INTRODUCTION

1

WHY ARE WE INTERESTED IN THE UNFOLDED PEPTIDES AND PROTEINS?

Vladimir N. Uversky and A. Keith Dunker

1.1. INTRODUCTION

In addition to transmembrane, globular, and fibrous proteins, it is becoming increasingly recognized that the protein universe includes intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs). These IDPs and IDRs are biologically active and yet fail to form specific three-dimensional (3-D) structures, existing instead as collapsed or extended dynamically mobile conformational ensembles [1–7]. These floppy proteins and regions are known as pliable, rheomorphic [8], flexible [9], mobile [10], partially folded [11], natively denatured [12], natively unfolded [3, 13], natively disordered [6], intrinsically unstructured [2, 5], intrinsically denatured [12], intrinsically unfolded [13], intrinsically disordered [4], vulnerable [14], chameleon [15], malleable [16], four-dimensional (4D) [17], protein-clouds [18], and dancing proteins [19], among several other terms. The variability of terms used to describe such proteins and regions is a simple reflection of their highly dynamic nature and the lack of the unique 3-D structure. None of these terms or their combinations is completely appropriate, as the majority of them have been borrowed from the fields such as protein folding or crystallography, which

Protein and Peptide Folding, Misfolding, and Non-Folding, First Edition. Edited by Reinhard Schweitzer-Stenner.

are not directly related to the biologically active proteins that normally exist as structural ensembles.

Since these proteins are highly abundant in any given proteome [20], the role of disorder in determining protein functionality in organisms can no longer be ignored. Native biologically active proteins were conceptualized as parts of the "protein trinity" [21] or the "protein quartet" [22], models where functional protein might exist in one of the several conformations—ordered, collapsed–disordered (molten globule-like), partially collapsed–disordered (pre-molten globule-like), or extended–disordered (coil-like)—and protein function might be derived from any one of these states and/or from the transitions between them. Disordered proteins are typically involved in regulation, signaling, and control pathways [23–25], which complement the functional repertoire of ordered proteins, which have evolved mainly to carry out efficient catalysis [26].

1.2. WHY STUDY IDPS?

Ordered globular proteins are characterized by rigid 3-D structures. The presence of such rigid structures implies that the Ramachandran angles and backbone atoms of each residue undergo non-isotropic small-amplitude motions relative to their local neighborhood and are characterized by the equilibrium positions defined by their time-averaged values. The atom fluctuations are caused by two factors, random thermal motion and small cooperative conformational changes of the local sequence neighborhood, and are known to be correlated with local residue packing [27]. Contrarily to this very static behavior, intrinsically disordered or natively unfolded proteins exist as dynamic ensembles in which atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values and typically involve non-cooperative conformational changes [6].

The kindred of proteins and protein domains, which have been shown *in vitro* to have little or no ordered structure under physiological conditions, is rapidly amplifying. In fact, over the past decade there has been an exponential increase in the amount of studies dedicated to intrinsically disordered or natively unfolded proteins, starting from a few papers in the early 1990s, and ending with about 300 papers in 2011. A special database, DisProt, was created to keep information about these proteins [28]. There are currently more than 620 proteins in this database.

The growing interest in this class of proteins is determined by several factors. The first issue is the structure–function relationship. Although the importance of protein dynamics for protein functions was recognized long ago, the existence of biologically active but extremely flexible proteins questioned the cornerstone paradigm of a protein science according to which a rigid well-folded 3-D structure is required for protein function. In a recent review, Professor Livesay emphasized, "In the same way that static photos of a dance recital certainly fail to reflect the completeness and grandeur of the performance, discrete structural snapshots lack sufficient information to completely describe protein dynamics" [19]. Since the structure

and function of IDPs represent a "shape-shifting dance," new ways of analyzing protein structure analysis and investigating protein functionality are necessary.

The lack of rigid globular structure under physiological conditions was proposed to provide IDPs with a considerable functional advantage, as their large plasticity allows them to interact efficiently with several different targets [2–4, 7]. Furthermore, a disorder/order transition induced in disordered proteins during the binding to specific targets *in vivo* provides a unique mechanism for the decoupling of binding specificity and affinity and might represent a simple tool for the regulation of numerous cellular processes, including transcription, translation, and cell cycle control [2–4, 7, 29]. Evolutionary persistence of the IDPs represents additional confirmation of their importance and raises intriguing questions on the role of protein disorder in biological processes.

Second, biomedical aspects are also of great importance. It has been established that deposition of some natively unfolded proteins is related to the development of several neurodegenerative disorders [30, 31]. Examples include Alzheimer's disease (deposition of amyloid-β, tau-protein, α-synuclein fragment NAC); Niemann–Pick disease type C, sub-acute sclerosing panencephalitis, argyrophilic grain disease, myotonic dystrophy, and motor neuron disease with neurofibrillary tangles (accumulation of tau-protein in the form of neurofibrillary tangles); Down's syndrome (nonfilamentous amyloid-β deposits); Parkinson's disease, dementia with Lewy bodies (LBs), LB variant of Alzheimer's disease, multiple system atrophy, and Hallervorden–Spatz disease (deposition of α-synuclein in the form of LBs and Lewy neurites [LNs]).

Finally, IDPs represent an attractive subject for the biophysical characterization of unfolded polypeptide chain under physiological conditions. A large variety of biophysical and biochemical methods have been applied to the structural description of these proteins. This includes proton nuclear magnetic resonance (^{1}H-NMR), heteronuclear NMR, circular dichroism (CD), optical rotatory dispersion (ORD), Fourier transform infrared spectroscopy (FTIR), intrinsic and extrinsic fluorescence, small angle X-ray scattering (SAXS), small angle neutron scattering (SANS), dynamic and static light scattering, gel electrophoresis, gel filtration, sedimentation, viscometry, scanning calorimetry, proteolytic mapping, epitope mapping, and electron microscopy (summarized in Ref. [32]). The special term "natively unfolded" has been used since it was introduced in 1994 to describe the behavior of tau-protein [12]. Although large amounts of experimental data have been accumulated and several disordered proteins have been rather well characterized, no systematic analysis of structural data for the kindred of IDPs has yet been carried out. This lack of methodical inspection of the conformational behavior of IDPs led frequently to confusion.

1.3. *LESSON 1:* DISORDEREDNESS IS ENCODED IN THE AMINO ACID SEQUENCE AND CAN BE PREDICTED

Similar to ordered proteins, the correct folding of which into relatively rigid biologically active conformations is determined by their amino acid sequences, the lack of

rigid structure in IDPs is also encoded in the specific features of their amino acid sequences. Some of these proteins have been discovered due to their unusual amino acid sequence compositions. The absence of regular structure in these proteins has been explained by the specific features of their amino acid sequences including the presence of numerous uncompensated charged groups (often negative), that is, a high net charge at neutral pH, arising from the extreme pI values in such proteins [13, 33, 34], and a low hydrophobic amino acid residue content [33, 34].

The analysis of charge and hydropathy has been shown to be sufficient to distinguish structured and some disordered proteins [3]. In fact, by comparing 275 natively folded and 91 natively unfolded proteins (i.e., proteins that at physiological conditions have been reported to have the NMR chemical shifts of a random coil, and/or lack significant ordered secondary structure, as determined by CD or FTIR, and/or show hydrodynamic dimensions close to those typical of an unfolded polypeptide chain), it has been shown that the combination of low mean hydrophobicity and relatively high net charge represents an important prerequisite for the absence of a compact structure in proteins under physiological conditions [3]. This observation was used to develop a charge–hydropathy (CH) plot method for distinguishing ordered and extended disordered proteins based only on their net charges and hydropathies [3]. According to this approach, natively unfolded proteins are specifically localized within a specific region of CH phase space and are separated from compact ordered proteins by a linear boundary [3]. From the physical viewpoint, such a combination of low hydrophobicity and high net charge as a prerequisite for intrinsic unfoldedness makes perfect sense: a high net charge leads to charge–charge repulsion, and low hydrophobicity means less driving force for protein compaction. In other words, these features are characteristic of IDPs with coil-like (or close to coil-like) structures.

A more detailed analysis was carried out to gain additional information on the compositional difference between ordered and disordered proteins. Comparison of a non-redundant set of ordered proteins with several data sets of disorder (where proteins were grouped based on different techniques, such as X-ray crystallography, NMR, and CD, used to identify disorder) revealed that disordered regions share at least some common sequence features over many proteins [1, 35]. In fact, the disordered proteins/regions were shown to be significantly depleted in bulky hydrophobic (Ile, Leu, and Val) and aromatic amino acid residues (Trp, Tyr, and Phe), which would normally form the hydrophobic core of a folded globular protein, and also possess low content of Cys and Asn residues. The depletion of disordered protein in Cys is also crucial as this amino acid residue is known to have a significant contribution to the protein conformation stability via the disulfide bond formation or being involved in coordination of different prosthetic groups. These depleted residues, Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn, were proposed to be called order-promoting amino acids. On the other hand, IDPs were shown to be substantially enriched in Ala, polar, disorder-promoting, amino acids, namely Arg, Gly, Gln, Ser, Glu, and Lys, and also in the hydrophobic but structure-braking Pro [4, 26, 36–38]. Note that these biases in the amino acid compositions of disordered proteins are also

consistent with the low overall hydrophobicity and high net charge characteristic of the natively unfolded proteins.

In addition to amino acid composition, the disordered segments have also been compared with the ordered ones by various attributes such as hydropathy, net charge, flexibility index, helix propensities, strand propensities, and compositions for groups of amino acids such as Trp + Tyr + Phe (aromaticity). As a result, 265 property-based attribute scales [36] and more than 6000 composition-based attributes (e.g., all possible combinations having one to four amino acids in the group) have been compared [39]. It has been established that 10 of these attributes, including 14 Å contact number, hydropathy, flexibility, β-sheet propensity, coordination number, content of major disorder-promoting residues (Arg + Ser + Pro + Glu), bulkiness, content of major order-promoting residues (Cys + Trp + Tyr + Phe), volume, and net charge, provide fairly good discrimination between order and disorder [4]. Later, 517 amino acid scales (including a variety of hydrophobicity scales, different measures of side chain bulkiness, polarity, volume, compositional attributes, the frequency of each single amino acid, and so on) were analyzed to construct a new amino acid attribute, for example, a novel amino acid scale that discriminates between order and disorder [40]. This scale outperformed the other 517 amino acid scales for the discrimination of order and disorder and provided a new ranking for the tendencies of the amino acid residue to promote order or disorder (from order-promoting to disorder-promoting): Trp, Phe, Tyr, Ile Met, Leu, Val, Asn, Cys, Thr, Ala, Gly, Arg, Asp, His, Gln, Lys, Ser, Glu, and Pro [40].

The fact that the sequences of ordered and disordered proteins are noticeably different raised three important conclusions: (1) IDPs clearly constitute a separate entity inside the protein kingdom; (2) these proteins can be reliably predicted using various computational tools [41]; (3) since peculiarities of amino acid sequence determine protein structure, structurally, these proteins should be very different from ordered globular proteins.

1.4. *LESSON 2:* DISORDERED PROTEINS ARE HIGHLY ABUNDANT IN NATURE

Intrinsic disorder in proteins is a common phenomenon. Based on the assumption that the absence of rigid structure is encoded in the specific features of the amino acid sequence, several predictors of naturally disordered regions (PONDRs) have been developed [1, 4, 42, 43]. Using these predictors, IDPs were indicated to be widespread. In one experiment, more than 15,000 out of 91,000 proteins in the then-current Swiss Protein database were identified as having long regions of sequence that shared the distinguishing sequence attributes of known IDRs [42]. In a second experiment, the commonness of intrinsic disorder was estimated by predicting disorder for whole genomes, including both known and putative protein sequences. Such predictions have been published for 31 genomes that span the three kingdoms. The percentage of sequences in each genome with segments predicted to have ≥40 consecutive disordered residues was used to gain an overview of proteomic disorder.

For so many consecutive predictions of disorder, the false-positive error rate was estimated from ordered proteins to be less than 0.5% of the segments of 40 and less than 6% of the fully ordered proteins [4, 43]. The eukaryotes exhibited more disorder by these measures than either the prokaryotes or the archaea, with *Caenorhabditis elegans*; *Arabidopsis thaliana*; *Saccharomyces cerevisiae*; and *Drosophila melanogaster* predicted to have 52–67% of their proteins with such long predicted regions of disorder, and bacteria and archaea predicted to have 16–45% and 26–51% of their proteins with such long disorder regions, respectively [43]. The increased prediction of disorder in eukaryotes compared with the other kingdoms has been suggested to be a consequence of the increased need for cell signaling and regulation [4, 43].

To understand the level of abundance of intrinsic disorder in the preeminent source of protein structural information, the Protein Data Bank (PDB), the amino acid sequences of proteins whose structures are determined by X-ray crystallography were compared with the corresponding sequences from the Swiss-Prot database [44]. The analyzed data set included 16,370 structures, which represent 18,101 PDB chains and 5434 different proteins from 910 different organisms (2793 eukaryotic, 2109 bacterial, 288 viral, and 244 archaeal). The analysis revealed that the complete sequences of only approximately 7% of proteins were observed in the corresponding PDB structures, and only approximately 25% of the total data set had >95% of their lengths observed in the corresponding PDB structures [44]. This clearly showed that the vast majority of PDB proteins were shorter than their corresponding Swiss-Prot sequences and/or contained numerous residues, which were not observed in maps of electron density.

According to their appearance in corresponding PDB entries, the residues in the Swiss-Prot sequences were grouped into four general categories: "Observed" (residues in structured regions); "Not observed" (residues regions with missing electron density, potentially disordered); "Uncharacterized" (residues that were not in the PDB sequence but were present in the Swiss-Prot sequence); and "Ambiguous" (residues of a single PDB chain associated with multiple PDB structures, which are observed in some 3-D structures but not in others; these residues occur due to the high redundancy of PDB) [44]. Next, the amino acid compositions and disorder propensities of residues in these four categories were analyzed by four different disorder predictors (PONDR® VL-XT, VL3-BA, VSL1P, and IUPred). "Not observed," "Ambiguous," and "Uncharacterized" regions were shown to possess the amino acid compositions typical for IDPs. The vast majority of residues in the "Observed" data set were predicted to be ordered, whereas the "Not observed" regions were mostly disordered. The "Uncharacterized" regions possessed some tendency toward order. Disorder predictions for the short "Ambiguous" regions were ambiguous, whereas long "Ambiguous" regions (>70 amino acid residues) were mostly predicted to be ordered, suggesting that they are likely to be wobbly domains [44]. The major conclusion of this study was really surprising since completely ordered proteins were not highly abundant in PDB and many PDB sequences had disordered regions. In fact, ~10% of the PDB proteins contained long disordered or ambiguous regions (with length $L > 30$ amino acids), and ~40% of the PDB proteins

possessed short disordered or ambiguous regions (with $10 \leq L < 30$ amino acids long) [44].

1.5. *LESSON 3:* DISORDERED PROTEINS ARE GLOBALLY HETEROGENEOUS

Typically, structural classification of IDPs is based on the definitions and terms elaborated for conformations observed in the unfolding/refolding pathways of typical globular proteins, which are known to exist in at least four different conformations, ordered, molten globule, pre-molten globule, and unfolded states [45–50]. The structural properties of the *molten globule* are well known and have been systematized in a number of reviews [e.g., Ref. [47]]. The protein molecule in this intermediate state has a globular structure typical of ordered globular proteins, preserves high secondary-structure content, and possesses the native-like folding pattern, but has no (or has only a trace of) rigid cooperatively melted tertiary structure. Molten globules are more accessible to proteolysis and are characterized by high affinity to the hydrophobic fluorescence probes (such as 8-anilinonaphthalene-1-sulfonate [ANS]). The averaged value for the increase in the hydrodynamic radius of a protein molecule in the molten globule state compared with that of ordered state is 15–20%, which corresponds to a volume increase of ~50%. The protein molecule in the pre-molten globule has no rigid tertiary structure and does not possess globular structure but is characterized by a considerable secondary structure. The protein molecule in the pre-molten globule state is considerably less compact than in the molten globule or folded states, but it is still more compact than the random coil. It can effectively interact with the hydrophobic fluorescent probe ANS, although weaker than the molten globule does.

Based on their structural properties, IDPs are separated in two different groups. Members of the first group, despite their flexibility, are rather compact and possess a well-developed secondary structure; that is, they show properties typical of the molten globule [47]. Proteins from the other group behave almost as a random coil [51] or as pre-molten globules [48]. The proteins from this second group constitute a class of natively unfolded (or extended intrinsically disordered) proteins, which are extremely flexible, essentially non-compact (extended), and have little or no ordered secondary structure under physiological conditions. The general conformational properties of intrinsically unfolded proteins are summarized below. Here we will mostly focus on the structural characteristics, which make natively unfolded proteins exceptional among others. These are low compactness, absence of globularity, low secondary-structure content, and high flexibility.

1.5.1. Compactness

The most unambiguous characteristic of the conformational state of a globular protein remains the hydrodynamic dimensions of the macromolecule. It was noted long ago that hydrodynamic techniques may help to recognize when a globular

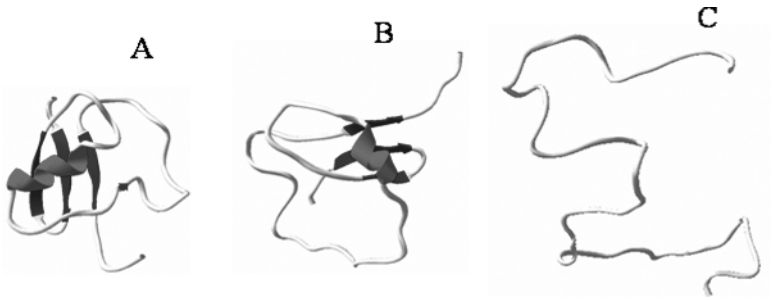

Figure 1.1. Illustration of the hydrodynamic volume variation for a polypeptide chain of 100 amino acids in different intrinsically disordered configurations: (A) collapsed (molten globule-like) disorder; (B) extended (pre-molten globule-like) disorder; (C) extended (coil-like) disorder.

protein has lost all of its non-covalent structure, that is, when it became unfolded [51]. This is because an essential increase in the hydrodynamic volume is associated with the unfolding of a protein molecule. Therefore, conformational states of a globular protein, ordered, molten globule, pre-molten globule, and unfolded, may be easily discriminated by the degree of compactness of the polypeptide chain. Similarly, different types of IDPs (molten globule-like, pre-molten globule-like, and coil-like) may be distinguished by their degree of compactness (Fig. 1.1). Furthermore, ordered and unfolded conformations of globular proteins possess very different molecular mass dependencies of their hydrodynamic radii, R_S [22, 51, 52]. Comparison of the $\log(R_S)$ versus $\log(M)$ curves for natively unfolded proteins with the same dependencies for the ordered, molten globular, pre-molten globular, and urea- or GdmCl-unfolded globular proteins revealed that these $\log(R_S)$ versus $\log(M)$ dependencies for different conformations of globular proteins are described by straight lines [22, 53]. This analysis revealed that according to their $\log(R_S)$ versus $\log(M)$ dependencies, natively unfolded proteins may be divided in two groups, with one of these groups behaving as random coils in poor solvent, and with protein in the second group being close, with respect to their hydrodynamic characteristics, to pre-molten globules (see Fig. 1.2).

This is a very important observation, which may definitely help in understanding the physical nature of the natively unfolded proteins. In fact, it is well established that the behavior of unfolded proteins obeys the theoretical and empirical rules that apply to linear random coils [51]. Particularly, it is known that the hydrodynamic dimensions of random coils depend essentially on the quality of solvent. A poor solvent encourages the attraction of macromolecular segments and hence a chain has to squeeze. On the other hand, in a good solvent, repulsive forces act primarily between the segments and the macromolecule conforms to a loose fluctuating coil [49, 51]. Water is a poor solvent, whereas solutions of urea and GdmCl are rather good solvents, with GdmCl being closer to the ideal one. This difference in solvent quality may account for the observed divergence in $\log(R_S)$ versus $\log(M)$ dependen-

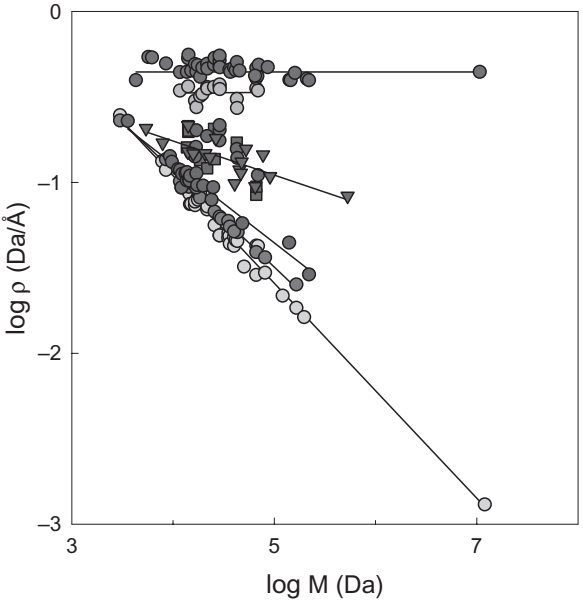

Figure 1.2. Variation of the density of protein molecules, ρ, with protein molecular weight, M, for ordered (red circles), molten globule (green circles), pre-molten globules (dark yellow symbols, where intermediates accumulated during the unfolding by urea or GdmCl are shown by circles); proteins with intact disulate bridges in 8 M urea or 6 M GdmCl are shown as squares; native pre-molten globules are shown as reversed triangles); native coils (blue circles); proteins without cross-links or with reduced cross-links unfolded in 8 M urea (pink circles); proteins without cross-links or with reduced cross-links unfolded in 6 M GdmCl (turquoise circles). The solid lines represent the best fit of the data. See color insert.

cies for the coil-like part of IDPs. The existence of a well-defined difference between the log(R_S) versus log(M) dependencies for globular proteins unfolded by urea and GdmCl should also be noted in this respect (see Fig. 1.2).

1.5.2. Globularity

Another very important structural parameter is the degree of globularization, which reflects the presence or absence of a tightly packed core in the protein molecule. In fact, it has been shown that the globular proteins in pre-molten globule-like conformations are characterized by low (coil-like) intramolecular packing density [46, 49]. This information could be extracted from the analysis of SAXS data (Kratky plot), whose shape is sensitive to the conformational state of the scattering protein molecules [54, 55]. It has been shown that a scattering curve in the Kratky coordinates has a characteristic maximum when the globular protein is in the native state or in the molten globule state (i.e., has a globular structure): if a protein is completely

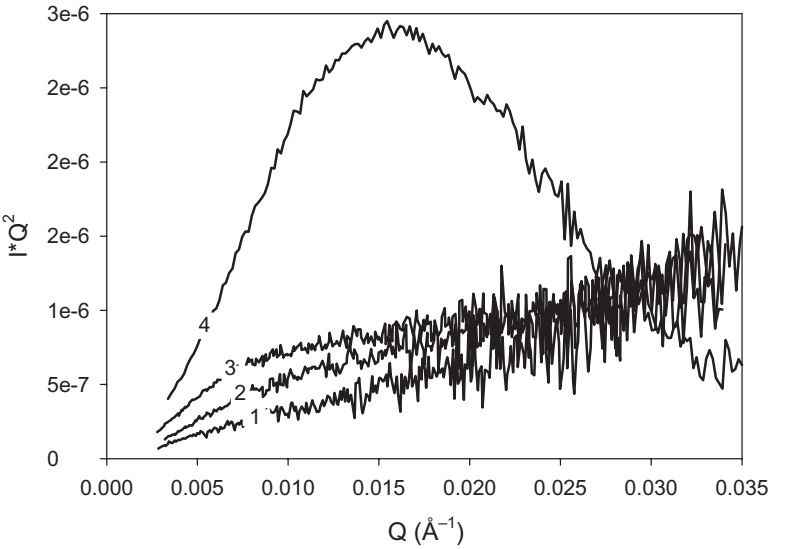

Figure 1.3. Kratky plots of SAXS data for natively unfolded α-synuclein [1], prothymosin α [2], and caldesmon 636–771 fragment [3]. The Kratky plot of a typical ordered protein SNase is shown for comparison [4].

unfolded or a pre-molten globule conformation (has no globular structure), such a maximum will be absent on the respective scattering curve. Analysis of several natively unfolded proteins by this technique revealed that representatives of both classes of natively unfolded proteins (coil-like and pre-molten globule-like) are characterized by the absence of a rigid globular core (see Fig. 1.3).

1.5.3. Secondary Structure

Far-ultraviolet CD spectra of a number of natively unfolded proteins (such as α-synuclein, prothymosin α, phosphodiesterase γ-subunit, caldesmon 636–771 fragment, and many others) possess distinctive spectral features with characteristic deep minima in the vicinity of 200 nm, and relatively low ellipticity at 220 nm (Fig. 1.4). This characteristic shape of far-UV CD spectra is a useful criterion for the selection of natively unfolded proteins. Figure 1.5 represents a "double wavelength" plot, $[\theta]_{222}$ versus $[\theta]_{200}$ dependence, which may be used to assort natively unfolded proteins into two non-overlapping groups. Approximately half of the studied proteins was characterized by far-UV CD spectra characteristic of almost completely unfolded polypeptide chains, with $[\theta]_{200} = -(18{,}900 \pm 2800)$ degree·cm^2/dmol and $[\theta]_{222} = -(1700 \pm 700)$ degree·cm^2/dmol, whereas the other half possessed spectra typical of the pre-molten globule state of globular proteins, being consistent with the existence of some residual secondary structure (with $[\theta]_{200} = -(10{,}700 \pm 1300)$ degree·cm^2/dmol and $[\theta]_{222} = -(3900 \pm 1100)$ degree·cm^2/dmol) [22, 53]. Some of

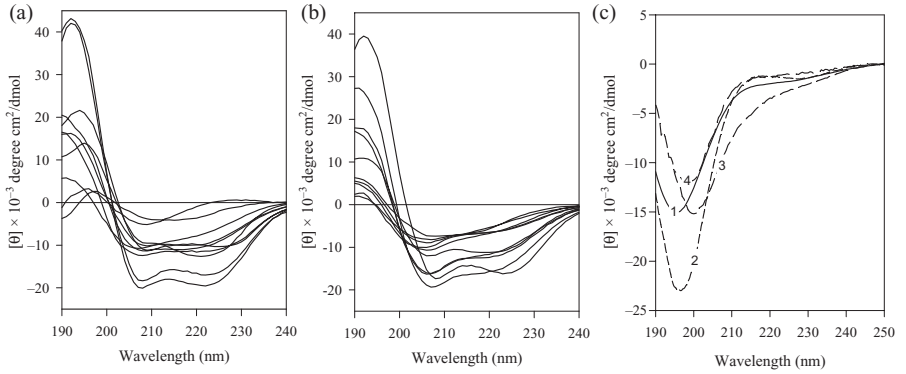

Figure 1.4. Far-UV CD spectra of nine ordered proteins: α-lactalbumin (bovine), α-lactalbumin (human), β-lactamase, ribonuclease A, retinol binding protein, apo-form, carbonic anhydrase B, phosphoglycerate kinase, cytochrome *c*, apomyoglobin (a), their molten globule folding intermediate states (b), and extended IDPs, α-synuclein [1], prothymosin α [2], caldesmon 636–771 fragment [3], and phosphodiesterase γ-subunit [4] (c).

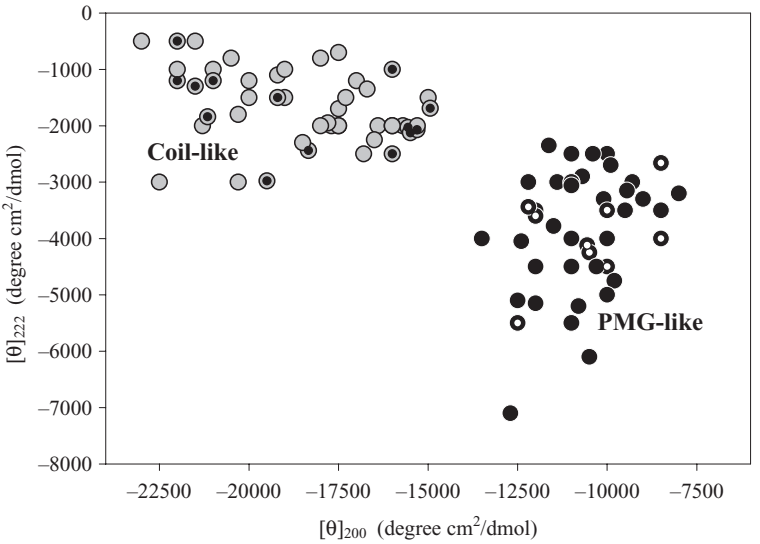

Figure 1.5. Analysis of far-UV CD spectra in terms of double wavelength plot, $[\theta]_{222}$ versus $[\theta]_{200}$, allows the natively unfolded proteins division on coil-like (gray circles) and pre-molten globule-like subclasses (black circles). Intrinsic pre-molten globules (PMG) and intrinsic coils for which the hydrodynamic parameters were measured are marked by white-dotted and black-dotted symbols, respectively.

the natively unfolded proteins were simultaneously characterized by CD and hydrodynamic methods. Intrinsic pre-molten globules and intrinsic coils studied by both techniques are indicated in Figure 1.5 as white-dotted and black-dotted symbols, respectively. These data are consistent with the important conclusion that more compact polypeptides (with pre-molten globule-like hydrodynamic characteristics) possess larger amounts of ordered secondary structure than less compact coil-like natively unfolded proteins. Thus, the simultaneous application of CD and hydrodynamic techniques leaves no doubt that natively unfolded proteins should be subdivided into two structurally distinct groups: intrinsic coils and intrinsic pre-molten globules [22, 53].

The analysis of these spectra yielded a low content of ordered secondary structure (α-helices and β-sheets).

This lack of ordered secondary structure is also confirmed by the FTIR spectroscopy of secondary-structure composition of natively unfolded proteins, such as tau-protein [12], α-synuclein [13, 56], β- and γ-synucleins [57], α_s-casein [58], cAMP-dependent protein kinase inhibitor [59], nucleoporins (Nups) [60, 61], and many others. Importantly, even the caldesmon 636–771 fragment, which was shown to have hydrodynamic properties typical of the pre-molten globule, possesses far-UV CD characteristic of essentially distorted polypeptide chains. Thus, the low overall content of ordered secondary structure could be considered a general property of extended IDPs.

1.5.4. High Flexibility

The fact that IDPs are characterized by an increased intramolecular flexibility may be easily derived from a large amount of NMR studies (summarized in Refs. [2–4, 6, 32, 62–66]). Furthermore, recent advances in NMR technology (especially the use of heteronuclear multidimensional approach) have even opened the way to detailed structural and dynamic description of these proteins [62–66]. Increased flexibility of natively unfolded proteins is indirectly confirmed by their extremely high sensitivity to protease degradation *in vitro* [4, 6, 32]. The functional implication of large intramolecular flexibility has also been recognized. Particularly, it has been noted that the intrinsic lack of structure and high flexibility may be considered a major functional advantage of a protein, which in this way attains the ability to bind several different targets. Precise control over the thermodynamics of the binding process may also be achieved in this way.

1.6. *LESSON 4:* HYDRODYNAMIC DIMENSIONS OF NATIVELY UNFOLDED PROTEINS ARE CHARGE DEPENDENT

Since extended intrinsically disordered (or natively unfolded) proteins are characterized by the low proportion of hydrophobic residues combined with the high content of charged amino acids, they represent an ideal system for the analysis of the role of charge interactions for the conformational properties of unfolded proteins, and

for testing the quantitative descriptions and predictions of polymer theory for the influence of charged amino acids on chain dimensions [67]. Recently, using single-molecule Förster resonance energy transfer (FRET), long-range distance distributions and dynamics of several natively unfolded proteins were analyzed. These proteins with very different properties were chosen to sample the range of sequence compositions found in natural proteins [67], as represented by the CH-plot discussed above [3]: the globular cold shock protein Csp*Tm*, which is stably folded even in the absence of ligands; the N-terminal domain of HIV-1 integrase, which folds only upon binding of Zn^{2+} ions and is otherwise denatured; and human prothymosin α, one of the natively unfolded proteins with the largest fraction of charged amino acids identified so far, which does not assume a well-defined folded structure under any known conditions and does not contain regular secondary structure, but plays crucial roles in different biological processes including cell proliferation, transcriptional regulation, and apoptosis. These three proteins were specifically labeled with a donor (Alexa Fluor 488, Molecular Probes, Eugene, OR) and an acceptor (Alexa Fluor 594, Molecular Probes) fluorophore, and investigated with confocal single-molecule fluorescence spectroscopy under a variety of conditions to evaluate the efficiency of energy transfer for individual molecules freely diffusing through the focal spot of the laser beam [67]. This analysis allows finding subpopulations of proteins characterized by different energy transfer efficiencies and distinguishing changes in the conformational properties within one of these subpopulations from a change in their relative abundances. The analysis revealed very peculiar and characteristic responses of the unfolded states of these three proteins to the increasing GdmCl concentration. The unfolded state of the heat shock protein Csp*Tm*, which was detectable by FRET even in the absence of GdmCl, continuously expanded with the increase in the denaturant concentration, whereas the GdmCl dependence on the energy transfer efficiency for the intrinsically disordered HIV-1 integrase and human prothymosin α possessed a more complex behavior, indicating the presence of partial collapse of the unfolded polypeptide chain at low denaturant concentrations followed by the denaturant-induced expansion similar to Csp*Tm* at higher GdmCl concentrations [67]. This conclusion was based on the presence of the remarkable "rollover" of the efficiency of the energy transfer below approximately 0.5 M GdmCl, which was absent when the uncharged denaturant urea was used for the analogous experiments. Importantly, the amplitude of the mentioned rollover of the efficiency of the energy transfer was proportional to the protein charge density. Based on these observations, it has been concluded that natively unfolded proteins can exhibit a prominent expansion at low ionic strengths, which correlate with their net charge and that these pronounced effects of charges on the dimensions of unfolded proteins might have important implications for their cellular functions [67].

This conclusion was further supported by a comprehensive analysis of phenylalanine-glycine (FG) repeat-containing Nups [68], which contain large intrinsically disordered domains with multiple phenylalanine–glycine repeats (FG domains) and form the nuclear pore complex gating the nucleocytoplasmic transport in eukaryotes. A systematic analysis of the hydrodynamic dimensions revealed that under the physiological conditions *in vitro* the FG domains of nucleporins adopt distinct categories

of intrinsically disordered structures, such as molten globule, pre-molten globule, relaxed coil, extended coil (as in urea), or very extended coil (as in GdmCl) [68]. Furthermore, the category of intrinsically disordered structure in a given FG domain was related to its amino acid composition, namely to the content of charged residues, where more charged FG domains possessed larger hydrodynamic dimensions. FG nucleporins with higher charge densities were shown to be more dynamic than the collapsed-coil FG domains. Furthermore, these relaxed- or extended-coil FG domains were shown to repel other FG domains, whereas the collapsed-coil FG domains were able to bind each other to form oligomers, clearly suggesting that there is a functional need in cells to have some FG domains aggregate and other FG domains repel, therefore providing a molecular basis for two different gating mechanisms operating at the nuclear pore complex at distinct locations, one acting as a hydrogel and the other as an entropic brush [68]. Therefore, the abundance and peculiarities of the charged residues distribution within the protein sequences might determine the physical and biological properties of natively unfolded proteins.

1.7. *LESSON 5:* POLYMER PHYSICS EXPLAINS HYDRODYNAMIC BEHAVIOR OF DISORDERED PROTEINS

Each protein is believed to be a unique entity that has a quite unmatched primary sequence, which governs its 3-D structure (or luck thereof) and ensures specific biological functions. Therefore, understanding the effect of sequence variance on the biological performance presents a difficult challenge. However, natural polypeptides have originated as random copolymers of amino acids, which were adjusted or "selected" over evolution based on their functional capacity [69, 70]. Despite their differences in primary amino acid sequences, protein molecules in a number of conformational states behave as polymer homologues, suggesting that the volume interactions can be considered a major driving force responsible for the formation of equilibrium structures or structural ensembles [71]. For example, ordered globular proteins and molten globules (both as folding intermediates of globular proteins and as examples of collapsed IDPs) exhibit key properties of polymer globules, where the fluctuations of the molecular density are expected to be much less than the molecular density itself. Extended IDPs (both intrinsic coils and intrinsic pre-molten globules) and ordered proteins in the pre-molten globule intermediate state possess properties of squeezed coils since water is a poor solvent for a polypeptide. In fact, even high concentrations of strong denaturants (e.g., urea and GdmCl) are very likely to be poor solvents for protein chains, resulting in the preservation of an extensive residual structure even under these harsh denaturing conditions [71].

Based on these and related observations, and given the fact that IDPs (especially their extended forms) are characterized by significant amino acid composition biases, the behavior of low-complexity polypeptides (e.g., homopolypeptide and block copolypeptides) can mimic reasonably well the overall polymeric behavior of more complex systems (e.g., IDPs and IDRs). Intriguingly, recent studies revealed that polar homopolypeptides without hydrophobic groups (e.g., polyglutamine or

glycine–serine block copolypeptides), used to model IDPs, prefer collapsed ensembles in aqueous media [72–77]. Furthermore, even polyglycine tends to form heterogeneous ensembles of collapsed structures in water [77]. These results clearly suggest that water is a poor solvent for polypeptide backbone alone and for the IDPs containing long tracts of polar amino acid residues.

A systematic analysis of conformational behavior of protamines, arginine-rich IDPs involved in the condensation of chromatin during spermatogenesis, and protamine-like peptides by a combination of molecular simulations and fluorescence experiments revealed that there is a charge-driven coil-to-globule transition in these highly charged polypeptides, where the net charge per residue serves as the discriminating order parameter (Rohit Pappu, personal communication). Overall, the increase in the size of a polypeptide chain with increasing net charge per residue can be attributed to the increase in the intramolecular electrostatic repulsions between similarly charged side chains and the favorable solvation of these moieties. Pappu also pointed out that there are at least three different classes of globule-forming polar/charged IDPs. The first class is composed of polar tracts, which collapse due to the water being a poor solvent for a backbone and non-charged side chains. The second class is represented by weak polyelectrolites and weak polyampholytes, which have low per residue net charge and low fractions of positively and/or negatively charged residues. These IDPs form collapsed structures since the driving force responsible for the collapse is not overcome by the intramolecular electrostatic repulsion between the charged side chains and by their favorable free energies of solvation. Furthermore, if such IDPs possess a polyampholytic nature, their globular state is additionally stabilized by electrostatic interactions between the oppositely charged side chains. Finally, IDPs from the third class are strong polyampholytes characterized by high fractions of positively and/or negatively charged residues but low per residue net charges. Such IDPs can form collapsed structures stabilized mostly by multiple electrostatic interactions between solvated side chains of opposite sign (Rohit Pappu, personal communication).

Clearly, IDPs with very high net charges are expected to be more extended and behave more similar to random coils (i.e., similar to conformations adopted by proteins in the denaturant GdmCl). The validity of this hypothesis was recently illustrated via the analysis of the set of Nups containing long natively unfolded domains with phenylalanine–glycine repeats (FG domains). These Nups constitute a gate of the nuclear pore complex (NPC), where the FG domains form a malleable network of disordered polypeptides, which selects and size-discriminates against diffusing macromolecules [68]. In this study, most Nup FG domains were shown to adopt collapsed molten-globular configurations and were characterized by a low content of charged amino acids. Others adopted more extended, coil-like conformations, were structurally more dynamic, and were characterized by a high content of charged amino acids. Many Nups contained both types of structures in a biphasic distribution along their polypeptide chain. For example, the N-terminus of Nsp1 (AA 1–172; Nsp1n) had a low charged-AA content of 2% and adopted molten globular configurations, whereas the remainder of the Nsp1 FG domain (AA 173–603; Nsp1m) had a charged AA content of 36% and adopted extended-coil configurations [68].

1.8. *LESSON 6:* NATIVELY UNFOLDED PROTEINS ARE PLIABLE AND VERY SENSITIVE TO THEIR ENVIRONMENT

Although the mechanisms of function of any biological subject (including proteins) cannot be understood without appropriate consideration of this subject's environment, IDPs represent an extreme case of environmental pliability. Amino acid biases characteristic of IDPs determine their structural variability and lack of a rigid well-folded structure. This structural plasticity is necessary for the unique functional repertoire of IDPs, which is complementary to the catalytic activities of ordered proteins. Amino acid biases also drive atypical responses of IDPs to changes in their environment. In general, the conformational behavior of IDPs is characterized by the low cooperativity (or the complete lack thereof) of the denaturant-induced unfolding, lack of the measurable excess heat absorption peak(s) characteristic of the melting of ordered proteins, "turned out" response to heat and changes in pH, and the ability to gain structure in the presence of various binding partners [78].

1.8.1. Effects of Strong Denaturants

IDPs, being highly dynamic, are characterized by low conformational stability, which is reflected in low steepness of the transition curves describing their unfolding induced by strong denaturants or even in the complete lack of the sigmodal shape of the unfolding curves. This is in strict contrast to the solvent-induced unfolding of ordered globular proteins, which is known to be a highly cooperative process. In fact, we can find here an extreme case of the cooperative transition, which is an all-or-none transition where a cooperative unit includes the whole molecule; that is, no intermediate states can be observed in the transition region. Based on the analysis of the unfolding transitions in ordered globular proteins, it has been concluded that the steepness of urea- or GdmCl-induced unfolding curves depends strongly on whether a given protein has a rigid tertiary structure (i.e., it is ordered) or is already denatured and exists as a molten globule [79, 80]. Although the denaturant-induced unfolding of a native molten globule can be described by a shallow sigmoidal curve (e.g., see Ref. [81]), urea- or GdmCl-induced structural changes in native pre-molten globules or native coils are non-cooperative and typically seen as monotonous featureless changes in the studied parameters. This is due to the low content of the residual structure in these species.

1.8.2. Temperature Effects

The analysis of the temperature effects on structural properties of several extended IDPs revealed that native coils and native pre-molten globules possess so-called "turned out" response to heat [78]. Figure 1.6A depicts the temperature-dependence of $[\theta]_{222}$ for α-synuclein, caldesmon 636–771 fragment, and phosphodiesterase γ-subunit and clearly shows that these three proteins partially fold as the temperature is increased. Therefore, temperature-induced changes in the far-UV CD spectrum of several extended IDPs were interpreted in terms of the temperature-induced forma-

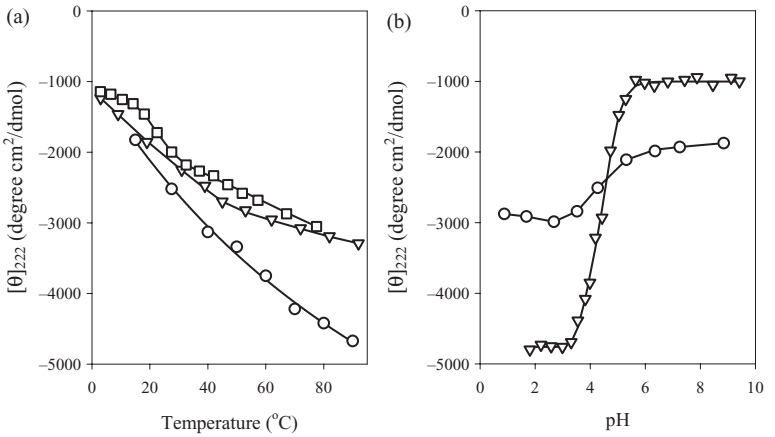

Figure 1.6. Effect of environmental factors on conformational properties of natively unfolded proteins. (a) Temperature-induced changes in far-UV CD spectrum ($[\theta]_{222}$ versus temperature dependence) measured for α-synuclein (triangles), phosphodiesterase γ-subunit (squares), and caldesmon 636–771 fragment (circles). (b) pH-induced structure formation ($[\theta]_{222}$ versus pH dependence) in the natively unfolded α-synuclein (circles) and prothymosin α (triangles).

tion of secondary structure [56, 82, 83]. These heating-induced structural changes in extended IDPs were completely reversible and consistent with the partial. The effects of elevated temperatures may be attributed to the increased strength of the hydrophobic interaction at higher temperatures, leading to a stronger hydrophobic attraction, which is the major driving force for folding.

1.8.3. Structure-Promoting Effects of Extreme pH

Extended IDPs are also characterized by the "turned out" response to changes in pH [56, 84–87], where a decrease (or increase) in pH induced partial folding of extended IDPs due to a decrease in their high net charge present at neutral pH, thereby decreasing charge–charge intramolecular repulsion and permitting hydrophobicity-driven collapse to the partially folded conformation [78]. Figure 1.6B illustrates this unusual conformational behavior of extended IDPs representing the pH dependence of $[\theta]_{222}$ for α-synuclein and prothymosin α. These data are consistent with the conclusion that changes in pH can induce partial folding of natively unfolded proteins.

1.8.4. Protein Chameleons

In general, the pliable nature of the IDPs allows them to be highly adjustable and to morph their structure. For example, α-synuclein was shown to adopt a series of different conformations depending on its environment, being able to stay substantially unfolded, or fold into an amyloidogenic partially folded conformation, or into α-helical or β-structured species, both monomeric and oligomeric. Furthermore, it

was shown to form several morphologically different types of aggregates, including oligomers (spheres, doughnuts, and worms), amorphous aggregates, and amyloid-like fibrils. Based on this astonishing conformational behavior and structural plasticity, the concept of a protein chameleon was introduced [15].

1.9. *LESSON 7:* WHEN BOUND, NATIVELY UNFOLDED PROTEINS CAN GAIN UNUSUAL STRUCTURES

Due to their lack of rigid structure, combined with the high level of intrinsic dynamics and almost unrestricted flexibility at various structure levels in the non-bound state, and due to their unique capability to adjust to the structure of the binding partner, IDPs are characterized by a diverse range of binding modes, creating a multitude of unusual complexes, many of which are not attainable by ordered proteins [88]. Some of these complexes are relatively static, resemble complexes of ordered proteins, and therefore are suitable for structure determination by X-ray crystallography. Among these static complexes are molecular recognition features (MoRFs), wrappers, chameleons, penetrators, huggers, intertwined strings, long cylindrical containers, connectors, armature, tweezers and forceps, grabbers, tentacles, pullers, and stackers or β-arcs [88]. These binding modes are shown in Figure 1.7 and briefly described below.

1.9.1. MoRFs

MoRFs are short, interaction-prone IDP segments that undergo disorder-to-order transitions upon binding and are abundantly involved in molecular recognition [89–91]. MoRFs, being identified as short structured fragments of disordered proteins involved in interaction with globular partners, were structurally classified according to their structures in the bound state: α-MoRFs form α-helices, β-MoRFs form β-strands, and ι-MoRFs form structures without a regular pattern of backbone hydrogen bonds [89, 91]. MoRF typically constitutes one contiguous segment fitted into a grove at the surface of the ordered partner.

1.9.2. Flexible Wrappers

Some IDPs wrap around their ordered binding partners. Complexes of this type are polyvalent ordered complexes where several ordered segments of a disordered protein bind to disjoint and spatially distant binding sites on the surface of the globular protein. Typically, the ordered segments of such wrapping IDPs are connected by the disordered regions. Secondary-structure elements of wrappers almost do not possess intramolecular interactions, forming very intensive intermolecular contacts with the binding partner [92–94]. Many proteins interacting with DNA or RNA are flexible wrappers. For example, numerous transcription factors, regulatory proteins, and other proteins that interact with DNA contain multiple zinc finger motifs. The zinc finger motifs act as independently folded globular domains that are

separated by flexible linker regions. Zinc finger domains are disordered in the absence of zinc. Proteins often contain multiple zinc fingers connected by flexible linkers and wrap around the DNA in a spiral manner. The zinc finger-containing proteins typically interact with the major groove along the double helix of DNA. In the bound state, the zinc fingers are arranged around the DNA strand in such a way that the α-helix of each finger contacts the DNA, forming an almost continuous stretch of α-helices around the DNA molecule [95].

1.9.3. Penetrators

In complexes of some IDPs with other proteins or RNA, significant parts of IDPs penetrate deep inside the structure of their binding partners. For example, in the crystal structure of the 30S ribosome, many ribosomal proteins, in addition to globular domains, contained extended internal loops or long N- or C-terminal extensions that were not seen in structures of the isolated proteins but which were associated intimately with the RNA inside the ribosome [96]. The most illustrative example of this penetrating mode is S12, which had a globular domain at the interface side and a long N-terminal extension that threaded its way through the 30S subunit to emerge on the back side to interact with proteins S8 and S17 [96].

1.9.4. Huggers

Typically, monomers constituting the oligomers formed as a result of folding coupled to binding are highly intertwined [97–99], kind of hugging each other. Typically, complexes of this kind are binary. However, there are several examples of group huggers, where monomers clasp more than one partner.

1.9.5. Intertwined Strings

Coiled coils represent a common structural motif in proteins, where up to seven long α-helices intertwined together similar to strings of a rope [100, 101]. This motif is formed by approximately 3–5% of all amino acids in proteins [102], with the most common members of this family being dimers and trimers. Individual α-helices in coiled coil are wrapped around each other into a left-handed helix to form a supercoil. In addition to the left-handed coiled coils, there are right-handed coiled coils [103, 104]. Coiled coils represent a relatively simple but tightly packed structure. Importantly, partners involved in the coiled-coil formation are typically disordered in the non-bound form.

1.9.6. Long Cylindrical Containers

Multichain coiled coils can assemble into long hollow cylinders containing a continuous axial pore with binding capacities for several hydrophobic compounds [105].

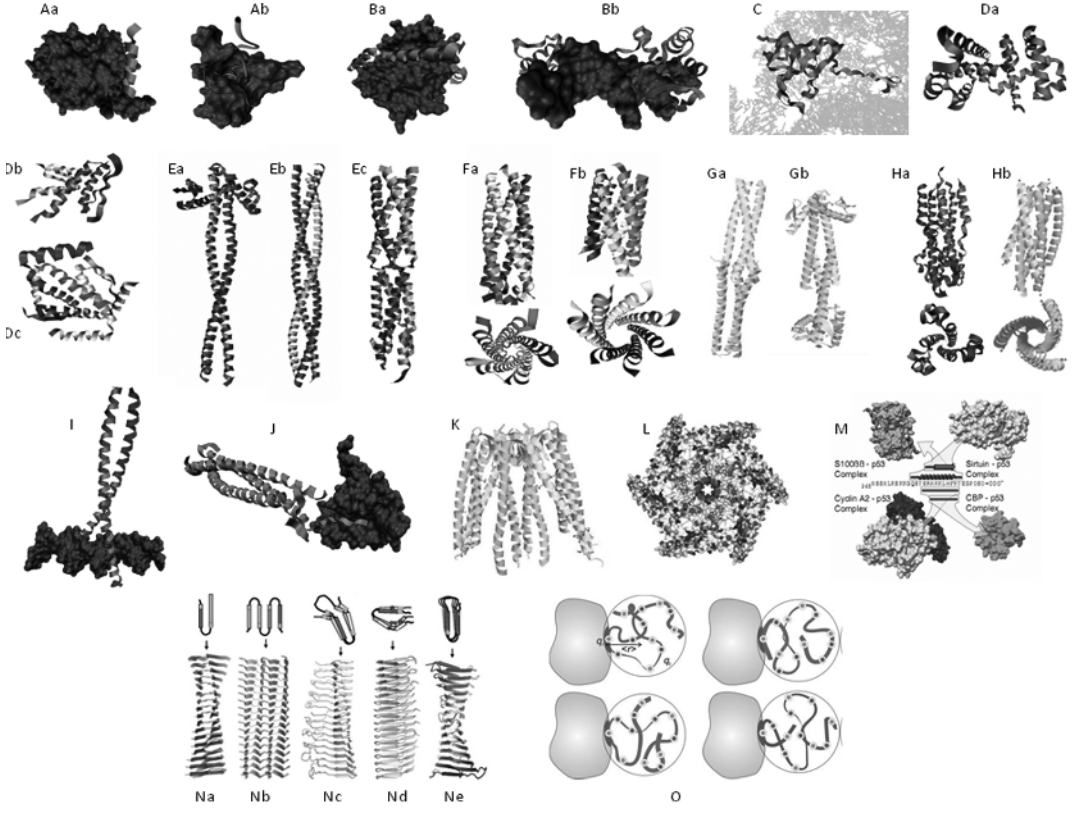
Aa
Ab
Ba
Bb
C
Da
Db
Dc
Ea
Eb
Ec
Fa
Fb
Ga
Gb
Ha
Hb
I
J
K
L
M
S100B - p53 Complex
Sirtuin - p53 Complex
Cyclin A2 - p53 Complex
CBP - p53 Complex
Na
Nb
Nc
Nd
Ne
O

Figure 1.7. A portrait gallery of disorder-based complexes. Illustrative examples of various interaction modes of intrinsically disordered proteins are shown. (A) **MoRFs:** (Aa) α-MoRF, a complex between the botulinum neurotoxin (red helix) and its receptor (a blue cloud) (PDB ID: 2NM1); (Ab) i-MoRF, a complex between an 18-mer cognate peptide derived from the α1 subunit of the nicotinic acetylcholine receptor from *Torpedo californica* (red helix) and α-cobratoxin (a blue cloud) (PDB ID: 1LXH). (B) **Wrappers:** (Ba) rat PP1 (blue cloud) complexed with mouse inhibitor-2 (red helices) (PDB ID: 2O8A); (Bb) a complex between the paired domain from the *Drosophila* paired (prd) protein and DNA (PDB ID: 1PDN). (C) **Penetrator:** Ribosomal protein S12 embedded into the rRNA (PDB ID: 1N34). (D) **Huggers:** (Da) *E. coli trp* repressor dimer (PDB ID: 1ZT9); (Db) tetramerization domain of p53 (PDB ID: 1PES); (Dc) tetramerization domain of p73 (PDB ID: 2WQI). (E) **Intertwined strings:** (Ea) dimeric coiled coil, a basic coiled-coil protein from *Eubacterium eligens* ATCC 27750 (PDB ID: 3HNW); (Eb) trimeric coiled coil, salmonella trimeric autotransporter adhesin, SadA (PDB ID: 2WPQ); (Ec) tetrameric coiled coil, the virion-associated protein P3 from Caulimovirus (PDB ID: 2O1J). (F) **Long cylindrical containers:** (Fa) pentameric coiled coil, side and top views of the assembly domain of cartilage oligomeric matrix protein (PDB ID: 1FBM); (Fb) side and top views of the seven-helix coiled coil, engineered version of the GCN4 leucine zipper (PDB ID: 2HY6). (G) **Connectors:** (Ga), human heat shock factor binding protein 1 (PDB ID: 3CI9); (Gb) the bacterial cell division protein ZapA from *Pseudomonas aeruginosa* (PDB ID: 1W2E). (H) **Armature:** (Ha) side and top views of the envelope glycoprotein GP2 from Ebola virus (PDB ID: 2EBO); (Hb) side and top views of a complex between the N- and C-terminal peptides derived from the membrane fusion protein of the Visna (PDB ID: 1JEK). (I) **Tweezers or forceps:** A complex between c-Jun, c-Fos, and DNA. Proteins are shown as red helices, whereas DNA is shown as a blue cloud (PDB ID: 1FOS). (J) **Grabbers:** Structure of the complex between βPIX coiled coil (red helices) and Shank PDZ (blue cloud) (PDB ID: 3L4F). (K) **Tentacles:** Structure of the hexameric molecular chaperone prefoldin from the archaeum *Methanobacterium thermoautotrophicum* (PDB ID: 1FXK). (L) **Pullers:** Structure of the ClpB chaperone from *Thermus thermophilus* (PDB ID: 1QVR). (M) **Chameleons:** The C-terminal fragment of p53 gains different types of secondary structure in complexes with four different binding partners, cyclinA (PDB ID: 1H26), sirtuin (PDB ID: 1MA3), CBP bromo domain (PDB ID: 1JSP), and s100$\beta\beta$ (PDB ID: 1DT7). (N) **Stackers or β-arcs:** (Na) stack of β-arches, β-amyloid; (Nb), superpleated β-structure (Sup35p, Ure2P, α-synuclein); (Nc) stack of β-solenoids (prion); (Nd) stack of β-arch dimers (insulin); (Ne) β-solenoids. Modified from Ref. [116]. (O) **Dynamic complexes:** Schematic representation of the polyelectrostatic model of Sic1–Cdc4 interaction. Schematic of an IDP (ribbon) interacting with a folded receptor (gray shape) through several distinct binding motifs and an ensemble of conformations (indicated by four representations of the interaction). The IDP possesses positive and negative charges (depicted as blue and red circles, respectively), giving rise to a net charge q_l, while the binding site in the receptor (light blue) has a charge q_r. The effective distance $<r>$is between the binding site and the centre of mass of the IDP. Reproduced from Ref. [119]. See color insert.

1.9.7. Connectors and Armature

Being formed, coiled coils can be used in the subsequent formation of higher order oligomers, where segments of coiled coils are used to bring oligomeric partners together [106–108]. Coiled coil can serve as an armature, around which a more complex structure is built [109, 110].

1.9.8. Tweezers and Forceps

Many transcription factors form coiled-coil dimers that interact with DNA. Here, the coiled-coil dimer grips the major grove of DNA in a forceps-like manner [111].

1.9.9. Grabbers

In several instances, the ends of the coiled coil form an extensive β-sheet interaction with binding partners [112].

1.9.10. Tentacles

In its crystal structure, the hexameric molecular chaperone prefoldin resembles a jellyfish with a body consisting of a double β-barrel assembly, from which six long tentacle-like coiled coils are protruding. The distal regions of the coiled coils contain hydrophobic patches, which are utilized in the multivalent binding of non-native proteins [113].

1.9.11. Pullers

The *Thermus thermophilus* chaperone ClpB is a two-tiered hexameric ring with a set of 85-Å-long and mobile coiled coils that are located on the outside of the hexamer and act as mechanical pullers [114]. Here, the concerted motions of these coiled coils cause adjacent subunits to move in opposite directions, generating the mechanical force required to pull aggregates apart [114].

1.9.12. Chameleons

One of the most unique features of IDPs is their ability to gain, in a template-dependent manner, very different structures in the bond form. This capability is illustrated by the C-terminal binding region of p53, the same short segment of which binds to four unrelated partners adopting different conformations (an α-helix, a β-sheet, and two differently laid irregular structures) when bound to the different partners [7, 115].

1.9.13. Stackers or β-arcs

The unifying model of the amyloid fibrils is "β-arcades," which are the columnar structures produced by in-register stacking of "β-arcs," strand-turn-strand motifs in which the two β-strands interact via their side chains, not via the polypeptide backbone as in a conventional β-hairpin [116].

1.10. *LESSON 8:* IDPS CAN FORM DISORDERED OR FUZZY COMPLEXES

In addition to the static complexes considered above, where bound partners have fixed structures, some IDPs do not fold even in their bound state, forming so-called disordered, dynamic, or fuzzy complexes with ordered proteins [82, 117–121], other disordered proteins [122–124], or biological membranes [125, 126]. In complexes of some of these IDPs with their binding partners, the disordered regions flanking the interaction interface but not the interface itself remain disordered. Such mode of interaction was recently described as "the flanking fuzziness" in contrast to "the random fuzziness" when the disordered protein remains entirely disordered in the bound state [127, 128]. It is also expected that a similar binding mode can be utilized by disordered proteins while interacting with nucleic acids and other biological macromolecules [88].

Physically, binding is considered as joining objects together, and suggests spatial and temporal fixation of bound partners. The formation of protein complexes with specific binding partners is expected to bring some fixation (at least at the binding site). Therefore, the mentioned disordered complexes, where interaction of a disordered protein with the binding partners is not accompanied by a disorder-to-order transition within the interaction interface clearly, cannot be described by the classical binding paradigm. The contradiction can be resolved assuming that the ordered binding partner or disordered protein contains multiple low-affinity binding sites. The existence of several similar binding sites combined with a highly flexible and dynamic structure of disordered protein creates a unique situation where any binding site of disordered protein can interact with any binding site of its partner with almost equal probability, in a staccato manner. The low affinity of each individual contact implies that each of them is not stable and can be readily broken. Therefore, such disordered or fuzzy complex can be envisioned as a highly dynamic ensemble in which a disordered protein does not present a single binding site to its partner but resembles a "binding cloud," in which multiple identical binding sites are dynamically distributed in a diffuse manner (see Fig. 1.7O). In other words, in this staccato-type interaction mode, a disordered protein rapidly changes multiple binding sites while probing the binding site(s) of its partner [88]. An additional factor that can help holding dynamic complex together could be a weak long-range attraction between protein molecules [129]. This long-range attraction is universal for all protein solutions and has a range several times that of the diameter of the protein molecule, much greater than the range of the screened electrostatic repulsion [129].

1.11. *LESSON 9:* INTRINSIC DISORDER IS CRUCIAL FOR RECOGNITION, REGULATION, AND SIGNALING

The functional importance of being disordered has been intensively analyzed [2, 4, 6, 21, 23–25, 130–136]. The majority of IDPs undergo a disorder-to-order transition upon functioning [2, 4, 6, 21, 23–25, 130–137]. When disordered regions bind to signaling partners, the free energy required to bring about the disorder-to-order

transition takes away from the interfacial, contact free energy, with the net result that a highly specific interaction can be combined with a low net free energy of association [4, 130]. High specificity coupled with low affinity seems to be a useful pair of properties for a signaling interaction so that the signaling interaction is reversible. In addition, a disordered protein can readily bind to multiple partners by changing shape to associate with different targets [4, 138, 139]. In addition to decoupled specificity and strength of binding, disorder has several clear advantages for functions in signaling, regulation, and control [1, 2, 4, 140–143]:

1. Increased speed of interaction due to greater capture radius and the ability to spatially search through interaction space.
2. Strengthened encounter complex allows for less stringent spatial orientation requirements.
3. Efficient regulation via rapid degradation.
4. Increased interaction (surface) area per residue.
5. A single disordered region may bind to several structurally diverse partners.
6. Many distinct (structured) proteins may bind to a single disordered region.
7. Intrinsic disorder provides the ability to overcome steric restrictions, enabling larger interaction surfaces in protein–protein and protein–ligand complexes than those obtained with rigid partners.
8. Unstructured regions fold to specific bound conformations, which can be very different, according to the template provided by structured partners.
9. Efficient regulation via posttranslational modification (PTM); that is, phosphorylation, methylation, ubiquitination, sumoylation, and so on.
10. Ease of regulation/redirection and production of otherwise diverse forms by alternative splicing (AS).
11. The possibility of overlapping binding sites due to extended linear conformations.
12. Diverse evolutionary rates with some IDPs being highly conserved and other IDPs possessing high evolutionary rates. The latter can evolve into sophisticated and complex interaction centers (scaffolds) that can be easily tailored to the needs of divergent organisms,
13. Flexibility that allows masking (or not) of interaction sites or that allows interaction between bound partners.
14. Binding fuzziness where different binding mechanisms (e.g., via stabilizing the binding-competent secondary-structure elements within the contacting region, or by establishing the long-range electrostatic interactions, or being involved in transient physical contacts with the partner, or even without any apparent ordering) can be employed to accommodate peculiarities of interaction with various partners.

This clearly suggests that there is a new two-pathway protein structure–function paradigm, with sequence-to-structure-to-function for enzymes and membrane trans-

port proteins, and sequence-to-disordered ensemble-to-function for proteins and protein regions involved in signaling, regulation, and control [4, 21, 22, 53, 132, 133].

Among the various functions found for disordered regions, even superficial analysis of natively unfolded proteins revealed that many of them undergo disorder-to-order transitions when stabilized by binding with specific targets [3]. In fact, for the majority of proteins described in a previous study, the existence of ligand-induced folding was established. Examples include induced structure formation upon binding with DNA (or RNA) for protamines, Max protein, high-mobility group proteins HMG-14 and HMG-17, osteonectine, Ser-Asp (SD) repeat protein D (SdrD protein), chromatogranins A and B, Δ131Δ fragment of SNase, and histone H1. Other examples include the folding of cytochrome *c* in the presence of heme, the folding of osteocalcine induced by cations, secondary-structure formation in parathyroid hormone-related protein induced by membrane association, structure formation in glucocorticoid receptor brought about by association with trimethylamine N-oxide, folding of histidine-rich protein II induced by heme; and structure formation and compaction of prothymosin α mediated by zinc [3]. Therefore, among the major functions of these unstructured, IDPs are nucleic acid binding, metal ion binding, heme binding, and interaction with membrane bilayers [3].

More than 150 proteins have been identified in early studies as containing functional disordered regions, or being completely disordered, yet performing vital cellular roles [4, 132]. Twenty-eight separate functions were assigned for these disordered regions, including molecular recognition via binding to other proteins, or to nucleic acids [132, 133]. An alternative view is that functional disorder fits into at least six broad classes based on their mode of action [5, 144].

A computational study was carried out for the evaluation of a correlation between the functional annotations in the Swiss-Prot database and the predicted intrinsic disorder [145–147]. The approach is based on the hypothesis that if a function described by a given keyword relies on intrinsic disorder, then the keyword-associated protein would be expected to have a greater level of predicted disorder than the protein randomly chosen from the Swiss-Prot. To test this hypothesis, functional keywords associated with 20 or more proteins in Swiss-Prot were found and corresponding keyword-associated data sets of proteins were assembled. For each keyword-associated set, 1000 length-matching and number-matching sets of random proteins were drawn from Swiss-Prot. Order–disorder predictions were carried out for the keyword-associated sets and for the matching random sets. If a function described by a given keyword were carried out by a long region of disordered protein, one would expect the keyword-associated set to have a greater amount of predicted disorder compared with the matching random sets. The keyword-associated set would be expected to have a lower amount of predicted disorder compared with the random sets if the keyword-associated function were carried out by a structured protein. Given the predictions for the function-associated and matching random sets, it is possible to calculate the *P*-values, where a *P*-value > 0.95 was used to define a disorder-associated function and a *P*-value < 0.05 was used to define an order-associated function. Intermediate *P*-values are ambiguous [145–147]. The

application of this approach revealed that out of 710 Swiss-Prot keywords, 310 functional keywords were associated with ordered proteins, 238 functional keywords were attributed to disordered proteins, and the remaining 162 keywords yield ambiguity in the likely function–structure associations [145–147].

Interestingly, most of the structured protein-associated keywords were related to enzyme activities, which is a result in accordance with the previous discussion indicating that structure formation is for enzyme catalysis rather than just for molecular recognition. As for the disordered protein-associated keywords, most were related to signaling and regulation, again in accordance with the arguments given above that the thermodynamics of disorder-to-order upon binding are favorable to binding reversibility and thus to signaling.

Based on the analysis of then-known data the Protein Trinity paradigm was formulated [4, 21]. According to this model, functional intracellular proteins (or their functional regions) can exist in any of the three thermodynamic states: ordered, molten globule, and random coil. Function can arise from any of the three conformations and transitions between them. "In this view, not just the ordered state, but any of the three states can be the native state of a protein" [4]. Data presented in this review show that natively unfolded proteins, which were originally considered as random coils, are not uniform and can be grouped into two structurally different subclasses, which were designated as intrinsic coils and intrinsic pre-molten globules. These observations bring a new player, the native pre-molten globule, to the protein function field [22]. Therefore, it has been suggested that the Protein Trinity should be extended to the Protein Quartet model, with function arising from four specific conformations (ordered forms, molten globules, pre-molten globules, and random coils) and transitions between any two of the states [22].

1.12. *LESSON 10:* PROTEIN POSTTRANSLATIONAL MODIFICATIONS OCCUR AT DISORDERED REGIONS

In a study of the functions associated with more than 100 long disordered regions, many were found to contain sites of protein PTM [132, 133]. These PTMs included phosphorylation, acetylation, fatty acylation, methylation, glycosylation, ubiquitination, and ADP-ribosylation, suggesting the possibility that protein modifications commonly occur in regions of disorder. A particular advantage of disorder for regulatory and signaling regions is that changes, such as protein modification, lead to large-scale disorder-to-order structural transitions: such large-scale structural changes are not subtle and therefore could be an advantage for signaling and regulation as compared with the much smaller changes that would be expected from the decoration of an ordered protein structure.

Protein phosphorylation and dephosphorylation are crucial for signaling. Indeed, about one-third of eukaryotic proteins are phosphorylated [148]. Many sites of protein phosphorylation were found to be in regions structurally characterized as intrinsically disordered [132, 133]. This conclusion was based on several lines of evidence, such as a very small number of PDB structures for both the unphosphory-

lated and phosphorylated forms of the same protein [149, 150]; the fact that the residues of the phosphorylation site often have extended, irregular conformation consistent with disordered structure [150]; the fact that the segments containing phosphorylation sites not only lack secondary structure but are held in place by side chain burial and also by backbone hydrogen bonds to the surrounding kinase side chains [151–156]; the fact that regions flanking the sites of phosphorylation are enriched in the disorder-promoting amino acids [150]; the fact that the sequence complexity distribution of the residues flanking phosphorylation sites matches almost exactly the complexity distribution obtained for IDPs [150]; and the fact there is a high correspondence between the prediction of disorder and the occurrence of phosphorylation [150].

Ubiquitination, the reversible modification of proteins by the covalent attachment of ubiquitin, is implicated in the regulation of a variety of cellular processes and is involved in many diseases. Recently, 141 new ubiquitination sites were identified using a combination of liquid chromatography, mass spectrometry, and mutant yeast strains [157]. The detailed analysis of the sequence biases and structural preferences around known ubiquitination sites indicated that the properties of these sites were similar to those of IDRs. In agreement with this computational study, structural information about the ubiquitination sites is sparse. In fact, despite the large size of PDB, only 7% of currently known ubiquitination sites in yeast could be confidently mapped to protein structures. The analysis of 3-D structures of 32 homologous protein chains (with 15 of them being 100% identical with query proteins) containing 28 ubiquitination sites revealed that 10 ubiquitination sites were in crystal or interchain/intrachain contacts, and therefore the assignment of these sites to a specific structural element should be made with caution. Of the 18 sites that could be confidently assigned to ordered regions, 11 were located within coils (two of which were close to the observed disordered regions), four within helices, and three within strands. The majority of the sites within coils and helices were surface exposed and had high B-factor values indicating high flexibility [157]. It has also been pointed out that along with the lack of structural information for the majority of experimentally detected ubiquitination sites, there were several examples of ubiquitination sites located in the experimentally confirmed disordered regions [157]. Based on these observations it has been concluded that the involvement of flexible and disordered protein regions into various aspects of ubiquitination process provides a strong support for the functional importance of such regions.

In addition to protease digestion, ubiquitination, and phosphorylation, several other types of PTMs, such as acetylation, fatty acid acylation, and methylation, have also been observed to occur in regions of intrinsic disorder [132, 133, 147]. From these findings, it is tempting to suggest that sites of protein modification in eukaryotic cells universally or at least very commonly exhibit a preference for IDRs. For all of the examples discussed earlier, the modifying enzyme has to bind to and modify similar sites in a wide variety of proteins. If all the regions flanking these sites are disordered before binding to the modifying enzyme, it is easy to understand how a single enzyme could bind to and modify a wide variety of protein targets.

1.13. *LESSON 11:* DISORDERED REGIONS ARE PRIMARY TARGETS FOR AS

Alternative splicing is a process by which two or more mature mRNAs are produced from a single precursor pre-mRNA by the inclusion and omission of different segments [158, 159]. The "exons" are joined to form the mRNA and the "introns" are left out [160]. So far AS has been commonly observed only in multicellular eukaryotes [161]. For humans and other mammals, multiple proteins are often produced from a single gene since 40–60% the genes yield proteins via the AS mechanism [162–164]. It was hypothesized that AS very likely provides an important mechanism for enhancing protein diversity in multicellular eukaryotes [165]. AS has affects on a diversity of protein functions such as protein–protein interactions, ligand binding, and enzymatic activity [166–168]. Therefore, it comes as no surprise that abnormal AS has been associated with numerous human diseases, including myotonic dystrophy [169], Axoospermia [170], Alzheimer's disease [171], Parkinson's disease [172, 173], and cancer [174].

Removal of a piece of sequence from a structured protein would often lead to dysfunctional protein folding, most often causing loss of function (sometimes, however, the AS isoform of structured proteins can maintain function, albeit typically with a reduction in activity). Why, then, is the AS phenomenon so common in nature? The analysis of the effect of AS on structured proteins revealed that AS-induced alterations are generally small in size, are usually located on the protein surface, and are most often located in coil regions [175]. Given the small sizes and locations of the changes resulting from AS, the different splice variants were predicted to fold into the same overall structures, with only slight structural perturbations that could be functionally important [175, 176].

The structural implications given above are interesting, but only a small fraction of AS events have been mapped to structured proteins. Since 40–60% of mammalian (human) genes are estimated to undergo AS, and since there are several thousand mammalian proteins in PDB [44], we would expect to find several thousand examples to study. So far, however, despite exhaustive searches of PDB, only 20 examples have been reported [175]. Based on the failure to find a significant number of examples of AS that map to regions of structure, it was hypothesized that the protein folding problems discussed earlier would be solved for different isoforms if the alternatively spliced regions of mRNA were to code for regions of IDP. If AS were to map to ID regions, both multiple and long splice variants would be allowed because structural perturbation would not be a problem.

To test this hypothesis, a collection of human proteins with structurally characterized regions of order and disorder was built and an exhaustive search on AS for all of these proteins was performed. This generated a set of 46 human proteins with 75 alternatively spliced segments, all of which were located in structurally characterized regions [177]. Importantly, of these 75 alternatively spliced regions of RNA, 43 (57%) coded for entirely disordered protein regions, 18 (24%) coded for both ordered and disordered protein regions (with the splice boundaries very often in, or very near to, the disordered regions), and just 14 (19%) coded for fully structured

regions [177]. Next, to increase the number of examples, all the Swiss-Prot proteins labeled as having AS isoforms were identified. This approach generated a set of 558 proteins with 1266 regions that are absent from one isoform due to AS. Disorder/order propensities of these AS proteins and regions were predicted together with the disorder/order propensities of the 46 structurally characterized proteins and for their 75 regions that were affected by AS. This analysis revealed a perfect correlation between predictions and observations of disorder in the 46 structurally characterized proteins. For the 1266 regions from Swiss-Prot, the predicted abundance of disorder closely matched the corresponding predictions for the 75 with known structure. These data strongly suggest that AS occurs mostly in regions of RNA that code for disordered protein [177].

These findings have crucial functional implications. Since disorder plays various roles in protein functions and in protein–protein interaction networks, modification of such functions could be readily accomplished by AS within disordered regions. Thus, a linkage between AS and signaling by disordered regions provides a novel and plausible mechanism for understanding the origins of cell differentiation, which ultimately gives rise to multicellular organisms in nature [177].

1.14. *LESSON 12:* DISORDERED PROTEINS ARE TIGHTLY REGULATED IN THE LIVING CELLS

It is clear now that the IDPs are real, abundant, diversified, and vital. Functions of IDPs are complementary to the catalytic activities of ordered proteins [2–4, 21–26, 29, 53, 132, 133, 135, 145–147, 178, 179]. Many disorder-related functions (e.g., signaling, control, regulation, and recognition) are incompatible with well-defined, stable 3-D structures [2–5, 21, 22, 24–26, 29, 53, 135, 145–147, 179, 180]. Intrinsic disorder is assumed to provide several functional advantages for its carriers, including increased interaction surface area, structural plasticity to interact with several targets, high specificity for given partners combined with high k_{on} and k_{off} rates that enable rapid association with the partner without an excessive binding strength, and the ability to fold upon binding and accessible PTM sites.

The highly dynamic nature of IDPs is a visual illustration of the chaos. However, the evolutionary persistence of these highly dynamic proteins, their unique functionality, and involvement in all the major cellular processes evidence that this chaos is tightly controlled [181]. To answer the question of how these proteins are governed and regulated inside the cell, Gsponer et al. conducted a detailed study focused on the intricate mechanisms of the IDP regulation [182]. To this end, all the *S. cerevisiae* proteins were grouped into three classes using one of the available disorder predictors, DisoPred2 [183]: (1) 1971 highly ordered proteins containing 0–10% of the predicted disorder; (2) 2711 moderately disordered proteins with 10–30% predicted disordered residues; and (3) 2020 highly disordered proteins containing 30–100% of the predicted disorder. Then, the correlations between intrinsic disorder and the various regulation steps of protein synthesis and degradation were evaluated.

To examine the transcription of genes encoding IDPs and ordered proteins, the transcriptional rates and the degradation rates of the corresponding transcripts were compared [182]. This analysis revealed that the transcriptional rates of mRNAs encoding IDPs and ordered proteins were comparable. However, the IDP-encoding transcripts were generally less abundant than transcripts encoding ordered proteins due to the increased decay rates of the former.

Tight regulation of IDP abundance was also observed at the protein level. IDPs were shown to be less abundant than ordered proteins due to the lower rate of protein synthesis and shorter protein half-lives. As the abundance and half-life in a cell of certain proteins can be further modulated via their PTMs such as phosphorylation [184], the experimentally determined yeast kinase–substrate network was analyzed next. IDPs were shown to be substrates of twice as many kinases as were ordered proteins. Furthermore, the vast majority of kinases whose substrates were IDPs were either regulated in a cell-cycle dependent manner, or activated upon exposure to particular stimuli or stress [182]. Therefore, PTMs may not only serve as an important mechanism for the fine-tuning of the IDP functions, but possibly they are necessary to tune the IDP availability under different cellular conditions.

In addition to *S. cerevisiae*, similar regulation trends were also found in *Schizosaccharomyces pombe* and *Homo sapiens* [182]. Based on these observations it has been concluded that both unicellular and multicellular organisms appear to use similar mechanisms to regulate the availability of IDPs. Overall, the study by Gsponer et al. clearly demonstrated that there is an evolutionarily conserved tight control of synthesis and clearance of most IDPs. This tight control is directly related to the major roles of IDPs in signaling, where it is crucial to be available in appropriate amounts and not to be present longer than needed [182]. It has been also pointed out that although the abundance of many IDPs is under strict control, some IDPs could be present in cells in large amounts and/or for long periods either due to specific PTMs or via interaction with other factors, which could promote changes in cellular localization of IDPs or protect them from the degradation machinery [4, 147, 150, 184, 185]. Overall, it has been concluded that the chaos seemingly introduced into the protein world by the discovery of IDPs is under tight control [181].

In an independent study, a global-scale relationship between the predicted fraction of protein disorder and RNA and protein expression in *E. coli* was analyzed [186]. It has been shown that the fraction of protein disorder were positively correlated with both measured RNA expression levels of *Escherichia coli* genes in three different growth media (lysogeny broth (LB) rich medium and nitrogen- and carbon-free (N-C-) minimal media supplemented with glycerol as a carbon source and either ammonium chloride or arginine as a nitrogen source) and predicted abundance levels of *E. coli* proteins. When a subset of 216 *E. coli* proteins that are known to be essential for the survival and growth of this bacterium was analyzed, the correlation between protein disorder and expression level became even more evident. In fact, essential proteins had on average a much higher fraction of disorder (0.36), had a higher number of proteins classified as completely disordered (19% vs. 2% for *E. coli* proteome), and were expressed at a higher level in all three media than an average *E. coli* gene [186]. To better understand the function–disorder relationship

for highly expressed *E. coli* proteins, a literature search was carried out for a group of proteins that had high levels of predicted intrinsic disorder. This study revealed that the disorder predictions matched well with the experimentally elucidated regions of protein flexibility and disorder [186]. A direct link between protein disorder and protein level in *E. coli* cells could be because disordered proteins may carry out essential control and regulation functions that are needed to respond to various environmental conditions. Another possibility is that IDPs might undergo more rapid degradation compared with structured proteins, which cells can counter by increasing the mRNA levels of the corresponding genes. In this case, higher synthesis and degradation rates could make the levels of these proteins very sensitive to the environment, with slight changes in either production or degradation leading to significant shifts in protein levels [186].

More evidence for the tight control of IDPs inside the cell came from the analysis of cellular regulation of so-called vulnerable proteins [14]. The integrity of soluble protein functional structures is maintained in part by a precise network of hydrogen bonds linking the backbone amide and carbonyl groups. In a well-ordered protein, hydrogen bonds are shielded from water attack, preventing backbone hydration and the total or partial denaturation of the soluble structure under physiological conditions [187, 188]. Since soluble protein structures may be more or less vulnerable to water attack depending on their packing quality, a structural attribute, protein vulnerability, was introduced as the ratio of solvent-exposed backbone hydrogen bonds, which represent local weaknesses of the structure, to the overall number of hydrogen bonds [14]. It has also been pointed out that structural vulnerability can be related to protein intrinsic disorder as the inability of a particular protein fold to protect intramolecular hydrogen bonds from water attack may result in backbone hydration leading to local or global unfolding. Since binding of a partner can help to exclude water molecules from the microenvironment of the preformed bonds, a vulnerable soluble structure gains extra protection of its backbone hydrogen bonds through the complex formation [187]. To understand the role of structure vulnerability in transcriptome organization, the relationship between the structural vulnerability of a protein and the extent of co-expression of genes encoding its binding partners was analyzed [14]. This study revealed that structural vulnerability can be considered a determinant of transcriptome organization across tissues and temporal phases [14]. Finally, by interrelating vulnerability, disorder propensity, and co-expression patterns, the role of protein intrinsic disorder in transcriptome organization was confirmed since the correlation between the extent of intrinsic disorder of the most disordered domain in an interacting pair and the expression correlation of the two genes encoding the respective interacting domains was evident [14].

1.15. *LESSON 13:* NATIVELY UNFOLDED PROTEINS ARE FREQUENTLY ASSOCIATED WITH HUMAN DISEASES

Because IDPs play crucial roles in numerous biological processes, it was not too surprising to find that some of them are involved in human diseases. For example,

a number of human diseases originate from the deposition of stable, ordered, filamentous protein aggregates, commonly referred to as amyloid fibrils. In each of these abnormal protein deposition-mediated pathological states, a specific protein or protein fragment changes from its natural soluble form into insoluble fibrils, which accumulate in a variety of organs and tissues [189–195]. Approximately 20 different proteins are known so far to be involved in protein deposition diseases. These proteins are unrelated in terms of sequence or starting structure. Several IDPs are found in this list of 20 proteins, and they are associated with the development of several neurodegenerative diseases [195, 196]. An incomplete list of disorders associated with IDPs includes Alzheimer's disease (deposition of amyloid-β, tau-protein, α-synuclein fragment NAC [197–200]), Niemann–Pick disease type C, subacute sclerosing panencephalitis, argyrophilic grain disease, myotonic dystrophy, and motor neuron disease with neurofibrillary tangles (accumulation of tau-protein in the form of neurofibrillary tangles [199]); Down's syndrome (nonfilamentous amyloid-β deposits [201]); Parkinson's disease, dementia with LB, diffuse LB disease, LB variant of Alzheimer's disease, multiple system atrophy, and Hallervorden–Spatz disease (deposition of α-synuclein in the form of LB, or Lewy neuritis [202]); prion diseases (deposition of PrP^{SC} [203]); and a family of polyQ diseases, a group of neurodegenerative disorders caused by expansion of GAC trinucleotide repeats coding for polyQ in the gene products [204]. Furthermore, most mutations in rigid globular proteins associated with accelerated fibrillation and protein deposition diseases have been shown to destabilize the native structure, increasing the steady-state concentration of partially folded (disordered) conformers [189–195].

The disorders just mentioned are called conformational diseases, as they are characterized by conformational changes, misfolding, and aggregation of an underlying protein. However, there is another side to this coin: protein functionality. In fact, many of the proteins associated with the conformational disorders are also involved in recognition, regulation, and cell signaling. For example, functions ascribed to α-synuclein, a protein involved in several neurodegenerative disorders, include binding fatty acids and metal ions; regulation of certain enzymes, transporters, and neurotransmitter vesicles; and regulation of neuronal survival (reviewed in [202]). Overall, about 50 proteins and ligands interact and/or co-localize with this protein. Furthermore, α-synuclein has remarkable structural plasticity and adopts a series of different monomeric, oligomeric, and insoluble conformations (reviewed in Ref. [15]). The choice between these conformations is determined by the peculiarities of the protein environment, assuming that α-synuclein has an exceptional ability to fold in a template-dependent manner. Based on these observations, we hypothesize that the development of the conformational diseases may originate from the misidentification, misregulation, and mis-signaling, accompanied by misfolding. In other words, mutations and/or changes in the environment may result in protein confusion, for which the identity of the protein becomes lost, thus reducing its capability to recognize proper binding partners and leading to the formation of nonfunctional and deadly aggregates.

Recent analysis of so-called polyglytamine diseases gives support to this hypothesis [205]. Polyglytamine diseases are a specific group of hereditary neurodegenera-

tion caused by expansion of CAG triplet repeats in an exon of disease genes that leads to the production of a disease protein containing an expanded polyglutamine, polyQ, stretch. Nine neurodegenerative disorders, including Kennedy's disease, Huntington's diseases, spinocerebellar atrophy -1, -2, -3, -6, -7, and -17, and dentatorubral pallidoluysian atrophy belong to this class of diseases [206–209]. In most polyQ diseases, expansion to over 40 repeats leads to their onset [209]. It has been emphasized that molecular processes such as unfolded protein response, protein transport, synaptic transmission, and transcription are implicated in the pathology of polyQ diseases [205]. Importantly, more than 20 transcription-related factors have been reported to interact with pathological polyQ proteins. Furthermore, these interactions were shown to repress the transcription, leading eventually to neuronal dysfunction and death (reviewed in [205]). These results suggest that polyQ diseases represent kind of transcriptional disorder [205], supporting our misidentification hypothesis for at least some of the conformational disorders.

Disorder is also very common in cancer-associated proteins. In 2002, a study found that 79% of cancer-associated proteins and 66% of cell-signaling proteins contain predicted regions of disorder of 30 residues or longer [23]. In contrast, only 13% of a set of proteins with well-defined ordered structures contained such long regions of predicted disorder. Here, cancer-associated proteins were defined as those human proteins in Swiss-Prot containing the keyword "oncogene" (this included anti- and proto-oncogenes) or containing the word "tumor" in the description field. In experimental studies, the presence of disorder has been directly observed in several cancer-associated proteins, including p53 [210], $p57^{kip2}$ [211], Bcl-X_L and Bcl-2 [212], c-Fos [213], and most recently, a thyroid cancer-associated protein, TC-1 [214].

1.16. *LESSON 14:* NATIVELY UNFOLDED PROTEINS ARE ATTRACTIVE DRUG TARGETS

Since many proteins associated with various human diseases are either completely disordered or contain long disordered regions [215, 216], and since some of this disease-related IDPs/regions are involved in recognition, regulation, and signaling, these proteins/regions clearly represent novel potential drug targets. It is recognized now that the possibility of interrupting the action of disease-associated proteins (including through modulation of protein–protein interactions) presents an extremely attractive objective for the development of new drugs. The rational design of enzyme inhibitors depends on the classical view of protein function, which states that 3-D structure is an obligatory prerequisite for function. While generally applicable to many enzymatic domains, this view has persisted to influence views concerning all protein functions despite numerous examples to the contrary. Due to failure to recognize the important role of disorder in protein function, current and evolving methods of drug discovery suffer from an overly rigid view of protein function. This is most apparent in the observation that the vast majority of currently available drugs target the active site of enzymes, presumably since these are the only proteins for which the order-function paradigm is generally applicable.

Disordered proteins often bind their partners with contiguous residues of relatively short length, which become ordered upon binding [90, 217, 218]. Targeting small molecules to the disordered regions of proteins should enable the development of more effective drug discovery techniques. Drugs targeting these regions will likely function through inducing the disordered region to form an ordered structure dissimilar to its structure in complex with its binding partner, thereby preventing binding. The principles of small molecule binding to disordered regions have not been well studied, but sequence-specific, small-molecule binding to short peptides has been observed [219]. An interesting twist of this disorder-based approach for drug discovery is that using disordered regions as drug targets can be described as inducing order to *prevent* function.

In agreement with above-mentioned concepts, small molecules, "Nutlins," have been recently discovered that inhibit the p53–Mdm2 interaction by mimicking the helix in p53 that binds to Mdm2 [220, 221]. The tumor suppressor protein p53 is at the center of a large signaling network. It regulates expression of genes involved in numerous cellular processes, including cell cycle progression, apoptosis induction, DNA repair, as well as others involved in responding to cellular stress [222]. When p53 function is lost, either directly through mutation or indirectly through several other mechanisms, the cell often undergoes cancerous transformation [223, 224]. Cancers showing mutations in p53 are found in the colon, lung, esophagus, breast, liver, brain, reticuloendothelial tissues, and hemopoietic tissues [223].

p53 is regulated by several different mechanisms including inhibition of its activity by binding to E3 ubiquitin ligase Mdm2, which binds to a short stretch of p53, residues 13–29. This region of p53 is within the transactivation domain; thus p53 cannot activate or inhibit other genes when Mdm2 is bound. Mdm2 ubiquitinates p53 and thus targets it for destruction. Mdm2 also contains a nuclear export signal that causes p53 to be transported out of the nucleus. Although X-ray crystallographic studies of the p53–Mdm2 complex reveal that the Mdm2 binding region of p53 forms a helical structure that binds into a deep groove on the surface of Mdm2 [225], NMR studies show that the unbound N-terminal region of p53 lacks a fixed structure, although it does possess an amphipathic helix that forms a secondary structure part of the time [210] and therefore represents an illustrative example of the α-MoRF concept. This amphipathic helix seen in the unbound state is the same helix that binds to Mdm2. A close examination of the interface between the proteins reveals that Phe^{19}, Trp^{23}, and Leu^{26} of p53 are the major contributors to the interaction, with the side chains of these three amino acids pointing down into a crevice on the Mdm2 surface.

Because of the apparent simplicity of the interface, as well as the importance of the p53-Mdm2 interaction, this protein–protein interaction has been investigated as a possible drug target by many researchers. Several successful peptide inhibitors of the interaction have been created [226–229]. These peptides were all derived from the region of p53 that binds to Mdm2. Additionally all successful peptide inhibitors contained the three crucial residues involved in the interaction, Phe^{19}, Trp^{23}, and Leu^{26} [221].

Several small molecules were recently found to block the p53–Mdm2 interaction [221, 230–232]. While some of these were natural products, others were from a class

of *cis*-imidazolines called "Nutlins." These latter molecules increased the level of p53 in cancer cell lines. This drastically decreased the viability of these cells, causing most of them to undergo apoptosis. When one of the Nutlin compounds was given orally to mice, researchers saw a 90% inhibition of tumor growth compared with the control. The structure of Nutlin-2 was shown to mimic the crucial residues of p53, with two bromophenyl groups fitting into Mdm2 in the same pockets as Trp^{23} and Leu^{26}, and an ethyl–ether side chain filling the spot normally taken by Phe^{19} [230–232].

This research demonstrates that finding small molecules to target regions of proteins normally bound by disordered proteins is a feasible approach. It is anticipated that by studying this drug interaction, it will be possible to identify regions of other proteins that can be mimicked by small molecules. Remarkably, the disorder prediction for p53 using PONDR VL-XT software showed a sharp downward spike indicating predicted ordered region near the N-terminus of the protein. Furthermore, the α-MoRF identifier was able to recognize the region of p53 that binds to Mdm2 as a region of molecular recognition [115].

This successful Nutlin story marks the potential beginning of a new era, *the signaling–modulation era*, in targeting drugs to protein–protein interactions. Importantly, this druggable p53–Mdm2 interaction involves a disorder-to-order transition. Principles of such a transition are generally understood and therefore can use to find similar drug targets [233]. In addition to Nutlins, seven types of promising drug molecules that act by blocking protein–protein interactions have been described [234, 235]. While protein disorder is not mentioned in any of the papers describing how a small molecule can block protein–protein interactions, the disorder-based analysis revealed that four of these interactions involve one structured partner and one disordered partner, with 3 of the 4 disordered segments becoming helix upon binding. Therefore, the p53–Mdm2 complex is not the only member of this class currently known to be blocked by a small drug-like molecule. We fully expect many more examples to appear shortly, and for some of these examples to lead to useful drug molecules. Since p53–Mdm2-like interactions are likely to be very common, they clearly define a cornucopia of new drug targets that would operate by blocking disorder-based protein–protein interactions.

For these examples, the drug molecules mimic a critical region of the disordered partner, which folds upon binding, and compete with this region for its binding site on the structured partner. We argue that these druggable sites are likely to operate by the coupled binding and folding mechanism and utilize interaction sites that are small and compact enough to be easily mimicked by small molecules. We have developed methods for predicting such binding sites in disordered regions [236] and have elaborated the bioinformatics tools to identify which disordered binding regions can be easily mimicked by small molecules [233].

A complementary approach to finding small molecules inhibiting disorder-based protein–protein interactions has been developed in the laboratory of Prof. Steven Metallo (see, e.g., [237]). Deregulation of the c-Myc transcription factor is involved in many types of cancer, making this oncoprotein an attractive target for drug discovery. In order to bind DNA, regulated target gene expression, and function in most

biological contexts, c-Myc must dimerize with its obligate heterodimerization partner, Max, which lacks a transactivation segment. c-Myc, which is intrinsically disordered as a monomer, undergoes coupled binding and folding of its basic-helix-loop-helix-leucine zipper domain (bHLHZip) upon heterodimerization with its partner protein Max. One approach to c-Myc inhibition has been to disrupt this dimeric complex. In a search for effective inhibitors of the c-Myc–Max interactions, a series of small molecules was analyzed to find seven Myc inhibitors, which were shown to bind to one of three discrete sites within the 85-residue bHLHZip domain of c-Myc. These binding sites are composed of short contiguous stretches of amino acids that can selectively and independently bind small molecules. Inhibitor binding induces only local conformational changes, preserves the overall disorder of c-Myc, and inhibits dimerization with Max. Furthermore, binding of inhibitors to c-Myc was shown to occur simultaneously and independently on the three independent sites. Based on these observations it has been concluded that a rational and generic approach to the inhibition of protein–protein interactions involving IDPs may therefore be possible through the targeting of intrinsically disordered sequence [237].

Ideally, a drug that targets a given protein–protein interaction should be tissue specific. Although some proteins are unique for a given tissue, many more proteins have very wide distribution, being present in several tissues and organs. How can one develop tissue-specific drugs targeting such abundant proteins? Often, tissue specificity for many of the abundant proteins is achieved via the AS of the corresponding pre-mRNAs, which generates two or more protein isoforms from a single gene. Estimates indicate that between 35% and 60% of human genes yield protein isoforms by means of alternatively spliced mRNA [162]. The added protein diversity from AS is thought to be important for tissue-specific signaling and regulatory networks in multicellular organisms. Recently, it has been established that the regions of AS in proteins are enriched in intrinsic disorder [177]. Since disorder is frequently utilized in protein binding regions, having AS of pre-mRNA coupled to regions of protein disorder was proposed to lead to tissue-specific signaling and regulatory diversity [177]. Therefore, associating AS with protein disorder enables the time- and tissue-specific modulation of protein function. These findings open up a unique opportunity to develop tissue-specific drugs modulating the function of a given IDP/IDR (with a unique profile of disorder distribution) in a target tissue and not affecting the functionality of this same protein (with different disorder distribution profile) in other tissues.

1.17. *LESSON 15:* BRIGHT FUTURE OF FUZZY PROTEINS

The last 10–15 years witnessed a real revolution in our understanding of the protein structure–function relationships. The fact that there is an entire class of polypeptides that do not have rigid structures but possess crucial biological function was underappreciated and ignored for a long time despite numerous examples scattered in the literature. The work that started as an attempt to understand what is special about several natively unfolded proteins characterized in our group produced a real explo-

sion of interest in non-folded proteins with biological functions. A new field was created and a lot of intriguing information was produced related to the IDPs in general and to the natively unfolded proteins in particular. This chapter discussed some general concepts related to the natively unfolded proteins and represents some of our data to support these concepts.

Based on the data summarized in this chapter, natively unfolded proteins are characterized by low overall hydrophobicity and high net charge. They possess large hydrodynamic volumes and low contents of ordered secondary structure. They are characterized by the absence of a tightly packed core. They are very flexible but may adopt relatively rigid conformations in the presence of natural ligands. In comparison with ordered globular proteins, natively unfolded polypeptides possess "turn out" responses to changes in the environment, as their structural complexities increase at high temperatures or at extreme pHs.

An intriguing property of natively unfolded proteins is their ability to undergo disorder-to-order transition upon function. The degree of these structural rearrangements varies over a very wide range, from coil to pre-molten globule transitions to formation of rigid ordered structures.

Multiple roles of natively unfolded proteins in pathogenesis of human diseases should not be ignored. Because of the numerous structural adjustments, perturbations, interactions, and functions ascribed to natively unfolded proteins, it is reasonable to suggest that they are potentially prone to misfold. In such cases, the development of different diseases associated with natively unfolded proteins (e.g., various synucleinopathies) may originate from the misregulation, missignaling, and misidentification of a corresponding protein, accompanied by or resulting from its misfolding. In fact, mutations and/or changes in the environment may reduce the capability of a natively unfolded protein to recognize proper binding partners, thus leading to the formation of nonfunctional and deadly aggregates.

More than a decade of intensive studies on IDPs revealed a number of unique features related to their structural properties, abundance, distribution, functional repertoire, regulation, involvement in disease pathogenesis, and so on. However, the amount of data produced so far is just a small tip of a humongous iceberg. IDPs continue to bring discoveries on a regular basis. More discoveries and breakthroughs are expected in future due to advances in novel experimental and computational tools for studies focused on IDPs. Modern protein science is at a turning point.

ACKNOWLEDGMENTS

This work would be impossible without numerous collaborators whose enthusiasm and help drove the studies on natively unfolded proteins for many years. In no particular order the list of people contributed to this project at its different stages includes: Eugene Permyakov, Joel Gillespie, Vyacheslav Abramov, Anthony Fink, Larissa Munishkina, Pierre Souillac, Sebastian Doniach, Ian Millett, Daniel Denning, Kevin Lee, Alexander Sigalov, Michael Rexach, Sergei Permyakov, Keith Oberg, Stefan Winter, Jie Li, Oxana Galzitskaya, Leonhard Kittler, Gunter Lober, Olga

Tcherkasskaya, Seung-Jae Lee, Min Zhu, Amy Manning-Bog, Alison McCormack, Donato Di Monte, Kiowa Bower, I-Hsuan Liu, Gregory Cole, John Goers, Ghiam Yamin, Charles Glaser, Elisa Cooper, Jeffrey Cohlberg, Mark Hokenson, Sudha Rajamani, Joseph Zbilut, Alessandro Giuliani, Alfredo Colosimo, Julie Mitchell, Mauro Colafranceschi, Norbert Marwan, Charles Webber, Jr., Christopher Oldfield, Yugong Cheng, Marc Cortese, Celeste Brown, Mikhail Karymov, Yuri Lyubchenko, Pedro Romero, Lilia Iakoucheva, Zoran Obradovic, Véronique Receveur-Bréchot, Jean-Marie Bourhis, Bruno Canard, Sonia Longhi, Chad McAllister, Yoshiko Kawano, Alexander Lushnikov, Andrew Mikheikin, Andrey Vartapetyan, Predrag Radivojac, Slobodan Vucetic, Timothy R. O'Connor, Jag Bhalla, Geoffrey Storchan, Caitlin MacCarthy, Maureen Harrington, Tanguy LeGall, Patrizia Polverino de Laureto, Laura Tosatto, Erica Frare, Oriano Marin, Angelo Fontana, Megan Sickmeier, Justin Hamilton, Vladimir Vacic, Agnes Tantos, Beata Szabo, Peter Tompa, Jake Chen, King Pan Ng, Gary Potikyan, Rupert Savene, Christopher Denny, Vinay Singh, Yue Zhou, Joseph Marsh, Julie Forman-Kay, Jingwen Liu, Zongchao Jia, Millie Georgiadis, Amrita Mohan, Ann Roman, Jiangang Liu, Narayanan Perumal, Eric Su, Fei Ji, Niels Klitgord, Michael Cusick, Marc Vidal, Chad Haynes, Saima Zaidi, Ya Yin Fang, and Jessica Chen. We are grateful to all former and current colleagues for their priceless contributions, assistance, and support. This work was supported in part by grants R01 LM007688-01A1 (to A. K. D and V. N. U.) and GM071714-01A2 (to A. K. D and V. N. U.) from the National Institutes of Health and the Program of the Russian Academy of Sciences for Molecular and Cellular Biology (to V. N. U.). We gratefully acknowledge the support of the Indiana University—Purdue University Indianapolis Signature Centers Initiative.

REFERENCES

1 Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: Theory, predictions and observations, *Pac Symp Biocomput*, 473–484.

2 Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm, *J Mol Biol 293*, 321–331.

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

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

5 Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem Sci 27*, 527–533.

6 Daughdrill, G. W., Pielak, G. J., Uversky, V. N., Cortese, M. S., and Dunker, A. K. (2005) Natively disordered proteins. In: Buchner, J. and Kiefhaber, T. (eds.), *Handbook of protein folding*. Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 271–353.

7 Uversky, V. N. and Dunker, A. K. (2010) Understanding protein non-folding, *Biochim Biophys Acta 1804*, 1231–1264.

8 Holt, C. and Sawyer, L. (1993) Caseins as rheomorphic proteins: Interpretation of primary and secondary structures of the as1-, b-, and k-caseins, *J Chem Soc Faraday Trans 89*, 2683–2692.

9 Pullen, R. A., Jenkins, J. A., Tickle, I. J., Wood, S. P., and Blundell, T. L. (1975) The relation of polypeptide hormone structure and flexibility to receptor binding: The relevance of X-ray studies on insulins, glucagon and human placental lactogen, *Mol Cell Biochem 8*, 5–20.

10 Cary, P. D., Moss, T., and Bradbury, E. M. (1978) High-resolution proton-magnetic-resonance studies of chromatin core particles, *Eur J Biochem 89*, 475–482.

11 Linderstrom-Lang, K. and Schellman, J. A. (1959) Protein structure and enzyme activity. In: Boyer, P. D., Lardy, H., and Myrback, K. (eds.), *The enzymes*, 2nd ed., Academic Press, New York, pp. 443–510.

12 Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994) Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure, *J Biol Chem 269*, 24290–24297.

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

14 Chen, J., Liang, H., and Fernandez, A. (2008) Protein structure protection commits gene expression patterns, *Genome Biol 9*, R107.

15 Uversky, V. N. (2003) A protein-chameleon: Conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders, *J Biomol Struct Dyn 21*, 211–234.

16 Fuxreiter, M., Tompa, P., Simon, I., Uversky, V. N., Hansen, J. C., and Asturias, F. J. (2008) Malleable machines take shape in eukaryotic transcriptional regulation, *Nat Chem Biol 4*, 728–737.

17 Tsvetkov, P., Asher, G., Paz, A., Reuven, N., Sussman, J. L., Silman, I., and Shaul, Y. (2008) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome, *Proteins 70*, 1357–1366.

18 Dunker, A. K. and Uversky V. N. (2010) Drugs for "protein clouds": targeting intrinsically disordered transcription factors, *Curr Opin Pharmacol 10*(6), 782–788.

19 Livesay, D. R. (2010) Protein dynamics: Dancing on an ever-changing free energy stage, *Curr Opin Pharmacol 10*, 706–708.

20 Uversky, V. N. (2010) The mysterious unfoldome: Structureless, underappreciated, yet vital part of any given proteome, *J Biomed Biotechnol 2010*, 568068.

21 Dunker, A. K. and Obradovic, Z. (2001) The protein trinity—Linking function and disorder, *Nat Biotechnol 19*, 805–806.

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

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

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

25 Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling, *J Mol Recognit 18*, 343–384.

26 Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophys J 92*, 1439–1456.

27 Halle, B. (2002) Flexibility and packing in proteins, *Proc Natl Acad Sci U S A 99*, 1274–1279.

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

29 Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., Vacic, V., Obradovic, Z., and Uversky, V. N. (2008) The unfoldomics decade: An update on intrinsically disordered proteins, *BMC Genomics 9*, Suppl 2, S1.

30 Uversky, V. N. (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases, *Front Biosci 14*, 5188–5238.

31 Uversky, V. N. (2008) Alpha-synuclein misfolding and neurodegenerative diseases, *Curr Protein Pept Sci 9*, 507–540.

32 Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding, *Proteins 62*, 24–45.

33 Hemmings, H. C., Jr, Nairn, A. C., Aswad, D. W., and Greengard, P. (1984) DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus, *J Neurosci 4*, 99–110.

34 Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Muller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) Prothymosin alpha: A biologically active protein with random coil conformation, *Biochemistry 34*, 13211–13218.

35 Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting disordered regions from amino acid sequence: Common themes despite differing structural characterization, *Genome Inform Ser Workshop Genome Inform 9*, 201–213.

36 Williams, R. M., Obradovi, Z., Mathura, V., Braun, W., Garner, E. C., Young, J., Takayama, S., Brown, C. J., and Dunker, A. K. (2001) The protein non-folding problem: Amino acid determinants of intrinsic order and disorder, *Pac Symp Biocomput*, 89–100.

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

38 Vacic, V., Uversky, V. N., Dunker, A. K., and Lonardi, S. (2007) Composition profiler: A tool for discovery and visualization of amino acid composition differences, *BMC Bioinformatics 8*, 211.

39 Li, X., Obradovic, Z., Brown, C. J., Garner, E. C., and Dunker, A. K. (2000) Comparing predictors of disordered protein, *Genome Inform Ser Workshop Genome Inform 11*, 172–184.

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

41 He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N., and Dunker, A. K. (2009) Predicting intrinsic disorder in proteins: An overview, *Cell Res 19*, 929–949.

42 Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions, *Pac Symp Biocomput*, 437–448.

43 Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inform Ser Workshop Genome Inform 11*, 161–171.

44 Le Gall, T., Romero, P. R., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder in the Protein Data Bank, *J Biomol Struct Dyn 24*, 325–342.

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

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

47 Ptitsyn, O. B. (1995) Molten globule and protein folding, *Adv Protein Chem 47*, 83–229.

48 Uversky, V. N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: Which way to go? *Cell Mol Life Sci 60*, 1852–1871.

49 Tcherkasskaya, O. and Uversky, V. N. (2001) Denatured collapsed states in protein folding: Example of apomyoglobin, *Proteins 44*, 244–254.

50 Ptitsyn, O. B. (1994) Kinetic and equilibrium intermediates in protein folding, *Protein Eng 7*, 593–596.

51 Tanford, C. (1968) Protein denaturation, *Adv Protein Chem 23*, 121–282.

52 Uversky, V. N. (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry 32*, 13288–13298.

53 Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur J Biochem 269*, 2–12.

54 Glatter, O. and Kratky, O. (1982) *Small angle X-ray scattering*, Academic Press, London.

55 Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., Ikura, T., and Kuwajima, K. (1996) Protein globularization during folding. A study by synchrotron small-angle X-ray scattering, *J Mol Biol 262*, 559–574.

56 Uversky, V. N., Li, J., and Fink, A. L. (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation, *J Biol Chem 276*, 10737–10744.

57 Uversky, V. N., Li, J., Souillac, P., Millett, I. S., Doniach, S., Jakes, R., Goedert, M., and Fink, A. L. (2002) Biophysical properties of the synucleins and their propensities to fibrillate: Inhibition of alpha-synuclein assembly by beta- and gamma-synucleins, *J Biol Chem 277*, 11970–11978.

58 Bhattacharyya, J. and Das, K. P. (1999) Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein, *J Biol Chem 274*, 15505–15509.

59 Thomas, J., Van Patten, S. M., Howard, P., Day, K. H., Mitchell, R. D., Sosnick, T., Trewhella, J., Walsh, D. A., and Maurer, R. A. (1991) Expression in *Escherichia coli*

and characterization of the heat-stable inhibitor of the cAMP-dependent protein kinase, *J Biol Chem 266*, 10906–10911.

60 Denning, D. P., Uversky, V., Patel, S. S., Fink, A. L., and Rexach, M. (2002) The *Saccharomyces cerevisiae* nucleoporin Nup2p is a natively unfolded protein, *J Biol Chem 277*, 33447–33455.

61 Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., and Rexach, M. (2003) Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded, *Proc Natl Acad Sci U S A 100*, 2450–2455.

62 Dyson, H. J. and Wright, P. E. (2001) Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states, *Methods Enzymol 339*, 258–270.

63 Dyson, H. J. and Wright, P. E. (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance, *Adv Protein Chem 62*, 311–340.

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

65 Eliezer, D. (2007) Characterizing residual structure in disordered protein States using nuclear magnetic resonance, *Methods Mol Biol 350*, 49–67.

66 Eliezer, D. (2009) Biophysical characterization of intrinsically disordered proteins, *Curr Opin Struct Biol 19*, 23–30.

67 Muller-Spath, S., Soranno, A., Hirschfeld, V., Hofmann, H., Ruegger, S., Reymond, L., Nettels, D., and Schuler, B. (2010) From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins, *Proc Natl Acad Sci U S A 107*, 14609–14614.

68 Yamada, J., Phillips, J. L., Patel, S., Goldfien, G., Calestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V. V., Newsam, S., Gopinathan, A., Lau, E. Y., Colvin, M. E., Uversky, V. N., and Rexach, M. F. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins, *Mol Cell Proteomics 9*, 2205–2224.

69 Lau, K. F. and Dill, K. A. (1990) Theory for protein mutability and biogenesis, *Proc Natl Acad Sci U S A 87*, 638–642.

70 Ptitsyn, O. (1995) Molten globule and protein folding, *Adv Protein Chem 47*, 83–229.

71 Tcherkasskaya, O. and Uversky, V. N. (2003) Polymeric aspects of protein folding: A brief overview, *Protein Pept Lett 10*, 239–245.

72 Vitalis, A. and Pappu, R. V. (2009) ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions, *J Comput Chem 30*, 673–699.

73 Vitalis, A., Wang, X., and Pappu, R. V. (2007) Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories, *Biophys J 93*, 1923–1937.

74 Vitalis, A., Wang, X., and Pappu, R. V. (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization, *J Mol Biol 384*, 279–297.

75 Wang, X., Vitalis, A., Wyczalkowski, M. A., and Pappu, R. V. (2006) Characterizing the conformational ensemble of monomeric polyglutamine, *Proteins 63*, 297–311.

76 Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin, *J Mol Biol 396*, 1295–1309.

77 Tran, H. T., Mao, A., and Pappu, R. V. (2008) Role of backbone–solvent interactions in determining conformational equilibria of intrinsically disordered proteins, *J Am Chem Soc 130*, 7380–7392.

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

79 Ptitsyn, O. B. and Uversky, V. N. (1994) The molten globule is a third thermodynamical state of protein molecules, *FEBS Lett 341*, 15–18.

80 Uversky, V. N. and Ptitsyn, O. B. (1996) All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins, *Fold Des 1*, 117–122.

81 Neyroz, P., Zambelli, B., and Ciurli, S. (2006) Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy, *Biochemistry 45*, 8918–8930.

82 Permyakov, S. E., Millett, I. S., Doniach, S., Permyakov, E. A., and Uversky, V. N. (2003) Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin, *Proteins 53*, 855–862.

83 Uversky, V. N., Permyakov, S. E., Zagranichny, V. E., Rodionov, I. L., Fink, A. L., Cherskaya, A. M., Wasserman, L. A., and Permyakov, E. A. (2002) Effect of zinc and temperature on the conformation of the gamma subunit of retinal phosphodiesterase: A natively unfolded protein, *J Proteome Res 1*, 149–159.

84 Uversky, V. N., Gillespie, J. R., Millett, I. S., Khodyakova, A. V., Vasiliev, A. M., Chernovskaya, T. V., Vasilenko, R. N., Kozlovskaya, G. D., Dolgikh, D. A., Fink, A. L., Doniach, S., and Abramov, V. M. (1999) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH, *Biochemistry 38*, 15009–15016.

85 Konno, T., Tanaka, N., Kataoka, M., Takano, E., and Maki, M. (1997) A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I, *Biochim Biophys Acta 1342*, 73–82.

86 Lynn, A., Chandra, S., Malhotra, P., and Chauhan, V. S. (1999) Heme binding and polymerization by *Plasmodium falciparum* histidine rich protein II: Influence of pH on activity and conformation, *FEBS Lett 459*, 267–271.

87 Johansson, J., Gudmundsson, G. H., Rottenberg, M. E., Berndt, K. D., and Agerberth, B. (1998) Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37, *J Biol Chem 273*, 3718–3724.

88 Uversky, V. N. (2011) Multitude of binding modes attainable by intrinsically disordered proteins: A portrait gallery of disorder-based complexes, *Chem Soc Rev 40*, 1623–1634.

89 Mohan, A. (2006) MoRFs: A dataset of molecular recognition features. *Master's thesis*, Indiana University, School of Informatics, p. 59.

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

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

92 Galea, C. A., Nourse, A., Wang, Y., Sivakolundu, S. G., Heller, W. T., and Kriwacki, R. W. (2008) Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1, *J Mol Biol 376*, 827–838.

93 Galea, C. A., Wang, Y., Sivakolundu, S. G., and Kriwacki, R. W. (2008) Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits, *Biochemistry 47*, 7598–7609.

94 Graham, T. A., Weaver, C., Mao, F., Kimelman, D., and Xu, W. (2000) Crystal structure of a beta-catenin/Tcf complex, *Cell 103*, 885–896.

95 Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes, *Genome Biol 1*, REVIEWS001.

96 Brodersen, D. E., Clemons, W. M., Jr, Carter, A. P., Wimberly, B. T., and Ramakrishnan, V. (2002) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16 S RNA, *J Mol Biol 316*, 725–768.

97 Teschke, C. M. and King, J. (1992) Folding and assembly of oligomeric proteins in *Escherichia coli*, *Curr Opin Biotechnol 3*, 468–473.

98 Xu, D., Tsai, C. J., and Nussinov, R. (1998) Mechanism and evolution of protein dimerization, *Protein Sci 7*, 533–544.

99 Gunasekaran, K., Tsai, C. J., and Nussinov, R. (2004) Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers, *J Mol Biol 341*, 1327–1341.

100 Mason, J. M. and Arndt, K. M. (2004) Coiled coil domains: Stability, specificity, and biological implications, *Chembiochem 5*, 170–176.

101 Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., and Lu, M. (2006) A seven-helix coiled coil, *Proc Natl Acad Sci U S A 103*, 15457–15462.

102 Wolf, E., Kim, P. S., and Berger, B. (1997) MultiCoil: A program for predicting two- and three-stranded coiled coils, *Protein Sci 6*, 1179–1189.

103 Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) High-resolution protein design with backbone freedom, *Science 282*, 1462–1467.

104 Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J., and Kammerer, R. A. (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer, *Nat Struct Biol 7*, 772–776.

105 Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: The crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J 21*, 5960–5968.

106 Low, H. H., Moncrieffe, M. C., and Lowe, J. (2004) The crystal structure of ZapA and its modulation of FtsZ polymerisation, *J Mol Biol 341*, 839–852.

107 Zhao, X., Ghaffari, S., Lodish, H., Malashkevich, V. N., and Kim, P. S. (2002) Structure of the Bcr-Abl oncoprotein oligomerization domain, *Nat Struct Biol 9*, 117–120.

108 Liu, X., Xu, L., Liu, Y., Tong, X., Zhu, G., Zhang, X. C., Li, X., and Rao, Z. (2009) Crystal structure of the hexamer of human heat shock factor binding protein 1, *Proteins 75*, 1–11.

109 Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milhollen, M. A., Pang, J. X., and Kim, P. S. (1999) Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution, *Proc Natl Acad Sci U S A 96*, 2662–2667.

110 Malashkevich, V. N., Singh, M., and Kim, P. S. (2001) The trimer-of-hairpins motif in membrane fusion: Visna virus, *Proc Natl Acad Sci U S A 98*, 8502–8506.

111 Glover, J. N. and Harrison, S. C. (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA, *Nature 373*, 257–261.

112 Im, Y. J., Kang, G. B., Lee, J. H., Park, K. R., Song, H. E., Kim, E., Song, W. K., Park, D., and Eom, S. H. (2010) Structural basis for asymmetric association of the betaPIX coiled coil and shank PDZ, *J Mol Biol 397*, 457–466.

113 Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U., and Moarefi, I. (2000) Structure of the molecular chaperone prefoldin: Unique interaction of multiple coiled coil tentacles with unfolded proteins, *Cell 103*, 621–632.

114 Lee, S., Sowa, M. E., Watanabe, Y. H., Sigler, P. B., Chiu, W., Yoshida, M., and Tsai, F. T. (2003) The structure of ClpB: A molecular chaperone that rescues proteins from an aggregated state, *Cell 115*, 229–240.

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

116 Kajava, A. V., Baxa, U., and Steven, A. C. (2010) Beta arcades: Recurring motifs in naturally occurring and disease-related amyloid fibrils, *FASEB J 24*, 1311–1319.

117 Borg, M., Mittag, T., Pawson, T., Tyers, M., Forman-Kay, J. D., and Chan, H. S. (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity, *Proc Natl Acad Sci U S A 104*, 9650–9655.

118 Mittag, T., Orlicky, S., Choy, W. Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., and Forman-Kay, J. D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, *Proc Natl Acad Sci U S A 105*, 17772–17777.

119 Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010) Protein dynamics and conformational disorder in molecular recognition, *J Mol Recognit 23*, 105–116.

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

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

122 Sigalov, A., Aivazian, D., and Stern, L. (2004) Homooligomerization of the cytoplasmic domain of the T cell receptor zeta chain and of other proteins containing the immunoreceptor tyrosine-based activation motif, *Biochemistry 43*, 2049–2061.

123 Sigalov, A. B., Zhuravleva, A. V., and Orekhov, V. Y. (2007) Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form, *Biochimie 89*, 419–421.

124 Pometun, M. S., Chekmenev, E. Y., and Wittebort, R. J. (2004) Quantitative observation of backbone disorder in native elastin, *J Biol Chem 279*, 7982–7987.

125 Sigalov, A. B. and Hendricks, G. M. (2009) Membrane binding mode of intrinsically disordered cytoplasmic domains of T cell receptor signaling subunits depends on lipid composition, *Biochem Biophys Res Commun 389*, 388–393.

126 Sigalov, A. B., Aivazian, D. A., Uversky, V. N., and Stern, L. J. (2006) Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits, *Biochemistry 45*, 15731–15739.

127 Tompa, P. and Fuxreiter, M. (2008) Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions, *Trends Biochem Sci 33*, 2–8.

128 Hazy, E. and Tompa, P. (2009) Limitations of induced folding in molecular recognition by intrinsically disordered proteins, *Chemphyschem 10*, 1415–1419.

129 Liu, Y., Fratini, E., Baglioni, P., Chen, W. R., and Chen, S. H. (2005) Effective long-range attraction between protein molecules in solutions studied by small angle neutron scattering, *Phys Rev Lett 95*, 118102.

130 Schulz, G. E. (1979) Nucleotide binding proteins. In: Balaban, M. (ed.), *Molecular mechanism of biological recognition*. Elsevier/North-Holland Biomedical Press, New York, pp. 79–94.

131 Pontius, B. W. (1993) Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association, *Trends Biochem Sci 18*, 181–186.

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

133 Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins, *Adv Protein Chem 62*, 25–49.

134 Dyson, H. J. and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr Opin Struct Biol 12*, 54–60.

135 Dyson, H. J. and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol 6*, 197–208.

136 Plaxco, K. W. and Gross, M. (1997) Cell biology. The importance of being unfolded, *Nature 386*, 657, 659.

137 Spolar, R. S. and Record, M. T., Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777–784.

138 Karush, F. (1950) Heterogeneity of the binding sites of bovine serum albumin, *J Am Chem Soc 72*, 2705–2713.

139 Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: Conformational disorder mediates binding diversity, *Proc Natl Acad Sci U S A 93*, 11504–11509.

140 Dunker, A. K., Obradovic, Z., Romero, P., Kissinger, C., and Villafranca, E. (1997) On the importance of being disordered, *PDB Newsletter 81*, 3–5.

141 Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Protein Struct Funct Genet 42*, 38–49.

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

143 Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2008) Intrinsic disorder in scaffold proteins: Getting more from less, *Prog Biophys Mol Biol 98*, 85–106.

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

145 Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Uversky, V. N., and Obradovic, Z. (2007) Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions, *J Proteome Res 6*, 1882–1898.

146 Vucetic, S., Xie, H., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (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*, 1899–1916.

147 Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins, *J Proteome Res 6*, 1917–1932.

148 Marks, F. (1996) *Protein phosphorylation*, Wiley, VCH Weinheim, New York, Basel, Cambridge, Tokyo.

149 Johnson, L. N. and Lewis, R. J. (2001) Structural basis for control by phosphorylation, *Chem Rev 101*, 2209–2242.

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

151 Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H., and Huber, R. (1993) Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24), *EMBO J 12*, 849–859.

152 Narayana, N., Cox, S., Shaltiel, S., Taylor, S. S., and Xuong, N. (1997) Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5-24) and adenosine, *Biochemistry 36*, 4438–4448.

153 Lowe, E. D., Noble, M. E., Skamnaki, V. T., Oikonomakos, N. G., Owen, D. J., and Johnson, L. N. (1997) The crystal structure of a phosphorylase kinase peptide substrate complex: Kinase substrate recognition, *Embo J 16*, 6646–6658.

154 ter Haar, E., Coll, J. T., Austen, D. A., Hsiao, H. M., Swenson, L., and Jain, J. (2001) Structure of GSK3beta reveals a primed phosphorylation mechanism, *Nat Struct Biol 8*, 593–596.

155 Hubbard, S. R. (1997) Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog, *Embo J 16*, 5572–5581.

156 McDonald, I. K. and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins, *J Mol Biol 238*, 777–793.

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

158 Sambrook, J. (1977) Adenovirus amazes at Cold Spring Harbor, *Nature 268*, 101–104.

159 Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing, *Annu Rev Biochem 72*, 291–336.

160 Gilbert, W. (1978) Why genes in pieces? *Nature 271*, 501.

161 Ast, G. (2004) How did alternative splicing evolve? *Nat Rev Genet 5*, 773–782.

162 Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T. A., and Soreq, H. (2005) Function of alternative splicing, *Gene 344*, 1–20.

163 Brett, D., Hanke, J., Lehmann, G., Haase, S., Delbruck, S., Krueger, S., Reich, J., and Bork, P. (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms, *FEBS Lett 474*, 83–86.

164 Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., Santos, R., Schadt, E. E., Stoughton, R., and Shoemaker, D. D. (2003) Genome-wide

survey of human alternative pre-mRNA splicing with exon junction microarrays, *Science 302*, 2141–2144.

165 Graveley, B. R. (2001) Alternative splicing: Increasing diversity in the proteomic world, *Trends Genet 17*, 100–107.

166 Minneman, K. P. (2001) Splice variants of G protein-coupled receptors, *Mol Interv 1*, 108–116.

167 Thai, T. H. and Kearney, J. F. (2004) Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants, *J Immunol 173*, 4009–4019.

168 Scheper, W., Zwart, R., and Baas, F. (2004) Alternative splicing in the N-terminus of Alzheimer's presenilin 1, *Neurogenetics 5*, 223–227.

169 Roberts, R., Timchenko, N. A., Miller, J. W., Reddy, S., Caskey, C. T., Swanson, M. S., and Timchenko, L. T. (1997) Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice, *Proc Natl Acad Sci U S A 94*, 13221–13226.

170 Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W. A., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K., et al. (1993) A Y chromosome gene family with RNA-binding protein homology: Candidates for the azoospermia factor AZF controlling human spermatogenesis, *Cell 75*, 1287–1295.

171 Lovestone, S., Reynolds, C. H., Latimer, D., Davis, D. R., Anderton, B. H., Gallo, J. M., Hanger, D., Mulot, S., Marquardt, B., Stabel, S., et al. (1994) Alzheimer's disease-like phosphorylation of the microtubule-associated protein tau by glycogen synthase kinase-3 in transfected mammalian cells, *Curr Biol 4*, 1077–1086.

172 Beyer, K., Domingo-Sabat, M., Humbert, J., Carrato, C., Ferrer, I., and Ariza, A. (2008) Differential expression of alpha-synuclein, parkin, and synphilin-1 isoforms in Lewy body disease, *Neurogenetics 9*, 163–172.

173 Beyer, K., Domingo-Sabat, M., Lao, J. I., Carrato, C., Ferrer, I., and Ariza, A. (2008) Identification and characterization of a new alpha-synuclein isoform and its role in Lewy body diseases, *Neurogenetics 9*, 15–23.

174 Venables, J. P. (2004) Aberrant and alternative splicing in cancer, *Cancer Res 64*, 7647–7654.

175 Wang, P., Yan, B., Guo, J. T., Hicks, C., and Xu, Y. (2005) Structural genomics analysis of alternative splicing and application to isoform structure modeling, *Proc Natl Acad Sci U S A 102*, 18920–18925.

176 Furnham, N., Ruffle, S., and Southan, C. (2004) Splice variants: A homology modeling approach, *Proteins 54*, 596–608.

177 Romero, P. R., Zaidi, S., Fang, Y. Y., Uversky, V. N., Radivojac, P., Oldfield, C. J., Cortese, M. S., Sickmeier, M., LeGall, T., Obradovic, Z., and Dunker, A. K. (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, *Proc Natl Acad Sci U S A 103*, 8390–8395.

178 Dunker, A. K. and Uversky, V. N. (2008) Signal transduction via unstructured protein conduits, *Nat Chem Biol 4*, 229–230.

179 Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins, *Curr Opin Struct Biol 18*, 756–764.

180 Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics 21*, 3433–3434.

181 Uversky, V. N. and Dunker, A. K. (2008) Biochemistry. Controlled chaos, *Science 322*, 1340–1341.

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

183 Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, *J Mol Biol 337*, 635–645.

184 Grimmler, M., Wang, Y., Mund, T., Cilensek, Z., Keidel, E. M., Waddell, M. B., Jakel, H., Kullmann, M., Kriwacki, R. W., and Hengst, L. (2007) Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases, *Cell 128*, 269–280.

185 Tompa, P. (2005) The interplay between structure and function in intrinsically unstructured proteins, *FEBS Lett 579*, 3346–3354.

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

187 Fernandez, A. and Scheraga, H. A. (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions, *Proc Natl Acad Sci U S A 100*, 113–118.

188 Fernandez, A. (2004) Keeping dry and crossing membranes, *Nat Biotechnol 22*, 1081–1084.

189 Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways, *Curr Opin Struct Biol 8*, 101–106.

190 Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem Sci 24*, 329–332.

191 Bellotti, V., Mangione, P., and Stoppini, M. (1999) Biological activity and pathological implications of misfolded proteins, *Cell Mol Life Sci 55*, 977–991.

192 Uversky, V. N., Talapatra, A., Gillespie, J. R., and Fink, A. L. (1999) Protein deposits as the molecular basis of amyloidosis. I. Systemic amyloidoses, *Med Sci Monit 5*, 1001–1012.

193 Uversky, V. N., Talapatra, A., Gillespie, J. R., and Fink, A. L. (1999) Protein deposits as the molecular basis of amyloidosis. II. Localized amyloidosis and neurodegenerative disordres, *Med Sci Monit 5*, 1238–1254.

194 Rochet, J. C. and Lansbury, P. T., Jr (2000) Amyloid fibrillogenesis: Themes and variations, *Curr Opin Struct Biol 10*, 60–68.

195 Uversky, V. N. and Fink, A. L. (2004) Conformational constraints for amyloid fibrillation: The importance of being unfolded, *Biochim Biophys Acta 1698*, 131–153.

196 Uversky, V. N. and Fink, A. L. (2005) Pathways to amyloid fibril formation: Partially folded intermediates in fibrillation of unfolded proteins. In: Sipe, J. D. (ed.), *Amyloid proteins: The beta pleated sheet conformation and disease*. Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 247–265.

197 Glenner, G. G. and Wong, C. W. (1984) Alzheimer's disease and Down's syndrome: Sharing of a unique cerebrovascular amyloid fibril protein, *Biochem Biophys Res Commun 122*, 1131–1135.

198 Masters, C. L., Multhaup, G., Simms, G., Pottgiesser, J., Martins, R. N., and Beyreuther, K. (1985) Neuronal origin of a cerebral amyloid: Neurofibrillary tangles of Alzheimer's disease contain the same protein as the amyloid of plaque cores and blood vessels, *Embo J 4*, 2757–2763.

199 Lee, V. M., Balin, B. J., Otvos, L., Jr, and Trojanowski, J. Q. (1991) A68: A major subunit of paired helical filaments and derivatized forms of normal Tau, *Science 251*, 675–678.

200 Ueda, K., Fukushima, H., Masliah, E., Xia, Y., Iwai, A., Yoshimoto, M., Otero, D. A., Kondo, J., Ihara, Y., and Saitoh, T. (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease, *Proc Natl Acad Sci U S A 90*, 11282–11286.

201 Wisniewski, K. E., Dalton, A. J., McLachlan, C., Wen, G. Y., and Wisniewski, H. M. (1985) Alzheimer's disease in Down's syndrome: Clinicopathologic studies, *Neurology 35*, 957–961.

202 Dev, K. K., Hofele, K., Barbieri, S., Buchman, V. L., and van der Putten, H. (2003) Part II: Alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease, *Neuropharmacology 45*, 14–44.

203 Prusiner, S. B. (2001) Shattuck lecture—Neurodegenerative diseases and prions, *N Engl J Med 344*, 1516–1526.

204 Zoghbi, H. Y. and Orr, H. T. (1999) Polyglutamine diseases: Protein cleavage and aggregation, *Curr Opin Neurobiol 9*, 566–570.

205 Okazawa, H. (2003) Polyglutamine diseases: A transcription disorder? *Cell Mol Life Sci 60*, 1427–1439.

206 Cummings, C. J. and Zoghbi, H. Y. (2000) Fourteen and counting: Unraveling trinucleotide repeat diseases, *Hum Mol Genet 9*, 909–916.

207 Gusella, J. F. and MacDonald, M. E. (2000) Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease, *Nat Rev Neurosci 1*, 109–115.

208 Orr, H. T. (2001) Beyond the Qs in the polyglutamine diseases, *Genes Dev 15*, 925–932.

209 Fischbeck, K. H. (2001) Polyglutamine expansion neurodegenerative disease, *Brain Res Bull 56*, 161–163.

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

211 Adkins, J. N. and Lumb, K. J. (2002) Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2, *Proteins 46*, 1–7.

212 Chang, B. S., Minn, A. J., Muchmore, S. W., Fesik, S. W., and Thompson, C. B. (1997) Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2, *Embo J 16*, 968–977.

213 Campbell, K. M., Terrell, A. R., Laybourn, P. J., and Lumb, K. J. (2000) Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos, *Biochemistry 39*, 2708–2713.

214 Sunde, M., McGrath, K. C., Young, L., Matthews, J. M., Chua, E. L., Mackay, J. P., and Death, A. K. (2004) TC-1 is a novel tumorigenic and natively disordered protein associated with thyroid cancer, *Cancer Res 64*, 2766–2773.

215 Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: Introducing the D2 concept, *Annu Rev Biophys 37*, 215–246.

216 Uversky, V. N., Oldfield, C. J., Midic, U., Xie, H., Xue, B., Vucetic, S., Iakoucheva, L. M., Obradovic, Z., and Dunker, A. K. (2009) Unfoldomics of human diseases: Linking protein intrinsic disorder with diseases, *BMC Genomics 10*, Suppl 1, S7.

217 Garner, E., Romero, P., Dunker, A. K., Brown, C., and Obradovic, Z. (1999) Predicting binding regions within disordered proteins, *Genome Inform Ser Workshop Genome Inform 10*, 41–50.

218 Cheng, Y., Oldfield, C. J., Meng, J., Romero, P., Uversky, V. N., and Dunker, A. K. (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments, *Biochemistry 46*, 13468–13477.

219 Chen, C. T., Wagner, H., and Still, W. C. (1998) Fluorescent, sequence-selective peptide detection by synthetic small molecules, *Science 279*, 851–853.

220 Chene, P. (2004) Inhibition of the p53-hdm2 interaction with low molecular weight compounds, *Cell Cycle 3*, 460–461.

221 Chene, P. (2004) Inhibition of the p53-MDM2 interaction: Targeting a protein–protein interface, *Mol Cancer Res 2*, 20–28.

222 Anderson, C. W. and Appella, E. (2004) Signaling to the p53 tumor suppressor through pathways activated by genotoxic and nongenotoxic stress. In: Bradshaw, R. A. and Dennis, E. A. (eds.), *Handbook of cell signaling*. Academic Press, New York, pp. 237–247.

223 Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. (1991) p53 mutations in human cancers, *Science 253*, 49–53.

224 Balint, E. E. and Vousden, K. H. (2001) Activation and activities of the p53 tumour suppressor protein, *Br J Cancer 85*, 1813–1823.

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

226 Bottger, A., Bottger, V., Sparks, A., Liu, W. L., Howard, S. F., and Lane, D. P. (1997) Design of a synthetic Mdm2-binding mini protein that activates the p53 response in vivo, *Curr Biol 7*, 860–869.

227 Wasylyk, C., Salvi, R., Argentini, M., Dureuil, C., Delumeau, I., Abecassis, J., Debussche, L., and Wasylyk, B. (1999) p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53, *Oncogene 18*, 1921–1934.

228 Chene, P., Fuchs, J., Bohn, J., Garcia-Echeverria, C., Furet, P., and Fabbro, D. (2000) A small synthetic peptide, which inhibits the p53-hdm2 interaction, stimulates the p53 pathway in tumour cell lines, *J Mol Biol 299*, 245–253.

229 Garcia-Echeverria, C., Chene, P., Blommers, M. J., and Furet, P. (2000) Discovery of potent antagonists of the interaction between human double minute 2 and tumor suppressor p53, *J Med Chem 43*, 3205–3208.

230 Klein, C. and Vassilev, L. T. (2004) Targeting the p53-MDM2 interaction to treat cancer, *Br J Cancer 91*, 1415–1419.

231 Vassilev, L. T. (2004) Small-molecule antagonists of p53-MDM2 binding: Research tools and potential therapeutics, *Cell Cycle 3*, 419–421.

232 Vassilev, L. T., Vu, B. T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., and Liu, E. A. (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2, *Science 303*, 844–848.

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

234 Arkin, M. R. and Wells, J. A. (2004) Small-molecule inhibitors of protein–protein interactions: Progressing towards the dream, *Nat Rev Drug Discov 3*, 301–317.

235 Arkin, M. (2005) Protein–protein interactions and cancer: Small molecules going in for the kill, *Curr Opin Chem Biol 9*, 317–324.

236 Cochran, A. G. (2000) Antagonists of protein–protein interactions, *Chem Biol 7*, R85–R94.

237 Hammoudeh, D. I., Follis, A. V., Prochownik, E. V., and Metallo, S. J. (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc, *J Am Chem Soc 131*, 7390–7401.