Intrinsically disordered proteins from A to Z

Vladimir N. Uversky*

Department of Molecular Medicine, University of South Florida, FL 33612, USA;

Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia;

* To whom correspondence should be addressed at Department of Molecular Medicine, University of South Florida, College of Medicine, 12901 Bruce B. Downs Blvd, MDC07, Tampa, FL 33612, USA ; Phone: 1-813-974-5816; Fax: 1-813-974-7357; E-mail: vuversky@health.usf.edu

Abstract

The ideas that proteins might possess specific functions without being uniquely folded into rigid 3D-structures and that these floppy polypeptides might constitute a noticeable part of any given proteome would have been considered as a preposterous fiction 15 or even 10 years ago. The situation has changed recently, and the existence of functional yet intrinsically disordered proteins and regions has become accepted by a significant number of protein scientists. These fuzzy objects with fuzzy structures and fuzzy functions are among the most interesting and attractive targets for modern protein research. This review summarizes some of the major discoveries and breakthroughs in the field of intrinsic disorder by representing related concepts and definitions.

Key words: Intrinsically disordered proteins; protein-protein interaction; protein structure; protein function; protein non-folding

Abbreviations: AA, amino acid; BRCA1, breast cancer 1; CDK, cyclin-dependent kinase; CPD Cdc4 phosphodegron; ELM, eukaryotic linear motif; IDP, intrinsically disordered protein; IDR, intrinsically disordered region; LM, linear motif; MBP, Myelin basic protein; MeCP2, methyl-CpG binding protein 2; MoRF, molecular recognition feature; Nup, nucleoporin; PCS, primary contact site; PONDR, predictor of natural disordered regions; PTM, posttranslational modification; SLiM, short linear motif; XPA, Xeroderma pigmentosum group A-complementing protein;

1. Rise of intrinsically disordered proteins: Another Cinderella story

Despite the widely-accepted protein structure-function paradigm, which dominated scientific minds for more than 100 years, intrinsically flexible but biologically active proteins have been repeatedly observed during the modern history of protein science. However, for a long time, information about such proteins was very sparse, and the idea that these non-rigid proteins are not mere rare and obscure exceptions but instead might constitute an important and novel structure-function class failed to hold. As a result the concept of biological function originating from conformationally flexible proteins was re-discovered many times. This repeated re-discovery gave rise to a multitude of terms introduced to describe these "non-traditional" proteins. A short list of names used by different authors over the years includes: pliable [1], floppy [2], rheomorphic [3], flexible [4], mobile [5], partially folded [6], natively denatured [7], natively unfolded [8, 9], natively disordered [10], intrinsically unstructured [11, 12], intrinsically denatured, [7] intrinsically unfolded [8], intrinsically disordered [13], vulnerable [14], chameleon [15], malleable [16], 4D [17], and dancing proteins [18], as well as protein clouds [19, 20] and even 3^2 proteins [15]. The situation changed at the turn of the century, when, based on the exhaustive analysis of available then literature data, the important hypothesis was formulated that naturally flexible proteins, instead of being just rare exceptions, represent a unique and very broad class of proteins [9, 11-13]. Although these proteins are still described by different names, over the recent years the terms "intrinsically disordered protein" (IDP) is becoming more widely used than other terms. The accepted meaning of the "intrinsic disorder" is based on notion that a corresponding protein (or protein region) is biologically active, yet exists as collapsed or extended, dynamically mobile conformational ensemble, either at the level of secondary or tertiary structure [9-13, 21, 22].

Currently, IDP-related research is one of the mostly dynamic and popular fields in modern protein science. In essence, the development of this field represents another Cinderella story, where hard working housekeeping girl (IDPs) was originally neglected, but suddenly became a star, when her little glass slipper (natural abundance and biological importance) was found.

2. Intrinsically disordered proteins from A to Z

*2.1. **A**: Amino acid sequence determinants of intrinsic disorder*

Alphabet starts with A. Protein starts with amino acid sequence. Therefore, the description of some of the IDP sequence peculiarities (determinants of intrinsic disorder), which are at the root of difference between "normal" (ordered) and "abnormal" (intrinsically disordered) proteins, seems to be a very logical start of this review.

The ability of a protein to fold or not to fold under the physiologic conditions is encoded in its amino acid sequence [9, 13, 21-24]. Based on the analysis of 275 natively folded and 91 natively unfolded proteins (i.e., native coils and native pre-molten globules, which are highly disordered proteins possessing little or no tertiary and secondary structure under physiologic conditions) it

has been concluded that the combination of low mean hydrophobicity (leading to low driving force for protein compaction) and high net charge (leading to strong electrostatic repulsion) represents an important prerequisite for the absence of compact structure in a protein [9]. Figure 1 represents the resulting dependence of the absolute mean charge on the mean normalized hydrophobicity for these two protein sets and shows that the extended IDPs (native coils or native pre-molten globules) occupy unique region of the charge-hydropathy phase space and that they are reliably separated from the compact globular proteins by a simple linear boundary.

More detailed analysis revealed that in comparison with "normal" ordered proteins, IDPs were significantly depleted in so-called order-promoting amino acids such as Ile, Leu, Val, Trp, Tyr, Phe, Cys, and Asn, being substantially enriched in disorder-promoting amino acids, Ala, Arg, Gly, Gln, Ser, Glu, Lys, and Pro [13, 25-28]. Furthermore, several other attributes of disordered segments stood along, and ten of these attributes, such as 14 Å contact number, hydropathy, flexibility, β-sheet propensity, coordination number, Arg+Glu+Ser+Pro, bulkiness, Cys+Phe+Tyr+Trp, volume, and net charge, provided rather reliable discrimination between ordered and disordered proteins [13]. Since disordered regions shared at least some common sequence features over many proteins [21, 29], these features made them recognizable and were used to develop specific predictors of intrinsic disorder. Development of such predictors provided a direct support for the hypothesis that intrinsic disorder is encoded in protein amino acid sequences. Currently, design of algorithms for finding regions lacking ordered structure is a very active area of research, and more than 50 predictors of disorder have been developed [30].

*2.2. **B**: Binding promiscuity*

Many IDPs are known to be promiscuous, being involved in numerous interactions with multiple partners. They frequently serve as the popular nodes, or hubs, in protein interaction networks [31-38]. Hubs are central to the normal function and stability of the protein-protein interaction network in any organism, and the deletion of a hub has been shown to be lethal to the organism [39]. Therefore, many important IDPs involved in the pathogenesis of human diseases are hubs. Illustrative examples of disordered hubs with tens and even hundreds of binding partners are p53, α-synuclein, BRCA1, XPA, p21, p27, estrogen receptor, etc. (see below, Section 2.4, for more detail). The analysis of the potential roles of intrinsic disorder in protein network architecture revealed that (i) intrinsic disordered regions (IDRs) can serve as the structural basis for hub protein promiscuity promoting the ability of one IDP-hub to interact with many binding partners; (ii) multiple IDPs can bind to same structured hub proteins, and (iii) IDRs can serve as flexible linkers between functional domains with the linkers enabling mechanisms that facilitate binding diversity and promiscuity [31, 32]. Therefore, in disordered hub proteins, IDRs may be present in one of at least two forms [40], flexible linker that connects two ordered domains allowing their unrestricted movement with respect to each other, and direct binding site that often undergoes disorder-to-order transition at binding (see Sections 2.9 and 2.13 for more details). However, binding of IDPs to

their partners is not always associated with folding and one of the proposed binding modes includes highly dynamic complexes [41-50] (see Section 2.3. for more detail).

*2.3. **C**: Charge rules*

Charged residues are obviously important for defining structure and function of IDPs (see Figure 1). Section 2.1 mentioned that one of the most striking features of highly disordered proteins (native coils or native pre-molten globules) is their high content of charged residues [9]. This high net charge is crucial for extended conformation to occur [51], since sequences that are rich in uncharged, polar amino acids and devoid of canonical hydrophobic residues are repeatedly shown to form heterogeneous ensembles of collapsed structures in aqueous solutions [52-58]. Based on the analysis of a set of highly charged polypeptides, it has been concluded that charge content can modulate the intrinsic preference of polypeptide backbones for collapsed structures [51]. Furthermore, the extensive simulation of coarse-grained polypeptides over a wide range of sequence hydrophobicities, charges, and lengths revealed that stability of a natively unfolded protein can be described as a coil-to-globule transition in a charge/hydropathy sequence composition space [59]. In agreement with these conclusions, the recent structural analysis revealed that, depending on the content of charged residues, nucleoporins (Nups) containing large intrinsically disordered domains with multiple Phe-Gly repeats (FG domains) adopt distinct categories of intrinsically disordered structures. Here, Nups with low charge contents possessed globular configurations, whereas highly charged Nups adopted more dynamic, extended coil conformations [60]. Figure 2A illustrates these observations and shows that the FG domains can indeed be segregated into two distinct categories based on their amino acid (AA) compositions and their measured Stokes radii [60].

In addition to being decisive for the collapsed/extended status of an IDP, charged residues are crucial for IDP function. This statement is illustrated by two examples below. The intrinsically disordered cyclin-dependent kinase (CDK) inhibitor Sic1 contains 9 suboptimal phosphorylated binding sites, known as Cdc4 phosphodegrons (CPDs), that are able to interacts with a single site on its receptor, the F-box protein Cdc4 [43, 45]. High-affinity binding of Sic1 to Cdc4 requires at least six Sic1 sites to be phosphorylated [41]. This binding is mediated via the formation of a highly dynamic complex where multiple CPDs of pSic1 rapidly exchange on and off of the Cdc4-binding surface, with each individual CPD interacting transiently with the core binding pocket of Cdc4 [43, 45]. Such interaction mode enables exchange among different sites and insures high affinity polyvalent interaction, since CPDs not directly bound in the core binding site at any given moment still can contribute to the binding energy via a second binding site or via long-range electrostatic interactions [45]. Based on the analysis of a mean-field statistical mechanical model for the electrostatic interactions between a single receptor site and a conformationally disordered polyvalent ligand, where phosphorylation sites were considered as negative contributions to the total charge of the ligand, it has been concluded that ultrasensitivity in the Sic1-Cdc4 system may be driven by cumulative electrostatic interactions, and that polyelectrostatic interactions, in

general, may provide a powerful mechanism for modulation of protein interactions by multiple phosphorylation sites in disordered protein regions [43]. Figure 2B represents an oversimplified model of such highly dynamic polyelectrostatic interaction

Another illustrative example of the role of charged residues in functionality of IDPs is DNA search by DNA-binding proteins via the intersegment transfer of DNA-binding proteins between different DNA regions. This interaction mode represents the so-called brachiation mechanisms, where "the protein jumps between two DNA molecules through an intermediate in which the recognition helix of the protein is adsorbed to one DNA fragment while the disordered N-tail is adsorbed to the other, using a motion that resembles that of a child brachiating along monkey-bars" [61]. Based on the exhaustive computational analysis it was shown that highly positively charged tails of DNA-binding proteins have higher numbers of intersegment transfers, and the brachiation efficiency is further increased when the charges were clustered and concentrated in the middle of the tail or closer to its N terminus [62]. Furthermore, the protein-DNA interactions can be further modulated by strategic positioning of a few negatively charged residues within the intrinsically disordered DNA-binding tails [62].

*2.4. **D**: D^2 concept (disorder in diseases)*

Due to their important biological functions (see Section 2.6.) and central positions in protein interaction networks (see Section 2.2), many IDPs are implicated in various human disease [63, 64]. This conclusion is based on analysis of the pathological roles of several individual IDPs and extensive computational/bioinformatics studies addressing the abundance of IDPs in various pathological conditions. By the assembly of specific datasets of proteins associated with a given disease and the subsequent computational analysis of these datasets using a number of disorder predictors, the prevalence of intrinsic disorder was shown in proteins associated with cancer, cardiovascular disease, neurodegenerative diseases, amyloidoses, and several other maladies as well as in proteins from pathogenic microbes and viruses [63, 65-73]. An analysis of the diseasome, a network that systematically links the human disease phenome with the human disease genome, uncovered an unfoldome (see Section 2.21) associated with human genetic diseases and revealed that intrinsic disorder is common in proteins associated with many human genetic diseases [74]. The evaluation of the association between a particular protein function (including the disease-specific functional keywords) with the level of intrinsic disorder in a set of proteins known to carry out this function showed that many diseases show strong correlations with proteins predicted to be disordered [75-77]. Based on these and many other observations it has been concluded that intrinsic disorder is highly abundant among proteins associated with various human diseases likely due to the importance of signaling to wide range of diseases. All this gave rise to the "disorder in disorders" or D^2 concept, according to which protein intrinsic disorder is a common player in various human disorders [63]. Some of the most recent examples illustrating the D^2 concept include a large set of oncogenic chimeric proteins [78], such as the EWS-FLI1 fusion protein in Ewing's sarcoma [79]; a highly abundant chromatin architectural protein, methyl-

CpG binding protein 2 (MeCP2), mutations in which are associated with developmental abnormalities, such as Rett syndrome [80, 81]; myelin basic protein (MBP), dysregulations of which are associated with the human autoimmune demyelinating disease multiple sclerosis [82], tau protein whose misfolding and aggregation are related to Alzheimer's disease and other taupathies [83], a highly conserved tissue-specific transcriptional regulator, Limb-bud and heart (LBH) protein, aberrant gain-of-function of which is associated with partial trisomy 2p syndrome [84], and many others.

*2.5. **E**: Ensembles of structures (protein clouds)*

IDPs and intrinsically disordered regions (IDRs) exist as dynamic ensembles, resembling "protein clouds" [19], in which the atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values. Typically, these positions and angles undergo non-cooperative conformational changes. Although these protein clouds are highly dynamic, often their structures can be described rather well by a fairly limited number of lower-energy conformations [85, 86]. As a result, both extended regions (random coil-like) as well as partially collapsed regions with residual secondary structure (pre-molten globule-like), domains with noticeable secondary structure and compaction (partially folded or molten globule-like), and domains with poorly packed side chains are included in the current view of intrinsic disorder [9, 22-24, 87].

*2.6. **F**: Functional anthology of IDPs*

Even rather superficial analysis of the first set of 91 "natively unfolded" proteins revealed that functionally they were very different from ordered proteins. In fact, this set almost did not contain enzymes and many IDPs were involved in nucleic acid binding, metal ion binding, heme binding and interaction with other proteins and membrane bilayers [9]. Later, a focused study identified more than 150 proteins either as containing functional disordered regions, or being completely disordered, yet performing vital cellular functions [87-89]. Based their mode of action, IDPs were grouped into six broad functional classes [assemblers, chaperones, display sites, effectors, entropic chains, and scavengers [12, 90, 91]], and twenty-eight separate functions were assigned for them, including molecular recognition via binding to other proteins, or to nucleic acids [88, 89].

The normal physiology and function of any organism are based on a set of highly coordinated protein interactions. This coordinated action is controlled by the recognition of unique identification regions, which are frequently located inside the IDRs [31, 32]. Therefore, many IDPs are commonly involved in regulation, recognition, signaling and control pathways where high-specificity/low-affinity interactions with multiple partners are necessary prerequisites. The functional diversity provided by disordered regions complements those of ordered protein regions [22, 75-77]. This is illustrated by Figure 3, which represents the extended sequence-to-structure-to-function scheme that includes functions of ordered and disordered proteins [22].

It is important to remember, however, that although biological actions of IDPs frequently involve functional disorder-to-order transitions, some important activities ascribed to IDPs depend on the flexibility, pliability and plasticity of the backbone and do not directly involve coupled binding and folding. These are so-called "entropic chain activities", since they rely entirely on the constant motion of an extended random-coil-like polypeptides [13] (see Figure 3).

*2.7. **G**: Grabbers, huggers, pullers, "binding staccato", and other interaction modes*

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, intrinsically disordered proteins 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 [20]. 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: molecular recognition features (MoRFs), wrappers, chameleons, penetrators, huggers, intertwined strings, long cylindrical containers, connectors, armature, tweezers and forceps, grabbers, tentacles, pullers, and stackers or β-arcs [20]. In addition to the static complexes, where bound partners have fixed structures, some intrinsically disordered proteins do not fold even in their bound state, forming so-called disordered, dynamic, or fuzzy complexes with ordered proteins [43, 45, 46, 48, 49, 92], other disordered proteins [44, 93, 94], or biological membranes [95, 96]. Here, the complex is likely to be formed via 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. In other words, in this staccato-type interaction mode, an disordered protein rapidly changes multiple binding sites while probing binding site(s) of its partner [20]. Figure 4 summarizes information about peculiarities of the IDP interactions with their binding partners by showing the portrait gallery of disorder-based protein complexes [20].

*2.8. **H**: Hydration*

Structure and function of any protein cannot be understood without taking into account its environment. In fact, protein interaction with environment drives all biological processes. Although many modulating factors (such as substrates, ions, ligands, cofactors, allosteric regulators, modifying enzymes, targeting proteins, etc.) are rather specific for a given protein, one environmental component, water, is inevitably present in any biological system, since life started in water, and water is still a major component of the majority of living beings. Therefore, elucidation of the intricate network of protein-water interactions and the peculiarities of a protein

molecule hydration is crucial for understanding of protein structure and function [97]. It is expected that due to the lack of a folded, compact structures and largely exposed interaction surfaces, the hydration of IDPs is significantly different from that of ordered, globular proteins. In agreement with this hypothesis, recent studies clearly showed that hydration of IDPs is significantly higher than that of globular proteins of similar size [97-99]. Furthermore, fully and partially disordered IDPs were shown to possess different degree of hydration [98, 99]. IDPs are also able to bind large amount of charged solute ions [97]. Nature uses these intrinsic propensities of IDRs and IDPs for having high hydration capacities and high capacities for absorption of charged solute ions in dehydrins and other dehydration-related proteins that provide protection to plants and free-living insects against deleterious consequences of water loss and freezing under dehydration stress conditions. It was proposed that the dehydration stress function of these proteins probably results from their simultaneous action of retaining water in the drying cells and preventing an adverse increase in ionic strength, thus countering deleterious effects such as protein denaturation [97]. IDRs found in the enzyme nudix hydrolase from the desiccation- and radiation-resistant bacterium *D. radiodurans* were shown to enlarge the overall surface of the enzyme, and increase its overall affinity for water, therefore increasing the chances of the protein to be located in the remaining water patches in the desiccated cell, where it is protected from the desiccation effects and can function normally [100].

*2.9. **I**: Induced folding*

The majority of intrinsically disordered proteins undergo a disorder-to-order transition upon interaction with specific binding partners [10, 11, 13, 31, 32, 65, 87-89, 101-106]. 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 [13, 102]. High specificity coupled with low affinity seems to be a useful pair of properties for a signaling interaction ensuring their reversibility. In addition, a disordered protein can readily bind to multiple partners by changing shape to associate with different targets [13, 107, 108]. Furthermore, IDPs were shown to contain a “conformational preference” for the structures they will adopt upon binding [109]. This research supported previous findings for proteins involved in individual protein-protein interactions, such as $p27^{Kip1}$ [110, 111] and p53 [112], which were shown to have disordered regions with significant residual helical structure that form α-helices upon binding to their partners. Since foldable-upon-binding IDRs possess specific amino acid features, several computational tools were developed for their prediction [113-115].

*2.10. **J**: Janus chaperones and other moonlighting IDPs*

Several dozen of eukaryotic proteins were shown to fulfill more than one, apparently unrelated, function [116-118]. These multitasking or moonlighting proteins obviously constitute a logical

challenge to the classical lock-and-key mechanism of protein action. In fact, many of the moonlighting proteins in addition to catalytic function possess regulatory, signaling, structural, recognition or chaperone functions [119]. Functional switch is mediated by multiple factors such as differential expression, change in the localization (outside rather than inside a cell, in the different cell type, different cellular compartments), or changes in the oligomeric state, complexation, ligand binding or posttranslational modifications [116-118].

Moonlighting IDPs were shown to have unprecedented multitasking properties, being able to possess opposing (inhibiting and activating) action [119]. The intrinsic flexibility determine the unique capabilities of these proteins to utilize the same or overlapping regions to exert distinct effects and to switch function by adopting different conformations upon binding [119]. Janus chaperones (or two-faced chaperones which can assist the folding of both RNA and protein substrates) occupy a standalone position among moonlighting IDPs [120, 121]. Illustrative examples of such Janus chaperones are L15, L16, L18 and L19 ribosomal proteins of the large subunit of *Escherichia coli* ribosome, which in addition to act as RNA chaperones by promoting the formation of the native RNA fold and/or stabilizing native RNA structure, has genuine protein-chaperone activity in preventing protein aggregation and promoting the reactivation of inactive enzymes [120, 121]. A likely molecular means of such Janus chaperone action is the "entropy-transfer" mechanism, according to which the ordering of the originally disordered chaperone is accompanied by a concomitant unfolding of the originally (mis)folded substrate [90].

*2.11. **K**: Kinetic advantage of intrinsic disorder*

The intrinsic lack of structure and function-related disorder-to-order transitions provide IDPs with various functional advantages, such as [11, 13, 21, 22, 26, 122-124]: (i) Decoupled specificity and strength of binding provides mechanistic grounds for the high-specificity-low-affinity interactions; (ii) Increased speed of interaction due to greater capture radius and the ability to spatially search through interaction space; (iii) Increased interaction (surface) area per residue; (iv) The ability for one-to-many and many-to-one interactions; (v) Increased capture radius for a specific binding site in comparison with that of ordered protein with its restricted conformational freedom, (vi) Fast binding kinetics, etc. The mentioned fast binding kinetics and increased capture radius represent the essence of the so-called "fly-casting mechanism" of protein binding, according to which the unfolded polypeptide first binds weakly at a relatively large distance from the actual binding site and then folds as the protein approaches the binding site [125].

Recently, a detailed two-step analysis of the "fly-casting mechanism" was performed [126]. First, the available experimental data on the binding kinetics of ordered proteins and IDPs were compared to show that generally IDPs bind faster than ordered proteins. Second, coarse-grained molecular dynamics simulations of the pKID–KIX complex were performed to show that interactions of IDPs with binding partners were characterized by high on- and high off-rates resulting from the lower binding free-energy barriers [126]. Contrarily to the expectations, the

capture rates of IDPs were shown to be comparable to those of ordered proteins. Based on the results of these analyses, the authors concluded that lower free-energy barriers which provide encounter complexes with a greater probability to evolve into the final bound states, rather than the predicted by the "fly-casting" model grater capture radii of IDPs (whose actual role in the increase in the binding rates is almost completely eliminated by the slower diffusion of extended IDPs in the encounter processes) represent the major kinetic advantage of IDPs [126].

*2.12. **L**: Linear motifs and other recognition mechanisms*

IDPs recognize and interact with binding partners via several mechanisms. In many cases, intrinsic conformational preferences of IDPs/IDRs can be utilized to predict potential sites capable for the disorder-to-order transitions [113-115]. In fact, in IDPs, potential binding regions able to fold into α-helices upon binding to specific partners are frequently seeing as sharp dips flanked by long regions of predicted disorder in the PONDR® VLXT plots [113, 114, 127]. Based on this correlation between the characteristic shapes of the disorder distribution plots and potential binding sites, specialized predictors of helical binding motifs, molecular recognition features (MoRFs), were developed [113, 114]. Intrinsic disorder also opens unique ways for using the palindromic sequences, retro-MoRFs, as potential binding sites, further extending promiscuity of IDPs since the set of retro-MoRF binding partners likely overlap the partner-sets of the originally identified MoRFs [128]. Inherent conformational preferences of some IDPs for their bound conformations, where the elements they use for recognition are partially/transiently pre-formed in the unbound state, is conceptualized as preformed structural element hypothesis [109]. Another IDP-specific binding mechanism utilizes the primary contact site (PCS) concept. Here PCSs are short recognition motifs within the disordered ensemble that are more exposed than other regions, and thereby may serve as the first sites of contact with the partner and therefore are kinetically distinguished in binding [98]. Finally, some recognition elements are often detected as short motifs of discernible conservation, also denoted as "consensus" sequences. The generality of this relation has led to the concept of linear motifs (LMs, also denoted as eukaryotic linear motifs, ELMs and short linear motifs, SLiMs). LMs were suggested to fall into locally disordered regions [129-131].

*2.13. **M**: Many-to-one and one-to-many binding modes*

In protein-protein interaction networks, disorder is utilized via at least two different mechanisms, where one IDP binds to many partners (the one-to-many mechanism), and where many IDPs bind to one (often ordered) partner (the many-to-one mechanism) [31, 32]. Figure 5 shows that p53 represents a great illustration of how intrinsic disorder is used by a hub protein and how disorder

confers amazing structural plasticity to interacting protein allowing it to adopt different conformations when binding to the different partners. In fact, the C-terminal regulation domain of p53, being disordered in its free form, adopts four different structures, a short α-helix, a β-strand, and two differently shaped coils, upon binding to its four different partners [132]. Furthermore, the same amino acids are buried to very different extends in each of the complexes. In essence, these pliable unfolded regions allow IDPs to be "poly-linguistic" and communicate to different partners using different "languages" [132]. In essence, polymorphic C-terminal region of p53 act as an intermolecular chameleon sequence, i.e., a segment of a protein that can adopt different secondary structures depending on the context (for more detail, see plot M in Figure 4). The concept of chameleon sequences was introduced to describe an intriguing phenomenon, where many unrelated ordered proteins have identical subsequences (up to length of eight residues) which adopt different secondary structures in different contexts [133-135]. Such sequences have been called chameleons for their ability to adopt different structures in different environments [134-140]. In application to IDPs, chameleon behavior could be an important feature that enables one disordered region to bind to multiple partners. With different secondary structures and with different side chain participation in the different complexes, it is as if one sequence can be "read" in multiple ways by the various binding partners.

*2.14. **N**: Natural abundance of IDPs*

Since for a very long time the available information on IDPs was very sparse, the abundance of such proteins or proteins with long IDRs was not clear. In fact, based on the time spent on collection the first sets of experimentally characterized IDPs, it seemed unlikely that such proteins would be highly abundant. However, already first applications of disorder predictors to the available then protein databases produced stunning results: more than 15,000 SwissProt proteins were predicted to contain disordered regions of at least 40 consecutive amino acids, with more than 1,000 having especially high disorder scores [141], and, at the genome level, intrinsic disorder appeared to be a common phenomenon, with eukaryotes having a higher percentage of proteins with long disordered regions (27-41%) than archaea (4-24%) or bacteria (2-21%) [142]. More recent bioinformatics studies indicated that about 25 to 30% of eukaryotic proteins are mostly disordered [143], that more than half of eukaryotic proteins have long regions of disorder [142, 143], and that more than 70% of signaling proteins have long disordered regions [65]. In an independent study, long (>30 residue) disordered segments were found to occur in 2.0% of archaean, 4.2% of eubacterial and 33.0% of eukaryotic proteins [144]. This high natural abundance of IDPs clearly suggests that, although intrinsically disordered proteins fail to form fixed 3D-structures under physiological conditions, they likely carry out important biological functions [9, 11-13, 23, 24, 31, 65, 89, 104, 105, 145].

*2.15. **O**: Oligomerization of IDPs: Obligate or two-state oligomers, fuzzy complexes, interacting clouds and coiled-coils*

Different types of proteinaceous oligomers or protein complexes have been defined in the literature [146]. These definitions include the distinction between obligate and non-obligate oligomers, where the former cover complexes of protomers that cannot exist independently, as opposed to non-obligate complexes. Another distinction can be made between permanent and transient complexes. Permanent protein-protein interactions are strong and irreversible, whereas a complex qualifies as transient if it readily undergoes changes in the oligomeric state [147]. Based on the differences in the molecular mechanisms of association, oligomers were divided into two-state and three-state oligomers [148-150]. In two-state complex formation, the unbound monomeric chains do not populate the folded state. In the three-state model, the monomeric chains are independently folded prior the complex formation. Therefore, monomers of oligomers that are formed *via* a two-state mechanism are intrinsically disordered in their unbound form [150]. In other words, two-state oligomers represent an extreme case of the induced folding mechanism where mutual folding and binding induce and stabilize the final structure of the complex.

However, complex formation is not necessary accompanied by complete folding and rigidifying of binding partners. In fact, some IDPs preserve significant disorder even in the bound state, forming so-called fuzzy complexes [151]. Other IDPs remain almost completely disordered in the bound state, and 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 [20].

Coiled-coils are common structural motif in proteins and protein complexes, were up to 7 long α-helices intertwined together similar to strings of a rope [152-154]. Approximately 3-5 % of all amino acids in proteins are involved in coiled-coil formation [155], and intermolecular coiled coils are a frequently used to form and support protein complexes. Regions corresponding to these intermolecular coiled-coils are typically highly disordered in their monomeric states but gain helical structure upon coiled-coil formation. Therefore, despite the fact that these intermolecular coiled-coils are ordered under physiological conditions, they are typical IDRs representing an abundant illustration of the folding upon binding concept [20].

*2.16. **P**: Posttranslational modifications*

Several signaling IDRs and IDPs were found to contain sites of various posttranslational modifications (PTMs), such as phosphorylation, acetylation, fatty acylation, methylation, glycosylation, ubiquitination, and ADP-ribosylation, suggesting that PTMs commonly occur in IDRs [89, 156]. This high prevalence of PTM sites in disordered regions is likely due to the fact

that PTMs can induce significant structural reorganization of an IDP/IDR, dramatically exciding structural consequences of "the decoration" of an ordered protein structure. Computational studies revealed that there is a strong correlation between the prediction of disorder and the occurrence of phosphorylation [157] and ubiquitination [158], and several other types of PTMs, such as digestion sites, acetylation, fatty acid acylation, and methylation, were shown to occur in regions of intrinsic disorder [77, 89, 156]. Therefore, it is tempting to hypothesize that sites of enzymatically catalyzed posttranslational modifications in eukaryotic proteins universally (or at least very commonly) exhibit a preference for intrinsically disordered regions. For all the enzymatically catalyzed PTMs, a modifying enzyme has to bind to and modify sites in a wide variety of proteins. If all the regions flanking these sites are disordered before binding to the modifying enzyme, it is easy to understand how a single enzyme could bind to and modify a wide variety of protein targets.

*2.17. **Q**: Quartet/trinity model*

From everything discussed in this brief review and elsewhere in literature, it is very clear that IDPs are very different from "normal", ordered proteins. These differences are seeing at different levels, starting from protein structures and ending with their functions. Obviously, a 117 years-old lock-and-key model of protein action, where enzyme and substrate have to fit to each other like a lock and key in order to exert a chemical effect on each other [159], and even much younger "induced fit" theory, where the enzyme is partially flexible and the substrate does not simply bind to the active site but it has to bring about changes to the shape of the active site to activate the enzyme and make the reaction possible [160] do not explain the functional peculiarities of IDPs and IDRs, such as binding promiscuity and the bound state polymorphism. In fact, it was recognized that the existence of IDPs calls for a reassessment of the existing protein–structure–function paradigm [11, 23, 87]. To express a new view on the protein functionality and on the native, biologically active state of a protein molecule, "The Protein Trinity" model was proposed 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, and that protein function can arise from any one of the three states or from transitions between them [87]. Since extended IDPs, also known as natively unfolded proteins, were later split into two structurally different subclasses, native coils and native pre-molten globules [24], this "Protein Trinity" model was subsequently expanded to "The Protein Quartet" model to include a fourth state (pre-molten globule), with function arising from four specific conformations (ordered forms, molten globules, pre-molten globules, and coils) and transitions between any two of the states [23].

*2.18. **R**: Regulation in cell (controlled chaos)*

The highly dynamic nature of IDPs is an obvious illustration of the chaos [161]. However, the evolutionary persistence of these highly dynamic proteins, their unique functionality and involvement in all the major cellular processes evidence that this chaos is tightly controlled [161].

This conclusion is supported by the results of a comprehensive analysis of the intricate mechanisms of the intrinsically disordered protein regulation inside the cell [162]. This analysis revealed that the transcriptional rates of mRNAs encoding intrinsically disordered proteins and ordered proteins were comparable. However the intrinsically disordered protein-encoding transcripts were generally less abundant than transcripts encoding ordered proteins due to the increased decay rates of the formers [162]. The existence of tight regulation of the IDP abundance was also established at the protein level. In fact, IDPs were shown to be less abundant than ordered proteins due to the lower rate of protein synthesis and shorter protein half-lives [162]. The abundance and half-life in a cell of IDPs was shown to be further modulated via their PTMs [162, 163]. 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 [162]. Thus, there is an evolutionarily conserved tight control of synthesis and clearance of most IDPs inside the cell. 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 [162].

In agreement with these conclusions, independent studies revealed that there is a global-scale correlation between the predicted fraction of protein disorder and RNA and protein expression in *E. coli* [164], and that the lifetimes and conformations of IDPs and their mRNAs are orchestrated to ensure precision, speed and flexibility in biological control [165].

*2.19. **S**: Scaffold proteins*

Scaffold proteins selectively bring together specific proteins within a signaling pathway and provide selective spatial orientation and temporal coordination to facilitate and promote interactions among interacting proteins [153]. Scaffold proteins contain several domains for protein-protein interaction, and their typical design includes a set of short globular domains (~80 amino acids on average) connected by longer linker regions (~150 residues on average) with crucial binding functions [166]. Often, scaffold proteins interact with most of their partners at the same time, being located at the centers of functional complexes and therefore acting as “party hubs” [34]. Scaffolds can potentially act as modulators of alternative pathways by promoting interactions between their signaling proteins [124]. Many signaling scaffold proteins are highly disordered and utilize the various features of highly flexible IDRs to obtain more functionality from less structure [124].

*2.20. **T**: Targets for drugs*

It is clear that IDPs and IDRs should be considered now as very promising targets for therapeutic drugs since targeting small molecules to the IDRs of proteins should enable the development of more effective drug discovery techniques. The attractiveness of IDPs as novel drug targets is based

on their high natural prevalence, enormous biological importance and frequent involvement into pathogenesis of various diseases. The difficulties come from several directions including the inability to use traditional "rational drug design" approaches, which are based on an overly rigid view of protein function and therefore utilize either known or modeled 3D structure of a target protein. The binding promiscuity of IDPs and the intrinsically disordered nature of their binding sites, which often resemble "protein clouds", make the traditional structure-based approaches for funding potential drugs highly ineffective or even completely infeasible. Therefore, new strategies are needed for using IDRs as novel drug targets [167]. Currently, there are at least two disorder-based approaches for drug discovery. One of these approaches, blocking of the IDR binding to ordered partners, represents a disorder-based rational drug design, which although benefits from information on the 3D structure of the ordered partner, but utilizes knowledge of intrinsic disorder with respect to the specific IDP-partner interaction that is being targeted. In this approach, 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 [167]. The second approach is a direct targeting of IDRs by small molecules, where small molecules act through forcing the IDR to form an ordered structure dissimilar to its structure in complex with binding partner, thereby preventing binding [168, 169].

*2.21. **U**: Unfoldome and unfoldomics*

Since IDPs constitute a significant portion of any given proteome, and because IDPs and IDRs have amazing structural variability and possess a very wide variety of functions the concepts of the unfoldome and unfoldomics were recently introduced [64, 170-174]. The unfoldome and unfoldomics concepts are built on the ideas that the suffixes –ome and –omics imply a new layer of knowledge, especially when one is dealing with the large masses of data, including results of the high throughput experiments and the computational/bioinformatics analyses of the large datasets [174]. Unfoldome is attributed to a portion of proteome which includes a set of IDPs (also known as natively unfolded proteins, therefore unfoldome) and proteins with IDRs. Unfoldomics focuses on the unfoldome and considers not only the identities of the set of proteins and protein regions in the unfoldome of a given organism, but also their functions, structures, interactions, evolution, etc. [174].

*2.22. **V**: Vulnerable proteins*

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 denaturation of the soluble structure under physiological conditions [175, 176]. Since soluble protein structures may be more or less vulnerable to water attack depending on their packing quality, a structural attribute, protein vulnerability, was introduced as the ratio of solvent-

exposed backbone hydrogen bonds (which represent local weaknesses of the structure) to the overall number of hydrogen bonds [14]. It has 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 [175]. 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 [14]. Finally, by interrelating vulnerability, disorder propensity and co-expression patterns, the role of protein intrinsic disorder in transcriptome organization was confirmed, since the correlation between the extent of intrinsic disorder of the most disordered domain in an interacting pair and the expression correlation of the two genes encoding the respective interacting domains was evident [14].

*2.23. **W**: Where you almost never see disorder*

Although intrinsic disorder is an abundant phenomenon and IDRs were found in many proteins, there are some functional categories which almost completely devoid disorder. Recently, a comprehensive bioinformatics study on a correlation between the functional annotations in the Swiss-Prot database and the predicted intrinsic disorder revealed that that out of 710 Swiss-Prot keywords, 310 functional keywords were associated with ordered proteins, 238 functional keywords were attributed to disordered proteins, and the remainder 162 keywords yield ambiguity in the likely function-structure associations [75-77]. An examination of order-correlated keywords suggested the presence of 5 major functional categories containing mostly ordered proteins [75]: catalysis (oxidoreductase, transferase, hydrolase, lyase, glycosidase, kinase, isomerase, ligase, decarboxylase, glycosyltransferase, protease, acyltransferase, monooxygenase, aminotransferase, metalloprotease, methyltransferase, aminoacyl-tRNA synthetase, aminopeptidase, and dioxygenase); transport (electron transport, sugar transport, etc.), biosynthesis (amino-acid biosynthesis, GMP biosynthesis, gluconeogenesis, amino-acid biosynthesis, pyrimidine biosynthesis, peptidoglycan synthesis, lipopolysaccharide biosynthesis, aromatic amino acid biosynthesis, branched-chain amino acid biosynthesis, purine biosynthesis, lipid a biosynthesis, and lipid synthesis); metabolism (carbohydrate metabolism, tricarboxylic acid cycle, aromatic hydrocarbons catabolism, and one-carbon metabolism); and trans-membrane proteins (porins). Note that proteins with catalytic, biosynthetic and metabolic functions are mostly enzymes, and therefore they are “normal” ordered proteins which are structured for catalysis. The proteins associated with transport are almost all membrane proteins and are necessarily structured so that their backbone hydrogen bonds are formed in the low dielectric environment of the membrane. Proteins from the fifth category, porins, represent transmembrane proteins that are large enough to facilitate passive diffusion [75]. The smaller disorder content and the greater amount of structural

information (*e.g.* a greater PDB coverage) have previously been reported for proteins from the first four functional categories as compared to highly disordered signaling and cancer-associated proteins [65].

*2.24. **X**: X-ray crystallography and intrinsic disorder: When something is missing*

In Protein Data Bank, IDRs are often seeing as regions with missing electron densities, since they fail to scatter X-rays coherently due to variation in position from one protein to the next. As a result, they are missing in electron density maps of X-ray crystallography experiments and are not present in the 3-D structures based on those electron density maps [177]. Although some missing density correspond to wobbly, structured domains rather than to intrinsically disordered ensembles, such wobbly domains are evidently not very common among the long regions of missing electron density [177]. Intriguingly, only ~7% of proteins in PDB completely devoid disorder and only ~25% of PDB entries have >95% of their lengths observed in the corresponding PDB structures [178].

Besides the regions of missing electron density, crystallized proteins often contain regions with high B-factor, which reflects the uncertainty in atom positions in the model and often represents the combined effects of thermal vibrations and static disorder [179]. In order to differentiate between flexible but ordered regions and IDRs, comparisons were made among four categories of protein flexibility: low-B-factor ordered regions, high-B-factor ordered regions, short disordered regions, and long disordered regions (with two last categories being selected as the short and long regions of missing electron density, respectively) [177]. The high-B-factor regions are more similar to IDRs than to ordered regions with low-B-factor. Furthermore, the observed distinctive amino acid biases of high-B-factor ordered regions, short disordered regions, and long disordered regions clearly indicated that the sequence determinants for these flexibility categories differ from one another, suggesting that the amino acid attributes that specify flexibility and intrinsic disorder are distinct and not merely quantitative differences on a continuum [177].

*2.25. **Y**: Y-W-F conundrum*

One of the most characteristic features of IDP and IDR sequences is their low content of aromatic and aliphatic residues. On the other hand, functional sites located within long IDRs do not obey this rule. For example, although amino acid compositions of linear motifs were reported to resemble the characteristic composition of IDPs, they were enriched in certain hydrophobic residues (Trp, Phe, Tyr, and Leu) [131]. Similarly, a systematic analysis of MoRFs revealed that these recognition motifs are typically enriched in Phe and Tyr [180]. Furthermore, aromatic residues are typically located at the MoRF interfaces, suggesting that they are directly involved in MoRF interaction with its binding partners [181]. It has been also pointed out that the variation of NMR dipolar couplings and heteronuclear relaxation rates in IDPs (α-synuclein and tau protein)

closely followed the variations of the bulkiness of amino acids along the polypeptide chain. Therefore, the bulkiness of amino acids was proposed to define the local deviations from the random coil behavior, determining local conformations and dynamics of IDPs providing insight into residual secondary structure and long-range transient interactions in IDPs [182]. Thus, the Y-W-F conundrum can be solved by taking into account the mentioned above obervations: in IDPs and IDRs, aromatic residues are sparse, but they are strategically positioned within the sequence to ensure proper functionality and local structure.

*2.26. **Z**: Zinc and other metal ions as IDP structure modulators*

Metal ions are among most common modifiers of IDP structure. For example, a small set of metal binding proteins, including ostecalcin, osteonectin, α-casein, HPV16 E7 protein, calsequestrin, manganese stabilizing protein, and HIV-1 integrase, was used to calculate the effect of ion binding on the mean net charge of proteins in order to evaluate the hypothesis that interaction of a natively unfolded protein with natural ligand that affects protein's mean net charge, mean hydrophobicity, or both, may change these parameters in such a way that they will approach those typical of ordered proteins [9]. This analysis revealed that, as expected, the interaction of at least these seven proteins with their natural ligands resulted in a shift in their parameters to those characteristics of ordered proteins [9]. These observations were not too surprising, as literature contains numerous facts clearly showing that binding of various metal ions can induce dramatic structural changes in IDPs.

3. Concluding remarks: Bright future of fuzzy protein clouds

Last 10-15 years witnessed a real revolution in our understanding of the protein structure-function relationships. The fact that there is an entire class of polypeptides 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 as an attempt to understand what is so special about several natively unfolded proteins produced a real explosion of interest in non-folded proteins with biological functions. A new field was created and a lot of intriguing information was produced related to the IDPs in general. IDPs are characterized by several unique amino acid features, such as low overall hydrophobicity and high net charge. They do not have rigid structure, being very flexible. An intriguing property of IDPs is their ability to undergo disorder-to-order transition upon function. The degree of these structural rearrangements varies over a very wide range. Multiple roles of IDPs in pathogenesis of human diseases should not be ignored too, as well as their intriguing and very promising roles as new and almost completely unexploited drug targets. Although more than a decade of intensive studies on IDPs revealed a number of unique features related to their structural properties, abundance, distribution, functional repertoire, regulation, involvement into the disease pathogenesis, etc., the mass of data produced so far is just a small tip of a humongous iceberg. Intrinsically disordered proteins continue to bring discoveries on a regular basis. More discoveries and breakthroughs are

expected in future due to the elaboration of novel experimental and computational tools for focused studies on the intrinsically disordered proteins. Modern protein science is at the turning point.

References

[1] H. J. Vogel, The Merck Frosst Award Lecture 1994. Calmodulin: a versatile calcium mediator protein, Biochem Cell Biol, 72 (1994) 357-376.

[2] J. Lee, D. J. O'Kane, A. J. Visser, Spectral properties and function of two lumazine proteins from Photobacterium, Biochemistry, 24 (1985) 1476-1483.

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

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

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

[6] K. Linderstrom-Lang, J. A. Schellman, Protein structure and enzyme activity, in: P. D. Boyer, H. Lardy, K. Myrback (Ed.)^(Eds.), The Enzymes, Academic Press, New York, 1959, pp. 443-510.

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

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

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

[10] G. W. Daughdrill, G. J. Pielak, V. N. Uversky, M. S. Cortese, A. K. Dunker, Natively disordered proteins, in: J. Buchner, T. Kiefhaber (Ed.)^(Eds.), Handbook of Protein Folding, Wiley-VCH, Verlag GmbH & Co. KGaA, Weinheim, Germany, 2005, pp. 271-353.

[11] P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, J Mol Biol, 293 (1999) 321-331.

[12] P. Tompa, Intrinsically unstructured proteins, Trends Biochem Sci, 27 (2002) 527-533.

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

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

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

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

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

[18] D. R. Livesay, Protein dynamics: dancing on an ever-changing free energy stage, Curr Opin Pharmacol, 10 (2010) 706-708.
[19] A. K. Dunker, V. N. Uversky, Drugs for 'protein clouds': targeting intrinsically disordered transcription factors, Curr Opin Pharmacol, 10 (2010) 782-788.
[20] V. N. Uversky, Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, Chem Soc Rev, 40 (2011) 1623-1634.
[21] A. K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, Z. Obradovic, C. Kissinger, J. E. Villafranca, Protein disorder and the evolution of molecular recognition: theory, predictions and observations, Pac Symp Biocomput, (1998) 473-484.
[22] V. N. Uversky, A. K. Dunker, Understanding protein non-folding, Biochim Biophys Acta, 1804 (2010) 1231-1264.
[23] V. N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci, 11 (2002) 739-756.
[24] V. N. Uversky, What does it mean to be natively unfolded?, Eur J Biochem, 269 (2002) 2-12.
[25] R. M. Williams, Z. Obradovi, V. Mathura, W. Braun, E. C. Garner, J. Young, S. Takayama, C. J. Brown, A. K. Dunker, The protein non-folding problem: amino acid determinants of intrinsic order and disorder, Pac Symp Biocomput, (2001) 89-100.
[26] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, A. K. Dunker, Sequence complexity of disordered protein, Proteins, 42 (2001) 38-48.
[27] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, A. K. Dunker, Intrinsic disorder and functional proteomics, Biophys J, 92 (2007) 1439-1456.
[28] V. Vacic, V. N. Uversky, A. K. Dunker, S. Lonardi, Composition Profiler: a tool for discovery and visualization of amino acid composition differences, BMC Bioinformatics, 8 (2007) 211.
[29] E. Garner, P. Cannon, P. Romero, Z. Obradovic, A. K. Dunker, Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization, Genome Inform Ser Workshop Genome Inform, 9 (1998) 201-213.
[30] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, A. K. Dunker, Predicting intrinsic disorder in proteins: an overview, Cell Res, 19 (2009) 929-949.
[31] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, V. N. Uversky, Flexible nets: The roles of intrinsic disorder in protein interaction networks, FEBS Journal, 272 (2005) 5129-5148.
[32] V. N. Uversky, C. J. Oldfield, A. K. Dunker, Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, J Mol Recognit, 18 (2005) 343-384.
[33] A. Patil, H. Nakamura, Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks, FEBS Lett, 580 (2006) 2041-2045.
[34] D. Ekman, S. Light, A. K. Bjorklund, A. Elofsson, What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae?, Genome Biol, 7 (2006) R45.
[35] C. Haynes, C. J. Oldfield, F. Ji, N. Klitgord, M. E. Cusick, P. Radivojac, V. N. Uversky, M. Vidal, L. M. Iakoucheva, Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes, PLoS Comput Biol, 2 (2006) e100.
[36] Z. Dosztanyi, J. Chen, A. K. Dunker, I. Simon, P. Tompa, Disorder and sequence repeats in hub proteins and their implications for network evolution, J Proteome Res, 5 (2006) 2985-2995.

[37] G. P. Singh, M. Ganapathi, D. Dash, Role of intrinsic disorder in transient interactions of hub proteins, Proteins, 66 (2007) 761-765.
[38] G. P. Singh, D. Dash, Intrinsic disorder in yeast transcriptional regulatory network, Proteins, 68 (2007) 602-605.
[39] H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, Lethality and centrality in protein networks, Nature, 411 (2001) 41-42.
[40] A. Patil, K. Kinoshita, H. Nakamura, Hub promiscuity in protein-protein interaction networks, Int J Mol Sci, 11 (2010) 1930-1943.
[41] P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, M. Tyers, Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication, Nature, 414 (2001) 514-521.
[42] J. M. Baker, R. P. Hudson, V. Kanelis, W. Y. Choy, P. H. Thibodeau, P. J. Thomas, J. D. Forman-Kay, CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices, Nat Struct Mol Biol, 14 (2007) 738-745.
[43] M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay, H. S. Chan, Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity, Proc Natl Acad Sci U S A, 104 (2007) 9650-9655.
[44] A. B. Sigalov, A. V. Zhuravleva, V. Y. Orekhov, Binding of intrinsically disordered proteins is not necessarily accompanied by a structural transition to a folded form, Biochimie, 89 (2007) 419-421.
[45] T. Mittag, S. Orlicky, W. Y. Choy, X. Tang, H. Lin, F. Sicheri, L. E. Kay, M. Tyers, J. D. Forman-Kay, Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, Proc Natl Acad Sci U S A, 105 (2008) 17772-17777.
[46] A. B. Sigalov, SCHOOL model and new targeting strategies, Adv Exp Med Biol, 640 (2008) 268-311.
[47] A. B. Sigalov, Signaling chain homooligomerization (SCHOOL) model, Adv Exp Med Biol, 640 (2008) 121-163.
[48] T. Mittag, L. E. Kay, J. D. Forman-Kay, Protein dynamics and conformational disorder in molecular recognition, J Mol Recognit, 23 (2010) 105-116.
[49] T. Mittag, J. Marsh, A. Grishaev, S. Orlicky, H. Lin, F. Sicheri, M. Tyers, J. D. Forman-Kay, Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase, Structure, 18 (2010) 494-506.
[50] A. B. Sigalov, Protein intrinsic disorder and oligomericity in cell signaling, Mol Biosyst, 6 (2010) 451-461.
[51] A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, R. V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins, Proc Natl Acad Sci U S A, 107 (2010) 8183-8188.
[52] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, R. V. Pappu, Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions, Proc Natl Acad Sci U S A, 103 (2006) 16764-16769.
[53] X. Wang, A. Vitalis, M. A. Wyczalkowski, R. V. Pappu, Characterizing the conformational ensemble of monomeric polyglutamine, Proteins, 63 (2006) 297-311.
[54] A. Moglich, K. Joder, T. Kiefhaber, End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation, Proc Natl Acad Sci U S A, 103 (2006) 12394-12399.
[55] A. Vitalis, X. Wang, R. V. Pappu, Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories, Biophys J, 93 (2007) 1923-1937.

[56] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures, Proc Natl Acad Sci U S A, 104 (2007) 2649-2654.
[57] L. Dougan, J. Li, C. L. Badilla, B. J. Berne, J. M. Fernandez, Single homopolypeptide chains collapse into mechanically rigid conformations, Proc Natl Acad Sci U S A, 106 (2009) 12605-12610.
[58] R. H. Walters, R. M. Murphy, Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation, J Mol Biol, 393 (2009) 978-992.
[59] H. S. Ashbaugh, H. W. Hatch, Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space, J Am Chem Soc, 130 (2008) 9536-9542.
[60] J. Yamada, J. L. Phillips, S. Patel, G. Goldfien, A. Calestagne-Morelli, H. Huang, R. Reza, J. Acheson, V. V. Krishnan, S. Newsam, A. Gopinathan, E. Y. Lau, M. E. Colvin, V. N. Uversky, M. F. Rexach, A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins, Mol Cell Proteomics, 9 (2010) 2205-2224.
[61] D. Vuzman, A. Azia, Y. Levy, Searching DNA via a "Monkey Bar" mechanism: the significance of disordered tails, J Mol Biol, 396 (2010) 674-684.
[62] D. Vuzman, Y. Levy, DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail, Proc Natl Acad Sci U S A, 107 (2010) 21004-21009.
[63] V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annu Rev Biophys, 37 (2008) 215-246.
[64] U. Midic, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, Protein disorder in the human diseasome: unfoldomics of human genetic diseases, BMC Genomics, 10 Suppl 1 (2009) S12.
[65] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic, A. K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, J Mol Biol, 323 (2002) 573-584.
[66] V. N. Uversky, A. Roman, C. J. Oldfield, A. K. Dunker, Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs, J Proteome Res, 5 (2006) 1829-1842.
[67] Y. Cheng, T. LeGall, C. J. Oldfield, A. K. Dunker, V. N. Uversky, Abundance of intrinsic disorder in protein associated with cardiovascular disease, Biochemistry, 45 (2006) 10448-10460.
[68] V. N. Uversky, Intrinsic disorder in proteins associated with neurodegenerative diseases, Front Biosci, 14 (2009) 5188-5238.
[69] A. Mohan, W. J. Sullivan, Jr., P. Radivojac, A. K. Dunker, V. N. Uversky, Intrinsic disorder in pathogenic and non-pathogenic microbes: discovering and analyzing the unfoldomes of early-branching eukaryotes, Mol Biosyst, 4 (2008) 328-340.
[70] B. Xue, R. W. Williams, C. J. Oldfield, G. K. Goh, A. K. Dunker, V. N. Uversky, Viral disorder or disordered viruses: do viral proteins possess unique features?, Protein Pept Lett, 17 (2010) 932-951.
[71] G. K. Goh, A. K. Dunker, V. N. Uversky, Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses, Virol J, 6 (2009) 69.
[72] G. K. Goh, A. K. Dunker, V. N. Uversky, A comparative analysis of viral matrix proteins using disorder predictors, Virol J, 5 (2008) 126.
[73] G. K. Goh, A. K. Dunker, V. N. Uversky, Protein intrinsic disorder toolbox for comparative analysis of viral proteins, BMC Genomics, 9 Suppl 2 (2008) S4.

[74] U. Midic, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, Protein disorder in the human diseasome: Unfoldomics of human genetic diseases, PLoS Computational Biology, (2008) In press.
[75] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, V. N. Uversky, Z. Obradovic, Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions, J Proteome Res, 6 (2007) 1882-1898.
[76] S. Vucetic, H. Xie, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions, J Proteome Res, 6 (2007) 1899-1916.
[77] H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins, J Proteome Res, 6 (2007) 1917-1932.
[78] H. Hegyi, L. Buday, P. Tompa, Intrinsic structural disorder confers cellular viability on oncogenic fusion proteins, PLoS Comput Biol, 5 (2009) e1000552.
[79] H. V. Erkizan, V. N. Uversky, J. A. Toretsky, Oncogenic partnerships: EWS-FLI1 protein interactions initiate key pathways of Ewing's sarcoma, Clin Cancer Res, 16 (2010) 4077-4083.
[80] R. P. Ghosh, T. Nikitina, R. A. Horowitz-Scherer, L. M. Gierasch, V. N. Uversky, K. Hite, J. C. Hansen, C. L. Woodcock, Unique physical properties and interactions of the domains of methylated DNA binding protein 2, Biochemistry, 49 (2010) 4395-4410.
[81] C. Yang, M. J. van der Woerd, U. M. Muthurajan, J. C. Hansen, K. Luger, Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes, Nucleic Acids Res, (2011).
[82] L. Homchaudhuri, M. De Avila, S. B. Nilsson, K. Bessonov, G. S. Smith, V. V. Bamm, A. A. Musse, G. Harauz, J. M. Boggs, Secondary structure and solvent accessibility of a calmodulin-binding C-terminal segment of membrane-associated myelin basic protein, Biochemistry, 49 (2010) 8955-8966.
[83] R. L. Narayanan, U. H. Durr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, Automatic assignment of the intrinsically disordered protein Tau with 441-residues, J Am Chem Soc, 132 (2010) 11906-11907.
[84] H. Al-Ali, M. E. Rieger, K. L. Seldeen, T. K. Harris, A. Farooq, K. J. Briegel, Biophysical characterization reveals structural disorder in the developmental transcriptional regulator LBH, Biochem Biophys Res Commun, 391 (2010) 1104-1109.
[85] W. Y. Choy, J. D. Forman-Kay, Calculation of ensembles of structures representing the unfolded state of an SH3 domain, J Mol Biol, 308 (2001) 1011-1032.
[86] A. Huang, C. M. Stultz, The effect of a DeltaK280 mutation on the unfolded state of a microtubule-binding repeat in Tau, PLoS Comput Biol, 4 (2008) e1000155.
[87] A. K. Dunker, Z. Obradovic, The protein trinity--linking function and disorder, Nat Biotechnol, 19 (2001) 805-806.
[88] A. K. Dunker, C. J. Brown, Z. Obradovic, Identification and functions of usefully disordered proteins, Adv Protein Chem, 62 (2002) 25-49.
[89] A. K. Dunker, C. J. Brown, J. D. Lawson, L. M. Iakoucheva, Z. Obradovic, Intrinsic disorder and protein function, Biochemistry, 41 (2002) 6573-6582.
[90] P. Tompa, P. Csermely, The role of structural disorder in the function of RNA and protein chaperones, Faseb J, 18 (2004) 1169-1175.

[91] P. Tompa, The functional benefits of protein disorder, J. Mol. Struct. THEOCHEM, 666-667 (2003) 361-371.
[92] S. E. Permyakov, I. S. Millett, S. Doniach, E. A. Permyakov, V. N. Uversky, Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin, Proteins, 53 (2003) 855-862.
[93] A. Sigalov, D. Aivazian, L. Stern, 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 (2004) 2049-2061.
[94] M. S. Pometun, E. Y. Chekmenev, R. J. Wittebort, Quantitative observation of backbone disorder in native elastin, J Biol Chem, 279 (2004) 7982-7987.
[95] A. B. Sigalov, G. M. Hendricks, Membrane binding mode of intrinsically disordered cytoplasmic domains of T cell receptor signaling subunits depends on lipid composition, Biochem Biophys Res Commun, 389 (2009) 388-393.
[96] A. B. Sigalov, D. A. Aivazian, V. N. Uversky, L. J. Stern, Lipid-binding activity of intrinsically unstructured cytoplasmic domains of multichain immune recognition receptor signaling subunits, Biochemistry, 45 (2006) 15731-15739.
[97] P. Tompa, P. Banki, M. Bokor, P. Kamasa, D. Kovacs, G. Lasanda, K. Tompa, Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects, Biophys J, 91 (2006) 2243-2249.
[98] V. Csizmok, M. Bokor, P. Banki, E. Klement, K. F. Medzihradszky, P. Friedrich, K. Tompa, P. Tompa, Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2, Biochemistry, 44 (2005) 3955-3964.
[99] M. Bokor, V. Csizmok, D. Kovacs, P. Banki, P. Friedrich, P. Tompa, K. Tompa, NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins, Biophys J, 88 (2005) 2030-2037.
[100] O. Awile, A. Krisko, I. F. Sbalzarini, B. Zagrovic, Intrinsically disordered regions may lower the hydration free energy in proteins: a case study of nudix hydrolase in the bacterium Deinococcus radiodurans, PLoS Comput Biol, 6 (2010) e1000854.
[101] R. S. Spolar, M. T. Record, Jr., Coupling of local folding to site-specific binding of proteins to DNA, Science, 263 (1994) 777-784.
[102] G. E. Schulz, Nucleotide Binding Proteins, in: M. Balaban (Ed.)^(Eds.), Molecular Mechanism of Biological Recognition, Elsevier/North-Holland Biomedical Press, New York, 1979, pp. 79-94.
[103] B. W. Pontius, Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association, Trends Biochem Sci, 18 (1993) 181-186.
[104] H. J. Dyson, P. E. Wright, Coupling of folding and binding for unstructured proteins, Curr Opin Struct Biol, 12 (2002) 54-60.
[105] H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions, Nat Rev Mol Cell Biol, 6 (2005) 197-208.
[106] K. W. Plaxco, M. Gross, Cell biology. The importance of being unfolded, Nature, 386 (1997) 657, 659.
[107] F. Karush, Heterogeneity of the binding sites of bovine serum albumin, J. Am. Chem. Soc., 72 (1950) 2705-2713.
[108] R. W. Kriwacki, L. Hengst, L. Tennant, S. I. Reed, P. E. Wright, Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity, Proc. Natl. Acad. Sci. U. S. A., 93 (1996) 11504-11509.

[109] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, J Mol Biol, 338 (2004) 1015-1026.
[110] E. A. Bienkiewicz, J. N. Adkins, K. J. Lumb, Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1), Biochemistry, 41 (2002) 752-759.
[111] E. R. Lacy, I. Filippov, W. S. Lewis, S. Otieno, L. Xiao, S. Weiss, L. Hengst, R. W. Kriwacki, p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding, Nat Struct Mol Biol, 11 (2004) 358-364.
[112] H. Lee, K. H. Mok, R. Muhandiram, K. H. Park, J. E. Suk, D. H. Kim, J. Chang, Y. C. Sung, K. Y. Choi, K. H. Han, Local structural elements in the mostly unstructured transcriptional activation domain of human p53, J Biol Chem, 275 (2000) 29426-29432.
[113] C. J. Oldfield, Y. Cheng, M. S. Cortese, P. Romero, V. N. Uversky, A. K. Dunker, Coupled folding and binding with alpha-helix-forming molecular recognition elements, Biochemistry, (2005) In press.
[114] Y. Cheng, C. J. Oldfield, J. Meng, P. Romero, V. N. Uversky, A. K. Dunker, Mining alpha-helix-forming molecular recognition features with cross species sequence alignments, Biochemistry, 46 (2007) 13468-13477.
[115] Z. Dosztanyi, B. Meszaros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics, 25 (2009) 2745-2746.
[116] C. J. Jeffery, Moonlighting proteins, Trends Biochem Sci, 24 (1999) 8-11.
[117] C. J. Jeffery, Multifunctional proteins: examples of gene sharing, Ann Med, 35 (2003) 28-35.
[118] C. J. Jeffery, Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins, Curr Opin Struct Biol, 14 (2004) 663-668.
[119] P. Tompa, C. Szasz, L. Buday, Structural disorder throws new light on moonlighting, Trends Biochem Sci, 30 (2005) 484-489.
[120] D. Kovacs, M. Rakacs, B. Agoston, K. Lenkey, K. Semrad, R. Schroeder, P. Tompa, Janus chaperones: assistance of both RNA- and protein-folding by ribosomal proteins, FEBS Lett, 583 (2009) 88-92.
[121] P. Tompa, D. Kovacs, Intrinsically disordered chaperones in plants and animals, Biochem Cell Biol, 88 (2010) 167-174.
[122] A. K. Dunker, Z. Obradovic, P. Romero, C. Kissinger, E. Villafranca, On the importance of being disordered., PDB Newsletter, 81 (1997) 3-5.
[123] C. J. Brown, S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams, A. K. Dunker, Evolutionary rate heterogeneity in proteins with long disordered regions, J. Mol. Evol., 55 (2002) 104-110.
[124] M. S. Cortese, V. N. Uversky, A. K. Dunker, Intrinsic disorder in scaffold proteins: getting more from less, Prog Biophys Mol Biol, 98 (2008) 85-106.
[125] B. A. Shoemaker, J. J. Portman, P. G. Wolynes, Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, Proc Natl Acad Sci U S A, 97 (2000) 8868-8873.
[126] Y. Huang, Z. Liu, Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism, J Mol Biol, 393 (2009) 1143-1159.
[127] E. Garner, P. Romero, A. K. Dunker, C. Brown, Z. Obradovic, Predicting Binding Regions within Disordered Proteins, Genome Inform Ser Workshop Genome Inform, 10 (1999) 41-50.

[128] B. Xue, A. K. Dunker, V. N. Uversky, Retro-MoRFs: Identifying protein binding sites by normal and reverse alignment and intrinsic disorder prediction, International Journal of Molecular Sciences, 11 (2010) 3725-3747.
[129] R. Linding, R. B. Russell, V. Neduva, T. J. Gibson, GlobPlot: Exploring protein sequences for globularity and disorder, Nucleic Acids Res, 31 (2003) 3701-3708.
[130] P. Puntervoll, R. Linding, C. Gemund, S. Chabanis-Davidson, M. Mattingsdal, S. Cameron, D. M. Martin, G. Ausiello, B. Brannetti, A. Costantini, F. Ferre, V. Maselli, A. Via, G. Cesareni, F. Diella, G. Superti-Furga, L. Wyrwicz, C. Ramu, C. McGuigan, R. Gudavalli, I. Letunic, P. Bork, L. Rychlewski, B. Kuster, M. Helmer-Citterich, W. N. Hunter, R. Aasland, T. J. Gibson, ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins, Nucleic Acids Res, 31 (2003) 3625-3630.
[131] M. Fuxreiter, P. Tompa, I. Simon, Structural disorder imparts plasticity on linear motifs, Bioinformatics, 23 (2007) 950-956.
[132] C. J. Oldfield, J. Meng, J. Y. Yang, M. Q. Yang, V. N. Uversky, A. K. Dunker, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, BMC Genomics, 9 Suppl 1 (2008) S1.
[133] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers, 22 (1983) 2577-2637.
[134] D. L. Minor, Jr., P. S. Kim, Context-dependent secondary structure formation of a designed protein sequence, Nature, 380 (1996) 730-734.
[135] I. Jacoboni, P. L. Martelli, P. Fariselli, M. Compiani, R. Casadio, Predictions of protein segments with the same aminoacid sequence and different secondary structure: a benchmark for predictive methods, Proteins, 41 (2000) 535-544.
[136] M. Mezei, Chameleon sequences in the PDB, Protein Eng, 11 (1998) 411-414.
[137] C. A. Smith, V. Calabro, A. D. Frankel, An RNA-binding chameleon, Mol Cell, 6 (2000) 1067-1076.
[138] S. Yoon, H. Jung, Analysis of chameleon sequences by energy decomposition on a pairwise per-residue basis, Protein J, 25 (2006) 361-368.
[139] J. T. Guo, J. W. Jaromczyk, Y. Xu, Analysis of chameleon sequences and their implications in biological processes, Proteins, 67 (2007) 548-558.
[140] K. Takano, Y. Katagiri, A. Mukaiyama, H. Chon, H. Matsumura, Y. Koga, S. Kanaya, Conformational contagion in a protein: Structural properties of a chameleon sequence, Proteins, (2007).
[141] P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, A. K. Dunker, Thousands of proteins likely to have long disordered regions, Pac Symp Biocomput, (1998) 437-448.
[142] A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, C. J. Brown, Intrinsic protein disorder in complete genomes, Genome Inform Ser Workshop Genome Inform, 11 (2000) 161-171.
[143] C. J. Oldfield, Y. Cheng, M. S. Cortese, C. J. Brown, V. N. Uversky, A. K. Dunker, Comparing and combining predictors of mostly disordered proteins, Biochemistry, 44 (2005) 1989-2000.
[144] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, Prediction and functional analysis of native disorder in proteins from the three kingdoms of life, J Mol Biol, 337 (2004) 635-645.
[145] A. L. Fink, Natively unfolded proteins, Curr Opin Struct Biol, 15 (2005) 35-41.
[146] I. M. Nooren, J. M. Thornton, Diversity of protein-protein interactions, EMBO J, 22 (2003) 3486-3492.

[147] J. R. Perkins, I. Diboun, B. H. Dessailly, J. G. Lees, C. Orengo, Transient protein-protein interactions: structural, functional, and network properties, Structure, 18 (2010) 1233-1243.
[148] C. M. Teschke, J. King, Folding and assembly of oligomeric proteins in Escherichia coli, Curr Opin Biotechnol, 3 (1992) 468-473.
[149] D. Xu, C. J. Tsai, R. Nussinov, Mechanism and evolution of protein dimerization, Protein Sci, 7 (1998) 533-544.
[150] K. Gunasekaran, C. J. Tsai, R. Nussinov, Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers, J Mol Biol, 341 (2004) 1327-1341.
[151] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, Trends Biochem Sci, 33 (2008) 2-8.
[152] J. M. Mason, K. M. Arndt, Coiled coil domains: stability, specificity, and biological implications, Chembiochem, 5 (2004) 170-176.
[153] W. Liu, H. Rui, J. Wang, S. Lin, Y. He, M. Chen, Q. Li, Z. Ye, S. Zhang, S. C. Chan, Y. G. Chen, J. Han, S. C. Lin, Axin is a scaffold protein in TGF-beta signaling that promotes degradation of Smad7 by Arkadia, Embo J, 25 (2006) 1646-1658.
[154] J. Liu, Q. Zheng, Y. Deng, C. S. Cheng, N. R. Kallenbach, M. Lu, A seven-helix coiled coil, Proc Natl Acad Sci U S A, 103 (2006) 15457-15462.
[155] E. Wolf, P. S. Kim, B. Berger, MultiCoil: a program for predicting two- and three-stranded coiled coils, Protein Sci, 6 (1997) 1179-1189.
[156] A. K. Dunker, C. J. Brown, Z. Obradovic, Identification and functions of usefully disordered proteins, Adv. Protein Chem., 62 (2002) 25-49.
[157] L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, A. K. Dunker, The importance of intrinsic disorder for protein phosphorylation, Nucleic Acids Res., 32 (2004) 1037-1049.
[158] P. Radivojac, V. Vacic, C. Haynes, R. R. Cocklin, A. Mohan, J. W. Heyen, M. G. Goebl, L. M. Iakoucheva, Identification, analysis, and prediction of protein ubiquitination sites, Proteins, 78 (2010) 365-380.
[159] E. Fischer, Einfluss der configuration auf die wirkung der enzyme., Ber. Dt. Chem. Ges., 27 (1894) 2985-2993.
[160] D. E. Koshland, Jr., Application of a theory of enzyme specificity to protein synthesis, Proc Natl Acad Sci USA, 44 (1958) 98-104.
[161] V. N. Uversky, A. K. Dunker, Biochemistry. Controlled chaos, Science, 322 (2008) 1340-1341.
[162] J. Gsponer, M. E. Futschik, S. A. Teichmann, M. M. Babu, Tight regulation of unstructured proteins: from transcript synthesis to protein degradation, Science, 322 (2008) 1365-1368.
[163] M. Grimmler, Y. Wang, T. Mund, Z. Cilensek, E. M. Keidel, M. B. Waddell, H. Jakel, M. Kullmann, R. W. Kriwacki, L. Hengst, Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases, Cell, 128 (2007) 269-280.
[164] O. Paliy, S. M. Gargac, Y. Cheng, V. N. Uversky, A. K. Dunker, Protein disorder is positively correlated with gene expression in Escherichia coli, J Proteome Res, 7 (2008) 2234-2245.
[165] B. Ma, R. Nussinov, Regulating highly dynamic unstructured proteins and their coding mRNAs, Genome Biol, 10 (2009) 204.

[166] A. Balazs, V. Csizmok, L. Buday, M. Rakacs, R. Kiss, M. Bokor, R. Udupa, K. Tompa, P. Tompa, High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1, FEBS J, 276 (2009) 3744-3756.
[167] Y. Cheng, T. LeGall, C. J. Oldfield, J. P. Mueller, Y. Y. Van, P. Romero, M. S. Cortese, V. N. Uversky, A. K. Dunker, Rational drug design via intrinsically disordered protein, Trends Biotechnol, 24 (2006) 435-442.
[168] D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc, J Am Chem Soc, 131 (2009) 7390-7401.
[169] S. J. Metallo, Intrinsically disordered proteins are potential drug targets, Curr Opin Chem Biol, 14 (2010) 481-488.
[170] M. S. Cortese, J. P. Baird, V. N. Uversky, A. K. Dunker, Uncovering the unfoldome: enriching cell extracts for unstructured proteins by Acid treatment, J Proteome Res, 4 (2005) 1610-1618.
[171] A. K. Dunker, C. J. Oldfield, J. Meng, P. Romero, J. Y. Yang, J. W. Chen, V. Vacic, Z. Obradovic, V. N. Uversky, The unfoldomics decade: an update on intrinsically disordered proteins, BMC Genomics, 9 Suppl 2 (2008) S1.
[172] U. Midic, C. J. Oldfield, A. K. Dunker, Z. Obradovic, V. N. Uversky, Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome, Protein Pept Lett, 16 (2009) 1533-1547.
[173] V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, S. Vucetic, B. Xue, L. M. Iakoucheva, Z. Obradovic, A. K. Dunker, Unfoldomics of human diseases: Linking protein intrinsic disorder with diseases, BMC Genomics, 10 (2009) S7.
[174] V. N. Uversky, The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome, J Biomed Biotechnol, 2010 (2010) 568068.
[175] A. Fernandez, H. A. Scheraga, Insufficiently dehydrated hydrogen bonds as determinants of protein interactions, Proc Natl Acad Sci U S A, 100 (2003) 113-118.
[176] A. Fernandez, Keeping dry and crossing membranes, Nat Biotechnol, 22 (2004) 1081-1084.
[177] P. Radivojac, Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, A. K. Dunker, Protein flexibility and intrinsic disorder, Protein Sci, 13 (2004) 71-80.
[178] T. Le Gall, P. R. Romero, M. S. Cortese, V. N. Uversky, A. K. Dunker, Intrinsic disorder in the Protein Data Bank, J Biomol Struct Dyn, 24 (2007) 325-342.
[179] G. Rhodes, Crystallography made crystal clear: A guide for users of macromolecular models, Editor (Ed.)^(Eds.), Academic Press, San Diego, CA, 1993,
[180] A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, V. N. Uversky, Analysis of molecular recognition features (MoRFs), J Mol Biol, 362 (2006) 1043-1059.
[181] V. Vacic, C. J. Oldfield, A. Mohan, P. Radivojac, M. S. Cortese, V. N. Uversky, A. K. Dunker, Characterization of molecular recognition features, MoRFs, and their binding partners, J Proteome Res, 6 (2007) 2351-2366.
[182] M. K. Cho, H. Y. Kim, P. Bernado, C. O. Fernandez, M. Blackledge, M. Zweckstetter, Amino acid bulkiness defines the local conformations and dynamics of natively unfolded alpha-synuclein and tau, J Am Chem Soc, 129 (2007) 3032-3033.
[183] A. V. Kajava, U. Baxa, A. C. Steven, Beta arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils, Faseb J, 24 (2010) 1311-1319.

Figure legends

Figure 1. Peculiarities of amino acid composition of IDPs. Comparison of the mean net charge and the mean hydrophobicity for a set of 275 folded (open circles) and 91 natively unfolded proteins (black diamonds). The solid line represents the border between extended IDPs and compact globular proteins. Illustrative examples of extended IDPs (native coils and native pre-molten globules) and some ordered globular proteins are shown.

Figure 2. Role of charges in IDP structure and function.

A. Distinct categories of disordered structures adopted by FG nups. Stokes radii values experimentally measured for the purified FG domains were divided by the values predicted for each nup based on its mass assuming a relaxed-coil configuration (y-axis). The amino acid (AA) composition of each domain was expressed as the % content of charged AAs *(i.e.* DEKR) divided by the % content of hydrophobic AAs (*i.e.* AILFWV) (x-axis). Gray boxes highlight the two main categories of intrinsically-disordered structures. Nup domains with a low or high charge-content are highlighted in blue or red boxes, respectively. The green line highlights the correlation between category of intrinsically-disordered structure and charged-to-hydrophobic AAs ratio. Reproduced from ref. [60].

B. Illustrative example of the polyelectrostatic model describing the dynamic complex between an extended IDP (Sic1, shown by a ribbon) and its ordered partner (Cdc4, shown as a gray shape). Here, interaction involves distinct binding motifs and an ensemble of conformations (indicated by four representations of the interaction). The intrinsically disordered protein possesses positive and negative charges (depicted as blue and red circles, respectively) giving rise to a net charge q_l, while the binding site in the receptor (light blue) has a charge q_r. The effective distance $<r>$ is between the binding site and the centre of mass of the intrinsically disordered protein. Reproduced from ref. [48].

Figure 3. Involvement of intrinsic disorder in protein function. Note that the classical structure-function paradigm cannot describe many of the function proteins perform. Reproduced from ref. [22].

Figure 4. A portrait gallery of disorder-based complexes. Illustrative examples of various interaction modes of intrinsically disordered proteins are shown. **A. MoRFs**. **a**, α-MoRF, a complex between the botulinum neurotoxin (red helix) and its receptor (a blue cloud) (PDB ID: 2NM1); **b**, i-MoRF, a complex between an 18-mer cognate peptide derived from the α1 subunit of the nicotinic acetylcholine receptor from *Torpedo californica* (red helix) and α-cobratoxin (a blue cloud) (PDB ID: 1LXH). **B. Wrappers**. **a**, rat PP1 (blue cloud) complexed with mouse

inhibitor-2 (red helices) (PDB ID: 2O8A); **b**, a complex between the paired domain from the Drosophila paired (prd) protein and DNA (PDB ID: 1PDN). **C. Penetrator**. Ribosomal protein s12 embedded into the rRNA (PDB ID: 1N34). **D. Huggers**. **a**, *E. coli trp* repressor dimer (PDB ID: 1ZT9); **b**, tetramerization domain of p53 (PDB ID: 1PES); **c**, tetramerization domain of p73 (PDB ID: 2WQI). **E. Intertwined strings**. **a**, dimeric coiled-coil, a basic coiled-coil protein from *Eubacterium eligens* ATCC 27750 (PDB ID: 3HNW); **b**, trimeric coiled-coil, salmonella trimeric autotransporter adhesin, SadA (PDB ID: 2WPQ); **c**, tetrameric coiled-coil, the virion-associated protein P3 from Caulimovirus (PDB ID: 2O1J). **F. Long cylindrical containers**. **a**, pentameric coiled coil, side and top views of the assembly domain of cartilage oligomeric matrix protein (PDB ID: 1FBM); **b**, side and top views of the seven-helix coiled-coil, engineered version of the GCN4 leucine zipper (PDB ID: 2HY6). **G. Connectors**. **a**, human heat shock factor binding protein 1 (PDB ID: 3CI9); **b**, the bacterial cell division protein ZapA from *Pseudomonas aeruginosa* (PDB ID: 1W2E). **H. Armature**. **a**, side and top views of the envelope glycoprotein GP2 from Ebola virus (PDB ID: 2EBO); **b**, side and top views of a complex between the N- and C-terminal peptides derived from the membrane fusion protein of the Visna (PDB ID: 1JEK). **I. Tweezers or forceps**. A complex between c-Jun, c-Fos and DNA. Proteins are shown as red helices, whereas DNA is shown as a blue cloud (PDB ID: 1FOS). **J. Grabbers**. Structure of the complex between βPIX coiled coil (red helices) and Shank PDZ (blue cloud) (PDB ID: 3L4F). **K. Tentacles**. Structure of the hexameric molecular chaperone prefoldin from the archaeum *Methanobacterium thermoautotrophicum* (PDB ID: 1FXK). **L. Pullers**. Structure of the ClpB chaperone from *Thermus thermophilus* (PDB ID: 1QVR). **M. Chameleons**. The C-terminal fragment of p53 gains different types of secondary structure in complexes with four different binding partners, cyclinA (PDB ID: 1H26), sirtuin (PDB ID: 1MA3), CBP bromo domain (PDB ID: 1JSP), and s100ββ (PDB ID: 1DT7). **N. Stackers or β-arcs**. **a**, stack of β-arches, β-amyloid; **b**, superpleated β-structure (Sup35p, Ure2P, α-synuclein); **c**, stack of β-solenoids (prion); **d**, stack of β-arch dimers (insulin); **e**, β-solenoids. Modified from ref. [183] **O. Dynamic complexes**. Schematic representation of the polyelectrostatic model of Sic1-Cdc4 interaction [45]. Schematic of an IDP (ribbon) interacting with a folded receptor (gray shape) through several distinct binding motifs and an ensemble of conformations (indicated by four representations of the interaction). The intrinsically disordered protein possesses positive and negative charges (depicted as blue and red circles, respectively) giving rise to a net charge q_l, while the binding site in the receptor (light blue) has a charge q_r. The effective distance $<r>$ is between the binding site and the centre of mass of the intrinsically disordered protein. Reproduced from ref. [48].

This figure is based on the data represented in ref. [20].

Figure 5. p53 interaction with different binding partners illustrate peculiarities of one-to-many signaling. A structure versus disorder prediction on the p53 amino acid sequence is shown in the center of the figure (up = disorder, down = order) along with the structures of various regions of p53 bound to fourteen different partners. The predicted central region of structure with the predicted amino and carbonyl termini as being disordered have been confirmed experimentally for

p53. The various regions of p53 are color coded to show their structures in the complex and to map the binding segments to the amino acid sequence. Starting with the p53-DNA complex (top, left, magenta protein, blue DNA), and moving in a clockwise direction, the Protein Data Bank IDs and partner names are given as follows for the fourteen complexes: (1tsr – DNA), (1gzh – 53BP1), (1q2d – gcn5), (3sak – p53 (tetrametization domain)), (1xqh – set9), (1h26 – cyclinA), (1ma3 – sirtuin), (1jsp – CBP bromo domain), (1dt7 – s100ββ), (2h1l – sv40 Large T antigen), (1ycs – 53BP2), (2gs0 – PH), (1ycr – MDM2), and (2b3g – rpa70). Reproduced from ref. [22].

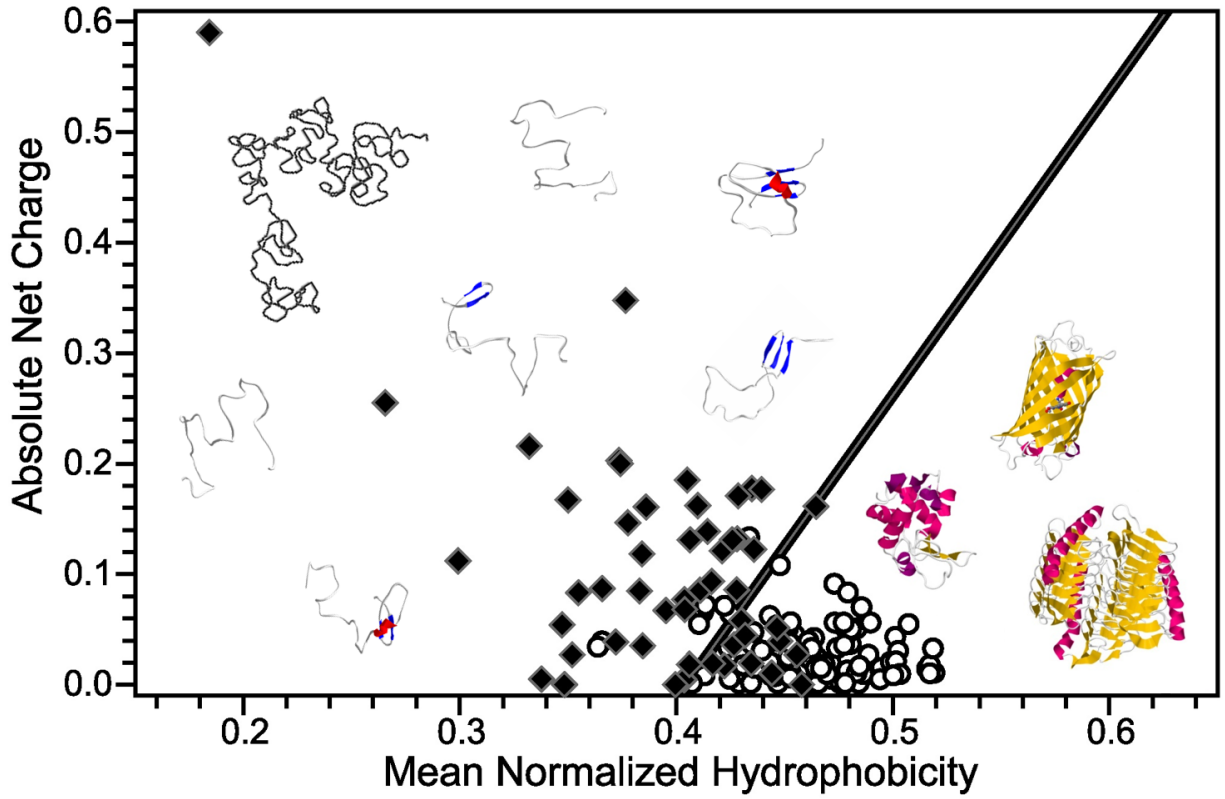
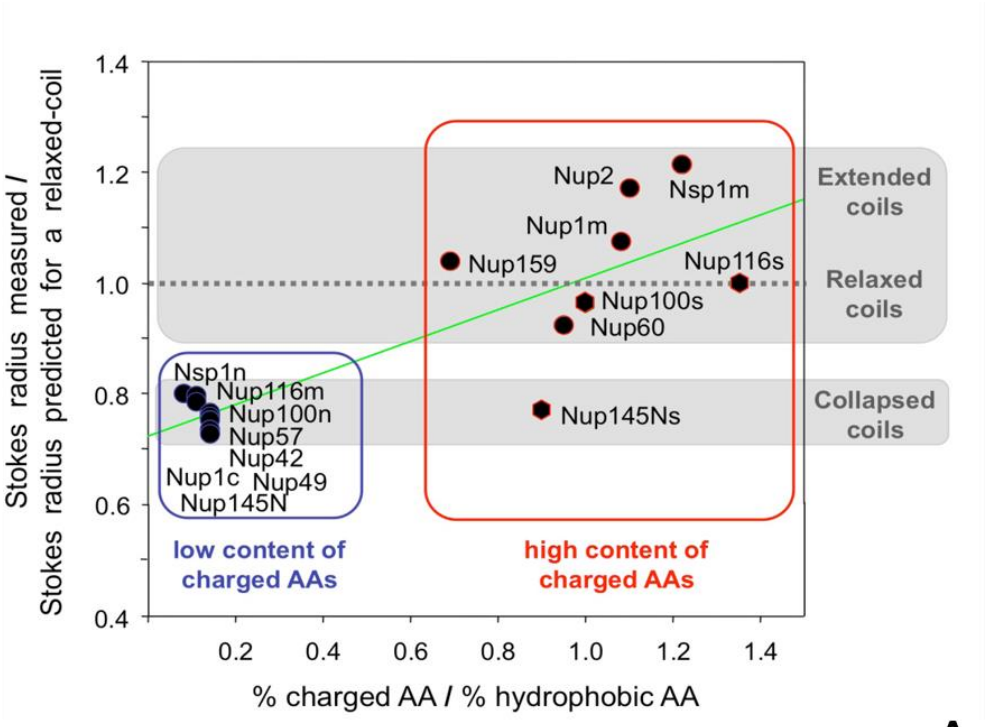
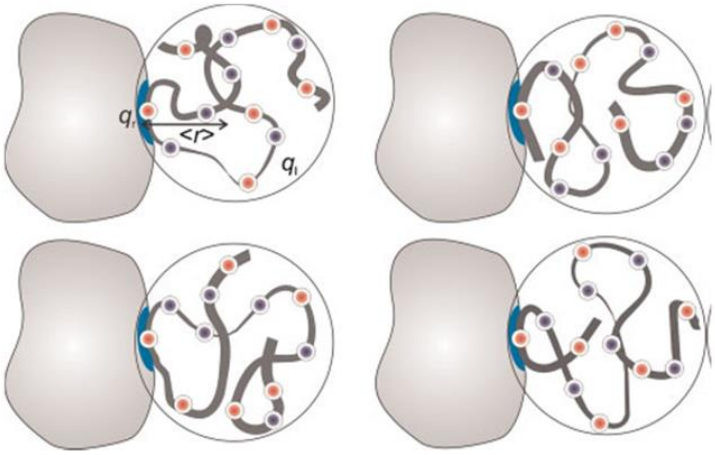

Absolute Net Charge
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.2
0.3
0.4
0.5
0.6
Mean Normalized Hydrophobicity

Figure 1


Stokes radius measured / Stokes radius predicted for a relaxed-coil
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.4
0.6
0.8
1.0
1.2
1.4
% charged AA / % hydrophobic AA
Nup2
Nsp1m
Nup1m
Nup159
Nup116s
Nup100s
Nup60
Nup145Ns
Nsp1n
Nup116m
Nup100n
Nup57
Nup42
Nup1c
Nup49
Nup145N
Extended coils
Relaxed coils
Collapsed coils
low content of charged AAs
high content of charged AAs

A


<r>

B

Figure 2

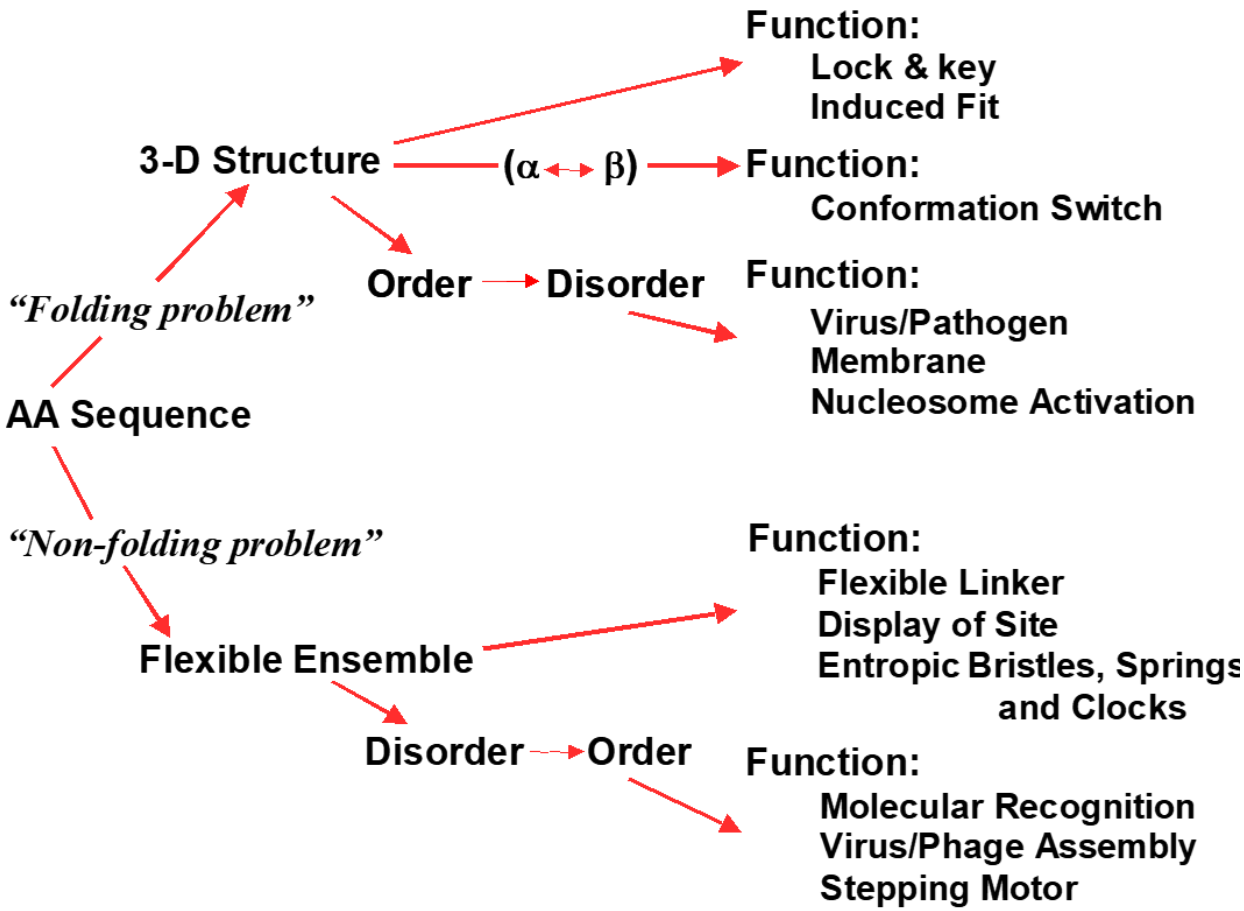

Function:
Lock & key
Induced Fit
3-D Structure
(α ↔ β)
Function:
Conformation Switch
Order → Disorder
Function:
Virus/Pathogen
Membrane
Nucleosome Activation
"Folding problem"
AA Sequence
"Non-folding problem"
Function:
Flexible Linker
Display of Site
Entropic Bristles, Springs
and Clocks
Flexible Ensemble
Disorder → Order
Function:
Molecular Recognition
Virus/Phage Assembly
Stepping Motor

Figure 3

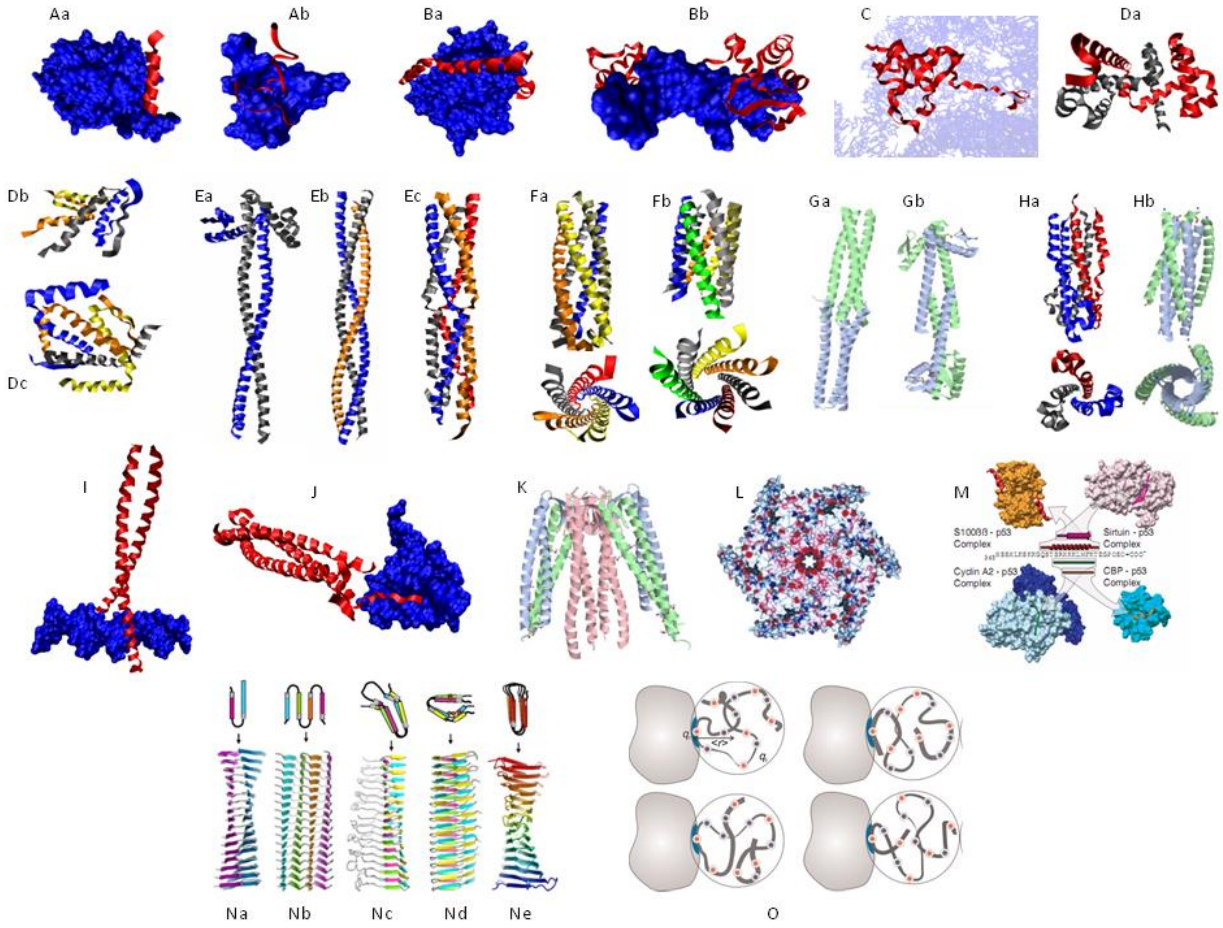

Aa
Ab
Ba
Bb
C
Da
Db
Dc
Ea
Eb
Ec
Fa
Fb
Ga
Gb
Ha
Hb
I
J
K
L
M
S100β - p53 Complex
Sirtuin - p53 Complex
Cyclin A2 - p53 Complex
CBP - p53 Complex
Na
Nb
Nc
Nd
Ne
O

Figure 4

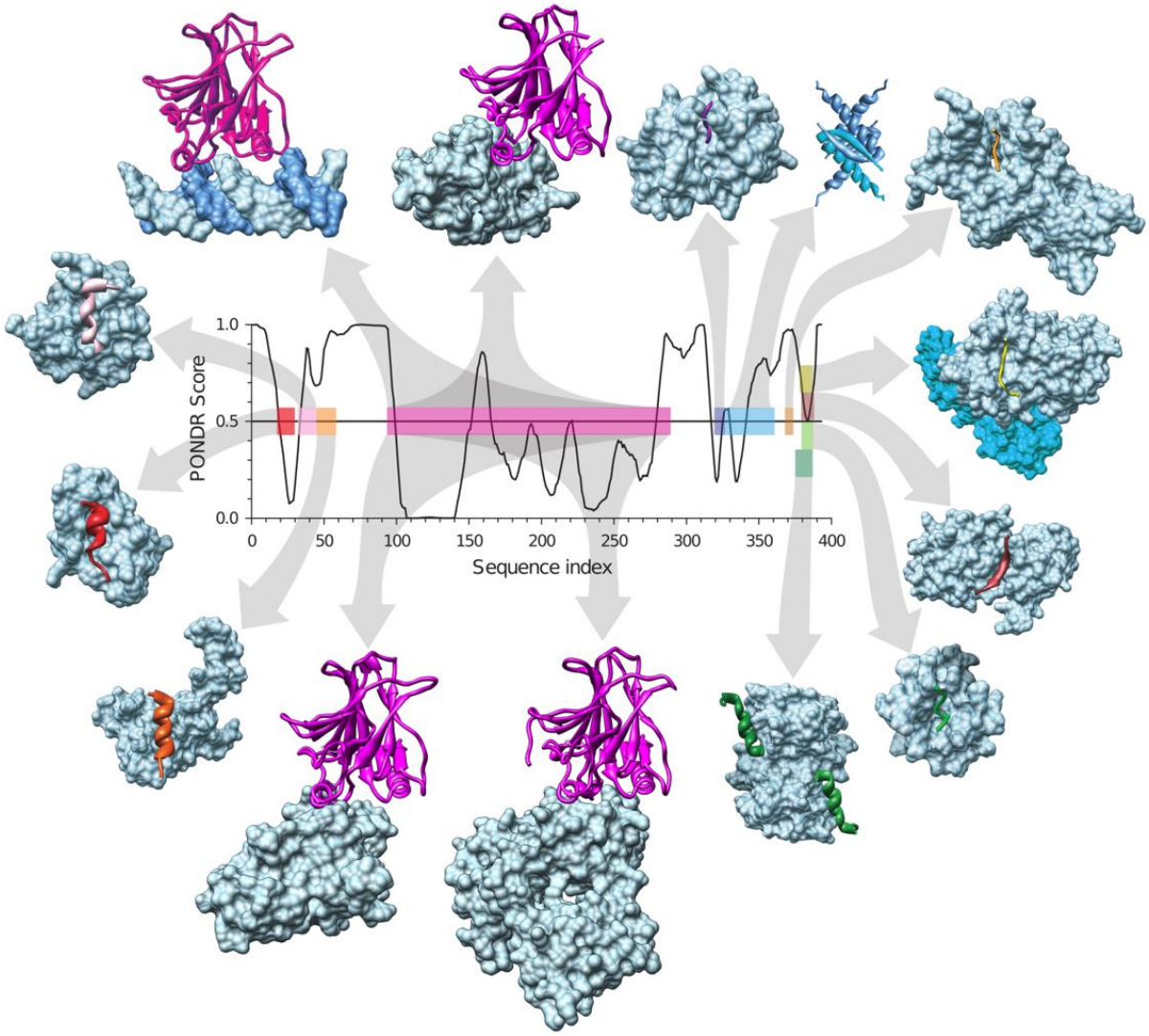

1.0
0.5
0.0
PONDR Score
0
50
100
150
200
250
300
350
400
Sequence index

Figure 5