

Review

# Order, Disorder, and Everything in Between

Shelly DeForte<sup>1</sup> and Vladimir N. Uversky<sup>1,2,3,\*</sup>

<sup>1</sup> Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA; sdeforte@health.usf.edu

<sup>2</sup> USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

<sup>3</sup> Laboratory of Structural Dynamics, Stability and Folding of Proteins, Institute of Cytology, Russian Academy of Sciences, St. Petersburg 194064, Russia

\* Correspondence: vuvversky@health.usf.edu; Tel.: +1-813-974-5816; Fax: +1-813-974-7357

Academic Editor: Derek J. McPhee

Received: 21 July 2016; Accepted: 11 August 2016; Published: 19 August 2016

**Abstract:** In addition to the “traditional” proteins characterized by the unique crystal-like structures needed for unique functions, it is increasingly recognized that many proteins or protein regions (collectively known as intrinsically disordered proteins (IDPs) and intrinsically disordered protein regions (IDPRs)), being biologically active, do not have a specific 3D-structure in their unbound states under physiological conditions. There are also subtler categories of disorder, such as conditional (or dormant) disorder and partial disorder. Both the ability of a protein/region to fold into a well-ordered functional unit or to stay intrinsically disordered but functional are encoded in the amino acid sequence. Structurally, IDPs/IDPRs are characterized by high spatiotemporal heterogeneity and exist as dynamic structural ensembles. It is important to remember, however, that although structure and disorder are often treated as binary states, they actually sit on a structural continuum.

**Keywords:** unfolded; flexible; unstructured; intrinsically disordered; protein function; structural heterogeneity; multi-functionality

---

## 1. The Dominant Paradigms in Protein Science

The dominant paradigms in protein science were, in many ways, shaped by the earliest experiments in the field. Those early experiments were constrained by many of the same limitations we have today. Can this protein be purified? Does it have consistency and simplicity in its function? Ultimately, the measures of success in the dominant methods of experimentation can direct scientific thought regarding what is most important in the study of proteins. As an example, early protein studies in the nineteenth century revolved around the easily obtained and easily crystallized protein hemoglobin. Myosin, also easily available and identified around the same time as hemoglobin, was largely ignored because it was not crystalline. Because of this, it would be another 100 years before we understood Myosin at even the most basic level [1].

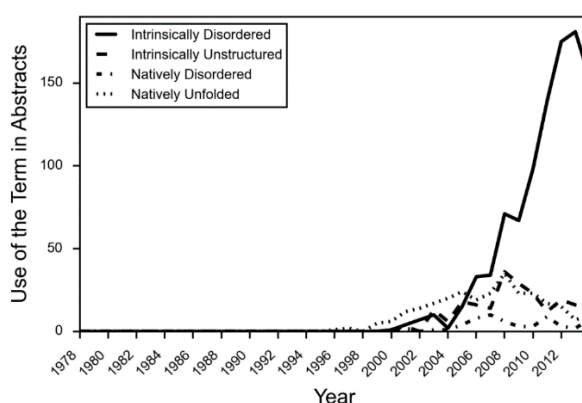
Jacob Berzelius coined the term “catalysis” in 1836 [2], amidst intense interest in enzymes, which at that point had not yet been shown to be proteins. Emil Fischer demonstrated enzyme specificity, and established his seminal lock and key model in 1894 [3], a model that still dominates our understanding of enzyme catalysis today. Interestingly, Fischer also proposed that proteins would prove to have a maximum length of 4000 amino acids [1], demonstrating the intuitive, and incorrect, speculations that follow from a strict adherence to the lock and key model. Experimental confirmation of the structure-function relationship continued with the Anfinsen's demonstration that RNase A could be re-natured in vitro from the completely unfolded state with an accompanying restoration of function [4], followed by early X-ray crystal studies of enzymes such as lysozyme [5] and ribonuclease-S [6]. It is undeniable that the success of these early experiments helped to shape dominant ideas of well-behaved

protein behavior, where well-behaved was synonymous with well-structured, with one singular function, and one mechanism of action.

In many ways, these early experiments represented canonical examples of how proteins should be, which all further experiments were then held against. Therefore, protein behaviors that ran counter to the expected results were considered anomalies. That a protein could take an extended form with minimal residual structure was well understood due to numerous denaturation experiments. However, it was assumed that the native and functional state of a protein must have a stable structure. Therefore, results that ran counter to this assumption were typically considered a problem with the experiment or the experimenter, and not a result of the intrinsic properties of the protein. As these anomalies accumulated to the point where they could no longer be ignored, these problem proteins and problem regions were often seen as functionally irrelevant, and in many cases, removed before experimentation.

## 2. Defining Intrinsically Disordered Proteins

Even while the structure–function paradigm was strengthened in protein science, examples of intrinsically disordered proteins (IDPs) regularly appeared in the literature. Prompted by the increased application of optical rotary dispersion in the 1950s and 1960s to the investigation of protein structure, Jirgensons suggested a classification scheme that included a category called “disordered” [7]. By this time, phosphovitin [7], casein [8], and histones [9] had been shown to have unusual structural properties. In 1971, it was proposed that two regions of missing electron density in the X-ray crystal structure of staphylococcus nuclease were “disordered” [10] as well. However, as more IDPs were uncovered, a wider variety of terms were applied to describe the phenomenon. Tau was initially referred to as “natively denatured” [11], while  $\alpha$ -synuclein was called “natively unfolded” [12]. Early reviews and theoretical work in the field used various terms as well, such as “intrinsically unstructured” [13], “natively disordered” [14] and “loopy” [15], among others. Vague terms such as “flexible” [16] and “mobile” [17] have a long history of hiding in the literature as well. In fact, until about 2005, the four most common terms “intrinsically disordered”, “intrinsically unstructured”, “natively disordered”, and “natively unfolded”, were all used about equally in the literature (Figure 1) [18]. However, after that time, due in part to a concerted effort in the field to use consistent terminology, the term “intrinsically disordered” became the predominant and agreed upon term.



**Figure 1.** The usage of IDP terminology in PubMed abstracts. The occurrence of each IDP term was counted for each year in abstracts of articles in PubMed associated with 1127 known IDPs.

The field of un-structural biology has arisen to try to explain all cases of proteins that fall outside of the structure–function paradigm, and it is necessarily broad in scope because of this. Therefore, just as there have been challenges in reaching a consensus on terminology, there have also been similar challenges in defining protein intrinsic disorder. However, despite these challenges, several common definitions have emerged.

Different definitions of IDPs emphasize different experimental and theoretical perspectives. An IDP may be described as having little or no ordered secondary or tertiary structure. This definition emphasizes an IDP's difference from proteins as understood by the tools of structural biology. An IDP may also be described as under-folded, or as failing to fold independently. This definition emphasizes that an IDP may have the same physical properties as the unfolded state or as a folding intermediate of an ordered protein, such as random coil, molten globule, or pre-molten globule. Finally, it has become increasingly common to focus on the ensemble nature of IDPs when defining them. This places IDPs in the context of behavior that can increasingly be measured by NMR. The properties of IDPs are typically described as being present *in vivo*, *in vitro*, or under functional conditions. This is to emphasize that IDPs display their structural properties in a functional, native state.

It is interesting to note that all of these definitions do not place disordered proteins in a single opposite position from ordered proteins, but instead place ordered and disordered proteins at different points on a continuum. This is clearer, when we understand that both ordered and disordered proteins have movement at the atomic level and about their Ramachandran angles. However, in ordered proteins this motion is sufficiently small that a consensus position can be inferred. On the other hand, an intrinsically disordered protein has movement that precludes the collapse into a single point, both at the ensemble and individual protein level.

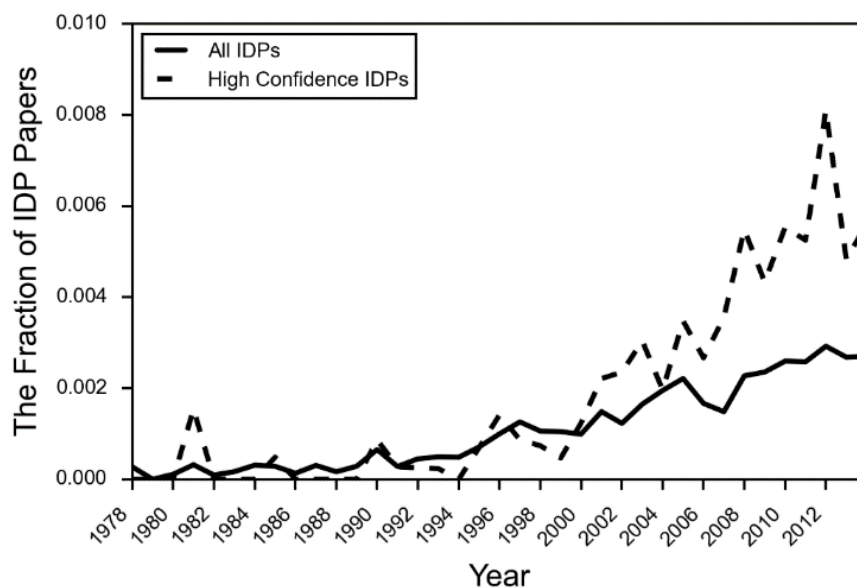
From these definitions, one can extrapolate a theoretical definition based on theories of protein folding. If a protein folds into its lowest energy conformation, then we can define an IDP as a protein that does not have a single global minimum in conformational space [19] or, alternately, IDPs can be described as having a relatively flat free energy surface [20]. Unfortunately, this theoretical definition cannot currently be characterized experimentally for structured or disordered proteins.

Finally, it is necessary to distinguish proteins that are mostly or fully disordered from proteins with isolated regions of disorder. The term IDP is used to refer to proteins that are fully disordered, or contain long, defining regions of disorder. In contrast, when a protein is mostly structured but displays some regions of disorder, it is said to have intrinsically disordered protein regions (IDPRs). Proteins that contain a mix of ordered and disordered regions are also called hybrid proteins.

### 3. The Subtler Side of Disorder

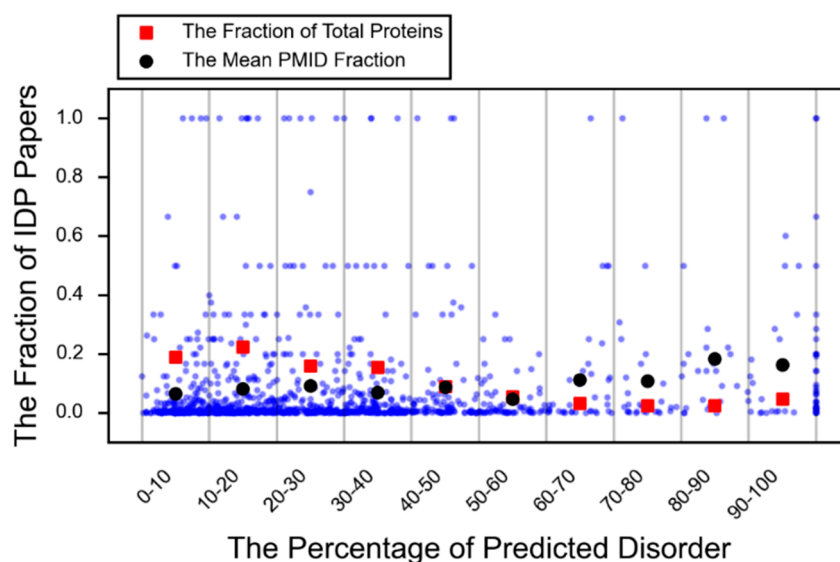
In much the same way that protein science has been shaped by the dominant experimental methods, so too has the field which specializes in studying protein intrinsic disorder. Early measurements of protein intrinsic disorder were obtained by low-resolution methods such as optical rotary dispersion and circular dichroism. These methods cannot measure individual regions of disorder, but only the structural properties over the whole protein. X-ray crystallography can indicate small regions of possible disorder by their absence in the resolved three-dimensional structure, but cannot establish the cause. Therefore, the early emphasis in the field was on proteins that are mostly or fully disordered *in vitro* or *in vivo*, such as Myelin Basic Protein [21],  $\alpha$ -synuclein [12], MAP2 [22], and tau [11].

Aside from the experimental challenges, there are additional reasons why the IDP literature has remained focused on highly and consistently disordered proteins. Acceptance of the relevance, and even the existence of IDPs and IDPRs has not yet solidified in the literature. Discussions of IDPs in textbooks are still largely absent, with a few exceptions in the last five years [23]. The citation aggregator PubMed did not add "intrinsically disordered proteins" to its MeSH (Medical Subject Headings) terms until 2014. The number of papers in the body of literature covering IDPs that actually use IDP terminology is still a fraction of a percent (Figure 2). Despite this, the rise of NMR and sequence-based bioinformatics methods has greatly expanded the experimental and theoretical literature on the topic of IDPs [23]. We now have the experimental tools to begin to characterize subtler categories of disorder, such as conditional disorder and partial disorder.



**Figure 2.** The fraction of PubMed IDs using IDP terminology by year. The fraction for each year is calculated by the number of PubMed IDs associated with IDPs that use IDP language, divided by the total number of PMIDs associated with the IDP proteins in the set. High confidence IDPs are those that have an extensive amount of experimental evidence verifying that the protein is intrinsically disordered.

A protein that is conditionally disordered is either ordered or disordered based on the environmental context or interaction partner. It is a term that encompasses both disorder-to-order transitions and transient (or cryptic) disorder, which is functionally relevant disorder that arises from structured regions (order-to-disorder) [24]. Furthermore, an increasing number of examples of small, but important IDPRs are appearing in the literature as can be seen in Figure 3 (red square).



**Figure 3.** The fraction of predicted disorder versus the fraction of PubMed IDs that use IDP language. Each blue dot represents a protein. The percent predicted disorder is plotted against the fraction of PubMed IDs that use intrinsic disorder language divided by all PubMed IDs associated with that protein search term. For each fraction of predicted disorder interval (0%–10%, 10%–20%, etc.), the fraction of the total proteins in that interval is plotted in red. The mean of the fraction of disorder PubMed IDs is plotted for each fraction of disorder interval in black.

#### 4. The Thin Line between Order and Disorder

Of special interest are proteins that display the properties of both structure and disorder, specifically proteins that have been at least partially crystallized, and enzymes, which will have a structured catalytic region in most cases. While structure and disorder are often treated as binary states, they actually sit on a continuum. Since many proteins might exist in the middle of the structure–disorder continuum, it is useful to identify the conceptual line between a structured and disordered protein or protein region, and explore why it is necessary to use the tools and language of disorder when structure is present.

When attempting to semantically separate structure from disorder, it becomes clear that neither the term “structured” nor “disordered” is precisely correct. All proteins have some movement, and no protein is completely chaotic. Because these are conceptual frameworks that do not point to precise biological realities, the tools of structural or un-structural biology should be applied when most useful to solve the problem. The separating line between order and disorder is therefore drawn not by theoretical descriptions, but by practical considerations.

When a protein can no longer be adequately described by a single three-dimensional structure or a series of snapshots in three dimensions, then the language of disorder has now become useful. IDPs are defined by conformational uncertainty, and are typically characterized by a combination of sequence level features and a description of the overall shape, i.e., extended, random coil, molten globule, or pre-molten globule. Additionally, an IDP that changes shape must also be described using the axis of time, and it is along this axis that structural biology and un-structural biology most acutely diverge. The introduction of a time variable greatly increases the possibilities when describing IDPs and IDPRs, and it is change over time that allows us to describe the mechanisms that IDPs may employ and the advantages imparted by disorder.

As an example, short segments of disorder are commonly observed in the form of hinges that move a domain in a controlled way, or loops that have an open and closed conformation, such as the WPD loop in the bacterial protein tyrosine phosphatase YopH [25]. While these regions are technically disordered, the ability to describe the movement as a series of structural snapshots typically places these dynamic movements within the realm of structural biology. However, other small disordered segments called Molecular Recognition Features (MoRFs) [26], which undergo a contextual transition between disorder and order upon binding, have a function that is defined by the presence of disorder and the transition to an ordered state, and not by a specifically defined three-dimensional structure. Therefore, even despite their short lengths and disorder–order transitions, MoRFs fall within the realm of un-structural biology.

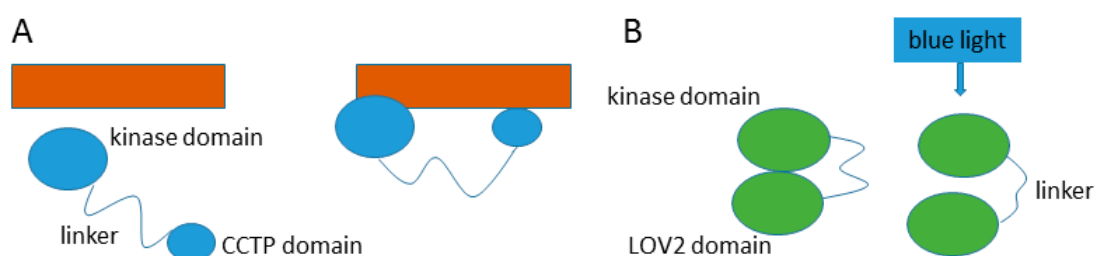
Furthermore, whether a protein is considered to be an IDP or to have an IDPR is largely determined by the functional significance assigned to the disorder. Disordered regions that have no known function are often considered to be functionally neutral sequence noise. An a priori assumption that a disordered region is function-neutral can create circular support for itself if this region is removed before experimentation; therefore, the identification of disorder specific functions is of critical importance.

#### 5. The Mechanisms of Disorder

A great deal of experimental and theoretical work has been done in order to illuminate how IDPs and IDPRs fit within a functional protein universe. When viewed within the context of finely regulated interaction and signaling networks, where proteins may need to display multiple context dependent behaviors, the advantages of disorder begin to become clear. While many historical examples of the functional properties of IDPs and IDPRs have come from studies of non-enzymes, an increasing number of more recent studies have shown that these functional advantages are also demonstrable in enzymes.

### 5.1. Entropy

Entropic functions make use of the advantages inherent in dynamic movement within a disordered protein chain. Entropic chains can provide precise spacing between functional domains, creating a less restricted search space, maintaining separation between domains, or creating the opportunity for two or more domains to interact with each other, or with another partner. Two enzymes with kinase domains demonstrate the utility of a disordered interdomain linker. The kinase Yck2 has a long disordered interdomain linker that allows the kinase domain and a conserved C-terminal peptide (CCTP) domain to interact with two separate Akr1 domains simultaneously [27], while a disordered interdomain linker in Phototropin 2 becomes elongated when exposed to blue light irradiation, preventing the LOV2 domain from making contact with and activating the kinase domain [28] (Figure 4). Additionally, there are non-enzyme examples of entropic clock functions, such as the voltage-gated potassium channel of nerve axons which uses a ball and chain mechanism to inactivate the channel [29]. Entropic bristles use entropy to fill space as is seen in the gating of the nuclear pore complex through the repulsion caused by disordered nucleoporins [30].



**Figure 4.** Entropic chain functions. (A) Yck2 (in blue) uses a disordered interdomain linker to bind to two separate domains on Akr1 (in orange); (B) The disordered interdomain linker in Phototropin 2 (in green) becomes elongated when irradiated with blue light, causing the activating LOV2 domain to separate from the kinase domain.

### 5.2. Accessibility

Posttranslational modification (PTM) requires site accessibility, so it is therefore not surprising that many PTM sites are embedded in disordered regions, which can provide a large surface area with a limited number of residues. Phosphorylation sites in particular have been shown to be enriched in disordered residues [31]. Most well-known IDPs have phosphorylation sites, and they have been demonstrated in the IDPRs of enzymes as well. For example, the intrinsically disordered Cap region of both Abl and Arg non-receptor tyrosine kinases is rich in phosphorylation sites that regulate multiple domains [32].

Site accessibility is also required for proteolytic processing that generates protein fragments with altered activity. For example, the phosphatase calcineurin contains an intrinsically disordered regulatory domain that is susceptible to proteolytic cleavage *in vitro* [33] and *in vivo* [34] that significantly increases its activity.

### 5.3. Plasticity

Some of the most striking functional advantages of IDPs and IDPRs come from their ability to contextually change due to binding or environmental cues. IDPs and IDPRs may change shape in response to interactions with proteins, nucleic acids, and other ligands, allowing them to specifically bind to a wide variety of partners. For example, the disordered loop near the active site of Mitochondrial 2,4-Dienoyl-CoA reductase allows this enzyme to accommodate a wide range of fatty acids [35]. Protein Kinase R has two intrinsically disordered interdomain regions that may allow it to effectively dimerize when interacting with a RNA activators of varying size and shape [36]. Plasticity may also be helpful in identifying and negotiating disordered regions in substrates, and appears to

play an important role in ubiquitination pathways. For instance, E3 ubiquitin ligases that bind both substrates and E2 ubiquitin conjugating enzymes are significantly more disordered than those that engage in single interactions [37].

## 6. Disorder-Related Biological Functions

The biophysical mechanisms used by disordered regions are disproportionately connected to particular biological roles, such as signaling and molecular and cellular regulation [38]. This is not surprising, considering the ability of disordered regions to change over time and to adapt based on the environmental context.

### 6.1. Signaling

Signaling pathways provide a way for complicated biological systems to coordinate physiological activities and responses. Signaling pathways can usually be described as a linear cascade of interactions that triggers some kind of change in the cell. Frequently PTMs such as phosphorylation play a key role in this, and disorder frequently plays a role both in the accessibility of PTM sites, and the activity of kinases [39]. Key sequence signals may be encoded in disordered sequences, such as in sulfhydryl oxidase ALR which has an IDPR that acts as a mitochondrial targeting signal in the cytosol and a recognition site in the disulfide relay system of the intermembrane space [40]. Receptors, which are often a starting point for a signaling cascade, may have IDPRs that expose phosphorylation sites or bind to signaling molecules as is seen in the receptor tyrosine-protein kinase ErbB2 [41].

### 6.2. Regulation

IDPs are frequently involved in regulation at the cellular level through involvement in gene transcription [42] and protein degradation [37,43], and at the protein level, through allosteric effects or PTMs that result in the masking and unmasking of interaction sites. As an example, phosphorylation of the IDP 4E-BP2 acts as a regulatory switch by inducing a disorder-order transition and preventing binding with eIF4E [44]. Conversely, regulation of glucokinase is facilitated by an order-disorder transition that causes a time-delay when glucose is low [45].

IDPs are also abundant in protein degradation pathways. There are a number of E3 ubiquitin-protein ligases which have long stretches of disorder that appear to mediate interactions with a variety of mostly disordered substrates [37]. For example, San1 is an E3 ubiquitin-protein ligase which has extended disorder in its N and C terminal substrate binding regions. Interestingly, San1 avoids auto-ubiquitination through the absence of lysines in its disordered binding regions [46]. Ubiquitin-independent protein degradation pathways also involve disordered protein regions. The enzymes thymidylate synthase and ornithine decarboxylase both contain IDPRs that appear to contain the sole requirements for ubiquitin independent degradation [43].

## 7. Disorder and Protein Evolution

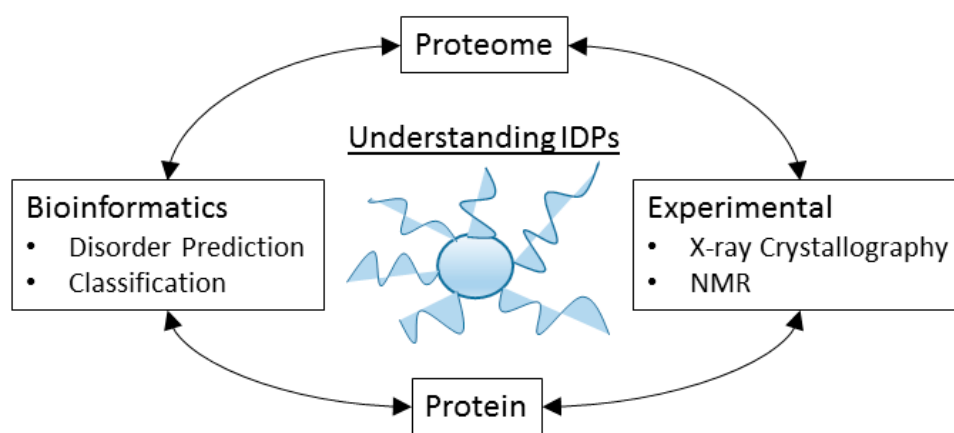
The evolution of protein coding regions in genomes presents a fundamental mystery. The human genome, the genome of the flowering plant *Arabidopsis thaliana*, and the genome of the protozoa *Tetrahymena* are all estimated to have approximately 27,000–29,000 protein coding regions. On the other hand, *Danio rerio* (zebrafish) and *Mus musculus* have close to 40,000 protein coding regions each [47]. It is clear that protein coding regions do not scale linearly with organismic complexity. One compelling explanation for this is that protein intrinsic disorder combined with alternative splicing facilitates tightly regulated and context specific multi-functional behavior that allows more complicated organisms to make the most of a limited genome [42,48].

There are several pieces of evidence to support a hypothesis of evolutionarily directed functional disorder. Bioinformatics analyses show that eukaryotes are more disordered than prokaryotes [49], natural sequences are more disordered than random sequences even with the same amino acid composition [50], and disorder within natural sequences is non-random in its patterns [51]. Disordered

regions tend to evolve more rapidly while maintaining their physiological functions [52], therefore, functional disorder may provide an advantage by buffering a genome against mutations. Finally, it can be argued that complex signaling networks and finely tuned regulatory mechanisms are themselves a response to organismic complexity; therefore, the overrepresentation of IDPs and IDPRs in signaling and regulation also supports the hypothesis of the directed evolution of disorder in genomes. Interestingly, some protists have more disorder than multicellular eukaryotes, suggesting there may be an optimal amount of disorder for an organism that is partly based on lifestyle [49].

## 8. The Tools of the Un-Structural Biologist

The study of IDPs and IDPRs requires a large number of experimental and computational methods, and typically the results of these experiments combine together to form a picture of the disorder properties over the protein and the proteome (Figure 5).



**Figure 5.** Experimental and bioinformatics techniques work together to describe the properties of disorder in proteomes and proteins.

### 8.1. Some Experimental Techniques

#### 8.1.1. X-ray Crystallography

It is somewhat surprising that X-ray crystal structures provide one of the largest datasets of experimentally indicated IDPRs, considering that X-ray crystallography is one of the primary tools of structural biology. However, missing regions in X-ray crystal structures are often caused by IDPRs. The challenge with using this data, however, is that missing regions are an imperfect indication of protein intrinsic disorder, as there are multiple possible explanations for a missing region, including experimental artifacts or annotation errors. Furthermore, authentic IDPRs identified in X-ray crystal structures will be non-representative in terms of the size of the region and the amino acid composition, due to their emergence from a very structured set of proteins. The decision to use X-ray crystal structure data as an indication of disorder must therefore be made by balancing the usefulness of a large amount of data against the imperfections in the data.

#### 8.1.2. Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is arguably the current best experimental technique for identifying protein intrinsic disorder and conformational ensembles. The key differences that make NMR superior to X-ray crystallography for identifying IDPs is that NMR does not require crystallization and NMR can provide direct observation of disorder instead of simply indicating a lack of structure. However, there are limitations in the size of the protein that restrict the applicability of NMR, and the amount of NMR data is still significantly less than the amount of X-ray crystal structure data.



Identifying IDPs and IDPRs using NMR can be accomplished through several different approaches. A collapsed HSQC NMR spectrum will indicate disorder over the entire protein, whereas a dispersed spectrum will indicate a structured protein. NMR techniques can also be used to generate conformational ensembles and multiple methods can be employed to measure the differences between ensembles [53,54]. Additionally, chemical shift and  $^{15}\text{N}$  ( $^1\text{H}$ ) NOE data can provide flexibility information without the requirement for any structural models [55].

### 8.1.3. Combining Experimental Techniques

The number of experimental techniques that can be used to study IDPs is extensive, and in fact multiple books [56,57] and reviews [58,59] have been dedicated to this topic. In practice, multiple techniques of varying resolution are typically employed and the aggregated evidence is used to create models of the disordered regions. These include low-resolution spectroscopic techniques such as circular dichroism, optical rotary dispersion, Fourier-transform infrared spectroscopy, and deep-UV resonance Raman spectroscopy. Additionally, the level of protein compaction can be measured by small angle X-ray scattering, small angle neutron scattering, gel-filtration, and viscometry. The properties of individual protein molecules can help identify ensemble properties and can be obtained via high speed atomic force microscopy (AFM), and single-molecule fluorescence resonance energy transfer (SM-FRET).

## 8.2. Bioinformatics Analysis

Bioinformatics tools and analysis have played a large part in the study of IDPs and the establishment of the field. The tools used to study IDPs typically focus on extracting information from protein sequences, however, genome studies focusing on the evolution of IDPs and the enrichment of splicing sites in disordered regions are common as well [60,61]. Several recent reviews have been written focusing on different aspects of bioinformatics analyses of IDPs such as the discovery of degenerate motifs in IDPs [62], predicting function in IDPs [63], and the prediction of IDPs by protein sequence [64]. Indeed, the computational tools used to analyze IDPs and IDPRs are as vast as the experimental tools.

### 8.2.1. Sequence Characteristics

Because of the lack of a stable three-dimensional structure, the computational study of IDPs and IDPRs is predominantly dependent on primary sequence information. Anfinsen's dogma suggests that the three-dimensional structure of a protein is encoded into the primary sequence [65], however, the accurate prediction of the folded structure from primary sequence remains elusive. Tools to predict disorder from primary sequences have been much more successful, however. This is intuitive from the perspective of entropy.

A three-dimensional structure has only one form that it can take, whereas the conformational fluctuations that can define a disordered protein are nearly infinite in their possibilities within steric limitations; therefore, predictors of disorder require less information than predictors of structure. IDPs have distinct sequence characteristics that facilitate the identification of disorder from sequence. IDPs are enriched in specific disorder promoting residues, such as alanine, glycine, serine, proline, glutamine, glutamic acid, lysine and arginine, and they are depleted in the order promoting residues isoleucine, valine, leucine, phenylalanine, cysteine, tryptophan, tyrosine, and asparagine [66,67]. These residues are roughly correlated with flexibility [68] and hydrophobicity scales [69] (Figure 6).

Low complexity regions are often disordered, and disordered regions are often enriched in low complexity motifs. However, neither disorder nor low complexity necessarily implies the other [70]. Disordered regions tend to have lower sequence conservation in families, however, there are also well conserved disordered domains [71]. Furthermore, in poorly conserved disordered regions, the chemical composition is often preserved [72,73].



**Figure 6.** Amino acid scales and disorder and order promoting residues. (**Top**) Ranking of the 20 amino acids by the Kyte-Doolittle hydrophobicity scale from most to least hydrophobic; (**Bottom**) Ranking of the amino acids from most to least flexible by Vihinen's flexibility scale.

### 8.2.2. Disorder Prediction

The distinct sequence features that are present in IDPs and IDPRs allow the construction of sequence based rules that can facilitate high performance disorder prediction. Over 70 predictors of disorder have been created since 1997 [73,74]. A favorable balance between true positives (TP) / true negatives (TN) and false positives (FP)/false negatives (FN) is the objective, and this is typically expressed by the Matthews Correlation Coefficient (MCC).

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) (TN + FP) (TP + FN) (TN + FN)}}$$

The CASP competition judges disorder predictors based on as yet unpublished disordered regions, which are usually obtained from missing regions in newly published X-ray crystal structures [75]. The highest ranking predictors in the CASP experiment have an MCC of approximately 0.5 and these results are usually achieved by slower predictors that use multiple sequence alignments along with sequence based features such as amino acid composition and the physicochemical properties of the amino acids. These slow but high performing predictors, such as PrDOS [76], SPINE-D [77], and DISOPRED3 [78], are best for small datasets and single protein prediction. Large datasets, however, require the use of faster predictors that can be run on a local computer, such as Espritz [79], and IUPred [80,81].

Disorder predictors must be trained and tested on datasets of experimentally indicated disordered residues. These datasets commonly come from Disprot [82], X-ray crystal structures in the PDB [83], or NMR data. Most disorder predictors will give a per-residue disorder score between 0.0 and 1.0, and the generally agreed upon threshold for disorder is greater than or equal to a score of 0.5. Some disorder predictors, such as SLIDER [84] or RAPID [85], will provide fast prediction for complete proteomes, by calculating a single score across the entire protein. Other predictors may provide a score that is calibrated differently, such as DynaMine which produces scores in the form of backbone N-H S<sup>2</sup> order parameter values [55,86]. When using DynaMine, a score below 0.7 is considered flexible.

Meta-predictors, which combine the outputs from multiple single predictors are also common, such as PONDR-FIT [87] and MetaDisorder [88]. A consensus of multiple predictors can provide improved results [89], as it will theoretically reduce the bias inherent in single predictors that were trained on limited datasets. However, despite a modest improvement through consensus methods, disorder prediction based on currently available datasets has likely hit a bottleneck in terms of the maximum possible MCC scores.

This limitation arises in part due to the imperfections in the testing and training sets for the development of disorder prediction. Experimental indications of disorder are gathered over a wide range of experimental techniques, including low resolution techniques such as circular dichroism which may not provide accurate estimates of the exact disordered residues, X-ray crystallography, which only provides an indication of disorder, but may be caused by other factors, and NMR, which requires significant, and therefore variable interpretation in order to assign disorder. A second issue, which further compounds this, is the presence of conditionally and partially disordered regions, which may be assigned as ordered or disordered, depending on the experiment. Finally, it is likely that there

are different flavors of disorder [90] with different sequence based markers, yet a clear classification scheme has not yet been created.

Disorder prediction should be employed in analysis with an appropriate awareness of the inherent level of error. However, despite these considerations, disorder prediction still provides a way to separate unique sequence regions that indicate a propensity towards disorder, and provides a useful level of biological accuracy, especially at the proteome level. While individual residues may not be assigned correctly in all cases, disorder prediction still provides an illuminating look into the propensity of the protein to be solvent exposed, to undergo dynamic transitions, or to be destabilized by environmental factors.

### 8.2.3. Classification of Functions

There appears to be a relationship between biophysical function, cellular function, and sequence characteristics, however, the identification and development of these relationships is still in its early stages [91]. Therefore, a major task in bioinformatics is to attach biological sequence information to physical behavior, biological functions, and cellular response. Functional classification based on sequence requires two sets of information. The first is sequence based features. These may be disorder prediction scores, calculations based on the physicochemical features, or amino acid motifs. The second set of information is functional annotation. Gene Ontology (GO) term assignments are one of the primary sources of annotations related to cellular components, biological processes, and molecular functions [92]. GO terms can be assigned based on experimental evidence, or can be inferred based on homology. Additionally, Enzyme Commission (EC) numbers provide a useful annotation protocol when classifying enzymes [93]. EC numbers are assigned based on the chemical reaction that is catalyzed by the enzyme. Similar to GO terms, EC designations can be made based on direct experimental evidence or can be inferred through sequence homology.

### 8.2.4. Proteome Level Studies

The early assumption in protein science was that protein intrinsic disorder represented an unusual and isolated phenomenon. In fact, it can be argued that this assumption is still held by many researchers today [18]. Therefore, the application of disorder prediction to whole proteomes has been critical to establishing the ubiquity and relevance of protein intrinsic disorder while the tools for large scale experimental identification are still nascent. Despite differences in the disorder predictors used, the proteomes they have been applied to, and the different measures applied, the consensus is that eukaryotes tend to have more predicted disorder than prokaryotes or archaea [49], and within eukaryotic proteomes especially, intrinsic disorder is exceptionally common [94]. For instance, Ward et al. found that 2.0% of archaean, 4.2% of eubacterial and 33.0% of eukaryotic proteins had disordered regions greater than 30 residues in length [95]. Estimates of the average fraction of disorder for eukaryotic proteomes tends to be between 20% and 30%, while prokaryotes tend to be closer to 5%–10%, however, there is significant variation and overlap in disorder prediction between the taxa [49,94].

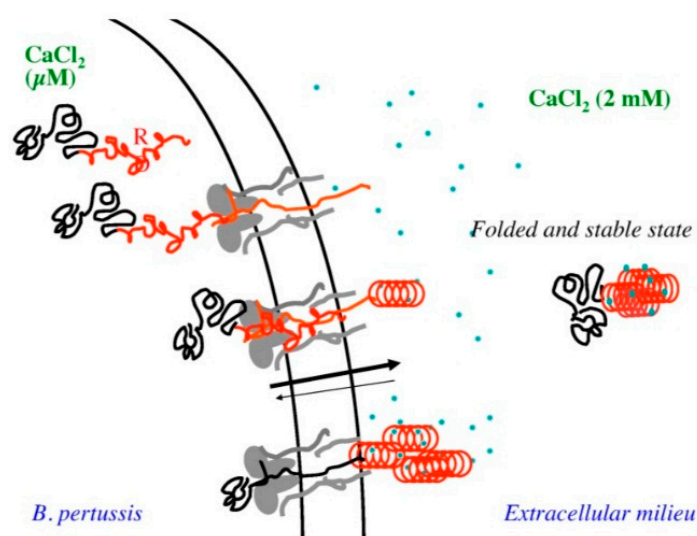
## 9. Protein Intrinsic Disorder and Disease

In 2008, Uversky et al. introduced the disorder in disorders ( $D^2$ ) concept, and showed that proteins with IDPRs greater than 30 residues in length are overrepresented in proteins involved with signaling, cancer, neurodegenerative diseases, cardiovascular diseases, and diabetes [96]. While this relationship may suggest innate pathogenicity in IDPs and IDPRs, studies suggest instead that IDPs and proteins with IDPRs perform tightly regulated [97] and necessary functions, many of which depend on the lack of a three-dimensional structure [98]. However, as is the case in structured proteins, genetic, environmental, or systemic perturbations can make IDPRs and IDPRs susceptible to misfolding and misregulation.

Some of the most well-known examples of disease related IDPs are implicated in neurodegeneration, such as tau,  $\alpha$ -synuclein, beta amyloid, and prion protein [99]. Flexibility in these proteins can facilitate perturbations into misfolded, aggregated states, and therefore the disease state is directly related to a structural transition facilitated by the disordered properties of the protein. However, there are also a myriad of potential roles that IDPs and IDPRs can play in disease processes. The pathogenic behavior of an IDP or protein with an IDPR may be triggered by genetic factors such as pathogenic mutations, alternative transcription, or aberrant splicing, or non-genetic factors such as altered protein expression levels, PTMs, or aberrant cleavage. These cellular changes may result in misfolding, loss of normal function, gain of toxic function, protein aggregation, misidentification, misregulation, or missignaling (reviewed in [96] and [100]).

While uncommon, there are some interesting examples of disease associated enzyme IDPs in the literature. For instance, virulence factors with catalytic domains may utilize long IDPRs to translocate or avoid host defense systems. The adenylate cyclase toxin in *Bordetella pertussis* provides an interesting example of this phenomenon. The adenylate cyclase toxin contains a Repeat in ToXin (RTX) motif, which is intrinsically disordered in the absence of calcium inside the bacterial cell.

This allows the enzyme to translocate the catalytic domain across the narrow type 1 secretion channel, and then transition to a globular structure when exposed to the calcium gradient on the bacterial cell wall [101] (Figure 7).



**Figure 7.** A schematic representation of the secretion of adenylate cyclase toxin through the type 1 secretion system. Reprinted under the creative commons license from ref. [101].

The human acetylcholinesterase variant, AChE-R, provides an interesting counter point to IDPR associated pathology. The AChE-R variant has an intrinsically disordered C-terminus, and the presence of this disordered region appears to provide neuroprotective effects in Alzheimer's disease as compared to the AChE-S variant, which has a helical C-terminus, and appears to accelerate the formation of amyloid fibrils [102]. It is likely that many examples of enzymes with IDPRs that are involved in disease processes in both pathogenic and protective capacities will emerge with the increased acceptance of the functional roles that IDPRs in enzymes may play.

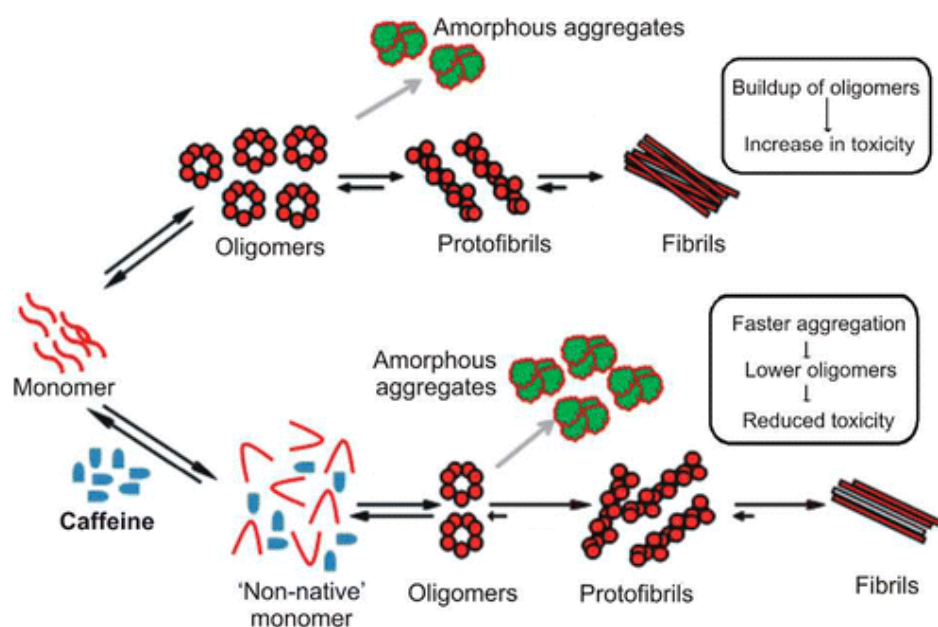
## 10. Protein Intrinsic Disorder and Drug Design and Discovery

The enrichment of long IDPRs in proteins associated with disease processes presents a rich source of potential drug targets, however, there are several challenges in drug design and discovery when IDPRs are targets. Secure binding to a small molecule requires stabilization in the binding site, and this may not be possible in all IDPRs. However, functional IDPRs will often take on a transient structure

due to binding or environmental factors which suggests that inducible structure may be possible in many cases.

In proteins that are mostly disordered in their native state, the disorder can be understood as a population of many interconverting conformations, with the potential to stabilize a single non-pathogenic conformation. The investigation of natural compounds with known effects on IDPs can provide a powerful route to discovery. For instance, the consumption of coffee appears to provide some protection against the development of Parkinson's disease.

A study on the effects of caffeine on  $\alpha$ -synuclein aggregation showed that caffeine modifies the conformation of the monomer form, thus accelerating the aggregation of a less toxic species [103] (Figure 8).

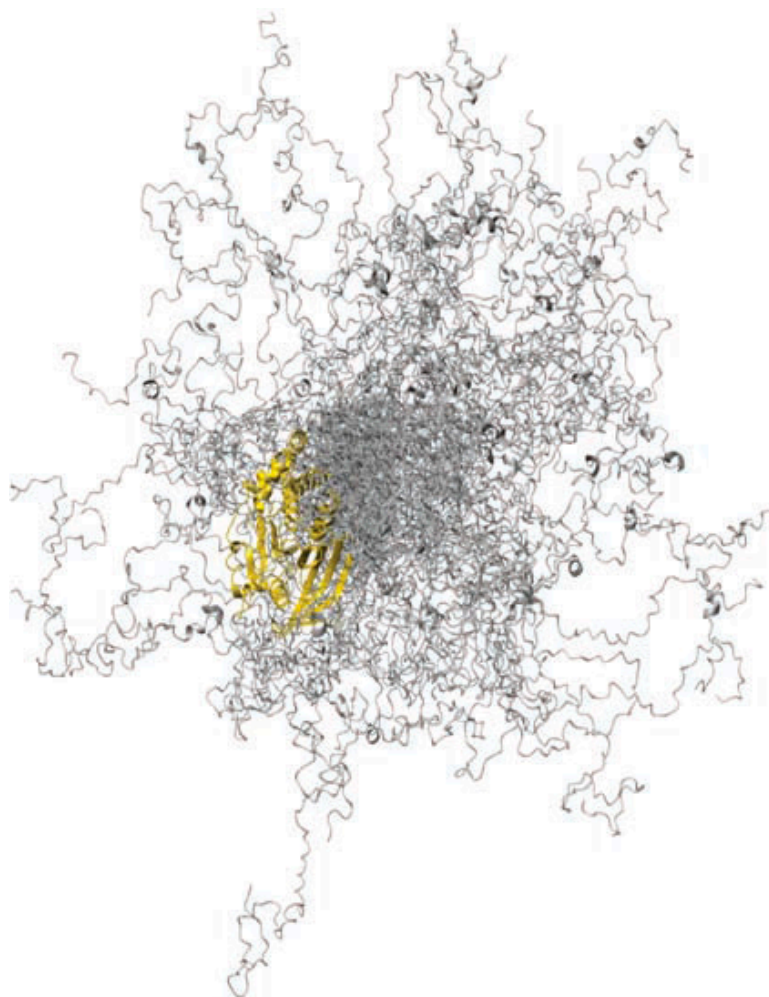


**Figure 8.** A schematic representation of the effects of caffeine on the aggregation properties of  $\alpha$ -synuclein. Reprinted with permission from ref. [103]. Copyright (2015) American Chemical Society.

Drug design and discovery in IDPRs may be more akin to navigating a handshake than placing a lock in a key, however, many of the principles of design and discovery are still the same. However, when stabilizing binding partners are not known, a clear challenge is the absence of a priori knowledge of the three-dimensional structure for in silico screening or rational drug design of favorable compounds. However, blind exploratory assays are standard in drug discovery, and the presence of an IDPR should not prohibit these kinds of screens. Instead, the biggest challenge in drug discovery for IDPRs may be the standard practice of the exclusion of IDPRs before drug screens.

#### *The Story of PTP1B*

The phosphatase PTP1B (Figure 9) provides an illustrative example of delayed drug discovery for an IDPR due to the consistent truncation of the region before drug screens.



**Figure 9.** A representative ensemble of 100 conformers for PTP1B. Reprinted with permission from Macmillan Publishers Ltd.: Nature Chemical Biology, part of Springer Nature. [104].

The catalytic region of PTP1B, encompassing residues 1–321 was purified in 1988 from the human placenta [105], and in 1990 the full length form of 435 residues was uncovered through cDNA cloning [106]. Despite knowledge of the full length form, and an early demonstration of its role in the regulation of PTP1B [107], studies on PTP1B between 1990 and 2014 focused almost exclusively on the originally purified form encompassing residues 1–321, therefore ignoring the disordered C-terminal region.

PTP1B became an attractive therapeutic target due to its involvement in multiple signaling pathways, including those implicated in obesity and diabetes [108]. However, the development of a small molecule inhibitor for the catalytic domain of PTP1B was frustrated by the highly charged nature of the catalytic site [109], and the practice of testing inhibitors against the truncated form. It was not until 2014 that MSI-1436, a known inhibitor of PTP1B *in vivo* [110], was screened against the full length form, and was found to be an effective inhibitor [104].

The C-terminal region of PTP1B is intrinsically disordered, moving within a wide range of three-dimensional space (Figure 9). However, as is often the case in IDPRs, there is residual secondary structure in the form of two small  $\alpha$ -helical regions. The most peripheral of these  $\alpha$ -helical regions provides one anchor point for MSI-1436, while the second anchor point is found close to the catalytic domain. Upon binding to these two regions, PTP1B becomes more compact, and  $V_{\max}$  is decreased.

While MSI-1436 provides a small amount of inhibition of the truncated form of the enzyme, the primary binding site is between residues 367 and 394, and therefore this region is required to observe

the full strength of the inhibition. To our knowledge, this was the first drug screen against the full length form; therefore, it is possible that MSI-1436 or other effective inhibitors had been tried and discarded previously.

The story of PTP1B demonstrates that it is sometimes assumptions about the lack of functional relevance of IDPRs in enzymes that creates one of the largest obstacles to understanding and utilizing these regions in disease intervention.

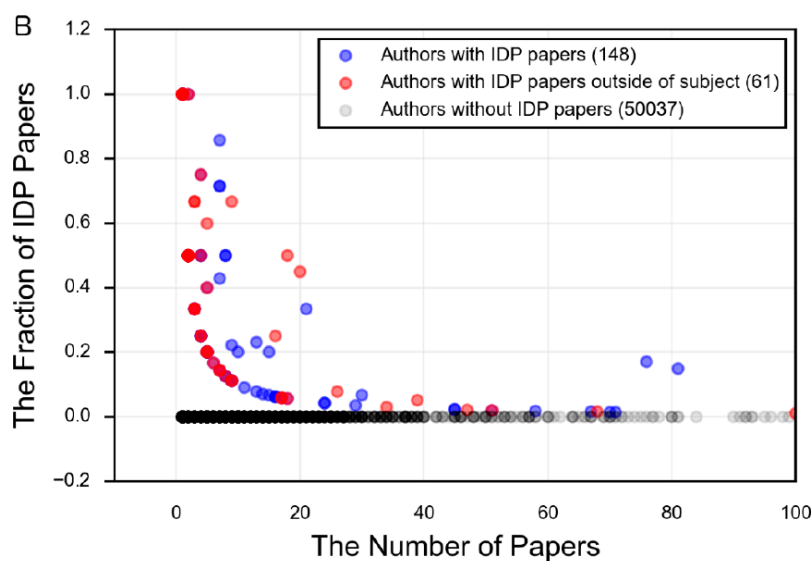
## 11. The Field of Protein Intrinsic Disorder

An un-structural biologist specializes in the tools and techniques used to study IDPs. Furthermore, a specialist in the field of IDPs must be aware of individual proteins identified as IDPs and the body of experimental, proteomic and bioinformatics literature validating the existence of disorder in these proteins. Due to the broad scope of the material covered by the field of protein intrinsic disorder, and the nascence and relative obscurity of the field, there are a number of researchers who focus almost exclusively on the study of IDPs from various perspectives.

The field of protein intrinsic disorder represents a powerful example of the productive relationship between experimental and bioinformatics techniques (Figure 5). For example, an experimentalist who notices unusual structural behavior in their protein may employ the use of disorder prediction to assess the propensity of their protein towards disorder. Using this information as a guide, they can target their research towards those regions predicted to be intrinsically disordered, and apply the appropriate experimental techniques. This experimental data yields information, which can then be used to revise and improve the prediction of disorder in other proteins, to extrapolate evolutionary information for the protein family, or to predict function or biophysical mechanisms. However, the number of researchers who specialize in the study of *specific* proteins that are intrinsically disordered, and who also embrace the language of disorder to describe these properties, is remarkably small (Figure 10), demonstrating that the tools and language of protein intrinsic disorder has not propagated far beyond those who specialize in studying IDPs.



Figure 10. Cont.



**Figure 10.** The number of papers per author for the search term in PubMed, plotted against the fraction of those papers that use IDP terminology. Each point represents an author on one or more papers associated with the given search term. The darker the dot, the larger the concentration of authors at that point. Blue dots are authors who have an IDP paper in the field in question ( $\alpha$ -synuclein or tau, in this case), while the red dots are authors who have an IDP paper in the field in question and also have an IDP paper in a different field. The fraction of IDP papers is the number of papers by that author that use IDP terminology divided by all papers for that author and search term. The following search terms were used: (A) “alpha synuclein”; (B) “tau AND (protein OR Alzheimer’s OR tauopathies OR neuronal)”.

## 12. Conclusions: Intrinsic Disorder Where You Least Expect It

As this new paradigm of un-structural biology becomes more accepted, it becomes clear that the disciplines of structural and un-structural biology can work in tandem to explain the dynamics of a protein over time. One should keep in mind though that from the viewpoint of complete lack of structure and complete sequence randomness, the structure-function relations in IDPs and IDPRs, where the structure is defined as a “low-resolution” ensemble of conformations, where the IDP/IDPR sequences are not random, and where IDPs/IDPRs have specific functions, the phrase “un-structural” biology does sound like a stretch. However, from the viewpoint of traditional “structural biology” that still serves as an operational term for many researchers working in the field of protein science, the aforementioned “low-resolution” ensemble of conformations cannot be considered as “unique structure” needed for a protein to have “unique function”. However, structural and functional characterization of ordered and disordered proteins does require very different methodological approaches, thus justifying and supporting the existence of two very different schools of thoughts. On the other hand, opposing “structural” and “un-structural” biology is counter-productive. In our view, these two concepts should not be opposed, but united, since they clearly complement each other. Furthermore, the line between order and disorder is a practical line, and the language and tools of protein intrinsic disorder become necessary when a protein or protein region can no longer be described in three dimensions. The acceptance of the presence and functional relevance of protein intrinsic disorder, however, remains relatively low, especially in protein populations that are expected to be structured. Therefore, this review highlights several challenges:

- There is a focus in the IDP literature on mostly or fully disordered proteins, and an incomplete understanding of the diverse mechanisms and functions employed by proteins that have regions of both order and disorder.



- The development and validation of disorder prediction depends on experimental datasets of missing regions from the PDB, however, it is not always clear what the cause and nature of the missing region is.
- There is an increasing number of enzymes with experimentally measured IDPRs in the literature, but no proteome level studies up to this point.
- There is limited acceptance of the language of protein intrinsic disorder outside of those who specialize in studying IDPs.
- Drug design and discovery frequently focuses on the truncated form of the protein, potentially resulting in missed opportunities.

Furthermore, it seems that future studies should be focused on addressing the question of how likely IDPRs are to occur in canonically structured proteins, and whether and how these regions are functionally significant.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Tanford, C.; Reynolds, J. *Nature's Robots, a History of Proteins*; Oxford University Press: Oxford, UK; New York, NY, USA, 2001.
2. Jorpes, J.E. *Jacob Berzelius, his Life and Work (Trans. B. Steele)*; Almqvist & Wiksell: Stockholm, Sweden, 1970.
3. Kunz, H. Emil fischer-unequaled classicist, master of organic chemistry research, and inspired trailblazer of biological chemistry. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 4439–4451. [[CrossRef](#)]
4. Anfinsen, C.B.; Redfield, R.R.; Choate, W.L.; Page, J.; Carroll, W.R. Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J. Biol. Chem.* **1954**, *207*, 201–210. [[PubMed](#)]
5. Blake, C.C.; Koenig, D.F.; Mair, G.A.; North, A.C.; Phillips, D.C.; Sarma, V.R. Structure of hen egg-white lysozyme. A three-dimensional fourier synthesis at 2 angstrom resolution. *Nature* **1965**, *206*, 757–761. [[CrossRef](#)] [[PubMed](#)]
6. Wyckoff, H.W.; Hardman, K.D.; Allewell, N.M.; Inagami, T.; Tsernoglou, D.; Johnson, L.N.; Richards, F.M. The structure of ribonuclease-S at 6 a resolution. *J. Biol. Chem.* **1967**, *242*, 3749–3753. [[PubMed](#)]
7. Jirgensons, B. Classification of proteins according to conformation. *Die Makromol. Chem.* **1966**, *91*, 74–86. [[CrossRef](#)]
8. Herriott, R.M. Isolation, crystallization, and properties of swine pepsinogen. *J. Gen. Physiol.* **1938**, *21*, 501–540. [[CrossRef](#)] [[PubMed](#)]
9. Jirgensons, B. Optical rotation and viscosity of native and denatured proteins. X. Further studies on optical rotatory dispersion. *Arch. Biochem. Biophys.* **1958**, *74*, 57–69. [[CrossRef](#)]
10. Arnone, A.; Bier, C.J.; Cotton, F.A.; Day, V.W.; Hazen, E.E.; Richardson, D.C.; Yonath, A.; Richardson, J.S. A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem.* **1971**, *246*, 2302–2316. [[PubMed](#)]
11. Schweers, O.; Schönbrunn-Hanebeck, E.; Marx, A.; Mandelkow, E. Structural studies of tau protein and alzheimer paired helical filaments show no evidence for  $\beta$ -structure. *J. Biol. Chem.* **1994**, *269*, 24290–24297. [[PubMed](#)]
12. Weinreb, P.H.; Zhen, W.; Poon, A.W.; Conway, K.A.; Lansbury, P.T. Nacp, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry* **1996**, *35*, 13709–13715. [[CrossRef](#)] [[PubMed](#)]
13. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331. [[CrossRef](#)] [[PubMed](#)]
14. Bailey, R.W.; Dunker, A.K.; Brown, C.J.; Garner, E.C.; Griswold, M.D. Clusterin, a binding protein with a molten globule-like region. *Biochemistry* **2001**, *40*, 11828–11840. [[CrossRef](#)] [[PubMed](#)]
15. Liu, J.; Tan, H.; Rost, B. Loopy proteins appear conserved in evolution. *J. Mol. Biol.* **2002**, *322*, 53–64. [[CrossRef](#)]
16. Pullen, R.A.; Jenkins, J.A.; Tickle, I.J.; Wood, S.P.; Blundell, T.L. 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.* **1975**, *8*, 5–20. [[CrossRef](#)] [[PubMed](#)]

17. Cary, P.D.; Moss, T.; Bradbury, E.M. High-resolution proton-magnetic-resonance studies of chromatin core particles. *Eur. J. Biochem.* **1978**, *89*, 475–482. [[CrossRef](#)] [[PubMed](#)]
18. DeForte, S.; Uversky, V.N. Intrinsically disordered proteins in pubmed: What can the tip of the iceberg tell us about what lies below? *RSC Adv.* **2016**, *6*, 11513–11521. [[CrossRef](#)]
19. Tompa, P. *Structure and Function of Intrinsically Disordered Proteins*; CRC Press: Boca Raton, FL, USA, 2010; p. 331.
20. Uversky, V.N. Introduction to intrinsically disordered proteins (IDPs). *Chem. Rev.* **2014**, *114*, 6557–6560. [[CrossRef](#)] [[PubMed](#)]
21. Thomas, W.H.; Weser, U.; Hempel, K. Conformational changes induced by ionic strength and pH in two bovine myelin basic proteins. *Hoppe Seyler's Z Physiol. Chem.* **1977**, *358*, 1345–1352. [[CrossRef](#)] [[PubMed](#)]
22. Hernández, M.A.; Avila, J.; Andreu, J.M. Physicochemical characterization of the heat-stable microtubule-associated protein map2. *Eur J. Biochem.* **1986**, *154*, 41–48. [[CrossRef](#)] [[PubMed](#)]
23. Dunker, A.K.; Oldfield, C.J. Back to the future: Nuclear magnetic resonance and bioinformatics studies on intrinsically disordered proteins. *Adv. Exp. Med. Biol.* **2015**, *870*, 1–34. [[PubMed](#)]
24. Jakob, U.; Kriwacki, R.; Uversky, V.N. Conditionally and transiently disordered proteins: Awakening cryptic disorder to regulate protein function. *Chem. Rev.* **2014**, *114*, 6779–6805. [[CrossRef](#)] [[PubMed](#)]
25. Hu, X.; Stebbins, C.E. Dynamics of the WPD loop of the Yersinia protein tyrosine phosphatase. *Biophys. J.* **2006**, *91*, 948–956. [[CrossRef](#)] [[PubMed](#)]
26. Vacic, V.; Oldfield, C.J.; Mohan, A.; Radivojac, P.; Cortese, M.S.; Uversky, V.N.; Dunker, A.K. Characterization of molecular recognition features, morfs, and their binding partners. *J. Proteome Res.* **2007**, *6*, 2351–2366. [[CrossRef](#)] [[PubMed](#)]
27. Roth, A.F.; Papanayotou, I.; Davis, N.G. The yeast kinase Yck2 has a tripartite palmitoylation signal. *Mol. Biol. Cell.* **2011**, *22*, 2702–2715. [[CrossRef](#)] [[PubMed](#)]
28. Takayama, Y.; Nakasako, M.; Okajima, K.; Iwata, A.; Kashojiya, S.; Matsui, Y.; Tokutomi, S. Light-induced movement of the LOV2 domain in an Asp720asn mutant LOV2-kinase fragment of arabidopsis phototropin 2. *Biochemistry* **2011**, *50*, 1174–1183. [[CrossRef](#)] [[PubMed](#)]
29. Bentrop, D.; Beyermann, M.; Wissmann, R.; Fakler, B. NMR structure of the “Ball-and-chain” Domain of KCNMB2, the  $\beta_2$ -subunit of large conductance  $\text{Ca}^{2+}$ - and voltage-activated potassium channels. *J. Biol. Chem.* **2001**, *276*, 42116–42121. [[CrossRef](#)] [[PubMed](#)]
30. Patel, S.S.; Belmont, B.J.; Sante, J.M.; Rexach, M.F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell* **2007**, *129*, 83–96. [[CrossRef](#)] [[PubMed](#)]
31. Iakoucheva, L.M.; Radivojac, P.; Brown, C.J.; O'Connor, T.R.; Sikes, J.G.; Obradovic, Z.; Dunker, A.K. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* **2004**, *32*, 1037–1049. [[CrossRef](#)] [[PubMed](#)]
32. Buffa, P.; Manzella, L.; Consoli, M.L.; Messina, A.; Vigneri, P. Modelling of the ABL and ARG proteins predicts two functionally critical regions that are natively unfolded. *Proteins* **2007**, *67*, 1–11. [[CrossRef](#)] [[PubMed](#)]
33. Manalan, A.S.; Klee, C.B. Activation of calcineurin by limited proteolysis. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 4291–4295. [[CrossRef](#)] [[PubMed](#)]
34. Mukerjee, N.; McGinnis, K.M.; Park, Y.H.; Gnegy, M.E.; Wang, K.K. Caspase-mediated proteolytic activation of calcineurin in thapsigargin-mediated apoptosis in SH-SY5Y neuroblastoma cells. *Arch. Biochem. Biophys.* **2000**, *379*, 337–343. [[CrossRef](#)] [[PubMed](#)]
35. Alphey, M.S.; Yu, W.; Byres, E.; Li, D.; Hunter, W.N. Structure and reactivity of human mitochondrial 2,4-dienoyl-CoA reductase: Enzyme-ligand interactions in a distinctive short-chain reductase active site. *J. Biol. Chem.* **2005**, *280*, 3068–3077. [[CrossRef](#)] [[PubMed](#)]
36. VanOudenhove, J.; Anderson, E.; Krueger, S.; Cole, J.L. Analysis of PKR structure by small-angle scattering. *J. Mol. Biol.* **2009**, *387*, 910–920. [[CrossRef](#)] [[PubMed](#)]
37. Bhowmick, P.; Pancsa, R.; Guharoy, M.; Tompa, P. Functional diversity and structural disorder in the human ubiquitination pathway. *PLoS ONE* **2013**, *8*, e65443. [[CrossRef](#)] [[PubMed](#)]
38. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell. Biol.* **2015**, *16*, 18–29. [[CrossRef](#)] [[PubMed](#)]
39. Kathiriyai, J.J.; Pathak, R.R.; Clayman, E.; Xue, B.; Uversky, V.N.; Davé, V. Presence and utility of intrinsically disordered regions in kinases. *Mol. Biosyst.* **2014**, *10*, 2876–2888. [[CrossRef](#)] [[PubMed](#)]

40. Banci, L.; Bertini, I.; Cefaro, C.; Ciofi-Baffoni, S.; Gajda, K.; Felli, I.C.; Gallo, A.; Pavelkova, A.; Kallergi, E.; Andreadaki, M.; et al. An intrinsically disordered domain has a dual function coupled to compartment-dependent redox control. *J. Mol. Biol.* **2013**, *425*, 594–608. [[CrossRef](#)] [[PubMed](#)]
41. Bornet, O.; Nouailler, M.; Feracci, M.; Sebban-Kreuzer, C.; Byrne, D.; Halimi, H.; Morelli, X.; Badache, A.; Guerlesquin, F. Identification of a Src kinase SH3 binding site in the C-terminal domain of the human ErbB2 receptor tyrosine kinase. *FEBS Lett.* **2014**, *588*, 2031–2036. [[CrossRef](#)] [[PubMed](#)]
42. Niklas, K.J.; Bondos, S.E.; Dunker, A.K.; Newman, S.A. Rethinking gene regulatory networks in light of alternative splicing, intrinsically disordered protein domains, and post-translational modifications. *Front. Cell. Dev. Biol.* **2015**, *3*, 8. [[CrossRef](#)] [[PubMed](#)]
43. Erales, J.; Coffino, P. Ubiquitin-independent proteasomal degradation. *Biochim. Biophys. Acta* **2014**, *1843*, 216–221. [[CrossRef](#)] [[PubMed](#)]
44. Bah, A.; Vernon, R.M.; Siddiqui, Z.; Krzeminski, M.; Muhandiram, R.; Zhao, C.; Sonenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109. [[CrossRef](#)] [[PubMed](#)]
45. Larion, M.; Salinas, R.K.; Bruschweiler-Li, L.; Miller, B.G.; Brüschweiler, R. Order-disorder transitions govern kinetic cooperativity and allostery of monomeric human glucokinase. *PLoS Biol.* **2012**, *10*, e1001452. [[CrossRef](#)] [[PubMed](#)]
46. Fredrickson, E.K.; Clowes Candadai, S.V.; Tam, C.H.; Gardner, R.G. Means of self-preservation: How an intrinsically disordered ubiquitin-protein ligase averts self-destruction. *Mol. Biol. Cell.* **2013**, *24*, 1041–1052. [[CrossRef](#)] [[PubMed](#)]
47. Hou, Y.; Lin, S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: Gene content estimation for dinoflagellate genomes. *PLoS ONE* **2009**, *4*, e6978. [[CrossRef](#)] [[PubMed](#)]
48. Dunker, A.K.; Bondos, S.E.; Huang, F.; Oldfield, C.J. Intrinsically disordered proteins and multicellular organisms. *Semin Cell. Dev. Biol.* **2015**, *37*, 44–55. [[CrossRef](#)] [[PubMed](#)]
49. Pancsa, R.; Tompa, P. Structural disorder in eukaryotes. *PLoS ONE* **2012**, *7*, e34687. [[CrossRef](#)] [[PubMed](#)]
50. Yu, J.F.; Cao, Z.; Yang, Y.; Wang, C.L.; Su, Z.D.; Zhao, Y.W.; Wang, J.H.; Zhou, Y. Natural protein sequences are more intrinsically disordered than random sequences. *Cell. Mol. Life Sci.* **2016**, *73*, 2949–2957. [[CrossRef](#)] [[PubMed](#)]
51. Teraguchi, S.; Patil, A.; Standley, D.M. Intrinsically disordered domains deviate significantly from random sequences in mammalian proteins. *BMC Bioinform.* **2010**, *11* (Suppl. 7), S7.
52. Brown, C.J.; Takayama, S.; Campen, A.M.; Vise, P.; Marshall, T.W.; Oldfield, C.J.; Williams, C.J.; Dunker, A.K. Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.* **2002**, *55*, 104–110. [[CrossRef](#)] [[PubMed](#)]
53. Potenza, E.; di Domenico, T.; Walsh, I.; Tosatto, S.C. Mobidb 2.0: An improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.* **2015**, *43*, D315–D320. [[CrossRef](#)] [[PubMed](#)]
54. Ota, M.; Koike, R.; Amemiya, T.; Tenno, T.; Romero, P.R.; Hiroaki, H.; Dunker, A.K.; Fukuchi, S. An assignment of intrinsically disordered regions of proteins based on nmr structures. *J. Struct. Biol.* **2013**, *181*, 29–36. [[CrossRef](#)] [[PubMed](#)]
55. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W.F. From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* **2013**, *4*, 2741. [[CrossRef](#)] [[PubMed](#)]
56. Uversky, V.N.; Dunker, A.K. *Intrinsically Disordered Protein Analysis*, 1st ed.; Humana Press: New York, NY, USA, 2012; Volume 1.
57. Uversky, V.N.; Dunker, A.K. *Intrinsically Disordered Protein Analysis*, 1st ed.; Humana Press: New York, NY, USA, 2012; Volume 2.
58. Uversky, V.N.; Dunker, A.K. Multiparametric analysis of intrinsically disordered proteins: Looking at intrinsic disorder through compound eyes. *Anal. Chem.* **2012**, *84*, 2096–2104. [[CrossRef](#)] [[PubMed](#)]
59. Uversky, V.N. Biophysical methods to investigate intrinsically disordered proteins: Avoiding an “Elephant and blind men” Situation. *Adv. Exp. Med. Biol.* **2015**, *870*, 215–260. [[PubMed](#)]
60. Khan, T.; Douglas, G.M.; Patel, P.; Nguyen Ba, A.N.; Moses, A.M. Polymorphism analysis reveals reduced negative selection and elevated rate of insertions and deletions in intrinsically disordered protein regions. *Genome Biol. Evol.* **2015**, *7*, 1815–1826. [[CrossRef](#)] [[PubMed](#)]
61. Smithers, B.; Oates, M.E.; Gough, J. Splice junctions are constrained by protein disorder. *Nucleic Acids Res.* **2015**, *43*, 4814–4822. [[CrossRef](#)] [[PubMed](#)]

62. Bhowmick, P.; Guharoy, M.; Tompa, P. Bioinformatics approaches for predicting disordered protein motifs. *Adv. Exp. Med. Biol.* **2015**, *870*, 291–318. [[PubMed](#)]
63. Varadi, M.; Vranken, W.; Guharoy, M.; Tompa, P. Computational approaches for inferring the functions of intrinsically disordered proteins. *Front. Mol. Biosci.* **2015**, *2*, 45. [[CrossRef](#)] [[PubMed](#)]
64. Punta, M.; Simon, I.; Dosztányi, Z. Prediction and analysis of intrinsically disordered proteins. *Methods Mol. Biol.* **2015**, *1261*, 35–59. [[PubMed](#)]
65. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [[CrossRef](#)]
66. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically disordered protein. *J. Mol. Graph. Model.* **2001**, *19*, 26–59. [[CrossRef](#)]
67. Williams, R.M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E.C.; Young, J.; Takayama, S.; Brown, C.J.; Dunker, A.K. The protein non-folding problem: Amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.* **2001**, 89–100.
68. Vihinen, M.; Torkkila, E.; Riikonen, P. Accuracy of protein flexibility predictions. *Proteins* **1994**, *19*, 141–149. [[CrossRef](#)] [[PubMed](#)]
69. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132. [[CrossRef](#)]
70. Kumari, B.; Kumar, R.; Kumar, M. Low complexity and disordered regions of proteins have different structural and amino acid preferences. *Mol. Biosyst.* **2015**, *11*, 585–594. [[CrossRef](#)] [[PubMed](#)]
71. Chen, J.W.; Romero, P.; Uversky, V.N.; Dunker, A.K. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J. Proteome Res.* **2006**, *5*, 879–887. [[CrossRef](#)] [[PubMed](#)]
72. Moesa, H.A.; Wakabayashi, S.; Nakai, K.; Patil, A. Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification. *Mol. Biosyst.* **2012**, *8*, 3262–3273. [[CrossRef](#)] [[PubMed](#)]
73. Li, J.; Feng, Y.; Wang, X.; Liu, W.; Rong, L.; Bao, J. An overview of predictors for intrinsically disordered proteins over 2010–2014. *Int. J. Mol. Sci.* **2015**, *16*, 23446–23462. [[CrossRef](#)] [[PubMed](#)]
74. He, B.; Wang, K.; Liu, Y.; Xue, B.; Uversky, V.N.; Dunker, A.K. Predicting intrinsic disorder in proteins: An overview. *Cell. Res.* **2009**, *19*, 929–949. [[CrossRef](#)] [[PubMed](#)]
75. Monastyrskyy, B.; Kryshchuk, A.; Moulton, J.; Tramontano, A.; Fidelis, K. Assessment of protein disorder region predictions in casp10. *Proteins* **2014**, *82* (Suppl. 2), 127–137. [[CrossRef](#)] [[PubMed](#)]
76. Ishida, T.; Kinoshita, K. Prdos: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **2007**, *35*, W460–W464. [[CrossRef](#)] [[PubMed](#)]
77. Zhang, T.; Faraggi, E.; Xue, B.; Dunker, A.K.; Uversky, V.N.; Zhou, Y. Spine-d: Accurate prediction of short and long disordered regions by a single neural-network based method. *J. Biomol. Struct. Dyn.* **2012**, *29*, 799–813. [[CrossRef](#)] [[PubMed](#)]
78. Jones, D.T.; Cozzetto, D. Disopred3: Precise disordered region predictions with annotated protein-binding activity. *Bioinform.* **2015**, *31*, 857–863. [[CrossRef](#)] [[PubMed](#)]
79. Walsh, I.; Martin, A.J.; di Domenico, T.; Tosatto, S.C. Espritz: Accurate and fast prediction of protein disorder. *Bioinform.* **2012**, *28*, 503–509. [[CrossRef](#)] [[PubMed](#)]
80. Dosztányi, Z.; Csizsók, V.; Tompa, P.; Simon, I. Iupred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinform.* **2005**, *21*, 3433–3434. [[CrossRef](#)] [[PubMed](#)]
81. Dosztányi, Z.; Csizsók, V.; Tompa, P.; Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **2005**, *347*, 827–839. [[CrossRef](#)] [[PubMed](#)]
82. Sickmeier, M.; Hamilton, J.A.; LeGall, T.; Vacic, V.; Cortese, M.S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V.N.; et al. Disprot: The database of disordered proteins. *Nucleic Acids Res.* **2007**, *35*, D786–D793. [[CrossRef](#)] [[PubMed](#)]
83. Touw, W.G.; Baakman, C.; Black, J.; te Beek, T.A.; Krieger, E.; Joosten, R.P.; Vriend, G. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **2015**, *43*, D364–D368. [[CrossRef](#)] [[PubMed](#)]
84. Peng, Z.; Mizianty, M.J.; Kurgan, L. Genome-scale prediction of proteins with long intrinsically disordered regions. *Proteins* **2014**, *82*, 145–158. [[CrossRef](#)] [[PubMed](#)]

85. Yan, J.; Mizianty, M.J.; Filipow, P.L.; Uversky, V.N.; Kurgan, L. Rapid: Fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. *Biochim. Biophys. Acta* **2013**, *1834*, 1671–1680. [[CrossRef](#)] [[PubMed](#)]
86. Cilia, E.; Pancsa, R.; Tompa, P.; Lenaerts, T.; Vranken, W.F. The dynamine webserver: Predicting protein dynamics from sequence. *Nucleic Acids Res.* **2014**, *42*, W264–W270. [[CrossRef](#)] [[PubMed](#)]
87. Xue, B.; Dunbrack, R.L.; Williams, R.W.; Dunker, A.K.; Uversky, V.N. Ponder-fit: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* **2010**, *1804*, 996–1010. [[CrossRef](#)] [[PubMed](#)]
88. Kozłowski, L.P.; Bujnicki, J.M. MetaDisorder: A meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinform.* **2012**, *13*, 111. [[CrossRef](#)] [[PubMed](#)]
89. Fan, X.; Kurgan, L. Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J. Biomol. Struct. Dyn.* **2014**, *32*, 448–464. [[CrossRef](#)] [[PubMed](#)]
90. Vucetic, S.; Brown, C.J.; Dunker, A.K.; Obradovic, Z. Flavors of protein disorder. *Proteins* **2003**, *52*, 573–584. [[CrossRef](#)] [[PubMed](#)]
91. Cozzetto, D.; Jones, D.T. The contribution of intrinsic disorder prediction to the elucidation of protein function. *Curr. Opin. Struct. Biol.* **2013**, *23*, 467–472. [[CrossRef](#)] [[PubMed](#)]
92. Consortium, G.O. Gene ontology consortium: Going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [[CrossRef](#)] [[PubMed](#)]
93. Bairoch, A. The enzyme database in 2000. *Nucleic Acids Res.* **2000**, *28*, 304–305. [[CrossRef](#)] [[PubMed](#)]
94. Peng, Z.; Yan, J.; Fan, X.; Mizianty, M.J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V.N.; Kurgan, L. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **2015**, *72*, 137–151. [[CrossRef](#)] [[PubMed](#)]
95. Ward, J.J.; Sodhi, J.S.; McGuffin, L.J.; Buxton, B.F.; Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **2004**, *337*, 635–645. [[CrossRef](#)] [[PubMed](#)]
96. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the d2 concept. *Annu. Rev. Biophys.* **2008**, *37*, 215–246. [[CrossRef](#)] [[PubMed](#)]
97. Gsponer, J.; Futschik, M.E.; Teichmann, S.A.; Babu, M.M. Tight regulation of unstructured proteins: From transcript synthesis to protein degradation. *Science* **2008**, *322*, 1365–1368. [[CrossRef](#)] [[PubMed](#)]
98. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochim. Biophys. Acta* **2010**, *1804*, 1231–1264. [[CrossRef](#)] [[PubMed](#)]
99. Uversky, V.N. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front. Aging Neurosci.* **2015**, *7*, 18. [[CrossRef](#)] [[PubMed](#)]
100. Uversky, V.N.; Davé, V.; Iakoucheva, L.M.; Malaney, P.; Metallo, S.J.; Pathak, R.R.; Joerger, A.C. Pathological unfoldomics of uncontrolled chaos: Intrinsically disordered proteins and human diseases. *Chem. Rev.* **2014**, *114*, 6844–6879. [[CrossRef](#)] [[PubMed](#)]
101. Sotomayor-Pérez, A.C.; Ladant, D.; Chenal, A. Disorder-to-order transition in the CyaA toxin RTX domain: Implications for toxin secretion. *Toxins (Basel)* **2015**, *7*, 1–20. [[CrossRef](#)] [[PubMed](#)]
102. Berson, A.; Soreq, H. It all starts at the ends: Multifaceted involvement of C- and N-terminally modified cholinesterases in Alzheimer's disease. *Rambam Maimonides Med. J.* **2010**, *1*, e0014. [[CrossRef](#)] [[PubMed](#)]
103. Kardani, J.; Roy, I. Understanding caffeine's role in attenuating the toxicity of  $\alpha$ -synuclein aggregates: Implications for risk of Parkinson's disease. *ACS Chem. Neurosci.* **2015**, *6*, 1613–1625. [[CrossRef](#)] [[PubMed](#)]
104. Krishnan, N.; Koveal, D.; Miller, D.H.; Xue, B.; Akshinthala, S.D.; Kragelj, J.; Jensen, M.R.; Gauss, C.M.; Page, R.; Blackledge, M.; et al. Targeting the disordered c terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.* **2014**, *10*, 558–566. [[CrossRef](#)] [[PubMed](#)]
105. Tonks, N.K.; Diltz, C.D.; Fischer, E.H. Purification of the major protein-tyrosine-phosphatases of human placenta. *J. Biol. Chem.* **1988**, *263*, 6722–6730. [[PubMed](#)]
106. Chernoff, J.; Schievella, A.R.; Jost, C.A.; Erikson, R.L.; Neel, B.G. Cloning of a cDNA for a major human protein-tyrosine-phosphatase. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2735–2739. [[CrossRef](#)] [[PubMed](#)]
107. Hao, L.; Tiganis, T.; Tonks, N.K.; Charbonneau, H. The noncatalytic C-terminal segment of the T cell protein tyrosine phosphatase regulates activity via an intramolecular mechanism. *J. Biol. Chem.* **1997**, *272*, 29322–29329. [[CrossRef](#)] [[PubMed](#)]
108. Tonks, N.K. Protein tyrosine phosphatases—From housekeeping enzymes to master regulators of signal transduction. *FEBS J.* **2013**, *280*, 346–378. [[CrossRef](#)] [[PubMed](#)]

109. Anderson, J.N.; Tonks, N.K. Protein tyrosine phosphatase-based therapeutics: Lessons from ptp1b. *Top. Curr. Genet.* **2004**, *5*, 201–230.
110. Lantz, K.A.; Hart, S.G.; Planey, S.L.; Roitman, M.F.; Ruiz-White, I.A.; Wolfe, H.R.; McLane, M.P. Inhibition of PTP1B by trodusquemine (MSI-1436) causes fat-specific weight loss in diet-induced obese mice. *Obesity (Silver Spring)* **2010**, *18*, 1516–1523. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).