Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions

Christopher J. Oldfield and A. Keith Dunker

Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202; email: cjoldfie@iupui.edu, kedunker@iupui.edu

Annu. Rev. Biochem. 2014. 83:553–84

First published online as a Review in Advance on March 5, 2014

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

This article's doi: 10.1146/annurev-biochem-072711-164947

Keywords

rheomorphic, natively, inherently, unstructured, unfolded, flexible, malleable, chameleon

Abstract

Intrinsically disordered proteins (IDPs) and IDP regions fail to form a stable structure, yet they exhibit biological activities. Their mobile flexibility and structural instability are encoded by their amino acid sequences. They recognize proteins, nucleic acids, and other types of partners; they accelerate interactions and chemical reactions between bound partners; and they help accommodate posttranslational modifications, alternative splicing, protein fusions, and insertions or deletions. Overall, IDP-associated biological activities complement those of structured proteins. Recently, there has been an explosion of studies on IDP regions and their functions, yet the discovery and investigation of these proteins have a long, mostly ignored history. Along with recent discoveries, we present several early examples and the mechanisms by which IDPs contribute to function, which we hope will encourage comprehensive discussion of IDPs and IDP regions in biochemistry textbooks. Finally, we propose future directions for IDP research.

Contents

CHARACTERISTICS

Intrinsically disordered proteins (IDPs) have been called many names (reviewed in Reference 1). We use the term "disordered" because of Jirgensons's (2) use of it for protein classification, because of Arnone et al.'s (3) use of it to describe ill-structured regions, and because this name has broad coverage (4). The word "intrinsically" indicates a sequence-dependent characteristic (5). Small-molecule ligands (6, 7), macromolecular binding partners, or posttranslational modifications (PTMs) (7) can induce IDPs or IDP regions to become structured or can cause structured domains to become IDPs (8).

Linkers (9), entropic springs (10) or elastomers (11), entropic bristles (12), and native molten globules (13, 14) all directly use flexibility to carry out function. Order–disorder transitions can underlie function (15), such as folding upon binding (16) or a chaperone's activation following unfolding (8).

Primarily because of their lack of structural constraints, IDPs and IDP regions facilitate several biological processes (17). Examples of such processes include alternative splicing (AS) (18); movement through narrow pores or channels (19); many but not all PTMs (20); overprinting (21), also known as dual coding in alternative reading frames (22); creation of oncogenic chimera proteins following aberrant gene fusion (23); and insertions and deletions (INDELs), especially long ones (24), arising from mutation (25).

IDPs with significant net charge are extended, with greater extension for greater net charge (26). IDPs with large but approximately equal numbers of oppositely charged side chains behave as polyampholytes, with greater extension if the charges are randomly distributed and with less extension if the plus and minus charges become more segregated (27). Neutral hydrophilic IDPs with few charged residues form collapsed, random structures because water is a poor solvent both for the backbone (28) and for polar, uncharged sequences, such as polyQ (29), or for sequences of native premolten globules (30). Finally, hydrophobic IDPs that have structure-forming sequence patterns, but insufficient folding energy, collapse into molten globules characterized by

mobile side chains and unstable tertiary structures (13). Premolten and molten globules have been observed as transient protein folding intermediates and as stable forms for several structured proteins under mildly denaturing conditions (31). For some sequences, molten and premolten globules represent folding end points. Such proteins can use their particular characteristics to perform biological functions (13, 14, 32). In summary, IDPs and IDP regions lack a stable structure; instead, they exist as conformational ensembles (30, 31) without equilibrium positions for their atom positions and bond angles (17, 35).

Structured proteins are grouped by their secondary structures (2, 36, 37). In contrast, IDPs and IDP regions are more difficult to partition into groups (38). Recent research has determined that, although all structured proteins can be unfolded by denaturing agents, some IDPs can be induced to fold but others cannot. This finding indicates a need to distinguish nonfoldable IDPs from foldable, conditional IDPs (39–41), including IDPs with a semidisordered form that have an increased tendency to form amyloids (41).

EARLY EXAMPLES

Beginning in the 1950s, many IDPs and IDP regions with biological functions were discovered; these include casein (42), phosvitin (43), fibrinogen (44), trypsinogen (45), and calcineurin (CaN) (46). The intrinsically disordered aspect of proteins had mostly been ignored until recently (**Figure 1**). The many recent publications about IDPs mean that this review can provide only a sparse sample of this literature; many outstanding papers have to be left out.

Current biochemistry textbooks either completely omit or contain only brief discussions of IDPs, IDP regions, and their roles in function. Extensive and integrated coverage of IDPs is needed to improve our understanding of the relationships between protein structure and function.

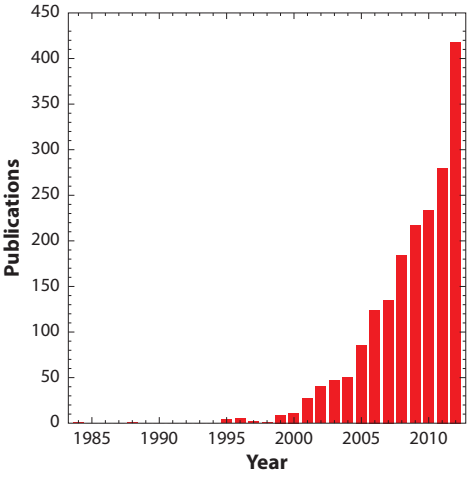

Figure 1

Number of publications related to intrinsically disordered proteins (IDPs) by year, from 1984 to 2012. Publications were retrieved from a search of PubMed (**http://www.ncbi.nlm.nih.gov/pubmed**) using IDP-related terms: "(inherently OR natively OR intrinsically) AND (disordered OR unfolded OR unstructured) AND protein." These numbers increase significantly when more terms are used in the search, but the false-positive rate also increases.

Casein

Eutherian milks contain multiple casein isoforms, such as α_{S1}, α_{S2}, β, and κ (47). Investigators have known since before 1950 that casein can survive prolonged heating; can survive treatment with denaturing agents such as urea, quanidine hydrochloride, or mildly acidic or basic solutions without significant change; and is highly sensitive to protease digestion. Specific optical rotation showed that casein closely resembles proteins that are unfolded by guanidine. Therefore, in 1952 McMeekin (42, p. 58) suggested that "casein occurs in milk in an unfolded configuration, which may be rapidly digested by proteolytic enzymes."

Thousands of casein molecules associate to form soluble aggregates termed micelles. Many different models for casein micelles have been proposed (48). One model (49) that is now gaining favor (50) considers a casein micelle to be an open, IDP-based matrix with stabilization

arising from several types of interactions, including critical contributions from phosphorylated side-chain attachments to calcium phosphate nanoclusters.

The separated κ and α_{S2} caseins form amyloid fibrils. Also, the caseins exhibit heat-shock protein–like molecular chaperone activity. The many alternative interactions associated with chaperone activity may compete with specific, fibril-forming interactions, thereby giving rise to amorphous aggregates rather than fibrils. Indeed, the prevention of fibril formation may be an important function of casein micelles (47). In general, avoidance of fibril formation is an important consideration for IDPs, IDP regions, and their evolution (41).

As a result of binding by casein, calcium phosphate reaches supersaturating levels in milk (51). Such supersaturation is crucial for the development and maintenance of bones and teeth. A collection of secreted calcium (phosphate)-binding phosphoproteins have been identified. These proteins may sequester calcium phosphate nanoclusters by casein-like mechanisms, so IDPs are probably important for sustaining bones and teeth (51, 52).

Phosvitin

In 1966, Jirgensons (2) studied the conformations of many proteins by using optical rotatory dispersion (ORD); these comparisons led him to propose that proteins should be classified by their conformations. His proposal preceded the Structural Classification of Proteins (SCOP) (36) and Class, Architecture, Topology, and Homology (CATH) (37) databases by ∼30 years. Unlike SCOP and CATH, however, Jirgensons's classification scheme included the category "disordered," which was based mainly on the protein phosvitin (2).

Jirgensons (43) demonstrated phosvitin's disorder in 1958 by viscosity and ORD; his results were later confirmed by circular dichroism (CD) and nuclear magnetic resonance (NMR). Phosvitin, from egg yolk, contains ∼57% serine, most of it phosphorylated, along with many polar residues and very few hydrophobic side chains (53). These sequence characteristics account for Jirgensons's (43) observation that phosvitin behaves as a flexible polyanion.

Fibrinogen to Fibrin

The fibrinogen hexamer ($\alpha_2\beta_2\gamma_2$) is a vertebrate blood plasma glycoprotein. Following injury, thrombin cleaves fibrinogen, converting fibrinogen into fibrin, which then forms an insoluble matrix (44). The thrombin cleavages create GPRP and GHRP termini for the α- and β-chains, respectively. These "knobs" dock into specific "holes" in the C domains of the fibrin β- and γ-subunits, respectively. The thrombin cut sites map to regions of missing electron density in human [Protein Data Bank (PDB) identifier 3GHG] and chicken (PDB 1EI3) fibrinogen, and additional flexible residues link the GPRP and GHRP knobs to their associated globular domains, enabling tethered searches that allow the knobs to dock into their respective holes.

Protease sensitivity and random-coil CD spectra of a large region near the carboxyl end of the α-chain, a region termed αC, led Doolittle (44) to suggest the existence of "free swimming appendages" within the complex. This αC region is a 273-residue segment and, to date, is the longest missing electron-density region of any PDB protein that is likely not be to a mobile, structured domain. The IDP regions of fibrinogen represent 28% of residues in the asymmetric unit, also the largest proportion of disorder observed so far for any PDB asymmetric unit (54).

The αC domain shows little sequence conservation (55) and is also very rich in INDELs (55). IDP regions often exhibit high sequence variability (56) and numerous INDELs (24, 25). Long insertions are frequently IDP regions (24, 57).

CD and NMR spectra suggest that the αC domains contain structured regions of limited size that are connected by flexible IDP linkers (58). Hydrogen-exchange studies on intact fibrinogen show rapid exchange for most of the αC region, indicating that significant parts

of this region remain unstructured in the fibrinogen assembly (59)—exactly as suggested by Doolittle (44) and Doolittle & Kollman (55).

In addition to the fibrin gel–like matrix, a blood clot contains several additional bound components. The αC IDP region binds to many of them (55, 58). Such binding to multiple partners is a common characteristic of IDPs that is enabled by their flexibility (60, 61).

Trypsinogen to Trypsin

Trypsinogen is activated by enteropeptidase cleavage after the K^+ in trypsinogen's VDDDDK amino terminal sequence (62), which is located in a 15-residue region of missing electron density (45). As for fibrinogen, the rapid digestion of trypsinogen is probably accelerated when the cleavage site is located in an IDP region.

After enteropeptidase cleavage, a two-residue IV knob appears at the end of an IDP region. Following a tethered search, this IV knob binds into a hole that promotes a disorder-to-order transition for trypsin's substrate binding pocket. Trypsinogen is inactive not because its catalytic triad is misaligned, but because the K^+/R^+ binding pocket is not structured (45).

Addition of the IV dipeptide and other hydrophobic dipeptides to solutions of trypsinogen activates it without protease digestion. As dipeptides become less similar to IV, weaker activation takes place, suggesting that specific docking occurs for the two-residue IV knob (63).

Calcineurin On/Off Switch

Calcium–calmodulin (CaM) binding activates CaN, a brain-abundant S,T phosphatase (64, 65). CaN's A-subunit contains the active site. CaN's B-subunit binds to an extension of the A-subunit (**Figure 2**) (66). CaN plays key signaling roles in multiple cell types by removing a phosphate from the nuclear factor of activated T cells (NFAT) that, in T cells, then proceeds into the nucleus and turns on cell proliferation (67).

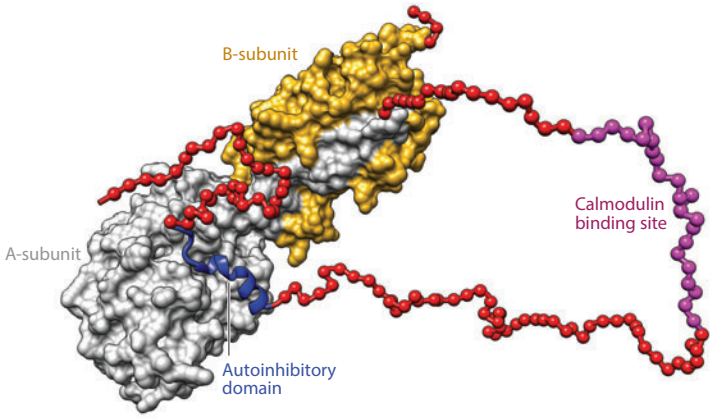

Figure 2

Intrinsic disorder, order, and function of calcineurin. The A-subunit (*gray surface*), B-subunit (*orange surface*), and autoinhibitory domain (*blue ribbon*) are rendered from an experimentally determined structure (Protein Data Bank identifier 1AUI). Missing electron-density regions, including the long disordered tail of the A-subunit, are also represented (*red strand with one ball per residue*), as are the residues corresponding to the calmodulin binding site (*lavender strand and balls*). Disordered regions are modeled to scale with a random conformation, without consideration of additional experimental data.

In vitro, CaN's phosphatase can be activated by trypsin digestion. CaM binding protects against trypsin digestion. Many other CaM-activated enzymes are also activated by trypsin and protected by CaM binding (68).

Upon binding, CaM completely surrounds its helical target (69), so CaM's binding site must be extremely accessible. CaM binding sites and flanking regions have IDP-like amino acid compositions. Their protease sensitivity, extreme accessibility, and amino acid composition all argue that CaM binding sites are located in IDP regions (68).

In crystal structures of CaN (66), the A-chain's CaM binding site is located in a region of missing electron density that is between the B-subunit binding site and the autoinhibitory domain (AID) bound to the active site (**Figure 2**). Both the AID and the binding site for CaM have sequences that exhibit amphipathic helices when folded, a finding that is consistent with the coil-to-helix transitions upon binding that have been observed for both segments. Hydrogen exchange demonstrates lack of protection for the CaM binding region

and the development of protection following addition of CaM, providing additional support for the IDP nature of the CaM binding site (70).

The flexible IDP tether localizes the AID in the vicinity of the active site, thereby favoring AID binding. Unlike trypsinogen and fibrinogen, protease digestion is not needed to expose a knoblike binding site. Instead, after binding, the AID binding motif is flanked by disordered regions on both sides (**Figure 2**) (66). Binding by CaM to its target causes AID displacement, thereby activating CaN's phosphatase.

CaN plays a key role in the rejection of transplanted organs. For T cells, following the binding of a foreign antigen to a cell-surface receptor, calcium levels rise, calcium binds to CaM, and the calcium–CaM complex binds to CaN's IDP region, activating CaN's phosphatase. The activated CaN removes a phosphate from the NFAT protein, which then translocates into the nucleus and turns on cell division, ultimately causing rejection of the transplant (67). The FK506/FK506-BP complex prevents organ-transplant rejection by binding next to CaN's active site and keeping it turned off by steric hindrance (66).

INCREASING ATTENTION

Interest in IDPs (**Figure 1**) began to increase in the late 1990s, following both the first NMR studies and the earliest computational biology experiments on IDPs. We speculate that the complementary findings of these two approaches played key roles in triggering the relatively recent recognition of the importance of IDPs and IDP regions.

Nuclear Magnetic Resonance

NMR is a versatile spectroscopy method for studying proteins (71) that, importantly, does not require crystallization and is well suited to the study of IDPs (34). For a structured protein, the individual resonance peaks tend to be spread out (e.g., there is a large chemical shift dispersion) because the local magnetic environments of the nuclei differ considerably. However, in an IDP the various peaks tend to be more closely spaced or even overlapping (e.g., there is a lack of chemical shift dispersion) because the local magnetic environments of the nuclei are indistinct (71).

Chemical shift dispersion provided critical information about the human cell-cycle control protein $p21^{Waf1/Cip1/Sdi1}$ and its interaction with a protein complex containing a cyclin bound to a cyclin-dependent protein kinase (Cdk) (60). Lack of chemical shift dispersion in the unbound state followed by increased dispersion upon binding indicated that $p21^{Waf1/Cip1/Sdi1}$ is an IDP when free and that it undergoes a disorder-to-order transition upon binding to the cyclin/Cdk complex. This conformational disorder allows $p21^{Waf1/Cip1/Sdi1}$ to bind to multiple partners, as indicated by its biological roles (60). Similar chemical shift dispersion experiments have shown that FlgM, an *Escherichia coli* protein involved in transcription regulation, is an IDP when alone but that its C-terminal half becomes structured upon binding to and inhibiting the transcriptional promoter σ28 (72).

NMR spectral data from IDP ensembles have provided conformational constraints. These constraints form the basis of two approaches to describing IDP ensembles. The first approach uses NMR-constrained molecular dynamics (MD) simulations over multiple copies of the protein (known as replicate exchange MD) (73). The second approach uses a broad sampling of possible structures, then applies the experimental data to select among them (33, 34). Regardless of the method used, agreement with NMR data does not ensure that the result is valid, given the many degrees of freedom compared with the paucity of constraints (34).

Structure determination by NMR involves fitting computationally generated structural models to collections of dihedral angle constraints obtained from through-bond H–H correlated spectroscopy and collections of through-space interatomic distances determined from nuclear Overhauser effect (NOE) spectroscopy (71). Insufficient constraints lead to underdetermined structures, providing

multiple structural possibilities. If pairs of structural possibilities are compared, structured regions typically provide small root-mean-square deviation (RMSD) values, whereas IDP regions provide large RMSD values. Also, structured regions sometimes provide large RMSD values due to insufficient data (71).

The NMR-determined three-dimensional (3D) structure of the DNA binding region of the Antennapedia homeodomain protein has a helix-turn-helix DNA binding motif. Associated experiments have shown that a six-residue IDP region undergoes a disorder-to-order transition following binding to the minor groove of DNA (74).

A recent approach to the identification of IDP regions from NMR-determined structures used an algorithm involving both backbone torsional angles and RMSD values to assign IDP regions to NMR structures (75). Another recent technique compared NMR-determined structures with X-ray-determined structures to identify the RMSD values that provided the greatest agreement between regions of missing electron density in the X-ray structures and regions with high RMSD values in the NMR structures (76).

NMR spectroscopy also has protocols that reveal motional information. One approach involves labeling the protein with ^{15}N and then measuring the NOE values arising from interactions between peptide ^{1}H and ^{15}N atoms. These ^{1}H–^{15}N heteronuclear NOE values exhibit small or even negative values for more rapidly tumbling peptides but strongly positive values for more slowly tumbling structured proteins and domains (77). Once assignments are made, plots of ^{1}H–^{15}N values versus residue numbers reveal structured and IDP regions in the same protein (78).

By adding isotopes such as ^{13}C or ^{15}N coupled with induction of expression of a specific protein, one can use NMR methods to investigate the structure and behavior of a protein residing inside a cell. Various isotopic labeling strategies can be used to explore different aspects of the structure. An alternative approach is to inject large cells (such as oocytes) with isotopically labeled proteins. Finally, isotopically labeled proteins can gain entry into tissue culture cells via penetration signals, such as polyR, or by addition of pore-forming toxins (79).

These in-cell NMR experiments involve significant perturbation, so they are not strictly in vivo experiments (79). Nevertheless, they have provided direct evidence that the extreme crowding and confinement inside cells do not cause the IDPs tested so far to fold into structure (79–82). Although IDPs provide high-quality in-cell NMR spectra, structured proteins often do not, evidently due to NMR line broadening that arises from significantly hindered motions (83). Under some conditions, proteins leak from cells; these leaked proteins provide strong NMR signals, so appropriate control experiments are crucial (84).

Computational Biology

Computational studies of IDPs began with investigations into why they do not fold. IDPs and IDP regions do not fold primarily because they are rich in polar residues and proline and depleted in hydrophobic residues (85–87). Additional factors can override sequence tendencies and induce an IDP region to become structured or a structured region to become an IDP.

Individual-residue computational biases (88) are determined by comparing residue compositions from sets of IDPs (**http://www.DisProt.org**) with compositions from representative sets of structured proteins (89). Through the use of previously published data (87, 88) and with structured protein amino acid compositions as the baseline, the enrichments and depletions of residues in IDPs and IDP regions can be compared with the enrichments and depletions of surface residues (**Figure 3**). With some exceptions, the hydrophobic residues are mostly depleted in both surface residues and IDPs, and the hydrophilic residues and proline are mostly enriched in both groups. The exceptions to these trends need to be studied further.

Overall, these data suggest that P, E, S, Q, and K are disorder-promoting amino acids; that C, W, I, Y, F, L, M, H, and N are

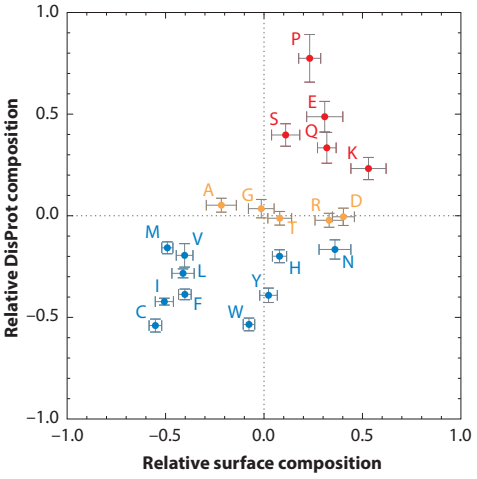

Figure 3

Comparison between the relative residue composition of the surfaces of ordered proteins and the residue composition of disordered proteins. Surface residues are taken from solvent-exposed residues in the structures of monomers (88). Disordered residues are taken from the Database of Protein Disorder (**http://www.DisProt.org**). Both compositions relate to those of a representative set of structured proteins (89) according to $(C_X - C_{Ref})/C_{Ref}$, where C_X is the composition of a residue type in the surface or disordered set and C_{Ref} is the composition of the same residue type in the structured protein set.

structure-promoting amino acids; and that A, G, T, R, and D are amino acids that are indifferent to disorder or structure. Thus, the compositional balance among these amino acids may determine whether a protein or region folds or whether it remains an IDP or IDP region.

Prediction of intrinsic disorder. Amino acid composition differences between structured proteins and IDPs (85) encouraged the development of disorder prediction algorithms (90–93). Two of the four seminal algorithms were based on machine learning methods (90, 91), and two were based on biophysical models of protein folding (92, 93).

Factors that were expected to affect protein folding, such as net charge, hydrophobicity, aromatic content, polarity, and sequence complexity, were tested to determine which had the greatest effects, either positive or negative, on structure formation (85). The most effective attributes were then combined using neural networks (90, 91). Additional refinements (94, 95) yielded a predictor of natural disordered regions (PONDR®). The merging of one predictor trained on variously characterized, long regions of internal disorder (VL) with another trained on X-ray-characterized, disordered termini (XT) resulted in PONDR VL-XT (95). Prediction accuracy exceeded expectations from chance (86, 90), supporting the concept that the sequence (or composition) contains information about whether a protein is folded or disordered. Another advance was the recognition that disorder compositional biases depend on sequence length (96). By now, many machine learning disorder predictors have been developed (97, 98). Investigators have recently advanced our understanding of the underlying similarities and differences among IDP predictors (99).

The second machine learning approach applied the same inputs and neural networks, but trained on data from the CaN family of sequences (91). Such homology-based predictors are potentially very accurate yet are underutilized.

The first biophysical model–based predictor used the concept that repulsion from net charge impedes folding and increased hydrophobicity promotes folding. This model was implemented as a plot of net charge versus hydropathy (92). A straight line on the charge–hydropathy (C–H) plot separates structured proteins from IDPs. The accuracy of the whole-protein C–H plot predictor was recently improved through the identification of better-performing hydropathy scales (100). The original predictor was modified to provide a per-residue predictor (101).

The second biophysical model–based predictor approximated a protein's folding energy using two inputs: (*a*) amino acid pairwise interaction energies and (*b*) a protein's amino acid composition (93). These composition-based folding energy estimates provide values that correlate with estimates calculated from 3D

structures. The composition-based folding energy differences between structured and IDP proteins provide the basis for this predictor of disorder, known as IUPred (93). This predictor has the advantage of being based directly on estimates of folding energy, and it is readily available (102).

Disorder prediction became part of the Critical Assessment of (Protein) Structure Prediction (CASP), beginning with CASP5 (2002), followed by published evaluations (103, 104); links to several of these predictors appear in the Database of Protein Disorder (DisProt; see **http://www.DisProt.org**). The CASP experiment provides unbiased comparisons between the IDP predictors, but the small sizes of the CASP data sets are a limitation.

Current training and testing sets contain many more structured residues than disordered residues, making performance evaluation a complex issue. One approach is to use balanced accuracy (ACC_{bal}), which is given by

$$ACC_{bal} = [(\%\ correct_{order}) + (\%\ correct_{disorder})]/2 \quad 1.$$

Estimates of ACC_{bal} range from 75% to 83% in various CASP experiments. ACC_{bal} has its own limitations, so other performance metrics are also used (103, 104).

Applications of disorder predictors. Many experiments with disorder predictors have been performed. Disorder predictors indicate that sequence databases contain far more IDPs and IDP regions than do structured protein databases (105), that eukaryotes have significantly more predicted disorder than do prokaryotes and archaea (106–108), that disease-associated proteins are rich in predicted disorder (109–112), and that several but not all types of PTMs occur in IDPs and IDP regions more often than in structured protein regions (20, 113, 114).

In a recent study, overall percentages of amino acids that were predicted to be disordered for various archaea proteomes (excluding the halophiles) ranged from ~12% to ~24%, the bulk of the common bacteria proteomes ranged from ~15% to ~30%, and the eukarya proteomes ranged from ~33% to ~50% (108). The halophiles are excluded here because their proteins have adapted to a high intracellular salt concentration by having a reduced hydrophobicity and a high negative surface charge (115), properties that would cause many false-positive predictions of disorder.

Depending on which predictor is used, the human proteome is estimated to have between ~35% (116) and ~50% (117) disordered residues. Another human proteome study focused on the modularity and distribution of predicted IDP regions; several findings indicated that IDP regions in the human proteome should be treated as functional units (118).

The eukaryotic nucleus is very rich in predicted disorder (107, 118). Indeed, eukaryotic transcription factors have been long known to contain nonfolding "acid blob" and "negative noodle" sequences (119). Eukaryotic transcription factors are extremely rich in predicted disorder (120); structured domains identified by homology were found to be largely separated from segments predicted to be IDP regions (121).

FURTHER CHARACTERIZATION OF INTRINSICALLY DISORDERED PROTEINS

A recent two-volume book contains 60 chapters describing biophysical studies of IDPs and IDP regions (122, 123). The blind-men-and-the-elephant parable from India tells us that more can be learned if two or more experimental methods are applied to the same IDP molecule. Examples of such studies are discussed here.

Disorder Prediction and Proteolysis

One of the earliest methods for studying protein stability was sensitivity to proteolysis (42). Accounting for digestion of both structured proteins and known IDPs as controls improves confidence in the results from using proteolysis

to identify IDP regions, and the use of multiple proteases with different specificities is also an advantage (124).

Comparing digestion results with disorder prediction allows each method to confirm the results of the other. Such combinations have been applied to the DNA repair enzyme XPA (125), the transcription factor Ubx (126), and a collection of proteins from the Protein Structure Initiative (PSI). Investigators found that structure determination successes and failures for PSI targets correlate with the absence and presence of IDP regions, respectively (127).

Chemical Modification and Electron Paramagnetic Resonance

Electron paramagnetic resonance (EPR) spectra arising from an unpaired electron in a spin-label moiety provides information about the localized motions on the millisecond (or faster) timescale (128). In studies of protein mobility by EPR spectroscopy, a cysteine is incorporated into the region of interest by site-directed mutagenesis, followed by covalent linkage of the spin-label moiety to the S–H group (128). With such labeling, EPR spectroscopy has shown that rhodopsin's C-terminal region, which is important for signaling, is highly mobile and becomes much less mobile upon antibody binding (129). Similarly, EPR spectroscopy has been used to characterize the disorder-to-order transition when the C-terminal IDP region of measles nucleoprotein binds to the measles virus phosphoprotein (130).

X-Ray Diffraction and Nuclear Magnetic Resonance

Several studies that compared the structured regions of proteins characterized by both X-ray diffraction and NMR have been performed. In contrast, too few experiments combining these two methods have focused on IDP regions.

In a combined X-ray–NMR study of Bcl-XL, NMR structure determination indicated that a functionally important 55-residue region of missing electron density in the X-ray structure was significantly unfolded. Furthermore, the ^{1}H–^{15}N heteronuclear NOE spectra provided negative values for the region with missing coordinates and strongly positive peaks for the structured regions. These NMR experiments showed that the region of missing density in the X-ray structure is both unstructured and highly mobile (78). Thus, the region of missing electron density in the X-ray structure of this protein does not arise from a mobile structured domain, which is always a possibility for large regions of missing coordinates in X-ray structures.

X-Ray Crystallography and Small-Angle X-Ray Scattering

The combination of protein crystallography and small-angle X-ray scattering (SAXS) provides insight into the shape adopted by the region of missing electron density (131). Recent advances have enabled more precise analyses of SAXS data for these purposes (132). Applications of these advanced methods to the human glycosylase NEIL1 have provided mechanistic insight into how NEIL1's flexible, disordered ~100-residue C-terminal domain becomes involved in multiple functional interactions (133).

Nuclear Magnetic Resonance and Small-Angle X-Ray Scattering

NMR provides highly localized structural information. Thus, IDP configurational ensembles developed by fitting sparse NMR data provide uncertain overall shape estimates. However, SAXS provides good shape estimates. Thus, the combination of data from both methods is advancing our understanding of the structures and motions of IDP ensembles (131, 134).

Characterization with Multiple Methods

Similar to many other scaffold and docking proteins (135, 136), AXIN is predicted to be mostly disordered (135), a hypothesis supported by multiple experiments (137). The G295–A500

region has multiple mutations associated with the development of cancer and has known binding sites to glycogen synthase kinase 3β (GSK3β) and β-catenin (138).

The G295–A500 region has been probed by multiple methods. A smooth decrease in tryptophan fluorescence was observed over the temperature range between 20°C and 80°C, whereas a globular protein domain showed a lateral fluorescence shift at ~60°C due to a thermal unfolding event, indicating the absence of a thermal unfolding for the G295–A500 segment over this temperature range. The CD spectrum indicated a random-coil structure. Size-exclusion chromatography provided a too-high molecular weight, indicating an extended shape. The ^{1}H–^{15}N heteronuclear single-quantum coherence NMR spectrum was significantly collapsed and remained almost unchanged with a shift to 6 M urea. Finally, 9 of 12 disorder predictors found this region to be mostly disordered. According to all these measures, the G295–A500 fragment is an IDP region.

In the Wnt pathway, AXIN binds both GSK3β and β-catenin, thereby accelerating the phosphorylation of β-catenin by the kinase. The G295–A500 fragment greatly accelerated this phosphorylation in vitro. Deletion of either binding site eliminates the acceleration. Thus, the G295–A500 segment performed its biological function while remaining completely disordered. The double binding of the kinase and its substrate raises the local concentration of each relative to the other, accelerating the phosphorylation by the mechanism of colocalization (139).

Investigators have suggested the term "stochastic machine" to describe a flexibly linked complex such as AXIN with its binding partners (140). Given the high frequency of disorder in scaffold and anchor proteins (135, 136), stochastic machines are likely to be very common. Binding of two or more partners to a single IDP region has an advantage compared with connection with a flexible linker arising from gene fusion because the separated components can be reutilized in multiple machines. Indeed, AXIN is the scaffold for at least three different pathways (140).

Both substrate-binding and ubiquitin-chain elongation are performed by a multiprotein E3 ligase that uses highly mobile parts based on IDP regions (141). This complex is a good candidate for another stochastic machine.

BIOLOGICAL FUNCTIONS

The various functions carried out by IDPs call for a fundamental reassessment of the protein structure–function paradigm (5). In this section, we discuss various methods used to identify IDP-associated functions. Overall, IDP functions complement those of structured proteins (142–144).

Computational Approaches

Because disorder can be predicted from sequence with fairly good accuracy, and because the large databases of amino acid sequences also contain functional information about these proteins, it seems reasonable to use computational approaches to search for potential relationships between function and disorder. Such computational studies examined the order and disorder tendencies of collections of proteins with the same Gene Ontology descriptors (107, 145–147). As had been shown for fibrinogen, trypsinogen, and CaN, these studies indicated that, on a very large scale, the functions of IDPs and IDP regions are heavily biased toward signaling, regulation, and control. However, the functions of structured proteins seem to be related mostly to catalysis (e.g., enzymes such as lysozyme); the controlled binding and release of small-molecule ligands (e.g., the binding and release of oxygen by myoglobin); and the movement of electrons, ions, or molecules across membranes (e.g., proton translocation by bacteriorhodopsin).

Regulation by disorder versus catalysis by structure is an imperfect partition. Many enzymes contain flexible, disordered loops or tails that fold onto the substrates when they bind, thereby contributing to catalysis

by helping to exclude water and sometimes by providing catalytic residues (148, 149). Also, enzymes often contribute to signaling and regulation, especially by PTMs such as phosphorylation, acetylation, or proteolysis. In several recent experiments, computational biologists have attempted to predict function from amino acid sequence (150). These experiments typically used amino acid sequence and 3D structure to provide the basis for inputs for the prediction of Gene Ontology annotations. In some of these experiments, the inclusion of disorder prediction combined with the various disorder-associated linear motifs (discussed in more detail below) proved useful for function prediction (151).

Manual Annotation

Although computational approaches have revealed trends and formed the basis for predicting functions of proteins from sequence, manually curated databases provide useful repositories. DisProt (152) currently contains manual annotations for 694 proteins with 1,539 experimentally characterized regions of disorder, almost all of which have identified biological functions. DisProt lists 39 different functions that have been associated with IDP regions, and many of them involve interconversions between order and disorder. Examples of these functions include flexible linkers, many types of PTMs, regulation of apoptosis, regulation of proteolysis, transport through narrow pores, nuclear localization signals, and binding to a wide variety of partners.

Databases

Five databases focus on different aspects of IDPs. These databases are the abovementioned DisProt, the Intrinsically Disordered Proteins with Extensive Annotations and Literature (IDEAL) (153; see **http://idp1.force.cs.is.nagoya-u.ac.jp/IDEAL/**), the Database of Protein Disorder and Mobility Annotations (MoBiDB) (154; see **http://MoBiDB.bio.unipd.it**), the Database of Disordered Protein Prediction (D^2P^2) (117; see **http://d2p2.pro/**), and the Protein Ensemble Database (pE-DB) (134; see **http://pedb.vib.be**).

IDEAL contains a link that enables predictions of structure and disorder (116) as well as indications of IDP involvement in protein–protein interaction (PPI) networks. MoBiDB includes IDP estimates from NMR structures as well as the outputs from several disorder predictors. Currently, MoBiDB lists 26,993 proteins with experimental annotations and 4,662,776 proteins with either experimental or prediction annotations. D^2P^2 contains data from nine different disorder predictors applied to >10 million sequences from >1,700 complete genomes representing >1,200 distinct species; it also presents applications of predictions of structured domains using hidden Markov models developed from SCOP SUPERFAMILY structures. Finally, pE-DB contains a collection of 39 IDP ensembles with a total of 3,973 structures. The various ensembles were developed by eight different research groups using different methodologies.

These databases provide experimenters with important sources of information against which to test the novelty of their latest IDP findings. Comparing information across these databases will yield new insights into IDP structure and function.

Solubility Enhancement

An analysis of PSI data showed that having predicted IDP regions increases the chances of soluble expression and purification but decreases the chances of successful structure determination. These results suggest that IDP regions generally help make proteins more soluble, thereby improving expression and purification, but then diminish the probability of crystallization and structure determination (54). Solubility enhancement is expected from polar and charged IDPs, but not necessarily from all types of IDPs, especially if they contain local regions with a tendency to aggregate. A direct indication that IDP regions can enhance solubility

has come from the increased solubility that results from the fusion of insoluble proteins with either natural IDPs (155, 156) or specially designed artificial IDP sequences (156).

Regulation of Protein Lifetimes

An oft-stated objection to the existence of IDPs and IDP regions in vivo is that they would rapidly degrade inside cells (157). Indeed, many IDP regions are highly sensitive to proteolytic digestion in vitro, but such sensitivity is not a compelling argument against IDP existence in vivo.

First, in vivo proteolysis is highly regulated. For example, for proteasome-mediated digestion most proteins are first ubiquitinated, which requires an exposed hydrophobic patch that is not common in IDPs (158). When present, such patches may be protected by chaperone-like bodyguard (143) or nanny (159) proteins, whose suggested purpose is protection from proteolysis, aggregation, and incorrect partner binding, rather than promotion of folding. Second, many disordered regions are involved in PPIs, either transiently or in stable complexes. Hydrophobic patches promote such interactions (see the next section), reducing an IDP's proteolytic sensitivity in vivo. Third, proteolytic digestion in vivo is an important regulatory mechanism. Some short-half-life proteins contain certain motifs, such as KEN-box, destruction-box, or PEST motifs (158). These motifs are associated with regions of disorder (158, 160, 161); therefore, rather than representing an argument against IDP existence in vivo, many IDP regions are probably important components of proteolytic regulatory mechanisms, just as for trypsinogen and fibrinogen activation (discussed above).

Protein–Protein Interactions

Several IDPs used to train the original disorder predictor contained short predictions of structure that matched binding sites to protein partners (162). The PONDR VL-XT in particular shows dips that relate to binding sites and are associated with local hydrophobic patches. We refer to these binding site dips as MoRFs (163).

As for association between globular proteins, burial of hydrophobic groups is important for the binding of IDPs to their protein partners (164). Very recent studies suggest that, especially for interactions involving disordered domains (discussed further below), polar interactions are also very important (165).

Figure 4*a* shows an original training-set example MoRF, 4E-BP1 (162), along with three additional examples of IDP regions involved in PPIs. These additional examples include two in which each IDP region uses three separate MoRFs to bind to one partner, namely phosphatase inhibitor II (**Figure 4*b***) and calpastatin (**Figure 4*c***), and one in which a long IDP wraps around its dimeric partner, namely $p27^{kip1}$ (**Figure 4*d***), thereby forming a disordered domain. The bound 4E-BP1 may contain additional regions that bind to its partner with multiple conformations (166), thereby forming a fuzzy complex. Disordered domains and fuzzy complexes are further discussed in below subsections.

Some protein segments map to regions of missing electron density in certain crystals but are structured in other crystals. These protein segments are referred to as ambiguous (167) or dual-personality (DP) (168) segments. DP segments have amino acid compositions that are intermediate between those of structure and disorder (168) and similar to those of helix-forming MoRFs plus their nearby flanking regions (169). Recent studies on semidisorder (41) are probably dealing with regions similar to these.

MoRFs that form irregular structure upon binding typically do not show short structure predictions with PONDR VL-XT. Recently developed predictors that identify all types of binding segments within IDP regions are useful (170, 171).

An alternative approach is to identify binding-associated sequence patterns known as ScanSite motifs (172), eukaryotic linear motifs (173), short linear motifs (174), or minimotifs (175). The motifs identified by these various

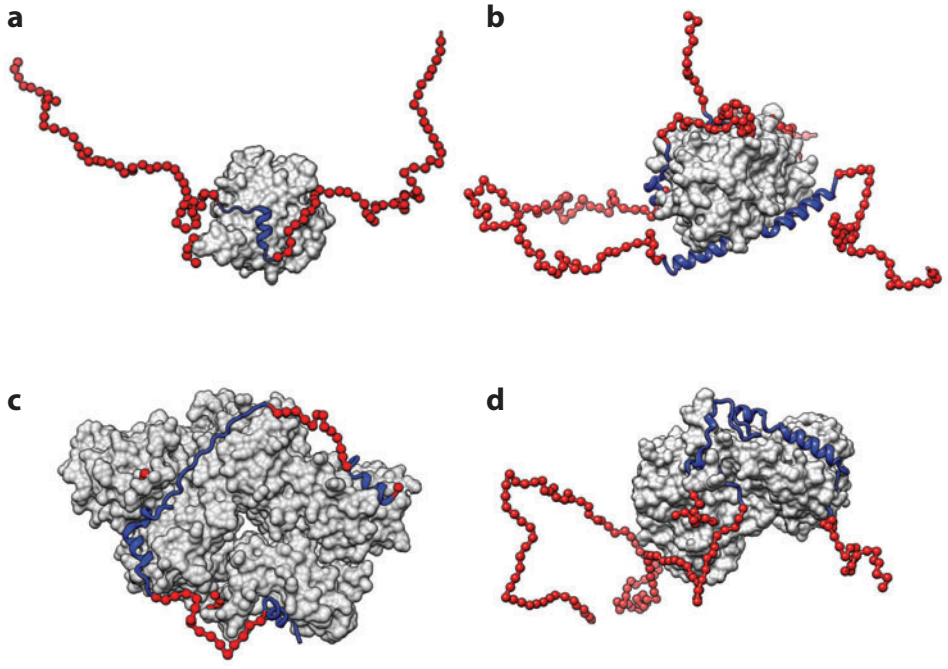

Figure 4

Examples of intrinsically disordered domains bound to partners. In all of these structures, a portion of the intrinsically disordered domains becomes structured upon binding (*blue ribbons*), whereas other portions remain disordered (*red strands with one ball per residue*). Disordered regions are modeled to scale with a random conformation, without consideration of additional experimental data. (*a*) The completely disordered protein 4E-BP1 binds to eukaryotic initiation factor 4E (eIF4E) through a few central residues [Protein Data Bank (PDB) identifier 1EJ4]. Additional evidence suggests that a significant fraction of the IDP regions form fuzzy complexes with eIF4E (166). (*b*) Protein phosphatase inhibitor 2 binds to type 1 protein phosphatase by using three discrete regions connected to and flanked by disordered regions (PDB 2O8A). (*c*) Calpastatin binds to calpain by using three discrete regions connected by disordered regions (PDB 3DF0). (*d*) $p27^{kip1}$ binds to the CDK2/cyclin A heterodimer, whereas the termini remain disordered (PDB 1JSU). Such large regions are termed disordered domains (203).

methods are located mostly in IDP regions (176), except for minimotifs, which map to both structured and disordered regions (177).

Given the crystal structure of a protein complex, a plot of monomer surface area versus buried surface area separates complexes that arose from two globular, structured proteins from complexes that arose from two IDPs. The former lie close to the origin of the plot, and the latter are located much farther from the origin (178). A modification of this plot separates hetero- and homodimer complexes into three types. In the first type, both monomers are structured before association. In the second type, one is structured and one is an IDP before association. In the third type, both monomers are IDPs before association (179). This modified approach was recently applied to the ribosome, and the results indicate that most ribosomal proteins have IDP regions that become structured upon binding (180)

The association between a structured protein and a flexible IDP or IDP region may involve preformed elements (181) or, alternatively, conformational selection, a concept first proposed in 1974 for binding to a flexible ligand (182). An alternative mechanism is known as induced fit (183) or coupled binding and

folding (184), wherein structure formation occurs concomitantly with binding. Any given association might involve a mixture of conformational selection and induced fit (185). Characterization of the mechanistic details of individual IDP-based PPIs is an active area of research (165, 186–193) that is expected to continue indefinitely.

Protein–protein interaction networks. Eukaryotic PPI networks contain hubs that bind to many partners, as well as other proteins that bind to only a few. Hubs may require a new principle for protein molecular recognition (194). An IDP is flexible, so it can bind to multiple partners (60, 142). Thus, IDP-based interactions have been proposed as the new principle to explain hub-associated binding to multiple partners (195). Multiple partner binding may involve two mechanisms. In the first, one region of disorder binds to many different partners (one-to-many binding). In the second, many different regions of disorder bind to one structured partner (many-to-one binding) (195).

One-to-many binding. An IDP region can bind to many partners by having one binding site after another (196). Alternatively, one IDP region can change its shape and thereby bind to many different partners (61, 144, 179). The interactions between p53 and its partners illustrate both of these IDP-based multiple partner binding mechanisms (**Figure 5**). These multiple partner capacities are further enhanced by PTMs and by AS (61), both of which frequently occur in IDPs and IDP regions (18, 20).

These structures (**Figure 5**) represent only a small fraction of the known protein interactions involving p53's IDP regions. The N-terminal IDP region probably binds to more than 40 different partners (C. Anderson, personal communication), and the C-terminal IDP region binds to an even larger number. As we determine more structures, the complexity of the overlap for the various binding sites will increase further. Also, with so many partners for each IDP region, it is important to consider not only that a single disordered region can bind to multiple partners, but that these multiple partners compete for binding to the same IDP region. This competition certainly has important biological consequences.

Many-to-one binding. Alternatively, many different IDPs or IDP regions of differing sequence can use their flexibility to bind to a single structured partner. The different MoRFs can bind to essentially the same binding site, to an overlapping but distinct binding site, or to completely different binding sites on the same protein (179, 197). When the sites are completely different, the binding regions can arise from the same protein. The PDB contains examples in which a single IDP has two or even three separate MoRF regions bound to the same partner with IDP linkers (**Figure 5**).

Tissue-specific rewiring of protein–protein interaction networks. Tissue-specific AS reveals that, in many specific examples, one AS isoform bears a MoRF-containing IDP region and another isoform in another tissue lacks the MoRF-containing IDP region. Thus, tissue-specific AS in messenger RNA (mRNA) regions that code for IDP regions can cause rewiring of the overall PPI network (198–200).

Mutual folding. Interaction between two IDPs can lead to mutual synergistic folding. The cyclic AMP (cAMP) response element–binding protein (CREB) interacts with the CREB-binding protein (CBP). CBP contains a molten globular IDP region that interacts with a fully unstructured segment from the p160 steroid coactivator ACTR. Upon interaction, these two IDPs mutually fold to yield an intertwined structured domain containing six helical segments, three from each IDP (14, 201).

Disordered domains. A structured protein domain evolves, functions, and exists independently of the rest of the protein chain (202). Some IDP regions longer than ~25 residues likewise evolve, function, and exist independently of the remainder of the protein chains, so such regions have been termed disordered domains (203). A common function

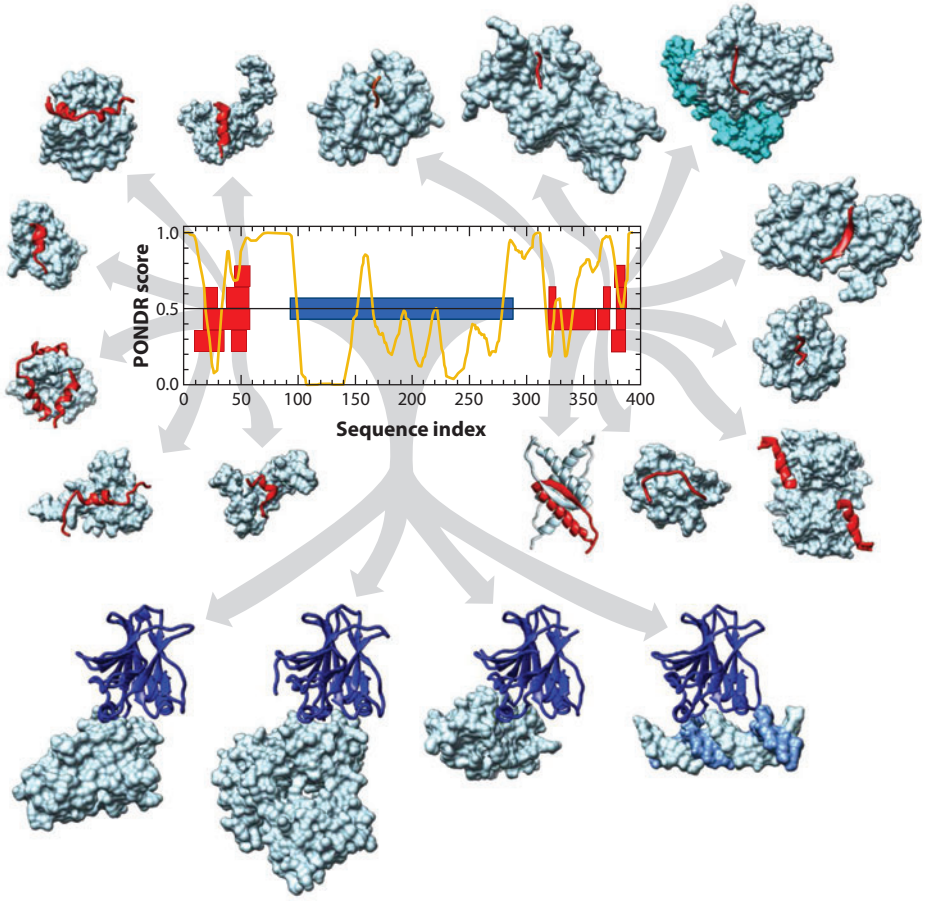

Figure 5

Intrinsic disorder and molecular interactions of the tumor suppressor p53. For the PONDR VL-XT (predictor of natural disordered regions, with one predictor trained on variously characterized, long regions of internal disorder and another trained on X-ray-characterized, disordered termini) prediction of intrinsic disorder for p53 (*center*), values above or below the 0.5 threshold indicate predictions of disorder or order, respectively. The prediction indicates that the N and C termini are largely disordered, whereas the central DNA binding domain is ordered. Also shown are the structures of several discrete regions of p53 (*ribbons*) that have been determined in complexes with partners (*surfaces*); the corresponding horizontal bars indicate the region of p53 that participates in each structure. Five partners of the N terminus (*clockwise from lower right*) are high-mobility group protein B1 [Protein Data Bank (PDB) identifier 2LY4], Taz2 domain of p300 (PDB 2K8F), nuclear coactivator–binding domain of p300 (PDB 2L14), MDM2 (PDB 1YCR), N terminus of replication protein A (PDB 2B3G), and PH domain of RNA polymerase II transcription factor B subunit 1 (PDB 2GS0). Eight partners of the C terminus (*clockwise from upper left*) are histone acetyltransferase domain of *Tetrahymena* general control nonderepressor 5 (PDB 1Q2D), SET9 (PDB 1XQH), CDK2/cyclin A (PDB 1H26), Sir2 (PDB 1MA3), bromodomain of CBP (PDB 1JSP), S100B(ββ) (PDB 1DT7), Tudor2 domain of PHF20 (PDB 2LDM), and p53 homotetramerization (PDB 3SAK). Four partners of the central DNA binding domain (*left to right*) are 53BP2 (PDB 1YCS), large T antigen of simian virus 40 (PDB 2H1L), BRCT domain of 53BP1 (PDB 1GZH), and DNA (PDB 1TSR).

of disordered domains is that they associate with and thereby regulate the activities of globular protein partners.

Researchers have devoted much effort to grouping proteins into homologous protein families. One such attempt, Pfam (204), uses hidden Markov models to find and group members of the same family by sequence pattern matching. For multicellular eukaryotes, many of these Pfam domains are IDPs or contain IDP regions (203, 205). Many of these IDP Pfams fit the characteristics of disordered domains.

These disordered domains are characterized by long, complex interfaces between the IDP region and its globular binding partner. Given the lengths of such interactions and the flexibilities of such IDP regions, the association and dissociation of these complexes can involve multiple localized steps, each regulated by separate signals, thereby enabling signal integration. For example, the protein $p27^{kip1}$ regulates the progression from G_1 into S phase. A disordered domain region of $p27^{kip1}$ causes cell-cycle arrest by binding to and shutting off the cyclin A/Cdk2 complex (**Figure 4*d***). Dissociation of $p27^{kip1}$ from this complex causes progression into S phase. This dissociation involves a signal conduit that includes both T and Y phosphorylations, ubiquitination, and proteasome digestion (206) in a series of steps that integrate signals from different pathways by using the advantages of IDP regions (207). Follow-up studies have suggested that IDPs are generally important in cell-cycle regulation (208).

Fuzzy complexes. Upon binding, many MoRFs contain regions, either internal or flanking, that remain disordered (209). Surprisingly, the nearby IDP region can affect the binding constant of the interaction between the MoRF and its partner, even though the IDP region fails to form structure upon binding. Complexes with IDP regions that affect binding are described as fuzzy (210). An interesting open question is whether an interaction can be fuzzy over the entire interaction surface or whether it requires at least some local region of structure (31).

The binding of the IDP protein 4E-BP1 to eukaryotic elongation factor 4E (eIF4E) involves a short region that becomes well structured (**Figure 4*a***) (211) and an additional region that is indicated to bind by NMR. The overall shape of the complex, as determined by fitting the X-ray structure to SAXS data, shows where the remaining regions of disorder project beyond the structured eIF4E surface and how the fuzzy binding region is probably organized (166).

Another interesting fuzzy complex forms between the IDP Sic1 and the cell division control protein 4 (Cdc4) in budding yeast. Phosphorylation of any six (or more) of nine suboptimal sites on Sic1 leads to tight binding of Cdc4 to a WD40 domain, followed by ubiquitination and degradation of Sic1, promoting the onset of DNA replication (212). Although multiple phosphorylations increase the apparent binding constant, the interaction is not polyvalent; rather, the WD40 domain has only one binding site. An electrostatic model shows that electrostatic interactions between Cdc4 and Sic1 lead to a fuzzy, non-structure-forming interaction that raises the apparent binding constant (213). Fuzzy interactions essentially maintain a high local concentration [e.g., colocalization (139)] of phosphorylated Sic1 sites, each of which can bind to one side on the WD40 domain of Cdc4, so as one hops off, another one rapidly hops on.

Binding affinities and kinetics. On the basis of a concept originally proposed in 1979 by Schulz (214) for protein–nucleic acid interactions, the energy needed to fold an IDP must be taken away from the interaction energy, which results in high specificity and low affinity. Experiments show that although many IDPs do indeed have weak affinities with their partners, the range of affinities for complexes involving IDPs strongly overlaps the range for complexes involving structured proteins; similar overlaps are also observed for on and off rates (190). Interactions involving IDPs can exhibit remarkably fast on rates (191), which probably occur via a fly-casting mechanism (192) or a dock-and-coalesce mechanism (193).

Protein–Nucleic Acid Interactions

In the late 1970s, the X-ray crystal structure of the tobacco mosaic virus (TMV) coat protein revealed a 25-residue, positively charged IDP region. Subsequent X-ray and NMR studies showed that this segment undergoes a disorder-to-order transition upon RNA binding during TMV assembly (16).

Later, in 1994, Spolar & Record (184) performed thermodynamic studies on several protein–DNA interactions. Their careful dissection of the thermodynamic data, along with analyses of known structural changes, led to the suggestion that these interactions often involve coupled binding and folding of proteins upon specific DNA binding.

In 1999, a collection of 75 crystal structures of protein–nucleic acid complexes were analyzed. Most of them contained DNA rather than RNA. Of them, investigators determined 24 crystal structures of the proteins without any nucleic acid, along with the structures of the complexes. For 8 of these 24 complexes, the proteins exhibited regions of disorder when alone, and these regions were structured in their complexes. The authors concluded that such examples of disorder in the proteins before binding were likely to be much more common but were not observed due to inhibition of crystallization by the disordered regions (215).

Protein–DNA interactions. Studies of DNA–protein interactions indicate that IDP regions are affinity tuners of such interactions. Also, via a so-called monkey-bar mechanism, IDP regions increase the overall capability of proteins to efficiently search DNA for specific binding regions (216).

For example, the important *Drosophila* regulatory transcription factor Ubx contains a homeobox (HOX) domain flanked by disordered regions. Ubx's HOX domain alone binds much more tightly than the HOX domain with the flanking IDP regions, a finding that is consistent with affinity tuning. Furthermore, intramolecular binding by an IDP-linked YPWM motif plays an especially important role in weakening Ubx's HOX binding to DNA (126).

Although Ubx has low binding specificity to DNA, interaction with Extradenticle (Exd, or Pbxl in mammals) causes significant gains in both affinity and specificity. These interactions are mediated by Ubx's IDP-linked YPWM motif, which binds into a pocket within Exd's HOX domain (217). Thus, the same IDP-linked YPWM can either weaken or strengthen Ubx binding to DNA, depending on the relative locations of Ubx and Exd's binding sites and the length of the IDP linker. Furthermore, the IDP linker between the Ubx HOX domain and the YPWM motif is encoded by several microexons (126), allowing the use of AS to change the length of this IDP linker; in turn, the use of AS can alter both Ubx's intrinsic DNA binding affinity and its interactions with Exd.

Protein–RNA interactions. Two important multisubunit RNA–protein complexes are the spliceosome (218) and the ribosome (219). The former excises the introns and splices the exons of precursor messenger RNA (pre-mRNA) molecules to yield mature mRNA, and the latter uses the coding instructions from mRNA to specifically link amino acids together to form polypeptide chains. For both assemblies, the catalytic steps are performed by RNA, not protein (218, 219).

Overall, the prokaryotic and eukaryotic ribosomes contain 52 and 79 proteins, respectively (220). Astoundingly, the 3D structure of the ribosome has been determined (219). A notable finding is that a large fraction of ribosomal proteins contain highly extended structures that play critical roles in ribosomal assembly (221). These extensions are also probably important for ribosome stability (219). A recent comprehensive analysis indicates that essentially all of the ribosomal proteins have large regions that are predicted to be IDPs both from their sequence and from their observed structure context (180) using a plot of monomer surface area versus buried surface area (178), as modified for generalized heterodimers (179). Correct ribosomal assembly probably depends on

disorder-to-order transitions of these long IDP regions (221).

Although the ribosome contains well-organized RNA and protein components, the spliceosome exhibits exceptionally dynamic composition and structure. Thus, identification of the key protein components has been a significant challenge (218). A recent analysis suggests that the ~100 proteins associated with the spliceosome are similar to the ribosomal proteins in that they are very rich in IDP regions (222). Thus, both major ribonucleoprotein assemblies in eukaryotes probably utilize IDPs and IDP regions for their assembly and function.

Allosteric Regulation

Several enzymes and binders of small molecules are regulated by effector molecules that bind to sites that are distant from the active site. Two models have been proposed to explain this allosteric regulation: the Monod–Wyman–Changeux (MWC) concerted model (223) and the Koshland–Némethy–Filmer (KNF) sequential model (224). Both models involve multisubunit proteins with two alternative states that differ in activity; one of the states is favored by effector binding. The MWC and KNF models differ in several interesting details with regard to the steps involved in coupling effector binding to changes in activity.

Recently, researchers proposed an alternative model for allosteric regulation in which the two structured states are replaced by an ensemble. In this model, effector binding alters the ensemble characteristics (225). IDP-region involvement in allosteric regulation has recently been identified for several proteins (226–228).

Single-molecule fluorescence energy transfer was recently used to study a mechanism by which an adenovirus hijacks cellular functions by using ternary interactions involving E1A (an adenovirus IDP that becomes a Hub), the retinoblastoma protein (pRB), and CBP (229). This study has shown that the E1A–pRB–CBP interactions can be induced to switch between positive and negative cooperativity, thereby changing whether the resulting complex is one of the two binary complexes or the ternary complex. A context-dependent switching among the various forms can, in turn, alter downstream signaling outputs. This IDP-dependent example of allosteric interactions has very broad implications (229).

Chaperone Function

IDPs and IDP regions have been implicated as chaperones for both RNA and protein folding (230, 231). Even for the mostly folded, multisubunit GROEL/GROES complex, IDP loops may be important for chaperone function (232). Various IDPs exhibit chaperone activity (233, 234).

The oxygen-sensor bacterial protein Hsp33 contains a zinc ligand that binds two cysteine S–H moieties among its ligands. Cysteine oxidation leads to disulfide-bond formation, which in turn triggers a major order-to-disorder transition. This order-to-disorder transition is associated with the development of chaperone activity (40).

The flexibility of IDP regions and their ability to bind to multiple partners may be important components of the mechanisms underlying chaperone activity (8, 231). More research is needed to understand the mechanisms by which IDP regions carry out chaperone activity.

EVOLUTION

Surface amino acids evolve faster than do buried residues (235). Shannon's entropy, calculated from the number of position-by-position changes in the sequence alignments of 130 protein families with 7,143 aligned sequences, increases linearly with the reciprocal of the packing density (236), indicating that protein 3D structure significantly reduces amino acid substitutions over evolutionary time.

Observations of structured proteins predict that, due to their lack of structure, IDP regions should show higher rates of change

compared with the structured regions of the same proteins. This hypothesis was tested in 26 protein families containing at least one experimentally characterized IDP region. As expected, the segments that aligned with the IDP regions showed statistically significant increased mutation rates compared with the segments that aligned with structure (56).

Unexpectedly, 3 of the 26 protein families revealed IDP regions with higher sequence conservation than that of the structured parts (56). Many additional predicted IDPs have high sequence conservation (237, 238). IDP regions with the expected lack of sequence conservation, known as flexible disorder, are associated with signaling pathways and multifunctionality, whereas those with unexpected sequence conservation, known as constrained disorder, are associated with RNA binding and chaperone activity (239). Constrained disorder is also associated with regions of proteins that are involved in PPIs and encoded by mRNA segments subject to tissue-specific AS (240). Both flexible disorder and constrained disorder are enriched in regions of proteins that undergo tissue-specific AS, but not general AS (199).

Replication protein A (RPA) contains an IDP linker that exhibits very high sequence variability, including multiple INDELs (56). Homologous RPA IDP segments from five widely divergent organisms were cloned, expressed, purified, and analyzed by NMR. Despite their sequence differences, the NMR data indicated very similar flexibilities. Thus, RPA's IDP linker exhibits well-conserved flexibility and disorder, whereas its sequence lacks significant conservation (241).

Comparing substitution matrices for structured and disordered proteins with similar degrees of sequence conservation shows that their patterns of amino acid substitution are clearly different (242). For example, substitutions of G for W or Y (or the reverse) are much more common in IDPs, whereas substitutions of N for E (or the reverse) are much more common in structured proteins. Overall, structured proteins favor structurally similar replacements, whereas IDPs and IDP regions show less favoritism. IDPs' greater tendency toward random substitutions becomes clearer from their lower rate of conservation, indicated by the identity elements of the substitution matrices (242).

As discussed above, IDP regions provide binding sites that are important for enabling the complexities observed for PPI networks in eukaryotes (195), especially with multiplexing by AS and PTM (61, 198). Several studies of the evolution of PPI networks suggest that changes in IDP regions—whether by point mutation, INDELs, changes in PTM sites, or changes in AS patterns—play fundamental roles in the evolutionary changes observed in these networks (243–246).

A recent study of p53 evolution found that disorder prediction values and evolutionary rates of change are significantly correlated. When combined with other observations about the evolution of IDP regions, these data suggest that IDP regions are likely to be disordered in vivo (247).

In silico studies suggest that random mutations readily convert IDP regions into structure, indicating that IDP regions require active maintenance (248), yet IDP regions are quite well conserved over evolutionary time (239, 249). Among yeast proteins that are harmful if overexpressed, the strongest determinant is IDP prediction, which suggests that overproduction of IDPs is often harmful (250). Consistent with this finding, IDPs are tightly regulated in yeast (251) and probably in other organisms as well (252). This tight regulation is probably needed, at least in part, because of the ability of IDPs to be involved in multiple processes, also known as moonlighting (253). Thus, an excess or dearth of a given IDP could give rise to unfavorable moonlighting activities or a lack of crucial ones, respectively. Viruses are proposed to hijack cell regulation by rapidly evolving IDP regions containing linear motifs that compete with similar cellular signaling motifs (254). All of these observations suggest that the amounts of IDPs in cells are necessarily maintained at specific levels because of their crucial regulatory activities.

The Myc proteins (c-Myc, MycN, and MycL in humans) are IDPs with so many INDELs and mutations that their sequences are essentially impossible to align. Investigators have recently attempted to construct a phylogenetic tree for this protein on the basis of PONDR VSL2P predictions (255). This tree seems reasonable overall and presents some interesting insights into the history of this protein family. We will learn over time whether this highly novel approach is generally useful.

FUTURE DIRECTIONS

A major shortcoming is that databases of PPIs are constructed as though each protein had a single binding site. For example, by using several PPI databases we identified more than 400 protein partners for the BRCA1 protein. None of the PPI databases indicate the BRCA1 sequence locations for partner binding, nor are the splice variants identified (257).

Tedious literature searches are needed to identify BRCA1's partner binding sites. Fewer than 40 of the 400 partner binding sites were identified, and most of these sites map to BRCA1's long central IDP region (256). Not one 3D structure has been determined for a BRCA1 MoRF, linear motif, or disordered domain bound to a partner.

How can we possibly understand tissue-specific rewiring around the BRCA1-based Hub in the overall PPI network unless we know both the regions of pre-mRNA AS and the sites of partner–MoRF interactions? Precision medicine will remain elusive until these MoRF–partner interaction sites and their 3D structures are determined on a very large scale. Concerted efforts are needed to perform this task.

Since the 1980s, investigators have known that eukaryotic transcription factors are rich in IDP regions (119), yet there have been surprisingly few studies that aimed to determine the roles of disorder in gene regulation, especially with regard to tissue-specific gene regulation. We suggest that, as for PPI networks, researchers will find that AS of pre-mRNA coding for transcription factor IDP regions, coupled with PTMs mapped to IDP regions containing MoRFs for proteins and nucleic acids, act to rewire gene regulation in a tissue-specific manner. The importance of this combination of factors for gene regulation has already become evident (126, 257).

We speculate that the orchestration and modulation of PPIs and protein–nucleic acid interactions involving IDP regions via tissue-specific alternative splicing and PTM provided the molecular basis for the original development of metazoans and that these same activities underlie cellular differentiation. Thus, an important future direction will be to test these hypotheses by performing experiments to identify and understand the various roles of IDP regions in gene regulation by transcription factors and in tissue-specific regulation of PPI networks and pathways.

NOTE ADDED IN PROOF

Matthew John Gage brought to our attention several recent biochemistry textbooks that contain brief discussions of IDPs (258, pp. 161–62; 259, pp. 104–5; 260, pp. 135–55; 261, pp. 60–61; 262, pp. 283–84; 263, pp. vii, 141–43), varying in extent from a couple of paragraphs to a small section. We look forward to seeing more extensive and integrated coverage of IDPs in future textbook editions.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Dunker AK, Babu MM, Barbar E, Blackledge M, Bondos SE, et al. 2013. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disord. Proteins* 1:e24157
2. Jirgensons B. 1966. Classification of proteins according to conformation. *Makromol. Chem.* 91:74–86
3. Arnone A, Bier CJ, Cotton FA, Day VW, Hazen EE Jr, et al. 1971. A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem.* 246:2302–16
4. Uversky VN, Oldfield CJ, Dunker AK. 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18:343–84
5. Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* 293:321–31
6. Uversky VN, Narizhneva NV, Ivanova TV, Tomashevski AY. 1997. Rigidity of human α-fetoprotein tertiary structure is under ligand control. *Biochemistry* 36:13638–45
7. Johnson LN. 1992. Glycogen phosphorylase: control by phosphorylation and allosteric effectors. *FASEB J.* 6:2274–82
8. Reichmann D, Jakob U. 2013. The roles of conditional disorder in redox proteins. *Curr. Opin. Struct. Biol.* 23:436–42
9. Qian YQ, Otting G, Furukubo-Tokunaga K, Affolter M, Gehring WJ, Wüthrich K. 1992. NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the *Drosophila* antennapedia protein. *Proc. Natl. Acad. Sci. USA* 89:10738–42
10. Trombitás K, Greaser M, Labeit S, Jin J-P, Kellermayer M, et al. 1998. Titin extensibility in situ: entropic elasticity of permanently folded and permanently unfolded molecular segments. *J. Cell Biol.* 140:853–59
11. Rauscher S, Pomès R. 2012. Structural disorder and protein elasticity. In *Fuzziness*, ed. M Fuxreiter, P Tompa, pp. 159–83. New York: Springer
12. Hoh JH. 1998. Functional protein domains from the thermally driven motion of polypeptide chains: a proposal. *Proteins* 32:223–28
13. Uversky VN, Narizhneva NV, Ivanova TV, Kirkitadze MD, Tomashevski AY. 1997. Ligand-free form of human α-fetoprotein: evidence for the molten globule state. *FEBS Lett.* 410:280–84
14. Ebert M-O, Bae S-H, Dyson HJ, Wright PE. 2008. NMR relaxation study of the complex formed between CBP and the activation domain of the nuclear hormone receptor coactivator ACTR. *Biochemistry* 47:1299–308
15. Dunker AK, Obradović Z. 2001. The protein trinity—linking function and disorder. *Nat. Biotechnol.* 19:805–6
16. Holmes KC. 1983. Flexibility in tobacco mosaic virus. *Ciba Found. Symp.* 93:116–38
17. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, et al. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59
18. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, et al. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA* 103:8390–95
19. Inobe T, Fishbain S, Prakash S, Matouschek A. 2011. Defining the geometry of the two-component proteasome degron. *Nat. Chem. Biol.* 7:161–67
20. Gao J, Xu D. 2012. Correlation between posttranslational modification and intrinsic disorder in protein. *Pac. Symp. Biocomput.* 1:94–103
21. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J. Virol.* 83:10719–36
22. Kovacs E, Tompa P, Liliom K, Kalmar L. 2010. Dual coding in alternative reading frames correlates with intrinsic protein disorder. *Proc. Natl. Acad. Sci. USA* 107:5429–34
23. Hegyi H, Buday L, Tompa P. 2009. Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins. *PLoS Comput. Biol.* 5:e1000552
24. Light S, Sagit R, Ekman D, Elofsson A. 2013. Long indels are disordered: a study of disorder and indels in homologous eukaryotic proteins. *Biochim. Biophys. Acta* 1834:890–97

25. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. 2011. Evolution and disorder. *Curr. Opin. Struct. Biol.* 21:441–46
26. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* 107:8183–88
27. Das RK, Pappu RV. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* 110:13392–97
28. Tran HT, Mao A, Pappu RV. 2008. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* 130:7380–92
29. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. USA* 103:16764–69
30. Uversky VN, Ptitsyn OB. 1996. Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride–induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* 255:215–28
31. Ptitsyn OB, Bychkova VE, Uversky VN. 1995. Kinetic and equilibrium folding intermediates. *Philos. Trans. R. Soc. Lond. B* 348:35–41
32. Srivastava AK, Sharma Y, Chary KVR. 2010. A natively unfolded βγ-crystallin domain from *Habella chejuensis*. *Biochemistry* 49:9746–55
33. Choy WY, Forman-Kay JD. 2001. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* 308:1011–32
34. Jensen MR, Ruigrok RWH, Blackledge M. 2013. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* 23:426–35
35. Dunker AK, Kriwacki RW. 2011. The orderly chaos of proteins. *Sci. Am.* 304:68–73
36. Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–40
37. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM. 1997. CATH—a hierarchic classification of protein domain structures. *Structure* 5:1093–108
38. Vucetic S, Brown CJ, Dunker AK, Obradović Z. 2003. Flavors of protein disorder. *Proteins* 52:573–84
39. Yegambaram K, Bulloch EM, Kingston RL. 2013. Protein domain definition should allow for conditional disorder. *Protein Sci.* 22:1502–18
40. Bardwell JCA, Jakob U. 2012. Conditional disorder in chaperone action. *Trends Biochem. Sci.* 37:517–25
41. Zhang T, Faraggi E, Li Z, Zhou Y. 2013. Intrinsically semi-disordered state and its role in induced folding and protein aggregation. *Cell Biochem. Biophys.* 67:1193–205
42. McMeekin TL. 1952. Milk proteins. *J. Food Prot.* 15:57–63
43. Jirgensons B. 1958. Optical rotation and viscosity of native and denatured proteins. X. Further studies on optical rotatory dispersion. *Arch. Biochem. Biophys.* 74:57–69
44. Doolittle RF. 1973. Structural aspects of the fibrinogen to fibrin conversion. *Adv. Protein Chem.* 27:1–109
45. Bode W, Fehlhammer H, Huber R. 1976. Crystal structure of bovine trypsinogen at 1–8 Åresolution. I. Data collection, application of patterson search techniques and preliminary structural interpretation. *J. Mol. Biol.* 106:325–35
46. Manalan AS, Klee CB. 1983. Activation of calcineurin by limited proteolysis. *Proc. Natl. Acad. Sci. USA* 80:4291–95
47. Holt C, Carver JA, Ecroyd H, Thorn DC. 2013. Caseins and the casein micelle: their biological functions, structures and behaviour in foods. *J. Dairy Sci.* 96:6127–46
48. Horne DS. 2006. Casein micelle structure: models and muddles. *Curr. Opin. Colloid Interface Sci.* 11:148–53
49. Holt C, Timmins PA, Errington N, Leaver J. 1998. A core-shell model of calcium phosphate nanoclusters stabilized by β-casein phosphopeptides, derived from sedimentation equilibrium and small-angle X-ray and neutron-scattering measurements. *Eur. J. Biochem.* 252:73–78

50. De Kruif CG, Huppertz T, Urban VS, Petukhov AV. 2012. Casein micelles and their internal structure. *Adv. Colloid Interface Sci.* 171/172:36–52
51. Holt C. 2013. Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Curr. Opin. Struct. Biol.* 23:420–25
52. Holt C, Sørensen ES, Clegg RA. 2009. Role of calcium phosphate nanoclusters in the control of calcification. *FEBS J.* 276:2308–23
53. Byrne BM, van het Schip AD, van de Klundert JA, Arnberg AC, Gruber M, Ab G. 1984. Amino acid sequence of phosvitin derived from the nucleotide sequence of part of the chicken vitellogenin gene. *Biochemistry* 23:4275–79
54. Oldfield CJ, Xue B, Van Y-Y, Ulrich EL, Markley JL, et al. 2013. Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim. Biophys. Acta* 1834:487–98
55. Doolittle RF, Kollman JM. 2006. Natively unfolded regions of the vertebrate fibrinogen molecule. *Proteins* 63:391–97
56. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, et al. 2002. Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* 55:104–10
57. Fukuchi S, Homma K, Minezaki Y, Nishikawa K. 2006. Intrinsically disordered loops inserted into the structural domains of human proteins. *J. Mol. Biol.* 355:845–57
58. Tsurupa G, Mahid A, Veklich Y, Weisel JW, Medved L. 2011. Structure, stability, and interaction of fibrin αC-domain polymers. *Biochemistry* 50:8028–37
59. Marsh JJ, Guan HS, Li S, Chiles PG, Tran D, Morris TA. 2013. Structural insights into fibrinogen dynamics using amide hydrogen/deuterium exchange mass spectrometry. *Biochemistry* 52:5491–502
60. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. 1996. Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. USA* 93:11504–9
61. Hsu W-L, Oldfield CJ, Xue B, Meng J, Huang F, et al. 2013. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* 22:258–73
62. Baratti J, Maroux S, Louvard D. 1973. Effect of ionic strength and calcium ions on the activation of trypsinogen by enterokinase. A modified test for the quantitative evaluation of this enzyme. *Biochim. Biophys. Acta* 321:632–38
63. Bode W, Huber R. 1976. Induction of the bovine trypsinogen–trypsin transition by peptides sequentially similar to the N-terminus of trypsin. *FEBS Lett.* 68:231–36
64. Klee CB, Crouch TH, Krinks MH. 1979. Calcineurin: a calcium- and calmodulin-binding protein of the nervous system. *Proc. Natl. Acad. Sci. USA* 76:6270–73
65. Stewart AA, Ingebritsen TS, Manalan A, Klee CB, Cohen P. 1982. Discovery of a Ca^{2+}- and calmodulin-dependent protein phosphatase: probable identity with calcineurin (CaM-BP80). *FEBS Lett.* 137:80–84
66. Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, et al. 1995. Crystal structures of human calcineurin and the human FKBP12–FK506–calcineurin complex. *Nature* 378:641–44
67. Martínez-Martínez S, Redondo JM. 2004. Inhibitors of the calcineurin/NFAT pathway. *Curr. Med. Chem.* 11:997–1007
68. Radivojac P, Vucetic S, O'Connor TR, Uversky VN, Obradović Z, Dunker AK. 2006. Calmodulin signaling: analysis and prediction of a disorder-dependent molecular recognition. *Proteins* 63:398–410
69. Meador WE, Means AR, Quiocho FA. 1993. Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science* 262:1718–21
70. Rumi-Masante J, Rusinga FI, Lester TE, Dunlap TB, Williams TD, et al. 2012. Structural basis for activation of calcineurin by calmodulin. *J. Mol. Biol.* 415:307–17
71. Cavanagh J, Fairbrother WJ, Palmer AG III, Rance M, Skelton NJ. 2010. *Protein NMR Spectroscopy: Principles and Practice*. San Diego: Elsevier
72. Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, Dahlquist FW. 1997. The C-terminal half of the anti-σ factor, FlgM, becomes structured when bound to its target, σ 28. *Nat. Struct. Mol. Biol.* 4:285–91
73. Allison JR, Varnai P, Dobson CM, Vendruscolo M. 2009. Determination of the free energy landscape of α-synuclein using spin label nuclear magnetic resonance measurements. *J. Am. Chem. Soc.* 131:18314–26

74. Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, et al. 1994. Homeodomain-DNA recognition. *Cell* 78:211–23
75. Martin AJM, Walsh I, Tosatto SCE. 2010. MOBI: a web server to define and visualize structural mobility in NMR protein ensembles. *Bioinformatics* 26:2916–17
76. Ota M, Koike R, Amemiya T, Tenno T, Romero PR, et al. 2013. An assignment of intrinsically disordered regions of proteins based on NMR structures. *J. Struct. Biol.* 181:29–36
77. Larsson G, Martinez G, Schleucher J, Wijmenga SS. 2003. Detection of nano-second internal motion and determination of overall tumbling times independent of the time scale of internal motion in proteins from NMR relaxation data. *J. Biomol. NMR* 27:291–312
78. Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, et al. 1996. X-ray and NMR structure of human Bcl-XL, an inhibitor of programmed cell death. *Nature* 381:335–41
79. Ito Y, Selenko P. 2010. Cellular structural biology. *Curr. Opin. Struct. Biol.* 20:640–48
80. Dedmon MM, Patel CN, Young GB, Pielak GJ. 2002. FlgM gains structure in living cells. *Proc. Natl. Acad. Sci. USA* 99:12681–84
81. Selenko P, Wagner G. 2007. Looking into live cells with in-cell NMR spectroscopy. *J. Struct. Biol.* 158:244–53
82. Binolfi A, Theillet F-X, Selenko P. 2012. Bacterial in-cell NMR of human α-synuclein: a disordered monomer by nature? *Biochem. Soc. Trans.* 40:950–54
83. Li C, Charlton LM, Lakkavaram A, Seagle C, Wang G, et al. 2008. Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. *J. Am. Chem. Soc.* 130:6310–11
84. Barnes CO, Pielak GJ. 2011. In-cell protein NMR and protein leakage. *Proteins Struct. Funct. Bioinform.* 79:347–51
85. Xie Q, Arnold G, Romero P, Obradović Z, Garner E, Dunker A. 1998. The sequence attribute method for determining relationships between sequence and protein disorder. *Workshop Genome Inform.* 9:193–200
86. Dunker AK, Brown CJ, Obradović Z. 2002. Identification and functions of usefully disordered proteins. *Adv. Protein Chem.* 62:25–49
87. Theillet FX, Kalmar L, Tompa P, Han KY, Selenko P, et al. 2013. The alphabet of intrinsic disorder. 1. Act like a pro: on the abundance and roles of proline residues in intrinsically disordered regions. *Intrinsically Disord. Proteins* 1:e24360
88. Vacic V, Uversky VN, Dunker AK, Lonardi S. 2007. Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinform.* 8:211
89. Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci.* 3:522–24
90. Romero P, Obradović Z, Kissinger C, Villafranca JE, Dunker AK. 1997. Identifying disordered regions in proteins from amino acid sequence. *Int. Conf. Neural Netw.* 1997:90–95
91. Romero P, Obradović Z, Dunker AK. 1997. Sequence data analysis for long disordered regions prediction in the calcineurin family. *Workshop Genome Inform.* 8:110–24
92. Uversky VN, Gillespie JR, Fink AL. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–27
93. Dosztányi Z, Csizmók V, Tompa P, Simon I. 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* 347:827–39
94. Li X, Romero P, Rani M, Dunker AK, Obradović Z. 1999. Predicting protein disorder for N-, C-, and internal regions. *Workshop Genome Inform.* 10:30–40
95. Romero P, Obradović Z, Li X, Garner EC, Brown CJ, Dunker AK. 2001. Sequence complexity of disordered protein. *Proteins* 42:38–48
96. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradović Z. 2006. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinform.* 7:208
97. Dosztányi Z, Sándor M, Tompa P, Simon I. 2007. Prediction of protein disorder at the domain level. *Curr. Protein Pept. Sci.* 8:161–71
98. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. 2009. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19:929–49

99. Jin F, Liu Z. 2013. Inherent relationships among different biophysical prediction methods for intrinsically disordered proteins. *Biophys. J.* 104:488–95
100. Huang F. 2013. *Optimizing hydropathy scale to improve IDP prediction and characterizing IDPs' functions*. PhD thesis. Indiana Univ., Indianapolis. 121 pp.
101. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, et al. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21:3435–38
102. Dosztányi Z, Csizmok V, Tompa P, Simon I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21:3433–34
103. Monastyrskyy B, Fidelis K, Moult J, Tramontano A, Kryshtafovych A. 2011. Evaluation of disorder predictions in CASP9. *Proteins Struct. Funct. Bioinform.* 79:107–18
104. Monastyrskyy B, Kryshtafovych A, Moult J, Tramontano A, Fidelis K. 2014. Assessment of protein disorder region predictions in CASP10. *Proteins* 82(Suppl. 2):127–37
105. Romero P, Obradović Z, Kissinger CR, Villafranca JE, Garner E, et al. 1998. Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* 3:437–48
106. Dunker AK, Obradović Z, Romero P, Garner EC, Brown CJ. 2000. Intrinsic protein disorder in complete genomes. *Workshop Genome Inform.* 11:161–71
107. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* 337:635–45
108. Xue B, Dunker AK, Uversky VN. 2012. Orderly order in protein intrinsic disorder distribution: disorder in 3,500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* 30:137–49
109. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323:573–84
110. Uversky VN, Oldfield CJ, Dunker AK. 2008. Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annu. Rev. Biophys.* 37:215–46
111. Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, et al. 2012. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput. Biol.* 8:e1002709
112. Vacic V, Iakoucheva LM. 2012. Disease mutations in disordered regions—exception to the rule? *Mol. Biosyst.* 8:27–32
113. Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, et al. 2010. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* 78:365–80
114. Gao J, Thelen JJ, Dunker AK, Xu D. 2010. Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Mol. Cell Proteomics* 9:2586–600
115. Allers T. 2010. Overexpression and purification of halophilic proteins in *Haloferax volcanii*. *Bioeng. Bugs* 1:288–90
116. Fukuchi S, Hosoda K, Homma K, Gojobori T, Nishikawa K. 2011. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct. Biol.* 11:29
117. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, et al. 2013. D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* 41:D508–16
118. Pentony MM, Jones DT. 2010. Modularity of intrinsic disorder in the human proteome. *Proteins* 78:212–21
119. Sigler PB. 1988. Transcriptional activation, acid blobs and negative noodles. *Nature* 333:210–12
120. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic disorder in transcription factors. *Biochemistry* 45:6873–88
121. Minezaki Y, Homma K, Kinjo AR, Nishikawa K. 2006. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J. Mol. Biol.* 359:1137–49
122. Uversky VN, Dunker AK. 2012. *Methods in Molecular Biology*, Vol. 895: *Intrinsically Disordered Protein Analysis. 1*. New York: Humana
123. Uversky VN, Dunker AK. 2012. *Methods in Molecular Biology*, Vol. 896: *Intrinsically Disordered Protein Analysis. 2*. New York: Humana
124. Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. 1997. Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol.* 266:223–30

125. Iakoucheva LM, Kimzey AL, Masselon CD, Bruce JE, Garner EC, et al. 2001. Identification of intrinsic order and disorder in the DNA repair protein XPA. *Protein Sci.* 10:560–71
126. Liu Y, Matthews KS, Bondos SE. 2008. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the *Drosophila* Hox protein ultrabithorax. *J. Biol. Chem.* 283:20874–87
127. Johnson DE, Xue B, Sickmeier MD, Meng J, Cortese MS, et al. 2012. High-throughput characterization of intrinsic disorder in proteins from the protein structure initiative. *J. Struct. Biol.* 180:201–15
128. Hubbell WL, Gross A, Langen R, Lietzow MA. 1998. Recent advances in site-directed spin labeling of proteins. *Curr. Opin. Struct. Biol.* 8:649–56
129. Langen R, Cai K, Altenbach C, Khorana HG, Hubbell WL. 1999. Structural features of the C-terminal domain of bovine rhodopsin: a site-directed spin-labeling study. *Biochemistry* 38:7918–24
130. Belle V, Rouger S, Costanzo S, Liquière E, Strancar J, et al. 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* 73:973–88
131. Receveur-Brechot V, Durand D. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* 13:55–75
132. Hura GL, Budworth H, Dyer KN, Rambo RP, Hammel M, et al. 2013. Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* 10:453–54
133. Hegde ML, Tsutakawa SE, Hegde PM, Holthauzen LMF, Li J, et al. 2013. The disordered C-terminal domain of human DNA glycosylase NEIL1 contributes to its stability via intramolecular interactions. *J. Mol. Biol.* 425:2359–71
134. Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, et al. 2014. pE-DB: the database of structural ensembles of intrinsically disordered and denatured proteins. *Nucleic Acids Res.* 42:D326–35
135. Cortese MS, Uversky VN, Dunker AK. 2008. Intrinsic disorder in scaffold proteins: getting more from less. *Prog. Biophys. Mol. Biol.* 98:85–106
136. Buday L, Tompa P. 2010. Functional classification of scaffold proteins and related molecules. *FEBS J.* 277:4348–55
137. Xue B, Dunker AK, Uversky VN. 2012. The roles of intrinsic disorder in orchestrating the Wnt pathway. *J. Biomol. Struct. Dyn.* 29:843–61
138. Noutsou M, Duarte AMS, Anvarian Z, Didenko T, Minde DP, et al. 2011. Critical scaffolding regions of the tumor suppressor axin1 are natively unfolded. *J. Mol. Biol.* 405:773–86
139. Kuriyan J, Eisenberg D. 2007. The origin of protein interactions and allostery in colocalization. *Nature* 450:983–90
140. Xue B, Romero PR, Noutsou M, Maurice MM, Rüdiger SGD, et al. 2013. Stochastic machines as a colocalization mechanism for scaffold protein function. *FEBS Lett.* 587:1587–91
141. Frye JJ, Brown NG, Petzold G, Watson ER, Grace CRR, et al. 2013. Electron microscopy structure of human APC/C^{CDH1}–EMI1 reveals multimodal mechanism of E3 ligase shutdown. *Nat. Struct. Mol. Biol.* 20:827–35
142. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, et al. 1998. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.* 3:473–84
143. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. 2002. Intrinsic disorder and protein function. *Biochemistry* 41:6573–82
144. Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208
145. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. 2007. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 6:1882–98
146. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. 2007. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J. Proteome Res.* 6:1899–916
147. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, et al. 2007. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* 6:1917–32

148. Fersht AR, Knill-Jones JW, Bedouelle H, Winter G. 1988. Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: A mobile loop envelopes the transition state in an induced-fit mechanism. *Biochemistry* 27:1581–87
149. McElheny D, Schnell JR, Lansing JC, Dyson HJ, Wright PE. 2005. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc. Natl. Acad. Sci. USA* 102:5032–37
150. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, et al. 2013. A large-scale evaluation of computational protein function prediction. *Nat. Methods* 10:221–27
151. Cozzetto D, Jones DT. 2013. The contribution of intrinsic disorder prediction to the elucidation of protein function. *Curr. Opin. Struct. Biol.* 23:467–72
152. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, et al. 2007. DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35:D786–93
153. Fukuchi S, Sakamoto S, Nobe Y, Murakami SD, Amemiya T, et al. 2012. IDEAL: intrinsically disordered proteins with extensive annotations and literature. *Nucleic Acids Res.* 40:D507–11
154. Di Domenico T, Walsh I, Martin AJ, Tosatto SC. 2012. MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* 28:2080–81
155. Singh J, Whitwill S, Lacroix G, Douglas J, Dubuc E, et al. 2009. The use of group 3 LEA proteins as fusion partners in facilitating recombinant expression of recalcitrant proteins in *E. coli*. *Protein Expr. Purif.* 67:15–22
156. Santner AA, Croy CH, Vasanwala FH, Uversky VN, Van Y-YJ, Dunker AK. 2012. Sweeping away protein aggregation with entropic bristles: Intrinsically disordered protein fusions enhance soluble expression. *Biochemistry* 51:7250–62
157. Janin J, Sternberg MJE. 2013. Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol. Rep.* 5:2
158. Tompa P, Prilusky J, Silman I, Sussman JL. 2008. Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* 71:903–9
159. Tsvetkov P, Reuven N, Shaul Y. 2009. The nanny model for IDPs. *Nat. Chem. Biol.* 5:778–81
160. Singh GP, Ganapathi M, Sandhu KS, Dash D. 2006. Intrinsic unstructuredness and abundance of pest motifs in eukaryotic proteomes. *Proteins* 62:309–15
161. Sandhu KS, Dash D. 2006. Conformational flexibility may explain multiple cellular roles of pest motifs. *Proteins* 63:727–32
162. Garner E, Romero P, Dunker AK, Brown C, Obradović Z. 1999. Predicting binding regions within disordered proteins. *Workshop Genome Inform.* 10:41–50
163. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. 2007. Mining α-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 46:13468–77
164. Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, et al. 2007. Characterization of molecular recognition features, morfs, and their binding partners. *J. Proteome Res.* 6:2351–66
165. Wong ETC, Na D, Gsponer J. 2013. On the importance of polar interactions for complexes containing intrinsically disordered proteins. *PLoS Comput. Biol.* 9:e1003192
166. Gosselin P, Oulhen N, Jam M, Ronzca J, Cormier P, et al. 2011. The translational repressor 4E-BP called to order by eIF4E: new structural insights by SAXS. *Nucleic Acids Res.* 39:3496–503
167. Le Gall T, Romero PR, Cortese MS, Uversky VN, Dunker AK. 2007. Intrinsic disorder in the Protein Data Bank. *J. Biomol. Struct. Dyn.* 24:325–42
168. Zhang Y, Stec B, Godzik A. 2007. Between order and disorder in protein structures: analysis of "dual personality" fragments in proteins. *Structure* 15:1141–47
169. Dunker AK. 2007. Another window into disordered protein function. *Structure* 15:1026–28
170. Dosztányi Z, Mészáros B, Simon I. 2009. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25:2745–46
171. Disfani FM, Hsu W-L, Mizianty MJ, Oldfield CJ, Xue B, et al. 2012. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 28:i75–83
172. Obenauer JC, Cantley LC, Yaffe MB. 2003. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.* 31:3635–41

173. Puntervoll P, Linding R, Gemünd C, Chabanis-Davidson S, Mattingsdal M, et al. 2003. ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* 31:3625–30
174. Davey NE, Edwards RJ, Shields DC. 2007. The SLiMDisc server: short, linear motif discovery in proteins. *Nucleic Acids Res.* 35:W455–59
175. Mi T, Merlin JC, Deverasetty S, Gryk MR, Bill TJ, et al. 2012. Minimotif Miner 3.0: database expansion and significantly improved reduction of false-positive predictions from consensus sequences. *Nucleic Acids Res.* 40:D252–60
176. Mészáros B, Dosztányi Z, Simon I. 2012. Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition. *PLoS ONE* 7:e46829
177. Sargeant DP, Gryk MR, Maciejewski MW, Thapar V, Kundeti V, et al. 2012. Secondary structure, a missing component of sequence-based Minimotif definitions. *PLoS ONE* 7:e49957
178. Gunasekaran K, Tsai C-J, Nussinov R. 2004. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J. Mol. Biol.* 341:1327–41
179. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. 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
180. Peng Z, Oldfield CJ, Xue B, Mizianty MJ, Dunker AK, et al. 2013. A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome. *Cell Mol. Life Sci.* 71:1477–504
181. Fuxreiter M, Simon I, Friedrich P, Tompa P. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* 338:1015–26
182. Burgen AS, Roberts GC, Feeney J. 1975. Binding of flexible ligands to macromolecules. *Nature* 253:753–55
183. Koshland DE Jr. 1959. Enzyme flexibility and enzyme action. *J. Cell Comp. Physiol.* 54:245–58
184. Spolar RS, Record MT Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777–84
185. Espinoza-Fonseca LM. 2009. Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem. Biophys. Res. Commun.* 382:479–82
186. Sugase K, Dyson HJ, Wright PE. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–25
187. Choy MS, Page R, Peti W. 2012. Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochem. Soc. Trans.* 40:969–74
188. Drobnak I, De Jonge N, Haesaerts S, Vesnaver G, Loris R, Lah J. 2013. Energetic basis of uncoupling folding from binding for an intrinsically disordered protein. *J. Am. Chem. Soc.* 135:1288–94
189. Gruet A, Dosnon M, Vassena A, Lombard V, Gerlier D, et al. 2013. Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis. *J. Mol. Biol.* 425:3495–509
190. Shammas SL, Rogers JM, Hill SA, Clarke J. 2012. Slow, reversible, coupled folding and binding of the spectrin tetramerization domain. *Biophys. J.* 103:2203–14
191. Shammas SL, Travis AJ, Clarke J. 2013. Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. *J. Phys. Chem. B* 117:13346–56
192. Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA* 97:8868–73
193. Zhou H-X, Bates PA. 2013. Modeling protein association mechanisms and kinetics. *Curr. Opin. Struct. Biol.* 23:887–93
194. Hasty J, Collins JJ. 2001. Protein interactions. Unspinning the web. *Nature* 411:30–31
195. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272:5129–48
196. Uversky VN, Dunker AK. 2010. Understanding protein non-folding. *Biochim. Biophys. Acta* 1804:1231–64
197. Hsu WL. 2013. *Mechanisms of binding diversity in protein disorder: molecular recognition features mediating protein interaction networks*. PhD thesis, Indiana Univ., Indianapolis. 100 pp.
198. Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, et al. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.* 23:443–50

199. Colak R, Kim T, Michaut M, Sun M, Irimia M, et al. 2013. Distinct types of disorder in the human proteome: functional implications for alternative splicing. *PLoS Comput. Biol.* 9:e1003030
200. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, et al. 2012. Tissue-specific alternative splicing remodels protein–protein interaction networks. *Mol. Cell* 46:884–92
201. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, et al. 2002. Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415:549–53
202. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 30:281–83
203. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. 2009. Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays* 31:328–35
204. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–301
205. Williams RW, Xue B, Uversky VN, Dunker AK. 2013. Distribution and cluster analysis of predicted disordered proteins in Pfam domains. *Intrinsically Disord. Proteins* 1:e25724
206. Galea CA, Nourse A, Wang Y, Sivakolundu SG, Heller WT, Kriwacki RW. 2008. Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, $p27^{kip1}$. *J. Mol. Biol.* 376:827–38
207. Dunker AK, Uversky VN. 2008. Signal transduction via unstructured protein conduits. *Nat. Chem. Biol.* 4:229–30
208. Follis AV, Galea CA, Kriwacki RW. 2012. Intrinsic protein flexibility in regulation of cell proliferation: advantages for signaling and opportunities for novel therapeutics. *Adv. Exp. Med. Biol.* 725:27–49
209. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, et al. 2006. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362:1043–59
210. Tompa P, Fuxreiter M. 2008. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* 33:2–8
211. Marcotrigiano J, Gingras AC, Sonenberg N, Burley SK. 1999. Cap-dependent translation initiation in eukaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* 3:707–16
212. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, et al. 2001. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 414:514–21
213. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS. 2007. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl. Acad. Sci. USA* 104:9650–55
214. Schulz GE. 1979. Nucleotide binding proteins. In *Moledular Mechanisms of Biological Recognition*, ed. M Balaban, pp. 79–94. Amsterdam: Elsevier/North-Holland
215. Nadassy K, Wodak SJ, Janin J. 1999. Structural features of protein–nucleic acid recognition sites. *Biochemistry* 38:1999–2017
216. Vuzman D, Levy Y. 2012. Intrinsically disordered regions as affinity tuners in protein–DNA interactions. *Mol. Biosyst.* 8:47–57
217. Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound ultrabithorax-extradenticle homeodomain complex. *Nature* 397:714–19
218. Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136:701–18
219. Moore PB. 2012. How should we think about the ribosome? *Annu. Rev. Biophys.* 41:1–19
220. Lecompte O, Ripp R, Thierry J-C, Moras D, Poch O. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res.* 30:5382–90
221. Timsit Y, Acosta Z, Allemand F, Chiaruttini C, Springer M. 2009. The role of disordered ribosomal protein extensions in the early steps of eubacterial 50S ribosomal subunit assembly. *Int. J. Mol. Sci.* 10:817–34
222. Coelho Ribeiro MdL, Espinosa J, Islam S, Martinez O, Thanki JJ, et al. 2013. Malleable ribonucleoprotein machine: protein intrinsic disorder in the *Saccharomyces cerevisiae* spliceosome. *PeerJ* 1:e2
223. Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118
224. Koshland DE Jr, Némethy G, Filmer D. 1966. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5:365–85

225. Hilser VJ. 2010. Biochemistry. An ensemble view of allostery. *Science* 327:653–54
226. Beckett D. 2009. Regulating transcription regulators via allostery and flexibility. *Proc. Natl. Acad. Sci. USA* 106:22035–36
227. Motlagh HN, Hilser VJ. 2012. Agonism/antagonism switching in allosteric ensembles. *Proc. Natl. Acad. Sci. USA* 109:4134–39
228. Kumar R, McEwan IJ. 2012. Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. *Endocr. Rev.* 33:271–99
229. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA. 2013. Modulation of allostery by protein intrinsic disorder. *Nature* 498:390–94
230. Ivanyi-Nagy R, Davidovic L, Khandjian EW, Darlix J-L. 2005. Disordered RNA chaperone proteins: from functions to disease. *Cell Mol. Life Sci.* 62:1409–17
231. Tompa P, Csermely P. 2004. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* 18:1169–75
232. Shewmaker F, Kerner MJ, Hayer-Hartl M, Klein G, Georgopoulos C, Landry SJ. 2004. A mobile loop order-disorder transition modulates the speed of chaperonin cycling. *Protein Sci.* 13:2139–48
233. Tompa P, Kovacs D. 2010. Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol.* 88:167–74
234. Thorn DC, Ecroyd H, Sunde M, Poon S, Carver JA. 2008. Amyloid fibril formation by bovine milk α_{S2}-casein occurs under physiological conditions yet is prevented by its natural counterpart, α_{S1}-casein. *Biochemistry* 47:3926–36
235. Bordo D, Argos P. 1990. Evolution of protein cores. Constraints in point mutations as observed in globin tertiary structures. *J. Mol. Biol.* 211:975–88
236. Jernigan RL, Kloczkowski A. 2007. Packing regularities in biological structures relate to their dynamics. *Methods Mol. Biol.* 350:251–76
237. Chen JW, Romero P, Uversky VN, Dunker AK. 2006. Conservation of intrinsic disorder in protein domains and families. I. A database of conserved predicted disordered regions. *J. Proteome Res.* 5:879–87
238. Chen JW, Romero P, Uversky VN, Dunker AK. 2006. Conservation of intrinsic disorder in protein domains and families. II. Functions of conserved disorder. *J. Proteome Res.* 5:888–98
239. Bellay J, Han S, Michaut M, Kim T, Costanzo M, et al. 2011. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 12:R14
240. Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, et al. 2012. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46:871–83
241. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. 2007. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J. Mol. Evol.* 65:277–88
242. Brown CJ, Johnson AK, Daughdrill GW. 2010. Comparing models of evolution for ordered and disordered proteins. *Mol. Biol. Evol.* 27:609–21
243. Mosca R, Pache RA, Aloy P. 2012. The role of structural disorder in the rewiring of protein interactions through evolution. *Mol. Cell Proteomics* 11:M111.014969
244. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338:1587–93
245. Merkin J, Russell C, Chen P, Burge CB. 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338:1593–99
246. Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, Davey NE. 2013. The switches.elm resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.* 6:rs7
247. Xue B, Brown CJ, Dunker AK, Uversky VN. 2013. Intrinsically disordered regions of p53 family are highly diversified in evolution. *Biochim. Biophys. Acta* 1834:725–38
248. Schaefer C, Schlessinger A, Rost B. 2010. Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics* 26:625–31
249. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. 2011. Protein disorder—a breakthrough invention of evolution? *Curr. Opin. Struct. Biol.* 21:412–18
250. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. 2009. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138:198–208

251. Gsponer J, Futschik ME, Teichmann SA, Babu MM. 2008. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* 322:1365–68
252. Uversky VN, Dunker AK. 2008. Biochemistry: controlled chaos. *Science* 322:1340–41
253. Tompa P, Szász C, Buday L. 2005. Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* 30:484–89
254. Davey NE, Travé G, Gibson TJ. 2011. How viruses hijack cell regulation. *Trends Biochem. Sci.* 36:159–69
255. Mahani A, Henricksson J, Wright APH. 2013. Origins of myc proteins—using intrinsic disorder predictors to trace distant relatives. *PLoS ONE* 8:e75057
256. Mark W-Y, Liao JCC, Lu Y, Ayed A, Laister R, et al. 2005. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein–protein and protein–DNA interactions? *J. Mol. Biol.* 345:275–87
257. Bondos SE, Hsiao H-C. 2012. Roles for intrinsic disorder and fuzziness in generating context-specific function in ultrabithorax, a Hox transcription factor. *Adv. Exp. Med. Biol.* 725:86–105
258. Voet D, Voet JG, Pratt CW. 2012. *Fundamentals of Biochemistry: Life at the Molecular Level*. Hoboken, NJ: Wiley. 1200 pp. 4th ed.
259. Pratt CW, Cornely K. 2013. *Essential Biochemistry*. Hoboken, NJ: Wiley 744 pp. 3rd ed.
260. McKee T, McKee JR. 2014. *Biochemistry: The Molecular Basis of Life*. New York: Oxford Univ. Press. 944 pp. 5th ed.
261. Tymoczko JL, Berg JM, Stryer L. 2011. *Biochemistry: A Short Course*. New York: Freeman. 800 pp. 2nd ed.
262. Voet D, Voet JG. 2010. *Biochemistry*. Hoboken, NJ: Wiley. 1428 pp. 4th ed.
263. Nelson DL, Cox MM. 2012. *Lehninger Principles of Biochemistry*. New York: Freeman. 1100 pp. 6th ed.

TABLE OF CONTENTS FOR VOLUME 1:

Annual Review of Biochemistry
Volume 83, 2014

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://www.annualreviews.org/errata/biochem