

# REVIEW

## A decade and a half of protein intrinsic disorder: Biology still waits for physics

Vladimir N. Uversky<sup>1,2\*</sup>

<sup>1</sup>Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612

<sup>2</sup>Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

Received 13 March 2013; Accepted 25 March 2013

DOI: 10.1002/pro.2261

Published online 2 April 2013 proteinscience.org

**The abundant existence of proteins and regions that possess specific functions without being uniquely folded into unique 3D structures has become accepted by a significant number of protein scientists. Sequences of these intrinsically disordered proteins (IDPs) and IDP regions (IDPRs) are characterized by a number of specific features, such as low overall hydrophobicity and high net charge which makes these proteins predictable. IDPs/IDPRs possess large hydrodynamic volumes, low contents of ordered secondary structure, and are characterized by high structural heterogeneity. They are very flexible, but some may undergo disorder to order transitions in the presence of natural ligands. The degree of these structural rearrangements varies over a very wide range. IDPs/IDPRs are tightly controlled under the normal conditions and have numerous specific functions that complement functions of ordered proteins and domains. When lacking proper control, they have multiple roles in pathogenesis of various human diseases. Gaining structural and functional information about these proteins is a challenge, since they do not typically “freeze” while their “pictures are taken.” However, despite or perhaps because of the experimental challenges, these fuzzy objects with fuzzy structures and fuzzy functions are among the most interesting targets for modern protein research. This review briefly summarizes some of the recent advances in this exciting field and considers some of the basic lessons learned from the analysis of physics, chemistry, and biology of IDPs.**

**Keywords:** intrinsically disordered protein; intrinsically disordered protein region; protein structure; protein function; protein evolution; protein–protein interaction; partially folded protein

### Introduction

A bit more than ten years ago, *Protein Science* published a review entitled “Natively unfolded proteins: a point where biology waits for physics” (*Protein Sci* 2002 11(4):739–756).<sup>1</sup> The major goal of that article

was to bring an intriguing protein family of natively unfolded proteins (which are recognized now to constitute a subset of a very broad class of intrinsically disordered proteins, IDPs) out of shadow, to emphasize their lack of ordered structure under physiological conditions (at least ordered structure that could be detected by traditional low resolution techniques), to systemize their major structural properties, and to highlight their biological significance. The introduction of such biologically active but essentially unstructured proteins was used to challenge the hitherto dominant structure-centric viewpoint

Additional Supporting Information may be found in the online version of this article.

\*Correspondence to: Vladimir N. Uversky, Department of Molecular Medicine, University of South Florida, Morsani College of Medicine, 12901 Bruce B. Downs Blvd, MDC07, Tampa, FL. E-mail: vuvversky@health.usf.edu

(structure–function paradigm), according to which a specific function of a protein is determined by its unique and rigid three-dimensional (3D) structure. The title of the review (“a point where biology waits for physics”) was inspired by the observations that many of such “structure-less” proteins analyzed by that time acted as “binders” that did undergo at least partial folding after interaction with their binding partners. These observations provoked an idea that these biologically important proteins with little or no ordered structure have to wait to become more folded (and functional) as a result of binding to their specific partners. In other words, for these proteins, “biology,” that is, the ability to have biological functions, seemed to wait for “physics” which is manifested in their ability to undergo binding-induced folding (at least partial), which is necessary to bring the functional state of these proteins to life.<sup>1</sup>

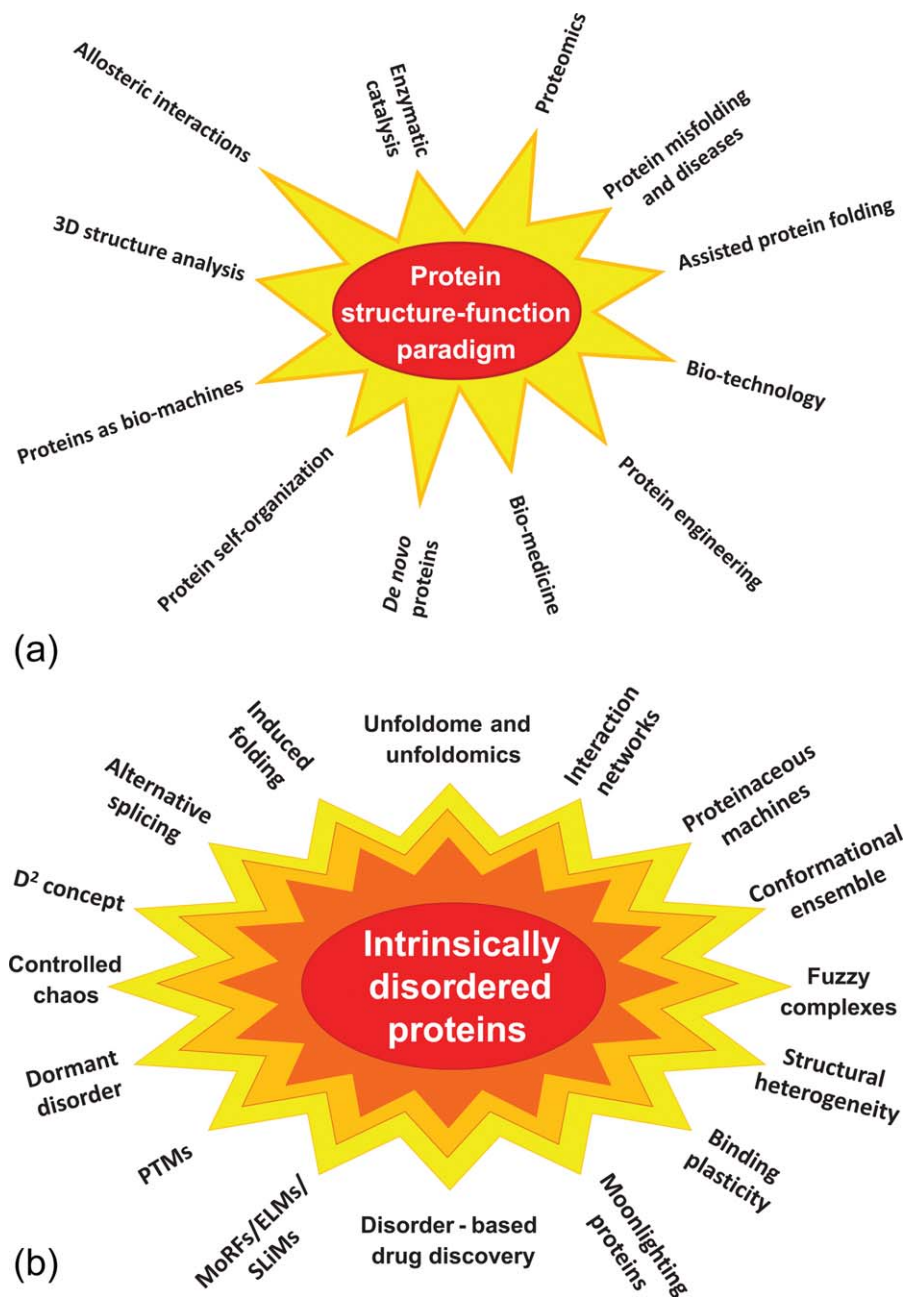
At the beginning, the idea that structure-less proteins can be biologically active was taken as a complete heresy by many researchers, since it was absolutely alien to then dominated structure–function paradigm which represented a foundation of the long-standing belief that the specific functionality of a given protein is determined by its unique 3-D structure. This structure–function paradigm that describes reasonably well the catalytic behavior of enzymes was based on the “lock-and-key” hypothesis formulated in 1894 by Emil Fischer.<sup>2</sup> This viewpoint was solidified by the successful solution of X-ray crystallographic structures of many proteins (as of February 26, 2013 there were 81,922 protein structures in the Protein Data Bank,<sup>3</sup> with 72,761 of these structures being determined by X-ray crystallography). These many crystal structures reinforced a static view of functional protein, where a rigid active site of an enzyme can be viewed as a sturdy lock that provides an exact fit to only one key, a specific and unique substrate.<sup>4</sup> Despite numerous limitations, this lock-and-key model was an extremely fruitful concept that was responsible for the creation of modern protein science.<sup>1</sup> Figure 1(A) shows some of the most obvious scientific consequences of the application of structure–function paradigm which is deservedly placed at the center of the “Big Bang” model that gives rise to the protein science universe.<sup>1</sup>

Obviously, the consideration of a protein as a rigid crystal-like entity is an oversimplification, since even the most stable and well-folded proteins are dynamic systems that possess different degrees of conformational flexibility. This is because of the simple fact that so-called conformational forces, that is, forces stabilizing the secondary structure of a protein and its tertiary fold, are weak and can be broken even at ambient temperatures due to thermal fluctuations.<sup>4</sup> The breaking of these weak interactions releases the groups that were involved in these interactions and gives them the possibility to be involved in the formation of new weak interactions of comparable energy.<sup>4</sup> Since these structural rearrangements

are of relatively small scale and since they occur typically in a time scale that is faster than the time required for structure determination by X-ray crystallography and many other physical techniques, the 3-D structures of proteins determined by these techniques represent averaged pictures.<sup>6</sup> Furthermore, one should keep in mind that not all proteins structures which are deposited to PDB are structured throughout their entire lengths. Instead, many PDB proteins have portions of their sequences missing from the determined structures (so-called regions of missing electron density)<sup>7,8</sup> due to the failure of the unobserved atom, side chain, residue, or region to scatter X-rays coherently caused by their flexible or disordered nature. Such flexible/disordered regions are rather common in the PDB, since only about 30% of protein structures deposited in the PDB do not have such regions of missing electron density.<sup>9</sup>

In addition to ordered proteins possessing disordered regions of varying length, the literature contains numerous examples of biologically active proteins with flexible structures.<sup>4</sup> Therefore, there is another class of functional proteins and protein regions that contain smaller or larger highly dynamic fragments, and some proteins are even characterized by a complete or almost complete lack of ordered structure under physiological conditions (at least *in vitro*) which appears to be a critical aspect of these proteins' function *in vivo*.<sup>4,10–15</sup> These proteins and protein regions (which are known now as IDPs and IDP regions (IDPRs)) have no single, well-defined equilibrium structure and exist as heterogeneous ensembles of conformers such that no single set of coordinates or backbone Ramachandran angles is sufficient to describe their conformational properties.

These proteins were independently discovered one-by-one over a long period of time and therefore they were considered as rare exceptions to the general rule. Although the phenomenon of biological functionality without stable structure was repeatedly observed, for a long time it was unnoticed by a wide audience because the authors frequently invented new terms to describe their protein of interest.<sup>16</sup> In fact, an incomplete list of terms coined in the literature to describe these proteins includes floppy, pliable, rheomorphic,<sup>17</sup> flexible,<sup>18</sup> mobile,<sup>19</sup> partially folded,<sup>20</sup> natively denatured,<sup>21</sup> natively unfolded,<sup>12,22</sup> natively disordered,<sup>15</sup> intrinsically unstructured,<sup>11,14</sup> intrinsically denatured,<sup>21</sup> intrinsically unfolded,<sup>22</sup> intrinsically disordered,<sup>13</sup> vulnerable,<sup>23</sup> chameleon,<sup>24</sup> malleable,<sup>25</sup> 4D,<sup>26</sup> protein clouds,<sup>27</sup> dancing proteins,<sup>28</sup> proteins waiting for partners,<sup>29</sup> and several other names often representing different combinations of “natively/naturally/inherently/intrinsically” with “unfolded/unstructured/disordered/denatured” among several others. Therefore, the majority of the names used in the early literature express that the “unfolded, unstructured, disordered, and denatured”



**Figure 1.** A: Protein structure–function paradigm is the “Big Bang” created universe of the modern protein science. Some major directions based on the consideration of protein function as lock-and-key mechanism are shown. Modified from Ref. 1. B: Paradigm shift caused by the introduction of the protein intrinsic disorder concept opened a wide array of new directions in protein science. In essence, introduction of this concept can be considered as a scientific revolution that, according to Kuhn,<sup>5</sup> “occurs when scientists encounter anomalies that cannot be explained by the universally accepted paradigm within which scientific progress has thereto been made” ([http://en.wikipedia.org/wiki/Paradigm\\_shift](http://en.wikipedia.org/wiki/Paradigm_shift)).

state is a “native, natural, inherent, and intrinsic” property of these proteins.<sup>16</sup>

Although protein intrinsic disorder is considered now as an established concept and PubMed contains hundreds and hundreds of papers talking about different aspects of IDPs/IDPRs, the route to recognizing these proteins as a novel functional entity was complex and lengthy. As it is often the case for new scientific concepts, the idea of structure-less functionality went through the stages of passive

ignorance and active denial to scrupulous examination and enthusiastic acceptance. For example, it took me more than a year to publish my first paper dedicated to the systematic analysis of such proteins, and the manuscript was successively rejected by 14 journals before it was finally accepted by *Proteins*.<sup>12</sup> However, time showed that the concept of protein intrinsic disorder was a useful invention and could be considered as a universal lock-pick that helps in solving many of the seemingly unsolvable

problems in protein science. One could say that this idea gave a new boost to the development of the protein science, generating a wide array of principally novel research directions [see Fig. 1(B)].

The goals of this review are: (i) to outline some recent advances in the field of IDPs/IDPRs; (ii) to illustrate the usefulness of intrinsic disorder for protein function; (iii) to show that intrinsic disorder can affect different levels of protein structural organization; (iv) to indicate intimate involvement of intrinsic disorder in pathogenesis of various maladies; (v) to emphasize the exceptional structural heterogeneity of IDPs/IDPRs and to show that IDPs are definitely much more structurally complex than random coil-like polypeptides; (vi) to accentuate that although this structural heterogeneity is very important for protein functionality, it represents a crucial hurdle for structural characterization of IDPs; (vii) to stress that new experimental and computational approaches and new theories and models are crucially needed for future progression of this field and protein science in general. These and other points highlight the current state of the field, where further advances in understanding of the “biology” of IDPs still waits for “physics,” with “physics” now being new theories, instrumentation, and analytical approaches.

### **Disorder Runs in their Blood: Disorderedness is Encoded in the Amino Acid Sequences of IDPs/IDPRs and can be Reliably Predicted**

Identification of IDPs as unique entities belonging to a new protein tribe is directly related to the recognition that their amino acid sequences are dramatically different from those of ordered proteins.<sup>10,12,13,30–32</sup> For example, it has been pointed out that the low content of hydrophobic residues combined with the high load of charged residues that often gives rise to high net charge of a polypeptide chain represents a characteristic feature of some IDPs (so called extended IDPs or natively unfolded proteins with coil-like or close to coil-like structures, see below).<sup>12</sup> Therefore, compact proteins and extended IDPs can be distinguished based only on their net charges and hydrophathies using a simple charge-hydrophathy (CH) plot, where the IDPs are specifically localized within a specific region of CH phase space and are reliably separated from compact ordered proteins.<sup>12</sup> More detailed comparison of amino acid sequences revealed that in comparison with ordered proteins and domains, the IDPs/IDPRs are significantly depleted in order-promoting amino acids (Trp, Tyr, Phe, Ile, Leu, Val, Cys, and Asn),<sup>10,33</sup> being instead enriched in disorder-promoting residues, such as Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro.<sup>13,31,32,34,35</sup>

Difference between ordered and disordered proteins goes far beyond these differences in their amino acid compositions. In fact, based on the

comparison of the 265 amino acid physico-chemical property-based scales (such as hydrophathy, net charge, flexibility index, helix propensities, strand propensities, aromaticity, etc.)<sup>34</sup> and more than 6000 composition-based attributes (e.g., all possible combinations having one to four amino acids in the group)<sup>36</sup> it has been concluded that ordered and disordered proteins and regions can be discriminated using many of these attributes.<sup>13</sup> Based on the analysis of 517 amino acid scales, a novel amino acid scale, Top-IDP (Trp, Phe, Tyr, Ile Met, Leu, Val, Asn, Cys, Thr, Ala, Gly, Arg, Asp, His, Gln, Lys, Ser, Glu, and Pro), was built to provide ranking for the tendencies of the amino acid residue to promote order or disorder.<sup>30</sup> The fact that the sequences of ordered and disordered proteins and regions are noticeably different suggested that IDPs clearly constitute a separate entity inside the protein kingdom, that these proteins can be reliably predicted using various computational tools,<sup>37–42</sup> and structurally, that IDPs should be very different from ordered globular proteins since peculiarities of amino acid sequence determine protein structure.

### **Natural Abundance of IDPs: Touching the Tip of the Iceberg**

Initial systematic analyses revealed that intrinsic disorder in proteins is a rather common phenomenon. In fact, as of 2002, the list of experimentally validated natively unfolded proteins with chain length greater than 50 amino acid residues contained more than 100 entries.<sup>1</sup> It was also pointed out that this list would probably be doubled if shorter polypeptides 30–50 residues long were included,<sup>1</sup> and that these 100 experimentally validated natively unfolded have at least 250 homologues, which are also expected to be natively unfolded.<sup>1,12</sup> It happened that these “large” numbers (which actually were large enough to make a crucial point that biologically active structure-less proteins represent the new rule and not mere rare exceptions) constitute just a small tip of an iceberg. In fact, using computational tools developed for sequence-based intrinsic disorder prediction the wide spread of IDPs and hybrid proteins containing IDPRs was convincingly shown.<sup>43–46</sup> For example, more than 15,000 out of 91,000 proteins in the then-current Swiss Protein database were identified as having long IDPRs.<sup>47</sup> The published in 2000 analysis of 31 whole genomes that span the 3 kingdoms of life revealed that many proteins contained segments predicted to have  $\geq 40$  consecutive disordered residues and that the eukaryotes exhibited more disorder by these measures than either the prokaryotes or the archaea.<sup>43</sup> Other studies on the abundance of intrinsic disorder in various evolutionary distant species supported these findings and consistently showed that the eukaryotic proteomes had higher

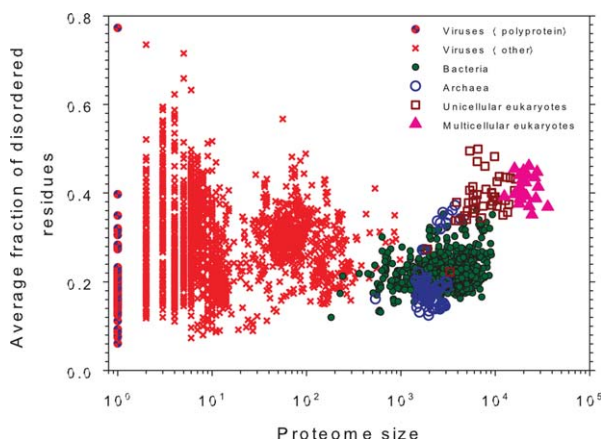


fraction of intrinsic disorder than prokaryotic proteomes.<sup>44,48–52</sup>

This conclusion is in line with the results of a comprehensive bioinformatics investigation of the disorder distribution in almost 3500 proteomes from viruses and three kingdoms of life, results of which are shown in Figure 2 as the correlation between the intrinsic disorder content and proteome size for 3484 species from viruses, archaea, bacteria, and eukaryotes.<sup>46</sup> Surprisingly, Figure 2 shows that there is a well-defined gap between the prokaryotes and eukaryotes in the plot of fraction of disordered residues on proteome size, where almost all eukaryotes have 32% or more disordered residues, whereas the majority of the prokaryotic species have 27% or fewer disordered residues.<sup>46</sup> Therefore, it looks like the fraction of 30% disordered residues serves as a boundary between the prokaryotes and eukaryotes and reflects the existence of a complex step-wise correlation between the increase in the organism complexity and the increase in the amount of intrinsic disorder. A gap in the plot of fraction of disordered residues on proteome size parallels a morphological gap between prokaryotic and eukaryotic cells which contain many complex innovations that seemingly arose all at once. In other words, this sharp jump in the disorder content in proteomes associated with the transition from prokaryotic to eukaryotic cells suggests that the increase in the morphological complexity of the cell paralleled the increased usage of intrinsic disorder.<sup>46</sup> The variability of disorder content in unicellular eukaryotes and rather weak correlation between disorder status and organism complexity (measured as the number of different cell types) is likely related to the wide variability of their habitats, with especially high levels of disorder being found in parasitic host-changing protozoa, the environment of which changes dramatically during their life-span.<sup>53</sup> The further support for this hypothesis came from the fact that the intrinsic disorder content in multicellular eukaryotes (which are characterized by more stable and less variable environment of individual cells) was noticeably less variable than that in the unicellular eukaryotes.<sup>46</sup>

### Complex Simplicity of IDPs

It was pointed out that IDPs possess noticeable amino acid biases, and many IDPs/IDPRs are characterized by sequence redundancy and low sequence complexity, containing long stretches of various repeats and being completely devoid of some (often many) types of amino acid residues. These observations seem to indicate that the sequence space of IDPs/IDPRs should be simpler than that of ordered proteins. However, the reality is more complex than conventional wisdom might suggest, and the sequence space attainable by simple IDPs/IDPRs is



**Figure 2.** Correlation between the intrinsic disorder content and proteome size for 3484 species from viruses, archaea, bacteria, and eukaryotes. Each symbol indicates a species. There are totally six groups of species: viruses expressing one polyprotein precursor (small red circles filled with blue), other viruses (small red circles), bacteria (small green circles), archaea (blue circles), unicellular eukaryotes (brown squares), and multicellular eukaryotes (pink triangles). Each viral polyprotein was analyzed as a single polypeptide chain, without parsing it into the individual proteins before predictions. The proteome size is the number of proteins in the proteome of that species and is shown in log base. The average fraction of disordered residues is calculated by averaging the fraction of disordered residues of each sequence over the all sequences of that species. Disorder prediction is evaluated by PONDR-VSL2B. Modified from Ref. 46.

more diversified than that of the structurally more sophisticated ordered proteins. In fact, a 100 residue-long protein in which any of the normally occurring 20 amino acids can be found has a sequence space of  $20^{100}$  ( $\sim 10^{130}$ ) sequences.<sup>54</sup> Obviously, not all random amino acid sequences can fold into unique structures. In other words, a sequence space of a foldable protein (or “foldable” sequence space) is noticeably smaller than the entire sequence space available for a random polypeptide chain. For decades, the actual size of “foldable” sequence space continues to be unsolved mystery despite a large body of theoretical, biochemical, and computational work that aims to unravel the relationship between a protein’s primary sequence and its resulting 3D structure.<sup>55</sup> However, the actual number of different amino acid residues in a given foldable sequence can be dramatically reduced,<sup>54</sup> since all twenty residues are not necessary for protein folding and the actual physicochemical identity of most of the amino acids in a protein is irrelevant.<sup>56–63</sup> In other words, folding alphabet can be noticeably reduced,<sup>55,64</sup> and amino acids can be clustered based on some shared features such as homolog substitution frequency,<sup>65</sup> local structural environments,<sup>66</sup> or peculiarities of the tertiary structural environments.<sup>67</sup> This simplified folding code further reduces the available “foldable” sequence space.<sup>68</sup>

Simply by virtue of their existence, IDPs/IDPRs add a new level of complexity to the sequence–structure relationship, dividing the population of protein sequences into two categories, sequences that yield natively ordered, and sequences that code natively disordered proteins.<sup>55</sup> IDPs/IDPRs cannot fold spontaneously and some of them require specific partners to gain more ordered structure. Therefore, they do not possess an entire folding code that defines the ability of foldable proteins to fold spontaneously into a unique biologically active structure. The missing portion of the IDP folding code (or at least part of it) is supplemented by binding partner(s). This defines a principal difference between structured proteins and IDPs/IDPRs: foldable proteins fold first and then bind to their partners whereas IDPs/IDPRs remain disordered until they interact with their partners.<sup>68,69</sup> Furthermore, many IDPs/IDPRs do not require folding to be functional,<sup>1,4,13,14,70–73</sup> and some of them form fuzzy complexes, in which they preserve significant amount of disorder.<sup>74,75</sup> All this suggests that the sequence space of IDPs (at least those which either do not fold at all or do not completely fold at binding) is noticeably greater than the “foldable” sequence space due to the removal of restrictions posed by the need to gain ordered structure spontaneously.<sup>68</sup> This represents one of the conundrums of intrinsic disorder, where the apparent sequence redundancy and simplicity are combined with the lack of structural restraints leading to the increase in the dimensions and complexity of the available sequence space.

Also, the existence of a noticeable sequence–structure heterogeneity of IDPs should be emphasized.<sup>68</sup> Since the unique 3D-structure of an ordered single-domain protein is defined by the interplay between all (or almost all) of its residues, one could expect that the structure-coding potential is homogeneously distributed within its amino acid sequence. On the other hand, a sequence of an IDP/IDPR contains multiple, relatively short functional elements and therefore represents a very complex structural and functional mosaic.<sup>68</sup> This important feature defines the known ability of an IDP/IDPR to interact, regulate, and be controlled by multiple structurally unrelated partners.<sup>76</sup> Such functional “anatomy” of IDPs/IDPRs is determined by the extremely high level of their sequence heterogeneity, which is further increased due to the ability of a single IDPR to bind to multiple partners gaining very different structures in the bound state.<sup>77</sup>

### Structural Heterogeneity of IDPs: Continuous Spectrum of Protein Structures

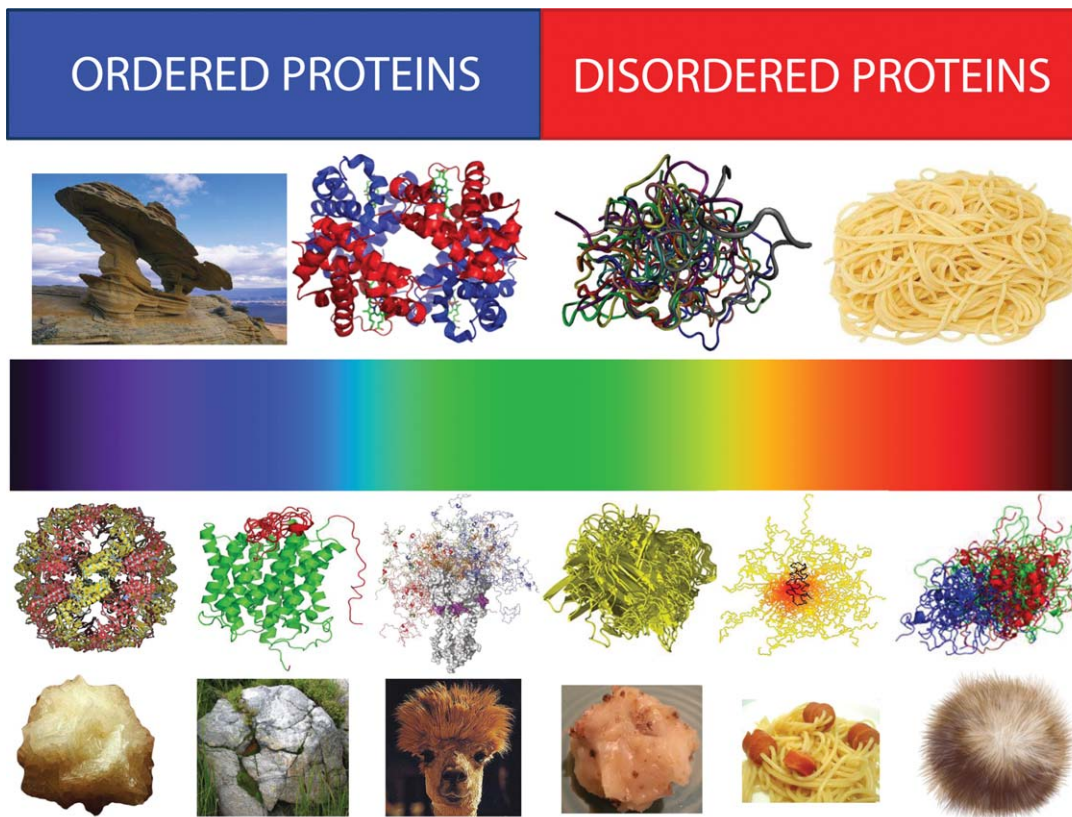
One of the crucial consequences of an extended sequence space and non-homogeneous distribution of foldability (or the structure-coding potential) within amino acid sequences of IDPs and IDPRs is their

astonishing structural heterogeneity. In fact, a typical IDP/IDPR contains a multitude of elements coding for potentially foldable, partially foldable, differently foldable, or not foldable at all protein segments.<sup>68</sup> As a result, different parts of a molecule are ordered (or disordered) to a different degree. This distribution is constantly changing in time where a given segment of a protein molecule has different structures at different time points. As a result, at any given moment, an IDP has a structure which is different from a structure viewed at another moment.<sup>68</sup>

Another level of structural heterogeneity is determined by the fact that many proteins are hybrids of ordered and disordered domains and regions, and this mosaic structural organization is crucial for their functions.<sup>16</sup> Also, even when they do not possess ordered domains, IDPs are known to have various levels and depth of disorder.<sup>78</sup> Over a few past years, an understanding of the available conformational space of IDPs/IDPRs underwent significant evolution. In fact, for a long time, IDPs were considered mostly “unstructured” or “natively unfolded” polypeptide chains. This was mostly due to the fact that the majority of IDPs analyzed at early stages of the field contained very little ordered structure, that is, they were really mostly unstructured or unfolded. Finding and characterization of such “structure-less” proteins was important to build up a strong case to counter-point the dominant view represented by the classical sequence-to-structure-to-function paradigm, especially since such fully unstructured, yet functional proteins clearly represented the other extreme of the protein structure–function spectrum.<sup>16</sup> The top half of the Figure 3 illustrates this situation by opposing rock-like ordered proteins and cooked spaghetti-like IDPs.

However, already in some early studies, it was indicated that IDPs/IDRs could be crudely grouped into two major structural classes, proteins with compact and extended disorder.<sup>1,4,12,13,73</sup> Based on these observations, the protein functionality was ascribed to at least three major protein conformational states, ordered, molten globular, and coil-like,<sup>13,79</sup> indicating that functional IDPs can be less or more compact and possess smaller or larger amount of flexible secondary/tertiary structure.<sup>1,4,12,13,73,79</sup> Roughly at the same time, it was emphasized that the extended IDPs (known as natively unfolded proteins) do not represent a uniform entity but contain two broad structural classes, native coils and native pre-molten globules.<sup>1</sup>

Currently available data suggest that intrinsic disorder possesses multiple flavors, can have multiple faces, and can affect different levels of protein structural organization, where whole proteins, or various protein regions can be disordered to a different degree.<sup>68</sup> This new view of structural space of



**Figure 3.** Structural heterogeneity of IDPs/IDPRs. Top half: Bi-colored view of functional proteins which are considered to be either ordered (folded, blue) or completely structure-less (disordered, red). Ordered proteins are taken as rigid rocks, whereas IDPs are considered as completely structure-less entities, kind of cooked noodles. Bottom half: A continuous emission spectrum representing the fact that functional proteins can extend from fully ordered to completely structure-less proteins, with everything in between. Intrinsic disorder can have multiple faces, can affect different levels of protein structural organization, and whole proteins, or various protein regions can be disordered to a different degree. Some illustrative examples includes ordered proteins that are completely devoid of disordered regions (rock-like type), ordered proteins with limited number of disordered regions (grass-on-the rock type), ordered proteins with significant amount of disordered regions (llama/camel hair type), molten globule-like collapsed IDPs (greasy ball type), pre-molten globule-like extended IDPs (spaghetti-and-sausage type), and unstructured extended IDPs (hairball type).

functional proteins can be visualized to form a continuous spectrum of differently disordered conformations extending from fully ordered to completely structure-less proteins, with everything in between (Fig. 3, bottom half). Here, functional proteins can be well-folded and be completely devoid of disordered regions (rock-like scenario). Other functional proteins may contain limited number of disordered regions (a grass-on-the rock scenario), or have significant amount of disordered regions (a llama/camel hair scenario), or be molten globule-like (a greasy ball scenario), or behave as pre-molten globules (a spaghetti-and-meatballs/sausage scenario), or be mostly unstructured (a hairball scenario).

Notably, in this representation, there is no boundary between ordered proteins and IDPs, and, the structure-disorder space of a protein is considered as a continuum. It is important to remember that even the most ordered proteins do not resemble “solid rocks” and have some degree of flexibility. In

fact, a protein molecule is an inherently flexible entity and the presence of this flexibility (even for the most ordered proteins) is crucial for its biological activity.<sup>80</sup> Also, another important point to remember is that due to their heteropolymeric nature, proteins are never random coils and always have some residual structure.<sup>68</sup>

### Polymeric Roots of the Unusual Biophysics of IDPs

Protein biophysicists/biochemists working on different aspects of ordered proteins (e.g., analyzing their structural properties, functions, folding, etc.) would find biophysical properties of functional IDPs/IDPRs to be rather unusual since these highly dynamic proteins do not follow the well-accepted wisdom that a protein has to be well-folded to be biologically functional. However, the unusualness is a subjective feature, and from the viewpoint of polymer physics the extended IDPs/IDPRs possess the expected behavior



of flexible and charged polymers, whereas the behavior of an ordered protein is rather unexpected (i.e., due to the existence of the native ensemble that for well-folded, ordered proteins can be approximated as a harmonic well around a unique, well-defined equilibrium structure). Therefore, one definitely should keep in mind that the “unusual” biophysics of extended IDPs/IDPRs has its roots in the usual polymer physics of highly charged and flexible polypeptides.

Each protein is believed to be a unique entity that has quite unique primary sequence which governs its 3D structure (or lack thereof) and ensures specific biological function(s). Therefore, understanding the effect of sequence variance on the biological performance presents a challenging task. However, natural polypeptides have originated as random copolymers of amino acids, which were adjusted or “selected” over evolution based on their functional capacities.<sup>56,81</sup> 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 as a major driving force responsible for the formation of equilibrium structures or structural ensembles.<sup>82</sup> For example, ordered globular proteins and molten globules (both as folding intermediates of globular proteins or 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 bad solvents for protein chains, resulting in the preservation of extensive residual structure even under these harsh denaturing conditions.<sup>82</sup>

Based on these and related observations, and taking into account the fact that many IDPs/IDPRs are characterized by significant amino acid composition biases, the overall polymeric behavior of these proteins and regions can be mimicked reasonably well by the behavior of low-complexity polypeptides (e.g., homopolypeptide and block copolypeptides). Following these ideas, it was shown that water is a poor solvent for polypeptide backbone alone and for the IDPs containing long tracts of polar amino acid residues since polar homo-polypeptides without hydrophobic groups (e.g., polyglutamine or glycine-serine block copolypeptides) were shown to prefer collapsed ensembles in aqueous media.<sup>83–88</sup> Furthermore, even polyglycine was shown to have a tendency to form heterogeneous ensembles of collapsed structures in water.<sup>88</sup>

A systematic analysis of the conformational behavior of protamines, arginine-rich IDPs involved in the condensation of chromatin during spermatogenesis, and protamine-like peptides 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.<sup>89</sup> Overall, the increase in the hydrodynamic dimensions of a polypeptide chain with increase in its 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.<sup>89</sup> Based on these premises, at least three different classes of globule-forming polar/charged IDPs were proposed. The first class is comprised by polar tracts which collapse due to water being a poor solvent for a backbone and non-charged side chains. The second class is represented by weak polyelectrolytes and weak polyampholytes, which have low per residue net charge and low fractions of positively and/or negatively charged residues. These IDPs/IDPRs 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/IDPRs possess polyampholytic nature, their globular state could be additionally stabilized by electrostatic interactions between the oppositely charged side-chains. Finally, IDPs/IDPRs from the third class are strong polyampholytes characterized by high fractions of positively and/or negatively charged residues but have low per residue net charge. Such intrinsically disordered protein can form collapsed structures stabilized mostly by multiple electrostatic interactions between solvated side-chains of opposite sign.<sup>89</sup>

The extended IDPs/IDPRs were used as a model system for the analysis of the effect of electrostatic interactions on conformational properties of unfolded proteins, and for testing the quantitative descriptions and predictions of polymer theory related to the influence of charged amino acids on chain dimensions.<sup>90</sup> For example, based on the analysis of the conformational equilibrium of coarse-grained polypeptides as a function of sequence hydrophobicity, charge, and length it has been concluded that the variations in sequence hydrophobicity and charge define a coil-to-globule transition comparable to that seen in the empirical CH-plot,<sup>12,91</sup> suggesting that a minimal, polymer physics-based model can capture the elements of global protein conformation.<sup>92</sup>

IDPs/IDPRs 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



analysis of the GdmCl-induced expansion of the unfolded states suggested that protein charge density plays a crucial role in defining the hydrodynamic behavior of the unfolded polypeptide chain.<sup>90</sup> Here, highly charged proteins can exhibit a prominent expansion at low ionic strength that correlates with their net charges.<sup>90</sup> It has been also hypothesized that the pronounced effect of charges on the dimensions of unfolded proteins might have important implications for their cellular functions.<sup>90</sup>

Similarly, a comprehensive analysis of the hydrodynamic dimensions of FG-nucleoporins containing large IDPRs with multiple phenylalanine-glycine repeats (FG-domains) revealed that under the physiologic conditions *in vitro* these domains adopt distinct categories of disordered structures, such as molten globule, pre-molten globule, relaxed-coil, extended-coil (as in urea), or very extended-coil (as in GdmCl).<sup>93</sup> 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.<sup>94</sup> Furthermore, FG-nucleoporins with higher charge density were shown to be more dynamic than the collapsed-coil FG-domains, being also prone to repel other FG-domains. On the other hand, the collapsed-coil FG-domains were prone to oligomerize. These observations suggested that different types of FG-domains with different aggregation propensities provide 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.<sup>94</sup> Therefore, the abundance and peculiarities of the charged residues distribution within the protein sequences might determine physical and biological properties of extended IDPs and IDPRs.

Also, simple polymer physics-based reasoning can give reasonably well-justified explanation of the conformational behavior of extended IDPs. 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 for 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.<sup>95</sup> The analysis of the temperature effects on structural properties of several extended IDPs revealed that native coils and native pre-molten globules partially fold as the temperature is increased.<sup>1,73,95–98</sup> These heating-induced structural changes in extended IDPs were 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. Similarly, extended IDPs/IDPRs are

characterized by the “turned out” response to changes in pH,<sup>96,99–102</sup> where a decrease (or increase) in pH induces their partial folding due to the minimization of their high net charges viewed at neutral pH, thereby decreasing charge/charge intramolecular repulsion and permitting hydrophobic-driven collapse to the partially folded conformation.<sup>95</sup>

### **Every Disordered Protein is Disordered in its Own Way**

Data accumulated so far indicate that intrinsic disorder exists at multiple structural levels and might differently affect different regions/domains of IDPs. This defines noted structural complexity and heterogeneity of IDPs/IDPRs which are further enhanced by the way different proteins/protein regions respond to their environments. Furthermore, since intrinsic disorder is crucial for many biological functions and therefore must prevail in different environments, the amino acid sequences and compositions of IDPs and IDPRs are specifically shaped by the peculiarities of their global and local environments. All this makes the protein intrinsic disorder phenomenon to be so broad that one can even assume that every disordered protein (or at least every family of disordered proteins) is disordered in its own way. This hypothesis has far-reaching consequences since it implies that a general disorder predictor has limited accuracy and cannot predict with equally high accuracy disorder status of all protein sequences due to their heterogeneity. It also implies that some environmental factors definitely should be taken into account when assessing intrinsic disorder in proteins. Several examples are presented below to support the overall validity of these statements.

### **Transmembrane proteins**

The first example is given by transmembrane (TM) proteins, in which disorder is widely observed (e.g., 40% of human integral plasma proteins were predicted to contain long IDPRs).<sup>103–107</sup> Furthermore, disorder is unevenly distributed between the cytoplasmic and the external surfaces of these proteins, with cytoplasmic domains being up to threefold more disordered than extracellular domains.<sup>105</sup> Although these analyses gave interesting hints on the abundance of disorder in TM proteins, the obvious weakness of such evaluations is in the fact that they were performed using the disorder predictors developed from structured and disordered regions found in water-soluble proteins.<sup>108</sup> However, the major physico-chemical properties of water-soluble and integral membrane proteins are very different due to the differences in their environments. For example, similar to typical water soluble proteins, the TM regions of membrane proteins are often

highly structured, containing  $\alpha$ -helices<sup>109</sup> or  $\beta$ -structure,<sup>110</sup> which are especially likely to occur due to the low dielectric constant values within the membrane bilayers.<sup>111,112</sup> On the other hand, the exterior regions of TM proteins are much more apolar than the exteriors of water-soluble proteins.<sup>113–115</sup> Therefore, the peculiarities of the membrane environment, with its highly nonpolar nature originating either from lipids or from protein interiors, are especially unfavorable for intrinsic disorder, since propensity for intrinsic disorder is typically encoded in a high content of polar and charged residues. Therefore, the IDPRs found in integral membrane proteins would be expected to be generally localized within the regions external to the membrane bilayer.<sup>108</sup> Also, the distinctive environment of the membrane bilayer imposes constraints on the amino acid composition of integral membrane proteins, even on the regions external to the membrane bilayer.<sup>116,117</sup> Comprehensive bioinformatics analysis revealed that integral membrane proteins commonly possess IDPRs defined as regions of missing electron density in their crystal structures.<sup>108</sup> Comparison of the IDPRs found in the  $\alpha$ -helical and the  $\beta$ -barrel bundle integral membrane proteins with the IDPRs viewed in typical water-soluble proteins revealed the existence of statistically distinct amino acid compositional biases characteristic for these three protein classes. Therefore, the use of specific amino acid signatures of IDPRs found in TM helical bundles and  $\beta$ -barrels can potentially lead to significantly more accurate disorder predictions for these two classes of integral membrane proteins.<sup>108</sup>

### **Proteins from archaea**

Another illustrative example of the specific disorder-related and environment-dependent sequence features is given by archaeal proteins.<sup>46,51</sup> Based on the levels of predicted disordered residues, archaeal proteins can be grouped into three classes, with ranges of the disordered residue content of 12–21%, 21%–32%, and 32%–38% (see Fig. 2). The archaeal proteomes with the highest disorder contents are halophiles and methanophiles.<sup>46,51</sup> Similar to TM proteins, the estimation of intrinsic disorder in the extremophilic proteins of the microorganisms surviving under hypersaline conditions using predictors developed for the “normal” non-halophilic proteins existing under the normal physiological conditions of 100–150 mM NaCl may not be accurate.<sup>46</sup> In fact, one of the strategies used by the halophilic archaea, which are salt-loving extremophilic organisms that grow optimally at high salt concentrations, to maintain proper osmotic pressure in their cytoplasm is a so-called “salt-in” strategy that involves accumulation of molar concentrations of potassium and chloride in their cytosoles.<sup>118</sup> This strategy requires extensive adaptation of the intracellular proteins to

the presence of near-saturating salt concentrations. The proteomes of such “salt-in” organisms are highly acidic,<sup>46,51</sup> and their proteins are characterized by remarkable instability at conditions of low salt concentrations and by maintaining soluble and active conformations in hypersaline conditions that are generally detrimental to the non-halophilic proteins.<sup>118–127</sup>

### **Viral proteins**

Finally, peculiarities of disorder distributions in viral proteins can be used to further support the importance of considering environmental factors.<sup>46,51</sup> Here, the comprehensive analysis of intrinsic disorder in various completed proteomes revealed that the viral proteomes have the largest variation of disorder content, which ranges from 7.3% disordered residues in the human coronavirus NL63 to 77.3% disordered residues in the *Avian carcinoma virus* proteome (see Fig. 2).<sup>46</sup> The high predicted intrinsic disorder content in viruses has multiple functional implications, where some IDPRs are used in the functioning of viral proteins and help viruses to hijack various pathways of the host cells, others likely have evolved to help viruses accommodate to their hostile habitats, and still others evolved to help viruses in managing their economic usage of genetic material via alternative splicing, overlapping genes, and anti-sense transcription.<sup>128</sup> These findings are in agreement with another study revealing that in comparison with archaea and bacteria, viral and bacteriophage proteins were significantly more enriched in polar residues and depleted in hydrophobic residues and were close to eukaryotic proteins in terms of their amino acid compositions and the reduced content of the order-promoting residues.<sup>129</sup>

### **Functional Manifoldness of IDPs**

#### **Functional protein clouds: Major functional advantages of being intrinsically disordered**

The high natural abundance of IDPs/IDPRs and their specific structural features indicate that these proteins and regions might carry out important biological functions. This hypothesis has been confirmed by several comprehensive studies,<sup>1,11–14,71–73,78,130–134</sup> which revealed that these structure-less members of the protein kingdom are abundantly involved in numerous biological processes, where they are frequently found to play different roles in regulation of the functions of their binding partners and in promotion of the assembly of supra-molecular complexes.<sup>1,4,11–15,31,70–73,76–79,131,132,134–149</sup> The conformational plasticity of IDPs/IDPRs provides them with a wide spectrum of exceptional functional advantages over the functional modes of ordered proteins and

domains.<sup>4,10,11,13,32,71,72,77,78,131,132,134,141,142,150,151</sup>

Some of these advantages are:

- 1 Increased speed of interaction due to greater capture radius and the ability to spatially search through interaction space;
- 2 Increased interaction (surface) area per residue;
- 3 Strengthened encounter complex allows for less stringent spatial orientation requirements;
- 4 Efficient regulation via rapid degradation;
- 5 The ability to be involved in one-to-many binding, where a single disordered region binds to several structurally diverse partners;
- 6 The ability to be involved in many-to-one binding, where many distinct (structured) proteins may bind a single disordered region;
- 7 The ability to overcome steric restrictions, enabling larger interaction surfaces in protein–protein and protein–ligand complexes than those obtained with rigid partners;
- 8 The ability to fold upon binding (completely or partially);
- 9 The ability of some IDPs/IDPRs to form very stable intertwined complexes;
- 10 The ability of some IDPs/IDPRs to stay substantially disordered in bound state;
- 11 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;
- 12 Binding plasticity, where an IDPR folds to specific bound conformations (which can be very different) according to the template provided by binding partners;
- 13 High accessibility of sites targeted for post-translational modifications (PTMs);
- 14 Efficient structural and functional regulation via PTMs such as phosphorylation, acetylation, lipidation, ubiquitination, sumoylation, and so forth, allowing for a simple means of modulation of their biological functions;
- 15 Efficient functional control via regulatory proteolytic attack sites of which are frequently associated with IDPRs;
- 16 Ease of regulation/redirection and production of otherwise diverse forms by alternative splicing (given the existence of multiple functions in a single disordered protein, and given that each functional element is typically relatively short, alternative splicing could readily generate a set of protein isoforms with a highly diverse set of regulatory elements<sup>152</sup>);
- 17 The possibility of overlapping binding sites due to extended linear conformation;
- 18 Decoupled binding affinity and specificity, where, due to the induced folding, IDP/IDPR can be involved in the formation of specific but weak complexes. In other words, IDP/IDPR might possess high specificity for given partners combined with high  $k_{\text{on}}$  and  $k_{\text{off}}$  rates that enable rapid association with the partner without an excessive binding strength. This combination of high specificity with low affinity defines the broad utilization of intrinsic disorder in regulatory interactions where turning a signal off is as important as turning it on;
- 19 Diverse evolutionary rates with some ID proteins being highly conserved and other ID proteins possessing high evolutionary rates. The latter ones can evolve into sophisticated and complex interaction centers (scaffolds) that can be easily tailored to the needs of divergent organisms;
- 20 Flexibility that allows masking (or not) of interaction sites or that allows interaction between bound partners;
- 21 The ability to be involved in the cascade interactions, where IDP binding to the first partner induces partial folding generating a new binding site suitable for interaction with the second partner, and so forth.

Functions of IDPs are complementary to the catalytic activities of ordered proteins.<sup>1,11–13,31,72,73,78,130,132,134,138–140,142,143</sup> Many disorder-related functions (e.g., signaling, control, regulation, and recognition) are incompatible with well-defined, stable 3-D structures.<sup>1,11–14,31,73,78,79,132,134,138–140,142,144,153</sup> Functions of many IDPs/IDPRs rely on interactions with specific binding partners, and many IDPs/IDPRs tend to undergo disorder-to-order transitions as a result of binding to their specific targets.<sup>12</sup>

Functionally, IDPs/IDPRs were grouped in at least six broad classes based on the mode of action.<sup>14,136</sup> These broad classes included protein and RNA chaperones, entropic chains, effectors, scavengers, assemblers and display sites,<sup>14,136</sup> and 28 separate functions, including molecular recognition via binding to other proteins, or to nucleic acids, were assigned for IDPRs in early studies.<sup>71,72</sup> Later, a rich spectrum of biological functions associated with IDPs/IDPRs was found based on a comprehensive computational study of a correlation between the functional annotations in the Swiss-Prot database and predicted intrinsic disorder.<sup>138–140</sup> The approach was 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 compared to the protein randomly chosen from the Swiss-Prot. This analysis revealed that ~44%

and ~34% of Swiss-Prot functional keywords were associated with ordered and disordered proteins, respectively, whereas 22% functional keywords yielded ambiguity in the likely function–structure associations.<sup>138–140</sup> Interestingly, most of the structured protein-associated key words were shown to be related to enzymatic activities, whereas the majority of the disordered protein-associated keywords were related to signaling and regulation. These results agree well with the notion that enzymatic catalysis requires ordered structure and that effectiveness of signaling is dependent on binding reversibility, a property directly associated with the thermodynamics of disorder-to-order transition induced by binding.<sup>138–140</sup>

Many IDPs and IDPRs undergo a disorder-to-order transition upon functioning.<sup>11,13,15,71,72,78,79,130–132,134,154–157</sup> 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.<sup>13,155</sup> High specificity coupled with low affinity is a useful pair of properties for a reversible signaling interaction. Furthermore, a disordered protein can readily bind to multiple partners by changing shape to associate with different targets.<sup>13,158,159</sup>

All this clearly suggests that there is a new two-pathway protein structure–function paradigm, with sequence-to-structure-to-function for enzymes and membrane transport proteins, and sequence-to-disordered ensemble-to-function for proteins and protein regions involved in signaling, regulation, and control.<sup>1,13,71,73,79</sup> One of the first generalization of this concept was given by The Protein Trinity Hypothesis, which suggested that native proteins can be in one of three states, the solid-like ordered state, the liquid-like collapsed-disordered state, or the gas-like extended-disordered state.<sup>79</sup> Function is then viewed to arise from any one of the three states or from transitions between them. This model was subsequently expanded to include a fourth state (pre-molten globule) and transitions between all four states.<sup>1</sup> In reality, based on the outlined above idea of the continuous spectrum of protein structures, functional proteins contain various amounts of intrinsic disorder and this continuous structural spectrum of protein defines their limitless functional variability.

#### **Lock-picking with Ariadne's thread: Illustrative examples of disorder-dependent protein functions**

Among intriguing protein functions relying on intrinsic disorder are moonlighting activities,<sup>137</sup> actions of hub proteins,<sup>78,93,134,160–164</sup> and scaffolding functions.<sup>141,165</sup> Since all these functions illustrate the notions that the intrinsic disorder concept

represents a universal skeleton key (or lock-pick) that helps unlocking seemingly unresolvable mysteries of protein science and therefore can be considered as a new Ariadne's thread that helps navigate the unusual twists of the sophisticated relationships between protein sequence, structure, and function, they are considered in some detail below.

**Moonlighting proteins.** Moonlighting is the ability of a protein to fulfill more than one function. Often, these functions are unrelated or at least are not obviously related to each other.<sup>137,166–168</sup> The capability of a protein to be involved in moonlighting or multi-tasking activities represent one of the solutions used by the Nature to increase the organism's complexity without the expansion of the genome size, where by acting differently at distinct points of metabolic networks proteins increase network complexity without increasing the actual size of the network.<sup>137,166–168</sup> Among various molecular mechanisms used by the moonlighting proteins to switch between functions are changes in cellular localization, changes in ligand binding, expression in different cell types, and variations of the oligomerization state.<sup>137</sup> In addition to these mechanisms that can be explained within the frames of the traditional structure–function paradigm, consideration of the intrinsic disorder phenomenon opens new possibilities.<sup>137</sup> In fact, one of the peculiar functional advantages of IDPs/IDPRs is their binding promiscuity and ability to be involved in one-to-many signaling, whereby an IDP/IDPR binds structurally different partners in a template-induced folding process.<sup>11,77,132,169</sup> Therefore, IDPs/IDPRs can use the same region or overlapping interaction regions/surfaces to exert distinct effects and employ the disorder-based mechanisms to switch function that relies on their capability to form different conformations upon binding.<sup>137</sup> Such structural malleability of IDPs/IDPRs defines their ability to participate in unprecedented moonlighting events, where these disordered moonlighting proteins or regions produce the opposing effects (inhibition and activation) on different partners or even the same partner molecule.<sup>137</sup>

**Hub proteins.** Signaling interactions inside the cell can be described as specific and complex networks that can be considered as “scale-free” or “small-world” networks, which have hubs, with many connections, and ends, that have the only connection to just one neighbor.<sup>170,171</sup> Such scale-free networks combine the local clustering of connections characteristic of regular networks with occasional long-range connections between clusters, as can be expected to occur in random networks. In other words, the distance between nodes in these scale-free networks follows a power-law distribution.<sup>172</sup> Based on their spatiotemporal peculiarities protein



hubs were grouped into two broad categories, “date hubs” that binds their numerous partners sequentially, and “party hubs” simultaneously interacting with their partners.<sup>173</sup> Since many IDPs are known to be involved in interaction with large number of distinct partners, they clearly can be considered as hubs in the scale-free protein–protein interaction networks.<sup>78,134</sup> Based on the systematic analysis of several know hub proteins<sup>134</sup> followed by a series of robust bioinformatics studies,<sup>93,160–164</sup> it was concluded that hubs commonly use disordered regions to bind to multiple partners and that there are at least two primary mechanisms by which disorder is utilized in protein–protein interaction networks where one disordered region binds to many partners or many disordered region bind to one partner.<sup>134</sup>

**Scaffold proteins.** Scaffold proteins constitute an important subclass of hubs that typically have a modest number of interacting partners and that are commonly found at the central parts of functional complexes, where they interact with most of their partners at the same time and therefore act as party hubs.<sup>160</sup> Besides being responsible for bringing together specific proteins within a signaling pathway and providing selective spatial orientation and temporal coordination to facilitate and promote interactions among interacting proteins, some scaffolds can influence the specificity and kinetics of signaling interactions via simultaneous binding to multiple participants in a particular pathway and facilitation and/or modifying the specificity of pathway interactions,<sup>174</sup> other scaffold can change conformations of individual proteins and thus modulate their activities,<sup>174</sup> still other scaffold proteins may modulate the activation of alternative pathways by promoting interactions between various signaling proteins.<sup>141</sup> Analysis of several well-characterized signaling scaffold proteins revealed that their large IDPRs are crucial for the successful scaffold function.<sup>141</sup> A more global bioinformatics analysis revealed that a typical design of a scaffold protein includes a set of short globular domains (~80 amino acids on average) connected by long linker regions (~150 residues on average) with crucial binding functions.<sup>165</sup> This gave further support to the notion that signaling scaffold proteins utilize the various features of highly flexible ID regions to obtain more functionality from less structure.<sup>141</sup>

**Disorder and transcription regulation.** Conformational plasticity and adaptability associated with intrinsic disorder are crucial for various protein functions. Among the proteins whose functional life is strongly disorder-dependent are transcription factors (TFs)<sup>175,176</sup> and other proteins involved in transcriptional regulation, such as the mediator complex,<sup>24,177</sup> core and linker histones,<sup>178</sup> and ribosomal proteins.<sup>179</sup> For example, from 83 to 94% of

TFs might possess long IDPRs, with the degree of disorder in eukaryotic TFs being significantly higher than in prokaryotic TFs.<sup>175,176</sup> Also, TFs were shown to be depleted in order-promoting residues and enriched in disorder-promoting residues, and were characterized by high levels of  $\alpha$ -molecular recognition feature (MoRF).<sup>175</sup> Furthermore, disorder is unevenly distributed within the TFs, with the degree of disorder in their activation regions being much higher than that in DNA-binding domains. However, the AT-hooks (which are DNA-binding motifs present in many proteins which binds to the (ATAA) and (TATT) repeats of DNA) and basic regions of TF DNA-binding domains are highly disordered suggesting that eukaryotes with their well-developed gene transcription machinery require transcription factor flexibility to be more efficient.<sup>175</sup> A number of interesting and important roles were also ascribed to intrinsic disorder in TFs related to the regulation of heat shock response (so called heat shock factors, HSFs)<sup>180</sup> and in the reprogramming TFs (the Yamanaka factors, namely Sox2, Oct3/4 (Pou5f1), Klf4, and c-Myc, and the Thomson factors, namely Sox2, Oct3, Lin28, and Nanog) overexpression of which is known to generate induced pluripotent stem (iPS) cells from terminally differentiated somatic cells.<sup>181</sup>

**Disorder in the regulation of cellular pathways.** Of special interests are the vital roles of intrinsic disorder in regulation and orchestration of various cellular pathways. One of the illustrative examples of this regulatory role of intrinsic disorder is the canonical Wnt-pathway that involves five proteins, Axin, CKI- $\alpha$ , GSK-3 $\beta$ , APC (adenomatous polyposis coli, also known as deleted in polyposis 2.5 protein), and  $\beta$ -catenin (all shown to contain long IDPRs). This pathway is known to play a number of crucial roles in the development of organism, and the malfunctions of which might lead to various diseases including cancer.<sup>182</sup> The comprehensive analysis of published data revealed that IDPRs found in Wnt-pathway proteins orchestrate protein–protein interactions, and facilitate PTMs and signaling.<sup>182</sup> Furthermore, the scaffold protein Axin and another large protein, APC, are heavily enriched in disorder and act as flexible concentrators in gathering together all other proteins involved in the Wnt-pathway.<sup>182</sup> Intriguingly, the multifarious roles of highly disordered APC in regulation of  $\beta$ -catenin function were established by showing that disordered APC helps the collection of  $\beta$ -catenin from cytoplasm, facilitates the  $\beta$ -catenin delivery to the binding sites on Axin, and controls the final detachment of  $\beta$ -catenin from Axin.<sup>182</sup>

Another important illustration of the involvement of intrinsic disorder in regulation of crucial pathway is given by the process of the programmed cell death (PCD), which is one of the most intricate cellular processes where the cell uses specialized

cellular machinery and intracellular programs to kill itself and which enables metazoans to control cell numbers and eliminate cells that threaten the animal's survival.<sup>183</sup> PCD includes several specific modules, such as apoptosis, autophagy, and programmed necrosis (necroptosis). These modules are not only tightly regulated but also intimately interconnected and are jointly controlled via a complex set of protein–protein interactions. Recently, several large sets of PCD-related proteins across 28 species were analyzed using a wide array of modern bioinformatics tools to understand the role of the intrinsic disorder in controlling and regulating the PCD.<sup>183</sup> This analysis revealed that proteins involved in regulation and execution of PCD possess substantial amount of intrinsic disorder and IDPRs were implemented in a number of crucial functions, such as protein–protein interactions, interactions with other partners including nucleic acids and other ligands, were shown to be enriched in post-translational modification sites, and were characterized by specific evolutionary patterns.<sup>183</sup>

### Molten globular Enzymes

Unique catalytic function of a protein is believed to be dictated by its unique 3D structure. This axiom constitutes a cornerstone of the lock-and-key paradigm and it seemed to be able to sustain the furious attack on protein structure–function relationship initiated by the discovery of IDPs and hybrid proteins with ordered domains and IDPRs. In fact, from the vast majority of experimental and computational studies a general conclusion was drawn over and over again, where the functional repertoire of IDPs complemented the functional arsenal of ordered proteins, with ordered proteins being mostly responsible for catalysis and transport and with IDPs doing the majority of other jobs in the cell.

On the other hand, all proteins (even the most ordered and tightly folded ones) are intrinsically flexible molecules that undergo conformational changes over a wide range of timescales and amplitudes.<sup>184</sup> In fact, the combination of active site reactivity with the dynamic character of proteins allows enzymes to be promiscuous and remarkably efficient at the same time.<sup>185</sup> Furthermore, in general, dynamic fluctuations are crucial for enzyme catalysis, since they can influence substrate binding and product release, and may even adjust the effective barriers of the catalyzed reactions.<sup>186–190</sup> Often, dynamic changes in the enzyme during the catalytic reaction can be described using the induced-fit model, where a conversion of one tight conformational ensemble (free enzyme) to another distinct ensemble (bound enzyme) takes place through a series of local substrate-mediated structural rearrangements.<sup>191</sup> Despite this crucial role of local flexibility in the enzymatic catalysis, enzymes are still

relatively stable molecules whose dynamic character is restricted to a small set of tightly folded conformations and whose unique (albeit locally flexible) structures are needed for efficient catalysis. From this viewpoint, the presence of intrinsic disorder is expected to be poorly compatible with enzymatic catalysis, which requires a well-organized environment in the active site of the enzyme in order to facilitate the formation of the transition state of the chemical reaction to be catalyzed.<sup>192</sup>

In a sharp contrast to this common wisdom supported by a wide array of specific examples, several enzymes were shown to be much more dynamic than the catalytic machines are expected to be, clearly possessing, in their precatalytic states, many characteristic properties of molten globules and retaining unusually high flexibility in structurally defined enzyme–ligand complexes. One of the best characterized examples of such molten globular enzymes is the engineered monomeric form of chorismate mutase from *Methanococcus jannaschii* (MjCM).<sup>184,193–195</sup> Here, a functional monomer (mMjCM) was created by inserting the hinge-loop sequence into the long, dimer-spanning N-terminal helix.<sup>193</sup> In its unbound form, mMjCM was shown to exist as a native molten globule that was described as a dynamic ensemble of  $\alpha$ -helical conformers rapidly interconverting on the millisecond timescale.<sup>193</sup> Interaction with natural ligand induced global conformational changes in the molten globular mMjCM promoting formation of a defined enzyme–ligand complex, which, however, preserved unusually high flexibility.<sup>184</sup> Catalytic mechanism of the molten globular mMjCM was described as follows: “Though probably stochastic in nature, internal motions in the complex may generate a collective dynamic matrix that samples catalytically active conformation(s) often enough to achieve rapid turnover in the presence of the true transition state.”<sup>184</sup> Therefore, some enzymes can represent a highly dynamic heterogeneous conformational ensemble which is still compatible with efficient catalysis.

In agreement with this hypothesis, a molten globular character was described for circularly permuted dihydrofolate reductase (DHFR),<sup>196,197</sup> and urease G from *Bacillus pasteurii* (BpUreG).<sup>198–200</sup> Of these three enzymatic molten globules UreG is the only natural molten globular enzyme known to date, since both circularly permuted DHFR and monomeric MjCM were obtained as a result of some genetic manipulations. Although the number of known native molten globules with enzymatic activity is small, their existence provides an interesting hint on early protein evolution. In fact, simple logics suggests that well-ordered enzymes appear as a result of long evolutionary process, whose very likely starting point was a partially folded polypeptide with some general properties of the molten globule.

### **Multifarious Interactivity of IDPs**

IDPs/IDPRs can form highly stable complexes, or be involved in signaling interactions where they undergo constant “bound–unbound” transitions, thus acting as dynamic and sensitive “on–off” switches. The ability of these proteins to return to the highly flexible conformations after the completion of a particular function, and their predisposition to gain different conformations depending on the environmental peculiarities, are unique physiological properties of IDPs which allow them to exert different functions in different cellular contests according to a specific conformational state.<sup>4</sup>

### **Static complexes**

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, as well as due to the unique capability to adjust to structure of the binding partner, IDPs are characterized by a very diverse range of binding modes, creating a multitude of unusual complexes, many of which are not attainable by ordered proteins.<sup>201</sup> Some of these complexes are relatively static, resemble complexes of ordered proteins, and, therefore are suitable for the structure determination by X-ray crystallography. Among these static complexes are: MoRFs, wrappers, chameleons, penetrators, huggers, intertwined strings, long cylindrical containers, connectors, armature, tweezers and forceps, grabbers, tentacles, pullers, and stackers or  $\beta$ -arcs.<sup>201</sup> These binding modes are shown in Supporting Information Figure 1S and briefly described in the Supporting Information Materials.

### **Disordered or fuzzy complexes of IDPs**

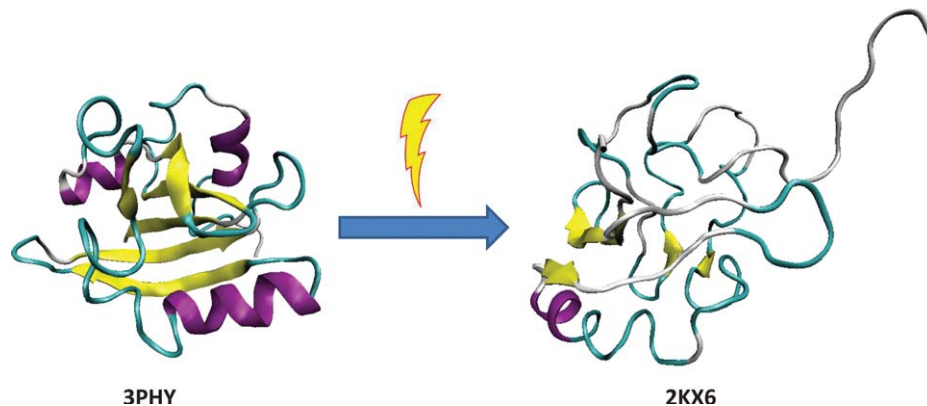
In addition to the static complexes, where bound partners have fixed structures, some IDPs/IDPRs do not fold even in their bound state, forming so-called disordered, dynamic, or fuzzy complexes with ordered proteins,<sup>97,202–206</sup> other disordered proteins,<sup>207–209</sup> or biological membranes.<sup>210,211</sup> 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.<sup>75,212</sup> It is also expected that the similar binding mode can be utilized by disordered protein while interacting with nucleic acids and other biological macromolecules.<sup>201</sup>

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, 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. This contradiction can be resolved assuming that the ordered binding partner and/or disordered protein contain 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 resemble a “binding cloud,” in which multiple identical binding sites are dynamically distributed in a diffuse manner. In other words, in this staccato-type interaction mode, a disordered protein rapidly changes multiple binding sites while probing binding site(s) of its partner.<sup>201</sup> An additional factor which can help holding a dynamic complex together could be a weak long-range attraction between protein molecules.<sup>213</sup> 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.<sup>213</sup>

### **Upside-Down Functionality: Functional Unfolding or Order–Disorder Transition**

The most common outcome of these function-related structural changes is the overall increase in the amount of ordered structure. However, functions of some ordered proteins require local or even global unfolding of a unique protein structure.<sup>68</sup> Among specific features of these structural alterations are their induced nature and transient character combined with a wide range of molecular mechanisms by which they can be promoted.<sup>68</sup> These functional unfolding-activating factors include light; mechanical force; changes in pH, temperature, or redox potential; interaction with membrane, ligands, nucleic acids, and proteins; various PTMs; release of autoinhibition due to the unfolding of autoinhibitory domains induced by their interaction with nucleic acids, proteins, membranes, PTMs, and so forth.<sup>68</sup> Among rather unusual factors used by nature to activate proteins via functional unfolding are light and mechanical force. For example, exposure to blue light results in the activation of the photoactive yellow protein (PYP), which is an ordered, water-soluble ~14 kDa protein that contains a thioester linked



**Figure 4.** Comparison of the ground state (left structure, PDB ID: 3PHY) and the transient light activated signaling state of the PYP (right structure, PDB ID: 2KX6). Ground state structure was determined by multidimensional NMR spectroscopy.<sup>218</sup> This structure is in agreement with an earlier published 1.4 Å crystal structure,<sup>219</sup> and modeled structure based on combined DEER, TR-SAXS/WAXS, and NMR data.<sup>217</sup> It consists of an open, twisted, 6-stranded, antiparallel  $\beta$ -sheet, which is flanked by four  $\alpha$ -helices on both sides.<sup>217-219</sup> On the contrary, the light-activated form is highly disordered. This structure satisfies DEER, SAXS/WAXS, and NMR data simultaneously.<sup>217</sup>

*p*-coumaric acid cofactor and serves as a photosensor in *Ectothiorhodospira halophila*.<sup>214,215</sup> PYP is a bacterial blue light sensor that undergoes conformational changes upon signal transduction. The absorption of a photon triggers substantial protein unfolding and leads to the formation of the transient signaling state that interacts with the partner molecules. This allows the swimming bacterium to operate the directional switch that protects it from harmful illumination. Comprehensive analysis combining double electron resonance spectroscopy (DEER), high resolution NMR, and time-resolved pump-probe X-ray solution scattering (TR-SAXS/WAXS) revealed that the transiently activated and short-lived signaling state of the PYP possessed a large degree of disorder and existed as an ensemble of multiple conformers that exchange on a millisecond time scale.<sup>216,217</sup> This unusual behavior is illustrated in Figure 4 that shows structures of inactive folded PYP and its light-activated functional form, which is highly disordered.<sup>68</sup> Some proteins undergo local unfolding induced by the mechanical force and therefore can serve as force sensors.<sup>68</sup> Among these natural force sensors are mechanosensitive ion channels that recognize and respond to the membrane tension, which is the mechanical forces applied along the plane of the cell membrane, rather than to the hydrostatic pressure perpendicular to the membrane plane.<sup>220</sup> These ion channels are activated via partial unfolding of some of their functional parts induced by membrane tension.<sup>221</sup>

For a long time, the fact that IDPs/IDPRs undergo disorder-to-order transitions either during their functions or in order to be functional was used as one of the strongest arguments against the idea of protein intrinsic disorder. It was stated that most IDPs (those which are not the artifacts of current

methods of protein production) are in fact proteins waiting for a partner (PWPs) that serve as parts of a multi-component complex and that do not fold correctly in the absence of other components.<sup>29</sup> Therefore, when folded after binding to their partners, these proteins are not too different from typical ordered proteins. However, one need to keep in mind that a portion of “folding code” that defines the ability of ordered proteins to spontaneously gain a unique biologically active structure is missing for IDPs/IDPRs since they cannot fold spontaneously. This missing portion of the “folding code” (or a part of it) can be supplemented by binding partner(s). As a result, ordered and disordered proteins can be discriminated on a simple basis of temporal correlation between their folding and binding: ordered proteins fold first and then bind to their partners while the IDPs/IDPRs remain disordered until they bind their partners and often preserve substantial disorder in the bound state.<sup>69</sup> Furthermore, numerous cases of functional unfolding (or transient disorder, or upside-down functionality) represent further support to the concept of functional disorder by clearly showing that many proteins possess dormant disorder that needs to be awakened in order to make these proteins functional.

### Specific Surveillance for Special Control or Controlled Chaos

It is clear now that the IDPs and IDPRs are real, abundant, diversified, and vital. The highly dynamic nature of IDPs and IDPRs is a visual illustration of the chaos. However, the evolutionary persistence of these highly dynamic proteins (see below), their unique functionality, and involvement in all the major cellular processes evidence that this chaos is tightly controlled.<sup>147</sup> To answer the question as to



how these proteins are governed and regulated inside the cell, Gsponer *et al.* conducted a detailed study focused on the intricate mechanisms of IDP regulation.<sup>222</sup> To this end, all the *Saccharomyces cerevisiae* proteins were grouped into three classes using one of the available disorder predictors, DisoPred2<sup>44</sup>: (i) 1971 highly ordered proteins containing 0–10% of the predicted disorder; (ii) 2711 moderately disordered proteins with 10–30% predicted disordered residues; and (iii) 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.

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 transcripts of genes encoding IDPs.<sup>222</sup> Furthermore, 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,<sup>223</sup> the experimentally determined yeast kinase–substrate network was also analyzed. 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.<sup>222</sup> Therefore, PTMs may not only serve as important mechanism for the fine-tuning of the IDP functions but possibly they are necessary to tune the IDP availability under the different cellular conditions.<sup>222</sup>

In addition to *S. cerevisiae*, similar regulation trends were also found in *Schizosaccharomyces pombe* and *Homo sapiens*.<sup>222</sup> Based on these observations it has been concluded that both unicellular and multicellular organisms appear to use similar mechanisms for regulation of the intrinsically disordered protein availability. Overall, this study clearly demonstrated that in eukaryotes, 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.<sup>222</sup> 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 or/and for long periods of time due to either specific PTMs or via interactions with other factors, which could promote changes in cellular localization of IDPs or protect them from the degradation machinery.<sup>13,70,138,223,224</sup> Overall, this study

clearly showed that the chaos seemingly introduced into the protein world by the discovery of IDPs is under the tight control.<sup>147</sup>

In an independent study, a global scale relationship between the predicted fraction of protein disorder and protein expression in *E. coli* was analyzed.<sup>225</sup> This study showed that the fraction of protein disorder was positively correlated with both measured RNA expression levels of *E. coli* genes in three different growth media and with predicted abundance levels of *E. coli* proteins.<sup>225</sup> When a subset of 216 *E. coli* proteins that are known to be essential for the survival and growth of this bacterium were 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.<sup>225</sup> The manual literature mining for a group of *E. coli* proteins that had high levels of predicted intrinsic disorder revealed that the disorder predictions matched well with the experimentally elucidated regions of protein flexibility and disorder.<sup>225</sup> A direct link between protein disorder and protein level in *E. coli* cells could be because the IDPs may carry out the essential control and regulation functions that are needed to respond to the various environmental conditions. Another possibility is that IDPs might undergo more rapid degradation compared to structured proteins, which cells can counter by increasing 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.<sup>225</sup>

Even more support for the tight control of IDPs inside the cell came from the analysis of cellular regulation of so-called “vulnerable” proteins.<sup>23</sup> The integrity of the 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 unfolding of the soluble structure under physiological conditions.<sup>226,227</sup> 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.<sup>23</sup> It has been also 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.<sup>226</sup> 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. This study revealed that structural vulnerability can be considered as a determinant of transcriptome organization across tissues and temporal phases.<sup>23</sup> 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.<sup>23</sup>

### IDPs in Diseases

Because of the fact that IDPs are highly abundant and 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 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.<sup>228–234</sup> Several unrelated proteins including many IDPs are known to be involved in these protein deposition diseases.<sup>234,235</sup> An illustrative examples of human neurodegenerative diseases associated with IDPs includes Alzheimer's disease (deposition of amyloid- $\beta$ , tau-protein,  $\alpha$ -synuclein fragment NAC)<sup>236–239</sup>; various tauopathies (accumulation of tau-protein in the form of neurofibrillary tangles)<sup>238</sup>; Down's syndrome (nonfilamentous amyloid- $\beta$  deposits)<sup>240</sup>; Parkinson's disease and other synucleinopathies (deposition of  $\alpha$ -synuclein)<sup>241</sup>; prion diseases (deposition of PrP<sup>SC</sup>)<sup>242</sup>; 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.<sup>243</sup> 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.<sup>228–234</sup>

The maladies given above have been called conformational diseases, as they are characterized by

the 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  $\alpha$ -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 Ref. 241). Overall, there are about 50 proteins and ligands that interact and/or co-localize with this protein. Furthermore,  $\alpha$ -synuclein has amazing structural plasticity and adopts a series of different monomeric, oligomeric, and insoluble conformations (reviewed in Ref. 24). The choice between these conformations is determined by the peculiarities of the protein environment, suggesting that  $\alpha$ -synuclein has an exceptional ability to fold in a template-dependent manner. Therefore, the development of the conformational diseases may originate not only from misfolding but also from the misidentification, misregulation, and missignaling of the related proteins.

Analysis of so-called polyglutamine diseases gives support to this hypothesis.<sup>244</sup> Polyglutamine diseases are a specific group of hereditary neurodegeneration caused by expansion of CAG triplet repeats in an exon of disease genes which 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, 17, and dentatorubral pallidoluysian atrophy are known to belong to this class of diseases.<sup>245–248</sup> In most polyQ diseases, expansion to over 40 repeats leads to the onset.<sup>248</sup> It has been emphasized that such molecular processes as unfolded protein response, protein transport, synaptic transmission, and transcription are implicated in the pathology of polyQ diseases.<sup>244</sup> 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 finally to the neuronal dysfunction and death (reviewed in Ref. 244). These results suggest that polyQ diseases represent kind of transcriptional disorder,<sup>244</sup> supporting our misidentification hypothesis for at least some of the conformational disorders.

Disorder is very common in cancer-associated proteins too. In a 2002 study, it was found that 79% of cancer-associated and 66% of cell-signaling proteins contain predicted regions of disorder of 30 residues or longer.<sup>130</sup> In contrast, only 13% of a set of proteins with well-defined ordered structures contained such long regions of predicted disorder.<sup>130</sup> In

experimental studies, the presence of disorder has been directly observed in several cancer-associated proteins, including p53,<sup>249</sup> p57<sup>kip2</sup>,<sup>250</sup> Bcl-X<sub>L</sub> and Bcl-2,<sup>251</sup> c-Fos,<sup>252</sup> a thyroid cancer associated protein TC-1,<sup>253</sup> EWS-FLI1 fusion protein that includes a potent transcriptional activator, the EWS domain, alongside the highly conserved DNA-binding domain FLI1,<sup>254,255</sup> among many other examples. The best characterized example of the important cancer-related IDP is the tumor suppressor protein p53, which occupies the center of a large signaling network. p53 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.<sup>256</sup> When p53 function is lost, either directly through mutation or indirectly through several other mechanisms, the cell often undergoes cancerous transformation.<sup>257,258</sup> Cancers showing mutations in p53 are found in colon, lung, esophagus, breast, liver, brain, reticuloendothelial tissues, and hemopoietic tissues.<sup>257</sup> p53 is regulated by several different mechanisms including inhibition of its activity by interaction with E3 ubiquitin ligase Mdm2, which binds to a short stretch of p53 located within the transactivation domain. Mdm2-bound p53 cannot activate or inhibit other genes. 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.<sup>259,260</sup>

### Druggability of IDPs

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. Since many proteins associated with various human diseases are either completely disordered or contain long disordered regions,<sup>261,262</sup> and since some of these disease-related IDPs/IDPRs are involved in recognition, regulation, and signaling, these proteins/regions clearly represent novel potential drug targets.<sup>27</sup> 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. In fact, the rational design of enzyme inhibitors depends on the classical view where 3D-structure is an obligatory prerequisite for function. While generally applicable to many enzymatic domains, this view has persisted to influence thinking concerning all protein functions despite numerous examples to the contrary. 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 “unique structure–unique function” paradigm is generally applicable.

IDPs often bind their partners with relatively short regions that become ordered upon binding.<sup>263–265</sup> Targeting disorder-based interactions should enable the development of more effective drug discovery techniques. There are at least two potential approaches for the inhibition of the disorder-based interactions, where small molecule either bind to the binding site of the ordered partner to outcompete the IDPs/IDPRs or interacts directly with the IDP/IDPR. The principles of small molecule binding to IDPRs have not been well studied, but sequence specific, small molecule binding to short peptides has been observed.<sup>266</sup> An interesting twist here is that small molecules can inhibit disorder-based protein–protein interactions via induction of the dysfunctional ordered structures in targeted IDPR, that is, via the drug-induced misfolding.

In agreement with these concepts, small molecules “Nutlins” have been discovered that inhibited the p53-Mdm2 interaction by mimicking the inducible  $\alpha$ -helix in p53 (residues 13–29) that binds to Mdm2.<sup>259,260</sup> Although X-ray crystallographic studies of the p53-Mdm2 complex revealed that the Mdm2 binding region of p53 forms an  $\alpha$ -helical structure that binds into a deep groove on the surface of Mdm2,<sup>267</sup> NMR studies showed that the unbound N-terminal region of p53 lacks fixed structure, although it does possess an amphipathic helix part of the time.<sup>249</sup> A close examination of the interface between the proteins reveals that Phe<sub>19</sub>, Trp<sub>23</sub>, and Leu<sub>26</sub> 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.<sup>259,260</sup> 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<sub>23</sub> and Leu<sub>26</sub>, and an ethyl-ether side chain filling the spot normally taken by Phe<sub>19</sub>.<sup>268–270</sup> Nutlins and related small 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 nutlins was given orally to mice, a 90% inhibition of tumor growth compared to the control was induced.<sup>260,268–270</sup>

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 transitions are generally understood and therefore can use to find similar drug targets, which are inducible  $\alpha$ -helices.<sup>271</sup> In addition to nutlins inhibiting p53-Mdm2 interaction, several other small molecules also act by blocking protein–protein interactions.<sup>272,273</sup> Some of these interactions involve one structured partner and one disordered partner, with disordered segments becoming  $\alpha$ -helix

upon binding.<sup>271</sup> Therefore, the p53–Mdm2 complex is not a unique exception and many other disorder-based protein–protein interactions are blocked by a small molecule. All this suggest that there is a cornucopia of new drug targets that would operate by blocking disorder-based protein–protein interactions.

For these p53–Mdm2-type 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. These druggable interaction sites operate by the coupled binding and folding mechanism. They are small enough and compact enough to be easily mimicked by small molecules.<sup>25</sup> Methods for predicting such binding sites in disordered regions have been developed<sup>274</sup> and the bioinformatics tools to identify which disordered binding regions can be easily mimicked by small molecules have been elaborated.<sup>271</sup>

A complementary approach for small molecules to inhibit the disorder-based protein–protein interactions relies on the direct binding of drugs to the IDPs/IDPRs, which is illustrated by the c-Myc–Max story.<sup>275</sup> In order to bind DNA, regulate expression of target genes, and function in most biological contexts, c-Myc transcription factor must dimerize with its obligate heterodimerization partner, Max, which lacks a transactivation segment. Both c-Myc and Max are intrinsically disordered in their monomeric forms. Upon heterodimerization, they undergo coupled binding and folding of their basic-helix-loop-helix-leucine zipper domains (bHLHZip). Since the deregulation of c-Myc is related to many types of cancer, the disruption of the c-Myc–Max dimeric complex is one of the approaches for c-Myc inhibition. Several small molecules were found to inhibit the c-Myc–Max dimer formation.<sup>275</sup> These molecules were shown to bind to one of the three discrete sites within the 85-residue bHLHZip domain of c-Myc, which are composed of short contiguous stretches of amino acids that can selectively and independently bind small molecules.<sup>275</sup> Inhibitor binding induces only local conformational changes, preserves the overall disorder of c-Myc, and inhibits interaction with Max.<sup>275</sup> 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.<sup>275</sup>

Recently, a functional misfolding concept was introduced to describe a mechanism preventing IDPs from unwanted interactions with non-native partners.<sup>276</sup> IDPs/IDPRs are characterized by high conformational dynamics and flexibility, the presence of sticky preformed binding elements, and the

ability to morph into differently-shaped bound configurations. However, detailed analyses of the conformational behavior and fine structure of several IDPs revealed that the preformed binding elements might be involved in a set of non-native intramolecular interactions. Based on these observations it was proposed that an intrinsically disordered polypeptide chain in its unbound state can be misfolded to sequester the preformed elements inside the non-interactive or less-interactive cage, therefore preventing these elements from the unnecessary and unwanted interactions with non-native binding partners.<sup>276</sup> It is important to remember, however, that the mentioned functional misfolding is related to the ensemble behavior of transiently populated elements of structure. In other words, it describes the behavior of a globally disordered polypeptide chain containing highly dynamic elements of residual structure, so-called interaction-prone preformed fragments, some of which could potentially be related to protein function.<sup>276</sup>

This ability of IDPRs/IDPRs to functionally misfold can be used for finding small molecules which would potentially stabilize different members of the functionally misfolded ensemble, and therefore prevent the targeted protein from establishing biological interactions.<sup>277</sup> This approach is very different from the discussed above direct targeting of short IDPRs since it is based on a small molecule binding to a highly dynamic surface created via the transient interaction of preformed interaction-prone fragments. In essence, this approach can be considered as an extension of the well-established structure-based rational drug design elaborated for ordered proteins. In fact, if the structure of a member(s) of the functionally misfolded ensemble can be guessed, then this structure can be used to find small molecules that are potentially able to interact with this structure, utilizing tools originally developed for the rational structure-based drug design for ordered proteins.<sup>277</sup>

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 alternative splicing 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.<sup>278</sup> The added protein diversity from alternative splicing is thought to be important for tissue-specific signaling and regulatory networks in the multicellular organisms. The regions of alternative splicing in proteins are



enriched in intrinsic disorder, and it was proposed that associating alternative splicing with protein disorder enables the time- and tissue-specific modulation of protein function.<sup>152</sup> Since disorder is frequently utilized in protein binding regions, having alternative splicing of pre-mRNA coupled to IDPRs can define tissue-specific signaling and regulatory diversity.<sup>152</sup> These findings open a unique opportunity to develop tissue-specific drugs modulating the function of a given ID protein/region (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.

## Evolution of Intrinsically Disordered Proteins

### *Wavy pattern of global evolution of intrinsic disorder*

IDPs/IDPRs are more common in eukaryotes than in less complex organisms.<sup>43,44,48–52</sup> This suggests that disorder, with its ability to be implemented in various signaling, recognition, and regulation pathways and networks, is important for the maintenance of life in eukaryotic and especially multicellular eukaryotic organisms.<sup>4,45,78,134</sup> Also, the finding that alternatively spliced regions of mRNA code for IDPRs much more often than for structured regions suggests that there is a linkage between alternative splicing and signaling by IDPRs that constitutes a plausible mechanism that could underlie and support cell differentiation, which ultimately gave rise to the multicellular eukaryotic organisms.<sup>152</sup>

Therefore, one can assume that intrinsic disorder represents a relatively recent evolutionary invention. However, this hypothesis obviously would be wrong if earlier stages of evolution would be taken into account. In fact, the chances that the first polypeptides that appeared in the primordial soup of the primitive Earth possessed well-developed and unique 3D structures are minimal. The Earth formed about 4.5 billion years ago. Scientists dated the first fossils to 3.85 billion years ago. There are still debates and different theories about what happened in those years between the time the earth was cool enough to spawn life and the time the first fossils were formed. At the beginning of the 20th century, Oparin<sup>279</sup> and Haldane<sup>280</sup> proposed that some organic molecules could have been spontaneously produced from the gases of the primitive Earth atmosphere, assuming that this primitive atmosphere was reducing (as opposed to oxygen-rich), and there was an appropriate supply of energy, such as lightning or ultraviolet light. Thirty year later, this hypothesis (that constitutes a cornerstone of the theory of molecular evolution) received strong support from the elegant experiments of Stanley L. Miller and Harold C. Urey who were able to

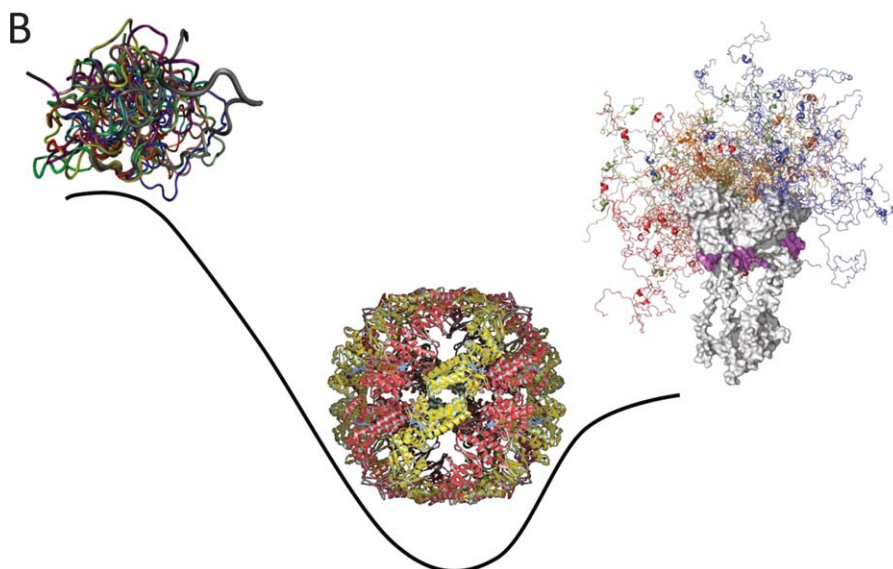
synthesize various organic compounds including some amino acids from non-organic compounds which were believed to represent the major components of the early Earth's atmosphere (water vapor, hydrogen, methane, and ammonia) by putting them into a closed system and running a continuous electric current through the system, to simulate lightning storms believed to be common on the early Earth.<sup>281,282</sup> However, the Miller–Urey experiment yielded only about half of the modern amino acids<sup>281,282</sup> suggesting that the first proteins on Earth may have contained only a few amino acids.

These findings go in parallel with the biosynthetic theory of the genetic code evolution suggesting that the genetic code evolved from a simpler form that encoded fewer amino acids,<sup>283</sup> probably paralleled by the invention of biosynthetic pathways for new and chemically more complex amino acids.<sup>284</sup> Furthermore, some additional support of the validity of this hypothesis can be found in the standard genetic code (that consists of  $4 \times 4 \times 4 = 64$  triplets of nucleotides, codons), which is redundant (64 codons encodes for 20 amino acids). In fact, with only two exceptions, codons encoding one amino acid may differ in any of their three positions. However, only the third positions of some codons may be fourfold degenerate, that is, any nucleotide at this position specifies the same amino acid and all nucleotide substitutions at this site are synonymous. Using these observations as a reflection of the evolutionary development, it was proposed that there was a period during code evolution where the third position was not needed at all and a doublet code preceded the triplet code, giving rise to  $4 \times 4 = 16$  codons encoding for 16 or fewer amino acids, if a termination codon is taken into account.<sup>285</sup>

Based on these and many other premises, one can discriminate evolutionary old and new amino acids. In 2000, Eduard N. Trifonov combined 40 different single-factor criteria into a consensus scale and proposed the following temporal order of addition for the amino acids: G/A, V/D, P, S, E/L, T, R, N, K, Q, I, C, H, F, M, Y, W.<sup>286</sup> Even superficial analysis of this sequence reveals that many of the early amino acids (such as G, D, E, P, and S) are disorder-promoting, as they are very abundant in modern IDPs. On the other hand, the major order-promoting residues (C, W, Y, and F) were added to the genetic code late. This observation is further illustrated by Figure 5(A) which represents modern genetic code, contains information on the early and late codons (shown by light red and light blue colors, respectively), and on corresponding disorder- and order-promoting residues (shown by red and blue colors, respectively). Codons with intermediate age and disorder-neutral residues are shown by light pink and pink colors, respectively. Figure 5(A)

**A**

	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC	UCC	UAC	UGC	C
	UUA Leu	UCA	UAA	UGA	A
	UUG	UCG	UAG	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CAU Arg	U
	CUC	CCC	CAC	CAC	C
	CUA	CCA	CAA Gln	CAA	A
	CUG	CCG	CAG	CAG	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC	ACC	AAC	AGC	C
	AUA Met	ACA	AAA Lys	AGA Arg	A
	AUG	ACG	AAG	AGG	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu	GGA	A
	GUG	GCG	GAG	GGG	G



**Figure 5.** Peculiarities of disorder evolution. A: Modern genetic code with information on the early and late codons (shown by light red and light blue colors, respectively) and disorder- and order-promoting residues (shown by red and blue colors, respectively). Codons with intermediate ages (i.e., those located between early and late codons) are shown by light pink color, whereas disorder-neutral residues are shown by pink color. B: Wavy pattern of the global disorder evolution. X-axis represents evolutionary time and Y-axis shows disorder content in proteins at given evolutionary time point. Here, primordial proteins are expected to be mostly disordered (left-hand side of the plot), proteins in LUA likely are mostly structured (center of the plot), whereas many protein in eukaryotes are either totally disordered or hybrids containing both ordered and disordered regions (right-hand side of the plot).

illustrates that there is relatively good agreement between the “age” of the residue and its disorder-promoting capacity, with early residues being mostly disorder-promoting, and with the majority of late residues being mostly order-promoting. This conclusion follows from the abundance of the matching colors (light red-red, light blue-blue, and light pink-pink). There are only two noticeable exceptions from these rule, valine and leucine, which are early but order-promoting residues. This strongly suggests that the primordial polypeptides were intrinsically disordered.

It is very unlikely that these disordered primordial polypeptides possessed catalytic activity.<sup>287</sup> This

hypothesis is in line with the RNA world theory suggesting that during the evolution of enzymatic activity, catalysis was transferred from RNA first to ribonucleoprotein (RNP) and only then to protein.<sup>288</sup> Therefore, the first proteins in the “breakthrough organism” (the first to have encoded protein synthesis) would be nonspecific chaperone-like proteins rather than catalysts.<sup>136,287</sup> Such RNA chaperone activities of early proteins conferred to their carriers a significant selective advantage in the RNA world, where RNA, which is especially prone to misfolding,<sup>289,290</sup> was used for both information storage and catalysis.<sup>291</sup> Since the variability of physicochemical properties of amino acids greatly exceeds that of

nucleotides and since protein structures are noticeably more stable than RNA structures, the transition from RNAs (ribozymes) to proteins as carriers of enzymatic activity was a logical evolutionary step. However, efficient catalysis relies on the proper spatial arrangement of catalytic residues which requires a stable structure.<sup>292</sup> Therefore, grafting of the enzymatic activity to proteins generated strong evolutionary pressure toward the well-folded structures.

In other words, the global evolution of intrinsic disorder is characterized by a wavy pattern [see Fig. 5(B)], where highly disordered primordial proteins with primarily RNA-chaperone activities were gradually substituted by the well-folded, highly ordered enzymes that evolved to catalyze the production of all the complex “goodies” crucial for the independent existence of the first cellular organisms. Due to its specific features crucial for the regulation of complex processes, protein intrinsic disorder was reinvented at the subsequent evolutionary steps leading to the development of more complex organisms from the last universal ancestor (i.e., the most recent organism from which all organisms now living on Earth descend<sup>293,294</sup>), and culminating in the appearance of the highly elaborated eukaryotic cells [see Fig. 5(B)].

#### ***Diversified evolution of IDPs in modern organisms***

There is no simple answer to the question on the comparative evolutionary rates of ordered and IDPs and regions in modern organisms. In fact, it looks like everything is possible, and intrinsically disordered sequences may evolve faster, slower or similar to ordered sequences. For example, disordered and ordered domains of the same protein (e.g., papillomavirus E7 oncoprotein) were shown to possess similar degrees of conservation and co-evolution.<sup>295</sup> Many other IDPs/IDPRs were shown to be characterized by high evolutionary rates<sup>151,296,297</sup> determined by the lack of specific structural restrictions. In fact, the analysis of calcineurins,<sup>10</sup> topoisomerase,<sup>298</sup> ribosomal protein S4,<sup>299</sup>  $\beta$ -subunits of the potassium channel Kv $\beta$ 1.1,<sup>300</sup> and many other proteins showed that disordered regions in these proteins contained more amino acid substitutions, insertions, and deletions than the ordered regions of the same proteins.<sup>151,301</sup> Furthermore, based on the observation that a significantly higher degree of positive Darwinian selection was observed in IDPRs of proteins compared to regions of  $\alpha$ -helix,  $\beta$ -sheet or tertiary structures, it was hypothesized that IDPRs may be required for the genetic variation with adaptive potential and that these regions may be of “central significance for the evolvability of the organism or cell in which they occur.”<sup>302</sup>

On the other hand, some IDPs and IDPRs are highly conserved. Human  $\alpha$ -synuclein (a canonical

IDP of 140 residues<sup>140,303</sup>) differs from its mouse counterpart by merely six residues (4%), and there are just 21 residue differences (12%, which include residue differences at 18 positions and 3 insertions/deletions) between the human and canary  $\alpha$ -synucleins.<sup>304</sup> In flagellin, the ordered central region has greater sequence diversity than its disordered termini.<sup>305</sup> Functionally important conserved regions of predicted disorder were shown to be rather common in proteins from all kingdoms of life, including viruses.<sup>306,307</sup> Furthermore, many functional domains of a significant size were shown to be intrinsically disordered.<sup>165</sup>

Overall, a systematic study of several families of proteins with at least one structurally characterized disordered region revealed that their IDPRs are characterized by highly heterogeneous evolutionary rates, with some disordered amino acid sequences evolving slowly, and others evolving more rapidly than ordered sequences.<sup>151</sup> Also, even different parts of the same disordered region can possess noticeable variability in their divergence during the evolutionary process.<sup>308</sup> Finally, in some disordered proteins, peculiarities of the amino acid composition, and not the amino acid sequence might be conserved.<sup>309,310</sup>

#### **Some Future Directions**

The last 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 which do not have rigid structures but possess crucial biological function was heavily underappreciated and ignored for a very long time despite numerous examples scattered in literature. The work which started in my group as an attempt to understand what is so special about several natively unfolded proteins produced a real explosion of interest to structure-less proteins with biological functions. A new field was created and a lot of intriguing information was produced related to structures and functions of IDPs/IDPRs. There is no need to list once again all the discoveries and findings made in this field—they are subjects of many recent reviews and some of them are briefly covered in this article. Although the amount of data generated during the past decade and a half on specific features related to the structural properties of IDPs and IDPRs, their abundance, distribution, functional repertoire, regulation, involvement into the disease pathogenesis, and so forth is vast, it seems that this mass of data produced so far is just a small tip of a humongous iceberg. IDPs/IDPRs continue to bring discoveries almost on a daily basis and even more breakthroughs are expected in future. Modern protein science is at the turning point, but biology still waits for physics. New models explaining various functions of IDPs, their evolution, and involvement in diseases are in great demand, together with the



general theory unifying current knowledge on protein structure and function, and with novel experimental and computational tools for focused studies of IDPs/IDPRs.

### Acknowledgments

This work would be impossible without numerous collaborators whose enthusiasm and help drove the studies on intrinsically disordered proteins for many years. Over the years I collaborated with more than 550 colleagues around the globe, and I am grateful to all former and current colleagues for their priceless contributions, assistance and support.

### References

- Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11:739–756.
- Fischer E (1894) Einfluss der configuration auf die wirkung der enzyme. *Ber Dt Chem Ges* 27:2985–2993.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Uversky VN, Dunker AK (2010) Understanding protein non-folding. *Biochim Biophys Acta* 1804:1231–1264.
- Kuhn TS (2012) The structure of scientific revolutions. Chicago: University Of Chicago Press.
- Petsko GA, Ringe D. 2004. Primers in biology. Protein structure and function. London: New Science Press Ltd., Sinauer Associates, Inc. Publishers, Blackwell Publishing.
- Bloomer AC, Champness JN, Bricogne G, Staden R, Klug A (1978) Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* 276:362–368.
- Bode W, Schwager P, Huber R (1978) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor complex and of its ternary complex with Ile-Val at 1.9 Å resolution. *J Mol Biol* 118:99–112.
- Le Gall T, Romero PR, Cortese MS, Uversky VN, Dunker AK (2007) Intrinsic disorder in the Protein Data Bank. *J Biomol Struct Dyn* 24:325–342.
- Dunker AK, Garner E, Guillot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput* 473–484.
- Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J Mol Biol* 293:321–331.
- Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
- Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. *J Mol Graph Model* 19:26–59.
- Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.
- Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK. Natively disordered proteins. In: Buchner J, Kiefhaber T, Eds. (2005) Handbook of protein folding. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA, pp. 271–353.
- Dunker AK, Babu MM, Barbar E, Blackledge M, Bondo SE, Dosztányi Z, Dyson HJ, Forman-Kay J, Fuxreiter M, Gsponer J, Han K-H, Jones DT, Longhi S, Metallo SJ, Nishikawa K, Nussinov R, Obradovic Z, Pappu R, Rost B, Selenko P, Subramaniam V, Sussman JL, Tompa P, Uversky VN (2013) What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins* 1:e24157.
- Holt C, Sawyer L (1993) Caseins as rheomorphic proteins: interpretation of primary and secondary structures of the  $\alpha$ 1-,  $\beta$ -, and  $\kappa$ -caseins. *J Chem Soc Faraday Trans* 89:2683–2692.
- Pullen RA, Jenkins JA, Tickle IJ, Wood SP, Blundell TL (1975) The relation of polypeptide hormone structure and flexibility to receptor binding: the relevance of X-ray studies on insulin, glucagon and human placental lactogen. *Mol Cell Biochem* 8:5–20.
- Cary PD, Moss T, Bradbury EM (1978) High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur J Biochem* 89:475–482.
- Linderstrom-Lang K, Schellman JA. Protein structure and enzyme activity. In: Boyer PD, Lardy H, Myrback K, Eds. (1959) The enzymes. New York: Academic Press, pp 443–510.
- Schweers O, Schonbrunn-Hanebeck E, Marx A, 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.
- Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT, Jr (1996) NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* 35:13709–13715.
- Chen J, Liang H, Fernandez A (2008) Protein structure protection commits gene expression patterns. *Genome Biol* 9:R107.
- Uversky VN (2003) A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J Biomol Struct Dyn* 21:211–234.
- Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ (2008) Malleable machines take shape in eukaryotic transcriptional regulation. *Nat Chem Biol* 4:728–737.
- Tsvetkov P, Asher G, Paz A, Reuven N, Sussman JL, Silman I, Shaul Y (2008) Operational definition of intrinsically unstructured protein sequences based on susceptibility to the 20S proteasome. *Proteins* 70:1357–1366.
- Dunker AK, Uversky VN (2010) Drugs for ‘protein clouds’: targeting intrinsically disordered transcription factors. *Curr Opin Pharmacol* 10:782–788.
- Livesay DR (2010) Protein dynamics: dancing on an ever-changing free energy stage. *Curr Opin Pharmacol* 10:706–708.
- Janin J, Sternberg MJE (2013) Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol Rep* 5:2.
- Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK (2008) TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* 15:956–963.
- Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. *Biophys J* 92:1439–1456.



32. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* 42:38–48.
33. Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK (1998) Predicting disordered regions from amino acid sequence: common themes despite differing structural characterization. *Genome Inform Ser Workshop Genome Inform* 9:201–213.
34. Williams RM, Obradovic Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker AK (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac Symp Biocomput* 89–100.
35. Vacic V, Uversky VN, Dunker AK, Lonardi S (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* 8:211.
36. Li X, Obradovic Z, Brown CJ, Garner EC, Dunker AK (2000) Comparing predictors of disordered protein. *Genome Inform Ser Workshop Genome Inform* 11:172–184.
37. Ferron F, Longhi S, Canard B, Karlin D (2006) A practical overview of protein disorder prediction methods. *Proteins* 65:1–14.
38. Bourhis JM, Canard B, Longhi S (2007) Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Curr Protein Pept Sci* 8:135–149.
39. Dosztanyi Z, Sandor M, Tompa P, Simon I (2007) Prediction of protein disorder at the domain level. *Curr Protein Pept Sci* 8:161–171.
40. Dosztanyi Z, Tompa P (2008) Prediction of protein disorder. *Methods Mol Biol* 426:103–115.
41. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res* 19:929–949.
42. Jin F, Liu Z (2013) Inherent relationships among different biophysical prediction methods for intrinsically disordered proteins. *Biophys J* 104:488–495.
43. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161–171.
44. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337:635–645.
45. Uversky VN (2010) The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol* 2010:568068.
46. Xue B, Dunker AK, Uversky VN (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* 30:137–149.
47. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK (1998) Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput* 437–448.
48. Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS (2006) Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Mol Biochem Parasitol* 150:256–267.
49. Tompa P, Dosztanyi Z, Simon I (2006) Prevalent structural disorder in *E. coli* and *S. cerevisiae* proteomes. *J Proteome Res* 5:1996–2000.
50. Galea CA, High AA, Obenauer JC, Mishra A, Park CG, Punta M, Schlessinger A, Ma J, Rost B, Slaughter CA, Kriwacki RW (2009) Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *J Proteome Res* 8:211–226.
51. Xue B, Williams RW, Oldfield CJ, Dunker AK, Uversky VN (2010) Archaic chaos: intrinsically disordered proteins in Archaea. *BMC Syst Biol* 4 (Suppl 1):S1.
52. Burra PV, Kalmar L, Tompa P (2010) Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS One* 5:e12069.
53. Mohan A, Sullivan WJ, Jr., Radivojac P, Dunker AK, Uversky VN (2008) Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes. *Mol Biosyst* 4:328–340.
54. Dryden DT, Thomson AR, White JH (2008) How much of protein sequence space has been explored by life on Earth? *J R Soc Interface* 5:953–956.
55. Wathen B, Jia Z (2009) Folding by numbers: primary sequence statistics and their use in studying protein folding. *Int J Mol Sci* 10:1567–1589.
56. Lau KF, Dill KA (1990) Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA* 87:638–642.
57. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 20:447–490.
58. Cordes MH, Davidson AR, Sauer RT (1996) Sequence space, folding and protein design. *Curr Opin Struct Biol* 6:3–10.
59. Riddle DS, Santiago JV, Bray-Hall ST, Doshi N, Grantcharova VP, Yi Q, Baker D (1997) Functional rapidly folding proteins from simplified amino acid sequences. *Nat Struct Biol* 4:805–809.
60. Plaxco KW, Riddle DS, Grantcharova V, Baker D (1998) Simplified proteins: minimalist solutions to the ‘protein folding problem’. *Curr Opin Struct Biol* 8:80–85.
61. Larson SM, England JL, Desjarlais JR, Pande VS (2002) Thoroughly sampling sequence space: large-scale protein design of structural ensembles. *Protein Sci* 11:2804–2813.
62. Guo HH, Choe J, Loeb LA (2004) Protein tolerance to random amino acid change. *Proc Natl Acad Sci USA* 101:9205–9210.
63. Doi N, Kakukawa K, Oishi Y, Yanagawa H (2005) High solubility of random-sequence proteins consisting of five kinds of primitive amino acids. *Protein Eng Des Sel* 18:279–284.
64. White SH, Jacobs RE (1990) Statistical distribution of hydrophobic residues along the length of protein chains. Implications for protein folding and evolution. *Biophys J* 57:911–921.
65. Murphy LR, Wallqvist A, Levy RM (2000) Simplified amino acid alphabets for protein fold recognition and implications for folding. *Protein Eng* 13:149–152.
66. Etchebest C, Benros C, Bornot A, Camproux AC, de Brevern AG (2007) A reduced amino acid alphabet for understanding and designing protein adaptation to mutation. *Eur Biophys J* 36:1059–1069.
67. Wang J, Wang W (1999) A computational approach to simplifying the protein folding alphabet. *Nat Struct Biol* 6:1033–1038.
68. Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys Acta* 1834:932–951.
69. Uversky VN, Dunker AK (2013) The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000 Biol Rep* 5:1.

70. Tompa P (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett* 579:3346–3354.
71. Dunker AK, Brown CJ, Obradovic Z (2002) Identification and functions of usefully disordered proteins. *Adv Protein Chem* 62:25–49.
72. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582.
73. Uversky VN (2002) What does it mean to be natively unfolded? *Eur J Biochem* 269:2–12.
74. Fuxreiter M, Tompa P (2012) Fuzzy complexes: a more stochastic view of protein function. *Adv Exp Med Biol* 725:1–14.
75. Tompa P, Fuxreiter M (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends Biochem Sci* 33:2–8.
76. Uversky VN (in press) Intrinsic disorder-based protein interactions and their modulators. *Curr Pharm Des*.
77. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9 (Suppl 1):S1.
78. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 18:343–384.
79. Dunker AK, Obradovic Z (2001) The protein trinity—linking function and disorder. *Nat Biotechnol* 19:805–806.
80. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. *Protein Eng* 12:713–720.
81. Ptitsyn O (1995) Molten globule and protein folding. *Adv Protein Chem* 47:83–229.
82. Tcherkasskaya O, Uversky VN (2003) Polymeric aspects of protein folding: a brief overview. *Protein Pept Lett* 10:239–245.
83. Vitalis A, Pappu RV (2009) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* 30:673–699.
84. Vitalis A, Wang X, Pappu RV (2007) Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys J* 93:1923–1937.
85. Vitalis A, Wang X, Pappu RV (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.
86. Wang X, Vitalis A, Wyczalkowski MA, Pappu RV (2006) Characterizing the conformational ensemble of monomeric polyglutamine. *Proteins* 63:297–311.
87. Williamson TE, Vitalis A, Crick SL, Pappu RV (2010) Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *J Mol Biol* 396:1295–1306.
88. Tran HT, Mao A, Pappu RV (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* 130:7380–7392.
89. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:8183–8188.
90. Muller-Spath S, Soranno A, Hirschfeld V, Hofmann H, Ruegger S, Reymond L, Nettels D, Schuler B (2010) From the cover: charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:14609–14614.
91. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44:1989–2000.
92. Ashbaugh HS, Hatch HW (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydrophobicity space. *J Am Chem Soc* 130:9536–9542.
93. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 580:2041–2045.
94. Yamada J, Phillips JL, Patel S, Goldfien G, Cales-tagne-Morelli A, Huang H, Reza R, Acheson J, Krishnan VV, Newsam S, Gopinathan A, Lau EY, Colvin ME, Uversky VN, Rexach MF (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol Cell Proteomics* 9:2205–2224.
95. Uversky VN (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.
96. Uversky VN, Li J, Fink AL (2001) Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* 276:10737–10744.
97. Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky VN (2003) Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins* 53:855–862.
98. Uversky VN, Permyakov SE, Zagranichny VE, Rodionov IL, Fink AL, Cherskaya AM, Wasserman LA, Permyakov EA (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.
99. Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM (1999) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* 38:15009–15016.
100. Konno T, Tanaka N, Kataoka M, Takano E, 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.
101. Lynn A, Chandra S, Malhotra P, Chauhan VS (1999) Heme binding and polymerization by *Plasmodium falciparum* histidine rich protein. II. Influence of pH on activity and conformation. *FEBS Lett* 459:267–271.
102. Johansson J, Gudmundsson GH, Rottenberg ME, Berndt KD, Agerberth B (1998) Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. *J Biol Chem* 273:3718–3724.
103. Magidovich E, Fleishman SJ, Yifrach O (2006) Intrinsically disordered C-terminal segments of voltage-activated potassium channels: a possible fishing rod-like mechanism for channel binding to scaffold proteins. *Bioinformatics* 22:1546–1550.
104. Roy S, Schnell S, Radivojac P (2006) Unraveling the nature of the segmentation clock: intrinsic disorder of clock proteins and their interaction map. *Comput Biol Chem* 30:241–248.
105. Minezaki Y, Homma K, Nishikawa K (2007) Intrinsically disordered regions of human plasma membrane

- proteins preferentially occur in the cytoplasmic segment. *J Mol Biol* 368:902–913.
106. De Biasio A, Guarnaccia C, Popovic M, Uversky VN, Pintar A, Pongor S (2008) Prevalence of intrinsic disorder in the intracellular region of human single-pass type I proteins: the case of the notch ligand Delta-4. *J Proteome Res* 7:2496–2506.
  107. Yang JY, Yang MQ, Dunker AK, Deng Y, Huang X (2008) Investigation of transmembrane proteins using a computational approach. *BMC Genomics* 9 (Suppl 1):S7.
  108. Xue B, Li L, Meroueh SO, Uversky VN, Dunker AK (2009) Analysis of structured and intrinsically disordered regions of transmembrane proteins. *Mol Biosyst* 5:1688–1702.
  109. Lenard J, Singer SJ (1966) Protein conformation in cell membrane preparations as studied by optical rotatory dispersion and circular dichroism. *Proc Natl Acad Sci USA* 56:1828–1835.
  110. Rosenbusch JP (1974) Characterization of the major envelope protein from *Escherichia coli*. Regular arrangement on the peptidoglycan and unusual dodecyl sulfate binding. *J Biol Chem* 249:8019–8029.
  111. Dunker AK, Jones TC (1978) Proposed knobs-into-holes packing for several membrane proteins. *Membr Biochem* 2:1–16.
  112. Paul C, Rosenbusch JP (1985) Folding patterns of porin and bacteriorhodopsin. *EMBO J* 4:1593–1597.
  113. Rees DC, DeAntonio L, Eisenberg D (1989) Hydrophobic organization of membrane proteins. *Science* 245:510–513.
  114. Stevens TJ, Arkin IT (1999) Are membrane proteins “inside-out” proteins? *Proteins* 36:135–143.
  115. Rees DC, Eisenberg D (2000) Turning a reference inside-out: commentary on an article by Stevens and Arkin entitled: “Are membrane proteins ‘inside-out’ proteins?” (*Proteins* 1999;36:135–143). *Proteins* 38:121–122.
  116. Von Heijne G (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the transmembrane topology. *EMBO J* 5:3021–3027.
  117. Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, Elofsson A (2008) Prediction of membrane-protein topology from first principles. *Proc Natl Acad Sci USA* 105:7177–7181.
  118. Siglioccolo A, Paiardini A, Piscitelli M, Pascarella S (2011) Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Struct Biol* 11:50.
  119. Zaccai G, Eisenberg H (1990) Halophilic proteins and the influence of solvent on protein stabilization. *Trends Biochem Sci* 15:333–337.
  120. Madern D, Ebel C, Zaccai G (2000) Halophilic adaptation of enzymes. *Extremophiles* 4:91–98.
  121. Mevarech M, Frolow F, Gloss LM (2000) Halophilic enzymes: proteins with a grain of salt. *Biophys Chem* 86:155–164.
  122. Wright DB, Banks DD, Lohman JR, Hilsenbeck JL, Gloss LM (2002) The effect of salts on the activity and stability of *Escherichia coli* and *Haloferax volcanii* dihydrofolate reductases. *J Mol Biol* 323:327–344.
  123. Bergqvist S, Williams MA, O'Brien R, Ladbury JE (2003) Halophilic adaptation of protein–DNA interactions. *Biochem Soc Trans* 31:677–680.
  124. Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K (2003) Unique amino acid composition of proteins in halophilic bacteria. *J Mol Biol* 327:347–357.
  125. Paul S, Bag SK, Das S, Harvill ET, Dutta C (2008) Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* 9:R70.
  126. Tadeo X, Lopez-Mendez B, Trigueros T, Lain A, Castano D, Millet O (2009) Structural basis for the amino acid composition of proteins from halophilic archaea. *PLoS Biol* 7:e1000257.
  127. Elcock AH, McCammon JA (1998) Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol* 280:731–748.
  128. Xue B, Williams RW, Oldfield CJ, Goh GK, Dunker AK, Uversky VN (2010) Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept Lett* 17:932–951.
  129. Berezovsky IN (2011) The diversity of physical forces and mechanisms in intermolecular interactions. *Phys Biol* 8:035002.
  130. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323:573–584.
  131. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:54–60.
  132. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
  133. Fink AL (2005) Natively unfolded proteins. *Curr Opin Struct Biol* 15:35–41.
  134. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J* 272:5129–5148.
  135. Uversky VN (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.
  136. Tompa P, Csermely P (2004) The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* 18:1169–1175.
  137. Tompa P, Szasz C, Buday L (2005) Structural disorder throws new light on moonlighting. *Trends Biochem Sci* 30:484–489.
  138. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z (2007) Functional anthology of intrinsic disorder. I. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* 6:1882–1898.
  139. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2007) Functional anthology of intrinsic disorder. III. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res* 6:1917–1932.
  140. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN (2007) Functional anthology of intrinsic disorder. II. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J Proteome Res* 6:1899–1916.
  141. Cortese MS, Uversky VN, Dunker AK (2008) Intrinsic disorder in scaffold proteins: getting more from less. *Prog Biophys Mol Biol* 98:85–106.
  142. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18:756–764.



143. Dunker AK, Uversky VN (2008) Signal transduction via unstructured protein conduits. *Nat Chem Biol* 4:229–230.
144. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN (2008) The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 9 (Suppl 2):S1.
145. Russell RB, Gibson TJ (2008) A careful disorderliness in the proteome: sites for interaction and targets for future therapies. *FEBS Lett* 582:1271–1275.
146. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN (2009) Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays* 31:328–335.
147. Uversky VN, Dunker AK (2008) *Biochemistry. Controlled chaos*. *Science* 322:1340–1341.
148. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
149. Uversky VN (2012) Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* 43:1090–1103.
150. Dunker AK, Obradovic Z, Romero P, Kissinger C, Villafraña E (1997) On the importance of being disordered. *PDB Newsletter* 81:3–5.
151. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol* 55:104–110.
152. Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK (2006) Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci USA* 103:8390–8395.
153. Dosztanyi Z, Csizmek V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21:3433–3434.
154. Spolar RS, Record MT, Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science* 263:777–784.
155. Schulz GE. Nucleotide binding proteins. In: Balaban M, Ed. (1979) *Molecular mechanism of biological recognition*. New York: Elsevier/North-Holland Biomedical Press, pp 79–94.
156. Pontius BW (1993) Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci* 18:181–186.
157. Plaxco KW, Gross M (1997) Cell biology. The importance of being unfolded. *Nature* 386:657, 659.
158. Karush F (1950) Heterogeneity of the binding sites of bovine serum albumin. *J Am Chem Soc* 72:2705–2713.
159. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21<sup>Waf1/Cip1/Sdi1</sup> in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 93:11504–11509.
160. Ekman D, Light S, Bjorklund AK, Elofsson A (2006) What properties characterize the hub proteins of the protein–protein interaction network of *Saccharomyces cerevisiae*? *Genome Biol* 7:R45.
161. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol* 2:e100.
162. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 5:2985–2995.
163. Singh GP, Ganapathi M, Dash D (2007) Role of intrinsic disorder in transient interactions of hub proteins. *Proteins* 66:761–765.
164. Singh GP, Dash D (2007) Intrinsic disorder in yeast transcriptional regulatory network. *Proteins* 68:602–605.
165. Balazs A, Csizmek V, Buday L, Rakacs M, Kiss R, Bokor M, Udupa R, Tompa K, Tompa P (2009) High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1. *FEBS J* 276:3744–3756.
166. Jeffery CJ (1999) Moonlighting proteins. *Trends Biochem Sci* 24:8–11.
167. Jeffery CJ (2003) Multifunctional proteins: examples of gene sharing. *Ann Med* 35:28–35.
168. Jeffery CJ (2004) Molecular mechanisms for multi-tasking: recent crystal structures of moonlighting proteins. *Curr Opin Struct Biol* 14:663–668.
169. Hsu WL, Oldfield CJ, Xue B, Meng J, Huang F, Romero P, Uversky VN, Dunker AK (2013) Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci* 22:258–273.
170. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393:440–442.
171. Goh KI, Oh E, Jeong H, Kahng B, Kim D (2002) Classification of scale-free networks. *Proc Natl Acad Sci USA* 99:12583–12588.
172. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–512.
173. Agarwal S, Deane CM, Porter MA, Jones NS (2010) Revisiting date and party hubs: novel approaches to role assignment in protein interaction networks. *PLoS Comput Biol* 6:e1000817.
174. Liu W, Rui H, Wang J, Lin S, He Y, Chen M, Li Q, Ye Z, Zhang S, Chan SC, Chen YG, Han J, Lin SC (2006) Axin is a scaffold protein in TGF-beta signaling that promotes degradation of Smad7 by Arkadia. *EMBO J* 25:1646–1658.
175. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK (2006) Intrinsic disorder in transcription factors. *Biochemistry* 45:6873–6888.
176. Minezaki Y, Homma K, Kinjo AR, Nishikawa K (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol* 359:1137–1149.
177. Toth-Petroczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M (2008) Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol* 4:e1000243.
178. Peng Z, Mizianty MJ, Xue B, Kurgan L, Uversky VN (2012) More than just tails: intrinsic disorder in histone proteins. *Mol Biosyst* 8:1886–1901.
179. Peng Z, Xue B, Mizianty MJ, Oldfield CJ, Kurgan L, Uversky VN (in press) A creature with hundred of waggly tails: intrinsically disordered proteins in ribosome. *Biochim Biophys Acta Mol Cell Res*.
180. Westerheide SD, Raynes R, Powell C, Xue B, Uversky VN (2012) HSF transcription factor family, heat shock response, and protein intrinsic disorder. *Curr Protein Pept Sci* 13:86–103.
181. Xue B, Oldfield CJ, Van YY, Dunker AK, Uversky VN (2012) Protein intrinsic disorder and induced pluripotent stem cells. *Mol Biosyst* 8:134–150.
182. Xue B, Dunker AK, Uversky VN (2012) The roles of intrinsic disorder in orchestrating the Wnt-pathway. *J Biomol Struct Dyn* 29:843–861.



183. Peng Z, Xue B, Kurgan L, Uversky VN (in press) Resilience of death: Intrinsic disorder in proteins involved in the programmed cell death. *Cell Death Differ*.
184. Pervushin K, Vamvaca K, Vogeli B, Hilvert D (2007) Structure and dynamics of a molten globular enzyme. *Nat Struct Mol Biol* 14:1202–1206.
185. Gatti-Lafranconi P, Hollfelder F (2013) Flexibility and reactivity in promiscuous enzymes. *ChemBiochem* 14:285–292.
186. Karplus M (1984) Dynamics of proteins. *Adv Biophys* 18:165–190.
187. Hammes GG (2002) Multiple conformational changes in enzyme catalysis. *Biochemistry* 41:8221–8228.
188. Boehr DD, Dyson HJ, Wright PE (2006) An NMR perspective on enzyme dynamics. *Chem Rev* 106:3055–3079.
189. Hammes-Schiffer S, Benkovic SJ (2006) Relating protein motion to catalysis. *Annu Rev Biochem* 75:519–541.
190. Olsson MH, Parson WW, Warshel A (2006) Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis. *Chem Rev* 106:1737–1756.
191. Koshland DE, Jr, Neet KE (1968) The catalytic and regulatory properties of enzymes. *Annu Rev Biochem* 37:359–410.
192. Vendruscolo M (2010) Enzymatic activity in disordered states of proteins. *Curr Opin Chem Biol* 14:671–675.
193. Vamvaca K, Vogeli B, Kast P, Pervushin K, Hilvert D (2004) An enzymatic molten globule: efficient coupling of folding and catalysis. *Proc Natl Acad Sci USA* 101:12860–12864.
194. Vamvaca K, Jelesarov I, Hilvert D (2008) Kinetics and thermodynamics of ligand binding to a molten globular enzyme and its native counterpart. *J Mol Biol* 382:971–977.
195. Woycechowsky KJ, Choutko A, Vamvaca K, Hilvert D (2008) Relative tolerance of an enzymatic molten globule and its thermostable counterpart to point mutation. *Biochemistry* 47:13489–13496.
196. Uversky VN, Kutyschenko VP, Protasova N, Rogov VV, Vassilenko KS, Gudkov AT (1996) Circularly permuted dihydrofolate reductase possesses all the properties of the molten globule state, but can resume functional tertiary structure by interaction with its ligands. *Protein Sci* 5:1844–1851.
197. Protasova N, Kireeva ML, Murzina NV, Murzin AG, Uversky VN, Gryaznova OI, Gudkov AT (1994) Circularly permuted dihydrofolate reductase of *E. coli* has functional activity and a destabilized tertiary structure. *Protein Eng* 7:1373–1377.
198. Zambelli B, Stola M, Musiani F, De Vriendt K, Samyn B, Devreese B, Van Beeumen J, Turano P, Dikiy A, Bryant DA, Ciurli S (2005) UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn<sup>2+</sup>. *J Biol Chem* 280:4684–4695.
199. Neyroz P, Zambelli B, Ciurli S (2006) Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy. *Biochemistry* 45:8918–8930.
200. Zambelli B, Cremades N, Neyroz P, Turano P, Uversky VN, Ciurli S (2012) Insights in the (un)structural organization of *Bacillus pasteurii* UreG, an intrinsically disordered GTPase enzyme. *Mol Biosyst* 8:220–228.
201. Uversky VN (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev* 40:1623–1634.
202. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS (2007) Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci USA* 104:9650–9655.
203. Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, Kay LE, Tyers M, Forman-Kay JD (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci USA* 105:17772–17777.
204. Mittag T, Kay LE, Forman-Kay JD (2010) Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* 23:105–116.
205. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD (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.
206. Sigalov AB, Kim WM, Saline M, Stern LJ (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.
207. Sigalov A, Aivazian D, 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.
208. Sigalov AB, Zhuravleva AV, Orekhov VY (2007) Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form. *Biochimie* 89:419–421.
209. Pometun MS, Chekmenev EY, Wittebort RJ (2004) Quantitative observation of backbone disorder in native elastin. *J Biol Chem* 279:7982–7987.
210. Sigalov AB, Hendricks GM (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.
211. Sigalov AB, Aivazian DA, Uversky VN, Stern LJ (2006) Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits. *Biochemistry* 45:15731–15739.
212. Hazy E, Tompa P (2009) Limitations of induced folding in molecular recognition by intrinsically disordered proteins. *Chemphyschem* 10:1415–1419.
213. Liu Y, Fratini E, Baglioni P, Chen WR, Chen SH (2005) Effective long-range attraction between protein molecules in solutions studied by small angle neutron scattering. *Phys Rev Lett* 95:118102.
214. Meyer TE, Yakali E, Cusanovich MA, Tollin G (1987) Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry* 26:418–423.
215. Hoff WD, van Stokkum IH, van Ramesdonk HJ, van Brederode ME, Brouwer AM, Fitch JC, Meyer TE, van Grondelle R, Hellingwerf KJ (1994) Measurement and global analysis of the absorbance changes in the photocycle of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Biophys J* 67:1691–1705.
216. Rubinstenn G, Vuister GW, Mulder FA, Dux PE, Boelens R, Hellingwerf KJ, Kaptein R (1998) Structural and dynamic changes of photoactive yellow protein

- during its photocycle in solution. *Nat Struct Biol* 5:568–570.
217. Ramachandran PL, Lovett JE, Carl PJ, Cammarata M, Lee JH, Jung YO, Ihee H, Timmel CR, van Thor JJ (2011) The short-lived signaling state of the photoactive yellow protein photoreceptor revealed by combined structural probes. *J Am Chem Soc* 133:9395–9404.
  218. Dux P, Rubinstenn G, Vuister GW, Boelens R, Mulder FA, Hard K, Hoff WD, Kroon AR, Crielaard W, Hellingwerf KJ, Kaptein R (1998) Solution structure and backbone dynamics of the photoactive yellow protein. *Biochemistry* 37:12689–12699.
  219. Borgstahl GE, Williams DR, Getzoff ED (1995) 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* 34:6278–6287.
  220. Gustin MC, Zhou XL, Martinac B, Kung C (1988) A mechanosensitive ion channel in the yeast plasma membrane. *Science* 242:762–765.
  221. Corry B, Hurst AC, Pal P, Nomura T, Rigby P, Martinac B (2010) An improved open-channel structure of MscL determined from FRET confocal microscopy and simulation. *J Gen Physiol* 136:483–494.
  222. Gsponer J, Futschik ME, Teichmann SA, Babu MM (2008) Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* 322:1365–1368.
  223. Grimmler M, Wang Y, Mund T, Cilensek Z, Keidel EM, Waddell MB, Jakel H, Kullmann M, Kriwacki RW, Hengst L (2007) Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell* 128:269–280.
  224. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32:1037–1049.
  225. Paliy O, Gargac SM, Cheng Y, Uversky VN, Dunker AK (2008) Protein disorder is positively correlated with gene expression in *Escherichia coli*. *J Proteome Res* 7:2234–2245.
  226. Fernandez A, Scheraga HA (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci USA* 100:113–118.
  227. Fernandez A (2004) Keeping dry and crossing membranes. *Nat Biotechnol* 22:1081–1084.
  228. Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* 8:101–106.
  229. Dobson CM (1999) Protein misfolding, evolution and disease. *Trends Biochem Sci* 24:329–332.
  230. Bellotti V, Mangione P, Stoppini M (1999) Biological activity and pathological implications of misfolded proteins. *Cell Mol Life Sci* 55:977–991.
  231. Uversky VN, Talapatra A, Gillespie JR, Fink AL (1999) Protein deposits as the molecular basis of amyloidosis. I. Systemic amyloidoses. *Med Sci Monitor* 5:1001–1012.
  232. Uversky VN, Talapatra A, Gillespie JR, Fink AL (1999) Protein deposits as the molecular basis of amyloidosis. II. Localized amyloidosis and neurodegenerative disorders. *Med Sci Monitor* 5:1238–1254.
  233. Rochet JC, Lansbury PT, Jr (2000) Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* 10:60–68.
  234. Uversky VN, Fink AL (2004) Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta* 1698:131–153.
  235. Uversky VN, Fink AL. Pathways to amyloid fibril formation: partially folded intermediates in fibrillation of unfolded proteins. In: Sipe JD, Ed. (2005) *Amyloid proteins: the beta pleated sheet conformation and disease*. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co. KGaA, pp 247–265.
  236. Glenner GG, Wong CW (1984) Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem Biophys Res Commun* 122:1131–1135.
  237. Masters CL, Multhaup G, Simms G, Pottgiesser J, Martins RN, 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.
  238. Lee VM, Balin BJ, Otvos L, Jr, Trojanowski JQ (1991) A68: a major subunit of paired helical filaments and derivatized forms of normal Tau. *Science* 251:675–678.
  239. Ueda K, Fukushima H, Masliah E, Xia Y, Iwai A, Yoshimoto M, Otero DA, Kondo J, Ihara Y, Saitoh T (1993) Molecular cloning of cDNA encoding an unrecognized component of amyloid in Alzheimer disease. *Proc Natl Acad Sci USA* 90:11282–11286.
  240. Wisniewski KE, Dalton AJ, McLachlan C, Wen GY, Wisniewski HM (1985) Alzheimer's disease in Down's syndrome: clinicopathologic studies. *Neurology* 35:957–961.
  241. Dev KK, Hofele K, Barbieri S, Buchman VL, van der Putten H (2003) Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease. *Neuropharmacology* 45:14–44.
  242. Prusiner SB (2001) Shattuck lecture—neurodegenerative diseases and prions. *N Engl J Med* 344:1516–1526.
  243. Zoghbi HY, Orr HT (1999) Polyglutamine diseases: protein cleavage and aggregation. *Curr Opin Neurobiol* 9:566–570.
  244. Okazawa H (2003) Polyglutamine diseases: a transcription disorder? *Cell Mol Life Sci* 60:1427–1439.
  245. Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* 9:909–916.
  246. Gusella JF, MacDonald ME (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat Rev Neurosci* 1:109–115.
  247. Orr HT (2001) Beyond the Qs in the polyglutamine diseases. *Genes Dev* 15:925–932.
  248. Fischbeck KH (2001) Polyglutamine expansion neurodegenerative disease. *Brain Res Bull* 56:161–163.
  249. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J Biol Chem* 275:29426–29432.
  250. Adkins JN, Lumb KJ (2002) Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. *Proteins* 46:1–7.
  251. Chang BS, Minn AJ, Muchmore SW, Fesik SW, Thompson CB (1997) Identification of a novel regulatory domain in Bcl-X(L) and Bcl-2. *EMBO J* 16:968–977.
  252. Campbell KM, Terrell AR, Laybourn PJ, Lumb KJ (2000) Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry* 39:2708–2713.
  253. Sunde M, McGrath KC, Young L, Matthews JM, Chua EL, Mackay JP, Death AK (2004) TC-1 is a

- novel tumorigenic and natively disordered protein associated with thyroid cancer. *Cancer Res* 64:2766–2773.
254. Ng KP, Potikyan G, Savene RO, Denny CT, Uversky VN, Lee KA (2007) Multiple aromatic side chains within a disordered structure are critical for transcription and transforming activity of EWS family oncoproteins. *Proc Natl Acad Sci USA* 104:479–484.
  255. Erkizan HV, Uversky VN, Toretzky JA (2010) Oncogenic partnerships: EWS-FLI1 protein interactions initiate key pathways of Ewing's sarcoma. *Clin Cancer Res* 16:4077–4083.
  256. Anderson CW, Appella E. Signaling to the p53 tumor suppressor through pathways activated by genotoxic and nongenotoxic stress. In: Bradshaw RA, Dennis EA, Eds. (2004) *Handbook of cell signaling*. New York: Academic Press, pp 237–247.
  257. Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991) p53 mutations in human cancers. *Science* 253:49–53.
  258. Balint EE, Vousden KH (2001) Activation and activities of the p53 tumour suppressor protein. *Br J Cancer* 85:1813–1823.
  259. Chene P (2004) Inhibition of the p53-hdm2 interaction with low molecular weight compounds. *Cell Cycle* 3:460–461.
  260. Chene P (2004) Inhibition of the p53-MDM2 interaction: targeting a protein-protein interface. *Mol Cancer Res* 2:20–28.
  261. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215–246.
  262. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* 10 (Suppl 1):S7.
  263. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z (1999) Predicting binding regions within disordered proteins. *Genome Inform Ser Workshop Genome Inform* 10:41–50.
  264. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 46:13468–13477.
  265. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44:12454–12470.
  266. Chen CT, Wagner H, Still WC (1998) Fluorescent, sequence-selective peptide detection by synthetic small molecules. *Science* 279:851–853.
  267. Kussie PH, Gorina S, Marechal V, Elenbaas B, Morneau J, Levine AJ, Pavletich NP (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274:948–953.
  268. Vassilev LT (2004) Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle* 3:419–421.
  269. Klein C, Vassilev LT (2004) Targeting the p53-MDM2 interaction to treat cancer. *Br J Cancer* 91:1415–1419.
  270. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu EA (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303:844–848.
  271. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YY, Romero P, Cortese MS, Uversky VN, Dunker AK (2006) Rational drug design via intrinsically disordered protein. *Trends Biotechnol* 24:435–442.
  272. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* 3:301–317.
  273. Arkin M (2005) Protein-protein interactions and cancer: small molecules going in for the kill. *Curr Opin Chem Biol* 9:317–324.
  274. Cochran AG (2000) Antagonists of protein-protein interactions. *Chem Biol* 7:R85–R94.
  275. Hammoudeh DI, Follis AV, Prochownik EV, Metallo SJ (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J Am Chem Soc* 131:7390–7401.
  276. Uversky VN (2011) Intrinsically disordered proteins may escape unwanted interactions via functional misfolding. *Biochim Biophys Acta* 1814:693–712.
  277. Uversky VN (2012) Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin Drug Discov* 7:475–488.
  278. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. *Gene* 344:1–20.
  279. Oparin AI (1924) *The origin of life (in Russian)*. Moscow: Moscow Worker publisher.
  280. Haldane JBS. *The origin of life*. In: Watts CA, Ed. (1929) *The rationalist annual for the year 1929*. London: Watts & Co, pp 3–10.
  281. Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* 117:528–529.
  282. Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. *Science* 130:245–251.
  283. Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38:367–379.
  284. Wong JT (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* 72:1909–1912.
  285. Jukes TH (1973) Possibilities for the evolution of the genetic code from a preceding form. *Nature* 246:22–26.
  286. Trifonov EN (2000) Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261:139–151.
  287. Poole AM, Jeffares DC, Penny D (1998) The path from the RNA world. *J Mol Evol* 46:1–17.
  288. Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. *J Mol Evol* 46:18–36.
  289. Treiber DK, Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* 11:309–314.
  290. Cristofari G, Darlix JL (2002) The ubiquitous nature of RNA chaperone proteins. *Prog Nucleic Acid Res Mol Biol* 72:223–268.
  291. Gilbert W (1986) Origin of life—the RNA world. *Nature* 319:618–618.
  292. Csermely P (1997) Proteins, RNAs and chaperones in enzyme evolution: a folding perspective. *Trends Biochem Sci* 22:147–149.
  293. Doolittle WF (2000) Uprooting the tree of life. *Sci Am* 282:90–95.
  294. Theobald DL (2010) A formal test of the theory of universal common ancestry. *Nature* 465:219–222.
  295. Chemes LB, Glavina J, Alonso LG, Marino-Buslje C, de Prat-Gay G, Sanchez IE (2012) Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein. *PLoS One* 7:e47661.
  296. Chen SC, Chuang TJ, Li WH (2011) The relationships among microRNA regulation, intrinsically disordered



- regions, and other indicators of protein evolutionary rate. *Mol Biol Evol* 28:2513–2520.
297. Lin YS, Hsu WL, Hwang JK, Li WH (2007) Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol* 24:1005–1011.
  298. Shaiu WL, Hu T, Hsieh TS (1999) The hydrophilic, protease-sensitive terminal domains of eucaryotic DNA topoisomerases have essential intracellular functions. *Pac Symp Biocomput*:578–589.
  299. Sayers EW, Gerstner RB, Draper DE, Torchia DA (2000) Structural preordering in the N-terminal region of ribosomal protein S4 revealed by heteronuclear NMR spectroscopy. *Biochemistry* 39:13602–13613.
  300. Wissmann R, Baukowitz T, Kalbacher H, Kalbitzer HR, Ruppertsberg JP, Pongs O, Antz C, Fakler B (1999) NMR structure and functional characteristics of the hydrophilic N terminus of the potassium channel beta-subunit Kvbeta1.1. *J Biol Chem* 274:35521–35525.
  301. Brown CJ, Johnson AK, Dunker AK, Daughdrill GW (2011) Evolution and disorder. *Curr Opin Struct Biol* 21:441–446.
  302. Nilsson J, Grahn M, Wright AP (2011) Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol* 12:R65.
  303. Breydo L, Wu JW, Uversky VN (2012) Alpha-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta* 1822:261–285.
  304. George JM (2002) The synucleins. *Genome Biol* 3:REVIEWS3002.
  305. Vonderviszt F, Kanto S, Aizawa S, Namba K (1989) Terminal regions of flagellin are disordered in solution. *J Mol Biol* 209:127–133.
  306. Chen JW, Romero P, Uversky VN, Dunker AK (2006) Conservation of intrinsic disorder in protein domains and families. I. A database of conserved predicted disordered regions. *J Proteome Res* 5:879–887.
  307. Chen JW, Romero P, Uversky VN, Dunker AK (2006) Conservation of intrinsic disorder in protein domains and families. II. functions of conserved disorder. *J Proteome Res* 5:888–898.
  308. Xue B, Brown CJ, Dunker AK, Uversky VN (2013) Intrinsically disordered regions of p53 family are highly diversified in evolution. *Biochim Biophys Acta* 1834:725–738.
  309. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ (2007) Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol* 65:277–288.
  310. Moesa HA, Wakabayashi S, Nakai K, Patil A (2012) Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Mol Biosyst* 8:3262–3273.