IDPology of the Living Cell: Intrinsic Disorder in the Subcellular Compartments of the Human Cell

Bi Zhao[1], Akila Katuwawala[1], Vladimir N. Uversky[2,3]* and Lukasz Kurgan[1]*
[1]Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
[2]Department of Molecular Medicine, USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL, USA
[3]Laboratory of New Methods in Biology, Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia

*corresponding authors:
Lukasz Kurgan: Department of Computer Science, Virginia Commonwealth University, 401 West Main Street, Room E4225, Richmond, VA 23284, USA; Email: lkurgan@vcu.edu; Phone: (804) 827-3986
Vladimir N. Uversky: Department of Molecular Medicine, University of South Florida, 12901 Bruce B. Downs Blvd. MDC07, Tampa, FL 33612, USA; Email: vuversky@usf.edu; Phone: (813) 974-5816

Abstract

Intrinsic disorder can be found in all proteomes of all kingdoms of life and in viruses, being particularly prevalent in the eukaryotes. We conduct a comprehensive analysis of the intrinsic disorder in the human proteins while mapping them into 24 compartments of the human cell. In agreement with previous studies, we show that human proteins are significantly enriched in disorder relative to a generic protein set that represents the protein universe. In fact, the fraction of proteins with long disordered regions and the average protein-level disorder content in the human proteome are about 3 times higher than in the protein universe. Furthermore, levels of intrinsic disorder in majority of human subcellular compartments significantly exceed the average disorder content in the protein universe. Relative to the overall amount of disorder in the human proteome, proteins in localized in nucleus and cytoskeleton have significantly increased amounts of disorder, measured by both high disorder content and presence of multiple long intrinsically disordered regions. We empirically demonstrate that, on average, human proteins are assigned to 2.3 subcellular compartments, with proteins localized to few subcellular compartments being more disordered than the proteins that are localized to many compartments. Functionally, the disordered proteins localized in the most disorder-enriched subcellular compartments are primarily responsible for interactions with nucleic acids and protein partners. This is the first-time disorder is comprehensively mapped into the human cell. Our observations add a missing piece to the puzzle of functional disorder and its organization inside the cell.

Keywords
Intrinsic disorder; subcellular location; human cell; intrinsically disordered proteins.

Funding
This research was supported in part by the National Science Foundation (grant 1617369) and the Robert J. Mattauch Endowment funds.

Conflicts of interest/Competing interests
The authors declare no conflicts of interests.

Availability of data and material
Data are available in the Supplement.

1 Introduction

Research suggests that protein universe, defined as the collection of all proteins of all organisms [1], includes two distinct subspaces composed of structured, mostly globular proteins [2-5] and disordered proteins [6-9]. The intrinsically disordered proteins (IDPs) or hybrid proteins that have ordered domains and intrinsically disordered protein regions (IDPRs) [10-13] are commonly found in all proteomes of all kingdoms of life and in viruses [14-18]. This high penetrance of intrinsic disorder in the protein universe stems from unique functional features of IDPs/IDPRs related to their high structural plasticity and lack of unique stable structure [10, 11, 19]. Structurally, these 'floppy' proteins and protein regions are described as highly dynamic ensembles of rapidly interconverting conformations [20, 21]. They complement the functional repertoire associated with the structured proteins by expanding the structure/disorder-to-function continuum concept [22, 23]. Being mostly incompatible with catalytic and transport functions, which are dominated by the structured proteins, IDPs/IDPRs are multifunctional entities with high binding promiscuity that is crucial for signaling, regulation, and recognition [10, 11, 21, 24-31]. Moreover, intrinsic disorder is interlinked with the pathogenesis of various human diseases. In fact, the majority of human cancer-associated proteins [32], as well as many proteins associated with neurodegeneration [33], diabetes [34], cardiovascular disease [35], and amyloidosis [36], are either intrinsically disordered or contain long IDPRs, giving raise to the D^2 (disorder in disorders) concept [37, 38].

It is recognized now that different organisms possess different levels of intrinsic disorder, where the prevalence of disorder increases with the increase in the organism complexity [14-17, 39]. In particular, several studies have shown that eukaryotic organisms are enriched in disorder when compared to prokaryotes [14, 16-18, 40-42]. Moreover, multiple works have analyzed abundance and specific functional features of the intrinsic disorder in the human proteome [43-49]. In spite of this substantial amount of interest, one of the key understudied topics is the distribution of the intrinsic disorder across the human cell. The subcellular locations of proteins are related to their function and knowledge of this relationship is essential to decipher the intricacies of the molecular machinery in the cell. The past studies that investigate the distribution of the proteins across subcellular compartments in the human cell excluded the aspect of the intrinsic disorder [50, 51]. However, it is clear that the disorder is not uniformly distributed over the cell. While some IDPs and proteins with IDPRs were shown to be specifically compartmentalized within selected organelles [52, 53], the peculiarities of the disorder distribution across various organelles in the living cell are poorly understood. The arguably closest study on this topic analyzes distribution of disorder in cellular compartment as a secondary objective while primarily characterizing abundance of disorder across the organismal taxonomy [14]. Consequently, this analysis covers only 16 subcellular compartments, quantifies the amount of disorder with a single measure, focuses solely on characterizing compartments enriched in disorder and contextualizes this enrichment against one baseline of specific taxonomic domain of life.

To fill this gap, we are reporting here the results of a large-scale computational analysis that comprehensively maps and functionally characterizes intrinsic disorder in the subcellular compartments of the human cell. We map the human proteome into 24 subcellular compartments, quantify disorder amount with multiple measures, investigate both enrichment and depletion of disorder across these compartments by using two key baselines, the human proteome and the protein universe. Beside these marked improvements, we incorporate several novel aspects including analysis of a relation between disorder and multi-compartment localization and examination of functional differences between disordered-enriched proteins localized in the disorder-enriched compartments vs. structured proteins localized in the disorder-depleted compartments. We also evaluate impact of using predictions by contrasting results secured using a more complete set of compartment annotations that rely on a combination of experimental and predicted data vs. solely experimental annotations.

2 Materials and Methods

We map the human proteins collected from UniProt into a comprehensive set of subcellular compartments, annotate them with putative intrinsic disorder, and perform detailed statistical analysis of the prevalence, sequence-level characteristics, and molecular functions of the intrinsically disordered and ordered proteins.

2.1 Mapping of Subcellular Compartments

We use the Gene Ontology (GO) annotation from the Gene Ontology Consortium [54, 55] to annotate cellular components of human proteins, which can be used to map proteins to the subcellular compartment. Total of 18,477 human proteins from the reference human proteins in UniProt [56] had cellular component GO terms. Furthermore, to improve coverage we supplement the GO-derived cellular component annotations with the experimentally-derived GO terms that we collect from the recently established COMPARTMENTS database [57]. This resulted in the availability of the cellular components GO terms for 19,797 human proteins.

We cover a comprehensive set of 24 subcellular compartments that cover all major organelles and membrane types in the human cell. They include cytoplasm (GO:0005737), nucleus (GO:0005694), nuclear membrane (GO:0031965), cell membrane (GO:0005886), cell junction (GO:0030054), cell projection (GO:0042995), cytoskeleton (GO:0005856), mitochondrion (GO:0005739), mitochondrial membrane (GO:0031966), endoplasmic reticulum (GO:0005783), endoplasmic reticulum membrane (GO:0005789), endosome (GO:0005768), endosome membrane (GO:0010008), Golgi apparatus (GO:0005794), Golgi membrane (GO:0000139), centrosome (GO:0005813), vacuole (GO:0005773), vacuolar membrane (GO:0005774), lysosome (GO:0005764), lysosomal membrane (GO:0005765), ribosome (GO:0005840), peroxisome (GO:0005777), peroxisomal membrane (GO:0005778), and other membrane (GO:0016020). This is comparable to the previous mapping efforts that considered 11 [57], 16 [50] and 30 [51] compartments in the human cell.

We map the 19,797 human proteins into these 24 subcellular compartments using their GO cellular component terms based on the "is_a" and "part_of" relations in the hierarchal structure of GO ontology. We categoriz the mapped cellular compartment annotations for a specific protein as either experimental or predicted. The experimental annotations require at least one GO term to be annotated experimentally in the GO database (keywords: Experiment, Direct assay, Physical interaction, Mutant phenotype, Genetic interaction and Expression pattern) or in the COMPARTMENTS database. The predicted annotations rely solely on the GO terms that were annotated via alignment/prediction in the GO database.

2.2 Annotation of Intrinsic Disorder

We investigate the feasibility of using the experimentally annotated intrinsic disorder in our analysis. We use version 3 of the MobiDB database [58], which provides access to a comprehensive collection of the experimentally-derived intrinsic disorder annotations. MobiDB aggregates experimental disorder data that is curated across multiple sources that include disorder databases, such as DisProt [59, 60] and FuzDB [61], as well as other databases where disorder information can be extracted indirectly, including Protein Data Bank (PDB) [62, 63] and Biological Magnetic Resonance Data Bank (BMRB) [64]. Using MobiDB 3 we collect the experimental disorder annotations for 6,029 human proteins that we previously mapped into subcellular compartments. This corresponds to approximately 30% coverage of our dataset. We test whether this coverage is adequate to uniformly represent the 24 compartments. **Supplementary Table S1** compares the actual coverage of the compartments by the experimentally annotated proteins with the corresponding confidence intervals of the expected coverage values. The coverage meets the expected range for only 8 out of the 24 compartments (33%), while in the other cases, it is either significantly

higher (9 times) or significantly lower (7 times). We conclude that at this point we cannot use of the experimentally annotated disorder due to the relatively low overall coverage (30% of proteins) and uneven distribution across compartments (only 33% are covered proportionally).

A feasible alternative is to use putative disorder, which can be obtained for virtually all proteins in our dataset, ensuring complete and even coverage across the subcellular compartments. There are over 60 predictors of intrinsic disorder [65-68]. Multiple recent studies have shown that some of these predictors provide high quality results [13, 69-74]. More specifically, the most accurate predictors achieve the Area Under the ROC Curve (AUC) values, which range between 0.5 and 1, at around 0.90 [13, 72]. Furthermore, research shows that consensus-based predictors, defined as the methods that combine multiple "base" predictors of disorder, generate particularly accurate predictions [75-79]. Correspondingly, we use a recently developed consensus predictor, MobiDB-lite [77]. This method combines predictions produced by several base disorder predictors including three versions of ESpritz [80], two versions of IUpred [81, 82], DisEMBL [83], RONN [84], PONDR VSL2B [85] and GlobPlot [86]. Results produced by MobiDB-lite are used across several popular resources including MobiDB [58], UniProt [87], InterPro [88], and SIFTS [89]. We also emphasize that multiple studies have used putative annotations of disorder to characterize the abundance and functional features of disorder. For instance, disorder predictions were used to map disorder into intra-nuclear compartments [90], to characterize dark proteomes [7, 91], to link GC content and disorder enrichment [92], to analyze typically unstructured cysteine-depleted proteins [93], ribosomal proteins [94], histones [95], beta catenin [96] and several viral proteomes [97-102], and to provide functional insights into protein-protein interactions [44, 103, 104], protein-nucleic acids interactions [105, 106], and programmed cell death and autophagy processes [107-109].

We collect the MobiDB-lite's predictions for the human proteins from the MobiDB database. Given that the GO data that we used to map subcellular compartments relies on a newer version of UniProt than the MobiDB resources that releases the MobiDB-lite's predictions, we could not obtain disorder predictions for 1,235 proteins. The resulting dataset of 18,562 human proteins has both the subcellular compartment annotations and high-quality putative disorder data. We share this dataset, including the information about the disorder and the compartment annotations, in the Supplement.

2.3 Representation of the Protein Universe

We contextualize the analysis of the intrinsic disorder in the human proteome by comparing these results against the abundance of the disorder in a protein universe, defined as a set of proteins collected across multiple organisms [1]. We use the collection of the manually reviewed and annotated proteins in the SwissProt database [110] to represent the protein universe. The disorder information for SwissProt, which includes 554,779 proteins from over 9500 organisms, was obtained from MobiDB 3.

2.4 Statistical Analysis

We quantify the amount of disorder in a given protein using two measures, disorder content and presence of long disordered regions (LDRs). The disorder content is computed as the total number of putative disordered amino acids divided the length of the given protein sequence. LDRs are defined as segments of at least 30 consecutive putative disordered amino acids [16, 42]. These segments are recognized as functional disordered domains [16, 47], and many of them were found to be implicated in the protein–protein recognition [111]. We quantify the number of proteins with LDRs as well as the average number of LDRs per protein when evaluating the amount of disorder across subcellular compartments.

We assess significance of differences in the amount of disorder in specific subcellular compartments against two reference sets: the human proteome and the protein universe. We also perform this analysis when comparing disorder among proteins that localize in a few vs. many compartments. To ensure that

these analyses are robust, we randomly subsample the reference sets to select the same number of proteins that have the similar sequence length (with ±10% margin) when compared with a set of proteins in a given subcellular compartment. We similarly subsample the larger set of proteins that localize in the few compartments. This sampling procedure accommodates for the bias in the disorder amount across proteins that have different chain length [14, 112] and is consistent with the similar analyses done in several recent studies [93, 109, 113-115]. We evaluate the significance with the Student *t*-test if the underlying data distribution is normal (as tested with the Anderson-Darling test at 0.01 significance); otherwise, we apply the Wilcoxon signed-rank test. We utilize the Bonferroni Correction (BC) to adjust *p*-value to minimize the type I errors when performing multiple tests [116], i.e., when comparing different compartments against the same reference set.

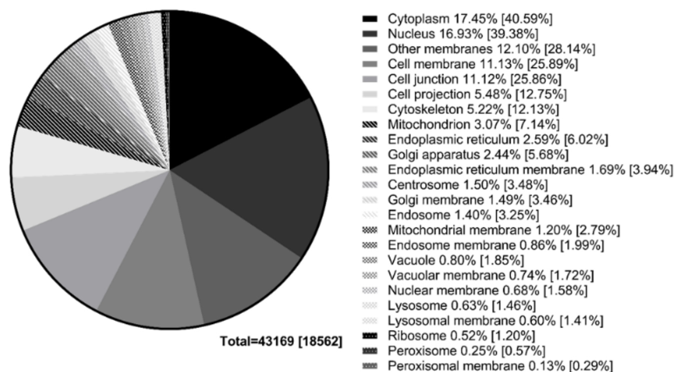

Figure 1. Fraction of proteins in subcellular compartments that were annotated based on the experimental and sequence-similarity derived evidence. The compartments are sorted in the descending order of the fractions. We report two numbers: the fraction among the 43,169 protein-compartment annotation pairs [the fraction among the 18,562 annotated proteins]. The difference is due to the fact that many proteins are associated with multiple subcellular compartments. Supplementary Figure S1 shows the corresponding results where subcellular compartments were annotated based solely on the experimental evidence.

3 Results

3.1 Abundance of the Intrinsic Disorder in the Subcellular Compartments of a Human Cell

Figure 1 summarizes the distribution of the human proteins across the 24 subcellular compartments. We report two numbers, the fraction among the 18,562 compartment-annotated human proteins (inside the square brackets) and the fraction among the 43,169 protein-compartment pairs. Our analysis reveals that on average human proteins are assigned to 43,169/18,562 = 2.3 subcellular compartments. We also provide the same analysis when limiting the protein set to the proteins that were assigned to the subcellular compartment using only the experimental annotations (by excluding sequence-similarity derived GO terms) in **Supplementary Figure S1**. We note that both results are very similar. The two compartments that have the largest number of proteins are the cytoplasm and nucleus, each hosting about 40% of the human proteins. On the other end of the spectrum are the lysosome, ribosome, and peroxisome, which are home for about 1.5%, 1.2%, and 0.6% of the human proteins, respectively.

Table 1 summarizes the number of proteins, the disorder content, and the number and fraction of proteins with the LDRs in the protein universe (represented with Swiss-Prot), the entire human proteome, and for each of the 24 subcellular compartments in the human cell. These measures are defined in Section 2.4. Our analysis suggests that 44.2% of human proteins contain putative LDRs. This result is in close agreement with multiple previous studies that have found a very similar rate of 44% [11, 117] and 44.1% [47]. The fraction of proteins with LDRs and the average protein-level disorder content in the human proteome are about 3 times higher than the disorder content in the protein universe. This is in line with

several past studies that show that eukaryotic organisms are substantially enriched in disorder, when compared with the proteins from the prokaryotic species [14, 16-18, 40-42].

Table 1. Summary of protein disorder annotation in protein universe, human genome, and each cellular compartment. Human subcellular compartments were sorted by the descending average of disorder content per protein in each given cellular compartment. The average disorder content was computed as the average of the protein-level content for a given protein set. The long disordered regions (LDRs) are defined in the literature as the sequence segments with at least 30 consecutive disordered amino acids.

Dataset		No. proteins	Average of protein-level disorder content	No. proteins with LDRs	% of protein with LDRs
Protein universe (Swiss-Prot)		554,779	0.047	74,639	13.5%
Human	Whole proteome	18,562	0.128	8,204	44.2%
	Nucleus	7,309	0.171	4,197	57.4%
	Cytoskeleton	2,252	0.151	1,266	56.2%
	Centrosome	646	0.146	391	60.5%
	Cytoplasm	7,535	0.133	3,691	49.0%
	Nuclear membrane	293	0.133	160	54.6%
	Cell projection	2,366	0.123	1,199	50.7%
	Golgi apparatus	1,054	0.101	467	44.3%
	Cell junction	4,800	0.097	1,733	36.1%
	Ribosome	223	0.096	57	25.6%
	Endosome	1,117	0.096	430	38.5%
	Cell membrane	4,805	0.095	1,952	40.6%
	Endosome membrane	731	0.088	271	37.1%
	Golgi membrane	642	0.085	233	36.3%
	Endoplasmic reticulum membrane	370	0.084	147	39.7%
	Endoplasmic reticulum	604	0.083	257	42.5%
	Other membranes	5,223	0.082	1,857	35.6%
	Vacuole	344	0.079	119	34.6%
	Mitochondrion	1,325	0.070	343	25.9%
	Lysosome	271	0.065	81	29.9%
	Mitochondrial membrane	518	0.059	106	20.5%
	Peroxisomal membrane	54	0.058	12	22.2%
	Vacuolar membrane	320	0.054	96	30.0%
	Lysosomal membrane	261	0.050	76	29.1%
	Peroxisome	106	0.035	20	18.9%

Our analysis reveals that the intrinsic disorder distributes unevenly across the subcellular compartments. In particular, we found that the disorder content in several compartments, such as nucleus, cytoskeleton, centrosome, cytoplasm, and nuclear membrane, is higher than the average disorder content in the human proteome. The most disorder-enriched nucleus has disorder content of 0.171, which is a 33% increase over the overall proteome-level disorder content of 0.128. Literature similarly points to the high amounts of disorder in the nucleus, in human and other eukaryotes [16, 109, 118]. The other 19 compartments are characterized by the lower than expected disorder content ranging between 0.035 (for peroxisome) and 0.123 (for cell projection). We note that while there are relatively few compartments where the amount of disorder is enriched (relative to the overall amount in the human proteome), they account for nearly half of the subcellular compartment annotations. As **Figure 1** reports, the five compartments with the higher-than-expected disorder content account for 17.5% (cytoplasm) + 16.9% (nucleus) + 5.2% (cytoskeleton) +

1.5% (centrosome) + 0.7% (nuclear membrane) = 41.8% of the compartment's annotations. Furthermore, we observe that a substantial fraction of human proteins, anywhere between 18.9% (for perixosome) and 60.5% (for nucleus), include LDRs. To compare, only 13.5% of proteins in the protein universe have LDRs. Importantly, **Table 1** also shows that the vast majority of human subcellular compartments are characterized by the increased disorder content as compared to protein universe. In fact, there is only one compartment (peroxisome) that has disorder content (0.035) lower than that of protein universe (0.047). Moreover, every single subcellular compartment in the human cell contains a larger fraction of proteins with LDRs, ranging between 20.5% and 60.5%, than the protein universe (13.5%).

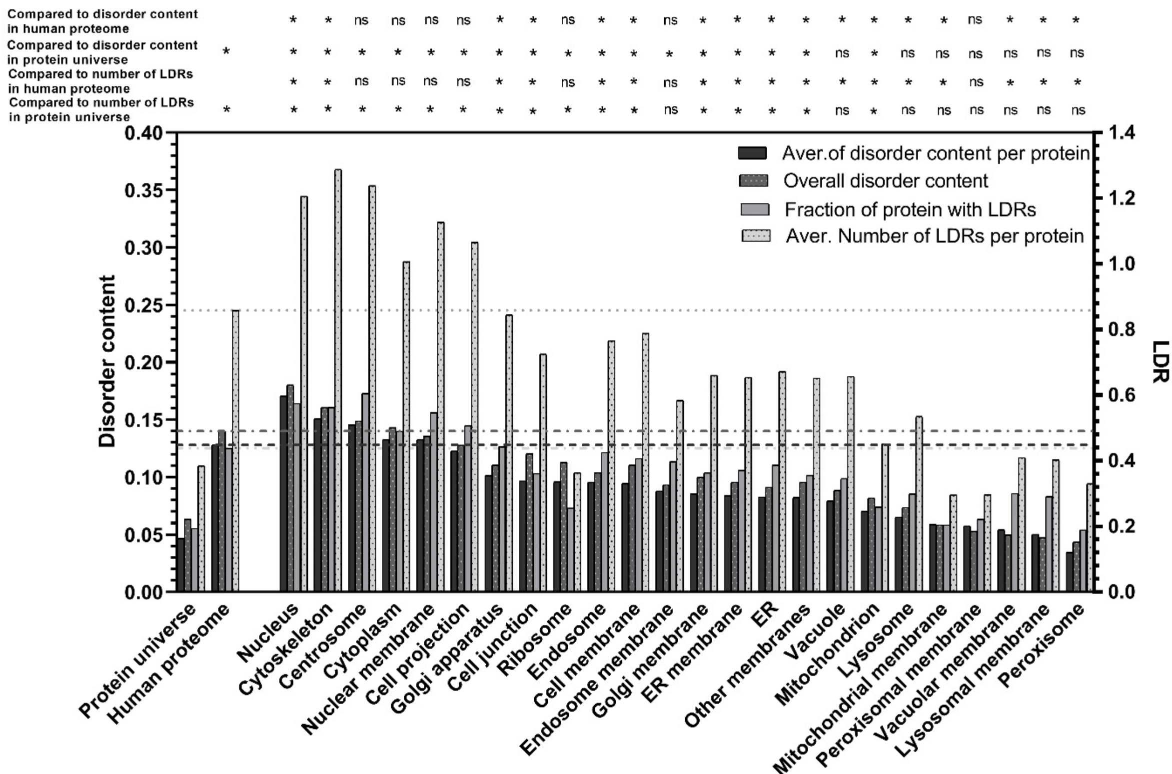

Figure 2. Disorder characteristics in the protein universe, human proteome and across subcellular compartments in the human cell that were annotated based on the experimental and sequence-similarity derived evidence. The left *y*-axis quantifies the disorder content, including the average of the protein-level content and the aggregate, over all proteins, content. The right *y*-axis quantifies the characteristics of the long disordered regions (LDRs) including the fraction of proteins with LDRs and the average number of LDRs per protein. The dashed horizontal lines correspond to the disorder content values (darker shades) and LDR-related characteristics (lighter shades) that were measures on the human proteome. The subcellular compartments are sorted based on their average of the protein-level disorder content. Statistical significance of the differences in the protein-level disorder content and LDR numbers between proteins in a given subcellular compartment and proteins in a given reference set (human proteome and protein universe) was evaluated with the Wilcoxon signed-rank test; the underlying data is not normal. Further details are explained in Section 2.4. The annotations of the significance are shown at the top of the plots where "*" denotes significant difference (p-value < 0.01) and "ns" denotes differences that are not significant.

3.2 Statistical Analysis of the Intrinsic Disorder in the Subcellular Compartments of a Human Cell

Figure 2 summarizes several key characteristics of the intrinsic disorder across the protein universe, human proteome and the 24 subcellular compartments. We analyze the overall disorder content (fraction of disordered residues across all proteins in a given protein set), the average of the protein-level disorder content, the fraction of proteins with LDRs, and the average number of LDRs per protein. Consistent with the observations in **Table 1**, we show that the disorder characteristics are enriched in the human proteome when compared to the protein universe; the differences in the disorder content and LDR numbers are

statistically significant (p-value < 0.01). The plot also highlights substantial differences in the disorder amounts across the subcellular compartments. We observe that the four disorder characteristics are mutually correlated. One exception from this trend is the ribosome, where the LDR-driven characteristics are unusually lower compared to the disorder content values. This stems from the fact that many ribosomal proteins are short, thus their likelihood of presence of LDRs is lower while they may still have substantial disorder content. Overall, the compartments with the low disorder content (e.g., peroxisome, lysosome, mitochondrion, vacuole, and their membranes) are also characterized by relatively low fractions of proteins with LDRs (between 19 and 35%) and low numbers of LDRs per protein (between 0.30 and 0.64). On the other hand, the disorder content-enriched compartments that include nucleus, cytoskeleton, centrosome, cytoplasm, and nuclear membrane, feature between 49 and 60% of the proteins with LDRs and, on average, more than one LDR per protein.

We quantify significance of the differences in the protein-level disorder content and LDR number between proteins in the 24 subcellular compartments and a given reference protein set: human proteome and protein universe. The results are summarized at the top of **Figure 2**. The proteins in the nucleus and cytoskeleton are significantly enriched in the intrinsic disorder measured with disorder content and number of LDRs when compared to the human proteome (p-value < 0.01). On the other end of the spectrum, proteins in 15 subcellular compartments (Golgi apparatus, endoplasmic reticulum, vacuole, mitochondrion, lysosome and their membranes, endosome, peroxisome, cell membrane, and cell junction) are significantly depleted in disorder content and number of LDRs relative to the human proteome (p-value < 0.01). While the number of the significantly disorder-enriched compartments is rather small, they cover 8,562 proteins (46%), while the 15 disorder-depleted compartments account for 12,153 proteins (65%); the sum is over 100% since majority of proteins are localized in multiple compartments. Furthermore, proteins in all but seven subcellular compartments (these exceptions include vacuole, lysosome, peroxisome, and their membranes, and mitochondrial membrane) are significantly enriched in disorder content when compared to the protein universe (p-value < 0.01; **Figure 2**). We visualize the corresponding distributions of the protein-level disorder content and the protein-level LDR numbers for the 24 subcellular compartments, the human proteome and the protein universe in **Supplementary Figure S2**.

Moreover, we observe that the above differences are virtually identical to the results produced when using the subcellular compartments annotated based solely on the experimental evidence (**Supplementary Figure S3**). More specifically, across 98 significance calculations, there are only two differences when comparing the disorder content and number of LDRs in the cytoskeleton against the human proteome. Overall, we observe that the distributions of the proteins (**Figure 1** and **Supplementary Figure S1**) and the comparative analysis of the disorder characteristics (**Figure 2** and **Supplementary Figure S3**) are very consistent across the two parallel analyses that rely on the experimental *vs*. the combined experimental and sequence-similarity derived evidence of subcellular locations. This suggests that the more comprehensive analysis that relies on the larger annotation pool is as accurate as the analysis that considers the purely experimental data.

3.3 Proteins that localize to fewer subcellular compartments are more disordered

Figures 2 shows that only a few subcellular compartments are enriched in disorder when compared to the larger number of the disorder-depleted compartments. This imbalanced relation can be explained by the fact that many proteins are localized in multiple compartments. **Figure 3** summarizes relation between the number of compartments per protein and the disorder content. It reveals that human proteins that are found in a few compartments are substantially more disordered than the proteins that localize in many compartments. To compare, the average disorder content for the large collection of proteins that localize to a single compartment is at about 0.12, whereas the content for a smaller set of proteins that localize at least eight compartments is at about 0.07. The difference in the disorder content is statistically significant

(p-value < 0.01). This interesting trend also explains the imbalanced nature of our results. The few disorder-enriched compartments include a large number of disorder-rich proteins that are associated with fewer subcellular compartments compared to the larger number of the disorder-depleted compartments that cover fewer and disproportionally less disordered proteins that are localized across many compartments. More specifically, the number of protein-compartment assignments for the proteins that have higher than average disorder content equals 11,568, while the number of the pairs with lower than average disorder is similar and equals 12,584.

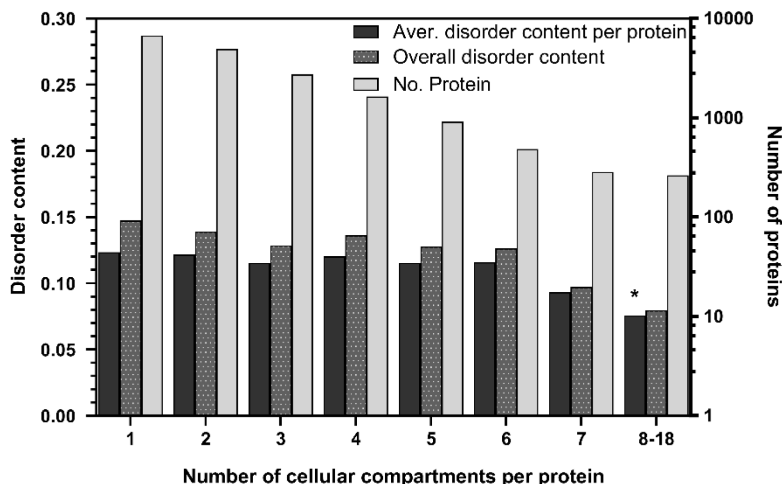

Figure 3. Disorder content for proteins associated with a given number of subcellular compartments that were annotated based on the experimental and sequence-similarity derived evidence. Proteins are grouped by the number of compartments they localize to, which is shown on the x-axis. The left y-axis quantifies the disorder content, including the average of the protein-level content and the aggregate, over all proteins, content. The right y-axis is in the logarithmic scale and quantifies the number of proteins that localize to a given number of subcellular compartments (light gray bars). "*" denotes significant difference in disorder content when compared to the proteins that localize in a single compartment (p-value < 0.01). Details of the statistical tests are explained in Section 2.4.

3.4 Functional Analysis of the Disordered and Structured Proteins

We elucidate the molecular functions associated with the disordered proteins (defined as proteins with one or more LDRs) localized in the disorder-enriched subcellular compartments. Similarly, we also extract the molecular functions associated with the structured proteins (defined as proteins with zero disorder content) localized in the disorder-depleted compartments. We use PANTHER to identify these functions [119, 120]. More specifically, we compare the set of disordered/structured proteins in a given compartment against the human proteome to identify significantly overrepresented molecular functions of the disordered/structured proteins. The statistical significance was assessed with the Fisher' exact test. We assume that a given function is significantly overrepresented if the enrichment rate > 4 and p-value < 0.01.

Table 2 summarizes the results of this analysis and provides four most enriched molecular functions for the disordered proteins in each of the two disorder-enriched compartments and for the structured proteins in each of the 15 disorder-depleted compartments. The complete set of overrepresented functions can be found in in the **Supplementary Table S2** (for the disorder-enriched compartments) and **Supplementary Table S3** (for the disorder-depleted compartments). The disordered/structured proteins provide many functions with very high enrichment rate. We identify between 4 (for peroxisome) and 95 (for cell membrane) functions that have rate = 2, which is equivalent to 200% increase compared to the expected value.

Table 2. Molecular functions associated with disordered protein in the disorder-enriched subcellular compartments and with structured proteins in the disorder-depleted compartments. We list top four functions that are significantly enriched when compared to the overall population of the human proteins based on the analysis with Panther software (p-value < 0.01 and enrichment rate > 4). Complete list of functions is provided in the Supplementary Table S2 (for disorder-enriched compartments) and Supplementary Table S3 (for disorder-depleted compartments). The disordered proteins are defined as proteins with at least one LDR while structured proteins have zero disorder content. The enrichment rate is defined as the rate between the observed and expected number of proteins with a specific molecular function in a given subcellular compartment. The p-values were computed based on the Fisher' exact test with the false disorder rate correction.

Subcellular locations	**Disorder/Structure enrichment (*p*-value)**	**No. of functions with a given enrichment rate**			**Top four significant molecular functions** (p-value < 0.01 & enrichment rate > 4) **[median of per protein disorder content]**
		1.5	**2**	**4**	
Nucleus	Disorder (p<0.0001)	83	76	21	activating transcription factor binding (GO:0033613) [0.184], nucleosome binding (GO:0031491) [0.466], K63-linked polyubiquitin modification-dependent protein binding (GO:0070530) [0.257], pre-mRNA binding (GO:0036002) [0.242]
Cytoskeleton	Disorder (p=0.0052)	33	32	16	dynein heavy chain binding (GO:0045504) [0.222], microtubule binding (GO:0008017) [0.155], motor activity (GO:0003774) [0.106], ATP-dependent microtubule motor activity, plus-end-directed (GO:0008574) [0.073]
Golgi apparatus	Structure (p<0.0001)	32	31	20	nucleotide-sugar transmembrane transporter activity (GO:0005338) [0], acetylglucosaminyltransferase activity (GO:0008375) [0], pyrimidine nucleotide-sugar transmembrane transporter activity (GO:0015165) [0], palmitoyltransferase activity (GO:0016409) [0]
Endosome	Structure (p=0.0023)	45	44	24	solute: proton antiporter activity (GO:0015299) [0], ATPase-coupled intramembrane lipid transporter activity (GO:0140326) [0], cyclin-dependent protein serine/threonine kinase regulator activity (GO:0016538) [0], cyclin-dependent protein serine/threonine kinase activity (GO:0004693) [0]
Cell membrane	Structure (p<0.0001)	101	95	30	transmembrane receptor protein serine/threonine kinase activity (GO:0004675) [0], activin binding (GO:0048185) [0], sugar transmembrane transporter activity (GO:0051119) [0], chemokine binding (GO:0019956) [0]
Golgi membrane	Structure (p=0.0005)	30	28	22	UDP-galactose transmembrane transporter activity (GO:0005459) [0], sialyltransferase activity (GO:0008373) [0], pyrimidine nucleotide-sugar transmembrane transporter activity (GO:0015165) [0], acetylglucosaminyltransferase activity (GO:0008375) [0]
Cell junction	Structure (p<0.0001)	68	62	23	fructose-bisphosphate aldolase activity (GO:0004332) [0], cytokine receptor binding (GO:0005126) [0], cytokine activity (GO:0005125) [0], hydrolase activity, acting on acid phosphorus-nitrogen bonds (GO:0016825) [0]
Endoplasmic reticulum membrane	Structure (p<0.0001)	10	10	7	oligosaccharyl transferase activity (GO:0004576) [0], mannosyltransferase activity (GO:0000030) [0], UDP-glycosyltransferase activity (GO:0008194) [0], transferase activity, transferring glycosyl groups (GO:0016757) [0]
Endoplasmic reticulum	Structure (p<0.0001)	19	18	16	palmitoyltransferase activity (GO:0016409) [0], mannosyltransferase activity (GO:0000030) [0], acetylglucosaminyltransferase activity (GO:0008375) [0], transferase activity, transferring hexosyl groups (GO:0016758) [0]
Other membrane	Structure (p<0.0001)	32	25	5	taste receptor activity (GO:0008527) [0], nucleotide transmembrane transporter activity (GO:0015215) [0], organophosphate ester transmembrane transporter activity (GO:0015605) [0], carbohydrate transmembrane transporter activity (GO:0015144) [0]
Vacuole	Structure (p<0.0001)	9	9	7	ubiquitin protein ligase binding (GO:0031625) [0], ubiquitin-like protein ligase binding (GO:0044389) [0], cysteine-type endopeptidase activity (GO:0004197) [0], cysteine-type peptidase activity (GO:0008234) [0]
Mitochondrion	Structure (p<0.0001)	30	26	19	proton-transporting ATP synthase activity, rotational mechanism (GO:0046933) [0], cytochrome-c oxidase activity (GO:0004129) [0], ligase activity, forming carbon-sulfur bonds (GO:0016877) [0], electron transfer activity (GO:0009055) [0]
Lysosome	Structure (p<0.0001)	7	7	5	cysteine-type endopeptidase activity (GO:0004197) [0], cysteine-type peptidase activity (GO:0008234) [0], endopeptidase activity (GO:0004175) [0], peptidase activity, acting on L-amino acid peptides (GO:0070011) [0]
Mitochondrial membrane	Structure (p<0.0001)	35	34	21	NADH dehydrogenase activity (GO:0003954) [0], cytochrome-c oxidase activity (GO:0004129) [0], proton-transporting ATP synthase activity, rotational mechanism (GO:0046933) [0], ATP transmembrane transporter activity (GO:0005347) [0]
Vacuolar membrane	Structure (p<0.0001)	20	20	20	ubiquitin protein ligase binding (GO:0031625) [0], ubiquitin-like protein ligase binding (GO:0044389) [0], proton transmembrane transporter activity (GO:0015078) [0], ATPase-coupled transmembrane transporter activity (GO:0042626) [0]
Lysosomal membrane	Structure (p<0.0001)	18	18	14	proton transmembrane transporter activity (GO:0015078) [0], ATPase-coupled transmembrane transporter activity (GO:0042626) [0], primary active transmembrane transporter activity (GO:0015399) [0], SNARE binding (GO:0000149) [0]
Peroxisome	Structure (p=0.0002)	4	4	3	flavin adenine dinucleotide binding (GO:0050660) [0], cofactor binding (GO:0048037) [0], oxidoreductase activity (GO:0016491) [0]

This suggests that the disordered proteins in the disorder-enriched compartments and the structured proteins in the disorder-depleted compartments are crucial for a wide range of cellular functions. For the disordered proteins, these functions are primarily related to the interactions with transcription factors, chromatin, RNA, and a variety of protein partners. This is consistent with literature data that similarly point to the importance of disorder in the context of protein-protein interactions [44, 104, 109, 121-123], protein-transcription factors interactions [124-126], chromatin condensing and organization [95, 118, 127], and protein-RNA interactions [94, 105, 128]. On the other hand, the structured proteins that are localized in the membranes are typically associated with molecular transport, while the other structured proteins are involved in the enzymatic activity and well as carbohydrate, nucleotide, cofactor and small ligand binding.

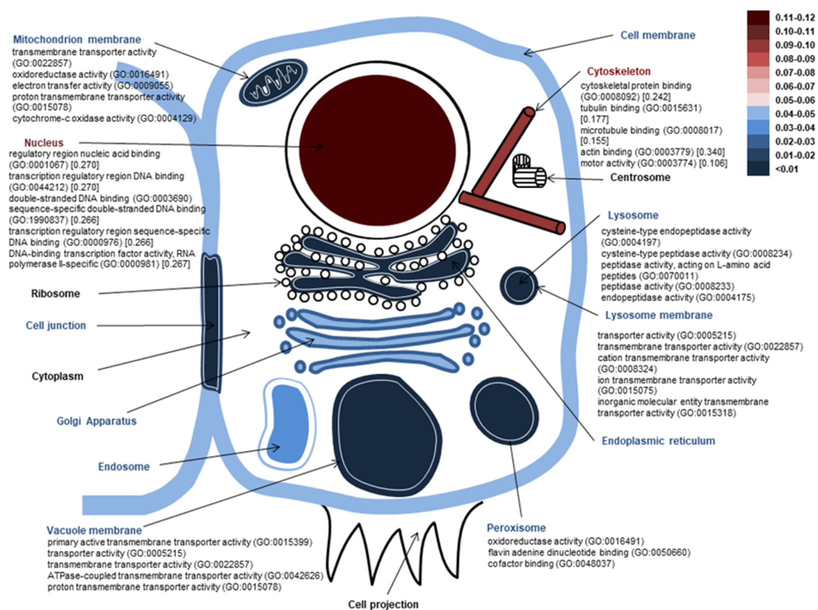

Figure 4. Intrinsic disorder in the human cell. Subcellular compartments are color-coded to reflect their disorder status where darker shade of red indicates the degree of the disorder enrichment while darker share of blue shows the degree of the disorder depletion, relative to the disorder content of the human proteome. Two significantly disorder-enriched compartments (nucleus and cytoskeleton) include a list of the top five molecular functions that are associated with the proteins in these compartments that have long disordered regions. Similarly, the five most disorder-depleted compartments (peroxisome, lysosome and its membrane, vacuolar membrane, and mitochondrial membrane) include the list of the top five molecular functions associated with the fully structured proteins (zero disorder content). Median protein-level disorder content for the proteins annotated with a given function is shown inside the square brackets; the functions for the fully structured proteins have the median content of zero. The complete list of cellular functions for the disorder- and order-enriched compartments is available in Supplementary Table S2.

3.5 Intrinsic Disorder in the Human Cell

Figure 4 summarizes the results of this study by mapping the intrinsic disorder into the human cell. The two significantly disorder-enriched subcellular compartments, relative to the overall amount of disorder in the human proteome, include nucleus (disorder content = 0.17 and 57% of the proteins with LDRs) and cytoskeleton (disorder content = 0.15 and 56% of proteins with LDRs). The 15 disorder-depleted compartments are colored in blue and include the cell junction, endosome, cell membrane, peroxisome, and well as Golgi apparatus, endoplasmic reticulum, mitochondrion, vacuole, lysosome and their membranes. The median protein-level disorder content in these compartments is either below 0.05 (Golgi apparatus and its membrane, endosome and cell membrane) or at zero (the remaining disorder-depleted compartments). We list the key cellular functions associated with the disordered proteins in the most disorder-enriched compartments and for the structured proteins in the most disorder-depleted compartments. As we mentioned above, the disordered proteins are crucial for the protein-nucleic acids

and protein-protein interactions in the cytoskeleton and nucleus. On the other hand, the structured proteins are primarily involved in molecular transport, enzymatic activity and interactions with smaller ligands.

4 Summary and Conclusions

Intrinsic disorder defines structural and functional heterogeneity of proteins [20, 23, 122, 129, 130]. It plays a crucial role in the existence of proteoforms (i.e., structurally and functionally different forms of a protein encoded by a given gene [131]) contributing to the notion of protein multifunctionality and protein structure-function continuum [22, 23]. The disorder-based heterogeneity and the presence of the proteoforms that can be (dis)ordered to different degree form the foundation of the "disorder-based heterogeneity pyramid". The next level of this pyramid is related to the inhomogeneous involvement of disorder in protein function, as a result of which, one can find order-specific (catalysis and transport) and disorder-specific (recognition, regulation, signaling) protein functions [23, 27, 104, 111, 122, 130, 132-142], which may co-exist within one hybrid protein molecule that includes both disordered and ordered regions [143, 144].

In the cell, proteins typically do not act alone, but are engaged in the formation of highly connected and regulated protein-protein interaction (PPI) networks [145]. The node degree distributions in these networks follows a power law with a relatively low number of highly-connected hubs; i.e., nodes with a number of links that greatly exceeds the network average number of links [145, 146]. Not surprisingly, many hubs of the PPI networks are IDPs/IDPRs [44, 147-152], which are known to be promiscuous binders. Furthermore, even if a given hub is an ordered protein, most of its partners were shown to be IDPs or hybrid proteins using their IDPRs to interact with such a hub [44].

Data reported in our study fill the gap between the well-studied distribution of disorder within the PPI networks vs. relatively poorly understood distribution of disorder at the subcellular level. In agreement with literature [14, 16-18], our comprehensive analysis shows that human proteins are substantially enriched in disorder, relative to a generic protein set that represents the protein universe. We demonstrate that IDPs/IDPRs are not homogeneously distributed within a cell. Instead, proteins located in several subcellular compartments (nucleus and its membrane, cytoskeleton, centrosome, and cytoplasm) are enriched in disorder, measured by both high disorder content and presence of multiple IDRs, relative to the remainder of the human proteins. This enrichment is statistically significant for nucleus and cytoskeleton. Moreover, we show that proteins localized to a few subcellular compartments are more disordered than proteins that are localized to many compartments. Our analysis also reveals that the disordered proteins localized in the most disorder-enriched compartments are primarily responsible for the interactions with DNA, RNA and a variety of protein partners. We believe that these are important observations that add a missing piece to the puzzle of multifunctionality of disorder.

Our study of the peculiarities of protein intrinsic disorder distribution within the generic human cell sets the grounds for the subsequent analyses of the differences in disorder distribution between the somatic and germ cells, peculiarities of disorder distribution within the gametes and stem cells, specific features of disorder distribution within the somatic cells from different organs and tissues, as well as changes in disorder distribution associated with cell cycle. Another logical extension of this study is a comparative analysis of the peculiarities of intracellular distribution of IDPs/IDPRs within plant, bacterial, and archaeal cells. These analyses should culminate in the creation of corresponding cellular disorder/order heat-maps, which would consider not only the distribution peculiarities of disordered and ordered protein within the various intracellular compartments, but also would pay attention to the levels of such compartment-specific disordered and ordered proteins and changes in these levels during the cell life.

References

1. Levitt, M., *Nature of the protein universe.* Proc Natl Acad Sci U S A, 2009. **106**(27): p. 11079-84.
2. Nepomnyachiy, S., N. Ben-Tal, and R. Kolodny, *Global view of the protein universe.* Proc Natl Acad Sci U S A, 2014. **111**(32): p. 11691-6.
3. Han, X., et al., *A global map of the protein shape universe.* PLoS Comput Biol, 2019. **15**(4): p. e1006969.
4. Andreeva, A., et al., *SCOP2 prototype: a new approach to protein structure mining.* Nucleic Acids Res, 2014. **42**(Database issue): p. D310-4.
5. Sillitoe, I., et al., *CATH: comprehensive structural and functional annotations for genome sequences.* Nucleic Acids Res, 2015. **43**(Database issue): p. D376-81.
6. Kulkarni, P. and V.N. Uversky, *Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome.* Proteomics, 2018. **18**(21-22): p. e1800061.
7. Hu, G., et al., *Taxonomic Landscape of the Dark Proteomes: Whole-Proteome Scale Interplay Between Structural Darkness, Intrinsic Disorder, and Crystallization Propensity.* Proteomics, 2018: p. e1800243.
8. Bhowmick, A., et al., *Finding Our Way in the Dark Proteome.* J Am Chem Soc, 2016. **138**(31): p. 9730-42.
9. Lieutaud, P., et al., *How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe.* Intrinsically Disord Proteins, 2016. **4**(1): p. e1259708.
10. Habchi, J., et al., *Introducing protein intrinsic disorder.* Chem Rev, 2014. **114**(13): p. 6561-88.
11. van der Lee, R., et al., *Classification of Intrinsically Disordered Regions and Proteins.* Chemical Reviews, 2014. **114**(13): p. 6589-6631.
12. A. Keith Dunker, M.M.B., Elisar Barbar, Martin Blackledge, Sarah E. Bondos, Zsuzsanna Dosztányi, H. Jane Dyson, Julie Forman-Kay, Monika Fuxreiter, Jörg Gsponer, Kyou-Hoon Han, David T. Jones, Sonia Longhi, Steven J. Metallo, Ken Nishikawa, Ruth Nussinov, Zoran Obradovic, Rohit V. Pappu, Burkhard Rost, Philipp Selenko, Vinod Subramaniam, Joel L. Sussman, Peter Tompa & Vladimir N Uversky, *What's in a name? Why these proteins are intrinsically disordered.* Intrinsically Disordered Proteins, 2013. **1**(1): p. e24157
13. Oldfield, C.J., et al., *Introduction to intrinsically disordered proteins and regions*, in *Intrinsically Disordered Proteins*, N. Salvi, Editor. 2019, Academic Press. p. 1-34.
14. Peng, Z., et al., *Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life.* Cell Mol Life Sci, 2015. **72**(1): p. 137-51.
15. Uversky, V.N., *The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome.* J. Biomed. Biotechnol., 2010. **2010**: p. 568068.
16. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.* J Mol Biol, 2004. **337**(3): p. 635-45.
17. Xue, B., A.K. Dunker, and V.N. Uversky, *Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life.* J. Biomol. Struct. Dyn., 2012. **30**(2): p. 137-49.
18. Schad, E., P. Tompa, and H. Hegyi, *The relationship between proteome size, structural disorder and organism complexity.* Genome Biol, 2011. **12**(12): p. R120.

19. Xie, H., et al., *Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions.* J. Proteome Res., 2007. **6**(5): p. 1882-98.
20. Uversky, V.N., *Unusual biophysics of intrinsically disordered proteins.* Biochim Biophys Acta, 2013. **1834**(5): p. 932-51.
21. Uversky, V.N., *The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini.* FEBS Lett, 2013. **587**(13): p. 1891-901.
22. Uversky, V.N., *p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept.* Int J Mol Sci, 2016. **17**(11): p. 1874.
23. Uversky, V.N., *Functional roles of transiently and intrinsically disordered regions within proteins.* FEBS J, 2015. **282**(7): p. 1182-9.
24. Dunker, A.K., et al., *Intrinsically disordered protein.* J Mol Graph Model, 2001. **19**(1): p. 26-59.
25. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 2000. **41**(3): p. 415-427.
26. Dunker, A.K., C.J. Brown, and Z. Obradovic, *Identification and functions of usefully disordered proteins.* Adv Protein Chem, 2002. **62**: p. 25-49.
27. Dunker, A.K., et al., *Intrinsic disorder and protein function.* Biochemistry, 2002. **41**(21): p. 6573-82.
28. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein Sci, 2002. **11**(4): p. 739-56.
29. Uversky, V.N., *What does it mean to be natively unfolded?* Eur J Biochem, 2002. **269**(1): p. 2-12.
30. Uversky, V.N., *Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go?* Cell Mol Life Sci, 2003. **60**(9): p. 1852-71.
31. Uversky, V.N., *Functions of short lifetime biological structures at large: the case of intrinsically disordered proteins.* Brief Funct Genomics, 2020. **19**(1): p. 60-68.
32. Iakoucheva, L.M., et al., *Intrinsic disorder in cell-signaling and cancer-associated proteins.* J Mol Biol, 2002. **323**(3): p. 573-84.
33. Uversky, V.N., *The triple power of D(3): protein intrinsic disorder in degenerative diseases.* Front Biosci (Landmark Ed), 2014. **19**: p. 181-258.
34. Du, Z. and V.N. Uversky, *A Comprehensive Survey of the Roles of Highly Disordered Proteins in Type 2 Diabetes.* Int J Mol Sci, 2017. **18**(10).
35. Cheng, Y., et al., *Abundance of intrinsic disorder in protein associated with cardiovascular disease.* Biochemistry, 2006. **45**(35): p. 10448-60.
36. Uversky, V.N., *Amyloidogenesis of natively unfolded proteins.* Curr Alzheimer Res, 2008. **5**(3): p. 260-87.
37. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Intrinsically disordered proteins in human diseases: introducing the D2 concept.* Annu Rev Biophys, 2008. **37**: p. 215-46.
38. Uversky, V.N., et al., *Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases.* Chem Rev, 2014. **114**(13): p. 6844-79.
39. Xue, B., et al., *Archaic chaos: intrinsically disordered proteins in Archaea.* BMC Syst. Biol., 2010. **4 Suppl 1**: p. S1.
40. Chen, J.W., et al., *Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions.* J Proteome Res, 2006. **5**(4): p. 879-87.

41. Yan, J., et al., *RAPID: fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale.* Biochim Biophys Acta, 2013. **1834**(8): p. 1671-80.
42. Peng, Z., M.J. Mizianty, and L. Kurgan, *Genome-scale prediction of proteins with long intrinsically disordered regions.* Proteins, 2014. **82**(1): p. 145-58.
43. Midic, U., et al., *Protein disorder in the human diseasome: unfoldomics of human genetic diseases.* BMC Genomics, 2009. **10 Suppl 1**: p. S12.
44. Hu, G., et al., *Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions.* Int J Mol Sci, 2017. **18**(12).
45. Colak, R., et al., *Distinct types of disorder in the human proteome: functional implications for alternative splicing.* PLoS Comput Biol, 2013. **9**(4): p. e1003030.
46. Korneta, I. and J.M. Bujnicki, *Intrinsic disorder in the human spliceosomal proteome.* PLoS Comput Biol, 2012. **8**(8): p. e1002641.
47. Pentony, M.M. and D.T. Jones, *Modularity of intrinsic disorder in the human proteome.* Proteins, 2010. **78**(1): p. 212-21.
48. Forcelloni, S. and A. Giansanti, *Evolutionary Forces and Codon Bias in Different Flavors of Intrinsic Disorder in the Human Proteome.* J Mol Evol, 2020. **88**(2): p. 164-178.
49. Fong, J.H., B.A. Shoemaker, and A.R. Panchenko, *Intrinsic protein disorder in human pathways.* Mol Biosyst, 2012. **8**(1): p. 320-6.
50. Fagerberg, L., et al., *Mapping the subcellular protein distribution in three human cell lines.* J Proteome Res, 2011. **10**(8): p. 3766-77.
51. Thul, P.J., et al., *A subcellular map of the human proteome.* Science, 2017. **356**(6340).
52. Meng, F., et al., *Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments.* Int J Mol Sci, 2015. **17**(1).
53. Darling, A.L., et al., *Intrinsically Disordered Proteome of Human Membrane-Less Organelles.* Proteomics, 2018. **18**(5-6): p. e1700193 (12 pages).
54. Gene Ontology, C., et al., *Gene Ontology annotations and resources.* Nucleic Acids Res, 2013. **41**(Database issue): p. D530-5.
55. Gene Ontology, C., *Gene Ontology Consortium: going forward.* Nucleic Acids Res, 2015. **43**(Database issue): p. D1049-56.
56. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.
57. Binder, J.X., et al., *COMPARTMENTS: unification and visualization of protein subcellular localization evidence.* Database (Oxford), 2014. **2014**: p. bau012.
58. Piovesan, D., et al., *MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins.* Nucleic Acids Res, 2018. **46**(D1): p. D471-D476.
59. Vucetic, S., et al., *DisProt: a database of protein disorder.* Bioinformatics, 2005. **21**(1): p. 137-40.
60. Hatos, A., et al., *DisProt: intrinsic protein disorder annotation in 2020.* Nucleic Acids Res, 2020. **48**(D1): p. D269-D276.
61. Miskei, M., C. Antal, and M. Fuxreiter, *FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies.* Nucleic Acids Res, 2017. **45**(D1): p. D228-D235.
62. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Res, 2000. **28**(1): p. 235-42.
63. Le Gall, T., et al., *Intrinsic disorder in the Protein Data Bank.* J Biomol Struct Dyn, 2007. **24**(4): p. 325-42.

64. Ulrich, E.L., et al., *BioMagResBank.* Nucleic Acids Res, 2008. **36**(Database issue): p. D402-8.
65. He, B., et al., *Predicting intrinsic disorder in proteins: an overview.* Cell Res, 2009. **19**(8): p. 929-949.
66. Meng, F., V.N. Uversky, and L. Kurgan, *Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions.* Cell Mol Life Sci, 2017. **74**(17): p. 3069-3090.
67. Meng, F., V. Uversky, and L. Kurgan, *Computational Prediction of Intrinsic Disorder in Proteins.* Curr Protoc Protein Sci, 2017. **88**: p. 2 16 1-2 16 14.
68. Liu, Y., X. Wang, and B. Liu, *A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction.* Brief Bioinform, 2019. **20**(1): p. 330-346.
69. Katuwawala, A., C.J. Oldfield, and L. Kurgan, *Accuracy of protein-level disorder predictions.* Brief Bioinform, 2020: p. doi: 10.1093/bib/bbz100.
70. Necci, M., et al., *A comprehensive assessment of long intrinsic protein disorder from the DisProt database.* Bioinformatics, 2018. **34**(3): p. 445-452.
71. Walsh, I., et al., *Comprehensive large-scale assessment of intrinsic protein disorder.* Bioinformatics, 2015. **31**(2): p. 201-8.
72. Monastyrskyy, B., et al., *Assessment of protein disorder region predictions in CASP10.* Proteins, 2014. **82 Suppl 2**: p. 127-37.
73. Peng, Z.L. and L. Kurgan, *Comprehensive comparative assessment of in-silico predictors of disordered regions.* Curr Protein Pept Sci, 2012. **13**(1): p. 6-18.
74. Katuwawala, A., C.J. Oldfield, and L. Kurgan, *DISOselect: Disorder predictor selection at the protein level.* Protein Sci, 2020. **29**(1): p. 184-200.
75. Peng, Z. and L. Kurgan, *On the complementarity of the consensus-based disorder prediction.* Pac Symp Biocomput, 2012: p. 176-87.
76. Zhao, B. and B. Xue, *Decision-Tree Based Meta-Strategy Improved Accuracy of Disorder Prediction and Identified Novel Disordered Residues Inside Binding Motifs.* Int J Mol Sci, 2018. **19**(10).
77. Necci, M., et al., *MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins.* Bioinformatics, 2017. **33**(9): p. 1402-1404.
78. Yan, J., M. Marcus, and L. Kurgan, *Comprehensively designed consensus of standalone secondary structure predictors improves Q3 by over 3%.* J Biomol Struct Dyn, 2014. **32**(1): p. 36-51.
79. Barik, A., et al., *DEPICTER: Intrinsic Disorder and Disorder Function Prediction Server.* J Mol Biol, 2020. **432**(11): p. 3379-3387.
80. Walsh, I., et al., *ESpritz: accurate and fast prediction of protein disorder.* Bioinformatics, 2012. **28**(4): p. 503-9.
81. Meszaros, B., G. Erdos, and Z. Dosztanyi, *IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding.* Nucleic Acids Res, 2018. **46**(W1): p. W329-W337.
82. Dosztanyi, Z., et al., *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, 2005. **21**(16): p. 3433-4.
83. Linding, R., et al., *Protein disorder prediction: implications for structural proteomics.* Structure, 2003. **11**(11): p. 1453-9.

84. Yang, Z.R., et al., *RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.* Bioinformatics, 2005. **21**(16): p. 3369-76.
85. Peng, K., et al., *Length-dependent prediction of protein intrinsic disorder.* BMC Bioinformatics, 2006. **7**: p. 208.
86. Linding, R., et al., *GlobPlot: Exploring protein sequences for globularity and disorder.* Nucleic Acids Res, 2003. **31**(13): p. 3701-8.
87. UniProt, C., *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.
88. Mitchell, A.L., et al., *InterPro in 2019: improving coverage, classification and access to protein sequence annotations.* Nucleic Acids Res, 2019. **47**(D1): p. D351-D360.
89. Dana, J.M., et al., *SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins.* Nucleic Acids Res, 2019. **47**(D1): p. D482-D489.
90. Meng, F., et al., *Compartmentalization and Functionality of Nuclear Disorder: Intrinsic Disorder and Protein-Protein Interactions in Intra-Nuclear Compartments.* Int J Mol Sci, 2016. **17**(1).
91. Kulkarni, P. and V.N. Uversky, *Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome.* Proteomics, 2018. **18**(21-22).
92. Peng, Z., V.N. Uversky, and L. Kurgan, *Genes Encoding Intrinsic Disorder in Eukaryota Have High GC Content.* Intrinsically Disordered Proteins, 2016. **4**(1): p. e1262225.
93. Yan, J., et al., *Structural and functional analysis of "non-smelly" proteins.* Cell Mol Life Sci, 2019.
94. Peng, Z., et al., *A creature with a hundred waggly tails: intrinsically disordered proteins in the ribosome.* Cell Mol Life Sci, 2014. **71**(8): p. 1477-504.
95. Peng, Z., et al., *More than just tails: intrinsic disorder in histone proteins.* Mol Biosyst, 2012. **8**(7): p. 1886-901.
96. Zhao, B. and B. Xue, *Self-regulation of functional pathways by motifs inside the disordered tails of beta-catenin.* BMC Genomics, 2016. **17 Suppl 5**: p. 484.
97. Charon, J., et al., *Protein intrinsic disorder within the Potyvirus genus: from proteome-wide analysis to functional annotation.* Molecular Biosystems, 2016. **12**(2): p. 634-652.
98. Dolan, P.T., et al., *Intrinsic disorder mediates hepatitis C virus core-host cell protein interactions.* Protein Sci, 2015. **24**(2): p. 221-35.
99. Fan, X., et al., *The intrinsic disorder status of the human hepatitis C virus proteome.* Mol Biosyst, 2014. **10**(6): p. 1345-63.
100. Xue, B., et al., *Protein intrinsic disorder as a flexible armor and a weapon of HIV-1.* Cell Mol Life Sci, 2012. **69**(8): p. 1211-59.
101. Xue, B., et al., *Structural disorder in viral proteins.* Chem Rev, 2014. **114**(13): p. 6880-911.
102. Meng, F., et al., *Unstructural biology of the Dengue virus proteins.* FEBS J, 2015. **282**(17): p. 3368-94.
103. Patil, A., K. Kinoshita, and H. Nakamura, *Hub promiscuity in protein-protein interaction networks.* Int J Mol Sci, 2010. **11**(4): p. 1930-43.
104. Fuxreiter, M., et al., *Disordered proteinaceous machines.* Chem Rev, 2014. **114**(13): p. 6806-43.

105. Wang, C., V.N. Uversky, and L. Kurgan, *Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea.* Proteomics, 2016. **16**(10): p. 1486-98.
106. Varadi, M., et al., *Functional Advantages of Conserved Intrinsic Disorder in RNA-Binding Proteins.* PLoS One, 2015. **10**(10): p. e0139731.
107. Peng, Z., et al., *Resilience of death: intrinsic disorder in proteins involved in the programmed cell death.* Cell Death Differ, 2013. **20**(9): p. 1257-67.
108. Uversky, A.V., et al., *On the intrinsic disorder status of the major players in programmed cell death pathways.* F1000Res, 2013. **2**: p. 190.
109. Na, I., et al., *Autophagy-related intrinsically disordered proteins in intra-nuclear compartments.* Mol Biosyst, 2016. **12**(9): p. 2798-817.
110. Boutet, E., et al., *UniProtKB/Swiss-Prot.* Methods Mol Biol, 2007. **406**: p. 89-112.
111. Tompa, P., et al., *Close encounters of the third kind: disordered domains and the interactions of proteins.* Bioessays, 2009. **31**(3): p. 328-35.
112. Howell, M., et al., *Not That Rigid Midgets and Not So Flexible Giants: On the Abundance and Roles of Intrinsic Disorder in Short and Long Proteins.* Journal of Biological Systems, 2012. **20**(4): p. 471-511.
113. Yan, J., et al., *Molecular recognition features (MoRFs) in three domains of life.* Mol Biosyst, 2016. **12**(3): p. 697-710.
114. Meng, F., et al., *Functional and structural characterization of osteocytic MLO-Y4 cell proteins encoded by genes differentially expressed in response to mechanical signals in vitro.* Sci Rep, 2018. **8**(1): p. 6716.
115. Ghadermarzi, S., et al., *Sequence-Derived Markers of Drug Targets and Potentially Druggable Human Proteins.* Front Genet, 2019. **10**: p. 1075.
116. Armstrong, R.A., *When to use the Bonferroni correction.* Ophthalmic and Physiological Optics, 2014. **34**(5): p. 502-508.
117. Oates, M.E., et al., *D(2)P(2): database of disordered protein predictions.* Nucleic Acids Res, 2013. **41**(Database issue): p. D508-16.
118. Sandhu, K.S., *Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins.* J Mol Recognit, 2009. **22**(1): p. 1-8.
119. Mi, H., et al., *Large-scale gene function analysis with the PANTHER classification system.* Nat Protoc, 2013. **8**(8): p. 1551-66.
120. Mi, H., et al., *Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0).* Nat Protoc, 2019. **14**(3): p. 703-721.
121. Uversky, V.N., *Intrinsic Disorder, Protein-Protein Interactions, and Disease.* Adv Protein Chem Struct Biol, 2018. **110**: p. 85-121.
122. Uversky, V.N., *Intrinsic disorder-based protein interactions and their modulators.* Curr Pharm Des, 2013. **19**(23): p. 4191-213.
123. Hsu, W.L., et al., *Intrinsic protein disorder and protein-protein interactions.* Pac Symp Biocomput, 2012: p. 116-27.
124. Staby, L., et al., *Eukaryotic transcription factors: paradigms of protein intrinsic disorder.* Biochem J, 2017. **474**(15): p. 2509-2532.
125. Liu, J., et al., *Intrinsic disorder in transcription factors.* Biochemistry, 2006. **45**(22): p. 6873-88.
126. Toth-Petroczy, A., et al., *Malleable machines in transcription regulation: the mediator complex.* PLoS Comput Biol, 2008. **4**(12): p. e1000243.

127. Lu, X., et al., *Chromatin condensing functions of the linker histone C-terminal domain are mediated by specific amino acid composition and intrinsic protein disorder.* Biochemistry, 2009. **48**(1): p. 164-72.
128. Mallik, S. and S. Kundu, *Molecular interactions within the halophilic, thermophilic, and mesophilic prokaryotic ribosomal complexes: clues to environmental adaptation.* J Biomol Struct Dyn, 2015. **33**(3): p. 639-56.
129. Uversky, V.N., *A decade and a half of protein intrinsic disorder: biology still waits for physics.* Protein Sci, 2013. **22**(6): p. 693-724.
130. Jakob, U., R. Kriwacki, and V.N. Uversky, *Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function.* Chem Rev, 2014. **114**(13): p. 6779-805.
131. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity.* Nat Methods, 2013. **10**(3): p. 186-7.
132. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm.* J Mol Biol, 1999. **293**(2): p. 321-331.
133. Dunker, A.K. and Z. Obradovic, *The protein trinity--linking function and disorder.* Nat Biotechnol, 2001. **19**(9): p. 805-6.
134. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Curr Opin Struct Biol, 2002. **12**(1): p. 54-60.
135. Dunker, A.K., et al., *Flexible nets: The roles of intrinsic disorder in protein interaction networks.* FEBS Journal, 2005. **272**(20): p. 5129-5148.
136. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
137. Tompa, P., *The interplay between structure and function in intrinsically unstructured proteins.* FEBS Lett, 2005. **579**(15): p. 3346-54.
138. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.* J Mol Recognit, 2005. **18**(5): p. 343-384.
139. Xie, H., et al., *Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions.* J Proteome Res, 2007. **6**(5): p. 1882-98.
140. Dunker, A.K., et al., *Function and structure of inherently disordered proteins.* Curr Opin Struct Biol, 2008. **18**(6): p. 756-64.
141. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Curr Opin Struct Biol, 2009. **19**(1): p. 31-8.
142. Uversky, V.N., *Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes.* Chem Soc Rev, 2011. **40**(3): p. 1623-34.
143. DeForte, S. and V.N. Uversky, *Not an exception to the rule: the functional significance of intrinsically disordered protein regions in enzymes.* Mol Biosyst, 2017. **13**(3): p. 463-469.
144. DeForte, S. and V.N. Uversky, *Order, Disorder, and Everything in Between.* Molecules, 2016. **21**(8).
145. Yook, S.H., Z.N. Oltvai, and A.L. Barabasi, *Functional and topological characterization of protein interaction networks.* Proteomics, 2004. **4**(4): p. 928-42.
146. Jeong, H., et al., *Lethality and centrality in protein networks.* Nature, 2001. **411**(6833): p. 41-2.
147. Dunker, A.K., et al., *Flexible nets. The roles of intrinsic disorder in protein interaction networks.* FEBS J, 2005. **272**(20): p. 5129-48.

148. Dosztanyi, Z., et al., *Disorder and sequence repeats in hub proteins and their implications for network evolution.* J Proteome Res, 2006. **5**(11): p. 2985-95.
149. Haynes, C., et al., *Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.
150. Oldfield, C.J., et al., *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC Genomics, 2008. **9 Suppl 1**: p. S1.
151. Patil, A. and H. Nakamura, *Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks.* FEBS Lett, 2006. **580**(8): p. 2041-5.
152. Patil, A., K. Kinoshita, and H. Nakamura, *Domain distribution and intrinsic disorder in hubs in the human protein-protein interaction network.* Protein Sci, 2010. **19**(8): p. 1461-8.