

RESEARCH ARTICLE

Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell

Antonio Deiana¹, Sergio Forcelloni¹, Alessandro Porrello², Andrea Giansanti^{1,3*}

1 Sapienza University of Rome, Department of Physics, Roma, Italy, **2** Lineberger Comprehensive Cancer Center (LCCC), University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America, **3** Istituto Nazionale di Fisica Nucleare, INFN, Roma, Italy

* andrea.giansanti@roma1.infn.it



OPEN ACCESS

Citation: Deiana A, Forcelloni S, Porrello A, Giansanti A (2019) Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. PLoS ONE 14(8): e0217889. <https://doi.org/10.1371/journal.pone.0217889>

Editor: Eugene A. Permyakov, Russian Academy of Medical Sciences, RUSSIAN FEDERATION

Received: May 20, 2019

Accepted: August 6, 2019

Published: August 19, 2019

Copyright: © 2019 Deiana et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: Publication of this study was funded in part by a basic research grant to AG from the Sapienza University of Rome (000008_17_RS_PICCOLI_2016_GIANSANTI-RICERCA 2016). There were no additional sources of funding. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Many studies about classification and the functional annotation of intrinsically disordered proteins (IDPs) are based on either the occurrence of long disordered regions or the fraction of disordered residues in the sequence. Taking into account both criteria we separate the human proteome, taken as a case study, into three variants of proteins: i) *ordered proteins* (ORDPs), ii) *structured proteins with intrinsically disordered regions* (IDPRs), and iii) *intrinsically disordered proteins* (IDPs). The focus of this work is on the different functional roles of IDPs and IDPRs, which up until now have been generally considered as a whole. Previous studies assigned a large set of functional roles to the general category of IDPs. We show here that IDPs and IDPRs have non-overlapping functional spectra, play different roles in human diseases, and deserve to be treated as distinct categories of proteins. IDPs enrich only a few classes, functions, and processes: *nucleic acid binding proteins*, *chromatin binding proteins*, *transcription factors*, and *developmental processes*. In contrast, IDPRs are spread over several functional protein classes and GO annotations which they partly share with ORDPs. As regards to diseases, we observe that IDPs enrich only cancer-related proteins, at variance with previous results reporting that IDPs are widespread also in cardiovascular and neurodegenerative pathologies. Overall, the operational separation of IDPRs from IDPs is relevant towards correct estimates of the occurrence of intrinsically disordered proteins in genome-wide studies and in the understanding of the functional spectra associated to different flavors of protein disorder.

Introduction

Over the last two decades, the concept of ‘*intrinsic disorder*’ has emerged as a prominent and influential topic in protein science [1–7]. The discovery of intrinsically disordered proteins (IDPs), i.e., functional proteins lacking a well-defined tertiary structure, has challenged the

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ORDPs, ordered proteins; IDPRs, structured proteins with intrinsically disordered regions; IDPs, intrinsically disordered proteins.

traditional sequence-structure-function paradigm [8,9]. In their seminal paper, Uversky, Gillespie, and Fink [2] introduced the “charge-hydrophobicity phase space”, which clarified that IDPs are characterized by a combination of “low overall hydrophobicity and large net charge”. IDPs constitute, in eukaryotes, a substantial part of the cellular proteome and are involved in many biological processes that complement the functional repertoires of ordered proteins [10,11]. Ever-increasing experimental evidence has revealed the presence of disordered regions also in well-structured proteins [12,13]. Then the distinction between intrinsically disordered proteins, i.e., proteins lacking a tertiary structure, and structured proteins containing intrinsically disordered regions gradually emerged [14–17]. An operational distinction of these two variants is the main objective of the present work. A note on terminology is in order at this point. The editors of the journal *Intrinsically Disordered Proteins* made an effort to disambiguate the semantics of protein intrinsic disorder [18]. They proposed “intrinsically disordered proteins” as a unifying term, recognizing that it is a compromise “far from being ideal”, and suggesting that “additional descriptors” would emerge, apt to clarify the many aspects of “structurelessness” that are included in one term. The terminology we adopt here is surely also not ideal. Let us try to make clear that we use here IDPs, when referring to previous studies, as the general unifying term proposed by Dunker et al. We use IDPs also to denote a subgroup of Dunker et al.’s IDPs, a variant that is operationally specified here as distinct from the variant of mostly structured proteins with long intrinsically disordered regions (we call here IDPRs).

The biological role of IDPs has been the focus of a growing number of publications, in the recent past. Xie et al., for example, in a series of three articles compiled an anthology of the functional roles of IDPs [19–21]. It has been reported that out of the 710 keywords recording biological functions in the Uniprot/Swissprot database, 310 (44%) are associated with ordered proteins and 238 (34%) with IDPs [22]. Several studies have reported lists of Gene Ontology functional classes that are enriched in IDPs, including but not limited to: cell regulation, transcription, translation, signaling, and alternative splicing [23–30]. Different studies highlight the occurrence of unfolded regions in proteins that function as chaperones for other proteins or RNA molecules, transcription factors, effectors, assemblers, and scavengers [14–16,27]. Importantly, many proteins with disordered regions have been associated with human diseases [9,31,32], and a D^2 concept (disorder in disorders) emerged [33]. However, in all these studies what we distinguish here as IDPs and IDPRs were considered as a unique variant, possibly aggregating proteins which are structurally and functionally different. This is still reflected, for example, in a recent paper by Darling and Uversky [34]. IDPRs, as we define them, may have a well-defined 3D structure that either can undergo conformational, allosteric changes (e.g., after a post-translational modification [35]) or accommodate functional, locally disordered domains, that are limited to less than 30% of the structure. The presence of unstructured loops or domains, e.g., linear motifs and Molecular Recognition Features (MORFs), can be important for these proteins to target low-affinity substrates enlarging the repertoire of protein interactions [15].

Let us remind that one of the features that distinguishes IDPs from structured proteins is their different interaction mechanism with target substrates. According to the classic *lock-and-key mechanism* [36], the specificity and affinity required for molecular interactions depend on the complementarity of the binding interfaces. However, even proteins that are disordered under physiological conditions or contain long unstructured regions can form stable complexes [9,37]. In this context, the lack of a well-defined 3D-structure represents a major functional advantage for IDPs, allowing them to interact with a broad range of substrates with relatively high-specificity and low-affinity, often undergoing a disorder-to-order transition upon binding [3,9,27,38,39]. However, despite that this description makes sense for proteins lacking a tertiary structure (IDPs), it is reasonable to suppose that structured proteins with

long disordered regions (IDPRs) still interact with substrates through a mechanism similar to the lock-and-key. It has been reported that disordered regions have many post-translational modification sites (PTMs) [9,30,40–42], which induce conformational changes leading to *one-lock-many-keys interactions* [34]. A recent study reviews different interaction mechanisms for unfolded and folded proteins [35]. Unfolded proteins are characterized by *induced-fit interactions*, whereas *conformational change interactions* characterize folded proteins. Although the recent literature does not explicitly separate IDPs from IDPRs, it is reasonable to hypothesize that IDPRs typically undergo conformational changes after post-translational modifications, whereas high-specificity and low-affinity induced fit mainly characterize IDPs.

Many studies aiming at a general classification of IDPs' functional roles were based on the required presence of long disordered regions in the sequence [19–21,43–46]. However, the presence of a long disordered segment does not necessarily imply that a protein lacks a three-dimensional structure. In support of this claim, Gsponer et al. showed that also the percentage of disordered residues plays an important role in determining the half-life of a protein [47]. Therefore, both the length of disordered domains and the overall fraction of disordered residues in the sequence are critical parameters to distinguish different types of protein disorder.

Following clear-cut criteria inspired by experimental observations [15,47], we compiled two lists of human IDPs and IDPRs. We identified IDPs as those proteins having more than 30% of disordered residues (implying a shorter half-life [47]). We defined IDPRs as proteins with at least one long disordered segment (suggesting a higher sensitivity to proteolytic degradation [15]) but less than 30% of disordered residues in their sequence. We have assessed both the over-representation and the enrichment of IDPRs and IDPs in different functional classes and gene ontologies. Moreover, we have investigated the relative occurrence of IDPRs and IDPs in proteins related to cancer, cardiovascular, and neurodegenerative diseases.

Our results indicate that IDPRs and IDPs are structurally different, have different functional spectra, and play different roles in human diseases.

Materials and methods

Dataset of protein sequences

We downloaded the human proteome from the UniProt-SwissProt database (<http://www.uniprot.org/uniprot>) [48], release 2018_07. We selected 20386 human proteins by searching for reviewed proteins belonging to Homo Sapiens (Organism ID: 9606, Proteome ID: UP000005640).

Dataset of protein sequences associated to diseases

To select the proteins associated to cancer, cardiovascular, and neurodegenerative diseases, we used the lists of keywords found in Iakoucheva *et al.* [49], Cheng *et al.* [50], and Uversky [51]. Then, we extracted the proteins annotated with those keywords from the UniProt-SwissProt database. The lists of disease-related proteins are available in [S2](#), [S3](#) and [S4 Texts](#).

Prediction of protein disorder

We identified disordered residues in the protein sequences using MobiDB (<http://mobidb.bio.unipd.it>) [52], a consensus database that combines experimental data (especially from X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (Cryo-EM)), curated data and disorder predictions based on various methods. From the initial dataset, we discarded proteins not annotated in MobiDB. Of the 20386 initial proteins, we analyzed 20030 proteins, i.e. 98.3% of the initial dataset.

Identification of IDPs and IDPRs

To partition the human proteome into different protein variants, we used two parameters: the percentage of disordered residues and the length of the longest disordered domain in the sequence. Based on these two parameters, we defined:

- i. *ordered proteins* (ORDPs): they have less than 30% of disordered residues, no C- or N-terminal segments longer than 30 consecutive disordered residues as well as no segments longer than 40 consecutive disordered residues in positions distinct from the N- and C-terminus;
- ii. *proteins with intrinsically disordered regions* (IDPRs): they have less than 30% of disordered residues in the polypeptide chain and at least either one C- or N-terminal segment longer than 30 consecutive disordered residues or one segment longer than 40 consecutive disordered residues in positions distinct from the N- and C-terminus;
- iii. *intrinsically disordered proteins* (IDPs): they have more than 30% of disordered residues in the polypeptide chain.

Generally speaking, ORDPs are proteins with a limited number of disordered residues and the absence of disordered domains. IDPRs, unlike ORDPs, are proteins with at least one long disordered segment (implying a short half-life [15]) accommodated in globally folded structures. The rationale of taking into account the location of the long disordered segments in the definitions i) and ii) is discussed below in the discussion section. IDPs are proteins with a significant percentage of disorder (implying an unfolded structure and a high susceptibility to proteolytic degradation [47]).

Clustering of the variants of disorder through functional profiles

The functional profiles of ORDPs, IDPRs and IDPs were defined considering several annotations, such as: biological processes, cellular components and molecular functions from the Gene Ontology database [53] and protein classes from the PANTHER database [54–56].

Functional profiles associated with each protein variant are unitary vectors whose components are estimated as the number of proteins of a variant in each functional class divided by the number of proteins in that specific variant. Functional distances between protein variants were evaluated through the Euclidean distance of their profiles. The protein variants were then clustered using the average linkage hierarchical clustering algorithm [57]. The stability of the clustering against noise in the attribution of proteins to the different functional classes was assessed through bootstrap [58]. In each bootstrap re-sampling, new lists of ORDPs, IDPRs, and IDPs were compiled by randomly picking-up (with repetition) proteins in each variant. The robustness of two variants clustering together was estimated by f , the number of times this happened, divided by the number of bootstrap replicates [59]. Similarly, we evaluated the probability that two variants cluster just by chance by evaluating the fraction f of times that the two variants are clustered using resampled lists of ORDPs, IDPRs, and IDPs in which proteins from the entire human proteome are randomly inserted.

Over-representation of a protein variant in a functional class

The fractions of ORDPs, IDPRs, and IDPs were evaluated in each functional class. A protein variant is over-represented in a functional class if the frequency of its proteins in such a class is higher than the frequency of the same variant in the human proteome. The statistical significance of the differences was assessed through a binomial test [54,60]. We did three statistical

analyses, one for each protein variant. A conservative p-value, accounting for the multiple testing, was obtained through Bonferroni correction [61].

Enrichment of a protein variant in a functional class

A functional class is enriched in the protein variant that has the highest frequency and depleted in the protein variant that has the lowest frequency. We evaluated the statistical significance of enrichment/depletion through a cascade of tests. Firstly, we tested the null-hypothesis that the three variants (ORDPs, IDPRs, and IDPs) have the same frequencies in a functional class. Through a goodness-of-fit (chi-square) test we rejected the null hypothesis when the p-value is lower than 0.05. Secondly, we tested the two pairs of variants with the highest and the two pairs with the lowest frequencies with two further goodness-of-fit tests of the hypothesis of equal probabilities. In each of the two tests, if the p-value is lower than 0.05 then the hypothesis of equal frequency of the pairs is discarded. Eventually we concluded that a given functional class is enriched in the protein variant that has the highest significant frequency and depleted in the protein variant that has the lowest significant frequency. Looking at S6 Table it is clear that, if the first test is not passed no conclusion on the enrichment can be made. Once the first test is passed then three cases are possible in each functional class: i) neither significant enrichment nor depletion; ii) either significant enrichment or significant depletion; iii) both significant enrichment and significant depletion.

Results

ORDPs, IDPRs and IDPs in the human proteome

In Table 1 we report numbers and frequencies of ORDPs, IDPRs, and IDPs in the human proteome. Most human proteins are ORDPs, followed by IDPs and IDPRs. The frequency of IDPs (proteins with more than 30% of disordered residues) is about 32%, a result consistent with Colak *et al.* [62]. According to previous estimates [63], the percentage of all proteins containing disorder (IDPRs and IDPs) is roughly 51%. Therefore, the classification of IDPRs either as ordered or disordered is important to correctly estimate the overall percentage of protein intrinsic disorder in the human proteome.

Protein variants: Distance of functional profiles

Functional distances between protein variants are reported in Fig 1 as hierarchical trees (distance matrices are reported in S1 Table) relative to gene ontology biological processes, cellular components, molecular functions, and to the PANTHER protein classes. IDPRs robustly cluster together with ORDPs. We conclude that, as a general trait, IDPRs have functional profiles which are more similar to those of ORDPs than to those of IDPs.

To further check that ORDPs and IDPRs do not cluster together by chance, we compared the hierarchical clusters of ORDPs, IDPRs, and IDPs in Fig 1 with the clusters obtained by

Table 1. Numbers and frequencies of ORDPs, IDPRs, and IDPs in the human proteome.

Variant	Number of proteins	%
ORDPs	9702	49
IDPRs	3862	19
IDPs	6466	32
Total	20030	100

Number and frequency of ORDPs, IDPRs and IDPs in the human proteome.

<https://doi.org/10.1371/journal.pone.0217889.t001>

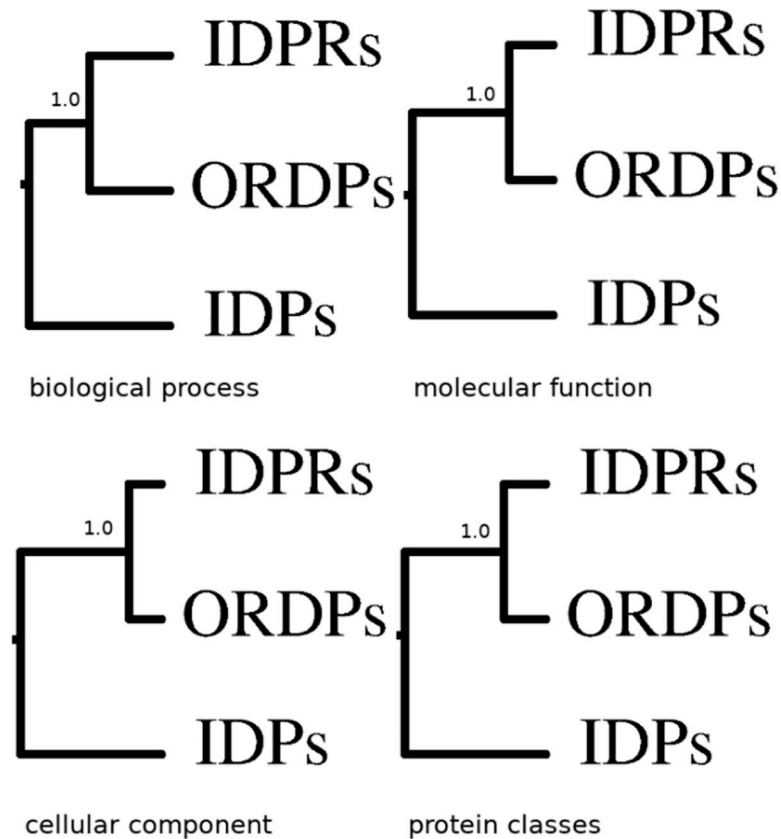


Fig 1. Clusters of the protein variants in the human proteome. Hierarchical trees of ORDPs, IDPRs, and IDPs, based on different functional profiles (i.e. biological processes, molecular functions, cellular components, and PANTHER protein classes) are here reported. The length of the branches are not identical and reflect the distance matrices in [S1 Table](#). In all the four cases, ORDPs and IDPRs are robustly separated, as indicated by bootstrap scores, the number of times these two groups clustered together divided by the number of bootstrap replicates (1000).

<https://doi.org/10.1371/journal.pone.0217889.g001>

repeatedly resampling each variant with proteins randomly selected from the human proteome. f values are comprised between 0.25 and 0.37 (see Figure A in [S1 Text](#)). We, therefore, concluded that the lower distance between the functional profiles of ORDPs and IDPRs, observed in the human proteome ([Fig 1](#)), is due to functional specialization and not by chance.

Protein variants: Over-representation in the functional classes

A protein variant is over-represented (under-represented) in a functional class if its frequency is significantly higher (lower) than the frequency it has in the human proteome. In [Fig 2](#) and [S2 Table](#), we report the over- (under-) representation of ORDPs, IDPRs, and IDPs in the functional classes of PANTHER.

As expected, ORDPs are over-represented among *enzymes* (oxidoreductases, isomerases, lyases, hydrolases, transferases, ligases), and in several other protein classes related to *storage, defense/immunity, transporters, receptors, transfer/carriers and calcium-binding proteins*. ORDPs are under-represented among cytoskeletal proteins, nucleic acid binding proteins and transcription factors.

IDPRs are over-represented among *transporters, cell-junction, cell adhesion, receptors, enzyme modulators, cytoskeletal proteins*. Moreover, they are over-represented in *enzymes* such as hydrolases and transferases. IDPRs are under-represented among transcription factors and

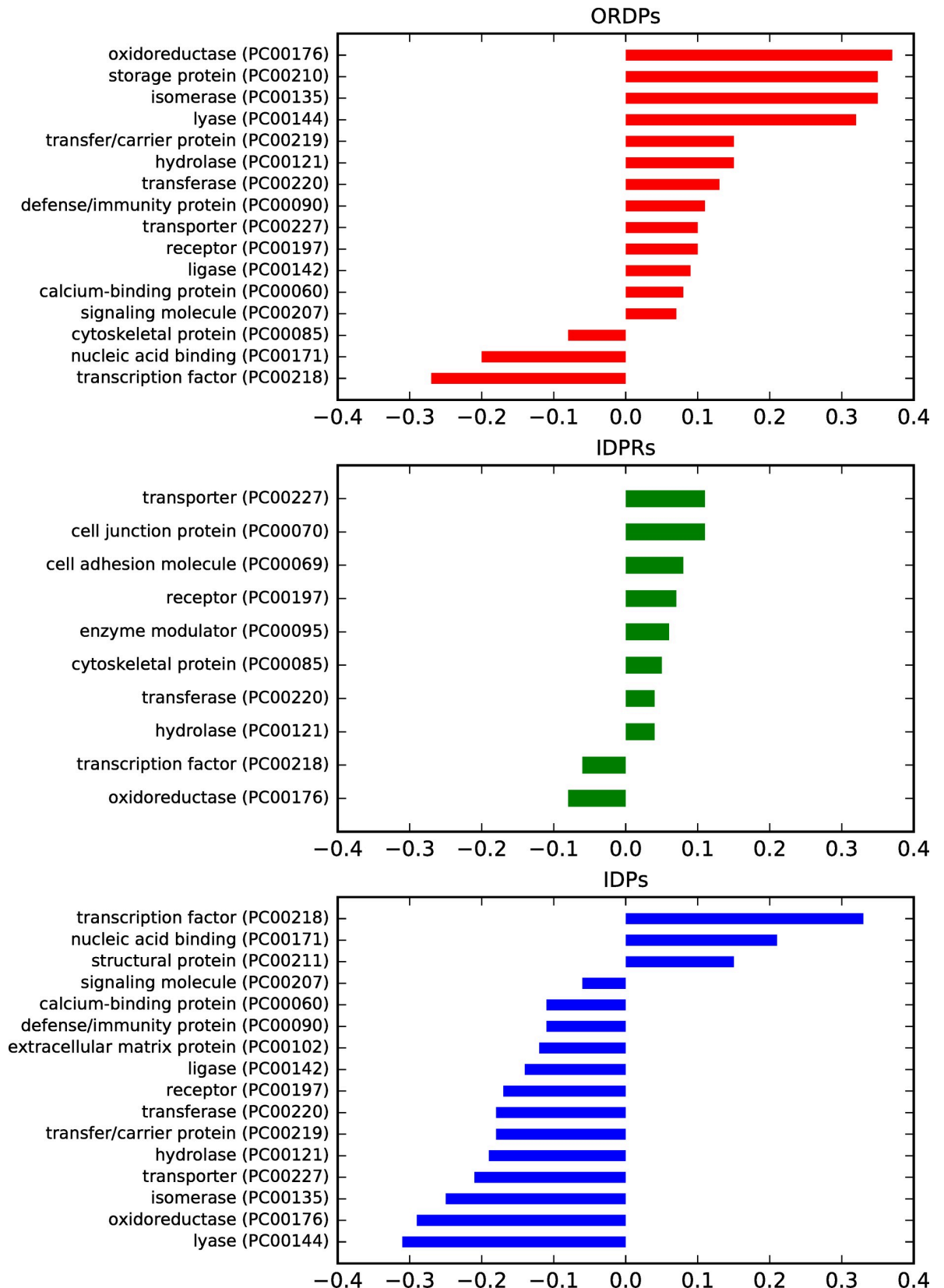


Fig 2. Over- (under-) representation of the protein variants in the PANTHER protein classes. Bar charts of the normalized differential occurrence of ORDPs, IDPRs, and IDPs in various protein classes, with respect to the human proteome (the reference). Only statistically significant differences are reported (p-value <0.05).

<https://doi.org/10.1371/journal.pone.0217889.g002>

oxidoreductases. Notably, both ORDPs and IDPRs are over-represented in transporters, receptors, hydrolases, and transferases; interestingly, IDPs are under-represented in all these latter functional classes.

IDPs are over-represented only among *structural proteins*, *transcription factors*, and *nucleic-acid binding proteins*; note that the two latter functions are not over-represented both in ORDPs and IDPRs.

We also compared the over- (under-) representation of the variants in Gene Ontology annotations related to biological processes, molecular functions and cellular components (for details see: [S1](#), [S2](#) and [S3](#) Figs; [S3](#), [S4](#) and [S5](#) Tables).

In a nutshell, it is worth noting that each variant displays articulated spectra of specific GO annotations, that are generally consistent with the closeness of IDPRs to ORDPs already shown above, based on functional distances and PANTHER classes. On the contrary, IDPs have a well characterized, peculiar, set of annotations. Indeed, IDPs are specifically over-represented in molecular functions related to *nucleic acid binding* and *chromatin binding*, and *developmental biological processes* (i.e., pattern specification processes, ectoderm, embryo, mesoderm and system development), where ORDPs are under-represented and IDPRs are not significantly represented.

Enrichment of functional classes in protein variants

In each protein functional class, the three protein variants (ORDPs, IDPRs, IDPs) are differently distributed ([Fig 3](#)). A functional class is enriched (depleted) in the variant that has the highest (lowest) relative frequency. The statistical significance of the enrichment (depletion) is evaluated with a specific test (see [Materials and Methods](#)). In [Fig 3](#) and [S6 Table](#), we report the protein classes that reached statistical significance in the assessment of their enrichment (depletion). Only two protein classes are enriched in IDPs: *nucleic acid binding proteins* and *transcription factors*. The other classes are enriched in ORDPs. No protein class is enriched in IDPRs.

The results on the enrichment of biological processes, cellular components, and molecular functions in the protein variants are reported in [S7](#), [S8](#) and [S9](#) Tables, respectively. Many development processes (*pattern specification*, *ectoderm*, *mesoderm*, *endoderm*, and *system development*), chromatin and nucleic acid binding are enriched in IDPs whereas ORDPs enrich *anatomical structure morphogenesis*, *death*, and *cell differentiation*. Notably, IDPRs enrich only *calcium-dependent phospholipid binding* and *helicase activity*. Finally, IDPs and IDPRs do not enrich any cellular component.

Over-representation and enrichment of protein variants in diseases

In [Fig 4](#) and [S10 Table](#), we report the over-representation of the protein variants (ORDPs, IDPRs, and IDPs) among the proteins related to cardiovascular diseases, neurodegenerative diseases, and cancer. IDPs are over-represented in cancer and under-represented in cardiovascular diseases. On the contrary, ORDPs are over-represented in cardiovascular diseases and under-represented in cancer. IDPRs are over-represented in all the diseases here considered, and particularly in neurodegenerative diseases.

In [Fig 5](#) and [S11 Table](#), we report the enrichment of protein variants in diseases. Notably, cancer-related proteins are specifically enriched in IDPs. Both, proteins related to cardiovascular diseases and those related to neurodegenerative diseases are enriched in ORDPs. IDPRs are

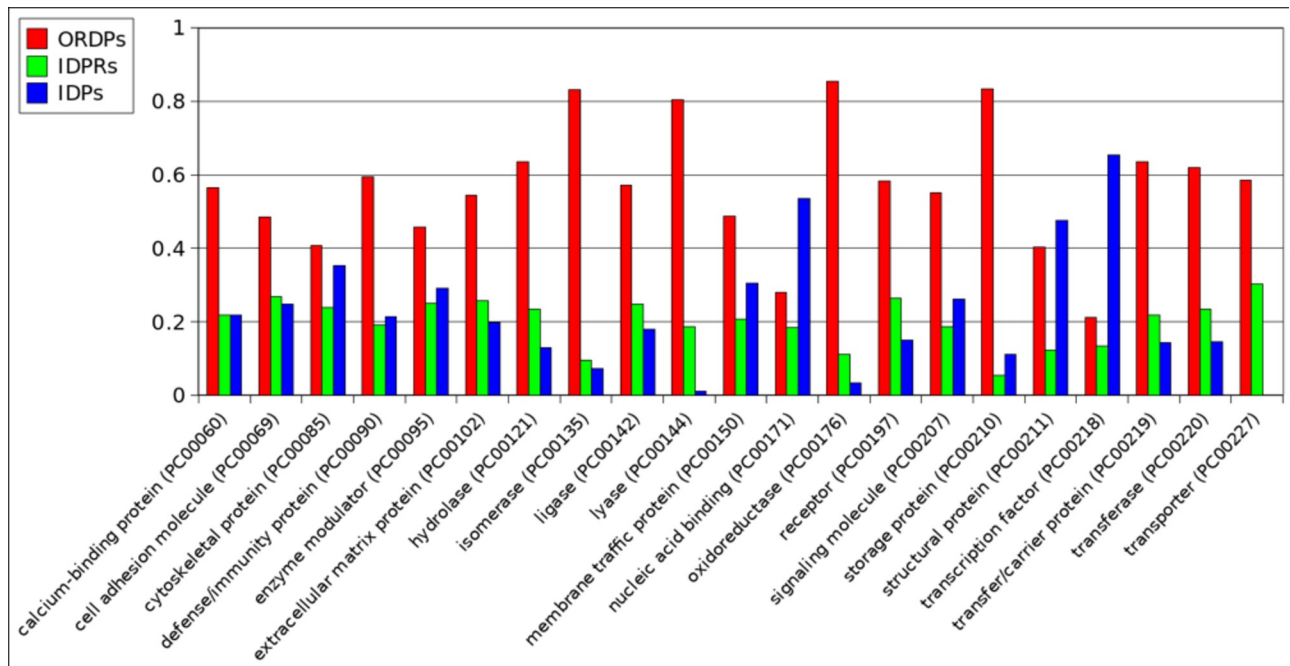


Fig 3. Functional classes are differently enriched in ORDPs, IDPRs, and IDPs. A functional class is enriched (depleted) in the protein variant that has significantly the highest (lowest) frequency. All the classes shown in this figure passed the enrichment (depletion) test for at least one variant (detailed numerical values can be found in S6 Table).

<https://doi.org/10.1371/journal.pone.0217889.g003>

depleted in all three categories of disease-related proteins (not significantly among proteins related to cardiovascular diseases, see S11 Table).

Discussion

In this study we operationally distinguish the functional spectra and biological roles of three broad variants of intrinsic protein disorder: i) *ordered proteins* (ORDPs), ii) *structured proteins with intrinsically disordered protein regions* (IDPRs), and iii) *intrinsically disordered proteins* (IDPs). The focus of this work is on the functional distinction and the different roles in diseases of IDPs and IDPRs, as we distinguish them.

Many studies aiming at a general classification of functional roles of protein intrinsic disorder were based on the required presence of long disordered regions in the sequence, thus considering IDPRs and IDPs as a single entity [19–21,43–46].

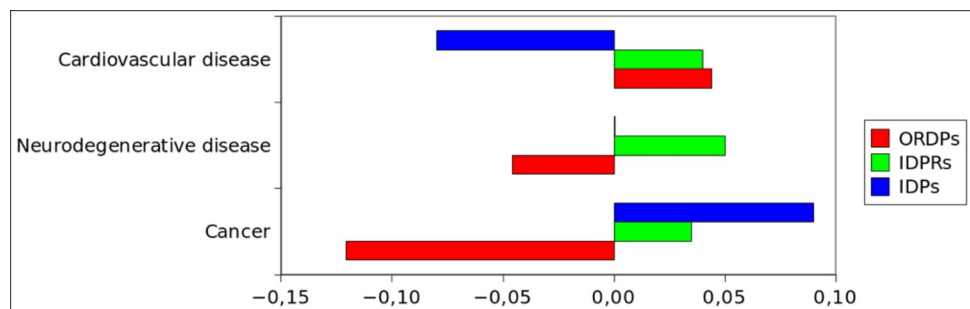


Fig 4. Over-representation of the variants of disorder in diseases. Bar charts of the normalized differential occurrence (with respect to the human proteome) of ORDPs, IDPRs, and IDPs in three groups of disease-related proteins.

<https://doi.org/10.1371/journal.pone.0217889.g004>

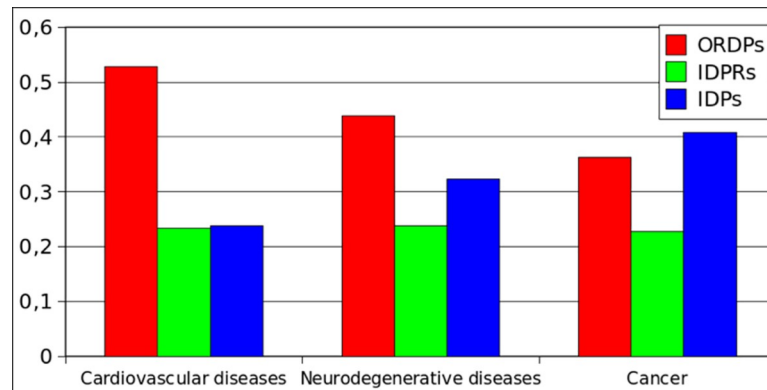


Fig 5. Enrichment of the protein variants in diseases. Different groups of disease-related proteins are enriched (depleted) in the protein variant that has significantly the highest (lowest) frequency. Each group is significantly enriched (depleted) in at least one of the variants (detailed numerical values can be found in S11 Table).

<https://doi.org/10.1371/journal.pone.0217889.g005>

However, the presence of just one long disordered segment (longer than 30 disordered residues) does not imply that a protein is globally intrinsically disordered, i.e., that it lacks a stable three-dimensional structure. With this criterion, it is impossible to discriminate between IDPs and IDPRs. Moreover, the general requirement of 30 consecutive disordered residues as a fingerprint of IDPs is neither experimentally justified nor universally accepted in the literature. Indeed, this length ranges from 20 residues in Necci et al. [64] up to 50 residues, as in Dunker et al. [3] and Ward et al. [44]. Our distinction is based precisely on this point: the difference between IDPs and IDPRs depends on the relative proportion of the unstructured regions. It has been shown that proteins with more than 30% of disordered residues are more prone to proteolytic degradation [47], likely due to the absence of a well-defined tertiary structure. In line with this argument and in agreement with other studies that adopt the same threshold [62,65], we have then defined proteins with more than 30% of disordered residues as IDPs. Finally, to identify IDPRs, we considered the study by van Der Lee et al. [15]. They report that a protein has a shorter half-life (due to the protease-vulnerability of its disordered regions) either if it has a segment of 30 consecutive disordered residues located at the C- or N-terminus or a segment of 40 consecutive disordered residues in a generic position of the polypeptide chain. Moreover, unstructured regions longer than ~25 residues have been considered to evolve, function, and exist as independent, disordered domains [16,66,67]. Thus, we identified IDPRs as those proteins with less than 30% of disordered residues but with at least one disordered domain.

The main result of this study is that intrinsically disordered proteins (IDPs) and structured proteins with intrinsically disordered protein regions (IDPRs) have different functional roles in the cell. We found that IDPs are mainly associated with nucleic acid binding proteins, chromatin binding proteins, transcription factors, and many developmental processes. The functional classes in which IDPRs are over-represented substantially overlap with those in which ORDPs are also over-represented (transporters, receptors, and catalytic activities). This outcome shows that the functional roles of IDPRs are more similar to those of ORDPs than to those of IDPs. Moreover, IDPs are specifically over-represented in PANTHER functional classes and GO annotations where both IDPRs and ORDPs are under-represented. We conclude that the broadly shared statement that IDPs are specifically involved in signaling and cell-regulation [68] should be taken with caution.

A few remarks are in order at this point of the discussion. We have reminded in the introduction that, as a general trait, IDPs do not contain sufficient hydrophobic amino acids [2], which are needed to mediate co-operative folding of polypeptide chains into a well-defined

tertiary structure. Moreover, intrinsic disorder can have different levels of protein structural organization, and whole proteins, or various protein regions can be disordered to a different degree [69]. To reconstruct a perspective, it is also worth reminding that since the time in which the first DNA and protein structures were determined, in the early fifties of the past century, molecular biology and biophysics have been dominated by the static picture brought by the crystallographers. Notwithstanding the importance of static equilibrium structures, the view that there are functional motions in proteins, spanning many scales of time, from femto-seconds to seconds, gradually emerged [70–74]. These works were influential in suggesting a thermodynamic scenario for proteins, in which fluctuations had a relevant role. In particular G. Careri [75] had the intuition that enzyme catalysis could be understood within a thermodynamic setting based on fluctuation-dissipation theorems and on a theory of correlated fluctuations.

The nowadays prevailing picture is that proteins acquire functions within a conformational continuum [15]. Thus, IDPs and IDPRs can be related to dynamic structures that interconvert on multiple timescales, covering a spectrum of states that are in dynamic equilibrium under physiological conditions [76].

ORDPs, IDPRs and IDPs are here coarse-grained operational classifications into which to accommodate different broad functional classes, that are influenced by the degree of disorder [77]. Now, it is quite clear that a polypeptide chain can encode for highly heterogeneous (in space and time) protein molecules containing unique sets of *foldons* (i.e. ordered segments), *semi-foldons* (i.e., partially folded protein segments with transient residual structure), *inducible foldons* (i.e., segments that are able to fold at interaction with specific binding partners), *non-foldons* (i.e., protein segments with functions depending on their intrinsically disordered nature), and *unfoldons* (i.e., segments that are able to unfold under specific conditions, becoming functional) [69,77]. Of this mosaic of units we do not have yet a unifying (equilibrium/ nonequilibrium) thermodynamic theory, based on basic statistical models. Finding these models, possibly in a mesoscopic stochastic theory of heterogeneous quasi-random polymers (see e.g. [78]), could provide an understanding of the operational rules we used to define ORDs, IDPRs and IDPs.

As regards to disease-related proteins, we observe that IDPs are abundant among cancer-related proteins but not among proteins associated with cardiovascular and neurodegenerative diseases, where ORDPs seem to dominate. This observation better specifies the quite established D^2 concept [33].

Overall, our approach is useful to discriminate between IDPs and IDPRs operationally, and we have used it to demonstrate the different functional roles of these two variants in the human proteome. As a conclusion, we believe that it is important to correctly evaluate the occurrence of disorder in proteomes, in evolutionary studies, in understanding functional aspects related to different flavors of protein disorder, and in deciphering the role of these variants in the emergence of diseases.

Supporting information

S1 Fig. Over- (under-) representation of the protein variants in the biological processes.

Bar charts of the normalized differential occurrence of ORDPs, IDPRs, and IDPs in various biological processes, with respect to the human proteome (the reference). Only statistically significant differences are reported (p -value < 0.05).

(TIF)

S2 Fig. Over- (under-) representation of the protein variants in the molecular functions.

Bar charts of the normalized differential occurrence of ORDPs, IDPRs, and IDPs in various

molecular functions, with respect to the human proteome (the reference). Only statistically significant differences are reported (p-value <0.05).

(TIF)

S3 Fig. Over- (under-) representation of the protein variants in the cellular components.

Bar charts of the normalized differential occurrence of ORDPs, IDPRs, and IDPs in various cellular components, with respect to the human proteome (the reference). Only statistically significant differences are reported (p-value <0.05).

(TIF)

S1 Table. Distance matrices between ORDPs, IDPRs and IDPs for different functional roles.

(DOCX)

S2 Table. Over-representation and under-representation of the protein variants in the protein classes of PANTHER.

(DOCX)

S3 Table. Over-representation and under-representation of the protein variants in the biological processes. The asterisk near some biological process indicates that the over-representation test is not statistically significant.

(DOCX)

S4 Table. Over-representation and under-representation of the protein variants in the molecular functions. The asterisk near some molecular function indicates that the over-representation test is not statistically significant.

(DOCX)

S5 Table. Over-representation and under-representation of the protein variants in the cellular components.

(DOCX)

S6 Table. Enrichment of the protein variants in the protein classes of PANTHER. The protein classes for which the enrichment test is not statistically significant are reported in red.

(DOCX)

S7 Table. Enrichment of the protein variants in the biological processes. The biological processes for which the enrichment test is not statistically significant are reported in red.

(DOCX)

S8 Table. Enrichment of the protein variants in the molecular functions. The molecular functions for which the enrichment test is not statistically significant are reported in red.

(DOCX)

S9 Table. Enrichment of the protein variants in the cellular components. The cellular components for which the enrichment test is not statistically significant are reported in red.

(DOCX)

S10 Table. Over-representation of the protein variants in diseases. The diseases for which the over-representation test is not statistically significant are reported in red.

(DOCX)

S11 Table. Enrichment of the protein variants in diseases. The diseases for which the enrichment test is not statistically significant are reported in red.

(DOCX)

S1 Text. Hierarchical clusters of protein functional profiles for protein variants populated with proteins randomly selected from the human proteome.

(DOCX)

S2 Text. UniProt ID of cancer-related proteins. Lists of the cancer-related proteins found in Iakoucheva *et al.* [49] and considered in this work.

(TXT)

S3 Text. UniProt IDs of cardiovascular-related proteins. Lists of the cardiovascular-related proteins found in Cheng *et al.* [50] and considered in this work.

(TXT)

S4 Text. UniProt IDs of neurodegenerative-related proteins. Lists of the neurodegenerative-related proteins found in Uversky [51] and considered in this work.

(TXT)

Author Contributions

Conceptualization: Antonio Deiana, Alessandro Porrello, Andrea Giansanti.

Data curation: Antonio Deiana, Sergio Forcelloni.

Formal analysis: Antonio Deiana, Alessandro Porrello, Andrea Giansanti.

Methodology: Antonio Deiana, Alessandro Porrello, Andrea Giansanti.

Software: Antonio Deiana, Sergio Forcelloni.

Supervision: Andrea Giansanti.

Validation: Sergio Forcelloni.

Visualization: Sergio Forcelloni.

Writing – original draft: Antonio Deiana, Andrea Giansanti.

Writing – review & editing: Sergio Forcelloni, Alessandro Porrello, Andrea Giansanti.

References

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol.* 1999; 31:293–321.
2. Uversky VN, Gillespie JR, Fink AL. Why Are “Natively Unfolded” Proteins Unstructured Under Physiologic Conditions? *PROTEINS: Structure, Function, and Genetics* 2000; 41:415–427.
3. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *J Mol Graph Model.* 2001; 19:26–59. PMID: [11381529](#)
4. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci.* 2002; 27:527–33. PMID: [12368089](#)
5. Fink AL. Natively unfolded proteins. *Curr Opin Struct Biol.* 2005; 15:35–41. <https://doi.org/10.1016/j.sbi.2005.01.002> PMID: [15718131](#)
6. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker KA. Natively Disordered Proteins. In: *Protein Folding Handbook.* Wiley-Blackwell; 2008.
7. Tompa P, Fersht A. Structure and function of intrinsically disordered proteins. New York: Chapman and Hall/CRC; 2009.
8. Schlessinger A, Schaefer C, Vicedo E, Schmidberger M, Punta M, Rost B. Protein disorder—a breakthrough invention of evolution?. *Curr Opin Struct Biol.* 2011; 21:412–418. <https://doi.org/10.1016/j.sbi.2011.03.014> PMID: [21514145](#)
9. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions.* 2016; 44:1185–1200. <https://doi.org/10.1042/BST20160172> PMID: [27911701](#)

10. Uversky VN. Intrinsic Disorder, Protein–Protein Interactions, and Disease. *Advances in Protein Chemistry and Structural Biology*. 2017;1876–1623.
11. Uversky VN. Paradoxes and wonders of intrinsic disorder: Stability of instability. *Intrinsically Disordered Proteins*. 2017; 5(1):e1327757. <https://doi.org/10.1080/21690707.2017.1327757> PMID: 30250771
12. Le Gall T, Romero PR, Cortese MS, Uversky VN, Dunker AK. Intrinsic disorder in the Protein Data Bank. *J Biomol Struct Dyn*. 2007; 24(4):325–42. <https://doi.org/10.1080/07391102.2007.10507123> PMID: 17206849
13. DeForte S, Uversky VN. Resolving the ambiguity: Making sense of intrinsic disorder when PDB structures disagree. *Protein Sci*. 2015; 25(3):676–688.
14. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta*. 2010; 1804:1231–64. <https://doi.org/10.1016/j.bbapap.2010.01.017> PMID: 20117254
15. Van der Lee R, Lang B, Kruse K, Gsponer J, Sanchez de Groot N, Huynen MA, et al. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep*. 2014; 8:1832–44. <https://doi.org/10.1016/j.celrep.2014.07.055> PMID: 25220455
16. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem*. 2014; 83:553–84. <https://doi.org/10.1146/annurev-biochem-072711-164947> PMID: 24606139
17. Uversky VN. Intrinsic disorder here, there, and everywhere, and nowhere to escape from it. *Cell Mol Life Sci*. 2017; 74:3065–3067. <https://doi.org/10.1007/s00018-017-2554-5> PMID: 28589440
18. Dunker K, Babu MM, Barbar E, Blackledge M, Bondos SE, Dosztányi Z et al. What's in a name? Why these proteins are intrinsically disordered. *Intrinsically Disordered Proteins*. 2013; 1:e24157. <https://doi.org/10.4161/idp.24157> PMID: 28516007
19. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res*. 2007; 6:1882–98. <https://doi.org/10.1021/pr060392u> PMID: 17391014
20. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, et al. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J Proteome Res*. 2007; 6:1899–916. <https://doi.org/10.1021/pr060393m> PMID: 17391015
21. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, et al. Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res*. 2007; 6:1917–32. <https://doi.org/10.1021/pr060394e> PMID: 17391016
22. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, et al. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*. 2008; 9 Suppl 2:S1.
23. Yan J, Mizianty MJ, Filipow PL, Uversky VN, Kurgan L. RAPID: fast and accurate sequence-based prediction of intrinsic disorder content on proteomic scale. *Biochim Biophys Acta*. 2013; 1834:1671–80. <https://doi.org/10.1016/j.bbapap.2013.05.022> PMID: 23732563
24. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005; 18:343–84. <https://doi.org/10.1002/jmr.747> PMID: 16094605
25. Garza AS, Ahmad N, Kumar R. Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sci*. 2009; 84:189–93. <https://doi.org/10.1016/j.lfs.2008.12.002> PMID: 19109982
26. Uversky VN. The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome. *J Biomed Biotechnol*. 2010; 2010:568068. <https://doi.org/10.1155/2010/568068> PMID: 20011072
27. Dyson HJ, Wright PE. Intrinsically Unstructured Proteins and Their Functions. *Nat Rev*. 2005; 6:197–208.
28. Buljan M, Chalancon G, Dunker AK, Bateman A, Balaji S, Fuxreiter M, et al. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr Opin Struct Biol*. 2013; 23:443–50. <https://doi.org/10.1016/j.sbi.2013.03.006> PMID: 23706950
29. Hegyi H, Kalmar L, Horvath T, Tompa P. Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. *Nucleic Acids Res*. 2011; 39:1208–19. <https://doi.org/10.1093/nar/gkq843> PMID: 20972208
30. Zhou J, Zhao S, Dunker AK. Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J Mol Biol*. 2018; 430:2342–59. <https://doi.org/10.1016/j.jmb.2018.03.028> PMID: 29626537
31. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, et al. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*. 2009; 10 Suppl 1:S7.

32. Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, et al. Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. *Chem Rev*. 2014; 114:6844–79. <https://doi.org/10.1021/cr400713r> PMID: 24830552
33. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys*. 2008; 37:215–46. <https://doi.org/10.1146/annurev.biophys.37.032807.125924> PMID: 18573080
34. Darling AL, Uversky VN. Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Front Genet*. 2018; 9:158. <https://doi.org/10.3389/fgene.2018.00158> PMID: 29780404
35. Arai M. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophys Rev*. 2018; 10:163–81. <https://doi.org/10.1007/s12551-017-0346-7> PMID: 29307002
36. Fisher E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft*. 1894, 27:2985–93.
37. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB et al. Extreme disorder in an ultra-high-affinity protein complex. *Nature*. 2018; 555(7694):61–66. <https://doi.org/10.1038/nature25762> PMID: 29466338
38. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002; 11:739–56. <https://doi.org/10.1110/ps.4210102> PMID: 11910019
39. Kulkarni PK, Uversky VN. Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *Proteomics*. 2018; 18:1800061.
40. Gao J, Xu D. Correlation between posttranslational modification and intrinsic disorder in protein. *Pac Symp Biocomput*. 2012;94–103. PMID: 22174266
41. Pejaver V, Hsu W-L, Xin F, Dunker AK, Radivojac P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci*. 2014; 23:1077–93. <https://doi.org/10.1002/pro.2494> PMID: 24888500
42. Burgi J, Xue B, Uversky VN, van der Goot FG. Intrinsic Disorder in Transmembrane Proteins: Roles in Signaling and Topology Prediction. *PLoS One*. 2016; 11:e0158594. <https://doi.org/10.1371/journal.pone.0158594> PMID: 27391701
43. Tompa P, Fersht A. Structure and function of intrinsically disordered proteins. CRC Press, Taylor & Francis Group; 2010.
44. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. 2004; 20:2138–2139. <https://doi.org/10.1093/bioinformatics/bth195> PMID: 15044227
45. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*. 2004; 337:635–45. <https://doi.org/10.1016/j.jmb.2004.02.002> PMID: 15019783
46. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci*. 2015; 72(1):137–51. <https://doi.org/10.1007/s00018-014-1661-9> PMID: 24939692
47. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*. 2008; 322:1365–8. <https://doi.org/10.1126/science.1163581> PMID: 19039133
48. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017; 45: D158–69. <https://doi.org/10.1093/nar/gkw1099> PMID: 27899622
49. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002; 323:573–84. [https://doi.org/10.1016/s0022-2836\(02\)00969-5](https://doi.org/10.1016/s0022-2836(02)00969-5) PMID: 12381310
50. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry*. 2006; 45:10448–60. <https://doi.org/10.1021/bi060981d> PMID: 16939197
51. Uversky VN. The triple power of D(3): protein intrinsic disorder in degenerative diseases. *Front Biosci (Landmark Ed)*. 2014; 19:181–258.
52. Piovesan D, Tabaro F, Paladin L, Necci M, Micetic I, Camilloni C, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res*. 2018; 46: D471–6. <https://doi.org/10.1093/nar/gkx1071> PMID: 29136219
53. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–9. <https://doi.org/10.1038/75556> PMID: 10802651

54. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc.* 2013; 8:1551–66. <https://doi.org/10.1038/nprot.2013.092> PMID: 23868073
55. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* 2016; 44:D336–342. <https://doi.org/10.1093/nar/gkv1194> PMID: 26578592
56. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 2017; 45:D183–9. <https://doi.org/10.1093/nar/gkw1138> PMID: 27899595
57. Webb AR. *Statistical Pattern Recognition*, 2nd Edition. John Wiley & Sons; 2002.
58. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Boca Raton, Florida, USA: Chapman & Hall/CRC; 1993.
59. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985; 39:783–91. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x> PMID: 28561359
60. Armitage P. *Statistical methods in medical research*. Blackwell Scientific Oxford; 1971.
61. Bland M. *An introduction to medical statistics*. Oxford University Press Oxford; 1987.
62. Colak R, Kim T, Michaut M, Sun M, Irimia M, Bellay J, et al. Distinct types of disorder in the human proteome: functional implications for alternative splicing. *PLoS Comput Biol.* 2013; 9:e1003030. <https://doi.org/10.1371/journal.pcbi.1003030> PMID: 23633940
63. Niklas KJ, Dunker AK, and Yruela I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *Journal of Experimental Botany*, 2018; 69,1437–1446. <https://doi.org/10.1093/jxb/erx493> PMID: 29394379
64. Necci M, Piovesan D, Tosatto SCE. Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Sci.* 2016; 25:2164–74. <https://doi.org/10.1002/pro.3041> PMID: 27636733
65. Bellay J, Han S, Michaut M, Kim T, Costanzo M, Andrews BJ, et al. Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* 2011; 12:R14. <https://doi.org/10.1186/gb-2011-12-2-r14> PMID: 21324131
66. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J Proteome Res.* 2006; 5,879–887. <https://doi.org/10.1021/pr060048x> PMID: 16602695
67. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, and Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays.* 2009; 31,328–3. <https://doi.org/10.1002/bies.200800151> PMID: 19260013
68. Wright PE, Dyson HJ. Intrinsically Disordered Proteins in Cellular Signaling and Regulation. *Nat Rev Mol Cell Biol.* 2015; 16:18–29. <https://doi.org/10.1038/nrm3920> PMID: 25531225
69. Uversky VN. Under-Folded Proteins: conformational Ensembles and Their Roles in Protein Folding, Function, and Pathogenesis. *Biopolymers* 2013; 99(11):870–887. <https://doi.org/10.1002/bip.22298> PMID: 23754493
70. Weber G. Energetics of ligand binding to proteins. *Adv. Prot. Chem.* 1975; 29:1–83 (1975).
71. Careri G, Gratton E, Fasella P. Enzyme Dynamics: The Statistical Physics Approach. *Ann. Rev. Biophysics Bioeng.* 8, 69–97 (1979).
72. Frauenfelder H, Petsko G, Tsernoglou D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* 1979; 280:5558–5563.
73. Brooks CI, Karplus M, Petitt B, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Wiley, New York, 1988.
74. Kay LE. Protein dynamics from NMR. *Nature. Struc. Biol. NMR Suppl.* 1998; S5:513–517.
75. Careri G. “The Fluctuating Enzyme” in *Quantum Statistical Mechanics in the Natural Sciences*, Springer US. New York; 1974, pp. 15–35.
76. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol.* 2007; 65(3):277–88. <https://doi.org/10.1007/s00239-007-9011-2> PMID: 17721672
77. Uversky VN. Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* 2015; 282(7):1182–9. <https://doi.org/10.1111/febs.13202> PMID: 25631540
78. Zhang X-J, Qian H, Qian M. Stochastic theory of nonequilibrium steady states and its applications. Part I. *Phys. Rep.* 2012; 510:1–86.