c-Myb on the other hand, showed that the difference between the binding was primarily in the large negative enthalpy generated in the complex with pKID (Parker *et al.* 1999; Zor *et al.* 2002). Phosphorylation of the N-terminal disordered region of human c-Src kinase also showed no change in the conformational ensemble of the free protein (Perez *et al.* 2009) and it was suggested that the primary role for phosphorylation of this region is to create a global electrostatic perturbation in the kinase that prompts its dissociation from the cell membrane.

Local interaction with membrane-spanning segments was also invoked to explain the effects of phosphorylation in the CFTR (Kanelis *et al.* 2010). Phosphorylation of a disordered regulatory region attached to the first nucleotide-binding domain (NBD1) changes the conformational propensities in this region (Baker *et al.* 2007) and modulates the transient interactions of this region with the NBD1 core, likely mediating gating for the chloride channel across the membrane (Kanelis *et al.* 2010). Deletion of Phe508 of the NBD1, the most common mutation that results in the disease cystic fibrosis, gives a protein where the interactions between the core and the regulatory region are not as effectively disrupted, providing a possible structural basis for the observed inefficiency of chloride transport when the ΔPhe508 mutation is present (Kanelis *et al.* 2010). Other groups have suggested that the disease primarily arises due to incomplete folding and degradation of the ΔPhe508 mutant protein in the endoplasmic reticulum (see Wang *et al.* 2008 and references therein). Nevertheless, a recent structural comparison of the wild-type and mutant proteins shows distinct changes in local structure and dynamics of the folded proteins that could also be invoked to explain the effects of the mutation in the disease state (Lewis *et al.* 2010).

Conformational changes upon phosphorylation have been documented in some systems. For example, a conformational change mediated by phosphorylation was observed using time-resolved fluorescence resonance energy transfer (FRET) of the phosphorylation domain of smooth muscle myosin (Kast *et al.* 2010). Specific conformational changes were also invoked to explain NMR observations showing that pseudo-phosphorylation (replacement of serine phosphorylation sites by glutamate) of the Alzheimer's disease-associated protein tau results in attenuation of binding to microtubules, mimicking the hyperphosphorylation effects observed in the brains of patients with Alzheimer's disease (Fischer *et al.* 2009). Phosphorylation of specific sites on tau by glycogen synthase kinase (GSK)3β causes structural changes that lead to interactions with apolipoprotein E, an interaction that does not occur in unphosphorylated tau or when tau has been phosphorylated by the cyclic-AMP dependent protein kinase A (PKA). This discrimination has important implications for the occurrence and onset of disease (Leroy *et al.* 2010). Pseudo-phosphorylation of mutant proteins was also used to show a conformational change in the tumor suppressor and cyclin-dependent kinase inhibitor $p19^{INK4d}$, which enabled the ubiquitination and degradation of the p19 protein (Löw *et al.* 2009).

9.3 Ubiquitination and neddylation

As well as its well-characterized role in targeting proteins for degradation, the reversible covalent addition of ubiquitin (and similar protein tags such as the small ubiquitin-like modifier (SUMO) and the cullin ring ligase regulator NEDD8) is also implicated in many other functions in the cell. Known ubiquitination sites frequently map to intrinsically disordered segments of target

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

molecules (Edwards *et al.* 2009; Radivojac *et al.* 2010) and this observation was used to formulate a sequence predictor program for ubiquitination sites (Radivojac *et al.* 2010). Structural plasticity and conformational dynamics were also observed for ubiquitin-interacting motifs, which serve to recognize ubiquitin-labeled target molecules (Sgourakis *et al.* 2010).

An extensive series of molecular dynamics simulations on the cullin-RING-E3 ligase system suggests a model where a disordered linker domain mediates correct interaction and positioning of the various domain components for efficient transfer of ubiquitin to substrates (Liu and Nussinov, 2009, 2010a). Further calculations also provided a rationale for the role of the NEDD8 attachment in priming the system for ubiquitin ligation (Liu & Nussinov, 2010b). These studies suggest interesting hypotheses for exploration with carefully designed experiments.

9.4 Ubiquitin-independent protein degradation

While the occurrence of disordered segments appears to be important in the ubiquitin-dependent proteasomal degradation of proteins, it has recently been observed that proteins that are intrinsically disordered or even mildly destabilized can be rapidly degraded by ubiquitin-independent mechanism (Jariel-Encontre *et al.* 2008). The ubiquitin pathway uses the 26S proteasome (with body and cap segments), and the ubiquitin-independent process utilizes the 20S proteasome (body segment only). Examples of disordered and unstable proteins that are degraded in ubiquitin-independent manner include p53 (Tsvetkov *et al.* 2009a) and IκBα (Mathes *et al.* 2010).

9.5 IDPs and metal ion binding

The role of metal ions in stabilizing the folded states of small proteins is well-established, for example, zinc finger proteins have an absolute requirement for the presence of zinc ions, otherwise they are not folded or functional (Lee *et al.* 1989; Parraga *et al.* 1988). Reversible binding of metal ions, where both the metal-free disordered form and the metal-bound ordered form are functional, occur for at least one zinc-containing protein, the redox-dependent chaperone Hsp33 (Graf *et al.* 2004; Won *et al.* 2004), but is much more widely observed among calcium-binding proteins. Examples include the RTX motif of the Type 1 secretion system of Gram-negative bacteria (Chenal *et al.* 2009; Sotomayor Perez *et al.* 2010) and the structural coupling between N- and C-terminal lobes in calmodulin (Chen *et al.* 2008). Since calcium signaling is such an important process in many metabolic systems, it is likely that this kind of reversible order–disorder equilibrium will be found to be quite common in the future.

10. Partly folded and molten globule-like states

Expanding the proteome to include intrinsically disordered states is only part of the story. Biophysical data on disordered proteins may be difficult to interpret, but are generally available from a variety of sources. Data on partly folded states, however, are much harder to obtain, and remain extremely difficult to interpret. Nevertheless, a number of groups have recently attempted to define partly folded or molten globular states using a number of different biophysical techniques.

It was clear in early folding studies that in addition to the 'unfolded' and 'folded' states of a protein, other intermediate states could exist along the folding pathway. The 'molten globule' state, originally proposed as a specific intermediate with a native-like fold but without native

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

tertiary structure (Ohgushi & Wada, 1983; Ptitsyn, 1973) provides a model for the intermediate state, although it is now recognized that such intermediates actually consist of a number of conformational states in equilibrium. Indeed, molten globule states are particularly difficult to study by NMR, as the exchange between the conformational states is frequently on an intermediate time scale compared to the chemical shift time scale, resulting in extensive broadening and disappearance of resonances. Methods of addressing these difficulties have recently been introduced (Hsu *et al.* 2009; Li & Palmer, 2010) but the structural characterization of these systems at a molecular level remains extremely challenging. Since molten globule-like domains are so difficult to characterize structurally by standard methods such as X-ray crystallography and NMR, other techniques such as fluorescence and CD spectroscopy, small-angle X-ray scattering and H/D exchange have been employed to give structural insights. In particular, the fluorescence of the dye 1-anilino-8-naphthalenesulfonic acid (ANS) changes in the presence of molten globule states, and has been extensively used as a signal of the presence of such a state (Cattoni *et al.* 2009).

An addition was made to the 'protein trinity' concept (folded, unfolded and molten globule) (Dunker & Obradovic, 2001; Romero *et al.* 2004) to include a fourth member, the pre-molten globule (Uversky and Ptitsyn, 1994, 1996), giving rise to the 'protein quartet' terminology (Uversky, 2002). A pre-molten globule is defined as a subset of unfolded proteins where there is a significant population of residual secondary structure in the free state (Uversky, 2002). Other terminology has been coined for complexes without fixed structures, such as 'cloud' or 'fuzzy' complexes (see Section 4) (Uversky, 2010).

10.1 AR proteins

AR proteins contain multiples of a short structured motif consisting of 33 amino acids arranged in a β-turn/loop-helix-loop-helix fold (Fig. 9). Multiples of between 3 and 24 repeats have been observed in over 3500 proteins, and the AR appears to function primarily as a protein recognition motif. The amino acid sequences of the repeats are homologous, but not identical, with the strongest homology in a consensus PLHLA sequence at the N-terminal end of the first short helix. AR proteins appear to have a wide repertoire of possible structured and partly structured forms. For example, the notch AR domain appears to be stably folded, although other parts of the notch protein are unstructured (Bertagna *et al.* 2008). As previously mentioned, the $p19^{INK4d}$ tumor suppressor undergoes a phosphorylation induced conformational change to an open position that may be coupled to ubiquitination (Barrick, 2009; Löw *et al.* 2009). By contrast, the AR domain of IκBα shows significant variation in the stability between its six ARs, and the presence of a partly folded region near the C-terminus of the free protein (Croy *et al.* 2004; Sue *et al.* 2008) appears to be required for some of its functions. In particular, mutation of the sequence of AR5 and AR6 to conform to the consensus sequence causes the partly folded region to become more stably folded (Ferreiro *et al.* 2007) but renders the IκBα less capable of stripping the NF-κB from DNA (Bergqvist *et al.* 2009).

10.2 Characterization of binding-induced molten globule-like state

It is perhaps intuitively satisfying (and analogous to the coupled folding and binding of completely unfolded IDPs) that a less-folded state present in a free protein should be more stably folded in complex with its partner, as in the case of IκBα. A different case has recently been characterized,

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

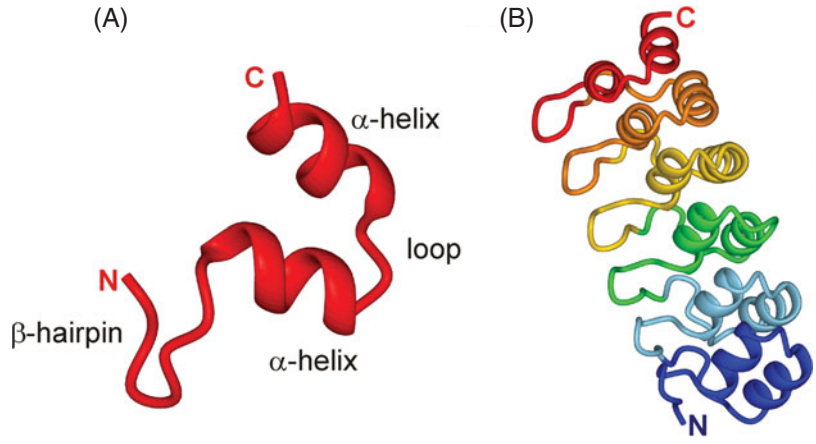


Fig. 9. (A) Single AR module from the Notch AR domain showing secondary structure elements. (B) Backbone structure of the Notch AR domain (PDB entry 1YYH) (Ehebauer *et al.* 2005), showing close connections between repeat subunits.

where binding of the partner appears to induce the formation of a molten globule-like state in a protein that is well folded in isolation. The p53 DNA binding domain interacts with the chaperone Hsp90. NMR, fluorescence and H/D exchange experiments all point to the formation of a loosened and flexible state of the p53 DBD in the presence of Hsp90 (Park *et al.* 2011). This observation has profound implications for the role of Hsp90 in many of its known cellular functions: if Hsp90 induces a similar loosened state in other client proteins, it would help explain the part that Hsp90 plays in the stabilization of the ligand-binding domains of nuclear hormone receptors in the absence of their hormone ligands (Pratt & Toft, 1997).

10.3 Coiled coil and leucine zipper proteins

Helical bundle structures would appear at first sight to be poor candidates for functions requiring flexibility. However, a number of examples have recently been reported where coiled-coil and leucine zipper structures show structural flexibility that enables them to bind many different partners (Miller, 2009). Examples include the coupled folding and binding reaction of c-Myc upon heterodimerization with its partner Max (Follis *et al.* 2009), and bacterial type III secretion systems, which show a variety of forms ranging from localized structural disorder to molten globule states (Gazi *et al.* 2009; Hamada *et al.* 2005).

11. Strategies for study of disordered proteins

Since the earliest recognition of the role of IDPs in biological processes, they have been the subject of numerous biochemical and biophysical analyses. Clearly some methods are completely unsuitable for use with IDPs, particularly those like crystallography that rely upon the formation of ordered uniform ensembles of molecules that all have the same structure. In this section, I will touch briefly on the techniques that have been used, with a brief analysis of the strengths and weaknesses of each approach. A recent issue of *Current Opinion in Structural Biology* was devoted to biophysical characterization of IDPs (Eliezer, 2009).

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

11.1 Bacterial expression of IDPs

In our experience, the overexpression of intrinsically disordered domains in bacteria can be problematic, with yields of expressed protein being reduced by degradation during expression, by aggregation within the bacterial cells, or because they are toxic to the cells. These disadvantages can be mitigated somewhat by co-expression of disordered domains with their binding partners, which yields a stable complex that can subsequently be purified and manipulated *in vitro* to produce, for example, NMR samples where the components of the complex are differentially labeled (Demarest *et al.* 2002). Alternatively, high yields of differentially labeled complex can be obtained by mixing bacterial lysates containing the overexpressed components, followed by purification of the complex (Sue *et al.* 2008). Interestingly, for the expression of *E. coli* proteins themselves, the presence of disorder in a protein sequence was positively correlated with expression, measured by both RNA and protein levels (Paliy *et al.* 2008), but the presence of large numbers of ubiquitination sites in disordered regions appears to regulate the expression levels of many highly expressed IDPs (Edwards *et al.* 2009).

11.2 Biochemistry of IDPs

As mentioned in Section 7, the proportion of IDPs present in a cell extract can be increased by heat treatment (Galea *et al.* 2009). Treatment with certain acids can also be used to enrich for IDPs (Cortese *et al.* 2005): structured proteins are normally precipitated by addition of trichloroacetic or perchloric acids, while IDPs remain in solution. The observation that IDPs are not generally precipitated by such treatments forms the basis of a 2D electrophoresis method to identify IDPs (Csizmok *et al.* 2006).

A number of studies aimed at a general understanding of the biochemical limitations on intrinsic disorder have characterized the behavior of several different IDPs under a variety of solution conditions. In general, it is found that the state of an IDP can be affected by changes in many different biochemical features of the solution (Uversky, 2009b). The populations of areas of an IDP with conformational preferences for local structured regions (helical turns, β-turns or hydrophobic clusters, for example), as well as the general level of compaction of the protein (Marsh & Forman-Kay, 2010) are highly dependent on solution conditions. In particular, IDPs appear to be sensitive to the presence of membranes (Bartels *et al.* 2010; Bonsor *et al.* 2008; de la Cruz *et al.* 2010; Reingewertz *et al.* 2009; Uversky & Eliezer, 2009), osmolytes (Chang & Oas, 2010) and macromolecular crowding (Bernado *et al.* 2004).

11.3 NMR techniques used for IDPs

Many of the solution NMR methods used to characterize folded proteins and complexes can be used for IDPs. Some experiments, indeed, may work better with disordered proteins than with ordered globular proteins, because the segments of a disordered protein may tumble independently, giving correlation times that approximate those of smaller proteins, with correspondingly narrower resonance lines. General techniques that are particularly appropriate for IDPs (and the unfolded forms of globular proteins) were reviewed several years ago (Dyson & Wright, 2004); since this time, a number of new NMR techniques have been developed specifically for the detection and characterization of IDPs. Two major problems occur in the NMR spectra of IDPs: the first problem is resonance overlap that may occur in a fully disordered polypeptide

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

chain, due to the similarity in the chemical environments of nuclei in amino acids that are all equally exposed to solvent. This problem can generally be overcome by the use of a variety of multi-dimensional triple resonance spectra, transverse relaxation optimization and deuteration (Motáckova *et al.* 2010), and by utilizing hardware improvements such as cryogenic probes (Wang *et al.* 2005). Less tractable is the second problem, resonance broadening that frequently occurs in IDPs due to exchange processes on an intermediate time scale compared to the NMR chemical shift time scale. This problem can be addressed to some extent by the design of pulse sequences and detection methods (Hsu *et al.* 2009; Li & Palmer, 2010). An extensive examination of the conformational ensemble of the Alzheimer's disease tau protein provides an illustration of the power of a concerted NMR approach to the characterization of an important IDP (Mukrasch *et al.* 2009).

Paramagnetic relaxation enhancement (PRE) detected by NMR is a venerable technique that has become one of the premier sources of structural and dynamic information on disordered proteins. The PRE method takes advantage of the r^{-6} distance dependence of the magnetic dipolar interaction between an unpaired electron and the NMR-active nucleus of interest. Because the magnetic moment of the unpaired electron is much larger than that of the nuclear spin, PRE effects are large, and can be observed at much longer distances than, for example, the nuclear Overhauser enhancement (NOE), the analogous dipolar interaction between nuclear spins. PRE is also ideally suited to the estimation of distance ranges in conformational ensembles such as those occur that with IDPs. The application of this method to the characterization of low-populated states (such as might occur, for example, in an IDP with a local conformational preference) has been recently reviewed (Clore & Iwahara, 2009), and the specific case of the description of a conformational ensemble has been reported (Iwahara & Clore, 2010). Examples of the use of PRE in specific systems include the drkN SH3 domain (Xue *et al.* 2009c), the p53 TAD (Lowry *et al.* 2008b) and the Parkinson's disease protein α-synuclein (Wu & Baum, 2010).

Another powerful tool that has recently been employed for the NMR characterization of IDPs is the measurement of residual dipolar couplings (RDC), which are measured under conditions of weak alignment (Dames *et al.* 2006). Examples include α-synuclein (Bertoncini *et al.* 2005; Sung & Eliezer, 2007) and β-synuclein (Bertoncini *et al.* 2007). Indeed, the combination of RDC and PRE experiments is frequently employed to give a more complete picture of the conformational ensemble of an unfolded or partly folded protein (Meier *et al.* 2008; Rospigliosi *et al.* 2009; Salmon *et al.* 2010).

NMR experiments are particularly suited to the estimation of polypeptide chain dynamics, and the dynamics of IDPs have been a subject of much study. The experimental techniques used to study dynamics, particularly in disordered systems, were recently reviewed (Mittermaier & Kay, 2009; Niu *et al.* 2011). One of the problems with classical methods such as the model-free approach (Lipari & Szabo, 1982) when applied to disordered proteins is that a number of assumptions that are valid for globular proteins are no longer appropriate. For example, the assumption of a uniform overall correlation time for all molecules is clearly inapplicable to a conformational ensemble. A particularly interesting case arises for proteins that consist of a structured domain and a long unstructured tail. Anomalies in the picosecond–nanosecond dynamics of such a system were early noted for the prion protein (Donne *et al.* 1997; Viles *et al.* 2001). A recent comprehensive analysis was able to resolve the anomalies and suggest a methodology for dealing with the 'tail wagging the dog' that occurs for molecules with long disordered tails (Bae *et al.* 2009).

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

11.4 Fluorescence methods

Transient contacts between non-adjacent parts of polypeptide chains are readily detected by FRET. Newer single-molecule techniques have been particularly popular in defining the populations of more- or less-compact conformers in disordered protein ensembles (reviewed in Lu, 2005). Examples of systems where these techniques have given valuable new insights include p53 (Huang *et al.* 2009), α-synuclein (Ferreon *et al.* 2009a) and three IDPs with contrasting properties (Müller-Späth *et al.* 2010).

11.5 Other experimental techniques

A variety of experimental techniques have been applied to the structural and dynamic characterization of IDPs. Some of these methods are really only applicable to small numbers of systems, while others are experimentally difficult to access or the results difficult to interpret for conformational ensembles. Among these specialized methods are: electron paramagnetic resonance (EPR) spectroscopy of spin-labeled proteins, which has been used to confirm the binding of the N-terminal tail of the measles virus to the rigid C-terminal domain (Belle *et al.* 2008; Kavalenka *et al.* 2010; Morin *et al.* 2006), infrared spectroscopy of specifically deuterated bonds (Cremeens *et al.* 2009), vibrational spectroscopy (Bischak *et al.* 2010), ^{19}F-NMR spectroscopy (Winkler *et al.* 2006), mass spectrometry (Frimpong *et al.* 2010), NMR relaxation studies of bound water (Bokor *et al.* 2005), high speed atomic force microscopy (Ando *et al.* 2001, 2007), small angle X-ray scattering (Bernado *et al.* 2007, 2008, 2009, 2010; Blobel *et al.* 2009), time-of-flight neutron scattering (Gaspar *et al.* 2008) and Raman optical activity (Zhu *et al.* 2008).

11.6 Computational studies of IDPs

Apart from the bioinformatic studies that use sequence data from genomic databases to predict the occurrence of IDPs and IDRs, there has also been a great deal of thought given to the theoretical underpinnings of the IDP phenomenon and the simulation of conformational ensembles by computational methods. It is impossible to do full justice to this rapidly burgeoning field; contributions to the description of energy landscapes have been made from a number of labs (Dosztanyi *et al.* 2005; Lobley *et al.* 2007; Lowry *et al.* 2008a; Mao *et al.* 2010; Tran *et al.* 2005, 2008; Tyka *et al.* 2010) and computational modeling of disordered states has been reported using various methods (Fisher *et al.* 2010; He *et al.* 2011; Staneva & Wallin, 2009; Zwier & Chong, 2010). Computational approaches to docking of small molecules and IDPs have also received significant attention (Andrusier *et al.* 2008; Follis *et al.* 2008; Hammoudeh *et al.* 2009; Mobley & Dill, 2009). Although systems in conformational equilibrium in solution may appear simple, it is clear from the studies quoted that challenges remain for attempts to define such ensembles in their entirety *in silico*.

12. Conclusions

It is clear from the foregoing that interest in IDPs spans a multitude of fields, and generates a multitude of opinions. It now seems that IDPs are incontrovertibly part of the landscape of protein structure and function. Motion and flexibility in proteins are ubiquitous, even in systems that we would normally think of as 'structured'. For example, defined motions are frequently

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

part of the mechanisms of highly specific enzymes, and disallowing these motions causes severe disruptions to the enzymatic activity. Disordered proteins may be seen as an extremum on the continuum from fully folded, entirely rigid proteins through flexibility and conformational heterogeneity. These observations clearly indicate that the proteome is vastly more versatile than we had imagined even only a few years ago, and we anticipate with eagerness further surprises in this field.

13. Acknowledgments

I thank Peter Wright and members of the Wright and Dyson groups, past and present, for helpful discussions, and Harold Scheraga for a critical reading of the manuscript. This work was supported by grants GM57374 and GM71862 from the National Institutes of Health.

14. References

Adda, C. G., Murphy, V. J., Sunde, M., Waddington, L. J., Schloegel, J., Talbo, G. H., Vingas, K., Kienzle, V., Masciantonio, R., Howlett, G. J., Hodder, A. N., Foley, M. & Anders, R. F. (2009). *Plasmodium falciparum* merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. *Molecular and Biochemical Parasitology* **166**, 159–171.

Alba, M. M., Tompa, P. & Veitia, R. A. (2007). Amino acid repeats and the structure and evolution of proteins. *Genome Dynamics* **3**, 119–30, 119–130.

Albert, R., Jeong, H. & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature* **406**, 378–382.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294.

Ando, T., Kodera, N., Takai, E., Maruyama, D., Saito, K. & Toda, A. (2001). A high-speed atomic force microscope for studying biological macromolecules. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12468–12472.

Ando, T., Uchihashi, T., Kodera, N., Yamamoto, D., Taniguchi, M., Miyagi, A. & Yamashita, H. (2007). High-speed atomic force microscopy for observing dynamic biomolecular processes. *Journal of Molecular Recognition* **20**, 448–458.

Andrusier, N., Mashiach, E., Nussinov, R. & Wolfson, H. J. (2008). Principles of flexible protein–protein docking. *Proteins* **73**, 271–289.

Arumugam, S., Gao, G., Patton, B. L., Semenchenko, V., Brew, K. & Van Doren, S. R. (2003). Increased backbone mobility in β-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. *Journal of Molecular Biology* **327**, 719–734.

Bae, S. H., Dyson, H. J. & Wright, P. E. (2009). Prediction of the rotational tumbling time for proteins with disordered segments. *Journal of the American Chemical Society* **131**, 6814–6821.

Bai, F., Branch, R. W., Nicolau, Jr., D. V., Pilizota, T., Steel, B. C., Maini, P. K. & Berry, R. M. (2010). Conformational spread as a mechanism for cooperativity in the bacterial flagellar switch. *Science* **327**, 685–689.

Baker, J. M., Hudson, R. P., Kanelis, V., Choy, W. Y., Thibodeau, P. H., Thomas, P. J. & Forman-Kay, J. D. (2007). CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nature Structural and Molecular Biology* **14**, 738–745.

Barrick, D. (2009). Biological regulation via ankyrin repeat folding. *Acs Chemical Biology* **4**, 19–22.

Bartels, T., Ahlstrom, L. S., Leftin, A., Kamp, F., Haass, C., Brown, M. F. & Beyer, K. (2010). The N-terminus of the intrinsically disordered protein α-synuclein triggers membrane binding and helix folding. *Biophysical Journal* **99**, 2116–2124.

Belle, V., Rouger, S., Costanzo, S., Liquiere, E., Strancar, J., Guigliarelli, B., Fournel, A. & Longhi, S. (2008). Mapping α-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Proteins* **73**, 973–988.

Bergqvist, S., Alverdi, V., Mengel, B., Hoffmann, A., Ghosh, G. & Komives, E. A. (2009). Kinetic enhancement of NF-κB·DNA dissociation by IκBα. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19328–19333.

Bergqvist, S., Croy, C. H., Kjaergaard, M., Huxford, T., Ghosh, G. & Komives, E. A. (2006). Thermodynamics reveal that helix four in the NLS of NF-κB p65 anchors IκBα, forming a very stable complex. *Journal of Molecular Biology* **360**, 421–434.

Bergqvist, S., Ghosh, G. & Komives, E. A. (2008). The IκBα/NF-κB complex has two hot spots, one at either end of the interface. *Protein Science* **17**, 2051–2058.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Bernado, P., Blackledge, M. & Sancho, J. (2006). Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophysical Journal* **91**, 4536–4543.

Bernado, P., Garcia, D. L. T. & Pons, M. (2004). Macromolecular crowding in biological systems: hydrodynamics and NMR methods. *Journal of Molecular Recognition* **17**, 397–407.

Bernado, P., Modig, K., Grela, P., Svergun, D. I., Tchorzewski, M., Pons, M. & Akke, M. (2010). Structure and dynamics of ribosomal protein L12: an ensemble model based on SAXS and NMR relaxation. *Biophysical Journal* **98**, 2374–2382.

Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* **129**, 5656–5664.

Bernado, P., Perez, Y., Blobel, J., Fernandez-Recio, J., Svergun, D. I. & Pons, M. (2009). Structural characterization of unphosphorylated STAT5a oligomerization equilibrium in solution by small-angle X-ray scattering. *Protein Science* **18**, 716–726.

Bernado, P., Perez, Y., Svergun, D. I. & Pons, M. (2008). Structural characterization of the active and inactive states of Src kinase in solution by small-angle X-ray scattering. *Journal of Molecular Biology* **376**, 492–505.

Bertagna, A., Toptygin, D., Brand, L. & Barrick, D. (2008). The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tuned disorder. *Biochemical Society Transactions* **36**, 157–166.

Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M. & Zweckstetter, M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured α-synuclein. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 1430–1435.

Bertoncini, C. W., Rasia, R. M., Lamberto, G. R., Binolfi, A., Zweckstetter, M., Griesinger, C. & Fernandez, C. O. (2007). Structural characterization of the intrinsically unfolded protein β-synuclein, a natural negative regulator of α-synuclein aggregation. *Journal of Molecular Biology* **372**, 708–722.

Bischak, C. G., Longhi, S., Snead, D. M., Costanzo, S., Terrer, E. & Londergan, C. H. (2010). Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines. *Biophysical Journal* **99**, 1676–1683.

Blobel, J., Bernado, P., Svergun, D. I., Tauler, R. & Pons, M. (2009). Low-resolution structures of transient protein–protein complexes using small-angle X-ray scattering. *Journal of the American Chemical Society* **131**, 4378–4386.

Bochkareva, E., Kaustov, L., Ayed, A., Yi, G. S., Lu, Y., Pineda-Lucena, A., Liao, J. C. C., Okorokov, A. L., Milner, J., Arrowsmith, C. H. & Bochkarev, A. (2005). Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15412–15417.

Bokor, M., Csizmok, V., Kovacs, D., Banki, P., Friedrich, P., Tompa, P. & Tompa, K. (2005). NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. *Biophysical Journal* **88**, 2030–2037.

Bonsor, D. A., Meenan, N. A. & Kleanthous, C. (2008). Colicins exploit native disorder to gain cell entry: a hitchhiker's guide to translocation. *Biochemical Society Transactions* **36**, 1409–1413.

Borg, M., Mittag, T., Pawson, T., Tyers, M., Forman-Kay, J. D. & Chan, H. S. (2007). Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9650–9655.

Bork, P. & Sudol, M. (1994). The WW domain: a signalling site in dystrophin? *Trends in Biochemical Sciences* **19**, 531–533.

Bothner, B., Aubin, Y. & Kriwacki, R. W. (2003). Peptides derived from two dynamically disordered proteins self-assemble into amyloid-like fibrils. *Journal of the American Chemical Society* **125**, 3200–3201.

Bothner, B., Lewis, W. S., Digiammarino, E. L., Weber, J. D., Bothner, S. J. & Kriwacki, R. W. (2001). Defining the molecular basis of Arf and Hdm2 interactions. *Journal of Molecular Biology* **314**, 263–277.

Bowman, P., Galea, C. A., Lacy, E. & Kriwacki, R. W. (2006). Thermodynamic characterization of interactions between $p27^{Kip1}$ and activated and non-activated Cdk2: intrinsically unstructured proteins as thermodynamic tethers. *Biochimica et Biophysica Acta* **1764**, 182–189.

Brown, C. J., Takayama, S., Campen, A. M., Vise, P., Marshall, T. W., Oldfield, C. J., Williams, C. J. & Dunker, A. K. (2002). Evolutionary rate heterogeneity in proteins with long disordered regions. *Journal of Molecular Evolution* **55**, 104–110.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21**, 167–195.

Burra, P. V., Kalmar, L. & Tompa, P. (2010). Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS ONE* **5**, e12069.

Campen, A., Williams, R. M., Brown, C. J., Meng, J., Uversky, V. N. & Dunker, A. K. (2008). TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein and Peptide Letters* **15**, 956–963.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Carbonell, P., Nussinov, R. & Del Sol, A. (2009). Energetic determinants of protein binding specificity: Insights into protein interaction networks. *Proteomics* **9**, 1744–1753.

Cattoni, D. I., Kaufman, S. B. & Gonzalez Flecha, F. L. (2009). Kinetics and thermodynamics of the interaction of 1-anilino-naphthalene-8-sulfonate with proteins. *Biochimica et Biophysica Acta* **1794**, 1700–1708.

Cervantes, C. F., Bergqvist, S., Kjaergaard, M., Kroon, G., Sue, S. C., Dyson, H. J. & Komives, E. A. (2010). The RelA nuclear localization signal folds upon binding to IκBα. *Journal of Molecular Biology*.

Chakrabortee, S., Meersman, F., Kaminski Schierle, G. S., Bertoncini, C. W., Mcgee, B., Kaminski, C. F. & Tunnacliffe, A. (2010). Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16084–16089.

Chang, Y. C. & Oas, T. G. (2010). Osmolyte-induced folding of an intrinsically disordered protein: folding mechanism in the absence of ligand. *Biochemistry* **49**, 5086–5096.

Chatr-Aryamontri, A., Ceol, A., Licata, L. & Cesareni, G. (2008). Protein interactions: integration leads to belief. *Trends in Biochemical Sciences* **33**, 241–242.

Chen, B., Lowry, D. F., Mayer, M. U. & Squier, T. C. (2008). Helix A stabilization precedes amino-terminal lobe activation upon calcium binding to calmodulin. *Biochemistry* **47**, 9220–9226.

Chen, J. (2009). Intrinsically disordered p53 extreme C-terminus binds to S100B($\beta\beta$) through "fly-casting". *Journal of the American Chemical Society* **131**, 2088–2089.

Chen, J. W., Romero, P., Uversky, V. N. & Dunker, A. K. (2006a). Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *Journal of Proteome Research* **5**, 879–887.

Chen, J. W., Romero, P., Uversky, V. N. & Dunker, A. K. (2006b). Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *Journal of Proteome Research* **5**, 888–898.

Chenal, A., Guijarro, J. I., Raynal, B., Delepierre, M. & Ladant, D. (2009). RTX calcium binding motifs are intrinsically disordered in the absence of calcium. *Journal of Biological Chemistry* **284**, 1781–1789.

Cheng, Y., Legall, T., Oldfield, C. J., Mueller, J. P., Van, Y. Y., Romero, P., Cortese, M. S., Uversky, V. N. & Dunker, A. K. (2006). Rational drug design via intrinsically disordered protein. *Trends in Biotechnology* **24**, 435–442.

Clore, G. M. & Iwahara, J. (2009). Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical Reviews* **109**, 4108–4139.

Cortese, M. S., Baird, J. P., Uversky, V. N. & Dunker, A. K. (2005). Uncovering the unfoldome: enriching cell extracts for unstructured proteins by acid treatment. *Journal of Proteome Research* **4**, 1610–1618.

Cortese, M. S., Uversky, V. N. & Dunker, A. K. (2008). Intrinsic disorder in scaffold proteins: getting more from less. *Progress in Biophysics and Molecular Biology* **98**, 85–106.

Cremeens, M. E., Zimmermann, J., Yu, W., Dawson, P. E. & Romesberg, F. E. (2009). Direct observation of structural heterogeneity in a β-sheet. *Journal of the American Chemical Society* **131**, 5726–+.

Croy, C. H., Bergqvist, S., Huxford, T., Ghosh, G. & Komives, E. A. (2004). Biophysical characterization of the free IκBα ankyrin repeat domain in solution. *Protein Science* **13**, 1767–1777.

Csermely, P., Palotai, R. & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**, 539–546.

Csizmok, V., Dosztanyi, Z., Simon, I. & Tompa, P. (2007). Towards proteomic approaches for the identification of structural disorder. *Current Protein and Peptide Science* **8**, 173–179.

Csizmok, V., Szollosi, E., Friedrich, P. & Tompa, P. (2006). A novel two-dimensional electrophoresis technique for the identification of intrinsically unstructured proteins. *Molecular and Cellular Proteomics* **5**, 265–273.

Dames, S. A., Aregger, R., Vajpai, N., Bernado, P., Blackledge, M. & Grzesiek, S. (2006). Residual dipolar couplings in short peptides reveal systematic conformational preferences of individual amino acids. *Journal of the American Chemical Society* **128**, 13508–13514.

Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J. & Wright, P. E. (2002). Structural basis for Hif-1α/CBP recognition in the cellular hypoxic response. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5271–5276.

Das, R. (2009). Allosteric activation of E2-RING finger-mediated ubiquitylation by a structurally defined specific E2-binding region of gp78. *Molecular Cell* **34**, 674–685.

Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T. & Dahlquist, F. W. (1997). The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, σ^{28}. *Nature Structural Biology* **4**, 285–291.

Daughdrill, G. W., Hanely, L. J. & Dahlquist, F. W. (1998). The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations. *Biochemistry* **37**, 1076–1082.

Davey, N. E., Travé, G. & Gibson, T. J. (2011). How viruses hijack cell regulation. *Trends in Biochemical Sciences* **36**, 159–169.

De Guzman, R. N., Goto, N. K., Dyson, H. J. & Wright, P. E. (2006). Structural basis for cooperative

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

transcription factor binding to the CBP coactivator. *Journal of Molecular Biology* **355**, 1005–1013.

De Guzman, R. N., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2004a). Interaction of the TAZ1 domain of the CREB-binding protein with the activation domain of CITED2: regulation by competition between intrinsically unstructured ligands for non-identical binding sites. *Journal of Biological Chemistry* **279**, 3042–3049.

De Guzman, R. N., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2004b). Structure and function of the CBP/p300 TAZ domains. In *Zinc Finger Proteins: from Atomic Contact to Cellular Function* (Eds. S. Iuchi & N. Kuldell), pp. 116–122. New York: Kluwer Academic/Plenum Publishers.

De Guzman, R. N., Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2005). CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry* **44**, 490–497.

De La Cruz, L., Bajaj, R., Becker, S. & Zweckstetter, M. (2010). The intermembrane space domain of Tim23 is intrinsically disordered with a distinct binding region for presequences. *Protein Science* **19**, 2045–2054.

Demarest, S. J., Martinez-Yamout, M., Chung, J., Chen, H., Xu, W., Dyson, H. J., Evans, R. M. & Wright, P. E. (2002). Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **415**, 549–553.

Deryusheva, E. I., Galzitskaya, O. V. & Serdyuk, I. N. (2008). Prediction of short loops in intrinsically disordered proteins. *Molecular Biology* **42**, 949–959.

Di Lello, P., Jenkins, L. M. M., Jones, T. N., Nguyen, B. D., Hara, T., Yamaguchi, H., Dikeakos, J. D., Appella, E., Legault, P. & Omichinski, J. G. (2006). Structure of the Tfb1/p53 complex: insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. *Molecular Cell* **22**, 731–740.

Donne, D. G., Viles, J. H., Groth, D., Mehlhorn, I., James, T. L., Cohen, F. E., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (1997). Structure of the recombinant full-length hamster prion protein PrP(29–231): the N terminus is highly flexible. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13452–13457.

Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I. & Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *Journal of Proteome Research* **5**, 2985–2995.

Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology* **347**, 827–839.

Drobnak, I., Korencic, A., Loris, R., Marianovsky, I., Glaser, G., Jamnik, A., Vesnaver, G. & Lah, J. (2009). Energetics of MazG unfolding in correlation with its structural features. *Journal of Molecular Biology* **392**, 63–74.

Duggan, B. M., Dyson, H. J. & Wright, P. E. (1999). Inherent flexibility in a potent inhibitor of blood coagulation, recombinant nematode anticoagulant protein c2. *European Journal of Biochemistry* **265**, 539–548.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582.

Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS Journal* **272**, 5129–5148.

Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C. & Villafranca, J. E. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pacific Symposium on Biocomputing* **3**, 473–484.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. & Obradovic, Z. (2001). Intrinsically disordered protein. *Journal of Molecular Graphics and Modeling* **19**, 26–59.

Dunker, A. K. & Obradovic, Z. (2001). The protein trinity–linking function and disorder. *Nature Biotechnology* **19**, 805–806.

Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Information Series Workshop on Genome Information* **11**, 161–171.

Dyson, H. J., Satterthwait, A. C., Lerner, R. A. & Wright, P. E. (1990). Conformational preferences of synthetic peptides derived from the immunodominant site of the circumsporozoite protein of *Plasmodium falciparum* by ^{1}H NMR. *Biochemistry* **29**, 7828–7837.

Dyson, H. J. & Wright, P. E. (2002). Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* **12**, 54–60.

Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chemical Reviews* **104**, 3607–3622.

Dyson, H. J. & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews in Molecular and Cellular Biology* **6**, 197–208.

Edwards, Y. J., Lobley, A. E., Pentony, M. M. & Jones, D. T. (2009). Insights into the regulation of intrinsically disordered proteins in the human proteome by analyzing sequence and gene expression data. *Genome Biology* **10**, R50.

Ehebauer, M. T., Chirgadze, D. Y., Hayward, P., Martinez, A. A. & Blundell, T. L. (2005). High-resolution crystal structure of the human Notch 1 ankyrin domain. *Biochemical Journal* **392**, 13–20.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Eliezer, D. (2009). Biophysical characterization of intrinsically disordered proteins. *Current Opinion in Structural Biology* **19**, 23–30.

Elkins, J. M., Hewitson, K. S., Mcneill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J. & Schofield, C. J. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *Journal of Biological Chemistry* **278**, 1802–1806.

Espinoza-Fonseca, L. M. (2009a). Reconciling binding mechanisms of intrinsically disordered proteins. *Biochemical and Biophysical Research Communications* **382**, 479–482.

Espinoza-Fonseca, L. M. (2009b). Thermodynamic aspects of coupled binding and folding of an intrinsically disordered protein: a computational alanine scanning study. *Biochemistry* **48**, 11332–11334.

Estrada, J., Bernado, P., Blackledge, M. & Sancho, J. (2009). ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinformatics* **10**, 104.

Feige, M. J., Groscurth, S., Marcinowski, M., Shimizu, Y., Kessler, H., Hendershot, L. M. & Buchner, J. (2009). An unfolded CH1 domain controls the assembly and secretion of IgG antibodies. *Molecular Cell* **34**, 569–579.

Feng, H., Jenkins, L. M. M., Durell, S. R., Hayashi, R., Mazur, S. J., Cherry, S., Tropea, J. E., Miller, M., Wlodawer, A., Appella, E. & Bai, Y. (2009). Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. *Structure* **17**, 202–210.

Ferreiro, D. U., Cervantes, C. F., Truhlar, S. M., Cho, S. S., Wolynes, P. G. & Komives, E. A. (2007). Stabilizing IκBα by "consensus" design. *Journal of Molecular Biology* **365**, 1201–1216.

Ferreon, A. C., Gambin, Y., Lemke, E. A. & Deniz, A. A. (2009a). Interplay of α-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5645–5650.

Ferreon, J. C., Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2009b). Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6591–6596.

Ferreon, J. C., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2009c). Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13260–13265.

Fischer, D., Mukrasch, M. D., Biernat, J., Bibow, S., Blackledge, M., Griesinger, C., Mandelkow, E. & Zweckstetter, M. (2009). Conformational changes specific for pseudophosphorylation at serine 262 selectively impair binding of tau to microtubules. *Biochemistry* **48**, 10047–10055.

Fisher, C. K., Huang, A. & Stultz, C. M. (2010). Modeling intrinsically disordered proteins with Bayesian statistics. *Journal of the American Chemical Society* **132**, 14919–14927.

Follis, A. V., Hammoudeh, D. I., Daab, A. T. & Metallo, S. J. (2009). Small-molecule perturbation of competing interactions between c-Myc and Max. *Bioorganic & Medicinal Chemistry Letters* **19**, 807–810.

Follis, A. V., Hammoudeh, D. I., Wang, H., Prochownik, E. V. & Metallo, S. J. (2008). Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules. *Chemistry and Biology* **15**, 1149–1155.

Fong, J. H., Shoemaker, B. A., Garbuzynskiy, S. O., Lobanov, M. Y., Galzitskaya, O. V. & Panchenko, A. R. (2009). Intrinsic disorder in protein interactions: insights from a comprehensive structural analysis. *PLoS Computational Biology* **5**, e1000316.

Forbes, J. G., Flaherty, D. B., Ma, K., Qadota, H., Benian, G. M. & Wang, K. (2010). Extensive and modular intrinsically disordered segments in *C elegans* TTN-1 and implications in filament binding, elasticity and oblique striation. *Journal of Molecular Biology* **398**, 672–689.

Frankel, A. D. & Kim, P. S. (1991). Modular structure of transcription factors: implications for gene regulation. *Cell* **65**, 717–719.

Freedman, S. J., Sun, Z. Y., Kung, A. L., France, D. S., Wagner, G. & Eck, M. J. (2003). Structural basis for negative regulation of hypoxia-inducible factor-1α by CITED2. *Nature Structural Biology* **10**, 504–512.

Freedman, S. J., Sun, Z. Y., Poy, F., Kung, A. L., Livingston, D. M., Wagner, G. & Eck, M. J. (2002). Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1α. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 5367–5372.

Frimpong, A. K., Abzalimov, R. R., Uversky, V. N. & Kaltashov, I. A. (2010). Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of α-synuclein. *Proteins* **78**, 714–722.

Fukuchi, S., Homma, K., Minezaki, Y., Gojobori, T. & Nishikawa, K. (2009). Development of an accurate classification system of proteins into structured and unstructured regions that uncovers novel structural domains: its application to human transcription factors. *BMC Structural Biology* **9**, 26.

Fuxreiter, M. & Tompa, P. (2009). Fuzzy interactome: the limitations of models in molecular biology. *Trends in Biochemical Sciences* **34**, 3.

Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C. & Asturias, F. J. (2008). Malleable

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

machines take shape in eukaryotic transcriptional regulation. *Nature Chemical Biology* **4**, 728–737.

Galea, C. A., High, A. A., Obenauer, J. C., Mishra, A., Park, C. G., Punta, M., Schlessinger, A., Ma, J., Rost, B., Slaughter, C. A. & Kriwacki, R. W. (2009). Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *Journal of Proteome Research* **8**, 211–226.

Galea, C. A., Pagala, V. R., Obenauer, J. C., Park, C. G., Slaughter, C. A. & Kriwacki, R. W. (2006). Proteomic studies of the intrinsically unstructured mammalian proteome. *Journal of Proteome Research* **5**, 2839–2848.

Galzitskaya, O. V., Garbuzynskiy, S. O. & Lobanov, M. Y. (2006). Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Computational Biology* **2**, e177.

Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De, G. H., Magnuson, R. D., Charlier, D., Van Nuland, N. A. & Loris, R. (2010). Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142**, 101–111.

Garner, E., Cannon, P., Romero, P., Obradovic, Z. & Dunker, A. K. (1998). Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. *Genome Information Series Workshop on Genome Information* **9**, 201–213.

Gaspar, A. M., Appavou, M. S., Busch, S., Unruh, T. & Doster, W. (2008). Dynamics of well-folded and natively disordered proteins in solution: a time-of-flight neutron scattering study. *European Biophysics Journal with Biophysics Letters* **37**, 573–582.

Gazi, A. D., Charova, S. N., Panopoulos, N. J. & Kokkinidis, M. (2009). Coiled-coils in type III secretion systems: structural flexibility, disorder and biological implications. *Cellular Microbiology* **11**, 719–729.

Gearhart, M. D., Holmbeck, S. M., Evans, R. M., Dyson, H. J. & Wright, P. E. (2003). Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *Journal of Molecular Biology* **327**, 819–832.

Gibson, T. J. (2009). Cell regulation: determined to signal discrete cooperation. *Trends in Biochemical Sciences* **34**, 471–482.

Goh, G. K., Dunker, A. K. & Uversky, V. N. (2008). Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics* **9** (Suppl. 2), S4.

Goh, G. K., Dunker, A. K. & Uversky, V. N. (2009). Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virology Journal* **6**, 1–12.

Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. (2002). Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *Journal of Biological Chemistry* **277**, 43168–43174.

Graf, P. C., Martinez-Yamout, M., Vanhaerents, S., Lilie, H., Dyson, H. J. & Jakob, U. (2004). Activation of the redox-regulated chaperone Hsp33 by domain unfolding. *Journal of Biological Chemistry* **279**, 20529–20538.

Grimmler, M., Wang, Y., Mund, T., Cilensek, Z., Keidel, E. M., Waddell, M. B., Jakel, H., Kullmann, M., Kriwacki, R. W. & Hengst, L. (2007). Cdk-inhibitory activity and stability of $p27^{Kip1}$ are directly regulated by oncogenic tyrosine kinases. *Cell* **128**, 269–280.

Gsponer, J. & Babu, M. M. (2009). The rules of disorder or why disorder rules. *Progress in Biophysics and Molecular Biology* **99**, 94–103.

Gsponer, J., Futschik, M. E., Teichmann, S. A. & Babu, M. M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* **322**, 1365–1368.

Gu, W. & Helms, V. (2005). Dynamical binding of proline-rich peptides to their recognition domains. *Biochimica et Biophysica Acta–Proteins and Proteomics* **1754**, 232–238.

Gunasekaran, K., Tsai, C. J., Kumar, S., Zanuy, D. & Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. *Trends in Biochemical Sciences* **28**, 81–85.

Hamada, D., Kato, T., Ikegami, T., Suzuki, K. N., Hayashi, M., Murooka, Y., Honda, T. & Yanagihara, I. (2005). EspB from enterohaemorrhagic *Escherichia coli* is a natively partially folded protein. *FEBS Journal* **272**, 756–768.

Hammes, G. G., Chang, Y. C. & Oas, T. G. (2009). Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13737–13741.

Hammoudeh, D. I., Follis, A. V., Prochownik, E. V. & Metallo, S. J. (2009). Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *Journal of the American Chemical Society* **131**, 7390–7401.

Han, J. D., Dupuy, D., Bertin, N., Cusick, M. E. & Vidal, M. (2005). Effect of sampling on topology predictions of protein-protein interaction networks. *Nature Biotechnology* **23**, 839–844.

Haynes, C. & Iakoucheva, L. M. (2006). Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Research* **34**, 305–312.

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M. & Iakoucheva, L. M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Computational Biology* **2**, e100.

He, Y., Liwo, A., Weinstein, H. & Scheraga, H. A. (2011). PDZ binding to the BAR domain of PICK1 is

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

elucidated by coarse-grained molecular dynamics. *Journal of Molecular Biology* **405**, 298–314.

Hegyi, H., Buday, L. & Tompa, P. (2009). Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Computational Biology* **5**, e1000552.

Hegyi, H., Kalmar, L., Horvath, T. & Tompa, P. (2011). Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. *Nucleic Acids Research* **39**, 1208–1219.

Hegyi, H. & Tompa, P. (2008). Intrinsically disordered proteins display no preference for chaperone binding in vivo. *PLoS Computational Biology* **4**, e1000017.

Hill, K. K., Roemer, S. C., Jones, D. N., Churchill, M. E. & Edwards, D. P. (2009). A progesterone receptor co-activator (JDP2) mediates activity through interaction with residues in the carboxyl-terminal extension of the DNA binding domain. *Journal of Biological Chemistry* **284**, 24415–24424.

Hilser, V. J. (2010). An ensemble view of allostery. *Science* **327**, 653–654.

Hirota, K. & Semenza, G. L. (2005). Regulation of hypoxia-inducible factor 1 by prolyl and asparaginyl hydroxylases. *Biochemical and Biophysical Research Communications* **338**, 610–616.

Holmbeck, S. M., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J. & Wright, P. E. (1998). High-resolution solution structure of the retinoid X receptor DNA-binding domain. *Journal of Molecular Biology* **281**, 271–284.

Holt, C. & Sawyer, L. (1993). Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the α_{S1}, β and κ-caseins. *Journal of the Chemical Society Faraday Transactions* **89**, 2683–2692.

Hsu, S. T., Bertoncini, C. W. & Dobson, C. M. (2009). Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. *Journal of the American Chemical Society* **131**, 7222–7223.

Huang, F., Rajagopalan, S., Settanni, G., Marsh, R. J., Armoogum, D. A., Nicolaou, N., Bain, A. J., Lerner, E., Haas, E., Ying, L. & Fersht, A. R. (2009). Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20758–20763.

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

Huang, Y. & Liu, Z. (2010b). Smoothing molecular interactions: The "kinetic buffer" effect of intrinsically disordered proteins. *Proteins* **78**, 3251–3259.

Hughes, S. & Graether, S. P. (2010). Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Science* **20**, 42–50.

Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z. & Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology* **323**, 573–584.

Iakoucheva, L. M. & Dunker, A. K. (2003). Order, disorder, and flexibility: prediction from protein sequence. *Structure* **11**, 1316–1317.

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research* **32**, 1037–1049.

Iwahara, J. & Clore, G. M. (2010). Structure-independent analysis of the breadth of the positional distribution of disordered groups in macromolecules from order parameters for long, variable-length vectors using NMR paramagnetic relaxation enhancement. *Journal of the American Chemical Society* **132**, 13346–13356.

Jaakola, V. P., Prilusky, J., Sussman, J. L. & Goldman, A. (2005). G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Engineering, Design and Selection* **18**, 103–110.

James, L. C., Roversi, P. & Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367.

James, L. C. & Tawfik, D. S. (2003). The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. *Protein Science* **12**, 2183–2193.

James, L. C. & Tawfik, D. S. (2005). Structure and kinetics of a transient antibody binding intermediate reveal a kinetic discrimination mechanism in antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12730–12735.

Jariel-Encontre, I., Bossis, G. & Piechaczyk, M. (2008). Ubiquitin-independent degradation of proteins by the proteasome. *Biochimica et Biophysica Acta* **1786**, 153–177.

Jowitt, T. A., Murdoch, A. D., Baldock, C., Berry, R., Day, J. M. & Hardingham, T. E. (2010). Order within disorder: Aggrecan chondroitin sulphate-attachment region provides new structural insights into protein sequences classified as disordered. *Proteins* **78**, 3317–3327.

Kanelis, V., Hudson, R. P., Thibodeau, P. H., Thomas, P. J. & Forman-Kay, J. D. (2010). NMR evidence for differential phosphorylation-dependent interactions in WT and ΔF508 CFTR. *EMBO Journal* **29**, 263–277.

Kast, D., Espinoza-Fonseca, L. M., Yi, C. & Thomas, D. D. (2010). Phosphorylation-induced structural changes in smooth muscle myosin regulatory light chain. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8207–8212.

Kavalenka, A., Urbancic, I., Belle, V., Rouger, S., Costanzo, S., Kure, S., Fournel, A., Longhi, S., Guigliarelli, B. & Strancar, J. (2010). Conformational analysis of the partially disordered measles virus N(TAIL)-XD complex by SDSL EPR spectroscopy. *Biophysical Journal* **98**, 1055–1064.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Kim, D. H., Lee, S. H., Nam, K. H., Chi, S. W., Chang, I. & Han, K. H. (2009). Multiple hTAF(II)31-binding motifs in the intrinsically unfolded transcriptional activation domain of VP16. *BMB Reports* **42**, 411–417.

Kim, P. M., Sboner, A., Xia, Y. & Gerstein, M. (2008). The role of disorder in interaction networks: a structural analysis. *Molecular Systems Biology* **4**, 179.

Kjaergaard, M., Teilum, K. & Poulsen, F. M. (2010). Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 12535–12540.

Kovacs, E., Tompa, P., Liliom, K. & Kalmar, L. (2010). Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5429–5434.

Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. (1996). Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 11504–11509.

Kucera, N. J., Hodsdon, M. E. & Wolin, S. L. (2011). An intrinsically disordered C terminus allows the La protein to assist the biogenesis of diverse noncoding RNA precursors. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1308–1313.

Kumar, R. & Thompson, E. B. (2010). Influence of flanking sequences on signaling between the activation function AF1 and DNA-binding domain of the glucocorticoid receptor. *Archives of Biochemistry and Biophysics* **496**, 140–145.

Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J. & Pavletich, N. P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948–953.

Lah, J., Simic, M., Vesnaver, G., Marianovsky, I., Glaser, G., Engelberg-Kulka, H. & Loris, R. (2005). Energetics of structural transitions of the addiction antitoxin MazE: is a programmed bacterial cell death dependent on the intrinsically flexible nature of the antitoxins? *Journal of Biological Chemistry* **280**, 17397–17407.

Laity, J. H., Dyson, H. J. & Wright, P. E. (2000). DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys_2–His_2 zinc fingers. *Journal of Molecular Biology* **295**, 719–727.

Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J. & Whitelaw, M. L. (2002). Asparagine hydroxylation of the HIF transactivation domain: a hypoxic switch. *Science* **295**, 858–861.

Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M. & Ryu, S. E. (2003). Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *Journal of Biological Chemistry* **278**, 7558–7563.

Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2009). Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry* **48**, 2115–2124.

Lee, C. W., Ferreon, J. C., Ferreon, A. C., Arai, M. & Wright, P. E. (2010a). Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 19290–19295.

Lee, C. W., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2010b). Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* **49**, 9964–9971.

Lee, H., Mok, K. H., Muhandiram, R., Park, K. H., Suk, J. E., Kim, D. H., Chang, J., Sung, Y. C., Choi, K. Y. & Han, K. H. (2000). Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *Journal of Biological Chemistry* **275**, 29426–29432.

Lee, J. O., Russo, A. A. & Pavletich, N. P. (1998). Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature* **391**, 859–865.

Lee, M. S., Gippert, G., Soman, K. Y., Case, D. A. & Wright, P. E. (1989). Three-dimensional solution structure of a single zinc finger binding domain. *Science* **245**, 635–637.

Leroy, A., Landrieu, I., Huvent, I., Legrand, D., Codeville, B., Wieruszeski, J. M. & Lippens, G. (2010). Spectroscopic studies of GSK3{beta} phosphorylation of the neuronal tau protein and its interaction with the N-terminal domain of apolipoprotein E. *Journal of Biological Chemistry* **285**, 33435–33444.

Lewis, H. A., Wang, C., Zhao, X., Hamuro, Y., Conners, K., Kearins, M. C., Lu, F., Sauder, J. M., Molnar, K. S., Coales, S. J., Maloney, P. C., Guggino, W. B., Wetmore, D. R., Weber, P. C. & Hunt, J. F. (2010). Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry. *Journal of Molecular Biology* **396**, 406–430.

Li, S., Iakoucheva, L. M., Mooney, S. D. & Radivojac, P. (2010). Loss of post-translational modification sites in disease. *Pacific Symposium on Biocomputing* **15**, 337–347.

Li, Y. & Palmer, A. G. (2010). Narrowing of protein NMR spectral lines broadened by chemical exchange. *Journal of the American Chemical Society* **132**, 8856–8857.

Lieutaud, P., Canard, B. & Longhi, S. (2008). MeDor: a metaserver for predicting protein disorder. *BMC Genomics* **9** (Suppl. 2), S25.

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003a). Protein disorder

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

prediction: implications for structural proteomics. *Structure* **11**, 1453–1459.

Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. (2003b). GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research* **31**, 3701–3708.

Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**, 4546–4559.

Liu, J., Faeder, J. R. & Camacho, C. J. (2009). Toward a quantitative theory of intrinsically disordered proteins and their function. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19819–19823.

Liu, J. & Nussinov, R. (2009). The mechanism of ubiquitination in the cullin-RING E3 ligase machinery: conformational control of substrate orientation. *PLoS Computational Biology* **5**, e1000527.

Liu, J. & Nussinov, R. (2010a). Molecular dynamics reveal the essential role of linker motions in the function of cullin-RING E3 ligases. *Journal of Molecular Biology* **396**, 1508–1523.

Liu, J. & Nussinov, R. (2010b). Rbx1 flexible linker facilitates cullin-RING ligase function before neddylation and after deneddylation. *Biophysical Journal* **99**, 736–744.

Liu, X. & Marmorstein, R. (2007). Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes and Development* **21**, 2711–2716.

Lobanov, M. Y., Furletova, E. I., Bogatyreva, N. S., Roytberg, M. A. & Galzitskaya, O. V. (2010a). Library of disordered patterns in 3D protein structures. *PLoS Computational Biology* **6**, e1000958.

Lobanov, M. Y., Shoemaker, B. A., Garbuzynskiy, S. O., Fong, J. H., Panchenko, A. R. & Galzitskaya, O. V. (2010b). ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. *Nucleic Acids Research* **38**, D283–D287.

Lobley, A., Swindells, M. B., Orengo, C. A. & Jones, D. T. (2007). Inferring function using patterns of native disorder in proteins. *PLoS Computational Biology* **3**, 1567–1579.

London, N., Movshovitz-Attias, D. & Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. *Structure* **18**, 188–199.

Love, J. J., Li, X., Chung, J., Dyson, H. J. & Wright, P. E. (2004). The LEF-1 high-mobility group domain undergoes a disorder-to-order transition upon formation of a complex with cognate DNA. *Biochemistry* **43**, 8725–8734.

Love, J. J., Li, X. A., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795.

Löw, C., Homeyer, N., Weininger, U., Sticht, H. & Balbach, J. (2009). Conformational switch upon phosphorylation: human CDK inhibitor $p19^{INK4d}$ between the native and partially folded state. *Acs Chemical Biology* **4**, 53–63.

Lowry, D. F., Hausrath, A. C. & Daughdrill, G. W. (2008a). A robust approach for analyzing a heterogeneous structural ensemble. *Proteins* **73**, 918–928.

Lowry, D. F., Stancik, A., Shrestha, R. M. & Daughdrill, G. W. (2008b). Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins* **71**, 587–598.

Lu, H. P. (2005). Probing single-molecule protein conformational dynamics. *Accounts of Chemical Research* **38**, 557–565.

Ma, B. & Nussinov, R. (2009). Regulating highly dynamic unstructured proteins and their coding mRNAs. *Genome Biology* **10**, 204.

Mackay, J. P., Sunde, M., Lowry, J. A., Crossley, M. & Matthews, J. M. (2007). Protein interactions: is seeing believing? *Trends in Biochemical Sciences* **32**, 530–531.

Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. (2010). Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8183–8188.

Marsh, J. A., Dancheck, B., Ragusa, M. J., Allaire, M., Forman-Kay, J. D. & Peti, W. (2010). Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* **18**, 1094–1103.

Marsh, J. A. & Forman-Kay, J. D. (2010). Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical Journal* **98**, 2383–2390.

Mathes, E., Wang, L., Komives, E. & Ghosh, G. (2010). Flexible regions within $I\kappa B\alpha$ create the ubiquitin-independent degradation signal. *Journal of Biological Chemistry* **285**, 32927–32936.

Matsushima, N., Yoshida, H., Kumaki, Y., Kamiya, M., Tanaka, T., Izumi, Y. & Kretsinger, R. H. (2008). Flexible structures and ligand interactions of tandem repeats consisting of proline, glycine, asparagine, serine, and/or threonine rich oligopeptides in proteins. *Current Protein and Peptide Science* **9**, 591–610.

Mayer, B. J., Hamaguchi, M. & Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. *Nature* **332**, 272–275.

Meier, S., Blackledge, M. & Grzesiek, S. (2008). Conformational distributions of unfolded polypeptides from novel NMR techniques. *Journal of Chemical Physics* **128**, 052204.

Merlini, G. & Bellotti, V. (2003). Molecular mechanisms of amyloidosis. *New England Journal of Medicine* **349**, 583–596.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Midic, U., Oldfield, C. J., Dunker, A. K., Obradovic, Z. & Uversky, V. N. (2009). Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. *Protein and Peptide Letters* **16**, 1533–1547.

Miller, M. (2009). The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Current Protein and Peptide Science* **10**, 244–269.

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

Mittermaier, A. K. & Kay, L. E. (2009). Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* **34**, 601–611.

Mobley, D. L. & Dill, K. A. (2009). Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **17**, 489–498.

Mohan, A., Oldfield, C. J., Radivojac, P., Vacic, V., Cortese, M. S., Dunker, A. K. & Uversky, V. N. (2006). Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology* **362**, 1043–1059.

Mohan, A., Sullivan, Jr., W. J., Radivojac, P., Dunker, A. K. & Uversky, V. N. (2008). Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Molecular Biosystems* **4**, 328–340.

Morin, B., Bourhis, J. M., Belle, V., Woudstra, M., Carriere, F., Guigliarelli, B., Fournel, A. & Longhi, S. (2006). Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling electron paramagnetic resonance spectroscopy. *Journal of Physical Chemistry B* **110**, 20596–20608.

Motáčková, V., Nováček, J., Zawadzka-Kazimierczuk, A., Kazimierczuk, K., Zídek, L., Sanderová, H., Krásný, L., Kozminski, W. & Sklenár, V. (2010). Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *Journal of Biomolecular NMR* **48**, 169–177.

Mukrasch, M. D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C., Mandelkow, E. & Zweckstetter, M. (2009). Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biology* **7**, e34.

Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D. & Schuler, B. (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14609–14614.

Narayanan, A. & Jacobson, M. P. (2009). Computational studies of protein regulation by post-translational phosphorylation. *Current Opinion in Structural Biology* **19**, 156–163.

Nguyen, D. X., Baglia, L. A., Huang, S. M., Baker, C. M. & Mccance, D. J. (2004). Acetylation regulates the differentiation-specific functions of the retinoblastoma protein. *EMBO Journal* **23**, 1609–1618.

Niu, X., Bruschweiler-Li, L., Davulcu, O., Skalicky, J. J., Bruschweiler, R. & Chapman, M. S. (2011). Arginine kinase: joint crystallographic and NMR RDC analyses link substrate-associated motions to intrinsic flexibility. *Journal of Molecular Biology* **405**, 479–496.

Nocula-Lugowska, M., Rymarczyk, G., Lisowski, M. & Ozyhar, A. (2009). Isoform-specific variation in the intrinsic disorder of the ecdysteroid receptor N-terminal domain. *Proteins* **76**, 291–308.

Noutsou, M., Duarte, A. M., Anvarian, Z., Didenko, T., Minde, D. P., Kuper, I., De, R. I., Oikonomou, C., Friedler, A., Boelens, R., Rudiger, S. G. & Maurice, M. M. (2011). Critical scaffolding regions of the tumor suppressor axin1 are natively unfolded. *Journal of Molecular Biology* **405**, 773–786.

O'Dea, E. L., Barken, D., Peralta, R. Q., Tran, K. T., Werner, S. L., Kearns, J. D., Levchenko, A. & Hoffmann, A. (2007). A homeostatic model of IκB metabolism to control constitutive NF-κB activity. *Molecular Systems Biology* **3**, 111.

Ohgushi, M. & Wada, A. (1983). 'Molten-globule state': a compact form of globular proteins with mobile side chains. *FEBS Letters* **164**, 21–24.

Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N. & Dunker, A. K. (2005). Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454–12470.

Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N. & Dunker, A. K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14–3-3 with their partners. *BMC Genomics* **9** (Suppl. 1), S1.

Olson, K. E., Narayanaswami, P., Vise, P. D., Lowry, D. F., Wold, M. S. & Daughdrill, G. W. (2005). Secondary structure and dynamics of an intrinsically unstructured linker domain. *Journal of Biomolecular Structure and Dynamics* **23**, 113–124.

Olzscha, H., Schermann, S. M., Woerner, A. C., Pinkert, S., Hecht, M. H., Tartaglia, G. G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F. U. & Vabulas, R. M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* **144**, 67–78.

Onitsuka, M., Kamikubo, H., Yamazaki, Y. & Kataoka, M. (2008). Mechanism of induced folding: Both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants. *Proteins* **72**, 837–847.

Paliy, O., Gargac, S. M., Cheng, Y., Uversky, V. N. & Dunker, A. K. (2008). Protein disorder is positively correlated with gene expression in *Escherichia coli*. *Journal of Proteome Research* **7**, 2234–2245.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Palmer, A. G. (2004). NMR characterization of the dynamics of biomacromolecules. *Chemical Reviews* **104**, 3623–3640.

Park, S. J., Borin, B. N., Martinez-Yamout, M. A. & Dyson, H. J. (2011). The client protein p53 forms a molten globule-like state in the presence of Hsp90. *Nature Structural and Molecular Biology* **18**, 537–541.

Parker, D., Rivera, M., Zor, T., Henrion-Caude, A., Radhakrishnan, I., Kumar, A., Shapiro, L. H., Wright, P. E., Montminy, M. & Brindle, P. K. (1999). Role of secondary structure in discrimination between constitutive and inducible activators. *Molecular and Cellular Biology* **19**, 5601–5607.

Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevit, R. E. (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* **241**, 1489–1492.

Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. (2006). Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208

Pentony, M. M. & Jones, D. T. (2010). Modularity of intrinsic disorder in the human proteome. *Proteins* **78**, 212–221.

Perez, Y., Gairi, M., Pons, M. & Bernado, P. (2009). Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *Journal of Molecular Biology* **391**, 136–148.

Petsko, G. A. (2001). Structural basis of thermostability in hyperthermophilic proteins, or "there's more than one way to skin a cat". *Methods in Enzymology* **334**, 469–478.

Pratt, W. B. & Toft, D. O. (1997). Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocrine Reviews* **18**, 306–360.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435–3438.

Privalov, P. L., Dragan, A. I., Crane-Robinson, C., Breslauer, K. J., Remeta, D. P. & Minetti, C. A. S. A. (2007). What drives proteins into the major or minor grooves of DNA? *Journal of Molecular Biology* **365**, 1–9.

Ptitsyn, O. B. (1973). Stages in the mechanism of self-organization of protein molecules. *Doklady Akademii nauk SSSR* **210**, 1213–1215.

Qin, B. Y., Liu, C., Srinath, H., Lam, S. S., Correia, J. J., Derynck, R. & Lin, K. (2005). Crystal structure of IRF-3 in complex with CBP. *Structure* **13**, 1269–1277.

Radhakrishnan, I., Perez-Alvarado, G. C., Dyson, H. J. & Wright, P. E. (1998). Conformational preferences in the Ser^{133}-phosphorylated and non-phosphorylated forms of the kinase inducible transactivation domain of CREB. *FEBS Letters* **430**, 317–322.

Radhakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R. & Wright, P. E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91**, 741–752.

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., Goebl, M. G. & Iakoucheva, L. M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78**, 365–380.

Radivojac, P., Vucetic, S., O'Connor, T. R., Uversky, V. N., Obradovic, Z. & Dunker, A. K. (2006). Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins* **63**, 398–410.

Rancurel, C., Khosravi, M., Dunker, A. K., Romero, P. R. & Karlin, D. (2009). Overlapping genes produce proteins with unusual sequence properties and offer insight into *de novo* protein creation. *Journal of Virology* **83**, 10719–10736.

Ravindranathan, S., Oberstrass, F. C. & Allain, F. H. (2010). Increase in backbone mobility of the VTS1p-SAM domain on binding to SRE-RNA. *Journal of Molecular Biology* **396**, 732–746.

Reingewertz, T. H., Benyamini, H., Lebendiker, M., Shalev, D. E. & Friedler, A. (2009). The C-terminal domain of the HIV-1 Vif protein is natively unfolded in its unbound state. *Protein Engineering Design and Selection* **22**, 281–287.

Ren, S., Uversky, V. N., Chen, Z., Dunker, A. K. & Obradovic, Z. (2008). Short Linear Motifs recognized by SH2, SH3 and Ser/Thr Kinase domains are conserved in disordered protein regions. *BMC Genomics* **9**(Suppl. 2), S26.

Romero, P., Obradovic, Z. & Dunker, A. K. (1999). Folding minimal sequences: the lower bound for sequence complexity of globular proteins. *FEBS Letters* **462**, 363–367.

Romero, P., Obradovic, Z. & Dunker, A. K. (2004). Natively disordered proteins: functions and predictions. *Applied Bioinformatics* **3**, 105–113.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E. & Dunker, A. K. (1997). Identifying disordered regions in proteins from amino acid sequences. *Proceedings of the IEEE International Conference on Neural Networks* **1**, 90–95.

Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S. & Dunker, A. K. (1998). Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing* **3**, 435–446.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. & Dunker, A. K. (2001). Sequence complexity of disordered protein. *Proteins* **42**, 38–48.

Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., Cortese, M. S., Sickmeier, M., Legall, T., Obradovic, Z. & Dunker, A. K. (2006). Alternative splicing in concert with

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8390–8395.

Rospigliosi, C. C., Mcclendon, S., Schmid, A. W., Ramlall, T. F., Barre, P., Lashuel, H. A. & Eliezer, D. (2009). E46K Parkinson's-linked mutation enhances C-terminal-to-N-terminal contacts in alpha-synuclein. *J. Mol. Biol.* **388**, 1022–1032.

Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M. & Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. *Journal of the American Chemical Society* **132**, 8407–8418.

Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Roder, L., Euzenat, J., Rechenmann, F. & Jacq, B. (1999). Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database. *Nucleic Acids Research* **27**, 89–94.

Sandhu, K. S. (2009). Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *Journal of Molecular Recognition* **22**, 1–8.

Schlessinger, A., Punta, M. & Rost, B. (2007). Natively unstructured regions in proteins identified from contact predictions. *Bioinformatics* **23**, 2376–2384.

Sgourakis, N. G., Patel, M. M., Garcia, A. E., Makhatadze, G. I. & Mccallum, S. A. (2010). Conformational dynamics and structural plasticity play critical roles in the ubiquitin recognition of a UIM domain. *Journal of Molecular Biology* **396**, 1128–1144.

Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 8868–8873.

Sickmeier, M., Hamilton, J. A., Legall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z. & Dunker, A. K. (2007). DisProt: the database of disordered proteins. *Nucleic Acids Research* **35**, D786–D793.

Sigalov, A. B. (2010). Protein intrinsic disorder and oligomericity in cell signaling. *Molecular Biosystems* **6**, 451–461.

Sigalov, A. B., Kim, W. M., Saline, M. & Stern, L. J. (2008). The intrinsically disordered cytoplasmic domain of the T cell receptor zeta chain binds to the nef protein of simian immunodeficiency virus without a disorder-to-order transition. *Biochemistry* **47**, 12942–12944.

Singh, G. P. & Dash, D. (2007). Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* **68**, 602–605.

Singh, G. P., Ganapathi, M. & Dash, D. (2007). Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* **66**, 761–765.

Sivakolundu, S. G., Nourse, A., Moshiach, S., Bothner, B., Ashley, C., Satumba, J., Lahti, J. & Kriwacki, R. W. (2008). Intrinsically unstructured domains of Arf and Hdm2 form bimolecular oligomeric structures in vitro and in vivo. *Journal of Molecular Biology* **384**, 240–254.

Smock, R. G. & Gierasch, L. M. (2009). Sending signals dynamically. *Science* **324**, 198–203.

Solt, I., Magyar, C., Simon, I., Tompa, P. & Fuxreiter, M. (2006). Phosphorylation-induced transient intrinsic structure in the kinase-inducible domain of CREB facilitates its recognition by the KIX domain of CBP. *Proteins* **64**, 749–757.

Sotomayor Perez, A. C., Karst, J. C., Davi, M., Guijarro, J. I., Ladant, D. & Chenal, A. (2010). Characterization of the regions involved in the calcium-induced folding of the intrinsically disordered RTX motifs from the bordetella pertussis adenylate cyclase toxin. *Journal of Molecular Biology* **397**, 534–549.

Srere, P. A. (2000). Macromolecular interactions: tracing the roots. *Trends in Biochemical Sciences* **25**, 150–153.

Stahl, M. L., Ferenz, C. R., Kelleher, K. L., Kriz, R. W. & Knopf, J. L. (1988). Sequence similarity of phospholipase C with the non-catalytic region of src. *Nature* **332**, 269–272.

Staneva, I. & Wallin, S. (2009). All-atom Monte Carlo approach to protein-peptide binding. *Journal of Molecular Biology* **393**, 1118–1128.

Stein, A., Pache, R. A., Bernado, P., Pons, M. & Aloy, P. (2009). Dynamic interactions of proteins in complex networks: a more structured view. *The FEBS Journal* **276**, 5390–5405.

Sue, S. C., Cervantes, C., Komives, E. A. & Dyson, H. J. (2008). Transfer of flexibility between ankyrin repeats in IκBα upon formation of the NF-κB complex. *Journal of Molecular Biology* **380**, 917–931.

Sue, S.-C., Alverdi, V., Komives, E. A. & Dyson, H. J. (2011). Detection of a ternary complex of NF-κB and IκBa with DNA provides insights into how IκBα removes NF-κB from transcription sites. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1367–1372.

Sugase, K., Dyson, H. J. & Wright, P. E. (2007a). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025.

Sugase, K., Lansing, J. C., Dyson, H. J. & Wright, P. E. (2007b). Tailoring relaxation dispersion experiments for fast-associating protein complexes. *Journal of the American Chemical Society* **129**, 13406–13407.

Sung, Y. & Eliezer, D. (2007). Residual Structure, Backbone Dynamics, and Interactions within the Synuclein Family. *Journal of Molecular Biology* **372**, 689–707.

Tainer, J. A., Getzoff, E. D., Alexander, H., Houghten, R. A., Olson, A. J., Lerner, R. A. & Hendrickson, W. A. (1984). The reactivity of anti-peptide antibodies is a function of the atomic mobility of sites in a protein. *Nature* **312**, 127–133.

Takenaga, K. (2011). Angiogenic signaling aberrantly induced by tumor hypoxia. *Frontiers in Bioscience* **16**, 31–48.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V. M., Byeon, I. J. L., Anjum, D. H., Kodali, R., Creamer, T. P., Conway, J. F., Gronenborn, A. M. & Wetzel, R. (2009). Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nature Structural and Molecular Biology* **16**, 380–389.

Tokuriki, N. & Tawfik, D. S. (2009). Protein dynamism and evolvability. *Science* **324**, 203–207.

Tompa, P. & Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB Journal* **18**, 1169–1175.

Tompa, P., Dosztanyi, Z. & Simon, I. (2006). Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *Journal of Proteome Research* **5**, 1996–2000.

Tompa, P. & Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences* **33**, 2–8.

Tompa, P. & Kalmar, L. (2010). Power law distribution defines structural disorder as a structural element directly linked with function. *Journal of Molecular Biology* **403**, 346–350.

Tran, H. T., Mao, A. & Pappu, R. V. (2008). Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *Journal of the American Chemical Society* **130**, 7380–7392.

Tran, H. T., Wang, X. & Pappu, R. V. (2005). Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* **44**, 11369–11380.

Truhlar, S. M., Torpey, J. W. & Komives, E. A. (2006). Regions of IκBα that are critical for its inhibition of NF-κB.DNA interaction fold upon binding to NF-κB. *Proceedings of the National Academy of Science of the United States of America* **103**, 18951–18956.

Tsai, C. J., Del, S. A. & Nussinov, R. (2009a). Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular Biosystems* **5**, 207–216.

Tsai, C. J., Ma, B. & Nussinov, R. (2009b). Protein-protein interaction networks: how can a hub protein bind so many different partners? *Trends in Biochemical Sciences* **34**, 594–600.

Tsunaka, Y., Toga, J., Yamaguchi, H., Tate, S., Hirose, S. & Morikawa, K. (2009). Phosphorylated intrinsically disordered region of FACT masks its nucleosomal DNA binding elements. *Journal of Biological Chemistry* **284**, 24610–24621.

Tsvetkov, P., Reuven, N., Prives, C. & Shaul, Y. (2009a). Susceptibility of p53 unstructured N terminus to 20 S proteasomal degradation programs the stress response. *Journal of Biological Chemistry* **284**, 26234–26242.

Tsvetkov, P., Reuven, N. & Shaul, Y. (2009b). The nanny model for IDPs. *Nature Chemical Biology* **5**, 778–781.

Turjanski, A. G., Gutkind, J. S., Best, R. B. & Hummer, G. (2008). Binding-induced folding of a natively unstructured transcription factor. *PLoS Computational Biology* **4**, e1000060.

Tyagi, M., Shoemaker, B. A., Bryant, S. H. & Panchenko, A. R. (2009). Exploring functional roles of multibinding protein interfaces. *Protein Science* **18**, 1674–1683.

Tyka, M. D., Keedy, D. A., Andre, I., Dimaio, F., Song, Y., Richardson, D. C., Richardson, J. S. & Baker, D. (2010). Alternate states of proteins revealed by detailed energy landscape mapping. *Journal of Molecular Biology* **405**, 607–618.

Uversky, V. N. (2002). Natively unfolded proteins: a point where biology waits for physics. *Protein Science* **11**, 739–756.

Uversky, V. N. (2009a). Intrinsic disorder in proteins associated with neurodegenerative diseases. *Frontiers in Bioscience* **14**, 5188–5238.

Uversky, V. N. (2009b). Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein Journal* **28**, 305–325.

Uversky, V. N. (2010). Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chemical Society Reviews* **40**, 1623–1634.

Uversky, V. N. & Eliezer, D. (2009). Biophysics of Parkinson's disease: structure and aggregation of α-synuclein. *Current Protein and Peptide Science* **10**, 483–499.

Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.

Uversky, V. N., Oldfield, C. J. & Dunker, A. K. (2008). Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annual Review of Biophysics* **37**, 215–246.

Uversky, V. N. & Ptitsyn, O. B. (1994). "Partly folded" state, a new equilibrium state of protein molecules: Four-state guanidinium chloride-induced unfolding of β-lactamase at low temperature. *Biochemistry* **33**, 2782–2791.

Uversky, V. N. & Ptitsyn, O. B. (1996). Further evidence on the equilibrium "pre-molten globule state": Four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *Journal of Molecular Biology* **255**, 215–228.

Uversky, V. N., Roman, A., Oldfield, C. J. & Dunker, A. K. (2006). Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *Journal of Proteome Research* **5**, 1829–1842.

Vacic, V., Oldfield, C. J., Mohan, A., Radivojac, P., Cortese, M. S., Uversky, V. N. & Dunker, A. K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. *Journal of Proteome Research* **6**, 2351–2366.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Vallee-Belisle, A., Ricci, F. & Plaxco, K. W. (2009). Thermodynamic basis for the optimization of binding-induced biomolecular switches and structure-switching biosensors. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13802–13807.

Viles, J. H., Donne, D. G., Kroon, G. J. A., Prusiner, S. B., Cohen, F. E., Dyson, H. J. & Wright, P. E. (2001). Local structural plasticity of the prion protein. *Analysis of NMR relaxation dynamics. Biochemistry* **40**, 2743–2753.

Vuzman, D., Azia, A. & Levy, Y. (2010a). Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails. *Journal of Molecular Biology* **396**, 674–684.

Vuzman, D., Polonsky, M. & Levy, Y. (2010b). Facilitated DNA search by multidomain transcription factors: cross talk via a flexible linker. *Biophysical Journal* **99**, 1202–1211.

Wang, F., Marshall, C. B., Li, G. Y., Yamamoto, K., Mak, T. W. & Ikura, M. (2009). Synergistic interplay between promoter recognition and CBP/p300 coactivator recruitment by FOXO3a. *Acs Chemical Biology* **4**, 1017–1027.

Wang, X., Koulov, A. V., Kellner, W. A., Riordan, J. R. & Balch, W. E. (2008). Chemical and biological folding contribute to temperature-sensitive ΔF508 CFTR trafficking. *Traffic* **9**, 1878–1893.

Wang, X., Zhang, S., Zhang, J., Huang, X., Xu, C., Wang, W., Liu, Z., Wu, J. & Shi, Y. (2010). A large intrinsically disordered region in SKIP and its disorder–order transition induced by PPIL1 binding revealed by NMR. *Journal of Biological Chemistry* **285**, 4951–4963.

Wang, Y., Filippov, I., Richter, C., Luo, R. & Kriwacki, R. W. (2005). Solution NMR studies of an intrinsically unstructured protein within a dilute 75 kDa eukaryotic protein assembly; probing the practical limits for efficiently assigning polypeptide backbone resonances. *ChemBiochem: A European Journal of Chemical Biology* **6**, 2242–2246.

Ward, J. J., Sodhi, J. S., Mcguffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of Molecular Biology* **337**, 635–645.

Waters, L., Yue, B., Veverka, V., Renshaw, P., Bramham, J., Matsuda, S., Frenkiel, T., Kelly, G., Muskett, F., Carr, M. & Heery, D. M. (2006). Structural diversity in p160/CREB-binding protein coactivator complexes. *Journal of Biological Chemistry* **281**, 14787–14795.

Weathers, E. A., Paulaitis, M. E., Woolf, T. B. & Hoh, J. H. (2004). Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Letters* **576**, 348–352.

Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A. & Lansbury, P. T. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **35**, 13709–13715.

Welch, G. R. (2009). The "fuzzy" interactome. *Trends in Biochemical Sciences* **34**, 1–2.

Wilkins, M. R. & Kummerfeld, S. K. (2008). Sticking together? Falling apart? Exploring the dynamics of the interactome. *Trends in Biochemical Sciences* **33**, 195–200.

Williams, R. M., Obradovic, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J. & Dunker, A. K. (2001). The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pacific Symposium on Biocomputing* **6**, 89–100.

Wilson, I. A., Haft, D. H., Getzoff, E. D., Tainer, J. A., Lerner, R. A. & Brenner, S. (1985). Identical short peptide sequences in unrelated proteins can have different conformations: a testing ground for theories of immune recognition. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 5255–5259.

Winkler, G. R., Harkins, S. B., Lee, J. C. & Gray, H. B. (2006). α-synuclein structures probed by 5-fluorotryptophan fluorescence and ^{19}F NMR spectroscopy. *Journal of Physical Chemistry B* **110**, 7058–7061.

Wojciak, J. M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. (2009). Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains. *EMBO Journal* **28**, 948–958.

Won, H. S., Low, L. Y., Guzman, R. D., Martinez-Yamout, M., Jakob, U. & Dyson, H. J. (2004). The zinc-dependent redox switch domain of the chaperone Hsp33 has a novel fold. *Journal of Molecular Biology* **341**, 893–899.

Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293**, 321–331.

Wright, P. E. & Dyson, H. J. (2009). Linking folding and binding. *Current Opinion in Structural Biology* **19**, 31–38.

Wu, K. P. & Baum, J. (2010). Detection of transient interchain interactions in the intrinsically disordered protein α-synuclein by NMR paramagnetic relaxation enhancement. *Journal of the American Chemical Society* **132**, 5546–5547.

Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M. & Wright, P. E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *Journal of Molecular Biology* **273**, 183–206.

Xu, J. & Xia, J. (2006). Structure and function of PICK1. *Neurosignals* **15**, 190–201.

Xue, B., Li, L., Meroueh, S. O., Uversky, V. N. & Dunker, A. K. (2009a). Analysis of structured and intrinsically disordered regions of transmembrane proteins. *Molecular Biosystems* **5**, 1688–1702.

Xue, B., Oldfield, C. J., Dunker, A. K. & Uversky, V. N. (2009b). CDF it all: consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions. *FEBS Letters* **583**, 1469–1474.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press

Xue, B., Williams, R. W., Oldfield, C. J., Dunker, A. K. & Uversky, V. N. (2010a). Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Systems Biology* **4**(Suppl 1), S1.

Xue, B., Williams, R. W., Oldfield, C. J., Goh, G. K., Dunker, A. K. & Uversky, V. N. (2010b). Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein and Peptide Letters* **17**, 932–951.

Xue, Y., Podkorytov, I. S., Rao, D. K., Benjamin, N., Sun, H. & Skrynnikov, N. R. (2009c). Paramagnetic relaxation enhancements in unfolded proteins: theory and application to drkN SH3 domain. *Protein Science* **18**, 1401–1424.

Zhu, F., Kapitan, J., Tranter, G. E., Pudney, P. D. A., Isaacs, N. W., Hecht, L. & Barron, L. D. (2008). Residual structure in disordered peptides and unfolded proteins from multivariate analysis and ab initio simulation of Raman optical activity data. *Proteins* **70**, 823–833.

Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R. & Wright, P. E. (2002). Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *Journal of Biological Chemistry* **277**, 42241–42248.

Zwier, M. C. & Chong, L. T. (2010). Reaching biological timescales with all-atom molecular dynamics simulations. *Current Opinion in Pharmacology* **10**, 745–752.

https://doi.org/10.1017/S0033583511000060 Published online by Cambridge University Press